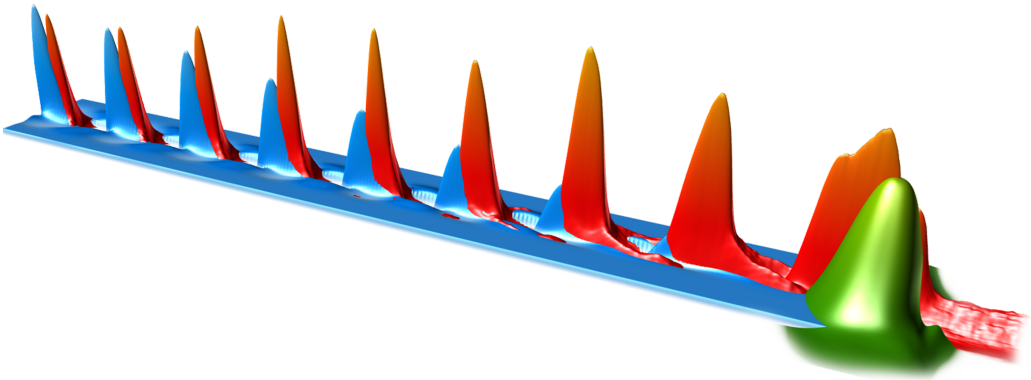
TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO



Ultra-fast plasma based acceleration modelling: on reduced algorithms to predict the future generation of particle accelerators

Anton Helm

Supervisors: Doctor Ricardo Parreira de Azambuja Fonseca
Doctor Jorge Miguel Ramos Domingues Ferreira Vieira
Doctor Warren Bicknell Mori

Thesis approved in public session to obtain the PhD Degree in

Physics

Jury final classification: **Pass with Distinction**

2021

Obra entregue para apreciação da prova de Doutoramento em Física para obtenção do grau de Doutor concedido pela Universidade de Lisboa, no Instituto Superior Técnico, ao abrigo do D.L. 74/2006, de 24 de março.

Lisboa, 18 de fevereiro de 2021

O presidente do júri
Homologo

TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Ultra-fast plasma based acceleration modelling: on reduced algorithms to predict the future generation of particle accelerators

Anton Helm

Supervisors: Doctor Ricardo Parreira de Azambuja Fonseca
Doctor Jorge Miguel Ramos Domingues Ferreira Vieira
Doctor Warren Bicknell Mori

Thesis approved in public session to obtain the PhD Degree in

Physics

Jury final classification: **Pass with Distinction**

Chairperson: Doctor Luís Paulo da Mota Capitão Lemos Alves,
Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor Ricardo Parreira de Azambuja Fonseca, Escola de Tecnologias e Arquitetura, ISCTE - Instituto Universitário de Lisboa

Doctor Jens Osterhoff, Deutsches Elektronen-Synchrotron DESY, Germany

Doctor Eduardo Paulo Jorge da Costa Alves, High Energy Density Science Division, SLAC National Accelerator Laboratory, USA

Doctor Marta Leitão Mota Fajardo, Instituto Superior Técnico, Universidade de Lisboa

Doctor Marija Vranic, Instituto Superior Técnico, Universidade de Lisboa

Funding Institutions

Fundação para a Ciência e Tecnologia (FCT)

2021

Abstract

Plasma-based accelerators (PBA) are strong candidates for the next generation of particle accelerators as conventional accelerators reached their limitations in terms of accelerating gradients. To understand the underlying physics and to model those accelerators, the particle-in-cell (PIC) method is commonly used. While accurate, simulations based on PIC require to resolve the shortest and fastest scales, making it computationally expensive. In the case when an intense laser pulse is propagating through a plasma, the shortest scales are associated with the laser wavelength, typically in µm-range, and are required to be resolved over acceleration distances, in meter-ranges. Reduced solvers can gain high computational speedup.

In this thesis, the implementation of a ponderomotive guiding center (PGC) solver is presented. In the case of PGC, only the laser envelope instead of the short laser wavelength is resolved. The relaxed resolution allows for high computational speedups. For PGC, the implementation, the numerical stability, and the parallel scalability are obtained, discussed, and analyzed. As PGC requires less computational resources than a full PIC simulation, it is a prime candidate to perform parametric studies. By comparing physical equivalent scenarios of particle for injection, the applicability for that kind of simulation is discussed.

Additionally, PGC allows to model scenarios that are not feasible with PIC. Such an example is the ionization seeding for the self-modulation instability for the AWAKE experiment at CERN. For the AWAKE experiment, a laser is used to seed the self-modulation instability and cut a long proton bunch in small beamlets which drive accelerating wakes inside a plasma. For this, the PGC implementation is extended to include laser-induced ionization to allow to model the self-modulation instability for the AWAKE experiment self-consistently. More specifically, a study for small misalignment in the order of the width of the proton bunch is presented.

Keywords: plasma-based accelerators, particle-in-cell simulation, parallel computing, envelope model

Resumo

Os plasma-based accelerators (PBA) são fortes candidatos à próxima geração de aceleradores de partículas, visto que os aceleradores convencionais atingiram suas limitações em termos de gradientes de aceleração. Para entender a física subjacente e modelar esses aceleradores, o método de particle-in-cell (PIC) é comumente usado. Embora precisas, as simulações baseadas no PIC exigem resolver as escalas mais curtas e mais rápidas, tornando-o computacionalmente caro. No caso em que um pulso intenso de laser está se propagando através de um plasma, as escalas mais curtas são associadas ao comprimento de onda do laser, tipicamente na faixa de µm, e precisam ser resolvidas nas distâncias de aceleração, na faixa de metros.

Nesta tese, é apresentada a implementação de um ponderomotive guiding center (PGC) solver. No caso do PGC, apenas o envelope do laser, em vez do comprimento de onda curto do laser, é resolvido. A resolução relaxada permite altas acelerações computacionais. Para o PGC, a implementação, a estabilidade numérica e a escalabilidade paralela são obtidas, discutidas e analisadas. O PGC é o principal candidato para realizar estudos paramétricos. Ao comparar cenários físicos equivalentes de partículas para injeção, é discutida a aplicabilidade para esse tipo de simulação. Além disso, o PGC permite modelar cenários que não são viáveis com o PIC. Um exemplo é a propagação de ionização para a instabilidade de auto-modulação (SMI) para o experimento AWAKE no CERN. Para o experimento AWAKE, um laser é usado para propagar a SMI e cortar um longo grupo de prótons em pequenas feixes de luz que conduzem a acelerações de vigília dentro de um plasma. Para isso, a implementação do PGC é estendida para incluir a ionização induzida por laser para permitir modelar a SMI para o experimento AWAKE de forma consistente. Mais especificamente, é apresentado um estudo para pequenos desalinhamentos na ordem da largura do bunch de prótons.

Keywords: aceleradores baseados em plasma, simulação particle-in-cell, computação paralela, algoritmos reduzidos, modelo de envelope

Acknowledgment

I think of science as a community of people.
A community of people bound by certain ethical precepts.

—

Lee Smolin
transcribed from
"AI Podcast #79 with Lex Fridman"

Throughout my life, my experience during my doctoral program was, by far, an immersive adventure. Part of it was the possibility to explore different cultures and dive deep into examining diverse academic topics. My Ph.D. let me acquire experiences and knowledge in an amount I could have never imagined before. In this sense, naming all the interactions would exceed any limits by a vast amount. Nevertheless, I would like to mention some in particular.

I have to thank Ricardo Fonseca. His contribution, expertise, and guidance were what I required to succeed. This thesis would not be as it is without his contributions. I enjoyed the in-depth discussions about PIC codes, numerics, and computer-related topics. Thank you for this.

I have to thank Jorge Vieira. His expertise in plasma-based accelerators is outstanding. No matter what kind of question I would ask him. He always had an answer and even a paper for me to study more precisely.

I have to thank Warren Mori. For one year, I was able to wander around the paths

of Dawson and explore UCLA. I appreciated all his kindness at UCLA, the effort, and all the trust he put in my work.

I have to thank Prof. Luis Silva. His dedication to science and the Group for Lasers and Plasmas (GoLP) makes it stand out on a worldwide level.

I have to thank my colleagues, Thales S., Joana M., and Lígia A., for the trust in my work, exploring PGC, and all the bugs that got uncovered.

I have to thank my colleagues, Marija V., Thomas G., Giannandrea I., Fabrizio G., and Fábio C., for the time at the pool with the exciting discussions.

I have to thank all of the members of the GoLP group. For being part of a fantastic group and making it as such. I also have to thank all the members of the UCLA Plasma Simulation Group. For all the discussions and all the help I received during my time at UCLA.

I have to thank the APPLAuSE program for allowing me to move to Lisbon, attend IST, and visit UCLA for one year.

Also, I would like to thank my family for the big and strong support. In particular, my mother who supported me not just throughout my Ph.D. and most likely she is the reason why I ended up in physics in the first place. In diesem Sinne, vielen Dank, Mama. Für die ganze Unterstützung möchte ich mich bei dir aus ganzem Herzen bedanken. Ohne dich und deine Unterstützung, wäre dies alles hier nicht möglich gewesen.

Lastly, from the bottom of my heart, I would acknowledge the strong support from my girlfriend, Ana. She dealt with me as an individual over the past months, and I know I was not much of a help, but she also supported me strongly. I can not express how grateful I am for having her on my side.

Thank you to everyone!

Contents

Introduction

Dass ich erkenne, was die Welt
Im Innersten zusammenhält.
(eng.: So that I may perceive whatever holds
The world together in its inmost folds.)

Faust. Eine Tragödie (Vers 382 f.),
Johann Wolfgang von Goethe

The idea of knowing what happens on the smallest scales has always been part of human curiosity. Rutherford showed one way of achieving it in 1911. By utilizing a beam of charged particles, he was able to calculate the nuclei's size [1]. In his well-known gold foil experiment, the particles were emitted by a radioactive element. Probing matter on even smaller scales requires higher energies with significant control. More recently, the collision of proton beams at the Large Hardon Collider (LHC) confirmed the Higgs boson [2] and explained why particles have a mass. While Higgs's work was acknowledged with a Nobel prize, it is clear it would not be achievable without decades of development of particle accelerators and, in particular, raising the energy of the particle beams. Particle accelerators became a fundamental tool not only in the realm of physics but also in hospitals to diagnose and treat patients [3], in the semiconductor industry [4] and as a light source for X-ray beams [5].

Due to the limitation in acceleration gradients, current accelerator technologies require progressively larger facilities to accelerate particles. In the case of the LHC, particles are accelerated in a 27 km-ring to a cumulative energy of 13 TeV. With directly allocated resources of € 1.0 billion for the years 2009–2012 [6], the LHC is an expensive scientific endeavor. The reason for requiring such great resources comes down to the point that with high accelerating gradients, around 100 MV/m, the metallic walls of a conventional accelerator are getting destroyed. An alternative approach is to use plasma. A plasma is a quasi-neutral cloud of ionized gas, and being an already disrupted gas, it can sustain higher accelerating gradients. This thesis focuses more on a complementary particle acceleration technique, called the plasma-based acceleration (PBA). It promises to reduce the size of particle acceleration by a thousand-fold or more, which inherently reduces the cost dramatically. Given the wide range of applications for particle accelerators in all the different science branches, cheaper and smaller accelerator facilities have a great potential of a broad impact on a wide range of scientific discoveries.

Particle accelerators

Charged particles can be accelerated when they are placed in an electric field. The gain of energy for the particles is equivalent to the crossed potential difference of the accelerating region. One of the first generations of particle accelerators were the Cockcroft-Walton generators [7]. Such generators were first used in 1932 by the scientist John Cockcroft and Ernest Walton and consisted of a capacitor bank to generate a high direct current (DC) voltage. While scientists were able to carry out an atom-splitting experiment [8], which was awarded a noble prize, the approach is not scalable for high voltages as spark discharges occur. Ising already proposed an alternative approach in 1924 [9]. He planned to apply consequently the same voltage to a particle through alternating fields. Even though Ising himself was not successful, it is a building block for conventional accelerators. The rapid

development in accelerator technology in the following years, including advances in superconducting magnet design and radiofrequency (RF) cavities, has led to the high peak energies available today.

However, the large electric fields stored in RF cavities lead to electrons being liberated from the surface walls and limiting the accelerating field [10]. Nevertheless, the particle energy can be increased by using a circular geometry for the accelerator where the same fields are applied multiple times to the particle beam as it circulates and gains energy. Such machines are known as synchrotrons and hold the record for the highest particle energies, e.g. the Large Electron Positron Collider (LEP) with 209 GeV [11] for electrons and positrons and the Large Hardon Collider (LHC) with 13 TeV for protons. Due to continues transverse acceleration a particle with a mass m traveling at energy E radiates synchrotron radiation of power

$$P \propto E^4 / \left(m^4 R^2 \right) \tag{0.1}$$

to stay on a circular trajectory of radius R. As the radiated power scales with m^{-4}, electrons tend to radiate at a power $1836^4 \approx 10^{13}$ times higher than of protons. With the inverse dependence of the synchrotron radius in (0.1), increasing the radius seems to be an alternative, but requires an accelerator to circle the planet to achieve a few TeV for electrons[†]. Hence, only linear geometries for electron acceleration are feasible. Therefore, the proposed International Linear Collider (ILC) is based on a linear scheme and aims to deliver 1 TeV electron bunches [13]. It will use state-of-the-art accelerating structures based on superconducting RF cavities. Unfortunately, the electric field inside such cavities is limited by resistive dissipation of electromagnetic energy in the walls. In case of the ILC the average acceleration field is intended to be around 31.5 MV/m which would require to have a 20.5 km

[†]On January 29, 1954, Fermi projected as part of his speech as retiring president of the American Physical Society that by 1994 a "Globatron" could be available. His prediction was based on the evolution of particle accelerators. The Globotron would span over the entire globe, delivering particles with an energy of up to 5×10^{15} eV [12].

long site to accelerate electrons and positrons to 250 GeV cumulative energy[‡]. The estimated construction cost for ILC is €7.2 billion.

Although extensive research in the field of RF cavities has been conducted, which led to the development of superconducting Niobium cavities [14], the average accelerating field is currently limited to 45 MV/m. Moreover, due to the disruption of the accelerator cavity walls, the highest measured accelerating field in RF cavities is limited to 200 MV/m [10]. Given the limitations, alternative methods for acceleration have been proposed, such as direct-laser acceleration [15], inverse free-electron laser [16] and plasma-based acceleration (PBA). The latter one has received the most attention.

Plasma-based accelerators

The use of plasmas as an accelerating medium has been proposed to overcome the limitation of the available acceleration gradient. Given that a plasma is a quasi-neutral cloud of electrons and ions, disruption has already occurred and would allow for higher accelerating gradients.

In 1979, Tajima and Dawson proposed using a laser pulse to generate a laser-driven electron accelerator. When an intense laser pulse propagates through a plasma, its radiation pressure can sweep the electrons aside while the heavy ions remain approximately static. Afterward, as the intensity of the pulses eases, the expelled electrons will feel an electrostatic restoring force caused by the remaining ions. This force will accelerate the electrons back, but like a harmonic oscillator, the electrons overshoot and continue in oscillatory motion at the characteristic plasma frequency ω_p. This motion introduces a density wave and, as a charge imbalance is present, a longitudinal electric field is established, which trails together with the density wave of the laser pulse. The acceleration scheme is known as laser-wakefield acceleration (LWFA).

[‡]Additional upgrades of the machine could deliver 1 TeV by expending the site length to 40 km.

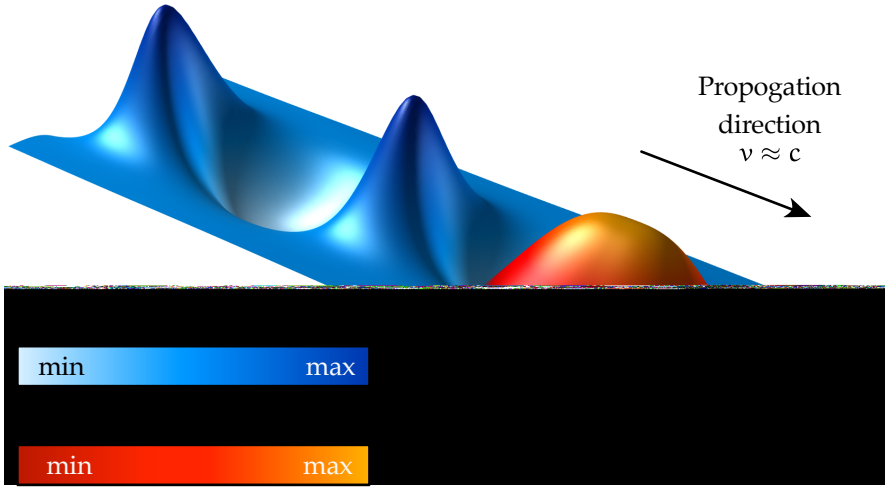


Figure 0.1: Schematic illustration of a laser-wakefield acceleration (LWFA) using simulation data. Shown here is the propagation of a laser pulse through a plasma. The laser pulse excites a density wave which is trailing the laser pulse. Associate with the charge imbalance is an electric field.

Moreover, a similar principle applies if a charged particle bunch is used to drive the wakes [18], known as plasma wakefield acceleration (PWFA). In this case, the space-charge forces of the particle bunch are responsible for perturbing the plasma. Both LWFA and PWFA are subgroups of PBA

Motivation and thesis outline

The physics behind PBA is highly complex and nonlinear. Hence, a purely theoretical description lacks to capture all observable phenomena, and numerical simulations based on the particle-in-cell (PIC) algorithm are required. While accurate, codes based on the PIC algorithm are computationally expensive. This thesis aims to implement, verify, and study a reduced solver with the potential to reduce the computational cost significantly. A reduction of the computational cost enables efficient exploration of vast parameter space, enables end-to-end full-scale simula-

tions, and promotes scientific discoveries.

The thesis is split in five chapters. In chapter 1, the introduction is extended to provide a theoretical background for the topics discussed later on. The purpose for the theoretical background is to give a brief overview instead of covering each detail individually.

The main focus of this thesis is to introduce and describe the implementation of the reduced pondermotive guiding center (PGC) solver based on the propagation of the envelope rather than the fast oscillating laser field in chapter 2. This allows to obtain computational speedups for laser-driven PBAs in the order of $\left(\omega_0/\omega_p\right)^2$ where ω_0 is the laser frequency, ω_p is the plasma frequency, so that the relation $\omega_0 > \omega_p$ applies. The presented implementation focuses mainly on the 3D implementation as in these geometries, the full physical aspects are recovered, and the impact arising from computational gains is the greatest. The discussion is concluded with a full derivation of the stability condition for the 3D version.

A major part of running numerical simulations is to take the full advantage of the available computational resources. For this, in chapter 3 the scalability of PGC on modern high-performance computing (HPC) systems is presented and discussed. By utilizing shared and distributed memory parallelization, the scalability of PGC for utilizing full HPC systems is shown.

One great advantage which arises with the reduced computational cost is the possibility of performing parametric studies. Hence, it is beneficial for designing future generations of particle accelerators as it allows them to find optimal parameters. Chapter 4 includes a review of using parametric studies for finding optimal parameters for down-ramp injection in the European Plasma Research Accelerator with eXcellence in Applications (EuPRAXIA) framework.

As the speedup of the PGC algorithm scales with the ratio between laser frequency and the plasma frequency, cases with $\omega_0/\omega_p > 10^3$ become feasible to be studied. One of those examples is the ionization seeding of the self-modulation instability (SMI). In chapter 5, the use of PGC combined with an ionization

model to model the full-scale experimental setup of the Advanced WAKEfield Experiment (AWAKE) experiment is presented. The studies are extended to include a discussion on the growth of the hosing instability during misalignment scenarios.

Original contributions

- Helm, A. et al., "Self-consistent modeling of relativistic ionization front for seeded self-modulation", *to be submitted to J. Plasma Phys.*
- Helm, A. et al., "Reduced laser wakefield acceleration modeling based on ponderomotive guidingcenter solver using particle-in-cell code OSIRIS", *submitted to J. Comput. Phys.*
- Miller, K. et al., "Dynamic load balancing with enhanced shared-memory parallelism for particle-in-cell codes," Comput. Phys. Commun. **259**, 107633 (2021).
- Assmann, R. W. et al., "EuPRAXIA Conceptual Design Report," Eur. Phys. J. Spec. Top. **229**, 3675–4284 (2020).
- Gorn, A. et al., "Proton beam defocusing in AWAKE: comparison of simulations and measurements," Plasma Phys. Control. Fusion **62**, 125023 (2020).
- Turner, M. et al., "Experimental study of wakefields driven by a self-modulating proton bunch in plasma," Phys. Rev. Accel. Beams **23**, 081302 (2020).
- Silva, T. et al., "On the use of the envelope model for down-ramp injection in laser-plasma accelerators", Plasma Phys. Control. Fusion **62**(2), 024001 (2020).

- Nghiem, P. et al., "Eupraxia, A Step Toward A Plasma-Wakefield Based Accelerator With High Beam Quality," J. Phys.: Conf. Ser., **1350**, 012068–10 (2019).
- Gschwendtner, E. et al., "Proton-driven plasma wakefield acceleration in AWAKE" Phil. Trans. R. Soc. A, **377**(2151), 20180418–9 (2019).
- Adli, E. et al. "Experimental Observation of Proton Bunch Modulation in a Plasma at Varying Plasma Densities" Phys. Rev. Lett., **122**(5), 054802 (2019).
- Turner, M. et al., "Experimental Observation of Plasma Wakefield Growth Driven by the Seeded Self-Modulation of a Proton Bunch," Phys. Rev. Lett., **122**(5), 054801 (2019).
- Adli, E. et al., "Acceleration of electrons in the plasma wakefield of a proton bunch," Nature, **561**(7723), 363–367 (2018).

1 Theoretical background: plasma-based accelerators

A Physicist is just an atom's way of looking at itself.

Niels Bohr

Using laser pulses to drive a plasma-based particle accelerator was initially proposed about four decades ago by the pioneering work of Tajima and Dawson [17]. Due to the usage of a laser pulse as a driver, it is referred to as LWFA. Since the initial proposal, LWFA and the field of PBA, in general, has seen tremendous progress, experimentally, theoretically, and computationally. Most of its growth connects to the development of the chirped-pulse amplification (CPA) laser technology [19]. CPA allows to generate short and intense laser pulse for various application, e.g. laser eye surgery [20–22], and manufacturing [23]. Since it is crucial for so many fundamental applications, the CPA technology was awarded a Nobel prize in 2018.

While before 2004 several experiments demonstrated the acceleration of electrons [24–30], the quality of the accelerated electron bunch was less than desired. Typically, the accelerated electron bunch was characterized by an exponential energy distribution and only a small fraction of the electron bunch reached energies above 100 MeV. This status changed significantly in 2004 when three groups presented high-quality electron bunches [31–33]. These works resulted from a higher

degree of control of the laser and plasma parameters, improved diagnostics techniques, and a greater understanding of the underlying physics. For this, numerical simulations are a great pillar to provide insight and allow to understand the complex nature of those systems.

The generation of plasma waves with large amplitudes to accelerate electrons inside a plasma is not exclusive to laser pulses. An alternative approach is to use a particle bunch, and it was initially proposed by Chen [18] in 1985. For this, a relativistic electron beam is used to drive a wakefield and is known as PWFA. While the two schemes show differences, they are built around the same idea, that a driver can perturb the quasi-neutrality of a plasma which creates large accelerating fields $E_0 = cm_e\omega_p/e$ or

$$E_0\,[\mathrm{V/m}] \simeq 96\sqrt{n_0\,[\mathrm{cm}^{-3}]} \tag{1.1}$$

where $\omega_p = \left(4\pi n_0 e^2/m_e\right)^{1/2}$ is the electron plasma frequency, n_0 is the ambient electron number density, m_e and e are the electron rest mass and charge, respectively, and c is the speed of light in vacuum. The interesting aspect about a PBA is the dependence on the plasma density. For common scenarios a plasma with a density of $n_0 = 1 \times 10^{18}\,\mathrm{cm}^{-3}$ is used, leading to field strength of $E_0 \sim 96\,\mathrm{GV/m}$.

Over the years, several theoretical concepts supported by experiments and simulations have been examined and proposed. Some of which allow us to give insight into the complex and intertwined physics happening in PBA. The focus of this chapter lies in reviewing several concepts associated with LWFA and proton-driven PWFA. Rather than providing a deep dive and rigorous analytical calculations for the underlying physics, the aim is to provide a brief overview. In the first part of this chapter, the focus lies on LWFA. The aim is to give insights into the generation of wakes and propagation inside a plasma of a short laser pulse, in a way, that the work on the reduced model presented in chapter 2 will become evident. Also, the theoretical background for LWFA is extended to complement the discussion on the use of reduced modeling for particle trapping and acceleration. Another cor-

nerstone of the thesis presents full-scale modeling of the AWAKE experiment. For this, section 1.2 accompanies a review on proton-driven PWFA. All the work presented in the thesis utilizes the PIC algorithm and, more specifically, is connected to the implementation of a reduced solver in OSIRIS [34]. The PIC algorithm is discussed in section 1.3, and the general overview of extending OSIRIS is presented in section 1.4.

1.1 Laser-driven plasma wakes

The development of LWFA has created much excitement in recent decades. Much of it came from successful experiments and theoretical understanding. In laser-driven PBA, an intense short laser pulse pushes the light electrons away, leaving a positive channel of heavy ions. The high inertia of the ions results in a space-charge force that pulls the electrons back. However, as the electrons try to establish a quasi-neutrality, they overshoot and create a periodic wake structure behind the laser pulse. Commonly, an analogy to a boat moving over water is made. Inside the periodic wake structure, fields of several GV/m are reached.

In general, a laser pulse has to be strong enough to be able to push electron such that a clear separation is noticeable. Its field strength is commonly expressed in terms of the normalized vector potential $a_0 = eA/m_e c^2$ where $A = |\mathbf{A}|$ is the laser vector potential. When the normalized vector potential is above unity, the electron quiver motion in a fast-oscillating laser field becomes relativistic. Such laser strengths are commonly reached in typical LWFA scenarios where lasers intensities are above $\gtrsim 10^{18}\,\mathrm{W\,cm^{-2}}$ with a laser wavelength of $\sim$1 µm. More common, the normalized vector potential is expressed in an engineering formula

$$a_0^2 \simeq 7.3 \times 10^{-19} \left(\lambda_0\,[\mu\mathrm{m}]\right)^2 I_0[\mathrm{W/cm^2}] \tag{1.2}$$

and given the mentioned example, the normalized vector potential is $a_0 = 0.85$.

The response of the plasma to a laser driver is inherently nonlinear, and therefore challenging to understand. Nevertheless, several theoretical descriptions exist.

The force characterizing the displacement of the electrons, the ponderomotive force, is associated with the light pressure. In the limit of small laser amplitudes, when $a_0 \ll 1$ is fulfilled, the ponderomotive force of a linear polarized laser pulse is given as

$$\mathbf{F}_{\mathrm{p}} = -m_e c^2 \nabla(\hat{a}^2/2) \tag{1.3}$$

where $a^2 = a^2(\mathbf{r}, t)$ is the slow-varying laser envelope squared. Here, the force is associated with the envelope's gradient, and the negative sign indicates that electrons located along the propagation will be pushed towards the laser propagation. In the non-linear regime, the ponderomotive force is

$$\mathbf{F}_{\mathrm{p,\,nl}} = -m_e c^2 \nabla\gamma, \tag{1.4}$$

where γ corresponds to the relativistic Lorentz factor of an electron fluid element. While it implicitly depends on the envelope, the leading motion is the relativistic quiver motion. The effect of the light pressure onto the electrons is the strong modulation of their density.

Throughout the propagation of the laser pulse inside a plasma, the locally different plasma densities influence the propagation. Those effects refer to the non-linear optical behavior of a plasma [35]. Three different effects are important to be mentioned. These effects are

- modulation of the pulse length,
- modulation of the pulse width,
- and modulation of the laser frequency.

For pulse compression, the head and the tail of the laser pulse propagate at different group velocities due to different plasma densities in the longitudinal direction. More specifically, the head has a lower group velocity as the local plasma density is

higher than the density at the tail of the laser pulse. In the transverse direction, the effect of the plasma density gradient leads to self-focusing. The natural diffraction can counteract this effect. One further modulation can occur due to longitudinal variations of the phase velocity and leads to local frequency changes. The combination of all these processes, their evolution, and the laser propagation contribution are important to be considered. In particular, those effects can affect the local energy distribution of the laser pulse, modify the ponderomotive force of the laser, and hence change the resulting wake.

As the laser propagates inside the plasma, its strength expels a significant portion of the plasma electrons, while leaving the heavy and static ions behind. Due to the electron deficiency, large electric fields are generated, pulling electrons back, and giving them a transverse momentum. While the electrons try to reestablish the quasi-neutrality, they overshoot and continue oscillating. Several theoretical models have been developed to characterize the wakes in LWFA scenarios [36–41]. In the 1D linear limit, the plasma perturbation is determined by

$$\frac{\delta n}{n_0} = \frac{c^2}{\omega_p} \int_0^t dt' \sin\left[\omega_p \left(t - t'\right)\right] \frac{\nabla^2 \hat{a}^2(\mathbf{r}, t')}{2} \tag{1.5}$$

and the corresponding field is

$$\frac{\mathbf{E}}{E_0} = -c \int_0^t dt' \sin\left[\omega_p \left(t - t'\right)\right] \frac{\nabla \hat{a}^2(\mathbf{r}, t')}{2}. \tag{1.6}$$

From equations (1.5) and (1.6), it is noticeable that the density modulations and the corresponding fields exhibit a sinusoidal shape with a frequency corresponding to ω_p. Moreover, the modulation's main contribution arises from the ponderomotive force (1.3), seen by the gradient of the laser envelope in the integral. Extending this theoretical model to a non-linear regime was done only in 1D [39–41] and in the 3D non-linear regime, numerical calculations are usually required. The most significant change for the non-linear regime is the modification of the wake struc-

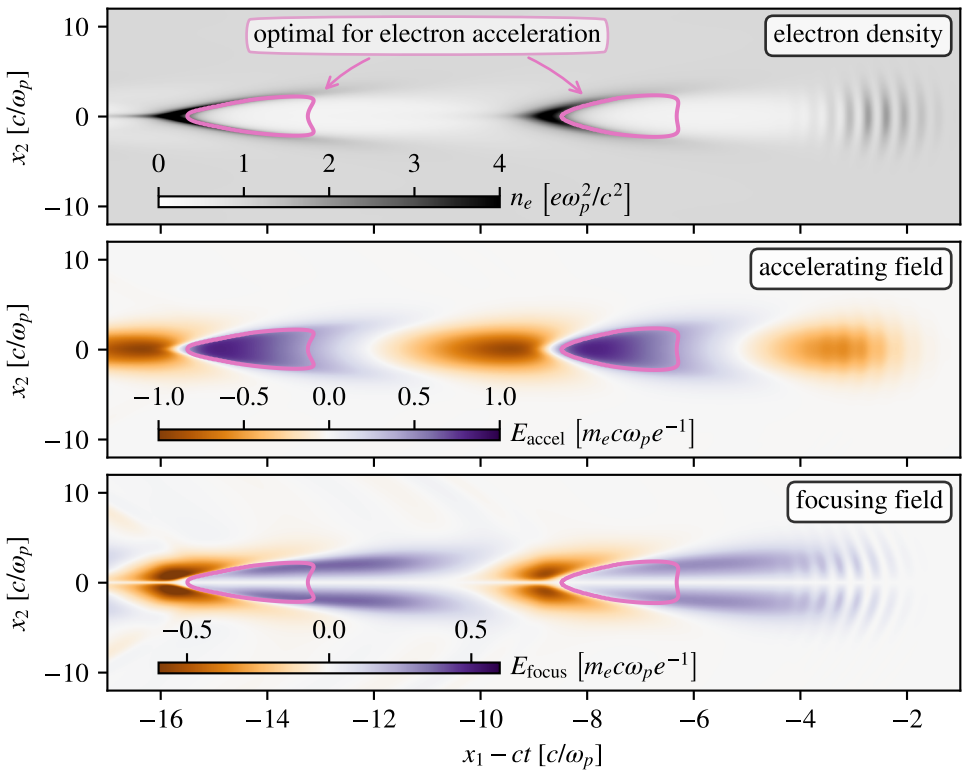


Figure 1.1: PIC simulation of a non-linear LWFA case. Shown are the areas where electron acceleration is favorable. It is favorable as an electron bunch would be accelerated and focused.

ture. The sinusoidal density modulation transforms into density spikes leading to a sawtooth-like electric field. Due to the absence of analytical models for the 2D and 3D non-linear cases, simulations are vital to model LWFA in non-linear regimes.

Moreover, higher dimensions are required to model transverse fields accurately and to recover full geometrical properties. In Figure 1.1, the plasma density, the accelerating, and the transverse fields are shown. Given an injected electron bunch, the generated plasma wakes allow focusing during the acceleration period. The challenging question becomes, how to inject an electron bunch for the acceleration.

While in general, particles can be injected externally and PBA can be used to in-

crease the energy, a more practical scenario is to combine it to a particle injector and particle accelerator. This approach allows us to build compact accelerator facilities with moderate particle energies. In 2002 Pukhov [42] showed through 3D simulations that a highly intense laser pulse with $a_0 = 10$ could generate a wake to produce mono-energetic electron beam without an external injection. This regime is known as a highly non-linear broken wave regime or blow-out regime. Using a laser pulse to self-trap electrons is experimentally challenging, as it requires precise control of the laser and plasma parameters. Hence, alternative injection approaches were suggested, such as ionization injection [43], injection through utilizing external magnetic fields [44], or by evolving plasma bubble [45]. In the case of ionization injection, a gas mixture of two gases is used. One of these two gases (with a lower ionization threshold) is used to generate a background plasma, while the other gas can be used to ensure ionization occurring at the higher laser field values. It allows for high fidelity control of particle trapping. Another approach is to use external magnetic fields. As the magnetic field is present, the wake structure is shortened due to the more forceful bending of the electrons towards the center as they slip behind the laser pulse. When the magnetic field is turned off, plasma electrons are trapped due to an expanding plasma bubble. A similar physical principle is achieved during a density down-ramp by making the plasma bubble expand. As the laser pulse propagates through a density transition, the wake structure expands from a higher density to a lower density, allowing for trapping of electrons.

There are important physical limits to laser-driven PBA. Those limits are associated with the laser propagation inside the plasma. One such limit is the depletion length. The depletion length is characterized by the laser's energy loss due to the production of the plasma waves required to accelerate particles. Another limitation of LWFA is described as the dephasing length. The phase velocity of the wake causes it, as it is less than the speed of light. As the longitudinal momentum of the electrons increases, their velocity can be higher than the wake's phase velocity. This momentum gain leads to electrons ending up in the decelerating region.

1.2 Proton-driven plasma waves

While the limitations like depletion and dephasing present in LWFA can be overcome through staging [46], further progress in terms of higher charge coupling efficiency has still to be shown. An alternative approach to LWFA is to use another way to drive the wakes. Such an alternative driver is a bunch of charged particles, and it is known as plasma wakefield acceleration (PWFA). Chen initially proposed it in 1985 and doubling of the 42 GeV Stanford Linear Accelerator Center (SLAC) beam was shown in the meter-scale was shown by Blumenfeld et al. [47] in 2007. PWFA would allow us to overcome the limitations of LWFA. However, the generation of a particle bunch requires large facilities to accelerate the bunch to moderate energies and contradicts the idea of building compact accelerators. In PWFA, the transformer limit determines similar to depletion in LWFA, the reachable energy of a witness bunch [48]. It arises because an electric field of the same order drags the driver as the one that accelerates a witness beam. Though, the energy achievable with conventional accelerators are orders of magnitude higher compared to a laser pulse. More specifically, the 42 GeV electron bunch of SLAC carries a total energy of 0.1 kJ and the 400 GeV proton bunch of Super Proton Synchrotron (SPS) at CERN carries a total energy of 20 kJ. The SPS bunch is the driver in the AWAKE experiment, and it was used to accelerate an electron bunch to 2 GeV recently [49]. In the following, as part of later discussions, the focus lies on proton drivers, but certain aspects apply to electron bunches as well.

While PWFA and LWFA share common physical aspects, such as perturbation of the quasi-equilibrium and the generation of the wake, the underlying physics is significantly different. In PWFA, rather than the ponderomotive force causing the electron displacement, the space charge force is felt by the plasma electrons. For an azimuthally symmetric bunch with density

$$n_b(\xi = ct - x_1, r) = n_{b,0}\, n_{b,\parallel}(\xi)\, n_{b,\perp}(r), \tag{1.7}$$

an analytical description in the linear regime exists [48]. For the bunch density in (1.7), the coordinate ξ is the position along the bunch which is moving with the velocity $v_b \simeq c$, $n_{b,0}$ is the charge density, $n_{b,\parallel}$ is the longitudinal and $n_{b,\perp}$ is the transverse bunch distribution. The longitudinal plasma wakefield $W_\parallel$ is given as

$$W_\parallel(\xi, r) = \frac{n_{b,0}\, e}{\varepsilon_0} \int_{-\infty}^{\xi} n_{b,\parallel}\left(\xi'\right) R(r) \cos\left(k_p \left(\xi - \xi'\right)\right) d\xi' \tag{1.8}$$

and the radial wake field $W_\perp$ is

$$W_\perp(\xi, r) = \frac{-n_{b,0}\, e}{\varepsilon_0 k_p} \int_{-\infty}^{\xi} n_{b,\parallel}\left(\xi'\right) \frac{dR(r)}{dr} \sin\left(k_p \left(\xi - \xi'\right)\right) d\xi' \tag{1.9}$$

where the $R(r)$ determines the radial dependency of the wake fields and is

$$\begin{aligned} R(r) = {} & k_p^2\, K_0\left(k_p r\right) \int_0^r r' n_{b,\perp} I_0\left(k_p r'\right) dr' \\ & + k_p^2\, I_0\left(k_p r\right) \int_r^\infty r' n_{b,\perp} K_0\left(k_p r'\right) dr' \end{aligned} \tag{1.10}$$

with I_0 and K_0 being the modified Bessel functions of the first and second kind. Comparing this equation to (1.6), a convolution of the driver density is noticeable. Like for the LWFA case, due to the causality condition, no fields exist ahead of the bunch. The term $R(r)$ has its maxima at $r = 0$ due to symmetry, which leads to vanishing radial wakefields at zero on-axis. On the other hand, assuming a Gaussian distribution in the longitudinal and transverse direction, the most effective wakefield excitation by a drive bunch with a bunch length σ_z and radial with σ_r the condition

$$k_p \sigma_z = \sqrt{2} \tag{1.11}$$

for the longitudinal distribution and $k_p \sigma_r = 1$ for the transverse distribution of the bunch have to be met [50].

Using a proton bunch to drive plasma wakes effectively becomes quite challenging as they usually tend to be long. More specifically, the SPS bunch used in the

AWAKE experiment has a length of 12 cm and requires a plasma with a density of $4 \times 10^9\,\mathrm{cm}^{-3}$ to drive plasma wakes effectively. Using a plasma with such low densities leads to acceleration gradients in the order of a few MV/m which is not sufficient to replace any conventional accelerators. A better approach would be to split the long proton bunch into beamlets to drive plasma waves periodically. This split into beamlets is achieved by seeding an instability, the self-modulation instability (SMI). The SMI arises from the transverse wakefields created by the front of the bunch, and it modulates the trailing proton bunch by pinching it into microbunches. Each beamlet is separated by the plasma wavelength, allowing to drive the wakefields resonantly. In the linear regime, the growth rate of the SMI [51] is given by

$$\Gamma = \frac{3\sqrt{3}}{4}\omega_p \left(\frac{n_{b,0} m_e}{2 n_e m_p \gamma_b} \frac{\xi}{ct}\right)^{1/3} \tag{1.12}$$

where m_p is the proton mass, n_e is the ambient plasma density, and γ_b the Lorentz factor of the proton beam.

1.3 Particle-in-cell algorithm

Over the past decades, the PIC algorithm has been established as a viable tool to model the behavior of plasmas under extreme scenarios. Moreover, it refers to a method to solve the fundamental kinetic equations for plasmas. Thanks to the pioneering work of Buneman, Dawson, Birdsall, Langdon, and others, PIC simulations allow us to provide real insights in astrophysical scenarios [52–55] and PBA accelerators [56–58].

The main governing idea of PIC is to reduce the computational needs required to advance plasma particles. Instead of calculating particle-particle interactions, it solves the equations for the particle motion based on surrounding fields. After advancing the particles, the associated currents with the particle motion are calculated. Then, the fields are calculated with the updated currents on a grid. Af-

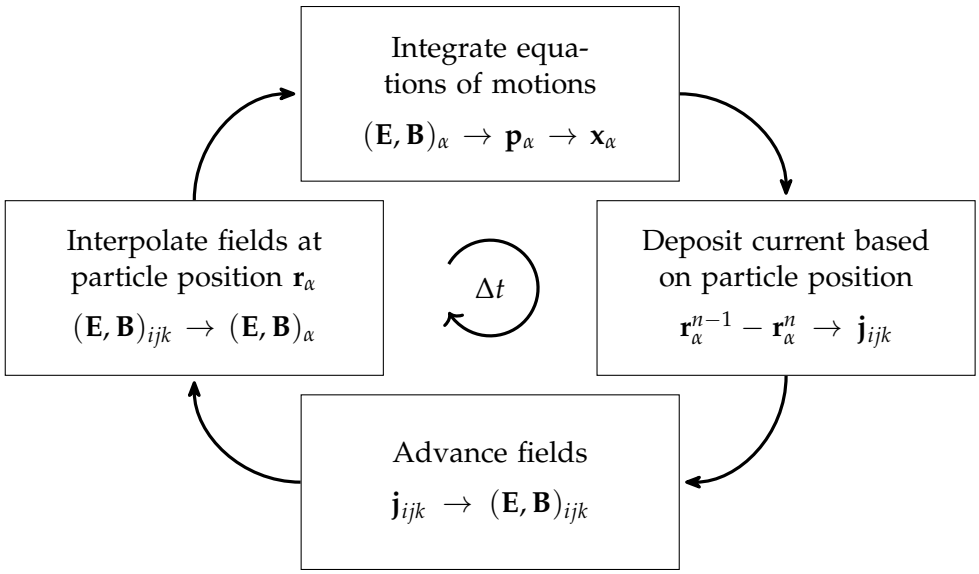


Figure 1.2: Simplified illustration of the PIC loop.

terward, the updated fields are then used for advancing the particles. These steps continue in a loop until a particular simulation time is reached, as shown in Figure 1.2.

The governing equations for the electric **E** and magnetic fields **B**, are Maxwell's equations (in cgs-units)

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 4\pi\rho \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{B} &= \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{J} \end{aligned} \tag{1.13}$$

where ρ and **J** are the charge and the current density distribution. The particles are advanced through a relativistic Boris-pusher [59] for the Lorentz-force equation:

$$\frac{d\mathbf{p}}{dt} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \tag{1.14}$$

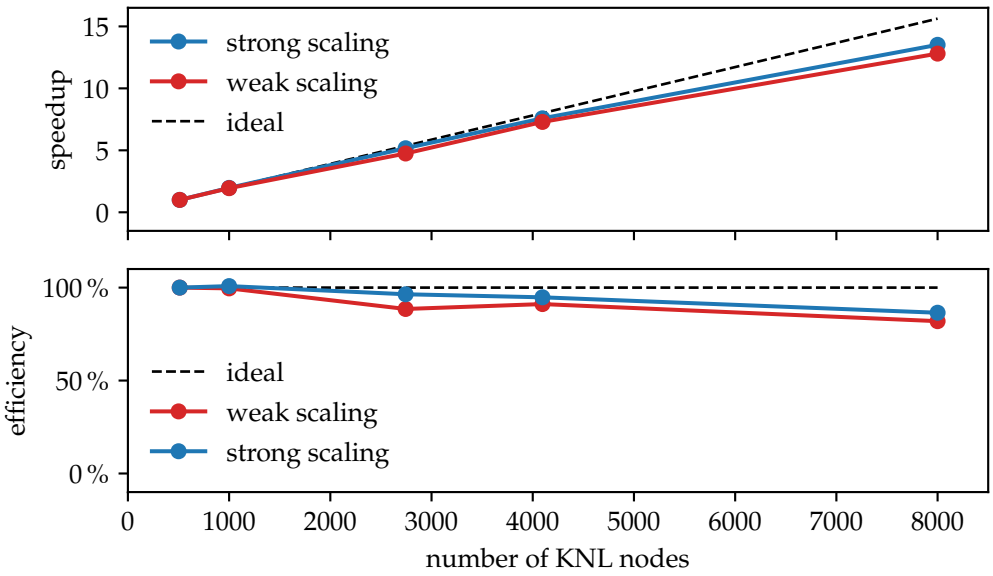


Figure 1.3: Weak and strong scaling test of OSIRIS on NERSC's Cori system using a hybrid parallelization scheme with OpenMP and MPI. The tests were done on the Intel Knights Landing nodes using AVX-512 vector pipelines. Each node has 68 cores, with each core having two vector processing units (VPUs).

where q is the charge of the particle and v is the velocity of the particle. In equation (1.13), (1.14), and in the following all quantities will be normalized. Spatial coordinates $\mathbf{r} = (x, y, z)$ are normalized to c/ω_p where c is the speed of light and $\omega_p = (\hat{n}_0 e^2/m_e)^{1/2}$ is the plasma frequency to the background plasma density $\hat{n}_0$ with e, m_e being the electron charge and mass, respectively. Time t is normalized to ω_p^{-1}, frequencies ω to ω_p, wavenumbers $\mathbf{k}$ to $k_p = \omega_p/c$, and velocities $\mathbf{v} = (v_x, v_y, v_z)$ to c. Moreover, momenta $\mathbf{p}$ are normalized to $m_e c$. Electrical fields $\mathbf{E}$ are normalized to $m_e c\omega_p/e$, while magnetic fields $\mathbf{B}$ to $m_e\omega_p/e$, and the vector potential $\mathbf{A}$ (or $\mathbf{a}$), to $m_e c^2/e$. Densities n are normalized to $\hat{n}$, charges q to e, and masses m to m_e.

While the equations (1.13) and (1.14) allow to model plasmas with large number

of particles, e.g. 10^{18} particles per cubic centimeter for LWFA, it is still computationally expensive. For this, PIC codes have to be developed to utilize modern HPC systems. In this thesis, the three-dimensional, relativistic, massively parallel, object-oriented PIC code OSIRIS [34] was used. Figure 1.3 shows the scaling of OSIRIS from a few computational nodes to almost the full NERSC Cori Machine. Even if it scales efficiently over many cores and modern hardware architectures, full-scale modeling is still computationally expensive. To overcome this, the focus of this thesis is the implementation of a reduced solver pondermotive guiding center (PGC). The advantage of OSIRIS 4.0, the latest version, is the increased flexibility for incorporating custom solvers into the PIC algorithm.

1.4 Implementing custom algorithms in OSIRIS

OSIRIS, with its latest version, is designed to utilize the object-oriented programming (OOP) paradigm. It allows developing classes to leverage composition, inheritance, and delegation. One possible approach is to incorporate classes dealing with communication, file reading, and writing, from classes used to solve the partial differential equations numerically. Inheritance can then be used to create custom classes to extend the functionality without adapting other parts of the code. In the following, a brief description of the OOP model in OSIRIS will be presented. While it was not developed as part of the thesis, it was extensively used in the implementation of the pondermotive guiding center (PGC) solver.

OSIRIS is constructed on top of the simulation object `t_simulation` which stores other objects, e.g. `t_node_conf` (required for handling communications), `t_emf` (object connected to electromagnetic fields), and `t_particles` (object storing particle information), as seen in Figure 1.4. Using the core object, custom solvers or "simulation modes" can be created by inheriting from the `t_simulation` object. Doing so allows us to customize the creation and the type of other objects. At the ini-

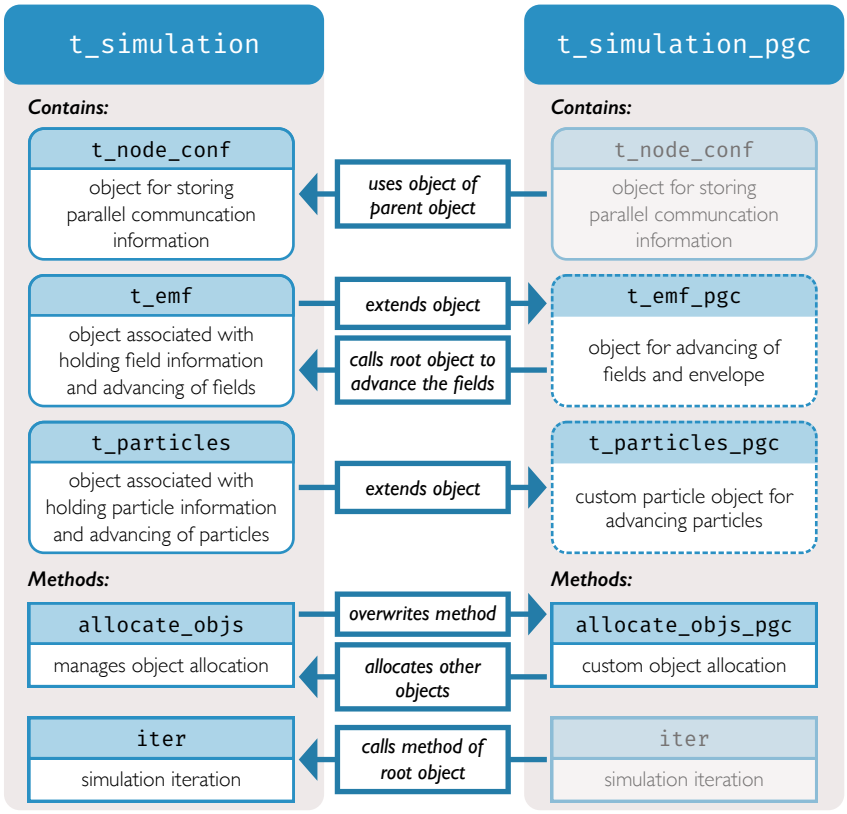


Figure 1.4: Object-oriented model for adapting custom algorithms in OSIRIS. The `t_simulation` class is the base class which which contains other important objects, e.g. `t_node_conf` for handling communications, `t_emf` for handling with electromagnetic fields, and `t_particles` for handling particles. By inheriting the base class, any class can overwrite specific simulation objects or methods and customize them to change the program flow. Transparent objects and methods represent the use of parent objects and methods.

tialization step of an OSIRIS simulation, the object method* `allocate_objs()` is responsible to allocate all required objects. In the case of PGC, a custom method `allocate_objs_pgc` is defined and provides custom objects for particles `t_emf_pgc` and electromagnetic fields `t_particles_pgc` which are extensions to `t_emf` and `t_particles`, respectively. Also, every inherited object can access methods and objects associated with the parent object. Therefore other object which do not require to be extended or specified can be directly allocated. Similarly, methods which are not specified but are available are going to be used directly, e.g. the method `iter()` of the simulation object. OSIRIS calls `iter()` to iterate the whole simulation by one time step which calls advance methods of individual objects for fields and particles. Any specification in those objects gives the flexibility of extending the object better.

As seen in this brief example, the OOP paradigm provides high flexibility to customize any OSIRIS simulation at runtime. It grants the possibility to perform a facet of different scenarios with no or minimal code adaptations. On the other hand, due to the cross-connection of the different objects, code fragmentation can increase, leading to increased maintainability.

*In modern Fortran this is known as type-bound procedures.

2 Reduced modeling on the plasma scale

I insist upon the view that "all is waves"

Erwin Schrödinger
Schrödinger: Life and Thought (1989)
Book by Walter Moore

Modeling the behavior of plasmas through computer simulation consist of vast numerical approaches. For some applications, simplified expressions may be derived for plasma under particular circumstances, e.g., cold fluid equations and magnetohydrodynamics. The LWFA is an example of a physical problem, more specific the non-linear regimes with $a_0 \geq 1$, for which no analytical expressions for a realistic, fully 3D problem exist. Kinetic modeling with the utilization of the PIC algorithm plays a critical role in these scenarios. Typically it involves self-consistent modeling of the electromagnetic fields, using Maxwell equations, and solving the particle motion under consideration of the ambient electromagnetic fields through the Lorentz force. PIC showed great success, especially in describing strong non-linear regimes, such as the highly non-linear broken-wave regime [42]. Unfortunately, it needs to resolve the shortest and the fastest scales. In LWFA, the shortest time scales are associated with the laser frequency, and the spatial scales are connected to the laser wavelength. These scales have to be resolved over the full propagation distances. It means that for a plasma accelerator with the density of

10^{17} cm^{-3} a laser pulse with a wavelength in the µm-range has to be propagated for 3.2 cm so that electrons can gain an energy of 1 GeV and 32 m to gain 1 TeV. Due to the disparity between the shortest scales associated with the laser wavelength and the longest scales associated with the acceleration distances makes modeling of plasma-based accelerators computationally challenging. Therefore, simulations are commonly performed in reduced geometries to overcome this disparity. While these simulations can be useful for carrying out parameter scans and providing insights into the underlying physics, many of the relevant physics is inherently three-dimensional.

Over the past years, numerous techniques and methods have been developed to overcome the disparity and reduce the computational cost of modeling LWFA scenarios. One example is the use of the moving window method [60], where the simulation box is co-propagating with the laser pulse, and only a fraction of the whole physical setup is considered. Following a similar approach led to the development of simulations in a Lorentz boosted frames [61–63]. By moving in a Lorentz boosted frame, where the laser pulse is stretched and the plasma is streaming towards the laser pulse, significant computational speedups can be obtained. However, those types of simulation are prone to numerical Cherenkov instability (NCI) which introduces artificial numerical noise and introducing unphysical scenarios. Despite the recent progress on understanding the NCI, there is still research left on the use of Lorentz boosted frames to model nonlinear regimes of laser acceleration [64–66]. Another approach is based on utilizing a cylindrical mode expansion of the electric and magnetic fields in a simulation [67]. The initial computational cost of a 2D setup is multiplied by including additional harmonic modes and leading to computational costs comparable to 2D simulations. Though, in cases of an optical-thin plasma where $\omega_0/\omega_p \gg 10$, those simulations are still challenging. An alternative method is to use a self-consistent description of the interaction between the particles and the laser field in terms of the slowly varying laser envelope and the associated ponderomotive force [68, 69]. This approach is known as ponder-

motive guiding center (PGC). For a PGC solver, the electric and the magnetic fields are governed by Maxwell's equations, and the Lorentz force is extended to include the ponderomotive force. As only the plasma fields are required to be resolved, a factor $(\omega_0/\omega_p)^2$ performance increase is obtainable.

In this chapter, the governing equations for the envelope and the Lorentz force extension are presented for the PGC algorithm. The main aim is to derive the finite difference equation for the laser envelope, in section 2.2, and for the extension of the Lorentz force, in section 2.5, which are implemented into OSIRIS. A discussion on the boundary conditions of the envelope solver is mentioned in section 2.3. An important aspect when performing numerical simulation is to ensure numerical stability. Moreover, in the case of PGC, numerical stability for Maxwell's equations and the envelope equation has to be guaranteed. In section 2.4, an analytical description of the numerical stability for the PGC algorithm in the presence of plasma is presented.

2.1 Ponderomotive Equations

The main idea behind the PGC algorithm is to introduce a self-consistent way of modeling the propagation of a laser pulse, focusing solely on its envelope and not requiring to solve for the rapid oscillations of a laser field. For this approximation to be valid, the condition $1 - v_{\parallel} \gg \omega_p/\omega_0$ for the normalized electron velocity $v_{\parallel}$ in the direction to the laser propagation has to be satisfied. Moreover, the condition requires that an electron passes through a laser field wavelength in time sufficiently short such that the radial motion is negligible. Given this and the Coulomb gauge $\nabla \cdot \tilde{\mathbf{A}} = 0$ for a general vector potential $\tilde{\mathbf{A}}$, the wave equation with fast and slow varying components can be expressed as

$$\left(\nabla^2 - \frac{\partial^2}{\partial t^2}\right)(\mathbf{A} + \mathbf{a}) = -\mathbf{J} - \mathbf{j} + \nabla\frac{\partial \Phi}{\partial t} \tag{2.1}$$

following the description by [69] where the normalized code units, as mentioned in section 1.3 are used. For (2.1), the capitalized symbols, such as current density **J**, a scalar potential Φ, and vector potential **A**, are associated with slow varying components of the fields and the lowercase symbols, such as the vector potential **a** and current density **j**, are connected to the fast varying laser field. While the slow-varying components are associated with the plasma response, the fast varying current density driven by the laser if the spot size was infinite can be defined as

$$\mathbf{j} = -\mathbf{a} \sum_i \frac{q_i \rho_i}{m_i} \tag{2.2}$$

where the sum is evaluated over particles with their charge q_i, charge density ρ_i, and the relativistic mass m_i. Therefore, we can express the evolution of our laser field to follow

$$\left(\nabla^2 - \frac{\partial}{\partial t^2}\right) \mathbf{a} = -\mathbf{j}. \tag{2.3}$$

Here, the laser field is still in its general form. To obtain the evolution of the envelope, we can assume a linear polarized laser pulse

$$\mathbf{a} = \frac{a(x,y,z,t)}{2} e^{i\omega_0(t-x)} \mathbf{e}_y + c.c. \tag{2.4}$$

with a polarization direction in y-direction. The laser pulse (2.4) corresponds to a monochromatic wave with the frequency ω_0 and a complex-valued envelope $a = a(x,y,z,t)$ which depends on the spatial and temporal coordinates. The evolution of the envelope, derived by Mora and Antonsen Jr [68], is governed by

$$\left(\nabla_\perp^2 - 2i\omega_0 \frac{\partial}{\partial \tau} + 2\frac{\partial}{\partial \tau \partial \xi}\right) a(\xi,y,z,\tau) = -\chi a(\xi,y,z,\tau)\,. \tag{2.5}$$

In (2.5), the envelope is now given in the light frame coordinates $(\tau, \xi) = (t, x-t)$ and includes the transverse Laplacian $\nabla_\perp^2 = \left(\partial_y^2 + \partial_z^2\right)$. The right-hand-side characterizes the plasma susceptibility and the plasma response to the propagation. It

is given as

$$\chi = -\sum_i \frac{q_i^2 n_i}{\langle m_i \rangle} \tag{2.6}$$

and determines the non-linear effects of the plasma onto the envelope with cycled averaged mass

$$\langle m \rangle = \sqrt{m_0^2 + \mathbf{p}^2 + \frac{(q\,|a|)^2}{2}} \tag{2.7}$$

where m_0 is the rest mass, $\mathbf{p}$ is the momentum of a charged particle induced by the wake, and the envelope a interpolated at the particle position. Expression (2.5) depends on a mixed temporal and spatial derivatives. The terms with $1/\omega_0^2$ can be dropped, reducing (2.5) to

$$\frac{\partial a}{\partial \tau} = \frac{1}{2i\omega_0}\left(1 + \frac{\partial_\xi}{i\omega_0}\right)\left(\chi a + \nabla_\perp^2 a\right). \tag{2.8}$$

The dropped term is used to reduce the axial group velocity of waves propagating at an angle to the axis. However, if Raman sidescatter is not essential and only forward Raman scattering is considered, this term can be dropped [69]. Also, the exclusion of mixed derivatives makes equation (2.8) of the form of a heat equation. Surprisingly, we moved from a wave description of light to a way of characterizing light using the principles of heat transport.

While the evolution of the envelope is influenced by the ambient plasma through the plasma suscebility χ, the influence of the envelope onto the motion of the particles is done through an extension of the Lorentz force [68]

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = q\left(\mathbf{E} + \frac{\mathbf{p}}{\gamma} \times \mathbf{B} - \frac{1}{4}\frac{q}{\langle m \rangle}\nabla |a|^2\right) \tag{2.9}$$

with the Lorentz factor γ. The field quantities $\mathbf{E}$ and $\mathbf{B}$ correspond to the slow-varying electric and magnetic fields, respectively.

With this set of equations, we can see that a PIC code has to solve the envelope

equation separately while the Lorentz force has to be extended by an additional term. In the next sections we focus mainly on the envelope equation and derive a finite difference scheme which can be incorporated next to a general Maxwell solver.

2.2 The Envelope Equation for the PGC algorithm

To include the envelope equation in three dimensions into OSIRIS, we require a set of algebraic equations. A finite difference method is used to transform the differential equation into a system of algebraic equations. This section is split into three subsections, containing a detailed derivation of the finite difference equation for each spatial dimension. The 2D cartesian, 2D cylindrical, and 3D cartesian version of the envelope equation are currently incorporated into OSIRIS. Other groups incorporated the PGC algorithm in 1D and 2D cartesian geometries [69] and in 2D cylindrical geometries [70]. Alternatively, the envelope equation (2.5) can be solved in the lab frame [71] rather than in a co-moving frame, but requires additional treatment of large phase errors or spurious dissipation [72]. While [73] included a 3D cartesian version of PGC in the light frame coordinates, their implementation scales quadratically with the transverse simulation box size and limits cases where the simulation box is large in the transverse direction.

2.2.1 Envelope Equation in 1D

As a limiting case, the one-dimensional version of the envelope equation allows us to characterize the plasma's influence on the evolution of the envelope. The equation for the envelope in this scenario is

$$\frac{\partial a}{\partial \tau} = \frac{1}{2i\omega_0}\left(1 + \frac{\partial_\xi}{i\omega_0}\right)\chi a \tag{2.10}$$

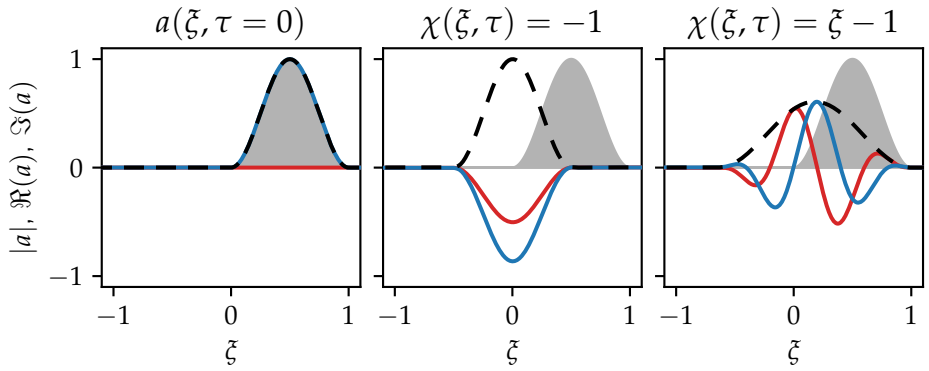


Figure 2.1: One-dimensional propagation of an envelope for the case of a constant χ and with a constant longitudinal gradient for χ. For the evolution, equation (2.13) is used and the initial value of the envelope follows $a(\xi, \tau = 0) = \sin^2(\xi\pi)$, shown in the left panel. Two scenarios are presented, a case with a constant plasma susceptibility (middle panel) and with a constant longitudinal gradient (right panel). The dashed black line represents the absolute value $|a|$, the blue line represents the real part $\Re(a)$, and the red line represents the imaginary part $\Im(a)$ of the envelope.

where the transverse Laplacian $\nabla_\perp^2$ in equation (2.8) has been dropped. In the case of a vacuum propagation, when $\chi = 0$ and therefore $\partial_\xi \chi = 0$, the envelope in the light frame coordinates would not evolve. It highlights that higher dimensions have to be included to have a correct physical representation of an envelope propagation.

To convert (2.10) into a set of equations, a central finite difference in time can be used. The usage of a central difference is favorable as its second order in time. In a finite-difference notation, we have

$$\left[\frac{\partial a}{\partial \tau}\right]_i^n = \frac{\mathcal{D}^\tau a_i^n}{2\Delta\tau} = \frac{a_i^{n+1} - a_i^{n-1}}{2\Delta\tau} \tag{2.11}$$

where the square brackets show the evaluation of a differentiation on time and space grid with i representing the spatial coordinate $\xi_i = i\Delta\xi$, n representing the

temporal coordinate $\tau_n = n\Delta\tau$, and the grid spacing for space $\Delta\xi$ and time $\Delta\tau$. The finite-difference operator $\mathcal{D}$ characterizes the difference on the grid. The superscript represents the finite difference in time, and the subscript indicates the difference in space. For longitudinal derivatives, a central difference scheme

$$\left[\frac{\partial a}{\partial \xi}\right]_i^n = \frac{\mathcal{D}_\xi a_i^n}{2\Delta\xi} = \frac{a_{i+1}^n - a_{i-1}^n}{2\Delta\xi} \tag{2.12}$$

is applied as well. The full finite difference expression of (2.10) in an expression for the envelope at a time step $n+1$ is

$$a_i^{n+1} = a_i^{n-1} + \frac{\Delta\tau}{i\omega_0}\left[\chi_i^n a_i^n + \frac{1}{2i\omega_0\Delta\xi}\left(a_i^n\left(\chi_{i+1}^n - \chi_{i-1}^n\right) + \chi_i^n\left(a_{i+1}^n - a_{i-1}^n\right)\right)\right]. \tag{2.13}$$

As seen in Fig. 2.1 or by inspecting (2.13), the effect of the plasma susceptibility is an introduction of phase rotation, and its gradient introduces a spatially different phase rotation. The slip back of the envelope arises from having the longitudinal gradient of the envelope and χ being negative while the gradient of the plasma susceptibility stretches or compresses the envelope.

2.2.2 Envelope Equation in 2D

For the 2d case, the diffusion in one transverse direction is added compared to the 1d case. The equation which characterizes the envelope evolution for cartesian geometries is

$$\frac{\partial a}{\partial \tau} = \frac{1}{2i\omega_0}\left(1 + \frac{\partial_\xi}{i\omega_0}\right)\left(\chi a + \frac{\partial^2}{\partial y^2}a\right) \tag{2.14}$$

where the second-order derivatives in y-direction transforms the envelope equation into a heat-type equation. In the case of cylindrical geometries, the transverse Laplacian is given as $\nabla_r^2 a = 1/r\partial_r a + \partial_r^2 a$ where r is the radial coordinate. As the governing equation is a heat-type equation, we require an implicit method for fi-

nite differentiation. Commonly, an implicit Crank-Nicolson (CN) method is used for this type of equations. Here, we consider an approach where only the second-order derivative in y-direction is solved implicitly. For this, time step $n+1$ and $n-1$ are used. The other terms are evaluated at time step n. It leads to the finite difference expression

$$\left(\alpha - \beta_y \mathcal{D}_{yy}\right) a_{ij}^{n+1} = \mathcal{S}_{ij} \tag{2.15}$$

with the right-hand side

$$\mathcal{S}_{ij} = \left(\alpha + \beta_y \mathcal{D}_{yy}\right) a_{ij}^{n-1} + \chi_{ij}^{n} a_{ij}^{n} + \gamma \mathcal{D}_{\tilde{\xi}} \chi_{ij}^{n} a_{ij}^{n} + 2\gamma\beta_y \mathcal{D}_{\tilde{\xi}} \mathcal{D}_{yy} a_{ij}^{n}, \tag{2.16}$$

$\alpha = i\omega_0/\Delta\tau$, $\gamma = 1/2i\omega_0$ and $\beta_y = 1/2\Delta y^2$. In above expression, the finite difference operator

$$\left[\partial_y^2 a\right]_{ijk}^{n} = \frac{\mathcal{D}_{yy} a_{ijk}^{n}}{\Delta y^2} = \frac{1}{\Delta y^2}\left(a_{i,j+1,k}^{n} - 2a_{i,j,k}^{n} + a_{i,j-1,k}^{n}\right) \tag{2.17}$$

for the second-order central difference is used. The right-hand side depends only on the normalized vector potential from previous time steps and can be calculated before calculating the envelope at the new time step $n+1$. Due to the finite difference operator in expression (2.15), a matrix inversion of a sparse tridiagonal matrix must be performed. Currently, three different methods are incorporated for the matrix inversion. The first method inverts the algebraic problem using the Thomas algorithm, and the second method utilizes QR-Factorization. Both methods are directly implemented in OSIRIS. They are provided to the PGC algorithm to be available at any point in time. Also, the third method uses specialized routines from LAPACK [74]. Using an external library like LAPACK is favorable as it is in general, optimized for numerical stability, and is widely available on computing platforms. All of the mentioned methods have a linear compute complexity $\mathcal{O}(N_y)$ with N_y being the number of cells in y-direction.

2.2.3 Envelope Equation in 3D

In the 3d case, the envelope can diffuse over the two transverse dimensions and can be compared to solving a heat equation in two dimensions. For solving heat equations in higher dimensions, the alternating-direction-implicit (ADI) method is frequently applied [75]. It is an example of an operator splitting method. ADI allows us to solve equation (2.8) numerically over two steps with the benefit that the diffusion operator for one transverse direction is evaluated at each step and allowing to keep a linear compute complexity. Hence, the two finite difference equations for the 3d case are

$$\left(2\alpha - \beta_y \mathcal{D}_{yy}\right) a_{ijk}^{n+1/2} = S_{ijk}^{\star} \tag{2.18a}$$

$$\left(2\alpha - \beta_z \mathcal{D}_{zz}\right) a_{ijk}^{n+1} = S_{ijk}^{\star\star} \tag{2.18b}$$

with the right-hand sides

$$\begin{aligned} S_{ijk}^{\star} = &\left(1 + \gamma \mathcal{D}_{\xi}\right) \chi_{ijk}^{n} a_{ijk}^{n} + \left(2\alpha + \beta_y \mathcal{D}_{yy}\right) a_{ijk}^{n-1/2} \\ &+ \left(2\beta_z \mathcal{D}_{zz} + 2\gamma \mathcal{D}_{\xi} \left(\beta_y \mathcal{D}_{yy} + \beta_z \mathcal{D}_{zz}\right)\right) a_{ijk}^{n} \end{aligned} \tag{2.19a}$$

$$\begin{aligned} S_{ijk}^{\star\star} = &\left(1 + \gamma \mathcal{D}_{\xi}\right) \chi_{ijk}^{n} a_{ijk}^{n+1/2} + \left(2\alpha + \beta_z \mathcal{D}_{zz}\right) a_{ijk}^{n} \\ &+ \left(2\beta_y \mathcal{D}_{yy} + 2\gamma \mathcal{D}_{\xi} \left(\beta_y \mathcal{D}_{yy} + \beta_z \mathcal{D}_{zz}\right)\right) a_{ijk}^{n+1/2} \end{aligned} \tag{2.19b}$$

where first the diffusion in the y-direction and then in the z-direction has been evaluated. For the right-hand side, we assumed that that the plasma susceptibility $\chi_{ijk}^{n} \simeq \chi_{ijk}^{n+1/2}$ does not change over a sequential half time step as it is quasi-static in the light frame coordinates. This assumption allows us to compute the envelope's evolution without advancing the particles over two consecutive half time steps. Each equation in (2.18) can be solved by inverting the matrix expression, similar to the 2D case.

2.3 Boundary Conditions for the Envelope Equation

Boundary conditions are an essential element in the discussion of differential equations. Their correct treatment is required, as otherwise, non-physical solutions can be obtained. As shown in the previous section, the finite difference equation set requires a matrix inversion, which depends on the transverse direction. Accordingly, we treat the boundary conditions for longitudinal and transverse separately. The finite-difference operator $\mathcal{D}_{\xi}$ appears only on the right-hand side of (2.18). Hence, a masking function onto the envelope after advancing and along the longitudinal boundaries can be applied. As a masking function, the first quarter of a $\sin^2(\xi)$-function over several boundary cells is used. It ensures a smooth transition and makes the envelope vanish at the boundaries. For the transverse direction, without further modification of the first and the last line of the matrices in (2.18a) and (2.18b) a Dirichlet boundary condition for the envelope is introduced. As in most scenarios, the envelope does not extend to the transverse boundaries; we considered this as a sufficient condition for the boundaries. In the case of periodic boundaries, the matrices' tri-diagonal structure is broken, and the additional side-diagonal entry appears at the upper-right and lower-left corner of the matrix. For this, we utilize the Sherman-Morrison formula [75], reducing it to an auxiliary problem.

2.4 Numerical stability of the envelope equation

The numerical stability of a finite difference scheme is necessary and sufficient for convergence. When solving Maxwell's equations, the Courant–Friedrichs–Lewy (CFL) condition has to be fulfilled to achieve stability. Here, we have to consider, in addition to the CFL condition, a stability equation for the envelope solver characterized by equations (2.18) and (2.19).

For obtaining the stability condition, a von-Neumann analysis can be applied,

which characterizes the growth of a numerical error. The derivation requires the governing finite difference scheme (2.18) and (2.19) to be rewritten in a matrix form

$$\mathbf{L}\,\mathbf{a}_{ijk}^{n+1/2} = \mathbf{R}\,\mathbf{a}_{ijk}^{n} \tag{2.20}$$

where the vector

$$\begin{aligned} \mathbf{a}_{ijk}^{n+1/2} &= \left(a_{ijk}^{n}, a_{ijk}^{n+1/2}, a_{ijk}^{n+1}\right)^{T} \\ \mathbf{a}_{ijk}^{n} &= \left(a_{ijk}^{n-1/2}, a_{ijk}^{n}, a_{ijk}^{n+1/2}\right)^{T} \end{aligned} \tag{2.21}$$

are characterizing the envelope values on a temporal and spatial grid. As seen, each component of the vectors is evaluated at different times steps and is shifted by a half time step. The matrix $\mathbf{L}$ is a diagonal matrix

$$\mathbf{L} = \mathrm{diag}(L_{11}, L_{22}, L_{33}) \tag{2.22}$$

with the entries

$$\begin{aligned} L_{11} &= 1 \\ L_{22} &= 2\alpha - \beta_y \mathcal{D}_{yy} \\ L_{33} &= 2\alpha - \beta_z \mathcal{D}_{zz}. \end{aligned} \tag{2.23}$$

while $\mathbf{R}$ is a sparse matrix

$$\mathbf{R} = \begin{pmatrix} 0 & R_{12} & 0 \\ R_{21} & R_{22} & 0 \\ 0 & R_{32} & R_{33} \end{pmatrix} \tag{2.24}$$

where each entry

$$\begin{aligned}
R_{12} &= 1 \\
R_{21} &= 2\alpha + \mathcal{D}_{yy}\beta_y \\
R_{22} &= \gamma\delta\chi + \chi\left(\gamma\mathcal{D}_{\xi} + 1\right) + 2\mathcal{D}_{zz}\beta_z + 2\gamma\mathcal{D}_{\xi}\left(\mathcal{D}_{yy}\beta_y + \mathcal{D}_{zz}\beta_z\right) \\
R_{32} &= 2\alpha + \mathcal{D}_{zz}\beta_z \\
R_{33} &= \gamma\delta\chi + \chi\left(\gamma\mathcal{D}_{\xi} + 1\right) + 2\mathcal{D}_{yy}\beta_y + 2\gamma\mathcal{D}_{\xi}\left(\mathcal{D}_{yy}\beta_y + \mathcal{D}_{zz}\beta_z\right)
\end{aligned} \tag{2.25}$$

can be obtained by comparing with the finite difference scheme. The finite difference operators $\mathcal{D}_{\xi}$, $\mathcal{D}_{yy}$ and $\mathcal{D}_{zz}$ are acting directly on the envelope values in (2.21). In above expressions, the plasma susceptibility χ and its longitudinal difference $\delta\chi = \mathcal{D}_{\xi}\chi = \chi_{i+1,j,k} - \chi_{i-1,j,k}$ are treated as constant parameters neglecting any spatial and temporal dependencies. The error for a finite difference scheme can be defined as

$$\epsilon^n_{ijk} = a^n_{ijk} - a(n\Delta\tau, i\Delta\xi, j\Delta y, k\Delta z) \tag{2.26}$$

representing the difference of the envelope evolution using finite difference, characterized by a^n_{ijk} and the analytical solution $a(n\Delta\tau i\Delta\xi, j\Delta y, k\Delta z)$ to (2.8). Given this, we can use the ansatz that the error can be decomposed into

$$\hat{\epsilon}^n = g^n \mathrm{e}^{ik_{\xi}\xi}\mathrm{e}^{ik_y y}\mathrm{e}^{ik_z z} \tag{2.27}$$

with the unknown error growth rate

$$g = \frac{\hat{\epsilon}^{n+1}}{\hat{\epsilon}^n} \tag{2.28}$$

and the different numerical modes characterized by k_{ξ}, k_y and k_z. Due to the usage of Fourier transformations, periodic or spatially bounded solutions are required. Both cases are present as, in general, the laser pulse is not extended to the boundaries unless a periodic case is studied. With the ansatz (2.27) we can evaluate the

finite difference operators as

$$\begin{aligned}\mathcal{D}_{\tilde{\xi}}\hat{\epsilon}^n &= 2i\sin\left(k_{\tilde{\xi}}\Delta\tilde{\xi}\right)\hat{\epsilon}^n = 2i\mathcal{S}_{\tilde{\xi}}\hat{\epsilon}^n \\ \mathcal{D}_{yy}\hat{\epsilon}^n &= 4\sin^2\left(\frac{k_y\Delta y}{2}\right)\hat{\epsilon}^n = 4\mathcal{S}_{yy}\hat{\epsilon}^n \\ \mathcal{D}_{zz}\hat{\epsilon}^n &= 4\sin^2\left(\frac{k_z\Delta z}{2}\right)\hat{\epsilon}^n = 4\mathcal{S}_{zz}\hat{\epsilon}^n\end{aligned} \tag{2.29}$$

with the properties $|\mathcal{S}_{\tilde{\xi}}| \leq 1$, $0 \leq \mathcal{S}_{yy} \leq 1$ and $0 \leq \mathcal{S}_{zz} \leq 1$. Together with (2.20) and (2.27), the error growth rate can be found by solving an eigenvalue problem

$$\sqrt{g}\hat{\boldsymbol{\epsilon}}^n = \hat{\mathbf{L}}^{-1}\hat{\mathbf{R}}\hat{\boldsymbol{\epsilon}}^n \tag{2.30}$$

with the error vector $\hat{\boldsymbol{\epsilon}}^n = (\hat{\epsilon}^{n-1/2}, \hat{\epsilon}^n, \hat{\epsilon}^{n+1/2})^T$ and the matrices $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$ being representations of the matrices $\mathbf{L}$ and $\mathbf{R}$ in the k-space. The eigenvalues to the eigenvalue problem (2.30) are

$$\begin{aligned}\lambda_1 &= \frac{R_{33}}{L_{33}} \\ \lambda_2 &= \frac{R_{22}}{2L_{22}} - \frac{\sqrt{L_{11}R_{22}^2 + 4L_{22}R_{12}R_{21}}}{2\sqrt{L_{11}}L_{22}} \\ \lambda_3 &= \frac{R_{22}}{2L_{22}} + \frac{\sqrt{L_{11}R_{22}^2 + 4L_{22}R_{12}R_{21}}}{2\sqrt{L_{11}}L_{22}}\end{aligned} \tag{2.31}$$

where the last two eigenvalues differ only by the sign of the last term. The eigenvalues are connected to the error growth rate through $\sqrt{g} = \lambda$. The given finite difference scheme is stable if the following condition

$$|g| = \left|\frac{\hat{\epsilon}^{n+1}}{\hat{\epsilon}^n}\right| \leq 1 \tag{2.32}$$

is fulfilled which corresponds to

$$\max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} \leq 1 \tag{2.33}$$

in the eigenvalue notation. In general, only the highest growing error mode will dominate, and therefore an upper limit given a set of parameters is required to be considered. The eigenvalues in (2.31) depend not only on the spatial and temporal resolutions but also on the numerical modes, the plasma susceptibility, its longitudinal gradient, and the laser frequency. In the following paragraphs, the discussion is separated into two parts. First, we focus on the stability of the finite difference scheme for propagation in a vacuum, and second, we discuss the stability of this scheme in the presence of a plasma. For both cases, square cells will be used. Assuming square cells allows studying of the numerical stability without loss of generality.

The case of vacuum and the absence of a plasma can be characterized by a vanishing plasma susceptibility χ. Using expressions (2.31), we can find the highest growing eigenvalue

$$\lambda_{\max} = -\frac{2\Delta\tau\,(\Delta\,\omega_0 + 2)}{\Delta\,\omega_0\,(\Delta\tau - i\Delta^2\omega_0)} \tag{2.34}$$

where $\mathcal{S}_\xi = \mathcal{S}_{yy} = \mathcal{S}_{zz} = 1$ for the eigenvalue λ_1 is used. With this, the condition for stability for vacuum is

$$\Delta\tau \leq \frac{\Delta^2\omega_0}{\sqrt{(1 + 4/\Delta\,\omega_0)\,(3 + 4/\Delta\,\omega_0)}} \tag{2.35}$$

with Δ being the spatial resolution. Note that it depends not only on the spatial and temporal resolution but also on the laser frequency ω_0, which is normalized to a characteristic frequency $\tilde{\omega}$. Furthermore, the upper limit of the stability condition increases with the grid size but also with the laser frequency, as seen in Fig. 2.2(a).

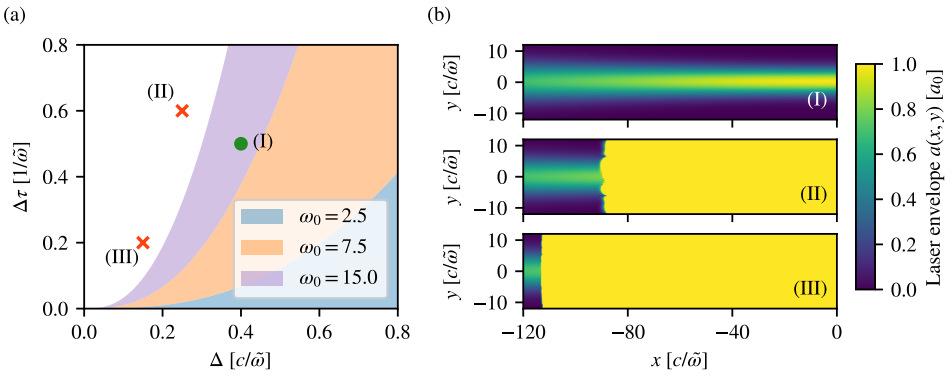


Figure 2.2: Figures are representing the stability condition for the envelope equation in a vacuum. Different stable regions for different temporal and longitudinal resolution are shown in (a). The regions differ based on the laser frequency ratio to a normalization constant $\tilde{\omega}$. The scatter points represent three simulations with different resolutions where red represents numerical unstable and green numerical stable simulations. The resolution is connected to the simulation results in (b), which shows the evolution of a laser pulse center towards the focal plane. The laser follows a Gaussian beam profile.

For $\omega_0 \gg 1/\Delta$, the condition reduces to

$$\Delta\tau < \frac{\Delta^2\omega_0}{\sqrt{3}}. \tag{2.36}$$

The above expression is an oversimplification of the stability condition, and in general, the full equation (2.35) should be used. The simplification is used merely for better interpretation of the stability condition. Examining the stability condition (2.36), it is noticeable that higher laser frequency loosens the temporal resolution restriction and scales linearly with the laser frequency. This result is not surprising, as seen in the governing equation for the envelope (2.8) the laser frequency reduces the contribution of the right-hand side. For verification of the stability condition (2.35), three simulations were performed with a laser pulse propagating towards the focal plane. The transverse component of the laser pulse follows the

description of a Gaussian beam

$$a(x,r) = \frac{a_0}{\sqrt{1+(x/x_R)^2}} \exp\left(i\phi\right) \exp\left(-r^2/w^2\right) \tag{2.37}$$

where

$$\begin{aligned} r &= \sqrt{y^2+z^2} \\ \phi &= \arctan(x/x_R) - (x/x_R)\, r^2/w^2 \\ w &= w_0\sqrt{1+(x/x_R)^2} \\ R &= x\left(1+(x_R/x)^2\right) \end{aligned} \tag{2.38}$$

are the common parameters for a Gaussian beam. In the study, the laser pulse has a pulse width of $w_0 = 4.0$, a laser frequency of $\omega_0 = 15.0$, and a Rayleigh length of $x_R = 120.0$ and only the spatial and temporal resolution is modified. In Fig. 2.2(b), the results of three different test cases are shown. If the stability condition is not satisfied, then the envelope erupts numerically after several iterations.

For the case when a plasma is present, a non-vanishing plasma susceptibility is required to be considered. Hence, all eigenvectors in (2.31) are considered. In Fig. 2.3, the highest error growth rate of all the possible growing modes is plotted against the laser frequency for a case without any longitudinal plasma gradients and including plasma gradients. If a plasma without a plasma gradient is present, then the error growth rate is bound by the stability criteria of the vacuum. On the contrary, including longitudinal plasma gradients, by the means of the longitudinal difference $\delta\chi$, introduces non-vanishing error growth rates. Particularly, using (2.31) and $\mathcal{S}_\xi = \mathcal{S}_{yy} = \mathcal{S}_{zz} = 0.0$, the error growth rate is given by

$$\hat{g} = 1 + \frac{\Delta\tau\delta\chi\sqrt{64\Delta^2\omega_0^4 + \Delta\tau^2\delta\chi^2}}{32\Delta^2\omega_0^4} \tag{2.39}$$

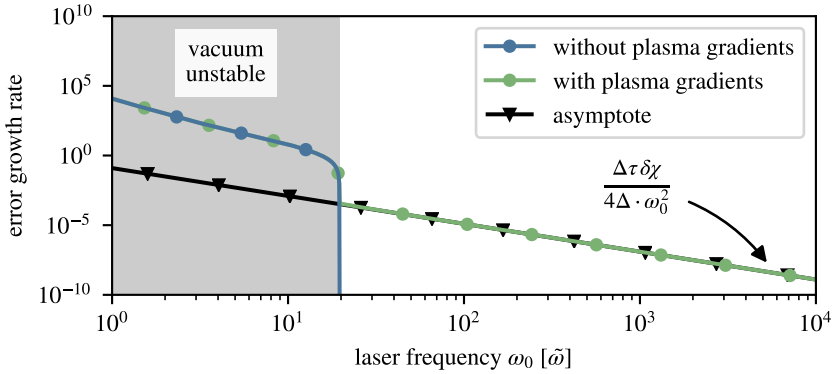


Figure 2.3: Double-log plot of the numerical error growth rate $|g| - 1$ as a function of the laser frequency ω_0. In this representation, a non-vanishing growth rate represent an unstable scenario. Shown here are cases with the time step of $\Delta\tau = 0.15$ and the spatial resolution of $\Delta\xi = \Delta y = \Delta z = 0.2$. In each different case, the proxy parameters for plasma susceptibility χ and its longitudinal gradient $\delta\chi$ were changed. For both cases $\chi = 1.0$ is used. The blue line represents the case $\delta\chi = 0.0$ and the green line represents $\delta\chi = 1.0$. The black line shows the asymptote for the stability if a gradient is present.

and for $\omega_0 \gg 1$, the error growth rate reduces to

$$\hat{g} = 1 + \frac{\Delta\tau}{4\Delta} \cdot \frac{\delta\chi}{\omega_0^2} \tag{2.40}$$

which does never fulfill the stability criteria (2.32) unless longitudinal plasma gradients $\delta\chi$ vanish. A longitudinal gradient is generally given in our simulations, and a non-vanishing envelope will push electrons inducing gradients in the plasma susceptibility. This fact makes the presented finite difference scheme for the PGC solver unconditionally unstable. However, as seen in Fig. 2.3, the error growth rate is close to unity implying small growth rates. Assuming the asymptotic limit, we

can find the incubation step

$$\mathcal{N} = \frac{\log f}{\log\left(1 + \frac{\Delta\tau}{4\Delta} \cdot \frac{\delta\chi}{\omega_0^2}\right)}. \tag{2.41}$$

after which an initial error is amplified by a amplification factor $f = |\epsilon^{\mathcal{N}}|/|\epsilon^0|$. Using above equation and considering a case where $\Delta = 0.1$, $\Delta\tau = 0.05$, $\delta\chi = 1.0$, $\omega_0 = 30$ and $f = 10^8$, we obtain an incubation step of $\mathcal{N} \sim 130\,000$. After this number of iteration, an error in the double-precision range has been amplified to a single-precision range. A further doubling of the incubation step implies an error growth to the order of unity from the initial range of double precision. Converting this to a real scenario, then this corresponds to a laser pulse with a wavelength of 1 µm propagating through a plasma with an electron density of $n_e = 1.24 \times 10^{18}\,\mathrm{cm}^{-3}$ for 6.32 cm such that the numerical error in double precision range reached unity. Here, we have to point out that a high gradient for the plasma susceptibility $\delta\chi = 1.0$ has been assumed. These high gradients are expected to be present at the vacuum-plasma interface, and more common throughout the propagation, gradients with at least an order of magnitude lower are expected. As the error growth scales with $\sim 1/\omega_0^2$, PGC is preferable for simulations where the ratio between the laser frequency and the plasma frequency is extremely high. These particular cases would not just lead to higher computational speedups but also stabilize the envelope propagation. Also, artificial numerical noise in the plasma susceptibility can be reduced by smoothing the plasma susceptibility and inherently reducing the numerical error growth.

2.5 Particle advancing for the ponderomotive guiding center solver

After having an equation governing the dynamics of the envelope as it propagates through the plasma, we now focus on the equation governing the dynamics of the plasma in response to the envelope. This inclusion is required to model the plasma response correctly, and as the envelope evolution is affected by it through the plasma susceptibility. The envelope can affect the particles through its ponderomotive force [68]. It can be included through an extension to the Lorentz force

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = q\left(\mathbf{E} + \frac{\mathbf{p}}{\gamma} \times \mathbf{B} - \frac{1}{4}\frac{q}{\langle m \rangle}\nabla|a|^2\right) \tag{2.42}$$

as discussed in section 2.1. At a given time n, the grid quantities $\mathbf{E}^n$, $\mathbf{B}^n$ and the ponderomotive force $\mathbf{F}^n = \nabla|a^n|^2$ are known. To evaluate the time derivative in (2.42), an implicit equation for the mass is required to be solved:

$$\langle m \rangle^n = \langle m \rangle^{n-1/2} + q\left(\mathbf{E}^n - \frac{q}{4\langle m \rangle^n}\mathbf{F}^n\right)\frac{\mathbf{p}^{n-1/2}}{\langle m \rangle^{n-1/2}}\frac{\Delta t}{2}. \tag{2.43}$$

This can be done with a simple quadratic formula. Then, the momentum can be updated using a Boris push [59]. The effect of the ponderomotive force can be seen as a correction to the electric field and leads to

$$\mathbf{p}^{n+1/2} = \mathcal{R}\left(\mathbf{p}^{n-1/2} + \frac{\mathbf{I}}{2}\right) + \frac{\mathbf{I}}{2} \tag{2.44}$$

with

$$\mathbf{I} = \left(q\mathbf{E}^n - \frac{1}{4}\frac{q^2}{\langle m \rangle^n}\mathbf{F}^n\right)\Delta t \tag{2.45}$$

where the operator $\mathcal{R}$ rotates the momentum vector around the magnetic field by an angle

$$\Theta = -\frac{qB^n}{\langle m \rangle^n} \Delta t. \tag{2.46}$$

In the absence of an envelope $a^n \to 0$ and therefore an absence of the ponderomotive force $\mathbf{F}^n \to 0$ the unmodified Boris push is recovered.

To update the particle position, the knowledge of $\langle m \rangle^{n+1/2}$ is required. We can use

$$\left(\langle m \rangle^{n+1/2}\right)^2 = m_0^2 + \left(\mathbf{p}^{n+1/2}\right)^2 + \frac{q^2 a^2}{2} \tag{2.47}$$

together with

$$a^2 = \left(a^{n+1/2}\right)^2 + \nabla \left(a^{n+1/2}\right)^2 \cdot \frac{\mathbf{p}^{n+1/2}}{\langle m \rangle^{n+1/2}} \frac{\Delta t}{2} \tag{2.48}$$

to solve for $\langle m \rangle^{n+1/2}$ using a quadratic formula. Afterwards, we can advance the particle position via

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \frac{\mathbf{p}^{n+1/2}}{\langle m \rangle^{n+1/2}} \Delta t \tag{2.49}$$

where $\mathbf{x} = (x, y, z)$ is the particle position.

2.6 Physical benchmark of the implementation

For testing the implementation of the PGC module, two examples can be examined. Throughout the thesis, the focus will be on the 3D implementation, as this is the case with the highest computational gain, and which recovers all geometrical aspects of laser-driven PBA. In the first example, a vacuum propagation of a laser pulse is examined to ensure that the implementation of the envelope propagation is correct and follows a Gaussian beam description. As a second example, the wake excitation by a laser pulse compared to full PIC* is presented.

*Throughout the thesis, full PIC refers to a simulation without the PGC module.

2.6.1 Vacuum propagation

A propagating laser pulse diffracts while moving away from the focal spot due to its finite width. On the contrary, a laser pulse focuses while moving towards the focal spot. In a vacuum, it is possible to obtain an analytical expression for the laser pulse by solving the electromagnetic wave equation in the paraxial approximation. Therefore it is an ideal physical benchmark for the implementation of the envelope solver.

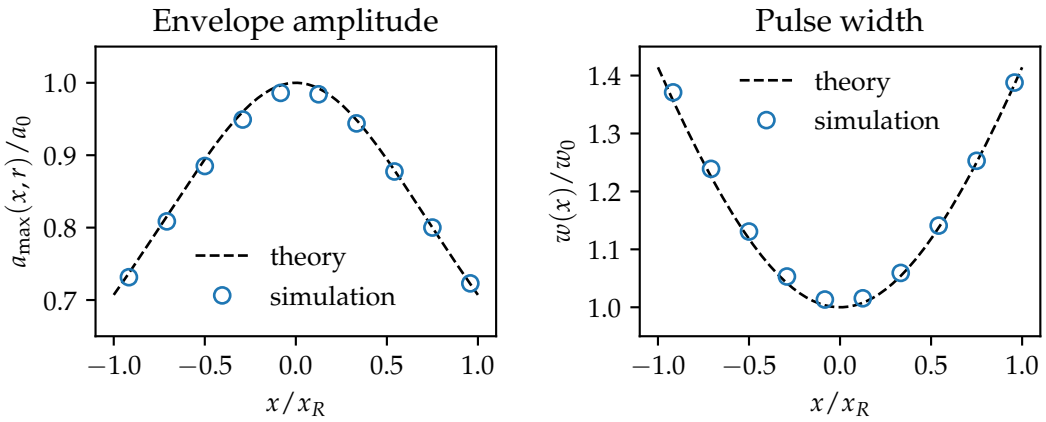


Figure 2.4: Comparison of the evolution of the peak envelope (left) and the pulse width (right) for a laser pulse propagating in a vacuum using the PGC solver. In both cases, the scatter points represent the obtained values from a simulation, and the dashed line represents the theoretically expected value.

Given a laser pulse, the envelope solver should recover the propagation properties of a Gaussian beam. The traversal component of the laser pulse follows the Gaussian beam profile, and is described by

$$a(x,r) = \frac{a_0}{\sqrt{1+(x/x_R)^2}} \exp\left(i\phi\right) \exp\left(-r^2/w^2\right) \tag{2.50}$$

with each beam parameter being defined by (2.38).

As a test case, a laser pulse with pulse width of $w_0 = 4.0$, a laser frequency $\omega_0 = 15.0$, and $a_0 = 1.0$ is considered. In addition, a pulse length of $\tau_L = 4.0$ is used. Figure 2.4 shows a comparison for the peak amplitude and pulse width for different propagation distances against the expected theoretical value. As the beam waist shrinks during the propagation towards the focal plane, the amplitude has to increase. Afterward, as the laser pulse diffracts and the beam waist increases, and due to energy conservation, the field amplitude has to decrease. As seen, PGC models the laser envelope propagation properly without characterizing the laser pulse's fast oscillating field.

2.6.2 Wakefield excitation

An important aspect is to ensure that the PGC implementation models the wake excitation in a plasma correctly. Analytical models for the non-linear behavior of a plasma in a non-linear 3D regime have not been found yet. The best results for simulating this non-linear behavior are obtained by utilizing PIC simulations. Furthermore, PIC simulations are commonly used to support experimental findings. Hence, the best approach to compare wake generation in plasmas by PGC is a comparison with full PIC simulation for the same physical parameters and only change numerical parameters like the grid resolution.

For the second physical benchmark, a laser with a frequency of $\omega_0 = 30.0$, a beam waist of $w_0 = 4.0$, and a normalized vector potential of $a_0 = 1.0$ is used. The longitudinal laser pulse profile follows a 5^{th} order polynomial [76] with a pulse duration of $\tau_{\text{FWHM}} = 3.0$, while the plasma profile follows a uniform distribution. These physical parameters are used both for the full PIC and for the PGC simulation, and only a short propagation inside the plasma is modeled. Later in chapter 4 longer propagation studies are performed.

While the transverse resolution is equal for both cases with $\Delta y = \Delta z = 0.075$, the difference between both simulations is in the resolution in longitudinal direc-

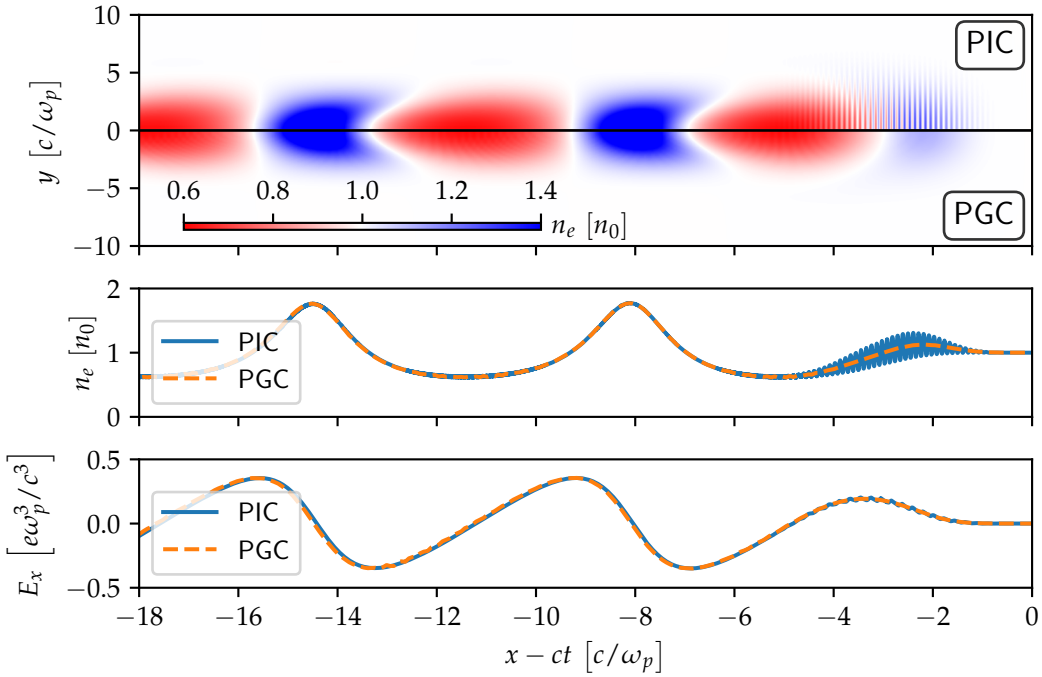


Figure 2.5: Comparison of wakefield excitation in 3D for full PIC and PGC. In the top plot, a slice electron density for the full PIC and PGC is shown. The middle plot represents a lineout for both cases along the laser propagation axis, while the bottom plot shows the resulting longitudinal field for both cases.

tion and time. For the case of PIC, the longitudinal resolution is $\Delta\xi = 2 \times 10^{-3}$ and $\Delta\tau = 2.497 \times 10^{-3}$, and for PGC it is $\Delta\xi = 7.5 \times 10^{-2}$ and $\Delta\tau = 3.75 \times 10^{-2}$, respectively. With two orders of magnitude reduced resolution, the example presented in Figure 2.5 reduced the computation cost by 9.74×10^1 in terms of CPUh.

As seen in figure 2.5, PGC can recover the wakefield excited by an intense laser pulse both in longitudinal and in the transverse direction. Furthermore, in this example, a non-linear regime is present with $a_0 = 1.0$, which lets the plasma wakes to deviate from the sinusoidal shape expected in the linear regime. With the envelope description of the PGC solver, the rapid oscillations of the laser field are averaged out. This is best seen in middle plot of the figure 2.5.

2.7 Conclusion

This chapter's focus is to provide the fundamental steps for implementing the PGC algorithm, and those are required to enable PGC in OSIRIS. As only the envelope is used to model the non-linear interaction of a laser pulse with a plasma, speedups in the order ω_0/ω_p are expected. In general, the speedup will be determined by the reduction of the spatial and temporal resolution. For the finite-difference equation for the evolution of the envelope, a stability condition was derived. It restricts the numerical resolution for a given laser frequency. More specifically, if the vacuum stability condition is not satisfied, a numerical eruption occurs.

On the other hand, the propagation of a laser envelope through a plasma makes the PGC algorithm, as described here, unconditionally unstable. Nevertheless, the PGC algorithm, as shown in the following chapters, can be used to model LWFA scenarios. In section 2.4, the incubation time step was introduced. It characterizes the growth of a numerical error by a specific amplification factor. The incubation time step estimates that for typical LWFA scenarios, the error growth rate is sufficiently low that numerical disruption of the simulation is not present, and the presented finite-difference scheme still can be used.

As discussed in this chapter, the PGC algorithm requires data locality to invert a matrix and to solve the finite difference equation for the envelope's evolution. Even if the presented algorithm's compute complexity scales linearly with the number of cells, it hinders the parallel scalability on modern HPC systems. In the next chapter, different parallelization schemes will be adopted to ensure the parallel scaling of PGC.

3 Parallelization of the ponderomotive guiding center solver

For over a decade prophets have voiced the contention
that the organization of a single computer has reached its limits
and that truly significant advances can be made only
by interconnection of a multiplicity of computers

Gene M. Amdahl (1967)

Even with computational benefits arising from reduced models for plasma-based acceleration, such as PGC, the simulations are still computationally intensive and require large "supercomputers." The use of high-performance computers is not exclusive to plasma sciences, but is also crucial for many other areas such as weather forecasting [77–80], climate research [81], various physics disciplines and artificial intelligence [82–85]. For computer simulations, performance or better time-to-solution plays an essential role, and low performance can restrict the studied cases or the results' accuracy. In most scenarios, simulations are carried out on parallel systems comprised of many computing nodes, typically more than 10^3, connected by an advance interconnect. To be able to utilize such supercomputers, the computations must be able to be portioned and assigned to parallel resources for execution. In PIC, the standard parallelization strategy is to divide the simu-

lation domain into computational sub-domains known as domain-decomposition. Hence, each sub-domain can be solved independently, and only the boundaries are required to exchange information with the neighboring domains.

The bottleneck for PGC arises from a required matrix inversion to advance the envelope. The matrix inversion requires data locality in the transverse direction as a forward elimination, and a backward substitution is performed. Consequently, domain decomposition can not be directly applied to PGC, and other steps have to be considered.

In this chapter, we focus on the parallelization of the PGC solver. First, basic concepts and measures for parallelization are introduced, followed by an introduction of shared memory parallelization (SMP) and distributed memory parallelization (DMP). Both concepts are the main approaches taken to ensure that PGC can take the full advantage of modern cluster systems. Finally, results for scaling PGC over thousands of cores are presented.

3.1 Parallelization and scalability

Since the initial observation that the number of transistors doubles approximately every two years by Moore [86], it was believed that performance gains for computations would follow the same trend as the operational frequency would increase, known as Moore's law. This law endured until microprocessors reached a frequency of a few GHz. Due to overheating issues at higher clock speeds, the development of multi-core processors was necessary. Nowadays, even desktop computers can be found with several cores. On the other hand, cluster systems are a collection of computers interconnected over a network. With such heterogeneous configurations, the trend of having interconnected and large core counts requires computer simulations to be split into parallel chunks, and algorithms need to be designed to take full advantage of these parallel computing architectures.

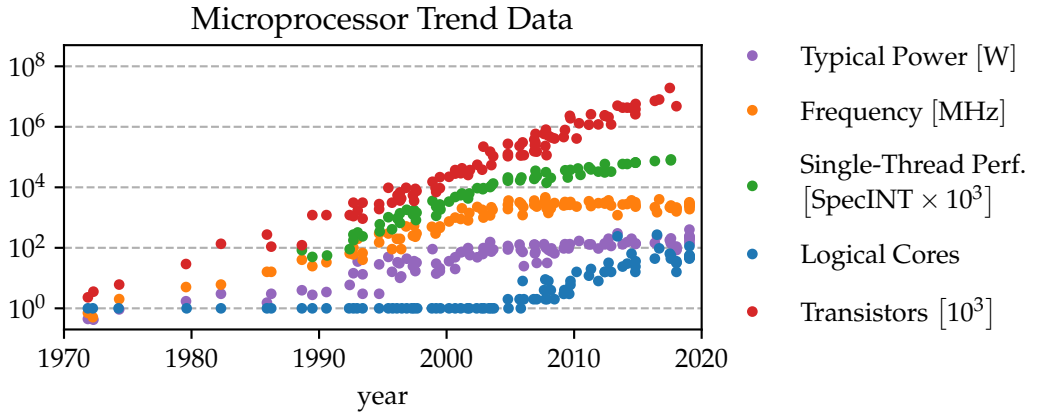


Figure 3.1: Microprocessor Trend Data for the past 44 years. Courtesy of Karl Rupp and Leonardo Suriano*for gathering and providing the data.

While in general several paths in parallelizing an algorithm can be taken, a quantitative measure is essential. To establish a benefit arising by parallelization, the speedup

$$S = \frac{T_{\text{serial}}}{T_{\text{parallel}}} \tag{3.1}$$

as a ratio between the time T_{serial} it took to perform a computation on a single computational unit and the time T_{parallel} it took to perform the same task in parallel is an appropriate measure. Theoretically, the speedup scales with the number of available parallel units N and the speedup is $S = N$ in an optimal case. An alternative approach is to characterize the efficiency

$$\mathcal{E} = \frac{S}{N} = \frac{T_{\text{serial}}}{N \times T_{\text{parallel}}} \tag{3.2}$$

of the parallel scaling. In optimal scenarios, the efficiency is 1, and in realistic cases, it is <1 as simulation has to wait for acquiring resources, retrieving information from neighboring computational units or if parts of the algorithm can not be paral-

*This data and plotting script is provided under the permissive "Creative Commons Attribution 4.0 International Public License". For more information and for the script used to generate figure, visit `https://github.com/ahelm/microprocessor-trend-data`.

lelized. Therefore any parallel application has a limited speedup. This notation is characterized by Amdahl's law [87] through a realistic speedup

$$S_{\text{real}} = \frac{1}{1 - p + \frac{p}{N}} \tag{3.3}$$

with p being the fraction of a simulation which benefits from parallelization. Furthermore,

$$\lim_{N \to \infty} S_{\text{real}} = \frac{1}{1 - p} \tag{3.4}$$

shows the limit of parallel applications. Considering a simulation that consists of 5 % not benefiting from parallelization, then the highest achievable speedup is 20.

3.2 Parallelization based on memory organization

The possibility of executing a task in parallel strongly depends on the underlying architecture and the available resources. As a broad definition, a parallel computer can be described as a collection of processing units to compute a solution for a given problem. Such processing units require memory to store information. Depending on how the memory is partitioned, two different parallelization models can be applied, distributed memory parallelization (DMP) and shared memory parallelization (SMP).

In the case of DMP, a distributed memory organization where each computational unit has a dedicated addressable memory is a prerequisite. As only one process can change the memory, a data race condition[†]can not appear. If a process requires information from neighboring processes, communication by message passing is required. Two potential bottlenecks can appear with communications

[†]Data race condition characterizes the substantial behavior of two or more systems competing for resources. The outcome depends on the sequence or timing of the events occurring. One example could be two processors who store information in the same location. In this example, the last value of the last processor will be stored in the memory location.

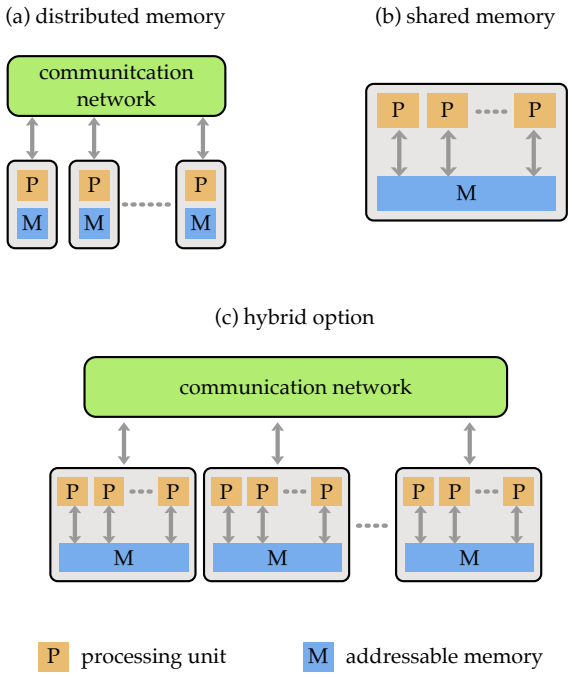


Figure 3.2: Schematic overview of different memory orchestration commonly found in high-performance computing. In a distributed memory case (a), each processing unit has its addressable memory, and information has to be passed through by communication. In a shared memory orchestration (b), memory is shared between processing units. Both orchestrations can be combined used to a hybrid option (c).

between two processes. As a communication requires to send data over a network, the communication's performance is bounded by the latency of the network and is typically slower than direct memory access. On the other hand, processes require to be load-balanced, and the slowest process's performance determines the total computation time.

As multi-core systems became a standard, SMP can easily be applied. In SMP, individual processing units share memory between each other. Therefore, no additional communication is required, removing the bottleneck caused by network latency. However, as writing and reading data synchronization is not guaranteed,

a data race condition can occur. Synchronization between the processes becomes crucial for critical operations and reduces parallel efficiency arising from additional overhead.

Moreover, both parallelization strategies can be applied as a hybrid option. A hybrid approach allows for more fine-grained control of memory access and helps mitigate the bottlenecks between both cases or scale over a higher number of processing units.

3.3 Shared memory parallelization for ponderomotive guiding center solver

For the algebraic problem (2.18), data locality is required. A SMP solution, e.g. based on `openMP`, seems favorable. While advancing the envelope can be parallelized through `openMP` directives for each dimension, additional steps are required for advancing particles. Based on the map-reduce idiom, the particle buffer can be split by chunks over several threads. Each thread processes the given particles individually, first, by interpolating the grid quantities onto the particle position. Second, advancing the particles based using the extended leap-frog method, presented in section 2.5. Third, depositing the grid-related quantities based on the updated positions and momentum. Each thread has a local buffer with the domain size for deposition, which avoids data race conditions as particles are not necessarily ordered, and deposition happens over several grid points depending on the interpolation order. After particles have been pushed, a thread based reduce operation is required before using the deposited quantities for advancing the fields and the envelope. The overall process for advancing particles is schematically shown in Fig. 3.3.

The parallel scalability of the SMP implementation can be measured through benchmarks. For this, benchmarks on an Intel Xeon Phi Knights Landing (KNL)

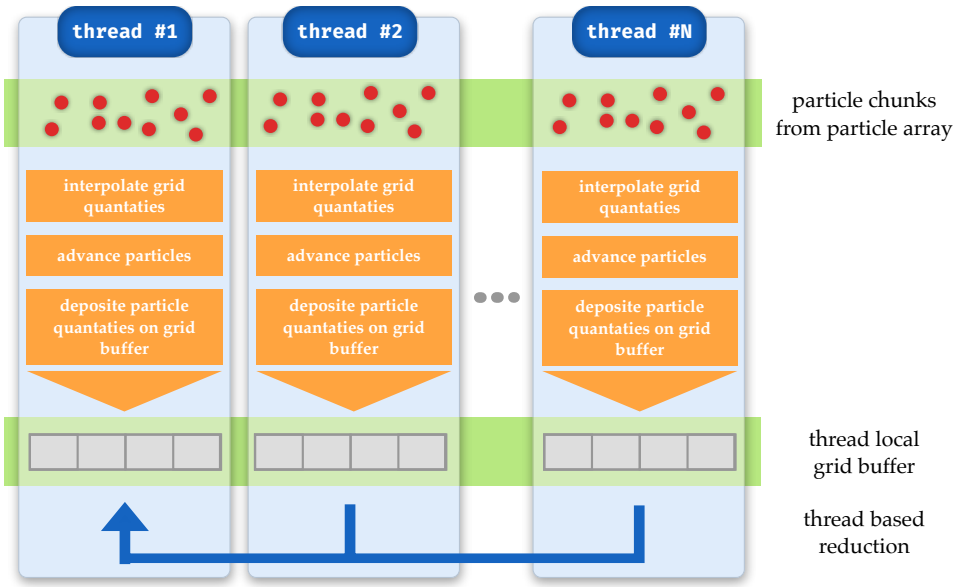


Figure 3.3: Thread-based particle advancing which is used to advance particles for shared memory parallelization.

architecture were carried out. The simulation time with a single thread is used as a reference point, and then the thread count is doubled in each step. For the presented case, 500 iterations were performed with periodic boundaries and a plane wave profile in transverse direction, a laser frequency of $\omega_0 = 100.0$, a pulse length of $L_{\mathrm{FWHM}} = 2.0$ and $a_0 = 1.0$. After 500 iterations, any overhead from initialization can be neglected, and the main contribution to the performance comes from the applied algorithms. The simulation box size is $50 \times 50 \times 50$ cells with 8 particles per cell. For the benchmark, the simulation time, the advancing of the envelope, and the particles' advancing are measured individually. The latter two are subtracted from the total time to deduce the other parts of the code's timing. Fig. 3.4(a) shows near-perfect scaling for the advancing of the particles.

On the contrary, the efficiency of the envelope drops significantly for the presented scaling. While some computations can be performed independently, e.g., the longitudinal directions, the bottleneck arises from solving the expressions (2.18a)

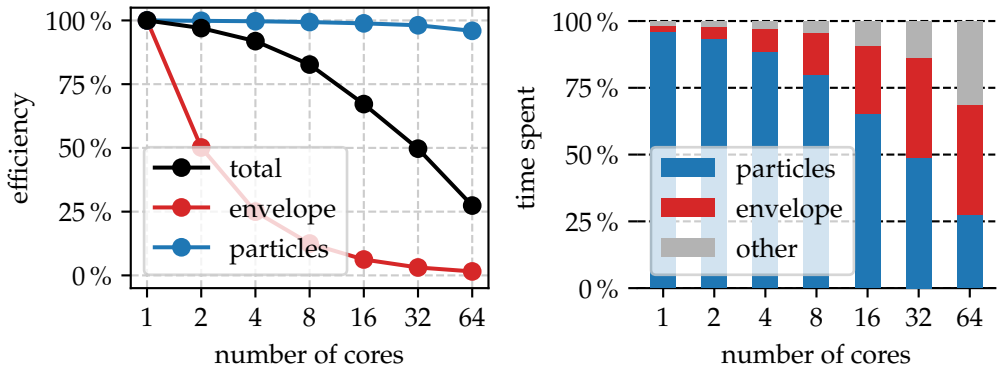


Figure 3.4: Benchmark for shared memory parallelization for the PGC solver. The benchmark was carried out on one Intel Knights lading (KNL) node using OpenMP.

and (2.18b) by inverting a matrix which, as mentioned before, requires forwards and backward substitution and can not be parallelized. An alternative approach might be to use cyclic reduction [75]. It allows for the step of inverting a matrix to be better parallelized over different threads or algorithms based on the Woodbury matrix identity [75]. In general, the total efficiency only drops below 80 % for more than 8 threads. As the total efficiency consists of several independent parts, the trend follows Amdahl's law, where the theoretical achievable limit for parallelism depends on the fraction of serial code. Here, the serial part corresponds to inverting the algebraic problem. As seen in Fig. 3.4(b), particle advancing dominates the total loop time for smaller thread counts, and with higher thread, other parts start to dominate. Thus, using SMP for parallelization over a small core would be applicable without significant efficiency drop. However, different approaches for parallelization are required to allow for scaling over a high number of cores.

3.4 Distributed memory parallelization for ponderomotive guiding center solver

For scaling over higher core counts, DMP is desired, especially when the simulation box is wide in the transverse direction. Generally, for a standard PIC algorithm, a domain-decomposition is applied. The total simulation domain is split into subregions, and the PIC algorithm is performed over the subregions. The advantage is that only information between the neighboring subregions must be shared to solve for the whole simulation box. In the case of PGC, matrix inversion is applied, e.g., utilizing the Thomas algorithm [75]. It consists of a forward and a backward sweep over the matrix equation to advance the envelope. For DMP, this would lead to a bottleneck as only one computational unit for inverting the matrix is used while the others are idle. Other matrix solvers, e.g., algorithms based on the Woodbury formula[75], could overcome the limitations by transforming it into auxiliary matrix equations. We choose an alternative approach inspired by parallel algorithms for Fast Fourier Transform (FFT) [88].

The required data locality can be ensured through a transpose operation by constructing the grid's local slices required for the matrix inversion. The construction of the local slices requires node-to-node communication in the transverse direction. After the slice is constructed locally, equation (2.18) is solved and is followed by another transpose operation to reconstruct the previous domain decomposition. Overall, as two matrix equations for 3D have to be solved, four node-to-node communications are required per time step. Compared to nearest neighbor communication in the typical domain decomposition of a PIC code, this can lead to scaling issues. Node to node communication is, in general, less favorable for scalability due to penalization by the latency of the communication.

The impact of additional communications have been tested by scaling studies on the JUQUEEN supercomputer in Jülich, Germany. The JUQUEEN supercomputer was a Blue Gene/Q system, which was online until mid 2018 and consisted of

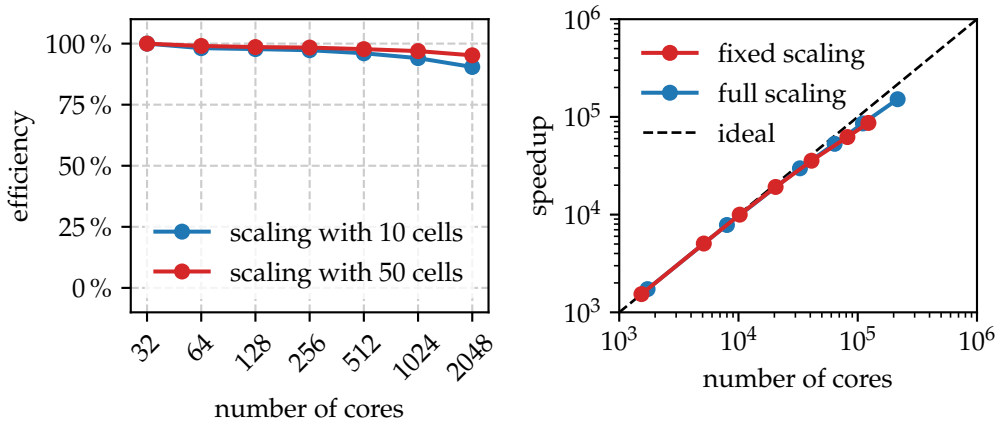


Figure 3.5: Distributed memory parallelization benchmark for parallel scalability of the PGC solver. The scaling tests was carried out on the JUQUEEN supercomputer.

458 752 cores. A weak scaling test was performed for the first test to analyze the efficiency drop arising from node-to-node communication. For the scaling test an initial box of $2048 \times 10 \times 50$ cells with 8 particles per cell is used. The scaling is done from 32 cores (one domain) to 2048 cores (64 domains) with each transverse direction being scaled individually. As shown in Fig. 3.5(a), the implementation has less than 10 % efficiency drop for a 64 times bigger domain. Also, having transversely larger domains is better in terms of efficiency. It is expected that this is an artifact of the network configuration. A strong scaling study was performed and presented in Fig. 3.5(b). In this case, a near-ideal scaling while increasing the number of utilized cores by 10^3 times is achievable. We use a simulation box of $15\,360 \times 240 \times 240$ cells with periodic boundaries in transverse direction and 8 particles per cell. Here, to check for impacts on increased node-to-node communication, two scenarios are distinguished. First, where the number of transverse domains stays constant and with an alternating number of domains, both scenarios show similar scaling. The drop towards the end is expected, and as the number of particles per core reduces, and communication becomes dominant [89].

3.5 Conclusion

In this chapter, parallelization schemes using SMP and DMP for the PGC solver were presented. As a matrix inversion in the transverse has to be performed for advancing the laser envelope, parallelizing PGC becomes challenging. A way to overcome this can be done through SMP. The presented algorithm requires special treatment of the particles. Here, splitting the particles over a different number of threads was proposed and presented. While this approach scales well for the particles, the solution is serial and therefore scales poorly for the envelope. Scaling over a small number of processing units is possible, but for simulations with larger simulation boxes or to utilize a full cluster system, a DMP approach was proposed and incorporated. While with this approach, data locality forces additional steps. By utilizing a transpose operation similar to FFT algorithms, it was shown that PGC could be scaled over modern HPC systems. Moreover, scaling over the full JUQUEEN machine showed the near-ideal scaling.

In general, the two presented algorithms for SMP and DMP can be utilized together in a hybrid parallelization scheme. While the DMP scheme can be used to parallelize PGC over almost the whole system, SMP can allow one order of magnitude more cores to be utilized.

4 Parametric studies of laser-plasma accelerators using PGC

The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.

Sir William Lawrence Bragg

Plasma-based accelerators are promising candidates for electron acceleration in the context of a future generation of light sources and particle colliders. Much progress has been made to demonstrate high acceleration gradients [24, 90], generation of high-quality beams [31–33] (with promising recent progress [91]) and applications and diagnostics [92]. It is believed that plasma-based accelerators may have a profound impact on basic research, medical applications, and material science by providing a compact electron beam source. However, applications that require excellent beam quality, such as free-electron lasers (FELs), are still a challenge for plasma acceleration. The typical percent-level energy spreads obtained at the moment would not be enough to drive FELs, which typically require slice energy spreads of 0.01 % to 0.1 % in order to lase in the X-rays domain [93].

The path towards high-quality beams leans heavily on experiments and even more on simulations to probe new regimes and ideas. Simulations not only laid the groundwork for the whole field of PBA by the work of Tajima and Dawson [17],

but predicted the bubble regime [42] two years before its experimental observation [31–33]. From the simulation point of view, the PIC method is an established tool for modeling plasma-based accelerators. In this method, the grid cells must be sufficiently small to resolve the system's shortest physical lengths. Usually, a plasma for experiments can be extended up to 15 diffraction lengths on the order of tens of cm [91]. However, this leads to a scale disparity of at least a few orders of magnitude between the acceleration length and the grid resolution. Consequently, full-scale three-dimensional PIC simulations are computationally expensive, making parametric studies of LWFA challenging and computationally demanding.

The use of PGC is especially interesting as it allow large parametric scans. For typical LWFA parameters, a speedup of at least two orders of magnitude can be achieved. However, one must be careful and understand the limitations and applicability of these algorithms to different physical scenarios. While reducing computational cost is essential, a reduced model should ensure that the most important physical processes are recovered. The challenging aspect with reduced solvers is modeling the acceleration process, and even more challenging is to model injection and trapping of background electrons in the accelerating structure. One way of quantifying a reduced solver's performance, on that matter, is to inspect the different phase-space properties associated with the bunch quality (e.g., energy spread, beam emittance, charge, mean energy). While previously proposed techniques, such as the quasi-static algorithm [68], are known not to model the self-injection accurately, for PGC, this has to be examined in detail.

This chapter compares the performance of the PGC algorithm with full PIC in the context of the EuPRAXIA project [94]. A key goal of EuPRAXIA is to produce 5 GeV, 30 pC electron beams from plasma accelerators with high 6D-brightness [95], that could be used to drive FELs among other applications. Simulations for several different schemes and design configurations have been investigated [96–99]. Here, investigations of a LWFA injector for EuPRAXIA using density down-ramp injection is reported. For this component, the electron beam energy required is 150 MeV,

and this beam would be transported to another plasma stage for further acceleration to 5 GeV [97]. Therefore, studies comparing the PGC algorithm with three-dimensional PIC simulations for density down-ramp injection are the main focus of this chapter. This approach enabled us to examine the computational performance, accuracy, and limitations of the PGC algorithm for this injection scheme. In particular two cases are studied, for very short density transitions, such as in shock-front injection [100], and for lower plasma densities ($\omega_0/\omega_p \gg 10$, where the computational gain is significant) that the PGC algorithm can accurately predict the injected beam charge and the energy gain of full PIC simulations. While reduced models are expected to be less accurate, they can still be used as guides to predict key bunch properties, such as its energy and charge. The quality of these predictions by comparing reduced simulation results with full PIC in 3D is assessed.

4.1 Parametric studies for down-ramp injection

Reduced modeling allows for efficiently exploring vast parametric spaces. With the computational savings achievable with PGC, broader studies can be performed, giving insight into the underlying physics of complex processes, such as the nonlinear nature of LWFA. The examination of larger parametric spaces is beneficial for projects in their initial design stage, like the EuPRAXIA project [94]. It would potentially narrow down possible paths and set a range of parameters for experimental setups.

As the first test, the parametric space of a short density transition is examined. Figure 4.1 shows a sketch of the density profile. After a short plateau with a density of $2.9 \times 10^{18}\,\mathrm{cm}^{-3}$, the density drops down to $1.9 \times 10^{18}\,\mathrm{cm}^{-3}$. The ramp size is on the order of the plasma wavelength, which is typical for shock-front injection schemes [100]. During the density transition, some electrons at the back of the bubble will be injected in the wakefield as the plasma wave expands, forming a bunch. The beam evolution is followed for around 1.75 mm. The chosen density

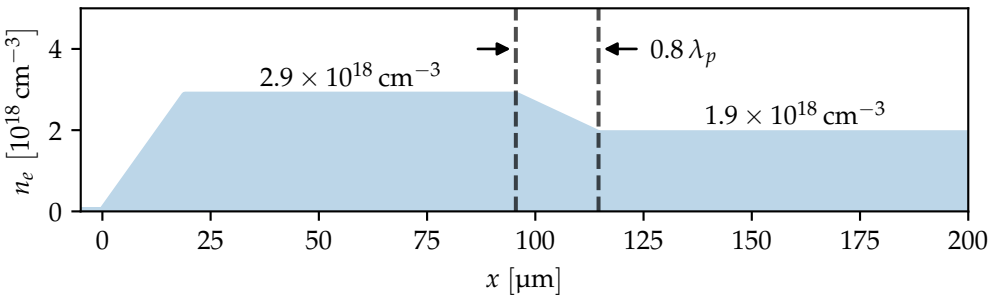


Figure 4.1: Density profile used for the parametric studies with a short density transition.

profile is commonly used to study the injection process in density down-ramps (see, e.g., Refs. [101, 102]). Moreover, the benefit of using a short plateau reduces the effects of envelope evolution and its contribution to the particle injection.

As for the laser parameters, a central wavelength of 800 nm and a spot size of $w_0 = 17.2\,\mu\text{m}$ are used. For the parametric study, the initial peak envelope amplitude a_{init} is scanned from 0.5 to 4.5 and the pulse duration τ_{d} from 19.1 fs to 57.3 fs. Overall this leads to 63 full 3D simulations for the given parameter ranges. The laser pulse's longitudinal profile follows a 5^{th} order polynomial [76], which is used to mimic a Gaussian profile. The injected charged for different laser strengths, and pulse duration is shown in Figure 4.2.

From the parametric sweep, a clear lower threshold for particle injection to occur is observable. The envelope has to fulfill the condition $a_{\text{init}} > 2.0$ to inject particles for the chosen setup. If the laser strength is increased, the charge increases, leading to an injected charge above 1 nC. This behavior is supported by the theoretical model for down-ramp injection [45]. With increasing laser strength, the amplitude of the wake increases, and during the bubble expansion, more particles can obtain a high enough longitudinal momentum to be trapped. As previously mentioned, the first density plateau is short, and no significant modification to the laser profile is expected. Hence, a polynomial dependency has to be expected and is supported by

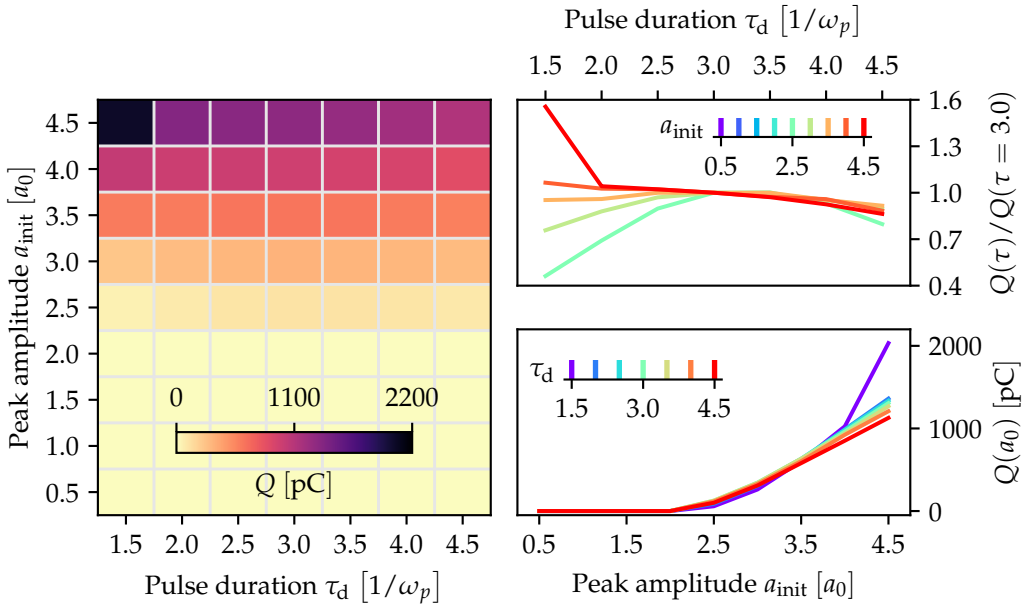


Figure 4.2: 3D parametric study of particle injection for a short density transition. The laser amplitude and the pulse duration were changed for the study, and the injected charge after 1.75 mm of propagation distance was measured. The left plot shows a parameter matrix for the different physical parameters used with the corresponding injected charge. Plots for lineouts for the different parameters are shown on the right.

theoretical descriptions of the wake amplitude [103]. Here, the polynomial trend for increasing peak amplitude is recovered. Similar to this, the simulations with $a_{\text{init}} = 2.5$ shows the highest charge for a pulse duration close to half a plasma oscillation $\tau_d = 3.0$. However, for higher laser amplitudes, this trend is not obeyed. The observation for the case of $a_{\text{init}} = 2.5$ is similar to 1D linear theory, where the highest wake amplitude is achievable when the laser duration is comparable to half a plasma oscillation. This trend is not observable for higher laser amplitudes as the laser pulse propagates through the plasma and gets compressed, the local laser field amplitude exceeds in certain scenarios values above $a > 10$, which leads to additional injection. Hence, the highest observable charge of 2 nC is obtainable

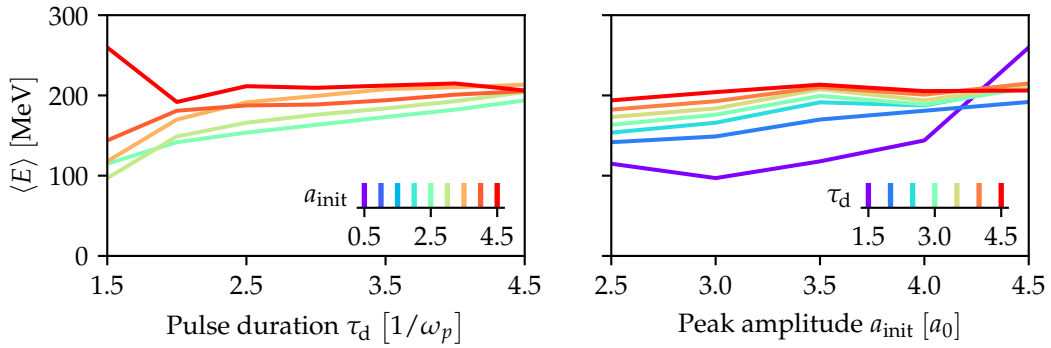


Figure 4.3: Comparison of the mean energy for varying pulse duration and peak amplitude for a short density transition. Shown are 35 different parametric studies where injection occurred. The left plot shows the mean energy dependence on the pulse duration, with each line corresponding to a fixed peak amplitude value. Similarly, the right plot shows the mean energy dependence on the peak amplitude for a given pulse duration.

for a laser pulse with an initial pulse duration of 19.1 fs and a peak envelope of $a_{\text{init}} = 4.5$.

The comparison can be extended further, to compare the averaged energy of the injected particle bunch. In figure 4.3, all the simulations, which include injected particles, are shown. In comparison, a trend for the mean particle energy is recovered – as the pulse duration and initial peak amplitude increase, the mean energy increases. This trend is associated with the higher energy carried by the laser pulse, which drives the wake over larger distances. Nevertheless, for the short laser pulse with $\tau_d = 1.5$, the trend is not followed. Here, the approximation of slow-varying amplitude over one laser cycle breaks down. More specifically, for the cases $\tau_d = 1.5$ the laser pulse contains less than 8 optical cycles and a discrepancy is to be expected.

In summary, the study presented shows the potential of performing simulations using a reduced solver like PGC. The total computational time for all 63 full 3D simulations was 0.5×10^6 CPUh, which is comparable only one full PIC simulation.

This reduction means that a full parametric scan can be performed for the same costs as a full PIC simulation. Undoubtedly, one has to ensure that the physical properties of the injected beam are recovered. To this end, we perform detailed one-to-one comparisons between PGC and full PIC in the following sections.

4.2 Comparison between PGC and PIC for short density transitions

As the PGC algorithm is the right candidate for performing parametric studies, verification of how well the reduced algorithm performs against full PIC simulations is required. For this, the discussion is extended and different beam parameters, such as energy, energy spread, and emittance are compared against a full PIC.

For the simulation, the density profile from the previous Section is used. The dimensions of the simulation box for the full PIC and PGC simulations are 69.2 µm × 215.4 µm × 215.4 µm with 280 cells in each transverse direction. It has to be noted that while a better transverse resolution may be needed to describe the bunch transverse dynamics accurately, our focus is to compare both setups without investigating the numerical resolution required for establishing convergence of the simulation towards a real physical scenario. The number of longitudinal cells and the time step is given by resolving each simulation's shortest scales. For full PIC, the fastest oscillations are associated with the laser pulse. Hence, the number of cells in the longitudinal direction is 5400, and the time step is $\Delta t = 41.1$ as, to resolve the laser wavelength and period. For the PGC simulations, the number of cells in longitudinal direction is 180 with $\Delta t = 1.03$ fs, corresponding to 63 points per plasma wavelength. In both cases, the Yee electromagnetic field solver [104] is used.

In a direct comparison of the injected charge results in 249 pC for the full PIC simulation and 227 pC for the PGC, i.e., a 9 % relative difference. The injected charge is high when compared with LWFA experiments. However, it is the typical

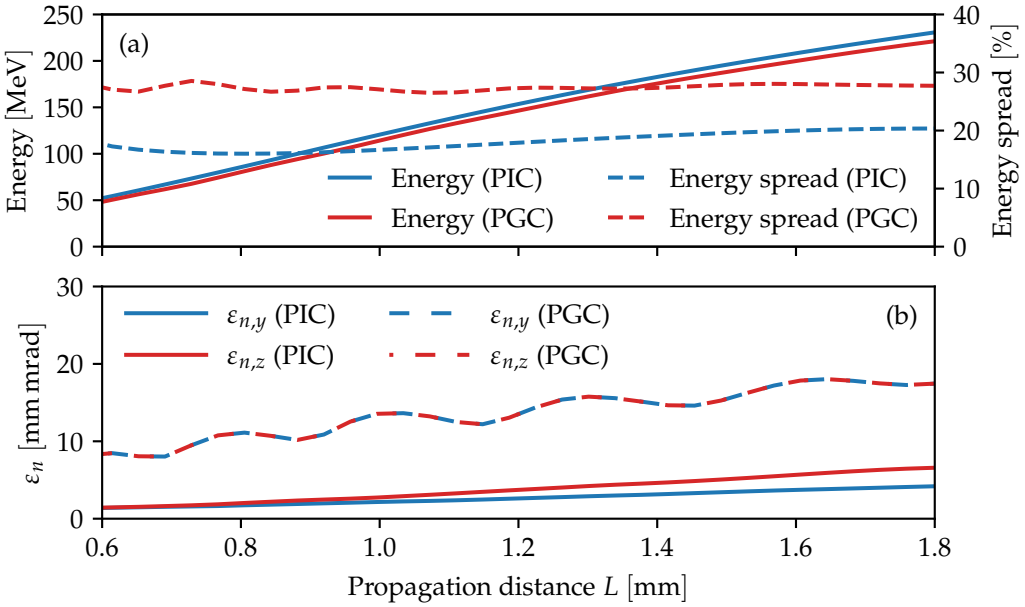


Figure 4.4: Comparison between the beam parameters in full PIC and PGC with a sharp density gradient. The propagation distance corresponds to the laser pulse propagation inside the plasma. The upper panel (a) shows the evolution of the mean energy [solid curves] and energy spread [dashed curves] for full PIC [blue curves] and PGC (red curves). The lower panel (b) displays the normalized emittance for full PIC [solid curves] and PGC [dashed curves] in the y-direction [blue curves] and the laser polarization z-direction [red curves].

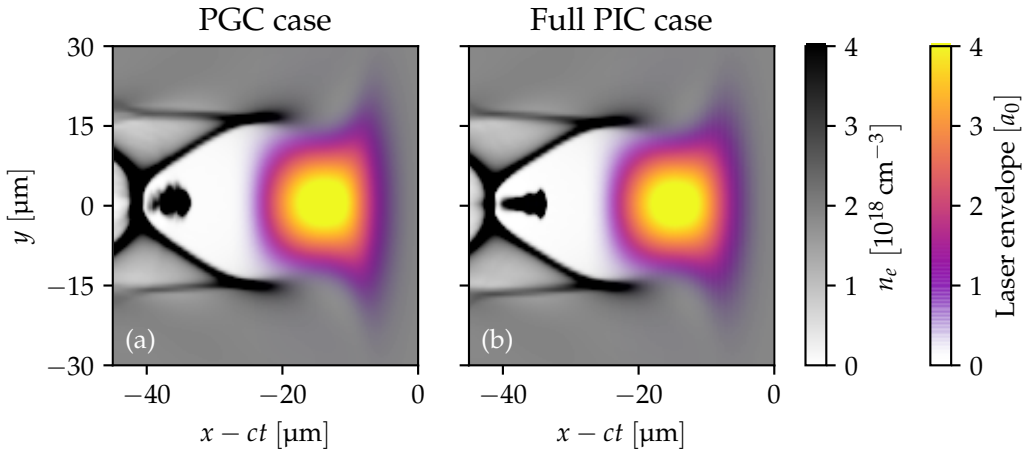


Figure 4.5: Electron density and laser envelope showing a comparison of the injected bunch in the PGC [panel (a)] and full PIC [panel (b)] simulations for the short density transition case. The laser propagated for both scenarios for a distance of $L = 1.2\,\mathrm{mm}$.

charge expected in down-ramp injection scenarios using a steep density gradient, which maximizes the injected charge per unit of length during the density transition [101] and values of this order have been observed in recent experiments [105]. Figure 4.4(a) shows a comparison of the electron bunch mean energy and the relative energy spread during acceleration from PGC and full PIC simulations. An excellent agreement in the acceleration gradient with a relative error of 5 % is noticeable. While the numerical value of the energy spread is different, the overall trend is similar. In general, the values of the energy spread and emittance can differ in simulations using reduced models compared to full PIC.

Figure 4.4(b) shows a comparison of the normalized emittance evolution between PGC (dashed lines) and full PIC (solid lines). The normalized emittance is defined as

$$\epsilon_{n,y} = \frac{1}{m_e c}\left(\langle y^2\rangle \left\langle p_y^2\right\rangle - \langle y p_y\rangle^2\right)^{1/2}, \tag{4.1}$$

where p_y is the canonical momentum to the coordinate y. The solid curves refer to

the full PIC simulations (blue in the y direction and red in the z direction, which is also the laser polarization direction), and the dashed curves represent the PGC simulations. Again, a similar trend of the emittance variation with time is found, but the emittance in the PGC case is higher by a factor of 5. This difference can be attributed to the injection process, which depends on the time step used in the simulation. In the PGC simulation, the transverse particle momentum at the injection point exhibits larger variations than in full PIC because the time step is higher in the PGC case. This increased time step increases the emittance in the PGC simulation and makes the injected bunch wider, as seen in the comparison in Fig. 4.5(a) for PGC and (b) for full PIC. Increasing the temporal resolution could reduce the emittance difference arising from the injection. A time step reduction without reducing the spatial resolution would increase the growth of the numerical Cherenkov instability. The other remark regarding the comparison is the lack of the inclusion of the laser polarization direction in PGC. This exclusion leads to both transverse emittances being equal for PGC, compared to full PIC where the normalized emittance is higher in the laser polarization direction.

Due to the reduced temporal and spatial resolution, the speedup of PGC over PIC was 480 times, making it a viable candidate for studying down-ramp injection for short density transitions. Even though PGC shows promising results, the injected beam does not fulfill the goals set for an injector stage, e.g., the required energy spread and emittance. Here, different density profiles with longer density down-ramps are shown to offer better control.

4.3 Electron injection for long density transitions

Now we consider longer ramps, of the order of hundreds of µm, and the plasma profile similar to ones obtained by hydrodynamic simulations of gas cells [106]. Fig. 4.6 displays the density profile used in our simulations. This profile is the optimal case found in our study for the EuPRAXIA project [94]. Here we compare

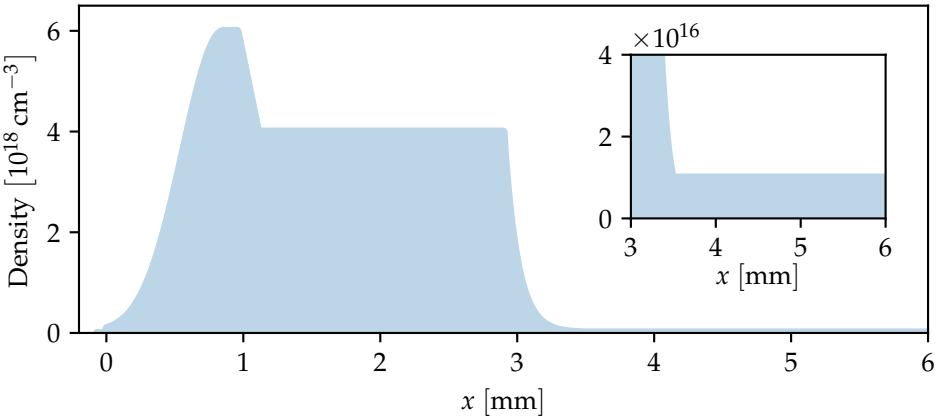


Figure 4.6: Sketch of density profile used in the simulations. The inset shows a zoom of the plasma lens region.

PGC and full PIC simulations.

The profile contains an up-ramp with a Gaussian profile and the total length of 1 mm, where the density goes from 0 cm^{-3} to 6×10^{18} cm^{-3}. The ramp is followed by a down-ramp, where injection occurs, with 150 µm length, which brings the density down to 4×10^{18} cm^{-3}, where it reaches a plateau. This beam is accelerated in the plateau for 1.8 mm. Finally, a plasma exit is considered with exponential profile and total length of 500 µm coupled with a 2 mm long 10^{16} cm^{-3} plasma [inset of Fig. 4.6], which works as a plasma lens [107].

The central laser wavelength in vacuum is $\lambda = 800$ nm, the beam waist and normalized vector potential at the focus are $w_0 = 18$ µm and $a_0 = 1.8$, respectively, and the pulse duration is $\tau_{\text{FWHM}} = 30$ fs. The laser is linearly polarized in the y direction and focused 217 µm into the plasma profile of Fig. 4.7.

For the PGC simulations, the simulation box is 52 µm × 86 µm × 86 µm and it is divided in 180 × 400 × 400 cells, and each cell starts with 16 macro-particles. The time step used for the PIC loop is $\Delta t = 0.225$ fs and we use the Yee solver for the plasma electromagnetic fields. For the full PIC, the dimensions of the box are the same but it is divided in 3200 × 400 × 400 cells, each cell starts with 4 macro-

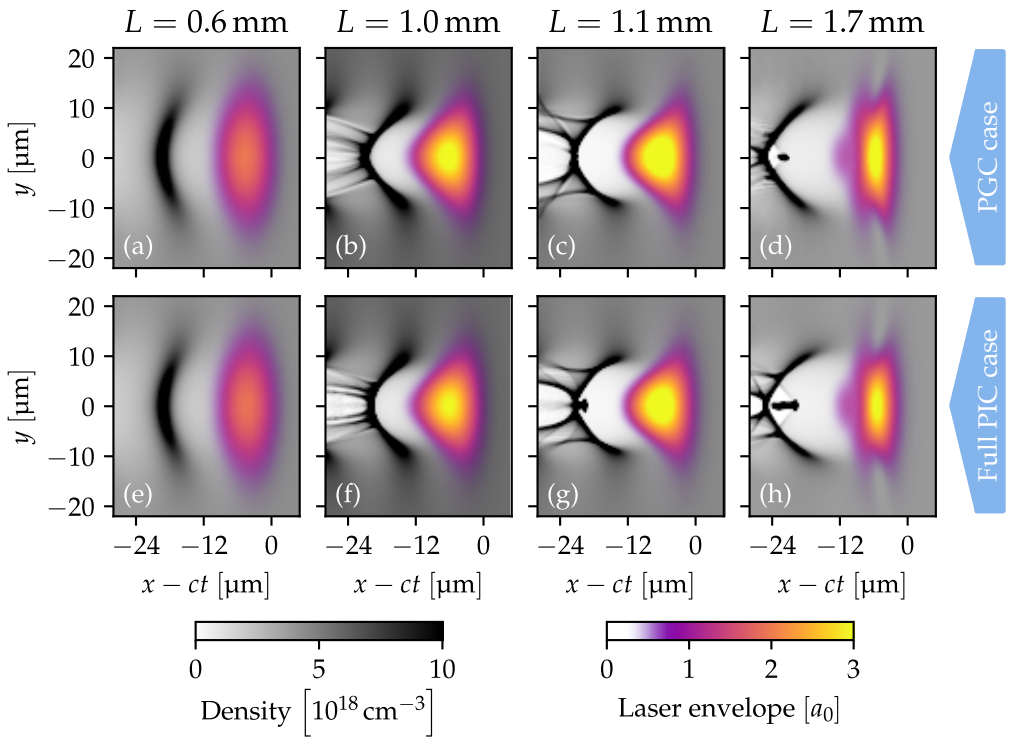


Figure 4.7: Comparison of PGC [(a)-(d)] and full PIC [(e)-(h)] simulations for the electron density and the laser envelope at four different propagation distances L. The laser envelope and plasma structures are well described in the PGC simulations. However, the injection during the density transition is not accurately described for long ramps.

particles, and time step $\Delta t = 51.235$ as. To mitigate NCI in the full PIC simulations, a electromagnetic field solver developed by [108] is used.

Figure 4.7 shows a comparison between PGC and full PIC simulations for the laser envelope and electron density at four different times during the system evolution. Figure 4.7(a) and 4.7(e) present the laser as it propagates in the density up-ramp, where it is being self-focused ($a_0 \approx 2.3$ at $z \simeq 0.7$ mm, while its maximum value in vacuum is $a_0 = 1.8$). The laser is thus greatly affected by the plasma – it self-focuses as it propagates in the plasma up-ramp. At $z = 0.68$ mm, the wakefield resembles the quasi-linear regime, and there is no beam in the accelerating structure. There is an excellent agreement between PGC and full PIC simulation. Figure 4.7(b) and Figure 4.7(f) display the system at the beginning of the density down-ramp, where we now observe a non-linear plasma wave (bubble) and the laser strongly self-focused, which is well described by the PGC algorithm. Before injection, around 1 mm into the plasma, the normalized vector potential gets as high as $a_0 \approx 4.0$.

While the laser envelope and plasma structure are still well described in Figure 4.7(c) and 4.7(g), the comparison with full PIC [panel (g)] shows the first significant disagreement as the latter shows injection at the rear of the bubble which is not present in the PGC case [panel (c)]. Finally, Figures 4.7(d) and 4.7(h) display the system after it reaches the density plateau. There is still an excellent agreement of the laser envelope between the two cases. However, the injected beam is quite different. For example, the total charge is 4 times higher in the full PIC case, which leads to a much stronger beam loading. The difference can be observed by carefully examining the region near the back of the bubble in panels (c) and (g). We conjecture this discrepancy is due to the lack of spatial and temporal resolution to resolve the injection process if the ramp is too smooth. If that is the case, one could infer that the computational gain of the PGC will be lower, as one needs to resolve the injection scale, instead of the plasma scales. Additionally, the algorithm becomes numerically unstable for a high longitudinal resolution, which limits how much

one can increase it. One possible solution is the use of sub-cycling of the particle pusher.

Overall, the laser evolution and the plasma structures are well described by the PGC model. However, in longer ramps, the injection details during the density transition are not fully recovered. The electromagnetic fields are also in good agreement between the two cases before injection. These results show that we can use the PGC in the context of EuPRAXIA, for example, to probe the laser evolution quickly, the plasma response, and the global plasma wakefield structure and dynamics at the time of injection.

4.4 Conclusion

In this chapter, the focus is on using PGC for reduced modeling. The reduced computational cost associated with PGC allows us to perform extensive parameter studies in 3D geometries with a fraction of the computational cost associated with full PIC. It was shown that PGC describes the laser evolution and the plasma dynamics for propagation well. In particular, the two studied scenarios for down-ramp injection of a short and a long density transition agree well for the envelope evolution and the plasma response to the laser. While in the case of a short density transition, the mean electron energy was in a good agreement, the emittance was by a factor of 5 greater. This difference was arising from a higher time step in the PGC case.

On the other hand, in the case of longer density transitions, a more significant difference in particle injection is observed, and a delayed injection in the case of PGC occurs. Here, similar to a short density transition, the time step and the spatial resolution affect the injection process. While higher resolution might model the injection better and give better agreement with PIC simulations, two points have to be considered. First, the stability of the envelope solver has to be guaranteed. Second, performing simulations at a higher resolution will reduce the speed-up

obtained by PGC and the purpose of performing reduced model simulations in the first place.

In summary, no significant differences in the evolution of the laser and the generation of the plasma waves were seen. This fact makes PGC an excellent tool to perform large parametric scans. However, due to the reduced numerical resolution, large density spikes are not adequately resolved and could introduce numerical artifacts.

5 Full-scale modeling of the relativistic ionization front for the AWAKE experiment

Was besteht und wirkt,
muß einen Grund
seines Bestehens und Wirkens haben.
(eng.: What exists and acts,
has to have a reason for
its existence and its action.)

Gottfried Wilhelm Leibniz

While in the original concept of PBA by Tajima and Dawson, a laser driver is used, this scheme holds some limitations. It is particularly evident if it is used for accelerating particles to energies in the TeV-range. For this, LWFA requires staging [109] as a laser pulse carries a small amount of energy. An alternative approach to LWFA is to use these high energy particle bunches to drive plasma wakefields. It allows the use of already existing facilities to increase the peak energy of those conventional accelerators. Driving wakes utilizing a particle bunch was already shown experimentally by different groups [47, 110–112].

An electron bunch, as an example for a charged particle bunch, repels electrons in an electron plasma while propagating through it. The cavity left behind the beam attract the repelled electrons toward the axis. On the other side, if the driver has a positive charge, e.g., a positron beam, the plasma's electrons are pulled towards the center axis and overshoot as the particle beam passes. Similar to a laser driver, this results in an electron deficiency and electron surplus areas.

Proton bunches are leading in terms of energy stored in a bunch. For example, the LHC proton bunch carries an energy of 1.25×10^5 J [6] compared to 1.6×10^1 J for the drive bunch planed for FACET II [113] and up to 2×10^1 J for a laser pulse at the Extreme Light Infrastructure (ELI) beamlines. Hence, using proton bunches to drive plasma wakes in plasma would allow us to accelerate electrons to TeV-range in a single stage. Indeed, this was suggested by Caldwell et al. [50]. While in the proposed setup, a proton bunch energy of 1 TeV, corresponding to the energy achieved by the LHC, was assumed, it also required the proton bunch to be 100 µm short. Such short bunches are favorable as it increases the peak electric field in a linear regime

$$E_{\max} = 240 \left[\mathrm{MV\,m^{-1}}\right] \frac{N}{4 \times 10^{10}} \left(\frac{0.6}{\sigma_x}\right)^2 \tag{5.1}$$

where N is the number of particles in the driver and σ_x is the root mean square (rms) length of the driver [50]. On the other hand, the requirement of short bunches imposes an additional challenge. A typical proton bunch for the LHC has a length of 7.55 cm which is orders of magnitudes longer than the bunch proposed by Caldwell et al. [50]. Using plasmas can help to overcome this disparity. Inspired by the self-modulated LWFA [114], it was proposed that self-modulation instability (SMI) could lead to the breakup of a long proton bunch into small beamlets. These beamlets would then drive the wakefields and accelerate injected electrons which is the main idea behind Advanced WAKEfield Experiment (AWAKE) at CERN.

In AWAKE, the proton bunch is coming from the SPS accelerator and is cut into beamlets, schematically presented in Fig. 5.1. This is achieved by utilizing 10 m

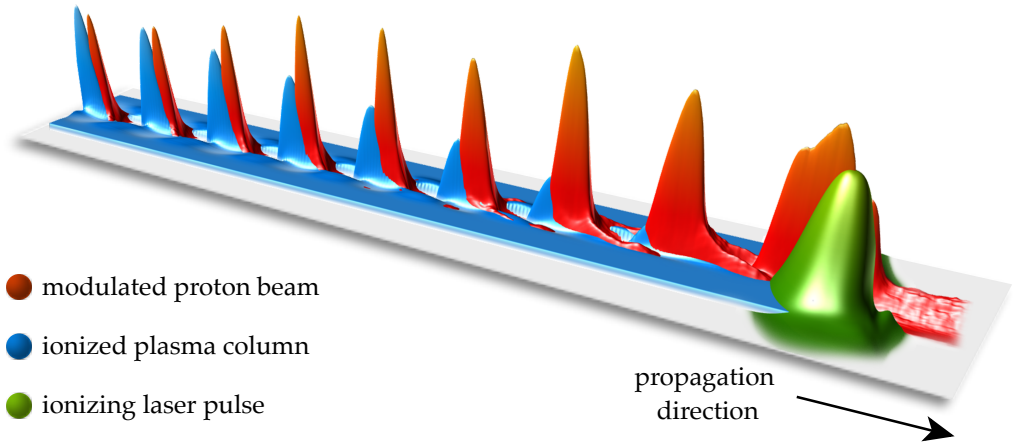


Figure 5.1: Sketch of seeding the self-modulation instability by using a electron plasma which is generated by an ionizing laser. The laser and the proton beam are co-propagating in a neutral gas, e.g. Rubidium. The plasma column of electrons is used to modulate a long proton bunch into beamlets which drive plasma wakes.

long neutral Rubidium gas cell with a gas density $10^{14}\,\mathrm{cm}^{-3}$ to $10^{15}\,\mathrm{cm}^{-3}$ and a laser pulse with a wavelength of 800 nm. Co-moving with the long proton bunch is the laser pulse. It generates a plasma by ionizing the Rubidium gas. Due to the transverse wakefields on the bunch itself, a modulation is triggered and gives rise to the SMI [115]. In recent experiments, an acceleration of up to two GeV in a single stage by using externally injected electrons was shown [49].

While numerous simulations and theoretical works support the AWAKE collaboration results, self-consistent and full-scale modeling are not present. The most prominent aspect of full modeling is the inclusion of ionization and its contribution to the growth of the SMI. As the SMI grows exponentially, the noise level is a determining factor and an uncontrolled seeding source for the instability. In general, noise could arise from the plasma's non-uniformity, the non-uniformity of the proton bunch itself, or be subject to the density and thermal fluctuations. Such a scenario would not be favorable for an accelerator as experimental results would

fluctuate significantly depending on the initial conditions. Hence, controlled initial seeding for the SMI is needed. In the AWAKE experiment, this is achieved using an ionizing laser to create a sharp ionization front, which allows us to effectively seed the SMI. However, PIC simulations of the AWAKE experiment exclude the ionizing laser pulse because the ratio between the laser frequency and the plasma frequency is $\omega_0/\omega_p > 1000$ and are merely impossible due to high computational costs. To circumvent this, the SMI is commonly seeded by assuming a sharply cut proton beam, which propagates into a pre-ionized plasma column of a fixed size. Here, the PGC solver could bridge the gap and allow it to include an ionizing laser.

In the next section, PGC as a tool for self-consistently modeling of the SMI is explored. First, the incorporation of an ionization model into PGC is described and discussed. An example setup is then presented and modeled in 3D using the incorporated ionization model in PGC. To compare the ionization model against standard simulation setups of the AWAKE experiment, a sharply cut proton bunch with similar parameters is used and compared against the case where ionization modeling is discussed in section 5.4. In addition to the SMI, another instability can occur the so-called hosing instability. The transverse hosing instability arises due to the beam's coupling to the plasma electrons as it propagates through an underdense plasma [116, 117]. While the hosing instability is suppressed if the SMI is seeded, asymmetries can reduce the suppression. Therefore this chapter will conclude with a case study where misalignment between a laser pulse and the proton beam is introduced.

5.1 Ionization model for the envelope

As the ionizing laser plays a significant role in the AWAKE to seed the SMI, acurate modeling of the ionization physics is necessary. In PIC codes, the Amosov, Delone, and Krainov (ADK) [118] model is most commonly used for ionization and will be the basis for including ionization into PGC. The ADK model assumes tun-

neling as the dominant ionization mechanism. Due to large scale disparities in the AWAKE experiment, the ionization laser propagation must be modeled using the PGC algorithm. Indeed, PGC enables 10^7 speedup compared to full PIC. Moreover, a full 3D simulation over the full propagation distance of the AWAKE experiment using PGC requires millions of CPUh, which is on the order of the largest possible computational allocations awarded for HPC allocations.

For the ADK model, ionization is characterized by the ionization rates

$$W\left[\mathrm{s}^{-1}\right] = \mathcal{A}E^{-\mathcal{C}} \times e^{-\mathcal{B}/E} \tag{5.2}$$

where E is the absolute value of the electric field in $\mathrm{GV\,m^{-1}}$, and the parameters $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are calculated for each ionization level. Above expression (5.2) characterizes a general approach. It can be used to model ionization in the case of a laser field and for using an averaged description of a laser. More specifically in the case of an envelope, for each ionization level i the following relations

$$\begin{aligned} A_i &= \frac{1}{2\tau_a}\left(\frac{2e}{n_s}\right)^{n_s}\left(\frac{1+2(n_s-1)}{\sqrt{2\pi n_s}}\right)\left(\frac{\xi_i}{\xi_a}\right)\left(2\left(\frac{\xi_i}{\xi_a}\right)^{3/2}E_a\right)^{2n_s-1} \\ B_i &= \frac{2}{3}\left(\frac{\xi_i}{\xi_a}\right)^{3/2}E_a \\ C_i &= 2n_s - 1 \\ n_s &= \frac{Z}{\sqrt{2\xi_i/\xi_a}} \end{aligned} \tag{5.3}$$

apply with the atomic number Z, the ionization energy in eV for each ionization level ξ_i, the atomic energy scale $\xi_a = 13.6\,\mathrm{eV}$, the atomic time interval $\tau_a = 2.42 \times 10^{-17}\,\mathrm{s}$, and the atomic field $E_a = 514.22\,\mathrm{GV\,m^{-1}}$.

One important aspect to discuss is the differences between an ADK model using a full description of the laser field versus a reduced one. For this, a rubidium vapor is considered. As a benchmark, a laser pulse with a normalized vector potential of

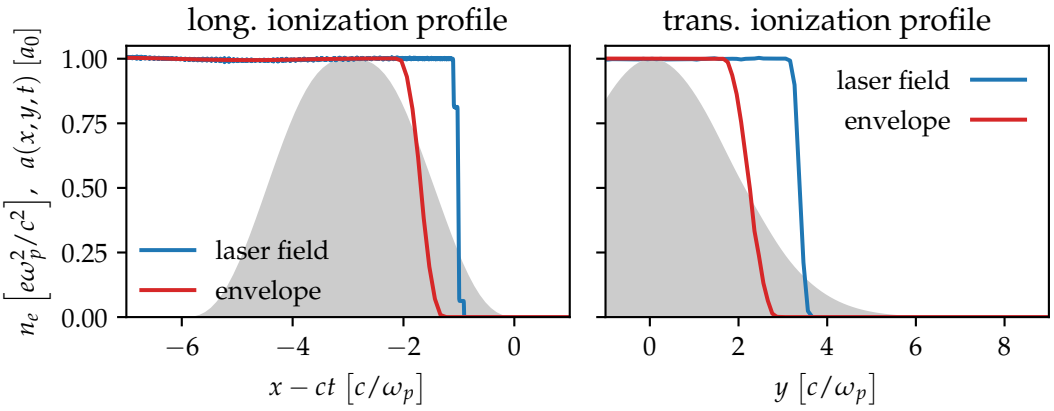


Figure 5.2: Comparison of ionization profiles using a full description of the laser pulse and the envelope. Shown are lineouts of the ionization profile in longitudinal direction (left) and transverse direction (right) for the case of rubidium vapor being ionized. The blue line represents a model where the full laser field was used. While for the red line, ionization rates for an envelope as characterized by (5.3) are used. The gray area shows the envelope of the laser pulse for both cases.

0.1, a frequency of 30, a pulse width of 2.5, and a pulse duration of 3.0 is considered. With the described laser parameters, a pulse is propagated through a rubidium vapor that will only ionize the rubidium until the first level. In a direct comparison between both approaches, an agreement in the level of ionization is verified. Though, the ionization rates for the envelope model are generally lower [119] leading to a narrow ionization profile, as shown in Fig. 5.2. For the study of the self-modulation instability, this results in a smaller volume of the electron column. However, for the development of the SMI, this reduced ionization profiles, especially in the longitudinal direction, will result in a small phase difference in the longitudinal direction. It has to be pointed out, that the difference is below 10 % and is purely relative, meaning the seeding will occur slightly delayed. On the other hand, the ionization front propagation speed is expected to have a higher contribution to the SMI. This effect is self-consistently modeled and is equal in both scenarios.

In the case of PGC, the generated plasma will only affect the laser pulse's tail characterized by the plasma susceptibility. The head of the pulse propagates like through vacuum and is not affected by the generated plasma. Though, the envelope equation can be extended to include additional terms to characterize the dispersion in rubidium or nonlinear self-focusing [120, 121]. Moreover, models that consider the nonlinear optical response of the Rubidium atoms [122] are promising but require extensive work to be generalized so that it can be incorporated into the PIC algorithm. However, the development of the SMI is determined by the seed and its propagation velocity. Assuming a linear slip back of the ionization front, then the relativistic factor is given as

$$\gamma_{\mathrm{if}} = \frac{1}{\sqrt{1-(1+\Delta\xi/\Delta d)^2}} \tag{5.4}$$

where $\Delta\xi$ is the slip back distance in the light-frame coordinates and Δd the propagation distance. With (5.4) and assuming a slip back on the order of a plasma wavelength over 10 m propagation distance would result in $\gamma_{\mathrm{if}} = 50$. For an ionization front propagating below this relativistic factor, we would obtain a phase mixing for the seed.

5.2 Simulation parameters and modeled scenario

Modeling of the SMI self-consistently with PGC consists of three independent components, the proton beam, the laser, and the neutral gas vapor. In this section, the parameters used in this chapter, are presented and discussed. The main goal is to stay as close as possible to the realistic parameters available at the AWAKE experiment. Though shot-to-shot fluctuations may be present in the experiment, such fluctuations are not considered.

In general, three spatial dimensions are considered and a moving window is used. The simulation box has the dimensions of 52.5 mm $\times$ 9.0 mm $\times$ 9.0 mm with

1656 cells in longitudinal direction and 640 cells in each transverse direction; the longitudinal resolution is $\Delta\xi = 31.7\,\mu\text{m}$ and the transverse resolution is $\Delta y = \Delta z = 14.1\,\mu\text{m}$. The time step is $\Delta\tau = 31.6\,\text{fs}$ and the total propagation time corresponds to the 10 m of the gas cell in the AWAKE experiment.

In these studies a fraction of the full SPS proton beams is considered. The profile of the beam in the light frame coordinates ($\xi = x - ct, \tau = t$) is determined as

$$n_b(\xi, r) = \begin{cases} n_{b,0}\left[1 + \cos\left(\sqrt{\frac{\pi}{2\sigma_x^2}}\left(\xi - \xi_h\right)\right)\right] \exp\left(\frac{r^2}{2\sigma_r^2}\right), & \xi_t \leq \xi \leq \xi_h \\ 0, & \text{otherwise} \end{cases} \tag{5.5}$$

with the beam radius $r = \sqrt{y^2 + z^2}$, the beam density $n_{b,0} = 4.6 \times 10^{12}\,\text{cm}^{-3}$, the rms bunch length $\sigma_x = 6.0\,\text{cm}$, and the rms bunch width $\sigma_r = 159.5\,\mu\text{m}$. Here, the beam is cut at the front which is determined by the head of the beam $\xi_h = 0.0\,\text{mm}$ and at the tail of the beam $\xi_t = -50.2\,\text{mm}$. This parameters are picked so, that the beam does not touch the boundaries of the box. Moreover, the relativistic factor of the beam is $\gamma = 480$.

The gas vapor has a simple step profile which is 10 m long and has a density of $2 \times 10^{14}\,\text{cm}^{-3}$. It is only used for the generation of the electron column based on the ionization rate. Gas kinematics are not considered. In addition, Rubidium absorption lines are also not included in our model. The proton beam and the laser are co-propagating through the vapor but are initialized outside.

The laser has a normalized vector potential of $a_0 = 1.02 \times 10^{-2}$, a spot size of 1 mm, and a mean laser wavelength of 0.8 µm. These parameters are associated with parameters obtained in the AWAKE experiment [123]. For pulse duration, a value of 12.5 ps was picked. Here, the laser pulse duration is artificially increased compared to the AWAKE experiment to allow lower longitudinal resolution. For this particular modeling, an envelope approach is required as $\omega_0/\omega_p = 3000$, and modeling using full PIC would be at least six orders of magnitude more computa-

tionally expensive. The focal plane of the laser is at the midpoint of the gas vapor. This focal plane position ensures that a plasma column is generated over the whole propagation distance. Also, the laser center is shifted backward by 3.7 mm compared to the head of the proton beam. This additional shift allows a clear separation of the instability's seeding, leaving 1.9 mm of the proton beam unperturbed.

5.3 Results for seeding the self-modulation instability using a relativistic ionization front

Fig. 5.3 shows the evolution of the axial electric fields, and SMI in the laser-ionized rubidium vapor plasma. During the initial propagation, the particle bunch self-modulates, and the longitudinal fields grow exponentially, known as the linear stage of the SMI [51, 124, 125]. The initial bunching can be seen clearly after 3.6 m of propagation, shown with Fig. 5.3(II). After 5 m of propagation the non-linear stage of the SMI is reached. It is characterized by the absence of further growth of the axial field. This transition to a non-linear stage is also associated with a phase shift of the axial field. While currently no theoretical model is found to describe the underlying physics, it is believed that the phase shift arises due to the superposition of the wakes generated by each beamlet [126].

On the other hand, comparing the electron density profile throughout the propagation reveals a narrow distribution around the focal plane and broader distribution at the beginning and the end of the propagation. This narrowing is connected to the laser propagation obeying the paraxial Helmholtz equation for propagation in a vacuum. It is important to understand if the evolution of the electron column and its description by the laser envelope have any effect on the growth and characteristic of the SMI. This question will be addressed in the following section.

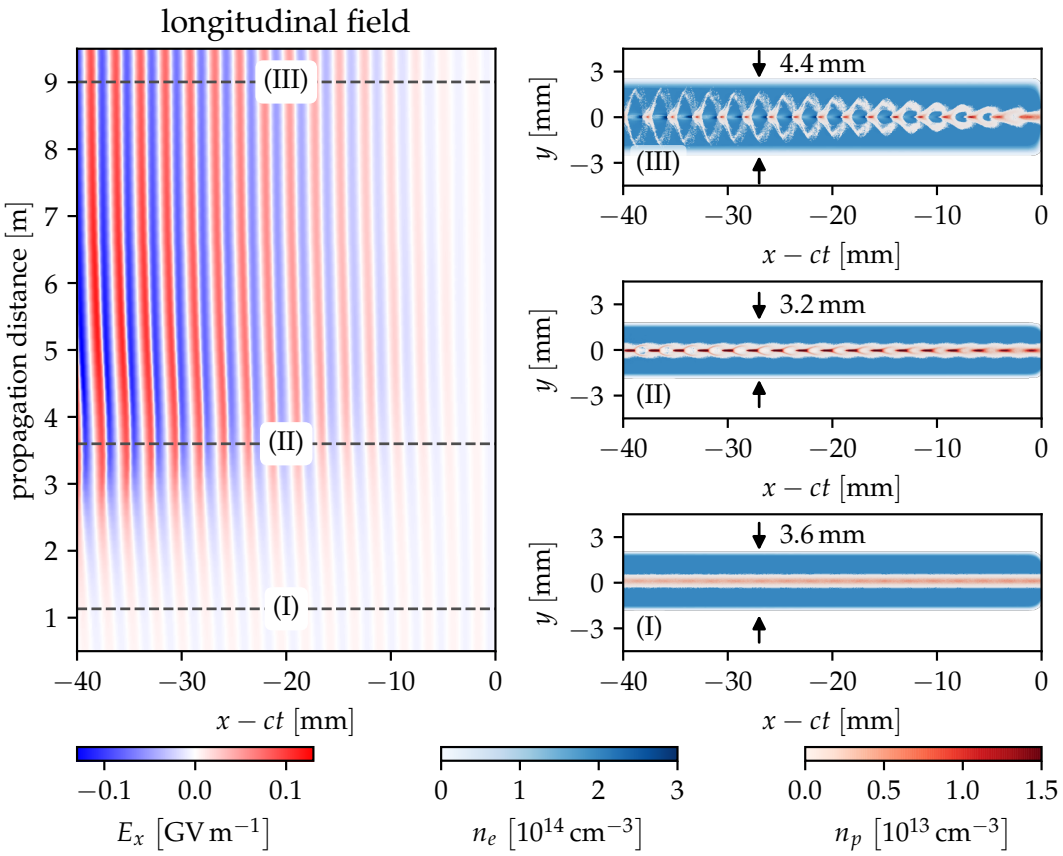


Figure 5.3: Full evolution of the self-consistent longitudinal plasma field along the protons beam center axis over the 10 m propagation distance (left panel) with three different propagation distances, slices of the ionized plasma column and the proton beam (right panels).

5.4 Study of the effect of ionization on the self-modulation instability

For establishing, how the ionization front can influence the growth of the SMI, a comparison against a pre-ionized case can be examined. Here, we consider the same physical parameters as in the previous section except for an ionizing laser pulse. The seed of the SMI is associated with the sharp front of the half-cut proton bunch propagating through a fixed-sized plasma column. The cut in the proton profile behaves similarly to the laser pulse; it seeds the growth of the SMI instability. There are two distinct differences. The first difference is the absence of any transverse effects due to a fixed-sized plasma column. The second difference is that the seed of the SMI propagates with a constant velocity compared to the ionization front in the previous example. For the pre-ionized case, the velocity of the seed is determined by the Lorentz gamma of the proton bunch.

Based on the pre-requisites, the case presented and discussed in the previous section is used as a base setup. For the pre-ionized case, a plasma column with a radius of 1.5 mm is added, and all other parameters are kept unchanged. The simulation is performed in 2D cylindrical geometries to reduce the computational cost, and the difference for the accelerating field at two distances is shown in Fig. 5.4. For better comparison, the case where laser ionization is used is shifted such that the tip of the proton beam and the ionization front overlap in the initial stages.

For the pre-ionized case, the seed of the SMI is determined by the half-cut front of the proton beam. Therefore, the propagation of the seeding front depends on the Lorentz gamma of the proton beam, which was $\gamma = 480$. Throughout the propagation, no significant slowdown of the proton bunch was observed. Complementary to this, the case for a laser induced ionization front exhibits different propagation velocities for the ionization front. While in the first stage, during the propagation towards the focal plane, the ionization front propagates superluminal, as seen in Figure 5.4. This effect of superluminal propagation arises due to the locally in-

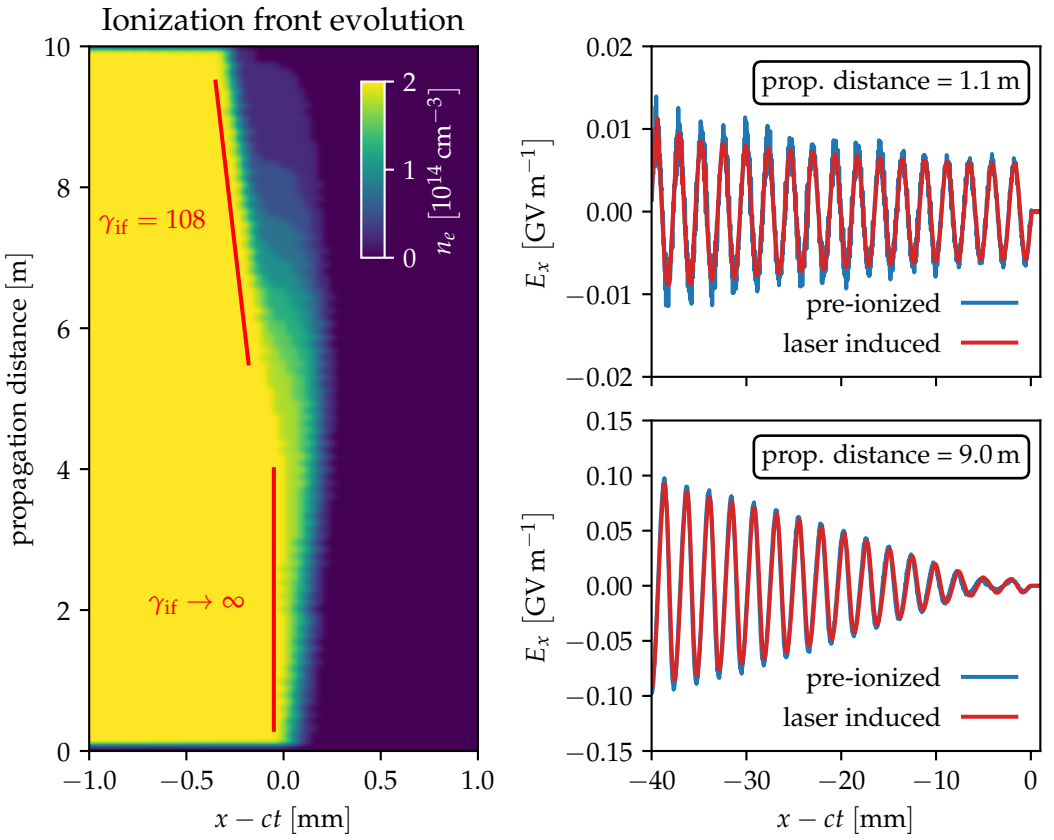


Figure 5.4: Evolution of the ionization front (left plot) and comparision of the axial fields for the cases with a pre-ionized plasma and a self-consistent ionization front (right plots). For better comparision, the ionization front was shifted by 2.7 mm to overlap with the cut proton bunch.

creased field amplitude. Afterwards, as the laser pulse defocuses, the ionization front propagates at $\gamma_{\text{if}} = 108$.

Comparing the axial fields, some fluctuations of the field after 1.1 m are observed. It has to be noted that the proton beam is sharply cut in the pre-ionized case, while for the laser case, the ionization shows a smoother density transition. However, both cases show the same phase. After a long propagation of 9.0 m, when the SMI fully developed and moved to a non-linear regime, no significant difference in phase and amplitude for the axial field is observed.

In summary, no significant differences in seeding the SMI between using a half-cut proton beam or a laser-induced ionization profile are seen even though the ionization front evolves over the whole propagation distance. For the second half, the ionization front is even propagating slower than the proton beam; over the whole propagation distance, the ionization front slipped back only by 0.3 mm which is only a fraction of the whole plasma wavelength.

5.5 Transverse effects on the development of the self-modulation instability

A self-consistent model for the generation of a plasma column allows us to model transverse effects more closely and study its influence on the development of the SMI. Here, two particular effects are discussed in this section (i) the effect of a shift between the proton beam center axis and the laser axis and (ii) the effect of a seeded hosing instability (HI). The latter is important to consider since the SMI and HI have comparable growth rates [127], and HI can cause a beam-breakup.

For the first study, the laser axis is shifted downwards along the y-axes with respect to the proton bunch's central beam axis. In the study, a shift of 0.32 mm is considered. Greater shifts, e.g., on the order of $\sim$1 mm, might result in additional issues arising from the edging of the proton beam and are not discussed

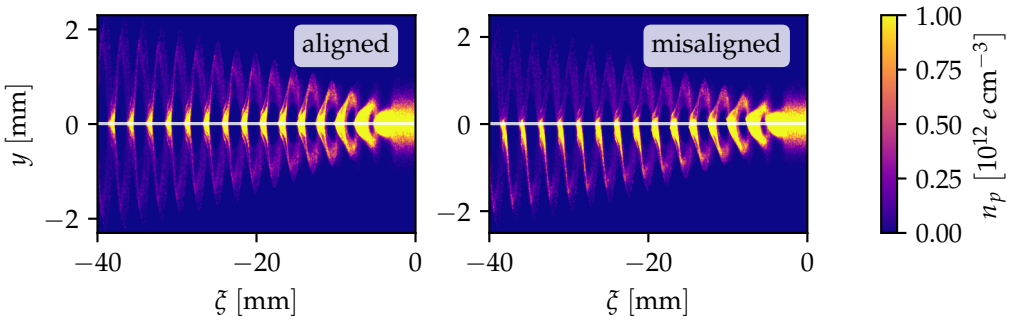


Figure 5.5: Comparison of proton density after 9.5 m of propagation for a case where the proton axis and the laser axis are aligned (left) and misaligned (right). In the misalignment case, the laser pulse is located 0.32 mm below the proton beam axis. In both scenarios, the white line represents the central beam axis.

here. Fig. 5.5 shows the perturbed proton bunch after 9.5 m of propagation. Only after a full propagation distance, a transverse distortion of the proton beam is observed. Before that, no clear difference with respect to the base case presented in the previous sections is not noticeable. After 9.5 m, the protons tend to be dragged towards the laser axis – the electron column center. Nevertheless, after the whole propagation distance, the overall morphology of the beams are very similar.

A slight decrease in the axial field is seen along the central beam axis in direct comparison. The reduction of the axial field is expected as a fraction of the driving bunch is being dragged downwards and reduces the wake driven by the beamlets. However, the transverse field in the misaligned case is more substantial, leading to enhanced transverse deflection of the proton beam driver.

The presented scenario of misalignment shows that the development of the SMI is stable enough even to withstand small misalignment and in the presented case, no evidence of a hosing instability is seen. However, the asymmetric beamlets, as seen in figure 5.5, introduce a shift of the beam centroid and further propagation (e.g., the second stage), which could result in the growth of hosing.

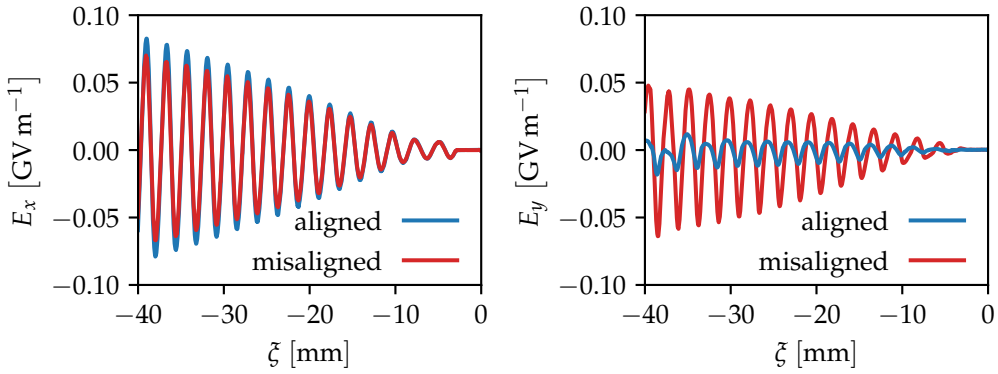


Figure 5.6: This comparison shows the axial field (left) and the E_y field (right) driven by the proton bunch for the aligned and the misaligned scenario. The presented fields correspond to the proton density profile shown in Figure 5.5.

For the second study, the beam centroid is shifted downwards along the y-axis to seed a hosing instability. More specifically, the shift is picked in such a way that the beam at the ionization front is aligned with the laser axis, while the tail for the 50 mm long simulated beam is shifted by the beam width σ_r of 0.2 mm, as seen in figure 5.7(b). The reasonably large tilt in the beam profile ensures strong seeding for the HI as only a fraction of the protons is aligned with the wake generated by the head of the beam.

As the beam propagates through the plasma, it self-modulates. The focusing fields cause a fraction of the protons to be dragged towards the center of the beam's wake. These protons result in a train of bunches aligned with the wake driven by the front. On the other hand, another fraction of the proton bunch is pushed away due to defocusing fields. As the SMI saturates, no substantial transverse displacements are observed, and the beam continues to propagate stably, as seen in Figure 5.7(c). These results agree with the previous work by [128], which showed hosing suppression for a self-modulated electron bunch without ionization modeling. More specifically, the frequency associated with the centroid displacement

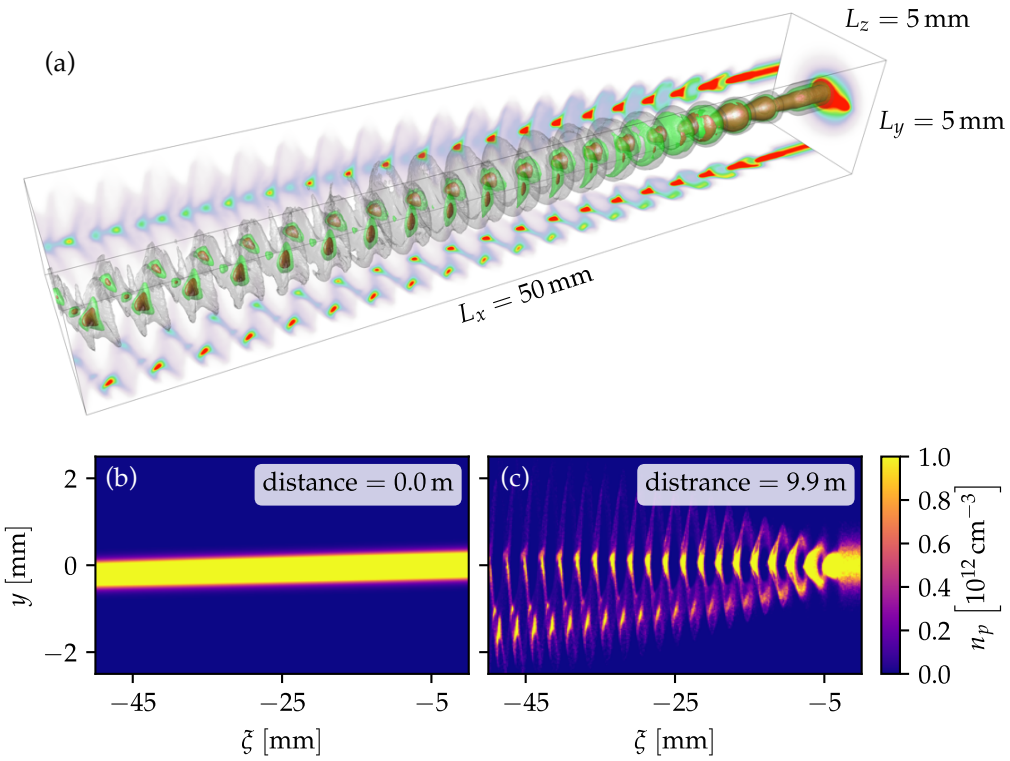


Figure 5.7: Density distribution of the beam driver for the study of seeded hosing instability. The upper figure (a) shows isosurfaces and density projections of the proton bunch after 9.9 m of propagation distance. The lower figures show slices of the beam density at the initial state (a) and after 9.9 m of propagation distance (b).

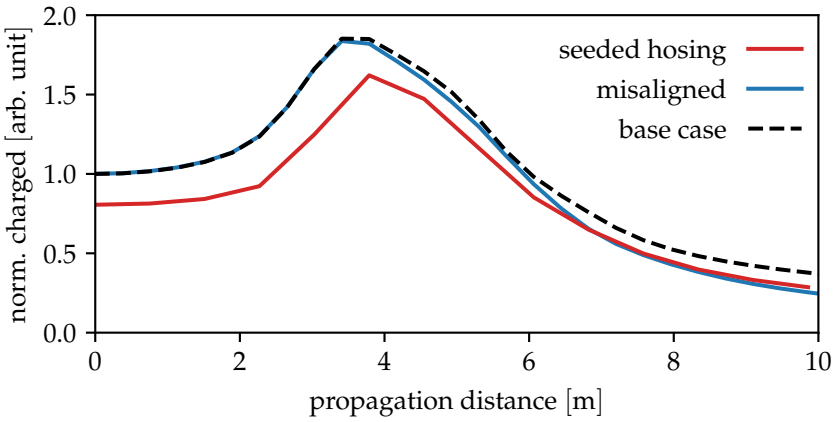


Figure 5.8: Comparision of the normalized beam driver charge over the propagation distance for the base case (black dashed line), for case of misalignment between the laser and beam axis (blue solid line), and for the case of seeded hosing (red solid line). The normalization factor is the integrated charge for the base case before the proton beam enters the gas vapour.

varies for each beamlet in a modulated bunch. As each beamlet resides in the focusing region, the resonant condition for hosing to occur is not fulfilled. Hence, the growth of HI is suppressed.

To further analyze the hosing suppression, the charge of the driver beam is compared for three cases. The first case is based on the results presented in section 5.2. In contrast, the two other scenarios correspond to the misaligned scenario and the seeded hosing case presented in this section. The driver charge is obtained by integrating a cylindrical region of the proton density where the center axis is aligned with the wake driven by the beam's head, and the radius for the integration is $r = \sigma_r/2 = 100\,\mu\text{m}$. The normalized charge of the driver is shown in figure 5.8. The charge for all three cases is normalized to the base case's initial charge in the integration region.

Comparing the charge evolution of each driver, two different progressions with a transition around 4 m are observed. The transition is associated with the saturation

of the SMI. In the first phase, the driver's charge increases, caused by the growth of the SMI. Afterward, the beam charge in the integration region drops with the propagation distance d. This drop comes from the saturation of the SMI and the associated phase transition for the driven wakefield. The phase transition yields radial defocusing of the beam driver.

Comparing the base case with the misaligned case, overlap with the base case in terms of the normalized charge is seen. However, as the protons get defocused, the plasma column's narrowness starts to play a role, and a fraction of the beam charge is pushed towards the laser axis and the associated center of the plasma column, which leads to a h igher reduction of the beam driver charge.

On the other side, comparing the base case with the seeded hosing case, a similar evolution of the beam driver charge is observed. However, the proton beam's tilt introduces a lower charge as the considered region aligns with the base case. As previously described, HI is suppressed for a self-modulated proton bunch resulting in a comparable evolution for the charge. Nevertheless, as only a fraction of the beam is driving the wake, the saturation of the SMI occurs later compared to the base case.

5.6 Conclusion

In this chapter, the modeling of self-consistent seeding of the self-modulation instability (SMI) was presented. More specifically, laser-ionization induced seeding of the SMI was shown. Due to a scale disparity between the laser wavelength and the plasma skin depth, the only computationally feasible approach is to model only the plasma scales. Considering only the plasma scales comes with the downside of excluding the laser pulse evolution. As shown in this chapter, the PGC solver with ionization can model the laser's evolution on the plasma scales.

As presented in this chapter, the ionization modeling using averaged ionization rates gives a self-consistent ionization profile to seed the SMI. Comparing it with a

model where ionization is excluded, but the SMI is seeded through a sharp density jump shows excellent agreement even though the seed in the ionization case is not static in the light-frame coordinates. However, the difference of the dynamic ionization front is only a fraction of the plasma wavelength, which does not introduce phase mixing or phase shifting to significantly modify the seed and the growth of the SMI. Hence, good agreement for both cases can be seen.

With a self-consistent ionization model, transverse effects can be included without further assumptions. As the growth rate of hosing instability (HI) is similar to the growth rate SMI, the inclusion of transverse effects from the plasma channel is essential. Here, two cases have been discussed. For the case of possible misalignment between the proton driver and the laser, the growth of the SMI was unperturbed. However, as the proton beamlets defocus post SMI saturation, an additional drag towards the laser axis was observed. Complementary to it, a seeded hosing case was investigated where the proton beam profile was tilted. In agreement with previous theoretical work, SMI suppresses the HI. Moreover, only a fraction of the charge contributes to the growth of the SMI and delaying the saturation. This reduction is associated with the tilt of the profile.

Conclusion and Outlook

The main focus of this thesis was to develop a reduced numerical solver to face the main challenge of bridging the scale disparity in particle-in-cell codes. With it, challenges for designing and studying the future generation of particle accelerators becomes more feasible. It is clear that with higher speedup and reduced time-to-solution, the space of exploration becomes greater and allows us to push the scientific community to a greater extend.

Based on the ponderomotive guiding center approximation, the presented solver enables to model the propagation of a laser pulse inside a plasma with great computational benefits. As part of the thesis, 2D cartesian, 2D cylindrical, and 3D cartesian geometries were incorporated into OSIRIS. In this work, the numerical stability for the 3D case was discussed, which is commonly neglected for newer implementations. It was shown that the PGC solver is unconditionally unstable and can affect any results. However, an upper limit for the PGC solver was determined. For most LWFA scenarios, PGC can be used before the numerical errors become dominant. The developed methods to characterize the numerical stability can be extended towards a path where different discretization schemes can be considered, next to the one incorporated here. Particularly explicit schemes can be promising as their stability is determined by a CFL condition similar to a Maxwell solver. However, additional steps have to be considered when using explicit schemes. Spectral methods, on the other hand, could mitigate numerical instabilities. Nonetheless, spectral methods display limitations in parallel scalability and cause additional overhead, which decreases computational efficiency.

As presented, the reduced PGC solver can further be used to perform parametric studies. Compared to full PIC simulations, PGC and PIC showed agreement regarding the laser evolution and the wake profiles. On the other hand, lower numerical resolution affected the injected particle bunches significantly. While this can be overcome with an increased temporal or spatial resolution, it would lead to higher computational costs.

A regime where PGC stands out is for cases where the ratio between the laser wavelength and the plasma skin depth reaches large values, e.g., modeling of relativistic seeding for the development of the SMI. Here, an ionization model was presented, which allows self-consistent modeling. It was clearly shown that using a self-consistent ionization model does not cause any additional phase shift and does not affect the seed for the SMI. This study helps the AWAKE experiment as it presents that modeling the ionization is not apriori required to seed the SMI. However, the developed work provides the required tooling for future experiments in the AWAKE framework.

Due to the high computational savings, PGC will be a vital tool for tasks associated with high computational costs, e.g., generating simulation output for machine learning models. With the current progress of developing data-driven discovery of governing equations [129, 130], PGC will provide valuable insights for developing a full 3D non-linear description of the plasma waves, which is currently unavailable. Moreover, PGC allows having a clear separation between the laser field and the plasma fields.

Bibliography

[1] E. Rutherford, "The scattering of α and β particles by matter and the structure of the atom", Philosophical Magazine **92**, 379–398 (2012).

[2] A. Collaboration et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Physics Letters B **716**, 1–29 (2012).

[3] M. Silari, "Applications of particle accelerators in medicine", Radiation Protection Dosimetry **146**, 440–450 (2011).

[4] J. F. Gibbons, "Ion Implantation in Semiconductors—Part I Range Distribution Theory and Experiments", Proceedings of the IEEE **56**, 295–319 (1968).

[5] D. H. Bilderback et al., "Review of third and next generation synchrotron light sources", Journal of Physics B **38**, 773–797 (2005).

[6] CERN, Facts and figures about the LHC .

[7] J. D. Cockcroft and E. T. S. Walton, "Experiments with High Velocity Positive Ions. (I) Further Developments in the Method of Obtaining High Velocity Positive Ions", Proceedings of the Royal Society **A136**, 619–630 (1932).

[8] J. D. Cockcroft and E. T. S. Walton, "Experiments with high velocity positive ions; 2, The disintegration of elements by high velocity protons", Proceedings of the Royal Society A **137**, 229–242 (1932).

[9] G. Ising, "Prinzip einer Methode zur Herstellung von Kanalstrahlen hoher Voltzahl", Arkiv för matematik, astronomi och fysik **18**, 1–4 (1924).

[10] H. H. Braun et al., "Frequency and Temperature Dependence of Electrical Breakdown at 21, 30, and 39GHz", Physical Review Letters **90**, 224801 (2003).

[11] G. Bachy et al., "The LEP collider: construction, project status and outlook", Physical Review Accelerators and Beams **26**, 19–32. 15 p (1989).

[12] J. W. Cronin, Fermi remembered (University of Chicago Press, Chicago, IL, 2004).

[13] P. Bambade et al., "The International Linear Collider: A Global Project", arXiv (2019).

[14] K. Saito, "Critical field limitation of the niobium superconducting RF cavity", 10th Workshop on RF Superconductivity, 583–587 (2001).

[15] E. A. Peralta et al., "Demonstration of electron acceleration in a laser-driven dielectric microstructure", Nature **503**, 1–11 (2013).

[16] R. B. Palmer, "Interaction of Relativistic Particles and Free Electromagnetic Waves in the Presence of a Static Helical Magnet", Journal of Applied Physics **43**, 3014–3023 (1972).

[17] T. Tajima and J. M. Dawson, "Laser Electron Accelerator", Physical Review Letters **43**, 267–270 (1979).

[18] P. Chen et al., "Acceleration of Electrons by the Interaction of a Bunched Electron-Beam with a Plasma", Physical Review Letters **54**, 693–696 (1985).

[19] D. Strickland and G. Mourou, "Compression of amplified chirped optical pulses", Optics Communications **56**, 219–221 (1985).

[20] A. Vogel and V. Venugopalan, "Mechanisms of pulsed laser ablation of biological tissues", Chemical Reviews **103**, 577–644 (2003).

[21] D. V. Palanker et al., "Femtosecond laser-assisted cataract surgery with integrated optical coherence tomography", Science Translational Medicine **2**, 58ra85 (2010).

[22] D. Z. Reinstein et al., "Small incision lenticule extraction (SMILE) history, fundamentals of a new refractive surgery technique and clinical outcomes", Eye and Vis **1**, 467 (2014).

[23] M. D. Shirk and P. A. Molian, "A review of ultrashort pulsed laser ablation of materials", Journal of Laser Applications **10**, 18–28 (1998).

[24] A. Modena et al., "Electron Acceleration From the Breaking of Relativistic Plasma-Waves", Nature **377**, 606–608 (1995).

[25] K. Nakajima et al., "Observation of Ultrahigh Gradient Electron Acceleration by a Self-Modulated Intense Short Laser Pulse", Physical Review Letters **74**, 4428–4431 (1995).

[26] D. Umstadter et al., "Nonlinear optics in relativistic plasmas and laser wake field acceleration of electrons", Science **273**, 472–475 (1996).

[27] A. Ting et al., "Plasma wakefield generation and electron acceleration in a self-modulated laser wakefield accelerator experiment", Physics of Plasmas **4**, 1889–1899 (1997).

[28] C. Gahn et al., "Multi-MeV electron beam generation by direct laser acceleration in high-density plasma channels", Physical Review Letters **83**, 4772–4775 (1999).

[29] W. P. Leemans et al., "Electron-Yield Enhancement in a Laser-Wakefield Accelerator Driven by Asymmetric Laser Pulses", Physical Review Letters **89**, 2049–4 (2002).

[30] V. Malka, "Electron Acceleration by a Wake Field Forced by an Intense Ultrashort Laser Pulse", Science **298**, 1596–1600 (2002).

[31] J. Faure et al., "A laser-plasma accelerator producing monoenergetic electron beams", Nature **431**, 541–544 (2004).

[32] C. G. R. Geddes et al., "High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding", Nature **431**, 538–541 (2004).

[33] S. Mangles et al., "Monoenergetic beams of relativistic electrons from intense laser-plasma interactions", Nature **431**, 535–538 (2004).

[34] R. A. Fonseca et al., "Osiris: a three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators", in Computational science — iccs 2002 (2002), pp. 342–351.

[35] W. B. Mori, "The physics of the nonlinear optics of plasmas at relativistic intensities for short-pulse lasers", IEEE Journal of Quantum Electronics **33**, 1942–1953 (1997).

[36] L. M. Gorbunov and V. I. Kirsanov, "Excitation of plasma waves by an electromagnetic wave packet", Journal of Experimental and Theoretical Physics **66**, 290 (1987).

[37] P. Sprangle et al., "Laser wakefield acceleration and relativistic optical guiding", Applied Physics Letters **53**, 2146–2148 (1988).

[38] E. Esarey et al., "Optically guided laser wake-field acceleration", Physics of Fluids B **5**, 2690–2697 (1993).

[39] S. V. Bulanov et al., "Nonlinear depletion of ultrashort and relativistically strong laser pulses in an underdense plasma", Physics of Fluids B **4**, 1935–1942 (1992).

[40] V. I. Berezhiani and I. G. Murusidze, "Relativistic wake-field generation by an intense laser pulse in a plasma", Physics Letters A **148**, 338–340 (1990).

[41] P. Sprangle et al., "Nonlinear interaction of intense laser pulses in plasmas", Physical Review A **41**, 4463–4469 (1990).

[42] A. Pukhov and J. Meyer-Ter-Vehn, "Laser wake field acceleration: the highly non-linear broken-wave regime", Applied Physics B **74**, 355–361 (2002).

[43] A. Pak et al., "Injection and Trapping of Tunnel-Ionized Electrons into Laser-Produced Wakes", Physical Review Letters **104**, 1191 (2010).

[44] J. Vieira et al., "Magnetically assisted self-injection and radiation generation for plasma-based acceleration", Plasma Physics and Controlled Fusion **54**, 124044 (2012).

[45] S. Kalmykov et al., "Electron Self-Injection and Trapping into an Evolving Plasma Bubble", Physical Review Letters **103**, 135004 (2009).

[46] S. Steinke et al., "Multistage coupling of independent laser-plasma accelerators", Nature **530**, 190–193 (2016).

[47] I. Blumenfeld et al., "Energy doubling of 42GeV electrons in a metre-scale plasma wakefield accelerator", Nature **445**, 741–744 (2007).

[48] R. D. Ruth et al., "A plasma wake field accelerator", Particle Accelerators **17**, 171–189 (1985).

[49] E. Adli et al., "Acceleration of electrons in the plasma wakefield of a proton bunch", Nature **561**, 363–367 (2018).

[50] A. Caldwell et al., "Proton-driven plasma-wakefield acceleration", Nature Physics **5**, 363–367 (2009).

[51] A. Pukhov et al., "Phase Velocity and Particle Injection in a Self-Modulated Proton-Driven Plasma Wakefield Accelerator", Physical Review Letters **107**, 145003–5 (2011).

[52] R. A. Bamford et al., "3D PIC simulations of collisionless shocks at lunar magnetic anomalies and their role in forming lunar swirls", Astrophysical Journal **830**, 146 (2016).

[53] F. Cruz et al., "Formation of collisionless shocks in magnetized plasma interaction with kinetic-scale obstacles", Physics of Plasmas **24**, 022901–15 (2017).

[54] A. Rigby et al., "Electron acceleration by wave turbulence in a magnetized plasma", Nature Physics **14**, 475–479 (2018).

[55] E. P. Alves et al., "Efficient Nonthermal Particle Acceleration by the Kink Instability in Relativistic Jets", Physical Review Letters **121**, 245101 (2018).

[56] A. Pukhov and J. Meyer-Ter-Vehn, "Relativistic magnetic self-channeling of light in near-critical plasma: Three-dimensional particle-in-cell simulation", Physical Review Letters **76**, 3975–3978 (1996).

[57] K. C. Tzeng et al., "Anomalous absorption and scattering of short-pulse high-intensity lasers in underdense plasmas", Physical Review Letters **76**, 3332–3335 (1996).

[58] T. Kluge et al., "High proton energies from cone targets: electron acceleration mechanisms", New Journal of Physics **14**, 023038 (2012).

[59] J. P. Boris, "Relativistic plasma simulation-optimization of a hybrid code", in Proceedings 4th conference on numerical simulation of plasmas (1970), pp. 3–67.

[60] C. D. Decker et al., "Particle-in-cell simulations of Raman forward scattering from short-pulse high-intensity lasers", Physical Review E **50**, R3338–R3341 (1994).

[61] J. L. Vay, "Noninvariance of space- and time-scale ranges under a Lorentz transformation and the implications for the study of relativistic interactions", Physical Review Letters **98**, 130405 (2007).

[62] S. F. Martins et al., "Modeling laser wakefield accelerator experiments with ultrafast particle-in-cell simulations in boosted frames", Physics of Plasmas **17**, 056705 (2010).

[63] J. L. Vay et al., "Effects of hyperbolic rotation in Minkowski space on the modeling of plasma accelerators in a Lorentz boosted frame", Physics of Plasmas **18**, 030701–5 (2011).

[64] X. Xu et al., "Numerical instability due to relativistic plasma drift in EM-PIC simulations", Computer Physics Communications **184**, 2503–2514 (2013).

[65] P. Yu et al., "Modeling of laser wakefield acceleration in Lorentz boosted frame using EM-PIC code with spectral solver", Journal of Computational Physics **266**, 124–138 (2014).

[66] B. B. Godfrey and J.-L. Vay, "Numerical stability of relativistic beam multi-dimensional PIC simulations employing the Esirkepov algorithm", Journal of Computational Physics **248**, 33–46 (2013).

[67] A. Davidson et al., "Implementation of a hybrid particle code with a PIC description in r-z and a gridless description in ϕ into OSIRIS", Journal of Computational Physics **281**, 1063–1077 (2015).

[68] P. Mora and T. M. Antonsen Jr, "Kinetic modeling of intense, short laser pulses propagating in tenuous plasmas", Physics of Plasmas **4**, 217 (1997).

[69] D. F. Gordon et al., "A ponderomotive guiding center particle-in-cell code for efficient modeling of laser-plasma interactions", IEEE Transactions on Plasma Science **28**, 1135–1143 (2000).

[70] C. Benedetti et al., "An accurate and efficient laser-envelope solver for the modeling of laser-plasma accelerators", Plasma Physics and Controlled Fusion **60**, 014002–8 (2017).

[71] D. Terzani and P. Londrillo, "A fast and accurate numerical implementation of the envelope model for laser–plasma dynamics", Computer Physics Communications **242**, 49–59 (2019).

[72] D. F. Gordon, "Improved Ponderomotive Guiding Center Algorithm", IEEE Transactions on Plasma Science **35**, 1486–1488 (2007).

[73] B. M. Cowan et al., "Characteristics of an envelope model for laser–plasma accelerator simulation", Journal of Computational Physics **230**, 61–86 (2011).

[74] E. Anderson et al., LAPACK Users' Guide, Third Edition (SIAM, Jan. 1999).

[75] W. H. Press et al., Numerical Recipes 3rd Edition, The Art of Scientific Computing (Cambridge University Press, Sept. 2007).

[76] F. S. Tsung et al., "Simulation of monoenergetic electron generation via laser wakefield accelerators for 5-25 TW lasers", in Physics of plasmas (University of California, Los Angeles, Los Angeles, United States, May 2006), p. 056708.

[77] W. Hazeleger et al., "Tales of future weather", Nature Climate Change **5**, 107–113 (2015).

[78] F. Vitart et al., Extended-range Prediction (2019).

[79] H. Carr, "ECMWF press kit: Bologna to host ECMWF's new data centre", 1–7 (2017).

[80] B. Rose and E. Floehr, "Analysis of High Temperature Forecast Accuracy of Consumer Weather Forecasts from 2005-2016", 1–15 (2019).

[81] O. Peckham, "WCRP's New Strategic Plan for Climate Research Highlights the Importance of HPC", HPCwire (2018).

[82] K. Matouš et al., "A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials", Journal of Computational Physics **330**, 192–220 (2017).

[83] M. Raissi et al., "Machine learning of linear differential equations using Gaussian processes", Journal of Computational Physics **348**, 683–693 (2017).

[84] B. O'Brien et al., "Leveraging HPC resources for high energy physics", in Journal of physics: conference series (Ludwig-Maximilians-Universität München, Munich, Germany, Jan. 2014), p. 032104.

[85] R. L. Jiang et al., "A new MHD code with adaptive mesh refinement and parallelization for astrophysics", Computer Physics Communications **183**, 1617–1633 (2012).

[86] G. E. Moore, "Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff.", IEEE Solid-State Circuits Society Newsletter **11**, 33–35 (2008).

[87] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities", in Afips conference proceedings - 1967 spring joint computer conference, afips 1967 (International Business Machines, Armonk, United States, Apr. 1967), pp. 483–485.

[88] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3", in Proceedings of the ieee (2005), pp. 216–231.

[89] R. A. Fonseca et al., "Exploiting multi-scale parallelism for large scale numerical modelling of laser wakefield accelerators", Plasma Physics and Controlled Fusion **55**, 124011–10 (2013).

[90] M. J. Hogan et al., "Multi-GeV Energy Gain in a Plasma-Wakefield Accelerator", Physical Review Letters **95**, 472–4 (2005).

[91] A. J. Gonsalves et al., "Petawatt Laser Guiding and Electron Beam Acceleration to 8 GeV in a Laser-Heated Capillary Discharge Waveguide", Physical Review Letters **122**, 084801 (2019).

[92] M. C. Downer et al., "Diagnostics for plasma-based electron accelerators", Reviews of Modern Physics **90**, 035002 (2018).

[93] B. W. J. McNeil and N. R. Thompson, "X-ray free-electron lasers", Nature Photonics **4**, 814–821 (2010).

[94] P. A. Walker et al., "Horizon 2020 EuPRAXIA design study", Journal of Physics: Conference Series **874**, 012029 (2017).

[95] B. J. Claessens et al., "Ultracold electron source", Physical Review Letters **95**, 164801 (2005).

[96] P. Tomassini et al., "High-quality 5 GeV electron bunches with resonant multi-pulse ionization injection", Plasma Physics and Controlled Fusion **62**, 014010 (2020).

[97] X. Li et al., "Toward low energy spread in plasma accelerators in quasilinear regime", Physical Review Accelerators and Beams **21**, 111301 (2018).

[98] X. Li et al., "Design of a 5 GeV laser–plasma accelerating module in the quasi-linear regime", Nuclear Instruments and Methods in Physics Research Section A **909**, 49–53 (2018).

[99] A. F. Pousa et al., "Compact Multistage Plasma-Based Accelerator Design for Correlated Energy Spread Compensation", Physical Review Letters **123**, 054801 (2019).

[100] A. Buck et al., "Shock-Front Injector for High-Quality Laser-Plasma Acceleration", Physical Review Letters **110**, 185006–185005 (2013).

[101] H. Ekerfelt et al., "A tunable electron beam source using trapping of electrons in a density down-ramp in laser wakefield acceleration", Scientific Reports **7**, 1–9 (2017).

[102] X. L. Xu et al., "High quality electron bunch generation using a longitudinal density-tailored plasma-based accelerator in the three-dimensional blowout regime", Physical Review Accelerators and Beams **20**, 111303–8 (2017).

[103] W. Lu et al., "Generating multi-GeV electron bunches using single stage laser wakefield acceleration in a 3D nonlinear regime", Physical Review Accelerators and Beams **10**, 81 (2007).

[104] K. S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media", IEEE Transactions on Antennas and Propagation **14**, 302–307 (1966).

[105] M. F. Gilljohann et al., "Direct Observation of Plasma Waves and Dynamics Induced by Laser-Accelerated Electron Beams", Physical Review X **9**, 011046 (2019).

[106] O. Kononenko et al., "2D hydrodynamic simulations of a variable length gas target for density down-ramp injection of electrons into a laser wakefield accelerator", Nuclear Instruments and Methods in Physics Research Section A **829**, 125–129 (2016).

[107] J. J. Su et al., "Plasma Lenses for Focusing Particle Beams", Physical Review A **41**, 3321–3331 (1990).

[108] R. Lehe et al., "Numerical growth of emittance in simulations of laser-wakefield acceleration", Physical Review Accelerators and Beams **16**, 021301–8 (2013).

[109] M. Scisciò et al., "Parametric study of transport beam lines for electron beams accelerated by laser-plasma interaction", Journal of Applied Physics **119**, 094905 (2016).

[110] P. Muggli and M. J. Hogan, "Review of high-energy plasma wakefield experiments", Comptes Rendus Physique **10**, 116–129 (2009).

[111] M. J. Hogan et al., "Plasma wakefield acceleration experiments at FACET", New Journal of Physics **12**, 055030 (2010).

[112] M. Litos et al., "High-efficiency acceleration of an electron beam in a plasma wakefield accelerator", Nature **515**, 92–95 (2014).

[113] C. Joshi et al., "Plasma wakefield acceleration experiments at FACET II", Plasma Physics and Controlled Fusion **60**, 034001 (2018).

[114] D. F. Gordon et al., "Seeding of the forward Raman instability by ionization fronts and Raman backscatter", Physical Review E **64**, 046404 (2001).

[115] E. Gschwendtner et al., "AWAKE, The Advanced Proton Driven Plasma Wakefield Acceleration Experiment at CERN", Nuclear Instruments and Methods in Physics Research Section A **829**, 76–82 (2016).

[116] D. H. Whittum et al., "Electron-Hose Instability in the Ion-Focused Regime", Physical Review Letters **67**, 991–994 (1991).

[117] T. J. Mehrling et al., "Mitigation of the Hose Instability in Plasma-Wakefield Accelerators", Physical Review Letters **118**, 174801 (2017).

[118] M. V. Ammosov et al., "Tunnel ionization of complex atoms and of atomic ions in an altemating electromagnetic field", Journal of Experimental and Theoretical Physics **64**, 1191 (1986).

[119] M. Chen et al., "Numerical modeling of laser tunneling ionization in explicit particle-in-cell codes", Journal of Computational Physics **236**, 220–228 (2013).

[120] P. Sprangle et al., "Propagation of intense short laser pulses in the atmosphere", Physical Review E **66**, 73–21 (2002).

[121] J. R. Peñano et al., "Propagation of ultra-short, intense laser pulses in air", Physics of Plasmas **11**, 2865–2874 (2004).

[122] G. Demeter, "Propagation of ultrashort resonant ionizing laser pulses in rubidium vapor", Physical Review A **99**, 1–9 (2019).

[123] P. Muggli et al., "AWAKE readiness for the study of the seeded self-modulation of a 400 GeV proton bunch", Plasma Physics and Controlled Fusion **60**, 014046 (2018).

[124] N. Kumar et al., "Self-Modulation Instability of a Long Proton Bunch in Plasmas", Physical Review Letters **104**, 255003 (2010).

[125] C. B. Schroeder et al., "Growth and Phase Velocity of Self-Modulated Beam-Driven Plasma Waves", Physical Review Letters **107**, 145002 (2011).

[126] K. V. Lotov, "Physics of beam self-modulation in plasma wakefield accelerators", Physics of Plasmas **22**, 103110–11 (2015).

[127] C. B. Schroeder et al., "Coupled beam hose and self-modulation instabilities in overdense plasma", Physical Review E **86**, 026402 (2012).

[128] J. Vieira et al., "Hosing Instability Suppression in Self-Modulated Plasma Wakefields", Physical Review Letters **112**, 205001–5 (2014).

[129] K. Champion et al., "Data-driven discovery of coordinates and governing equations", Proceedings of the National Academy of Sciences **116**, 22445–22451 (2019).

[130] E. P. Alves and F. Fiuza, "Data-driven discovery of reduced plasma physics models from fully-kinetic simulations", arXiv (2020).