Bayesian Optimization of Laser-Wakefield Accelerators

Faran Irshad

München 2025

Bayesian Optimization of Laser-Wakefield Accelerators

Faran Irshad

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Faran Irshad
aus Karachi
München, den 08.09.2025

Erstgutachter: Prof. Dr. Stefan Karsch
Zweitgutachter: Dr. habil. Andreas Döpp
Tag der mündlichen Prüfung: 11.11.2025

In loving memory of my grandparents Chaudhry Muhammad Ismail, Shakoor Ahmed, Zubaida Begum, and Khurshid Begum, whose love and guidance shaped my life. You are deeply missed and forever remembered.

Zusammenfassung

Laser-Wakefield-Beschleuniger (LWFA) sind kompakte Teilchenbeschleuniger mit hohen Feldgradienten, die in der Lage sind, ultra-relativistische Elektronenstrahlen über Distanzen im Zentimeterbereich zu erzeugen. Durch das Anregen großamplitudiger Plasmaschwingungen mittels intensiver Laserpulse können LWFAs Beschleunigungsfelder erzeugen, die um mehrere Größenordnungen stärker sind als jene konventioneller Hochfrequenzbeschleuniger. Trotz ihres Potenzials für Anwendungen in Medizin, Industrie und der ultraschnellen Wissenschaft werden LWFAs derzeit durch Herausforderungen in Bezug auf Reproduzierbarkeit, Stabilität und Steuerbarkeit eingeschränkt. Ihre Leistung ist äußerst empfindlich gegenüber kleinsten Änderungen der Laser- und Plasmabedingungen und zeigt ein komplexes, nichtlineares Verhalten. Herkömmliche manuelle Abstimmung und heuristische Optimierungsstrategien reichen nicht aus, um den hochdimensionalen Parameterraum zuverlässig zu erkunden und mehrere konkurrierende Strahleigenschaften wie Energie, Ladung und spektrale Bandbreite gleichzeitig zu optimieren.

Diese Arbeit handelt diese Herausforderungen durch die Entwicklung und Implementierung von Optimierungsstrategien basierend auf Methoden des maschinellen Lernens, insbesondere der Bayesschen Optimierung (BO), zur systematischen Steuerung und Feinabstimmung von LWFAs. Nach einer Einführung in die theoretischen Grundlagen der Bayesschen Optimierung wird ein neuartiger Multi-Objective-Multi-Fidelity-Optimierungsrahmen namens Trust-MOMF vorgestellt. Diese Methode erlaubt eine stichprobeneffiziente Exploration komplexer, hochdimensionaler Parameterräume unter Verwendung von Datenquellen unterschiedlicher Genauigkeit. Im Gegensatz zu klassischen Ansätzen unterstützt diese Methode kontinuierliche Treuestufen (Fidelities), ohne dass ein monotoner Leistungszuwachs zwischen den Stufen vorausgesetzt werden muss, was sie besonders geeignet für Systeme mit verrauschten, teuren und inkonsistenten Messungen macht.

Die Leistungsfähigkeit von Trust-MOMF wird zunächst in numerischen Simulationen von LWFAs mit Hilfe von Particle-in-Cell-Codes demonstriert. Die Methode zeigt eine um eine Größenordnung schnellere Konvergenz und reduziert damit die Rechenkosten erheblich im Vergleich zu herkömmlichen multiobjektiven Optimierungsverfahren. Das während der Optimierung gelernte Surrogatmodell wird anschließend invertiert, um eine gezielte Einstellung der Elektronenstrahlenergie

zu ermöglichen und Zielkonflikte zwischen konkurrierenden Anforderungen wie Bandbreite und Effizienz zu untersuchen.

Im zweiten Teil der Arbeit wird das Verfahren auf reale LWFA-Experimente am ATLAS-Lasersystem angewendet, einem Laser der multi-PW-Leistungsklasse. Die Anzahl der Laserschüsse pro Konfiguration wird dabei als Fidelity-Variable interpretiert, wodurch die Optimierung den Informationsgewinn gegen die experimentellen Kosten abwägen kann. Erstmals wird in dieser Arbeit eine multiobjektive, multifidele bayessche Optimierung in LWFA-Experimenten demonstriert, wobei qualitativ hochwertige Elektronenstrahlen mit deutlich weniger Laserschüssen als bei manueller Optimierung erzielt werden. Darüber hinaus werden die trainierten Modelle für Feinabstimmungen und a-posteriori-Einzelzieloptimierungen eingesetzt, was eine reproduzierbare Abstimmung der Spitzenenergie im Bereich von 150 bis 400 MeV innerhalb eines achtdimensionalen Parameterraums ermöglicht.

Insgesamt etabliert diese Arbeit einen verallgemeinerbaren und effizienten Ansatz zur Steuerung von LWFAs und ermöglicht eine automatisierte, datengestützte Betriebsweise, die auf die Anforderungen spezifischer Nutzer zugeschnitten ist. Die vorgestellten Methoden stellen einen wichtigen Schritt in Richtung des praktischen Einsatzes von Laser-Plasma-Beschleunigern dar, insbesondere in Nutzerzentren, in denen eine flexible und schnelle Re-Optimierung von Strahleigenschaften erforderlich ist. Dies ebnet den Weg für Echtzeit-Abstimmungsstrategien, automatisierte Regelungen und die zukünftige Integration von LWFAs in wissenschaftliche Großforschungsanlagen.

Abstract

Laser Wakefield Accelerators (LWFAs) are compact, high-gradient particle accelerators capable of producing ultra-relativistic electron beams over centimeter-scale distances. By driving large amplitude plasma waves with intense laser pulses, LWFAs can achieve accelerating fields orders of magnitude greater than those in conventional radiofrequency accelerators. Despite their potential for applications in medicine, industry, and ultrafast science, LWFAs remain limited by challenges in reproducibility, stability, and control. Their performance is highly sensitive to small changes in laser and plasma conditions, and they exhibit complex, nonlinear behavior. Traditional manual tuning and heuristic optimization strategies are insufficient to reliably access the high-dimensional parameter space and optimize multiple competing beam metrics such as energy, charge, and bandwidth.

This thesis addresses these challenges by developing and implementing machine learning-based optimization strategies, specifically Bayesian optimization (BO), for the systematic control and tuning of LWFAs. After reviewing the theoretical foundations of Bayesian optimization, a novel multi-objective, multi-fidelity optimization framework termed Trust-MOMF is introduced. This method enables sample-efficient exploration of complex, high-dimensional parameter spaces using data sources with varying accuracy. Unlike traditional approaches, this method supports continuous fidelities without requiring monotonic improvement across fidelity levels, making it especially suitable for systems where measurements are noisy, expensive, and inconsistent.

The performance of Trust-MOMF is first demonstrated in numerical simulations of LWFAs using particle-in-cell codes. The method shows an order of magnitude improvement in convergence times, substantially reducing the computational cost compared to standard multi-objective optimization techniques. The surrogate model learned during the optimization process is then inverted to enable energy tuning of the accelerator and uncover trade-offs between competing objectives such as bandwidth and efficiency.

In the second part of the thesis, this framework is applied to real-world LWFA experiments using ATLAS which is a petawatt-class laser system. The number of laser shots per configuration is interpreted as a fidelity variable, allowing the optimization to balance information gain with experimental cost. For the first time, multi-objective multi-fidelity Bayesian optimization is demonstrated in LWFA experiments, yielding high-quality electron beams with significantly fewer laser shots. Furthermore, the trained models are used to perform fine-tuning and a posteriori single-objective optimization, enabling reproducible energy tuning between 150 and 400 MeV by navigating an 8-dimensional parameter space.

Together, these results establish a generalizable and efficient approach for controlling LWFAs, enabling automated, data-driven operation tailored to specific user requirements. This work represents a significant step toward deploying laser-plasma accelerators in practical settings, including user facilities that require flexible and rapid re-optimization of beam parameters. It will further enable real-time tuning strategies, automated control, and eventual deployment of LWFAs in scientific user facilities.

Contents

1. Introduction

1.1. Particle accelerators

Particle accelerators have been essential tools in advancing our understanding of the fundamental structure of matter. Since their inception in the early 20th century, particle accelerators have become indispensable tools in both fundamental and applied sciences. The concept of accelerating charged particles using time-varying electromagnetic fields was first realized by Rolf Wideröe in 1928 with the development of the linear accelerator based on radio-frequency (RF) drift tubes [1]. Shortly thereafter, based on Wideröe's ideas, Ernest Lawrence and M. Stanley Livingston introduced the cyclotron in 1931, revolutionizing compact particle acceleration [2].

Over the decades, these devices have evolved into large-scale scientific instruments, culminating in monumental projects such as the Large Hadron Collider (LHC) at CERN. The LHC uses superconducting magnets and high-frequency RF cavities to accelerate protons to energies of 6.5 TeV per beam in a 27-kilometer circular tunnel, facilitating groundbreaking discoveries such as the Higgs boson [3].

Beyond high-energy physics, particle accelerators are indispensable tools across various scientific and technological domains. In photon science, synchrotron radiation facilities and Free-Electron Lasers play a crucial role. The European XFEL [4] and SLAC's Linac Coherent Light Source (LCLS) [5], both powered by linear accelerators, generate femtosecond X-ray pulses of unprecedented brightness. These facilities have revolutionized structural biology through X-ray crystallography of proteins [6–8] and enabled real-time studies of ultrafast chemical reactions [9]. In medicine, conventional cyclotrons and synchrotrons are employed for proton and carbon ion therapy, offering superior dose localization for treating tumors with minimal damage to surrounding tissue [10, 11].

Despite their immense utility, conventional radio-frequency accelerators are often prohibitively large and expensive, limiting their broader adoption outside of large-scale national labs. This limitation arises fundamentally from the breakdown threshold of metallic RF cavities, which are responsible for accelerating charged particles via oscillating electric fields. These structures cannot sustain arbitrarily high field strengths and depending on the surface quality of the cavities, a vacuum breakdown occurs [12] when the fields exceed $100\,\mathrm{MVm^{-1}}$. Because of this constraint, the energy gain per unit length is limited, requiring either long linear accelerators called linacs composed of many sequential RF cavities or circular machines termed synchrotrons, such as the Large Hadron Collider (LHC), which circulate particles repeatedly through accelerating modules to reach the desired energies. However, synchrotrons are prone to the problem of relativistic particles emitting synchrotron radiation when bent along curved trajectories, resulting in energy losses that scale with the fourth power of the beam energy and inversely with the radius of curvature [13]. Larger ring sizes are required to compensate, significantly increasing construction costs and land requirements. As a result, high-energy accelerators become megaprojects costing billions of dollars and occupying vast physical footprints. This challenge has led to significant interest in exploring alternative acceleration mechanisms to develop compact, table-top accelerators that can democratize access to high-brightness electron beams and secondary radiation sources [14, 15].

1.2. Plasma wakefield acceleration

Wakefield acceleration, first proposed by Tajima and Dawson in 1979 [16], presents a promising route toward compact, high-gradient accelerators. In this approach, an ultra-intense laser pulse or a dense, relativistic charged particle bunch, referred to as the *driver*, propagates through an underdense plasma, expelling electrons from its path via its electromagnetic field. Since the much heavier ions remain essentially stationary on the timescale of the interaction, this displacement of electrons forms a positively charged ion cavity behind the driver. This three-dimensional co-moving structure formed behind the driver is called a plasma wakefield and supports field strengths that exceed the breakdown limits of conventional RF cavities by three to four orders of magnitude. For instance, at a typical plasma electron density of $10^{18}\mathrm{cm^{-3}}$, the longitudinal electric field can reach approximately $100\,\mathrm{GVm^{-1}}$, enabling electron acceleration to the GeV scale within just a few centimeters of plasma [17, 18].

Although the concept of using lasers as the driver for wakefield acceleration was proposed as early as 1979, its experimental realization was delayed by limitations in laser technology. The intensities required to drive substantial wakefields demand ultrashort laser pulses with Terawatt to Petawatt peak powers, which were beyond the capabilities of lasers in the 1980s. A transformative breakthrough came with the invention of chirped pulse amplification (CPA) in 1985 by Strickland and Mourou [19], a technique that enables the amplification of ultrashort pulses without damaging the gain medium. This method paved the way for the development of high-power lasers capable of delivering sub-picosecond pulses with peak powers in the Terawatt regime and led to the first demonstration of electrons accelerated by laser wakefields (LWFA) [20]. Following this, another milestone in the field came in 2004 when three independent groups reported the production of quasi-monoenergetic electron bunches from laser-driven wakefields[21–23]. Advances in laser technology, target design, and diagnostics have enabled the production of multi-GeV electron beams [24].

1.3. Challenges in stability and control of LWFA

Despite the significant progress made over the past two decades, several challenges remain before LWFA can transition from a laboratory environment to a robust tool for applied science and industrial use. The LWFA process is highly nonlinear and sensitive to small variations in laser and plasma parameters. Parameters such as plasma density, position of the laser focus and the laser dispersion intricately affect the resulting electron beam properties. Moreover, laser systems operating at powers of multi-Terawatt often exhibit fluctuations that further complicate experimental repeatability. These factors collectively give rise to pronounced shot-to-shot fluctuations as well as long-term drifts in beam parameters such as charge, energy, and divergence that are difficult to model, predict, or control with conventional manual tuning. These instabilities are compounded by the high dimensionality and nonlinearity of the parameter space, where small changes in laser or target settings can produce disproportionately large effects on accelerator performance. In this context, traditional heuristic optimization, such as grid scans or manual tuning, is often insufficient, particularly when multiple, often competing and correlated, objectives must be satisfied simultaneously.

A further complication arises in the context of user-driven facilities, where the demand for specific beam properties can change on short timescales depending on the experimental needs. For instance, one user may request high-charge, broadband electron beams to drive ultrafast X-ray absorption spectroscopy [25], while another may require narrow-bandwidth, quasi-monoenergetic beams for precision radiotherapy [26]. In such scenarios, the LWFA must be re-optimized—ideally within hours rather than days—to meet new performance targets. Achieving this level of operational flexibility is currently hindered by the absence of fast, reliable tuning strategies. Fine control over single-beam parameters, such as adjusting the peak energy while keeping charge and bandwidth constant, remains particularly challenging, underscoring the need for more automated, intelligent, and data-driven approaches to controlling LWFA systems.

Lastly, an often overlooked challenge in automation lies in the formulation of objective functions that are mathematical descriptions of the goal of the optimization. In many cases, it is not straightforward to express the desired beam characteristics, such as ”high quality,” ”low divergence,” or ”sufficient stability,” in precise mathematical terms suitable for direct optimization. Different applications may demand trade-offs between conflicting criteria, for example, maximizing beam charge while minimizing energy spread or optimizing stability at the expense of peak energy. Moreover, the users may find that the initially chosen objective functions do not fully capture what is desirable in practice, necessitating iterative redefinition. This ambiguity in specifying quantitative targets further complicates the development of conventional optimization workflows. Addressing this challenge requires not only efficient optimization algorithms but also building flexible models that allow for rapidly deriving different objective functions as new insights emerge during experimentation.

In principle, various optimization strategies could be envisioned to address this challenge, including evolutionary algorithms and gradient-based methods. However, several practical constraints render these approaches largely unsuitable for LWFA. Evolutionary algorithms typically require thousands of function evaluations across multiple generations to converge on an optimum. Given that Petawatt-class laser systems operate at low repetition rates, often delivering no more than one shot per minute, acquiring sufficient data for such algorithms would take an infeasibly long time and consume significant experimental beam time. Gradient-based optimization is likewise impractical because the inherently stochastic and chaotic behavior of laser–plasma interactions precludes the existence of smooth, differentiable objective functions. These limitations rule out conventional gradient descent or quasi-Newton methods, motivating the development of alternative optimization techniques capable of efficiently exploring complex, high-dimensional parameter spaces with a limited number of experimental shots.

1.4. Bayesian optimization as a solution

Bayesian optimization (BO) has emerged as a powerful tool for sample-efficient optimization of expensive-to-evaluate noisy objective functions such as an LWFA. It has recently gained popularity since the 2010s in the context of hyperparameter tuning of various machine learning methods [27] as it offers a principled strategy to outperform random or grid search by orders of magnitude in computational efficiency. Over the past decade, its versatility has led to widespread adoption across diverse scientific and engineering domains, including materials science [28], robotics [29] among others. At its core, BO operates by constructing a probabilistic surrogate model to capture beliefs about the objective function based on prior measurements. This model is updated iteratively as new evaluations are performed. Acquisition functions such as Expected Improvement, Probability of Improvement, or Upper Confidence Bound then determine where to query next, carefully balancing the trade-off between exploring uncertain regions and exploiting promising candidates. Eventually, the optimization is terminated either when the desired target performance is achieved or when the available budget of evaluations is exhausted.

The successful application of BO to other experimental disciplines naturally attracted interest from the laser wakefield community to assess whether it could improve the efficiency and reproducibility of laser-wakefield accelerators. The first demonstration of Bayesian optimization in this context was reported by Shalloo et al. [30], who demonstrated that BO could autonomously optimize a single objective, where the authors performed two experimental runs using only the total electron charge as an objective that was varied by adjusting the laser and target parameters in real time. This pioneering study established that BO could drastically reduce the number of required shots to reach optimal operating conditions compared to manual tuning, showing clear potential to improve the stability and reproducibility of LWFA performance. However, the approach was limited to scalar objectives and did not address the practical reality that accelerator users often care about multiple, sometimes competing, beam characteristics. For example, maximizing charge alone can degrade energy spread or maximum achievable mean energy, resulting in operating points that are suboptimal for different applications. Furthermore, prior work [31] in this domain did not consider the additional dimension of measurement fidelity[1] or explore strategies for fine-tuning specific beam parameters such as peak energy. These limitations were the main motivators behind the developments presented in this dissertation, which introduces multi-objective and multi-fidelity Bayesian optimization methods tailored to the unique challenges of LWFA experiments. By systematically addressing these gaps, this work aims to advance the field toward more versatile, reliable, and user-configurable plasma accelerators.

1.5. Scope and contributions of this thesis

This thesis aims to bridge the gap between algorithmic advances in Bayesian optimization and the experimental challenges of laser wakefield acceleration. The main contributions of this work are as follows:

- **Algorithmic Development:** The first multi-objective multi-fidelity (MOMF) Bayesian optimization method that can handle continuous fidelity levels was

[1] Fidelity refers to the accuracy of estimating beam properties from the number of shots, with a higher number of shots indicating higher fidelity data or estimation.

presented in this work. It is the first method that does not assume monotonicity of objective values across fidelities as highlighted in a later chapter. This approach significantly accelerates convergence, as validated on synthetic test functions and simulations.

- **Numerical Demonstration:** The first demonstration of using multi-fidelity Bayesian optimization to numerical simulations using the pseudo-spectral Particle-In-Cell code FBPIC [32]. An order-of-magnitude reduction in overall evaluation cost was achieved while maintaining optimization performance by utilizing low-fidelity simulations to guide high-fidelity simulations. Moreover, it was shown that the surrogate models constructed during optimization can later be reused for *a posteriori* tuning of the electron beam properties such as peak energies.

- **Experimental Application:** The MOMF optimization strategy is applied for the first time in laser wakefield acceleration experiments. The fidelity is defined by the number of laser shots per parameter setting, and the optimizer dynamically allocates resources based on cost and expected improvement. A model-based exploitation step is introduced to refine solutions to desired performance targets.

- **Energy Tuning in Experiments:** One of the most significant outcomes is the real-time tuning of peak energies using surrogate models built during optimization in a single iteration. Energy tuning has already been demonstrated using different schemes within the realm of LWFA. However, so far, demonstration of simultaneously moving eight different parameters to yield electrons at desired energies has always required new optimization cycles. Using the model generated during an optimization process enables tuning of electron beams at target energies within a few additional iterations, offering a powerful route toward control in future LWFA applications.

- **Physical Insights:** The models built by Bayesian optimization reveal deeper insights into the operational regimes of LWFA at CALA based on the underlying physics. Specifically, the underlying surrogate model can help elucidate the physics behind the generation of electron beams with different properties. Some of these insights are highlighted in chapter 5.

Together, these results demonstrate that Bayesian optimization, when appropriately tailored, can transform laser wakefield accelerator operation from a manual and empirical endeavor into a data-driven and controllable process. This paradigm shift is essential for scaling LWFA into practical applications across scientific and industrial domains.

1.6. Outline of the thesis

The remainder of the thesis is structured into the following chapters:

- **Chapter 2: Fundamentals of Laser-Wakefield Acceleration**
  This chapter reviews the core physics underlying LWFA, including laser-plasma interactions, wakefield generation, injection mechanisms such as self-injection and density downramp injection, and the resulting beam dynamics. It provides the theoretical background necessary to understand both the experimental and computational aspects of plasma accelerators.

- **Chapter 3: Bayesian Optimization and the MOMF Framework**
  This chapter introduces the mathematical foundations of Bayesian optimization, covering surrogate modeling with Gaussian processes, acquisition functions, and optimization strategies. It also surveys recent advances in multi-objective and multi-fidelity Bayesian optimization, culminating in the development of the proposed Multi-Objective Multi-Fidelity (MOMF) framework tailored for accelerator applications.

- **Chapter 4: Application to Numerical Simulations**
  This chapter demonstrates the use of the MOMF framework in high-fidelity numerical simulations of LWFA using the FBPIC code. It shows how the proposed approach accelerates the search for optimal parameters, compares different objective formulations, and highlights the ability to efficiently explore trade-offs in simulation-based studies.

- **Chapter 5: Experimental Implementation and Results**
  This chapter describes the integration of the MOMF approach into experimental campaigns on a Petawatt-class laser system. It details the experimental setup, data acquisition pipeline, and optimization results, including demonstrations of multi-objective multi-fidelity optimization, stability improvements, and the fine-tuning of electron beam energies.

- **Chapter 6: Conclusion and Outlook**
  This chapter summarizes the core results of this work and discusses the broader implications of these findings for plasma accelerator development. It also gives an outlook on future research directions that are currently underway to further enhance control, reproducibility, and performance in LWFA systems.

2. Physics of Laser-plasma Interactions

In this chapter, the fundamentals of laser–plasma interactions, which are crucial to understanding a Laser Wakefield accelerator (LWFA) will be discussed. The main components of an LWFA are the highly intense laser pulses and a gas that is ionized by the laser pulses to generate underdense plasmas. The chapter will begin by discussing laser light that can be described by Maxwell's equations. Essential to this work would be an understanding of the dispersion of the laser pulses that can be described using Taylor's expansion. Afterwards, the basics of plasma waves will be outlined and the generation of the plasma wakefields by the intense laser pulses. At the end, we would discuss the injection of the electrons into these wakefields to accelerate them to the MeV regime. More detailed theoretical descriptions, analytical derivations of the physics described in this chapter can be found in textbooks such as [33–40] among others.

2.1. High intensity laser pulses

$$\nabla \times \boldsymbol{H}(\boldsymbol{r},t) = \boldsymbol{J}(\boldsymbol{r},t) + \frac{\partial \boldsymbol{D}(\boldsymbol{r},t)}{\partial t}, \qquad \text{(Faraday's Equation)} \tag{2.1a}$$

$$\nabla \times \boldsymbol{E}(\boldsymbol{r},t) = -\frac{\partial \boldsymbol{B}(\boldsymbol{r},t)}{\partial t}, \qquad \text{(Ampère's Equation)} \tag{2.1b}$$

$$\nabla \cdot \boldsymbol{B}(\boldsymbol{r},t) = 0, \qquad \text{(Gauss's equation)} \tag{2.1c}$$

$$\nabla \cdot \boldsymbol{D}(\boldsymbol{r},t) = \rho_{ext}(\boldsymbol{r},t), \qquad \text{(Coulomb's Equation)} \tag{2.1d}$$

where $\boldsymbol{E}(\boldsymbol{r},t)$ is the electric field, $\boldsymbol{H}(\boldsymbol{r},t)$ is the magnetic field, $\boldsymbol{D}(\boldsymbol{r},t)$ is the electric flux density, $\boldsymbol{B}(\boldsymbol{r},t)$ is the magnetic flux density, $\boldsymbol{J}(\boldsymbol{r},t)$ is the electric current density, $\rho_{ext}(\boldsymbol{r},t)$ is the external charge density and $\boldsymbol{r}$ is a generalized vector in space. The pure microscopic form of Maxwell's equations can only hold in a vacuum for a limited number of external charges and currents. For the fields in condensed matter, the description of the fields taking into account all the point charges would exceed all available computational resources. Thus, the macroscopic form of Maxwell's equations averages the fields over subatomic distances (between electrons or ions), which is acceptable in the field of optics since all of the available experimental detectors are only able to resolve the fields on larger than atomic scales. The material response is also an averaged response instead of individual point charges or currents. Consequently, this results in two additional fields $\boldsymbol{H}(\boldsymbol{r},t)$ and $\boldsymbol{D}(\boldsymbol{r},t)$ that are dependent on the material response through the following equations [33]

$$\boldsymbol{D}(\boldsymbol{r},t) = \varepsilon_0\boldsymbol{E}(\boldsymbol{r},t) + \boldsymbol{P}(\boldsymbol{r},t), \tag{2.2a}$$

$$\boldsymbol{H}(\boldsymbol{r},t) = \frac{1}{\mu_0}[\boldsymbol{B}(\boldsymbol{r},t) - \boldsymbol{M}(\boldsymbol{r},t)], \tag{2.2b}$$

where $\boldsymbol{P}(\boldsymbol{r},t)$ is the dielectric polarization and $\boldsymbol{M}(\boldsymbol{r},t)$ is the magnetic polarization, ε_0 is the vacuum permittivity and μ_0 is the vacuum permeability. The vacuum permittivity and vacuum permeability are related to the speed of light c via the following relation

$$c^2 = \frac{1}{\mu_0\varepsilon_0}. \tag{2.3}$$

In the case of vacuum, assuming no condensed matter, the dielectric polarization and the magnetic polarization become zero resulting in the following modified form of equations 2.2a and 2.2b [36]

$$\boldsymbol{D}(\boldsymbol{r},t) = \varepsilon_0\boldsymbol{E}(\boldsymbol{r},t), \tag{2.4a}$$

$$\boldsymbol{H}(\boldsymbol{r},t) = \frac{1}{\mu_0}\boldsymbol{B}(\boldsymbol{r},t). \tag{2.4b}$$

Substituting equations 2.4a and 2.4b in the equations 2.1a-2.1d and using 2.3 we get the following modified Maxwell's equations [35]

$$\nabla \times \boldsymbol{B} = \mu_0\boldsymbol{J} + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t}, \tag{2.5a}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \tag{2.5b}$$

$$\nabla \cdot \boldsymbol{B} = 0, \tag{2.5c}$$

$$\nabla \cdot \boldsymbol{E} = \frac{\rho_{ext}}{\varepsilon_0},. \tag{2.5d}$$

Please note that we dropped the explicit dependence of the fields on space and time for the sake of brevity, however the fields are still dependent on these quantities.

We can reduce the four first-order coupled Maxwell's equations into two second-order differential equations through the use of scalar Φ and vector $\boldsymbol{A}$ potentials. We can make use of the fact that $\nabla \cdot (\nabla \times \boldsymbol{F}) = 0$ for vector fields $\boldsymbol{F}$ to define $\boldsymbol{B}$ in terms of a vector potential $\boldsymbol{A}$

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}. \tag{2.6}$$

Substituting 2.12 into 2.5b and using the fact that a vanishing curl of a vector field implies the existence of a scalar potential we can write the following equation

$$\boldsymbol{E} = -\frac{\partial \boldsymbol{A}}{\partial t} - \nabla\Phi. \tag{2.7}$$

Since the potentials are constructed using the homogeneous Maxwell's equations 2.5b and 2.5c, these are inherently fulfilled. The dynamics of the two potentials are then determined by the two other inhomogeneous equations which contain the source of charges and currents. Inserting Equations 2.12 and 2.7 into 2.5a and 2.5d, using the vector identity $\nabla \times (\nabla \times \boldsymbol{A}) = \nabla(\nabla.\boldsymbol{A}) - \nabla^2\boldsymbol{A}$ results in two second-order differential equations for the scalar and the vector potential

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Phi - \nabla^2\Phi = \frac{\rho_{ext}}{\varepsilon}, \tag{2.8a}$$

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\boldsymbol{A} - \nabla^2\boldsymbol{A} = \mu_0\boldsymbol{J}. \tag{2.8b}$$

The above two equations together form the wave equations for the scalar and vector potentials. We can solve these equations for the special case of vacuum assuming no free charges and currents the simplest solution to these equations is of the form

$$\Phi = 0, \tag{2.9a}$$

$$\boldsymbol{A} = -\boldsymbol{A_0}\sin(\omega_L t - \boldsymbol{kr} + \varphi), \tag{2.9b}$$

$$\tag{2.9c}$$

where $\omega_L = 2\pi c/\lambda_L$ is the angular frequency with the central wavelength λ_L, $\boldsymbol{k}$ is the wave vector and φ is a general phase offset. From the definitions of the vector and scalar potential we can calculate the electric and magnetic fields of this particular solution

$$\boldsymbol{E} = \omega_L\boldsymbol{A_0}\cos(\omega_L t - \boldsymbol{kr} + \varphi), \tag{2.10a}$$

$$\boldsymbol{B} = \boldsymbol{k} \times \boldsymbol{A_0}\cos(\omega_L t - \boldsymbol{kr} + \varphi). \tag{2.10b}$$

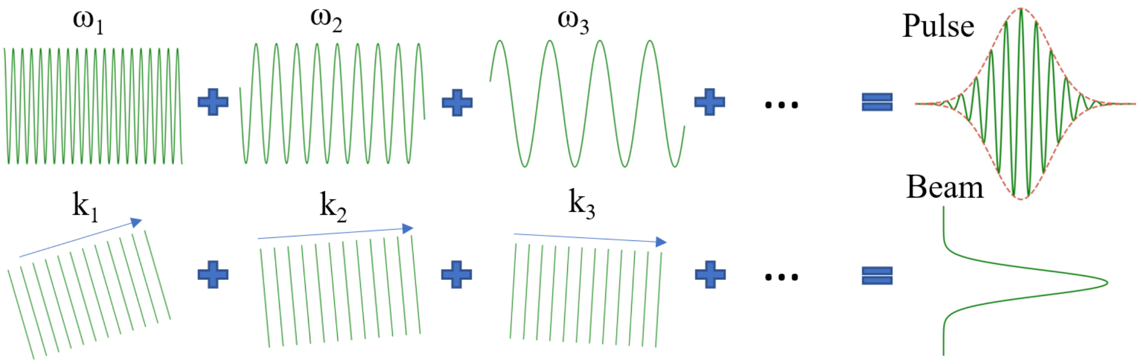

Figure 2.1.: At the top we can see that the superposition of plane waves with different temporal frequencies leads to a pulse in time domain. Similarly the superposition of plane waves with different spatial frequencies or wave vectors leads to a beam in the spatial domain. A pulse can also be described using an envelope function shown here in red and a carrier frequency shown here in green.

These solutions are known as plane waves, which travel in the direction of $\boldsymbol{k}$ which is connected to the angular frequency through the dispersion relation in vacuum

$$\mid k \mid = \frac{\omega_L}{c}. \tag{2.11}$$

The electric and magnetic fields are perpendicular to each other and both are perpendicular to the wave vector, the amplitude of the fields is related by

$$\mid \boldsymbol{A} \mid = \frac{c}{\omega_L} \mid \boldsymbol{B} \mid = \frac{1}{\omega_L} \mid \boldsymbol{E} \mid . \tag{2.12}$$

The magnitude of the electric field is c times larger than the magnetic field. Since Maxwell's equations are linear with respect to the electric and magnetic fields, thus an infinite linear superposition of plane waves with different angular frequencies is also a solution to Maxwell's equations and represents a more general solution. These solutions can be written in the form of complex amplitudes

$$\boldsymbol{E(r,t)} = \int_{-\infty}^{\infty} \boldsymbol{E}(\boldsymbol{k},\omega) \exp[i(\boldsymbol{kr} - \omega t)]\, d^3k d\omega + c.c, \tag{2.13}$$

where $\boldsymbol{E}(\boldsymbol{k},\omega)$ represents the amplitude of plane waves with different frequency and different propagation direction in frequency domain and c.c represents the complex conjugate. The above description is used to describe pulsed beams which have a finite spatial and temporal width. A solution consisting of a single propagation direction and a single frequency results in a temporally and spatially infinite wave. Thus, to form a short pulse the superposition of plane waves of different frequencies is needed and to form a beam, the superposition of plane waves with different wave vectors is needed as shown in Figure 2.1.

To simplify the description of ultrashort femtosecond pulses, the Slowly Varying Envelope Approximation (SVEA) is used, where a pulse is described by an envelope function $\boldsymbol{V}(r,t)$ with rapid oscillations at a carrier frequency ω_L [33]

$$\boldsymbol{E(r,t)} = \frac{1}{2}\boldsymbol{V}(\boldsymbol{r},t)\exp[i(\boldsymbol{kr} - \omega_L t)] + c.c. \tag{2.14}$$

Here the pulse is considered to be propagating in a medium, hence $\mid k \mid = n(\omega)\frac{\omega_L}{c}$. The addition of a refractive index that is dependent on the carrier frequency $n(\omega)$ depicts the propagation of the pulse in a medium where the different wave vectors experience different retardation factors. The SVEA holds if the fast carrier frequency is at least an order of magnitude greater than the frequency with which the envelope changes. For the laser pulses from ATLAS which typically have a duration of 27 fs with an optical cycle of 2.6 fs, the SVEA can be applied without measurable deviations.

If we assume that the pulse with a wide spectrum is centered around the frequency ω_0, then we can apply Taylor's expansion on the frequency dependent wave vector $k(\omega)$ [34]

$$\begin{aligned} k(\omega) = k(\omega_0) + \frac{\partial k}{\partial \omega}\Big|_{\omega_0}(\omega - \omega_0) + \frac{1}{2}\frac{\partial^2 k}{\partial \omega^2}\Big|_{\omega_0}(\omega - \omega_0)^2 \\ + \frac{1}{6}TOD(\omega - \omega_0)^3 + \frac{1}{24}FOD(\omega - \omega_0)^4 + \ldots \quad . \end{aligned} \tag{2.15}$$

The first coefficient of the above equation describes the phase velocity of the laser pulse

$$v = \frac{\omega_0}{k(\omega_0)} = \frac{c}{n(\omega_0)} \tag{2.16}$$

where $n(\omega_0)$ is the refractive index of the dispersive material at the central frequency ω_0, TOD is known as the third order dispersion and FOD refers to the fourth order dispersion. The second coefficient defines the group velocity of the pulse which can be thought of as the velocity of the pulse envelope

$$v = \left[\frac{\partial k}{\partial \omega}\Big|_{\omega_0}\right]^{-1} = \frac{c}{n(\omega_0) + \omega\frac{\partial n(\omega)}{\partial \omega}\Big|_{\omega_0}} = \frac{c}{n_g(\omega)} \tag{2.17}$$

where $n_g(\omega)$ is known as the group index. The third coefficient is known as the group velocity dispersion or GVD.

$$GVD = \frac{\partial^2 k}{\partial \omega^2}\Big|_{\omega_0} \tag{2.18}$$

When a pulse passes through a medium having a non-zero GVD, the pulse shape changes. Similarly, we can define the third and fourth orders of dispersion which affect the pulse shape in different ways. We will see later that the GVD, TOD and FOD were parameters that were controlled during the optimization of the laser-wakefield accelerator since they play a vital role in the evolution of the laser pulse within the plasma. This will be demonstrated in Section 2.2.

So far we have restricted our discussion to the Gaussian pulse in time domain. Another important consideration is the spatial profile and the propagation of this profile in the spatial domain. For simplicity we will fix our coordinate system such that the propagation of the beam is in z-direction and assume that the polarization is in x-direction. We also assume that the spatial frequencies of the beam are much smaller than the propagation distance leading to Fresnel approximation. Solving the wave equation under Fresnel approximation leads us to the formalism of Gaussian beams [36]

$$\boldsymbol{E}(x,y,z) = \hat{x}E(z)\exp - \frac{x^2+y^2}{w^2(z)} \exp i\frac{k(x^2+y^2)}{2R(z)} \exp i\varphi(z), \tag{2.19}$$

where $w(z)$ is called the width of the Gaussian beam, $E(z)$ is the amplitude, $R(z)$ is the radius of curvature and $\varphi(z)$ is the Gouy phase shift. From the above equation we can see that the Gaussian beam keeps its profile while its width, amplitude, phase curvature and phase shift changes with propagation in z. Each of these components can be written in the form of z as

$$E(z) = E_0 \frac{1}{\sqrt{1+\left(\frac{z}{z_0}\right)^2}}, \tag{2.20a}$$

$$w(z) = w_0\sqrt{1+\left(\frac{z}{z_0}\right)^2}, \tag{2.20b}$$

$$R(z) = z\left[1+\left(\frac{z}{z_0}\right)^2\right], \tag{2.20c}$$

$$\varphi(z) = \arctan\left(-\frac{z}{z_0}\right), \tag{2.20d}$$

where $z_0 = \frac{\pi w_0^2}{\lambda}$ is called the Rayleigh length defined where the peak intensity of the Gaussian beam drops to $\frac{1}{2}$ of its maximum value. The width of the Gaussian beam reaches a minimum value at $z = 0 \implies w(z) = w_0$ called the waist of the beam. The radius of curvature at the waist or the focus of the Gaussian beam becomes infinite implying flat wavefronts, and it reaches a minimum at z_0. For a large distance z the amplitude decreases and the width increases linearly with an asymptotic angle of $\theta = \frac{2w_0}{z_0}$. Finally, we will now define the laser intensity in terms of the Poynting vector and also outline how it is practically determined.

The energy flux of the electromagnetic field is given by the Poynting vector $\boldsymbol{S}$ and in practice is always measured through a surface of a detector $\boldsymbol{S} \cdot \boldsymbol{n}$ where $\boldsymbol{n}$ is the normal vector of the detector surface. The Poynting vector is related to the electric and magnetic fields through the following relation

$$\boldsymbol{S} = \frac{1}{\mu_0}\boldsymbol{E} \times \boldsymbol{B}. \tag{2.21}$$

The optical intensity of the electromagnetic field is defined as the temporal average of the Poynting vector and is given by

$$I = \langle|\boldsymbol{S}|\rangle = \frac{1}{\mu_0}\langle|\boldsymbol{E}| \times \boldsymbol{B}\rangle = \varepsilon_0 c\langle\boldsymbol{E^2}\rangle = \frac{\varepsilon_0 c}{2}\tilde{E}_t^2, \tag{2.22}$$

where I is the optical intensity, $\tilde{E_t}$ is the slowly varying envelope function of the electric field. The intensity explicitly depends on the electric field magnitude because it temporally averages over the fast oscillations and hence any phase information of the electric field is lost. In practice, a simple yet useful engineering formula for calculating the laser intensity is given by

$$I = \frac{pulse\ energy}{pulse\ duration \times focal\ area}. \tag{2.23}$$

where the pulse duration is FWHM duration of the laser pulse and focal area is calculated by considering the FWHM spatial width of the beam. This formula assumes uniform distribution of the pulse energy within the temporal and spatial widths, which is strictly not true as seen in Figure 2.1. For a Gaussian pulse there is an additional scalar factor of 0.94 and for pulses shaped differently, this factor would change. In the field of laser-plasma physics, the intensity of the laser pulses is generally defined in units of Wcm^{-2}.

2.2. Plasma wakefields

After describing the laser light, we will now move to discuss the second important component in building up a laser wakefield accelerator. We will start with a few basic plasma definitions and then move onto discussing how the laser light behaves when it propagates through a plasma. The evolution of the laser pulse and the different focusing or defocusing effects that it is subjected to will be discussed. Once we have all of these basics covered, we will move onto how the wakefields are generated, which will be covered in the Section 2.3.

2.2.1. Basic plasma definitions

We will start with a few basic concepts from plasma physics that describe this often-called fourth state of matter. The plasma consists of ionized particles and appears to be quasi-neutral and depicts a collective behavior dominated by the electro-magnetic interaction of the ionized particles. Most of the behavior of the plasma occurs because of this long-range Coulomb interaction between the charged electrons and ions, which is not affected by the short-range interaction between the neutral atoms or molecules. This interaction can be between single particles or between an ensemble of particles which is then described by a distribution function of particles.

One of the most important collective behavior of the plasma relevant for this work is the plasma oscillations. Let us consider a plasma that has been formed from a gas, as in the case of this work by a highly intense laser pulse via ionization [39, 41]. The electrons comprising the plasma are displaced by a distance and move against a positive and uniform ion background under the influence of an applied external field. This displaced sheath of electrons and the uniform ion background is similar to a plate capacitor with a charge density $\sigma = en_e d$ where d is the distance between the electron sheath and the ionic background. This charge separation induces an electric field that pulls electrons back to the ionic background. By using the equation of motion and the electric field

$$\ddot{d} = -\frac{eE}{m_e} = -\frac{e^2 n_e}{m_e \epsilon} d, \tag{2.24}$$

an equation for the harmonic oscillator is derived.

The characteristic frequency of such an oscillation is called the plasma frequency and follows the following equation [38]

$$\omega_p = \sqrt{\frac{n_e e^2}{\varepsilon_0 m_e}}, \tag{2.25}$$

The frequency depends on the density and the mass of electrons, and therefore decreases when the electrons experience a relativistic mass increase. The different effects arising from this dependency will be discussed in Section 2.2.3.

2.2.2. Electromagnetic fields in plasma

In this section, we will outline the basic equations that describe an electromagnetic field and move on to the dispersion relation and the refractive index of the plasma.

At the end we will look at the laser pulsed beam evolution when it travels through the plasma.

The evolution of a plasma can be defined using a kinetic model, where the electrons and the ions are defined using distribution functions $f_e(\boldsymbol{r}, \boldsymbol{v}, t)$ and $f_i(\boldsymbol{r}, \boldsymbol{v}, t)$ respectively. The time evolution of these distribution functions can be described by the following system of equations [42]

$$\frac{\partial f_e}{\partial t} + \boldsymbol{v_e} \cdot \nabla_x f_e - e\left[\boldsymbol{E} + \frac{\boldsymbol{v_e}}{c} \times \boldsymbol{B}\right] \cdot \nabla_p f_e = 0, \tag{2.26a}$$

$$\frac{\partial f_i}{\partial t} + \boldsymbol{v_i} \cdot \nabla_x f_i + Z_i e\left[\boldsymbol{E} + \frac{\boldsymbol{v_i}}{c} \times \boldsymbol{B}\right] \cdot \nabla_p f_i = 0. \tag{2.26b}$$

The above equations can then be coupled to Maxwell's equations through the source terms of charges and currents

$$\boldsymbol{J} = e\int (Z_i f_i \boldsymbol{v_i} - f_e \boldsymbol{v_e}) d^3 p, \tag{2.27a}$$

$$\rho = e\int (Z_i f_i - f_e) d^3 p. \tag{2.27b}$$

This set of equations along with Maxwell's equations completely describes the behavior of the plasma where the binary collisions are not dominant. These equations are solved numerically by the different varieties of particle-in-cell codes. We will later on see one such code termed Fourier-Bessel Particle-In-Cell (FBPIC) [43] that was used extensively in this work. For now, we will restrict ourselves to simpler analytical cases whereby we can gain insights into the interaction of electromagnetic fields with the plasma.

We will consider a monochromatic plane wave of the form $\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}(\boldsymbol{k}, \omega)\exp[i(\boldsymbol{kr} - \omega t)]$, which acts like a driving field for a charge particle. The equation of motion of this charged particle is given by

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{F_L} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}), \tag{2.28}$$

where $\boldsymbol{p}$ is the momentum of the charged particle and $\boldsymbol{F_L}$ is the Lorentz force. When we are dealing with particles with velocities much less than the speed of light $\mid v \mid \ll c$, we can ignore the force component $\boldsymbol{v} \times \boldsymbol{B}$ since $\mid B \mid \approx \frac{|E|}{c}$ and are left with

$$m\frac{d\boldsymbol{v}}{dt} = q\boldsymbol{E}. \tag{2.29}$$

Taking the curl of Equation (2.5b) and replacing the curl of $\boldsymbol{B}$ from Equation (2.5a) we get the following equation

$$\nabla \times \nabla \times \boldsymbol{E} = -\mu_0 \frac{\partial \boldsymbol{J}}{\partial t} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2}. \tag{2.30}$$

The current can be derived by integrating Equation (2.27a) in velocity space and results in $\boldsymbol{J} = -en_e v$, ignoring the contributions from ions since we assume that ions are much heavier and not displaced by the electric field. This relation along with Equation (2.29) allows us to evaluate the first term in Equation (2.30) and using the definition of the monochromatic plane wave we get

$$\boldsymbol{k} \times \boldsymbol{k} \times \boldsymbol{E} = \left(\frac{\omega^2}{c^2} - \frac{\mu_0 e^2 n_e}{m_e}\right)\boldsymbol{E}. \tag{2.31}$$

For the above equation to hold we need to satisfy the dispersion relation

$$(kc)^2 = \omega^2 - \omega_p^2. \tag{2.32}$$

Using the dispersion relation we can define the phase and group velocity similar to how we defined it for a laser pulse in Equation (2.16) and Equation (2.17) respectively

$$v_p = \frac{\omega}{k(\omega)} = \frac{c}{\sqrt{1 - \frac{{\omega_p}^2}{\omega^2}}} = \frac{c}{\eta}, \tag{2.33}$$

where we have Equation (2.32) and η is termed as the refractive index of the plasma. Similarly the group velocity can be defined as

$$v = \left[\frac{\partial k}{\partial \omega}\right]^{-1} = \frac{c^2 k}{\omega} = c\eta, \tag{2.34}$$

where we have used Equation (2.33) in the last step. From these expressions we can see that the plasma refractive index is only real-valued for $\omega > \omega_p$ implying a traveling wave inside the plasma. On the other hand when $\omega < \omega_p$, then the refractive index is imaginary we have exponentially decaying evanescent waves. Physically, we can imply that for frequencies above the plasma frequency the electrons are too slow to follow these frequencies and hence the field penetrates the plasma. For frequencies below the plasma frequency, the disturbance is perfectly shielded and hence the electromagnetic wave is reflected.

Usually it is convenient to express the plasma refractive index in terms of the plasma electron density. In this case, an electromagnetic wave of a given frequency ω can only propagate in a plasma with a density n_e less than

$$n_{crit} = \frac{\omega^2 \varepsilon_0 m_e}{e^2}, \tag{2.35}$$

which is called the critical density of the plasma for a given frequency ω. Using the critical and the plasma electron density the refractive index η becomes

$$\eta = \sqrt{1 - \frac{n_e}{n_{crit}}}. \tag{2.36}$$

A plasma with an electron density higher than the critical density is termed as an overdense plasma while those with an electron density less than the critical is called an underdense plasma and allows an electromagnetic wave to propagate through it. For the frequency of the laser light dealt with in this work, the critical density is on the order of 10^{21}cm^{-3} while the hydrogen gas used in the work has a plasma electron density of $n_e \approx 10^{18}\text{cm}^{-3}$. Hence, the plasma in this work is categorized as underdense and the laser light can propagate through it.

2.2.3. Relativistic non-linear effects

Once the motion of the electrons which are being driven by the laser light becomes relativistic, we need to modify the equations derived earlier to take into account the relativistic mass increase of electrons. Another important consideration is the dependence of the refractive index on the local intensity $\eta \implies \eta(I_L)$ changes which takes us into the domain of non-linear optics. Both of these effects are seen when a highly intense laser pulse is propagating through the plasma in this work. Since the plasma frequency is dependent on the mass of an electron we can directly introduce the Lorentz factor into the equation of the refractive index of plasma

$$\eta = \sqrt{1 - \frac{{\omega_p}^2}{\langle\gamma\rangle\omega^2}} = \eta(I_L), \tag{2.37}$$

where $\langle\gamma\rangle$ is the temporally averaged Lorentz factor and I_L is the local intensity of the laser. The local intensity $I_L((r), t)$ at the focus of a laser pulse varies with position due to the Gaussian nature of the beam and changes over time because the pulse envelope also follows a Gaussian shape. Equation (2.37) implies that the phase velocity and the group velocity of the laser pulse is also a function of position and time. We can now expand Equation (2.37) assuming relativistically underdense plasma $\frac{\omega_p^2}{\langle\gamma\rangle} \ll \omega^2$ and assuming all perturbations of density, Lorentz factor and the local light frequency to be small [44]

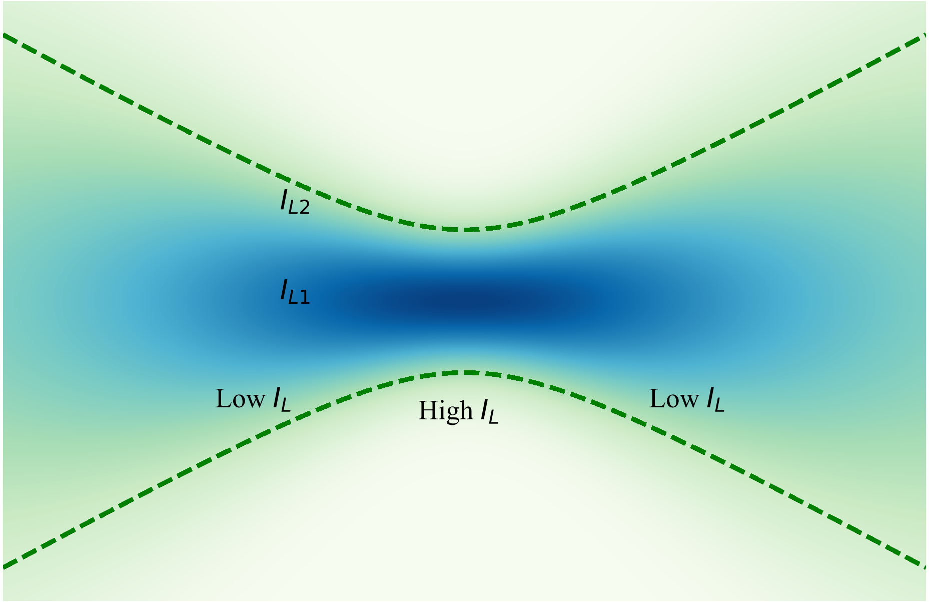

Figure 2.2.: The intensity profile in the focus of a laser beam is shown. The regions closer to the central axis have a higher intensity than at the edges of the laser focus. Also in the longitudinal direction a higher intensity is reached closer to the waist than at a region further away from the focus. This variation of intensity in the laser focus consequently leads to a variation of the plasma refractive index.

$$\eta \approx 1 - \frac{1}{2}\frac{{\omega_p}^2}{\omega^2}\left(1 + \frac{\delta n_e}{n_e} - 2\frac{\delta\omega}{\omega} - \frac{a_0^2}{4}\right). \tag{2.38}$$

We will briefly look at the different spatio-temporal effects that result from this modified refractive index [45].

Spatial focusing effects

The first transverse focusing effect that we discuss is known as the ionization-induced defocusing. The focusing laser beam results in a higher on axis intensity I_{L1} than at the edges I_{L2} as shown in Figure 2.2 owing to the properties of a propagating Gaussian beam (see Section 2.1). As a result, the plasma density is higher at the center compared to the edges $n_{e1} > n_{e2}$ resulting in a lower refractive index at the center compared to the edges $\eta_1 < \eta_2$. This implies that the phase velocity of the laser pulse is higher at the center resulting in initially flat wave fronts to bulge. This leads to a defocusing of the beam [46] and hence the beam would not converge to the achievable waist. This is one of the reasons that a high vacuum is needed before a high-powered beam can be focused down to reach high intensities. In this work, the gas used is primarily hydrogen, which has a low ionization threshold and hence this effect is not seen by the main laser beam since the edges of the beam also fully ionize the hydrogen gas. Moreover, for dense gases the ionization-induced defocusing is countered by the effects of ponderomotive and relativistic self-focusing that we discuss next.

Second spatial focusing effect is a result of a ponderomotive force [47–49] that the electrons feel due to the inhomogeneity of the laser focus. Consider a particle situated in the focus of the laser beam, it experiences transversely a higher magnitude of the electric field oscillation in the first cycle since it is in an area of stronger fields. The force in the second cycle of an area of weaker fields is not enough to offset the force in the first cycle. Hence, over a complete cycle the particle ends up being pushed away from regions of higher intensity. This reduces the density of the of the plasma in the center where the laser has the stronger fields. Consequently, the refractive index is higher in the center than the edges from Equation (2.36), implying a higher phase velocity of the laser pulse at the edges compared to the center. The flat wave fronts of the pulse are bent inward and lead to a focusing effect termed as the ponderomotive self-focusing, and it counters the ionization-induced defocusing in dense plasmas.

The third and final spatial focusing effect that we will discuss is called relativistic self-focusing. Similar to ponderomotive self-focusing, it counteracts ionization defocusing but the mechanism and the physics behind it is different from ponderomotive self-focusing. While the ponderomotive force change the plasma electron density in different regions, the relativistic self-focusing arises due to the relativistic mass increase of the electrons. This effect depends on the power of the laser pulse rather than the intensity, and the threshold power can be found by balancing the natural diffraction with the self-focusing effect and is given by [50–52]

$$P_{crit} \approx \frac{n_{crit}}{n_e} 17.4\text{GW}. \tag{2.39}$$

The relativistic self-focusing increases if for a similar plasma electron density the power in the laser pulse is increased. Alternatively, the effect can also be increased with a constant power by increasing the plasma electron density using a higher gas pressure. Taking into account this relativistic self-focusing, Equation (2.20b) is modified and to first order approximation can be written as

$$w(z)^2 = w_0^2 \left[1 + \left(1 - \frac{P}{P_{crit}} \right) \frac{z^2}{z_0^2} \right]. \tag{2.40}$$

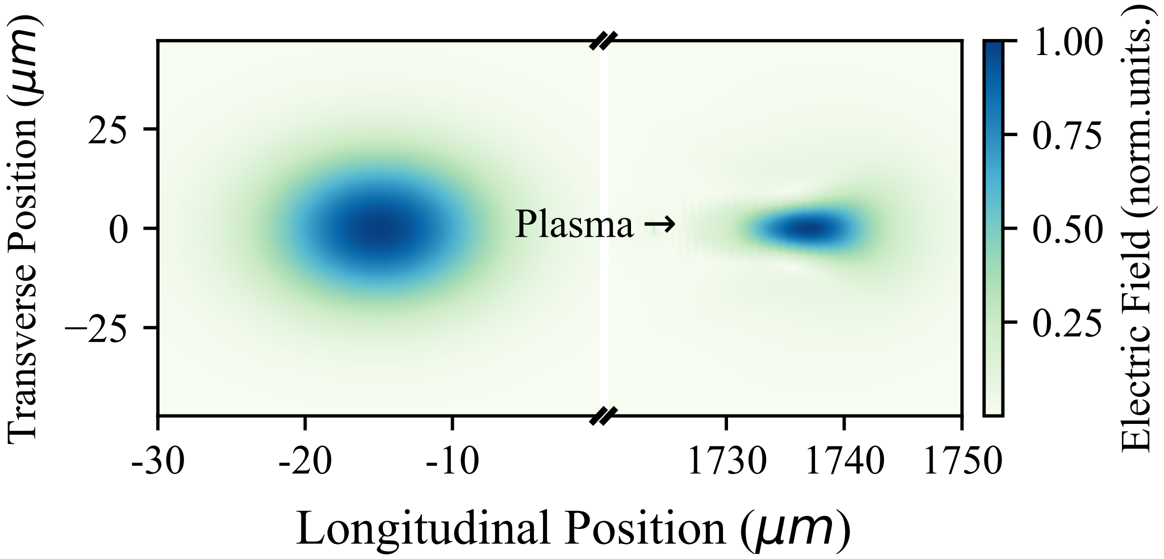

Figure 2.3.: The evolution of the 2D envelope of a laser pulse is shown from a FBPIC simulation after propagating $1.7\mu\mathrm{m}$ in a plasma with an electron density of $n_e = 5.5e18\mathrm{cm}^{-3}$. On the left is the laser envelope at the start of the simulation before entering the plasma at $z = 0\mu\mathrm{m}$. On the right is the laser pulse after propagating $1.7\mathrm{mm}$ into the plasma. We see that the laser pulsed beam has shrunk both transversely and longitudinally. The relativistic and ponderomotive focusing effects dominate over the ionization defocusing effects.

For laser powers much greater than the critical power, higher orders need to be taken into account that prevent the laser beam from converging to a non-physically small spot size. For these much higher powers the dominant effect of focusing is due to the ponderomotive force while for powers around the critical power, the effect of relativistic self-focusing is more important [14, 52]. Some of these effects can be seen from the result of an FBPIC simulation Figure 2.3 where the 2D envelope of the laser pulse is shown. The laser power considered in this simulation is on the order of 50TW while the critical power for the plasma electron density in this case is on the order of $P_{crit} = 5\mathrm{TW}$.

Temporal compression effects

The effects that we described above also hold true for the longitudinal variations in the plasma electron density and consequently, the laser pulse. This results in local variations in the pulse frequency and also pulse steepening or elongation effects.

The first effect that is a result of the longitudinal variation of the plasma electron density arises from varying ionization degrees. In a particular gas of high atomic number where a higher number of electrons can be ionized the front of the pulse will experience a lower plasma electron density. The peak of the pulse would see

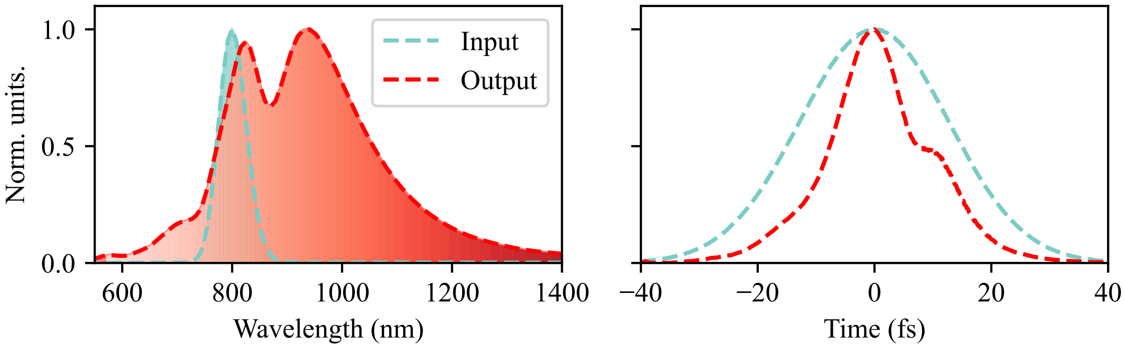

Figure 2.4.: The effect of red-shift on the spectrum of the laser pulse before and after propagating in the plasma is shown on the left. The ponderomotive and relativistic effect dominate over the ionization effect leading to a net red shift. This leads to Self-Phase modulation increasing the bandwidth of the input pulse. This effect also leads to a pulse compression and distortion as can be seen on the right sub figure.

the highest plasma electron density since it is the most intense and can ionize the gas to a higher degree. The back of the pulse (after the peak) will see a constant plasma density since the recombination time is significantly longer than the pulse duration under 100fs. Since the front of the pulse sees a higher refractive index the phase velocity at the front is smaller than the phase velocity at the back. This increased phase velocity at the back of the pulse results in a chirp being introduced that blue shifts the laser pulse, which is why this effect is termed as the ionization blue-shift. For experiments in this work, this ionization-induced effect is not significant since hydrogen was primarily used, which can be ionized by the leading edge of the pulse. The main pulse sees a more uniform plasma density, and only the head of the pulse could be affected by the ionization blue-shift. In our experiments the relativistic and ponderomotive effects are more prominent which as in the spatial case work against the ionization induced effects and cause a red shift. This can be seen in Figure 2.4 where the initial spectrum and the output spectrum of a pulse propagating in a plasma are shown. Moreover, because of the red or blue shifts the spectrum of the pulse broadens which is known as self-phase modulation (SPM) as can be seen in Figure 2.4.

The second effect is due to the expulsion of electrons by the ponderomotive force in the forward direction from the leading edge of the laser pulse and in the backward direction by the trailing edge of the laser pulse. This effect means that the rising edge of the pulse will experience a density higher than that of the peak of the pulse. Additionally the remaining electrons will have gained relativistic mass. Both of these effects essentially mean that the trailing edge of the pulse would see a plasma of higher refractive index and consequently would result in a higher group velocity according to Equation (2.34). On the other hand, the front of the pulse is slowed down and leads to a steepening of the pulse front, consequently compressing the pulse [53] as seen in Figure 2.4. This effect reduces the longitudinal length of the pulse and can be seen in Figure 2.3 where the pulse envelope has shrunk.

2.3. Laser wakefield acceleration

In this section, we will discuss the basics of a laser wakefield accelerator. We have already discussed how the axial and radial ponderomotive forces of the laser field displace electrons into regions of low intensity. The much heavier ions of the gas are not displaced directly by the field, so we create a charge separation. This results in a potential that forces the electrons to return to the axis when the laser has moved forward. Thus, the electrons oscillate collectively around the laser axis resulting in a plasma wave excitation. This particular plasma wave is called a wakefield owing to its similarity to a wake generated by a boat moving in water. The axial fields in the wake can exceed TV/m which is many orders of magnitude higher than conventional RF cavities and is the main motivation behind using plasma waves for electron acceleration. Another advantage of this behavior of the electrons is the radial fields that keep the electrons inside the structure while they are being accelerated by the axial fields. In this section, we will go over the generation of the wakefields, followed by mechanisms of injecting electrons into this structure and then we will finally look at limitations of this process.

2.3.1. Linear wakefield generation

As previously described, an external field applied to a plasma can excite collective electron oscillations in a plasma. A laser pulse can provide this external field and for some cases this process can be described and solved analytically using a set of three differential equations

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\varepsilon_0}, \qquad \text{Gauss's Law} \tag{2.41a}$$

$$\frac{\partial \rho}{\partial t} + \nabla j = 0, \qquad \text{Continuity Equation} \tag{2.41b}$$

$$\frac{\partial \boldsymbol{p}}{\partial t} = \boldsymbol{F}_E + \boldsymbol{F}_{pond}, \qquad \text{Equation of Motion} \tag{2.41c}$$

where $\boldsymbol{F}_E$ and $\boldsymbol{F}_{pond}$ are the Lorentz and the ponderomotive forces respectively. For the derivations and the discussions for the wakefields we will assume that the ions are immobile and create a stationary background which can be justified owing to their much higher mass than the electrons. We also assume that the density perturbations δn are small and hence can be written as $n_e = n_0 + \delta n$, where n_0 is the constant, unperturbed background plasma density and n_e is the total plasma electron density. The direction of travel for the laser pulse is assumed to be z. The wave equation for the density perturbations δn in the case of a weak laser driver ($a_0 \ll 1$) is [39, 54]

$$\left(\frac{\partial^2}{\partial t^2} + \omega_p^2\right)\delta n = n_0 \frac{c^2}{4}\nabla^2 a^2, \tag{2.42}$$

where $\nabla = \frac{\partial}{\partial z}\hat{z} + \frac{\partial}{\partial r}\hat{r}$. For simplicity, we will apply a coordinate transformation [50, 55] to a co-moving frame with the phase velocity of the plasma wakefield $\xi = z - ct$ and $\tau = t$. The derivatives are then transformed to $\partial_\xi = \partial_z$ and $\partial_t = \partial_\tau - c\partial_{xi} \approx -c\partial_\xi$. The ∂_τ vanishes since we assume quasi-static approximation. The wave equation for the axial field in the direction of the laser propagation can be attained using Equation (2.41a) and the equations for the density perturbations and the axial field in the co-moving frame are

$$\left(\frac{\partial^2}{\partial \xi^2} + k_p^2\right)\delta n = \frac{n_0}{4}\nabla^2 a^2 \tag{2.43a}$$

$$\left(\frac{\partial^2}{\partial \xi^2} + k_p^2\right)\boldsymbol{E}(r,\xi) = k_p^2 \nabla \Phi_p(r,\xi), \tag{2.43b}$$

where $\Phi_p(r,\xi) = -\frac{m_e c^2}{2e}a^2(r,\xi)$ is the ponderomotive potential associated with the ponderomotive force of the laser pulse. We can recognize this equation as being the equation of a harmonic oscillator driven by the ponderomotive potential $\nabla^2 a^2$. The solution to these equations for a driver with a transverse Gaussian profile and a sinusoidal squared axial profile was derived in [50]. The result also proposed that the wakefield can be driven resonantly when the laser pulse axial length is half the length of the plasma wavelength. The radial size of the wakefield is on the order of the waist of the focused laser pulse. For a linearly polarized pulse, the axial electric field and the density perturbation are given by [56]

$$\delta n = -n_0 \frac{\pi a_0^2}{8} \left[1 + \frac{8}{k_p^2 w(z)^2} \left(1 - \frac{2r^2}{w(z)^2} \right) \right] \exp \left(-\frac{2r^2}{w(z)^2} \right) \sin k_p \xi, \tag{2.44a}$$

$$E_z = -E_0 \frac{\pi a_0^2}{8} \exp \left(-\frac{2r^2}{w(z)^2} \right) \cos k_p \xi. \tag{2.44b}$$

The radial electric field can be derived from the axial field using the Panofsky-Wenzel theorem and follows the same dependence as the density perturbations. From Equation (2.44a) and Equation (2.44b) we can highlight some of the salient features of the Wakefield. The Wakefield has a sinusoidal shape behind the laser pulse with a wavelength associated with the plasma frequency. Hence, a higher plasma electron density increases the plasma frequency, which decreases the plasma wavelength. The density perturbations and the radial field are $\frac{\pi}{2}$ phase-shifted with respect to the axial field. The axial field provides regions $\frac{\lambda_p}{2}$ for both acceleration and deceleration. However, because of the phase shift between the radial and axial fields off-axis electrons would only see a focusing and accelerating field for $\frac{\lambda_p}{4}$, which is the favorable regime for electron acceleration.

2.3.2. Nonlinear wakefields and bubble regime

The laser energies considered in this work have a vector potential $a_0 > 1$ and hence drive nonlinear wakefields where the assumption of small density perturbations no longer hold. Nevertheless, we can still follow the same procedure as we did in the linear case and derive the wave equation for the nonlinear wakefields [56, 57] using Equations (2.41a) to (2.41c)

$$\frac{\partial^2 \phi}{\partial \xi^2} = k_p^2 \gamma_p^2 \left[\beta_p \left(1 - \frac{1 + a^2/2}{\gamma_p^2 (1+\phi)^2} \right)^{-1/2} - 1 \right], \tag{2.45}$$

where ϕ is the potential associated with the wakefield, k_p is the plasma wave vector, $\gamma_p = \left(1 - \beta_p^2\right)^{-1/2}$ is the Lorentz factor associated with the phase velocity v_p of the plasma wave and $\beta_p = v_p/c$. If the velocity of the plasma electrons which are the constituents of the plasma wave becomes ultra relativistic we can simplify Equation (2.45) by using the following approximations

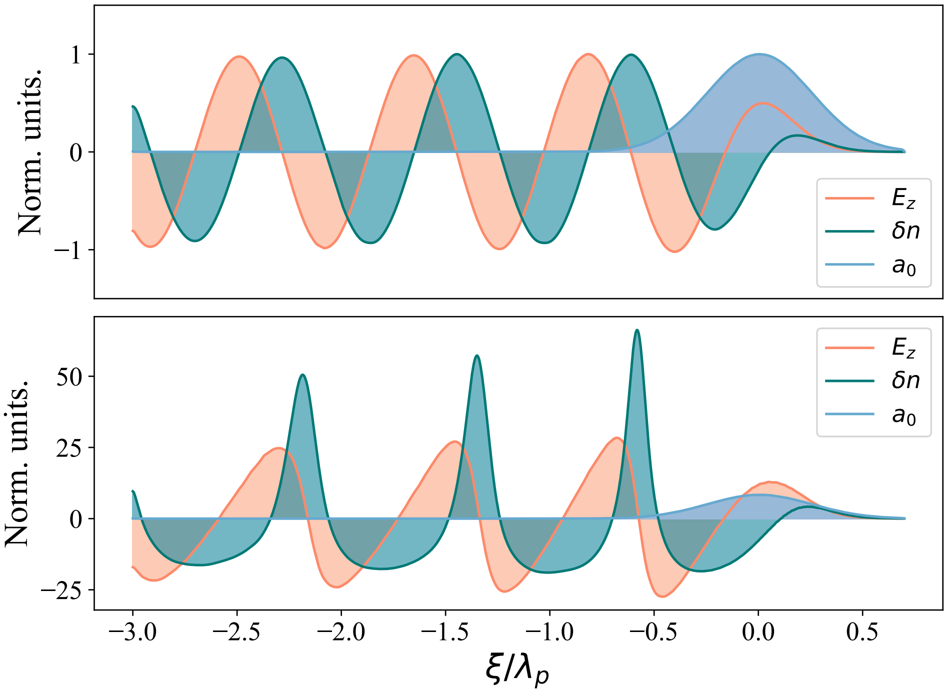

Figure 2.5.: The difference between the two regimes of linear and non-linear wakefields is depicted. The results are derived from FBPIC simulations where a laser with an a_0 of 0.3 is driving the wakefields. On the other hand the laser a_0 for the bottom figure is around 3.0, hence driving nonlinear wakefields. Both figures have the quantities normalized by those of the linear case. Hence for the nonlinear case the density perturbations are much larger than the nonlinear case resulting in a sawtooth-like axial field.

$$\left(1 - \frac{1 + a^2/2}{\gamma_p^2 (1+\phi)^2}\right) \approx 1 + \frac{1 + a^2/2}{2\gamma_p^2 (1+\phi)^2} \tag{2.46a}$$

$$\beta_p = \sqrt{1 - \gamma_p^{-2}} \approx 1 - \frac{1}{2\gamma_p^2}. \tag{2.46b}$$

Using Equation (2.46a) and Equation (2.46b) leads to a modified Equation (2.45)

$$\frac{\partial^2 \phi}{\partial \xi^2} = \frac{-k_p^2}{2}\left(1 - \frac{1 + a^2}{(1+\phi)^2}\right), \tag{2.47}$$

We can solve the Equations (2.45) and (2.47) numerically for a Gaussian shaped laser pulse and use the solution of the potential to calculate the radial and axial fields. In Figure 2.5 we see the difference between the linear and nonlinear wakefield regime where the axial fields and the density distribution is shown. The figure

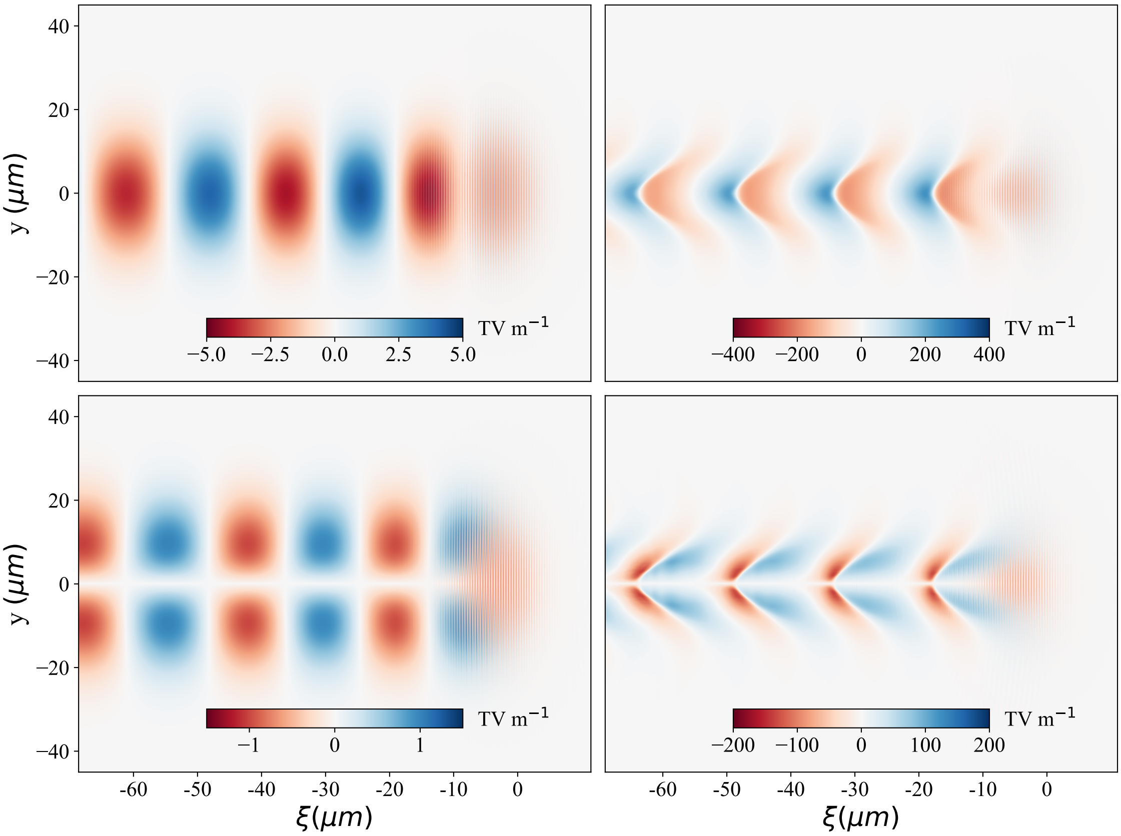

Figure 2.6.: Radial and axial electric fields are depicted in this figure with the left half showing the linear wakefield case ($a_0 = 0.3$) while the figures on the right half depicting the nonlinear wakefield case ($a_0 = 1.5$). The laser for the nonlinear case has been normalized for the sake of clarity, since on a similar scale it would appear stronger than the laser for the linear case.

was generated from simulating the wakefields using a Particle-in-cell code known as FBPIC that would be described in subsequent sections. The a_0 for the two cases is 0.3 and 1.5, however, for the second case the power is above the critical power required for self focusing and the pulse also compresses. This results in an a_0 of about 3.0 within the plasma that results in non-linear wakefields. Compared to the linear case the axial fields take on a "sawtooth" like shape where there is a linear increase of the field between the peaked density perturbations. In contrast to the linear wakefield, the nonlinear regime has a half plasma wavelength where the electrons are both focused and accelerated. This can be seen in more detail in Figure 2.6 where both axial and the radial fields are depicted. We also see that the wakefields in the nonlinear case are curved towards the laser. We previously discussed in Section 2.2.3 how the local variations in intensity can result in local variation in the relativistic mass of the electrons. With an increase in the relativistic mass the plasma wavelength increases. Hence, the electrons forming the plasma wakefield on axis have a larger plasma wavelength than the electrons which are radially far from the axis. This results in the curvature seen in the wakefields for the nonlinear case.

For a highly nonlinear wakefield with an $a_0 \gg 1$, all of the electrons around the laser axis are pushed out by the ponderomotive force of the laser pulse. The following analysis is a phenomenological treatment since a full 3D analytical derivation is not possible in the nonlinear case. When the area behind the laser pulse becomes devoid of electrons, it results in a blow-out or a bubble regime [58]. This ionic bubble is surrounded by a thin sheath of electrons that fall back to the central axis. This regime was studied using PIC simulations by Lu et al. [59–61] who found the extent of the bubble to be related to the a_0 of the driving pulse and the plasma wavelength

$$R_b = \frac{2\sqrt{a_0}}{k_p}. \tag{2.48}$$

This extent of the bubble radius is valid in the case when the laser waist is matched to the bubble radius $waist \approx R_b$. The axial and the radial fields inside the bubble vary linearly with z and r respectively and were also derived from PIC simulations

$$E_{z,max} = \frac{\omega_p m_e c}{e}\sqrt{a_0} \tag{2.49a}$$

$$E_r \approx \frac{\omega_p m_e c k_p}{4e} r. \tag{2.49b}$$

It can be seen from the above equations that the maximum axial fields can be increased by a higher laser intensity or by increasing the plasma electron density, which results in an increased plasma frequency. This particular regime is attractive for electron acceleration since the axial and radial fields are linear in the bubble with a highly nonlinear behaviour at the edges of the bubble, resulting in less beam distortion during acceleration. We will discuss the limitations of the energy that can be transferred to the electrons in Section 2.3.4.

2.3.3. Electron injection

Once the laser pulse has established the plasma wakefield structure, a mechanism is required to inject electrons into this accelerating region in order to exploit it for particle acceleration. In the experiments presented in this work, the electrons that are accelerated originate from the background plasma itself. These plasma electrons are trapped and injected into the wakefield under particular conditions that will be described in this section. Since an equation of the wakefield potential was derived in the last section Equation (2.47), we can use it to derive a Hamiltonian to examine the dynamics of individual electrons in the co-moving frame of the plasma wave [1]. In this frame, the wakefield appears quasi-static, and the electron trajectories can be described using the Hamiltonian of the system. The Hamiltonian $\mathcal{H}(\xi, p_z)$ governing the motion of a test electron in a one-dimensional, electrostatic plasma wave with normalized quantities is given by [62]

$$\mathcal{H}(\xi, p_z) = \sqrt{1 + a^2(\xi) + p_z^2} - \beta_p p_z - \phi(\xi) \tag{2.50}$$

where $p_z = p_z'/m_e c$ is the normalized longitudinal momentum, β_p is the normalized phase velocity and $\phi(\xi)$ is the normalized scalar potential of the plasma wave. This Hamiltonian describes the phase space trajectories in the (ξ, p_z) plane, and the boundary between trapped and untrapped orbits is known as the separatrix. Only electrons that cross into the region bounded by the separatrix can become trapped in the Wakefield and subsequently accelerated. The condition for trapping can be derived by comparing the electron's Hamiltonian $\mathcal{H}$ to the value of the Hamiltonian at the separatrix $\mathcal{H}_s$ given by

$$\mathcal{H}_s = \frac{\sqrt{1 + a^2}}{\gamma_p} - \phi_{min} \tag{2.51}$$

where $\gamma_p = 1/\sqrt{1 - \beta_p^2}$ and ϕ_{min} is the minimum of the normalized scalar potential. When the condition $\mathcal{H} \leq \mathcal{H}_s$ is achieved, the electrons are trapped in the Wakefield. This implies that the electron needs to have a longitudinal velocity that is greater than the velocity of the trailing end of the wakefield to be able to co-propagate with it. In Figure 2.7, the phase space structure of the wakefield along with the wakefield potential and the longitudinal field is shown in the comoving frame of reference. The separatrix is shown in red while a trapped orbit is shown in violet. Trapped electrons oscillate within the wake structure, periodically gaining and losing kinetic energy as they are alternately accelerated and decelerated.

In contrast, electrons whose initial Hamiltonian exceeds or falls short of the separatrix threshold, represented in black, remain untrapped. These electrons either outrun the wake or lag behind it in the co-moving frame. Their collective motion constitutes the plasma oscillations that sustain the wakefield structure. To make use of the accelerating structure of the wakefield, the acceleration process needs to be terminated before the electrons enter the decelerating phase of the wake.

Many of the injection mechanisms change the trajectories of the electrons to trap them inside the wakefield structure. From Equation (2.51), we can see a slower

[1] However, this analysis fails to describe practical laser-wakefield accelerators, which are often subjected to global perturbations that can change the structure of the wakefields.

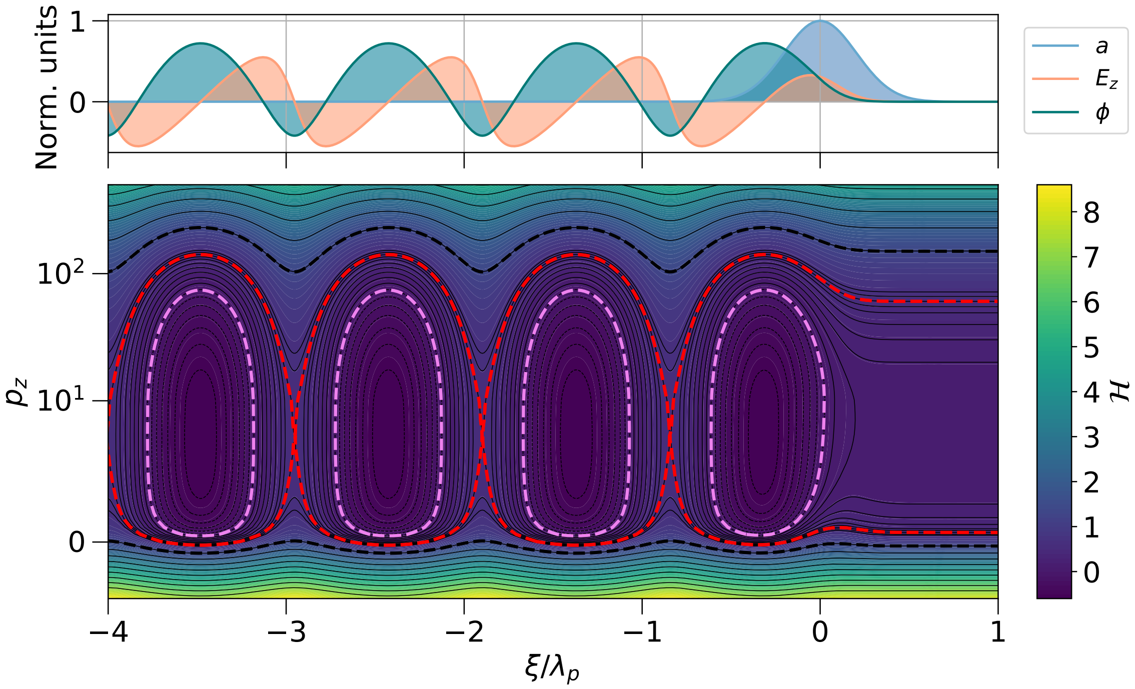

Figure 2.7.: The top part of the graph shows the different wakefield quantities with $a_0 = 1$. In the bottom part of the graph the phase space is depicted in the comoving frame of reference. The red contour defines the separatrix that distinguishes between trapped (violet) and untrapped (black) orbits.

phase velocity (smaller γ_p) and a higher wakefield amplitude (higher ϕ_{min}) both decrease the value of the separatrix Hamiltonian. This implies that trapping is easier when a larger wakefield is driven by a higher intensity laser (higher ϕ_{min}) or when the plasma electron density is increased, leading to lower phase velocities (see Equation (2.34)).

The injection mechanisms can both be uncontrolled, like self-injection [63], or controlled, such as density down-ramp injection [64], shock injection [65], ionization injection [66] and colliding pulse injection [67, 68]. The most relevant injection mechanisms will be described in the following section, which were used in this work.

Self-injection

Previously, we examined how increasing the laser intensity drives the plasma wake into the nonlinear regime (see Section 2.3.2). However, this process reaches a physical limit when the velocity of electrons oscillating in the plasma wave exceeds the phase velocity of the wakefield. At this point, the wakefield undergoes a breakdown, analogous to the breaking of a water wave near the shoreline, wherein the electric field reaches its maximum amplitude and can no longer sustain coherent wave motion. This phenomenon enables a subset of fast-moving electrons to acquire sufficient momentum to enter the accelerating phase of the wake, initiating the process known as *self-injection*. The mechanism outlined above corresponds to longitudinal self-injection, in which electrons originate near the laser axis and possess negligible transverse momentum. On the other hand, transverse self-injection occurs when the electrons that were pushed away by the ponderomotive force travel around the sheath of the electrons before being injected. This process of injection is experimentally easy to implement since only a laser of sufficient power is required to achieve the wakebreaking threshold. As we have already seen the laser pulse can self-focus and compress thereby increasing the a_0 which can in turn lead to self-injection. A threshold value can be derived for this self-injection process from the following expression [69, 70]

$$R_b k_p > 2\sqrt{\ln\left[\frac{2n_{crit}}{3n_e}\right] - 1}. \tag{2.52}$$

Using Equation (2.48) we can reformulate this expression to

$$2\sqrt{a_0} > 2\sqrt{\ln\left[\frac{2n_{crit}}{3n_e}\right] - 1} \implies a_0 > a_{0,si} = \ln\left[\frac{2n_{crit}}{3n_e}\right] - 1. \tag{2.53}$$

For experiments involving lasers with a central wavelength of $\lambda = 800nm \implies n_{crit} \approx 1.7 \times 10^{21}\text{cm}^{-3}$ and plasma electron densities in the range of $5 \times 10^{18}\text{cm}^{-3}$, the threshold value becomes $a_{0,si} \approx 4.5$. This value can be routinely achieved in typical LWFA experiments when taking into account the increase in the normalized vector potential due to self-focusing and self-compression within the plasma during pulse propagation. Since self-injection threshold depends on the plasma electron density, it can be minimized or completely stopped by lowering the electron density.

While this is one of the simplest injection mechanisms requiring no special target systems, it has a couple of disadvantages. Firstly, once the threshold value for the laser vector potential is achieved, the electrons continue to inject at the back of the bubble during the propagation of the laser pulse. The electrons injected at earlier stages of propagation remain in the bubble for a longer duration and hence are accelerated to higher energies in contrast to the electrons injected at later stages. This naturally creates electron beams with a relatively large energy bandwidth. Secondly, even more critical is the dependence of this scheme on the laser pulse evolution in the plasma and the consequent nonlinear process of wave breaking. This process is sensitive to small variations that arise from shot-to-shot fluctuations of the laser parameters making it unstable and difficult to control.

The controlled schemes of injecting electrons try to address these challenges by localizing the injection to reduce bandwidth and making the injection process independent of the pulse evolution in the plasma.

Density down-ramp injection

When a laser pulse is propagating in a density down ramp $\frac{dn}{dz} < 0$, it leads to an increasing bubble size from Equation (2.48) and a reduction in the local phase velocity of the wake from Equation (2.33). These effects can be exploited to inject an electron beam into the bubble and can be understood by looking at the local phase velocity [14]. The local phase of the wake is given by $\phi = k_p(z)(z-ct)$, where the influence of the density downramp on the group velocity of the laser pulse is considered to be minimal $\nu_g \approx c$. This can be used alongside the definitions of effective plasma frequency $\omega_{p,eff} = \partial_t \phi$ and wavenumber $k_{p,eff} = \partial_z \phi$ to yield the following local phase velocity

$$\nu_p = \frac{\omega_{p,eff}}{k_{p,eff}} \approx \frac{c}{1 + \frac{\xi}{k_p}\partial_z k_p} \approx c\left(1 - \frac{\xi}{k_p}\partial_z k_p\right) \approx c\left(1 - \frac{\xi}{2n}\partial_z n\right). \tag{2.54}$$

From the above equation, it can be seen that the phase velocity behind the laser pulse ($\xi < 0$) will decrease for a density downramp ($\partial_z n < 0$). This effect of reduction in the phase velocity is more prominent the larger the distance is behind the laser pulse and the steeper the density gradient. When the decreased local phase velocity approaches the velocity of the plasma electrons, injection can occur. This results in an injection mechanism that is independent of the pulse evolution. This density downramp can occur naturally at the end of the gas jet when the gas density tapers off [71, 72] or it can also be introduced through a shock wave in the gas density flow that leads to a special case of density down-ramp injection known as shock front injection.

Shock front injection

Shock front injection [65] is a specific type of density down-ramp injection where a sharp, localized decrease in plasma density is used to inject electrons into the Wakefield bubble. The localized reduction in the plasma density is achieved using a shock front introduced by perturbing the supersonic flow of gas from a de Laval nozzle [73] by a razor-sharp blade or a wire. This technique has a potential for stable production of narrow bandwidth and high charge electron beams [74, 75]. The characteristic of a narrow bandwidth arises because of the small timescales

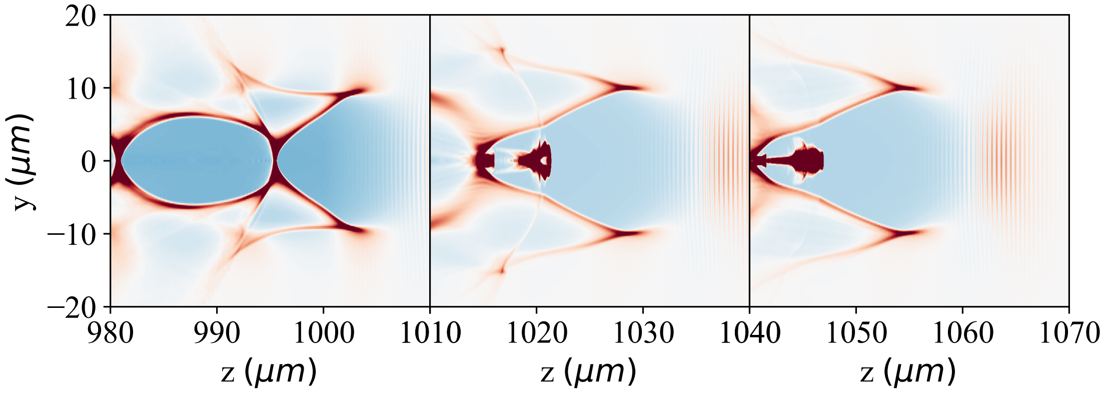

Figure 2.8.: The evolution of the bubble and injection of an electron beam is shown using three time snapshots when propagating through a density shock. In this instance, the normalized vector potential is approximately $a_0 \approx 2.5$. The density downramp from $8.8 \times 10^{18}\mathrm{cm}^{-3}$ to $5.5 \times 10^{18}\mathrm{cm}^{-3}$ occurs from $1000\ \mu m$ to $1010\ \mu m$. The time elapsed between each snapshot is about $89fs$. The plasma wavelength increases because of lower density and the electrons forming the sheath of the first bubble are trapped and then accelerated.

involved in the process and the small local region where the injection takes place. From Equation (2.48), we see that the size of the bubble is inversely proportional to the plasma electron density that implies a shorter bubble where the density is higher and a longer bubble when the density is low. The density in a shock front injection increases to a point of the shock after which it rapidly decreases. The electrons that make up the sheath at the back of the bubble in a high density region suddenly are captured at the back of the longer bubble as the density decreases in the sharp downramp.

Figure 2.8 shows this behavior where three snapshots from a laser Wakefield accelerator simulation is shown depicting the mechanism of shock front injection. In the first picture a bubble in the region of high density region is formed before the shock. The sharp density downramp occurs in the second graph where we see the electrons forming the back sheath of the bubble being captured. In the last image, these electrons are then accelerated. It can be noted that the length of the bubble is longer in the last image compared to the first one. If the density is tuned such that there is no self-injection before and after the shock, then the injection is only restricted to this small spatial and temporal region. The energy of the electron beam can be tuned by moving the shock along the gas jet resulting in variable acceleration length. Other parameters of interest that influence the shock front injection are the backing pressure of the gas jet and the movement of the blade or wire that induces the shock in the gas flow. These parameters can be actively controlled during an experiment resulting in beams with variable bunch properties such as energy spread, mean energy and pointing [76].

2.3.4. Acceleration limits

In the sections before, we outlined the generation of the plasma wakefields and the consequent electron injection and acceleration in the trailing bubble. In this section, we will outline the limitations of the process of electron acceleration in the laser wakefields. Since the group velocity of the laser is lower in the plasma than vacuum, eventually the electrons start to move forward in the bubble till they reach the middle of the bubble after which they start to decelerate since the axial field reverses sign (see Figure 2.6) in a process termed as the dephasing. Another effect known as depletion of the laser field is the reduction in the potential of the laser because of the transfer of energy to the wakefields. Lastly, the acceleration and the plasma wakefields can only be sustained when the laser has a high enough intensity. The fundamental process of diffraction limits the distance over which the wakefields can be maintained. These three effects will be elaborated in this section.

Dephasing

To describe the effect of dephasing, we need to look at the typical velocities of the laser pulses used in this work. The group velocity of a laser pulse with a central wavelength ($\lambda_L \approx 800$nm) propagating in a plasma with a plasma electron density $n_e \approx 10^{18}\text{cm}^{-3}$ is around $v_g \approx 0.9997c$ from Equation (2.36). In the comoving frame, an electron beam with an energy greater than 30 MeV would eventually pass the middle of the bubble and enter the region of the wakefield where it would experience the decelerating fields. The length in the laboratory frame which the electron travels before it enters the second half of the bubble is termed as the *dephasing length* L_{deph}. For a highly underdense plasma and the bubble regime results in an expression for the dephasing length [77]

$$L_{deph} \approx \frac{\omega^2}{\omega_p^2}\lambda_p \begin{cases} 1, & a_0 \ll 1 \\ \dfrac{\sqrt{2}a_0}{\pi}. & a_0 \gg 1 \end{cases} \tag{2.55}$$

The different regimes result from the increase of the plasma wavelength in highly nonlinear wakefields. From the above expression and the definition of the plasma frequency Equation (2.25), the dephasing length is inversely proportional to the plasma electron density $L_{deph} \propto n_e^{-3/2}$. Hence by decreasing the electron density, one can increase the dephasing length. For case of nonlinear wakefields $a_0 \gg 1$, the dephasing length is also increased by increasing the laser intensity. Although dephasing is an inherent problem to the laser Wakefield accelerators, attempts have been made to circumvent it by introducing a density upramp after injection [78, 79]. This shrinks the plasma wavelength allowing the electron to remain in the accelerating part of the bubble because of the accelerating wakefield phase.

Depletion

When the laser pulse excites the plasma wakefields, it constantly transfers energy from the laser to the plasma. This process eventually drains the driving laser potential to a point where it is no longer strong enough to drive plasma wakefields. As shown already in Section 2.2.3, the front part of the laser pulse driver sees the highest plasma electron density and hence it is the part that loses the most energy and etches away. The velocity at which this happens is given by [80]

$$v_{etch} \approx c\frac{\omega_p^2}{\omega^2}. \tag{2.56}$$

The etching starts from the front of the laser and propagates to the back. The time it takes for this effect to completely erode away a laser pulse is denoted by t_{etch} and the length is called the *Depletion Length* L_{dep}. Assuming a laser pulse with an axial length $c\Delta\tau_L$ and a highly underdense plasma $v_g \approx c$ the depletion length is given by

$$L_{dep} = ct_{etch} = c\frac{c\Delta\tau}{v_{etch}} = \frac{\omega^2}{\omega_p^2}c\Delta\tau. \tag{2.57}$$

From Equation (2.57), we can see that the depletion length is directly proportional to the pulse duration and is inversely proportional to the plasma electron density. Hence by reducing the plasma electron density, one can increase the depletion length. Another important effect of the etching is the decrease of the group velocity of the laser pulse, resulting in the electrons entering dephasing earlier and reducing the dephasing length. Both the depletion length and the dephasing length for the parameters considered in this work is on the order of a few mm.

Diffraction

Finally, the last limiting factor in accelerating the electrons is the natural diffraction of the laser beam. To attain high intensities that can generate the plasma wakefields the laser beam is focused and its size evolves according to Equation (2.20b), implying increasing size once the laser beam passes the focus spot at $z = 0$. The increase in size diminishes the normalized vector potential, thereby inhibiting the generation of wakefields. The focal spot size is generally not tunable during an experiment and is a fixed parameter of the facility, where it is chosen to result in high intensities and matched resonance conditions to drive wakefields. The acceleration can still occur over several Rayleigh lengths, beyond which the laser intensity becomes insufficient to sustain nonlinear wakefields. However, as discussed in Section 2.2.3 the laser can experience self-focusing if it is above a certain power threshold. This self-focusing effect can counteract natural diffraction, allowing the laser to maintain values of high a_0 for an extended length. Due to the laser depletion, the laser pulse eventually drops below the critical power required for self-focusing. After this point, the natural diffraction of the laser pulse dominates, reducing the intensities at the laser axis. One mechanism to overcome this limitation is to guide the laser pulse through the plasma using external guiding structures[81, 82]. The highest peak electron energies have been achieved using these guiding structures that were able to maintain the focal spot size of the laser on the order of tens of cm in length. In this work, to keep the experiment simpler, no complex guiding structures were employed since the main goal was to optimize the current available free parameters.

One primary goal of this work is to automate the procedure of finding good regimes of laser wakefield acceleration. From this chapter, we can see that the complex process of driving wakefields, injection, and acceleration can affect the resulting electron beams from the LWFA. The laser normalized vector potential a_0 is one of the most important parameters that has an impact on the acceleration of the electrons. This parameter locally varies in the region of a laser pulse focus relative to the target in experiments, one can change the value of a_0 at different positions within the plasma. Another important parameter that influences the acceleration of the electrons and the evolution of the laser pulse in the plasma is the plasma electron density. A lower density increases the dephasing and depletion length at the expense of weaker fields. In an experiment setting the density can be changed by changing the backing pressure of a gas nozzle or by changing the distance of the gas nozzle to the laser axis. In this work, all of the electrons are injected into the plasma wakefields using shock front injection, hence controlling the upramp and downramp profiles can also lead to widely different electron beams. Lastly, introducing chirp in the laser pulse by changing the second, third or fourth dispersion can lead to different temporal profiles of the laser pulse. This in turn can affect the value of a_0 at different positions within the plasma that can affect the electron beams. From the theoretical framework laid in this chapter, we identified these different parameters to optimize the electron beams coming out of a LWFA. This is by far not an exhaustive list of parameters but the ones that have a significant effect. In the next section we will discuss the basics of different components of Bayesian optimization that is the other major part of this work.

3. Bayesian Optimization with Multiple Objectives and Fidelities

In this chapter, the mathematical underpinnings of Bayesian optimization (BO) starting with basics of conditional probabilities are described. We then discuss the Bayes' rule, its importance in constructing Gaussian processes (GP) that serve as the surrogate model for our optimization schemes. After that, acquisition functions, which are the second core component of Bayesian optimization would be discussed. We will then introduce the novel trust-based optimization method that was proposed during this work and show its effectiveness compared to other advanced state-of-the-art BO methods. This chapter aims to help the readers, who encounter BO for the first time, understand the gist of BO methods.

The roots of Bayesian probability theory trace back to two early pioneers of statistics: Thomas Bayes and Pierre-Simon Laplace. Thomas Bayes is credited with formulating the foundational ideas of what we now call Bayesian inference [83]. He introduced a method for updating probabilities based on new evidence, a principle that underpins modern Bayesian analysis. Pierre-Simon Laplace greatly extended and applied Bayes' ideas to a wide range of scientific and practical problems, from celestial mechanics to demography [84, 85].

These developments were crucial in the early foundations of statistics, which today are dominated by two broad approaches, namely Bayesian statistics and frequentist statistics. In frequentist statistics, developed in the early 20th century, probabilities are interpreted as the long-run frequencies of events. This approach focuses on repeated sampling from a population and defines probabilities as limiting frequencies as the number of trials approaches infinity. Frequentist inference often emphasizes hypothesis testing, p-values, and confidence intervals, with no direct way to incorporate prior beliefs or uncertainty about parameters [86].

In contrast, Bayesian statistics interpret probabilities as degrees of belief, which can be updated as new evidence is acquired. In this framework, the probability of a hypothesis reflects the amount of our belief in the hypothesis, and these beliefs are continuously updated via Bayes' rule, which we will discuss later. The Bayesian approach offers a coherent way to model uncertainty and incorporate prior knowledge alongside new data [87].

In this chapter, we will focus on the Bayesian approach, particularly because it relates directly to Bayesian Optimization. By combining prior beliefs about a model or system with new observations, Bayesian methods provide a powerful framework for inference and decision making, which makes them well suited for optimizing expensive-to-evaluate black-box functions such as a Laser Wakefield accelerator.

3.1. Conditional probabilities

We will start by describing a formal model of a random process known as the probability space. This probability space consists of the sample space denoted by Ω, which is the set of all the possible outcomes of a random process. An event space $\mathcal{F}$ is another feature of the probability space, consisting of all events, where each event is a subset of the sample space. The final component is a probability function that assigns a probability to each event within the event space. Assuming an event $A \in \mathcal{F}$, we denote this probability function by $p(A) \in [0, 1]$, where 0 means event A is impossible and 1 means event A is certain. We can continue to define another event B similarly with a probability $p(B)$. In this context, we can now talk about the joint probability distribution $p(A, B)$ that quantifies the probability that both events A and B will occur simultaneously. If both events A and B are independent, we end up with $p(A, B) = p(A)p(B)$. However, in cases where event A and event B are not independent, the calculation of $p(A, B)$ would depend on the relationship between the two events. This leads us to the concept of conditional probability.

Conditional probability refers to the probability of event A occurring given that event B has already occurred. The joint probability then results in the expression [88]

$$p(A|B) = \frac{p(A, B)}{p(B)}. \tag{3.1}$$

Another concept that is related to the joint distribution is known as marginal probability distribution. Marginalization is the process of summing over the joint probability distribution to recover the individual probabilities of events. Continuing the example of events A and B, we can recover the marginal probability distribution of A by summing over all possible outcomes of B [89]

$$p(A) = \sum_{B} p(A, B). \tag{3.2}$$

For the case of a continuous probability distribution, the summation becomes an integral [90]

$$p(A) = \int p(A, B)dB = \int p(A|B)p(B)dB. \tag{3.3}$$

Marginalization effectively sums over all the possible ways B can occur while still considering A, giving us the total probability of A while marginalizing out B. From Equation (3.3), we can also think of marginal distribution as an average of conditional probability distribution over all possible values of event B. We can repeat the same procedure and get the probability distribution of B by marginalizing out A. This concept of marginalization is fundamental in Bayesian statistics, where we often deal with joint distributions of multiple variables and use marginalization to compute probabilities for subsets of variables.

To demonstrate the concept of conditional and marginal probabilities, consider an experiment in which we measure two related physical quantities, such as the position X and the momentum P_{mom} of a particle. We can write the joint probability distribution for these two random variables as $p(X = x, P_{mom} = p_{mom})$, where x is the realized or measured value of the random variable X and p_{mom} is the realized value of momentum. The probability that the particle is found at any position x while having a particular momentum p_{mom1} is given by the conditional probability distribution $p(X = x|P_{mom} = p_{mom1})$. If we are only interested in finding the particle at any position x regardless of the momentum it has, we can marginalize out the momentum. This is done by averaging out the conditional probability distribution for all possible values of P_{mom}, resulting in $p(X = x) = \int p(X = x, P_{mom} = p_{mom})dp_{mom}$.

Bayes' rule

Next, we will derive the Bayes' rule that is the underlying principle of the models used in this work. The rule generally outlines the procedure for updating the probability of a particular event A occurring, given some data labelled as the event B. We can repeat the application of the Bayes' rule to update the probabilities after attaining more data sequentially. From Equation (3.1) we can write the conditional probability of event B given A as $p(B|A) = \dfrac{p(A, B)}{p(A)}$. Since both equations have the joint probability distribution, we can use both equations to eliminate it and write [91]

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}, \tag{3.4}$$

where $p(A)$ is known as the prior probability that represents our belief of event A occurring before any data B has been observed. $p(B|A)$ is the likelihood that represents how likely the observed data B is given that A is true and $p(A|B)$ is the posterior distribution that represents the updated probability of A after taking into account the data B. The $p(B)$ is the marginal likelihood that ensures normalization of the posterior over all possibilities of A. Equation (3.4) is essential for Bayesian inference and modeling because it allows us to update our belief (the prior distribution) after observing new data (event B) to yield the new belief (the posterior distribution). Please note that in deriving Equation (3.4), we utilized the fact that the joint probability distribution is commutative by definition.

We can see the effects of applying Bayes' rule in the context of particle physics where we consider a colliding experiment to detect a new particle. Assume, we have a theoretical model that predicts the existence of a new particle with a very small probability. Based on this model and previous experimental results, the probability of the particle existing is $p(E) = 1\% = 0.01$, that is the prior probability. Moreover, the experimental detector system has a true positive rate of 90%, that implies that it can detect the particle if it exists with a probability of $p(D|E) = 0.9$. Naturally, this means that the detector can also signal a detection even when the particle does not exist $p(D|\neg E) = 0.1$.

Now, we assume that the experiment results in the detector giving a signal that could indicate the presence of the particle. We can use the Bayes' rule to update the probability that the particle exists based on this new detection signal. The only missing component for applying the Bayes' rule is $p(D)$, the marginal probability of the detector giving out a signal. This can be calculated by averaging the conditional probability of the detector giving a signal both when the particle exists and when it does not $p(D) = p(D|E)p(E) + p(D|\neg E)p(\neg E) = (0.9)(0.01) + (0.1)(0.99) = 0.009 + 0.099 = 0.108$. Now from Equation (3.4), we get $p(E|D) = 0.9 * 0.01/0.108 = 8.3\%$, thus after detecting a signal, the probability that the particle exists has increased from 1% to about 8.33%. While this is a significant increase, it still shows that, due to the high false positive rate and the initial low prior probability, the detection alone does not provide overwhelming evidence for the particle's existence. By incorporating both the prior knowledge from theory and the new data from the experiment, Bayesian inference provides a systematic way to quantify and refine uncertainty in scientific discovery.

If we repeat the experiment, we can now use this updated posterior distribution as the prior distribution for the next experiment. Assuming another detection, the probability that the particle exists rises from 8.3% to about 45%. This demonstrates how accumulating evidence for the particle's existence causes the posterior probability to increase and shift our belief.

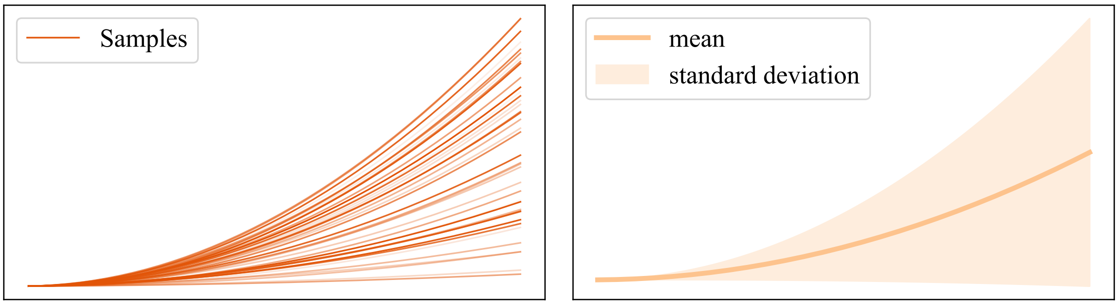

Figure 3.1.: The figure on the left shows a number of sample functions derived from $f_\theta(x) = \theta_1 x^2 + \theta_2 x$ after sampling the vector parameter $\boldsymbol{\theta}$ from a standard normal distribution $\mathcal{N}(0,1)$. For each particular value of vector $\boldsymbol{\theta_1}$ a complete function $f_1(x)$ is generated. This two step method of generating functions could be replaced by getting rid of the parametrization and directly defining a prior distribution over the functions as shown on the right. The complete distribution of the functions can be defined as a prior mean and an uncertainty which is shown here in the form of 2 standard deviation confidence intervals.

3.2. Gaussian process regression

In the last section, we explored the foundational concepts of conditional probability and Bayes' rule, which provide the theoretical framework for updating beliefs based on new data. These principles are at the heart of Bayesian inference, and their power becomes even more evident when applied to more complex models. One such model, the Gaussian process (GP), allows us to extend the Bayesian framework to continuous function spaces, enabling us to make predictions with uncertainty. In this section, we will delve into Gaussian process regression (GPR), a key tool in Bayesian Optimization, where we use probabilistic modeling to predict unknown functions and guide optimization strategies efficiently.

GPR is a non-parametric Bayesian approach to regression that allows us to make predictions about unknown functions in a probabilistic manner. At its core, GPR provides a flexible way to model complex relationships between inputs and outputs without assuming a fixed functional form. Instead, GPR models the distribution over possible functions that fit the observed data, capturing both the predicted mean values and the uncertainty around them. This makes GPR particularly powerful for tasks such as optimization, where understanding the uncertainty in predictions is crucial for making informed decisions. Before venturing further into the details of the GPR, we can first make an attempt to understand what a distribution of functions entails.

Assume that we have a function $f_\theta(x) = \theta_1 x^2 + \theta_2 x$ that is parameterized by a vector $\boldsymbol{\theta}$. A different parameter value $\boldsymbol{\theta}$ results in a slightly different version of the polynomial function. We can generate a random sample of functions by sampling the values of $\boldsymbol{\theta}$ from a standard normal distribution which has a zero mean and a unit variance $\mathcal{N}(0,1)$. These different values of $\boldsymbol{\theta}$ are then used in the polynomial equation to generate a function sample. This process of generating functions in two steps can be replaced by generating the functions directly as shown in Figure 3.1. We can define a prior distribution of functions and then sample from this distribution to obtain the same samples. This prior distribution is also non-parametric and as we will see later the shape of the functions is derived from the data itself instead of assuming a parameterized functional form.

A Gaussian process in the context of regression, defines a prior over functions that can be then updated using the available data. The GPR is fully characterized by a mean function $m(x)$ and a kernel function $k(x,x')$ [92]

$$f(x) \sim \mathcal{GP}(m(x), k(x,x')). \tag{3.5}$$

The mean function $m(x)$ captures our prior expectation of the value of the function at each point.

$$m(x) = \mathbb{E}[f(x)] \tag{3.6}$$

In many applications, the prior expectation of the function value is assumed to be zero $m(x) = 0$, if no specific prior knowledge exists about the function's behavior. This assumption that the function decays to zero at the boundaries simplifies the model and centers the GP around zero, allowing the data to inform the function's shape entirely. However, if prior knowledge about the function exists, it can be incorporated into a non-zero mean function. As a simplest case, if a linear trend is expected from the underlying function, then a mean function $m(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$, where $\boldsymbol{w}$ is a vector of coefficients and $\boldsymbol{x}$ is a vector of inputs, can be incorporated into the GPR.

3.2.1. Kernel functions

The kernel function, also known as the covariance function, is a fundamental component of GPR. It encodes our assumptions about the function that we want to learn, dictating the shape, smoothness, and general behavior of the functions that the GP can model. Essentially, the choice of kernel determines the class of functions that the GP can represent, and thus, GPR can only estimate functions that are consistent with the properties implied by the selected kernel. The kernel function $k(x,x')$ defines the covariance between the function values at two input points x and x'

$$k(x,x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))], \tag{3.7}$$

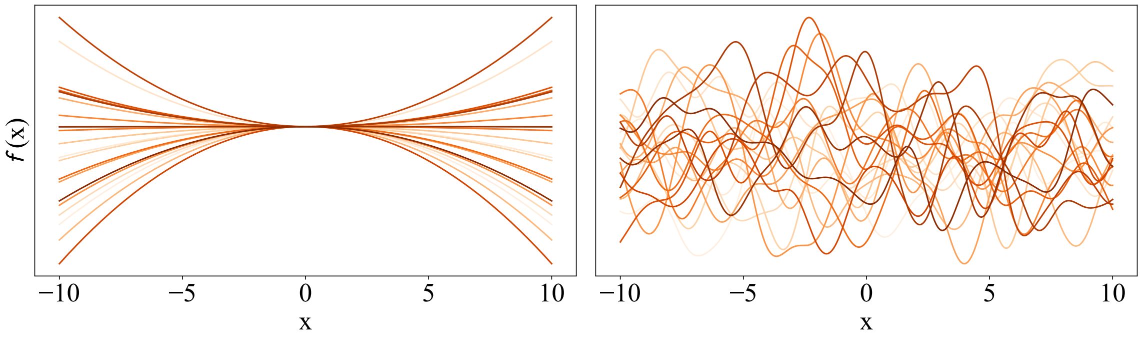

Figure 3.2.: The figure on the left shows a number of sample functions derived from a GPR that has a non-stationary kernel while the figure on the right shows the sample functions derived from a GPR with a stationary kernel. We can see that for the linear kernel case, the statistical properties of the function depend on the location in the input space since the further we are from $x = 0$, the higher the variance. In contrast with the non-stationary case, the GPR samples with the stationary kernel oscillate around the zero mean with similar variance.

where $m(x)$ is the mean function. It dictates how changes in the input space translate to changes in the function values, thereby shaping the GP's ability to fit data and generalize to new inputs. By choosing an appropriate kernel, we can tailor the GP to capture specific characteristics of the data, leading to more accurate modeling and prediction.

Kernel functions can be broadly classified into stationary and non-stationary kernels based on whether their properties depend solely on the relative positions of the input points or on their absolute positions. This implies that the statistical properties of the process are invariant under translations in the input space. More formally, a kernel $k(x, x')$ is stationary if it depends only on the difference $|x - x'|$ and not on the actual values of x or x'. Examples of stationary kernels include squared exponential, periodic and Matern kernels, among others. On the other hand, a kernel is non-stationary if it depends on the absolute positions of x and x'. Non-stationary kernels can model processes where the behavior changes over different regions of the input space and include linear or polynomial kernels. In Figure 3.2, we can see a global trend that arises through the use of a nonstationary kernel on the left portion of the figure. Another phenomenon is the increase in variance as the distance from $x = 0$ is increased. This implies varying levels of noise that is input-space dependent. On the other hand, with the use of stationary kernel we see that the function samples are smooth and always vary with a similar magnitude around the constant zero mean.

The kernel functions obey a number of nice properties that allows us to create new kernel functions from existing ones (see [92] for more details). The addition and product of two kernel functions result in a new valid kernel function. This applies even if the domain of both kernel functions is different since the new kernel function then operates on the product space. As an example, if we suspect the function that we are modeling to have some periodicity and a smooth structure on top, we can model it using an addition of Matern and a periodic kernel. We will now discuss some of the basic and simplest kernel functions that are used frequently.

The first of the kernels is the Squared Exponential (SE) also known as the Radial Basis Function (RBF) that is defined as [93]

$$k_{RBF}(x, x') = \sigma^2 \exp\left[\frac{-(x - x')^2}{2l^2}\right], \tag{3.8}$$

where σ is the signal variance and l is the length-scale hyperparameter that determines how quickly the correlations between function values decays with distance. The functions that result from this kernel function are infinitely differentiable and always smooth. Another reason for the popularity of the RBF kernel arises from its property of being a universal function approximator [94]. This implies that with appropriate hyperparameters, it can approximate any continuous function on a compact domain to arbitrary precision.

In some applications, a control over the smoothness of the functions is required which is the motivation for the Matern class of covariance functions defined as

$$k_{Matern}(x, x') = \sigma^2 \frac{2^{1-v}}{\Gamma(v)} \left(\frac{\sqrt{2v}|x - x'|}{l}\right)^v K_v \left(\frac{\sqrt{2v}|x - x'|}{l}\right), \tag{3.9}$$

where K_v is the modified Bessel function, $\Gamma(v)$ is the gamma function and $v > 0$ is the smoothness parameter. The degree of smoothness is governed by the parameter v which defines the number of times the function can be differentiated. The most common values for v are $3/2$ and $5/2$ implying once and twice differentiable functions and are denoted as Matern $3/2$ and Matern $5/2$ kernels respectively. For the case of $v = \infty$ the Matern kernel approaches the RBF kernel.

Finally we will outline another kernel that has a repeated correlation that does not vanish to zero with increasing distance between x and x'. This is the periodic kernel and is defined as

$$k_{Periodic}(x, x') = \sigma^2 \exp\left[-\frac{2\sin^2(\pi|x - x'|/p)}{l^2}\right], \tag{3.10}$$

where p is the period of the function, l is the length-scale parameter and σ is the signal variance. This type of kernel can be used with other kernels to model a

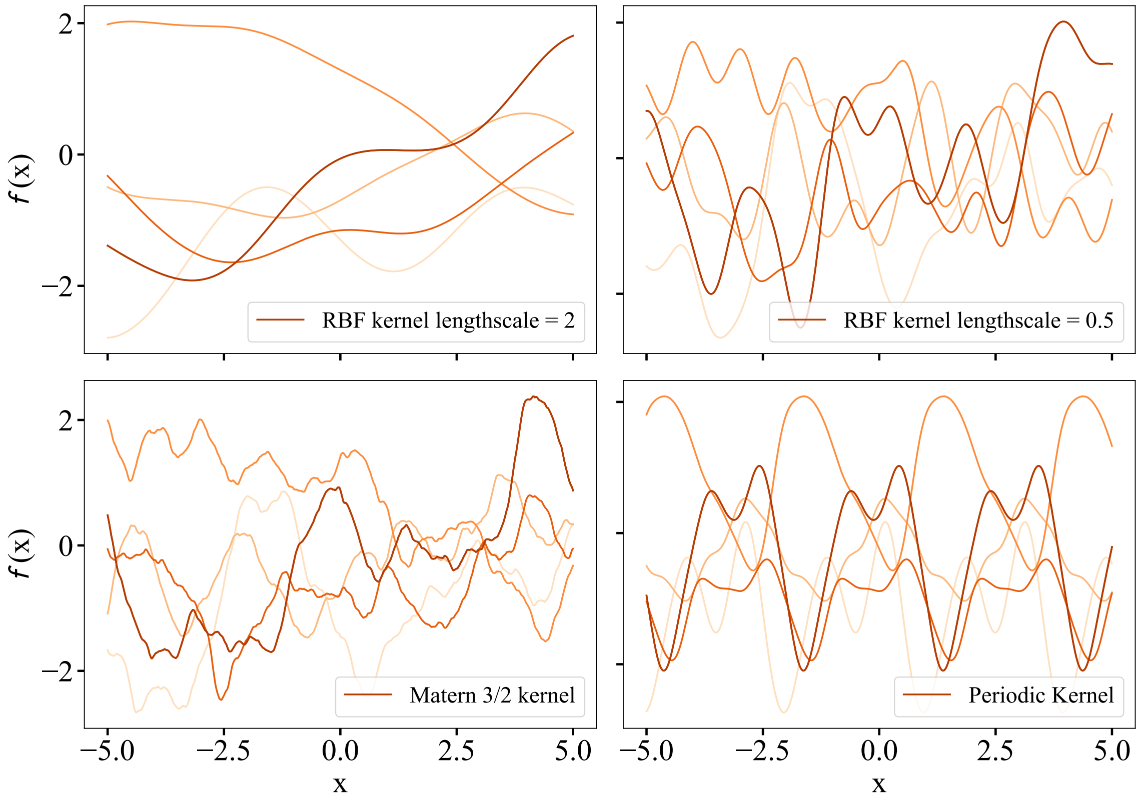

Figure 3.3.: The effect of choosing different kernels and a kernel with a different hyperparameter of lengthscale is shown. On the top left is a RBF kernel with a large lengthscale compared to the top right where a smaller lengthscale RBF is shown. We can see that even though the smaller lengthscale RBF varies faster than the larger lengthscale RBF, it is still quite smooth. On the bottom left is the Matern 3/2 which can describe less smooth functions. The periodic kernel on the bottom right shows functions that repeat after a certain period.

smooth function that is periodic on a longer timescale and smooth on a shorter timescale.

Some samples derived from a GPR using the kernels described above can be seen in Figure 3.3 where also the effect of a different length scale in the RBF kernel can be seen. All of these sample functions are derived from the prior distributions governed by the kernel functions. However, we are more interested in how these functions change when we acquire data from our true function through evaluations.

3.2.2. Posterior distribution

In previous section, we talked about how our prior beliefs can be incorporated in the framework of a GPR. In this section, we will consider observations through which we update our prior distribution to result in a posterior distribution, which reflects both our prior beliefs and the information provided by the data. Let us consider a dataset consisting of input-output pairs $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i=1}^n$ where $\boldsymbol{x}_i \in \mathbb{R}^d$ is the i-th input vector defined on a d-dimensional space and $y_i \in \mathbb{R}$ is a noisy observation of the underlying black-box function $f(\boldsymbol{x}_i)$. We will consider the case where y becomes a vector of output values in a later section when discussing multi-objective problems. For now, we restrict ourselves to the scalar output case

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{3.11}$$

where ϵ_i represents independent and identically distributed (i.i.d.) Gaussian noise with zero mean and variance σ^2. The goal for the GPR is to use the existing dataset $\mathcal{D}$ to make predictions about the function f at new input points. Since the prior distribution is used to build a GP, the function values follow a multivariate normal distribution. Specifically, for the input points $\mathbf{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}]$, the prior distribution of function values $\mathbf{f} = [f(\boldsymbol{x_1}), f(\boldsymbol{x_2}), ..., f(\boldsymbol{x_n})]$ is

$$\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K}), \tag{3.12}$$

where $\mathbf{m} = [m(\boldsymbol{x_1}), m(\boldsymbol{x_2}), ..., m(\boldsymbol{x_n})]$ is the mean vector and $\mathbf{K}$ is the $n \times n$ covariance matrix with entries $K_{ij} = k(x_i, x_j)$ derived from the kernel function. Since we assume that there is inherent or external noise present in the system, we can deduce the distribution for the noisy evaluations from Equation (3.11), resulting in

$$\mathbf{y} \sim \mathcal{N}(\mathbf{m}, \mathbf{K} + \sigma^2\mathbf{I}), \tag{3.13}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix.

Now, we consider a new position in the input space $\boldsymbol{x_*}$ where we want to probe our black-box function. Using GPR, we can calculate the distribution of $f_* = f(\boldsymbol{x_*})$ at this point given the observed data $\mathcal{D}$

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ m_* \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma^2\mathbf{I} & \mathbf{k}_* \\ \mathbf{k}_*{}^{\mathrm{T}} & k_{**} \end{bmatrix}\right), \tag{3.14}$$

where m_* is the mean function evaluated at $\boldsymbol{x_*}$, $\mathbf{k}_* = [k(x_1, x_*), k(x_2, x_*), ..., k(x_n, x_*)]$ is the covariance vector between x_* and the training inputs $\mathbf{X}$ and k_{**} is the variance at position x_*. To find the posterior predictive distribution of f_*, given $\mathbf{y}$, we make use of the conditional distribution

$$f_* | \mathbf{y}, \mathbf{X}, \boldsymbol{x_*} \sim \mathcal{N}(\mu_*, \sigma_*^2), \tag{3.15}$$

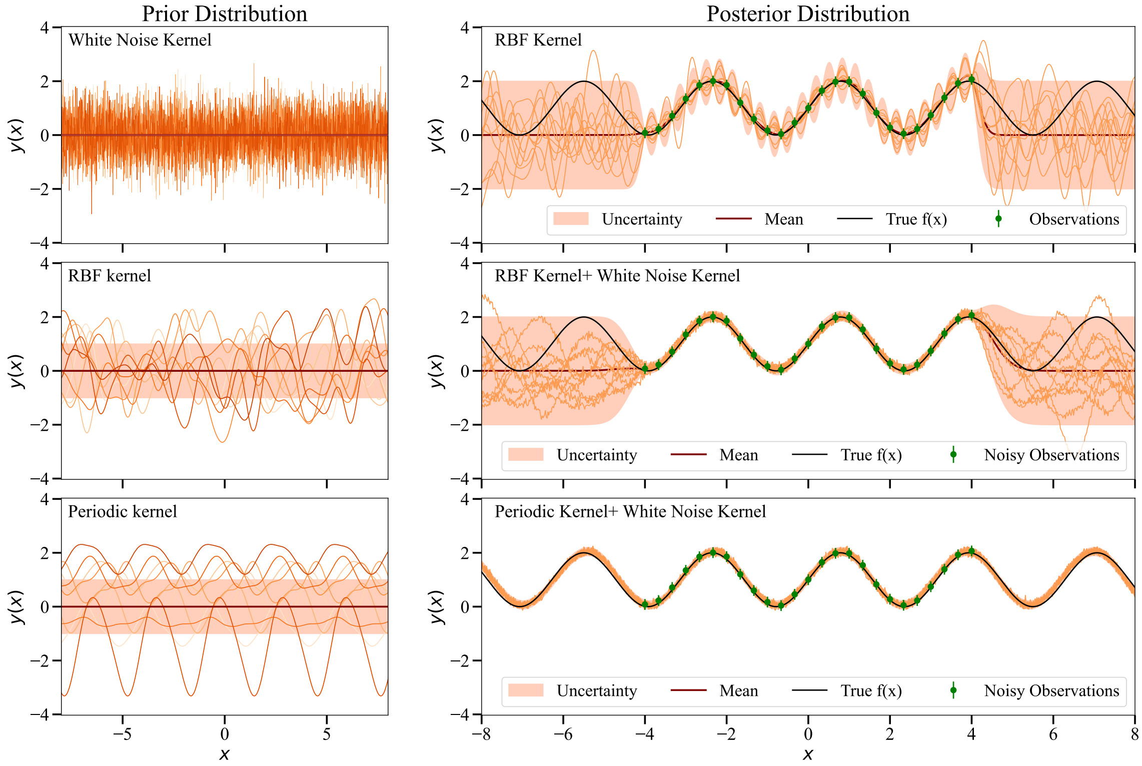

Figure 3.4.: The effect of choosing different prior kernel functions and the data on the predictive mean and the predictive variance is outlined. On the left are the prior distributions generated from different kernel functions. The top is the white noise kernel that assumes no correlation between two points x and x'. The second is an RBF kernel that generates smooth functions while the periodic kernel has functions with repetitive behavior. On the right are the posterior distributions with top with a RBF only kernel and we can see that the variance grows rapidly because the observations are noisy. By including a noise kernel with the RBF we see an improvement in the middle graph. In the regions where no data is observed we see that the periodic kernel is better since it can learn the periodic shape and extrapolate it to regions with scarce or no data.

where the μ_* is the predictive mean and σ_*^2 is the predictive variance and are given as

$$\mu_* = m_* + \mathbf{k_*}^{\mathrm{T}}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{m}), \tag{3.16}$$

$$\sigma_*^2 = k_{**} - \mathbf{k_*}^{\mathrm{T}}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{k}_*. \tag{3.17}$$

The predictive mean represents our best estimate of the function value at $\boldsymbol{x}_*$ given the observed data and the prior whereas the predictive variance quantifies the uncertainty associated with this prediction. This variance includes both the uncertainty inherent in the predictions of the GP model and also includes the uncertainty arising from the noise in the observations. An example of using different kernels in the GPR and how that influences the prior and the posterior distribution is shown in Figure 3.4. The figure shows how the data and the prior both influence the posterior distribution. Since our assumption of the underlying function being periodic is correct, the choice of periodic kernel is better for extrapolation outside the boundaries where no data is available. The RBF kernel also fits this periodic function nicely because of its universal estimator properties but fails to extrapolate into the regions where no data is available. The uncertainty between data points also increases for the only RBF kernel case, but it is considerably reduced when a white noise kernel is added to the RBF kernel shown in the middle plot.

3.2.3. Hyperparameter selection

As we saw in the last section, the performance and predictive capabilities of the GP model are heavily influenced by the choice of hyperparameters associated with the mean and kernel functions. Hyperparameters include parameters such as the length-scale l, signal variance σ in the kernel function, as well as any parameters in the mean function. Selecting appropriate hyperparameters is crucial for accurately capturing the underlying patterns in the data and making reliable predictions. In general, smaller length-scales can allow the GPR to model rapid changes while a larger length-scale enforces smoother variation. The signal variance controls the magnitude of the deviations from the mean function.

The hyperparameters of a Gaussian process are typically learned by maximizing the marginal likelihood (also known as the evidence) of the observed data under the GP model. The marginal likelihood integrates over all possible functions that could explain the data, weighted by their prior probability under the GP. Again considering a dataset $\mathcal{D} = (\boldsymbol{x}_i, y_i)_{i=1}^n$ and for simplicity assuming a zero mean function, the marginal likelihood is the probability of observing the outputs $\mathbf{y} = [\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_n}]$ given the inputs $\mathbf{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}]$, and the hyperparameters $\boldsymbol{\theta}$ of the kernel function

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{f}, \mathbf{X}, \boldsymbol{\theta}) p(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta}) d\mathbf{f}, \tag{3.18}$$

where $p(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta})$ is the GP prior over the true function values $\mathbf{f}$ and $p(\mathbf{y}|\mathbf{f}, \mathbf{X}, \boldsymbol{\theta})$ is the likelihood of the data given the function values accounting for the noisy observation. The logarithm of the marginal likelihood usually termed as the log marginal likelihood is often used because of numerical stability and computational convenience and is given by [95]

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^{\mathrm{T}}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \sigma^2\mathbf{I}| - \frac{n}{2}\log 2\pi \tag{3.19}$$

where $|\mathbf{K} + \sigma^2\mathbf{I}|$ denotes the determinant of the covariance and noise matrix. The first term in the above expression measures the distance of the data from the model mean (here assumed 0) and becomes zero when the outputs align perfectly with the model predictive mean. This term can be affected by changing the hyperparameters since the matrix $\mathbf{K}$ is different for different hyperparameters. The second term penalizes complex models since it prevents the problem of overfitting where the model can perfectly model the observed data but is unable to generalize and predict new data accurately. Models with smaller length scales yield small variances that can result in small determinant values. The log of such small determinants result in large negative values hence decreasing the log marginal likelihood values. The third term is a normalizing term since it does not depend on the hyperparameters of the model. Since we need a model that is relatively simple and can model the data accurately the log marginal likelihood is maximized. The final choice of the hyperparameters are those that maximize the log marginal likelihood [96]

$$\theta^* = \arg\max \log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}). \tag{3.20}$$

In practice, generally the gradients of the log marginal likelihood are available and hence many of the gradient-based methods can be used to solve this optimization problem.

We will end this section on GPR by outlining one core problem with using GPs for functions which have a high dimensionality ($n \approx 100$) in the range of hundreds. To find the predictive mean and the predictive variance from the GPR we have to invert a $n \times n$ matrix. This requires a time that scales with the number of samples $O(n^3)$, where O implies the asymptotic time complexity. For higher dimensions, a larger number of samples is required to accurately model the underlying function hence the calculations involved in building a GP become prohibitively expensive [92]. There have been attempts to implement GPs for such high spaces using sparse approximation methods [97] but those are out of the scope of this work since the dimensionality in this work is restricted to under 10.

3.3. Acquisition functions

Bayesian Optimization leverages the predictive capabilities of GPR to make informed decisions about where to next sample the black-box function also known as the objective function. The key idea is to use the posterior distribution provided by the Gaussian process to quantify the uncertainty in the function's behavior across the input space. This uncertainty quantification enables the optimization process to balance the exploration of unexplored regions with the exploitation of areas likely to contain the optimum. This exploration-exploitation tradeoff is a key feature of the acquisition functions. We already discussed how the GPR are constructed and can efficiently model the underlying objective function. The next core component of the Bayesian optimization (BO) is called the acquisition function, which are a class of functions that act on the predictive mean and variance from the GPR and result in a new position x^*. We will briefly outline the general BO loop here before diving into the specific flavours of acquisition functions that are traditionally used and the new functions proposed during this thesis work.

The BO loop consists of initializing by taking n random or grid measurements of the objective function resulting in a dataset consisting of n input pairs (x_n, y_n). Using this dataset a GPR is fitted to it by optimizing the log marginal likelihood and generating optimized hyperparameters. This GP provides a posterior mean function and a posterior variance at each point in the input space. An acquisition function then uses the posterior mean and variance to determine the next point x_{n+1} at which to evaluate the objective function. The objective function is then evaluated at this selected point to yield the output y_{n+1}. This new data point (x_{n+1}, y_{n+1}) is then added to the dataset and the GPR model is augmented with this new dataset. Another optimization of the hyperparameters is performed to better fit the data and the process of selecting a point is then carried out. This process is repeated iteratively until either the optimization budget is fulfilled or some predefined stopping criterion is reached. This general loop is outlined in Figure 3.5 where also the different acquisition functions values are displayed that would be discussed in the next sections.

The acquisition function plays an important role in the BO loop as demonstrated by the different recommendations in Figure 3.5. It evaluates potential points in the input space and quantifies their utility in terms of contributing to the optimization goal. In the following sections, we will explore various acquisition functions in detail and examine how they influence the optimization process. Understanding acquisition functions is crucial for tailoring Bayesian Optimization to specific problems and achieving optimal results.

3.3.1. Single-objective, single-fidelity acquisition functions

Acquisition functions encompass a broad class of functions that can be constructed using the posterior distribution of the GP model. This acquisition function is then optimized, often using gradient methods, to identify its maximum, which corresponds to the next evaluation point for the black-box objective function. The choice of the acquisition function significantly impacts the convergence of Bayesian optimization towards the global optimum. This section provides an overview of some widely recognized acquisition function policies in the existing literature. We would like to note that the development of acquisition functions is still an active area of research, with some recent contributions using for instance distance corre-

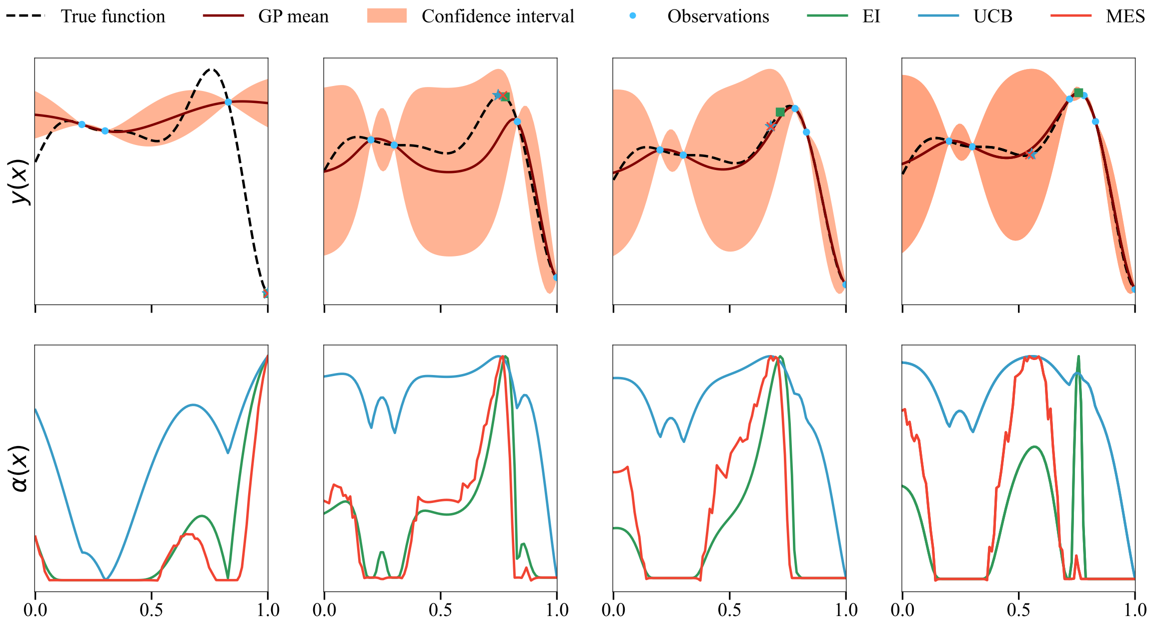

Figure 3.5.: Iterations of a Bayesian optimization loop are shown in this figure. On the top the true function, posterior GP mean, posterior GP variance, observations of the objective function and the maximum of three different acquisition functions is shown. On the bottom graphs the values of the acquisition functions across the input domain is shown. The maximum of these acquisition functions shown in the top plot are the positions that are recommended by the BO loop. In this loop the recommendation of the EI acquisition function is taken. We can see that the BO was able to reconstruct the true function near the peak in about 3 iterations starting from 3 random data points.

lation [98] or deep neural networks [99]. The design of the acquisition functions is also influenced by the type of the problem that they intend to solve. We consider first the simplest case of an objective function that is scalar-valued and has the same cost of evaluation. Advanced cases where the objective function output is a vector or the case where approximate values of the objective function can be gained with a relatively cheaper and varied cost will be discussed later.

Assuming that the objective function has been evaluated n times generating a data set $D_n = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), ..., (\boldsymbol{x}_n, y_n)\}$, these acquisition functions propose the optimal choice for the subsequent evaluation point, denoted here as $\boldsymbol{x}_{n+1}$. This selection is achieved through the optimization of a derived metric, taking into account the information from previously evaluated points. Specifically, the next evaluation point is given by $\boldsymbol{x}_{n+1} = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \alpha(\boldsymbol{x})$, where $\mathcal{X}$ is the input parameter domain and $\alpha(\boldsymbol{x})$ represents the acquisition function.

One of the widely used acquisition function is the upper confidence bound (UCB) policy [100], which is alternatively referred to as the lower confidence bound for minimization tasks. The UCB acquisition function is expressed as:

$$\alpha_{UCB}(\boldsymbol{x}) = \mu_n(\boldsymbol{x}) + \kappa\sigma_n(\boldsymbol{x}), \tag{3.21}$$

where κ is a parameter that balances exploration and exploitation, $\mu_n(\boldsymbol{x})$ and $\sigma_n(\boldsymbol{x})$ denote the GP mean and the GP standard deviation under the posterior distribution $p(y|\boldsymbol{x}, D_n)$, respectively. The hyperparameter κ is used to balance exploration and exploitation. A larger value of this hyperparameter encourages regions with a higher degree of uncertainty while a smaller value prefers further exploitation of explored regions. UCB is popular due to its simplicity, lower computational cost and effectiveness in practice.

Another well-known acquisition function in Bayesian optimization is Expected Improvement (EI) [101], which suggests the next evaluation point based on the expected improvement over the current optimal objective value f^*. The EI acquisition function is expressed as:

$$\alpha_{EI}(\boldsymbol{x}) = \mathbb{E}_{f(x)}[\max(f(\boldsymbol{x}) - f^*, 0)], \tag{3.22}$$

and with the GP posterior distribution $f(x) \sim \mathcal{N}(\mu(x), \sigma^2(x))$, α_{EI} can be computed analytically

$$\alpha_{EI}(\boldsymbol{x}) = (\mu(x) - f^*)\Phi\left(\frac{\mu(x) - f^*}{\sigma(x)}\right) + \sigma(x)\phi\left(\frac{\mu(x) - f^*}{\sigma(x)}\right), \tag{3.23}$$

where, $\Phi(.))$ is the cumulative distribution function (CDF) and $\phi(.))$ is the probability density function. The term $\mu(x) - f^*$ measures the magnitude of improvement that can be gained over the current best value f^* while the CDF measures the probability of improvement. Higher uncertainty can also increase the value of α_{EI} depicting that the acquisition function can potentially find better values of the objective functions by exploring. Like UCB, EI also is computationally cheaper to evaluate since it has a closed-form analytical expression.

A variation of EI is the Knowledge Gradient (KG) acquisition function [102, 103], which relies entirely on the posterior model. KG selects the next evaluation point based not on the best observable value but on the best value of the posterior mean. The KG acquisition function is given by:

$$\alpha_{KG}(\boldsymbol{x}) = \mathbb{E}\left[\max_{\boldsymbol{x}'\in\mathcal{X}}(\mu_{n+1}(\boldsymbol{x}')) - \max_{\boldsymbol{x}'\in\mathcal{X}}(\mu_n(\boldsymbol{x}'))\right] \tag{3.24}$$

where the terms $\mu_{n+1}(\boldsymbol{x})$ and $\mu_n(\boldsymbol{x})$ represent the posterior mean of the Gaussian process (GP) after $n+1$ and n evaluations, respectively. The term $\mu_{n+1}(\boldsymbol{x})$ represents the posterior mean of the GP after including an additional data point

$(\boldsymbol{x}, f(\boldsymbol{x}))$, where $f(\boldsymbol{x})$ is the realization of the stochastic process at location $\boldsymbol{x}$. This captures the updated belief about the function based on the new data point. The term $\mu_n(\boldsymbol{x})$ represents the prior belief before the new data point is included. KG is more exploratory than EI as it is influenced by posterior changes throughout the domain. On the other hand, maximizing the α_{KG} requires solving an inner optimization problem for each candidate x and thus is associated with higher computational cost.

Information-theoretic acquisition functions utilize the mutual information $I(\boldsymbol{x}; \boldsymbol{x}^*)$ between a specific location in the parameter domain and the observed data set. One such function is Entropy Search (ES) [104], represented mathematically as:

$$\alpha_{ES}(\boldsymbol{x}) = H(p(\boldsymbol{x}^*|D_n)) - \mathbb{E}(H(p(\boldsymbol{x}^*|D_n, \boldsymbol{x}, f(\boldsymbol{x})))), \tag{3.25}$$

where H denotes Shannon's entropy and $p(\boldsymbol{x}^*|D_n)$ refers to the probability distribution for the position of the maximum given currently observed data. Moreover, $p(\boldsymbol{x}^*|D_n, \boldsymbol{x}, f(\boldsymbol{x}))$ is the probability distribution for the position of the maximum given data D_n as well as $\boldsymbol{x}$ and $f(\boldsymbol{x})$, where $f(\boldsymbol{x})$ is drawn from the GP trained on D_n. The left term in the equation represents the entropy of the posterior distribution of the maximizing location $\boldsymbol{x}^*$, while the right term depicts an expectation over the entropy of the posterior after an additional sample. In higher-dimensional input spaces, evaluating the mutual information between the point to be queried and the maximizing location $\boldsymbol{x}^*$ becomes challenging. To address this, computationally more efficient variations have been introduced. Max-value entropy search (MES) [105] utilizes the mutual information between the maximum value y^* rather than the maximizing location $\boldsymbol{x}^*$. The MES acquisition function is formulated as:

$$\alpha_{MES}(\boldsymbol{x}) = H(p(\boldsymbol{y}^*|D_n)) - \mathbb{E}(H(p(\boldsymbol{y}^*|D_n, \boldsymbol{x}, f(\boldsymbol{x})))). \tag{3.26}$$

MES offers comparable or even better performance than ES while being significantly faster to compute [105]. It should be noted that MES may perform suboptimally in certain conditions since it assumes noiseless observations only [106]. In Figure 3.5, we can see the three different acuqisition functions EI, UCB and MES and how they differ when suggesting the new point. One can observe that the UCB and MES favor exploration of the input space more than EI, which is a greedier acquisition function and thus in higher dimensions with large spaces, MES and UCB would perform better.

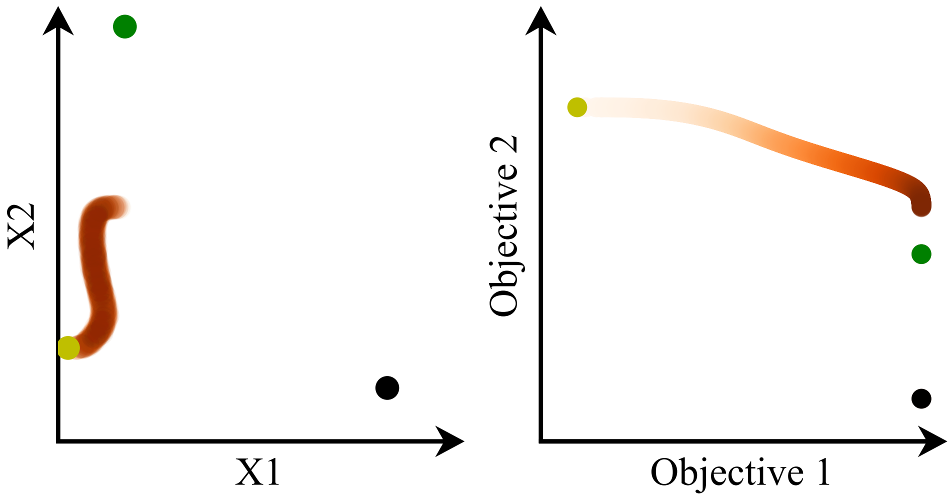

Figure 3.6.: Pareto front. Illustration of how a multi-objective function $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{y}$ acts on a two-dimensional input space $\boldsymbol{x} = (x_1, x_2)$ and transforms it to the objective space $\boldsymbol{y} = (y_1, y_2)$ on the right. The Pareto front shown on the right in blue is the ensemble of points that dominate others, meaning points that give the highest combination of y_1 and y_2. The corresponding set of coordinates in the input space is called the Pareto set shown in red on the left. Note that both the Pareto front and the Pareto set may be continuously defined locally, but can also contain discontinuities when local maxima get involved. In this example, f is a modified version of the Branin-Currin function from [107, 108] that exhibits a single, global maximum in y_2 but multiple local maxima in y_1. The different colors denote these optimums and have the same color in the input space as in the output space.

3.3.2. Multi-objective, single-fidelity acquisition functions

As we outlined at the start of this section, there exist many use cases in which the objective consists of multiple sub-goals. In this case, the function that is being maximized can be expressed as a vector of functions

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} f^{(1)}(\boldsymbol{x}) \\ f^{(2)}(\boldsymbol{x}) \\ \dots \end{pmatrix}$$

that are evaluated to yield output vectors

$$\boldsymbol{y}(\boldsymbol{x}) = \begin{pmatrix} y^{(1)}(\boldsymbol{x}) \\ y^{(2)}(\boldsymbol{x}) \\ \dots \end{pmatrix} .$$

These multi-objective optimization problems involve trade-offs between conflicting objectives, and the goal is to find solutions that balance these objectives effectively.

One can think of these sub-goals as a vector of solutions, which has to be reduced to a scalar number to be compatible with conventional single-objective acquisition functions. Before exploring the different ways to scalarize or to directly solve the multi-objective optimization problem, we will first introduce some concepts that make meaning of multi-objective optimization clearer. The optimum solution to the multi-objective problem consists of a set of solution vectors each of which defines the value of individual functions. A useful concept to describe the set of solutions is thinking of all the solutions being points in the multi-dimensional output objective space. This concept is termed as the *Pareto efficiency* [109], which is visualized as the *Pareto front* ($\mathcal{P}$). To describe the Pareto front, the notion of *domination* is defined as following:

Definition 1. A point $\mathbf{p}_1 = (y_1, y_2, \ldots, y_n)$ in an n-dimensional output space is defined as *non-dominated* if there does not exist another point $\mathbf{p}'_2 = (y'_1, y'_2, \ldots, y'_n)$ such that $\mathbf{p}'_2$ satisfies both:

$$\bigwedge_{i=1}^{n} (y'_i \geq y_i) \qquad \text{and} \qquad \bigvee_{i=1}^{n} (y'_i > y_i),$$

where $\bigwedge_{i=1}^{n}$ represents the logical AND operation applied across all n conditions, indicating that all inequalities $y'_i \geq y_i$ must hold simultaneously for $i = 1, 2, \ldots, n$. The $\bigvee_{i=1}^{n}$ represents the logical OR operation, meaning that at least one of these inequalities must be strict.

The Pareto front is composed of a set of *non-dominated* points in the output space as shown in Figure 3.6. All of the points and solutions that lie under the Pareto front are termed as the dominated solutions. Thus the solution of a multi-objective problem is to find the complete Pareto front. Several different methodologies exist to tackle the problem of multi-objective optimization, the simplest of these is scalarization of multiple objectives. Scalarization methods in general transform the vector of multiple objectives into a single scalar objective function. This allows the use of traditional single-objective optimization techniques to solve multi-objective problems. Here, we provide a detailed description of some common scalarization methods, including the weighted sum, weighted product, and epsilon-constraint method.

The first method to scalarize a vector objective function with m outputs is known as the weighted sum method [110]

$$y_{scalar}(x) = \sum_{m=1}^{M} w_m f_m(x), \tag{3.27}$$

where w_m are the non-negative weights reflecting the importance of each objective. In weighted sum methods, differences in the scales of objectives can result in unfair weighting. One of the simplest algorithms using this approach for multi-objective optimization is ParEGO (Pareto Efficient Global Optimization) proposed by Knowles [111]. In this algorithm, scalarization is achieved using a weighted sum of the objective functions with random weights generated at each iteration. The main advantage of ParEGO over other scalarization methods is that the random weights allow for exploring the Pareto front for a diverse set of solutions without predefining the weights or relying on user preferences. By iteratively running this algorithm with new weights, it becomes possible to approximate a convex Pareto optimal set. The general problem with the weighted sum method is not being able to find the non-convex regions of the Pareto front and running optimizations with different weights is often computationally prohibitive.

When the objectives are of different scales and normalization is not possible, a scalarization method that could be employed is the weighted product defined as [112]

$$y_{scalar}(x) = \prod_{m=1}^{M} [f_m(x)]^{w_m}, \tag{3.28}$$

or alternatively for computational ease the logarithm of the objective could be used

$$\log y_{scalar}(x) = \sum_{m=1}^{M} w_m \log f_m(x). \tag{3.29}$$

This is generally more scale invariant compared to the weighted sum methodologies, but retains the problem of running the optimization multiple times to find the points on the Pareto front.

Finally the last scalarization method that we discuss here is known as the epsilon-constraint method where the problem of multi-objective optimization is reframed as a constraint optimization problem. Here, only one objective is optimized subject to the constraint that the other objective values have at least a value greater than epsilon. This method has the flexibility of controlling the acceptable values of the different objectives but on the other hand this flexibility can result in challenging optimization scenarios where prior knowledge about the different objective functions does not exist. There are other methods that are employed to scalarize the objective that are not outlined here and the reader is referred to [113] for more details.

In general the scalarization process, is often not straightforward and the weights or the epsilon values given to each sub-goal are usually assigned empirically. Moreover, one optimization with the scalarized objective results in a single point on the Pareto front and to reconstruct the Pareto front a number of scalarized objective optimizations need to be performed. This becomes computationally prohibitive and the running one scalarized optimization results in a point on the Pareto front that has an inherent bias attached to it. This bias comes with the choice of the weights that was made during the scalarization process. A more potent strategy for solving the class of multi-objective problems is directly trying to increase the diversity of solutions during the optimization process.

The Hypervolume Indicator and the Expected Hypervolume Improvement (EHVI) acquisition function offer powerful approaches to address this challenge by directly working with the Pareto front and quantifying improvements in a way that considers all objectives simultaneously.

Hypervolume (HV) is defined as the n-dimensional volume of the output subspace covered from a reference point, always taken to be zero in this work, to a set of points in the objective space. This metric considers both the convergence of solutions towards the true Pareto front and the diversity of solutions along the front and by integrating over the objective space, it inherently accounts for all objectives without requiring weights or preferences. Using this definition of HV, we can then proceed to define Hypervolume Improvement (HVI) as

$$\mathrm{HVI}(\mathcal{P}, \boldsymbol{y}) = \mathrm{HV}(\mathcal{P} \cup \boldsymbol{y}) - \mathrm{HV}(\mathcal{P}), \tag{3.30}$$

where $\mathcal{P}$ is the current Pareto front and $\boldsymbol{y}$ is a new vector valued output. Equation 3.30 describes the difference between the current hypervolume and one with an additional output point $\boldsymbol{y}$ [114]. If the set of points making up $\mathcal{P}$ already dominate $\boldsymbol{y}$ then $HVI = 0$, because there is no hypervolume gained by adding the point $\boldsymbol{y}$. The hypervolume and the hypervolume improvement are shown in Figure 3.7 along with the true Pareto front and different solutions.

HVI can be used to generalize the expected improvement policy described in Section 3.3.1 to the multi-objective scenario. First proposed by Emmerich et al. [115], this method is called *Expected Hypervolume Improvement (EHVI).* Following the definition from Yang et al. [114], we can write this as

$$\mathrm{EHVI}(\boldsymbol{x}) = \mathbb{E}(\mathrm{HVI}(\mathcal{P}, \boldsymbol{y})) = \int \mathrm{HVI}(\mathcal{P}, \boldsymbol{y})\, p(\boldsymbol{y}|\boldsymbol{x}, D_n)\, dy. \tag{3.31}$$

This criterion has been demonstrated to achieve a good convergence to the true Pareto front [116–118]. This method assumes a GP model for each different objective m that can provide a posterior mean and a posterior variance. Another point that needs to be considered when using EHVI is the assignment of the reference point from which the hypervolume is calculated. This point is usually chosen such that it is dominated by all potential solutions.

A common criticism of the EHVI acquisition function has been the time complexity involved in calculating it. A first closed-form calculation of EHVI was implemented

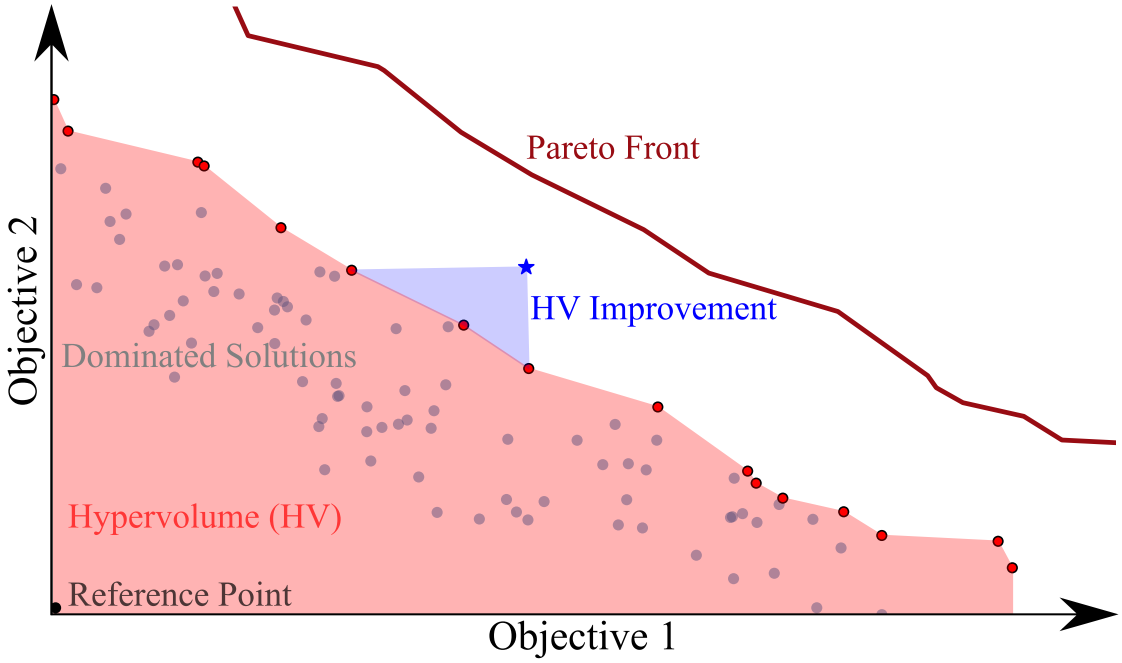

Figure 3.7.: The concept of hypervolume and hypervolume improvement is shown in this figure. The area covered between the set of non-dominated solutions here shown in red and the reference point is the current amount of hypervolume. A potential new point shown in blue can result in a hypervolume improvement which is the blue area that would be increased when the blue solution is achieved.

by Emmerich et al. [119] with a computational complexity $\mathcal{O}(n^3 \log n)$ for a 2-D output space. Over the years with efforts by Hupkens et al. [120], Emmerich et al. [121] and Yang et al. [122] the time complexity for 2-D and 3-D case has been reduced to $\mathcal{O}(n \log n)$. In this work an implementation of EHVI available on BoTorch based on estimating gradients using auto-differentiation is used as described by Daulton et al. [123]. This exploits the high number of cores that are available with modern GPUs to make EHVI optimization fast and applicable to real-world scenarios. While we have focused our discussion on EHVI, it should noted that information theoretic acquisition functions such as entropy search can also be adapted to multi-objective scenarios [124, 125].

3.3.3. Single-objective, multi-fidelity acquisition functions

In many real-world optimization problems, multiple information sources with varying degrees of fidelity are available. These sources can provide different levels of accuracy, typically with a trade-off between data fidelity and cost. By intelligently combining information from low-fidelity and high-fidelity sources, we can accelerate the optimization process while controlling computational resources. [126–131].

As before, we consider the input space with the vector inputs $\boldsymbol{x} \in \mathcal{X}$, but now we additionally include special input parameters that we call the fidelity parameters denoted by $\boldsymbol{s} \in \mathcal{S}$. Here $\mathcal{X}$ and $\mathcal{S}$ are the input and fidelity spaces, respectively. Importantly, in our Gaussian process (GP) model, fidelity $\boldsymbol{s}$ is treated as a regular input dimension alongside $\boldsymbol{x}$. This means that fidelity $\boldsymbol{s}$ is accounted for in the kernel of the GP, allowing us to make probabilistic inferences conditioned on both $\boldsymbol{x}$ and $\boldsymbol{s}$. We can also choose a different kernel to model the fidelity parameter and then use a product of kernels as the total kernel for the GP. This method of different kernels is often times required since the fidelity parameter is usually discrete. This prohibits the use of kernels like the RBF which assumes the a continuous space. On the other hand if the fidelity parameter can be treated as a continuous parameter then a single kernel can be defined over the augmented input space. At the end, the goal is to build a surrogate model that incorporates information from $f(\boldsymbol{x}, \boldsymbol{s})$. The challenge in multi-fidelity optimization is to effectively balance the trade-off between information and cost while ultimately finding a global maximum at the target fidelity. This target fidelity usually corresponds to the highest fidelity which is also the most expensive information source. To effectively utilize multi-fidelity models, specialized acquisition functions are required to decide not only where to sample next but also at which fidelity level.

One intuitive solution to this problem is the two-step approach proposed by Lam et al. [132]. In this approach, the selection of the next point to probe in $\boldsymbol{x}$ is done separately from the fidelity choice $\boldsymbol{s}$. To achieve this, Lam et al. use an Expected Improvement (EI) policy that is evaluated at the target fidelity only. Lower fidelity measurements are implicitly incorporated into this process, as they affect the surrogate model at the highest fidelity. Once a suitable position has been identified in $\boldsymbol{x}$, the ideal fidelity for probing this point is chosen by comparing the predicted reduction in uncertainty or gain in knowledge with the computational cost involved.

An alternative approach is to combine both the selection of the next point and the weighting by the expected knowledge gain per unit cost in a single step. Notably, Max-Value Entropy Search (MES) and Knowledge Gradient (KG) acquisition functions can be adapted with minor changes for this kind of multi-fidelity optimization. For KG [133, 134], the acquisition function is conditioned on the best value of the posterior mean at the target fidelity $\boldsymbol{s}^*$ ($\max_{\boldsymbol{x} \in \mathcal{X}}(f(\boldsymbol{x}, \boldsymbol{s}^*))$) rather than the best value of the posterior mean. In the case of MES, the mutual information between the maximum value y^* at the highest fidelity and the data set is maximized. This gain of information is then divided by the computational cost that is a function of fidelity $\boldsymbol{s}$ [135]. The resulting multi-fidelity MES acquisition function is then given by a simple modification to Equation (3.26):

$$\text{MF-MES}(\boldsymbol{x}, \boldsymbol{s}|D_n) = \frac{\text{MES}(\boldsymbol{x}|D_n)}{\text{cost}(\boldsymbol{s})}. \tag{3.32}$$

Similar multi-fidelity policies can be developed for other exploratory acquisition functions, such as the Upper Confidence Bound [136, 137].

3.4. Multi-objective multi-fidelity optimization

So far we have reviewed established techniques for MOMF optimization. However, these methods are typically applied independently. As we have observed, MO optimization addresses problems where the objective consists of multiple sub-objectives, aiming to find solutions that balance trade-offs between conflicting goals. In contrast, MF optimization leverages data from various information sources with differing levels of fidelity to optimize a single objective function more efficiently.

Both techniques offer unique advantages in their respective domains. MO optimization enables decision-makers to consider multiple criteria simultaneously, while MF optimization reduces computational costs by utilizing a hierarchy of models with varying accuracies. Despite these benefits, many real-world problems would greatly benefit from an approach that integrates both techniques.

This leads us to the concept of joint Multi-Objective Multi-Fidelity (MOMF) optimization—a method that combines the strengths of MO and MF optimization into a unified framework. By doing so, we can simultaneously handle multiple conflicting objectives and efficiently utilize resources by incorporating models of varying fidelity. Several compelling reasons motivate the pursuit of this integrated optimization approach, including improved efficiency, robustness and improved decision-making by the users. While efficiency is a feature of a MF technique, it becomes much more important when multiple objectives are being optimized since a number of low fidelity simulations can allow the optimizer to explore the different objectives without incurring high cost. The robustness is a benefit imparted by the MO technique since a prevalent issue in single-objective MF optimization arises because of over-reliance on lower fidelity data. By incorporating multiple objectives into the optimization process, the search strategy becomes more diversified, leading to a more robust and reliable optimization outcome. Lastly, integrating multi-objective and multi-fidelity optimization offers a deeper and more comprehensive understanding of the trade-offs inherent in complex optimization tasks. By enabling decision-makers to examine the interactions between different objectives across varying levels of fidelity, it provides valuable insights into how these factors influence one another. This enriched understanding facilitates more informed and holistic decision-making, ultimately leading to improved optimization outcomes and solutions that better align with the overall goals of the problem.

Despite these apparent benefits, MOMF optimization remains an emerging area of research. A first paper reporting such an approach for discrete fidelity levels was recently published by Belakaria et al [138]. Their MES for MO Bayesian optimization is based on the maximization of mutual information between the Pareto front and the search domain. Being mostly motivated by applications in neural network training, the approach from this paper is however limited to scenarios where higher fidelities yield higher objective values. The assumption that the objective values at lower fidelities are always upper-bounded by the values at the highest fidelity does not hold true for many use cases, such as numerical simulations of physical systems.

During the course of this work, a method that targets simultaneous multi-objective multi-fidelity optimization was proposed. This technique is able to work with continuous fidelity spaces and is furthermore much simpler to implement for practitioners. To benchmark this novel method, another method that selects the input parameters x and the fidelity parameter s sequentially was also proposed.

3.4.1. Trust - MOMF

Our proposed optimization scheme hinges on the *joint* optimization of the objective function and our *trust* in the information source that produces the results. We use 'trust' to represent our confidence level in the outcomes generated by the information source at varying levels of fidelity. This approach allows us to reframe multi-fidelity optimization problem as a multi-objective optimization problem, where trust $\theta(\boldsymbol{s})$ and output objectives $f(\boldsymbol{x})$ become the two focal points of optimization. Consequently, we aim to optimize the following function output:

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{s}) = \begin{pmatrix} \boldsymbol{f}(\boldsymbol{x}) \\ \theta(\boldsymbol{s}) \end{pmatrix}, \tag{3.33}$$

where $\boldsymbol{f}(\boldsymbol{x})$ is the output vector of the objectives.

The trust objective $\theta(\boldsymbol{s})$ is fundamental to the proposed optimization scheme. High fidelity sources, usually more costly, are inherently more trustworthy than their low fidelity counterparts due to their superior accuracy and reliability. Consequently, trust is modeled as a function that increases monotonically with fidelity. In the simplest case, one might assume a linear relationship where trust equals the fidelity parameter itself, $\theta(s) = s$. A more rigorous approach might involve quantifying trust through measures like mutual information, thereby linking the notion of trust to the information shared between fidelity levels. In such cases, the trust objective can be seen as an approximation that reflects the average mutual information shared across fidelities. For numerous circumstances, such as simulations in which the outputs at increasing fidelity converge, an appropriate trust curve may be approximated by $\theta(s) \approx \tanh(s)$. I would discuss the influence and design of the trust objective on the optimizer in more detail in Section 3.5.2.

As discussed in Section 3.3.2, we can optimize a problem of the form Eq.3.33 using, for instance, Expected Hypervolume Improvement (EHVI). This acquisition function seeks to increase the joint hypervolume encompassed by the Pareto front of $f(\boldsymbol{x}, \boldsymbol{s})$. Due to the increasing nature of the trust function, the optimizer tends to explore points with higher trust values. In a multi-fidelity context where computational or experimental costs vary, the acquisition function can be normalized by incorporating a cost penalty, specifically by dividing the value of the acquisition function by the corresponding cost similar to what we discussed in Section 3.3.3. This modification ensures that the optimization process selectively investigates the point that maximizes the ratio of expected hypervolume improvement to the associated cost. As a result, the acquisition function optimizes our knowledge of the Pareto front and Pareto set on a per-unit-cost basis. The final acquisition function is then written as

$$\text{Trust-MOMF}(\boldsymbol{x}, \boldsymbol{s}|D_n) = \frac{\text{EHVI}(\boldsymbol{x}, \boldsymbol{s}|D_n)}{\text{cost}(\boldsymbol{s})}. \tag{3.34}$$

where EHVI$(\boldsymbol{x}, \boldsymbol{s})$ signifies that the complete objective comprises both the normal output objectives and the trust objective. The entire structure of this approach for Trust-based Multi-Objective Multi-Fidelity (Trust-MOMF) optimization is outlined in Algorithm 1.

The algorithm starts with a random initialization of n points that result in a data set $D_n = \{(\boldsymbol{x}_1, \boldsymbol{s}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{s}_2, \boldsymbol{y}_2), ..., (\boldsymbol{x}_n, \boldsymbol{s}_n, \boldsymbol{y}_n)\}$. Assuming that the number of output objectives is k, we can use the data set D_n to build k Gaussian process models, fitted corresponding to k output objectives. The initial cost C_{init} utilized until this point is the cost undertaken to generate the initial dataset D_n. After the initialization, the acquisition function is optimized to yield both the new input $\boldsymbol{x}_{n+1}$ and fidelity position $\boldsymbol{s}_{n+1}$ where the objective function is then evaluated resulting in a new objective vector $\boldsymbol{y}_{n+1}$. This generates a new data sample $(\boldsymbol{x}_{n+1}, \boldsymbol{s}_{n+1}, \boldsymbol{y}_{n+1})$ that is used to update the data set and re-fit the Gaussian process models. The last step is to update the spent cost C_{spent} and check whether the computational budget C_{total} was exceeded. The process is terminated when the available computation budget $C_{\text{avail}} = C_{\text{total}} - C_{\text{spent}}$ becomes 0.

Figure 3.8 reproduced from [139] exemplifies the optimization of a modified 1D Forrester function by integrating information from lower fidelity data that incur reduced cost. In this test scenario, we consider a linear trust function defined by

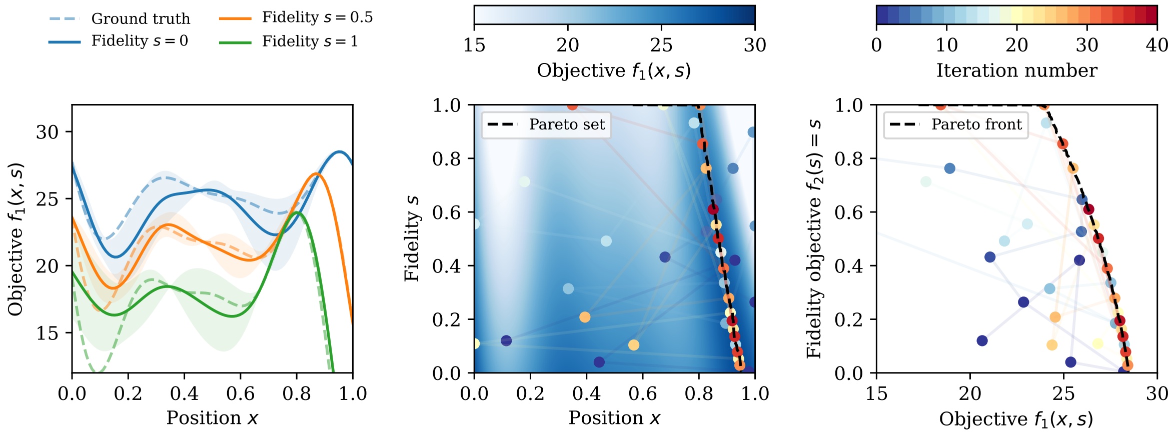

Figure 3.8.: Application of trust-MOMF to an optimization of a modified Forrester function. On the left is the mean and variance (shaded curve) of the fitted Gaussian process after 40 iterations of Bayesian optimization. The ground truth function curves at different fidelity levels are indicated as dashed lines. Note the small variance close to the maximum at all fidelity values. At the center the input space consisting of both the position and fidelity values is shown. The dotted black line depicts the Pareto set and the color indicates the iteration number of the sampled points. This illustrates how the optimizer first explores at low fidelity and then moves along the Pareto set. Finally at the right, the output objective space is shown, depicting how the optimizer tries to increase the hypervolume, which is simply the area in 2D, below the Pareto front. The figure has been reproduced from the work [139] conducted during this thesis.

$\theta(s) = s$ and model the cost function as $C(s) = \exp[a \cdot s]$, with $a = 5$. This formulation results in an evaluation at the maximum fidelity level ($s = 1$) being approximately 150 times more expensive than at the minimum fidelity level $s = 0$. It is noteworthy that this cost function can be replaced with any other monotonically increasing function as per the requirements of the specific application and the effect of choosing different trust functions would be discussed in a later section. A fundamental assumption of the Trust-MOMF method is that the Pareto set, which underlies the Pareto front, remains within a similar region of the search space across different fidelity levels. This enables efficient transfer of information between different fidelities. However, it is important to acknowledge that this method, akin to all multi-fidelity approaches that rely on knowledge transfer, becomes less efficient if the Pareto set undergoes significant shifts between fidelity levels.

Lastly, another important feature of this method is that it can be expanded to include m fidelity dimensions $s^{(m)}$. This approach is especially advantageous for multi-dimensional numerical models that have individual resolution parameters for each dimension. Depending on the problem at hand, these individual fidelity dimensions can be managed either through a single, unified trust objective or by treating them as separate trust objectives within the optimization framework.

Algorithm 1 Trust-based Multi-Objective Multi-Fidelity Optimization (Trust-MOMF)

```
Inputs:
  • Probed Dataset $D_n$
  • Gaussian Process Models $GP_1, \dots, GP_k$
  • Fidelity function $\boldsymbol{s}$, Cost function $C(\boldsymbol{s})$
  • Total computational budget $C_{\text{total}}$
Outputs: Optimal Pareto front $\mathcal{P}^*$
Initialization:
  Generate $n_{\text{init}}$ initial data points $D_n$
  Fit Gaussian Process models $GP_1, \dots, GP_k$
  Initialize Pareto front $\mathcal{P} = \emptyset$
  Initialize spent budget $C_{\text{spent}} = C_{\text{init}}$
while $C_{\text{spent}} < C_{\text{total}}$ do
    Optimization Step:
  Optimize the Trust-MOMF acquisition function, Equation (3.34)
  $(\boldsymbol{x}_{n+1}, \boldsymbol{s}_{n+1}) \leftarrow \arg\max_{\boldsymbol{x},\boldsymbol{s}} \left[\frac{\text{EHVI}(\boldsymbol{x},\boldsymbol{s})|D)}{C(\boldsymbol{s})}\right]$
    Probe problem at selected position and fidelity:
  $\boldsymbol{y}_{n+1} \leftarrow \text{Problem}(\boldsymbol{x}_{n+1}, \boldsymbol{s}_{n+1})$
    Data and Model Updates:
  Update Dataset $D_{n+1} \leftarrow D_n \cup \{(\boldsymbol{x}_{n+1}, \boldsymbol{s}_{n+1}, \boldsymbol{y}_{n+1})\}$
  Re-fit Gaussian Process models $GP_1, \dots, GP_k$
  Update Pareto front $\mathcal{P}$
  Increment $C_{\text{spent}}$ by $C(\boldsymbol{s}_{n+1})$ and $n$ by 1
Terminate
```

However, it is important to note that the Expected Hypervolume Improvement (EHVI) algorithm does not scale efficiently with a high number of output dimensions, as previously discussed. When numerous fidelity dimensions need to be optimized individually, the sequential Multi-Objective Multi-Fidelity (MOMF) method introduced in the next section may serve as a more suitable alternative. Moreover, the different fidelity dimensions may be discrete or categorical in nature, assuming finite values. While the trust-MOMF method remains applicable in these cases, the optimization of the acquisition functions requires adaptation. For instance, mixed acquisition function optimization, as implemented in BoTorch, can handle input spaces that include both continuous and discrete input parameters, accommodating the unique characteristics of each fidelity dimension.

3.4.2. Sequential optimization

In the last section, we discussed the benefits and implementation of joint Trust-based Multi-Objective Multi-Fidelity (Trust-MOMF) optimization. An alternative approach that merits investigation is a sequential version of the Multi-Objective Multi-Fidelity (MOMF) optimization. This sequential optimization scheme, inspired by the work of Lam et al. [132] on single-objective optimization, decouples the selection of the next position to probe from the choice of fidelity level, effectively partitioning them into two separate steps. In this section, this sequential optimization scheme will be presented, detailing its fundamental principles, operational procedures, and potential advantages. This *sequential* MOMF approach serves as a benchmark to evaluate the extent to which different multi-objective multi-fidelity problems benefit from a joint optimization strategy.

In this scheme, we assume an initial dataset after which, the first step involves selecting a candidate position in the input space, denoted as $\boldsymbol{x}_{n+1}$. This input position selection process is identical to that in conventional multi-objective optimization (see section 3.3.2). In Algorithm 2 and the examples used throughout this work, we employ the Expected Hypervolume Improvement (EHVI) method for this purpose. It is still noteworthy to point out that this mechanism of selecting the input position is agnostic to the method used to get this input position. One can use any other method that is used in the multi-objective optimization to result in a input position. Once the candidate point has been identified, we proceed to the fidelity selection phase.

For fidelity selection, the Gaussian process (GP) model needs to be scalarized for the usual single-objective multi-fidelity approaches to be applicable. In this work, the scalarization was achieved by summing all the objectives with equal weights, thereby avoiding any preference between objectives. This method is justified by the fact that all objective functions vary on a similar scale, due to a normalization step in data processing that scales the outputs of all objectives to the range $[0, 1]$. Alternatively, other scalarization techniques such as the weighted Chebyshev method or penalty boundary intersection [140] could be utilized.

After obtaining a scalar output from the GP, we apply a multi-fidelity acquisition function (see Section 3.3.3) to determine an appropriate fidelity setting s_{n+1}. In this context, we use the Multi-Fidelity Max-value Entropy Search (MF-MES) method, which aims to maximize the information gain per unit cost as already outlined in Section 3.3.3. Other methods based on knowledge gradient are also suitable alternatives. Following the evaluation of the problem at the selected position and fidelity (x_{n+1}, s_{n+1}), we update the dataset and GP models. The optimization loop then restarts and continues until the allotted computational budget is exhausted.

As we will discuss in Section 3.5, this relatively straightforward implementation of a sequential MOMF policy already yields a significantly faster convergence compared to pure multi-objective optimization. One notable advantage of this scheme is the minimal computational overhead relative to pure MO optimization, since the information gain across fidelities only needs to be computed at the already selected candidate point.

Algorithm 2 Sequential Multi-Objective Multi-Fidelity Optimization (Seq. MOMF)

```
Inputs:
  • Probed Dataset $D_n$
  • Gaussian Process Models $GP_1, \dots, GP_k$
  • Fidelity function $\boldsymbol{s}$, Cost function $C(\boldsymbol{s})$
  • Total computational budget $C_{\text{total}}$
Outputs: Optimal Pareto front $\mathcal{P}$
Initialization:
  Generate $n_{\text{init}}$ initial data points $D_n$
  Fit Gaussian Process models $GP_1, \dots, GP_k$
  Initialize Pareto front $\mathcal{P} = \emptyset$
  Initialize spent budget $C_{\text{spent}} = C_{\text{init}}$
while $C_{\text{spent}} < C_{\text{total}}$ do
    Position Selection Step:
      Perform expected hypervolume improvement, Equation (3.31), at maximum fidelity:
      $\boldsymbol{x}_{n+1} \leftarrow \arg\max_{\boldsymbol{x} \in \mathcal{X}}[\text{EHVI}(\boldsymbol{f}(\boldsymbol{x}) | D_n, s = 1)]$
    Fidelity Selection Step:
      Normalize Data $\boldsymbol{y} \to \boldsymbol{y}'$
      Scalarize output objectives: $y_{\text{scalar}} = \sum_{i=1}^{k} a_i \cdot y'_i$
      Fit a Gaussian Process model $GP_{\text{scalar}}$ on scalarized output objective
      Perform max-value entropy search, Equation (3.32), on $GP_{\text{scalar}}$ at selected position $\boldsymbol{x}_{n+1}$
      $\boldsymbol{s}_{n+1} \leftarrow \arg\max_{s \in \mathcal{S}}[\text{MF-MES}(\boldsymbol{x}_{n+1})]$
    Probe problem at selected position and fidelity:
      $\boldsymbol{y}_{n+1} \leftarrow \text{Problem}(\boldsymbol{x}_{n+1}, \boldsymbol{s}_{n+1})$
    Data and Model Updates:
      Update Dataset $D_{n+1} \leftarrow D_n \cup \{(\boldsymbol{x}_{n+1}, \boldsymbol{s}_{n+1}, \boldsymbol{y}_{n+1})\}$
      Re-fit Gaussian Process models $GP_1, \dots, GP_k$
      Update Pareto front $\mathcal{P}$
      Increment $C_{\text{spent}}$ by $C(\boldsymbol{s}_{n+1})$ and $n$ by 1
Terminate
```

3.5. Comparison and benchmark

In this section, we will describe the results of our proposed Trust-based and sequential MOMF algorithms on synthetic test functions before using them to accelerate simulations and experiments. All of the benchmarking in this section was performed by modifying and extending existing implementations of MES and EHVI functions in the BoTorch package [141] to the multi-objective, multi-fidelity problem.

A brief discussion of the general procedure used to generate the benchmarks is first described, after which a discussion of the results is presented. A description of the synthetic objective functions used to demonstrate the effectiveness of the trust-MOMF is given first, following which the initialization and hypervolume calculation procedure is outlined.

Test functions. To assess the performance of the methods and estimate the cost reduction factors, we use multi-fidelity modifications of the popular maximizing Branin-Currin (2-D) and Park (4-D) test functions. The function definitions for the Branin-Currin and Park are given in the appendix. Since this is one of the first study that deals with simultaneous multi-fidelity and multi-objective problems, many of the modifications of the synthetic functions are original contributions resulting from this work. For the results shown in this section, the cost function for the different fidelities is modeled as an exponential function of the form $C(s) = \exp[a \cdot s]$ with $a = 4.7$ to result in a ratio of about 120:1 between the highest ($s = 1$) and lowest ($s = 0$) fidelity. The number of iterations for the MOMF algorithms was fixed to 120 while the multi-objective single-fidelity optimization referred to as MO ran for 80 iterations. For the MO optimization, the total cost was 9600 while the MOMF algorithms stopped at variable costs ranging from 1500 to 4000.

Initialization. The MOMF algorithms start with five initial points randomly distributed within the input search space. In contrast, the single-fidelity multi-objective optimizer is configured with just one starting point at the highest fidelity level $s = 1$, resulting in an initial cost of approximately $C(1) \approx 120$. For the MOMF optimizers, the fidelity levels of the initial points are selected based on a probability distribution that is weighted inversely by the cost $p(s) \propto 1/C(s)$, and on average, reduces the initialization cost by a factor of five $C(1) \approx 25$. Acknowledging that initial points can significantly impact optimization performance, each optimization was performed 10 times with different initializations to gather more robust statistics on algorithm convergence.

Hypervolume calculation. After completing an optimization run, the hypervolume attained by each of the algorithm at every iteration is calculated. For this calculation, the points obtained during the optimization run are used as training inputs for a GP model. This GP model is then probed with a random sample of 10,000 high-fidelity input points, from which the set of non-dominated points is generated. This Pareto set is then used to compute the hypervolume at each iteration.

It's important to highlight that this hypervolume estimate is based on the expected values from the GP model rather than on samples drawn from it. The model confidence is not incorporated in this estimation, unlike methods such as expected hypervolume improvement. Moreover, since all of the test points are fixed at the highest fidelity and the GP model might have not many high fidelity points in its training data, the hypervolume resulting from the calculation is stochastic in nature. Consequently, the hypervolume estimate may fluctuate or even decrease as new points update the model, see for instance sequential MOMF method at cost ~ 600 in Figure 3.9. This variability reduces as the GP model becomes more confident in its predictions.

In the first two iterations, the GP model does not have enough training points, resulting in a large number of non-dominated points. This abundance makes the hypervolume calculation computationally intensive and impractical at this early stage. This calculation of the hypervolume is performed after the fourth iteration to address this issue without sacrificing critical information. This adjustment explains why, in the subsequent figures, the graphs for multi-objective optimizations start at a cumulative cost of 480. By delaying the calculation, we ensure computational efficiency while maintaining the integrity of the optimization process.

3.5.1. Results

Branin-Currin test function. In Figure 3.9 an optimization of the multi-fidelity versions of Branin and Currin [107, 108] functions is shown. From the ten trials conducted, we selected a single representative trial based on how closely its hypervolume trajectory, plotted against cost, matched the mean hypervolume trajectory of all trials. In the multi-objective (MO) optimization, the cost axis exhibits regular increments, reflecting a fixed evaluation cost at the highest fidelity level. In contrast, both MOMF optimization methods display irregular step sizes along the cost axis. This irregularity arises because the MOMF algorithms often select several points at intermediate fidelity levels before conducting evaluations at the highest fidelity. This can also be seen in the top right of the figure where the distribution of selected fidelities for each algorithm is illustrated. Interestingly, we observe distinct behaviors between the two MOMF algorithms. The sequen-

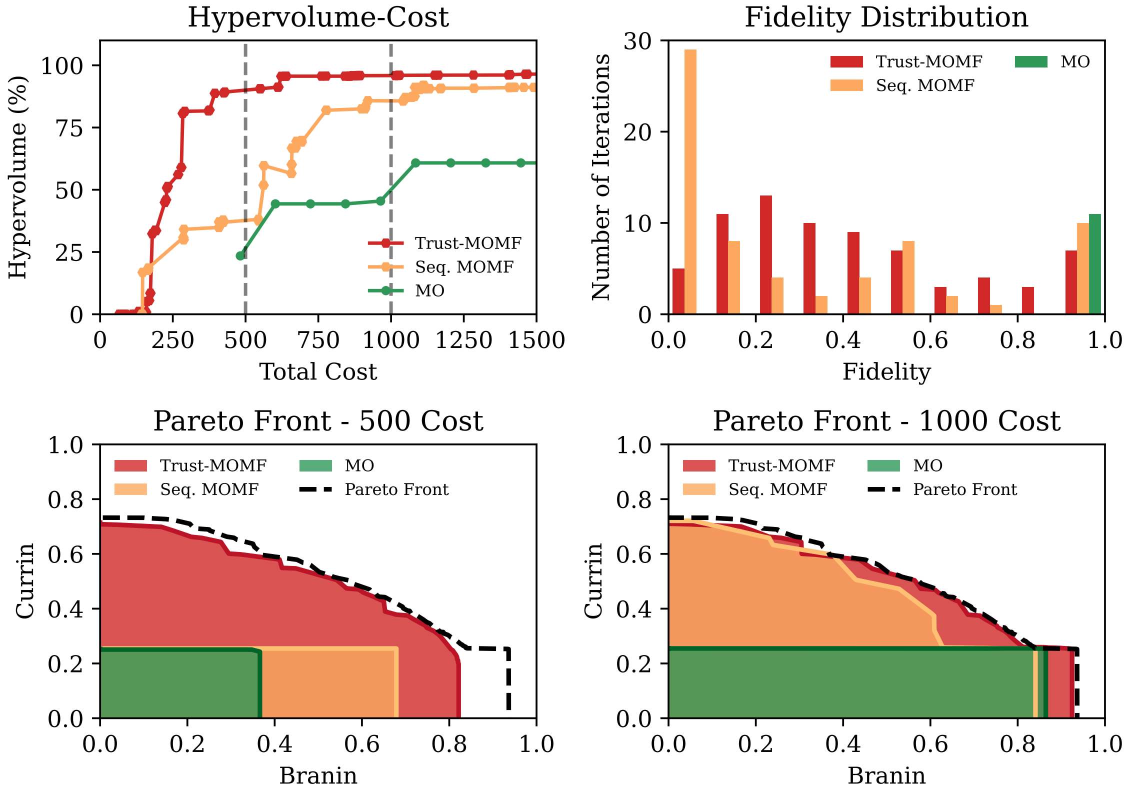

Figure 3.9.: Benchmark with 2-D-Branin-Currin problem. *Top Left:* The hypervolume progression of a single representative trial is presented, showing the percentage of the total hypervolume achieved versus the total cost for both versions of the MOMF optimization and the single-fidelity MO optimization.The greyed out dashed line represents the cost at which the two bottom Pareto fronts are generated. *Top Right:* This subfigure shows the distribution of fidelities for the representative trial run. The Trust-MOMF method allocates more iterations to intermediate fidelity levels compared to the sequential MOMF method, which predominantly selects points at the minimum and maximum fidelity levels.*Bottom Left:* The Pareto fronts obtained by each of the three algorithms at a cumulative cost of 500 are depicted here. The dashed black line represents the estimated true Pareto front. *Bottom Right:* At a cumulative cost of 1,000 for a single representative trial, the Pareto front clearly demonstrates that the Trust-MOMF method has achieved an accurate approximation of the estimated true Pareto front. Meanwhile, the conventional single-fidelity MO algorithm has only uncovered a small portion of the Pareto front at this cost level. This highlights the efficiency of the Trust-MOMF method in exploring and approximating the Pareto front more effectively within the same or even reduced computational budget.

tial MOMF method selects significantly more points at the lowest fidelity level, whereas the Trust-MOMF optimization favors more points at intermediate fidelity levels. This difference may be attributed to the Trust-MOMF method's joint optimization of both the input variables and the fidelity levels.

The bottom part of Figure 3.9 depicts the behavior of the Pareto front at two different costs. The black dashed line denotes the true Pareto front, calculated

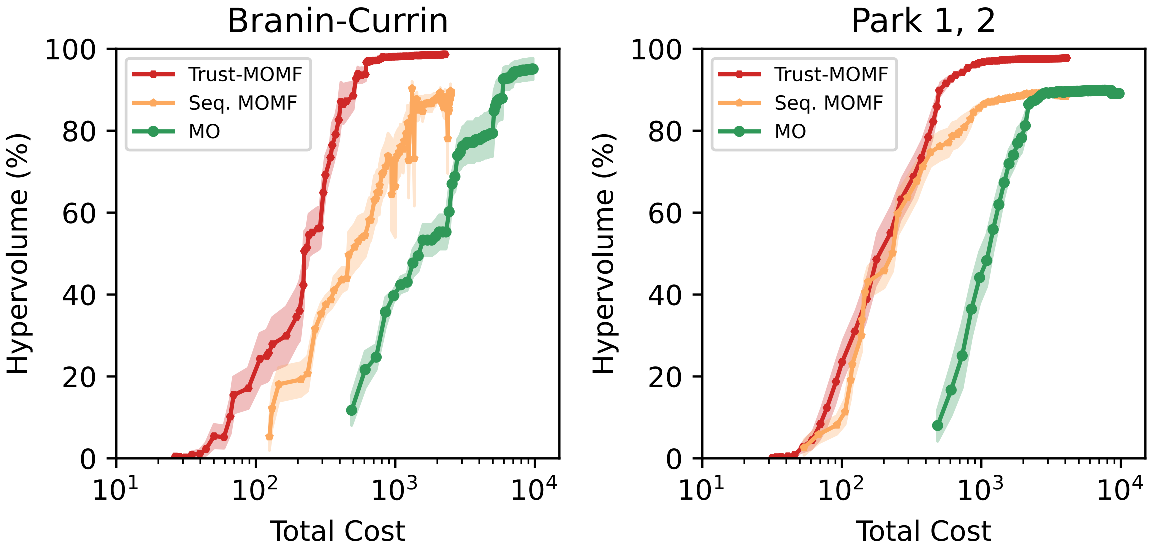

Figure 3.10.: Mean Hypervolume Results for 10 trials of optimization of Branin-Currin and Park Functions. *Left:* Results for the Branin-Currin optimization is illustrated with the total cost shown on a logarithmic scale to effectively represent the significant differences in cost magnitude. The noticeable dips in the curve around a cost of 900 are due to the stochastic nature of the hypervolume calculation. At a convergence hypervolume threshold of 90%, the Trust-MOMF optimization exhibits a cost advantage of approximately one order of magnitude compared to the MO optimization. *Right:* The mean hypervolume of 10 trials for the Park functions is shown. Here again, the Trust-MOMF optimization demonstrates an order of magnitude more cost efficiency considering a 90% threshold of hypervolume.

using 10,000 random sampling points. From this figure, it is evident that the Trust-MOMF method has already achieved substantial coverage of the trade-off region and has identified the maximum of the Currin function. In contrast, at a cumulative cost of 500, both the standard MO optimization and the sequential MOMF algorithms display significantly less hypervolume coverage. When the cost increases to 1000, the MO optimization has successfully optimized the Branin function but has yet to explore the trade-off region between objectives. The sequential MOMF algorithm, at a cost of 1000, reaches a state comparable to that of the Trust-MOMF optimization at a cost of 500. Meanwhile, the Trust-MOMF optimization has attained nearly 97% hypervolume coverage.

To evaluate the cost advantage, we examined the mean hypervolume cost curves obtained from ten independent runs, as illustrated on the left side of Figure 3.10. Setting a convergence threshold at 90%, we found that the Trust-MOMF optimization method demonstrates a cost advantage of approximately one order of magnitude over the standard MO optimization. Specifically, the MO algorithm reached 90% convergence at an estimated cost of 6000, while the Trust-MOMF method achieved the same level of convergence at a cost of 530. This results in a cost reduction factor of approximately 11. Furthermore, the Trust-MOMF method progressed to a higher final hypervolume percentage of 99%, compared to the MO optimization's final convergence at 94%.

Park test function. Figure 3.11 shows the results of a representative trial of the three algorithms when they are used to optimize modified Park functions [142]. The Park functions optimization problem is a more difficult problem than the Branin-Currin functions, since they have two more input dimensions.

The top-left subplot in Figure 3.11 shows the percentage of hypervolume covered as a function of cost for both the MOMF and MO optimization runs. Similar to the Branin-Currin problem, the sequential MOMF method focuses on evaluating points at the minimum ($s = 0$) and maximum ($s = 1$) fidelity levels, whereas the Trust-MOMF method selects a greater number of points at intermediate fidelity levels.

The bottom two plots in Figure 3.11 depict the evolution of the Pareto fronts for the three optimization methods. Notably, the Trust-MOMF method has already achieved a hypervolume coverage of 94% at a cost of 750, resulting in a Pareto front that closely approximates the true Pareto front. The sequential MOMF method has also begun to explore the trade-off region resulting in a Pareto front with slightly less coverage than the trust-MOMF. In contrast, the MO optimization, at a cost of 750, has located points near the maximum of the second Park function and is just begining to explore the trade-off region at a cost of 1500.

Similar to the Branin-Currin, we can assess the cost advantage, by looking at the mean hypervolume versus total cost curves generated from ten independent trials presented in Figure 3.10. The MO algorithm converged to a hypervolume coverage of 89.8% at an approximate cost of 7600, while the Trust-MOMF optimization reached 90% hypervolume coverage at a significantly lower cost of about 560. This results in a cost reduction factor of approximately 13, which is consistent with the cost savings observed in the Branin-Currin problem.

3.5.2. Understanding trust objective

Optimization of synthetic functions depicts a significant reduction in the computational cost required to find the Pareto front when using joint trust-based or sequential MOMF optimization methods. The trust-MOMF outperformed the sequential approach, suggesting that it more effectively utilizes the available information within the combined search domain.

It is important to note that the extent of cost reduction is intrinsically linked

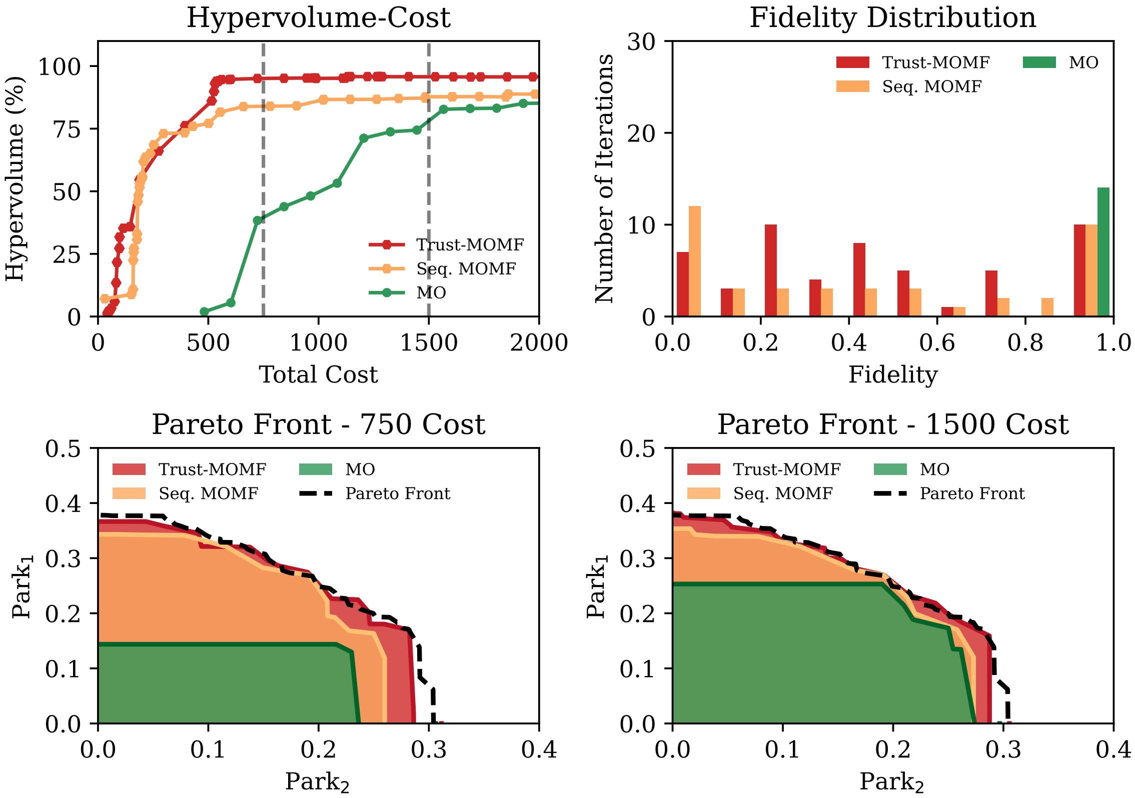

Figure 3.11.: Benchmark with 4-D-Park$_{1,2}$ problem. *Top Left:* Hypervolume of a single representative trial expressed as a percentage of the total hypervolume versus total cost for both MOMF versions and single-fidelity MO. *Top Right:* The number of points taken at different fidelites for the representative trials shown on the left. The Trust-MOMF method in this case has a higher number of points taken at the highest fidelity. This is because once it has converged it takes 6 points at the highest fidelity to increase hypervolume. The sequential MOMF optimization as seen in the Branin-Currin case takes fewer intermediate fidelity points when compared to the Trust-MOMF optimization. *Bottom Left:* The Pareto front for each of the three algorithms for the same trial at a cost of 750 (indicated by the dashed line in the figure on the top left). The area represents the amount of Pareto front covered by each algorithm. The Trust-MOMF method has already converged to almost 95% of the total hypervolume. The sequential MOMF optimization also converged but to a lower overall hypervolume. The MO optimization at this cost has only found a maximum of Park 2 function.*Bottom Right:* The Pareto front for a cost of 1500 shows little changes in both MOMF Pareto fronts, but a better coverage for the MO Pareto front. At this cost, the MO optimization still has not reached the hypervolume that the Trust-MOMF optimization reached at a cost of 750

to the cost ratio between the lowest and highest fidelity levels. In the examples shown before, this cost ratio was 1:120, implying that the maximum possible cost reduction by exclusively using the lowest fidelity data points would be a factor of 120. Mean fidelity values of $\overline{s} \simeq 0.30$ and $\overline{s} \simeq 0.32$ for both the Branin-Currin and Park functions optimizations respectively was achieved by averaging across 10 trials and 150 iterations. This results in average costs per iteration

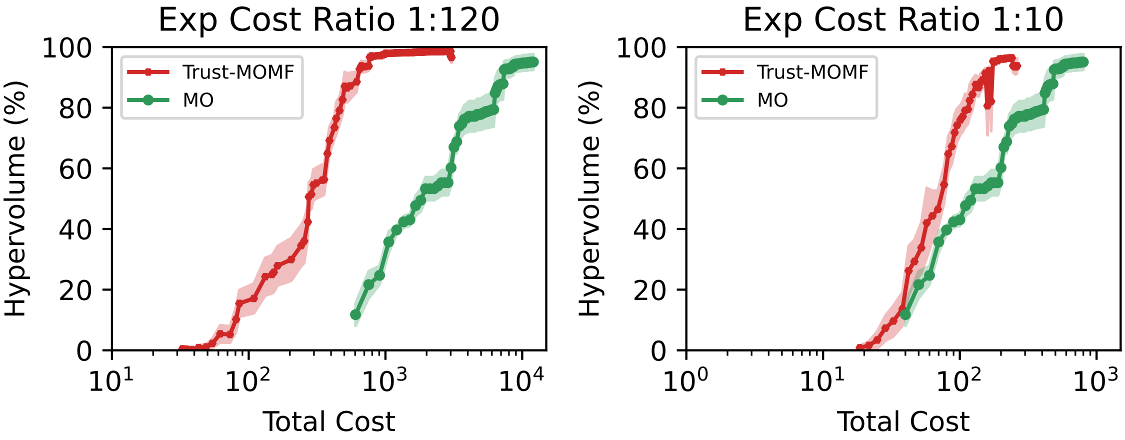

Figure 3.12.: Benchmarks with Branin-Currin functions with two different cost ratios of an exponential cost function. On the left the cost ratio between the lowest and the highest fidelity is 1:120 while it is 12 times less for the right graph. It is apparent from the figures that for such a cost function, a decrease in cost benefit is seen.

of approximately 4.3 and 4.6, whereas the conventional MO optimization had a fixed cost of 120 per iteration. Therefore, the maximum achievable cost reduction, assuming the information gain is identical across all fidelity levels, is about 26. However, this assumption generally does not hold, as lower-cost approximations typically do not provide as much information as the highest fidelity evaluations. In this instance, a cost reduction of roughly half of this theoretical maximum was observed, resulting in a cost reduction factor of 13 in both test cases. The cost reduction factor can be even greater when the ratio between the highest and lowest fidelity costs is increased. For instance, with a maximum cost ratio of 1200:1 for the Branin-Currin problem, a cost reduction factor of 44 was measured.

One important aspect of the trust-MOMF scheme is the trust objective that was not yet discussed. In the case of the optimizations of Branin-Currin and the Park functions, the trust objective was defined as a linear function of the fidelity parameter. The aim of the trust objective is to quantify the amount of information that can be obtained through probing different fidelities. A linear function implies a constant increase in information as the fidelity parameter is incrementally raised. Combining a linear function $\theta(s) = s$ for the trust objective and an exponential cost function $cost(s) = \exp(-5s)$, we get a trust-versus-cost trade-off in the Trust-MOMF objective of the form $s \cdot \exp(-5s)$. This trade-off reaches its analytical maximum at $s = 0.2$, which is close to the mean fidelity adopted by the Trust-MOMF method during optimization. However, the mean fidelity during optimization was higher because the optimizer probed the objective functions for a fixed number of iterations resulting in variance and would improve as the

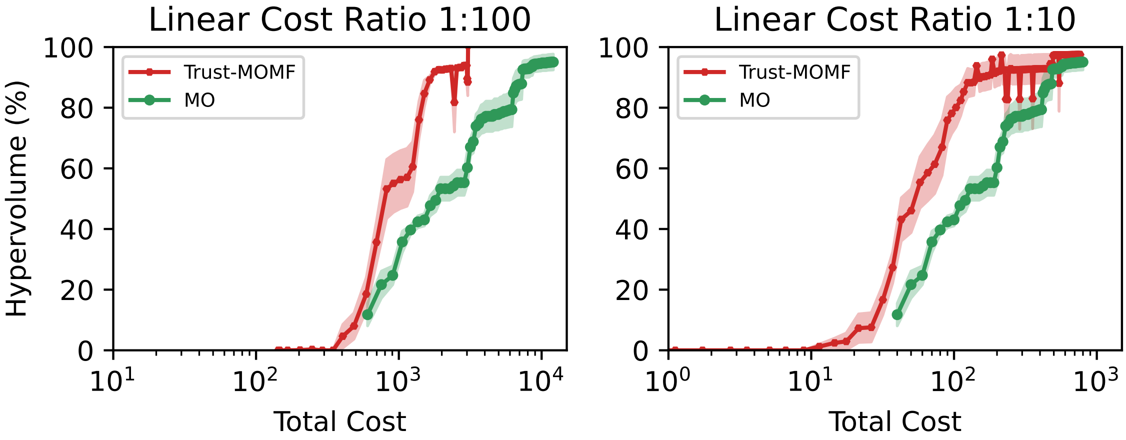

Figure 3.13.: Benchmarks with Branin-Currin functions with two different cost ratios of a linear cost function. Since the linear cost function combined with the trust objective results in an analytical expression independent of the fidelity, hence we see no effect with different cost ratios.

number of iterations is increased. We can think of the trust objective as giving an approximate information content as a function of fidelity that the user assumes for the optimization problem. This approximate behavior can still be rejected by the optimizer given enough iterations, resulting in a lower cost benefit.

This trust objective can also be used to encapsulate the convergence behavior of the optimization problem. As an example, a converging trust function such as $\theta(s) = s\exp(-s)$ defined over the domain $s \in [0, 1]$ can be used. This function is monotonically increasing but with diminishing returns. Note that the combination with a cost function of the form $cost(s) = \exp(-4s)$ yields the same trust-versus-cost trade-off and thus, the same optimum at $s = 0.2$. In all cases, this optimum serves as an estimate for the potential speedup attainable when employing Trust-MOMF with a specific trust objective function and cost function. We can see this effect in practice when additional benchmarks with different cost functions and various cost ratios were generated. All of these tests were conducted only on the Branin-Currin functions and the results for the exponential cost function can be seen in Figure 3.12. On the left side of this figure, the ratio between the maximum and minimum cost was kept to 120 as in the previous benchmarks while for the graph on the right, this ratio was reduced to 10. In this instance, the analytical maximum is shifted to a higher value because of a lower cost ratio and thus results in a higher mean fidelity. This results in an overall higher cost and a lower speedup when using the Trust-MOMF method. In Figure 3.13, the cost function is no longer an exponential but a linear function with two different cost ratios between the maximum and minimum cost. The difference in the advantage of using trust-MOMF between the cost ratios with a linear cost function is not as pronounced as in the case of exponential functions. In the case of a linear cost function, there is no analytical maximum and hence the cost ratio does not affect the outcome of the optimization. The choice of the fidelity in this instance would only depend on the behavior of the objection functions as a function of the fidelity.

4. Numerical Experiments

In the domain of laser-plasma interaction, numerical simulations and methods play a crucial role in understanding and predicting complex plasma behaviors that are often inaccessible through analytical methods. The analytical framework is usually either restricted to a linear regime with laser intensities below the relativistic limits or is based on simplifications such as reduction to a single dimension. For this reason, numerical tools are indispensable and one of the most widely used numerical techniques in this domain is the Particle-in-Cell (PIC) method [143–145]. The Particle-in-Cell method is a computational technique designed to simulate the dynamics of charged particles interacting with electromagnetic fields in a self-consistent manner. Initially developed for plasma physics applications [146–148], the PIC method has become a cornerstone in the simulation of a wide range of phenomena, including astrophysical processes [149], nuclear fusion [150], accelerator physics [151], and space plasma interactions [152].

For the purpose of optimization, a PIC simulation can serve as an objective function by modeling complex plasma behaviors and providing high-dimensional outputs based on a set of input parameters. Bayesian optimization treats the PIC simulation as a "black-box" function, where each simulation run yields an electron beam on which different performance metrics can be defined. By using a surrogate model, such as a Gaussian process, Bayesian optimization efficiently explores the parameter space of the PIC simulation, minimizing the number of expensive simulation runs while identifying the optimal conditions. This approach is particularly effective for expensive, high-fidelity simulations like PIC, where each evaluation is computationally intensive. Moreover, this is an apt testbed for the Multi-fidelity Multi-objective scheme developed in this work, since lower-fidelity simulations can be executed quite easily as it would be discussed in the next sections. In the following, the general PIC method would be discussed and its variants particularly FBPIC that was used in this work. A discussion on the optimization of a numerical LWFA using FBPIC simulations and how multi-objective multi-fidelity optimization benefits current state-of-the-art in the domain of numerical experiments will follow.

4.1. PIC method

Macroparticles. At its core, the PIC method combines aspects of both particle-based and grid-based computational approaches. The plasma is represented by a large number of macroparticles, each symbolizing a large collection of physical particles such as electrons or ions. The basic rationale behind using a macroparticle is to reduce the computational load since tracking each particle for plasma densities in the range of 10^{19}cm^{-3} is prohibitively expensive. Hence, a large number of identical real particles in phase space are bunched together into a macroparticle. The macroparticle is a computational particle and in itself is not a physical observable. These macroparticles carry properties like position, velocity, charge, and mass similar to the large collection of particles that they represent. The particles have a definite momentum and the position is the average over the positions of the real particles indicating a certain spatial extent around the average position. The macroparticles follow the same trajectory as the real particles it represents, provided the scattering processes between the real particles do not have a chaotic behavior. This is an approximation since the individual momentum and position of the particles can change; however, if the particles remain close in the phase space then the approximation remains valid. The Lorentz force experienced by the computational particle is similar to the real particles since it is only dependent on the charge-to-mass ratio.

Fields. The electromagnetic fields are computed on a spatial grid that discretizes the simulation domain, with field quantities such as electric and magnetic fields defined at grid points. By coupling the motion of particles with the evolution of fields, the PIC method self-consistently solves the Vlasov-Maxwell (see Equations (2.26a) and (2.26b)) system of equations, capturing the intricate interplay between particles and fields in plasmas. In the context of LWFA, PIC simulations are particularly well-suited for modeling because they can capture the highly non-linear interactions between the laser pulse and plasma. This involves resolving fine-scale phenomena such as particle trapping, beam loading, and handling complex geometries inherent in plasma environments.

The PIC algorithm advances the simulation through a series of computational steps, repeated at each time increment to evolve the system over time. These steps are designed to model the self-consistent evolution of particles and fields accurately and the temporal resolution of the simulation is determined by this time-step. In each of these time-steps the macroparticle momenta and positions are calculated along with the magnetic and electric field. Each of these time-steps is called a PIC cycle, and the important steps are outlined below while the complete PIC cycle is shown in Figure 4.1.

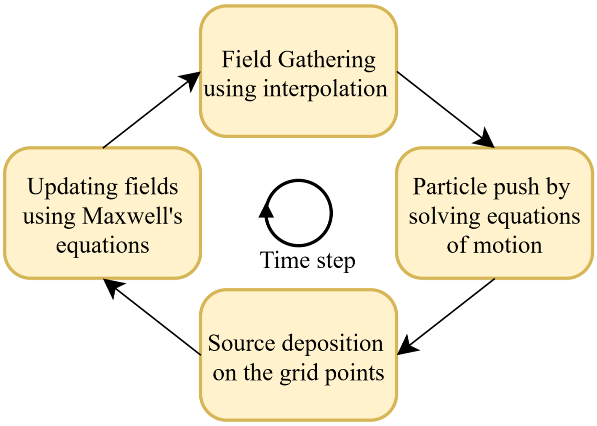
Field Gathering using interpolation
Particle push by solving equations of motion
Source deposition on the grid points
Updating fields using Maxwell's equations
Time step

Figure 4.1.: Illustration of the important steps in a PIC cycle. Completing the four steps of the PIC cycle represents an increase in the time step. One can recursively repeat the steps to simulate the self-consistent time evolution of the particles and the fields. Additional steps such as filtering source terms or applying external fields are often used between the four fundamental steps.

Initialization. To describe a single PIC cycle, we assume that both the grid and the particles have been initialized. The spatial domain is defined and discretized into a grid appropriate for the problem's dimensionality. Particles are then initialized by assigning positions and velocities that reflect the desired plasma density and temperature profiles. Each macroparticle is assigned a charge and mass, representing a group of real particles. We assume that the fields are known on the grid and the next step is to determine the fields and the position, momenta of the particles at a next time step.

PIC cycle. The first step in the PIC cycle is determining the fields at the position of the macroparticles. To achieve this, the electromagnetic fields on the grid are interpolated to the positions of the particles. This field gathering process ensures that each particle experiences the appropriate local fields based on its position within the grid. The Lorentz force acting on each particle is calculated using the interpolated fields, incorporating both electric and magnetic contributions. Once the Lorentz force is determined from the fields, the second step is to use the equations of motion to advance the particles in time and is generally called the particle pusher.

Once the particles have been pushed, the third step is to project their charges onto the grid to compute the charge density at each grid point. This process, known as charge deposition, uses weighting schemes such as the nearest grid point method or linear weighting to distribute the particle charges onto the grid. Similarly, the current density resulting from particle motions is calculated and assigned to the grid. With the charge and current densities known on the grid, the last step involves solving Maxwell's equations to update the electromagnetic fields. Numerical methods such as the finite-difference time-domain method or spectral methods are employed to compute the electric and magnetic fields at each grid point. This step is crucial, as it determines how the fields evolve in response to the charge and current distributions in the plasma. After the fields have been calculated on the grid, the process is repeated recursively at each time step to determine the time evolution of the plasma and the electromagnetic fields.

Throughout the PIC simulation process, diagnostics are performed to collect data for analysis. This includes recording particle distributions and snapshots of the field distributions at each time step. These diagnostics are essential for interpreting the results of the simulation and for validating the model against experimental or theoretical predictions. A number of different PIC codes implementing different methods for solving the electromagnetic fields and for particle pushers are available. Some of them that are relevant for the field of laser-plasma interactions include OSIRIS [153], WarpX [154], EPOCH [155], QuickPIC [156], CALDER-Circ [157], PIConGPU [158] and FBPIC [32]. In this work and chapter FBPIC was used as a test bench to showcase the advantage of the trust-MOMF technique. Hence a brief introduction to FBPIC is given in the following section.

4.2. FBPIC

Traditional fully 3D PIC codes typically discretize the spatial simulation domain using a Cartesian grid, representing electromagnetic fields and particles across all three spatial dimensions. While this approach is general and accurate, it becomes computationally expensive for high-resolution simulations, especially when modeling relativistic particles that require fine temporal and spatial discretization to maintain accuracy.

FBPIC addresses this challenge by exploiting the quasi-rotational symmetry often present in laser-plasma interactions. Instead of a full 3D Cartesian grid, FBPIC uses a cylindrical coordinate system (r,ϕ,z) and decomposes the fields into azimuthal modes using a Fourier-Bessel series. This method captures deviations from perfect rotational symmetry by including higher-order modes, effectively reducing a 3D problem to a series of 2D problems. The computational savings are significant. In cylindrical coordinates, the number of grid points scales as $N_{total} \approx m \times N_r \times N_z$ where m is the number of modes, N_r is the number of radial grid points and N_z is the number of longitudinal grid points. Typically, only a few modes $m \approx 2$ to 4 are sufficient to accurately model the physical system, leading to orders-of-magnitude reductions in memory usage and computational time compared to Cartesian grids, which scale as $N_{total} \approx N_x \times N_y \times N_z$. Another effect usually seen in common PIC codes using finite-difference time-domain (FDTD) as field solver relates to the handling of particles and fields propagating close to the speed of light. These solvers usually result in spurious vacuum dispersion relation [159] where they artificially slow down the group velocity of the laser pulse leading to spurious dephasing of the electron beam if the spatial and temporal resolution is not chosen carefully [160]. This results in an artifact known as Numerical Cherenkov radiation (NCR) [161]. This effect can be mitigated by increasing the spatial resolution of the simulation and adhering to the Courant-Friedrichs-Lewy (CFL) condition [162] that defines the largest time step possible.

Spectral Solvers. As opposed to the FDTD solvers, FBPIC employs spectral solvers that take a different approach to solving Maxwell's equations. The core innovation of FBPIC lies in its use of the Fourier-Bessel decomposition, which expresses the fields as a sum of orthonormal Bessel functions. The fields are then transformed to the spectral space by employing Fourier transform in the longitudinal direction and Hankel transform in the radial direction. In the spectral space the time integration can be performed analytically and hence this field solver is not subject to the CFL limit allowing the time step to be chosen freely. At each PIC cycle the FBPIC transforms the fields to the spectral space, advances them in time and reverts them to the real space. Due to all of the features of FBPIC discussed thus far, it was chosen to represent the numerical accelerator since all of the laser-plasma interactions discussed in this work are quasi-rotationally symmetric. FBPIC is optimized for modern computational architectures, including multi-core CPUs and Graphics Processing Units (GPUs). Certain steps of the PIC cycle are easily parallelizable like particle pushing since each push of a macroparticle is independent. The same is true for field solvers and thus GPUs offer massive parallelism with thousands of cores capable of handling simultaneous computations. FBPIC leverages this by offloading computationally intensive tasks, such as spectral transforms and particle pushing, to GPUs, resulting in significant speedups over CPU-only implementations. All of the FBPIC simulations in this work were run on GPUs either on Tesla V100 or RTX 3090 GPUs from NVIDIA.

Another important factor to consider when working with FBPIC is the weighting factor of the macroparticles. As already discussed before, macroparticles represent a large number of physical particles to make computations tractable. In a cylindrical coordinate system, the volume element increases with radius r, meaning that macroparticles located further from the axis represent a larger physical volume and, consequently, more real particles. This weighting needs to be taken into account when analysing particle beams resulting from FBPIC simulations. This weighting also ensures that the particle density and current are correctly calculated when mapping particles to the grid and when computing electromagnetic fields.

FBPIC simulations: computational parameters. Setting up an FBPIC simulation involves specifying a range of technical and physical parameters that define both the computational framework and the physical phenomena being modeled. The technical parameters establish the computational foundation of the simulation. One of the primary considerations is the spatial resolution that can be set by the grid parameters, which involves determining the number of grid points in the radial N_r and longitudinal N_z directions, as well as the number of azimuthal modes m. Selecting higher resolutions in these dimensions enhances the accuracy of the simulation but also increases the computational cost. In the case of laser-plasma acceleration, the spatial resolution can be set by recognizing the smallest length scale of interest. Longitudinally, this is the wavelength of the driving laser pulse and radially is the dimension of the bubble.

Another crucial parameter is the time step size Δt, which dictates the temporal resolution of the simulation. Although the spectral methods employed in FBPIC permit larger time steps compared to traditional finite-difference methods, the time-step must still be chosen carefully to accurately capture the fast dynamics inherent in relativistic plasma phenomena. For full 3D simulations, the number of macroparticles is usually on the order of 10^9 while for quasi-3D simulations such as FBPIC it is typically 10^6 macroparticles. As already described before, the particle pusher is a very parallelizable operation, hence a huge speedup is observed when running FBPIC on GPUs instead of CPUs. Correctly, determining all of the parameters of the FBPIC simulations can be done using convergence studies where reproduction of known physical phenomena can be used to judge the accuracy of the simulations. One example could be looking at the total charge of the accelerated beam while changing the amount of macroparticles used or while changing the grid resolution. When the total charge converges to a particular value, increasing the macroparticle count further does not result in significant improvement in accuracy. This particular value of macroparticles is then enough to simulate the relevant physics and further increases in macroparticles result in diminishing returns at the cost of significantly higher computational cost. Moreover, the nature of simulations can be exploited to further reduce the computational cost. In this work, the primary mechanism to inject electrons is shock-assisted injection therefore, it is sufficient to sample the region around the shock with a large number of macroparticles to accurately simulate the accelerated beam [163]. Similarly, the linear wakefields forming in regions of lower laser intensity result in a nearly laminar flow of particles, allowing us to decrease the macroparticle density far away from the laser axis [164]. The outer regions of the bubble are then sampled with a lower number of macroparticles resulting in different macroparticle density in different regions of the plasma. This reduces the computational cost while still maintaining the accuracy of the simulations. Since the trust-MOMF technique was used to optimize the numerical accelerator, a simulation fidelity parameter χ

was also introduced to control the spatial resolution of the simulation. The lowest possible value of this parameter 1 results in simulations that take around $40 - 60$ seconds to complete, while the highest possible value 4 can result in simulations of over 2 hours. This parameter along with other physical parameters that are discussed in the next section are controlled by the Bayesian optimizer.

Lastly, another important concept in reducing the computational load is simulating the LWFA in a different Lorentz frame, that moves with the laser and is generally called the boosted frame. The boosted frame is characterized by a boost factor γ_b which is also used to transform the different space metrics into the boosted frame. In the boosted frame of reference, the Lorentz transformation causes the laser pulse to become elongated and its wavelength to increase. This elongation permits the PIC loop to utilize larger time steps and increased spatial grid spacing, as the stretched laser fields can be adequately resolved with a coarser grid. Moreover, in the boosted frame the plasma becomes shorter, hence the time needed for the laser to propagate through the plasma is also reduced. In general, the total number of iterations required in the boosted frame is reduced by a factor of $\approx (1 + \beta_b)^2\gamma_b^2$ [165]. All of the simulations performed in this work were conducted using the boosted Lorentz frame.

FBPIC simulations: physical parameters. The physical parameters directly influence the behavior of the simulated plasma and laser interactions. Laser properties are particularly significant in simulations of laser-plasma interactions. The intensity and amplitude of the laser determine the strength of the electromagnetic field and its capability to drive plasma waves as already discussed in 2. The spot size and focusing parameters control the transverse profile and peak intensity of the laser beam, which are critical for both injection and subsequent acceleration of the electron beams.

Many of the plasma parameters are equally crucial. The density profile specifies how the electron density varies spatially within the simulation domain, which is vital for matching conditions in wakefield acceleration and for accurately modeling plasma dynamics. Information about the species and ionization states defines the types of particles present in the simulation and their initial charge states, influencing the plasma's response to electromagnetic fields. Since many of the laser parameters are not easily modified such as Laser power, laser waist, laser duration, laser wavelength, they are kept fixed in this work to the parameters of the ATLAS-3000 laser. Other parameters that are often variable in an experiment setting are allowed to be modified by the Bayesian optimizer. These parameters include the plateau plasma density, the position of laser focus, as well as the lengths of the up- and downramps of the plasma density close to the density transition. The ranges of these parameters are chosen so that different kinds of electron beams can be generated from the numerical accelerator. A summary of all free and dependent

Variable input parameters			
		min. value	*max. value*
Plateau Plasma density	n_e	$2 \cdot 10^{18}\,\mathrm{cm}^{-3}$	$9 \cdot 10^{18}\,\mathrm{cm}^{-3}$
Upramp length	l_{up}	$0.25\,\mathrm{mm}$	$1.75\,\mathrm{mm}$
Downramp length	l_{down}	$0.0\,\mathrm{\mu m}$	$50\,\mathrm{\mu m}$
Focus position	z_0	$-0.5\,\mathrm{mm}$	$2.5\,\mathrm{mm}$
Simulation fidelity	χ	1	4
Fixed input parameters			
Laser wavelength	λ_0	$800\,\mathrm{nm}$	
Laser power	P	$50\,\mathrm{TW}$	
Laser waist (FWHM)	w_0^{FWHM}	$20\,\mathrm{\mu m}$	
Laser duration (FWHM)	Δt	$30\,\mathrm{fs}$	
Dependent variables			
Plasma wavelength	λ_p	$2\pi c\sqrt{m_e\epsilon_0/e^2 n_e}$	
Plasma wavenumber	k_p	$2\pi/\lambda_p$	
Critical density	n_c	$(2\pi c/\lambda_0)^2(m_e\epsilon_0/e^2)$	
Critical power	P_c	$2m_e c^3 n_c/(r_e n_e)$	
Peak intensity	I_0	$2P/(\pi w_0^2)$	
Peak potential	a_0	$\sqrt{2I_0/\epsilon_0 c}\cdot(e/k_p m_e c^2)$	
Matched peak potential	$a_0^{matched}$	$2(P/P_c)^{1/3}$	
Matched bubble radius	r_b	$\sqrt{2a_0^{matched}/k_p}$	
Rayleigh length	z_R	$\pi w_0^2/\lambda_0$	
Waist (Gaussian beam in vacuum)	w	$\sqrt{1+(z-z_0)/z_R)^2}$	
Simulation mesh parameters			
Transverse box size	l_r	$2.5\cdot w(z=0)$	
Longitudinal box size	l_z	$25\,\mathrm{\mu m}+r_b$	
Simulation length	$l_{z,max}$	$3.5\,\mathrm{mm}$	
Transverse resolution	Δr	$600\,\mathrm{nm}/\chi$	
Longitudinal resolution	Δz	$60\,\mathrm{nm}/\chi$	
Boost factor	γ_{boost}	$\sqrt{l_{z,max}/l_z}/\chi$	

Table 4.1.: **Simulation and scan parameters.** The top section shows the four simulation parameters and their ranges that are used in the optimization problem. In addition to those, a fidelity parameter χ is introduced that allows the optimizer to choose between low and high numerical resolution. Based on the scan parameters and the fixed problem parameters, several dependent variables are calculated that can provide estimations for the correct box size for the simulations.

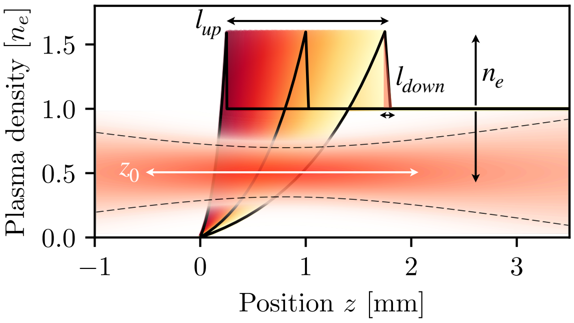

Figure 4.2.: Illustration of the four variable input parameters from Table 4.1. The three different ramps show the minimum and the maximum values of upramp length l_{up} and the downramp length l_{down}. The value after the downramp is the plasma plateau density n_e and the focus position of the laser can be modified in the region depicted by the white line z_0. The figure is reproduced from the work of the author [166].

parameters of the simulations and the range in which the variable parameters are scanned is given in Table 4.1 and illustrated in Figure 4.2.

4.3. Single-objective optimization of FBPIC simulations

This section will present and discuss the results from the numerical optimization of the LWFA. The advantages of Bayesian optimization, in general, and trust-MOMF, in particular, will also be highlighted. One of the main strengths of Bayesian optimization compared to other methods is its ability to efficiently locate the global optimum of a function using a minimal number of samples. Moreover, Bayesian optimization is adaptable; by modifying the acquisition function and the model, it can transition from optimizing a single objective to handling multiple objectives simultaneously. This adaptability is realized in multi-objective optimization (see Section 3.3.2), where Bayesian optimization can implicitly optimize multiple combinations of objectives by optimizing the *expected hypervolume improvement* [114, 123].

Despite the high sample efficiency of Bayesian optimization, solving a multi-dimensional problem still demands a substantial number of evaluations. In the case of the laser-plasma accelerator with four adjustable input parameters, 100 evaluations were typically needed to find the optimum. Considering that each simulation, as discussed in Section 4.2, requires several hours to run, a complete optimization process would take several days to compute.

To address this challenge and enable multi-dimensional optimization, the optimization process can be accelerated by utilizing low-resolution simulations that employ a coarser numerical grid and a higher boost factor (see Table 4.1). These simulations capture the essential physics of injection and acceleration but have not fully converged regarding final charge, energy, and other parameters. Despite their approximate nature, these solutions can be computed in just a few minutes on a GPU, providing valuable insights for the optimization process. Crucially, the variability in simulation resolution, and thus its *fidelity*, can be directly incorporated into the optimization process by introducing a new *fidelity variable*, χ (see Table 4.1). In a process called multi-fidelity optimization (see Section 3.4), a Gaussian process that models the objective function over the four input dimensions (n_e, l_{up}, l_{down}, z_0) as well as the fidelity parameter χ is constructed. The decision regarding the next position to probe is taken by the trust-MOMF acquisition function defined earlier in Section 3.4, which is based on the common optimization of the different objectives and an additional trust objective. Regarding the latter, the algorithm also considers the computational cost associated with the fidelity parameter. By conducting a convergence study of PIC simulations, it was found that the computing time of these simulations approximately scales with $cost(\chi) \propto \chi^{3.5}$. By integrating this cost and fidelity information into the optimization process, an average speed-up of about an order of magnitude was achieved in this work. Consequently, a complete multi-fidelity optimization run typically requires around 10 hours to compute.

Although the MOMF acquisition function is primarily designed for multi-objective optimization, it is also capable of optimizing a single objective across multiple fidelity levels as demonstrated in Section 3.4.1. To ensure a fair comparison between different optimization schemes, the results in this and the following sections, are thus obtained using the same algorithm. The specific optimization parameters used in these simulations are summarized in Table 4.2. Notably, both types of objectives were subjected to the same constraints in terms of the maximum number of iterations and the computational budget. Additionally, to assess typical performance, each optimization was executed five times, starting from five different random initial points.

4.3.1. Defining single-objectives

The optimization conducted in this study aims to produce quasi-monoenergetic electron beams with a high total charge concentrated around a specific target energy E_0. From a statistical perspective, achieving this involves minimizing the difference between the beam's central energy tendency and the target energy, reducing the statistical dispersion or energy spread, and maximizing the total charge, which is represented by the integral of the electron beam spectrum. These

Optimization parameters		
Number of Trials	n_{TRIALS}	5
Max. number of Iterations	n_{BATCH}	150
Maximum Cost	C_{total}	50 GPU hours
Number of initial points	n_{INIT}	5
Input Dimensions	dim_x	5
Output Dimensions	dim_y	1 (single objective) or 3 (multi-objective)
Cost Function	$cost(\chi)$	$\propto \chi^{3.5}$

Table 4.2.: **Summary of optimization parameters used for numerical accelerator.** Some of the key parameters when running the trust-MOMF algorithm are summarized here. To terminate the optimization process efficiently, two upper thresholds have been established: the maximum number of iterations and the total computational cost. The optimization was concluded once either of these limits was reached. In the case of single-objective runs, the output dimension was set to one, whereas for the multi-objective case, trust-MOMF algorithm simultaneously optimized three objectives. It's important to note that, due to the use of adaptive meshes in this work, the cost function was approximated rather than calculated exactly.

characteristics can be quantified using various statistical measures, including standard deviation, median absolute deviation, mean energy, median energy, and total charge [167]. This variety means there is considerable flexibility in how these objectives can be encoded into a single *scalarized* objective function. Each choice of objective function tends to emphasize certain outcomes, which can lead to significant differences in the optimization results. In the subsequent paragraphs, several different objective functions are introduced, that are designed to achieve the same overarching goal: to simultaneously maximize the total charge, minimize the spectral width, and reduce the deviation from the target energy. By exploring these different formulations, the influence of scalarization on the optimization process and the resulting beam characteristics can be highlighted.

Examples. An objective function based on the mean energy and the standard deviation can be defined as follows

$$O_1 = \frac{Q^{\frac{1}{2}}}{\Delta\bar{E}^2 \cdot \sigma_E} \tag{4.1}$$

where Q denotes the total charge of the electron beam and σ_E represents the standard deviation of the energy spectrum. The term $\Delta\bar{E}^2$ is calculated as $|\bar{E} -$

$E_0|^2 + \epsilon$, where $\bar{E}$ is the mean energy of the spectrum, E_0 is the target energy and ϵ is a small offset introduced to prevent the objective function from becoming infinite when the mean energy approaches the target energy. Specifically, I used $\epsilon = 1\,\mathrm{MeV}$, considering that beams within 1 MeV of the target energy are deemed sufficiently optimized.

A characteristic property of the mean is that it tends to be influenced by values that are far from the target, especially in datasets with outliers. As a result, in the presence of noise, it is often more suitable to use median-based descriptors instead of mean-based ones. An objective function that leverages median statistics can be formulated as:

$$O_2 = \frac{Q^{\frac{1}{2}}}{|\Delta\tilde{E}| \cdot E_{MAD}}, \tag{4.2}$$

where Q represents the total charge of the electron beam, $\tilde{E}$ is the median energy of the spectrum, E_0 is the target energy, $|\Delta\tilde{E}| = |\tilde{E} - E_0| + \epsilon$ is their absolute distance (plus offset) and E_{MAD} is the median absolute deviation around the median. In this instance, a square root is applied to the total charge to decrease its emphasis within the objective function. This scaling ensures that while charge remains an important factor, it does not overwhelmingly dominate the optimization process.

The use of $Q^{1/2}$ in the previous objective function is entirely empirical, and one can just as effectively define alternative versions by applying different exponential weights to the charge Q. For instance, the following two objective functions can be considered:

$$O_{2,a} = \frac{Q^2}{|\Delta\tilde{E}| \cdot E_{MAD}}, \tag{4.3}$$

and

$$O_{2,b} = \frac{Q^3}{|\Delta\tilde{E}| \cdot E_{MAD}}, \tag{4.4}$$

which should incentivize the optimizer to find beams with higher total charge.

Previously, we touched upon the general issue that objective functions involving division can become unbounded as the denominator approaches zero. Instead of mitigating this problem by introducing offsets, it may be more advantageous to reformulate the objective function by eliminating the division operation. One way to achieve this is by implicitly optimizing for the target energy and energy spread by maximizing the charge within a specific energy window. This approach can be expressed as

$$O_3 = 2Q_{in} - Q, \tag{4.5}$$

Objective definitions		
Objective 1	O_1	$Q^{0.5}((\Delta\bar{E}^2)\sigma_E)^{-1}$, Equation (4.1)
Objective 2	O_2	$Q^{0.5}((\lvert\Delta\tilde{E}\rvert)E_{MAD})^{-1}$, Equation (4.2)
Objective 2_a	O_{2a}	$Q^{2}((\lvert\Delta\tilde{E}\rvert)E_{MAD})^{-1}$, Equation (4.3)
Objective 2_b	O_{2b}	$Q^{3}((\lvert\Delta\tilde{E}\rvert)E_{MAD})^{-1}$, Equation (4.4)
Objective 3	O_3	$2Q_{in} - Q$, Equation (4.5)
Charge-related metrics		
Q	Total integrated charge	
Q_{in}	Charge within an energy interval $E_0 \pm \Delta E$	
Central tendency metrics		
$\bar{E}$	Mean energy	
$\tilde{E}$	Median energy	
E_0	Target energy (300 MeV)	
$\Delta\bar{E}^2$	Mean-squared difference of median and target energy	
$\lvert\Delta\tilde{E}\rvert$	Absolute difference of median and target energy	
Statistical dispersion metrics		
σ_E	standard deviation	
E_{MAD}	median absolute deviation	

Table 4.3.: **Summary of single-objective functions.** The five single-objective scalarized functions that are optimized in this study are shown at the top. The middle and lower parts of the summary display the metrics related to charge, central tendency, and statistical dispersion employed to construct these single objectives. These are also used later in the multi-objective multi-fidelity optimization.

where Q_{in} denotes the charge within a specified energy range ΔE centered around the target energy E_0. Mathematically, Q_{in} is given by

$$Q_{in} = \int_{E_0-\Delta E/2}^{E_0+\Delta E/2} Q(E)dE. \tag{4.6}$$

A summary of the single objectives and definitions of metrics used to define them is outlined in Table 4.3.

4.3.2. Optimization results

The various “logical” objectives described in the preceding section were each applied to optimize the simulated laser wakefield accelerator, and the outcomes of

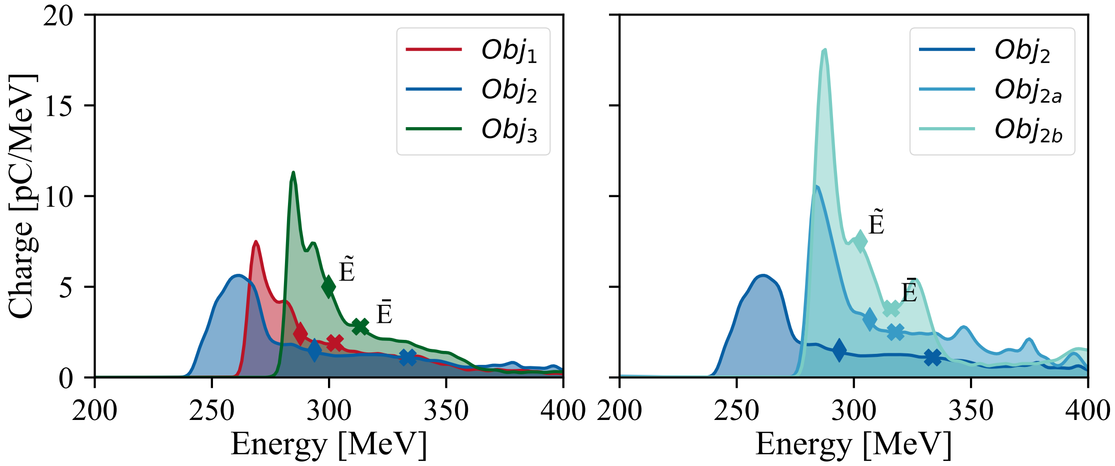

Figure 4.3.: Single-objective optimization spectra. On the left is the final spectra obtained after single-objective optimization of FBPIC simulations using three different objectives (O_1, O_2 and O_3) to optimize beam charge, beam distance from target energy (300 MeV) and energy spread. Median energy $\tilde{E}$ and mean energy $\bar{E}$ of each spectrum is indicated using diamond and cross markers respectively. On the right is a demonstration of the effect of changes in objective weight. Variations of O_2 objective with charge squared ($O_{2,a}$) or charge to the power of three ($O_{2,b}$) was used, leading to higher overall charge in the beam and - without explicit optimization - more peaked spectra.

these optimizations are presented here. In Figure 4.3, the final spectrum of the three scalarized objectives O_1, O_2 and O_3 is shown on the left sub figure. The first objective relies on mean energy, making it more sensitive to outliers when compared with the second objective. This explains why the first spectrum tends to have a suppressed high or low-energy tail. On the other hand, the second objective using median energy and median absolute deviation allows for a high-energy tail while still keeping the median near 300 MeV. Since this spectrum has a longer tail the mean of this spectrum is higher than 300 MeV. The third objective in Figure 4.3 yields a noticeably higher peak charge because it places additional implicit emphasis on the total charge near this region. The beam here has an even shorter tail since it explicitly penalizes any beam charge outside the $250-350$ MeV range. Overall, these results demonstrate how different ways of scalarizing the statistical measures produce distinct energy spectra. Furthermore, when beams are implicitly optimized to remain near the target energy similar to the third objective, it often yields more favorable outcomes than explicitly targeting that energy.

In the next step, the comparison of different forms of the second objective (O_2, $O_{2,a}$ and $O_{2,b}$) where the total charge Q is weighted by $Q^{1/2}$,Q^2 and Q^3 respectively is shown in Figure 4.3. As expected by design of the objective, the higher weight increases the total charge in the optimized beam spectrum. We also see that this choice of hyperparameters results in objectively better beams than the O_2

and O_{2a} variations when considering the energy and bandwidth of the beams. However, it is not possible to know the right choice of hyperparameters in advance. This choice of different hyperparameters adds another layer of complexity to the optimization process. Consequently, for each new problem, operators or users must carry out multiple optimization runs to identify the most appropriate objectives and parameter combinations when using single-objective optimization schemes.

Another notable aspect of these spectra is that the energy with the highest spectral charge density, which is referred to as peak energy E_{peak}, is farther from the target energy than either the mean or median energies. This discrepancy arises because highly charged electron beams cause beam loading effects in laser wakefield accelerators, leading to skewed spectra [163]. In such asymmetric distributions, the peak does not coincide with the mean or median, making explicit optimization of the peak energy necessary. The problems of hyper-parameter choice and the explicit optimization of the peak energy are discussed in the next section.

4.4. Multi-objective optimization of numerical accelerator

As demonstrated in the previous section, a central difficulty in single-objective optimization of complex systems is that the optimal weighting of hyperparameters is not known *a priori*. Achieving a higher value in one particular objective often involves a trial-and-error process of adjusting weights. In addition, multi-objective problems can exhibit trade-offs among different objectives. Consequently, changing the weight for one objective can inadvertently benefit or harm other objectives. Single-objective optimization is thus invariably biased toward one particular trade-off, and it is generally hard or impossible to detect this bias beforehand. As a result, the final solution may fail to deliver the balance of parameters that a user or operator desires.

A more versatile strategy is to directly explore the trade-off among different objectives and then select the most appropriate combination of objectives *a posteriori*.

This approach produces the Pareto front (see Section 3.3.2) in the output space, and the Pareto set in the input space. As a reminder, a point is said to dominate another if it is at least as good in every objective and strictly better in at least one. Consequently, the Pareto front consists of all non-dominated solutions. The region covered by these points in objective space is quantified by the hypervolume, which serves as an indirect measure of solution diversity. In Bayesian optimization the expected hypervolume improvement can therefore be used to optimize different objectives simultaneously. In this work, the mean energy difference $\Delta\bar{E} = |\bar{E} - E_0|$,

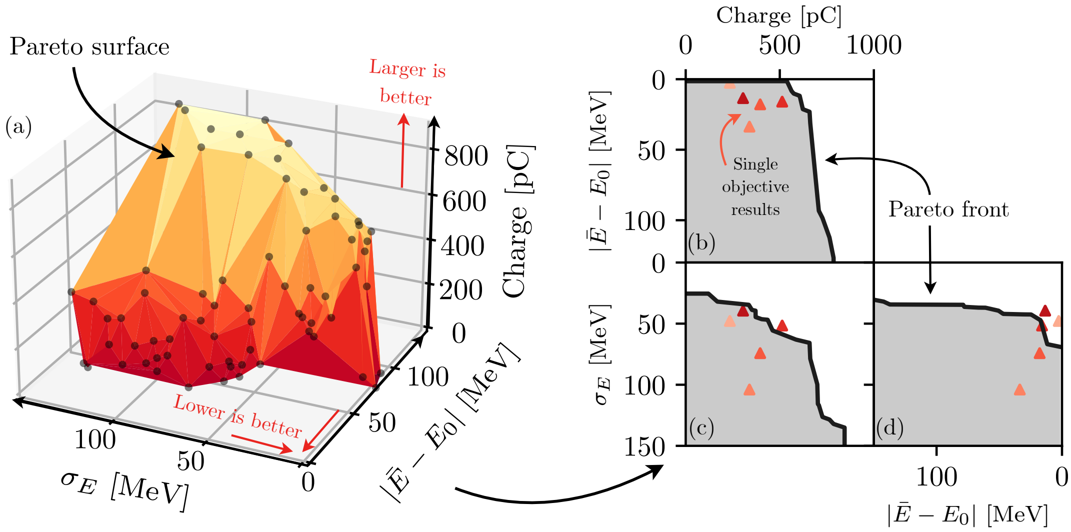

Figure 4.4.: Multi-objective optimization. (a) This figure illustrates the Pareto surface defined by the non-dominated solutions across the three objectives. Figures (b–d) show 2-D projections of this surface, showing the Pareto front for the following objective pairs: charge versus energy distance (b), charge versus energy spread (c), and energy spread versus energy distance (d). This result demonstrates that a single multi-objective optimization produces either similar or better outcomes than all of the single-objective runs. (The figure has been reproduced from a publication generated during this work by the first author) [166]

the standard deviation σ_E and total charge in the beam Q were chosen as individual objectives spanning the output space.

Results and discussion. Figure 4.4 presents the outcome of a representative run of the multi-objective Bayesian optimization. By querying the Gaussian process model, a collection of solutions were acquired that can be displayed as a Pareto surface, representing all non-dominated points in the three-dimensional output space. Panels 4.4(b–d) in the figure illustrate the corresponding projections of the Pareto surface resulting in Pareto fronts for the different objective pairs. The red triangles indicate the beam parameters obtained from the optimizations discussed in the preceding section. These results demonstrate that multi-objective optimization achieves performance on par with the combinations of objectives introduced in Section 4.3.1.

Figure 4.4 illustrates several typical trade-offs found in multi-objective problems, many of which can be understood through underlying physical processes. One clear example is visible in Figure 4.4b, where an increase in the distance to target energy is seen when the total charge exceeds 500 pC. This effect arises primarily from beam loading [163]: as the charge of the electron bunch increases, it reduces the strength of the wakefields, which consequently leads to lower mean energy and

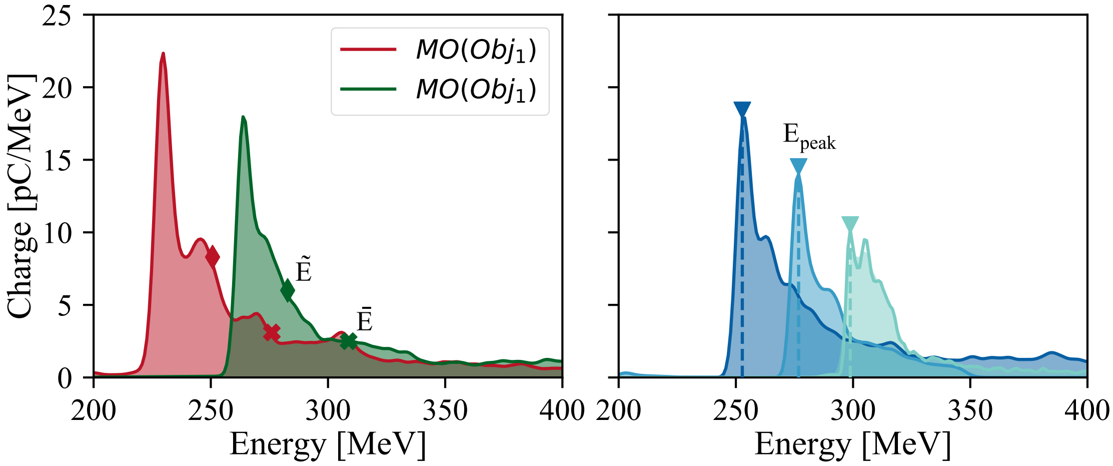

Figure 4.5.: Selected spectra obtained via one multi-objective optimization run. On the left are the spectra selected as optimal lower confidence bound solutions for the objectives O_1 and O_2. On the right are solutions optimized for peak energies of 250 MeV, 275 MeV and 300 MeV.

thus, an increase in the distance to the target energy.

where higher total charge leads to a broader energy spectrum, reflecting the different input parameters needed to produce a high-charge beam as opposed to a quasi-monoenergetic beam. This phenomenon, also reported in earlier studies (e.g., Götzfried et al.[163]), underscores the difficulty of simultaneously maximizing charge and maintaining a narrow energy spread.

Another trade-off is apparent in Figure 4.4c where higher total charge leads to a broader energy spectrum. This indicates that the input parameters that yield a beam with a higher total charge are different from the ones that produce quasi-monoenergetic beams, an effect reported in earlier studies, e.g. in Götzfried et al.[163]. Finally, another notable result, albeit not directly visible from the plots, is the absence of any high-energy beams with low charge. This is a result of the design of the three objectives used in this work. None of the objectives benefit from high-energy beams with lower charge. The total charge objective worsens and the optimization prefers higher charged low energy beams. Higher mean energy also has a negative impact on the distance to the target energy objective since as the energy increases, the distance to $E_0 = 300\,\text{MeV}$ also increases. Hence, most beams are restricted to energies near or lower than the target energy.

Multi-objective Bayesian optimization proves particularly valuable because it uncovers the inherent trade-offs within a system, such as a laser-plasma accelerator, and provides two key advantages. Firstly, it allows for a thorough characterization of the system's performance with regard to each objective. Secondly, it delivers a set of solutions that are not strongly biased toward specific objective combinations. By using hypervolume as the objective measure, this approach avoids issues that can arise in single-objective optimization, such as the inclusion of offset values in denominators. Since hypervolume only increases modestly when focusing on a single objective, multi-objective optimization does not disproportionately exploit any one objective at the expense of the others.

As noted earlier, a key advantage of this optimization approach is that the Gaussian process model can be queried efficiently, providing immediate predictions of the means and variances for each individual objective (here Q, σ_E and $\Delta\bar{E}$) based on a chosen set of input parameters x (i.e. n_e, l_{down}, l_{up} and z_0). These predictions can be combined to form a new objective function $O(x)$ and its overall uncertainty can be estimated by propagating the variances from the individual objectives.[1] A conservative solution candidate $\hat{x}$ can then be identified by considering the lower confidence bound

$$\hat{x} = \underset{x}{\operatorname{argmax}}\{\mu(O(x)) - \sigma(O(x))\}. \tag{4.7}$$

Figure 4.5 depicts such inferred solutions for the previously defined objectives O_1 and O_2, see Equation (4.1) and Equation (4.2), respectively. Due to the increased charge in these beams, the value of O_1 is approximately 40% higher than its single-objective counterpart result (see Figure 4.3). Meanwhile, the result for O_2 aligns with prior observations, with the multi-objective approach reaching about 90% of the corresponding single-objective value. This lower value is likely explained by the fact that the optimizer focuses on mean energy as an objective rather than exclusively targeting the median energy.

It is worth noting that in these candidate solutions, the spectral peaks occur at approximately 230 MeV and 272 MeV, respectively, placing them significantly below the "target" energy of $E_0 = 300\,\text{MeV}$. As previously discussed, this happens because, in highly skewed spectra, neither the mean nor the median coincides with the peak energy E_{peak}. This issue can be addressed without starting a fresh optimization by leveraging the existing multi-objective scan to build a Gaussian process that predicts E_{peak} for any given input x. Then, suitable candidates can be selected by using

$$\hat{x} = \underset{x}{\operatorname{argmin}}\{\|E_0 - \mu(E_{peak}(x))\| + \sigma(E_{peak}(x))\}. \tag{4.8}$$

where the lower confidence bound is adopted for minimization, making use of both prior solutions and uncertainty estimates.

[1] For a generic objective of the form $O(x) = x_1/(x_2 \cdot x_3)$ the uncertainties $\sigma(x_i)$ propagate as $\frac{\sigma(O(x))}{\mu(O(x))} \approx \sqrt{\left(\frac{\sigma(x_1)}{\mu(x_1)}\right)^2 + \left(\frac{\sigma(x_2)}{\mu(x_2)}\right)^2 + \left(\frac{\sigma(x_3)}{\mu(x_3)}\right)^2}$

In most cases, promising results are found immediately or after $1 - 2$ additional iterations, which refine the Gaussian process with the outcomes of a new candidate. The results of this process are shown in Figure 4.5b, showing that the multi-objective results can even translate to objectives that differ substantially from the three objectives directing the hypervolume search. Consequently, multi-objective Bayesian optimization greatly simplifies the process of locating both optimal parameter settings and optimal objective formulations. The latter can be evaluated *a posteriori* at negligible computational cost and can inform subsequent single-objective optimizers focused on refining a specific objective. This procedure is especially suitable for systems that are not yet well understood, such as newly established experiments or simulations. For systems that are already well characterized, however, a carefully chosen objective—like Equation (4.5) with a suitably defined energy window—can deliver competitive results.

Finally, the benefit of incorporating lower-fidelity simulations becomes clear in Figure 4.6. With the same computational budget, the Trust-MOMF algorithm attains a hypervolume of about 85%, while the multi-objective (MO) optimization only reached roughly 58%. Even when the MO optimization is extended by an additional 480 minutes, it fails to exceed 68% hypervolume. The fidelity distribution displayed in Figure 4.6 further demonstrates that Trust-MOMF frequently uses simulations at fidelities below 0.5, thereby making efficient use of faster, low-resolution simulations. Comparing costs shows that Trust-MOMF reduces the overall cost by a factor of approximately 7, closely matching the improvements observed with test functions in Section 3.5.1. These promising outcomes highlight the value of joint MOMF for computationally expensive numerical simulations in physics.

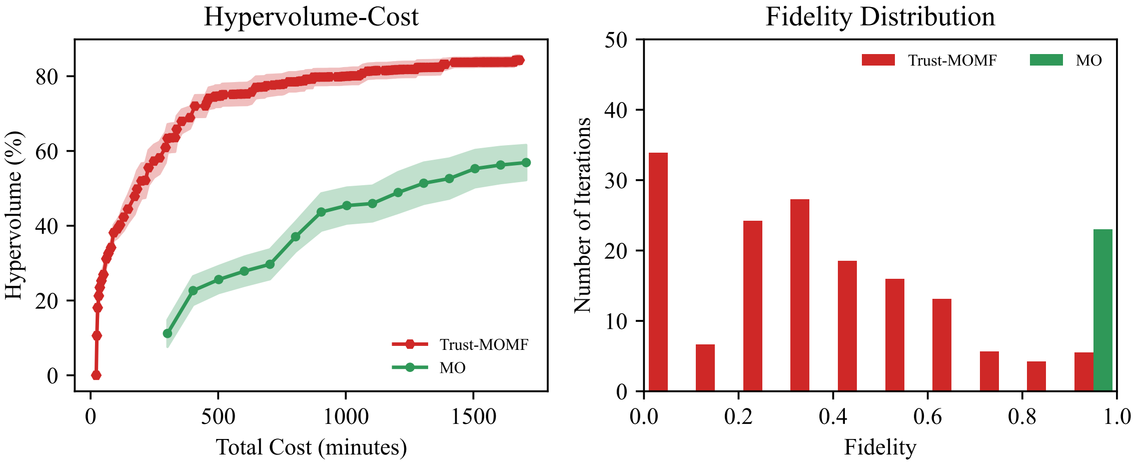

Figure 4.6.: Benchmark with PIC simulations of laser wakefield accelerator using a 4-D input space. On the left, the mean hypervolume, across ten trials, is plotted for Trust-MOMF and MO optimizations as a percentage, plotted against the elapsed time in minutes. The shaded area denotes the standard deviation over those ten trials, and each optimization run was capped at a 30-hour computational budget. On the right, the number of sampled points at varying fidelity levels for a single trial of both Trust-MOMF and MO optimizations is depicted.

4.5. Input-space analysis

In the previous section, the focus has mainly been on how single- and multi-objective optimization perform. Another important strength of Bayesian optimization lies in analyzing the model that emerges from the optimization itself, which can reveal physical insights and parameter dependencies of the system. In this section, we examine how different input parameters specifically affect each objective by training a Gaussian process model using data from our multi-objective runs.

In Figure 4.7, the influence of input parameters on the objectives, either viewed in pairs or considered individually, is depicted. To create them, the pairwise combinations of input parameters are taken while averaging over the remaining two, consequently producing six two-dimensional plots. The color scale in each plot represents the corresponding objective's value. The effect of individual parameters can also be investigated by averaging over the other three parameters and repeating this process for each objective and parameter. From these visualizations, certain trends emerge, which can in some cases be explained through physical reasoning.

Density (n_e): Increasing the plasma electron density initially increases the total beam charge, but then, somewhat unexpectedly, starts to drop. Examining the underlying PIC simulations confirms that while higher densities initially lead to greater injection, part of the injected electron population is lost toward the end of

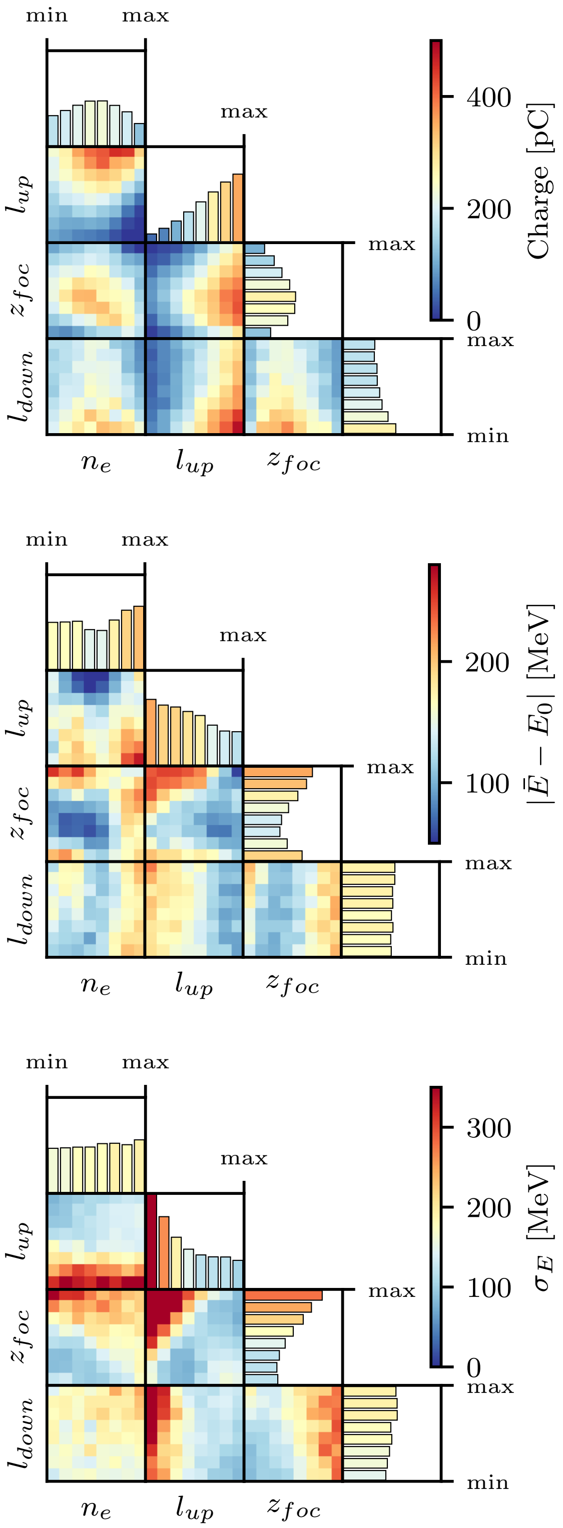

Figure 4.7.: Input space visualization. The pair plots visualize how the four input parameters (n_e,l_{down},l_{up} and z_0) influence the three output metrics (charge, energy distance and energy spread). Meanwhile, the one-dimensional histograms illustrate the averaged effect of each individual input parameter. The figure is reproduced from the work of the author [166].

the accelerator region due to dephasing and defocusing fields at higher densities. Interestingly, the density yielding optimal energy aligns with the density yielding optimal charge, implying that there is a density at which beam loading is ideal for achieving the target energy of 300 MeV. Meanwhile, the energy spread tends to grow with density and exhibits an intriguing correlation with the focal plane.

Upramp length (l_{up}): A nearly linear increase in injected charge is observed as the upramp length increases. This phenomenon is likely connected to laser self-focusing, since a more extended upramp allows the laser to self-focus more strongly, producing higher laser intensities at the injection point. The data also suggest that having a longer upramp can help the beam reach the target energy. A larger energy spread observed at shorter upramp lengths likely arises from broadband electron beams lacking a pronounced spectral peak, as evidenced by the notably low charge at these shorter upramp lengths.

Downramp length (l_{down}): Varying the downramp length directly affects the injected charge, since shorter downramps translate to a faster expansion of the wakefield and thereby enhance electron injection. Meanwhile, within the parameter ranges explored in this study, the downramp length does not appear to influence the mean energy significantly, which makes sense given that the downramp primarily influences the injection point rather than the effective acceleration length.

Focus position (z_{foc}): There is a particular focus position that maximizes the total charge and yields a beam energy closer to the target energy. However, the energy spread at this position is relatively larger than the smallest possible values, likely due to the increased beam loading that accompanies a higher charge.

For some parameter combinations, some trends can also be observed. For instance, when the plasma density is higher or the upramp length is longer, the optimal focus position for achieving maximum charge shifts upstream. These combinations can be detected by observing the 2D pair plots and can offer insights into the acceleration process for the users running the accelerators.

5. Experiment Optimization

After establishing the utility of Bayesian optimization for numerical FBPIC simulations in the previous chapter, I will now extend these methods to an experimental laser wakefield accelerator setting. As in the numerical case, the experimental setup can be modeled as a black-box system with adjustable input parameters that elicit a measurable system response, which is subsequently fed back into the Bayesian optimizer. A schematic illustrating the feedback loop is provided in Figure 5.1, where a closed-loop feedback diagram outlines the optimization cycle. Given the basic similarities between the numerical and experimental workflows, the optimization strategies developed in the computational domain can be applied directly to experimental studies with relatively few modifications.

However, key distinctions exist between numerical LWFA simulations and real-world experiments, particularly in the nature of controllable inputs, the types of measurable outputs, and the presence of experimental fluctuations arising from variations in the laser and plasma conditions. In contrast to simulations, where parameters such as the upramp and downramp lengths can be precisely controlled, experimental setups typically allow only indirect manipulation of these quantities. Furthermore, while simulations operate under idealized and repeatable conditions, real experiments are subject to shot-to-shot variations, particularly in high-power laser systems. These fluctuations necessitate the development of noise-mitigation techniques to ensure robust optimization performance.

To address these challenges, the chapter begins with an overview of the ATLAS-3000 laser system, which is built upon the ATLAS-300 platform. This discussion aims to provide context for understanding the sources of experimental noise and variability. For a more comprehensive technical description of the ATLAS-300 and ATLAS-3000 systems, readers are referred to the dissertations of M. Gilljohann [168] and F.M. Foerster [169]. Following this, I describe the experimentally controlled input parameters and how they correspond to the variables optimized in the numerical framework. This section also provides a detailed account of the communication protocol used to interface the Bayesian optimizer with the experimental hardware, ensuring seamless data exchange and real-time feedback for efficient optimization. A crucial aspect of experimental optimization is the measurement of electron beams, which requires the use of an electron spectrometer for objective

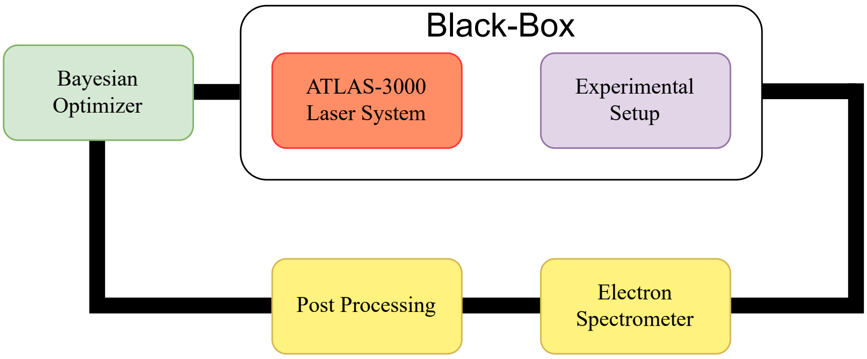

Figure 5.1.: The general experimental Bayesian loop where the optimizer treats the experimental setup and the laser system as a black-box function. The electron spectrometer captures the system's response, while a post-processing module extracts relevant features that inform the optimizer. These extracted features are then structured as a vectorized objective function, guiding the optimization process toward improved performance.

evaluation. Under certain conditions, electron beams appear in pairs, necessitating the implementation of a data-cleaning approach based on a Gaussian Mixture Model to accurately cluster and distinguish beam features.

The subsequent sections present the results of various multi-objective optimization strategies applied to the experimental LWFA, followed by an extension to multi-fidelity, multi-objective optimization. Finally, I will highlight one of the most significant contributions of this work—the demonstration of systematic peak energy tuning in an experimental LWFA by simultaneously adjusting eight different parameters within 1-3 iterations, marking an important step toward precision control in laser-plasma acceleration.

5.1. ATLAS-3000 laser system

The experiments were conducted using the ATLAS-3000 Titanium-Sapphire chirped-pulse amplification (CPA) laser system, at the Center of Advanced Laser Applications (CALA) in Garching, Germany. Chirped Pulse Amplification [19] is a fundamental technique employed in laser science to generate ultra-intense laser pulses while mitigating the risk of damage to optical components during amplification. The process begins with pulse stretching, where a short laser pulse is temporally expanded using a dispersive optical system, such as diffraction gratings. This stretching spreads the pulse's frequency components over time, significantly lowering its peak power and minimizing nonlinear effects or material damage in subsequent amplification stages. Once stretched, the pulse undergoes amplification in a laser gain medium, where its energy is increased through multiple stages of amplifiers while maintaining a reduced power. Finally, the amplified pulse is recompressed using a second dispersive optical system with opposite dispersion in a process known as pulse compression. This restores the pulse to its original short duration while preserving the amplified energy, resulting in an ultra-high-power laser pulse. CPA is the foundation of modern high-intensity laser systems, enabling petawatt-scale peak powers that are crucial for driving nonlinear wakefields in plasma acceleration. This technique is fully implemented in the ATLAS-3000 laser system, illustrated in Figure 5.2.

The system begins with seed pulses of approximately 6 fs generated from a mode-locked Titanium-Sapphire oscillator operating at a repetition rate of 80 MHz. A Pockels cell reduces this repetition rate to 10 Hz, after which the pulse energy is increased to 500 µJ through an initial multipass amplifier. Before undergoing further amplification, the pulse is stretched to approximately 800 ps using a grating stretcher. To enable fine dispersion control and spectral shaping, the stretched pulse is then directed through an acousto-optic programmable dispersive filter known as DAZZLER [170]. This filter allows for pre-compensation of high-order dispersion accumulated throughout the amplifier chain, ensuring optimal pulse quality at the end. The dispersion orders of the laser pulse, which influence its shape and ultimately its duration, are among the key parameters optimized using Bayesian optimization. The DAZZLER is controlled via proprietary software provided by Fastlite, through which the Bayesian optimizer adjusts the dispersion parameters in real time.

Following dispersion management, the pulse undergoes further amplification in a regenerative amplifier and a multipass amplifier before passing through a second Pockels cell, which further reduces the repetition rate to the final 1 Hz. The pulse energy is subsequently increased to its final value of 20 J through a series of multipass amplifiers. To maintain beam quality, spatial filters are incorporated after the third, fourth, and fifth multipass amplifiers to remove high-spatial-frequency components. These filters are housed within telescopes, which also expand the beam to a diameter of 9 cm before entering the last amplifier, thereby reducing the fluence and preventing optical damage. In addition, deformable mirrors coupled with wavefront sensors are utilized within a closed-loop system to correct wavefront distortions and ensure a high-quality beam profile.

The pulse is expanded to a final diameter of 27 cm and then compressed in a vacuum using a grating compressor, achieving a final full-width-at-half-maximum (FWHM) duration of 30 fs. Diagnostics of the compressed pulse are performed using a self-referenced spectral interferometer, WIZZLER [171], and a frequency-resolved optical gating (FROG) device, GRENOUILLE [172]. To optimize pulse compression, a closed feedback loop is implemented between the WIZZLER and DAZZLER, wherein the spectral phase measured by the WIZZLER is corrected by adjusting the dispersion settings in the DAZZLER. Once the pulse has been optimally compressed, it is directed via a series of mirrors into the Electron and Thomson Test Facility (ETTF) beamline, where laser-plasma acceleration experiments are conducted.

One issue identified during the course of this study was the severe blackening of beamline mirrors, which progressively reduced the transmitted energy reaching the ETTF. Since transmission measurements were not performed before each experimental session, a conservative estimate of transmission was assumed. However, this did not significantly affect the Bayesian optimization process, as the degradation occurred over several experimental days, whereas the optimizer operates on an hourly timescale during each session. The primary consequence of mirror degradation on the electron beams in this work was a reduction in the total charge and energy of the accelerated electron beam rather than a fundamental limitation of the optimization approach. Based on available transmission measurements, the overall energy transmission through the compressor and beamline was estimated to be $(50 \pm 10)\%$. Despite these challenges, the optimization process remained effective in adapting to experimental conditions and improving the performance of laser wakefield acceleration experiments.

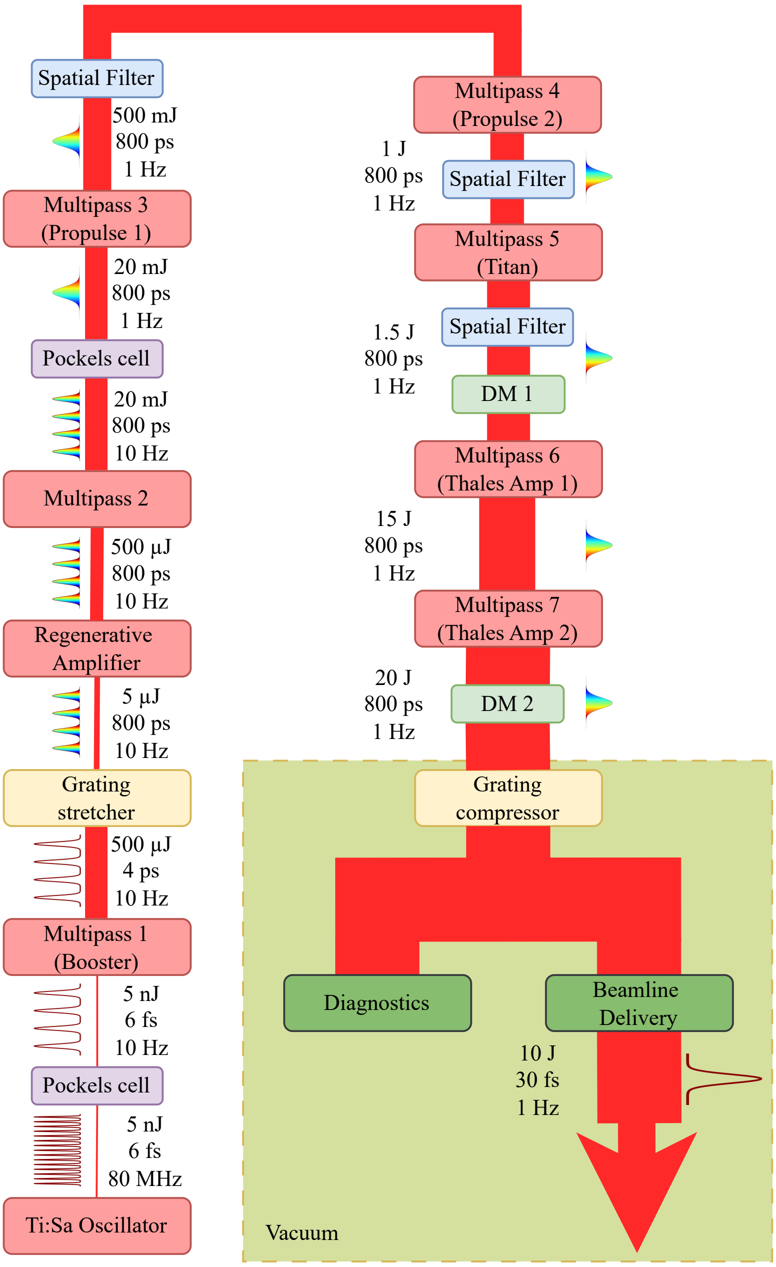

Figure 5.2.: Sketch of the ATLAS 3000 Laser system adapted from [169]. The DM refers to the deformable mirrors within the laser chain to correct wavefront errors accumulated during propagation.

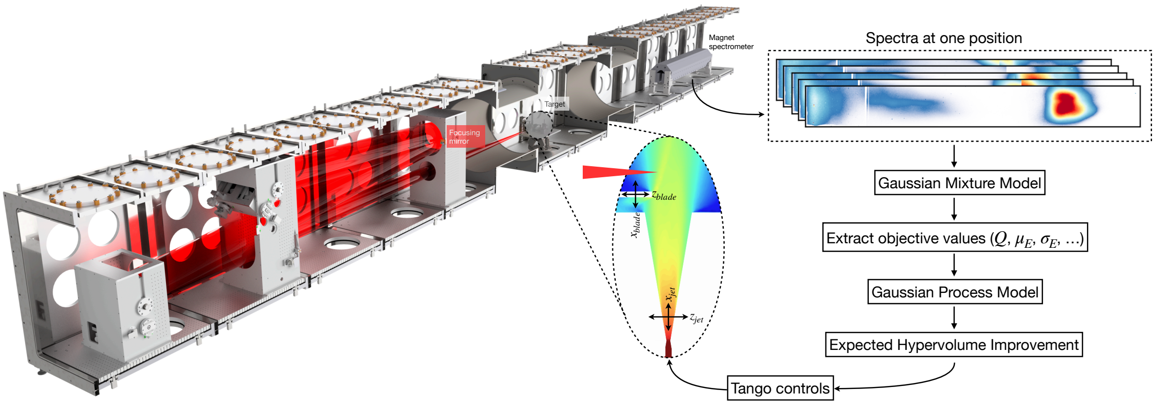

Figure 5.3.: Schematic of the experimental setup and optimization workflow. The laser is focused in an f/44 geometry onto a movable gas target and shock-inducing blade, with an inset showing a fluid simulation of the perturbed flow. A dipole magnet spectrometer serves as the main diagnostic. For each setting, 5–10 shots are recorded, cleaned using a Gaussian Mixture Model, and objective values are extracted. A Gaussian Process model is updated and queried for the next input via expected hypervolume improvement, with the optimal setting applied through the Tango control system.

5.2. Experimental setup and diagnostics

This section describes the experimental setup used to generate relativistic electron beams and the associated diagnostics. A general schematic of the setup and the data processing pipeline is shown in Figure 5.3. Following an overview of the beam focus and target system, the electron spectrometer is introduced as the primary diagnostic for characterizing the accelerated beams. The spectral data recorded by the spectrometer is then processed through a dedicated pipeline that extracts key beam parameters from raw images. Lastly, the various optimization strategies employed during the experiments are outlined, providing the foundation for the subsequent analysis and discussion of performance results.

5.2.1. Beam focusing

The beam from ATLAS-3000 is directed through a series of turning mirrors and ultimately focused onto the target using a spherical mirror with a focal length of 10 meters, resulting in an f-number of 37. Moreover, to improve beam quality and experimental stability, an aperture of 9.2 cm is positioned 6 m downstream of the spherical mirror. At this point, the beam diameter is estimated to be 10.8 cm, meaning the aperture symmetrically clips about 0.8 cm from each side of the beam, leading to a modified f-number of 43. This clipping serves to remove the outermost, higher-order spatial modes, which in this work degraded focus quality and reduced the reproducibility of the electron beam generation. Although the

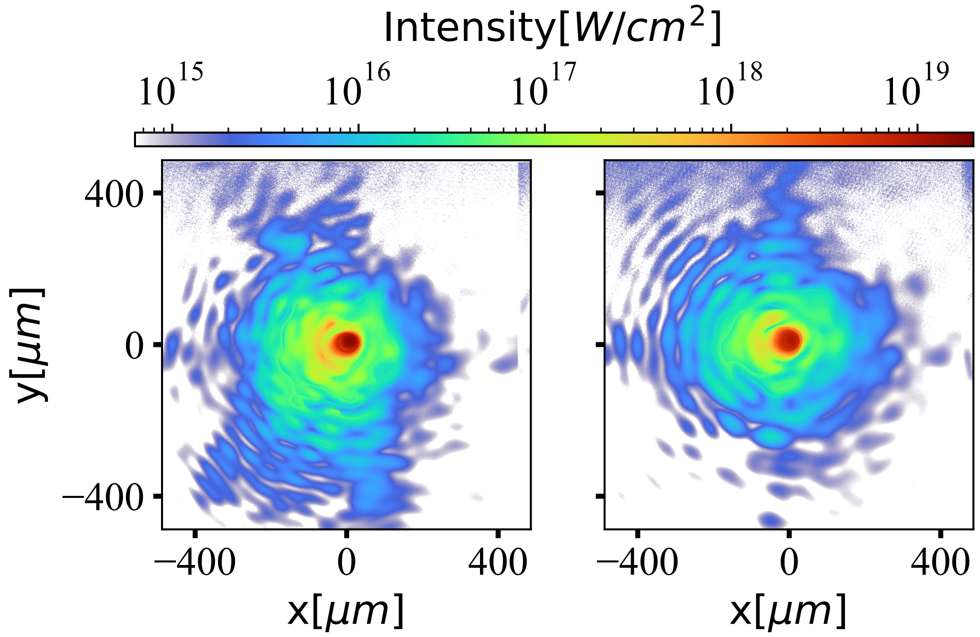

Figure 5.4.: Two representative reconstructed HDR images of focus are shown in this figure. The left image of the focus is without an aperture in the beam path while the right focus represents a focus with a 9.2 cm aperture in the beam path. The focus with the aperture has more uniform widths in both x and y dimensions compared to the focus without the aperture.

aperture reduces the total transmitted laser energy by approximately 27.7%, this has no adverse effect on the Bayesian optimization process. Instead, it primarily limits the achievable charge and energy of the resulting electron beams, while enhancing shot-to-shot consistency.

Before each experimental run, the wavefront and focus quality of the laser beam were measured under vacuum conditions and actively corrected using a closed-loop feedback system comprising a deformable mirror and a Shack-Hartmann wavefront sensor. While using the focus diagnostics, the beam is attenuated and sent to a beamsplitter that directs the two parts of the beam to a far-field camera and a Shack-Hartmann wavefront sensor. The far-field camera recorded the spatial intensity distribution of the focal spot using different neutral density filters with varying attenuation levels. These individual exposures were subsequently stitched together to produce a high dynamic range (HDR) image of the focus, capturing intensity variations across four orders of magnitude. This HDR image was then used to determine the focal spot size and estimate the corresponding on-target intensity.

Two representative HDR images of the focus are shown in Figure 5.4, comparing the f/37 focusing geometry without an aperture (left) and with a 9.2 cm aperture introduced into the beam path (right). The impact of the aperture on beam quality is clearly visible with a more symmetric focus profile. Without the aperture, the full-width at half maximum (FWHM) spot size was measured to be 41 µm and 38 µm in the x and y direction, respectively. When the aperture was introduced, the FWHM spot size slightly increased to 45 µm and 46 µm in the x and y direction, respectively. However, this modest increase in spot size was accompanied by a more uniform and symmetric focal profile, highlighting the aperture's beneficial effect on focus quality.

5.2.2. Targetry

The target used in the experiment was a 7-mm-long axisymmetric supersonic de Laval nozzle, commonly referred to in this work as a gas jet, which was filled with either pure hydrogen or a mixture containing 4% nitrogen dopant. This gas jet was mounted on a hexapod stage, enabling precise positioning in three dimensions relative to the laser focus. To introduce a shock within the supersonic gas flow, a silicon wafer was positioned atop the nozzle and mounted on a motorized translation stage with degrees of freedom in both the x and z directions. The insertion of this wafer into the gas flow creates a shock front, which plays a critical role in the electron injection process.

Translating the gas jet along the laser propagation axis (z) enables control over the peak intensity of the laser at the point of interaction with the gas. Movement in the vertical direction (x) adjusts both the plateau plasma density and the local density profile near the shock, which strongly influences the injection mechanism and thereby the characteristics of the resulting electron beam. The position of the silicon blade along the gas jet determines the injection point, effectively tuning the acceleration length available to the electrons. Moving the blade further into the gas jet, the acceleration length is shortened; retracting it increases the length, thereby providing control over the final electron energy. However, the change in the electron energy by changing the acceleration length can influence other parameters of the beam. Hence, there is a need to establish a method for changing the energy while keeping other beam properties constant, and is later demonstrated using the surrogate model. The blade can also be vertically translated to modify the shape of the shock front, a technique shown to impact beam quality and stability [173]. The Hexapod (PI H-824.V) is used to control the position of the gas jet and the blade both while the translation stages (Newport NSA12) only affect the blade positions. These motors and the hexapod have minimal backlash (on the order of a micrometer or less), which is insignificant compared to the step sizes that influence accelerator performance in this work.

Variable input parameters			
		min. value	*max. value*
Gas Jet Longitudinal	z_{jet}	0 mm	9 mm
Gas Jet Transverse	x_{jet}	0 mm	4 mm
Blade Longitudinal	z_{blade}	0 mm	2.5 mm
Blade Transverse	x_{blade}	0 mm	1.5 mm
Backing Pressure Gas Jet	p_{gas}	2 bar	8 bar
Second Order Dispersion	$\beta^{(2)}$	$\beta^{(2)}_{short}$ - 150 fs^2	$\beta^{(2)}_{short}$ + 150 fs^2
Third Order Disperion	$\beta^{(3)}$	$\beta^{(3)}_{short}$ - 2500 fs^3	$\beta^{(3)}_{short}$ + 2500 fs^3
Fourth Order Dispersion	$\beta^{(4)}$	$\beta^{(4)}_{short}$ - 50 000 fs^4	$\beta^{(4)}_{short}$ + 50 000 fs^4
Number of Shots	n_{shots}	3	15
Fixed input parameters			
Laser wavelength	λ_0	800 nm	
Laser power	P	200 TW	
Laser waist (FWHM)	w_0^{FWHM}	46 µm	
Laser duration (FWHM)	Δt	30 fs	

Table 5.1.: **Experimental scan, fixed and dependent parameters.** The top section shows the different input parameters that were controlled and scanned during the experimental optimization run. The bottom section shows some of the fixed laser parameters that could not be controlled and are a function of the laser facility being used.

The plasma density is further tunable via the backing pressure applied to the nozzle, while the temporal characteristics of the laser pulse are controlled through an acousto-optic programmable dispersive filter DAZZLER, which adjusts the second-, third-, and fourth-order dispersion terms. Together, these experimental controls define an eight-dimensional input parameter space optimized using Bayesian optimization. All of the different parameters along with the ranges in which they were moved are summarized in Table 5.1.

The entire system is integrated into the Tango Controls framework, allowing for automated and remote adjustment of all relevant parameters, including the gas jet longitudinal and transverse positions, blade position and height, backing pressure, and the three dispersion orders of the laser pulse [174, 175]. This level of control enables precise, real-time exploration of the complex parameter space governing laser wakefield acceleration.

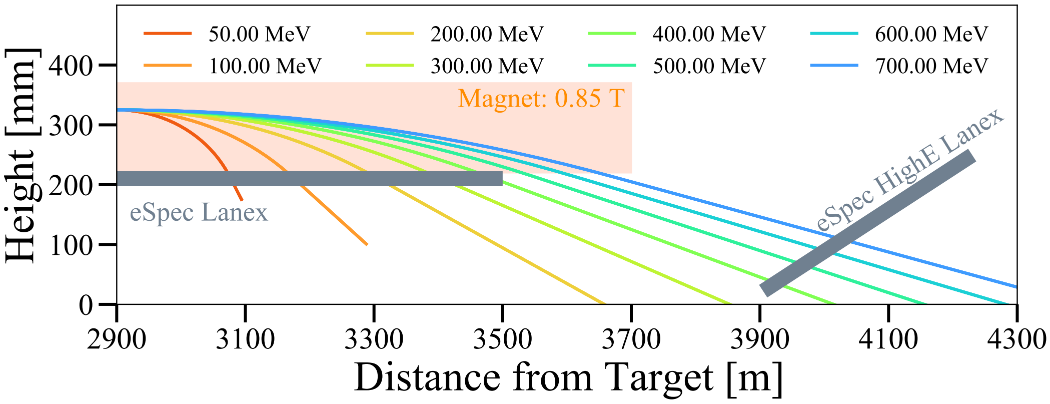

Figure 5.5.: This is a basic schematic of the electron spectrometer that was used in this work. The trajectories of different monoenergetic electron beams are shown. Three cameras view the two lanex screens shown in this figure and the images from all three are calibrated and stitched together to give the final complete spectra.

5.2.3. Electron spectrometer

After the target area, a magnetic dipole spectrometer situated 2.9 m downstream is used as the main diagnostic tool to ascertain the charge, energy, bandwidth and the divergence of the electron beam as shown in Figure 5.5. A 80 cm long permanent magnet with a magnetic field strength of 0.85 T is used to deflect electron beams entering it at a nominal height of 325 mm. The electron trajectories are governed by the Larmor radius r_L given by

$$r_L = \frac{\gamma m_e \nu_e}{eB_\perp} \approx \frac{\gamma m_e c}{eB_\perp} \tag{5.1}$$

where γ is the Lorentz factor, $B_\perp$ is the bending magnetic field and the ν_e is the velocity of the electrons. Based on this relation, electrons with different energies follow distinct curved paths and are spatially separated according to their momenta. Their impact positions are recorded using scintillating screens (lanex) placed at specific locations along the deflection axis. In this work, two lanex screens were employed: one placed directly beneath the magnet to capture lower-energy electrons, and another positioned downstream to detect higher-energy electrons. When electrons strike the screens, they emit visible fluorescence light, which is imaged using 12-bit CMOS cameras. This arrangement allows for the reconstruction of the full energy spectrum of the electron beam across a wide energy range.

One significant source of measurement uncertainty arises from fluctuations in the vertical pointing of the electron beam. A vertical deflection of the beam prior to entering the magnet can lead to a systematic shift in the detected energy. This effect can be mitigated by introducing a movable pointing screen to monitor the electron beam trajectory before entering the magnet. For the electron beams and the spectrometer used in this work, a change of ±1 mrad in the electron pointing resulted in an error of about $2-3\%$ [176] in the energy ranges considered in this work. Since this error is energy-dependent and increases with electron energy, the effect of the pointing variation needs to be taken into account for accurate electron energy calculations from the spectrometer at higher energies [177]. However, since the primary objective of this work was to demonstrate the effectiveness of Bayesian optimization, and the pointing-related errors were within acceptable limits for the energy range studied, the pointing screen was omitted to simplify the experimental setup.

5.2.4. Post-processing spectra

Once an electron spectrum was obtained from the individual images of the lanex screens, it was passed through a dedicated data post-processing pipeline before being fed to the Bayesian optimizer. During this step, key beam output parameters such as the total beam charge Q, median energy $\overline{E}$, peak energy E_{peak} and energy spread σ_E are extracted from the spectrum. While these quantities are sufficient to characterize an idealized, normally distributed energy spectrum, real laser-accelerated electron beams often exhibit strong deviations from normality. In particular, the actual energy distributions are frequently multimodal or skewed, which reduces the accuracy and interpretability of aggregate metrics such as the mean or standard deviation. For instance, in gas mixture targets, shock-assisted ionization injection [178] commonly leads to spectra with multiple distinct electron bunches, often featuring low-energy tails and secondary peaks. In such cases, computing global statistical metrics over the entire spectrum can obscure the characteristics of the most relevant bunch and mislead the optimizer.

To address this challenge, a Gaussian mixture model (GMM) approach [179] was employed to isolate the contributions of individual electron bunches in real time. This method enables more accurate estimation of metrics like the median, mean, and energy spread in the presence of complex or multi-peaked distributions. Figure 5.6 illustrates a representative example of such a multimodal spectrum, generated from a LWFA experiment using a gas mixture. Here, shock-assisted ionization injection produces an unstable combination of low- and high-energy features. While it would appear to the optimizer that this particular position in the input space x_i produced stable electron beams at 350 MeV, the spectrum among different shots is quite unstable, with most charge contained in a beam fluctuating in energy between 200 MeV and 350 MeV. The GMM-based decomposition enables a robust and efficient separation of these spectral components. For training the GMM, 1000 samples were drawn from a probability distribution derived from the raw spectrum.

Optimization parameters		
Number of shots per position	n_{shots}	3-15 (fixed 10 for 4D)
Max. number of Iterations	n_{iter}	User-Defined
Number of initial points	n_{INIT}	5-10
Input Dimensions	dim_x	4-8
Output Dimensions	dim_y	3-4 (multi-objective)
Cost Function	$cost(n_{shots})$	$\frac{n_{shots}}{rep.rate} + fixed.cost$

Table 5.2.: **Summary of optimization parameters used for experiments** Some of the key parameters used during the experimental optimization are summarized here. In this instance, the optimization process was terminated by the user owing to limited experimental beam time. In the case of 4D optimization number of shots was kept fixed to 10 while for the 8D optimization the optimizer adaptively selected between 3 and 15 shots. The fixed cost in the cost function is simply the time taken by the slowest motor to reach a set value.

The number and position of local intensity peaks within a 100 MeV window were used to initialize the model. Subsequently, the Expectation-Maximization (EM) algorithm [180] was employed to fit a multi-component Gaussian distribution to the data. The EM algorithm iteratively estimates the likelihood that each portion of the spectrum belongs to a particular Gaussian mode, effectively partitioning the total charge across distinct energy bands. The effect of this filtering procedure is demonstrated in the lower sub-figures of Figure 5.6, where both low- and high-energy features are clearly separated and can be selectively retained or excluded.

Following this decomposition, physical metrics are recalculated for the targeted electron bunch, enabling a more precise and meaningful evaluation of beam quality. This selective post-processing step not only prevents contamination of objective metrics by irrelevant features but also enhances the reliability of the feedback provided to the Bayesian optimizer. As shown in Figure 5.6, the resulting GMM-segmented spectra allow for cleaner identification of performance trends and support more accurate analysis for the Bayesian optimizer.

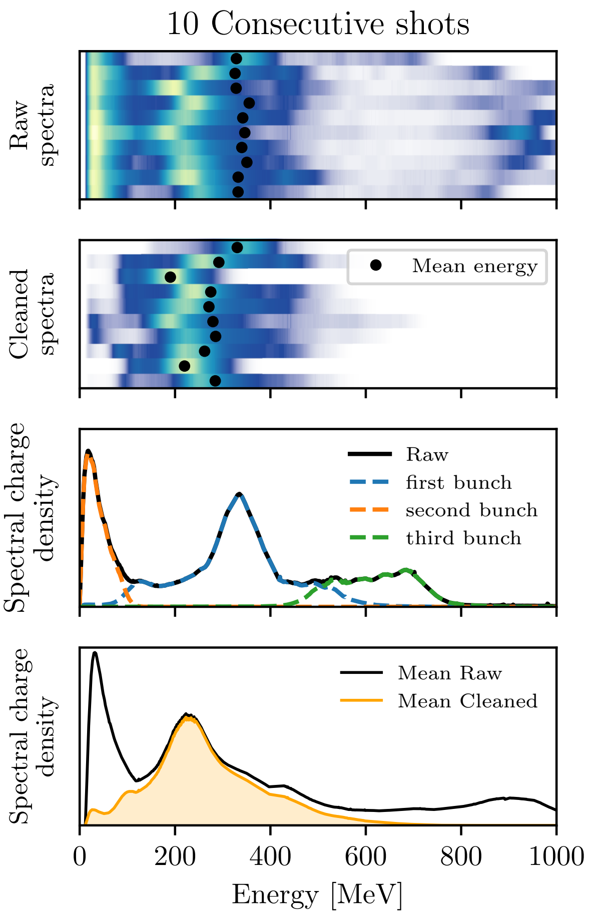

Figure 5.6.: Reconstructed raw electron spectra from the lanex images (top) and spectra after 'cleaning' using a Gaussian mixture model. The third plot shows the last spectrum among the 10 shots as an example, demonstrating the different electron bunches that the Gaussian mixture model predicts from the raw black spectrum. The orange shaded line in the bottom plot shows the average cleaned spectrum versus the average original black spectrum.

5.3. Multi-objective optimization of LWFA

This section presents the results of various multi-objective optimization strategies applied to experimental laser wakefield acceleration. It begins with a brief overview of the general optimization parameters and procedures used across multiple experimental campaigns, followed by a discussion of specific results organized by optimization type. In the final part of the section, I will demonstrate how results from a multi-objective optimization can be repurposed for single-objective exploitation, allowing targeted refinement within specific regions of the input space. These results serve as the basis for energy tuning of LWFA electron beams, which is one of the key outcomes of this work.

A summary of the experimental optimization parameters is provided in Table 5.2.

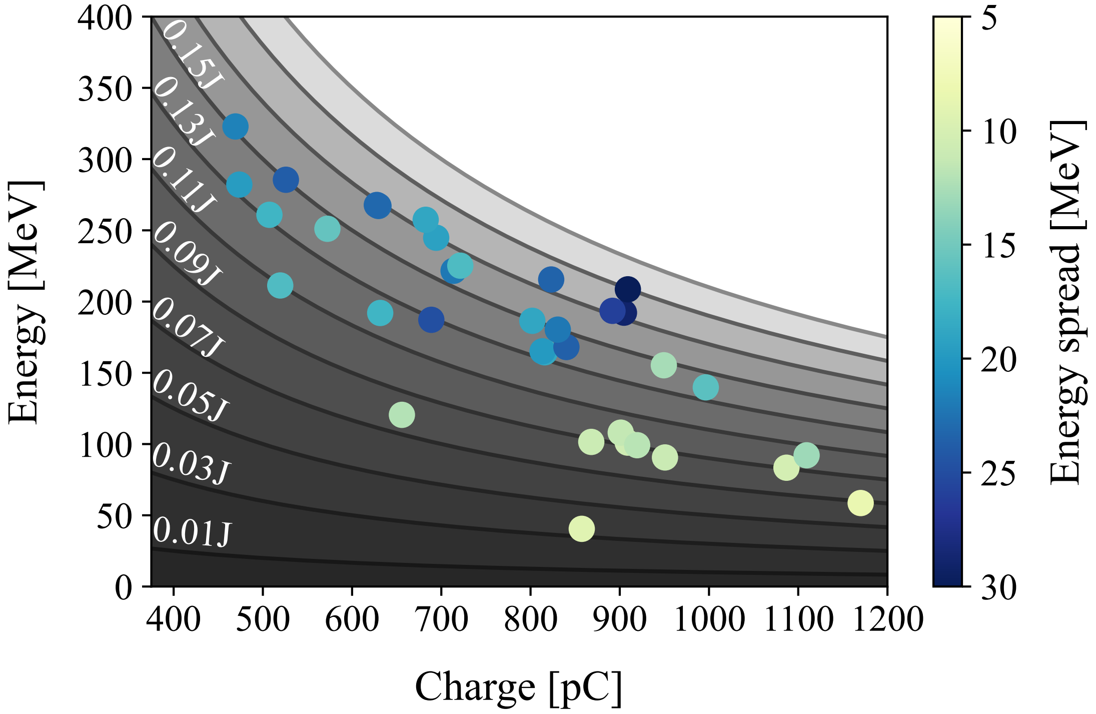

Figure 5.7.: Pareto-optimal configurations of the laser wakefield accelerator obtained from 4D multi-objective optimization of mean energy, charge and energy spread. The absolute energy spread is indicated by color, while grey shaded regions denote areas of equal total beam energy, corresponding to constant accelerator efficiency. Although the Pareto front lies on a 3D surface, only its 2D projection onto the charge–mean energy plane is shown.

At each setting $\vec{x}$ in the input space $\mathcal{X}$, the optimizer collects n shots, denoted as $(\vec{y}_1(\vec{x}), \ldots, \vec{y}_n(\vec{x}))$, and computes the average response $\vec{\mu}(\vec{x}) = n^{-1} \sum_{i=1}^{n} \vec{y}_i(\vec{x})$. This averaging step is critical for mitigating the impact of shot-to-shot fluctuations, which are inherent in high-power laser-plasma experiments. In the first optimization run, which involved only multi-objective Bayesian optimization (MO), the number of shots per input parameter setting was fixed at 10. In all subsequent experiments, trust-MOMF was employed, hence, the number of shots was adaptively chosen by the optimizer within a predefined range.

At each evaluated input setting, multiple beam parameters such as total charge, energy spread, and mean or median energy were extracted, making each measurement inherently multi-dimensional. The results that follow provide a detailed account of how these vector-valued outputs were optimized to explore and exploit the complex, high-dimensional performance landscape of LWFA.

5.3.1. 4D optimization

As an initial test, a multi-objective optimization was conducted in a reduced 4-dimensional input space. The optimization comprised a total of 60 evaluated configurations $\vec{x}$, including an initial set of 5 randomly selected points, which are necessary to initialize the Bayesian optimization process. Upon completion of the run, 35 configurations were identified as Pareto-optimal, and are visualized in Figure 5.7. The product of total beam charge and mean energy yields the total energy of the beam and can therefore be used to measure the energy transfer efficiency from the laser to the electron bunch.

This initial experiment clearly recovered the well-known trade-off between the beam charge and energy. As shown in Figure 5.7, most of the Pareto-optimal solutions cluster around contours of similar energy transfer efficiency ($\sim$ 4%). Two different regions of operation can be identified by looking at the beams on both sides of 900 pC. Below this threshold, higher-charge solutions tend to exhibit slightly greater efficiency, likely due to increased beam loading, which enhances energy extraction from the wakefield. However, as the charge approaches 1 nC, a substantial drop in efficiency is observed, suggesting a transition into a different acceleration regime. Interestingly, this high-charge regime is also where electron beams with the lowest absolute energy spread are found. A plausible physical interpretation is that energy spread accumulates over longer acceleration distances. Thus, minimizing energy spread under these experimental conditions required a shorter acceleration length, which in turn led to a lower mean energy $\overline{E}(\vec{x})$. This outcome illustrates the optimizer's ability to identify and exploit complex trade-offs in beam quality metrics across multiple physical regimes.

5.3.2. 8D optimization

After demonstrating the usefulness of Bayesian multi-objective optimization for experimental LWFA, the input parameter space was expanded from 4D to 8D by including gas pressure and the three orders of laser dispersion terms. Alongside a higher input dimensionality, the optimization strategy was further enhanced by introducing a multi-fidelity formulation, allowing the optimizer to balance precision and time cost more effectively. In the context of experimental LWFA, fidelity was interpreted as the number of laser shots taken at a given input configuration. This approach is based on the assumption that averaging over a greater number of shots provides a more reliable estimate of the beam metrics, albeit at the expense of increased time per evaluation. Since experimental beam time is inherently limited, there is a trade-off between data quality and the number of unique configurations that can be explored. To manage this trade-off, the Trust-MOMF algorithm [139] was employed, which has previously demonstrated accelerated convergence in numerical simulations (see Chapter 4).

In this implementation, the trust objective was defined as $\sqrt(n)$, a monotonically increasing function of the number of shots n, capturing the idea that additional measurements increase confidence in the result. The cost model combined three components: the number of shots, the laser repetition rate, and a fixed cost associated with hardware actuation. The latter was estimated empirically as the time required for the slowest motor in the system to traverse its full range between two consecutive input configurations. For this experimental run, the minimum and maximum number of shots per input setting were user-defined and set to 3 and 15, respectively, allowing the optimizer to dynamically allocate measurement resources based on the expected increase in utility and the total cost.

In this 8D optimization run, a total of 95 parameter configurations were evaluated, including 10 randomly sampled initial points used to initialize the Bayesian optimizer. This increased number of evaluations was necessary due to the larger dimensionality of the 8D input space compared to the earlier 4D case. The resulting Pareto front consisted of 35 non-dominated configurations, shown in Figure 5.8. When compared with the 4D optimization results presented in Figure 5.7, a clear improvement in laser-to-beam energy transfer efficiency is observed. On average, the efficiency increased from approximately 4% to 6.5%, with a maximum value of 8.25% achieved in one configuration.

As in the 4D case, a trade-off between mean energy and energy spread, where low-bandwidth beams tend to be produced at lower beam energies. Additionally, in the case of LWFA at CALA, the efficiency must be decreased with an overall lower injected charge to attain beams with low energy bandwidth. This highlights the subtle interplay between charge, energy, and spectral quality in laser-plasma acceleration.

The four right-hand panels of Figure 5.8 display representative spectra corresponding to four selected Pareto-optimal configurations, illustrating the diversity of beam profiles produced by a single multi-objective optimization. These results demonstrate a practical advantage of this approach: users can select solutions *a posteriori* based on specific application requirements, such as beam energy or spectral width, without needing to re-run the optimization. Additionally, the error bars in the left panel of Figure 5.8 represent the shot-to-shot standard deviations in beam charge and energy, providing a measure of experimental stability. These fluctuations motivated subsequent optimization tasks that explicitly included stability as an optimization objective.

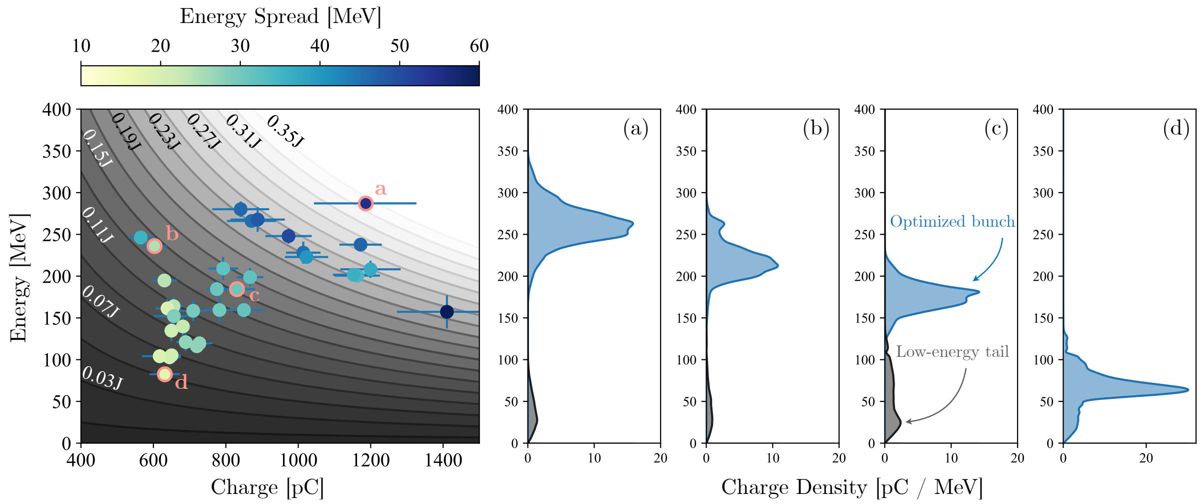

Figure 5.8.: Pareto-optimal configurations of the laser wakefield accelerator obtained from an 8-dimensional multi-objective optimization with beam energy, charge, and energy spread as outputs. The absolute energy spread is color-coded, while error bars indicate the standard deviation of charge and energy due to shot-to-shot fluctuations. Grey shaded regions represent contours of constant beam energy, corresponding to equal accelerator efficiency. Compared to the 4D case, the 8D optimization yielded higher efficiencies by exploiting additional degrees of freedom—namely gas pressure and laser dispersion parameters. Panels (a–d) show representative electron spectra corresponding to marked points on the Pareto front, each exhibiting beam properties close to the respective local average. Greyed-out regions in the spectra denote low-energy tails, identified and removed using the Gaussian Mixture Model described earlier.

5.3.3. 8D optimization with target energy

In addition to optimizing standard beam quality metrics such as charge and energy spread, multi-objective optimization can also be extended to target a specific desired beam energy. To achieve this, an additional objective defined as the distance to a target energy $\Delta E(\vec{x}) = |E_0 - \overline{E}(\vec{x})|$ was introduced, which is minimized the closer the expectation value of beam energy is to the target energy. Moreover, as noted at the end of the previous section, shot-to-shot fluctuations in beam properties can be explicitly addressed within the optimization framework by introducing an objective function that rewards input configurations yielding more stable electron beams. In this work, this was achieved through the definition of an instability parameter, which quantifies relative fluctuations in key beam metrics and serves as an additional objective in the multi-objective optimization. The instability parameter $S(\vec{x})$ is computed as the root sum of squares of the relative standard deviations for the total charge Q, energy spread σ_E and deviation from the target energy ΔE :

$$S(\vec{x}) = \sqrt{\left(\frac{\sigma_Q(\vec{x})}{\mu_Q(\vec{x})}\right)^2 + \left(\frac{\sigma_{\sigma_E}(\vec{x})}{\mu_{\sigma E}(\vec{x})}\right)^2 + \left(\frac{\sigma_{\Delta E}(\vec{x})}{\mu_{\Delta E}(\vec{x})}\right)^2} \,. \tag{5.2}$$

This formulation can be interpreted as a measure of aggregate relative variability across the three beam quality metrics. Minimizing $S(\vec{x})$ effectively reduces the joint uncertainty in charge, energy spread, and distance to target energy, thereby favoring input conditions that produce more consistent output. However, it is important to acknowledge that this definition has limitations. One known issue is that relative standard deviations tend to increase as absolute values of parameters such as energy spread or distance to the target energy become small. This can result in disproportionately large contributions to the instability parameter in otherwise favorable operating regimes. While this definition provides a practical and interpretable starting point for incorporating stability into the optimization process, future work could benefit from more robust formulations that better account for scale-dependent behavior in beam metrics.

The full objective vector is then given by

$$\vec{y}(\vec{x}) = (Q(\vec{x}), \sigma_E(\vec{x}), \Delta E(\vec{x}), S(\vec{x})).$$

with the optimizer aiming to maximize Q and minimize the remaining three quantities. The number of shots per input configuration was chosen dynamically via multi-fidelity optimization, allowing for improved efficiency by allocating more measurements to regions of interest while conserving beam time.

This optimization run consisted of 65 evaluated input positions, including 5 random initial points, and yielded a Pareto front composed of 20 non-dominated configurations. Due to the complexity of visualizing four-dimensional objective spaces, a three-dimensional projection is shown in Figure 5.9, illustrating trade-offs between total charge, energy spread, and instability.

The results confirm a persistent trade-off between beam charge and energy spread, consistent with well-documented beamloading effects in LWFA [163, 181, 182]. In addition, the optimization reveals a secondary trade-off between performance and stability: more stable beams tend to have lower charge or broader spectra, whereas higher-performance beams exhibit increased shot-to-shot variability. These findings highlight the necessity of explicitly including stability and target energy as optimization objectives when precision and reproducibility are essential.

5.3.4. Exploitation phase

After a successful multi-objective optimization run that yielded Pareto optimal points, the model can be used to exploit a certain region of a Pareto front to obtain a desired solution. While multi-objective optimization is ideal for exploring trade-offs and identifying Pareto-optimal configurations, there are situations where a

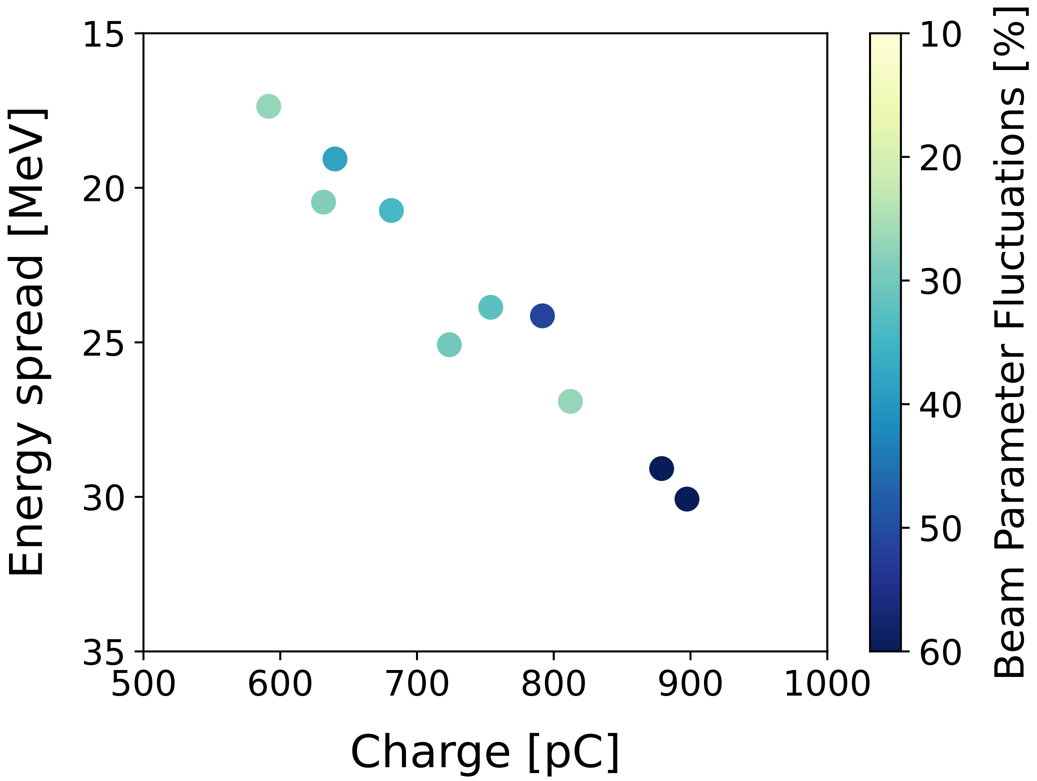

Figure 5.9.: Pareto-optimization of a laser wakefield accelerator targeting a beam energy of 250 MeV. A clear trade-off is observed between energy spread and total charge, consistent with beamloading effects in the wakefield. Additionally, the results reveal a compromise between high-performance configurations with greater shot-to-shot variability and lower-performance configurations exhibiting improved stability.

single, refined solution is desired, particularly if none of the Pareto-optimal points fully satisfy a user's specific performance criteria. In such cases, the Pareto front serves as a rich database of solutions that can be selectively refined through single-objective optimization. To perform this exploitation phase, a custom scalarized objective function is defined as

$$O_{\text{sing}}(\vec{x}) = Q(\vec{x}) + a \cdot \sigma_E(\vec{x}) + b \cdot \Delta E(\vec{x}), \tag{5.3}$$

where a and b are the scalar weights applied to each objective. Importantly, a is constrained to be a negative real number, while b is restricted to be a positive real number. This choice of positive or negative weights depends on whether the objective is maximized or minimized. This formulation allows us to convert a multi-objective landscape into a tailored single-objective function that emphasizes user-defined priorities.

The process begins by selecting a specific point $\vec{p} \in \mathcal{P} \subset \mathcal{Y}$ on the Pareto front that the user wishes to locally exploit. The aim is to determine the values of

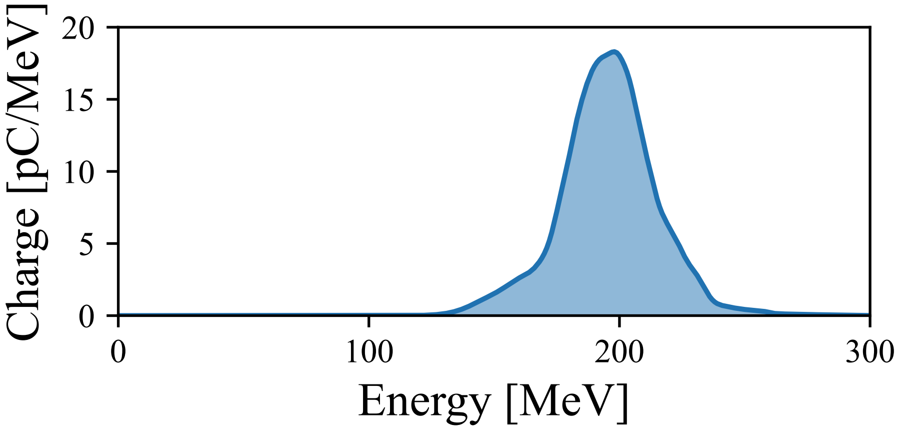

Figure 5.10.: An example spectrum obtained during the exploitation phase following multi-objective optimization. This run was initialized by selecting a Pareto-optimal point near a median energy of 200 MeV. The resulting shot exhibits a median energy of 208 MeV with a relative energy spread of 5% defined by the median energy and the median absolute deviation.

a, b that make this point optimal for the scalarized objective. This problem can be reformulated into a loss function which, when optimized, yields the suitable hyper-parameters.The loss function was defined as

$$\ell(\vec{p}, \vec{p}') = \max_{\vec{p}' \in \mathcal{P}} \{O_{\text{sing}}(\vec{p}')\} - O_{\text{sing}}(\vec{p}), \tag{5.4}$$

which measures the discrepancy between the scalar objective at the chosen point and the maximum attainable value across the entire Pareto front. By minimizing this loss function, the values of a and b that ensure $\vec{p}$ is the most favorable solution according to the constructed single-objective criterion.

Once the hyperparameters are identified, they are used to define a custom single-objective function, as in Equation (5.3), and initiate a new round of Bayesian single-objective optimization. For this, a lower confidence bound (LCB) acquisition function [166] was used, which emphasizes exploitation by converging toward local optima near the region of interest on the Pareto front.

The result of this focused optimization is shown in Figure 5.10. It yields a high-quality electron beam with a charge density of up to 18 pC/MeV at a median energy of 208 MeV, and a relative energy spread of only 5% defined by the median energy and the median absolute deviation. This demonstrates the effectiveness of combining multi-objective exploration with targeted single-objective refinement to achieve application-specific performance in laser wakefield acceleration.

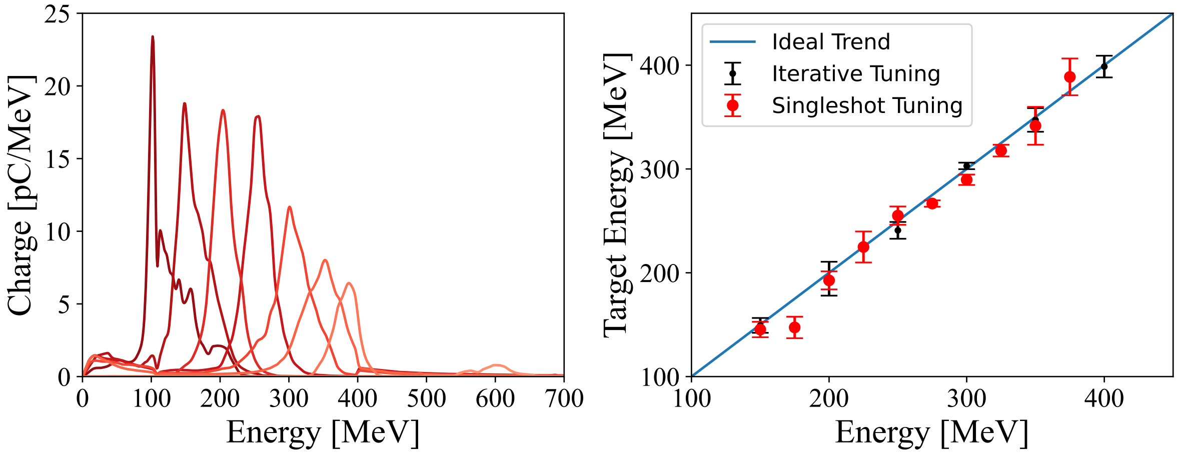

Figure 5.11.: Energy tuning of the LWFA. Left: Example electron spectra obtained using the iterative tuning strategy, producing beams with mean energies ranging from 100 MeV to 400 MeV. An additional spectrum at 600 MeV illustrates the optimizer's attempt to extrapolate beyond its training range. Right: Comparison of experimental tuning results with the ideal target energy (blue line). Iterative tuning points are shown in black, while single-shot tuning points are shown in red with the error bars indicating the standard error of the mean. The results demonstrate that the optimizer reliably tunes the electron beam energy across the range of 100 MeV to 400 MeV.

5.4. Tuning of a LWFA

Once the Pareto front is established to a sufficient degree, the trained model can be leveraged for its predictive capabilities to cater to changing user preferences dynamically without running new optimizations from the start. In contrast to initiating a single-objective optimization from scratch, this strategy is both more efficient and more flexible, as it is informed by a multi-objective landscape that avoids bias toward any particular solution. By clearly decoupling the phases of multi-objective exploration and single-objective exploitation, a versatile methodology was established that is capable of meeting diverse experimental goals such as energy tuning or targeting specific beam parameters. Specifically, different solutions can be selected via *a posteriori* scalarization of the objectives, guided by user-defined performance criteria. Again, acquisition functions using confidence bounds (UCB in this case) were used, which consider both the probability of yielding beams at a particular target energy and the uncertainty of the prediction. This acquisition function is minimized using a gradient-based optimizer [183], yielding an updated 8-dimensional configuration that is predicted to produce electron beams near a user-defined energy target.

To demonstrate the effectiveness of this approach, an iterative tuning was performed by acquiring 10 shots at each recommended configuration, appending the results to the GP training data, and repeating this procedure three times for each target energy. In parallel, a single-shot tuning strategy was also evaluated in which no new data was incorporated after the initial prediction. The results are shown in Figure 5.11, where iterative (and single-shot) scans were performed for target energies ranging from 150 MeV to 400 MeV, in steps of 50 MeV (and 25 MeV for single-shot). The tuning results demonstrate that our model-based approach enables accurate and continuous energy control over an octave in energy, spanning from 150 to 400 MeV. Also shown on the left panel of Figure 5.11 are example spectra from the iterative tuning, where the energy was tuned from 100 MeV to 400 MeV.

The success of the single-shot tuning highlights the efficiency of the model in enabling low-latency accelerator control, while the iterative tuning demonstrates its adaptability to evolving system conditions, such as long-term drifts or environmental changes. Notably, the optimizer navigates an 8-dimensional parameter space—beyond what could reasonably be explored by even the most experienced human operators.

To assess the generalizability of the GP model, it was asked to predict a configuration for a target energy of 600 MeV, well beyond its original training range centered around 250 MeV. Remarkably, although the model was not explicitly optimized for this regime, it successfully produced several shots with energies approaching the 600 MeV target as shown in the left panel of Figure 5.11. On average, however, the resulting beam energy was around 470 MeV. This outcome underscores both the predictive power and limitations of the model. While extrapolation beyond the training distribution remains challenging, significant shifts in energy output can still be achieved using the existing multi-objective optimization framework.

Beyond achieving the desired energy, the GP model can also be used to provide physical insights into how different experimental parameters were adjusted to reach the desired energies, as illustrated in Figure 5.12. Several systematic trends can be observed. For instance, the gas pressure and consequently the plasma electron density is increased steadily to reach higher target energies. This higher plasma electron density enhances the axial fields (see Section 2.3.2), resulting in the electrons gaining higher energy in the same length[1]. The subtle movement of the other parameters is harder to explain because of the complex nature of the physics involved. However, these patterns underline the fact that the model is not merely interpolating between known configurations, but actively navigating a complex, high-dimensional control space to achieve the desired goal.

A more general and intuitive approach to tuning beam energy or other performance metrics can be achieved by inverting the mapping between the input and

[1] In this instance, the dephasing length is estimated to be larger than the length of the gas jet; otherwise, increasing the gas pressure would result in lower energy beams.

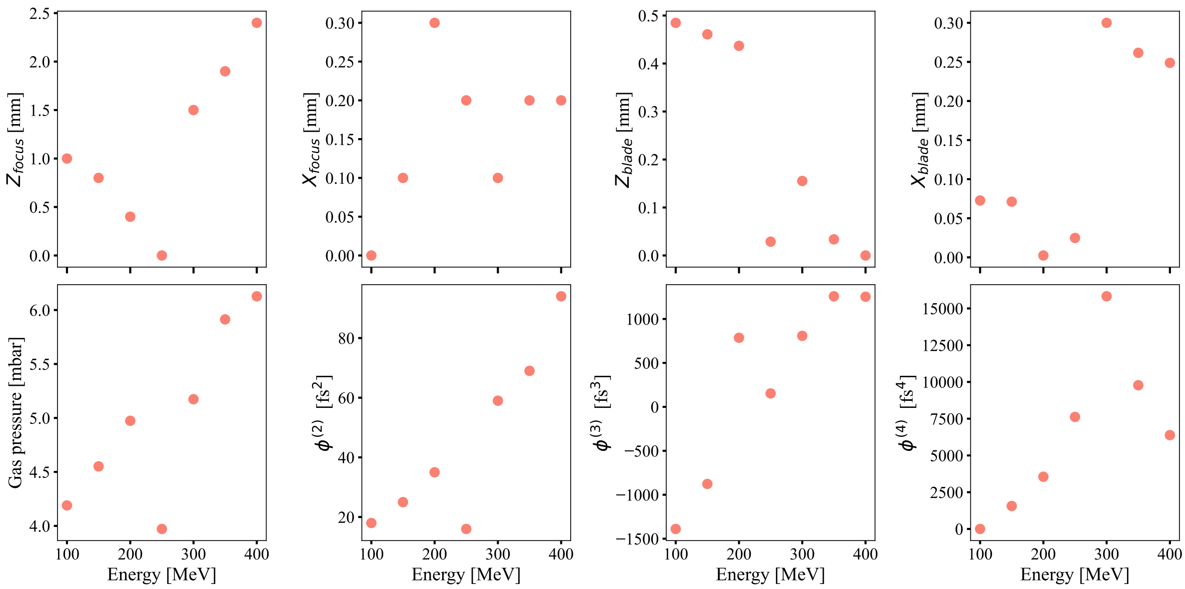

Figure 5.12.: Energy tuning inputs. Shown are the input configurations proposed by the GP model for tuning the median electron energy across the range 100–400 MeV. Each subplot corresponds to one of the eight experimental control parameters. A positive displacement of Z_{blade} corresponds to the silicon blade moving further into the gas jet. An increase in the x dimension of either the gas jet or the blade brings them closer to the laser axis. The zero points of the different orders of dispersion represent the values that resulted in the shortest possible pulse.

output domains. In this framework, the desired beam properties—such as total charge, energy, and energy spread—are treated as inputs, while the experimental control parameters—the eight tunable degrees of freedom in the system—are treated as outputs. A model is then trained to learn the inverse relationship from the objective space $\mathcal{Y}$ to the input parameter space $\mathcal{X}$ which is termed the inverse model.

This inversion offers a more natural and user-friendly perspective for experimental control. Instead of exploring the input space to discover acceptable outcomes, the user can directly specify a desired set of beam parameters, and the inverse model will return one or more configurations of control parameters likely to yield the requested performance. This approach not only simplifies the tuning process but also allows for rapid adaptation to new experimental goals without requiring full re-optimization.

This inverse modeling approach represents a promising direction for future work and is currently the focus of an ongoing research project. While the present study demonstrates forward modeling and optimization using Gaussian processes, extending the framework to learn the inverse mapping opens the door to a more interactive and goal-oriented control strategy. By enabling users to query desired beam outcomes and receive corresponding experimental configurations in real time, this method has the potential to significantly enhance both usability and efficiency in laser-plasma accelerator operation.

6. Summary and outlook

Particle accelerators have long been at the heart of scientific and technological progress, driving discoveries across high-energy physics, materials science, medicine, and beyond. Yet despite their transformative impact, conventional radio-frequency accelerators face fundamental limits to the achievable accelerating gradients imposed by vacuum breakdown, resulting in facilities that are often prohibitively large and expensive. LWFA has emerged as a promising candidate for overcoming these constraints, offering orders of magnitude higher field gradients and the potential for compact, cost-effective accelerators. However, while proof-of-principle experiments have demonstrated GeV-class acceleration in centimeter-scale plasmas, realizing the full potential of LWFA in applied settings has been hindered by challenges in stability, reproducibility, and controllability. This dissertation was motivated by the observation that conventional heuristic tuning strategies are increasingly inadequate for such complex systems, and demonstrates how machine learning approaches, specifically Bayesian optimization, can be leveraged to systematically and efficiently improve LWFA performance.

The first major contribution of this work is the development of the Trust-MOMF optimization method. Prior research applying Bayesian optimization to accelerator control had primarily focused on single-objective formulations, which are often too restrictive to capture the real-world trade-offs between competing performance metrics, such as beam energy, charge, and energy spread. The Trust-MOMF framework unifies multi-objective and multi-fidelity Bayesian optimization in a single coherent approach that enables efficient exploration of the Pareto front and intelligent allocation of limited experimental resources across different fidelity levels. This is especially important in LWFA, where measurement fidelity can be traded off against cost by averaging over different numbers of shots.

In controlled numerical experiments, this framework was shown to significantly outperform conventional optimization strategies. When applied to synthetic test functions with cost ratios exceeding two orders of magnitude between fidelity levels, Trust-MOMF consistently converged to higher-quality solutions at substantially reduced computational expense (see Section 3.5.1). In practice, we observed over an order of magnitude reduction in the effective cost to achieve comparable hypervolume performance compared to single-fidelity strategies (see Section 4.4). These results underline the generality and scalability of Trust-MOMF and suggest its applicability to a wide range of physics and engineering optimization problems.

The second major result of this work was applying Trust-MOMF to the optimization of LWFA electron beams in variable-fidelity numerical simulations using the FBPIC code. This implementation allowed for systematic exploration of the multidimensional accelerator parameter space, yielding detailed Pareto surfaces that map out the trade-offs between beam charge, mean energy, and bandwidth while also revealing important physical insights. For example, it was observed that maximum efficiency was consistently associated with operation at higher charge and higher energy spread, while producing more monoenergetic beams required a deliberate reduction of injected charge. Critically, the Gaussian process surrogate models produced during optimization also enabled post-hoc interrogation of the system, allowing inverse predictions of input settings that are likely to yield target beam characteristics. This capability is particularly relevant for applications requiring fine control, such as tuning the peak energy without substantially altering other beam properties, or for supporting user-driven operation where target specifications may change dynamically.

The experimental realization of multi-objective multi-fidelity Bayesian optimization constitutes perhaps the most significant and novel aspect of this dissertation. While previous demonstrations had shown that BO could optimize scalar objectives such as total beam charge in LWFA, no prior study had integrated multi-objective and multi-fidelity optimization into a live experimental accelerator campaign using the ATLAS-3000 laser system. In these experiments, performed using the ATLAS-3000 laser system, the number of shots per parameter configuration was treated as an adjustable fidelity variable, reflecting the improved statistical confidence from averaging more measurements. The optimizer dynamically allocated more shots to promising regions, achieving faster convergence while making efficient use of limited beam time. In practice, this strategy enabled rapid identification of Pareto-optimal solutions trading off charge, energy, and stability, with cost reductions and convergence rates surpassing simpler Bayesian optimization (BO) baselines.

A particularly noteworthy outcome of this experimental work is the demonstration of precise energy tuning of the electron beam from approximately 150 MeV to 400 MeV. By inverting the forward model constructed during the multi-objective optimization, it was shown that it is possible to propose new parameter configurations achieving user-specified target energies while preserving other beam properties within acceptable bounds. Furthermore, this approach to energy tuning represents a departure from previous methods, as it achieved target energies in at most three iterations, with many instances reaching the desired values on the initial attempt. This capability is highly relevant for user-driven facilities, where experimental demands can shift from day to day and operators must be able to reconfigure accelerator performance without resorting to manual trial-and-error tuning.

Beyond these primary contributions, this work carries broader implications for the role of data-driven methods in experimental science. As laser-plasma accelerators mature, the parameter spaces involved in their operation will only grow more complex, and the expectations for reproducibility and flexibility will become more stringent. The success of Trust-MOMF demonstrates that Bayesian optimization is not merely a theoretical construct but also a practical tool that can be deployed in real time to improve performance and accelerate discovery. Moreover, the methods demonstrated here can also be adapted to other experimental domains characterized by expensive, noisy, and high-dimensional optimization problems.

Several promising directions for future research emerge from this work. One important limitation of current Bayesian optimization strategies, including Trust-MOMF, lies in their reliance on Gaussian processes as surrogate models. While GPs offer excellent uncertainty quantification in small-data regimes, they scale poorly with large datasets and high-throughput experiments. Addressing this challenge will require exploring alternative surrogate models such as Bayesian neural networks or scalable approximations to GPs that retain uncertainty estimates while improving computational efficiency. Developing techniques making GPs more scalable through sparse approximations also remains an active and promising area of research.

Another direction lies in moving beyond fixed, user-specified objectives. In real-world applications, users may struggle to define the mathematical form of their ideal beam properties precisely. Future optimization frameworks could incorporate interactive learning of objectives that dynamically refine the optimization targets through dialogue or preference elicitation with the experimenter. This would make Bayesian optimization even more flexible and accessible to non-expert users.

A further avenue is the use of warm-starting, where optimization can be accelerated by leveraging prior data from simulations or previous experimental campaigns. For example, historical data collected on similar plasma densities or laser parameters could be used to inform priors and thereby reduce the number of shots required to converge in a new session. This integration of offline and online learning has the potential to significantly improve efficiency and consistency across experiments.

In summary, this dissertation has introduced and validated a comprehensive framework for the multi-objective, multi-fidelity Bayesian optimization of laser-wakefield accelerators. Through a combination of theoretical development, extensive numerical benchmarking, and experimental demonstration, it has established that intelligent, data-driven optimization can meaningfully advance the reproducibility, flexibility, and performance of LWFA systems. As the field continues to evolve, these methods and insights provide a foundation for the next generation of compact plasma accelerators and illustrate the broader potential of machine learning to transform experimental science.

A. Test Functions

In this section, the different analytical functions that were used to benchmark the performance of the trust-MOMF technique are described. Since this work was one of the first studies on combined multi-objective and multi-fidelity optimization, many of the test functions from the literature could not be used. They all had to be modified to incorporate an additional fidelity input dimension and exhibit trade-off behavior between the different objectives.

Modified Multi-Fidelity Forrester Function

The original Forrester function is

$$f(x) = (6x-2)^2 \cdot \sin(12x-4) + 7.025$$

and it was modified to a multi-fidelity use case:

$$f(x,s) = D(s) \cdot [E - g(x,s)]$$

where

$$g(x,s) = A(s) \cdot f[x - 0.2(1 - x \cdot s)] + B(s) \cdot (x - 0.5) - C(s)$$

with $A = 0.5 + 0.5s$, $B = 2 - 2s$, $C = 5s - 5$, $D = 1.5 - 0.5s$ and $E = 25$. The most important differences to other multi-fidelity versions of this function are that it contains a fidelity- and position-dependent shift $\tilde{x}(x,s) = x - 0.2(1 - x \cdot s)$, is inverted for maximization ($E - g(x,s)$ term), the value of the maxima along the fidelity is decreasing (D(s) term) and the maxima are continuously connected from the low to high fidelity.

Modified Multi-Fidelity, Multi-objective Branin-Currin Function

Two popular functions that are used for benchmarking optimization techniques are the Branin-Currin functions. The usual form of the Branin function is

$$B(\boldsymbol{x}) = a(x_2 - bx_1^2 + cx_1 - r)^2 + p(1-t)\cos(x_1) + p,$$

where values of the constants were taken to be $a = 1$, $b = 5.1/(4\pi^2)$, $c = 5/\pi$, $r = 6$, $p = 10$, $t = 1/(8\pi)$ and the form of Currin function is

$$C(\boldsymbol{x}) = \left[1 - \exp\left(-\frac{1}{2x_2}\right)\right] \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20}.$$

Both of these functions were modified to restrict the domain and range to a unit hypercube implying $x_i, y_i \in [0, 1]\, \forall i = [1, 2]$. The modified form that was used for the Branin is

$$B(\boldsymbol{x}, s) = -[a(x_{22} - b(s)x_{11}^2 + c(s)x_{11} - r)^2 + p(1 - t(s))\cos(x_{11}) + p]$$

where $x_{11} = 15x_1 - 5$, $x_{22} = 15x_2$, $a = 1$, $b(s) = 5.1/(4\pi^2) - 0.01(1 - s)$, $c(s) = 5/\pi - 0.1(1 - s)$, $r = 6$, $p = 10$, $t(s) = (1/(8\pi)) + 0.05(1 - s)$ was used. The rescaled x_{11} and x_{22} are used to normalize the original domain of the Branin to [0,1]. The main difference is the addition of fidelity parameter s which for a value of 1 yields the original Branin function and results in approximations to the original Branin function for values less than 1. Since we are maximizing the problem the function is also inverted here. The modified form for the Currin function is

$$C(\boldsymbol{x}, s) = -\left[\left[1 - (0.1)(1 - s)\exp\left(-\frac{1}{2x_2}\right)\right] \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20}\right]$$

where again the main difference is the addition of fidelity term $1 - s$ and the inversion necessary for maximization.

Modified Multi-Fidelity Multi-Objective Park Functions

For bechmarking the multi-objective multi-fidelity problem in higher dimensions, multi-fidelity versions of Park 1 and Park 2 functions were used. The original form of the Park functions is

$$P_1(\boldsymbol{x}) = \frac{x_1}{2}\left[\sqrt{1 + (x_2 + x_3^2)\frac{x_4}{x_1^2}} - 1\right] + (x_1 + 3x_4)\exp\left[1 + \sin(x_3)\right]$$

$$P_2(x) = \frac{2}{3}\exp(x_1 + x_2) - x_4\sin(x_3) + x_3$$

As with previous modifications, the fidelity dimension s was also introduced here. To achieve a reasonable Pareto front for optimization, the two functions were also slightly modified. The location of the Pareto set was also modified to not have all the optimizing points in the corners of the 4-D hypercube. A last modification is shifting the Pareto front of the Park functions by subtraction to place a higher importance on the trade-off region. The final form of the two modified Park functions is

$$P_1(\boldsymbol{x}, s) = A(s)\left[T_1 + T_2 - B(s)\right]/22 - 0.8$$

$$T_1 = \left[\frac{x_1 + 0.001(1-s)}{2}\right] \cdot \left[\sqrt{1 + (x_2 + x_3^2)\frac{x_4}{x_1^2}}\right]$$

$$T_2 = (x_1 + 3x_4)\exp[1 + \sin(x_3)]$$

$$P_2(\boldsymbol{x}, s) = A(s)\left[5 - \frac{2}{3}\exp\left(x_1 + x_2\right) - (x_4)\sin\left(x_3\right)A(s) + x_3 - B(s)\right]/4 - 0.7$$

where $A(s) = (0.9 + 0.1s)$ and $B(s) = 0.1(1-s)$. Both Park functions now contain a fidelity parameter s. These Park functions are evaluated on a transformed input space

$$[x_1, x_2, x_3, x_4] \rightarrow [1 - 2(x_1 - 0.6)^2, x_2, 1 - 3(x_3 - 0.5)^2, 1 - (x_4 - 0.8)^2].$$

Bibliography

[1] Rolf Wideröe. “Über ein neues Prinzip zur Herstellung hoher Spannungen.” In: *Archiv für Elektrotechnik* 21 (1928), pp. 387–406.

[2] Ernest O Lawrence and M Stanley Livingston. “The production of high speed light ions without the use of high voltages.” In: *Physical Review* 40.1 (1932), p. 19.

[3] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.” In: *Physics Letters B* 716.1 (2012), pp. 1–29.

[4] Massimo Altarelli et al. “The European X-ray free-electron laser.” In: *Technical Design Report, DESY 2006-097* (2007).

[5] Paul Emma et al. “First lasing and operation of an ångstrom-wavelength free-electron laser.” In: *Nature Photonics* 4.9 (2010), pp. 641–647.

[6] Henry N Chapman et al. “Femtosecond X-ray protein nanocrystallography.” In: *Nature* 470.7332 (2011), pp. 73–77.

[7] Petra Fromme and John CH Spence. “Femtosecond nanocrystallography using X-ray lasers for membrane protein structure determination.” In: *Current opinion in structural biology* 21.4 (2011), pp. 509–516.

[8] Andrew Aquila et al. “Time-resolved protein nanocrystallography using an X-ray free-electron laser.” In: *Optics express* 20.3 (2012), pp. 2706–2716.

[9] Linda Young et al. “Roadmap of ultrafast x-ray atomic and molecular physics.” In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 51.3 (2018), p. 032003.

[10] Radhe Mohan and David Grosshans. “Proton therapy—present and future.” In: *Advanced Drug Delivery Reviews* 109 (2017), pp. 26–44.

[11] Radhe Mohan. “A review of proton therapy–Current status and future directions.” In: *Precision radiation oncology* 6.2 (2022), pp. 164–176.

[12] X Xu et al. “RF breakdown studies in X-band klystron cavities.” In: *Proceedings of the 1997 Particle Accelerator Conference (Cat. No. 97CH36167).* Vol. 3. IEEE. 1997, pp. 3045–3047.

[13] Helmut Wiedemann. *Particle accelerator physics.* Springer Nature, 2015.

[14] E. Esarey, C. B. Schroeder, and W. P. Leemans. “Physics of laser-driven plasma-based electron accelerators.” In: *Reviews of Modern Physics* 81.3 (2009), pp. 1229–1285. DOI: 10.1103/revmodphys.81.1229.

[15] Simon Martin Hooker. “Developments in laser-driven plasma accelerators.” In: *Nature Photonics* 7.10 (2013), pp. 775–782.

[16] Toshiki Tajima and John M Dawson. “Laser electron accelerator.” In: *Physical review letters* 43.4 (1979), p. 267.

[17] Wim P Leemans et al. “GeV electron beams from a centimetre-scale accelerator.” In: *Nature physics* 2.10 (2006), pp. 696–699.

[18] Chris E Clayton et al. “Self-guided laser wakefield acceleration beyond 1 GeV using ionization-induced injection.” In: *Physical review letters* 105.10 (2010), p. 105003.

[19] Donna Strickland and Gerard Mourou. “Compression of amplified chirped optical pulses.” In: *Optics communications* 55.6 (1985), pp. 447–449.

[20] François Amiranoff et al. “Observation of laser wakefield acceleration of electrons.” In: *Physical Review Letters* 81.5 (1998), p. 995.

[21] Jérôme Faure et al. “A laser–plasma accelerator producing monoenergetic electron beams.” In: *Nature* 431.7008 (2004), pp. 541–544.

[22] Stuart PD Mangles et al. “Monoenergetic beams of relativistic electrons from intense laser–plasma interactions.” In: *Nature* 431.7008 (2004), pp. 535–538.

[23] CGR Geddes et al. “High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding.” In: *Nature* 431.7008 (2004), pp. 538–541.

[24] AJ Gonsalves et al. “Petawatt laser guiding and electron beam acceleration to 8 GeV in a laser-heated capillary discharge waveguide.” In: *Physical review letters* 122.8 (2019), p. 084801.

[25] Brendan Kettle et al. “Extended X-ray absorption spectroscopy using an ultrashort pulse laboratory-scale laser-plasma accelerator.” In: *Communications Physics* 7.1 (2024), p. 247.

[26] Yannick Glinec et al. “Radiotherapy with laser-plasma accelerators: Monte Carlo simulation of dose deposited by an experimental quasimonoenergetic electron beam.” In: *Medical physics* 33.1 (2006), pp. 155–162.

[27] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. “Practical bayesian optimization of machine learning algorithms.” In: *Advances in neural information processing systems* 25 (2012).

[28] Turab Lookman et al. “Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design.” In: *npj Computational Materials* 5.1 (2019), p. 21.

[29] Roberto Calandra et al. “Bayesian optimization for learning gaits under uncertainty: An experimental comparison on a dynamic bipedal walker.” In: *Annals of Mathematics and Artificial Intelligence* 76 (2016), pp. 5–23.

[30] RJ Shalloo et al. “Automation and control of laser wakefield accelerators using Bayesian optimization.” In: *Nature communications* 11.1 (2020), pp. 1–8.

[31] Sören Jalas et al. “Bayesian Optimization of a Laser-Plasma Accelerator.” In: *Physical review letters* 126.10 (2021), p. 104801.

[32] Rémi Lehe et al. “A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm.” In: *Computer Physics Communications* 203 (2016), pp. 66–82.

[33] Bahaa EA Saleh and Malvin Carl Teich. *Fundamentals of photonics*. john Wiley & sons, 2019.

[34] Andrew M Weiner. *Ultrafast optics*. John Wiley & Sons, 2011.

[35] John David Jackson. *Classical electrodynamics*. John Wiley & Sons, 2021.

[36] Peter W Milonni and Joseph H Eberly. *Laser physics*. John Wiley & Sons, 2010.

[37] Jean-Claude Diels and Wolfgang Rudolph. *Ultrashort laser pulse phenomena*. Elsevier, 2006.

[38] William Kruer. *The physics of laser plasma interactions*. crc Press, 2019.

[39] Paul Gibbon. *Short pulse laser interactions with matter: an introduction*. World Scientific, 2005.

[40] Francis F Chen et al. *Introduction to plasma physics and controlled fusion*. Vol. 1. Springer, 1984.

[41] LV184662 Keldysh. “Ionization in the field of a strong electromagnetic wave.” In: *Selected Papers of Leonid V Keldysh*. World Scientific, 2024, pp. 56–63.

[42] Anatoliĭ Aleksandrovich Vlasov. “The vibrational properties of an electron gas.” In: *Soviet Physics Uspekhi* 10.6 (1968), p. 721.

[43] Rémi Lehe et al. “A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm.” In: *Computer Physics Communications* 203 (2016), pp. 66–82. DOI: 10.1016/j.cpc.2016.02.007. eprint: 1507.04790.

[44] Johannes Götzfried. “Beam loading in high-charge laser wakefield accelerators.” PhD thesis. lmu, 2023.

[45] MJV Streeter et al. “Observation of laser power amplification in a self-injecting laser wakefield accelerator.” In: *Physical Review Letters* 120.25 (2018), p. 254801.

[46] WP Leemans et al. “Experiments and simulations of tunnel-ionized plasmas.” In: *Physical Review A* 46.2 (1992), p. 1091.

[47] Brice Quesnel and Patrick Mora. “Theory and simulation of the interaction of ultraintense laser pulses with electrons in vacuum.” In: *Physical Review E* 58.3 (1998), p. 3719.

[48] CS Liu, VK Tripathi, and Bengt Eliasson. *High-power laser-plasma interaction.* Cambridge university press, 2019.

[49] Peter Mulser and Dieter Bauer. *High power laser-matter interaction.* Vol. 238. Springer Science & Business Media, 2010.

[50] P Sprangle et al. “Laser wakefield acceleration and relativistic optical guiding.” In: *Applied Physics Letters* 53.22 (1988), pp. 2146–2148.

[51] P Sprangle, Cha-Mei Tang, and E Esarey. “Relativistic self-focusing of short-pulse radiation beams in plasmas.” In: *IEEE transactions on plasma science* 15.2 (1987), pp. 145–153.

[52] Guo-Zheng Sun et al. “Self-focusing of short intense pulses in plasmas.” In: *The Physics of fluids* 30.2 (1987), pp. 526–532.

[53] Jérôme Faure et al. “Observation of laser-pulse shortening in nonlinear plasma waves.” In: *Physical review letters* 95.20 (2005), p. 205003.

[54] LM Gorbunov and VI Kirsanov. “Excitation of plasma waves by an electromagnetic wave packet.” In: *Zh. Eksp. Teor. Fiz* 93 (1987), pp. 509–518.

[55] P Sprangle, Eric Esarey, and A Ting. “Nonlinear theory of intense laser-plasma interactions.” In: *Physical review letters* 64.17 (1990), p. 2011.

[56] Eric Esarey et al. “Overview of plasma-based accelerator concepts.” In: *IEEE Transactions on plasma science* 24.2 (1996), pp. 252–288.

[57] CB Schroeder et al. “Trapping, dark current, and wave breaking in nonlinear plasma waves.” In: *Physics of Plasmas* 13.3 (2006).

[58] Alexancer Pukhov and Jürgen Meyer-ter-Vehn. “Laser wake field acceleration: the highly non-linear broken-wave regime.” In: *Applied Physics B* 74 (2002), pp. 355–361.

[59] Wei Lu et al. “Nonlinear theory for relativistic plasma wakefields in the blowout regime.” In: *Physical review letters* 96.16 (2006), p. 165002.

[60] W Lu et al. “A nonlinear theory for multidimensional relativistic plasma wave wakefields.” In: *Physics of Plasmas* 13.5 (2006).

[61] Wei Lu et al. “Generating multi-GeV electron bunches using single stage laser wakefield acceleration<? format?> in a 3D nonlinear regime.” In: *Physical Review Special Topics—Accelerators and Beams* 10.6 (2007), p. 061301.

[62] Eric Esarey and Mark Pilloff. “Trapping and acceleration in nonlinear plasma waves.” In: *Physics of Plasmas* 2.5 (1995), pp. 1432–1436.

[63] Sébastien Corde et al. “Observation of longitudinal and transverse self-injections in laser-plasma accelerators.” In: *Nature Communications* 4.1 (2013), p. 1501.

[64] Sergei Bulanov et al. “Particle injection into the wave acceleration phase due to nonlinear wake wave breaking.” In: *Physical Review E* 58.5 (1998), R5257.

[65] A. Buck et al. “Shock-Front Injector for High-Quality Laser-Plasma Acceleration.” In: *Physical Review Letters* 110.18 (2013), p. 185006. DOI: `10.1103/physrevlett.110.185006`.

[66] Arthur Pak et al. “Injection and trapping of tunnel-ionized electrons into laser-produced wakes.” In: *Physical review letters* 104.2 (2010), p. 025003.

[67] E Esarey et al. “Electron injection into plasma wakefields by colliding laser pulses.” In: *Physical Review Letters* 79.14 (1997), p. 2682.

[68] Jérôme Faure et al. “Controlled injection and acceleration of electrons in plasma wakefields by colliding laser pulses.” In: *Nature* 444.7120 (2006), pp. 737–739.

[69] Alexander George Roy Thomas. “Scalings for radiation from plasma bubbles.” In: *Physics of Plasmas* 17.5 (2010).

[70] Stuart PD Mangles et al. “Self-injection threshold in self-guided laser wakefield accelerators.” In: *Physical Review Special Topics—Accelerators and Beams* 15.1 (2012), p. 011302.

[71] AJ Gonsalves et al. “Tunable laser plasma accelerator based on longitudinal density tailoring.” In: *Nature Physics* 7.11 (2011), pp. 862–866.

[72] CGR Geddes et al. “Plasma-density-gradient injection of low absolute-momentum-spread electron bunches.” In: *Physical review letters* 100.21 (2008), p. 215004.

[73] John David Anderson. “Modern compressible flow: with historical perspective.” In: *(No Title)* (1990).

[74] Wentao Wang et al. “Free-electron lasing at 27 nanometres based on a laser wakefield accelerator.” In: *Nature* 595.7868 (2021), pp. 516–520.

[75] Karl Schmid et al. “Density-transition based electron injector for laser driven wakefield accelerators.” In: *Physical Review Special Topics-Accelerators and Beams* 13.9 (2010), p. 091301.

[76] KK Swanson et al. “Control of tunable, monoenergetic laser-plasma-accelerated electron beams using a shock-induced density downramp injector.” In: *Physical Review Accelerators and Beams* 20.5 (2017), p. 051301.

[77] E Esarey et al. “Nonlinear pump depletion and electron dephasing in laser wakefield accelerators.” In: *Aip conference proceedings*. Vol. 737. 1. Citeseer. 2004, pp. 578–584.

[78] W Rittershofer et al. “Tapered plasma channels to phase-lock accelerating and focusing forces in laser-plasma accelerators.” In: *Physics of Plasmas* 17.6 (2010).

[79] Emilien Guillaume et al. “Electron rephasing in a laser-wakefield accelerator.” In: *Physical review letters* 115.15 (2015), p. 155002.

[80] Iris Kock, T Edler, and Stefan G Mayr. “Growth behavior and intrinsic properties of vapor-deposited iron palladium thin films.” In: *Journal of Applied Physics* 103.4 (2008).

[81] CG Durfee Iii and HM Milchberg. “Light pipe for high intensity laser pulses.” In: *Physical review letters* 71.15 (1993), p. 2409.

[82] Robert J Shalloo et al. “Hydrodynamic optical-field-ionized plasma channels.” In: *Physical Review E* 97.5 (2018), p. 053203.

[83] Thomas Bayes. “LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S.” In: *Philosophical transactions of the Royal Society of London* 53 (1763), pp. 370–418.

[84] Pierre Simon Laplace. *Théorie analytique des probabilités*. Courcier, 1820.

[85] Stephen M Stigler. *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press, 1990.

[86] José M Bernardo and Adrian FM Smith. *Bayesian theory*. Vol. 405. John Wiley & Sons, 2009.

[87] Andrew Gelman et al. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[88] William Feller. *An introduction to probability theory and its applications, Volume 2*. Vol. 81. John Wiley & Sons, 1991.

[89] Athanasios Papoulis. *Random variables and stochastic processes*. McGraw Hill, 1965.

[90] Morris H DeGroot. “Probability and statistics.” In: *(No Title)* (2002).

[91] Geoffrey Grimmett and David Stirzaker. *Probability and random processes*. Oxford university press, 2020.

[92] Carl Edward Rasmussen, Christopher KI Williams, et al. *Gaussian processes for machine learning.* Vol. 1. Springer, 2006.

[93] Charles A Micchelli, Yuesheng Xu, and Haizhang Zhang. "Universal Kernels." In: *Journal of Machine Learning Research* 7.12 (2006).

[94] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2018.

[95] David JC MacKay. "Comparison of approximate methods for handling hyperparameters." In: *Neural computation* 11.5 (1999), pp. 1035–1068.

[96] Kevin P Murphy. *Machine learning: a probabilistic perspective.* MIT press, 2012.

[97] Joaquin Quinonero-Candela and Carl Edward Rasmussen. "A unifying view of sparse approximate Gaussian process regression." In: *The Journal of Machine Learning Research* 6 (2005), pp. 1939–1959.

[98] Takuya Kanazawa. *Using Distance Correlation for Efficient Bayesian Optimization.* 2025. arXiv: `2102.08993 [cs.LG]`. URL: `https://arxiv.org/abs/2102.08993`.

[99] Shibo Li et al. "Multi-Fidelity Bayesian Optimization via Deep Neural Networks." In: *Advances in Neural Information Processing Systems.* Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 8521–8531. URL: `https://proceedings.neurips.cc/paper_files/paper/2020/file/60e1deb043af37db5ea4ce9ae8d2c9ea-Paper.pdf`.

[100] Dennis D Cox and Susan John. "A statistical method for global optimization." In: *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics.* IEEE. 1992, pp. 1241–1246.

[101] Jonas Mockus. "Application of Bayesian approach to numerical methods of global and stochastic optimization." In: *Journal of Global Optimization* 4.4 (1994), pp. 347–365.

[102] Peter I Frazier, Warren B Powell, and Savas Dayanik. "A knowledge-gradient policy for sequential information collection." In: *SIAM Journal on Control and Optimization* 47.5 (2008), pp. 2410–2439.

[103] Warren Scott, Peter Frazier, and Warren Powell. "The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression." In: *SIAM Journal on Optimization* 21.3 (2011), pp. 996–1026.

[104] Philipp Hennig and Christian J Schuler. "Entropy Search for Information-Efficient Global Optimization." In: *Journal of Machine Learning Research* 13.6 (2012).

[105] Zi Wang and Stefanie Jegelka. “Max-value entropy search for efficient Bayesian optimization.” In: *International Conference on Machine Learning*. PMLR. 2017, pp. 3627–3635.

[106] Quoc Phong Nguyen, Bryan Kian Hsiang Low, and Patrick Jaillet. *Rectified Max-Value Entropy Search for Bayesian Optimization*. 2022. arXiv: 2202.13597 [cs.LG]. URL: https://arxiv.org/abs/2202.13597.

[107] Laurence Charles Ward Dixon. “The global optimization problem. an introduction.” In: *Toward global optimization* 2 (1978), pp. 1–15.

[108] Carla Currin et al. “Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments.” In: *Journal of the American Statistical Association* 86.416 (1991), pp. 953–963.

[109] Jürgen Branke et al. *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Vol. 5252. Springer Science & Business Media, 2008.

[110] R Timothy Marler and Jasbir S Arora. “The weighted sum method for multi-objective optimization: new insights.” In: *Structural and multidisciplinary optimization* 41 (2010), pp. 853–862.

[111] Joshua Knowles. “ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems.” In: *IEEE Transactions on Evolutionary Computation* 10.1 (2006), pp. 50–66.

[112] Evangelos Triantaphyllou and Evangelos Triantaphyllou. *Multi-criteria decision making methods*. Springer, 2000.

[113] Kaisa Miettinen. *Nonlinear multiobjective optimization*. Vol. 12. Springer Science & Business Media, 1999.

[114] Kaifeng Yang et al. “Multi-objective Bayesian global optimization using expected hypervolume improvement gradient.” In: *Swarm and evolutionary computation* 44 (2019), pp. 945–956.

[115] Michael TM Emmerich, Kyriakos C Giannakoglou, and Boris Naujoks. “Single-and multiobjective evolutionary optimization assisted by Gaussian random field metamodels.” In: *IEEE Transactions on Evolutionary Computation* 10.4 (2006), pp. 421–439.

[116] Ivo Couckuyt, Dirk Deschrijver, and Tom Dhaene. “Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization.” In: *Journal of Global Optimization* 60.3 (2014), pp. 575–594.

[117] Chang Luo, Koji Shimoyama, and Shigeru Obayashi. “Kriging model based many-objective optimization with efficient calculation of expected hypervolume improvement.” In: *2014 IEEE Congress on Evolutionary Computation (CEC)*. IEEE. 2014, pp. 1187–1194.

[118] Koji Shimoyama, Shinkyu Jeong, and Shigeru Obayashi. “Kriging-surrogate-based optimization considering expected hypervolume improvement in non-constrained many-objective test problems.” In: *2013 IEEE Congress on Evolutionary Computation*. IEEE. 2013, pp. 658–665.

[119] Michael TM Emmerich, André H Deutz, and Jan Willem Klinkenberg. “Hypervolume-based expected improvement: Monotonicity properties and exact computation.” In: *2011 IEEE Congress of Evolutionary Computation (CEC)*. IEEE. 2011, pp. 2147–2154.

[120] Iris Hupkens et al. “Faster exact algorithms for computing expected hypervolume improvement.” In: *international conference on evolutionary multi-criterion optimization*. Springer. 2015, pp. 65–79.

[121] Michael Emmerich et al. “A multicriteria generalization of bayesian global optimization.” In: *Advances in Stochastic and Deterministic Global Optimization*. Springer, 2016, pp. 229–242.

[122] Kaifeng Yang et al. “Computing 3-D expected hypervolume improvement and related integrals in asymptotically optimal time.” In: *International Conference on Evolutionary Multi-Criterion Optimization*. Springer. 2017, pp. 685–700.

[123] Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. “Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization.” In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 9851–9864.

[124] Shinya Suzuki et al. “Multi-objective Bayesian optimization using Pareto-frontier entropy.” In: *International Conference on Machine Learning*. PMLR. 2020, pp. 9279–9288.

[125] Ben Tu et al. “Joint Entropy Search for Multi-Objective Bayesian Optimization.” In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., 2022, pp. 9922–9938. URL: `https://proceedings.neurips.cc/paper_files/paper/2022/file/4086fe59dc3584708468fba0e459f6a7-Paper-Conference.pdf`.

[126] Deng Huang et al. “Sequential kriging optimization using multiple-fidelity evaluations.” In: *Structural and Multidisciplinary Optimization* 32.5 (2006), pp. 369–382.

[127] Victor Picheny et al. “Quantile-based optimization of noisy computer experiments with tunable precision.” In: *Technometrics* 55.1 (2013), pp. 2–13.

[128] Kevin Swersky, Jasper Snoek, and Ryan P Adams. “Multi-Task Bayesian Optimization.” In: *Advances in Neural Information Processing Systems*. Ed. by C. J. C. Burges et al. Vol. 26. Curran Associates, Inc., 2013.

[129] Aaron Klein et al. "Fast bayesian optimization of machine learning hyperparameters on large datasets." In: *Artificial Intelligence and Statistics.* PMLR. 2017, pp. 528–536.

[130] Mark McLeod, Michael A Osborne, and Stephen J Roberts. *Practical bayesian optimization for variable cost objectives.* 2017. arXiv: `1703.04335 [stat.ML]`.

[131] Yehong Zhang et al. "Information-based multi-fidelity Bayesian optimization." In: *NIPS Workshop on Bayesian Optimization.* 2017.

[132] Rémi Lam, Douglas L Allaire, and Karen E Willcox. "Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources." In: *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference.* 2015, p. 0143.

[133] Jian Wu and Peter I Frazier. "Continuous-fidelity Bayesian optimization with knowledge gradient." In: *NIPS Workshop on Bayesian Optimization.* 2017.

[134] Jian Wu et al. "Practical multi-fidelity bayesian optimization for hyperparameter tuning." In: *Uncertainty in Artificial Intelligence.* PMLR. 2020, pp. 788–798.

[135] Shion Takeno et al. "Multi-fidelity Bayesian optimization with max-value entropy search and its parallelization." In: *International Conference on Machine Learning.* PMLR. 2020, pp. 9334–9345.

[136] Kirthevasan Kandasamy et al. "The multi-fidelity multi-armed bandit." In: *Advances in neural information processing systems* 29 (2016), pp. 1777–1785.

[137] Kirthevasan Kandasamy et al. "Multi-fidelity bayesian optimisation with continuous approximations." In: *International Conference on Machine Learning.* PMLR. 2017, pp. 1799–1808.

[138] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. "Multi-fidelity multi-objective bayesian optimization: an output space entropy search approach." In: *Proceedings of the thirty-fourth AAAI Conference on artificial intelligence.* 2020, pp. 10035–10043.

[139] Faran Irshad, Stefan Karsch, and Andreas Döpp. "Leveraging trust for joint multi-objective and multi-fidelity optimization." In: *Machine Learning: Science and Technology* 5.1 (2024), p. 015056.

[140] Tinkle Chugh. "Scalarizing Functions in Bayesian Multiobjective Optimization." In: *2020 IEEE Congress on Evolutionary Computation (CEC).* 2020, pp. 1–8. DOI: `10.1109/CEC48606.2020.9185706`.

[141] Maximilian Balandat et al. “BoTorch: A framework for efficient Monte-Carlo Bayesian optimization.” In: *Advances in Neural Information Processing Systems (NeurIPS)* (2020).

[142] Jeong Soo Park. “Tuning complex computer codes to data and optimal designs.” PhD thesis. University of Illinois at Urbana-Champaign, 1991.

[143] Charles K Birdsall and A Bruce Langdon. *Plasma physics via computer simulation*. CRC press, 2018.

[144] John M Dawson. “Particle simulation of plasmas.” In: *Reviews of modern physics* 55.2 (1983), p. 403.

[145] Toshi Tajima. *Computational plasma physics: with applications to fusion and astrophysics*. CRC press, 2018.

[146] John Dawson. “One-dimensional plasma model.” In: *The Physics of Fluids* 5.4 (1962), pp. 445–459.

[147] Oscar Buneman. “Dissipation of currents in ionized media.” In: *Physical Review* 115.3 (1959), p. 503.

[148] Yu N Grigoryev, Vitaliĭ Andreevich Vshivkov, and Mikhail Petrovich Fedoruk. *Numerical” particle-in-cell” methods: theory and applications*. Walter de Gruyter, 2012.

[149] Anatoly Spitkovsky. “Particle acceleration in relativistic collisionless shocks: Fermi process at last?” In: *The Astrophysical Journal* 682.1 (2008), p. L5.

[150] A Bruce Langdon and Barbara F Lasinski. “Electromagnetic and relativistic plasma simulation models.” In: *Methods in Computational Physics* 16 (1976), pp. 327–366.

[151] J-L Vay et al. “Numerical methods for instability mitigation in the modeling of laser wakefield accelerators in a Lorentz-boosted frame.” In: *Journal of Computational Physics* 230.15 (2011), pp. 5908–5929.

[152] R Koch. “Wave–particle interactions in plasmas.” In: *Plasma physics and controlled fusion* 48.12B (2006), B329.

[153] Ricardo A Fonseca et al. “OSIRIS: A three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators.” In: *Computational Science—ICCS 2002: International Conference Amsterdam, The Netherlands, April 21–24, 2002 Proceedings, Part III 2*. Springer. 2002, pp. 342–351.

[154] Luca Fedeli et al. “Pushing the frontier in the design of laser-based electron accelerators with groundbreaking mesh-refined particle-in-cell simulations on exascale-class supercomputers.” In: *SC22: international conference for high performance computing, networking, storage and analysis*. IEEE. 2022, pp. 1–12.

[155] TD Arber et al. “Contemporary particle-in-cell approach to laser-plasma modelling.” In: *Plasma Physics and Controlled Fusion* 57.11 (2015), p. 113001.

[156] Chengkun Huang et al. “QUICKPIC: A highly efficient particle-in-cell code for modeling wakefield acceleration in plasmas.” In: *Journal of Computational Physics* 217.2 (2006), pp. 658–679.

[157] Agustin F Lifschitz et al. “Particle-in-cell modelling of laser–plasma interaction using Fourier decomposition.” In: *Journal of Computational Physics* 228.5 (2009), pp. 1803–1814.

[158] Heiko Burau et al. “PIConGPU: a fully relativistic particle-in-cell code for a GPU cluster.” In: *IEEE Transactions on Plasma Science* 38.10 (2010), pp. 2831–2839.

[159] R Lehe et al. “Numerical growth of emittance in simulations of laser-wakefield acceleration.” In: *Physical Review Special Topics—Accelerators and Beams* 16.2 (2013), p. 021301.

[160] Benjamin M Cowan et al. “Generalized algorithm for control of numerical dispersion in explicit<? format?> time-domain electromagnetic simulations.” In: *Physical Review Special Topics—Accelerators and Beams* 16.4 (2013), p. 041303.

[161] Brendan B Godfrey. “Numerical Cherenkov instabilities in electromagnetic particle codes.” In: *Journal of Computational Physics* 15.4 (1974), pp. 504–521.

[162] Richard Courant, Kurt Friedrichs, and Hans Lewy. “Über die partiellen Differenzengleichungen der mathematischen Physik.” In: *Mathematische annalen* 100.1 (1928), pp. 32–74.

[163] J. Götzfried et al. “Physics of High-Charge Electron Beams in Laser-Plasma Wakefields.” In: *Physical Review X* 10.4 (2020), p. 041015. DOI: 10.1103/physrevx.10.041015. eprint: 2004.10310.

[164] H. Ding et al. “Nonlinear plasma wavelength scalings in a laser wakefield accelerator.” In: *Physical Review E* 101.2 (2020), p. 023209. DOI: 10.1103/physreve.101.023209. eprint: 2001.09507.

[165] *FBPIC: Running boosted-frame simulations.* https://fbpic.github.io/advanced/boosted_frame.html. Accessed: 2024-11-26.

[166] F. Irshad, S. Karsch, and A. Döpp. “Multi-objective and multi-fidelity Bayesian optimization of laser-plasma acceleration.” In: *Phys. Rev. Res.* 5 (1 2023), p. 013063. DOI: 10.1103/PhysRevResearch.5.013063. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.5.013063.

[167] David Ruppert and David S Matteson. *Statistics and data analysis for financial engineering*. Vol. 13. Springer, 2011. DOI: 10.1007/978-1-4939-2614-5.

[168] Max Gilljohann. “Towards hybrid wakefield acceleration.” PhD thesis. lmu, 2022.

[169] Moritz Foerster. “Hybrid Wakefield acceleration.” PhD thesis. lmu, 2024.

[170] Pierre Tournois. “Acousto-optic programmable dispersive filter for adaptive compensation of group delay time dispersion in laser systems.” In: *Optics communications* 140.4-6 (1997), pp. 245–249.

[171] A Moulet et al. “Single-shot, high-dynamic-range measurement of sub-15 fs pulses by self-referenced spectral interferometry.” In: *Optics letters* 35.22 (2010), pp. 3856–3858.

[172] Patrick O’shea et al. “Highly simplified device for ultrashort-pulse measurement.” In: *Optics letters* 26.12 (2001), pp. 932–934.

[173] Emilien Guillaume. “Control of electron injection and acceleration in Laser-Wakefield Accelerators.” PhD thesis. École Polytechnique, 2015.

[174] Nils Weiße et al. “Tango Controls and Data Pipeline for Petawatt Laser Experiments.” In: *High Power Laser Science and Engineering* 11 (2022), pp. 1–8.

[175] TANGO Controls Collaboration. *TANGO Controls*. https://www.tango-controls.org. Accessed: 2024-05-22. 2024.

[176] Moritz Foerster. “Hybrid Wakefield acceleration: a source of stable and high-density electron beams.” PhD thesis. Ludwig Maximilians Universität München, 2024.

[177] Katinka Grafenstein. “Laser wakefield acceleration in the GeV regime using optically-induced shock injection.” PhD thesis. Ludwig Maximilians Universität München, 2024.

[178] Cédric Thaury et al. “Shock assisted ionization injection in laser-plasma accelerators.” In: *Scientific reports* 5.1 (2015), pp. 1–7.

[179] Andreas Döpp et al. “Data-driven Science and Machine Learning Methods in Laser-Plasma Physics.” In: *High Power Laser Science and Engineering* 11 (2023), pp. 55–96. DOI: https://doi.org/10.1017/hpl.2023.47.

[180] Arthur P Dempster, Nan M Laird, and Donald B Rubin. “Maximum likelihood from incomplete data via the EM algorithm.” In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22.

[181] Manuel Kirchen et al. “Optimal Beam Loading in a Laser-Plasma Accelerator.” In: *Physical Review Letters* 126.17 (2021), p. 174801. DOI: 10.1103/physrevlett.126.174801.

[182] FM Foerster et al. “Stable and High-Quality Electron Beams from Staged Laser and Plasma Wakefield Accelerators.” In: *Physical Review X* 12.4 (2022), p. 041016.

[183] Richard H Byrd et al. “A limited memory algorithm for bound constrained optimization.” In: *SIAM Journal on scientific computing* 16.5 (1995), pp. 1190–1208.

List of Publications by the Author

As first author

- **F. Irshad**, C. Eberle, F.M. Foerster, K.v. Grafenstein, F. Haberstroh, E. Travac, N. Weisse, S. Karsch, and A. Döpp, *Pareto Optimization and Tuning of a Laser Wakefield Accelerator*, Physical Review Letters 133, 085001 (2024).
- **F. Irshad**, S. Karsch, and A. Döpp, *Multi-objective and multi-fidelity Bayesian optimization of laser-plasma acceleration*, Physical Review Research 5.1, 013063 (2023)
- **F. Irshad**, S. Karsch, and A. Döpp, *Leveraging trust for joint multi-objective and multi-fidelity optimization*, Machine Learning: Science and Technology, 5.1, 015056 (2024).

As co-author

- A. Döpp, C. Eberle, S. Howard, **F. Irshad**, J. Lin, and M. Streeter, *Data-driven science and machine learning methods in laser–plasma physics.*, High Power Laser Science and Engineering 11, (2023).
- K.v. Grafenstein, F. M. Foerster, F. Haberstroh, D. Campbell, **F. Irshad**, F. C. Salgado, G. Schilling, E. Travac, N. Weiße, M. Zepf, A. Döpp and S. Karsch, *Laser-accelerated electron beams at 1 GeV using optically-induced shock injection*, Scientific Reports, 13(1), 11680, (2023).
- F.M. Foerster, A. Döpp, F. Haberstroh, K.v. Grafenstein, D. Campbell, Y.-Y. Chang, S. Corde, J.P. Couperus Cabadağ, A. Debus, M.F. Gilljohann, A.F. Habib, T. Heinemann, B. Hidding, A. Imran, **F. Irshad**, A. Knetsch, O. Kononenko, A. Martinez de la Ossa, A. Nutter, R. Pausch, G. Schilling, A. Schletter, S. Schöbel, U. Schramm, E. Travac, P. Ufer and S. Karsch, *Stable and High-Quality Electron Beams from Staged Laser and Plasma Wakefield Accelerators*, Physical Review X 12(4), (2022).

- N. Weiße, L. Doyle, J. Gebhard, F. Balling, F. Schweiger, F. Haberstroh, L.D. Geulig, J. Lin, **F. Irshad**, J. Esslinger and S. Gerlach, *Tango Controls and data pipeline for petawatt laser experiments.*, High Power Laser Science and Engineering 11, (2023).

Acknowledgements

I would like to start by thanking all the colleagues and people I met during my time at CALA.

Thanks to my doctoral advisor **Prof. Dr. Stefan Karsch** who gave me the chance to work at the forefront of laser wakefield acceleration. Your technical expertise and experience in laser-plasma interaction were a great source of learning for me. Your skepticism of machine learning also pushed me to develop and present my ideas in a more understandable manner and greatly improved my scientific communication.

I am deeply grateful to **Dr. Andreas Döpp** who introduced me to Bayesian optimization and motivated me to pursue this particular topic. The support and advice that you gave me throughout my years at CALA ranging from debugging code to theory of Bayesian to trying to help me through all my personal problems, was a big help. Your experience and skill at presenting results and forming arguments was truly a learning experience.

Many of the experiment days would never have gone smoothly if **Gregor Schilling** had not been there to constantly tune the ATLAS laser system. Thanks for the many hours you spent fixing and turning on the laser during experiment days. On this note, thanks to **Andreas Münzer** for stepping in whenever Gregor was not available.

I now want to thank my colleagues who formed the core of ETTF and with whom I spent most of my time working. Thanks to **Moritz Foerster**, **Florian Haberstroh** and **Katinka von Grafenstein** for helping out and teaching me so much about ATLAS and ETTF. You were always there to help me put together my experimental setup and to answer dozens of questions that I had about it. Thanks to **Enes Travac** for making beam times fun with his numerous never-ending jokes. **Christoph Eberle** for helping me in numerous tasks, from writing and debugging my optimizer to proofreading my thesis. All of you spent many hours with me on the experiment evenings and made working at CALA enjoyable for me. I also want to mention the very different and interesting conversations I had about work or otherwise with all of you.

I also want to thank **Nils Weisse, Jinpu Lin, Johannes Zirkelbach** who made my experiments possible and helped out in many tasks during these beamtimes. I would also like to acknowledge many people with whom I worked at different times in the group, including **David Campbell, Jannik Esslinger, Max Gilljohann, Johannes Götzfried, Albert Schletter, Jakob Schröder**. I also want to thank the newer members of ETTF: **Mohamed Ayache, Felipe Peña, Sanchita Sharan, Maurice Zeuner, Bi Zewu** and wish them the best for their future experimental campaigns.

Many thanks to the other group leaders working at CALA: **Jörg Schreiber, Peter Thierolf, Andreas Döpp** and their members, especially to **Leonard Doyle and Florian Schweiger** who helped me in setting up the Bayesian optimizer and the different devices in Tango.

A lot of thanks to **Martin Groß, Klaus Franke, Felix Rauscher, Katharina Adler, Hans-Friedrich Wirth, Oliver Gosau, Nik Gjotev and Florian Saran** for their administrative and technical support at CALA.

I am also grateful to the Max Planck School of Photonics (MPSP) for the financial support and interesting spring and fall schools that helped me a lot during my doctorate. On this note thanks to the people at IMPRS-APS for making the Ringberg meetings fun.

At the end I would like to thank my family without whom I would not be at the place I am today. I am infinitely grateful to my **father** for all his hard work over the years that opened up so many opportunities for me.

میں اپنی والدہ کا تہہ دل سے شکر گزار ہوں جن کی دعاؤں اور محبت نے مجھے اس سفر میں استقامت بخشی۔
ماں جی، مجھے پتا ہے کہ آپ میری اتنی اچھی اردو پر حیران ہوں گی۔

To all my siblings for their constant encouragement. Finally, but above all, thanks to my wife **Mishal Rizwan** for your unwavering support and love through thick and thin all these years, especially during countless beamtime nights. You mean the world to me. Lastly, to my daughter **Amyra “meeru”** who makes every day a joy for me.

آپ فکر نہ کریں، بابا آپ کو اکیلا نہیں سونے دیں گے اور ہمیشہ آپ کے پاس رہیں گے۔