Development of a longitudinal phase space density reconstruction algorithm at KARA

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von der KIT-Fakultät für Elektrotechnik und Informationstechnik
des Karlsruher Instituts für Technologie (KIT)

angenommene
Dissertation

von
M.Sc. Felipe Antonio Donoso Aguirre
geboren in Santiago (Chile)

Tag der mündlichen Prüfung:	29. Oktober 2026
Hauptreferent:	Prof. Dr. Theo Scherer
Korreferentin:	Prof. Dr. Anke-Susanne Müller
Korreferent:	Prof. Dr. Martin Frank

Abstract

The fast longitudinal dynamics of electron bunches in synchrotron storage rings, particularly the micro-bunching instability, can limit accelerator performance, making their mitigation and control an important challenge for next generation synchrotron light sources.

This work presents the first experimental application of a new longitudinal phase space reconstruction method constrained by the Vlasov–Fokker–Planck equation (VFPE) using electro-optical spectral decoding (EOSD) measurements at the Karlsruhe Research Accelerator (KARA). A surrogate model is developed for the EOSD measurements, in which the system impulse response is obtained from electromagnetic simulations of the diagnostic setup and convolved with the longitudinal charge density profile. Combined with the VFPE model for the longitudinal phase space density evolution, this formulation leads to a nonlinear partial differential equation (PDE)-constrained inverse problem for reconstructing the longitudinal phase space density.

The nonlinear inverse problem is solved using an adjoint-based optimization method. In this formulation, the adjoint provides the gradient of the reconstruction objective, while iterative linearization accounts for the nonlinear collective beam dynamics. This approach is applied across different dynamical regimes, ranging from stable synchrotron motion to the onset of the fast dynamics of the micro-bunching instability.

The reconstruction method is verified and validated using simulated and experimental data. Numerical verification with synthetic data uses the known ground-truth phase space density as a direct reference for evaluating the reconstruction. Experimental validation is then performed using turn-by-turn EOSD measurements for the reconstruction, together with synchronized measurements from two independent diagnostics to relate the reconstructed longitudinal dynamics to the horizontal bunch size and the power of the emitted coherent synchrotron radiation.

The results show that the reconstructed non-equilibrium longitudinal phase space density dynamics are consistent with the expected physical evolution of the electron bunch. As a result, the new method enables efficient and reliable reconstruction of the longitudinal phase space dynamics from measurements of the longitudinal bunch profile.

Kurzfassung

Die schnelle longitudinale Dynamik von Elektronenpaketen in Synchrotron-Speicherringen, insbesondere die Mikrobunching-Instabilität, kann die Leistungsfähigkeit des Beschleunigers begrenzen. Ihre Abschwächung und Kontrolle stellt daher eine wichtige Herausforderung für Synchrotronlichtquellen der nächsten Generation dar.

Diese Arbeit präsentiert die erste experimentelle Anwendung einer neuen Methode zur Rekonstruktion des longitudinalen Phasenraums, bei der die Vlasov-Fokker-Planck-Gleichung (VFPE) als Nebenbedingung verwendet wird und Messungen mittels elektrooptischer spektraler Dekodierung (EOSD) am Karlsruhe Research Accelerator (KARA) einbezogen werden. Für die EOSD-Messungen wird ein Ersatzmodell entwickelt, bei dem die Impulsantwort des Systems aus elektromagnetischen Simulationen des Diagnoseaufbaus bestimmt und mit dem longitudinalen Ladungsdichteprofil gefaltet wird. In Kombination mit dem VFPE-Modell für die Entwicklung der longitudinalen Phasenraumdichte führt diese Formulierung zu einem nichtlinearen inversen Problem mit einer partiellen Differentialgleichung (PDE) als Nebenbedingung zur Rekonstruktion der longitudinalen Phasenraumdichte.

Das nichtlineare inverse Problem wird mithilfe einer Optimierungsmethode auf Basis des adjungierten Problems gelöst. Dabei liefert die adjungierte Formulierung den Gradienten der Zielfunktion der Rekonstruktion, während eine iterative Linearisierung die nichtlineare kollektive Strahldynamik berücksichtigt. Dieser Ansatz wird auf unterschiedliche dynamische Regime angewendet, die von stabiler Synchrotronbewegung bis zum Einsetzen der schnellen Dynamik der Mikrobunching-Instabilität reichen.

Die Rekonstruktionsmethode wird anhand simulierter und experimenteller Daten verifiziert und validiert. Bei der numerischen Verifikation mit synthetischen Daten dient die bekannte Phasenraumdichte als direkte Referenz zur Bewertung der Rekonstruktion. Die experimentelle Validierung erfolgt anschließend anhand umlaufaufgelöster EOSD-Messungen für die Rekonstruktion sowie synchronisierter Messungen zweier unabhängiger Diagnosesysteme, um die rekonstruierte longitudinale Dynamik mit der horizontalen Paketgröße und der Leistung der emittierten kohärenten Synchrotronstrahlung in Beziehung zu setzen.

Die Ergebnisse zeigen, dass die rekonstruierte Nichtgleichgewichtsdynamik der longitudinalen Phasenraumdichte mit der erwarteten physikalischen Entwicklung des Elektronenpakets übereinstimmt. Damit ermöglicht die neue Methode eine effiziente und zuverlässige Rekonstruktion der longitudinalen Phasenraumdynamik auf Grundlage von Messungen des longitudinalen Ladungsprofils.

Acknowledgments

Funding

I acknowledge the support of the KIT Center MathSEE (Mathematics in Sciences, Engineering, and Economics) with its Seed Projects (Bridge PhDs) program, KIT Graduate School Computational and Data Science (KCDS), the KIT Scientific Computing Center (SCC), and the Institute for Beam Physics and Technology (IBPT) at KIT for supporting and funding this research.

Professors

I was fortunate to have three outstanding professors supervising and supporting me throughout my PhD. I would like to thank Prof. Theo Scherer for his kindness, guidance, and constant support. I am also deeply grateful to Prof. Anke-Susanne Müller for her leadership and vision. I would also like to thank Prof. Martin Frank for his scientific guidance, for introducing me to the world of applied mathematics, for his encouraging words, and for always being there when I needed his support.

Colleagues

There were many colleagues who contributed to my work and provided support throughout my PhD. I apologize in advance for not being able to mention everyone individually.

Dr. Stefan Funkner, however, deserves special recognition as my scientific advisor. His meticulous way of thinking is reflected throughout this thesis, and his guidance was essential during these years.

I am equally grateful to my boss, Dr. Erik Bründermann, for bringing balance to my research and helping guide its direction.

I owe my sincere gratitude to Micha for helping me understand the EOSD diagnostics and for his incredibly useful observations. I am also deeply grateful to Miriam, whose key insights into beam dynamics and beam measurements provided the exact spark needed to finally solve the core problem of this work. A particular thanks to Markus for his endless patience with my early misunderstandings of phase space evolution, expertly guiding me toward the correct physical interpretation.

Finally, many thanks to Andrea, Julian, Chenran, Gudrun, Alex, Joseph, Patrick, Johannes, Nathan, David, Matthias, Dima, Houssam, Thanos, and Bastian for their support, conversations, beers, and laughs, and for hopefully not being too upset that I left out all the doctoral titles before their names.

AI tools

Generative AI was used as a tool for grammatical and stylistic revision of text passages. It was also used occasionally for code generation and as a learning aid.

Family

Para mi familia, todo mi amor.
Gracias, Andre, por tu amor, apoyo y paciencia.
Gracias por sostener a nuestra familia mientras estuve absorto en el estudio.
Gracias a mis dos hermosos hijos, Joaquín y Violeta, por su amor incondicional, por entender tanto a pesar de su corta edad y por abrazar con alegría una vida lejos de la tierra que los vio nacer.

Mi camino en la ciencia empezó con mi padre y mi madre.
Gracias por su amor, que trasciende la distancia y el tiempo.
Por supuesto, mis hermanos y hermanas están siempre en mi corazón.

A mi familia en Ecuador, a mis suegros y a mis cuñadas, gracias por todo su cariño y apoyo.

Contents

1. Introduction **1**
1.1. State of the art in longitudinal phase space tomography 2
1.2. Thesis objective and structure 3

2. Longitudinal phase space density dynamics **5**
2.1. Particle density dynamics 5
2.1.1. Hamiltonian phase space density evolution 6
2.1.2. Fokker–Planck phase space density evolution 9
2.2. Electromagnetic forces in the co-moving reference frame 13
2.2.1. Reference orbit and co-moving coordinate system 13
2.2.2. Charged particle dynamics under the Lorentz force 15
2.3. Longitudinal beam dynamics 16
2.3.1. Synchrotron motion and momentum compaction factor 18
2.3.2. Energy and phase deviation equations 19
2.3.3. Wake fields and longitudinal impedances 22
2.3.4. Longitudinal phase space 26

3. Beam diagnostics for longitudinal beam dynamics at KARA **31**
3.1. Visible-light monitor 32
3.2. THz detector 33
3.3. Electro-optical spectral decoding 34
3.3.1. The Pockels effect and laser phase shift 35
3.3.2. Electromagnetic field simulations 37
3.3.3. EOSD simulation 41
3.3.4. EOSD impulse response 48

4. The Vlasov–Fokker–Planck equation solution **53**
4.1. Operator splitting based on particle conservation 53
4.2. Inovesa solver 56
4.2.1. Phase space grid coordinates 56
4.2.2. Lagrange polynomial interpolation 56
4.2.3. Inverse-map problem 59
4.2.4. Implementation of Hamiltonian operators 60
4.2.5. Implementation of the Fokker–Planck operator 64
4.3. Sparse matrix representation of the split operators 65

5. Longitudinal phase space density reconstruction **69**
5.1. Inverse problem formulation 69
5.2. PDE-constrained optimization solution 70
5.2.1. One-shot formulation 71
5.2.2. Optimization based on sensitivity equations 72
5.2.3. Optimization based on adjoint equations 73

5.2.4. Discretize-then-differentiate approach . . . 73
5.2.5. The adjoint-based optimization algorithm . . . 74
5.2.6. The inner–outer loop optimization algorithm . . . 76
5.3. Discrete operator adjoint formulation . . . 76
5.4. Optimization algorithms . . . 80
5.4.1. Line-search and trust-region strategies . . . 80
5.4.2. First-order methods . . . 80
5.4.3. Gradient descent . . . 81
5.4.4. Steepest descent . . . 81
5.4.5. Conjugate-gradient methods . . . 81
5.4.6. Constrained methods . . . 82
5.5. Linear least-squares formulation . . . 83
5.5.1. Equivalence of least-squares and adjoint gradients . . . 84
5.5.2. Matrix-free implementation . . . 85
5.5.3. Linear least-squares optimization algorithms . . . 86
5.5.4. Projected-restarted non-negative conjugate gradient for least-squares . . . 87

6. Inovesalib . . . 91
6.1. Architecture . . . 92
6.1.1. Inovesa core . . . 92
6.1.2. Parameters . . . 93
6.1.3. Observers . . . 93
6.1.4. Controllers . . . 93
6.1.5. Plugins . . . 93
6.2. Algorithmic differentiation . . . 94
6.3. C++ adaptation of IR Tools . . . 95

7. Reconstruction algorithm verification and validation . . . 97
7.1. Selection of the reconstruction algorithm . . . 98
7.1.1. Simulated datasets and evaluation protocol . . . 98
7.1.2. Static tomography as a limiting case . . . 101
7.1.3. Evolution of the reconstruction error . . . 102
7.1.4. Influence of regularization on the inner–outer loop . . . 103
7.1.5. Warm-start initialization of the inner–outer loop . . . 105
7.1.6. Stabilization by non-negative projected restarts . . . 107
7.1.7. Comparison of nonlinear and linearized dynamics reconstruction . . . 108
7.2. Reconstruction from simulated EOSD data . . . 111
7.2.1. Sensitivity to bunch current mismatch . . . 113
7.2.2. Sensitivity to RF voltage amplitude mismatch . . . 114
7.2.3. Sensitivity to impedance model mismatch . . . 115
7.3. Reconstruction from EOSD measurements . . . 116
7.3.1. Beam diagnostic data . . . 117
7.3.2. Phase space size and initial bunch length . . . 120
7.3.3. Consistency with EOSD measurements . . . 121
7.3.4. Comparison of derived observables . . . 123
7.3.5. Sensitivity to the EOSD impulse response . . . 126
7.3.6. Sensitivity to model parameters . . . 128
7.4. Discussion . . . 131

8. Conclusions **133**

A. Appendix **135**
A.1. Phase retardation algorithm calculation example 135
A.2. Derivation of the evolved quadrilateral area 139
A.3. Intermediate steps in the derivation of the Vlasov equation 141
A.4. Approximate second-order methods 142
A.4.1. The BFGS method 142
A.4.2. Symmetric Rank-One (SR1) 142
A.4.3. Limited-memory BFGS (L-BFGS) 143
A.5. Second-order methods 143
A.5.1. Exact Newton method 143
A.5.2. Gauss–Newton method 143
A.5.3. Matrix-free Newton–Krylov methods 144

Acronyms **145**

Symbols **147**

Glossary **153**

Publications **157**

Bibliography **159**

1. Introduction

The Karlsruhe Research Accelerator (KARA) is an electron storage ring operated both as a synchrotron radiation (SR) source and as a test facility for accelerator research. A central focus of its research program is the investigation of electron-bunch dynamics, particularly in non-equilibrium regimes.

The study of non-equilibrium bunch dynamics is fundamental for accelerator physics. Non-equilibrium dynamics affect beam stability, constrain operation near machine limits, and govern the generation of coherent synchrotron radiation (CSR). A detailed understanding of these dynamics is therefore essential for the design, operation, and control of present and future accelerators.

The KARA storage ring has a circumference of 110.4 m, stores electron beams with energy from 0.5 GeV to 2.5 GeV, and accelerates them with a maximum voltage of 1 600 kV with a radio frequency (RF) system close to 500 MHz. The RF system defines a nominal bucket spacing of about 2 ns, with the harmonic number of 184, while the revolution period of approximately 368 ns corresponds to a turn frequency of 2.7 MHz.

As an accelerator test facility, KARA can be operated over a broad range of beam energies, RF voltages, bunch currents, filling patterns, and momentum compaction factors [8]. In particular, low-momentum compaction operation produces shorter bunches and strengthens the influence of wake fields and CSR, which can become dominant in the longitudinal beam dynamics [9]. This operational flexibility provides controlled access to non-equilibrium regimes and makes KARA a suitable laboratory for cutting-edge experimental investigation [10].

The strong scientific interest in longitudinal beam dynamics, at KARA, has motivated sustained theoretical and experimental research [11, 12, 13], together with the development of dedicated diagnostics [14, 15, 16, 17]. These sensors provide access to key observables, including the longitudinal bunch profile, the emitted CSR power, and the horizontal beam profile.

Although these observables contain valuable information about the longitudinal dynamics, none of them provides a complete description of the beam state. A more comprehensive description is provided by the longitudinal phase space distribution, which relates the longitudinal position of the electrons within the bunch to their relative energy deviation. Since no diagnostic can directly resolve this distribution for all electrons, the longitudinal phase space cannot be measured directly and has to be inferred from beam observables.

The longitudinal phase space distribution can be reconstructed indirectly from sequences of beam observables. In particular, longitudinal phase space tomography uses bunch profiles as projections of the phase space distribution and reconstructs the underlying distribution using established tomographic techniques. These methods can provide reliable reconstructions when the phase space evolution is sufficiently stable and can be approximated as a rigid rotation [18]. In non-equilibrium regimes, however, the distribution may evolve rapidly under nonlinear collective effects, violating this assumption and limiting the applicability of conventional tomography.

This motivates the central problem addressed in this thesis: the reconstruction of longitudinal phase space densities in regimes where the beam distribution evolves under nonlinear collective effects and cannot be treated as a rigidly rotating object.

1.1. State of the art in longitudinal phase space tomography

Phase space tomography can be understood as a diagnostic technique in which beam observations are used to infer a higher-dimensional phase space distribution. The reconstructed phase space density is therefore not independent of the diagnostic system used to observe the beam, nor of the dynamical model that relates the particle evolution to the measurements. In accelerator physics, its development has therefore progressed alongside advances in beam diagnostics and in the dynamical models used to interpret their measurements.

An early experimental application of tomography in accelerator physics was reported by Fraser [19], who used algebraic reconstruction techniques and known transfer matrices to reconstruct both transverse and longitudinal phase space distributions from measured beam profiles. Mottershead [20] used transverse beam-profile measurements obtained from light emitted through interactions with residual gas at several known positions along a linac. Since the beam optics between the measurement stations were known, the corresponding phase space transformations could be described by transfer matrices and reconstructed by a maximum-entropy tomography.

Jackson [21] transferred the diagnostic principle to the longitudinal plane by combining a resistive-wall monitor and a fast digitizer with synchrotron motion in a synchrotron, which supplies different projections of the bunch. The projections are recorded at a fixed observation point at different times, so the evolving bunch profile provides different effective viewing angles of the phase space distribution, enabling reconstruction with a standard Radon transform approach.

Hancock et al. [22] addressed the rigid rotation limitation by incorporating particle tracking into the phase space reconstruction, allowing non-rigid phase space evolution to be included in the projection model. The reconstruction itself was performed using an algebraic reconstruction technique (ART) [23] in a proton synchrotron. This established the central idea of modern longitudinal tomography: the useful information delivered by the diagnostic depends on how faithfully the reconstruction represents the dynamics during acquisition.

Building on Hancock's work, longitudinal phase space tomography developed into a mature tracking-based framework [24, 25]. The original software was later re-implemented to enable parallel tracking, reuse of particle trajectories, and coupling to external tracking programs such as BLonD [26, 27]. This lineage extended tomography to RF manipulations and selected collective effects in proton machines. For electron storage rings, however, a finite macroparticle ensemble may undersample low density regions and produce a fragmented representation.

Progress in longitudinal phase space tomography at KARA has followed advances in high-resolution diagnostics for short bunches. Electro-optical spectral decoding (EOSD) [14] provides non-destructive, single-shot longitudinal bunch profiles on successive turns at MHz rates. Funkner et al. [18] used profiles spanning half a synchrotron period to reconstruct the longitudinal phase space density by filtered backprojection (FBP), treating synchrotron motion as a rigid rotation. This work established the direct experimental basis for the present thesis by demonstrating longitudinal phase space tomography from turn-by-turn EOSD measurements of a stored electron bunch. Its assumption also exposes the central limitation: under collective effects, profile changes contain both rotation and deformation of the density, which FBP cannot distinguish.

Recent developments use machine learning not only to accelerate tomography but also to increase its dimensionality and temporal resolution. Roussel et al. [28] combined neural representations of the beam distribution with differentiable particle tracking to reconstruct detailed four-dimensional phase space from quadrupole and screen measurements. It later extended this generative approach to experimental six-dimensional reconstruction [29]. Hoover and Wong [30] incorporated normalizing flows into a maximum-entropy formulation, making conservative six-dimensional inference computationally tractable. For longitudinal dynamics, Luo et al. [31] introduced a spatio-temporal convolutional neural network trained on BLonD simulations to reconstruct sequences of two-dimensional phase space distributions from one-dimensional current profiles with real-time inference. Together, these methods demonstrate the potential of learned priors, differentiable dynamics, and temporal networks.

Their reliability nevertheless depends on how accurately the reconstructed phase space represents the relevant beam dynamics. Since the true experimental phase space is not known, validation is commonly based on synthetic data with known ground truth, or error between predicted and real measurements. Validation is therefore a central issue: reproducing the measurements used in the reconstruction demonstrates consistency with the observation model, but it does not by itself prove that the reconstructed phase space represents the true beam dynamics.

This thesis is a natural continuation of the beam dynamics and advanced diagnostics studies at KARA. It builds on the work of Funkner et al. [18] and develops the reconstruction approach introduced by the author in Donoso et al. [1, 2, 3]. Together, these works define a methodological progression from the initial formulation and validation of the reconstruction algorithm with synthetic measurements, through increasingly realistic modeling of the diagnostic response, to its application and validation with experimental data.

Within this progression, the reconstruction framework is developed in three main steps:

- First, the rigid-rotation assumption is replaced by a reconstruction constrained by the Vlasov–Fokker–Planck equation (VFPE), which describes the evolution of the longitudinal phase space. The reconstruction determines an initial phase space density whose subsequent Hamiltonian transport, collective self-interaction, radiation damping, and diffusion reproduce a sequence of measured profiles.
- Second, the observation model is extended to include the physical response of the EOSD diagnostic. Since an EOSD measurement does not correspond directly to an ideal charge profile, the electromagnetic propagation through the crystal geometry and the optical system is represented by an efficient diagnostic impulse response, allowing the reconstruction to predict the measured EOSD signal.
- Finally, the complete reconstruction framework is applied to experimental EOSD measurements. The reconstructed phase space density is then independently evaluated using observables from synchronized diagnostics that are not included as inputs to the reconstruction.

1.2. Thesis objective and structure

The objective of this thesis is the development of a longitudinal phase space density reconstruction algorithm at KARA. The algorithm reconstructs the hidden two-dimensional longitudinal distribution of an electron bunch from measured one-dimensional observables while enforcing the accelerator physics model that describes the turn-by-turn evolution. The initial phase space density serves as the

optimization variable, while the central unknown is the phase space trajectory that evolves from this initial state and is required to reproduce the sequence of measured diagnostic signals.

The first part of the thesis, comprising Chapters 2 and 3, develops the physical, mathematical, and experimental foundations of the phase space reconstruction problem. It introduces phase space density dynamics, Hamiltonian flow, damping and diffusion, longitudinal synchrotron motion, wake fields, impedances, and the VFPE, together with the diagnostics used to observe the longitudinal beam dynamics. These foundations are required to describe the mechanisms governing the evolution of the longitudinal phase space density and its experimental observation.

The second part, comprising Chapters 4, 5, and 6, turns these foundations into a computational reconstruction framework. Chapter 4 develops the numerical solution of the VFPE using operator splitting, providing the evolution model required for the reconstruction. Chapter 5 formulates the reconstruction as a partial differential equation (PDE)-constrained optimization problem, while Chapter 6 presents the software framework that provides the modular simulation and algorithmic infrastructure needed to combine the forward model with optimization, diagnostics, and validation.

The final part, comprising Chapters 7 and 8, presents the verification and validation (V&V) of the reconstruction method and summarizes the resulting reconstruction capability. Chapter 7 first uses synthetic data with a known ground-truth phase space density to quantify reconstruction errors and assess algorithmic choices. The method is then tested with simulated EOSD measurements, where the diagnostic response is included through the impulse response of the EOSD setup. Finally, the algorithm is applied to experimental KARA data, using the EOSD system as the reconstruction input and the visible-light monitor (VLM) and terahertz (THz) detectors as independent validation observables. Chapter 8 summarizes the main results and discusses the relevance of the developed reconstruction method for future diagnostics and control of non-equilibrium beam dynamics.

2. Longitudinal phase space density dynamics

This chapter introduces the Vlasov–Fokker–Planck equation as the governing model for the evolution of phase space densities. The derivation starts from generic particle motion in canonical coordinates, where conservative dynamics are generated by Hamiltonian forces. From particle conservation, this deterministic motion leads to Liouville- or Vlasov-transport of the density.

The Hamiltonian description is then extended by including dissipative and stochastic effects. In electron storage rings, these effects are mainly associated with synchrotron radiation, where radiation damping acts as a systematic dissipative force and quantum excitation introduces stochastic fluctuations.

The discussion is restricted to relativistic electrons in circular storage rings. Single-pass dynamics and non-relativistic kinematics, such as those found in FELs and hadron colliders, are not considered. The main focus is the reduced two-dimensional longitudinal phase space, described by the longitudinal position with respect to the synchronous particle and the corresponding energy deviation.

The longitudinal dynamics are governed by conservative contributions from RF focusing and collective effects, together with dissipative and diffusive contributions arising from radiation effects. This model is required to describe micro-bunching regimes and provides the physical basis for phase space reconstruction under the non-equilibrium conditions addressed in the following chapters.

2.1. Particle density dynamics

A general dynamical system describes how the state of a particle evolves according to a governing law [32, p. 39]. At any instant, the particle state is represented by a point $\mathbf{x}$, and the set of all admissible particle states forms the state space $\mathcal{M}$. The evolution rule

$$f^{\Delta t} : \mathcal{M} \to \mathcal{M}, \tag{2.1}$$

shows that an initial state evolves to another state in the same state space $\mathcal{M}$ after the time interval Δt [32, p. 41]. Therefore, the evolution of any point in the state space can be written as

$$\mathbf{x}(t + \Delta t) = f^{\Delta t}\big(\mathbf{x}(t)\big). \tag{2.2}$$

The same evolution rule applies to sets of states: an initial region $\mathcal{M}_i \subseteq \mathcal{M}$ is mapped to the evolved region $f^{\Delta t}(\mathcal{M}_i)$. In this way, the dynamics can be viewed as both the motion of individual states and the deformation of regions in $\mathcal{M}$, as illustrated in Figure 2.1.

In classical dynamical systems, the general state space can take the structured form of a phase space. In Hamiltonian mechanics, this phase space is represented by canonical conjugate coordinates: generalized positions q and their corresponding generalized momenta p [33]. In this case, the evolution rule is specified by the Hamiltonian flow generated by the Hamiltonian $H(q, p, t)$. The Hamiltonian encodes

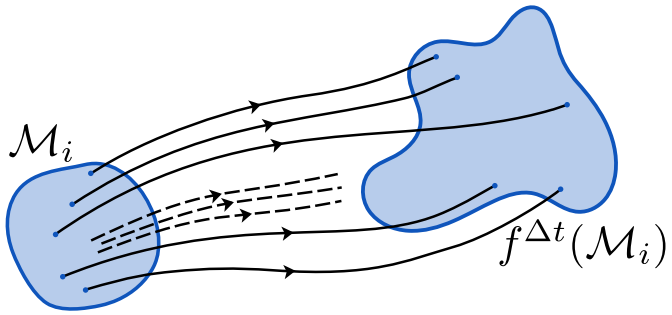


Figure 2.1.: Evolution of an initial region $\mathcal{M}_i \subseteq \mathcal{M}$ into $f^{\Delta t}(\mathcal{M}_i)$ under the map $f^{\Delta t}$ over the time step Δt.

the conservatives forces acting on the system and defines the equations of motion through Hamilton's canonical equations:

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}. \tag{2.3}$$

The analysis of dynamical systems benefits from a distribution-based description rather than from tracking individual trajectories [32, p. 371]. In this formulation, the density $\psi(q, p, t)$ represents the local particle concentration in phase space. The analysis of the phase space density extends the single particle picture to a statistical description of the bunch. In accelerators, the complete motion is described in a six-dimensional phase space. However, for many storage-ring applications, transverse and longitudinal dynamics can be treated separately.

A description of the longitudinal dynamics is obtained by studying how the phase space density evolves under the equations of motion. The analysis first considers conservative forces, where the transport is Hamiltonian. This provides the Liouville description used as the basis for the Vlasov equation before stochastic effects are included.

2.1.1. Hamiltonian phase space density evolution

This section derives the evolution of the phase space density for a Hamiltonian flow on a continuous manifold, following the approach of Wiedemann [34]. Each initial condition defines distinct flow lines therefore do not intersect. Under these assumptions, an initial phase space region $\mathcal{M}_i$ is continuously deformed into the mapped region $f^{\Delta t}(\mathcal{M}_i)$, as shown in Figure 2.2.

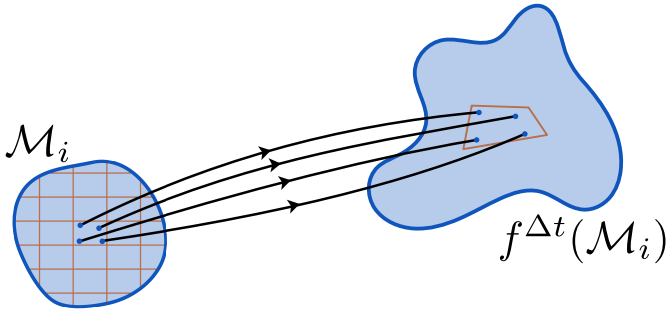


Figure 2.2.: Hamiltonian flow of a phase space region $\mathcal{M}_i$. A rectangular area element within the initial distribution is transported by the flow lines into a deformed quadrilateral within the final distribution $f^{\Delta t}(\mathcal{M}_i)$.

The derivation is presented in two dimensions only for geometric clarity, where the deformation of a phase space area can be visualized. Since the argument is based on local particle conservation, the resulting transport equation remains valid in arbitrary dimensions.

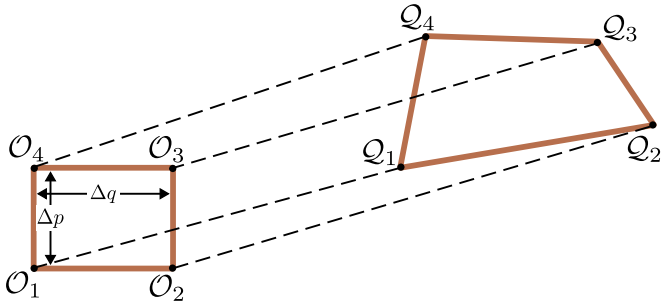


Figure 2.3.: Evolution of an infinitesimal rectangular region in the initial phase space $\mathcal{M}_i$ into a deformed quadrilateral in $f^{\Delta t}(\mathcal{M}_i)$.

Consider an infinitesimal rectangular region of phase space at time t, as shown in Figure 2.3, defined by the vertices $\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3, \mathcal{O}_4$, with side lengths Δq and Δp. The area of this initial region, at time t, is given by:

$$\mathcal{A}_t = \Delta q \Delta p. \tag{2.4}$$

The system is assumed to be closed, so the number of particles is conserved. The number of particles contained within the area element is defined as

$$N_t = \psi(q, p, t)\mathcal{A}_t. \tag{2.5}$$

As the system evolves over a small time interval Δt, the particles located at the vertices $\mathcal{O}_i$ move to new positions $\mathcal{Q}_i$ according to the equations of motion:

$$\frac{dq}{dt} = f^q(q, p, t), \quad \frac{dp}{dt} = f^p(q, p, t). \tag{2.6}$$

The mapping of each initial vertex to its new position is approximated by the first-order displacement in the two phase space coordinates. For a vertex

$$\mathcal{O}_i = \begin{pmatrix} q_i \\ p_i \end{pmatrix},$$

the transported vertex is

$$\mathcal{Q}_i = \begin{pmatrix} q_i + f^q(q_i, p_i, t)\,\Delta t \\ p_i + f^p(q_i, p_i, t)\,\Delta t \end{pmatrix}. \tag{2.7}$$

Equivalently, the first-order update is given by

$$q_i(t + \Delta t) = q_i(t) + f^q(q_i, p_i, t)\,\Delta t, \quad p_i(t + \Delta t) = p_i(t) + f^p(q_i, p_i, t)\,\Delta t. \tag{2.8}$$

Since the rates of change f^q and f^p generally vary across phase space, the different vertices of the initial rectangle are displaced by different amounts. The initially rectangular element therefore deforms into a quadrilateral with vertices $\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4$, as shown in Fig. 2.3.

The number of particles contained in this deformed region at time $t + \Delta t$ is given by

$$N_{t+\Delta t} = \psi(q + \Delta q, p + \Delta p, t + \Delta t)\mathcal{A}_{t+\Delta t}. \tag{2.9}$$

Since the system is conservative, particles are neither created nor destroyed. Thus, the number of particles within the volume must be conserved across the interval Δt:

$$\frac{dN}{dt} = 0 \implies N_t = N_{t+\Delta t}. \tag{2.10}$$

Equating the expressions for the particle number at both time steps yields the fundamental conservation law:

$$\psi(q, p, t)\mathcal{A}_t = \psi(q + \Delta q, p + \Delta p, t + \Delta t)\mathcal{A}_{t+\Delta t}. \tag{2.11}$$

To solve this equation, the area of the deformed quadrilateral, $\mathcal{A}_{t+\Delta t}$, must be determined. The full geometric derivation, based on decomposing the quadrilateral into two triangles and computing their areas with vector products, is given in Appendix A.2. The resulting first-order expression is

$$\mathcal{A}_{t+\Delta t} = \Delta q \Delta p \left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t\right] \tag{2.12}$$

Substituting the evolved area from Eq. (2.12) into the conservation law Eq. (2.11), and eliminating the common differential area factor, yields:

$$\psi(q, p, t) = \psi(q + \Delta q, p + \Delta p, t + \Delta t)\left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t\right]. \tag{2.13}$$

Expanding the displaced density $\psi(q+\Delta q, p+\Delta p, t+\Delta t)$ in a first-order Taylor series about (q, p, t):

$$\psi(q + \Delta q, p + \Delta p, t + \Delta t) \approx \psi(q, p, t) + \frac{\partial \psi}{\partial q} f^q(q, p, t)\Delta t + \frac{\partial \psi}{\partial p} f^p(q, p, t)\Delta t + \frac{\partial \psi}{\partial t}\Delta t. \tag{2.14}$$

Substituting Eq. (2.14) into Eq. (2.13) and dividing by Δt, and taking the continuum limit $\Delta t \to 0$ leads to:

$$\frac{\partial \psi}{\partial t} + f^q(q, p, t)\frac{\partial \psi}{\partial q} + f^p(q, p, t)\frac{\partial \psi}{\partial p} = -\psi(q, p, t)\left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right) \tag{2.15}$$

The intermediate algebra leading from Eq. (2.13) to Eq. (2.15) is given in Appendix A.3.

For a Hamiltonian system, the equations of motion are given by Hamilton's canonical equations:

$$f^q(q, p, t) = \frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad f^p(q, p, t) = \frac{dp}{dt} = -\frac{\partial H}{\partial q} \tag{2.16}$$

When these are substituted into the divergence term, it vanishes identically:

$$\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p} = \frac{\partial^2 H}{\partial q \partial p} - \frac{\partial^2 H}{\partial p \partial q} = 0 \tag{2.17}$$

Substituting Eq. (2.17) into Eq. (2.15) yields the Vlasov equation:

$$\frac{\partial \psi}{\partial t} + \frac{\partial H}{\partial p}\frac{\partial \psi}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial \psi}{\partial p} = 0 \tag{2.18}$$

This equation states that, for Hamiltonian motion, the phase space density is conserved along the trajectories generated by H. Moreover, inserting the vanishing divergence condition in Eq. (2.12) gives

$$\mathcal{A}_{t+\Delta t} = \mathcal{A}_t, \tag{2.19}$$

which confirms Liouville's theorem: in a conservative Hamiltonian system, the phase space volume is invariant under the motion.

2.1.2. Fokker–Planck phase space density evolution

The geometric derivation presented in Section 2.1.1 relies on the Hamiltonian nature of the flow, where the evolution could be visualized by tracking the boundary of a specific phase space volume forward in time. However, realistic physical systems are rarely governed by conservative dynamics alone [35]. Additional mechanisms can introduce systematic dissipation and small, rapidly varying perturbations whose cumulative effect can be modeled as random "kicks".

A classical example is Brownian motion, where a particle experiences an enormous number of microscopic collisions within a time interval that is still short on the scale of the macroscopic evolution [36, p. 32]. In this regime, the dynamics can be approximated as Markovian, meaning that the statistics of the next increment depend only on the current state. In accelerator physics, analogous stochastic perturbations arise from quantum excitation. Because these effects consist of many small, rapidly decorrelating momentum kicks, the resulting dynamics are mathematically similar to Brownian motion and can be treated statistically.

Over a short interval Δt, let $p = p(t)$ denote the momentum coordinate at the initial time and let $p' = p(t + \Delta t)$ denote its value at the next time step. The total momentum change is decomposed into a conservative contribution, and a Fokker–Planck contribution:

$$p' - p = \Delta p + \delta p(\Delta t), \tag{2.20}$$

where Δp is the deterministic increment generated by the Hamiltonian dynamics, and $\delta p(\Delta t)$ represents the additional non-Hamiltonian increment due to deterministic dissipative effects and random perturbations. This contribution is defined as

$$\delta p(\Delta t) = -\beta_d \, p \Delta t + B(\Delta t), \tag{2.21}$$

where the term $-\beta_d \, p$ represents a linear friction (damping) force. The term $B(\Delta t)$ denotes a stochastic momentum kick during the interval Δt and is specified only through its statistics, e.g. $\langle B(\Delta t) \rangle = 0$ and $\langle B(\Delta t)^2 \rangle = 2D \, \Delta t$ for constant diffusion coefficient D.

The change in the particle state over the interval Δt is therefore decomposed into a Hamiltonian contribution, a non-conservative dissipative contribution, and random fluctuations:

$$q' = q + \Delta q, \tag{2.22}$$
$$p' = p + \Delta p + \delta p(\Delta t). \tag{2.23}$$

Here, particles starting at (q, p) within an initial rectangular element may be transported inside or outside the corresponding target quadrilateral at $t + \Delta t$, as illustrated in Figure 2.4.

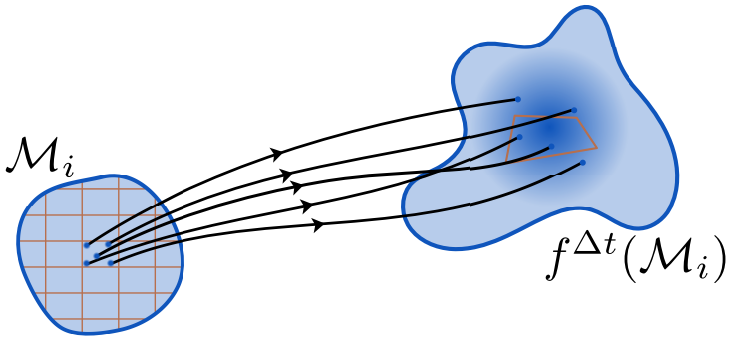


Figure 2.4.: Visualization of flow lines diffusing by Fokker–Planck effects.

In the presence of dissipative and stochastic effects, an initial state no longer determines a unique final state. Dissipative drift modifies the motion, while random fluctuations spread the possible final states. The density evolution is therefore described by fixing a target phase space element around (q', p') at time $t + \Delta t$ and accounts for all previous states (q, p) that can contribute to it through a transition probability.

Adopting the inverse viewpoint, instead of following one initial element to its final state, the analysis focuses on the target element $\mathcal{A}_{t+\Delta t}$ and identifies the source elements at time t that may contribute to it. This concept is illustrated in Figure 2.5 where the one-to-one correspondence is replaced by a distribution of transition probabilities, so the target element can receive contributions from multiple source elements.

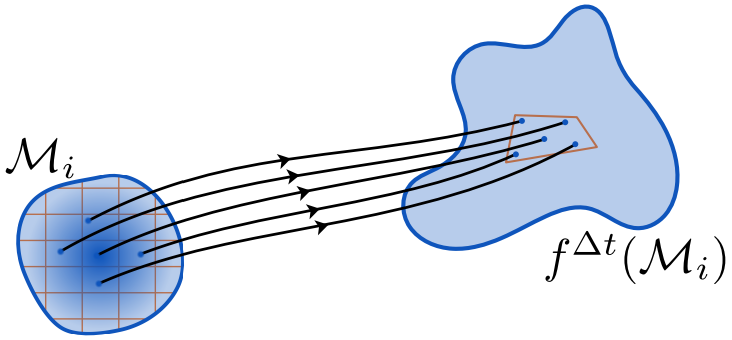


Figure 2.5.: Inverse view of Fokker–Planck transport: multiple source cells contribute probabilistically to one target cell.

For the drift–diffusion model in Eq. (2.21), and following Chandrasekhar [35, p. 22], the occurrence of different values of the stochastic increment $B(\Delta t)$ is described through the transition probability density

$$P(\delta p \mid p, \Delta t) = \frac{1}{\sqrt{4\pi D \Delta t}} \exp\left(-\frac{(\delta p + \beta_d \, p \, \Delta t)^2}{4 D \Delta t} \right), \tag{2.24}$$

where $\delta p + \beta_d p\,\Delta t = B(\Delta t)$.

To determine the number of particles ending up within the target area $\mathcal{A}_{t+\Delta t}$, it is necessary to take into consideration all source area elements $\mathcal{A}_t$ which, at time t, are located at a distance δp away from the final element [37, p. 39].

The total particle number at the new time step is obtained by integrating over all possible momentum displacements δp. Each value of δp selects a source coordinate $(q, p - \delta p)$, whose density contributes to the target element with the corresponding transition probability:

$$N_{t+\Delta t} \;=\; \mathcal{A}_t \int_{-\infty}^{+\infty} \psi\big(q,\, p-\delta p,\, t\big)\, P\big(\delta p \,|\, p-\delta p,\, \Delta t\big)\, d(\delta p). \tag{2.25}$$

Equating Eq. (2.9) with Eq. (2.25) gives the local particle conservation under stochastic momentum transport:

$$\psi(q+\Delta q, p+\Delta p, t+\Delta t)\,\mathcal{A}_{t+\Delta t} = \mathcal{A}_t \int_{-\infty}^{+\infty} \psi(q, p-\delta p, t)\, P(\delta p \mid p-\delta p, \Delta t)\, d(\delta p). \tag{2.26}$$

Since δp is small over a short time step, the source point lies close to p, allowing the integrand to be expanded locally about the target momentum. Rather than expanding ψ and P separately, it is convenient to treat their product as a single function:

$$F(p-\delta p) := \psi(q, p-\delta p, t)\, P(\delta p \mid p-\delta p, \Delta t). \tag{2.27}$$

The integrand can then be approximated by expanding $F(p-\delta p)$ in a Taylor series about p:

$$F(p-\delta p) = F(p) - \delta p\,\frac{\partial F}{\partial p} + \frac{1}{2}\delta p^2\,\frac{\partial^2 F}{\partial p^2} + \mathcal{O}(\delta p^3). \tag{2.28}$$

Writing F explicitly gives the compact "product" form often referred to as the Kramers-Moyal expansion:

$$\begin{aligned}\psi(q, p-\delta p, t)\, P(\delta p \mid p-\delta p, \Delta t) = \psi(q,p,t)\, P(\delta p \mid p, \Delta t) \\ - \delta p\,\frac{\partial}{\partial p}\Big[\psi(q,p,t)\, P(\delta p \mid p, \Delta t)\Big] \\ + \frac{1}{2}\delta p^2\,\frac{\partial^2}{\partial p^2}\Big[\psi(q,p,t)\, P(\delta p \mid p, \Delta t)\Big] + \mathcal{O}(\delta p^3).\end{aligned} \tag{2.29}$$

This expansion shifts all dependence on the state coordinate to derivatives evaluated at the target momentum p, while the integration remains over the variable δp. As a result, the integral reduces to a combination of the moments of the transition probability, which encode the systematic drift and the diffusive spread.

Substituting this expansion back into the integral of Eq. (2.26) allows the right-hand side to be evaluated term by term:

$$\int_{-\infty}^{+\infty} \psi(q, p-\delta p, t)\, P(\delta p \mid p-\delta p, \Delta t)\, d(\delta p) = \psi(q,p,t) \int_{-\infty}^{+\infty} P(\delta p \mid p, \Delta t)\, d(\delta p) - \frac{\partial}{\partial p}\left[\psi(q,p,t) \int_{-\infty}^{+\infty} \delta p\, P(\delta p \mid p, \Delta t)\, d(\delta p)\right] + \frac{1}{2}\frac{\partial^2}{\partial p^2}\left[\psi(q,p,t) \int_{-\infty}^{+\infty} \delta p^2\, P(\delta p \mid p, \Delta t)\, d(\delta p)\right]. \tag{2.30}$$

The bracketed integrals are the moments of the momentum increment. For the Gaussian kernel in Eq. (2.24), they are

1. Normalization:
$$\int_{-\infty}^{+\infty} P(\delta p \mid p, \Delta t)\, d(\delta p) = 1. \tag{2.31}$$

2. First Moment, $\langle \delta p \rangle$:
$$\int_{-\infty}^{+\infty} \delta p\, P(\delta p \mid p, \Delta t)\, d(\delta p) = -\beta_d\, p\, \Delta t. \tag{2.32}$$

3. Second Moment, $\langle \delta p^2 \rangle$:
$$\int_{-\infty}^{+\infty} \delta p^2\, P(\delta p \mid p, \Delta t)\, d(\delta p) = 2D\Delta t. \tag{2.33}$$

Substituting these moments into Eq. (2.30) gives

$$\int_{-\infty}^{+\infty} \psi(q, p-\delta p, t)\, P(\delta p \mid p-\delta p, \Delta t)\, d(\delta p) = \psi(q,p,t) + \beta_d \Delta t \frac{\partial}{\partial p}\left[p\, \psi(q,p,t)\right] + D\Delta t \frac{\partial^2 \psi(q,p,t)}{\partial p^2}. \tag{2.34}$$

Using the Taylor expansion of $\psi(q+\Delta q, p+\Delta p, t+\Delta t)$ from Eq. (2.14) together with Eq. (2.34), and subtracting $\psi(q,p,t)$ from both sides, gives

$$\left[\frac{\partial \psi}{\partial t} + f^q(q,p,t)\frac{\partial \psi}{\partial q} + f^p(q,p,t)\frac{\partial \psi}{\partial p} + \psi(q,p,t)\left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\right]\Delta t = \beta_d \Delta t \frac{\partial}{\partial p}\left[p\, \psi(q,p,t)\right] + D\Delta t \frac{\partial^2 \psi(q,p,t)}{\partial p^2} + \mathcal{O}(\Delta t^2). \tag{2.35}$$

Dividing Eq. (2.35) by Δt and taking the limit $\Delta t \to 0$ gives the VFPE with drift and diffusion in p:

$$\frac{\partial \psi}{\partial t} + f^q \frac{\partial \psi}{\partial q} + f^p \frac{\partial \psi}{\partial p} + \psi\left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right) = \beta_d \frac{\partial}{\partial p}(p\, \psi) + D\, \frac{\partial^2 \psi}{\partial p^2}. \tag{2.36}$$

When the conservative part is Hamiltonian, Eq. (2.16) reduces the left-hand side to

$$\frac{\partial \psi}{\partial t} + \frac{\partial H}{\partial p}\frac{\partial \psi}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial \psi}{\partial p} = \beta_d \frac{\partial}{\partial p}(p\,\psi) + D\,\frac{\partial^2 \psi}{\partial p^2}. \tag{2.37}$$

This section derived the VFPE by combining conservative Hamiltonian transport with non-conservative and stochastic Fokker–Planck contributions.

The resulting equation provides the density evolution model used in this thesis: conservative motion is represented by the Vlasov contribution, while radiation damping and quantum excitation are represented by drift and diffusion in momentum.

The next section connects this abstract phase space formulation to beam dynamics in an electron storage ring, where the relevant variables, forces, and reference trajectory are introduced.

2.2. Electromagnetic forces in the co-moving reference frame

The formal VFPE introduced in the previous section defines the evolution of a particle density in a canonical phase space. To apply this formulation to electrons in a storage ring, the canonical variables must be connected to the coordinates frames, forces, and reference trajectory used in accelerator physics.

The fundamental description of particle beams in a synchrotron relies on a reference frame that simplifies the equations of motion. Instead of tracking absolute laboratory trajectories, a curvilinear coordinate system is attached to the reference particle moving along the reference orbit. With this coordinate system, the motion of the beam is described through small spatial and momentum, or energy, deviations from the reference particle.

The evolution of the phase space coordinates is governed by the Lorentz force, which describes the interaction of the particles with the accelerator's electromagnetic fields. These fields serve distinct roles: magnetic fields primarily guide particles along the reference trajectory and provide transverse confinement, whereas electric fields are responsible for changing the particle energy and sustaining acceleration.

2.2.1. Reference orbit and co-moving coordinate system

The transverse and longitudinal motion of particles in a storage ring are most conveniently described in a co-moving reference frame defined relative to the ideal design orbit $\mathbf{r}_0(s)$, where s is the arc length along the reference path. In the longitudinal plane, the reference particle is the synchronous particle, whose arrival is synchronized with the RF system at a fixed phase.

The moving frame is mathematically described as the Frenet–Serret frame, also referred to as the tangent–normal–binormal (TNB) basis, which consists of three orthonormal unit vectors: the Tangent ($\hat{\mathrm{T}}$), the Normal ($\hat{\mathrm{N}}$), and the Binormal ($\hat{\mathrm{B}}$).

The basis vectors are defined by the geometry of the reference curve parameterized by the arc length:

$$\hat{\mathrm{T}}(s) = \frac{d\mathbf{r}_0}{ds}, \quad \hat{\mathrm{N}}(s) = \frac{\frac{d\hat{\mathrm{T}}}{ds}}{\left|\frac{d\hat{\mathrm{T}}}{ds}\right|}, \quad \hat{\mathrm{B}}(s) = \hat{\mathrm{T}}(s) \times \hat{\mathrm{N}}(s). \tag{2.38}$$

In accelerator physics, the coordinate axes are typically chosen as a local right-handed triad $(\hat{z}, \hat{x}, \hat{y})$ attached to the design orbit, where $\hat{z}$ is tangential to the orbit, $\hat{x}$ is the horizontal (radial) direction, and $\hat{y}$ is vertical. The principal normal $\hat{\mathrm{N}}$, however, points toward the local center of curvature (inward), whereas the horizontal coordinate $\hat{x}$ is commonly taken positive outward:

$$\hat{z} \equiv \hat{\mathrm{T}}, \qquad \hat{y} \equiv \hat{\mathrm{B}}, \qquad \hat{x} \equiv -\hat{\mathrm{N}}. \tag{2.39}$$

The evolution of the TNB basis along s is governed by the Frenet–Serret relations [34, p. 87]:

$$\begin{aligned} \frac{d\hat{z}}{ds} &= -\kappa(s)\,\hat{x}, \\ \frac{d\hat{x}}{ds} &= \kappa(s)\,\hat{z} - \tau(s)\,\hat{y}, \\ \frac{d\hat{y}}{ds} &= \tau(s)\,\hat{x}, \end{aligned} \tag{2.40}$$

which introduce the two geometric invariants of the reference trajectory: the curvature $\kappa(s)$ and the torsion $\tau(s)$:

$$\begin{aligned} \kappa(s) &:= \left\| \frac{d\hat{\mathrm{T}}}{ds} \right\|, \\ \tau(s) &:= -\frac{d\hat{x}}{ds} \cdot \hat{y} = \frac{d\hat{y}}{ds} \cdot \hat{x}. \end{aligned} \tag{2.41}$$

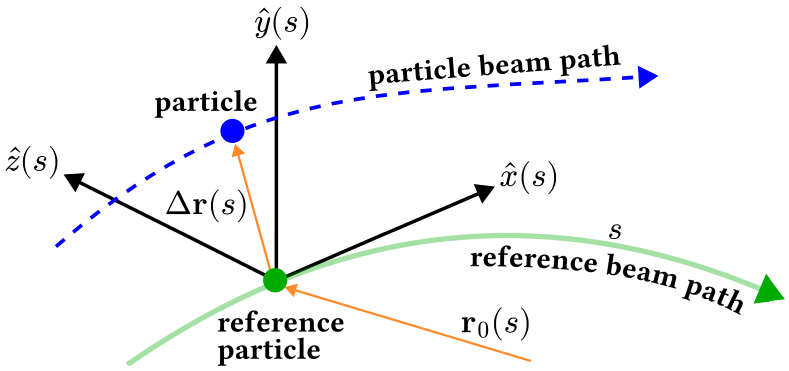


Figure 2.6.: Particle coordinate reference frame.

A particle close to the design orbit is described by its offsets in the local co-moving basis, as illustrated in Figure 2.6. In general, the Frenet–Serret (TNB) frame provides a three-dimensional description with respect to each point of the reference curve. In the present formulation, however, the longitudinal coordinate s is chosen as the independent variable and labels the position along the design orbit. The particle position $\mathbf{r}(s)$ is therefore defined at the same longitudinal location s as the reference point $\mathbf{r}_0(s)$:

$$\mathbf{r}(s) = \mathbf{r}_0(s) + \Delta\mathbf{r}(s). \tag{2.42}$$

Accordingly, $\Delta\mathbf{r}(s)$ is the displacement from $\mathbf{r}_0(s)$ within the transverse plane orthogonal to $\hat{z}(s)$ and it is written purely in terms of the transverse coordinates (x, y):

$$\Delta\mathbf{r}(s) = x(s)\,\hat{x}(s) + y(s)\,\hat{y}(s), \tag{2.43}$$

where $x(s)$ and $y(s)$ are the horizontal and vertical offsets.

The longitudinal dynamics is not obtained from these transverse kinematic relations, but from the arrival-time or RF-phase deviation with respect to the synchronous particle and the corresponding relative energy deviation, which follow from the time-of-flight dependence and the RF energy exchange.

2.2.2. Charged particle dynamics under the Lorentz force

The dynamics of a charged particle in an accelerator are determined by the Lorentz force. For a particle of charge q, position $\mathbf{r}(t)$, and velocity $\mathbf{v}(t)$, the force exerted by the external electromagnetic fields is

$$\mathbf{F} = \mathrm{q}\left[\mathbf{E}(\mathbf{r},t) + \mathbf{v}(t) \times \mathbf{B}(\mathbf{r},t)\right], \tag{2.44}$$

where $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$ are the electric and magnetic fields evaluated at the particle position and time, t.

From Newton's second law, the force acting on a moving particle is defined as the rate of change of its momentum:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}. \tag{2.45}$$

For a relativistic particle with rest mass m_0, the momentum is given by

$$\mathbf{p}(t) = \gamma m_0 \mathbf{v}(t), \tag{2.46}$$

where γ is the Lorentz factor. Combining Eq. (2.45) and Eq. (2.46) with Eq. (2.44) yields the fundamental equation of motion used throughout accelerator physics:

$$\frac{d}{dt}(\gamma m_0 \mathbf{v}(t)) = \mathrm{q}\left[\mathbf{E}(\mathbf{r},t) + \mathbf{v}(t) \times \mathbf{B}(\mathbf{r},t)\right]. \tag{2.47}$$

Equation (2.47) shows that the Lorentz force governs the full three-dimensional evolution of the particle momentum and therefore determines the particle trajectory. The magnetic contribution mainly produces acceleration perpendicular to the direction of motion, changing the trajectory without changing the particle energy, whereas the electric field contribution can accelerate the particle along its motion and thereby change its energy.

2.3. Longitudinal beam dynamics

Longitudinal beam dynamics describes the motion of particles relative to the synchronous particle. The longitudinal state is described by two quantities: the arrival-time deviation with respect to the synchronous particle, equivalently represented as an RF-phase or longitudinal-position deviation, and the corresponding energy deviation.

The synchronous particle arrives at the RF cavity with the same phase on every turn and gains, on average, the energy needed to compensate the loss due to synchrotron radiation. Particles that are not synchronous arrive earlier or later than this reference particle and carry a slightly different energy.

Energy and momentum are intrinsically coupled through the relativistic energy-momentum relation:

$$E^2 = p^2c^2 + m_0^2c^4. \tag{2.48}$$

Differentiating with respect to time gives

$$\frac{dE}{dt} = \frac{pc^2}{E}\frac{dp}{dt}. \tag{2.49}$$

The factor pc^2/E can be expressed in terms of the particle velocity using the relativistic definitions of momentum and total energy:

$$p = \gamma m_0 v, \qquad E = \gamma m_0 c^2. \tag{2.50}$$

Therefore,

$$\frac{pc^2}{E} = \frac{\gamma m_0 v\, c^2}{\gamma m_0 c^2} = v = \beta c, \tag{2.51}$$

where $\beta := v/c$. Substitution into Eq. (2.49) gives

$$\frac{dE}{dt} = \beta c \frac{dp}{dt}, \tag{2.52}$$

or, equivalently, in differential form:

$$dE = \beta c\, dp. \tag{2.53}$$

The particle energy changes only through the component of the applied force along the direction of motion. For a particle with velocity $\mathbf{v}(t)$ and total force $\mathbf{F}$, the corresponding energy rate is

$$\frac{dE}{dt} = \mathbf{F} \cdot \mathbf{v}. \tag{2.54}$$

Substituting the expression for the Lorentz force from Eq. (2.44):

$$\frac{dE}{dt} = \mathbf{F} \cdot \mathbf{v} = \mathrm{q}\,\mathbf{E}(\mathbf{r},t) \cdot \mathbf{v}(t) + \mathrm{q}\left(\mathbf{v}(t) \times \mathbf{B}(\mathbf{r},t)\right) \cdot \mathbf{v}(t). \tag{2.55}$$

The magnetic contribution vanishes identically because $\mathbf{v} \times \mathbf{B}$ is orthogonal to $\mathbf{v}$, so that

$$\left(\mathbf{v} \times \mathbf{B}\right) \cdot \mathbf{v} = 0, \tag{2.56}$$

and the energy rate of change reduces to

$$\frac{dE}{dt} = \mathrm{q}\,\mathbf{v}(t) \cdot \mathbf{E}(\mathbf{r},t)\,. \tag{2.57}$$

Equation (2.57) shows that only the electric-field component parallel to the particle trajectory contributes to the energy change.

In the moving coordinate system, this corresponds to the component along z.

$$\frac{dE}{dt} = \mathrm{q} v_z E_z. \tag{2.58}$$

To connect this instantaneous rate to the macroscopic energy gain:

$$dE = \mathrm{q} v_z E_z dt. \tag{2.59}$$

Change variables from time (t) to position (z). Since $dz = v_z dt$, and substituting this into the equation:

$$dE = \mathrm{q} E_z dz. \tag{2.60}$$

The energy change is obtained by integrating along the reference trajectory, where the local coordinate z coincides with the path coordinate s, so that $dz = ds$. Therefore:

$$\int dE = \mathrm{q} \int E_z dz \quad \Rightarrow \quad \Delta E = \mathrm{q} \underbrace{\left(\int_{s_1}^{s_2} E_z(s)\, ds \right)}_{V}. \tag{2.61}$$

The integral on the right side is identified as the longitudinal potential:

$$V(t) = \int_{s_1}^{s_2} E_z(s,t)\, ds. \tag{2.62}$$

Consequently, for a particle traversing a determined path at a given instance t, the total energy change is directly proportional to the potential:

$$\Delta E = \mathrm{q} V(t). \tag{2.63}$$

In this description, the time t denotes the particle arrival time at the RF cavity. During the finite crossing time, the particle samples the oscillating field at different phases. This effect is represented by the transit-time factor, which relates the field amplitude and cavity length to the effective cavity voltage. In the thin-gap approximation used here, it is absorbed into the effective voltage, so the cavity is modeled by a single voltage evaluated at the particle arrival phase.

In an electron storage ring, the beam loses energy each turn mainly through synchrotron radiation [34]. This loss is compensated in the RF cavities, where an oscillating voltage changes the energy of the passing particles. The synchronization between the RF field and the particle arrival time is the basis of longitudinal beam dynamics.

2.3.1. Synchrotron motion and momentum compaction factor

The longitudinal dynamics is governed by the balance between radiative energy loss, RF energy gain, and the arrival phase of the particles. The mean energy loss per turn U_0 must be compensated to maintain the reference energy E_0. This is provided by the RF system, whose phase-dependent voltage restores the lost energy and produces the stabilizing mechanism of phase focusing.

Using the thin-gap model introduced above, the RF cavity is described by a sinusoidal voltage evaluated at the particle arrival phase $\phi(t)$. The corresponding energy kick is

$$\Delta E = eV_{RF}\sin(\phi(t)), \tag{2.64}$$

where e denotes the elementary charge magnitude, used instead of q to avoid confusion with the canonical coordinate. The sign of the electron charge is absorbed into the RF-phase convention. The phase is defined as

$$\phi(t) = 2\pi f_{\mathrm{RF}}\, t = \omega_{\mathrm{RF}}\, t, \tag{2.65}$$

where f_{RF} and ω_{RF} are the frequency and the angular frequency of the RF field, respectively.

For a particle to experience a consistent energy gain turn after turn, the RF field must return to the same phase by the time the particle completes one revolution. This implies that the frequency of the RF field must be an exact integer multiple of the revolution frequency f_{rev}. This integer is the harmonic number h:

$$f_{\mathrm{RF}} = h \cdot f_{\mathrm{rev}} = h \cdot \frac{\beta c}{L_0}, \tag{2.66}$$

where L_0 is the length of the design orbit and βc is the particle velocity along the orbit. The integer h is the harmonic number and gives the number of RF periods, or equivalently the number of RF buckets, around one revolution.

For the synchronous particle to maintain a constant energy on average, the energy gained from the cavity must exactly balance the energy loss per turn due to SR, U_0. The amount of energy restored is determined by the synchronous phase ϕ_s:

$$U_0 = eV_{\mathrm{RF}}\sin(\phi_s). \tag{2.67}$$

The synchronous phase ϕ_s is chosen from the two RF phases that satisfy Eq. (2.67). Stability requires the phase that provides a restoring correction, but which of the two phases is stable depends on the sign of the momentum. For the usual case of positive momentum compaction, a higher-energy particle travels a longer path, arrives later at the cavity, and must gain less energy than the synchronous particle; a lower-energy particle arrives earlier and must gain more. If the sign of this dependence is reversed, the stable and unstable synchronous phases are interchanged.

This coupling between longitudinal position and momentum is determined by the magnetic lattice configuration. It is quantified by the momentum compaction factor α_{c}, defined as the relative change in closed-orbit length per relative momentum deviation:

$$\alpha_c \;:=\; \frac{\Delta L/L_0}{\Delta p/p_0}. \tag{2.68}$$

For small momentum offsets it is convenient to introduce the relative momentum deviation

$$\delta \;:=\; \frac{\Delta p}{p_0}, \tag{2.69}$$

so that

$$\frac{\Delta L}{L_0} = \alpha_c\,\delta. \tag{2.70}$$

The revolution period can be written as

$$T = \frac{L}{\beta c}.$$

Consider a reference, synchronous, particle with (L_0, β_0) and a nearby particle with small deviations $\Delta L := L - L_0$ and $\Delta\beta := \beta - \beta_0$. Using a first-order expansion:

$$\frac{\Delta T}{T_0} = \frac{T - T_0}{T_0} = \frac{L/(\beta c) - L_0/(\beta_0 c)}{L_0/(\beta_0 c)} \simeq \frac{\Delta L}{L_0} - \frac{\Delta\beta}{\beta_0},$$

For ultra-relativistic electrons $\beta_0 \simeq 1$ and $\Delta\beta/\beta_0$ is typically negligible compared to $\Delta L/L_0$, so that

$$\frac{\Delta T}{T_0} \;\simeq\; \frac{\Delta L}{L_0} \;=\; \alpha_c\,\delta. \tag{2.71}$$

If the velocity variation is retained, it is customary to introduce the phase-slip factor

$$\eta \;:=\; \frac{1}{\delta}\frac{\Delta T}{T_0} \;=\; \alpha_c - \frac{1}{\gamma^2}, \tag{2.72}$$

which reduces to $\eta \simeq \alpha_c$ in the highly relativistic limit $\gamma \to \infty$.

2.3.2. Energy and phase deviation equations

The longitudinal single particle dynamics is described in terms of deviations with respect to the synchronous particle at discrete turns. The energy deviation at turn n is defined as

$$\varepsilon_n \;:=\; E_n - E_0, \tag{2.73}$$

and the phase deviation as

$$\varphi_n \;:=\; \phi_n - \phi_s. \tag{2.74}$$

The index n labels successive passages through the RF cavity, i.e. successive turns of the storage ring.

The turn-by-turn energy update of a particle is given by the RF gain minus the mean radiative loss:

$$E_{n+1} = E_n + eV_{\mathrm{RF}}\sin(\phi_n) - U_0. \tag{2.75}$$

Subtracting the reference energy E_0 from both sides and using the definition (2.73) yields

$$\varepsilon_{n+1} = \varepsilon_n + eV_{\mathrm{RF}}\sin(\phi_n) - U_0. \tag{2.76}$$

Substituting the synchronous energy balance from Eq. (2.67) and the particle phase from Eq. (2.74) gives the turn-by-turn evolution of the energy deviation:

$$\varepsilon_{n+1} = \varepsilon_n + eV_{\mathrm{RF}}[\sin(\phi_s + \varphi_n) - \sin(\phi_s)]. \tag{2.77}$$

Since the cavity voltage oscillates at frequency f_{RF}, the RF phase advances linearly in time with ω_{RF}. For a particle crossing the cavity at times t_n and t_{n+1}, the RF-phase advance over one turn is

$$\phi_{n+1} - \phi_n = \omega_{\mathrm{RF}}(t_{n+1} - t_n) = \omega_{\mathrm{RF}} T_n, \tag{2.78}$$

where $T_n := t_{n+1} - t_n$ is the particle's revolution period.

For the synchronous particle, the corresponding phase advance is

$$\Delta\phi_s = \omega_{\mathrm{RF}} T_0 = 2\pi h. \tag{2.79}$$

This advance corresponds to an integer number of RF periods and is therefore equivalent to zero in RF-phase. Hence, the synchronous arrival phase ϕ_s remains constant from turn to turn.

The phase deviation φ_n is the difference between the particle RF phase and the synchronous RF-phase. Therefore, its change from one turn to the next is obtained by subtracting the synchronous phase advance from the particle phase advance:

$$\begin{aligned} \varphi_{n+1} - \varphi_n &= (\phi_{n+1} - \phi_n) - \Delta\phi_s \\ &= \omega_{\mathrm{RF}}(T_n - T_0). \end{aligned} \tag{2.80}$$

Using the slip-factor relation derived previously:

$$\frac{T - T_0}{T_0} = \eta\,\delta, \tag{2.81}$$

Eq. (2.80) becomes

$$\varphi_{n+1} - \varphi_n = \omega_{\mathrm{RF}} T_0\,\eta\,\delta, \tag{2.82}$$

so that the phase evolution is directly coupled to the relative momentum deviation δ.

Using the energy–momentum relation in Eq. (2.53) for small deviations around the reference particle, the energy deviation can be identified as $\varepsilon \approx \Delta E$. Together with the relative momentum deviation defined in Eq. (2.69), this gives

$$\varepsilon = \beta c\,\Delta p = \beta c\, p_0\,\delta, \tag{2.83}$$

and therefore

$$\delta = \frac{\varepsilon}{\beta c\, p_0}. \tag{2.84}$$

Using the relativistic identity evaluated at the reference particle, one obtains $p_0 = \beta E_0/c$. Substitution yields

$$\delta = \frac{1}{\beta^2 E_0}\,\varepsilon. \tag{2.85}$$

Combining Eqs. (2.80)–(2.85) yields

$$\varphi_{n+1} - \varphi_n = \omega_{\mathrm{RF}} T_0\,\eta\,\delta = 2\pi h\,\eta\,\frac{\varepsilon}{\beta^2 E_0}. \tag{2.86}$$

Therefore, the phase deviation evolves turn by turn according to

$$\varphi_{n+1} = \varphi_n + \frac{2\pi h\,\eta}{\beta^2 E_0}\,\varepsilon. \tag{2.87}$$

For the subsequent derivation in continuous time, it is assumed that φ and ε vary slowly from turn to turn. Under this smooth-variation approximation, the discrete updates are converted into differential equations by replacing $\Delta/T_0 \to d/dt$. From Eq. (2.87)

$$\dot{\varphi} = \frac{\varphi_{n+1} - \varphi_n}{T_0} = \frac{2\pi h\eta}{\beta^2 E_0 T_0}\varepsilon = \frac{2\pi h\eta f_{\mathrm{rev}}}{\beta^2 E_0}\varepsilon, \tag{2.88}$$

and from Eq. (2.77)

$$\dot{\varepsilon} = \frac{\varepsilon_{n+1} - \varepsilon_n}{T_0} = \frac{eV_{\mathrm{RF}}}{T_0}[\sin(\phi_s + \varphi) - \sin(\phi_s)] \approx f_{\mathrm{rev}} eV_{\mathrm{RF}} \cos(\phi_s)\, \varphi, \tag{2.89}$$

where the linearization $\sin(\phi_s + \varphi) - \sin(\phi_s) \approx \varphi \cos(\phi_s)$ is applied for small phase deviations. Finally, the coupled equations of motion are:

$$\begin{aligned} \frac{d\varphi}{dt} &= \frac{2\pi h\eta f_{\mathrm{rev}}}{\beta^2 E_0}\varepsilon, \\ \frac{d\varepsilon}{dt} &= e\,\hat{V} f_{\mathrm{rev}}\, \varphi, \end{aligned} \tag{2.90}$$

where $\hat{V} = V_{\mathrm{RF}} \cos(\phi_s)$ denotes the effective linear RF-voltage slope around the synchronous phase.

Taking the time derivative of Eq. (2.88) and substituting $\dot{\varepsilon}$ from Eq. (2.89) yields the second-order differential equation describing the synchrotron motion:

$$\ddot{\varphi} = \frac{2\pi h\eta f_{\mathrm{rev}}}{\beta^2 E_0}\dot{\varepsilon} = \left(\frac{2\pi h\eta f_{\mathrm{rev}}^2 e\hat{V}}{\beta^2 E_0}\right)\varphi. \tag{2.91}$$

This takes the standard form of a harmonic oscillator:

$$\ddot{\varphi} + \Omega_s^2 \varphi = 0, \tag{2.92}$$

where $\Omega_s = 2\pi f_s$ is the angular synchrotron frequency defined by:

$$f_s = f_{\mathrm{rev}} \sqrt{\frac{h\eta e\hat{V}}{2\pi\beta^2 E_0}}. \tag{2.93}$$

Since the energy and phase are linearly coupled, the energy deviation ε satisfies an identical differential equation:

$$\ddot{\varepsilon} + \Omega_s^2 \varepsilon = 0. \tag{2.94}$$

The coupled equations derived above show that, in the absence of dissipative effects, the longitudinal variables execute intrinsic oscillatory motion about the synchronous fixed point, in direct analogy to betatron oscillations in transverse dynamics.

However, in the continuous formulation above, the energy loss per turn was treated as a constant U_0. In reality, the SR loss depends on the particle energy, so an off-energy particle does not lose exactly the same amount of energy per turn. Close to the synchronous energy E_0, this dependence can be linearized:

$$U(E) \;\approx\; U(E_0) + \left.\frac{dU}{dE}\right|_{E_0} \varepsilon, \tag{2.95}$$

The derivative quantifies how sensitively the radiative loss changes with energy, particles with $\varepsilon > 0$ radiate more strongly and therefore lose energy faster than the synchronous particle, while particles with $\varepsilon < 0$ radiate less and catch up. This provides a dissipative restoring action in the energy coordinate. This simplified model neglects damping-partition effects; a complete treatment is given in [38, p. 110].

Starting from the continuous energy deviation equation and inserting Eq. (2.95) gives

$$\dot{\varepsilon} = f_{\mathrm{rev}}\Big(eV_{\mathrm{RF}}\sin(\phi_s+\varphi) - U_0 - \left.\frac{dU}{dE}\right|_{E_0}\varepsilon\Big). \tag{2.96}$$

Applying the small-angle approximation, the evolution of the energy deviation is linearized to:

$$\dot{\varepsilon} \simeq f_{\mathrm{rev}}\Big(eV_{\mathrm{RF}}\cos\phi_s\,\varphi - \left.\frac{dU}{dE}\right|_{E_0}\varepsilon\Big). \tag{2.97}$$

Substituting $\dot{\varepsilon}$ from Eq. (2.97) into Eq. (2.91), the equation can be written entirely in terms of the phase deviation φ. The term proportional to the energy deviation is expressed through $\dot{\varphi}$, leading to the standard form of a damped harmonic oscillator:

$$\ddot{\varphi} + 2\alpha_z\dot{\varphi} + \Omega_s^2\varphi = 0, \tag{2.98}$$

where the longitudinal damping rate is defined as

$$2\alpha_z := \frac{1}{\tau_d} = \frac{f_{\mathrm{rev}}}{2}\left.\frac{dU}{dE}\right|_{E_0}. \tag{2.99}$$

This result confirms that the energy dependence in radiation losses produces an effective friction term in the phase dynamics, proportional to $\dot{\varphi}$.

An analogous damped oscillator equation follows for ε, since ε and φ remain linearly coupled in the harmonic approximation:

$$\ddot{\varepsilon} + 2\alpha_z\dot{\varepsilon} + \Omega_s^2\varepsilon = 0. \tag{2.100}$$

2.3.3. Wake fields and longitudinal impedances

The longitudinal dynamics presented in previous sections has been governed exclusively by externally applied electromagnetic fields, such as those generated by RF cavities and magnets. In addition, the beam itself acts as a source of electromagnetic fields. The interaction of these self-generated fields with the particles gives rise to a broad class of phenomena known as collective effects, which can significantly affect both the transverse and longitudinal beam dynamics.

Self-fields are commonly separated into space-charge effects and wake-field effects [34]. Space-charge effects arise from the direct electromagnetic interaction within the bunch, including image charges in an ideal conducting beam pipe. Wake fields instead result from non-ideal beam–environment interactions, such as finite wall conductivity, resonant structures, cavities, collimators, and geometric discontinuities. These fields remain after the source particle has passed and can influence trailing particles, either within the same bunch or in later bunches, producing energy changes, potential-well distortion, and longitudinal instabilities. The following discussion focuses on intra-bunch wake fields and their frequency-domain representation by longitudinal impedances.

The interaction of a particle beam with its environment begins with its own electromagnetic field. As illustrated in Figure 2.7, a static point charge generates a radial Coulomb field extending equally in all directions. However, for ultra-relativistic particles ($\gamma \gg 1$), this field is transformed by Lorentz contraction. It becomes compressed into a thin, pancake-like disc perpendicular to the direction of motion over the vertical and horizontal planes, with an opening angle on the order of $1/\gamma$.

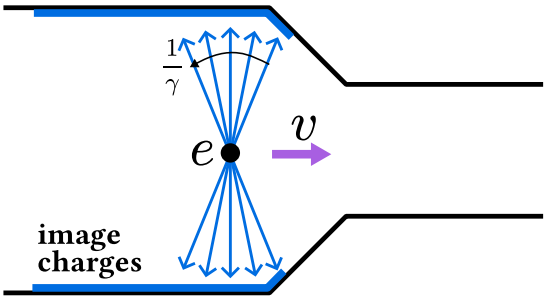


Figure 2.7.: Schematic representation of the electric-field lines of an ultrarelativistic charged particle inside a beam pipe. The field is compressed into a transverse disc (pancake shape) due to Lorentz contraction. The image charges are represented by the thin blue shading on the pipe walls.

In free space, the fields of a particle moving in a straight line co-propagate with it. Inside an accelerator, the conducting vacuum chamber imposes boundary conditions through image charges and currents. For an ideal smooth, perfectly conducting pipe these fields follow the beam, but finite wall resistivity or geometric discontinuities, such as cavities, transitions, or collimators, disturb this propagation and leave induced fields behind the source particle. These trailing induced fields are referred to as wake fields.

To connect wake fields with the longitudinal dynamics, the energy change produced by the total longitudinal wake field must be determined. This is described by the wake potential, which represents the accumulated effect of the longitudinal wake field $E_{\parallel}$, generated by the bunch through its interaction with the accelerator environment, on a test particle located at longitudinal position z within the bunch. The wake potential is defined as

$$V_c(z,t) = \int_{L_0} E_{\parallel}(z,s)\ ds, \tag{2.101}$$

where s is the coordinate along the reference trajectory over an accelerator length L_0, and z labels the test particle position within the bunch. The variable t denotes the revolution time at which the wake field is evaluated.

Directly computing the full wake field $E_{\parallel}$ for every bunch configuration is generally impractical. Instead, the integrated response of the accelerator environment to a target charge is described by the longitudinal wake function $W(z)$. More precisely, $W(z)$ is the longitudinal voltage per unit of charge induced at the position of a target particle, after the wake field has been integrated along the section. Consequently, $W(z)$ carries units of $\mathrm{V/C}$ and is defined for a specific path length L.

Let $\rho(z,t)$ denote the normalized longitudinal bunch profile at time t, where z is the longitudinal position within the bunch. The normalization is chosen such that

$$\int_{-\infty}^{\infty} \rho(z,t)\, dz = 1. \tag{2.102}$$

The wake potential follows from the superposition of the wakes generated by all particles in the bunch. Writing the total bunch charge as $Q = eN_b$, where e is the elementary charge and N_b is the number of electrons in the bunch, the collective wake potential is

$$V_{\mathrm{c}}(z,t) = eN_b \int_{-\infty}^{\infty} W(z-z')\,\rho(z',t)\,dz'. \tag{2.103}$$

Causality restricts the convolution in Eq. (2.103) to source particles located ahead of the test particle. Since positive z is defined toward the bunch tail, the integration therefore extends from the bunch head to the test particle position z.

The resulting energy change experienced by a particle at longitudinal position z within the bunch is obtained by summing the wake contributions from all source particles:

$$\Delta E(z,t) = e^2 N_b \int_{-\infty}^{\infty} W(z-z')\,\rho(z',t)\,dz'. \tag{2.104}$$

Unlike the RF voltage, which is approximately linear across the bunch length, for short bunches, the wake-induced potential introduces energy variations and is generally non-linear. Consequently, it can distort the equilibrium bunch profile, modify the longitudinal focusing, and generate fine longitudinal structures. Under suitable conditions, these self-induced distortions can grow and ultimately lead to the micro-bunching instability.

Although the wake-function formalism provides a direct physical interpretation of the beam–environment interaction, the frequency-domain representation is often better suited for analysis and computation. Accelerator components are naturally characterized by their frequency response, and the convolution in Eq. (2.103) becomes a multiplication between the beam-current spectrum at time t, $I_b^t(\omega)$, and the longitudinal impedance, in analogy with an electrical circuit. The wake potential is then obtained by applying the inverse Fourier transform to this product:

$$V_{\mathrm{c}}(z,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} I_b^t(\omega)\, Z_{\parallel}(\omega)\, e^{i\omega z/v}\, d\omega, \tag{2.105}$$

where the beam current spectrum $I_b(\omega)$ is obtained from the longitudinal charge distribution of the bunch. For a bunch containing N_b electrons and a normalized profile $\rho(z,t)$, the beam current is

$$I_b(z,t) = eN_b\, v\, \rho(z,t), \tag{2.106}$$

where v is the particle velocity. Its frequency spectrum is given by

$$I_b^t(\omega) = \int_{-\infty}^{\infty} I_b(z,t)\, e^{-i\omega z/v}\, \frac{dz}{v} = eN_b \int_{-\infty}^{\infty} \rho(z,t)\, e^{-i\omega z/v}\, dz. \tag{2.107}$$

The longitudinal coupling impedance $Z_{\parallel}(\omega)$ is defined as the Fourier transform of the wake function:

$$\begin{aligned} Z_{\parallel}(\omega) &= \mathcal{F}[W(z)], \\ &= \int_{-\infty}^{\infty} W(z)\, e^{-i\omega z/v}\, \frac{dz}{v}. \end{aligned} \tag{2.108}$$

Therefore, the wake potential expressed in terms of the longitudinal impedance and the Fourier transform of the bunch profile at time t, $\tilde{\rho}^t(\omega)$, is

$$V_{\mathrm{c}}(z,t) = \frac{eN_b}{2\pi} \int_{-\infty}^{\infty} \tilde{\rho}^t(\omega)\, Z_{\parallel}(\omega)\, e^{i\omega z/v}\, d\omega, \tag{2.109}$$

Equation (2.109) provides a practical evaluation of the wake potential defined in Eq. (2.101). The impedance characterizes the accelerator environment and is fixed for a given machine configuration, while the time dependence enters through the bunch spectrum. Therefore, the wake potential, and consequently the collective energy change per turn, evolves with the bunch profile.

The total longitudinal impedance of an accelerator is the sum of several distinct contributions arising from different machine elements and physical mechanisms [39].

Impedances are typically categorized according to their physical origin [40]:

- **Resistive-Wall Impedance:** This arises from the finite conductivity of the vacuum chamber material. In a perfectly conducting pipe, image charges and currents move synchronously with the beam without resistance. However, finite conductivity leads to the skin effect, where fields penetrate the wall and are dragged behind the particle, dissipating energy and creating a trailing wake. This effect is significant even in smooth pipes and is often the dominant source of heating and instability in long storage rings.
- **Geometric Impedance:** This category encompasses wake fields generated by variations in the vacuum chamber cross-section, such as cavities, bellows, transitions, and collimators. As the beam field encounters a change in geometry, it diffracts at the aperture, exciting electromagnetic modes that can be trapped within the structure (resonant modes) or propagate down the beam pipe. Collimators, in particular, are critical sources of geometric impedance as they require small apertures to protect accelerator components, often leading to strong short-range wake fields.
- **CSR Impedance:** CSR differs from conventional wake-field mechanisms because the emitted field can interact with particles that are located *ahead* of the source when ordered along the reference trajectory. This does not violate causality, since the field acts only after it has been emitted. The effect arises in dipole magnets, where the radiation propagates approximately along straight lines while the electrons follow the longer curved orbit. As a result, radiation emitted by one part of the bunch can reach particles located ahead of the emitting particles in the longitudinal bunch coordinate. The effect is enhanced when the emitted wavelength is comparable to the longitudinal structures within the bunch, so that radiation from many particles adds coherently [41, 42].

  This interaction is modeled using two primary approaches:

  1. The **Free Space model**, which describes the interaction of a bunch moving on a curved path in a vacuum, leading to a strong "overtake" potential.
  2. The **Parallel Plates model**, which accounts for the shielding effect of the vacuum chamber boundaries. The conducting walls suppress the propagation of radiation with wavelengths comparable to or larger than the pipe aperture, effectively cutting off the impedance at low frequencies while converging to the free-space model at high frequencies.

A detailed overview of impedance models commonly used for synchrotrons is given in [34]. The purpose of this section is not to review these models individually, but to show how the total longitudinal impedance modifies the equations of motion and affects the phase space evolution.

2.3.4. Longitudinal phase space

The VFPE presented in Section 2.1 provides a general description of the evolution of a particle distribution in canonical phase space coordinates. To apply Eq. (2.37) to longitudinal beam dynamics, the phase deviation φ and the energy deviation ε, introduced in Section 2.3.2, must be related to canonical Hamiltonian coordinates.

The equations of motion derived previously describe the single particle longitudinal dynamics under the RF system and collective effects. To connect these physical variables to the VFPE, the dimensional equations are first rewritten in terms of the longitudinal position z and the energy deviation ε. These equations are then normalized to dimensionless variables (q, p), for which the longitudinal motion takes canonical oscillator form. The external Hamiltonian used in the VFPE is obtained from these normalized equations of motion.

The phase deviation is converted into a longitudinal position coordinate by relating it to the arrival time deviation Δt with respect to the synchronous particle. Following Schönfeldt [43, p. 4], this relation is written as

$$\varphi = -\omega_{\mathrm{RF}} \Delta t. \tag{2.110}$$

The same arrival-time deviation is written as

$$z := \beta c \, \Delta t, \tag{2.111}$$

where $z > 0$ is chosen toward the bunch tail. Combining both relations gives

$$z = -\frac{\beta c}{\omega_{\mathrm{RF}}} \, \varphi = -\frac{\beta c T_0}{2\pi h} \, \varphi. \tag{2.112}$$

where the harmonic condition $\omega_{\mathrm{RF}} = 2\pi h / T_0$ derived from Eq. (2.66) was used. Differentiating Eq. (2.112) and inserting Eq. (2.88) gives

$$\dot{z} = -\frac{\beta c}{\omega_{\mathrm{RF}}} \dot{\varphi} = \frac{\eta c}{\beta E_0} \, \varepsilon. \tag{2.113}$$

Substituting the expression for φ from Eq. (2.112) into the energy equation, Eq. (2.89), gives

$$\dot{\varepsilon} = -\frac{\omega_{\mathrm{RF}} e \hat{V} f_{\mathrm{rev}}}{\beta c} \, z. \tag{2.114}$$

Equations (2.113) and (2.114) show that the longitudinal motion is a harmonic oscillator in the physical variables (z, ε). Differentiating Eq. (2.113) with respect to time and substituting Eq. (2.114) yields

$$\ddot{z} + \left(\frac{\eta c}{\beta E_0} \frac{\omega_{\mathrm{RF}} e \hat{V} f_{\mathrm{rev}}}{\beta c} \right) z = 0. \tag{2.115}$$

The coefficient multiplying z defines the synchrotron angular frequency already introduced in Eq. (2.93). Equation (2.115) can therefore be written as

$$\ddot{z} + \Omega_s^2 z = 0, \tag{2.116}$$

with

$$\Omega_s^2 = \frac{\eta c}{\beta E_0} \frac{\omega_{\mathrm{RF}} e \hat{V} f_{\mathrm{rev}}}{\beta c} = \frac{2\pi h \eta e \hat{V} f_{\mathrm{rev}}^2}{\beta^2 E_0}. \tag{2.117}$$

Equation (2.116) has the form of a linear harmonic oscillator. The longitudinal variables are normalized using the zero-current root mean square (RMS) bunch length [1], σ_{z0}, and the zero-current RMS energy spread, $\sigma_{E,0}$. This gives the dimensionless coordinates

$$q := \frac{z}{\sigma_{z,0}}, \quad p := \frac{\varepsilon}{\sigma_{E,0}}. \tag{2.118}$$

The two normalization scales are related by [38, p. 142]

$$\sigma_{z,0} = \frac{|\eta| c}{\beta E_0 \Omega_s} \sigma_{E,0}. \tag{2.119}$$

Using Eqs. (2.113) and (2.114), the normalized variables satisfy

$$\dot{q} = \Omega_s p, \quad \dot{p} = -\Omega_s q. \tag{2.120}$$

Introducing the angular time[2]

$$\theta = \Omega_s t = 2\pi f_{s,0} t, \tag{2.121}$$

the equations become

$$\frac{dq}{d\theta} = p, \quad \frac{dp}{d\theta} = -q. \tag{2.122}$$

In canonical form, the Hamilton's equations are

$$\frac{dq}{d\theta} = \frac{\partial H_e}{\partial p}, \quad \frac{dp}{d\theta} = -\frac{\partial H_e}{\partial q}. \tag{2.123}$$

Comparing Eqs. (2.122) and (2.123) gives

$$\frac{\partial H_e}{\partial p} = p, \quad \frac{\partial H_e}{\partial q} = q. \tag{2.124}$$

Integrating these relations yields the dimensionless Hamiltonian generated by the external RF focusing:

$$H_e(q,p) = \frac{1}{2}\left(q^2 + p^2\right). \tag{2.125}$$

Collective effects enter the Hamiltonian through the wake potential $V_{\mathrm{c}}(z,t)$. The wake potential does not directly represent a potential energy of a single particle. Instead, it gives the collective energy kick produced by the fields generated by the bunch. As derived in Eq. (2.104), the energy change experienced by a test particle at longitudinal position z is

$$\Delta E(z,t) = \Delta\varepsilon(z,t) = eV_c(z,t). \tag{2.126}$$

This energy kick modifies the energy deviation ε. Because the normalized energy coordinate is defined by Eq. (2.118), the wake kick produces the following change in p:

$$\Delta p(z,t) = \frac{\Delta\varepsilon_{\mathrm{c}}(z,t)}{\sigma_{E,0}} = \frac{eV_c(z,t)}{\sigma_{E,0}}. \tag{2.127}$$

[1] Also called natural bunch length

[2] Schönfeldt [43, p. 12] defines the dimensionless time as $\theta = f_s t$, which counts synchrotron periods rather than angular phase. In this thesis, θ denotes the angular time, $\theta = 2\pi f_{s,0} t$.

Since the revolution frequency is f_{rev}, the particle receives f_{rev} such kicks per unit physical time. In the smooth-time approximation, the corresponding rate is therefore

$$\frac{dp}{dt} = f_{\mathrm{rev}}\Delta p = \frac{e f_{\mathrm{rev}}}{\sigma_{E,0}} V_{\mathrm{c}}(z,t). \tag{2.128}$$

Using the angular normalized time $\theta = 2\pi f_{s,0} t$, this becomes

$$\frac{dp}{d\theta} = \frac{1}{2\pi f_{s,0}} \frac{dp}{dt} = \frac{e f_{\mathrm{rev}}}{2\pi \sigma_{E,0} f_{s,0}} V_c(z,t). \tag{2.129}$$

Using the impedance representation introduced in Eq. (2.109), the wake potential can be expressed in the dimensionless canonical variables as

$$V_{\mathrm{c}}(q,\theta) = \frac{eN_b}{2\pi} \int_{-\infty}^{\infty} \tilde{\rho}^{\,\theta}(\omega)\, Z_{\parallel}(\omega)\, e^{i\omega\sigma_{z,0}q/v}\, d\omega, \tag{2.130}$$

where $\tilde{\rho}^{\,\theta}(\omega)$ denotes the Fourier transform of the bunch profile at the revolution time $t = \theta/(2\pi f_{s,0})$. The factor σ_{z0} appears because the impedance is defined with respect to the physical longitudinal coordinate $z = \sigma_{z0} q$. In dimensionless canonical variables, this spectrum is written as

$$\tilde{\rho}^{\,\theta}(\omega) = \int_{-\infty}^{\infty} \rho(q,\theta) e^{-i\omega\sigma_{z,0}q/v}\, dq, \tag{2.131}$$

where the bunch profile, $\rho(q,\theta)$, is determined by the phase space density at θ:

$$\rho(q,\theta) = \int_{-\infty}^{\infty} \psi(q,p,\theta)\, dp. \tag{2.132}$$

From Eq. (2.129), a change of the energy coordinate is generated by the derivative of the Hamiltonian with respect to the longitudinal coordinate:

$$\frac{dp}{d\theta} = \frac{e f_{\mathrm{rev}}}{2\pi \sigma_{E,0} f_{s,0}} V_c(z,t) = -\frac{\partial H_c}{\partial q}. \tag{2.133}$$

Therefore, the collective kick is represented by adding a q-dependent term to the Hamiltonian, obtained as an antiderivative of Eq. (2.133):

$$H_{\mathrm{c}}(q,\theta) = \frac{e f_{\mathrm{rev}}}{2\pi \sigma_{E,0} f_{s,0}} \int_{q}^{\infty} V_{\mathrm{c}}(q',\theta)\, dq'. \tag{2.134}$$

The longitudinal Hamiltonian in normalized phase space coordinates is therefore

$$H(q,p,\theta) = H_e(q,p) + H_{\mathrm{c}}(q,\theta). \tag{2.135}$$

The remaining terms in Eq. (2.37) account for radiation damping and quantum excitation, which act on the energy coordinate. For the zero-current equilibrium, the stationary phase space density satisfies the VFPE in the absence of collective effects. This condition fixes the relation between the damping coefficient β_d and the diffusion coefficient D as [44]

$$D = \beta_d. \tag{2.136}$$

From Eq. (2.99), the physical damping rate entering the Fokker–Planck drift term is $1/\tau_d$, where the damping time τ_d has units of seconds. Expressing this rate with respect to the angular time gives the damping coefficient

$$\beta_d = \frac{1}{2\pi f_{s,0}\tau_d}. \tag{2.137}$$

Finally, the dimensionless VFPE for longitudinal beam dynamics takes the form

$$\frac{\partial\psi}{\partial\theta} + p\frac{\partial\psi}{\partial q} - \left[q - \frac{e f_{\mathrm{rev}}}{2\pi\sigma_{E,0} f_{s,0}} V_{\mathrm{c}}(q,\theta)\right]\frac{\partial\psi}{\partial p} = \beta_d \frac{\partial}{\partial p}\left(p\psi + \frac{\partial\psi}{\partial p}\right). \tag{2.138}$$

3. Beam diagnostics for longitudinal beam dynamics at KARA

The previous chapter established the theoretical framework of beam dynamics in synchrotrons, progressing from longitudinal motion to the nonlinear collective effects that drive the micro-bunching instability and the emission of CSR, ultimately reshaping the beam's phase space density. The beam dynamics can be investigated experimentally through observables that reflect the evolution of the longitudinal phase space of the electron bunch. This chapter therefore turns from theoretical beam dynamics to experimental diagnostics and introduces the measurement systems at KARA that provide the data required for longitudinal phase space reconstruction and validation.

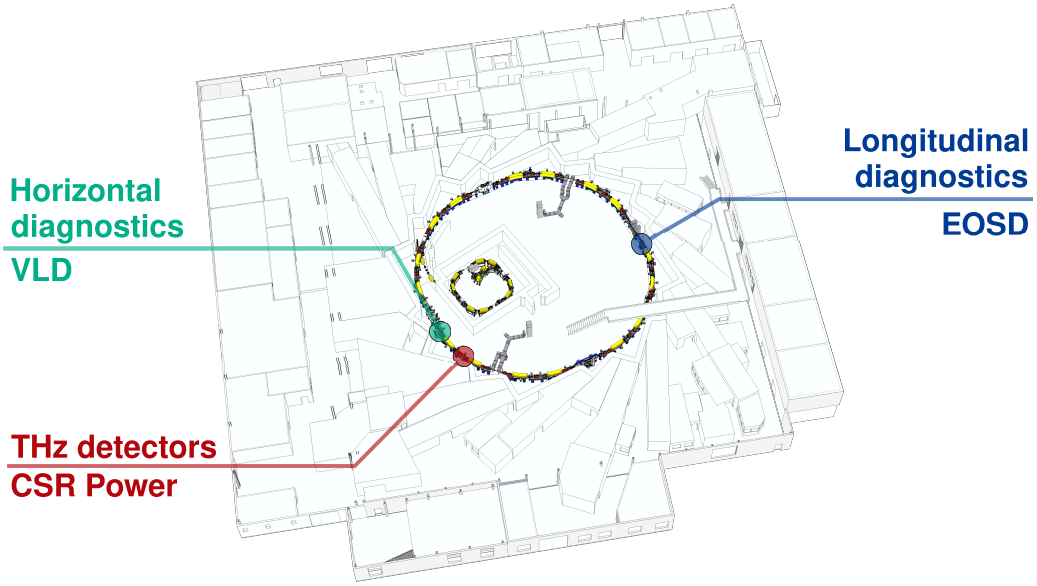


Figure 3.1.: Schematic layout of KARA showing the locations of the three diagnostics used to validate the reconstructed longitudinal phase space density.

The experimental setup shown in Fig. 3.1 combines synchronized turn-by-turn measurements from three complementary diagnostics [16]. The principal diagnostic used in this work is the EOSD system, which provides information on the longitudinal bunch profile used in the reconstruction [45]. Complementary diagnostics include fast THz detectors, which measure the emitted CSR power [15], and the VLM, which provides the horizontal beam profile [46]. These additional diagnostics supply independent experimental measurements against which the physical consistency of the reconstructed phase space density can be assessed.

This combination of diagnostics is particularly well suited for the study of micro-bunching dynamics [16]. EOSD provides the longitudinal bunch profile information with sub-picosecond resolution, allowing fine structures in the charge distribution to be resolved. From the reconstructed phase space density, the emitted CSR power can be calculated and directly compared with the signal measured by the fast

THz detectors. Likewise, the energy-spread information inferred from the reconstructed phase space can be compared with the horizontal bunch size measured by the VLM. Their synchronized acquisition therefore makes it possible to evaluate the reconstructed phase space not only through the EOSD measurements that drives the reconstruction, but also through independent experimental observables linked to the same underlying longitudinal dynamics [13, 10].

This chapter first introduces the two diagnostics used as independent validation channels for the reconstructed longitudinal dynamics: the VLM, which provides horizontal beam-size information related to the energy spread, and the THz detector system, which measures the CSR emission associated with micro-bunching formation. The central reconstruction diagnostic, EOSD, is then described in detail, including the measurement principle and the simulation methods used to model its transfer function.

3.1. Visible-light monitor

The VLM is one of two diagnostics used to validate the reconstructed longitudinal dynamics. It measures the horizontal beam profile to derive the horizontal bunch length, which is then compared against the energy-spread evolution inferred from the reconstructed phase space.

The horizontal bunch profile was measured at the visible-light diagnostics (VLD) port of KARA, where SR in the visible range is used for beam diagnostics [46, 17]. The visible SR is extracted at a dipole beam port and transported through an optical beamline composed of two off-axis paraboloid mirrors and two planar mirrors, which form an intermediate image of the source point rotated by 90° with respect to the beam. To permit simultaneous operation of different diagnostics, the beamline was designed to distribute the incoming light into separate spectral branches by means of consecutive short-pass filters. In this way, the same visible-light source can support several complementary instruments, including filling-pattern diagnostics, streak-camera measurements, and transverse bunch size measurements.

For turn-by-turn measurements of the horizontal bunch profile at the VLD port [47], Karlsruhe line array detector for MHz-repetition rate spectroscopy (KALYPSO) is an ultra-fast line-array detector system with 256 pixels and frame rates up to 2.7 MHz, thereby enabling single-turn-resolved measurements in single-bunch operation [48]. Because the system was designed for high continuous data throughput, long acquisition times are possible, with the main limitations arising from data transfer and the available memory of the readout computer.

The setup uses incoherent visible SR emitted at the VLD source point. At the optical table, the radiation is deflected by a mirror and focused onto the KALYPSO sensor by a cylindrical lens, while the readout system is contained in a compact shielded housing. The resulting detector signal consists of a sequence of horizontal bunch profiles acquired turn by turn, so that the temporal evolution of the horizontal bunch distribution can be followed directly over many revolutions. A comprehensive technical description of this VLMing system and its implementation at KARA is detailed in [49, 50].

Since the observation point is located in a dispersive section of the storage ring, the measured horizontal bunch size contains contributions from both the horizontal beta function and the energy spread. For fixed lattice optics, turn-by-turn variations of the horizontal bunch size can therefore be used as an indicator of changes in the energy spread [13, 16]. The horizontal bunch size can be written as

$$\sigma_x^2 = \beta_x \varepsilon_x + \left(D_x \sigma_E\right)^2 , \tag{3.1}$$

where β_x is the horizontal beta function at the source point, ε_x is the horizontal emittance, D_x is the horizontal dispersion, and σ_E denotes the relative energy spread. Under fixed optical conditions and

an approximately constant beta function contribution, $\beta_x \varepsilon_x$ remains nearly unchanged, so variations in the horizontal bunch size derived from the measured profile can be interpreted as an indicator of changes in the energy spread. Figure 3.2 illustrates the comparison process of the horizontal bunch size derived from the measurements and the derivation of the energy spread from the phase space distribution.

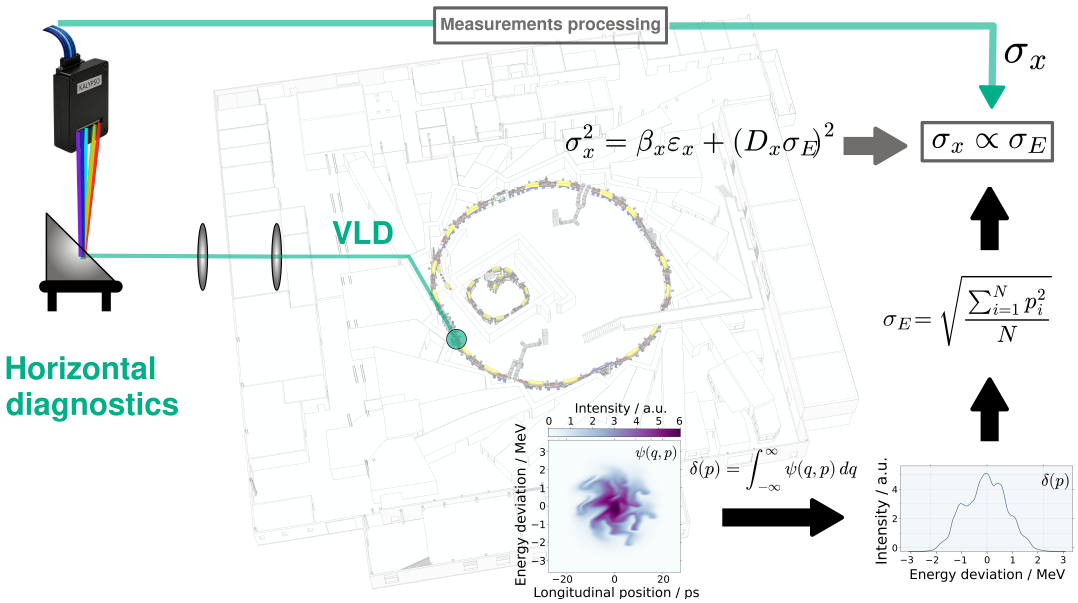


Figure 3.2.: Sketch of the comparison between the horizontal bunch size obtained from the VLM diagnostic and the energy spread derived from the phase space density.

This diagnostic is not used to reconstruct the longitudinal phase space directly. Instead, it serves as a consistency check for the reconstructed dynamics. Synchronized measurements at KARA have shown that longitudinal and transverse observables can be recorded simultaneously and compared on a turn-by-turn basis [16]. During burst phases of the micro-bunching instability, the reconstructed phase space is expected to show an increase in energy spread, which should manifest itself as an enlarged horizontal bunch profile in the dispersive section. The VLD port thus supplies an independent observable for validating the reconstructed longitudinal dynamics [13, 10].

3.2. THz detector

While the VLM provides access to horizontal beam size variations related to the energy spread, the THz detector measures the CSR power emitted by the bunch. The THz detector system therefore provides a second independent observable for validating the reconstructed longitudinal dynamics. This signal is particularly relevant in the micro-bunching regime [12], where longitudinal substructures on sufficiently short length scales lead to a strong enhancement of CSR [51, 10].

The emitted CSR power was recorded with fast THz detectors based on room temperature zero-bias Schottky barrier diodes at the Infrared 2 beamline (IR2) [52]. These detectors measure the incident radiation power integrated over their frequency dependent sensitivity range, which extends over several hundred gigahertz. Their response times of about 100 ps enable bunch-to-bunch discrimination of fast CSR/THz signals at KARA [53].

The detector signals were read out using the Karlsruhe Pulse Taking Ultra-fast Readout Electronics (KAPTURE) system [54], a dedicated 500 MHz acquisition system developed for continuous bunch-by-bunch sampling of the fast signals from the THz detectors.

The THz power serves as an independent observable for micro-structure formation in the bunch. The reconstructed longitudinal phase space density, obtained from EOSD measurements, can be used to derive the longitudinal bunch profile and from it the expected CSR power, P_{CSR}, in the terahertz range. The predicted CSR power can then be compared with the THz power measured experimentally. Figure 3.3 illustrates the comparison between the terahertz power measured by the THz detector and the CSR power derived from the reconstructed phase space density.

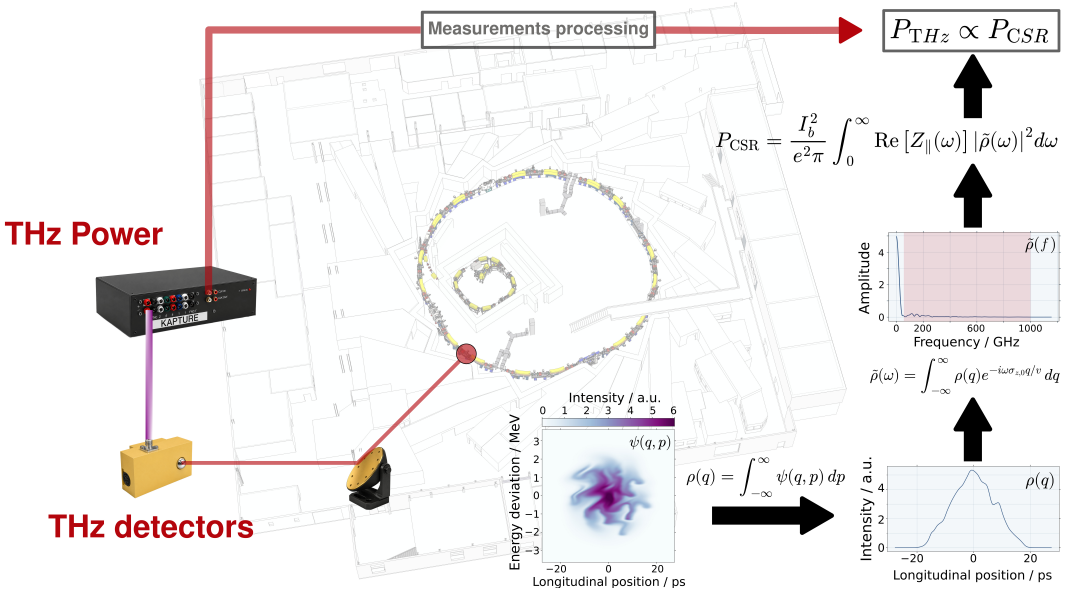


Figure 3.3.: Sketch of the comparison between the terahertz power measured by the THz detector and the CSR power derived from the phase space density. The quantity P_{CSR} is calculated from the bunch current I_b, the longitudinal impedance, and the spectral range from 70 GHz to 1.1 THz covered by the THz detector, highlighted in red in the Fourier spectrum of the bunch profile.

The THz power is not part of the reconstruction itself, but rather an independent observable used for its physical validation. This validation relies on the expected relation between longitudinal substructure and CSR emission: the formation of fine structures increases the emitted power, whereas their decay or smoothing leads to a reduction in CSR power. Therefore, agreement between the predicted CSR power and measured THz power provides an independent validation of whether the reconstructed phase space correctly captures the fine longitudinal structures responsible for CSR emission.

3.3. Electro-optical spectral decoding

The central diagnostic for phase space density reconstruction is the EOSD system, which provides single-shot measurements of the longitudinal bunch profile with sub-picosecond resolution [14, 55]. At KARA, a near-field monitor was developed for turn-by-turn longitudinal bunch profile measurements in the storage ring and implemented as an in-vacuum diagnostics system close to the beam [14, 55, 56, 57].

The core of this near-field setup is an electro-optical crystal positioned in-vacuum close to the beam path[55, 56]. As the electron bunch passes, its highly compressed Coulomb field modifies the intrinsic birefringence of the crystal via the Pockels effect [58, 59].

A chirped laser pulse, whose frequency varies along the pulse length, is sent through the crystal. The induced change in birefringence modifies its time-dependent polarization, which is then converted into an intensity modulation using waveplates and a beam splitter in a nearly crossed configuration.

The modulated laser pulse is then analyzed with a fast spectrometer based on a grating and the KALYPSO line-array detector [60, 61], enabling turn-by-turn single-shot measurements of the longitudinal bunch profile with sub-picosecond resolution [45, 58, 59, 55, 56]. A comprehensive overview of the electro-optical sampling (EOS) principle and its application to very short electron bunches can be found in [62, 14, 63].

Although the EOSD monitor provides the most direct information on the longitudinal charge distribution and shows that the appearance of micro-bunching substructures coincides with intense CSR bursts, the measured signal is not identical to the bunch charge profile [2]. Instead, it is shaped by the electro-optical response of the crystal, the laser propagation, the setup geometry, and the spectrometer readout.

3.3.1. The Pockels effect and laser phase shift

The electro-optical modulation of the laser pulse propagating through the crystal is governed by the Pockels effect, which models the linear change in the refractive index of the crystal in the presence of an external electric field. In an isotropic medium, the material responds equally in all directions, so the applied electric field induces a polarization in the material that is aligned with the field. The optical response can therefore be described by a scalar susceptibility. In an anisotropic crystal, or in a crystal whose optical properties are modified by an external field, the susceptibility and therefore the dielectric response are no longer described by a scalar but by a higher-rank tensor. The susceptibility describes how the material polarizes in response to the field, while the dielectric response describes how this polarization modifies the field inside the crystal.

Consequently, the refractive index depends on both the propagation direction and the polarization of the light. Different polarization components therefore propagate through the crystal at different phase velocities and accumulate a relative phase delay, referred to as the phase shift. This phenomenon is known as birefringence and constitutes the central mechanism underlying the EOSD signal.

The electro-optic crystal considered in this study is made from Gallium Phosphide (GaP). In the absence of an external field, GaP is optically isotropic to a good approximation[1] [59], meaning the refractive index is identical for all polarization directions and can be denoted by n_0. When the electric field of the passing electron bunch reaches the crystal, it alters the optical response of the material through the Pockels effect. This change is described as a linear perturbation of the inverse dielectric tensor and appears experimentally as a small modification in the refractive indices of the two orthogonal polarization components, also referred to as the eigenpolarizations (or eigenmodes) of the crystal.

The GaP crystal is mounted at a specific orientation relative to the electric field to maximize the observable electro-optic effect. With this orientation, the Pockels response is governed predominantly by the electro-optic coefficient r_{41} [62]. If E denotes the electric-field component responsible for the electro-optic modulation, the induced birefringence causes the two laser polarization components along the crystals eigenpolarization to experience slightly different refractive indices:

[1] In the experimental setup at KARA, intrinsic birefringence was observed and compensated using waveplates.

$$n_1 \approx n_0 + \frac{1}{2} n_0^3 r_{41} E, \quad n_2 \approx n_0 - \frac{1}{2} n_0^3 r_{41} E. \tag{3.2}$$

Since these components propagate with slightly different phase velocities, the electro-optical effect can be observed via the phase retardation accumulated between them. Subtracting the two expressions in Eq. (3.2) yields their refractive-index difference:

$$\Delta n = n_1 - n_2 \approx n_0^3 r_{41} E. \tag{3.3}$$

For a laser pulse that propagates through a crystal slice of thickness d, the accumulated phase retardation Γ is the optical path difference expressed in units of the laser wavelength, λ:

$$\Gamma = \frac{2\pi d}{\lambda} \Delta n. \tag{3.4}$$

Inserting Eq. (3.3) into Eq. (3.4) gives

$$\Gamma \approx \frac{2\pi d}{\lambda} n_0^3 r_{41} E. \tag{3.5}$$

This expression represents the maximum phase retardation, assuming that the electric field is oriented at 45° with respect to crystal's eigenmodes and remains approximately constant over the considered slice during the interaction time.

Furthermore, the electro-optic coefficient r_{41} is frequency-dependent. Depending on the crystal material, this relationship may be approximately linear only within a certain frequency region, beyond which the response becomes non-linear.

In the actual experimental setup, however, the electric field generated by the bunch varies across the crystal during the passage of the laser pulse because of the setup geometry, as discussed in M. Reissig's thesis [63] and described in the following section.

The laser does not experience a static or spatially uniform field because the bunch field varies across the crystal during its passage. For a chirped laser pulse, each temporal position τ within the pulse is associated with a different frequency, so the calculation must determine which local value of the space- and time-dependent transverse electric field is sampled by each laser temporal slice as it propagates through the crystal.

To model this interaction, the crystal is divided into n sufficiently small longitudinal slices of thickness d_i. A temporal slice τ of the laser pulse intersects crystal slice i at the interaction time $t_i(\tau)$, where it samples the local transverse field $E_y(z_i, t_i(\tau))$. Physically, the resulting retardation is the phase difference accumulated between the two polarization components of that laser slice. The total phase retardation is therefore approximated by

$$\Gamma(\tau) \approx \sum_{i=1}^{n} \frac{2\pi d_i}{\lambda} n_0^3 r_{41} E_y(z_i, t_i(\tau)) \,. \tag{3.6}$$

The phase difference is converted into an intensity modulation by passing the laser through a $\lambda/4$ waveplate, a $\lambda/2$ waveplate, and a polarizing beam splitter. The modulated chirped laser pulse is then analyzed by a spectrometer consisting of a grating and the KALYPSO line-array detector. Since the laser pulse is chirped, its spectrum encodes the longitudinal coordinate of the pulse: each detector pixel corresponds to a different optical frequency and therefore to a different time slice of the laser pulse.

In practice, the measured phase retardation, denoted by $\hat{\Gamma}$, can be written as

$$\hat{\Gamma} = \frac{I_M(f)}{I_U(f)} - 1, \tag{3.7}$$

where $I_M(f)$ is the measured modulated laser spectrum and $I_U(f)$ is the corresponding unmodulated reference spectrum, obtained without electro-optical modulation. The quantity $\hat{\Gamma}$ therefore represents the experimentally extracted measured EOSD signal and is related to the physical phase retardation [45]. To improve the signal-to-noise ratio, the mean channel background noise, measured without laser light, is subtracted from both spectra before the normalization.

Once the experimental $\hat{\Gamma}$ is extracted, it must be mathematically mapped back to the physical charge distribution of the electron bunch. In a simplified analytical model, the longitudinal bunch profile is assumed to be Gaussian, while the electric field inside the crystal is modeled as a plane wave with an amplitude determined by the Coulomb field of a near-relativistic charged particle. Combined with a constant approximate frequency dependence of r_{41}, this model allows the electric field to be calculated explicitly at each point along the laser path [62]. Under these idealized assumptions, estimating the bunch profile reduces to determining the standard deviation of the Gaussian profile that best reproduces the measured signal [62, p. 71].

Although the analytical reduction is useful for illustrating the measurement principle when the bunch profile is assumed to be Gaussian, it does not adequately represent the reconstruction of non-Gaussian profiles under the experimental conditions considered in this work. In particular, in the presence of micro-bunching, the bunch profile can develop sub-structures, while the electric field inside the crystal is also influenced by additional electromagnetic effects, including geometry-dependent wakefield contributions. Estimating the time-dependent field distribution inside the crystal therefore requires an external electromagnetic solver capable of computing the field response for a given beam distribution and EOSD system geometry.

3.3.2. Electromagnetic field simulations

As shown in Equation 3.6, calculating the total phase retardation of the laser pulse requires sampling the electric field at discrete probe points along its path through the crystal. The probes are spaced at intervals d_i, representing thin slices over which the electric field is assumed constant.

An electromagnetic solver is used to obtain the electric field at these probe positions. The solver must account for the full three-dimensional geometry of the in-vacuum part of the EOSD setup and for the field generated by a charged particle bunch moving at nearly the speed of light. For this purpose, the Wakefield Solver within the Computer Simulation Technology (CST) Particle Studio Suite [64] is utilized. This time-domain solver relies on the finite integration technique (FIT) [65], which evaluates Maxwell's equations on a discretized grid using a leap-frog approach for time advancement.

The Wakefield Solver can provide wake potentials and impedances, which are commonly used, for example, in accelerator-cavity and impedance studies [66]. For the EOSD forward model considered here, however, the time-dependent electric field at the probe positions inside the crystal is required. The geometry and orientation of the electro-optical (EO) crystal are typically chosen to maximize the phase retardation for the transverse component of the electron bunch's electric field $E_y(t)$. Thus the transverse component is of particular interest for the simulation.

The CST Studio 3D model representing the EOSD crystal setup and beam pipe is a geometric simplification of the original computer-aided design (CAD) design. Nevertheless, the macroscopic dimensions

and primary structural shapes accurately reflect the physical hardware installed at KARA, as depicted in the full model shown in Figure 3.4.

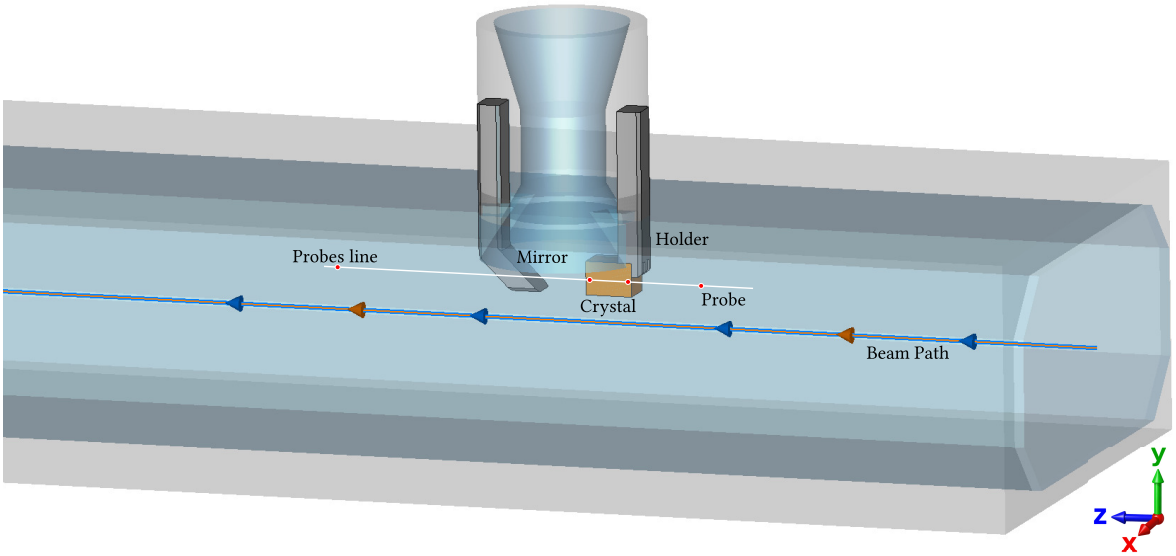


Figure 3.4.: 3D CST Studio Suite® model of the EOSD setup, including the beam pipe, crystal holder, mirror, and electro-optic crystal. The field probes are placed along a straight line through the center of the crystal. Four representative probe locations are highlighted in red: one before the EOSD structure, two inside the crystal near the entrance and exit facets, and one after the crystal along the beam direction.

To illustrate how the electric field evolves at different probe positions, Fig. 3.5 shows an example obtained with a Gaussian beam profile.

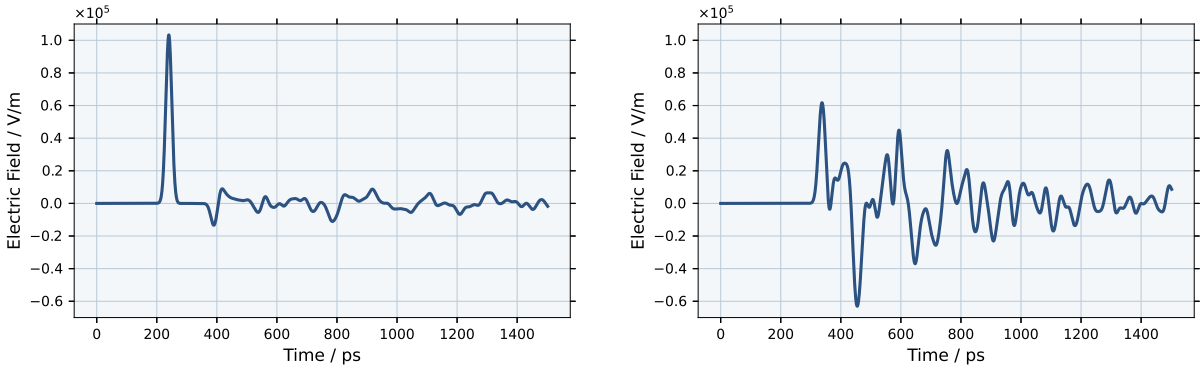
Figure 3.5a shows the electric field before the crystal. The dominant peak follows the Gaussian bunch profile used as input to the CST Wakefield Solver. The smaller oscillations at later times correspond to trailing wakefield contributions reaching the probe position after the passage of the main field. A reflected contribution from the entrance face of the crystal is also visible around 400 ps.

Figure 3.5b shows the field inside the crystal near the entrance face. At the vacuum–dielectric interface, part of the incident field is reflected, while the transmitted component enters the crystal with a modified amplitude. As the field propagates through the dielectric, its amplitude and polarization evolve further, as observed near the exit face in Fig. 3.5c. At the second interface, part of the field is again reflected and the transmitted component propagates back into vacuum.

Figure 3.5d shows the field after the crystal, where the dominant Gaussian-like structure remains similar to that observed before the crystal despite the modifications introduced during propagation through the dielectric.

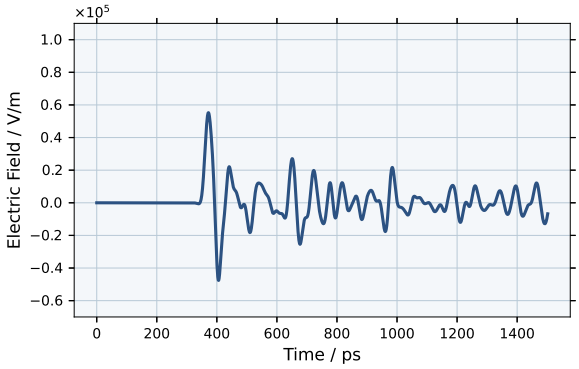
As demonstrated in [62], the transitional behavior of the main bunch field across the vacuum-dielectric boundaries can be analytically modeled using a plane wave approximation. However, the trailing wakefields are strongly coupled to the macroscopic geometry of the crystal setup and the beam pipe. As a result, the numerical wakefield solver is used, which includes the simplified geometries and material properties of the crystal and its surrounding structures. Figure 3.6 shows the wakefield isolines remaining inside and around the crystal after the bunch and its Coulomb field have passed the EO assembly.

The maximum simulation frequency in the CST Wakefield Solver must be chosen consistently with the spectral content of the selected bunch charge profile to preserve as much of its frequency content, and

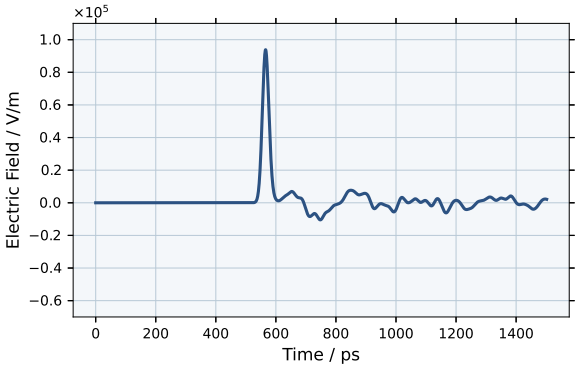


(a) Probe before the crystal, first red point on the right.

(b) Probe inside the crystal near the entrance face.



(c) Probe inside the crystal near the exit face.



(d) Probe after the crystal along the beam direction.

Figure 3.5.: Examples of the time-dependent electric field evaluated at the four probe positions indicated in Fig. 3.4, computed using the CST Wakefield Solver. The simulation assumes a Gaussian bunch profile and illustrates the temporal evolution of the electric field before the electro-optic crystal, inside the crystal near its entrance and exit faces, and after the crystal.

thus its sub-picosecond structures, as possible. For the simulations previously shown in Figure 3.5, the solver was configured with an upper frequency limit of 50 GHz. This choice is appropriate because the input charge profile is Gaussian with $\sigma = 9.87\,\mathrm{ps}$, whose spectral content is concentrated well below this frequency. The corresponding time-domain charge profile and its Fourier spectrum are shown in Fig. 3.7.

Consider now a custom bunch profile generated with Inovesa. This profile is approximately 30 % shorter than the Gaussian profile discussed above and contains micro-bunching substructures. Its longitudinal charge distribution and corresponding Fourier spectrum are shown in Fig. 3.8. In contrast to the Gaussian case, the micro-bunching introduces significant high-frequency content. A cutoff at 50 GHz is therefore insufficient to represent the relevant spectral components; the simulation bandwidth must be extended ideally above 1 THz, to capture the fine structures of the bunch profile.

High-frequency CST wakefield simulations require finer spatial and temporal discretization, which substantially increases the required memory and computational time. To reduce this computational cost, the full model shown in Fig. 3.4 is replaced by a reduced geometry that retains only the electro-optic crystal, its support structure, and a shorter section of the beam pipe, as shown in Fig. 3.9.

Figure 3.10 shows the simulated electric field at a probe positioned inside the crystal near its starting edge. To focus on simulations of the EOSD bunch profile measurements, the electric field of the probe is cropped to the primary field induced by the bunch's Coulomb field, thus omitting trailing wakefields.

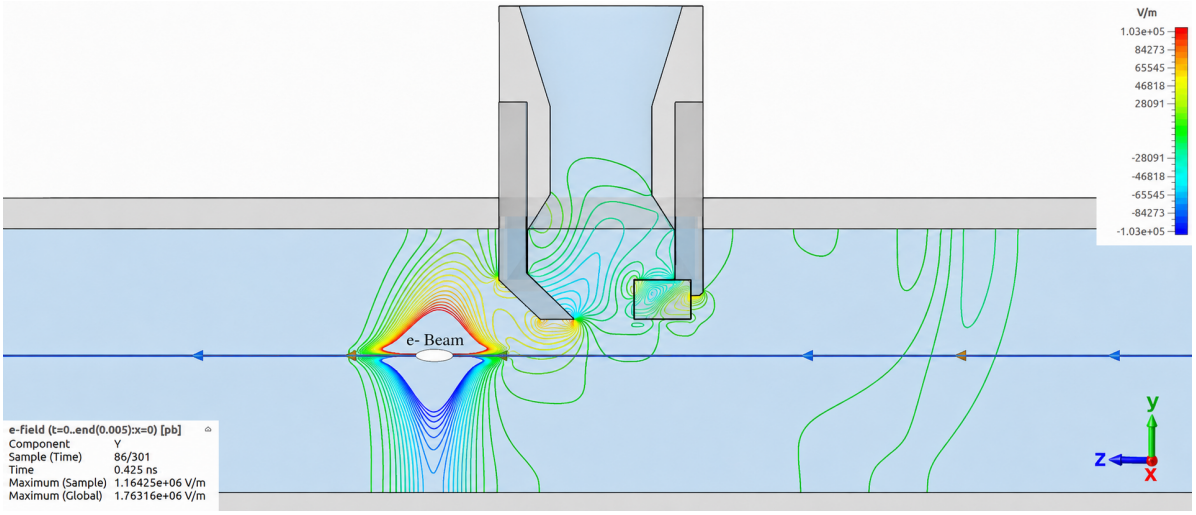
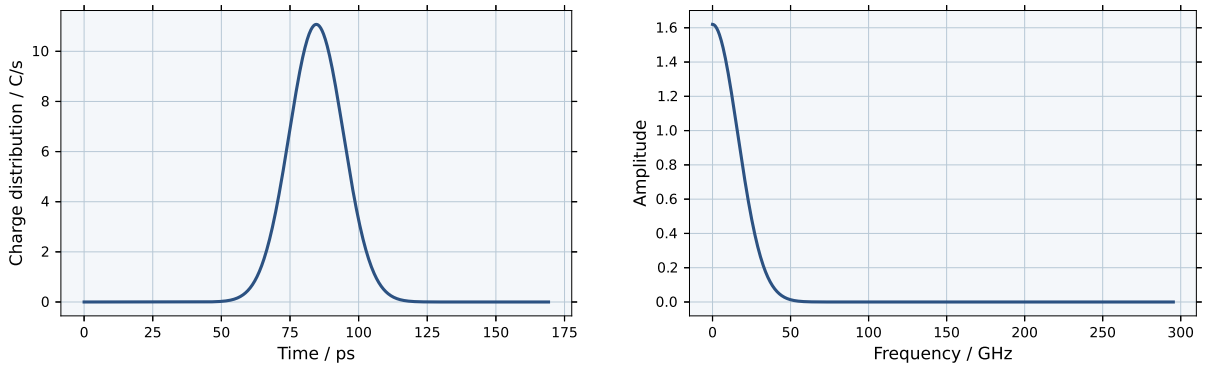


Figure 3.6.: Example of trailing wakefields generated after the electron bunch passes the EOSD setup. The field isolines illustrate the geometry-dependent wakefield pattern inside and around the electro-optic crystal.



(a) Longitudinal Gaussian charge profile.

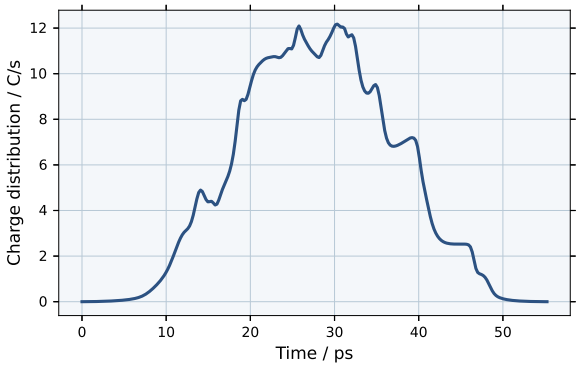
(b) Corresponding frequency spectrum.

Figure 3.7.: Gaussian bunch profile ($\sigma = 9.87$ ps) and its Fourier transform. The spectrum shows that an upper simulation limit of 50 GHz effectively captures the dominant spectral components of the bunch.

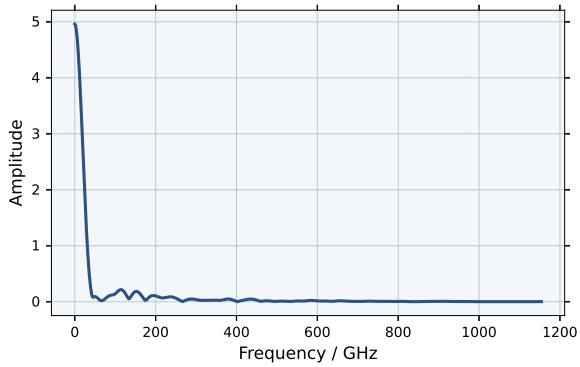
The results are compared across four different CST frequency cutoffs. By referencing the original micro-bunching profile in Figure 3.8a, it is clear that a 50 GHz cutoff is insufficient and smooth out the micro-bunching substructures. However, at cutoffs above 600 GHz, the simulation accuracy improves considerably, capturing the high-frequency micro-structures.

For a more quantitative analysis, Fig. 3.12 tracks the electric field across four specific probe locations inside the crystal. The data show significant signal degradation, with the micro-bunching structures encoded in the electric field becoming substantially smoothed after propagating through approximately one third of the crystal.

The simulations show that the electric field sampled at the probe positions differs significantly when comparing a Gaussian charge profile with a custom profile exhibiting micro-bunching substructures. The next subsection examines how these differences propagate into the phase retardation calculation.



(a) Longitudinal charge profile generated with Inovesa, showing micro-bunching substructures.



(b) Magnitude of the corresponding Fourier spectrum, showing the high-frequency components introduced by the micro-bunching substructures.

Figure 3.8.: Custom longitudinal bunch profile with micro-bunching and its spectral content. Compared with the Gaussian profile, the substructures generate significant high-frequency components, requiring a substantially larger simulation bandwidth than 50 GHz.

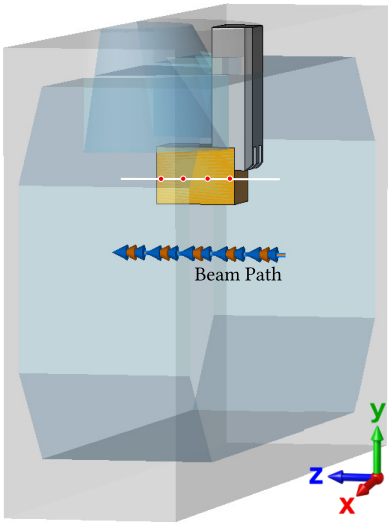
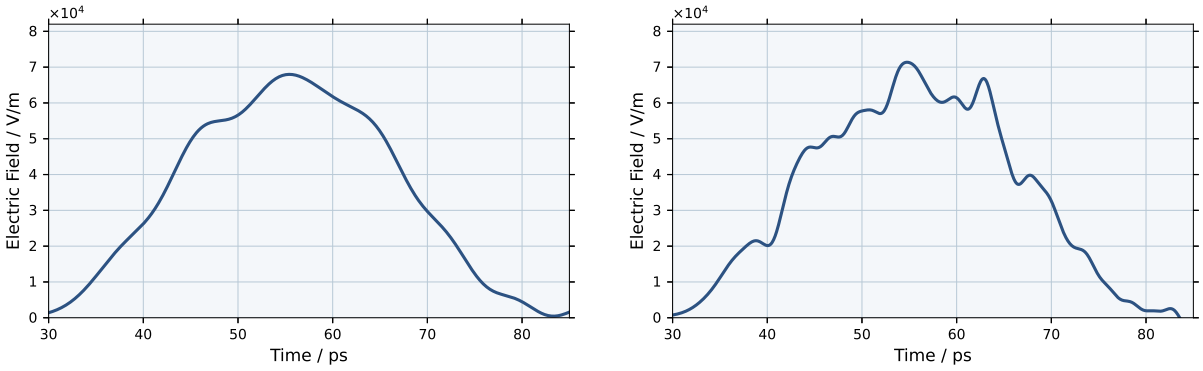


Figure 3.9.: Reduced 3D CST model of the EOSD setup. To minimize computational requirements for high-frequency simulations, the geometry is truncated to include only the electro-optic crystal, the crystal holder, and the immediate surrounding section of the beam pipe.

3.3.3. EOSD simulation

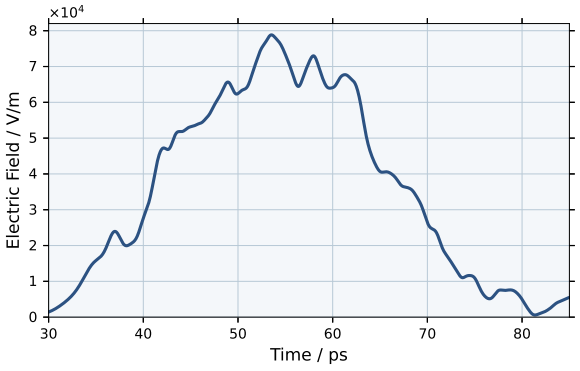
The phase retardation expression in Eq. (3.6) assumes an idealized geometry in which the laser pulse and the bunch electric field enter the crystal simultaneously from the same side and propagate together. The EOSD setup at KARA, however, follows a different optical configuration. As illustrated in Fig. 3.13, the laser pulse is injected from above, reflected by a mirror at approximately a right angle, and then enters the crystal from the side opposite to the incoming bunch field. It propagates toward the far end of the crystal, where a reflective coating redirects the pulse. After this reflection, the laser pulse and the bunch field co-propagate through the crystal:

This geometry naturally defines two interaction regions. In the upstream region, the laser pulse propagates opposite to the bunch electric field, whereas in the downstream region, after reflection,

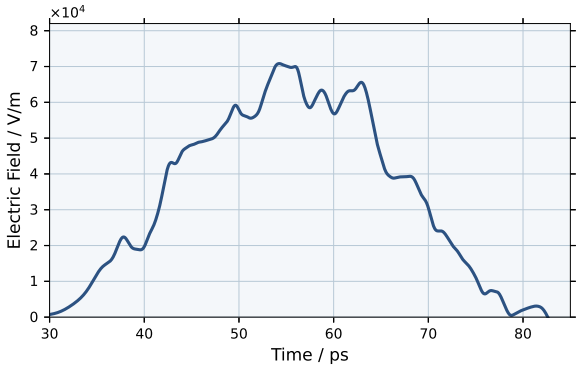


(a) Upper frequency limit of 50 GHz.

(b) Upper frequency limit of 200 GHz.



(c) Upper frequency limit of 600 GHz.



(d) Upper frequency limit of 1 THz.

Figure 3.10.: Comparison of the bunch-induced electric field at a probe located inside the electro-optic crystal near the entrance face for different CST Wakefield Solver frequency cutoffs. The time window is restricted to the dominant electric-field components of 55 ps; smaller wakefield-related oscillations are omitted. The 50 GHz cutoff smooths out the micro-bunching substructures, whereas higher cutoffs preserve the fast temporal variations of the field more accurately.

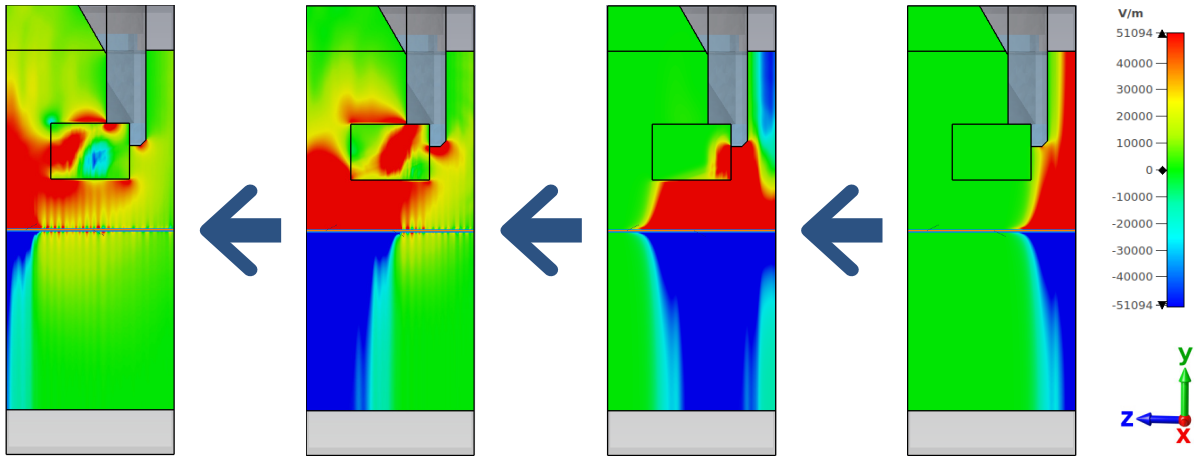
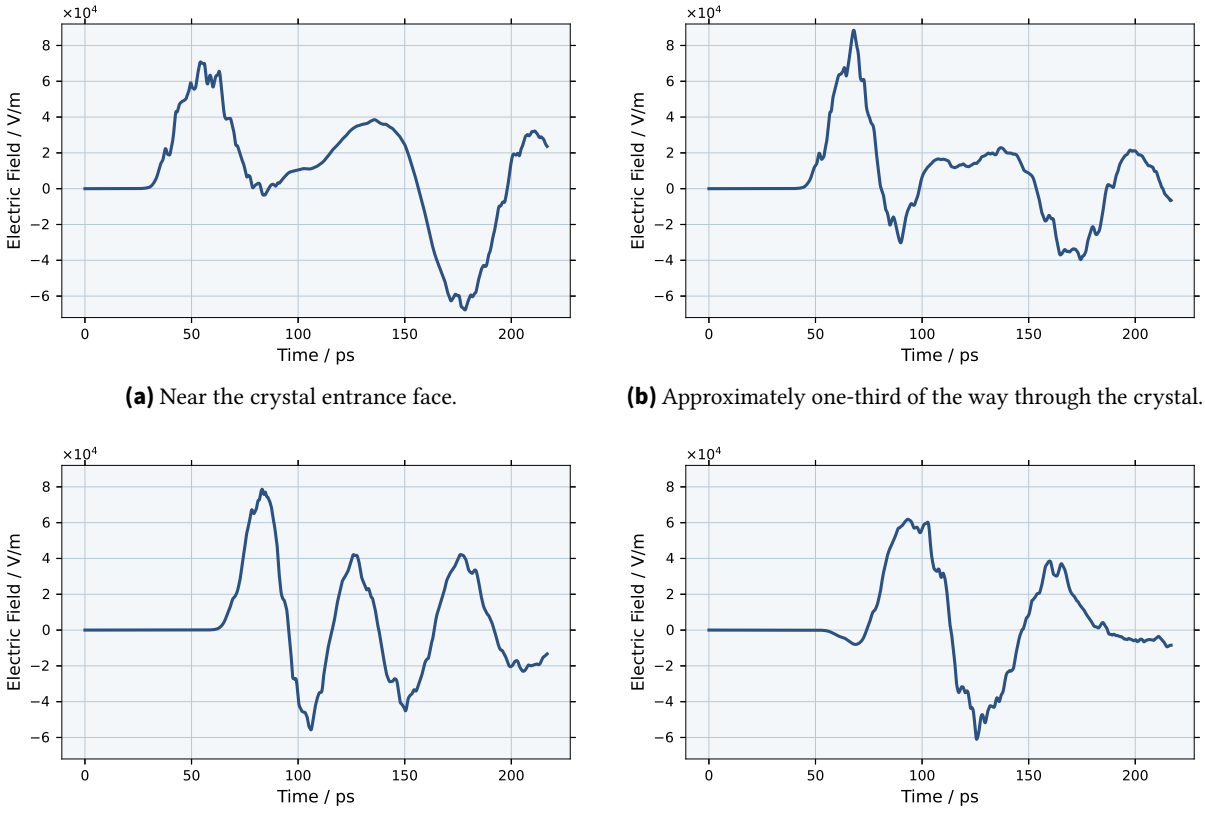


Figure 3.11.: CST simulation of the electric field generated by an electron bunch, with a maximum simulation frequency of 200 GHz. The bunch exhibits micro-bunching as it passes the electro-optic crystal from right to left. The resulting electric-field distribution is modified by the crystal and the surrounding geometry. In the final frame, the blue regions inside the crystal indicate a reversal of the field direction in the trailing part of the signal.



(a) Near the crystal entrance face.

(b) Approximately one-third of the way through the crystal.

(c) Approximately two-thirds of the way through the crystal.

(d) Near the crystal exit face.

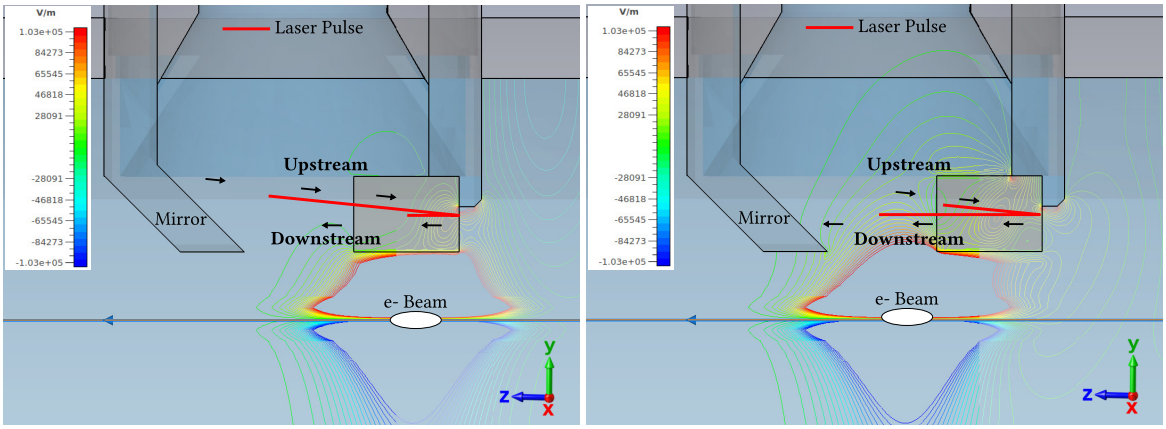
Figure 3.12.: Evolution of the bunch-induced electric field at four probe positions inside the electro-optic crystal for a micro-bunched input profile in a 1 THz simulation. The comparison shows how the fine temporal structures are progressively modified and smoothed as the field propagates through the crystal and interacts with the surrounding EOSD geometry.

both propagate in the same direction. These two regimes must be taken into account when modeling the accumulated phase retardation.

The phase retardation in Eq. (3.6) represents the accumulated phase shift along the laser pulse. By parameterizing the position along the pulse with the local coordinate τ and accounting for both the upstream and downstream propagation paths, this expression can be generalized. In this form, the phase retardation is obtained by evaluating the electric field along the trajectory followed by each point of the laser pulse within the crystal, leading to

$$\Gamma(\tau) = \frac{2\pi}{\lambda} n_0^3 r_{41} \int_{\mathcal{C}} E_y\big(z, t(z;\tau)\big)\, \mathrm{d}z, \tag{3.8}$$

where τ denotes the temporal delay associated with a frequency component of the chirped laser pulse, such that each value of τ corresponds to a specific position along the laser pulse. Furthermore, $\mathcal{C}$ denotes the trajectory followed by the laser inside the crystal, while $t(z;\tau)$ represents the time at which the laser coordinate τ reaches the position z along its path.



(a) Laser pulse entering the crystal opposite to the bunch field and propagating toward the reflective coating, with a substantial upstream component still present when the bunch field reaches the crystal.

(b) Laser pulse co-propagating with the bunch electric field along the downstream path. As the laser head exits the crystal, the tail continues to sample the trailing residual field.

Figure 3.13.: Upstream and downstream laser pulse trajectories at two distinct moments. The upstream path is shown with an exaggerated angle only for visual separation of the two trajectories; this angle is not part of the simulation geometry.

The total number of simulation samples is limited by the temporal resolution of the electromagnetic solver, which is determined by the constant time step Δt. The simulation time is therefore discretized as

$$t_i = i\,\Delta t, \quad i = 0, \ldots, N_{\max} - 1, \tag{3.9}$$

where $N_{\max}$ is the total number of available time samples. Since the electromagnetic fields are evaluated at these discrete simulation times, the position of each point of the laser pulse within the crystal must be determined at every time step.

For this purpose, the laser pulse is discretized along its temporal coordinate into M sampling points:

$$\tau_m = m\,\Delta\tau, \quad m = 0, \ldots, M - 1, \tag{3.10}$$

where $\Delta\tau = L_{\text{laser}}/M$ is the temporal spacing between consecutive sampling points and L_{laser} denotes the temporal length of the laser pulse. In practice, L_{laser} is determined experimentally [56].

To describe the propagation of these sampling points, the coordinate z is defined along the crystal axis. The laser is assumed to propagate in the longitudinal z-direction along a trajectory passing through the center of the crystal, i.e., at $x = 0$ and $y = 0$. The coordinate z_{start} denotes the crystal face through which the bunch electric field enters. This face also contains the reflective coating for the chirped laser pulse. The influence of the coating on the bunch field entering the crystal is assumed to be negligible. The opposite crystal face is denoted by z_{end}.

The temporal reference for the laser trajectory is defined by t_{head}, which denotes the instant at which the head of the laser pulse, corresponding to $\tau = 0$, reaches the reflective coating at z_{start} and begins its downstream propagation. The remaining sampling points reach the coating at the shifted times $t_{\text{head}} + \tau_m$.

The laser trajectory consists of an upstream path before reflection and a downstream path after reflection. For the numerical construction, both paths are represented by straight lines extending along the propagation axis, while their physically valid portions are subsequently restricted to the crystal boundaries. As illustrated in Fig. 3.13, before reflection the laser follows the upstream path from

left to right, opposite to the propagation direction of the electron bunch. The position z_i of the laser coordinate τ at the simulation time t_i is given by

$$z_i^{\mathrm{up}} = z_{\mathrm{start}} + v_{\mathrm{L}}\big[(t_{\mathrm{head}} + \tau) - t_i\big], \tag{3.11}$$

where $v_{\mathrm{L}} = c/n_0$ is the laser velocity in the electro-optic medium [2]. At $t_i = t_{\mathrm{head}} + \tau$, the corresponding laser coordinate reaches the reflective coating at z_{start}.

After reflection, the propagation direction is reversed, and the laser follows the downstream path from right to left, co-propagating with the electron bunch, as shown in Fig. 3.13. Its position is then given by

$$z_i^{\mathrm{down}} = z_{\mathrm{start}} + v_{\mathrm{L}}\big[t_i - (t_{\mathrm{head}} + \tau)\big]. \tag{3.12}$$

A calculated position contributes to the phase retardation only when it lies within the physical boundaries of the crystal. Consequently, for each simulation frame indexed by i, the calculated laser position is tested against these boundaries. For the downstream trajectory, the valid position range is

$$z_{\mathrm{start}} \leq z_i^{\mathrm{down}} \leq z_{\mathrm{end}}. \tag{3.13}$$

For the upstream trajectory, the boundary at z_{start} is excluded to prevent the reflection point from being counted twice in the subsequent integration:

$$z_{\mathrm{start}} < z_i^{\mathrm{up}} \leq z_{\mathrm{end}}. \tag{3.14}$$

Each position satisfying the corresponding boundary condition is retained as a valid trajectory point and relabeled using the index j. After this relabeling, t_j denotes the simulation time associated with the retained frame, while $z_j(\tau_m)$ denotes the position inside the crystal of the discrete laser coordinate τ_m. The retained points form the discrete trajectory sets $\mathcal{C}^{\Delta}_{\mathrm{down}}$ and $\mathcal{C}^{\Delta}_{\mathrm{up}}$.

The positions $z_j(\tau_m)$ do not generally coincide with the fixed probe locations used in the CST simulation. The local electric field must therefore be evaluated by interpolation between the two nearest bounding probes. Let z_k and z_{k+1} denote the neighboring probe positions satisfying

$$z_k \leq z_j(\tau_m) \leq z_{k+1}.$$

The interpolated field is then given by

$$E_y\big(z_j(\tau_m), t_j\big) = E_y(z_k, t_j) + \frac{E_y(z_{k+1}, t_j) - E_y(z_k, t_j)}{z_{k+1} - z_k}\left(z_j(\tau_m) - z_k\right). \tag{3.15}$$

$$\Gamma(\tau_m) \approx \frac{2\pi}{\lambda} n_0^3 r_{41} \left(\sum_{j \in \mathcal{C}^{\Delta}_{\mathrm{down}}} \|\Delta z_j\| \, E_y\big(z_j(\tau_m), t_j\big) + \sum_{j \in \mathcal{C}^{\Delta}_{\mathrm{up}}} \|\Delta z_j\| \, E_y\big(z_j(\tau_m), t_j\big) \right). \tag{3.16}$$

For each valid trajectory point j, the quantity $\|\Delta z_j\|$ represents the effective interaction length associated with the interpolated field value $E_y\big(z_j(\tau_m), t_j\big)$.

This discretization defines the sampling of the EOSD signal and is independent of the electromagnetic solver. However, the electric field at these sampling points is obtained from probe data, so the probe spacing must be sufficiently fine to avoid repeated interpolation between the same probe pair. To ensure

[2] The Pockels effect slightly modifies the propagation velocity; this variation is neglected and a constant velocity is assumed.

this, the spatial spacing of the probes, $\Delta z_{\mathrm{probe}}$, must be smaller than or equal to the spatial separation between consecutive laser sampling points, Δz_τ:

$$\Delta z_{\mathrm{probe}} \leq \Delta z_\tau. \tag{3.17}$$

Accordingly, the spatial separation between consecutive laser sampling points is set by the laser propagation velocity inside the crystal, $\Delta z_\tau = v_{\mathrm{L}} \Delta\tau$, with $v_{\mathrm{L}} = c/n_0$. For the central laser wavelength $\lambda_0 \approx 1\,050\,\mathrm{nm}$, the simulation uses the GaP refractive index $n_0 \approx 3.24$ [67], yielding $v_{\mathrm{L}} \approx 9.25 \times 10^7\,\mathrm{m\,s^{-1}}$.

Furthermore, the probe spacing is geometrically defined as $\Delta z_{\mathrm{probe}} = L_{\mathrm{crystal}}/N_{\mathrm{probes}}$, where L_{crystal} is the crystal length. Combining these definitions with Eq. (3.17) yields the condition for the minimum number of required probes:

$$N_{\mathrm{probes}} \geq \frac{L_{\mathrm{crystal}}}{v_{\mathrm{L}} \Delta\tau}. \tag{3.18}$$

For example, for a laser pulse width $L_{\mathrm{laser}} = 47.6\,\mathrm{ps}$ [56] and sampling points $M = 128$, the temporal spacing is $\Delta\tau \approx 0.371\,\mathrm{ps}$, corresponding to a spatial separation of $\Delta z_\tau \approx 3.43 \times 10^{-5}\,\mathrm{m}$. With $L_{\mathrm{crystal}} = 7 \times 10^{-3}\,\mathrm{m}$, the condition in Eq. (3.18) evaluates to

$$N_{\mathrm{probes}} \gtrsim 204. \tag{3.19}$$

Thus, at least 204 probes are required to ensure that each laser sampling point is associated with a distinct interpolation region and to avoid artificial linear correlations in the reconstructed field.

The chosen laser sampling also defines the maximum frequency that can be represented by the discretized signal. The sampling frequency is $f_{\mathrm{s}} = 1/\Delta\tau \approx 2.69\,\mathrm{THz}$, corresponding to the Nyquist frequency

$$f_{\mathrm{Nyquist}} = \frac{f_{\mathrm{s}}}{2} = \frac{1}{2\Delta\tau} \approx 1.34\,\mathrm{THz}. \tag{3.20}$$

The 1 THz upper-frequency limit used in the CST simulations therefore remains below the Nyquist limit of the laser sampling, so the simulated frequency content can be represented without aliasing.

Finally, for each laser coordinate τ_m and each simulation frame t_i, the upstream and downstream positions are computed from Eqs. (3.12) and (3.11). Positions outside the crystal are discarded. For valid positions, the field is obtained by linear interpolation between the neighbouring CST probes and accumulated into the corresponding phase-retardation sum. For a detail explanation of the algorithm, the reader is referred to Appendix A.1.

The phase retardation simulation for the Gaussian bunch profile shown in Fig. 3.7, using the reduced 3D EOSD model shown in Fig. 3.9 and simulated up to 50 GHz, is presented in Fig. 3.14.

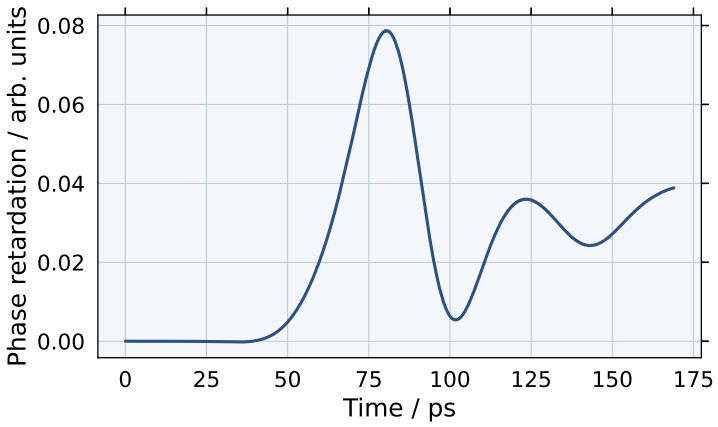


Figure 3.14.: Simulated phase retardation for a Gaussian bunch profile using CST electric fields evaluated at 1000 probe positions inside the crystal. The laser pulse is assumed to be synchronized such that the bunch electric field enters the crystal when the laser pulse reaches the reflective coating.

Figure 3.15 shows the phase retardation simulation for the micro-bunched profile presented in Fig. 3.8a, using the reduced 3D EOSD, simulated up to 1 THz.

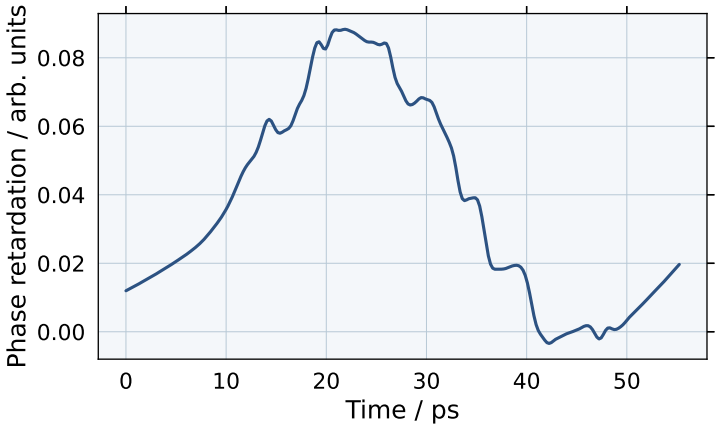


Figure 3.15.: Simulated phase retardation for a bunch profile containing micro-bunching substructures, using CST electric fields evaluated at 1000 probe positions inside the crystal and the reduced 3D EOSD model.

In contrast to the Gaussian case, where the phase retardation closely follows the bunch profile, the response to the micro-bunched distribution shows clear deviations from the original shape, as shown in Fig. 3.16. Although some of the microstructures remain visible in the phase retardation, their shape is modified by the response of the diagnostic setup.

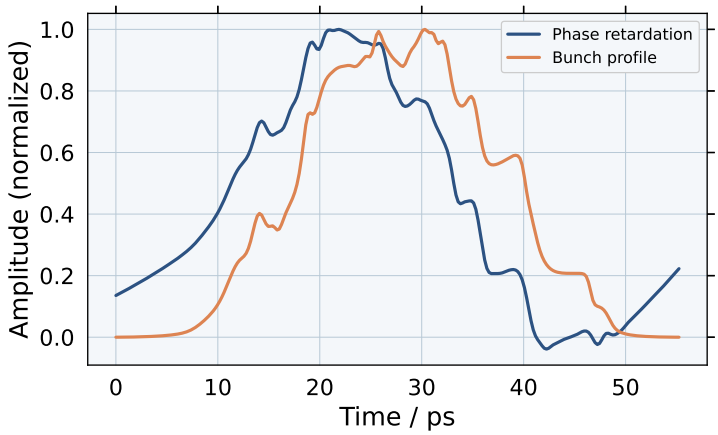


Figure 3.16.: Comparison between the micro-bunched bunch profile and the corresponding phase retardation. The vertical scale is normalized to facilitate comparison of their shapes.

The electric-field data used to generate Fig. 3.15 were computed on the bwUniCluster 3 high-performance computing (HPC) system using one node with 80 central processing unit (CPU) cores. The simulation required a wall-clock time of 15 h, 34 min, 52 s and used 261.58 GB of memory. This illustrates the high computational cost of resolving the full electromagnetic response directly for a short micro-bunched profile, whose fine temporal structures require a large simulated frequency bandwidth. The following section therefore introduces the impulse-response formulation, which replaces repeated full-wave simulations by a convolution based forward model.

3.3.4. EOSD impulse response

Evaluating the simulated phase retardation for arbitrary longitudinal electron bunch profiles using a full three-dimensional electromagnetic solver is computationally expensive and therefore impractical within the phase space reconstruction algorithm. The objective of this subsection is to replace this calculation with an impulse response formulation that directly maps the longitudinal bunch profile, $\rho(t)$, to the phase retardation signal, $\Gamma(\tau)$. In this formulation, the EOSD geometry, material properties, laser trajectory, synchronization, and simulation bandwidth are kept fixed.

For a relativistic electron bunch propagating close to the speed of light, Lorentz contraction strongly compresses the Coulomb field in the longitudinal direction, so the temporal structure of the incident transverse field is closely related to the longitudinal bunch profile, $\rho(t)$. Under the linear approximation of Maxwell's equations for a fixed geometry and linear dielectric media, the generated electromagnetic fields satisfy the principle of superposition. Furthermore, because the geometry and material properties of the EOSD setup are static, a temporal delay in the bunch profile produces the corresponding temporal delay in the internal electric field. Within the linear Pockels approximation and for a fixed laser trajectory, the mapping from $\rho(t)$ to $\Gamma(\tau)$ is therefore treated as approximately linear and time invariant.

Under these assumptions, the phase retardation generated by an arbitrary bunch profile can be expressed as the convolution

$$\Gamma(\tau) = \int_{-\infty}^{\infty} \rho(t')\, h(\tau - t')\, \mathrm{d}t' = (\rho * h)\,(\tau), \tag{3.21}$$

where $h(\tau)$ denotes the impulse response of the complete EOSD forward model. Physically, $h(\tau)$ is the phase retardation signal produced when the longitudinal bunch profile is replaced by a sufficiently short, delta-like excitation.

To derive this impulse response, the longitudinal bunch profile is first written as a continuous superposition of shifted delta functions,

$$\rho(t) = \int_{-\infty}^{\infty} \rho(t')\, \delta(t - t')\, \mathrm{d}t'. \tag{3.22}$$

Let ${E_y}^{\delta}(z, t)$ denote the effective electric field inside the crystal generated by the delta-like bunch profile $\rho(t) = \delta(t)$. By time invariance, the response to a shifted delta function $\delta(t-t')$ is ${E_y}^{\delta}(z, t-t')$. By linearity and superposition, the effective field generated by an arbitrary bunch profile is therefore

$$E_y(z, t) = \int_{-\infty}^{\infty} \rho(t')\, E_y^{\delta}(z, t - t')\, \mathrm{d}t'. \tag{3.23}$$

Equation (3.23) expresses the field inside the crystal as the convolution of the bunch profile with the electromagnetic response to a delta-like excitation. This response naturally includes propagation effects, refraction at interfaces, and reflections from the crystal boundaries and surrounding metallic structures.

To relate the internal field to the diagnostic signal, Eq. (3.23) is substituted into Eq. (3.8), yielding

$$\Gamma(\tau) = \frac{2\pi}{\lambda} n_0^3 r_{41} \int_{\mathcal{C}} \int_{-\infty}^{\infty} \rho(t')\, E_y^{\delta}\big(z, t(z;\tau) - t'\big)\, \mathrm{d}t'\, \mathrm{d}z. \tag{3.24}$$

For a fixed laser trajectory, the laser propagation introduces the shift relation

$$t(z;\tau) = t(z;0) + \tau. \tag{3.25}$$

Consequently,

$$t(z;\tau) - t' = t(z;0) + \tau - t' = t(z;\tau - t'). \tag{3.26}$$

Using this relation and exchanging the order of integration, Eq. (3.24) becomes

$$\Gamma(\tau) = \int_{-\infty}^{\infty} \rho(t') \left[\frac{2\pi}{\lambda} n_0^3 r_{41} \int_{\mathcal{C}} E_y^{\delta}\big(z, t(z;\tau - t')\big)\, \mathrm{d}z \right] \mathrm{d}t'. \tag{3.27}$$

The term in brackets is the phase retardation produced by the delta-like bunch profile and therefore defines the EOSD impulse response:

$$h(\tau) = \frac{2\pi}{\lambda} n_0^3 r_{41} \int_{\mathcal{C}} E_y^{\delta}\big(z, t(z;\tau)\big)\, \mathrm{d}z. \tag{3.28}$$

Substituting Eq. (3.28) into Eq. (3.27) recovers the convolutional forward model in Eq. (3.21).

Equation (3.21) is the central result of the impulse response formulation. It shows that, within the linear response and time invariance approximations, the EOSD phase retardation can be obtained by convolving the longitudinal bunch profile with the impulse response of the complete EOSD system. The impulse response h encodes the electromagnetic transfer characteristics of the crystal and surrounding geometry, together with the laser trajectory and pulse length used in the phase-retardation calculation.

In the numerical implementation, the impulse response is obtained from a finite width, impulse-like bunch simulated with the CST Wakefield Solver. Its validity is therefore limited by the selected simulation bandwidth, temporal resolution, and finite simulation window. The bunch profiles used in the convolution must be represented with the same temporal sampling, while the EOSD geometry, material properties, laser trajectory, and synchronization must remain unchanged. In addition, edge regions affected by the finite convolution window are excluded from the valid part of the calculated signal.

3.3.4.1. Numerical evaluation of the impulse response

In practice, the EOSD impulse response $h(\tau)$ is obtained directly from the CST Wakefield solver simulations by applying the same phase retardation integration algorithm used for arbitrary bunch profiles. The main technical limitation of this numerical calculation is the finite simulation bandwidth, which is constrained by the spatial mesh resolution, the solver time step, and the available computational resources. Consequently, the computed impulse response is not an ideal, infinitely broadband delta response, but a band-limited approximation.

This limitation is consistent with the EOSD simulations discussed above. A simulation driven by a given bunch profile can only contain the field components resolved within the simulated frequency range. Consequently, an impulse response computed with a given bandwidth can only be used to simulate EOSD signals for bunch profiles whose relevant spectral content lies within the same bandwidth. In addition, the same laser trajectory must be used, because the impulse response includes both the electromagnetic response of the EOSD geometry and the laser propagation itself.

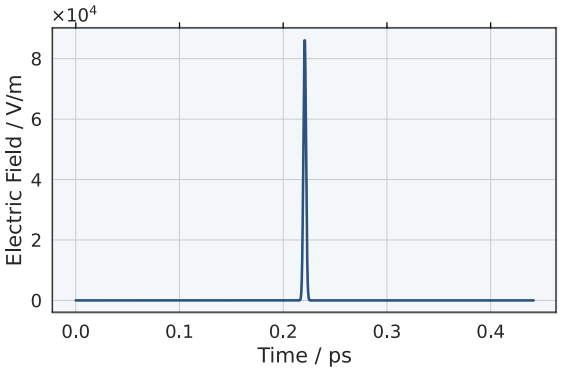
To compute the phase retardation by discrete convolution of the bunch profile with the impulse response, both signals must be sampled with the same temporal spacing $\Delta\tau$. The phase retardation signal must then be extracted from the *valid* part of the convolution, where the bunch profile fully overlaps with the impulse response and no artificial contribution from zero-padding enters the result.

The impulse response signal used in the convolution is evaluated over a temporal window longer than the effective experimental laser window. In this work, the impulse response window is chosen to be twice the laser length:

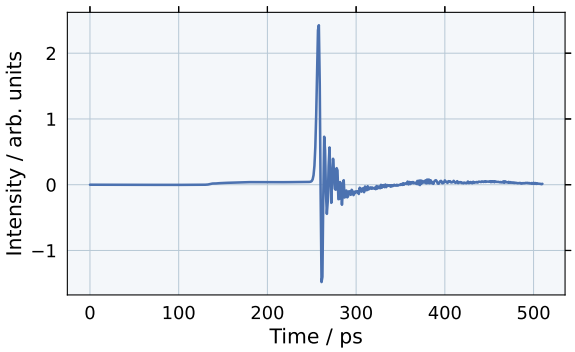
$$L_{\mathrm{impulse}} = 2L_{\mathrm{laser}}. \tag{3.29}$$

This choice ensures that the phase retardation recovered from the convolution is taken from the *valid* region, where the bunch profile fully overlaps with the impulse response. The desired signal has length L_{laser} and is extracted from the center of the full convolution result. The doubled impulse response window keeps trailing field contributions inside the valid convolution region, avoiding truncation and boundary padding artifacts.

Figure 3.17b shows the impulse response obtained for the reduced 3D EOSD model shown in Fig. 3.9, using a CST simulation bandwidth of 1 THz. The response is computed from the impulse-like Gaussian excitation shown in Fig. 3.17a. The effective laser length used in the impulse-response calculation is chosen to match the signal length of the bunch profile shown in Fig. 3.8a.



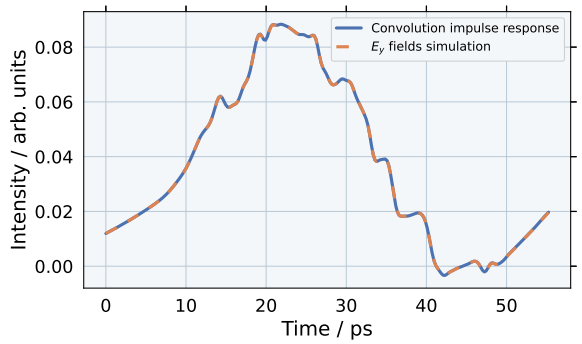
(a) Short Gaussian excitation used as a delta-like input, with $\sigma = 0.001\,25\,\mathrm{ps}$.



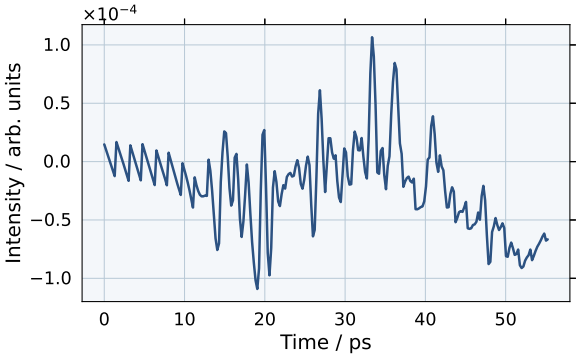
(b) EOSD impulse response h obtained by applying the phase-retardation calculation to the short Gaussian excitation.

Figure 3.17.: Numerical impulse-response evaluation for the reduced 3D EOSD model with a CST simulation bandwidth of 1 THz. The short Gaussian excitation approximates a delta-like bunch profile, and the resulting phase-retardation signal defines the impulse response h.

Figure 3.18 shows very good agreement between the phase retardation computed directly from the CST electric fields and the phase retardation obtained by convolving the bunch profile with the impulse response.



(a) Phase retardation computed by direct CST-based simulation compared with the result obtained by convolution with the EOSD impulse response.



(b) Pointwise difference between the direct CST-based phase retardation and the impulse-response convolution result.

Figure 3.18.: Comparison between the phase retardation obtained from the direct CST-based simulation and from the convolution with the EOSD impulse response. The RMS difference between the two calculations is 4.19×10^{-5}.

The discrete convolution can be efficiently formulated as a matrix-vector product, where the transfer operator G is a Toeplitz matrix [68] constructed from the discrete impulse response h. Thus, the phase retardation signal is given by

$$\Gamma = G\rho, \tag{3.30}$$

where the matrix G represents the valid convolution with the EOSD impulse response.

In an unoptimized Python implementation, the computation of the convolution using the Toeplitz matrix required approximately 34 ms, including the construction of G from h. This is several orders of

magnitude faster than the direct CST-based approach, where the electric field must first be recomputed with a full electromagnetic simulation before the phase retardation can be evaluated.

Finally, the matrix representation defines a linear EOSD transfer operator, allowing the convolution-based phase retardation model to be directly integrated into the phase space density reconstruction algorithm.

4. The Vlasov–Fokker–Planck equation solution

The phase space density formalism and the dimensionless Vlasov–Fokker–Planck equation for longitudinal beam dynamics in a storage ring were introduced in Chapter 2. The Hamiltonian part describes the conservative transport generated by the RF system and the collective wake potential, while the Fokker–Planck (FP) terms account for radiation damping, and quantum excitation in the energy coordinate.

This chapter introduces an operator-splitting discretization of the VFPE, in which the phase space density is advanced through a sequence of Hamiltonian transport and Fokker–Planck operators.

4.1. Operator splitting based on particle conservation

The solution of the VFPE is formulated as the time evolution of the phase space density under Hamiltonian and Fokker–Planck transport mechanisms. Equation (2.10) expresses the conservation of particle content within a transported phase space region, whereas Eq. (2.26) enforces the same conservation law for drift and diffusion transport by integrating the contributions from all possible initial momenta, each weighted by its transition density probability. These relations lead to an operator formulation of the VFPE evolution, following the approach of Warnock and Ellison [44].

To make the notation explicit, let the vector

$$\mathbf{x} = (q, p) \tag{4.1}$$

denote a point in the two-dimensional phase space. Using the flow notation introduced in Eqs. (2.1) and (2.2), the Hamiltonian flow over a time interval Δt maps an initial state $\mathbf{x}_0$ to

$$\mathbf{x} = f^{\Delta t}(\mathbf{x}_0), \tag{4.2}$$

where, for compact notation, $\mathbf{x}_0 = \mathbf{x}(t)$ denotes the initial state and $\mathbf{x} = \mathbf{x}(t + \Delta t)$ the state after the time interval Δt. Accordingly, an arbitrary initial region $\mathcal{M}_i \subset \mathcal{M}$ is mapped into $f^{\Delta t}(\mathcal{M}_i)$:

$$\int_{f^{\Delta t}(\mathcal{M}_i)} \psi(\mathbf{x}, t + \Delta t)\, \mathrm{d}\mathbf{x} = \int_{\mathcal{M}_i} \psi(\mathbf{x}_0, t)\, \mathrm{d}\mathbf{x}_0, \tag{4.3}$$

where $\mathrm{d}x = \mathrm{d}q\, \mathrm{d}p$.

Using Eq. (4.2) to change the integration variable on the left-hand side from the evolved point $x \in f^{\Delta t}(\mathcal{M}_i)$ to the initial point $\mathbf{x}_0 \in \mathcal{M}_i$ maps the integration domain back to the initial phase space region:

$$\int_{\mathcal{M}_i} \psi\big(f^{\Delta t}(\mathbf{x}_0), t + \Delta t\big)\, \big|\det J^{\Delta t}(\mathbf{x}_0)\big|\, \mathrm{d}\mathbf{x}_0 = \int_{\mathcal{M}_i} \psi(\mathbf{x}_0, t)\, \mathrm{d}\mathbf{x}_0, \tag{4.4}$$

with

$$J^{\Delta t}(\mathbf{x}_0) = \frac{\partial f^{\Delta t}(\mathbf{x}_0)}{\partial \mathbf{x}_0} \tag{4.5}$$

the Jacobian J of the flow map. Since $\mathcal{M}_i$ is arbitrary, the local density relation is

$$\psi(\mathbf{x}, t + \Delta t) = \frac{\psi(\mathbf{x}_0, t)}{|\det J^{\Delta t}(\mathbf{x}_0)|}, \qquad \mathbf{x} = f^{\Delta t}(\mathbf{x}_0). \tag{4.6}$$

This relation defines the Perron–Frobenius density transport [69]. For the Hamiltonian flow this relation is deterministic, invertible, and volume preserving [44]. Therefore:

$$\left|\det J^{\Delta t}(\mathbf{x}_0)\right| = 1, \tag{4.7}$$

and the Hamiltonian flow reduces to the inverse-map rule

$$\psi(\mathbf{x}, t + \Delta t) = \psi\big(f^{-\Delta t}(\mathbf{x}), t\big)\,. \tag{4.8}$$

The Hamiltonian transport operator is denoted by $\mathcal{P}_{\Delta t}$, where Δt is the time interval to advances the density from time t to time $t + \Delta t$:

$$\psi(\mathbf{x}, t + \Delta t) = \mathcal{P}_{\Delta t} \circ \psi(\mathbf{x}_0, t). \tag{4.9}$$

Here, the flow $f^{\Delta t}$ may be nonlinear, so the transported state $f^{\Delta t}(\mathbf{x}_0)$ depends on the initial point $\mathbf{x}_0$.

For the Hamiltonian flow, the density at the target point x is obtained from the density at the corresponding source point:

$$\mathcal{P}_{\Delta t} \circ \psi(\mathbf{x}_0, t) = \psi(f^{-\Delta t}(\mathbf{x}), t). \tag{4.10}$$

This inverse-map relation is the basis of the backward interpolation operator used to advance the Vlasov part.

For the Fokker–Planck momentum increment, the density update is written in terms of the transition probability for the momentum displacement. Applying particle conservation to an arbitrary region and rewriting Eq. (2.26) with the notation introduced in this section gives

$$\int_{f^{\Delta t}(\mathcal{M}_i)} \psi(\mathbf{x}, t + \Delta t)\,\mathrm{d}x = \int_{\mathcal{M}_i} \int_{-\infty}^{+\infty} \psi(\mathbf{x}_0 - \delta\mathbf{x}_p, t)\, P(\delta\mathbf{x}_p \mid \mathbf{x}_0 - \delta\mathbf{x}_p, \Delta t)\,\mathrm{d}(\delta\mathbf{x}_p)\,\mathrm{d}\mathbf{x}_0. \tag{4.11}$$

Here, $P(\delta\mathbf{x}_p \mid \mathbf{x}_0 - \delta\mathbf{x}_p, \Delta t)$ is the transition probability density that a particle initially located at $\mathbf{x}_0 - \delta\mathbf{x}_p$ undergoes the momentum increment δx_p during the interval Δt.

Since Eq. (4.11) expresses this particle balance for an arbitrary region, the corresponding local density relation is

$$\psi(\mathbf{x}, t + \Delta t) = \int_{-\infty}^{+\infty} \psi(\mathbf{x}_0 - \delta\mathbf{x}_p, t)\, P(\delta\mathbf{x}_p \mid \mathbf{x}_0 - \delta\mathbf{x}_p, \Delta t)\,\mathrm{d}(\delta\mathbf{x}_p). \tag{4.12}$$

Unlike the Hamiltonian operator, this transition operator is not expressed in terms of the inverse flow. Its action is defined by

$$D_{\Delta t} \circ \psi(\mathbf{x}_0) = \int_{-\infty}^{+\infty} \psi(\mathbf{x}_0 - \delta\mathbf{x}_p, t)\, P(\delta x_p \mid \mathbf{x}_0 - \delta\mathbf{x}_p, \Delta t)\, \mathrm{d}(\delta\mathbf{x}_p). \tag{4.13}$$

Following [44], $D_{\Delta t}$ is identified as the FP operator for time interval Δt. Using Eq. (2.34), Eq. (4.13) becomes:

$$D_{\Delta t} \circ \psi(\mathbf{x}_0) \approx \psi(\mathbf{x}_0, t) + \beta_d \Delta t \frac{\partial}{\partial p} \left(p\psi(\mathbf{x}_0, t) + \frac{\partial \psi(\mathbf{x}_0, t)}{\partial p} \right), \tag{4.14}$$

where the coefficient β_d follows from the dimensionless damping–diffusion balance in Eq. (2.136).

The operator-splitting approach is motivated by the Hamiltonian structure of the equations of motion, which allows the flow to be separated into simpler Hamiltonian substeps.

If the Hamiltonian is decomposed into several contributions:

$$H = H^1 + \cdots + H^l, \tag{4.15}$$

then the corresponding equations of motion split into the sum of l Hamiltonian subsystems generated by each H^i. Each subproblem is therefore itself Hamiltonian, and its phase space transport defines a corresponding Perron–Frobenius operator $\mathcal{P}_{\Delta t}{}^i$. This is the Hamiltonian splitting principle described by Blanes et al. [70, p. 10].

Applying the Hamiltonian substeps sequentially gives a first-order Lie–Trotter approximation of the deterministic density advance:

$$\mathcal{P}_{\Delta t} \approx \mathcal{P}^l_{\Delta t} \circ \cdots \circ \mathcal{P}^1_{\Delta t}. \tag{4.16}$$

The result of each substep provides the input state for the next one. Since the sub-operators generally do not commute, the ordering fixes the particular first-order splitting sequence used for the Hamiltonian transport.

The same splitting principle is then extended to FP density update. The FP contribution is not represented by a phase space flow, because it redistributes the density through drift and diffusion rather than mapping each initial point to a unique final point. It is therefore included as a density operator $D_{\Delta t}$. Applying the deterministic Perron–Frobenius transports first and the FP update afterwards gives the one-step splitting approximation

$$\psi(\mathbf{x}, t + \Delta t) \approx D_{\Delta t} \circ \mathcal{P}^l_{\Delta t} \circ \cdots \circ \mathcal{P}^1_{\Delta t} \circ \psi(\mathbf{x}_0, t). \tag{4.17}$$

This is one possible splitting sequence. The next section presents the operators and ordering used in Inovesa.

4.2. Inovesa solver

Inovesa is an efficient Vlasov–Fokker–Planck (VFP) solver that enables accurate simulations of phase space evolution, capturing phenomena such as CSR and the micro-bunching instability on standard desktop hardware. The original C++ code base [71], developed by Schönfeldt at Karlsruhe Institute of Technology (KIT) [72], combined a discretized longitudinal phase space model with optimized numerical operators and parallelization, making systematic micro-bunching studies practical on standard workstations within minutes.

This section presents the mathematical foundations of the numerical operator-based phase space advance in Inovesa, together with a brief overview of its practical implementation.

4.2.1. Phase space grid coordinates

The continuous phase space coordinates are mapped to grid coordinates. For a point (q, p), the corresponding grid values are defined as

$$x_r = \frac{q - q_{min}}{\Delta q}, \qquad y_r = \frac{p - p_{min}}{\Delta p}, \qquad x_r, y_r \in \mathbb{R}_{\geq 0} \tag{4.18}$$

where q_{min} and p_{min} denote the canonical coordinates at the origin of the discrete grid, corresponding to $x_r = 0$ and $y_r = 0$, respectively, while Δq and Δp are the grid spacings along the two coordinate axes.

The phase space density is stored at integer grid indices

$$x, y \in \mathbb{N}_{\geq 0}. \tag{4.19}$$

The canonical coordinates represented by a grid node (x, y) are

$$q_x = q_{\min} + x\Delta q, \qquad p_y = p_{\min} + y\Delta p. \tag{4.20}$$

For a general real position (x_r, y_r), the integer grid nodes

$$x = \lfloor x_r \rfloor, \qquad y = \lfloor y_r \rfloor \tag{4.21}$$

identify the lower grid node of the cell containing that position.

4.2.2. Lagrange polynomial interpolation

The inverse flow maps each grid node (x, y) to a real-valued source position $({x_r}', {y_r}')$, which generally does not coincide with a node of the discrete phase space grid. The density at this position must therefore be reconstructed from the surrounding grid values, as shown in Fig. 4.1.

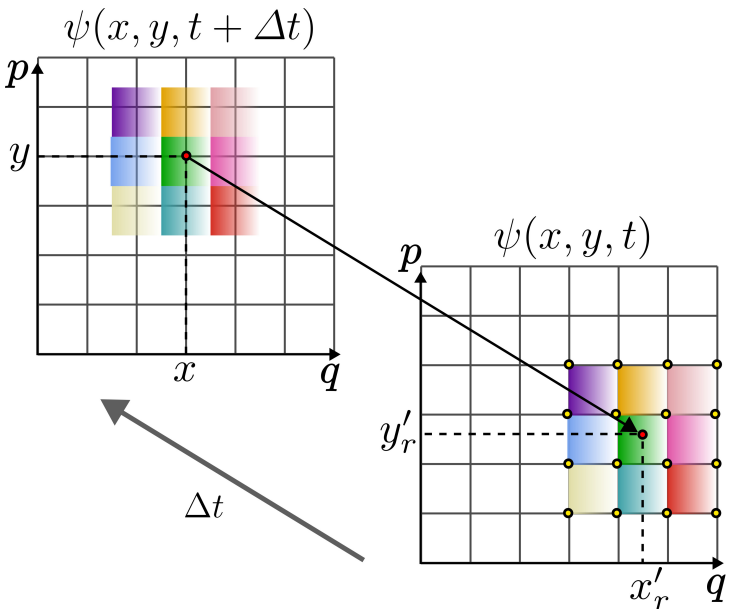


Figure 4.1.: Inverse mapping of the target grid node (x, y) at $t + \Delta t$ to the corresponding source position $({x_r}', {y_r}')$ at time t. The density at the inverse-mapped position is reconstructed by interpolation from neighboring grid points. The yellow nodes illustrate one possible selection of interpolation points.

Inovesa reconstructs the density at the inverse-mapped position using spline interpolation, implemented through cubic Lagrange polynomial interpolation with four interpolation points. Although linear and quadratic interpolation schemes are also available in Inovesa, the cubic scheme is used here to illustrate the procedure.

For clarity, the interpolation is first described along the x-axis while the y-coordinate is held fixed. Cubic Lagrange interpolation uses four neighboring grid nodes x_i[1], with $i = 0, \ldots, 3$, to construct the interpolation polynomial $P(x_r)$.

The polynomial is written as a linear combination of the corresponding Lagrange basis functions $L_i(x_r)$, weighted by the phase space density values $\psi(x_i, y)$:

$$P(x_r, y) = \sum_{i=0}^{3} \psi(x_i, y, t)\, L_i(x_r). \tag{4.22}$$

Each basis polynomial $L_i(x_r)$ is defined by

$$L_i(x_r) = \prod_{\substack{j=0 \\ j \neq i}}^{3} \frac{x_r - x_j}{x_i - x_j}, \tag{4.23}$$

and is constructed to equal unity at its associated grid node x_i and zero at the other three points. Consequently:

$$L_i(x_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{4.24}$$

[1] For nodes outside the grid boundary, Inovesa sets the value equal to the nearest edge value.

The two-dimensional interpolation is constructed by applying the one-dimensional interpolation procedure successively along each coordinate direction. For cubic interpolation, this results in a bi-cubic interpolation.

$$P(x_r, y_r) = \sum_{j=0}^{3} \sum_{i=0}^{3} \psi(x_i, y_j, t)\, L_i(x_r)\, L_j(y_r) \tag{4.25}$$

The phase space density at time $t + \Delta t$ is reconstructed at each target grid point (x, y) from the 4×4 neighborhood surrounding the inverse-mapped position $(x_r{}', y_r{}')$:

$$\psi(x, y, t + \Delta t) \approx \sum_{i=0}^{3} \sum_{j=0}^{3} \psi(x_i, y_j, t)\, L_i(x_r')\, L_j(y_r'). \tag{4.26}$$

As illustrated in Fig. 4.2, cubic interpolation is first applied along the x-direction to each of the four rows in the 4×4 neighborhood. This produces four intermediate values $\psi(x_r{}', y_j, t)$, shown as gray points. These values are then interpolated along the y-direction at $y_r{}'$, yielding the phase space density $\psi(x_r{}', y_r{}', t)$ at the inverse-mapped position.

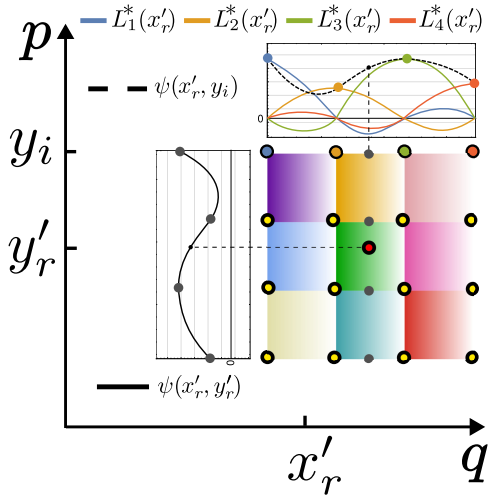


Figure 4.2.: Bi-cubic Lagrange interpolation at the inverse-mapped position $(x_r{}', y_r{}')$. Cubic interpolation is first performed along the x-direction for each of the four rows, producing the intermediate values $\psi(x_r{}', y_j, t)$ shown in gray. These values are subsequently interpolated along the y-direction to obtain $\psi(x_r{}', y_r{}', t)$. The upper inset shows the Lagrange basis contributions evaluated at $x_r{}'$, where the superscript $*$ indicates that each basis function has been weighted by the phase space density at its corresponding grid node.

The implementation of the Fokker–Planck operator requires the first and second derivatives of the Lagrange basis along the y-direction. Applying the product rule to the cubic Lagrange basis gives the following first and second derivatives [73],

$$\frac{\mathrm{d}L_i}{\mathrm{d}y}(y_r) = \sum_{\substack{k=0 \\ k \neq i}}^{3} \left[\frac{1}{y_i - y_k} \prod_{\substack{j=0 \\ j \neq i,k}}^{3} \frac{y_r - y_j}{y_i - y_j} \right], \tag{4.27}$$

$$\frac{\mathrm{d}^2 L_i}{\mathrm{d}y^2}(y_r) = \sum_{\substack{k=0 \\ k\neq i}}^{3} \frac{1}{y_i - y_k} \left[\sum_{\substack{\ell=0 \\ \ell\neq i,k}}^{3} \left(\frac{1}{y_i - y_\ell} \prod_{\substack{j=0 \\ j\neq i,k,\ell}}^{3} \frac{y_r - y_j}{y_i - y_j} \right) \right]. \tag{4.28}$$

As a final implementation detail, cubic Lagrange interpolation is not guaranteed to preserve either non-negativity or the total particle number. Inovesa therefore provides optional safeguards: negative interpolated values can be clamped to zero, and the phase space density can be re-normalized at a configurable interval [72, 74].

4.2.3. Inverse-map problem

The inverse-map rule in Eq. (4.8) requires the source coordinate of every target grid node. In the following, the target node at $\theta_{n+1} = \theta_n + \Delta\theta$ is denoted by the arrival point (q_a, p_a), and its source coordinate at θ_n by the departure point (q_d, p_d). Using the flow notation of Eq. (2.6), the exact departure point satisfies

$$\begin{pmatrix} q_d \\ p_d \end{pmatrix} = \begin{pmatrix} q_a \\ p_a \end{pmatrix} - \int_{\theta_n}^{\theta_{n+1}} \begin{pmatrix} f^q(q,p,\theta) \\ f^p(q,p,\theta) \end{pmatrix} \mathrm{d}\theta. \tag{4.29}$$

Although the rates of change f^q and f^p are known, their evaluation inside the integral requires the trajectory connecting the departure and arrival points. In the inverse-mapping formulation, the arrival point is fixed by the grid node, whereas the departure point is unknown. Consequently, the trajectory cannot be evaluated directly without further approximation.

Evaluating its rate only at the arrival point would give the first-order Taylor approximation

$$\begin{pmatrix} q_d \\ p_d \end{pmatrix} = \begin{pmatrix} q_a \\ p_a \end{pmatrix} - \begin{pmatrix} f^q(q_a, p_a, \theta + \Delta\theta) \\ f^p(q_a, p_a, \theta + \Delta\theta) \end{pmatrix} \Delta\theta + \mathcal{O}(\Delta\theta^2). \tag{4.30}$$

Keeping only the first-order term gives the reverse-time mapping

$$\begin{pmatrix} q_d \\ p_d \end{pmatrix} \approx \begin{pmatrix} q_a \\ p_a \end{pmatrix} - \left. \begin{pmatrix} \mathrm{d}q/\mathrm{d}\theta \\ \mathrm{d}p/\mathrm{d}\theta \end{pmatrix} \right|_{(q_a, p_a, \theta_{n+1})} \Delta\theta. \tag{4.31}$$

Substituting the Hamiltonian equations stated in Eq. (2.123) into Eq. (4.31) gives

$$\begin{pmatrix} q_d \\ p_d \end{pmatrix} \approx \begin{pmatrix} q_a \\ p_a \end{pmatrix} - \left. \begin{pmatrix} \partial H/\partial p \\ -\partial H/\partial q \end{pmatrix} \right|_{(q_a, p_a, \theta_{n+1})} \Delta\theta. \tag{4.32}$$

For the linear synchrotron Hamiltonian of Eq. (2.125), the exact simultaneous flow over the angular step $\Delta\theta$ is

$$\begin{pmatrix} q_a \\ p_a \end{pmatrix} = \begin{pmatrix} \cos\Delta\theta & \sin\Delta\theta \\ -\sin\Delta\theta & \cos\Delta\theta \end{pmatrix} \begin{pmatrix} q_d \\ p_d \end{pmatrix}. \tag{4.33}$$

The departure point required by the inverse-map update is obtained with the inverse rotation:

$$\begin{pmatrix} q_d \\ p_d \end{pmatrix} = \begin{pmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{pmatrix} \begin{pmatrix} q_a \\ p_a \end{pmatrix}. \tag{4.34}$$

Expanding the components gives

$$q_d = q_a \cos\Delta\theta - p_a \sin\Delta\theta, \tag{4.35}$$
$$p_d = q_a \sin\Delta\theta + p_a \cos\Delta\theta. \tag{4.36}$$

For small angular steps:

$$\sin\Delta\theta \approx \tan\Delta\theta \approx \Delta\theta, \qquad \cos\Delta\theta \approx 1. \tag{4.37}$$

The inverse rotation then reduces to the first-order inverse-map relation

$$q_d \approx q_a - p_a\Delta\theta, \tag{4.38}$$
$$p_d \approx p_a + q_a\Delta\theta. \tag{4.39}$$

The same result is obtained by substituting Eq. (2.125) into Eq. (4.32). Thus, the Hamiltonian formulation and the small-angle expansion of the exact rotation give the same first-order inverse mapping.

In the split-map implementation used in Inovesa, the linear inverse mapping is written as

$$q_d \approx q_a - p_a\Delta\theta, \qquad p_d \approx p_a + q_a \tan\Delta\theta. \tag{4.40}$$

The factor $\tan\Delta\theta$ is the finite-step coefficient used for the energy-coordinate shear, where q is kept fixed and p is updated. Since $\tan\Delta\theta = \Delta\theta + \mathcal{O}(\Delta\theta^3)$, this form remains first-order consistent with the small-angle inverse rotation. For the nonlinear inverse mapping, the first-order backward trace in Eq. (4.31) is used.

4.2.4. Implementation of Hamiltonian operators

The Hamiltonian operators in Inovesa combine the inverse mapping described in Section 4.2.3 with the Lagrange interpolation introduced in Section 4.2.2. For each target grid node, the inverse Hamiltonian map determines a generally real-valued source coordinate in the previous density. The density at this source coordinate is then reconstructed using the Lagrange polynomial in Eq. (4.22).

Following the Hamiltonian decomposition of Eq. (2.135), the implementation uses the collective effects (wake potential) operator K, and the decoupled synchrotron motion operators; the RF operator R_ϵ, and the drift operator R_φ. The subscript Δt, indicating the corresponding time increment, is omitted from the operator notation for simplicity. The collective kick is followed by the RF kick and then by the drift:

$$\mathcal{P} = R \circ K, \qquad R = R_\varphi \circ R_\epsilon. \tag{4.41}$$

The inverse maps are constructed from displacements expressed in grid coordinates. A displacement in canonical coordinates, $(\delta q, \delta p)$, corresponds to the grid displacement

$$\Delta x_r = \frac{\delta q}{\Delta q}, \qquad \Delta y_r = \frac{\delta p}{\Delta p}. \tag{4.42}$$

For a first-order step in the angular time θ, these displacements are

$$\Delta x_r = \frac{1}{\Delta q}\frac{\mathrm{d}q}{\mathrm{d}\theta}\Delta\theta, \qquad \Delta y_r = \frac{1}{\Delta p}\frac{\mathrm{d}p}{\mathrm{d}\theta}\Delta\theta. \tag{4.43}$$

Since Δx_r and Δy_r describe the forward motion, the source coordinates for the inverse map are obtained by subtraction:

$$x'_r = x - \Delta x_r, \qquad y'_r = y - \Delta y_r. \tag{4.44}$$

The following subsections specify these grid displacements for the linear and nonlinear synchrotron-motion terms and for the collective contribution.

4.2.4.1. Linear synchrotron motion

For the linear synchrotron motion implemented in Inovesa, the canonical displacements are given by Eq. (4.40):

$$\delta q = p_y \Delta\theta, \qquad \delta p = -q_x \tan\Delta\theta, \tag{4.45}$$

where p_y is the canonical momentum associated with the row y, and q_x is the longitudinal position associated with the column x on the target gride, as defined in Eq. (4.20).

The grid displacements are given by the relation of Eq. (4.43). For the drift map R_φ, the displacement at grid row y is

$$\Delta x_r^R(y) = \frac{p_y}{\Delta q}\Delta\theta. \tag{4.46}$$

For the RF map R_ϵ, the signed forward displacement is

$$\Delta y_r^R(x) = -\frac{q_x}{\Delta p}\tan\Delta\theta. \tag{4.47}$$

Since the forward synchrotron motion update applies R_ϵ before R_φ, the inverse-mapping performs the drift first and then the RF update. Therefore, the longitudinal coordinate used in the RF inverse-map is the coordinate obtained after the inverse-drift step. For a target node (x, y), this gives

$$\begin{aligned} x'_r &= x - \Delta x_r^R(y) = x - \frac{p_y}{\Delta q}\Delta\theta, \\ q_x &= q_{\min} + x'_r \Delta q, \\ y'_r &= y - \Delta y_r^R(x) = y + \frac{q_x}{\Delta p}\tan\Delta\theta. \end{aligned} \tag{4.48}$$

For a centered matched grid, the indices x_0 and y_0 denote the grid positions corresponding to $q = 0$ and $p = 0$, respectively. The canonical coordinates are then written as

$$q_x = (x - x_0)\Delta q, \qquad p_y = (y - y_0)\Delta p, \qquad \Delta q = \Delta p. \tag{4.49}$$

In this case, the linear synchrotron-motion displacements reduce to

$$\Delta x_r^R(y) = (y - y_0)\Delta\theta, \qquad \Delta y_r^R(x) = -(x - x_0)\tan\Delta\theta. \tag{4.50}$$

These are the linear position and energy displacements used to construct the interpolation maps [72, 75].

4.2.4.2. Nonlinear synchrotron motion

For the nonlinear synchrotron motion implemented in Inovesa, the canonical displacements are given by Eq. (4.31):

$$\delta q = \frac{dq}{d\theta}\Delta\theta, \qquad \delta p = \frac{dp}{d\theta}\Delta\theta, \tag{4.51}$$

Using the canonical coordinates defined in Eq. (2.118) and the nonlinear rate of change of the energy deviation in Eq. (2.89), the rate of change of the momentum coordinate with respect to the angular time is

$$\frac{\mathrm{d}p}{\mathrm{d}\theta} = \frac{f_{\mathrm{rev}}}{2\pi f_{s,0}\sigma_{E,0}}\left[eV_{\mathrm{RF}}\sin(\phi_s + \varphi(q_x)) - U_0\right]. \tag{4.52}$$

The nonlinear RF map R_ϵ retains the sinusoidal dependence of the RF voltage instead of replacing it by its slope at the synchronous phase. The corresponding energy-grid displacement is

$$\Delta y_r^R(x) = \frac{f_{\mathrm{rev}}}{2\pi f_{s,0}\sigma_{E,0}\Delta p}\left[eV_{\mathrm{RF}}\sin(\phi_s + \varphi(q_x)) - U_0\right]\Delta\theta. \tag{4.53}$$

Here $\varphi(q_x)$ is the RF-phase deviation associated with the canonical coordinate q_x.

The drift map R_φ can be extended beyond the linear relation governed by the phase-slip factor by including the dependence of momentum compaction on the relative momentum deviation. Starting from Eq. (2.88), the longitudinal phase rate is

$$\dot{\varphi} = \frac{2\pi h f_{\mathrm{rev}}}{\beta^2 E_0}\eta(\delta)\,\varepsilon, \tag{4.54}$$

where the slip factor depends on the relative momentum deviation through $\eta(\delta)$. The rate of change of the canonical position coordinate is

$$\frac{\mathrm{d}q}{\mathrm{d}\theta} = \frac{2\pi h f_{\mathrm{rev}}\sigma_{E,0}}{\beta^2 E_0\sigma_{\varphi,0}\Omega_s}\eta(\delta)\,p. \tag{4.55}$$

The normalization is chosen such that the linear reference case satisfies $\mathrm{d}q/\mathrm{d}\theta = p$. Therefore:

$$\frac{2\pi h f_{\mathrm{rev}}\sigma_{E,0}}{\beta^2 E_0\sigma_{\varphi,0}\Omega_s}\eta_0 = 1, \tag{4.56}$$

where η_0 is the reference slip factor. Hence, for nonlinear momentum compaction:

$$\frac{\mathrm{d}q}{\mathrm{d}\theta} = p\,\frac{\eta(\delta)}{\eta_0}. \tag{4.57}$$

In the ultra-relativistic approximation, $\eta(\delta) \simeq \alpha_{\mathrm{c}}(\delta)$ and $\eta_0 \simeq \alpha_0$. Using the momentum-compaction expansion

$$\alpha_c(\delta) = \alpha_0 + \alpha_1\delta + \alpha_2\delta^2, \tag{4.58}$$

together with the ultra-relativistic approximation of Eq. (2.85):

$$\delta = \frac{\varepsilon}{E_0} = \frac{\sigma_{E,0}}{E_0} p, \tag{4.59}$$

the rate of change of the longitudinal coordinate with respect to the angular time is

$$\frac{\mathrm{d}q}{\mathrm{d}\theta} = p \left[1 + \frac{\alpha_1}{\alpha_0} \frac{\sigma_{E,0}}{E_0} p + \frac{\alpha_2}{\alpha_0} \left(\frac{\sigma_{E,0}}{E_0} p \right)^2 \right]. \tag{4.60}$$

The corresponding position grid displacement is therefore

$$\Delta x_r^R(y) = \frac{p_y \Delta\theta}{\Delta q} \left[1 + \frac{\alpha_1}{\alpha_0} \frac{\sigma_{E,0}}{E_0} p_y + \frac{\alpha_2}{\alpha_0} \left(\frac{\sigma_{E,0}}{E_0} p_y \right)^2 \right]. \tag{4.61}$$

Setting $\alpha_1 = \alpha_2 = 0$ recovers the linear drift displacement in Eq. (4.46).

With the synchrotron motion ordering $R = R_\varphi \circ R_\epsilon$, the drift relation first determines the longitudinal source coordinate associated with the target node. This coordinate is then used in the RF relation, where q is kept fixed. For a target node (x, y), the nonlinear inverse-map coordinates are

$$\begin{aligned} x'_r &= x - \Delta x_r^R(y), \\ q_x &= q_{\min} + x'_r \Delta q, \\ y'_r &= y - \Delta y_r^R(x). \end{aligned} \tag{4.62}$$

The resulting real-valued source coordinates are evaluated with the Lagrange interpolation described in Section 4.2.2.

4.2.4.3. Collective effects

For the collective contribution, the Hamiltonian H_c defined in Eq. (2.134) is treated using the first-order Hamiltonian displacement in Eq. (4.32):

$$\delta q = \frac{\partial H_c}{\partial p} \Delta\theta, \qquad \delta p = -\frac{\partial H_c}{\partial q} \Delta\theta. \tag{4.63}$$

Since H_c depends only on the longitudinal coordinate q, the wake kick leaves q unchanged and modifies only the energy coordinate p. Using Eq. (2.134), the energy-grid displacement is

$$\Delta y_r^K(x) = \frac{e f_{\mathrm{rev}}}{2\pi f_{s,0} \sigma_{E,0} \Delta p} V_{\mathrm{c}}(q_x, \theta_n) \Delta\theta. \tag{4.64}$$

Inovesa evaluates the wake potential from the density at the beginning of the wake-kick operator. Therefore, $V_{\mathrm{c}}(q_x, \theta_n)$ is known before the inverse map is constructed. Since the collective kick leaves q fixed, the potential is evaluated at the target-column coordinate q_x, defined in Eq. (4.20). The corresponding source coordinate is

$$x'_r = x, \qquad y'_r = y - \Delta y_r^K(x). \tag{4.65}$$

4.2.5. Implementation of the Fokker–Planck operator

Beyond conservative density transport, the Lagrange representation is also used to evaluate the FP contribution to the VFPE. The damping and diffusion terms require the first and second derivatives of the phase space density with respect to the momentum coordinate. These derivatives are evaluated by differentiating the interpolation polynomial along the momentum-grid coordinate y, using the derivative expressions introduced in Section 4.2.2.

The derivatives with respect to the canonical momentum p are related to the derivatives with respect to the grid coordinate y by

$$\frac{\partial}{\partial p} = \frac{1}{\Delta p}\frac{\partial}{\partial y}, \qquad \frac{\partial^2}{\partial p^2} = \frac{1}{\Delta p^2}\frac{\partial^2}{\partial y^2}. \tag{4.66}$$

Transforming the FP operator in Eq. (4.14) to grid coordinates, and omitting the subscript Δt for simplicity, gives

$$(D \circ \psi)(x,y) = \psi(x,y) + \beta_d \Delta\theta \left[\psi(x,y) + \frac{p_y}{\Delta p}\frac{\partial \psi}{\partial y}(x,y) + \frac{1}{\Delta p^2}\frac{\partial^2 \psi}{\partial y^2}(x,y)\right]. \tag{4.67}$$

For a centered grid, $p_y/\Delta p = y - y_0$, and therefore

$$(D \circ \psi)(x,y) = \psi(x,y) + \beta_d \Delta\theta \left[\psi(x,y) + (y-y_0)\frac{\partial \psi}{\partial y}(x,y) + \frac{1}{\Delta p^2}\frac{\partial^2 \psi}{\partial y^2}(x,y)\right]. \tag{4.68}$$

The cubic derivative approximations are selected according to the sign of the momentum coordinate [72]. Inovesa splits the evaluation at the zero-momentum index y_0, because the damping term transports density toward $p = 0$. For $y < y_0$, the density is transported toward increasing y, and the derivative is evaluated using the points $y-2$, $y-1$, y, and $y+1$. For $y \geq y_0$, the density is transported toward decreasing y, and the points $y-1$, y, $y+1$, and $y+2$ are used.

For $y < y_0$, the first derivative is obtained from Eq. (4.27) as

$$\frac{\partial \psi}{\partial y}(x,y) \approx \frac{1}{6}\left[\psi(x,y-2) - 6\psi(x,y-1) + 3\psi(x,y) + 2\psi(x,y+1)\right], \tag{4.69}$$

while the second derivative is obtained from Eq. (4.28) as

$$\frac{\partial^2 \psi}{\partial y^2}(x,y) \approx \psi(x,y-1) - 2\psi(x,y) + \psi(x,y+1). \tag{4.70}$$

Substitution into Eq. (4.68) gives

$$(D \circ \psi)(x,y) \approx \xi_{-2}^{(-)}(y)\psi(x,y-2) + \xi_{-1}^{(-)}(y)\psi(x,y-1) + \xi_{0}^{(-)}(y)\psi(x,y) + \xi_{1}^{(-)}(y)\psi(x,y+1), \tag{4.71}$$

with

$$\xi_{-2}^{(-)}(y) := \beta_d \Delta\theta \frac{y-y_0}{6}, \tag{4.72}$$

$$\xi_{-1}^{(-)}(y) := \beta_d \Delta\theta \left[\frac{1}{\Delta p^2} - (y-y_0)\right], \tag{4.73}$$

$$\xi_{0}^{(-)}(y) := 1 + \beta_d \Delta\theta \left[1 + \frac{y-y_0}{2} - \frac{2}{\Delta p^2}\right], \tag{4.74}$$

$$\xi_{1}^{(-)}(y) := \beta_d \Delta\theta \left[\frac{1}{\Delta p^2} + \frac{y-y_0}{3}\right]. \tag{4.75}$$

For $y \geq y_0$, the first derivative is

$$\frac{\partial \psi}{\partial y}(x,y) \approx \frac{1}{6}\left[-2\psi(x,y-1) - 3\psi(x,y) + 6\psi(x,y+1) - \psi(x,y+2)\right], \tag{4.76}$$

with the same second derivative approximation from Eq. (4.70). The corresponding update is

$$(D \circ \psi)(x,y) \approx \xi_{-1}^{(+)}(y)\psi(x,y-1) + \xi_0^{(+)}(y)\psi(x,y) + \xi_1^{(+)}(y)\psi(x,y+1) + \xi_2^{(+)}(y)\psi(x,y+2), \tag{4.77}$$

where

$$\xi_{-1}^{(+)}(y) := \beta_d \Delta\theta \left[\frac{1}{\Delta p^2} - \frac{y-y_0}{3}\right], \tag{4.78}$$

$$\xi_0^{(+)}(y) := 1 + \beta_d \Delta\theta \left[1 - \frac{y-y_0}{2} - \frac{2}{\Delta p^2}\right], \tag{4.79}$$

$$\xi_1^{(+)}(y) := \beta_d \Delta\theta \left[\frac{1}{\Delta p^2} + (y-y_0)\right], \tag{4.80}$$

$$\xi_2^{(+)}(y) := -\beta_d \Delta\theta \frac{y-y_0}{6}. \tag{4.81}$$

At $y = y_0$, the momentum vanishes and the first-derivative contribution to the damping term is zero. Both point selections therefore reduce to the same centered diffusion contribution at the zero-momentum grid point.

4.3. Sparse matrix representation of the split operators

Under the adopted operator-splitting scheme, the evolution of the phase space density is decomposed into four sub-operators. Three describe the deterministic Hamiltonian transport, while the fourth represents the damping and diffusion generated by the FP contribution. In Inovesa, these operators are applied sequentially as numerical functions acting on the discretized phase space density. Each sub-operator can, however, also be represented by a sparse matrix.

For the Hamiltonian operators, the density at each grid point is reconstructed from the preceding state by evaluating the cubic Lagrange interpolation at the corresponding inverse-mapped position. The FP operator does not involve an inverse mapping, but its discrete formulation likewise expresses each updated density value as a weighted combination of neighboring momentum-grid values. Consequently, all four operators admit the same general sparse-matrix representation [1], although their coefficients are obtained differently.

To construct this representation, the two-dimensional phase space indices are mapped to a one-dimensional index according to

$$n := Nx + y, \tag{4.82}$$

where N is the number of grid points along each side of the square phase space grid and

$$x, y \in \{0, \ldots, N-1\}, \qquad n \in \{0, \ldots, N^2 - 1\}. \tag{4.83}$$

Using this flattened representation, the updated phase space density at the grid point n can be written as

$$\hat{\psi}(n) = \sum_{i=0}^{3} \psi(n_i) L_i(x'_r), \tag{4.84}$$

where n_i are the flattened indices of the contributing grid points and $L_i(\hat{n})$ are their corresponding coefficients. For the Hamiltonian operators, $\hat{n}$ denotes the one-dimensional position in the preceding state[2] determined by the inverse dynamics of the operator. The coefficients $L_i(\hat{n})$ are then the values of the Lagrange polynomials at this position. For the FP operator, the coefficients are those obtained from the differentiated Lagrange polynomials derived in the preceding section.

The discretized phase space density is arranged into the vector

$$\boldsymbol{\psi} := \begin{bmatrix} \psi(1) \\ \psi(2) \\ \vdots \\ \psi(N^2) \end{bmatrix}, \tag{4.85}$$

and the updated density is defined analogously by

$$\hat{\boldsymbol{\psi}} := \begin{bmatrix} \hat{\psi}(1) \\ \hat{\psi}(2) \\ \vdots \\ \hat{\psi}(N^2) \end{bmatrix}. \tag{4.86}$$

Equation (4.84) can then be written simultaneously for all grid points in the matrix form

$$\begin{bmatrix} \hat{\psi}(1) \\ \hat{\psi}(2) \\ \vdots \\ \hat{\psi}(N^2) \end{bmatrix} = \overbrace{\begin{bmatrix} L_1(\hat{n}_1) & L_2(\hat{n}_1) & L_3(\hat{n}_1) & L_4(\hat{n}_1) & 0 & \cdots & 0 \\ 0 & L_1(\hat{n}_2) & L_2(\hat{n}_2) & L_3(\hat{n}_2) & L_4(\hat{n}_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & L_4(\hat{n}_{N^2}) \end{bmatrix}}^{L(\hat{\boldsymbol{n}})} \begin{bmatrix} \psi(1) \\ \psi(2) \\ \vdots \\ \psi(N^2) \end{bmatrix}, \tag{4.87}$$

Here, $\hat{n}_n$ denotes the preceding-state position associated with the target grid point n. Each row contains only the coefficients associated with the grid points n_i contributing to the corresponding updated density value. All remaining matrix entries are zero.

The locations of the nonzero coefficients depend on the direction in which the operator acts. For an operator acting along the momentum-grid coordinate, neighboring points are adjacent in the flattened representation. For an operator acting along the position coordinate, neighboring points are separated by N entries. Equation (4.87) is therefore schematic: the four nonzero coefficients in each row are not necessarily located in consecutive columns.

The action of a generic split operator can consequently be written compactly as

$$\hat{\boldsymbol{\psi}} = L(\hat{\boldsymbol{n}}) \cdot \boldsymbol{\psi}. \tag{4.88}$$

Representing the four split operators by the matrices K, R_ϵ, R_φ, and D, the evolution of the discretized phase space density over one normalized time interval $\Delta\theta$ is written as

$$\boldsymbol{\psi}(n, \theta + \Delta\theta) = \mathrm{D} \cdot \mathrm{R}_\varphi \cdot \mathrm{R}_\epsilon \cdot \mathrm{K}(\boldsymbol{\psi}(\theta)) \cdot \boldsymbol{\psi}(n, \theta). \tag{4.89}$$

[2] For this derivation the time of the phase space density is implicit in the notation.

The coefficients of $\mathrm{K}(\boldsymbol{\psi}(\theta)))$ are recalculated from the current phase space density through the collective wake potential. The complete transformation from θ to $\theta+\Delta\theta$ is therefore represented by the composite operator

$$\mathrm{M}_\theta = \mathrm{D} \cdot \mathrm{R}_\varphi \cdot \mathrm{R}_\epsilon \cdot \mathrm{K}(\boldsymbol{\psi}(\theta)), \tag{4.90}$$

where M_θ denotes the matrix operator evaluated at θ for a time step $\Delta\theta$. The phase space density is then advanced according to

$$\boldsymbol{\psi}(n, \theta + \Delta\theta) = \mathrm{M}_\theta \cdot \boldsymbol{\psi}(n, \theta). \tag{4.91}$$

Equation (4.87) makes the sparsity of the individual operator matrices explicit. For cubic Lagrange interpolation, each target grid point depends on only four grid points from the preceding phase space state. Each operator matrix is therefore sparse, with only a few nonzero entries per row. However, successive matrix multiplications increase the coupling between grid points, causing the resulting matrix to become progressively less sparse.

For fixed grid and machine parameters, the matrices D, R_φ, and R_ϵ remain constant. The wake-kick matrix K varies with time because the wake potential is recalculated from the phase space density at each time step. Consequently, although each matrix operation is linear once its coefficients are fixed, the complete evolution remains nonlinear.

5. Longitudinal phase space density reconstruction

This chapter addresses the reconstruction of the longitudinal phase space density of an electron bunch from a sequence of EOSD measurements. The unknown is a non-negative density

$$\psi = \psi(q, p, \theta) \geq 0, \tag{5.1}$$

where q denotes the normalized longitudinal position and p the normalized energy deviation. The objective is to identify an initial density ψ° whose turn-by-turn evolution follows the longitudinal beam dynamics and from which the predicted EOSD response reproduces the measured signals [1, 2].

This differs from rigid-rotation tomography [18], where the phase space is assumed to remain approximately unchanged while it rotates, making tools such as FBP applicable. In the non-equilibrium regimes considered here, collective effects and micro-bunching can deform the phase space during the measurement sequence [76, 13]. The reconstruction must therefore account for this deformation by incorporating the beam dynamics directly into the inverse problem.

A second improvement introduced in this chapter is the inclusion of the EOSD observation model. Since the EOSD measurement is not treated as an ideal longitudinal charge profile, the relation between the bunch profile and the measured signal is modeled through an impulse response that acts as a surrogate transfer function of the EOSD setup.

5.1. Inverse problem formulation

The phase space density reconstruction problem is formulated as a PDE-constrained inverse problem,

$$\begin{aligned}
\min_{\psi^\circ \geq 0} \quad & J(\psi) = \frac{1}{2}\sum_{n=1}^{m} \|\Gamma_n - \hat{\Gamma}_n\|^2 + \gamma R(\psi^\circ) \\
\text{s.t.} \quad & \frac{\partial \psi}{\partial \theta} + \frac{\partial H}{\partial p}\frac{\partial \psi}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial \psi}{\partial p} = \beta_d \frac{\partial}{\partial p}\left(p\psi + \frac{\partial \psi}{\partial p}\right), \\
& H(q, p, \theta) = \frac{1}{2}\left(q^2 + p^2\right) + \frac{e f_{\mathrm{rev}}}{2\pi\sigma_{E,0} f_{s,0}} \int_q^\infty V_{\mathrm{c}}(q', \theta)\, dq', \\
& V_{\mathrm{c}}(q, \theta) = \frac{eN_b}{2\pi} \int_{-\infty}^{\infty} \tilde{\rho}^\theta(\omega) Z_\parallel(\omega) e^{i\omega\sigma_{z,0} q/v} \mathrm{d}\omega, \\
& \rho(q, \theta) = \int \psi(q, p, \theta) \mathrm{d}p, \\
& \tilde{\rho}^\theta(\omega) = \int_{-\infty}^{\infty} \rho(q, \theta) e^{-i\omega\sigma_{z,0} q/v} \mathrm{d}q, \\
& \Gamma_n = h * \rho(q, \theta_n), \\
& \psi^\circ = \psi(q, p, \theta_1).
\end{aligned} \tag{5.2}$$

The cost function $J(\psi)$ measures the mismatch between the predicted EOSD signals Γ_n and the measured signals $\hat{\Gamma}_n$ over the measurement sequence. The term $R(\psi^\circ)$ regularizes the initial density to stabilize the reconstruction against measurement noise, model mismatch, and the ill-posedness of the inverse problem [77, 78].

The predicted measurements Γ_n are generated by evolving the initial density ψ° according to the VFPE, which provides the dynamical constraint of the inverse problem. This evolution is driven by the Hamiltonian H, which contains the harmonic synchrotron-motion term and the collective wake contribution V_c.

At the measurement times θ_n, related to the simulation time t through Eq. (2.121), the bunch profile $\rho(q, \theta_n)$ is obtained by projecting the phase space density over p. The predicted EOSD signal is then computed by applying the impulse response h introduced in Chapter 3. The inverse problem therefore connects the initial density to the measured signals through VFPE evolution, phase space projection, and the EOSD surrogate observation model.

Finally, the problem is nonlinear because the wake potential $V_\mathrm{c}(q, \theta)$ depends on the bunch profile and therefore on the phase space density ψ.

5.2. PDE-constrained optimization solution

The reconstruction problem introduced in Eq. (5.2) can be written in terms of the control variable u and the state variable x, following the notation of control theory. Let $\Omega \subset \mathbb{R}^2$ denote the phase space domain with coordinates (q, p). The control variable is the initial non-negative phase space density:

$$u : \Omega \to \mathbb{R}_{\geq 0}, \qquad u(q, p) = \psi(q, p, \theta_1). \tag{5.3}$$

The state variable is the complete density evolution generated from this initial condition under the VFPE dynamics:

$$x : \Omega \times [\theta_1, \theta_m] \to \mathbb{R}_{\geq 0}, \qquad x(q, p, \theta) = \psi(q, p, \theta). \tag{5.4}$$

For readability, the explicit dependence on (q, p) and θ is omitted in the following.

The constraints are written compactly as

$$F(x, u) = 0, \tag{5.5}$$

where the dependence on time is implicit in the state trajectory x. Using this notation, the optimization problem reads [1]

$$\begin{aligned} \min_{x,\, u} \quad & J(x, u) \\ \text{s.t.} \quad & F(x, u) = 0. \end{aligned} \tag{5.6}$$

The cost function J is formulated in full-space form, treating the state trajectory x and the control u as mathematically independent optimization variables. The coupling between state and control is enforced entirely through the constraint $F(x, u) = 0$, which includes the forward and observation models.

[1] The non-negativity constraint is not included explicitly in this formulation; it is enforced directly by the optimization algorithm used to solve the unconstrained problem.

Following the classification in Gunzburger [79, p. 22], the constrained problem can be solved using three generic approaches: one-shot methods, optimization methods based on sensitivity equations, and optimization methods based on adjoint equations. These three approaches can be understood as not strictly independent alternatives, but rather successive reorganizations of the same optimality structure.

In the following subsections, the logic flow of these three methods is explicitly derived to demonstrate how the problem naturally converges to the adjoint-based approach, which stands as the only computationally tractable strategy for high-dimensional phase space density reconstruction. The derivation is carried out directly in the discrete formulation, consistent with the numerical reconstruction problem considered in this thesis. Although the three approaches can also be formulated in continuous form, the reader can find the corresponding continuous derivations in [79].

5.2.1. One-shot formulation

The first approach consists of enforcing the physical constraints by means of Lagrange multipliers, replacing the constrained minimization problem with an unconstrained stationary problem. The corresponding discrete Lagrangian is defined as

$$\mathcal{L}(x, u, \lambda) = J(x, u) - \lambda^\top F(x, u), \tag{5.7}$$

where λ is the vector of Lagrange multipliers associated with the discrete constraints and has the same dimension as $F(x, u)$. The one-shot problem requires finding x, u, and λ such that $\mathcal{L}$ is stationary, a mathematical requirement equivalent to setting $\nabla\mathcal{L} = 0$. Differentiation of $\mathcal{L}$ with respect to λ recovers the constraints:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -F(x, u) = 0. \tag{5.8}$$

Differentiation with respect to the state yields the adjoint equation:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial J}{\partial x} - \lambda^\top \frac{\partial F}{\partial x} = 0, \tag{5.9}$$

and differentiation with respect to the control gives the optimality condition:

$$\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial J}{\partial u} - \lambda^\top \frac{\partial F}{\partial u} = 0. \tag{5.10}$$

Equations (5.8)–(5.10) form the full first-order optimality system.

Although the one-shot method is the most direct formulation of the optimization problem, it is not feasible for phase space density reconstruction. It requires solving for the full state trajectory and the adjoint variables simultaneously, rather than advancing them sequentially. For a high-resolution phase space grid over many time steps, this leads to a very large coupled system. Since the dynamics are nonlinear, the system would also require repeated assembly and solution of a global Jacobian, making the memory and computational cost prohibitive.

Full-space one-shot approaches are nevertheless viable in other optimization contexts [79]. In this thesis, the one-shot formulation is therefore used only to make the constrained structure of the problem explicit and to motivate the adjoint equations. It serves as an analytical reference point, not as the computational strategy used for the reconstruction.

5.2.2. Optimization based on sensitivity equations

The second approach begins by assuming that, for a given control u, the constraints can be solved by forward calculation. In this way, the state is regarded as a function of the control, $x = f(u)$, and the cost function may be viewed in a reduced form depending on u alone:

$$\begin{aligned} \min_{u} \quad & J(x,u) \\ \text{s.t.} \quad & F(x,u) = 0. \end{aligned} \tag{5.11}$$

Strictly speaking, the reduced formulation replaces the constraint $F(x,u) = 0$ by the control-to-state relation $x = f(u)$. In practice, this mapping is rarely known explicitly in closed-form and is instead defined only through the forward solver. The constraint notation is therefore retained to make clear that the gradient is still obtained from the underlying dynamical model, as it cannot be computed directly from the implicit mapping $f(u)$.

The total derivative of J with respect to the control is obtained via the chain rule:

$$\frac{dJ}{du} = \frac{\partial J}{\partial u} + \frac{\partial J}{\partial x}\frac{dx}{du}. \tag{5.12}$$

To determine the state sensitivity $\frac{\mathrm{d}x}{\mathrm{d}u}$, the constraint $F(x,u) = 0$, is differentiated with respect to the control, yielding the sensitivity system:

$$\frac{dF}{du} = \frac{\partial F}{\partial x}\frac{dx}{du} + \frac{\partial F}{\partial u} = 0. \tag{5.13}$$

The gradient $\frac{\mathrm{d}J}{\mathrm{d}u}$ can therefore be obtained, in principle, by first solving the forward problem, then solving Eq. (5.13) for the state sensitivity, and finally evaluating Eq. (5.12).

Sensitivity-based methods are often attractive when the number of control variables is small compared with the dimension of the state. A useful practical criterion can be expressed in terms of the forward map

$$F : \mathbb{R}^n \to \mathbb{R}^m, \qquad n \ll m. \tag{5.14}$$

For a low-dimensional control space, with n much smaller than the observation dimension m, sensitivities can be computed efficiently by solving one linearized problem for each control direction. The computational cost therefore scales primarily with the number of control variables and remains manageable when n is small.

In the present reconstruction problem, this condition is not satisfied. After spatial discretization, the control u contains one parameter per phase space grid cell. Computing the full sensitivity matrix $\frac{\mathrm{d}x}{\mathrm{d}u}$ would therefore require one linearized PDE solve for each control degree of freedom, at each iteration. Although this formulation makes the gradient structure explicit, its cost scales with the phase space grid size and becomes prohibitive.

5.2.3. Optimization based on adjoint equations

The third approach combines the Lagrangian formulation with the reduced viewpoint introduced in Subsection 5.2.2. Although the adjoint method is derived from the full-space constrained formulation, its purpose is precisely to eliminate the explicit dependence of Eq. (5.12) on the state sensitivity $\frac{\mathrm{d}x}{\mathrm{d}u}$, thereby avoiding the need to compute the large sensitivity matrix directly. In this way, the adjoint relation obtained from the Lagrangian provides the gradient of the reduced functional $J(u) = J(x(u), u)$ in a computationally tractable form.

From Eq. (5.9), the adjoint variable λ is constructed precisely such that:

$$\frac{\partial J}{\partial x} = \lambda^\top \frac{\partial F}{\partial x}. \tag{5.15}$$

Substituting Eq. (5.15) into Eq. (5.12) gives

$$\frac{dJ}{du} = \frac{\partial J}{\partial u} + \lambda^T \left[\frac{\partial F}{\partial x} \frac{dx}{du} \right]. \tag{5.16}$$

Using the sensitivity equation (5.13) into the square brackets of Eq. (5.16), the state sensitivity cancels out entirely:

$$\frac{dJ}{du} = \frac{\partial J}{\partial u} - \lambda^\top \frac{\partial F}{\partial u}. \tag{5.17}$$

Equation (5.17) is the key result for the present formulation, because it replaces the full sensitivity matrix $\frac{\mathrm{d}x}{\mathrm{d}u}$ by a single adjoint variable λ. As a consequence, at each optimization iteration, the gradient can be obtained by solving the forward problem once and the adjoint problem once, regardless of the number of control degrees of freedom. This is precisely the advantage of adjoint-based optimization emphasized in [79].

The adjoint equation propagates error information backward through the time sequence, starting from the measurement residuals and carrying their influence back to the initial condition. This backward propagation is analogous to error backpropagation in neural-network training, where sensitivity information is transferred through the computational graph of the forward model.

The one-shot and sensitivity-based formulations are not practical for this reconstruction problem: the former requires solving one global space-time system, while the latter requires one linearized solve per phase space degree of freedom. The adjoint formulation avoids both costs by preserving the sequential structure of the solver: the state is propagated forward, the adjoint is propagated backward, and the gradient is assembled afterward. For a control variable given by a high-resolution phase space density, this reduction to a single adjoint solve per iteration is what makes the optimization computationally tractable.

5.2.4. Discretize-then-differentiate approach

The optimization solution is derived from a discretize-then-differentiate viewpoint. This choice follows the structure of the numerical problem: the VFPE model is solved as a sequence of discrete updates on a discretized phase space, and the cost function is evaluated from discrete measurements and discrete forward predictions. The adjoint system is therefore obtained by differentiating the algebraic equations of the implemented forward solver.

The alternative is the differentiate-then-discretize approach, where the continuous adjoint equations are derived first and discretized afterward [79, p. 57]. This route preserves the analytical structure of the continuous optimization problem, clarifies boundary and terminal conditions, and is well suited for mesh-independent analysis. However, because differentiation and discretization do not generally commute [79], the resulting discrete gradient may differ from the one associated with the implemented solver. The discretize-then-differentiate approach instead gives a gradient consistent with the numerical update sequence of the implemented forward solver. Ultimately, neither approach is absolutely superior; both are well established and successful, making the distinction as much methodological as it is technical.

5.2.5. The adjoint-based optimization algorithm

The essence of the adjoint-based optimization algorithm for nonlinear optimization lies in computing the exact gradient of the cost function with respect to the optimization variables efficiently. At each iteration, the current controlled phase space is used to calculate the state trajectory by solving the forward dynamics and quantifying the measurement mismatch. The corresponding adjoint system is then solved backward to propagate this error back to the control space. The assembled gradient can subsequently be used directly by a broad class of first-order optimization methods, such as steepest descent, conjugate gradient, projected gradient, or quasi-Newton schemes.

The adjoint-based optimization algorithm can be described by the following sequence of operations:

1. **Initialization:** Choose an initial guess u for the control variable.
2. **Forward state solve:** Solve the forward dynamical model for the current control u,

$$F(x, u) = 0, \tag{5.18}$$

   to generate the complete state trajectory x over the computational domain. This step might itself be an iterative process and the entire trajectory is usually preserved in memory.
3. **Residual evaluation:** Use the simulated state trajectory to compute the predicted observables, evaluate the cost function $J(x, u)$, and determine the residuals with respect to the measured quantities.
4. **Adjoint seeding:** Construct the state derivative of the cost function,

$$\frac{\partial J}{\partial x}, \tag{5.19}$$

   which measures the sensitivity of J with respect to the state. This term contains the measurement residuals and provides the starting values for the backward adjoint solve.
5. **Backward adjoint solve:** Solve the adjoint equation, obtained by transposing Eq. (5.15),

$$\left(\frac{\partial F}{\partial x}\right)^{\top} \lambda = \frac{\partial J}{\partial x}^{\top}, \tag{5.20}$$

   to compute the adjoint variable λ. This solve is performed backward through the time sequence. The adjoint variable carries the influence of the measurement residuals back to the initial control variable.

6. **Gradient assembly:** Use the current state trajectory x and the adjoint variable λ to assemble the full gradient of the cost function. By evaluating the local partial derivatives along the current trajectory, the gradient with respect to the control is:

$$\frac{dJ}{du} = \frac{\partial J}{\partial u} - \lambda^\top \frac{\partial F}{\partial u}. \tag{5.21}$$

7. **Optimization algorithm:** Provide the computed gradient to an external optimization method, which determines a search direction, a step size, and the updated control u. Depending on the chosen optimizer, this step may itself involve additional internal iterations before the new control is accepted.

8. **Convergence check:** Update the control variable and repeat the optimization loop if the convergence criteria are not satisfied.

The adjoint method can therefore be viewed as an efficient gradient engine embedded in a broader iterative optimization framework. A similar computational workflow can also be constructed from a sensitivity-based formulation. Both approaches share the same nonlinear forward state solve (Steps 1–3) and the same external optimization update (Steps 7–8), but they differ fundamentally in the way the gradient is evaluated (Steps 4–6). Whereas the adjoint method requires only a single backward linear solve to obtain the exact gradient, the sensitivity-based approach requires a separate forward linear solve for each control degree of freedom.

A more detailed flow of the adjoint-based optimization procedure is shown in Fig. 5.1.

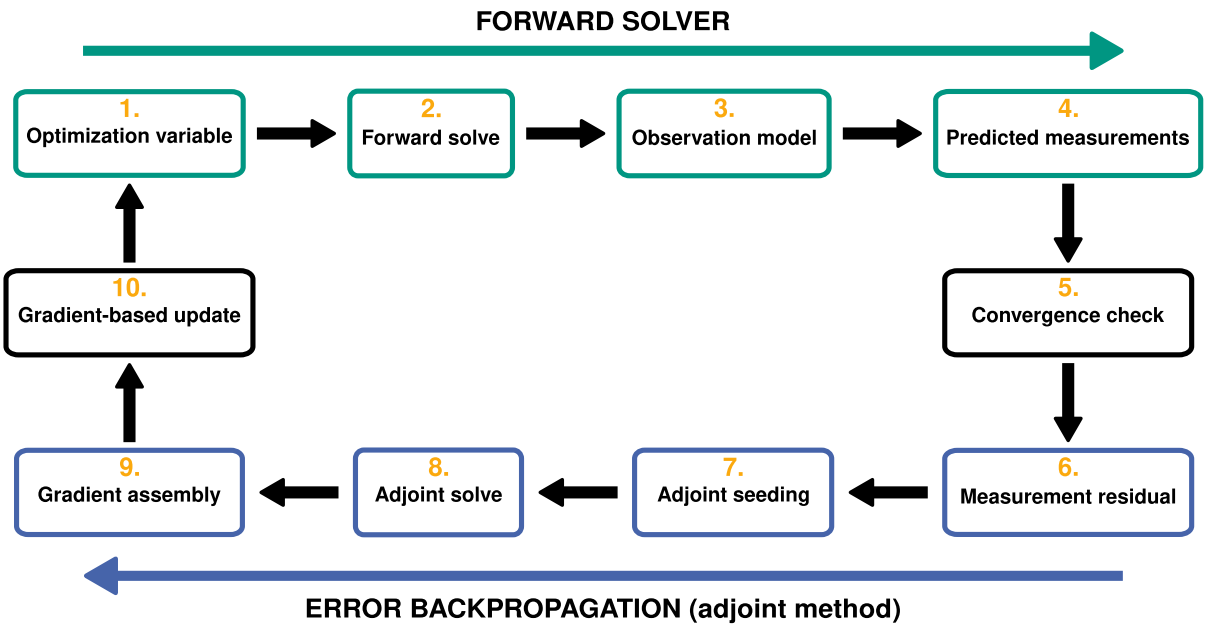


Figure 5.1.: Computational workflow of the adjoint-based optimization method. The algorithm alternates between solving the forward governing dynamics to evaluate the measurement residual, and solving the linear adjoint equation backward to propagate this error. The combined state and adjoint variables assemble the exact gradient, which dictates the control variable update for the subsequent iteration.

5.2.6. The inner–outer loop optimization algorithm

The phase space density reconstruction problem is inherently nonlinear, since the forward dynamics and the corresponding adjoint depend on the current control through the state-dependent operators. A practical strategy for handling this nonlinearity is to use two nested levels: an outer loop updates the physics model and the adjoint gradient, while an inner loop solves the resulting linearized optimization problem with a gradient-based method.

In the split-operator formulation used to solve the VFPE in Chapter 4, the outer loop updates the state-dependent operators with the current wake potential, while the inner loop keeps these operators fixed and performs the minimization. The basic steps combining the physical model update with the mathematical optimization can therefore be summarized as follows:

1. **Outer initialization:** Initialize the algorithm with an initial guess for the phase space density.
2. **Forward-operator update:** Evolve the current phase space density and use it to evaluate the state-dependent forward operators.
3. **Gradient-operator assembly:** Assemble the sparse-matrix operators required for the gradient calculation. These matrices define the fixed forward propagation, while their transposes define the corresponding adjoint propagation used inside the inner optimizer.
4. **Inner optimization loop:** Solve the corresponding gradient-based optimization problem using the assembled gradient operator. The gradient is recalculated at each inner iteration, and the loop returns an updated phase space density.
5. **Outer convergence check:** Compare the updated phase space density returned by the inner optimization loop with the density from the previous outer iteration. If the change is below a fixed tolerance, or if the maximum number of outer iterations has been reached, terminate the algorithm. Otherwise, use the updated density to recompute the wake potential, rebuild the state-dependent operators, and start the next outer iteration.

This inner–outer loop separates the roles of the two iteration levels. The outer loop updates the dynamics, while the inner loop computes the phase space update for the linearized dynamics.

5.3. Discrete operator adjoint formulation

This section derives the gradient $\frac{\mathrm{d}J}{\mathrm{d}u}$ following the three steps introduced in Section 5.2.5: the adjoint seeding, the backward solution of the adjoint equation, and the final gradient assembly.

The reconstruction problem is formulated using the discrete operator solution of the VFPE introduced in Chapter 4. In Inovesa, the phase space density is advanced through a split-operator scheme, in which each time step is represented by a sequence of discrete operators acting on the state.

This formulation provides a direct basis for both the forward model and the adjoint calculation through the transposed derivatives of the corresponding evolution operators. The reconstruction problem of Eq. (5.2) can therefore be reformulated as:

$$\begin{aligned} \min_{\{x_n\}_{n=1}^m,\, u \geq 0} \quad & J(\{x_n\}_{n=1}^m, u) := \frac{1}{2}\sum_{n=1}^{m} \|G(x_n) - \hat{\Gamma}_n\|^2 + \gamma R(u) \\ \text{s.t.} \quad & x_1 = u, \\ & x_{n+1} = M_n(x_n), \quad n = 1, \ldots, m-1. \end{aligned} \tag{5.22}$$

In this formulation, u is the initial phase space density targeted for reconstruction, and x_n is the state density at discrete time step n. The one-step state transition M_n encapsulates the four discrete operators, while the observation map G transforms each state into the corresponding estimated measurement.

For the adjoint derivation, it is convenient to collect the complete state trajectory into a single stacked vector $\mathbf{X}$. The forward dynamics can then be represented by a constraint vector $F(\mathbf{X}, u)$, while the estimated observations are collected in $\mathbf{G}(\mathbf{X})$. This matrix formulation makes the structure of the gradient calculation explicit and reveals the backward propagation of the adjoint variables. The regularization term is retained in the cost function but does not contribute to the adjoint seeding because it depends only on u; its contribution is added during the final gradient assembly.

The reconstruction problem becomes

$$\begin{aligned} \min_{\mathbf{X},\, u \geq 0} \quad & J(\mathbf{X}, u) := \frac{1}{2}\left\|\mathbf{G}(\mathbf{X}) - \hat{\mathbf{\Gamma}}\right\|^2 + \gamma R(u) \\ \text{s.t.} \quad & \mathbf{F}(\mathbf{X}, u) = \mathbf{0}. \end{aligned} \tag{5.23}$$

The stacked state trajectory, observation vector, and measurement vector are defined, respectively, as

$$\mathbf{X} := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad \mathbf{G}(\mathbf{X}) := \begin{bmatrix} G(x_1) \\ G(x_2) \\ \vdots \\ G(x_m) \end{bmatrix}, \quad \hat{\mathbf{\Gamma}} := \begin{bmatrix} \hat{\Gamma}_1 \\ \hat{\Gamma}_2 \\ \vdots \\ \hat{\Gamma}_m \end{bmatrix}. \tag{5.24}$$

The complete forward dynamics are represented by

$$\mathbf{F}(\mathbf{X}, u) = \begin{bmatrix} x_1 - u \\ x_2 - M_1(x_1) \\ \vdots \\ x_m - M_{m-1}(x_{m-1}) \end{bmatrix} = \mathbf{0}, \tag{5.25}$$

The adjoint seeding is obtained from the derivative of the cost function with respect to the stacked state vector:

$$\frac{\partial J}{\partial \mathbf{X}} = \begin{bmatrix} \dfrac{\partial J}{\partial x_1} & \dfrac{\partial J}{\partial x_2} & \cdots & \dfrac{\partial J}{\partial x_m} \end{bmatrix}. \tag{5.26}$$

Evaluating Eq. (5.26) and taking its transpose gives

$$\frac{\partial J}{\partial \mathbf{X}}^{\top} = \begin{bmatrix} \frac{\partial G(x_1)}{\partial x_1}^{\top} r_1 \\ \frac{\partial G(x_2)}{\partial x_2}^{\top} r_2 \\ \vdots \\ \frac{\partial G(x_m)}{\partial x_m}^{\top} r_m \end{bmatrix}. \tag{5.27}$$

The measurement residual at step n is defined as

$$r_n = G(x_n) - \hat{\Gamma}_n. \tag{5.28}$$

The stacked adjoint vector is defined as

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix}. \tag{5.29}$$

The Jacobian of the constraint vector with respect to the stacked state vector has the lower block-triangular form

$$\frac{\partial \mathbf{F}}{\partial \mathbf{X}} = \begin{bmatrix} I & 0 & \cdots & 0 \\ -\frac{\partial M_1(x_1)}{\partial x_1} & I & \cdots & 0 \\ 0 & -\frac{\partial M_2(x_2)}{\partial x_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\frac{\partial M_{m-1}(x_{m-1})}{\partial x_{m-1}} & I \end{bmatrix}. \tag{5.30}$$

The discrete adjoint equation, Eq. (5.20), is now expressed in terms of the stacked vectors

$$\frac{\partial \mathbf{F}}{\partial \mathbf{X}}^{\top} \boldsymbol{\lambda} = \frac{\partial J}{\partial \mathbf{X}}^{\top}. \tag{5.31}$$

Substituting Eqs 5.30, 5.29 and 5.27 in Eq. (5.31)

$$\begin{bmatrix} I & -\frac{\partial M_1(x_1)}{\partial x_1}^{\top} & \cdots & 0 \\ 0 & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\frac{\partial M_{m-1}(x_{m-1})}{\partial x_{m-1}}^{\top} \\ 0 & \cdots & 0 & I \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix} = \begin{bmatrix} \frac{\partial G(x_1)}{\partial x_1}^{\top} r_1 \\ \frac{\partial G(x_2)}{\partial x_2}^{\top} r_2 \\ \vdots \\ \frac{\partial G(x_m)}{\partial x_m}^{\top} r_m \end{bmatrix}. \tag{5.32}$$

Because this system is upper triangular, it is solved by backward substitution. This provides the algebraic origin of the backward adjoint propagation. The terminal adjoint variable is seeded by

$$\lambda_m = \frac{\partial G(x_m)}{\partial x_m}^\top r_m, \tag{5.33}$$

and the remaining adjoint variables are obtained recursively as

$$\lambda_n = \frac{\partial G(x_n)}{\partial x_n}^\top r_n + \frac{\partial M_n(x_n)}{\partial x_n}^\top \lambda_{n+1}, \qquad n = m-1, \ldots, 1. \tag{5.34}$$

Once the backward substitution reaches the first state, λ_1 contains the accumulated sensitivity of all m measurement residuals propagated back to the initial condition. The gradient of the reduced objective function is then obtained from

$$\frac{dJ}{du} = \frac{\partial J}{\partial u} - \frac{\partial \mathbf{F}}{\partial u}^\top \boldsymbol{\lambda}. \tag{5.35}$$

Since the explicit dependence of J on u arises only from the regularization term:

$$\frac{\partial J}{\partial u} = \gamma \frac{\partial R(u)}{\partial u}, \tag{5.36}$$

while the control enters the constraint vector only through the first block, $x_1 - u = 0$. Therefore,

$$\frac{\partial \mathbf{F}}{\partial u} = \begin{bmatrix} -I \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \frac{\partial \mathbf{F}}{\partial u}^\top = \begin{bmatrix} -I & 0 & \cdots & 0 \end{bmatrix}. \tag{5.37}$$

Substitution into Eq. (5.35) gives the final gradient

$$\frac{dJ}{du} = \gamma \frac{\partial R(u)}{\partial u} + \lambda_1, \tag{5.38}$$

where λ_1 is explicitly:

$$\lambda_1 = \frac{\partial G(x_1)}{\partial x_1}^\top r_1 + \frac{\partial M_1(x_1)}{\partial x_1}^\top \frac{\partial G(x_2)}{\partial x_2}^\top r_2 + \cdots + \left(\prod_{j=1}^{m-1} \left(\frac{\partial M_j(x_j)}{\partial x_j} \right)^\top \right) \frac{\partial G(x_m)}{\partial x_m}^\top r_m. \tag{5.39}$$

The resulting vector λ_1 represents the sensitivity of the complete measurement sequence to the reconstructed initial condition, and without the regularization term, corresponds directly to the gradient supplied to the external optimization algorithm.

This derivation establishes the adjoint formulation used in the reconstruction algorithm directly from the discrete split-operator representation of the Inovesa VFPE solver. The resulting backward recursion propagates the measurement sensitivities through the dynamical model and provides the gradient of the cost function with respect to the initial phase space density.

5.4. Optimization algorithms

The central role of the adjoint method is to provide the gradient information required by an external optimizer, whereas the optimization algorithm determines the search direction, the step acceptance rule, and the globalization strategy that govern the actual convergence of the reduced problem [79, 80, 81].

Let the cost function be denoted by

$$J(u) := J\big(x(u), u\big), \tag{5.40}$$

where $u \in \mathbb{R}^{N_q \times N_p}$ is the initial phase space density and $x(u)$ is the corresponding state trajectory produced by the forward solver. At the current iterate u_k, the adjoint derivation developed above yields the gradient g_k

$$g_k := \nabla J(u_k). \tag{5.41}$$

This section introduces the basic concepts of gradient-based optimization using the notation of *Numerical Optimization* [81], with emphasis on the algorithmic families compatible with the adjoint formulation and the requirements of the present phase space density reconstruction problem.

5.4.1. Line-search and trust-region strategies

Before discussing individual optimization methods, it is useful to distinguish the two standard strategies used to determine the update at each iteration.

Consider a generic optimization step in which the control variable is updated as

$$u_{k+1} = u_k + \alpha_k p_k, \tag{5.42}$$

where p_k is the search direction and α_k determines the step length along that direction.

In line-search methods, the search direction p_k is chosen first, and the step length α_k is then determined along that direction. In trust-region methods, by contrast, the search direction, which need not coincide with the gradient, is determined within a prescribed neighborhood around the current iterate.

In this sense, line-search methods are direction first and step size second, whereas trust-region methods are radius first and direction second. In both cases, a subproblem is solved to determine either an appropriate step length or search direction [81, p. 19].

5.4.2. First-order methods

The methods presented in this section are limited to first-order approaches, for which the optimization update is constructed directly from gradient information. This makes them particularly well suited to the present formulation, since the adjoint method provides the required gradient with respect to the reconstructed phase space density.

Approximate second-order and second-order methods instead seek to incorporate Hessian information, either indirectly through approximations built from gradient evaluations or directly through second-order derivatives. These approaches are discussed in detail in Appendix A.4.

For the present reconstruction problem, three representative first-order methods are considered. Although based on relatively simple update rules, they provide sufficiently powerful optimization strategies for the high-dimensional phase space density reconstruction and differ mainly in how the available gradient information is transformed into a descent direction.

5.4.3. Gradient descent

The simplest representative is gradient descent, for which the search direction is chosen as the negative gradient:

$$p_k = -g_k, \qquad u_{k+1} = u_k - \alpha_k g_k. \tag{5.43}$$

Thus, each iteration moves directly opposite to the current gradient of the reduced cost function. Its main advantage is conceptual and computational simplicity: once the adjoint gradient has been assembled, no additional operations are required apart from the selection of the step length.

Its main drawback is its sensitivity to poor scaling in the gradient components. When the partial derivatives of the reduced cost function with respect to different control variables differ strongly in magnitude, the negative gradient may point toward a direction that is locally steep but poorly aligned with the route to the minimizer. In such cases, successive iterations tend to over-correct in one set of components and under-correct in others, producing the characteristic zig-zag behavior and consequently slow convergence.

5.4.4. Steepest descent

A natural extension of gradient descent is obtained by allowing the descent direction to be measured with respect to a weighted inner product in the control space. In that case, the direction of steepest decrease is no longer the negative gradient itself, but a scaled version of it. Introducing a symmetric positive definite operator M_k, the steepest-descent direction is defined by

$$p_k = -M_k^{-1} g_k, \qquad u_{k+1} = u_k - \alpha_k M_k^{-1} g_k. \tag{5.44}$$

Thus, gradient descent appears as the special case $M_k = I$. The role of M_k is to rescale the components of the gradient so that variables with very different sensitivities do not dominate the update. In practice, M_k may be chosen in several ways: as a fixed diagonal scaling based on the expected magnitude or units of the control variables, as a preconditioner derived from the dominant diagonal or block-diagonal structure of the linearized problem, or as approximation of local curvature. The simplest choice is often a diagonal matrix whose entries are estimated from the relative scale of the gradient components or from prior physical normalization of the control.

5.4.5. Conjugate-gradient methods

A more effective first-order alternative is obtained by retaining information from previous search directions. Conjugate-gradient methods define the new direction by the recurrence

$$p_k = -g_k + \beta_k p_{k-1}, \qquad u_{k+1} = u_k + \alpha_k p_k, \tag{5.45}$$

with initialization $p_0 = -g_0$. In this way, the current search direction combines the new gradient with a controlled amount of memory from the previous step, rather than discarding all information from earlier iterations as in plain gradient descent.

The motivation for this construction comes from the linear least-squares case, where conjugate-gradient directions can exploit the structure of the problem much more efficiently than plain gradient descent. In the present reconstruction problem, this exact linear structure is lost because the reduced map from the control to the state is nonlinear. Nevertheless, the same memory effect often remains beneficial in practice, so that conjugate-gradient methods typically require fewer gradient evaluations than plain gradient descent.

Several classical choices of the parameter β_k are commonly used:

$$\beta_k^{\mathrm{FR}} = \frac{g_k^\top g_k}{g_{k-1}^\top g_{k-1}}, \qquad \text{Fletcher–Reeves,} \tag{5.46}$$

$$\beta_k^{\mathrm{PRP}} = \frac{g_k^\top (g_k - g_{k-1})}{g_{k-1}^\top g_{k-1}}, \qquad \text{Polak–Ribière–Polyak,} \tag{5.47}$$

$$\beta_k^{\mathrm{HS}} = \frac{g_k^\top (g_k - g_{k-1})}{p_{k-1}^\top (g_k - g_{k-1})}, \qquad \text{Hestenes–Stiefel,} \tag{5.48}$$

$$\beta_k^{\mathrm{DY}} = \frac{g_k^\top g_k}{p_{k-1}^\top (g_k - g_{k-1})}, \qquad \text{Dai–Yuan.} \tag{5.49}$$

Modern variants such as Hager–Zhang introduce additional safeguards to improve descent properties under inexact line searches [82].

5.4.6. Constrained methods

The physical nature of phase space density introduces a strict feasibility requirement: the density cannot become negative. If the optimizer is allowed to update u_{k+1} freely into negative regions, the resulting intermediate states may become physically meaningless and can destabilize the forward solver. Constrained methods therefore incorporate projection operators, active-set identification, or specialized bound-handling algorithms into the update process in order to enforce admissibility at every step [83, 84]. In addition, the constraint mechanism may also include restart sequences, which can themselves play a regularizing role by preventing the accumulation of unstable search directions and restoring the iteration to a more robust admissible regime.

The most direct mechanism to enforce boundary conditions is to truncate the tentative first-order update by mapping it back onto the feasible set $\mathcal{C}$:

$$\widetilde{u}_{k+1} = u_k - \alpha_k g_k, \qquad u_{k+1} = \Pi_{\mathcal{C}}(\widetilde{u}_{k+1}), \tag{5.50}$$

where $\Pi_{\mathcal{C}}$ denotes the projection operator. For the strict non-negativity constraint of the physical density array, this projection reduces to a trivial component-wise truncation:

$$\Pi_{\mathcal{C}}(u_{k+1}) = \max(u_{k+1}, 0). \tag{5.51}$$

Although the projection itself is straightforward, abrupt truncation can disrupt the conjugate descent direction. A restart is therefore required to recompute the search direction from the projected iterate.

5.5. Linear least-squares formulation

The discrete-operator formulation introduced in Section 5.3 describes the original nonlinear constrained optimization problem. As discussed in Section 5.2.6, this problem is addressed using an inner–outer loop strategy where the nonlinear dynamics are linearized within each inner iteration. The inner loop utilizes this linearized evolution to evaluate the gradient. By applying the matrix operators derived in Section 4.3, this step is formulated as a linear least-squares problem. The state evolution is then given by

$$\begin{aligned} x_1 &= u \\ x_2 &= \mathrm{M}_1 x_1 \\ &\vdots \\ x_{n+1} &= \mathrm{M}_\mathrm{n} x_n \quad n = 1, \dots, m-1, \end{aligned} \tag{5.52}$$

where M_n is the matrix operator that advances the phase space state from step n to $n+1$. This evolution is linear, and any state can be related directly to the control variable as

$$x_n = \prod_{i=1}^{n-1} \mathrm{M}_i u. \tag{5.53}$$

The observation model can likewise be represented by a matrix G. Combining it with Eq. (5.53) gives the estimated measurements at each step:

$$\begin{aligned} \mathrm{G} \cdot u &= \Gamma_1 \\ \mathrm{GM}_1 \cdot u &= \Gamma_2 \\ \mathrm{GM}_2\mathrm{M}_1 \cdot u &= \Gamma_3 \\ &\vdots \\ \mathrm{GM}_{m-1} \cdots \mathrm{M}_1 \cdot u &= \Gamma_m \end{aligned} \tag{5.54}$$

The relation between the control variable and the complete sequence of observations can therefore be expressed through the system matrix $\mathbf{A}$

$$\mathbf{A} = \begin{bmatrix} \mathrm{G} \\ \mathrm{GM}_1 \\ \mathrm{GM}_2\mathrm{M}_1 \\ \vdots \\ \mathrm{GM}_{m-1} \cdots \mathrm{M}_1 \end{bmatrix} \in \mathbb{R}^{mN \times N^2}, \tag{5.55}$$

and the stacked measurement vector $\mathbf{b}$

$$\mathbf{b} = \begin{bmatrix} \hat{\Gamma}_1 \\ \hat{\Gamma}_2 \\ \vdots \\ \hat{\Gamma}_m \end{bmatrix} \in \mathbb{R}^{mN}. \tag{5.56}$$

The reconstruction problem of Eq. (5.22) can then be condensed into the regularized linear least-squares formulation

$$\min_{u \geq 0} \ \widehat{J}(u) = \frac{1}{2}\|\mathbf{A}u - \mathbf{b}\|^2 + \gamma R(u). \tag{5.57}$$

This representation is particularly useful because it allows the inner reconstruction problem to be treated with least-squares methods from linear algebra.

5.5.1. Equivalence of least-squares and adjoint gradients

Linear least-squares problems such as Eq. (5.57) have well-established numerical solution methods. Many of these methods can be interpreted in terms of gradient-based optimization, although this terminology is less common in numerical linear algebra. For the present reconstruction problem, the least-squares gradient has the same structure as the discrete adjoint method derived for the nonlinear formulation in the previous sections. This equivalence can be seen by considering the exact gradient of the least-squares objective:

$$\nabla \widehat{J}(u) = \mathbf{A}^\top(\mathbf{A}u - \mathbf{b}) + \gamma \nabla R(u). \tag{5.58}$$

Defining the residual vector, r, as:

$$\mathbf{A}u - \mathbf{b} = \mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}. \tag{5.59}$$

Applying the block-transpose $\mathbf{A}^\top$:

$$\mathbf{A}^\top = \begin{bmatrix} \mathrm{G}^\top & \mathrm{M}_1^\top \mathrm{G}^\top & \mathrm{M}_1^\top \mathrm{M}_2^\top \mathrm{G}^\top & \cdots & \mathrm{M}_1^\top \cdots \mathrm{M}_{m-1}^\top \mathrm{G}^\top \end{bmatrix}. \tag{5.60}$$

the matrix-vector product yields the exact summation:

$$\mathbf{A}^\top(\mathbf{A}u - \mathbf{b}) = \mathrm{G}^\top r_1 + \mathrm{M}_1^\top \mathrm{G}^\top r_2 + \mathrm{M}_1^\top \mathrm{M}_2^\top \mathrm{G}^\top r_3 + \cdots + \left(\prod_{j=1}^{m-1} \mathrm{M}_j^\top \right) \mathrm{G}^\top r_m. \tag{5.61}$$

This expression can now be directly compared to the initial adjoint variable λ_1 derived from Section 5.3. Recalling Eq. (5.39), the exact backpropagation of the adjoint variables yields:

$$\lambda_1 = \frac{\partial G(x_1)}{\partial x_1}^\top r_1 + \frac{\partial M_1(x_1)}{\partial x_1}^\top \frac{\partial G(x_2)}{\partial x_2}^\top r_2 + \cdots + \left(\prod_{j=1}^{m-1} \left(\frac{\partial M_j(x_j)}{\partial x_j} \right)^\top \right) \frac{\partial G(x_m)}{\partial x_m}^\top r_m. \tag{5.62}$$

Under the linearized dynamics of the inner loop, the local state-transition Jacobian reduces to the fixed matrix operator:

$$\frac{\partial M_n(x_n)}{\partial x_n} = \frac{\partial (M_n x_n)}{\partial x_n} = \mathrm{M}_n. \tag{5.63}$$

Similarly, the observation operator can be expressed in matrix form:

$$\frac{\partial G(x_n)}{\partial x_n} = \frac{\partial (G x_n)}{\partial x_n} = \mathrm{G}. \tag{5.64}$$

Substituting these linear equivalences into the λ_1 expression perfectly maps every term of the adjoint backpropagation to the least-squares gradient. The initial adjoint variable λ_1 is therefore mathematically indistinguishable from $\mathbf{A}^\top r$.

5.5.2. Matrix-free implementation

In general, solving the least-squares problem requires access to the system matrix $\mathbf{A}$. Its explicit construction can be computationally expensive because each block row requires propagation of the full phase space dynamics, while successive state-transition products become increasingly dense. As a result, explicit construction introduces substantial computational and memory costs.

A matrix-free method avoids constructing $\mathbf{A}$. Instead, the solver stores only the operators M_n and G, together with the required state, measurement, and adjoint vectors. The computation is then reduced to repeated applications of the forward and transposed operators, with a lower memory requirement.

To leverage this architecture, the optimization algorithm must interface with the physics solver through two standardized functions that evaluate the matrix-vector products $\mathbf{A}v$ and $\mathbf{A}^\top w$ for arbitrary input vectors $v \in \mathbb{R}^{N^2}$ and $w \in \mathbb{R}^{mN}$. Following the algebraic structure established in the preceding sections, these mathematical products translate directly into the forward dynamic solver and the adjoint calculation.

1. **The forward action** ($\mathbf{A}v$): The input vector v is interpreted as an initial phase space density and propagated forward through the dynamics. At each step, the corresponding estimated observable is evaluated, and the resulting sequence is stacked into the output vector:

$$\mathbf{A}v = \begin{bmatrix} \mathrm{G}v \\ \mathrm{G}\mathrm{M}_1 v \\ \mathrm{G}\mathrm{M}_2\mathrm{M}_1 v \\ \vdots \\ \mathrm{G}\mathrm{M}_{m-1}\cdots\mathrm{M}_1 v \end{bmatrix}. \tag{5.65}$$

2. **The transposed action** ($\mathbf{A}^\top w$): The input vector w is interpreted as the error of the least-squares formulation. The transposed action is then given by the first adjoint vector,

$$\mathbf{A}^\top w = \lambda_1. \tag{5.66}$$

   The vector λ_1 is evaluated using the backward adjoint recursion of Eq. (5.61), based on the sequence of operators M_n and G associated with the current phase space state. These operators are evaluated in the outer loop and then used to describe the linearized dynamics within the inner-loop calculation.

This modular structure allows the physics solver to be used as a black-box linear operator within matrix-free optimization algorithms. The matrix-free approach is not limited to operators that are explicitly represented as matrices; it only requires a forward solver and the calculation of the adjoint.

Regardless of how the transposed action is implemented, the resulting operation is the same. The reconstruction problem can therefore be solved using iterative methods for large-scale linear least-squares problems without explicitly assembling the system matrix. The algorithms that exploit this structure are introduced in the following subsection.

5.5.3. Linear least-squares optimization algorithms

For the matrix-free formulation developed above, a suitable linear least-squares algorithm must be expressible entirely through vector operations and repeated applications of $\mathbf{A}v$ and $\mathbf{A}^\top w$. Any method that requires explicit access to individual entries, rows, or columns of the system matrix $\mathbf{A}$ is therefore not compatible with this setting.

Consequently, classical direct least-squares solvers, such as those based on singular value decomposition (SVD), QR factorization, or explicit Cholesky decomposition of the normal equations, cannot be implemented matrix-free. These direct methods fundamentally rely on dense matrix manipulations and orthogonal transformations that require full access to the matrix entries. Conversely, Krylov subspace iterative methods are natively designed to build approximate solutions relying solely on sequential matrix-vector products. Algorithms such as the conjugate gradient for least squares (CGLS), least-squares QR algorithm (LSQR) method, and the generalized minimal residual method (GMRES) method perfectly satisfy the matrix-free condition and are the standard choices for massive inverse problems.

Figure 5.2 illustrates the implementation of the linear least-squares optimization algorithm using matrix-free operators within the inner–outer loop strategy.

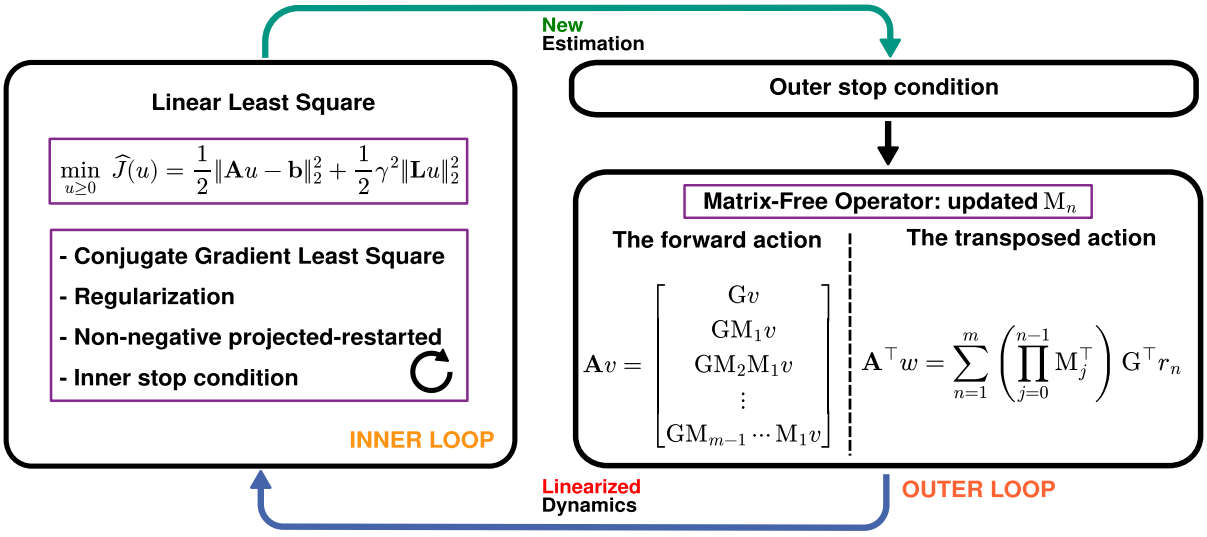


Figure 5.2.: Outer–inner loop with the matrix-free operator for linear least-squares inner loop optimization. After each estimate of the phase space density, the forward dynamics operators M_n are recomputed in the outer loop.

5.5.4. Projected-restarted non-negative conjugate gradient for least-squares

The CGLS is similar to the non-linear conjugate gradient versions discussed in Sec. 5.4.5, but it differs fundamentally in its algebraic structure.

The derivation begins by taking the least-squares formulation from Eq. (5.57) and making it fully linear by explicitly defining the regularization term as a matrix operator acting on the control variable:

$$\min_{u\geq 0} \ \widehat{J}(u) = \frac{1}{2}\|\mathbf{A}u - \mathbf{b}\|_2^2 + \frac{1}{2}\gamma^2\|\mathbf{L}u\|_2^2. \tag{5.67}$$

In this formulation, the choice of the L_2-norm is mathematically essential because the Euclidean norm yields a convex, continuously differentiable quadratic functional with a linear stationary condition. Furthermore, the regularization operator is explicitly written as a linear matrix $\mathbf{L} \in \mathbb{R}^{N_L \times N^2}$, leading to the standard Tikhonov regularization term $\|\mathbf{L}u\|_2^2$. This explicit linearity is required so that the entire penalized objective remains a perfect quadratic form. It ensures that the gradient of the regularization term,

$$\nabla(\|\mathbf{L}u\|_2^2) = 2\mathbf{L}^\top\mathbf{L}u, \tag{5.68}$$

cleanly integrates into the linear matrix-free framework without introducing nonlinear perturbations that would break the Krylov subspace generation.

Introducing the augmented operator and augmented right-hand side

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \gamma\,\mathbf{L} \end{bmatrix}, \qquad \bar{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}, \tag{5.69}$$

the problem in Eq. (5.67) can be rewritten as

$$\min_{u\geq 0} \ \widehat{J}(u) = \frac{1}{2}\|\bar{\mathbf{A}}u - \bar{\mathbf{b}}\|_2^2. \tag{5.70}$$

If the non-negativity constraint is temporarily ignored, the cost function becomes a standard quadratic least-squares functional in the augmented variables. Differentiating with respect to u gives

$$\nabla\widehat{J}(u) = \bar{\mathbf{A}}^\top\big(\bar{\mathbf{A}}u - \bar{\mathbf{b}}\big). \tag{5.71}$$

Imposing the stationary condition

$$\nabla\widehat{J}(u) = 0 \tag{5.72}$$

yields the normal equations

$$\bar{\mathbf{A}}^\top\bar{\mathbf{A}}u = \bar{\mathbf{A}}^\top\bar{\mathbf{b}}. \tag{5.73}$$

At this point the parallel with the nonlinear conjugate-gradient method introduced in Section 4 becomes direct. In both cases, the algorithm is built from the same ingredients: a current iterate $u^{(k)}$, a gradient $g^{(k)}$, a search direction $p^{(k)}$, a step length α_k, and a recurrence coefficient β_k.

The least-squares residual is defined as

$$r^{(k)} := \bar{\mathbf{b}} - \bar{\mathbf{A}}u^{(k)} \in \mathbb{R}^{mN+N_L}, \tag{5.74}$$

and the corresponding gradient is

$$g^{(k)} := \bar{\mathbf{A}}^\top r^{(k)} = \bar{\mathbf{A}}^\top \bar{\mathbf{b}} - \bar{\mathbf{A}}^\top \bar{\mathbf{A}} u^{(k)} \in \mathbb{R}^{N^2}. \tag{5.75}$$

Therefore, $g^{(k)}$ plays exactly the same role as the gradient in the nonlinear conjugate-gradient method. The difference is that here it is also the residual of the linear system in Eq. (5.73).

Starting from an initial guess $u^{(0)}$, the exact CGLS iteration is initialized by

$$r^{(0)} = \bar{\mathbf{b}} - \bar{\mathbf{A}} u^{(0)}, \tag{5.76}$$

$$g^{(0)} = \bar{\mathbf{A}}^\top r^{(0)}, \tag{5.77}$$

$$p^{(0)} = g^{(0)}. \tag{5.78}$$

This is the exact analogue of the nonlinear algorithm: the first search direction is chosen as the gradient itself.

The forward operator is applied to the current search direction:

$$q^{(k)} := \bar{\mathbf{A}} p^{(k)}. \tag{5.79}$$

The step length α_k is then

$$\alpha_k = \frac{\langle g^{(k)}, g^{(k)} \rangle}{\langle q^{(k)}, q^{(k)} \rangle} = \frac{\|g^{(k)}\|_2^2}{\|q^{(k)}\|_2^2}. \tag{5.80}$$

As in the nonlinear line-search interpretation, α_k determines how far one moves along the current search direction. Here, however, it is not chosen by an external line-search procedure: because the objective is quadratic, the optimal value can be written explicitly. In other words, α_k is the exact minimizer of the functional along the line $u^{(k)} + \alpha p^{(k)}$.

With α_k determined, the iterate and the least-squares residual are updated by

$$u^{(k+1)} = u^{(k)} + \alpha_k p^{(k)}, \tag{5.81}$$

$$r^{(k+1)} = r^{(k)} - \alpha_k q^{(k)}. \tag{5.82}$$

The new gradient is then

$$g^{(k+1)} = \bar{\mathbf{A}}^\top r^{(k+1)}, \tag{5.83}$$

and the recurrence coefficient is

$$\beta_k = \frac{\langle g^{(k+1)}, g^{(k+1)} \rangle}{\langle g^{(k)}, g^{(k)} \rangle} = \frac{\|g^{(k+1)}\|_2^2}{\|g^{(k)}\|_2^2}. \tag{5.84}$$

Thus, exactly as in the nonlinear conjugate-gradient algorithm, β_k controls how much of the previous search direction is retained. The new direction is then

$$p^{(k+1)} = g^{(k+1)} + \beta_k p^{(k)}. \tag{5.85}$$

These equations define the CGLS iteration for the regularized least-squares problem. The parallel with the nonlinear conjugate-gradient method is therefore complete: in both cases the update is built from the pair $(g^{(k)}, p^{(k)})$, a step length α_k, and a recurrence coefficient β_k. The essential simplification in the present linear least-squares setting is that the gradient is available in closed form, the step length can be computed exactly, and the entire iteration is determined directly by the operators $\bar{\mathbf{A}} v$ and $\bar{\mathbf{A}}^\top w$, without requiring a separate nonlinear line-search procedure.

The actual reconstruction algorithm is formulated in a correction, or error updating, formulation rather than as a direct solution. Instead of recomputing an entirely new phase space density from scratch, the algorithm starts from the current feasible approximation $u^{(\ell)} \in \mathbb{R}^{N^2}$ and asks only for the correction needed to fit the measurements. Therefore, the next iterate is written as

$$u = u^{(\ell)} + e, \tag{5.86}$$

where $u^{(\ell)} \in \mathbb{R}^{N^2}$ denotes the current feasible approximation and $e \in \mathbb{R}^{N^2}$ is the correction to be computed. Here the index ℓ labels the projected-restart loop associated with the linearized linear least-squares problem.

Inside each projected-restart step ℓ, the correction e is computed approximately by an inner CGLS iteration.

Substituting Eq. (5.86) into Eq. (5.67) yields the correction problem

$$\min_e \ \frac{1}{2}\|\mathbf{A}e - r^{(\ell)}\|_2^2 + \frac{1}{2}\gamma^2\|\mathbf{L}(u^{(\ell)} + e)\|_2^2, \tag{5.87}$$

where

$$r^{(\ell)} := \mathbf{b} - \mathbf{A}u^{(\ell)} \tag{5.88}$$

is the residual associated with the current feasible iterate. Defining the augmented residual

$$\bar{r}^{(\ell)} = \begin{bmatrix} r^{(\ell)} \\ -\gamma\,\mathbf{L}u^{(\ell)} \end{bmatrix}, \tag{5.89}$$

the correction problem becomes

$$\min_e \ \frac{1}{2}\|\bar{\mathbf{A}}e - \bar{r}^{(\ell)}\|_2^2. \tag{5.90}$$

Equation (5.90) is solved approximately by an inner CGLS iteration. Starting from

$$e_0 = 0, \tag{5.91}$$

the inner solver produces a correction e_k after a prescribed number k of iterations. The corresponding tentative update of the phase space density is then

$$\widetilde{u}^{(\ell+1)} = u^{(\ell)} + e_k. \tag{5.92}$$

Lastly, the non-negative condition is enforced explicitly by projection

$$u^{(\ell+1)} = \Pi_{\mathcal{C}}\left(\widetilde{u}^{(\ell+1)}\right). \tag{5.93}$$

The projected-restarted non-negative conjugate gradient method (nn-CGLS) therefore follows naturally from the structure of the reconstruction problem, combining the discrete VFPE solver and its adjoint, regularization for noise and ill-posedness, and the physical non-negativity constraint of the phase space density.

6. Inovesalib

As introduced in Section 4.2, Inovesa provides an efficient numerical framework for simulating the evolution of the longitudinal phase space density. Since its introduction, Inovesa has been applied in a wide range of scientific studies. Steinmann et al. [12] used Inovesa to compare time-resolved THz spectroscopic measurements at KARA with VFP simulations, relating measured bursting dynamics to simulated longitudinal substructures due to the micro-bunching instability instability. Schönfeldt et al. [85] used Inovesa to investigate how the CSR impedance contributes to the synchronous phase shift at KARA, connecting the solver to an experimentally accessible machine observable. In a related study, Funkner et al. [18] developed a non-destructive longitudinal phase space density tomography method, qualitatively validating their reconstructed dynamics against Inovesa simulations driven by KARA machine parameters.

Inovesa has also been used to interpret parameter dependent instability studies. Brosi et al. [51] analyzed the effect of a damping wiggler on the micro-bunching instability at KARA and used Inovesa simulations to study how changes in damping time modify the temporal burst pattern of the emitted CSR. Boltz et al. [42] employed Inovesa for passive particle tracking in order to study how CSR wake fields perturb synchrotron motion and may seed micro-structures in longitudinal phase space. In another application, Brosi et al. [86] used Inovesa as one of two VFP solvers in a cross-code study of SOLEIL and KARA. Most recently, Brosi et al. [87] used Inovesa at MAX IV to investigate microwave instabilities under varying intra-beam scattering (IBS) conditions, finding good qualitative agreement between simulation and experiment. These studies highlight the role of Inovesa as a reference tool for comparative and interpretive numerical studies, extending its use beyond individual simulation cases.

Inovesa has additionally been used for hardware design, impedance manipulation, and studies of machine learning for control. Maier et al. [11] simulated the impact of corrugated structures on the longitudinal beam dynamics and emitted CSR power at KARA, and subsequently studied how these effects depend on the bunch length under low momentum compaction operation [88]. Wang et al. [89] used Inovesa as the CSR simulation environment for accelerated deep reinforcement learning aimed at fast feedback of beam dynamics at KARA, showing that Inovesa can also serve as a simulation backend for low latency control system development.

These applications demonstrate the broad scientific use of Inovesa and motivated its transition from a dedicated solver to a more general computational framework. Inovesalib preserves the validated numerical methods of Inovesa while reorganizing them into a modular C++ library with explicit interfaces for model parameters, observers, controllers, and dynamically loaded algorithms. The same physical problem can therefore be solved through using a compact C++ application programming interface (API), interchangeable algorithm plugins, or Python bindings, facilitating integration, extension, and maintenance.

One motivation for this development was to investigate whether algorithmic differentiation (AD) could provide derivatives directly from the VFPE forward model. A partial implementation using CoDiPack was developed as a proof of concept, but it was not used to produce the reconstruction results reported in Chapter 7.

This chapter presents the design and implementation of Inovesalib, with particular emphasis on its modular architecture, its experimental support for AD, and its application to the phase space reconstruction developed in this thesis.

6.1. Architecture

Inovesalib is organized as a modular scientific software framework built around the validated numerical foundations of Inovesa. Its architecture separates solver execution, parameter models, diagnostics, runtime control, and algorithmic extension into distinct layers, ensuring that each component can evolve independently while remaining connected through stable interfaces. An overview of the architecture is illustrated in Fig. 6.1. All components shown in the figure were implemented and used in constructing and executing the reconstruction workflow presented in this thesis. The framework should nevertheless be regarded as research software that is not yet ready for production because software stabilization, API consolidation, and broader robustness testing remain ongoing.

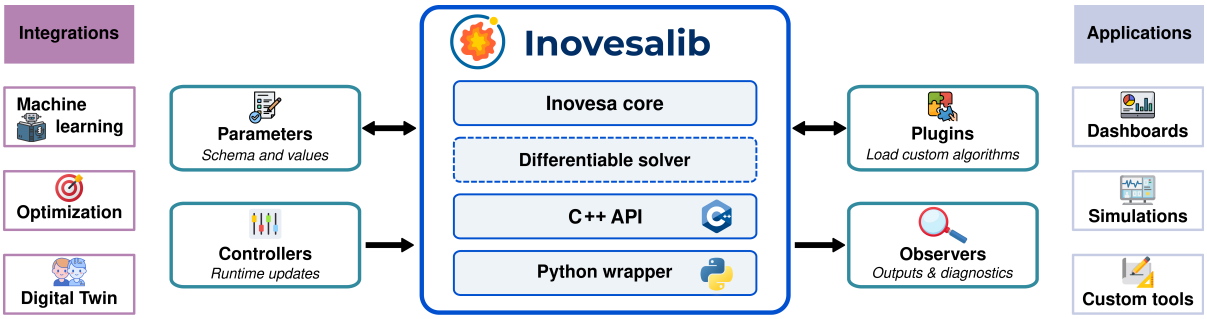


Figure 6.1.: Overview of the modular Inovesalib architecture, showing the separation between solver execution, parameter models, observers, controllers, algorithm plugins, and the available C++ and Python interfaces. All depicted components were implemented and used in the reconstruction workflow presented in this thesis. Further stabilization is required for production use.

6.1.1. Inovesa core

At the core of Inovesalib lies the existing Inovesa VFPE solver, now encapsulated in the `Inovesa class` as a reusable library component. This class follows a Facade design pattern that preserves the validated Inovesa numerical workflow while exposing it through a compact object interface for external applications, analysis pipelines, and Python-based workflows. In this way, the established solver kernel is reorganized into a form with explicit and reusable state management, runtime control, and data access.

The time evolution is carried out in the main iteration loop within the `Inovesa class`. For each step, the wake potential is updated, and the transport is performed by sequentially applying the numerical operators to the phase space state in the order of wake, RF, drift, and Fokker–Planck [43, 1]. This realizes the operator splitting scheme in a form that remains close to the validated implementation while making the solver accessible through a stable library API. In addition, Inovesalib supports selective operator refresh during runtime, so that controller-driven parameter changes can update the operators without rebuilding the full solver instance.

6.1.2. Parameters

Inovesalib introduces an explicit parameter layer that separates parameter configuration from solver execution. The parameters describe different aspects of the computation, including physical machine settings, numerical solver choices, reconstruction parameters, optimizer settings, and execution options.

All parameter models derive from the abstract `ModelParameters`, which provides a structured interface for describing, validating, and accessing parameters at runtime through an explicit schema. This schema defines the stable interface through which a concrete parameter model interacts with the framework. In this way, parameter models define a clear boundary between parameter specification and algorithmic implementation, while allowing different parameter groups to be organized according to their role in the simulation, reconstruction, or optimization workflow.

6.1.3. Observers

Inovesalib separates output generation, diagnostics, and external communication from the numerical solver through an observer pattern. Observer components can be attached to a solver run without modifying the solver logic itself. This keeps the physics implementation independent from file output, visualization, benchmarking, and communication mechanisms.

The observer layer provides extension points for different output-related tasks. For example, observers can be implemented to save simulation data, stream results through asynchronous communication interfaces, generate plots, record metadata, or monitor performance during execution. This allows the same solver core to be used in batch simulations, interactive applications, optimization workflows, and diagnostic studies while keeping the output logic outside the VFPE solver implementation.

6.1.4. Controllers

Inovesalib also implements a layer for influencing the simulation during runtime. Runtime control interfaces modify parameters, trigger updates, or steer the simulation while it is running. This capability supports the reconstruction workflow, interactive studies, feedback control studies, and optimization studies. Using it in online accelerator applications would require further development and validation.

In the current implementation, this functionality is realized through controller components. A controller can receive external commands, validate requested parameter changes against the active parameter model, and apply accepted updates during the solver iteration. Depending on the modified quantities, the affected numerical operators can then be refreshed consistently. This keeps runtime steering outside the numerical VFPE core while allowing the active simulation state to be modified in a controlled way.

6.1.5. Plugins

Inovesalib uses a plugin architecture to extend Inovesa beyond a fixed solver implementation. The original solver can therefore remain available as a reference implementation, while additional algorithms can be introduced, tested, and compared within the same computational framework. The reconstruction algorithms used in this thesis were integrated and executed through this plugin layer.

This extensibility is especially relevant for accelerator-physics applications where the numerical solver is only one component of a larger workflow. New plugins can be used to implement experimental algorithms, modified solver strategies, optimization routines, methods for parameter identification, interfaces for machine learning, or surrogate models based on neural networks. In this way, Inovesalib provides an architecture in which physics-based simulation, algorithm development, and data-driven methods can be combined without changing the core Inovesa code.

The plugin concept therefore supports a separation between the validated beam-dynamics solver and the algorithms built on top of it. This separation is important for maintaining the reliability of the established VFPE simulation while enabling rapid development of reconstruction, optimization, and control methods for accelerator studies.

6.2. Algorithmic differentiation

Derivative information is central to many optimization problems in accelerator physics, including parameter identification, model calibration, optimization, control, and phase space reconstruction. In this thesis, the reconstruction algorithms require derivatives of the simulated dynamics with respect to the optimization variables.

Finite-difference approximations are not well suited for this task, since they introduce truncation and cancellation errors and scale poorly with the number of parameters. AD avoids these limitations by applying the chain rule directly to the elementary operations of the program, producing derivatives that are accurate up to floating point precision.

AD can be applied in forward or reverse mode. Forward mode propagates parameter sensitivities together with the simulation state and is efficient when the number of differentiated inputs is small. Reverse mode, also known as adjoint mode, is more suitable for the reconstruction problem, where a scalar objective function depends on a high-dimensional initial phase space density. In this case, the forward simulation is first evaluated and the required operations are recorded on a tape. The tape is then traversed backward to propagate adjoint information from the objective function to the input variables.

For the implementation, two main AD strategies are possible. Source transformation tools generate derivative code at the compiler level and can be highly efficient, but they require deeper integration into the compilation process. Operator overloading instead replaces the scalar type of the simulation with an active type that records the required derivative information at runtime. This approach is easier to integrate into an existing C++ codebase with complex solver logic.

In Inovesalib, a partial AD implementation was developed through operator overloading with CoDiPack [90]. This implementation was developed as a proof of concept and applies active scalar arithmetic and chain rule evaluation to selected solver operations while preserving the existing Inovesa architecture. It does not yet provide AD support for the complete solver and was not used to generate the reconstruction examples presented in this thesis.

As an implementation check, the partial differentiation path implemented with CoDiPack was tested on a tomography case against the independently derived adjoint based on sparse matrix operators. The two approaches produced closely matching reconstructions, providing evidence of consistency between the implementations. Because the AD implementation was not part of the reconstruction techniques evaluated in this thesis, this comparison is not included in the reported results.

Extending the CoDiPack implementation to the complete solver and to sensitivities with respect to machine and model parameters remains future work.

6.3. C++ adaptation of IR Tools

The original MATLAB IR Tools package [91, 92] provides a comprehensive suite of iterative least-squares algorithms and regularization methods for large-scale inverse problems. Because the original package is implemented in MATLAB, it cannot be integrated directly into the Inovesalib C++ framework. Consequently, a separate C++ adaptation was developed as part of this thesis by translating and rewriting the core logic required for the reconstruction framework. This adaptation is not an official C++ version of MATLAB IR Tools. It is an independent implementation tailored to couple the Inovesalib VFPE solver to matrix-free C++ solver interfaces.

The original MATLAB package covers a broad range of iterative methods, including algorithms regularized through semi-convergence, explicitly penalized Tikhonov formulations, and constrained solvers that impose bounds on the reconstructed solution. The C++ adaptation follows the matrix-free organization of the original package but implements the components used by the reconstruction method developed in this thesis.

The C++ adaptation implements the matrix-free formulation derived in Section 5.5.2. The least-squares solver obtains the products $\mathbf{A}v$ and $\mathbf{A}^\top w$ from callable C++ objects that provide the forward and adjoint actions, respectively. This arrangement connects the VFP solver implemented in Inovesalib directly to the reconstruction algorithm without assembling the complete system matrix.

Together, these developments establish the software basis for the reconstruction method developed in this thesis, with the modular Inovesalib framework and the C++ adaptation of IR Tools linking the VFPE forward and adjoint operators to the CGLS reconstruction. The reconstruction implemented in Inovesalib is evaluated in Chapter 7 using synthetic data for verification and experimental observations for validation, while extending AD support to the complete solver and preparing the broader framework for production use remain future work.

7. Reconstruction algorithm verification and validation

The previous chapters established the theoretical framework and methodological foundations required for the reconstruction of the longitudinal phase space density from EOSD measurements. Chapter 2 introduced the VFPE used in this thesis to model the evolution of the longitudinal phase space density in beam dynamics. Chapter 3 then presented the experimental diagnostics at KARA and the EOSD forward model, including the impulse response formulation used as the observation model for the measured phase retardation signal. Chapter 5 formulated the reconstruction as a constrained inverse problem in which the unknown is the initial phase space density and the subsequent states are linked by the VFPE. Chapters 4 and 6 then presented the computational framework used to solve this problem: Inovesa provides the VFPE solver for the longitudinal phase space density evolution, while Inovesalib exposes this solver through modular interfaces that enable the development of the reconstruction algorithms.

This chapter brings these components together through the verification and validation (V&V) of the reconstruction framework. Verification examines whether the implemented reconstruction method correctly recovers known solutions under controlled simulations, whereas validation examines whether the reconstructed dynamics are consistent with experimental observations. The implementation of the reconstruction algorithm is not defined by a single algorithmic choice, but by a hierarchy of nested algorithms, heuristics, parameter selections, and implementation techniques that directly influence the performance and the reconstructed results. Moreover, the observation model itself can affect the behavior of the optimization procedure and must therefore be included in the validation of the complete reconstruction framework [93].

However, an exhaustive exploration of all possible combinations is neither practical nor necessary. Instead, a representative set of reconstruction strategies is first compared under controlled conditions in order to establish the rationale for the projected-restarted non-negative conjugate gradient least squares method used in the remainder of the chapter. These comparisons are performed using synthetic data with a known ground-truth phase space density. The comparison is organized as a progressive evaluation of the main algorithmic choices.

The first step uses projection-only reconstruction as a limiting case to show the need for dynamical constraints. It then examines the evolution of the reconstruction error, the effect of regularization in the inner–outer loop, the benefit of a warm start, and the role of non-negative projected restarts as an implicit regularization mechanism. Finally, the fully updated nonlinear reconstruction is compared with the linearized-dynamics approximation to assess the trade-off between reconstruction quality and computational cost.

The second step is numerical verification. Once the algorithm has been fixed, it is tested on synthetic EOSD data generated from Inovesa simulations. In this setting, the ground-truth phase space density is known, so the reconstruction can be evaluated directly through phase space errors, profile residuals, and trajectory consistency. These simulations are not restricted to an ideal matched case. Instead, they

also include controlled deviations in machine and beam-dynamics parameters, since the experimental reconstruction cannot rely on a perfectly known physical model of the longitudinal beam dynamics.

The third step is the experimental validation. The final section applies the selected method to synchronized KARA measurements and compares the reconstructed dynamics with the independent observables introduced in Chapter 3: the EOSD bunch-length information, the horizontal bunch-size signal from the VLM as a proxy for the energy spread, and the THz power detector signal compared to CSR power. The validation is therefore not limited to fitting the EOSD measurements used in the inverse problem. It also examines whether the reconstructed phase space produces physically consistent signatures in diagnostics that are not included in the optimization objective.

7.1. Selection of the reconstruction algorithm

The purpose of this section is to present the dynamic least-squares reconstruction method based on conjugate-gradient least squares (CGLS) as the strategy used for the simulated and experimental evaluations. The baseline implementation follows a matrix-free approach, so the full reconstruction matrix is not assembled explicitly. Instead, the required forward and adjoint operations are evaluated using the sparse-matrix representation of the VFPE evolution and the impulse response model of the EOSD diagnostic, which maps the simulated bunch profile to the measured phase retardation signal. The principal algorithmic choices of this reconstruction method are summarized in Table 7.1.

Property	Choice
Nonlinear strategy	Outer iterations with updated forward and adjoint operators
Optimization problem	Dynamic least squares
Iterative method	CGLS
Implementation	Matrix-free
Forward model	Sparse-matrix VFPE evolution with EOSD impulse response
Adjoint model	Sparse-matrix adjoint operators
Regularization	Tikhonov regularization using a 2D Laplacian matrix
Non-negativity	Projection of negative density values to zero
Restart strategy	Restart after projection
Stopping criteria	Maximum iterations; residual and update stagnation

Table 7.1.: Principal algorithmic choices of the reconstruction method.

The objective is not to establish a universal ranking of optimization algorithms. Instead, the following comparisons focus on the algorithmic decisions that are most relevant to the present inverse problem. These include the accuracy of the reconstructed initial phase space density with respect to the ground truth and the algorithm performance in terms of iterations and runtime.

7.1.1. Simulated datasets and evaluation protocol

The reconstruction algorithm is evaluated through controlled numerical experiments before it is applied to experimental data. This separation is important because real EOSD measurements do not provide direct access to the true phase space density. In simulations, by contrast, the same Inovesa run can provide the full phase space trajectory, the corresponding longitudinal bunch profiles, the

energy profiles, the bunch length, the energy spread, and the CSR-related quantities. The algorithmic comparison can therefore be based on quantities that are unavailable in the experimental case.

Inovesa provides trajectories of discretized phase space densities and bunch profiles, which are the main simulation observables used for the evaluation, since most other relevant quantities can be derived from the phase space density. Each entry in the data stream corresponds to one turn of the electron bunch in the storage ring, and the state is assumed to be observed at the location of the EOSD diagnostic.

The simulation parameters are deliberately chosen to differ from those used later for the reconstruction of the experimental measurements. The main beam and machine parameters follow the Inovesa simulation setting used in the earlier phase space tomography study by Funkner et al. [18], providing a connection to that non-equilibrium test case and facilitating comparison with the previous rigid-rotation-based reconstruction. The physics parameters used for the Inovesa simulation can be found in Table 7.2.

Parameter	Value
Beam energy	1.285 GeV
Bunch current	0.85 mA
Synchrotron frequency (f_s)	8.256 kHz
Accelerating voltage	799.257 kV
Revolution frequency (f_r)	2.715 92 MHz
Relative natural energy spread	$4.7 \cdot 10^{-4}$
Bending radius	5.559 m
Damping time	10.4 ms
Harmonic number	184
Vacuum gap	0.032 m
Linear RF	True
Phase space size (dimensionless)	12
Grid size	256 x 256 cells

Table 7.2.: Physics-related Inovesa configuration parameters.

The simulated dataset is chosen in a regime in which the bunch is in non-equilibrium and micro-bunching structures develop during the measurement window. This choice is intentional: a nearly Gaussian stable bunch would make the inverse problem artificially simple and would not test the time-dependent dynamics that motivate the VFPE-constrained reconstruction.

For each reconstruction, the measurement window is chosen as half a synchrotron period. This choice follows a legacy consideration from static tomography: under the rigid-rotation limit, this interval spans $180°$ of unique projection angles, while the second half-period repeats the same projection directions. For the dynamic reconstruction, this restriction is not strictly required because collective effects deform the phase space. Nevertheless, the half-period window is retained as a reference length to allow direct comparison with the projection-based reconstruction case. Using the parameters in Table 7.2, the number of turns per synchrotron period is

$$N_s = \frac{f_r}{f_s} = 328.96, \tag{7.1}$$

$$m = \frac{N_s}{2} \approx 164. \tag{7.2}$$

The full sinogram of the simulated EOSD data is shown in Fig. 7.1, together with the CSR intensity obtained from Inovesa. The reconstruction dataset is selected from a region where micro-bunching instabilities are present, indicated by the increased CSR intensity. In this case, m consecutive measurements are taken starting at turn 11000, marked by the green line in the figure. The initial part of the simulation is excluded because the simulation starts from an ideal two-dimensional Gaussian phase space density, which can introduce artificial transient oscillations [43, p. 75].

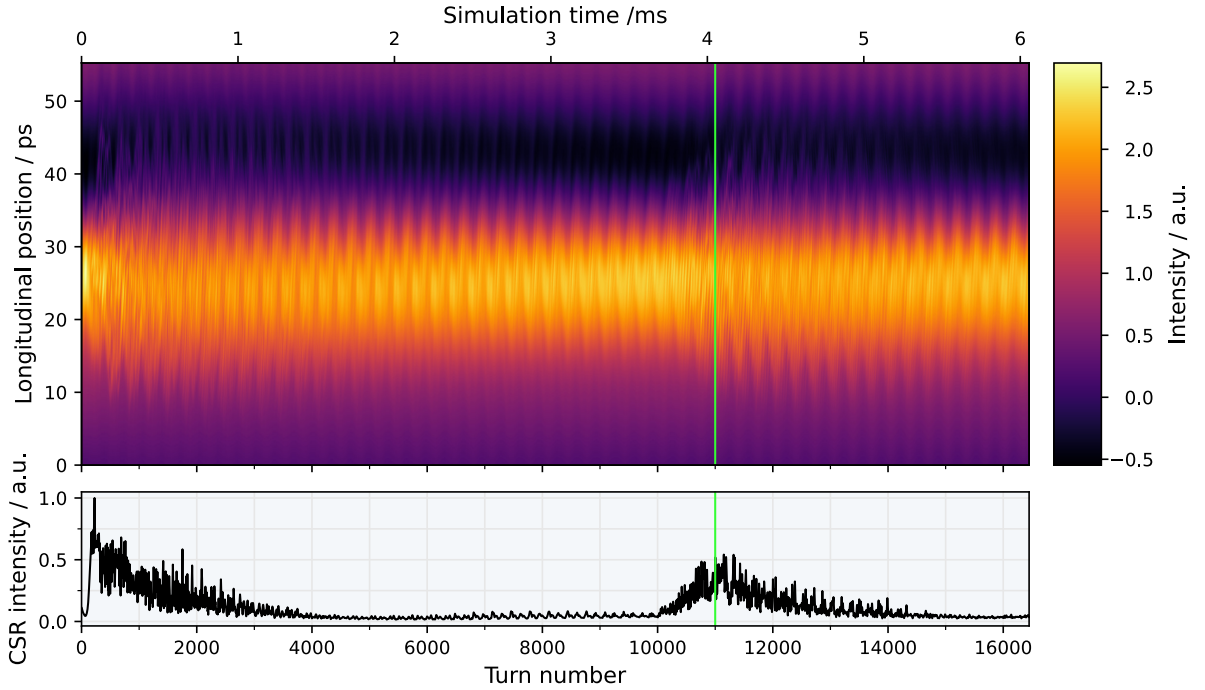


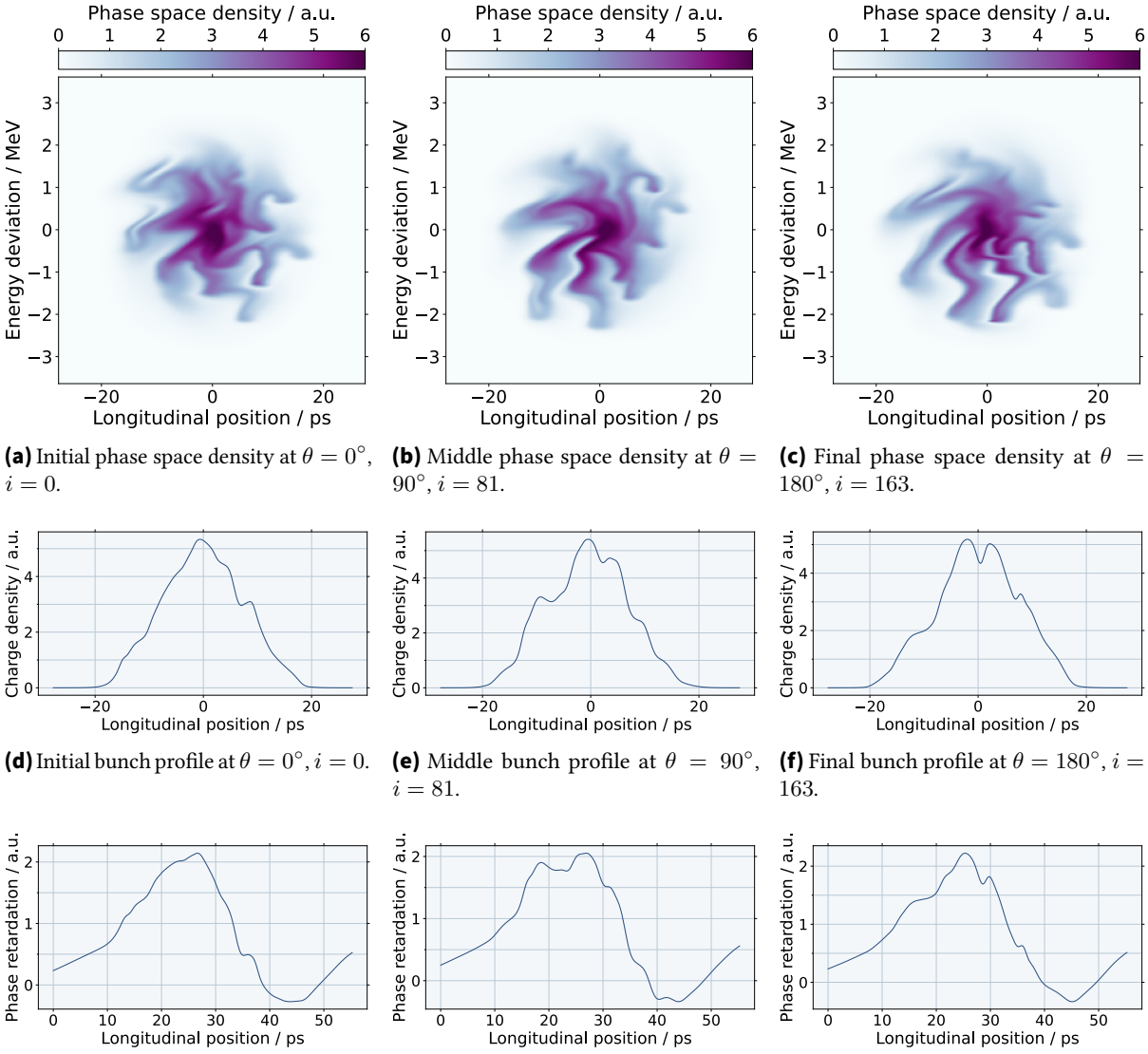
Figure 7.1.: Simulated EOSD sinogram obtained from the Inovesa trajectory, shown together with the corresponding CSR intensity. The reconstruction window is selected in the micro-bunching regime, where the CSR intensity is increased. The green line marks the starting turn of the dataset used for reconstruction, $i = 0$, corresponding to turn 11000 of the simulation. From this point, (m) consecutive measurements are used.

Figure 7.2 shows three representative points of the test trajectory used for the reconstruction studies. The upper row shows samples of the phase space density, the middle row shows the corresponding bunch profiles obtained by projection onto the longitudinal coordinate, and the lower row shows the simulated EOSD data obtained by convolving the bunch profiles with the impulse response of the EOSD setup used for the experimental measurements.

For the simulated studies, the reconstructed phase space density can be compared directly with the reference distribution. Let $\boldsymbol{\psi}_i^{\mathrm{ref}}$ and $\boldsymbol{\psi}_i$ denote the vectorized reference and reconstructed phase space densities at turn i, respectively. The relative phase space error is then evaluated using the relative Euclidean norm (L_2) as

$$\hat{e}_{\psi,i} = \frac{\left\|\boldsymbol{\psi}_i - \boldsymbol{\psi}_i^{\mathrm{ref}}\right\|_2}{\left\|\boldsymbol{\psi}_i^{\mathrm{ref}}\right\|_2}. \tag{7.3}$$

Unless stated otherwise, $i = 0$ denotes the initial state of the test trajectory and the symbol use is $\hat{e}_\psi$. For the dataset used in this section, this initial state corresponds to turn 11000 of the Inovesa simulation.



(a) Initial phase space density at $\theta = 0°$, $i = 0$.

(b) Middle phase space density at $\theta = 90°$, $i = 81$.

(c) Final phase space density at $\theta = 180°$, $i = 163$.

(d) Initial bunch profile at $\theta = 0°$, $i = 0$.

(e) Middle bunch profile at $\theta = 90°$, $i = 81$.

(f) Final bunch profile at $\theta = 180°$, $i = 163$.

(g) Initial EOSD measurement at $\theta = 0°$, $i = 0$.

(h) Middle EOSD measurement at $\theta = 90°$, $i = 81$.

(i) Final EOSD measurement at $\theta = 180°$, $i = 163$.

Figure 7.2.: Representative samples of the simulated test trajectory used for the reconstruction studies. The upper row shows the phase space density at three rotation angles, where the phase space rotates counterclockwise. The middle row shows the corresponding bunch profiles obtained by projection onto the longitudinal coordinate, and the lower row shows the simulated EOSD data obtained by applying the EOSD impulse response model to the bunch profiles.

7.1.2. Static tomography as a limiting case

In static tomography the phase space density is assumed to remain sufficiently unchanged while the synchrotron motion rotates the distribution, so that the measured bunch profiles can be interpreted as projections from different angles. This assumption is the basis of FBP and related algebraic reconstruction techniques used in longitudinal phase space tomography [24, 18].

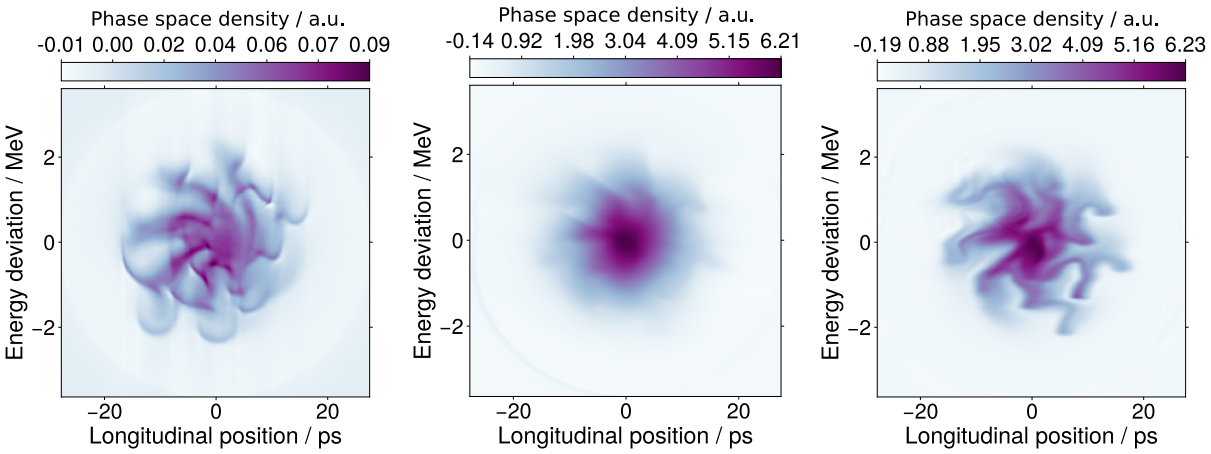
For the micro-bunching dataset used in this chapter, shown in Fig. 7.2, the static reconstruction assumption is expected to fail. The phase space does not only rotate, but also deforms under the influence of collective fields. Therefore, the measured profiles used here for the reconstruction do

not correspond to projections of a single fixed object. Applying FBP to this data mixes temporal evolution with angular sampling and can introduce artifacts that resemble physical structures without representing a feasible phase space reconstruction. For the FBP reconstruction, a Shepp–Logan filter with a cutoff frequency of 0.1 is used. Table 7.3 shows the error score obtained with FBP and CGLS when the bunch profiles, rather than the simulated EOSD data, are used as input. The CGLS reconstruction is evaluated without regularization, without non-negativity projection, and without an additional stopping criterion.

Method	Iteration 10^0	Iteration 10^1	Iteration 10^2	Iteration 10^3	Iteration 10^4	Iteration 10^5
FBP	0.988613	-	-	-	-	-
CGLS	0.239000	0.098804	0.041002	0.025675	0.016695	0.011351

Table 7.3.: Error score $\hat{e}_\psi$ for the reconstruction of the initial phase space density shown in Fig. 7.2a. FBP gives the static-tomography baseline, while CGLS is evaluated after 1, 10, 100, 1000, 10000 and 100000 iterations. Values closer to 0 indicate better agreement with the ground truth.

The phase space densities reconstructed with FBP and CGLS are shown in Fig. 7.3. Compared with the ground-truth phase space density in Fig. 7.2a, the FBP reconstruction shows a substantially incorrect density scale.



(a) FBP reconstruction.

(b) CGLS reconstruction after 1 iteration.

(c) CGLS reconstruction after 100 iterations.

Figure 7.3.: Comparison of the initial phase space density reconstructed with FBP and CGLS using the simulated bunch profiles as input data. The reconstructions are compared with the ground-truth initial phase space density shown in Fig. 7.2a.

Figure 7.3 also shows that, even under ideal simulated conditions, the reconstruction algorithm can produce negative values in the phase space density. Since such values are not physically meaningful for a density distribution, non-negativity must be enforced explicitly. This motivates the use of the projected-restarted variant of the reconstruction algorithm.

7.1.3. Evolution of the reconstruction error

Least-squares minimization methods are inherently iterative, presenting a direct trade-off between reducing the objective function and increasing computational cost. It is therefore necessary to evaluate when further iterations yield only marginal improvements. Using bunch profiles as the input for the

reconstruction algorithm, the following analysis examines this convergence behavior by comparing the iterative output against the known ground-truth phase space density of the simulated test case. This evaluation is not intended to establish a universal stopping criterion; the exact convergence profile will vary for different initial parameters or the structure of the input data. The comparison is intended to illustrate the characteristic convergence behavior rather than provide an optimized performance benchmark, since the internal parameters of the methods were not individually tuned.

Figure 7.4 shows the evolution of the error score after each reconstruction iteration for the three optimization methods. The fixed-step gradient descent (GD) method reduces the error steadily but converges more slowly toward a stable minimum because the fixed step size does not adapt to the local behavior of the objective function. In this comparison, a step size of 10^{-3} is used. Steepest descent (SD) adapts the step length at each iteration and converges more rapidly than fixed-step GD. CGLS converges fastest, requiring significantly fewer iterations to achieve a lower error score. This behavior is expected, since steepest descent minimizes the objective only along the current descent direction, whereas CGLS constructs conjugate search directions and minimizes the residual over the subspace generated by the directions explored so far.

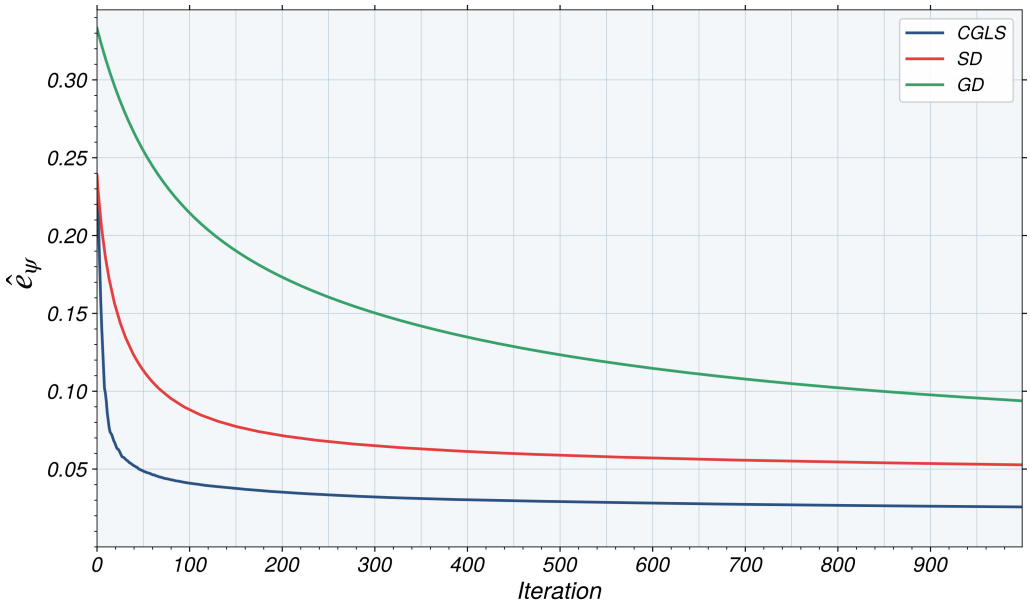


Figure 7.4.: Evolution of the reconstruction error. At turn 11000, the score is shown over 1000 reconstruction iterations for CGLS (blue), fixed-step GD (green), and SD (red). Since the error score is recorded after each iteration, the three curves do not start from the same initial value. CGLS provides the fastest convergence among the three methods, while fixed-step GD converges more gradually toward a stable minimum error.

7.1.4. Influence of regularization on the inner–outer loop

The phase space reconstruction by simulated bunch profiles provides a simplified baseline in which regularization has only a limited effect, since the dataset is free of measurement noise and systematic bias. In this subsection, the reconstruction is instead tested with noise-free simulated EOSD data. Examples of simulated EOSD data are shown in the third row of Fig. 7.2. Although the nonlinear dependence of the VFPE evolution on the phase space density is also present when bunch profiles are used as input, in that case the longitudinal density is directly available from the input data. For EOSD measurements, this density must instead be inferred through the measurement model, making the nonlinear reconstruction more sensitive to the current phase space estimate.

The nonlinearity enters through the collective wake potential, which must be evaluated from the currently estimated phase space. Consequently, a non-physical reconstruction can produce an inconsistent wake potential, yielding an adjoint gradient that fails to properly reduce the error.

To avoid recomputing the gradient at every CGLS iteration, the linearized dynamics approach presented in Chapter 5 is applied. The sparse-matrix operators are kept fixed during 50 inner CGLS iterations and are then updated in an outer loop.

Figure 7.5 shows the evolution of the ground-truth error score $\hat{e}_\psi$ over the inner–outer CGLS iterations, comparing reconstructions without regularization (red) and with regularization (blue). The initial phase space density is taken as a two-dimensional Gaussian distribution. The corresponding wake potentials are estimated from Gaussian bunch profiles fitted to the measurements.

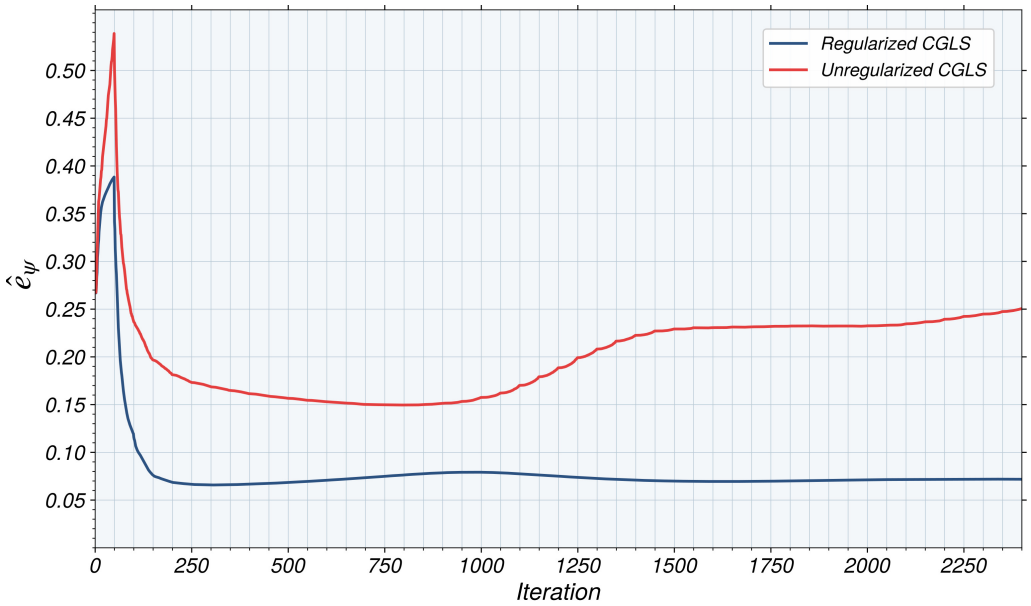


Figure 7.5.: Influence of regularization. Evolution of the ground-truth error score for the inner–outer CGLS reconstruction, comparing the Tikhonov regularized case in blue with the unregularized case in red. The linearized dynamics are over 50 inner CGLS iterations. The regularized reconstruction reaches a stable error level after approximately 2100 iterations, whereas the unregularized reconstruction does not reach a comparable stabilization within the evaluated range.

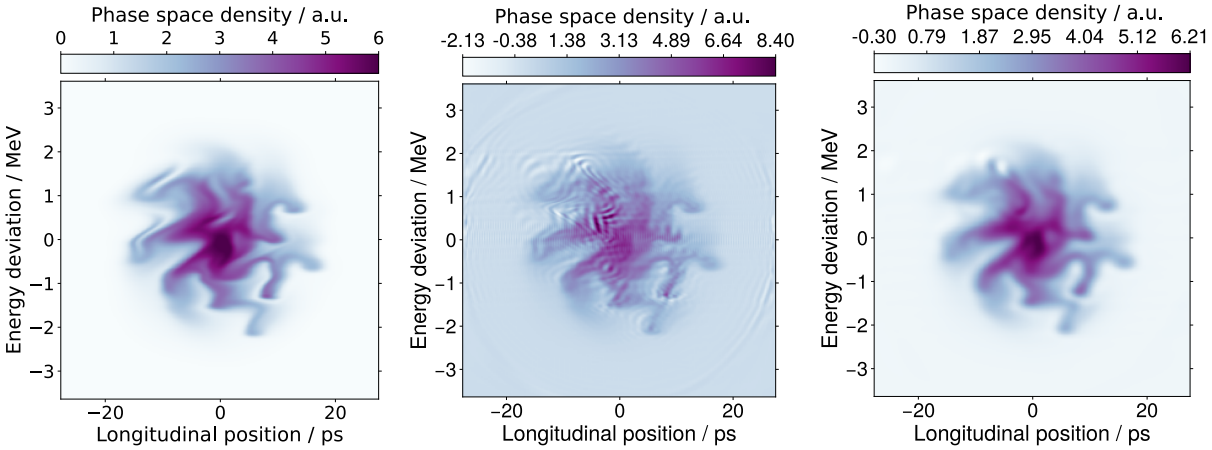
As shown in Fig. 7.5, the regularized reconstruction reaches a lower error score than the reconstruction without regularization. Nevertheless, both curves show a similar initial behavior. During the first inner loop, up to 50 CGLS iterations, the ground-truth error score increases. In this first outer iteration, the dynamics operators are computed from the initial two-dimensional Gaussian phase space and remain fixed throughout the inner loop. After 50 iterations, the dynamics are recomputed from the updated phase space estimate. The updated estimate provides a new longitudinal density and therefore a new wake potential, resulting in a more representative local approximation of the nonlinear dynamics and allowing the ground-truth error to decrease in the following outer iteration.

The main difference is the effect of regularization on the reconstructed phase space density itself. The Tikhonov regularization with a 2D Laplacian matrix penalizes rapid changes in the reconstructed phase space density. As a result, it reduces sharp oscillations and isolated artifacts in the reconstruction. This produces a smoother and more controlled reconstruction during the first inner loop, even when the initial phase space estimate is inaccurate. However, excessive smoothing can also suppress small scale

features. For the synthetic data, a regularization parameter of 1 was selected empirically by trial and error, without further fine tuning.

Once the linearized dynamics are updated, the phase space estimate becomes more consistent with the VFPE evolution. Consequently, the search direction improves, and both reconstructions approach better error scores after approximately four outer iterations, corresponding to 200 inner CGLS iterations.

A direct visual comparison is provided in Fig. 7.6. The unregularized reconstruction shows a stronger deviation from the ground-truth density than the regularized result. However, the regularized phase space density also appears smoother, since the Laplacian regularization reduces high-frequency components in the reconstruction.



(a) Ground-truth phase space density at turn 11000.

(b) Unregularized CGLS reconstruction after 1000 iterations.

(c) Regularized CGLS reconstruction after 1000 iterations.

Figure 7.6.: *Influence of regularization.* Comparison of phase space density reconstructions at turn 11000, illustrating its impact on the final CGLS estimate.

7.1.5. Warm-start initialization of the inner–outer loop

The previous subsection raises the question of how the first inner loop of the linearized dynamics reconstruction can be improved. A direct strategy is to start from an initial phase space density that is closer to the target distribution. For sequential reconstruction trajectories, this can be achieved by using the final reconstruction from the preceding trajectory as the initial estimate for the next one.

Figure 7.7 shows the same reconstruction methods and dataset used in Fig. 7.5, but with a warm-start initialization. The initial phase space density is taken from the reconstruction of the previous trajectory. This estimate is close to the target phase space density of the current reconstruction, but it is rotated by approximately $\theta = -1.09^\circ$ and slightly deformed.

The regularized reconstruction, shown in blue, has a more stable behavior. The error score decreases only weakly during the first inner loop and reaches a lower error score after approximately two outer iterations. The unregularized reconstruction, shown in red, presents unstable behavior. A better initial phase space estimate can reduce the number of outer iterations required by the regularized algorithm to reach an accurate reconstruction, while the subsequent oscillations indicate the need for an appropriate stopping criterion.

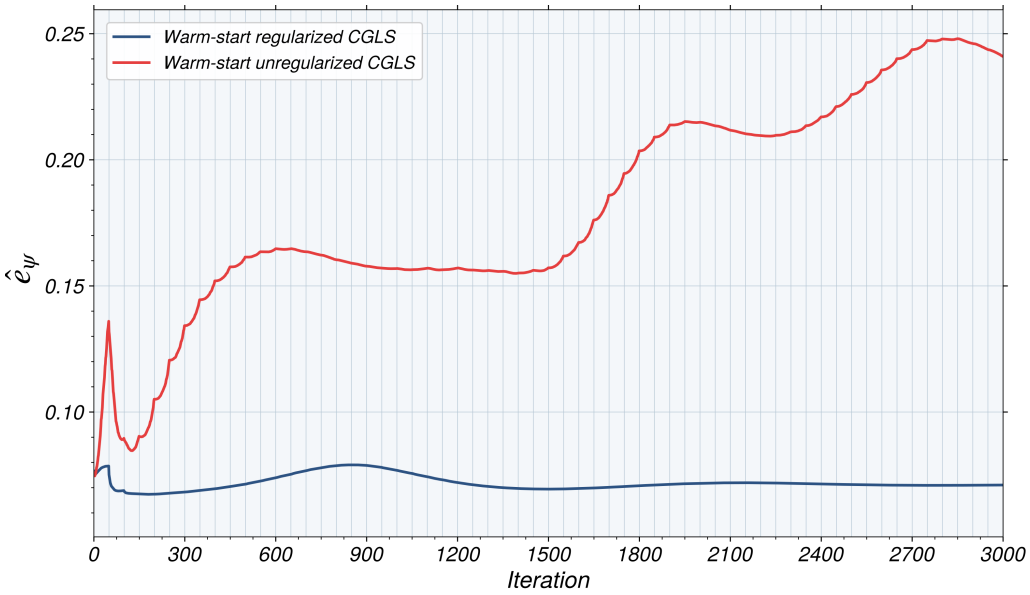
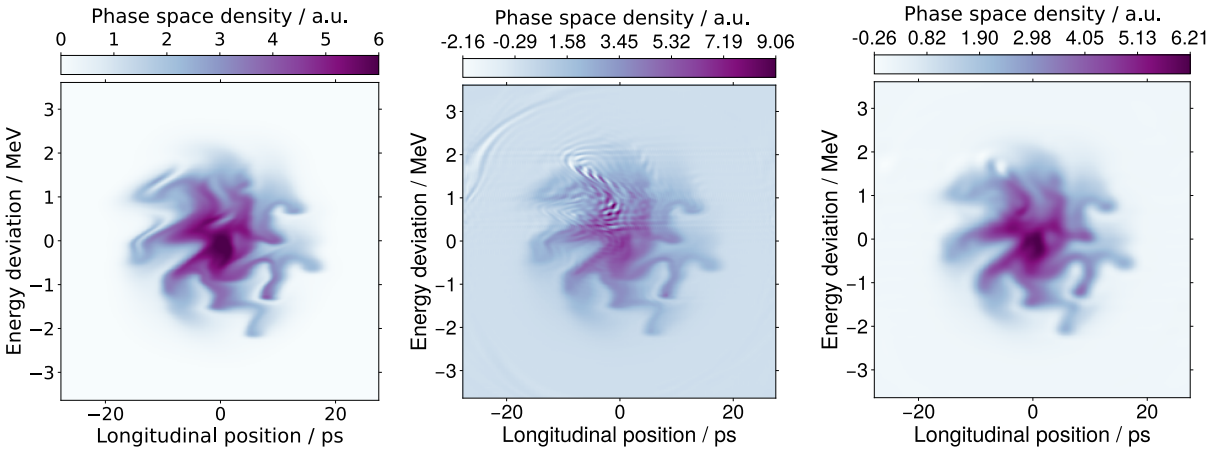


Figure 7.7.: Warm-start initialization. Evolution of the ground-truth error score $\hat{e}_\psi$ for the inner–outer CGLS reconstruction. The plot compares Tikhonov-regularized CGLS (blue) with the unregularized case (red). The initial phase space density is taken from the final regularized reconstruction of the preceding trajectory at turn $i = 10999$. The nonlinear dynamics are updated over 50 inner CGLS iterations. The regularized reconstruction reaches a stable error level after approximately 2900 iterations, whereas the unregularized reconstruction does not reach a comparable stabilization within the evaluated range.

Figure 7.8 provides a direct visual comparison illustrating the effects of a warm-start initialization. As shown, providing an initial phase space estimate improves the reconstruction quality for both the regularized and unregularized CGLS methods.



(a) Ground-truth phase space density at turn 11000.

(b) Warm-start unregularized CGLS reconstruction after 1000 iterations.

(c) Warm-start regularized CGLS reconstruction after 1000 iterations.

Figure 7.8.: *Warm-start initialization.* Comparison of phase space density reconstructions at turn 11000, illustrating its impact on both the unregularized and regularized final CGLS estimates.

7.1.6. Stabilization by non-negative projected restarts

Iterative least-squares methods such as CGLS do not impose non-negativity constraints in their standard formulation. Consequently, intermediate or final reconstructions may contain negative values, even though the phase space density is physically non-negative. This non-physical artifact is evident in all phase space reconstructions presented so far. Additional negative contributions may also arise from the high-order Lagrange-polynomial interpolation used in the Inovesa operators. In the context of numerical Vlasov equation solvers, Besse et al. [74] note that such interpolation can lead to zero or negative values.

To mitigate these non-physical contributions, a non-negative projected-restart strategy is employed [94]. After a prescribed number of CGLS iterations, the phase space vector is projected onto the non-negative constraint set by setting all negative components to zero (Sec. 5.4.6). The CGLS algorithm is then restarted using this projected solution as the new initial guess, with recomputed residuals and search directions. This effectively introduces an additional restart loop within the linearized dynamics reconstruction.

Figure 7.9 compares the inner–outer CGLS reconstruction with and without the non-negative projected-restart strategy. Both reconstructions use a Gaussian initial phase space density, and neither includes Tikhonov regularization. The projected-restarted reconstruction, shown in blue, achieves a lower error score than the CGLS reconstruction without projection, shown in red. This indicates that the non-negativity projection-restarted provides regularization by incorporating the prior physical information that a phase space density cannot be negative [95]. However, non-negativity alone does not replace explicit regularization or an appropriate stopping criterion.

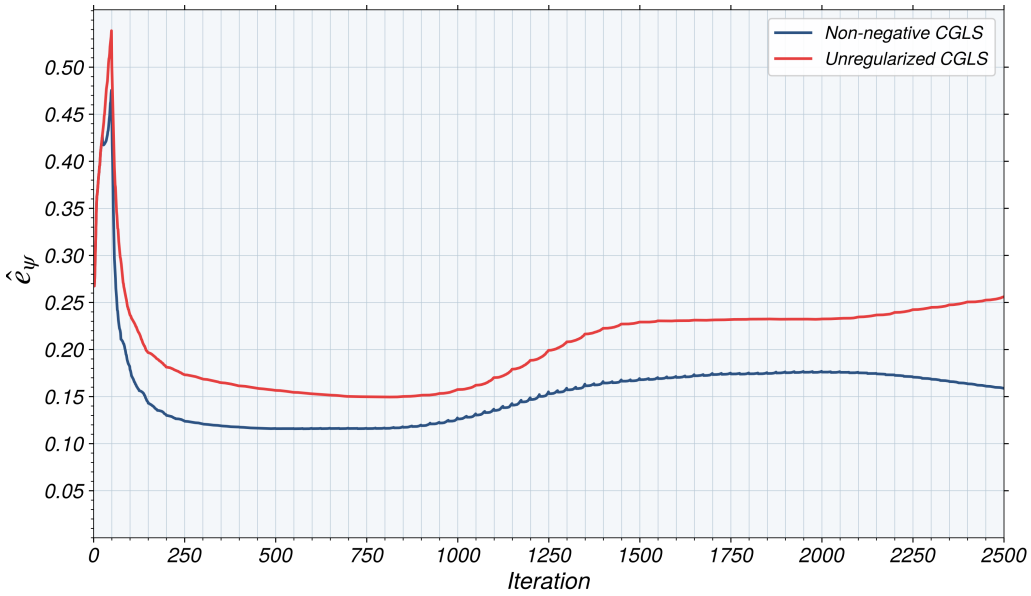
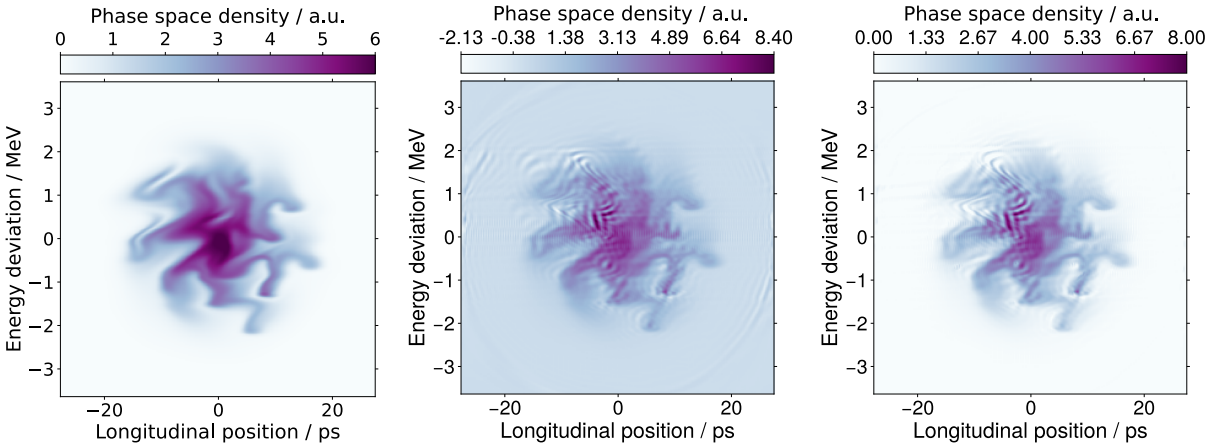


Figure 7.9.: Non-negative projected-restarted. Evolution of the ground-truth error score $\hat{e}_\psi$ for the inner–outer CGLS reconstruction. The plot compares projected-restarted non-negative CGLS (blue) with unregularized CGLS without projection (red). Neither method uses Tikhonov regularization. The non-negative projected-restarted CGLS performs two restart cycles of 25 inner iterations each, so that the linearized dynamics are recomputed after every 50 iterations.

Figure 7.10 illustrates that the projected-restart strategy yields a visible improvement. While the overall reconstruction does not achieve high accuracy in the absence of regularization, the resulting phase space successfully consists of non-negative values.



(a) Ground-truth phase space density at turn 11000.

(b) Unregularized inner–outer CGLS reconstruction without a warm-start initialization after 1000 iterations.

(c) Unregularized non-negative projected-restarted CGLS reconstruction without a warm-start initialization after 1000 iterations.

Figure 7.10.: *Non-negative projected-restarted.* Comparison of phase space density reconstructions at turn 11000. The figure illustrates how the projected-restart strategy restricts non-physical excursions and improves the final estimate when neither Tikhonov regularization nor warm-start initialization is applied.

7.1.7. Comparison of nonlinear and linearized dynamics reconstruction

A direct strategy for solving the nonlinear phase space density reconstruction problem is to recompute the gradient at every optimization iteration. In this case, the forward and adjoint operators are updated after each reconstruction update using the same sparse-matrix formulation described above. This provides the most direct treatment of the nonlinear dependence of the VFPE evolution on the phase space density, but it also increases the computational cost substantially.

Figure 7.11 first compares the nonlinear conjugate gradient approach with the linearized dynamics inner–outer CGLS method, where neither non-negative projected restarts nor Tikhonov regularization are applied. In this setting, the nonlinear method reaches the minimum stable error in less iterations, because the forward and adjoint operators are recomputed after every iteration. As a result, the gradient is continuously updated with the current VFPE trajectory, allowing the algorithm to rapidly correct deviations introduced by the initial phase space estimate.

In contrast, the linearized dynamics method requires more iterations to reach its minimum error score, since the operators remain fixed within each inner loop and are only updated in the outer loop. The main difference, however, lies in the computational cost, as shown in Fig. 7.12. For the same number of iterations, the linearized-dynamics inner–outer method is approximately 50 times faster than the nonlinear conjugate gradient algorithm.

The comparison is then repeated for the complete reconstruction strategy, including non-negative projected restarts and Tikhonov regularization. The results are shown in Fig. 7.13 in a time logarithmic scale plot. The linearized dynamics method reaches a small error score faster over the tested iteration range, because the projection and regularization stabilize the inner CGLS updates and reduce the sensitivity to the linearized trajectory.

In the nonlinear method, the same regularization leads to a slower decrease in the error score, but still guides the reconstruction toward a better solution than the unregularized case within the evaluated range.

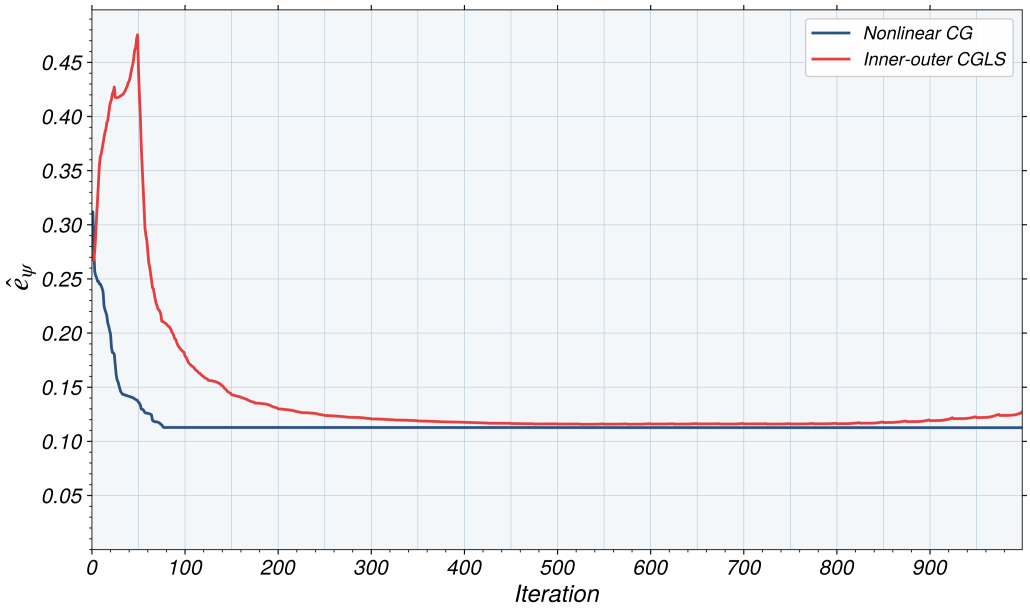


Figure 7.11.: Nonlinear conjugate gradient comparison. Evolution of the ground-truth error score $\hat{e}_\psi$ for the nonlinear conjugate gradient reconstruction (blue) and the linearized dynamics inner–outer CGLS reconstruction (red). Neither method includes non-negative projected restarts or Tikhonov regularization. In the nonlinear method, the forward and adjoint operators are recomputed after every optimizer iteration. In the linearized dynamics method, the operators are updated only in the outer loop; each outer iteration contains two restart cycles of 25 CGLS iterations.

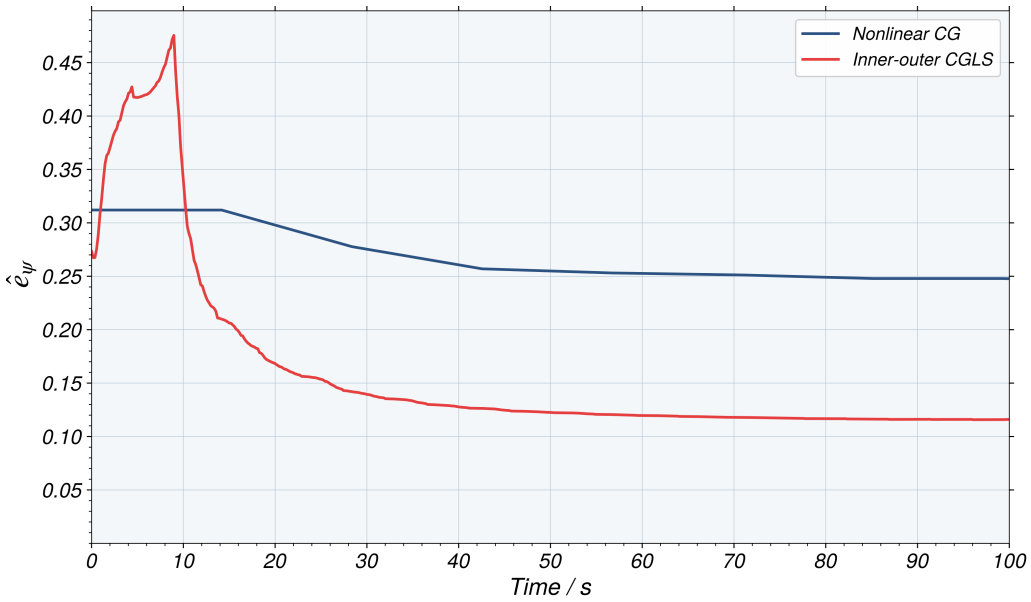


Figure 7.12.: Nonlinear conjugate gradient comparison. Evolution of the ground-truth error score over computation time for the linearized-dynamics inner–outer method (red) and the nonlinear conjugate gradient method (blue). Since the error score is recorded after each iteration, the two curves do not start from the same initial value. For the same number of iterations, the inner–outer method reaches lower error scores in substantially less computation time.

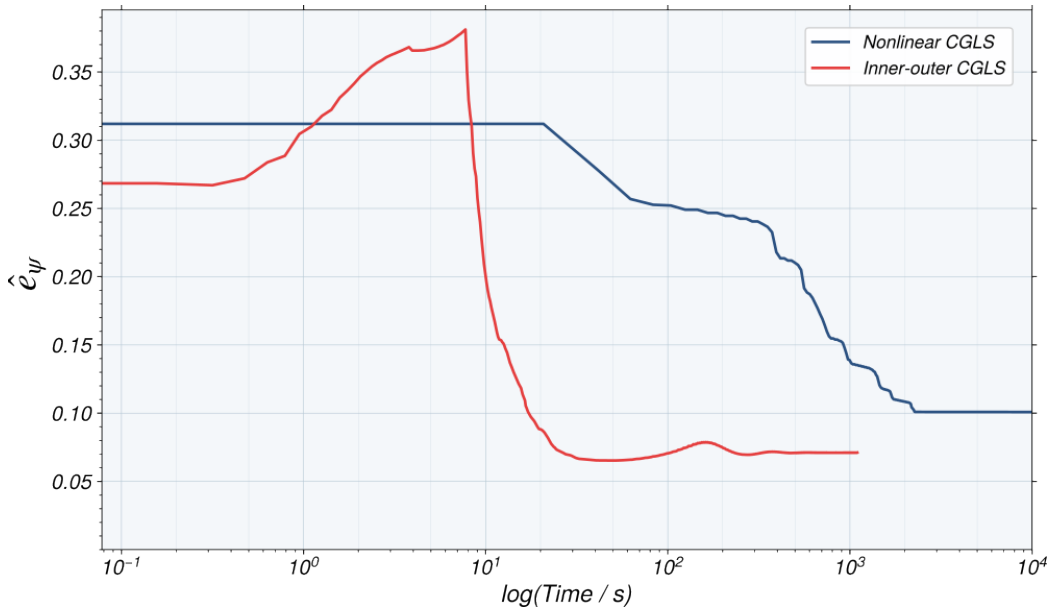
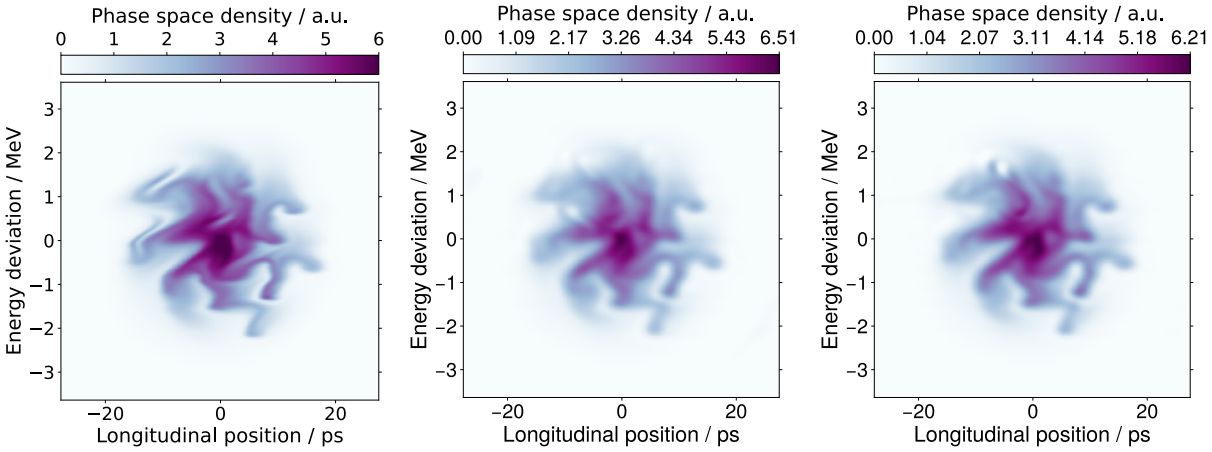


Figure 7.13.: Nonlinear conjugate gradient comparison. Evolution of the ground-truth error score $\hat{e}_\psi$ for the regularized nonlinear conjugate gradient reconstruction (blue) and the linearized dynamics inner–outer CGLS reconstruction (red) with non-negative projected restarts and Tikhonov regularization. In the nonlinear method, the forward and adjoint operators are recomputed after every optimizer iteration. In the linearized dynamics method, the operators are updated only in the outer loop; each outer iteration contains two restart cycles of 25 CGLS iterations.

Figure 7.14 illustrates that the inner–outer loop using linearized dynamics, Tikhonov regularization, and non-negative projected restarts yields a phase space estimate comparable to the nonlinear conjugate gradient method, while requiring significantly less computational time.



(a) Ground-truth phase space density at turn 11000.

(b) Regularized nonlinear conjugate gradient reconstruction with non-negative projected restarts and without a warm-start initialization after 1000 iterations.

(c) Regularized inner–outer CGLS reconstruction with non-negative projected restarts and without a warm-start initialization after 1000 iterations.

Figure 7.14.: Nonlinear conjugate gradient comparison. Comparison of phase space density reconstructions at turn 11000. Both reconstructions utilize Tikhonov regularization and non-negative projected restarts without a prior initial estimate.

The reported computational advantage of the inner–outer CGLS method applies to the tested discretization and phase space configuration. Since the relative performance of both methods may depend on the grid size and phase space extent, its transferability to other discretizations has not been established and remains to be investigated.

The following section evaluates the reconstructions across an extended sequence of EOSD measurements, detailing the relationship between the derived beam observables and the ground-truth data.

7.2. Reconstruction from simulated EOSD data

The objective of this section is to verify the phase space density reconstruction using consecutive simulated EOSD data for which the ground truth is known. This allows the reconstructed phase space densities to be compared directly with the reference Inovesa trajectory.

The reconstruction is performed with the regularized inner–outer CGLS method, including non-negative projected restarts and warm-start initialization from the reconstruction of the previous turn. The simulation parameters used in the reconstruction are the same as those used to generate the simulated EOSD data, as listed in Table 7.2.

The subset of simulated EOSD data starts at the turn highlighted by the green line in Fig. 7.1. It is chosen to cover the transition from a regime without visible micro-bunching, through a developed micro-bunching instability, and back toward a more stable regime. This selection contains 5200 consecutive measurements and follows the corresponding CSR intensity evolution.

Figure 7.15 shows the error score between the reconstructed phase space densities and the ground-truth trajectory for this measurement subset. The agreement is worst in the region where micro-bunching is present and improves considerably when the phase space evolution becomes more stable.

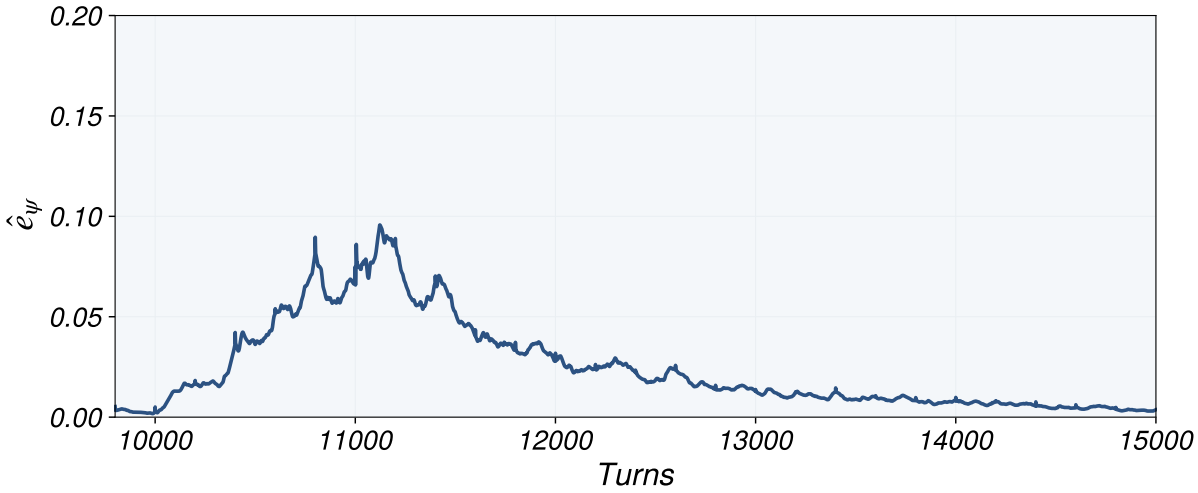


Figure 7.15.: Error score between the reconstructed phase space densities and the ground-truth trajectory for 5200 consecutive measurements selected around the micro-bunching regime. The reconstruction is obtained with the regularized non-negative inner–outer CGLS method. Higher error values correspond to regions with stronger micro-bunching dynamics, while lower values are obtained when the phase space evolution is more stable.

Figure 7.16 compares the ground-truth observables with the corresponding observables derived from the reconstructed phase space density and shows their error scores. The error scores are calculated analogously to Eq. (7.3), but using the scalar values of the CSR power, bunch length, and energy spread

relative to their corresponding ground-truth values. The comparison includes the longitudinal bunch length, the energy spread, and the CSR intensity. The errors remain low, with mean values of 0.0007, 0.0020, and 0.0302, respectively. The larger CSR-intensity error is expected, since the CSR intensity depends quadratically on the longitudinal bunch spectrum [12]. Therefore, small deviations in the reconstructed bunch profile can produce comparatively larger discrepancies in the CSR signal. The spikes observed in the bunch length and energy spread errors may be related to the synchrotron oscillation period, although their exact origin cannot be determined from the present analysis.

Overall, the error curves in Fig. 7.16 show that the reconstructed phase space captures the main longitudinal and energy dynamics, while the CSR intensity remains more sensitive to small differences in the longitudinal structure.

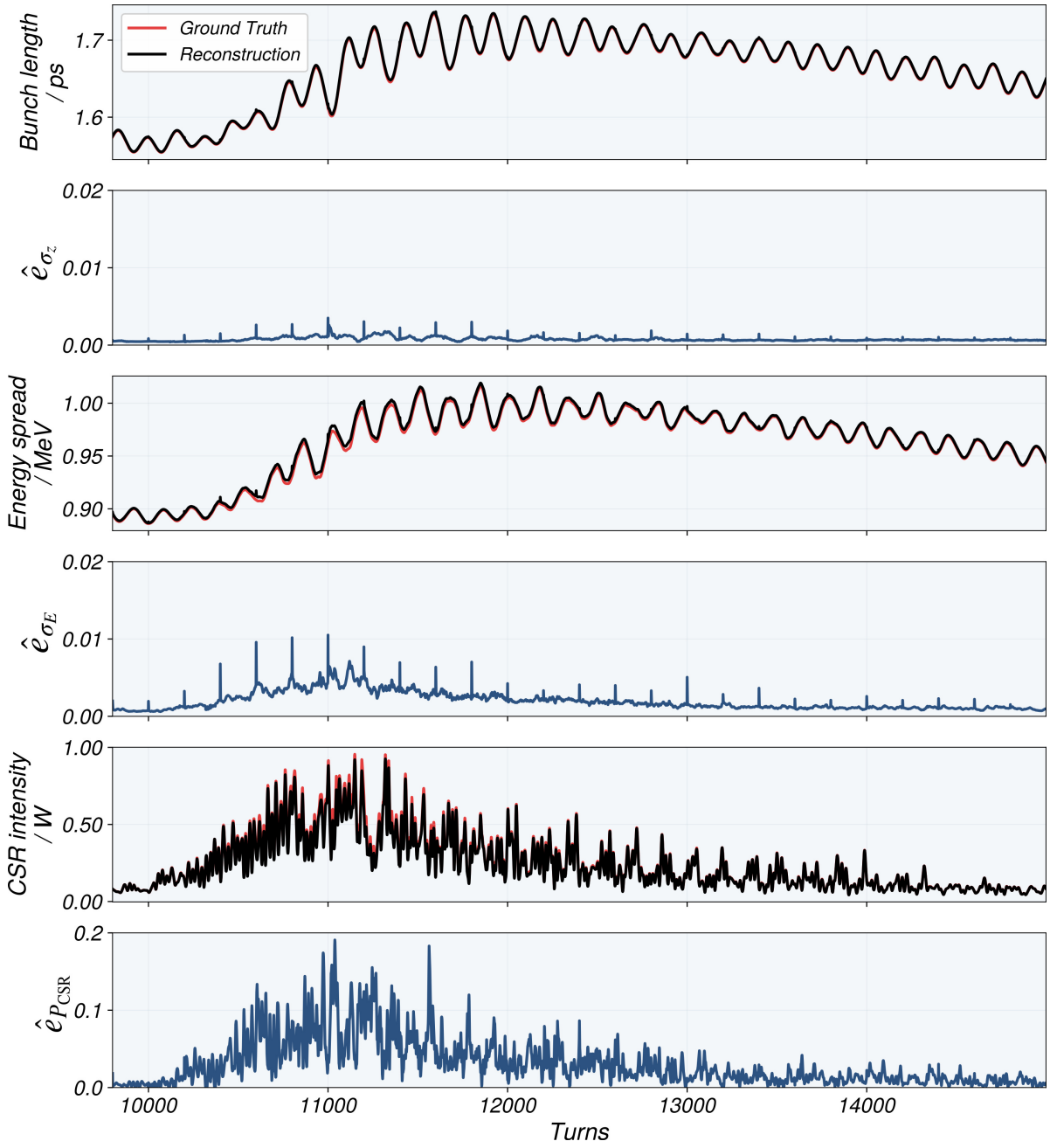


Figure 7.16.: Comparison between ground-truth observables and observables derived from the reconstructed phase space density, together with the corresponding error scores. The ground truth is shown in red and the reconstruction in black. The longitudinal bunch length, energy spread, and CSR intensity have mean error scores of 0.0007, 0.0020, and 0.0302, respectively.

7.2.1. Sensitivity to bunch current mismatch

In experimental operation, the effective bunch current used in the reconstruction model may differ from the value that best represents the measured bunch dynamics. In a storage ring, the beam current also decreases continuously if no additional bunches or electrons are injected. Therefore, tracking the bunch

current and providing an appropriate value to the reconstruction model is important. A controlled current offset is therefore a direct and physically meaningful stress test for the reconstruction.

Figure 7.17 shows the error score between the reconstructed phase space density and the ground-truth phase space density along the tested dataset for five current values used in the reconstruction. The blue curve corresponds to the baseline case presented in the previous section, with a bunch current of $I_b = 0.85\,\mathrm{mA}$. The green and yellow curves correspond to $I_b = 0.75\,\mathrm{mA}$ and $I_b = 0.95\,\mathrm{mA}$, respectively, representing offsets of $\pm 0.1\,\mathrm{mA}$ from the nominal value. These cases show only a moderate deviation from the baseline reconstruction.

The pink curve corresponds to a much larger current mismatch, $I_b = 1.85\,\mathrm{mA}$, and shows the strongest model error among the tested cases. At such a high current, the collective wake contribution is overestimated with respect to the ground-truth trajectory, so the reconstruction model drives stronger longitudinal phase space deformation, including enhanced micro-bunching structures and changes in the bunch length and energy spread.

Finally, the purple curve corresponds to the non-physical case $I_b = 0$, where the wake potential is zero and the self-interaction effects are neglected. As expected, this case produces larger errors in the micro-bunching regime.

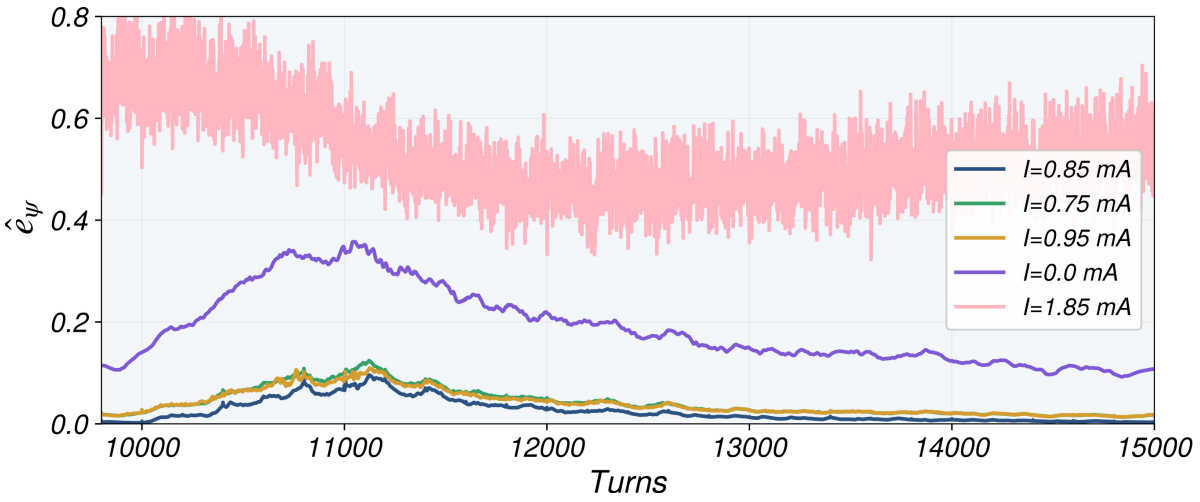


Figure 7.17.: Sensitivity of the phase space reconstruction to bunch-current mismatch. The error score is evaluated between the reconstructed phase space density and the ground-truth phase space density for different current values used in the reconstruction model. The blue curve shows the nominal case, $I_b = 0.85\,\mathrm{mA}$; the green and yellow curves show moderate offsets, $I_b = 0.75\,\mathrm{mA}$ and $I_b = 0.95\,\mathrm{mA}$; the pink curve shows a large overestimated current, $I_b = 1.85\,\mathrm{mA}$; and the purple curve shows the case without collective wake interaction, $I_b = 0$.

7.2.2. Sensitivity to RF voltage amplitude mismatch

The RF voltage, V_{RF}, directly affects the synchrotron motion and therefore the rotation rate of the longitudinal phase space. Errors in this parameter modify the modeled phase space evolution and can make it more difficult for the reconstruction to follow the EOSD measurements. For this reason, an RF-voltage mismatch provides an important validation case. Only mismatches in the amplitude of V_{RF} are considered here.

Figure 7.18 shows the error score between the reconstructed phase space density and the ground-truth phase space density along the tested dataset for three RF-voltage values. The blue curve corresponds to the baseline case with $V_{\mathrm{RF}} = 799.257\,\mathrm{kV}$. The green curve shows a -10% deviation, with $V_{\mathrm{RF}} = 719.331\,\mathrm{kV}$, while the yellow curve shows a $+10\%$ deviation, with $V_{\mathrm{RF}} = 879.182\,\mathrm{kV}$. Both offsets lead to a clear deviation from the ground truth, especially in dynamically active regions, although the reconstruction remains reasonably stable in regimes with weaker phase space deformation.

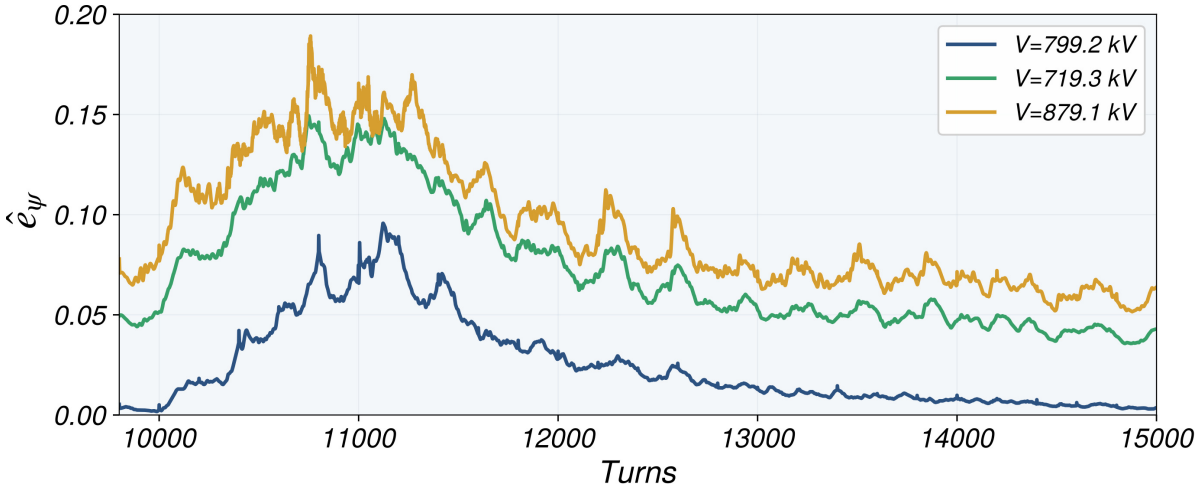


Figure 7.18.: Sensitivity of the phase space reconstruction to RF-voltage amplitude mismatch. The error score is evaluated between the reconstructed phase space density and the ground-truth phase space density for three RF-voltage values used in the reconstruction model. The blue curve shows the nominal case, $V_{\mathrm{RF}} = 799.257\,\mathrm{kV}$; the green curve shows a -10% voltage offset, $V_{\mathrm{RF}} = 719.331\,\mathrm{kV}$; and the yellow curve shows a $+10\%$ voltage offset, $V_{\mathrm{RF}} = 879.182\,\mathrm{kV}$.

7.2.3. Sensitivity to impedance model mismatch

The impedance model determines the wake potential used in the VFPE dynamics. In simulations, this impedance can be specified consistently with the model used to generate the data. In experimental reconstructions, however, the machine impedance is only approximately known. Testing different impedance models on synthetic data therefore provides a controlled evaluation of how sensitive the reconstruction is to this source of model mismatch.

Figure 7.19 shows the error score between the reconstructed phase space density and the ground-truth phase space density along the tested dataset for two impedance models. The blue curve corresponds to the baseline reconstruction using the parallel-plates CSR impedance, while the green curve corresponds to the free-space CSR impedance. As expected, the free-space model does not reproduce the shielding effects introduced by the vacuum chamber geometry. Consequently, the reconstructed dynamics deviate substantially from the ground truth, showing that the impedance model is a critical component of the reconstruction algorithm.

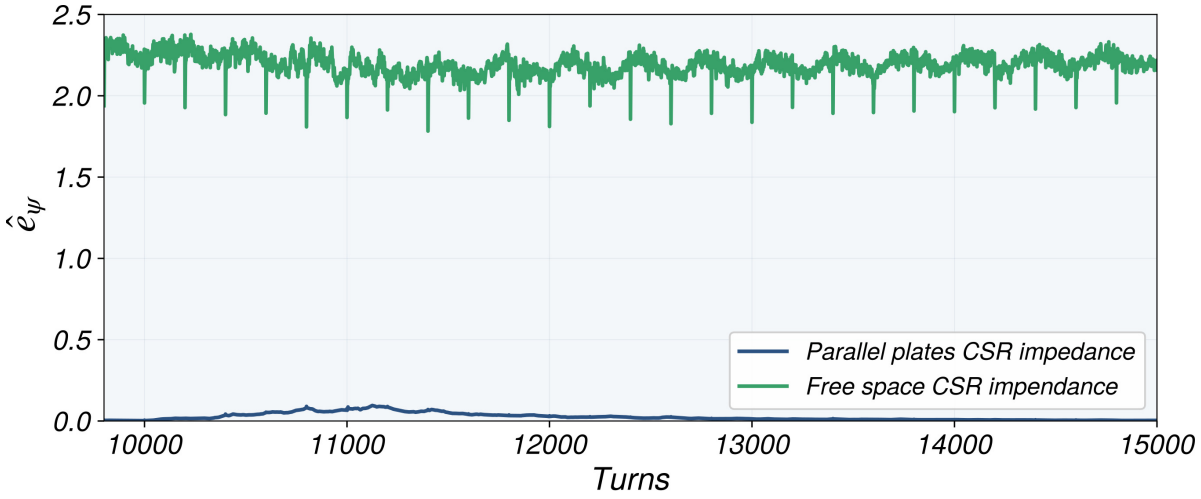


Figure 7.19.: Sensitivity of the phase space reconstruction to impedance model mismatch. The error score is evaluated between the reconstructed phase space density and the ground-truth phase space density for two impedance models used in the reconstruction. The blue curve shows the baseline case using the parallel-plates CSR impedance, while the green curve shows the reconstruction obtained with the free-space CSR impedance.

7.3. Reconstruction from EOSD measurements

The final section applies the selected algorithm to diagnostic measurements at KARA. In contrast to the synthetic validation, the true phase space density is unknown, so the reconstructed phase space must be validated indirectly. The first requirement is data consistency: the EOSD signals predicted from the reconstructed phase space must reproduce the EOSD measurements. The second requirement is physical consistency: observables derived from the reconstructed phase space should agree with diagnostics that were not used directly in the optimization objective.

The reconstruction is performed with the non-negative projected-restarted CGLS method in sequential mode. In this mode, each reconstruction in the sequence is initialized with the phase space density reconstructed in the previous step. This provides a warm-start strategy that improves the initialization of consecutive reconstructions.

The experimental data consist of synchronized EOSD, VLM, and THz power measurements, as introduced in Chapter 3. The EOSD diagnostic provides the longitudinal measurement sequence used in the reconstruction. The VLM signal provides turn-by-turn horizontal bunch-size information, while the THz power signal measures CSR emission associated with short longitudinal structures and micro-bunching dynamics.

The simulation parameters corresponding to the machine settings at the time of the measurements are listed in Table 7.4.

From the machine parameters listed in Table 7.4, the number of turns per synchrotron period is obtained as

$$N_s = \frac{f_r}{f_s} = 261.11, \tag{7.4}$$

$$m = \frac{N_s}{2} \approx 130. \tag{7.5}$$

Parameter	Value
Beam energy	1.3 GeV
Bunch current	0.886 mA
Synchrotron frequency (f_s)	10.4 kHz
Accelerating voltage	1 300 kV
Revolution frequency (f_r)	2.715 51 MHz
Relative natural energy spread	$4.7 \cdot 10^{-4}$
Bending radius	5.559 m
Damping time	10.0 ms
Harmonic number	184
Vacuum gap	0.032 m
Linear RF	True
Phase space size (dimensionless)	14.0
Grid size	91 x 91 cells

Table 7.4.: Physic related Inovesa configuration parameters.

7.3.1. Beam diagnostic data

Each diagnostic first provides a sensor specific signal that is processed into the quantity used for the analysis. For the EOSD system, this quantity is the phase retardation,[1] for the VLM it is the horizontal beam profile, and for the Schottky diode it is the measured THz power. From these diagnostic signals, the observables used for comparison are obtained: the longitudinal bunch length σ_z from EOSD, the horizontal bunch size σ_x from VLM, and the THz power as an indicator of CSR emission.

Figure 7.20 shows a synchronized subset of the diagnostic data acquired by P. Schönfeldt, B. Kehrer, J. Steinmann, and M. Brosi [13]. The subset considered in this thesis covers a continuous window of 73,000 consecutive turns of the electron bunch in the storage ring.

The high-frequency structures visible in the EOSD and VLM sinograms, appearing as darker regions near the center, are correlated with increased THz power. This correlation is characteristic of micro-bunching instability and identifies the regime studied in this thesis. Figure 7.21 shows a detailed view of the micro-bunching instability regime.

The derived observables used for comparison are the longitudinal bunch length from EOSD and the horizontal bunch size from VLM. They are obtained by first calculating the center of mass of each measurement and then evaluating the root-mean-square distance from this center. Since the measured distributions approximately reproduce the corresponding bunch profiles, this quantity provides an estimate of their standard deviation and therefore of the bunch length or horizontal bunch size. This procedure was implemented and validated in [13] and is used to compare the diagnostic data with the quantities derived from the reconstructed phase space density.

Alternative extraction methods are also possible. For VLM measurements, Kehrer [47] used a filament beam spread function (FBSF), simulated with OpTaliX, convolved with a Gaussian profile to fit the measured data. This is conceptually similar to the EOSD treatment used in this thesis, where the

[1] In some works, EOSD measurements are referred to directly as longitudinal bunch profiles. In this thesis, however, the measured quantity is treated as the EOSD phase retardation, which is related to the longitudinal bunch profile through the EOSD observation model.

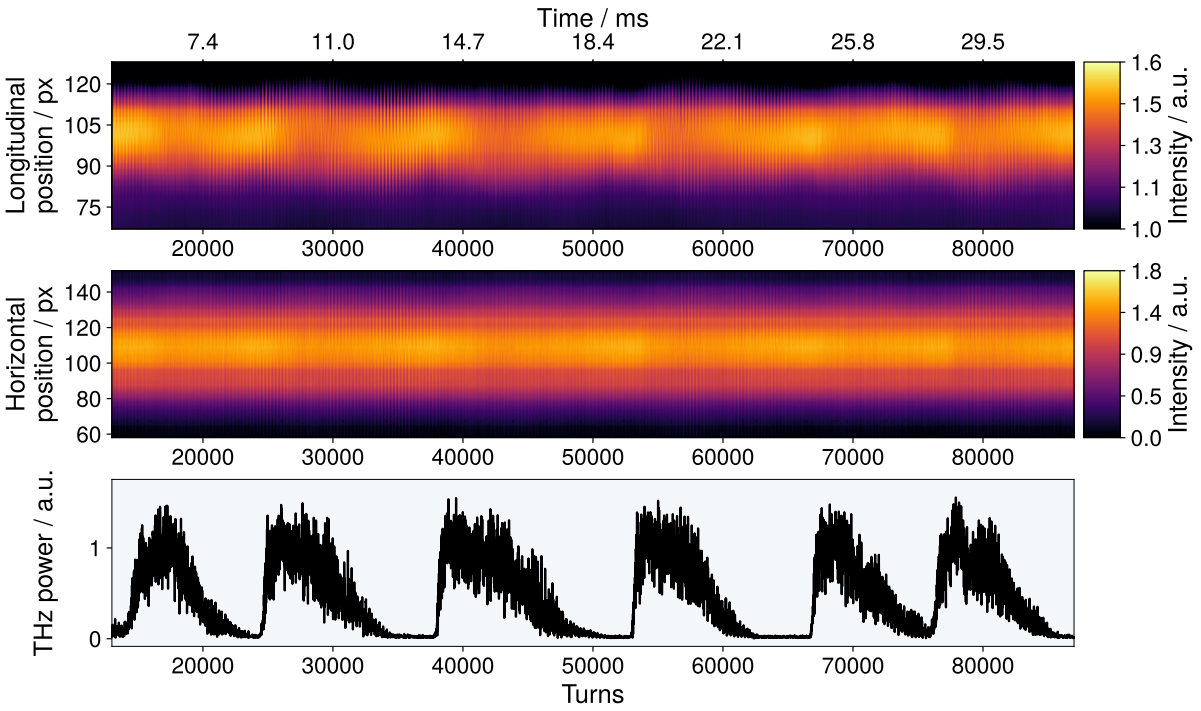


Figure 7.20.: Subset of the synchronized diagnostic data acquired by M. Brosi et al. [13], covering 73 000 consecutive turns used for the experimental reconstruction. The upper panel shows the EOSD phase-retardation sinogram, the middle panel the VLM horizontal beam-profile sinogram, and the lower panel the THz power measurements.

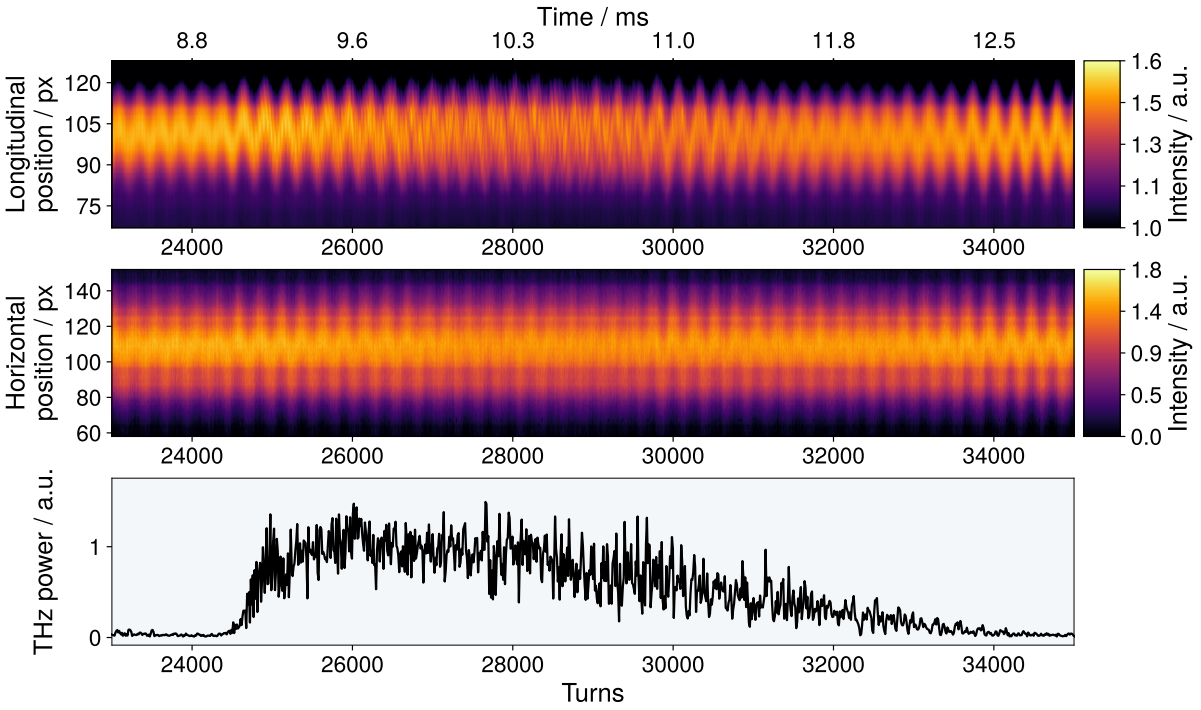


Figure 7.21.: Detailed view of the synchronized base diagnostic data in the micro-bunching instability regime. The upper panel shows the EOSD phase-retardation sinogram, the middle panel shows the VLM horizontal beam-profile sinogram, and the lower panel shows the THz power measurements.

impulse response of the EOSD setup is used to relate a bunch profile to the measured phase retardation signal.

This distinction is important for the validation strategy used in this chapter. The diagnostic observables are not treated as absolute ground truth for the reconstructed phase space density. Instead, they provide validated experimental indicators that follow the expected beam dynamics at KARA during the measurements and can therefore be used for consistency checks.

The observables contain both the slower evolution of the instability and a synchrotron oscillation with a measured period of approximately $N_s = 261$ turns. To compare the slower trends between diagnostics, the observables are smoothed with a Hamming-window filter, reducing the synchrotron-scale contribution while preserving the slower envelope.

The Hamming-window filter is a weighted moving average in which the central points of the window receive larger weights than the edges. The filtered signal is obtained as a normalized weighted average, and the reduced edge weights avoid abrupt cutoffs, providing smoother suppression of the synchrotron-frequency component than a simple moving average.

The window length is chosen so that the amplitude of the synchrotron frequency component is reduced to at most 10% of its original value. Because the THz signal is sampled every 11.568 turns, the THz window is selected first. The shortest integer window satisfying this criterion is $N_{\mathrm{THz}} = 36$, corresponding to approximately 416.45 turns. The turn-by-turn observables are therefore filtered with the matched Hamming-window length $N_{\mathrm{EOSD}} = N_{\mathrm{VLM}} = 416$.

With this choice, all diagnostic observables are filtered over approximately the same physical time interval. The filtered signals are used to compare the slower observable trends, whereas the unfiltered signals retain the full turn-by-turn dynamics.

Figure 7.22 shows the diagnostic observables before and after applying the Hamming-window filter. The pink curves show the unfiltered quantities, while the red curves show the filtered trends used for comparison with the reconstructed phase space observables in the following analysis.

The filtered derived observables are not used as input to the phase space reconstruction. The reconstruction uses only the EOSD sinogram, without applying the window filtering used for the comparison signals. The sinogram is, however, restricted to the valid region of 91 pixels, which determines the phase space grid size used in the reconstruction.

For visualization only, the measured and reconstructed observables are min–max normalized over the displayed interval. The quantitative metrics are computed from the filtered, non-normalized signals as defined below.

For an observable measurement z_i, the normalized quantity is defined as

$$\tilde{z}_i = \frac{z_i - \min(z)}{\max(z) - \min(z)}, \tag{7.6}$$

where the minimum and maximum are evaluated over the full set of measurements in the considered window. The same min–max normalization is applied to the observables derived from the reconstruction.

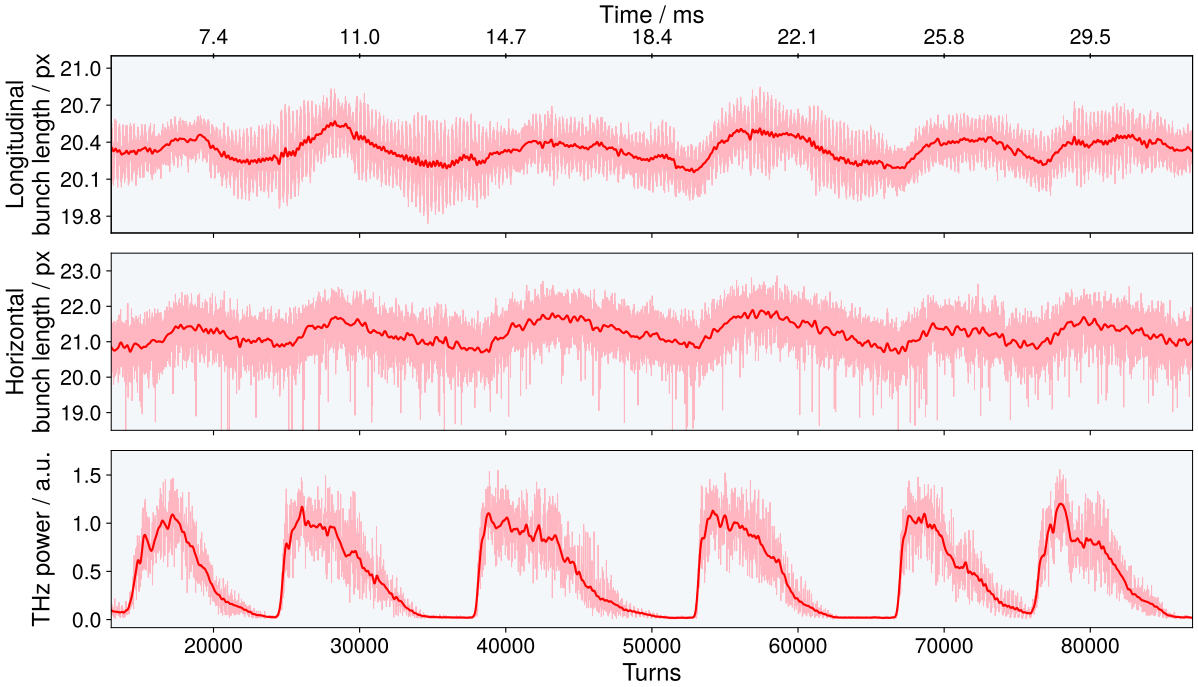


Figure 7.22.: Diagnostic observables before and after Hamming-window filtering. The pink curves show the unfiltered observables, and the red curves show the filtered trends. The upper panel shows the EOSD longitudinal bunch length σ_z filtered with $N_{\mathrm{EOSD}} = 416$, the middle panel shows the VLM horizontal bunch size σ_x filtered with $N_{\mathrm{VLM}} = 416$, and the lower panel shows the THz power filtered with $N_{\mathrm{THz}} = 36$.

7.3.2. Phase space size and initial bunch length

Section 7.1.5 showed that the initial phase space estimate influences the performance of the reconstruction algorithm. For the experimental reconstruction, both the phase space size and the initial bunch length are therefore estimated from the measurements before starting the optimization.

The phase space size is chosen so that the simulation grid spacing matches the temporal sampling of the EOSD measurements. With a fixed number of grid points, the EOSD sampling directly determines the grid spacing and therefore the phase space size in Inovesa, ensuring that the simulated bunch profile and the EOSD impulse response are represented on the same temporal grid as the measurements.

The initial longitudinal bunch length is estimated from the first EOSD measurement of the reconstruction dataset, denoted by Γ_1. The purpose is to obtain an approximate bunch size consistent with the measured signal and use it to initialize the reconstruction.

A test Gaussian bunch profile is first convolved with the EOSD impulse response in full-convolution mode. Since Γ_1 contains only a finite portion of this response, its position within the full convolution is determined by maximizing the cross-correlation between the two signals. This alignment defines the valid time window for the subsequent profile estimation.

Using the aligned window, the corresponding convolution operator is assembled as a Toeplitz matrix H. The bunch profile ρ associated with Γ_1 is then estimated from the constrained least-squares problem

$$\min_{\rho \geq 0} \|H\rho - \Gamma_1\|_2^2 \,. \tag{7.7}$$

This provides a non-negative bunch profile consistent with the first EOSD measurement. Since a single longitudinal bunch profile is insufficient to determine the two-dimensional phase space density, the deconvolved profile cannot be used directly for initialization. Instead, a Gaussian is fitted to estimate the characteristic bunch size and used to construct a rotationally symmetric initial phase space density.

7.3.3. Consistency with EOSD measurements

A direct measure of reconstruction performance is the agreement between the measured EOSD signal and the EOSD signal predicted from the first phase space density of the reconstructed trajectory. However, a good fit to the EOSD measurements alone does not guarantee a physically consistent reconstruction, since non-physical phase space densities may still reproduce the measured signals. The EOSD fit is therefore used as a data consistency criterion and is later complemented by comparisons with independent beam diagnostics.

The reconstruction obtained using the algorithmic choices summarized in Table 7.1 is taken as the reference reconstruction for the experimental analysis. The corresponding phase space evolution represents the main experimental reconstruction result, while the subsequent sensitivity studies examine variations around this reference.

To compare the experimental EOSD measurements with the signals derived from the reconstructed phase space density, the normalized error definition introduced in Eq. (7.3) is used. Here, the experimental EOSD measurement is taken as the reference, and the EOSD signal predicted from the reconstructed phase space density is evaluated relative to it.

The first evaluation considers the EOSD error score over the full experimental dataset. Figure 7.23 shows the evolution of this score over the 73000 reconstructed measurements. The unfiltered values are shown in light blue, the window-filtered trend in dark blue, and the mean value in black.

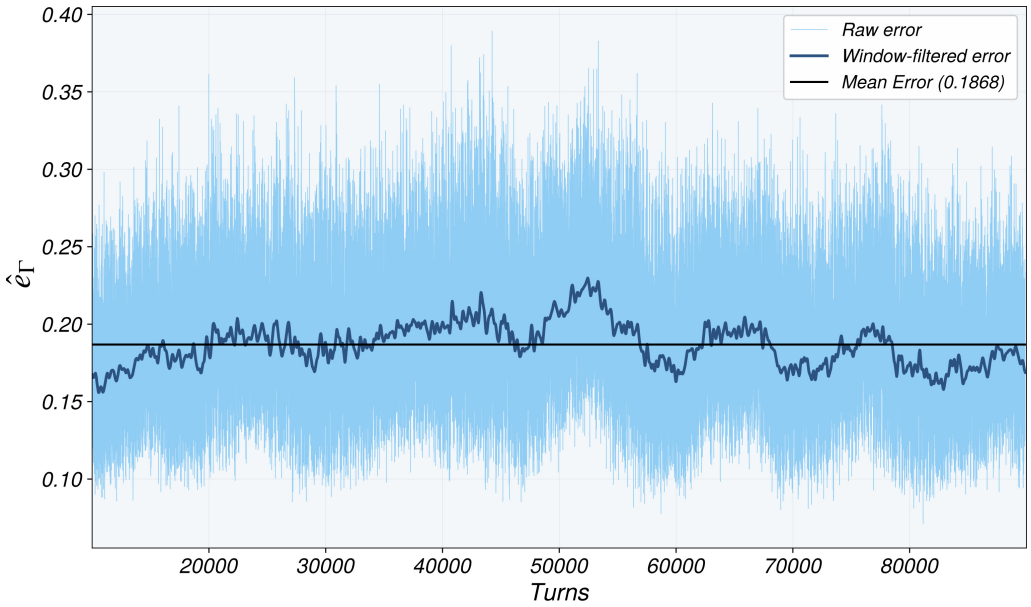
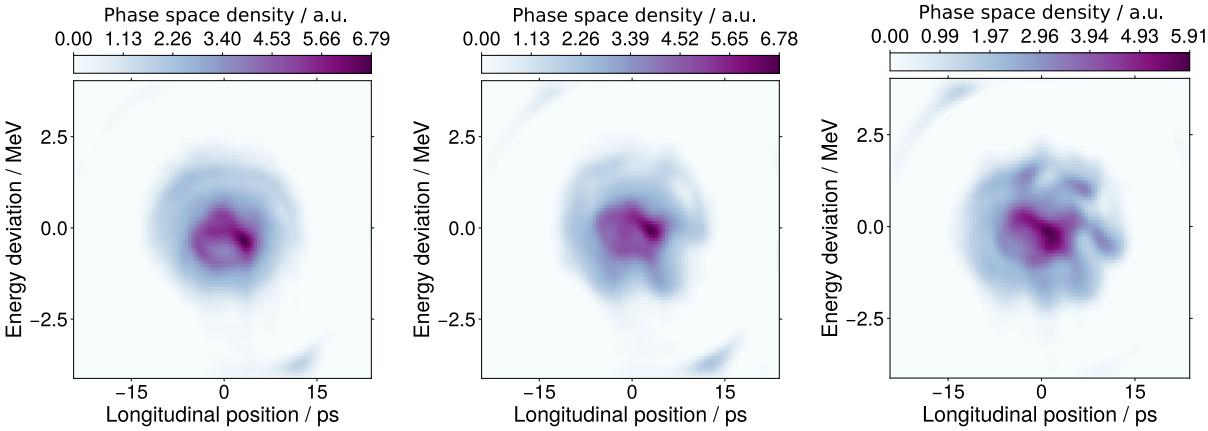


Figure 7.23.: EOSD error score over the full reconstructed dataset. The light blue curve shows the unfiltered error values, the dark blue curve shows the window-filtered trend using 416 data points, and the black line indicates the mean value. The dataset contains 73000 measurements. The mean error score is 0.1868, with a standard deviation of 0.0408.

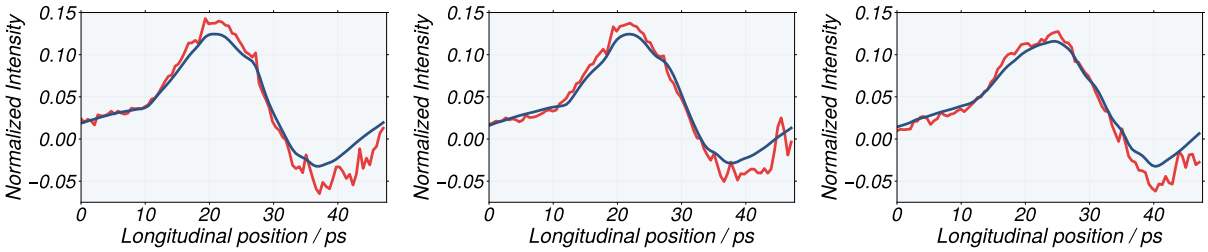
Figure 7.24 shows representative EOSD measurement comparisons together with the corresponding reconstructed phase space densities. The examples are selected from different dynamical regimes: a region without visible micro-bunching, the onset of micro-bunching, and a fully developed micro-bunching regime. The ring-like structure visible away from the center of the reconstructed phase space density is associated with the mismatch on the right-hand side of the EOSD signal comparisons in Figs. 7.24d, 7.24e, and 7.24f. This mismatch is hypothesized to arise partly from impulse-response mismatch. The influence of impulse-response mismatch is examined in Sec. 7.3.5.



(a) Phase space reconstruction at turn 23000.

(b) Phase space reconstruction at turn 25000.

(c) Phase space reconstruction at turn 27000.



(d) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 23000.

(e) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 25000.

(f) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 27000.

Figure 7.24.: Comparison of experimental EOSD measurements and reconstructed phase space densities at three selected turns. The upper row shows the reconstructed phase space densities, while the lower row compares the experimental EOSD measurements in red with the EOSD signals predicted from the reconstructed phase space in blue. The selected turns illustrate different stages of the measured dynamics, from no visible micro-bunching to the onset of micro-bunching and a developed micro-bunching regime.

A complementary evaluation examines the EOSD agreement along the reconstructed trajectory. Each reconstruction uses m EOSD measurements, and an error score is computed for each measurement index, from Γ_1 to Γ_m. To obtain a statistical description over the full dataset, the scores corresponding to the same measurement index are averaged across all reconstruction windows; for example, all errors associated with Γ_1 are averaged together.

Figure 7.25 shows the resulting mean EOSD error score along the reconstructed trajectory, together with the standard deviation over all reconstruction windows. The mean error is lower during the first quarter of the trajectory, increases toward the middle, and partially decreases near the end. Because the standard deviation is large, this variation should be interpreted cautiously rather than as a definitive trajectory dependent trend.

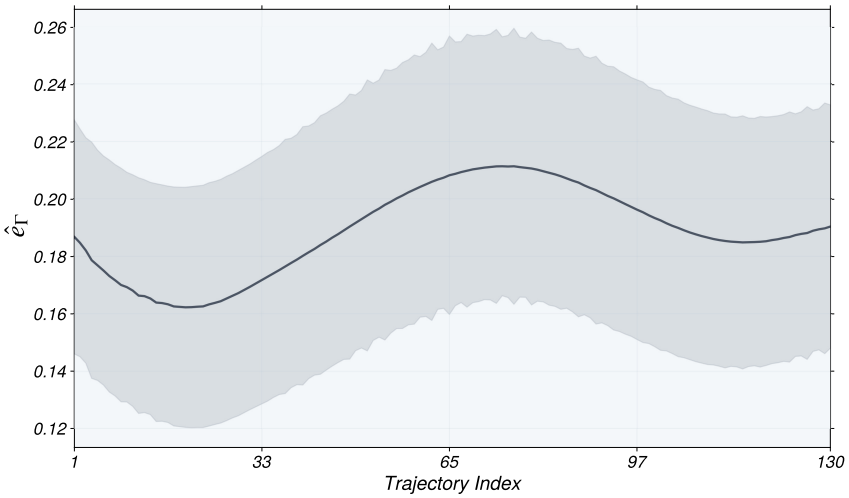


Figure 7.25.: Mean EOSD error score along the reconstructed trajectory, evaluated over the full dataset of 73000 measurements. The solid curve shows the average error score at each relative measurement position within the trajectory, and the shaded region indicates the corresponding standard deviation. The mean error score over all trajectory positions is 0.1825.

7.3.4. Comparison of derived observables

In this section, the phase space density reconstruction is applied to the full experimental dataset, producing one reconstructed phase space density for each EOSD measurement. From each reconstructed distribution, several observables are derived, including the longitudinal bunch length, the energy spread, and the CSR power.

The bunch length and energy spread are obtained from the RMS values of the corresponding phase space projections. Since the phase space coordinates are centered around zero, these RMS values provide the equivalent widths of the projected distributions. However, as shown in Fig. 7.24, the reconstructed phase space can contain an outer ring-like noise contribution. This contribution also appears in the projected distributions and can bias the RMS evaluation. To reduce this effect, the RMS calculation is restricted to a window whose length is 10% larger than the full width at half maximum (FWHM) of the initial Gaussian bunch profile fitted to the EOSD measurement. This truncation reduces the influence of noise contributions near the edges of the reconstructed phase space.

The CSR power is derived from the impedance model and the Fourier transform of the reconstructed bunch profile. While the numerical calculation can be performed over the full frequency range, the experimental THz signal is measured with a Schottky diode whose sensitivity is limited to a finite frequency band. Therefore, the frequency range used for the CSR comparison is restricted to the detector bandwidth, here taken as 70 GHz to 1.1 THz.

Figure 7.26 shows the observables derived from the reconstructed phase space density. The unfiltered values are shown in light blue, while the window-filtered trends are shown in blue. A visual comparison with Fig. 7.22 shows that the reconstructed observables follow the same qualitative behavior as the diagnostic data: the bunch length, energy spread, and CSR power increase when micro-bunching appears. The unfiltered observables also show fast oscillatory components that are reduced by the window filtering.

Figure 7.27 compares the measured observables with those derived from the reconstructed phase space density. The error score used for signals representing the same quantity is not appropriate here

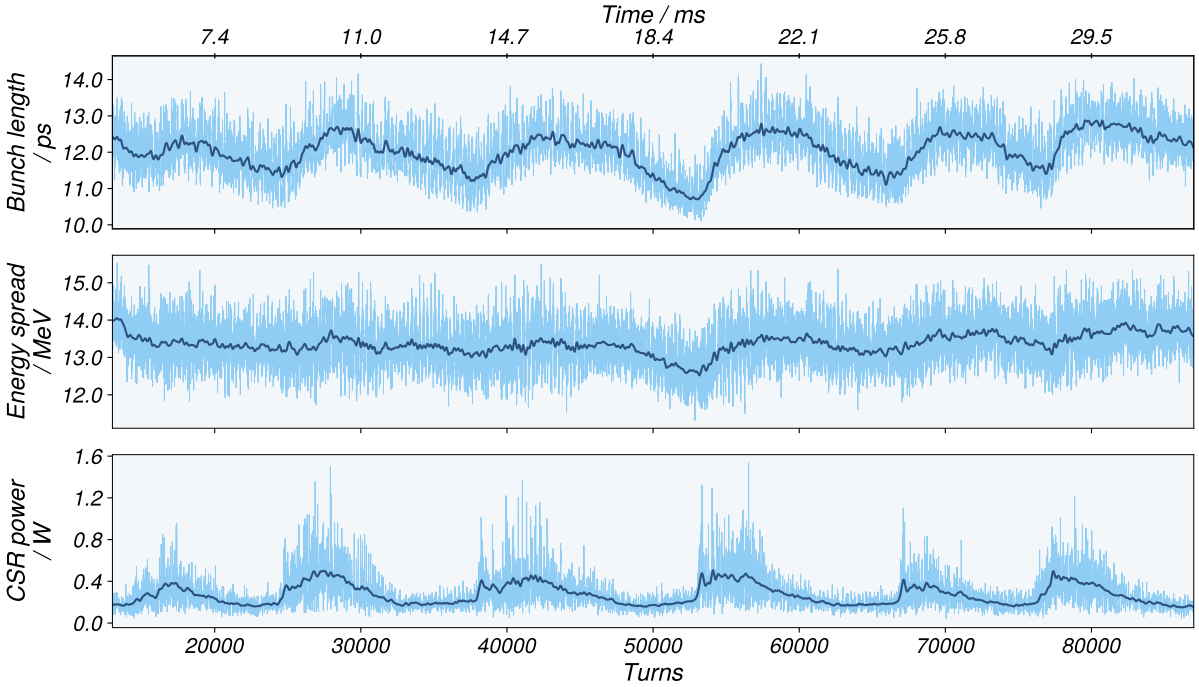


Figure 7.26.: Observables derived from the reconstructed phase space density. The light blue curves show the unfiltered reconstructed observables, while the blue curves show the corresponding window-filtered trends. The upper panel shows the longitudinal bunch length, the middle panel shows the energy spread, and the lower panel shows the CSR power.

because the reconstructed and measured observables have different units, gains, offsets, and diagnostic responses. Therefore, the comparison is quantified using the zero-lag Pearson correlation $r(0)$, the Spearman rank correlation $\rho(0)$, and a robust normalized root-mean-square error (NRMSE).

The Pearson and Spearman coefficients are calculated directly from the paired signals, without applying an additional normalization. Pearson correlation measures linear trend agreement, while Spearman correlation measures monotonic agreement based on the rank ordering of the values. Both coefficients range from -1 to 1, where values close to 1 indicate strong positive agreement, values close to 0 indicate little association, and values close to -1 indicate an inverse relation.

For the NRMSE calculation, an affine fit is first applied to map each reconstructed observable to the scale of the corresponding measurement. The fit uses a constant offset and gain determined over the complete evaluation interval. This step accounts for differences in units, detector gain, and baseline, but it does not provide an absolute calibration. The remaining disagreement is then quantified with a robust NRMSE, normalized by the 5–95 percentile range of the measured signal, following the standard use of RMSE-based error measures for model evaluation [96]. The NRMSE quantifies the remaining disagreement after a constant scale and offset have been removed, but it does not provide an absolute calibration.

Table 7.5 shows that all three observables exhibit high trend agreement. The EOSD comparison mainly tests data consistency, since the EOSD signal enters the reconstruction directly. In contrast, the THz signal provides an independent validation that the reconstructed phase space captures the temporal evolution of the CSR emission. The VLM comparison shows the weakest agreement among the three diagnostics. Its slightly higher Spearman coefficient indicates that the correspondence is marginally stronger in terms of monotonic ordering than in terms of a strictly linear relation. This may be explained

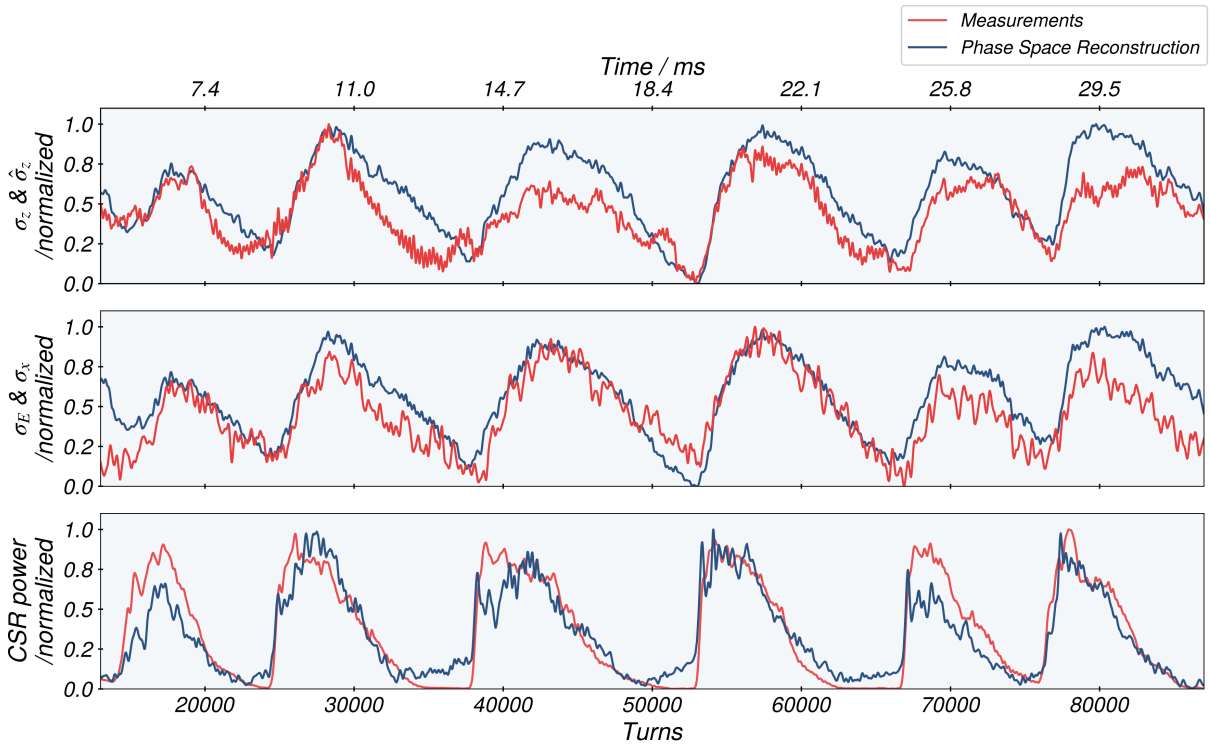


Figure 7.27.: Comparison between measured observables (red) and observables derived from the reconstructed phase space density (blue) after min–max normalization. The upper panel compares the reconstructed longitudinal bunch length σ_z with the measured EOSD bunch length $\hat{\sigma_z}$. The middle panel compares the reconstructed energy spread σ_E with the measured horizontal bunch size σ_x from VLM. The lower panel compares the reconstructed CSR power with the measured THz power signal.

by the fact that the horizontal bunch size is only an indirect proxy for the energy spread, so their relation is not expected to be strictly one-to-one.

Table 7.5.: Full-range agreement between reconstructed and measured observables. All scores are calculated from the filtered, non-normalized signals; min–max normalization is used only to display the curves together.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8702	0.8723	0.1546
VLM energy-spread	0.8341	0.8514	0.1732
THz CSR	0.9131	0.8855	0.1519

A clear advantage of the phase space reconstruction is that, once the model parameters are set to the accelerator conditions, the reconstructed phase space is obtained with physical units. The derived observables can therefore be expressed directly in the corresponding units. However, the reconstructed phase space size, and consequently the bunch length and energy spread, depend on the effective laser pulse length used in the EOSD model. For the dataset used in this thesis, this value was not measured directly. Therefore, the experimentally derived value reported in [56] is used, with $\sigma = 20.2\,\mathrm{ps}$, corresponding to $\mathrm{FWHM} = 2.355\sigma = 47.6\,\mathrm{ps}$. The sensitivity to the EOSD impulse response is evaluated in the following section.

7.3.5. Sensitivity to the EOSD impulse response

The experimental reconstruction depends directly on the impulse response used to model the EOSD diagnostic. This impulse response defines how a longitudinal bunch profile is transformed into the measured phase retardation signal. Therefore, deviations in its shape, temporal alignment, or effective laser pulse length can modify the reconstructed phase space density.

This sensitivity is especially important because the impulse response is obtained from a simulated model and cannot represent the experimental EOSD setup perfectly. Small differences in the effective laser pulse length, the electro-optical response, or the timing between the laser pulse and the electron bunch can produce systematic discrepancies between the measured and reconstructed EOSD signals. These discrepancies can then appear as artificial structures in the reconstructed phase space density.

The purpose of this subsection is to evaluate how robust the reconstruction is with respect to the chosen EOSD impulse response. Different impulse-response configurations are tested while keeping the reconstruction algorithm and beam-dynamics parameters fixed. The resulting changes in the EOSD fit, reconstructed phase space density, and derived observables are then compared.

The first sensitivity test compares the full EOSD impulse response with the simplified laser-trajectory model presented by M. Reissig [56], in which only the downstream path through the crystal is included. The effective laser pulse length, corresponding to the measurement width, is kept equal to the value used in the previous section, $\mathrm{FWHM} = 47.6\,\mathrm{ps}$.

Table 7.6.: Full-range agreement between reconstructed and measured observables for an impulse response calculated using only the downstream laser trajectory through the EOSD crystal.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8489	0.8669	0.1659
VLM energy-spread	0.2748	0.3103	0.3019
THz CSR	0.8544	0.8629	0.1937

Table 7.6 shows a notable result. Although the agreement scores for the EOSD bunch length and THz CSR are close to those reported in Table 7.5, the VLM energy-spread proxy shows substantially weaker agreement. In addition, the mean EOSD measurement error over the full dataset is 0.1497, which is lower than the value shown in Fig. 7.23.

This lower EOSD error, together with the poorer agreement with the independent VLM observable, suggests that the downstream-only impulse response may fit the EOSD measurements more closely while reducing the physical consistency of the reconstructed phase space. This result highlights the importance of the independent-diagnostic validation used in this study: a reduction of the measurement error does not necessarily imply a better physical reconstruction, and can otherwise lead to a false positive if the independent diagnostics are not considered.

The second impulse-response sensitivity test modifies the effective laser pulse length used in the EOSD model. Two values are considered: a shorter pulse length of $35\,\mathrm{ps}$ and a longer pulse length of $65\,\mathrm{ps}$, relative to the experimental estimate of $47.6\,\mathrm{ps}$ used in the baseline reconstruction.

Table 7.7 shows the agreement metrics obtained with the impulse response calculated for a laser pulse length of $\mathrm{FWHM} = 35\,\mathrm{ps}$. The EOSD bunch-length agreement remains high, but the VLM energy-spread proxy shows weaker agreement, and the THz CSR signal becomes anticorrelated with

Table 7.7.: Full-range agreement between reconstructed and measured observables for an impulse response calculated with a laser pulse length of $\mathrm{FWHM} = 35\,\mathrm{ps}$.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8466	0.8786	0.1670
VLM energy-spread proxy	0.4987	0.5407	0.2722
THz CSR	-0.3003	-0.3715	0.3555

the measured trend. The mean EOSD measurement error over the full dataset is 0.2090, which by itself does not reveal the strong loss of physical agreement indicated by the independent diagnostics.

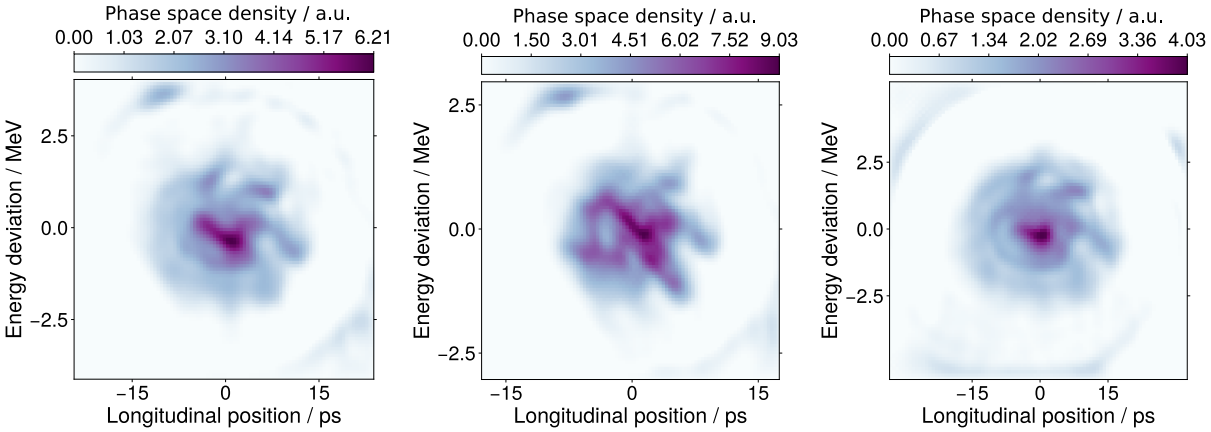
Table 7.8 shows the corresponding results for a laser pulse length of $\mathrm{FWHM} = 65\,\mathrm{ps}$. In this case, the EOSD bunch-length and THz CSR agreement remain high, while the VLM energy-spread proxy is still weaker than in the baseline case. However, the mean EOSD measurement error increases to 0.4147, showing that the impulse-response mismatch is also visible in the direct EOSD fit. These results illustrate the difficulty of assessing reconstruction quality from a single metric: a model can improve or preserve agreement in one observable while degrading consistency with another.

Table 7.8.: Full-range agreement between reconstructed and measured observables for an impulse response calculated with a laser pulse length of $\mathrm{FWHM} = 65\,\mathrm{ps}$.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8620	0.8678	0.1591
VLM energy-spread proxy	0.7003	0.7381	0.2241
THz CSR	0.8718	0.8332	0.1826

Finally, Fig. 7.28 illustrates the effect of the impulse-response variations discussed above on the reconstructed phase space density at turn 27000. This turn is selected because micro-bunching structures are visible in the experimental dataset. The figure compares the downstream case with the two modified laser pulse lengths, showing how the different impulse-response assumptions affect both the reconstructed phase space density and the corresponding EOSD signal fit.

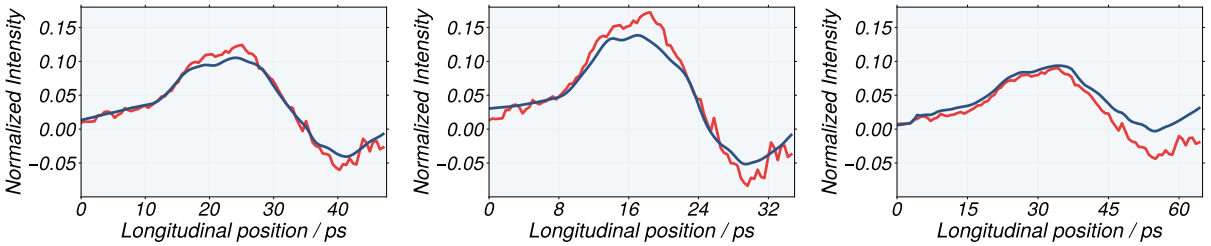
In the downstream case, Fig. 7.28d shows a small mismatch between the measured and reconstructed EOSD signals. This mismatch is associated with a stronger outer ring-like noise contribution in the reconstructed phase space. For the modified laser pulse lengths, Figs. 7.28e and 7.28f show changes in both the width and amplitude of the reconstructed EOSD signal. The shorter pulse length produces a narrower signal, while the longer pulse length produces a broader signal with a poorer fit to the measurement. Since the inferred bunch length directly affects the simulated longitudinal dynamics, an incorrect laser pulse length can lead to a substantially different beam response. In both cases, the reconstructed signal intensity deviates substantially from the baseline reconstruction.



(a) Phase space reconstruction at turn 27000 using an EOSD impulse response calculated with only the downstream laser trajectory through the crystal.

(b) Phase space reconstruction at turn 27000 using an EOSD impulse response calculated with a laser pulse length of 35 ps.

(c) Phase space reconstruction at turn 27000 using an EOSD impulse response calculated with a laser pulse length of 65 ps.



(d) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 27000, using the impulse response calculated with only the downstream laser trajectory.

(e) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 27000, using the impulse response calculated with a laser pulse length of 35 ps.

(f) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 27000, using the impulse response calculated with a laser pulse length of 65 ps.

Figure 7.28.: Comparison of reconstructed phase space densities and EOSD signal fits for three tested EOSD impulse-response models at turn 27000, where micro-bunching structures are present. The upper row shows the reconstructed phase space densities obtained with an impulse response calculated using only the downstream laser trajectory, with a laser pulse length of 35 ps, and with a laser pulse length of 65 ps. The lower row compares the corresponding experimental EOSD measurements in red with the EOSD signals predicted from the reconstructed phase space densities in blue.

These examples also show that similar qualitative structures in the reconstructed phase space density are not sufficient to guarantee a physically correct reconstruction. Even if the distributions exhibit expected features, such as micro-bunching structures, their connection to the real particle dynamics can be compromised when the EOSD impulse-response model is not accurate. Therefore, the physical interpretation of the reconstructed phase space must always be tied to its consistency with the diagnostic model and with independent beam measurements.

7.3.6. Sensitivity to model parameters

The experimental reconstruction depends on several accelerator parameters that are not known with perfect accuracy, including, among others, the bunch current, RF voltage, machine impedance model, momentum compaction factor, radiation damping time, and natural energy spread. The simulation

validation in Sec. 7.2 showed the expected sensitivity to the bunch current, RF voltage, and impedance model. In this section, the experimental study is restricted to two directly relevant machine parameters: the bunch current used in the reconstruction model and the RF voltage.

Each parameter is varied within a physically plausible range while the remaining parameters are kept fixed. The resulting changes in the reconstructed phase space density and in the derived observables, including bunch length, energy spread, and CSR-related signal, are then compared. The goal is not to tune the parameters until all diagnostics agree optimally, but to assess which reconstruction features remain robust under reasonable modeling variations.

Table 7.9 shows the agreement metrics obtained when the bunch current in the reconstruction model is reduced to $I_b = 0.786\,\mathrm{mA}$, compared with the baseline value $I_b = 0.886\,\mathrm{mA}$ used in the previous section. The agreement remains high for all three observables, with the strongest correlation obtained for the THz CSR signal. The mean EOSD measurement error over the full dataset is 0.1827.

Table 7.9.: Full-range agreement between reconstructed and measured observables for a bunch current of $I_b = 0.786\,\mathrm{mA}$. The baseline reconstruction in the previous section used $I_b = 0.886\,\mathrm{mA}$.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8712	0.8728	0.1541
VLM energy-spread proxy	0.8427	0.8591	0.1691
THz CSR	0.9068	0.8966	0.1571

Table 7.10 shows the corresponding metrics for an increased bunch current of $I_b = 0.986\,\mathrm{mA}$. The EOSD bunch-length agreement remains comparable to the lower-current case, while the VLM energy-spread proxy and THz CSR agreement decrease. This trend is consistent with the experimental conditions, since the bunch current decreases continuously during the measurement. A lower current is therefore more likely to be closer to the effective current during the reconstructed interval than a larger current. The mean EOSD measurement error over the full dataset is 0.1906.

Table 7.10.: Full-range agreement between reconstructed and measured observables for a bunch current of $I_b = 0.986\,\mathrm{mA}$. The baseline reconstruction in the previous section used $I_b = 0.886\,\mathrm{mA}$.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8697	0.8715	0.1549
VLM energy-spread proxy	0.8248	0.8431	0.1775
THz CSR	0.8626	0.7736	0.1885

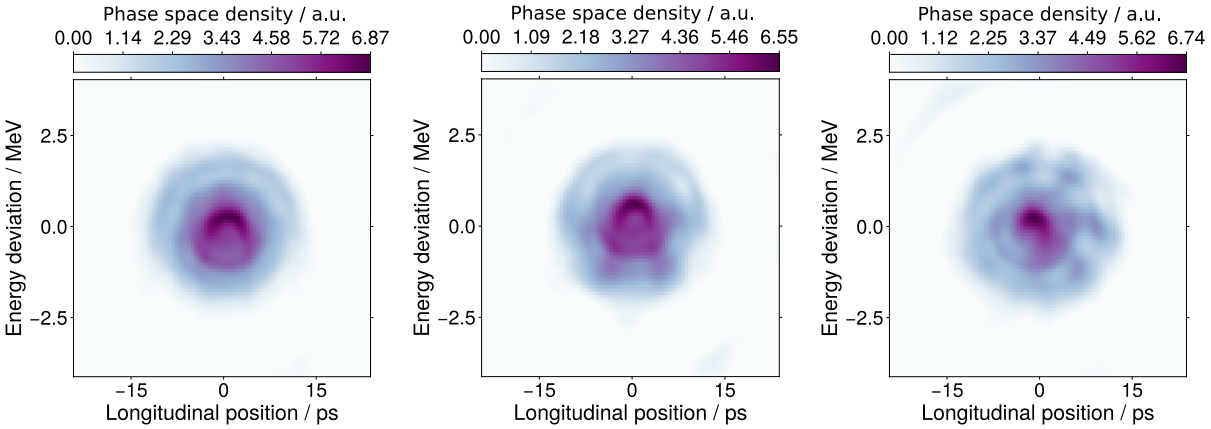
Table 7.11 shows the metrics obtained when the collective current contribution is removed by setting $I_b = 0\,\mathrm{mA}$. In this case, the EOSD bunch-length and VLM energy-spread proxy still show high trend agreement, and the mean EOSD measurement error is 0.1733, close to the baseline value. The THz CSR comparison is not reported because the reconstructed bunch profile does not produce a valid frequency contribution in the THz range used for the CSR evaluation.

This is a noteworthy result. The EOSD measurement error remains close to the baseline, and the direct comparison between measured and reconstructed EOSD signals still appears reasonably good. In addition, the reconstructed phase space density in Fig. 7.29 shows structures that resemble micro-bunching. However, since the collective wake contribution is absent in this reconstruction model, these structures should not be interpreted as physical micro-bunching dynamics. Instead, they indicate that the optimization algorithm can reduce the EOSD measurement error by introducing patterns

that emulate the measured trend. Although the reconstruction remains partially consistent with the measured EOSD trend, the structures appear less physically natural than in the baseline case, reinforcing the need for validation beyond the EOSD fit alone.

Table 7.11.: Full-range agreement between reconstructed and measured observables for the current-free case, $I_b = 0\,\mathrm{mA}$. The baseline reconstruction in the previous section used $I_b = 0.886\,\mathrm{mA}$.

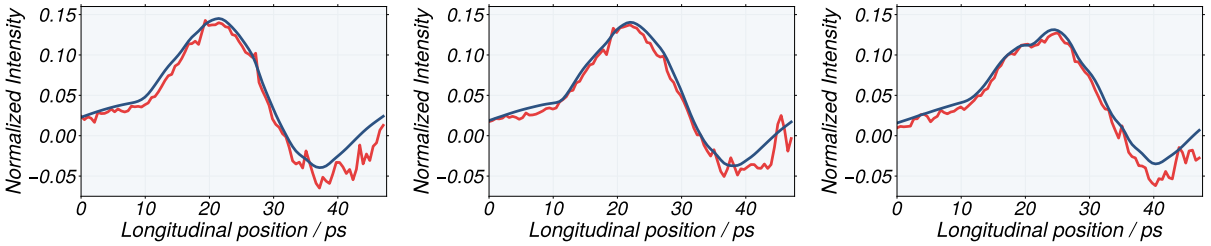
Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8515	0.8498	0.1646
VLM energy-spread proxy	0.8909	0.8994	0.1427
THz CSR	-	-	-



(a) Phase space reconstruction at turn 23000.

(b) Phase space reconstruction at turn 25000.

(c) Phase space reconstruction at turn 27000.



(d) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 23000.

(e) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 25000.

(f) Experimental EOSD measurement (red) and reconstructed EOSD signal (blue) at turn 27000.

Figure 7.29.: Comparison of experimental EOSD measurements and reconstructed phase space densities at three selected turns for $I_b = 0\,\mathrm{mA}$, with all other reconstruction parameters kept fixed. The upper row shows the reconstructed phase space densities, while the lower row compares the experimental EOSD measurements in red with the EOSD signals predicted from the reconstructed phase space densities in blue.

Table 7.12 shows the agreement metrics obtained for a modified RF voltage of $V_{\mathrm{RF}} = 1\,170\,\mathrm{kV}$. The EOSD bunch-length and THz CSR observables retain strong correlation, while the VLM energy-spread proxy shows a lower correlation and larger NRMSE. The mean EOSD measurement error over the full dataset is 0.1954.

Table 7.13 shows the agreement metrics obtained for an increased RF voltage of $V_{\mathrm{RF}} = 1\,430\,\mathrm{kV}$. The EOSD bunch length and the VLM energy-spread proxy still show strong correlation, while the THz

Table 7.12.: Full-range agreement between reconstructed and measured observables for a modified RF voltage of $V_{\mathrm{RF}} = 1\,170\,\mathrm{kV}$. The baseline reconstruction in the previous section used $V_{\mathrm{RF}} = 1\,300\,\mathrm{kV}$.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8740	0.8792	0.1525
VLM energy-spread proxy	0.8049	0.8254	0.1863
THz CSR	0.9036	0.8909	0.1596

CSR observable shows a lower correlation and a larger NRMSE. The mean EOSD measurement error over the full dataset is 0.1851.

Table 7.13.: Full-range agreement between reconstructed and measured observables for an increased RF voltage of $V_{\mathrm{RF}} = 1\,430\,\mathrm{kV}$. The baseline reconstruction in the previous section used $V_{\mathrm{RF}} = 1\,300\,\mathrm{kV}$.

Observable	Pearson $r(0)$	Spearman $\rho(0)$	NRMSE_{95-5}
EOSD bunch length	0.8673	0.8680	0.1562
VLM energy-spread proxy	0.8403	0.8569	0.1702
THz CSR	0.8032	0.6984	0.2220

The results of this section show that moderate deviations in the bunch current and RF voltage, on the order of 10%, still produce acceptable reconstruction behavior for the operating conditions considered here. This indicates a degree of robustness with respect to small parameter uncertainties, although this tolerance may vary across operating regimes, particularly when the dynamics change rapidly. Independent diagnostics therefore remain necessary to identify cases where the EOSD fit alone is not sufficient.

7.4. Discussion

The reconstructions with simulated Inovesa data establish the basis for interpreting the experimental results. They show that the algorithm can recover non-equilibrium phase space densities from EOSD signals when the VFPE provides a consistent description of the dynamics. They also show that moderate offsets in bunch current and RF voltage remain tolerable, whereas an inappropriate impedance model produces clearer deviations from the ground truth.

The experimental reconstruction therefore extends this validation to the case in which the true phase space density is not available. In this setting, the quality of the reconstruction cannot be assessed by a direct ground-truth comparison. Instead, the reconstructed phase space must satisfy two complementary requirements: it must reproduce the EOSD measurements used in the inverse problem, and it must remain consistent with independent diagnostics that were not included in the optimization objective. This consistency was evaluated using the zero-lag Pearson correlation $r(0)$, the Spearman rank correlation $\rho(0)$, and the robust NRMSE.

The results show that the reconstructed phase space densities reproduce the main temporal trends of the measured observables. The EOSD bunch length provides a direct data consistency check, while the VLM horizontal bunch size and the THz CSR signal provide independent validation of the reconstructed dynamics. The Pearson and Spearman coefficients quantify the linear and monotonic trend agreement, respectively, while the robust NRMSE measures the remaining discrepancy after removing a constant

gain and offset. In particular, the high agreement with the THz signal indicates that the reconstructed longitudinal structures capture relevant features of the CSR-emission dynamics. The VLM comparison is weaker, but still shows a consistent trend, which is reasonable because the horizontal bunch size is only an indirect proxy for the energy spread.

The sensitivity studies also show why independent validation is necessary. In several cases, the EOSD measurement error alone does not fully reflect the physical quality of the reconstruction. A modified impulse response or a current-free reconstruction can still reproduce parts of the EOSD signal while producing phase space structures whose physical interpretation is less reliable. These cases would be difficult to identify from the EOSD fit alone and could lead to false positive conclusions if the independent diagnostics were not considered.

The EOSD impulse response is therefore a critical component of the reconstruction framework. The tested variations show that changes in the laser-trajectory model and in the effective laser pulse length can modify both the EOSD fit and the derived physical observables. A more accurate, experimentally validated impulse-response model would therefore be expected to reduce systematic reconstruction artifacts, improve the agreement with independent diagnostics, and strengthen the physical interpretation of the reconstructed phase space density.

The parameter studies indicate that moderate variations of the bunch current and RF voltage still lead to acceptable reconstruction behavior. This suggests a degree of robustness with respect to small uncertainties in accelerator parameters. At the same time, the results show that the reconstruction remains sensitive to modeling choices that directly affect the longitudinal dynamics or the observation model. This reinforces the need to evaluate reconstruction quality using both data consistency and physical consistency.

Further improvements are still possible. The algorithmic implementation can be refined to improve convergence, reduce the EOSD measurement residual, and increase computational performance. Possible directions include improved stopping criteria, adjusted regularization, more robust initialization, and a more accurate treatment of the nonlinear dynamics during the reconstruction. Nevertheless, the experimental results are promising: the method reconstructs phase space densities that reproduce the measured EOSD data and show consistent trends with independent diagnostics, demonstrating the feasibility of VFPE-constrained phase space reconstruction from experimental EOSD measurements.

8. Conclusions

This thesis investigated whether the longitudinal phase space density of an electron bunch can be reconstructed in a non-equilibrium regime from turn-by-turn, one-dimensional measurements while accounting for the dynamics that deform the density during the measurement time interval. The central objective was to move beyond the rigid-rotation assumption to develop a reconstruction in which the measured sequence is connected to a single initial phase space density through the VFPE.

The results demonstrate that VFPE-constrained optimization can recover non-equilibrium longitudinal dynamics from EOSD measurements when the accelerator parameters and diagnostic models provide an adequate representation of the experimental setup. This is shown not only by the reproduction of the EOSD measurements, but also by the consistency of the reconstructed phase space with trends observed in synchronized diagnostic measurements that were not used as reconstruction inputs.

A surrogate observation model was developed to relate the EOSD measurements to the longitudinal bunch profile. Since the EOSD diagnostic does not provide an ideal charge profile, the electromagnetic propagation through the crystal and optical system was modeled as an effective impulse response. The convolution of this impulse response with the underlying bunch profile generates the measured phase-retardation signal, establishing the forward diagnostic model used for the reconstruction.

The reconstruction framework developed in this thesis is based on an operator-splitting formulation of the VFPE, in which the different physical contributions to the beam dynamics are treated separately. This separation provides a natural basis for expressing the forward evolution as a sequence of discrete propagation steps. By reformulating these steps as sparse matrix operations, the corresponding adjoint formulation can be expressed in matrix form, enabling a direct calculation of the gradient.

The matrix formulation provides an efficient representation of the reconstruction problem, whereas the complete beam dynamics remain nonlinear due to the dependence of the wake potential on the evolving phase space density. This dependence is handled by repeatedly solving a linearized least-squares problem with fixed wake potentials and then updating the collective dynamics from the newly reconstructed trajectory. In this way, the method preserves the essential nonlinear beam physics while allowing each reconstruction step to be solved with efficient iterative linear algebra methods.

The linear least-squares optimization is solved with an inner–outer CGLS algorithm, combined with regularization and non-negative projected restarted. This specific algorithmic combination proved highly effective for the reconstruction problem considered in this thesis. It achieved high accuracy for synthetic datasets and produced promising results when applied to experimental measurements.

The current implementation is used for diagnostic analysis, comparison of operating regimes, and model assessment. Within this scope, it provides a practical basis for fine-tuning heuristics, such as adjusting the impulse response, varying machine parameters within plausible ranges, selecting regularization strengths, and defining stopping criteria based on consistency with independent diagnostics.

Beyond reproducing the EOSD measurements, the reconstructed phase space density provides a direct connection to the beam dynamics during the measurement sequence. It can be propagated, analyzed, and compared across turns to study the evolution of longitudinal structures and their relation to

collective effects. In addition, quantities derived from the reconstructed phase space can provide access to beam observables that are not yet directly available experimentally.

The use of the presented phase space reconstruction algorithm for online enhanced diagnostics or feedback control would require further improvements in speed, calibration, robustness, and uncertainty quantification. Achieving the necessary speed requires accelerating the forward and adjoint evaluations to match operational time scales. Additionally, system calibration must be improved. Parameters such as the EOSD impulse response, timing alignment, and diagnostic gains must be known or updated dynamically with high accuracy. The algorithm must also demonstrate robustness, remaining stable under measurement noise, incomplete data, and sudden shifts in the accelerator regime. Finally, rigorous uncertainty quantification is essential to distinguish reliable reconstructed structures from artifacts caused by model mismatch, diagnostic errors, or regularization choices.

Ongoing development of Inovesalib focuses on improving its readiness for production use, including software stabilization, API consolidation, and broader testing of its implemented modular interfaces. The framework already provides modular connections between the VFPE solver and the software components used for the reconstruction in this thesis. Further development aims to prepare these research components for broader workflows in accelerator control, machine learning, and data science. In parallel, the implementation of full AD of the solver is being pursued to provide sensitivities with respect to machine and model parameters for extending the framework toward broader inverse problems and accelerator fine-tuning tasks. In addition, performance improvements through modern parallelization techniques and HPC-ready implementation are intended to support large sensitivity studies, scalable reconstruction of extensive datasets, and efficient integration into digital-twin initiatives.

Future work should focus first on establishing an experimental validation of the simulated EOSD impulse response used in the phase space reconstruction. This should be complemented by new synchronized measurement campaigns that incorporate as many of the available diagnostics and sensors at KARA as possible, providing a stronger experimental basis for improving calibration and robustness. In-depth uncertainty quantification studies are also needed, both for the current experimental dataset and for future datasets, in order to assess the sensitivity of the reconstruction to diagnostic errors, model mismatch, and regularization choices. Finally, the efficient implementation of the algorithm on dedicated accelerator hardware should be investigated as an important step toward online enhanced diagnostics and feedback applications.

This thesis demonstrates that VFPE-constrained reconstruction can transform turn-by-turn EOSD measurements into a physically interpretable estimate of the longitudinal phase space density in a non-equilibrium bunch regime. By combining beam dynamics, a realistic EOSD observation model, sparse adjoint-based optimization, and validation with independent diagnostics, the work establishes a methodological basis for using phase space reconstruction as an advanced diagnostic tool for longitudinal beam dynamics at KARA.

A. Appendix

A.1. Phase retardation algorithm calculation example

To further clarify the phase retardation algorithm calculation, three representative situations are considered.

In the following cases, the index i labels the simulation frame used to compute a candidate laser position. If the candidate position satisfies the corresponding boundary condition, it is retained and relabeled with the valid trajectory index j. The accumulation step is therefore written in terms of z_j and t_j.

Case 1: At the instant of reflection:

- **Simulation Time & Slice:** The algorithm evaluates the interaction exactly at the simulation step $t_i = t_{\mathrm{head}}$.
- **Trajectory Calculation:** The kinematic equations, Eq. (3.12) and Eq. (3.11), evaluate to:

$$\begin{aligned} z_i^{\mathrm{down}} &= z_{\mathrm{start}} + v_{\mathrm{L}}\big[-\tau_m\big], \\ z_i^{\mathrm{up}} &= z_{\mathrm{start}} + v_{\mathrm{L}}\big[\tau_m\big]. \end{aligned}$$

  For $\tau_0 = 0$, the position of the laser head is

$$\begin{aligned} z_i^{\mathrm{down}} &= z_{\mathrm{start}}, \\ z_i^{\mathrm{up}} &= z_{\mathrm{start}}. \end{aligned}$$

- **Boundary Validation:** For $\tau_0 = 0$, the downstream and upstream coordinates both equal z_{start}. Because Eq. (3.14) requires the upstream position to be strictly greater than the entrance edge, this upstream point is rejected. This prevents the field at the reflection coordinate from being summed twice. For any other laser point $\tau_m > 0$, there are no valid points in the downstream trajectory; all such points lie on the upstream trajectory and must be checked against the crystal boundaries.
- **Field Interpolation & Accumulation:** The algorithm identifies the nearest discrete probes bounding the valid position, for example $p_0 \leq z_j < p_1$. It calculates the local electric field $E_y(z_j, t_j)$ by linear interpolation and accumulates this value into the corresponding downstream or upstream phase retardation sum.

Case 2: Before reflection:

- **Simulation Time & Slice:** The algorithm evaluates an early simulation step t_i, before the laser head has reached the reflective edge, such that $0 \leq t_i < t_{\mathrm{head}}$.
- **Trajectory Calculation:** The kinematic equations evaluate to:

$$\begin{aligned} z_i^{\mathrm{down}} &= z_{\mathrm{start}} + v_{\mathrm{L}}\big[(t_i - t_{\mathrm{head}}) - \tau_m\big], \\ z_i^{\mathrm{up}} &= z_{\mathrm{start}} + v_{\mathrm{L}}\big[(t_{\mathrm{head}} - t_i) + \tau_m\big]. \end{aligned}$$

- **Boundary Validation:** The downstream time is always negative, yielding $z_i^{\mathrm{down}} < z_{\mathrm{start}}$. Therefore, there are no valid laser points on the downstream path, and this contribution is rejected. The upstream time is positive, yielding $z_i^{\mathrm{up}} > z_{\mathrm{start}}$, so the laser point lies on the upstream trajectory. Each point must still be checked to ensure that it lies within the crystal boundaries.
- **Field Interpolation & Accumulation:** For each valid coordinate z_j^{up}, the algorithm locates the corresponding bounding probes, $p_k \leq z_j^{\mathrm{up}} < p_{k+1}$, interpolates the electric field $E_y(z_j^{\mathrm{up}}, t_j)$, and adds this result to the upstream phase retardation sum for the slice τ_m. However, if the laser head co-propagates with the electric field of the bunch inside the crystal, these values might be zero. This formulation is flexible enough to model scenarios where the laser pulse and the bunch profile have an offset, and the modulation starts in the upstream trajectory.

Case 3: After reflection:

- **Simulation Time & Slice:** The algorithm evaluates a later simulation step t_i, after the laser head has reached the reflective edge, such that $t_i > t_{\mathrm{head}}$.
- **Trajectory Calculation:** The kinematic equations evaluate to:

$$\begin{aligned} z_i^{\mathrm{down}} &= z_{\mathrm{start}} + v_{\mathrm{L}}\big[t_i - (t_{\mathrm{head}} + \tau_m)\big], \\ z_i^{\mathrm{up}} &= z_{\mathrm{start}} + v_{\mathrm{L}}\big[(t_{\mathrm{head}} + \tau_m) - t_i\big]. \end{aligned}$$

  Depending on the value of τ_m, the laser point may contribute to either the downstream or upstream trajectory.
- **Boundary Validation:** To determine whether a laser point belongs to the downstream or upstream trajectory at a given t_i, the following conditions are evaluated. A point belongs to the downstream trajectory if

$$t_i \geq t_{\mathrm{head}} + \tau_m. \tag{A.1}$$

  A point belongs to the upstream trajectory if

$$t_i < t_{\mathrm{head}} + \tau_m. \tag{A.2}$$

  The resulting coordinate is then checked against the corresponding crystal-boundary condition.
- **Field Interpolation & Accumulation:** The algorithm dynamically identifies the closest probes required to evaluate $E_y(z_j, t_j)$ and accumulates the corresponding valid contribution into the downstream or upstream phase retardation sum.

Finally, the spatial distance over which the calculated field interacts during a single frame depends on the simulation time step:

$$\Delta z_j = v_{\mathrm{L}} \Delta t. \tag{A.3}$$

Although Δt is usually constant, writing the interaction length as Δz_j keeps the notation flexible enough to support non-uniform simulation steps without changing the core equations.

A particularly relevant special case arises when the spatial discretization of the laser pulse matches the probe spacing in the crystal, and the temporal update of the electromagnetic solver is chosen accordingly. In this configuration, the simulation time step is

$$\begin{aligned} \Delta t &= \frac{\Delta d}{v_L}, \\ t_i &= t_{\mathrm{head}} + i\frac{\Delta d}{v_L}, \quad \text{for } i = 0, \ldots, N_{\max} - 1, \end{aligned} \tag{A.4}$$

where Δd denotes the probe spacing, and $N_{\max}$ is the total number of simulation time steps required for the laser pulse to fully traverse the crystal.

To achieve this alignment, the laser pulse is discretized with the same temporal resolution,

$$\begin{aligned} \Delta\tau &= \frac{\Delta d}{v_L}, \\ \tau_m &= m\frac{\Delta d}{v_L}, \quad \text{for } m = 0, \ldots, M-1, \end{aligned} \tag{A.5}$$

with $\tau_0 = 0$ and M the number of laser sampling points. In general, $M \leq N_{\max}$, since the laser pulse duration is shorter than the total simulation time.

Under these conditions, the laser positions along the trajectory reduce to simple integer multiples of the probe spacing. For the downstream path,

$$\begin{aligned} z_i^{\text{down}} &= z_{\text{start}} + v_L\big[(i-m)\frac{\Delta d}{v_L} + t_{\text{head}} - t_{\text{head}}\big], \\ z_i^{\text{down}} &= z_{\text{start}} + (i-m)\Delta d. \end{aligned} \tag{A.6}$$

Conversely, the upstream positions (counter-propagating toward the mirror) evaluate to:

$$\begin{aligned} z_i^{\text{up}} &= z_{\text{start}} + v_L\big[t_{\text{head}} - t_{\text{head}} + (m-i)\frac{\Delta d}{v_L}\big], \\ z_i^{\text{up}} &= z_{\text{start}} + (m-i)\Delta d. \end{aligned} \tag{A.7}$$

The valid trajectory points are determined by both the propagation direction and the crystal boundaries. By identifying the trajectory index with the probe index j, the downstream contribution corresponds to

$$j = i - m, \qquad 0 \leq j \leq N-1, \tag{A.8}$$

where N denotes the total number of probe positions along the crystal. Solving this inequality for the simulation frame i yields the downstream temporal bounds: $m \leq i \leq m + N - 1$.

For the upstream contribution,

$$j = m - i, \qquad 1 \leq j \leq N-1, \tag{A.9}$$

where the lower bound excludes $j = 0$ to avoid double counting at the reflection point. Solving this inequality establishes the temporal bounds for the upstream trajectory as $m - N + 1 \leq i \leq m - 1$.

These constraints ensure that only trajectory points lying within the crystal and associated with valid probe indices contribute to the phase retardation sum.

In this special case, the spatial increment along the trajectory is constant,

$$\Delta z^{\text{down}} = \Delta d, \qquad \Delta z^{\text{up}} = -\Delta d, \tag{A.10}$$

so that the effective interaction length is

$$\|\Delta z_i\| = \Delta d. \tag{A.11}$$

The key advantage of this configuration is that the laser sampling points coincide exactly with the probe positions. As a result, the electric field is evaluated directly at the probes, ${E_y}^{(j)}$, and no interpolation is required.

The explicit form of the phase retardation sum depends intrinsically on the relationship between the number of laser samples M, the discrete probes N, and the total simulation steps $N_{\max}$. The valid index ranges for the upstream and downstream contributions must strictly satisfy both the spatial probe bounds, $0 \leq j \leq N-1$, and the temporal solver bounds, $0 \leq i \leq N_{\max}-1$. Specifically, for the final slice of the pulse, $m = M-1$, to complete its downstream journey, the simulation must satisfy $N_{\max} \geq M+N-1$ to avoid truncating the physical data. In the general case, however, not every laser sample has access to the full set of N probe positions in both directions; the upper limits must therefore be restricted by the smaller of the spatial and temporal bounds. The generic phase retardation relation associated with the two trajectories is

$$\Gamma_m = \frac{2\pi}{\lambda} n_0^3 r_{41}\, \Delta d \left(\underbrace{\sum_{j=0}^{\min(N-1,\, N_{\max}-1-m)} E_y^{(j)}(t_{m+j})}_{\text{downstream}} + \underbrace{\sum_{\substack{j=1 \\ m>0}}^{\min(N-1,\, m)} E_y^{(j)}(t_{m-j})}_{\text{upstream}} \right). \tag{A.12}$$

A simpler expression is obtained in the restricted case $M = N$ and $N_{\max} = 2N$. In this configuration, the laser sampling, probe spacing, and simulation time step are aligned, and the simulation window is long enough for all laser samples to complete the downstream traversal. Therefore, for $m = 0, \ldots, N-1$, Eq. (A.12) reduces to

$$\Gamma_m = \frac{2\pi}{\lambda} n_0^3 r_{41}\, \Delta d \left(\underbrace{\sum_{j=0}^{N-1} E_y^{(j)}(t_{m+j})}_{\text{downstream}} + \underbrace{\sum_{\substack{j=1 \\ m>0}}^{m} E_y^{(j)}(t_{m-j})}_{\text{upstream}} \right). \tag{A.13}$$

Equivalently, written directly in terms of the simulation-time index i, the same expression becomes

$$\Gamma_m = \frac{2\pi}{\lambda} n_0^3 r_{41}\, \Delta d \left(\underbrace{\sum_{i=m}^{m+N-1} E_y^{(i-m)}(t_i)}_{\text{downstream}} + \underbrace{\sum_{\substack{i=0 \\ m>0}}^{m-1} E_y^{(m-i)}(t_i)}_{\text{upstream}} \right). \tag{A.14}$$

The explicit condition $m > 0$ in the upstream term ensures that the laser head is never counted in the upstream contribution. Consequently, the upstream sum is empty for $m = 0$, while for later laser samples it contains m valid contributions. The downstream term always contains N contributions, corresponding to the full traversal of the crystal after reflection.

A.2. Derivation of the evolved quadrilateral area

As illustrated in Figure A.1, this area can be calculated as the sum of two triangular sub-regions Δ_1 and Δ_2 using the vector cross product:

$$\begin{aligned}\mathcal{A}_{t+\Delta t} &= \Delta_1 + \Delta_2 \\ &= \frac{\left|\overrightarrow{\mathcal{Q}_4\mathcal{Q}_1} \times \overrightarrow{\mathcal{Q}_4\mathcal{Q}_3}\right|}{2} + \frac{\left|\overrightarrow{\mathcal{Q}_2\mathcal{Q}_1} \times \overrightarrow{\mathcal{Q}_2\mathcal{Q}_3}\right|}{2} \\ &= \frac{|(\mathcal{Q}_1 - \mathcal{Q}_4) \times (\mathcal{Q}_3 - \mathcal{Q}_4)|}{2} + \frac{|(\mathcal{Q}_1 - \mathcal{Q}_2) \times (\mathcal{Q}_3 - \mathcal{Q}_2)|}{2}.\end{aligned} \tag{A.15}$$

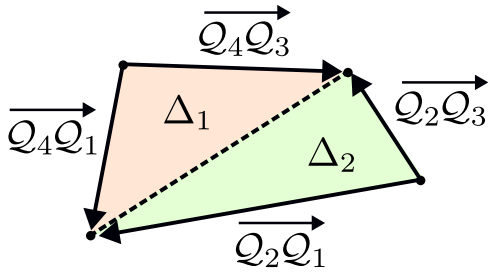


Figure A.1.: Area of a generic quadrilateral as $\mathcal{A} = \Delta_1 + \Delta_2$.

The vertices of the initial rectangle at time t are defined as:

$$\begin{aligned}\mathcal{O}_1 &= \begin{pmatrix} q \\ p \end{pmatrix}, \quad \mathcal{O}_2 = \begin{pmatrix} q + dq \\ p \end{pmatrix}, \\ \mathcal{O}_3 &= \begin{pmatrix} q + dq \\ p + dp \end{pmatrix}, \quad \mathcal{O}_4 = \begin{pmatrix} q \\ p + dp \end{pmatrix}.\end{aligned} \tag{A.16}$$

Applying the flow map, the new vertices $\mathcal{Q}_i$ are obtained.

$$\begin{aligned}\mathcal{Q}_1 &= \begin{pmatrix} q + f^q(q,p,t)\Delta t \\ p + f^p(q,p,t)\Delta t \end{pmatrix}, \\ \mathcal{Q}_2 &= \begin{pmatrix} q + dq + f^q(q+dq,p,t)\Delta t \\ p + f^p(q+dq,p,t)\Delta t) \end{pmatrix}, \\ \mathcal{Q}_3 &= \begin{pmatrix} q + dq + f^q(q+dq,p+dp,t)\Delta t \\ p + dp + f^p(q+dq,p+dp,t)\Delta t \end{pmatrix}, \\ \mathcal{Q}_4 &= \begin{pmatrix} q + f^q(q,p+dp,t)\Delta t \\ p + dp + f^p(q,p+dp,t)\Delta t \end{pmatrix}.\end{aligned} \tag{A.17}$$

Assuming small spatial variations in velocity across the element, the flow field at the displaced corners is approximated using a first-order Taylor series expansion:

$$\mathbf{f}(q+\Delta q, p+\Delta p, t) \approx \mathbf{f}(q,p,t) + \frac{\partial \mathbf{f}}{\partial q}\Delta q + \frac{\partial \mathbf{f}}{\partial p}\Delta p. \tag{A.18}$$

Using the Taylor expansion A.18 on $f = (f^q, f^p)$, the vertices become:

$$\begin{aligned}
\mathcal{Q}_1 &= \begin{pmatrix} q + f^q(q,p,t)\Delta t \\ p + f^p(q,p,t)\Delta t \end{pmatrix}, \\
\mathcal{Q}_2 &= \begin{pmatrix} q + dq + f^q(q,p,t)\Delta t + \frac{\partial f^q}{\partial q}dq\Delta t \\ p + f^p(q,p,t)\Delta t) + \frac{\partial f^p}{\partial p}dp\Delta t \end{pmatrix}, \\
\mathcal{Q}_3 &= \begin{pmatrix} q + dq + f^q(q,p,t)\Delta t + \frac{\partial f^q}{\partial q}dq\Delta t + \frac{\partial f^q}{\partial p}dp\Delta t \\ p + dp + f^p(q,p,t)\Delta t + \frac{\partial f^p}{\partial q}dq\Delta t + \frac{\partial f^p}{\partial p}dp\Delta t \end{pmatrix}, \\
\mathcal{Q}_4 &= \begin{pmatrix} q + f^q(q,p,t)\Delta t + \frac{\partial f^q}{\partial q}dq\Delta t \\ p + dp + f^p(q,p,t)\Delta t + \frac{\partial f^p}{\partial p}dp\Delta t \end{pmatrix}.
\end{aligned} \tag{A.19}$$

From these coordinates, the edge vectors required for the area calculation are derived:

$$\begin{aligned}
\mathcal{Q}_1 - \mathcal{Q}_4 &= \begin{pmatrix} -\frac{\partial f^q}{\partial p}dp\Delta t \\ -dp - \frac{\partial f^p}{\partial p}dp\Delta t \end{pmatrix}, \\
\mathcal{Q}_3 - \mathcal{Q}_4 &= \begin{pmatrix} dq + \frac{\partial f^q}{\partial q}dq\Delta t \\ \frac{\partial f^p}{\partial q}dq\Delta t \end{pmatrix}, \\
\mathcal{Q}_1 - \mathcal{Q}_2 &= \begin{pmatrix} -dq - \frac{\partial f^q}{\partial q}dq\Delta t \\ -\frac{\partial f^p}{\partial q}dq\Delta t \end{pmatrix}, \\
\mathcal{Q}_3 - \mathcal{Q}_2 &= \begin{pmatrix} \frac{\partial f^q}{\partial p}dp\Delta t \\ dp + \frac{\partial f^p}{\partial p}dp\Delta t \end{pmatrix}.
\end{aligned} \tag{A.20}$$

Substituting these vectors into Eq. (A.15) , the area of the first triangle is found to be:

$$\Delta_1 = \frac{dpdq}{2}\left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t + \left(\frac{\partial f^q}{\partial q}\frac{\partial f^p}{\partial p} - \frac{\partial f^q}{\partial p}\frac{\partial f^p}{\partial q}\right)\Delta t^2\right] \tag{A.21}$$

Similarly, for the second triangle formed by $\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3$:

$$\Delta_2 = \frac{dpdq}{2}\left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t + \left(\frac{\partial f^q}{\partial q}\frac{\partial f^p}{\partial p} - \frac{\partial f^p}{\partial q}\frac{\partial f^q}{\partial p}\right)\Delta t^2\right] \tag{A.22}$$

Since the contributions from both triangles are identical, the total area of the deformed quadrilateral is given by the sum:

$$\mathcal{A}_{t+\Delta t} = dqdp\left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t + \left(\frac{\partial f^q}{\partial q}\frac{\partial f^p}{\partial p} - \frac{\partial f^q}{\partial p}\frac{\partial f^p}{\partial q}\right)\Delta t^2\right] \tag{A.23}$$

This relation shows that the initially rectangular element is mapped, to first order, into a parallelogram.

To determine the instantaneous evolution, the limit $\Delta t \to 0$ is considered. In this regime, terms of order Δt^2 vanish strictly faster than linear terms. Consequently, the area evolution is governed solely by the first-order expansion:

$$\mathcal{A}_{t+\Delta t} \approx dqdp \left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right) \Delta t\right] \tag{A.24}$$

A.3. Intermediate steps in the derivation of the Vlasov equation

Substituting the evolved area Eq. (2.12) into the conservation law Eq. (2.11) yields:

$$\mathrm{d}q\mathrm{d}p\, \psi(q,p,t) = \mathrm{d}q\mathrm{d}p\, \psi(q+\Delta q, p+\Delta p, t+\Delta t) \left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right) \Delta t\right]. \tag{A.25}$$

Eliminating the common factor $\mathrm{d}q\mathrm{d}p$, the relation simplifies to:

$$\psi(q,p,t) = \psi(q+\Delta q, p+\Delta p, t+\Delta t) \left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right) \Delta t\right]. \tag{A.26}$$

The phase space density at time $t + \Delta t$ is then expanded using a Taylor series around (q, p, t):

$$\psi(q+\Delta q, p+\Delta p, t+\Delta t) \approx \psi(q,p,t) + \frac{\partial \psi}{\partial q} f^q(q,p,t)\Delta t + \frac{\partial \psi}{\partial p} f^p(q,p,t)\Delta t + \frac{\partial \psi}{\partial t}\Delta t. \tag{A.27}$$

Substituting Eq. (A.27) into Eq. (A.26) results in:

$$\psi(q,p,t) = \left[\psi(q,p,t) + \left(\frac{\partial \psi}{\partial t} + f^q(q,p,t)\frac{\partial \psi}{\partial q} + f^p(q,p,t)\frac{\partial \psi}{\partial p}\right)\Delta t\right] \left[1 + \left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t\right] \tag{A.28}$$

Expanding the product and grouping terms by powers of Δt:

$$\begin{aligned}\psi(q,p,t) = \psi(q,p,t) &+ \psi(q,p,t)\left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t \\ &+ \left(\frac{\partial \psi}{\partial t} + f^q(q,p,t)\frac{\partial \psi}{\partial q} + f^p(q,p,t)\frac{\partial \psi}{\partial p}\right)\Delta t \\ &+ \left(\frac{\partial \psi}{\partial t} + f^q(q,p,t)\frac{\partial \psi}{\partial q} + f^p(q,p,t)\frac{\partial \psi}{\partial p}\right)\left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p}\right)\Delta t^2\end{aligned} \tag{A.29}$$

Subtracting ψ from both sides and grouping the terms by powers of Δt:

$$0 = \left[\frac{\partial \psi}{\partial t} + f^q(q,p,t)\frac{\partial \psi}{\partial q} + f^p(q,p,t)\frac{\partial \psi}{\partial p} + \psi(q,p,t)\left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p} \right) \right] \Delta t + O(\Delta t^2) \tag{A.30}$$

Dividing by Δt and taking the limit as $\Delta t \to 0$, the second-order terms vanish, leaving the general continuity equation:

$$\frac{\partial \psi}{\partial t} + f^q(q,p,t)\frac{\partial \psi}{\partial q} + f^p(q,p,t)\frac{\partial \psi}{\partial p} = -\psi(q,p,t)\left(\frac{\partial f^q}{\partial q} + \frac{\partial f^p}{\partial p} \right) \tag{A.31}$$

A.4. Approximate second-order methods

Between pure first-order methods and full second-order methods lies an intermediate family known as Quasi-Newton methods, which reconstruct curvature information approximately from gradient differences rather than from explicit Hessians. These methods are especially attractive in reduced-space inverse problems because they preserve the gradient-only interface of the adjoint method while capturing substantially more curvature information than descent methods [81, 97]. Their central idea is to replace the exact inverse Hessian by an approximation H_k, and to define the search direction as

$$p_k = -H_k g_k. \tag{A.32}$$

A.4.1. The BFGS method

The canonical representative of this family is the Broyden–Fletcher–Goldfarb–Shanno method (BFGS) method, where H_k is a dense matrix approximating the inverse Hessian.

The canonical representative is BFGS. With

$$s_k := u_{k+1} - u_k, \qquad y_k := g_{k+1} - g_k, \qquad \rho_k := \frac{1}{y_k^\top s_k}, \tag{A.33}$$

the inverse-Hessian approximation is updated by

$$H_{k+1} = \left(I - \rho_k s_k y_k^\top \right) H_k \left(I - \rho_k y_k s_k^\top \right) + \rho_k s_k s_k^\top. \tag{A.34}$$

Provided the curvature condition $y_k^\top s_k > 0$ holds, this rank-two update mathematically guarantees that H_{k+1} remains symmetric and positive definite, ensuring that p_k is always a valid descent direction.

A.4.2. Symmetric Rank-One (SR1)

A second classical representative is the symmetric rank-one method (SR1) method (SR1),

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^\top}{(s_k - H_k y_k)^\top y_k}, \tag{A.35}$$

Unlike BFGS, the SR1 update does not automatically force the approximate Hessian to remain positive definite. While this requires more delicate trust-region handling, it allows the SR1 method to map indefinite curvature much more faithfully, making it highly useful when navigating highly non-convex regions of the reduced objective landscape.

A.4.3. Limited-memory BFGS (L-BFGS)

For the present high-dimension reconstruction problem, allocating memory for the dense H_k matrix is computationally prohibitive. The Limited-memory BFGS (limited-memory Broyden–Fletcher–Goldfarb–Shanno method (L-BFGS)) algorithm elegantly circumvents this constraint. Instead of storing a monolithic matrix, L-BFGS securely stores only a small, rolling history of the n_{mem} most recent curvature pairs:

$$\{(s_{k-j}, y_{k-j})\}_{j=0}^{n_{mem}-1}. \tag{A.36}$$

The algorithm computes the necessary matrix-vector product $H_k g_k$ implicitly through a highly efficient recursive two-loop algorithm [81, 97]. Consequently, L-BFGS serves as the ultimate bridge between first-order memory limits and second-order acceleration in high-dimensional tomography.

A.5. Second-order methods

Second-order methods deviate from the previous families by utilizing the explicit Hessian $\nabla^2 J(u_k)$, or its exact action via matrix-vector products, to formulate the search direction. In the context of large-scale inverse problems, this family provides the fastest local convergence rates and the most accurate topological map of the objective landscape. However, because these methods demand additional inner systems to be solved at every outer optimization step, they carry the highest per-iteration analytical and computational costs [98, 99].

A.5.1. Exact Newton method

The theoretical baseline for this family is the exact Newton step, obtained by minimizing the local quadratic Taylor approximation of the reduced cost function around the current iterate u_k. Writing

$$J(u_k + p) \approx J(u_k) + g_k^\top p + \frac{1}{2} p^\top \nabla^2 J(u_k)\, p,$$

with $g_k := \nabla J(u_k)$, the stationarity condition with respect to the step p yields the linear system

$$\nabla^2 J(u_k)\, p_k = -g_k. \tag{A.37}$$

When the Hessian is available and sufficiently regular, this step provides the classical local quadratic convergence of Newton's method. However, in the present reconstruction problem, deriving and assembling the full reduced Hessian, or the corresponding second-order adjoint model, is highly restrictive because of the state-dependent dynamics. For this reason, pure Newton steps are generally impractical in the present setting.

A.5.2. Gauss–Newton method

When the inverse problem is framed as a nonlinear least-squares minimization, defined by $J(u) = \frac{1}{2}\|r(u)\|^2$ with residual vector $r(u)$ and Jacobian $J_r(u)$, the exact Hessian decomposes precisely as:

$$\nabla^2 J(u) = J_r(u)^\top J_r(u) + \sum_{i=1}^{m} r_i(u)\, \nabla^2 r_i(u). \tag{A.38}$$

The Gauss–Newton method assumes that near the solution, the residuals $r_i(u)$ become small enough that the second-order residual term can be safely neglected. The algorithm instead solves the highly structured system:

$$\Big(J_r(u_k)^\top J_r(u_k)\Big)\, p_k = -J_r(u_k)^\top r(u_k). \tag{A.39}$$

This approximation captures the dominant curvature perfectly using only first-order Jacobian data, making it particularly natural for the linearized inner-loop iterations discussed earlier.

A.5.3. Matrix-free Newton–Krylov methods

To avoid explicit Hessian assembly entirely, the exact Newton system in Eq. (A.37) can be solved approximately using an inner Krylov subspace method, such as linear Conjugate Gradient. These Newton–CG methods rely strictly on Hessian-vector products:

$$p_k \approx \arg\min_{p \in \mathcal{K}_\ell} \left\{ \tfrac{1}{2} p^\top \nabla^2 J(u_k) p + g_k^\top p \right\}, \tag{A.40}$$

where $\mathcal{K}_\ell$ is the Krylov subspace generated by the gradients. When an inverse problem exhibits ill-posedness, solving this inner Krylov iteration inexactly (truncated Newton) serves as a potent implicit regularizing mechanism [98].

Acronyms

AD	algorithmic differentiation 91, 92, 94, 95, 134
API	application programming interface 91, 92, 134
ART	algebraic reconstruction technique 2
BFGS	Broyden–Fletcher–Goldfarb–Shanno method 142, 143
BLonD	beam longitudinal dynamics 2, 3
CAD	computer-aided design 37
CGLS	conjugate gradient for least squares 86–89, 95, 98, 102–111, 116, 133
CPU	central processing unit 48
CSR	coherent synchrotron radiation 1, 25, 31–35, 56, 91, 98–100, 111–113, 115–117, 123–127, 129–132
CST	Computer Simulation Technology 37–42, 45–47, 50–52
EO	electro-optical 35–38, 126
EOS	electro-optical sampling 35
EOSD	electro-optical spectral decoding 2–4, 31, 32, 34, 35, 37–41, 43, 45–52, 69, 70, 97–103, 111, 114, 116–134
FBP	filtered backprojection 2, 69, 101, 102
FBSF	filament beam spread function 117
FIT	finite integration technique 37
FP	Fokker–Planck 9, 10, 13, 29, 53–55, 58, 64–66, 92
FWHM	full width at half maximum 123
GaP	Gallium Phosphide 35, 46
GD	gradient descent 81, 82, 103
GMRES	generalized minimal residual method 86
HDF5	Hierarchical Data Format version 5
HPC	high-performance computing 48, 134
IBS	intra-beam scattering 91
Inovesa	Inovesa Numerical Optimized Vlasov Equation Solver Application 39, 41, 55–57, 59–65, 76, 79, 91–94, 97–100, 107, 111, 117, 120, 131
IR2	Infrared 2 beamline 33
KALYPSO	Karlsruhe line array detector for MHz-repetition rate spectroscopy 32, 35, 36
KAPTURE	Karlsruhe Pulse Taking Ultra-fast Readout Electronics 34
KARA	Karlsruhe Research Accelerator 1–4, 31–35, 38, 41, 91, 97, 98, 116, 119, 134

KIT	Karlsruhe Institute of Technology 56
L-BFGS	limited-memory Broyden–Fletcher–Goldfarb–Shanno method 143
L-BFGS-B	bound-constrained limited-memory Broyden–Fletcher–Goldfarb–Shanno method
LSQR	least-squares QR algorithm 86
MAX IV	MAX IV Laboratory 91
NRMSE	normalized root-mean-square error 124, 130, 131
PDE	partial differential equation 4, 69, 72
RF	radio frequency 1, 2, 5, 13, 15–22, 24, 26, 27, 53, 60–63, 92, 99, 114, 115, 117, 128–132
RMS	root mean square 27, 51, 123
SD	steepest descent 74, 103
SOLEIL	Source Optimisée de Lumière d'Énergie Intermédiaire du LURE 91
SR	synchrotron radiation 1, 5, 17, 18, 21, 32
SR1	symmetric rank-one method 142
SVD	singular value decomposition 86
THz	terahertz 4, 31–34, 43, 91, 98, 116–120, 123–127, 129–132
TNB	tangent–normal–binormal 13, 14
VFP	Vlasov–Fokker–Planck 56, 91, 95
VFPE	Vlasov–Fokker–Planck equation 3–5, 12, 13, 26, 28, 29, 53, 64, 70, 73, 76, 79, 89, 91–95, 97–99, 103, 105, 108, 115, 131–134
VLD	visible-light diagnostics 32, 33
VLM	visible-light monitor 4, 31–33, 98, 116–118, 120, 124–127, 129–132

Symbols

Beam dynamics

Sign	Description	Unit
$\mathbf{B}$	Magnetic field	T
c	Speed of light in vacuum	$\mathrm{m\,s^{-1}}$
e	Elementary particle charge magnitude	C
E_0	Reference particle energy	eV
$\mathbf{E}$	Electric field	$\mathrm{V\,m^{-1}}$
$\mathbf{F}$	Lorentz force	N
f_{rev}	Revolution frequency	Hz
f_s	Synchrotron frequency	Hz
$f_{s,0}$	Reference synchrotron frequency	Hz
h	RF harmonic number	
L_0	Design orbit length	m
m_0	Rest mass of the particle	kg
N_s	Number of turns per synchrotron oscillation	turns
p	Normalized longitudinal energy-deviation coordinate	
q	Normalized longitudinal position coordinate	
s	Path length along the reference orbit	m
t	Continuous laboratory frame time	s
T_0	Reference particle revolution period	s
t_n	Particle arrival time at the RF cavity on turn n	s
U_0	Mean energy loss per turn	eV
v	Particle speed	$\mathrm{m\,s^{-1}}$
$\hat{V}$	Effective linear RF voltage slope	V
v_{L}	Laser propagation velocity in the electro-optic crystal	$\mathrm{m\,s^{-1}}$
V_{RF}	RF voltage amplitude	V
$\mathbf{v}$	Particle velocity vector	$\mathrm{m\,s^{-1}}$
v_z	Longitudinal particle velocity component	$\mathrm{m\,s^{-1}}$
W	Longitudinal wake function	$\mathrm{V\,C^{-1}}$
x	Horizontal position	m
y	Vertical position	m
z	Longitudinal position	m
$Z_{\parallel}$	Longitudinal coupling impedance	Ω
α_0	Zeroth order momentum compaction coefficient	
α_1	First order momentum compaction coefficient	
α_2	Second order momentum compaction coefficient	
α_{c}	Momentum compaction factor	
β	Relativistic velocity factor v/c	
β_x	Horizontal beta function	m

Sign	Description	Unit
ε	Energy deviation	eV
$\dot{\varepsilon}$	Time derivative of the energy deviation	eV s^{-1}
ε_x	Horizontal emittance	m
η	Phase-slip factor	
γ	Relativistic Lorentz factor	
κ	Curvature of the reference orbit	m^{-1}
ω	Angular frequency	rad s^{-1}
ω_{RF}	RF angular frequency	rad s^{-1}
Ω_s	Synchrotron angular frequency	rad s^{-1}
ϕ_s	Synchronous phase	rad
ρ	Longitudinal bunch profile	
ρ_0	Reference bending radius	m
σ	RMS width (standard deviation)	
$\sigma_{E,0}$	Zero-current RMS energy spread	eV
σ_x	Horizontal beam size	m
σ_z	Longitudinal bunch length	m
τ	Geometric torsion of the reference orbit	m^{-1}
τ_d	Longitudinal damping time	s
θ	Normalized synchrotron angular time	rad
θ_n	Normalized synchrotron angular time at measurement index n	rad
φ	RF phase coordinate	rad
$\dot{\varphi}$	Time derivative of the RF phase coordinate	rad s^{-1}

Vlasov–Fokker–Planck equation

Sign	Description	Unit
D	Diffusion coefficient	
$f^{\Delta t}$	Finite-time flow map over the time interval Δt	
f^p	Hamiltonian flow component in the p direction	s^{-1}
F	Kramers–Moyal product integrand	
f^q	Hamiltonian flow component in the q direction	s^{-1}
H	Longitudinal Hamiltonian	eV
$\mathcal{M}_i$	Indexed region of the dynamical system state space	
$\mathcal{M}$	Dynamical system state space	
P	Transition probability density	
$\mathbf{x}$	Point in the dynamical system state space	
β_d	Longitudinal damping coefficient	s^{-1}
ψ	Phase space density	normalized
ψ°	Initial phase space density	normalized

EOSD and beam diagnostics

Sign	Description	Unit
d	Electro-optic crystal thickness	m

Sign	Description	Unit
$\Delta\tau$	Temporal spacing between laser sampling points	s
E_y	Vertical electric-field component	$\mathrm{V\,m^{-1}}$
G	EOSD forward observation operator	
h	EOSD impulse response	
I_M	Measured optical intensity	
I_U	Reference optical intensity	
m	Number of measurements index	
M	Number of laser sampling points	
n	Discrete turn index	
n_0	Unperturbed crystal refractive index	
$N_{\max}$	Maximum number of time samples	
N	Number of probe positions along the electro-optic crystal	
N_{probes}	Number of electro-optic simulation probes	
r_{41}	Pockels electro-optic coefficient	$\mathrm{m\,V^{-1}}$
τ	Temporal coordinate within the chirped laser pulse	s
τ_m	Temporal coordinate of laser sampling point m	s
t_{head}	Time at which the laser pulse head reaches the reflective coating	s
t_i	Simulation time at frame i	s
$t(z;\tau)$	Laser/crystal interaction time at position z for laser coordinate τ	s
t_j	Simulation time of the retained trajectory point j	s
t	Simulation time	s
Γ	Electro-optic phase retardation signal	rad
$\hat{\Gamma}$	Measured EOSD signal	
λ	Optical wavelength	m

Numerical grid and transport operators

Sign	Description	Unit
x	Longitudinal-position grid-node index	
y	Momentum grid-node index	
x_r	Continuous longitudinal-position grid coordinate	
y_r	Continuous momentum grid coordinate	
n	Flattened phase space grid index	
N	Number of phase space grid points per coordinate	
Δq	Longitudinal-position grid spacing	
Δp	Momentum grid spacing	
Δt	Physical time step	s
$\Delta\theta$	Normalized time step	
L_i	Lagrange interpolation basis function	
χ	Finite-difference coefficient	
J	Jacobian matrix of the phase space map	
$\mathcal{P}_{\Delta t}$	Hamiltonian density-transport operator over the time interval Δt	

Sign	Description	Unit
$D_{\Delta t}$	Fokker–Planck evolution operator over the time interval Δt	
D	Fokker–Planck evolution operator in simplified notation	
K	Collective wake-kick map in simplified notation	
R_ϵ	RF energy-kick map	
R_φ	Longitudinal-drift map	
R	Composite synchrotron-motion map	
M_n	One-step state transition operator at discrete step n	
Π	State-to-observable projection operator	
D	Matrix representation of the Fokker–Planck evolution operator	
K	Matrix representation of the collective wake-kick map	
R_ϵ	Matrix representation of the RF energy-kick map	
R_φ	Matrix representation of the longitudinal-drift map	
M_θ	Matrix representation of the one-step state transition operator at angular time θ	
M_n	Matrix representation of the one-step state transition operator at discrete step n	

Optimization algorithms and evaluation metrics

Sign	Description	Unit
$\mathbf{A}$	Global forward operator	
$\mathbf{b}$	Stacked measurement vector	
B_k	Approximate Hessian at iteration k	
$\mathcal{C}$	Admissible non-negative set	
F	Dynamical constraint residual	
g_k	Objective-function gradient at iteration k	
H_k	Approximate inverse Hessian	
J	Least-squares objective functional	
$\mathcal{K}_\ell$	Krylov subspace at inner iteration ℓ	
$\mathbf{L}$	Regularization operator	
M_k	Positive-definite steepest-descent metric operator	
$\frac{\mathrm{d}J}{\mathrm{d}u}$	Gradient of the objective functional with respect to the control variable	
p_k	Optimization search direction	
r	Data residual vector	
$r(0)$	Zero-lag Pearson correlation coefficient	
R	Regularization functional	
s_k	Optimization step difference	
$\frac{\mathrm{d}x}{\mathrm{d}u}$	State sensitivity matrix with respect to the control variable	
u	Initial phase space density used as the control variable	
α_k	Optimization step length	
β_k	Conjugate-gradient coefficient	
Δ_k	Trust-region radius	
ε_n	Local misfit derivative acting as the adjoint source	

Sign	Description	Unit
γ	Regularization weight	
λ	Adjoint variable	
$\rho(0)$	Zero-lag Spearman rank correlation coefficient	

Glossary

adjoint method Method for efficiently computing the gradient of an objective function by solving an adjoint problem associated with the forward model. 73, 75, 80, 84, 142

birefringence The dependence of a material's refractive index on polarization direction. An applied electric field changes this birefringence, producing the optical phase retardation measured by the EOSD system. 35

chirped laser pulse A laser pulse whose instantaneous optical frequency varies with time. In EOSD, this frequency–time relation maps the temporal bunch-field information onto the measured optical spectrum. 35, 36, 43, 44

co-moving reference frame A coordinate system following the reference orbit, used to describe transverse and longitudinal particle motion relative to the accelerator lattice. 13

collective effects Beam-dynamics effects caused by the electromagnetic fields generated by the particle bunch itself. These fields couple the motion of the particles to the evolving bunch distribution. 1, 2, 5, 22, 26–28, 31, 60, 69, 99, 134

convolution A mathematical operation that combines two functions by evaluating their overlap as one is shifted relative to the other. It is used to model the EOSD signal from the longitudinal bunch profile and the system impulse response. 24, 48–52, 120, 133

dynamical system A mathematical model describing how the state of a system evolves with time. 5, 6

electro-optic crystal A crystal whose refractive properties change in response to an electric field. In the EOSD setup, the bunch field changes the crystal birefringence and the laser polarization. 35, 38–43

emittance A measure of the occupied area or phase space quality of a beam distribution. 32

Fokker–Planck equation A kinetic equation describing the evolution of a probability density under drift and diffusion. In this thesis it represents radiation damping and quantum-excitation effects in the longitudinal dynamics.

forward solver The numerical procedure that propagates an initial state and model parameters through the governing equations to predict the state trajectory and corresponding measurements. 72–74, 80, 82, 86

Hamiltonian A function that describes the energy of a dynamical system and determines its evolution through Hamilton's equations. Here, it describes the conservative longitudinal beam dynamics, including the RF and collective wake effects. 3–6, 8, 9, 13, 26–28, 53–55, 59, 60, 63, 65, 66, 70

ill-posedness Ill-posedness is the property of a problem for which the existence, uniqueness, or stability of a solution is not guaranteed. Phase space reconstruction can exhibit ill-posedness because limited and noisy measurements must determine a high-dimensional density. 70, 89, 144

impulse response The output of a linear measurement system for an impulse-like input. In this thesis it maps a short bunch excitation to the EOSD phase-retardation response and enables convolution-based forward modeling. 3, 4, 48–51, 69, 70, 97, 98, 100, 101, 119, 120, 122, 125–128, 132–134

Inovesalib Modular framework wrapper for the Inovesa solver. 91–95, 97, 134, 155

inverse problem A problem in which unknown controls or states are inferred from measured effects. Here, the initial phase space density is inferred from a sequence of EOSD signals and the known beam-dynamics model. 69, 70, 86, 95, 97–99, 131, 134, 142–144

Krylov subspace The vector space spanned by a starting vector and the vectors obtained by repeatedly applying a linear operator to it. It is used by CGLS to iteratively solve the linearized least-squares reconstruction problem. 86, 87, 144

linearized-dynamics approximation The linearization in which state-dependent evolution operators are held fixed during an inner least-squares solve and updated only during an outer iteration. 76, 85, 97, 104, 105, 107–110

Liouville's theorem A principle of Hamiltonian mechanics stating that phase space volume is preserved during conservative evolution. It provides the basis for the transport of the phase space density under Hamiltonian dynamics. 9

longitudinal impedance The frequency-domain representation of the longitudinal wake function, related to it through a Fourier transform. It characterizes the collective electromagnetic interaction between the bunch and its environment. 22, 24, 25, 34

longitudinal phase space The phase space describing the longitudinal motion of particles within the bunch, represented here by the longitudinal position and relative energy-deviation coordinates. 1–5, 31, 33, 34, 56, 69, 91, 97, 101, 114, 133, 134

manifold A continuous mathematical space that locally resembles Euclidean space and can be described using local coordinates. In dynamical systems, it provides the space in which the system state evolves. 6

matrix-free method A numerical method that applies a matrix and its transpose to vectors without explicitly storing the matrix entries. The reconstruction uses forward and adjoint operator actions in this form. 85–87, 95, 98

micro-bunching instability A collective instability in which small longitudinal density modulations grow and form fine substructures within the bunch. These structures strongly affect the coherent radiation and the measured signals. 5, 24, 31–33, 35, 37, 39–43, 47, 48, 56, 69, 91, 99–101, 111, 114, 116–118, 122, 123, 127–129

momentum compaction The dependence of the closed-orbit circumference, and therefore of the revolution period, on the relative momentum deviation. 1, 18, 62, 91, 128

non-negativity constraint The physical requirement that every value of a phase space density be greater than or equal to zero. The reconstruction enforces it through projection and restart steps. 59, 69, 70, 82, 87, 89, 97, 98, 102, 107–111, 116, 121, 133

observation operator The mapping from a model state to the quantity measured by a diagnostic. In the thesis it combines phase space projection with the EOSD response model. 85

observer pattern A software design pattern in which independent observer components receive updates from a simulation without changing the solver's core logic. 93

operator splitting A time-integration strategy that advances a complex evolution equation by applying separate operators for distinct physical processes. The solver separates wake, RF, transport, and Fokker–Planck actions. 4, 53, 55, 65, 76, 79, 92, 133

Perron–Frobenius operator An operator that transports a density through a deterministic map by accounting for the map's change of volume. It is used as a conceptual basis for the numerical evolution of phase space densities. 55

phase retardation The relative optical phase shift accumulated between orthogonal polarization components after propagating through the electro-optic crystal. It is the measured quantity modeled in the reconstruction. 36, 37, 40, 41, 43, 45–52, 97, 98, 117–119, 126, 133, 135–138

phase space The space formed by the position and momentum coordinates of particles. The thesis uses the longitudinal coordinates q and p to represent this space. 1–7, 9, 10, 13, 25, 26, 28, 31–34, 48, 52, 53, 55–59, 64–67, 69–74, 76, 77, 79–81, 83, 85, 86, 89, 92, 94, 97–108, 110–112, 114–117, 119–134, 141

phase space density A non-negative distribution describing the population of particles over phase space. Its time evolution is the central forward model and reconstruction target of this thesis. 56, 58, 65–67, 69, 70, 82, 101, 102, 105–108, 110, 111, 113–117, 122, 124, 125, 128, 130

phase space tomography The reconstruction of a multidimensional phase space distribution from lower-dimensional projections or measurements. The thesis extends this idea by incorporating the time-dependent beam dynamics into the reconstruction. 1–3, 69, 91, 94, 99, 101, 102, 143

phase-slip factor The factor describing the change of revolution phase with relative momentum deviation; it combines momentum compaction with the relativistic velocity change. 19, 62

plugin architecture A software architecture in which algorithms can be loaded and exchanged dynamically through stable interfaces. Inovesalib uses it to separate the validated solver from reconstruction and analysis algorithms. 93

Pockels effect The linear electro-optic effect in which the refractive-index difference of a crystal changes proportionally to an applied electric field. 35, 45

reference orbit The ideal design trajectory about which particle deviations and accelerator coordinates are defined. 13

regularization The introduction of additional mathematical constraints or penalties to stabilize an ill-posed inverse problem and suppress solutions dominated by noise or unphysical oscillations. 77, 79, 87, 89, 95, 97, 98, 102–105, 107, 108, 132–134

semi-convergence The behavior in which an iterative inverse-problem solution initially approaches the desired solution but later begins fitting measurement noise or model error. Early stopping can therefore act as implicit regularization. 95

sinogram An array of measured projections indexed by projection coordinate and angle or time. In this thesis, sequences of bunch-profile or EOSD measurements. 100, 117–119

spline interpolation A piecewise-polynomial interpolation method used to estimate values between discrete grid points. In the Inovesa transport operators, interpolation evaluates the phase space density at inverse-mapped positions. 57

state sensitivity The derivative of a model state with respect to a control variable or parameter. It quantifies how a perturbation of the control changes the state produced by the forward model. 72, 73

state space The set of all admissible states of a dynamical system. Here, a state is a two-dimensional longitudinal phase space density. 5

synchrotron motion The oscillatory longitudinal motion of a particle around the synchronous particle, produced by the combined action of the radio-frequency system and the ring momentum compaction. 2, 4, 21, 60–63, 70, 91, 101, 114

Tikhonov regularization A quadratic regularization strategy that penalizes the norm of a transformed solution, such as a discrete Laplacian applied to the reconstructed phase space density. 87, 98, 104, 106–110

Toeplitz matrix A matrix whose entries are constant along each diagonal. The discretized EOSD convolution is represented by a Toeplitz observation matrix. 51, 120

Vlasov equation A collisionless kinetic equation that transports a particle density along the self-consistent Hamiltonian flow in phase space. 6, 9, 107

wake field The electromagnetic field left behind by a charged bunch and experienced by particles that follow it. 1, 4, 22, 23, 25, 37–40, 42, 91

wake potential The voltage-like longitudinal effect obtained by convolving a bunch profile with the longitudinal wake function. It is the collective contribution to the longitudinal Hamiltonian. 23–25, 27, 28, 37, 53, 60, 63, 67, 70, 76, 92, 104, 114, 115, 133

warm start An initialization strategy that uses a previously computed solution as the initial guess for a related problem. Consecutive experimental reconstructions use the preceding phase space estimate in this way. 97, 105, 106, 108, 110, 111, 116

Publications

[1] F. Donoso et al. “Longitudinal Phase Space Density Tomography Constrained by the Vlasov–Fokker–Planck Equation”. In: *Proceedings of the 15th International Particle Accelerator Conference (IPAC’24)*. JACoW Publishing, July 1, 2024, pp. 2350–2353. DOI: 10.18429/JACoW-IPAC2024-WEPG55.

[2] F. Donoso et al. “Extended Phase Space Tomography for EOSD Simulation Considering Crystal Geometry Effects”. In: *Proceedings of the 16th International Particle Accelerator Conference (IPAC’25)*. JACoW Publishing, Nov. 5, 2025, pp. 2818–2821. DOI: 10.18429/JACoW-IPAC2025-THPM063.

[3] F. Donoso et al. “Phase Space Tomography Constrained by the Vlasov–Fokker–Planck Equation Using EOSD Diagnostics at KARA”. In: *Proceedings of the 17th International Particle Accelerator Conference (IPAC’26)*. JACoW Publishing, July 20, 2026, pp. 3245–3248. DOI: 10.18429/JACoW-IPAC2026-WEP5146.

[4] F. Donoso et al. “Inovesalib: A Modular Vlasov–Fokker–Planck Solver Framework”. In: *Proceedings of the 17th International Particle Accelerator Conference (IPAC’26)*. JACoW Publishing, July 20, 2026, pp. 3673–3676. DOI: 10.18429/JACoW-IPAC2026-WEP6163.

[5] F. Donoso et al. “Providing Reproducibility and Accessibility of Research Software Development for Accelerator Physics at KARA”. In: *Proceedings of the 17th International Particle Accelerator Conference (IPAC’26)*. JACoW Publishing, July 20, 2026, pp. 3677–3680. DOI: 10.18429/JACoW-IPAC2026-WEP6164.

[6] S. Funkner et al. “First Measurements of the Transient Electrical Near-Field of Electron Bunches in the FLUTE In-Air Section”. In: *Proceedings of the 17th International Particle Accelerator Conference (IPAC’26)*. JACoW Publishing, May 22, 2026, pp. 3036–3039. DOI: 10.18429/JACoW-IPAC2026-WEP6049.

[7] S. Funkner et al. “Reconstructing Wake Functions Using Haïssinski Distributions from Multiple Bunch Charges”. In: *Proceedings of the 16th International Particle Accelerator Conference (IPAC’25)*. JACoW Publishing, Nov. 5, 2025, pp. 2419–2422. DOI: 10.18429/JACoW-IPAC2025-WEPS102.

Bibliography

[8] Alexander Papash et al. “Different Operation Regimes at the KIT Storage Ring KARA (Karlsruhe Research Accelerator)”. In: *Proceedings of the 12th International Particle Accelerator Conference* IPAC2021 (2021). Ed. by Lin Liu (Ed.) et al., 4 pages, 0.735 MB. DOI: 10.18429/JACOW-IPAC2021-MOPAB036. URL: https://jacow.org/ipac2021/doi/JACoW-IPAC2021-MOPAB036.html (visited on 09/13/2026).

[9] Tobias Boltz et al. *On the Perturbation of Synchrotron Motion in the Micro-Bunching Instability*. Version 1. 2022. DOI: 10.48550/ARXIV.2211.15086. Pre-published.

[10] Miriam Brosi. “Overview of the Micro-Bunching Instability in Electron Storage Rings and Evolving Diagnostics”. In: *Proceedings of the 12th International Particle Accelerator Conference*. JACoW Publishing, 2021. DOI: 10.18429/JACOW-IPAC2021-THXA02.

[11] Sebastian Maier et al. “Simulation of the Effect of Corrugated Structures on the Longitudinal Beam Dynamics at KARA”. In: *Proceedings of the International Particle Accelerator Conference (IPAC2022)*. JACoW Publishing, 2022. DOI: 10.18429/JACOW-IPAC2022-WEPOMS006.

[12] Johannes L. Steinmann et al. “Continuous Bunch-by-Bunch Spectroscopic Investigation of the Micro-Bunching Instability”. In: *Physical Review Accelerators and Beams* 21.11 (Nov. 30, 2018), p. 110705. DOI: 10.1103/PhysRevAccelBeams.21.110705.

[13] Miriam Brosi et al. “Synchronous Measurements of Electron Bunches Under the Influence of the Microbunching Instability”. In: *Proceedings of the 10th Int. Particle Accelerator Conf.* JACoW Publishing, 2019. DOI: 10.18429/JACOW-IPAC2019-WEPTS015.

[14] Nicole Hiller. “Electro-Optical Bunch Length Measurements at the ANKA Storage Ring”. Doctoral thesis. Karlsruhe Institute of Technology, 2013. DOI: 10.5445/IR/1000041159.

[15] Michele Caselle et al. “Commissioning of an Ultra-fast Data Acquisition System for Coherent Synchrotron Radiation Detection”. In: *Proceedings of the 5th Int. Particle Accelerator Conf.* JACoW Publishing, 2014. DOI: 10.18429/JACOW-IPAC2014-THPME113.

[16] Benjamin Kehrer et al. “Synchronous Detection of Longitudinal and Transverse Bunch Signals at a Storage Ring”. In: *Physical Review Accelerators and Beams* 21.10 (Oct. 23, 2018), p. 102803. DOI: 10.1103/PhysRevAccelBeams.21.102803.

[17] Meghana Patil et al. “Status and Upgrade of the Visible Light Diagnostics Port for Energy Spread Measurements at KARA”. In: *Proceedings of the International Particle Accelerator Conference (IPAC2023)*. JACoW Publishing, 2023, pp. 4764–4767. DOI: 10.18429/JACoW-IPAC2023-THPL123.

[18] Stefan Funkner et al. “Revealing the Dynamics of Ultrarelativistic Non-Equilibrium Many-Electron Systems with Phase Space Tomography”. In: *Scientific Reports* 13.1 (Mar. 21, 2023), p. 4618. DOI: 10.1038/s41598-023-31196-5.

[19] J. S. Fraser. “Beam Analysis Tomography”. In: *IEEE Transactions on Nuclear Science* 26.1 (Feb. 1979), pp. 1641–1645. DOI: 10.1109/TNS.1979.4330456.

[20] C. T. Mottershead. “Maximum Entropy Beam Diagnostic Tomography”. In: *IEEE Transactions on Nuclear Science* 32.5 (1985), pp. 1970–1972. DOI: 10.1109/TNS.1985.4333784.

[21] G. Jackson. "Design, Implementation, and Results from a Longitudinal Phase Space Tomography (PST) Monitor in the Fermilab Main Ring". In: *Proceedings of International Conference on Particle Accelerators*. May 17, 1993, pp. 2418–2419. DOI: 10.1109/pac.1993.309341.

[22] S. Hancock et al. "Tomographic Measurements of Longitudinal Phase Space Density". In: *Computer Physics Communications* 118.1 (Apr. 1, 1999), pp. 61–70. DOI: 10.1016/S0010-4655(99)00194-0.

[23] Richard Gordon. "A Tutorial on Art (Algebraic Reconstruction Techniques)". In: *IEEE Transactions on Nuclear Science* 21.3 (June 1974), pp. 78–93. DOI: 10.1109/TNS.1974.6499238.

[24] S. Hancock, M. Lindroos, and S. Koscielniak. "Longitudinal Phase Space Tomography with Space Charge". In: *Physical Review Special Topics - Accelerators and Beams* 3.12 (Dec. 12, 2000), p. 124202. DOI: 10.1103/PhysRevSTAB.3.124202.

[25] S. Hancock and J.-L. Sanchez Alvarez. *A Pedestrian Guide to Online Phase Space Tomography in the CERN PS Complex*. Technical report PS/RF/Note 2001-010. CERN, 2001.

[26] Christoffer Hjerto Grindheim and Simon Albright. *Longitudinal Phase Space Tomography Version 3*. Technical report CERN-ACC-NOTE-2021-0004. CERN, Jan. 29, 2021. URL: https://cds.cern.ch/record/2750116.

[27] Simon Albright et al. "Recent Developments in Longitudinal Phase Space Tomography". In: *Proceedings of the 13th International Particle Accelerator Conference*. JACoW Publishing, 2022. DOI: 10.18429/JACOW-IPAC2022-MOPOPT043.

[28] Ryan Roussel et al. *Phase Space Reconstruction from Accelerator Beam Measurements Using Neural Networks and Differentiable Simulations*. Jan. 26, 2023. URL: http://arxiv.org/abs/2209.04505 (visited on 01/31/2023). Pre-published.

[29] Ryan Roussel et al. "Efficient Six-Dimensional Phase Space Reconstructions from Experimental Measurements Using Generative Machine Learning". In: *Physical Review Accelerators and Beams* 27.9 (Sept. 11, 2024), p. 094601. DOI: 10.1103/PhysRevAccelBeams.27.094601.

[30] Austin Hoover and Jonathan C. Wong. *High-Dimensional Maximum-Entropy Phase Space Tomography Using Normalizing Flows*. May 31, 2024. URL: http://arxiv.org/abs/2406.00236 (visited on 06/05/2024). Pre-published.

[31] H. Y. Liu et al. "Real-Time Tomography of Synchrotron Longitudinal Phase Space Based on Spatio-Temporal Neural Networks". In: *Phys. Rev. Accel. Beams* 27.5 (May 21, 2024), p. 052804. DOI: 10.1103/PhysRevAccelBeams.27.052804.

[32] Predrag Cvitanovic et al. *Chaos: Classical and Quantum*. 17.6.4. ChaosBook.org, May 6, 2023. URL: https://chaosbook.org.

[33] Yannis Papaphilippou. *Particle Motion in Hamiltonian Formalism*. Feb. 26, 2024. DOI: 10.48550/arXiv.2402.16474. Pre-published.

[34] Helmut Wiedemann. *Particle Accelerator Physics*. New York, NY: Springer Berlin Heidelberg, 2015.

[35] S. Chandrasekhar. "Stochastic Problems in Physics and Astronomy". In: *Reviews of Modern Physics* 15.1 (Jan. 1, 1943), pp. 1–89. DOI: 10.1103/RevModPhys.15.1.

[36] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. 2nd ed. Springer Series in Synergetics. New York: Springer-Verlag, 1996.

[37] Nelson Wax and M. H. Cohen. "*Selected Papers on Noise and Stochastic Processes*". In: *Physics Today* 8.5 (May 1, 1955), pp. 19–19. DOI: 10.1063/1.3062012.

[38] Matthew Sands. *The Physics of Electron Storage Rings: An Introduction*. Technical report SLAC-R-121. SLAC National Accelerator Laboratory, Nov. 1, 1970. DOI: 10.2172/1453884.

[39] Giovanni Iadarola et al. “A Generalized Tool to Compute Wake Potential and Impedance from Electromagnetic Time Domain Simulations”. In: *Proceedings of the International Particle Accelerator Conference (IPAC2023)*. JACoW Publishing, Sept. 26, 2023, pp. 3522–3525. DOI: 10.18429/JACOW-IPAC2023-WEPL170.

[40] Martin Dohlus and Rainer Wanzenberg. “An Introduction to Wake Fields and Impedances”. In: *Proceedings of the CAS–CERN Accelerator School: Intensity Limitations in Particle Beams*. Ed. by Werner Herr. Geneva: CERN, 2017, pp. 15–41. DOI: 10.23730/CYRSP-2017-003.15.

[41] Marco Venturini et al. “Coherent Synchrotron Radiation and Bunch Stability in a Compact Storage Ring”. In: *Physical Review Special Topics - Accelerators and Beams* 8.1 (Jan. 28, 2005), p. 014202. DOI: 10.1103/PhysRevSTAB.8.014202.

[42] Tobias Boltz et al. “Perturbation of Synchrotron Motion in the Micro-Bunching Instability”. In: *Proceedings of the 10th Int. Particle Accelerator Conf.* JACoW Publishing, 2019. DOI: 10.18429/JACOW-IPAC2019-MOPGW018.

[43] Patrik Schönfeldt. “Simulation and Measurement of the Dynamics of Ultra-Short Electron Bunch Profiles for the Generation of Coherent THz Radiation”. PhD thesis. Karlsruhe Institute of Technology, 2018. DOI: 10.5445/IR/1000084466.

[44] Robert L. Warnock and James A. Ellison. *A General Method for Propagation of the Phase Space Distribution, with Application to the Sawtooth Instability*. Technical report SLAC-PUB-8404. SLAC National Accelerator Laboratory, 2000. DOI: 10.2172/753322.

[45] Stefan Funkner et al. “High Throughput Data Streaming of Individual Longitudinal Electron Bunch Profiles”. In: *Physical Review Accelerators and Beams* 22.2 (Feb. 8, 2019), p. 022801. DOI: 10.1103/PhysRevAccelBeams.22.022801.

[46] Benjamin Kehrer et al. “Visible Light Diagnostics at the ANKA Storage Ring”. In: *Proceedings of the 6th Int. Particle Accelerator Conf.* JACoW Publishing, 2015. DOI: 10.18429/JACOW-IPAC2015-MOPHA037.

[47] Benjamin Kehrer et al. “Turn-by-Turn Horizontal Bunch Size and Energy Spread Studies at KARA”. In: *Proceedings of the 10th Int. Particle Accelerator Conf.* JACoW Publishing, 2019. DOI: 10.18429/JACOW-IPAC2019-WEPGW016.

[48] L. Rota et al. “KALYPSO: Linear Array Detector for High-Repetition Rate and Real-Time Beam Diagnostics”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 936 (Aug. 2019), pp. 10–13. DOI: 10.1016/j.nima.2018.10.093.

[49] Benjamin Kehrer et al. “Time-Resolved Energy Spread Studies at the ANKA Storage Ring”. In: *Proceedings of the 8th Int. Particle Accelerator Conf.* JACoW Publishing, 2017. DOI: 10.18429/JACOW-IPAC2017-MOOCB1.

[50] Meghana Mahaveer Patil. “Development and Integration of High Throughput Detector Systems for Photon-Based Diagnostics at Karlsruhe Research Accelerator”. Doctoral thesis. Karlsruhe Institute of Technology, Oct. 23, 2023. DOI: 10.5445/IR/1000163059.

[51] M. Brosi et al. “Studies of the Micro-Bunching Instability in the Presence of a Damping Wiggler”. In: *Journal of Physics: Conference Series* 1067 (Sept. 2018), p. 062017. DOI: 10.1088/1742-6596/1067/6/062017.

[52] Y.-L. Mathis, B. Gasharova, and D. Moss. “Terahertz Radiation at ANKA, the New Synchrotron Light Source in Karlsruhe”. In: *Journal of Biological Physics* 29.2–3 (June 2003), pp. 313–318. DOI: 10.1023/A:1024429801191.

[53] Johannes Leonhard Steinmann. “Diagnostics of Short Electron Bunches with THz Detectors in Particle Accelerators”. Doctoral thesis. Karlsruhe Institute of Technology, 2019. DOI: 10.5445/KSP/1000090017.

[54] M. Caselle et al. “KAPTURE-2. A Picosecond Sampling System for Individual THz Pulses with High Repetition Rate”. In: *Journal of Instrumentation* 12.01 (Jan. 2017), p. C01040. DOI: 10.1088/1748-0221/12/01/C01040.

[55] Micha Reißig et al. “Development of an Electro-Optical Longitudinal Bunch Profile Monitor at KARA Towards a Beam Diagnostics Tool for FCC-ee”. In: *Proceedings of the 13th International Particle Accelerator Conference*. JACoW Publishing, 2022. DOI: 10.18429/JACOW-IPAC2022-MOPOPT025.

[56] M. Reissig et al. “Simulations of an Electro-Optical in-Vacuum Bunch Profile Monitor and Measurements at KARA for Use in the FCC-ee”. In: *Proceedings of the International Particle Accelerator Conference (IPAC2024)*. JACoW Publishing, July 1, 2024, pp. 2354–2357. DOI: 10.18429/JACOW-IPAC2024-WEPG56.

[57] M. Reissig et al. “First Two-Bunch Measurements Using the Electro-Optical near-Field Monitor at KARA”. In: *Proceedings of the International Particle Accelerator Conference (IPAC2023)*. JACoW Publishing, Sept. 26, 2023, pp. 4756–4759. DOI: 10.18429/JACOW-IPAC2023-THPL121.

[58] Markus Hüning. “Investigations of the Longitudinal Charge Distribution in Very Short Electron Bunches”. In: *Proceedings of DIPAC 2001*. Grenoble, France: JACoW, 2001, pp. 56–60.

[59] S. Casalbuoni et al. “Numerical Studies on the Electro-Optic Detection of Femtosecond Electron Bunches”. In: *Physical Review Special Topics - Accelerators and Beams* 11.7 (July 9, 2008), p. 072802. DOI: 10.1103/PhysRevSTAB.11.072802.

[60] Lorenzo Rota et al. “KALYPSO: A Mfps Linear Array Detector for Visible to NIR Radiation”. In: *Proceedings of the 5th Int. Beam Instrumentation Conf.* JACoW Publishing, 2017. DOI: 10.18429/JACOW-IBIC2016-WEPG46.

[61] Christopher Gerth et al. “KALYPSO: Linear Array Detector with Continuous Read-Out at MHz Frame Rates”. In: *Proceedings of the 8th International Beam Instrumentation Conference*. JACoW Publishing, 2019. DOI: 10.18429/JACOW-IBIC2019-TUPP001.

[62] Bernd Richard Steffen. “Electro-Optic Methods for Longitudinal Bunch Diagnostics at FLASH”. Doctoral thesis. University of Hamburg, 2007.

[63] Micha Reissig. “Development of an Electro-Optical Longitudinal Bunch Profile Monitor for the Future Electron-Positron Collider FCC-ee”. PhD thesis. Karlsruhe Institute of Technology, July 31, 2026. DOI: 10.5445/IR/1000197280.

[64] Dassault Systems. *CST Studio Suite*. 2024. URL: https://www.cst.com.

[65] T. Weiland. “A Discretization Model for the Solution of Maxwell’s Equations for Six-Component Fields”. In: *Archiv Elektronik und Uebertragungstechnik* 31 (Apr. 1, 1977), pp. 116–120. URL: https://ui.adsabs.harvard.edu/abs/1977ArElU..31..116W (visited on 04/27/2026).

[66] B. Kehrer et al. “Numerical Wakefield Calculations for Electro-Optical Measurements”. In: *4th International Particle Accelerator Conference (IPAC 2013), Shanghai, China, May 12-17, 2013 Proceedings Publ.Online Paper MOPME015 JACoW, 2013*. 2013, pp. 503–505.

[67] Bang Wu et al. “Comparison of the Detection Performance of Three Nonlinear Crystals for the Electro-optic Sampling of a FEL-THz Source”. In: *Proceedings of the 5th Int. Particle Accelerator Conf.* JACoW Publishing, 2014. DOI: 10.18429/JACOW-IPAC2014-THPRO017.

[68] Per Christian Hansen. “Deconvolution and Regularization with Toeplitz Matrices”. In: *Numerical Algorithms* 29.4 (2002), pp. 323–378. DOI: 10.1023/A:1015222829062.

[69] Andrzej Lasota and Michael C. Mackey. *Chaos, Fractals, and Noise*. Ed. by J. E. Marsden and L. Sirovich. Vol. 97. Applied Mathematical Sciences. New York, NY: Springer New York, 1994. DOI: 10.1007/978-1-4612-4286-4.

[70] Sergio Blanes, Fernando Casas, and Ander Murua. “Splitting Methods for Differential Equations”. In: *Acta Numerica* 33 (July 2024), pp. 1–161. DOI: 10.1017/S0962492923000077.

[71] Patrik Schönfeldt et al. *Inovesa/Inovesa: The Ones*. Version v1.1.1. Zenodo, Jan. 17, 2021. DOI: 10.5281/ZENODO.597356.

[72] Patrik Schönfeldt et al. “Parallelized Vlasov-Fokker-Planck Solver for Desktop Personal Computers”. In: *Physical Review Accelerators and Beams* 20.3 (Mar. 14, 2017), p. 030704. DOI: 10.1103/PhysRevAccelBeams.20.030704.

[73] Wikipedia contributors. *Lagrange Polynomial*. Wikipedia. 2026. URL: https://en.wikipedia.org/wiki/Lagrange_polynomial (visited on 08/19/2026).

[74] N. Besse and E. Sonnendrücker. “Semi-Lagrangian Schemes for the Vlasov Equation on an Unstructured Mesh of Phase Space”. In: *Journal of Computational Physics* 191.2 (Nov. 2003), pp. 341–376. DOI: 10.1016/S0021-9991(03)00318-8.

[75] Patrik Schönfeldt et al. “Elaborated Modeling of Synchrotron Motion in Vlasov-Fokker-Planck Solvers”. In: *Journal of Physics: Conference Series* 1067 (Sept. 2018), p. 062025. DOI: 10.1088/1742-6596/1067/6/062025.

[76] C. Evain et al. “Spatio-Temporal Dynamics of Relativistic Electron Bunches during the Micro-Bunching Instability in Storage Rings”. In: *EPL (Europhysics Letters)* 98.4 (May 1, 2012), p. 40006. DOI: 10.1209/0295-5075/98/40006.

[77] Bernadette N. Hahn. “Dynamic Linear Inverse Problems with Moderate Movements of the Object: Ill-posedness and Regularization”. In: *Inverse Problems & Imaging* 9.2 (2015), pp. 395–413. DOI: 10.3934/ipi.2015.9.395.

[78] Richard C. Aster, Clifford H. Thurber, and Brian Borchers. *Parameter Estimation and Inverse Problems*. International Geophysics Series. Amsterdam ; Boston: Elsevier Academic Press, 2005.

[79] Max D. Gunzburger. *Perspectives in Flow Control and Optimization*. Advances in Design and Control. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2003.

[80] Harbir Antil and Dmitriy Leykekhman. “A Brief Introduction to PDE-Constrained Optimization”. In: *Frontiers in PDE-Constrained Optimization*. Ed. by Harbir Antil et al. Vol. 163. New York, NY: Springer New York, 2018, pp. 3–40. DOI: 10.1007/978-1-4939-8636-1_1.

[81] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research. New York: Springer, 1999.

[82] William W. Hager and Hongchao Zhang. “A Survey of Nonlinear Conjugate Gradient Methods”. In: *Pacific Journal of Optimization* 2.1 (2006), pp. 35–58.

[83] Dimitri P. Bertsekas. “Projected Newton Methods for Optimization Problems with Simple Constraints”. In: *SIAM Journal on Control and Optimization* 20.2 (Mar. 1982), pp. 221–246. DOI: 10.1137/0320018.

[84] Richard H. Byrd et al. “A Limited Memory Algorithm for Bound Constrained Optimization”. In: *SIAM Journal on Scientific Computing* 16.5 (Sept. 1995), pp. 1190–1208. DOI: 10.1137/0916069.

[85] Patrik Schönfeldt et al. “Study of the Influence of the CSR Impedance on the Synchronous Phase Shift at KARA”. In: *Proceedings of the 9th Int. Particle Accelerator Conf.* JACoW Publishing, 2018. DOI: 10.18429/JACOW-IPAC2018-WEPAL028.

[86] Miriam Brosi et al. “Simulations of the Micro-Bunching Instability for SOLEIL and KARA Using Two Different VFP Solver Codes”. In: *Proceedings of the 13th International Particle Accelerator Conference*. JACoW Publishing, 2022. DOI: 10.18429/JACOW-IPAC2022-WEPOMS005.

[87] M. Brosi et al. “Investigation of the Microwave Instability at MAX IV Laboratory in Combination with Intra-Beam Scattering”. In: *Proceedings of the International Particle Accelerator Conference (IPAC2025)*. JACoW Publishing, Nov. 5, 2025, pp. 2207–2210. DOI: 10.18429/JACOW-IPAC2025-WEPM099.

[88] Sebastian Maier et al. “Simulation Studies on Longitudinal Beam Dynamics Manipulated by Corrugated Structures under Different Bunch Length Conditions at KARA”. In: *Proceedings of the International Particle Accelerator Conference (IPAC2023)*. JACoW Publishing, Sept. 26, 2023, pp. 3570–3573. DOI: 10.18429/JACOW-IPAC2023-WEPL189.

[89] Weija Wang et al. “Accelerated Deep Reinforcement Learning for Fast Feedback of Beam Dynamics at KARA”. In: *IEEE Transactions on Nuclear Science* 68.8 (Aug. 2021), pp. 1794–1800. DOI: 10.1109/TNS.2021.3084515.

[90] M. Sagebaum, T. Albring, and N. R. Gauger. “Expression Templates for Primal Value Taping in the Reverse Mode of Algorithmic Differentiation”. In: *Optimization Methods and Software* 33.4–6 (Nov. 2, 2018), pp. 1207–1231. DOI: 10.1080/10556788.2018.1471140.

[91] Silvia Gazzola, Per Christian Hansen, and James G. Nagy. *IR Tools: A MATLAB Package of Iterative Regularization Methods and Large-Scale Test Problems*. July 1, 2018. URL: http://arxiv.org/abs/1712.05602 (visited on 06/01/2023). Pre-published.

[92] Silvia Gazzola, Per Christian Hansen, and James G. Nagy. “IR Tools: A MATLAB Package of Iterative Regularization Methods and Large-Scale Test Problems”. In: *Numerical Algorithms* 81.3 (July 2019), pp. 773–811. DOI: 10.1007/s11075-018-0570-7.

[93] F. A. Donoso, K. J. Austin, and P. R. McAree. “How Do ICP Variants Perform When Used for Scan Matching Terrain Point Clouds?” In: *Robotics and Autonomous Systems* 87 (Jan. 2017), pp. 147–161. DOI: 10.1016/j.robot.2016.10.011.

[94] James G. Nagy and Zdenek Strakos. “Enforcing Nonnegativity in Image Reconstruction Algorithms”. In: *Mathematical Modeling, Estimation, and Imaging*. Vol. 4121. San Diego, CA: SPIE, Oct. 4, 2000, pp. 182–190. DOI: 10.1117/12.402439.

[95] Silvia Gazzola and Yves Wiaux. “Fast Nonnegative Least Squares Through Flexible Krylov Subspaces”. In: *SIAM Journal on Scientific Computing* 39.2 (Jan. 2017), A655–A679. DOI: 10.1137/15M1048872.

[96] Alexei Botchkarev. “A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms”. In: *Interdisciplinary Journal of Information, Knowledge, and Management* 14 (2019), pp. 045–076. DOI: 10.28945/4184.

[97] Dong C. Liu and Jorge Nocedal. “On the Limited Memory BFGS Method for Large Scale Optimization”. In: *Mathematical Programming* 45.1–3 (Aug. 1989), pp. 503–528. DOI: 10.1007/BF01589116.

[98] Martin Hanke. “Regularizing Properties of a Truncated Newton-Cg Algorithm for Nonlinear Inverse Problems”. In: *Numerical Functional Analysis and Optimization* 18.9–10 (Jan. 1997), pp. 971–993. DOI: 10.1080/01630569708816804.

[99] George Biros and Omar Ghattas. “Parallel Lagrange–Newton–Krylov–Schur Methods for PDE-Constrained Optimization. Part I: The Krylov–Schur Solver”. In: *SIAM Journal on Scientific Computing* 27.2 (Jan. 2005), pp. 687–713. DOI: 10.1137/S106482750241565X.