A Discrete Vlasov Approach to Coherent Collective Instabilities in FEL-Injectors via Perturbation Theory and Semi-Lagrangian Simulation

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Philipp Amstutz

Hamburg
2023

GutachterInnen der Dissertation:	Prof. Dr. Gudrid Moortgat-Pick Dr. Mathias Vogt
Zusammensetzung der Prüfungskommission:	Prof. Dr. Gudrid Moortgat-Pick Dr. Mathias Vogt Prof. Dr. Peter Hauschildt Prof. Dr. Wolfgang Hillert Prof. Dr. Michael Potthoff
Vorsitzender der Prüfungskommission:	Prof. Dr. Michael Potthoff
Datum der Disputation:	22.08.2023
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

In free-electron laser injectors the so-called microbunching instability can severely degrade the beam quality of electron bunches and therefore negatively impact the performance of the free-electron laser. Precise understanding of the microbunching process is required to develop mitigation strategies that minimize the negative impact of the instability. In this work, two approaches of investigating the microbunching instability are presented – one analytical and one numerical. Analytically, a perturbation theory for the instability is derived which is based on a Fréchet–Taylor expansion of the phase-space density propagation operators – the so-called Perron–Frobenius operators – with respect to their dependence on the initial condition of the phase-space density. With this perturbation theory, important problems regarding the microbunching instability can be treated, such as the impact of two-color density modulations, multi-stage bunch compression, and the generation of higher-harmonics due to non-linear effects. Generalized microbunching gain functions for the contributions of the first and higher-order perturbation terms are derived from this theory, which depend only on normalized beam- and machine parameters. Numerically, the microbunching instability is simulated using the semi-Lagrangian code SelaV$_{\mathrm{1D}}$, which was developed during the course of this work. SelaV$_{\mathrm{1D}}$ employs tree-based domain decomposition to represent phase-space densities numerically on a grid. With this, it is possible to efficiently simulate the exotic phase-space densities prevalent in free-electron laser injectors, which exhibit strong non-linear correlations in the longitudinal phase-space, using the grid-based semi-Lagrangian method. Both approaches are applied to study the microbunching instability for the FLASH2020+ upgrade project, in particular with respect to the microbunching mitigation potential of a laser-heater and different bunch compression schemes.

Kurzfassung

In den Injektoren von Freie-Elektronen Lasern kann die sogenannte Microbunchinginstabilität die Strahlqualität der Elektronenpakete negativ beeinflussen und damit die Leistungsfähigkeit des Freie-Elektronen Lasers einschränken. Ein präzises Verständnis des Microbunchingprozesses ist nötig, um Strategien zu entwickeln, die den negativen Einfluss der Instabilität reduzieren. In der vorliegenden Arbeit werden zwei Ansätze ausgearbeitet, um die Microbunchinginstabilität zu untersuchen – ein analytischer und ein numerischer. Der analytische Ansatz besteht in der Herleitung einer Störungstheorie, basierend auf der Fréchet–Taylor Entwicklung der Phasenraumdichtepropagationsoperatoren – den sogenannten Perron–Frobenius Operatoren – im Bezug auf ihre Abhängigkeit von dem Anfangswert der Phasenraumdichte. Mit dieser Störungstheorie können wichtige Fragestellungen bezüglich der Microbunchinginstabilität bearbeitet werden, wie der Einfluss von zweifarbigen Dichtemodulationen, mehrstufige Kompression des Elektronenpakets und die Erzeugung von höheren Harmonischen durch nicht-lineare Effekte. Generalisierte Microbunchingverstärkungsfunktionen für die Beiträge der ersten und höherer Störungsordnungen werden aus dieser Theorie hergeleitet, welche nur von normalisierten Strahl- und Maschinenparametern abhängen. Numerisch wird die Microbunchinginstabilität mit dem semi-Lagrangian Code SelaV_{1D} simuliert, der im Rahmen dieser Arbeit entwickelt wurde. SelaV_{1D} zerlegt die Simulationsdomäne in eine Baumstruktur, um die Phasenraumdichte numerisch effizient auf einem Gitter abzubilden. Damit ist es möglich, die exotischen Phasenraumdichten, welche starke nicht-lineare Korrelationen im longitudinalen Phasenraum aufweisen, wie sie in Freie-Elektronen Lasern häufig vorkommen, mit der gitterbasierten semi-Lagrangian Methode zu simulieren. Beide Ansätze werden angewandt, um die Microbunchinginstabilität für das FLASH2020+ Projekt zu simulieren, mit besonderem Augenmerk auf die Möglichkeit, die Microbunchinginstabilität durch Einsatz eines Laserheaters und verschiedener Bunchkompressionsschemata zu unterdrücken.

List of Publications

Parts of the results presented in this work are published in the following articles:

(a) Philipp Amstutz and Mathias Vogt. Advances in a Perturbation Theory for the Microbunching Instability in Free-Electron Laser Injectors. In *Proc. IPAC'23*, pages 3546–3549. JACoW Publishing, Geneva, Switzerland, May 2023.

(b) Philipp Amstutz. Comparison of Eulerian, Lagrangian and Semi-Lagrangian Simulations of Phase-Space Density Evolution. In *Proc. FEL'22*. JACoW Publishing, Geneva, Switzerland, Aug 2022. in preparation by the editors. (**invited talk**)

(c) Philipp Amstutz and Mathias Vogt. Microbunching Studies for the FLASH2020+ Upgrade Using a Semi-Lagrangian Vlasov Solver. In *Proc. IPAC'22*, pages 2334–2337. JACoW Publishing, Geneva, Switzerland, Jul 2022.

(d) Philipp Amstutz and Mathias Vogt. Investigation of Vlasov Systems with a Certain Class of Linearly-Collective Hamiltonians. In *Proc. IPAC'21*, pages 3157–3160. JACoW Publishing, Geneva, Switzerland, Aug 2021.

(e) Lucas Schaper, Sven Ackermann, Enrico Allaria, Philipp Amstutz, Karolin Baev, Martin Beye, Christopher Gerth, Ingmar Hartl, Wolfgang Hillert, Katja Honkavaara, Mehdi Mohammad Kazemi, Tino Lang, Pardis Niknejadi, Fabian Pannek, Juliane Rönsch-Schulenburg, Dmitrii Samoilenko, Evgeny Schneidmiller, Siegfried Schreiber, Markus Tischer, Mathias Vogt, Mikhail Yurkov, and Johann Zemella. Flexible and Coherent Soft X-ray Pulses at High Repetition Rate: Current Research and Perspectives. *Applied Sciences*, 11(20), 2021.

In (a),(c), and (d) PhA derived the analytical results, conducted the numerical simulations, and wrote the papers; MV provided guidance, corrections, and partial results. In (e) PhA contributed microbunching gain studies, similar to those shown in Section 6.3. (a) is a summary of the perturbation theory presented in Chapter 5. Parts of (b) are summarized in Sections 2.3 and 2.5 of this work. The numerical results of (c) are presented in this work in Sections 6.4.4 and 6.4.5. (d) is summarized in Section 2.4.

List of Key Results

Hurried expert readers might find the following list helpful, which contains pointers to some of the novel results and concepts presented in this thesis.

- Derivation of a class of collective Hamiltonians that are self-preserving under the Vlasov equation (sec. 2.4, p. 32 ff.).
- Symplectic LDU decomposition of the linear single-particle map of a single bunch compression stage (sec. 3.4, p. 54 ff.), multiple stages (sec. 3.4.3, p. 57 ff.), and the resulting notion of the dechirped frame (sec. 3.4.4, p. 57 ff.).
- Recursion relation for the higher-order dispersion coefficients of a magnetic chicane (sec. 3.5, p. 58 ff.) and their small angle limit (sec. 3.5.3, p. 60 ff.).
- Algorithm to calculate RF-parameters for linear bunch compression via streaked functions (sec. 3.6, p. 61 ff.).
- Collective Perron–Frobenius operators and their expansion via Fréchet derivatives (sec. 5.1, p. 83 ff.).
- Perturbation theory for the discrete Vlasov equation via a Fréchet–Taylor expansion of collective Perron–Frobenius operators (sec. 5.2, p. 84 ff.).
- Fréchet derivatives of homomorphic collective Perron–Frobenius operators (sec. 5.4, p. 87 ff.).
- Arbitrary-order phase-space density perturbation term after a single microbunching step and a monochromatic initial perturbation (sec. 5.6.1, p. 93 ff.).
- Expression for the arbitrary-order perturbed charge-density term, determined by applying the symplectic LDU decomposition of the single-particle dynamics (sec. 5.6.2, p. 98 ff.).
- Second-order phase-space density and charge density perturbation for two-color perturbations, showing wave-mixing (sec. 5.6.3, p. 100 ff.).
- General expression for higher-order phase-space density perturbation terms after multiple microbunching steps (sec. 5.7, p. 102 ff.).
- Exact first-order gain function for two microbunching steps (sec. 5.7.1, p. 104 ff.), (sec. 5.8.5, p. 124 ff.).
- Generalized higher-order microbunching gain functions, depending only on normalized beam- and machine parameters (sec. 5.8, p. 107 ff.).
- Development of $\mathrm{SelaV}_{\mathrm{1D}}$ a semi-Lagrangian Vlasov simulation code, employing tree-based domain decomposition (sec. 6, p. 127 ff.).

- Benchmarking SelaV$_{\mathrm{1D}}$ against theory (sec. 6.3, p. 140 ff.) and well-established codes (sec. 6.4.3, p. 147 ff.), (sec. 6.4.6, p. 156 ff.).
- Simulation of the full longitudinal phase-space density of an FEL bunch (sec. 6.4, p. 146 ff.).
- Quantification of higher-order microbunching contributions via two-parametric gain functions (sec. 6.4.4.1, p. 152 ff.).
- Numerical microbunching gain and shot-noise studies for the FLASH2020+ project, regarding possible mitigation schemes using via a laser heater and optimized compression working points (sec. 6.4.4, p. 148 ff.), (sec. 6.4.5, p. 153 ff.).

Contents

1 Introduction **17**

2 Theoretical Foundations **23**
2.1 Hamiltonian Mechanics . . . 23
2.2 Phase-Space Densities . . . 25
2.3 Time-Evolution of Phase-Space Densities . . . 25
2.3.1 Macroscopic Liouville Equation . . . 25
2.3.2 Method of Characteristics . . . 26
2.3.3 Perron–Frobenius Operators . . . 27
2.3.4 Linear Automorphisms on $\mathcal{W}$ . . . 28
2.3.5 Lie Transformation Representation . . . 29
2.3.6 Interacting Particles . . . 30
2.3.6.1 Bogoliubov-Born-Green-Kirkwood-Yvon Hierarchy . . . 30
2.3.6.2 Boltzmann Equation . . . 30
2.3.6.3 Vlasov Equation . . . 31
2.3.7 Collective Perron–Frobenius Operators . . . 31
2.4 Vlasov Equation with Self-Preserving Hamiltonians . . . 32
2.4.1 Convolution-Type Collective Hamiltonians . . . 33
2.4.2 Poisson-Type Kick Hamiltonians . . . 35
2.5 Simulation Methods . . . 35
2.5.1 Lagrangian Approach . . . 35
2.5.2 Eulerian Methods . . . 37
2.5.3 Semi-Lagrangian Method . . . 37
2.6 Beam Dynamics Basics . . . 38
2.6.1 General Electromagnetic Hamiltonian . . . 38
2.6.2 Frenet–Serret Coordinates . . . 38
2.6.3 Canonical Transformation to Frenet–Serret Coordinates . . . 39
2.6.4 Change of the Independent Variable . . . 40
2.6.5 Accelerator Hamiltonian with Independent Coordinate s . . . 41
2.6.6 Typical Accelerator Physics Coordinates . . . 42
2.6.6.1 Conventional Coordinates . . . 42
2.6.6.2 Rose-Hoffstätter and COSY INFINITY Coordinates . . . 42
2.6.6.3 Coordinates used in elegant . . . 42
2.6.6.4 Unscaled Coordinates . . . 43
2.6.7 Paraxial Approximation . . . 43
2.6.8 Magnetic Potentials of Common Magnets . . . 43
2.6.9 Linear Maps of Common Beamline Elements . . . 44
2.6.9.1 Drift Space . . . 44
2.6.9.2 Upright Bend Magnet . . . 45
2.6.9.3 Edge Focusing . . . 46

2.6.9.4 Upright Quadrupole 46

3 Single-Particle Dynamics in High-Gain FEL Injectors 49
3.1 Radio-Frequency Cavity 50
3.2 Drift Space 50
3.3 C-Type Magnetic Chicane 51
3.3.1 Thin-Lens Approximation 52
3.3.2 Thick Bend Magnets 53
3.4 Linear Bunch Compression 54
3.4.1 Initially Chirped Bunch 55
3.4.2 Two-Stage Compression 56
3.4.3 Multi-Stage Compression 57
3.4.4 Dechirped Frame 57
3.5 Non-Linear Longitudinal Dispersion in a Magnetic Chicane 58
3.5.1 Bend Magnets 58
3.5.2 Drift Sections 60
3.5.3 Small Angle Limit 60
3.6 Phase-Space Linearization Algorithm 61
3.6.1 Streaked Functions 64
3.6.2 Inversion of RF Curvature Coefficients 66
3.6.2.1 Single Sine Function 66
3.6.2.2 Double Sine Functions 67
3.6.3 Calculation Procedure 68

4 Collective Effects in FEL Injectors 71
4.1 Longitudinal Space Charge in a Relativistic Bunch 71
4.2 Coherent Synchrotron Radiation 75
4.2.1 Radiation Term 76
4.2.2 Static Term 78
4.2.3 Construction of the Retarded Angle 79
4.2.4 Straight-Line Coulomb Renormalization 80
4.2.5 CSR Kick Function 80
4.2.6 Transient Effects 82

5 Microbunching Instability 83
5.1 Expansion of Collective Perron–Frobenius Operators 83
5.2 Perturbation Theory 84
5.3 Collective-Kick Perron–Frobenius Operators 86
5.4 Fréchet Derivatives of Homomorphic Perron–Frobenius Operators 87
5.4.1 Linearly Collective Hamiltonians 89
5.5 Fréchet Derivatives of Collective-Kick Operators 90
5.6 Expansion of a Single Microbunching Step 91
5.6.1 Monochromatic Perturbation 93
5.6.1.1 General Identities 93
5.6.1.2 First Order 95
5.6.1.3 Second Order 95
5.6.1.4 Higher Order 96
5.6.2 Charge Densities 98
5.6.3 Two-Color Perturbation 100
5.6.3.1 First Order 100
5.6.3.2 Second Order 100

5.7 Multiple Microbunching Steps . . . 102
5.7.1 Two Steps . . . 104
5.8 Application to LSC-driven Microbunching . . . 107
5.8.1 First-Order Propagated Density . . . 109
5.8.1.1 Normalized First-Order Gain Function . . . 110
5.8.1.2 Wavelength-Dependent First-Order Gain Function . . . 112
5.8.2 Second-Order Propagated Density . . . 116
5.8.2.1 Second-Order Gain Function . . . 116
5.8.2.2 Wavelength-Dependent Second-Order Gain Function . . . 121
5.8.3 Third-Order Gain Functions . . . 121
5.8.4 Phase-Space Density Approximations . . . 124
5.8.5 First-Order Two-Step Gain Function . . . 124

6 Semi-Lagrangian Vlasov Simulations 127
6.1 Tree-Based Domain Decomposition . . . 129
6.2 Enhancements of the Semi-Lagrangian Vlasov Code . . . 129
6.2.1 Grid Layout and Inter-Leaf Interpolation . . . 131
6.2.2 Higher-Order Interpolation . . . 131
6.2.2.1 Spline Interpolation . . . 132
6.2.2.2 Local Polynomial Interpolation . . . 132
6.2.2.3 Smooth Local Polynomial Interpolation . . . 133
6.2.3 Input Language Parser . . . 135
6.2.4 Import and Export of Particle Ensembles . . . 136
6.2.4.1 Export . . . 136
6.2.4.2 Import . . . 137
6.2.5 Local Poisson Shot Noise . . . 138
6.3 Discrete-Time Long-Bunch Model . . . 140
6.3.1 Single-Stage Gain Functions . . . 141
6.3.1.1 Small Perturbation . . . 142
6.3.1.2 Large Perturbation . . . 142
6.3.2 Two-Stage Gain Functions . . . 144
6.4 Full Phase-Space Density Simulations for FLASH2020+ . . . 146
6.4.1 Simulation Model and Setup . . . 146
6.4.2 Compression Working Points . . . 146
6.4.3 Verification of the Large-Scale Dynamics . . . 147
6.4.4 Numerical Gain Functions . . . 148
6.4.4.1 Two-Parametric Gain Functions . . . 152
6.4.5 Shot-Noise Simulations . . . 153
6.4.6 Cross-Comparison of the Shot-Noise Results . . . 156

7 Summary 161

A Coordinates for a Type of Curl-Free Potentials 163

B SelaV$_{1D}$ Input Files for Discrete-Time Long-Bunch Model 167

C Phase-Space Linearization Script 169

1 Introduction

Free-electron lasers (FELs) are light sources that can generate brilliant, ultra-short pulses of electro-magnetic radiation with wavelengths ranging from the hard X-ray [1, 2] to the THz spectrum [3]. The conception and first experimental demonstration of an FEL in 1971 [4] marked the starting point of the development of a multitude of accelerator-based light sources around the world that employ variations of the FEL principle [5–9]. Today, FELs are among the most valuable tools that drive scientific progress in many areas of research. X-ray crystallography experiments highly profit from short, yet intense small-bandwidth X-ray pulses provided by FELs, which allow for single-shot measurements unattainable with longer, less intense pulses [10]. Ultrafast processes, such as for instance chemical reactions or charge-transfer processes in molecular systems, can be studied with ultra-short FEL photon pulses on femtosecond timescales [11–13].

Among the most powerful FEL-based light sources are single-pass high-gain FELs. Such FELs are driven by a single-pass linear accelerator – the so-called *injector* – which is capable of supplying high-quality electron bunches for the subsequent generation of radiation. The efficiency of the FEL process and the resulting photon-pulse properties depend crucially on the properties of the electron bunches. Most generally, in order for an electron bunch to successfully drive the FEL process, it has to have a sufficiently high transverse and longitudinal charge density, a small enough energy spread, and an adequately small transverse size. A small transverse electron beam size in the undulator is not only required to achieve the necessary transverse charge density, but also to ensure transverse coherence of the radiation [1]. Typically, the required charge densities exceed what can be produced directly by the electron source – the so-called *gun*. Here, the limiting factor results from the interaction between the electrons via the electrostatic force, also known as *space-charge forces*. If the charge density, and therefore the magnitude of the space-charge forces, is too high, the beam quality at the end of the low-energy section is significantly degraded, in particular with respect to the transverse and longitudinal emittance.

Due to relativistic effects, space-charge forces decrease the higher the central energy of the bunch is, see Section 4.1. Therefore, space-charge forces affect the particle dynamics particularly strongly in the gun region, before the particles are accelerated. Following the gun section, the particles are accelerated in accelerating RF-modules, each consisting of several coupled RF-cavities. Due to resulting increase of the bunch energy, the space-charge forces are reduced. It is this circumstance that allows electron bunches with the required charge density to be produced in the injector. At the gun, the bunch is generated with an intermediate longitudinal charge density, as to keep the space-charge driven degradation of the bunch quality within acceptable limits. Only after the bunch is accelerated to a higher energy, its charge density is increased by compressing the bunch longitudinally. This way, the relativistic suppression the of space-charge forces counteracts their increase resulting from the higher charge density, so that the a good beam quality can be preserved.

In application, longitudinal bunch compression is achieved by the combined effects of

an accelerating RF-module and a subsequent magnetic chicane. Using the RF-module, a correlation between the longitudinal position of a particle within the bunch and the relative deviation of its energy from the central energy of the bunch is generated, which is referred to as an *energy chirp*. The path length of a particle traversing a magnetic chicane depends on the energy of the particle – the higher the particle energy the shorter the path length. This effect is called *longitudinal dispersion*. The sign of energy chirp is chosen so that particles in the back of the bunch have a larger energy than those in the front. Then, due to the longitudinal dispersion, electrons from both, the front and the back of the bunch, move towards the center of the bunch, resulting in a compression of the bunch and consequently in an increase of the charge density.

A magnetic chicane consists of a sequence of dipole magnets that bend the trajectory of the bunch into a curved path. Due to the deflection of the particles in these magnets, they emit synchrotron radiation. This has two unfavorable effects on the beam dynamics in the chicane: Firstly, the stochastic nature of the *incoherent synchrotron radiation* (ISR) emission process causes the energy spread of the bunch to increase, in particular at high energies [14]. Secondly, electrons in the front of the bunch can be affected by the radiation emitted by the back electrons, causing them to gain or lose energy, or change their transverse momentum in the bend plane – an effect typically referred to by the term *coherent synchrotron radiation* (CSR) [15–17], which depends only weakly on the central beam energy. Generally, both effects degrade the overall quality of the electron bunch. The impact of both, ISR and CSR, increases with the deflection angle of the chicane magnets.

In order to keep ISR and CSR effects at bay, the deflection angle – and consequently the longitudinal dispersion – of a magnetic chicane cannot be chosen arbitrarily large. It is therefore impractical to perform the total compression after the bunch was accelerated to its final energy, as either the required deflection angle or the chirp of the bunch would have to be exceedingly large. Moreover, since the impact of ISR and CSR increases with the beam energy, it is advisable to limit the beam energy at the magnetic chicanes. Therefore, high-gain, single-pass FELs typically implement *staged* bunch compression schemes, in which the bunch is compressed not in a single, but multiple magnetic chicanes. Between the chicanes the beam is successively accelerated. Two to three of such bunch compression stages are common. This approach allows to compress the bunch successively to the required final charge density, using magnetic chicanes with intermediate strengths at intermediate beam energies.

Bunch compression can give rise to a collective instability of the electron bunch, known as the *microbunching instability* (MBI) [18–20]. If the longitudinal charge density of an electron bunch contains inhomogeneities, these will generate a space-charge field within the bunch. Interaction with the longitudinal space-charge (LSC) field results in electrons gaining or losing energy, proportional on the magnitude of the LSC field at their position within the bunch. Via this effect, an initial modulation of the longitudinal charge density translates into an energy modulation of the bunch. Conversely, due to the longitudinal dispersion of a subsequent chicane, this energy modulation causes local changes in the charge density. Using the terms introduced above, this can be understood by interpreting an energy modulation as a position-dependent chirp, which results in a *local* compression or decompression of the bunch. This is the LSC-driven microbunching process, which can lead to an amplification of inhomogeneities in the longitudinal charge density. A similar effect can occur in the magnetic chicanes, where it is the interaction with the CSR field that drives the translation of charge density inhomogeneities into energy inhomogeneities.

Due to the microbunching instability, even small initial inhomogeneities in the charge

density can be amplified and cause significant inhomogeneities in the longitudinal phase-space of the bunch – that is, the joint distribution of the energy and longitudinal position of the particles. Further, the microbunching-induced energy modulations effectively increase the energy spread of the bunch. Both effects negatively impact the performance of the bunch in the radiation-generating FEL process.

The FEL process exploits a collective instability that occurs in magnetic devices called *undulators*, which consist of a sequence of small, alternately poled dipole magnets. This FEL instability relies on the resonant exchange of energy between the electron bunch and the synchrotron radiation it produces in the undulator. Similar to the microbunching process in the injector, this causes an energy modulation of the electron bunch. Here, however, the wavelength of the modulation is determined by the wavelength of the undulator radiation. Due to the longitudinal dispersion generated by the undulator dipoles, the energy modulation is converted into modulations of the charge density with the same wavelength. During the evolution of the FEL instability this effect therefore causes the bunch to develop charge density modulations with the wavelength λ of the undulator radiation

$$\lambda = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2}\right), \tag{1.0.1}$$

where λ_u is the undulator period and K is the dimensionless *undulator parameter* [1]. It is this equality of the wavelength of the charge-density modulation and the radiation that enables the generation of coherent radiation in an FEL. Analysis shows that the power of the radiation field increases exponentially during parts of the FEL process, which is quantified by the *gain length*. If, however, the longitudinal phase-space of the bunch contains large inhomogeneities induced by the microbunching instability, the development of the aforementioned coherent charge-density modulations during the FEL instability is hampered, which has a negative impact on the gain length, the quality of the produced photon pulses, and therefore the overall performance of the FEL.

A basic operation mode of an FEL is based on the self-amplification of spontaneous emission (SASE), where the spontaneously emitted undulator radiation triggers and is amplified by the FEL instability. Advanced operation modes of FELs employ sophisticated manipulations of the longitudinal phase-space of the electron bunch before it enters the undulator – an approach known as external *seeding* – to control the initial condition of the instability and consequently control and improve the properties of the generated photon pulses. In particular, the *echo-enabled harmonic generation* (EEHG) seeding scheme is based on the creation of fine substructures in the longitudinal phase-space [21]. If microbunching-induced inhomogeneities are present in the bunch, the seed-substructures are disturbed, which can result in a reduced efficacy of the seeding. Moreover, if the seeded region is small compared to the size of the microbunches, then the shot-to-shot variation of the resulting radiation might be particularly high, as the charge density in the seeded region can vary significantly due to stochastic variations of the initial conditions of the microbunching instability. Overall, due to the resulting stochastic variations of the longitudinal phase-space, the microbunching instability can can contribute significantly to the fluctuations of the photon-pulse properties, not only in seeded operations modes, but also in SASE operation [22–24].

Analysis shows that the growth rate of the microbunching instability depends critically on the energy spread of the electron bunch [18]. To exploit this effect to suppress the microbunching instability, it was proposed to implement a so-called *laser heater* [18]. A laser heater consists of an undulator in which the electron bunch interacts with a laser pulse in the optical or infra-red spectrum, which results in an energy modulation of the bunch at the same wavelength as the laser pulse. In a subsequent magnetic chicane,

these modulations are overlapped with each other in the longitudinal phase-space so that their individual contributions to the charge density average out and the energy spread of the bunch is increased. To achieve optimal performance of the FEL, a setting of the laser heater has to be found so that the total energy spread after the heating process is large enough to suppress the microbunching instability but is still small enough to not impede the FEL process in the undulators.

Another way to manage the microbunching instability is to chose a favorable bunch-compression working point. A multi-stage bunch compression setup allows to achieve a given total target compression factor with different combinations of settings of the relevant operational parameters – namely the settings of the RF-modules and the deflection angles of the magnetic chicanes. Within technical limits, the compression can be variably distributed between the stages. As the non-collective longitudinal dynamics of the bunch are largely governed by these settings, some working points might facilitate the mircobunching instability more than others.

Overall, to successfully design and predictively operate a high-gain single-pass free-electron laser, a good understanding of the microbunching instability and its mitigation methods is required. To this end, in this work two approaches to investigating the microbunching instability are presented. In both approaches, the electron bunch is described by its longitudinal *phase-space density* (PSD) – a smooth function describing the stochastic distribution of the particles in phase-space, which will be introduced in more detail in Section 2.2. Both approaches use the same model of the single-particle dynamics and the collective effects occurring in free-electron laser injectors, which are introduced in Section 3 and Section 4, respectively. Together with Section 2, these sections give a concise introduction to the underlying beam dynamical principles and the theoretical approaches that are used throughout this thesis.

As the microbunching instability results from small inhomogeneities in the longitudinal phase-space density of the bunch, a natural analytical approach is to treat it as a perturbation problem, where the inhomogeneities are viewed as a perturbation to an otherwise homogeneous phase-space density. A perturbative approach has the goal to derive a formal power series in terms of this perturbation for the phase-space density of the bunch at a given time. The time evolution of a phase-space density of interacting particles is governed by the *Vlasov equation*, which is a non-linear, first order integro-partial-differential equation, see Section 2.3. Therefore, in the Vlasov picture, initial inhomogeneities can be treated as a perturbation of the initial conditions of the Vlasov equation. A particularly useful way to describe solutions of the Vlasov equation is via *Perron–Frobenius* (PF) operators, which are the propagation operators of the phase-space density. By expanding the Perron–Frobenius operator of a system with respect to its dependence on the initial conditions, a formal power series for the time-evolved phase-space density can be obtained. Deriving this expansion of collective Perron–Frobenius operators with the help of Fréchet derivatives and formulating the resulting perturbation theory are major parts of this work, which is presented in Section 5.

Complementing the analytical approach, a numerical approach is presented, which is based on the semi-Lagrangian method [25–28]. In the semi-Lagrangian method, the phase-space density is represented numerically by storing values of the phase-space density function on a numerical grid. As described in Section 2.5, an advantage of grid-based methods compared to particle-tracking methods is the absence of artificial numerical shot-noise, which allows to simulate small-scale collective effects – such as the microbunching instability – with higher fidelity. Unfortunately, the longitudinal phase-space densities prevalent in free-electron laser injectors are typically hard to represent efficiently on a numerical grid. This is due to the *exotic* shape of their support, which

regularly features not only a large energy chirp, but additionally also strong non-linear correlations. For such a phase-space density, a rectangular homogeneously spaced grid would cover mostly empty areas of phase-space and only small parts of it would actually intersect the support of the phase-space density. Implementing the semi-Lagrangian approach with a homogeneous grid in a simulation code is therefore not viable, as a prohibitively large amount of memory and computation time would have to be spent on representing and propagating empty areas of phase space. In Section 6, the development of a *se*mi-*La*grangian *V*lasov code called SelaV$_{\mathrm{1D}}$ is presented, which overcomes the numerical challenge posed by exotic phase-space densities by employing tree-based domain decomposition [29]. Here, a rectangular simulation domain is decomposed into a hierarchy of nested sub-rectangles, which forms a tree structure. The hierarchy is generated in a way, so that only those rectangles intersecting the support of the phase-space density have descendants. Only on the *leaf* rectangles – those furthest down the hierarchy – the values of the phase space density are stored on a homogeneous grid covering the leaf. This allows to numerically neglect empty areas of phase-space and store only the non-zero values of the phase-space density.

Both approaches, analytical and numerical, are applied to investigate the microbunching instability and potential mitigation strategies for the FLASH2020+ upgrade project of the free-electron laser FLASH in Hamburg [30].

2 Theoretical Foundations

In this section, a brief summary of the theoretical concepts that are used throughout this work is given. Firstly, a formulation of Hamiltonian systems and solutions of their equations of motion, based on Lie-operators, is recapitulated [31–33]. Subsequently, the concept of a phase-space density is introduced. A method to solve their evolution equation – the Liouville equation – is presented which is based on the method of characteristics. It becomes apparent that solutions of the Liouville equation can be expressed in terms of propagation operators, the so-called *Perron–Frobenius* operators, which play a central role in this work. A summary of the Hamiltonian formulation of accelerator beam dynamics is given, introducing concepts and terminology specific to particle accelerators. The section is concluded by an overview of different simulation methods to simulate the evolution of a phase-space density.

2.1 Hamiltonian Mechanics

In a time-dependent Hamiltonian system the dynamics are governed by single function $C^1 \ni H\colon \mathbb{R}\times\mathbb{R}^{2n} \to \mathbb{R}$, which is called the *Hamiltonian*, where n is the number of degrees of freedom of the system. In addition to the time dependence, the Hamiltonian depends on $2n$ arguments, called the *canonical variables*, which are denoted by $\boldsymbol{q} = (q_1, \ldots, q_n)^{\mathrm{T}}$, and their *conjugate momenta* $\boldsymbol{p} = (p_1, \ldots, p_n)^{\mathrm{T}}$. Together, $\boldsymbol{q}$ and $\boldsymbol{p}$ span the *phase-space* of the system. A complete set of phase-space variables is denoted by the symbol $\boldsymbol{z} = (q_1, p_1, \ldots, q_n, p_n)^{\mathrm{T}}$.

Phase-space variables evolve in time according to the equation of motion generated by the Hamiltonian, which reads

$$\frac{\mathrm{d}\boldsymbol{z}_i}{\mathrm{d}t} = -\left\{H, \boldsymbol{z}_i\right\}\big|_{t,\boldsymbol{z}}\,, \tag{2.1.1}$$

for $i = 1, \ldots, n$, where $\{\cdot,\cdot\}$ denotes the *Poisson bracket*, which is defined as

$$\{f, g\} \equiv \sum_{j=1}^{n} \frac{\partial f}{\partial q_j}\frac{\partial g}{\partial p_j} - \frac{\partial g}{\partial q_j}\frac{\partial f}{\partial p_j}, \tag{2.1.2}$$

for any two functions $C^1 \ni f, g\colon \mathbb{R}^{2n} \to \mathbb{R}$. Equivalently, the Poisson bracket can be written in terms of the gradients of its arguments

$$\{f, g\} = (\nabla_{\boldsymbol{z}} f)^{\mathrm{T}}\, \underline{J}_{2n}\, \nabla_{\boldsymbol{z}} g, \tag{2.1.3}$$

where $\underline{J}_{2n} \in \mathbb{R}^{2n\times 2n}$ is given by the direct sum $\underline{J}_{2n} = J \oplus \overset{\times n}{\cdots} \oplus J$ of the matrix

$$\underline{J} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.1.4}$$

Using this vector notation, the component-wise formulation of the equations of motion (2.1.1) can be brought into the more compact form

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{z} = \underline{J}_{2n}\,\nabla_{\boldsymbol{z}} H|_{t,\boldsymbol{z}}\,, \tag{2.1.5}$$

where the term on the right hand side is called the *vector field* $\boldsymbol{F}(t,\boldsymbol{z}) = \underline{J}_{2n}\,\nabla_{\boldsymbol{z}} H|_{t,\boldsymbol{z}}$ generated by the Hamiltonian.

By defining the *Lie operator* of a phase-space function f via

$$:f:\cdot \equiv \{f,\cdot\} \tag{2.1.6}$$

Equation (2.1.1) can be written as

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}t} = -:H(t,\boldsymbol{z}):\,\boldsymbol{z}. \tag{2.1.7}$$

If the Hamiltonian is actually time-independent $\frac{\mathrm{d}H}{\mathrm{d}t} = 0$, the system is called *autonomous* and the solution of Equation (2.1.7) is given by

$$\boldsymbol{z}(t) = \boldsymbol{\phi}_t(\boldsymbol{z}_0) = \exp(-t:H(\boldsymbol{z}):)\boldsymbol{z}|_{\boldsymbol{z}_0}\,, \tag{2.1.8}$$

for the initial condition $\boldsymbol{z}(t=0) = \boldsymbol{z}_0$. This forms a family of solutions, with the initial condition $\boldsymbol{z}_0$ as the family parameter, which is called the *flow* $\boldsymbol{\phi}_t(\boldsymbol{z})$ of the system. It can be seen that the flow of an autonomous system fulfills the semi-group property with respect to composition

$$\boldsymbol{\phi}_{t_0} \circ \boldsymbol{\phi}_{t_1} \equiv \boldsymbol{\phi}_{t_0}(\boldsymbol{\phi}_{t_1}(\cdot)) = \boldsymbol{\phi}_{t_0+t_1}, \tag{2.1.9}$$

for all $t_0, t_1 \in \mathbb{R}$. This implies that flows are invertible, as it is $\boldsymbol{\phi}_t \circ \boldsymbol{\phi}_{-t} = \boldsymbol{\phi}_0 = \mathrm{Id}$. If the Hamiltonian does depend on time, the system is non-autonomous and the equations of motion are not solved by Equation (2.1.8). Nevertheless, it can be seen that for a sufficiently regular H also in the non-autonomous case a flow $\boldsymbol{\phi}_{t\leftarrow t_0}(\boldsymbol{z})$ exists in some domain around z_0, which additionally depends on the time t_0 of the initial condition and fulfills the relation

$$\boldsymbol{\phi}_{t_2\leftarrow t_1} \circ \boldsymbol{\phi}_{t_1\leftarrow t_0} = \boldsymbol{\phi}_{t_2\leftarrow t_0}, \tag{2.1.10}$$

for all $t_0, t_1, t_2 \in \mathbb{R}$.

Flows define, for a fixed initial and final time, *maps* from the phase-space onto itself, which are, by virtue of the invertibility of flows, in fact *automorphisms* $\boldsymbol{M} \in \mathrm{aut}(\mathbb{R}^{2n})$. A map $\boldsymbol{M}$ is called *symplectic* iff its Jacobian $\underline{D}(\boldsymbol{z}) \equiv \nabla_{\boldsymbol{z}} \boldsymbol{M}|_{\boldsymbol{z}}$ fulfills the condition

$$\underline{D}(\boldsymbol{z})^{\mathrm{T}}\,\underline{J}_{2n}\,\underline{D}(\boldsymbol{z}) = \underline{J}_{2n}, \tag{2.1.11}$$

in which case $\underline{D}(\boldsymbol{z})$ is called a *symplectic matrix*, for all $\boldsymbol{z} \in \mathbb{R}^{2n}$. It can be seen that a flow of a Hamiltonian system always defines symplectic maps, for all values of the time parameter.

While maps arise naturally from the flow of Hamiltonian systems, it can be convenient, especially when modeling physical systems, to be able to define maps ab initio, without knowing the Hamiltonian that generates them explicitly. When defining such a map, it is mandatory to ensure that the map fulfills the symplectic condition (2.1.11), as it would otherwise violate a basic property of Hamiltonian systems. In this work, we make heavy use of this approach to model the longitudinal beam dynamics in an FEL injector.

2.2 Phase-Space Densities

The dynamical state of a single particle is completely determined by its position in phase-space. Consequently, the total state of an ensemble of N particles is exactly described by the set of the phase-space coordinates $\boldsymbol{Z} \in \mathbb{R}^{2nN}$ of its constituents, which we will refer to as the system's *microscopic state*. If the number of particles in the considered ensemble is large, as for instance in a particle bunch where N can easily reach values of $10^8 \dots 10^{12}$, it becomes unwieldy to keep track of the phase-space trajectories of all particles individually. Apart from the fact that the sheer number of particles prohibits any attempt to actually compute the trajectories, it is also typically impossible to know the initial conditions required to describe the initial microscopic state of the system. A statistical approach might therefore be better suited to treat many-particle systems. In such an approach, the phase-space coordinates of the particles are assumed random variables that are distributed following a distribution function that is called the microscopic *phase-space density* (PSD). A microscopic phase-space density is a function $f_N : \mathbb{R}^{2nN} \to \mathbb{R}$ which describes the probability $f_N(\boldsymbol{Z})$ to find the system in the microscopic state $\boldsymbol{Z}$ [34]. Macroscopical properties of the system can then be derived from the statistical properties of this distribution function. While this approach removes the need to specify the precise initial microscopic state of an ensemble, in application it is still a most likely unsurmountable challenge to describe the evolution of the $(2nN)$-parametric microscopic phase-space density, due to the large number of its parameters.

A less intricate description of a many-particle system is possible, if the phase-space coordinates of the particles can be assumed to be *independent, and identically distributed* (IID) [35]. In that case, the initial phase-space coordinates of all particles are distributed following a single, common phase-space density $\Psi : \mathbb{R}^{2n} \to \mathbb{R}$, which is a function of a single set of phase-space coordinates only. Such a function Ψ is commonly referred to as the *macroscopic* phase-space density. This is in contrast to the microscopic phase-space density $f_N : \mathbb{R}^{2nN} \to \mathbb{R}$, which is a function of the phase-space coordinates of all particles. Due to the reduced number of parameters, one-particle phase-space densities are much easier to handle than their microscopic counterparts, making them a viable option to describe IID many-particle systems analytically and puts them within the reach of numerical treatability. Throughout this work, we use the term "phase-space density" exclusively to refer to a one-particle phase-space density and use the modifier "microscopic" where distinction between the two concepts is necessary.

In the following, we will consider phase-space densities that live on the Banach space of Lebesque integrable functions $\mathcal{W} = L^p(\mathbb{R}^{2n}, \mathbb{R})$, with some $1 \leqslant p < \infty$.

2.3 Time-Evolution of Phase-Space Densities

Having introduced the concept of a phase-space density, we will now investigate how to determine the time-evolution of the phase-space density of a many-particle system.

2.3.1 Macroscopic Liouville Equation

If the particles do not interact with one another, it can be seen that the time-dependent phase-space density $\Psi(\boldsymbol{z}, t)$ adheres to the evolution equation

$$\frac{\partial}{\partial t}\Psi = :H: \Psi = (\underline{J}_{2n} \nabla_{\boldsymbol{z}} H)^{\mathrm{T}} \nabla_{\boldsymbol{z}} \Psi \tag{2.3.1}$$

which is known as the *Liouville equation*, with an initial condition $\Psi(t_0, \boldsymbol{z}) = \Psi_0(\boldsymbol{z})$ and where $H : \mathbb{R}^{2n} \to \mathbb{R}$ is the single-particle Hamiltonian of the system [36, 37].

The Liouville equation can be derived by showing that the total derivative of the phase-space density vanishes

$$\frac{\mathrm{d}}{\mathrm{d}t}\Psi = \frac{\partial}{\partial t}\Psi + \left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{z}\right)^{\mathrm{T}} \nabla_{\boldsymbol{z}}\Psi = 0, \tag{2.3.2}$$

which is known as *Liouville's theorem* [37], from which Equation (2.3.1) follows immediately. A proof of Liouville's theorem can be obtained by showing that a phase-space density fulfills the continuity equation

$$\frac{\partial}{\partial t}\Psi + \nabla_{\boldsymbol{z}}^{\mathrm{T}}\left(\Psi\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{z}\right) = 0, \tag{2.3.3}$$

from which Equation (2.3.2) follows after noticing that $\nabla_{\boldsymbol{z}}^{\mathrm{T}}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{z} = 0$ [32].

Along the same lines, it can be seen that the time evolution of a microscopic phase-space density is given by the $2nN$-dimensional equivalent of the Liouville equation

$$\frac{\partial}{\partial t}f_N = :H_{2nN}: f_N = \left(\underline{J}_{2nN}\,\nabla_{\boldsymbol{Z}}H_{2nN}\right)^{\mathrm{T}}\nabla_{\boldsymbol{Z}}f_N, \tag{2.3.4}$$

where the Hamiltonian $H_{2nN}\colon \mathbb{R}^{2nN} \to \mathbb{R}$ is a function of all $2nN$ phase-space variables.

2.3.2 Method of Characteristics

It can be seen from Equation (2.3.1) that the Liouville equation (2.3.1) is a linear first-order partial differential equation (PDE). More specifically, the Liouville equation belongs to a class of PDEs that can be solved by the so-called *method of characteristics* [38]. The method of characteristics works by transforming the PDE into a system of two ordinary differential equations. One is an equation of motion for a trajectory through the domain of the PDE, which in the case of the Liouville equation is the phase space. A solution of this ODE is called a *characteristic*, which gives the method its name. The second ODE describes the rate of change of the value of the solution of the original PDE along its characteristics. From the solutions of both ODEs, a solution of the original PDE can be constructed.

Consider the initial value problem

$$\begin{cases} \frac{\partial}{\partial t}u(t,\boldsymbol{z}) + \boldsymbol{F}(t,\boldsymbol{z})^{\mathrm{T}}\nabla_{\boldsymbol{z}}u(t,\boldsymbol{z}) = S(t,\boldsymbol{z}) \\ u(t_0,\boldsymbol{z}) = u_0(\boldsymbol{z}), \end{cases} \tag{2.3.5}$$

for $u(t,\boldsymbol{z}), u : \mathbb{R}\times\mathbb{R}^{2n} \to \mathbb{R}$, with the known initial condition $u_0(\boldsymbol{z}) : \mathbb{R}^{2n} \to \mathbb{R}$, the vector field $\boldsymbol{F}(t,\boldsymbol{z}) : \mathbb{R}\times\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ and the source term $S(t,\boldsymbol{z}) : \mathbb{R}\times\mathbb{R}^{2n} \to \mathbb{R}$. Let the characteristic $\boldsymbol{\xi}\colon \mathbb{R} \to \mathbb{R}^{2n}$ be a C^1 path through $\mathbb{R}^{2n}$, which originates at the initial time at $\boldsymbol{\xi}(t_0) = \boldsymbol{z}_0$. Let further $v(t), v : \mathbb{R} \to \mathbb{R}$ be the solution of the initial value problem (2.3.5), at time t evaluated at the position $\boldsymbol{\xi}(t)$, that is

$$v(t) \equiv u(t,\boldsymbol{\xi}(t)). \tag{2.3.6}$$

Comparing the total time-derivative of v

$$\frac{\mathrm{d}}{\mathrm{d}t}v(t) = \frac{\partial}{\partial t}u(t,\boldsymbol{z}) + \left(\frac{\partial}{\partial t}\boldsymbol{\xi}(t)\right)^{\mathrm{T}}\nabla_{\boldsymbol{z}}u(t,\boldsymbol{z}) \tag{2.3.7}$$

with the original PDE (2.3.5), it can be seen that the two equations can be made equal by setting

$$\begin{cases} \frac{\partial}{\partial t}\boldsymbol{\xi}(t) = \boldsymbol{F}(t,\boldsymbol{z}), & \boldsymbol{\xi}(t_0) = \boldsymbol{z}_0 \\ \frac{\mathrm{d}}{\mathrm{d}t}v(t) = S(t,\boldsymbol{z}), & v(t_0) = u_0(\boldsymbol{z}_0). \end{cases} \tag{2.3.8}$$

This defines a system of ODEs that is equivalent to the initial PDE.

It can be seen that the Liouville equation (2.3.1) is of the form (2.3.5), where the vector field is given by the Hamiltonian vector field $\boldsymbol{F}(t, \boldsymbol{z}) = \underline{J}_{2n} \nabla_{\boldsymbol{z}} H$ and the source term vanishes $S(t, \boldsymbol{z}) = 0$. Hence, the ODE for the characteristic $\boldsymbol{\xi}$ is equivalent to Hamilton's equations of motion. As a result, we see that the characteristics of the Liouville equation are given by the Hamiltonian flow $\phi_{t \leftarrow t_0}$ of the system, as introduced in Section 2.1, so that $\boldsymbol{\xi}(t) = \phi_{t \leftarrow t_0}(\boldsymbol{z}_0)$. As the source term vanishes, the ODE for v has the trivial solution $v(t) = \Psi(t, \boldsymbol{\xi}(t)) = \Psi_0(\boldsymbol{z}_0)$. Using the invertibility of the flow, it can be seen that the origin of a characteristic can be reconstructed via

$$\boldsymbol{\xi}(t) = \phi_{t \leftarrow t_0}(\boldsymbol{z}_0) \Leftrightarrow \boldsymbol{z}_0 = \phi^{-1}_{t \leftarrow t_0}(\boldsymbol{\xi}(t)) = \phi_{t_0 \leftarrow t}(\boldsymbol{\xi}(t)). \tag{2.3.9}$$

Therefore, an expression for the solution can be obtained from $\Psi(t, \boldsymbol{\xi}(t)) = \Psi_0(\boldsymbol{z}_0)$ by applying the inverse flow to the arguments on both sides which yields

$$\Psi(t, \boldsymbol{z}) = \Psi_0(\phi_{t_0 \leftarrow t}(\boldsymbol{z})). \tag{2.3.10}$$

In summary, we see that the value of the phase-space density is preserved along the characteristics, which are given by the Hamiltonian flow of the system. This well-known result is also commonly referred to as Liouville's theorem.

2.3.3 Perron–Frobenius Operators

In the following, we will describe the dynamics of a system by using maps instead of Hamiltonian flows, keeping in mind the aforementioned close relationship between the two concepts. We have seen that the solution $\Psi_1(\boldsymbol{z}) = \Psi(t_1, \boldsymbol{z})$ of the Liouville equation at any given time t_1 can be constructed from the initial condition $\Psi_0 = \Psi(t_0, \cdot)$, by evaluating the initial condition at the position of the characteristic at time t_0 that reaches the point $\boldsymbol{z}$ at time t_1. Let $\boldsymbol{M} \colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be an invertible, measure preserving map that propagates the phase-space coordinates from t_0 to t_1, which for example could – but does not necessarily have to be – given by a symplectic flow $\boldsymbol{\phi}_{t_1 \leftarrow t_0}$. Via Equation (2.3.10) it can be seen that Ψ_1 is then given by the composition of the initial condition with the inverse of $\boldsymbol{M}$. As this operation is so fundamental for the description of the evolution of phase-space densities, it is convenient to introduce an operator notation for it. To this end [26, 39], the so-called *Perron–Frobenius* (PF) operator $\mathcal{M} \colon \mathcal{W} \to \mathcal{W}$ associated to the map $\boldsymbol{M} \colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is defined by

$$\mathcal{M}\Psi \equiv \Psi \circ \boldsymbol{M}^{-1} = \Psi(\boldsymbol{M}^{-1}(\cdot)), \tag{2.3.11}$$

which is closely related to the Ruelle transfer operator [40]. Perron–Frobenius operators are the propagation operators for phase-space densities in the sense that

$$\Psi_1 = \mathcal{M}\Psi_0. \tag{2.3.12}$$

In this work, we use the calligraphic version of the symbol used for a map to denote its associated Perron–Frobenius operator. The rare exceptions from this convention are pointed out explicitly.

As a PF operator is defined via the inverse of a given invertible map, it immediately follows that a PF operator is invertible itself, where its inverse is given by the composition operation with the associated non-inverted map

$$\mathcal{M}^{-1}\Psi \equiv \Psi \circ \boldsymbol{M}. \tag{2.3.13}$$

With this, it is

$$\mathcal{M}^{-1}\mathcal{M}\,\Psi = \Psi \circ \boldsymbol{M} \circ \boldsymbol{M}^{-1} = \Psi \circ \mathrm{Id} = \Psi \tag{2.3.14}$$

and analogously $\mathcal{M}\mathcal{M}^{-1}\Psi = \Psi$. We note that the PF operator $\mathcal{I}$ associated to the identity map $\mathrm{Id} = \mathrm{Id}^{-1} : \boldsymbol{z} \mapsto \boldsymbol{z}$ is the identity operator in the sense that for any PF operator it is

$$\mathcal{I}\,\mathcal{M}\,\Psi = \Psi \circ \boldsymbol{M}^{-1} \circ \mathrm{Id}^{-1} = \Psi \circ \boldsymbol{M}^{-1} = \mathcal{M}\,\Psi, \tag{2.3.15}$$

and analogously $\mathcal{M}\,\mathcal{I}\,\Psi = \mathcal{M}\,\Psi$, which implies that

$$\mathcal{I}\,\mathcal{M} = \mathcal{M}\,\mathcal{I} = \mathcal{M}. \tag{2.3.16}$$

Further, it can be seen from the associativity of the composition operation that consequently PF operators are also associative in the sense that for any three PF operators $\mathcal{A}, \mathcal{B}, \mathcal{C}$ it is

$$\begin{aligned}(\mathcal{A}\,\mathcal{B})\,\mathcal{C}\,\Psi &= \Psi \circ C^{-1} \circ (A \circ B)^{-1} = \Psi \circ C^{-1} \circ B^{-1} \circ A^{-1} \\ &= \Psi \circ (B \circ C)^{-1} \circ A^{-1} = \mathcal{A}\,(\mathcal{B}\,\mathcal{C})\,\Psi,\end{aligned} \tag{2.3.17}$$

which shows that

$$(\mathcal{A}\,\mathcal{B})\,\mathcal{C} = \mathcal{A}\,(\mathcal{B}\,\mathcal{C}). \tag{2.3.18}$$

From the existence of an inverse element (2.3.13), the existence of an identity element (2.3.16) and the associativity (2.3.18), it follows that Perron–Frobenius operators form a group, which we will denote with the symbol $\mathcal{PF}$, where the group operation is the subsequent application of the operators.

2.3.4 Linear Automorphisms on $\mathcal{W}$

For all $\Psi, \Phi \in \mathcal{W}$ and $\mu, \nu \in \mathbb{R}$ it is

$$\mathcal{M}\,(\mu\Psi + \nu\Phi) = (\mu\Psi + \nu\Phi) \circ \boldsymbol{M}^{-1} = \mu\Psi \circ \boldsymbol{M}^{-1} + \nu\Phi \circ \boldsymbol{M}^{-1} = \mu\mathcal{M}\,\Psi + \nu\mathcal{M}\,\Phi, \tag{2.3.19}$$

which shows that PF operators are a subset of the linear automorphisms on the Banach space of the phase-space densities

$$\mathcal{PF} \subset \mathrm{aut}(\mathcal{W}, \mathcal{W}), \tag{2.3.20}$$

$$\mathcal{PF} \subset \mathrm{lin}(\mathcal{W}, \mathcal{W}). \tag{2.3.21}$$

For any two $\mathbb{K}$-vector spaces U and V, the set of linear maps $\mathrm{lin}(U, V)$ between them is defined to consist of those maps that, for all $x, y \in U$ and all $\mu, \nu \in \mathbb{K}$, fulfill the relation

$$f(\mu x + \nu y) = \mu f(x) + \nu f(y). \tag{2.3.22}$$

In the following it will be shown that $\mathrm{lin}(U, V)$ forms a $\mathbb{K}$-vector space itself. Let $f, g, h \in \mathrm{lin}(U, V)$, $x, y \in U$ and $\mu, \nu \in \mathbb{K}$. From the associativity and commutativity of V it can be seen via

$$[f + (g + h)](x) = f(x) + [g(x) + h(x)] = [f(x) + g(x)] + h(x) = [(f + g) + h](x) \tag{2.3.23}$$

and

$$[f + g](x) = f(x) + g(x) = g(x) + f(x) = [g + f](x), \tag{2.3.24}$$

that $\mathrm{lin}(U, V)$ is associative, $f + (g + h) = (f + g) + h$, and commutative, $f + g = g + f$, as well. Denoting the additive neutral element of V with 0_V, the additive neutral element

of $\mathrm{lin}(U,V)$ is given by the map $0_{\mathrm{lin}(U,V)}: x \mapsto 0_V$, which maps all elements of U to 0_V. With this, it is

$$[f + 0_{\mathrm{lin}(U,V)}](x) = f(x) + 0_{\mathrm{lin}(U,V)}(x) = f(x) + 0_V = f(x), \tag{2.3.25}$$

which implies $f + 0_{\mathrm{lin}(U,V)} = f$. For each element f of $\mathrm{lin}(U,V)$, an additive inverse $-f$ exists, which is given by the map $-f: x \mapsto -(f(x))$, which takes x to the additive inverse of $f(x)$, as this implies

$$[f + (-f)](x) = f(x) + (-f(x)) = f(x) + (-(f(x))) = 0_V. \tag{2.3.26}$$

In the same vein, the remaining vector-space axioms of $\mathrm{lin}(U,V)$ relating to scalar multiplications with elements of $\mathbb{K}$ are directly implied by the corresponding axioms of V and we omit the proofs.

When considering endomorphisms of a vector space – where the domain and codomain of the considered maps is the same, $U = V$ – in addition to vector-space addition, a second inner operation can be defined on the set of linear maps $\mathrm{lin}(U,U)$, namely operator composition, that is subsequent application of maps, which we will denote with $\circ$. For $f, g \in \mathrm{lin}(U,U)$ it is

$$\begin{aligned}[g \circ f](\mu x + \nu y) &= g(f(\mu x + \nu y)) = g(\mu f(x) + \nu f(y)) \\ &= \mu g(f(x)) + \nu g(f(y)) = \mu[g \circ f](x) + \nu[g \circ f](y),\end{aligned} \tag{2.3.27}$$

which implies that $[g \circ f] \in \mathrm{lin}(U,U)$. Operator composition is therefore indeed an inner operation on the linear maps

$$\circ: \mathrm{lin}(U,U) \times \mathrm{lin}(U,V) \to \mathrm{lin}(U,U). \tag{2.3.28}$$

Further, with $f, g, h \in \mathrm{lin}(U,U)$ it can be seen from

$$h \circ [g \circ f] = h([g \circ f](\cdot)) = h(g(f(\cdot))) = h(g(\cdot)) \circ f = [h \circ g] \circ f \tag{2.3.29}$$

that operator composition is associative. It is evident, that the identity map $\mathrm{Id}_{\mathrm{lin}(U,U)}: x \mapsto x$ is linear in the sense of Equation (2.3.22) and is therefore contained in $\mathrm{lin}(U,U)$. For all $f \in \mathrm{lin}(U,U)$ it is

$$\mathrm{Id}_{\mathrm{lin}(U,U)} \circ f = f \circ \mathrm{Id}_{\mathrm{lin}(U,U)} = f, \tag{2.3.30}$$

which shows that $\mathrm{Id}_{\mathrm{lin}(U,U)}$ is the identity element with respect to operator composition. The existence of such an identity element makes $\mathrm{lin}(U,U)$ a *unitary* algebra.

Above we have shown that the set of linear maps between two vector spaces $\mathrm{lin}(U,U)$, forms a vector space itself and it admits a second, associative inner operation, which is operator composition. Therefore, $\mathrm{lin}(U,U)$ forms an *associative algebra*. It is important to note that as Perron–Frobenius operators are a subset of $\mathrm{lin}(\mathcal{W},\mathcal{W})$, they are *contained* within the algebra but the group of PF operators $\mathcal{PF}$ itself is neither a vector space nor an algebra. In Section 5.1, we will exploit the fact that PF operators are contained within the algebra of linear maps to derive a series expansion of a certain type of PF operator.

2.3.5 Lie Transformation Representation

At the time t_1 the phase-space density can be written formally as the power series

$$\Psi(\boldsymbol{z}, t_1) = \sum_{n=0} \frac{(t_1 - t_0)^n}{n!} \left.\frac{\partial^n \Psi}{\partial t^n}\right|_{t=t_0}. \tag{2.3.31}$$

From the Liouville equation (2.3.1), it can be seen that the higher-order partial derivatives appearing in the power series can be expressed by repeated application of the Lie operator $:H:$

$$\frac{\partial^n \Psi}{\partial t^n} =: H :^n \Psi. \tag{2.3.32}$$

By defining the *Lie transformation*

$$\exp(:H:) \equiv \sum_{n=0} \frac{1}{n!} :H:^n \tag{2.3.33}$$

the final phase-space density can be written as

$$\Psi_1(\boldsymbol{z}) = \exp([t_1 - t_0] :H:) \Psi_0(\boldsymbol{z}). \tag{2.3.34}$$

This shows the equivalence of the Perron–Frobenius operator $\mathcal{M}$ and the Lie transformation

$$\mathcal{M} = \exp([t_1 - t_0] :H:). \tag{2.3.35}$$

We will make use of this representation in Section 5.4.1 to derive the Fréchet derivative of a certain class of collective Perron–Frobenius operators, which will be introduced later. An extensive treatise of Lie transformations, their properties and their convergence properties can be found in reference [33].

2.3.6 Interacting Particles

Until now we have assumed that the particles of a many-particle system do not interact with each other, in the sense that the time-evolution of any particle is independent on the phase-space position of all other particles in the ensemble. We will now introduce approaches that allow to take into account certain types of interaction between the particles [41].

2.3.6.1 Bogoliubov-Born-Green-Kirkwood-Yvon Hierarchy

If the interaction between the particles can be described by adding a pair-wise interaction Hamiltonian to the external single-particle Hamiltonian, then the time-evolution of the system is fully described by the resulting microscopic N-particle Liouville equation (2.3.4). Due to the interaction terms, an analytic solution of this collective microscopic Liouville equation is typically not obtainable. Bogoliubov, Born, Green, Kirkwood, and Yvon [42–44] conceived the result that the evolution equation of a s-particle density f_s can be obtained by integrating the N-particle Liouville equation over the phase-space coordinates of the remaining $N - s$ particles. This results in an inhomogeneous Liouville equation for f_s, where the inhomogeneity can be calculated from the $s + 1$-particle density f_{s+1}. As a result, a chain of equations is obtained, relating an s-particle density to the $s + 1$-particle density, which is called the *BBGKY hierarchy*, after the aforementioned authors [42]. Following the BBGKY hierarchy down to its end leads back to the original microscopic Liouville equation. However, truncating the chain at a certain depth, for instance by modeling all following contributions instead of solving them explicitly, is a viable approach to obtain actually treatable kinetic equations.

2.3.6.2 Boltzmann Equation

A prominent example of a kinetic equation that can be interpreted as the result of a truncated BBGKY hierarchy is the *Boltzmann equation* [37, 45]

$$\frac{\partial}{\partial t}\Psi - :H: \Psi = C[\Psi], \tag{2.3.36}$$

which corresponds to an inhomogeneous Liouville equation, where the inhomogeneity is given by the so-called *collision* term $C[\Psi]$, that depends on the current phase-space density Ψ. The Boltzmann equation describes the evolution of the phase-space density of an ensemble of particles when binary *scattering* between the particles can occur. It takes the form

$$C[\Psi](\boldsymbol{q},\boldsymbol{p}) = \int_{\mathbb{R}^{2n}} W(\boldsymbol{p}_a,\boldsymbol{p}_b,\boldsymbol{p}_c,\boldsymbol{p})\big[\Psi(\boldsymbol{q},\boldsymbol{p}_a)\,\Psi(\boldsymbol{q},\boldsymbol{p}_b) - \Psi(\boldsymbol{q},\boldsymbol{p}_c)\,\Psi(\boldsymbol{q},\boldsymbol{p})\big]\,\mathrm{d}\boldsymbol{p}_a\,\mathrm{d}\boldsymbol{p}_b\,\mathrm{d}\boldsymbol{p}_c, \tag{2.3.37}$$

with the scattering cross-section $W(\boldsymbol{p}_a,\boldsymbol{p}_b,\boldsymbol{p}_c,\boldsymbol{p})$, which describes the probability of two particles to change their momenta from $\boldsymbol{p}_a$ and $\boldsymbol{p}_b$ before the collision to $\boldsymbol{p}_c$ and $\boldsymbol{p}$ after the collision. With this it can be seen that the Boltzmann equation (2.3.36) defines a first order, non-linear, integro-partial differential equation.

2.3.6.3 Vlasov Equation

In many physically relevant systems, the particles dominantly interact with one another via *long-range* forces, such as the electrostatic force or gravity. In that case, it is appropriate to employ the so-called *mean-field approximation*, in which the combined interactions with all particles is described via a smooth field [34]. Collective interactions of this type can be accounted for by augmenting the single-particle Hamiltonian H_0 of a system with a term $H_{\mathrm{col}}[\Psi]$ that reflects the collective interaction potential and therefore depends on the current phase-space density Ψ. This yields the so-called *Vlasov equation*

$$\frac{\partial}{\partial t}\Psi(t,\boldsymbol{z}) - :H_0 + H_{\mathrm{col}}[\Psi(t,\cdot)]:\,\Psi(t,\boldsymbol{z}) = 0, \tag{2.3.38}$$

which is the evolution equation of the phase-space density of a system of particles whose interactions can be described by a smooth interaction potential [46].

In many cases the interaction potential is given by the solution of a Poisson-type equation, with the configuration-space density $\rho(\boldsymbol{q}) = \int_{\mathbb{R}^n} \Psi(\boldsymbol{q},\boldsymbol{p})\,\mathrm{d}^n\boldsymbol{p}$ as its source term. Then, the collective Hamiltonian can be determined from the phase-space density via a projection along $\boldsymbol{p}$ followed by a convolution with the Green's function G of the Poisson equation with respect to $\boldsymbol{q}$

$$H_{\mathrm{col}}[\Psi] = \int_{\mathbb{R}^{2n}} G(\boldsymbol{q},\boldsymbol{q}')\,\Psi(\boldsymbol{q}',\boldsymbol{p}')\,\mathrm{d}^n\boldsymbol{p}'\,\mathrm{d}^n\boldsymbol{q}' \tag{2.3.39}$$

Just as the Boltzmann equation, this Vlasov equation is a first order, non-linear, integro-partial differential equation, with the difference that in the former, the non-linearity results from a collective dependence of the inhomogeneity, while in the latter it is a result of the collective dependence of the Hamiltonian.

The Vlasov equation is the appropriate kinetic equation to describe the evolution of the phase-space density of an electron bunch in a particle accelerator, where scattering processes can be neglected and collective interactions can occur for instance via space-charge forces, wake fields, and coherent synchrotron radiation.

2.3.7 Collective Perron–Frobenius Operators

Due to the non-linearity of the Vlasov equation, an explicit analytic solution generally cannot be derived. In particular, the method of characteristics cannot be applied directly to the Vlasov equation. Nevertheless, we will here make the assumption that for well-behaved collective and single-particle Hamiltonians a unique solution of the Vlasov equation does exist, at least on a finite time interval. Let $\Psi[\Psi_0](t,\boldsymbol{z})$ be the solution of

the Vlasov equation (2.3.38), with the initial condition $\Psi(t_0, \boldsymbol{z}) = \Psi_0(\boldsymbol{z})$. We note that $\Psi[\Psi_0](t, \boldsymbol{z})$ is fully determined by Ψ_0, as fixing an initial condition $\Psi_0(\boldsymbol{z})$ directly implies the time-evolution $\Psi[\Psi_0](t, \boldsymbol{z})$ of the system. By plugging in the solution $\Psi[\Psi_0](t, \boldsymbol{z})$ into the argument of the collective Hamiltonian, we obtain the *resultant* collective Hamiltonian

$$\hat{H}[\Psi_0](t, \boldsymbol{z}) = H_0(t, \boldsymbol{z}) + H_{\mathrm{col}}[\Psi[\Psi_0](t, \cdot)](t, \boldsymbol{z}), \tag{2.3.40}$$

which formally depends only on the initial condition Ψ_0, as the current phase-space density $\Psi[\Psi_0]$ is, as mentioned above, fully determined by Ψ_0. Hence, resultant collective Hamiltonians can be viewed as a family of time-dependent Hamiltonians, whose family parameter is the initial condition of the phase-space density Ψ_0. Once this parameter is fixed, the resultant collective Hamiltonian is an ordinary time-dependent Hamiltonian. As a result, also the symplectic flow $\boldsymbol{\phi}[\Psi_0]_{t \leftarrow t_0}$ generated by this Hamiltonian depends on the initial condition parameter. Further, one can define the resultant collective Perron–Frobenius operator $\mathcal{M}[\Psi_0]$ with the associated map $\boldsymbol{\phi}[\Psi_0]_{t_1 \leftarrow t_0}$, which propagates the initial condition to a given time t_1 via

$$\Psi(t_1, \cdot) = \mathcal{M}[\Psi_0]\, \Psi_0. \tag{2.3.41}$$

Formally, a collective Perron–Frobenius operator $\mathcal{M}[\cdot]$ is, with respect to its dependence on the initial condition, a mathematical entity that takes a phase-space density and produces a Perron–Frobenius operator, $\mathcal{M}[\cdot] \colon \mathcal{W} \to \mathcal{PF}$.

It needs to be stressed, that this description of the solution of the Vlasov equation via collective PF operators is generally of formal nature only, as determining a closed-form expression for the resultant PF operator of course still requires a closed-form solution of the Vlasov equation in the first place. Collective PF operators therefore do not provide the means to *solve* a Vlasov equation.

If the solution of a Vlasov equation is known or can be approximated, then collective PF operators provide an excellent formalism to study the influence of changes of the initial condition on the dynamics of the system. Developing this formalism and applying it to derive a perturbation theory for the microbunching instability is a major part of this work, which is presented in Section 5.

2.4 Vlasov Equation with Self-Preserving Hamiltonians

In general, the Vlasov equation (2.3.38) cannot be solved analytically for an arbitrary single-particle Hamiltonian H_0 and collective Hamiltonian H_{col}. Only in special cases an analytical solution can be obtained. In this section we will define a class of Hamiltonians for which the Vlasov equation indeed does admit such an exact solution. This section constitutes a summary of the key results presented in reference [47] by the author.

By defining

$$H[\Psi] \equiv H_0 + H_{\mathrm{col}}[\Psi], \tag{2.4.1}$$

the Vlasov equation (2.3.38) can be written as the initial value problem

$$\begin{cases} \frac{\partial \Psi}{\partial t} - \{H[\Psi], \Psi\} = 0 \\ \Psi(0, \boldsymbol{z}) = \Psi_0(\boldsymbol{z}). \end{cases} \tag{2.4.2}$$

Integrating the Vlasov equation over time yields a formal solution for the phase-space density

$$\Psi(t, \boldsymbol{z}) = \Psi_0(\boldsymbol{z}) + \int_0^t \frac{\partial \Psi}{\partial t'}\, \mathrm{d}t' = \Psi_0(\boldsymbol{z}) + \int_0^t \{H[\Psi], \Psi\}|_{t'}\, \mathrm{d}t', \tag{2.4.3}$$

which is not an explicit solution, as Ψ occurs on both sides of the equation. Substituting the collective argument of the Hamiltonian in the Vlasov equation (2.4.2) with the formal solution (2.4.3) yields

$$\frac{\partial \Psi}{\partial t} - \left\{ H\left[\Psi_0 + \int_0^t \{H[\Psi], \Psi\}|_{t'} \, \mathrm{d}t' \right], \Psi \right\} = 0. \tag{2.4.4}$$

Without making further assumptions about the Hamiltonian, this expression cannot be further simplified. In the following we will consider the case of Hamiltonians that are *affine linear* with respect to their collective dependence – that is, they satisfy the relation

$$H[\mu\Psi + \nu\Phi] = H_{\mathrm{col}}[\mu\Psi + \nu\Phi] + H_0 = \mu H_{\mathrm{col}}[\Psi] + \nu H_{\mathrm{col}}[\Phi] + H_0, \tag{2.4.5}$$

for any $\mu, \nu \in \mathbb{R}$ and $\Psi, \Phi \in \mathcal{W}$. Using this and the fact that the Poisson bracket is bilinear, Equation (2.4.4) can be written as

$$\frac{\partial \Psi}{\partial t} - \{H\left[\Psi_0\right], \Psi\} = \int_0^t \{H_{\mathrm{col}}\left[\{H[\Psi], \Psi\}|_{t'}\right], \Psi\} \, \mathrm{d}t'. \tag{2.4.6}$$

We note that the Hamiltonian on the left-hand side of this equation *depends only on the initial condition* Ψ_0 of the phase-space density – crucially, it does *not* depend on the current phase-space density. Formally, Equation (2.4.6) can be viewed as a Boltzmann equation with a generalized "collision" term which is given by a time integral involving the history of the phase-space density up to the current time t. Therefore, if the right-hand side of Equation (2.4.6) vanishes for all Ψ, the Vlasov equation reduces to the Liouville equation

$$\frac{\partial \Psi}{\partial t} - \{H[\Psi_0], \Psi\} = 0. \tag{2.4.7}$$

The only way the right-hand side of Equation (2.4.6) can vanish for all phase-space densities is that the Hamiltonian in the first argument of the Poisson bracket vanishes

$$H_{\mathrm{col}}\left[\{H[\Psi], \Psi\}\right] = 0, \quad \forall \Psi. \tag{2.4.8}$$

Expressing $H[\cdot]$ via Equation (2.4.1), it can be seen that this condition is fulfilled if both,

$$H_{\mathrm{col}}\left[\{H_{\mathrm{col}}[\Psi], \Psi\}\right] = 0 \tag{2.4.9}$$

and

$$H_{\mathrm{col}}\left[\{H_0, \Psi\}\right] = 0, \tag{2.4.10}$$

hold simultaneously for all Ψ. We call Hamiltonians that fulfill this condition *self-preserving*.

2.4.1 Convolution-Type Collective Hamiltonians

In the following, we consider collective Hamiltonians $H_{\mathrm{col}} = H_K$ that depend on the phase-space density via a convolution with a kernel function $K(\cdot,\cdot)\colon \mathbb{R}^{2n} \times \mathbb{R}^{2n} \to \mathbb{R}$:

$$H_K[\Psi](t, \boldsymbol{z}) \equiv (K * \Psi)(t, \boldsymbol{z}) \equiv \int_{\mathbb{R}^{2n}} K(\boldsymbol{z}, \boldsymbol{z}') \, \Psi(t, \boldsymbol{z}') \, \mathrm{d}\boldsymbol{z}' \tag{2.4.11}$$

In that case, the left-hand side of Equation (2.4.9) can be written as

$$H_K\left[\{H_K[\Psi],\Psi\}\right] = \int_{\mathbb{R}^{2n}} K(\boldsymbol{z},\boldsymbol{z}')\left\{\int_{\mathbb{R}^{2n}} K(\boldsymbol{z}',\boldsymbol{z}'')\Psi(t,\boldsymbol{z}'')\,\mathrm{d}\boldsymbol{z}'', \Psi(t,\boldsymbol{z}')\right\}_{\boldsymbol{z}'}\mathrm{d}\boldsymbol{z}' \quad (2.4.12)$$

$$= \iint\limits_{\mathbb{R}^{2n}\times\mathbb{R}^{2n}} \Psi(t,\boldsymbol{z}'')\; K(\boldsymbol{z},\boldsymbol{z}')\left\{K(\boldsymbol{z}',\boldsymbol{z}''),\Psi(t,\boldsymbol{z}')\right\}_{\boldsymbol{z}'}\mathrm{d}\boldsymbol{z}'\,\mathrm{d}\boldsymbol{z}'' \quad (2.4.13)$$

$$= \iint\limits_{\mathbb{R}^{2n}\times\mathbb{R}^{2n}} \Psi(t,\boldsymbol{z}'')\;\Psi(t,\boldsymbol{z}')\left\{K(\boldsymbol{z},\boldsymbol{z}'),K(\boldsymbol{z}',\boldsymbol{z}'')\right\}_{\boldsymbol{z}'}\mathrm{d}\boldsymbol{z}'\,\mathrm{d}\boldsymbol{z}'', \quad (2.4.14)$$

where the notation $\{\cdot,\cdot\}_{\boldsymbol{z}'}$ is used to denote that the Poisson bracket acts on the coordinates $\boldsymbol{z}'$. Equation (2.4.13) follows from the fact that the Poisson bracket is bilinear, so that the integral over $\boldsymbol{z}'$ can be taken out of the Poisson bracket, assuming that both Ψ and K are sufficiently regular. The validity of Equation (2.4.14) can be seen from Theorem 1, assuming Ψ and K fulfill the smoothness condition required by the theorem.

Theorem 1 (Poisson Bracket Integral)**.** *Let $u \in C^1(\mathbb{R}^{2n},\mathbb{R})$ and $v, w \in C^2(\mathbb{R}^{2n},\mathbb{R})$ with*

$$\lim_{z_i\to\infty} u\,\frac{\partial(v\,w)}{\partial z_j} = 0 \quad (2.4.15)$$

for all $i,j \in \{1,\ldots,2n\}$, then it is

$$\int_{\mathbb{R}^{2n}} \{u,v\}\; w\,\mathrm{d}\boldsymbol{z} = \int_{\mathbb{R}^{2n}} \{w,u\}\; v\,\mathrm{d}\boldsymbol{z}. \quad (2.4.16)$$

Proof. Let J_{ij} be the components of the matrix $\underline{J}_{2n}$ and $f = v\,w$. It is

$$\int_{\mathbb{R}^{2n}} \{u,f\}\;\mathrm{d}\boldsymbol{z} = \sum_{i,j=1}^{2n} J_{ij}\int_{\mathbb{R}^{2n}} \frac{\partial u}{\partial z_i}\frac{\partial f}{\partial z_j}\,\mathrm{d}\boldsymbol{z} \quad (2.4.17)$$

$$= \sum_{i,j=1}^{2n} J_{ij}\int_{\mathbb{R}^{2n-1}}\left[u\,\frac{\partial f}{\partial z_j}\bigg|_{z_i=-\infty}^{z_i=\infty} - \int_{\mathbb{R}} u\,\frac{\partial^2 f}{\partial z_i\partial z_j}\,\mathrm{d}z_i\right]\mathrm{d}(\boldsymbol{z}\backslash z_i) \quad (2.4.18)$$

$$= -\sum_{i,j=1}^{2n} J_{ij}\int_{\mathbb{R}^{2n}} u\,\frac{\partial^2 f}{\partial z_i\partial z_j}\,\mathrm{d}\boldsymbol{z} = 0, \quad (2.4.19)$$

where the last equality follows from the antisymmetry of $\underline{J}_{2n}$. To denote integration over all except the ith coordinates the notation $\int_{\mathbb{R}^{2n-1}}\ldots\mathrm{d}(\boldsymbol{z}\backslash z_i)$ is used.

The Poisson bracket obeys the Leibniz rule $\{u,v\}\,w = \{u,vw\} - \{u,w\}\,v$. Therefore, it is

$$\int_{\mathbb{R}^{2n}} \{u,v\}\; w\,\mathrm{d}z = \int_{\mathbb{R}^{2n}} \{u,v\,w\}\;\mathrm{d}\boldsymbol{z} - \int_{\mathbb{R}^{2n}} \{u,w\}\; v\,\mathrm{d}\boldsymbol{z} = \int_{\mathbb{R}^{2n}} \{w,u\}\; v\,\mathrm{d}\boldsymbol{z}. \quad (2.4.20)$$

□

From Equation (2.4.14), it can be seen that H_K fulfills Equation (2.4.9) if

$$\{K(\boldsymbol{x},\boldsymbol{z}),K(\boldsymbol{z},\boldsymbol{y})\}_{\boldsymbol{z}} = 0 \quad \forall \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{2n}. \quad (2.4.21)$$

Analogously, for the single-particle Hamiltonian the left-hand side of Equation (2.4.10) can be written as

$$H_K\left[\{H_0,\Psi\}\right] = \int_{\mathbb{R}^{2n}} K(\boldsymbol{z},\boldsymbol{z}')\left\{H_0,\Psi(t,\boldsymbol{z}')\right\}_{\boldsymbol{z}'}\mathrm{d}\boldsymbol{z}' \quad (2.4.22)$$

$$= \int_{\mathbb{R}^{2n}} \Psi(t,\boldsymbol{z}')\left\{K(\boldsymbol{z},\boldsymbol{z}'),H_0\right\}_{\boldsymbol{z}'}\mathrm{d}\boldsymbol{z}'. \quad (2.4.23)$$

Therefore, if it is

$$\{K(\boldsymbol{x}, \boldsymbol{z}), H_0\}_{\boldsymbol{z}} = 0 \quad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^{2n}, \tag{2.4.24}$$

then H_0 fulfills the condition (2.4.10).

2.4.2 Poisson-Type Kick Hamiltonians

Of particular importance for this work are Poisson-type kick Hamiltonians, which will be introduced in more detail in Section 5.5. Poisson-type kick Hamiltonians are a subclass of the more general convolution Hamiltonians introduced above. Their kernel function, which arises as the Green's function $G\colon \mathbb{R}^n \times \mathbb{R}^n$ of a Poisson equation, depends only on the configuration-space coordinates $\boldsymbol{q} \in \mathbb{R}^n$ instead of the complete set of phase-space coordinates $\boldsymbol{z} \in \mathbb{R}^{2n}$

$$H_G[\Psi](\boldsymbol{q}, \boldsymbol{p}) = \iint\limits_{\mathbb{R}^n \times \mathbb{R}^n} G(\boldsymbol{q}, \boldsymbol{q}')\, \Psi(\boldsymbol{q}', \boldsymbol{p}')\, \mathrm{d}\boldsymbol{p}'\, \mathrm{d}\boldsymbol{q}'. \tag{2.4.25}$$

As the Green's function does not depend on the conjugate momenta, it immediately follows that the Poisson bracket in Equation (2.4.21) vanishes, which implies that the condition holds. By the same argument, if the single-particle Hamiltonian H_0 also does not depend on the conjugate momenta, Equation (2.4.22) is fulfilled.

This most remarkable result shows that Poisson-type kick Hamiltonians together with drift-type single-particle Hamiltonians are self-preserving. Hence, for this type of Hamiltonians, the Vlasov equation reduces to a Liouville equation, as shown in Equation (2.4.7), where the collective Hamiltonian depends only on the initial condition of the system. In fact, the collective Hamiltonian does not change at all during the evolution of the system – it remains the same as it was at the initial condition of the system. Using the notation of Section 2.3.7 this result reads

$$H_G[\Psi[\Psi_0](t, \cdot)] = H_G[\Psi_0]. \tag{2.4.26}$$

As the resulting Liouville equation – in which the initial condition Ψ_0 is merely a parameter of the Hamiltonian – can be solved explicitly via the method of characteristics, also the associated collective Perron–Frobenius operator can be determined explicitly. Studying the dependence of such a Perron–Frobenius operator on the initial condition is the subject of Section 5.1.

2.5 Simulation Methods

In general, the partial differential equation that describe the time-evolution of the phase-space density of an ensemble of interacting particles cannot be solved analytically, so that a solution can only be obtained numerically. Different approaches to simulating a collective many-body system exist, which can be broadly divided into three classes. In the following, a brief introduction of the three simulation types is given, which constitutes a summary of parts of reference [41] by the author.

2.5.1 Lagrangian Approach

In the *Lagrangian approach*, the many-particle system is represented by an ensemble of simulated particles. As the number of particles present in physical particle bunch can be very large, it is computationally hugely expensive to simulate a particle ensemble with the same number of particles. Therefore, typically only a smaller ensemble of so-called

macroparticles is simulated. Each macroparticle represents multiple physical particles. At the start of the simulation, the phase-space coordinates of the macroparticles are generated following an initial distribution function $\chi\colon \mathbb{R}^{2n} \to \mathbb{R}$, yielding an ensemble $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$. Following [48], it can be seen how the macroparticle approach can be used to represent the phase-space density of a system.

From this macroparticle ensemble, the initial phase-space density Ψ can be approximated by the *Klimontovich density*

$$\Psi_K(t_0, \boldsymbol{z}) = \sum_{i=1}^{N} w_i \, \delta(\boldsymbol{z} - \boldsymbol{z}_i), \tag{2.5.1}$$

where the weights, $w_i \equiv \Psi(t_0, \boldsymbol{z}_i)/[N\,\chi(\boldsymbol{z}_i)]$, depend on the actual initial phase-space density of the system Ψ_0 and the chosen initial macroparticle distribution function χ. An approximation of the expected value of a function f on $\mathbb{R}^{2n}$ with respect to the actual phase-space density can be computed from the Klimontovich density via

$$E_\Psi[f] \equiv \int_{\mathbb{R}^{2n}} f(\boldsymbol{z})\, \Psi(\boldsymbol{z}) \mathrm{d}^{2n} \approx \int_{\mathbb{R}^{2n}} f(\boldsymbol{z})\, \Psi_K(\boldsymbol{z}) \mathrm{d}^{2n} = \sum_{\boldsymbol{z}_i} w_i \, f(\boldsymbol{z}_i), \tag{2.5.2}$$

which is an *unbiased and consistent estimate* for the exact expected value [48].

Let $\boldsymbol{M}\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be the map that maps phase-space coordinates from the time t_0 to t_1. Applying the method of characteristics shows that the propagated Klimontovich density is given by

$$\Psi_K(t_1, \boldsymbol{z}) = \Psi_K(t_0, \boldsymbol{M}^{-1}(\boldsymbol{z})) = \sum_{i=1}^{N} w_i \, \delta(\boldsymbol{z} - \boldsymbol{M}(\boldsymbol{z}_i)), \tag{2.5.3}$$

which shows that a Klimontovich density can be propagated by mapping its underlying macroparticles forward in time and keeping their weights unchanged.

While the initial distribution function χ can in principle be chosen arbitrarily, there are at least two natural choices. Choosing χ to be equal to the initial phase-space density $\Psi(t_0, \cdot)$ results in the weights $w_i = 1/N$ being equal for all macroparticles. Alternatively, the particles might be distributed uniformly in a phase-space region $U \subset \mathbb{R}^{2n}$, covering the support of $\Psi(t_0, \cdot)$, in which case it is $\chi = |U|^{-1}$ and therefore $w_i = \Psi(t_0, \boldsymbol{z}_i)\,|U|/N$. A uniform distribution can be achieved either by generating the particles randomly with uniform distribution or by generating the particles on a regular grid spanning U. Due to the different weights of the macroparticles, this approach is referred to as *weighted macroparticle tracking* [49].

A most significant advantage of Lagrangian methods is their resilience to the curse of dimensions, as their computational demand grows only linearly with the number of phase-space dimensions considered. Increasing the number of phase-space dimensions by 2 only has the consequence that for each particle two more phase-space coordinates have to be stored and propagated. This makes Lagrangian methods a prime choice for systems with a high number of phase-space dimensions.

However, the particulate nature of the phase-space density representation inevitably introduces inherent inhomogeneities in the phase-space density. This phenomenon is known as *artificial shot noise*. Only if the number of macroparticles is close to the number of physical particles, the artificial shot noise matches the actual shot noise of the ensemble. For a smaller number of macroparticles, the artificial shot noise can result in more pronounced inhomogeneities than what would be expected from the natural shot noise of the physical ensemble. Therefore, great care has to be taken to ensure

artificial shot noise does not spuriously drive collective instabilities. This is typically achieved by smoothing the resulting charge density heuristically. If done carelessly, this can potentially remove or damp also the actual non-artificial inhomogeneities that are being studied, potentially invalidating the results of the simulation.

2.5.2 Eulerian Methods

In *Eulerian* methods, a phase-space density is represented numerically by storing its values on discrete points on a grid which covers a region of the phase-space.

A plethora of different Eulerian methods exists to determine the time evolution of such a representation of the phase-space density. Prominent examples are finite volumes, finite elements, and finite differences methods [50]. In a finite differences method, based on the numerical phase-space density values, the gradients of the phase-space density are approximated by finite difference schemes, which allows to determine the local rate of change of the phase-space density from the underlying kinetic equation. Employing a discrete time-integration scheme, the new local values of the phase-space density after a time step are approximated.

A significant advantage of Eulerian methods compared to Lagrangian methods is that, by storing the actual values of the phase-space density on a grid, they produce – when used properly – a smooth numerical representation of the phase-space density. In particular, a grid-based representation does not add numerical shot noise.

Due to their reliance on a numerical grid, Eulerian methods are susceptible to the curse of dimensions, as the number of grid points that is required to achieve a given resolution grows exponentially with the number of phase-space dimensions. Further, finite-difference schemes have to fulfill strict criteria to ensure their stability, consistency and convergence. A particularly critical one is the Courant-Friedrichs-Lewy (CFL) condition, which poses an upper limit for the length of a time-step to ensure the convergence of the numerical solution. The finer the spatial resolution of the grid, the smaller is the maximum time step allowed by the CFL condition. Especially when studying effects that occur on small scales, this can lead to prohibitively small time steps being required. Due to these restrictions Eulerian methods are less commonly employed in beam-dynamics simulations than Lagrangian methods.

2.5.3 Semi-Lagrangian Method

In the *semi-Lagrangian method*, the phase-space density is represented numerically by storing its values on a grid, similar to an Eulerian method. The distinguishing feature of the semi-Lagrangian method is that the phase-space density is propagated by direct application of the method of characteristics (2.3.10), which states that given an initial phase-space density $\Psi_0 = \Psi(t_0, \cdot)$, the value of the propagated phase-space density $\Psi_0 = \Psi(t_1, \cdot)$ at a position $\boldsymbol{z}$ can be calculated by evaluating Ψ_0 at the origin of the characteristic that ends at $\boldsymbol{z}$:

$$\Psi_1(\boldsymbol{z}) = \Psi_0(\boldsymbol{M}^{-1}(\boldsymbol{z})), \tag{2.5.4}$$

where $\boldsymbol{M} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is the map that maps the phase-space coordinates from t_0 to t_1 [25–28]. This can be directly implemented in a simulation code. To calculate the new phase-space density numerically, first an empty grid is set up. To determine the value of the new phase-space density $\Psi_1(\boldsymbol{z}_i)$ at a grid point, the coordinates of the point $\boldsymbol{z}_i$ are tracked backwards in time, which yields their origin $\boldsymbol{M}^{-1}(\boldsymbol{z}_i)$ at the initial time. Subsequently, the numerical representation of Ψ_0 is evaluated at that origin $\boldsymbol{M}^{-1}(\boldsymbol{z}_i)$, which finally yields the new value $\Psi_1(\boldsymbol{z}_i) = \Psi_0(\boldsymbol{M}^{-1}(\boldsymbol{z}_i))$. Generally, the origin $\boldsymbol{M}^{-1}(\boldsymbol{z}_i)$

of $\boldsymbol{z}_i$ will not coincide with a grid point of the numerical representation of Ψ_0. As the values of Ψ_0 are known only on the grid points, an interpolation scheme needs to be employed to evaluate Ψ_0 at such a off-grid point. This process is repeated for each grid point of Ψ_1, which in the end results in a complete numerical representation of the new phase-space density.

As the phase-space density is represented on a grid, but the update step is based on tracking coordinates, the semi-Lagrangian method can be considered a hybrid between a Lagrangian and an Eulerian method. A significant advantage of the semi-Lagrangian method over a Lagrangian method is the absence of numerical shot-noise, which allows this method to be used to study systems in which relevant effects occur on small scales, as for instance the microbunching instability. Compared to finite difference methods, the semi-Lagrangian method has the advantage that it allows for arbitrarily long time steps without becoming inherently instable. As long as the phase-space map M is known, the update step can be executed, independent on the length of the time step it represents. As any grid-based method, the semi-Lagrangian method is susceptible to the curse of dimensions, which is why it is usually only used to simulate systems with a small number of degrees of freedom [26, 51–53].

In Section 6, the development of the semi-Lagrangian Vlasov simulation code SelaV_{1D} is presented. It overcomes a challenge that grid-based simulation codes face, when it comes to simulating a phase-space density whose support is shaped in a way that does not allow it to be captured efficiently on a homogeneous grid, via a tree-based domain decomposition method.

2.6 Beam Dynamics Basics

In particle accelerators, charged particles are guided and accelerated by electromagnetic fields. To describe the dynamics of a particle traversing these fields, the Hamiltonian formalism has proven itself to be a valuable tool. In the following, the Hamiltonian formalism for charged particles in electromagnetic fields is recapitulated, including the coordinate transformations commonly used in accelerator physics [54–56].

2.6.1 General Electromagnetic Hamiltonian

In a fixed Cartesian coordinate system with coordinates $\boldsymbol{q} = (q_x, q_y, q_z)^{\mathrm{T}}$, the Hamiltonian of a charged particle with mass m and charge e in an electromagnetic field with four-potential $(\phi, \boldsymbol{A})$, $\boldsymbol{A} = (A_x, A_y, A_y)^{\mathrm{T}}$ reads

$$H(\boldsymbol{q}, \boldsymbol{p}; t) = e\phi(q) + c\sqrt{(\boldsymbol{p} - e\boldsymbol{A}(q))^2 + m^2c^2}, \tag{2.6.1}$$

where $\boldsymbol{p} = (p_x, p_y, p_z)$ is the canonical momentum $\boldsymbol{p} = \gamma m\boldsymbol{v} + e\boldsymbol{A}$ and t is the independent time coordinate. Assuming that both, $\boldsymbol{A}$ and ϕ, are time-independent, also H does not depend explicitly on time and hence is autonomous.

2.6.2 Frenet–Serret Coordinates

In an accelerator beamline, all particles move on similar trajectories and individual particles deviate only slightly from a reference path $\boldsymbol{r}(s)$, which is commonly referred to as the *reference orbit*. The variable s is called the *arc length*, which is defined as the distance traveled along the reference orbit. It is sensible to describe the particle motion in coordinates relative to this reference orbit instead of in a fixed coordinate system. One such coordinate system is defined by the so-called *Frenet–Serret* apparatus [57].

The origin of this system follows the reference orbit and its basis vectors are the tangent $\boldsymbol{\tau}$, normal $\boldsymbol{n}$ and binormal $\boldsymbol{b}$ of the reference orbit defined by

$$\boldsymbol{\tau} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}, \quad \boldsymbol{n} = \left\| \frac{\mathrm{d}\boldsymbol{\tau}}{\mathrm{d}s} \right\|^{-1} \frac{\mathrm{d}\boldsymbol{\tau}}{\mathrm{d}s}, \quad \text{and} \quad \boldsymbol{b} = \boldsymbol{\tau} \times \boldsymbol{n}. \tag{2.6.2}$$

These basis vectors define a right-handed coordinates system and satisfy the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}s} \begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{n} \\ \boldsymbol{b} \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \iota \\ 0 & -\iota & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{n} \\ \boldsymbol{b} \end{pmatrix}, \tag{2.6.3}$$

where the quantity κ is commonly called the local *curvature*, which is equivalent to the inverse of the local bending radius $\kappa = \frac{1}{R}$ of the orbit. The quantity ι is the *torsion* of the orbit. In most cases, accelerators are designed to be close to planar so that the torsion of their orbit vanishes. With this, any position $\boldsymbol{q}$ in the vicinity of the orbit can be uniquely expressed in the new coordinates $\boldsymbol{Q} = (Q_x, Q_y, Q_s)^{\mathrm{T}}$ as

$$\boldsymbol{q} = \boldsymbol{r}(Q_s) + Q_x \boldsymbol{n}(Q_s) + Q_y \boldsymbol{b}(Q_s), \tag{2.6.4}$$

where we have introduced $Q_s \equiv s$ for the sake of uniform notation.

2.6.3 Canonical Transformation to Frenet–Serret Coordinates

To transform the Hamiltonian from the fixed Cartesian coordinate system to the co-moving Frenet–Serret coordinate system, a canonical coordinate transformation with the time independent generating function

$$F = F_3(\boldsymbol{Q}, \boldsymbol{p}; t) + \boldsymbol{q} \cdot \boldsymbol{p} \quad \text{with} \quad F_3(\boldsymbol{Q}, \boldsymbol{p}; t) = -\boldsymbol{p} \cdot [\boldsymbol{r}(Q_s) + Q_x \boldsymbol{n}(Q_s) + Q_y \boldsymbol{b}(Q_s)] \tag{2.6.5}$$

has to be performed [55, 56]. The relation between the old and new coordinates is then given by

$$q_x \equiv -\frac{\partial F_3}{\partial p_x}, \quad q_y \equiv -\frac{\partial F_3}{\partial p_y}, \quad \text{and} \quad q_z \equiv -\frac{\partial F_3}{\partial p_z}, \tag{2.6.6}$$

which indeed gives the correct expression (2.6.4) when plugging in (2.6.5). For the new canonical momenta $P = (P_x, P_y, P_s)$ this transformation yields

$$P_x \equiv -\frac{\partial F_3}{\partial Q_x} = \boldsymbol{p} \cdot \boldsymbol{n}(Q_s) \tag{2.6.7}$$

$$P_y \equiv -\frac{\partial F_3}{\partial Q_y} = \boldsymbol{p} \cdot \boldsymbol{b}(Q_s) \tag{2.6.8}$$

$$P_s \equiv -\frac{\partial F_3}{\partial Q_s} = \boldsymbol{p} \cdot \left[\boldsymbol{\tau}(Q_s) + Q_x \left\{-\kappa \boldsymbol{\tau}(Q_s) + \iota \boldsymbol{n}(Q_s)\right\} - \iota Q_y \boldsymbol{n}(Q_s)\right]. \tag{2.6.9}$$

In the following, we assume that the torsion of the orbit is zero $\iota = 0$. This fixes the bend plane to the x-plane. In this case, P_s simplifies to

$$P_s = \boldsymbol{p} \cdot \boldsymbol{\tau}(Q_s) \left[1 - Q_x \kappa\right]. \tag{2.6.10}$$

By defining $\underline{T} \in \mathrm{SO}(3)$ to be the matrix with row vectors $\boldsymbol{n}$, $\boldsymbol{b}$, and $\boldsymbol{\tau}$, the new momenta can be written as

$$\left(P_x, P_y, \frac{P_s}{1 - Q_x \kappa}\right)^{\mathrm{T}} = \underline{T} \boldsymbol{p}, \tag{2.6.11}$$

which can be inverted to yield an explicit expression for the old momenta

$$\boldsymbol{p} = \underline{T}^{\mathrm{T}} \left(P_x, P_y, \frac{P_s}{1 - Q_x \kappa} \right)^{\mathrm{T}}. \tag{2.6.12}$$

Similarly, the vector potential $\boldsymbol{A}^* = (A_x^*, A_y^*, A_s^*)^{\mathrm{T}}$ in the new basis is given by

$$\left(A_x^*, A_y^*, \frac{A_s^*}{1 - Q_x \kappa} \right)^{\mathrm{T}} = \underline{T} \boldsymbol{A}, \tag{2.6.13}$$

which implies

$$\boldsymbol{A} = \underline{T}^{\mathrm{T}} \left(A_x^*, A_y^*, \frac{A_s^*}{1 - Q_x \kappa} \right)^{\mathrm{T}}. \tag{2.6.14}$$

The scalar potential is invariant under this transformation, so that we have

$$\phi^* = \phi. \tag{2.6.15}$$

Generally, the new Hamiltonian H^* after a canonical transformation with generating function F is given by

$$H^* = H + \frac{\partial F}{\partial t}, \tag{2.6.16}$$

which for the case at hand simplifies to $H^* = H$, as the generating function is not explicitly time dependent. Hence, H^* can be generated from Equation (2.6.1) by expressing the old coordinates and potential in terms of their new counterparts, which gives

$$H^*(\boldsymbol{Q}, \boldsymbol{P}; t) = e\phi^* + c\sqrt{(P_x - eA_x^*)^2 + \left(P_y - eA_y^*\right)^2 + \left(\frac{P_s - eA_s^*}{1 - Q_x \kappa} \right)^2 + m^2 c^2}, \tag{2.6.17}$$

where we have suppressed the arguments of the vector potential $\boldsymbol{A}^* = \boldsymbol{A}^*(\boldsymbol{q}(\boldsymbol{Q}))$ and scalar potential $\phi^* = \phi^*(\boldsymbol{q}(\boldsymbol{Q}))$.

2.6.4 Change of the Independent Variable

Up to now, the system was described with the time t as the independent variable. For the application to accelerator physics, it is however much more convenient to use the arc length s as the independent variable, because while the physical distance between elements of the beam line along the orbit is given by the geometry of the machine, the arrival time of a particle at those elements varies between particles [55]. Unfortunately, this change of the independent variable cannot be achieved directly via a generating function as before. Instead, in order to derive the new Hamiltonian and canonical variables we have to resort to the principle of least action

$$\delta S \equiv \delta \int_{t_0}^{t_1} \left[\boldsymbol{p} \cdot \frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t} - H(\boldsymbol{q}, \boldsymbol{p}; t) \right] \mathrm{d}t = 0, \tag{2.6.18}$$

with $\boldsymbol{q} = (q_1, \ldots, q_n)^{\mathrm{T}}$ and $\boldsymbol{p} = (p_1, \ldots, p_n)^{\mathrm{T}}$. Assume we want the coordinate q_n to act as the new independent variable. To achieve this, we start by taking its conjugate variable p_n and consider it as a function of the remaining canonical variables as well as the Hamiltonian H and time t

$$p_n = p_n(\boldsymbol{q}, p_1, \ldots, p_{n-1}, t, H). \tag{2.6.19}$$

Equation (2.6.18) then gives

$$\int_{t_0}^{t_1}\Bigg(\sum_{i=1}^{n-1}\Big[\delta p_i\frac{\mathrm{d}q_i}{\mathrm{d}t}+p_i\frac{\mathrm{d}\delta q_i}{\mathrm{d}t}+\frac{\partial p_n}{\partial q_i}\delta q_i\frac{\mathrm{d}q_n}{\mathrm{d}t}+\frac{\partial p_n}{\partial p_i}\delta p_i\frac{\mathrm{d}q_n}{\mathrm{d}t}\Big]+ \\ \frac{\partial p_n}{\partial q_n}\delta q_n\frac{\mathrm{d}q_n}{\mathrm{d}t}+\frac{\partial p_n}{\partial t}\delta t\frac{\mathrm{d}q_n}{\mathrm{d}t}+\frac{\partial p_n}{\partial H}\delta H\frac{\mathrm{d}q_n}{\mathrm{d}t}-\delta H-H\frac{\mathrm{d}\delta t}{\mathrm{d}t}\Bigg)\mathrm{d}t=0, \tag{2.6.20}$$

which – after applying integration by parts to the terms containing time derivatives of varied quantities, using $\frac{\partial p_n}{\partial q_n}=0$ and reordering the terms – yields

$$\int_{t_0}^{t_1}\Bigg(\sum_{i=1}^{n-1}\Big[\delta p_i\Big\{\frac{\mathrm{d}q_i}{\mathrm{d}t}+\frac{\partial p_n}{\partial p_i}\frac{\mathrm{d}q_n}{\mathrm{d}t}\Big\}+\delta q_i\Big\{\frac{\partial p_n}{\partial q_i}\frac{\mathrm{d}q_n}{\mathrm{d}t}-\frac{\mathrm{d}p_i}{\mathrm{d}t}\Big\}\Big] \\ +\delta t\Big\{\frac{\partial p_n}{\partial t}\frac{\mathrm{d}q_n}{\mathrm{d}t}+\frac{\mathrm{d}H}{\mathrm{d}t}\Big\}+\delta H\Big\{\frac{\partial p_n}{\partial H}\frac{\mathrm{d}q_n}{\mathrm{d}t}-1\Big\}\Bigg)\mathrm{d}t=0. \tag{2.6.21}$$

As Equation (2.6.21) has to hold for any variation of the trajectory, all terms in curly braces have to vanish. This gives the new equations of motion

$$\frac{\mathrm{d}q_i}{\mathrm{d}q_n}=-\frac{\partial p_n}{\partial p_i},\quad \frac{\mathrm{d}p_i}{\mathrm{d}q_n}=+\frac{\partial p_n}{\partial q_i},\quad \frac{\mathrm{d}H}{\mathrm{d}q_n}=-\frac{\partial p_n}{\partial t}\quad\text{and}\quad \frac{\mathrm{d}t}{\mathrm{d}q_n}=+\frac{\partial p_n}{\partial H}. \tag{2.6.22}$$

These have the structure of Hamilton's equations of motion, where q_n is the independent variable, $-p_n$ is the Hamiltonian, and $q_1,\dots q_{n-1},H$ are the new canonical coordinates and $p_1,\dots,p_{n-1},t$ are their conjugate momenta [58].

2.6.5 Accelerator Hamiltonian with Independent Coordinate s

As we have seen, by choosing $Q_s\equiv s$ as the new independent coordinate, $-P_s$ becomes the new Hamiltonian. Solving Equation (2.6.17) for P_s gives

$$P_s=eA_s^*+(1-Q_x\kappa)\sqrt{\frac{(H^*-e\phi^*)^2}{c^2}-\left(P_x-eA_x^*\right)^2-\left(P_y-eA_y^*\right)^2-m^2c^2}, \tag{2.6.23}$$

so that the new Hamiltonian $K=-P_s$ in the Frenet–Serret coordinate system, with s as the independent variable, is

$$K=-P_s=\frac{Q_x\kappa-1}{c}\sqrt{(H-e\phi)^2-c^2\left(P_x-eA_x\right)^2-c^2\left(P_y-eA_y\right)^2-m^2c^4}-eA_s, \tag{2.6.24}$$

where we have – for the sake of simpler notation – replaced all starred symbols by their un-starred variants. All those symbols now shall refer to the respective quantities inside the Frenet–Serret coordinate system. The new canonical coordinates are Q_x, Q_y, and H, with the respective conjugate momenta P_x, P_y, and t and we write

$$\boldsymbol{z}\equiv\left(Q_x,\quad P_x,\quad Q_y,\quad P_y,\quad H,\quad t\right)^{\mathrm{T}}. \tag{2.6.25}$$

For the further investigation, it is helpful to introduce some auxiliary symbols. We define the total energy E and kinetic energy T of a particle as

$$E\equiv H-e\phi\quad\text{and}\quad T\equiv E-mc^2 \tag{2.6.26}$$

and its total momentum p via

$$c^2p^2\equiv E^2-mc^2. \tag{2.6.27}$$

It needs to be stressed that the quantities introduced above are neither canonical variables, nor are they integrals of motion.

2.6.6 Typical Accelerator Physics Coordinates

While the canonical coordinates with the conjugate pairs (Q_x, P_y), (Q_y, P_y) and (H, t), which were used above, provide a complete description of the dynamics in a particle accelerator, beam dynamics is usually formulated in different sets of coordinates. Until now, the orbit $r(s)$ was an arbitrary path through configuration space. Especially, it did not necessarily fulfill any kind of equations of motion. In typical coordinates used in accelerator physics, this orbit is now considered to be the *trajectory* of what is referred to as a *reference particle*. Further, the transverse momenta are normalized, for instance to the longitudinal momentum of the individual particle or the total momentum of the reference particle, which enables the expansion of the Hamiltonian around small values of the resulting normalized momenta. A multitude of different coordinates, which can be derived from the coordinates above, are used in literature and in simulation codes. In the following, a selection of the most regularly encountered coordinates is briefly presented without rigorous derivations. In Appendix A, a less well-known set of coordinates is defined, which is applicable when the electromagnetic potentials fulfill a condition on their curl with respect to their transverse and time dependence.

2.6.6.1 Conventional Coordinates

In textbooks [54, 55, 59], regularly the coordinates

$$\left(x \equiv Q_x,\ p_x \equiv \frac{P_x}{P_0},\ y \equiv Q_y,\ p_y \equiv \frac{P_x}{P_0},\ \tau \equiv \frac{s}{\beta_0} - ct,\ \delta \equiv \frac{H}{cP_0} - \frac{1}{\beta_0} = \frac{H - E_0}{cP_0}\right) \tag{2.6.28}$$

are encountered, where P_0 and E_0 are the total momentum and total energy of the reference particle and it is $\beta_0 \equiv \frac{cP_0}{E_0}$. Here, τ takes the notion of the path-length distance of a particle to the reference particle and the coordinate δ is its relative energy deviation. Explicit derivation of these coordinates via a generating function shows that the Hamiltonian picks up a new term δ/β_0 due to the dependence of the τ coordinate on the independent variable s. The new Hamiltonian reads

$$\begin{aligned} K_{\text{conv}} = {} & \frac{\delta}{\beta_0} - \frac{eA_s}{P_0} \\ & + (Q_x\kappa - 1)\sqrt{\left(\frac{1}{\beta_0} + \delta - \frac{e\phi}{P_0 c}\right)^2 - \left(p_x - \frac{eA_x}{P_0}\right)^2 - \left(p_x - \frac{eA_x}{P_0}\right)^2 - \frac{1}{\beta_0^2\gamma_0^2}}. \end{aligned} \tag{2.6.29}$$

These are the coordinates used in the accelerator simulation and design program MAD8 [60].

2.6.6.2 Rose-Hoffstätter and COSY INFINITY Coordinates

Another set of useful coordinates is used by Rose and Hoffstätter [61] and in the beam physics code COSY INFINITY [62]. These coordinates are defined by

$$\left(Q_x, \quad a \equiv \frac{P_x}{P_0}, \quad Q_y, \quad b \equiv \frac{P_y}{P_0}, \quad \tau \equiv -[t - t_0(s)]\frac{T_0}{P_0}, \quad \eta \equiv \frac{T - T_0}{T_0}\right), \tag{2.6.30}$$

where T_0 is the kinetic energy of the reference particle.

2.6.6.3 Coordinates used in elegant

In the particle tracking code `elegant` the coordinates

$$\left(Q_x, \quad \frac{x'(1+\delta)}{\sqrt{1 + x'^2 + y'^2}}, \quad Q_y, \quad \frac{y'(1+\delta)}{\sqrt{1 + x'^2 + y'^2}}, \quad s \equiv \beta ct, \quad \delta \equiv \frac{P - P_0}{P_0}\right) \tag{2.6.31}$$

are used [63]. The primed variables x' and y' refer to the transverse slopes of the particle trajectories, which are defined by $x' \equiv P_x/P_s$ and $y' \equiv P_y/P_s$ and $\beta = \frac{cp}{E}$.

2.6.6.4 Unscaled Coordinates

Especially when investigating parts of the beam line in which the beam is accelerated, it can be convenient to work in the unscaled coordinates defined by

$$\left(x \equiv Q_x, \quad p_x \equiv P_x, \quad y \equiv Q_y, \quad p_y \equiv P_x, \quad z \equiv \frac{s}{\beta_0} - ct, \quad E \equiv H - E_0\right). \quad (2.6.32)$$

In these coordinates, the canonical momenta p_x, p_y, and E do not require rescaling, when the reference momentum P_0 changes.

2.6.7 Paraxial Approximation

In its complete form (2.6.29), the accelerator Hamiltonian yields intricate equations of motion, which generally cannot be solved analytically. As the normalized transverse momenta p_x, p_y, the transverse displacement x,y, and the energy deviation δ are usually small, it is justified to assume that the principle dynamics of the system are preserved if the kinematic part of the Hamiltonian is expanded to second order in these variables. This is called the *paraxial* approximation. In a section with vanishing electric potential, the Hamiltonian expanded to second order in this way, reads

$$K_{\text{conv},2} = \kappa x + \frac{1}{2}\left(p_x - \frac{eA_x}{P_0}\right)^2 + \frac{1}{2}\left(p_y - \frac{eA_y}{P_0}\right)^2 + \frac{1}{2\beta_0^2\gamma_0^2}\delta^2 + \frac{\kappa}{\beta_0}x\delta - \frac{eA_s}{P_0}. \quad (2.6.33)$$

2.6.8 Magnetic Potentials of Common Magnets

Magnetic elements typically encountered in most particle accelerators include *dipole* magnets, used for steering the beam; *quadrupole* magnets, which provide transverse focusing; and *sextupole* magnets, which provide chromatic correction. These magnets belong to a class of magnets for which a gauge fixing exists so that the scalar potential and the transverse components of the vector potential vanish: $\phi = A_x = A_y = 0$. In this case, the magnetic fields are fully determined by the longitudinal component A_s of the vector potential.

In the laboratory system, the longitudinal magnetic potential of a dipole magnet whose magnetic fields B points in the y-direction is given by

$$A_s = -Bx \text{ (upright)} \quad A_s = -By \text{ (skew)}. \quad (2.6.34)$$

It can be shown that in the Frenet–Serret frame this potential reads [55]

$$A_s^* = -B\,(1 - \kappa x)\left(x + \frac{1}{2}\frac{\kappa x^2}{1 - \kappa x}\right) = -B\left(x - \frac{1}{2}\kappa x^2\right). \quad (2.6.35)$$

For an upright or skew quadrupole magnet with magnetic gradient G, the potential reads

$$A_s = \frac{G}{2}(y^2 - x^2) \text{ (upright)} \quad A_s = Gyx \text{ (skew)}. \quad (2.6.36)$$

Upright sextupole magnets produce a potential of the form

$$A_s = S\left(\frac{1}{2}xy^2 - \frac{1}{6}x^3\right), \quad (2.6.37)$$

where S is the sextupole strength parameter [14]. See Figure 2.1 for a plot of the these potentials.

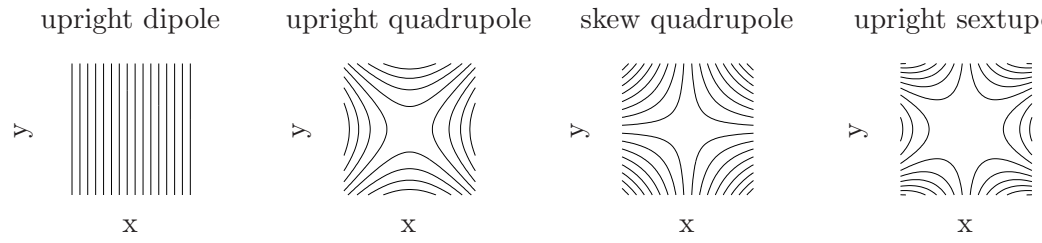

Figure 2.1: Equipotential lines of the longitudinal magnetic potential of common beamline elements.

2.6.9 Linear Maps of Common Beamline Elements

In many cases, it is expedient to consider a beamline as a sequence of elements in which the electromagnetic potentials are piecewise constant. This is known as the sharp-cutoff (SCOFF) approximation. Using this approximation allows to derive solutions of the equations of motion for each beamline element separately in form of a map. The solution of the whole beamline is then given as the composition of the maps of its constituent elements.

Assuming vanishing electric and transverse magnetic potentials, the second order Hamiltonian (2.6.33) yields the general equations of motion

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}s} = \underline{J}_6 \left(\kappa + \frac{\kappa}{\beta_0}\delta - \frac{e}{P_0}\frac{\partial A_s}{\partial x}, \quad p_x, \quad -\frac{e}{P_0}\frac{\partial A_s}{\partial y}, \quad p_y, \quad 0, \quad \frac{\kappa}{\beta_0}x + \frac{1}{\beta_0^2\gamma_0^2}\delta \right)^{\mathrm{T}}, \tag{2.6.38}$$

where $\boldsymbol{z}$ are the coordinates defined in Equation (2.6.28). It can be seen that if also A_s is expanded to second order in the transverse coordinates, Equation (2.6.38) defines an affine-linear ordinary differential equation.

2.6.9.1 Drift Space

A drift space is a straight section of the beamline in which all electromagnetic potentials vanish. With $\kappa = A_x = A_y = A_s = 0$, the second order Hamiltonian (2.6.33) yields the equations of motion

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}s} = \underline{J}_6 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 \\ 0 & +\frac{1}{\beta_0^2\gamma_0^2} \end{pmatrix} \boldsymbol{z} \tag{2.6.39}$$

$$= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & +\frac{1}{\beta_0^2\gamma_0^2} \\ 0 & 0 \end{pmatrix} \boldsymbol{z} \equiv \underline{d}(\gamma_0)\boldsymbol{z}. \tag{2.6.40}$$

As $\underline{d}(\gamma_0)$ is a second-order nilpotent matrix, the transfer map of a drift space of length l is given by

$$\boldsymbol{z} \mapsto [\underline{1}_6 + l\underline{d}(\gamma_0)]\boldsymbol{z} \equiv \underline{D}(l, \gamma_0)\boldsymbol{z}. \tag{2.6.41}$$

In the ultra-relativistic limit, $(\beta\gamma)^{-2} \to 0$, this becomes

$$\boldsymbol{z} \mapsto \left[\underline{1}_6 + l\hat{\underline{d}}\right] \boldsymbol{z} \equiv \hat{\underline{D}}(l)\boldsymbol{z}, \tag{2.6.42}$$

where

$$\hat{\underline{d}} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.6.43}$$

2.6.9.2 Upright Bend Magnet

In an upright bend magnet, the curvature is non-zero and the longitudinal magnetic potential is given by Equation (2.6.35), which reads

$$A_s = -B_0 \left(x - \frac{\kappa}{2} x^2 \right). \tag{2.6.44}$$

This yields the equations of motion

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}s} = \underline{J}_6 \left(\kappa + \frac{\kappa}{\beta_0}\delta + \frac{eB_0}{P_0}(1-\kappa x), \quad p_x, \quad 0, \quad p_y, \quad 0, \quad \frac{\kappa}{\beta_0} x + \frac{1}{\beta_0^2\gamma_0^2}\delta \right)^{\mathrm{T}}. \tag{2.6.45}$$

Note that until now the only assumption that was made about the orbit is that its torsion is zero and it lies in the x-plane. Apart from that, the choice of the orbit was completely arbitrary. From equation (2.6.45) it can be seen, that it is sensible to choose the trajectory of a particle with reference momentum P_0 as the orbit through a dipole. By choosing the curvature of the reference trajectory to be equal to the curvature of the reference particle $\kappa = -eB_0/P_0$, the equation becomes linear

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}s} = \underline{J}_6 \begin{pmatrix} \kappa^2 & 0 & 0 & 0 & 0 & \frac{\kappa}{\beta_0} \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\kappa}{\beta_0} & 0 & 0 & 0 & 0 & \frac{1}{\beta_0^2\gamma_0^2} \end{pmatrix} \boldsymbol{z} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\kappa^2 & 0 & 0 & 0 & 0 & -\frac{\kappa}{\beta_0} \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\kappa}{\beta_0} & 0 & 0 & 0 & 0 & \frac{1}{\beta_0^2\gamma_0^2} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \boldsymbol{z} \equiv \underline{b}(\gamma_0, \kappa)\boldsymbol{z}. \tag{2.6.46}$$

Hence, the transfer map of a dipole magnet is

$$\boldsymbol{z} \mapsto \exp(l\underline{b}(\gamma_0, \kappa))\boldsymbol{z} \equiv \underline{B}(\gamma_0, \kappa), \tag{2.6.47}$$

where l is the arc length of the trajectory in the magnet. The matrix $\underline{B}$ indeed can be expressed in closed form [64]:

$$\underline{B}(l, \kappa, \gamma_0) \equiv \begin{pmatrix} \cos(l\kappa) & \frac{\sin(l\kappa)}{\kappa} & 0 & 0 & 0 & \frac{1-\cos(l\kappa)}{\beta_0\kappa} \\ -\kappa\sin(l\kappa) & \cos(l\kappa) & 0 & 0 & 0 & \frac{\sin(l\kappa)}{\beta_0} \\ 0 & 0 & 1 & l & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -\frac{\sin(l\kappa)}{\beta_0} & -\frac{1-\cos(l\kappa)}{\beta_0\kappa} & 0 & 0 & 1 & \frac{l}{\beta_0^2\gamma_0^2} - \frac{1}{\beta_0}\left(l - \frac{\sin(l\kappa)}{\beta_0\kappa}\right) \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.6.48}$$

In the ultra-relativistic limit this matrix becomes

$$\hat{\underline{B}}(l, \kappa) \equiv \begin{pmatrix} \cos(l\kappa) & \frac{\sin(l\kappa)}{\kappa} & 0 & 0 & 0 & \frac{1-\cos(l\kappa)}{\kappa} \\ -\kappa\sin(l\kappa) & \cos(l\kappa) & 0 & 0 & 0 & \sin(l\kappa) \\ 0 & 0 & 1 & l & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -\sin(l\kappa) & -\frac{1-\cos(l\kappa)}{\kappa} & 0 & 0 & 1 & \frac{\sin(l\kappa)}{\kappa} - l \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.6.49}$$

In some cases, it is however more appropriate to treat a dipole in the so-called *thin-lens* approximation. In this approximation, the bend radius $R = \kappa^{-1}$ is considered to be much larger than the path length of the orbit in the dipole magnet l. This approximation

can be constructed by taking the limit $l \to 0$ of $l\underline{b}$, while keeping terms proportional to $l\kappa$ constant. The resulting matrix

$$l\underline{b}_{\mathrm{tl}} \equiv \lim_{\substack{l\to 0 \\ l\kappa=\mathrm{const.}}} l\underline{b} = l \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -\kappa^2 & 0 & 0 & 0 & 0 & -\frac{\kappa}{\beta_0} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\kappa}{\beta_0} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.6.50}$$

is second-order nilpotent, so that the thin-lens map for a dipole magnet is

$$\boldsymbol{z} \mapsto [\underline{1}_6 + l\underline{b}_{\mathrm{tl}}]\boldsymbol{z} \equiv \underline{B}_{\mathrm{tl}}(\beta_0, \kappa, l)\boldsymbol{z}. \tag{2.6.51}$$

If the bending radius is much larger than the path length in the magnet, we can approximate $l\kappa^2 \approx 0$ and set $l\kappa = \alpha$, where α is the bending angle of the magnet. In this small-angle approximation, the transfer map is

$$\boldsymbol{z} \mapsto [\underline{1}_6 + \alpha\beta_0^{-1}\hat{\underline{b}}_{\mathrm{tl}}]\boldsymbol{z} \equiv \hat{\underline{B}}_{\mathrm{tl}}(\alpha\beta_0^{-1})\boldsymbol{z}, \tag{2.6.52}$$

where

$$\hat{\underline{b}}_{\mathrm{tl}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.6.53}$$

2.6.9.3 Edge Focusing

If the the orbit is not perpendicular to the pole face of a dipole magnet, electrons on the side towards which the pole face is inclined are deflected stronger than those on the opposite side, as they traverse more of the magnetic field. In addition, due to the magnetic field lines protruding from the pole face, an inclination of the pole face causes a horizontal field component to appear, which focuses or defocuses the beam in the vertical plane. This effect is called *edge focusing* [54]. In linear approximation, the transfer map of edge focusing is

$$\boldsymbol{z} \mapsto \begin{pmatrix} 1 & 0 \\ \kappa\tan(\alpha) & 1 \end{pmatrix} \oplus \begin{pmatrix} 1 & 0 \\ -\kappa\tan(\alpha) & 1 \end{pmatrix} \oplus \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{z} \equiv \underline{F}(\kappa, \alpha)\boldsymbol{z}, \tag{2.6.54}$$

where α is the angle between the orbit and the pole face.

2.6.9.4 Upright Quadrupole

The longitudinal vector potential of an upright quadrupole magnet is given in Equation (2.6.36), which yields the equations of motion

$$\frac{\mathrm{d}\boldsymbol{z}}{\mathrm{d}s} = \underline{J}_6 \left(-\frac{eG}{P_0}x, \quad p_x, \quad +\frac{eG}{P_0}y, \quad p_y, \quad 0, \quad \frac{1}{\beta_0^2\gamma_0^2}\delta \right)^{\mathrm{T}} \tag{2.6.55}$$

$$= \begin{pmatrix} 0 & 1 \\ k_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ -k_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \frac{1}{\beta_0^2\gamma_0^2} \\ 0 & 0 \end{pmatrix} \boldsymbol{z} \equiv \underline{q}(\gamma_0, k_1)\boldsymbol{z}, \tag{2.6.56}$$

where we introduced the *quadrupole-strength* parameter $k_1 \equiv eG/P_0$. The exact solution of Equation (2.6.56) for a thick quadrupole of length l is

$$\boldsymbol{z} \mapsto \underline{\hat{R}}_2\left(l, \sqrt{k_1}\right) \oplus \underline{\hat{R}}_2\left(l, \sqrt{-k_1}\right) \oplus \begin{pmatrix} 1 & \frac{l}{\beta_0^2\gamma_0^2} \\ 0 & 1 \end{pmatrix} \boldsymbol{z} \equiv \underline{Q}(\gamma_0, k_1, l)\boldsymbol{z}, \tag{2.6.57}$$

where $\underline{\hat{R}}_2 : \mathbb{R} \times \mathbb{C} \to \mathbb{R}^{2\times 2}$ is the scaled rotation matrix

$$\underline{\hat{R}}_2(x, y) = \begin{pmatrix} \cos(xy) & \frac{1}{y}\sin(xy) \\ -y\sin(xy) & \cos(xy) \end{pmatrix}. \tag{2.6.58}$$

Also for a quadrupole, a thin-lens approximation can be constructed from Equation (2.6.56) by keeping only those terms in $\underline{q}$ that depend on the strength parameter

$$\underline{q}_{\text{tl}} \equiv \begin{pmatrix} 0 & 0 \\ k_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 \\ -k_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.6.59}$$

which yields the thin-lens map

$$\boldsymbol{z} \mapsto [\underline{1}_6 + l\underline{q}_{\text{tl}}(k_1)]\boldsymbol{z} \equiv \underline{Q}_{\text{tl}}(k_1, l)\boldsymbol{z}. \tag{2.6.60}$$

3 Single-Particle Dynamics in High-Gain FEL Injectors

High-energy electron bunches with a high current density are required to drive the radiation generation process in a high-gain free-electron laser. Accelerating electron bunches to the required energy, compressing them to the required current density and measuring and controlling the bunch quality, is the task of a linear accelerator section upstream of the FEL undulator beamlines, commonly referred to as the *injector*. The injector is to a large extent responsible for the quality of the bunches in the undulator.

Acceleration is achieved by means of radio-frequency (RF) cavities. Within these cavities, an electron gains or loses energy due to the interaction with an electro-magnetic wave, depending on the phase of the wave. In addition to the overall acceleration of the bunch, RF cavities are used to imprint a correlation between the longitudinal position of an electron within the bunch and its energy. This so-called *energy chirp*, or *correlated energy spread*, is necessary to compress the beam in magnetic chicanes.

Magnetic chicanes consist of a series of dipole magnets that deflect the beam so that the orbit in the chicane deviates from a straight. Typical choices for the chicane geometry include C-type chicanes, see Figure 3.1, and less commonly S-type chicanes, in which the orbit resembles the shape of the letters C and S, respectively. The working principle of a magnetic chicane is based on the fact that in a bending magnet a particle with a lower energy is deflected stronger than a particle with higher energy. As a result, the path length of the trajectory of a low-energy particle through the chicane is longer than that of a higher-energy particle. This dependence of the path length on the energy is called *longitudinal dispersion*. Together with the chirp imprinted by the RF cavities the longitudinal dispersion generated in the magnetic chicanes results in a compression of the bunch. Bunch compression is crucial for the operation of an FEL, as the maximum bunch current that can be produced directly by the electron source – while preserving a good bunch quality – is limited.

As described in Section 1, compressing the bunch to its final current density in a single strong chicane is usually not viable, due to degrading effects caused by the consequently intense synchrotron radiation. Many facilities therefore implement a *staged* bunch compression scheme, where multiple compression stages consisting of an acceleration section and a subsequent chicane are staged one after another.

Typically, the magnetic lattice of an FEL injector is designed without coupling between the dynamics in the two transverse planes. Non-linear magnetic elements, such as for instance sextupoles, are typically avoided and are therefore not discussed here. Hence, the principle transverse dynamics can be described to good approximation by the optical functions defined in the linear transverse uncoupled theory.

In the following, we describe the function of the different beamline elements. The results are used to motivate a one-dimensional ultra-relativistic model of the longitudinal dynamics, which is used to investigate the microbunching instability.

3.1 Radio-Frequency Cavity

When a particle traverses an electric field with a component parallel to its direction of motion, its kinetic energy changes. In particular, an electron – being a negatively charged particle – experiences a positive energy gain when it traverses an electric field that is anti-parallel to its direction of motion. Hence, electric fields anti-parallel to the longitudinal direction of the orbit are required to accelerate electron bunches in an FEL injector. From first principles it can be seen that the maximum gradient achievable in electrostatic fields is limited due to electrical breakdown effects. Higher gradients can be achieved by employing time-dependent fields. These can be generated in resonant radio-frequency (RF) cavities [54].

The amount of energy a particle gains in an RF-cavity depends on the amplitude and the phase ϕ of the RF-field when the particle enters the cavity. As this phase varies in time, the energy gain depends on the arrival time of the particle at the cavity. Therefore, the energy gain can be modeled by

$$\Delta E = eV_{\mathrm{cav}} \cos(k_{\mathrm{cav}}\tau + \phi_0), \tag{3.1.1}$$

where V_{cav} is the effective cavity voltage, the wave number k_{cav} is given via the cavity frequency $k_{\mathrm{cav}} = 2\pi f_{cav}/c$, and ϕ_0 is the phase of the rf-field when the reference particle enters. Using the unscaled coordinates (2.6.32), the transfer map of a rf-cavity is then given by

$$\begin{pmatrix} z \\ E \end{pmatrix} \mapsto \begin{pmatrix} z \\ E + eV_{\mathrm{cav}}[\cos(k_{\mathrm{cav}} z + \phi_0) - \cos(\phi_0)] \end{pmatrix} \tag{3.1.2}$$

and the reference energy changes according to

$$E_0 \mapsto E_0 + \cos(\phi_0). \tag{3.1.3}$$

For simplicity, this model neglects the fact that the amplitude of the rf-field depends on the transverse position and omits the treatment of rf-wakefields. In the linear approximation this map reads

$$\begin{pmatrix} z \\ E \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ -eV_{\mathrm{cav}} \sin(\phi_0) & 1 \end{pmatrix} \begin{pmatrix} z \\ E \end{pmatrix}. \tag{3.1.4}$$

In the transverse planes, an rf-cavity has a focusing effect [65, 66], which has to be taken into account when calculating the optical functions of the lattice. In fact, the electro-magnetic wave in an rf-cavity is not a plane wave, so that its amplitude depends on the distance to the central axis of the cavity. In principle, the energy gain of a particle therefore depends on its transverse position when traversing the cavity. However, typically the transverse extent of the electron bunch is sufficiently small compared to the transverse scale of the rf-field so that the variation of the rf-amplitude within the bunch is negligible.

3.2 Drift Space

The longitudinal component of the transfer map of a drift space of length l is given by

$$D\colon \begin{pmatrix} z \\ E \end{pmatrix} \mapsto \begin{pmatrix} 1 & \frac{l}{\beta_0^2\gamma_0^2 E_0} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z \\ E \end{pmatrix}, \tag{3.2.1}$$

which in the ultra-relativistic limit $\gamma_0 \to \infty$ approaches the identity map. When analyzing the microbunching instability, we investigate phase-space densities with a modulation of the energy coordinate

$$\Psi_0(z, E) = \alpha f\left(E - A\cos\left(2\pi z/\lambda\right)\right), \tag{3.2.2}$$

where $f\colon \mathbb{R} \to \mathbb{R}$ is the energy distribution function, $\alpha \in \mathbb{R}^+$ is a normalization constant and $A \in \mathbb{R}^+$ is the amplitude of the modulation at a wavelength $\lambda \in \mathbb{R}^+$. After a drift of length l the PSD is then reads

$$\Psi_1 = \mathcal{D}\Psi_0 = \alpha f(E - A\cos(2\pi z/\lambda - \Omega E)), \tag{3.2.3}$$

where we defined $\Omega \equiv \frac{2\pi l}{\beta_0^2\gamma_0^2 E_0 \lambda}$. Expanding Ψ_1 around $\Omega = 0$ gives

$$\Psi_1 = \Psi_0 - \alpha A\Omega E \sin(2\pi z/\lambda) f'(E - A\cos(2\pi z/\lambda)) + O(A^2\Omega^2). \tag{3.2.4}$$

We now consider the norm $\|\cdot\| = \int_{\mathbb{R}} |\cdot| \mathrm{d}E$ of a slice of the PSD at a fixed z position. This yields an upper bound for the relative deviation of the new PSD from the initial PSD:

$$\alpha^{-1}\|\Psi_1 - \Psi_0\| \leqslant |A\Omega| \|E f'(E - A\cos(2\pi z/\lambda))\| + O(A^2\Omega^2). \tag{3.2.5}$$

It can be seen that

$$\begin{aligned} \|E f'(E - A\cos(2\pi z/\lambda))\| &= \|(E + A\cos(2\pi z/\lambda)) f'(E)\| && (3.2.6)\\ &\leqslant \|E f'(E)\| + A\|f'(E)\|. && (3.2.7) \end{aligned}$$

In case of a Gaussian energy distribution $f(E) = 1/\sqrt{2\pi\sigma_E^2}\exp\left(-\frac{1}{2}x^2/\sigma_E^2\right)$, we get

$$\|f'(E)\| = \sqrt{\frac{2}{\pi\sigma_E^2}} \quad \text{and} \quad \|E f'(E)\| = 1. \tag{3.2.8}$$

Therefore, the modulation is approximately preserved if

$$\frac{2\pi l A}{\beta_0^2\gamma_0^2\lambda E_0}\left(\frac{A}{\sigma_E}\sqrt{\frac{2}{\pi}} + 1\right) \ll 1. \tag{3.2.9}$$

In a typical scenario the modulation amplitude is in the order of the energy spread but much less than the central energy of the beam. Estimating, for instance, $A/\sigma_E \approx 1$, $A/E_0 \approx 10^{-4}$ and $\beta_0^2\gamma_0^2 \approx 10^6$ the condition on the wavelength becomes $\lambda \gg 10^{-10} l$. Therefore, in this parameter regime, wavelengths larger than 1 nm are preserved in a 1 m drift space. As the microbunching instability typically occurs at wavelengths in the order of a few micrometer, the condition (3.2.9) is usually fulfilled for all wavelengths that could be amplified by instability. In that case, for the purpose of investigating the microbunching instability, the longitudinal single-particle transport map of a drift space can be approximated by the identity map.

3.3 C-Type Magnetic Chicane

A magnetic chicane is typically laid-out as depicted in Figure 3.1. Four dipole magnets bend the orbit of the beam into a C-shaped trajectory. As the bending angle of a particle depends on its energy, also the path length a particle covers while traversing is energy dependent. To quantify this effect, we derive the linear transfer map of the chicane in this section. Special attention is paid to the $\underline{M}_{56}$ component of the transfer matrices, as it determines the coupling between the energy and the longitudinal coordinate of a particle. A more detailed discussion of this topic can be found in reference [67].

3.3.1 Thin-Lens Approximation

In the thin-lens approximation, the transfer map between the entrance of the first dipole magnet and the position a distance d after the second magnet is

$$\boldsymbol{z} \mapsto \underline{M} \equiv \underline{\hat{D}}_{\mathrm{tl}}(d)\underline{\hat{B}}_{\mathrm{tl}}(-\alpha)\underline{\hat{D}}_{\mathrm{tl}}(l)\underline{\hat{B}}_{\mathrm{tl}}(\alpha)\,\boldsymbol{z}, \tag{3.3.1}$$

where l is the length of the drift space between the two magnets. Explicit calculation yields

$$\underline{M} = \begin{pmatrix} 1 & l+d & 0 & 0 & 0 & -\alpha\, l \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & l+d & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -\alpha\, l & 0 & 0 & 1 & \alpha^2\, l \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.3.2}$$

For the transverse plane, it can be seen that this section of the chicane acts as a drift space but with an additional term coupling the energy to the displacement in the bending plane of the magnets. This effect is called *transverse dispersion* and is quantified by the term

$$\underline{M}_{16} = -\alpha l. \tag{3.3.3}$$

It can be seen that this map features coupling between the transverse and longitudinal planes. There is a non-vanishing $\underline{M}_{56}$ component given by

$$\underline{M}_{56} = \alpha^2 l. \tag{3.3.4}$$

This term means that an energy deviation will cause a shift of the longitudinal coordinate of the particle. This effect is typically referred to as *longitudinal dispersion*. In particular, the $\underline{M}_{56}$ is positive independent on the sign of the bending angle. Hence, a positive energy deviation will cause an increase in the longitudinal coordinate and vice versa. Further, there is an $\underline{M}_{52} = -\alpha l$ component, which quantifies how much an momentum deviation of a particle in the bending plane affects the longitudinal coordinate.

The transfer map of the whole chicane is given by

$$\boldsymbol{z} \mapsto \underline{M} \equiv \underline{\hat{B}}_{\mathrm{tl}}(\alpha)\,\underline{\hat{D}}_{\mathrm{tl}}(l)\,\underline{\hat{B}}_{\mathrm{tl}}(-\alpha)\,\underline{\hat{D}}_{\mathrm{tl}}(l_c)\,\underline{\hat{B}}_{\mathrm{tl}}(-\alpha)\,\underline{\hat{D}}_{\mathrm{tl}}(l)\,\underline{\hat{B}}_{\mathrm{tl}}(\alpha)\,\boldsymbol{z}, \tag{3.3.5}$$

where l_c is the length of the drift space between the inner magnets. Calculating $\underline{M}$ explicitly yields

$$\underline{M} = \begin{pmatrix} 1 & 2\,l+l_c & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2\,l+l_c & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2\,\alpha^2\, l \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.3.6}$$

As can be seen, the coupling between the transverse and longitudinal planes vanishes. This shows that a symmetric C-type chicane induces no transverse dispersion. Hence, in the thin-lens approximation, the only component that distinguishes a chicane from a drift space is the non-zero longitudinal dispersion given by

$$\underline{M}_{56} = 2\alpha^2 l. \tag{3.3.7}$$

We may add that in the non-relativistic case, additional longitudinal dispersion occurs due to velocity effects, which is quantified by the term

$$\underline{M}_{56} = \frac{2l+l_c}{\beta_0^2\gamma_0^2}. \tag{3.3.8}$$

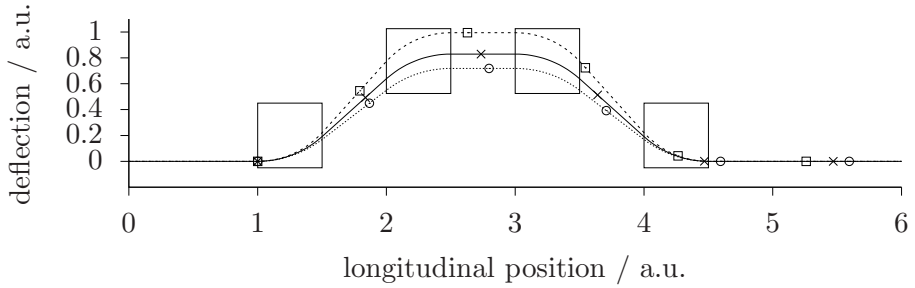

Figure 3.1: Trajectories of particles with different momenta traversing a C-shaped magnetic chicane. The bend radius of the particle with the trajectory marked with the solid line is 1.5 times the length of a bend magnet, represented by the rectangles. The momenta of the dashed and dotted trajectories differ from the solid trajectory by -10% and 10%, respectively. Markers are plotted at equidistant path lengths along each trajectory.

3.3.2 Thick Bend Magnets

Using the transfer matrix (2.6.49) for the thick bend magnets and including edge focusing effects given by (2.6.54), the transfer map of a C-shaped chicane with bending angle $\alpha \equiv s_B\kappa$ is in the ultra-relativistic limit given by

$$\begin{aligned}\boldsymbol{z} \mapsto \underline{M}\boldsymbol{z} \equiv &\underline{\hat{B}}(s_B,\kappa)\,\underline{F}(\kappa,\alpha)\,\underline{\hat{D}}(s_D)\,\underline{F}(-\kappa,-\alpha)\,\underline{\hat{B}}(s_B,-\kappa)\,\underline{\hat{D}}(s_C)\\ &\underline{\hat{B}}(s_B,-\kappa)\,\underline{F}(-\kappa,-\alpha)\,\underline{\hat{D}}(s_D)\,\underline{F}(\kappa,\alpha)\,\underline{\hat{B}}(s_B,\kappa)\,\boldsymbol{z},\end{aligned} \tag{3.3.9}$$

where s_B, s_D, and s_C, are the path lengths of the trajectory within a bend magnet, between the first two and last two magnets, and between the inner magnets, respectively. Executing the matrix multiplication yields

$$\begin{aligned}\underline{M} = &\begin{pmatrix}1 & 4\frac{\tan(\alpha)}{\kappa} + 2\frac{s_D}{\cos(\alpha)^2} + s_C\\ 0 & 1\end{pmatrix} \oplus \begin{pmatrix}1 & 4s_B + 2s_D + s_C\\ 0 & 1\end{pmatrix}\\ &\oplus \begin{pmatrix}1 & 4\frac{\tan(\alpha)-\alpha}{\kappa} + 2s_D\tan(\alpha)^2\\ 0 & 1\end{pmatrix}.\end{aligned} \tag{3.3.10}$$

This shows that the transverse dispersion vanishes, even when the bend magnets are treated as thick elements. The longitudinal dispersion coefficient of the thick-magnet chicane is

$$\underline{M}_{56} = 4\frac{\tan(\alpha)-\alpha}{\kappa} + 2s_D\tan(\alpha)^2. \tag{3.3.11}$$

It is noteworthy that the longitudinal dispersion does not increase monotonically throughout the chicane. Figure 3.2 shows an example of the evolution of the $\underline{M}_{56}$ coefficient in dependence on the path length in a chicane of typical dimensions. As can be seen, the longitudinal dispersion decreases in the first bend magnet. From the transfer matrix of a single bend magnet (2.6.49) in the ultra-relativistic limit, it can be seen that the longitudinal dispersion of a single bend magnet is indeed negative, independent on its bending radius and length. Only in the inner two magnets the dispersion increases, with the total increase being distributed equally between the two. Finally, in the last magnet, the $\underline{M}_{56}$ decreases by the same amount as in the first magnet of the chicane. In the drift spaces between the magnets, the longitudinal dispersion is constant.

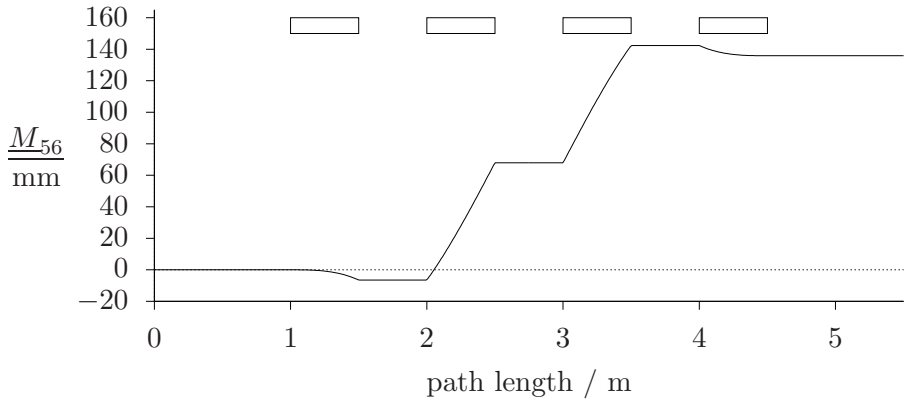

Figure 3.2: Evolution of the $\underline{M}_{56}$ through a C-shaped chicane. The bending angle is $\alpha = 16^\circ$ and the path length in the magnets as well as in all drift spaces between the magnets is 0.5 m. Positions of the magnets are indicated by the boxes at the top of the graph.

3.4 Linear Bunch Compression

After introducing the longitudinal dynamics in an rf-cavity and a magnetic chicane, we now investigate how the combination of both can be used to compress an electron bunch. To that end, we use the unscaled longitudinal coordinates $\boldsymbol{z} = (q, p)^{\mathrm{T}} \equiv (z, E)^{\mathrm{T}}$ as defined in (2.6.32).

In the linear approximation, a series of consecutive rf-cavities act as a kick map

$$\boldsymbol{z} \mapsto \underline{K}(h)\boldsymbol{z} \equiv \begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \boldsymbol{z}. \tag{3.4.1}$$

on the longitudinal coordinates. The quantity h is commonly referred to as the *chirp* induced by the cavity. It depends on the amplitude and phase settings of the cavity.

A magnetic chicane acts as a drift map

$$\boldsymbol{z} \mapsto \underline{D}(\beta)\boldsymbol{z} \equiv \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \boldsymbol{z}, \tag{3.4.2}$$

which *shears* the beam in the longitudinal direction. We define the *shearing parameter* β, which is given by the longitudinal dispersion of the chicane in conventional coordinates $\underline{M}_{56}$ normalized to the reference momentum of the beam

$$\beta \equiv \frac{M_{56}}{c\,P_0}. \tag{3.4.3}$$

Wherever confusion with the relativistic quantity $\beta_{\mathrm{rel}} \equiv v/c$ is possible, the symbol β_{rel} is used to resolve the ambiguity. The optical beta functions are denoted with β_x and β_y.

In total, the effect of a single bunch compression stage, consisting of an accelerating section followed by a magnetic chicane, is therefore given by

$$M(h,\beta)\colon \boldsymbol{z} \mapsto \underline{M}(h,\beta)\boldsymbol{z} \equiv \underline{D}(\beta)\,\underline{K}(h)\boldsymbol{z} = \begin{pmatrix} 1+h\beta & \beta \\ h & 1 \end{pmatrix} \boldsymbol{z}. \tag{3.4.4}$$

Unfortunately, in this form of the map, some important pieces of information about the dynamics are not directly accessible. In particular, the amount of compression a bunch

experiences cannot directly be inferred from $\underline{M}$. Further, the total amount of shearing, which plays an important role for the microbunching instability, is obscured. A more revealing form of the matrix can obtained by writing $\underline{M}$ as the product of a symplectic lower-triangular matrix, a symplectic diagonal matrix, and a symplectic upper-triangular matrix. It can be seen that such a symplectic LDU decomposition of $\underline{M}$ is given by

$$\underline{M}(h,\beta) = \begin{pmatrix} 1 & 0 \\ Ch & 1 \end{pmatrix} \begin{pmatrix} C^{-1} & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} 1 & C\beta \\ 0 & 1 \end{pmatrix} \tag{3.4.5}$$

$$= \underline{K}(Ch)\,\underline{S}(C)\,\underline{D}(C\beta), \tag{3.4.6}$$

where we introduced the *compression factor* $C \equiv (1+h\beta)^{-1}$, and the symplectic scaling matrix

$$\underline{S}(C) \equiv \begin{pmatrix} C^{-1} & 0 \\ 0 & C \end{pmatrix}. \tag{3.4.7}$$

From this decomposition, it becomes clear that the effect of a bunch compression stage on an initially unchirped bunch can be understood more intuitively as a three-step process:

Firstly, the bunch is sheared via the drift map given by the matrix $\underline{D}(C\beta)$. This step is particularly important for microbunching considerations, because this shearing is what drives the conversion of inhomogeneities of the energy distribution into charge-density inhomogeneities. We note that the magnitude of the shearing effect is given by the product of the longitudinal dispersion of the chicane and the compression factor. This shows that compression enhances the amount of shearing a bunch experiences.

Next, the bunch is compressed in a measure-preserving manner via the scaling map $S\colon \boldsymbol{z} \mapsto \underline{S}(C)\boldsymbol{z}$. If $|C| > 1$, this implies that the bunch becomes shorter, while simultaneously its energy distribution is broadened. It can be seen that such a symplectic scaling of an initial phase-space distribution Ψ_0 results in an increase of the local charge density by a factor C

$$\rho_1(q) = \int_{\mathbb{R}} \mathcal{S}(C)\Psi_0(q,p)\mathrm{d}p = \int_{\mathbb{R}} \Psi_0(C\,q, p/C)\mathrm{d}p \tag{3.4.8}$$

$$= C\int_{\mathbb{R}} \Psi_0(C\,q,p)\mathrm{d}p = C\rho_0(C\,q). \tag{3.4.9}$$

We see that this is the step in which the actual compression of the bunch occurs.

Lastly, an energy chirp Ch is imprinted on the bunch via the kick matrix $\underline{K}(Ch)$. For microbunching considerations this step is of no immediate importance, as it does not affect the longitudinal position of the particles, so that the charge density is conserved.

3.4.1 Initially Chirped Bunch

Consider a bunch with initial phase-space density Ψ_0 and covariance matrix

$$\Sigma_0 = \begin{pmatrix} \langle q^2\rangle & \langle qp\rangle \\ \langle qp\rangle & \langle p^2\rangle \end{pmatrix}, \tag{3.4.10}$$

where $\langle\cdot\rangle$ denotes the expected value with respect to Ψ_0. Diagonalizing this covariance matrix yields

$$\Sigma_0 = \begin{pmatrix} 1 & 0 \\ \frac{\langle qp\rangle}{\langle q^2\rangle} & 1 \end{pmatrix} \begin{pmatrix} \langle q^2\rangle & 0 \\ 0 & \langle p^2\rangle - \frac{\langle qp\rangle^2}{\langle q^2\rangle} \end{pmatrix} \begin{pmatrix} 1 & \frac{\langle qp\rangle}{\langle q^2\rangle} \\ 0 & 1 \end{pmatrix} \tag{3.4.11}$$

$$= \underline{K}(h_0)\,\mathrm{diag}\big(\langle q^2\rangle, \langle p^2\rangle - h_0\langle qp\rangle\big)\,\underline{K}(h_0)^{\mathrm{T}}, \tag{3.4.12}$$

where we have introduced the initial chirp of the bunch $h_0 \equiv \frac{\langle qp \rangle}{\langle q^2 \rangle}$. We see that the general covariance matrix of a bunch with an position-energy correlation $\langle qp \rangle$ is therefore equivalent to that of an initially uncorrelated bunch that was propagated by a linear kick map with magnitude h_0. The quantity $\langle p^2 \rangle - h_0 \langle qp \rangle$ is commonly referred to as the *uncorrelated energy spread* of the bunch. In this form it is immediately apparent how a general covariance matrix is propagated by a bunch compression stage. Using $\underline{M}$ in the form of Equation (3.4.4) ,we see that

$$\underline{\Sigma}_0 \mapsto \underline{M}(h, \beta)\, \underline{\Sigma}_0\, \underline{M}(h, \beta)^{\mathrm{T}} \tag{3.4.13}$$

$$= \underline{D}(\beta)\underline{K}(h)\underline{K}(h_0)\, \mathrm{diag}\big(\langle q^2 \rangle, \langle p^2 \rangle - h_0 \langle qp \rangle\big)\, \underline{K}(h_0)^{\mathrm{T}} \underline{K}(h)^{\mathrm{T}} \underline{D}(\beta)^{\mathrm{T}} \tag{3.4.14}$$

$$= \underline{D}(\beta)\underline{K}(h + h_0)\, \mathrm{diag}\big(\langle q^2 \rangle, \langle p^2 \rangle - h_0 \langle qp \rangle\big)\, \underline{K}(h + h_0)^{\mathrm{T}} \underline{D}(\beta)^{\mathrm{T}} \tag{3.4.15}$$

$$= \underline{M}(h + h_0, \beta)\, \mathrm{diag}\big(\langle q^2 \rangle, \langle p^2 \rangle - h_0 \langle qp \rangle\big)\, \underline{M}(h + h_0, \beta)^{\mathrm{T}}. \tag{3.4.16}$$

This shows that the final covariance matrix after a bunch compression stage can be calculated by transferring the initial chirp of the bunch to the chirp induced by the acceleration stage and propagating the uncorrelated diagonal covariance matrix via the resulting transfer matrix.

3.4.2 Two-Stage Compression

In many facilities the total required compression of the bunch is not achieved in a single compression stage, but is distributed over typically two or three stages. The transfer matrix of a two-stage setup is

$$\underline{M}_{2 \leftarrow 0} = \underline{M}(h_2, \beta_2)\underline{M}(h_1, \beta_1). \tag{3.4.17}$$

By writing the transfer matrix of the first stage in the LDU form given in Equation (3.4.6) and writing the matrix of the second stage in the form shown in Equation (3.4.4) we get

$$\underline{M}_{2 \leftarrow 0} = \underline{D}(\beta_2)\, \underline{K}(h_2)\, \underline{K}(C_1 h_1)\, \underline{S}(C_1)\, \underline{D}(C_1 \beta_1) \tag{3.4.18}$$

$$= \underline{D}(\beta_2)\, \underline{K}(h_2 + C_1 h_1)\, \underline{S}(C_1)\, \underline{D}(C_1 \beta_1) \tag{3.4.19}$$

$$= \underline{M}(h_2 + C_1 h_1, \beta_2)\, \underline{S}(C_1)\, \underline{D}(C_1 \beta_1). \tag{3.4.20}$$

By introducing the total chirp

$$h_2^* \equiv h_2 + C_1 h_1 \tag{3.4.21}$$

and the effective compression

$$C_2^* \equiv (1 + h_2^* \beta_2)^{-1} \tag{3.4.22}$$

of the second stage and employing again the LDU decomposition of $\underline{M}(h_2^*, \beta_2)$, we see that $\underline{M}_{2 \leftarrow 0}$ can be written as two pairs of drift and scale maps, followed by a single kick map

$$\underline{M}_{2 \leftarrow 0} = \underline{M}(h_2^*, \beta_2)\, \underline{S}(C_1)\, \underline{D}(C_1 \beta_1) \tag{3.4.23}$$

$$= \underline{K}(C_2^* h_2^*)\, \underline{S}(C_2^*)\, \underline{D}(C_2^* \beta_2)\, \underline{S}(C_1)\, \underline{D}(C_1 \beta_1). \tag{3.4.24}$$

Note that we introduced the notion of the *effective* compression of the second stage C_2^* to differentiate it from what could be called the *individual* compression $C_2 \equiv (1 + h_2 \beta_2)^{-1}$. The effective compression is the total compression occurring in the stage – this depends on the total chirp of the bunch, including the contribution to the chirp from the previous stage. On the other hand, the individual compression would be the compression produced by the stage, if the bunch had no initial chirp at the beginning of the stage.

It can be seen that for any two pairs of scale- and drift maps the identity

$$\underline{S}(C_2)\,\underline{D}(\beta_2)\,\underline{S}(C_1)\,\underline{D}(\beta_1) = \begin{pmatrix} \frac{1}{C_1 C_2} & \frac{\beta_1 + C_1^2 \beta_2}{C_1 C_2} \\ 0 & C_1 C_2 \end{pmatrix} = \underline{S}(C_1 C_2)\,\underline{D}(\beta_1 + C_1^2 \beta_2) \tag{3.4.25}$$

holds. Hence, we arrive at the convenient result that also a two-stage transfer map can be written in LDU-form

$$\underline{M}_{2\leftarrow 0} = \underline{K}(C_2^* h_2^*)\,\underline{S}(C_2^* C_1)\,\underline{D}(C_1\beta_1 + \beta_2 C_2^* C_1^2). \tag{3.4.26}$$

3.4.3 Multi-Stage Compression

The strategy we used to derive the two-stage transfer matrix can be generalized to an n-stage compression setup

$$\underline{M}_{n\leftarrow 0} \equiv \prod_{i=1}^{n} \underline{M}(h_i, \beta_i) = \prod_{i=3}^{n} \underline{M}(h_i, \beta_i)\underline{M}(h_2, \beta_2)\underline{M}(h_1, \beta_1) \tag{3.4.27}$$

$$= \prod_{i=3}^{n} \underline{M}(h_i, \beta_i)\underline{M}(h_2^*, \beta_2)\,\underline{S}(C_1^*)\,\underline{D}(C_1^* \beta_1) \tag{3.4.28}$$

$$= \underline{M}(h_n^*, \beta_n) \prod_{i=1}^{n-1} \underline{S}(C_i^*)\,\underline{D}(C_i^* \beta_i) = \underline{K}(C_n^* h_n^*) \prod_{i=1}^{n} \underline{S}(C_i^*)\,\underline{D}(\beta_i^*), \tag{3.4.29}$$

where the *effective* kick-, compression- and drift-parameters of the ith stage are given by

$$h_i^* \equiv h_i + C_{i-1}^* h_{i-1}^*, \quad C_i^* \equiv (1 + h_i^* \beta_i)^{-1} \quad \text{and} \quad \beta_i^* = C_i^* \beta_i \tag{3.4.30}$$

respectively, for $i > 0$. By defining $C_0 \equiv 1$, the initial chirp of the bunch before the first compression stage is given by h_0^*. The scale- and drift-maps in the product can be concatenated pairwise as shown before. By introducing the quantities

$$h_n^\dagger = C_n^* h_n^*, \quad C_n^\dagger \equiv C_n^* C_{n-1}^\dagger = \prod_{i=1}^{n} C_i^*, \quad \text{and} \quad \beta_n^\dagger \equiv \beta_{n-1}^\dagger + C_{n-1}^{\dagger\,2} \beta_n^*, \tag{3.4.31}$$

which define the *total* kick-, compression and drift-parameters, the product can be written as a single scale- and drift-map:

$$\underline{M}_{n\leftarrow 0} = \underline{K}\left(h_n^\dagger\right)\,\underline{S}\left(C_n^\dagger\right)\,\underline{D}\left(\beta_n^\dagger\right). \tag{3.4.32}$$

3.4.4 Dechirped Frame

We see from Equation (3.4.29) that the longitudinal transfer matrix of an n-stage compression setup can be expressed as series of n pairs of drift and scale matrices followed by a single kick matrix. For many studies, however, the total chirp of the bunch, which is induced by this final kick matrix, is not particularly relevant. In particular with regard to numerical simulations, a strong chirp can make it difficult to represent a phase-space density efficiently on a numerical grid. Therefore, it can be useful to study the longitudinal dynamics using the *dechirped frame*. In this frame, the linear chirp after each stage is removed by concatenating a compensating kick matrix to the transfer map

$$\bar{\underline{M}}_{n\leftarrow 0} \equiv \underline{K}(-h_n^\dagger)\underline{M}_{n\leftarrow 0} = \underline{S}(C_n^\dagger)\,\underline{D}(\beta_n^\dagger). \tag{3.4.33}$$

In this dechirped frame, the transfer map between any two stages is just a series of drift and scale matrices

$$\underline{\bar{M}}_{m\leftarrow n} = \underline{\bar{M}}_{m\leftarrow 0}\underline{\bar{M}}^{-1}_{n\leftarrow 0} = \prod_{i=n+1}^{m} \underline{S}(C_i^*)\,\underline{D}(C_i^*\beta_i), \tag{3.4.34}$$

which again could be concatenated into a single scale-drift pair as shown before.

3.5 Non-Linear Longitudinal Dispersion in a Magnetic Chicane

In the linear approximation, the longitudinal dispersion coefficient $\underline{M}_{56}$ can be calculated via matrix multiplication as shown in Section 3.3. However, the path length through a magnetic chicane in fact does not decrease exactly linearly with increasing particle energy. The additional non-linear components of the longitudinal dispersion play an important role, especially when large compression factors are to be achieved. Compared to the linear case, it is unfortunately much more challenging to derive the non-linear coefficients directly from the Hamiltonian formulation of the dynamics. This would require deriving higher-order maps from the non-linear Hamiltonian, while explicitly taking into account the edge geometry of the dipole magnets.

Instead, in this section we will follow a different strategy to derive the non-linear dispersion coefficients. We will derive the analytic expression for the path length of the orbit through the chicane by purely geometric means. The longitudinal dispersion coefficients can then be calculated from the derivatives of that expression with respect to the particle momentum. A similar approach was used in reference [67] to derive only the linear dispersion coefficients of a chicane. Below, we expand upon that work by showing that this approach can also be used to derive the non-linear dispersion coefficients and that those can can be conveniently expressed in form of a recursion relation. This geometric approach yields closed form expressions for the dispersion coefficients of arbitrary order.

3.5.1 Bend Magnets

Inside the homogeneous magnetic field of a dipole magnet, the trajectory of a particle describes a circle with radius

$$R(P) = \frac{P}{eB}, \tag{3.5.1}$$

where P is the momentum of the beam and B is the magnetic field strength. The path length of the part of the trajectory inside a rectangular bend magnet depends on the length l_B of the magnet, the bend radius R, and the entrance angle α, see Figure 3.3. For an arbitrary entrance angle, the exit angle can be seen to be

$$\alpha' = \operatorname{asin}\left(\frac{l_B}{R} - \sin(\alpha)\right), \tag{3.5.2}$$

so that the path length is given by

$$s_b = R[\alpha + \alpha'] = R\left[\alpha + \operatorname{asin}\left(\frac{l_B}{R} - \sin(\alpha)\right)\right]. \tag{3.5.3}$$

In particular, a particle that enters or exits the magnet perpendicular to the pole face will traverse the path length

$$s_B(R) = R\,\operatorname{asin}\left(\frac{l_B}{R}\right). \tag{3.5.4}$$

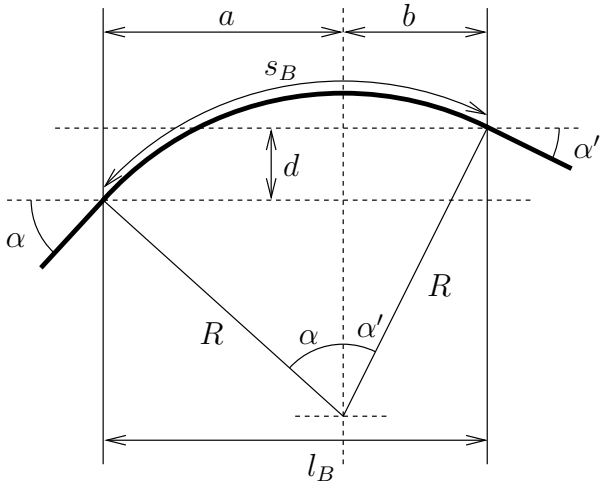

Figure 3.3: Illustration of the trajectory of a particle in a rectangular dipole magnet. Shown are the entrance and exit angles α and α', the bend radius R, the length l_B of the dipole magnet, and the path length s_B. a,b, and d are auxiliary quantities.

Due to a trigonometric function being involved, the path length depends non-linearly on the particle momentum. In order to quantify the non-linear contributions to the longitudinal dispersion function, its Taylor expansion around the nominal momentum P_0 is calculated.

For the following discussion, it is convenient to introduce the function $f(x) = x\,\mathrm{asin}(x^{-1})$. Given that $\frac{\mathrm{d}}{\mathrm{d}x}\,\mathrm{asin}(x) = (1-x^2)^{-1/2}$, the first two derivatives of f can be determined directly

$$f'(x) = \mathrm{asin}\left(x^{-1}\right) - \frac{1}{(x^2-1)^{1/2}} \quad \text{and} \quad f''(x) = \frac{1}{x(x^2-1)^{3/2}}. \tag{3.5.5}$$

For the higher-order derivatives, a recurrence relation can be derived. Consider the function family

$$F(a,b;x) = x^b\left(x^2-1\right)^{\frac{a}{2}} \tag{3.5.6}$$

where $a, b \in \mathbb{R}$. Differentiating with respect to x yields

$$\begin{aligned} F'(a,b;x) &= bF(a,b-1;x) + aF(a-2,b+1;x) && (3.5.7)\\ &= (a+b)F(a-2,b+1;x) - bF(a-2,b-1;x). && (3.5.8) \end{aligned}$$

The nth derivative $F^{(n)}$ is therefore given by the recurrence relation

$$F^{(n)}(a,b;x) = (a+b)F^{(n-1)}(a-2,b+1;x) - bF^{(n-1)}(a-2,b-1;x), \quad n \geqslant 1. \tag{3.5.9}$$

With this, the second and higher derivatives of f can be written as

$$f^{(n)}(x) = F^{(n-2)}(-3,-1;x), \quad n \geqslant 2. \tag{3.5.10}$$

Using $s_B(R) = l_B f(R/l_B)$ and $\frac{\mathrm{d}}{\mathrm{d}P}R(P) = 1/(eB)$, the derivatives of the path length with respect to the momentum are given by

$$\frac{\mathrm{d}^n s_B(R(P))}{\mathrm{d}P^n} = \left(\left.\frac{\mathrm{d}R(P)}{\mathrm{d}P}\right|_P\right)^n \left.\frac{\mathrm{d}^n s_B(R)}{\mathrm{d}R^n}\right|_{R(P)} = \frac{l_B}{(eBl_B)^n} f^{(n)}\left(\frac{R(P)}{l_B}\right). \tag{3.5.11}$$

The Taylor series around a reference momentum P_0 then reads

$$s_B = l_B \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{P - P_0}{eBl_B} \right)^n f^{(n)} \left(R_0/l_B \right), \tag{3.5.12}$$

which can also be written in terms of the relative momentum deviation $\delta = (P - P_0)/P_0$:

$$s_B = l_B \sum_{n=0}^{\infty} \delta^n \, D_{B,n}(R_0/l_B), \tag{3.5.13}$$

with

$$D_{B,n}(x) \equiv \frac{x^n f^{(n)}(x)}{n!}, \tag{3.5.14}$$

where $R_0 = P_0/(eB)$ is the nominal bending radius. The nth order non-linear dispersion coefficient of a bend magnet is therefore given by $l_b \, D_{B,n}(R_0/l_b)$. See Figure 3.4 for a plot of the linear coefficient $D_{B,1}$ and Figure 3.5 for plots of the higher-order terms normalized to $D_{B,1}$.

3.5.2 Drift Sections

Given the bend angle $\alpha(R) = \text{asin}(l_B/R)$, the path length s_D in the drift section between the first and second magnets is

$$s_D(R) = \frac{l_D}{\cos(\alpha(R))} = l_D \frac{R/l_B}{\sqrt{(R/l_B)^2 - 1}} = l_D F(-1, 1, R/l_B), \tag{3.5.15}$$

where l_D is the distance between the magnets. In analogy to the derivation above, the Taylor expansion of $s_D(R(P))$ around P_0 can be seen to be

$$s_D = l_D \sum_{n=0}^{\infty} \delta^n \, D_{D,n}(R_0/l_B) \quad \text{with} \quad D_{D,n}(x) \equiv \frac{x^n F^{(n)}(-1, 1, x)}{n!}. \tag{3.5.16}$$

Plots of the linear coefficient $D_{D,1}$ and the higher-order terms normalized to $D_{B,1}$ are shown in Figure 3.4 and Figure 3.6, respectively.

3.5.3 Small Angle Limit

As can be seen in Figure 3.5 and Figure 3.6, the normalized higher-order dispersion coefficients converge to a fixed value in limit of small bending angles. In the following, these limits will be derived. Using Equation (3.5.9), it can be seen that

$$\lim_{x \to \infty} \frac{x^{n+1} F^{(n+1)}(a, b; x)}{F(a, b; x)} \tag{3.5.17}$$

$$= \lim_{x \to \infty} \left[(a + b) \frac{x^{n+1} F^{(n)}(a - 2, b + 1; x)}{F(a, b; x)} - b \frac{x^{n+1} F^{(n)}(a - 2, b - 1; x)}{F(a, b; x)} \right] \tag{3.5.18}$$

$$= \lim_{x \to \infty} \left[(a + b) \frac{x^2}{x^2 - 1} \frac{x^n F^{(n)}(a - 2, b + 1; x)}{F(a - 2, b + 1; x)} - b \frac{1}{x^2 - 1} \frac{x^n F^{(n)}(a - 2, b - 1; x)}{F(a - 2, b - 1; x)} \right], \tag{3.5.19}$$

so that if the limit $\lim_{x \to \infty} x^n F^{(n)}(a, b; x)/F(a, b; x)$ exists for all a and b, also the limit $\lim_{x \to \infty} x^{n+1} F^{(n+1)}(a, b; x)/F(a, b; x)$ exists for all a and b and is given by

$$\lim_{x \to \infty} \frac{x^{n+1} F^{(n+1)}(a, b; x)}{F(a, b; x)} = (a + b) \lim_{x \to \infty} \frac{x^n F^{(n)}(a - 2, b + 1; x)}{F(a - 2, b + 1; x)}. \tag{3.5.20}$$

For $n = 0$ the limit exists and is equal to unity, so that Equation (3.5.20) holds for all $n \geqslant 0$. Applying Equation (3.5.20) n times yields

$$\lim_{x\to\infty} \frac{x^n F^{(n)}(a,b;x)}{F(a,b;x)} = \prod_{i=0}^{n-1} a + b - i. \tag{3.5.21}$$

With this, the small-angle limits of the normalized higher-order coefficients $D_{B,n}/D_{B,1}$ and $D_{D,n}/D_{D,1}$ can be evaluated. The dispersion coefficients of a bend magnet are

$$L_{B,n} \equiv \lim_{x\to\infty} \frac{D_{B,n}(x)}{D_{B,1}(x)} = \frac{1}{n!} \lim_{x\to\infty} \frac{x^{n-1} f^{(n)}(x)}{f^{(1)}(x)} \tag{3.5.22}$$

$$= \frac{1}{n!} \lim_{x\to\infty} \left[(n-1) \frac{x^{n-2} f^{(n)}(x)}{f^{(2)}(x)} + \frac{x^{n-1} f^{(n+1)}(x)}{f^{(2)}(x)} \right] \tag{3.5.23}$$

$$= \frac{1}{n!} \lim_{x\to\infty} \left[(n-1) \frac{x^{n-2} F^{(n-2)}(a,b;x)}{F(a,b;x)} + \frac{x^{n-1} F^{(n-1)}(a,b;x)}{F(a,b;x)} \right]. \tag{3.5.24}$$

Setting $a = -3$ and $b = -1$ and using Equation (3.5.21) yields

$$L_{B,n} \equiv \frac{1}{n!} \left[(n-1) \prod_{i=0}^{n-3} (-4-i) + \prod_{i=0}^{n-2} (-4-i) \right] = \frac{-3}{n!} \prod_{i=0}^{n-3} (-4-i) \tag{3.5.25}$$

$$= \frac{(-1)^{n+1}(n+1)}{2}, \tag{3.5.26}$$

which agrees with the limits that can be estimated graphically from Figure 3.5.

For the drift sections the limits of the relative dispersion coefficients are

$$L_{D,n} \equiv \lim_{x\to\infty} \frac{D_{D,n}(x)}{D_{D,1}(x)} = \frac{1}{n!} \lim_{x\to\infty} \frac{x^{n-1} F^{(n)}(-1,1,x)}{F^{(1)}(-1,1,x)} \tag{3.5.27}$$

$$= \frac{1}{n!} \lim_{x\to\infty} \frac{x^{n-1} F^{(n-1)}(-3,0,x)}{F^{(0)}(-3,0,x)} = \frac{1}{n!} \prod_{i=0}^{n-2} (-3-i) \tag{3.5.28}$$

$$= \frac{(-1)^n (n+1)}{2}. \tag{3.5.29}$$

3.6 Phase-Space Linearization Algorithm

Usually, it is desirable that the bunch is compressed as uniformly as possible, in the sense that all parts of the bunch are compressed by the same amount. Perfectly uniform compression can only be achieved if the longitudinal transport map of the bunch-compression setup is linear. However, the transfer maps of both constituents of a bunch-compression stage – namely the RF-module and the magnetic chicane – are intrinsically non-linear. In the accelerating rf-modules, the energy gain of an electron depends sinusoidally on its longitudinal position. Magnetic chicanes introduce non-linear dispersion as described in Section 3.5. To minimize the non-linearity of the total transfer map, it is therefore necessary to choose the compression parameters of the beamline elements in a way so that their individual non-linear contributions cancel each other out as far as possible.

In the following, an algorithm is presented which calculates the appropriate RF settings for a two-stage bunch compression setup, including a higher-harmonic cavity, which

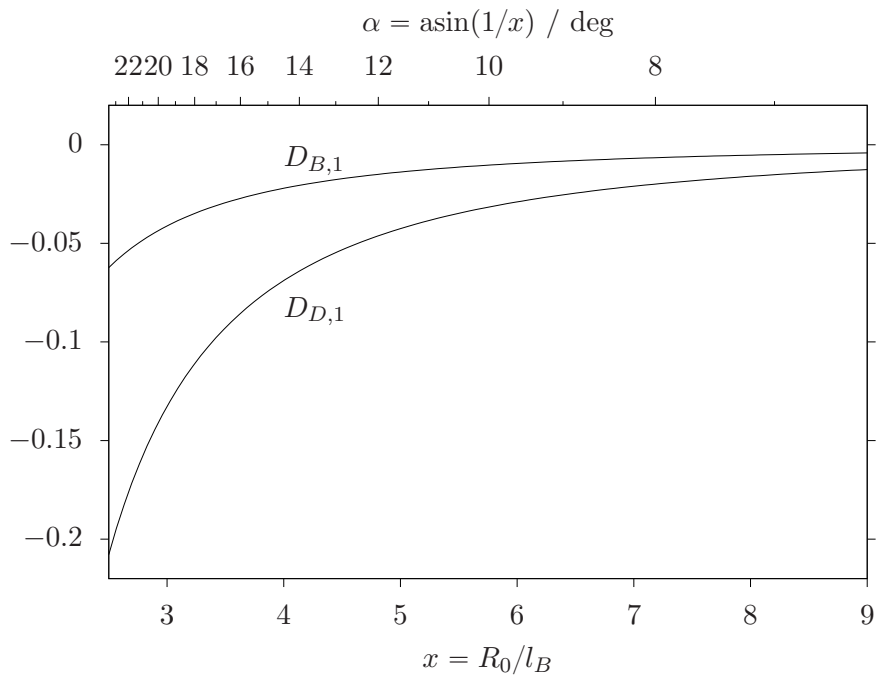

Figure 3.4: Plots of the linear dispersion coefficient of a bend magnet $D_{B,1}$ and drift sections $D_{D,1}$ of a magnetic chicane.

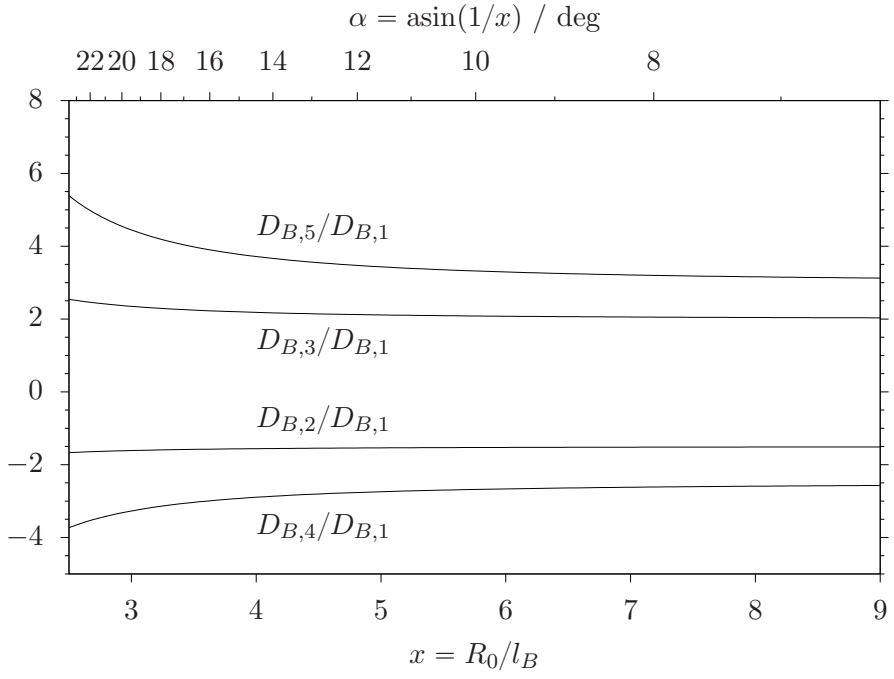

Figure 3.5: Plots of the non-linear dispersion coefficients of a bend magnet relative to the linear coefficient.

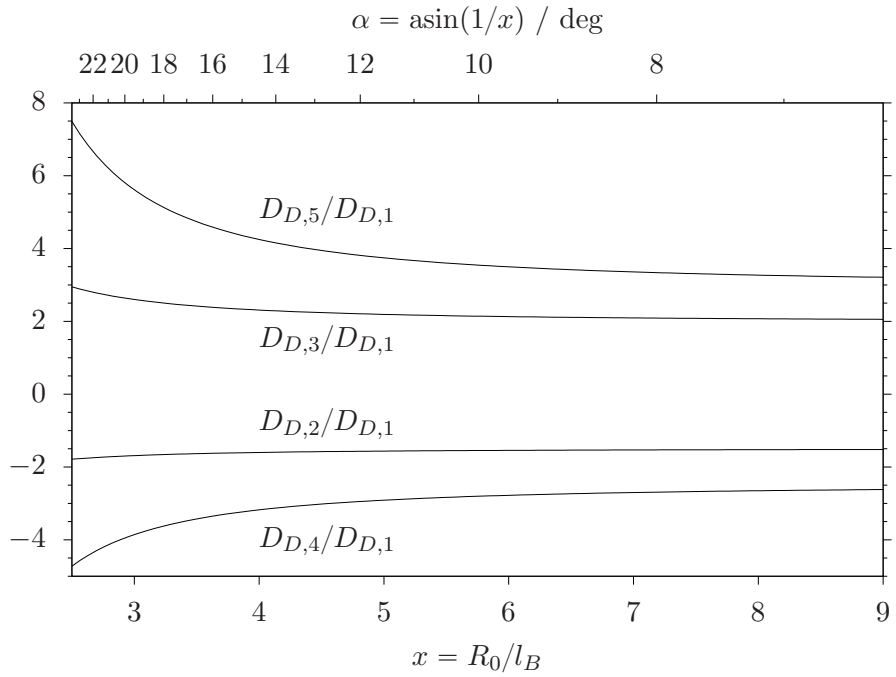

Figure 3.6: Plots of the non-linear dispersion coefficients of a drift section in a magnetic chicane relative to the linear coefficient.

minimize the total compression non-linearities. Our algorithm is derived by investigating the effect of the involved transport maps on a test phase-space density of the form

$$\Psi[\rho, f](q,p) \equiv \rho(q)\,\delta(p - f(q)), \tag{3.6.1}$$

which corresponds to a bunch with vanishing energy spread, charge density $\rho(q)$, and a local energy deviation that is given by $f(q)$. The central idea behind the algorithm is to keep f as linear as possible in the central part of the bunch, around $q = 0$. This is achieved by ensuring that the second and third derivatives of f vanish at $q = 0$. In references [68, 69], approaches with a similar objective but a different mathematical foundation are presented.

To derive our algorithm, first it needs to be understood how the energy function f and in particular its derivatives are affected by the two involved map types: kick- and drift-maps. It can be seen that applying a general kick map $K_\kappa\colon (q,p) \mapsto (q, p + \kappa(q))$ to a test density, as defined above, yields a density of the same form

$$\mathcal{K}_\kappa \Psi[\rho, f] = \rho(q)\,\delta(p - \kappa(q) - f(q)) = \Psi[\rho, f + \kappa]. \tag{3.6.2}$$

This shows that the new energy function is the sum of the old energy function and the kick function and the charge density remains unchanged.

Turning to drift maps, the derivation is less straight-forward. Applying a drift map $D_\beta\colon (q,p) \mapsto (q + \beta(p), p)$ yields

$$\mathcal{D}_\beta \Psi[\rho, f] = \rho(q - \beta(p))\,\delta(p - f(q - \beta(p))). \tag{3.6.3}$$

While not immediately obvious, it is indeed possible to bring this expression into the form shown in Equation (3.6.1). Assume, a unique function $g\colon U \to \mathbb{R}$ exists in a suitable region $U \subset \mathbb{R}$ around $q = 0$ which fulfills the condition

$$g(q) - f(q - \beta(g(q))) = 0 \tag{3.6.4}$$

and therefore also $g(q) = f(q - \beta(g(q)))$. For each longitudinal position $q \in U$, this function yields a value $g(q)$ so that the argument of the δ distribution vanishes at $p = g(q)$. Therefore, the delta distribution can be equivalently written as

$$\delta(p - f(q - \beta(p)) = \delta(p - g(q)). \tag{3.6.5}$$

The charge density of the propagated phase-space density can then be determined to be

$$\int_{\mathbb{R}} \mathcal{D}_\beta \Psi[\rho, f] \mathrm{d}p = \int_{\mathbb{R}} \rho(q - \beta(p))\, \delta(p - g(q))\, \mathrm{d}p = \rho(q - \beta(g(q))). \tag{3.6.6}$$

As a result, we find that the new phase-space density can be written as

$$\mathcal{D}_\beta \Psi[\rho, f] = \rho(q - \beta(g(q)))\, \delta(p - g(q)) = \Psi[\rho', g], \tag{3.6.7}$$

with $\rho'(q) \equiv \rho(q - \beta(g(q)))$. We note that there can be many subsets U on which solutions of Equation (3.6.4) exist and that the solution on a given subset is not necessarily unique. If the solution is not unique, the above derivation holds for each solution independently and the total phase-space density is the sum of the individual solutions. In the following, solutions of Equation (3.6.4) are reintroduced as *streaked functions*. A geometric construction of streaked functions is shown and their first three derivatives are calculated.

3.6.1 Streaked Functions

Consider a function g which is constructed from an initial function $f \in C^1(\mathbb{R}, \mathbb{R})$ and a *streak function* $\beta \in C^1(\mathbb{R}, \mathbb{R})$ so that g attains the value of the initial function $f(x)$ at the *streaked* coordinate given by $x + \beta(f(x))$:

$$g(x + \beta(f(x))) = f(x). \tag{3.6.8}$$

This construction is depicted in Figure 3.7. Conversely, it can be seen that g also satisfies

$$f(x - \beta(g(x))) = g(x), \tag{3.6.9}$$

which is equivalent to Equation (3.6.4). Hence, g is defined only implicitly and an explicit solution in general cannot be constructed. In fact, the domain of g might be restricted to certain intervals. The rigorous conditions for the existence of g are slightly beyond the scope of this investigation. For our purposes, it is sufficient to assume that $g(x)$ exists in a finite region U around $x = 0$ and consider only x so that $x + \beta(f(x))$ is within the domain of g.

It can be shown that despite the fact that g is defined only implicitly, its derivatives can be calculated explicitly nonetheless. To this end, we introduce the function

$$C(x) = \frac{1}{1 + f'(x)\beta'(f(x))}, \tag{3.6.10}$$

which will help to keep the notation concise. With this, the first derivative of g at the position $x + \beta(f(x))$ can be calculated from Equation (3.6.8) and written as

$$g'(x + \beta(f(x))) = C(x) f'(x). \tag{3.6.11}$$

As C determines the ratio of the derivative of f and the derivative of g at the translated coordinate, it might be aptly referred to as the *local compression function*. The second

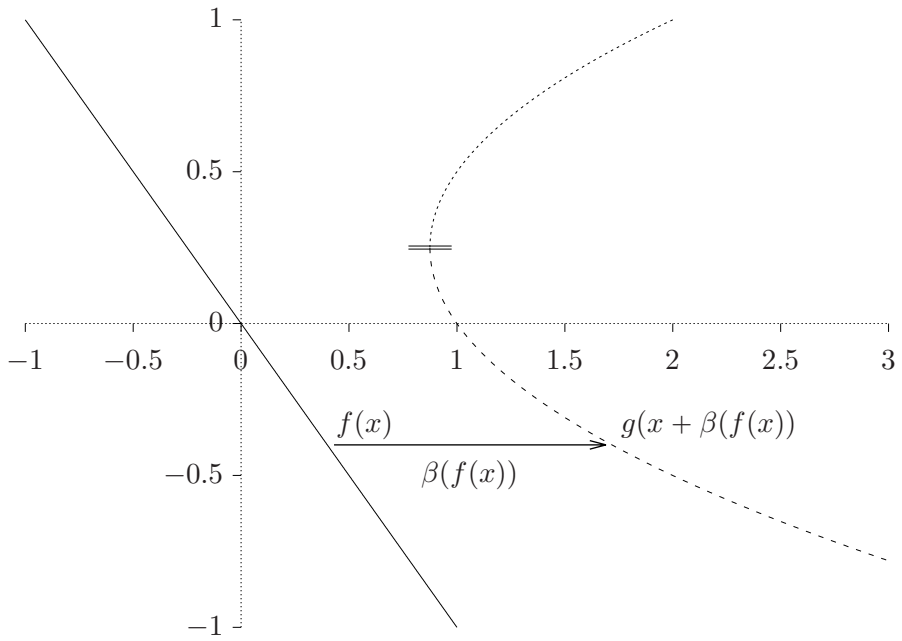

Figure 3.7: Construction of a function g (dashed) which fulfills $g(x+\beta(f(x))) = f(x)$ from an initial function f (solid) and a streaking function β. The functions used in this illustration are $f(x) = -x$ and $\beta(y) = 1 + 2y^2$. A second function that also fulfills the same condition is shown by the dotted curve.

and third derivatives can then be written in terms of the derivatives of C and f

$$g''(x+\beta(f(x))) = C\left[C'f' + Cf''\right] \tag{3.6.12}$$

$$\begin{aligned} g'''(x+\beta(f(x))) &= CC'\left[C'f' + Cf''\right] + C^2\left[C''f' + C'f'' + C'f'' + Cf'''\right] \\ &= CC'^2f' + 3C^2C'f'' + C^2C''f' + C^3f''', \end{aligned} \tag{3.6.13}$$

where the function arguments have been dropped for the sake of notational brevity. Turning to the derivatives of the compression function, we see

$$C' = -C^2\left[f''\beta' + f'^2\beta''\right] \tag{3.6.14}$$

and

$$C'' = 2\frac{C'}{C} - C^2\left[f'''\beta' + f''\beta'' + 2f'f''\beta'' + f'^2\beta'''\right], \tag{3.6.15}$$

where the arguments of the derivatives of the streak function were dropped: $\beta^{(n)} \equiv \beta^{(n)}(f(x))$. This explicitly shows that the value of g and its first three derivatives at the streaked coordinate $x-\beta(f(x))$ depend only on the values of f and β and their first three derivatives evaluated at, respectively, the initial coordinate x and $f(x)$. More generally, it becomes apparent that the nth derivative of g depends on the first n derivatives of f and β. Hence, we find that the Taylor series of g can be constructed explicitly up to arbitrary order. It is noteworthy that in general this Taylor series does not truncate at finite order, even if f and β are both of finite order.

Without loss of generality, in the following it is assumed that the coordinate system has been chosen so that at $q = 0$ the energy deviation vanishes: $f(0) = 0$. Further, it is assumed without loss of generality that the streak function β is origin-preserving: $\beta(0) = 0$, which also can be accomplished by a suitable choice of the coordinates. In that

case, g and its derivatives can be evaluated at $x = 0$, from Equations (3.6.8), (3.6.11), (3.6.12) and (3.6.13), as it is $g^{(n)}(0) = g^{(n)}(x + \beta(f(x))|_{x=0}$. This then yields the final result for the first three derivatives of g at $x = 0$:

$$\begin{aligned} g(0) &= f(0) \\ g'(0) &= \left(Cf'\right)\big|_{x=0} \\ g''(0) &= \left(CC'f' + C^2f''\right)\big|_{x=0} \\ g'''(0) &= \left(CC'^2f' + 3C^2C'f'' + C^2C''f' + C^3f'''\right)\big|_{x=0}. \end{aligned} \tag{3.6.16}$$

3.6.2 Inversion of RF Curvature Coefficients

As shown in Equation (3.6.2), applying a kick map to a test phase-space density as defined in Equation (3.6.1) simply adds the underlying kick function to the energy function. Kick maps therefore provide the means to manipulate the derivatives of the energy function directly. In bunch-compression systems, the energy kicks are generated by accelerating cavity modules, which, due to their operating principle, inherently produce sinusoidal energy kicks. The change of the derivatives of the energy function therefore depends on the arrival time of the bunch at the module, which is typically quantified by the so-called *rf-phase* ϕ. In many FEL injectors, a cavity is installed that operates at a harmonic of the main rf-frequency. The purpose of these "linearizer" cavities, which are typically installed after the first linac section, is not to accelerate the beam, but to counteract the non-linearities in the energy distribution introduced by the accelerating cavities.

In order to make use of the rf-cavities to manipulate the energy function, it is necessary to calculate the appropriate amplitudes and phases so that the cavities produce a kick function with the requested derivatives. This inverse problem is solved in the following. First, the case of a single cavity is considered. While in the single-cavity case the inversion can also be achieved by direct calculation using trigonometric identities, we introduce here an approach that can be naturally extended to multiple cavities. This approach is then used to invert the two-cavity case, which allows to include aforementioned linearizer cavities in the algorithm.

3.6.2.1 Single Sine Function

Consider the function $f(x) = A\cos(k\,x + \phi)$, where $k \in \mathbb{R}^+$ is a fixed wave number, while the amplitude $A \in \mathbb{R}^+$ and phase $\phi \in [0, 2\pi)$ are free parameters. Given two values $h_0, h_1 \in \mathbb{R}$, the goal is to find A and ϕ so that f and its first derivative assume these values at $x = 0$: $f(0) = h_0$ and $f'(0) = h_1$. Writing f' explicitly, this yields the non-linear system of equations

$$\begin{cases} h_0 = A\cos(\phi) \\ h_1 = -kA\sin(\phi) \end{cases}. \tag{3.6.17}$$

This system can by inverted directly by algebraic means using trigonometric identities, which yields

$$\begin{cases} \phi = \operatorname{atan2}(-\frac{h_1}{k}, h_0) \\ A = \sqrt{\frac{h_1^2}{k^2} + h_0^2} \end{cases}. \tag{3.6.18}$$

Another approach is to reparameterize the system first by introducing the complex amplitudes $a = A\exp(i\phi)$ and its complex conjugate $\bar{a} = A\exp(-i\phi)$, which allows to

rewrite the system (3.6.17) as the linear equation

$$\begin{pmatrix} h_0 \\ h_1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ ik & -ik \end{pmatrix} \begin{pmatrix} a \\ \bar{a} \end{pmatrix}. \tag{3.6.19}$$

Then, a solution for a and $\bar{a}$ can be obtained by matrix inversion, which yields

$$\begin{pmatrix} a \\ \bar{a} \end{pmatrix} = \frac{1}{-ik} \begin{pmatrix} -ik & -1 \\ -ik & 1 \end{pmatrix} \begin{pmatrix} h_0 \\ h_1 \end{pmatrix} = \begin{pmatrix} 1 & -i\frac{1}{k} \\ 1 & i\frac{1}{k} \end{pmatrix} \begin{pmatrix} h_0 \\ h_1 \end{pmatrix}, \tag{3.6.20}$$

so that the complex amplitude is given by

$$a = h_0 - i\frac{h_1}{k}. \tag{3.6.21}$$

Extracting the amplitude and phase from a shows that this solution is equivalent to Equation (3.6.18)

$$\begin{cases} \phi = \arg(a) = \operatorname{atan2}(\Im(a), \Re(a)) = \operatorname{atan2}(-\frac{h_1}{k}, h_0) \\ A = |a| = \sqrt{\frac{h_1^2}{k^2} + h_0^2} \end{cases}. \tag{3.6.22}$$

3.6.2.2 Double Sine Functions

Now consider the sum of two sine functions

$$f(x) = A_1 \cos(k_1 x + \phi_1) + A_2 \cos(k_2 x + \phi_2), \tag{3.6.23}$$

with amplitudes $A_1, A_2 \in \mathbb{R}^+$, phases $\phi_1, \phi_2 \in [0, 2\pi)$, and fixed wave numbers $k_1, k_2 \in \mathbb{R}^+$. This function has the four free parameters A_1, A_2, ϕ_1, and ϕ_2. Again, the goal is to find solutions for these parameters so that f and its first three derivatives assume certain values $h_0, h_1, h_2, h_3 \in \mathbb{R}$ at $x = 0$:

$$\begin{cases} h_0 = f^{(0)}(0) = & A_1 \cos(\phi_1) + & A_2 \cos(\phi_2) \\ h_1 = f^{(1)}(0) = & -k_1 A_1 \sin(\phi_1) - & k_2 A_2 \sin(\phi_2) \\ h_2 = f^{(2)}(0) = & -k_1^2 A_1 \cos(\phi_1) - & k_2^2 A_2 \cos(\phi_2) \\ h_3 = f^{(3)}(0) = & k_1^3 A_1 \sin(\phi_1) + & k_2^3 A_2 \sin(\phi_2). \end{cases} \tag{3.6.24}$$

It is not immediately obvious how this system can be solved directly by trigonometric means. As before, however, it can be transformed into a linear system by rewriting it in terms of the complex amplitudes $a_n = A_n \exp(i\phi_n)$ with $n = 1, 2$ and their complex conjugates $\bar{a}_n$:

$$\begin{pmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ ik_1 & -ik_1 & ik_2 & -ik_2 \\ -k_1^2 & -k_1^2 & -k_2^2 & -k_2^2 \\ -ik_1^3 & ik_1^3 & -ik_2^3 & ik_2^3 \end{pmatrix} \begin{pmatrix} a_1 \\ \bar{a}_1 \\ a_2 \\ \bar{a}_2 \end{pmatrix}. \tag{3.6.25}$$

After inverting the matrix we get

$$\begin{pmatrix} a_1 \\ \bar{a}_1 \\ a_2 \\ \bar{a}_2 \end{pmatrix} = \frac{1}{k_2^2 - k_1^2} \begin{pmatrix} k_2^2 & -i\frac{k_2^2}{k_1} & 1 & -i\frac{1}{k_1} \\ k_2^2 & i\frac{k_2^2}{k_1} & 1 & i\frac{1}{k_1} \\ -k_1^2 & i\frac{k_1^2}{k_2} & -1 & i\frac{1}{k_2} \\ -k_1^2 & -i\frac{k_1^2}{k_2} & -1 & -i\frac{1}{k_2} \end{pmatrix} \begin{pmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \end{pmatrix}, \tag{3.6.26}$$

which yields the solution for the complex amplitudes

$$\begin{cases} a_1 = \dfrac{k_2^2 h_0 + h_2}{k_2^2 - k_1^2} - i\dfrac{k_2^2 h_1 + h_3}{k_1(k_2^2 - k_1^2)} \\ a_2 = \dfrac{k_1^2 h_0 + h_2}{k_1^2 - k_2^2} - i\dfrac{k_1^2 h_1 + h_3}{k_2(k_1^2 - k_2^2)} \end{cases} . \tag{3.6.27}$$

From this, the absolute value of the amplitude and the rf-phases can be calculated equivalently to Equation (3.6.22) for both complex amplitudes independently. It becomes apparent that – as one might have intuitively expected – a solution exists only if the frequencies of the two cavities are unequal.

3.6.3 Calculation Procedure

With the mathematical tools developed above, it is now possible to specify the computation procedure to calculate the rf-parameters. We will assume that the bunch compression setup consist of a cavity module doublet, a chicane, a single cavity module, and a second chicane, in that order.

In the following, we will refer to the value of the energy function and its first three initial derivatives as the *curvature coefficients* and denote them by

$$\boldsymbol{b} = (E, h, h', h'') \equiv (f(0), f'(0), f''(0), f'''(0)). \tag{3.6.28}$$

For the procedure, we specify the initial curvature coefficients $\boldsymbol{b}_0$ and the final nonlinear coefficients h_2', h_2'' after the second chicane, as the input beam parameters.

The input machine parameters for the procedure are: target beam energies E_1, E_2 at the two chicanes and the respective compression factors C_1, C_2; frequencies of the first cavity k_1, linearizer cavity k_h, and second cavity k_2; plus the layout of the chicanes: magnet lengths, drift lengths, deflection angles.

In the first step of the procedure, the first three dispersion coefficients are calculated from the chicane parameters as described in Section 3.5 for both chicanes. These correspond to the derivatives of the streak function as defined in Section 3.6.1, which we will denote by $\boldsymbol{\beta} = (\beta'(0), \beta''(0), \beta'''(0))$.

By these input parameters, the required beam chirps $\hat{h}_1$ and $\hat{h}_2$ at the entrance of the two chicanes are fixed and are given by

$$\hat{h}_i = \frac{1}{(1 - \frac{1}{C_i})\beta_i'}. \tag{3.6.29}$$

Also the chirps h_i after the respective chicanes can be calculated immediately via $h_i = C_i \hat{h}_i$. With this and the fact that the final energy is equal to E_2, which was specified as an input parameter, all final curvature coefficients after the second chicane $\boldsymbol{b}_2 = (E_2, h_2, h_2', h_2'')$ are known.

With the final curvature coefficients $\boldsymbol{b}_2$ fully specified, the procedure works by tracking the coefficients backwards and solving for the required rf-parameters on the way. Using the relations given in Equation (3.6.16), the curvature coefficients $\hat{\boldsymbol{b}}_2$ at the entrance of the second chicane can be calculated from $\boldsymbol{b}_2$ using the streak coefficients $\boldsymbol{\beta}_2$, which could be notated as $\hat{\boldsymbol{b}}_2 \overset{\boldsymbol{\beta}_2}{\leftarrow} \boldsymbol{b}_2$.

As the energy and chirp before and after the second cavity are known, the required rf-parameters of the second cavity can be determined using the single-cavity inversion formula given in Equation (3.6.22): $(E_2 - E_1, h_2 - h_1) \overset{\text{sine inv.}}{\rightarrow} (A_2, \phi_2)$. Using these rf-parameters the derivatives of the kick function of the second cavity can be calculated.

As a result of Equation (3.6.2), tracking $\hat{\boldsymbol{b}}_2$ backwards through the cavity then amounts to merely subtracting the kick function derivatives, yielding the curvature coefficients at the entrance of the second cavity $\boldsymbol{b}_1$ via $\boldsymbol{b}_1 \overset{A_2,\phi_2}{\leftarrow} \hat{\boldsymbol{b}}_2$.

With the coefficients $\boldsymbol{b}_1$ known, the coefficients at the entrance of the first chicane $\hat{\boldsymbol{b}}_1 \overset{\boldsymbol{\beta}_1}{\leftarrow} \boldsymbol{b}_1$ can be calculated in the same manner as before, using Equation (3.6.16) with the chicane's streak coefficients $\boldsymbol{\beta}_1$.

Now that the coefficients after the first cavity $\hat{\boldsymbol{b}}_1$ are known, as well as the initial curvature coefficients β_0 – which were specified as input parameters – the required parameters of the cavity doublet can be determined using the two-sine inversion formula (3.6.27). This yields the final set of parameters $(\hat{\boldsymbol{b}}_1 - \boldsymbol{b}_0) \overset{\text{2-sine inv.}}{\rightarrow} (A_1, \phi_1, A_h, \phi_h)$, namely the amplitudes and phases of the first cavity and the linearizer cavity.

Appendix C contains an implementation of this algorithm. This implementation was used to derive the bunch compression working points studied in Section 6.4.

4 Collective Effects in FEL Injectors

4.1 Longitudinal Space Charge in a Relativistic Bunch

An important contributor to the collective dynamics of an electron bunch in FEL injectors is the interaction between the electrons via longitudinal space-charge forces. As the electrons are accelerated to energies that largely exceed their rest energy, it is mandatory to properly take into account relativistic effects, when determining the electric potentials within an electron bunch. In the following, we consider the electric potential of an relativistic electron bunch following a straight trajectory in the lab frame with constant velocity. As the "lab frame" we define the frame of reference in which the accelerator components are at rest. This is in contrast to the "bunch frame", which is the rest frame of a reference particle with reference energy $E_0 = \gamma\, m_\mathrm{e}\, c^2$. For the purpose of this derivation, the velocity spread within the bunch is neglected. The goal is to derive an expression for the electro-magnetic potentials in the lab frame, as this is the frame of reference in which the equation of motion are formulated. To this end, we define the four-current in the lab frame, transform it into the bunch frame, solve Maxwell's equation in the bunch frame, and finally transform the resulting potentials back into the lab frame.

Consider a bunch that travels in the direction labeled by the z coordinate. Then, assuming the charge density is not explicitly time-dependent $\partial_t \rho(c\,t, x, y, z) = 0$, the charge density $\rho(c\,t, x, y, z)$ of such a bunch in the lab frame is given by

$$\rho(c\,t, x, y, z) = \rho(0, x, y, z - \beta ct), \tag{4.1.1}$$

with the corresponding four-current

$$J^\mu(c\,t, x, y, z) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \beta \end{pmatrix} c\,\rho(c\,t, x, y, z). \tag{4.1.2}$$

Here, t denotes the time coordinate, while x, y, and z are the coordinates along an Cartesian coordinate system in the lab frame. Together they form the lab frame coordinates $x^\mu = (c\,t, x, y, z)^\mathrm{T}$. In order to transform this four-current from the lab frame into the bunch frame, a Lorentz-transform given by

$$\Lambda^\mu_\nu = \begin{pmatrix} \gamma & 0 & 0 & -\beta\,\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\,\gamma & 0 & 0 & \gamma \end{pmatrix} \tag{4.1.3}$$

has to be applied. This yields the four-current in the bunch frame in dependence on the

lab frame coordinates

$$J'^{\mu}(x^{\mu}) = \Lambda^{\mu}_{\nu} J^{\nu}(x^{\mu}) = \begin{pmatrix} \gamma - \beta^2\gamma \\ 0 \\ 0 \\ -\beta\gamma + \beta\gamma \end{pmatrix} c\,\rho(x^{\mu}) = \begin{pmatrix} 1-\beta^2 \\ 0 \\ 0 \\ 0 \end{pmatrix} c\,\gamma\,\rho(x^{\mu}) = \begin{pmatrix} c \\ 0 \\ 0 \\ 0 \end{pmatrix} \frac{1}{\gamma}\,\rho(x^{\mu}). \tag{4.1.4}$$

To solve Maxwell's equations in the bunch frame, this four-current needs to be expressed as a function of the bunch frame coordinates $x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\mu}$. Replacing the lab coordinates in Equation (4.1.4) by

$$x^{\mu} = (\Lambda^{\mu}_{\nu})^{-1}\, x'^{\mu} = \begin{pmatrix} \gamma & 0 & 0 & \beta\,\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta\,\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} c\,t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma(c\,t' + \beta\,z') \\ x' \\ y' \\ \gamma(\beta\,c\,t' + z') \end{pmatrix} \tag{4.1.5}$$

and using Equation (4.1.1), this yields for the bunch frame four-current

$$J'^{\mu}\left(c\,t', x', y', z'\right) = \begin{pmatrix} c \\ 0 \\ 0 \\ 0 \end{pmatrix} \frac{1}{\gamma}\,\rho\left(\gamma[c\,t' + \beta\,z'], x', y', \gamma[\beta\,c\,t' + z']\right) \tag{4.1.6}$$

$$= \begin{pmatrix} c \\ 0 \\ 0 \\ 0 \end{pmatrix} \frac{1}{\gamma}\,\rho\left(0, x', y', \gamma[\beta\,c\,t' + z'] - \beta\,\gamma\,[c\,t' + \beta\,z']\right) \tag{4.1.7}$$

$$= \begin{pmatrix} c \\ 0 \\ 0 \\ 0 \end{pmatrix} \frac{1}{\gamma}\,\rho\left(0, x', y', \left[1-\beta^2\right]\,\gamma\,z'\right) = \begin{pmatrix} c \\ 0 \\ 0 \\ 0 \end{pmatrix} \frac{1}{\gamma}\,\rho\left(0, x', y', \frac{z'}{\gamma}\right). \tag{4.1.8}$$

It becomes apparent that in the bunch frame, the current density vanishes and only the charge density component remains.

Maxwell's equations can now be solved in the bunch frame, which yields the proper four-potential $A'^{\mu} \equiv (\phi'/c, A'_x, A'_y, A'_z)^{\mathrm{T}}$. In Lorenz gauge, the covariant formulation of Maxwell's equations reads

$$\Box' \begin{pmatrix} \frac{\phi'}{c} \\ A'_x \\ A'_y \\ A'_z \end{pmatrix} = \mu_0 J'^{\mu} \tag{4.1.9}$$

with $\Box' \equiv \frac{1}{c^2}\partial^2_{t'} - \boldsymbol{\nabla}'^2$ and the vacuum magnetic permeability μ_0. The prime denotes that this differential operator acts on the bunch-frame coordinates. As the current density vanishes in the bunch frame, a solution for the vector potential is $\left(A'_x, A'_y, A'_z\right) = (0,0,0)$. For the scalar potential, this yields

$$\Box'\phi' = c\,\mu_0\,J'^0, \tag{4.1.10}$$

where $J'^0 = \frac{c}{\gamma}\rho(0, x', y', z'/\gamma)$ is given by Equation (4.1.8). Due to the vanishing time dependence of the charge density, a time-independent solution for the scalar potential

exists, which we will consider in the following. With $\partial_{t'}^2 \phi' = 0$ Equation (4.1.10) reduces to Poisson's equation for the scalar potential

$$-\boldsymbol{\nabla}'^2 \phi'(t', x', y', z') = \frac{1}{\epsilon_0 \gamma} \rho\left(0, x', y', \frac{z'}{\gamma}\right), \tag{4.1.11}$$

with the vacuum permittivity $\epsilon_0^{-1} \equiv \mu_0 c^2$. This shows that the source function of Poisson's equation in the bunch frame is the time-independent charge density that was elongated in a measure-preserving along the z direction by the factor γ. Poisson's equation can be solved by convolving the source function with the Green's function $G = G_{\boldsymbol{\nabla}^2} = \frac{-1}{4\pi\sqrt{x^2+y^2+z^2}}$.

$$\phi'(t', x', y', z') = -\frac{1}{\epsilon_0 \gamma} \left[G(u_1, u_2, u_3) * \rho(0, u_1, u_2, u_3/\gamma)\right](t', x', y', z'). \tag{4.1.12}$$

Convolutions of functions with modified arguments are cumbersome to formulate unambiguously. In the notation above, the auxiliary argument variables u_i are used in order to avoid ambiguity. The convolution operation is executed with respect to the u_i, which yields a function that is subsequently evaluated at (t', x', y', z').

As we want to solve the equations of motion in the lab frame, the potentials have to be transformed into the lab frame and be expressed as a function of the lab frame coordinates. Transforming the four potential via an inverse Lorenz boost along the z direction yields

$$A^\mu = \left(\Lambda^\mu_\nu\right)^{-1} A'^\nu = \begin{pmatrix} \gamma & 0 & 0 & \beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} \frac{\phi'}{c} \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \beta \end{pmatrix} \frac{\gamma}{c} \phi'. \tag{4.1.13}$$

Transforming the bunch frame coordinates back into the lab frame using

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \Lambda^\mu_\nu x'^\nu = \begin{pmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \gamma[ct - \beta z] \\ x \\ y \\ \gamma[-\beta ct + z] \end{pmatrix}, \tag{4.1.14}$$

the lab-frame potentials can be seen to be

$$A^\mu = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \beta \end{pmatrix} \frac{\gamma}{c} \phi'\left(\gamma[t - \beta z/c], x, y, \gamma[-\beta ct + z]\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \beta \end{pmatrix} \frac{\gamma}{c} \phi'\left(0, x, y, \gamma[-\beta ct + z]\right), \tag{4.1.15}$$

where the time coordinate has been set to zero, as the scalar potential is not explicitly time-dependent.

These electromagnetic potentials determine the Hamiltonian

$$H(\boldsymbol{q}, \boldsymbol{p}; t) = e\phi(\boldsymbol{q}) + c\sqrt{(\boldsymbol{p} - e\boldsymbol{A}(q))^2 + m^2 c^2}, \tag{4.1.16}$$

as already shown in Equation (2.6.1), where the canonical coordinates are $\boldsymbol{q} = (x, y, z)$ and conjugate momentum is $\boldsymbol{p} = m_e \gamma c \boldsymbol{\beta} + Z\boldsymbol{A}$. For the equation of motion of the ith momentum coordinate, this yields

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\mathrm{d}H}{\mathrm{d}q_i} = -e\frac{\mathrm{d}\phi(\boldsymbol{q})}{\mathrm{d}q_i} + c\, e\frac{\mathrm{d}\boldsymbol{A}(\boldsymbol{q})}{\mathrm{d}q_i} \cdot \frac{(\boldsymbol{p} - e\boldsymbol{A})}{\sqrt{(\boldsymbol{p} - e\,\boldsymbol{A}(\boldsymbol{q}))^2 + m^2\, c^2}}. \tag{4.1.17}$$

It can be seen that in the ultra-relativistic limit, the right-most fraction is equal to $\boldsymbol{\beta}$. In that case, it is

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -e\left[\frac{\mathrm{d}\phi(\boldsymbol{q})}{\mathrm{d}q_i} - c\,\boldsymbol{\beta}\cdot\frac{\mathrm{d}\boldsymbol{A}(\boldsymbol{q})}{\mathrm{d}q_i}\right]. \tag{4.1.18}$$

Plugging in the potentials from Equation (4.1.15) and $\boldsymbol{\beta} = (0, 0, \beta)$, this gives

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -e\,\gamma\left(1-\beta^2\right)\frac{\mathrm{d}}{\mathrm{d}q_i}\phi'(0,x,y,\gamma[z-\beta ct]) = -\frac{e}{\gamma}\frac{\mathrm{d}}{\mathrm{d}q_i}\phi'(0,x,y,\gamma[z-\beta ct]). \tag{4.1.19}$$

In particular, it is

$$\frac{\mathrm{d}p_{x,y}}{\mathrm{d}t} = -\frac{e}{\gamma}\frac{\partial\phi'}{\partial q_{x,y}}(0,x,y,\gamma[z-\beta ct]), \quad \text{and} \quad \frac{\mathrm{d}p_z}{\mathrm{d}t} = -e\frac{\partial\phi'}{\partial q_z}(0,x,y,\gamma[z-\beta ct]). \tag{4.1.20}$$

Substituting the bunch-frame potential with Equation (4.1.12) we get

$$\frac{\mathrm{d}p_{x,y}}{\mathrm{d}t} = \frac{e}{\epsilon_0\,\gamma^2}\frac{\partial}{\partial q_{x,y}}\left[G(u_1,u_2,u_3) * \rho(0,u_1,u_2,u_3/\gamma)\right](0,x,y,\gamma[z-\beta ct]) \tag{4.1.21}$$

$$\frac{\mathrm{d}p_z}{\mathrm{d}t} = \frac{e}{\epsilon_0\,\gamma}\frac{\partial}{\partial q_z}\left[G(u_1,u_2,u_3) * \rho(0,u_1,u_2,u_3/\gamma)\right](0,x,y,\gamma[z-\beta ct]). \tag{4.1.22}$$

Using the differentiation rules for convolutions, this can be written using the derivative of the Green's function

$$\frac{\mathrm{d}p_{x,y}}{\mathrm{d}t} = \frac{e}{\epsilon_0\,\gamma^2}\left[\frac{\partial G(u_1,u_2,u_3)}{\partial u_{1,2}} * \rho\left(0,u_1,u_2,\frac{u_3}{\gamma}\right)\right](0,x,y,\gamma[z-\beta ct]) \tag{4.1.23}$$

$$\frac{\mathrm{d}p_z}{\mathrm{d}t} = \frac{e}{\epsilon_0\,\gamma}\left[\frac{\partial G(u_1,u_2,u_3)}{\partial u_3} * \left(0,u_1,u_2,\frac{u_3}{\gamma}\right)\right](0,x,y,\gamma[z-\beta ct]). \tag{4.1.24}$$

In this form, the elongated charge density appears in the equations of motion. In certain cases, it can be beneficial to reformulate the equations of motion in a way so that only the unmodified charge density appears. This can be achieved by applying a theorem concerning convolutions of functions with scaled arguments. Given two functions $f, g\colon \mathbb{R}\to\mathbb{R}$, it can be seen that when scaling the argument of g by a factor α, in the sense $g(x)\to g(\alpha x) = g\circ(x\mapsto\alpha x)$, their convolution can be written as

$$f * (g\circ(x\mapsto\alpha x)) = \frac{1}{\alpha}\left\{[f\circ(x\mapsto x/\alpha)] * g\right\}\circ(x\mapsto\alpha x). \tag{4.1.25}$$

With this we see that the relativistic effect can be equivalently accounted for by contracting the argument of the Green's function

$$\frac{\mathrm{d}p_{x,y}}{\mathrm{d}t} = \frac{e}{\epsilon_0\,\gamma}\left[\frac{\partial G(u_1,u_2,\gamma\,u_3)}{\partial u_{1,2}} * \rho(0,u_1,u_2,u_3)\right](0,x,y,z-\beta ct) \tag{4.1.26}$$

$$\frac{\mathrm{d}p_z}{\mathrm{d}t} = \frac{e}{\epsilon_0}\left[\frac{\partial G(u_1,u_2,\gamma\,u_3)}{\partial u_3} * \rho(0,u_1,u_2,u_3)\right](0,x,y,z-\beta ct), \tag{4.1.27}$$

instead of elongating the charge density.

This thesis revolves around a one-dimensional model of the longitudinal beam dynamics. A one-dimensional equivalent of Equation (4.1.27) can be constructed by specifying the transverse bunch shape explicitly, which allows to remove the transverse dependence of the force term by integrating over the transverse dimensions [70,71]. Then, the convolution with respect to the transverse coordinates can be executed analytically, yielding an effective one-dimensional longitudinal Green's function $G_\parallel(z)$ so that

$$\frac{\mathrm{d}p_z}{\mathrm{d}t}(t,z) = \frac{e}{\epsilon_0}\left[G_\parallel(\gamma u) * \rho_z(u)\right](z-\beta ct), \tag{4.1.28}$$

with the line charge density $\rho_z(z) = \iint_{\mathbb{R}^2} \rho(0, x, y, z)\,\mathrm{d}x\,\mathrm{d}y$. This can be solved in Fourier space via the convolution theorem, which yields

$$\frac{\mathrm{d}p_z}{\mathrm{d}t}(t,z) = \frac{e}{\epsilon_0}\left\{\mathcal{F}^{-1}_{z\leftarrow k}\left[\widetilde{G_\parallel(\gamma u)}(k)\ \widetilde{\rho_z}(k)\right]\right\}(z-\beta ct). \tag{4.1.29}$$

As it is $\widetilde{G_\parallel(\gamma u)}(k) = \frac{1}{\gamma}\widetilde{G_\parallel(u)}\left(\frac{k}{\gamma}\right)$, we finally arrive at

$$\frac{\mathrm{d}p_z}{\mathrm{d}t}(t,z) = \frac{e}{\epsilon_0\,\gamma}\left\{\mathcal{F}^{-1}_{z\leftarrow k}\left[\widetilde{G_\parallel}\left(\frac{k}{\gamma}\right)\ \widetilde{\rho_z}(k)\right]\right\}(z-\beta ct). \tag{4.1.30}$$

This shows that in the one-dimensional model, relativistic effects can be accounted for by modifying the non-relativistic Fourier-transformed Green's function – which might be known ab initio – according to

$$\widetilde{G_\parallel}(k) \mapsto \frac{1}{\gamma}\widetilde{G_\parallel}\left(\frac{k}{\gamma}\right). \tag{4.1.31}$$

4.2 Coherent Synchrotron Radiation

When relativistic charged particles travel on a bent trajectory they emit synchrotron radiation. In most parts of a free-electron lasers injector, the trajectory of the electrons is designed to be straight. Typically, the design electron orbit deviates significantly from a straight line only in the bunch compression chicanes, where it is deflected by the chicane magnets. Within the chicane magnets, the electrons travel on a circular trajectory. Due to this curved trajectory, it is possible that the synchrotron radiation emitted by the electrons closer to the *tail* – the upstream end – of the bunch can reach electrons closer to the *head* – the downstream end – of the bunch. Emitting this radiation causes the electrons to lose energy. In turn, electrons further downstream that interact with the emitted radiation can gain or lose energy in the process. Also the transverse momentum of the electrons in the bend plane can be affected by this interaction. For the purpose of this thesis, we are mainly interested in the longitudinal aspects of this effect, which is the exchange of energy that the electrons experience.

In the following, to determine the energy exchange between a source and a observer electron via this radiative interaction, first the radiation fields generated by the source electron are calculated using the Liénard-Wiechert fields [15–17, 72–76]. The rate of change of the energy of the observer electron can then be determined by examining its motion relative to these fields. The derivation presented here is heavily inspired by reference [75], but deviates from that guideline in certain aspects. In particular, we will postpone certain approximations until the end of the derivation and consider only the interaction of two particles, which travel on a common circular trajectory within the magnet. In reference [75], also the interaction between electrons upstream and downstream of the magnet with those inside the magnet is considered. A more general approach is presented in reference [76], which allows to calculate the radiative interaction of electrons following an arbitrary trajectory.

The electric Liénard-Wiechert field observed at a position $\boldsymbol{r}$, caused by a source electron that follows the trajectory $\boldsymbol{r}_s(t)$, can be split into two parts

$$\boldsymbol{E}(\boldsymbol{r}_s, \boldsymbol{r}, t) = \boldsymbol{E}_{\mathrm{rad}}(\boldsymbol{r}, t_r) + \boldsymbol{E}_{\mathrm{stat}}(\boldsymbol{r}, t_r), \tag{4.2.1}$$

where $\boldsymbol{E}_{\mathrm{rad}}$ is the *radiation* term, which encapsulates the contribution of the emitted radiation, and $\boldsymbol{E}_{\mathrm{stat}}$ is the *static* term, which describes the contribution of the Coulomb

force. Both terms need to be evaluated at the *retarded time* t_r. At a given observation time t, the retarded time t_r at an observation point $\boldsymbol{r}$ is defined as the time at which the source electron was separated from the observation point by a distance equal to the difference between the observation time and the retarded time times the speed of light

$$|\boldsymbol{r} - \boldsymbol{r}_s(t_r)| = c\,(t - t_r)\,. \tag{4.2.2}$$

The solution of this implicit equation depends on the actual source trajectory. In the following, we will consider a source electron moving on a circular trajectory and evaluate the fields at observation points along the same trajectory.

4.2.1 Radiation Term

The radiation term reads

$$\boldsymbol{E}_{\text{rad}} = \frac{-e}{4\pi\,\epsilon_0\,c}\frac{\boldsymbol{n}\times\left[(\boldsymbol{n}-\boldsymbol{\beta}_r)\times\dot{\boldsymbol{\beta}}_r\right]}{(1-\boldsymbol{n}\cdot\boldsymbol{\beta}_r)^3\,|\boldsymbol{r}-\boldsymbol{r}_r|}, \tag{4.2.3}$$

where $\boldsymbol{r}_r \equiv \boldsymbol{r}_s(t_r)$ is the position of the source electron at the retarded time, $\boldsymbol{\beta}_r \equiv \boldsymbol{\beta}_s(t_r)$ is the velocity vector $\boldsymbol{\beta}_s(t) \equiv \frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{r}_s(t)$ at the retarded time with the time derivative $\dot{\boldsymbol{\beta}}_r \equiv \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\beta}_r$, and $\boldsymbol{n} \equiv \frac{\boldsymbol{r}-\boldsymbol{r}_r}{|\boldsymbol{r}-\boldsymbol{r}_r|}$ is the normalized vector pointing from the observation point to the source [72]. The rate of change of the energy of the observer electron due to the interaction with the radiation field is given by

$$\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{\text{rad}} = -e\,c\,\boldsymbol{\beta}\cdot\boldsymbol{E}_{\text{rad}}, \tag{4.2.4}$$

where $\boldsymbol{\beta}$ is the velocity vector of the observer electron. Note that in the following a bold $\boldsymbol{E}$ denotes an electric field and E denotes the energy of an electron. Using a well-known identity for the vector triple product the numerator of Equation (4.2.3) can be written as

$$\boldsymbol{n}\times\left[(\boldsymbol{n}-\boldsymbol{\beta}_r)\times\dot{\boldsymbol{\beta}}_r\right] = (\boldsymbol{n}-\boldsymbol{\beta}_r)\,\boldsymbol{n}\cdot\dot{\boldsymbol{\beta}}_r - \dot{\boldsymbol{\beta}}_r\,\boldsymbol{n}\cdot(\boldsymbol{n}-\boldsymbol{\beta}_r). \tag{4.2.5}$$

With this, Equation (4.2.4) yields

$$\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\text{rad}} = \frac{e^2}{4\pi\,\epsilon_0}\frac{(\boldsymbol{\beta}\cdot\boldsymbol{n}-\boldsymbol{\beta}\cdot\boldsymbol{\beta}_r)(\boldsymbol{n}\cdot\dot{\boldsymbol{\beta}}_r)-(\boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}}_r)(1-\boldsymbol{n}\cdot\boldsymbol{\beta}_r)}{(1-\boldsymbol{n}\cdot\boldsymbol{\beta}_r)^3\,|\boldsymbol{r}-\boldsymbol{r}_r|}. \tag{4.2.6}$$

We will now consider a circular source trajectory and evaluate the electric fields that an electron observes which travels on the same trajectory but in front or behind the source electron. This configuration is depicted in Figure 4.1. As the electrons travel on a common circle with radius R, the arc length along that trajectory can be used to specify their positions. In Figure 4.1, s denotes the arc position of the observer electron and s_r is the position of the source electron at the retarded time. It is sensible to introduce the *retarded angle*, which is the angle between the observer electron and the source electron

$$\mu \equiv \frac{s-s_r}{R}. \tag{4.2.7}$$

Using this retarded angle, the vector products in Equation (4.2.6) can be evaluated explicitly

$$\begin{gathered}\boldsymbol{n}\cdot\boldsymbol{\beta} = \boldsymbol{n}\cdot\boldsymbol{\beta}_r = \beta\cos\left(\frac{\mu}{2}\right), \quad \boldsymbol{\beta}\cdot\boldsymbol{\beta}_r = \beta^2\cos(\mu), \quad \boldsymbol{n}\cdot\dot{\boldsymbol{\beta}}_r = \dot{\beta}\sin\left(\frac{\mu}{2}\right)\\ \boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}}_r = \beta\dot{\beta}\sin(\mu), \quad |\boldsymbol{r}-\boldsymbol{r}_r| = 2R\sin\left(\frac{\mu}{2}\right),\end{gathered} \tag{4.2.8}$$

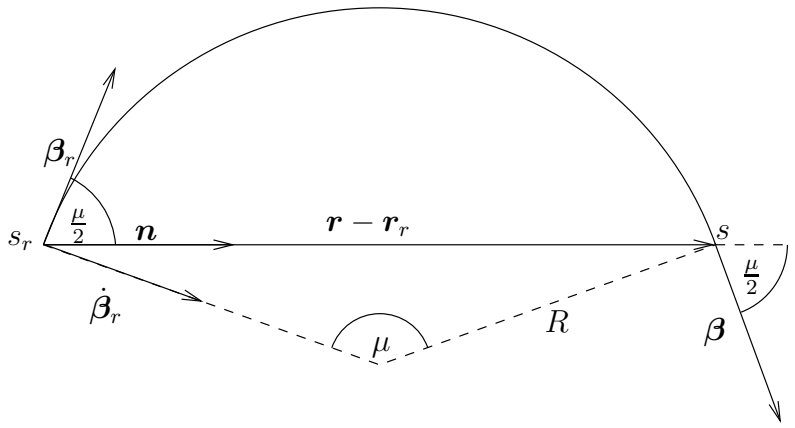

Figure 4.1: Geometric relations used in the calculation of the Liénard-Wiechert field, generated by a charged particle on a circular trajectory.

where $\beta = |\boldsymbol{\beta}| = |\boldsymbol{\beta}_r|$, $\dot{\beta} = |\dot{\boldsymbol{\beta}}| = |\dot{\boldsymbol{\beta}}_r|$, and a dot denotes differentiation with respect to t. Substituting this and applying well-known trigonometric identities, Equation (4.2.6) yields

$$\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{\mathrm{rad}} = \frac{e^2}{4\pi\,\epsilon_0}\frac{(\boldsymbol{\beta}\cdot\boldsymbol{n}-\boldsymbol{\beta}\cdot\boldsymbol{\beta}_r)(\boldsymbol{n}\cdot\dot{\boldsymbol{\beta}}_r)-(\boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}}_r)(1-\boldsymbol{n}\cdot\boldsymbol{\beta}_r)}{(1-\boldsymbol{n}\cdot\boldsymbol{\beta}_r)^3\,|\boldsymbol{r}-\boldsymbol{r}_r|} \tag{4.2.9}$$

$$= \frac{e^2}{4\pi\,\epsilon_0}\frac{\beta\dot{\beta}\left\{\left[\cos\left(\frac{\mu}{2}\right)-\beta\cos(\mu)\right]\sin\left(\frac{\mu}{2}\right)-\sin(\mu)\left[1-\beta\cos\left(\frac{\mu}{2}\right)\right]\right\}}{2R\sin\left(\frac{\mu}{2}\right)\left[1-\beta\cos\left(\frac{\mu}{2}\right)\right]^3} \tag{4.2.10}$$

$$= \frac{e^2\beta\dot{\beta}}{8\pi\,\epsilon_0 R}\,\frac{\beta-\cos\left(\frac{\mu}{2}\right)}{\left[1-\beta\cos\left(\frac{\mu}{2}\right)\right]^3}. \tag{4.2.11}$$

From first principles, it can be seen that the absolute value of the acceleration vector $\dot{\boldsymbol{\beta}}$ of a particle that travels with velocity βc on a circular trajectory with radius R is $\dot{\beta} = \frac{\beta^2 c}{R}$. With this, we finally arrive at

$$\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\mathrm{rad}} = \frac{e^2}{4\pi\,\epsilon_0\,R^2}\,\beta^3\frac{\beta-\cos\left(\frac{\mu}{2}\right)}{2\left[1-\beta\,\cos\left(\frac{\mu}{2}\right)\right]^3} = \frac{e^2}{4\pi\,\epsilon_0\,R^2}\,F_{\mathrm{rad}}\left(\beta,\frac{\mu}{2}\right) \tag{4.2.12}$$

where we define the auxiliary function

$$F_{\mathrm{rad}}(\beta,\mu) \equiv \frac{\beta-\cos(\mu)}{2\left[\frac{1}{\beta}-\cos(\mu)\right]^3}. \tag{4.2.13}$$

Of particular interest is the ultra-relativistic limit. To this end, we scale the auxiliary function by a factor of γ^{-4} and its angle argument by γ^{-1}

$$\hat{F}_{\mathrm{rad}}(\beta,\mu) = \frac{1}{\gamma^4}F_{\mathrm{rad}}\left(\beta,\frac{\mu}{\gamma}\right) \tag{4.2.14}$$

where $\gamma \equiv 1/\sqrt{1-\beta^2}$. Repeated application of L'Hospital's rule [77] shows after the third iteration that the limit $\lim_{\beta\to 1}\hat{F}_{\mathrm{rad}}$ exists and is given by

$$\lim_{\beta\to 1}\hat{F}_{\mathrm{rad}}(\beta,\mu) = 2\frac{\mu^2-1}{\left[\mu^2+1\right]^3} \equiv \Phi_{\mathrm{rad}}(\mu). \tag{4.2.15}$$

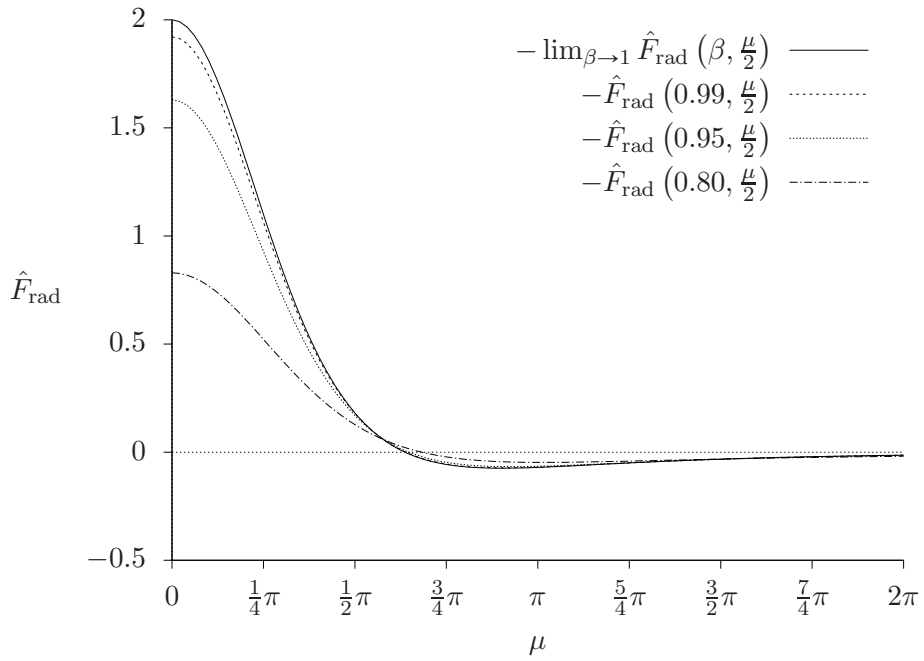

Figure 4.2: Plots of the auxiliary function $\hat{F}_{\mathrm{rad}}(\beta,\mu)$ which describes the rate of change of the energy resulting from the radiative part of the Liénard-Wiechert field, which an observer electron experiences that is separated from the source electron by the retarded angle μ.

In Figure 4.2, plots of $\hat{F}_{\mathrm{rad}}(\beta,\mu)$ are shown for multiple values of β as well as the ultrarelativistic limit.

In conclusion, we find that the rate of change of the energy of the observer electron due to radiation emitted by the source electron is given by Equation (4.2.12), which in the ultrarelativistic limit takes the form

$$\lim_{\beta\to 1}\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\mathrm{rad}} = \frac{e^2}{4\pi\epsilon_0\,R^2}\,\gamma^4\,\Phi_{\mathrm{rad}}\left(\frac{\gamma\mu}{2}\right) = \frac{e^2}{2\pi\epsilon_0\,R^2}\,\gamma^4\,\frac{\left(\frac{\gamma\mu}{2}\right)^2-1}{\left[\left(\frac{\gamma\mu}{2}\right)^2+1\right]^3}, \tag{4.2.16}$$

where μ is the retarded angle between the electrons.

4.2.2 Static Term

In addition to the radiative part, there is a static part $\boldsymbol{E}_{\mathrm{stat}}$ that contributes to the total Liénard-Wiechert field

$$\boldsymbol{E}_{\mathrm{stat}} = \frac{-e}{4\pi\,\epsilon_0\,\gamma^2}\,\frac{\boldsymbol{n}-\boldsymbol{\beta}_r}{[1-\boldsymbol{n}\cdot\boldsymbol{\beta}_r]^3\,|\boldsymbol{r}-\boldsymbol{r}_r|^2}. \tag{4.2.17}$$

It encapsulates the contribution of the electrostatic Coulomb potential of the source electron, taking into account the potentially curved trajectory as well as the retardation effect. Analogously to the radiation term, the resulting rate of change of the energy of the observer electron is

$$\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{\mathrm{stat}} = -e\,c\boldsymbol{\beta}\cdot\boldsymbol{E}_{\mathrm{stat}}. \tag{4.2.18}$$

Using again Equation (4.2.8), it can be seen that

$$\frac{\boldsymbol{\beta}\cdot[\boldsymbol{n}-\boldsymbol{\beta}_r]}{[1-\boldsymbol{n}\cdot\boldsymbol{\beta}_r]^3\,|\boldsymbol{r}-\boldsymbol{r}_r|^2}=\frac{1}{R^2}\frac{\beta\left[\cos\left(\frac{\mu}{2}\right)-\beta\cos(\mu)\right]}{4\left[1-\beta\cos\left(\frac{\mu}{2}\right)\right]^3\,\sin\left(\frac{\mu}{2}\right)^2}. \tag{4.2.19}$$

Introducing the auxiliary function

$$F_{\text{stat}}(\beta,\mu)=\frac{\beta\left[\cos(\mu)-\beta\cos(2\mu)\right]}{4\gamma^2[1-\beta\cos(\mu)]^3\,\sin(\mu)^2}, \tag{4.2.20}$$

Equation (4.2.18) can be written as

$$\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\text{stat}}=\frac{e^2}{4\pi\,\epsilon_0\,R^2}\,F_{\text{stat}}\left(\beta,\frac{\mu}{2}\right). \tag{4.2.21}$$

To investigate the ultra-relativistic limit, we again define a scaled auxiliary function

$$\hat{F}_{\text{stat}}(\beta,\mu)\equiv\frac{1}{\gamma^4}F_{\text{stat}}\left(\beta,\frac{\mu}{\gamma}\right). \tag{4.2.22}$$

The ultra-relativistic limit of this function exists and is given by

$$\lim_{\beta\to1}\hat{F}_{\text{stat}}(\beta,\mu)=\frac{3\mu^2+1}{\mu^2(\mu^2+1)^3}\equiv\Phi_{\text{stat}}(\mu). \tag{4.2.23}$$

Therefore, the contribution of the static term to the rate of change of the energy of the observer particle is in the ultra-relativistic limit

$$\lim_{\beta\to1}\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\text{stat}}=\frac{3e^2}{4\pi\epsilon_0\,R^2}\gamma^4\frac{\left(\frac{\gamma\mu}{2}\right)^2+1}{\left(\frac{\gamma\mu}{2}\right)^2\left[\left(\frac{\gamma\mu}{2}\right)^2+1\right]^3}. \tag{4.2.24}$$

4.2.3 Construction of the Retarded Angle

Above, the interaction between the source and the observer electrons moving on a common circle was derived in terms of the retarded angle between both electrons. In application, it is however more meaningful to formulate the interaction in terms of the path-length distance of the electrons along the arc. Let s and s' be the arc positions of the observer and the source electron at the observation time t, respectively, and s_r the arc position of the source electron at the retarded time t_r. As the construction in Figure 4.3 depicts, the retarded angle is given as the solution of the equation system

$$\begin{cases}\sin\left(\frac{\mu}{2}\right)=\frac{c\Delta t}{2R}\\ \mu=\frac{\beta c\Delta t}{R}+\frac{s-s'}{R}\end{cases}, \tag{4.2.25}$$

with $\Delta t\equiv t-t_r$. By solving one equation for $c\Delta t/R$ and substituting in the other, this can be seen to be equivalent to

$$\frac{s-s'}{R}=\mu-2\beta\sin\left(\frac{\mu}{2}\right). \tag{4.2.26}$$

Seeing that the limit

$$\lim_{\beta\to1}\gamma^3\left[\frac{\mu}{\gamma}-2\beta\sin\left(\frac{\mu}{2\gamma}\right)\right]=\frac{\mu}{2}+\frac{\mu^3}{24} \tag{4.2.27}$$

exists, Equation (4.2.26) yields for the retarded angle in the ultra-relativistic limit

$$\frac{s-s'}{R}=\frac{\mu}{2\,\gamma^2}+\frac{\mu^3}{24}. \tag{4.2.28}$$

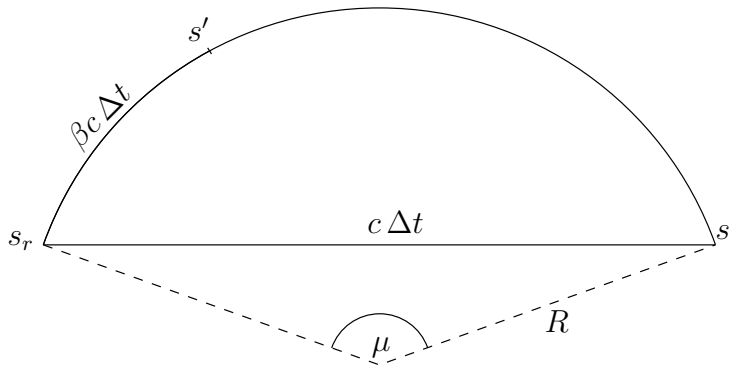

Figure 4.3: Construction of the retarded angle μ from the position s of an observer particle, the position s' of a source particle at the observation time, and the position s_r of the source particle at the retarded time. Observer and source particle both travel on the same circular trajectory.

4.2.4 Straight-Line Coulomb Renormalization

Unfortunately, the electrostatic interaction term (4.2.24) contains a non-integrable singularity at $\mu = 0$. A remedy proposed in [75] is to subtract from (4.2.24) the contribution of the Coulomb term that the observer electron at path length s would experience from the source electron at path length s', if they were traveling not on a circle, but on a straight line with the same separation $s - s'$, which is given by

$$\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\mathrm{str}} = \frac{e^2}{4\pi\epsilon_0}\frac{1}{\gamma^2\,(s-s')^2} \tag{4.2.29}$$

The separation $s - s'$ can be expressed in terms of the retarded angle using Equation (4.2.26), or, in the ultrarelativistic limit, Equation (4.2.28). In the latter case, the straight-line term reads

$$\lim_{\beta\to 1}\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\mathrm{str}} = \frac{e^2}{4\pi\epsilon_0\,R^2}\,\gamma^4\,\frac{1}{(\frac{\gamma\mu}{2}+\frac{\gamma^3\mu^3}{24})^2} = \frac{e^2}{4\pi\epsilon_0\,R^2}\,\gamma^4\,\Phi_{\mathrm{str}}\left(\frac{\gamma\mu}{2}\right), \tag{4.2.30}$$

with the auxiliary function

$$\Phi_{\mathrm{str}}\,(\mu) \equiv \frac{1}{\mu^2\left(\frac{1}{3}\mu^2+1\right)^2}. \tag{4.2.31}$$

Subtracting Equation (4.2.30) from Equation (4.2.24) yields the renormalized static interaction term

$$\left(\frac{\mathrm{d}E^*}{\mathrm{d}ct}\right)_{\mathrm{stat}} = \frac{e^2}{4\pi\epsilon_0\,R^2}\,\gamma^4\,\Phi^*_{\mathrm{stat}}\left(\frac{\gamma\mu}{2}\right), \tag{4.2.32}$$

with

$$\Phi^*_{\mathrm{stat}}(\mu) \equiv \Phi_{\mathrm{stat}}(\mu) - \Phi_{\mathrm{str}}(\mu) = \frac{1}{\mu^2}\left[\frac{3\mu^2+1}{(\mu^2+1)^3} - \frac{1}{(\frac{1}{3}\mu^2+1)^2}\right]. \tag{4.2.33}$$

4.2.5 CSR Kick Function

The total CSR induced rate of change of the observer electron's energy is given by the sum of the radiative part (4.2.24) and the renormalized electrostatic part (4.2.33)

$$\left(\frac{\mathrm{d}E}{\mathrm{d}ct}\right)_{\mathrm{CSR}} = \frac{e^2}{4\pi\epsilon_0\,R^2}\,\gamma^4\,\Phi\left(\frac{\gamma\mu}{2}\right), \tag{4.2.34}$$

with

$$\Phi(\mu) \equiv \Phi_{\text{rad}} + \Phi^*_{\text{stat}} = 2\frac{\mu^2-1}{\left[\mu^2+1\right]^3} + \frac{1}{\mu^2}\left[\frac{3\mu^2+1}{(\mu^2+1)^3} - \frac{1}{(\frac{1}{3}\mu^2+1)^2}\right]. \tag{4.2.35}$$

See Figure 4.4 for a plot of $\Phi(\mu)$.

In order to determine the total local rate of change of an observer electron at a position s, the contributions of all source electrons behind the observer have to be integrated. Electrons in front of the observer do not contribute, as their emitted radiation cannot reach the observer. Given the line charge density $\rho(s)$ of the bunch, the total differential CSR kick function is

$$\left(\frac{\mathrm{d}E}{\mathrm{d}s}\right)_{\text{CSR}}\Bigg|_s = \frac{e^2}{4\pi\epsilon_0\, R^2}\gamma^4 \int_{s_0}^{s} \rho(s')\, \Phi\left(\frac{\gamma\,\mu(s-s')}{2}\right) \mathrm{d}s', \tag{4.2.36}$$

where $\mu(s-s')$ is the solution of Equation (4.2.28) and s_0 is an appropriately chosen lower integration limit. This expression can be directly used to determine the CSR kick function numerically.

It can be helpful to determine the integral in Equation (4.2.36) using integration by parts. The integral of the kernel function Φ_{CSR} can be determined by introducing $\hat{\mu} = \gamma\mu/2$, which is related to the path-length difference via Equation (4.2.28)

$$\frac{1}{3}\hat{\mu}^3 + \hat{\mu} = \gamma^3\,\frac{s-s'}{R}. \tag{4.2.37}$$

For the differential element, this yields

$$\mathrm{d}s' = -\frac{R}{\gamma^3}\left(\hat{\mu}^2+1\right)\,\mathrm{d}\hat{\mu}, \tag{4.2.38}$$

so that the integral can be evaluated:

$$\int_{\mathbb{R}} \Phi\left(\frac{\gamma\,\mu(s-s')}{2}\right)\mathrm{d}s' = -\frac{R}{\gamma^3}\int_{\mathbb{R}} (\hat{\mu}^2+1)\Phi\left(\hat{\mu}\right)\mathrm{d}\hat{\mu} = \frac{R}{\gamma^3}\,\Theta\left(\hat{\mu}\right), \tag{4.2.39}$$

where the auxiliary function $\Theta\left(\mu\right)$ is defined as

$$\Theta\left(\mu\right) \equiv \frac{2\mu\left(\mu^2+2\right)}{(\mu^2+2)^2-1}. \tag{4.2.40}$$

In Figure 4.4, a plot of $\Theta\left(\mu\right)$ is shown. With this, the integral can can be written as

$$\begin{aligned}\int_{s_0}^{s} \rho(s')\,\Phi_{\text{CSR}}\left(\frac{\gamma\,\mu(s-s')}{2}\right)\mathrm{d}s' = &-\frac{R}{\gamma^3}\,\rho(s_0)\,\Theta\left(\hat{\mu}(s-s_0)\right)\\ &-\frac{R}{\gamma^3}\int_{s_0}^{s}\frac{\mathrm{d}\rho(s')}{\mathrm{d}s}\,\Theta\left(\hat{\mu}(s-s_0)\right)\mathrm{d}s',\end{aligned} \tag{4.2.41}$$

where the identity $\Theta\left(\hat{\mu}(0)\right) = 0$ was used.

From the above expression, an approximation for the case of large particle distances and large lower integration limit s_0 can be derived, as suggested in [75]. Considering a line charge density with compact support, the term proportional to $\rho(s_0)$ in Equation (4.2.41) vanishes for a sufficiently large value of s_0. If the particle distance is much larger than the scaled bending radius $s-s' \gg \frac{R}{\gamma^3}$, then Equation (4.2.37) yields for the scaled retarded angle the approximation

$$\hat{\mu} \approx 3^{\frac{1}{3}}\,\gamma\,\left(\frac{s-s'}{R}\right)^{\frac{1}{3}}. \tag{4.2.42}$$

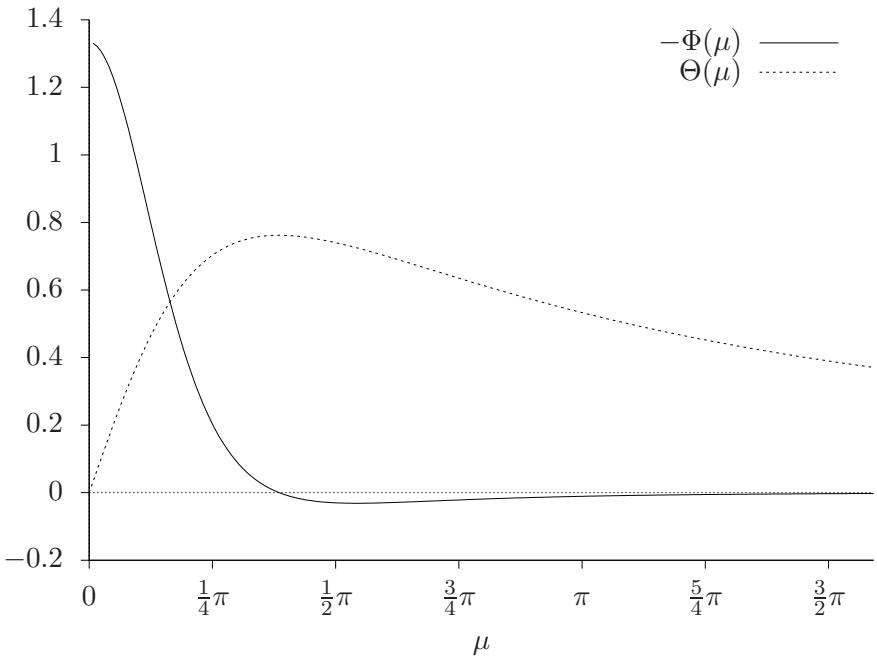

Figure 4.4: Plots of the CSR kernel function $\Phi(\hat{\mu})$ and its anti-derivative $\Theta(\hat{\mu})$.

For $\hat{\mu} \gg 1$, the function $\Theta(\hat{\mu})$ approaches $\Theta(\hat{\mu}) \approx 2\hat{\mu}^{-1}$. Hence, for large particle distances and $s_0 \to \infty$ the total CSR interaction term given in Equation (4.2.41) can be approximated by

$$\left(\frac{\mathrm{d}E}{\mathrm{d}s}\right)_{\mathrm{CSR}}\Bigg|_s \approx -\left(\frac{8}{3}\right)^{\frac{1}{3}} \frac{e^2}{4\pi\epsilon_0\, R^{\frac{2}{3}}} \int_{\infty}^{s} \frac{\mathrm{d}\rho(s')}{\mathrm{d}s} \left(s-s'\right)^{-\frac{1}{3}} \mathrm{d}s', \quad s - s' \gg \frac{R}{\gamma^3}. \tag{4.2.43}$$

It needs to be stressed that this approximation is valid only for large particle distances. Interactions between particles with separation smaller than R/γ^3 are not reflected correctly by this approximation. To treat such small-scale interactions correctly, the original Equation (4.2.36) has to be used.

4.2.6 Transient Effects

Consider the situation close to the entrance of the dipole. Before entering the dipole, the electrons travel on a straight trajectory and therefore do not emit radiation. Only after they enter the dipole field, they start to radiate. For the sake of simplicity, assume that all electrons enter the dipole simultaneously so that all electrons start to radiate at the same time. Close to the entrance of the magnet, the radiation emitted by the electrons further back in the bunch did not yet reach all electrons further in front of the bunch, due to the finite propagation speed. By definition, only the radiation emitted by electrons that are closer to the observer electron than the so-called *slippage length* did have enough time to reach the observer electron. It can be seen that this length corresponds to the difference between the path length which the bunch has traveled within the magnet and the straight-line distance to its entry point into the magnet [75, 78]. Thus, the slippage length increases as the bunch travels through the magnet. If the angular position of the bunch within the magnet is denoted by ϕ, it can be seen from Equation (4.2.28) that for large angles ϕ, the slippage length is $s_L = \frac{R\phi^3}{24}$. This circumstance can be accounted for in the calculation of the differential CSR kick function by replacing the lower integration limit in Equation (4.2.36) by $s_0 \to s - s_L$.

5 Microbunching Instability

In the previous sections, the single-particle dynamics as well as the collective effects that occur in free-electron laser injector beamlines were introduced. The longitudinal single-particle dynamics are dominated by the compression of the bunch in the bunch-compression stages. Collective effects occur due to space-charge forces, radiative interaction in dipole magnets, and wakefields in the acceleration cavities. In this section, it is investigated how the combination of bunch compression and collective interaction can drive a beam instability in the longitudinal phase-space, which can lead to pronounced inhomogeneities in the longitudinal charge-density of the electron bunch, as well as modulations of the particle energy along the bunch. These inhomogeneities take the form of a series of localized regions of high charge-density separated by regions of low charge-density. Due to this partitioning of the bunch into much smaller *microbunches* the process is referred to as the *microbunching instability* (MBI). To study the microbunching instability, we start by developing a perturbation theory based on the Fréchet–Taylor expansion of collective Perron–Frobenius operators. A brief summary of parts of the results presented in this section is given in reference [79] by the author.

5.1 Expansion of Collective Perron–Frobenius Operators

As shown in section 2.3.7, a collective Perron–Frobenius operator $\mathcal{M}[\cdot]$ can be interpreted as an operator that takes a phase-space density, which lives on a Banach space of Lesbeque integrable functions $\mathcal{W}$, and produces an element of the group of Perron–Frobenius operators

$$\mathcal{M}[\cdot]\colon \mathcal{W} \to \mathcal{PF}. \tag{5.1.1}$$

The following perturbative approach is based on the expansion of a collective PF operator with respect to this collective dependence utilizing its Fréchet derivatives.

Definition 1 (Fréchet derivative)**.** *Let V, W be normed $\mathbb{K}$-vector spaces and $U \subseteq V$ open. Let $f : V \to W$. If for all $x \in U$ a linear and bounded operator $Df(x) \in \mathrm{lin}(V, W)$ exists so that for any $h \in V$ it is*

$$\lim_{\|h\|_V \to 0} \frac{\|f(x+h) - f(x) - Df(x)\,h\|_W}{\|h\|_V} = 0, \tag{5.1.2}$$

then $Df(x)$ is unique and is called the Fréchet derivative of f at x and f is called Fréchet differentiable on U [80].

From Definition 1, it follow that Fréchet derivatives are defined only for mappings between two vector spaces. However, the codomain of collective PF operators is the group of PF operators, which in fact do not form a vector space. PF operators, however, form a subset of the linear maps between phase-space densities

$$\mathrm{lin}(\mathcal{W}, \mathcal{W}) \supset \mathcal{PF}, \tag{5.1.3}$$

which indeed do form a $\mathbb{K}$-vector space with respect to addition and scalar multiplication with any $\lambda \in \mathbb{K}$. Moreover, together with operator composition as the multiplicative operation, $\mathrm{lin}(\mathcal{W}, \mathcal{W})$ forms an unitary associative $\mathbb{K}$-algebra, as shown in Section 2.3.3, which we will refer to as an "algebra" without explicitly mentioning the unitarity and associativity.

A collective PF operator can therefore also be considered a mapping between $\mathcal{W}$ and the larger space of linear endomorphisms of $\mathcal{W}$, which is a superset of the PF operators:

$$\mathcal{M}[\cdot]\colon \mathcal{W} \to \mathrm{lin}(\mathcal{W}, \mathcal{W}). \tag{5.1.4}$$

By enlarging the codomain in this way, $\mathcal{M}[\cdot]$ is now a map between two vector spaces, so that we can now meaningfully assign Fréchet derivatives to collective PF operators.

If $\mathcal{M}[\cdot]$ is Fréchet differentiable on the open subset $U \subset \mathcal{W}$, it can be seen that the collective PF operator of a sum of two PSDs $\Psi, \phi \in U$ can be written as

$$\mathcal{M}[\Psi + \phi] = \mathcal{M}[\Psi] + D\mathcal{M}[\Psi]\,\phi + o(\|\phi\|). \tag{5.1.5}$$

Definition 2 (Higher order Fréchet derivatives)**.** *Let V, W be normed $\mathbb{K}$-vector spaces and $U \subseteq V$ open. Let $f : V \to W$. f is Fréchet differentiable n times on U, exactly if the Fréchet derivative Df exists and is Fréchet differentiable $(n-1)$ times on U.*

If the collective PF operator $\mathcal{M}[\cdot]$ is N times Fréchet differentiable at Ψ, then $\mathcal{M}[\Psi + \phi]$ can be expanded into a truncated Taylor series

$$\mathcal{M}[\Psi + \phi] = \sum_{n=0}^{N} \frac{1}{n!} D^n \mathcal{M}[\Psi]\,\phi^n + o(\|\phi\|^N), \tag{5.1.6}$$

where $D^n M[\Psi]$ is the nth Fréchet derivative of $\mathcal{M}[\cdot]$ at Ψ and the remainder term $o(\|\phi\|^N)$ denotes all terms of order higher than N [80].

From the definition of a Fréchet derivative, it follows that $D^n M[\Psi]$ is a linear map of the form

$$\mathcal{W} \to \underbrace{\mathrm{lin}(\mathcal{W}, \mathrm{lin}(\mathcal{W}, \dots)\dots)}_{\times n}. \tag{5.1.7}$$

This can equivalently be interpreted as a multi-linear map $\mathcal{W}^n \to \mathrm{lin}(\mathcal{W}, \mathcal{W})$. The expression $D^n\mathcal{M}[\Psi]\,\phi^n$ then denotes the evaluation of that map at $(\underbrace{\phi, \cdots, \phi}_{\times n})$.

5.2 Perturbation Theory

It is important to note that the microbunching instability is an amplification process, in the sense that it can can only amplify or attenuate *preexisting* inhomogeneities in the PSD. In turn, this implies that if a PSD does not feature inhomogeneities initially, there is nothing to be amplified and the microbunching instability will not occur. In particular, the microbunching instability cannot induce inhomogeneities into an initially homogeneous PSD.

This circumstance opens up an opportunity of systematically studying the MBI by means of a perturbative approach. One can envision a PSD Ψ_0 which is sufficiently homogeneous, so that it is not affected by the instability. Most generally, in a collective system, any PSD propagates via a collective Perron–Frobenius operator

$$\Psi_0 \mapsto \mathcal{M}[\Psi_0]\Psi_0 \equiv \Psi_1. \tag{5.2.1}$$

In this perturbation theory, the behavior of a PSD $\Psi_0 + \epsilon\phi_0$ is studied that consists of a homogeneous PSD Ψ_0 which is perturbed by a second PSD ϕ_0, where the strength of the perturbation is governed by the perturbation parameter $\mathbb{R} \ni \epsilon \ll 1$. This perturbation introduces the initial inhomogeneities which seed the microbunching instability. The effect induced by the MBI can then be studied by comparing the evolution of the undisturbed PSD with that of the disturbed PSD. Of particular interest is the evolution of the perturbation function ϕ_0.

Just as in the unperturbed case, also the perturbed PSD is propagated by a collective PF operator

$$\Psi_0 + \epsilon\phi_0 \mapsto \mathcal{M}[\Psi_0 + \epsilon\phi_0](\Psi_0 + \epsilon\phi_0). \tag{5.2.2}$$

It is important to note that while a collective PF operator is – just as any other PF operator – linear with respect to the PSD that it propagates

$$\mathcal{M}[\Psi_0 + \epsilon\phi_0](\Psi_0 + \epsilon\phi_0) = \mathcal{M}[\Psi_0 + \epsilon\phi_0]\Psi_0 + \epsilon\mathcal{M}[\Psi_0 + \epsilon\phi_0]\phi_0, \tag{5.2.3}$$

it is generally non-linear with respect to its collective dependence

$$\mathcal{M}[\Psi_0 + \epsilon\phi_0](\Psi_0 + \epsilon\phi_0) \neq \mathcal{M}[\Psi_0](\Psi_0 + \epsilon\phi_0) + \epsilon\mathcal{M}[\phi_0](\Psi_0 + \epsilon\phi_0). \tag{5.2.4}$$

Due to this non-linearity, it is not immediately possible to isolate the term corresponding to the propagated unperturbed part, $\mathcal{M}[\Psi_0]\Psi_0$, in Equation (5.2.2) so that an expression for the propagated disturbance function cannot be extracted.

As a remedy, the collective PF operator is expanded with respect to its collective dependence into the Fréchet–Taylor series (5.1.6). Then, the right-hand side of Equation (5.2.2) reads

$$\mathcal{M}[\Psi_0 + \epsilon\phi_0](\Psi_0 + \epsilon\phi_0) = \left(\sum_{n=0}^{N} \frac{1}{n!}\epsilon^n D^n \mathcal{M}[\Psi_0]\phi_0^n + \epsilon^N o\left(\|\phi_0\|^N\right)\right)(\Psi_0 + \epsilon\phi_0). \tag{5.2.5}$$

Rearranging the sum by equal orders of the perturbation parameter ϵ yields

$$\begin{aligned}\mathcal{M}[\Psi_0 + \epsilon\phi_0](\Psi_0 + \epsilon\phi_0) =& \mathcal{M}[\Psi_0]\Psi_0 + \sum_{n=1}^{N} \frac{1}{n!}\epsilon^n \left(D^n \mathcal{M}[\Psi_0]\phi_0^n\Psi_0 + n\, D^{n-1}\mathcal{M}[\Psi_0]\phi_0^n\right) \\ &+ \epsilon^N o\left(\|\phi_0\|^N\right).\end{aligned} \tag{5.2.6}$$

We see that in Equation (5.2.6), the term of the propagated unperturbed PSD $\mathcal{M}[\Psi_0]\Psi_0$ is now isolated. The contribution of the perturbation ϕ_0 is fully contained in the terms in the sum. By defining the nth order propagated perturbation density

$$\phi_{1,n} \equiv D^n \mathcal{M}[\Psi_0]\phi_0^n\Psi_0 + n\, D^{n-1}\mathcal{M}[\Psi_0]\phi_0^n \tag{5.2.7}$$

Equation (5.2.6) can be written more concisely as

$$\mathcal{M}[\Psi_0 + \epsilon\phi_0](\Psi_0 + \epsilon\phi_0) = \Psi_1 + \sum_{n=1}^{N} \frac{1}{n!}\epsilon^n \phi_{1,n} + \epsilon^N o\left(\|\phi_0\|^N\right). \tag{5.2.8}$$

This is as far as we will formulate the perturbation theory for an arbitrary unperturbed initial PSD, an arbitrary perturbation density, and an arbitrary collective PF operator. In order to arrive at a closed-form expression for the propagated perturbation densities, it is necessary to derive an explicit expression for the Fréchet derivatives in Equation (5.2.7).

5.3 Collective-Kick Perron–Frobenius Operators

Of particular interest for the microbunching instability is the collective interaction via longitudinal space-charge forces and via coherent synchrotron radiation. Both of these interactions manifest themselves as a kick map that effects the energy coordinate of a particle, depending on its longitudinal position. The local strength of the kick is given by a function that depends on the charge-density function of the bunch. In particular, this kick function is generally given as the solution of a Poisson equation with the charge density as the source term. A collective kick map of this type can then be written as

$$K[\Psi]\colon (q,p)^{\mathrm{T}} \mapsto (q, p + k[\Psi](q))^{\mathrm{T}}, \tag{5.3.1}$$

where the collective kick function $k[\Psi]$ is given by the convolution of the charge density of the PSD with the Green's function G of the underlying Poisson equation

$$k[\Psi](q) = \iint\limits_{\mathbb{R}^2} G(q,q')\Psi(q',p')\,\mathrm{d}q'\,\mathrm{d}p' \equiv [G \circledast \Psi](q), \tag{5.3.2}$$

where we have introduced the symbol $\circledast$ as a short hand for the convolution with respect to the q dimension together with a simultaneous projection along the p dimension, dropping the function arguments to avoid overly lengthy notation. It can be shown that the Green's function $G(q,q')$ of the free-space Poisson equation actually depends only on the difference of its arguments, $q-q'$. Here, the symbol G shall be used for both, the one-parameter and the two-parameter versions of the Green's function

$$G(q,q') \equiv G(q-q'). \tag{5.3.3}$$

As the integral in Equation (5.3.2) is linear with respect to both, Ψ and G, the $\circledast$-operation is bilinear. Consequently, also the kick function is linear in its dependence on the PSD: For all $\Psi, \Phi \in \mathcal{W}$ and $\mu, \nu \in \mathbb{R}$ it is

$$k[\mu\Psi + \nu\Phi] = G \circledast (\mu\Psi + \nu\Phi) = \mu(G \circledast \Psi) + \nu(G \circledast \Phi) = \mu k[\Psi] + \nu k[\Phi]. \tag{5.3.4}$$

From Equation (5.3.1) we see that the linearity of the kick function implies that the kick map generated by the sum of any two PSDs $\Psi, \Phi \in \mathcal{W}$ is equal to the composition of the kick maps generated by the individual PSDs

$$\begin{aligned} K[\Psi+\Phi] &= (q,p)^{\mathrm{T}} \mapsto (q, p + k[\Psi](q) + k[\Phi](q))^{\mathrm{T}} && (5.3.5)\\ &= \left[(q,p)^{\mathrm{T}} \mapsto (q, p + k[\Psi](q))^{\mathrm{T}}\right] \circ \left[(q,p)^{\mathrm{T}} \mapsto (q, p + k[\Phi](q))^{\mathrm{T}}\right] && (5.3.6)\\ &= K[\Psi] \circ K[\Phi] = K[\Phi] \circ K[\Psi], && (5.3.7) \end{aligned}$$

from which we note that all kicks commute. Consequently, the same relation holds for the Perron–Frobenius operator corresponding to this map: For any $\phi \in \mathcal{W}$, it is

$$\begin{aligned} \mathcal{K}[\Psi+\Phi]\phi(z) &= \phi\left(K[\Psi+\Phi]^{-1}(z)\right) = \phi\left(\{K[\Psi] \circ K[\Phi]\}^{-1}(z)\right) && (5.3.8)\\ &= \phi\left(\{K[\Phi]^{-1} \circ K[\Psi]^{-1}\}(z)\right) = \phi\left(K[\Phi]^{-1}(K[\Psi]^{-1}(z))\right) && (5.3.9)\\ &= \mathcal{K}[\Psi]\phi\left(K[\Phi]^{-1}(z)\right) && (5.3.10)\\ &= \mathcal{K}[\Psi]\mathcal{K}[\Phi]\phi\,(z) = \mathcal{K}[\Phi]\mathcal{K}[\Psi]\phi\,(z)\,. && (5.3.11) \end{aligned}$$

In general, the relation

$$\mathcal{M}[\Psi+\Phi] = \mathcal{M}[\Psi]\mathcal{M}[\Phi] \tag{5.3.12}$$

defines a class of collective PF operators, which we call *homomorphic* collective PF operators. Another fitting term for this property would be "quasi-exponential". In the following, some general properties of the Fréchet derivatives of members of this class – and therefore in particular of collective kick operators – are derived.

5.4 Fréchet Derivatives of Homomorphic Perron–Frobenius Operators

In general, a collective PF operator $\mathcal{M}[\cdot] : \mathcal{W} \to \mathrm{lin}(\mathcal{W}, \mathcal{W})$ is a mapping from the Banach space $\mathcal{W}$ of the phase-space densities to the space of linear automorphism of the same space. By definition, the underlying vector space of $\mathcal{W}$, which is the domain of $\mathcal{M}[\cdot]$, constitutes an Abelian group under addition. It can be seen that $\mathrm{lin}(\mathcal{W}, \mathcal{W})$ forms an associative algebra over $\mathbb{R}$, where pointwise addition and operator concatenation are the addition and multiplication operation, respectively. With this in mind, we see that Equation (5.3.12) describes the condition for $\mathcal{M}[\cdot]$ to be a homomorphism between the vector space $\mathcal{W}$ and the algebra $\mathrm{lin}(\mathcal{W}, \mathcal{W})$, with respect to its multiplication operation. We therefore define the following notion of a homomorphic maps from vector spaces to algebras.

Definition 3 (Vector-Space-to-Algebra Homomorphism)**.** *Let $(V, \mathbb{K}, +)$ and $(W, \mathbb{K}, +)$ be vector spaces over a field $\mathbb{K}$. Let A be the algebra $(W, \mathbb{K}, +, *)$ defined by $(W, \mathbb{K}, +)$ together with the multiplicative operation $* : W \times W \to W$. A map $f : V \to A$ is called homomorphic exactly if $f(x + y) = f(x) * f(y)$ for all $x, y \in V$.*

For these homomorphisms, it is possible to prove a powerful statement regarding their Fréchet derivatives. It states that the local Fréchet derivative of a homomorphism is given by its local value multiplied with its Fréchet derivative evaluated at the zero element of the domain vector space.

Theorem 2 (Fréchet Derivative of a Vector-Space-to-Algebra Homomorphism.)**.** *Let $(V, \mathbb{K}, +), (W, \mathbb{K}, +)$ be a normed vector spaces. Let A be the algebra $(W, \mathbb{K}, +, *)$. Let $f : V \to A$ be a bounded homomorphism. If f is Fréchet differentiable at 0, then f is Fréchet differentiable on V and its Fréchet derivative at $x \in V$ is given by*

$$Df(x) = f(x) * Df(0) \tag{5.4.1}$$

Proof. From the fact that f is homomorphic, we see

$$\|f(x + h) - f(x) - Df(x)\, h\|_W = \|f(x) * (f(h) - \mathrm{Id}_W) - Df(x)\, h\|_W. \tag{5.4.2}$$

Using $f(h) = f(0) + Df(0)\, h + o(\|h\|_V)$ and $f(0) = \mathrm{Id}_W$, this yields

$$\begin{aligned}
&\|f(x) * (f(h) - \mathrm{Id}_W) - Df(x)\, h\|_W && (5.4.3)\\
=&\|f(x) * Df(0)h - Df(x)\, h + f(x) * o(\|h\|_V)\|_W && (5.4.4)\\
\leqslant&\|f(x) * Df(0)h - Df(x)\, h\|_W + \|f(x) * o(\|h\|_V)\|_W && (5.4.5)\\
\leqslant&\|f(x) * Df(0)h - Df(x)\, h\|_W + \|f(x)\|_W\, o(\|h\|_V). && (5.4.6)
\end{aligned}$$

Therefore, it is

$$Df(x) = f(x) * Df(0) \iff \lim_{\|h\|_V \to 0} \frac{\|f(x + h) - f(x) - Df(x)\, h\|_W}{\|h\|_V} = 0. \tag{5.4.7}$$

□

Further, it is possible to show that higher-order Fréchet derivatives of homomorphic Vector-Space-to-Algebra maps are given by repeated application of the first-order Fréchet derivative.

Theorem 3 (Higher-order Fréchet Derivatives of Vector-Space-to-Algebra Homomorphisms.)**.** *Let $(V,+),(W,+)$ be normed vector spaces. Let A be the algebra $(W,+,*)$. Let $f\colon V\to A$ be a bounded Fréchet differentiable homomorphism. Then, for $n\geqslant 1$ the nth Fréchet derivative of f is given by*

$$(D^n f(x))y^n = f(x) * (Df(0)\,y)^n. \tag{5.4.8}$$

Proof. For $n=1$ the theorem is proven in Theorem 2. For $n>1$, the theorem is proven by induction as follows. For the $(n+1)$st Fréchet derivativ, it is with $x,h,g\in V$

$$\begin{aligned}
0 &= \lim_{\|h\|_V\to 0}\frac{1}{\|h\|_V}\|D^n f(x+h)\,g^n - D^n f(x)\,g^n - D^{n+1}f(x)\,g^n h\|_W && (5.4.9)\\
&= \lim_{\|h\|_V\to 0}\frac{1}{\|h\|_V}\|(f(x+h)-f(x)) * (Df(0)\,g)^n - D^{n+1}f(x)\,g^n h\|_W && (5.4.10)\\
&= \lim_{\|h\|_V\to 0}\frac{1}{\|h\|_V}\|(Df(x)\,h+o(\|h\|_V)) * (Df(0)\,g)^n - D^{n+1}f(x)\,g^n h\|_W && (5.4.11)\\
&\leqslant \lim_{\|h\|_V\to 0}\frac{1}{\|h\|_V}\|Df(x)\,h * (Df(0)\,g)^n - D^{n+1}f(x)\,g^n h\|_W && (5.4.12)\\
&= \lim_{\|h\|_V\to 0}\frac{1}{\|h\|_V}\|f(x) * Df(0)\,h * (Df(0)\,g)^n - D^{n+1}f(x)\,g^n h\|_W, && (5.4.13)
\end{aligned}$$

which implies

$$D^{n+1}f(x)g^n\,h = f(x) * Df(0)\,h * (Df(0)\,g)^n. \tag{5.4.14}$$

Setting $h=g=y$, this shows that

$$D^{n+1}f(x)y^{n+1} = f(x) * (Df(0)\,y)^{n+1}. \tag{5.4.15}$$

□

It can also be seen how Fréchet derivatives of Vector-Space-to-Algebra maps behave under application of the binary operation of the codomain algebra.

Theorem 4. *Let $(V,+),(W,+)$ be normed vector spaces. Let A be the algebra $(W,+,*)$. Let $f\colon V\to A$ be Fréchet differentiable at $x\in V$. Then, for all $y\in V$, $g\in A$*

$$D(f(x)*g)\,y = \big(Df(x)\,y\big) * g \tag{5.4.16}$$

and

$$D(g*f(x))\,y = g * Df(x)\,y. \tag{5.4.17}$$

Proof. First, Equation (5.4.16) is proven. The proof for Equation (5.4.17) follows by analogy. Let $u(x)=f(x)*g$, then

$$\begin{aligned}
0 &= \lim_{\|h\|_V\to 0}\frac{\|u(x+h)-u(x)-Du(x)\,h\|_W}{\|h\|_V} && (5.4.18)\\
&= \lim_{\|h\|_V\to 0}\frac{\|[f(x+h)-f(x)]*g-Du(x)\,h\|_W}{\|h\|_V} && (5.4.19)\\
&= \lim_{\|h\|_V\to 0}\frac{\|[Df(x)\,h+o(\|h\|_V)]*g-Du(x)\,h\|_W}{\|h\|_V} && (5.4.20)\\
&\leqslant \lim_{\|h\|_V\to 0}\frac{\|Df(x)\,h*g-Du(x)\,h\|_W}{\|h\|_V}, && (5.4.21)
\end{aligned}$$

which implies

$$D(f(x)*g)\,y = Df(x)\,y*g. \tag{5.4.22}$$

□

Note that Theorem 4 holds for general Vector-Space-to-Algebra maps, not only for homomorphisms.

5.4.1 Linearly Collective Hamiltonians

Consider a system with a homomorphic collective Perron–Frobenius operator $M[\cdot]\colon \mathcal{W} \to \mathrm{lin}(\mathcal{W}, \mathcal{W})$ that can be expressed in terms of a Lie series, as defined in Section 2.3.5, with the collective Hamiltonian $H[\cdot]\colon \mathcal{W} \to C^1(\mathbb{R}^2, \mathbb{R})$

$$\mathcal{M}[\Psi] = \exp(:H[\Psi]:). \tag{5.4.23}$$

If this Hamiltonian depends linearly on the phase-space density, in the sense that for all $\mu, \nu \in \mathbb{R}$ and $\Psi, \Phi \in \mathcal{W}$ it is

$$H[\mu\Psi + \nu\Phi] = \mu H[\Psi] + \nu H[\Phi], \tag{5.4.24}$$

then it can be seen that the Fréchet derivative of operators of this type can be calculated explicitly.

Theorem 5 (Fréchet Derivative of homomorphic Perron–Frobenius Operators with linearly collective Hamiltonians)**.** *Let $\mathcal{M}[\cdot] : \mathcal{W} \to \mathrm{lin}(\mathcal{W}, \mathcal{W})$ be homomorphic. Let $\mathcal{M}[\cdot]\colon \Psi \mapsto \exp(:H[\Psi]:)$, where $H[\cdot]\colon \mathcal{W} \to C^1(\mathbb{R}^2, \mathbb{R})$ is linear. Then for all $\Psi \in \mathcal{W}$*

$$D\mathcal{M}[\Psi]\cdot = \mathcal{M}[\Psi] :H[\cdot]:, \tag{5.4.25}$$

in the sense that $D\mathcal{M}[\Psi]\,\phi = \mathcal{M}[\Psi] :H[\phi]:$

Proof.

$$0 = \lim_{\|\phi\|_{\mathcal{W}} \to 0} \frac{\|\mathcal{M}[\Psi + \phi] - \mathcal{M}[\Psi] - D\mathcal{M}[\Psi]\,\phi\|_{\mathrm{lin}(\mathcal{W},\mathcal{W})}}{\|\phi\|_{\mathcal{W}}} \tag{5.4.26}$$

$$= \lim_{\|\phi\|_{\mathcal{W}} \to 0} \frac{\|\mathcal{M}[\Psi](\mathcal{M}[\phi] - \mathrm{Id}) - D\mathcal{M}[\Psi]\,\phi\|_{\mathrm{lin}(\mathcal{W},\mathcal{W})}}{\|\phi\|_{\mathcal{W}}} \tag{5.4.27}$$

$$= \lim_{\|\phi\|_{\mathcal{W}} \to 0} \frac{\|\mathcal{M}[\Psi](\exp(:H[\phi]:) - \mathrm{Id}) - D\mathcal{M}[\Psi]\,\phi\|_{\mathrm{lin}(\mathcal{W},\mathcal{W})}}{\|\phi\|_{\mathcal{W}}} \tag{5.4.28}$$

$$= \lim_{\|\phi\|_{\mathcal{W}} \to 0} \frac{\|\mathcal{M}[\Psi](:H[\phi]: + o(\|\phi\|)) - D\mathcal{M}[\Psi]\,\phi\|_{\mathrm{lin}(\mathcal{W},\mathcal{W})}}{\|\phi\|_{\mathcal{W}}} \tag{5.4.29}$$

$$\leqslant \lim_{\|\phi\|_{\mathcal{W}} \to 0} \frac{\|\mathcal{M}[\Psi] :H[\phi]: - D\mathcal{M}[\Psi]\,\phi\|_{\mathrm{lin}(\mathcal{W},\mathcal{W})}}{\|\phi\|_{\mathcal{W}}}, \tag{5.4.30}$$

which implies

$$D\mathcal{M}[\Psi]\,\phi = \mathcal{M}[\Psi] :H[\phi]: . \tag{5.4.31}$$

□

We note that the requirement for $H[\cdot]$ to be linear was not used explicitly in the proof of Theorem 5. Nevertheless, the Hamiltonian indeed does need to be linear in its collective dependence as otherwise the Fréchet derivative $D\mathcal{M}[\Psi]\cdot \equiv \mathcal{M}[\Psi] :H[\cdot]:$ would not be a linear operator, which is required by its definition.

Neither does the linearity of the collective Hamiltonian imply that the collective Perron–Frobenius operator is homomorphic, nor vice versa. It can, however, be shown that if in addition to its linearity the Lie operators of the Hamiltonians commute for all phase-space densities, then the Perron–Frobenius operator is homomorphic.

Theorem 6. *Let $\mathcal{M}[\cdot]\colon \Psi \mapsto \exp(:H[\Psi]:)$, where $H[\cdot]\colon \mathcal{W} \to C^1(\mathbb{R}^2,\mathbb{R})$ is linear and let H be such that for all $\Psi,\Phi \in \mathcal{W}$ it is*

$$[:H[\Psi]:,:H[\Phi]:] = 0, \tag{5.4.32}$$

where $[\cdot,\cdot]$ denotes the Lie commutator $[A,B] = AB - BA$. Then $\mathcal{M}[\cdot]$ is homomorphic in the sense of Definition 3.

Proof. From the linearity of the Hamiltonian and the Poisson bracket, we see

$$\mathcal{M}[\Psi+\Phi] = \exp(:H[\Psi+\Phi]:) = \exp(:H[\Psi]: + :H[\Phi]:). \tag{5.4.33}$$

Using the Baker-Campbell-Hausdorff identity $\exp(A+B) = \exp(A)\exp(B)$ for operators A, B with vanishing commutator, $[A,B] = 0$, it can be seen that

$$\mathcal{M}[\Psi+\Phi] = \exp(:H[\Psi]:)\exp(:H[\Phi]:) = \mathcal{M}[\Psi]\mathcal{M}[\Phi]. \tag{5.4.34}$$

□

Together with Theorem 3, we see that if $H[\cdot]$ is linear and self-commutative – as required by Theorem 6 – then the higher order Fréchet derivatives of the associated Perron–Frobenius operator $\mathcal{M}[\Psi] = \exp(:H[\Psi]:)$ are given by repeated application of the Lie operator of the collective Hamiltonian

$$D^n\mathcal{M}[\Psi]\,\phi^n = \mathcal{M}[\Psi](D\mathcal{M}[0]\,\phi)^n = \mathcal{M}[\Psi](\mathcal{M}[0]:H[\phi]:)^n = \mathcal{M}[\Psi]:H[\phi]:^n, \tag{5.4.35}$$

where we have used the identity $\mathcal{M}[0] \equiv \mathrm{Id}$, which follows from the fact that the Perron–Frobenius operator is homomorphic. The notation $:H[\cdot]:^n$ is used in the sense that

$$:H[\cdot]:^n\ \phi^n \equiv :H[\phi]:^n, \tag{5.4.36}$$

which allows the higher-order derivatives to be written in operator form

$$D^n\mathcal{M}[\Psi] = \mathcal{M}[\Psi]:H[\cdot]:^n\ . \tag{5.4.37}$$

5.5 Fréchet Derivatives of Collective-Kick Operators

After having derived some general statements about homomorphic Perron–Frobenius operators, these results are in the following applied to our specific case at hand – namely collective kicks as introduced in Section 5.3. We have already seen that Perron–Frobenius operators of collective kicks are homomorphic in their collective dependence. In order to make use of Theorem 5, it is additionally necessary to show that collective kick maps are generated by linearly collective Hamiltonians.

It can be seen that a kick map of the form

$$K\colon (q,p)^{\mathrm{T}} \mapsto (q, p + k(q))^{\mathrm{T}} \tag{5.5.1}$$

is generated by the Hamiltonian $H(q,p) = -\kappa(q)$, where $\kappa(q)$ is the anti-derivative of the kick function $k(q)$ with

$$\frac{\partial\kappa(q)}{\partial q} = k(q). \tag{5.5.2}$$

With this Hamiltonian, it is

$$:H:\,q = 0, \quad :H:\,p = -\frac{\partial\kappa(q)}{\partial q} = -k(q) \tag{5.5.3}$$

and consequently for $n > 1$

$$:H:^n q = 0, \quad :H:^n p = 0. \tag{5.5.4}$$

Therefore, the exponential map $\exp(- :H:)$ is exactly the map given in Equation (5.5.1)

$$\exp(- :H:)q = q, \quad \exp(- :H:)p = p - :H: p = p + k(q). \tag{5.5.5}$$

This derivation is also applicable the kick function has a collective dependence on the phase-space density as in Equation (5.3.2). Here, the phase-space density is merely a parameter used to determine the kick function and does not add any functional dependencies on the phase-space coordinates. Using the anti-derivative Γ of the Green's function G, which is defined by

$$\frac{\partial \Gamma(q, q')}{\partial q} = G(q, q') \tag{5.5.6}$$

and assuming that both, $G(q, q')$ and $\Gamma(q, q')$, exist and are continuous everywhere, then the anti-derivative of k can be written as

$$\kappa[\Psi](q) = [\Gamma \circledast \Psi](q). \tag{5.5.7}$$

This follows from the Leibniz integral rule [81], which in this case states that the differentiation and integral operation in

$$\frac{\partial}{\partial q}\kappa[\Psi](q) = \frac{\partial}{\partial q}[\Gamma \circledast \Psi](q) = \frac{\partial}{\partial q}\iint\limits_{\mathbb{R}^2} \Gamma(q, q')\Psi(q', p') \,\mathrm{d}q' \,\mathrm{d}p' \tag{5.5.8}$$

$$= \iint\limits_{\mathbb{R}^2} \frac{\partial}{\partial q}\Gamma(q, q')\Psi(q', p') \,\mathrm{d}q' \,\mathrm{d}p' = \iint\limits_{\mathbb{R}^2} G(q, q')\Psi(q', p') \,\mathrm{d}q' \,\mathrm{d}p' \tag{5.5.9}$$

$$= [G \circledast \Psi](q) = k[\Psi](q) \tag{5.5.10}$$

are interchangeable. Analogously to the non-collective case, we therefore see that the collective Hamiltonian

$$H[\Psi](q, p) \equiv -[\Gamma \circledast \Psi](q) \tag{5.5.11}$$

generates the collective kick map shown in Equation (5.3.1)

$$\exp(- :H[\Psi]:) = K[\Psi]. \tag{5.5.12}$$

The linearity of this collective Hamiltonian with respect to its phase-space density parameter follows immediately from the bilinearity of the $\circledast$ operator.

This shows that Perron–Frobenius operators of collective kicks of this type are generated by a linearly collective Hamiltonian. Together with the fact that these Perron–Frobenius operators are homomorphic – which was shown in Section 5.3 – this implies the theorems introduced in Section 5.4 are indeed applicable to these operators and can be used to determine their Fréchet derivatives.

5.6 Expansion of a Single Microbunching Step

After having laid much groundwork regarding Fréchet derivatives of Perron–Frobenius operators, we are now in a good position to actually apply our findings to a concrete physically relevant scenario. Consider a single bunch compression stage, consisting of

a long linac section followed by a magnetic chicane. In an ultra-relativistic model, the dynamics of the longitudinal phase-space can be reduced to the kick introduced by the accelerating cavities, which we will denote by the map $K_{\text{acc}} : \mathbb{R}^2 \to \mathbb{R}^2$, plus a collective kick, $K[\Psi]_{\text{coll}} : \mathbb{R}^2 \to \mathbb{R}^2$, generated by the self-interaction due to collective forces such as longitudinal space-charge forces or wake fields. We assume that the $K[\Psi]_{\text{coll}}$ is generated by a Hamiltonian $H[\cdot]$ with a linear collective dependence as in Equation (5.5.11). As all kick maps commute, the order of these kicks is not relevant. Velocity effects are neglected here, which is a valid approximation in the ultra-relativistic limit. Omitting CSR-kicks, the longitudinal dynamics in a magnetic chicane are governed by the longitudinal dispersion it produces, which can be represented by a single drift map $D_{\text{chic}} : \mathbb{R}^2 \to \mathbb{R}^2$. The Perron–Frobenius operator $M[\Psi]$ of a single bunch compression stage is then given by

$$\mathcal{M}[\Psi] = \mathcal{D}_{\text{chic}}\, \mathcal{K}_{\text{acc}}\, \mathcal{K}[\Psi]_{\text{coll}} = \mathcal{M}_0 \mathcal{K}[\Psi]_{\text{coll}}, \tag{5.6.1}$$

where Ψ is the phase-space density at the beginning of the stage and we have assigned the symbol $\mathcal{M}_0 \equiv \mathcal{D}_{\text{chic}}\, \mathcal{K}_{\text{acc}}$ to the non-collective part of the total Perron–Frobenius operator. We make the approximation that both, the kick-function of the RF-modules and the dispersive drift-function of the magnetic chicane, are linear. This allows to use the symplectic LDU decomposition (3.4.6) of the non-collective map so that $\mathcal{M}_0$ can be written as

$$\mathcal{M}_0 = \mathcal{K}_{Ch}\, \mathcal{S}_C\, \mathcal{D}_{C\beta}, \tag{5.6.2}$$

where $\mathcal{K}, \mathcal{S}$, and $\mathcal{D}$ are the Perron–Frobenius operators of the kick-, compression-, and drift-maps defined in Section 3.4, respectively. Using Theorem 4, the Fréchet derivative of this operator can be seen to be

$$D\mathcal{M}[\Psi] = \mathcal{M}_0\, D\mathcal{K}[\Psi]_{\text{coll}}. \tag{5.6.3}$$

By repeated application of the theorem it follows that the higher-order derivatives are given by

$$D^n\mathcal{M}[\Psi] = \mathcal{M}_0\, D^n\mathcal{K}[\Psi]_{\text{coll}}. \tag{5.6.4}$$

As $K[\Psi]$ is generated by a linear Hamiltonian $K[\Psi]_{\text{coll}} = \exp(- :H_{\text{coll}}[\Psi]:)$, the Perron–Frobenius operator is given by $\mathcal{K}[\Psi]_{\text{coll}} = \exp(:H_{\text{coll}}[\Psi]:)$ and we see via Theorem 5 and Equation (5.4.37) that

$$D^n\mathcal{M}[\Psi] = \mathcal{M}_0\, \mathcal{K}_{\text{coll}}[\Psi]\ :H_{\text{coll}}[\cdot]:^n\ . \tag{5.6.5}$$

Using the definition of $\mathcal{M}[\Psi]$, this can be written concisely as

$$D^n\mathcal{M}[\Psi] = \mathcal{M}[\Psi]\ :H[\cdot]:^n, \tag{5.6.6}$$

where here and in the following we drop the subscript of the collective Hamiltonian for notational brevity and set $H[\cdot] \equiv H_{\text{coll}}[\cdot]$.

With this expression for the higher-order derivatives at hand, we can now turn to the evaluation of the perturbation theory introduced in Section 5.2. From the definition of the propagated perturbation densities (5.2.7)

$$\phi_{1,n} \equiv D^n\mathcal{M}[\Psi_0]\, \phi_0^n \Psi_0 + n\, D^{n-1}\mathcal{M}[\Psi_0]\, \phi_0^n, \tag{5.6.7}$$

we see by plugging in Equation (5.6.6)

$$\begin{aligned} \phi_{1,n}[\phi_0] &= \mathcal{M}[\Psi_0]\left(:H[\phi_0]:^n\ \Psi_0 + n\ :H[\phi_0]:^{n-1}\ \phi_0\right) && (5.6.8)\\ &= \mathcal{M}[\Psi_0]\ :H[\phi_0]:^{n-1}\ \left(:H[\phi_0]:\ \Psi_0 + n\phi_0\right), && (5.6.9) \end{aligned}$$

where we additionally introduced the notation $\phi_{1,n}[\cdot]$ to denote the functional dependence on the perturbation density.

5.6.1 Monochromatic Perturbation

One of the goals of this investigation of the microbunching instability is to provide insights on how the amplification factor depends on longitudinal scale of the initial perturbation. To this end, initial perturbation ϕ_0 are considered that are equal to the initial undisturbed phase-space density Ψ_0 but modulated sinusoidally on a single wavelength $2\pi/k$, with $k \in \mathbb{R}^+$, along the longitudinal dimension:

$$\varphi_k(q,p) \equiv \cos(kq)\Psi_0(q,p) = \frac{1}{2}\left[e^{i(kq)} + e^{-i(kq)}\right]\Psi_0(q,p). \tag{5.6.10}$$

Introducing $a_k(q) \equiv \frac{1}{2}\, e^{ikq}$, this can be written more concisely as

$$\varphi_k(q,p) = \left[a_k(q) + \bar{a}_k(q)\right]\Psi_0(q,p). \tag{5.6.11}$$

Further, it is assumed that the initial unperturbed phase-space density does not depend on the longitudinal coordinate and is determined solely by an energy distribution $\psi(p)$

$$\Psi_0(q,p) = \psi(p), \tag{5.6.12}$$

where $\int_{\mathbb{R}} \psi(p)\mathrm{d}p = \rho_0$ and ρ_0 denotes the value of the constant longitudinal charge density.

Physically, this corresponds to the case of an infinitely long, homogeneous bunch or, equivalently, a homogeneous bunch with finite longitudinal length, where edge effects are neglected. If we take l_b to be this finite length, then it is $\rho_0 \approx l_b^{-1}$. We note that the function Ψ_0 as defined in Equation (5.6.12) is not integrable over $\mathbb{R}^2$ and therefore is not a member of $\mathcal{W}$. This technicality can be circumvented formally by truncating Ψ_0 at an arbitrarily large but finite value of q, using a smooth cutoff function $\chi_{l_b}(q)$ so that $\Psi_0(q,p) = \psi(q)\chi_{l_b}(q)$. A suitable $\chi_{l_b}(q)$ can be constructed for instance by convolving the rectangle function with length l_b with an appropriate kernel function. The original phase-space density (5.6.12) can then always be recovered by taking the limit $l_b \to \infty$. Therefore, the results based on the assumption that $\Psi_0 \in \mathcal{W}$ are still valid in the case of an infinitely long bunch.

It can be seen that the unperturbed density (5.6.12) does not produce collective kicks, as the collective Hamiltonian is constant

$$H[\Psi_0] = -\Gamma \circledast \Psi_0 = -\iint_{\mathbb{R}^2} \Gamma(q-q')\psi(p')\,\mathrm{d}q'\,\mathrm{d}p' = -\rho_0 \int_{\mathbb{R}} \Gamma(q-q')\,\mathrm{d}q' = \text{const.} \tag{5.6.13}$$

For this choice of Ψ_0, it therefore is $\mathcal{K}_{\text{coll}}[\Psi_0] = \mathrm{Id}$ and consequently

$$\mathcal{M}[\Psi_0] = \mathcal{M}_0. \tag{5.6.14}$$

5.6.1.1 General Identities

Before continuing, we will derive some identities, which will be useful in the further evaluation. From the definition of $a_k(q)$, it can be seen that

$$a_{k_1}(q)a_{k_2}(q) = \frac{1}{2}a_{k_1+k_2}(q) \tag{5.6.15}$$

and

$$\bar{a}_k = a_{-k}. \tag{5.6.16}$$

Hence, in particular it is

$$a_k \bar{a}_k = \frac{1}{2} a_0 = \frac{1}{4}. \tag{5.6.17}$$

We note that a Hamiltonian of the form $H[a_k \psi]$ can be written as

$$\begin{aligned} H[a_k(q)\psi(p)] &= -\int_{\mathbb{R}} \Gamma(q-q') a_k(q') \,\mathrm{d}q' \int_{\mathbb{R}} \psi(p') \,\mathrm{d}p' = -\rho_0 \int_{\mathbb{R}} \Gamma(q-q') a_k(q') \,\mathrm{d}q' \\ &= \rho_0 \int_{\mathbb{R}} \Gamma(q') a_k(q-q') \,\mathrm{d}q' = \rho_0 a_k(q) \int_{\mathbb{R}} \Gamma(q') e^{-ikq'} \,\mathrm{d}q' \qquad (5.6.18) \\ &= \rho_0 a_k(q) \tilde{\Gamma}(k), \qquad (5.6.19) \end{aligned}$$

where $\tilde{\Gamma}$ denotes the Fourier transform of Γ. Analogously, $H[\bar{a}_k \psi]$ is given by

$$H[\bar{a}_k \psi] = \rho_0 \bar{a}_k(q) \int_{\mathbb{R}} \Gamma(q') e^{ikq'} \,\mathrm{d}q' = \rho_0 \bar{a}_k(q) \tilde{\Gamma}(-k) = \rho_0 \bar{a}_k(q) \overline{\tilde{\Gamma}(k)} = \overline{H[a_k \psi]}. \tag{5.6.20}$$

At this point, we define the function

$$W(k) \equiv \rho_0 \int_{\mathbb{R}} G(q) e^{-ikq} \,\mathrm{d}q = ik\rho_0 \int_{\mathbb{R}} \Gamma(q) e^{-ikq} \,\mathrm{d}q = ik\rho_0 \tilde{\Gamma}(k), \tag{5.6.21}$$

which – in a slight abuse of the term – we refer to as an *impedance function*. Conventionally, an impedance function is denoted with the symbol $W(k)$ and its unit is Ohm. In our definition, the impedance W has the same unit as the conjugate momentum p, which is eV. Our impedance function is related to the conventional impedance function by multiplication with the bunch charge Q_b and the elementary charge

$$W(k) = e\, Q_b\, Z(k). \tag{5.6.22}$$

As both definitions differ only by a constant factor and describe the same physical concept, calling $W(k)$ an impedance function is justified and will help the comprehensibility more than introducing a new term for it. With this definition of the impedance function, we get for the partial derivatives of the Hamiltonian

$$\frac{\partial}{\partial q} H[a_k \psi] = ika_k(q) \rho_0 \tilde{\Gamma}(k) = a_k(q) W(k) \tag{5.6.23}$$

and

$$\frac{\partial}{\partial p} H[a_k \psi] = 0, \tag{5.6.24}$$

so that the Lie operator $:H[a_k\psi]:$ can be written as

$$:H[a_k\psi]: = \frac{\partial H[a_k\psi]}{\partial q} \frac{\partial}{\partial p} - \frac{\partial H[a_k\psi]}{\partial p} \frac{\partial}{\partial q} = W(k) a_k(q) \frac{\partial}{\partial p}. \tag{5.6.25}$$

Composing two Lie operators therefore yields

$$:H[a_{k_2}\psi]::H[a_{k_1}\psi]: = \frac{1}{2} W(k_2) W(k_1) a_{k_2+k_1}(q) \frac{\partial^2}{\partial p^2}, \tag{5.6.26}$$

or more generally for any number N of Lie operators

$$\prod_{i=1}^{N} :H[a_{k_i}\psi]: = \frac{1}{2^{N-1}} \left(\prod_{i=1}^{N} W(k_i) \right) a_{\sum_{i=1}^{N} k_i}(q) \frac{\partial^n}{\partial p^n}. \tag{5.6.27}$$

Explicitly, for all functions $f, g \in C^1(\mathbb{R}, \mathbb{R})$ it is

$$:H[a_k\psi]: f(q)g(p) = f(q) W(k) a_k(q) \frac{\partial g}{\partial p}(p), \tag{5.6.28}$$

and in particular

$$:H[a_{k_1}\psi]: a_{k_0}(q)\psi(p) = \frac{1}{2} W(k_1) a_{k_1+k_0}(q) \frac{\partial \psi}{\partial p}(p). \tag{5.6.29}$$

5.6.1.2 First Order

From Equation (5.6.9), it can be seen that the general expression of the first-order propagated perturbation function reads

$$\phi_{1,1}[\varphi_k] = \mathcal{M}_0 \left(:H[\varphi_k]: \Psi_0 + \varphi_k\right), \tag{5.6.30}$$

which we note to be linear with respect to φ_k. Plugging in φ_k from Equation (5.6.11) together with Ψ_0 from Equation (5.6.12) yields

$$\phi_{1,1}[\varphi_k] = \phi_{1,1}[a_k\psi + \bar{a}_k\psi] = \phi_{1,1}[a_k\psi] + \phi_{1,1}[\bar{a}_k\psi]. \tag{5.6.31}$$

Using Equation (5.6.25) to evaluate the Lie operator, the first term can be written as

$$\phi_{1,1}[a_k\psi] = \mathcal{M}_0(:H[a_k\psi]: \psi + a_k\psi) = \mathcal{M}_0\left(a_k(q)\left[1 + W(k)\frac{\partial}{\partial p}\right]\psi(p)\right). \tag{5.6.32}$$

With this, the total first-order propagated perturbation function $\phi_{1,1}[\varphi_k]$ can be reconstructed from Equation (5.6.31):

$$\phi_{1,1}[\varphi_k] = \phi_{1,1}[a_k\psi] + \phi_{1,1}[\bar{a}_k\psi] = \mathcal{M}_0\left(a_k(q)\left[1 + W(k)\frac{\partial}{\partial p}\right]\psi(p)\right) + \text{c.c.}\,. \tag{5.6.33}$$

Therefore, the first-order propagated perturbation density has the same frequency in the q-dimension as the initial perturbation.

5.6.1.3 Second Order

For the second-order propagated perturbation, Equation (5.6.9) yields

$$\phi_{1,2}[\varphi_k] = \mathcal{M}_0 :H[\varphi_k]: \left(:H[\varphi_k]: \Psi_0 + 2\varphi_k\right). \tag{5.6.34}$$

As before, we set $\varphi_k(q,p) = [a_k(q) + \bar{a}_k(q)]\Psi_0(q,p)$ and $\Psi_0(q,p) = \psi(p)$. Here, however, unlike before the second-order propagated perturbation density does not depend linearly on the initial perturbation, so that $\phi_{1,2}[\varphi_k]$ cannot be constructed from the sum of $\phi_{1,2}[a_k\psi]$ and $\phi_{1,2}[\bar{a}_k\psi]$. Instead, using the linearity of $:H[\cdot]:$ we get

$$\begin{aligned}
\phi_{1,2}[\varphi_k] &= \mathcal{M}_0 :H[a_k\psi + \bar{a}_k\psi]: \left(:H[a_k\psi + \bar{a}_k\psi]: \psi + 2a_k\psi + 2\bar{a}_k\psi\right) && (5.6.35)\\
&= \mathcal{M}_0 \left(:H[a_k\psi]: + :H[\bar{a}_k\psi]:\right)\left(:H[a_k\psi]: \psi + :H[\bar{a}_k\psi]: \psi + 2a_k\psi + 2\bar{a}_k\psi\right) && (5.6.36)\\
&= \mathcal{M}_0\Big(:H[a_k\psi]:^2 \psi + :H[a_k\psi]::H[\bar{a}_k\psi]: \psi \\
&\qquad + 2 :H[a_k\psi]: a_k\psi + 2 :H[a_k\psi]: \bar{a}_k\psi\Big) + \text{c.c.}\,. && (5.6.37)
\end{aligned}$$

The terms in the parenthesis can be evaluated using Equations (5.6.25)–(5.6.29):

$$:H[a_k\psi]:^2 \psi = \frac{1}{2}W(k)^2 a_{2k}(q)\frac{\partial^2\psi}{\partial p^2} \tag{5.6.38}$$

$$:H[a_k\psi]::H[\bar{a}_k\psi]: \psi = \frac{1}{4}W(k)W(-k)\frac{\partial^2\psi}{\partial p^2} = \frac{|W(k)|^2}{2}\frac{\partial^2\psi}{\partial p^2} \tag{5.6.39}$$

$$2 :H[a_k\psi]: a_k\psi = W(k)a_{2k}(q)\frac{\partial\psi}{\partial p} \tag{5.6.40}$$

$$2 :H[a_k\psi]: \bar{a}_k\psi = \frac{1}{2}W(k)\frac{\partial\psi}{\partial p}. \tag{5.6.41}$$

We see that the second order perturbation density consists of a term that is independent on the q coordinate, and a term with two times the frequency of the initial perturbation with respect to q:

$$\begin{aligned}\phi_{1,2}[\varphi_k] = \mathcal{M}_0\Bigg(a_{2k}(q) \left[\frac{W(k)^2}{2} \frac{\partial^2}{\partial p^2} + W(k) \frac{\partial}{\partial p} \right] + \\ \left[\frac{|W(k)|^2}{4} \frac{\partial^2}{\partial p^2} + \frac{W(k)}{2} \frac{\partial}{\partial p} \right] \Bigg) \psi + \text{c.c.} \, .\end{aligned} \tag{5.6.42}$$

5.6.1.4 Higher Order

To investigate propagated perturbation densities of higher order, we write Equation (5.6.9) as

$$\phi_{1,n}[\varphi_k] = \mathcal{M}_0 \left({:}H[\varphi_k]{:}^n \, \psi + n \, {:}H[\varphi_k]{:}^{n-1} \, \varphi_k \right), \tag{5.6.43}$$

which is a sum of two terms involving powers of ${:}H[\varphi_k]{:}$. Plugging in the monochromatic perturbation (5.6.11), it can be seen that ${:}H[\varphi_k]{:}^n$ can be written as

$${:}H[\varphi_k]{:}^n = {:}H[(a_k + \bar{a}_k)\psi]{:}^n = ({:}H[a_k\psi]{:} + {:}H[\bar{a}_k\psi]{:})^n. \tag{5.6.44}$$

Because the above Lie operators commute, this expression can be evaluated using the binomial formula:

$${:}H[\varphi_k]{:}^n = \sum_{m=0}^{n} \binom{n}{m} {:}H[a_k\psi]{:}^{n-m} \, {:}H[\bar{a}_k\psi]{:}^m \tag{5.6.45}$$

$$= \frac{1}{2^{n-1}} \sum_{m=0}^{n} \binom{n}{m} W(k)^{n-m} \overline{W(k)^m} \; a_{(n-2m)k}(q) \frac{\partial^n}{\partial p^n} \tag{5.6.46}$$

$$= \sum_{m=0}^{n} S_{n,m}(k) \, a_{[n-2m]k}(q) \frac{\partial^n}{\partial p^n}, \tag{5.6.47}$$

where we define

$$S_{n,m}(k) \equiv \frac{1}{2^{n-1}} \binom{n}{m} W(k)^{n-m} \overline{W(k)^m}. \tag{5.6.48}$$

We see that ${:}H[\varphi_k]{:}^n$ contains terms with frequencies that are multiples of the initial perturbation frequency $j\,k$, where j goes from $-n$ to n in steps of 2. It is worthwhile to rewrite ${:}H[\varphi_k]{:}^n$ with the multiple of the initial frequency as the summation index. A concise way to rewrite the sum in that manner is to let j run from $-n$ to n in steps of 1, but set to zero all those terms, where $j \notin \{-n, -n+2, \ldots, n-2, n\}$. This can be achieved by defining

$$\hat{S}_{n,j} \equiv \begin{cases} S_{n,(n-j)/2} & \text{if } j \in \{-n, -n+2, \ldots, n-2, n\} \\ 0 & \text{otherwise} \, . \end{cases} \tag{5.6.49}$$

We note that $\hat{S}_{n,-j} = \overline{\hat{S}_{n,j}}$. The values of $\hat{S}_{n,j}$ up to $n = 5$ are summarized in Table 5.1. The sum can then be written as

$${:}H[\varphi_k]{:}^n = \sum_{j=-n}^{n} \hat{S}_{n,j}(k) \, a_{jk}(q) \frac{\partial^n}{\partial p^n}. \tag{5.6.50}$$

Applying this to the monochromatic perturbation density yields

$$:H[\varphi_k]:^n \varphi_k =:H[\varphi_k]:^n (a_k + \bar{a}_k)\psi \tag{5.6.51}$$

$$= \frac{1}{2}\sum_{j=-n}^{n} \hat{S}_{n,j}(k)\left[a_{[j+1]k}(q) + a_{[j-1]k}(q)\right]\frac{\partial^n}{\partial p^n}\psi. \tag{5.6.52}$$

Because it is $\hat{S}_{n-1,n} = \hat{S}_{n-1,-n} = 0$, also the term $:H[\varphi_k]:^{n-1} \varphi_k$ in Equation (5.6.43) can be written as a sum from $-n$ to n

$$:H[\varphi_k]:^{n-1} \varphi_k = \frac{1}{2}\sum_{j=-(n-1)}^{n-1} \hat{S}_{n-1,j}(k)\left[a_{[j+1]k}(q) + a_{[j-1]k}(q)\right]\frac{\partial^{n-1}}{\partial p^{n-1}}\psi \tag{5.6.53}$$

$$= \frac{1}{2}\sum_{j=-n}^{n} \hat{S}_{n-1,j}(k)\left[a_{[j+1]k}(q) + a_{[j-1]k}(q)\right]\frac{\partial^{n-1}}{\partial p^{n-1}}\psi. \tag{5.6.54}$$

This sum can again be reordered with respect to terms of equal periodicity. It can be seen that

$$\sum_{j=-n}^{n} \hat{S}_{n-1,j}\, a_{[j+1]k} = \sum_{j=-n+1}^{n+1} \hat{S}_{n-1,j-1}\, a_{jk} \tag{5.6.55}$$

$$= \sum_{j=-n}^{n} \hat{S}_{n-1,j-1}\, a_{jk} + \hat{S}_{n-1,n}\, a_{[n+1]k} - \hat{S}_{n-1,-n-1}\, a_{-nk} = \sum_{j=-n}^{n} \hat{S}_{n-1,j-1}\, a_{jk}, \tag{5.6.56}$$

where we used that $\hat{S}_{n-1,n} = \hat{S}_{n-1,-n-1} = 0$. Analogously, it is

$$\sum_{j=-n}^{n} \hat{S}_{n-1,j}\, a_{[j-1]k} = \sum_{j=-n}^{n} \hat{S}_{n-1,j+1}\, a_{jk}. \tag{5.6.57}$$

Plugging everything back into Equation (5.6.43), an explicit expression for the nth-order propagated perturbation density is obtained

$$\phi_{1,n}[\varphi_k] = \mathcal{M}_0\left(:H[\varphi_k]:^n \Psi_0 + n:H[\varphi_k]:^{n-1} \varphi_k\right) \tag{5.6.58}$$

$$= \mathcal{M}_0 \sum_{j=-n}^{n} a_{jk}(q)\left(\hat{S}_{n,j}(k)\frac{\partial^n}{\partial p^n} + \frac{n}{2}\left[\hat{S}_{n-1,j-1}(k) + \hat{S}_{n-1,j+1}(k)\right]\frac{\partial^{n-1}}{\partial p^{n-1}}\right)\psi \tag{5.6.59}$$

$$= \mathcal{M}_0 \sum_{j=-n}^{n} a_{jk}(q)\left(\hat{S}_{n,j}(k)\frac{\partial^n}{\partial p^n} + \hat{T}_{n,j}(k)\frac{\partial^{n-1}}{\partial p^{n-1}}\right)\psi, \tag{5.6.60}$$

where we define

$$\hat{T}_{n,j}(k) \equiv n\,\frac{\hat{S}_{n-1,j-1}(k) + \hat{S}_{n-1,j+1}(k)}{2}. \tag{5.6.61}$$

We note that $\hat{T}_{n,-j} = \overline{\hat{T}_{n,j}}$. The values of $\hat{T}_{n,j}$ up to $n = 5$ are summarized in Table 5.2. This shows that each $\phi_{1,n}$ can be written as the sum of factorizable functions, to which the unperturbed Perron–Frobenius operator is applied:

$$\phi_{1,n}(q,p) = \sum_{j=-n}^{n} \mathcal{M}_0\big(a_{jk}(q)\, f_{n,j}(k,p)\big), \tag{5.6.62}$$

where the q-dependent factor a_{jk} of each summand is periodic with a multiple of the initial perturbation frequency, and the p-dependent factor is a combination of the nth and the $(n-1)$th derivative of the energy distribution

$$f_{n,j}(k,p) = \hat{S}_{n,j}(k)\frac{\partial^n}{\partial p^n}\psi(p) + \hat{T}_{n,j}(k)\frac{\partial^{n-1}}{\partial p^{n-1}}\psi(p). \tag{5.6.63}$$

Table 5.1: Values of $\hat{S}_{n,j}$ for $n \in [1, \ldots, 5]$.

$\hat{S}_{n,j}$	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
$n=1$	0	W	0	0	0	0
$n=2$	$W\,\overline{W}$	0	$\frac{W^2}{2}$	0	0	0
$n=3$	0	$\frac{3\,W^2\,\overline{W}}{4}$	0	$\frac{W^3}{4}$	0	0
$n=4$	$\frac{3\,W^2\,\overline{W}^2}{4}$	0	$\frac{W^3\,\overline{W}}{2}$	0	$\frac{W^4}{8}$	0
$n=5$	0	$\frac{5\,W^3\,\overline{W}^2}{8}$	0	$\frac{5\,W^4\,\overline{W}}{16}$	0	$\frac{W^5}{16}$

Table 5.2: Values of $\hat{T}_{n,j}$ for $n \in [1, \ldots, 5]$.

$\hat{T}_{n,j}$	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
$n=1$	0	1	0	0	0	0
$n=2$	$\overline{W}+W$	0	W	0	0	0
$n=3$	0	$\frac{6\,W\,\overline{W}+3\,W^2}{4}$	0	$\frac{3\,W^2}{4}$	0	0
$n=4$	$\frac{3\,W\,\overline{W}^2+3\,W^2\,\overline{W}}{2}$	0	$\frac{3\,W^2\,\overline{W}+W^3}{2}$	0	$\frac{W^3}{2}$	0
$n=5$	0	$\frac{15\,W^2\,\overline{W}^2+10\,W^3\,\overline{W}}{8}$	0	$\frac{20\,W^3\,\overline{W}+5\,W^4}{16}$	0	$\frac{5\,W^4}{16}$

5.6.2 Charge Densities

Having derived expressions for the propagated perturbation densities, in the following it is shown how the corresponding charge densities can be determined. Generally, the charge density $\rho[\Psi]$ is defined as the integral of phase-space density $\Psi(q,p)$ over the conjugate momentum

$$\rho[\Psi](q) \equiv \int_{\mathbb{R}} \Psi(q,p)\,\mathrm{d}p. \tag{5.6.64}$$

As this projection operation is linear, it can be seen that the charge density of the Fréchet–Taylor expansion (5.2.8) of a perturbed phase-space density is given by the sum of the propagated perturbation densities plus that of the propagated unperturbed density

$$\rho[\mathcal{M}[\Psi_0+\epsilon\phi_0](\Psi_0+\epsilon\phi_0)] = \rho[\mathcal{M}_0\Psi_0] + \sum_{n=1}^{N}\frac{1}{n!}\epsilon^n\rho[\phi_{1,n}] + \epsilon^N o\left(\|\phi_0\|^N\right). \tag{5.6.65}$$

In Section 5.6.1, we have seen that all propagated perturbation densities $\phi_{1,n}$ can be written as a sum of factorizable functions, to which the undisturbed Perron–Frobenius operator is applied, so that it is

$$\rho[\phi_{1,n}] = \sum_{j=-n}^{n}\rho[\mathcal{M}_0\big(a_{jk}(q)\,f_{n,j}(k,p)\big)]. \tag{5.6.66}$$

The unperturbed Perron–Frobenius operator $\mathcal{M}_0$ can be expressed as the composition of a linear drift-, symplectic scaling-, and a linear kick-operator, according to Equation (5.6.2).

Under general kicks – and therefore also for the linear linear kicks considered here– the charge density is invariant: for any $h \in \mathbb{R}$ and any $\Psi \in \mathcal{W}$ it is

$$\rho[\mathcal{K}_h\Psi] = \int_{\mathbb{R}}\Psi(q,p-hq)\mathrm{d}p = \int_{\mathbb{R}}\Psi(q,p)\mathrm{d}p = \rho[\Psi]. \tag{5.6.67}$$

Scaling the phase-space density also scales the charge density in a measure-preserving way:

$$\rho[\mathcal{S}_C\Psi](q) = \int_{\mathbb{R}} \Psi(q\,C, p/C)\mathrm{d}p = C\int_{\mathbb{R}} \Psi(q\,C, p)\mathrm{d}p = C\,\rho[\Psi](q\,C). \tag{5.6.68}$$

Only drift maps influence the charge density in a more intricate way. For an arbitrary phase-space density, the expression

$$\rho[\mathcal{D}_\beta\Psi] = \int_{\mathbb{R}} \Psi(q-\beta p, p)\mathrm{d}p, \tag{5.6.69}$$

does not admit any immediately useful simplifications. However, for a factorizable density of the form $\chi(q,p) = a_k(q)f(p)$, it is

$$\rho[\mathcal{D}_\beta\chi] = \int_{\mathbb{R}} a_k(q-\beta p)\,f(p)\mathrm{d}p = a_k(q)\int_{\mathbb{R}} e^{-ik\beta p}\,f(p)\mathrm{d}p = a_k(q)\tilde{f}(k\beta). \tag{5.6.70}$$

As expected, the charge density retains the spatial frequency k of the phase-space density. The amplitude of the charge density depends on the drift factor β and the frequency k via the Fourier transform of the function f.

With this, we see that

$$\begin{aligned}\rho[\mathcal{M}_0\chi](q) &= \rho[\mathcal{K}_{Ch}\,\mathcal{S}_C\,\mathcal{D}_{C\beta}\chi](q) = \rho[\mathcal{S}_C\,\mathcal{D}_{C\beta}\chi](q) &(5.6.71)\\ &= C\rho[\mathcal{D}_{C\beta}\chi](q\,C) &(5.6.72)\\ &= Ca_k(q\,C)\tilde{f}(kC\beta). &(5.6.73)\end{aligned}$$

From Equation (5.6.63), it can be seen that the Fourier transform of $f_{n,j}(k,p)$ with respect to p is given by

$$\tilde{f}_{n,j}(k,\omega) = \left[(i\omega)^n\hat{S}_{n,j}(k) + (i\omega)^{n-1}\hat{T}_{n,j}(k)\right]\tilde{\psi}(\omega). \tag{5.6.74}$$

Using this, the summands in Equation (5.6.66) can be written as

$$\begin{aligned}&\rho[\mathcal{M}_0\big(a_{jk}(q)\,f_{n,j}(k,p)\big)] = C\,a_{jk}(q\,C)\,\tilde{f}_{n,j}(k, jkC\beta) &(5.6.75)\\ &= a_{jkC}(q)\,C\left[(ijkC\beta)^n\hat{S}_{n,j}(k) + (ijkC\beta)^{n-1}\hat{T}_{n,j}(k)\right]\tilde{\psi}(jkC\beta). &(5.6.76)\end{aligned}$$

Therefore, Equation (5.6.66) yields an explicit expression for the charge density of the nth order propagated phase-space density in form of an Fourier series:

$$\begin{aligned}\rho[\phi_{1,n}](q) &= C\sum_{j=-n}^{n}\frac{1}{2}e^{ijkCq}\left[(ijkC\beta)^n\hat{S}_{n,j}(k) + (ijkC\beta)^{n-1}\hat{T}_{n,j}(k)\right]\tilde{\psi}(jkC\beta) &(5.6.77)\\ &= C\sum_{j=-n}^{n}\frac{1}{2}e^{ijkCq}A_{n,j}(k, C\beta), &(5.6.78)\end{aligned}$$

with the Fourier coefficients

$$A_{n,j}(k, C\beta) \equiv \left[(ijkC\beta)^n\hat{S}_{n,j}(k) + (ijkC\beta)^{n-1}\hat{T}_{n,j}(k)\right]\tilde{\psi}(jkC\beta). \tag{5.6.79}$$

We see that all Fourier coefficients are scaled by the value of the Fourier transform $\tilde{\psi}$ of the energy distribution, evaluated at $jkC\beta$, independent on the order n. With this, the Fourier coefficients of the charge density – which in certain contexts are also referred to as the *bunching coefficients* – can be constructed up arbitrary order.

5.6.3 Two-Color Perturbation

In the following, the case is considered where the initial perturbation is not given by a single monochromatic modulation but by two modulations with two individual wavelengths. In the monochromatic case, the phase of the modulation was of no concern, as the unperturbed density is independent on the q-coordinate. When two modulations are considered, their phases need to be accounted for explicitly. To this end, the definition (5.6.11) of a monochromatic perturbation φ_{k_i} is amended to include a phase factor $c_i \in \mathbb{C}, |c_i| = 1$

$$\varphi_{k_i}(q,p) \equiv \left[c_i a_{k_i}(q) + \overline{c}_i \overline{a}_{k_i}(q)\right] \Psi_0(p). \tag{5.6.80}$$

A two-color perturbation function can then be defined as

$$\varphi_{k_1,k_2}(q,p) \equiv \frac{\varphi_{k_1}(q,p) + \varphi_{k_2}(q,p)}{2}. \tag{5.6.81}$$

5.6.3.1 First Order

Using the fact that the first-order propagated density function $\phi_{1,1}[\cdot]$ depends linearly on the perturbation, we immediately see that

$$\phi_{1,1}[\varphi_{k_1,k_2}] = \frac{\phi_{1,1}[\varphi_{k_1}] + \phi_{1,1}[\varphi_{k_2}]}{2}. \tag{5.6.82}$$

This shows that to first order, the two modulations to not interact with each other: the total propagated perturbation density is just the sum of two monochromatic perturbations, propagated individually.

5.6.3.2 Second Order

Inserting the two-color perturbation into the expression (5.6.34) for the second-order propagated perturbation $\phi_{1,2}[\cdot]$ yields

$$\phi_{1,2}[\varphi_{k_1,k_2}] = \mathcal{M}_0 :H[\varphi_{k_1,k_2}]: (:H[\varphi_{k_1,k_2}]: \psi + 2\varphi_{k_1,k_2}) \tag{5.6.83}$$

$$= \frac{1}{4}\mathcal{M}_0 (:H[\varphi_{k_1}]: + :H[\varphi_{k_2}]:) (:H[\varphi_{k_1}]: \psi + :H[\varphi_{k_2}]: \psi + 2\varphi_{k_1} + 2\varphi_{k_2}) \tag{5.6.84}$$

$$= \frac{1}{4}\mathcal{M}_0\Big(:H[\varphi_{k_1}]:^2 \psi + 2 :H[\varphi_{k_1}]: \varphi_{k_1} + :H[\varphi_{k_2}]:^2 \psi + 2 :H[\varphi_{k_2}]: \varphi_{k_2} \tag{5.6.85}$$

$$+ :H[\varphi_{k_1}]::H[\varphi_{k_2}]: \psi + 2 :H[\varphi_{k_1}]: \varphi_{k_2} + :H[\varphi_{k_2}]::H[\varphi_{k_1}]: \psi + 2 :H[\varphi_{k_2}]: \varphi_{k_1}\Big),$$

where we set $\Psi_0 = \psi$ as before. We see that this expression for $\phi_{1,2}[\varphi_{k_1,k_2}]$ contains those terms that correspond to the second-order densities of the two perturbations propagated individually:

$$\mathcal{M}_0 \left(:H[\varphi_{k_i}]:^2 \psi + 2 :H[\varphi_{k_i}]: \varphi_{k_i}\right) \equiv \phi_{1,2}[\varphi_{k_i}]. \tag{5.6.86}$$

To capture the remaining terms, we define

$$\phi^{\mathrm{II}}[\varphi_{k_i}, \varphi_{k_j}] \equiv \mathcal{M}_0 \left(:H[\varphi_{k_i}]::H[\varphi_{k_j}]: \psi + 2 :H[\varphi_{k_i}]: \varphi_{k_j}\right). \tag{5.6.87}$$

With this, the total second-order propagated density can be expressed concisely as

$$\phi_{1,2}[\varphi_{k_1,k_2}] = \frac{1}{4}\left(\phi_{1,2}[\varphi_{k_1}] + \phi_{1,2}[\varphi_{k_2}] + \phi^{\mathrm{II}}[\varphi_{k_1}, \varphi_{k_2}] + \phi^{\mathrm{II}}[\varphi_{k_2}, \varphi_{k_1}]\right). \tag{5.6.88}$$

In this form, the terms corresponding to the solitary evolution of the two perturbations, $\phi_{1,2}[\varphi_{k_1}]$ and $\phi_{1,2}[\varphi_{k_2}]$, are isolated from the from those that contain the interaction

between the two, namely $\phi^{\mathrm{II}}[\varphi_{k_2},\varphi_{k_1}]$ and $\phi^{\mathrm{II}}[\varphi_{k_2},\varphi_{k_1}]$. The $\phi_{1,2}[\varphi_{k_i}]$ terms have already been studied in Section 5.6.1.3. In the following, attention is paid only to the interaction terms ϕ^{II}.

In Equation (5.6.87), the first term contains two Lie operators acting on Ψ_0 and therefore describes the combined effect of the two perturbations on the evolution of the unperturbed density. It can be evaluated to

$$:H[\varphi_{k_i}]::H[\varphi_{k_j}]:\,\Psi_0 = \frac{1}{2}\Big[c_i c_j W(k_i)W(k_j)a_{k_i+k_j}(q) + c_i\overline{c}_j W(k_i)\overline{W(k_j)}a_{k_i-k_j}(q)\Big]\frac{\partial^2}{\partial p^2}\psi(p) + \text{c.c.}\,. \tag{5.6.89}$$

The second term describes the influence of one modulation on the evolution of the other:

$$:H[\varphi_{k_i}]:\,\varphi_{k_j} = c_i W(k_i)a_{k_i}(q)\big[c_j a_{k_j}(q) + \overline{c}_j\overline{a}_{k_j}(q)\big]\frac{\partial}{\partial p}\psi(p) + \text{c.c.} \tag{5.6.90}$$

$$= \frac{1}{2}\big[c_i c_j W(k_i)a_{k_i+k_j}(q) + c_i\overline{c}_j W(k_i)a_{k_i-k_j}(q)\big]\frac{\partial}{\partial p}\psi(p) + \text{c.c.}\,. \tag{5.6.91}$$

By reordering the sum of both terms with respect to subterms of equal periodicity, $a_{k_i+k_j}(q)$ and $a_{k_i-k_j}(q)$, Equation (5.6.87) can be written as

$$\phi^{\mathrm{II}}[\varphi_{k_i},\varphi_{k_j}] = \frac{1}{2}\mathcal{M}_0\bigg(c_i c_j W(k_i)\bigg[W(k_j)\frac{\partial^2\psi}{\partial p^2} + 2\frac{\partial\psi}{\partial p}\bigg]a_{k_i+k_j}(q) \tag{5.6.92}$$

$$+ c_i\overline{c}_j W(k_i)\bigg[\overline{W(k_j)}\frac{\partial^2\psi}{\partial p^2} + 2\frac{\partial\psi}{\partial p}\bigg]a_{k_i-k_j}(q)\bigg) + \text{c.c.}\,. \tag{5.6.93}$$

For the the sum of the two interaction terms in Equation (5.6.88), we finally get

$$\Phi^{\mathrm{II}}[\varphi_{k_i},\varphi_{k_j}] \equiv \phi^{\mathrm{II}}[\varphi_{k_i},\varphi_{k_j}] + \phi^{\mathrm{II}}[\varphi_{k_j},\varphi_{k_i}] = \tag{5.6.94}$$

$$\mathcal{M}_0\bigg(c_i c_j\bigg[W(k_i)W(k_j)\frac{\partial^2\psi}{\partial p^2} + \{W(k_i)+W(k_j)\}\frac{\partial\psi}{\partial p}\bigg]a_{k_i+k_j}(q) \tag{5.6.95}$$

$$+ c_i\overline{c}_j\bigg[W(k_i)\overline{W(k_j)}\frac{\partial^2\psi}{\partial p^2} + \Big\{W(k_i)+\overline{W(k_j)}\Big\}\frac{\partial\psi}{\partial p}\bigg]a_{k_i-k_j}(q)\bigg) + \text{c.c.}\,. \tag{5.6.96}$$

Using Equation (5.6.71), the charge density of Φ^{II} can be calculated. This yields

$$\rho\big[\Phi^{\mathrm{II}}[\varphi_{k_i},\varphi_{k_j}]\big](q) = c_i c_j C A^{\mathrm{II}}(k_i,+k_j,C\beta)a_{k_i+k_j}(Cq) \tag{5.6.97}$$

$$+ c_i c_j C A^{\mathrm{II}}(k_i,-k_j,C\beta)a_{k_i-k_j}(Cq) + \text{c.c.}\,. \tag{5.6.98}$$

where we define

$$A^{\mathrm{II}}(k_i,k_j,C\beta) \equiv \tag{5.6.99}$$
$$\big[(i[k_i+k_j]C\beta)^2 W(k_i)W(k_j) + i[k_i+k_j]C\beta\,\{W(k_i)+W(k_j)\}\big]\,\tilde{\psi}([k_i+k_j]C\beta).$$

Note that $A^{\mathrm{II}}(k_i,-k_j,C\beta)$ indeed produces the correct expression, because it is $W(-k_j) = \overline{W(k_j)}$. These results show that the non-linear terms allow initial modulation waves to interact with each other. Due to the interaction, modulation waves are generated with different frequencies are generated. This concept of *wave mixing* is extensively studied in the context of plasma physics [82, 83]. A similar result – more specific to LSC driven microbunching – for the second-order charge density is presented in [84] using a different approach.

5.7 Multiple Microbunching Steps

In Section 5.6, a single microbunching step was analyzed, which consists of a single collective kick, followed by a shearing and compression of the bunch. In the following, it is investigated how multiple consecutive microbunching steps affect the microbunching instability. Generally, the challenge here lies in the fact that – as we have seen – the first microbunching step changes the charge density of the bunch, which causes higher-harmonics of an initial modulation to appear. These harmonics then contribute to the collective kick that is part of the second mircobunching step. Due to this complication, already after the second step, a closed-form exact solution for the phase-space densities and charge densities can generally not be obtained. We will therefore make use of our perturbative approach to derive an expansion of the densities beyond the first step.

Here, the total perturbed phase-space density of the sth step is denoted by $\hat{\Psi}_s$. In the following, we will consider the expansion of $\hat{\Psi}_s$ with respect to the perturbation parameter:

$$\hat{\Psi}_s = \sum_{n=0}^{N} \frac{1}{n!} \epsilon^n \phi_{s,n} + o\left(\epsilon^N\right). \tag{5.7.1}$$

Note that here the 0th order term $\phi_{s,0}$ is included in the expansion. Before, when only a single microbunching step was considered, this term was taken out of the expansion sum and given it the symbol $\Psi_1 \equiv \phi_{1,0}$ to highlight the fact that it represents the unperturbed phase-space density. In the following, both notations are used interchangeably and we define $\Psi_s \equiv \phi_{s,0}$. Further, it is convenient to also assign a symbol to the perturbation part

$$\Phi_s \equiv \sum_{n=1}^{N} \frac{1}{n!} \epsilon^n \phi_{s,n} + o\left(\epsilon^N\right), \tag{5.7.2}$$

which allows to express the total phase-space density as the sum of the unperturbed and the part contributed by the perturbation

$$\hat{\Psi}_s = \Psi_s + \Phi_s. \tag{5.7.3}$$

The total Perron–Frobenius operator of the sth step, $\mathcal{M}_s[\cdot]$, includes both, the single-particle and the collective part. The total density after the sth step is then given by

$$\hat{\Psi}_s = \mathcal{M}_s[\hat{\Psi}_{s-1}]\hat{\Psi}_{s-1}. \tag{5.7.4}$$

The expression $\mathcal{M}_s[\hat{\Psi}_{s-1}]$ can be expanded using Equation (5.2.8), which yields

$$\mathcal{M}_s[\hat{\Psi}_{s-1}] = \mathcal{M}_s[\Psi_{s-1} + \Phi_{s-1}] = \sum_{m=0}^{M} \frac{1}{m!} D^m \mathcal{M}_s[\Psi_{s-1}]\, \Phi_{s-1}^m + o\left(\|\Phi_{s-1}\|^M\right). \tag{5.7.5}$$

In this expression, the perturbation density after the $(s-1)$st step Φ_{s-1} is substituted by its expansion with respect to the perturbation parameter ϵ. It is important to keep in mind that the term Φ_{s-1}^m in Equation (5.7.5) denotes the m arguments of the m-multi-linear bounded function $D^m\mathcal{M}_s[\Psi_{s-1}]$. It can be seen that for any m-multi-linear bounded mapping $f\colon V^m \to W$, where V, W are normed vector spaces, it is

$$f\left(x + o\left(\|y\|^N\right)\right)^m = f\, x^m + o\left(\|y\|^N\right). \tag{5.7.6}$$

Therefore, we can write

$$\mathcal{M}_s[\hat{\Psi}_{s-1}] = \sum_{m=0}^{M} \frac{1}{m!} D^m \mathcal{M}_s[\Psi_{s-1}] \left(\sum_{n=1}^{N} \frac{1}{n!} \epsilon^n \phi_{s-1,n}\right)^m + o\left(\epsilon^N\right) + o\left(\|\Phi_{s-1}\|^M\right). \tag{5.7.7}$$

Recognizing that $o\left(\|\Phi_{s-1}\|^M\right) = o\left(\epsilon^M\right)$, the sum of the two remainder terms can be written as

$$o\left(\epsilon^N\right) + o\left(\|\Phi_{s-1}\|^M\right) = o\left(\epsilon^{\min(N,M)}\right). \tag{5.7.8}$$

By replacing the occurrence of $\hat{\Psi}_{s-1}$ on which the Perron–Frobenius operator acts, we get

$$\hat{\Psi}_s = \mathcal{M}_s[\hat{\Psi}_{s-1}]\hat{\Psi}_{s-1} = \mathcal{M}_s[\hat{\Psi}_{s-1}]\left(\sum_{l=0}^{N}\frac{1}{l!}\epsilon^l\phi_{s-1,l} + o\left(\epsilon^N\right)\right). \tag{5.7.9}$$

Plugging in the expanded form of $\mathcal{M}_s[\hat{\Psi}_{s-1}]$, this then yields

$$\hat{\Psi}_s = \sum_{m=0}^{M}\frac{1}{m!}D^m\mathcal{M}_s[\Psi_{s-1}]\left(\sum_{n=1}^{N}\frac{1}{n!}\epsilon^n\phi_{s-1,n}\right)^m\sum_{l=0}^{N}\frac{1}{l!}\epsilon^l\phi_{s-1,l} + o\left(\epsilon^{\min(N,M)}\right). \tag{5.7.10}$$

In order to arrive at an expansion of Φ_s, this expression now has to be reordered with respect to equal orders of ϵ. To this end, the sums over n and m is rewritten using a multi-index notation. Using a multi-index $\mathbb{N}_1^m \ni \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, the notation

$$\phi_{s,\boldsymbol{\alpha}} \equiv \phi_{s,\alpha_1}\ldots\phi_{s,\alpha_m} \tag{5.7.11}$$

can be used to construct the arguments of the multi-linear Fréchet derivative. The symbol $\mathbb{N}_0$ refers to the set of natural numbers including 0, and $\mathbb{N}_1$ for the same set without 0. With this multi-index notation, the expression for the evaluated mth Fréchet derivative can be written as a sum over terms with equal power of ϵ:

$$D^m\mathcal{M}_s[\Psi_{s-1}]\left(\sum_{n=1}^{N}\frac{1}{n!}\epsilon^n\phi_{s-1,n}\right)^m = \sum_{n=0}^{N}\epsilon^n\sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_1^m\\|\boldsymbol{\alpha}|=n}}\frac{1}{\boldsymbol{\alpha}!}D^m\mathcal{M}_s[\Psi_{s-1}]\,\phi_{s-1,\boldsymbol{\alpha}}, \tag{5.7.12}$$

where $\boldsymbol{\alpha}! \equiv \prod_{i=1}^m \alpha_i!$ and $|\boldsymbol{\alpha}| \equiv \sum_{i=1}^m \alpha_i$. Substituting this in Equation (5.7.10) yields

$$\hat{\Psi}_s = \sum_{n=0}^{N}\epsilon^n\sum_{m=0}^{M}\sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_1^m\\|\boldsymbol{\alpha}|=n}}\frac{1}{\boldsymbol{\alpha}!}D^m\mathcal{M}_s[\Psi_{s-1}]\,\phi_{s-1,\boldsymbol{\alpha}}\sum_{l=0}^{N}\frac{1}{l!}\epsilon^l\phi_{s-1,l} \tag{5.7.13}$$

$$= \sum_{n=0}^{N}\sum_{l=0}^{N}\epsilon^{n+l}\sum_{m=0}^{M}\sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_1^m\\|\boldsymbol{\alpha}|=n}}\frac{1}{\boldsymbol{\alpha}!\,l!}D^m\mathcal{M}_s[\Psi_{s-1}]\,\phi_{s-1,\boldsymbol{\alpha}}\,\phi_{s-1,l} \tag{5.7.14}$$

$$= \sum_{n=0}^{N}\epsilon^n\sum_{m=0}^{M}\sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_1^m\times\mathbb{N}_0\\|\boldsymbol{\alpha}|=n}}\frac{1}{\boldsymbol{\alpha}!}D^m\mathcal{M}_s[\Psi_{s-1}]\,\phi_{s-1,\boldsymbol{\alpha}}, \tag{5.7.15}$$

where in the last equation we have expanded the domain of the multi-index by $\mathbb{N}_0$ so that also the term $\phi_{s-1,l}$ – on which the evaluated Fréchet derivative acts – can be included in the multi-index notation. From this we see that the terms $\phi_{s,n}$ of the expansion of $\hat{\Psi}_s$ are given by

$$\phi_{s,n} = \sum_{m=0}^{M}\sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_1^m\times\mathbb{N}_0\\|\boldsymbol{\alpha}|=n}}\frac{|\boldsymbol{\alpha}|!}{\boldsymbol{\alpha}!}D^m\mathcal{M}_s[\phi_{s-1,0}]\,\phi_{s-1,\boldsymbol{\alpha}}. \tag{5.7.16}$$

In particular, this shows that terms of the phase-space density of order n after the sth step are given by finite sums containing only combinations of terms of order $\leqslant n$ from the previous step.

For $n = 0$, there is only a single multi-index that contributes to the sum in Equation (5.7.16), namely $\boldsymbol{\alpha} = (0)$. Therefore, the 0th-order term reads

$$\phi_{s,0} = \mathcal{M}_s[\phi_{s-1,0}]\phi_{s-1,0}, \tag{5.7.17}$$

which shows that the unperturbed density after the s-step is given by the unperturbed density at the previous step, propagated with the unperturbed PF operator.

There are two multi-indices that contribute to the first order term $n = 1$ in Equation (5.7.16): $\boldsymbol{\alpha} = (1)$ and $\boldsymbol{\alpha} = (1, 0)$. This gives

$$\phi_{s,1} = \mathcal{M}_s[\phi_{s-1,0}]\,\phi_{s-1,1} + D\mathcal{M}_s[\phi_{s-1,0}]\,\phi_{s-1,1}\,\phi_{s-1,0}. \tag{5.7.18}$$

The two terms correspond to the first-order term propagated with the unperturbed PF operator and the unperturbed density propagated with the first-order perturbed PF operator, respectively.

For orders higher than $n = 1$, the number of allowed multi-indices grows rapidly. For instance, the third-order density is generated by the multi-indices $(3, 0)$, $(2, 1, 0)$, $(1, 2, 0)$, $(1, 1, 1, 0)$, $(2, 1)$, $(1, 1, 1)$, $(1, 2)$, and (3).

5.7.1 Two Steps

In the following, the first-order phase-space density after two microbunching steps is derived, assuming a monochromatic modulation as the initial perturbation. The total Perron–Frobenius operator of each step consists of a collective kick $\mathcal{K}_i[\cdot]$ followed by a non-collective, single-particle part $\mathcal{N}_i$:

$$\mathcal{M}_i[\cdot] = \mathcal{N}_i\,\mathcal{K}_i[\cdot]. \tag{5.7.19}$$

Again, we consider

$$\mathcal{K}_i[\cdot] = \exp(:H_i[\cdot]:) \tag{5.7.20}$$

with a convolution-type Hamiltonian, as in Equation (5.5.11). As shown in Equation (5.6.2), the non-collective part of each step can be decomposed into a drift-, scale-, and a kick-operator

$$\mathcal{N}_i = \mathcal{K}_{C_i h_i}\,\mathcal{S}_{C_i}\,\mathcal{D}_{C_i\beta_i}, \tag{5.7.21}$$

each with its own set of drift-, chirp- and compression-parameters.

As before, a density with no dependence on the q coordinate is chosen for the unperturbed initial phase-space density

$$\Psi_0(q, p) = \phi_{0,0}(q, p) = \psi(p), \tag{5.7.22}$$

with $\int_{\mathbb{R}} \psi(p)\mathrm{d}p = \rho_0$. Equation (5.7.17) shows that the 0th order density after the second step is given by

$$\phi_{2,0} = \mathcal{M}_2[\phi_{1,0}]\phi_{1,0} = \mathcal{M}_2\left[\mathcal{M}_1[\phi_{0,0}]\phi_{0,0}\right]\,\mathcal{M}_1[\phi_{0,0}]\phi_{0,0}. \tag{5.7.23}$$

We saw that $\phi_{0,0}$ does not produce collective kicks: $\mathcal{K}_1[\phi_{0,0}] = \mathrm{Id}$. Hence, it is

$$\phi_{2,0} = \mathcal{M}_2\left[\mathcal{N}_1\phi_{0,0}\right]\,\mathcal{N}_1\phi_{0,0}. \tag{5.7.24}$$

Also the unperturbed density after the first step

$$\phi_{1,0} = \mathcal{N}_1\phi_{0,0} = \mathcal{N}_1\psi(p) = \mathcal{K}_{C_1 h_1}\psi(p/C_i) \tag{5.7.25}$$

produces no collective kicks, as its charge density is still homogeneous

$$\rho[\mathcal{K}_{C_1 h_1}\psi(p/C_i)] = \rho[\psi(p/C_i)] = C_i\rho_0 = const.\,. \tag{5.7.26}$$

Therefore, it is $\mathcal{M}_2[\phi_{1,0}] = \mathcal{N}_2$ and

$$\phi_{2,0} = \mathcal{N}_2\,\mathcal{N}_1\,\phi_{0,0}. \tag{5.7.27}$$

For the first-order density after the second step, Equation (5.7.18) yields

$$\phi_{2,1} = \mathcal{M}_2[\phi_{1,0}]\,\phi_{1,1} + D\mathcal{M}_2[\phi_{1,0}]\,\phi_{1,1}\,\phi_{1,0} = \mathcal{N}_2\,(\phi_{1,1} + :H_2[\phi_{1,1}]:\,\phi_{1,0})\,. \tag{5.7.28}$$

If the initial perturbation is a monochromatic density modulation φ_k, the first-order perturbation density after the first step is given by

$$\phi_{1,1}[\varphi_k] = \phi_{1,1}[a_k\psi] + \phi_{1,1}[\bar{a}_k\psi] = \mathcal{N}_1\,\left(a_k(q)\left[1 + W_1(k)\frac{\partial}{\partial p}\right]\psi(p)\right) + \text{c.c.}\,, \tag{5.7.29}$$

see Equation (5.6.33). Using Equation (5.6.71), the corresponding charge density can be seen to be

$$\rho\left[\phi_{1,1}[\varphi_k]\right] = C_1\,a_k(C_1\,q)\,\mathcal{F}\left\{\left[1 + W_1(k)\frac{\partial}{\partial p}\right]\psi(p)\right\}(kC_1\beta_1) + \text{c.c.} \tag{5.7.30}$$

$$= C_1\,a_{C_1 k}(q)\;\left[1 + i\,kW_1(k)C_1\beta_1\right]\tilde{\psi}(kC_1\beta_1) + \text{c.c.} \tag{5.7.31}$$

$$= \rho_0\,C_1\,g_1\,a_{C_1 k}(q) + \text{c.c.}\,, \tag{5.7.32}$$

where the symbol

$$g_1 \equiv \frac{1}{\rho_0}\left[1 + i\,kW_1(k)C_1\beta_1\right]\tilde{\psi}(kC_1\beta_1) \tag{5.7.33}$$

was introduced. Analogous to the derivation in Section 5.6.1.1, using this expression for the charge density, the Lie operator in Equation (5.7.28) can be seen to be

$$:H_2[\phi_{1,1}[\varphi_k]]:= C_1\,g_1\,W_2(C_1\,k)\,a_{C_1 k}(q)\,\frac{\partial}{\partial p} + \text{c.c.}\,. \tag{5.7.34}$$

For the differential operator $\frac{\partial}{\partial p}$, it can be seen that the following commutation relations hold for any phase-space density $\Psi(q,p)$:

$$\frac{\partial}{\partial p}\mathcal{K}_h\Psi = \frac{\partial}{\partial p}\left[\Psi(q, p - hq)\right] = \left[\frac{\partial}{\partial p}\Psi\right](q, p - hq) = \mathcal{K}_h\frac{\partial}{\partial p}\Psi \tag{5.7.35}$$

$$\frac{\partial}{\partial p}\mathcal{S}_C\Psi = \frac{\partial}{\partial p}\left[\Psi(Cq, p/C)\right] = \frac{1}{C}\left[\frac{\partial}{\partial p}\Psi\right](Cq, p/C) = \frac{1}{C}\mathcal{S}_C\frac{\partial}{\partial p}\Psi \tag{5.7.36}$$

$$\frac{\partial}{\partial p}\mathcal{D}_\beta\Psi = \frac{\partial}{\partial p}\left[\Psi(q - \beta p, p)\right] = \left[-\beta\frac{\partial}{\partial q} + \frac{\partial}{\partial p}\Psi\right](q - \beta p, p) = \mathcal{D}_\beta\left[\frac{\partial}{\partial p} - \beta\frac{\partial}{\partial q}\right]\Psi. \tag{5.7.37}$$

With this, we determine

$$\frac{\partial}{\partial p}\phi_{1,0} = \frac{\partial}{\partial p}\mathcal{N}_1\phi_{0,0} = \frac{\partial}{\partial p}\mathcal{K}_{C_1 h_1}\,\mathcal{S}_{C_1}\,\mathcal{D}_{C_1\beta_1}\,\psi(p) \tag{5.7.38}$$

$$= \frac{1}{C_1}\mathcal{K}_{C_1 h_1}\,\mathcal{S}_{C_1}\,\mathcal{D}_{C_1\beta_1}\frac{\partial}{\partial p}\psi(p) = \frac{1}{C_1}\mathcal{N}_1\frac{\partial}{\partial p}\psi(p). \tag{5.7.39}$$

Therefore, the term $:H_2[\phi_{1,1}]:\,\phi_{1,0}$ in Equation (5.7.28) is given by

$$:H_2[\phi_{1,1}]:\,\phi_{1,0} = g_1\, W_2(C_1\, k)\, a_{C_1 k}(q)\, \mathcal{N}_1 \frac{\partial}{\partial p}\psi(p) + \text{c.c.}\,. \tag{5.7.40}$$

Note that in this expression, the single-particle operator of the first stage $\mathcal{N}_1$ acts only on the derivative of the energy distribution and in particular not on the $a_{C_1 k}(q)$ term. For reasons that become apparent in a moment, this term is rewritten in a way so that $\mathcal{N}_1$ acts on both, the p-dependent term as well as the q-dependent term. This can be achieved by noticing that it is

$$a_{C_1 k}(q)\, \mathcal{N}_1 \frac{\partial}{\partial p}\psi(p) = \mathcal{N}_1 \left(\left[\mathcal{N}_1^{-1} a_{C_1 k}(q)\right] \frac{\partial}{\partial p}\psi(p) \right). \tag{5.7.41}$$

Applying the inverse operator $\mathcal{N}_1^{-1}$ to $a_{C_1 k}(q)$ yields

$$\begin{aligned} \mathcal{N}_1^{-1} a_{C_1 k}(q) &= \mathcal{D}_{-C_1\beta_1}\, \mathcal{S}_{1/C_1}\, \mathcal{K}_{-C_1 h_1}\, a_{C_1 k}(q) = \mathcal{D}_{-C_1\beta_1}\, \mathcal{S}_{1/C_1}\, a_{C_1 k}(q) && (5.7.42)\\ &= \mathcal{D}_{-C_1\beta_1}\, a_k(q) = a_k(q)\, \exp(ik\, C_1\beta_1 p). && (5.7.43) \end{aligned}$$

Plugging everything back into Equation (5.7.28), we finally arrive at

$$\phi_{2,1} = \mathcal{N}_2\mathcal{N}_1 \left(a_k(q) \left[1 + \left\{ W_1(k) + g_1\, W_2(C_1\, k)\, \exp(ik\, C_1\beta_1 p) \right\} \frac{\partial}{\partial p} \right] \psi(p) \right) + \text{c.c.}\,. \tag{5.7.44}$$

Following the same approach as in Section 3.4.2, the two-stage Perron–Frobenius operator can be expressed in terms of a single kick-, scale-, and drift-operator

$$\begin{aligned} \mathcal{N}_{2\leftarrow 0} \equiv \mathcal{N}_2\mathcal{N}_1 &= \mathcal{K}_{C_2 h_2}\, \mathcal{S}_{C_2}\, \mathcal{D}_{C_2\beta_2}\, \mathcal{K}_{C_1 h_1}\, \mathcal{S}_{C_1}\, \mathcal{D}_{C_1\beta_1} && (5.7.45)\\ &= \mathcal{K}_{C_2^* h_2^*}\, \mathcal{S}_{C_2^\dagger}\, \mathcal{D}_{\beta_2^\dagger}, && (5.7.46) \end{aligned}$$

where

$$\begin{aligned} h_2^* &= h_2 + C_1 h_1 && (5.7.47)\\ C_2^* &= \frac{1}{1 + h_2^*\beta_2} && (5.7.48)\\ C_2^\dagger &= C_2^* C_1 && (5.7.49)\\ \beta_2^\dagger &= C_1\beta_1 + C_1^2 C_2^*\beta_2. && (5.7.50) \end{aligned}$$

Hence, $\mathcal{N}_{2\leftarrow 0}^{-1}\phi_{2,1}$ is factorizable with respect to its dependence on q and p. The charge density of $\phi_{2,1}$ can therefore be determined using Equation (5.6.71), which yields

$$\rho[\phi_{2,1}] = C_2^\dagger\; a_k\left(C_2^\dagger q\right)\; \tilde{f}\left(k\,\beta_2^\dagger\right), \tag{5.7.51}$$

where $\tilde{f}$ is the Fourier transform of the p-dependent part of $\mathcal{N}_{2\leftarrow 0}^{-1}\phi_{2,1}$:

$$\tilde{f}(\omega) = \mathcal{F}_{\omega\leftarrow p}\left(\left[1 + \left\{ W_1(k) + g_1\, W_2(C_1\, k)\, \exp(ik\, C_1\beta_1 p) \right\} \frac{\partial}{\partial p} \right] \psi(p) \right). \tag{5.7.52}$$

Using the identities

$$\mathcal{F}_{\omega\leftarrow p} \frac{\partial}{\partial p}\psi(p) = i\omega\, \tilde{\psi}(\omega) \tag{5.7.53}$$

and

$$\mathcal{F}_{\omega\leftarrow p}\, \exp(ikp) X(p) = \int_{\mathbb{R}} \delta(k - \omega')\, \tilde{X}(\omega - \omega')\, \mathrm{d}\omega' = \tilde{X}(\omega - k) \tag{5.7.54}$$

gives

$$\tilde{f}(\omega) = [1 + i\omega W_1(k)]\,\tilde{\psi}(\omega) + i(\omega - k\,C_1\beta_1)\; g_1\, W_2(C_1\,k)\;\tilde{\psi}(\omega - k\,C_1\beta_1). \tag{5.7.55}$$

Noticing that $\beta_2^\dagger - C_1\beta_1 = C_1^2 C_2^* \beta_2 = C_2^\dagger C_1 \beta_2$ we then arrive at the final expression for the first-order charge density after the second step

$$\rho[\phi_{2,1}] = C_2^\dagger\; a_k\Big(C_2^\dagger q\Big)\;\; g_2^\dagger + \text{c.c.}\;, \tag{5.7.56}$$

where the symbol $g_2^\dagger$ is defined as

$$g_2^\dagger \equiv \big[1 + ik\,\beta_2^\dagger W_1(k)\big]\tilde{\psi}(k\,\beta_2^\dagger)\;\; + i\,g_1\,C_2^\dagger C_1\beta_2 k\; W_2(C_1\,k)\;\tilde{\psi}(C_2^\dagger C_1\beta_2 k). \tag{5.7.57}$$

In Section 5.8.5, a two-step microbunching gain function is derived from this expression for $g_2^\dagger$, which is plotted in Figure 6.12.

5.8 Application to LSC-driven Microbunching

Until now, the results obtained from the perturbation theory are of strictly mathematical nature and only little context for the underlying physical process was given. It is therefore expedient to further investigate the resulting expressions, and determine the implications for the microbunching instability. While the perturbation theory was formulated for an arbitrary impedance function, in the following the focus is on the effect of longitudinal space charge.

In the perturbation theory, no assumptions about the energy distribution ψ were made. In the following, the case is considered where the initial energy distribution is a Gaussian distribution with zero mean and variance σ_p^2:

$$\psi(p) = \rho_0 \frac{\xi(p/\sigma_p)}{\sigma_p}, \quad \text{where } \xi(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \tag{5.8.1}$$

The parameter σ_p is referred to as the *energy spread*. Derivatives of the energy distribution can then be expressed via the probabilist's Hermite polynomials

$$\frac{\mathrm{d}^n\psi}{\mathrm{d}p^n}(p) = \frac{(-1)^n \mathrm{He}_n(p/\sigma_p)\psi(p)}{\sigma_p^n}, \tag{5.8.2}$$

where

$$\mathrm{He}_n(x) \equiv (-1)^n \xi(x)^{-1} \frac{\mathrm{d}^n\xi}{\mathrm{d}x^n}(x). \tag{5.8.3}$$

The Fourier transform of ψ is given by

$$\tilde{\psi}(\omega) \equiv \int_{\mathbb{R}} \psi(p) e^{-i\omega p}\mathrm{d}p = \frac{\rho_0}{\sigma_p}\int_{\mathbb{R}} \xi(p/\sigma_p) e^{-i\omega p}\mathrm{d}p = \rho_0 \int_{\mathbb{R}} \xi(x) e^{-i\sigma_p\omega x}\mathrm{d}x = \rho_0 \tilde{\xi}(\sigma_p\omega). \tag{5.8.4}$$

Using the well-known identity

$$\tilde{\xi}(x) = \sqrt{2\pi}\xi(x) \tag{5.8.5}$$

shows that

$$\tilde{\psi}(\omega) = \sqrt{2\pi}\rho_0 \xi(\sigma_p\omega). \tag{5.8.6}$$

Until now, also the impedance function $W(k)$ was considered an arbitrary function and not further specified. As we work in a one-dimensional, longitudinal model,

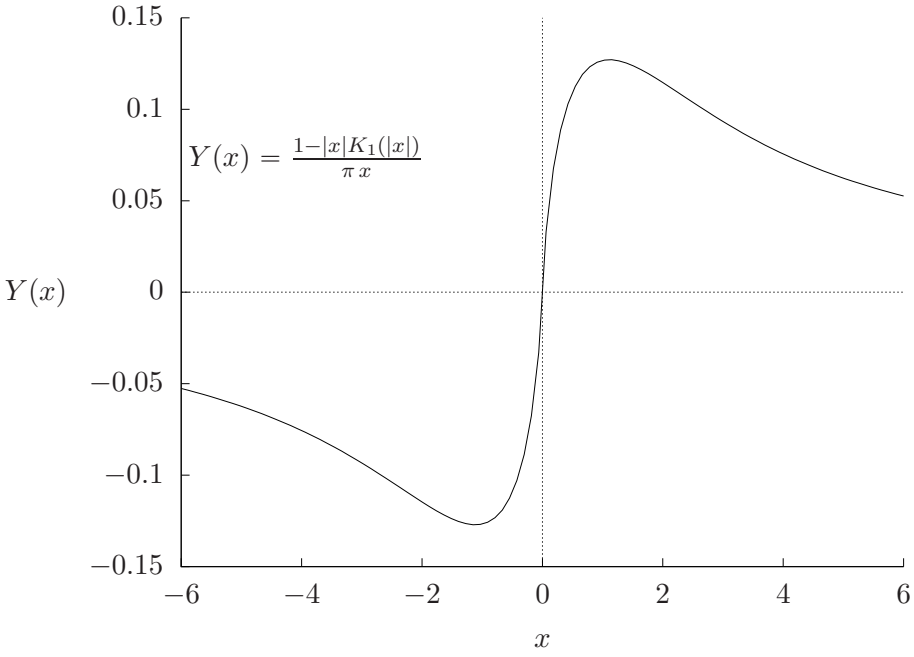

Figure 5.1: Plot of the longitudinal impedance function of a bunch with radially uniform charge distribution.

assumptions about the transverse shape of the bunch have to be made to derive this longitudinal impedance function. For certain physically relevant models of the transverse bunch shape, an impedance function can be determined analytically, see [70, 71]. One frequently considered model is that of a cylindrical bunch with a transverse charge distribution that is radially uniform. In that case, the impedance function of a non-relativistic beam is given by

$$W(k) = i\,\frac{l\rho_0 N_b e^2}{\varepsilon_0}\,\frac{1-|rk|K_1(|rk|)}{\pi\, r^2\, k} = i\, eU_0\, Y(rk), \tag{5.8.7}$$

where $l \in \mathbb{R}^+$ is the path length of the section, $r \in \mathbb{R}^+$ is the radius of the transverse distribution. The quantity U_0 is defined as

$$U_0 \equiv \frac{l\rho_0 N_b e}{r\varepsilon_0} = \frac{l}{r}\, I_0\, Z_0, \tag{5.8.8}$$

where $I_0 = c\rho_0 N_b e$ is the beam current, with N_b being the number of electrons in the bunch, $Z_0 \equiv \frac{1}{\varepsilon_0 c} \approx 376.73\,\text{Ohm}$ is the free-space impedance. Further, the auxiliary function Y is defined as

$$Y(x) \equiv \frac{1-|x|K_1(|x|)}{\pi\, x}, \tag{5.8.9}$$

which is plotted in Figure 5.1. When relativistic effects are to be taken into account, the argument of the impedance function has to be scaled by the Lorentz factor γ:

$$W(k) \to \frac{1}{\gamma}\, W\left(\frac{k}{\gamma}\right). \tag{5.8.10}$$

With the energy distribution and impedance function fixed, we can now have a closer look at the results for the propagated perturbation densities.

5.8.1 First-Order Propagated Density

The first-order propagated perturbation density of resulting from a monochromatic initial perturbation is given by Equation (5.6.33) and reads, after plugging in ψ and Z,

$$\phi_{1,1}[\varphi_k] = \frac{\rho_0}{\sigma_p}\mathcal{M}_0 \left\{ a_k(q) \left[1 - i\frac{eU_0}{\gamma\sigma_p} Y\left(\frac{rk}{\gamma}\right) \mathrm{He}_1\left(\frac{p}{\sigma_p}\right) \right] \xi\left(\frac{p}{\sigma_p}\right) \right\} + \text{c.c.}\,. \tag{5.8.11}$$

As shown in Equation (5.6.2), the single-particle Perron–Frobenius operator is given by $\mathcal{M}_0 = \mathcal{K}_{Ch}\,\mathcal{S}_C\,\mathcal{D}_{C\beta}$. To investigate the microbunching effect, the chirp imprinted on the phase-space density by the $\mathcal{K}_{Ch}$ operator is, in this context, irrelevant. We therefore define the *dechirped* phase-space density

$$\grave{\phi}_{1,1} = \mathcal{K}_{-Ch}\,\phi_{1,1}, \tag{5.8.12}$$

where this correlation is removed. As already shown in Equation (5.6.67), the charge-density is equal for both, the chirped and the unchirped phase-space density:

$$\rho[\grave{\phi}_{1,1}] = \rho[\phi_{1,1}]. \tag{5.8.13}$$

Applying the drift and scaling parts of $\mathcal{M}_0$, $\grave{\phi}_{1,1}$ can be written explicitly as

$$\grave{\phi}_{1,1}[\varphi_k] = \frac{\rho_0}{\sigma_p} e^{iCkq} e^{ik\beta p} \left[1 - iA\,\mathrm{He}_1\left(\frac{p}{C\sigma_p}\right) \right] \xi\left(\frac{p}{C\sigma_p}\right) + \text{c.c.}\,, \tag{5.8.14}$$

where the relative modulation amplitude

$$A \equiv \frac{eU_0}{\gamma\sigma_p} Y\left(\frac{rk}{\gamma}\right) \tag{5.8.15}$$

was introduced. This dimensionless parameter describes the amplitude of the energy modulation that a bunch with unity density modulation would experience – normalized to the energy spread of the bunch. We therefore refer to it as the *amplitude parameter*. Further, the dimensionless parameter

$$B \equiv kC\beta\sigma_p \tag{5.8.16}$$

is defined, which describes the dispersive shift of the longitudinal coordinate a particle with an energy deviation equal to the energy spread of the bunch would experience – normalized with respect to the wavenumber of the modulation. It is a measure for the amount of shearing the magnetic chicane induces on the phase-space density. Hence, it shall be called the *shearing parameter*. Note that both, the shearing and the amplitude parameter, depend on the wavenumber k of the modulation. With this, the phase-space density can be written most concisely as

$$\grave{\phi}_{1,1}[\varphi_k] = \frac{\rho_0}{\sigma_p} e^{iCkq} e^{iB\frac{p}{C\sigma_p}} \left[1 - iA\,\mathrm{He}_1\left(\frac{p}{C\sigma_p}\right) \right] \xi\left(\frac{p}{C\sigma_p}\right) + \text{c.c.}\,. \tag{5.8.17}$$

This representation of the phase-space density is particularly handy, as it depends only on 4 dimensionless quantities: the longitudinal coordinate Ckq, compressed and normalized with the wavenumber of the modulation; the energy coordinate $\frac{p}{C\sigma_p}$, normalized to the compressed energy spread; the amplitude parameter A; and the shearing parameter B.

An expression for the charge-density is obtained from Equation (5.6.78):

$$\rho[\phi_{1,1}] = \frac{C}{2}A_{1,0}(k, C\beta) + C\left[\frac{1}{2}e^{ikCq}A_{1,1}(k, C\beta) + \text{c.c.}\right] \tag{5.8.18}$$

$$= \frac{C}{2}e^{ikCq}A_{1,1}(k, C\beta) + \text{c.c.} = \frac{C}{2}e^{ikCq}[ikC\beta\hat{S}_{1,1} + \hat{T}_{1,1}]\tilde{\psi}(kC\beta) + \text{c.c.} \tag{5.8.19}$$

$$= \frac{C}{2}e^{ikCq}[ikC\beta W(k) + 1]\tilde{\psi}(kC\beta) + \text{c.c.} \tag{5.8.20}$$

$$= \sqrt{\frac{\pi}{2}}C\rho_0 e^{ikCq}[1 - kC\beta\, eU_0 Y(rk)]\xi(kC\beta\sigma_p) + \text{c.c.}\,, \tag{5.8.21}$$

where the coefficients $A_{n,j}$ from Equation (5.6.78) are not to be confused with the dimensionless quantity A defined in Equation (5.8.15). By substituting A and B, this can finally be written as

$$\rho[\phi_{1,1}] = \frac{1}{2}C\rho_0 e^{ikCq}[1 - AB]e^{-\frac{1}{2}B^2} + \text{c.c.} \tag{5.8.22}$$

$$= C\rho_0 \cos(kCq)[1 - AB]e^{-\frac{1}{2}B^2}. \tag{5.8.23}$$

5.8.1.1 Normalized First-Order Gain Function

Equation (5.8.23) shows that the amplitude of the propagated charge-density modulation depends on the parameters A and B via the function

$$G_1(A, B) \equiv (1 - AB)e^{-\frac{1}{2}B^2}, \tag{5.8.24}$$

which we will call the first-order *normalized gain function*. It describes the amplification – or suppression – of the microbunching amplitude in dependence on the perturbation wavelength. We call it the normalized gain function to distinguish it from the more conventional notion of a gain function $G_1(k)$ which describes the dependence of the gain on the wavenumber k. Figure 5.2 show plots of $G_1(A, B)$ for different values of A. The conventional gain function can be constructed from the normalized gain function by explicitly resolving the dependence of the normalized parameters on k, which is investigated in more detail in later sections.

In the following, we investigate some features of the normalized gain function and see how they can be understood in context with the underlying phase-space dynamics. For this discussion, it is with out loss of generality assumed that A is positive. To facilitate better understanding, Figures 5.5 and 5.6 show plots of $\grave{\phi}_{1,1}[\varphi_k]$ together with the charge density $\rho[\phi_{1,1}[\varphi_k]]$ for two values of A and multiple values of B.

The center-left subplots in both Figures show the $B = 0$ case, which means that no shearing of the phase-space has occurred. For both, $A = 1$ and $A = 3$, $\phi_{1,1}$ can be visualized as a periodic sequence of connected regions, alternating between positive and negative values. In the $A = 1$ case, the shape of these regions is close to elliptical, with some positive chirp. Whereas for $A = 3$, the regions are hour-glass shaped, with two pronounced extrema connected by a "waist" of lower absolute density. Also in this case, the regions show positive chirp.

We note that the normalized gain function has a zero crossing at $AB = 1$, which implies that the microbunching completely vanishes at first order for this combination of parameters. This situation is shown in the center-right subplots of Figures 5.5 and 5.6. Due to the shearing, regions of positive and negative density start to overlap more. Figuratively speaking, for $B > 0$, the upper part of a positive region moves in the direction of the lower part of the adjacent negative region, and vice versa. If $AB = 1$,

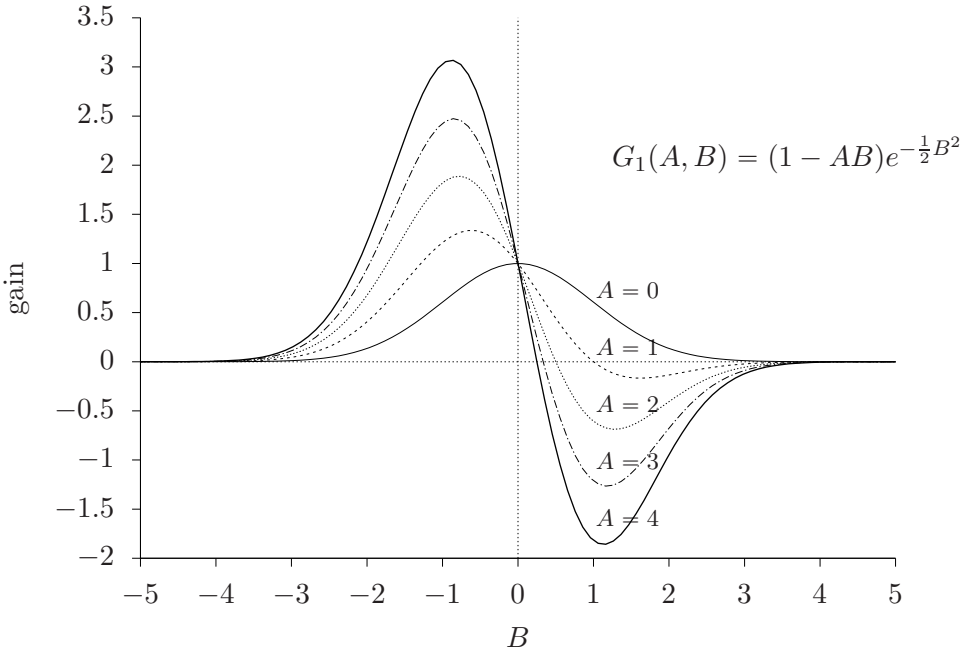

Figure 5.2: Plot of $G_1(A, B)$ as a function of B for various values of A.

the positive and negative regions overlap precisely so that their contributions to the charge density cancel out exactly at all longitudinal positions.

For $A \neq 0$, the normalized gain function features two extrema. For a given value of A, the extrema are located at

$$B_{\pm}(A) = \frac{1 \pm \sqrt{4A^2 + 1}}{2A}. \tag{5.8.25}$$

It can be seen that one of the extrema occurs at a positive value of the shearing parameter, while the other occurs at negative shearing. In Figure 5.3, a plot of the value of the two extrema is shown.

The extremum at B_- is always a maximum. In Figures 5.5 and 5.6, this case is depicted in the top-right subplots. Here, the negative shearing causes the initially positively chirped regions to stand upright in phase-space. Regions of positive and negative density overlap only marginally. This results in the maximum amplification of the microbunching amplitude, as only regions of the same sign contribute to the charge density at each longitudinal position and virtually no cancellation occurs. For $B < B_-$, the regions start to overlap again, and the gain decreases. We note that the absolute value of $G_1(A, B_-)$ is always larger than unity. This implies that at $B = B_-$, the microbunching is always amplified, independent on the value of A.

At B_+, the extremum is a minimum, indicating a π phase-shift of the microbunching. It can be seen that the absolute value of $G_1(A, B_+(A))$ is smaller than unity, if $A < 2.54$. This indicates that if the normalized energy modulation amplitude is smaller than 2.54, then the microbunching instability is always suppressed for all positive values of the shearing parameter. If A is larger than this threshold, amplification of the microbunching amplitude occurs. Figures 5.5 and 5.6 help to understand this effect qualitatively: If A is small enough, then the shape of the density regions resembles ellipses. Shearing these ellipses by any positive value causes positive and negative regions to overlap, which results in a reduction of the charge density modulation amplitude. In case A is large,

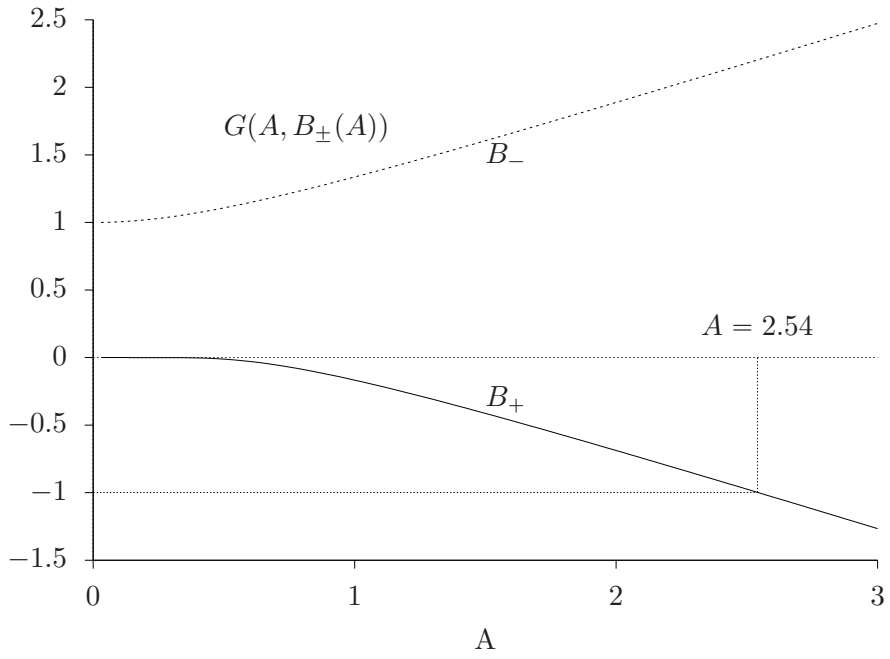

Figure 5.3: Plot of $G_1(A, B_+(A))$ and $G_1(A, B_+(A))$.

the regions resemble hour-glasses more than ellipses: the difference between the value of the phase-space density at the extrema and the value at the waist is large. This geometry allows for an amplifying effect to occur. At a certain value of the shearing parameter, the upper and lower extrema of two neighboring regions of the same sign fully overlap. In particular, this full overlap of the maxima occurs at the position of the waist of the interjacent region of opposite sign. This clearly results in an amplification of the microbunching amplitude, as in the projection the two extrema sum up and the waist of the opposite sign contributes only marginally. This situation is shown in the lower-left subplot of Figure 5.6. If B is larger than this critical value, the regions start to overlap more uniformly, and the microbunching amplitude decreases. Figure 5.4 shows a contour plot of the normalized gain function.

5.8.1.2 Wavelength-Dependent First-Order Gain Function

The normalized gain function $G_1(A, B)$ describes how the geometry of the phase-space dynamics influences the charge-density inhomogeneities, depending on the normalized modulation amplitude A and the shearing parameter B. This description provides valuable, abstract insights into the fundamental microbunching process. In practice, however, one is often times particularly interested in how strong the microbunching instability is amplified – or suppressed – at a certain wavelength $\lambda = \frac{2\pi}{k}$. As hinted at before, the dependence of the microbunching gain on the wavenumber k can be obtained from the normalized gain function by explicitly resolving the dependence of the normalized parameters on k. This then defines the wavenumber-dependent gain function

$$G_1(k) \equiv G_1\big(A(k), B(k)\big). \tag{5.8.26}$$

The normalized shearing parameter, as defined in Equation (5.8.16), depends linearly on the wavenumber $B(k) = B'k$, where we introduce

$$B' \equiv C\beta\sigma_p = \frac{C\underline{M}_{56}\sigma_p}{cP_0}, \tag{5.8.27}$$

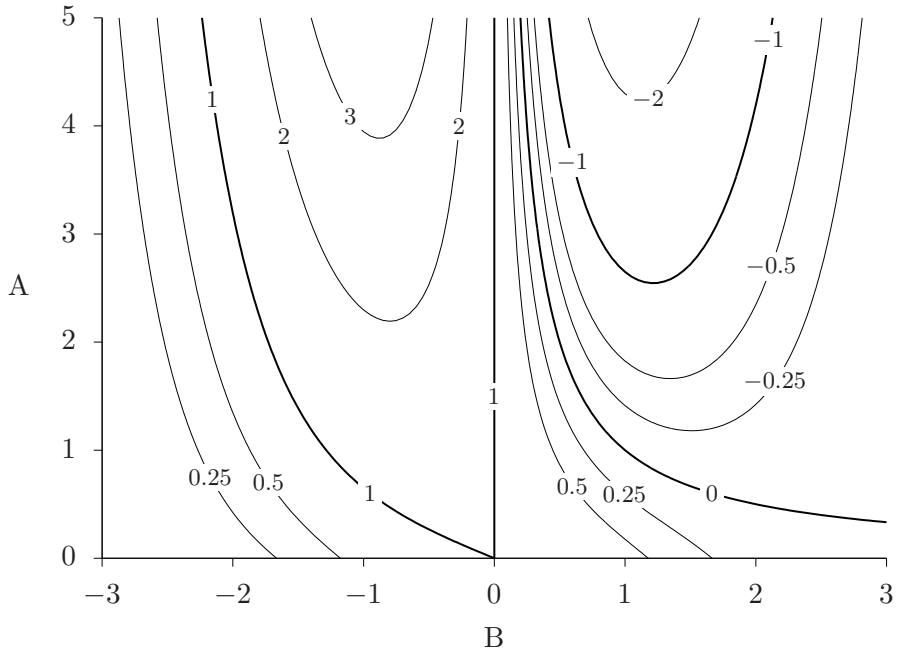

Figure 5.4: Contour plot of $G_1(A, B)$. The contours of $G_1 = -1$, $G_1 = 0$, $G_1 = 1$ are marked with thicker lines.

using the definition of β given in Equation (3.4.3). In practice, typical values of B' lie in the range 0 µm to 100 µm. For instance, the parameter combination $C = 4$, $\underline{M}_{56} = 140\,\text{mm}$, $\sigma_p = 5\,\text{keV}$ and $cP_0 = 146\,\text{MeV}$ yields $B' \approx 19.2\,\mu\text{m}$.

In Equation (5.8.15), the normalized amplitude parameter is defined. By defining

$$A' \equiv \frac{eU_0}{\gamma\sigma_p} = \frac{l}{r}\frac{eI_0Z_0}{\gamma\sigma_p} \quad \text{and} \quad r' \equiv \frac{r}{\gamma} \tag{5.8.28}$$

the amplitude parameter can be written as

$$A(k) = A'\,Y(r'k). \tag{5.8.29}$$

For the parameter combination $l = 25\,\text{m}$, $r = 500\,\mu\text{m}$, $I_0 = 31\,\text{A}$, $cP_0 = 146\,\text{MeV}$($\Longrightarrow$ $\gamma \approx 285.7$) and $\sigma_p = 5\,\text{keV}$, it is $A' \approx 408.8$ and $r' \approx 1.75\,\mu\text{m}$.

Figures 5.7 and 5.9 show plots of the curve $(B(k), A(k))$ for parameters similar to the example above and multiple values of B' and r', respectively, on top of the contour lines of $G_1(A, B)$. In Figures 5.8 and 5.10, plots of $|G_1(k)|$ are shown, which correspond to the value of $|G_1(A, B)|$ along the curves in Figures 5.7 and 5.9, respectively. These Figures can help to get a more intuitive understanding of the effect the parameters A',r', and B' have on the wavelength-dependent gain curve, by investigating how the parameters affect the location of a given wavelength in the A,B space:

The parameter A' scales the $(B(k), A(k))$ curve in the A direction. Increasing A' will therefore increase the maximum gain value and push the location of the maximum towards longer wavelengths.

The parameter B' stretches or shrinks the curve in the B direction. Increasing B' will push a given wavelength towards larger values of B, causing a reduction of the maximum attainable gain. This effect is illustrated by the three curves in Figure 5.7, which correspond to different values of B'.

Variation of the r' parameter will not affect the B-coordinate of a given wavelength, but will change its position along the A axis only, corresponding to a stretching or

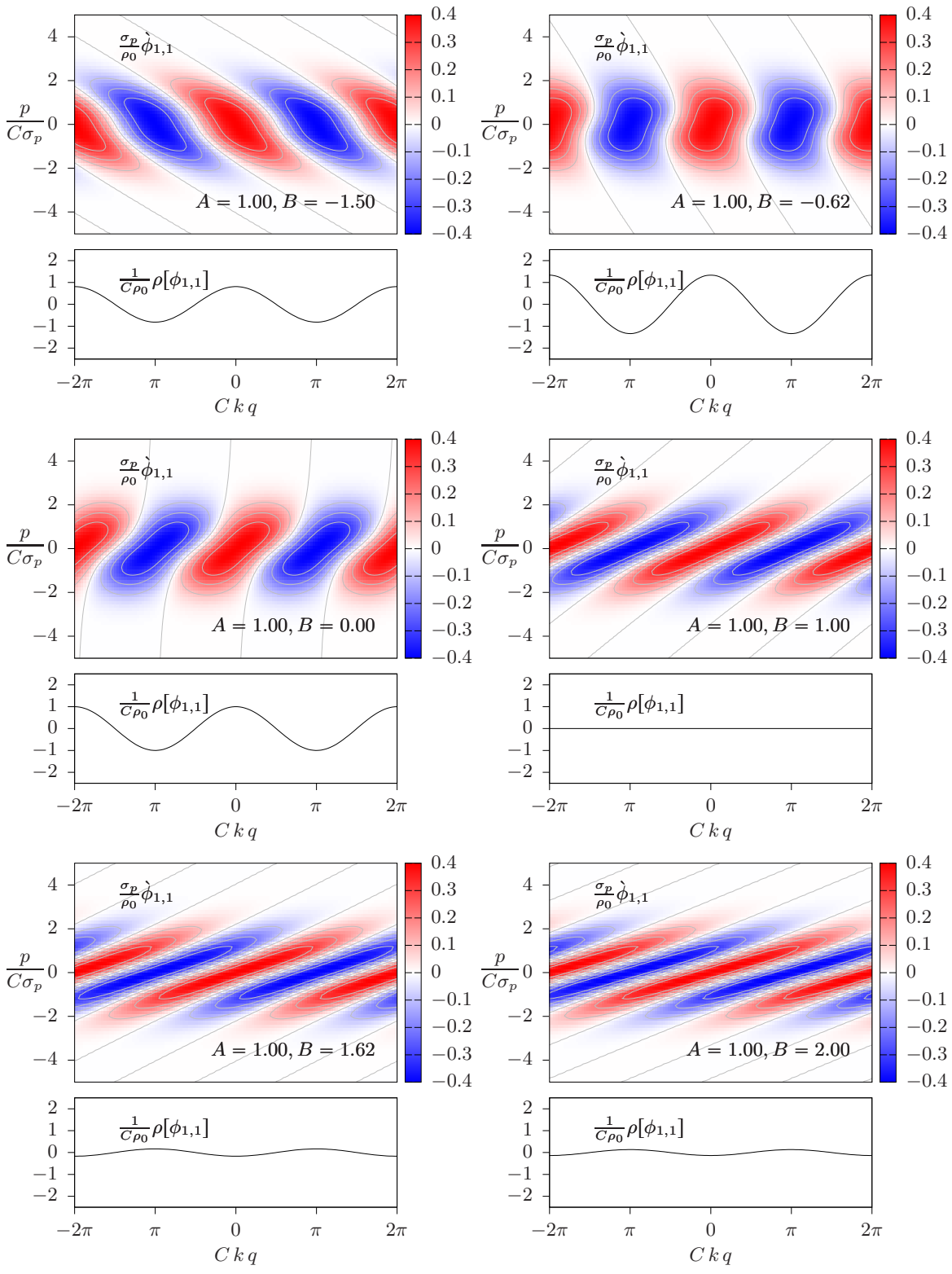

Figure 5.5: Plots of $\dot{\phi}_{1,1}$ and $\rho_{1,1}$ for $A = 1$ and multiple values of B.

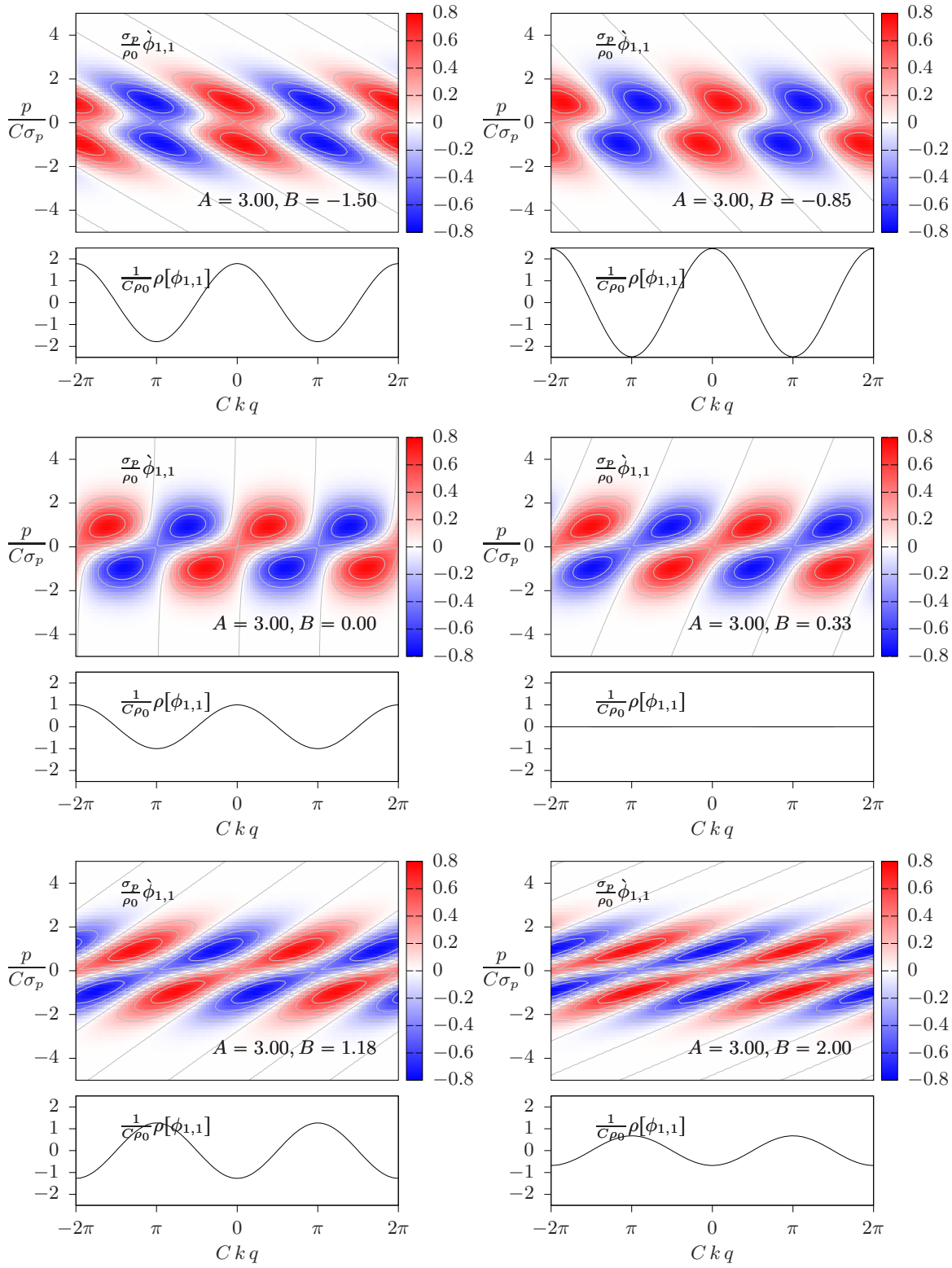

Figure 5.6: Plots of $\dot{\phi}_{1,1}$ and $\rho_{1,1}$ for $A = 3$ and multiple values of B.

shrinking of the $A(k)$ function, as shown in Figure 5.9. This will only slightly affect the location of the maximum. The value of the maximum, however, increases as r' decreases and vice versa.

5.8.2 Second-Order Propagated Density

After analyzing the first-order perturbation density and the resulting gain function, now the second-order perturbation density is considered, which is given in Equation (5.6.42). Plugging in the derivatives (5.8.2) of the energy distribution yields

$$\phi_{1,2}[\varphi_k] = \frac{\rho_0}{\sigma_p}\mathcal{M}_0\Bigg\{a_{2k}(q)\left[\frac{W(k)^2}{2\sigma_p^2}\mathrm{He}_2\left(\frac{p}{\sigma_p}\right) - \frac{W(k)}{\sigma_p}\mathrm{He}_1\left(\frac{p}{\sigma_p}\right)\right] + \frac{1}{2}\left[\frac{|W(k)|^2}{2\sigma_p^2}\mathrm{He}_2\left(\frac{p}{\sigma_p}\right) - \frac{W(k)}{\sigma_p}\mathrm{He}_1\left(\frac{p}{\sigma_p}\right)\right]\Bigg\}\xi\left(\frac{p}{\sigma_p}\right) + \text{c.c.}\,. \tag{5.8.30}$$

As before, the single-particle Perron–Frobenius operator is applied explicitly. Using the parameters A, and B, as defined before, this can be written as

$$\grave{\phi}_{1,2}[\varphi_k] = \frac{\rho_0}{\sigma_p}\Bigg\{e^{i2Ckq}e^{i2B\frac{p}{C\sigma_p}}\left[-\frac{1}{2}A^2\,\mathrm{He}_2\left(\frac{p}{C\sigma_p}\right) - iA\,\mathrm{He}_1\left(\frac{p}{C\sigma_p}\right)\right] \tag{5.8.31}$$

$$+\frac{1}{2}\left[\frac{1}{2}\,|A|^2\,\mathrm{He}_2\left(\frac{p}{C\sigma_p}\right) - iA\,\mathrm{He}_1\left(\frac{p}{C\sigma_p}\right)\right]\Bigg\}\xi\left(\frac{p}{C\sigma_p}\right) + \text{c.c.}\,, \tag{5.8.32}$$

where $\grave{\phi}_{1,2} \equiv \mathcal{K}_{-Ch}\,\phi_{1,2}$ denotes the unchirped variant of $\phi_{1,2}$.

5.8.2.1 Second-Order Gain Function

For the charge density, Equation (5.6.78) yields

$$\rho[\phi_{1,2}](q) = \frac{C}{2}\left\{A_{2,0}(k,C\beta) + \left[e^{ikCq}A_{2,1}(k,C\beta) + e^{i2kCq}A_{2,2}(k,C\beta) + \text{c.c.}\right]\right\}. \tag{5.8.33}$$

With $A_{2,0} = A_{2,1} = 0$, and

$$A_{2,2} = \left[-2(kC\beta Z)^2 + 2ikC\beta Z\right]\tilde{\psi}(2kC\beta) \tag{5.8.34}$$

$$= \sqrt{2\pi}\rho_0\left[2(AB)^2 - 2AB\right]\xi(2B) \tag{5.8.35}$$

$$= 2\rho_0\left[(AB)^2 - AB\right]e^{-2B^2}, \tag{5.8.36}$$

the charge density can be written as

$$\rho[\phi_{1,2}](q) = 2\,C\rho_0\cos(2kCq)\,G_2(A,B), \tag{5.8.37}$$

where the *second-order normalized gain function* G_2 is defined as

$$G_2(A,B) \equiv AB\left[AB-1\right]e^{-2B^2}. \tag{5.8.38}$$

It is sensible to not include the factor $2 = 2!$ in the definition of the gain function, as the contribution of the second-order charge density to the total charge density is $\frac{\epsilon^2}{2!}\rho[\phi_{1,2}]$. This way, G_2 reflects the effective contribution of the second-order charge

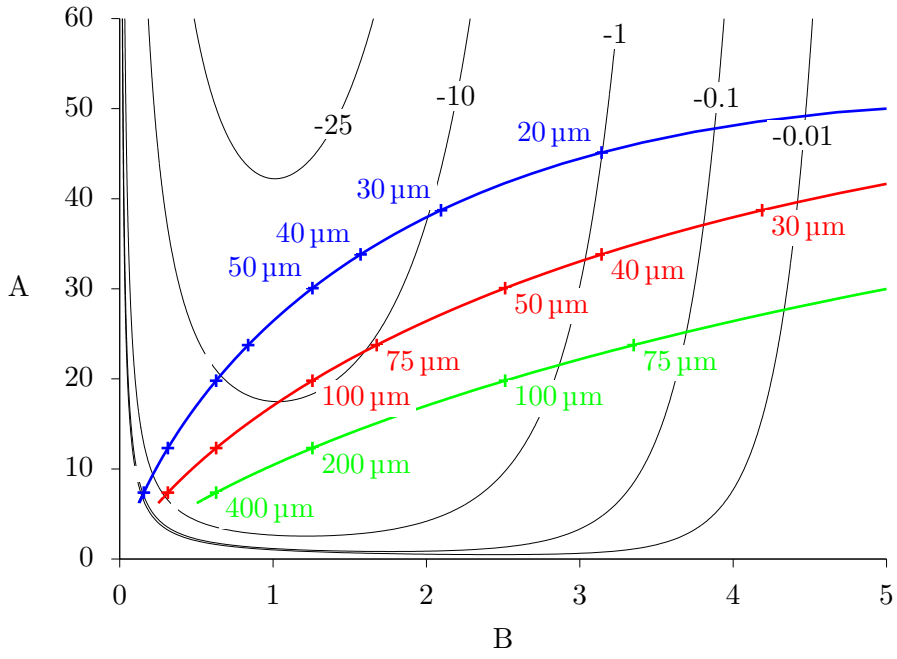

Figure 5.7: Plot of the curves (B(k),A(k)) with the parameters $A' = 400, r' = 1.75\,\mu\text{m}$, and $B' = 10\,\mu\text{m}$ (blue), $B' = 20\,\mu\text{m}$ (red), $B' = 40\,\mu\text{m}$ (green), on top of a contour plot of $G_1(A, B)$.

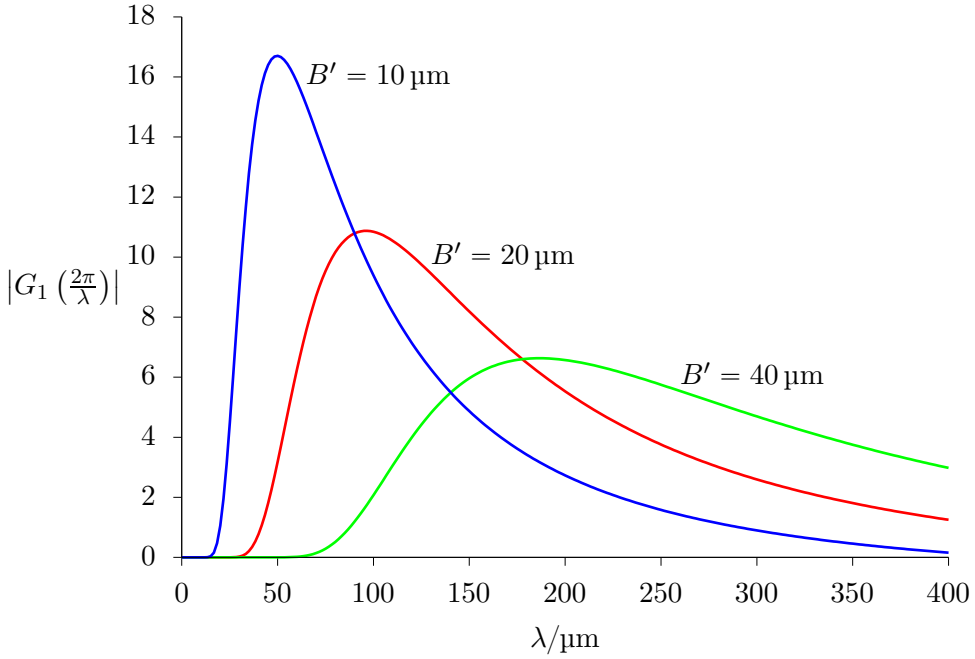

Figure 5.8: Plot of $|G_1(k)|$ for the same parameters as in Figure 5.7.

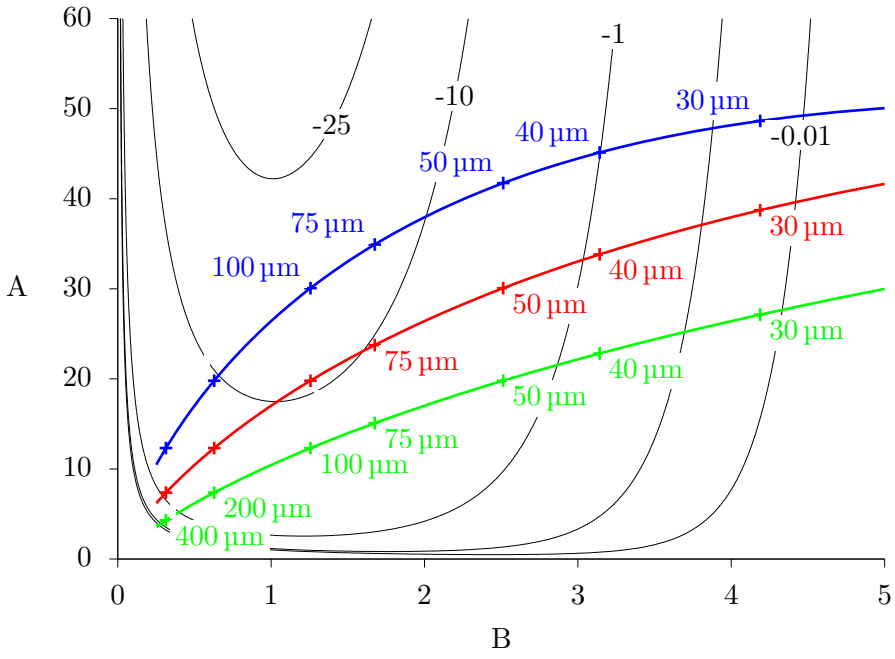

Figure 5.9: Plot of the curves (B(k),A(k)) with the parameters $A' = 400, B' = 20\,\mu\text{m}$, and $r' = 3.50\,\mu\text{m}$ (blue), $r' = 1.75\,\mu\text{m}$ (red), $r' = 0.875\,\mu\text{m}$ (green), on top of a contour plot of $G_1(A, B)$.

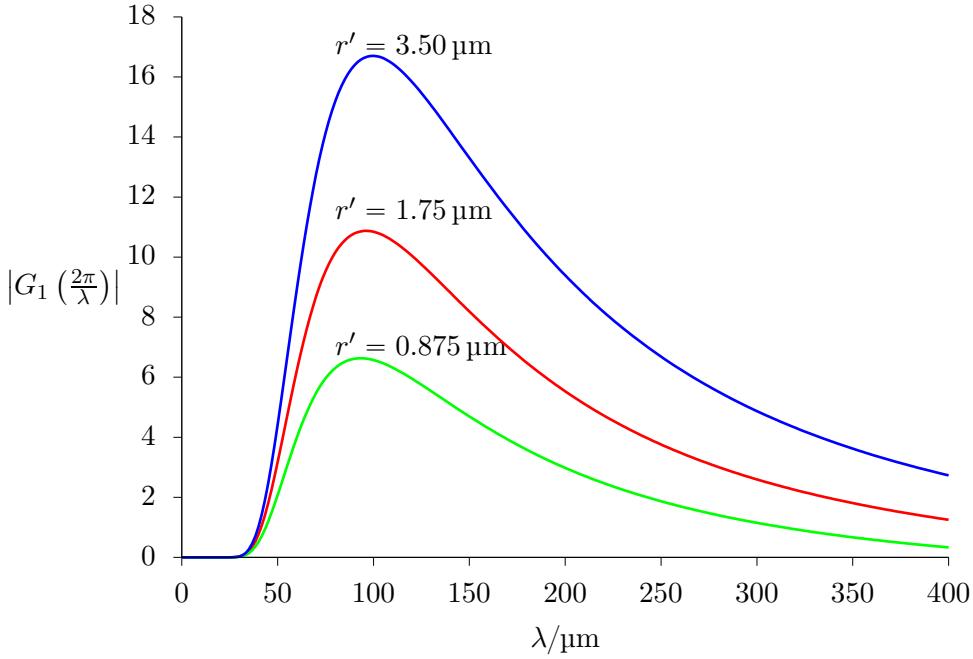

Figure 5.10: Plot of $|G_1(k)|$ for the same parameters as in Figure 5.9.

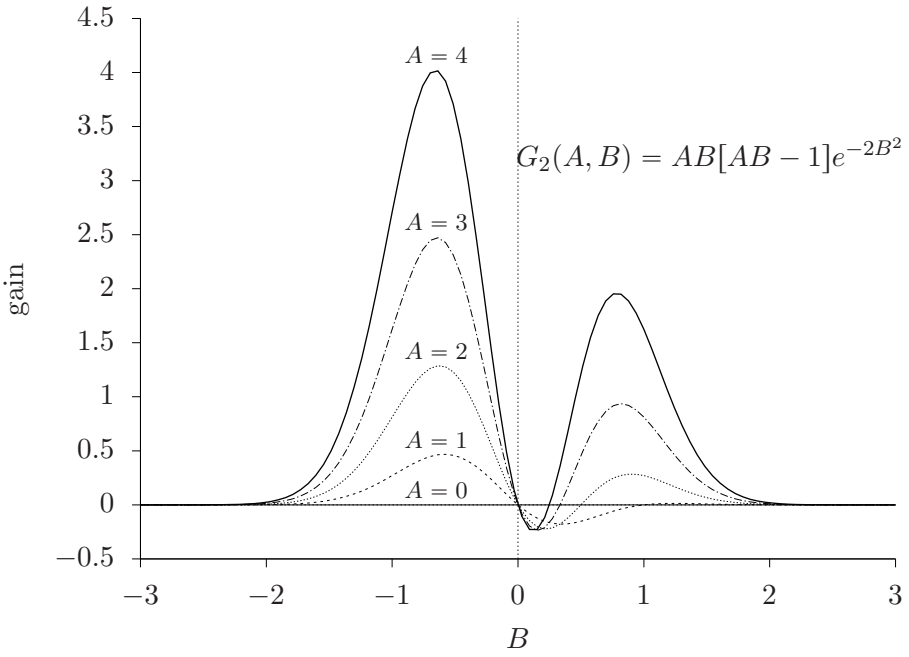

Figure 5.11: Plot of $G_2(A, B)$ as a function of B for various values of A.

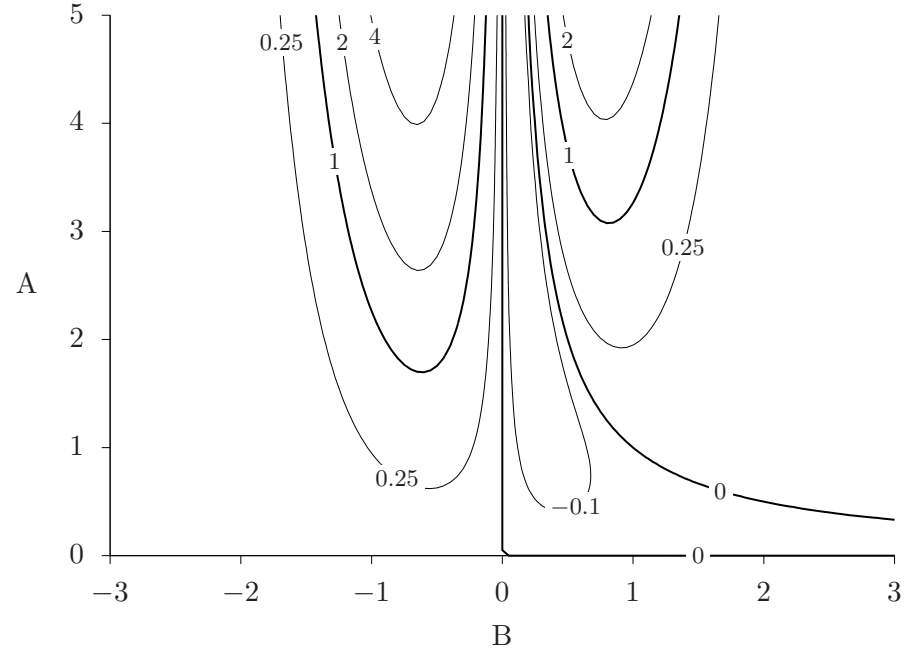

Figure 5.12: Contour plot of $G_2(A, B)$. The contours of $G_2 = 0$, $G_2 = 1$ are marked with thicker lines.

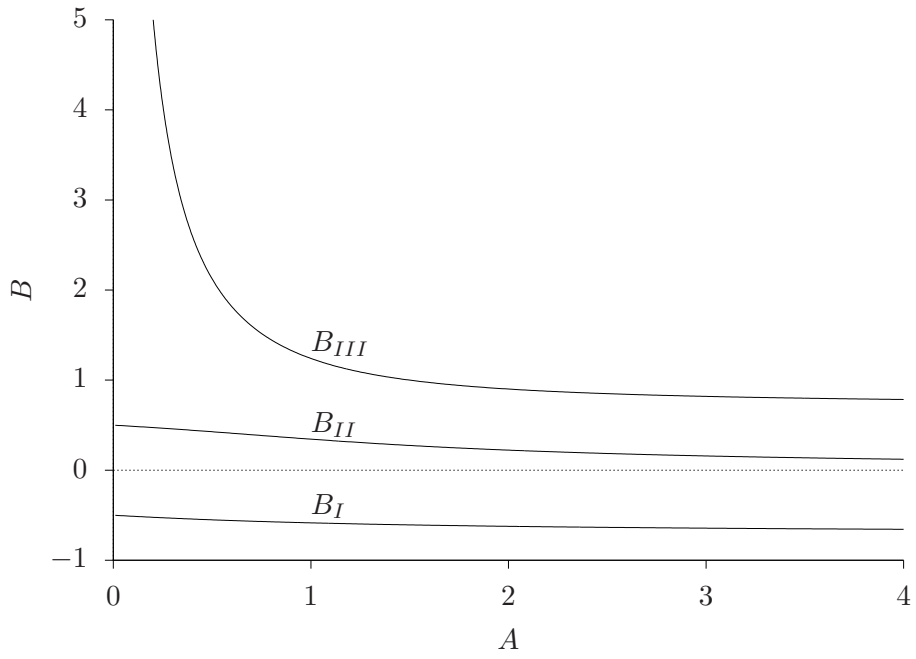

Figure 5.13: Location of the local extrema of $G_2(A, B)$ as a function of A.

density, normalized to ϵ^2. Figure 5.11 shows plots of $G_2(A, B)$ for multiple values of A with B as the independent variable. A contour plot of $G_2(A, B)$ is shown in Figure 5.12.

We note that G_2 has a root at $AB = 1$ – just as G_1. This indicates that at $AB = 1$, both the first-order and the second-order microbunching gain vanish. At this parameter combination the microbunching instability is suppressed up to at least second order.

For a fixed value of A, the three local extrema of G_2 occur at the roots of

$$4\,A^2\,B^3 - 4\,A\,B^2 - 2\,A^2\,B + A. \tag{5.8.39}$$

These roots shall be denoted with B_I, B_{II}, and B_{III}, where $B_I < B_{II} < B_{III}$. In principle, closed-form solutions for the roots of (5.8.39) can be directly obtained as it is a third-order polynomial. Their closed-form expressions, however, are rather complicated and provide only little additional insight. We will therefore omit them here. Figure 5.13 shows a plot of the roots, which were obtained numerically.

Plots of $\grave{\phi}_{1,2}$ and $\rho_{1,2}$ are shown in Figures 5.14 and 5.15, which illustrate the origin of the roots and maxima of the second-order gain function. The second-order phase-space density consists of a periodic sequence of negative-valued "central" regions around $\frac{p}{C\sigma_p} = 0$, each accompanied by two adjacent positive "satellite" regions with around $\frac{p}{C\sigma_p} = \pm 2$ that are slightly shifted towards larger and smaller values of q, respectively. The amount of integrated absolute density contained in the two satellites is larger than that in the single central. At $B = 0$, the satellites overlap with the center in a way so that the charge density vanishes, see center-left subplot. For $B = B_I$, which is shown in the top-right subplots, the two satellites align, leaving the center isolated, which results in the maximum charge density amplitude. If $B = B_{II}$, the satellites align with the central region, see center-right subplots. As the total amount of integrated absolute density is larger in the satellites than in the center, this constellation causes a local extremum, albeit with a smaller amplitude compared to the other two extrema. The extremum at $B = B_{III}$, as shown in the bottom-right subplots, can again be explained by two

satellites aligning between two central region. For small values of A, the amplitude of this extremum is comparatively small, as in this case, the center and satellite regions are less well localized, which results in more overlap. Between B_{II} and B_{III} lies the case where $AB = 1$, see the bottom-left subplots. Here, the satellites are shifted from their alignment with the central region so that their contributions to the charge density cancel out exactly, resulting in the root of the gain function.

5.8.2.2 Wavelength-Dependent Second-Order Gain Function

Analogous to the first-order case, also the wavelength-dependent second-order gain function can be derived by substituting the dependence of the parameters A and B on the wavenumber k in the normalized second-order gain function:

$$G_2(k) \equiv G_2\big(A(k), B(k)\big). \tag{5.8.40}$$

Exemplary plots of G_2 are shown in Figure 5.16.

5.8.3 Third-Order Gain Functions

Above, it was shown how Equations (5.6.58) and (5.6.78) yield explicit expressions for the first- and second-order perturbation densities and gain functions. In principle, the same approach can be used for any perturbation order. However, we restrict ourselves and conclude our analytical investigations in this direction at the third order, for which only the gain function is derived explicitly.

For the third-order coefficients, we get $A_{3,0} = A_{3,2} = 0$ and

$$A_{3,1} = -\frac{3}{4}\left[i(kC\beta)^3 Z^2\overline{Z} + (kC\beta)^2(2Z\overline{Z} + Z^2)\right]\tilde{\psi}(kC\beta) = \rho_0\,\frac{3}{4}(AB)^2(AB-1)e^{-\frac{1}{2}B^2} \tag{5.8.41}$$

and

$$A_{3,3} = -\frac{27}{4}\left[i(kC\beta)^3 Z^3 + (kC\beta)^2 Z^2\right] = -\rho_0\,\frac{27}{4}\,(AB)^2(AB-1)e^{-\frac{9}{2}B^2}. \tag{5.8.42}$$

We see that the third-order charge density contains modulations on the first and third harmonic:

$$\rho[\phi_{1,3}](q) = 3!\,\left[C\rho_0\cos(kCq)\,G_3^I(A,B) + C\rho_0\cos(3kCq)\,G_3^{III}(A,B)\right], \tag{5.8.43}$$

where

$$G_3^I(A,B) \equiv \frac{1}{8}(AB)^2(AB-1)e^{-\frac{1}{2}B^2} \tag{5.8.44}$$

and

$$G_3^{III}(A,B) \equiv -\frac{9}{8}\,(AB)^2(AB-1)e^{-\frac{9}{2}B^2} \tag{5.8.45}$$

describe third-order contribution to the first- and third harmonic, respectively. As before, we see that both gain functions have a root at $AB = 1$.

This shows that at $AB = 1$ the microbunching gain vanishes completely, up to at least the third perturbation order.

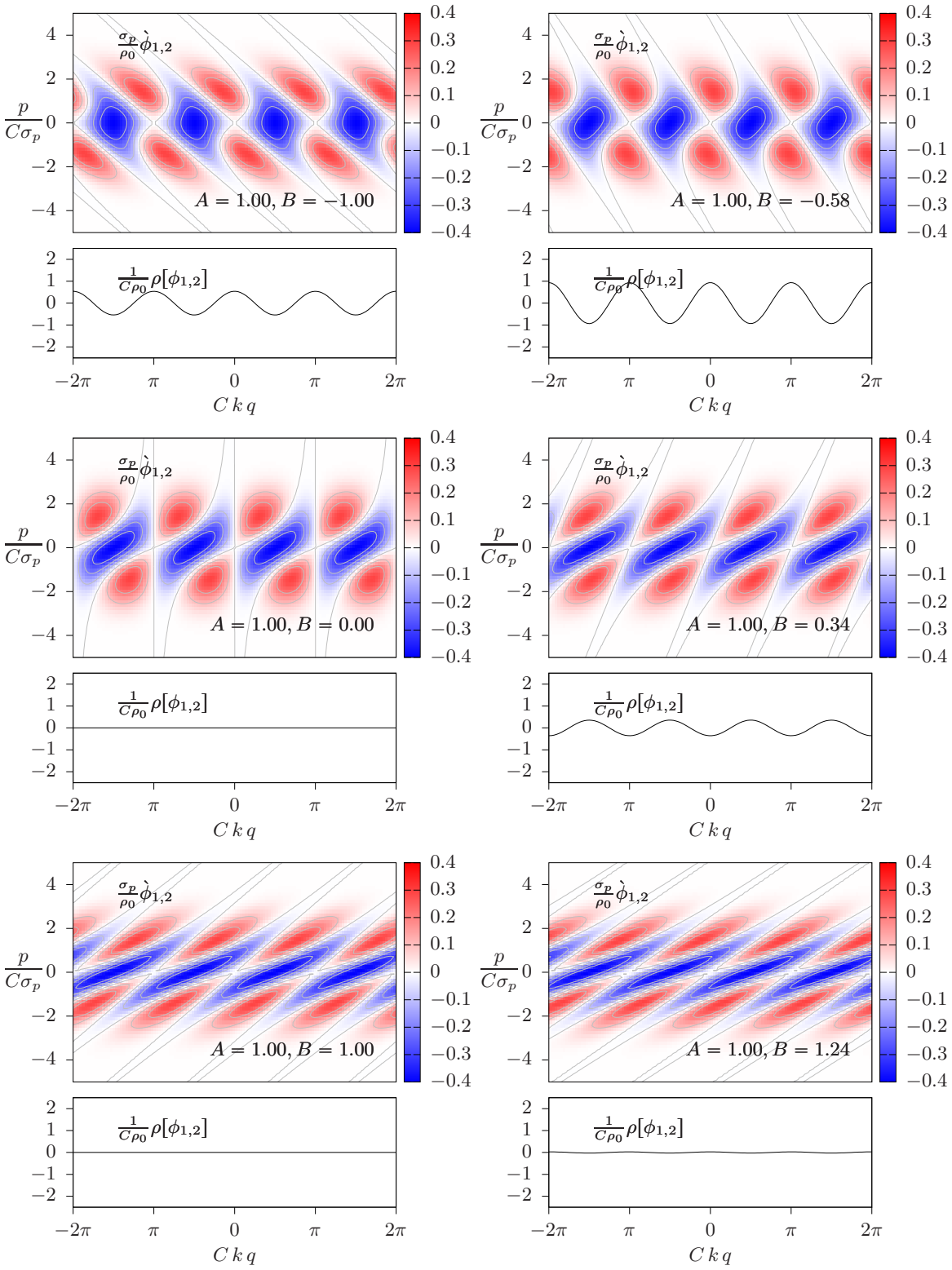

Figure 5.14: Plots of $\dot{\phi}_{1,2}$ and $\rho_{1,2}$ for $A = 1$ and multiple values of B.

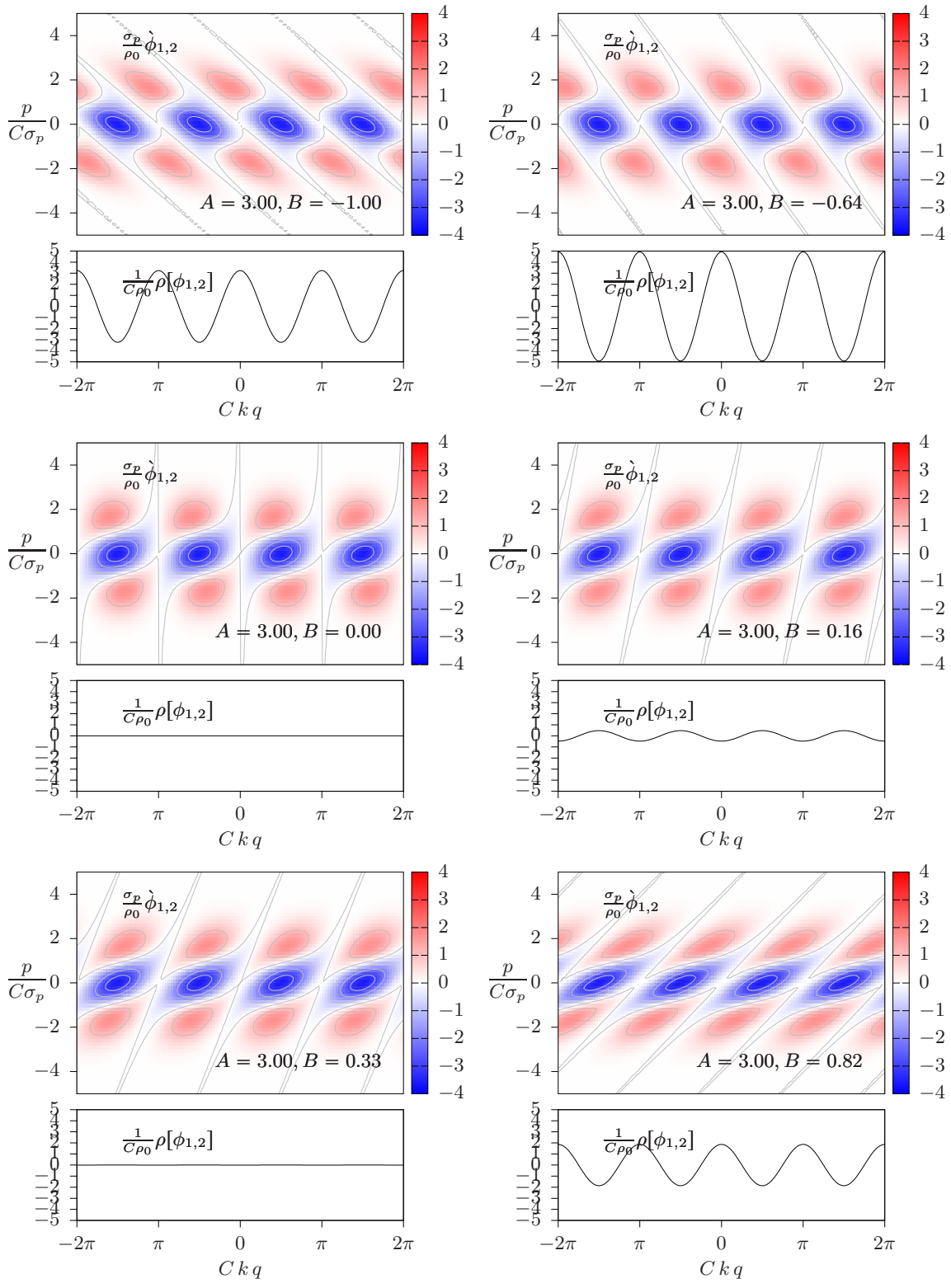

Figure 5.15: Plots of $\grave{\phi}_{1,2}$ and $\rho_{1,2}$ for $A = 3$ and multiple values of B.

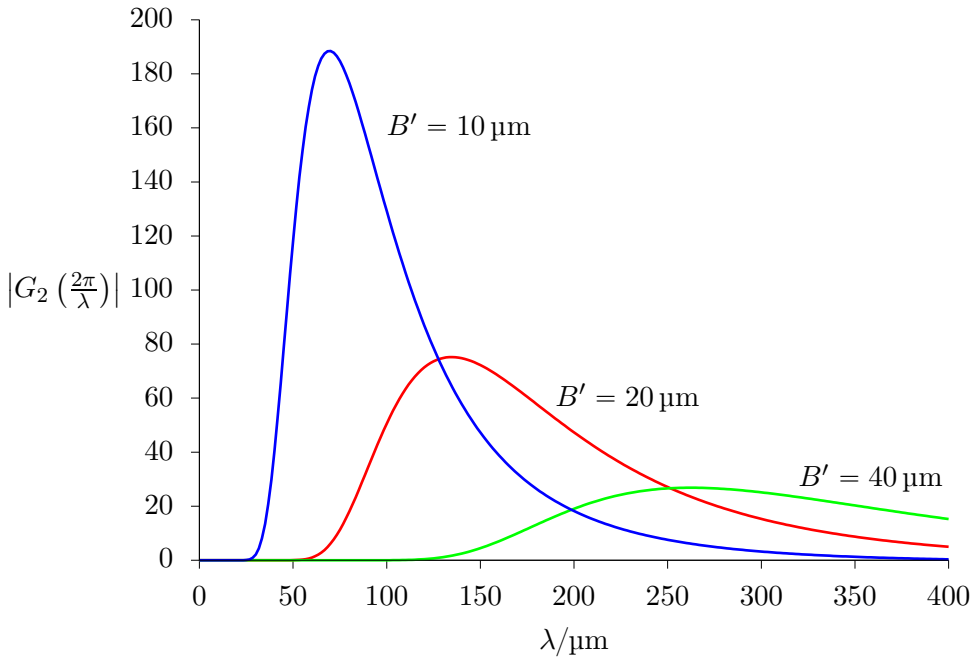

Figure 5.16: Plot of $|G_2(k)|$ for the same parameters as in Figure 5.7.

5.8.4 Phase-Space Density Approximations

With expressions for the nth-order perturbation densities at hand, approximations of the total density can be constructed. For a given perturbation parameter ε, we denote the approximation of the total phase-space density *up to* nth-order with

$$\Phi_{1,N} \equiv \sum_{n=0}^{N} \frac{\varepsilon^n}{n!} \phi_{1,n} \tag{5.8.46}$$

and the corresponding approximate charge density with

$$\mathrm{P}_{1,N} \equiv \sum_{n=0}^{N} \frac{\varepsilon^n}{n!} \rho_{1,n}, \tag{5.8.47}$$

where P here is to be taken as the Greek capital letter Rho. Figure 5.17 shows an illustrative example of the convergence of $\Phi_{1,N}$ and $\mathrm{P}_{1,N}$ towards the exact solution. For the chosen parameters, the fourth-order approximate phase-space density is visually indistinguishable from the exact solution. Already the second-order approximation of the charge density is close to the exact solution. We note that the rate of convergence depends on the normalized parameters A and B, and the value chosen for the perturbation parameter ε. In particular, when the product εA is much larger than unity, an adequate approximation might only be achieved at a much larger order N.

5.8.5 First-Order Two-Step Gain Function

In Section 5.7.1 an expression for the charge density after two microbunching steps was derived, see Equation (5.7.56). It can be seen that the wavenumber of the modulation of this charge density is equal to the wavenumber of the initial perturbation multiplied by the total compression factor $C_2^\dagger$ of the two steps. The complex amplitude $g_2^\dagger$, as defined in Equation (5.7.57), involves two impedance functions: the impedance of the

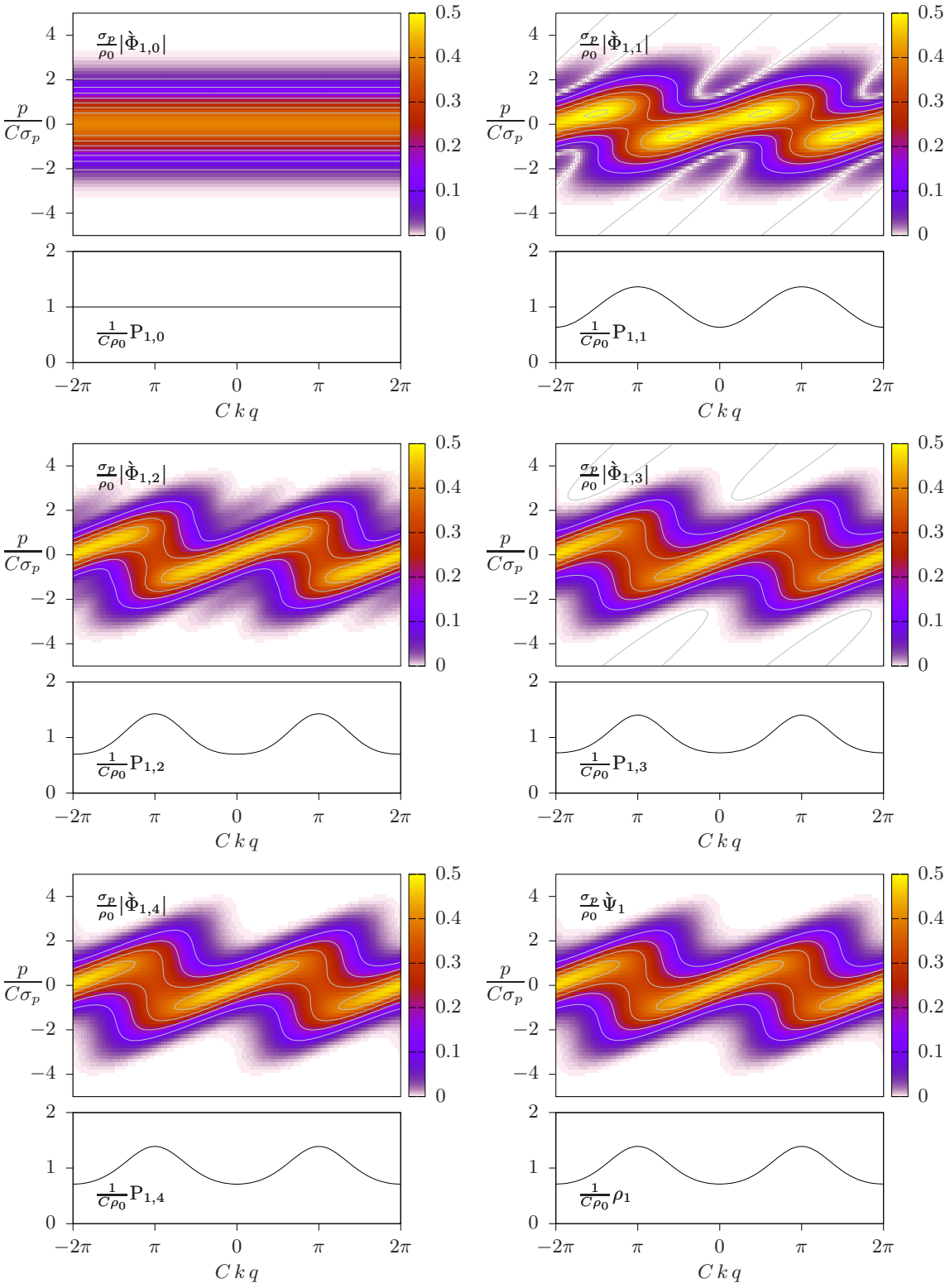

Figure 5.17: Plots of the approximate phase-space densities and charge-densities up to a given order. The bottom-right subplot shows the exact solution of the phase-space density and the charge density that was obtained from it numerically. In this example, the normalized parameters are $A = 4$, $B = 1$ and the perturbation parameter is $\varepsilon = 0.2$.

first step, evaluated at the initial wavenumber $W_1(k)$ and the impedance of the second step, evaluated at the initial wavenumber compressed by the first step $W(C_1 k)$. Now, both of these occurrences are replaced by the LSC impedance

$$W_n(k) = i\frac{eU_n\, Y(r'_n k)}{\gamma_n} \tag{5.8.48}$$

with $n \in \{1, 2\}$, where the value of the Lorentz factor, effective beamsize and space-charge amplitude might differ between the two steps. As before, a Gaussian energy distribution is assumed so that it is $\tilde{\Psi}(\omega) = \sqrt{2\pi}\,\rho_0\,\xi(\sigma_p\omega)$. Firstly, it can be seen that the auxiliary symbol g_1 introduced in Equation (5.7.33) is exactly the normalized first-order gain-function of a single stage, as defined in Equation (5.8.24)

$$g_1 \equiv \frac{1}{\rho_0}\left[1 + i\,kW_1(k)C_1\beta_1\right]\tilde{\psi}(kC_1\beta_1) = G_1(A_1, B_1), \tag{5.8.49}$$

with $A_1 \equiv \frac{eU_1}{\gamma_1\sigma_p}Y(r'_1 k)$ and $B_1 \equiv kC_1\beta_1\sigma_p$. With this, Equation (5.7.56) yields

$$\rho[\phi_{2,1}](q) = \rho_0\, C_2^\dagger\, \cos(C_2^\dagger q)\, G_2^\dagger, \tag{5.8.50}$$

where $G_2^\dagger$ might be referred to as the *first-order, two-step gain function*, which can be written explicitly as

$$G_2^\dagger = \left[1 - k\,\beta_2^\dagger\frac{eU_1\, Y(r'_1 k)}{\gamma_1}\right]\frac{\tilde{\psi}(k\,\beta_2^\dagger)}{\rho_0} - G_1(A_1, B_1)\, C_2^\dagger C_1\beta_2 k\,\frac{eU_2\, Y(r'_2 C_1 k)}{\gamma_2}\,\frac{\tilde{\psi}(C_2^\dagger C_1\beta_2 k)}{\rho_0}. \tag{5.8.51}$$

After defining $A_2 \equiv \frac{eU_2}{\gamma_2\sigma_p}Y(r'_2 k)$, $B_2^\dagger \equiv k\beta_2^\dagger\sigma_p$ and substituting $C_2^\dagger C_1\beta_2 = \beta_2^\dagger - C_1\beta_1$, the gain function in dependence on the normalized parameters can be written as

$$G_2^\dagger(A_1, B_1, A_2, B_2^\dagger) = G_1(A_1, B_2^\dagger) + A_2\, G_1(A_1, B_1)\,\left(B_2^\dagger - B_1\right)\, e^{-\frac{1}{2}\left(B_2^\dagger - B_1\right)^2}. \tag{5.8.52}$$

See Figure 6.12 for a plot of this function.

6 Semi-Lagrangian Vlasov Simulations

As outlined in Section 2.5, simulating large collective many-body systems is not a trivial endeavor. Apart from developing a code in a way that it is actually capable of computing the evolution such a system in a sensible amount of time, another challenge is to have it be tolerant of unavoidable numerical artifacts. Numerical artifacts are differences between the numerical representation of the system and the exact solution, that stem solely from deficiencies of the simulation code. In particular when investigating instabilities, it is important to ensure that the evolution of the instability is not affected significantly by numerical artifacts. Every simulation approach has its own class of potential artifacts. For a given instable system, it is therefore advisable to choose an approach whose artifacts couple to the instability as little as possible.

The microbunching instability is driven by the self-amplification of inhomogeneities in the charge density of the electron bunch. A good simulation approach is therefore one that introduces no or only little charge-density inhomogeneities artificially. Further, it would be advantageous if the method allowed to simulate a system without *any* significant inhomogeneities at all.

Lagrangian simulation approaches represent the many-body system as an ensemble of particles. Due to the particulate nature of this representation, it inevitably introduces inhomogeneities in the phase-space density, and therefore in the charge density. This phenomenon is referred to as *artificial shot-noise*. Coupling of the artificial shot-noise to the instability might be subdued using smoothing techniques. Great care has to be taken when applying smoothing to the charge density, in order not to smooth out the contributions of the actual instability to the charge-density inhomogeneities. This becomes especially troublesome when trying to simulate the emergence of the instability from actual small-scale, small-amplitude inhomogeneities that were added to the system on purpose. Then, the artificial shot-noise can easily dominate the total charge-density inhomogeneities. Smoothing out the artificial contributions then also eradicates the purposefully added "seed" inhomogeneities, which invalidates the approach. The only way to lessen the artificial shot-noise without smoothing is to use a sufficiently large amount of macro particles. In particular when simulating instabilities that occur on short length scales compared to the overall extent of the system, a substantial amount of macro particles might be required. Super-computers of the early 21st century – with many hundreds or thousands of cores – do make it possible to simulate many-body systems using close to as many macroparticles as actual particles present in the physical system [85]. But even then, typical computation times amount to days, if not weeks. Overall, it is not hard to find arguments against using a Lagrangian approach to simulate the microbunching instability.

Eulerian and Semi-Lagrangian methods, as introduced in Sections 2.5.2 and 2.5.3, do not use particles, but represent the phase-space density by storing its values on a grid. As no particles are used, these methods do not suffer from artificial shot-noise and are therefore better suited to simulate the microbunching instability. Of course, also Eulerian and Semi-Lagrangian methods are not free of artifacts. In Semi-Lagrangian

codes, the phase-space density has to be interpolated to obtain its values between grid points, which introduces interpolation errors. Eulerian codes typically need to determine gradients of the phase-space density, for instance by finite-difference schemes, which also introduces errors. Compared to artificial shot-noise, the magnitude of these artifacts is typically much smaller. Further, these artifacts are typically not pseudo-randomly distributed and have well-defined error bounds.

Grid-based methods, however, have to overcome a challenge that Lagrangian methods do not: the amount of grid points required – and therefore the amount of computer memory required – to represent a phase-space density with high enough resolution, grows exponentially with the dimension of the phase-space. This is the *curse of dimensions*. In particle-based methods, there is no such dependence on the dimension of the phase-space. There, only the number of coordinates stored for each particle increases with the dimension. Nevertheless also in particle-based methods, the number of particles has to be increased exponentially to achieve constant resolution in phase-spaces of increasing dimension. One, however, always has the option to choose fewer particles and accept the reduced resolution. Grid-based methods are therefore especially suited for low-dimensional systems. Simulations of full six-dimensional phase-spaces using grid-based methods are not common.

Another challenge that grid-based methods face is the fact that some phase-space densities do not lend themselves readily to be represented on a simple homogeneous grid, due to their shape. In particular, if the volume of the support – the region in which the phase-space density attains a non-zero value – is much smaller than the volume of the minimum bounding rectangle of the support, a homogeneous grid will mainly cover empty areas of phase-space and only a small amount of grid points will actually be in the support. As a result, computer memory will be wasted on storing a large amount of zeros, representing the empty regions of phase space, which makes this approach numerically unfeasible. We call such phase-space densities *exotic*. A metric for the exoticness is the ratio between the volume of the support and the volume of its minimum bounding rectangle. In some cases, it is possible to reduce the exoticness of a phase-space density by a suitable coordinate transformation [51]. This strategy is particularly advantageous, if the exoticness stems mainly from linear correlations between the dimensions. Then, the system can be treated in the decorrelated coordinates, in which the less exotic phase-space density can be represented efficiently on a homogeneous grid, ensuring of course proper transformation of the equations of motion. If, however, the exoticness is caused by non-linearities in the shape of the support, this procedure is more cumbersome, as determining the appropriate transformation is generally not straight-forward and applying it most likely results in overly intricate equations of motion. In particular, the longitudinal phase-space density of electron bunches in FELs usually has such a non-linear shape, as it picks up non-linearities from collective effects, rf-acceleration, and magnetic chicanes. For this type of non-linear, exotic phase-space densities, it is advisable not to use a homogeneous grid, but rather to employ some form of domain decomposition in order to make the numerical representation more efficient.

In reference [71] the author of this thesis presents a proof-of-principle, showing that tree-based domain decomposition is suitable to facilitate efficient semi-Lagrangian Vlasov simulation of the longitudinal phase-space densities in FELs. During the course of this PhD project, the proof-of-principle code was largely extended and made serviceable to properly simulate actual physical systems. The result is SelaV$_{1D}$ – a one-dimensional **se**mi-**La**grangian **V**lasov simulation code [53].

In the following, the concept of tree-based domain decomposition is briefly outlined, which is presented in reference [71] in much more detail. Subsequently, the key im-

provements and some of the new features that were integrated in SelaV$_{1D}$ are presented. Finally, numerical results for the microbunching instability obtained from SelaV$_{1D}$ are presented and compared with the perturbation theory derived before.

6.1 Tree-Based Domain Decomposition

In tree-based domain decomposition, the simulation window – the part of phase-space which contains the phase-space density – is subdivided into a hierarchy of nested hyper-rectangles. Each rectangle in the hierarchy can be the *parent* of smaller *child* rectangles contained within it. If a rectangle is not a parent, it is called a *leaf*. Only the outermost rectangle, which covers the whole simulation window, does not have a parent, which makes it the *root* of the tree. In SelaV$_{1D}$, the geometry of the child rectangles is chosen so that they contain the union of the half-phase-spaces of their parent rectangle along each dimension, see Figure 6.1. The graph of the parent-child relationships of these rectangles forms what is known as a *tree* [86], which gives the method its name.

Such a tree can be used to create a data structure that minimizes the number of grid points outside of the support of the phase-space density. To that end, the tree is generated in a way so that only those rectangles that intersect the support $\mathrm{supp}(\Psi)$ of the phase-space density Ψ are parents. Mathematically, the support of many typical phase-space densities actually covers the entire phase-space $\mathrm{supp}(\Psi) = \mathbb{R}^{2n}$, as Ψ might become arbitrarily small but is nowhere exactly zero. Numerically, this issue can be circumvented by using the truncated ϵ-support

$$\mathrm{supp}_\epsilon(\Psi) \equiv \left\{ \boldsymbol{z} \in \mathbb{R}^{2n} \mid \Psi(z) \geqslant \epsilon \right\}, \tag{6.1.1}$$

where $\epsilon \in \mathbb{R}$ is a suitably chosen threshold value. Each leaf of the resulting tree is assigned a homogeneous grid which covers it. Only on these leaf-grids, the values of the phase-space density are stored in memory. No values are stored for phase-space regions not covered by leafs. As by definition each leaf intersects the support $\mathrm{supp}_\epsilon(\Psi)$ of the phase-space density at least partially, this procedure minimizes the amount of wasted computer memory.

This data structure and the associated routines – in particular the interpolation routines – provide the key functionality for efficient semi-Lagrangian simulation of exotic phase-space densities. We therefore deemed it prudent to implement them in a separate library `libselav`, as they might be useful for future projects, apart from SelaV$_{1D}$. It is noteworthy that the hyper-tree structures offered by `libselav` can be of any dimension. As SelaV$_{1D}$ simulates systems with one-degree of freedom it uses `libselav` for its two-dimensional trees. In the following, most of the enhancements are explained for a two-dimensional phase-space, as this is what SelaV$_{1D}$ uses. Nevertheless, all of the presented concepts and routines are applicable in phase-spaces of arbitrary dimension and are, in fact, implemented in `libselav` in a dimension-agnostic way.

6.2 Enhancements of the Semi-Lagrangian Vlasov Code

SelaV$_{1D}$ has its origins in the unnamed proof-of-concept code described in [71]. To make SelaV$_{1D}$ suitable for actual simulation work, large parts of the existing code were reworked and many new features were implemented in both SelaV$_{1D}$ and `libselav`. In the following, some of the most important enhancements will be highlighted.

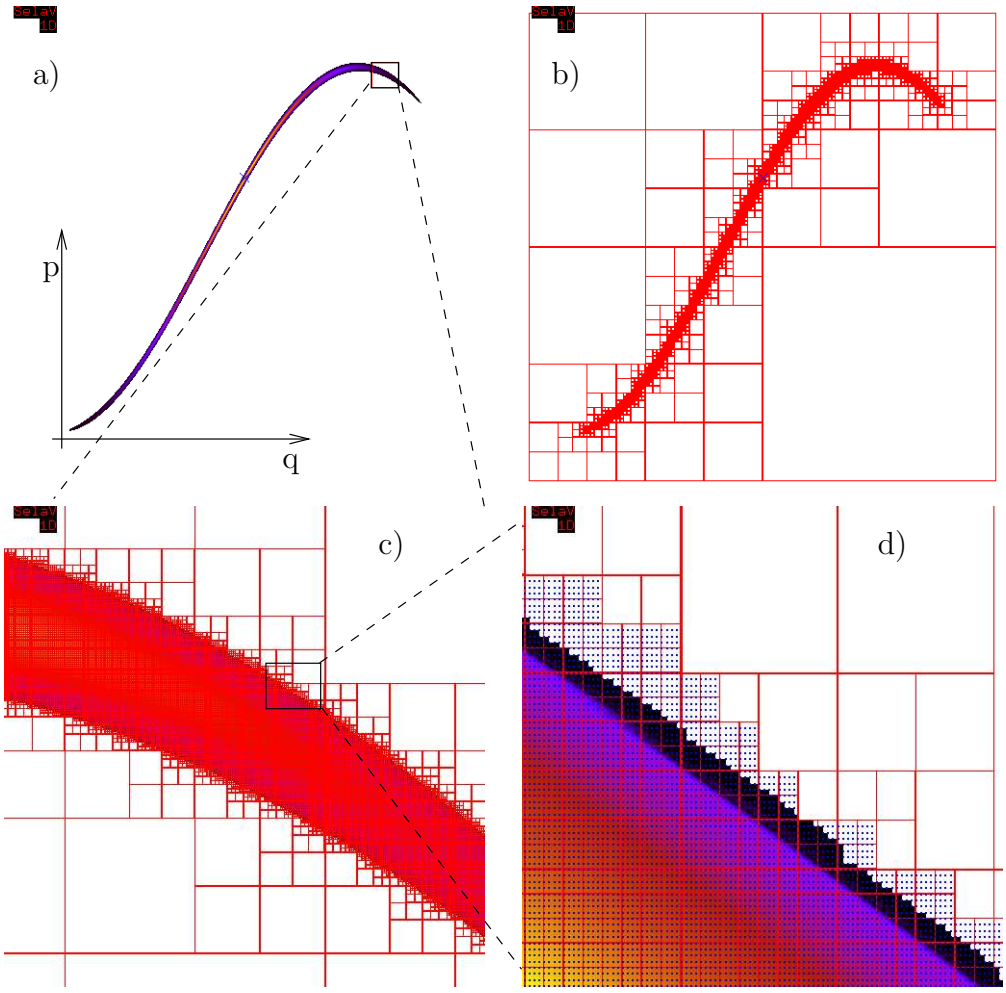

Figure 6.1: Tree-based domain decomposition of an exotic phase-space density. a) the phase-space density; b) hierarchy of rectangles covering the support of the phase-space density; c) part of tree at an intermediate magnification; d) detail showing leafs of the tree and the location of the grid points covering them (blue dots).

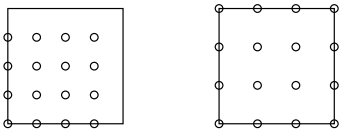

Figure 6.2: Two possible layouts of the grid covering a leaf rectangle. Circles mark the position of the grid points. The new layout used by SelaV$_{1D}$ is on the left. On the right is the old layout used in the proof-of-principle code. This examples depicts the case of four grid points per dimension. In the actual implementation, the number of grid points per dimension can be any power of two.

6.2.1 Grid Layout and Inter-Leaf Interpolation

A major shortcoming of the proof-of-concept implementation was an unfavorable layout of the grid that covers the leaf rectangles. In Figure 6.2, the old and the new layout are depicted. The advantage of the old layout is that it allows to linearly interpolate the phase-space density anywhere within a leaf, using only the grid points contained in that leaf. This simplified the implementation in the proof-of-concept code significantly. However, using the old layout phase-space points on the edges and the corners of a leaf rectangles were sampled multiple times by neighboring leafs, as illustrated in Figure 6.3. This redundant sampling of grid points is clearly not optimal for memory efficiency.

In the new layout, the grid points of neighboring do not overlap, so that all grid points are sampled only once. While the illustrations shown here depict only the two-dimensional case, this approach also works in higher dimensions, which is also implemented in `libselav`. As in the new layout the grid on a single leaf misses points on half of its edges, it is no longer possible to interpolate every point within a single leaf using only its own grid points. Interpolation in the region between the outer-most grid points and the edges of the leaf requires grid points from the neighbors of the leaf. This complicates the interpolation routine, as the appropriate grid points have to be gathered from the neighbors and be put into a temporary grid, based on which the actual interpolation is then calculated. A routine is implemented in `libselav` that constructs such a temporary, *local* grid of selectable size around any given phase-space point.

The tree-structures implemented in `libselav` automatically keep track of the neighbors of any leaf *while the tree is constructed*. Determining a leaf neighbor is therefore a constant time look-up operation, which is crucial for the performance of this approach.

Being able to use the grid points of neighbor leafs enables another improvement implemented in SelaV$_{1D}$: higher-order interpolation schemes. For linear interpolation, only the four grid points closest to the interpolation point are required – the so-called *nearest neighbors*. Higher-order interpolation schemes, however, need more than the nearest neighbors. In the proof-of-principle code, only linear interpolation was possible, as each leaf did contain the required nearest-neighbor grid points of of all points in the phase-space region it covers, but higher-order interpolation was impossible as that requires points outside outside of the leaf – which were inaccessible in that implementation. Thanks to the new ability to construct local grids of any size, it was possible to implement higher-order interpolation schemes in SelaV$_{1D}$.

6.2.2 Higher-Order Interpolation

In semi-Lagrangian codes, the new value of a phase-space density at a given point after a time step is determined by tracking the phase-space coordinates of that point backwards in time and evaluating the old phase-space density at the resulting point. As the phase-

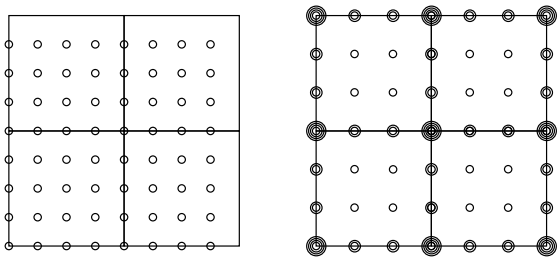

Figure 6.3: Grid points covered by neighboring leafs. Using the old layout (right), phase-space points on the edges of the leaf are sampled two times, those at the corners four times, as indicated by multiple circles. With the new layout (left), every point in phase-space is sampled only once.

space density is represented numerically on a grid, its value is known exactly only on the grid points. However, the back-tracked coordinates generally do not coincide with a grid point. To determine a value for the phase-space density at such off-grid points, interpolation, based on the surrounding grid points, is necessary. Interpolation therefore plays a central role in semi-Lagrangian simulation codes and the interpolation scheme determines important performance parameters of the code.

6.2.2.1 Spline Interpolation

One well-known interpolation method, that is often used in various applications, is the so-called *spline interpolation*, and in particular *cubic* spline interpolation [87, 88]. With cubic spline interpolation, a one-dimensional function is interpolated between the grid points piecewise by third-order polynomials. The coefficients of these polynomials are chosen in a way that the first and second derivatives of the interpolating function are continuous at the grid points. Coefficients that fulfill these requirements can only be calculated by taking into account *all* grid points and not only those closest to the interpolation point. Spline interpolation is therefore a *global* interpolation method, as all grid points influence the value of the interpolation function everywhere.

Extending one-dimensional spline interpolation to arbitrary dimensions is not trivial, as for each dimension beyond the first, spline coefficients have to be recalculated based on interpolated values between all grid points along the previous dimension. Spline interpolation has been used successfully in codes which use a model of the phase-space dynamics, in which in each time-step map the phase-space coordinates change only along a single phase-space dimension [51, 89]. In two dimensions, this correspond to kick- and drift maps. In that case, multi-dimensional interpolation is not needed, as only interpolation along that single dimension is necessary.

In SelaV_{1D}, spline interpolation is *not* used, as we decided to allow for arbitrary maps, not only kick and drift maps.

6.2.2.2 Local Polynomial Interpolation

Another approach, better suited for SelaV_{1D}, is the *local* polynomial interpolation, which we already hinted at, when we discussed the grid layout used in SelaV_{1D}. In local interpolation schemes, only a small number of grid points around the point to be interpolated is used to construct an interpolating polynomial. As a result, each grid point contributes only within a finite region around it to the overall piecewise interpolating polynomial.

A one-dimensional local polynomial interpolation scheme can be easily extended to multiple dimensions. This is achieved by decomposing the multi-dimensional interpola-

tion into multiple one-dimensional interpolations along each dimension [87,90]. For each (hyper-)row along the ith dimension of the local grid, the value of the one-dimensional interpolating function is calculated at the ith coordinate of the interpolation point. These interpolated points in turn form a grid whose number of dimensions is reduced by 1 compared to the previous grid. The process is repeated for the $(i+1)$st dimension of the new grid. At its end, this recursive procedure yields the final result for the interpolated value. For example, given a homogeneous 4×4 grid with known grid constants and with the values $\mathbb{R}^{4\times4} \ni y^{\#} = \{y_{0,0}, \ldots, y_{3,3}\}$, a four-point one-dimensional interpolation scheme INTERP1D : $\mathbb{R} \times \mathbb{R}^4 \to \mathbb{R}$ is extended to the two-dimensional interpolation scheme INTERP2D : $\mathbb{R}^2 \times \mathbb{R}^{4\times4} \to \mathbb{R}$ via

$$\begin{aligned}\text{INTERP2D}\left(x_0, x_1; y^{\#}\right) =& \text{INTERP1D}(x_0; \\ & \text{INTERP1D}(x_1; y_{0,0}, y_{0,1}, y_{0,2}, y_{0,3}), \\ & \text{INTERP1D}(x_1; y_{1,0}, y_{1,1}, y_{1,2}, y_{1,3}), \\ & \text{INTERP1D}(x_1; y_{2,0}, y_{2,1}, y_{2,2}, y_{2,3}), \\ & \text{INTERP1D}(x_1; y_{3,0}, y_{3,1}, y_{3,2}, y_{3,3})).\end{aligned} \tag{6.2.1}$$

This way, the multi-dimensional interpolation function inherits many of the properties of the underlying one-dimensional interpolation function. In particular, it is typically desirable for the interpolating function to be continuously differentiable to some degree. From the construction of the tensor interpolation, it becomes apparent that, if the one-dimension interpolation function is continuously differentiable, then so is the multi-dimensional interpolation function.

To achieve a continuously differentiable interpolant, the one-dimensional interpolating polynomial has to be constructed in a way so that the derivatives of the two polynomials of two neighboring grid cells – the region enclosed by two grid points – attain the same value at the location of their common grid point. Lagrange polynomials are ill-suited for this task as they are fully defined by the values grid points alone and therefore do not provide a handle on the derivatives of the resulting interpolation function, see Figure 6.4a. Better suited are interpolation schemes that take the derivatives the grid points into account explicitly. One example is Hermite interpolation [88], which allows to specify the value of the interpolant as well as its derivatives up to a given order at all grid points. It is well-known how to construct Hermite interpolation polynomials of any order for any number of grid points. If only a small number of grid points and derivatives is considered, the interpolating polynomials can also be constructed directly. With such an interpolation scheme, the derivatives of the interpolating polynomials of two neighboring grid cells at their common grid point can be chosen freely. By choosing the same values for the derivatives of the two polynomials, continuous differentiability is achieved.

6.2.2.3 Smooth Local Polynomial Interpolation

In SelaV_{1D}, only the values of the phase-space density is known at the grid points, whereas the values of its derivatives are not known explicitly. Hence, the question arises, which values to choose for the derivatives. Without any additional knowledge about the analytical derivatives, the only way is to approximate them, based on the values of the phase-space density, using *finite differences* [91]. Following this approach, it is imperative to choose a finite-difference scheme that yields the same value for *both* grid cells at their common grid point, to achieve continuous differentiability.

Consider an evenly spaced grid with grid constant h, so that the locations of the grid points fulfill $x_i - x_j = h(i-j)$. The goal is to construct a polynomial of degree

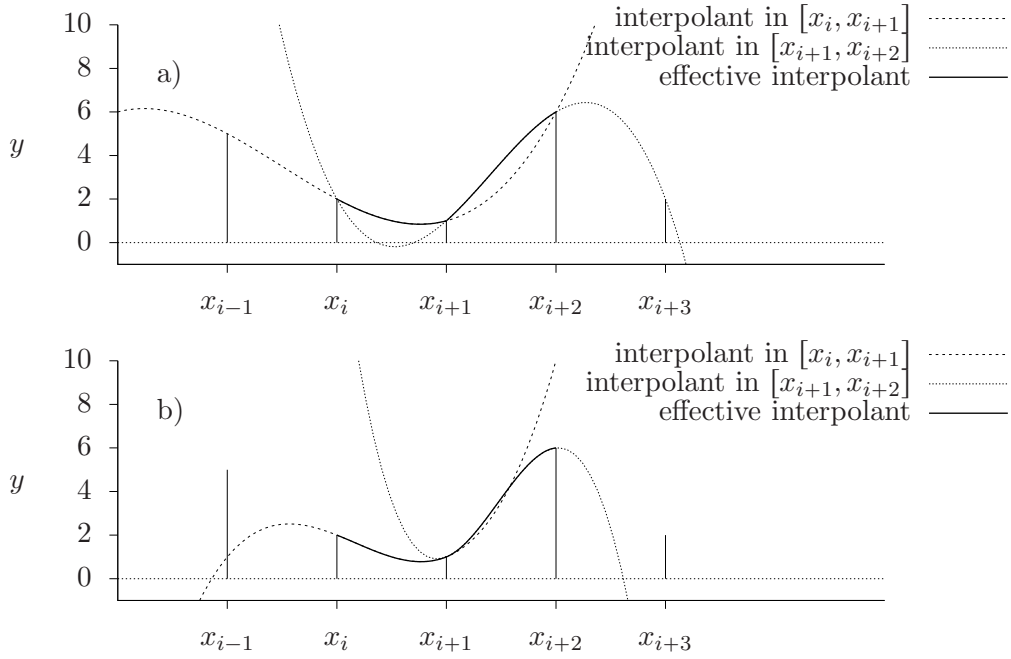

Figure 6.4: Comparison of a) third-order Lagrangian interpolation and b) cubic C^1 interpolation. Note the kink in the effective interpolant at x_{i+1} in a), indicating a discontinuity in the derivative. Using cubic C^1 interpolation, the derivative is continuous at the grid cell boundaries. In b), the values of the interpolant of a cell does not match the grid value at the outermost of its four grid points. This is acceptable, because the interpolant is only ever evaluated between the two innermost grid points while the outer grid points are only used to determine the derivatives at the cell boundaries.

3 that is C^1 at the grid points using finite differences. To this end, we take the four grid points at $x_{i-1}, \ldots, x_{i+2}$, with $x_{i-1} < x_i < x < x_{i+1} < x_{i+2}$, where x is the point to be interpolated, together with their values $y_{i-1}, \ldots, y_{i+2}$. The interval $[x_i, x_{i+1}]$ is the grid cell in which the interpolation polynomial is evaluated. In order for the derivative of the interpolation function $f'(x)$ to be continuous across grid cells, we have to choose a finite difference scheme to determine values for the derivatives at the lower boundary $y_i'^- \equiv f'(x_i)$ and the upper boundary $y_i'^+ \equiv f'(x_{i+1})$, so that

$$y_i'^+ = y_{i+1}'^-. \tag{6.2.2}$$

This condition can be fulfilled by using the central differences at the cell boundaries

$$y_i'^- \equiv \frac{y_{i+1} - y_{i-1}}{2h} \quad \text{and} \quad y_i'^+ \equiv \frac{y_{i+2} - y_i}{2h}, \tag{6.2.3}$$

which approximate the derivatives accurately to second order. Using this scheme, two neighboring grid cells agree on the same value for the derivative at their common grid point, while using only their own respective sets of four grid points.

This approach can be extend to increase the smoothness of the effective interpolant further by fixing also its higher derivatives to values determined by finite difference schemes. To achieve C^2 interpolation, the values for the second derivatives at the cell boundaries $y_i''^- \equiv f''(x_i)$ and $y_i''^+ \equiv f''(x_{i+1})$ can be determined using the second-degree finite differences

$$y_i''^- \equiv \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} \quad \text{and} \quad y_i''^+ \equiv \frac{y_i - 2y_{i+1} + y_{i+2}}{h^2}, \tag{6.2.4}$$

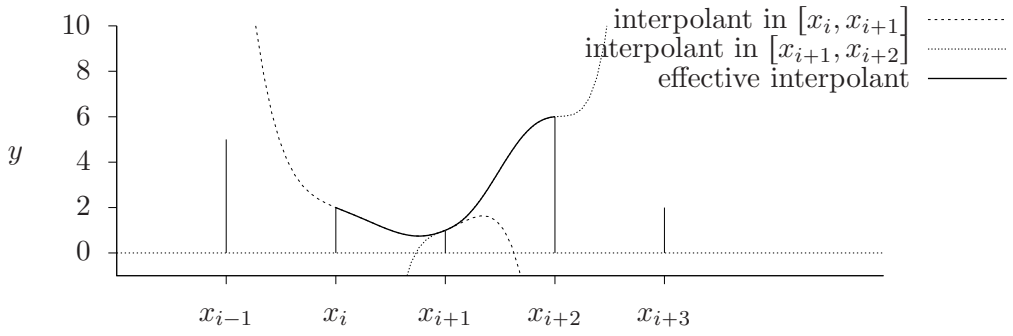

Figure 6.5: Quintic C^2 interpolation based on four grid points.

which are accurate to second order. Remarkably, both the first and second central finite differences can be calculated using four grid points and yield second-order approximations for the respective derivative. Hence, it is possible to construct a C^2 interpolant using local grids of size four. An example of the resulting interpolation function is shown in Figure 6.5.

Implemented in SelaV$_{1D}$ is bilinear and bicubic Lagrange interpolation as well as bicubic C^1 and biquintic C^2 interpolation, as described above, plus nearest neighbor interpolation for testing purposes. See Figure 6.6 for a visual comparison of the methods.

6.2.3 Input Language Parser

In order to allow for flexible simulation setups, SelaV$_{1D}$ is controlled by a simple input language. Notably, it allows phase-space densities and Perron–Frobenius operators to be assigned to variables, so that they can be handled as objects. In addition, many of the usual floating-point arithmetic operations, string manipulation and flow-control features commonly found in contemporary scripting languages are implemented. The following example shows one the most elementary operations, namely initializing a phase-space density object; generating a collective Perron–Frobenius operator object; executing Perron–Frobenius steps, which yield more phase-space density objects; and finally writing output generated from the resulting densities.

```
psi0 = psd_gauss(                // define a bi-gaussian density
          sig_q=2, sig_p=0.5,    // set RMS values
          correlation=0.1,       // correlation parameter
          limits=[-10,-10,10,10] // size  of the root  rectangle
);

drift = map_driftl(1);           // PF op. of a linear drift-map
kick = map_poisson1d(psi0);      // collective PF operator,
                                 // calculated from psi0

psi1 = propagate(psi0, kick);    // apply coll. PF operator to psi0
psi2 = propagate(psi1, drift);   // apply drift PF operator to psi1

write_localmoments(psi0, "localmoments-psi0.dat", 1); // write output
write_localmoments(psi1, "localmoments-psi2.dat", 1);
```

A description of the features of the input language, as well as a complete list of the

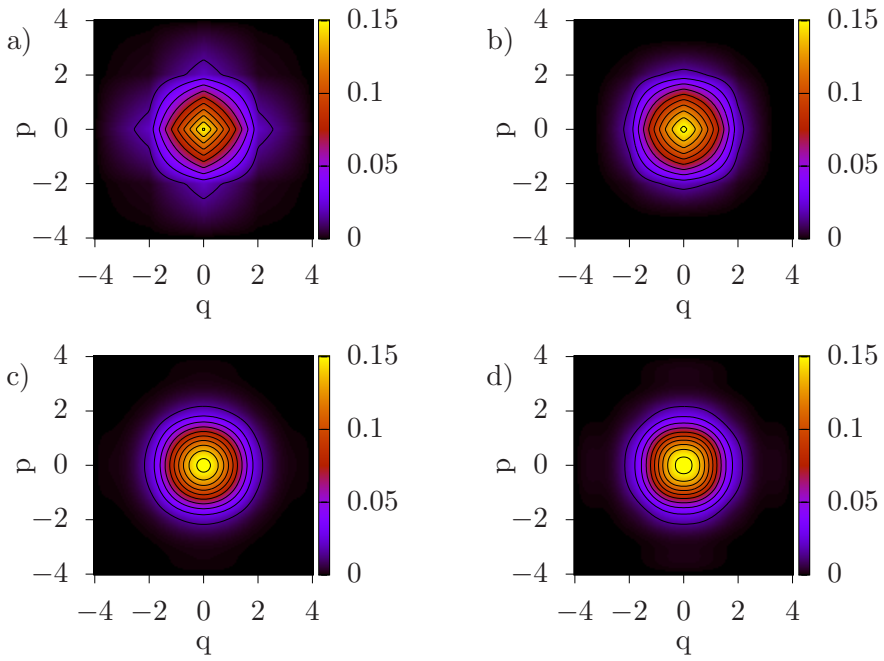

Figure 6.6: Visual comparison of interpolation methods implemented in SelaV$_{1D}$: a) bilinear Lagrangian, b) bicubic Lagrangian, c) bicubic C^1 , and d) biquintic C^2. The test function is a Gaussian distribution with zero mean and $\sigma_q = \sigma_p = 1$, sampled on a 4×4 grid.

available phase-space density initialization routines and Perron–Frobenius operators can be found in the documentation of SelaV$_{1D}$ [53].

6.2.4 Import and Export of Particle Ensembles

Most beam-dynamics simulation codes follow the Lagrangian approach and simulate the evolution of a bunch by tracking an ensemble of macroparticles. To enable some degree of interoperability between SelaV$_{1D}$ and these particle-based codes, routines have been implemented in SelaV$_{1D}$ that generate particle ensembles from the numerical representation of the smooth phase-space density and vice versa. However, as SelaV$_{1D}$ is a one-dimensional code, when importing an ensemble only the longitudinal coordinates can be taken into account even if the input particle ensemble is six-dimensional. Also, when exporting a particle ensemble, only the longitudinal conjugate coordinates can be generated.

6.2.4.1 Export

To generate a particle ensemble that is distributed with the probability density represented on the tree, the following algorithm is used: First, the integrated probability density of each leaf is calculated and stored in that leaf in an auxiliary variable. This represents the probability to find any particle inside a given leaf. These probabilities are then recursively propagated upwards in the tree hierarchy, by exploiting the fact that the integrated probability-density contained in a parent rectangle is equal to the sum of the probabilities of its child rectangles. At the end of this process, every rectangle in the tree has its integrated probability-density stored in an auxiliary variable. This results in what is known as a *weighted* graph [86].

After this setup step, the actual particle ensemble is generated. A single particle is generated by conducting a weighted descent of the rectangle tree: Starting with the root of the tree as the *current* rectangle, the *next* rectangle is randomly selected among the children of the current rectangle. This random selection is done in a way so that the probability of the child rectangle with index i to be selected as the next rectangle is given by its relative integrated probability $\tilde{P}_i \equiv \frac{P_i}{\sum_{n=1}^{N} P_n}$, where $P_1, \ldots, P_N$ are the integrated probability densities of the child rectangles of the current rectangle. Note that in two-dimensional trees it is $N = 4$. This process is continued recursively with the next rectangle as the current rectangle, until the current rectangle is a leaf. Then, based on the probability density stored on the grid of that leaf, the coordinates of the particle to be generated are determined via rejection sampling [87], where the domain of the rejection sampling is limited to the phase-space area covered by the leaf. This process is repeated until the requested number of particles is generated. It needs to be stressed that this weighted descent method is crucial for the efficiency of the sampling routine: Applying rejection sampling directly to the root rectangle, would be highly inefficient for exotic phase-space densities, because the vast majority of samples would be rejected immediately, as they fall out of the support. This would cause an exceedingly large number of samples to be generated in the process.

6.2.4.2 Import

To import a particle ensemble – which was for instance generated as the output of a particle-tracking code – into SelaV$_{\mathrm{1D}}$, it is necessary to generate a smooth approximation of the phase-space density from it .

One method to achieve this is *kernel density estimation* (KDE) [92, 93]. Using KDE, an *estimator* $\hat{\Psi}$ for the phase-space density is given by

$$\hat{\Psi}(\boldsymbol{z}) = \frac{1}{N} \sum_{i=1}^{N} \hat{K}(\boldsymbol{z} - \boldsymbol{z}_i), \tag{6.2.5}$$

where $\hat{K} \colon \mathbb{R}^2 \to \mathbb{R}$ is a suitable kernel function, and $\boldsymbol{z}_i$ are the phase-space coordinates of the N particles. Conceptually, this is equivalent to treating the particle ensemble as a Klimontovich density which is convolved with the kernel function

$$\hat{\Psi}(\boldsymbol{z}) = \hat{K}(\boldsymbol{z}) * \sum_{i=1}^{N} \delta(\boldsymbol{z} - \boldsymbol{z}_i). \tag{6.2.6}$$

Usually, the kernel function is chosen as $\hat{K}(\boldsymbol{z}) = K(\|\boldsymbol{z}\|/\alpha)$, where $\|\cdot\|$ is an appropriate norm, $K \colon \mathbb{R} \to \mathbb{R}$ is a one-dimensional kernel with compact support, and α is a *smoothing parameter*. This approach was evaluated during the early development of SelaV$_{\mathrm{1D}}$, but was dismissed after some testing. It became apparent that phase-space densities generated with this method still show a considerable level of artificial density modulations, stemming from the shot-noise of the input particle ensemble. Increasing the smoothing parameter would partially suppress these artifacts. Suppressing these artifacts to a level where they would not seed the microbunching instability, however, required value a for the smoothing parameter that was large enough to significantly affect the macroscopic shape of the charge density as well as the energy spread. Due to these issues with KDE, a different approach was implemented in SelaV$_{\mathrm{1D}}$.

In certain cases, the phase-space density can be approximated by a function of the form

$$\Psi(q, p) = \rho(q)\, \psi(q; p), \tag{6.2.7}$$

where $\rho(\cdot)$ is the charge density and $\psi(q;\cdot)$ is a one-parametric family of energy distributions, where q is the family parameter. Such a distribution family can be constructed, for instance, from a Gaussian distribution whose mean μ_p and standard deviation σ_p depend on q:

$$\psi(q;p) = \xi\left(\mu_p(q), \sigma_p(q); p\right) \equiv \frac{1}{\sqrt{2\pi}\sigma_p(q)} \exp\left(-\frac{1}{2}\left[\frac{p-\mu_p(q)}{\sigma_p(q)}\right]^2\right). \tag{6.2.8}$$

The functions $\mu_p(q)$ and $\sigma_p(q)$ take the notion of the *local* mean energy and the *local* energy spread, respectively. Pertaining to beam-dynamics simulations, it became apparent that the phase-space density of a bunch immediately after the gun can be approximated sufficiently well by a function of this form. Of course, not all phase-space densities can be approximated in this manner. Especially, if the support of the phase-space density undercuts itself, it is typically not possible to find an appropriate distribution family for this approach.

Assuming the actual phase-space density can be approximated by Equations (6.2.7) and (6.2.8), it is possible to construct a smooth numerical representation from the local statistical moments of a given particle ensemble. To that end, first the particles are binned with respect to their longitudinal coordinates $Q = \{q_1, \ldots, q_N\}$, into equidistant bins $b_i = [\hat{q}_i - \frac{h}{2}, \hat{q}_i + \frac{h}{2})$ with center points $\hat{q}_i$ and bin width h. Estimates for the local mean and standard deviation of the energy at $\hat{q}_i$ of the ensemble can then be determined from the first and second statistical moments of the sample in the ith bin

$$\mu_p(\hat{q}_i) \approx \hat{\mu}_{p,i} \equiv \frac{1}{|Q \cap b_i|} \sum_{n|q_n \in b_i} p_n \tag{6.2.9}$$

and

$$\sigma_p(\hat{q}_i)^2 \approx \hat{\sigma}^2_{p,i} \equiv \frac{1}{|Q \cap b_i|} \sum_{n|q_n \in b_i} \left(p_n - \hat{\mu}_{p,i}\right)^2, \tag{6.2.10}$$

respectively. Based on these estimates, Equation (6.2.8) can be approximated by choosing interpolating functions for the local mean energy $\mu_p(q) \to \hat{\mu}_p(q)$ and standard deviation $\sigma_p(q) \to \hat{\sigma}_p(q)$, which interpolate between the respective bin values $\hat{\mu}_p(\hat{q}_i) = \hat{\mu}_{p,i}$ and $\hat{\sigma}_p(\hat{q}_i) = \hat{\sigma}_{p,i}$. To ensure the smoothness of the resulting phase-space density, the one-dimensional C^1 or C^2 interpolation schemes described in Section 6.2.2.3 can be employed. The result is a smooth phase-space density that – by construction – does not support inhomogeneities on scales much smaller than h. See Figure 6.7 for an example of a particle ensemble and the phase-space density generated by this method of local moments.

Using this method, it was possible to generate phase-space densities from particle ensembles with sufficiently small residual shot-noise artifacts which do not seed the microbunching instability. This method can be extended to include higher-order statistical moments, if an analytical expression for a distribution family is available that is appropriately parameterized in terms of its statistical moments up to the requested order.

6.2.5 Local Poisson Shot Noise

Potentially, the microbunching instability can be seeded by the shot noise present in the bunch. In macroparticle tracking codes, shot-noise effects can be accounted for straightforwardly by choosing a sufficiently large number of simulated macroparticles, which is close to the actual number of electrons in the bunch. In grid-based simulation methods, such as the semi-Lagrangian method, there is no inherent shot noise. To study

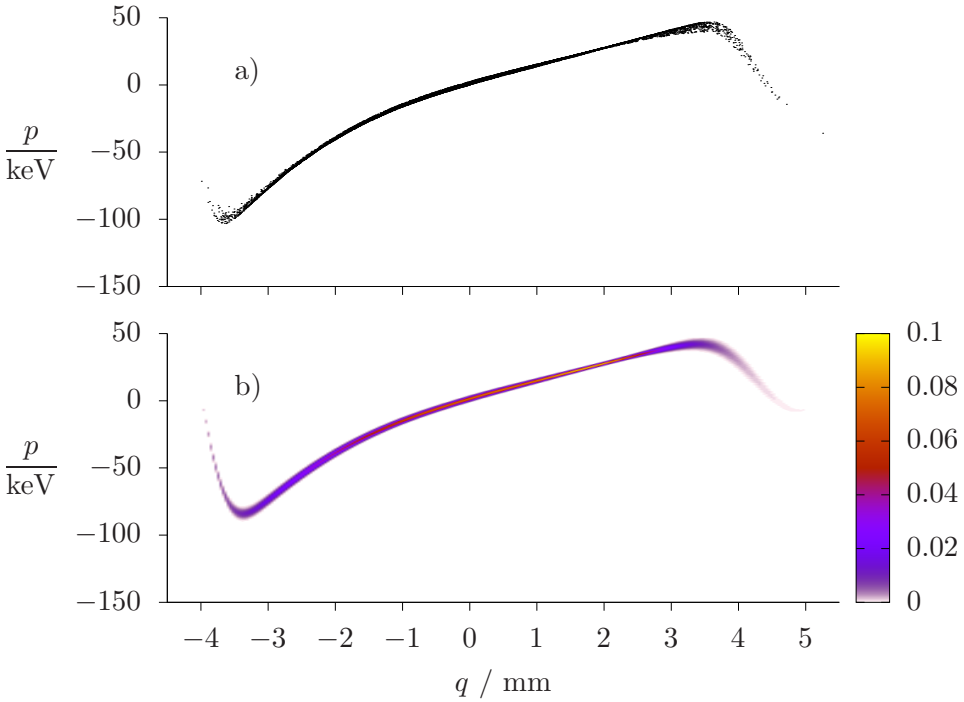

Figure 6.7: a) Subset ($N = 10^4$) of the output of a particle-tracking simulation of the FLASH gun [94]. b) smooth phase-space density generated from the ensemble a) using the method of local moments.

shot-noise effects using a semi-Lagrangian code, the inhomogeneities induced by shot noise have to be added explicitly to the initial condition of the phase-space density.

In SelaV$_{\mathrm{1D}}$, a routine was implemented to add shot-noise inhomogeneities to the grid-based representation of a phase-space density. This is achieved by modifying the grid values of an initially homogeneous phase-space density Ψ_0. For each grid point $\boldsymbol{z}_{ij}$, the local integrated phase-space density is approximated by $u_{ij} \equiv \Psi_0(\boldsymbol{z}_{ij})\,\mathrm{d}^2\boldsymbol{z}_{ij}$, where $\mathrm{d}^2\boldsymbol{z}_{ij}$ is the phase-space area covered by the grid cell. With this, the expected number of particles in a grid cell is then given by

$$\lambda(\boldsymbol{z}_{ij}) \equiv \lambda_{ij} = \frac{|Q_b|\,u_{ij}}{e}, \tag{6.2.11}$$

where Q_b is the total charge of the bunch. Each grid value $\Psi_0(\boldsymbol{z}_{ij})$ is replaced by a realization of a random variable X_{ij}, normalized to the cell volume

$$\Psi_0(\boldsymbol{z}_{ij}) \leftarrow \frac{X_{ij}}{\mathrm{d}^2\boldsymbol{z}_{ij}}. \tag{6.2.12}$$

Each of the random variables X_{ij} follows a Poisson distribution, so that the probability of it attaining the value $k \in \mathbb{N}^{0,+}$ is given by

$$P(X_{ij} = k) = \frac{\lambda_{ij}^k\,\mathrm{e}^{-\lambda_{ij}}}{k!}, \tag{6.2.13}$$

where λ_{ij} is the local expected number of particles as defined above. After this modification, the local statistics of the modified phase-space density matches the particle-number statistics of a locally Poisson-distributed particle ensemble. See Figure 6.8 for plots of two examples of phase-space densities that were modified in this way. In [95, 96], a conceptually similar method is proposed, where the assumption is made that the expected number of particles in a grid cell is large so that the Poisson distribution can be approximated by a Gaussian distribution.

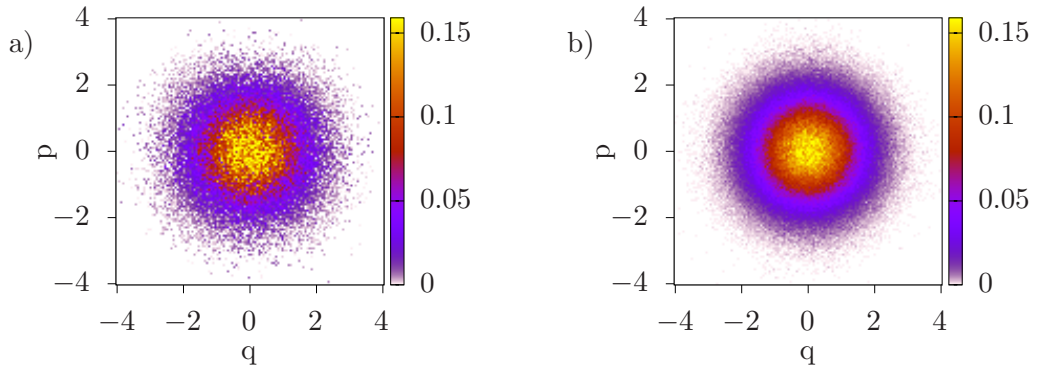

Figure 6.8: Initially Gaussian phase-space densities ($\sigma_q = \sigma_p = 1$) sampled on a 128×128 grid after adding local Poisson shot noise, for bunches with is a) $|Q_b|/e = 10^4$ and b) $|Q_b|/e = 10^6$.

6.3 Discrete-Time Long-Bunch Model

For a first benchmark, SelaV$_{1D}$ is applied to the ultra-relativistic discrete-time bunch compression model and the results are compared with the perturbation theory derived in Section 5. In this model each bunch-compression stage is treated as a single microbunching step. The long linac section is treated as a single collective kick, caused by longitudinal space charge forces, plus the kick of the accelerating cavities. After the linac section follows a magnetic chicane, which acts via a drift map on the longitudinal phase-space.

Locally, the strength of the longitudinal space-charge kicks is determined by an impedance function, which depends on the beam energy and the transverse beam size. These parameters vary along the linac section due to the acceleration of the bunch and the transverse dynamics governed by the optics in the section. In a more detailed model of the dynamics, the variation of these parameters would be accounted for by integrating the local impedance function along the beamline, yielding an *integrated* impedance function. As this approach requires detailed knowledge of the energy profile and optics along the linac, the number of free parameters is large. To keep the number of free parameters small for this benchmark, the impact of the space-charge kicks is instead approximated by using an *characteristic* impedance function. In this approximation, the impedance function is considered to be constant along the linac section, and a single *characteristic* value is used for each of its free parameters, namely the beam energy and transverse beam size. Typically, the mean value of the respective parameters along the linac is a suitable choice for this characteristic value. The only remaining free parameter is the length of the linac section.

The bunch compression chicane is modeled as a single linear drift map. Collective kicks that occur in the chicane due to space-charge and coherent synchrotron radiation are neglected in this model. Working in the dechirped frame introduced in Section 3.4.4, the drift map of the chicane can then be parameterized with respect to three parameters: the compression factor, the chicane strength, and the beam energy at the chicane.

For the phase-space density, we will use the *long bunch* approximation. In this approximation, it is assumed that the bunch is longitudinally much longer than the microbunching structures that are investigated. Edge effects, in particular the collective kicks caused by the rising and falling flanks of the charge density at the front and the back of the bunch, do not affect the microbunching process in the core of the bunch, in

that case. Therefore, the microbunching instability can be studied by simulating only a portion of the full phase-space density taken from the core part of the bunch. The initial phase-space density in the core of the bunch is considered to have a Gaussian energy distribution with standard deviation σ_p, and initially no dependence on the longitudinal coordinate q. The scaling of the charge density is fixed by specifying the initial bunch current I_0. As we work in the dechirped frame, the energy chirp of the bunch is accounted for by the choice of the coordinates and does not have to be included explicitly in the phase-space density. This allows also chirped bunches to be represented by a q-independent phase-space density.

In total, this model has six free parameters per bunch compression stage: the characteristic energy $\bar{E}$, the characteristic beam size $\bar{r}$, the length of the linac section L, the chicane energy E_{BC}, the compression factor C, and the chicane strength $\underline{M}_{56}$; plus two free parameters specifying the initial conditions of the phase-space density: the initial energy spread σ_p and current I_0.

To seed the microbunching instability in the simulation, the initially homogeneous phase-space density is perturbed by a modulation with a given wavelength λ and perturbation amplitude ϵ, via $\Psi(q,p) \mapsto (1+\epsilon)\cos(2\pi q/\lambda)\Psi(q,p)$.

Setting the longitudinal size of the simulation window equal to the perturbation wavelength, the total phase-space density is given by the *periodic continuation* of the part contained in the simulation window. Support for this type of phase-space topology has been implemented in SelaV$_{\mathrm{1D}}$, including the correct treatment of resulting periodic boundary conditions in the routines that compute the collective maps. When the bunch is compressed during the course of the simulation, the simulation window is resized accordingly.

The perturbation theory derived in Section 5 can be applied to this model by treating each bunch compression stage as a single microbunching step. This provides an excellent opportunity to cross-validate the numerical results provided by SelaV$_{\mathrm{1D}}$ against the theory and vice versa.

6.3.1 Single-Stage Gain Functions

To determine the microbunching gain function numerically, multiple simulation runs are needed. At the beginning of each run, the initial phase-space density is perturbed on a wavelength λ with a perturbation strength ϵ, as described above. In each run, the perturbation wavelength is set to a different value, while all other parameters constant are kept constant. The perturbed phase-space density is then propagated according to the discrete-time model. After the bunch-compression chicane, the charge density of the phase-space density is calculated and written to a file. This file then contains the values of the charge density $\rho_j = \rho(q_j)$ at equidistant longitudinal positions $q_0, \ldots, q_{N-1}$, where N is determined by the effective resolution chosen for the simulation. As the simulation window is chosen so that it contains exactly one perturbation wavelength, the grid constant is $q_{j+1} - q_j = \lambda/N$. This process is repeated for an array of initial wavelengths λ_m, which cover the critical part of the gain curve, around the region of maximum gain. In a post-processing step, the charge-densities are further evaluated. Using discrete Fourier transform (DFT) [87], the harmonic content of the charge-densities can be determined:

$$\tilde{\rho}_n(\lambda_m) \equiv \sum_{j=0}^{N-1} \rho_j(\lambda_m)\, e^{-2\pi\, i\, j\, n/N}. \tag{6.3.1}$$

The amplitude of the nth harmonic is equal to double the absolute value of the nth component $\tilde{\rho}_n$ of the DFT. A numerical approximation $G^{[n]}$ of the gain of the nth

harmonic of the initial wavelength λ_m can therefore be obtained via

$$G_{[n]}(\lambda_m) = \frac{2}{\epsilon^n} \frac{|\tilde{\rho}_n(\lambda_m)|}{|\tilde{\rho}_0(\lambda_m)|}, \tag{6.3.2}$$

where the factor $|\tilde{\rho}_0(\lambda_m)|^{-1}$ provides the required normalization to the DC part of the charge density.

From the free parameters of the discrete-time model, the normalized parameters A', B', and r' of the perturbation theory can be calculated directly:

$$A' = \frac{L}{\bar{r}} \frac{e\, I_0\, Z_0}{\gamma\, \sigma_p}, \quad B' = C \underline{M}_{56} \frac{\sigma_p}{E_{\mathrm{BC}}}, \quad \text{and } r' = \frac{\bar{r}}{\gamma}, \tag{6.3.3}$$

with $\gamma = \frac{\bar{E}}{\mathrm{m_e}\, c^2}$. With this, the theoretical gain functions given in Equations (5.8.26),(5.8.40), and (5.8.45) can be evaluated.

6.3.1.1 Small Perturbation

Figure 6.9 shows the comparison of the numerical gain function, calculated from $\mathrm{SelaV_{1D}}$ simulations as described above, and the analytical gain functions derived from the perturbation theory, up to third order, for an example parameter set. These parameters are comparable to a possible working point of the first bunch compression stage of FLASH2020+. In these simulations, the perturbation parameter was set to a small value $\epsilon = 10^{-3}$. As can be seen, the numerical and analytic gain functions show excellent agreement. This agreement provides strong verification not only for the validity of the perturbation theory, but also for the numerical correctness of the simulation code $\mathrm{SelaV_{1D}}$.

6.3.1.2 Large Perturbation

In the previous example, the perturbation parameter ϵ was chosen to a small value. The gain function of the nth harmonic is, in that case, dominated by the nth order terms of the perturbation series. If ϵ is not much smaller than unity, the nth harmonic is also affected by terms in the perturbation series of order larger than n. For instance, in Equation (5.8.44), the contribution of the third-order perturbation term to the gain of the first harmonic was derived explicitly. Figure 6.10 shows the results of a $\mathrm{SelaV_{1D}}$ simulation with the same parameters as in Figure 6.9 but with $\epsilon = 0.05$. It can be seen that the first-order analytic gain functions does not agree well with the numerical results. Better agreement is achieved when the third-order corrections given by Equation (5.8.44) are included, which yields the third-order gain function for the first harmonic $G_1(2\pi/\lambda) + \epsilon^2\, G_3^I(2\pi/\lambda)$.

For wavelengths larger than 80 µm, this third-order gain function matches the numerical results well. For wavelengths smaller than that, the third-order gain function predicts an extremum which is not reflected in the numerical results. This disagreement in the small wavelength regime is the result of an effect that is illustrated in Figure 6.11. In Figure 6.11a, the phase-space density after the first bunch-compression chicane is depicted that results from a initial perturbation at the wavelength $\lambda \approx 49\,\mu\mathrm{m}$. This perturbation wavelength is in the region of maximum disagreement of theory and simulation depicted in Figure 6.10. In this case, multiple periods of the energy-modulated phase-space density overlap and contribute to the phase-space density in the simulation window, which longitudinally covers one compressed initial modulation wavelength. This is due to the relatively large value of the impedance function at this perturbation wavelength, resulting in large energy modulations, combined with the fact that the

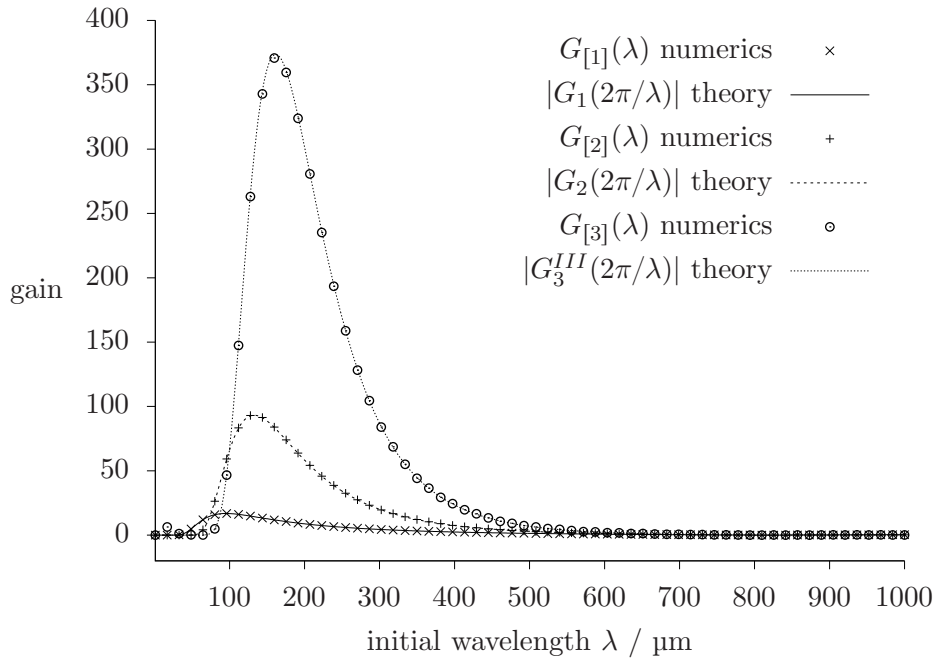

Figure 6.9: Comparison of the numerical gain functions up to third order, obtained from $\mathrm{SelaV_{1D}}$ simulations and the analytic gain functions resulting from the perturbation theory. For this example, the following parameters of the discrete-time model were chosen: $I_0 = 25\,\mathrm{A}$, $\sigma_p = 5\,\mathrm{keV}$, $L = 25\,\mathrm{m}$, $\bar{E} = 100\,\mathrm{MeV}$, $\bar{r} = 300\,\mu\mathrm{m}$, $C = 4$, $\underline{M}_{56} = 140\,\mathrm{mm}$, and $E_{\mathrm{BC}} = 145\,\mathrm{MeV}$. In the simulations, the perturbation parameter was set to $\epsilon = 10^{-3}$.

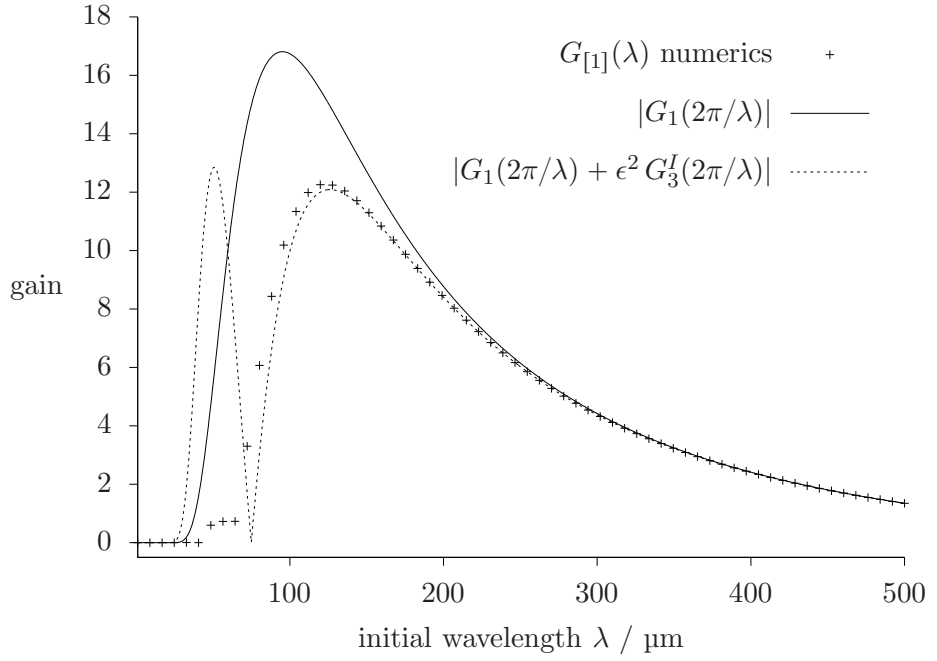

Figure 6.10: Gain functions of the first harmonic, evaluated using the same parameters as in Figure 6.10, but with $\epsilon = 0.05$. Plotted are the numerical gain function (plus signs), the first-order analytic gain function (solid), and the analytic gain function including third-order corrections (dotted).

transverse motion caused by the dispersion of the chicane relative to the modulation wavelength is larger at small wavelengths. If this *over-shearing* effect occurs, the instability shows a strong non-linear response to the perturbation so that already the low harmonics are affected by very high-order terms of the perturbation series. In that case, even the third-order gain function does not quantify the actual gain properly. At longer wavelengths, such as $\lambda \approx 120\,\mu\text{m}$ shown in Figure 6.11b, this over-shearing effect does not occur. Then, the third-order gain function is sufficient to properly approximate the actual microbunching gain. This explains the good agreement between theory and numerics for long wavelengths shown in Figure 6.10.

This example illustrates important caveats that have to be kept in mind when interpreting analytic gain functions: Only when the perturbation amplitude is small, non-linear contributions of higher-order terms of the perturbation series to lower-order harmonics can be neglected. The instability is then well-described by the gain functions derived from the perturbation theory at all wavelengths. If the perturbation amplitude is not small, non-linear contributions can become significant. This significance can depend on the perturbation wavelength, so that the analytic gain function might predict the gain correctly only in certain wavelength regions. Including higher-order corrections can extend the range of validity of the analytic gain function.

6.3.2 Two-Stage Gain Functions

The discrete-time long-bunch model can be easily extended to the case of two bunch-compression stages. The second stage is described by a second set of parameters $\bar{E}_2$, $\bar{r}_2$, L_2, $E_{\text{BC},2}$, $\underline{M}_{56,2}$, and C_2. From these, the parameters A_2 and $B_2^\dagger$ can be calculated, which appear in the analytic first-order two-stage gain function $G_2^\dagger$ given by Equation (5.8.52), as described in that section. To simulate the two-stage case, the simulation is set up exactly as in the single-stage case but with an additional microbunching step, reflecting the second compression stage. As before, the charge-density is written to a file after the second compression stage and the numerical gain function of the first harmonic $G_{[1],2}$ is calculated using the same method as in the single-stage case. In Figure 6.12, a comparison of the numerical gain function $G_{[1],2}$ and the analytic first-order two-stage gain function $G_2^\dagger$ is shown. As can be seen, they show excellent agreement. This result verifies that the perturbation theory is valid also in the two-stage case.

In literature, a method might be encountered which aims to to estimate the two-stage gain function without resorting to the full perturbation theory [97]. There, the gain functions of both compression stages are calculated independently, and the total two-stage gain function is estimated via the product of the two independent single-stage gain functions. For the second stage, the initial beam parameters are amended to take into account the compression from the first stage: $I_{0,2} = C_1\, I_{0,1}$ and $\sigma_{p,2} = C_1\, \sigma_{p,1}$. If the first-order gain functions of the first and second stage calculated in this way are denoted by $\check{G}_{1,1}(\lambda)$ and $\check{G}_{1,2}(\lambda)$, respectively, the estimated two-stage gain function is

$$\check{G}(\lambda) = \check{G}_{1,1}(\lambda)\, \check{G}_{1,2}(\lambda/C_1). \tag{6.3.4}$$

Of course, this is a rather crude approximation, as it neglects the influence of the first stage on the phase-space density completely, apart from the fact that it is compressed. As can be seen from Figure 6.12, the estimate resulting from this "method of multiplied gains" is indeed not a good approximation of the actual first-order gain function and should be used with great caution, if at all.

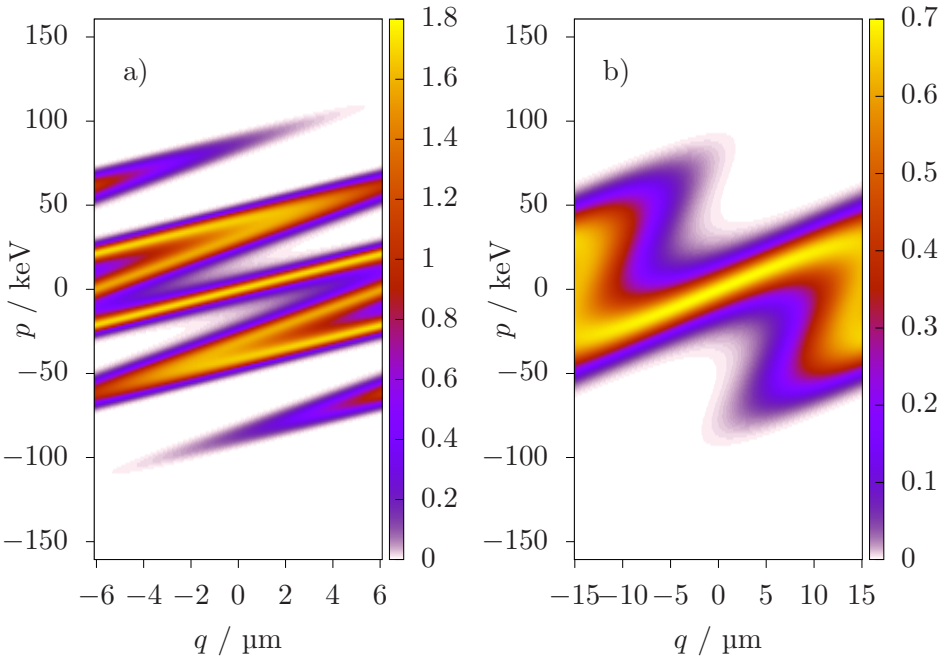

Figure 6.11: Phase-space densities after a bunch compression stage with compression factor $C = 4$, simulated by SelaV$_{1D}$ using the same parameters as in Figure 6.10. The initial perturbation wavelength is a) $\lambda \approx 49\,\mu m$ and b) $\lambda \approx 120\,\mu m$.

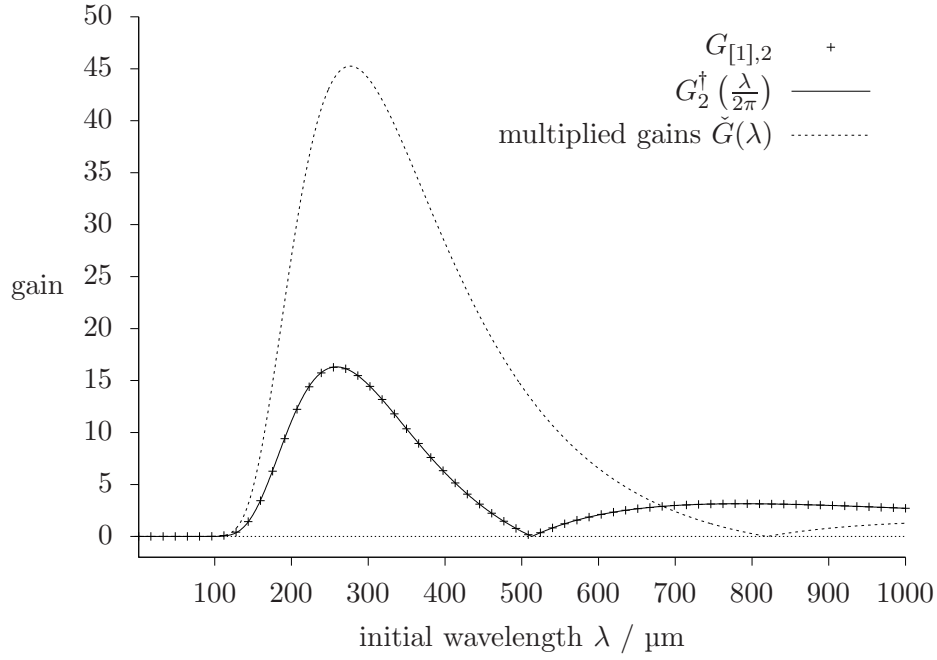

Figure 6.12: Gain functions of the first harmonic after two bunch compression stages: simulated by SelaV$_{1D}$ (plus signs), first-order two-stage perturbation theory (solid), and the result of the method of multiplied gains (dotted). The parameters of the first stage are the same as in Figure 6.9 and the parameters of the second stage are $L_2 = 50\,m$, $\bar{E}_2 = 350\,MeV$, $\bar{r}_2 = 200\,\mu m$, $C_2 = 8$, $\underline{M}_{56,2} = 70\,mm$, and $E_{BC,2} = 550\,MeV$. The perturbation amplitude was set to $\epsilon = 10^{-3}$.

6.4 Full Phase-Space Density Simulations for FLASH2020+

Thanks to its tree-based domain decomposition approach, SelaV$_{1D}$ is not only capable of simulating simplified models with periodic boundary conditions – such as the discrete-time long-bunch model shown in Section 6.3 – but also full phase-space densities with a compact support can be simulated. In the discrete-time long-bunch model, non-linear effects of neither the rf-cavities nor the magnetic chicanes could be included as that would have broken the periodicity condition. Such restrictions do not exist when the full phase-space density is simulated. Full-bunch simulations therefore enable not only a more detailed description of the bunch but also allow to treat a more detailed model of the longitudinal dynamics in the accelerator beamline.

In the following, the beamline model for full-bunch simulations with SelaV$_{1D}$ is described. Further, full-bunch simulation results are presented that relate to the upgrade project FLASH2020+ of the free-electron laser FLASH in Hamburg [30, 98]. Some of the results presented in this section are also published in reference [99] by the author.

6.4.1 Simulation Model and Setup

In an FEL injector beamline, the longitudinal single-particle dynamics are typically dominated by the rf-cavities and the magnetic chicanes. The longitudinal dynamics therefore governed by the operating parameters of these elements and their positions within the lattice. Coupling between the transverse and longitudinal dynamics can typically be neglected, apart from the longitudinal dispersion that is created in the chicanes. Regarding collective effects, the longitudinal dynamics are influenced by the transverse charge distribution, as it in principle affects the longitudinal space-charge forces. In the one-dimensional space-charge model introduced in Section 4.1, this dependence is fully described by a single parameter, namely the characteristic local transverse beam size.

To set up a full-bunch simulation with SelaV$_{1D}$, the user provides a description of the accelerator lattice and data for the optical functions $\beta_{x,y}$ along the beamline. A lattice description consists of the positions, lengths, and default operating parameters of the rf-cavities and chicane magnets. Beam optics data is specified as tuples of the longitudinal position and the horizontal and vertical optical β functions. Together with the transverse emittances – which are also specified by the user – this determines the characteristic transverse beamsizes along the beamline.

From this information, a preprocessing script generates an input file for SelaV$_{1D}$ with appropriate instructions to simulate the propagation of a bunch through the lattice. This input file contains a number of parameters that can be adjusted by the user, such as the operating parameters of the rf-cavities and the chicane magnets. Crucially, the input file implements operator splitting to treat collective effects with higher fidelity: drift spaces, cavities and chicane magnets can be split into multiple microbunching steps, each comprising a collective kick followed by a single-particle map. The length of these steps is selectable by the user. With respect to the representation of drift spaces, SelaV$_{1D}$ does not rely on the ultra-relativistic approximation so that longitudinal dispersion due to velocity effects can be taken into account.

6.4.2 Compression Working Points

One of the goals of the simulations presented below is to determine the influence of different compression schemes on the microbunching instability at FLASH after the FLASH2020+ upgrade. To this end, two compression working points (WPs) were studied. The first working point, referred to as WP1, was conceived to provide final beam

				BC1			BC2		
	I_0	C_1	C_2	E	α	M_{56}	E	α	M_{56}
WP1	31.2 A	4	4	146 MeV	16°	139.9 mm	550 MeV	5°	72.6 mm
WP2	20.0 A	5	5	143 MeV	16°	139.9 mm	550 MeV	5°	72.6 mm

Table 6.1: Compression parameters of two compression working points.

	ACC1		ACC39		ACC23	
	A	ϕ	A	ϕ	A	ϕ
WP1	163.8 MeV	11.87°	19.94 MeV	−175.39°	415.1 MeV	13.27°
WP2	161.3 MeV	12.91°	19.92 MeV	−174.23°	413.3 MeV	10.07°

Table 6.2: RF settings calculated for two compression working points.

parameters that are particularly suited for the seeded operation of the FLASH1 beamline. Based on the results presented below, the microbunching performance of WP1 was deemed improvable and a second working point, WP2, was specified. Table 6.1 summarizes the key parameters of both working points. The rf-settings that achieve the compression goals were calculated for both working points using the algorithm described in Section 3.6 and are shown in Table 6.2. For WP2, a slightly lower energy of 143 MeV at the first chicane was chosen, as it became apparent that this enables a more linear compression, while keeping the RF amplitudes within their technical limits. In fact, 143 MeV is the established nominal energy at the first chicane since the FLASH2020+ related shutdown in 2021/2022. In Figure 6.13, the resulting phase-space densities after the second chicane are depicted. They were calculated with SelaV$_{1D}$, assuming an initially Gaussian phase-space density. In both cases, the central parts of the bunch show a nearly linear dependence between the longitudinal position and the energy coordinate, which verifies the validity of the algorithm used to calculate the rf-settings. As a consequence, the final charge densities remain close to a Gaussian distribution. In particular, the charge densities do not feature an unwanted current spike in the head of the bunch, which can easily occur if the rf-settings are chosen less carefully.

6.4.3 Verification of the Large-Scale Dynamics

In a first effort to verify the validity of the results of simulations with SelaV$_{1D}$, a comparison with the well-tried 6D particle tracking code `elegant` [63,100] was made. In this section, we show comparisons of what could be called the *large-scale dynamics*, namely those that determine the overall shape of the longitudinal phase-space distribution on length scales in the order of the bunch length. This is in contrast to the *small-scale dynamics*, which occur on smaller length scales, such as the microbunching instability.

Both `elegant` and SelaV$_{1D}$ were set up with the same lattice and the same initial conditions for the bunch, for which a smooth Gaussian distribution was chosen, as not to seed the microbunching instability. The optics data needed for SelaV$_{1D}$ was precalculated for the lattice by MAD8 [60]. Both simulations start at the entrance of first rf-module in order to exclude the low-energy gun region. As a figure of merit, the bunch properties – in particular the charge density, the local mean energy deviation and the local energy spread along the bunch – are compared after the second bunch-compression chicane. To determine these quantities from the particle-based output of `elegant`, the particles are binned into longitudinal slices, from which the local beam properties are estimated via the statistical momenta of their energy distribution.

Figure 6.14 shows the case where no collective effects are included in the simulation

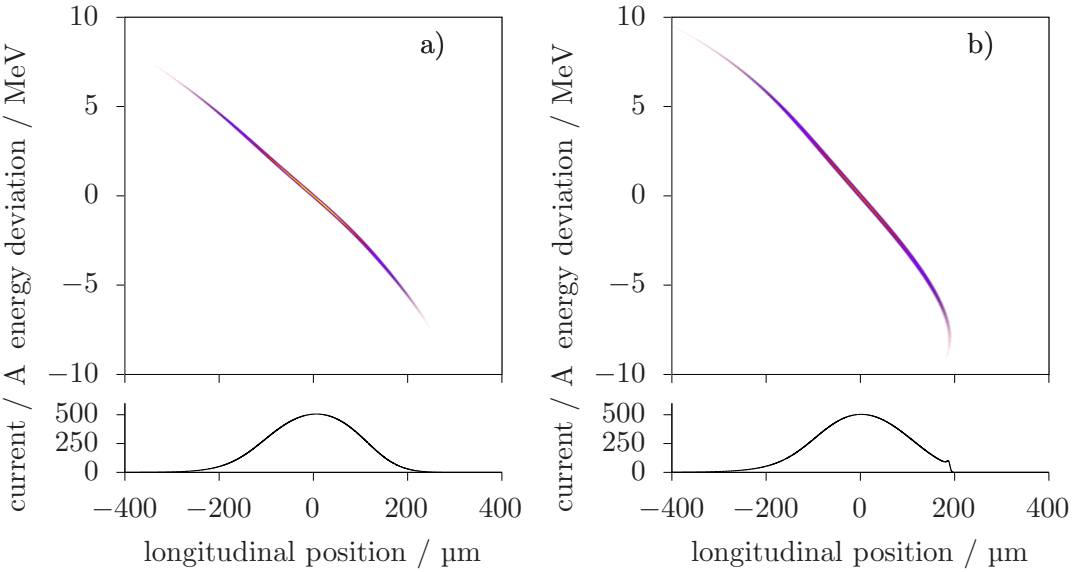

Figure 6.13: Plots of the phase-space distributions after the second bunch compression chicane and the corresponding charge densities for a) WP1 and b) WP2.

so that only the single-particle dynamics of the two codes are compared. The excellent agreement indicates that the model of the longitudinal dynamics implemented in SelaV$_{1D}$ is compatible with that used in `elegant`. As `elegant`, in contrast to SelaV$_{1D}$, simulates also the transverse single-particle dynamics, these results further verify that the longitudinal dynamics are indeed not strongly coupled to the transverse dynamics, and the longitudinal-only model of SelaV$_{1D}$ is adequate.

Figure 6.15 shows the results, when longitudinal space-charge forces are enabled in both simulation codes. The rf-settings were tweaked slightly to counteract the reduced compression that result from the LSC forces and achieve the same compression as before. While the agreement – especially in the head and tail parts of the bunch – is still very good, the results do deviate slightly in the central parts of the bunch. It is easily conceivable that this deviation results from slight differences in the transverse beam sizes considered by the two codes, which influence the longitudinal dynamics via the LSC impedance. Both SelaV$_{1D}$ and `elegant` use the same model for LSC in which a one-dimensional longitudinal impedance function is determined from the characteristic transverse beam size. In `elegant`, this characteristic beamsize is calculated from the standard deviations of the particles in the transverse configuration spaces and features are implemented that allow the user to smooth the longitudinal charge density to reduce the impact of the numerical shot-noise. In SelaV$_{1D}$, the beamsize is determined from the optical functions calculated by MAD8 and no smoothing is required. Slight differences in this order of magnitude are therefore to be expected.

Overall, it can be concluded that the agreement of the large-scale beam dynamics predicted by the two codes is more than satisfactory.

6.4.4 Numerical Gain Functions

After the validity of the full-bunch simulation setup was established, it is now used to investigate the microbunching instability numerically. A straight-forward approach – which was also followed in Section 6.3 – to quantify the microbunching instability is to imprint a modulation onto the initial charge density of an otherwise smooth bunch and

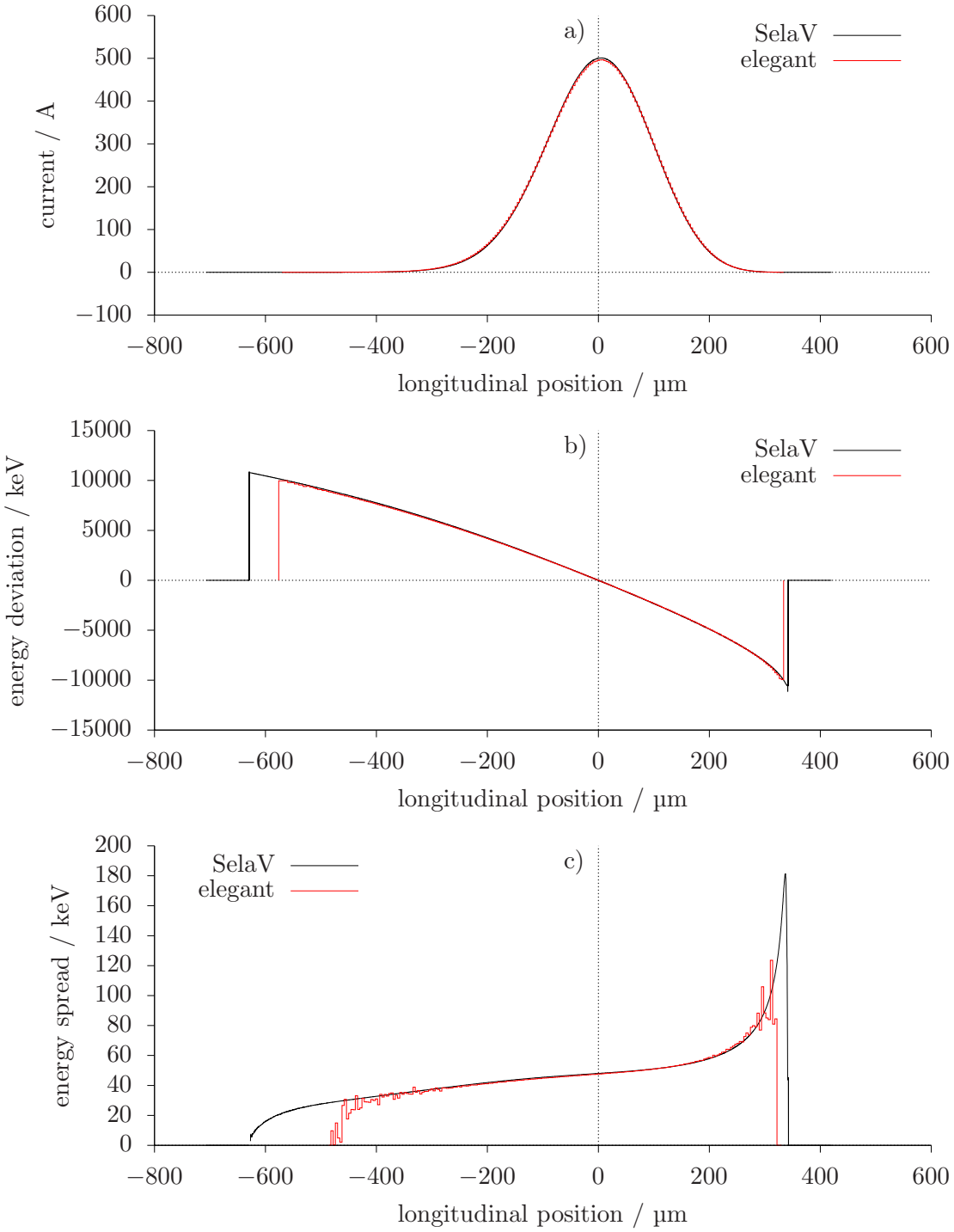

Figure 6.14: Comparison of simulation results by SelaV_{1D}, and `elegant` for the bunch properties after the second compression chicane for WP1. Collective effects are turned off in the simulations.

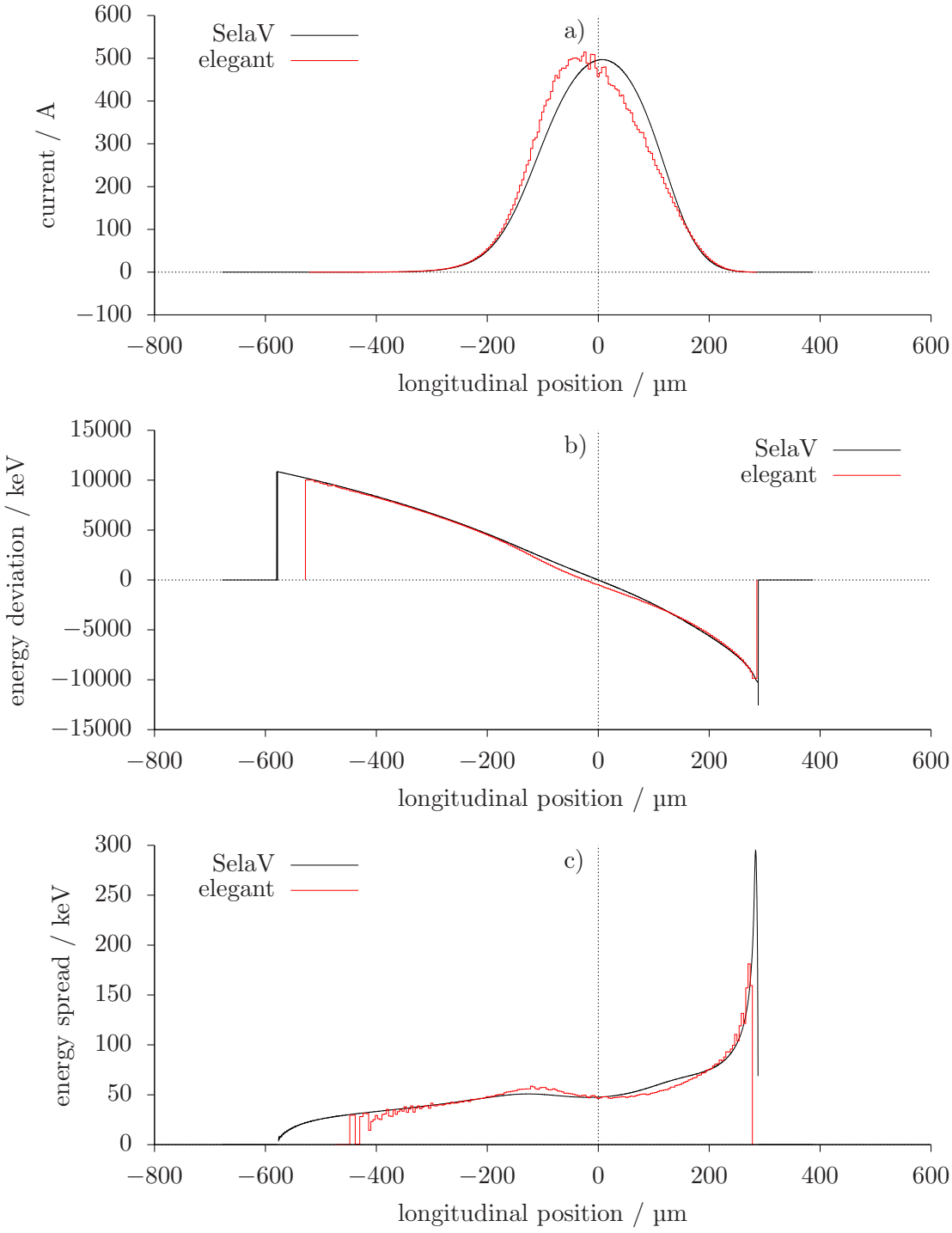

Figure 6.15: Comparison of simulation results by SelaV_{1D}, and `elegant` for the bunch properties after the second compression chicane for WP1. Longitudinal space-charge effects are enabled in the simulation.

determine the relative increase of the modulation amplitude after the bunch traversed the beamline. By defining $\rho_f[A,k](q)$ as the final charge density that results from an initial modulation with amplitude A at a wavenumber k, then the gain function can be defined via the Fourier transform $\tilde{\rho}_f[A,k](\omega) \equiv \mathcal{F}_{\omega \leftarrow q}$ as

$$G_A(C,k) = \frac{|\tilde{\rho}_f[A,k](C\,k)| - |\tilde{\rho}_f[0,0](C\,k)|}{A}, \tag{6.4.1}$$

where C is the total compression factor of the beamline. The arguments of the Fourier transforms are scaled by the compression factor in order to account for the fact that the modulation is shifted towards smaller wavelengths due to the compression. In this form, $G_A(C,k_0)$ is the gain of a modulation at a given *initial* wavenumber k_0. In many cases, it is more instructive to report the gain at a given *final* wavenumber k_f, which is given by $G_A(C,k_f/C)$. By subtracting the term $\tilde{\rho}_f[0,0]$, contributions stemming from the charge-density spectrum of an unperturbed bunch – which might be referred to as the *form factor* of the bunch – are removed. The linear gain function is then defined by the limit $G_{\mathrm{lin}} \equiv \lim_{A\to 0} G_A(C,k)$. Numerically, this limit can be approximated by choosing a sufficiently small perturbation amplitude. In practice it became apparent that by setting $A = 10^{-4}$ a good approximation of the linear gain function can be achieved.

For the FLASH2020+ project, it is of particular interest to determine to which degree the laser heater that is installed as part of the upgrade [101] can suppress the microbunching instability. Situated upstream of the first bunch compression chicane, the laser heater increases the energy spread of a bunch by inducing an energy modulation at an optical wavelength in a modulator undulator, which is subsequently over-folded in the chicane. For the purpose of studying the microbunching instability – which typically occurs on the micrometer scale – the optical-scale energy modulation can be neglected so that the effect of the laser heater can be treated by considering just the induced increase of the initial the energy spread. To resolve optical-scale modulations on the full phase-space density of the bunch, a prohibitively fine effective grid resolution would be required. If at all, studying optical-scale modulations is therefore only possible using simplified models, such as the aforementioned long-bunch approximation, which employs periodic continuation of the phase-space density. Figure 6.16 shows the numerical linear gain curves calculated using $\mathrm{SelaV}_{\mathrm{1D}}$ for both working points and three different values of the initial energy spread σ_p. The smallest considered energy spread of 3 keV corresponds to the value that can expected from the FLASH gun [102]. Energy spreads of 5 keV and 7 keV are well achievable with the laser-heater setup, requiring only a fraction of the maximum available laser-pulse energy [101,103]. For the unperturbed initial phase-space density, a bi-variate Gaussian distribution was chosen.

It can be seen that at WP2, the microbunching gain is approximately four times smaller compared to WP1, independent on the initial energy spread. Between the two working points, the locations of the maxima are comparable, with WP2 showing only a small shift towards longer wavelengths. Maxima occur at final wavelengths of approximately 8 µm, 11 µm and 15 µm, for 3 keV, 5 keV and 7 keV energy spread, respectively. The results for the 3 keV case are consistent with experimental observations obtained at the pre-upgrade FLASH lattice [104]. At WP2, the microbunching instability is completely suppressed at the highest considered laser heater setting of 7 keV.

These results show that the microbunching instability can be successfully suppressed using intermediate settings of the laser heater, in conjunction with a well-chosen compression working point. It has to be kept in mind when interpreting these results that increasing the initial energy spread as well as choosing a larger total compression factor leads to an increase of the final energy spread of the bunch. As WP2 generates a total

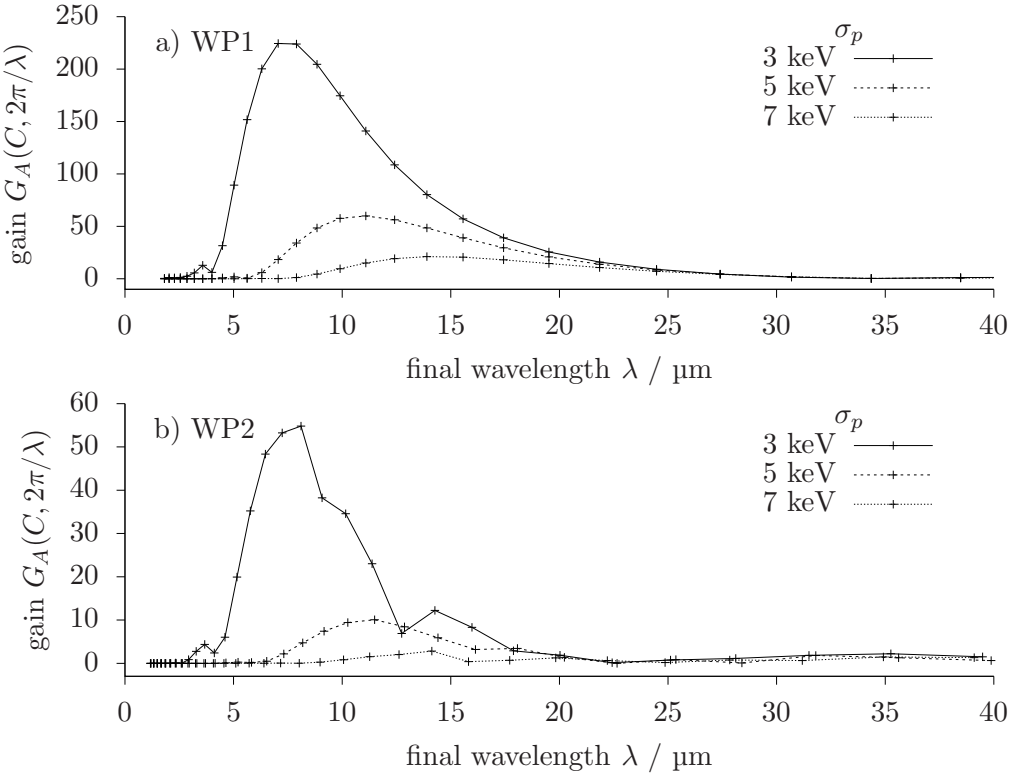

Figure 6.16: Numerical linear gain functions obtained with SelaV_{1D}, for two compression working points of the FLASH2020+ lattice and multiple values of the initial bunch energy spread σ_p. The perturbation amplitude is $A = 10^{-4}$.

compression factor of $C = 25$, an initial energy spread of 7 keV will result in an final energy spread of at least 175 keV, even without taking into account collective effects, while the smaller compression factor of WP1, $C = 16$, would result in a smaller lower bound for the final energy spread, namely 112 keV. It is well-known that the FEL process requires bunches with sufficiently small energy spread. At the same time, an excessive amount of microbunching can severely degrade the FEL performance. It is therefore to be expected that optimal FEL performance is achieved at an intermediate final energy spread which is large enough to sufficiently suppress the microbunching instability but small enough to drive the FEL instability.

6.4.4.1 Two-Parametric Gain Functions

In the definition of the gain function in Equation (6.4.1), the natural compression of the initial modulation wavelength is accounted for by scaling of the arguments of the involved charge-density spectra by the known compression factor C of the lattice. An even more revealing way to quantify the impact of the microbunching instability on the charge-density spectrum is to keep this scaling factor as a free parameter – instead of fixing it to the theoretical compression factor. The resulting *two-parametric gain function* $G_A(h, k)$ then describes the contribution of an initial modulation at the initial wavenumber k to the amplitude of the final charge-density modulation at the multiple $h\,k$ of the initial wavenumber.

Figures 6.17a and 6.17b show plots of the two-parametric gain functions for the two working points, respectively. In this case, the perturbation parameter was chosen small enough, $A = 10^{-4}$, to keep the instability in the linear regime. It can be seen that the

maxima of the gain function occur at the multiples $h \approx 16 = C_{\mathrm{WP1}}$ and $h \approx 25 = C_{\mathrm{WP2}}$ of the initial wavelength, respectively, which corresponds to the compression factors of the two working points. The broadening of the peaks in the h-direction is a result of the finite support of the charge density. With respect to the final modulation wavelength, the gain maxima occur at $\lambda_{\mathrm{max,WP1}} \approx 130\,\mu\mathrm{m}$ and $\lambda_{\mathrm{max,WP2}} \approx 200\,\mu\mathrm{m}$, respectively. Both correspond to an initial modulation wavelength of $\lambda_{\mathrm{max,WP1}}/C_{\mathrm{WP1}} \approx \lambda_{\mathrm{max,WP2}}/C_{\mathrm{WP2}} \approx 8\,\mu\mathrm{m}$, which is consistent with the results of the one-parametric gain functions.

By increasing the perturbation parameter, non-linear effects of the instability can be studied. Figures 6.17c and 6.17d show the two-parametric gain functions for $A = 10^{-2}$. It can be seen that additional gain-peaks appear around multiples of the compression factors of the respective working points, namely $h \approx 16, 32, 48$ for WP1 and $h \approx 25, 50$ for WP2. These peaks quantify the contributions to the higher harmonics of the compressed initial perturbation wavelength. In addition, it can be seen that the wavelength of maximum gain is shifted towards longer wavelengths. For WP1, the maximum gain at the fundamental now occurs at $\lambda_{\mathrm{max,WP1}} \approx 230\,\mu\mathrm{m}$ and for WP2 at $\lambda_{\mathrm{max,WP2}} \approx 250\,\mu\mathrm{m}$, which correspond to initial modulation wavelengths of approximately $14\,\mu\mathrm{m}$ and $11\,\mu\mathrm{m}$, respectively. We note that while in the linear regime, the initial modulation wavelengths that results in the maximum microbunching gain were the same for both working points, in the non-linear regime they differ. The shift towards longer wavelengths can be explained by the fact that the larger density modulation amplitude results in large energy modulation amplitudes induced by the LSC forces, which in turn lead to a stronger overfolding of the modulations so that modulations at smaller wavelengths are more likely to be suppressed. As a result, maximum gain occurs at larger wavelengths. Stronger overfolding also explains the overall reduced gain values, compared to the linear case.

6.4.5 Shot-Noise Simulations

Until now, the microbunching instability was investigated by modulating the initial charge density at a single wavelength and analyzing the resulting amplification of the modulation amplitude. This approach yields gain functions which give us a good picture of the wavelength regions in which modulations are most significantly amplified by the microbunching instability. However, in order to determine the level of microbunching that is to be expected for a given bunch it is necessary to compare the gain function with the spectrum of the density inhomogeneities that are actually present in the bunch, which we will call the *noise spectrum*. After all, it could be that the gain function peaks in a wavelength region which is well outside of the noise-spectrum bandwidth of the bunches under consideration. In that case, even if the gain function features a large peak, the observed level of microbunching will be small, as the bunch does not contain inhomogeneities that could actually be amplified by the microbunching instability, no matter how large the gain is.

Naturally, the question arises what causes the initial inhomogeneities in the charge density of a bunch. While it is conceivable that one might naively suspect modulations of the intensity of the photo-cathode laser, or even the intricacies of the emission process at the cathode as potential sources of the inhomogeneities, these can typically be ruled out if a small-bandwidth laser system and a slow-response-time cathode material is used [105].

A source of inhomogeneity that undoubtedly is at play results from the circumstance that an electron bunch is in fact an ensemble of single particles and as such is inherently inhomogeneous. Inhomogeneities of the charge density that arise from this particulate nature of the bunch are referred to as *shot noise* or *Poisson noise* [106]. It becomes

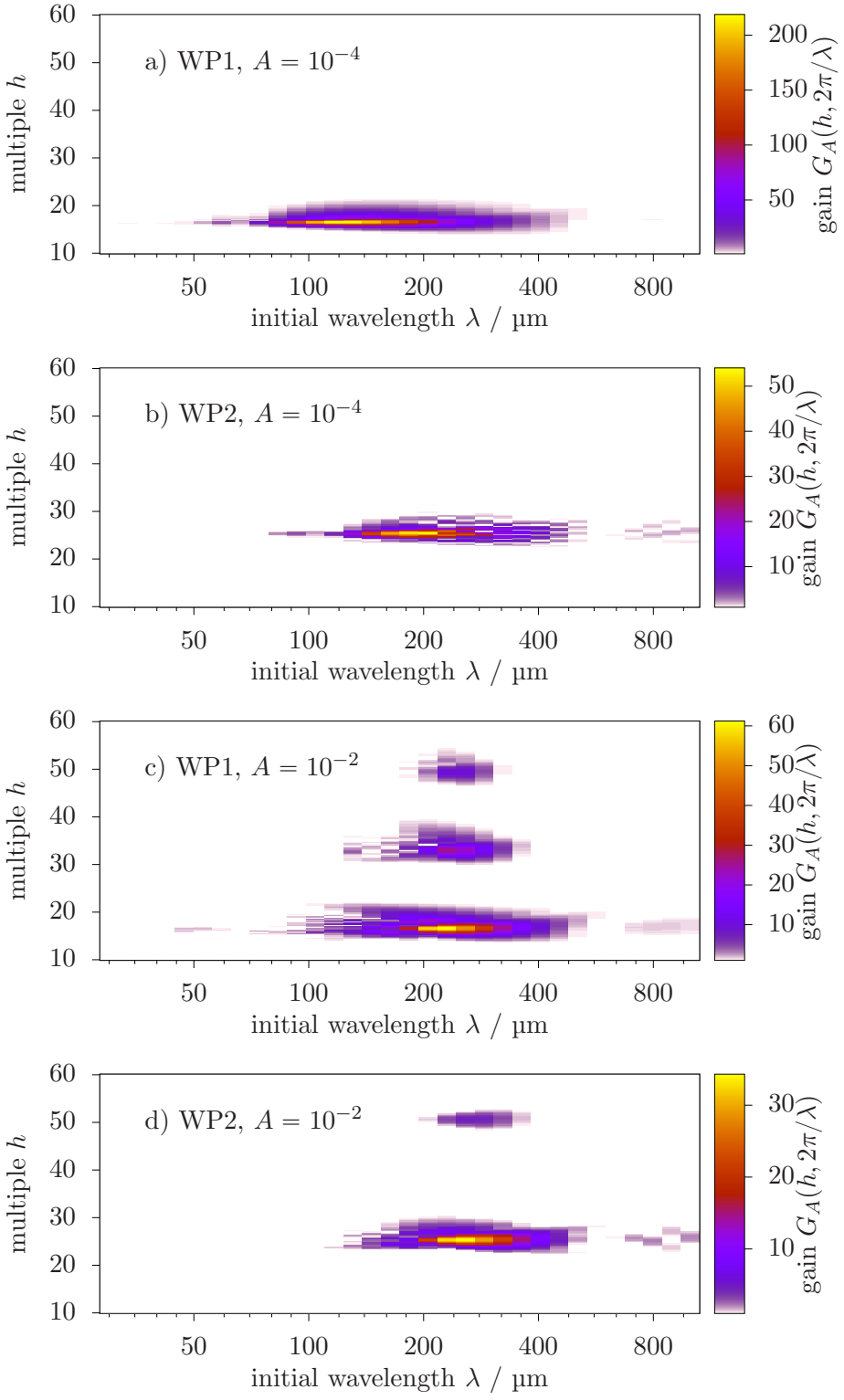

Figure 6.17: Two-parametric gain functions for two compression working points and two values of the perturbation amplitude A.

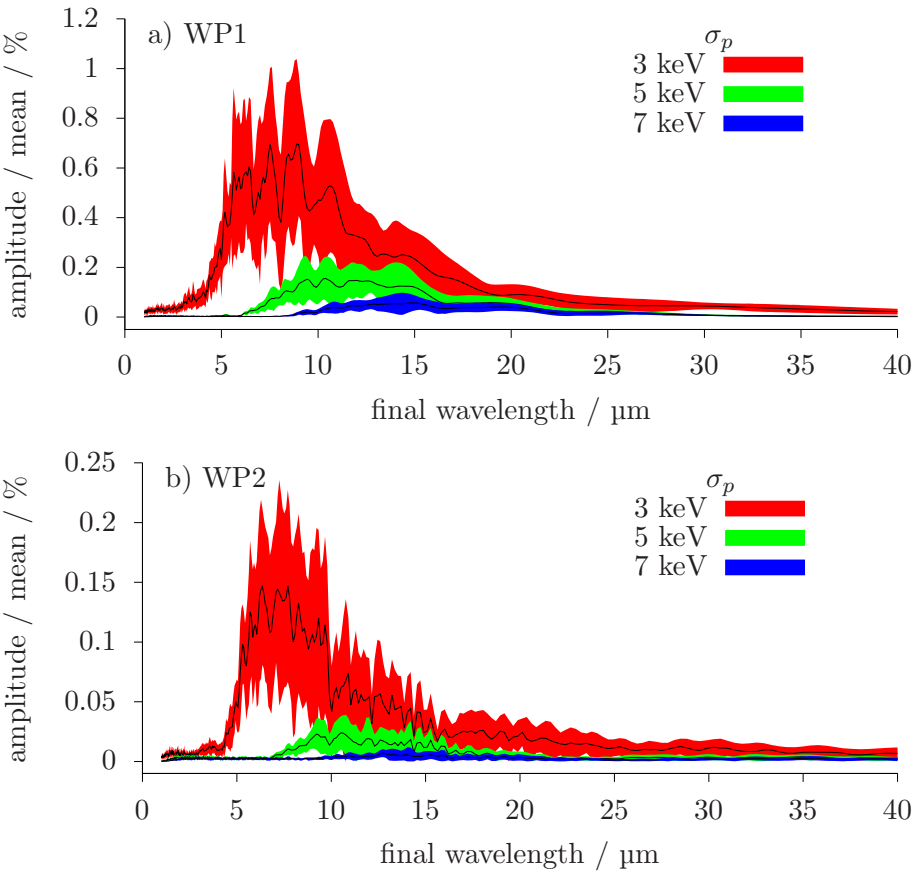

Figure 6.18: Charge-density spectra resulting from shot noise for a) WP1 and b) WP2, normalized to the mean charge density. Solid lines show the mean value of 16 different realizations of the initial shot noise and filled areas depict their standard deviations. The form factor of an unperturbed bunch has been subtracted.

apparent that the noise spectrum of shot noise is wavelength independent. As in this case modulations on all frequencies are present in the bunch with the same amplitude, the gain function is indeed a good approximation of the shape of the expected final microbunching spectrum.

In particle-tracking simulations, shot noise is inherently included as the bunch is actually represented as a particle ensemble. However, as in a simulation typically less particles are included than present in the physical bunch, care has to be taken to ensure that the shot-noise amplitude of the simulated bunch matches that of a bunch with the actual number of particles. In semi-Lagrangian simulation codes, such as SelaV_{1D}, shot noise is not automatically present, as a bunch is represented by a smooth numerical representation of the phase-space density, instead of single particles. To be able to study the microbunching amplification of shot noise, a routine was implemented in SelaV_{1D} that allows to add noise to an otherwise smooth phase-space density, by adding appropriately distributed random values to the grid samples, as described in Section 6.2.5. Figures 6.18a and 6.18b show the charge density spectra after the second bunch compression chicane of bunches that have been seeded this way for WP1 and WP2, respectively. Again, multiple values of the initial energy spread σ_p have been studied to investigate the efficacy of the laser heater. It can be seen that the shot-noise spectra are very much comparable to the gain functions shown in Figure 6.16 with respect to the location and

width of their peaks, as well as the relative reduction of the maximum with increasing energy spread. This good agreement with the linear gain function indicates that at the considered working points the magnitude of the shot-noise induced inhomogeneities is small enough to keep the microbunching instability in the linear regime.

In Figure 6.19, plots of the central parts of the phase-space densities after the second bunch compression chicane are shown for the different working points and initial energy spreads. These plots illustrate the fact that a reduction of the microbunching amplitude can only be achieved by paying the price of an increased final energy spread.

6.4.6 Cross-Comparison of the Shot-Noise Results

As part of the start-to-end simulation efforts for the FLASH2020+ project [107, 108], simulations with the particle-tracking code IMPACT-Z [109] were conducted [110]. In IMPACT-Z, a variety of collective effects is implemented, in particular a 3D model of the space-charge forces, a 1D model of CSR effects based on references [75, 76], and cavity wakefields. This gave the opportunity to compare $\mathrm{SelaV_{1D}}$ with a well-tried code with respect to the microbunching instability. Both codes were set up with the same lattice and the simulation was started after the first cavity of the first RF-module, again in order to avoid simulating the low-energy gun region. As the initial condition of the longitudinal phase-space density a bi-variate Gaussian distribution with $\sigma_p = 3\,\mathrm{keV}$ was chosen, with an additional chirp equivalent to the linearized kick of the first cavity. For the IMPACT-Z simulation, the transverse distribution was generated based on the theoretical optical function at that location. The number of particles used in the IMPACT-Z simulation was chosen to 620 million, which corresponds to a quarter of the particles present in a bunch with a charge of 400 pC. It was assumed that this ratio between simulated and physical particles is small enough to adequately reflect the shot noise.

Figure 6.20 shows a comparison of the phase-space densities after the second bunch compressor calculated by the respective codes for working point WP1. It can be seen from Figures 6.20a and 6.20c that the large-scale shape of the phase-space densities and the charge density distributions differs only slightly between the codes, predominantly in the head and tail parts of the bunch. For the IMPACT-Z simulation, the rf-settings had to be tweaked slightly in order to produce the correct final peak current, which indicates that the model IMPACT-Z uses for the longitudinal single-particle beam dynamics is not completely compatible to that implemented in $\mathrm{SelaV_{1D}}$. Also, it has to be noted that while in the IMPACT-Z simulation all implemented collective-effects were enabled, in the $\mathrm{SelaV_{1D}}$ simulation only LSC effects were included. With these differences in mind, the agreement of the small-scale bunch shape between the two codes is more than satisfactory. Figures 6.20b and 6.20d show close-ups of parts of the phase-space density around the center of the bunch, which show that also the form of the microbunching structures agrees well between the codes.

In Figure 6.21, a comparison of the normalized charge density spectra after the second bunch compressor predicted by $\mathrm{SelaV_{1D}}$ and IMPACT-Z is shown. As is is not possible to simulate an unperturbed bunch with IMPACT-Z due to the unavoidable shot-noise, in contrast to Figure 6.18 the form factor of the unperturbed bunch was not subtracted from the spectra and is therefore visible in this plot. It is apparent that the wavelength region and amplitude of the contributions of the microbunching instability predicted by the two codes agree very well. The aforementioned differences in the charge-density distribution are reflected in the form factors.

We see that the results produced by $\mathrm{SelaV_{1D}}$ and IMPACT-Z agree in general very well. However, much larger differences lie in the run times of the two codes. An IMPACT-Z run with the aforementioned number of simulated particles takes more than a day wall-

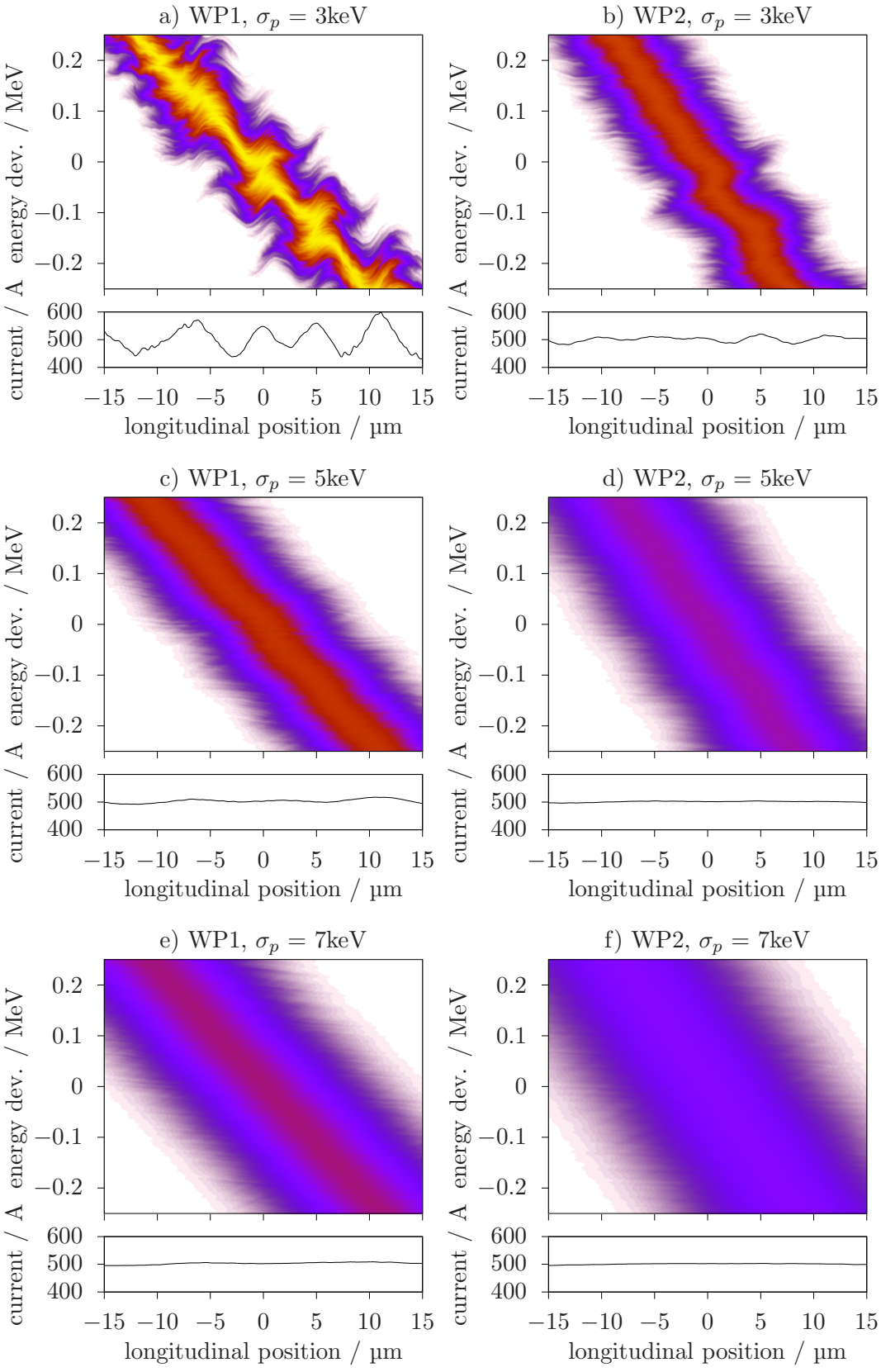

Figure 6.19: Plots of the central parts of the phase-space densities obtained from SelaV_{1D} after the second bunch compression chicane, depicting shot-noise inhomogeneities amplified by the microbunching instability for two compression working points and multiple values of the initial energy spread σ_p.

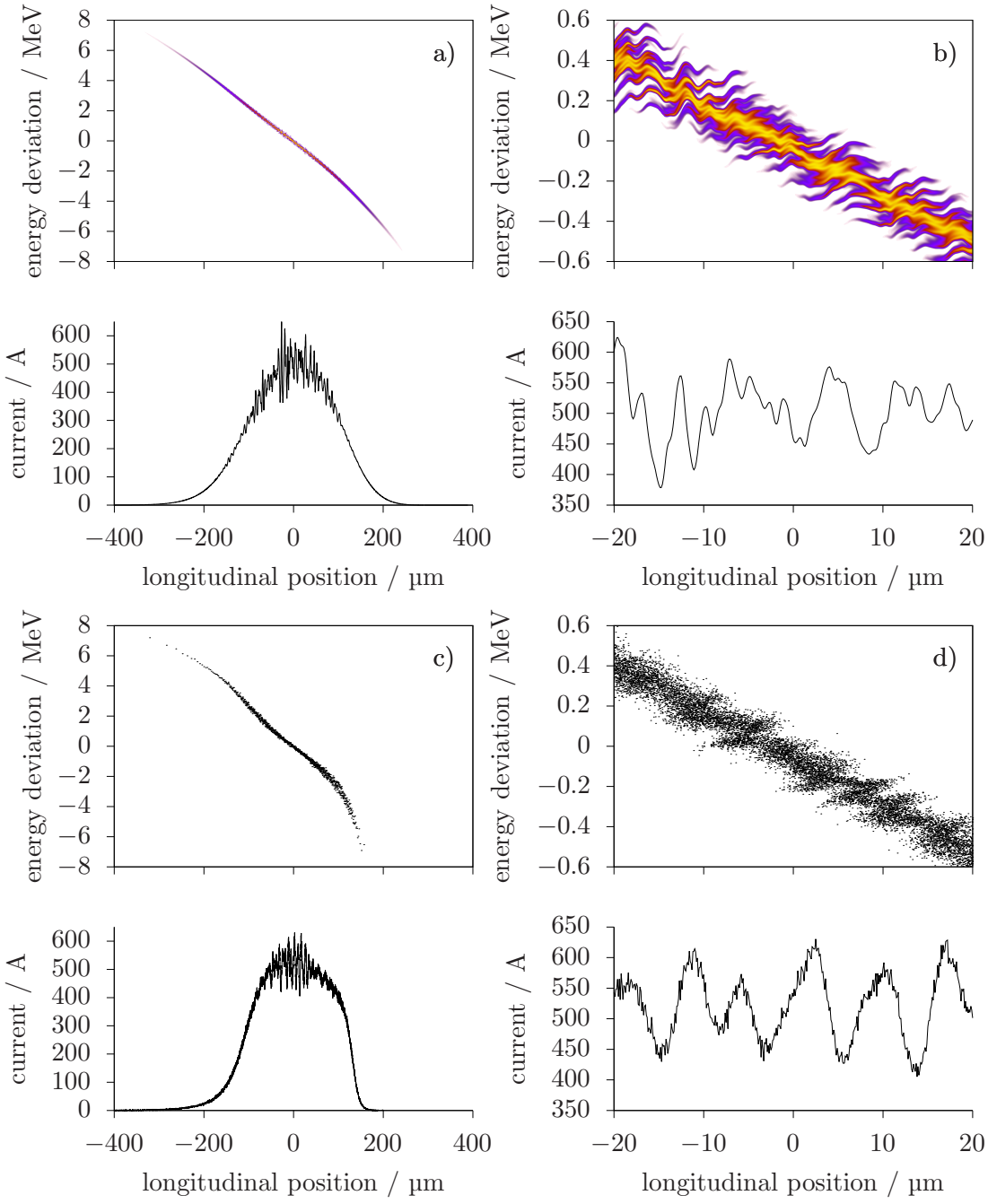

Figure 6.20: Phase-space densities and particle distribution resulting from a),b) SelaV$_{1D}$ and c),d) IMPACT-Z simulations, respectively, and the corresponding charge densities. In c) every 500.000th and in d) every 10.000th of the 620 million simulated particles is plotted in order to keep the graphic from saturating.

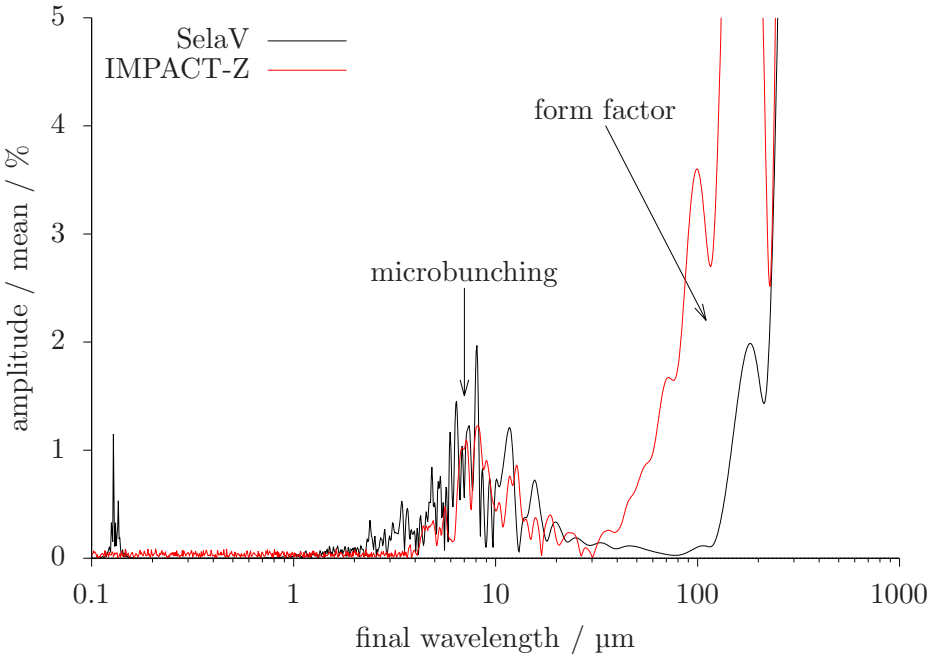

Figure 6.21: Comparison of charge-density spectra after the second bunch compressor resulting from simulations with SelaV_{1D} and IMPACT-Z.

clock time on a computer cluster with 64 CPUs [110]. A SelaV_{1D} run at an appropriate resolution takes less than 10 minutes on a personal computer with 4 cores.

7 Summary

In this work, the microbunching instability occurring in free-electron laser injectors was investigated via analytical as well as numerical methods.

On the analytical side, a perturbation theory for the microbunching instability was developed based on the formalism of collective Perron–Frobenius operators, which are the propagation operators of a phase-space density whose time-evolution is determined by a kinetic equation similar to Liouville's equation, which preserves the Lebesque measure and is sufficiently regular. By expanding the collective Perron–Frobenius operators with respect to their dependence on the initial phase-space density by means of their Fréchet derivatives, a Fréchet–Taylor series representation of the Perron–Frobenius operator was obtained. Using the expanded form of the operator, a perturbation series for a propagated phase-space density was derived. For a certain class of collective Perron–Frobenius operators, namely those that fulfill a homomorphy criterion, the Fréchet derivatives were derived explicitly. These findings were used to study the amplification of an initial charge density modulation due to the microbunching instability in a bunch compression stage. Considering a modulation on a single wavelength, this approach allowed not only to derive the linear gain function, but also the non-linear contributions which result in the generation of higher harmonics of the initial modulation of the charge density. Further, the approach allowed to study initial perturbations that consist of two modulations on two different wavelengths. It was shown that in this case wave mixing occurs. In particular, the second order perturbation term contains terms with a periodicity corresponding to the sum and the difference of the initial perturbation. A general approach to extend the perturbation theory to multiple microbunching stages was shown. For two microbunching stages, an explicit expression for the linear microbunching gain function was derived. By fixing the energy distribution function of the bunch and the collective impedance function, closed-form expressions of the terms of the phase-space density perturbation series and the associated charge densities were derived up to arbitrary order. These expressions were parameterized with respect to a set of normalized parameters, describing the energy amplitude and shearing of a microbunching modulation, which allowed to describe the gain process in a general manner. It further allowed to introduce universal normalized gain functions, which where derived explicitly up to third order.

A particularly insightful representation of the linear longitudinal single-particle beam dynamics in one and more bunch compression stages was derived in the form of a symplectic LDU decomposition. This finding crucially facilitated the application of the aforementioned perturbation theory, as it allowed to treat the system in a decorrelated frame.

On the numerical side, the development of the semi-Lagrangian Vlasov simulation code $\mathrm{SelaV_{1D}}$ was presented. A distinguishing feature of $\mathrm{SelaV_{1D}}$ is that it allows to represent the exotic longitudinal phase-space densities that are prevalent in FEL injectors – characterized by a strong non-linear correlation between the energy deviation and the longitudinal position – numerically efficient by employing tree-based domain decomposition. With $\mathrm{SelaV_{1D}}$ it is therefore possible to simulate the whole phase-space density of a

bunch using the semi-Lagrangian approach, which – in contrast to particle-tracking approaches – is free of artificial shot-noise. The absence of this class of numerical artifacts allows a precise numerical analysis of the microbunching instability, as the inhomogeneity of the initial phase-space density can be specified precisely, without unwanted contributions resulting from imperfections of the numerical representation. The basic working principle of SelaV$_{1D}$ was described as well as key features that facilitate its use for the simulation of collective beam dynamics in accelerator beamlines.

A simplified model of the microbunching process, the discrete-time long-bunch model, was simulated using SelaV$_{1D}$ and the results compared to analytical results obtained from the perturbation theory, with which they agreed excellently.

SelaV$_{1D}$ was further used to simulate two compression working points of FLASH after the FLASH2020+ upgrade. A primary goal of these simulation was to determine to which degree the microbunching instability could be suppressed by the laser heater that is part of the upgrade. These simulations where conducted with the full phase-space density and a more refined model of the beam dynamics. From these simulation results, numerical gain functions were calculated and compared between the two working points. As an extension of the linear one-parametric gain function, the concept of a two-parametric gain function was introduced, which allows to quantify the generation of higher harmonics in the non-linear case. Further, the emergence of the microbunching instability from shot-noise was investigated. Shot-noise spectra and phase-space features predicted by the semi-Lagrangian code SelaV$_{1D}$ agree well with simulation results from a particle-tracking code. In general, it was shown that SelaV$_{1D}$ is a reliable tool to investigate the mircobunching instability, that was successfully used to aid the design of bunch compression schemes for the FLASH2020+ upgrade project. Overall, the results indicate that the microbunching instability at FLASH after the FLASH2020+ upgrade can be successfully suppressed, even with intermediate laser heating, if a suitable compression working point is chosen. Investigating the impact of such microbunching mitigation efforts on the FEL performance was, however, beyond the scope of this work.

For the aforementioned compression working points it is mandatory to compress the bunch as linear as possible. To that end, an algorithm was developed to determine appropriate rf-settings which provide linear bunch compression.

Overall, in this thesis two powerful tool sets, one analytical and one numerical, were developed to determine the impact of the microbunching instability on the beam quality in free-electron laser injectors. By applying both tool sets, a better understanding of the microbunching instability – in particular with respect to non-linear contributions, the formation of higher harmonics, and its behaviour in multiple compression stages – was obtained and microbunching mitigation schemes for the FLASH2020+ project were investigated.

A Coordinates for a Type of Curl-Free Potentials

In addition to the regularly used coordinates outlined in Section 2.6.6, we want to introduce another set of coordinates. The distinguishing feature of these coordinates is that the transverse and electric potentials are included in the coordinates. As a result, the Hamiltonian is independent on the potentials and only a dependence on the longitudinal vector potential remains. These coordinates are defined by

$$\tilde{\boldsymbol{z}} \equiv \begin{pmatrix} \tilde{x} \\ \tilde{p}_x \\ \tilde{y} \\ \tilde{p}_y \\ \tilde{\tau} \\ \tilde{\delta} \end{pmatrix} \equiv \begin{pmatrix} Q_x \\ \frac{P_x - eA_x}{P_0} \\ Q_y \\ \frac{P_y - eA_y}{P_0} \\ ct_0(s) - ct\frac{\sqrt{(H-e\phi)^2 - m^2c^4}}{H-e\phi} \\ \frac{1}{c}\frac{\sqrt{(H-e\phi)^2 - m^2c^4}}{P_0} - 1 \end{pmatrix}. \tag{A.0.1}$$

In the following, it is shown that this indeed defines a canonical transformation if the potentials fulfill a certain condition. To this end we introduce the intermediate coordinates

$$\hat{\boldsymbol{z}} \equiv \begin{pmatrix} \hat{Q}_x \\ \hat{P}_x \\ \hat{Q}_y \\ \hat{P}_y \\ \hat{H} \\ \hat{t} \end{pmatrix} = \begin{pmatrix} Q_x \\ P_x - eA_x \\ Q_y \\ P_y - eA_y \\ \sqrt{(H-e\phi)^2 - m^2c^4} \\ t\frac{\sqrt{(H-e\phi)^2 - m^2c^4}}{H-e\phi} \end{pmatrix}. \tag{A.0.2}$$

Defining the auxiliary symbol $\Delta \equiv \frac{H-e\phi}{\sqrt{(H-e\phi)^2 - m^2c^4}}$, the Jacobian $\underline{M} \equiv \partial\hat{\boldsymbol{z}}_i/\partial\boldsymbol{z}_j$ of the coordinate transformation (A.0.2) is given by

$$\underline{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -e\frac{\partial A_x}{\partial Q_x} & 1 & -e\frac{\partial A_x}{\partial Q_y} & 0 & 0 & -e\frac{\partial A_x}{\partial t} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -e\frac{\partial A_y}{\partial Q_x} & 0 & -e\frac{\partial A_y}{\partial Q_y} & 1 & 0 & -e\frac{\partial A_y}{\partial t} \\ -e\frac{\partial \phi}{\partial Q_x}\Delta & 0 & -e\frac{\partial \phi}{\partial Q_y}\Delta & 0 & \Delta & -e\frac{\partial \phi}{\partial t}\Delta \\ \frac{et\left(\Delta^{-1}-\Delta\right)}{H-e\phi}\frac{\partial \phi}{\partial Q_x} & 0 & \frac{et\left(\Delta^{-1}-\Delta\right)}{H-e\phi}\frac{\partial \phi}{\partial Q_y} & 0 & \frac{t\left(\Delta^{-1}-\Delta\right)}{H-e\phi} & \frac{et\left(\Delta^{-1}-\Delta\right)}{H-e\phi}\frac{\partial \phi}{\partial t} + \frac{1}{\Delta} \end{pmatrix}. \tag{A.0.3}$$

For (A.0.2) to be a canonical transformation, $\underline{M}$ has to preserve the symplectic structure $\underline{M}^{\mathrm{T}}\underline{J}_6\underline{M} = \underline{J}_6$, where $\underline{J}_6 \equiv \underline{J}_2 \oplus \underline{J}_2 \oplus \underline{J}_2$ and

$$\underline{J}_2 \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{A.0.4}$$

After some calculation, it can be seen that

$$\underline{M}^{\mathrm{T}}\underline{J}_6\underline{M} = \begin{pmatrix} 0 & 1 & -e\left(\frac{\partial A_x}{\partial Q_y} - \frac{\partial A_y}{\partial Q_x}\right) & 0 & 0 & -e\left(\frac{\partial \phi}{\partial Q_x} + \frac{\partial A_x}{\partial t}\right) \\ -1 & 0 & 0 & 0 & 0 & 0 \\ e\left(\frac{\partial A_x}{\partial Q_y} - \frac{\partial A_y}{\partial Q_x}\right) & 0 & 0 & 1 & 0 & -e\left(\frac{\partial \phi}{\partial Q_y} + \frac{\partial A_y}{\partial t}\right) \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ e\left(\frac{\partial \phi}{\partial Q_x} + \frac{\partial A_x}{\partial t}\right) & 0 & e\left(\frac{\partial \phi}{\partial Q_y} + \frac{\partial A_y}{\partial t}\right) & 0 & -1 & 0 \end{pmatrix}. \tag{A.0.5}$$

Therefore, the transformation (A.0.2) is symplectic if the potentials fulfill the conditions

$$\frac{\partial A_x}{\partial Q_y} = \frac{\partial A_y}{\partial Q_x}, \quad \frac{\partial \phi}{\partial Q_x} = -\frac{\partial A_x}{\partial t}, \quad \frac{\partial \phi}{\partial Q_y} = -\frac{\partial A_y}{\partial t}. \tag{A.0.6}$$

Alternatively, this condition can also be written more concisely in the form of the curl of the vector field $(-\phi, A_x, A_y)$ with respect to the coordinates (t, Q_x, Q_y):

$$\begin{pmatrix} \frac{\partial}{\partial t} \\ \frac{\partial}{\partial Q_x} \\ \frac{\partial}{\partial Q_y} \end{pmatrix} \times \begin{pmatrix} -\phi \\ A_x \\ A_y \end{pmatrix} = 0. \tag{A.0.7}$$

In a next step, the coordinates $\hat{\boldsymbol{z}}$ are scaled via the improper scaling transformation

$$\underline{S} \equiv \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{P_0} \end{pmatrix} \oplus \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{P_0} \end{pmatrix} \oplus \begin{pmatrix} 0 & -c \\ \frac{1}{cP_0} & 0 \end{pmatrix}, \tag{A.0.8}$$

which yields another set of intermediate coordinates

$$\bar{\boldsymbol{z}} \equiv \begin{pmatrix} \bar{Q}_x \\ \bar{P}_x \\ \bar{Q}_y \\ \bar{P}_y \\ \bar{\tau} \\ \bar{\delta} \end{pmatrix} \equiv \underline{S} \begin{pmatrix} \hat{Q}_x \\ \hat{P}_x \\ \hat{Q}_y \\ \hat{P}_y \\ \hat{H} \\ \hat{t} \end{pmatrix}. \tag{A.0.9}$$

We note that $\underline{S}$ is symplectic with multiplier P_0^{-1} in the sense that

$$\underline{S}^{\mathrm{T}}\underline{J}_6\underline{S} = P_0^{-1}\underline{J}_6. \tag{A.0.10}$$

Hence, also the Hamiltonian has to be scaled with a factor P_0^{-1} when applying the transformation $\underline{S}$

$$\bar{K} = \frac{K}{P_0}. \tag{A.0.11}$$

To arrive at the final coordinates A.0.1, we subtract unity from $\bar{\delta}$ and add the time coordinate $ct_0(s)$ of the reference particle to $\bar{\tau}$

$$\tilde{\boldsymbol{z}} \equiv \begin{pmatrix} \tilde{x} \\ \tilde{p}_x \\ \tilde{y} \\ \tilde{p}_y \\ \tilde{\tau} \\ \tilde{\delta} \end{pmatrix} \equiv \begin{pmatrix} \bar{Q}_x \\ \bar{P}_x \\ \bar{Q}_y \\ \bar{P}_y \\ \bar{\tau} + ct_0(s) \\ \bar{\delta} - 1 \end{pmatrix}. \tag{A.0.12}$$

This shows that the coordinates $\boldsymbol{z}$ defined in (2.6.25) are connected to the coordinates $\tilde{\boldsymbol{z}}$, defined in (A.0.2), via a symplectic transformation, if the condition (A.0.6) holds. In these new coordinates the Hamiltonian reads

$$\tilde{K} = (\tilde{x}\kappa - 1)\sqrt{\left(\tilde{\delta} + 1\right)^2 - \tilde{p}_x^2 - \tilde{p}_y^2} - \frac{eA_s}{P_0}. \tag{A.0.13}$$

B SelaV$_{1D}$ Input Files for Discrete-Time Long-Bunch Model

In the following, input files for SelaV$_{1D}$ are presented which implement the discrete-time long-bunch model as shown in section 6.3. They also serve as an example for the syntax used by SelaV$_{1D}$'s parser. In this example the parameters to be used in the simulation are defined in the file `example-a.inp`. At the end of the parameter file, the actual simulation is started by including the file `DTLB.inp`, which implements the model and loops over the perturbation wavelengths.

example-a.inp

```
name="example-a";

// Beam Parameters
I0   = 25;     // Initial Current [A]
SIGE = 5e3;    // Inital Energy Spread [eV]

// First Linac
E1   = 100e6;  // Characteristic Energy [eV]
L1   = 25;     // Length [m]
BS1  = 0.3e-3; // Characteristic Beamsize [m]

// First BC
C1   = 4.00;   // Compression Factor [1]
R561 = 140e-3; // Chicane Strength [m]
EBC1 = 145e6;  // Energy [eV]

// Second Linac
E2   = 350e6;  // Characteristic Energy [eV]
L2   = 50;     // Length [m]
BS2  = 0.2e-3; // Characteristic Beamsize [m]

// Second BC
C2   = 8;      // Compression Factor [1]
R562 = 70e-3;  // Chicane Strength [m]
EBC2 = 550e6;  // Energy [eV]


// start simulation
include("DTLB.inp");
```

DTLB.inp

```
// Make base directory
mkdir("output");  chdir("output");  mkdir(name);  chdir(name);  basedir=getcwd();

NEXP:=4;  NSTEPS:=64;  LAMBDAMIN:=1e-6;  LAMBDAMAX:=10e-4;
LAMBDAS:=linspace(LAMBDAMIN,LAMBDAMAX,NSTEPS);
A0 := 1e-3;
SAVEGRIDS:=0;
```

```
i=0;
for LAMBDA in LAMBDAS {
        // Set up directory
        chdir(basedir);
        dir=strcat(basedir,format("/%04g/",i));
        mkdir(dir); chdir(dir);
        i=i+1;

        QSIM = I0 * LAMBDA/2.9979246e+08;

        limits=[-LAMBDA/2,-SIGE*8,LAMBDA/2,SIGE*8];
        psi = psd_bump(width=2*LAMBDA, sig_p=SIGE, weight=QSIM,
                limits=limits, dropoff=0.1*LAMBDA, topology=1, nexp=NEXP);

        modify(psi,format("psi*(1+%e*cos(q*2*pi/%e))",A0,LAMBDA));
        normalize(psi);
        write_localmoments(psi,"INI-localmoments.dat",1);

        // LSC1
        LSC1 = map_spacecharge(psi,
                beamsize=BS1, energy=E1,length=L1,npad=1,file="LSC1-field.dat");

        // BC1
        BC1 = map_compose(
                map_driftl(C1*R561/EBC1),
                map_hyperbolic(1./C1));

        STAGE1 = map_compose(LSC1,BC1);

        limits=[limits[0]/C1,limits[1]*C1,limits[2]/C1,limits[3]*C1];
        psi = propagate(psi,STAGE1,limits=limits);
        write_localmoments(psi,"DBC1-localmoments.dat",1);
        pvar_DBC1 = variance(psi)[3];
        if(SAVEGRIDS) {
                write_grid(psi,format("DBC1-grid-%e.dat",LAMBDA),npts=[256,256]);
        };

        // LSC2
        LSC2 = map_spacecharge(psi,
                beamsize=BS2, energy=E2, length=L2, npad=1, file="LSC2-field.dat");

        // BC2
        BC2 = map_compose(
                map_driftl(C2*R562/EBC2),
                map_hyperbolic(1./C2));

        STAGE2 = map_compose(LSC2,BC2);

        limits=[limits[0]/C2,limits[1]*C2,limits[2]/C2,limits[3]*C2];
        psi = propagate(psi,STAGE2,limits=limits);
        write_localmoments(psi,"DBC2-localmoments.dat",1);
        pvar_DBC2 = variance(psi)[3];

        { who(); } > "vars.dat";
};
```

C Phase-Space Linearization Script

The following is a Python implementation of the phase-space linearization approach described in Section 3.6.

rfcalc.py

```
import numpy as np

# functions for chicane dispersion coefficients
def F(m,a,b,lamb):
   if(m==0):
      return (lamb**2-1)**(-a) * lamb**b
   else:
      return -2*a*F(m-1, a+1, b+1, lamb) + b*F(m-1, a, b-1, lamb)

def dsB(lB, angle, n):
   lamb = 1./sin(angle)
   if(n==0):
      return lB * lamb * asin(1./lamb)
   elif(n==1):
      return lB * lamb * ( asin(1./lamb) - 1./sqrt(lamb**2-1))
   else:
      return lB * lamb**n * F(n-2, 3./2., -1., lamb)

def dsD(lD, angle, n):
   lamb = 1./sin(angle)
   return lD * lamb**n * F(n, 1./2., 1, lamb)

def chicanecoeff(lB,lD,angle,n):
   return 4*dsB(lB, angle, n) + 2*dsD(lD, angle, n)


# phase-space linearization functions
def solve_rf(b, k):
   E, h, _, _ = b
   z = E - 1j * h/k
   return z

def solve_rf_doublett(b, k1, k2):
   def coefs(k1,k2):
      return np.array([ (k2**2), -1j*k2**2/k1, 1, -1j/k1 ])/(k2**2-k1**2)
   z1 = b @ coefs(k1,k2)
   z2 = b @ coefs(k2,k1)
   return np.array([z1,z2])

def RF(z, k):
   zbar = z.conjugate()
   dE   = z + zbar
   dh   =  1j*k**1 * (z - zbar)
   dhp  =    -k**2 * (z + zbar)
   dhpp = -1j*k**3 * (z - zbar)
   return np.real(np.array([dE,dh,dhp,dhpp])/2.)
```

```
def BC(b0, R56, R566, R5666):
   f, fp, fpp, fppp = b0

   bp   = R56/f;  bpp  = R566/f**2;  bppp = R5666/f**3

   C   = 1./(1.+fp*bp);  Cp  = -C**3 * (fpp*bp + fp**2*bpp)
   Cpp = 3*Cp**2/C - C**4 * (fppp*bp + 3*fp*fpp*bpp + fp**3*bppp)

   g    = f;  gp   = C*fp;  gpp  = Cp*fp + C**2*fpp
   gppp = Cpp*fp + 3*C*Cp*fpp + C**3*fppp
   return np.array([g, gp, gpp, gppp])

def rfcalc( stage1, stage2, freqh, b0, nlf=(0.,0.)):
   # unpack compression stage (cavity & chicane)parameters
   E1, freq1, C1, lb1, ld1, angle1 = stage1
   E2, freq2, C2, lb2, ld2, angle2 = stage2

   hpf, hppf = nlf # unpack final non-linearities

   # constants
   c = 299792458;  kh =2*np.pi*freqh/c;  k1 = 2*np.pi*freq1/c;  k2 = 2*np.pi*freq2/c

   # calculate chicane dispersion coefficients
   R561, R5661, R56661 = \
      [-chicanecoeff(lb1, ld1, angle1, n) for n in [1,2,3]]
   R562, R5662, R56662 = \
      [-chicanecoeff(lb2, ld2, angle2, n) for n in [1,2,3]]

   # determine required chirps before chicances
   h1 = (1./C1-1)/R561 * E1;  h2 = (1./C2-1)/R562 * E2

   # propagate final beam coefs back through BC2
   b2 = [E2, C2*h2, hpf, hppf]
   b2hat = BC(b2, -R562, -R5662, -R56662)

   # complex amplitude of second cavity
   b1 = np.array([E1, h1*C1, 0, 0])
   z2 = solve_rf(b2hat-b1, k2)
   b1 = b2hat - RF(z2, k2)

   # propagate b1 back through first chicane
   b1hat = BC(b1, -R561, -R5661, -R56661)

   # solve for settings of first and harmonic cavity
   z1, zh = solve_rf_doublett(b1hat-np.array(b0), k1, kh)
   return (zh, z1, z2)

if __name__ == '__main__':
   # compression stage parameters
   #          E[eV] freq[Hz]  C   lb[m]   ld[m]    angle[rad]
   stage1 = ( 146e6, 1.3e9, 4.0, 0.5000, 0.5000, 16*np.pi/180. )
   stage2 = ( 550e6, 1.3e9, 4.0, 0.4463, 4.4300,  5*np.pi/180. )
   freqh = 3.9e9

   b0 = (5.6e6, 1.46283e+07, -1.76782e+09, 5.03274e+11) # initial curvature
   nlf= (-85.4186*1e12, -843.043*1e15)                  # final non-linearities

   sol = rfcalc( stage1, stage2, freqh, b0, nlf )
   for z in sol:
      print(f'{np.abs(z)/1e6:8.3f} {np.angle(z)/np.pi*180:7.2f}')
```

Bibliography

[1] P. Schmüser, M. Dohlus, J. Rossbach, and C. Behrens. *Free-Electron Lasers in the Ultraviolet and X-Ray Regime: Physical Principles, Experimental Results, Technical Realization*. Springer Tracts in Modern Physics. Springer International Publishing, 2014.

[2] W. Decking et al. A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator. *Nature Photonics*, 14(6):391–397, Jun 2020. doi:10.1038/s41566-020-0607-z.

[3] M. Krasilnikov et al. THz SASE FEL at PITZ: lasing at a wavelength of 100µm. In *Proc. IPAC'23*, number 14 in IPAC'23 - 14th International Particle Accelerator Conference, pages 3855–3859. JACoW Publishing, Geneva, Switzerland, 05 2023. doi:10.18429/JACoW-IPAC2023-THZG2.

[4] John M. J. Madey. Stimulated Emission of Bremsstrahlung in a Periodic Magnetic Field. *Journal of Applied Physics*, 42(5):1906–1913, 12 1971. doi:10.1063/1.1660466.

[5] W. Ackermann et al. Operation of a free-electron laser from the extreme ultraviolet to the water window. *Nature Photonics*, 1(6):336–342, Jun 2007. doi:10.1038/nphoton.2007.76.

[6] Brian McNeil. First light from hard x-ray laser. *Nature Photonics*, 3(7):375–377, Jul 2009. doi:10.1038/nphoton.2009.110.

[7] Zhirong Huang and Ingolf Lindau. Sacla hard-x-ray compact fel. *Nature Photonics*, 6(8):505–506, Aug 2012. doi:10.1038/nphoton.2012.184.

[8] Heung-Sik Kang et al. Hard x-ray free-electron laser with femtosecond-scale timing jitter. *Nature Photonics*, 11(11):708–713, Nov 2017. doi:10.1038/s41566-017-0029-8.

[9] Eduard Prat et al. A compact and cost-effective hard x-ray free-electron laser driven by a high-brightness and low-energy electron beam. *Nature Photonics*, 14(12):748–754, Dec 2020. doi:10.1038/s41566-020-00712-8.

[10] Henry N. Chapman. X-ray free-electron lasers for the structure and dynamics of macromolecules. *Annual Review of Biochemistry*, 88(1):35–58, 2019. doi:10.1146/annurev-biochem-013118-110744. PMID: 30601681.

[11] B. Pfau et al. Ultrafast optical demagnetization manipulates nanoscale spin structure in domain walls. *Nature Communications*, 3(1):1100, Oct 2012. doi:10.1038/ncomms2108.

[12] Artem Rudenko and Daniel Rolles. Time-resolved studies with fels. *Journal of Electron Spectroscopy and Related Phenomena*, 204:228–236, 2015. doi:10.1016/j.elspec.2015.07.010. Gas phase spectroscopic and dynamical studies at Free-Electron Lasers and other short wavelength sources.

[13] W. Roseker et al. Towards ultrafast dynamics with split-pulse x-ray photon correlation spectroscopy at free electron laser sources. *Nature Communications*, 9(1):1704, Apr 2018. doi:10.1038/s41467-018-04178-9.

[14] Alexander Wu Chao, Karl Hubert Mess, Maury Tigner, and Frank Zimmermann. *Handbook of Accelerator Physics and Engineering*. WORLD SCIENTIFIC, 2nd edition, 2013. doi:10.1142/8543.

[15] L.V. Iogansen and M.S. Rabinovich. Coherent Electron Radiation in a Synchrotron. I. *JETP*, 8(4):708, 1959.

[16] L.V. Iogansen and M.S. Rabinovich. Coherent Electron Radiation in the Synchrotron. II. *JETP*, 10(1):83, 1960.

[17] M.S. Rabinovich and L.V. Iogansen. Coherent Electron Radiation in the Synchrotron. III. *JETP*, 11(4):856, 1960.

[18] E.L. Saldin, E.A. Schneidmiller, and M.V. Yurkov. An analytical description of longitudinal phase space distortions in magnetic bunch compressors. *Nucl. Instrum. Methods Phys. Res. A*, 483(1):516–520, 2002. doi:10.1016/S0168-9002(02)00372-8. Proceedings of the 23rd International Free Electron Laser Confere nce and 8th FEL Users Workshop.

[19] E.L. Saldin, E.A. Schneidmiller, and M.V. Yurkov. Klystron instability of a relativistic electron beam in a bunch compressor. *Nucl. Instrum. Methods Phys. Res. A*, 490(1):1–8, 2002. doi:10.1016/S0168-9002(02)00905-1.

[20] T Limberg, Ph Piot, and E.A Schneidmiller. An analysis of longitudinal phase space fragmentation at the tesla test facility. *Nucl. Instrum. Methods Phys. Res. A*, 475(1):353–356, 2001. doi:10.1016/S0168-9002(01)01538-8. FEL2000: Proc. 22nd Int. Free Electron Laser Conference and 7th F EL Users Workshop.

[21] G. Stupakov. Using the beam-echo effect for generation of short-wavelength radiation. *Phys. Rev. Lett.*, 102:074801, Feb 2009. doi:10.1103/PhysRevLett.102.074801.

[22] Christoph Lechner et al. Suppression of FEL Lasing by a Seeded Microbunching Instability. In *37th International Free Electron Laser Conference*, page TUA02, 2015. doi:10.18429/JACoW-FEL2015-TUA02.

[23] E. Hemsing, B. Garcia, Z. Huang, T. Raubenheimer, and D. Xiang. Sensitivity of echo enabled harmonic generation to sinusoidal electron beam energy structure. *Phys. Rev. Accel. Beams*, 20:060702, Jun 2017. doi:10.1103/PhysRevAccelBeams.20.060702.

[24] Primož Rebernik Ribič, Eléonore Roussel, Gregory Penn, Giovanni De Ninno, Luca Giannessi, Giuseppe Penco, and Enrico Allaria. Echo-enabled harmonic generation studies for the fermi free-electron laser. *Photonics*, 4(1), 2017. doi:10.3390/photonics4010019.

[25] Eric Sonnendrücker, Jean Roche, Pierre Bertrand, and Alain Ghizzo. The Semi-Lagrangian Method for the Numerical Resolution of Vlasov Equations. Research Report RR-3393, INRIA, 1998.

[26] Robert L. Warnock and James A. Ellison. A General method for propagation of the phase space distribution, with application to the sawtooth instability. In *2nd ICFA Advanced Accelerator Workshop on the Physics of High Brightness Beams*, pages 322–348, 3 2000.

[27] N. Besse and E. Sonnendrücker. Semi-Lagrangian schemes for the Vlasov equation on an unstructured mesh of phase space. *Journal of Computational Physics*, 191(2):341–376, 2003. doi:10.1016/S0021-9991(03)00318-8.

[28] Eric Sonnendrücker. Numerical methods for the Vlasov equations. Lecture notes. Archived at `https://web.archive.org/web/20180727045746id_/http://www-m16.ma.tum.de/foswiki/pub/M16/Allgemeines/NumMethVlasov/Num-Meth-Vlasov-Notes.pdf`, 2013.

[29] R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, Mar 1974. doi:10.1007/BF00288933.

[30] Martin Beye and Stephan Klumpp, editors. *FLASH2020+ - upgrade of FLASH: conceptual design report*. Verlag Deutsches Elektronen-Synchrotron, Hamburg, 2020. doi:10.3204/PUBDB-2020-00465.

[31] Herbert Goldstein, John L. Safko, and Charles P. Poole. *Classical mechanics; Third edition, Pearson new international edition*. Pearson, Harlow, 2014.

[32] Vladimir Igorevich Arnold. *Mathematical methods of classical mechanics; 2nd edition*, volume 60 of *Graduate texts in mathematics*. Springer, New York, 1989.

[33] Alex J. Dragt. Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics. `https://www.physics.umd.edu/dsat/dsatliemethods.html`.

[34] Franz Schwabl. *Statistical mechanics; 2nd edition*. Springer, Berlin, 2006.

[35] Richard Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.

[36] J. Liouville. Note sur la Théorie de la Variation des constantes arbitraires. *Journal de Mathématiques Pures et Appliquées*, 1e série, 3, 1838.

[37] E. M. Lifshitz and L. P. Pitaevskii. *Physical kinetics*, volume 10 of *Course of theoretical physics*. Pergamon, Oxford, 1981.

[38] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 2010.

[39] Fern Y Hunt. Unique ergodicity and the approximation of attractors and their invariant measures using Ulam's method . *Nonlinearity*, 11:307, 1998.

[40] David Ruelle. Dynamical Zeta Functions and Transfer Operators. *Notices of the AMS*, 49(8):887–895, 2002.

[41] Philipp Amstutz. Comparison of Eulerian, Lagrangian and Semi-Lagrangian Simulations of Phase-Space Density Evolution. In *Proc. FEL'22*. JACoW Publishing, Geneva, Switzerland, Aug 2022. in preparation by the editors.

[42] Max Born and H. S. Green. A general kinetic theory of liquids i. the molecular distribution functions. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 188(1012):10–18, 1946. doi:10.1098/rspa.1946.0093.

[43] John G. Kirkwood. The Statistical Mechanical Theory of Transport Processes I. General Theory. *The Journal of Chemical Physics*, 14(3):180–201, 12 1946. doi:10.1063/1.1724117.

[44] John G. Kirkwood. The Statistical Mechanical Theory of Transport Processes II. Transport in Gases. *The Journal of Chemical Physics*, 15(1):72–76, 11 1946. doi:10.1063/1.1746292.

[45] Carlo Cercignani. *The Boltzmann equation and its applications*, volume 67 of *Applied mathematical sciences*. Springer, New York, 1988.

[46] A A Vlasov. The vibrational properties of an electron gas. *Soviet Physics Uspekhi*, 10(6):721, jun 1968. doi:10.1070/PU1968v010n06ABEH003709.

[47] Philipp Amstutz and Mathias Vogt. Investigation of Vlasov Systems with a Certain Class of Linearly-Collective Hamiltonians. In *Proc. IPAC'21*, pages 3157–3160. JACoW Publishing, Geneva, Switzerland, Aug 2021. doi:10.18429/JACoW-IPAC2021-WEPAB226.

[48] A. Y. Aydemir. A unified Monte Carlo interpretation of particle simulations and applications to non-neutral plasmas . *Physics of Plasmas*, 1(4):822–831, 04 1994. doi:10.1063/1.870740.

[49] M. Vogt, T. Sen, and J. A. Ellison. Simulations of three one-dimensional limits of the strong-strong beam-beam interaction in hadron colliders using weighted macroparticle tracking. *Phys. Rev. ST Accel. Beams*, 5:024401, Feb 2002. doi:10.1103/PhysRevSTAB.5.024401.

[50] Sören Bartels. *Numerical approximation of partial differential equations*, volume 64 of *Texts in applied mathematics*. Springer, Cham, 2016. doi:10.1007/978-3-319-32354-1.

[51] Marco Venturini, Robert Warnock, and Alexander Zholents. Vlasov solver for longitudinal dynamics in beam delivery systems for x-ray free electron lasers. *Phys. Rev. ST Accel. Beams*, 10:054403, May 2007. doi:10.1103/PhysRevSTAB.10.054403.

[52] Patrik Schönfeldt, Miriam Brosi, Markus Schwarz, Johannes L. Steinmann, and Anke-Susanne Müller. Parallelized Vlasov-Fokker-Planck solver for desktop personal computers. *Phys. Rev. Accel. Beams*, 20:030704, Mar 2017. doi:10.1103/PhysRevAccelBeams.20.030704.

[53] Philipp Amstutz. Selav_{1D}. `https://selav.desy.de/`. Accessed June 2023. Archived at `https://web.archive.org/web/20221223134722/https://selav.desy.de/`.

[54] Helmut Wiedemann. *Particle accelerator physics; Fourth edition*. Graduate texts in physics. Springer, Cham, 2015. doi:10.1007/978-3-319-18317-6.

[55] Andrzej Wolski. *Beam Dynamics in High Energy Particle Accelerators*. IMPERIAL COLLEGE PRESS, 2014. doi:10.1142/p899.

[56] D. A. Edwards and M. J. Syphers. An introduction to the physics of particle accelerators. *AIP Conference Proceedings*, 184(1):2–189, 04 1989. doi:10.1063/1.38066.

[57] Wolfgang Kühnel. *Differential geometry: Curves - surfaces - manifolds; 3rd ed.*, volume 77 of *Student mathematical library*. AMS, Providence, 2015.

[58] Vladimir N. Litvinenko, Yichao Jing, and Gang Wang. Advanced Accelerator Physics Lecture 5. Archived at `https://web.archive.org/web/20220629094237/http://case.physics.stonybrook.edu/images/5/50/PHY564_Lecture_5.pdf`, 2015.

[59] Klaus Wille. *The physics of particle accelerators: an introduction*. Oxford University Press, Oxford, 2000.

[60] H. Grote and F. Christoph Iselin. The MAD Program (Methodical Accelerator Design): Version 8.10 ; User's Reference Manual. Technical Report CERN-SL-90-13-AP-rev-3, Geneva, 1993.

[61] Georg Heinz Hoffstätter. *Rigorous bounds on survival times in circular accelerators and efficient computation of fringe field transfer maps*. Dr., Michigan State University, 1994. Ph.D. Thesis; Michigan State University, Diss., 1994.

[62] M. Berz and K. Makino. COSY INFINITY Version 8.1 User's Guide and Reference Manual. 2001.

[63] Michael Borland. `elegant`. `https://ops.aps.anl.gov/elegant.html`. Accessed June 2023. Archived at `https://web.archive.org/web/20220929134213/https://ops.aps.anl.gov/elegant.html`.

[64] F. Christoph Iselin. The MAD Program (Methodical Accelerator Design) Version 8.13: Physical Methods Manual. Technical Report CERN-SL-92-?? (AP), CERN, Sep 1994.

[65] S. C. Hartman and J. B. Rosenzweig. Ponderomotive focusing in axisymmetric rf linacs. *Phys. Rev. E*, 47:2031–2037, Mar 1993. doi:10.1103/PhysRevE.47.2031.

[66] J. Rosenzweig and L. Serafini. Transverse particle motion in radio-frequency linear accelerators. *Phys. Rev. E*, 49:1599–1602, Feb 1994. doi:10.1103/PhysRevE.49.1599.

[67] P. Castro. Beam trajectory calculations in bunch compressors of TTF2. Technical Report DESY-TECHNICAL-NOTE-2003-01, 4 2003.

[68] Igor Zagorodnov and Martin Dohlus. Semianalytical modeling of multistage bunch compression with collective effects. *Phys. Rev. ST Accel. Beams*, 14:014403, Jan 2011. doi:10.1103/PhysRevSTAB.14.014403.

[69] Andy Wolski. A Short Introduction to Bunch Compressors for Linear Colliders. Lecture notes. Archived at `https://web.archive.org/web/20220121022804/http://www.desy.de/~njwalker/uspas/coursemat/notes/unit_5_notes.pdf`, 2003.

[70] Marco Venturini. Models of longitudinal space-charge impedance for microbunching instability. *Phys. Rev. ST Accel. Beams*, 11:034401, Mar 2008. doi:10.1103/PhysRevSTAB.11.034401.

[71] Philipp Amstutz. Vlasov Simulation of Exotic Phase-Space Densities via Tree-Based Domain-Decomposition. Masterarbeit, Universität Hamburg, Hamburg, 2018.

[72] John David Jackson. *Classical electrodynamics; Third edition.* John Wiley & Sons, Inc., New York, 1999.

[73] E. Wiechert. Elektrodynamische Elementargesetze. *Annalen der Physik*, 309(4):667–689, 1901. doi:10.1002/andp.19013090403.

[74] J.B. Murphy, S. Krinksy, and Gluckstern R.L. Longitudinal wakefield for synchrotron radiation. In *Proc. Part. Accel. Conf. 1995*, May 1995.

[75] E.L. Saldin, E.A. Schneidmiller, and M.V. Yurkov. On the coherent radiation of an electron bunch moving in an arc of a circle. *Nucl. Instrum. Methods Phys. Res. A*, 398(2):373–394, 1997. doi:10.1016/S0168-9002(97)00822-X.

[76] E.L. Saldin, E.A. Schneidmiller, and M.V. Yurkov. Radiative interaction of electrons in a bunch moving in an undulator. *Nucl. Instrum. Methods Phys. Res. A*, 417(1):158–168, 1998. doi:10.1016/S0168-9002(98)00623-8.

[77] Walter Rudin. *Real and complex analysis.* McGraw-Hill series in higher mathematics. McGraw-Hill, New York, 1966.

[78] M. Borland. Simple method for particle tracking with coherent synchrotron radiation. *Phys. Rev. ST Accel. Beams*, 4:070701, Jul 2001. doi:10.1103/PhysRevSTAB.4.070701.

[79] Philipp Amstutz and Mathias Vogt. Advances in a Perturbation Theory for the Microbunching Instability in Free-Electron Laser Injectors. In *Proc. IPAC'23*, pages 3546–3549. JACoW Publishing, Geneva, Switzerland, May 2023. doi:10.18429/JACoW-IPAC2023-WEPL183.

[80] J. Dieudonné. *Foundations of modern analysis; 6th print*, volume 10 of *Pure and applied mathematics.* Academic, New York, 1960.

[81] Wilfred Kaplan. *Advanced calculus; 4th ed.* Addison-Wesley, Redwood City, 1991.

[82] J.F. Federici. Review of four-wave mixing and phase conjugation in plasmas. *IEEE Transactions on Plasma Science*, 19(4):549–564, 1991. doi:10.1109/27.90319.

[83] H Bindslev. Three-wave mixing and thomson scattering in plasmas. *Plasma Physics and Controlled Fusion*, 35(11):1615, nov 1993. doi:10.1088/0741-3335/35/11/009.

[84] Marco Venturini. Shot-noise seeded microbunching instability: Second-order correction to the gain function. *Nucl. Instrum. Methods Phys. Res. A*, 599(2):140–145, 2009. doi:10.1016/j.nima.2008.11.004.

[85] J. Qiang, Y. Ding, P. Emma, Z. Huang, D. Ratner, T. O. Raubenheimer, M. Venturini, and F. Zhou. Start-to-end simulation of the shot-noise driven microbunching instability experiment at the linac coherent light source. *Phys. Rev. Accel. Beams*, 20:054402, May 2017. doi:10.1103/PhysRevAccelBeams.20.054402.

[86] Ronald Gould. *Graph theory.* Benjamin/Cummings, Menlo Park, 1988.

[87] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition, 2007.

[88] Kendall E. Atkinson. *An Introduction to Numerical Analysis*. Wiley, New York, 1978.

[89] C.Z Cheng and Georg Knorr. The integration of the vlasov equation in configuration space. *Journal of Computational Physics*, 22(3):330–351, 1976. doi:10.1016/0021-9991(76)90053-X.

[90] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of mathematical functions with formulas, graphs, and mathematical tables; 9th print*. Dover books on advanced mathematics. Dover, New York, 1970.

[91] Bengt Fornberg. Generation of finite difference formulas on arbitrarily spaced grids. *Math. Comp.*, 51:699–706, 1988.

[92] Murray Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956. doi:10.1214/aoms/1177728190.

[93] Emanuel Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962. doi:10.1214/aoms/1177704472.

[94] Igor Zagorodnov. Start-to-End Simulations, FLASH I, 500 pC. `https://www.desy.de/fel-beam/s2e/flash/Nominal/flashI/nom500pC.html`. Accessed June 2023. Archived at `https://web.archive.org/web/20220121045807/https://www.desy.de/fel-beam/s2e/flash/Nominal/flashI/nom500pC.html`.

[95] Marco Venturini. Microbunching instability in single-pass systems using a direct two-dimensional vlasov solver. *Phys. Rev. ST Accel. Beams*, 10:104401, Oct 2007. doi:10.1103/PhysRevSTAB.10.104401.

[96] M. Venturini and A. Zholents. Modeling microbunching from shot noise using vlasov solvers. *Nucl. Instrum. Methods Phys. Res. A*, 593(1):53–56, 2008. doi:10.1016/j.nima.2008.04.036.

[97] Cheng-Ying Tsai. Concatenated analyses of phase space microbunching in high brightness electron beam transport. *Nucl. Instrum. Methods Phys. Res. A*, 940:462–474, 2019. doi:10.1016/j.nima.2019.06.061.

[98] Lucas Schaper, Sven Ackermann, Enrico Allaria, Philipp Amstutz, Karolin Baev, Martin Beye, Christopher Gerth, Ingmar Hartl, Wolfgang Hillert, Katja Honkavaara, Mehdi Mohammad Kazemi, Tino Lang, Pardis Niknejadi, Fabian Pannek, Juliane Rönsch-Schulenburg, Dmitrii Samoilenko, Evgeny Schneidmiller, Siegfried Schreiber, Markus Tischer, Mathias Vogt, Mikhail Yurkov, and Johann Zemella. Flexible and Coherent Soft X-ray Pulses at High Repetition Rate: Current Research and Perspectives. *Applied Sciences*, 11(20), 2021. doi:10.3390/app11209729.

[99] Philipp Amstutz and Mathias Vogt. Microbunching Studies for the FLASH2020+ Upgrade Using a Semi-Lagrangian Vlasov Solver. In *Proc. IPAC'22*, pages 2334–2337. JACoW Publishing, Geneva, Switzerland, Jul 2022. doi:10.18429/JACoW-IPAC2022-WEPOMS037.

[100] M. Borland. Elegant: A flexible sdds-compliant code for accelerator simulation. Technical Report LS-287 TRN: US0004540, 8 2000.

[101] C. Gerth, E. Allaria, A. Choudhuri, L. Schaper, E. Schneidmiller, S. Schreiber, M. Tischer, P. Vagin, M. Vogt, L. Winkelmann, M.V. Yurkov, and J. Zemella. Layout of the Laser Heater for FLASH2020+. In *Proc. IPAC'21*, number 12 in International Particle Accelerator Conference, pages 1647–1650. JACoW Publishing, Geneva, Switzerland, 08 2021. doi:10.18429/JACoW-IPAC2021-TUPAB111.

[102] Siegfried Schreiber. private communication.

[103] D. Samoilenko, C. Gerth, P. Amstutz, C. Mahnke, P. Niknejadi, L. Schaper, E. Schneidmiller, A. Schoeps, S. Schreiber, S. Telawane, L. Winkelmann, M. V. Yurkov, J. Zemella, W. Hillert, and C. Mai. First Beam Heating Studies with the Laser Heater for FLASH2020+. In *Proc. IPAC'23*, number 14, pages 1910–1914. JACoW Publishing, Geneva, Switzerland, 05 2023. doi:10.18429/JACoW-IPAC2023-TUPL098.

[104] B. Schmidt, C. Behrens, H. Delsim-Hashemi, P. Schmüser, and S. Wesch. Coherent micro-bunching radiation from electron bunches at flash in the 10 micrometer wavelength range. In *Proc. FEL'08*.

[105] S. Bettoni, M. Csatari Divall, R. Ganter, M. Pedrozzi, E. Prat, S. Reiche, T. Schietinger, A. Trisorio, C. Vicario, and V. Goryashko. Impact of laser stacking and photocathode materials on microbunching instability in photoinjectors. *Phys. Rev. Accel. Beams*, 23:024401, Feb 2020. doi:10.1103/PhysRevAccelBeams.23.024401.

[106] W. Schottky. Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. *Annalen der Physik*, 362(23):541–567, 1918. doi:10.1002/andp.19183622304.

[107] P. Niknejadi. in preparation.

[108] P. Niknejadi et al. Improving the realistic modeling of the eehg seed section in start to end simulations. In *Proc. FEL'22*, 2022. in preparation by the editors.

[109] Ji Qiang. Impact: Integrated map and particle accelerator tracking code. `https://amac.lbl.gov/~jiqiang/IMPACT/`. Accessed June 2023.

[110] Martin Dohlus. private communication.

Acknowledgments

Dr. Mathias Vogt has supported me greatly during this PhD project by sharing his expert knowledge and providing reassuring guidance; I am deeply grateful for his efforts and kindness. Go raibh míle maith agat, mif.f !

Prof. Dr. Gudrid Moortgat-Pick kindly allowed me to join her research group and provided helpful advice at all times. Thank you!

I had the pleasure to work with a large number of colleagues, who helped me with enlightening discussions, interesting suggestions or providing opportunities to present my work. Thanks are due to Dr. Ilya Agapov, Dr. Martin Dohlus, Dr. Christopher Gerth, Dr. Katja Honkavaara, Carsten Mai, Dr. Pardis Niknejadi, Fabian Pannek, Dr. Juliane Rönsch-Schulenburg, Dmitrii Samoilenko, Dr. Lucas Schaper, Dr. Evgeny Schneidmiller, Dr. Siegfried Schreiber and Dr. Johann Zemella.

Special thanks are due to Dr. Immo Bahns, Dr. Florian Christie, Dr. Thorsten Hellert, Dr. Christian Henkel, Marius Hoffmann, Dr. Max Kellermeier, Dr. Willi Kuropka and Dr. Patrick Rauer, with whom I had the pleasure to share also the less-strictly scientific yet most delightful experiences that are an integral part of every healthy academic endeavor.

My mother Bettina Amstutz and my father Hubertus Amstutz with Carola Kunkel, supported me unconditionally as always. For that, I am truly thankful.

Anne Kirschner had patience and gave me confidence; for that and so much more, I admire her deeply.

Eidesstattliche Versicherung

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

Hamburg, den 14.07.2023

Philipp Amstutz