

Novel Concepts for the Simulation and Experimental Demonstration of High-Quality Laser-Plasma Acceleration

by

MANUEL KIRCHEN

A Thesis Submitted to the Faculty of
Mathematics, Informatics and Natural Sciences
Department of Physics
for the Degree of

DOCTOR RERUM NATURALIUM

UNIVERSITÄT HAMBURG

August 2021

Gutachter der Dissertation:

Dr. Andreas Maier

Prof. Dr. Wim Leemans

Zusammensetzung der Prüfungskommission:

Dr. Andreas Maier

Prof. Dr. Wim Leemans

Dr. Jens Osterhoff

Prof. Dr. Peter Hauschildt

Prof. Dr. Daniela Pfannkuche

Vorsitzende der Prüfungskommission:

Prof. Dr. Daniela Pfannkuche

Datum der Disputation:

17. Dezember 2021

Vorsitzender des Fach-Promotionsausschusses Physik: Prof. Dr. Wolfgang Hansen

Leiter des Fachbereichs Physik:

Prof. Dr. Günter H. W. Sigl

Dekan der Fakultät MIN:

Prof. Dr. Heinrich Graener

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium.

Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 17.08.2021

Manuel Kirchen

Abstract

In a laser-plasma accelerator (LPA), an intense laser pulse drives a plasma wave that can trap and accelerate electrons at orders of magnitude larger accelerating gradients than modern radiofrequency technology. LPAs promise compact sources of high-brightness electron beams that would benefit a variety of applications in science, medicine and industry. Most of these applications demand low energy spread beams and high-efficiency operation. Achieving both requires to flatten the accelerating fields via controlled beam loading, which refers to an effect by which the shape of the plasma wave changes as the electrons absorb energy from it.

This thesis reports on the experimental demonstration of such optimal beam loading. A novel plasma source was developed that uses localized ionization injection to generate high-quality electron bunches. By balancing laser and plasma parameters the setup allows optimizing beam loading through the combined effect of injected bunch current profile and plasma wave evolution. Operation at optimal conditions leads to flat average accelerating fields, which enabled the generation of one percent level energy spread electron bunches with tens of pC charge that were efficiently accelerated to a few hundred MeV energy. Using machine learning, a predictive model was built that quantifies how shot-to-shot variations in beam loading, and thus the beam quality, depend on subtle fluctuations of the drive laser. The model allows to assess how specific improvements to the laser performance would affect the LPA stability.

A central component of the design and evaluation of the experiment were extensive simulation studies using the spectral, quasi-3D particle-in-cell code `FBPIC` whose development this work also contributed to. To run simulations in an optimal Lorentz-boosted frame and with that reduce the computational cost by orders of magnitude, a novel algorithm was invented that represents the particle-in-cell equations in Galilean (comoving) coordinates and thereby eliminates the numerical Cherenkov instability when modeling relativistically streaming plasmas. In summary, the numerical concepts presented in this thesis enable faster and more accurate simulations that, together with the insights from the experiment, provide a basis for designing and operating LPAs with an application-relevant quality and stability.

Zusammenfassung

In einem Laser-Plasma-Beschleuniger (LPA) treibt ein intensiver Laserpuls eine Plasmawelle an, die Elektronen einfangen und mit um Größenordnungen stärkeren Gradienten beschleunigen kann als moderne Hochfrequenztechnik. LPB versprechen kompakte Quellen von Elektronenstrahlen hoher Helligkeit, die für eine Vielzahl von Anwendungen in Wissenschaft, Medizin und Industrie von Nutzen wären. Die meisten dieser Anwendungen setzen Strahlen mit einer geringen Energiebreite und einen hocheffizienten Betrieb voraus. Um beides zu erreichen, müssen die beschleunigenden Felder durch kontrolliertes Beam Loading abgeflacht werden. Beam Loading bezeichnet einen Effekt, bei dem sich die Form der Plasmawelle durch Energieabsorption der Elektronen ändert.

Diese Arbeit berichtet über die experimentelle Demonstration von solch optimalem Beam Loading. Es wurde eine neuartige Plasmaquelle entwickelt, die mittels lokalisierter Ionisationsinjektion hochwertige Elektronenpakete erzeugt. Durch die Abstimmung von Laser- und Plasmaparametern ermöglicht der Aufbau Beam Loading durch den kombinierten Effekt aus injiziertem Stromprofil des Elektronenpakets und Evolution der Plasmawelle zu optimieren. Der Betrieb bei optimalen Bedingungen führt zu flachen mittleren Beschleunigungsfeldern, was die Erzeugung von Elektronenpaketen mit einer Energiebreite von ungefähr einem Prozent rms und einigen zehn pC Ladung ermöglichte, die effizient auf eine Energie von einigen hundert MeV beschleunigt wurden. Mithilfe von maschinellem Lernen wurde ein Modell erstellt, das quantifiziert, wie Schuss-zu-Schuss-Schwankungen des Beam Loading, und damit der Strahlqualität, von geringfügigen Fluktuationen des Antriebslasers abhängen. Anhand des Modells lässt sich beurteilen, wie sich spezifische Verbesserungen der Laserleistung auf die Stabilität des LPB auswirken würden.

Ein zentraler Bestandteil der Konzeption und Auswertung des Experiments waren umfangreiche Simulationsstudien mit dem spektralen, quasi-3D Particle-in-Cell-Code FBPIC, zu dessen Entwicklung diese Arbeit ebenfalls beigetragen hat. Um die Simulationen in einem optimalen Lorentz-geboosteten Bezugssystem durchführen zu können, und damit die Rechenkosten um Größenordnungen zu reduzieren, wurde ein neuartiger Algorithmus entwickelt, der die Particle-In-Cell-Gleichungen in Galileischen (sich mitbewegenden) Koordinaten darstellt und dadurch die numerische Cherenkov Instabilität bei der Modellierung relativistisch strömender Plasmen eliminiert. Zusammengekommen ermöglichen die in dieser Arbeit vorgestellten numerischen Konzepte schnellere und genauere Simulationen, die gemeinsam mit den Erkenntnissen aus dem Experiment eine Grundlage für die Entwicklung und den Betrieb von LPB mit einer anwendungsrelevanten Qualität und Stabilität bilden.

Contents

Introduction	17
Publications	23
1 Physics of laser-plasma acceleration	25
1.1 Laser-driven plasma waves	25
1.1.1 Acceleration in plasma	25
1.1.2 Ponderomotive force	27
1.1.3 Laser pulse propagation in plasma	29
1.1.4 Plasma wave excitation	32
1.1.5 Regimes of laser-plasma acceleration	34
1.2 Electron acceleration	37
1.2.1 Injection	38
1.2.2 Acceleration limits	42
1.2.3 Beam loading	43
I Generation of stable high-quality electron beams	49
2 LUX laser-plasma accelerator	51
2.1 Overview	51
2.2 Drive laser	52
2.2.1 ANGUS laser system and beam delivery	52
2.2.2 Laser diagnostics	54
2.2.3 Low-power characterization	56
2.2.4 High-power online measurements	60
2.3 Electron beam line	62
2.3.1 Target area and electron beam transport	62
2.3.2 Electron beam diagnostics	63
3 Development of a localized ionization injection plasma source	67
3.1 Plasma source	67
3.1.1 Motivation and previous work	67
3.1.2 Conceptual design	70
3.1.3 Experimental realization	70
3.1.4 Gas supply and monitoring system	72

3.2	Gas and plasma profile	73
3.2.1	CFD simulations	73
3.2.2	Dopant gas localization	77
4	Demonstration of optimal beam loading	79
4.1	Overview	79
4.2	Injection and acceleration dynamics	82
4.2.1	Particle-in-cell simulation setup	82
4.2.2	Electron beam generation at the optimum	83
4.2.3	Influence of laser and plasma parameters	84
4.3	Beam loading physics	86
4.3.1	Phase space shearing and optimal beam loading	87
4.3.2	Beam loading correlations	88
4.3.3	Characterization of the optimum	89
4.4	Influence of drive laser instabilities	90
4.4.1	Laser-electron correlations	90
4.4.2	Machine-learning-based surrogate model	92
4.4.3	Beam quality prediction	93
4.5	Summary and outlook	95
II	Development of novel particle-in-cell methods	97
5	The electromagnetic particle-in-cell method	99
5.1	Overview of the particle-in-cell algorithm	99
5.1.1	Discretization	101
5.1.2	Macroparticles and shape factors	103
5.2	The particle-in-cell cycle	104
5.2.1	Field gathering and particle push	104
5.2.2	Charge and current deposition	106
5.2.3	Field solving	107
5.3	Maxwell solvers	108
5.3.1	Finite-difference time-domain	108
5.3.2	Pseudospectral time-domain	111
5.3.3	Pseudospectral analytical time-domain	113
5.3.4	Arbitrary-order pseudospectral solvers	114
5.4	Numerical Cherenkov	117
5.4.1	Numerical Cherenkov radiation	117
5.4.2	Numerical Cherenkov instability	120

6	FBPIC – A spectral, quasi-3D particle-in-cell code	123
6.1	Spectral quasi-cylindrical algorithm	123
6.1.1	Fourier-Hankel representation	124
6.1.2	Overview of the algorithm	127
6.2	Numerical implementation	129
6.2.1	Discretization and transformations	129
6.2.2	Field gathering and particle push	131
6.2.3	Charge and current deposition	133
6.2.4	Quasi-cylindrical pseudospectral analytical time domain	135
6.2.5	Open boundary conditions and moving window	137
6.2.6	Advanced features	140
6.2.7	Practical implementation	141
6.3	Parallelization	142
6.3.1	Intra-device parallelization	143
6.3.2	Inter-device parallelization	147
6.3.3	Parallel scalability	150
7	Elimination of the numerical Cherenkov instability	153
7.1	Stable discrete representation of relativistically streaming plasmas . . .	153
7.1.1	Previous NCI mitigation strategies	154
7.1.2	Galilean scheme	154
7.2	Galilean-PSATD solver of arbitrary order	156
7.2.1	Particle-in-cell equations in Galilean coordinates	156
7.2.2	Derivation of the discretized Maxwell equations	156
7.2.3	Practical implementation	159
7.2.4	Locality with a finite-order stencil	160
7.2.5	Asymmetry of the vacuum dispersion relation	161
7.3	Theoretical stability analysis	163
7.3.1	Numerical dispersion relation of a cold relativistic plasma . . .	163
7.3.2	Evaluation of the NCI growth rate	165
8	Application to plasma acceleration	171
8.1	Stable Lorentz-boosted frame simulations	171
8.1.1	Lorentz-boosted frame technique	171
8.1.2	NCI suppression with the Galilean-PSATD solver	173
8.2	Comparison of FBPIC to a conventional 3D algorithm	177
Conclusion		183
Bibliography		187
Acknowledgement		209

List of Figures

1.1	1D laser-driven wakefield	34
1.2	Regimes of laser-plasma acceleration	36
1.3	Electron motion in a laser-driven plasma wakefield	39
1.4	Principle of ionization injection	41
1.5	Beam loading in the linear regime	45
1.6	Beam loading in the nonlinear regime	47
2.1	LUX laser-plasma accelerator	52
2.2	Laser diagnostics	54
2.3	Focal spot evolution	57
2.4	Flattened Gaussian laser model	58
2.5	Pulse duration statistics	59
2.6	Laser ionization trace	59
2.7	Drive laser stability	60
2.8	Wavefront stability	61
2.9	Spectrometer charge calibration	65
3.1	Plasma source for localized ionization injection	69
3.2	Rendering and images of the plasma source	71
3.3	CFD simulation of the plasma source	75
3.4	Gas and plasma density profile	76
3.5	Dopant gas localization	78
4.1	High-quality electron beam spectrum and profile	81
4.2	Particle-in-cell simulation of the injection and acceleration process	83
4.3	Influence of the laser and plasma source parameters	85
4.4	Phase space shearing and optimal beam loading	87
4.5	Beam loading correlations	88
4.6	Electron beam quality at the optimal load	89
4.7	Influence of the drive laser	91
4.8	Neural network training loss	93
4.9	Beam quality prediction	94
4.10	Optimal beam loading scaled to GeV energies	96
5.1	Overview of the particle-in-cell method	100
5.2	Space and time discretization	102
5.3	Particle-in-cell cycle	105
5.4	Numerical dispersion relation of the Yee solver	109
5.5	Numerical dispersion relation of the PSTD solver	112
5.6	Numerical dispersion relation of the arbitrary-order PSATD solver	116

5.7	Emission of numerical Cherenkov radiation	118
5.8	Degradation of the beam quality due to numerical Cherenkov radiation	119
5.9	Example of a numerical Cherenkov instability	121
6.1	Overview of the FBPIC algorithm	127
6.2	Grid layout and discretization	129
6.3	Open boundary conditions and moving window	138
6.4	Two-level parallelization strategy	142
6.5	Degree of data parallelism of the particle-in-cell routines	144
6.6	Exchange of information between sub-domains	147
6.7	Effective stencil width of the finite-order PSATD solver	149
6.8	Strong scaling of FBPIC	151
6.9	Weak scaling efficiency of FBPIC	152
7.1	Illustration of the Galilean scheme	154
7.2	Effective stencil width of the finite-order Galilean-PSATD solver	161
7.3	Asymmetric dispersion relation of the Galilean-PSATD solver	162
7.4	Distorted electromagnetic and spurious beam mode intersections	166
7.5	Theoretically predicted and simulated NCI growth rates	170
8.1	Lorentz-boosted frame technique	172
8.2	Stability of Lorentz-boosted frame simulations of plasma acceleration	174
8.3	Lorentz-boosted frame comparison: Plasma wakefield	175
8.4	Lorentz-boosted frame comparison: Laser and beam parameters	176
8.5	Comparison of FBPIC with WARPX 3D: Gas density profile	178
8.6	Comparison of FBPIC with WARPX 3D: Plasma wakefield	179
8.7	Comparison of FBPIC with WARPX 3D: Laser and beam parameters	180
8.8	Comparison of FBPIC with WARPX 3D: Computational cost	181
9.1	Usage of FBPIC in the plasma acceleration community	185

List of Tables

6.1	Formal analogy spectral Cartesian and spectral quasi-cylindrical	136
7.1	Numerical parameters of the NCI growth rate analysis	167
8.1	Lorentz-boosted frame comparison: Simulation parameters	173

Introduction

Many discoveries in science are intimately connected to the development of machines for imaging and probing physical processes on their natural length and time scales. Brilliant light sources, such as synchrotrons or free-electron lasers, produce high brightness radiation, from the terahertz to the X-ray regime, that is used to study the dynamics and structure of matter at the atomic and molecular level, bringing advances in life sciences and chemistry. Particle colliders go one step further as they allow to explore the subatomic realm of matter and provide answers to the most fundamental questions about the universe. At the heart of these machines are accelerators that provide highly energetic beams of charged particles, such as electrons, positrons or protons.

The quest for resolving ever shorter length and time scales has led to an exponential growth in beam energy since the advent of particle accelerators a century ago—a development that entailed a steady increase of the size and cost of these machines. The latest generation of high-energy particle accelerators built or envisioned are already on the edge of feasibility, as they are kilometer-scale facilities that cost billions of dollars. To a large extent, this increase in size and cost can be attributed to the fact that the accelerating gradient in conventional accelerators, which are based on radiofrequency cavities, is limited to ~ 100 MV/m mainly due to material breakdown. Hence, pushing the energy frontier further or reducing the footprint of future machines inevitably requires the development of novel accelerator technology capable to deliver much higher gradients.

Besides its relevance for fundamental research, a more compact and cost-efficient technology would also have a positive impact on the availability of smaller-scale accelerators and derived radiation sources for applications in industry, medicine, and materials science. These include, for example, medical imaging and radiotherapy, radionuclide generation and non-destructive material inspection.

A promising path towards more compact accelerators lies in the use of plasma as accelerating medium. Since plasma is ionized gas that is already broken down into free electrons and ions, it can sustain much larger electric fields than the metallic structures in conventional accelerators. Four decades ago, Tajima and Dawson [1] proposed to accelerate electrons in the wake of an intense laser pulse that propagates through an underdense plasma. Shortly thereafter, Chen et al. [2] extended the idea by proposing the use of a charged particle beam as the driver. The radiation pressure of the laser pulse (or the space-charge force of the particle beam) pushes aside the

lighter plasma electrons and induces a charge separation on the sub-mm scale. As the displaced electrons are pulled back by the heavier and thus immobile ions, a traveling plasma density wave forms that supports large longitudinal and transverse electric fields. Similar to a surfer riding the wake of a motorboat, a carefully injected bunch of electrons (or positrons) can get transported and accelerated by this plasma wakefield. Today, plasma accelerators are routinely operated at accelerating gradients of ~ 100 GV/m [3] which is a thousand times larger than what is possible with conventional technology.

The experimental realization of laser-driven plasma acceleration [4] became possible only through the invention of chirped-pulse amplification [5] and its breakthrough finally came with the availability of high-power laser systems that could deliver Joule-level energies in less than a hundred femtoseconds of pulse duration. In 2004, three groups reported [6–8] on the generation of electron beams that for the first time featured a narrow spectrum at ~ 100 MeV energy and with ~ 100 pC of charge. From then on, the field has made rapid progress. Laser-plasma accelerators have produced beams of few-femtosecond duration with kiloampere peak currents [9, 10] and mm mrad transverse emittances [11, 12]. A steady increase in beam energy into the GeV range has been achieved [13–15], which culminated in the recent production of 8 GeV electrons in a 20-cm-long plasma guiding channel [16].

In virtually all laser-plasma accelerators, the injected and accelerated electrons stem from the plasma background. While early experiments relied on a rather uncontrollable self-trapping mechanism [17], more advanced injection techniques [18–21] led to a reduction in energy spread, down to the 1%-level [22–24], while improving the tunability and reliability of the bunch generation process.

In many measures, the electron beams from plasma accelerators are now on par with conventional machines. However, several challenges remain. Among the most pressing are further reducing the energy spread, improving the reproducibility and stability of the generated beams, and operating at an increased repetition rate and overall efficiency. Future applications, such as a plasma-based free-electron laser [25, 26] or linear collider [27], demand per-mille energy spreads and operation at kHz repetition rates with a few percent total efficiency. This requires not only a next generation of high average and peak power laser systems with a high wall-plug efficiency, but also a refinement of the injection methods and more control over the acceleration process.

Of particular importance is control over an effect known as beam loading [28–31], as it allows to flatten the otherwise steep accelerating fields while maximizing the transfer of energy from the plasma wave to the electron beam. Past experiments characterized effects from beam loading [32] and could already demonstrate its use to accelerate a large amount of charge while reducing the energy spread [33, 34].

For less demanding applications, a poorer beam quality and efficiency may suffice, but a highly stable and reliable operation would nonetheless have to be ensured [35]. While

past experimental results were often based on only a few tens of selected shots, there is now a transition towards more reproducible measurements with high statistics. In a recent experiment related to the present work, continuous and stable electron beam generation over more than 24 hours at 1 Hz repetition rate could be demonstrated [36]. Despite this progress, a few-percent stability in charge and central energy is still the norm. Further improvements on this front depend on gaining a better understanding of the influence of shot-to-shot drive laser fluctuations. These insights can then be used for targeted laser development and as an input to active stabilization systems when progressing to higher repetition rates.

The first part of this thesis contributes to this experimental progress towards stable high-quality electron beams. A novel continuous-flow plasma source for the reliable generation of high quality electron beams via localized ionization injection has been developed. The setup offers control over the injected bunch current profile and tunability of the acceleration process, which enables operation at optimal beam loading conditions. The resulting flattening of the average accelerating fields has been demonstrated experimentally and led to the production of beams with one percent level energy spreads [37]. The measurements were performed with high statistics, which allowed building a machine learning based surrogate model of the experiment that quantifies the influence of drive laser fluctuations.

The experimental progress described above has been fueled by an ever-increasing theoretical understanding of laser-plasma acceleration. Since the underlying physics often goes beyond what can be described by analytical models, numerical methods take a central role in this development. The most versatile and widespread approach in this regard is the electromagnetic particle-in-cell technique [38–40]. In this method, Maxwell’s equations are solved on a discrete grid and coupled to the motion of individual particles in continuous phase space. Particle-based simulations were already employed in the pioneering work on plasma acceleration by Dawson and colleagues [1, 2] and, to this day, remain the tool of choice for modeling the complex relativistic and kinetic plasma phenomena at stake.

A major drawback, however, is the high computational cost associated with particle-in-cell modeling. This circumstance restricted simulations in early work to one or two spatial dimensions. Eventually, due to the exponential growth of computing power, the increased availability of supercomputers, and the development of parallel numerical algorithms, fully three-dimensional simulations became feasible. Yet even today, the immense cost of such a simulation can prevent parameter scans and quantitative studies of experiments or concepts. Moreover, despite their superior parallel scalability, the finite difference based field solvers established in this process are prone to numerical errors [41–44].

Several recent developments opened up new avenues to increase the accuracy and drastically reduce the computational cost of plasma acceleration simulations. A method for the efficient parallelization of spectral field solvers enables more accurate results, while retaining a good scalability [45, 46]. The use of a cylindrical geometry with azimuthal mode decomposition allows capturing the important three-dimensional physics at the expense of only a few two-dimensional simulations [47, 48]. The Lorentz-boosted frame technique promises orders of magnitude speedup by modeling the laser-plasma interaction in a relativistic frame of reference [49].

The vision to combine all of these advancements in a single algorithm led to the development of the open source particle-in-cell code FBPIC [43, 50] which the present thesis has significantly contributed to. In particular, contributions were made to various improvements of the original algorithm, the code implementation, and the parallelization on modern compute hardware. Another central result of the second part of this thesis is the co-invention of a novel particle-in-cell algorithm [51–53] that intrinsically eliminates the numerical Cherenkov instability when simulating relativistically streaming plasmas. The occurrence of this instability is a long-standing problem in simulations of relativistic beams and plasmas [54, 55] and the lack of satisfying solutions for its mitigation prevented a widespread adoption of the Lorentz-boosted frame technique. It is demonstrated that the combination of all of these developments can reduce the cost of a typical plasma acceleration simulation by several orders of magnitude when compared to a traditional three-dimensional simulation in the laboratory frame, and without sacrificing accuracy.

Overview of Chapters

Physics of laser-plasma acceleration – Chapter 1 gives a theoretical introduction to the physics of laser-plasma acceleration. The plasma wave excitation by an intense laser pulse, the injection of electrons and the acceleration process are discussed.

LUX laser-plasma accelerator – Chapter 2 provides an overview of the LUX accelerator at which the experiments have been performed. The ANGUS laser system, the target area, the electron beam line and relevant diagnostics are described. Comprehensive measurements of the laser pulse properties are also presented.

Development of a localized ionization injection plasma source – Chapter 3 discusses the development of a novel plasma source for the stable generation of high-quality electron bunches using localized ionization injection. The gas density profile is derived from multi-species computational fluid dynamics simulations and the localization of the dopant gas is verified with measurements.

Demonstration of optimal beam loading – Chapter 4 reports on the experimental demonstration of optimal beam loading in a laser-plasma accelerator, resulting in the production of 1%-level energy spread electron beams. A detailed analysis of the measured beam loading effects and generated beam quality is presented. The influence of drive laser fluctuations is investigated using machine learning. The results are explained with particle-in-cell simulations.

The electromagnetic particle-in-cell method – Chapter 5 introduces the electromagnetic particle-in-cell method for relativistic plasma physics simulations. An overview of the particle-in-cell cycle is given, followed by a detailed discussion of finite-difference and pseudospectral methods for solving Maxwell's equations. Numerical Cherenkov radiation and instability are also explained.

FBPIC - A spectral, quasi-3D particle-in-cell code – Chapter 6 presents the spectral, quasi-3D particle-in-cell code FBPIC. An up-to-date description of the original algorithm is given, followed by a description of the implementation on modern hardware and parallelization by domain decomposition. Scaling tests are also presented.

Elimination of the numerical Cherenkov instability – Chapter 7 reports on a novel particle-in-cell algorithm, the Galilean scheme, that intrinsically eliminates the numerical Cherenkov instability (NCI) in simulations of relativistically drifting plasmas. The Galilean-PSATD solver and its extension to arbitrary order are discussed in detail. The elimination of the NCI is explained by a theoretical stability analysis.

Application to laser-plasma acceleration – Chapter 8 presents the application of the developed PIC methods to simulations of laser-plasma acceleration. The Galilean-PSATD solver is shown to enable stable Lorentz-boosted frame simulations, and FBPIC is compared to a conventional 3D Cartesian algorithm in terms of cost and accuracy.

Publications

The main results of this thesis have been peer reviewed and published in the following first-authored journal articles.

1. **M. Kirchen**, S. Jalas, P. Messner, P. Winkler, T. Eichner, L. Hübner, T. Hülsenbusch, L. Jeppe, T. Parikh, M. Schnepf, and A. R. Maier, *Optimal beam loading in a laser-plasma accelerator*, Phys. Rev. Lett. 126, 174801 (2021) [37]
2. **M. Kirchen**, R. Lehe, S. Jalas, O. Shapoval, J.-L. Vay, and A. R. Maier, *Scalable spectral solver in Galilean coordinates for eliminating the numerical Cherenkov instability in particle-in-cell simulations of streaming plasmas*, Phys. Rev. E 102, 013202 (2020) [53]
3. **M. Kirchen**, R. Lehe, B. B. Godfrey, I. Dornmair, S. Jalas, K. Peters, J.-L. Vay, and A. R. Maier, *Stable discrete representation of relativistically drifting plasmas*, Phys. Plasmas 23, 100704 (2016) [51]

The content of these articles, which were published under the Creative Commons Attribution 4.0 International license, has been created by the present author and is reproduced in this thesis, in an expanded form. Contributions of co-authors have been highlighted at the beginning of the respective chapters.

The following co-authored publications are also related to this thesis but report on work primarily done by others.

4. S. Jalas, **M. Kirchen**, P. Messner, P. Winkler, L. Hübner, J. Dirkwinkel, M. Schnepf, R. Lehe, and A. R. Maier, *Bayesian optimization of a laser-plasma accelerator*, Phys. Rev. Lett. 126, 104801 (2021) [56]
5. A. R. Maier, N. M. Delbos, T. Eichner, L. Hübner, S. Jalas, L. Jeppe, S. W. Jolly, **M. Kirchen**, V. Leroux, P. Messner, M. Schnepf, M. Trunk, P. A. Walker, C. Werle, and P. Winkler, *Decoding Sources of Energy Variability in a Laser-Plasma Accelerator*, Phys. Rev. X 10, 031039 (2020) [36]
6. S. Jalas, I. Dornmair, R. Lehe, H. Vincenti, J.-L. Vay, **M. Kirchen**, and A. R. Maier, *Accurate modeling of plasma acceleration with arbitrary order pseudo-spectral particle-in-cell methods*, Phys. Plasmas 24, 033115 (2017) [44]

7. R. Lehe, **M. Kirchen**, B. B. Godfrey, A. R. Maier, and J.-L. Vay, *Elimination of Numerical Cherenkov Instability in flowing-plasma Particle-In-Cell simulations by using Galilean coordinates*, Phys. Rev. E 94, 053305 (2016) [52]
8. R. Lehe, **M. Kirchen**, I. A. Andriyash, B. B. Godfrey, and J.-L. Vay, *A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm*, Comput. Phys. Commun. 203, 66 (2016) [43]

Contributions have also been made to the following publications that are not directly related to the content of this thesis.

9. V. Leroux, S. Jolly, M. Schnepf, T. Eichner, S. Jalas, **M. Kirchen**, P. Messner, C. Werle, P. Winkler, and A. Maier, *Wavefront degradation of a 200 TW laser from heat-induced deformation of in-vacuum compressor gratings*, Opt. Express 26, 13061-13071 (2018) [57]
10. N. Delbos, C. Werle, I. Dornmair, T. Eichner, L. Hübner, S. Jalas, S. W. Jolly, **M. Kirchen**, V. Leroux, P. Messner, M. Schnepf, M. Trunk, P. A. Walker, P. Winkler, A. R. Maier, *LUX – A Laser-Plasma Driven Undulator Beamline*, Nucl. Instr. Meth. Phys. Res. A 909, 318 (2018) [58]
11. R. Brinkmann, N. Delbos, I. Dornmair, R. Assmann, C. Behrens, K. Floettmann, J. Grebenyuk, M. Gross, S. Jalas, **M. Kirchen**, T. Mehrling, A. Martinez de la Ossa, J. Osterhoff, B. Schmidt, V. Wacker, and A. R. Maier, *Chirp mitigation of plasma-accelerated beams using a modulated plasma density*, Phys. Rev. Lett. 118, 214801 (2017) [59]

1 Physics of laser-plasma acceleration

This chapter gives a theoretical introduction to the physics of laser-plasma acceleration. In Sec. 1.1, we motivate the use of plasma as an accelerating medium and describe the excitation of a plasma wave by the ponderomotive force of an intense, ultrashort laser pulse. In Sec. 1.2, we discuss the injection and acceleration of an electron bunch in the plasma wave, with a particular emphasis on ionization injection and beam loading. The content of this introduction is largely based on [60] and [4].

1.1 Laser-driven plasma waves

1.1.1 Acceleration in plasma

Plasma refers to ionized matter in a wide range of density and temperature regimes. It is commonly defined as “a *quasineutral* gas of charged and neutral particles which exhibits *collective behavior*.”[61]

Quasineutrality refers to a plasma’s charge-neutral appearance from afar, even though the constituent particles carry charge on the microscopic level. Collective behavior then arises from the long-range Coulomb interaction involving many of these particles. It is the resulting electromagnetic plasma dynamics and large fields associated with it that our interest lies in and that we seek to exploit for particle acceleration.

In the following, we consider a gas of fully ionized atoms and free electrons and summarize the criteria that characterize it as an ideal collisionless plasma. To simplify matters, the heavier ions are assumed to be stationary and form a homogeneous positively charged background n_0 . This is justified by the fact that the phenomena regarded here occur on much shorter time scales than ion motion.

The occurrence of quasineutrality can be quantified by means of the *Debye shielding length* [61, 62]

$$\lambda_D = \sqrt{\frac{\epsilon_0 k_B T_e}{n_e e^2}} \propto \sqrt{\frac{T_e}{n_e}}, \quad (1.1)$$

with ϵ_0 the vacuum permittivity, k_B the Boltzmann constant, T_e the electron plasma temperature, n_e the electron plasma density and e the elementary charge. It is the characteristic distance over which a local charge perturbation is screened by the

surrounding plasma. More precisely, λ_D equals the radius r at which the unshielded Coulomb potential of a point-like charge q , $\Phi_0 \sim q/r$, has been damped by a factor $1/e$, $\Phi \sim \Phi_0 \exp(-r/\lambda_D)$. To an outside observer, a plasma appears charge neutral on length scales $L \gg \lambda_D$.

The collective electrostatic interaction between particles dominates over individual effects if many particles reside within a sphere of radius λ_D . Hence, the definition of an ideal plasma entails that the *Debye number* [61, 62]

$$N_D = \frac{4}{3}\pi\lambda_D^3 n_e \propto n_e^{-\frac{1}{2}}, \quad (1.2)$$

should be much larger than unity, $N_D \gg 1$. The response time over which the plasma recovers quasineutrality after an external perturbation can be defined as [62]

$$t_D \equiv \frac{\lambda_D}{v_{te}} = \sqrt{\frac{\epsilon_0 k_B T_e}{e^2 n_e}} \cdot \sqrt{\frac{m_e}{k_B T_e}} \quad (1.3)$$

with m_e the electron mass and $v_{te} \equiv \sqrt{k_B T_e / m_e}$ the plasma thermal velocity. For collective processes to dominate, t_D should be shorter than the average time between individual Coulomb collisions, which is also satisfied by the constraint $N_D \gg 1$.

From the above considerations follows that when displacing a sheath of electrons with respect to the static ion background, the restoring Coulomb force invokes a collective oscillation around the equilibrium position at the characteristic *electron plasma frequency* [61, 62]

$$\omega_p = t_D^{-1} = \sqrt{\frac{n_e e^2}{\epsilon_0 m_e}}. \quad (1.4)$$

The resonant excitation of such plasma oscillations gives rise to large-amplitude electron density waves [63] which can sustain electric fields on the order of the *cold non-relativistic wave breaking limit* [64]

$$E_0 = \frac{m_e c \omega_p}{e} \propto \sqrt{n_e}. \quad (1.5)$$

For a practical example, consider field-ionized hydrogen at 20 mbar pressure, which yields a plasma density of $n_e = 10^{18} \text{ cm}^{-3}$ at temperatures of $\sim 10 \text{ eV}$ [4]. In this case, $\lambda_D \simeq 24 \text{ nm}$, $N_D \simeq 54$, and the plasma electrons respond collectively to perturbations within a characteristic time $\omega_p^{-1} \simeq 18 \text{ fs}$. This allows for the excitation of plasma waves supporting electric fields on the order of $E_0 \simeq 100 \text{ GV/m}$, which exceeds the accelerating gradients in conventional accelerators by three orders of magnitude.

Plasma-based accelerators [64, 1, 2, 65, 4] make use of these extremely large fields to drastically reduce the distance required to accelerate charged particles to ultrarelativistic

energies. In the most common schemes [1, 2], an intense laser- or particle beam driver resonantly ($L_{\text{drv}} \sim \lambda_p/2$) excites a trailing electron density wave that carries large electric fields, the *plasma wakefield*, in which a short ($L_{\text{bunch}} < \lambda_p$), co-propagating electron bunch can get accelerated over an extended distance $L_{\text{acc}} \gg \lambda_p$, with $\lambda_p = 2\pi c/\omega_p = 2\pi/k_p$ the characteristic wavelength of the plasma modulation. For the example given above, $\lambda_p \simeq 33 \mu\text{m}$, which allows for the acceleration of electron bunches with durations $< 100 \text{ fs}$ to GeV energy levels over only centimeter distances.

The following sections summarize the physics of a laser-plasma accelerator driven by a single, short laser pulse, which is commonly referred to as *laser-wakefield accelerator* (LWFA) [4].

1.1.2 Ponderomotive force

In order to arrive at an understanding for the mechanism of plasma wave excitation by a laser pulse, we first consider the relativistic interaction of an electron with a linearly polarized plane-wave light pulse for which we impose a slowly varying but finite temporal envelope.

The motion of a charged particle in an electric and magnetic field is determined by the Lorentz force $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$. It is convenient to express those fields in terms of a scalar Φ and vector \mathbf{A} potential [66]:

$$\mathbf{E} = -\nabla\Phi - \frac{\partial\mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}. \quad (1.6)$$

The Lorentz force equation for an electron with charge e and mass m_e then reads

$$\frac{d\mathbf{p}}{dt} = -e \left[-\nabla\Phi - \frac{\partial\mathbf{A}}{\partial t} + \mathbf{v} \times (\nabla \times \mathbf{A}) \right], \quad (1.7)$$

where t is the time, $\mathbf{p} = \gamma m \mathbf{v}$ is the momentum with \mathbf{v} the velocity and $\gamma = 1/\sqrt{1 - v^2/c^2}$ the relativistic factor of the particle, where c is the vacuum speed of light. The transversely polarized light pulse can be defined as $\mathbf{A} = \hat{A} \cos(kz - \omega t) \mathbf{e}_\perp$ with temporal envelope $\hat{A}(z, t)$.

Transverse motion

As the wave passes, the electron starts to oscillate in the electric field and performs a *quiver motion* in the transverse direction. For sub-relativistic velocities $v \ll c$,

the motion is dominated by the electric component of the Lorentz force (qE_{\perp}). By integrating Eq. (1.7) in time, we obtain the *transverse momentum equation* [60]

$$\gamma\beta_{\perp} = a, \quad (1.8)$$

where we introduced the *normalized vector potential* $a = eA_{\perp}(m_e c)^{-1}$ and $\beta_{\perp} = v_{\perp}/c$. We find that the electron is accelerated to relativistic velocities as the right hand side of Eq. (1.8) approaches unity, $\gamma\beta_{\perp} \sim 1$.

In the context of laser-plasma acceleration, the onset of relativistic effects and different regimes of wave excitation are therefore usually classified in terms of the peak amplitude a_0 of the normalized vector potential. It is related to the intensity $I = (\epsilon_0 c/2)E^2$ via

$$I_0 = \frac{a_0^2}{\lambda^2} \frac{\pi m_e c^3}{2r_e}, \quad (1.9)$$

where r_e is the classical electron radius. The present generation of laser-plasma accelerators is driven by titanium-sapphire laser systems that produce pulses with a central wavelength $\lambda \simeq 800$ nm, for which relativistic effects become important at intensities of $\sim 2 \times 10^{18}$ W cm⁻².

Longitudinal motion

Consequently, when the incident pulse reaches relativistic intensities ($a \rightarrow 1$), the influence of the magnetic component of the Lorentz force ($\mathbf{v} \times \mathbf{B}$) can no longer be neglected. Similar to before, we can derive the *longitudinal momentum equation* [60]

$$\frac{d(\gamma\beta_z)}{dt} = c \left(\frac{\partial\phi}{\partial z} - \frac{1}{2\gamma} \frac{\partial a^2}{\partial z} \right) \quad (1.10)$$

where we introduced the *normalized scalar potential* $\phi = e\Phi(m_e c^2)^{-1}$ and substituted Eq. (1.8) for the transverse velocity component. In vacuum ($\phi = 0$), the axial motion of the electron is described by the second term on the right, which corresponds to a nonlinear force exerted by the magnetic field that scales $\propto a^2$.

In addition to the transverse quiver motion, the electron performs longitudinal oscillations while gaining an average drift velocity in the propagation direction of the pulse. However, as the intensity decreases along the trailing edge of the pulse, the acceleration reverses, and albeit being displaced longitudinally, the electron is at rest again at the end of the interaction [60].

For radially non-uniform pulses, similar drift effects occur in the transverse direction, which can be discussed qualitatively: Consider the quiver motion in the electric field of a focused laser pulse. As the oscillating electron moves into regions of lower intensity,

it experiences a reduced restoring force and effectively drifts away from the region of highest intensity in the center [60].

A general expression that describes the acceleration in an inhomogeneous laser field is given by the *ponderomotive force* [67], which, in the limit $a_0 \ll 1$, reads [4]

$$\mathbf{F}_p = -m_e c^2 \frac{\nabla \mathbf{a}^2}{2\gamma}, \quad (1.11)$$

and can be understood as a radiation pressure that pushes electrons towards regions of lower intensity. In principle, the ponderomotive force allows for direct electron acceleration in vacuum. However, the $1/\gamma$ scaling of Eq. (1.11) fundamentally limits the energy gain of ponderomotive acceleration schemes. Among other reasons, this motivates the use of a plasma in which the ponderomotive force transforms the transverse laser field into the axial electric field of a plasma wave, which can then be used to efficiently accelerate a particle beam [4].

1.1.3 Laser pulse propagation in plasma

Before progressing to a theoretical description of the wakefield excitation mechanism, we briefly discuss the propagation of a laser pulse in plasma. Once again, we can express the fields in terms of the potentials and choose the Coulomb gauge ($\nabla \cdot \mathbf{A} = 0$) to write Maxwell's equations as [66]

$$\nabla^2 \Phi = -\frac{\rho}{\epsilon_0}, \quad (1.12)$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} + \nabla \frac{1}{c^2} \frac{\partial \Phi}{\partial t}, \quad (1.13)$$

with ϵ_0 and μ_0 the vacuum permittivity and permeability, and c the vacuum speed of light, respectively. In the absence of a charge ρ and current density \mathbf{J} , solutions of Eqs. (1.12) and (1.13) are electromagnetic modes of the form $e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}$ that fulfill the vacuum dispersion relation [66]

$$\omega^2 = c^2 k^2, \quad (1.14)$$

with ω the angular frequency and $k = |\mathbf{k}| = 2\pi/\lambda$ the wavenumber.

In the presence of plasma—represented here as a cold ($k_B T_e \simeq 0$) electron fluid—the fields are coupled to the fluid density and velocity via $\rho = e(n_0 - n_e)$ and $\mathbf{J} = -en_e \mathbf{v}$. Consequently, for a plane-wave light pulse that propagates through plasma, the *electromagnetic wave equation* [Eq. (1.13)] becomes [60, 4]

$$c^2 \nabla^2 a - \frac{\partial^2 a}{\partial t^2} = \omega_p^2 \frac{na}{\gamma}, \quad (1.15)$$

with $n = n_e/n_0$ and where we used Eq. (1.8) express the transverse fluid velocity v_\perp in terms of the normalized vector potential a . Further note that the term $\nabla\partial\phi/\partial t$ can be neglected [4].

For non-relativistic intensities and in the limit of small density perturbations, we can set $\gamma = n \simeq 1$ and immediately see that the vacuum dispersion relation [Eq. (1.14)] expands to [61, 62]

$$\omega^2 = \omega_p^2 + c^2 k^2. \quad (1.16)$$

Correspondingly, the phase (v_ϕ) and group (v_g) velocity of a wave packet that travels through plasma are given by [66]

$$v_\phi = \frac{\omega}{k} = \frac{c}{\eta}, \quad (1.17)$$

$$v_g = \frac{d\omega}{dk} = c \cdot \eta, \quad (1.18)$$

where $\eta = (1 - \omega_p^2/\omega^2)^{1/2}$ is the plasma refractive index as evident from Eq. (1.16).

From these expressions, we infer that a high-frequency light pulse with $\omega > \omega_p$ propagates through a plasma with a superluminal phase ($v_\phi > c$) and subluminal group ($v_g < c$) velocity. In this situation, the plasma acts like a nonlinear refractive medium and is called *underdense*. In contrast, an electromagnetic oscillation that is slower than the characteristic response time t_D is shielded out by the plasma. Hence, if $\omega < \omega_p$, the refractive index becomes imaginary and the incident wave is reflected or absorbed. In this case, the plasma behaves similar to a mirror and is called *overdense*.

The transition from under- to overdense behaviour ($\omega = \omega_p$) can be defined in terms of a critical density [61, 62]

$$n_c = \frac{\omega^2 \epsilon_0 m_e}{e^2}. \quad (1.19)$$

Revisiting the previous example with $\lambda = 800 \text{ nm}$ yields a critical density $n_c \simeq 1.74 \times 10^{21} \text{ cm}^{-3}$, which is several orders of magnitude larger than the $10^{17-19} \text{ cm}^{-3}$ plasma densities laser-wakefield accelerators are typically operated at. Accordingly, this work is concerned with the physics of underdense plasmas only.

Nonlinear effects

The plasma refractive index from above, $\eta = (1 - \omega_p^2/\omega^2)^{1/2}$, applies for non-relativistic intensities ($\gamma = 1$) and negligible perturbations to the plasma density ($n \simeq 1$). In a more general form, the refractive index can be written as [68]

$$\eta \simeq 1 - \frac{\omega_p^2}{2\omega^2} \frac{n}{\gamma} \quad (1.20)$$

assuming $\omega_p^2/\omega^2 \ll 1$. Here, the influence of density variations, $n = n_e/n_0$, resulting either from the laser-plasma interaction itself or from external measures, is explicitly included. In addition, the factor γ^{-1} accounts for an effective decrease in plasma frequency as the fluid electrons gain relativistic mass.

In the presence of an inhomogeneous refractive index profile $\eta(r, z)$, a laser pulse is modulated transversely and longitudinally during its propagation through the plasma. Of particular interest are radial profiles that satisfy $\partial\eta(r)/\partial r < 0$, since the resulting focusing effect can increase the intensity or optically guide a laser pulse.

For example, self-focusing of a laser pulse occurs as a result of the relativistic quiver motion of the plasma fluid electrons [69, 68]. Consider a laser pulse with a Gaussian radial envelope, $\hat{a}(r) = a_0 \exp(-r^2/w_0^2)$, with spot size w_0 . During the interaction with the laser, the relativistic factor of the oscillating electrons is, to first order, determined by the vector potential, $\gamma(r) \simeq (1 + \hat{a}(r)^2)^{1/2}$. The refractive index thus follows the intensity profile and focuses the laser.

In the limit $a_0 \ll 1$, one can define a critical power at which this *relativistic self-focusing* effect compensates the natural diffraction of a long ($L \gg \lambda_p$) pulse [69, 68]:

$$P_c = \frac{2m_e c^3}{r_e} \left(\frac{\omega}{\omega_p} \right)^2. \quad (1.21)$$

Reconsidering our previous example with $\lambda = 800 \text{ nm}$ and $n_e = 10^{18} \text{ cm}^{-3}$, we find that optical guiding due to relativistic self-focusing becomes possible at laser powers $P = (\pi/2)Iw_0^2 \simeq 30 \text{ TW}$.

Further modifications of the refractive index profile emerge from the plasma wave excitation process that is discussed in detail in the next section. The longitudinal ponderomotive force of a short ($L_{\text{drv}} \sim \lambda_p$) laser pulse induces a plasma density response that counteracts relativistic self-focusing at the head of the pulse, but may enhance focusing of its tail [70, 68]. Moreover, the transverse ponderomotive force of a narrow ($k_p w_0 \sim 1$) laser pushes plasma electrons outwards, which creates a density depleted region on the axis that enhances guiding of the pulse [71, 72, 68].

Irrespective of the nonlinear interaction with the plasma electrons, optical guiding of a laser pulse is also possible in preformed plasma channels. A parabolic density variation of the form $\delta n_e(r) = r^2/(\pi r_e w_0^4)$ exactly compensates the diffraction of a Gaussian laser in the limit $P \ll P_c$, $a_0 \ll 1$ [73, 72, 68].

In the longitudinal direction, an intense, short ($L_{\text{drv}} \sim \lambda_p$) pulse is also affected by the nonlinear effects discussed so far and undergoes *self-phase modulation* [74–76]. The plasma response can lead to self-steepening and an initial self-compression of the intensity profile. At the same time, overall frequency redshifting occurs, in line with the fact that energy is deposited into the plasma.

1.1.4 Plasma wave excitation

When an intense laser pulse propagates through an underdense plasma, its ponderomotive force pushes the lighter electrons forward and aside, while the heavier ions remain immobile due their large inertia. In consequence, electrons accumulate at the leading edge of the pulse and an electron-depleted region forms behind the pulse.

The resulting charge separation induces an electrostatic field that acts as a restoring force and eventually pulls back the displaced electrons. At the same time, the electrons experience a backward directed ponderomotive force along the trailing edge of the pulse. The excess momentum causes them to overshoot their initial position and then oscillate around it at the plasma frequency ω_p . From the superposition of the individual oscillations, an electron density wave forms that travels at a phase velocity v_p equal to the group velocity $v_g \simeq c$ of the driving laser pulse and thus has a wavelength of $\simeq 2\pi c/\omega_p \equiv \lambda_p$.

Resonant excitation of the plasma wave requires the two ponderomotive kicks to add in phase, $\tau_{\text{drv}} \simeq \omega_p^{-1}$, where τ_{drv} corresponds to the rms intensity duration for a Gaussian laser pulse. For a plasma density $n_e = 10^{18} \text{ cm}^{-3}$, as in our previous example, the resonance condition is fulfilled for a pulse duration of ~ 40 fs FWHM.

Quasistatic 1D cold fluid theory

A solution for the resulting electric fields behind the driver, the so-called *plasma wakefield*, can be obtained by solving a set of 1D cold fluid equations [70, 77, 78]. Their derivation is summarized in the following. Details can be found in [60, 4].

The derivation starts with the previous expressions for the transverse and longitudinal electron motion in a plane-wave light pulse [Eqs. (1.8) and (1.10)], and the electromagnetic wave equation [Eq. (1.15)]. Together with the *continuity equation* for the density n_e of a cold ($k_B T_e \simeq 0$) electron fluid that is modulated in the longitudinal direction [60, 4]

$$\frac{\partial n_e}{\partial t} + c \frac{\partial}{\partial z} (n_e \beta_z) = 0, \quad (1.22)$$

and *Poisson's equation* for the electrostatic potential [60, 4]

$$\frac{\partial^2 \phi}{\partial z^2} = k_p^2 (n - 1), \quad (1.23)$$

they form a closed set of equations describing the excitation of nonlinear plasma waves by a driver pulse with arbitrary amplitude a and group velocity $v_g = v_p$. Note that

$n = n_e/n_0$ with n_0 the density of the static ion background. It is also convenient to split the relativistic factor into a transverse and longitudinal component:

$$\gamma = \gamma_\perp \gamma_z = (1 + a^2)^{\frac{1}{2}} (1 - \beta_z^2)^{-\frac{1}{2}}. \quad (1.24)$$

The equations are then simplified by (1) performing a coordinate transformation ($\zeta = z - v_p t, \tau = t$), such that the system is represented in a frame that is comoving with the driver (at the wave phase velocity v_p), and by (2) applying the *quasi-static approximation* (QSA) [79, 70], which assumes that the driver envelope does not change significantly over the characteristic plasma response time.

For one thing, it follows that the evolution of the driver that is described by the electromagnetic wave equation [Eq. (1.15)] can be treated separately and is therefore disregarded in the following. But more importantly, the QSA allows us to neglect the slowly varying time derivatives in the fluid equations that describe the plasma response to the laser pulse excitation.

Consequently, the partial derivatives in the comoving coordinates become [60, 4]

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial \zeta}, \quad \frac{\partial}{\partial t} = \frac{\partial}{\partial \tau} - v_p \frac{\partial}{\partial \zeta} \simeq -v_p \frac{\partial}{\partial \zeta}, \quad (1.25)$$

so that we only have to integrate over $\partial/(\partial\zeta)$ in the longitudinal momentum equation [Eq. (1.10)] and the continuity equation [Eq. (1.22)]. This then yields the conservation equations [60, 4]

$$n = \frac{\beta_p}{\beta_p - \beta_z}, \quad (1.26)$$

$$\phi = \gamma(1 - \beta_p \beta_z) - 1. \quad (1.27)$$

Finally, the above equations can be combined with Poisson's equation [Eq. (1.23)] to arrive at expressions for the normalized density n , scalar potential ϕ and longitudinal electric field $\varepsilon = E_z/E_0$ that only depend on the laser amplitude a [60, 4]:

$$n = \gamma_p^2 \beta_p \left(\frac{1}{\psi} - \beta_p \right), \quad \frac{\partial^2 \phi}{\partial \zeta^2} = k_p^2 \gamma_p^2 \left(\frac{\beta_p}{\psi} - 1 \right), \quad \varepsilon = -k_p^{-1} \frac{\partial \phi}{\partial \zeta}, \quad (1.28)$$

$$\text{with } \psi = \left(1 - \frac{1 + a^2}{\gamma_p^2 (1 + \phi)^2} \right)^{\frac{1}{2}}. \quad (1.29)$$

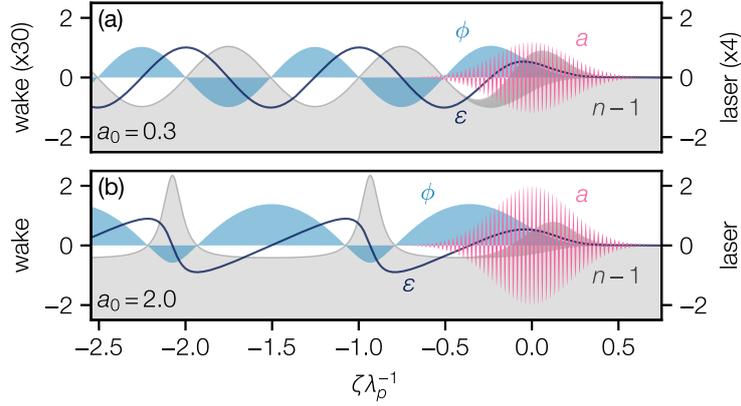


Figure 1.1 – 1D plasma wave driven by a Gaussian laser with peak amplitude (a) $a_0 = 0.3$ and (b) $a_0 = 2.0$. The pulse propagates to the right, is centered at $\zeta = 0$ and has a length $c\tau_0 = 2k_p^{-1}$. Shown are the normalized laser vector potential a , plasma wave scalar potential ϕ , axial electric field $\varepsilon = E_z/E_0$ and electron density modulation $n - 1$. Note that for the case $a_0 = 0.3$, the laser and wakefield quantities have been scaled by a factor 4 and 30, respectively.

Analytical solutions to Eq. (1.28) exist for square-shaped laser pulses [79, 70]. In this case, the excitation is optimal for a driver length $L_{\text{drv}} = \lambda_{np}/2$, with λ_{np} the nonlinear plasma wavelength as defined below. The maximum wakefield amplitude is [4]

$$\varepsilon_{\text{max}} = \frac{E_{\text{max}}}{E_0} = \frac{a_0^2}{2} \left(1 + \frac{a_0^2}{2} \right)^{-\frac{1}{2}}, \quad (1.30)$$

assuming linear polarization. Equation (1.30) reduces to $\varepsilon_{\text{max}} \simeq a_0^2/2$ in the limit $a_0 \ll 1$ and $\varepsilon_{\text{max}} \simeq a_0/\sqrt{2}$ in the limit $a_0 \gg 1$.

For arbitrary laser profiles, the second-order differential equation for the wakefield potential [Eq. (1.28)] has to be integrated numerically. Figure 1.1 shows the wakefield quantities assuming a Gaussian pulse [$a = a_0 \exp(-\zeta^2/(c\tau_0)^2) \cos(k\zeta)$] with a peak amplitude (a) $a_0 = 0.3$ and (b) $a_0 = 2$. Based on these 1D results and additional 3D simulations, the different regimes of laser-wakefield excitation are discussed in the following.

1.1.5 Regimes of laser-plasma acceleration

Depending on the normalized driver strength a_0 and spot size $k_p w_0$, different regimes of laser-plasma acceleration are distinguished, whose description requires theoretical models of varying complexity.

In the *linear regime* ($a_0 \ll 1$), analytical solutions exist for the linearized cold fluid equations that readily extend to 3D [80]. In the *nonlinear regime* ($a_0 \sim 1$), the 1D cold fluid equations from above are sufficient to describe the 3D wakefield structure in the broad driver limit ($k_p w_0 \gg 1$). The latter criteria entails that the excitation is dominated by the longitudinal component of the ponderomotive force. However, for a general description of nonlinear plasma waves excited by radially confined drivers ($k_p w_0 \sim 1$) one has to rely on numerical solutions of a 3D fluid model [81, 82]. In the highly nonlinear *blowout* or *bubble regime* ($a_0 \gtrsim k_p w_0$, where usually $a_0 \gg 1$), the fluid picture eventually breaks down and a 3D kinetic description based on particle simulations is required [83, 84, 17].

The following discussion of these different regimes is therefore supported by results from particle-in-cell simulations with FBPIC. Examples of a laser-driven plasma wave in each regime are collated in Fig. 1.2.

In the *linear regime* ($a_0 \ll 1$), the electron motion remains non-relativistic, which leads to the formation of a sinusoidal plasma wave [compare Fig. 1.1(a)]. Since the excitation occurs mainly in the longitudinal direction, the wakefield amplitude inherits the intensity profile of the driver in the radial direction. It can also be observed that the longitudinal and transverse electric fields are shifted in phase by $\pi/2$ with respect to each other [Fig. 1.2(d)]. An electron beam is accelerated if placed in the second half of a plasma wave bucket. Given the requirement that the beam simultaneously experiences focusing forces, this region reduces to $\lambda_p/4$.

When transitioning to the *nonlinear regime* ($a_0 \sim 1$), the electron motion becomes relativistic, which has a significant effect on the shape of the plasma wave. The electron density modulation becomes peaked and the accelerating field develops a sawtooth-like profile [compare Fig. 1.1(b)].

In addition, the plasma wave period increases. Physically, this is explained by the electrons gaining relativistic mass, which effectively decreases the plasma frequency and in turn elongates the wavelength of the laser-driven oscillation. An analytical expression for the nonlinear plasma wavelength is given by [4]

$$\lambda_{np} = \lambda_p \times \begin{cases} 1 + (3/16) \varepsilon_{\max}^2 & \text{for } \varepsilon_{\max} \ll 1 \\ (2/\pi)(\varepsilon_{\max} + \varepsilon_{\max}^{-1}) & \text{for } \varepsilon_{\max} \gg 1, \end{cases} \quad (1.31)$$

which assumes $\gamma_p \gg 1$. For a wakefield driven in resonance by a linearly polarized, square-shaped laser pulse [Eq. (1.30)], we find $\lambda_{np} \simeq \lambda_p(\sqrt{2}/\pi) a_0$ in the limit $a_0 \gg 1$.

An important radial feature of plasma waves in the nonlinear regime can be observed in Fig. 1.1(c). Since the drive laser has a Gaussian intensity profile in the radial direction, the wakefield excitation is strongest on the axis and becomes weaker towards the outer radii. As a result, the plasma wavelength decreases with distance from the axis and

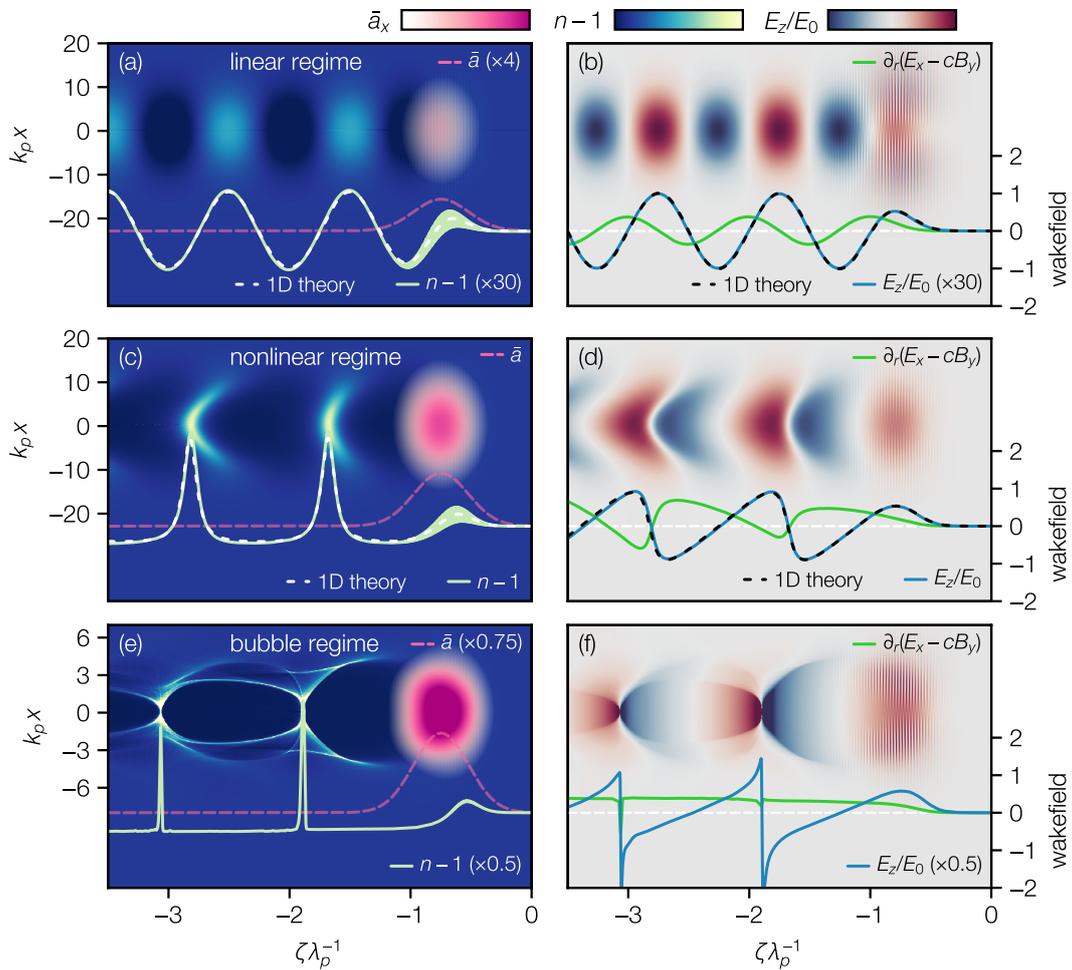


Figure 1.2 – Laser-driven plasma waves in different regimes. Results were obtained from quasi-3D particle-in-cell simulations with FBPIC and are compared to the 1D cold fluid theory. The plasma density is $n_e = 10^{18} \text{ cm}^{-3}$. The linearly polarized laser ($\lambda = 800 \text{ nm}$) propagates to the right and has a Gaussian profile [$\hat{a} = a_0 \exp(-\zeta^2/(c\tau_0)^2) \exp(-r^2/w_0^2)$] with a fixed pulse duration $c\tau_0 = 2k_p^{-1}$ and varying spot size w_0 and peak amplitude a_0 . (a), (b) Linear regime with $a_0 = 0.3$ and $k_p w_0 = 12$; (c), (d) Nonlinear regime with $a_0 = 2$ and $k_p w_0 = 12$; (e), (f) Blowout regime with $a_0 = 4$ and $k_p w_0 = 4$. Electron density modulation $n - 1$ (left column) and longitudinal electric field E_z/E_0 (right column). On-axis ($r = 0$) lineouts comparing simulation (solid) and 1D theory (dashed) results. Time-averaged laser vector potential \bar{a} . Focusing fields $\partial_r(E_x - cB_y)$ (arb. u.). Note that the physical quantities have been scaled by different factors for better visibility.

the wavefronts become curved. With regard to the acceleration and transport of an electron beam, one also finds that the region where the transverse fields are focusing is larger in the nonlinear regime.

If the driver amplitude is increased further ($a_0 \gg 1$) and provided that the spot size is comparable to the plasma wavelength ($k_p w_0 \sim 1$), the impact of the radial component of the ponderomotive force can cause all plasma electrons to be expelled from the axis. In consequence, the electrons accumulate in a thin sheath and flow around the cavitating region that forms behind the driver. A condition for the occurrence of such electron cavitation that also applies in the case $a_0 < 1$ is given by [4]

$$\frac{a_0^2}{\sqrt{1+a_0^2}} \geq \frac{k_p^2 w_0^2}{4}. \quad (1.32)$$

An example for a plasma wave driven in this so-called *blowout* or *bubble regime* [84, 17, 85] is shown in Fig. 1.2(e) and (f). Within the bubble, the longitudinal and transverse fields vary approximately linearly with the axial and radial coordinate. Assuming a completely blown-out, spherical ion cavity with radius $r_B \gg k_p^{-1}$ that is centered at $\zeta = r = 0$, the fields can be approximated as [85]

$$E_z \simeq \frac{k_p \zeta}{2} E_0, \quad (1.33)$$

$$E_r - cB_\theta \simeq \frac{k_p r}{2} E_0, \quad (1.34)$$

for $|\zeta|, r \leq r_B$.

Theoretical considerations [85–87] for the highly nonlinear regime ($a_0 \geq 4$) furthermore predict that an ideal spherical bubble of radius $r_B \simeq (2/k_p)\sqrt{a_0}$ is excited when the laser spot size satisfies $k_p w_0 \simeq 2\sqrt{a_0}$. According to Eq. (1.33), the maximum accelerating field in this case is $E_{\max}/E_0 \simeq \sqrt{a_0}$. However, these scalings are accurate in a narrow range of parameters only [88]. Moreover, the structure of a laser-driven bubble can be quite complex and is usually not adequately described by existing analytical theories [89]. Therefore, particle-in-cell simulations are indispensable for modeling laser-wakefield acceleration in this regime.

1.2 Electron acceleration

Having established a theoretical basis for describing the excitation of a laser-driven plasma wave, we proceed to discuss the injection and acceleration of electrons therein.

The motion of an electron in a 1D nonlinear plasma wakefield can be described by means of a Hamiltonian [90, 91, 4]

$$H(\zeta, \gamma\beta_z) = \sqrt{1 + \gamma^2\beta_z^2 + a^2(\zeta)} - \phi(\zeta) - \gamma\beta_z\beta_p, \quad (1.35)$$

where ϕ is the wakefield potential that is obtained by numerically solving the quasistatic plasma fluid equations [Eq. (1.28)] for a given laser envelope a . Bear in mind that the coordinate system is comoving with the driver pulse ($\zeta = z - v_p t$), assuming $v_p = v_g$.

Eq. (1.35) comprises the total kinetic and potential energy of an electron and is conserved along a given trajectory in phase space. Hence, for a constant value $H \equiv H_0$, one can derive an expression for the evolution of the electron's longitudinal momentum $\gamma\beta_z$ as a function of the wakefield position ζ [90, 4]:

$$\gamma\beta_z = \beta_p\gamma_p^2(H_0 + \phi) \pm \gamma_p\sqrt{\gamma_p^2(H_0 + \phi)^2 - \gamma_\perp^2}. \quad (1.36)$$

Examples for possible electron trajectories in phase space $(\zeta, \gamma\beta_z)$, each of which correspond to a different value H_0 , are shown in Fig. 1.3.

Depending on their initial position and momentum, electrons either move on open or closed orbits. Electrons that are initially at rest before the interaction with the laser ($\gamma\beta_z = 0$ at $\zeta = +\infty$) move on the trajectory that satisfies $H_0 = 1$. These electrons are untrapped and take part in the cold fluid oscillations that form the plasma wave. Electrons can get trapped in a plasma wave bucket provided they have a sufficiently high initial momentum or are born at a position where the wakefield potential is large. Once trapped, they move on a closed orbit in phase space and oscillate longitudinally as they periodically gain and lose energy. Electrons with a very high initial momentum also interact with the plasma wakefield but remain untrapped and eventually outrun the laser.

The boundary separating trapped from untrapped orbits in phase space corresponds to the trajectory of an electron having a momentum equal to the wake phase velocity ($\gamma\beta_z = \gamma_p\beta_p$) at the minimum of the potential ($\phi(\zeta_{\min}) \equiv \phi_{\min}$). The Hamiltonian of this *separatrix* is given by [90, 4]

$$H_s \equiv H(\zeta_{\min}, \gamma_p\beta_p) = \gamma_\perp/\gamma_p - \phi_{\min}. \quad (1.37)$$

1.2.1 Injection

Acceleration of an electron bunch in the plasma wave requires injecting electrons into trapped phase space orbits that satisfy $H_t < H_s$, while ensuring that the initial position and momentum of the bunch are such that it is focused by the transverse fields and that large energy gains are possible. In addition, to minimize the spread in final energy,

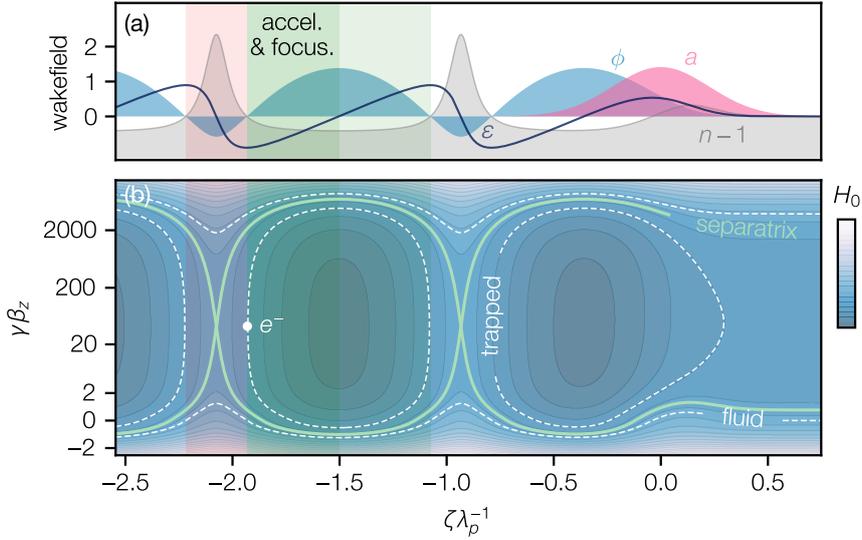


Figure 1.3 – Single electron motion in a laser-driven plasma wakefield. The parameters are the same as in Fig. 1.2(c),(d) ($n_e = 10^{18} \text{ cm}^{-3}$, $a_0 = 2$, $c\tau_0 = 2k_p^{-1}$). (a) Time-averaged laser vector potential \bar{a} , excited wakefield potential ϕ , density modulation $n - 1$ and longitudinal electric field ϵ . Focusing and defocusing regions are indicated as green- and red-shaded areas. (b) Possible electron trajectories in phase space ($\zeta, \gamma\beta_z$) and contours indicating constant levels of H [Eq. (1.35)]. Plasma background electrons populate the cold fluid orbit ($H_0 = 1$) and effectively stream towards the left in the comoving frame. Electrons that are injected onto a trapped orbit ($H_t < H_s$) can get accelerated in the plasma wave. Trapped and untrapped regions in phase space are separated by a separatrix (H_s).

the populated phase space region should be small, which also implies that the injection process has to be localized in space and time. The above can either be realized via external injection of a pre-accelerated bunch or through various internal injection processes, which are summarized in the following.

Self-trapping of plasma background electrons can occur as a result of wave breaking, meaning that the excitation is so strong that the fluid momentum exceeds the phase velocity of the plasma wave ($\gamma \geq \gamma_p$) [4]. Within the 1D theory considered here, this threshold is reached when the wakefield amplitude approaches the *cold relativistic wave breaking limit* [63]

$$E_{\text{WB}} = \sqrt{2(\gamma_p - 1)}E_0. \quad (1.38)$$

where $\gamma_p \simeq \omega/\omega_p$ is the relativistic factor associated with the wake phase velocity and E_0 is the cold non-relativistic wave breaking limit that was defined earlier [Eq. (1.5)]. Revisiting the previous example, where $n_e = 10^{18} \text{ cm}^{-3}$ and $\lambda = 800 \text{ nm}$, we find $\gamma_p \simeq 41$ and $E_{\text{WB}} \simeq 9E_0$.

In reality, the wave breaking limit is much lower as additional effects that are not captured by the 1D cold fluid theory, such as the radial shape and evolution of the plasma wave, and the finite plasma temperature [92], have to be considered. In the nonlinear regime, the curvature of the wakefield eventually leads to *transverse wave breaking* [93]. In addition, driver amplitude or plasma density variations can lower the threshold for *longitudinal wave breaking* [94] by temporarily reducing the wake phase velocity. It is for the same reasons, however, that electron injection due to wave breaking is difficult to control and sensitive to fluctuations in laser and plasma parameters. As a consequence, advanced injection techniques with greater control over the trapped phase space have been proposed.

This includes, for example, injection in tailored density profiles—either by triggering longitudinal wave breaking in short ($L \sim \lambda_p$) density transitions [95], or by continuously lowering the wake phase velocity in long ($L \gg \lambda_p$) density downramps [94]. Other popular injection techniques are colliding pulse injection [96, 97] and ionization-induced injection [98, 99]. In the first case, plasma background electrons are pre-accelerated in the beat wave of two colliding laser pulses and thereby lifted onto trapped phase space orbits. In the second case, new electrons are born at trapped wakefield phases from ionizing the more strongly bound levels of a dopant gas, either with the help of a secondary laser pulse [98] of higher frequency [100], or in the high-intensity region of the driver itself [99]. The latter mechanism was used in this work and is therefore explained in more detail in the following.

Ionization injection

So far, we have assumed that the plasma consists of fully pre-ionized atoms and electrons. In practice, the plasma is usually created via field-ionization of a low atomic number gas (e.g. hydrogen) by the drive laser itself. In this case, all electrons are already ionized prior to the interaction with the main pulse, because the threshold for ionization is much lower than the intensity required to drive the plasma wave.

This situation changes, however, when considering a high atomic number gas with more strongly bound electrons. As an example, Fig. 1.4(a) shows the ionization degree of the different energy levels of nitrogen throughout the interaction with a Gaussian laser pulse with $a_0 = 2$ and $\lambda = 800$ nm. Here, tunnel ionization is the dominant process [99] and was calculated based on the Ammosov-Delone-Krainov (ADK) model [101] using the formulas given in [102]. While the first five levels of nitrogen (N^{1-5+}) are ionized early, in the low-intensity pedestal of the pulse, the inner K shell electrons of nitrogen ($N^{6,7+}$) are released only later, in the high-intensity region.

An electron that is ionized at this delayed position (ζ_{ion}) can become trapped, as we see in the following. For the sake of simplicity, the electron is assumed to be born at

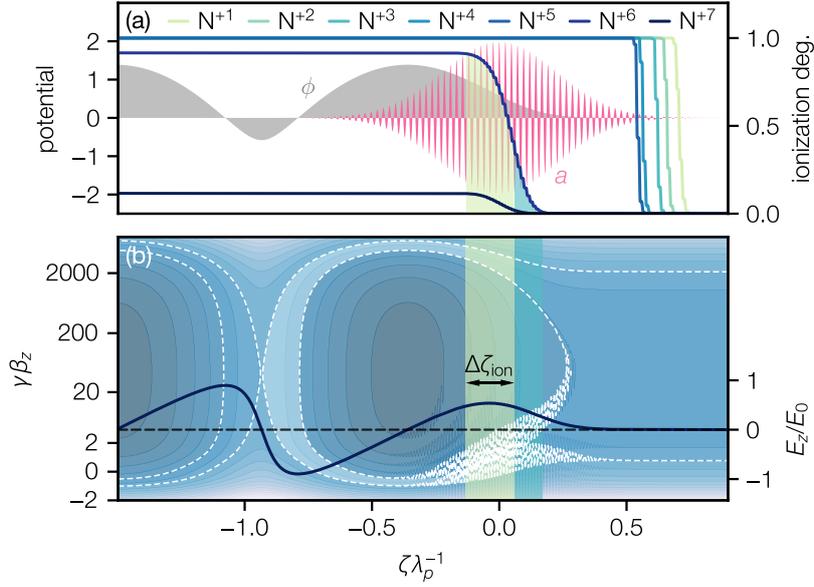


Figure 1.4 – Principle of ionization injection. Ionization degree of Nitrogen in the vector potential a of a Gaussian laser pulse ($a_0 = 2$, $\lambda = 800$ nm, $c\tau_0 = 2k_p^{-1}$) that excites a plasma wave with scalar potential ϕ ($n_e = 10^{18}$ cm $^{-3}$). The ionization degree increases as the atoms pass through the laser field from right to left in the comoving frame. The first five energy levels (N^{1-5+}) are fully ionized in the low-intensity pedestal. The last two levels ($N^{6,7+}$) are ionized in regions of highest intensity (blue/green area). (b) Possible electron trajectories in phase space $(\zeta, \gamma\beta_z)$ and wakefield amplitude E_z/E_0 . In a region of length $\Delta\zeta_{\text{ion}}$ (green area), where $H_{\text{ion}} < H_s$, the ionized electrons are born into trapped orbits (white) and can get accelerated in the plasma wave.

rest ($\beta_z = \beta_{\perp} = 0$). This is a good approximation, since the ionization probability is highest at the peak of the electric field where $a(\zeta_{\text{ion}}) \simeq 0$ for a linearly polarized laser. Note, however, that some electrons get ionized off-peak and thereby accumulate residual transverse momentum, which degrades the beam emittance, i.e., increases the occupied volume in transverse phase space [99].

The initial Hamiltonian of the electron is then simply given by [99]

$$H_{\text{ion}} = 1 - \phi(\zeta_{\text{ion}}). \quad (1.39)$$

Consequently, under consideration of Eq. (1.37), the condition for trapping ($H_{\text{ion}} < H_s$) can be rewritten as [99]

$$1 - \gamma_p^{-1} \leq \phi_{\text{ion}} - \phi_{\text{min}}, \quad (1.40)$$

where we assume $\gamma_{\perp} = 1$ ($a_{\perp} = 0$) at the position where the electron becomes trapped (ζ_{min}). The principle of ionization injection is illustrated in Fig. 1.4(b). Injection occurs in the phase space region $\Delta\zeta_{\text{ion}}$ where (1) the laser amplitude is such that a strongly

bound electron is ionized and (2) the potential ϕ_{ion} is large enough so that the electron resides on a trapped orbit.

The latter requirement is usually only fulfilled in nonlinear plasma waves for a driver strength $a_0 \geq 1.7$ [99]. This roughly coincides with the minimum intensity (assuming $\lambda = 800$ nm) required to ionize the more strongly bound levels of nitrogen, which is why nitrogen is often chosen as the dopant gas in practice. Note that simultaneously reaching and operating near the threshold for ionization and trapping is desirable, since the injected phase space volume is then minimized in both the transverse ($a(\zeta_{\text{ion}}) \simeq 0$) and longitudinal ($\Delta\zeta_{\text{ion}} \rightarrow 0$) direction [99, 103]. Equation (1.40) also highlights that the condition for trapping may be lowered by reducing the wake phase velocity γ_p , for example, by introducing a density downramp.

Finally, it should be noted that although the 1D theory discussed here is useful to describe the key features of ionization-induced trapping, 3D particle-in-cell simulations are usually required to accurately model the injection process in practice.

1.2.2 Acceleration limits

Once a bunch has been injected into the wakefield it gets accelerated to an energy

$$W \simeq eE_0 \varepsilon_{\text{avg}} L_{\text{acc}}, \quad (1.41)$$

where $\varepsilon_{\text{avg}} = E_{\text{avg}}/E_0$ is the normalized average accelerating gradient experienced by the electrons and L_{acc} is the total acceleration length. The value of E_{avg} depends on the plasma density, driver amplitude and size, and might be reduced due to beam loading as discussed below. The maximum value of L_{acc} is limited by electron dephasing, laser energy depletion and laser diffraction [4]. These limits are explained in the following.

Consider a short ($L_{\text{bunch}} \ll \lambda_p$) electron bunch that is injected at the back of a plasma wave bucket, in the region where the transverse fields are focusing. During the process of gaining energy, the electrons catch up with the laser as their velocity exceeds that of the wake, $\gamma_{\text{beam}} \gg \gamma_p$. As soon as the bunch slipped forward by approximately a quarter plasma wavelength ($\lambda_{np}/4$), the axial electric field switches sign and the electrons are decelerated again. The acceleration distance is thus limited to a *dephasing length* [104]

$$L_d \simeq \gamma_p^2 \frac{\lambda_{np}}{2} = \gamma_p^2 \frac{\lambda_p}{2} \begin{cases} 1 & \text{for } a_0 \ll 1 \\ (\sqrt{2}/\pi)a_0 & \text{for } a_0 \gg 1, \end{cases} \quad (1.42)$$

assuming a square-shaped driver with resonant length $L_{\text{drv}} = \lambda_{np}/2$ and $\beta_{\text{beam}} \simeq 1$. Since $L_d \propto n_e^{-3/2}$ and $E_0 \propto n^{-1/2}$, the maximum energy gain is inversely proportional

to the density, $W \propto n^{-1}$. Achieving higher energies in a single laser-plasma accelerator stage therefore generally involves operating at lower densities.

To give an example, we again consider a laser-plasma accelerator that is operated at a density $n_e = 10^{18} \text{ cm}^{-3}$ and driven by a laser pulse with $\lambda = 800 \text{ nm}$, $a_0 = 1.5$ and $L_{\text{drv}} \simeq 62 \text{ fs}$. In this case, dephasing limits the acceleration length to $L_{\text{acc}} \sim L_d \simeq 32 \text{ mm}$. Setting $\varepsilon_{\text{avg}} = \varepsilon_{\text{max}}/2$, we find for the final beam energy $W \simeq 1.2 \text{ GeV}$.

Another limiting factor is the driver losing energy as it excites the wakefield. In the broad driver limit, we can estimate a *pump depletion length* L_{pd} by equating the energy transferred to the wakefield with the energy contained in the pulse, $E_{\text{max}}^2 L_{pd} \simeq E_l^2 L_{\text{drv}}$ [4], where $E_l = a_0(\omega c m_e e)^{-1}$ is the peak electric laser field. Again, for a square-shaped pulse with length $L_{\text{drv}} = \lambda_{np}/2$, the 1D pump depletion length is given by [105]

$$L_{pd} \simeq \gamma_p^2 a_0^2 \varepsilon_{\text{max}}^{-2} \frac{\lambda_{np}}{2} = \gamma_p^2 \lambda_p \begin{cases} 2/a_0^2 & \text{for } a_0 \ll 1 \\ (\sqrt{2}/\pi)a_0 & \text{for } a_0 \gg 1, \end{cases} \quad (1.43)$$

Note that $L_{pd} \gg L_d$ for $a_0 \ll 1$ since the coupling between laser and plasma wave increases $\propto a_0^2$ in the linear regime. Conversely, $L_{pd} \sim L_d$ for $a_0 \gg 1$, which means that the acceleration process is more efficient in the nonlinear regime. In the previous example, laser energy depletion limits the acceleration length to $L_{\text{acc}} \sim L_{pd}/2 \simeq 61 \text{ mm}$.

To efficiently accelerate electrons to GeV energies and beyond, the previous scalings imply tens of centimeter acceleration lengths and driver strengths $a_0 \sim 1$. To reach such high intensities at reasonable laser powers, the pulse is usually focused to a small spot size of a few $k_p w_0$. Given a spot size $w_0 = 4k_p^{-1} \simeq 21 \mu\text{m}$, the peak amplitude $a_0 = 1.5$ in the example above is reached for a laser power of $\sim 70 \text{ TW}$.

The intensity of a Gaussian laser decreases by a factor of two at a *Rayleigh length* $z_R = \pi w_0^2/\lambda$ away from the focus position. The acceleration distance is thus primarily limited by diffraction to $z_R \simeq 1.8 \text{ mm} \ll L_d \leq L_{pd}/2$. Overcoming this limit requires to optically guide the laser pulse, either by relying on self-focusing effects or through the use of preformed plasma channels [96].

1.2.3 Beam loading

To illustrate the process of injection and acceleration, we have so far assumed a non-evolving driver and neglected any feedback from the electron bunch to the wakefield. Due to its space-charge force, however, the accelerated electron bunch drives its own wake which alters the laser-driven plasma wave. Through this process, which is referred to as *beam loading* [28–31], the bunch reduces the accelerating gradient along its current profile and thereby effectively absorbs energy from the wakefield.

Beam loading has fundamental implications for plasma acceleration. It limits the amount of charge that can get accelerated in a wakefield and determines the energy-transfer efficiency from plasma wave to electron bunch [106]. Moreover, it can serve as a mechanism for preserving the electron beam quality: A properly shaped bunch current profile can flatten the otherwise steep longitudinal electric field of the laser-driven plasma wave and thereby prevent the accumulation of correlated [28–30] (and uncorrelated [107]) energy spread.

Beam loading was first studied in the linear regime ($a_0 \ll 1$, $n_b/n_0 \ll 1$, where n_b is the peak electron bunch density) [28, 29]. In the linear regime, beam loading can be viewed as a superposition of a laser- and beam-driven wakefield. In the latter case, instead of the ponderomotive force, the plasma fluid electrons are displaced by the space-charge force of the more rigid relativistic electron beam [4].

A limit for the amount of charge that can be loaded into a linear wakefield of amplitude $\varepsilon_{\max} = E_{\max}/E_0 < 1$ and over a transverse area $A \gg k_p^{-1}$ is given by [29]

$$Q_{\max} = en_0 k_p^{-1} \varepsilon_{\max} A. \quad (1.44)$$

This assumes that the charge is contained in a short ($\ll k_p^{-1}$) bunch that is placed at the peak of the accelerating field. In this situation, all of the energy stored in the plasma wave gets absorbed by the bunch, corresponding to a *beam loading efficiency*

$$\eta_b = 1 - \varepsilon_b^2 / \varepsilon_{\max}^2 \quad (1.45)$$

of 100%, where $\varepsilon_b = E_b/E_0 = 0$ is the wakefield amplitude behind the bunch. However, since the accelerating gradient falls off to zero throughout the bunch, this situation also leads to 100% energy spread.

In contrast, by using a properly phased bunch of finite length and with a triangular current profile (head more charge than tail), it is possible to compensate for the longitudinal variation of the laser-driven wakefield [28, 29]. As shown in Fig. 1.5(a), such *optimal beam loading* conditions lead to flat accelerating fields that imprint no energy spread (assuming a non-evolving plasma wave and neglecting dephasing effects).

For this optimally loaded case, the constant accelerating field along the bunch E_b and the loaded charge Q_b are given by [29]

$$E_b = \cos(k_p \zeta_b) E_0, \quad (1.46)$$

$$Q_b = \frac{\sin^2(k_p \zeta_b)}{2 \cos(k_p \zeta_b)} Q_{\max}, \quad (1.47)$$

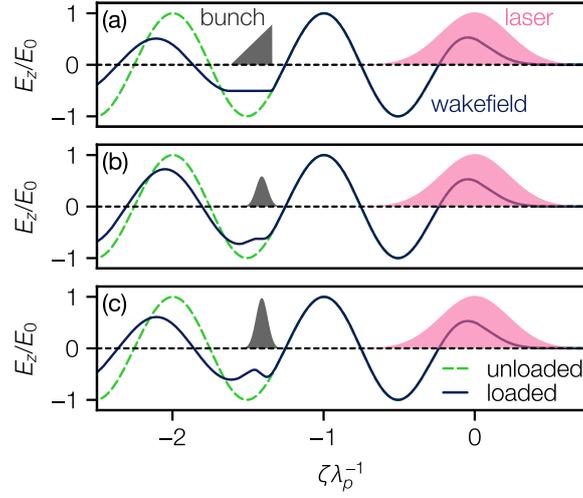


Figure 1.5 – Beam loading in the linear regime. Examples of a 1D laser-driven wakefield loaded with different bunch current profiles. (a) Flattening of the accelerating field E_z/E_0 due to optimal beam loading with a triangular profile ($k_p\zeta_b = \pi/3$, $k_p\Delta\zeta_b = \sqrt{3}$, $Q_b/Q_{\max} = 3/4$). (b) Partial flattening of the wakefield for a Gaussian profile. (c) Reversal of the accelerating field slope due to overloading.

with $k_p\zeta_b$ the position ahead of the wakefield minimum where the current profile begins. Accordingly, we can write for the energy absorbed per unit length [31]

$$Q_b E_b = E_0 \frac{en_0}{2k_p} \sin^2(k_p\zeta_b) \varepsilon_{\max}^2 A, \quad (1.48)$$

where $\eta_b = \sin^2(k_p\zeta_b)$ corresponds to the beam loading efficiency. Equation (1.48) illustrates that depending on the bunch length, there is a trade-off between charge and accelerating gradient, with the efficiency decreasing for shorter bunches. For the example shown in Fig. 1.5(a), the bunch head position, length and charge is $k_p\zeta_b = \pi/3$, $k_p\Delta\zeta_b = \sqrt{3}$ and $Q_b/Q_{\max} = 3/4$ [29]. Hence, the bunch is accelerated at 50% of the maximum unloaded gradient with an efficiency of 75%.

Note that non-ideal current profiles also flatten the fields to some degree, yet a residual curvature remains. Besides, a higher than optimal charge density locally reverses the slope of the acceleration field. Both effects are illustrated in Fig. 1.5(b) and (c) for a Gaussian bunch current profile of varying peak density.

As evident from Eq. (1.48), the total charge that can be loaded is a function of the cross-sectional area A out of which the beam absorbs energy. In consequence, the beam loading efficiency in the linear regime additionally depends on the transverse overlap between laser- and beam-driven wakefield [29]. For example, consider a wakefield with $\varepsilon_{\max} = 1$ driven by a Gaussian laser with radius $k_p w_0 = 4$ that is loaded according

to the optimal conditions defined above. For a wide beam with $A \sim \pi w_0^2/2$ and at a plasma density $n_e = 10^{18} \text{ cm}^{-3}$, we find for the charge $Q_b \sim 450 \text{ pC}$.

However, these considerations neglect that for a realistic radial driver profile, a wide beam would experience transversely varying longitudinal fields and nonlinear focusing forces that increase the energy spread and degrade the emittance [29, 31]. To preserve the beam quality, a beam cross-section $\ll k_p^{-2}$ is usually assumed. For a beam radius $\sigma_r < k_p^{-1}$, the effective wakefield area is $A_{\text{eff}} \sim k_p^{-2}$ [29]. In our example, this significantly reduces the efficiency as the charge would be limited to $\sim 20 \text{ pC}$. Note that a possible solution to this problem is the use of hollow plasma channels [108].

Beam loading was also studied analytically in the blowout regime, assuming an idealized near-spherical bubble with radius $k_p r_B \gg 1$. The physical interpretation of beam loading in this regime is as follows [30, 31]: The space-charge force of the loaded bunch repels the sheath electrons and reduces their transverse momentum as they return towards the axis in the rear half of the bubble. Thereby, the accelerating gradient is reduced and energy is absorbed from the wave. If the bunch current is too large, however, the transverse motion of the sheath is reversed, causing the axial field to become decelerating. The process of energy absorption is most efficient if the sheath electrons are slowly brought to rest so that they reach the axis with almost zero transverse momentum.

As in the linear regime, a current profile can be derived that results in flat accelerating fields along the bunch. The optimal profile in this case is trapezoidal (highest current at the bunch head) and the energy absorbed per unit length, i.e., the equivalent to Eq. (1.48), is given by [30, 31]

$$Q_b E_b = E_0 \frac{en_0}{2k_p} \frac{\pi}{8k_p^2} (k_p r_B)^4 = E_0 \frac{en_0}{2k_p} \left(\frac{k_p r_B}{2} \right)^2 \frac{\pi r_B^2}{2}. \quad (1.49)$$

The alternative notation on the right side can be derived from the linear expression [Eq. (1.48)] by setting $\varepsilon_{\text{max}} = k_p r_B/2$, $A = \pi r_B^2/2$ and $\sin^2(k_p \zeta_b) = 1$.

Equation (1.49) illustrates that, under ideal conditions, the beam loading efficiency can approach 100% in the blowout regime and does not depend on the bunch length and transverse size, which is an advantage over the linear regime [31]. Given our previous example and a blowout radius $k_p r_B = 4$, we find $Q_b \simeq 1200 \text{ pC}$ for $\varepsilon_b = 1$.

Note that the above analytical descriptions of beam loading are not applicable to many practical cases of laser-plasma acceleration. Figure 1.6 presents two examples of a beam-loaded wakefield that cannot be described with the existing theories [29, 30, 89]. Figure 1.6(a) shows a nonlinear plasma wave ($a_0 = 2$, $k_p w_0 = 4$) in which a narrow beam is accelerated that is dense enough ($n_b \gg n_0$) to drive a blowout and thereby significantly changes the laser-driven wakefield structure. Figure 1.6(b) shows a plasma

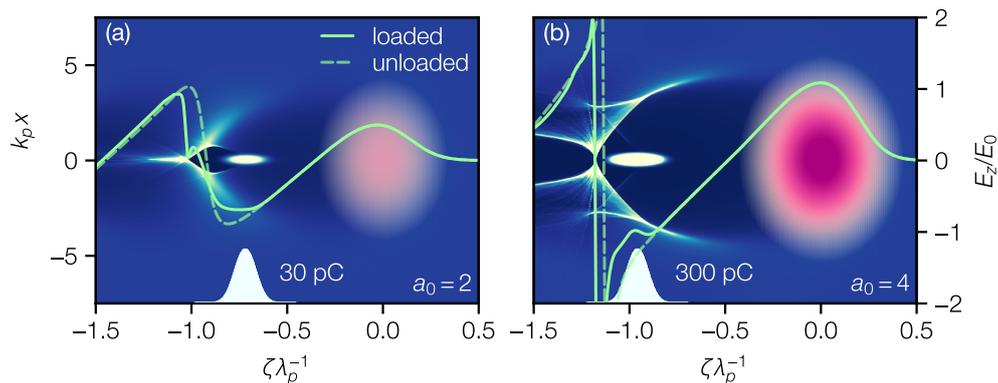


Figure 1.6 – Example of a beam loaded plasma wave in the (a) nonlinear and (b) blowout regime. Results were obtained from quasi-3D particle-in-cell simulations with FBPIC. In both cases, the wakefield is driven by a Gaussian laser ($\lambda = 800$ nm) with duration $c\tau_0 = 2k_p^{-1}$, spot size $k_p w_0 = 4$ and peak amplitude a_0 . The plasma density is $n_e = 10^{18}$ cm $^{-3}$. A Gaussian electron bunch with length $k_p \sigma_z = 1/3$, radius $k_p \sigma_r = 1/8$ and charge Q is positioned at ζ_0 . (a) $a_0 = 2$, $\zeta_0 = -24$ μ m, $Q = 30$ pC. (b) $a_0 = 4$, $\zeta_0 = -32$ μ m, $Q = 300$ pC $\simeq Q_b/4$ [compare Eq. (1.49)].

wave in the blowout regime ($a_0 = 4$, $k_p w_0 = 4$) [85, 87]. Note that the resulting sheath structure is much more complex [89] than the simple spherical bubble assumed in the corresponding analytical theory. Moreover, the laser does not drive a complete blowout and the remaining electrons are expelled by the electron bunch, which further complicates the situation.

As a consequence, it is difficult to predict the bunch current profile that flattens a given laser-driven wakefield. In particular, the optimal profile is not always trapezoidal, but various shapes, sometimes with the current increasing from head to tail, can be required [109]. Furthermore, operation at optimal beam loading conditions has to be viewed as an averaged process. Due to driver evolution, the wakefield can transition through different regimes during the acceleration process. In addition, dephasing of the electron bunch has to be taken into account. For these reasons, a comprehensive description of beam loading generally requires the use of particle-in-cell simulations.

Part I

Generation of stable high-quality electron beams

2 LUX laser-plasma accelerator

This chapter provides an overview of the LUX laser-plasma accelerator, focused on aspects relevant to the present work. In Sec. 2.1, we introduce the LUX facility and summarize the experimental setup. The laser system and diagnostics, as well as measurements of the pulse properties, are presented in Sec. 2.2. The target area, electron beam transport and diagnostics are described in Sec. 2.3.

2.1 Overview

LUX [58, 36] (Laser-driven undulator x-ray source) is an experimental facility for laser-plasma acceleration, with a focus on using the generated electron bunches to drive high brightness undulator radiation in the extreme ultraviolet and x-ray regime. It is located at Deutsches Elektronen-Synchrotron (DESY) and is built and operated in cooperation with the University of Hamburg. The overarching goal is the demonstration of a plasma-driven free-electron laser [25, 26], which imposes stringent requirements on the generated electron beam quality and stability.

The experiment is composed of four main elements: A high-power laser system, a target area for plasma acceleration, an undulator, and a transport beam line that includes laser, electron and x-ray diagnostics.

The present work introduces a novel plasma source design and provides new insights into the injection and acceleration of low energy spread electron beams. We therefore focus on the laser-plasma accelerator and restrict ourselves to describing the parts used for the corresponding experimental activities. We refer to more detailed descriptions of the laser system [110] and final focusing [111], the target area [112, 113], the electron beam transport [114] and the undulator and x-ray diagnostics [111].

Figure 2.1 shows a schematic overview of the LUX accelerator in the configuration used for the experiments in the context of this thesis. At the time of the measurements, the Ti:sapphire laser system ANGUS delivered pulses at 1 Hz, with an on-target energy of up to 3 J at pulse durations of ~ 35 fs FWHM, corresponding to peak powers of up to 80 TW. The laser was focused by a 2 m focal length off-axis parabolic mirror (OAP) at $f/25$ to a 25 μm FWHM spot size inside a 10-mm-long gas-filled target, resulting in a peak normalized vector potential a_0 of up to 2.4. A set of online laser diagnostics

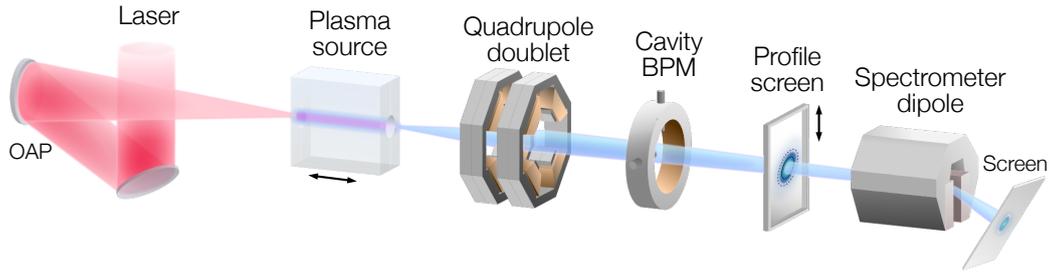


Figure 2.1 – Schematic overview of the LUX laser-plasma accelerator. High-power laser pulses are focused inside a gas-filled target where electrons are injected and accelerated in a nonlinear plasma wave. Electron bunches leaving the plasma source are captured by an electromagnetic quadrupole doublet and transported to a set of diagnostics. A cavity beam position monitor (BPM) measures the bunch charge and position, a scintillating screen the transverse beam profile, and a permanent dipole magnet disperses the beam to measure the energy spectrum.

measured the pulse energy and spectrum before the final compressor as well as the near- and far-field and the wavefront behind the final focusing mirror.

Electron bunches with tens of pC of charge were generated from localized ionization injection and accelerated to a few hundred MeV at plasma densities of $\sim 10^{18} \text{ cm}^{-3}$. Bunches leaving the plasma were either focused by a compact electromagnetic quadrupole doublet or allowed to naturally diverge in a free drift to a permanent magnet dipole spectrometer. Transverse beam profiles were measured on a retractable scintillator and the center of mass and bunch charge were non-invasively measured by a cavity beam position monitor (BPM). The latter was used to charge-calibrate the spectrometer.

The following sections provide a more detailed description of the drive laser and electron beam line. The plasma source design is discussed in Ch. 3. The experimental results are presented in Ch. 4.

2.2 Drive laser

2.2.1 ANGUS laser system and beam delivery

The ANGUS laser has been designed to deliver 25 fs short pulses at a peak power of 200 TW [115, 110]. The system is based on chirped pulse amplification (CPA) [5] in Titanium-doped sapphire at a central wavelength of 800 nm [116, 117]. The crystals are pumped by frequency-doubled neodymium lasers at $\sim 530 \text{ nm}$.

Broadband ($\sim 100 \text{ nm}$) and ultrashort (15 fs) pulses are generated in a mode-locked oscillator (Venteon). The seed pulses are then stretched ($> 100 \text{ ps}$) to reduce the

intensity and amplified from a few nJ to 400 μ J in a regenerative amplifier. In the process, the repetition rate is reduced from 83 MHz to 1 kHz with a Pockels cell.

Subsequently, the pulses are compressed and sent through a cross-polarized wave (XPW) generation stage to enhance the temporal contrast [118]. The nonlinear process also recovers spectral bandwidth that was lost from gain narrowing in the previous amplifier [110]. After stretching the pulses a second time, they are sent through an acousto-optic programmable dispersive filter [119] (Fastlite Dazzler HR 800) which allows shaping the spectral intensity and phase. The energy lost in the various processes is recovered in a double-pass amplifier before reducing the repetition rate to ≤ 5 Hz by another pulse picker. Both Ti:sapphire amplifiers are driven by a single diode-pumped Nd:YLF laser that delivers 25 mJ pulses at 1 kHz (Thales JADE).

The 50 μ J pulses generated in the front end are further amplified in three multi-pass amplifiers that are pumped by frequency-doubled Nd:YAG lasers. These lasers are in turn pumped by gas discharge flash lamps, which limits their repetition rate to 5 Hz as most of the radiation is converted to thermal energy. The repetition rate of the main Ti:sapphire amplifiers has been further limited to 1 Hz to mitigate heat-induced deformation of the gold-coated diffraction gratings in the final in-vacuum compressor [57].

The first two stages are pumped with a total of 3×1.4 J (Thales SAGA), increasing the seed pulse energy to 35 mJ and 1.4 J, respectively. The last stage is pumped with 12 J (Thales GAIA) and amplifies the beam to its final energy of 6 J. Throughout amplification, the beam size is gradually increased to operate close to the saturation fluence while avoiding damage of optical elements. The pump lasers furthermore imprint a transversely flat-top beam profile (6th-order supergaussian) [110].

The output of each individual section of the laser chain is monitored with a set of online diagnostics. This involves at least a measurement of the near- and far-field (camera), power (photodiode) or pulse energy (pyroelectric sensor), and spectrum (fiber-optic spectrometer). The retrieved beam position and pointing is used to actively stabilize the input of the next stage. In addition, the temperature of all amplifier crystals is stabilized to ± 0.02 K.

Once fully amplified, a motorized half-wave plate in combination with a pair of thin-film polarizers allows variable attenuation of the pulse energy down to a minimum of ~ 100 mJ. For low-power characterization of the spatio-temporal pulse properties on target, the energy is reduced further to ≤ 1 mJ using neutral-density filters.

Prior to the final compression in vacuum, the beam is magnified once more to 72 mm FWHM to reduce the fluence and sent onto a deformable mirror (Imagine Optics ILAO Star 52) to correct for wavefront aberrations. A motorized lens in the beam expanding

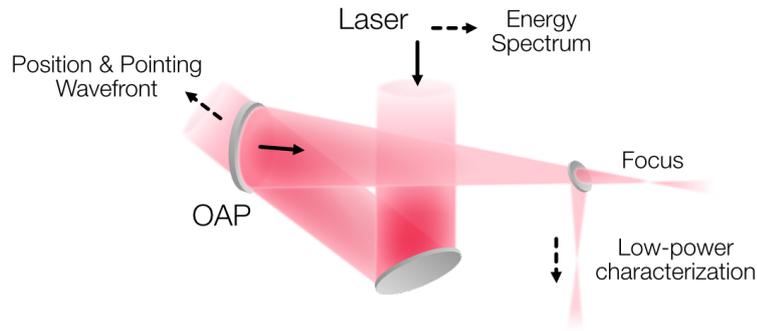


Figure 2.2 – Location of the main laser diagnostics used in the experiment. During high-power operation, the pulse energy and spectrum were monitored after the final amplifier using the part of the beam that was rejected in the attenuator. The beam position, pointing and wavefront were deduced from the leakage of the off-axis parabolic (OAP) mirror. For the characterization and optimization of the spatio-temporal pulse properties prior to experiments, the focused beam was deflected to a low-power diagnostic just before reaching the target.

telescope furthermore allows fine adjustment of the divergence and, thus, longitudinal position of the focused laser spot.

Finally, the compressed pulses are delivered to the accelerator tunnel via a 35-m-long in-vacuum transport beam line. Using a pair of high-precision mirror mounts, the beam is aligned onto an off-axis (15° offset angle) parabolic mirror (OAP) ($f = 2025$ mm) [111] that focuses the laser inside the target chamber. With respect to the electron beam line, the laser is polarized in the horizontal x direction, which coincides with the dispersion plane of the compressor.

2.2.2 Laser diagnostics

The ANGUS system and the accelerator beam line have been equipped with a plethora of laser diagnostics that serve different purposes. Some are used for offline characterization and optimization of the pulse properties, others for online monitoring of the system's performance and as input for active stabilization, and still others for synchronized shot-to-shot measurements of the key parameters that control, or are affected by, the laser-plasma interaction. Comprehensive overviews of these diagnostic capabilities are given in [110] and [111].

The present work primarily relied on two sets of diagnostics, whose location is schematically shown in Fig. 2.2:

- (1) A low-power diagnostic close to the interaction chamber that was used to optimize the spatio-temporal pulse shape prior to experiments.

- (2) A set of devices for high-power online measurements of the single-shot drive laser properties before the interaction with the plasma.

For the low-power characterization, the laser was attenuated to 1 mJ and deflected away from the propagation axis with a retractable mirror immediately before reaching the target. The beam was still focused in vacuum and exited the chamber through a thin window.

A pair of lenses ($f = 300$ mm) then collimated and refocused the divergent beam in the image plane of a $4\times$ microscope objective that was attached to a 12-bit camera (Basler acA1300-30gm). The entire setup was mounted on a motorized translation stage and allowed 1:1 imaging of the evolution of the transverse intensity profile in the target chamber.

In a second diagnostic branch, an additional lens ($f = 100$ mm) imaged the plane of the deformable mirror onto a Shack-Hartmann wavefront sensor [120] (Imagine Optics HASO 3), enabling closed-loop optimization of the focal spot quality. In yet another branch, the temporal pulse properties were retrieved either with self-referenced spectral interferometry [121] (Fastlite Wizzler 800) or frequency-resolved optical gating [122] (Swamp Optics Grenouille 8-20). The former measurement was fed back to the acousto-optical pulse shaper (Dazzler) to optimize the spectral phase.

To obtain accurate online measurements during high-power operation, great care was taken to minimize systematic errors and avoid a coupling between the measured properties when sampling the laser beam. As a consequence, the online diagnostics have been placed at two different locations.

The pulse energy and spectrum were monitored directly after the final amplifier using the part of the beam that was rejected in the attenuator and reflected by the second thin-film polarizer. The pulse energy was measured with a pyroelectric energy sensor (Gentec-EO QE50). The spectrum was measured with a fiber-coupled near-infrared spectrometer (Oceanoptics STS-NIR) that collected the diffuse light scattered off the ceramic surface of the energy sensor.

The setup had several advantages. (i) No additional optics had to be placed in the beam path. (ii) Because the attenuator was operated below full transmission, a significant fraction of the energy was available for the measurement which increased its robustness and accuracy. (iii) Since the sampling method relied on the polarization-dependent reflectivity, it did not alter the spectrum of the probe beam. In contrast, this is usually a problem when using the leakage of a dielectric mirror as probe beam, since in this case small variations of the wavelength-dependent reflectivity can lead to large variations in the transmitted spectrum [110].

The wavefront, beam profile and focal spot were simultaneously measured directly behind the off-axis parabolic (OAP) mirror. An in-vacuum lens collected the transmitted

leakage of the OAP and focused it through a window. Outside the chamber, the beam was split twice and either reached its focus in the image plane of a $4\times$ microscope objective that was attached to a camera (Basler acA1300-30gm), or, was collimated by additional lenses ($f = 25/50$ mm) that imaged the mirror surface onto a wavefront sensor (Imagine Optics HASO4 First) and a near-field camera (Basler acA1920-40gm), respectively.

Because the propagation through the curved mirror introduced aberrations, the setup allowed for relative measurements only. However, this was outweighed by the advantage of measuring the spatial beam properties as close as possible to the plasma source. This was indeed critical because, for example, heating of the compressor and beam line optics could slowly change the wavefront properties, in particular the defocus, over time.

Through calibration with two permanent laser alignment screens that were positioned up- and downstream of the plasma source, the near- and far-field measurement also served as a permanent alignment reference that defined the laser beam position and pointing with respect to the accelerator design axis.

In the following, we present measurements that are representative for the performance of the ANGUS laser system during the time of the experiments. The measurements either have been performed beforehand at the pre-target diagnostic in low-power mode (1 mJ pulse energy), or in parallel with the electron beam measurements using the high-power online diagnostics. For detailed explanations of some of the standard methods for characterizing high-power laser systems, we recommend [123].

2.2.3 Low-power characterization

The focal spot quality has been characterized with the wavefront sensor and the mode imaging system. Closed-loop optimization with the deformable mirror reduced the absolute wavefront aberrations to below 70 nm rms, which resulted in a Strehl ratio of > 0.85 .

The longitudinal evolution of the focal spot in vacuum is shown in Fig. 2.3. The measurement is composed of 100 images that have been obtained from an automated camera position scan through the focal plane. Figure 2.3(d) shows the intensity distribution at the focus position ($z = 0$ mm). The measured spot size was $24.1\ \mu\text{m}$ and $25.5\ \mu\text{m}$ in x and y , respectively. The distorted Airy pattern around the central spot originates from the high spatial frequency components that are present in the flat-top intensity profile before focusing.

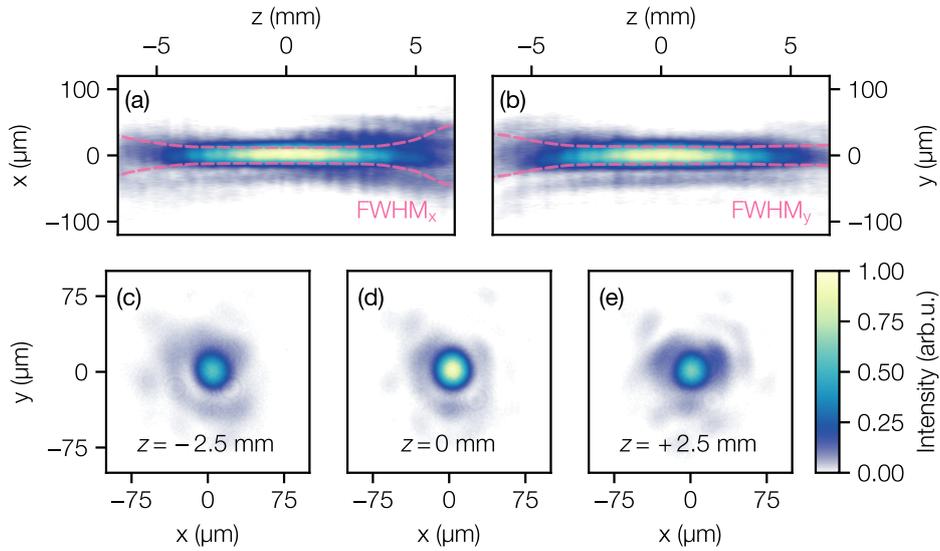


Figure 2.3 – Measurement of the evolution of the focused laser spot in vacuum. Transverse intensity distribution for different longitudinal positions z in (a) the horizontal plane x ($y = 0$) and (b) the vertical plane y ($x = 0$). FWHM spot size (magenta dashed line). Images of the focused spot (c) 2.5 mm before the focal plane, (d) in the focal plane, and (e) 2.5 mm after the focal plane. The step size of the scan was $\Delta z = 156.25 \mu\text{m}$.

In addition, we performed a high dynamic range measurement of the focal spot by combining multiple images at different camera saturation levels. The relative intensity level decreased to below 10^{-3} within a radius of $\sim 250 \mu\text{m}$ around the central peak.

Figure 2.3(c) and (e) show the intensity distribution 2.5 mm before and after the focal plane. In this range, the spot keeps a distinct round shape suitable to drive the laser-plasma interaction. As the pulse propagates further out of focus, diffraction and residual wavefront aberrations cause the intensity distribution to become increasingly distorted. This effectively limits the maximum interaction length to $\sim 7.5 \text{ mm}$ to ensure a homogeneous and axisymmetric plasma wave excitation throughout the entire injection and acceleration process. Guiding of the laser, for example due to relativistic self-focusing, can further extend this distance during the propagation through the plasma.

In particle-in-cell simulations of the laser-plasma interaction, the measured intensity distribution was approximated by a so-called flattened Gaussian laser model [124]. The model provides analytical solutions to the evolution of focused flat-top intensity profiles. A flattened Gaussian laser is defined as a weighted sum of cylindrically symmetric Laguerre-Gauss modes. It is characterized by a waist and an order N , which we choose such that the profile approximates the experimentally measured evolution of the focal spot. Figure 2.4 compares the transverse laser intensity through the focus for

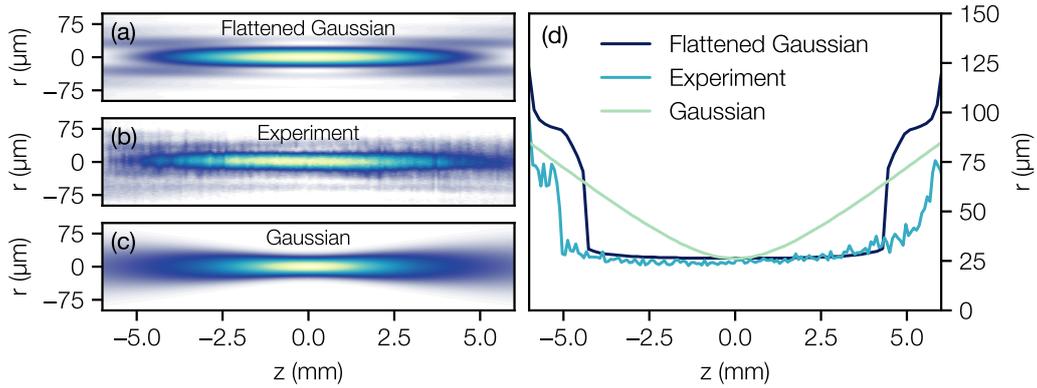


Figure 2.4 – Comparison of different analytical laser models for approximating the measured longitudinal evolution of the focal spot. Intensity trace (a) of a flattened Gaussian laser of order $N = 100$, (b) from an experimental measurement and (c) of a standard Gaussian laser ($N = 1$). (d) FWHM spot size corresponding to (a)-(c).

a flattened Gaussian laser ($N = 100$) that was used in simulations, for an experimental measurement, and for a standard Gaussian laser ($N = 1$). The transverse structure and the evolution of the FWHM spot size in the vicinity of the focus position is much better represented by the flattened Gaussian laser than by the standard Gaussian laser.

The temporal pulse properties and stability were characterized with the Wizzler. The reconstructed intensity profiles of 300 consecutive laser shots are shown in Fig. 2.5(a). On average, the FWHM pulse duration of this series was (36 ± 0.5) fs, corresponding to a relative rms stability of 1.3%. The preceding optimization of the spectral intensity and phase with the Dazzler furthermore resulted in a symmetric pulse shape and a sub-picosecond contrast ratio of at least 1000:1. The average intensity profile is shown in Fig. 2.5(b). It drops to relative intensity levels in the per-mille range within ± 100 fs around the peak. A complementary Grenouille measurement yielded an FWHM pulse duration of 37.9 fs.

Coarse tuning of the pulse front tilt and spatial chirp in the dispersion plane of the compressor was possible using the Grenouille. More precise adjustments of the angular dispersion were based on measurements of the focal spot evolution. Chromatic aberrations in the amplified beam introduced a small wavelength-dependent spread of the longitudinal focus position. In the presence of angular chirp, this would lead to a tilt of the measured intensity trace [Fig. 2.3(a)]. Minimization of the effect reduced the pulse front tilt to < 8 μrad .

The longitudinal focus position with respect to the plasma source was determined by measuring the recombination light that was emitted after the laser ionized a low-density background gas. For the measurement, the pulse energy was increased to ~ 250 mJ

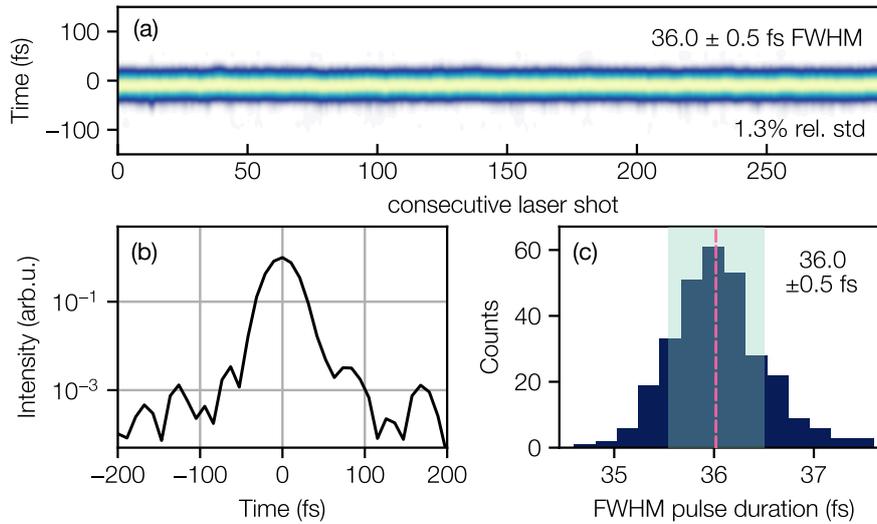


Figure 2.5 – Measurement of the temporal pulse shape in low-power mode and pulse duration statistics. (a) Temporal intensity profile of 300 consecutive laser shots. (b) Average intensity profile in logarithmic scale. (c) Histogram of the measured pulse durations (FWHM of intensity). Average pulse duration (magenta dashed line) and standard deviation (green shaded area).

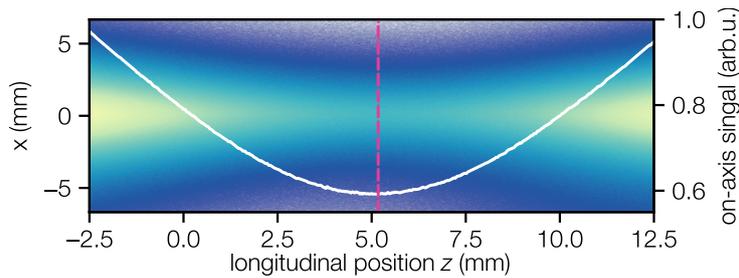


Figure 2.6 – Side view of the recombination light of laser-ionized hydrogen in the interaction chamber. The width and strength of the signal is a measure of the laser spot size. The focus position (magenta dashed line) is assumed at the minimum of the on-axis signal (white line).

and the laser was focused in the interaction chamber that was filled with hydrogen at a few mbar pressure. From the side, a camera observed the recombination light of the laser-ionized plasma (Fig. 2.6). Since the intensity was well above the ionization threshold of hydrogen, the width and strength of this signal is an indirect measure of the laser spot size. Consequently, the focus was assumed at the position where the measured trace has its minimum ($z_{\text{foc}} = 5.15$ mm). The resolution of this method was estimated to be on the order of ± 1 mm.

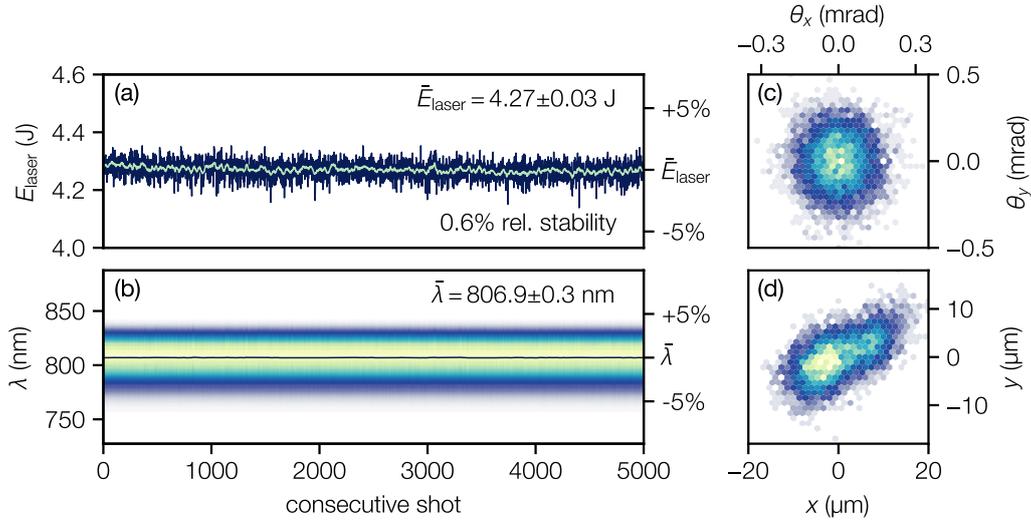


Figure 2.7 – Drive laser stability. High-power online measurements of key laser properties and corresponding statistics over 5000 consecutive shots. (a) Pulse energy E_{laser} after full amplification. Single-shot (blue) and 25-shot moving average (green). (b) Laser spectrum λ and 25-shot moving average of the central wavelength $\bar{\lambda}$ (line). (c) Laser pointing angle θ through the plasma source in the horizontal x and vertical y direction. (d) Transverse position of the focal spot in the plasma.

2.2.4 High-power online measurements

Despite the widespread assumption that the primary sources of instability in today’s plasma accelerators originate from the driver, the laser input parameters are rarely measured on a shot-to-shot basis in such experiments. In contrast, the online diagnostics of our setup provided shot-synchronized records of the most important drive laser properties. Based on these measurements, we summarize the laser stability below. In Ch. 4, we then use the same data to explain the interplay between drive laser and electron beam stability.

Figure 2.7 shows the laser performance over 5000 consecutive shots, i.e., ~ 1.4 hours of operation. The pulse energy was (4.27 ± 0.3) J, which corresponds to a relative rms stability of 0.6%. Taking into account the compressor efficiency and losses during the beam transport, this translated to an on-target energy of 2.6 J (60% beam line transmission). The amplified spectrum was centered around a mean wavelength of (806.9 ± 0.3) nm and had a bandwidth of (32.2 ± 0.2) nm FWHM.

The laser pointing through the target and the transverse position of the focus are shown in Fig. 2.7(c) and (d). The propagation angle varied by 0.08 and 0.14 mrad rms, the focus position by 6.2 and 4.2 μm rms in x and y , respectively. Although having a direct effect on it, the laser pointing could only partially explain the measured electron

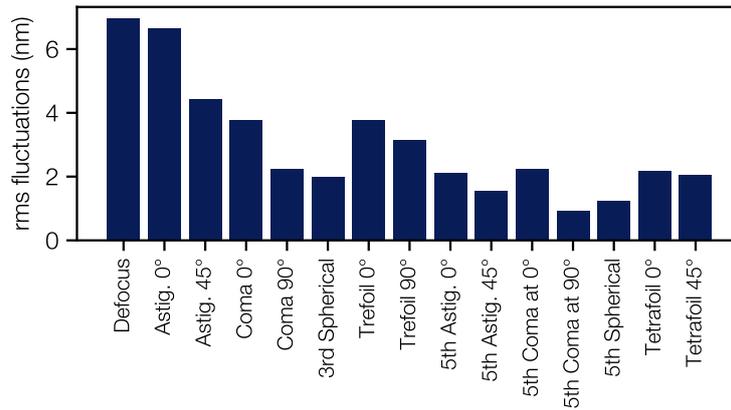


Figure 2.8 – Laser wavefront stability. Standard deviation (rms fluctuations) of the first 15 Zernike polynomials (excluding piston and tilts) measured over 5000 consecutive laser shots.

beam pointing jitter. Spatio-temporal couplings, such as pulse front tilt, were most likely responsible for the residual fluctuations [125, 126].

The recorded wavefront slopes were fitted with Zernike polynomials [127] (24 terms, Wyant indexing, RMS normalization) using a fixed circular pupil. Figure 2.8 shows the rms jitter of the first 15 Zernike polynomials (excluding piston and tilts) for the same 5000 shots that are shown in Fig. 2.7. The tilts and defocus were used to calculate the transverse and longitudinal focus position in the plasma source. For the given focusing setup, the measured defocus variations of 6.9 nm rms translated to a longitudinal focus position jitter of $\sim 162 \mu\text{m}$. Separate measurements indicated that the observed fast changes of the wavefront are caused by air density fluctuations, and may also originate from pump-induced variations in the amplification process. The latter mechanism, however, has not yet been verified.

It should be noted that the stability of the recorded parameters was close to the specified repeatabilities of the used measurement devices. The repeatability of the wavefront sensor was $< \lambda/200$, that of the energy sensor was $< 0.5\%$. However, the presence of meaningful correlations between these parameters and the electron beam properties (compare Ch. 4) indicates that the accuracy of the measurements exceeded these specifications. Still, for future experiments aiming at even better stability, devices with higher precision will be needed.

2.3 Electron beam line

In the following, we give a brief overview of the target area, which accommodated the plasma source, the electromagnets that were used to transport the accelerated bunch, and the electron beam diagnostics.

2.3.1 Target area and electron beam transport

The plasma source was located in a compact target chamber and installed on top of a 5-axis positioning system that was actuated by out-of-vacuum motors [112]. Diagnostic windows provided optical access from multiple sides. A set of five identical metal pipe feedthroughs, of which three could be used to supply gas and two were reserved for online pressure measurements, connected the in-vacuum setup to an outside gas system [113]. The plasma source and gas system are described in detail in Ch. 3.

The vacuum system [112] supported continuous-flow operation of the plasma source at supply rates of up to $\sim 1 \text{ l}_n/\text{min}$ of hydrogen gas. A tapered beam pipe design and additional apertures whose diameter was matched to the laser waist minimized the gas flow between adjacent beam line sections that were differentially pumped by turbomolecular pumps. The system was able to reduce the pressure by 8 orders of magnitude over a distance of only two meters in the upstream direction, from $\sim 100 \text{ mbar}$ inside the plasma source to 10^{-6} mbar at the position of the final focusing mirror [112].

Upon acceleration, the electron bunches were captured and transported downstream the beam line. The large focusing forces of the plasma wakefield entail small matched beam sizes and cause the electrons to diverge strongly after leaving the plasma. Consequently, to mitigate chromatic emittance growth [128, 129] during the beam transport, the diverging beam was focused by an electromagnetic quadrupole doublet [114] (customized DESY XQA [130, 131]) that was situated directly behind the target chamber, at a distance of only $\sim 100 \text{ mm}$ from the plasma source. With gap sizes of 12 and 22 mm, the 10-cm-long magnets reached maximum focusing strengths of 150 and 70 T/m, respectively, and were capable of capturing beams with energies of up to 450 MeV over a distance of only 400 mm [114].

In addition, two pairs of corrector dipoles [114] that were installed along the beam line could be used to steer the electron beam. These magnets were, however, not needed as small adjustments to the laser alignment provided sufficient control over the electron beam position and pointing.

2.3.2 Electron beam diagnostics

Cavity beam position and charge monitor

The beam position and charge were measured with a cavity beam position monitor (BPM) [111] that was originally developed for the European XFEL [132]. In this non-invasive measurement, the beam passes through a resonator cavity and excites electromagnetic modes that are picked up by a fast readout electronic [133, 134]. The total bunch charge and transverse center of mass position are derived from the TM_{01} (monopole) and TM_{11} (dipole) mode signal, respectively.

Being located 2.7 m behind the plasma source, the BPM had a half-angle acceptance of ~ 5 mrad for unfocused electrons. The position measurement had a resolution of ~ 10 μm and was calibrated against the direct beam profile measurements described below. The estimated relative accuracy of the charge measurement was $\sim 10\%$.

Beam profile screen

Transverse phase space characteristics, such as the divergence and pointing out of the plasma source, were derived from unfocused measurements of the beam profile with a screen station [114]. As with the BPM, the design of the screen station was based on developments for the European XFEL [135, 132].

For the measurement, the beam passed through a scintillating screen, a 400 μm thick LYSO:Ce crystal, that was driven into the beam path. The emitted light was collected by a 2x magnifying objective (Schneider Makro/Symmar $f = 120$ mm) that imaged the backside of the crystal onto a camera (Basler avA2300-25gm). To prevent a falsification of the measurement from optical transition radiation, the screen was oriented perpendicular to the beam axis and observed at a 45° angle using a Scheimpflug arrangement for the optics.

The screen station was located 3.1 m behind the plasma source. Considering both transverse axes simultaneously, the measurement had a maximum radial acceptance of ± 2.8 mrad and a spatial resolution of at least 60 μm . During the analysis, the recorded images were processed to reduce artifacts from noise and the signal was fitted with a 2D Gaussian to derive the relevant statistical properties.

Note that a second pair of transverse beam diagnostics, i.e., a combination of BPM and profile screen, was installed towards the end of the beam line. However, only the setup that was closer to the plasma source was used in the experiment, because its larger acceptance angle increased the signal-to-noise ratio of the beam profile measurement and ensured that the whole charge was detected when the beams were not focused by the quadrupole doublet.

Electron spectrometer

The electron spectrometer [111, 114] allowed a seamless measurement of the energy between 50–2500 MeV. A 40-cm-long permanent dipole magnet with a peak field strength of 0.96 T dispersed the electron beam in the vertical direction onto a scintillating screen (CAWO OG 16) that was mounted at 45° on the backside of a thin-walled vacuum chamber. Two objectives (Kowa LM12HC, $f = 12.5$ mm) each imaged a different section of the 16×800 mm wide screen onto a camera (Basler acA1920-40gm), covering a low- (50–123 MeV) and high-energy (123–1000 MeV) range, respectively.

The spectrometer entrance was located 8.2 m from the plasma source and had an acceptance angle of ± 1.3 mrad. Along the non-dispersive x -axis, the acceptance was additionally limited by the screen width. Dipole edge focusing furthermore resulted in a demagnification (magnification) of the beam by a factor of 2–3 in the vertical (horizontal) direction. As a consequence, the signal could be truncated in the non-dispersive plane when measuring unfocused electron beams. This had, however, only a negligible influence on the reconstructed spectrum in the energy range of interest.

The raw camera images were corrected for optical distortion, processed to reduce noise, fused together, and calibrated in terms of size and energy. The energy resolution was $\leq 1\%$ between 220 and 310 MeV for unfocused beams with rms divergences ≤ 2 mrad and $\sim 0.1\%$ for energies that were focused onto the screen. More details on the dipole magnet characterization, the design of the imaging system, and the energy calibration and analysis can be found in [111, 114, 113].

Determining the spectral charge density additionally required knowledge of the absolute charge associated with the energy-calibrated spectrum. If the bunch charge is not measured independently, this relation has to be derived by charge-calibrating the spectrometer against a known electron beam source, which is demanding. An alternative is to at least calibrate the imaging system and use measured estimates for the photon yield of the scintillating screen [136, 137]. This method, however, can be prone to error. In contrast, we derived the calibration directly from the BPM charge measurement. Since this still required making some assumptions, we summarize the analysis that underlies the results of Ch. 4 below.

For the main measurement, the electron beams were unfocused and the spectrometer was sensitive in the energy range from 50 to 1000 MeV. Because the measured spectra did not extend to low energies, only the high-energy camera signal (> 123 MeV) was used for the analysis of the electron beam properties. While no significant spectral content was measured in the low-energy range (50–123 MeV), the spectrometer could not detect very low energy (< 50 MeV) electron background that was still detected by the BPM for unfocused beams.

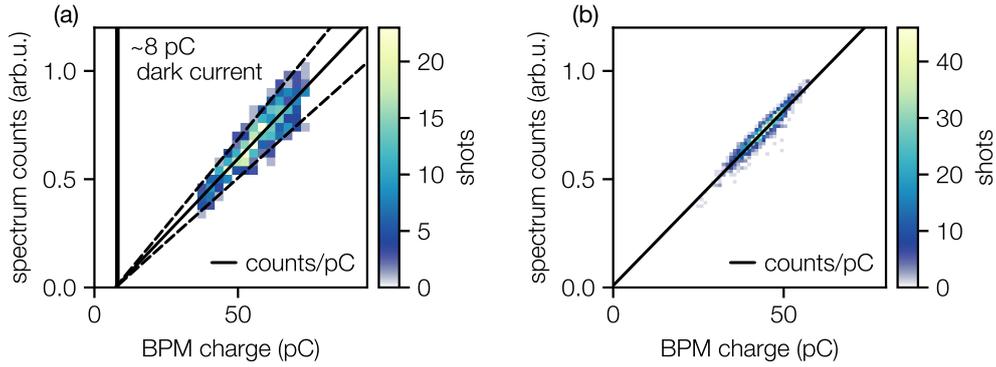


Figure 2.9 – Spectrometer charge calibration. (a) Correlation of BPM charge and summed electron spectrum counts (> 123 MeV) for unfocused electron beams with < 0.25 mrad pointing deviation from the central beam line axis and charges in the range $\bar{Q} \pm 1.5 \times \sigma_Q$ (54 ± 19 pC); Linear fit (solid line) and $\pm 15\%$ slope error (dashed lines). The charge offset indicates that ~ 8 pC of low-energy background electrons were detected by the BPM, which were not measured on the spectrometer. (b) Correlation of BPM charge and summed electron spectrum counts (> 123 MeV) for focused electron beams with median energies in the range (280 MeV $\pm 5\%$) (± 7 MeV); Linear fit (solid line). The correlation indicates that for focused beams all charge measured by the BPM was measured on the spectrometer.

Figure 2.9(a) shows a correlation of the measured BPM charge with the summed spectrum in the high-energy range (> 123 MeV) of the spectrometer, from which we estimated a total low-energy dark current of ~ 8 pC. This offset was subsequently subtracted from all measured charges for the unfocused measurements. The charge calibration of the spectrometer (counts/pC) was derived from the slope of the correlation.

For the focused measurements, the entire spectrum of the narrow energy spread beams was imaged onto the spectrometer screen. Low-energy electrons were over-focused and not transported to the diagnostics. Figure 2.9(b) shows the correlation of the measured BPM charge with the summed spectrum for beams in the vicinity of the focused energy (280 MeV). No charge offset is present, which confirms the absence of low-energy electron background. The charge measured by the BPM was used to calculate the spectral density of each individual spectrum. As a cross-check, we used the slope of the correlation (counts/pC) as a charge calibration of the spectrometer instead of using the single-shot measured charge. The peak spectral density statistics obtained with either method agree to the percent level. Because of the accuracy of the BPM charge measurement and the assumptions made to calibrate the spectrometer, we estimate the measured peak spectral densities to be accurate within $\pm 20\%$.

3 Development of a localized ionization injection plasma source

This chapter presents the development of a novel plasma source for the stable generation of high-quality electron beams via localized ionization injection. In Sec. 3.1, we review previous work to motivate our approach, followed by a description of the plasma source and the gas supply system. In Sec. 3.2, we present computational fluid dynamics simulations used to derive the gas density profile, as well as a measurement of the dopant gas localization.

Co-author contributions: P. Messner and S. Jalas contributed to the realization of the plasma source, and the dopant gas localization measurement. T. Parikh helped performing the CFD simulations.

3.1 Plasma source

The plasma source is arguably the most important component of a laser-plasma accelerator. Ideally, its design unites several features in a single device. This includes, for example, support for a controlled injection technique, tunability of parameters, steady-state operating conditions, external laser guiding, and optical accessibility. While simple designs already meet several of these criteria, plasma sources are constantly evolving and more complex designs are being invented, targeting ever increasing functionality.

In the following, we briefly review existing plasma sources to motivate the development of a novel design for the stable generation of high-quality electron beams from localized ionization injection. We then describe the experimental realization of the plasma source and the gas supply and monitoring system.

3.1.1 Motivation and previous work

In general, two types of plasma sources are used for laser-plasma acceleration, namely gas jets [138] and gas-filled volumes, such as cells or capillaries [139, 140]. With the former type, a supersonic nozzle forms a directed flow of gas that emerges freely into vacuum. In gas-filled volumes, on the other hand, pressure builds up due to the flow

resistance at small entrance and exit apertures. While gas jets are usually supplied with short bursts of highly pressurized gas, cells and capillary-type structures are operated with a continuous flow of gas.

Usually, a low atomic number gas, such as hydrogen or helium, is used and the plasma is formed instantaneously through optical field ionization in the low-intensity pedestal of the laser pulse [60]. With capillary-type plasma sources, the gas is sometimes pre-ionized by an electrical discharge, which, through Ohmic heating and subsequent cooling at the channel walls, can imprint a transverse plasma profile that guides the laser [139, 141]. Such plasma guiding channels were essential to demonstrate multi-GeV energy gains in a laser-plasma accelerator [142, 16]. Guiding channels in gas cells or jets have been realized from hydrodynamic expansion of a plasma column created by a secondary laser [143, 144].

As mentioned before, a critical feature of any plasma source is its ability to integrate or support controlled injection mechanisms. Optical injection techniques, such as colliding pulse injection [96, 97], rely on multiple laser pulses that are incident from different directions. Since the plasma is easily accessible from all sides, gas jets have so far been used in these experiments [18, 22]. Injection from a sudden decrease in wake phase velocity [94, 95, 145] requires steep plasma density downramps. Shock-induced transitions have been created by inserting a thin blade into the supersonic flow of a gas jet [146, 21, 24]. Density downramps have also been realized by embedding a supersonic nozzle into a capillary-type structure [20].

Both injection methods have been successfully used for generating high-quality electron beams with 1% level energy spreads [22–24]. However, a general drawback of the supersonic gas jets used in these experiments is that the highly dynamic change in gas pressure and velocity can potentially lead to instabilities during the injection and acceleration process [147]. Furthermore, in order to limit the gas load on the vacuum system, these types of sources are usually limited to repetition rates < 10 Hz. This contrasts continuous-flow designs, which promise very stable operating conditions [140, 112, 36] and by definition allow for much higher repetition rates. An exception worth mentioning, however, is the demonstration of few-MeV electron beams at 1 kHz repetition rate using a continuous-flow gas jet [148, 149].

Electron beam generation in gas cells or discharge capillaries has so far relied on self-trapping from wave breaking [93, 3, 13] or ionization-induced injection [150, 19, 151]. As it has proven difficult to control and to lack reproducibility, application of the former technique is in decline. The latter technique, on the other hand, is widely used in practice, as it is easy to implement and relatively robust against fluctuations [36].

To use ionization injection, it is enough to add a small amount of dopant gas, a high atomic number gas such as nitrogen or argon, to the primary gas. Because the higher

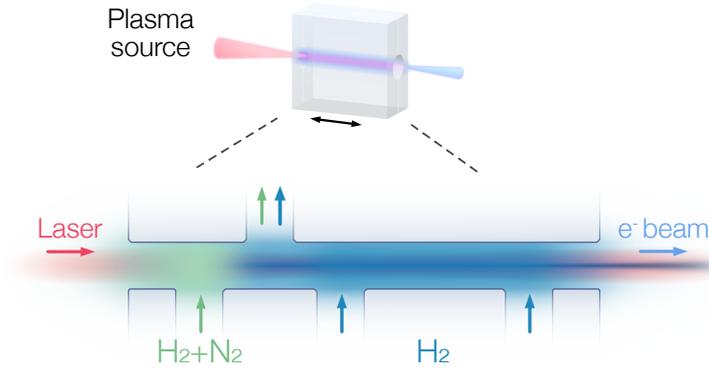


Figure 3.1 – Plasma source design with a dedicated injection and acceleration section. The plasma source consists of a micromachined structure that has been milled into a sapphire crystal. Three inlets supply a 5-mm-long interaction channel with hydrogen (H_2). In the front, the gas is doped with nitrogen (N_2) for localized ionization injection of electron bunches into the laser-driven plasma wave. An additional vacuum outlet separates both sections.

atomic levels are ionized only in the intense region of the drive laser, electrons are released at the correct wake phase to get trapped [99].

However, compared to the previously mentioned injection techniques, this straightforward implementation of ionization injection can result in a poor beam quality. Since the dopant gas is usually distributed along the entire propagation distance, ionization injection tends to occur continuously, which leads to large energy spreads. Moreover, the injection volume and initial transverse momentum spread scale with the intensity of the laser, which typically has to be large ($a_0 \gtrsim 2$) to facilitate trapping [99]. This can result in large transverse emittances [103] and an increased slice energy spread.

Hence, generating a high quality electron bunch from ionization injection requires a plasma source design that localizes the dopant gas and, ideally, facilitates trapping at low laser intensities. First experiments that could demonstrate lower energy spreads (5-10%) from a confinement of the dopant gas relied on two gas cells or jets that were placed immediately behind each other [152–154]. Recently, localization of the dopant gas in a discharge capillary has also been shown [155]. Instead of localizing the dopant gas, a reduction in energy spread has also been demonstrated from shock-assisted [156] and self-truncated [157] ionization injection. Finally, trapping of electrons ionized by a secondary laser of higher frequency than the drive laser promises very low emittances [100], but has yet to be realized experimentally.

3.1.2 Conceptual design

As part of this work, we developed a novel continuous-flow plasma source that integrates a dedicated injection and acceleration section in a single capillary-type design. It consists of a micro-machined channel structure that has been milled into a sapphire crystal. A schematic of the plasma source is shown in Fig. 3.1.

A first inlet continuously supplies the entrance of a 5-mm-long interaction channel with a mixture of hydrogen (H_2) and nitrogen (N_2). In this section, a high-quality electron bunch is produced from ionization injection and then accelerated to higher energies in a following section, which is supplied with pure hydrogen by a pair of inlets with equal mass flow. An integrated vacuum outlet separates both sections.

Operation at a balanced pressure ratio prevents the diffusion of nitrogen, which results in the formation of a short ($\sim 700\ \mu\text{m}$) transition region. Usage of a high dopant concentration (5-20 %) furthermore leads to the formation of an ionization-induced plasma density downramp that lowers the trapping threshold and maximizes the injection rate at low laser intensities. By varying the concentration, we can additionally influence the shape this downramp, which provides control over the injected bunch length and current profile.

In combination with the tunable laser parameters, the setup offers comprehensive control over the injection and acceleration dynamics. In the experiment described in Ch. 4, this allowed precisely optimizing beam loading, resulting in 1% level energy spreads.

3.1.3 Experimental realization

As is common for capillary-type plasma sources, sapphire was chosen as the manufacturing material because of its high melting point, thermal conductivity and electrical resistance [112]. While the high mechanical hardness of sapphire is an advantage, it also presents a challenge when it comes to machining the required sub-mm scale channel structures. In preceding work [112, 113], a reliable manufacturing process has been developed that allows complex designs to be realized within a reasonable time frame.

The process involves milling the 3D design into a pair of 3-mm-thick slabs of sapphire crystal ($10\times 10\ \text{mm}$), using a commercial CNC milling machine with diamond heads. The slabs are then sandwiched together to form a closed volume and mounted in a target holder. Once installed, the polished surfaces provide sufficient sealing to prevent the gas from leaking at the interface. The assembly is connected to the outside through a system of channels and pipes that are integrated into the in-vacuum alignment stage.

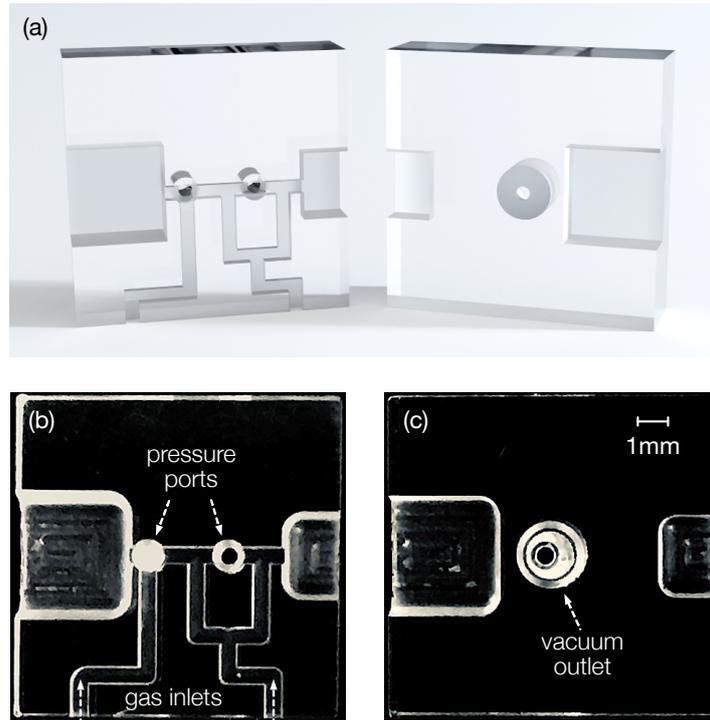


Figure 3.2 – (a) Rendering of the CAD model used for manufacturing. Image of the plasma source (b) back and (c) front plate. The 3D design is milled into a pair of 3-mm-thick slabs of sapphire crystal (10x10 mm). The slabs are sandwiched together to create the final structure. The square-shaped channels are 500 μm wide. Additional holes drilled perpendicular to the propagation axis serve as pressure ports and gas-separating vacuum outlet.

Figure 3.2 shows a rendering of the CAD model and images of the machined sapphire plates. The main structure consists of square-shaped channels with an edge length of 500 μm that have been milled into the back plate [Fig. 3.2(b)]. The 5-mm-long interaction channel is supplied via two inlet channels that are connected to the bottom of the target holder. Through the first inlet, the injection section is supplied with a pre-mixed blend of hydrogen and nitrogen. The second inlet is split symmetrically to provide an equal flow of hydrogen to the beginning and end of the acceleration section.

Additional holes of 500 μm diameter that have been drilled perpendicular to the propagation axis serve as pressure ports and gas-separating vacuum outlet. The outlet in the front plate [Fig. 3.2(c)] is positioned between the two left inlets. Towards the front, its diameter increases to 1.5 mm to enhance the release of gas into the vacuum. The pressure ports in the back plate connect the mixed gas inlet and the center of the pure hydrogen region to out-of-vacuum pressure gauges.

As the gas flows through the channels, the applied pressure drops, causing it to expand and accelerate towards the exits. In between the inlets, however, static pressure builds up, which translates to a density profile that is mostly constant along the main channel but decreases rapidly along the outlets. To ensure that the interaction length is in the range of the Rayleigh length, larger rectangular volumes have been added that connect the channel entrance and exit to the vacuum. The residual flow resistance within these volumes counteracts the rapid gas expansion. By choosing a smaller size for the exit volume, we could, to some degree, tailor the plasma-to-vacuum density transition such that it leads to an adiabatic reduction of the focusing forces. This can reduce the beam divergence and mitigate chromatic emittance growth [158, 159].

3.1.4 Gas supply and monitoring system

The supply of gas to each of the two inlets was controlled by a combination of motorized regulation valve (Pfeiffer Vacuum EVR 116) and mass flow meter (Bronkhorst F111-B) [112]. In addition, one channel could be bypassed and a commercial gas mixer (QCAL GMS_2CH) could be used instead [113]. The device blended the hydrogen gas with a variable concentration of nitrogen. It comprises two mass flow controllers (Sensirion SFC5400) that have been calibrated for a range of gas species. In the flow range used, the controllers had an accuracy $< 1\%$ and a repeatability $< 0.2\%$. In practice, the achievable flow stability was measured to be better than 0.2% rms [112]. At a mixing ratio of 95% H_2 and 5% N_2 , the fluctuations of the N_2 concentration were measured to be below 0.1% rms [113].

As discussed previously, operation with a continuous flow of gas leads to a non-evolving pressure and velocity distribution inside the plasma source. This circumstance not only benefits the stability of the laser-plasma interaction, it also allows us to monitor the plasma density by measuring the gas pressure at selected locations along the main channel.

More precisely, the pressure is measured with two capacitive pressure gauges (Pfeiffer CMR361) that are located outside the vacuum environment, but are connected directly to the channel via the access ports in the sapphire back plate [113]. The pressure gauges are absolutely calibrated and measure pressure independently of the gas species with a relative accuracy of 0.2% . Consequently, the gas number density n can be inferred from the ideal gas law, which relates it to the static pressure p according to [160]

$$p = nk_{\text{B}}T, \tag{3.1}$$

where k_{B} is the Boltzmann constant and T is the temperature. For diatomic molecules, such as hydrogen, the plasma density n_e is then simply given as twice the gas density n multiplied by the ionization level.

In the given arrangement, the sensors measure the static pressure that builds up in the pressure ports. Using an estimate for the initial gas temperature, Eq. (3.1) directly relates this value to the on-axis density in regions where the net flow is zero, which is the case in between the two inlets that define the acceleration section. In regions where the fluid is in motion, such as in the injection section, variations in temperature have to be considered when deriving the density from the measured pressure. Moreover, the measurement can suffer from systematic errors during high-power operation. This includes, for example, temperature differences due to heating of the sapphire from the laser, which can lead to deviations in the conditions measured at the sensors and the actual conditions in the plasma source.

3.2 Gas and plasma profile

During the design process and for particle-in-cell simulations of the laser-plasma interaction, it is critical to know the gas distribution, in particular the longitudinal gas density profile, that forms inside the plasma source. We therefore performed multi-species computational fluid dynamics (CFD) simulations of the plasma source that explicitly took into account the mixing of hydrogen and nitrogen.

To approximate the gas profile in the experiment as close as possible, these simulations were referenced to the online mass flow and pressure measurements. A direct experimental confirmation of the simulated gas density profile was not feasible, because the particularities of the design restricted the optical access to the interaction channel. We refer to previous work [112, 113], in which the applicability of this modeling approach was confirmed by comparison with interferometric measurements [161] and Raman spectroscopy [162] for similar plasma sources.

We did, however, experimentally verify the localization of the dopant gas, which is a crucial prerequisite for the injection mechanism. In remainder of this section, we present spectrally resolved measurements of the plasma recombination light after optical field ionization of the gas with the drive laser.

3.2.1 CFD simulations

The gas inside capillary-type plasma sources can be described as a compressible fluid exhibiting viscous, laminar and subsonic flow that transitions to more complex regimes at the outlets [112, 160]. In general, such fluid flow is modeled by numerically solving the discretized Navier–Stokes equations, which is referred to as computational fluid dynamics (CFD) simulations [163].

In the following, we present CFD simulation results that correspond to the optimized setpoint used during the experiment described in Ch. 4. 3D steady-state, compressible RANS (Reynolds-Averaged Navier-Stokes) simulations have been performed with the commercial software ANSYS FLUENT [164]. A turbulent flow model (realizable $k-\epsilon$) [165] was used, even though matching results were also obtained with a laminar flow model. The numerical solver was based on the SIMPLE algorithm with a second order upwind scheme [166]. Modeling of the non-reactive mixing of gas species was based on kinetic theory, using the Chapman-Enskog equation for the diffusion coefficient [167]. This required defining the characteristic length and energy parameter of the Lennard-Jones potential for both species, which were taken from [168].

Figure 3.3(a) shows the simulation domain that has been discretized as a polyhedral mesh with an adaptive resolution. Note that, when compared to Figure 3.2, this visualization shows the plasma source upside down.

The mesh comprises the inlet and interaction channels, the larger volumes at the entrance and exit, the two pressure ports, as well as the vacuum outlet in between the injection and acceleration section. By including the pressure ports, we could directly compare and match the experimentally measured pressures to the simulated values at the end face of these volumes.

The steady-state solution of the simulation was constrained by initial and boundary conditions. A constant mass flow was specified at the entrance of the inlets and the gas was assumed to have room temperature (293.15 K). The vacuum-facing domain walls were modeled as open boundary, assuming a pressure of 10^{-6} mbar outside of the domain and no back flow. The remaining walls fulfilled the conventional no-slip condition for viscous flow, which means that the fluid is assumed to have zero velocity exactly at the boundary.

In the experiment, the first inlet was supplied with 16.2×10^{-8} kg/s of hydrogen and 25.0×10^{-8} kg/s of nitrogen, corresponding to a combined normal flow rate of 120 ml_n/min (at 1 atm, 0° C) and an N₂ number concentration of 10%. The second inlet was supplied with 51.2×10^{-8} kg/s of hydrogen, which corresponds to a normal flow rate of 342 ml_n/min. The direct measurement of the static pressure yielded 19.9 mbar at the mixed gas inlet and 24.9 mbar in the center of the pure hydrogen region.

In order to converge the simulation to the experimentally measured pressures, the inlet flow rates were equally reduced by 27%. The deviation could be explained by losses in the gas supply chain. Following the adjustment, the simulated pressures agreed with the measurement to the few-percent level ($< 4\%$ error). The residual deviation was asymmetric with the simulated pressure being slightly higher in the injection section and slightly lower in the acceleration section. In the final profile, we compensated for this asymmetry such that the simulated pressure ratio would exactly match the experimental reference.

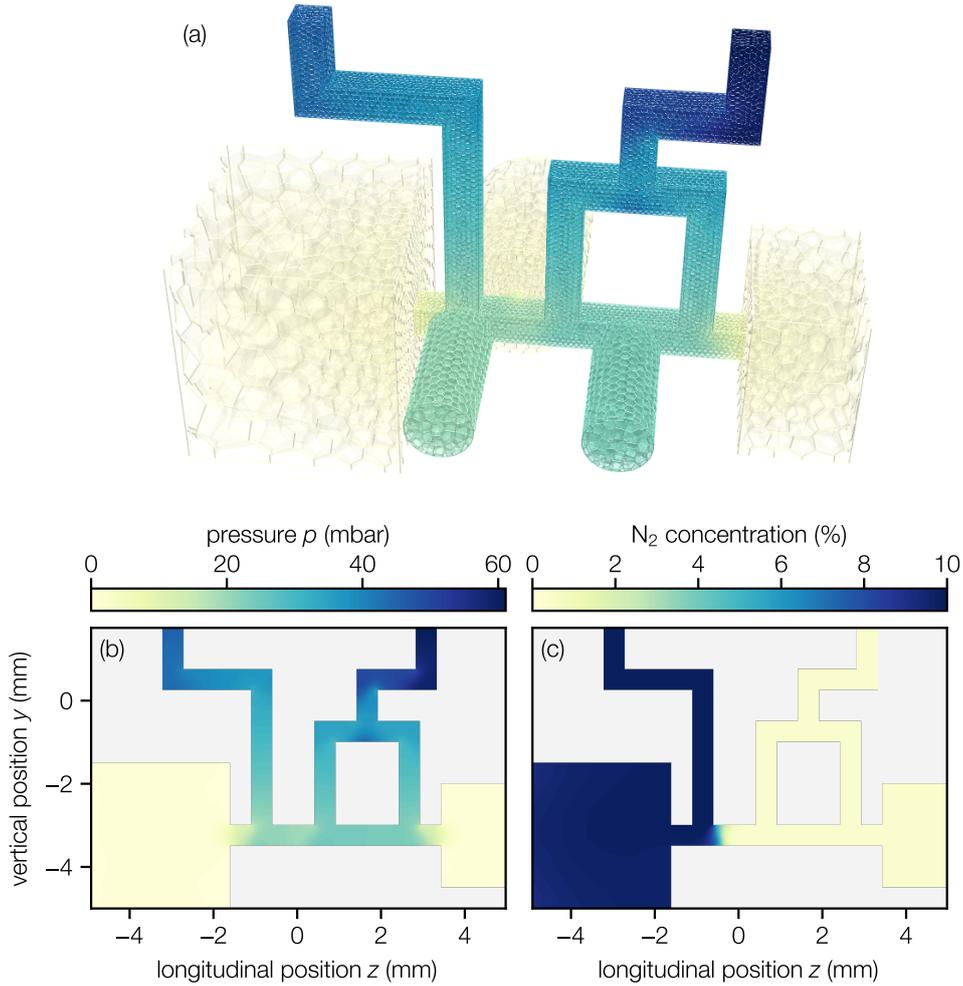


Figure 3.3 – Computational fluid dynamics (CFD) simulation of the plasma source using ANSYS Fluent. (a) 3D plot of the discrete simulation mesh showing the gas pressure. Transverse slices ($x = 0$ mm) showing (a) the pressure and (b) the nitrogen (N_2) concentration along the propagation axis. The initial concentration at the mixed gas (left) inlet was 10%.

Figure 3.3(b) and (c) presents the simulation results, showing the pressure and nitrogen concentration along the propagation axis for a transverse slice that is centered in the horizontal direction ($x = 0$ mm). In the central region of the interaction channel, a plateau-like pressure profile develops that drops towards the outlets. At the same time, the differential pressures between the inlets are balanced in such a way that the resulting velocity gradients minimize the diffusion of nitrogen molecules into the acceleration section. As a consequence, the mixed gas is strictly confined to the injection section and the initial 10% N_2 concentration falls off to zero over a short distance of only ~ 700 μm . From these results, we then extracted the number density

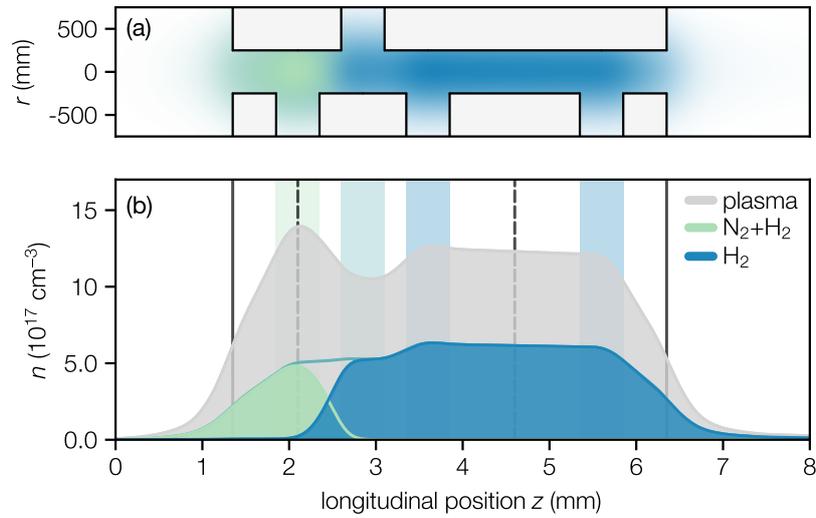


Figure 3.4 – Longitudinal gas and plasma density profile that has been derived from the CFD simulation. (a) Simplified representation of the channel layout. (b) Mixed gas (green), hydrogen (blue) and plasma electron (grey) density n . Location of the inlets and vacuum outlet (colored bars), pressure ports (dashed lines) and interaction channel entrance and exit (solid lines).

of each species and used it as a direct input to particle-in-cell simulations. For this, we only consider the on-axis profile and thereby neglect the small transverse asymmetries that are present in the three-dimensional gas distribution.

Figure 3.4 shows the longitudinal gas density profile, as well as the plasma density profile that is produced once the laser has ionized the lower energy levels. Since the nitrogen atoms yield five times more electrons than the hydrogen atoms, a plasma density peak forms at the beginning of the profile. Owing to the variations in wake phase velocity, the trapping of electrons is enhanced along the density downramp of this peak, whereas the injection is suppressed along the adjacent upramps. As described at the beginning of this chapter, this effect is desirable because it further localizes the injection and can minimize the initial momentum spread. Acceleration to higher energies then mainly occurs throughout the flat plateau region before the plasma transitions to vacuum along the exit downramp. At the very end of the plasma source, the profile asymptotically decreases towards zero density, which is a consequence of the small exit volume. As mentioned earlier, the beam will experience residual focusing forces in this region, which can help to preserve the transverse beam quality.

3.2.2 Dopant gas localization

In this final section, we experimentally demonstrate the localization of nitrogen within the injection section of the plasma source. We recorded spectrally resolved images of the plasma recombination light after optical field ionization of the gas with the drive laser at moderate intensities. As the plasma emission is much shorter (ns time scales) than the time scale of the fluid motion, the measurement effectively reveals the neutral gas distribution at the laser arrival time [169, 152, 155]. However, for the measurement to work, we had to use an adapted design that omitted the vacuum outlet in the front plate, as it would otherwise impair the optical access from the side. Therefore, the results represent a qualitative verification of the mechanism only.

The plasma emission was observed perpendicular to the propagation axis through a diagnostic window in the target vacuum chamber. A 1:1 optical imaging system relayed a 10-mm-wide side view of the plasma source onto the slit entrance of a commercial Czerny-Turner spectrometer (PI Acton Series SP-2150i). The imaging system comprised four identical achromatic lens doublets (50.8 mm and $f = 150$ mm) in a double $4f$ configuration. Inside the spectrometer, the incident signal was dispersed by a reflective grating (1200 grooves/mm with a 500 nm blaze) and imaged onto a camera (Basler acA1920-40gm) that was mounted at the exit plane.

The gas was ionized by the drive laser, which was attenuated to an energy of ~ 25 mJ and compressed to a pulse duration of 41 fs FWHM. To improve the transverse and longitudinal uniformity of the ionized plasma column, the beam was radially cropped to a 40 mm diameter by a motorized iris before the compressor. The operating conditions of the plasma source were similar to those during the electron experiment. The injection section was supplied with a 90:10 mixture of hydrogen and nitrogen at a flow rate of 100 ml_n/min. The acceleration section was filled with hydrogen at a variable flow rate.

For the measurement, the spectrometer was tuned to a strong persistent Nitrogen emission line at 567.9 nm [170]. Figure 3.5 shows the spectrally resolved plasma emission signal, which is directly related to the longitudinal distribution of nitrogen molecules along the interaction channel. Note that the measurement becomes unreliable at the beginning and end of the interaction channel, as the machined surfaces distort and reduce the transmitted light.

In the first image of the series, only the left inlet was supplied and the mixed gas spread along the entire interaction channel. However, when the hydrogen flow into the acceleration section was increased, the mixed gas was pushed back, which resulted in a displacement of the dopant species towards the injection section. With increasing flow difference, the pressure imbalance counteracted the diffusion of nitrogen molecules causing the mixing interface to become steeper. Eventually, the region of nitrogen was

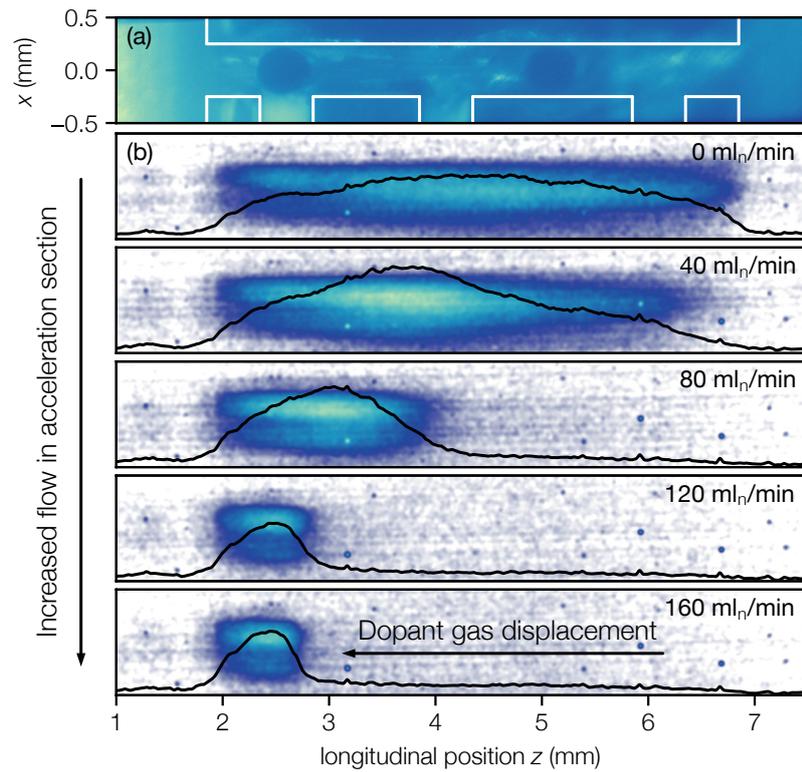


Figure 3.5 – Measurement of the localization of Nitrogen in the injection section. (a) Image of the plasma source overlaid with the channel layout. (b) Spectrally resolved images of the plasma emission after optical ionization. Signal in the vicinity of 567.9 nm (strong persistent Nitrogen line) and projections (black solid line). The left inlet was supplied with a mixture of 90% H₂ and 10% N₂ at a flow rate of 100 ml_n/min. The right inlets were filled with 90% H₂ at a combined flow rate of up to 160 ml_n/min (increasing from top to bottom).

compressed to sub-mm length. In summary, the measurement clearly demonstrates the possibility to localize the dopant gas at the beginning of the plasma source.

4 Demonstration of optimal beam loading

This chapter presents an experimental study for the demonstration of optimal beam loading in a laser-plasma accelerator, which was published in [37]. The following sections are mostly reproductions of [37], though some have been expanded to include additional details. In Sec. 4.1, we provide an overview of the experiment and place it in context with previous work. Based on simulations, Sec. 4.2 describes the injection and acceleration dynamics, followed by an analysis of the measured beam loading physics in Sec. 4.3. Moreover, in Sec. 4.4, we study the influence of drive laser instabilities. Finally, in Sec. 4.5, we summarize the results and provide an outlook for future work.

Co-author contributions: S. Jalas, P. Messner and P. Winkler helped performing the experiment. M. Schnepf, T. Eichner and T. Hülsenbusch maintained and operated the laser system. P. Winkler and L. Hübner developed the emittance measurement method. S. Jalas and L. Jeppe contributed to the particle-in-cell simulation study.

4.1 Overview

Laser-plasma accelerators produce high-brightness electron bunches of few-femtosecond duration with kiloampere peak currents [9, 10] and sub-micron emittances [11, 12], which makes them promising candidates to drive a compact free-electron laser [25, 26] or future linear collider [27]. However, delivering the small energy spreads required by applications remains a challenge, in particular since the steep gradients of the fields—which are inherent to the small size of the accelerating structure—can imprint large correlated momentum spreads onto the accelerated bunch.

At the same time, to operate LPAs efficiently [106], the accelerated electrons have to absorb a significant fraction of the energy stored in the plasma wave, which reduces the accelerating field along the bunch through an effect known as beam loading [28–31]. By shaping the bunch current profile, beam loading can be used to locally flatten the longitudinal wakefield. To effectively prevent the accumulation of correlated energy spread, however, it is important to further consider the full dynamic evolution of the plasma wakefield and phase of the injected bunch. Carefully optimizing those dynamics can result in optimal beam loading conditions that enable both, efficient acceleration and the preservation of a small initial energy spread.

Previous experiments first observed and characterized effects from beam loading in a laser-plasma accelerator [32, 171], studied the limit of fully loaded wakefields [172], presented first evidence of loading optimum total charges [33] and investigated the transition from beam-loaded to beam-dominated laser wakefields [34]. Albeit producing high charge beams (~ 300 pC) by operating in regimes of stronger blowout, previous work [33, 34] did not yet fully exploit the mechanism that leads to flat accelerating fields and consequently reported large energy spreads ($\sim 15\%$).

Significant flattening of the accelerating fields ($< 3\%$) and high efficiency ($\sim 40\%$) has been demonstrated in a beam-driven plasma accelerator, by pre-shaping and matching the current profile of an externally injected electron bunch to the wakefield [173]. Achieving a similar level of control over beam loading using internal injection in a laser-plasma accelerator can be considered more difficult, because the process of bunch generation is less flexible and intrinsically coupled to the acceleration.

Broadly speaking, to achieve optimal beam loading in this case, the following three steps must be performed:

- (1) Populate a small phase space volume during injection such that the initial slice energy spread and transverse emittance is small.
- (2) Control the longitudinal distribution of charge during the trapping process so as to inject a current profile that cancels the laser-driven wakefield.
- (3) Balance the wakefield evolution throughout the subsequent acceleration process such that each particle along the bunch experiences the same net acceleration.

Note that the latter two steps are coupled as the ideal bunch current profile and phase is specific to a given wakefield amplitude and shape. Since the wakefield can dynamically change, an average optimum has to be found. However, this can also be seen as an additional degree of freedom during optimization, which can be helpful when control over the internally injected bunch current profile is limited.

In the following, we describe the experimental realization of this strategy at the LUX laser-plasma accelerator. Electron bunches with a small initial slice energy spread were generated from localized ionization injection [19, 150, 152] and accelerated to hundreds of MeV using the plasma source described in Ch. 3. By tuning key laser and plasma parameters, we could optimize the combination of injected bunch current profile and wakefield evolution. This led to an on-average flattening of the accelerating field experienced by the electrons to the $\sim 1\%$ level.

As a result, electron bunches with exceptional quality have been produced. Figure 4.1 shows an example of the best measured transverse beam profiles and energy spectra. When fulfilling optimal beam loading conditions, electron bunches with ~ 44 pC of charge were accelerated to ~ 282 MeV while minimizing the accumulated energy spread to the

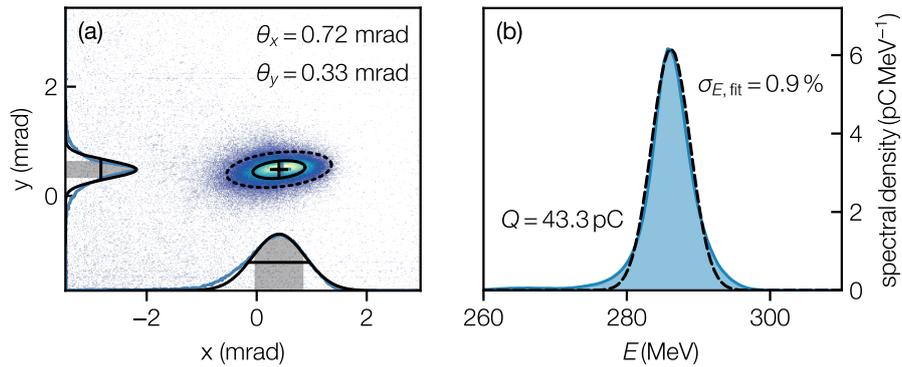


Figure 4.1 – Example of the achievable beam quality after optimizing the laser-plasma accelerator to operate at optimal beam loading conditions. (a) Transverse beam profile of an unfocused electron beam with 54 pC of charge showing the rms divergence out of the plasma source. Projections of the raw signal (blue) and Gaussian fits (black). (b) Energy spectrum (blue) and Gaussian fit (black) of a beam with sub-percent energy spread.

1% level. For these beams, the spectral charge density was on the order of 5 pC MeV^{-1} and the divergence out of the plasma source was in the sub-mrad range.

The laser and plasma source properties at this optimized set-point were the same as described in Sec. 2.2 and Sec. 3.2, respectively. The on-target energy was 2.6 J and the pulse duration was 34 fs FWHM, corresponding to a peak power of 69 TW. The focused spot size was $25 \mu\text{m}$ FWHM, resulting in a peak normalized vector potential of $a_0 = 2.1$. The gas density derived from the measured pressures was around $6 \times 10^{17} \text{ cm}^{-3}$ and the nitrogen concentration in the injection region was 10%.

In the following sections, this set-point is studied in detail in terms of quality and stability using both simulations and experimental data. In Sec. 4.2, we describe the injection and acceleration dynamics based on particle-in-cell simulations and explain the influence of laser and plasma parameters on the longitudinal phase space.

In Sec. 4.3, we present high-statistic measurements of the beam loading physics. An analysis of shot-to-shot variations reveals a shearing of the longitudinal phase space for varying loads. We also characterize the average beam quality at the optimum.

Together with the electron beam properties, we acquired shot-synchronized records of the most important laser properties that affect the injection and acceleration process. In section Sec. 4.4, we correlate this data and apply machine learning techniques to build a surrogate model that predicts the beam quality as a function of the drive laser.

4.2 Injection and acceleration dynamics

4.2.1 Particle-in-cell simulation setup

The laser-plasma interaction was modeled with particle-in-cell simulations using the spectral, quasi-cylindrical code FBPIC [50] that is described in detail in second part of this work (see Ch. 6). The first two modes of the azimuthal decomposition were included. The longitudinal and transverse grid resolution was $\Delta z^{-1} = 50 \mu\text{m}^{-1}$ and $\Delta r^{-1} = 3 \mu\text{m}^{-1}$, respectively. We used 8 particles per cell (64 for the nitrogen ions) and cubic particle shapes. The time step was set to $c\Delta t = \Delta z$.

To reduce the computational costs, the simulations were performed in a Lorentz-boosted frame ($\gamma_{\text{boost}} = 5$) [49]. Optical field ionization of the gas atoms was based on the Ammosov-Delone-Krainov (ADK) model [99, 102]. Except for the inner two levels of nitrogen, the atoms were assumed to be pre-ionized. As described in Sec. 2.2, the measured radial laser intensity evolution was approximated by a flattened Gaussian laser model ($N = 100$) [124].

Defining accurate and unbiased simulation input parameters that reproduce the experiment is a complex task. Systematic measurement errors are unavoidable as the relevant properties are difficult to measure, especially in absolute numbers. Additional errors can arise when making assumptions in the simulation that only approximate the experimental conditions. Despite agreeing on a qualitative level, these errors resulted in a quantitative difference between the particle-in-cell simulations and the experimental measurement. As an unbiased strategy to get quantitative agreement between simulation and experiment, we used a Bayesian optimization algorithm [174, 56] that converged the simulated beam loading dynamics to the measurement.

The optimizer adjusted the laser energy, pulse duration, spot size, longitudinal focus position and the gas density. For each set of input parameters, simulations for 5 focus positions were performed. The energy spread, beam charge, and energy from such a scan were compared to the actual correlations measured in the experiment. The optimizer then varied the input parameters to reduce the difference between simulation and experiment. It converged after ~ 300 iterations, which corresponds to a total of ~ 1500 particle-in-cell simulations.

Consequently, the matched simulations had a 4% smaller laser energy, 5% larger spot size, 2% shorter pulse duration, 22% lower gas density and a focus shift of -0.4 mm compared to the estimated values in the experiment. These differences are within the error of the respective measurements.

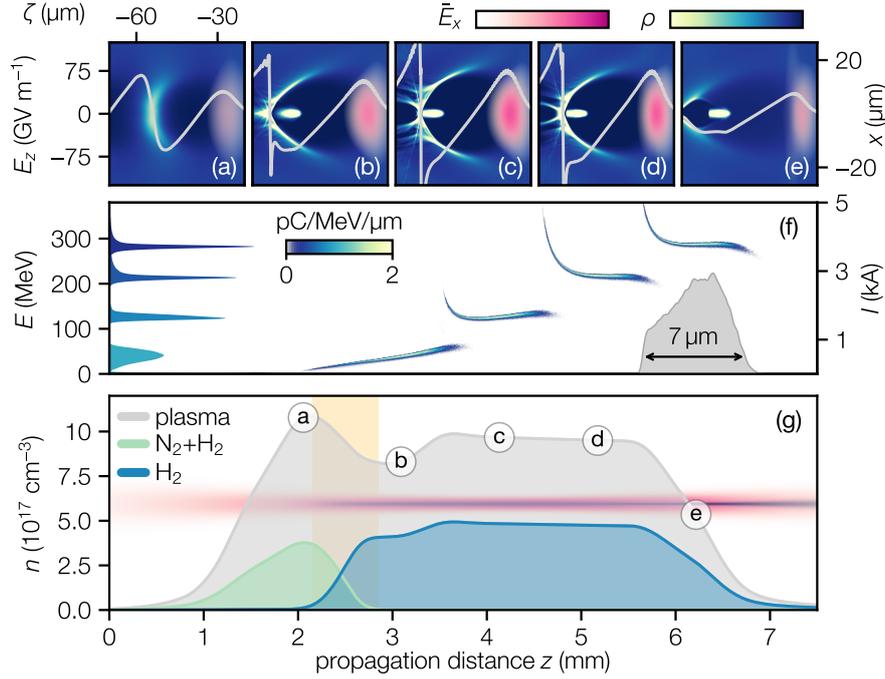


Figure 4.2 – Particle-in-cell simulation of the injection and acceleration process. (a)-(e) 2D snapshots ($\zeta = z - ct, x, y = 0$) of the laser, plasma wave and electron bunch at different propagation distances z , showing the charge density ρ , laser field envelope \vec{E}_x and on-axis longitudinal wakefield E_z . (f) Evolution of the longitudinal phase space (ξ, E) corresponding to (b)-(e) and projections to the energy axis; Current profile at position (e). (g) Molecular gas (10% N₂ + 90% H₂; 100% H₂) and resulting plasma electron density profile; z positions of (a)-(e) (markers); injection region (yellow); laser (red) and electron beam (blue) waist evolution.

4.2.2 Electron beam generation at the optimum

Figure 4.2 shows a simulation of the injection and acceleration process that corresponds to the optimized set-point in the experiment. The laser pulse is initially focused at $z_{\text{foc}} = 4.75$ mm. As it propagates through the plasma source, the low-intensity pedestal pre-ionizes the weakly bound energy levels of the gas atoms (H⁺, N¹⁻⁵⁺), forming a background plasma of density $n_e \simeq 9.6 \times 10^{17} \text{ cm}^{-3}$. As described previously in Ch. 3, the higher electron yield from the lower nitrogen levels leads to the formation of a plasma density peak at the beginning of the profile [Fig. 4.2(g)]. The inner K -shell electrons of nitrogen (N^{6,7+}), however, are tunnel-ionized only in regions of high laser intensity at later wake phases and therefore can get trapped [99].

At the beginning of the interaction, the laser intensity is still low and the contraction of the plasma wavelength along the initial upramp suppresses the trapping of electrons

[Fig. 4.2(a)]. As the pulse then reaches higher intensities, the plasma wave transitions to the nonlinear blowout regime [85] and localized injection of electrons occurs over a short distance of $\sim 700 \mu\text{m}$ along the density downramp, which lowers the trapping threshold by decreasing the wake phase velocity [94, 20]. Throughout the trapping process, the peak normalized vector potential stays below $a_0 = 2.1$ and close to the ionization threshold of the higher nitrogen levels, which minimizes the initial momentum spread of the electrons.

In combination with the decrease in nitrogen doping, the injection mechanism leads to the formation of a bunch with a ramped current profile that decreases from head to tail. The length of this profile is mainly determined by the difference in plasma wavelength at the beginning and end of the downramp [Fig. 4.2(b)]. Consequently, the initial phase of the head of the bunch roughly coincides with the position of the crest of the wave at the density peak [Fig. 4.2(a)]. Eventually, the tail of the current profile gets truncated in the second density upramp and an isolated bunch with 49 pC of charge distributed along a $\Delta\xi \simeq 7 \mu\text{m}$ (23 fs) long quasi-trapezoidal current profile (3 kA peak current) is subsequently accelerated to an energy of 288 MeV throughout the 4-mm-long pure hydrogen section [Fig. 4.2(c)].

Immediately upon injection, the longitudinal phase space (ξ, E) is positively correlated as a result of the retarded trapping of electrons and beam loading along the current profile cancels the slope of the laser-driven wakefield [Fig. 4.2(b)]. Since the laser exceeds the critical power for relativistic self-focusing ($P/P_c \simeq 2$) [4] a maximum intensity ($a_0 \simeq 3$) is thereafter reached in the center of the density plateau where the stronger blowout reverses the initial positive energy chirp [Fig. 4.2(c)]. Diffraction of the laser then again leads to a decrease of the accelerating field until the beam-induced wakefield dominates at the end of the plasma [Fig. 4.2(d)-(e)]. Although the beam loading varies continuously, the averaged dynamics result in a flattened core of the longitudinal phase space at the exit of the plasma source, with a projected relative energy spread of $\sim 1\%$.

4.2.3 Influence of laser and plasma parameters

As emphasized at the beginning of this chapter, beam loading has to be viewed as a dynamic and coupled process. In the experiment, we could optimize this process by controlling the laser and plasma source parameters that mainly influenced the injected bunch current profile and the wakefield evolution. More specifically, we primarily tuned the overall gas density n , the nitrogen concentration C_{N_2} , the laser energy E_{laser} and the focus position z_{foc} .

Figure 4.3 illustrates the effect of these parameters by means of their influence on the longitudinal phase space (E, ξ) and charge Q at the end of the acceleration process.

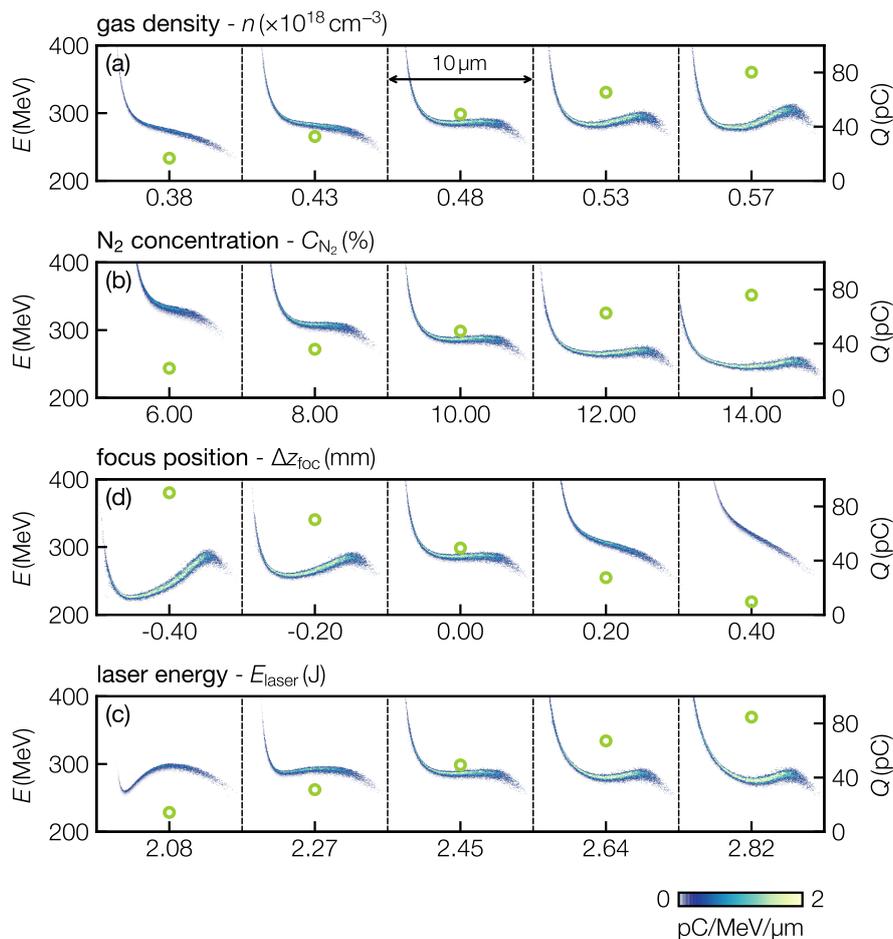


Figure 4.3 – Influence of the laser and plasma source parameters used to optimize beam loading. Longitudinal phase (E, ξ) at the end of the acceleration process and bunch charge Q (green dots) from a scan of simulations, varying the (a) gas density n , (b) Nitrogen concentration C_{N_2} , (c) longitudinal focus position z_{foc} and (d) laser energy E_{laser} . The central column corresponds to the optimized set-point.

Each row corresponds to a simulation scan of a single parameter around the optimum (central column).

The overall gas density n [Fig. 4.3(a)] scales the nominal accelerating gradient as well as the number of electrons that get trapped. Owing to relativistic self-focusing, a larger density also leads to higher laser intensities throughout the interaction. This further increases the wakefield amplitude and injected charge. Consequently, the energy at the head of the bunch increases, whereas the energy at the tail remains almost constant due to the higher load.

The dopant gas concentration C_{N_2} [Fig. 4.3(b)] scales the number of electrons that get released per unit volume in the injection section. As a result, the bunch current increases with the fraction of nitrogen molecules. At the same time, the peak of the electron density profile gets elevated. This in turn increases the bunch length, since it depends on the difference in plasma wavelength at the beginning and end of the trapping process. In combination, both effects increase the total charge while lowering the central beam energy.

The longitudinal position of the focus z_{foc} [Fig. 4.3(c)] has a direct influence on the laser intensity at the point of injection and, in consequence, the size of the injection volume. Shifting the focus varies the injected charge without significantly changing the wakefield evolution. Over- or underloading of the plasma wave then imprints an almost linear positive or negative chirp onto the accelerated phase space.

For the same reasons as with the focus position, the injected charge increases with laser energy E_{laser} [Fig. 4.3(d)]. However, a higher laser energy also drives a stronger plasma wave, which partially compensates a decrease in electron energy due to the higher load. Moreover, variations in laser energy affect the nonlinearity of the wakefield which in turn influences the curvature of the longitudinal phase space.

In the experiment, we would balance the contribution of each parameter to the beam loading dynamics until we minimized the measured energy spread while at the same time maintaining a high bunch charge. Indirectly, this also led to a high beam loading efficiency since the injected current profile then covered a significant fraction of the plasma wave.

4.3 Beam loading physics

When fulfilling the ideal conditions as described above, the average accelerating fields were flattened to the 1% level and the accelerator delivered beams with an exceptional quality, as exemplified in Figure 4.1. However, drive laser fluctuations caused the injection and acceleration process to deviate from this optimum from shot to shot, primarily via the amount of injected charge. By analyzing these variations, we can infer the underlying phase space dynamics directly from the measured data.

For this analysis we recorded a series of 5207 consecutive shots. Events where no beam was generated or a laser or electron diagnostic was missing were excluded (<3% of all shots). The electron bunches were unfocused and had an rms beam divergence of (0.8 ± 0.1) mrad and (0.6 ± 0.2) mrad in x and y , respectively. The pointing jitter was 0.5 and 0.3 mrad rms. Note that these values were inferred from a subsequent measurement (~ 1000 consecutive shots) of the transverse beam profiles.

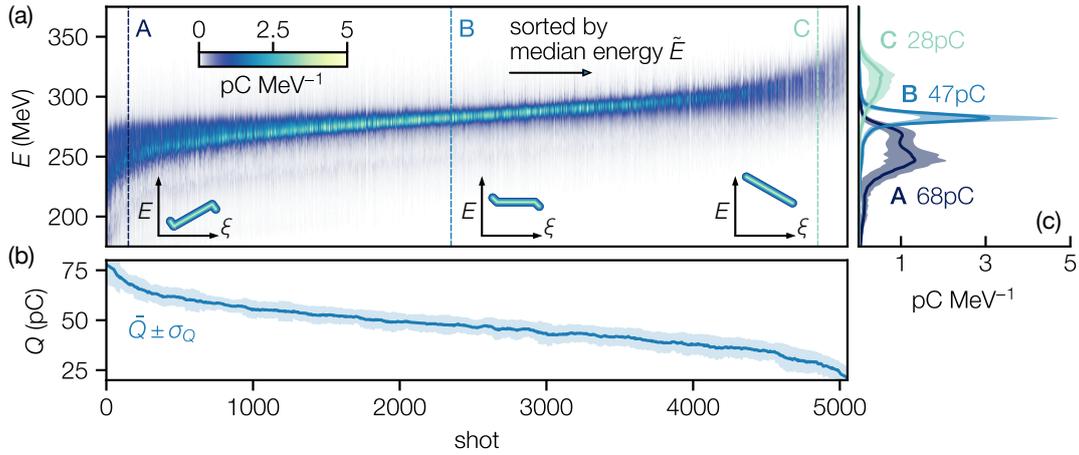


Figure 4.4 – Experimental results showing the phase space shearing and optimal beam loading. (a) Measured series of energy spectra sorted by median energy \tilde{E} ; (b) 100-shot moving average \bar{Q} (solid line) and standard deviation $\pm\sigma_Q$ (shaded area) of the corresponding sorted charges Q . (c) 100-shot average spectrum (solid lines) and standard deviation (shaded area) in regions of strong (A), optimal (B) and weak (C) beam loading. Locations of (A)-(C) and a schematic interpretation of the corresponding longitudinal phase spaces (ξ, E) are shown in (a).

4.3.1 Phase space shearing and optimal beam loading

Figure 4.4 shows the measured (a) energy spectra and (b) bunch charges Q sorted by the median bunch energy \tilde{E} , as well as the (c) characteristic spectrum in regions of (A) strong, (B) optimal and (C) weak beam loading. The spectra and charges shown in Fig. 4.4(c) were obtained from averaging over ± 50 shots at the respective locations in the sorted dataset.

In this sorted representation, the beam loading physics become particularly evident. At higher charges (A), overly strong beam loading locally reversed the slope of the wakefield and imprinted a positive chirp onto the longitudinal phase space (ξ, E) . The characteristic energy spectrum was almost flat-top and slightly peaked at the edges. This can be explained by the shape of the overloaded wakefield, which caused the phase space to turn over at the head and tail of the bunch.

With decreasing charge, the decelerating beam loading effects got weaker, and effectively sheared the tail of the bunch towards higher energies. The highest spectral density and lowest energy spread was reached when the core temporal slices overlapped in energy at the optimal load (B). In this case, the extraction of energy along the injected current profile exactly compensated the gradient of the average accelerating field [29, 30]. The resulting spectrum resembled a Gaussian distribution with slightly wider tails, in part

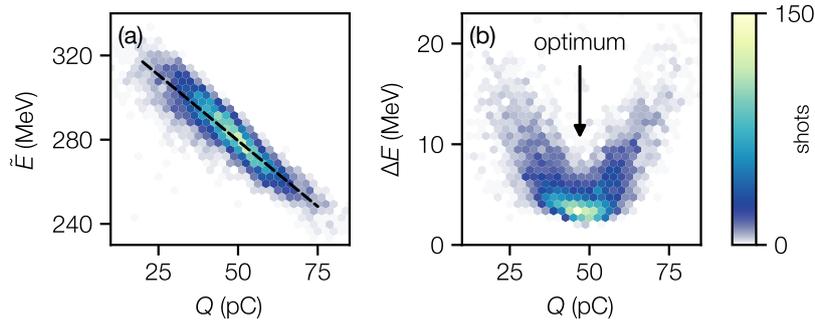


Figure 4.5 – Effect of the variations in beam loading on the electron beam properties. (a) Correlation of bunch charge Q with median energy \tilde{E} and fitted linear decrease of energy as a function of charge (dashed line). (b) Correlation of bunch charge Q with energy spread (median absolute deviation) ΔE . The phase space is flattened at ~ 47 pC. The data is the same as in Fig. 4.4.

because the beams were unfocused, but also because the wakefield was not perfectly flattened at the head and tail of the bunch, as indicated by simulations.

For lower charges (C), effects from beam loading diminished causing the energy to monotonically increase from head to tail, as the electrons then mainly experienced the undisturbed laser-driven wakefield. As expected, the corresponding spectrum was again broad and had the lowest peak spectral density, due to the combination of low charge and large energy spread.

4.3.2 Beam loading correlations

The same physics is reflected in the direct correlations of energy and energy spread with charge, Fig. 4.5. Within the range covered by the measurement, an increase in charge Q linearly decreased the median energy,

$$\tilde{E} = (342 - 1.25 \times Q[\text{pC}]) \text{ MeV}, \quad (4.1)$$

and resulted in a distinct V-shaped distribution of the energy spread ΔE around a minimum at ~ 47 pC. Note that we use the median absolute deviation as a robust statistical measure of the energy spread to account for the differently shaped spectra.

Over the full range of loads measured, we observed a ~ 7 -fold reduction of the projected energy spread down to $\sim 1\%$. Such a large variation might create the impression that the regime of operation is particularly unstable. However, it simply highlights the level of precision required to flatten the accelerating fields over a significant fraction of the plasma wavelength. If the beam loaded wakefield would have residual curvature, if the slice energy spread would be larger, or if the bunch would be much shorter, the clear V-shaped correlation would diminish.

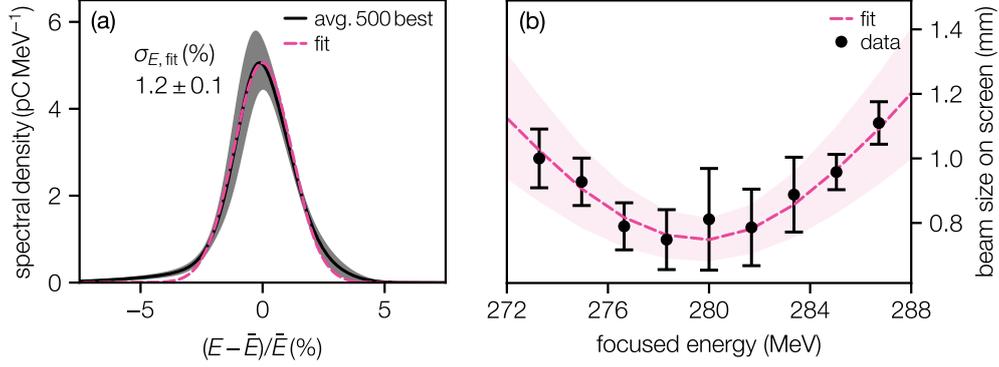


Figure 4.6 – Characterization of the electron beam quality at the optimal load. (a) Relative energy spectrum at optimal beam loading. 500-shot average (black solid line) and standard deviation (shaded area) of beams with lowest energy spread $\sigma_{E,\text{fit}}$; Gaussian fit (magenta dashed line). (b) Results of the energy-resolved quadrupole scan. Reconstructed beam size on the spectrometer screen for different focused energies around a central energy of 280 MeV. Measured data (black dots) and fit (magenta dashed line).

Moreover, we can use the measured correlations to estimate the bunch length in a similar fashion as was done in [171]. In the limit of vanishing charge, the bunch effectively samples the undisturbed accelerating field and acquires an approximately linear chirp. In this case, we can relate the median energy to the average accelerating field at the center of the bunch. Consequently, the relative difference between the accelerating field at the head and tail of the bunch ($\Delta\Xi$) approximately corresponds to twice the relative difference in median energy between the optimally loaded (\tilde{E}_{opt}) and unloaded (\tilde{E}_0) case,

$$\Delta\tilde{E} = (\tilde{E}_0 - \tilde{E}_{\text{opt}}), \quad (4.2a)$$

$$\Delta\Xi = 2 \times \Delta\tilde{E}/(\tilde{E}_0 + \Delta\tilde{E}). \quad (4.2b)$$

Using Equation (4.1), we can extrapolate the median energy in the unloaded case to be $\tilde{E}_0 \simeq 342$ MeV and, in consequence, $\Delta\tilde{E} \simeq 60$ MeV. Considering a bunch position close to the end of the first plasma wave bucket, this means that the bunch extended over approximately one third ($\Delta\Xi \simeq 0.3$) of the accelerating field, which is in good agreement with the simulation [compare Fig. 4.2(f)].

4.3.3 Characterization of the optimum

To fully characterize the accelerator performance at the optimal load, we repeated the measurement with the bunches focused onto the spectrometer screen. This increased the energy resolution of the spectrometer to $\sim 0.1\%$.

Figure 4.6(a) presents the average beam quality of the 500 shots with lowest energy spread, i.e., about 10% of this dataset (5130 consecutive shots), which we consider representative for operation at optimal beam loading conditions. On average, these bunches had an energy of (282 ± 5.3) MeV with (44.3 ± 4) pC of charge. The relative energy spread was (1.2 ± 0.1) % rms (Gaussian fit) or (7.1 ± 0.8) MeV FWHM. The peak spectral density was (5.3 ± 0.7) pC MeV⁻¹.

In addition, an energy-resolved quadrupole scan [12] was performed to measure the transverse phase space properties in the non-dispersive (laser polarization) plane. The measurement procedure and analysis was developed in [114]. At each step, we varied the focusing strength to image a different energy onto the spectrometer screen and recorded ~ 300 shots. For the analysis, only beams with a small energy spread were considered. From the remaining shots, we selected all beams within a given energy interval to reconstruct the dependence of the beam size on the focused energy. As an example, Fig. 4.6(b) shows the measured beam size on the screen around a central energy of 280 MeV.

From a fit to the analyzed data, we could then infer the transverse properties of the beam at its virtual source position in the target [114]. Consistent with the injection mechanism and in agreement with simulations, the normalized emittance at 282 MeV was (1.9 ± 0.3) mm mrad with a reconstructed rms source size and beam divergence of (4.0 ± 0.1) μm and (0.9 ± 0.1) mrad, respectively.

The beam loading efficiency at the optimal load was estimated from the simulation. We derived a wake-to-beam energy-transfer efficiency of $\sim 19\%$ by relating the energy loss of the laser (48 mJ) to the energy gain of the electrons (9 mJ) along the 2-mm-long plateau region [compare Fig. 4.2(g)]. Higher beam loading efficiencies would require injecting more charge, by using longer bunches or by operating in a regime of stronger blowout.

4.4 Influence of drive laser instabilities

In the following, we again use the dataset of unfocused electron beams to study the origin of the shot-to-shot beam loading variations that have been observed in the experiment. For that, we combine the electron data with the synchronized data from the high-power laser diagnostics presented in Ch. 2.

4.4.1 Laser-electron correlations

In a first step, we studied the pairwise correlations between drive laser and electron beam parameters. The analysis revealed that the observed variations in charge, which lead

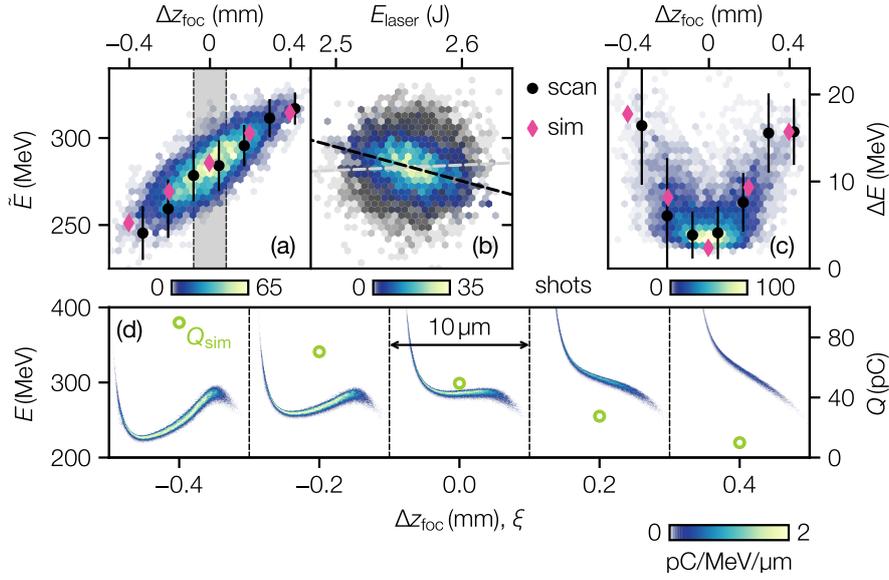


Figure 4.7 – Influence of the drive laser. Correlation of median electron energy \tilde{E} with (a) laser focus position Δz_{foc} and (b) pulse energy E_{laser} with (colored; black dashed line) and without (grayed out) filtering the data by $\pm 0.5\sigma_{z_{\text{foc}}}$ [gray area in (a)]. (c) Correlation of energy spread ΔE with Δz_{foc} . Overlaid are results from a longitudinal target position scan (black dots; 100-shot average) and particle-in-cell simulations (magenta diamonds). (d) Corresponding longitudinal phase spaces (ξ, E) and charges Q_{sim} (green circles) from the simulations.

to the beam loading variations, are directly related to fluctuations of the laser intensity at the point of injection. To leading order, small variations of the wavefront defocus of only 6.9 nm rms translated to a longitudinal focus position jitter $\sigma_{z_{\text{foc}}} = 162 \mu\text{m}$ rms, which scaled the intensity and thus the ionization and trapping rate along the density downramp.

In Fig. 4.7(a) and (c), we plot the median energy \tilde{E} and energy spread ΔE as a function of the relative focus position Δz_{foc} . The results look almost identical to the direct correlations of the beam energy and energy spread with charge, Fig. 4.5. More or less charge was injected as the focus position fluctuated, with only minor effect on the overall laser evolution. As discussed in Sec. 4.3, this shifted the beam loading conditions away from the optimum and, in consequence, varied the energy while increasing the energy spread.

The mechanism was verified by a dedicated scan of the longitudinal target position (black dots) and reproduced in the particle-in-cell simulations (magenta diamonds). Figure 4.7(d) visualizes how a downstream (upstream) shift of the focus position decreases (increases) the injected charge and thereby shears the final longitudinal phase space to higher (lower) energies.

Of course, other laser properties also vary the sensitive injection and acceleration process. However, because they have less influence or are more stable, their effect can remain hidden. For example, a subtle correlation between laser pulse energy E_{laser} and electron energy \tilde{E} [Fig. 4.7(b)] becomes visible only after filtering the data for shots within $\bar{z}_{\text{foc}} \pm 0.5\sigma_{z_{\text{foc}}}$. An otherwise expected weak positive correlation of electron energy with laser energy due to an increase in wakefield strength is overcompensated by stronger beam loading due to an increase of injected charge.

4.4.2 Machine-learning-based surrogate model

Ideally, we would like to study the effect of each laser parameter onto the electron beam quality. However, this is a non-trivial task, since it requires disentangling the interplay of hidden and nonlinear dependencies. For example, uncovering the subtle correlation in Fig. 4.7(b) requires either extensive filtering or a more advanced multivariate statistical analysis. Moreover, nonlinear dependencies, such as the correlation between focus position and energy spread in Fig. 4.7(b), cannot be represented by simple linear models. Consequently, this motivated us to apply advanced machine learning techniques [175, 176] to build a multivariate, nonlinear regression model of the experiment.

Using all of the available diagnostic data, we trained an artificial neural network (ANN) to predict, for each single shot, the beam energy, charge and energy spread from 15 different laser parameters: the laser pulse energy, central wavelength and spectral bandwidth, position and pointing into the target, longitudinal focus position, and higher order wavefront aberrations (astigmatism, coma, 3rd order spherical, trefoil).

The neural network was a simple multilayer perceptron (MLP) that was implemented with Python, using the Keras [177] package and additional features from scikit-learn [178]. It consisted of four fully connected layers; an input layer (15 units), followed by two hidden layers (30 units) with leaky ($\alpha = 0.2$) rectifier activation and an output layer (3 units) with linear activation. A Gaussian Dropout layer (20% rate) was added after each hidden layer as regularization to avoid overfitting. The Adam optimizer was used for training.

The data (~ 5000 shots) was split into a (chronologically ordered) training and test dataset using a common 80/20 split ratio. The model learned from the first 80% of the shots (training dataset) to predict the following 20% of the shots (test dataset). Each variable of the training dataset was normalized to a mean of zero and a standard deviation of 1. The normalization factors of the training data were used to normalize the test dataset. The ANN was trained in batch sizes of 100 samples for a maximum of 500 epochs, while monitoring overfitting. Figure 4.8 shows the loss evolution during training. Training was stopped early if the accuracy did not increase and the so-far best model was selected as final result.

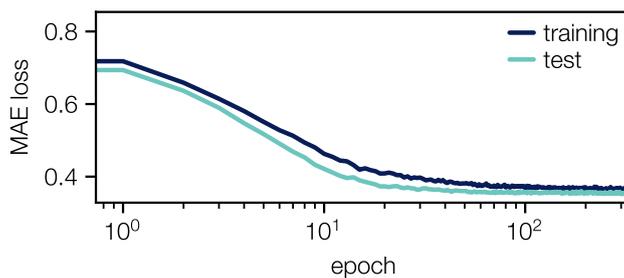


Figure 4.8 – Loss [mean absolute error (MAE)] evolution during training of the artificial neural network for the training (dark blue) and test (light blue) dataset. The final model was obtained after 330 epochs.

As expected from the analysis of the pairwise correlations in the previous section, the focus position [Fig. 4.7(a) and (c)] was the most important feature to predict the electron beam properties. Despite having less influence, all other laser diagnostics that were included further increased the predictive power of the model when compared to a model that excluded any of them. In particular, the laser central wavelength, laser energy, position and pointing through the plasma source in the horizontal (polarization) plane, wavefront astigmatism and coma had significant importance for the prediction. Note, however, that the feature importance for the prediction is naturally coupled to the magnitude of fluctuation of each parameter and does not necessarily reflect their physical importance for the injection and acceleration dynamics.

4.4.3 Beam quality prediction

In Fig. 4.9(a) and (c), we demonstrate the predictive power of the model by comparing the measured and predicted energy \tilde{E} and energy spread ΔE . For better visualization of the single-shot predictive capabilities, we show only 50 out of the ~ 1000 available consecutive shots. In addition, Fig. 4.9(c) compares the measured and predicted energy for entire test dataset.

Remarkably, the limited set of laser diagnostics (an energy sensor, a near- and far-field camera, a spectrometer and a wavefront sensor as described above) is already sufficient to accurately determine the electron beam quality. Over the entire test dataset, the model achieves a coefficient of determination (R^2) of 0.84, 0.74 and 0.73 for the energy, energy spread and charge, respectively. These results indicate that we not only identified the main laser parameters affecting stability, but also that the plasma source itself seems not to be a dominant source of instability.

By training over a shorter and testing over a longer period, we could additionally confirm that the predictive power remained constant for at least one hour. Even longer-term applicability of the surrogate model is expected as long as the accelerator

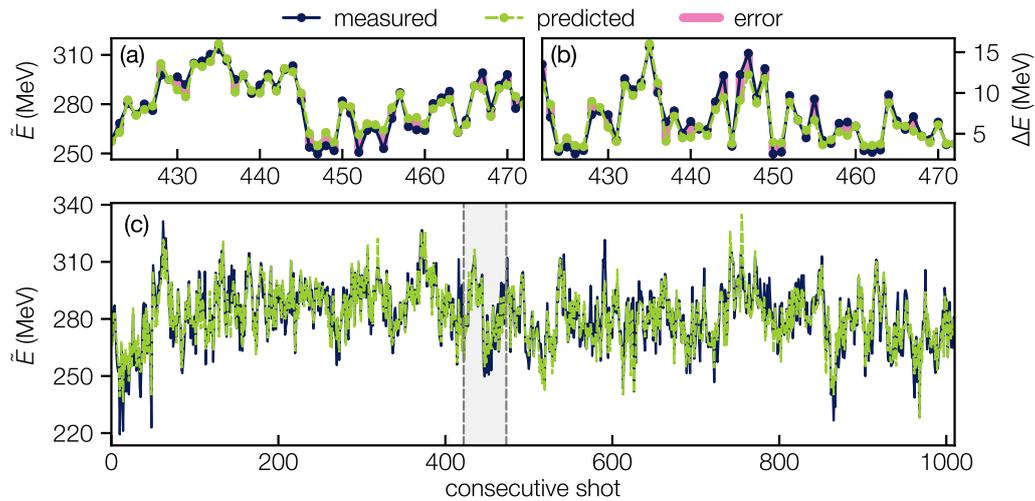


Figure 4.9 – Single-shot electron beam quality prediction by a neural network and from a set of input laser parameters. Measured (blue) and predicted (green) (a) median energy \tilde{E} and (b) energy spread ΔE for a series of 50 consecutive shots. Prediction error (red). (c) Prediction of the median energy for the entire 1000-shot test dataset. The 50 exemplary shots shown in (a) and (b) correspond to the gray selection marked in (c).

does not drift away significantly from the central working point that was used for training. For the focus position, laser energy and pointing, for example, we ensured that this did not happen by actively stabilizing slow drifts. In fact, the accelerator conditions remained stable during the 6 hours of continuous operation when recording the datasets analyzed in this chapter. Note that it would also be possible to retrain the model at a shifted working point or to pre-train it from a set of scans. The latter would ensure validity of the model over a wider range of input parameter variations.

One of the great advantages of a surrogate model is that we can use it to explore the physics beyond the actual measurement and evaluate strategies for future improvement. For example, we can benchmark the effect of a hypothetical active feedback system operating at 1 Hz: One could envision to stabilize each laser parameter to a target value, by correcting for 75% of the deviation from the previous shot. In this case, the model predicts a reduction of the rms jitter of all electron beam properties by more than 28%, which highlights the potential of fast feedback systems, especially when considering the transition to kHz repetition rates in next generation high power laser systems.

Similarly, we can quantify how specific improvements to the laser system could benefit the accelerator stability and derive guidelines for future laser development. For example, it seems feasible to (i) reduce the longitudinal focus position jitter by a factor of four to $41\mu\text{m}$ rms; (ii) halve the jitter of higher order wavefront aberrations to $\sim 2\text{ nm}$ rms;

and (iii) halve the central wavelength jitter to 0.15 nm rms. With those improvements the electron energy jitter would decrease by a factor of ~ 3 (then 1.9% rms) and 90% of all shots would have a relative energy spread $\Delta E/\tilde{E} < 1.7\%$.

The model can also be used to further optimize the beam quality. Within the fluctuations of the measurement, it already predicts a local minimum of the energy spread ($\sim 10\%$ smaller) at a slightly shifted set of parameters compared to the central working point. Wide-range online training of the model in future experiments could be used to efficiently find global optima.

4.5 Summary and outlook

In summary, we have demonstrated laser-plasma acceleration at optimal beam loading conditions, which resulted in a flattening of the average accelerating fields to $\sim 1\%$ over a substantial fraction of the plasma wavelength ($\sim \lambda_p/5$). This enabled the production of 1.2% rms energy spread electron bunches at 282 MeV with a peak spectral density of 5.3 pC MeV^{-1} (44 pC charge) and a transverse emittance of 1.9 mm mrad. Simulations indicate a beam loading efficiency of $\sim 19\%$.

Applications demand highly reproducible beams, which motivated us to study the origin of shot-to-shot variations of the injection and acceleration process. Using machine learning, we could build a surrogate model of the experiment that accurately (average $R^2 = 0.77$) predicts how deviations from the optimal beam loading conditions dilute the single-shot beam quality as a function of key drive laser properties. The model identifies and quantifies sources of instability, serves as a basis for active feedback and global optimization techniques, or can simply be used as a non-invasive virtual diagnostic for the electron beam. We expect such models to become an essential tool for the realization of plasma accelerators with application-relevant stability and quality.

With our plasma source design, we could demonstrate precise control over the injected phase space. Our approach combines the advantages of other highly localized injection techniques, such as shock-front injection, with the robustness of ionization injection and continuous flow operation.

Our strategy to achieve optimal beam loading conditions, i.e., injecting a suitable bunch current profile along an ionization-induced density downramp in combination with optimizing the beam loaded wakefield evolution during subsequent acceleration, has general validity. The strategy can also be applied to longer (or shorter) propagation distances, different densities or laser parameters.

For example, combining our approach with external laser guiding will make it possible to scale the beam energy to the GeV level. In fact, a combination of localized ionization injection and external guiding has recently been demonstrated in experiments [155].

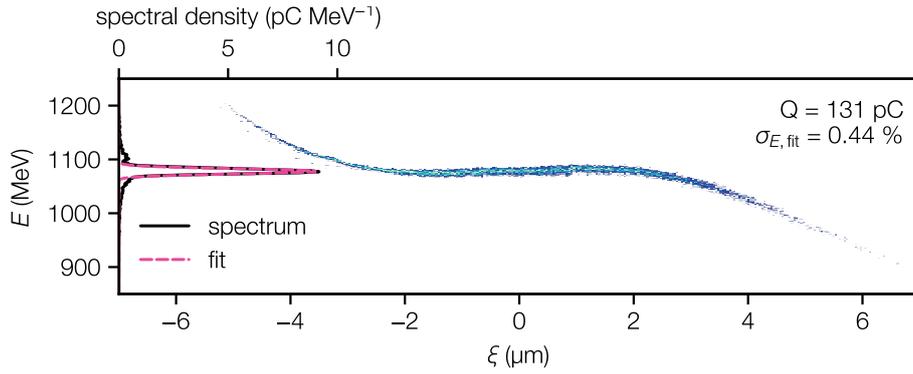


Figure 4.10 – Results from a simulation where the strategy of achieving optimal beam loading has been scaled to GeV beam energies by adding a plasma guiding channel to the plasma source. Longitudinal phase space of the accelerated electron beam with a charge of 131 pC and an rms energy spread of 0.4%. Projected spectrum (black) and Gaussian fit (magenta).

During acceleration to higher energies, we can assume that the initial absolute slice energy spread of the injected bunches is preserved. Operation at optimal beam loading conditions, as we have demonstrated, would then allow for projected energy spreads on the per-mille level and application-relevant energy-transfer efficiencies. The demonstration of per-mille energy spreads, however, will require to control the flattening of the fields with a similar precision. Of course, with our present experimental results, we could only demonstrate an upper limit of $\sim 1\%$ flattening at $\sim 20\%$ efficiency.

As an example that our approach can be scaled to higher energies, resulting in even lower relative energy spreads, Fig. 4.10 presents the results of a simulation of our plasma source where we extended the acceleration region to 25 mm by adding a plasma guiding channel. The driver was a Gaussian laser pulse with 10 J total energy, 41 μm FWHM focal spot size and 59 fs FWHM pulse length, leading to an a_0 of 2.0, while operating at a plasma density of $\sim 2 \times 10^{-17} \text{ cm}^{-3}$. At the final energy of $\sim 1 \text{ GeV}$, the electron bunch had an rms energy spread (Gaussian fit) of 0.4%, already marking an improvement. Further optimizations of such a setup with simulations and in experiments will be the subject of future work.

Part II

Development of novel particle-in-cell methods

5 The electromagnetic particle-in-cell method

This chapter introduces the fully explicit electromagnetic particle-in-cell algorithm, with an emphasis on its application for simulations of relativistic beams and plasmas. In Secs. 5.1 and 5.2, we give an overview of the method and describe the fundamental steps of the particle-in-cell cycle. We then discuss the most commonly used finite-difference and pseudospectral field solving methods in more detail in Sec. 5.3. Finally, in Sec. 5.4, we describe a class of numerical artifacts referred to as numerical Cherenkov, which can lead to erroneous results when simulating relativistic particles. This chapter borrows from earlier reviews on this topic [179, 180, 40].

5.1 Overview of the particle-in-cell algorithm

The modeling of many phenomena in plasma physics requires a fully kinetic description. In the context of plasma acceleration, this includes, for example, the excitation of highly nonlinear plasma waves or the self-injection of electrons. The tool of choice to model these processes is the electromagnetic particle-in-cell (PIC) method [38–40]. It efficiently approximates solutions of the Maxwell-Vlasov equation, by combining an Eulerian representation of the fields and their sources with a Lagrangian treatment of individual particles.

An electromagnetic particle-in-cell algorithm [38–40] iteratively solves Maxwell’s equations on a discrete grid

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}, \quad (5.1a)$$

$$\frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{B} - \mu_0 \mathbf{J}, \quad (5.1b)$$

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (5.1c)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (5.1d)$$

where \mathbf{E} and \mathbf{B} are the electric and magnetic field, ρ and \mathbf{J} are the charge and current densities, t is the time and $c^2 \epsilon_0 \mu_0 = 1$, respectively.

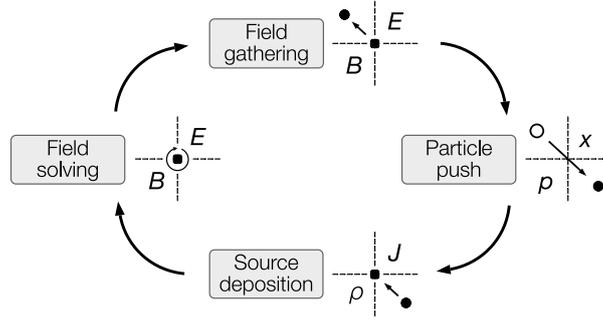


Figure 5.1 – Basic principle of the electromagnetic particle-in-cell method. (1) Particles gather electromagnetic fields from a discrete grid; (2) Their positions and momenta are advanced in time; (3) Updated charge and current densities are projected onto the grid; (4) The electromagnetic fields are advanced in time. The process is repeated for many iterations to simulate the time evolution of the system.

The electromagnetic fields act on freely moving particles that are advanced in a continuous phase space by solving the Newton-Lorentz equations of motions

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{\gamma m}, \quad (5.2a)$$

$$\frac{d\mathbf{p}}{dt} = q \left(\mathbf{E} + \frac{\mathbf{p}}{\gamma m} \times \mathbf{B} \right), \quad (5.2b)$$

where m , q , \mathbf{x} are the mass, charge, position, $\mathbf{p} = \gamma m \mathbf{v}$ is the momentum with \mathbf{v} the velocity and $\gamma = 1/\sqrt{1 - v^2/c^2}$ the relativistic factor of the particle, respectively.

Maxwell's equations [Eqs. (5.1a) to (5.1d)] and the equations of motion [Eqs. (5.2a) and (5.2b)] are coupled by *projecting* (or *depositing*) the particles charge ρ and current \mathbf{J} density distributions onto a spatial grid and by *interpolating* (or *gathering*) the \mathbf{E} and \mathbf{B} fields from this grid to the particle positions \mathbf{x} .

The coupled system is then evolved in time by recursively updating the fields and particles for many time steps Δt . One such iteration is called a *PIC cycle* and is conceptually shown in Fig. 5.1:

- (1) **Field gathering:** First, the \mathbf{E} and \mathbf{B} fields are interpolated from the grid to the particle positions.
- (2) **Particle push:** The particle momenta \mathbf{p} and positions \mathbf{x} are then advanced in time by solving the discretized Newton-Lorentz equations of motion.
- (3) **Charge and current deposition:** At the updated positions, the particles charge ρ and current \mathbf{J} density distributions are projected onto the grid.

- (4) **Field solving:** The updated sources are then used to integrate Maxwell's equations and advance the \mathbf{E} and \mathbf{B} fields in time.

In between these fundamental steps, additional physical or numerical operations are often inserted into the PIC cycle, such as a filtering of the deposited source terms or fields, the application of external forces or the evaluation of elementary particle processes. The core elements of a particle-in-cell algorithm are presented in detail in the next section.

5.1.1 Discretization

A variety of methods exist for numerically solving Maxwell's equations. In the following, we consider the most commonly used second-order finite-difference [181] and the class of pseudospectral [38, 182, 183, 45] field solving methods. Although detailed explanations of these Maxwell solvers will be given only later, it is useful to define their discretization in advance.

In these algorithms, the fields are discretized on a spatial grid and Maxwell's equations are integrated in the time domain, either directly in real space or after transforming the field quantities to Fourier space. Two different discretization schemes are typically used in the spatial domain. Finite-difference methods often use a *staggered* grid where each field component is shifted by half a grid cell in each direction such that all discrete derivatives are centered and therefore second-order accurate. With spectral methods, on the other hand, it is more common to define all fields at the same grid positions, which is referred to as *cell-centered* or *nodal* grid.

The numerical integration in time is typically based on an explicit leapfrog scheme, where \mathbf{E} and ρ are defined at integer time steps $n\Delta t$ and \mathbf{B} and \mathbf{J} are defined at half-integer time steps $(n + 1/2)\Delta t$.

In the following, we consider a regular spaced 3D Cartesian grid with $N_x \times N_y \times N_z$ grid points

$$\mathbf{x}_j = j_x \Delta x \hat{\mathbf{u}}_x + j_y \Delta y \hat{\mathbf{u}}_y + j_z \Delta z \hat{\mathbf{u}}_z, \quad \begin{aligned} j_x &\in [0, N_x - 1] \\ j_y &\in [0, N_y - 1] \\ j_z &\in [0, N_z - 1] \end{aligned} \quad (5.3)$$

where j_x , j_y and j_z are integers, Δx , Δy and Δz are the cell sizes and $\hat{\mathbf{u}}_x$, $\hat{\mathbf{u}}_y$ and $\hat{\mathbf{u}}_z$ is the unit vector in each direction. For simplicity and unless defined otherwise, the fields $F(\mathbf{x})$ are all defined on the grid nodes \mathbf{x}_j , where F is either \mathbf{E} , \mathbf{B} , \mathbf{J} or ρ and $F^n \equiv F(\mathbf{x}_j, n\Delta t)$.

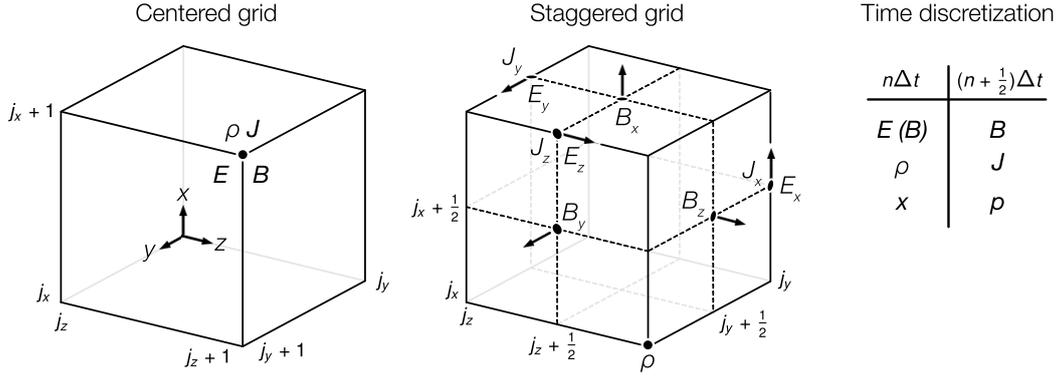


Figure 5.2 – Space and time discretization of the physical quantities in a particle-in-cell algorithm. The electromagnetic fields are usually discretized on a spatial grid with a cell-centered or staggered (Yee) configuration. The leapfrog integration in time implies a staggering in time.

When referring to a staggered configuration of the fields, we denote the shifted positions of the field quantities as $\mathbf{x}_{j,F} = \mathbf{x}_j + \Delta\mathbf{x}_F$, where $\Delta\mathbf{x}_F$ is a half-cell translation of the grid that depends on F .

With pseudospectral methods, Maxwell's equations are solved in Fourier space where the fields are represented as a sum of Fourier modes

$$F(\mathbf{x}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} d\mathbf{k} \hat{\mathcal{F}}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (5.4)$$

with the Fourier components being defined as

$$\hat{\mathcal{F}}(\mathbf{k}) = \int_{\mathbb{R}^3} d\mathbf{x} F(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}. \quad (5.5)$$

Accordingly, the discrete Fourier transform of $F(\mathbf{x}_j)$ and its inverse read

$$\hat{\mathcal{F}}(\mathbf{k}_j) = \frac{1}{N_x N_y N_z} \sum_j e^{-i\mathbf{k}_j \cdot \mathbf{x}_j} F(\mathbf{x}_j), \quad (5.6)$$

$$F(\mathbf{x}_j) = \sum_{\hat{j}} e^{i\mathbf{k}_{\hat{j}} \cdot \mathbf{x}_j} \hat{\mathcal{F}}(\mathbf{k}_{\hat{j}}), \quad (5.7)$$

with the vectors of the reciprocal lattice of \mathbf{x}_j given by

$$\mathbf{k}_{\hat{j}} = \hat{j}_x \frac{2\pi}{N_x \Delta x} \hat{\mathbf{u}}_x + \hat{j}_y \frac{2\pi}{N_y \Delta y} \hat{\mathbf{u}}_y + \hat{j}_z \frac{2\pi}{N_z \Delta z} \hat{\mathbf{u}}_z. \quad (5.8)$$

$\hat{j}_x \in [-N_x/2, N_x/2 - 1]$
 $\hat{j}_y \in [-N_y/2, N_y/2 - 1]$
 $\hat{j}_z \in [-N_z/2, N_z/2 - 1]$

The equations of motion, Eqs. (5.2a) and (5.2b), are discretized and integrated in time over successive time steps $n\Delta t$ using a leapfrog method with the positions \mathbf{x} defined at integer time steps $n\Delta t$ and the momenta \mathbf{p} at half-integer time steps $(n + 1/2)\Delta t$, respectively. Figure 5.2 visualizes the different grid configurations and summarizes the discretization in time.

5.1.2 Macroparticles and shape factors

Another fundamental aspect of the particle-in-cell method is the concept of finite size *macroparticles* [40]. Considering all individual physical particles in a particle-in-cell algorithm would create an unfeasible amount of computational cost, without providing any advantage for modeling their large-scale collective behavior that is of interest in simulations of collisionless plasmas.

Clusters of real particles in phase space are therefore represented as macroparticles of charge $q_\mu = w_\mu \cdot q$ and with a mass $m_\mu = w_\mu \cdot m$, where w_μ is the *weight* of each macroparticle with index μ relating it to the number of physical particles and q and m are the charge and mass of a single real particle, respectively.

Each macroparticle carries a momentum \mathbf{p}_μ and is attributed a finite spatial extent, or *shape*, around its position \mathbf{x}_μ that typically covers one or up to a few cells. A macroparticle thus represents a cloud of physical particles that all have the same momentum and whose spatial distribution is approximated by a fixed shape function that defines the macroparticle's range of interaction with closest grid points.

Consequently, when interpolating from or projecting to the numerical grid, the individual field quantities are weighted by *shape factors* $S(\mathbf{x})$. A common definition of these interpolation functions is based on splines of order l [184]. For example, first order ($l = 1$) shape factors corresponding to a linear weighting between the two nearest grid points in a single direction z can be defined as

$$S^1(z) = \left(1 - \frac{|z|}{\Delta z}\right) \Theta(\Delta z - |z|). \quad (5.9)$$

Higher order splines ($l > 1$) take into account more neighboring grid points. This can reduce the noise in a simulation, but also increases the computational cost of the grid interpolation and projection algorithms [39]. A convenient generalization of these shape factors to arbitrary order is given by their Fourier space representation

$$\hat{S}^l(k_z) = \frac{1}{\Delta z} \int_{-\infty}^{\infty} dz e^{-ik_z z} S^l(z) = \text{sinc}^{l+1}\left(\frac{k_z \Delta z}{2}\right), \quad (5.10)$$

where $\text{sinc}(x) = \sin(x)/x$.

5.2 The particle-in-cell cycle

5.2.1 Field gathering and particle push

In the following, we describe the individual steps of the PIC cycle [40] as shown in Fig. 5.3. In the first step [Fig. 5.3(a)], the electromagnetic fields are interpolated from the numerical grid to the particle positions \mathbf{x}_μ^n . Here, we assume that the \mathbf{E}^n and \mathbf{B}^n fields are known initially. However, this requires prior interpolation between $\mathbf{B}^{n-1/2}$ and $\mathbf{B}^{n+1/2}$ for algorithms where the magnetic field is defined at half-integer time steps. Expressions for the fields gathered at the particle position \mathbf{x}_μ then read

$$\mathbf{E}_{\text{interp}}^n(\mathbf{x}_\mu^n) = \sum_j S(\mathbf{x}_\mu^n - \mathbf{x}_j) \mathbf{E}^n(\mathbf{x}_j), \quad (5.11)$$

$$\mathbf{B}_{\text{interp}}^n(\mathbf{x}_\mu^n) = \sum_j S(\mathbf{x}_\mu^n - \mathbf{x}_j) \mathbf{B}^n(\mathbf{x}_j), \quad (5.12)$$

where $\mathbf{E}^n(\mathbf{x}_j)$ and $\mathbf{B}^n(\mathbf{x}_j)$ are the fields defined at the grid points \mathbf{x}_j at time $n\Delta t$ and $S(\mathbf{x}) = S(x) \times S(y) \times S(z)$ is a product of identical shape factors in each direction.

With the electromagnetic fields defined at the particle positions, discretized versions of Eqs. (5.2a) and (5.2b) can be solved [Fig. 5.3(b)]. First, the momenta are updated, for example, by using the *Vay pusher* [185],

$$\frac{\mathbf{p}^{n+\frac{1}{2}} - \mathbf{p}^{n-\frac{1}{2}}}{\Delta t} = q \mathbf{E}_{\text{interp}}^n + q \left(\frac{\mathbf{p}^{n+\frac{1}{2}}}{2m\gamma^{n+\frac{1}{2}}} + \frac{\mathbf{p}^{n-\frac{1}{2}}}{2m\gamma^{n-\frac{1}{2}}} \right) \times \mathbf{B}_{\text{interp}}^n. \quad (5.13)$$

Note that in practice, solving this implicit equation for $\mathbf{p}^{n+1/2}$ requires a more involved explicit procedure, which is not detailed here. The new momenta can then be used to push the particles to their updated positions \mathbf{x}^{n+1}

$$\frac{\mathbf{x}^{n+1} - \mathbf{x}^n}{\Delta t} = \frac{\mathbf{p}^{n+\frac{1}{2}}}{m\gamma^{n+\frac{1}{2}}}, \quad (5.14)$$

where $\gamma^n = \sqrt{1 + (\mathbf{p}^n/mc)^2}$. This operation is usually split into two separate half-pushes in order to have the particle positions and momenta synchronized at $(n + 1/2)\Delta t$.

Another common algorithm to update the momentum is the *Boris pusher* [186]. Both algorithms preserve the kinetic energy of a particle in the absence of electric fields ($\mathbf{E} = 0$). The Vay pusher additionally ensures that the force acting on a particle is exactly zero when the electric and magnetic fields compensate ($\mathbf{E} + \mathbf{v} \times \mathbf{B} = 0$) [187]. Because this avoids errors in simulations of relativistic beams or plasmas, it is

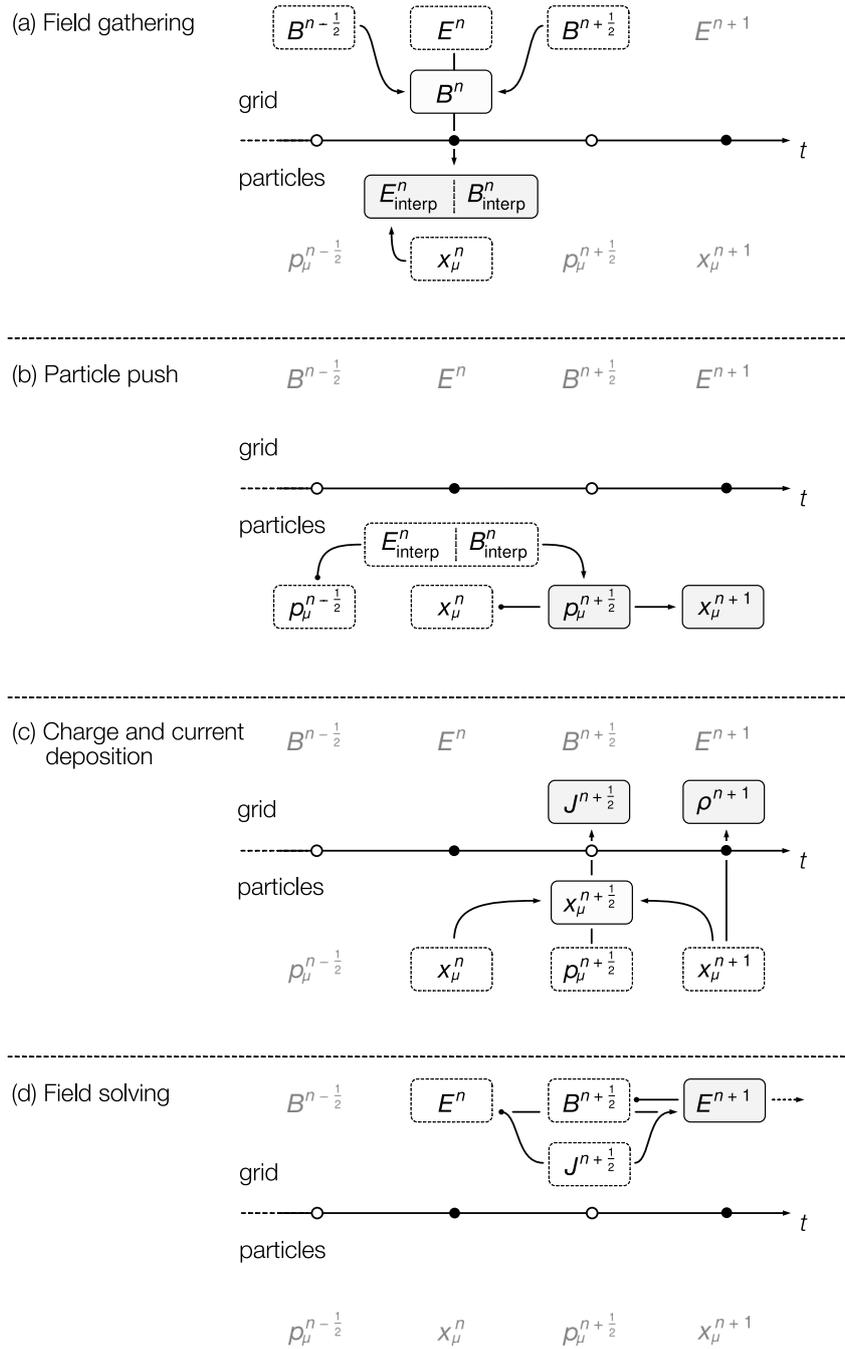


Figure 5.3 – Main steps of the particle-in-cell cycle. (a) The synchronized $\mathbf{E}(\mathbf{x}_j)$ and $\mathbf{B}(\mathbf{x}_j)$ fields are interpolated to the macroparticle positions \mathbf{x}_μ at $t/\Delta t = n$. (b) $\mathbf{E}_{\text{interp}}$ and $\mathbf{B}_{\text{interp}}$ are used to first advance the particle momenta \mathbf{p}_μ to $t/\Delta t = n + \frac{1}{2}$. The positions \mathbf{x}_μ are then advanced to $t/\Delta t = n + 1$. (c) The current density is deposited at $t/\Delta t = n + \frac{1}{2}$, the charge density at $t/\Delta t = n + 1$. (d) Usually, the electric field is first advanced to $t/\Delta t = n + 1$ and then used to advance the magnetic field to $t/\Delta t = n + \frac{3}{2}$. With pseudospectral algorithms, the integration is performed in Fourier space. The illustration is inspired by a figure from [180].

the natural choice for the work presented here. Further note that the more recently developed *Higuera pusher* [187] unites advantages of the former two algorithms, but is not considered here.

5.2.2 Charge and current deposition

Consequently, in order to solve Maxwell's equations and advance the fields in time, it is required to update the source terms by projecting the particle quantities onto the grid [Fig. 5.3(c)]. The charge density at the next time step $(n + 1)\Delta t$ is obtained from *direct* deposition onto the grid

$$\rho^{n+1}(\mathbf{x}_j) = \frac{1}{\Delta x \Delta y \Delta z} \sum_{\mu} qw_{\mu} S(\mathbf{x}_j - \mathbf{x}_{\mu}^{n+1}), \quad (5.15)$$

where the value at each grid point \mathbf{x}_j is the cell-averaged contribution of all macroparticles weighted by the shape factors S . Similarly, one can obtain the current density at time $(n + 1/2)\Delta t$,

$$\mathbf{J}_d^{n+1/2}(\mathbf{x}_j) = \frac{1}{\Delta x \Delta y \Delta z} \sum_{\mu} qw_{\mu} \mathbf{v}_{\mu}^{n+1/2} S(\mathbf{x}_j - \mathbf{x}_{\mu}^{n+1/2}). \quad (5.16)$$

However, for this directly projected current density, $\nabla \cdot \mathbf{J}_d^{n+1/2} \neq -(\rho^{n+1} - \rho^n)/\Delta t$, hence, in general, such a deposition does not satisfy the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0. \quad (5.17)$$

A common solution to this problem for finite-difference algorithms is to use more complex, *charge conserving* deposition algorithms, such as the widespread *Esirkepov* [188] deposition or other related methods [189, 190] instead of Eq. (5.16). An equivalent of the Esirkepov deposition has also been developed for pseudospectral solvers [45]. Alternatively, the directly deposited current density $\mathbf{J}_d^{n+1/2}$ can also be corrected, preferably in Fourier space, according to [191, 179]

$$\hat{\mathcal{J}}^{n+1/2} = \hat{\mathcal{J}}_d^{n+1/2} + \frac{i\mathbf{k}}{k^2} \hat{\mathcal{G}}, \quad (5.18a)$$

$$\hat{\mathcal{G}} = \frac{\hat{\varrho}^{n+1} - \hat{\varrho}^n}{\Delta t} + i\mathbf{k} \cdot \hat{\mathcal{J}}_d^{n+1/2}, \quad (5.18b)$$

with the corrected current $\hat{\mathcal{J}}^{n+1/2}$ then exactly satisfying the continuity equation, i.e., $(\hat{\varrho}^{n+1} - \hat{\varrho}^n)/\Delta t + i\mathbf{k} \cdot \hat{\mathcal{J}}^{n+1/2} = 0$. The operation changes the current without modifying its curl and is therefore often denoted *curl-free current correction*.

As a complement to using high-order shape factors, it is common practice to apply a digital filter to the deposited charge and current density. For example, a three-point filter of the form [40]

$$F_{j_z}^f = \alpha F_{j_z} + (1 - \alpha)(F_{j_z-1} + F_{j_z+1})/2 \quad (5.19)$$

is often applied, where $F_{j_z}^f$ is the filtered quantity at grid point j_z in a single direction z . For $\alpha = 1/2$, this operation is called a *binomial smoother* that acts as a low-pass filter in the frequency domain and thereby effectively suppresses the noisy high-frequency content of the particle distribution. Because the smoother alone reduces the amplitude of the signal over a wide range of frequencies, an additional compensation filter ($\alpha = 3/2$) can be added that counteracts the attenuation of lower frequencies.

In Fourier space, such a combination of n passes of a binomial smoother followed by a compensator is equivalent to multiplying the field $\hat{\mathcal{F}}(k_z)$ with a transfer function [40]

$$\hat{\mathcal{T}}^n(k_z) = \left(1 - \sin^2\left(\frac{k_z \Delta z}{2}\right)\right)^n \left(1 + n \sin^2\left(\frac{k_z \Delta z}{2}\right)\right). \quad (5.20)$$

5.2.3 Field solving

In the final step of the PIC cycle [Fig. 5.3(d)], Maxwell's equations are integrated in time to retrieve the updated electromagnetic fields, \mathbf{E}^{n+1} and $\mathbf{B}^{n+3/2}$ (or \mathbf{B}^{n+1} depending on the Maxwell solver).

Advancing the fields in time only requires to solve the first two of Maxwell's equations, i.e., the Maxwell-Faraday [Eq. (5.1a)] and the Maxwell-Ampere equation [Eq. (5.1b)]. This is because if Eq. (5.1c) ($\nabla \cdot \mathbf{B} = 0$) is satisfied initially, it will automatically be satisfied at later times. The same applies to Eq. (5.1d) ($\nabla \cdot \mathbf{E} = \rho/\epsilon_0$), provided that the algorithm obeys the continuity equation, Eq. (5.17).

For algorithms that violate charge conservation, an alternative solution is to add a *Poisson corrector* [40] to the cycle, which modifies the electric field such that Gauss's law [Eq. (5.1d)] holds nonetheless. For example, similar to the current correction from the previous section, a *Boris correction* [186] can be applied to the electric field. Another possibility is to gradually diffuse away errors, which is referred to as *Marder correction* [192]. Nonetheless, the preferred solution is to use a charge conserving current deposition or to correct the currents, in particular since these operations are more local than correcting the electric field and thus better suited for parallelization [179].

As mentioned at the beginning of this chapter, the discretization and numerical integration of Eq. (5.1a) and Eq. (5.1b) can vary a lot between different Maxwell solvers and will be discussed in detail in the next section.

5.3 Maxwell solvers

The choice of a particular field solving method can have a significant influence on the numerical accuracy and stability of a simulation. For these reasons, a variety of Maxwell solvers have been developed in the past, a few of them specifically designed for simulations of relativistic beams or plasma acceleration. However, while some of these methods provide superior accuracy, they can increase the computational cost or prevent the efficient parallelization of a particle-in-cell algorithm to distributed compute units. The development of novel or the improvement of existing Maxwell solvers therefore continues to be an active part of research in the field of computational plasma physics.

In the following section, we summarize the Maxwell solvers relevant to the new developments presented in this work. This includes the traditional finite-difference time-domain (FDTD) solver [181], the pseudospectral time-domain (PSTD) [38, 182] and the pseudospectral analytical time-domain (PSATD) solver [183], as well as their generalization to arbitrary order [45, 46].

5.3.1 Finite-difference time-domain

The standard finite-difference time-domain (FDTD) solver (also known as *Yee solver* [181]) is arguably the most popular and widely used Maxwell solver. It uses a simple and robust configuration of space- and time-centered numerical derivatives, resulting in a second-order accurate representation of Maxwell's equations.

As already mentioned before in Sec. 5.1.1, the discretization is realized by arranging the fields on the *Yee lattice*, i.e., $F^n \equiv F(\mathbf{x}_{j,F}, n\Delta t)$, with the staggered grid locations $\mathbf{x}_{j,F}$ varying for each individual component as depicted in Fig. 5.2. Consequently, the discretized versions of the Maxwell-Faraday and Maxwell-Ampere equation read [181]

$$\frac{\mathbf{B}^{n+\frac{1}{2}} - \mathbf{B}^{n-\frac{1}{2}}}{\Delta t} = -\hat{\nabla} \times \mathbf{E}^n, \quad (5.21a)$$

$$\frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{c^2\Delta t} = \hat{\nabla} \times \mathbf{B}^{n+\frac{1}{2}} - \mu_0 \mathbf{J}^{n+\frac{1}{2}}. \quad (5.21b)$$

The discrete differential operator is defined as $\hat{\nabla} = D_x \hat{\mathbf{u}}_x + D_y \hat{\mathbf{u}}_y + D_z \hat{\mathbf{u}}_z$ where the finite-difference operator along z is

$$(D_z F)_{j_x, j_y, j_z}^n := \frac{F_{j_x, j_y, j_z + \frac{1}{2}}^n - F_{j_x, j_y, j_z - \frac{1}{2}}^n}{\Delta z} \approx \frac{\partial F^n(x, y, z)}{\partial z}. \quad (5.22)$$

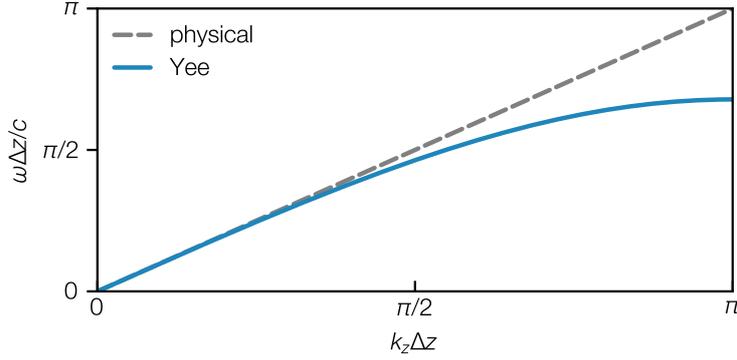


Figure 5.4 – Subluminal numerical dispersion relation of the standard FDTD (Yee) solver for waves propagating in a single direction z ($k_x = k_y = 0$) and for a time step at the CFL limit, $c\Delta t = \Delta z/\sqrt{3}$, for $\Delta x = \Delta y = \Delta z$.

Note that here the indices (j_x, j_y, j_z) refer to grid positions relative to $\mathbf{x}_{j,F}$ and that n has to be replaced with $n + 1$ in the first equation to update the magnetic field from $\mathbf{B}^{n+1/2}$ to $\mathbf{B}^{n+3/2}$ at the end of the PIC cycle.

As can be seen from Eq. (5.22), numerically calculating the derivatives depends only on information from the two nearest grid points. This makes updating the fields a highly localized operation in space, which is a prerequisite to efficiently parallelize the execution of a particle-in-cell algorithm via *domain decomposition*. This parallelization technique is based on spatially decomposing the simulation grid into smaller volumes that can be efficiently computed in parallel and on distributed compute units. Thanks to the locality of the FDTD solver, only the field information (and particles) in a thin boundary layer have to be exchanged between adjacent domains, which minimizes costly communication. FDTD solvers in combination with domain decomposition are therefore at the heart of most parallel particle-in-cell algorithms, especially since the present generation of supercomputers are built as clusters of many interconnected compute devices with fast shared memory but relatively slow channels of communication in between them.

However, the approximation of the spatial and temporal derivatives by low order finite differences introduces errors to the dispersion of electromagnetic waves. While electromagnetic modes of the form $e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$ should obey the physical dispersion relation in vacuum

$$\omega^2 = c^2 \mathbf{k}^2, \quad (5.23)$$

where \mathbf{k} and ω , respectively, are the wavevector and angular frequency, the *numerical* dispersion relation of the standard FDTD solver reads [193]

$$\frac{1}{c^2 \Delta t^2} \sin^2 \left(\frac{\omega \Delta t}{2} \right) = \frac{1}{\Delta x^2} \sin^2 \left(\frac{k_x \Delta x}{2} \right) + \frac{1}{\Delta y^2} \sin^2 \left(\frac{k_y \Delta y}{2} \right) + \frac{1}{\Delta z^2} \sin^2 \left(\frac{k_z \Delta z}{2} \right) \quad (5.24)$$

and reduces to the physical one only in the limit $\omega\Delta t \ll 1$ and $k_{x,y,z}\Delta(x,y,z) \ll 1$. The difference between Eq. (5.23) and Eq. (5.24) is shown in Fig. 5.4. The numerical dispersion relation becomes increasingly distorted at large wavenumbers $k = |\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2}$, causing electromagnetic waves to propagate at subluminal ($\leq c$) vacuum phase $v_\phi = \omega/k$ and group velocities $v_g = \partial\omega/\partial k$.

The presence of spurious dispersion can seriously degrade the accuracy and validity of plasma acceleration simulations. For example, the deviations from the physical group velocity cause a propagating laser pulse to slow down, potentially leading to an overestimation of the trapping probability and the dephasing of electrons in a laser-driven plasma wave [41, 43]. An accelerated electron beam can furthermore travel at relativistic speeds that surpass the phase velocity of the distorted electromagnetic modes, resulting in the artificial emission and resonant excitation of Cherenkov-like radiation [194, 54]. Because this so-called *numerical Cherenkov radiation* can have detrimental effects on the phase space of the beam [42, 44], its mitigation is of utmost importance to simulations of plasma acceleration and will be discussed in detail in the next section.

A straightforward approach to reduce these undesirable effects is to increase the spatial resolution of a simulation so as to improve the numerical dispersion of the physical frequencies of interest. However, applying this measure is further complicated by the presence of a *Courant-Friedrichs-Lewy* (CFL) condition [195]

$$\frac{1}{c^2\Delta t^2} \geq \frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2}, \quad (5.25)$$

which limits the maximum allowed propagation distance per time step $c\Delta t$ to a fraction of the grid resolution to ensure numerical stability of the field solver. This limitation emerges from the fact that ω in Eq. (5.24) becomes imaginary for time steps exceeding the CFL condition, which would allow for unphysical solutions of Eqs. (5.21a) and (5.21b) that grow exponentially. Thus, increasing the spatial resolution to mitigate problems from spurious numerical dispersion in combination with fulfilling the above requirement can drastically increase the computational cost of a simulation beyond practical feasibility.

Non-standard finite-difference (NSFDTD) [196–198, 41, 42, 193] solvers have been developed to circumvent some of the described problems by adjusting or—under certain conditions—even removing spurious dispersion along one or more axes. Yet, these solvers can introduce other complications.

Worth highlighting is the so-called Cole-Karkkainen solver [196, 198] with Cowan coefficients [41] (CKC). At the CFL limit, it allows for dispersion-less simulations of plasma acceleration in the axial direction. It will be used in chapter Ch. 8 for a

comparison between a finite-difference 3D Cartesian and a quasi-cylindrical spectral solver.

5.3.2 Pseudospectral time-domain

In contrast to FDTD solvers, pseudospectral methods follow an entirely different approach to solving Maxwell's equations. Instead of approximating the spatial derivatives by finite-differences in real space, the integration is done in Fourier space, where taking the derivative of a function $f(x)$ is as simple as multiplying its Fourier transform $\mathcal{F}(f(x))$ with ik ,

$$\mathcal{F}\left(\frac{df(x)}{dx}\right) = \int_{-\infty}^{\infty} \frac{df(x)}{dx} e^{-ikx} dx = ik \int_{-\infty}^{\infty} f(x) e^{-ikx} dx = ik\mathcal{F}(f(x)). \quad (5.26)$$

In Fourier space, Maxwell's equations [Eqs. (5.1a) to (5.1d)] become

$$\frac{\partial \hat{\mathbf{B}}}{\partial t} = -i\mathbf{k} \times \hat{\mathbf{E}}, \quad (5.27a)$$

$$\frac{1}{c^2} \frac{\partial \hat{\mathbf{E}}}{\partial t} = i\mathbf{k} \times \hat{\mathbf{B}} - \mu_0 \hat{\mathcal{J}}, \quad (5.27b)$$

$$i\mathbf{k} \cdot \hat{\mathbf{E}} = \frac{\hat{\rho}}{\epsilon_0}, \quad (5.27c)$$

$$i\mathbf{k} \cdot \hat{\mathbf{B}} = 0. \quad (5.27d)$$

These equations also have to be integrated in time, for example, by using the same leapfrog integration as above. In the following, $\hat{\mathcal{F}}^n \equiv \hat{\mathcal{F}}(\mathbf{k}_j, n\Delta t)$, where $\hat{\mathcal{F}}(\mathbf{k}_j)$ are the Fourier components of a field quantity $F(\mathbf{x}_j)$ (compare Sec. 5.1.1). We further omit the reference to discrete k -space by using $\mathbf{k} = \mathbf{k}_j$. The discretized version of Eqs. (5.27a) and (5.27b) are then given by [38, 182]

$$\frac{\hat{\mathbf{B}}^{n+\frac{1}{2}} - \hat{\mathbf{B}}^{n-\frac{1}{2}}}{\Delta t} = -i\mathbf{k} \times \hat{\mathbf{E}}^n, \quad (5.28a)$$

$$\frac{\hat{\mathbf{E}}^{n+1} - \hat{\mathbf{E}}^n}{c^2 \Delta t} = i\mathbf{k} \times \hat{\mathbf{B}}^{n+\frac{1}{2}} - \mu_0 \hat{\mathcal{J}}^{n+\frac{1}{2}}. \quad (5.28b)$$

While the discretization in time is the same as for the FDTD solver, the spatial derivatives in this so-called pseudospectral time-domain (PSTD) solver are equivalent to finite-difference approximations of *infinite order* [46].

This advantage, however, comes at the price of transforming the field quantities back and forth between their real space and Fourier space representation when advancing the fields within a cycle. These grid transformations are global operations. The resulting

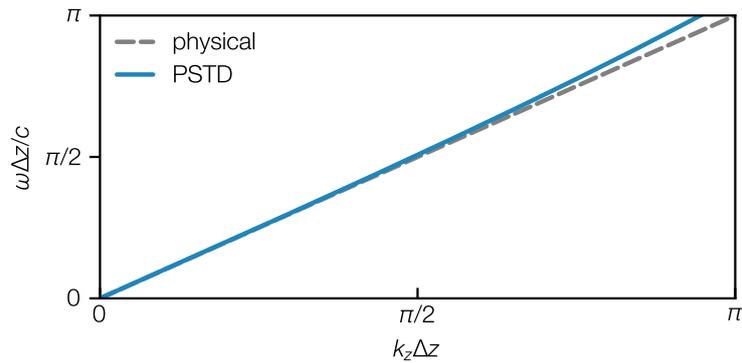


Figure 5.5 – Superluminal numerical dispersion relation of the PSTD solver for waves propagating in a single direction z ($k_x = k_y = 0$) and for a time step at the CFL limit, $c\Delta t = 2/(\sqrt{3}\pi)\Delta z$, for $\Delta x = \Delta y = \Delta z$.

overhead in communication across the entire simulation domain can prevent an efficient parallelization by domain decomposition. Pseudospectral solvers therefore either have to be executed on hardware that allows for a fast exchange of information across the grid or—as we will learn in Sec. 5.3.4—have to be adapted to add spatial locality to the field-solving operations.

Since the integration remains only second-order accurate in time, the propagation of electromagnetic waves is still modeled incorrectly with the PSTD solver, with a numerical dispersion relation [182]

$$\sin\left(\frac{\omega\Delta t}{2}\right) = k\frac{c\Delta t}{2}. \quad (5.29)$$

Consequently, the PSTD solver is also subject to a CFL stability criterion [182]

$$\frac{1}{c^2\Delta t^2} \geq \frac{\pi}{2} \left(\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2} \right). \quad (5.30)$$

As shown in Fig. 5.5 and contrary to the standard FDTD solver, Eq. (5.29) supports vacuum electromagnetic modes that propagate at superluminal ($\geq c$) phase velocities. The numerical dispersion relation furthermore converges to the correct one for $\omega\Delta t \ll 1$, i.e., the amount of spurious dispersion can be significantly reduced simply by choosing time steps much smaller than the CFL condition. Thus, when using the PSTD solver in practice, *subcycling* [40] of the field solving step can increase its accuracy, by updating Eqs. (5.28a) and (5.28b) multiple times for a fraction of the time step $\Delta t/m$ where m is the number of subcycles within one full PIC cycle.

5.3.3 Pseudospectral analytical time-domain

The increase in accuracy from the spectral representation of the spatial derivatives was still limited by the fact that the leapfrog integration is only second-order accurate in time in the PSTD solver. However, starting from the Fourier space representation of Maxwell's equations and under the assumption that the current density remains constant over one time step, the Maxwell-Faraday and the Maxwell-Ampere equation can also be integrated *analytically* in time, which results in the pseudospectral analytical time-domain (PSATD) solver [183, 45]

$$\hat{\mathbf{B}}^{n+1} = C\hat{\mathbf{B}}^n - \frac{S}{ck}i\mathbf{k} \times \hat{\mathbf{E}}^n + \frac{1-C}{\epsilon_0 c^2 k^2}i\mathbf{k} \times \hat{\mathcal{J}}^{n+\frac{1}{2}}, \quad (5.31a)$$

$$\begin{aligned} \hat{\mathbf{E}}^{n+1} = & C\hat{\mathbf{E}}^n + \frac{S}{k}c i\mathbf{k} \times \hat{\mathbf{B}}^n - \frac{S}{\epsilon_0 ck} \hat{\mathcal{J}}^{n+\frac{1}{2}} \\ & - \frac{i\mathbf{k}}{\epsilon_0 k^2} \left[\hat{\varrho}^{n+1} \left(1 - \frac{S}{ck\Delta t} \right) - \hat{\varrho}^n \left(C - \frac{S}{ck\Delta t} \right) \right], \end{aligned} \quad (5.31b)$$

where the shorthand notations $C \equiv \cos(ck\Delta t)$ and $S \equiv \sin(ck\Delta t)$ were used and $k = |\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2}$.

The PSATD solver provides exact solutions to Maxwell's equations in vacuum, it accurately reproduces the physical vacuum dispersion relation [Eq. (5.23)] and it is not subject to a CFL limit. While this means that in principle arbitrarily long time steps are possible, a time step $c\Delta t = \Delta z$ is typically chosen in simulations of relativistic beams or plasma acceleration, where Δz is the cell size in the direction of propagation.

The analytical integration furthermore allows for the \mathbf{E} and \mathbf{B} fields to be defined at the same point in time. This contrasts the usual time-staggering of the electromagnetic fields in the FDTD and PSTD solver, which can prevent a correct compensation of the electric and magnetic term in the Lorentz force [Eq. (5.2b)] [185]. The problem becomes most prevalent in situations where a relativistic electron beam co-propagates and overlaps with a laser pulse, as can be the case in laser-wakefield acceleration, for direct laser acceleration or in a free-electron laser. It was shown [199, 43] that the force experienced by the electron beam can be largely overestimated for such a configuration.

It should also be highlighted that the above expression for advancing the electric field [Eq. (5.31b)] explicitly depends on the charge density ρ , which was not the case for the expressions of the FDTD and PSTD solver. However, a ρ -free formulation can be derived

$$\hat{\mathbf{B}}^{n+1} = C\hat{\mathbf{B}}^n - \frac{S}{ck}i\mathbf{k} \times \hat{\mathbf{E}}^n + \frac{1-C}{\epsilon_0 c^2 k^2}i\mathbf{k} \times \hat{\mathcal{J}}^{n+\frac{1}{2}}, \quad (5.32a)$$

$$\begin{aligned} \hat{\mathcal{E}}^{n+1} = & C\hat{\mathcal{E}}^n + \frac{S}{k} c i \mathbf{k} \times \hat{\mathcal{B}}^n - \frac{S}{\epsilon_0 c k} \hat{\mathcal{J}}^{n+\frac{1}{2}} \\ & + \frac{\mathbf{k}}{\epsilon_0 k^2} \left[\left(\frac{S}{ck} - \Delta t \right) \mathbf{k} \cdot \hat{\mathcal{J}}^{n+\frac{1}{2}} + \epsilon_0 (1 - C) \mathbf{k} \cdot \hat{\mathcal{E}}^n \right], \end{aligned} \quad (5.32b)$$

where the continuity equation [Eq. (5.17)] and Gauss's law [Eq. (5.1c)] have been used to replace the charge densities in the last term with the divergence of the electric field and of the current density, respectively. In practice, the ρ -free formulation for updating the fields can lift the requirement to include the heavier ion species in a simulation, provided that only the current of the electrons contributes to the plasma dynamics.

The connection between the PSATD and the FDTD and PSTD solver becomes more evident when rewriting these equations in their time-staggered form

$$\hat{\mathcal{B}}^{n+\frac{1}{2}} - \hat{\mathcal{B}}^{n-\frac{1}{2}} = -\frac{2S_h}{ck} i \mathbf{k} \times \hat{\mathcal{E}}^n + \frac{1 - C_h}{\epsilon_0 c^2 k^2} i \mathbf{k} \times \left(\hat{\mathcal{J}}^{n+\frac{1}{2}} - \hat{\mathcal{J}}^{n-\frac{1}{2}} \right), \quad (5.33a)$$

$$\hat{\mathcal{E}}^{n+1} - \hat{\mathcal{E}}^n = \frac{2S_h}{k} c i \mathbf{k} \times \hat{\mathcal{B}}^{n+\frac{1}{2}} - \frac{2S_h}{\epsilon_0 c k} \hat{\mathcal{J}}^{n+\frac{1}{2}} + \frac{\mathbf{k}}{\epsilon_0 k^2} \left(\frac{2S_h}{ck} - \Delta t \right) \mathbf{k} \cdot \hat{\mathcal{J}}^{n+\frac{1}{2}}, \quad (5.33b)$$

with $C_h \equiv \cos(ck\Delta t/2)$ and $S_h \equiv \sin(ck\Delta t/2)$ denoting the integration over half a time step. Expanding these coefficients as a Taylor series about $\Delta t = 0$ and keeping only the first order terms, the above expressions reduce to the PSTD equations, Eqs. (5.28a) and (5.28b). Thus, the PSTD solver converges to the PSATD solver in the limit of a vanishing time step, $\Delta t \rightarrow 0$ [45].

Finally, it should again be noted that all equations in this chapter, except for the ones of the FDTD solver, were given assuming a centered grid configuration for the fields. Nevertheless, the pseudospectral equations can be reformulated for a staggered grid by multiplying them with appropriate phase factors in Fourier space that shift the fields from nodal to staggered grid positions in real space [200, 179].

5.3.4 Arbitrary-order pseudospectral solvers

Spectral methods used to be popular in the early decades of particle-in-cell modeling, but were eventually replaced by local finite-difference methods when the parallelization of a PIC code through domain decomposition became mandatory on modern supercomputers [179]. However, it has been shown that domain decomposition is still possible with pseudospectral solvers by using domain-local Fourier transformations and by exploiting the fact that physical field information travels only a finite distance in each time step [45]. Even though this approach introduces errors from truncating the fields at the domain boundaries, these errors can be neglected in practice when using sufficiently large guard regions between the domains.

In addition, it was later shown that by incorporating a finite-difference operator into the spectral Maxwell equations, arbitrary-order pseudospectral solvers can be derived [46]. At the expense of reducing the accuracy of the spatial derivatives to that of a high-order FDTD scheme, the field-solving operation becomes even more local, which allows for thinner boundary layers. Although this makes the solver again prone to spurious numerical dispersion, the order can be tuned to balance accuracy and locality which enables scalability to many compute units [201] while at the same time ensuring correct modeling of the physics [44, 202].

We recall that in Sec. 5.3.1 we introduced the discrete differential operator $\hat{\nabla}$ in Maxwell's equations, which represented centered finite differences of second order on a staggered grid. Here, we follow the same approach and extend $\hat{\nabla}$ to finite differences of arbitrary order n_{order} on a nodal grid. We then use the Fourier space representation of this operator to define a modified wavevector that incorporates spatial locality into the derivatives of the spectral Maxwell equations.

For a scheme with grid-centered fields, the finite-difference operator D applied to a field F at point j_z and along z can be defined as [203]

$$(D_z F)_{j_z} = \sum_{m=1}^n \alpha_{n,m} \frac{F_{j_z+m} - F_{j_z-m}}{2m\Delta z}, \quad (5.34a)$$

$$\alpha_{n,m} = (-1)^{m+1} \frac{2(n!)^2}{(n-m)!(n+m)!}. \quad (5.34b)$$

Equation (5.34a) takes into account the $2n \equiv n_{\text{order}}$ nearest grid points for calculating the spatial derivative. This set of grid points is often called a *finite-difference stencil* with stencil coefficients $\alpha_{n,m}$. The accuracy of this operation is inversely proportional to its locality.

The Fourier space representation of Eq. (5.34a) is defined as [46, 44]

$$(D_z F)_{j_z} = \sum_{k_z} i[k_z] \hat{\mathcal{F}}_{k_z} e^{ik_z j_z \Delta z}, \quad (5.35)$$

with $\hat{\mathcal{F}}$ the Fourier transform of F and where we introduced the modified wavenumber

$$[k_z] = \sum_{m=1}^n \alpha_{n,m} \frac{\sin(k_z m \Delta z)}{m \Delta z}. \quad (5.36)$$

Note that $[k_z] \rightarrow k_z$ for $n_{\text{order}} \rightarrow \infty$.

From Eq. (5.35) it becomes evident that the spatial derivative of finite order n_{order} in real space can be reproduced in Fourier space simply by replacing the wavevector \mathbf{k} with its modified counterpart $[\mathbf{k}]$. With such a modification the PSTD solver [Eqs. (5.28a)

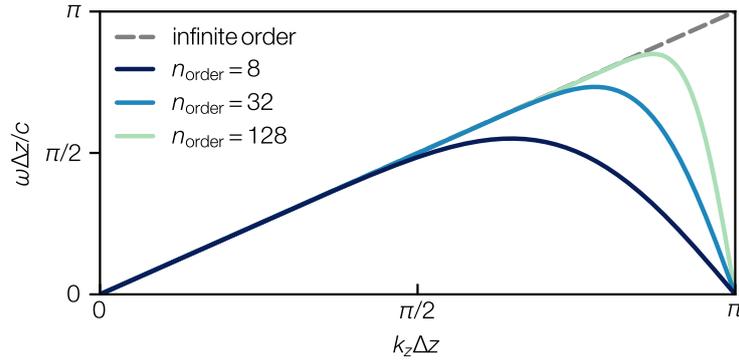


Figure 5.6 – Numerical dispersion relation of the arbitrary-order PSATD solver for waves propagating in a single direction z ($k_x = k_y = 0$) and for a time step $c\Delta t = \Delta z$. The relation is shown for different orders $n_{\text{order}} = [\infty, 8, 32, 128]$.

and (5.28b)] becomes mathematically equivalent to an FDTD scheme of arbitrary order [46]. In fact, using the equivalent of Eq. (5.35) for a staggered grid and with $n_{\text{order}} = 2$, one would obtain the standard FDTD (Yee) solver. As a result, the spectral field-solving operation becomes strictly local on the spatial grid, with an effective reach that equals the finite-difference stencil width.

In case of the PSATD solver, however, the effect of the discrete derivatives on locality is more complex, because the modified wavevector also appears in the coefficients C and S [compare Eqs. (5.31a) and (5.31b)]. As a result, the effective stencil of the PSATD solver is wider than that of a PSTD (or FDTD) solver of the same order.

The numerical dispersion relation of the arbitrary-order PSATD solver is given by

$$\omega^2 = c^2[\mathbf{k}]^2 \quad (5.37)$$

and is shown in Fig. 5.6 for different values of n_{order} in one dimension along the z -axis ($k_x = k_y = 0$). With decreasing order, the solver becomes more local but the amount of spurious numerical dispersion also increases. This lowers the phase velocity $v_{\phi,z} = \omega/[k_z]$ of modes with large k_z in a similar way as for the standard FDTD solver. In practice, one can freely choose the order of the solver to match the requirements in accuracy for a given application. While some applications can require very high-order stencils [202], it was shown in [44] and is shown again in Ch. 8 that relatively low orders are often sufficient to correctly model the physics of plasma accelerators.

The (arbitrary-order) PSATD solver forms the basis of the more advanced algorithms presented in the next chapters. In Ch. 6 we review its formulation in cylindrical geometry with azimuthal mode decomposition and in Ch. 7 we present its derivation in Galilean (comoving) coordinates. In the course of this, the locality of the solver for a given order n_{order} is studied in detail by numerical evaluation and in simulations.

5.4 Numerical Cherenkov

Numerical Cherenkov refers to a class of instabilities that arise from the unphysical coupling of relativistic particles to numerically distorted electromagnetic modes and aliases thereof.

One usually distinguishes between *numerical Cherenkov radiation* (NCR) and a *numerical Cherenkov instability* (NCI). However, both originate from the same underlying numerical issues, which were first described in [194, 54]. NCR usually refers to an effect where the emission of artificial radiation introduces errors in simulations of relativistic beams. It is furthermore often labeled a zero-order NCI effect. The term NCI is more general, but often refers specifically to a virulent instability that can develop in simulations of relativistic plasmas. In the following, we discuss the origin of these instabilities and their impact on simulations of plasma acceleration.

5.4.1 Numerical Cherenkov radiation

As explained in the previous sections, the presence of spurious dispersion causes waves of high spatial frequency to slow down and travel at a phase velocity that is smaller than the vacuum speed of light. A resonant coupling of these slower electromagnetic modes to relativistic particles results in the excitation of unphysical radiation. In analogy to its physical counterpart, this artificial radiation is commonly called numerical Cherenkov radiation.

NCR is a widespread problem in simulations of plasma acceleration using the standard FDTD solver [42]. However, it can be entirely mitigated by using the infinite-order PSATD solver and significantly suppressed by using the arbitrary-order PSATD solver with a moderately high order [44].

In the following, we discuss the effects of NCR by comparing a low-order PSATD solver ($n_{\text{order}} = 4$) to the dispersion-less one ($n_{\text{order}} = \infty$). For simplicity, we consider a 2D grid ($k_y = 0$) and modify the wavevector only in the longitudinal (propagation) direction z . In this case, the numerical dispersion relation of the arbitrary-order PSATD solver, Eq. (5.37), reduces to

$$\omega^2 = c^2(k_x^2 + [k_z]^2). \quad (5.38)$$

Frequencies supported by relativistic particles traveling at a velocity v_0 along z can be described by an artificial *beam mode* $v_0 k_z$. NCR is predominantly emitted at intersections of this beam mode with the distorted electromagnetic mode, i.e., whenever the main NCI resonance condition

$$\omega - v_0 k_z = 0 \quad (5.39)$$

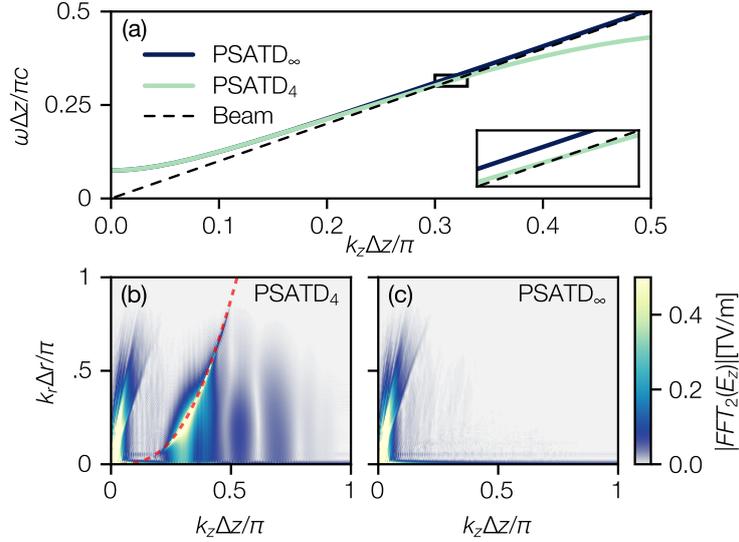


Figure 5.7 – Emission of numerical Cherenkov radiation. (a) Numerical dispersion relation of the (finite-order) PSATD solver for $k_x = (\pi/4)\Delta x^{-1}$ (solid lines); Spurious beam mode ($\gamma_{\text{beam}} = 50$) (gray dashed line). The inset shows the intersection of the beam mode with the distorted electromagnetic mode of the PSATD₄ solver (NCR resonance); Fourier transformed longitudinal electric field $|\text{FFT}_{2\text{D}}(E_z)|$ of a quasi-cylindrical simulation that contains a 25 MeV beam using (a) the PSATD₄ solver and (b) the PSATD_∞ solver; Mode intersection as a function of (k_z, k_x) (magenta dashed line). The figure reproduces results originally published in [44].

is fulfilled. The natural dispersion relation would never cross the beam mode as physical waves in vacuum obey $\omega = ck > k_z v_0$. In the presence of spurious numerical dispersion, however, the unphysical resonance condition [Eq. (5.39)] is fulfilled for all wavevectors (k_x, k_z) that satisfy

$$\frac{[k]}{k_z} = \frac{v_0}{c}. \quad (5.40)$$

Figure 5.7 shows the intersection of the numerical dispersion relation of the PSATD₄ solver with a beam mode ($\gamma_{\text{beam}} = 50$) for (a) a fixed transverse wavenumber $k_x = (\pi/4)\Delta x^{-1}$ and (b) as a function of k_x and k_z , i.e., it shows all possible solutions of Eq. (5.40). In addition, the Fourier-transformed longitudinal electric field $|\text{FFT}_{2\text{D}}(E_z)|$ of a plasma acceleration simulation that contains a 25 MeV beam is shown in Fig. 5.7(b) and (c) for the PSATD₄ and PSATD_∞ solver, respectively. The analysis visualizes the emission of artificial radiation at the main NCR resonances, Eq. (5.40), when using the low-order PSATD₄ solver. When this radiation acts back on the particles, it can alter the beam’s phase space and lead to a deterioration of the simulated beam quality [42, 44].

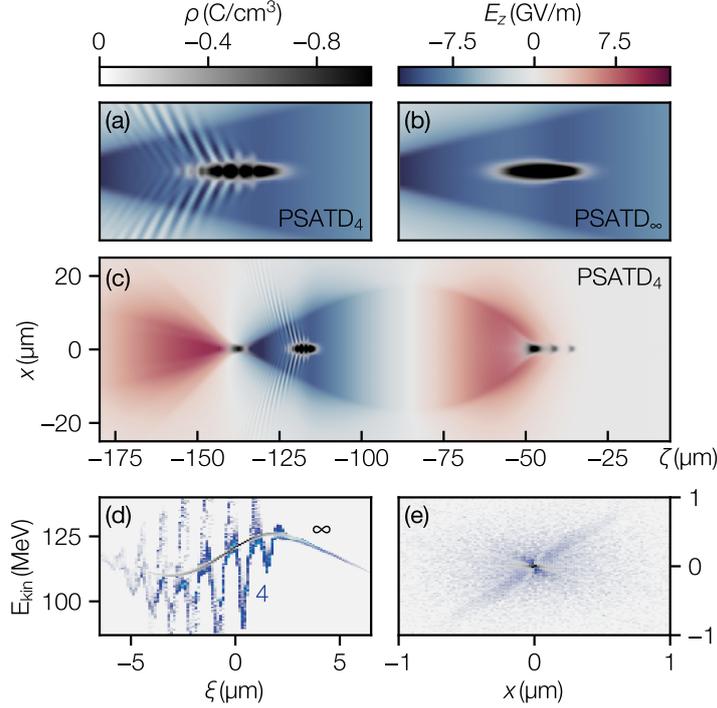


Figure 5.8 – Degradation of the beam quality due to numerical Cherenkov radiation in a simulation of plasma acceleration. (a)-(c) 2D snapshot of the longitudinal electric field E_z and the charge density ρ showing the beam-driven wakefield. Close-up comparison of the results obtained with (a) the PSATD₄ and (b) the PSATD_∞ solver in the vicinity of the accelerated beam. Comparison of the corresponding (d) longitudinal (ξ , E_{kin}) and (e) transverse (x , $p_x/m_e c$) phase spaces. The figure reproduces results originally published in [44].

To give a practical example, we reproduced FBPIC simulations of a beam-driven plasma accelerator that were originally published in [44], with a setup as follows. A 180 pC, 1 GeV energy electron beam of 8.4 μm rms length and 5 μm rms width propagates through a plasma of density $1 \times 10^{17} \text{ cm}^{-3}$ and excites a nonlinear plasma wave in the blowout regime. A second, 50 pC, 25 MeV energy beam of 2 μm rms length and 1 μm rms width is positioned in the back of the driver's wakefield and is accelerated over a distance of a few tens of mm. Both beams initially have Gaussian charge density profiles, zero emittance and no energy spread. The grid resolution is $\Delta z^{-1} = 5 \mu\text{m}^{-1}$ in the longitudinal direction and $\Delta r^{-1} = 1.5 \mu\text{m}^{-1}$ in the radial direction, with 16 macroparticles per cell.

Figure 5.8(c) shows the plasma wave after 10 mm of propagation, together with a close-up comparison between the low-order and infinite-order solver in the area of the accelerated beam, (a) and (b). In the PSATD₄ case, NCR is clearly visible in the form of high-frequency oscillations of the longitudinal electric field. At this point in time,

the radiation already affected the beam over the entire propagation distance, which strongly distorted its longitudinal and transverse charge density profile.

The phase space of the beam is compared in Fig. 5.8(d) and (e). In both cases, the electrons gain an average energy of ~ 120 MeV and beam loading of the plasma wave imprints a characteristic S-shape onto the longitudinal phase space. In the PSATD₄ case, however, additional high-frequency distortions of the phase space are visible, which are most pronounced in the central region where the current density is the highest. A similar distortion of the phase space is visible in the transverse plane, where the occupied phase space volume is increased from the interaction with NCR.

When compared to the infinite-order case, using $n_{\text{order}} = 4$ results in a significant growth of the (slice) energy spread and of the emittance. In [44], we could furthermore show a similar degradation of beam quality when using the standard FDTD solver. Nevertheless, it was also shown that moderately high orders of the PSATD solver are already sufficient to suppress NCR. For the setup considered here, choosing $n_{\text{order}} = 32$ preserves the beam quality to the level of the infinite-order simulation. For higher resolutions, even lower orders are already sufficient to suppress NCR.

5.4.2 Numerical Cherenkov instability

The presence of NCR in simulations of relativistic beams as described above usually does not lead to an instability. One reason for this is that the excited NCR modes have a group velocity that is smaller than the propagation velocity of the beam. The emitted radiation therefore quickly falls back and propagates out of the region of resonant excitation. Another reason is that the beam is small compared to the simulation box and not charge neutral and therefore does not create enough feedback to cause a fast growing instability [51–53].

This contrasts to simulations of relativistic plasmas which can suffer from an exponentially growing numerical Cherenkov instability [54, 55, 204, 205]. More importantly, even if the problem of NCR is mitigated, for example by using the infinite-order PSATD solver, an NCI can still develop through the resonant and non-resonant coupling of spatial and temporal aliases of the electromagnetic and the spurious beam mode.

The occurrence of an NCI can severely limit the applicability of the PIC method for many important applications. In astrophysics, for example, this includes simulations of relativistic collisionless shocks in counterstreaming plasmas as a potential source of high energy particles [206–208]. In accelerator physics, the modeling of relativistically drifting plasmas is required when simulating plasma accelerators in an optimal relativistic frame of reference [49, 209–211]. The *Lorentz-boosted frame technique* [49] allows for orders of magnitude reductions in run time, but its widespread adoption has been prevented due to a lack of satisfying solutions to eliminate the NCI.

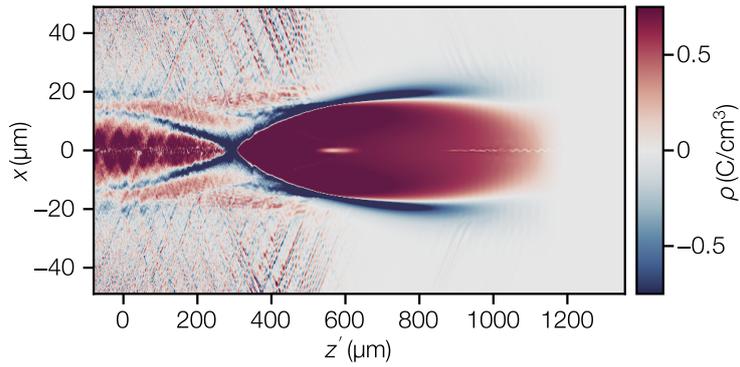


Figure 5.9 – Occurrence of a numerical Cherenkov instability in a Lorentz-boosted frame simulation of laser-plasma acceleration using the standard PSATD solver. High-frequency signatures of the instability are visible in the charge density ρ , on the left side of the simulation box.

Figure 5.9 shows an example where an NCI develops in a Lorentz-boosted frame simulation of a laser-plasma accelerator. In this simulation, the laser propagates to the right and the plasma streams relativistically towards the left edge of the box. Although a dispersion-less PSATD solver of infinite order is used, a strong instability is visible that corrupts the physical results of the simulation.

A novel method that intrinsically eliminates the numerical Cherenkov instability in simulations of relativistically streaming plasmas is presented in Ch. 7 and its application for Lorentz-boosted frame simulations in Ch. 8, respectively.

6 FBPIC – A spectral, quasi-3D particle-in-cell code

This chapter describes the spectral, quasi-3D particle-in-cell code FBPIC whose development this work contributed to as part of a larger collaborative effort [50]. An overview of the algorithm that was originally published in [43] is given in Sec. 6.1. Section 6.2 presents an up-to-date description of the numerical implementation, including several unpublished improvements [212–217]. Section 6.3 describes the parallelization of the code on modern high-performance computing infrastructure, which builds on previous work [218, 43, 219, 220].

6.1 Spectral quasi-cylindrical algorithm

Particle-in-cell algorithms are among the most computationally demanding numerical tools in the field of plasma physics. Therefore, one often seeks solutions to reduce the dimensionality of a problem in order to decrease the computational cost associated with it.

The physics of laser-plasma accelerators is three-dimensional by nature. This prevents the use of a 2D cylindrical geometry or a slab-like configuration in 2D Cartesian coordinates. Consequently, simulations are usually performed in 3D Cartesian coordinates, which can be impractically expensive.

In many cases, however, the physics is at least *close-to-cylindrically* symmetric. The most prominent example is a laser-driven plasma wave where the linear polarization of the drive pulse introduces an azimuthal dependence that only partially breaks the otherwise perfectly cylindrical symmetry.

It has been shown that such a configuration can be efficiently modeled by decomposing the fields into a Fourier series of 2D azimuthal modes in cylindrical coordinates [221, 47, 48]. Most importantly, this series can be truncated after only a few modes while still capturing the dominant 3D physics. This then drastically reduces the computational cost of an otherwise full 3D simulation to that of a few 2D simulations.

The original formulations [47, 48] of such *quasi-cylindrical* or *quasi-3D* algorithms are based on the standard FDTD Maxwell solver and therefore inherit all of its

disadvantages, namely the presence of spurious numerical dispersion, the susceptibility to numerical Cherenkov and the space-time staggering of the fields.

The quasi-cylindrical approach was later extended to work with pseudospectral solvers by using a Fourier-Hankel representation of the fields [43]. This allowed for the derivation of a spectral quasi-cylindrical particle-in-cell algorithm that retains the many advantages of the PSATD solver and combines them with the computational efficiency of the reduced geometry. The solver correctly reproduces the physical dispersion in the longitudinal and transverse direction, it is free of numerical Cherenkov radiation and the discretized fields are centered in space and time.

Moreover, we later extended the solver to arbitrary-order which enables parallelization by domain decomposition [45, 44, 218] and combined it with the Galilean technique to eliminate the numerical Cherenkov instability in Lorentz-boosted frame simulations [51–53]. The Galilean-PSATD solver will be presented in Ch. 7.

All of these developments form the basis of the open-source particle-in-cell code FBPIC [43, 50]. Together with the original developer, Rémi Lehe, this code was continuously co-developed as part of this thesis work. In the following, we give an overview of the algorithm, its practical implementation and the parallelization on modern high-performance computing architectures. In Ch. 8, we furthermore present a comparison between FBPIC and a traditional 3D FDTD algorithm for a typical use case of laser-plasma acceleration, in order to showcase the numerical accuracy and computational efficiency of the code.

6.1.1 Fourier-Hankel representation

In the previous chapter, we considered Maxwell’s equations in Cartesian coordinates, $\mathbf{x} = (x, y, z)$, and showed that they can be solved in Fourier space by representing the fields as a sum of Fourier modes. In cylindrical coordinates, $\mathbf{r} = (r, \theta, z)$, however, the individual modes no longer decouple when inserting this Fourier representation into Maxwell’s equations.

Instead, a *Fourier-Hankel* representation of the fields has to be used [43]. More precisely, in this Fourier-Hankel representation, a scalar field $S(\mathbf{r})$ is expressed as a sum of Fourier-Bessel modes [43]

$$S(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk_z \int_0^{\infty} \frac{k_{\perp} dk_{\perp}}{(2\pi)^2} \hat{S}_m(k_{\perp}, k_z) \tilde{J}_m(k_{\perp} r) e^{ik_z z - im\theta}, \quad (6.1)$$

where the spectral component $\hat{\mathcal{S}}_m$ is defined as

$$\hat{\mathcal{S}}_m(k_\perp, k_z) = \int_{-\infty}^{\infty} dz \int_0^{\infty} r dr \int_0^{2\pi} d\theta S(\mathbf{r}) \tilde{j}_m(k_\perp r) e^{im\theta - ik_z z}. \quad (6.2)$$

Similarly, a vector field $\mathbf{V}(\mathbf{r})$ is expressed as

$$V_r(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk_z \int_0^{\infty} \frac{k_\perp dk_\perp}{(2\pi)^2} \left(\hat{\mathcal{V}}_{+,m}(k_\perp, k_z) \tilde{j}_{m+1}(k_\perp r) + \hat{\mathcal{V}}_{-,m}(k_\perp, k_z) \tilde{j}_{m-1}(k_\perp r) \right) e^{ik_z z - im\theta}, \quad (6.3a)$$

$$V_\theta(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk_z \int_0^{\infty} \frac{k_\perp dk_\perp}{(2\pi)^2} i \left(\hat{\mathcal{V}}_{+,m}(k_\perp, k_z) \tilde{j}_{m+1}(k_\perp r) - \hat{\mathcal{V}}_{-,m}(k_\perp, k_z) \tilde{j}_{m-1}(k_\perp r) \right) e^{ik_z z - im\theta}, \quad (6.3b)$$

$$V_z(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk_z \int_0^{\infty} \frac{k_\perp dk_\perp}{(2\pi)^2} \hat{\mathcal{V}}_{z,m}(k_\perp, k_z) \tilde{j}_m(k_\perp r) e^{ik_z z - im\theta}, \quad (6.3c)$$

with the spectral components $\hat{\mathcal{V}}_{+,m}$, $\hat{\mathcal{V}}_{-,m}$ and $\hat{\mathcal{V}}_{z,m}$ given by

$$\hat{\mathcal{V}}_{+,m}(k_\perp, k_z) = \int_{-\infty}^{\infty} dz \int_0^{\infty} r dr \int_0^{2\pi} d\theta \frac{V_r(\mathbf{r}) - iV_\theta(\mathbf{r})}{2} \tilde{j}_{m+1}(k_\perp r) e^{im\theta - ik_z z}, \quad (6.4a)$$

$$\hat{\mathcal{V}}_{-,m}(k_\perp, k_z) = \int_{-\infty}^{\infty} dz \int_0^{\infty} r dr \int_0^{2\pi} d\theta \frac{V_r(\mathbf{r}) + iV_\theta(\mathbf{r})}{2} \tilde{j}_{m-1}(k_\perp r) e^{im\theta - ik_z z}, \quad (6.4b)$$

$$\hat{\mathcal{V}}_{z,m}(k_\perp, k_z) = \int_{-\infty}^{\infty} dz \int_0^{\infty} r dr \int_0^{2\pi} d\theta V_z(\mathbf{r}) \tilde{j}_m(k_\perp r) e^{im\theta - ik_z z}. \quad (6.4c)$$

Here, \tilde{j}_m is the Bessel function of order m , r is the radial distance, θ is the azimuth, z is the axial coordinate and k_\perp and k_z are wave vectors in the transverse and axial direction, respectively. The sum over m is a sum over azimuthal modes. Further note that the relation to Cartesian coordinates is given by $x = r \cos(\theta)$, $y = r \sin(\theta)$ and $k_x = k_\perp \cos(\theta)$, $k_y = k_\perp \sin(\theta)$.

When using this Fourier-Hankel representation for the fields F_r , F_θ and F_z , where F is either E , B or J , and inserting them into Maxwell's equations in cylindrical coordinates

$$\partial_t B_r = -\frac{1}{r} \partial_\theta E_z + \partial_z E_\theta, \quad \frac{1}{c^2} \partial_t E_r = \frac{1}{r} \partial_\theta B_z - \partial_z B_\theta - \mu_0 J_r, \quad (6.5a)$$

$$\partial_t B_\theta = -\partial_z E_r + \partial_r E_z, \quad \frac{1}{c^2} \partial_t E_\theta = \partial_z B_r - \partial_r B_z - \mu_0 J_\theta, \quad (6.5b)$$

$$\partial_t B_z = -\frac{1}{r} \partial_r r E_\theta + \frac{1}{r} \partial_\theta E_r, \quad \frac{1}{c^2} \partial_t E_z = \frac{1}{r} \partial_r r B_\theta - \frac{1}{r} \partial_\theta B_r - \mu_0 J_z, \quad (6.5c)$$

the different modes decouple and one obtains analogous equations for the spectral coefficients $\hat{\mathcal{F}}_{+,m}$, $\hat{\mathcal{F}}_{-,m}$ and $\hat{\mathcal{F}}_{z,m}$, where $\hat{\mathcal{F}}$ is either $\hat{\mathcal{E}}$, $\hat{\mathcal{B}}$ or $\hat{\mathcal{J}}$ [43]:

$$\partial_t \hat{\mathcal{B}}_{+,m} = \frac{ik_{\perp}}{2} \hat{\mathcal{E}}_{z,m} - k_z \hat{\mathcal{E}}_{+,m}, \quad \frac{1}{c^2} \partial_t \hat{\mathcal{E}}_{+,m} = -\frac{ik_{\perp}}{2} \hat{\mathcal{B}}_{z,m} + k_z \hat{\mathcal{B}}_{+,m} - \mu_0 \hat{\mathcal{J}}_{+,m}, \quad (6.6a)$$

$$\partial_t \hat{\mathcal{B}}_{-,m} = \frac{ik_{\perp}}{2} \hat{\mathcal{E}}_{z,m} + k_z \hat{\mathcal{E}}_{-,m}, \quad \frac{1}{c^2} \partial_t \hat{\mathcal{E}}_{-,m} = -\frac{ik_{\perp}}{2} \hat{\mathcal{B}}_{z,m} - k_z \hat{\mathcal{B}}_{-,m} - \mu_0 \hat{\mathcal{J}}_{-,m}, \quad (6.6b)$$

$$\partial_t \hat{\mathcal{B}}_{z,m} = -ik_{\perp} \hat{\mathcal{E}}_{+,m} - ik_{\perp} \hat{\mathcal{E}}_{-,m}, \quad \frac{1}{c^2} \partial_t \hat{\mathcal{E}}_{z,m} = ik_{\perp} \hat{\mathcal{B}}_{+,m} + ik_{\perp} \hat{\mathcal{B}}_{-,m} - \mu_0 \hat{\mathcal{J}}_{z,m}. \quad (6.6c)$$

Note that we used the shorthand notation $\partial_x \equiv \partial/\partial x$. Accordingly, the conservation equations, $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$ and $\nabla \cdot \mathbf{B} = 0$, become

$$k_{\perp}(\hat{\mathcal{E}}_{+,m} - \hat{\mathcal{E}}_{-,m}) + ik_z \hat{\mathcal{E}}_{z,m} = \frac{\hat{\rho}_m}{\epsilon_0}, \quad (6.7)$$

$$k_{\perp}(\hat{\mathcal{B}}_{+,m} - \hat{\mathcal{B}}_{-,m}) + ik_z \hat{\mathcal{B}}_{z,m} = 0 \quad (6.8)$$

and will be preserved provided that the current satisfies the continuity equation $\partial_t \rho + \nabla \cdot \mathbf{J} = 0$ in spectral space

$$\partial_t \hat{\rho}_m + k_{\perp}(\hat{\mathcal{J}}_{+,m} - \hat{\mathcal{J}}_{-,m}) + ik_z \hat{\mathcal{J}}_{z,m} = 0. \quad (6.9)$$

Just like in the Cartesian case (compare Sec. 5.3.3), Eqs. (6.6a) to (6.6c) can be integrated, either by using the PSTD [182] or the PSATD [183] framework, to advance the spectral coefficients in time.

However, the distinctive feature of the quasi-cylindrical formalism is that the sum over m in Eqs. (6.1) and (6.3a) to (6.3c) corresponds to a sum over an infinite set of azimuthal modes of the form $e^{-im\theta}$ that evolve independently of each other. Most importantly, this sum can be truncated for systems that are close-to-cylindrically symmetric, because modes with large $|m|$ tend to have an amplitude that is close or exactly zero in this case [47]. As a consequence, instead of the 3D Cartesian fields $\hat{\mathcal{F}}(k_x, k_y, k_z)$, only a few 2D matrices $\hat{\mathcal{F}}_m(k_{\perp}, k_z)$ are required to represent the complete physics.

The first mode ($m = 0$) represents fields that are independent of θ and higher order modes ($m \geq 1$) those that vary proportional to $\cos(m\theta)$ and $\sin(m\theta)$. For example, the first mode ($m = 0$) is sufficient to represent an idealized plasma wakefield that is driven by a radially symmetric particle beam, whereas one additional mode ($m = 1$) is required if the wave is driven by a linearly polarized laser pulse. This drastically reduces the computational costs of an otherwise full 3D simulation to that of a few 2D simulations.

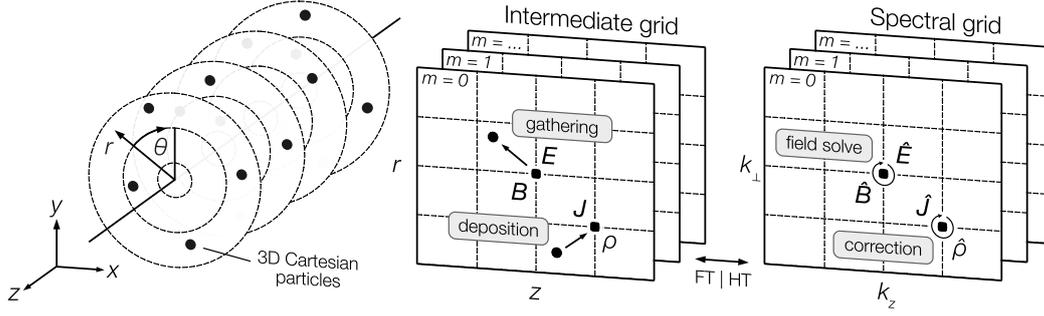


Figure 6.1 – Overview of the FBPIC algorithm. The 3D fields are defined in cylindrical coordinates (r, z, θ) , but are decomposed into a finite number of 2D modes m with azimuthal dependence. The equations of motions are integrated in 3D Cartesian coordinates (x, y, z) . The particles gather from and deposit to intermediate grids in spatial space. The current correction and field solving is performed on a corresponding grid in spectral space. The different representations are connected through a combination of Fourier (FT) and Hankel (HT) transformation.

6.1.2 Overview of the algorithm

The Fourier-Bessel decomposition of the fields from above provides the basis for the complete pseudospectral, quasi-cylindrical particle-in-cell algorithm [43] that is described in the following.

Figure 6.1 shows an overview of the algorithm. Even though the 3D representation of the fields is reduced to m independent 2D azimuthal modes, the macroparticles are still defined in 3D and the equations of motion are integrated in Cartesian (x, y, z) coordinates. Moreover, just like in any other pseudospectral algorithm, the field gathering and deposition is performed on 2D intermediate grids in real space (r, z) . This intermediate 2D representation of the fields mirrors the azimuthal decomposition in spectral space and is related to the 3D cylindrical representation as follows [47, 48]:

$$F_u(\mathbf{r}) = \sum_{m=-\infty}^{\infty} \hat{F}_{u,m}(r, z) e^{-im\theta}, \quad (6.10)$$

$$\hat{F}_{u,m}(r, z) = \frac{1}{2\pi} \int_0^{2\pi} d\theta F_u(\mathbf{r}) e^{im\theta}. \quad (6.11)$$

Here, F_u are the components of the 3D fields E, B or J , where u is either r, θ or z , and $\hat{F}_{u,m}$ is the projection onto the intermediate (r, z) grid of mode m . Note that scalar fields like ρ are treated in the same way as the z -component of a vector field.

The mode-dependent angular factors $e^{im\theta}$ are calculated from the particle positions using the relation $e^{im\theta} = (x + iy)^m / r^m$ [47]. These factors have to be well-resolved by the particle distribution. However, for a low number of modes this requires only a

few macroparticles per radial cell that are (usually randomly) distributed in θ . As a consequence, the total number of particles is significantly reduced compared to that of a conventional 3D simulation. The geometry furthermore implies that macroparticles have larger weights the farther away from the axis they are initialized, because the radial cell volume increases but the number of macroparticles per cell usually does not.

Prior to the integration of Maxwell's equations, each intermediate grid $\hat{F}_{u,m}$ is transformed to spectral space $\hat{\mathcal{F}}_{u,m}$ by applying a Fourier transform (FT) in the longitudinal direction ($z \rightarrow k_z$) followed by a Hankel transform (HT) in the radial direction ($r \rightarrow k_\perp$) to a combination of the intermediate field components $\hat{F}_{u,m}$ [compare Eqs. (6.4a) to (6.4c)]:

$$\hat{\mathcal{F}}_{+,m}(k_\perp, k_z) = \text{HT}_{m+1} \left[\text{FT} \left[\frac{\hat{F}_{r,m}(r, z) - i\hat{F}_{\theta,m}(r, z)}{2} \right] \right], \quad (6.12)$$

$$\hat{\mathcal{F}}_{-,m}(k_\perp, k_z) = \text{HT}_{m-1} \left[\text{FT} \left[\frac{\hat{F}_{r,m}(r, z) + i\hat{F}_{\theta,m}(r, z)}{2} \right] \right], \quad (6.13)$$

$$\hat{\mathcal{F}}_{z,m}(k_\perp, k_z) = \text{HT}_m [\text{FT} [\hat{F}_{z,m}(r, z)]]. \quad (6.14)$$

Correspondingly, to go back to the spatial grid, the inverse transformations read [compare Eqs. (6.3a) to (6.3c)]:

$$\hat{F}_{r,m}(r, z) = \text{IFT} \left[\text{IHT}_{m+1} [\hat{\mathcal{F}}_{+,m}(k_\perp, k_z)] + \text{IHT}_{m-1} [\hat{\mathcal{F}}_{-,m}(k_\perp, k_z)] \right], \quad (6.15a)$$

$$\hat{F}_{\theta,m}(r, z) = i \text{IFT} \left[\text{IHT}_{m+1} [\hat{\mathcal{F}}_{+,m}(k_\perp, k_z)] - \text{IHT}_{m-1} [\hat{\mathcal{F}}_{-,m}(k_\perp, k_z)] \right], \quad (6.15b)$$

$$\hat{F}_{z,m}(r, z) = \text{IFT} [\text{IHT}_m [\hat{\mathcal{F}}_{z,m}(k_\perp, k_z)]]. \quad (6.15c)$$

The Fourier transform (FT) and n -th order Hankel transform (HT_n) and their inverse (IFT, IHT_n) are defined as

$$\text{FT}[f](k_z) \equiv \int_{-\infty}^{\infty} dz e^{-ik_z z} f(z), \quad \text{IFT}[g](z) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} dk_z e^{ik_z z} g(k_z),$$

$$\text{HT}_n[f](k_\perp) \equiv 2\pi \int_0^{\infty} r dr \tilde{j}_n(k_\perp r) f(r), \quad \text{IHT}_n[g](r) \equiv \frac{1}{2\pi} \int_0^{\infty} k_\perp dk_\perp \tilde{j}_n(k_\perp r) g(k_\perp).$$

Note that n is either m for the transformation of the longitudinal component $\hat{\mathcal{F}}_{z,m}$ or $m+1$ or $m-1$ for the transformation of the transverse components $\hat{\mathcal{F}}_{+,m}$ and $\hat{\mathcal{F}}_{-,m}$, respectively.

The particle-in-cell cycle follows the same essential steps as described in Ch. 5 for the PSATD solver. Details on the individual elements of the cycle and the overall numerical implementation of the algorithm are presented in the next section.

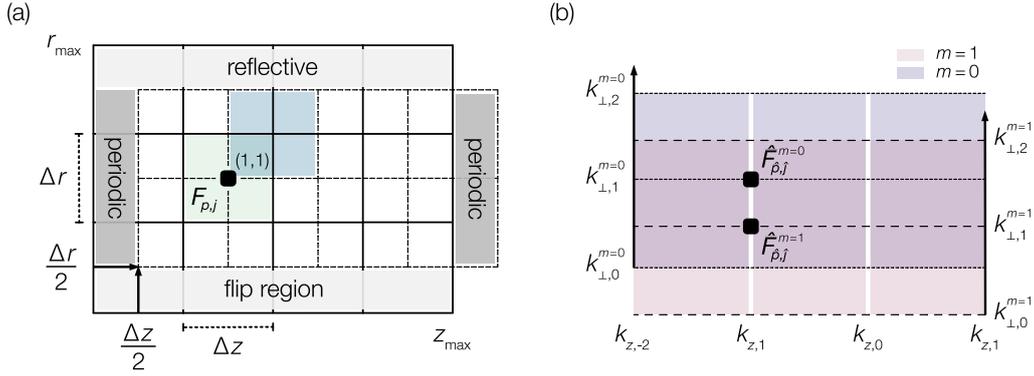


Figure 6.2 – Grid layout and discretization. (a) Intermediate grid discretization in spatial space for a single azimuthal mode m . All fields ($F_{p,j}$) are defined in the center of a *grid cell* (green) or, equivalently, at the lower left corner of a *field cell* (blue). The spectral representation imposes periodic boundaries in the longitudinal direction and reflective boundaries at $r = r_{\max}$. Particles that reside in the first radial half-cell (flip region) are treated differently, as their shape overlaps with a virtual grid point below $r = 0$. (b) Discretization of the spectral components $\hat{\mathcal{F}}_{\hat{p},j}$. The difference in the discretization of the transverse wavenumber is illustrated for modes 0 and 1.

6.2 Numerical implementation

6.2.1 Discretization and transformations

We begin with a description of the discretized transforms and the resulting intermediate and spectral grid layout, which is schematically shown in Fig. 6.2.

The discrete Fourier transform (DFT) in the longitudinal direction uses the 1D fast Fourier transform (FFT) algorithm, which efficiently scales $\propto N_z \log(N_z)$, where N_z is the number of grid points in z . The Fourier series representation imposes a regular spaced discretization of the intermediate grid with periodic boundary conditions. Note that the fields are defined at the center of a grid cell spanning from $z_j - \Delta z/2$ to $z_j + \Delta z/2$. The longitudinal discretization of the fields on the intermediate and spectral grid can be summarized as follows:

$$z_j = \Delta z \left(j + \frac{1}{2} \right) \quad j \in [0, N_z - 1] \quad \text{with} \quad \Delta z = \frac{z_{\max}}{N_z}, \quad (6.16)$$

$$k_{z,\hat{j}} = \hat{j} \frac{2\pi}{\Delta z N_z} \quad \hat{j} \in [-N_z/2, N_z/2 - 1].$$

The discrete Hankel transform (DHT) in the transverse direction is based on a customized quasi-discrete Hankel transform algorithm [222–224, 43], which is a linear

operation on a finite set of sampling points that can be expressed as a matrix M . This matrix has to be determined such that the discrete operation approximates the analytical transform as closely as possible. Because the DHT is a simple matrix multiplication, its computational cost scales $\propto N_r^2$, where N_r is the number of grid points in r .

As opposed to the FFT in the longitudinal direction, there is no universal standard for the discretization of the DHT in the transverse direction. To simplify the gathering and deposition procedures, FBPIC uses a regularly spaced discretization of the intermediate grid that matches the one in the longitudinal direction. Consequently, the fields are also defined at the center of a grid cell spanning from $r_p - \Delta r/2$ to $r_p + \Delta r/2$. The spectral grid, on the other hand, is irregularly spaced with the grid points given by the positive zeros of the Bessel function of order m , α_p^m , where m is the azimuthal mode. The transform is furthermore constrained to the region ($0 < r < r_{\max}$), where we impose the Dirichlet boundary condition $\mathbf{E} = \mathbf{B} = \mathbf{0}$ for $r \geq r_{\max}$. This allows representing the fields as a Fourier-Bessel series over a finite interval [222, 43]. The transverse discretization of the intermediate and spectral grid can be summarized as follows:

$$r_p = \Delta r \left(p + \frac{1}{2} \right) \quad p \in [0, N_r - 1] \quad \text{with} \quad \Delta r = \frac{r_{\max}}{N_r}, \quad (6.17)$$

$$k_{\perp, \hat{p}} = \frac{\alpha_{\hat{p}}^m}{\Delta r N_r} \quad \hat{p} \in [0, N_r - 1],$$

where $\alpha_0^m = 0$ for $m > 0$. Note that the k_{\perp} -discretization differs for each azimuthal mode m , which, however, is not an issue because the spectral components $\hat{\mathcal{F}}_m$ are independent of each other.

Finally, the discrete Hankel transform of order n is defined as

$$\text{DHT}_n^m[f](k_{\perp, \hat{p}}^m) = \sum_{p=0}^{N_r-1} (M_{n,m})_{\hat{p},p} f(r_p), \quad (6.18)$$

$$\text{IDHT}_n^m[g](r_p) = \sum_{\hat{p}=0}^{N_r-1} (M'_{n,m})_{p,\hat{p}} g(k_{\perp, \hat{p}}^m). \quad (6.19)$$

The transformation matrices $M_{n,m}$ are obtained from the analytical Hankel transform of the eigenmodes of a cavity with perfectly conducting boundaries at $r = r_{\max}$. A detailed derivation of these matrices is given in the Appendix of [43]. The original definition of the DHT supported only the first two modes ($m = 0$ and $m = 1$), but was later extended to higher order modes ($m > 1$) [213]. Here, we only summarize the final expression for the inverse matrices $M_{n,m}^{-1}$ and highlight the changes that were introduced by the higher order mode extension.

For the case $n = 0$ or $m = 0$, the inverse matrices $M_{n,m}^{-1}$ are completely determined by the analytical transformation of the eigenmodes and are given by

$$(M_{n,m}^{-1})_{p,\hat{p}} = \frac{\tilde{J}_n(k_{\perp,\hat{p}}^m r_p)}{\pi r_{\max}^2 \left[\tilde{J}_{n+\delta_{n,m}}(\alpha_{\hat{p}}^m) \right]^2}. \quad (6.20)$$

The forward transformation matrices $M_{n,m}$ are then obtained from numerically inverting $M_{n,m}^{-1}$, whereas the inverse transformation matrices $M'_{n,m}$ are set to equal $M_{n,m}^{-1}$. This guarantees that successively applying a DHT and IDHT retrieves the original function.

For the case $n \neq 0$ and $m \neq 0$, an empirical choice has to be made for the $\hat{p} = 0$ entry of the matrix. In the original paper [43], this choice was $(M_{n,m})_{0,p} = 0$ and $(M_{n,m}^{-1})_{p,0} = 0$, whereas in the more recent extension that supports modes with $m > 1$ [213], it is $(M_{n,m}^{-1})_{p,0} = r^{m-1}/(\pi r_{\max}^{m+1})$ for $m \neq 0$ and $n = m - 1$ before the numerical inversion, as well as $(M_{n,m})_{0,p} = 0$ for $m \neq 0$ and $n \neq m - 1$.

The discretization in time is the same as the one described in Ch. 5 for the PSATD solver. The fields \mathbf{E} , \mathbf{B} , ρ and the particle positions \mathbf{x} are defined at integer time steps $n\Delta t$. The current \mathbf{J} and the particle momenta \mathbf{p} are defined at half-integer time steps $(n + 1/2)\Delta t$. Note that in the following, the symbol n is again used exclusively for the iteration number.

6.2.2 Field gathering and particle push

The particle-in-cell cycle starts with the gathering of the \mathbf{E} and \mathbf{B} fields at $t = n\Delta t$. The fields interpolated to the position $\mathbf{r}_\mu = (r_\mu, \theta_\mu, z_\mu)$ of a macroparticle with index μ are given by a sum over all grid cells (p, j) and contributions of each mode m [43]

$$\begin{aligned} F_{u,\text{interp}}(\mathbf{r}_\mu) &= \sum_{p,j} S_{r,p} S_{z,j} \left[\sum_{m=-N_m}^{N_m} \hat{F}_{u,m}(r_p, z_j) e^{-im\theta_\mu} \right] \\ &= \sum_{p,j} S_{r,p} S_{z,j} \left[\hat{F}_{u,0}(r_p, z_j) + 2 \Re \left(\sum_{m=1}^{N_m} \hat{F}_{u,m}(r_p, z_j) e^{-im\theta_\mu} \right) \right], \end{aligned} \quad (6.21)$$

where F is either E or B and u is either r , θ or z , respectively. In the second equality, the relation $\hat{F}_{u,-m}(r, z) = \hat{F}_{u,m}^*(r, z)$ was used to constrain the sum over m to positive values of m . Consequently, only the modes with $m \geq 0$ have to be included in the numerical implementation, as they are sufficient to evaluate the fields and gather the forces acting on the particles [43].

The particle shape factors in the radial and longitudinal direction are given by $S_{r,p} = S(r_\mu - r_p)$ and $S_{z,j} = S(z_\mu - z_j)$, respectively. For linear shapes, an expression of

$S(x)$ is given by Eq. (5.9), but higher order shapes are possible as well. Both, linear and cubic particle shapes are implemented in FBPIC [212]. For simplicity, we restrict ourselves to the former in the following considerations. Further note that, in practice, for a given particle position \mathbf{r}_μ , the sum over p and j is only evaluated for the two nearest cells (linear shapes) or four nearest cells (cubic shapes) along each direction, because the shape factors are zero otherwise.

Special care has to be taken for particles that reside in the first half-cell of the radial grid, i.e., if $r_\mu < r_0 = \Delta r/2$. Strictly speaking, Eq. (6.21) would have to be evaluated for $p = -1$ as the particle shape extends below the axis in this case. However, the value $\hat{F}_{u,m}(r_{-1}, z_j)$ is not part of the grid.

As a solution, we explicitly set $\hat{F}_{u,m}(-\Delta r/2, z_j) = \pm \hat{F}_{u,m}(\Delta r/2, z_j)$, so that the contribution from this cell is either added or removed from the particle. The interpolated field is then either $\hat{F}_{u,m}(\Delta r/2, z_j)$ or zero for a particle that resides exactly on the axis ($r_\mu = 0$). The sign depends on the physical on-axis definition of each field component [47, 43]. For the transverse components, $\hat{F}_{r,m}$ and $\hat{F}_{\theta,m}$, it equals $\text{sgn}[-(-1)^m]$, whereas for the longitudinal component, $\hat{F}_{z,m}$, it equals $\text{sgn}[(-1)^m]$. Note that for cubic shape factors, the same procedure is also applied for the point at $r = -3\Delta r/2$.

If a particle leaves the box at the radial boundary ($r_\mu > r_{\max}$) it stops gathering forces from the grid. In contrast, periodic boundary conditions are applied to the particles in the longitudinal direction. If a particles leaves the simulation box at $z > z_{\max}$ it effectively experiences the fields at $z' = z - \lfloor z/z_{\max} \rfloor \times z_{\max}$, i.e., its position is temporarily shifted by a multiple of N_z cells. Note, however, that particles are removed before leaving the box when using open boundary conditions, as explained in Sec. 6.2.5.

In the final step of the interpolation routine we calculate the transverse Cartesian components from the cylindrical ones using the relations $F_x = \cos(\theta)F_r - \sin(\theta)F_\theta$ and $F_y = \sin(\theta)F_r + \cos(\theta)F_\theta$. These fields are then used to integrate the discretized equations of motion using the Vay algorithm [185] that was already explained in Sec. 5.2.

First, the momenta are pushed from $\mathbf{p}^{n-1/2}$ to the next half time step $\mathbf{p}^{n+1/2}$. The particle positions are then moved from \mathbf{x}^n to $\mathbf{x}^{n+1/2}$ to synchronize with the momenta for the deposition of the currents $\hat{\mathbf{J}}_d^{n+1/2}$. This is followed by a second half-push to \mathbf{x}^{n+1} , which is the starting point for the next iteration and where the charge density $\hat{\rho}^{n+1}$ is projected to the grid. The particularities of the deposition of the source terms are explained in the following.

6.2.3 Charge and current deposition

In line with Sec. 5.2, the expressions for projecting the charge ($\hat{\rho}_m$) and current ($\hat{\mathbf{J}}_{d,m}$) density onto the intermediate grid of mode m are given by [43, 47]

$$\hat{\rho}_m(r_p, z_j) = \frac{1}{V'_p} \sum_{\mu} qw_{\mu} S'(r_p - r_{\mu}) S(z_j - z_{\mu}) e^{im\theta_{\mu}}, \quad (6.22)$$

$$\hat{\mathbf{J}}_{d,u,m}(r_p, z_j) = \frac{1}{V'_p} \sum_{\mu} qw_{\mu} v_{u,\mu} S'(r_p - r_{\mu}) S(z_j - z_{\mu}) e^{im\theta_{\mu}}, \quad (6.23)$$

where u is either r , θ or z , w_{μ} and $v_{u,\mu}$ respectively are the weight and the cylindrical components of the velocity of a macroparticle with index μ , and V'_p is the volume of a cell with radial index p .

In the original formulation of the algorithm [43], V'_p was defined as the geometric cell volume, $V'_p \equiv V_p$, which is the natural choice for calculating the density,

$$V_p = \pi[(p+1)^2 - p^2] \Delta r^2 \Delta z. \quad (6.24)$$

Likewise, the radial shape factors $S'_{r,\mu}$ were set to equal the ones used for the field gathering, i.e., $S'(r) \equiv S(r)$, which is given by Eq. (5.9) for linear particle shapes.

However, we have recently shown [215] that the deposition in the radial direction needs to be slightly adjusted because the former two definitions of V' and S' do not exactly fulfill two fundamental requirements that we impose onto the algorithm:

- (1) The algorithm represents the charge density as a sum of Fourier-Bessel modes, which are not only defined at the grid points, but also in-between them. The integral over this continuous representation expressed as the discrete spectral transform of the deposited charge density should give the total charge of all macroparticles [215].
- (2) It is well known that when projecting onto cylindrical meshes, the density associated with a particle depends not only on the cell volume but also on the radial position within each cell. This should be taken into account by the deposition function, so that a radially uniform distribution of macroparticles also produces a uniform density when deposited onto the intermediate grid [225–227].

Without going into the details, we briefly summarize how to satisfy these two requirements. To meet the first requirement, we write the total charge as

$$Q = \int_0^{z_{\max}} dz \int_0^{r_{\max}} 2\pi r dr \rho_{m=0}(r, z). \quad (6.25)$$

where $\rho_{m=0}(r, z)$ is the continuous representation of the charge density of the first mode. When expressing $\rho(r, z)$ as the aforementioned sum over Fourier-Bessel modes and replacing the spectral components $\hat{\rho}(k_{\perp, \hat{p}}, k_{z, j})$ by the definition of the discrete Hankel transform [Eq. (6.18)] of the intermediate charge density $\hat{\rho}(r_p, z_j)$, we get an expression for the total deposited charge Q as represented in spectral space [215]

$$Q = \sum_{\mu} qw_{\mu} \sum_{j=0}^{N_z-1} \sum_{p=0}^{N_r-1} S'(r_p - r_{\mu}) S(z_j - z_{\mu}) \frac{\Delta z}{V'_p} \sum_{\hat{p}=0}^{N_r-1} (M_{0,0})_{\hat{p}, p} \frac{2}{\alpha_{\hat{p}}^0 \tilde{J}_1(\alpha_{\hat{p}}^0)}. \quad (6.26)$$

By comparing this expression to Eq. (6.22), we see that it only amounts to the real charge $Q = \sum_{\mu} qw_{\mu}$ when replacing V'_p by a modified cell volume [215]

$$V_{p, \text{mod}} = \Delta z \sum_{\hat{p}=0}^{N_r-1} (M_{0,0})_{\hat{p}, p} \frac{2}{\alpha_{\hat{p}}^0 \tilde{J}_1(\alpha_{\hat{p}}^0)}. \quad (6.27)$$

Consequently, we choose $V'_p \equiv V_{p, \text{mod}}$ in Eq. (6.22) when depositing the charge onto the intermediate grid of mode $m = 0$ and $V'_p \equiv V_p$ otherwise. This guarantees that the total charge is correctly represented in spectral space. Note that the same modification is used in the calculation of the current density.

We can meet the second requirement by adding a cell-dependent *Ruyten* correction [225] to the particle shape factors without changing their basic properties,

$$S_p(r) = S(r) + C_p(r), \quad (6.28)$$

where, for our grid layout, the correction factor C_p depends on the *field cell* that the particle resides in and is therefore given by

$$C_p(r) = \frac{|r|}{\Delta r} \left(1 - \frac{|r|}{\Delta r}\right) \times \begin{cases} \beta_p & \text{if } r \in [r_p, r_{p+1}], \\ -\beta_{p-1} & \text{if } r \in [r_{p-1}, r_p], \\ 0 & \text{otherwise.} \end{cases} \quad (6.29)$$

In order to preserve a uniform macroparticle density, it can be shown [215] that the Ruyten correction coefficients β_p for linear particle shapes have to be defined as

$$\beta_p^1 = \frac{6}{p+1} \left(\sum_{\ell=0}^p \frac{V'_{\ell}}{2\pi\Delta z\Delta r^2} - \frac{(p+1)^2}{2} - \frac{1}{24} \right), \quad (6.30)$$

where V'_l is either the geometric volume V_l [Eq. (6.24)] for $m \geq 1$, or the modified volume $V_{l, \text{mod}}$ [Eq. (6.27)] for $m = 0$. In summary, fulfilling the second requirement requires the use of $S'(r) \equiv S_p(r)$ in Eq. (6.22).

The addition or subtraction of contributions deposited below the axis, by particles that are located in the first half-cell of the grid, follows exactly the same rules as defined for the field gathering. Here again, the scalar field ρ behaves like the axial component of a vector field. In addition, the Ruyten correction factor is explicitly set to $\beta_{-1} = 0$ for particles that are below $r_0 = \Delta r/2$.

To suppress the accumulation of high frequency noise, a filter is applied to the deposited charge $\hat{\rho}$ and current $\hat{\mathbf{J}}_d$ densities after transforming them to spectral space ($\hat{\rho}$, $\hat{\mathbf{J}}_d$). As explained in Sec. 5.2.2, this can be done by multiplying the fields with a transfer function $\hat{\mathcal{T}}(k_\perp, k_z) = \hat{\mathcal{T}}(k_\perp) \times \hat{\mathcal{T}}(k_z)$, which in FBPIC defaults to a single-pass binomial smoother without a compensator [Eq. (5.20)].

Eventually, the filtered current is further modified to satisfy the continuity equation in spectral space [Eq. (6.9)], by using the curl-free correction [Eq. (5.18a)] [45, 43] that was introduced in Sec. 5.2.2,

$$\hat{\mathcal{J}}_{+,m}^{n+\frac{1}{2}} = \hat{\mathcal{J}}_{d,+,m}^{n+\frac{1}{2}} - \frac{k_\perp}{2k^2} \hat{\mathcal{G}}_m^{n+\frac{1}{2}}, \quad (6.31a)$$

$$\hat{\mathcal{J}}_{-,m}^{n+\frac{1}{2}} = \hat{\mathcal{J}}_{d,-,m}^{n+\frac{1}{2}} + \frac{k_\perp}{2k^2} \hat{\mathcal{G}}_m^{n+\frac{1}{2}}, \quad (6.31b)$$

$$\hat{\mathcal{J}}_{z,m}^{n+\frac{1}{2}} = \hat{\mathcal{J}}_{d,z,m}^{n+\frac{1}{2}} + \frac{ik_z}{k^2} \hat{\mathcal{G}}_m^{n+\frac{1}{2}}, \quad (6.31c)$$

where $k = \sqrt{k_\perp^2 + k_z^2}$ and

$$\hat{\mathcal{G}}_m^{n+\frac{1}{2}} = \frac{\hat{\mathcal{Q}}_m^{n+1} - \hat{\mathcal{Q}}_m^n}{\Delta t} + k_\perp (\hat{\mathcal{J}}_{d,+,m}^{n+\frac{1}{2}} - \hat{\mathcal{J}}_{d,-,m}^{n+\frac{1}{2}}) + ik_z \hat{\mathcal{J}}_{d,z,m}^{n+\frac{1}{2}}. \quad (6.32)$$

As an alternative to the curl-free correction from above, we recently developed [216] the so-called *cross-deposition* correction, which is inspired by the principles of the Esirkepov deposition for Cartesian solvers [188, 45]. It uses the directional change of the charge density along the longitudinal and transverse axes to correct the respective components of the current density independently of each other. As a consequence, the spatial locality of the current correction becomes comparable to the one of the Maxwell solver, which can be of advantage when parallelizing the algorithm via domain decomposition. A detailed presentation of this new correction will be given elsewhere.

6.2.4 Quasi-cylindrical pseudospectral analytical time domain

In the final step of the particle-in-cell cycle, the updated source terms $\hat{\mathcal{J}}^{n+\frac{1}{2}}$ and $\hat{\rho}^{n+1}$ are used to advance the electromagnetic fields with the PSATD solver. The updated spectral components, $\hat{\mathcal{E}}^{n+1}$ and $\hat{\mathcal{B}}^{n+1}$, are then transformed back to real space for the next iteration.

Table 6.1 – Formal analogy between the spectral Cartesian and spectral quasi-cylindrical representation of the differential operators [52].

Operator	Spectral Cartesian	Spectral cylindrical
Gradient $\mathbf{F} = \nabla S$	$\hat{\mathcal{F}} = i\mathbf{k}\hat{\mathcal{S}}$	
	$\hat{\mathcal{F}}_x = ik_x\hat{\mathcal{S}}$	$\hat{\mathcal{F}}_{+,m} = -k_\perp\hat{\mathcal{S}}_m/2$
	$\hat{\mathcal{F}}_y = ik_y\hat{\mathcal{S}}$	$\hat{\mathcal{F}}_{-,m} = k_\perp\hat{\mathcal{S}}_m/2$
	$\hat{\mathcal{F}}_z = ik_z\hat{\mathcal{S}}$	$\hat{\mathcal{F}}_{z,m} = ik_z\hat{\mathcal{S}}_m$
Curl $\mathbf{F} = \nabla \times \mathbf{V}$	$\hat{\mathcal{F}} = i\mathbf{k} \times \hat{\mathcal{V}}$	
	$\hat{\mathcal{F}}_x = ik_y\hat{\mathcal{V}}_z - ik_z\hat{\mathcal{V}}_y$	$\hat{\mathcal{F}}_{+,m} = k_z\hat{\mathcal{V}}_{+,m} - ik_\perp\hat{\mathcal{V}}_{z,m}/2$
	$\hat{\mathcal{F}}_y = ik_z\hat{\mathcal{V}}_x - ik_x\hat{\mathcal{V}}_z$	$\hat{\mathcal{F}}_{-,m} = -k_z\hat{\mathcal{V}}_{-,m} - ik_\perp\hat{\mathcal{V}}_{z,m}/2$
	$\hat{\mathcal{F}}_z = ik_x\hat{\mathcal{V}}_y - ik_y\hat{\mathcal{V}}_x$	$\hat{\mathcal{F}}_{z,m} = ik_\perp\hat{\mathcal{V}}_{+,m} + ik_\perp\hat{\mathcal{V}}_{-,m}$
Divergence $\mathbf{F} = \nabla \cdot \mathbf{V}$	$\hat{\mathcal{F}} = i\mathbf{k} \cdot \hat{\mathcal{V}}$	
	$\hat{\mathcal{F}} = ik_x\hat{\mathcal{V}}_x + ik_y\hat{\mathcal{V}}_y + ik_z\hat{\mathcal{V}}_z$	$\hat{\mathcal{F}}_m = k_\perp(\hat{\mathcal{V}}_{+,m} - \hat{\mathcal{V}}_{-,m}) + ik_z\hat{\mathcal{V}}_{z,m}$

The derivation of the quasi-cylindrical PSTD and PSATD equations from first principles is explained in detail in [43]. In [52] it was furthermore shown that expressions for the quasi-cylindrical version of these solvers can be readily derived from their Cartesian counterpart, by taking advantage of a formal analogy between the spectral representation of the differential operators in both geometries.

Following this methodology, which is summarized in Table 6.1, we can derive the quasi-cylindrical (arbitrary-order) PSATD scheme from the Cartesian scheme by replacing the differential operators in Eqs. (5.31a) and (5.31b):

$$\begin{aligned} \hat{\mathcal{B}}_{+,m}^{n+1} = C\hat{\mathcal{B}}_{+,m}^n - \frac{S}{c[k]} \left(-\frac{ik_\perp}{2}\hat{\mathcal{E}}_{z,m}^n + [k_z]\hat{\mathcal{E}}_{+,m}^n \right) \\ + \frac{1-C}{\epsilon_0 c^2 [k]^2} \left(-\frac{ik_\perp}{2}\hat{\mathcal{J}}_{z,m}^{n+\frac{1}{2}} + [k_z]\hat{\mathcal{J}}_{+,m}^{n+\frac{1}{2}} \right), \end{aligned} \quad (6.33a)$$

$$\begin{aligned} \hat{\mathcal{B}}_{-,m}^{n+1} = C\hat{\mathcal{B}}_{-,m}^n - \frac{S}{c[k]} \left(-\frac{ik_\perp}{2}\hat{\mathcal{E}}_{z,m}^n - [k_z]\hat{\mathcal{E}}_{-,m}^n \right) \\ + \frac{1-C}{\epsilon_0 c^2 [k]^2} \left(-\frac{ik_\perp}{2}\hat{\mathcal{J}}_{z,m}^{n+\frac{1}{2}} - [k_z]\hat{\mathcal{J}}_{-,m}^{n+\frac{1}{2}} \right), \end{aligned} \quad (6.33b)$$

$$\begin{aligned} \hat{\mathcal{B}}_{z,m}^{n+1} = C\hat{\mathcal{B}}_{z,m}^n - \frac{S}{c[k]} \left(ik_\perp\hat{\mathcal{E}}_{+,m}^n + ik_\perp\hat{\mathcal{E}}_{-,m}^n \right) \\ + \frac{1-C}{\epsilon_0 c^2 [k]^2} \left(ik_\perp\hat{\mathcal{J}}_{+,m}^{n+\frac{1}{2}} + ik_\perp\hat{\mathcal{J}}_{-,m}^{n+\frac{1}{2}} \right), \end{aligned} \quad (6.33c)$$

$$\begin{aligned}\hat{\mathcal{E}}_{+,m}^{n+1} = & C\hat{\mathcal{E}}_{+,m}^n + c\frac{S}{[k]} \left(-\frac{ik_{\perp}}{2}\hat{\mathcal{B}}_{z,m}^n + [k_z]\hat{\mathcal{B}}_{+,m}^n - \mu_0\hat{\mathcal{J}}_{+,m}^{n+\frac{1}{2}} \right) \\ & + \frac{k_{\perp}}{2\epsilon_0[k]^2} \left[\hat{\varrho}_m^{n+1} \left(1 - \frac{S}{c[k]\Delta t} \right) - \hat{\varrho}_m^n \left(C - \frac{S}{c[k]\Delta t} \right) \right],\end{aligned}\quad (6.34a)$$

$$\begin{aligned}\hat{\mathcal{E}}_{-,m}^{n+1} = & C\hat{\mathcal{E}}_{-,m}^n + c\frac{S}{[k]} \left(-\frac{ik_{\perp}}{2}\hat{\mathcal{E}}_{z,m}^n - [k_z]\hat{\mathcal{B}}_{-,m}^n - \mu_0\hat{\mathcal{J}}_{-,m}^{n+\frac{1}{2}} \right) \\ & - \frac{k_{\perp}}{2\epsilon_0[k]^2} \left[\hat{\varrho}_m^{n+1} \left(1 - \frac{S}{c[k]\Delta t} \right) - \hat{\varrho}_m^n \left(C - \frac{S}{c[k]\Delta t} \right) \right],\end{aligned}\quad (6.34b)$$

$$\begin{aligned}\hat{\mathcal{E}}_{z,m}^{n+1} = & C\hat{\mathcal{E}}_{z,m}^n + c\frac{S}{[k]} \left(ik_{\perp}\hat{\mathcal{B}}_{+,m}^n + ik_{\perp}\hat{\mathcal{B}}_{-,m}^n - \mu_0\hat{\mathcal{J}}_{z,m}^{n+\frac{1}{2}} \right) \\ & - \frac{i[k_z]}{2\epsilon_0[k]^2} \left[\hat{\varrho}_m^{n+1} \left(1 - \frac{S}{c[k]\Delta t} \right) - \hat{\varrho}_m^n \left(C - \frac{S}{c[k]\Delta t} \right) \right],\end{aligned}\quad (6.34c)$$

where $[k] = \sqrt{k_{\perp}^2 + [k_z]^2}$ and where we again used the shorthand notations $C \equiv \cos(c[k]\Delta t)$ and $S \equiv \sin(c[k]\Delta t)$. Here we already replaced the axial wave vector k_z with its modified counterpart $[k_z]$ to represent the spatial derivatives as finite-difference approximations of arbitrary order n_{order} (compare Sec. 5.3.4). Thereby we directly incorporated the needed locality to parallelize the solver via domain decomposition in the longitudinal direction. Note that the modified wavevector $[k_z]$ is also used for the current correction [Eq. (6.31)], but not for the filtering of the deposited source terms [Eq. (5.20)]. Further note that a similar modification of the transverse wave vector k_{\perp} is not possible, because the Hankel transform is an intrinsically non-local operation in the radial direction.

It should again be highlighted that this quasi-cylindrical PSATD solver retains all of the advantages of the Cartesian PSATD solver, i.e., it is dispersion-less in all directions for $n_{\text{order}} = \infty$, it avoids interpolation errors as the electromagnetic fields are centered in space and time, and it is in principle not subject to a CFL limit.

6.2.5 Open boundary conditions and moving window

In Sec. 6.2.1, we had to impose specific boundary conditions in order to be able to represent the fields within the finite simulation box as a discrete series of Fourier–Bessel modes. In practice, these boundary conditions cause electromagnetic waves to periodically wrap around the simulation box in the longitudinal direction and to get reflected at the transverse boundary. However, this is often not the desired behavior, especially not in simulations of plasma acceleration for which open boundary conditions are mandatory in the longitudinal and at least preferable in the transverse direction.

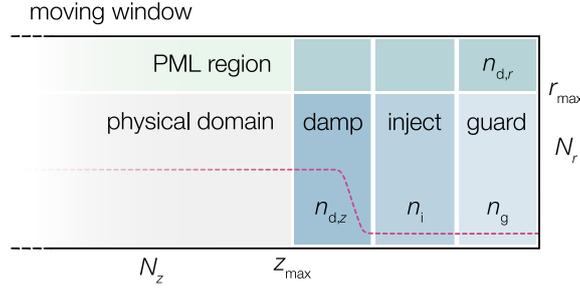


Figure 6.3 – Open boundary conditions and moving window. In the longitudinal direction, the physical domain (N_z cells) is extended by a field damping ($n_{d,z}$ cells), an injection (n_i cells) and a guard (n_g cells) region (blue); Illustration of the damping function (dashed magenta line). In the radial direction, the physical domain (N_r cells) is extended by a perfectly matched layer ($n_{d,r}$ cells) that gradually absorbs incident waves (green).

Nevertheless, alternative boundary conditions can be realized by applying them to the real-space fields on the intermediate grid. Particles are also treated differently in this case. Figure 6.3 illustrates the numerical implementation of open boundary conditions in the longitudinal and transverse direction.

Longitudinal boundary

When choosing open boundary conditions in the longitudinal direction, the grid is extended by $n_g + n_i + n_{d,z}$ cells on both sides. The outer n_g cells serve as a guard region that prevents fields and particles from wrapping around the periodic simulation box. The intermediate grid fields in this region are zeroed at each time step. In addition, particles inside this region are removed from the simulation at regular intervals, at latest when they have traveled halfway into it. The length of the guard region is set to a few tens of cells or is set to match the effective stencil width of the finite-order PSATD solver. The next inner $n_i = \lfloor n_g/2 \rfloor$ cells are used to inject new particles into the simulation. The fields are also zeroed in this region. Finally, waves that propagate out of the physical domain are gradually absorbed in a region of $n_{d,z}$ cells, by multiplying the fields with a damping function at end of each iteration [228, 43]. For example, at the left boundary,

$$\hat{F}_{u,m} = \sin\left(\pi \frac{z - (z_{\min} - n_{d,z}\Delta z)}{n_{d,z}\Delta z}\right)^2 \hat{F}_{u,m}, \quad (6.35)$$

in the interval $(z_{\min} - n_{d,z}\Delta z) \leq z < (z_{\min} - n_{d,z}/2\Delta z)$, where F is E or B and u is either r , θ or z . In practice, a few tens of cells are usually sufficient to properly absorb the fields in this damping region.

Transverse boundary

In the transverse direction, open boundary conditions are realized by adding a perfectly matched layer (PML), which is an artificial layer designed to attenuate and absorb incident waves [229]. Here, we give a brief overview of the spectral quasi-cylindrical PML formulation that is implemented in FBPIC [214].

When adding the PML, the radial grid is enlarged by $n_{d,r} = \Delta r_{\text{pml}}/\Delta r$ cells and four additional transverse PML components, \hat{E}_r^* , \hat{E}_θ^* , \hat{B}_r^* and \hat{B}_θ^* , are added to the physical fields. Just like the regular fields, these components are evolved in spectral space using reduced PSATD equations that only depend on the axial component of the fields [214]

$$\hat{\mathcal{B}}_{u,m}^{*,n+1} = C\hat{\mathcal{B}}_{u,m}^{*,n} + \frac{S}{c[k]} \frac{ik_\perp}{2} \hat{\mathcal{E}}_{z,m}^n, \quad (6.36)$$

$$\hat{\mathcal{E}}_{u,m}^{*,n+1} = C\hat{\mathcal{E}}_{u,m}^{*,n} - c \frac{S}{[k]} \frac{ik_\perp}{2} \hat{\mathcal{B}}_{z,m}^n, \quad (6.37)$$

where u is either $+$ or $-$. Eventually, after transforming the fields back to real space, anisotropic damping is applied in the PML region, $r_{\text{max}} < r \leq r_{\text{max}} + \Delta r_{\text{pml}}$, which has the effect of gradually absorbing waves without reflecting them. More precisely, at each time step, the azimuthal and axial components of the electromagnetic fields are modified according to [214]

$$\hat{F}_{\theta,m} = \hat{F}_{\theta,m} - (1 - e^{-\sigma}) \hat{F}_{\theta,m}^*, \quad (6.38)$$

$$\hat{F}_{z,m} = \hat{F}_{z,m} - (1 - e^{-\sigma}) \hat{F}_{z,m}^*, \quad (6.39)$$

where F is either E or B and where the damping factor σ is given by

$$\sigma = 4 \frac{c\Delta t}{\Delta r} \left(\frac{r - r_{\text{max}}}{\Delta r_{\text{pml}}} \right)^2. \quad (6.40)$$

The effectiveness of the PML increases with its thickness $n_{d,r}$. In practice, a few tens of cells are usually sufficient to prevent physical waves to reflect back towards the axis.

Moving window

The implementation of a *moving window* is yet another feature that is relevant to simulations of plasma acceleration. The technique allows reducing the computational cost by sliding a window across the interaction region, such that the simulation box always only covers the small region of interest that includes the driver and the immediate plasma wave behind it. In the following, we assume this window to follow a plasma wave that travels from left to right, i.e., towards positive z . In this picture, the driver

appears to be static and fresh plasma is continuously added at the right edge of the box, while excited plasma oscillations eventually leave the box at the left boundary.

The numerical realization of the moving window requires to continuously track its progression, starting from an initial position $z_{\text{win}} = z_{\text{max}}$,

$$z_{\text{win}} = z_{\text{win}} + v_{\text{win}}\Delta t, \quad (6.41)$$

where v_{win} is the velocity of the moving window. At each time step, the numerical grid is then virtually translated by an integer amount of cells, by updating the limits of the simulation box:

$$n_{\text{move}} = \lfloor (z_{\text{win}} - z_{\text{max}})/\Delta z \rfloor, \quad (6.42)$$

$$z_{\text{max}} = z_{\text{max}} + n_{\text{move}}\Delta z, \quad (6.43)$$

$$z_{\text{min}} = z_{\text{min}} + n_{\text{move}}\Delta z. \quad (6.44)$$

This automatically guarantees that particles are removed in the guard region at the left edge of the simulation box. In addition, new particles are continuously added in the injection region at the right side of the box.

Finally, the fields are shifted backwards whenever the box limits have been shifted forward. This operation is done in spectral space, where shifting the fields by n_{move} cells on the spatial grid is a simple multiplication by a phase factor,

$$\hat{\mathcal{F}}'_{u,m} = \hat{\mathcal{F}}_{u,m} \left(e^{ik_z\Delta z} \right)^{n_{\text{move}}}. \quad (6.45)$$

Note that this operation uses the true (unmodified) wavevector k_z . The moving window is handled right after the PSATD push and before damping the fields, which is the final step of the particle-in-cell cycle.

6.2.6 Advanced features

Apart from the core elements of the particle-in-cell cycle, a variety of additional features [217] have been implemented in FBPIC. These features complement the core functionalities of the code, with a clear focus on its ability to efficiently simulate the physics of plasma acceleration.

We provide functions to initialize arbitrary types of laser pulses, particle beams or external electromagnetic fields. We support pre-defined lasers with Gaussian intensity profiles with a temporal chirp or with a length of only a few cycles and those with more complex transverse intensity profiles that can be expressed as a combination of Laguerre-Gauss modes. In particular, we implement a flattened Gaussian laser model [124], which accurately reproduces the focal intensity of the transversely flat-top

profiles of modern high-power lasers. Lasers can either be initialized directly on the grid or can be emitted from a (moving) antenna plane.

Besides, the code supports the injection of particle beams with arbitrary density and velocity distributions, while ensuring a correct calculation of the initial space charge fields and with the option to inject the particles through a plane. Pre-defined functions for Gaussian particle distributions are provided, as well as helper functions to read 6D phase spaces from a file.

The code furthermore supports two essential elementary particle processes, namely optical field ionization of atoms based on the Ammosov-Delone-Krainov (ADK) [102, 99] model and inverse Compton scattering of photons based on the Klein-Nishina formula [230]. Notably, ADK ionization rates are evaluated instantaneously (unaveraged) and in the rest frame of a particle (Lorentz-invariant).

Output of simulation results adheres to the OpenPMD standard [231]. The code allows to restart a simulation from checkpoints and supports particle tracking for trajectory reconstruction. Standardized input of simulation cases across different particle-in-cell codes is made possible by supporting the PICMI interface [232].

Finally, all of the features support the transformation of the simulation from and to different relativistic frames of reference, which is an essential pre-requisite for the application of the Lorentz-boosted frame technique (see Sec. 8.1.1) [49].

6.2.7 Practical implementation

FBPIC is written entirely in Python, but follows several optimization strategies to reach a performance that is comparable to pre-compiled C code. Python is a dynamically-typed and interpreted language, which simplifies the implementation and debugging of the algorithm. The flexibility and readability of the language facilitates rapid prototyping and empowers users to quickly adapt or add new features to the code. In addition, Python allows us to expose an easy-to-use API and provide an object-oriented environment to setup a simulation with pre-defined physical objects, such as particle beams, laser pulses or plasmas.

However, pure Python code is converted to platform-independent bytecode that is executed by a virtual machine rather than directly by the processor. This can significantly limit the achievable performance of compute-intense programs. FBPIC overcomes this limitation by making use of pre-compiled libraries, such as Intel MKL [233] or FFTW [234], and Just-In-Time (JIT) compilation to generate native machine code at run time. The latter functionality is provided by the Numba package [235] that is based on the LLVM compiler infrastructure [236].

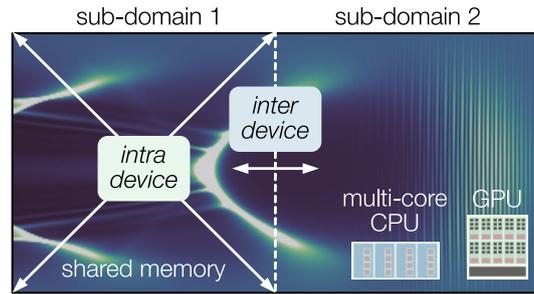


Figure 6.4 – Illustration of the two-level parallelization of FBPIC. On a lower level, the particle-in-cell routines are directly parallelized on a multi-CPU or GPU (intra-device communication through shared memory). On a higher level, the simulation box is split into sub-domains that exchange particle and field information at the boundaries (inter-device communication across distributed memory).

Moreover, the execution of the code is parallelized on two levels. On a first level, the algorithm is optimized to run efficiently on highly parallel hardware, in particular on CUDA-enabled GPUs and multi-core CPUs. The parallel implementation is based on advanced features provided by the Numba package. The GPU implementation additionally uses the CUDA libraries cuFFT, cuBLAS [237] and Thrust [238], that are exposed by the cupy [239] package. On a second level, the computation is scaled to distributed compute nodes via domain decomposition. The communication is handled by MPI (Message Passing Interface) [240] using the mpi4py package [241]. Details on the parallelization of the algorithm will be given in the next section.

6.3 Parallelization

Particle-in-cell simulations can be extremely demanding in terms of arithmetic operations and memory consumption. Typical simulations of plasma acceleration comprise of millions of cells and macroparticles that have to be evolved in time for millions of iterations. Therefore, the execution of a code is usually parallelized across many compute units that share the computational work.

As we have already noted in Sec. 5.3, a common strategy to parallelize a particle-in-cell algorithm is to spatially decompose the simulation box into multiple sub-domains:

- (1) Each sub-domain can be treated by a separate compute device. Minimal information at the boundaries is exchanged between neighboring sub-domains.
- (2) Within each sub-domain, the multiple cores of a single compute device can share the work by executing parallel versions of the particle-in-cell routines.

This two-level approach is schematically shown in Fig. 6.4. The lower level relies on the fast exchange of data through *shared memory* and is referred to as *intra-device* parallelization. The upper level takes into account the slower exchange of data across *distributed memory* and is referred to as *inter-device* parallelization. The implementation of these two parallelization levels in FBPIC builds on previous work [218] and is described in the following.

6.3.1 Intra-device parallelization

A single parallel compute device can be a multi-core CPU, GPU or other co-processor that consists of many cores. Parallelization on such a device is based on the simultaneous execution of independent threads that do not interact directly with each other. Communication between threads requires to synchronize their execution and exchange information through the collectively shared memory. Highest performance can be expected when all threads perform the same task but on different parts of the data.

Many steps of the particle-in-cell cycle consist of such *data parallel* operations, for example, when applying the same calculation to all macroparticles or discrete grid points. These operations can be efficiently parallelized without having to share information between threads.

However, some calculations involve multiple threads updating the same value stored in memory. If the read-modify-write cycle of a thread is interleaved with another one, it can happen that the thread reads a version of the value that is already outdated. In the presence of such *race conditions*, locks have to be used to prevent other threads from accessing the value until the currently unlocked thread has finished processing it. Because this violates the concept of data parallelism, it can severely degrade the performance of the computation. Consequently, an algorithm should circumvent possible race conditions, for example, by using data parallel primitives or by creating temporary copies of the data. If race conditions can not be avoided, hardware optimized atomic operations should be used, as they are more efficient than the lock-based procedure described above.

Figure 6.5 illustrates the degree of data parallelism of the individual steps of the particle-in-cell cycle. The particle push (a) can be parallelized very easily, with each thread updating the momentum and position of a single macroparticle. The same is true for operations that act on the discrete elements of the grid, including for example the field solving step (b). The situation becomes more complex when interpolating between grid points and particle positions.

When collecting the fields, a single grid point value is required multiple times for different particles (c). This is, however, less of an issue because individual read accesses are always serialized at the hardware level. Consequently, the field gathering can

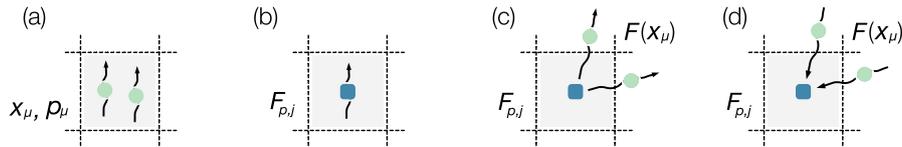


Figure 6.5 – Degree of data parallelism of the individual steps of the particle-in-cell cycle. A single independent thread (black arrow) per element can be used for (a) the particle push or (b) the field solving step. Multiple particles interact with the same grid node during (c) the gathering and (d) the deposition routines. The latter situation causes race conditions when parallelized on a per-particle basis.

also be parallelized on a per-particle basis. Nevertheless, performance may suffer as multiple threads have to wait for each other if they simultaneously try to access the same location in memory.

In contrast, multiple particles deposit to a single grid node when projecting the charge and current densities (d). If parallelized on a per-particle basis, race conditions would occur when concurrently executing threads attempt to update the same value of the field array. Hence, more creative parallelization strategies have to be applied to this method.

The fast Fourier transform and discrete Hankel transform are not shown in Fig. 6.5. These algorithms involve global communication in a sequence of parallel operations, for which pre-optimized libraries exist.

Even though the above considerations have general validity, the details of the parallel implementation depend on the underlying hardware. Up to now, two different intra-device parallelizations of FBPIC have been realized, a multi-threaded CPU implementation [219] and a highly parallel CUDA implementation that runs on Nvidia GPUs [218, 43, 220]. GPUs execute thousands of lightweight threads that access high bandwidth memory. They are especially well suited to accelerate the data parallel operations of a particle-in-cell algorithm. CPU threads can reach a higher performance than a single GPU thread, but only a few of them can be run efficiently at the same time. In the following, we describe the differences of these two implementations in more detail.

CPU implementation

The parallel CPU implementation utilizes the multi-threading features provided by Numba, in particular the possibility to write explicitly parallel for-loops. Elements of these loops are automatically distributed and processed by concurrently running threads. Although any number of threads can be launched, best performances are achieved when the number of threads matches the physical cores of the CPU. Internally,

the threading layer of Numba relies on the highly optimized Intel TBB [242] or OpenMP [243] library. Parallel versions of the fast Fourier transform algorithm are provided by the FFTW [234] or Intel MKL [233] library. The matrix multiplications of the discrete Hankel transform make use of optimized BLAS [244] implementations.

The particle push is implemented as a simple parallel loop over all macroparticles. The distribution of individual loop elements to the threads is handled internally by Numba.

A different approach is chosen for the implementation of the field gathering and deposition routines. Here, outer parallel loops iterate over the total number of threads and the particle arrays are divided into smaller chunks. Each thread then performs a sequential loop over its batch of particles.

For the deposition, we additionally create multiple copies (one for each thread) of the 2D charge and current density arrays. The particles of each thread then deposit into their local copy, which avoids any race conditions. Eventually, the copies are reduced to a single array by performing a parallel loop over the total number of cells in the longitudinal direction, with an inner sequential loop that sums the contributions of each copy in the transverse direction.

Element-wise operations on the 2D field arrays, such as the PSATD push or the filtering of the source terms, are also implemented as a combination of an outer parallel loop over the longitudinal dimension and an inner sequential loop over the transverse dimension.

Further note that each azimuthal mode is treated separately. Although this duplicates read accesses in the gathering and deposition methods, it greatly simplifies the structure of the code without significantly degrading its performance.

GPU implementation

The GPU implementation is based on the Numba [235] and cupy [239] package. Numba allows to directly compile Python functions into CUDA [237] kernels and execute them on Nvidia GPUs. For performance reasons, Numba is only used to translate the implemented particle-in-cell functions to low-level CUDA instructions. The allocation of device memory and the actual launching of kernels is managed by cupy. The latter package also exposes direct API calls to the cuFFT and cuBLAS library [237], which provide optimized routines for the FFT and DHT operations.

At the beginning of the program, all simulation data is initialized on the CPU and copied to the GPU. To avoid unnecessary data transfers between host and device memory, the simulation remains on the GPU for the entire particle-in-cell loop. Parts of it are moved back to the CPU when writing results to disk and additional transfers

occur when continuously injecting new particles or a laser. Because these additional transfers are small or do not happen every time step, they usually have a negligible impact on performance.

An important difference to the CPU implementation is that particles are sorted by cell location to avoid race conditions and to facilitate coalescing memory access patterns in the gathering and deposition routines. The latter means that a sequence of values accessed by a single or a group of threads should ideally be located next to each other in memory. The sorting also simplifies the resizing of particle arrays on the GPU, which is required when removing or adding new particles in the longitudinal direction.

Sorting of the particles happens at least twice per time step, prior to the charge and current deposition. The particles remain sorted until after the particle push of the next time step. First, the index of the cell that each particle resides in is calculated from its position and stored in a flattened 1D cell index array. This array is then sorted, using an efficient parallel Radix sort provided by the Thrust library, which is accessible through `copy`. During sorting, we keep track of the index change of each value and afterwards use this information to rearrange all other particle arrays. In a final step, we additionally infer the number of particles per cell from a parallel cumulative sum of the sorted cell indices.

The execution of CUDA kernels is based on launching a very large number of threads, ideally one for each data element that can be processed in parallel. For this reason, we use a single thread per macroparticle in the particle push and field interpolation routines. In the latter kernel, each thread additionally performs a loop over closest grid points to sequentially gather all field contributions. Kernels that act on the field arrays also execute with one thread per element.

The deposition again represents an exception as it is parallelized on a per-cell basis. Each thread iterates over the particles per cell to sequentially deposit their charge and current onto a thread-local representation of the surrounding grid points. This procedure already avoids the majority of possible race conditions. The summed-up contributions of each cell are then added to the global array. Because neighboring threads share the same grid points, atomic add operations are used in this final step of writing to device memory.

The maximum achievable performance of the GPU implementation is mainly limited by the available memory bandwidth. This is because the individual operations of the particle-in-cell algorithm are usually not very compute-intensive, but overall demanding in terms of reading and writing data from and to device memory.

To conclude, it should be highlighted that the code runs much faster on a modern GPU than on a multi-core CPU, assuming typical grid sizes and comparable hardware costs.

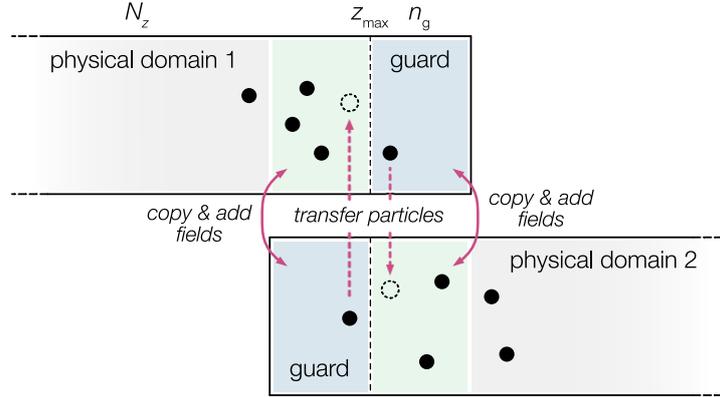


Figure 6.6 – Exchange of information between adjacent sub-domains. Each domain is extended by a guard region (blue) that overlaps with the physical domain of its neighbor (green). At each time step, the updated electromagnetic fields of the local domain are copied to the neighboring guard region. Similarly, the deposited quantities are added in the overlapping region. Particles that propagate into the guard region are exchanged at regular intervals.

6.3.2 Inter-device parallelization

Owing to the reduced geometry of FBPIC, many practical simulation cases can already be run efficiently on a single device. However, to further reduce the run time or when running very large setups, it is necessary to scale the execution to multiple devices.

As explained before, the inter-device parallelization of the code is based on domain decomposition in the longitudinal direction. The simulation box is split into multiple sub-domains, while ensuring continuity of the physics by exchanging particle and field information at the boundaries. Communication between sub-domains is handled by MPI using the mpi4py package [241]. The parallelization strategy is enabled by the finite-order PSATD solver that comes with a trade-off between accuracy and communication overhead (see Sec. 5.3.4 for details).

Figure 6.6 shows the grid layout at the interface of two domains and illustrates the exchange of information. Each sub-domain is extended by a guard region of n_g cells that overlap with the $n_p \equiv n_g$ outer cells of the physical region of the neighboring domain. Furthermore, we add the following communication steps to the particle-in-cell cycle:

- (1) Once filtered and corrected in spectral space, the current density is transformed back onto the spatial grid. The processed current density in the overlapping $n_p + n_g$ cells is then exchanged and added to the neighboring domain, as if particles would simultaneously reside in both domains. After the exchange, the current density is again transformed to spectral space. The procedure ensures

that each sub-domain sees the total current of all particles in the overlapping region when advancing the E and B fields. Note that by default, the deposited charge density is not exchanged as it is only used for the domain-local correction of the currents.

- (2) Eventually, the electromagnetic fields of each sub-domain are advanced in spectral space and transformed back to real space. Because the information from the neighboring domain is missing, the fields evolve incorrectly at the boundaries and wrap around the edges of the local grid. Owing to the locality of the finite-order Maxwell solver, however, these errors will be confined to the guard region. We therefore replace the guard region of each neighboring domain with the correct electromagnetic fields of the overlapping physical region. The cycle then continues as usual with possible damping of the fields at open boundaries, before transforming the fields back to spectral space.
- (3) Particles that travel into the guard region are removed and transferred to the physical region of the neighboring domain at the beginning of a cycle. To save communication overhead, particles are not exchanged at every time step, but only after a certain period that is given by the minimum time it would take for a particle to advance by $n_g/2$ cells. This ensures that particles never travel more than halfway into the guard region.

Locality of the finite-order PSATD solver

The size of the guard region depends on the effective stencil width of the finite-order PSATD solver. In section Sec. 5.3.4, we discussed that this width can not be inferred directly from the mathematical definition of the modified wavevector $[k]$.

However, it was shown in previous work [44] that we can study the locality of the PSATD solver by applying an inverse Fourier transform to the coefficients for advancing the fields, i.e., C and S in Eqs. (6.33) and (6.34). By that we get an equivalent to the finite-difference stencil of an FDTD solver.

For the cylindrical formulation and using finite-order derivatives in the longitudinal direction only, we can numerically calculate the effective stencil coefficients α^* from a sum of inverse Fourier transforms \mathcal{F}^{-1} along z :

$$\alpha^* = \sqrt{\left|\mathcal{F}^{-1}(C)\right|^2 + \left|\mathcal{F}^{-1}\left(\frac{S}{[k]}k_{\perp}\right)\right|^2 + \left|\mathcal{F}^{-1}\left(\frac{S}{[k]}[k_z]\right)\right|^2}, \quad (6.46)$$

where $[k] = \sqrt{[k_z]^2 + k_{\perp}^2}$.

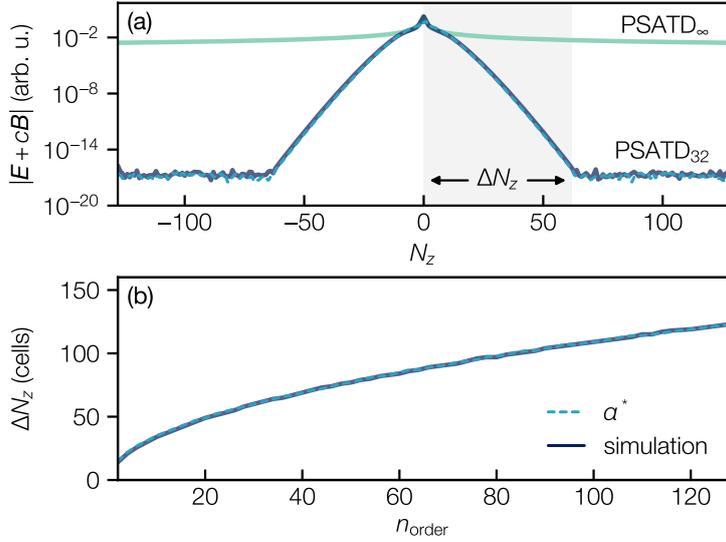


Figure 6.7 – (a) Numerically estimated [from Eq. (6.46)] and simulated stencil. The fields extend globally for the standard case (PSATD_∞), but remain localized for a finite order (PSATD₃₂). (b) Maximum stencil reach ΔN_z for different values of n_{order} .

Alternatively, we can retrieve the stencil directly from an FBPIC simulation. We initialize a δ -peak in the electromagnetic fields at the center of a grid and propagate it for a single time step $\Delta t = \Delta z/c$. The resulting spreading of the fields is a direct measure of the stencil. Note that the propagation of this signal is not physical, since the δ -peak does not satisfy $\nabla \cdot \mathbf{E} = 0$ and $\nabla \cdot \mathbf{B} = 0$ at $t = 0$. Instead, this is a simple empirical method to retrieve the locality of the solver from a simulation.

Figure 6.7 compares the stencils from the numerical evaluation [Eq. (6.46)] and the simulation. As expected, the infinite-order solver globally affects the fields with a stencil that extends over the entire grid. Contrarily, for $n_{\text{order}} = 32$, the stencil quickly falls off to machine precision over a distance $\Delta N_z \Delta z$. The lower plot additionally shows the maximum stencil extent for different values of n_{order} .

Consequently, we choose the guard region size to match this width, $n_g \equiv \Delta N_z$, to guarantee a seamless continuation of the physics at the boundaries. Using significantly fewer guard cells would lead to spurious distortions of the electromagnetic fields over time. It should be noted that the curl-free correction of the currents [Eq. (6.31)] remains non-local, yet truncating the stencil of this operation typically has no negative effect on the simulation. The cross-deposition correction, on the other hand, is strictly local with an effective stencil width of $\simeq \Delta N_z/2 \equiv n_g/2$, i.e., half the size of the guard region.

6.3.3 Parallel scalability

To evaluate the performance of the inter-device parallelization, we present scaling tests of FBPIC on the supercomputer JUWELS [245, 246]. Two separate benchmarks have been performed, respectively measuring the strong and weak scaling efficiency of the code.

The benchmarks have been performed on the Booster module [246], running the code on up to 256 compute nodes. Each node is equipped with $2 \times$ AMD EPYC Rome 7402 CPUs (24 cores each), 512 GB RAM and $4 \times$ Nvidia A100 GPUs (each with 40 GB of HBM2e memory). Nodes are interconnected via $4 \times$ InfiniBand HDR (Connect-X6) adapters.

It should be highlighted that FBPIC supports the use of Nvidia GPUDirect technology, which enables remote direct memory access (RDMA) between GPUs. On JUWELS, this allows for a direct exchange of GPU data between domains via MPI, without transferring the data to the CPU first.

Strong scaling

A strong scaling test measures the ability to reduce the time-to-solution by parallel computation of a problem with fixed size. In particular, it reveals the overhead that limits the minimum achievable time per iteration in a PIC code.

The strong scaling of FBPIC is evaluated based on a simulation of a nonlinear plasma wave that is driven by a laser pulse ($a_0 = 2$) at a density of $3 \times 10^{18} \text{cm}^{-3}$. While increasing the number of sub-domains, the total grid size (excluding guard cells) is kept constant at 16384×1024 (z, r) cells with $2 \times 2 \times 4$ (r, z, θ) particles per cell. Reflective transverse boundary conditions, two azimuthal modes and cubic particle shape factors are used. The order of the PSATD solver is set to $n_{\text{order}} = 8$, resulting in a guard region size of $n_g = 32$ cells. A particle exchange occurs every 6 iterations. The system is first evolved in time until the laser-driven plasma wave fully developed and then solved for another 256 iterations for measuring the timings.

Figure 6.8 shows the speedup (with respect to the run time on 4 GPUs) and the average time per step when scaling the computation from 4 to 64 GPUs. For the considered setup, the speedup is approximately linear up to 16 GPUs. Beyond that, the strong scaling efficiency is reduced from a combination of Python function call and MPI communication overhead, as well as a decrease in GPU utilization for small sub-domain sizes. By using a realistic LPA setup, we also introduce load imbalance between domains, which further limits the scalability.

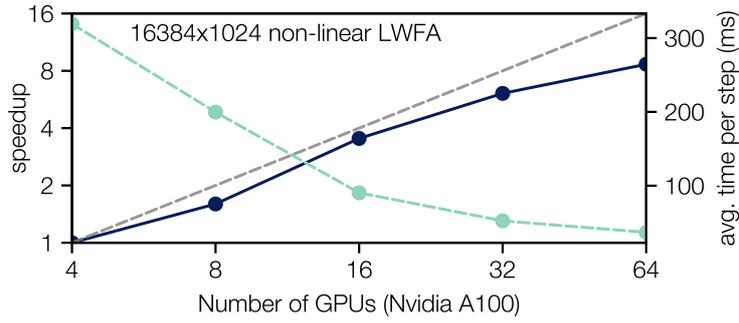


Figure 6.8 – Results of a strong scaling test of FBPIC on up to 64 Nvidia A100 GPUs on the supercomputer JUWELS. Average time per step (green dashed line) and speedup relative to the run time on 4 GPUs (blue solid line) for a simulation of a nonlinear laser-driven plasma wave. Ideal speedup (gray dashed line).

For comparable use cases, it is therefore most efficient to run FBPIC on 1-16 GPUs, or, viewed differently, not to scale the execution beyond the point where the average time per step is ~ 25 -50 ms.

Weak scaling

Complementary to the previous benchmark, a weak scaling test aims at demonstrating the ability to scale to arbitrarily large problem sizes, while keeping the amount of work per compute unit fixed.

For the benchmark, we simulate a hot plasma of electrons and ions. The electrons are initialized with a normalized momentum spread of $\sigma_{\beta\gamma} = 0.1$ rms in each direction. The initial grid has a size of 2048×1024 (z, r) cells with $2 \times 2 \times 4$ (r, z, θ) particles per cell. For each fourfold increase in number of GPUs, the grid resolution is increased by the same factor in the longitudinal direction, up to a total of ~ 2 million $\times 1024$ cells. The other numerical parameters are the same as for the strong scaling case above. The system is evolved for 256 iterations to measure the computation time.

Figure 6.9 shows the weak scaling efficiency with respect to the single-GPU simulation. Running the code in parallel increases the average time per step by $\sim 25\%$ compared to the serial execution. This is expected from the associated overhead in function calls and communication. Nevertheless, the efficiency remains at a high value of $\sim 80\%$ when further increasing the number of domains, showing optimum scaling up to a total of 1024 GPUs.

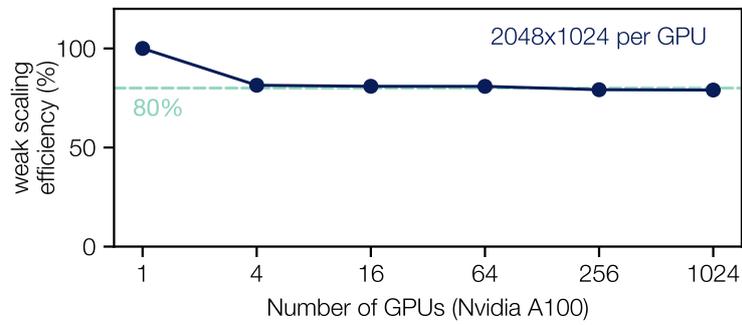


Figure 6.9 – Results of a weak scaling test of FBPIC on up to 1024 Nvidia A100 GPUs on the supercomputer JUWELS. The weak scaling efficiency is defined relative to the run time of the single-GPU simulation.

7 Elimination of the numerical Cherenkov instability

This chapter describes a novel pseudospectral particle-in-cell algorithm, the Galilean scheme, that intrinsically eliminates the numerical Cherenkov instability (NCI) in simulations of relativistically drifting plasmas. The method and its application were first published in [51], together with a formal derivation and stability analysis in [52]. More recently, a generalization of the solver to arbitrary order was published in [53]. The following sections closely follow [51, 52] and are in part reproductions of [53]. In Sec. 7.1, we describe the general concept and place it in context with previous NCI mitigation strategies. Section 7.2 presents the novel algorithm in detail, followed by a theoretical stability analysis in Sec. 7.3.

Co-author contributions: The Galilean-PSATD solver was co-invented by the present author and R. Lehe. The present author proposed the concept and studied its application to laser-plasma acceleration, and mainly worked on the later extension to arbitrary order. R. Lehe derived the solver equations and the plasma dispersion relation that forms the basis of the theoretical stability analysis. Both worked on the numerical implementation, with contributions from S. Jalas and O. Shapoval.

7.1 Stable discrete representation of relativistically streaming plasmas

Many physical systems that could in principle be modeled with the electromagnetic particle-in-cell method involve relativistic plasmas. A particularly relevant subclass of problems deals with relativistically streaming plasmas, including, for example, simulations of plasma accelerators in an optimal, Lorentz-boosted frame of reference [49, 209–211] or relativistic collisionless shocks [206–208] in astrophysics. However, as already discussed in Sec. 5.4, the occurrence of the numerical Cherenkov instability [54, 55, 204, 205] can fundamentally limit the applicability of the particle-in-cell method for such problems.

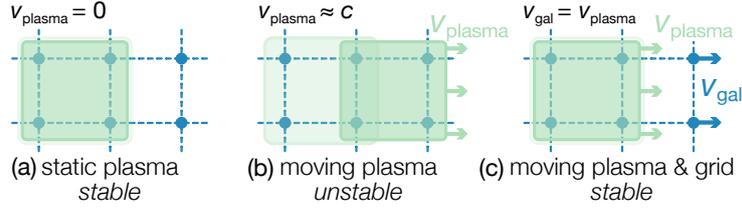


Figure 7.1 – Schematic illustrating the basic principle of the Galilean-PSATD solver. In contrast to a static plasma (a), particle-in-cell simulations of a relativistically streaming plasma (b) can suffer from the numerical Cherenkov instability. The instability can be eliminated by discretizing and solving Maxwell’s equations in Galilean coordinates that are comoving with the plasma (c).

7.1.1 Previous NCI mitigation strategies

Various strategies have been developed to suppress the NCI. For example, smoothing of the electric and magnetic fields [210], wide-band filtering of the currents [193, 209] or the use of specific time steps [193] can reduce the instability growth rate. More advanced methods mitigate the NCI by scaling the deposited currents with a wavenumber-dependent factor [247, 248], by slightly changing the ratio of the gathered E and B fields [249, 250], or by artificially modifying the vacuum dispersion in combination with using a small time step and low-pass spatial filtering [251–253].

Although these strategies can efficiently suppress the instability, they do not eliminate it. Furthermore, they introduce NCI-specific corrections that sometimes contradict the natural discretization of the underlying physical equations and therefore may alter the physical results. In addition, the solutions are typically restricted to a 2D or 3D Cartesian geometry, with the exception of the last listed method that was also successfully applied to a quasi-cylindrical algorithm [252, 253].

7.1.2 Galilean scheme

Compared to previous work, we take a fundamentally different approach to solve the problem of NCI. The idea is to change the mathematical representation of the physics prior discretization, such that, from a numerical point of view, a moving plasma appears to be at rest. The intuition is that the resulting discrete representation should recover the stability properties of a static plasma simulation.

As shown in Fig. 7.1, this is achieved by performing a Galilean change of coordinates of the form

$$\mathbf{x}' = \mathbf{x} - \mathbf{v}_{\text{gal}}t, \quad (7.1)$$

where \mathbf{v}_{gal} is the velocity of the moving Galilean frame. We furthermore choose $\mathbf{v}_{\text{gal}} = \mathbf{v}_{\text{plasma}}$, where $\mathbf{v}_{\text{plasma}}$ is the speed of the relativistic plasma. As a consequence,

the numerical grid comoves with the plasma in the standard coordinates \boldsymbol{x} , whereas both appear to be at rest in the Galilean coordinates \boldsymbol{x}' .

In fact, we could show [51, 52] that combining this heuristic approach with a spectral framework eliminates the NCI in simulations of relativistically streaming plasmas. More precisely, a stable algorithm is obtained by extending the derivation of the PSATD solver, starting with the fundamental particle-in-cell equations in the Galilean coordinates \boldsymbol{x}' . Following the usual discretization and analytical time integration under the assumption of comoving currents, this so-called *Galilean-PSATD solver* becomes intrinsically free of the NCI for a uniform plasma flowing at the velocity of the comoving coordinate system.

Unlike the previous strategies to suppress the NCI [247–253] that have been discussed above, the Galilean scheme does not require artificial corrections or numerical modifications of the underlying solver. Because the method is based on a simple change of coordinates it has no impact on the continuous space-time representation of the physics, yet positively affects the numerical stability of the discretized equations. It is furthermore independent of the geometry and therefore can be applied not only to 2D and 3D Cartesian algorithms, but also to the spectral, quasi-cylindrical algorithm that was presented in Ch. 6.

The Galilean scheme was originally applied to the standard PSATD solver with infinite-order accuracy [51, 52]. In this case, a qualitative explanation for the absence of the numerical Cherenkov instability is threefold:

- (1) Because the solver correctly models the electromagnetic vacuum dispersion relation, the NCI is mitigated to zeroth order (no NCR).
- (2) The immobility of the plasma with respect to the numerical grid suppresses the coupling of spatial aliases of the fields to the particles.
- (3) The analytical integration naturally takes into account the comoving currents in the direction of the plasma motion and thereby eliminates remaining sources of the instability.

Moreover, as we have shown later [53], the formulation can be extended to incorporate spatial derivatives of arbitrary order. Although this introduces spurious numerical dispersion, the Galilean transformation introduces an asymmetry that cancels the NCR resonance, which is why the first stability criteria holds true even in the finite-order case.

In the following, we present this more general, arbitrary-order Galilean-PSATD solver in detail and prove the elimination of the NCI based on a theoretical and numerical stability analysis. In Ch. 8, we demonstrate the practical usefulness of the solver by applying it to simulations of plasma acceleration in a Lorentz-boosted frame.

7.2 Galilean-PSATD solver of arbitrary order

7.2.1 Particle-in-cell equations in Galilean coordinates

In the Galilean coordinates, $\mathbf{x}' = \mathbf{x} - \mathbf{v}_{\text{gal}}t$, the Newton-Lorentz equations of motion read

$$\frac{d\mathbf{x}'}{dt} = \frac{\mathbf{p}}{\gamma m} - \mathbf{v}_{\text{gal}}, \quad (7.2a)$$

$$\frac{d\mathbf{p}}{dt} = q \left(\mathbf{E} + \frac{\mathbf{p}}{\gamma m} \times \mathbf{B} \right), \quad (7.2b)$$

and Maxwell's equations become

$$\left(\frac{\partial}{\partial t} - \mathbf{v}_{\text{gal}} \cdot \nabla' \right) \mathbf{B} = -\nabla' \times \mathbf{E}, \quad (7.3a)$$

$$\frac{1}{c^2} \left(\frac{\partial}{\partial t} - \mathbf{v}_{\text{gal}} \cdot \nabla' \right) \mathbf{E} = \nabla' \times \mathbf{B} - \mu_0 \mathbf{J}, \quad (7.3b)$$

$$\nabla' \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (7.3c)$$

$$\nabla' \cdot \mathbf{B} = 0, \quad (7.3d)$$

where ∇' denotes a spatial derivative with respect to the Galilean coordinates \mathbf{x}' . Accordingly, the continuity equation transforms to

$$(\partial_t - \mathbf{v}_{\text{gal}} \cdot \nabla') \rho + \nabla' \cdot \mathbf{J} = 0. \quad (7.4)$$

Just like for the arbitrary-order PSATD solver, the *spatial* derivatives ∇' are furthermore replaced by discretized derivatives $\hat{\nabla}'$ on a grid, which are accurate to order n_{order} and correspond, in Fourier space, to a multiplication by a modified wavevector $[\mathbf{k}]$ [compare Eq. (5.36)] [46, 44]. As explained previously in Sec. 5.3.4, we thereby introduce spatial locality to the evolution of the fields on the grid, which is the prerequisite for efficient parallelization by domain decomposition.

These equations are then discretized and integrated in time following the same steps as for the arbitrary-order PSATD particle-in-cell algorithm that was described in Chs. 5 and 6.

7.2.2 Derivation of the discretized Maxwell equations

In the PSATD scheme, Maxwell's equations are first transformed to Fourier space and then integrated *analytically* over one time step.

More explicitly, the Fourier space representation of Eqs. (7.3a) and (7.3b) with discretized derivatives is

$$\left(\frac{\partial}{\partial t} - i[\mathbf{k}] \cdot \mathbf{v}_{\text{gal}}\right) \hat{\mathcal{B}} = -i[\mathbf{k}] \times \hat{\mathcal{E}}, \quad (7.5a)$$

$$\frac{1}{c^2} \left(\frac{\partial}{\partial t} - i[\mathbf{k}] \cdot \mathbf{v}_{\text{gal}}\right) \hat{\mathcal{E}} = i[\mathbf{k}] \times \hat{\mathcal{B}} - \mu_0 \hat{\mathcal{J}}, \quad (7.5b)$$

where the Fourier transform is defined by $\hat{\mathcal{F}}(\mathbf{k}, t) = \int d^3\mathbf{x}' F(\mathbf{x}', t) e^{-i\mathbf{k} \cdot \mathbf{x}'}$.

These equations can be integrated analytically in time provided that the evolution of the source terms $\hat{\rho}$ and $\hat{\mathcal{J}}$ is known and that $\hat{\mathcal{E}}$, $\hat{\mathcal{B}}$, $\hat{\mathcal{J}}$ and $\hat{\rho}$ satisfy the conservation equations:

$$i[\mathbf{k}] \cdot \hat{\mathcal{B}} = 0, \quad i[\mathbf{k}] \cdot \hat{\mathcal{E}} = \frac{\hat{\rho}}{\epsilon_0}, \quad (\partial_t - i[\mathbf{k}] \cdot \mathbf{v}_{\text{gal}})\hat{\rho} + i[\mathbf{k}] \cdot \hat{\mathcal{J}} = 0. \quad (7.6)$$

However, because the deposited real-space quantities ρ and \mathbf{J} are only known at discrete times (at $n\Delta t$ and $(n+1)\Delta t$ for ρ and $(n+1/2)\Delta t$ for \mathbf{J}), one has to make explicit assumptions about their time evolution for the integration.

For the standard PSATD solver, one usually assumes the currents to be constant over one time step in the standard coordinates \mathbf{x} ,

$$\mathbf{J}(\mathbf{x}, t) = \mathbf{J}(\mathbf{x}, (n+1/2)\Delta t) \quad \forall t \in [n\Delta t, (n+1)\Delta t]. \quad (7.7)$$

Consequently, to further embed the idea of representing a moving plasma as numerically static, we also assume the currents to be constant in the Galilean coordinates \mathbf{x}' ,

$$\mathbf{J}(\mathbf{x}', t) = \mathbf{J}(\mathbf{x}', (n+1/2)\Delta t) \quad \forall t \in [n\Delta t, (n+1)\Delta t]. \quad (7.8)$$

Note that these two assumptions are fundamentally different. In fact, Eq. (7.8) is equivalent to assume that the currents are comoving with the bulk plasma in the standard coordinates. This so-called *comoving current assumption* [52, 51] is a key reason for the absence of the NCI with the Galilean-PSATD solver.

From the time evolution of the currents $\hat{\mathcal{J}}$ and under consideration of the continuity equation, Eq. (7.6), we can infer the time evolution of the charge density $\hat{\rho}$:

$$\hat{\mathcal{J}}(\mathbf{k}, t) = \hat{\mathcal{J}}(\mathbf{k}, (n+1/2)\Delta t), \quad (7.9a)$$

$$\begin{aligned} \hat{\rho}(\mathbf{k}, t) &= \hat{\rho}(\mathbf{k}, (n+1)\Delta t) \frac{1 - e^{i\mathbf{k} \cdot \mathbf{v}_{\text{gal}}(t-n\Delta t)}}{1 - e^{i\mathbf{k} \cdot \mathbf{v}_{\text{gal}}\Delta t}} \\ &\quad - \hat{\rho}(\mathbf{k}, n\Delta t) \frac{e^{i\mathbf{k} \cdot \mathbf{v}_{\text{gal}}\Delta t} - e^{i\mathbf{k} \cdot \mathbf{v}_{\text{gal}}(t-n\Delta t)}}{1 - e^{i\mathbf{k} \cdot \mathbf{v}_{\text{gal}}\Delta t}}. \end{aligned} \quad (7.9b)$$

Equations (7.9a) and (7.9b) are then combined with Eqs. (7.5a) and (7.5b) to derive the expressions for the time evolution of $\hat{\mathcal{E}}$ and $\hat{\mathcal{B}}$. For the mathematical details of this analytical integration we refer to [52].

The final update equations of the Galilean-PSATD solver are given by

$$\hat{\mathcal{B}}^{n+1} = \theta^2 C \hat{\mathcal{B}}^n - \frac{\theta^2 S}{c[k]} i[\mathbf{k}] \times \hat{\mathcal{E}}^n + \frac{\theta \chi_1}{\epsilon_0 c^2 [k]^2} i[\mathbf{k}] \times \hat{\mathcal{J}}^{n+\frac{1}{2}}, \quad (7.10a)$$

$$\begin{aligned} \hat{\mathcal{E}}^{n+1} &= \theta^2 C \hat{\mathcal{E}}^n + \frac{\theta^2 S}{[k]} ci[\mathbf{k}] \times \hat{\mathcal{B}}^n + \frac{i\nu\theta\chi_1 - \theta^2 S}{\epsilon_0 c [k]} \hat{\mathcal{J}}^{n+\frac{1}{2}} \\ &\quad - \frac{1}{\epsilon_0 [k]^2} (\chi_2 \hat{\varrho}^{n+1} - \theta^2 \chi_3 \hat{\varrho}^n) i[\mathbf{k}], \end{aligned} \quad (7.10b)$$

where we use the shorthand notations $\hat{\mathcal{E}}^n \equiv \hat{\mathcal{E}}(\mathbf{k}, n\Delta t)$, $\hat{\mathcal{B}}^n \equiv \hat{\mathcal{B}}(\mathbf{k}, n\Delta t)$ and

$$[k] = \sqrt{[\mathbf{k}]^2}, \quad C = \cos(c[k]\Delta t), \quad S = \sin(c[k]\Delta t), \quad (7.11a)$$

$$\nu = \frac{[\mathbf{k}] \cdot \mathbf{v}_{\text{gal}}}{c[k]}, \quad \theta^{(*)} = e^{(-)i[\mathbf{k}] \cdot \mathbf{v}_{\text{gal}} \Delta t / 2}, \quad (7.11b)$$

$$\chi_1 = \frac{1}{1 - \nu^2} (\theta^* - C\theta + i\nu\theta S), \quad (7.11c)$$

$$\chi_2 = \frac{\chi_1 - \theta(1 - C)}{\theta^* - \theta}, \quad \chi_3 = \frac{\chi_1 - \theta^*(1 - C)}{\theta^* - \theta}. \quad (7.11d)$$

In addition, the discretized continuity equation reads

$$-i([\mathbf{k}] \cdot \mathbf{v}_{\text{gal}}) \frac{\theta^* \hat{\varrho}^{n+1} - \hat{\varrho}^n \theta}{\theta^* - \theta} + i[\mathbf{k}] \cdot \hat{\mathcal{J}}^{n+\frac{1}{2}} = 0. \quad (7.12)$$

Because the current \mathbf{J}_d from a *direct* deposition scheme (from the particles to the grid) does not automatically satisfy the continuity equation, this current can be corrected using the *curl-free* method described in Sec. 5.2.2,

$$\hat{\mathcal{J}}^{n+\frac{1}{2}} = \hat{\mathcal{J}}_d^{n+\frac{1}{2}} + \frac{i[\mathbf{k}]}{[k]^2} \hat{\mathcal{G}}, \quad (7.13a)$$

$$\hat{\mathcal{G}} = -i([\mathbf{k}] \cdot \mathbf{v}_{\text{gal}}) \frac{\theta^* \hat{\varrho}^{n+1} - \theta \hat{\varrho}^n}{\theta^* - \theta} + i[\mathbf{k}] \cdot \hat{\mathcal{J}}_d^{n+\frac{1}{2}}. \quad (7.13b)$$

Note that for $n_{\text{order}} \rightarrow \infty$, $[\mathbf{k}] \rightarrow \mathbf{k}$ and Eqs. (7.5a) to (7.13b) then equal the infinite-order formulation of the solver as originally derived in [51, 52]. In the limit $\mathbf{v}_{\text{gal}} = \mathbf{0}$, the above expressions reduce to the standard PSATD equations [183, 45]. For $\mathbf{v}_{\text{gal}} = \mathbf{v}_{\text{plasma}}$, on the other hand, the solver becomes intrinsically free of the NCI for a relativistically drifting plasma.

Apart from its resilience against the NCI, the arbitrary-order Galilean-PSATD solver inherits all of the advantageous properties of the standard PSATD solver. Due to the

analytical integration in time, the solver is not bound to a CFL condition - in principle allowing for arbitrarily long time steps. The fields are centered in space and - except for the currents \mathbf{J} - also in time.

Another fundamental advantage of the Galilean scheme is its independence of geometry. Analogous equations in cylindrical geometry with azimuthal mode decomposition can be readily derived from the equations listed above, by using the methodology presented in Sec. 6.2.4 [52]. It should be noted again that in this case finite-order derivatives are possible along the axial direction only.

7.2.3 Practical implementation

The Galilean scheme has been implemented in the particle-in-cell frameworks WARP [254] and WARPX [255] in 2D and 3D Cartesian coordinates and in the quasi-3D code FBPIC [43] that was presented in Ch. 6. Apart from updating the equations of the Maxwell solver, this required only minor changes to the other parts of the particle-in-cell cycle.

In principle, with the Galilean scheme, everything should be solved in the new coordinates $\mathbf{x}' = \mathbf{x} - \mathbf{v}_{\text{gal}}t$ and $-\mathbf{v}_{\text{gal}}$ should be added to the velocity of the particles in the discretized equations of motion. The Vay push [compare Eq. (5.14)] then reads

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \Delta t \left(\frac{\mathbf{p}^{n+1/2}}{\gamma^{n+1/2}m} - \mathbf{v}_{\text{gal}} \right), \quad (7.14)$$

where $\gamma^{n+1/2} = \sqrt{1 + (\mathbf{p}^{n+1/2}/mc)^2}$.

However, it is equivalent to, instead, shift the boundaries of the numerical grid. In the practical implementation of FBPIC, for example, the longitudinal simulation box boundaries are simply updated twice per iteration, after each half-push of the particle positions, according to

$$z_{\text{min}} = z_{\text{min}} + v_{\text{gal}}\Delta t/2, \quad (7.15a)$$

$$z_{\text{max}} = z_{\text{max}} + v_{\text{gal}}\Delta t/2. \quad (7.15b)$$

In this case, of course, the coordinate transformation is restricted to the axial direction z only.

Further note that the Galilean scheme should not be confused with a moving window [compare Sec. 6.2.5]. The latter technique works by (virtually) adding and removing columns of cells at the edges of the grid, which is not equivalent to the change of coordinates described here. Finally, it should be highlighted that the Galilean scheme adds, if at all, only negligible computational cost to the particle-in-cell cycle.

7.2.4 Locality with a finite-order stencil

To achieve locality of the field solver, we introduced the modified wavevector $[k]$ in Eqs. (7.10a) and (7.10b), representing spatial derivatives of arbitrary order n_{order} . However, as explained in Sec. 5.3.4, the analytical integration in time does not preserve the strict locality of the underlying finite-difference stencil. Mathematically, this is reflected by the fact that $[k]$ appears in the coefficients C , S and - if $|v_{\text{gal}}| \neq 0$ - also in θ , ν , χ_1 , χ_2 , χ_3 [Eqs. (7.11a) to (7.11d)].

We have shown in Sec. 6.3.2 that we can get an effective stencil width of the PSATD solver by applying an inverse Fourier transform to the set of coefficients for advancing the fields. Here, we repeat this analysis for the Galilean case.

Using again the cylindrical formulation with finite-order derivatives in z only, the stencil coefficients

$$\alpha^* = \sqrt{\left| \mathcal{F}^{-1}(\theta^2 C) \right|^2 + \left| \mathcal{F}^{-1}\left(\frac{\theta^2 S}{[k]} k_{\perp}\right) \right|^2 + \left| \mathcal{F}^{-1}\left(\frac{\theta^2 S}{[k]} [k_z]\right) \right|^2} \quad (7.16)$$

can be numerically calculated for arbitrary values of the transverse wavenumber k_{\perp} , Δt and $v_{\text{gal}} = v_{\text{gal}} \hat{u}_z$. Here, $[k] = \sqrt{[k_z]^2 + k_{\perp}^2}$. In addition, we use FBPIC to initialize a δ -peak and propagate it for a single time step $\Delta t = \Delta z/c$ in order to retrieve the stencil directly from a simulation.

Fig. 7.2(a) compares the results from the numerical evaluation [Eq. (7.16)] and the simulation. In accordance with the previous analysis in Sec. 6.3.2, the infinite-order stencil extents over the entire grid, whereas for $n_{\text{order}} = 32$ it falls off to machine precision at a distance $\Delta N_z \Delta z$. While symmetric for $v_{\text{gal}} = 0$, the stencil leans towards the direction of the grid velocity for $v_{\text{gal}} = c$. Hence, the stencil becomes wider but the solver remains local for $|v_{\text{gal}}| > 0$. This can also be seen in Fig. 7.2(b), which shows the maximum stencil extent ΔN_z for different values of n_{order} and v_{gal} .

The analysis demonstrates that the Galilean-PSATD solver can also be scaled to many compute units by spatially decomposing the simulation box into sub-domains that are computed in parallel. However, as each subgrid is extended by a boundary layer of ΔN_z cells to exchange the fields between neighboring domains, the communication overhead increases slightly for $|v_{\text{gal}}| > 0$.

It should again be noted that the charge conserving correction of the current density, Eq. (7.13a), remains non-local even for a finite order. However, as we have already discussed in Sec. 6.3.2, truncating its effective stencil function at ΔN_z only causes very small variations in the currents that typically have no negative effect on the simulation. Alternatively, it is possible to use a charge conserving current deposition that preserves

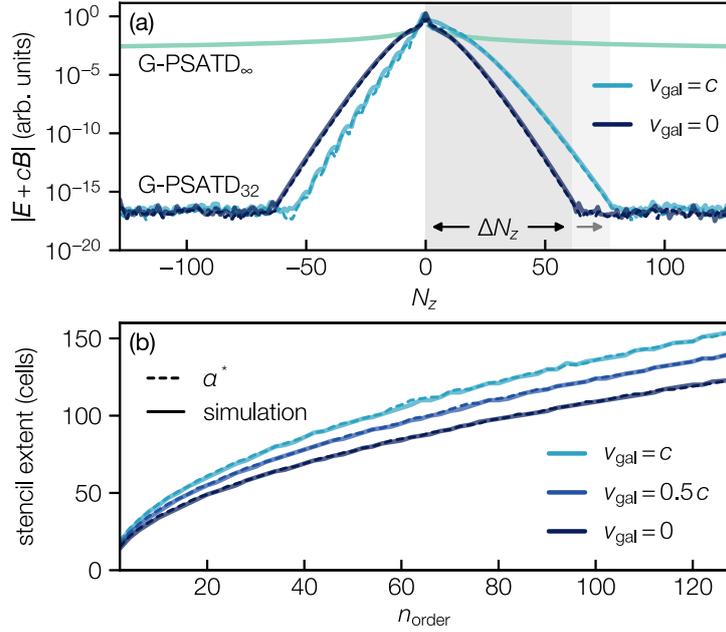


Figure 7.2 – (a) Numerically estimated [from Eq. (7.16)] and simulated stencil. The fields extent globally for the standard case (Galilean-PSATD_∞), but remain localized for a finite order (Galilean-PSATD₃₂). (b) Maximum stencil reach ΔN_z for different values of n_{order} and v_{gal} .

locality [45]. For the following analysis, we restrict ourselves to using the simple current correction from Eq. (7.13a).

7.2.5 Asymmetry of the vacuum dispersion relation

In the previous section we showed that the Galilean-PSATD solver behaves similarly to the standard PSATD solver and becomes more local for a finite order. As explained before, the increased locality comes at the price of decreasing the solver’s accuracy, which causes the vacuum dispersion relation to be altered.

Considering Eqs. (7.10a) and (7.10b) in vacuum (i.e., $\hat{\mathcal{J}} = 0$ and $\hat{\rho} = 0$), with electromagnetic modes of the form $e^{i\mathbf{k}\cdot\mathbf{x}-i\omega t} = e^{i\mathbf{k}\cdot\mathbf{x}'-i(\omega-\mathbf{k}\cdot\mathbf{v}_{\text{gal}})t}$ leads, after some algebra, to the numerical vacuum dispersion relation of the arbitrary-order Galilean-PSATD solver

$$\omega = \pm c[\mathbf{k}] + \mathbf{v}_{\text{gal}} \cdot (\mathbf{k} - [\mathbf{k}]), \quad (7.17)$$

which describes the propagation of electromagnetic waves on the discretized grid. Fig. 7.3 shows this dispersion relation for different values of v_{gal} in a single direction z (i.e., along k_z with $k_x = k_y = 0$).

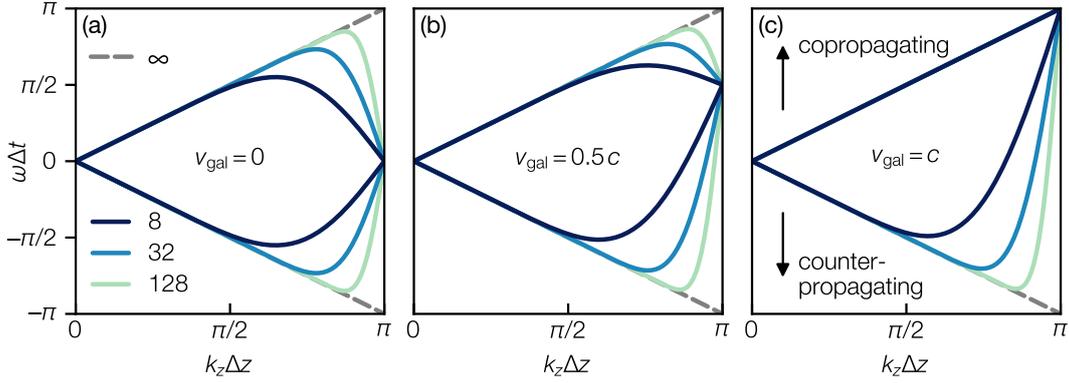


Figure 7.3 – Asymmetry of the numerical vacuum dispersion relation in one dimension. (a) The electromagnetic mode becomes distorted for finite orders of the field solver. (b) Increasing v_{gal} improves (deteriorates) the dispersion of copropagating (counterpropagating) waves. (c) At the limit $v_{\text{gal}} = c$, any spurious numerical dispersion is compensated in the direction of the Galilean frame.

For $v_{\text{gal}} = 0$, Eq. (7.17) reduces to the vacuum dispersion relation of the arbitrary-order PSATD solver [Eq. (5.37)] as defined in Sec. 5.3.4. As the solver becomes more local ($n_{\text{order}} = 8, 32, 128$), high frequency waves slow down and travel at a phase velocity $v_{\Phi} = \omega/k$ that is smaller than the vacuum speed of light.

As explained in Sec. 5.4, these slow electromagnetic modes couple resonantly to relativistic particles whenever the main NCI resonance condition [Eq. (5.39)],

$$\omega - k_z v_0 = 0, \quad (7.18)$$

is fulfilled. Recall that the term $k_z v_0$ represents an artificial beam (or streaming plasma) mode that describes the frequencies that are supported by relativistic particles that propagate along z at a velocity v_0 .

In simulations of relativistic beams, this unphysical resonance leads to the excitation of numerical Cherenkov radiation (NCR), whereas in simulations of relativistically streaming plasmas it causes a fast growing numerical Cherenkov instability (NCI).

However, once the comoving velocity v_{gal} is increased, the numerical dispersion relation of the Galilean-PSATD solver becomes asymmetrical. Increasing v_{gal} improves the dispersion of forward propagating waves (i.e., those having ω and k_z of the same sign) but worsens the dispersion of backward propagating waves. For $v_{\text{gal}} = v_0$, there will be no intersection between the electromagnetic mode and the beam mode. Inserting the vacuum dispersion relation, Eq. (7.17), into the resonance condition, Eq. (7.18), gives

$$c[k] - [k_z]v_0 = 0, \quad (7.19)$$

a condition which is never met. Moreover, at the limit $v_{\text{gal}} = c$, waves propagating in the direction of the comoving frame exhibit perfect physical dispersion regardless of the solver's order.

This unique property of the numerical dispersion relation is a direct consequence of the Galilean coordinate transformation and a fundamental prerequisite for the stability of the solver for finite orders. For $v_{\text{gal}} = v_0$, the asymmetry of the dispersion relation effectively suppresses any NCR in the direction of the beam or plasma and—in combination with the other stability properties—eliminates the NCI even in the finite-order case, as shown in Sec. 7.3.

The dispersion of backward propagating waves, on the other hand, worsens compared to the reference case ($v_{\text{gal}} = 0$), which has to be taken into account in practice. For example, it could be necessary to use a higher order n_{order} of the solver to mitigate problems from NCR in simulations that include a backward propagating relativistic beam. Nevertheless, we show in Ch. 8 that relatively low orders are still sufficient to not degrade such a counterpropagating beam in Lorentz-boosted frame simulations of plasma acceleration.

7.3 Theoretical stability analysis

7.3.1 Numerical dispersion relation of a cold relativistic plasma

In Sec. 7.2, we derived the arbitrary-order Galilean-PSATD solver and showed that the locality of the field evolution makes it suitable for parallelization. Furthermore, we have shown that even for a finite order, the asymmetric features of the vacuum dispersion relation mitigate the problem of NCR for a comoving relativistic beam or plasma.

In the following, we present a stability analysis that proves the elimination of the NCI for a relativistically streaming plasma. The analysis is based on a theoretical dispersion relation that includes the particle feedback of a moving plasma on the electromagnetic fields and that considers all of the approximations of the numerical scheme [55, 204, 205, 247]. The dispersion relation of the Galilean-PSATD solver was originally presented in [52] and was later extended to include the modified wavevector $[\mathbf{k}]$ [53].

For its derivation, we start with the discretized Vlasov equation of a drifting plasma and combine it with the discretized Maxwell equations. We then consider the evolution of small perturbations to the fields which are of the form

$$\mathbf{E}, \mathbf{B} \propto e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t} = e^{i\mathbf{k}\cdot\mathbf{x}' - (\omega - \mathbf{k}\cdot\mathbf{v}_{\text{gal}})t}, \quad (7.20)$$

where ω and \mathbf{k} relate to the natural interpretation of the fields that does not depend on \mathbf{v}_{gal} . From the coupled system one finally obtains the following expression, Eq. (7.21) with Eqs. (7.22a) to (7.23b), which represent the *numerical* dispersion relation of the arbitrary-order Galilean-PSATD solver, relating ω and \mathbf{k} , in the presence of a neutral, uniform plasma flowing at $\mathbf{v}_0 = v_0 \mathbf{u}_z$ (with Lorentz factor γ_0) on a periodic 2D Cartesian grid. The mathematical details of the derivation can be found in the Appendix of [52] and [53].

$$\begin{aligned} & (s_{\omega''}^2 - t_{c[k]}^2 c_{\omega''}^2) \left[1 - \frac{1}{[k]} \left([k_x] \xi_{3x} + \frac{[k_z] \xi_{3z}}{\gamma_0^2} \right) \right] - \xi_1 \left[\frac{\chi_5}{[k]^2} \left([k_z]^2 + \frac{[k_x]^2}{\gamma_0^2} \right) + \chi_5' \frac{[k_z] v_0}{c[k]} \right] \\ & + \frac{[k_x] v_0}{c[k]} \left[\frac{\chi_5}{[k]} \left([k_z] \xi_{2x} - \frac{[k_x] \xi_{2z}}{\gamma_0^2} \right) + \chi_5' \frac{\xi_{2x} v_0}{c} \right] \\ & + \frac{1}{\gamma_0^2} \left(\chi_5 + \frac{[k_z] v_0}{c[k]} \chi_5' \right) \left[\xi_1 \frac{[k_x] \xi_{3x} + [k_z] \xi_{3z}}{[k]} + \frac{[k_x] v_0}{c[k]} (\xi_{3x} \xi_{2z} - \xi_{3z} \xi_{2x}) \right] = 0 \end{aligned} \quad (7.21)$$

$$\xi_1 = \hat{\mathcal{T}}(\mathbf{k}) \frac{\omega_p^2}{\gamma_0 c[k]} \sum_{m_x, m_z = -\infty}^{\infty} \frac{1}{\frac{2}{\Delta t} \sin \left[\frac{\Delta t}{2} (\omega - k_z v_0 - m_z \frac{2\pi}{\Delta z} (v_0 - v_{\text{gal}})) \right]} \hat{\mathcal{S}}^2(\mathbf{k}_m) \quad (7.22a)$$

$$\xi_2 = \hat{\mathcal{T}}(\mathbf{k}) \frac{\omega_p^2}{\gamma_0 [k]} \sum_{m_x, m_z = -\infty}^{\infty} \frac{\cos \left[\frac{\Delta t}{2} (\omega - k_z v_0 - m_z \frac{2\pi}{\Delta z} (v_0 - v_{\text{gal}})) \right]}{\left(\frac{2}{\Delta t} \right)^2 \sin^2 \left[\frac{\Delta t}{2} (\omega - k_z v_0 - m_z \frac{2\pi}{\Delta z} (v_0 - v_{\text{gal}})) \right]} \hat{\mathcal{S}}^2(\mathbf{k}_m) \mathbf{k}_m \quad (7.22b)$$

$$\xi_3 = \hat{\mathcal{T}}(\mathbf{k}) \frac{\omega_p^2}{\gamma_0 [k]} \sum_{m_x, m_z = -\infty}^{\infty} \frac{1}{\left(\frac{2}{\Delta t} \right)^2 \sin^2 \left[\frac{\Delta t}{2} (\omega - k_z v_0 - m_z \frac{2\pi}{\Delta z} (v_0 - v_{\text{gal}})) \right]} \hat{\mathcal{S}}^2(\mathbf{k}_m) \mathbf{k}_m \quad (7.22c)$$

$$\chi_5 = \frac{c_{\omega''} c_{\nu c[k]}}{1 - \nu^2} (t_{\omega''} (t_{c[k]} - \nu t_{\nu c[k]}) - t_{c[k]} (t_{\nu c[k]} - \nu t_{c[k]})) \quad (7.23a)$$

$$\chi_5' = \frac{c_{\omega''} c_{\nu c[k]}}{1 - \nu^2} (t_{\omega''} (t_{\nu c[k]} - \nu t_{c[k]}) - t_{c[k]} (t_{c[k]} - \nu t_{\nu c[k]})) \quad (7.23b)$$

Note that we only consider the case where the Galilean transformation is along the direction of the moving plasma, i.e., $\mathbf{v}_{\text{gal}} = v_{\text{gal}} \mathbf{u}_z$ and that we use the following shorthand notations:

$$\begin{aligned} s_x &= \sin\left(\frac{x \Delta t}{2}\right), \quad c_x = \cos\left(\frac{x \Delta t}{2}\right), \quad t_x = \tan\left(\frac{x \Delta t}{2}\right), \\ \omega'' &\equiv \omega - v_{\text{gal}}(k_z - [k_z]), \quad \nu = \frac{[k_z] v_{\text{gal}}}{c[k]}. \end{aligned}$$

In fact, and as we have discussed in detail in [52], this rather complicated equation takes into account all numerical properties of the underlying algorithm, such as the finite temporal and spatial resolution, the discretization on the grid and the particularities of

the PIC cycle (e.g., the particle field gathering, charge and current deposition, current correction, etc.).

For a vanishing time step and cell size ($\omega\Delta t \ll 1$, $k\Delta x \ll 1$, $k\Delta z \ll 1$) and for any value of v_{gal} , it can be shown that Eq. (7.21) reduces to [52]

$$\frac{\Delta t^2}{4} \times \left(\omega^2 - c^2 k^2 - \frac{\omega_p^2}{\gamma_0} \right) \times \left(1 - \frac{\omega_p^2}{\gamma_0^3 (\omega - k_z v_0)^2} \right) = 0. \quad (7.25)$$

In the limit of infinite resolution, the dispersion relation thus supports the two physical modes of the relativistic plasma, the relativistic electromagnetic mode $\omega^2 = c^2 k^2 + \omega_p^2 / \gamma_0$ and the relativistic plasma mode $\omega = k_z v_0 \pm \omega_p \gamma_0^{-3/2}$. For finite resolutions, on the other hand, it recovers all of the unphysical couplings between distorted numerical modes that can lead to the numerical Cherenkov instability.

Especially the coefficients ξ are of interest, as they represent the numerical plasma response, taking into account the particle shape factor $\hat{\mathcal{S}}(\mathbf{k})$ [Eq. (5.10)], the spatial field smoothing $\hat{\mathcal{T}}(\mathbf{k})$ [Eq. (5.20)] and a sum over all spatial aliases \mathbf{K}_m ($\mathbf{K}_m = m_x \frac{2\pi}{\Delta x} \hat{\mathbf{u}}_x + m_z \frac{2\pi}{\Delta z} \hat{\mathbf{u}}_z$ and $\mathbf{k}_m = \mathbf{k} + \mathbf{K}_m$). The latter arise from the mismatch of sampling the continuous particle distribution on a discrete grid and play a key role in the occurrence of the NCI. It should again be noted that for infinite order, $[\mathbf{k}] \rightarrow \mathbf{k}$ and Eq. (7.21) then equals to the original dispersion relation as derived in [52]. Furthermore, setting $v_{\text{gal}} = 0$ recovers the dispersion relation of the standard PSATD solver that would be equivalent to the one formerly derived in [247].

7.3.2 Evaluation of the NCI growth rate

In the following section, we begin with an analytical evaluation of the dispersion relation, which closely follows [247] and reveals some of the sources of NCI growth. This analysis, however, is limited and Eq. (7.21) has to be solved numerically to obtain the NCI growth rate across all frequencies. Consequently, the results of this numerical evaluation are compared to actual 2D PIC simulations with WARPX. In the following, we consider three variants of the solver and use the same test case as in [52, 204, 247] with the exact parameters listed in Table 7.1.

For the sake of simplicity, we restrict ourselves to using a finite stencil in z only, thus $[k] = \sqrt{[k_z]^2 + k_x^2}$. Nevertheless, the main findings, such as the elimination of the NCI, are equally valid when using finite-order stencils along multiple axes.

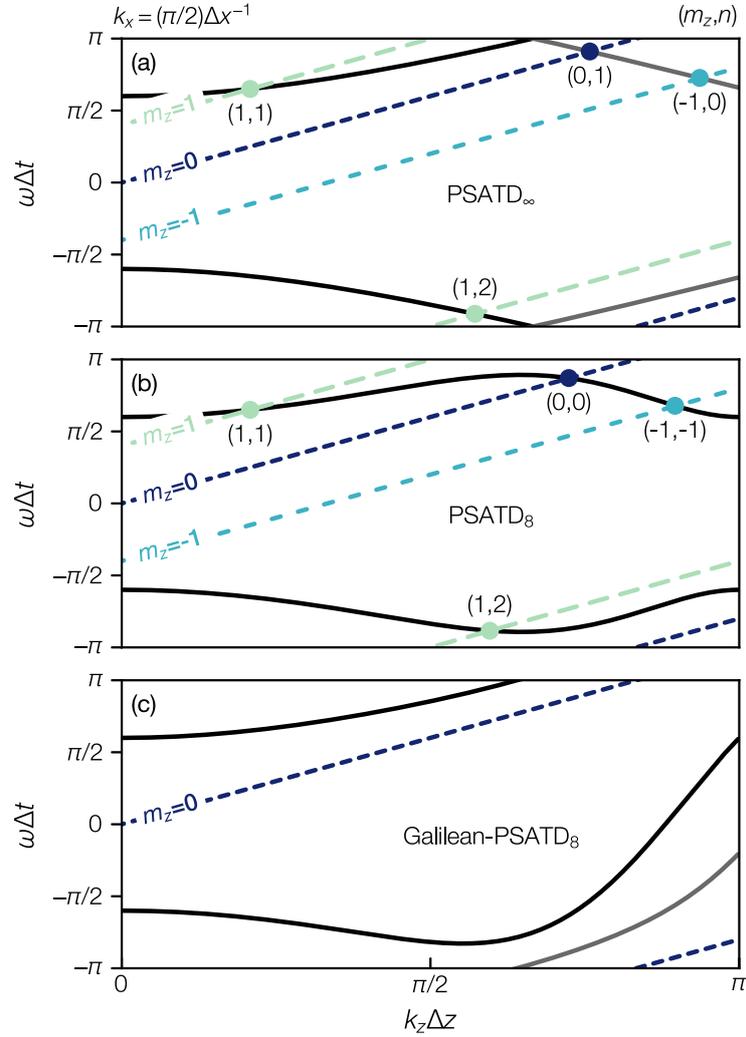


Figure 7.4 – Distorted electromagnetic modes (black solid lines) and temporal aliases (gray solid lines) for forward and backward ($\pm\omega\Delta t$) propagating waves and spurious beam modes (dashed lines) for different variants of the solver (see Table 7.1), at $k_x = (\pi/2)\Delta x^{-1}$. (a) PSATD $_{\infty}$; (b) PSATD $_8$; (c) Galilean-PSATD $_8$. The relativistic beam (or plasma) and the Galilean frame (panel (c) only) move in the forward ($+\omega\Delta t$) direction. Resonant growth of the NCI can occur at intersections of these modes and their spatial (m_z) and temporal (n) aliases.

Table 7.1 – Variants of the solver and parameters of the NCI growth rate simulations, scaled by $k_{p,r} \equiv k_p/\gamma_0^{1/2}$, with $k_p^2 = n_0 e^2/m_e \epsilon_0 c^2$.

PSATD _∞	$n_{\text{order}} = \infty, v_{\text{gal}} = 0$
PSATD ₈	$n_{\text{order}} = 8, v_{\text{gal}} = 0$
Galilean-PSATD ₈	$n_{\text{order}} = 8, v_{\text{gal}} = v_0$
Plasma density	n_0 (scales the simulation)
Lorentz factor	$\gamma_0 = 130$
Cell size	$\Delta z = \Delta x = 0.3868 k_{p,r}^{-1}$
Time step	$\Delta t = 1.2\Delta z/c$
Number of grid points	$N_x = N_z = 200$
Particle shape factor	3 (cubic)

Analytical evaluation

To begin with the analytical evaluation, it is useful to discard the plasma related terms of the dispersion relation and thereby recover the vacuum dispersion relation, i.e., the very first term of Eq. (7.21),

$$\tan^2\left(\frac{(\omega - v_{\text{gal}}(k_z - [k_z]))\Delta t}{2}\right) = \tan^2\left(\frac{c[k]\Delta t}{2}\right). \quad (7.26)$$

Although the vacuum dispersion relation was already given by Eq. (7.17) and discussed in Sec. 7.2, it now becomes evident that the above equation supports an infinite number of temporal aliases as seen by the particles,

$$\omega_n = \omega + n \frac{2\pi}{\Delta t}, \quad (7.27)$$

with n the temporal alias number. Even for the standard PSATD solver with infinite order accuracy (PSATD_∞), this causes a distortion of the electromagnetic mode at large k if the time step Δt is too large. This is because particles sample the fields at discrete times and at maximum see a frequency of $\pm\pi/\Delta t$. If the actual frequency of the fields supported by the grid exceeds this value, particles see a temporal alias of the electromagnetic mode. This can be seen in Fig. 7.4(a), where the electromagnetic mode (solid lines) of the PSATD_∞ solver is shown for a fixed transverse wavenumber $k_x = (\pi/2)\Delta x^{-1}$. In our example, $c\Delta t = 1.2\Delta z > \pi/k_{\text{max}}$, where $k_{\text{max}} = \pi(\Delta z^{-2} + \Delta x^{-2})^{1/2}$ is the maximum wavenumber of the grid. Hence, particles see a temporal alias of the fields (gray lines) and ω decreases once the frequencies supported by the grid, $c[k]$ exceed the maximum frequency supported by the time step, $\pi c/(1.2\Delta z)$. Conversely, in the finite order case (PSATD₈), the main electromagnetic mode is always distorted and eventually starts decreasing at high k_z .

These distorted electromagnetic modes can couple to spurious plasma or beam modes that emerge from the resonance condition in the plasma coefficients ξ ,

$$\omega + n \frac{2\pi}{\Delta t} - k_z v_0 - m_z \frac{2\pi}{\Delta z} (v_0 - v_{\text{gal}}) = 0, \quad (7.28)$$

i.e., whenever the sine-term in the denominator of these expressions becomes zero. Here, m_z is the spatial alias number in the direction of the moving plasma, and n is the index of the temporal alias. Not surprisingly, for $m_z = 0$ and $n = 0$ we get the NCR resonance condition that was already mentioned in Sec. 7.2.5, Eq. (7.18). The $m_z = 0$ beam mode and the first two spatial aliases $m_z = [-1, +1]$ are shown in Fig. 7.4. Intersections of these beam modes—or their temporal aliases—with the electromagnetic mode can support resonant growth of the NCI.

As expected, the $m_z, n = (0, 0)$ resonance only occurs for finite orders (PSATD₈), while the standard PSATD solver (PSATD_∞) solely suffers from aliased resonances. By setting $v_{\text{gal}} = v_0$ in Eq. (7.28) all resonances with spatial aliases $m_z \neq 0$ vanish [Fig. 7.4(c)], which is a direct consequence of the numerical grid following the moving plasma in the Galilean frame. In addition, the asymmetrical dispersion relation prevents the $m_z = 0$ beam mode to intersect with the main electromagnetic mode in case of the finite-order Galilean-PSATD solver (Galilean-PSATD₈).

Despite of that, Eq. (7.28) suggests that an NCI resonance with $m_z = 0$ and a temporal alias $n \neq 0$ of the fields could still occur for any order of the Galilean-PSATD solver. For example, the $m_z, n = (0, 1)$ intersection in Fig. 7.4(a) does not vanish for $v_{\text{gal}} = v_0$. However, as we will learn from the numerical analysis in the following, the ν -dependent terms in the coefficients χ_5 and χ'_5 seem to cancel this resonance for $v_{\text{gal}} = v_0$ and also prevent the occurrence of an otherwise typical non-resonant growth of the NCI for $m_z = 0$. These additional terms are a direct consequence of the analytical integration when deriving the PSATD equations and come from the fact that the currents are assumed to be constant over one time step in the comoving frame, which is equivalent to assume the currents to be comoving with the plasma in the original frame.

Numerical evaluation

Again, those final observations have been made by numerically solving Eq. (7.21), because it is not easily possible to extract more information from the analytical expressions. Fig. 7.5 shows the theoretical NCI growth rates $\text{Im}(\omega)$ for the three solver cases and compares them to WARPX simulations. The possible curves of resonant growth are shown for the different beam modes and are labeled according to their spatial and temporal alias number. The k_x locations of intersecting modes can be

derived as a function of k_z by inserting the vacuum dispersion relation Eq. (7.17) into the resonance condition Eq. (7.28) and solve it for k_x [247]:

$$k_{x,\text{res}} = \left[\left(k_z \frac{v_0}{c} - (k_z - [k_z]) \frac{v_{\text{gal}}}{c} + m_z \frac{2\pi}{\Delta z} \frac{(v_0 - v_{\text{gal}})}{c} - n \frac{2\pi}{c\Delta t} \right)^2 - [k_z]^2 \right]^{1/2}. \quad (7.29)$$

The standard PSATD_∞ case suffers from non-resonant growth of the NCI over a wide range of frequencies and in addition shows resonant growth along the $m_z, n = (0, 1)$ and $m_z, n = (-1, 0)$ resonance lines. A very similar pattern of growth can be seen for the finite-order PSATD₈ case, although here three lines of resonant growth, $m_z, n = (0, 0)$, $(0, 1)$ and $(-1, -1)$ are visible, with the $m_z, n = (0, 0)$ resonance being the dominant one. In contrast, numerically solving the dispersion relation for the Galilean-PSATD₈ case predicts absolutely no growth across all frequencies, proving the absence of the NCI for a uniformly streaming plasma. The estimated growth rates from the PIC simulations validate the above results and show excellent quantitative agreement for all three cases. Note that the visible residual background can be attributed to numerical noise in the simulation.

Although not shown here, another useful conclusion can be drawn by slightly modifying the dispersion relation before numerically solving it. Explicitly replacing $[k_z]$ by k_z in ω'' and ν removes the asymmetric properties of the dispersion relation and thereby reintroduces the $m_z, n = (0, 0)$ resonance for the otherwise stable case with $v_{\text{gal}} = v_0$. Indeed, when numerically solving this modified dispersion relation, NCI growth appears only along the $m_z, n = (0, 0)$ resonance line. This again highlights the self-consistency of the Galilean approach. Despite being a fundamental ingredient to cancel the resonance, the asymmetry of the vacuum dispersion relation was not artificially introduced to the scheme, but rather it is a natural consequence of deriving the PIC equations in the comoving coordinates.

In summary, the stability analysis unfolded the different sources of NCI growth and we demonstrated how the Galilean coordinate transformation introduces unique properties that lead to their compensation. Consequently, we proved the elimination of the numerical Cherenkov instability for a finite, localized stencil in the case of a cold relativistic plasma moving uniformly in one direction. In the next chapter, the practical usefulness of the solver is validated by applying it to simulations of laser-plasma acceleration using the Lorentz-boosted frame technique.

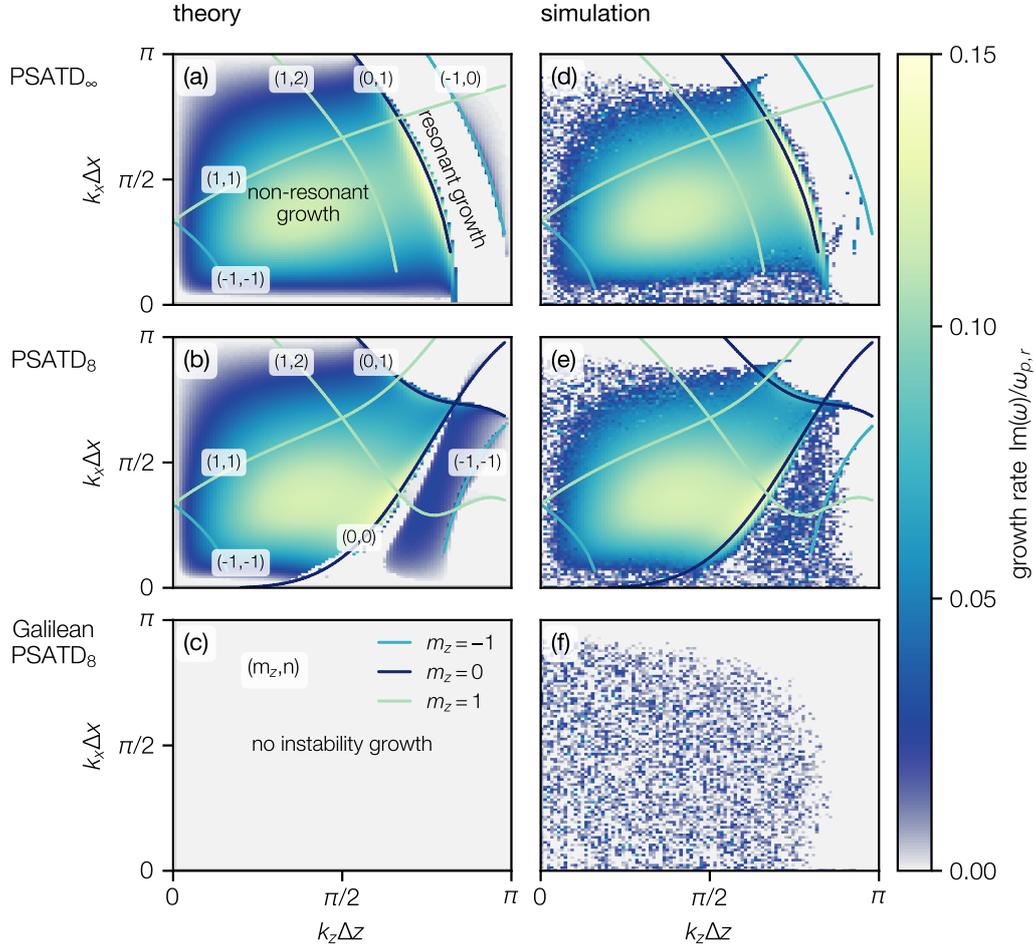


Figure 7.5 – (a)-(c) Theoretically predicted and (d)-(f) simulated NCI growth rates $\text{Im}(\omega)/\omega_{p,r}$ with the parameters as defined in Table 7.1. (a), (d) PSATD $_{\infty}$; (b), (e) PSATD $_8$; (c), (f) Galilean-PSATD $_8$. Theoretical growth rates are predicted by numerically solving Eq. (7.21) and simulated growth rates are estimated from the difference of the Fourier transformed fields between $\omega_{p,r}t \simeq 90$ and 120. Mode intersections that support resonant growth of the NCI are calculated from Eq. (7.29) and are overlaid as solid lines and labeled according to their spatial (m_z) and temporal (n) alias number.

8 Application to plasma acceleration

This chapter presents the application of the developed particle-in-cell methods to simulations of plasma acceleration. In Sec. 8.1, we review the Lorentz-boosted frame technique and demonstrate the stability and accuracy of the Galilean-PSATD solver in practice. The section is in part a reproduction of [53]. In Sec. 8.2, we consider a realistic simulation case and demonstrate a three orders of magnitude reduction in computational cost using FBPIC over a conventional full-3D particle-in-cell algorithm, without sacrificing in accuracy.

8.1 Stable Lorentz-boosted frame simulations

8.1.1 Lorentz-boosted frame technique

Particle-in-cell simulations of plasma accelerators are costly. Spatially resolving the smallest scales, such as the drive laser wavelength, demands fine grid resolutions. Consequently, many cells are required to cover the much larger plasma.

In the axial direction, the requirements on the grid size can be relaxed by sliding a window across the interaction region, such that the simulation box always only covers the small region of interest that includes the laser and the immediate plasma wave behind it.

Nonetheless, a disparity of longest to shortest space and time scales remains, because the total plasma length is usually orders of magnitude longer than the maximum allowed propagation distance within one time step. Hence, many iterations are required to simulate the complete interaction.

These considerations are valid in the laboratory frame which is the usual frame of reference for simulations. However, because we are only interested in resolving physical objects that propagate at relativistic velocities in a single direction, a much more efficient ratio of scales can be achieved by transforming the entire simulation from the laboratory frame to a relativistic frame of reference moving in the direction of the driver [49],

$$\begin{aligned} ct' &= \gamma(ct - \beta z), \\ z' &= \gamma(z - \beta ct), \end{aligned} \tag{8.1}$$

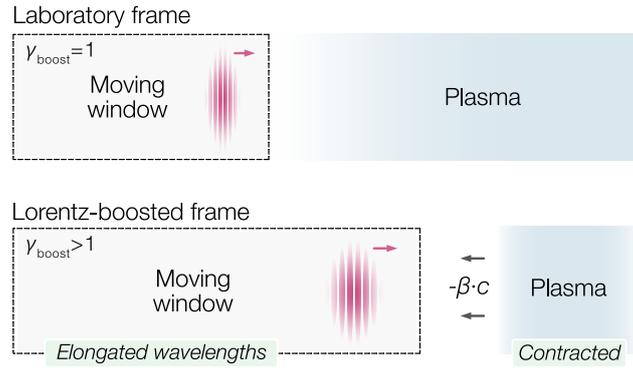


Figure 8.1 – Lorentz-boosted frame technique. In the laboratory frame ($\gamma = 1$), the simulation is usually confined to a short moving window that follows the laser-plasma interaction through a much longer plasma. When transforming such a system to a relativistic frame of reference ($\gamma > 1$), the copropagating objects elongate, whereas the plasma contracts and counterpropagates at a relativistic velocity $-\beta c$. This results in a much more efficient ratio of scales, which reduces the total number of iterations $\propto \gamma^2$.

where $\gamma = (1 - \beta^2)^{-1/2}$ and $\beta = v_{\text{boost}}/c$ is the speed of the relativistic frame along the axis of propagation z . The concept is illustrated in Fig. 8.1.

In this Lorentz-boosted frame, the length scales of copropagating relativistic objects, such as the laser pulse or the plasma wave, are elongated by a factor $\simeq \gamma(1 + \beta)$. In contrast, objects that have been initially at rest, such as the plasma, are contracted by a factor γ and counterpropagate at a relativistic velocity $v_{\text{plasma}} = -\beta c$ in the opposite direction.

As a consequence, the total simulation length decreases. At the same time the propagation distance per time step increases, because the cell size is elongated while the number of grid cells remains constant, i.e., the relative spatial resolution of the copropagating objects is unchanged. The combination of both effects reduces the overall number of iterations to complete the simulation by $\simeq (1 + \beta)^2 \gamma^2$.

Note that this estimate is approximately valid for $\gamma \ll \gamma_{\text{wake}}$, where γ_{wake} is the relativistic factor associated with the phase velocity of the plasma wave. More details on the calculation and limitations of the speedup are found in [209].

Meaningful choices for γ range between 2 and 100 for typical setups of plasma acceleration. The Lorentz-boosted frame technique thus allows for orders of magnitude speedups and a drastic reduction of computational costs in practice.

However, the applicability of the technique with conventional Maxwell solvers is limited because the counterstreaming plasma eventually gives rise to a numerical Cherenkov instability (see also Sec. 5.4). As a solution, we can use the Galilean-PSATD solver

Table 8.1 – Parameters of the laser-plasma acceleration simulation in cylindrical coordinates, decomposed into two azimuthal modes. Laser and electron beam quantities have Gaussian distributions.

On-axis plasma density	$n = 5 \cdot 10^{17} \text{ cm}^{-3}$
Laser wavelength & intensity ¹	$\lambda = 800 \text{ nm}, a_0 = 2$
Laser waist ² & pulse length ²	$w_0 = 20 \text{ }\mu\text{m}, c\tau = 12 \text{ }\mu\text{m}$
e ⁻ -beam energy	$\gamma = 2000, \sigma_\gamma = 10$
e ⁻ -beam charge	$Q = 100 \text{ pC}$
e ⁻ -beam size	$\sigma_z = 1.5 \text{ }\mu\text{m}, \sigma_r = 0.5 \text{ }\mu\text{m}$
e ⁻ -beam norm. emittance	$\epsilon_n = 1 \text{ mm mrad}$
Resolution	$\Delta z = \lambda/64, \Delta r = 0.25 \text{ }\mu\text{m}$
Time step	$\Delta t = \Delta z/c$
Particles per cell & shape	$2 \times 4 \times 4 (r, z, \theta), 3 \text{ (cubic)}$

¹peak normalized vector potential

² 2σ of intensity distribution

that was presented in Ch. 7 to perform the simulation in a coordinate system that is comoving with the relativistic plasma, hence $v_{\text{gal}} = -\beta c$, and thereby eliminate the NCI.

8.1.2 NCI suppression with the Galilean-PSATD solver

In the following, we show FBPIC simulations of a nonlinear plasma wakefield in the blowout regime that is used to double the energy of a 1 GeV electron beam with 100 pC of charge over a distance of $\simeq 30 \text{ mm}$. The plasma wave is resonantly driven by a laser pulse that is kept in focus by a matched plasma guiding channel.

The setup has been simulated in the laboratory frame and in a Lorentz-boosted frame with $\gamma_{\text{boost}} = 10$. The simulation parameters are summarized in Table 8.1. Except for the infinite-order simulation, the computation is parallelized to 8 GPUs by spatial domain decomposition in the axial direction.

Fig. 8.2 compares the stability of the Lorentz-boosted frame simulation for the three different variants of the solver in accordance with Sec. 7.3. At the time shown, the laser pulse, propagating to the right, has already reached the end of the counterstreaming plasma. The trailing plasma wave and the accelerated electron beam are visible as modulations of the charge density ρ' . While the onset of an NCI is visible in case of the standard PSATD_∞ solver, a strong NCI has already developed for the finite-order PSATD₈ solver. The distorted dispersion relation leads to a higher growth rate of the instability in this case.

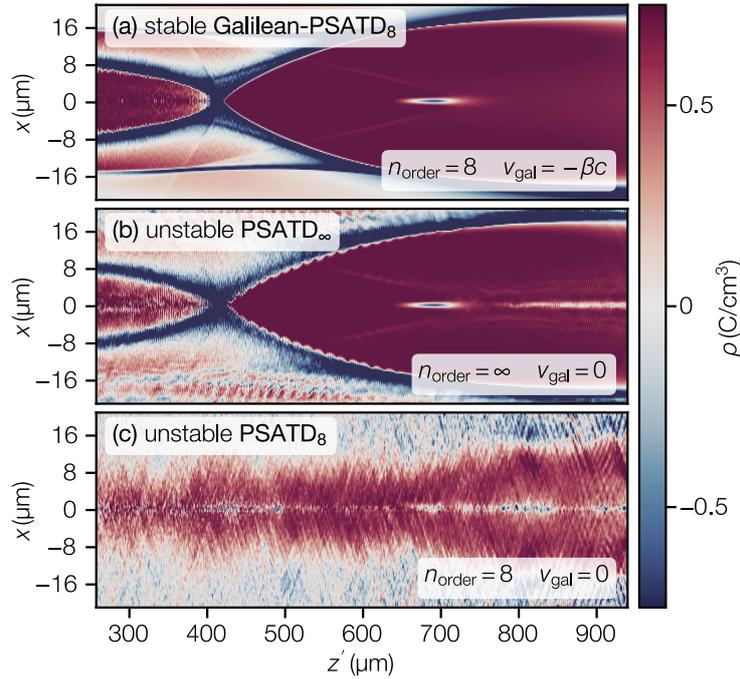


Figure 8.2 – Stability of the laser-plasma acceleration simulation in a relativistic frame with $\gamma_{\text{boost}} = 10$. The charge density ρ' is shown at $t' = 7800 \Delta z'/c$ for different variants of the solver (see Table 7.1). (a) Galilean-PSATD₈; (b) PSATD_∞; (c) PSATD₈. Simulation parameters are listed in Table 8.1. A destructive NCI develops unless setting $v_{\text{gal}} = v_{\text{plasma}}$.

In contrast, the Galilean-PSATD₈ solver remains completely stable for the entire simulation distance, while the only difference here is the use of $v_{\text{gal}} = v_{\text{plasma}}$. Although the excitation of a strong nonlinear plasma wave causes deviations from the initially uniform plasma flow, no signs of an NCI become apparent. The high density electron bunch that moves opposite to the Galilean frame is not degraded by NCR or other detrimental numerical effects. Despite the presence of a distorted vacuum dispersion relation in this direction, the accuracy of the solver for the relatively low order $n_{\text{order}} = 8$ is already sufficient to mitigate NCR [44] in this use case.

The results of this finite-order Lorentz-boosted frame simulation are back-transformed to the laboratory frame and benchmarked against a high accuracy ($n_{\text{order}} = 128$) reference simulation with $\gamma_{\text{boost}} = 1$, i.e., a lab-frame simulation.

Fig. 8.3 shows a direct comparison of the longitudinal wakefield E_z and the charge density ρ towards the end of the simulation after $\simeq 25$ mm of propagation. The guided laser expels almost all electrons from the axis and forms a trailing plasma cavity sustaining tens of GeV/m accelerating gradients. Within the blowout, the high charge electron bunch drives its own wakefield and absorbs energy from the wake. The

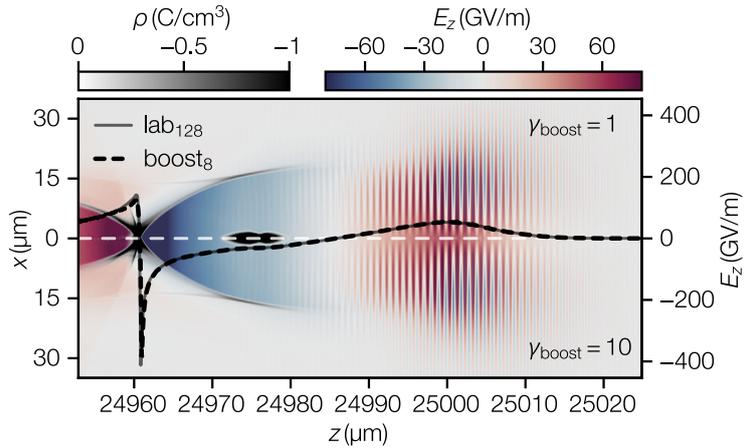


Figure 8.3 – Accelerating wakefields E_z (zx -plane at $y = 0$ and on-axis lineout) and overlaid charge density ρ after $\simeq 25$ mm of propagation. Results from the laboratory frame simulation ($n_{\text{order}} = 128$ and $\gamma_{\text{boost}} = 1$) are shown in the upper half (solid line) and results from the Lorentz-boosted frame simulation ($n_{\text{order}} = 8$ and $\gamma_{\text{boost}} = 10$) are shown in the lower half (dashed line), respectively.

beam loading flattens the accelerating fields along the bunch and thereby minimizes the accumulation of linearly correlated energy spread. The on-axis accelerating field agrees well between both simulations with only small deviations near the region of highest density at the back of the first plasma oscillation where the electron trajectories collapse.

A comparison of the evolution of characteristic laser and electron beam properties is shown in Fig. 8.4. The laser is initially focused $250 \mu\text{m}$ behind the entrance of the plasma channel into a 1 mm long sinusoidal density upramp. As it reaches relativistic intensities the nonlinear interaction with the plasma causes the pulse to self-modulate. As a consequence, the pulse length τ decreases and the transverse waist w oscillates along the propagation distance.

The electron bunch is initially focused at the end of the upramp and situated at the back of the first plasma oscillation, $32.5 \mu\text{m}$ behind the centroid of the laser. While slowly dephasing in the direction of the driver, it gets accelerated to $E_{\text{kin}} \simeq 2.1 \text{ GeV}$ and accumulates a relative energy spread of $\sigma_E \simeq 1.4\%$. A mismatch between the transverse beam phase space and the strong focusing forces leads to betatron decoherence causing a slight growth of the emittance that eventually saturates to $\varepsilon_n \simeq 1.3 \text{ mm mrad}$.

Both the nonlinear laser-plasma interaction and the electron beam dynamics are correctly reproduced in the Lorentz-boosted frame. At the end of the propagation distance, the shown quantities deviate by less than a percent between both simulations. The agreement is confirmed by a direct comparison of the longitudinal and transverse phase spaces at $z_{\text{prop}} = 30 \text{ mm}$.

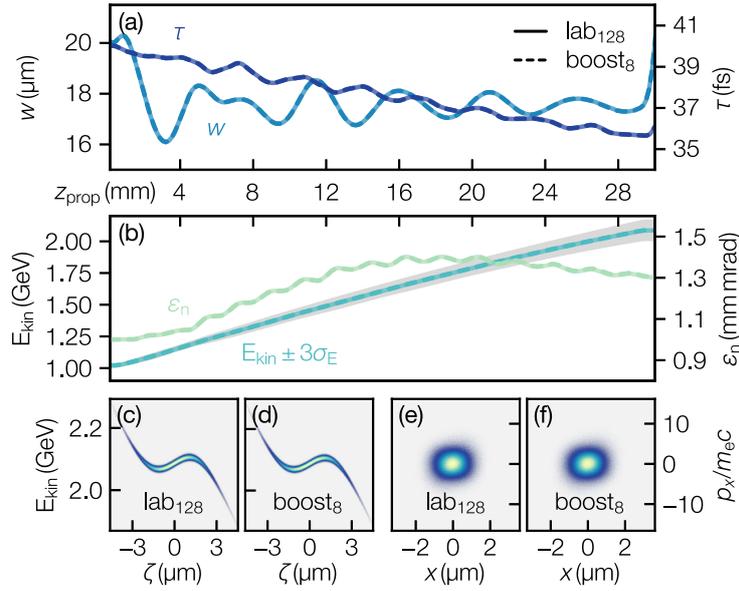


Figure 8.4 – Comparison of the results between the laboratory frame simulation with $n_{\text{order}} = 128$ and $\gamma_{\text{boost}} = 1$ (solid lines) and the Lorentz-boosted frame simulation with $n_{\text{order}} = 8$ and $\gamma_{\text{boost}} = 10$ (dashed lines). Evolution of (a) the laser waist w and pulse length τ and (b) the electron beam energy $E_{\text{kin}} \pm 3\sigma_E$ and emittance ε_n . Comparison of the (c), (d) longitudinal (ζ , E_{kin}) and (e), (f) transverse phase space (x , $p_x/m_e c$) of the electron bunch at the end of the acceleration distance, $z_{\text{prop}} = 30$ mm.

Summary

In summary, the results prove the applicability of the finite-order Galilean-PSATD solver for typical simulations of plasma acceleration in the Lorentz-boosted frame. For the example case considered here, the required number of PIC iterations is reduced by two orders of magnitude, while the use of a local stencil of order $n_{\text{order}} = 8$ allows for efficient parallelization by domain decomposition.

Although the excitation of a plasma wave leads to deviations from the uniform streaming velocity, we demonstrated that the solver remains stable and accurate. Furthermore, even when using low orders, a relativistic high-density beam that counterpropagated with respect to the Galilean frame did not show any signs of an NCI and was not degraded by NCR.

Nevertheless, large deviations from the bulk plasma velocity or simulations of counter-streaming plasmas, such as studied in astrophysics, can again lead to an instability. Likewise, a dense relativistic beam that counterpropagates with respect to the Galilean frame can still degrade from NCR when using very low orders and grid resolutions. In

these situations, however, the simultaneous combination of multiple comoving frames in more than one direction is a possible solution.

Apart from suppressing NCR, simulations including relativistic beams could further benefit from a Galilean frame. Recent work [256] showed that even for an ideal numerical dispersion and in the absence of Cherenkov-like fields, simulations of relativistic beams can suffer from erroneous space charge fields. However, by modeling such systems in a frame comoving with the beam, it should be possible to recover the correct physical self-fields with the arbitrary-order Galilean-PSATD solver.

8.2 Comparison of FBPIC to a conventional 3D algorithm

FBPIC promises to reduce the computational cost and increase the numerical accuracy and stability of simulations of plasma acceleration. In this final section, we present a comprehensive benchmark of the code by comparing it to a conventional 3D Cartesian FDTD algorithm that is implemented in the particle-in-cell code WARPX [255].

Based on a practical example, we demonstrate the validity of decomposing the three-dimensional physics into a few azimuthal modes and show that this leads to order of magnitude savings in computational cost. We prove again that the physics are preserved when applying the Lorentz-boosted frame technique, while further decreasing the run time of the simulation. In combination, we demonstrate a speedup by three orders of magnitude when compared to a full-3D reference simulation in the laboratory frame, without sacrificing accuracy.

Simulation setup

For the comparison, we present simulations of a laser-plasma accelerator that is operated in the blowout regime. An electron beam is generated from localized ionization injection and gets accelerated to an energy of ~ 300 MeV over a distance of 2 mm. The assumed laser and plasma parameters are similar to those of common experimental setups.

Figure 8.5 shows the gas and plasma density distribution in the simulation. A plateau of hydrogen gas (H_2) extends over a length of 1.6 mm with 0.2 mm long sinusoidal ramps at the beginning and end of the profile. In a first section, the gas is mixed with 0.75% nitrogen ($\text{N}_2 + \text{H}_2$) over a length of 400 μm . The gas density throughout the plateau region is $1 \times 10^{18} \text{ cm}^{-3}$, which, once the molecules are ionized by the laser, results in an electron density of $n_e = 2 \times 10^{18} \text{ cm}^{-3}$. The plasma density is slightly increased in the mixed gas region due to the five times higher electron yield from the outer shell of nitrogen.

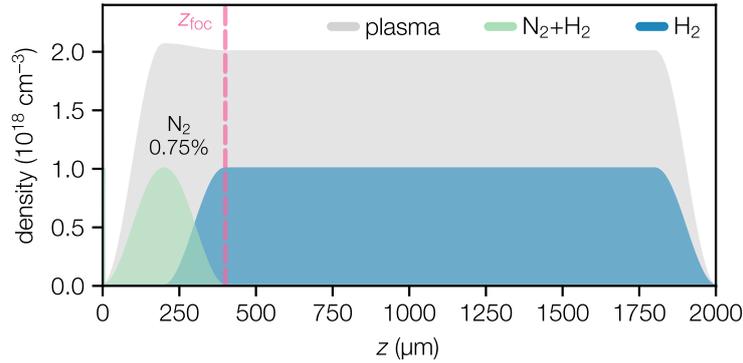


Figure 8.5 – Gas and plasma density profile of the simulation used to compare FBPIC and WARPX 3D. Plasma (gray), mixture of hydrogen and 0.75% nitrogen (green), pure hydrogen (blue); Laser focus position $z_{\text{foc}} = 0.4$ mm (magenta).

The plasma wave is driven by a Gaussian laser pulse ($\tau = 30$ fs) that is focused to a spot size of $w_0 = 15$ μm at the end of the mixed gas region ($z_{\text{foc}} = 0.4$ mm). The peak normalized vector potential in focus is $a_0 = 3$, which corresponds to a pulse energy of ~ 2.25 J. As the laser enters the plasma, the innermost electrons of the nitrogen atoms are ionized in regions of highest intensity and get injected into the plasma wave. The generated electron bunch is subsequently accelerated in the pure hydrogen section.

The 3D Cartesian reference simulation in the laboratory frame was performed with WARPX (WarpX 3D_{lab}), using the non-standard Cole-Karkkainen FDTD solver with Cowan coefficients (CKC) [196, 198, 41] and one macroparticle per cell. The longitudinal and transverse resolution was $\Delta z^{-1} = 60$ μm^{-1} and $\Delta x^{-1} = \Delta y^{-1} = 3$ μm^{-1} , respectively. The grid had a size of $300 \times 300 \times 3000$ (x, y, z) cells.

The FBPIC simulations were performed using the (Galilean-)PSATD solver with order $n_{\text{order}} = 8$ and $1 \times 2 \times 6$ (r, z, θ) particles per cell. Using the same spatial resolution, the grid had a size of 150×3000 (r, z) cells. Although the initial setup can be represented by only two azimuthal modes, three azimuthal modes were required to correctly reproduce the transverse structure of the injected beam. The quasi-3D simulation was performed in the laboratory frame (FBPIC_{lab}) and in a Lorentz-boosted frame with $\gamma_{\text{boost}} = 4$ (FBPIC_{boost}).

Both algorithms used cubic particle shape factors and reflective boundaries in the transverse direction. The time step was set to $c\Delta t = \Delta z$. Optical field ionization was modeled using the Ammosov-Delone-Krainov (ADK) [99, 102] model. Except for the inner two levels of nitrogen, the gas atoms were assumed to be pre-ionized.

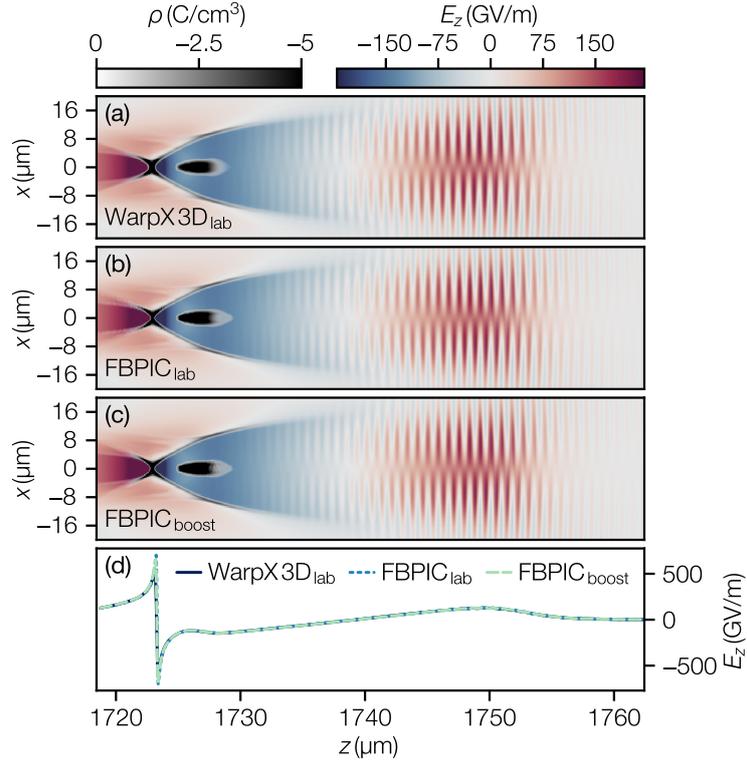


Figure 8.6 – Comparison of the plasma wakefield and the accelerated beam profile between a 3D Cartesian (WarpX 3D_{lab}), quasi-3D (FBPIC_{lab}), and quasi-3D (FBPIC_{boost}) Lorentz-boosted frame ($\gamma_{\text{boost}} = 4$) simulation. (a)-(c) Accelerating wakefield E_z (zx -plane at $y = 0$) and overlaid charge density ρ after ~ 1.75 mm of propagation. (d) On-axis lineouts of the longitudinal electric field E_z .

Accuracy of the physical results

Fig. 8.6(a)-(c) show 2D snapshots of the longitudinal electric field E_z overlaid with the charge density ρ towards the end of the acceleration process. All three simulations reproduce the structure of the wakefield and the injected beam profile with remarkable accuracy. The injected beam charge (146 pC) and peak current (18 kA) deviate by less than a percent between the different cases. The excellent agreement is confirmed by a direct comparison of the on-axis accelerating fields in Fig. 8.6(d).

The evolution of key laser and beam parameters is compared in Fig. 8.7. The relativistic interaction with the plasma leads to self-focusing of the laser and a shortening of the pulse length. This further increases the peak intensity and thereby the nonlinearity of the plasma wave throughout the acceleration process. The electron bunch is injected after ~ 250 μm and accelerated to a final energy of 324 MeV. The injected current profile leads to optimized beam loading conditions which limit the relative rms energy

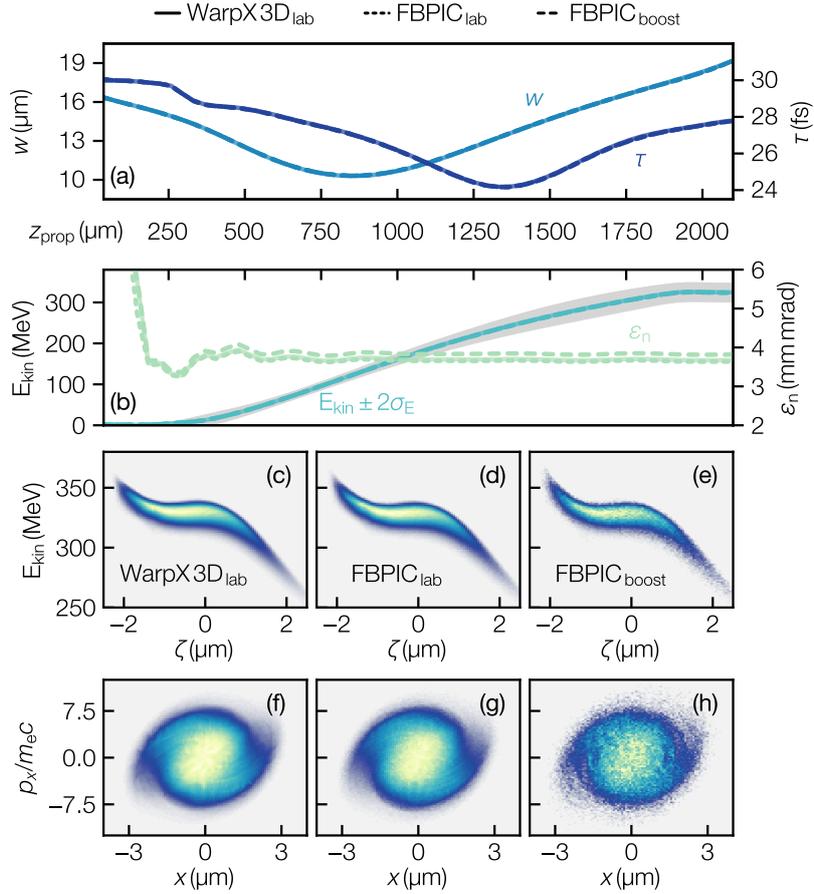


Figure 8.7 – Comparison of laser and beam parameters between a 3D Cartesian (WarpX 3D_{lab}), quasi-3D (FBPIC_{lab}), and quasi-3D (FBPIC_{boost}) Lorentz-boosted frame ($\gamma_{\text{boost}} = 4$) simulation. Evolution of (a) the laser waist w and pulse length τ and (b) the electron beam energy $E_{\text{kin}} \pm 2\sigma_E$ and emittance ε_n . Comparison of the (c)-(e) longitudinal (ζ , E_{kin}) and (f)-(h) transverse phase space (x , $p_x/m_e c$) of the electron bunch at the end of the acceleration distance, $z_{\text{prop}} \simeq 2$ mm.

spread to 3.6%. The normalized emittance in the laser polarization plane settles to ~ 3.6 mm mrad. Except for slight deviations in the evolution of the emittance in the Lorentz-boosted frame simulation, the agreement between the simulations is outstanding for all parameters.

The final phase space of the accelerated bunch is compared in Fig. 8.7(c)-(h). The core of the longitudinal phase space is flattened from beam loading, which reduces the projected energy spread. In the transverse dimension, the phase space shows a slightly complex structure at the radial edges. Remarkably, the different simulations agree well even for the most subtle details of the distributions.

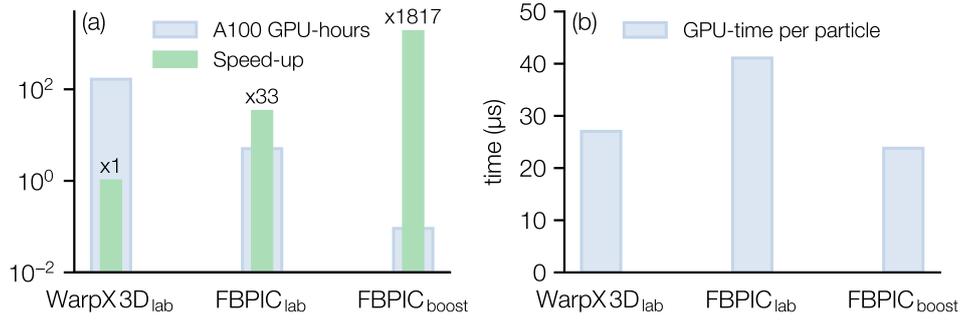


Figure 8.8 – Comparison of the computational cost between a 3D Cartesian (WarpX 3D_{lab}), quasi-3D (FBPIC_{lab}), and quasi-3D (FBPIC_{boost}) Lorentz-boosted frame ($\gamma_{\text{boost}} = 4$) simulation. (a) Computational cost measured in Nvidia A100 GPU hours (blue) and speedup (green) compared to the 3D reference simulation. (b) Total single-GPU computing time spent per macroparticle for the entire simulation.

Note that the number of macroparticles in the beam is reduced by a factor of 9 going from 3D to quasi-3D and by another factor of 31 for the Lorentz-boosted frame simulation. This is expected as the total number of macroparticles in the plasma is reduced by $2\gamma_{\text{boost}}^2 = 32$ in the latter case. The reduced number of particles can also explain the small differences for the beam emittance.

Computational cost

All three simulations were performed on the Booster module of the JUWELS super-computer [245, 246] that is equipped with $4 \times$ Nvidia A100 GPUs per compute node. Due to the small size of the simulation box, both FBPIC simulations were run on a single GPU. The WARPX simulation was parallelized on 16 GPUs (4 nodes), with 2 sub-domains per transverse dimension and 4 sub-domains in the axial direction. The decomposition ensured a large enough problem size per GPU to make efficient use of the available computing power.

Fig. 8.8(a) compares the computational cost of the different simulations. The achieved speedup is defined relative to the GPU hours consumed by the full-3D simulation. The quasi-cylindrical geometry of FBPIC leads to a reduction of the computational cost by a factor of 33. Further application of the Lorentz-boosted frame technique decreases the run time by an additional factor of 55, which is close to the theoretically predicted speedup of $4\gamma_{\text{boost}}^2 = 64$.

Fig. 8.8(b) compares the total GPU time spent per particle during the entire simulation. Using this metric, FBPIC achieves a performance that is comparable to the performance of the 3D FDTD implementation of WARPX, although the spectral algorithm adds

significant complexity to the computation. In the Lorentz-boosted frame simulation, the computing time per particle is further reduced because the plasma streams through the simulation box.

In summary, the advanced algorithm implemented in `FBPIC` enables a speedup by a factor of ~ 1800 compared to a conventional 3D simulation. Although `FBPIC`'s implementation is based on Python, the various optimizations and just-in-time compilation allow it to keep up with the compute performance of the C++ code `WARPX`.

It should be noted that comparable speedups are also possible with `WARPX`. The code also supports the Lorentz-boosted frame technique in combination with various solvers. The arbitrary-order Galilean-PSATD solver in cylindrical geometry is currently being implemented and will be fully supported in the future.

Conclusion

Laser-plasma acceleration could allow for compact and cost-efficient sources of high-brightness electron beams that would benefit a wide range of applications with enormous societal impact. Although the field has seen rapid progress over the past decade, further improvements of the beam quality, in particular the energy spread, the overall efficiency, as well as the stability of laser-plasma accelerators (LPAs) are necessary to close the gap to modern accelerator technology.

It is widely believed that progressing in these matters requires to move away from proof-of-principle simulations and experiments and toward a more systematic approach to designing and operating LPAs. However, two fundamental bottlenecks have held up this transition so far. Adequately modeling the relativistic and kinetic phenomena in LPAs requires computationally intensive particle-in-cell simulations. The associated high cost can prevent extensive design and tolerance studies which are crucial to further enhance the performance of these machines. Similarly, due to the low repetition rates and poor reproducibility of common experimental setups, it is often impossible to study the complex injection and acceleration process with sufficient fidelity to optimize an LPA design in practice.

The approach taken by this thesis addressed these challenges in a cohesive manner. Contributions to the development of a highly efficient particle-in-cell (PIC) code have made it possible to drastically reduce the cost of LPA simulations. Supported by these modeling capabilities, an advanced plasma source has been developed that not only incorporates a mechanism for injecting an electron bunch with a small slice energy spread, but also provides control over beam loading so as to efficiently accelerate this bunch without imprinting correlated energy spread. The plasma source has been embedded in an experimental environment with stable operating conditions which allowed precise optimization and analysis of this process. This not only led to the production of high-quality electron beams, but through high-statistics measurements also enabled the use of machine learning to gain unprecedented insights into the stability of the accelerator. In the following, we summarize these results in more detail and provide a perspective on how they may impact the field of laser-plasma acceleration in the future.

With several new ideas and numerical concepts, this thesis contributed to the development of the open-source particle-in-cell code `FBPIC`. The code uses a cylindrical geometry with azimuthal mode decomposition to capture the essential three-dimensional

physics of plasma acceleration while reducing the computational cost to that of a few two-dimensional simulations. Moreover, the algorithm uses the arbitrary-order pseudospectral analytical time-domain (PSATD) method for solving Maxwell's equations and thereby avoids numerical errors common to conventional finite-difference time-domain (FDTD) solvers. In particular, contributions were made to enhancements of the original algorithm, its numerical implementation on modern hardware, such as multi-core processors and graphics processing units (GPU), as well as the efficient parallelization across distributed compute units by spatial domain decomposition.

The core features of FBPIC alone already enable more efficient and accurate simulations of plasma acceleration compared to conventional algorithms. Still, in many cases, an even greater reduction in simulation run time is possible by modeling the interaction in a relativistic frame of reference. However, application of this Lorentz-boosted frame technique involves simulating a relativistically streaming plasma, which gives rise to the numerical Cherenkov instability (NCI). As a solution, this thesis proposed to represent and discretize the particle-in-cell equations in Galilean (comoving) coordinates, so that the streaming plasma appears to be at rest with respect to the numerical grid. When combined with the PSATD framework, it was shown that this approach eliminates the NCI and does so even when the solver is generalized to arbitrary order for parallelization purposes. Unlike previous strategies that could only suppress and not eliminate the instability, the Galilean scheme does not introduce artificial numerical modifications or corrections to the underlying solver, nor is it restricted to a particular geometry.

The stability and accuracy of the Galilean-PSATD solver was validated theoretically and with realistic LPA simulations. In addition, based on a comparison with a conventional 3D Cartesian FDTD algorithm, it was demonstrated that the combined feature set of FBPIC enables speedups of more than three orders of magnitude for typical cases of laser-plasma acceleration. Simulations that might have taken days on a supercomputer a decade ago, can now be run in minutes on a single GPU. This paves the way for detailed optimization and tolerance studies, which are crucial for designing a next generation of LPAs with application-relevant performance. The significance of these results is also reflected by the fact that the adoption of FBPIC by the plasma acceleration community is steadily increasing, as evidenced in Fig. 9.1.

Supported by simulations with FBPIC, a novel plasma source was developed that integrates a dedicated injection and acceleration stage in a single capillary-type structure. Localized ionization-induced trapping along a density downramp served as a robust and reliable injection mechanism to generate electron bunches with a small slice energy spread and competitive transverse emittance. A particular advantage of the injection method is that it permits operation with a continuous flow of gas, which implies stable plasma conditions and readiness for kHz repetition rates. Due to its capillary-type design, the plasma source might also be combined with a discharge-based plasma guiding channel to reach higher beam energies and overall efficiencies in the future.

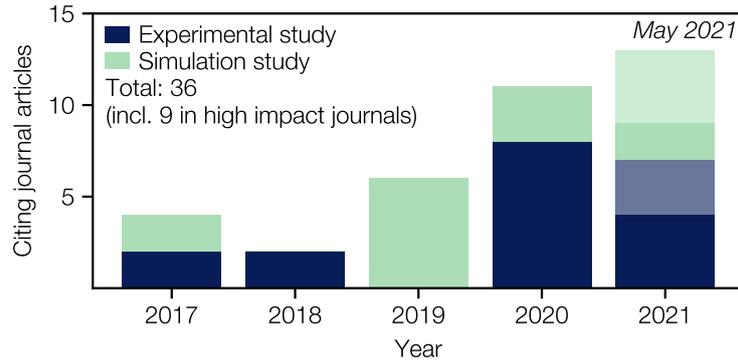


Figure 9.1 – Usage of FBPIC in the plasma acceleration community as of May 2021. Shown are the number of citations in peer-reviewed journal articles by a year, subdivided into experimental (to explain a measurement) and simulation (to study novel physics concepts or theories) studies. Citations in conference proceedings and numerical studies have been excluded. Out of a total of 36 publications, 9 have been published in high impact journals (impact factor > 8). The translucent fraction for 2021 corresponds to articles that have been submitted or accepted but not yet published.

The plasma source was embedded in the LUX accelerator that, driven by the ANGUS laser system, supported continuous and stable operation at 1 Hz repetition rate and thereby enabled reproducible measurements with thousands of individual shots. A comprehensive set of online diagnostics allowed precisely characterizing each individual laser pulse before the interaction with the plasma as well as the generated electron beam. By adjusting the laser energy, longitudinal focus position, plasma density and dopant gas concentration, the setup provided control over the injected bunch charge, current profile and evolution of the wakefield. This in turn affected the shape of the accelerating fields along the bunch via beam loading of the plasma wave.

Consequently, it was possible to tune the accelerator to optimal beam loading conditions which resulted in a flattening of the average accelerating fields to $\sim 1\%$ over a substantial fraction of the plasma wave ($\sim \lambda_p/5$). Operation at this optimum led to the production of electron bunches with a narrow energy spread of 1.2% rms, a central energy of 282 MeV, a charge of 44 pC and a peak spectral density of 5.3 pC MeV^{-1} . The electrons bunches exited the plasma with a sub-mrad divergence and normalized transverse emittance of 1.9 mm mrad. An energy-transfer efficiency of $\sim 19\%$ was estimated from simulations. A correlation-based analysis of shot-to-shot deviations from this optimum furthermore revealed the characteristic beam loading physics and allowed observing the shearing of the ~ 20 -fs-short bunch in longitudinal phase space due to under- and overloading of the plasma wave.

Compared to earlier work, effects from beam loading were diagnosed in more detail and much finer control over the accelerating fields was achieved. The demonstrated

beam quality already meets the targeted requirements of first applications such as a free-electron laser in the XUV range. Further optimizing and scaling the developed LPA design to higher energies could allow for per-mille energy spreads and tens-of-percent conversion efficiencies, which opens the door to a compact X-ray free-electron laser or linear collider. However, to reach these future milestones, the electron beam stability must also be improved, which motivated searching for the origin of the shot-to-shot variations that were observed in the experiment.

To this end, machine learning was used to build the first surrogate model of an LPA. Enabled by the rich set of electron and laser data that was collected during the measurements it was possible to train a neural network that accurately predicts the single-shot electron beam properties as a function of fundamental drive laser parameters. Consequently, this model can be used to identify and quantify sources of instability. With such information at hand, it is then possible to derive a clear roadmap for future laser development. For example, fluctuations of the wavefront, in particular the defocus jitter, were the strongest driver of instability—an effect that was not widely considered prior to this work. Hence, in a first step, it would make sense to concentrate efforts on wavefront stabilization, as this would benefit the LPA stability the most. Apart from this immediate use, surrogate models can also serve as a virtual diagnostic and provide the foundation for automatic control and global optimization techniques.

The field of laser-plasma acceleration has seen tremendous progress over the past two decades and continues to move at a fast pace. With the upcoming generation of high-power laser systems with kHz repetition rates, the next decade is likely to see another revolution in beam quality and stability. These high repetition rates offer the possibility for active stabilization and extensive optimization that, together with rapid simulations tools, will enable the design and operation of a next generation of LPAs capable to drive radiation sources with enough peak and average brilliance to allow for first user operation. The concepts and methods developed in this thesis will help to take the next step into this exciting future.

Bibliography

- [1] T. Tajima and J. M. Dawson, *Laser Electron Accelerator*, Phys. Rev. Lett. **43**, 267 (1979).
- [2] P. Chen, J. M. Dawson, R. W. Huff, and T. Katsouleas, *Acceleration of Electrons by the Interaction of a Bunched Electron Beam with a Plasma*, Phys. Rev. Lett. **54**, 693 (1985).
- [3] A. Modena, Z. Najmudin, A. E. Dangor, C. E. Clayton, K. A. Marsh, C. Joshi, V. Malka, C. B. Darrow, C. Danson, D. Neely, and F. N. Walsh, *Electron acceleration from the breaking of relativistic plasma waves*, Nature **377**, 606 (1995).
- [4] E. Esarey, C. B. Schroeder, and W. P. Leemans, *Physics of laser-driven plasma-based electron accelerators*, Rev. Mod. Phys. **81**, 1229 (2009).
- [5] D. Strickland and G. Mourou, *Compression of amplified chirped optical pulses*, Opt. Commun. **56**, 219 (1985).
- [6] J. Faure, Y. Glinec, A. Pukhov, S. Kiselev, S. Gordienko, E. Lefebvre, J. P. Rousseau, F. Burgy, and V. Malka, *A laser-plasma accelerator producing monoenergetic electron beams*, Nature **431**, 541 (2004).
- [7] C. G. R. Geddes, C. Toth, J. van Tilborg, E. Esarey, C. B. Schroeder, D. Bruhwiler, C. Nieter, J. Cary, and W. P. Leemans, *High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding*, Nature **431**, 538 (2004).
- [8] S. P. D. Mangles, C. D. Murphy, Z. Najmudin, A. G. R. Thomas, J. L. Collier, A. E. Dangor, E. J. Divall, P. S. Foster, J. G. Gallacher, C. J. Hooker, D. A. Jaroszynski, A. J. Langley, W. B. Mori, P. A. Norreys, F. S. Tsung, R. Viskup, B. R. Walton, and K. Krushelnick, *Monoenergetic beams of relativistic electrons from intense laser-plasma interactions*, Nature **431**, 535 (2004).
- [9] O. Lundh, J. Lim, C. Rechatin, L. Ammoura, A. Ben-Ismaïl, X. Davoine, G. Gallot, J. P. Goddet, E. Lefebvre, V. Malka, and J. Faure, *Few femtosecond, few kiloampere electron bunch produced by a laser-plasma accelerator*, Nat. Phys. **7**, 219 (2011).

- [10] A. Buck, M. Nicolai, K. Schmid, C. M. Sears, A. Sävert, J. M. Mikhailova, F. Krausz, M. C. Kaluza, and L. Veisz, *Real-time observation of laser-driven electron acceleration*, Nat. Phys. **7**, 543 (2011).
- [11] G. R. Plateau, C. G. R. Geddes, D. B. Thorn, M. Chen, C. Benedetti, E. Esarey, A. J. Gonsalves, N. H. Matlis, K. Nakamura, C. B. Schroeder, S. Shiraishi, T. Sokollik, J. van Tilborg, C. Toth, S. Trotsenko, T. S. Kim, M. Battaglia, T. Stöhlker, and W. P. Leemans, *Low-Emittance Electron Bunches from a Laser-Plasma Accelerator Measured using Single-Shot X-Ray Spectroscopy*, Phys. Rev. Lett. **109**, 064802 (2012).
- [12] R. Weingartner, S. Raith, A. Popp, S. Chou, J. Wenz, K. Khrennikov, M. Heigoldt, A. R. Maier, N. Kajumba, M. Fuchs, B. Zeitler, F. Krausz, S. Karsch, and F. Grüner, *Ultralow emittance electron beams from a laser-wakefield accelerator*, Phys. Rev. ST Accel. Beams **15**, 111302 (2012).
- [13] W. P. Leemans, B. Nagler, A. J. Gonsalves, C. Tóth, K. Nakamura, C. G. Geddes, E. Esarey, C. B. Schroeder, and S. M. Hooker, *GeV electron beams from a centimetre-scale accelerator*, Nat. Phys. **2**, 696 (2006).
- [14] X. Wang, R. Zgadzaj, N. Fazel, Z. Li, S. A. Yi, X. Zhang, W. Henderson, Y. Y. Chang, R. Korzekwa, H. E. Tsai, C. H. Pai, H. Quevedo, G. Dyer, E. Gaul, M. Martinez, A. C. Bernstein, T. Borger, M. Spinks, M. Donovan, V. Khudik, G. Shvets, T. Ditmire, and M. C. Downer, *Quasi-monoenergetic laser-plasma acceleration of electrons to 2 GeV*, Nature Communications **4**, 1988 (2013).
- [15] H. T. Kim, K. H. Pae, H. J. Cha, I. J. Kim, T. J. Yu, J. H. Sung, S. K. Lee, T. M. Jeong, and J. Lee, *Enhancement of Electron Energy to the Multi-GeV Regime by a Dual-Stage Laser-Wakefield Accelerator Pumped by Petawatt Laser Pulses*, Phys. Rev. Lett. **111**, 165002 (2013).
- [16] A. J. Gonsalves, K. Nakamura, J. Daniels, C. Benedetti, C. Pieronek, T. C. H. de Raadt, S. Steinke, J. H. Bin, S. S. Bulanov, J. van Tilborg, C. G. R. Geddes, C. B. Schroeder, C. Tóth, E. Esarey, K. Swanson, L. Fan-Chiang, G. Bagdasarov, N. Bobrova, V. Gasilov, G. Korn, P. Sasorov, and W. P. Leemans, *Petawatt Laser Guiding and Electron Beam Acceleration to 8 GeV in a Laser-Heated Capillary Discharge Waveguide*, Phys. Rev. Lett. **122**, 084801 (2019).
- [17] A. Pukhov and J. Meyer-ter Vehn, *Laser wake field acceleration: the highly non-linear broken-wave regime*, Applied Physics B **74**, 355 (2002).
- [18] J. Faure, C. Rechatin, A. Norlin, A. Lifschitz, Y. Glinec, and V. Malka, *Controlled injection and acceleration of electrons in plasma wakefields by colliding laser pulses*, Nature **444**, 737 (2006).

-
- [19] A. Pak, K. A. Marsh, S. F. Martins, W. Lu, W. B. Mori, and C. Joshi, *Injection and Trapping of Tunnel-Ionized Electrons into Laser-Produced Wakes*, Phys. Rev. Lett. **104**, 025003 (2010).
- [20] A. J. Gonsalves, K. Nakamura, C. Lin, D. Panasenko, S. Shiraishi, T. Sokollik, C. Benedetti, C. B. Schroeder, C. G. R. Geddes, J. van Tilborg, J. Osterhoff, E. Esarey, C. Toth, and W. P. Leemans, *Tunable laser plasma accelerator based on longitudinal density tailoring*, Nat. Phys. **7**, 862 (2011).
- [21] A. Buck, J. Wenz, J. Xu, K. Khrennikov, K. Schmid, M. Heigoldt, J. M. Mikhailova, M. Geissler, B. Shen, F. Krausz, S. Karsch, and L. Veisz, *Shock-Front Injector for High-Quality Laser-Plasma Acceleration*, Phys. Rev. Lett. **110**, 185006 (2013).
- [22] C. Rechatin, J. Faure, A. Ben-Ismaïl, J. Lim, R. Fitour, A. Specka, H. Videau, A. Tafzi, F. Burgy, and V. Malka, *Controlling the Phase-Space Volume of Injected Electrons in a Laser-Plasma Accelerator*, Phys. Rev. Lett. **102**, 164801 (2009).
- [23] W. T. Wang, W. T. Li, J. S. Liu, Z. J. Zhang, R. Qi, C. H. Yu, J. Q. Liu, M. Fang, Z. Y. Qin, C. Wang, Y. Xu, F. X. Wu, Y. X. Leng, R. X. Li, and Z. Z. Xu, *High-Brightness High-Energy Electron Beams from a Laser Wakefield Accelerator via Energy Chirp Control*, Phys. Rev. Lett. **117**, 124801 (2016).
- [24] K. K. Swanson, H.-E. Tsai, S. K. Barber, R. Lehe, H.-S. Mao, S. Steinke, J. van Tilborg, K. Nakamura, C. G. R. Geddes, C. B. Schroeder, E. Esarey, and W. P. Leemans, *Control of tunable, monoenergetic laser-plasma-accelerated electron beams using a shock-induced density downramp injector*, Phys. Rev. Accel. Beams **20**, 051301 (2017).
- [25] A. R. Maier, A. Meseck, S. Reiche, C. B. Schroeder, T. Seggebrock, and F. Grüner, *Demonstration Scheme for a Laser-Plasma-Driven Free-Electron Laser*, Phys. Rev. X **2**, 031019 (2012).
- [26] S. Corde, K. Ta Phuoc, G. Lambert, R. Fitour, V. Malka, A. Rousse, A. Beck, and E. Lefebvre, *Femtosecond x rays from laser-plasma accelerators*, Rev. Mod. Phys. **85**, 1 (2013).
- [27] C. B. Schroeder, E. Esarey, C. G. R. Geddes, C. Benedetti, and W. P. Leemans, *Physics considerations for laser-plasma linear colliders*, Phys. Rev. ST Accel. Beams **13**, 101301 (2010).
- [28] S. Van der Meer, *Improving the power efficiency of the plasma wakefield accelerator*, Tech. Rep. PS-85-65 (CERN, 1985).
- [29] T. C. Katsouleas, S. Wilks, P. Chen, J. M. Dawson, and J. J. Su, *Beam Loading in Plasma Accelerators*, Part. Accel. **22**, 81 (1987).

- [30] M. Tzoufras, W. Lu, F. S. Tsung, C. Huang, W. B. Mori, T. Katsouleas, J. Vieira, R. A. Fonseca, and L. O. Silva, *Beam Loading in the Nonlinear Regime of Plasma-Based Acceleration*, Phys. Rev. Lett. **101**, 145002 (2008).
- [31] M. Tzoufras, W. Lu, F. S. Tsung, C. Huang, W. B. Mori, T. Katsouleas, J. Vieira, R. A. Fonseca, and L. O. Silva, *Beam loading by electrons in nonlinear plasma wakes*, Phys. Plasmas **16**, 056705 (2009).
- [32] C. Rechatin, X. Davoine, A. Lifschitz, A. B. Ismail, J. Lim, E. Lefebvre, J. Faure, and V. Malka, *Observation of Beam Loading in a Laser-Plasma Accelerator*, Phys. Rev. Lett. **103**, 194804 (2009).
- [33] J. P. Couperus, R. Pausch, A. Köhler, O. Zarini, J. M. Krämer, M. Garten, A. Huebl, R. Gebhardt, U. Helbig, S. Bock, K. Zeil, A. Debus, M. Bussmann, U. Schramm, and A. Irman, *Demonstration of a beam loaded nanocoulomb-class laser wakefield accelerator*, Nat. Commun. **8**, 487 (2017).
- [34] J. Götzfried, A. Döpp, M. F. Gilljohann, F. M. Foerster, H. Ding, S. Schindler, G. Schilling, A. Buck, L. Veisz, and S. Karsch, *Physics of High-Charge Electron Beams in Laser-Plasma Wakefields*, Phys. Rev. X **10**, 041015 (2020).
- [35] F. Albert, M. E. Couprie, A. Debus, M. C. Downer, J. Faure, A. Flacco, L. A. Gizzi, T. Grismayer, A. Huebl, C. Joshi, M. Labat, W. P. Leemans, A. R. Maier, S. P. D. Mangles, P. Mason, F. Mathieu, P. Muggli, M. Nishiuchi, J. Osterhoff, P. P. Rajeev, U. Schramm, J. Schreiber, A. G. R. Thomas, J.-L. Vay, M. Vranic, and K. Zeil, *2020 roadmap on plasma accelerators*, New Journal of Physics **23**, 031101 (2021).
- [36] A. R. Maier, N. Delbos, T. Eichner, L. Hübner, S. Jalas, L. Jeppe, S. W. Jolly, M. Kirchen, V. Leroux, P. Messner, M. Schnepf, M. Trunk, P. A. Walker, C. Werle, and P. Winkler, *Decoding Sources of Energy Variability in a Laser-Plasma Accelerator*, Phys. Rev. X **10**, 031039 (2020).
- [37] M. Kirchen, S. Jalas, P. Messner, P. Winkler, T. Eichner, L. Hübner, T. Hülsenbusch, L. Jeppe, T. Parikh, M. Schnepf, and A. R. Maier, *Optimal Beam Loading in a Laser-Plasma Accelerator*, Phys. Rev. Lett. **126**, 174801 (2021).
- [38] J. M. Dawson, *Particle simulation of plasmas*, Rev. Mod. Phys. **55**, 403 (1983).
- [39] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (Taylor & Francis, 1988).
- [40] C. K. Birdsall and A. B. Langdon, *Plasma physics via computer simulation*, Series in Plasma Physics (Taylor & Francis, 2004).

-
- [41] B. M. Cowan, D. L. Bruhwiler, J. R. Cary, E. Cormier-Michel, and C. G. Geddes, *Generalized algorithm for control of numerical dispersion in explicit time-domain electromagnetic simulations*, Phys. Rev. ST Accel. Beams **16**, 41303 (2013).
- [42] R. Lehe, A. Lifschitz, C. Thaury, V. Malka, and X. Davoine, *Numerical growth of emittance in simulations of laser-wakefield acceleration*, Phys. Rev. ST Accel. Beams **16**, 21301 (2013).
- [43] R. Lehe, M. Kirchen, I. A. Andriyash, B. B. Godfrey, and J.-L. Vay, *A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm*, Comput. Phys. Commun. **203**, 66 (2016).
- [44] S. Jalas, I. Dornmair, R. Lehe, H. Vincenti, J. L. Vay, M. Kirchen, and A. R. Maier, *Accurate modeling of plasma acceleration with arbitrary order pseudo-spectral particle-in-cell methods*, Phys. Plasmas **24**, 033115 (2017).
- [45] J. L. Vay, I. Haber, and B. B. Godfrey, *A domain decomposition method for pseudo-spectral electromagnetic simulations of plasmas*, J. Comput. Phys **243**, 260 (2013).
- [46] H. Vincenti and J. L. Vay, *Detailed analysis of the effects of stencil spatial variations with arbitrary high-order finite-difference Maxwell solver*, Comput. Phys. Commun. **200**, 147 (2016).
- [47] A. Lifschitz, X. Davoine, E. Lefebvre, J. Faure, C. Rechatin, and V. Malka, *Particle-in-Cell modelling of laser-plasma interaction using Fourier decomposition*, J. Comput. Phys **228**, 1803 (2009).
- [48] A. Davidson, A. Tableman, W. An, F. Tsung, W. Lu, J. Vieira, R. Fonseca, L. Silva, and W. Mori, *Implementation of a hybrid particle code with a PIC description in r - z and a gridless description in ϕ into OSIRIS*, J. Comput. Phys **281**, 1063 (2015).
- [49] J. L. Vay, *Noninvariance of space- and time-scale ranges under a Lorentz transformation and the implications for the study of relativistic interactions*, Phys. Rev. Lett. **98**, 130405 (2007).
- [50] R. Lehe and M. Kirchen, *FBPIC - Fourier-Bessel Particle-In-Cell code* (2020).
- [51] M. Kirchen, R. Lehe, B. B. Godfrey, I. Dornmair, S. Jalas, K. Peters, J. L. Vay, and A. R. Maier, *Stable discrete representation of relativistically drifting plasmas*, Phys. Plasmas **23**, 100704 (2016).
- [52] R. Lehe, M. Kirchen, B. B. Godfrey, A. R. Maier, and J. L. Vay, *Elimination of numerical Cherenkov instability in flowing-plasma particle-in-cell simulations by using Galilean coordinates*, Phys. Rev. E **94**, 053305 (2016).

- [53] M. Kirchen, R. Lehe, S. Jalas, O. Shapoval, J.-L. Vay, and A. R. Maier, *Scalable spectral solver in Galilean coordinates for eliminating the numerical Cherenkov instability in particle-in-cell simulations of streaming plasmas*, Phys. Rev. E **102**, 013202 (2020).
- [54] B. B. Godfrey, *Numerical Cherenkov instabilities in electromagnetic particle codes*, J. Comput. Phys **15**, 504 (1974).
- [55] B. B. Godfrey, *Canonical momenta and numerical instabilities in particle codes*, J. Comput. Phys **19**, 58 (1975).
- [56] S. Jalas, M. Kirchen, P. Messner, P. Winkler, L. Hübner, J. Dirkwinkel, M. Schnepf, R. Lehe, and A. R. Maier, *Bayesian Optimization of a Laser-Plasma Accelerator*, Phys. Rev. Lett. **126**, 104801 (2021).
- [57] V. Leroux, S. W. Jolly, M. Schnepf, T. Eichner, S. Jalas, M. Kirchen, P. Messner, C. Werle, P. Winkler, and A. R. Maier, *Wavefront degradation of a 200 TW laser from heat-induced deformation of in-vacuum compressor gratings*, Opt. Express **26**, 13061 (2018).
- [58] N. Delbos, C. Werle, I. Dornmair, T. Eichner, L. Hübner, S. Jalas, S. W. Jolly, M. Kirchen, V. Leroux, P. Messner, M. Schnepf, M. Trunk, P. A. Walker, P. Winkler, and A. R. Maier, *LUX – A laser-plasma driven undulator beamline*, Nucl. Instr. Meth. Phys. Res. A **909**, 318 (2018).
- [59] R. Brinkmann, N. Delbos, I. Dornmair, M. Kirchen, R. Assmann, C. Behrens, K. Floettmann, J. Grebenyuk, M. Gross, S. Jalas, T. Mehrling, A. Martinez de la Ossa, J. Osterhoff, B. Schmidt, V. Wacker, and A. R. Maier, *Chirp Mitigation of Plasma-Accelerated Beams by a Modulated Plasma Density*, Phys. Rev. Lett. **118**, 214801 (2017).
- [60] P. Gibbon, *Short Pulse Laser Interactions with Matter* (World Scientific, 2005).
- [61] F. Chen, *Introduction to Plasma Physics and Controlled Fusion* (Springer, 1984).
- [62] R. Goldston and P. Rutherford, *Introduction to Plasma Physics* (Taylor & Francis, 1995).
- [63] A. I. Akhiezer and R. V. Polovin, *Theory of wave motion of an electron plasma*, Soviet Phys. JETP **3** (1956).
- [64] J. M. Dawson, *Nonlinear Electron Oscillations in a Cold Plasma*, Phys. Rev. **113**, 383 (1959).
- [65] E. Esarey, P. Sprangle, J. Krall, and A. Ting, *Overview of plasma-based accelerator concepts*, IEEE Transactions on Plasma Science **24**, 252 (1996).
- [66] J. Jackson and P. Jackson, *Classical Electrodynamics* (Wiley, 1998).

-
- [67] W. Kruer, *The Physics Of Laser Plasma Interactions* (Avalon Publishing, 1988).
- [68] E. Esarey, P. Sprangle, J. Krall, and A. Ting, *Self-focusing and guiding of short laser pulses in ionizing gases and plasmas*, IEEE Journal of Quantum Electronics **33**, 1879 (1997).
- [69] P. Sprangle, C.-M. Tang, and E. Esarey, *Relativistic Self-Focusing of Short-Pulse Radiation Beams in Plasmas*, IEEE Transactions on Plasma Science **15**, 145 (1987).
- [70] P. Sprangle, E. Esarey, and A. Ting, *Nonlinear theory of intense laser-plasma interactions*, Phys. Rev. Lett. **64**, 2011 (1990).
- [71] G. Sun, E. Ott, Y. C. Lee, and P. Guzdar, *Self-focusing of short intense pulses in plasmas*, The Physics of Fluids **30**, 526 (1987).
- [72] P. Sprangle, E. Esarey, J. Krall, and G. Joyce, *Propagation and guiding of intense laser pulses in plasmas*, Phys. Rev. Lett. **69**, 2200 (1992).
- [73] P. Sprangle and E. Esarey, *Interaction of ultrahigh laser fields with beams and plasmas*, Physics of Fluids B: Plasma Physics **4**, 2241 (1992).
- [74] F. S. Tsung, C. Ren, L. O. Silva, W. B. Mori, and T. Katsouleas, *Generation of ultra-intense single-cycle laser pulses by using photon deceleration*, Proceedings of the National Academy of Sciences **99**, 29 (2002).
- [75] D. F. Gordon, B. Hafizi, R. F. Hubbard, J. R. Peñano, P. Sprangle, and A. Ting, *Asymmetric Self-Phase Modulation and Compression of Short Laser Pulses in Plasma Channels*, Phys. Rev. Lett. **90**, 215001 (2003).
- [76] J. Vieira, F. Fiúza, L. O. Silva, M. Tzoufras, and W. B. Mori, *Onset of self-steepening of intense laser pulses in plasmas*, New Journal of Physics **12**, 045025 (2010).
- [77] V. Berezhiani and I. Murusidze, *Interaction of highly relativistic short laser pulses with plasmas and nonlinear wake-field generation*, Physica Scripta **45**, 87 (1992).
- [78] D. Teychenné, G. Bonnaud, and J.-L. Bobin, *Wave-breaking limit to the wake-field effect in an underdense plasma*, Phys. Rev. E **48**, R3248 (1993).
- [79] P. Sprangle, E. Esarey, and A. Ting, *Nonlinear interaction of intense laser pulses in plasmas*, Phys. Rev. A **41**, 4463 (1990).
- [80] P. Sprangle, E. Esarey, A. Ting, and G. Joyce, *Laser wakefield acceleration and relativistic optical guiding*, Applied Physics Letters **53**, 2146 (1988).
- [81] E. Esarey, P. Sprangle, J. Krall, A. Ting, and G. Joyce, *Optically guided laser wake-field acceleration*, Physics of Fluids B: Plasma Physics **5**, 2690 (1993).

- [82] B. Shadwick, G. Tarkenton, E. Esarey, and W. Leemans, *Fluid simulations of intense laser-plasma interactions*, IEEE Transactions on Plasma Science **30**, 38 (2002).
- [83] C. D. Decker, W. B. Mori, and T. Katsouleas, *Particle-in-cell simulations of Raman forward scattering from short-pulse high-intensity lasers*, Phys. Rev. E **50**, R3338 (1994).
- [84] P. Mora and T. M. Antonsen, *Electron cavitation and acceleration in the wake of an ultraintense, self-focused laser pulse*, Phys. Rev. E **53**, R2068 (1996).
- [85] W. Lu, C. Huang, M. Zhou, W. B. Mori, and T. Katsouleas, *Nonlinear Theory for Relativistic Plasma Wakefields in the Blowout Regime*, Phys. Rev. Lett. **96**, 165002 (2006).
- [86] W. Lu, C. Huang, M. Zhou, M. Tzoufras, F. S. Tsung, W. B. Mori, and T. Katsouleas, *A nonlinear theory for multidimensional relativistic plasma wave wakefields*, Physics of Plasmas **13**, 056709 (2006).
- [87] 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*, Phys. Rev. ST Accel. Beams **10**, 061301 (2007).
- [88] C. Benedetti, C. B. Schroeder, E. Esarey, F. Rossi, and W. P. Leemans, *Numerical investigation of electron self-injection in the nonlinear bubble regime*, Physics of Plasmas **20**, 103108 (2013).
- [89] 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**, 063103 (2021).
- [90] E. Esarey and M. Pilloff, *Trapping and acceleration in nonlinear plasma waves*, Physics of Plasmas **2**, 1432 (1995).
- [91] T. Esirkepov, S. V. Bulanov, M. Yamagiwa, and T. Tajima, *Electron, Positron, and Photon Wakefield Acceleration: Trapping, Wake Overtaking, and Ponderomotive Acceleration*, Phys. Rev. Lett. **96**, 014803 (2006).
- [92] C. B. Schroeder and E. Esarey, *Relativistic warm plasma theory of nonlinear laser-driven electron plasma waves*, Phys. Rev. E **81**, 056403 (2010).
- [93] S. V. Bulanov, F. Pegoraro, A. M. Pukhov, and A. S. Sakharov, *Transverse-Wake Wave Breaking*, Phys. Rev. Lett. **78**, 4205 (1997).
- [94] S. Bulanov, N. Naumova, F. Pegoraro, and J. Sakai, *Particle injection into the wave acceleration phase due to nonlinear wake wave breaking*, Phys. Rev. E **58**, R5257 (1998).

-
- [95] H. Suk, N. Barov, J. B. Rosenzweig, and E. Esarey, *Plasma Electron Trapping and Acceleration in a Plasma Wake Field Using a Density Transition*, Phys. Rev. Lett. **86**, 1011 (2001).
- [96] E. Esarey, R. F. Hubbard, W. P. Leemans, A. Ting, and P. Sprangle, *Electron Injection into Plasma Wakefields by Colliding Laser Pulses*, Phys. Rev. Lett. **79**, 2682 (1997).
- [97] G. Fubiani, E. Esarey, C. B. Schroeder, and W. P. Leemans, *Beat wave injection of electrons into plasma waves using two interfering laser pulses*, Phys. Rev. E **70**, 016402 (2004).
- [98] M. Chen, Z.-M. Sheng, Y.-Y. Ma, and J. Zhang, *Electron injection and trapping in a laser wakefield by field ionization to high-charge states of gases*, Journal of Applied Physics **99**, 056109 (2006).
- [99] M. Chen, E. Esarey, C. B. Schroeder, C. G. R. Geddes, and W. P. Leemans, *Theory of ionization-induced trapping in laser-plasma accelerators*, Phys. Plasmas **19**, 033101 (2012).
- [100] L.-L. Yu, E. Esarey, C. B. Schroeder, J.-L. Vay, C. Benedetti, C. G. R. Geddes, M. Chen, and W. P. Leemans, *Two-Color Laser-Ionization Injection*, Phys. Rev. Lett. **112**, 125001 (2014).
- [101] M. V. Ammosov, N. B. Delone, and V. P. Krainov, *Tunnel ionization of complex atoms and atomic ions in an electromagnetic field*, Sov. Phys. JETP **64**, 1191 (1986).
- [102] M. Chen, E. Cormier-Michel, C. Geddes, D. Bruhwiler, L. Yu, E. Esarey, C. Schroeder, and W. Leemans, *Numerical modeling of laser tunneling ionization in explicit particle-in-cell codes*, J. Comput. Phys **236**, 220 (2013).
- [103] C. B. Schroeder, J.-L. Vay, E. Esarey, S. S. Bulanov, C. Benedetti, L.-L. Yu, M. Chen, C. G. R. Geddes, and W. P. Leemans, *Thermal emittance from ionization-induced trapping in plasma accelerators*, Phys. Rev. ST Accel. Beams **17**, 101301 (2014).
- [104] E. Esarey, B. A. Shadwick, C. B. Schroeder, and W. P. Leemans, *Nonlinear Pump Depletion and Electron Dephasing in Laser Wakefield Accelerators*, AIP Conference Proceedings **737**, 578 (2004).
- [105] B. A. Shadwick, C. B. Schroeder, and E. Esarey, *Nonlinear laser energy depletion in laser-plasma accelerators*, Physics of Plasmas **16**, 056704 (2009).
- [106] C. B. Schroeder, E. Esarey, C. Benedetti, and W. P. Leemans, *Efficiency considerations for high-energy physics applications of laser-plasma accelerators*, AIP Conf. Proc. **1777**, 020001 (2016).

- [107] A. Ferran Pousa, A. Martinez de la Ossa, and R. W. Assmann, *Intrinsic energy spread and bunch length growth in plasma-based accelerators due to betatron motion*, Scientific Reports **9**, 17690 (2019).
- [108] C. B. Schroeder, C. Benedetti, E. Esarey, and W. P. Leemans, *Beam loading in a laser-plasma accelerator using a near-hollow plasma channel*, Physics of Plasmas **20**, 123115 (2013).
- [109] C. Benedetti, *Characterization of energy gain and efficiency for guided and unguided laser-plasma accelerator stages*, in 2018 IEEE Advanced Accelerator Concepts Workshop (2018).
- [110] V. Leroux, *On the wavefront of ultrahigh intensity lasers: spatial contrast and gratings deformation*, Ph.D. thesis, Universität Hamburg (2018).
- [111] C. Werle, *First Undulator Experiments at the LUX Beamline: Towards a Plasma-Based Accelerator as a Soft X-ray Source*, Ph.D. thesis, Universität Hamburg (2019).
- [112] N. Delbos, *High Repetition Rate Laser-Plasma Accelerator: 5 Hz Electron Beam Generation and Advanced Target Design*, Ph.D. thesis, Universität Hamburg (2017).
- [113] P. Messner, *Tunable Laser-Plasma Acceleration with Ionization Injection*, Ph.D. thesis, Universität Hamburg (2020).
- [114] P. Winkler, *Emittance Measurements at Laser-Wakefield Accelerators*, Ph.D. thesis, Universität Hamburg (2019).
- [115] Thales Optronique S.A., *200 TW Ti:Sa laser system with 5 Hz repetition rate* (2014).
- [116] K. F. Wall and A. Sanchez, *Titanium sapphire lasers*, The Lincoln laboratory journal **3**, 447 (1990).
- [117] C. N. Danson, C. Haefner, J. Bromage, T. Butcher, J.-C. F. Chanteloup, E. A. Chowdhury, A. Galvanauskas, L. A. Gizzi, J. Hein, D. I. Hillier, N. W. Hopps, Y. Kato, E. A. Khazanov, R. Kodama, G. Korn, R. Li, Y. Li, J. Limpert, J. Ma, C. H. Nam, D. Neely, D. Papadopoulos, R. R. Penman, L. Qian, J. J. Rocca, A. A. Shaykin, C. W. Siders, C. Spindloe, S. Szatmári, R. M. G. M. Trines, J. Zhu, P. Zhu, and J. D. Zuegel, *Petawatt and exawatt class lasers worldwide*, High Power Laser Sci. Eng. **7**, e54 (2019).
- [118] A. Jullien, O. Albert, F. Burgy, G. Hamoniaux, J.-P. Rousseau, J.-P. Chambaret, F. Augé-Rochereau, G. Chériaux, J. Etchepare, N. Minkovski, and S. M. Saitiel, *10-10 temporal contrast for femtosecond ultraintense lasers by cross-polarized wave generation*, Opt. Lett. **30**, 920 (2005).

-
- [119] P. Tournois, *Acousto-optic programmable dispersive filter for adaptive compensation of group delay time dispersion in laser systems*, Opt. Commun. **140**, 245 (1997).
- [120] R. V. Shack, *Production and use of a lecticular Hartmann screen*, in Program of the 1971 Spring Meeting of the Optical Society of America, Vol. 61 (J. Opt. Soc. Am., 1971) p. 656.
- [121] T. Oksenhendler, S. Coudreau, N. Forget, V. Crozatier, S. Grabielle, R. Herzog, O. Gobert, and D. Kaplan, *Self-referenced spectral interferometry*, Appl. Phys. B **99**, 7 (2010).
- [122] P. O'Shea, M. Kimmel, X. Gu, and R. Trebino, *Highly simplified device for ultrashort-pulse measurement*, Opt. Lett. **26**, 932 (2001).
- [123] K. Nakamura, H.-S. Mao, A. Gonsalves, H. Vincenti, D. Mittelberger, J. Daniels, A. Magana, C. Toth, and W. Leemans, *Diagnostics, Control and Performance Parameters for the BELLA High Repetition Rate Petawatt Class Laser*, IEEE Journal of Quantum Electronics **53**, 1200121 (2017).
- [124] M. Santarsiero, D. Aiello, R. Borghi, and S. Vicalvi, *Focusing of axially symmetric flattened Gaussian beams*, Journal of Modern Optics **44**, 633 (1997).
- [125] A. Popp, J. Vieira, J. Osterhoff, Z. Major, R. Hörlein, M. Fuchs, R. Weingartner, T. P. Rowlands-Rees, M. Marti, R. A. Fonseca, S. F. Martins, L. O. Silva, S. M. Hooker, F. Krausz, F. Grüner, and S. Karsch, *All-Optical Steering of Laser-Wakefield-Accelerated Electron Beams*, Phys. Rev. Lett. **105**, 215001 (2010).
- [126] M. Thévenet, D. E. Mittelberger, K. Nakamura, R. Lehe, C. B. Schroeder, J.-L. Vay, E. Esarey, and W. P. Leemans, *Pulse front tilt steering in laser plasma accelerators*, Phys. Rev. Accel. Beams **22**, 071301 (2019).
- [127] R. J. Noll, *Zernike polynomials and atmospheric turbulence*, J. Opt. Soc. Am. **66**, 207 (1976).
- [128] K. Floettmann, *Some basic features of the beam emittance*, Phys. Rev. ST Accel. Beams **6**, 034202 (2003).
- [129] I. Dornmair, *Advanced Beam Dynamics and Diagnostics Concepts for Laser-Plasma Accelerators*, Ph.D. thesis, Universität Hamburg (2017).
- [130] B. Krause and A. Petrov, *Technical specifications version A for quadrupole XQA* (2009).
- [131] I. Okunev, I. Morozov, and N. Nefedov, *X-FEL Quadrupole with Gradient of 100 T/m*, in Proc of Int. Conf. Synchrotron and FEL Radiation, Vol. 84 (2016) p. 101.

- [132] D. Lipka *et al.*, *Very First Experience with the Standard Diagnostics at the European XFEL*, in Proc. of 8th IPAC (Copenhagen, Denmark, 2017) p. 180.
- [133] R. Lorenz, *Cavity beam position monitors*, AIP Conf. Proc. **451**, 53 (1998).
- [134] D. Lipka *et al.*, *Cavity BPM designs, related electronics and measured performances*, in Proc. of 9th DIPAC (Basel, Switzerland, 2009).
- [135] C. Wiebers, M. Holz, G. Kube, D. Nölle, G. Priebe, and H.-C. Schröder, *Scintillating Screen Monitors for Transverse Electron Beam Profile Diagnostics at the European XFEL*, in Proc. of 2nd IBIC (2013).
- [136] A. Buck, K. Zeil, A. Popp, K. Schmid, A. Jochmann, S. D. Kraft, B. Hidding, T. Kudyakov, C. M. S. Sears, L. Veisz, S. Karsch, J. Pawelke, R. Sauerbrey, T. Cowan, F. Krausz, and U. Schramm, *Absolute charge calibration of scintillating screens for relativistic electron detection*, Rev. Sci. Instrum. **81**, 033301 (2010).
- [137] T. Kurz, J. P. Couperus, J. M. Krämer, H. Ding, S. Kuschel, A. Köhler, O. Zarini, D. Hollatz, D. Schinkel, R. D'Arcy, J.-P. Schwinkendorf, J. Osterhoff, A. Irman, U. Schramm, and S. Karsch, *Calibration and cross-laboratory implementation of scintillating screens for electron bunch charge determination*, Rev. Sci. Instrum. **89**, 093303 (2018).
- [138] K. Schmid and L. Veisz, *Supersonic gas jets for laser-plasma experiments*, Rev. Sci. Instrum. **83**, 053304 (2012).
- [139] D. J. Spence and S. M. Hooker, *Investigation of a hydrogen plasma waveguide*, Phys. Rev. E **63**, 015401 (2000).
- [140] J. Osterhoff, A. Popp, Z. Major, B. Marx, T. P. Rowlands-Rees, M. Fuchs, M. Geissler, R. Hörlein, B. Hidding, S. Becker, E. A. Peralta, U. Schramm, F. Grüner, D. Habs, F. Krausz, S. M. Hooker, and S. Karsch, *Generation of Stable, Low-Divergence Electron Beams by Laser-Wakefield Acceleration in a Steady-State-Flow Gas Cell*, Phys. Rev. Lett. **101**, 085002 (2008).
- [141] A. Butler, D. J. Spence, and S. M. Hooker, *Guiding of High-Intensity Laser Pulses with a Hydrogen-Filled Capillary Discharge Waveguide*, Phys. Rev. Lett. **89**, 185003 (2002).
- [142] W. P. Leemans, A. J. Gonsalves, H.-S. Mao, K. Nakamura, C. Benedetti, C. B. Schroeder, C. Tóth, J. Daniels, D. E. Mittelberger, S. S. Bulanov, J.-L. Vay, C. G. R. Geddes, and E. Esarey, *Multi-GeV Electron Beams from Capillary-Discharge-Guided Subpetawatt Laser Pulses in the Self-Trapping Regime*, Phys. Rev. Lett. **113**, 245002 (2014).
- [143] C. G. Durfee and H. M. Milchberg, *Light pipe for high intensity laser pulses*, Phys. Rev. Lett. **71**, 2409 (1993).

-
- [144] R. J. Shalloo, C. Arran, L. Corner, J. Holloway, J. Jonnerby, R. Walczak, H. M. Milchberg, and S. M. Hooker, *Hydrodynamic optical-field-ionized plasma channels*, Phys. Rev. E **97**, 053203 (2018).
- [145] C. G. R. Geddes, K. Nakamura, G. R. Plateau, C. Toth, E. Cormier-Michel, E. Esarey, C. B. Schroeder, J. R. Cary, and W. P. Leemans, *Plasma-Density-Gradient Injection of Low Absolute-Momentum-Spread Electron Bunches*, Phys. Rev. Lett. **100**, 215004 (2008).
- [146] K. Schmid, A. Buck, C. M. S. Sears, J. M. Mikhailova, R. Tautz, D. Herrmann, M. Geissler, F. Krausz, and L. Veisz, *Density-transition based electron injector for laser driven wakefield accelerators*, Phys. Rev. ST Accel. Beams **13**, 091301 (2010).
- [147] S. Kuschel, M. B. Schwab, M. Yeung, D. Hollatz, A. Seidel, W. Ziegler, A. Sävert, M. C. Kaluza, and M. Zepf, *Controlling the Self-Injection Threshold in Laser Wakefield Accelerators*, Phys. Rev. Lett. **121**, 154801 (2018).
- [148] D. Guénot, D. Gustas, A. Vernier, B. Beaurepaire, F. Böhle, M. Bocoum, M. Lozano, A. Jullien, R. Lopez-Martens, A. Lifschitz, and J. Faure, *Relativistic electron beams driven by kHz single-cycle light pulses*, Nature Photonics **11**, 293 (2017).
- [149] L. Rovige, J. Huijts, I. Andriyash, A. Vernier, V. Tomkus, V. Girdauskas, G. Raciukaitis, J. Dudutis, V. Stankevicius, P. Gecys, M. Ouille, Z. Cheng, R. Lopez-Martens, and J. Faure, *Demonstration of stable long-term operation of a kilohertz laser-plasma accelerator*, Phys. Rev. Accel. Beams **23**, 093401 (2020).
- [150] C. McGuffey, A. G. R. Thomas, W. Schumaker, T. Matsuoka, V. Chvykov, F. J. Dollar, G. Kalintchenko, V. Yanovsky, A. Maksimchuk, K. Krushelnick, V. Y. Bychenkov, I. V. Glazyrin, and A. V. Karpeev, *Ionization Induced Trapping in a Laser Wakefield Accelerator*, Phys. Rev. Lett. **104**, 025004 (2010).
- [151] C. E. Clayton, J. E. Ralph, F. Albert, R. A. Fonseca, S. H. Glenzer, C. Joshi, W. Lu, K. A. Marsh, S. F. Martins, W. B. Mori, A. Pak, F. S. Tsung, B. B. Pollock, J. S. Ross, L. O. Silva, and D. H. Froula, *Self-Guided Laser Wakefield Acceleration beyond 1 GeV Using Ionization-Induced Injection*, Phys. Rev. Lett. **105**, 105003 (2010).
- [152] B. B. Pollock, C. E. Clayton, J. E. Ralph, F. Albert, A. Davidson, L. Divol, C. Filip, S. H. Glenzer, K. Herpoldt, W. Lu, K. A. Marsh, J. Meinecke, W. B. Mori, A. Pak, T. C. Rensink, J. S. Ross, J. Shaw, G. R. Tynan, C. Joshi, and D. H. Froula, *Demonstration of a Narrow Energy Spread, ~ 0.5 GeV Electron Beam from a Two-Stage Laser Wakefield Accelerator*, Phys. Rev. Lett. **107**, 045001 (2011).

- [153] J. S. Liu, C. Q. Xia, W. T. Wang, H. Y. Lu, C. Wang, A. H. Deng, W. T. Li, H. Zhang, X. Y. Liang, Y. X. Leng, X. M. Lu, C. Wang, J. Z. Wang, K. Nakajima, R. X. Li, and Z. Z. Xu, *All-Optical Cascaded Laser Wakefield Accelerator Using Ionization-Induced Injection*, Phys. Rev. Lett. **107**, 035001 (2011).
- [154] G. Golovin, S. Chen, N. Powers, C. Liu, S. Banerjee, J. Zhang, M. Zeng, Z. Sheng, and D. Umstadter, *Tunable monoenergetic electron beams from independently controllable laser-wakefield acceleration and injection*, Phys. Rev. ST Accel. Beams **18**, 011301 (2015).
- [155] A. J. Gonsalves, K. Nakamura, C. Benedetti, C. V. Pieronek, S. Steinke, J. H. Bin, S. S. Bulanov, J. van Tilborg, C. G. R. Geddes, C. B. Schroeder, J. Daniels, C. Tóth, L. Obst-Huebl, R. G. W. van den Berg, G. Bagdasarov, N. Bobrova, V. Gasilov, G. Korn, P. Sasorov, W. P. Leemans, and E. Esarey, *Laser-heated capillary discharge plasma waveguides for electron acceleration to 8 GeV*, Physics of Plasmas **27**, 053102 (2020).
- [156] C. Thaury, E. Guillaume, A. Lifschitz, K. Ta Phuoc, M. Hansson, G. Grittani, J. Gautier, J. P. Goddet, A. Tafzi, O. Lundh, and V. Malka, *Shock assisted ionization injection in laser-plasma accelerators*, Sci. Rep. **5**, 16310 (2015).
- [157] M. Mirzaie, S. Li, M. Zeng, N. A. M. Hafz, M. Chen, G. Y. Li, Q. J. Zhu, H. Liao, T. Sokollik, F. Liu, Y. Y. Ma, L. M. Chen, Z. M. Sheng, and J. Zhang, *Demonstration of self-truncated ionization injection for GeV electron beams*, Sci. Rep. **5**, 14659 (2015).
- [158] K. Floettmann, *Adiabatic matching section for plasma accelerated beams*, Phys. Rev. ST Accel. Beams **17**, 054402 (2014).
- [159] I. Dornmair, K. Floettmann, and A. R. Maier, *Emittance conservation by tailored focusing profiles in a plasma accelerator*, Phys. Rev. ST Accel. Beams **18**, 041302 (2015).
- [160] T. E. Faber, *Fluid dynamics for physicists* (Cambridge University Press, 1995).
- [161] A. J. Gonsalves, T. P. Rowlands-Rees, B. H. P. Broks, J. J. A. M. van der Mullen, and S. M. Hooker, *Transverse Interferometry of a Hydrogen-Filled Capillary Discharge Waveguide*, Phys. Rev. Lett. **98**, 025002 (2007).
- [162] T. Weineisen, B. Göppner, K. Schmid, M. Fuchs, H. Schröder, S. Karsch, and F. Grüner, *Density measurement in a laser-plasma-accelerator capillary using Raman scattering*, Phys. Rev. ST Accel. Beams **14**, 050705 (2011).
- [163] T. Chung *et al.*, *Computational fluid dynamics* (Cambridge University Press, 2002).
- [164] *ANSYS Fluent Theory Guide, Release 12.0* (2009).

-
- [165] T.-H. Shih, W. W. Liou, A. Shabbir, Z. Yang, and J. Zhu, *A new k - ϵ eddy viscosity model for high reynolds number turbulent flows*, *Comput. Fluids*. **24**, 227 (1995).
- [166] S. Patankar, *Numerical heat transfer and fluid flow* (Taylor & Francis, 2018).
- [167] S. Chapman and T. G. Cowling, *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases* (Cambridge University Press, 1990).
- [168] E. L. Cussler and E. L. Cussler, *Diffusion: mass transfer in fluid systems* (Cambridge University Press, 2009).
- [169] N. H. Matlis, A. J. Gonsalves, S. Steinke, J. van Tilborg, E. H. Matlis, B. Shaw, D. E. Mittelberger, C. G. R. Geddes, and W. P. Leemans, *Dynamics and density distributions in a capillary-discharge waveguide with an embedded supersonic jet*, *J. Appl. Phys.* **118**, 204506 (2015).
- [170] NIST, *Strong lines of Nitrogen*.
- [171] C. Rechatin, J. Faure, X. Davoine, O. Lundh, J. Lim, A. Ben-Ismaïl, F. Burgy, A. Tafzi, A. Lifschitz, E. Lefebvre, and V. Malka, *Characterization of the beam loading effects in a laser plasma accelerator*, *New J. Phys.* **12**, 045023 (2010).
- [172] E. Guillaume, A. Döpp, C. Thaury, A. Lifschitz, J.-P. Goddet, A. Tafzi, F. Sylla, G. Iaquanello, T. Lefrou, P. Rousseau, K. T. Phuoc, and V. Malka, *Physics of fully-loaded laser-plasma accelerators*, *Phys. Rev. ST Accel. Beams* **18**, 061301 (2015).
- [173] C. A. Lindstrøm, J. M. Garland, S. Schröder, L. Boulton, G. Boyle, J. Chappell, R. D’Arcy, P. Gonzalez, A. Knetsch, V. Libov, G. Loisch, A. Martinez de la Ossa, P. Niknejadi, K. Pöder, L. Schaper, B. Schmidt, B. Sheeran, S. Wesch, J. Wood, and J. Osterhoff, *Energy-Spread Preservation and High Efficiency in a Plasma-Wakefield Accelerator*, *Phys. Rev. Lett.* **126**, 014801 (2021).
- [174] D. R. Jones, M. Schonlau, and W. J. Welch, *Efficient global optimization of expensive black-box functions*, *J. Global Optim.* **13**, 455 (1998).
- [175] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Machine learning and the physical sciences*, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [176] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, *Nature* **521**, 436 (2015).
- [177] F. Chollet *et al.*, *Keras* (2015).

- [178] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, *Scikit-learn: Machine Learning in Python*, J. Mach. Learn. Res. **12**, 2825 (2011).
- [179] J.-L. Vay and R. Lehe, *Simulations for Plasma and Laser Acceleration*, Rev. Accel. Sci. Tech. **9**, 165 (2016).
- [180] R. Lehe, *Improvement of laser-wakefield accelerators: towards a compact free electron laser*, Ph.D. thesis, Ecole Polytechnique (2014).
- [181] K. Yee, *Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media*, IEEE Trans. Antennas Propag. **14**, 302 (1966).
- [182] Q. H. Liu, *The PSTD algorithm: A time-domain method requiring only two cells per wavelength*, Microw. Opt. Technol. Lett. **15**, 158 (1997).
- [183] I. Haber, R. Lee, H. Klein, and J. Boris, in Proc. Sixth Conf. on Num. Sim. Plasmas (Lawrence Livermore Laboratory, 1973) pp. 46–48.
- [184] H. Abe, N. Sakairi, R. Itatani, and H. Okuda, *High-order spline interpolations in the particle simulation*, Journal of Computational Physics **63**, 247 (1986).
- [185] J.-L. Vay, *Simulation of beams or plasmas crossing at relativistic velocity*, Phys. Plasmas **15**, 056701 (2008).
- [186] J. P. Boris, *Relativistic plasma simulation-optimization of a hybrid code*, in Proc. Fourth Conf. Num. Sim. Plasmas (1970) p. 3.
- [187] A. V. Higuera and J. R. Cary, *Structure-preserving second-order integration of relativistic charged particle trajectories in electromagnetic fields*, Phys. Plasmas **24**, 052104 (2017).
- [188] T. Esirkepov, *Exact charge conservation scheme for Particle-in-Cell simulation with an arbitrary form-factor*, Comput. Phys. Commun. **135**, 144 (2001).
- [189] J. Villasenor and O. Buneman, *Rigorous charge conservation for local electromagnetic field solvers*, Comput. Phys. Commun. **69**, 306 (1992).
- [190] T. Umeda, Y. Omura, T. Tominaga, and H. Matsumoto, *A new charge conservation method in electromagnetic particle-in-cell simulations*, Comput. Phys. Commun. **156**, 73 (2003).
- [191] R. L. Morse and C. W. Nielson, *Numerical Simulation of the Weibel Instability in One and Two Dimensions*, Phys. Fluids. **14**, 830 (1971).
- [192] B. Marder, *A method for incorporating Gauss' law into electromagnetic PIC codes*, J. Comput. Phys **68**, 48 (1987).

-
- [193] J.-L. Vay, C. Geddes, E. Cormier-Michel, and D. Grote, *Numerical methods for instability mitigation in the modeling of laser wakefield accelerators in a Lorentz-boosted frame*, Journal of Computational Physics **230**, 5908 (2011).
- [194] J. P. Boris and R. Lee, *Nonphysical self forces in some electromagnetic plasma-simulation algorithms*, J. Comput. Phys **12**, 131 (1973).
- [195] R. Courant, K. Friedrichs, and H. Lewy, *Über die partiellen Differenzgleichungen der mathematischen Physik*, Math. Ann. **100**, 32 (1928).
- [196] J. B. Cole, *A high-accuracy realization of the yee algorithm using non-standard finite differences*, IEEE Trans. Microw. Theory Techn. **45**, 991 (1997).
- [197] A. Pukhov, *Three-dimensional electromagnetic relativistic particle-in-cell code VLPL (Virtual Laser Plasma Lab)*, J. Plasma Phys. **61**, 425 (1999).
- [198] M. Karkkainen, E. Gjonaj, T. Lau, and T. Weiland, *Low-Dispersion Wake Field Calculation Tools*, in Proc. ICAP, 1 (2006) p. 35.
- [199] R. Lehe, C. Thaury, E. Guillaume, A. Lifschitz, and V. Malka, *Laser-plasma lens for laser-wakefield accelerators*, Phys. Rev. ST Accel. Beams **17**, 121301 (2014).
- [200] Y. Ohmura and Y. Okamura, *Staggered grid pseudo-spectral time-domain method for light scattering analysis*, Piers Online **6**, 632 (2010).
- [201] H. Vincenti and J. L. Vay, *Ultrahigh-order Maxwell solver with extreme scalability for electromagnetic PIC simulations of plasmas*, Comput. Phys. Commun. **228**, 22 (2018).
- [202] G. Blaclard, H. Vincenti, R. Lehe, and J. L. Vay, *Pseudospectral Maxwell solvers for an accurate modeling of Doppler harmonic generation on plasma mirrors with particle-in-cell codes*, Phys. Rev. E **96**, 033305 (2017).
- [203] B. Fornberg, *Generation of Finite Difference Formulas on Arbitrarily Spaced Grids*, Math. Comput. **51**, 699 (1988).
- [204] B. B. Godfrey and J. L. Vay, *Numerical stability of relativistic beam multidimensional PIC simulations employing the Esirkepov algorithm*, J. Comput. Phys **248**, 33 (2013).
- [205] X. Xu, P. Yu, S. F. Martins, F. S. Tsung, V. K. Decyk, J. Vieira, R. A. Fonseca, W. Lu, L. O. Silva, and W. B. Mori, *Numerical instability due to relativistic plasma drift in EM-PIC simulations*, Comput. Phys. Commun. **184**, 2503 (2013).
- [206] Y. Kazimura, J. I. Sakai, T. Neubert, and S. V. Bulanov, *Generation of a Small-Scale Quasi-Static Magnetic Field and Fast Particles during the Collision of Electron-Positron Plasma Clouds*, Astrophys. J. Lett. **498**, L183 (1998).

- [207] L. O. Silva, R. A. Fonseca, J. W. Tonge, J. M. Dawson, W. B. Mori, and M. V. Medvedev, *Interpenetrating Plasma Shells: Near-Equipartition Magnetic Field Generation and Nonthermal Particle Acceleration*, *Astrophys. J.* **596**, L121 (2003).
- [208] A. Spitkovsky, *Particle Acceleration in Relativistic Collisionless Shocks: Fermi Process at Last?*, *Astrophys. J.* **682**, L5 (2008).
- [209] J. L. Vay, C. G. Geddes, E. Esarey, C. B. Schroeder, W. P. Leemans, E. Cormier-Michel, and D. P. Grote, *Modeling of 10 GeV-1 TeV laser-plasma accelerators using Lorentz boosted simulations*, *Phys. Plasmas* **18**, 123103 (2011).
- [210] S. F. Martins, R. A. Fonseca, L. O. Silva, W. Lu, and W. B. Mori, *Numerical simulations of laser wakefield accelerators in optimal Lorentz frames*, *Comput. Phys. Commun.* **181**, 869 (2010).
- [211] S. F. Martins, R. A. Fonseca, W. Lu, W. B. Mori, and L. O. Silva, *Exploring laser-wakefield-accelerator regimes for near-term lasers using particle-in-cell simulation in Lorentz-boosted frames*, *Nat. Phys.* **6**, 311 (2010).
- [212] K. Peters, M. Kirchen, and R. Lehe, *Implementation of high-order particle shape factors in FBPIC* (2017), unpublished.
- [213] R. Lehe, *Discrete Hankel transform formalism supporting high-order azimuthal modes in FBPIC* (2017), unpublished.
- [214] R. Lehe, *Implementation of a transverse perfectly matched layer (PML) in FBPIC* (2019), unpublished.
- [215] R. Lehe, M. Kirchen, and L. Jeppe, *Accurate spectral-cylindrical deposition using Ruyten-corrected shape factors and a modified radial cell volume* (2019), unpublished.
- [216] R. Lehe and M. Kirchen, *Cross-deposition scheme for charge conservation in a spectral-cylindrical algorithm* (2018), unpublished.
- [217] R. Lehe, M. Kirchen, S. Jalas, and L. Jeppe, *Implementation of advanced physics features in FBPIC* (2020), unpublished.
- [218] M. Kirchen, *Parallelization of a spectral, quasi-cylindrical Particle-In-Cell code*, Master's thesis, Universität Hamburg (2015).
- [219] M. Kirchen and R. Lehe, *Multi-threaded CPU implementation of FBPIC* (2017), unpublished.
- [220] L. Jeppe, M. Kirchen, and R. Lehe, *Improvements to the GPU implementation of FBPIC* (2020), unpublished.

-
- [221] B. Godfrey, *The IPR0P Three-Dimensional Beam Propagation Code* (Defense Technical Information Center, 1985).
- [222] M. Guizar-Sicairos and J. C. Gutiérrez-Vega, *Computation of quasi-discrete Hankel transforms of integer order for propagating optical wave fields*, J. Opt. Soc. Am. A **21**, 53 (2004).
- [223] M. Cree and P. Bones, *Algorithms to numerically evaluate the Hankel transform*, Comput. Math. with Appl. **26**, 1 (1993).
- [224] L. Yu, M. Huang, M. Chen, W. Chen, W. Huang, and Z. Zhu, *Quasi-discrete Hankel transform*, Opt. Lett. **23**, 409 (1998).
- [225] W. M. Ruyten, *Density-Conserving Shape Factors for Particle Simulations in Cylindrical and Spherical Coordinates*, J. Comput. Phys **105**, 224 (1993).
- [226] D. J. Larson, D. W. Hewett, and A. Langdon, *Correction factors for PIC accumulation on radial grids*, Comput. Phys. Commun. **90**, 260 (1995).
- [227] J. Verboncoeur, *Symmetric Spline Weighting for Charge and Current Density in Particle Simulation*, J. Comput. Phys **174**, 421 (2001).
- [228] P. Yu, A. Davidson, A. Tableman, V. K. Decyk, F. S. Tsung, and W. B. Mori, *Lorentz boosted frame simulation of Laser wakefield acceleration using hybrid Yee-fft solver in quasi-3d geometry*, in Proc. 6th Int. Part. Accel. Conf. (2015) p. 691.
- [229] P. Lee and J.-L. Vay, *Efficiency of the Perfectly Matched Layer with high-order finite difference and pseudo-spectral Maxwell solvers*, Comput. Phys. Commun. **194**, 1 (2015).
- [230] O. Klein and Y. Nishina, *Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac*, Z. Phys. **52**, 853 (1929).
- [231] A. Huebl, R. Lehe, J.-L. Vay, D. P. Grote, I. Sbalzarini, S. Kuschel, D. Sagan, F. Pérez, F. Koller, and M. Bussmann, *openPMD: A meta data standard for particle and mesh based data* (2018).
- [232] J.-L. Vay, R. Lehe, D. P. Grote, and A. Huebl, *PICMI: Particle-In-Cell Modeling Interface* (2020).
- [233] Intel, *Intel Math Kernel Library (MKL)* (2020).
- [234] M. Frigo and S. G. Johnson, *The Design and Implementation of FFTW3*, Proc. IEEE **93**, 216 (2005).
- [235] S. K. Lam, A. Pitrou, and S. Seibert, *Numba: A llvm-based python jit compiler*, in Proc. Sec. Workshop on LLVM Comp. Infra. HPC (2015) p. 1.

- [236] C. Lattner and V. Adve, *LLVM: A compilation framework for lifelong program analysis & transformation*, in International Symposium on Code Generation and Optimization (2004) p. 75.
- [237] NVIDIA, *CUDA* (2020).
- [238] N. Bell and J. Hoberock, *Thrust: A Productivity-Oriented Library for CUDA*, in GPU Computing Gems Jade Edition (2012) p. 359.
- [239] R. Okuta, Y. Unno, D. Nishino, S. Hido, and C. Loomis, *CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations* (2017).
- [240] Message Passing Interface Forum, *MPI: A Message-Passing Interface Standard, Version 3.1* (2015).
- [241] L. D. Dalcin, R. R. Paz, P. A. Kler, and A. Cosimo, *Parallel distributed computing using Python*, Adv. Water Resour. **34**, 1124 (2011).
- [242] Intel, *Intel Threading Building Blocks Library (TBB)* (2020).
- [243] L. Dagum and R. Menon, *OpenMP: an industry standard API for shared-memory programming*, IEEE Comput. Sci. Eng. **5**, 46 (1998).
- [244] L. S. Blackford, A. Petitet, R. Pozo, K. Remington, R. C. Whaley, J. Demmel, J. Dongarra, I. Duff, S. Hammarling, G. Henry, *et al.*, *An updated set of basic linear algebra subprograms (BLAS)*, ACM Trans. Math. Softw. **28**, 135 (2002).
- [245] Jülich Supercomputing Centre, *JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre*, Journal of large-scale research facilities **5**, A135 (2019).
- [246] Jülich Supercomputing Centre, *JUWELS Booster Module Configuration* (2020).
- [247] B. B. Godfrey, J. L. Vay, and I. Haber, *Numerical stability analysis of the pseudo-spectral analytical time-domain PIC algorithm*, J. Comput. Phys **258**, 689 (2014).
- [248] B. B. Godfrey, J. L. Vay, and I. Haber, *Numerical stability improvements for the pseudospectral EM PIC algorithm*, IEEE Trans. Plasma Sci. **42**, 1339 (2014).
- [249] B. B. Godfrey and J.-L. Vay, *Suppressing the numerical Cherenkov instability in {FDTD} {PIC} codes*, J. Comput. Phys **267**, 1 (2014).
- [250] B. B. Godfrey and J. L. Vay, *Improved numerical Cherenkov instability suppression in the generalized PSTD PIC algorithm*, Comput. Phys. Commun. **196**, 221 (2015).

-
- [251] P. Yu, X. Xu, V. K. Decyk, F. Fiuza, J. Vieira, F. S. Tsung, R. A. Fonseca, W. Lu, L. O. Silva, and W. B. Mori, *Elimination of the numerical Cerenkov instability for spectral EM-PIC codes*, *Comput. Phys. Commun.* **192**, 32 (2015).
- [252] P. Yu, X. Xu, A. Tableman, V. K. Decyk, F. S. Tsung, F. Fiuza, A. Davidson, J. Vieira, R. A. Fonseca, W. Lu, L. O. Silva, and W. B. Mori, *Mitigation of numerical Cerenkov radiation and instability using a hybrid finite difference-FFT Maxwell solver and a local charge conserving current deposit*, *Comput. Phys. Commun.* **197**, 144 (2015).
- [253] F. Li, P. Yu, X. Xu, F. Fiuza, V. K. Decyk, T. Dalichaouch, A. Davidson, A. Tableman, W. An, F. S. Tsung, R. A. Fonseca, W. Lu, and W. B. Mori, *Controlling the numerical Cerenkov instability in PIC simulations using a customized finite difference Maxwell solver and a local FFT based current correction*, *Comput. Phys. Commun.* **214**, 6 (2017).
- [254] J. L. Vay, D. P. Grote, R. H. Cohen, and A. Friedman, *Novel methods in the Particle-In-Cell accelerator Code-Framework Warp*, *Comput. Sci. Discov.* **5**, 014019 (2012).
- [255] J. L. Vay, A. Almgren, J. Bell, L. Ge, D. P. Grote, M. Hogan, O. Kononenko, R. Lehe, A. Myers, C. Ng, J. Park, R. Ryne, O. Shapoval, M. Thévenet, and W. Zhang, *Warp-X: A new exascale computing platform for beam-plasma simulations*, *Nucl. Instrum. Methods Phys. Res. A* **909**, 476 (2018).
- [256] X. Xu, F. Li, F. S. Tsung, T. N. Dalichaouch, W. An, H. Wen, V. K. Decyk, R. A. Fonseca, M. J. Hogan, and W. B. Mori, *On numerical errors to the fields surrounding a relativistically moving particle in PIC codes*, *J. Comput. Phys.* **413**, 109451 (2020).

Acknowledgement

First and foremost, I would like to thank Dr. Andreas Maier for giving me the opportunity to join his, back then still young, research group at the University of Hamburg and, in particular, for supervising this thesis. Over the past several years, Andi has supported me in every way possible, provided guidance whenever I needed it, and enabled me to build my own scientific career. Even more importantly, he created a fair and productive work environment and through his leadership enabled a group of young researchers to grow into an outstanding team that, through its accomplishments, has earned itself a place in the history of laser-plasma acceleration.

Second, I would like to thank Prof. Dr. Wim Leemans for examining this thesis, for his great support in this early phase of my career and for welcoming me as an affiliate scientist at the Lawrence Berkeley National Laboratory (LBNL) when he was still head of the BELLA Center. Wim's profound knowledge of virtually any aspect of laser-plasma acceleration, even when it comes to the smallest technical details, has always impressed me and struck me as something to strive for.

I would also like to thank Dr. Jean-Luc Vay for giving me the opportunity to intermittently join his research group at Berkeley Lab to collaborate on the development of new particle-in-cell methods and codes. Jean-Luc is not only one of the greatest experts in this field, but also an extremely likeable and supportive person, and I felt very honored to work in his group.

Moreover, I want to express my deepest gratitude to Remi Lehe, who introduced me to particle-in-cell modeling and taught me almost all the foundational knowledge that enabled me to work in this field. Remi is not only an extremely productive and talented physicist, but also one of the most humble and kind persons that I have ever met. I am glad to have Remi as a colleague and friend, and thankful that several years ago he invited me join the development of what has now become the widely used particle-in-cell code FBPIC.

I want to thank Sören Jalaš for being the best colleague and friend anyone could ask for, and especially for his support whenever I needed it. Many of the results of this work have benefited from the highly cooperative work style that Sören and I adopted during the last years, and therefore I hope that we can continue to work together in the future. In the same spirit, I would like to thank Paul Winkler, Philipp Messner

and Vincent Leroux for the enjoyable moments we spent together, not only at the LUX accelerator but also outside of work.

Furthermore, I want to thank all other past and present members of the LUX group. I really enjoyed being a part of this team. For their immediate contributions to the results of this thesis, I want to thank Matthias Schnepf, Chris Werle, Timo Eichner, Thomas Hülsenbusch, Lars Hübner and Laurids Jeppe. In addition, I want to thank Heike Kaminski and Fenna Reinholdt. I also thank the technical groups at Deutsches Elektronen-Synchrotron (DESY) and the University workshop for their support during the construction phases of the experiment.

Over the past years at DESY and during my visits to Berkeley Lab, I have had the pleasure of meeting many other great people. Some of them have given me good advice, others have taught me new things, and still others have engaged me in great discussions about physics and life over a cocktail on Elba. I would like to express my gratitude to all these people.

Finally, I want thank my parents, Rita and Günther, for their endless support during my whole life, all of my friends for the wonderful moments we spent together, and Moni for her unconditional love and understanding during these demanding past couple of years.