First Demonstration of Echo-Enabled Harmonic Generation in a Storage Ring and Deep Learning for Beam Diagnostics

Arjun Radha Krishnan

2026

Dissertation

zur Erlangung des Grades
Doktor der Naturwissenschaften

Fakultät Physik
der Technischen Universität Dortmund

technische universität
dortmund

Vorsitzender der Kommission: Prof. Dr. Mirko Cinchetti

Erstgutachter: Prof. Dr. Shaukat Khan

Zweitgutachter: Prof. Dr. Metin Tolan

weiteres Mitglied der Kommission: Prof. Dr. Christof Weitenberg

Datum der Prüfung: 20.03.2026

Abstract

The generation and characterization of ultrashort, coherent radiation pulses in the extreme ultraviolet (EUV) and X-ray regimes are critical for investigating ultrafast dynamic processes in matter. This thesis advances these capabilities through two complementary research directions: the development of deep-learning-based methods for advanced beam diagnostics and the first experimental implementation of echo-enabled harmonic generation (EEHG) in an electron storage ring.

To overcome the limitations of conventional diagnostics, deep-learning-based methods were developed for both storage ring and free-electron laser (FEL) applications. At the 1.5-GeV electron storage ring DELTA, operated by the Center for Synchrotron Radiation of TU Dortmund University, coherent harmonic generation (CHG) is employed, a technique where an external laser seed interacts with the electron beam to generate ultrashort radiation pulses at harmonics of the laser frequency. A convolutional neural network (CNN) was trained to extract the spectral phase properties of the seed laser pulses from the CHG radiation spectra. This method successfully identified a large uncompensated third-order dispersion in the seed pulse, explaining previously unresolved spectral asymmetries. Furthermore, two CNN-based methods were implemented to reconstruct photon pulse profiles from longitudinal electron phase space images at FLASH, an FEL in Hamburg. This work involved the application of a standard U-Net model and a specialized dual-input design (Y-Net) developed specifically for this task. The deep learning models demonstrate the capability to perform single-shot reconstruction of the photon pulse profiles without requiring frequent reference measurements and reduce non-physical artifacts.

In the experimental domain, this work presents the commissioning of the SPEED (Short-Pulse Emission via Echo at DELTA) facility at DELTA. This project utilizes the EEHG mechanism that employs a two-stage seeding scheme to manipulate electrons in the longitudinal phase space for harmonic generation. By reconfiguring a 4.75 m electromagnetic undulator, the world's first proof-of-principle demonstration of EEHG in a storage ring was achieved. Coherent emission was successfully observed and characterized up to the 11th harmonic (73 nm) of an 800 nm seed laser. These results confirm the robustness of the EEHG mechanism in a circular accelerator. Building on this success, a conceptual design for an optimized and compact permanent-magnet-based EEHG setup is presented. Simulation studies indicate that this design, fitting within a typical 5-meter straight section, can extend efficient harmonic generation up to the 29th harmonic ($\approx 27\,\mathrm{nm}$), showing a path forward for future short-pulse sources based on electron storage rings.

Kurzfassung

Die Erzeugung und Charakterisierung ultrakurzer, kohärenter Strahlungspulse im extremen Ultraviolett- (EUV) sowie im Röntgenbereich ist für die Untersuchung ultraschneller dynamischer Prozesse in Materie von entscheidender Bedeutung. In diesem Kontext beschreibt die vorliegende Arbeit zwei sich ergänzende Forschungsbeiträge: die Entwicklung von *Deep-Learning*-basierten Methoden für neue Formen der Strahldiagnose und die erste experimentelle Umsetzung des EEHG-Verfahrens (*Echo-Enabled Harmonic Generation*) in einem Elektronenspeicherring.

Um die Grenzen konventioneller Strahldiagnose zu überwinden, wurden *Deep-Learning*-basierte Verfahren sowohl für Anwendungen an Speicherringen als auch für Freie-Elektronen-Laser (FEL) entwickelt. Am 1,5-GeV-Elektronenspeicherring DELTA, der vom Zentrum für Synchrotronstrahlung der TU Dortmund betrieben wird, wurde das CHG-Verfahren (*Coherent Harmonic Generation*) eingesetzt, bei dem externe Laserpulse die Elektronen im Speicherring so manipulieren, dass ultrakurze Strahlungspulse bei Harmonischen der Laserfrequenz entstehen. Ein sog. *Convolutional Neural Network* (CNN) wurde trainiert, um Eigenschaften der spektralen Phase der Laserpulse aus den CHG-Strahlungsspektren zu extrahieren. Mit dieser Methode konnte eine hohe unkompensierte Dispersion dritter Ordnung erfolgreich identifiziert werden, wodurch zuvor unverstandene Asymmetrien in den Spektren erklärt werden konnten. Darüber hinaus wurden zwei CNN-basierte Methoden implementiert, um Photonenpulsprofile aus gemessenen Verteilungen der Elektronen im longitudinalen Phasenraum bei FLASH, einem FEL in Hamburg, zu rekonstruieren. Diese Arbeit umfasste die Anwendung eines Standard-*U-Net*-Modells und eines speziell für diese Aufgabe entwickelten *Y-Net*-Modells mit zwei Eingangsverteilungen. Diese *Deep-Learning*-Modelle ermöglichen eine Einzelpuls-Rekonstruktion der Photonenpulsform, ohne dass zahlreiche Referenzmessungen durchgeführt werden müssen. Zudem wird das Auftreten nicht-physikalischer Artefakte reduziert.

Im experimentellen Teil beschreibt diese Arbeit die Inbetriebnahme der Anlage SPEED (Short-Pulse Emission via Echo at DELTA) bei DELTA. Dieses Projekt nutzt den EEHG-Mechanismus, der mit einer zweistufigen Laser-Wechselwirkung die Elektronen im longitudinalen Phasenraum so manipuliert, dass Strahlung bei hohen Harmonischen erzeugt wird. Durch die Neukonfiguration eines 4,75 m langen elektromagnetischen Undulators konnte das EEHG-Verfahren weltweit erstmalig an einem Speicherring demonstriert werden. Kohärente Strahlungsemission wurde erfolgreich beobachtet und bis zur 11. Harmonischen (73 nm) der 800-nm-Laserpulse charakterisiert. Diese Ergebnisse bestätigen die Robustheit des EEHG-Mechanismus in einem Speicherring. Aufbauend auf diesen Erfolg wird das Konzept einer optimierten und kompakten EEHG-Anordnung auf der Basis von Permanentmagneten vorgestellt. Simulationsstudien belegen, dass die effiziente Erzeugung von Strahlung bis zur 29. Harmonischen (ca. 27 nm) in einer typischen freien Strecke von 5 Metern Länge möglich ist, womit ein Weg für zukünftige Kurzpulsquellen an Elektronenspeicherringen aufgezeigt wird.

To my father,
Who taught me to fix my gaze up at the stars,
and showed me the immense beauty of this universe.

Contents

Chapter 1

Introduction

The investigation of dynamic processes in matter requires pulsed short-wavelength radiation with a duration comparable to or shorter than the timescale of the observed phenomena. A vast range of critical physical and chemical processes, such as lattice vibrations, changes in electronic or magnetic structure, and chemical bond dynamics, occurs on the femtosecond (10^{-15} s) timescale [1]. Synchrotron radiation sources based on electron storage rings have long provided reliable, high-flux X-ray and extreme ultraviolet (EUV) radiation with excellent shot-to-shot stability. However, the radiation pulse duration from conventional storage rings is fundamentally determined by the electron bunch length. This is typically several tens of picoseconds (10^{-12} s) [2] which is orders of magnitude too long to resolve the dynamics in molecular and condensed-matter systems [1]. While specialized 'low-alpha' optics [3] can reduce the bunch length to several picoseconds, this often comes at the cost of reduced beam current and stability. Conventional optical lasers can easily generate femtosecond pulses, but their wavelengths are limited to the near-visible range, which lacks the short wavelengths required for atomic resolution. High harmonic generation (HHG) [4–6] sources can convert these lasers to EUV wavelengths with femtosecond duration, but with low photon flux [7]. Free-electron lasers (FELs) have overcome these limitations of pulse duration and photon flux, delivering femtosecond pulses in the X-ray regime [8] with peak brightness orders of magnitude higher than storage rings. However, there are currently only about ten short-wavelength FEL facilities worldwide compared to the approximately fifty storage ring facilities in operation [9]. Furthermore, storage rings serve as multi-user facilities with tens of beamlines functioning simultaneously and offer repetition rates in the hundreds of MHz. In contrast, while superconducting FELs have pushed repetition rates into the MHz regime, normal-conducting FELs are typically limited to repetition rates of roughly 100 Hz. Bringing femtosecond capabilities to storage rings would therefore allow a much broader community to access ultrafast EUV and X-ray science, leveraging the existing infrastructure and high stability of these sources.

To overcome the fundamental pulse duration limits of storage rings, several laser seeding techniques have been developed. The core principle relies on manipulating only a short slice of the electron bunch using a femtosecond laser pulse and extracting the radiation emitted from this region. The laser pulse imposes a sinusoidal energy modulation on the electrons within the slice. The technique called femtoslicing [10] spatially separates the energy-modulated electrons from the rest of the bunch using strong dispersive magnets. These displaced electrons then radiate incoherently in a subsequent undulator. While successful at generating femtosecond X-ray pulses, this method is limited by low photon flux. On the other hand, techniques such as coherent harmonic generation (CHG) [11] and echo-enabled harmonic generation (EEHG) [12, 13] use dispersive sections to achieve longitudinal density modulation. In these schemes, the laser-electron interaction occurs in a specialized undulator called a modulator, while a magnetic chicane converts the energy modulation into density modulation. This creates periodic microbunches that emit coherently at harmonics of the laser wavelength in a final undulator called the radiator. Because the resulting radiation intensity scales with the square of the number of contributing electrons (N^2), the coherent emission is significantly brighter than the incoherent background, which scales linearly with N, thereby eliminating the need for spatial separation.

The techniques CHG and EEHG are closely linked to seeding schemes in FELs. CHG is the storage ring equivalent of high-gain harmonic generation (HGHG) used in FELs. While HGHG utilizes long undulators to achieve exponential amplification, CHG operates without the exponential gain. EEHG was originally proposed to improve the harmonic reach of seeded FELs. CHG has been successfully implemented at several storage rings [14–16], including at DELTA, TU Dortmund [17, 18], where it has been operational since 2011. However, CHG efficiency decays exponentially with harmonic number, limiting it to low harmonics. EEHG reaches higher harmonics through a two-stage modulation process. While EEHG has been successfully demonstrated at several FEL facilities based on linear accelerators (linacs) [19–21], its implementation in an electron storage ring had never been attempted. Such an installation requires integrating a complex lattice comprising two modulators, two chicanes, and a radiator into the limited space of a standard storage ring straight section. In addition, storage rings have naturally larger beam emittance and energy spread compared to linac-based sources. A large beam energy spread reduces the relative strength of the laser-induced energy modulation, which lowers the efficiency of the EEHG microbunching process. Consequently, the viability of EEHG in a storage ring remained untested.

Beyond generating the ultrashort, short-wavelength pulses from CHG and EEHG, characterizing them is critical due to their complex spectro-temporal sensitivity to seed laser and electron beam dynamics [22]. A transverse deflecting structure (TDS) [23] is commonly used for this purpose, offering femtosecond longitudinal resolution by mapping the electron bunch longitudinal phase space distribution onto a screen. However, extracting the FEL

pulse profiles from the TDS measurements requires a reconstruction algorithm that typically compares the measured profiles of a lasing bunch to a non-lasing reference [24]. Because electron bunch properties fluctuate shot-to-shot, this reference comparison is often imperfect, introducing artifacts into the reconstructed temporal profile.

The rapidly advancing field of deep learning offers new approaches to the diagnostic challenges. The 2024 Nobel Prize in Physics, awarded to John Hopfield and Geoffrey Hinton for foundational work in artificial neural networks [25], highlights the transformative impact of machine learning across scientific disciplines. Deep learning architectures can learn complex, nonlinear relationships from data, making them well-suited for accelerator diagnostics. In light of these developments, this thesis addresses both the diagnostic and experimental challenges outlined above through two complementary research directions.

On the diagnostics side, deep learning frameworks to overcome fundamental limitations of conventional diagnostic techniques are developed. For CHG spectral analysis, a convolutional neural network (CNN) extracting the spectral phase properties of the seed laser directly from radiation spectra is discussed. For FEL pulse diagnostics, CNN-based architectures that reconstruct photon pulse profiles from longitudinal phase space images recorded using a TDS are presented. Unlike reference-subtraction methods, these networks avoid unphysical artifacts and can operate without requiring frequent recalibration measurements, making them suitable for real-time diagnostics.

On the experimental side, the commissioning of SPEED (Short-Pulse Emission via Echo at DELTA), the first-ever implementation of EEHG in an electron storage ring, is presented. This involved rewiring the U250 undulator, previously used for the CHG seeding scheme, to accommodate the two modulators, two chicanes, and a radiator within a single undulator. The project was conceived as a proof-of-principle experiment aimed to demonstrate the feasibility of the EEHG seeding scheme in a storage ring. Complementing the experimental work, a conceptual design study is presented for a dedicated, permanent-magnet-based EEHG setup. This study investigates the feasibility of integrating a compact EEHG source into the standard 5-meter straight sections available at many synchrotron facilities.

This thesis is structured as follows:

- Chapter 2 establishes the foundational principles of electron beam dynamics, synchrotron radiation, ultrashort laser physics, and artificial neural networks.
- Chapter 3 presents the CNN-based method for analyzing CHG spectra and extracting seed laser phase information.
- Chapter 4 develops two CNN-based architectures for FEL pulse profile reconstruction from the measured electron phase space distribution.

- Chapter 5 details the core experimental contribution of this work: the implementation and commissioning of the EEHG-based radiation source at DELTA. It describes the first experimental observation of EEHG radiation in a storage ring.
- Chapter 6 presents a conceptual study for a dedicated, permanent-magnet-based EEHG setup.
- Finally, Chapter 7 summarizes the key findings and gives an outlook.

Chapter 2

Basic Concepts

2.1 Electron Beam Dynamics

In a charged particle accelerator, the trajectory of a particle is manipulated by magnetic fields. The fundamental interaction is described by the Lorentz force, which acts on a particle with charge q moving with velocity $\vec{v}$ through electric $\vec{E}$ and magnetic $\vec{B}$ fields as

$$F = q(\vec{E} + \vec{v} \times \vec{B}) \tag{2.1}$$

Since $\vec{v}$ is relativistic in a storage ring, for the purpose of guiding and focusing the beam, the magnetic component of the force is dominant. The force exerted by the magnetic field can be expressed in terms of the momentum of the particle $\vec{p}$ as

$$\vec{F} = \frac{q}{\gamma m_0} \vec{p} \times \vec{B}. \tag{2.2}$$

Here m_0 is the rest mass of the particle and γ is the Lorentz factor.

In practice, the trajectories of all the particles are described in relation to the orbit of an ideal particle with nominal momentum p_0 and a co-moving coordinate system [26, 27] whose origin falls at the ideal particle, as shown in Fig. 2.1.

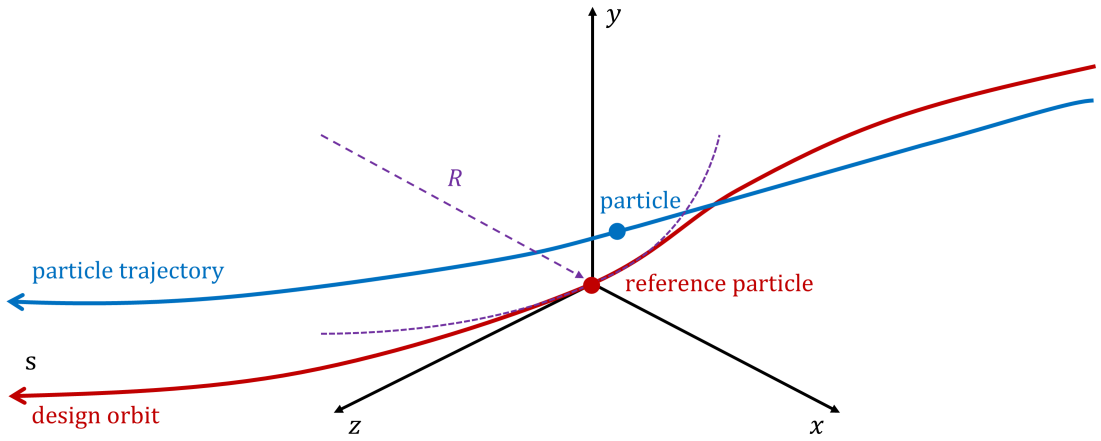

Figure 2.1: Sketch of the co-moving coordinate system. Here the reference particle (red) follows the orbit defined by the longitudinal coordinate s in the laboratory system. The relative positions of individual particles (blue) with respect to this reference electron are expressed by the transverse coordinates x (horizontal), y (vertical), and the longitudinal deviation z. The geometry of the orbit at any point s is described by its local curvature radius R.

A particle in this coordinate system can be described by a 6-dimensional (6D) coordinate vector,

$$\boldsymbol{X} = \begin{pmatrix} x \\ x' \\ y \\ y' \\ z \\ \Delta p/p_0 \end{pmatrix} \tag{2.3}$$

where x, y, and z are the particle position in the transverse plane and longitudinal direction, respectively, $x' = dx/ds$ and $y' = dy/ds$ are the derivatives of the particle trajectory relative to the reference orbit, and Δp is the momentum deviation from the reference momentum p_0.

The magnetic field in the transverse plane can be approximated as a multipole expansion, with each component having a different effect on the path of the particle. For example, in the horizontal plane, the vertical magnetic field is written as [26]

$$\begin{aligned} \frac{q}{p}B_y(x) &= \frac{q}{p}B_{y0} + \frac{q}{p}\frac{dB_y}{dx}x + \frac{1}{2!}\frac{q}{p}\frac{d^2B_y}{dx^2}x^2 + \frac{1}{3!}\frac{q}{p}\frac{d^3B_y}{dx^3}x^3 + \ldots \\ &= \text{dipole} + \text{quadrupole} + \text{sextupole} + \text{octupole} + \ldots. \end{aligned} \tag{2.4}$$

The dipole magnets have a constant magnetic field along the horizontal axis and only contribute to the bending of the particle trajectory. A quadrupole magnet has a magnetic field with a gradient $g = dB_y/dx$ resulting in a linearly increasing magnetic field with the trans-

verse displacement from the center of the magnet and is used for focusing the particles. A quadrupole magnet focuses particles in one plane but defocuses them in the other transverse plane. This means that, for an overall focusing, an arrangement of quadrupole magnets with alternating polarities needs to be used.

In a simplified accelerator model where the beam is guided and focused only by horizontal dipoles and quadrupoles, the particle dynamics are governed by linear differential equations. The equations of motion in the transverse phase space for a particle with momentum $p = p_0 + \Delta p$ are given by [26]

$$x''(s) + \left(\frac{1}{R(s)} + k(s)\right) x(s) = \frac{1}{R(s)} \frac{\Delta p}{p_0} \tag{2.5}$$

$$y''(s) - k(s)y(s) = 0 \tag{2.6}$$

where

$$R(s) = \frac{p_0}{qB(s)} \tag{2.7}$$

is the bending radius of the particle orbit given by the magnetic field of the dipole magnet, and

$$k(s) = \frac{q}{p_0} \frac{dB_y(s)}{dx} \tag{2.8}$$

is the quadrupole strength, which is given by the gradient of the magnetic field.

The trajectories of the particles execute a transverse oscillation about the reference orbit with a position-dependent amplitude and phase. This is called a betatron oscillation and is described by the transverse position $x(s)$ and the angle $x' = dx/ds$ given by [26]

$$x(s) = \sqrt{C_x \beta_x(s)} \cos\left(\Psi_x(s) + \phi_x\right) + D(s) \frac{\Delta p}{p_0} \tag{2.9}$$

$$x'(s) = -\sqrt{\frac{C_x}{\beta_x(s)}} \Big[\alpha(s) \cos(\Psi_x(s) + \phi_x) + \sin(\Psi_x(s) + \phi_x) + D'(s) \frac{\Delta p}{p_0}\Big], \tag{2.10}$$

where the amplitude function $\beta_x(s)$ is called the beta function and

$$\alpha_x(s) = -\frac{\beta_x'(s)}{2}. \tag{2.11}$$

The phase advance is given by

$$\Psi_x(s) = \int_0^s \frac{d\sigma}{\beta(\sigma)} \tag{2.12}$$

and C_x is the Courant-Snyder invariant. A particle within a periodic lattice, observed at the same position s during each revolution, follows an elliptical path in the transverse phase space, as shown in Fig. 2.2. The value πC_x gives the area of this ellipse and is a constant for an ultra-relativistic particle. For a Gaussian transverse distribution of particles, the

Courant-Snyder invariant of the particles that lie one standard deviation from the beam axis is called the beam emittance ϵ.

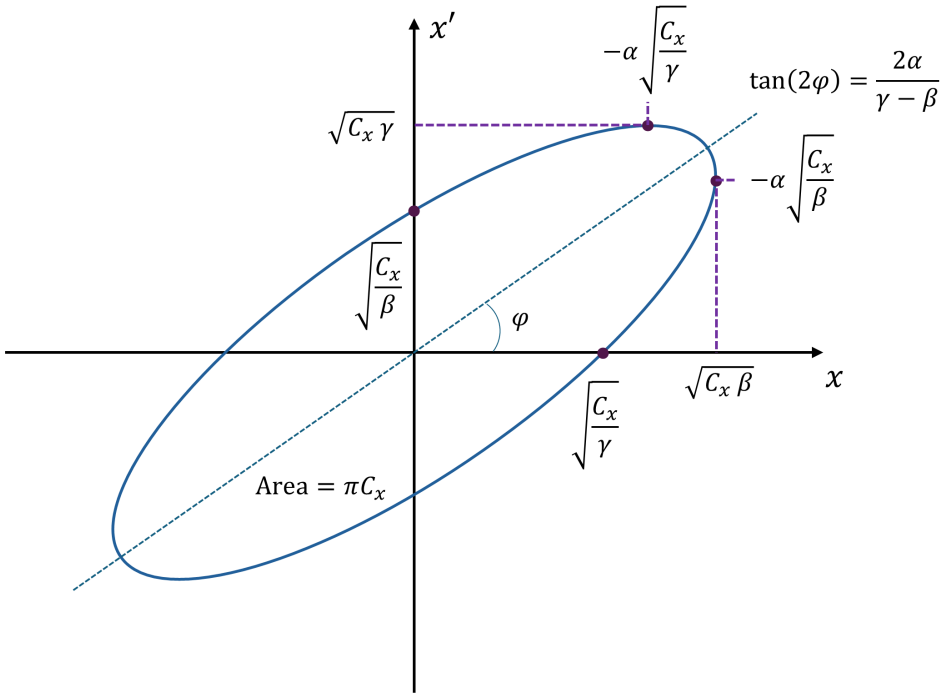

Figure 2.2: Phase space ellipse traced by a particle's betatron oscillation in the horizontal phase space (x, x') when observed at a fixed position s over many turns. The shape and orientation of the ellipse are determined by the Twiss parameters α, β, and γ. The area of the ellipse πC_x is an invariant of the motion.

In the horizontal phase space,

$$C_x = \gamma_x(s)x^2(s) + 2\alpha_x(s)x(s)x'(s) + \beta_x(s)x'^2(s), \tag{2.13}$$

where

$$\gamma_x(s) = \frac{1 + \alpha_x^2(s)}{\beta_x(s)}. \tag{2.14}$$

The functions $\alpha(s)$, $\beta(s)$ and $\gamma(s)$ are collectively called Twiss parameters or optical functions. In this model, the betatron oscillations in the horizontal and vertical planes are assumed to be uncoupled. This approximation neglects fields that couple the motion between the two planes and allows the dynamics to be described by independent equations. The same relations discussed for horizontal betatron oscillations hold true for vertical phase space with a different set of Twiss parameters and Courant-Snyder invariants.

The horizontal dispersion function $D_x(s)$ introduces a momentum-dependent horizontal offset Δx to the particle as

$$D_x(s) = \frac{\Delta x(s)}{\Delta p/p_0}. \tag{2.15}$$

The vertical dispersion and its derivative are zero for an accelerator with only horizontally deflecting dipoles.

While the dispersion function describes the momentum-dependent transverse offset, the overall effect on the path length of the closed orbit is characterized by the momentum compaction factor α_p. This parameter is defined as the ratio of relative change in path length to relative momentum deviation, given by [26]

$$\alpha_p = \frac{\Delta L/L_0}{\Delta p/p_0} = \frac{1}{L}\oint \frac{D(s)}{R(s)} ds. \tag{2.16}$$

A particle in a storage ring lattice can be tracked through a 6×6 transfer matrix whose elements depend on the radius of curvature R and quadrupole strength k which is structured for each element individually. The particle coordinates at any position $s+\Delta s$ is calculated by multiplying the transfer matrix corresponding to the length Δs with the 6D initial coordinate vector at position s. For betatron motion without coupling between the horizontal and vertical plane, this will result in

$$\boldsymbol{X}_{s+\Delta s} = \begin{pmatrix} R_{11} & R_{12} & 0 & 0 & R_{15} & R_{16} \\ R_{21} & R_{22} & 0 & 0 & R_{25} & R_{26} \\ 0 & 0 & R_{33} & R_{34} & 0 & 0 \\ 0 & 0 & R_{43} & R_{44} & 0 & 0 \\ R_{51} & R_{52} & 0 & 0 & 0 & R_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \boldsymbol{X}_s \tag{2.17}$$

with the matrix elements calculated for each element between s and $s + \Delta s$. For example, the 6×6 transfer matrix in the ultra-relativistic regime for a drift section of length L is given by

$$R_{\text{drift}} = \begin{pmatrix} 1 & L & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.18}$$

The matrix elements for different magnets can be found in accelerator physics textbooks [26, 28]. For storage rings with inter-plane coupling, the off-diagonal submatrix elements will be nonzero, and a more complex formalism is required to explain the betatron motion in that case [29, 30].

2.2 Synchrotron Radiation Sources

Particle accelerators have undergone significant evolution across generations. Starting from the earliest accelerators, such as cathode ray tubes [31], the field has now advanced to investigating novel techniques such as laser plasma acceleration [32]. The emergence of synchrotrons and storage rings as light sources was a significant advancement, revolutionizing research in various scientific disciplines and offering insight into the structure of matter at the atomic level. With free-electron lasers and fourth-generation synchrotron light sources such as diffraction-limited storage rings, the boundaries of spatial and temporal resolution are being pushed forward.

An electron with charge e subjected to acceleration will produce electromagnetic radiation with a radiated power given by the Larmor formula [33]

$$P_{\mathrm{s}} = \frac{e^2}{6\pi\varepsilon_0 m_0^2 c^3}\left(\frac{dp}{dt}\right)^2, \tag{2.19}$$

where ε_0 is the permittivity of free space, m_0 is the electron rest mass, and c is the velocity of light. In circular accelerators, this acceleration is applied perpendicularly to the particle's velocity by the dipole magnets, thereby bending the beam and maintaining its circular trajectory. For relativistic velocities, Eq. (2.19) becomes [34],

$$P_{\mathrm{s}} = \frac{e^2 c}{6\pi\varepsilon_0 (m_0 c^2)^4}\frac{E^4}{R^2}, \tag{2.20}$$

where R is the radius of curvature, and E is the kinetic energy of the electron. Since P_s is inversely proportional to the fourth power of rest mass energy (m_0c^2), the power emitted by electrons is much higher than that of protons with comparable energies. This is the reason why synchrotron light sources are exclusively based on electrons rather than heavier particles like protons.

The spectral distribution of the synchrotron radiation from a bending magnet is given by [27]

$$\frac{dP}{d\omega} = \frac{P_{\mathrm{s}}}{\omega_c} S\left(\frac{\omega}{\omega_c}\right) \quad \text{with} \quad S\left(\frac{\omega}{\omega_c}\right) = \frac{9\sqrt{3}}{8\pi}\frac{\omega}{\omega_c}\int_{\omega/\omega_c}^{\infty} K_{5/3}(x)\,dx, \tag{2.21}$$

where $K_{5/3}$ is the modified Bessel function of the second kind, $\omega_c = 3c\gamma^3/2R$ is the critical frequency that divides the synchrotron radiation spectrum into two regions of equal radiated power. The universal synchrotron radiation function S is shown in Fig. 2.3.

In the center-of-mass frame, the angular distribution of the radiation for an electron accelerated perpendicular to its motion is similar to that of the Hertz dipole. But in the laboratory frame, this pattern is modified by relativistic beaming. This effect, a consequence of the

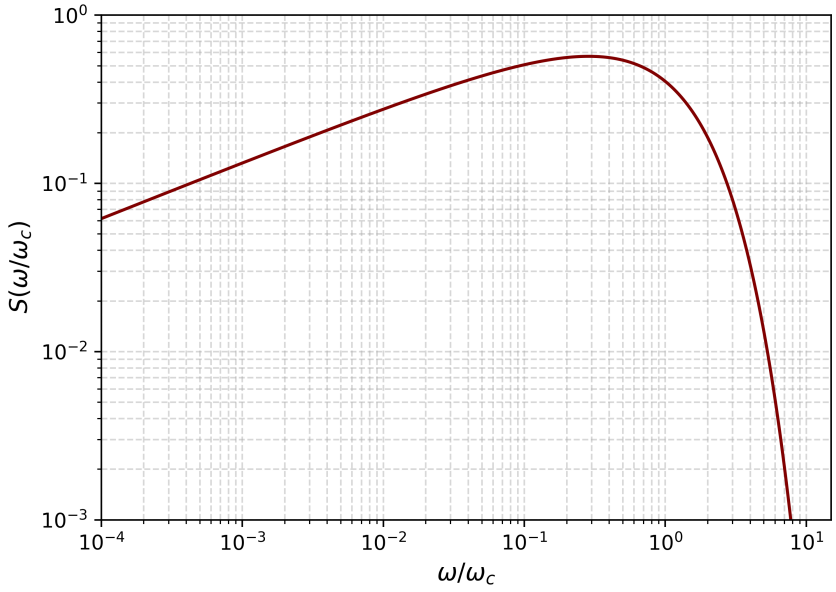

Figure 2.3: Universal synchrotron radiation function $S(\omega/\omega_c)$ for a bending magnet.

Lorentz transformation, causes the radiation to be concentrated into a forward-peaked cone with a half-opening angle of approximately $1/\gamma$.

Insertion Devices

Insertion devices are specialized magnetic structures that are inserted into the accelerator lattice, either in the available straight sections of a storage ring or along the beam path of a linear accelerator, to perform a specific function. In a storage ring, these straight sections are the dedicated drift spaces located between the magnetic lattice cells. The most common types of insertion devices are wigglers and undulators. These devices use a periodic magnetic structure with the primary function of forcing a relativistic electron beam to oscillate. This oscillation produces highly intense and collimated synchrotron radiation, far exceeding that from the ring's main bending magnets. Wigglers and undulators differ primarily in the strength of their magnetic fields and produce radiation with distinctly different properties, despite being based on the same physical principle.

A wiggler or undulator consists of a periodic array of magnets. The most common configuration is a planar device, where dipoles with alternating polarity create a sinusoidal vertical magnetic field along the direction s of the electron's motion. The vertical component of this field is represented as [26]

$$B(s) = B_0 \sin\left(\frac{2\pi s}{\lambda_\mathrm{u}}\right), \tag{2.22}$$

where λ_u is the length of one period and B_0 is the peak magnetic field. Helical undulators, which use a rotating field to induce a helical trajectory and produce circularly polarized

light, are also of key importance.

The fundamental difference between a wiggler and an undulator is determined by the magnitude of the electron's angular deflection compared to the natural opening angle of the synchrotron radiation ($1/\gamma$). This distinction is quantified by the dimensionless undulator parameter (K parameter) defined as [26]

$$K = \frac{eB_0\lambda_\mathrm{u}}{2\pi m_e c} \approx 0.934\, B_0\,[\mathrm{T}]\,\lambda_\mathrm{u}\,[\mathrm{cm}]. \tag{2.23}$$

When the K parameter is large ($K \gg 1$), the electron undergoes large-amplitude oscillations with a maximum deflection angle that is much larger than the natural emission angle of the radiation $1/\gamma$. In this regime, the radiation cones emitted at each period of the electron path are widely separated, and the radiation from the different poles adds incoherently. The resulting radiation spectrum is broad and continuous, similar to that from a bending magnet but significantly more intense. If a wiggler has $2N_\mathrm{u}$ poles, its flux is approximately $2N_\mathrm{u}$ times greater than that of a single bending magnet with the same field strength. The primary characteristic of wiggler radiation is its high flux over a wide energy range.

When the K parameter is moderate, the maximum deflection angle is comparable to or smaller than the natural radiation cone opening angle. The key advantage of undulators is that radiation emitted from different periods of the magnetic structure interferes constructively as the electron slips behind a photon it emitted by one radiation wavelength λ_r over the course of one magnetic period λ_u, and the waves add up in phase. This interference effect drastically modifies the emission spectrum. Instead of a broad continuum, the spectrum consists of a series of sharp peaks, or harmonics, at integer fractions of the fundamental wavelength. The fundamental wavelength λ_r of the radiation is given by the undulator equation [35]

$$\lambda_r = \frac{\lambda_\mathrm{u}}{2\gamma^2}\left(1 + \frac{K^2}{2} + \gamma^2\theta^2\right), \tag{2.24}$$

where θ is the observation angle. The wavelength is tuned by controlling B_0 (and hence K) or λ_u or the electron energy (γ). For a planar undulator, the on-axis spectrum consists of only the odd harmonics ($n = 1, 3, 5, ...$) of the fundamental wavelength. Due to the symmetry of the electron's sinusoidal motion, the intensity of even harmonics ($n = 2, 4, ...$) is zero on-axis, but these harmonics appear with nonzero intensity at specific off-axis angles.

The radiation is a result of the coherent addition of the light from each of the N_u magnetic periods. This adds up the amplitudes in phase, causing the peak intensity to scale as N_u^2, while the interference sharpens the spectral line, making the relative spectral width ($\Delta\lambda/\lambda$) inversely proportional to N_u. This combination of high intensity and narrow linewidth, along with the small source size and divergence, makes the undulator radiation more powerful for experiments than the radiation from a dipole magnet.

2.3 Ultrashort Laser Pulses

The short-pulse source at the DELTA facility uses a Ti:sapphire laser system [36] which consists of a mode-locked oscillator and a subsequent amplifier based on chirped-pulse amplification (CPA) [37]. The 800 nm Ti:sapphire laser oscillator installed at the DELTA short-pulse facility is based on an optical resonator incorporating a Ti:sapphire crystal as the gain medium. The crystal is optically pumped by a continuous-wave (CW) Nd:YVO_4 laser [36], which operates at a fundamental wavelength of 1064 nm and uses frequency-doubling to produce the 532 nm pump beam.

In an optical resonator, light forms standing waves that satisfy the resonance condition, requiring the optical path length of the cavity to be an integer multiple of half the wavelength of the standing wave. Without any phase correlation between the modes, the laser emits a CW output with each longitudinal mode oscillating independently and contributing incoherently to the total field. This results in a multi-mode emission with a relatively constant average intensity but rapid phase and amplitude fluctuations.

To generate ultrashort pulses, it is necessary to enforce a fixed phase relationship among the longitudinal modes through a process known as mode-locking [38]. When the modes are phase-locked, they interfere constructively at regular time intervals, resulting in a periodic train of pulses. This synchronization can be achieved through either active or passive methods. Active mode-locking utilizes external devices, such as acousto-optic or electro-optic modulators, to modulate the cavity losses at the round-trip frequency. In contrast, passive mode-locking relies on nonlinear optical effects, such as a saturable absorber or the Kerr-lens effect, which instantaneously reduce cavity losses for high-intensity pulses relative to CW operation [39]. The repetition rate f_{rep} of this pulse train is determined by the round-trip time in the cavity as

$$f_{\mathrm{rep}} = \frac{c}{2L}. \tag{2.25}$$

Ultrashort laser pulses are best described by their electric field as a function of time, or equivalently, by their spectral amplitude and phase. In the time domain, the electric field of a laser pulse is written as

$$E(t) = A(t)e^{i(\omega_0 t + \phi(t))}, \tag{2.26}$$

where $A(t)$ is the complex field envelope, ω_0 is the central angular frequency, and $\phi(t)$ is the time-dependent phase. The instantaneous frequency ω_i is given by

$$\omega_i(t) = \omega_0 + \frac{d\phi(t)}{dt}. \tag{2.27}$$

A pulse with a nonzero $d\phi(t)/dt$ is 'chirped' with its frequency varying with time along the pulse duration.

Taking the Fourier transform of $E(t)$ yields the spectral field given by

$$\tilde{E}(\omega) = \tilde{A}(\omega)e^{i\phi(\omega)} \tag{2.28}$$

where $\tilde{A}(\omega)$ is the spectral amplitude, and $\phi(\omega)$ is the spectral phase. The Taylor expansion of $\phi(\omega)$ around the central frequency ω_0 gives

$$\phi(\omega) = D_0 + D_1(\omega - \omega_0) + D_2(\omega - \omega_0)^2 + D_3(\omega - \omega_0)^3 + \cdots \tag{2.29}$$

where D_1 is the group delay, D_2 is the group delay dispersion (GDD) and D_3 is the third-order dispersion (TOD).

The time-bandwidth product (TBP) expresses the fundamental uncertainty between temporal and spectral widths. For a pulse with FWHM pulse length Δt and spectral FWHM $\Delta\nu$, the TBP is given by

$$\Delta t \Delta\nu \geq C \tag{2.30}$$

where C is a constant depending on the pulse shape. For a Gaussian pulse, $C = 0.441$. A pulse that satisfies $\Delta t \Delta\nu = C$ is said to be transform-limited or Fourier-limited. For a transform-limited pulse, the spectral phase would be flat or linear (with only nonzero D_0 and D_1). Any nonzero higher-order terms in the spectral phase broaden and distort the pulse.

When a femtosecond pulse propagates through a transparent material, its different frequency components travel at different group velocities due to the frequency dependence of the refractive index. This phenomenon is known as group velocity dispersion (GVD). The group velocity is defined as

$$v_g = \left(\frac{dk}{d\omega}\right)^{-1}, \tag{2.31}$$

where $k(\omega)$ is the wavenumber and ω is the angular frequency. The second derivative of the wavenumber with respect to frequency $d^2k/d\omega^2$ characterizes the GVD of the material and is usually expressed in units of fs^2/mm. At a central wavelength of 800 nm, typical optical materials such as fused silica exhibit positive GVD with values around $+36\ \text{fs}^2/\text{mm}$. The total GDD accumulated by the pulse while propagating through a medium of length L is given by

$$\text{GDD} = \text{GVD} \cdot L. \tag{2.32}$$

A nonzero GDD introduces a quadratic spectral phase, leading to temporal broadening of the pulse and a linear frequency chirp. After a Gaussian pulse passes through a transparent medium, it broadens from the transform-limited pulse length τ_0 according to [40]

$$\tau = \tau_0\sqrt{1 + \left(\frac{4\ln 2 \cdot \text{GDD}}{\tau_0^2}\right)^2}. \tag{2.33}$$

This expression shows that dispersion-induced broadening becomes more significant for shorter initial pulse durations. While group delay and GDD are responsible for temporal shifting and symmetric broadening of the pulse, higher-order dispersion introduces asymmetric features and distortion. Nonzero TOD can lead to the formation of satellite pulses.

In practical femtosecond systems, dispersion from all optical elements, including lenses, beam splitters, and windows, accumulates and must be carefully compensated using dispersive components such as prism pairs, gratings, or chirped mirrors.

Chirped Pulse Amplification (CPA)

Femtosecond pulses, due to their extremely short duration and correspondingly high peak power, cannot be directly amplified using conventional laser gain media without causing damage or nonlinear distortions. Chirped Pulse Amplification (CPA) [37] solves this problem by stretching, amplifying, and then recompressing the pulse, thereby reducing the peak intensity during amplification. A schematic of the CPA system used at DELTA is illustrated in Fig. 2.4.

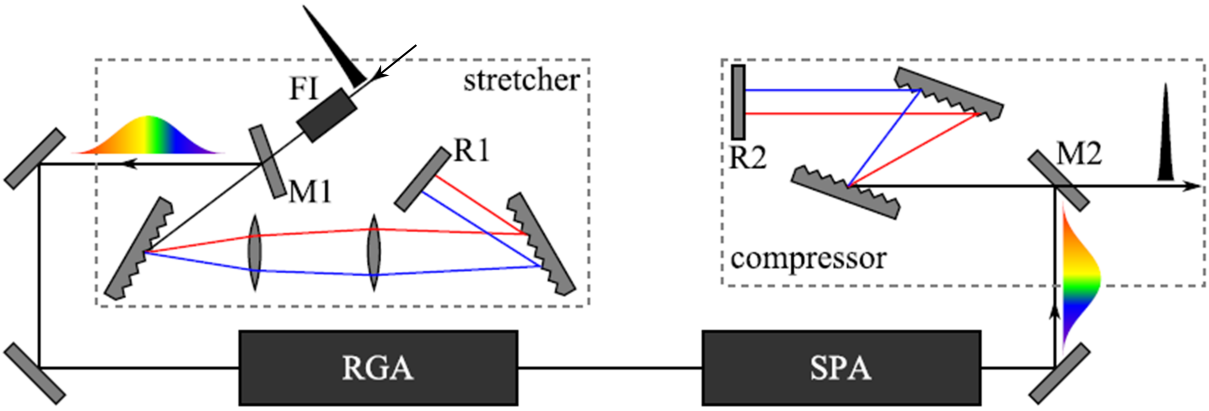

Figure 2.4: Design of a Ti:sapphire laser amplifier utilizing chirped-pulse amplification (CPA). Short-duration, high-intensity pulses (shown in black) are directed into the pulse stretcher via a Faraday isolator [41], which blocks unwanted optical feedback to the source oscillator. The stretcher configuration comprising diffraction gratings, focusing lenses, and a vertical reflecting element (R1) creates an extended optical pathway for shorter wavelengths compared to longer wavelengths. This wavelength-dependent path difference introduces positive GDD, temporally broadening the pulses with a frequency chirp (depicted as the multicolored pulse). Following amplification through regenerative (RGA) and single-pass (SPA) amplifier stages, the broadened pulses enter the compressor section. The compressor's grating and retroreflector arrangement (R2) applies a negative GDD to restore the original pulse duration. The recompressed pulses exit the amplifier via mirror M2. Image taken from [42].

Initially, the transform-limited femtosecond pulse from the oscillator is passed through a dispersive optical setup (such as a Martinez stretcher configuration, which uses a telescope placed between two gratings, as shown in the section labeled 'stretcher' in Fig. 2.4) that imposes a large positive GDD, resulting in a temporally stretched pulse. The pulse duration typically increases by several orders of magnitude, from tens of femtoseconds to hundreds of picoseconds, reducing the peak power by the same factor. The stretched pulse is then amplified using a regenerative or multipass amplifier. In a regenerative amplifier (RGA), the pulse is injected into an optical resonator and makes many passes through the gain medium. Its injection and ejection are actively controlled by an optical switch such as a Pockels cell. In contrast, a multipass amplifier directs the beam through the gain medium, where each pass follows a slightly different geometric path. Amplification can increase the pulse energy from nanojoules (oscillator) to millijoules or more, depending on the gain and number of passes through the gain medium.

After amplification, the chirped pulse is sent through a compressor that introduces negative dispersion to cancel the chirp imposed during stretching, which is typically achieved with a grating pair, as shown in Fig. 2.4. Ideally, the spectral phase is restored to a flat or linear profile, and the pulse is recompressed close to its original transform-limited duration. To ensure high-fidelity compression, the stretcher and compressor must be carefully matched not only in GDD, but also in higher-order dispersion terms such as TOD. Mismatches can result in residual chirp, asymmetric pulse shapes, and satellite pulses.

At the DELTA short-pulse facility, the CPA scheme is realized using a commercial Ti:sapphire amplifier system (Coherent Legend Elite Duo HP [36]) seeded by the Kerr-lens mode-locked oscillator (Coherent Micra-5). The oscillator delivers pulses with a minimum pulse duration of 45 fs and an energy of a few nanojoules, operating at a repetition rate of 83.3 MHz, which is synchronized to the 500 MHz storage ring radiofrequency with its 6th subharmonic. The amplifier system operates at a 1 kHz repetition rate, providing a final compressed output with pulse energies of up to 8 mJ, which corresponds to an average power of 8 W.

2.4 Gaussian Beam Optics

The propagation of laser beams in free space or homogeneous media can often be accurately described using the framework of Gaussian beam optics, which is based on solutions to the paraxial wave equation. A Gaussian beam represents the fundamental transverse mode (TEM_{00}) of laser radiation and is characterized by a bell-shaped intensity profile in the transverse direction.

The complex electric field of a Gaussian beam propagating in the $+s$ direction is typically

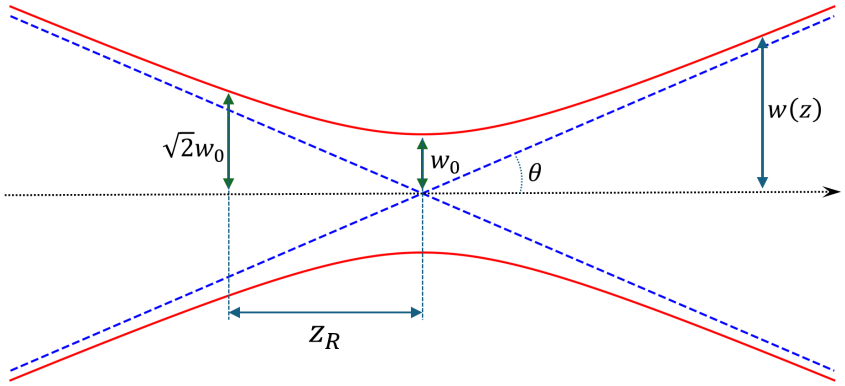

Figure 2.5: Sketch of a Gaussian beam waist.

written as [40]

$$\tilde{E}(x,y,s,t) = E_0 e^{i(\omega t - ks)} \frac{w_0}{w(s)} e^{i\zeta(s)} \exp\left(-\frac{r^2}{w^2(s)}\right) \exp\left(-i\frac{kr^2}{2R(s)}\right), \tag{2.34}$$

where $k = 2\pi/\lambda$ is the wavenumber, $r^2 = x^2 + y^2$ is the radial distance from the propagation axis, $w(s)$ is the beam radius, w_0 is the beam waist radius, $R(s)$ is the radius of curvature of the wavefronts, and $\zeta(s)$ is the Gouy phase shift. Some key parameters governing the Gaussian beam are described below and are depicted in Fig. 2.5.

Beam Waist and Beam Radius: The beam waist size, denoted by w_0, is the minimum radius of the beam, which occurs at the plane of focus ($s = 0$ by convention). The beam radius at any position s is defined as the radial distance where the electric field amplitude drops to $1/e$ of its on-axis value, which corresponds to the intensity dropping to $1/e^2$ (about 13.5%). The evolution of the beam radius is given by [40]

$$w(s) = w_0 \sqrt{1 + \left(\frac{s}{s_R}\right)^2}, \tag{2.35}$$

where s_R is the Rayleigh length (Eq. (2.36)).

Rayleigh Length: Rayleigh length s_R is a parameter which defines the beam's "depth of focus". It is the distance along the propagation axis from the waist to the position where the cross-sectional area of the beam, $\pi w(s)^2$, has doubled. At this point, the beam radius has increased to $w(s_R) = \sqrt{2}w_0$. The Rayleigh range is determined by the waist size and the wavelength [40]

$$s_R = \frac{\pi w_0^2}{\lambda}. \tag{2.36}$$

Wavefront Curvature: Except at the beam waist, the wavefronts of a Gaussian beam are approximately spherical. The radius of curvature of these wavefronts $R(s)$ is given by [40]

$$R(s) = s\left[1 + \left(\frac{s_R}{s}\right)^2\right]. \tag{2.37}$$

At the waist ($s = 0$), $R(s) \to \infty$, corresponding to a planar wavefront. For large distances from the waist ($|s| \gg s_R$), the radius of curvature approaches the distance from the waist, $R(s) \approx s$, meaning the wavefronts are nearly spherical, originating from the center of the waist.

Gouy Phase Shift: The Gouy phase $\zeta(s)$ is a phase shift that a Gaussian beam acquires relative to a uniform plane wave as it propagates through its focus. It is given by [40]

$$\zeta(s) = \arctan\left(\frac{s}{s_R}\right) \tag{2.38}$$

with a total phase shift of π across the entire focal region from $s = -\infty$ to $s = +\infty$.

Beam Divergence: In the far field ($|s| \gg s_R$), the divergence angle θ of the beam approaches

$$\theta = \lim_{s\to\infty} \frac{w(s)}{s} \approx \frac{w_0}{s_R} = \frac{\lambda}{\pi w_0}. \tag{2.39}$$

This implies that a tightly focused beam will have a large divergence angle, while a highly collimated beam cannot be focused to a very small spot.

Beam Quality (M^2 Factor): The ideal Gaussian beam described above is the fundamental TEM_{00} mode. However, real laser beams may be composed of a superposition of higher-order modes such as Hermite-Gaussian or Laguerre-Gaussian modes, or be distorted by imperfections in the laser cavity or optics.

The quality of a real laser beam is quantified by the beam quality factor, or M^2 factor. It is defined as the ratio of the beam parameter product (BPP) of the actual beam to that of an ideal Gaussian beam of the same wavelength.

$$M^2 = \frac{\mathrm{BPP}_{\mathrm{real}}}{\mathrm{BPP}_{00}} \geq 1. \tag{2.40}$$

The BPP is the product of the beam waist radius and the far-field divergence angle. For an ideal Gaussian beam, it is given by [40]

$$\mathrm{BPP}_{00} = w_0\theta = \lambda/\pi. \tag{2.41}$$

The equations for Rayleigh length and divergence are modified to account for this as [40]

$$s_R = \frac{\pi w_0^2}{M^2 \lambda} \quad \text{and} \quad \theta = \frac{M^2 \lambda}{\pi w_0}. \tag{2.42}$$

2.5 Free-Electron Lasers

While synchrotron radiation provides an excellent source of broadband, partially coherent radiation, the quest for higher brightness, full coherence, and shorter pulse duration has led to the development of free-electron lasers (FELs). Unlike conventional lasers, which rely on stimulated emission from atomic or molecular transitions, FELs generate radiation through the interaction of a relativistic electron beam with an undulator's magnetic field, where the interplay between the electrons and the radiation field drives an exponential intensity amplification. This enables FELs to achieve exceptional brightness, tunability, and coherence far surpassing conventional synchrotron radiation sources [8].

The basic operating principle of a high-gain FEL involves passing a highly focused electron beam with high peak current, a small energy spread, and low emittance through a long undulator. The undulator radiation then co-propagates with the electron bunch. Under the resonance condition, a coherent interaction between the electrons and their own radiation field leads to energy exchange that exponentially amplifies the radiation. This is termed self-amplified spontaneous emission (SASE) [43].

The fundamental quantity governing FEL performance is the Pierce parameter (also known as the FEL parameter) ρ, which characterizes the strength of the electron-radiation coupling given by [43]

$$\rho = \left[\frac{1}{16} \frac{I_e}{I_A} \frac{\widehat{K}^2}{\gamma^3 \sigma_x^2 k_\mathrm{u}^2} \right]^{1/3}, \tag{2.43}$$

where I_e is the peak current of an electron bunch, $I_A \approx 17$ kA is the Alfvén current, γ is the Lorentz factor, λ_u is the undulator period length, $k_\mathrm{u} = 2\pi/\lambda_\mathrm{u}$, and σ_x is the root-mean-square (rms) transverse size of the electron beam. Furthermore, $\widehat{K}$ is the modified undulator parameter given by

$$\widehat{K} = K \cdot \left[J_0\left(\frac{K^2}{4 + 2K^2} \right) - J_1\left(\frac{K^2}{4 + 2K^2} \right) \right], \tag{2.44}$$

where J_0 and J_1 are the Bessel functions of order 0 and 1 and K is the undulator parameter. In terms of the Pierce parameter, the one-dimensional power gain length L_{G0} is defined as [43]

$$L_{G0} = \frac{\lambda_\mathrm{u}}{4\pi\sqrt{3}\rho}. \tag{2.45}$$

The Pierce parameter also governs other key FEL characteristics such as the saturation efficiency $\eta_{\text{sat}} \propto \rho$ and the spectral bandwidth $\Delta\omega/\omega \propto \rho$ [44]. Typical values range from $\rho \sim 10^{-3}$ for soft X-ray facilities to $\rho \sim 10^{-4}$ for hard X-ray FELs.

2.5.1 Evolution of the FEL Process

The SASE process begins with shot noise in the electron beam, which provides the initial seed for radiation growth. The process is typically divided into three distinct regimes along the undulator: an initial start-up regime, an exponential high-gain regime, and a saturation regime. This characteristic evolution of FEL power is illustrated in Fig. 2.6.

In the initial part of the undulator under the start-up regime (blue shaded region in Fig. 2.6), electrons traverse the magnetic field and emit spontaneous, incoherent synchrotron radiation. This initial radiation, which co-propagates with the bunch, is weak and serves as the "seed" for the coherent amplification that follows.

The interaction between electrons and the electromagnetic field is described by a set of coupled equations. The rapid oscillation of the electrons in these fields gives rise to the ponderomotive force, which is the nonlinear force experienced by a charged particle in an inhomogeneous oscillating electromagnetic field. Arising here from the combined fields of

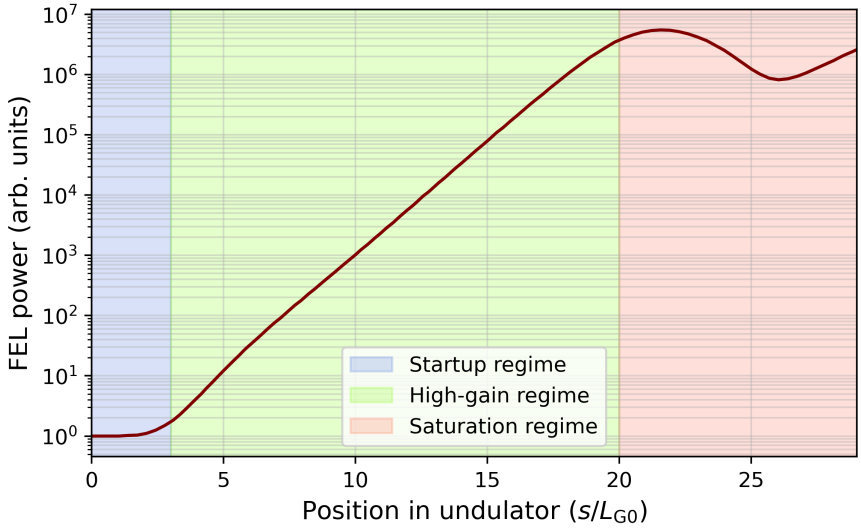

Figure 2.6: A typical FEL gain curve. The region between 0 and $3L_{G0}$ (startup regime, shaded in blue) shows only a small increase in the FEL power, while in the high-gain regime (shaded in green) the power increases exponentially until it saturates beyond around $20\,L_{G0}$ (saturation regime, shaded in red).

the undulator and the radiation, this force modulates the electron energy as [45, 46]

$$\frac{d\eta_n}{ds} = -\frac{e\widehat{K}}{2\gamma_0^2 m_e c^2}\,\mathrm{Re}\left[\mathcal{E}(s)\,e^{i\psi_n(s)}\right], \tag{2.46}$$

where $\mathcal{E}(s)$ is the complex amplitude of the radiation electric field, $\eta_n = \Delta E/E_0$ represents the relative energy deviation of the n-th electron from the resonant energy E_0, and $\psi_n(s)$ is the ponderomotive phase describing the longitudinal position of the electron relative to the radiation field. These two coordinates define the ponderomotive phase space.

Since electrons move slower than the radiation wavefront due to their oscillatory motion, electrons continuously slip behind the radiation wavefront. While a resonant electron maintains a constant phase despite this slippage, any energy deviation leads to a ponderomotive phase evolution described by [45]

$$\frac{d\psi_n}{ds} = 2k_\mathrm{u}\eta_n(s). \tag{2.47}$$

During this initial phase (typically extending to $s \lesssim 3L_{G0}$), electrons are randomly distributed in the ponderomotive phase space. The energy modulation is initially weak, resulting in inefficient energy transfer between the electrons and the electromagnetic field.

The two equations (2.46) and (2.47) are coupled to a third equation describing the growth of the field amplitude $\mathcal{E}$, which is driven by the collective emission from the electrons [45] given by

$$\frac{d\mathcal{E}}{ds} = -\frac{\mu_0 c\widehat{K}}{2\gamma_0}\cdot j_0\,\langle e^{-i\psi}\rangle. \tag{2.48}$$

Here, the term $b = \langle e^{-i\psi}\rangle$ is the bunching factor and j_0 is the current density. The brackets indicate averaging over all electrons. As the interaction progresses, the electrons with higher energy slip forward in phase while the lower energy electrons fall behind. This leads to the formation of microbunches at the radiation wavelength scale, creating electron density spikes with spacing equal to one radiation wavelength, which results in an increased bunching factor. A larger b causes the field $\mathcal{E}$ to grow much faster. This larger field, in turn, creates a stronger energy modulation, which deepens the bunching. This positive feedback loop results in an exponential power growth as shown in the green shaded region in Fig. 2.6. In this high-gain regime ($3 \lesssim s/L_{G0} \lesssim 20$), the power grows exponentially as [45]

$$P(s) \propto e^{s/L_{G0}}. \tag{2.49}$$

The spectral bandwidth σ_ω of the FEL radiation also narrows during amplification according

to [45]

$$\frac{\sigma_\omega}{\omega} \propto \rho\sqrt{\frac{L_{\mathrm{G0}}}{s}}. \tag{2.50}$$

As the FEL continues to extract energy from the electron beam, the mean beam energy decreases, and the energy spread increases significantly. The initially stable microbunches begin to oscillate in the ponderomotive phase space with some electrons moving to phase space regions where they absorb energy from the radiation field. At this saturation regime ($s \gtrsim 20L_{\mathrm{G0}}$, red shaded region in Fig. 2.6) a dynamic equilibrium is established where the net energy exchange averages to zero. The FEL radiation power saturates at [45]

$$P_{\mathrm{sat}} = \rho P_{\mathrm{beam}}, \tag{2.51}$$

where $P_{\mathrm{beam}} = I_0 E/e$ is the power contained in the electron beam with I_0 being the beam current and E being the energy of one electron.

2.5.2 Self-Amplified Spontaneous Emission (SASE) and Seeded FELs

Self-Amplified Spontaneous Emission (SASE): SASE represents the most common mode of FEL operation, where the amplification process begins from electron beam shot noise. The random nature of this initial seed leads to significant shot-to-shot fluctuations in both spectral and temporal properties. Due to the statistical nature of shot noise, the temporal coherence is fundamentally limited. The coherence length is given by [43]

$$\tau_c = \frac{\sqrt{\pi}}{\sigma_\omega}. \tag{2.52}$$

When the electron bunch duration T_b exceeds the coherence time ($T_b > \tau_c$), the radiation consists of multiple intensity spikes, each with different amplitude and spectral characteristics. The number of coherent temporal and spectral spikes is approximately $M = T_b/\tau_c$ [43].

Seeded FELs: To overcome SASE limitations, various seeding techniques have been developed to initiate the FEL process with coherent radiation [47], thereby eliminating the statistical fluctuations inherent in SASE operation. Direct seeding uses an external laser at the fundamental FEL wavelength. This approach is limited to infrared and visible wavelengths where conventional laser sources are available. Frequency up-conversion techniques such as High-Harmonic Generation (HHG) [4, 48, 49] can be used to provide seed radiation at shorter wavelengths. Self-seeding techniques use the FEL itself as the seed source through

spectral filtering of SASE radiation in a first undulator section, which then seeds a second undulator section [50].

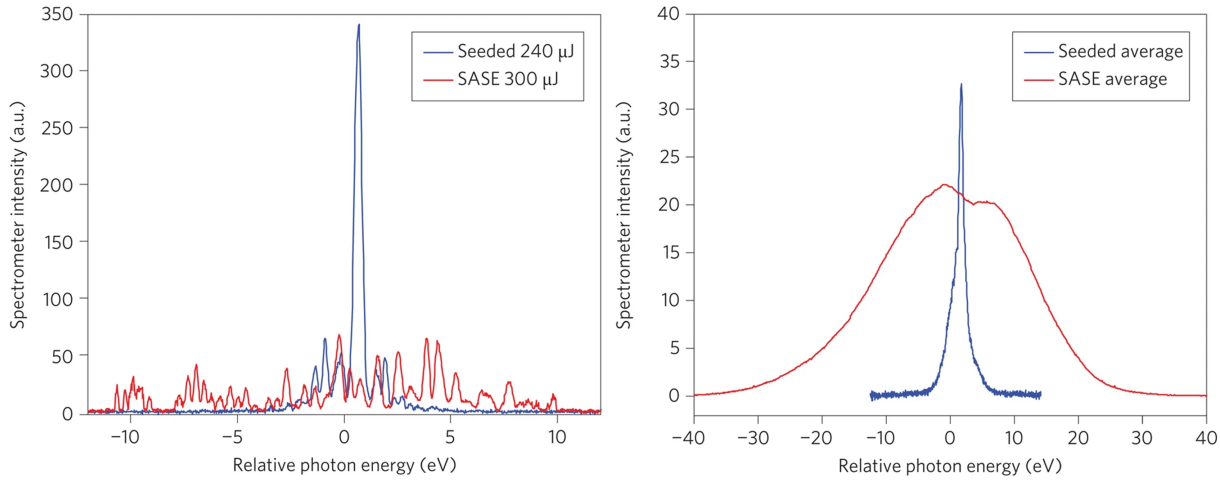

Figure 2.7: Single-shot (left) and averaged (right) X-ray spectrum in SASE mode (red) and self-seeded mode (blue). Image adapted with permission from [51].

Figure 2.7 shows the difference in the FEL radiation spectra between SASE and self-seeded operation. The spectra of the FEL pulses in seeded mode show a more stable and narrow bandwidth, while for the SASE mode, it is inherently dependent on the stochastic noise. However, the most common seeding schemes for X-ray FELs are High-Gain Harmonic Generation (HGHG) and Echo-Enabled Harmonic Generation (EEHG). These schemes, which are now in user operation [20, 21, 52–55], will be discussed in detail in the following section.

2.6 Short-Pulse Generation Techniques at Synchrotron Light Sources

The duration of the radiation pulses emitted by an electron beam in a storage ring is fundamentally constrained by the length of the electron bunches, which typically spans several tens of picoseconds. This limitation in pulse length makes it impossible to resolve ultrafast phenomena occurring on sub-picosecond timescales. These include phase transitions, chemical reactions, changes in molecular conformation, and electronic or magnetic structures. While operating modes such as 'low-alpha' [3, 56, 57] can compress the bunch length to a few picoseconds by reducing the momentum compaction factor (α_p) through specific tuning of the magnetic lattice, producing femtosecond-scale pulses from conventional synchrotron radiation sources requires further specialized techniques. These are discussed in the following sections.

2.6.1 Femtoslicing

The femtoslicing method, first introduced by Zholents and Zolotorev in 1996 [58], is an advanced technique for generating femtosecond-scale radiation pulses from conventional synchrotron sources. This method involves the interaction of a femtosecond laser pulse with a narrow slice of electrons within an electron bunch, within a modulator. The interaction induces an energy modulation in this selected slice of electrons. Following this energy modulation, the electron bunch passes through a dispersive section, which serves to angularly separate the modulated electrons from the rest of the bunch as illustrated in Fig. 2.8 [59]. The modulated and transversely separated electrons then emit radiation pulses whose duration closely matches that of the initial femtosecond laser pulse. This radiation can be selectively extracted using an aperture and beamline optics.

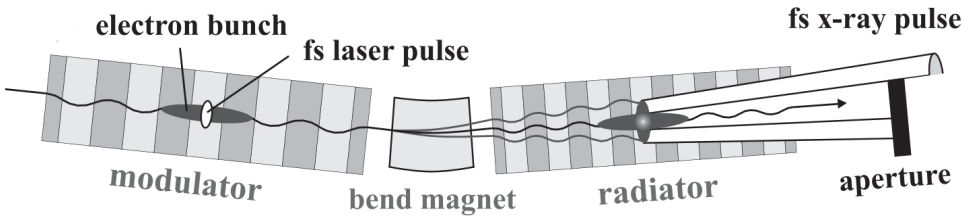

Figure 2.8: Schematics of the femtoslicing setup. An ultrashort laser pulse interacts with a long electron bunch in the modulator, inducing energy modulation. The subsequent bend magnet creates a spatial separation of the modulated slice. In the radiator, this slice emits femtosecond X-ray pulses, which are then isolated from the core beam background using an aperture. Figure adapted from [59].

While femtoslicing is a powerful technique for generating femtosecond-scale synchrotron radiation, it does have certain limitations. One significant drawback is that it produces incoherent synchrotron radiation, resulting in an intensity that is several orders of magnitude lower than what can be achieved with coherently generated synchrotron radiation. Additionally, only a small fraction of the electrons in the bunch participate in the radiation process, further reducing the overall efficiency of the technique. Nevertheless, femtoslicing offers the ability to tune the radiator to any arbitrary wavelength.

Femtoslicing was first demonstrated at the ALS in Berkeley, USA [10] using spatial separation, followed by BESSY II (Berlin, Germany) [59], SLS (Villigen, Switzerland) [60] and SOLEIL (Saint-Aubin, France) [61] using angular separation of femtosliced radiation from the rest of the bunch. Femtoslicing was also initially planned for DELTA [62], but was ultimately not realized.

2.6.2 Coherent Harmonic Generation (CHG)

The CHG scheme involves an initial undulator, known as the "modulator", where an ultrashort laser pulse interacts with the electrons in a short longitudinal slice of the bunch to induce a sinusoidal energy modulation within the slice. Secondly, a magnetic chicane, which introduces longitudinal dispersion, causes the formation of microbunches. Finally, the coherent emission of radiation occurs in an undulator called the "radiator". This scheme is conceptually similar to the HGHG seeding technique used in FELs. The difference in CHG is that the radiator is too short to provide exponential FEL gain. The schematics of the CHG configuration is depicted in Fig. 2.9.

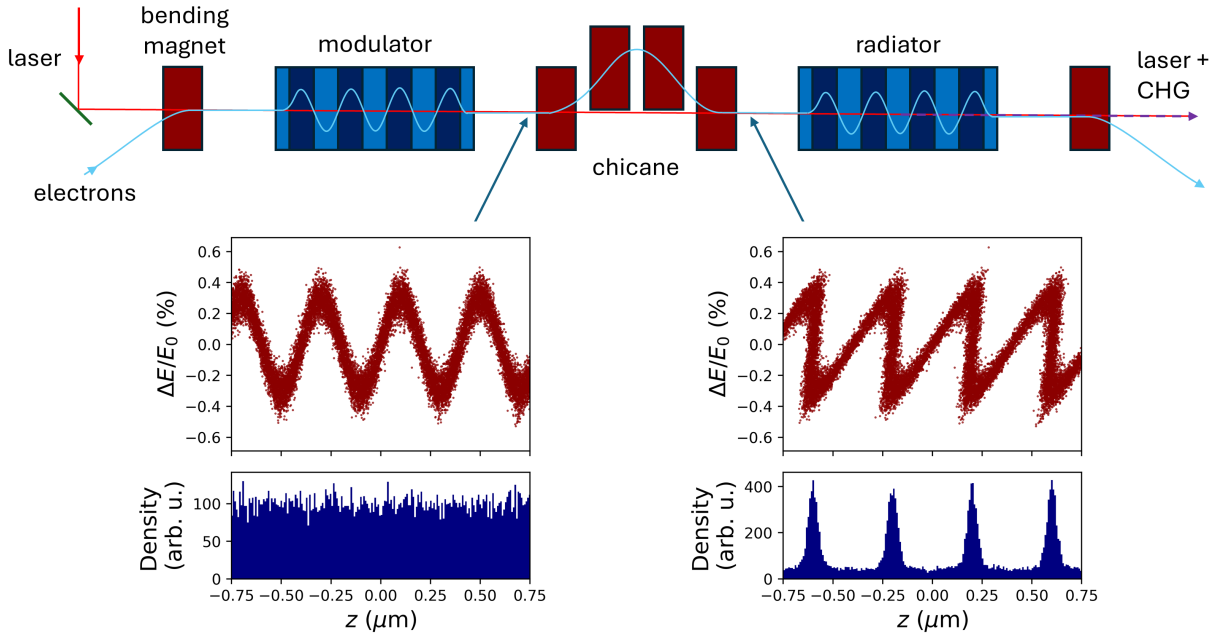

Figure 2.9: Top: Schematics of the CHG setup, which includes a seed laser, a modulator, a magnetic chicane, and a radiator. Bottom: Electron phase space distribution following the modulator and chicane. The energy modulation imparted by the laser is converted into a density modulation via the magnetic chicane, resulting in a series of closely spaced electron density maxima (known as microbunches).

With a modulator tuned to the laser wavelength, the electric field profile of the seed pulse is imprinted onto the longitudinal phase space distribution of the electrons. The periodicity of the resulting energy modulation corresponds to the laser wavelength, while the modulation envelope mirrors the shape of the laser pulse.

Considering a Gaussian laser pulse with an rms transverse size σ_r, the energy modulation amplitude for an electron located at radius r is given by [13, 63]

$$\Delta E(r) = \sqrt{\frac{P_\mathrm{L}}{P_0}} \frac{\widehat{K} L_\mathrm{u}}{\gamma \sigma_r} \cdot m_0 c^2 \exp\left(-\frac{r^2}{4\sigma_r^2}\right), \tag{2.53}$$

where $\widehat{K}$ is the modified undulator strength parameter from Eq. (2.44), L_u is the undulator length, P_L is the peak laser power, and $P_0 = I_\mathrm{A} m_0 c^2/e \approx 8.7\,\mathrm{GW}$ with the Alfvén current $I_\mathrm{A} \approx 17\,\mathrm{kA}$. The energy modulation depends on various factors, including the laser wavelength, the length and number of periods in the undulator, laser power, and the laser beam's waist size.

To convert the energy modulation into a density modulation, which results in the formation of microbunches, a dispersive section, typically a magnetic chicane, is employed. The chicane introduces a sub-wavelength path length difference between electrons with different energies. A standard chicane configuration consists of four dipole magnets, as illustrated in Fig. 2.9. The strength of a chicane is characterized by the transfer matrix element R_{56} [27] which quantifies the relationship between the path length difference Δz, and the relative energy deviation $\Delta E/E_0$ given by

$$\Delta z = R_{56}\frac{\Delta E}{E_0}. \tag{2.54}$$

By tuning the dispersion appropriately, the sinusoidal energy modulation is converted to density spikes with a spacing corresponding to the laser wavelength λ_L. As a result, a series of microbunches is created, which produces coherent emission at the laser wavelength and its higher harmonics [11].

The bunching factor $b(\omega)$ given by

$$b(\omega) = \frac{1}{N}\left|\sum_{k=1}^{N}\exp\left(-i\frac{\omega z_k}{c}\right)\right|, \tag{2.55}$$

is introduced to quantify the coherently emitted radiation power. The bunching factor accounts for the phase differences of individual electrons within the bunch and depends on the longitudinal position z_k of each electron. For a large ensemble of N randomly distributed electrons, the expectation value of the bunching factor is given by $b = 1/\sqrt{N}$ while the maximum theoretical value is 1, which corresponds to perfect phase alignment.

The incoherent emission power from the entire bunch, containing N_b electrons, is the sum of the power from individual electrons given by

$$P_\mathrm{inc}(\omega)\mathrm{d}\omega = N_\mathrm{b} P_e(\omega)\mathrm{d}\omega. \tag{2.56}$$

where $P_e(\omega)$ is the power spectrum from a single electron. This incoherently emitted power is proportional to the number of electrons in the bunch. The coherently emitted power from N_s electrons within an energy-modulated slice is proportional to the square of the number of electrons and the bunching factor $b(\omega)$. This is given by

$$P_{coh}(\omega)\mathrm{d}\omega = P_e(\omega)\mathrm{d}\omega N_\mathrm{s}^2 b^2(\omega). \tag{2.57}$$

The ratio of the coherent power from the slice to the incoherent power from the entire bunch is then given by

$$\frac{P_e(\omega)\mathrm{d}\omega N_\mathrm{s}^2 b^2(\omega)}{P_e(\omega)\mathrm{d}\omega N_\mathrm{b}} = \frac{N_\mathrm{s}^2}{N_\mathrm{b}} b^2(\omega). \tag{2.58}$$

For typical conditions where the electron bunch is on the order of tens of ps and the seed pulse and modulated slice are a few tens of fs, with an assumption of $N_b \approx 10^{10}$ and $N_s \approx 10^7$ along with a bunching factor of 0.1, the ratio of coherent to incoherent emission becomes

$$\frac{N_s^2}{N_b} b^2(\omega) \approx \frac{(10^7)^2}{10^{10}} \times (0.1)^2 = 10^2. \tag{2.59}$$

This shows that the coherent emission from a small fraction of the electrons can be significantly stronger than the incoherent emission from the entire bunch, by a factor of 100 in this example. The incoherent background produced by the rest of the bunch is negligible and does not contribute significantly to the overall signal.

The CHG technique was first demonstrated at the ACO (Anneau de Collisions d'Orsay) storage ring in Orsay, France [14]. Following this initial demonstration, CHG was implemented at several other storage rings to produce short radiation pulses including UVSOR in Okazaki, Japan [15], Elettra in Trieste, Italy [16] and DELTA in Dortmund, Germany [18].

One major limitation of the CHG technique is the difficulty of reaching high harmonics. Expressing the energy modulation amplitude and the longitudinal dispersion as dimensionless quantities

$$A \equiv \frac{\Delta E}{\sigma_E} \quad \text{and} \quad B \equiv \frac{2\pi\sigma_E R_{56}}{\lambda_\mathrm{L} E_0}, \tag{2.60}$$

with the rms electron energy spread σ_E, the bunching factor for the h-th harmonic of the seed laser can be shown to be [12, 64]

$$|b\,(\lambda_\mathrm{L}/h)| = \exp\left(-\frac{1}{2}h^2B^2\right) J_h(hAB). \tag{2.61}$$

This indicates that the bunching factor decreases exponentially with the harmonic number by $b \propto e^{-h^2}$.

A seeding technique called echo-enabled harmonic generation (EEHG) was proposed by Stupakov [12] to mitigate this limitation and reach shorter wavelengths.

2.6.3 Echo-Enabled Harmonic Generation (EEHG)

Echo-Enabled Harmonic Generation (EEHG) was proposed as a seeding scheme for FELs, but can also be employed for short-pulse generation in storage rings. This technique employs

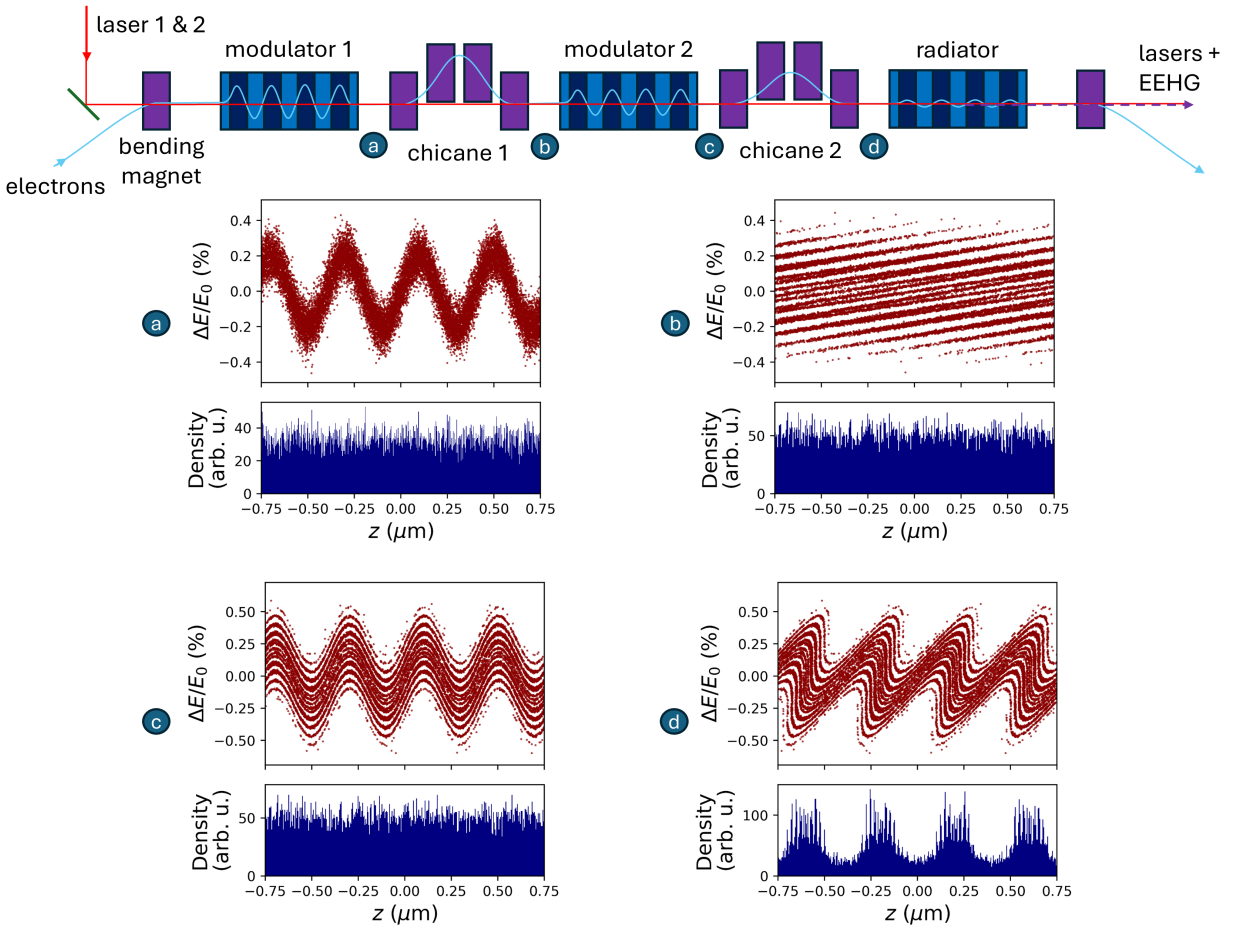

Figure 2.10: Top: Schematic illustration of the EEHG setup, with two seed lasers, two modulators, two magnetic chicanes, and a radiator. (a) A sinusoidal energy modulation is imposed on the electron beam. (b) The first chicane, characterized by a large R_{56} value, significantly shears the electron distribution in phase space. (c) A second laser interaction introduces another energy modulation to the electrons. (d) The second chicane transforms this energy-modulated distribution into a density modulation.

a two-stage process of energy modulation, utilizing two modulators and two corresponding chicanes. After the energy modulation in the initial modulator (Fig. 2.10(a)), the electron beam passes through the first chicane, which is characterized by a large value of R_{56}. This chicane induces significant shear in the longitudinal phase space distribution of the electrons, resulting in the formation of narrow stripes of low local energy spread as seen in Fig. 2.10(b). A second energy modulation is then applied to these stripes in the second modulator, imprinting an additional sinusoidal modulation (Fig. 2.10(c)). Subsequently, the second chicane converts this energy modulation into a density modulation (Fig. 2.10(d)), now with a much richer harmonic content than what can be achieved using CHG alone.

Consider a target harmonic h defined as $h = n + Km$ where $K = \lambda_1/\lambda_2$ represents the ratio

between the wavelengths of the first and second seed laser, and n and m are integers. The theoretical bunching factor achieved by the EEHG scheme is expressed as [12]

$$b_{n,m} = |e^{-\left(\frac{1}{2}[nB_1+(Km+n)B_2]^2\right)} J_m\left(-(Km+n)A_2B_2\right) J_n\left(-A_1[nB_1+(Km+n)B_2]\right)|\,, \tag{2.62}$$

where A_1, A_2 and B_1, B_2 are defined as in Eq. (2.60), representing the parameters for the first and second energy modulation and the first and second chicane, respectively. The optimized scenario is achieved when $|n| = 1$. Considering $n = -1$ and $m > 4$, the bunching factor simplifies to [13]

$$|b^{\text{opt}}_{-1,m}| \approx \frac{0.39}{m^{1/3}}. \tag{2.63}$$

This allows the generation of harmonics higher than CHG, as the optimum bunching factor scales down much more slowly compared to CHG with the target harmonic.

The optimum values of the parameters A_1, A_2, B_1, and B_2 that maximize the bunching factor in Eq. (2.62) at the target harmonic can be estimated analytically. The optimization is performed by first choosing the integers n and m that sum to the target harmonic ($h = n + Km$), and then tuning the machine parameters to maximize the bunching for that specific (n, m) combination. In practice, A_1 and A_2 are constrained by the laser power and its properties, while more flexible parameters are the chicane strengths, which can be tuned to maximize the bunching for given energy modulation amplitudes.

For $m > 4$, the maximum value of the Bessel function J_m is achieved when its argument is equal to $m + 0.81m^{1/3}$ with the value reaching roughly $0.67/m^{1/3}$. So, the term $|J_m\left(-(Km+n)A_2B_2\right)|$ is maximized when [13]

$$(Km+n)A_2B_2 = m + 0.81m^{1/3}. \tag{2.64}$$

From this, B_2 for a given energy modulation A_2 is

$$B_2 = \frac{m + 0.81m^{1/3}}{(Km+n)A_2}. \tag{2.65}$$

For the remaining terms, $|e^{-\left(\frac{1}{2}[nB_1+(Km+n)B_2]^2\right)} \times J_n\left(-A_1[nB_1+(Km+n)B_2]\right)|$, it is convenient to introduce a factor ξ given by

$$\xi = nB_1 + (Km+n)B_2, \tag{2.66}$$

and the term becomes [13]

$$e^{-\xi^2/2} J_n\left(A_1\xi\right). \tag{2.67}$$

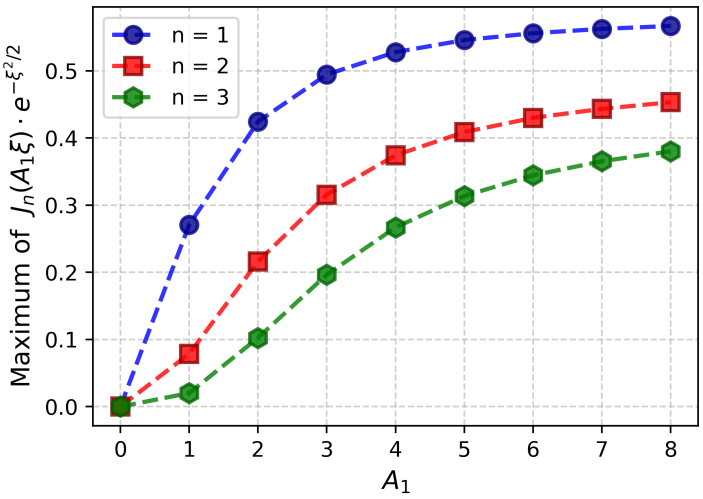

Figure 2.11: Maximum value of $J_n\left(A_1\xi\right)\cdot e^{-\xi^2/2}$ as a function of A_1 shown for three different values of n.

This expression can be numerically calculated for different values of A_1, ξ, and n. Figure 2.11 shows the maximum of Eq. (2.67) for different values of A_1 for 3 orders of the Bessel function. The function shows an asymptotic behavior, reaching close to the maximum by $A_1 \approx 5$. This means that choosing a large first energy modulation ($A_1 > 5$) does not have a significant advantage in improving the bunching factor. The maximum value is achieved for $|n| = 1$, so it is advisable to select the parameters accordingly. A further practical consideration is to select n and m with opposite signs (e.g., $n = -1$ and $m > 0$) such that B_1 and B_2 have the same sign, which leads to a simpler machine layout, such as using two chicanes as dispersion elements. This is why $n = -1$ has been selected for the optimization.

With A_1 fixed, one can find the value of ξ which maximizes the expression 2.67, which is approximately [13]

$$\xi \approx \frac{j'_{n,1}}{A_1}\left[1 + \frac{1}{A_1^2}\frac{1}{1-\left(n/j'_{n,1}\right)}\right]^{-1}, \tag{2.68}$$

where $j'_{n,1} \approx n + 0.81n^{1/3}$ is the argument of the Bessel function corresponding to its maximum. From the optimized B_2 from Eq. (2.65), the optimum value of B_1 can then be calculated as [13]

$$B_1 = \frac{\xi - (Km+n)B_2}{n}. \tag{2.69}$$

This means that the optimum value of the first chicane depends not only on the first energy modulation but also on the second chicane strength and, in effect, the second energy modulation. For setups with limited first chicane strength, such as in DELTA, the second energy modulation becomes a limiting factor since a too-low energy modulation would require a high first chicane strength, which is not achievable by the magnets. Preferably, the working points

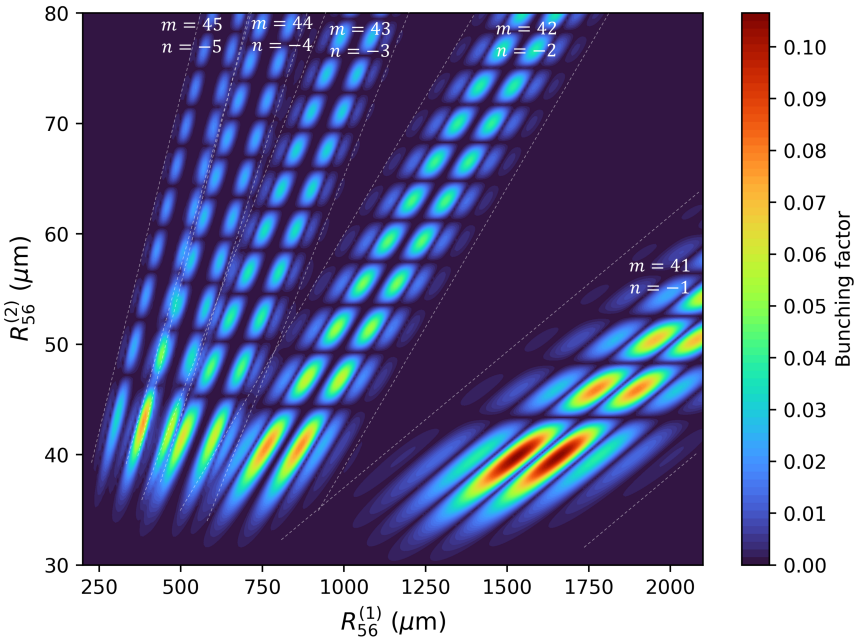

Figure 2.12: Theoretical EEHG bunching map showing the bunching factor at the 40th harmonic for different first ($R_{56}^{(1)}$) and second ($R_{56}^{(2)}$) chicane strengths. Here $K = 1$ and $A1 = A2 = 5$. The maxima corresponding to different combinations of m and n are marked and labeled in the figure.

are chosen such that $n = -1$, which results in the maximum bunching factor, but also requires the highest chicane currents. Using $|n| > 1$ relax the demand on the chicane strength, but result in a reduced bunching factor. Figure 2.12 shows the bunching map for EEHG at the 40th harmonic. It shows the bunching factors corresponding to values of n starting from -1 down to -4. Operating at reduced first chicane strength further complicates things because of different $b_{n,m}$ contributing to the final bunching factor, and it becomes necessary to consider the effect of the phase difference between the two laser pulses.

2.7 The Synchrotron Radiation Source DELTA

DELTA is a 1.5 GeV synchrotron light source operated by the Center for Synchrotron Radiation at TU Dortmund University [65, 66]. The facility serves both as a user-oriented synchrotron radiation source and as a research platform for accelerator physics. Figure 2.13 shows a detailed schematic of the facility.

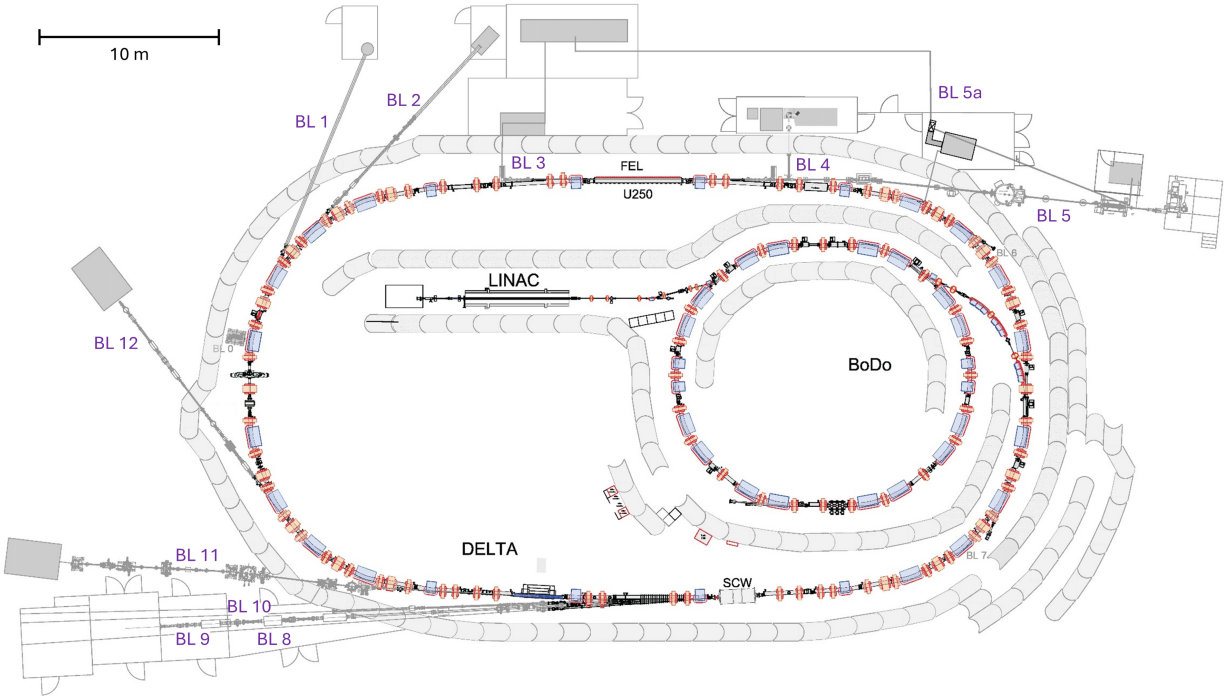

Figure 2.13: Schematics of the DELTA synchrotron radiation source with the electron source (Gun), the linear accelerator (Linac), the booster synchrotron (BoDo) and the storage ring (DELTA) with its beamlines BL0 to BL12. Image credits to Tanja Schulte-Eickhoff.

DELTA consists of an electron gun, a linear accelerator (linac), a booster synchrotron, and a storage ring. Electrons are initially emitted from a thermionic triode gun equipped with a dispenser cathode and accelerated to 90 keV. These low-energy electrons are subsequently bunched by a 3 GHz prebuncher cavity and accelerated to 4 MeV. The main S-band linac then boosts the energy further to approximately 80 MeV using 3 GHz radiofrequency (RF) fields powered by a pulsed klystron, resulting in a bunched beam with 0.33 nanosecond spacing [67].

The electron beam is then transferred to the booster synchrotron BoDo via transfer line T1. The booster features a FODO lattice comprising C-type dipole magnets and alternating quadrupole magnets over a circumference of 50.4 m. Acceleration is achieved through a 500 MHz DESY-type RF cavity, increasing the beam energy to the final value of 1.5 GeV within

approximately 4 seconds. Upon reaching the target energy, electrons are injected into the storage ring through transfer line T2.

The DELTA storage ring has a circumference of 115.2 m and a revolution frequency of about 2.6 MHz. It employs a triplet lattice structure and is equipped with two 500 MHz RF cavities to replenish the energy lost due to synchrotron radiation. The first resonator, a DORIS-type cavity, receives up to 65 kW of forward power from a klystron. In 2019, a second resonator, an EU-type cavity [68] powered by up to 75 kW of forward power from a solid-state amplifier, was added in preparation for the installation of the superconducting wiggler (SCW) [69]. Under typical multibunch operation, about 144 bunches are stored with a total beam current of 140 mA, gradually decreasing over a beam lifetime of over 50 hours. A single-bunch mode with beam currents up to 20 mA is used for specialized experiments, while a hybrid filling pattern enables the short-pulse experiments while in user operation.

Synchrotron radiation for user beamlines is produced by the dipole magnets in the storage ring, which supply beamlines BL 1, 2, and 12. There are three insertion devices accommodated in the straight sections of the storage ring. Hard X-rays for beamlines BL 8, 9, and 10 are generated by a superconducting wiggler (SCW), while soft X-rays are provided by a permanent-magnet undulator U55 to BL 11. The northern straight section accommodating the electromagnetic undulator U250 was originally designed for the low-gain free-electron laser experiment FELICITA [70]. Since 2011, it has been repurposed for the DELTA short-pulse facility. The main parameters of the storage ring are summarized in Table 2.1 [71].

Parameter	**Value**
Nominal beam energy	1.5 GeV
RF frequency	499.83 MHz
Circumference	115.2 m
Revolution frequency	2.60 MHz
Revolution time	384.1 ns
Bunch length	80 ps FWHM
Multibunch current	140 mA (max.)
Single-bunch current	10 to 15 mA (typ.)
Momentum compaction factor α	$5 \cdot 10^{-3}$
Horizontal beam emittance	15 nm rad
Vertical beam emittance	1.5 nm rad
Relative energy spread	$6.8 \cdot 10^{-4}$

Table 2.1: Main parameters of the DELTA storage ring.

2.8 Short-Pulse Facility at DELTA

In 2011, a CHG-based short-pulse facility was established at DELTA [18]. This facility generates ultrashort pulses in both the vacuum-ultraviolet (VUV) and THz wavelength regimes. The installation consists of several key components: a Ti:sapphire laser system, a laser radiation transport beamline (BL 3) [72], the modified undulator U250, a diagnostics beamline (BL 4), a dedicated THz beamline (BL 5a) [73], a pump-pulse beamline [74], and a beamline for pump-probe experiments (BL 5) operated by Forschungszentrum Jülich [75]. A schematic overview of the short-pulse facility is given in Fig. 2.14.

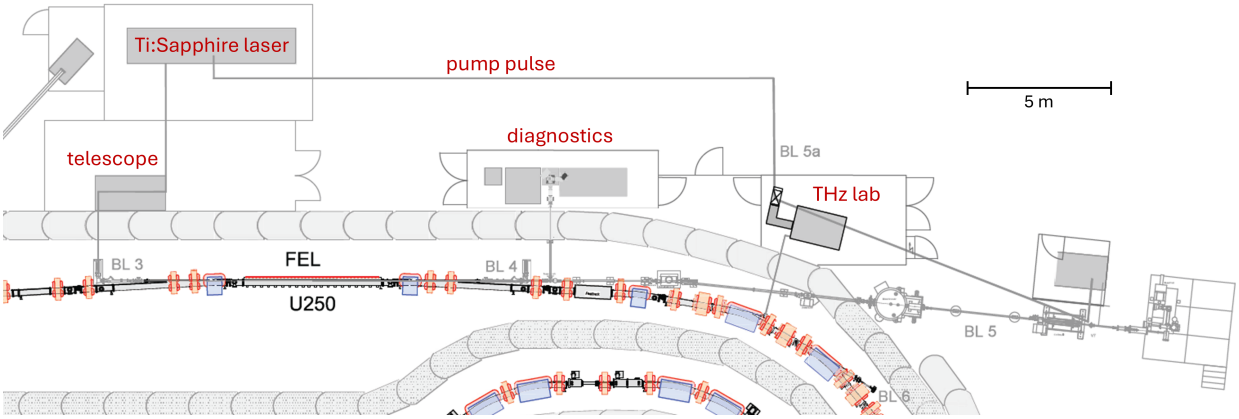

Figure 2.14: Sketch of the short-pulse facility at DELTA. The laser pulse is guided to the undulator U250 through BL 3 and the diagnostics beamline (BL 4) and the VUV beamline (BL 5) allow the detection of VUV radiation. The dedicated beamline (BL 5a) is used for THz radiation. For pump-probe experiments at BL 5, a portion of the laser pulse is used for pumping the sample. Image credits to Tanja Schulte-Eickhoff.

The femtosecond laser system delivers 800 nm radiation pulses with 8 mJ pulse energy and a minimum pulse duration of 40 fs at 1 kHz repetition rate. A second-harmonic generation (SHG) and third-harmonic generation (THG) unit converts the laser radiation to 400 nm or 267 nm wavelengths, producing pulse energies of 2.5 mJ and 1.0 mJ, respectively. After frequency conversion, approximately 2.5 mJ of residual 800 nm laser power remains available.

The laser repetition rate is synchronized with the 6th subharmonic of the DELTA RF frequency (83.33 MHz), ensuring sub-ps synchronization between laser pulse and electron bunch arrival. An RF phase shifter (vector modulator) provides ps-level control of the delay between the laser pulse and the electron bunch to achieve longitudinal overlap. Transverse overlap is accomplished using two remotely controlled mirrors positioned at the end of beamline BL 3. This dual-mirror configuration enables independent control of both the position and angle of the laser beams, ensuring collinear propagation with the electron beam for optimal interaction. A telescope system consisting of three lenses focuses the 800 nm laser beam into undulator U250, while a mirror telescope is used to focus the 400 nm beam [72].

The undulator U250 was originally operated as a storage ring FEL [76] but was subsequently reconfigured for CHG operation to achieve larger R_{56} values for the chicane [77]. In this modified arrangement, the first seven periods function as the modulator, the final seven periods serve as the radiator, and the three intermediate periods constitute the chicane section. The chicane has a maximum R_{56} value of 130 μm. The generated CHG radiation is subsequently directed to either BL 4 or BL 5.

Downstream of the U250, the energy modulation of electrons combined with dispersive effects from the subsequent magnets creates a sub-picosecond density dip in the longitudinal electron distribution. This dip produces coherent emission of THz radiation, which is extracted through beamline BL 5a [42].

BL 4 serves as a diagnostic beamline for CHG operation. This beamline facilitates diagnostics of the laser-electron interaction, optimization of laser-electron overlap, and characterization of the generated CHG radiation. The rough longitudinal overlap is found by monitoring photodiode signals induced by both laser pulses and synchrotron radiation from the electron bunch. Fine adjustment of the longitudinal overlap is then performed using a streak camera with approximately 2 ps time resolution [78].

Transverse overlap optimization employs a systematic scanning procedure where the angles of the two in-vacuum mirrors in BL 3 are adjusted in small steps. While separate scans are conducted for each mirror, the procedure typically begins with the second mirror to locate the overlap region. The quality of the laser-electron overlap is monitored through the THz signal intensity from BL 5a, where a higher THz signal indicates improved laser-electron overlap and consequently enhanced energy modulation. Once a measurable THz signal is established, optimization proceeds through a scanning procedure where both mirrors are moved in synchronized steps to explore different angle combinations. Figure 2.15 shows the result of such a scan, plotting the THz signal intensity as a function of the horizontal angles of the two in-vacuum BL 3 mirrors. This method involves scanning the first mirror across a range for several fixed positions of the second mirror. This 2D map allows for the identification of the optimal angles that produce the maximum THz signal and thus the best transverse overlap. This process is repeated independently for each transverse plane to achieve optimal overlap conditions.

The CHG radiation can be characterized using a Czerny-Turner spectrometer [79] (Acton Research Corporation SpectraPro-275) equipped with an image-intensified CCD camera (Andor iStar CCD [80]). The camera features a gate width as short as 2 ns, enabling selective recording of only those pulses of undulator radiation at which laser-electron overlap occurs. This gating capability is essential because laser interaction occurs only once per millisecond (1 kHz) while the pulse rate in single bunch mode is 2.6 MHz, making it crucial to filter out background synchrotron radiation from all other undulator pulses. The Czerny-Turner spectrometer provides spectral coverage down to 200 nm wavelength. For shorter wavelengths,

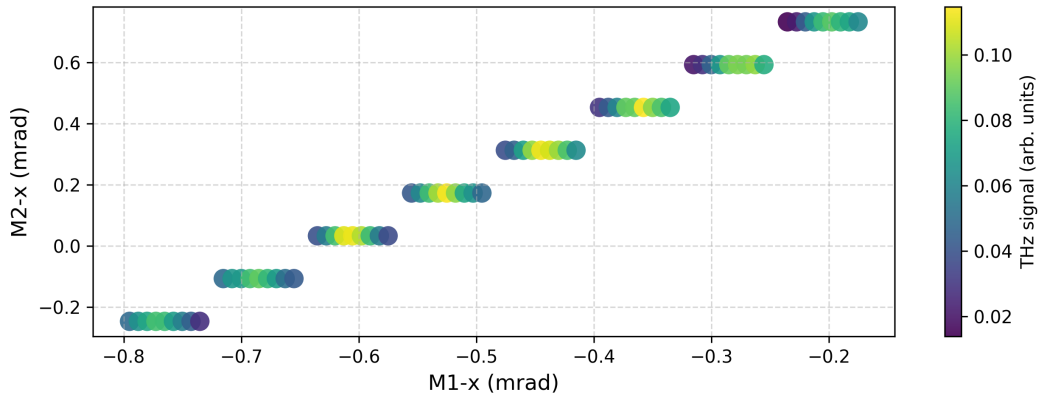

Figure 2.15: An example of transverse overlap optimization. The THz signal intensity (color scale) is mapped as a function of the horizontal angles of the in-vacuum BL 3 mirrors M1 and M2.

an in-vacuum grating spectrometer (HP Spectroscopy easyLIGHT XUV) employing a microchannel plate (MCP) imaged by a CCD camera extends the accessible range to 30-210 nm [81]. The MCP features a switchable gate width (minimum of 200 ns) that is synchronized with the arrival time of the modulated electron bunch, effectively rejecting radiation from other undulator pulses.

2.9 Introduction to Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational models inspired by the structure and function of biological neural systems, where information is processed through interconnected neurons. The origins of this idea can be traced back to the seminal work of McCulloch and Pitts (1943) [82], who proposed the first formal neuron model, followed by Rosenblatt's perceptron (1958) [83], which demonstrated the ability to learn linearly separable patterns. However, the limitations of single-layer perceptrons, as highlighted by Minsky and Papert (1969) [84], led to a significant decline in neural network research. Interest in the field was rekindled in the 1980s with the introduction of the backpropagation algorithm by Rumelhart, Hinton, and Williams [85], which made it possible to train multilayer networks effectively. Around the same period, John J. Hopfield introduced Hopfield networks (1982) [86], a type of network that could store patterns and recall them from partial or noisy information, offering a simple mathematical model of associative memory. This breakthrough, combined with the rise of large datasets and powerful computational resources in the 2000s, laid the foundation for the modern deep learning era. In recognition of their pioneering contributions, John J. Hopfield and Geoffrey E. Hinton were awarded the 2024 Nobel Prize in Physics, highlighting the profound scientific and interdisciplinary significance of neural network research.

The following sections outline foundational concepts of deep learning and neural networks, based on standard textbooks such as [87, 88].

2.9.1 The Perceptron

The perceptron represents the simplest form of a neural network, consisting of a single artificial neuron. Mathematically, a perceptron computes a weighted sum of its inputs and applies an activation function to produce an output

$$a = \sigma\left(\mathbf{w}^T\mathbf{x} + b\right), \tag{2.70}$$

where $\mathbf{x}$ is the input vector, $\mathbf{w}$ denotes the corresponding weight vector, b denotes the bias term, and σ is a nonlinear activation function. Common activation functions include

- **Sigmoid**: $\sigma(z) = \frac{1}{1+e^{-z}}$
- **Tanh**: $\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$
- **ReLU** (Rectified Linear Unit): $\sigma(z) = \max(0, z)$

The activation function introduces nonlinearity, allowing the network to learn complex nonlinear mappings beyond linear transformations. For instance, without a nonlinear activation,

a multi-layer network would mathematically collapse into a single linear model, incapable of capturing intricate relationships in the data.

2.9.2 Feedforward Neural Networks

A feedforward neural network (FNN), also known as a multilayer perceptron (MLP) [85], extends the basic perceptron by introducing multiple neurons and hidden layers between the input and output layers. This architectural enhancement enables the network to learn non-linear mappings through the composition of multiple linear transformations and non-linear activation functions.

Each layer in an FNN performs a linear transformation followed by a non-linear activation function. Formally, the transformation at layer l is given by

$$\mathbf{a}^{(l)} = \sigma^{(l)}\left(\mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}\right) \tag{2.71}$$

for layers $l = 1, 2, ..., L$, with L being the total number of layers, where $\mathbf{a}^{(0)} = \mathbf{x}$ is the input, $\mathbf{W}^{(l)}$ is the weight matrix, $\mathbf{b}^{(l)}$ is the bias vector, and $\sigma^{(l)}$ is the activation function for layer l.

This structure allows FNNs to approximate complex functions. According to the Universal Approximation Theorem [89], a feedforward network with a single hidden layer containing a sufficient number of neurons can approximate any continuous function. In practice, deeper architectures with multiple hidden layers are preferred due to their representational efficiency and improved generalization.

2.9.3 Training a Neural Network

Training a neural network involves adjusting its parameters—weights and biases—to minimize a loss function that quantifies the discrepancy between predicted outputs and true labels. Here, the term 'loss function' refers to the mathematical formula used to compute the loss, while 'loss' refers to the actual value computed for a given dataset or batch. Some commonly used loss functions include the mean squared error (MSE) [90],

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{2.72}$$

used for regression tasks and binary cross-entropy loss for binary classification described by [91]

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]. \tag{2.73}$$

Here, $\hat{y}$ and y denote the predicted outputs and true labels, respectively, and N is the total number of samples.

To minimize the loss, its gradient with respect to each parameter, $\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}}$ needs to be computed, which is done using the backpropagation algorithm [85], which applies the chain rule in reverse mode

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}} = \frac{\partial \mathcal{L}}{\partial \mathbf{a}^{(L)}} \frac{\partial \mathbf{a}^{(L)}}{\partial \mathbf{a}^{(L-1)}} \cdot \cdots \cdot \frac{\partial \mathbf{a}^{(l+1)}}{\partial \mathbf{a}^{(l)}} \frac{\partial \mathbf{a}^{(l)}}{\partial \mathbf{W}^{(l)}}, \tag{2.74}$$

and similarly for $\frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}}$. This algorithm works backward from the final layer, calculating the gradient of the loss with respect to the output and then propagating this error information layer by layer to compute the gradients for all parameters.

Training proceeds in two computational phases: the forward pass, where predictions are generated, and the backward pass, where gradients are computed for parameter updates.

Forward Pass Algorithm: The forward pass computes the output of the network by applying a sequence of linear transformations and nonlinear activation functions from the input to the output layer. For a fully connected neural network:

1. Initialize input activations: $\mathbf{a}^{(0)} = \mathbf{x}$
2. For each layer $l = 1, 2, \ldots, L$:

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)} \tag{2.75}$$

$$\mathbf{a}^{(l)} = \sigma^{(l)}(\mathbf{z}^{(l)}) \tag{2.76}$$

3. Compute the loss: $\mathcal{L}(\mathbf{a}^{(L)}, \mathbf{y})$, where $\mathbf{y}$ is the target output.

Intermediate values $\mathbf{z}^{(l)}$ and $\mathbf{a}^{(l)}$ are stored during this process for use in the backward pass.

Backward Pass Algorithm: The backward pass computes gradients of the loss function with respect to weights and biases using the chain rule. It propagates the errors from the output layer back to the input layer.

1. Compute the error at the output layer:

$$\boldsymbol{\delta}^{(L)} = \frac{\partial \mathcal{L}}{\partial \mathbf{z}^{(L)}} = \frac{\partial \mathcal{L}}{\partial \mathbf{a}^{(L)}} \odot \sigma'^{(L)}(\mathbf{z}^{(L)}), \tag{2.77}$$

where $\odot$ denotes element-wise multiplication and σ' is the derivative of the activation function.

2. Compute the error for all other layers $l = L-1, L-2, \ldots, 1$:

$$\boldsymbol{\delta}^{(l)} = \left(\mathbf{W}^{(l+1)}\right)^T \boldsymbol{\delta}^{(l+1)} \odot \sigma'^{(l)}(\mathbf{z}^{(l)}). \tag{2.78}$$

3. Compute gradients of the loss with respect to parameters:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}} = \boldsymbol{\delta}^{(l)} \left(\mathbf{a}^{(l-1)}\right)^T, \tag{2.79}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}} = \boldsymbol{\delta}^{(l)}. \tag{2.80}$$

The computed gradients are used to update the network parameters using an optimization algorithm.

In theory, the gradients could be computed over the entire dataset, but this is computationally expensive. The common practice is mini-batch gradient descent, where the training set is split into small, random subsets called mini-batches. The forward and backward passes are performed on this mini-batch, and the resulting average gradients $\frac{\partial \mathcal{L}_{\text{batch}}}{\partial \mathbf{W}^{(l)}}$ are used to update the parameters. At iteration t, the weights and biases are updated as

$$\mathbf{W}^{(l)}(t+1) = \mathbf{W}^{(l)}(t) - \eta \frac{\partial \mathcal{L}_{\text{batch}}}{\partial \mathbf{W}^{(l)}}, \tag{2.81}$$

$$\mathbf{b}^{(l)}(t+1) = \mathbf{b}^{(l)}(t) - \eta \frac{\partial \mathcal{L}_{\text{batch}}}{\partial \mathbf{b}^{(l)}}, \tag{2.82}$$

where η is the learning rate, typically ranging between 0.1 and 1×10^{-5}, and $\frac{\partial \mathcal{L}_{\text{batch}}}{\partial \mathbf{W}^{(l)}}$ and $\frac{\partial \mathcal{L}_{\text{batch}}}{\partial \mathbf{b}^{(l)}}$ are the average gradients over the current mini-batch.

While this version of gradient descent is the foundation, more advanced adaptive methods like RMSprop or Adam [92] are commonly used in practice as they often lead to faster convergence.

2.9.4 Batch Normalization

Training very deep neural networks is complicated by the fact that the parameters in each layer are updated during training. As a result, the distribution of the inputs to a given layer continuously changes as training progresses. This phenomenon, termed internal covariate shift [93], impedes training by forcing the network to adapt to these shifting distributions. Batch Normalization (BN) [93] is a technique designed to mitigate this issue by stabilizing the input distributions to each layer. It operates by normalizing the layer inputs using the statistics of the mini-batch of data. For a mini-batch $\mathcal{B} = \{x_1, \ldots, x_m\}$ of size m, the BN

transformation is applied to the layer's pre-activation inputs x_i as

$$\hat{x}_i = \frac{x_i - \boldsymbol{\mu}_{\mathcal{B}}}{\sqrt{\boldsymbol{\sigma}_{\mathcal{B}}^2 + \epsilon}} \tag{2.83}$$

where the mean $\boldsymbol{\mu}_{\mathcal{B}}$ and variance $\boldsymbol{\sigma}_{\mathcal{B}}^2$ of the mini-batch are defined as

$$\boldsymbol{\mu}_{\mathcal{B}} = \frac{1}{m}\sum_{i=1}^{m} x_i \ , \qquad \boldsymbol{\sigma}_{\mathcal{B}}^2 = \frac{1}{m}\sum_{i}^{m}(x_i - \boldsymbol{\mu}_{\mathcal{B}})^2. \tag{2.84}$$

Here, ϵ is a small constant added for numerical stability to prevent division by zero. This normalization step ensures the inputs to the subsequent activation function are in a consistent range, which helps stabilize the learning process.

However, enforcing a zero mean and unit variance can limit the network's representational power. To restore this flexibility, BN introduces a scale factor $\boldsymbol{\gamma}$ and a shift factor $\boldsymbol{\beta}$:

$$y_i = \boldsymbol{\gamma}\hat{x}_i + \boldsymbol{\beta}. \tag{2.85}$$

These parameters are learned during optimization, allowing the network itself to determine the optimal mean and variance for each layer's inputs.

During inference (when the model is used for prediction), inputs are often processed one at a time, so there is no mini-batch from which to calculate a mean and variance. In this case, the model uses statistics that were estimated during the training process. A common practice is to calculate an exponential moving average of $\boldsymbol{\mu}_{\mathcal{B}}$ and $\boldsymbol{\sigma}_{\mathcal{B}}^2$ as training progresses. After training is complete, these final averages are stored. These fixed values then serve as the mean and variance to normalize the new input data the model receives.

2.9.5 Regularization

Neural networks, particularly deep architectures, are susceptible to overfitting due to their high capacity and ability to memorize training data. Overfitting occurs when a model learns to perform exceptionally well on training samples it has encountered repeatedly during training, but fails to generalize to new, unseen data. In this scenario, the network essentially memorizes specific patterns and noise present in the training set rather than learning the underlying generalizable relationships. This leads to poor performance on validation and test datasets, when the model encounters data it has never seen before. Regularization techniques are essential for improving generalization performance, preventing models from fitting noise in the training data and encouraging it to learn more robust, generalizable representations rather than memorizing training-specific artifacts.

L2 Regularization (Weight Decay): L2 regularization discourages large weights by adding a penalty term to the loss function proportional to the sum of the squared weights [94]. This ensures that the penalty increases with weight magnitude and remains smooth for gradient-based optimization. The modified loss function $\mathcal{L}_{\text{reg}}$ becomes

$$\mathcal{L}_{\text{reg}} = \mathcal{L}_{\text{o}} + \lambda \sum_i w_i^2, \tag{2.86}$$

where $\mathcal{L}_{\text{o}}$ is the original loss, λ is the regularization strength and w_i denotes all the individual weights in the network. This technique, also known as weight decay, prevents the weights from growing too large, promoting simpler and more robust models.

Dropout: Dropout [95] is a regularization technique where, during training, a subset of neurons is randomly "dropped" (i.e., set to zero) with a given probability p. The activation of a single neuron in a layer after dropout is given by

$$a_i^{(l)} = \begin{cases} 0 & \text{with probability } p, \\ a_i^{(l)} & \text{with probability } (1-p). \end{cases} \tag{2.87}$$

During training, dropout effectively trains an ensemble of different network architectures, as each forward pass uses a different subset of neurons. At test time, all neurons are active, but their outputs are scaled by $(1-p)$ to maintain the expected output magnitude. This regularization technique prevents neurons from relying too heavily on specific other neurons and forces the network to learn more robust representations.

2.9.6 Hyperparameters and Tuning

In contrast to model parameters like weights and biases, which are learned during training, hyperparameters are external configuration settings that are chosen prior to the training process. Examples introduced in the preceding sections include the learning rate η, the L2 regularization strength λ, the dropout probability p, as well as architectural choices like the number of hidden layers and the number of units in each layer.

The ideal values of the hyperparameters depend on the model architecture, the dataset, and the specific task the network is designed for. The process of finding the optimal combination is known as hyperparameter tuning [96]. To conduct this search properly, the dataset is typically split into three sets: a training set to learn the model parameters, a validation set to evaluate the model's performance for a given set of hyperparameters, and a test set to provide an unbiased estimate of the final model's generalization ability.

Common tuning strategies include grid search, which exhaustively tries every combination of a predefined set of hyperparameter values, and random search [97], which samples configurations from a statistical distribution and is often more efficient. More advanced automated methods, such as Bayesian optimization [98], build a probabilistic model to select more promising hyperparameter values based on past results.

2.9.7 Convolutional Neural Networks

Traditional fully connected neural networks face significant challenges when applied to image processing tasks. For a typical 224×224 RGB image with 150,528 pixels, a fully connected network would require millions of parameters even for modest hidden layer sizes, leading to computational complexity and overfitting. Moreover, fully connected networks treat each pixel independently, ignoring crucial spatial relationships and lacking translation invariance (the ability to recognize the same feature regardless of its position in the input).

Convolutional Neural Networks (CNNs) [99] are inspired by the visual cortex by processing data within localized receptive fields [100]. This structure enables the network to learn in a hierarchical fashion, beginning with the identification of low-level features like edges and colors in its initial layers. As data propagates through the network, subsequent layers progressively combine these simple patterns to construct more abstract and complex representations, ultimately allowing for the recognition of whole objects.

Convolutional Layer: The fundamental building block of a CNN is the convolutional layer, where a learnable kernel (or filter) is applied across the spatial dimensions of the input. For a 2D input $\mathbf{X}$ and filter $\mathbf{K}$, the convolution output at position is given by

$$S(i,j) = (\mathbf{X} * \mathbf{K})(i,j) = \sum_m \sum_n \mathbf{X}(i+m, j+n) \cdot \mathbf{K}(m,n), \tag{2.88}$$

where the summation is over the dimensions of the filter. Multiple filters can be used in a single layer to extract different features such as edges, textures, or shapes. The convolution operation captures local patterns through the kernel while maintaining spatial relationships in the output feature map. The kernels are learnable parameters that act as feature detectors. The kernels in early layers detect simple features like edges and textures, while the deeper layers detect complex patterns and semantic features.

Figure 2.16 illustrates an example convolution operation done with an arbitrary input and kernel. Two example feature maps extracted using different kernels are depicted in Fig. 2.17. For the model, the kernels are not explicitly generated, but the model learns the best kernels through optimization algorithms such as gradient descent to extract the relevant information during training.

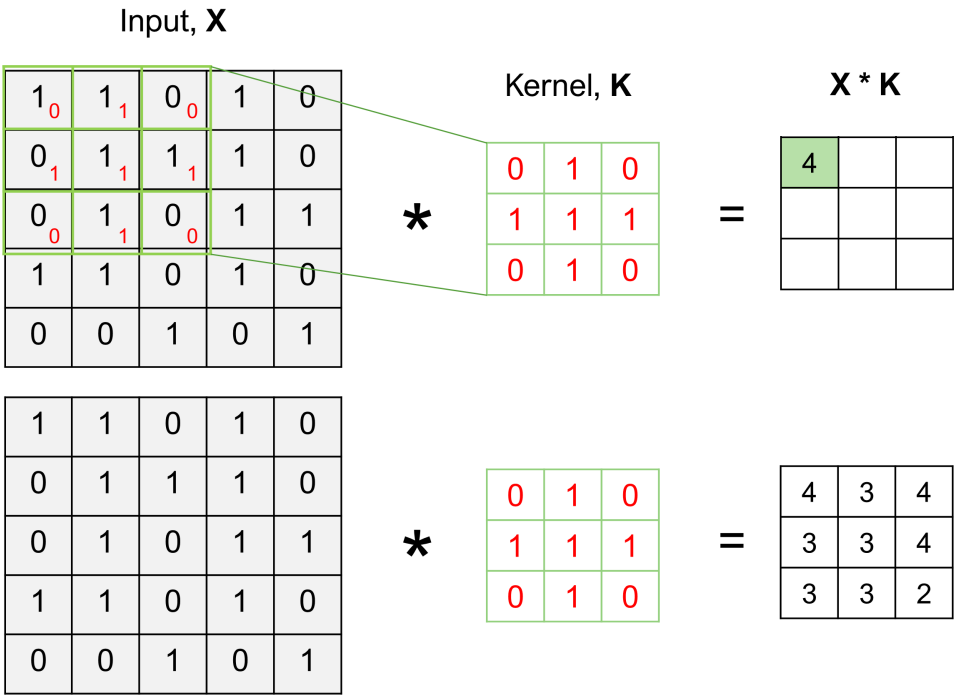

Figure 2.16: Illustration of the 2D convolution operation in a convolutional neural network (CNN). The top row shows how a 3×3 kernel $\mathbf{K}$ is applied to a 3×3 region of the input matrix $\mathbf{X}$. Each element of the kernel is multiplied element-wise with the corresponding input value, and the results are summed to produce a single output value (4 in this example). This value becomes the top-left entry of the resulting feature map. In the lower panel, the kernel slides across the input matrix with a stride of 1, performing the same operation at each step. The complete output feature map is shown on the right. This example uses a binary input and a simple kernel for illustrative clarity. In practice, both input and kernel values are typically real-valued.

Activation and Pooling: After convolution, a nonlinear activation function is applied element-wise to introduce nonlinearity. To reduce the spatial resolution and computational cost, pooling layers are employed. The most common is max pooling, which retains the maximum value in a local neighborhood, given by

$$\text{MaxPool}(i, j) = \max_{m,n} \{\mathbf{X}[i \cdot s + m,\ j \cdot s + n]\} \tag{2.89}$$

where s denotes the stride, which controls the step size with which the pooling window moves across the input.

Padding and Stride: To control the spatial dimensions of feature maps, CNNs use padding and stride. Padding involves adding zeros around the border of the input to preserve its size after convolution. Stride controls the step size of the kernel as it moves across the input, allowing for downsampling.

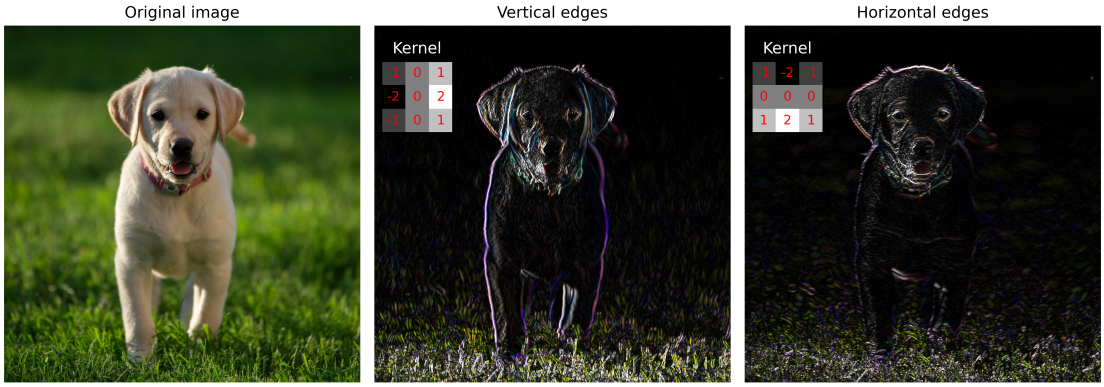

Figure 2.17: Two example kernels applied to an image for feature extraction. The vertical and horizontal edge detection kernels highlight different directional features of the puppy, demonstrating how convolutional filters capture useful visual patterns. While these specific edge-detection kernels can be hand-crafted, a CNN learns the optimal kernel values automatically during training to extract the most relevant features for a given task. The image of the puppy is credited to Ben Owen on Unsplash.

CNN Architecture Typical CNN architectures consist of repeated blocks of convolutional, activation, and pooling layers, followed by fully connected layers. A representative CNN architecture is illustrated in Fig. 2.18. CNNs achieve computational efficiency by sharing weights across spatial locations, which significantly reduces the total parameter count compared to fully connected networks. Additionally, the convolutional structure enables the network to learn features that remain consistent regardless of their position within the input. Popular architectures include LeNet [101], AlexNet [102], VGGNet [103], ResNet [104], U-Net [105], etc.

For detailed descriptions and theoretical foundations of convolutional architectures and deep learning techniques, the reader is referred to comprehensive textbooks such as [87, 88].

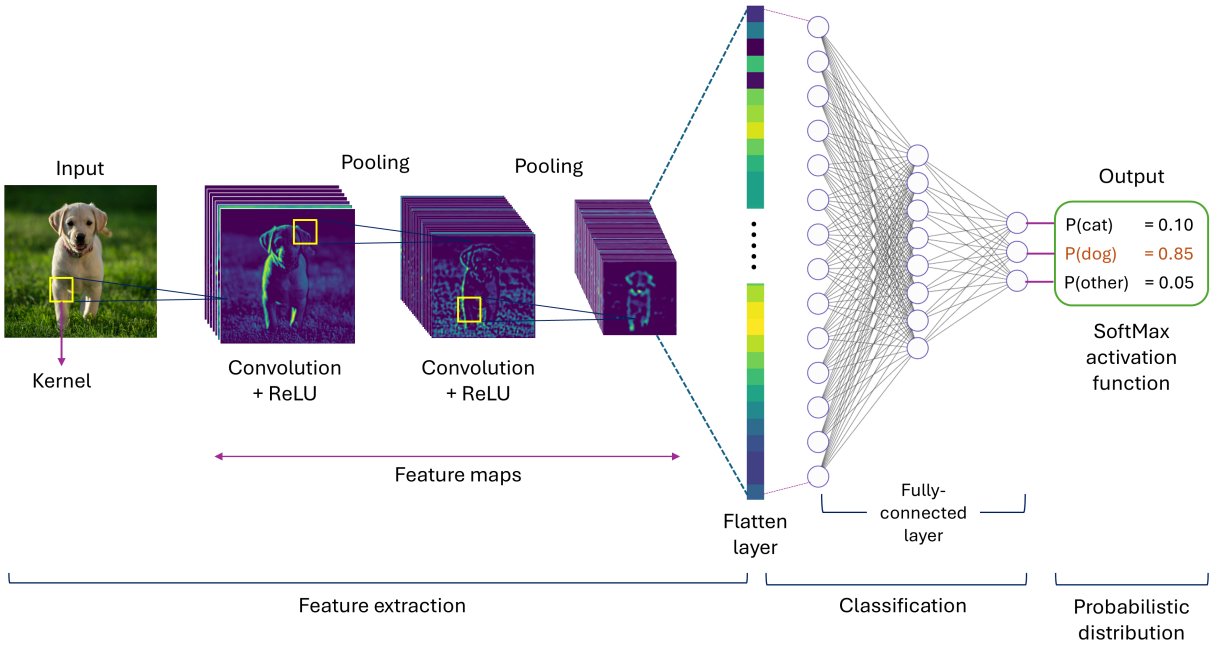

Figure 2.18: Schematic overview of a convolutional neural network (CNN) architecture for image classification. The input image is processed through a series of convolutional layers with ReLU activation and pooling operations, generating hierarchical feature maps. These features are then flattened, which converts the 2D array of features into a one-dimensional vector. This flattened vector is then passed through fully connected layers for classification. The output layer uses a SoftMax activation function to produce a probability distribution over possible classes (e.g., cat, dog, other).

Chapter 3

Analysis of CHG Spectra Using Convolutional Neural Networks

3.1 Spectral Properties of CHG Radiation

To simplify the analysis of CHG radiation, it is convenient to assume a constant electron density and seed pulse intensity. The assumption on electron density is reasonable for electron bunches with a duration of several tens of picoseconds. However, the assumption of a constant seed pulse intensity is a significant simplification. In reality, a Gaussian seed pulse exhibits a spatiotemporal intensity profile following a normal distribution. When the seed pulse waist size is significantly larger than the transverse cross section of the electron bunch, its transverse profile may be approximated as relatively uniform, while its longitudinal intensity profile is approximated as Gaussian. This simplification helps in analyzing the energy modulation within the electron bunch.

The energy modulation amplitude is proportional to the electric field of the seed pulse, which exhibits a Gaussian envelope with a temporal FWHM broadened by a factor of $\sqrt{2}$ relative to that of the seed pulse. To account for the slippage effect, where electrons lag behind the laser pulse by one seed wavelength per undulator period, the modulation profile must be convolved with a rectangular function of length $N_\mathrm{u}\lambda/c$, where N_u is the number of undulator periods, λ is the seed wavelength and c is the velocity of light.

Depending on the value of the chicane's dispersion strength R_{56}, the optimum bunching factor is obtained at different positions along the slice. Figure 3.1 shows a 2D map of the theoretical bunching factor, calculated using Eq. (2.61) at the second harmonic of the seed wavelength. The plot shows the dependence of the bunching factor on both the longitudinal coordinate and the R_{56} value. In this example, a normalized modulation amplitude of $A \equiv \Delta E_\mathrm{max}/\sigma_E = 5$ is assumed.

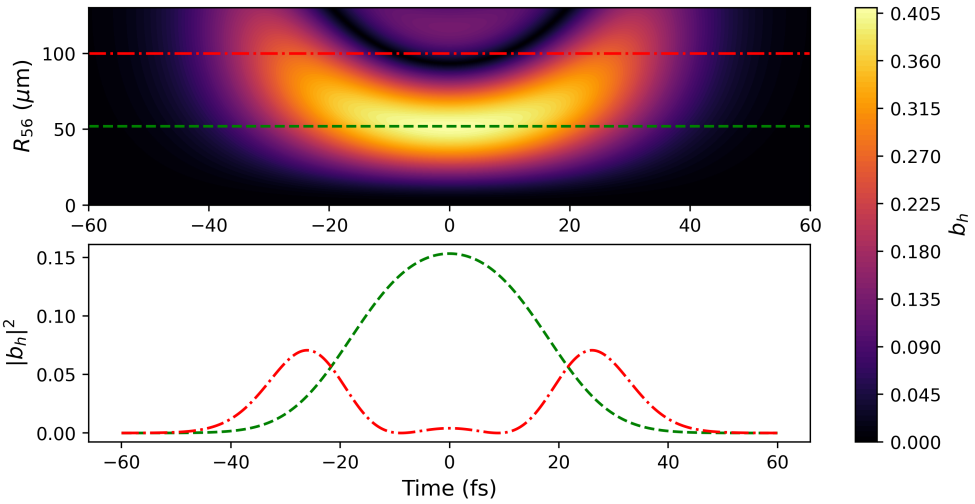

Figure 3.1: Top: 2D map of the theoretical bunching factor as a function of longitudinal position (in time units) and R_{56} value of the chicane between modulator and radiator. Bottom: Squared bunching factors for selected R_{56} values, indicated by the lines in the top plot.

The optimum R_{56} setting as a function of the longitudinal coordinate is approximately parabolic, with the largest bunching factor obtained by optimizing for the central position. Increasing R_{56} beyond this optimum leads to overbunching, where the phase space distribution is sheared past being vertical (parallel to the energy axis). This overbunching results in a reduction of the central peak and the appearance of multiple smaller maxima of the bunching factor along the modulated slice as illustrated in Fig. 3.2. At the optimal setting of $R_{56} = 30$ μm, microbunching is observed at the center of the slice. However, stronger chicanes overshear the distribution, disrupting the central microbunching and creating separated, lower-amplitude bunching envelopes at off-center longitudinal positions.

For a uniform electron density distribution, the intensity of the coherently emitted radiation is proportional to the square of the bunching factor (see Eq.(2.55)). This relationship is displayed in the lower panel of Fig. 3.1, which compares the squared bunching factors for two representative R_{56} values, indicated by the two lines. The green dashed line corresponds to the optimal case, showing a single pronounced peak at the bunch center. In contrast, the red dash-dotted line represents an overbunched case for the central portion of the modulated slice, where the central peak disappears and multiple secondary maxima form.

The sensitivity of the bunching factor to the chicane strength provides a means to manipulate the temporal structure of the coherently emitted pulses. Such control over pulse shaping has practical applications, as demonstrated in the context of High-Gain Harmonic Generation (HGHG) seeding at the FERMI free-electron laser facility in Trieste, Italy [22].

In addition to adjusting the chicane strength, the spectral phase of the seed laser pulse can also be employed to control the spectro-temporal characteristics of the CHG radiation pulses.

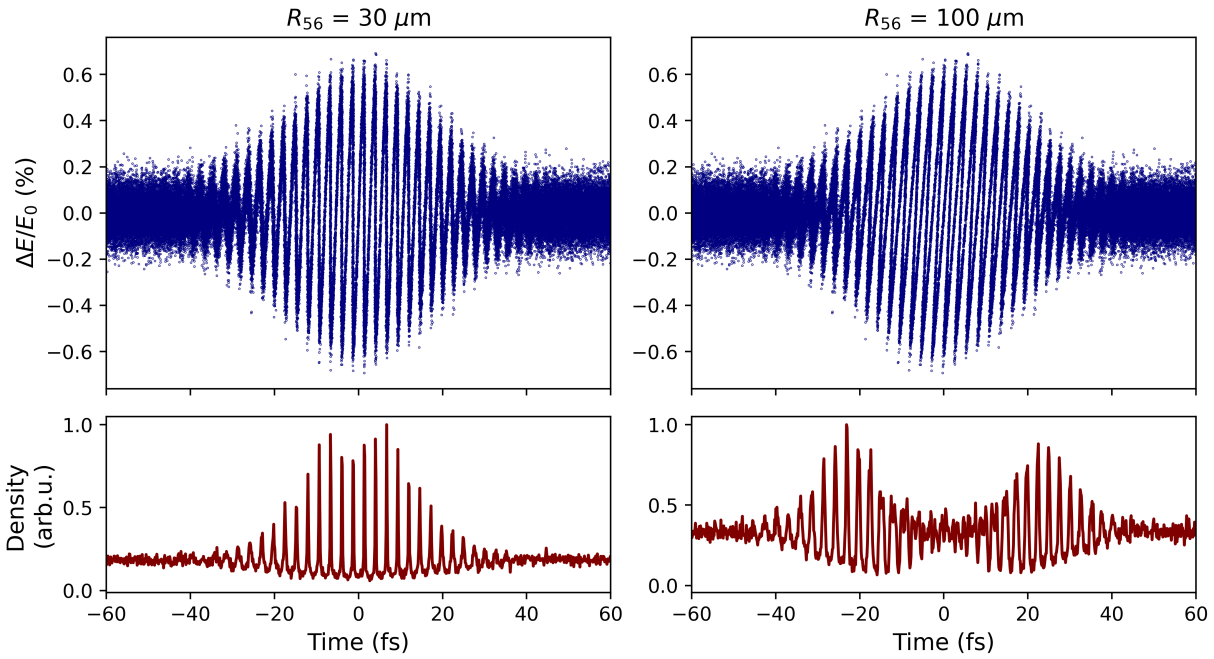

Figure 3.2: Electron phase space distribution (top) and density (bottom) for an energy modulation with a Gaussian pulse after a moderate chicane with R_{56} of 30 μm (left) and with a stronger chicane of R_{56} = 100 μm (right).

This is because the properties of the seed pulse are imprinted onto the energy modulation induced within the electron bunch. A laser pulse with a central frequency ω_0 can be expressed in the frequency domain in terms of its spectral amplitude $\tilde{E}(\omega)$ and spectral phase $\phi(\omega)$ as

$$\tilde{E}(\omega) = |\tilde{E}_0(\omega)|e^{-i\phi(\omega)}, \tag{3.1}$$

where the spectral phase $\phi(\omega)$ is expanded into a Taylor series as

$$\phi(\omega) = D_0 + D_1(\omega - \omega_0) + \frac{1}{2}D_2(\omega - \omega_0)^2 + \frac{1}{6}D_3(\omega - \omega_0)^3 + \cdots . \tag{3.2}$$

In this expansion, D_0 denotes the central phase, D_1 is the group delay, D_2 is the group delay dispersion (GDD), and D_3 corresponds to the third-order dispersion (TOD).

A Fourier-limited laser pulse corresponds to a scenario where the GDD and all higher-order dispersion terms are zero. A non-zero GDD stretches the pulse duration and introduces a linear frequency change (frequency chirp) along the pulse. The presence of a non-zero TOD adds further complexity by creating an asymmetry in the longitudinal pulse shape and generating side pulses that either precede or follow the main pulse.

In a laser system, a Treacy-type pulse compressor [106], commonly employed in chirped-pulse amplification (CPA) schemes [37], can be utilized to control the spectral phase properties of the amplified pulse. The group delay dispersion (GDD) and third-order dispersion (TOD)

are directly related to the parameters of the pulse compressor, as described by [107]

$$D_2 = -4 \cdot \frac{4\pi^2 cd}{\omega_0^3 g^2} \left\{1 - [G - \sin\gamma]^2\right\}^{\frac{3}{2}} \tag{3.3}$$

$$D_3 = -\frac{3D_2}{2\omega_0} \cdot \frac{1 + G\sin\gamma - \sin^2\gamma}{1 - [G - \sin\gamma]^2}, \tag{3.4}$$

where $G = \frac{2\pi c}{\omega_0 g}$, d is the separation between the two gratings, g is the grating constant, γ is the angle of incidence, and c is the speed of light. By adjusting the grating separation d, one can manipulate the GDD and TOD of the compressed laser pulse. However, it is important to note that changing d does not allow for independent control over GDD and TOD.

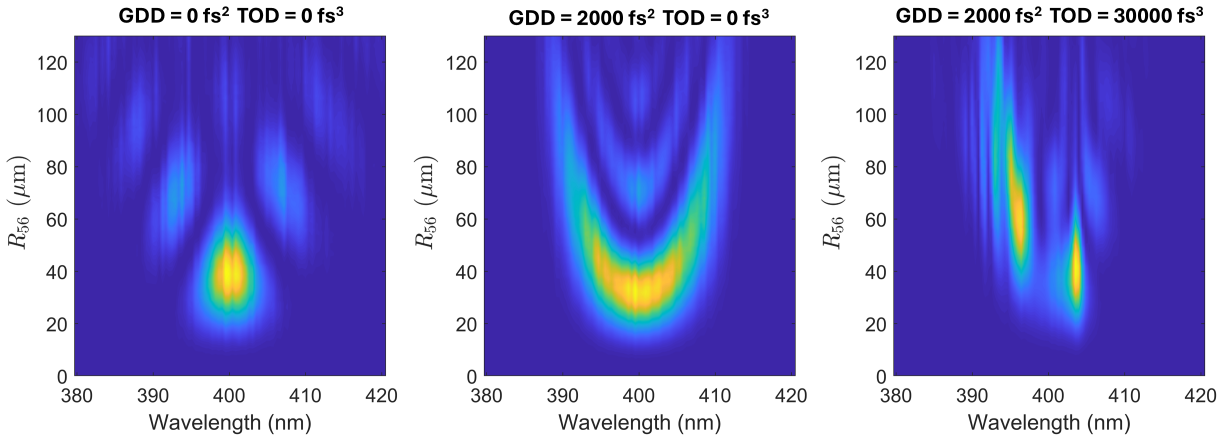

Figure 3.3: Simulated CHG spectral maps for a transform-limited seed pulse with zero frequency chirp (left), for a seed pulse with strong chirp corresponding to a GDD of 2000 fs^2 but with zero TOD (center), and for a seed pulse with large positive GDD and TOD (right).

Figure 3.3 shows the spectra of the CHG radiation for different values of R_{56}, where each row in the image represents the CHG spectrum for a corresponding chicane strength. This type of plot will be referred to as a spectral map in the following discussions. For a transform-limited seed pulse (with zero chirp), the longitudinally separated maxima of the coherently emitted pulses at high R_{56} values interfere with each other if they are of the same frequency, producing interference fringes in the CHG radiation spectra, as shown in Fig. 3.3 (left). Conversely, when the seed pulse is strongly chirped, the two maxima of the pulse produced at large values of R_{56} correspond to different frequencies. This frequency variation results in maxima at these frequencies in the CHG spectra. When only a non-zero group delay dispersion (GDD) is present, the chirped laser pulse generates a CHG spectral map with a symmetric parabolic profile as shown in Fig. 3.3 (center). However, introducing a non-zero third-order dispersion (TOD) results in longitudinal asymmetry of the seed laser pulse, which in turn makes the spectral map asymmetric. This effect is illustrated in Fig. 3.3

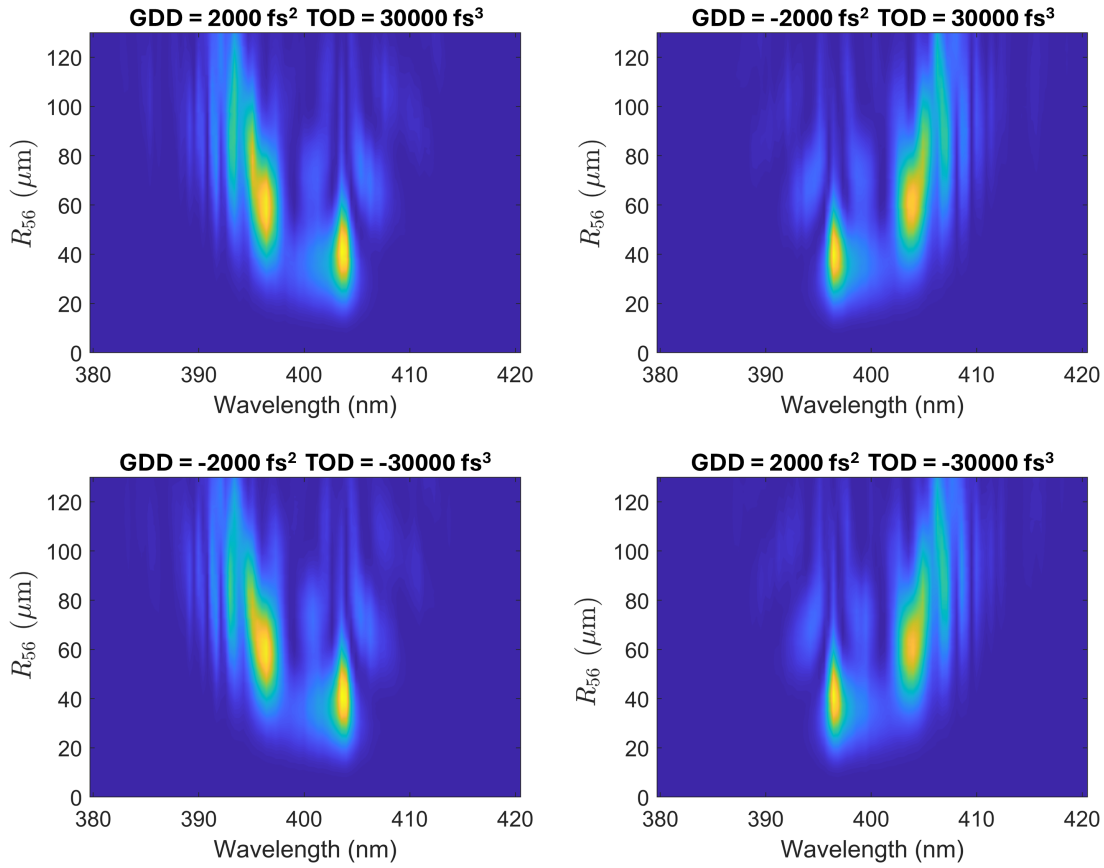

Figure 3.4: Simulated CHG spectral maps for combinations of different signs of GDD and TOD values. The plots in each column are identical irrespective of switching the signs of both GDD and TOD.

(right), where a seed pulse with a GDD of 2000 fs^2 and a TOD of 30000 fs^3 is assumed. The resulting spectral map exhibits a characteristic asymmetric shape with a shifted central peak and several secondary peaks, whose positions depend on the specific values of GDD and TOD. The side on which the secondary peak appears changes when the sign of either the GDD or TOD is reversed. Consequently, the spectral map remains identical when both the GDD and TOD have their signs flipped simultaneously. This is demonstrated in Fig. 3.4, where the spectral maps are shown for different sign combinations of GDD and TOD, but with the same absolute values.

3.2 Simulation of CHG Spectra

For simulating the CHG spectra under different spectral phase conditions, the laser pulse is modeled considering different values of GDD and TOD. A transform-limited pulse duration of $\Delta t_0 = 45$ fs (FWHM) is assumed. The spectral shape of the laser pulse is taken to be Gaussian, centered around a wavelength of 800 nm, which corresponds to an angular frequency of $\omega_0 = 2.35 \times 10^{15}$ s^{-1}. For transform-limited pulses, the time-bandwidth product (TBP) constrains the minimum spectral bandwidth, given by $\Delta t_0 \cdot \Delta\omega_0 \approx 2\pi \cdot 0.441$. This yields an FWHM angular frequency bandwidth of $\Delta\omega_0 = 6.16 \times 10^{13}$ s^{-1}.

In the simulations, the electric field $\mathcal{E}(t)$ is constructed by summing monochromatic components across a frequency range $\omega \in [\omega_0 - \Delta\omega, \omega_0 + \Delta\omega]$, where $\Delta\omega = 5 \times \Delta\omega_0$ s^{-1} was chosen to cover a large frequency range to capture the full spectral content of the pulse, including contributions from wings of the Gaussian spectrum that may influence higher-order phase effects. The field is expressed as

$$\mathcal{E}(t) = \sum_{\omega=\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} \tilde{\mathcal{E}}(\omega) e^{i(\omega t + \phi(\omega))}, \tag{3.5}$$

where $\tilde{\mathcal{E}}(\omega)$ represents the spectral amplitude, and $\phi(\omega)$ denotes the spectral phase (3.2).

To facilitate numerical simulations, the electric field amplitude is normalized to its maximum value, leading to

$$\mathcal{E}(t)_{\text{norm}} = \frac{\mathcal{E}(t)}{\mathcal{E}(t)_{\text{max}}}. \tag{3.6}$$

This normalized field is then used to model the energy modulation of an electron bunch interacting with the laser pulse. A few examples of the laser pulse profile for different combinations of GDD and TOD are shown in Fig. 3.5.

The electron distribution was initialized with 2.4×10^5 electrons representing a bunch slice of length 1200 fs, carrying a current of 100 nA with an energy spread of $\sigma_E/E_0 = 7 \times 10^{-4}$.

The laser-induced energy modulation is modeled by modifying the energy of each electron according to the normalized laser electric field as

$$E_f(t) = E_i(t) + \mathcal{E}(t)_{\text{norm}} \delta E, \tag{3.7}$$

where $E_i(t)$ and $E_f(t)$ are the initial and final electron energies, respectively, and δE is the peak modulation amplitude. This encodes the laser's electric field structure into the electron energy distribution.

The transfer of energy modulation into spatial density modulation was modeled using a dispersive chicane characterized by its longitudinal dispersion parameter R_{56}. The chicane

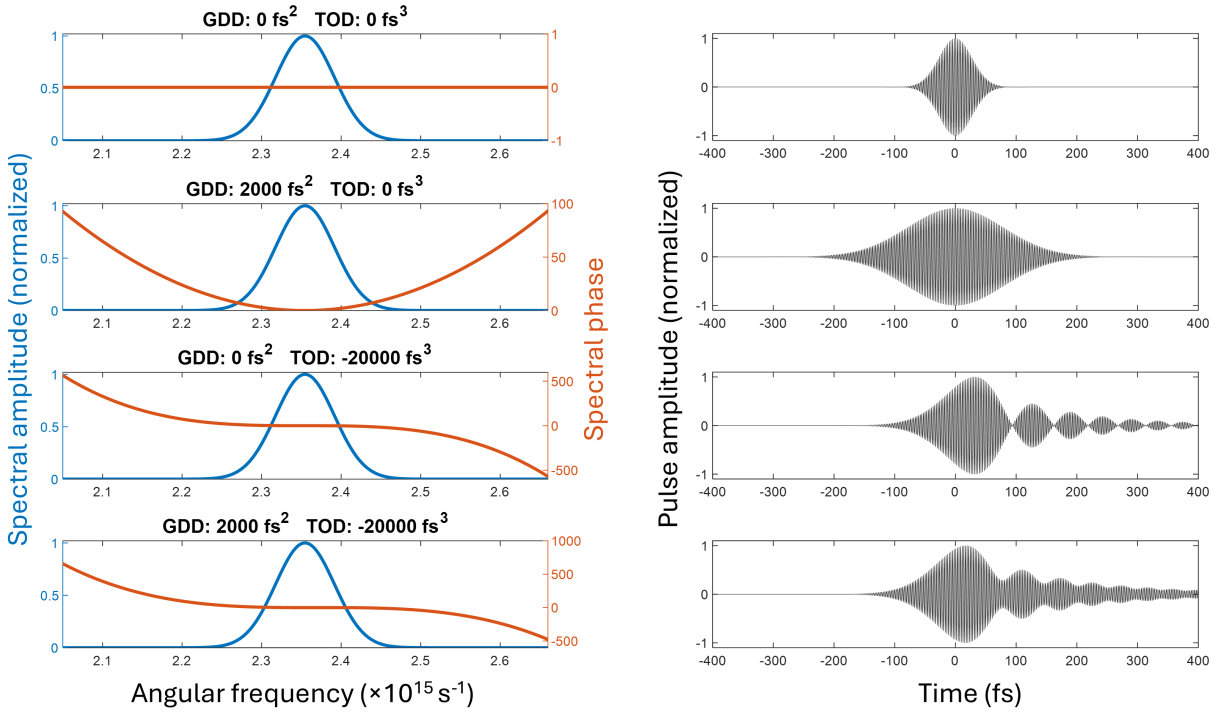

Figure 3.5: Examples of the laser pulse shapes with different combinations of GDD and TOD (see title). The left column shows the spectral amplitude (blue) and spectral phase (red) of the laser pulse, and the right column shows the respective pulse shape in the time domain.

imposes a path length difference dz proportional to the energy deviation

$$dz = R_{56}\frac{\Delta E}{E_0}. \tag{3.8}$$

This redistribution of electrons in longitudinal phase space results in a density modulation. The spectral characteristics of the CHG radiation are determined by evaluating the bunching factor $b(\omega)$ using Eq. (2.55). The spectral intensity of CHG is proportional to $b(\omega)^2$. To generate spectral maps, $b(\omega)^2$ was evaluated over corresponding wavelength ranges for different harmonics and for R_{56} values spanning 0 to 130 μm.

The simulation was implemented in *MATLAB*. The source code can be found on GitHub [1].

[1] https://github.com/arjun-krishnan/CHG-spectra-analyse

3.3 Analysis of CHG Spectra Using Convolutional Neural Networks

The CHG spectra observed in experiments at DELTA exhibit clear asymmetries along the wavelength axis. The primary goal of this work is to understand the physical mechanisms that produce these complex spectral features and to establish a method for their control. The central hypothesis is that these features are a direct result of the seed laser's spectral phase properties, specifically its GDD and TOD. To test this hypothesis, a method was needed to connect the observed spectral shapes with the GDD and TOD values of the seed laser. Conventionally, tools like Frequency-Resolved Optical Gating (FROG) [108] or Spectral Phase Interferometry for Direct Electric-field Reconstruction (SPIDER) [109] can be used to measure the spectral phase and calculate these values, but reliable measurements were not possible during the experiments. This led to the development of a new method that uses the CHG spectrum itself as a non-invasive diagnostic.

The initial strategy involved comparing the observed and simulated spectral maps to identify the GDD and TOD values that minimized the root mean square difference between the images. However, this approach was proven difficult due to the lack of pixel-to-pixel correspondence between the simulated and observed spectral maps. The experimental observations exhibited more complex features influenced by various factors, such as the noise in image acquisition, uncertainties in the energy modulation amplitude, etc. These complexities were difficult to accurately model in simulations, complicating direct comparisons. To address this, a method that could focus on the broader features present in the spectral map rather than relying on a pixel-by-pixel comparison with simulations was required. A convolutional neural network (CNN) [99] is a viable solution, as CNNs can be trained to detect features in images independent of their position. CNNs achieve a high degree of invariance to local translations and deformations, which avoids the need for a one-to-one correspondence between simulation and observation. This capability allows for a more robust comparison by focusing on the essential features of the spectral maps rather than their exact alignment.

3.3.1 Details of the Neural Network

To estimate the GDD and TOD of the seed laser pulse from the observed spectral maps, a CNN was designed and trained using *TensorFlow* [110] in *Python*. The CNN was trained on a dataset of simulated spectral maps of CHG at the 2nd, 3rd, and 4th harmonics of 800 nm, generated for various combinations of GDD and TOD. These spectral maps, which contain CHG spectra in the range of 380 nm to 420 nm for the 2nd harmonic, 255 nm to 275 nm for the 3rd harmonic, and 190 nm to 210 nm for the 4th harmonic and correspond to chicane

strengths ranging from 0 to 130 μm, were simulated using a variation of the *MATLAB* code with different GDD and TOD values [2].

A CNN architecture consisting of two convolutional blocks and three fully-connected layers [87] was used. The network takes 64×64 pixel images as input and outputs two values corresponding to GDD and TOD. These output values are normalized to lie within the range [0, 1].

The convolutional blocks consist of a convolutional layer, followed by a ReLU activation function and a max-pooling layer [87]. The convolutional layers use a 4×4 kernel with a 1×1 stride and zero-padding. The max-pooling layers utilize a 2×2 kernel with zero-padding. The three fully-connected layers have 800, 400, and 128 units, respectively. Figure 3.6 presents the schematics of the CNN used.

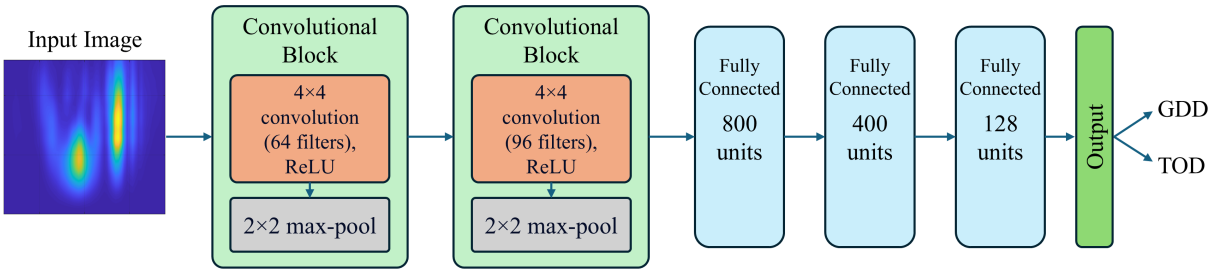

Figure 3.6: Architecture of the convolutional neural network used for the prediction of GDD and TOD from the observed CHG spectral maps. The network accepts an input CHG spectral map and processes it through two convolutional blocks, each consisting of a 4 × 4 convolution layer (with 64 and 96 filters, respectively), ReLU activation, and 2×2 max-pooling. The resulting features are flattened and passed through three fully connected layers (800, 400, and 128 units) to predict the output dispersion parameters GDD and TOD.

The specific architecture was chosen as a balance between model complexity and performance, tailored to the nature of the input data. Since the input images are of relatively small dimension (downsampled to 64×64 pixels) and the spectral features are not overly complex, a deep and computationally expensive model was not required. Initial experiments showed that a network with just two convolutional blocks was sufficient to learn the underlying patterns effectively. Adding a third convolutional layer did not yield a significant improvement in accuracy. The final parameters, such as the kernel sizes and the number of units in the fully-connected layers, were determined through hyperparameter tuning.

For training the CNN, the Huber loss function [111] was used, which is particularly well-suited for regression tasks where the data may contain outliers. The Huber loss function combines the advantages of both the mean squared error (MSE) and the mean absolute error

[2] https://github.com/arjun-krishnan/CHG-spectra-analyse

(MAE). It is defined as

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq \delta \\ \delta\left(|a| - \frac{1}{2}\delta\right) & \text{for } |a| > \delta \end{cases} \tag{3.9}$$

where a is the difference between the predicted and actual values, and δ is a threshold parameter. When the absolute error $|a|$ is less than δ, the Huber loss behaves like the MSE, giving a quadratic penalty to small errors. For $|a|$ greater than δ, it behaves like the MAE, which is less sensitive to outliers. This property makes the Huber loss more robust than MSE, as it prevents large errors of outliers from disproportionately influencing the model. For the CNN developed in this work, δ was chosen to be 0.5.

For error estimation in the CNN's predictions, the Monte Carlo dropout technique [112] was employed. Dropout is a regularization method originally introduced to prevent overfitting by randomly deactivating a subset of units within the network during training. In Monte Carlo dropout, this idea is extended to the prediction, where dropout is applied multiple times to the trained network, effectively creating an ensemble of slightly different models. During each forward pass, a random subset of activations is suppressed with a predefined probability p, and a prediction is generated. By repeating this process multiple times, a distribution of predictions for each input is obtained. The mean of this distribution is taken as the final prediction, while the standard deviation provides an estimate of the uncertainty in the prediction. For the network developed, p was chosen to be 0.3 and the forward passes were repeated 100 times to generate the distribution of predictions. This method allows one to quantify the model's confidence in its predictions and to account for the variability in the data.

3.3.2 Preparation of the Training Data

Since no labeled real-world measurements of CHG spectra were available, simulations were necessary to create the training samples. The simulated CHG spectral maps were generated for various combinations of GDD and TOD. Specifically, the GDD values ranged from -7000 fs^2 to 7000 fs^2, and the TOD values from 0 fs^3 to 30000 fs^3. Negative TOD values were intentionally excluded from the training set based on the observation that the spectral maps remain unchanged when the signs of both GDD and TOD are inverted. To resolve this sign ambiguity, a GDD sign convention based on the physical properties of the pulse compressor was first established. The 'zero GDD' setting, which yields the shortest pulse, was identified as a reference. Any compressor grating separation smaller than this reference was defined as positive net GDD (under-compensating the positive GDD from the stretcher), and any separation larger was defined as negative net GDD (over-compensation). A rough comparison of observed and simulated spectral maps suggested that the TOD in the real laser system

should be positive, leading to the decision to focus exclusively on positive TOD values in the training samples.

The simulated spectral maps were preprocessed to make them suitable for training the CNN. Each spectral map was normalized to its maximum value to ensure consistency across samples. The maps were then downsampled using linear interpolation to a resolution of 64×64 pixels. This reduction in size was done to balance the resolution of the spectral features with the computational efficiency of the CNN.

To better mimic the noise observed in real-world measurements, a Gaussian white noise overlay with a mean of 0 and standard deviation of 0.05 was added to each image to improve the model's robustness to noisy data. Additionally, to introduce some variability into the dataset, each image was randomly translated by a few pixels along both the vertical and horizontal axes. This data augmentation technique further improves the generalization capability of the CNN by preventing it from overfitting to position-dependent features or specific noise patterns in the training set.

The training set for each harmonic consisted of 43,310 samples, representing a comprehensive range of GDD and TOD values within the specified ranges.

3.3.3 Model Training and Validation

The dataset of 43,310 simulated spectral maps was partitioned into training, validation, and test sets using a standard 80-10-10 split. Here, 80% of the samples were used to train the network's parameters, while 10% served as a validation set to monitor for overfitting and fine-tune hyperparameters during the training process. The remaining 10% were reserved as a held-out test set, which the model never saw during training, to provide an unbiased assessment of its final predictive performance.

The finalized models for each harmonic demonstrated strong predictive accuracy and generalization, achieving a low Huber loss on their respective test sets of approximately 0.0015 for the 400 nm model, 0.0016 for 266 nm, and 0.0018 for 200 nm. This confirmed the reliability of the models before their application to experimental data.

3.4 Experimental Observation of CHG Spectra

In the absence of a direct method to measure the temporal shape of the coherent harmonic generation (CHG) pulses, the experimental analysis is limited to the time-integrated intensity and spectral properties of the CHG radiation. The spectra of CHG radiation down to 200 nm were measured using an in-air Czerny-Turner-type monochromator [79] (Acton

Research Corporation SpectraPro-275) equipped with an image-intensified CCD (iCCD) camera (Andor iStar 334T) at beamline 4 (BL 4). The short gating window (< 300 ns) of the iCCD camera allows the detector to be synchronized with the arrival of individual CHG pulses, which occur every 1 ms. This precise, short-duration gating effectively filters out the background noise from the 2600 incoherent synchrotron radiation pulses that occur between consecutive CHG pulses.

To observe radiation below 200 nm, an extreme ultraviolet (XUV) spectrometer (HP Spectroscopy easyLIGHT XUV) has been installed at BL 4. This instrument uses a diffraction grating to disperse the XUV light onto a microchannel plate (MCP) detector. The MCP is gated by a fast high-voltage switcher (Behlke FSWP 51-02) to ensure it only amplifies the signal synchronized with the arrival of the CHG pulse. The intensified signal from the MCP strikes a phosphor screen, converting it to visible light, which is then imaged by a CCD camera. This setup allows for recording spectra ranging from 210 nm down to 30 nm. However, the experimental setup at the time included a magnesium fluoride (MgF_2) window that separated the storage ring vacuum from the beamline. This window blocks wavelengths below 130 nm, thereby limiting the observation of higher harmonics to the 6th harmonic of the 800 nm seed, corresponding to 133 nm. This window was removed in 2023 in light of the implementation of the EEHG scheme at DELTA, which is discussed in detail in Chapter 5.

After the laser-electron overlap is optimized following the procedure outlined in the previous chapter, spectral measurements were performed at the 2nd, 3rd, and 4th harmonics using the Czerny-Turner setup and at the 4th, 5th, and 6th harmonics using the XUV spectrometer. To create the spectral maps, spectra were recorded while scanning the chicane current from 0 A to 700 A. The scan was conducted at a rate of 1 A/s, with one spectrum recorded every second, resulting in a total of 700 spectra for each spectral map.

In postprocessing, a region of interest (ROI) was selected in each image of the spectra. Within this ROI, the pixel intensities were summed across each column to reduce the 2D spectral image to a 1D array representing the intensity distribution across the wavelength range. An example for this step is shown in Fig. 3.7. These 1D projections from all 700 images were then stacked to form the final spectral map, which displays the spectra as a function of wavelength on the horizontal axis and chicane current (or equivalently R_{56}) on the vertical axis.

The same data acquisition procedure was repeated with different spectral phases of the seed laser pulse by altering the separation d between the compressor gratings inside the laser amplifier. The resulting GDD and TOD values are determined by Eq. (3.3) and (3.4). The specific parameters for the compressor at DELTA include a grating constant of $g = 1500$ mm^{-1} and an angle of incidence of $\gamma = 51.3°$. The grating distance corresponding to the shortest possible compressed pulse with a duration of approximately 40 fs is $d \approx 265$ mm.

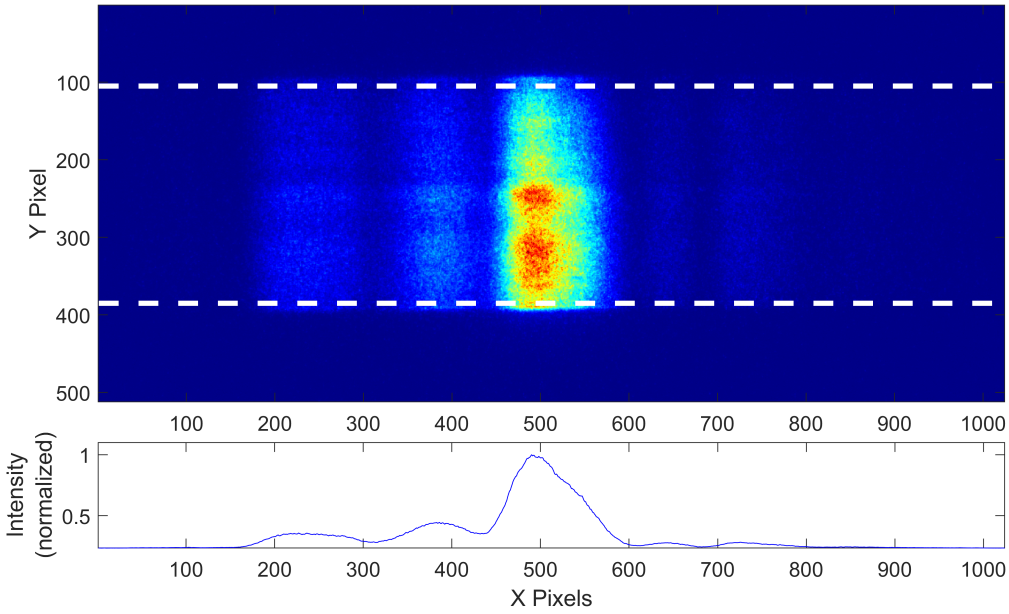

Figure 3.7: Top: Raw image from the iCCD camera of the CHG radiation around 400 nm with a chicane current of 400 A. The limits of ROI are shown in dashed white lines. Bottom: A projection of intensities within the ROI.

3.5 Experimental Results and Analysis

Figures 3.8 and 3.9 present the recorded CHG spectral maps, extending up to the 6th harmonic of the 800 nm seed wavelength. These spectral maps demonstrate the evolution of CHG spectra under varying frequency chirp of the seed laser pulse. The spectral maps corresponding to minimum frequency chirp (center column) display fringes arising from the interference of the two maxima at high R_{56} values, consistent with the simulations (Fig. 3.3 (left)). In contrast, with a large positive chirp (right column), the spectral maps reveal an asymmetric parabolic shape, attributed to high GDD and TOD in the seed pulse, as seen in simulations (Fig. 3.3 (right)). The left column shows intermediate spectral features, corresponding to moderate GDD values.

3.5.1 CNN Implementation for the Second Harmonic

The convolutional neural network described in Section 3.3 was used to predict the GDD and TOD values for 11 different compressor grating settings, given the 2nd harmonic spectral maps as input. The predictions were analyzed by fitting the data to a linear function, where the slope was fixed to the theoretical values derived from Eq. (3.3) and (3.4), while a constant vertical offset was treated as the only free parameter. The results of the fit are presented in Fig. 3.10 with the red dotted lines denoting the 95% confidence interval of the fit.

The GDD predictions closely match the theoretical expectations, demonstrating the reli-

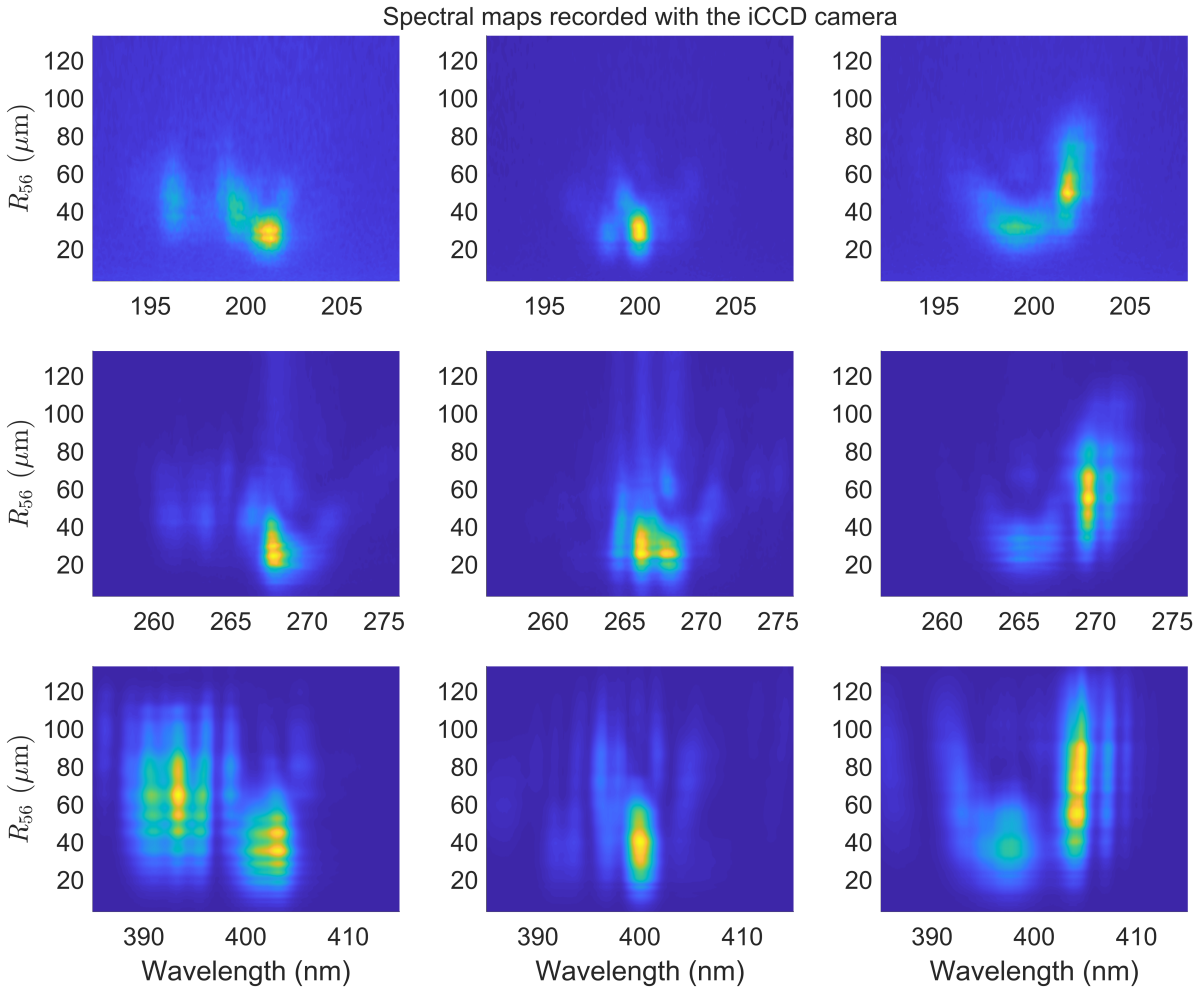

Figure 3.8: Observed spectral maps for the 2nd (bottom row), 3rd (middle row), and 4th (top row) harmonic of the 800 nm seed laser recorded using the iCCD camera. Each column displays the result for a different seed pulse frequency chirp: positive (left), zero (center), and negative (right). The datasets for different wavelengths were taken on different days, and as such, minor variations in other experimental conditions may be present between columns.

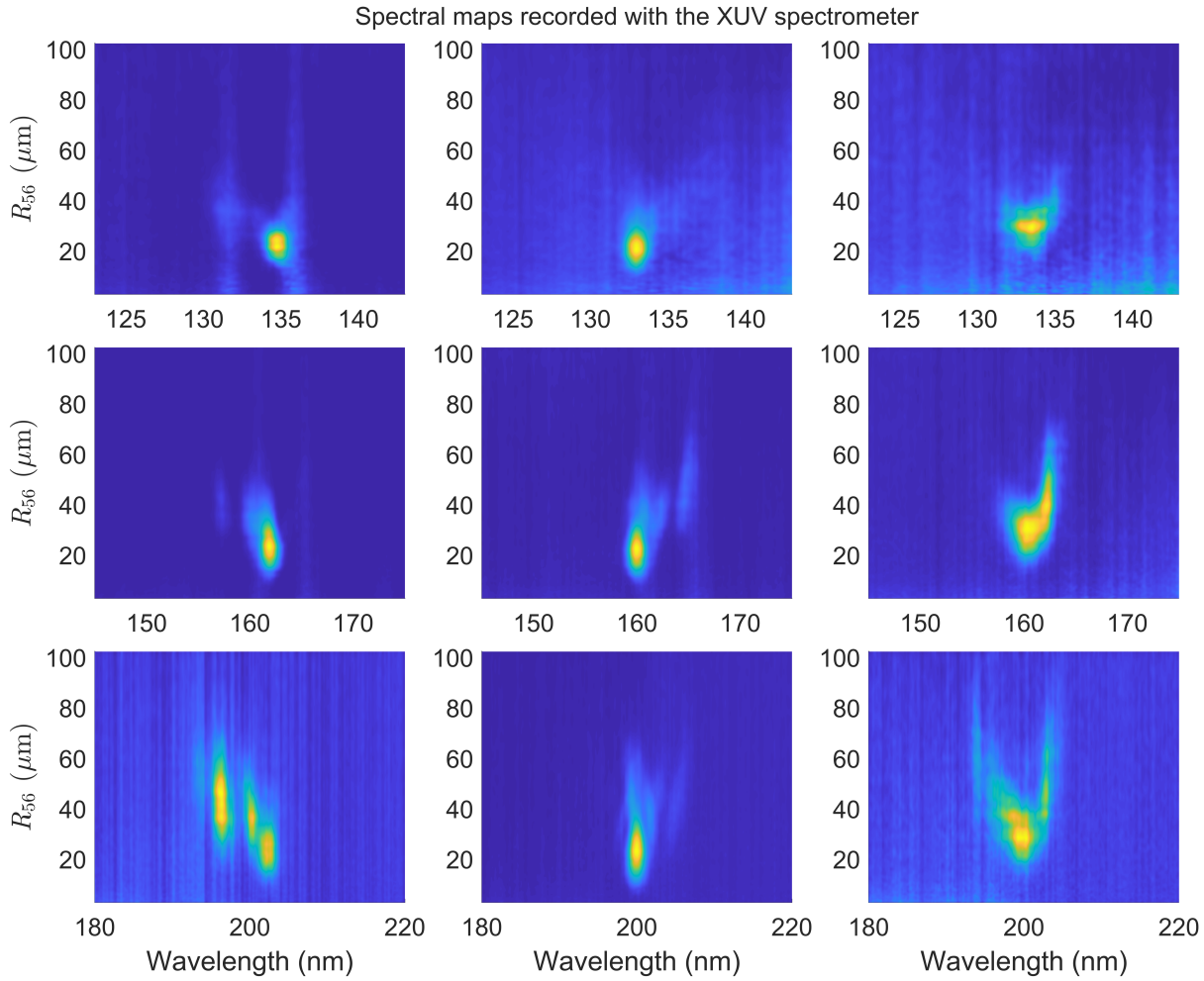

Figure 3.9: Observed spectral maps for the 4th (bottom row), 5th (middle row) and 6th (top row) harmonic of the 800 nm seed laser recorded using the XUV spectrometer. Each column displays the result for a different seed pulse frequency chirp: positive (left), zero (center), and negative (right). The datasets for different wavelengths were taken on different days, and as such, minor variations in other experimental conditions may be present between columns.

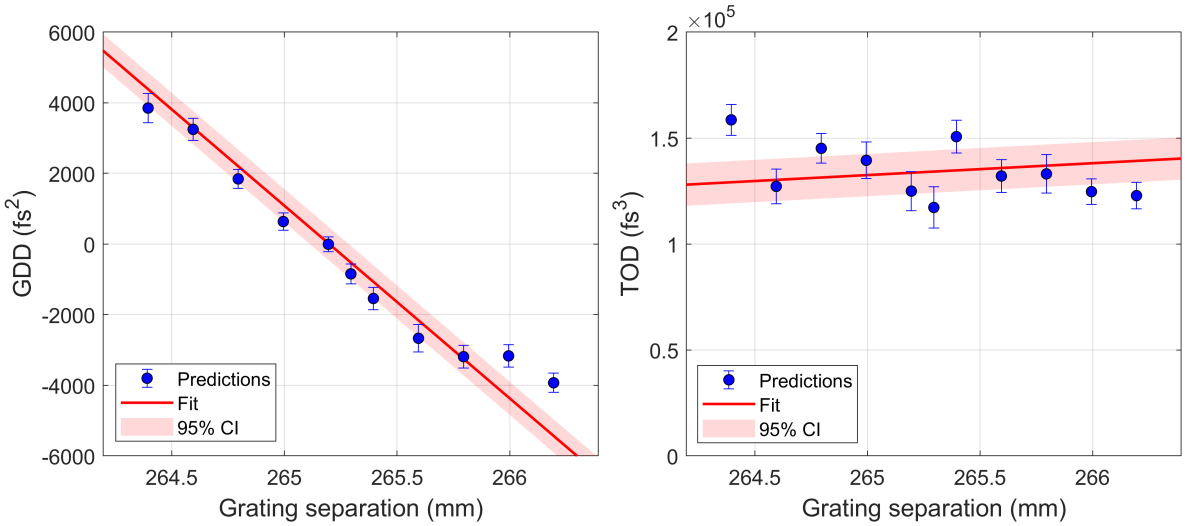

Figure 3.10: The blue dots indicate the CNN-predicted GDD (left) and TOD (right) as a function of the compressor grating separation. Error bars were calculated as the standard deviation of the distribution of predictions using the Monte Carlo dropout technique. Linear fits to the data are shown as red lines and the red shaded regions show the 95% confidence interval.

ability of the CNN in extracting GDD from the spectral features. In contrast, the TOD predictions exhibit large error bars, indicating a higher degree of uncertainty in the prediction. Despite this larger uncertainty, the model consistently predicts a large, positive TOD. The observed imprecision is primarily attributed to the relatively low sensitivity of harmonic spectral features to variations in TOD compared to GDD. While changes in GDD produce clear and distinguishable effects on the spectral maps, variations in TOD induce more subtle modifications, making it more challenging for the CNN to extract precise information. In addition, higher-order dispersion terms may be playing a role where the CNN interprets spectral modifications induced by these higher orders as a TOD effect. This inherent limitation affects the precision of TOD estimation, highlighting the need for improved approaches or additional spectral diagnostics to enhance sensitivity to TOD and higher-order spectral phase.

Figure 3.11 compares the observed, simulated, and interpolated CHG spectral maps for three different grating separations corresponding to positive, minimum, and negative GDD. They show a high degree of similarity, particularly in terms of overall shape, features, and peak positions.

It is important to note that the predicted values of TOD are found to be significantly large and positive, on the order of 20,000 fs^3. This result is in contrast to the assumption that the combination of a pulse stretcher and compressor in the chirped pulse amplification system should ideally yield an amplified pulse with near-zero GDD and TOD, where the dispersive effects of the stretcher are compensated by the compressor. The large positive TOD predicted

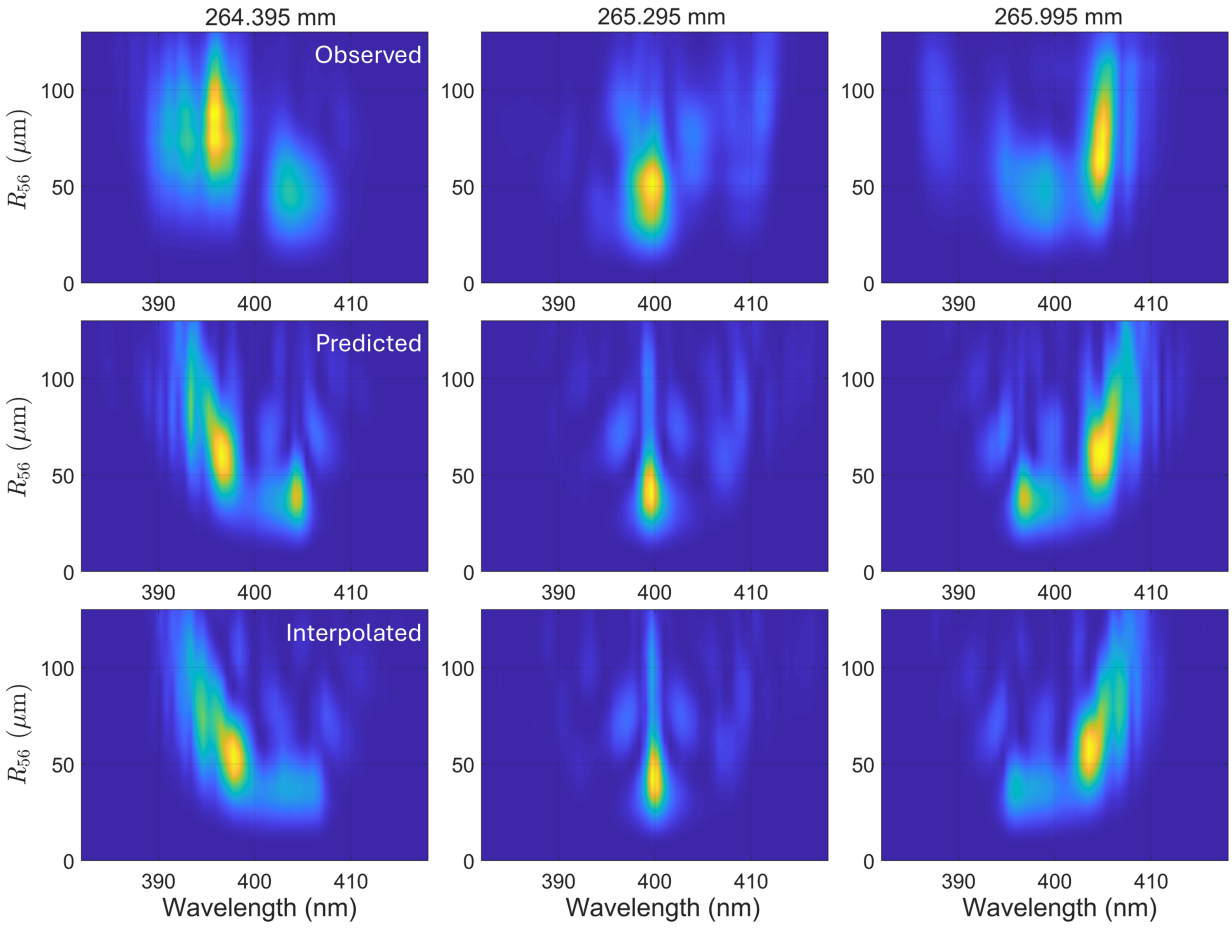

Figure 3.11: Observed CHG spectral maps (top row), simulated spectral maps from the GDD and TOD values predicted by CNN (middle row), and spectral maps for GDD and TOD values interpolated from the fit (bottom row), i.e., using the corresponding GDD and TOD values on the solid line in Fig. 3.10 for different compressor grating separations specified above each column.

in this study suggests that the compensation provided by the compressor may be incomplete, especially for the higher-order terms. Another possibility is the introduction of additional TOD from optical components within the laser system and on the beam path to the laser-electron interaction region, which may collectively contribute to the uncompensated TOD observed in the measurements.

The presence of a significant TOD suggests suboptimal conditions for achieving a transform-limited laser pulse, which directly impacts the efficiency and quality of the CHG output pulse. A transform-limited pulse represents the ideal case for maximizing the temporal coherence and spectral intensity of CHG emission. However, the observed TOD opens a possibility of leveraging it as a tunable parameter for shaping the temporal and spectral characteristics of the CHG emission.

By carefully controlling the TOD, it becomes possible to tailor the pulse shape and spectral content to suit specific experimental needs. For instance, when both TOD and GDD are

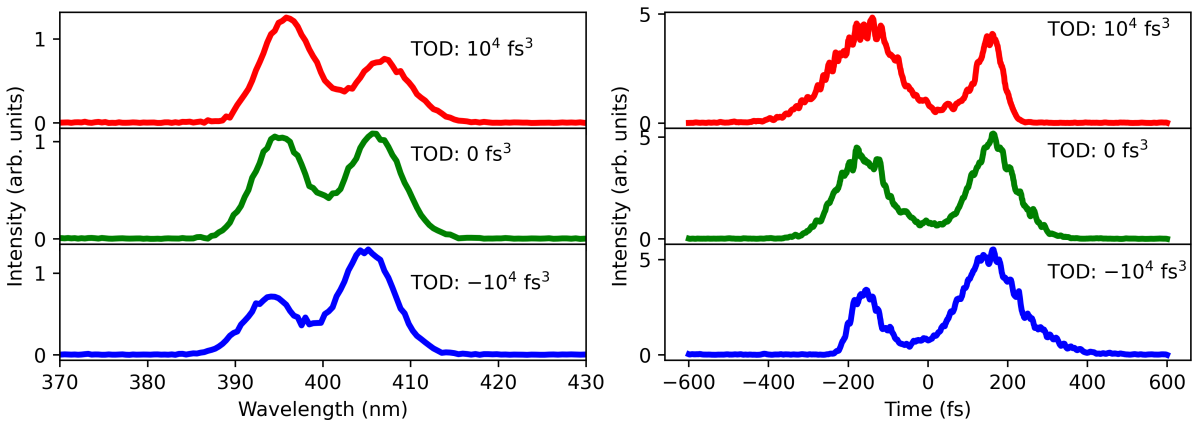

Figure 3.12: Spectra (left) and temporal shape (right) of the CHG pulse at different TOD values. All examples are for GDD = 5000 fs^2 and $R_{56} = 50$ μm. The simulations were performed using *SPECTRA* [113].

large, the CHG process can generate two temporally separated maxima with distinct central wavelengths and intensities at large R_{56} values. The magnitude of the TOD determines the degree of asymmetry between these maxima, with a zero TOD condition resulting in identical intensities, whereas increasing the TOD leads to larger asymmetry in the intensities between the maxima. The pulse shape and spectra for 3 different combinations of GDD and TOD are depicted in Fig. 3.12. It can be seen that by tuning the TOD, one can control the asymmetry between the two maxima and can tune the width of the individual maximum.

3.5.2 Comparing Predictions with Observations

To compare the spectral characteristics across the measured, predicted, and interpolated spectral maps, the positions of the two dominant peaks were analyzed for all 11 samples. The peak near the central wavelength of 400 nm was defined as the central peak, while the adjacent bright feature, appearing around 395 nm or 405 nm, was designated as the secondary peak. The location of the secondary peak relative to the central wavelength was observed to depend on the sign of the GDD.

Figure 3.13 (left) shows the relationship between the central peak position and the grating separation for the observed, predicted, and interpolated spectral maps. The data reveal a consistent trend across all three cases: as the grating separation increases, the central peak position transitions from longer to shorter wavelengths. Similarly, the secondary peak position is plotted as a function of grating separation in Fig. 3.13 (right). In this case, a notable trend emerges in which the secondary peak position undergoes a sharp transition from wavelengths less than 400 nm to larger than 400 nm. This shift corresponds to the GDD switching from positive to negative. The alignment between the observed, predicted,

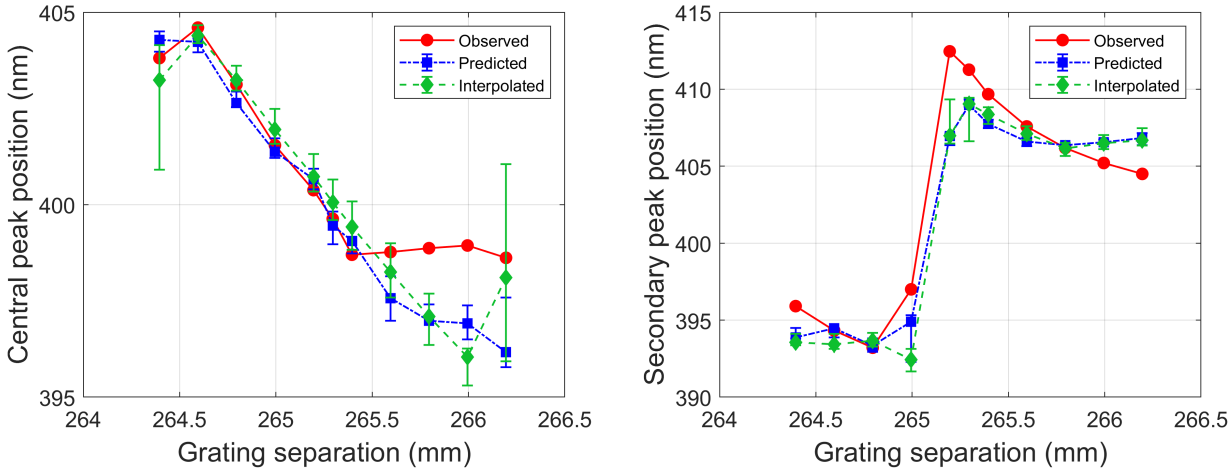

Figure 3.13: Central peak position (left) and secondary peak position (right) as a function of compressor grating separation for CHG spectra around 400 nm. The plots compare the observed data (red) with the CNN predictions (blue) and interpolated data (green). A strong agreement is observed across all datasets, validating the predictive model. The secondary peak position notably undergoes a sudden transition from shorter to longer wavelengths as the GDD changes sign around 265 mm.

and interpolated results across both comparisons provides strong evidence for the validity of the convolutional neural network.

To assess and quantify the asymmetry in the spectral features, a parameter was introduced that measures the modulus of the shift between the central peak position and the centroid of the intensity distribution along the wavelength axis. Let $I(\lambda, R_{56})$ be the full two-dimensional spectral map, which represents the intensity as a function of wavelength and chicane strength. The central peak position, λ_{peak}, is determined from the spectrum at the optimal chicane strength R_{56}^* which yields the maximum intensity for the central peak. A smoothing spline $f(\lambda)$ is first fitted to this one-dimensional intensity distribution $I(\lambda, R_{56}^*)$. The peak position $\hat{\lambda}$ is then defined as the wavelength that maximizes this fitted curve as

$$\hat{\lambda} = \arg\max_{\lambda} f(\lambda). \tag{3.10}$$

The weighted centroid $\bar{\lambda}$ is calculated from the projection of the entire 2D spectral map onto the wavelength axis. This projected spectrum $\bar{I}(\lambda)$ represents the total intensity at each wavelength, summed over all chicane strengths. The centroid is the intensity-weighted average of this projected spectrum given by

$$\bar{\lambda} = \frac{\int \lambda \cdot \bar{I}(\lambda)\, d\lambda}{\int \bar{I}(\lambda)\, d\lambda}. \tag{3.11}$$

The final asymmetry parameter is the absolute difference between the centroid of the projected map and the peak of the optimal spectral slice given by

$$\text{Asymmetry} = |\bar{\lambda} - \hat{\lambda}|. \tag{3.12}$$

This parameter provides a robust metric for evaluating deviations from symmetry. A symmetric spectrum will have its brightest point aligned with its centroid, yielding an asymmetry of zero. In contrast, for a spectrum with a significant side peak or lopsidedness, the centroid is pulled away from the central peak, resulting in a larger value of asymmetry. This method is sensitive to variations in spectral shape, such as those induced by changes in GDD.

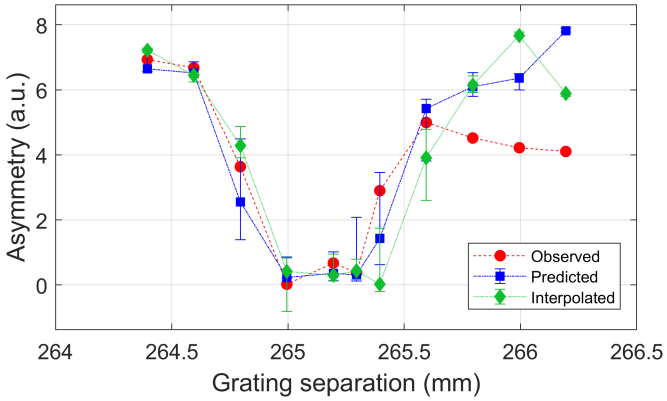

Figure 3.14: The asymmetry parameter for the CHG spectra around 400 nm as a function of compressor grating separation. The data exhibit a pronounced dip around 265.2 mm, which corresponds to the point of zero GDD where the spectrum is expected to be most symmetric.

The asymmetry parameter as a function of grating separation is presented in Fig. 3.14. A distinct trend is observed, characterized by a pronounced dip in the asymmetry parameter at a grating separation of approximately 265 mm. This point corresponds to the condition of zero GDD, where the spectral features are expected to be most symmetric. Away from this point, the asymmetry increases, reflecting the impact of nonzero GDD on the spectral features. Crucially, GDD alone does not produce the observed spectral asymmetry. A nonzero TOD is required, as this is the term responsible for creating an asymmetric temporal pulse shape, as illustrated in Fig. 3.5. The increase in the asymmetry parameter away from the zero-GDD point is therefore a direct result of the interplay between the varying GDD and the persistent TOD of the laser system. The asymmetry parameter derived from the observed spectral maps demonstrates excellent agreement with the predicted and interpolated values, underscoring the accuracy of the theoretical model and the efficiency of the convolutional neural network used for analysis.

3.5.3 CNN Implementation for Higher Harmonics

To extend the analysis to higher harmonics, the same model architecture was trained on datasets of 41310 samples for both 266 nm and 200 nm CHG spectra separately. Figure 3.15 shows the predicted GDD and TOD for all three wavelengths: 400, 266, and 200 nm. Predictions for GDD follow a linear trend, which is expected, and closely follow the theoretical slope dictated by Eq. (3.3). The TOD predictions, however, show a large spread and uncertainty. This can be attributed to the low sensitivity of the spectral features to the changes in TOD.

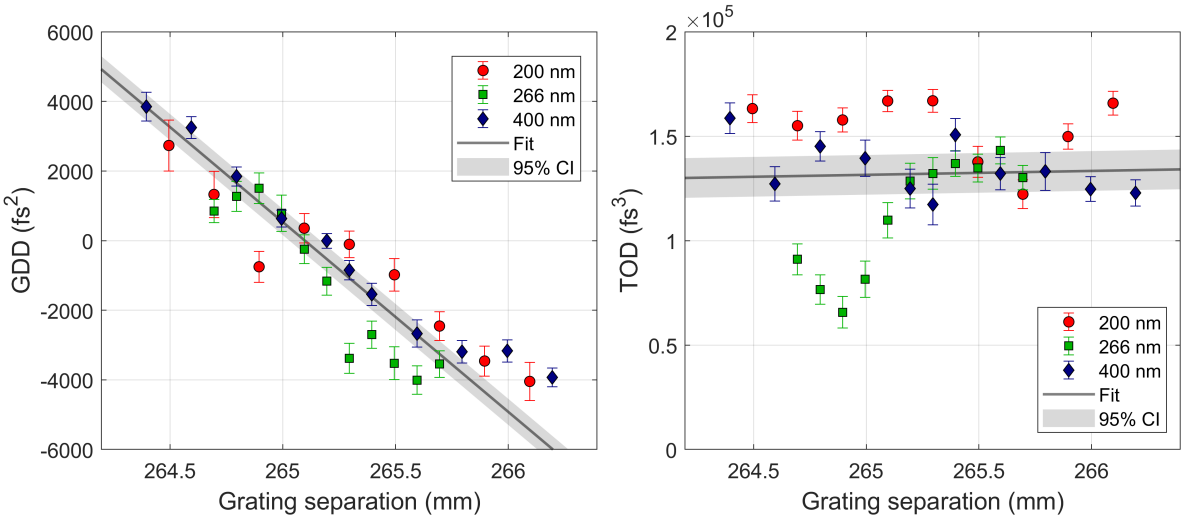

Figure 3.15: Predicted GDD and TOD as a function of compressor grating separation for the 2nd, 3rd, and 4th harmonics of 800 nm. The GDD predictions (left) show a strong linear correlation consistent with theory for all harmonics. The TOD predictions (right) exhibit scatter, indicating the lower sensitivity of the spectral features to this parameter.

A similar analysis of the central peak, secondary peak, and asymmetry parameter was performed for the 3rd and 4th harmonics, as depicted in Fig. 3.16 and Fig. 3.17, respectively. The peak position analyses show strong qualitative agreement between the observed data, the CNN predictions, and the interpolated results, with the secondary peak flipping sides at the expected grating separation.

However, in contrast to the 2nd harmonic (400 nm), the asymmetry parameter for the higher harmonics does not exhibit a pronounced symmetric dip. While the asymmetry approaches a minimum near 265 mm from larger grating separations, a corresponding increase is not observed towards smaller separations. This is primarily attributed to the limited experimental scan range, which was insufficient to capture the full dip. The lower signal-to-noise ratio inherent in the higher harmonics also makes it more challenging to resolve a well-defined minimum.

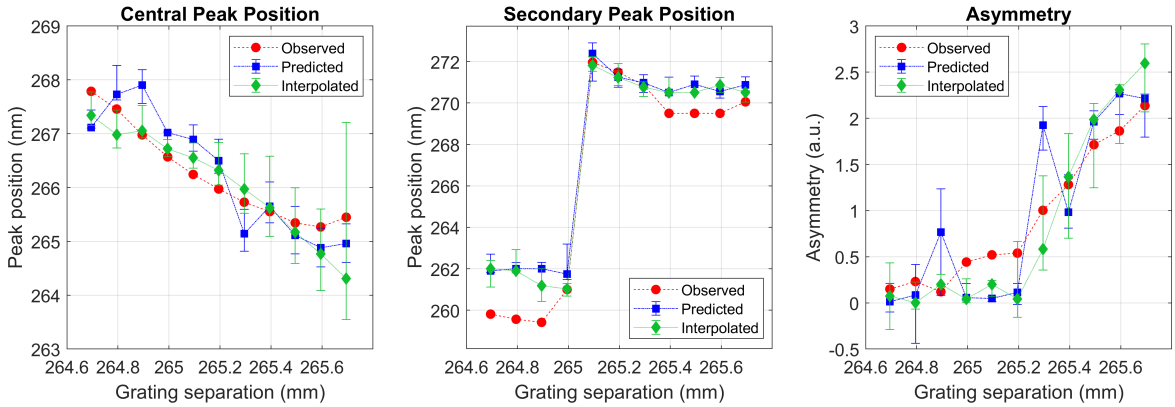

Figure 3.16: Comparison of CHG spectral features for the 3rd harmonic versus grating separation. Plotted are the central peak position (left), secondary peak position (right), and the asymmetry parameter (bottom).

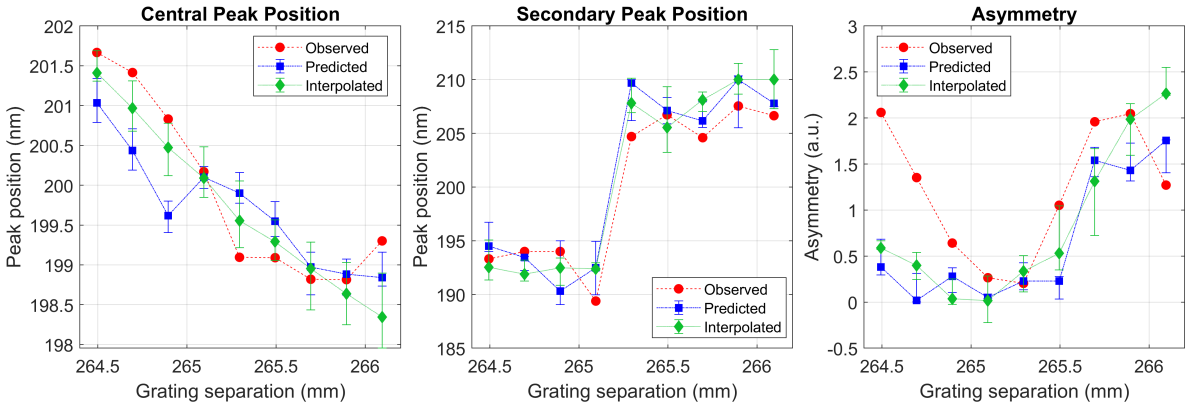

Figure 3.17: Comparison of CHG spectral features for the 4th harmonic versus grating separation. Plotted are the central peak position (left), secondary peak position (right), and the asymmetry parameter (bottom).

The 5th and 6th harmonics exhibit similar qualitative features as the other harmonics discussed here but are significantly less pronounced and suffer from a lower signal-to-noise ratio. Since including such noisy data is not ideal for CNN training, these harmonics are excluded from the analysis.

3.6 Summary

This chapter presented the development and validation of a novel diagnostic method based on a CNN to characterize the spectral phase of ultrashort laser pulses by analyzing CHG spectra. The traditional challenge of interpreting complex, noisy experimental data was overcome by training a CNN on simulated spectral maps, enabling it to recognize and quantify the underlying laser parameters from robust spectral features rather than relying on pixel-to-pixel comparisons.

This work successfully developed a novel diagnostic for the CHG process that operates by characterizing the spectral phase of the seed laser from the CHG spectra. The traditional challenge of interpreting complex, noisy experimental data was overcome by training a CNN on simulated spectral maps. This enables the network to recognize and quantify the underlying laser parameters that produced the harmonics, moving beyond pixel-to-pixel comparisons. This method provides a robust way to monitor the CHG process and understand its high sensitivity to the seed laser parameters.

The trained CNN model demonstrated high accuracy in predicting the GDD, with predictions showing excellent linear correlation with the physical adjustments of the pulse compressor in the laser system. More significantly, the model revealed a large, persistent positive TOD on the order of 20,000 fs^3. This finding is critical, as it suggests the chirped-pulse amplification in the laser system leaves a substantial uncompensated higher-order chirp.

The presence of a large TOD provides a clear explanation for the pronounced asymmetries observed in the experimental CHG spectral maps. While GDD governs the symmetric broadening and interference patterns, it is the interplay between GDD and the dominant TOD that dictates the asymmetric features, such as the location and intensity of secondary spectral peaks. This effect was quantitatively captured using a custom asymmetry parameter, which showed strong agreement between experimental data and model predictions. This large asymmetry, which was less apparent in previous work at facilities such as FERMI [22] likely due to their use of a near transform-limited laser, provided an ideal comparison for validating the model. The results thus provide a quantitative demonstration of how higher-order spectral phase, when present, can dominate the spectrotemporal properties of coherently emitted radiation.

The ability to quantify the influence of GDD and TOD on the spectra underscores the potential for advanced pulse shaping in CHG or HGHG seeding schemes by independently controlling these parameters. However, the experimental setup lacked a mechanism for direct TOD manipulation, preventing a systematic investigation of its effects on spectral properties. A controlled study of the influence of TOD could be achieved by tuning the stretcher of the CPA system or using an array of TOD-compensating mirrors. By selectively

introducing these mirrors into the beam path, one could modulate TOD and conduct a detailed examination of its role in spectral asymmetry.

Chapter 4

Deep Learning-Based Photon Profile Reconstruction for Free-Electron Lasers

Free-electron lasers (FELs) have become indispensable tools for investigating ultrafast processes in matter. Their ability to deliver intense, femtosecond-scale photon pulses has opened new frontiers in time-resolved studies of atomic, molecular, and solid-state systems. However, advancing toward even shorter pulse durations, approaching the sub-femtosecond regime, remains a major objective in the field. This challenge is closely tied to the production of highly compressed, high-current electron bunches, typically achieved through several bunch compression stages by RF acceleration combined with dispersive magnetic elements. However, space charge effects and other collective beam dynamics constrain the extent of achievable compression. To overcome this, several advanced laser-based bunch compression or bunch shaping schemes have been proposed and have shown promise in generating ultrashort FEL pulses [114–118]. However, fully exploiting these ultrashort-light-pulse sources for experiments requires precise knowledge of the temporal structure of each FEL pulse. One method is to infer the temporal profile of the FEL pulse from the measured longitudinal phase space density of the driving electron bunch [24]. The FEL process [45], and as a result the emitted FEL pulse profile, strongly depends on time-varying properties of the electron bunch, such as the current, energy spread, and emittance. In turn, the FEL emission leads to an energy loss and an increased energy spread in the electron bunch. By analyzing the phase space density of the electron bunch, particularly the FEL-induced energy loss and spread, it becomes possible to reconstruct the temporal envelope of the emitted FEL pulse on a shot-by-shot basis.

This chapter introduces a pulse-shaping method implemented at FLASH, Hamburg [118], and a deep learning-based method for reconstructing photon pulse profiles from the longitudinal

phase space density of electron bunches. A schematic layout of the FLASH facility is shown in Fig. 4.1.

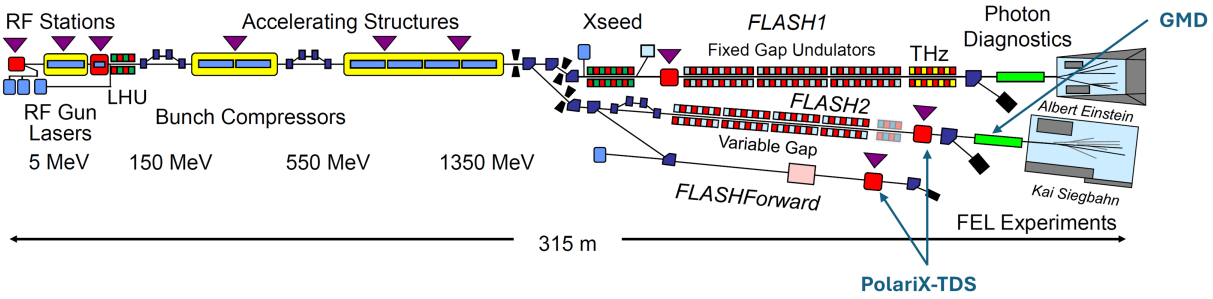

Figure 4.1: Schematic layout of the FLASH facility as of 2023 (not to scale). The two main SASE beamlines, FLASH1 and FLASH2, are operational and serve user experiments. A third beamline is dedicated to the FLASHForward plasma acceleration experiment. The locations of the laser heater undulator (LHU), PolariX-TDS diagnostics, and the gas monitor detector (GMD) for pulse energy measurements are indicated. Image source: https://flash.desy.de/, courtesy of Siegfried Schreiber.

4.1 Pulse Shaping Using Laser Heater

The laser heater is a critical subsystem for suppressing the microbunching instability (MBI) [119–121]. It imprints an energy modulation on electron bunches via interaction with a laser pulse in an undulator (LHU in Fig. 4.1). This modulation is converted into an increased energy spread after a subsequent magnetic chicane, i.e., the bunch is 'heated'. This early increase in the energy spread can suppress the MBI and enhance the performance of an FEL [120]. However, if the laser-induced energy spread is increased beyond a certain value, there will be a detrimental increase in the final energy spread of the beam. This is sometimes referred to as 'laser overheating'. Since the FEL process depends on the local charge density and energy spread, the lasing can be suppressed by increasing the energy spread beyond the lasing threshold, which is typically close to 0.1% at X-ray energies [8, 114]. The laser heater pulse profile can be manipulated so that energy spread over the threshold is introduced in parts of the beam, thus by suppressing the FEL process at these locations, while leaving certain parts of the beam unspoiled, which contribute to the FEL process. The temporal profile of the resulting FEL pulse can consequently be manipulated by tailoring the laser-heater pulse profile [114, 122].

The FLARE [118] setup at FLASH utilizes an interferometer to shape the laser heater pulse envelope and control the temporal profile of the resulting FEL pulse. Figure 4.2 shows the schematics of the FLARE setup where a laser pulse is split into two copies. A motorized delay stage in one of the interferometer arms introduces a delay between the two copies of the pulse. If the delay between the pulses is an integer multiple of the wavelength, the two

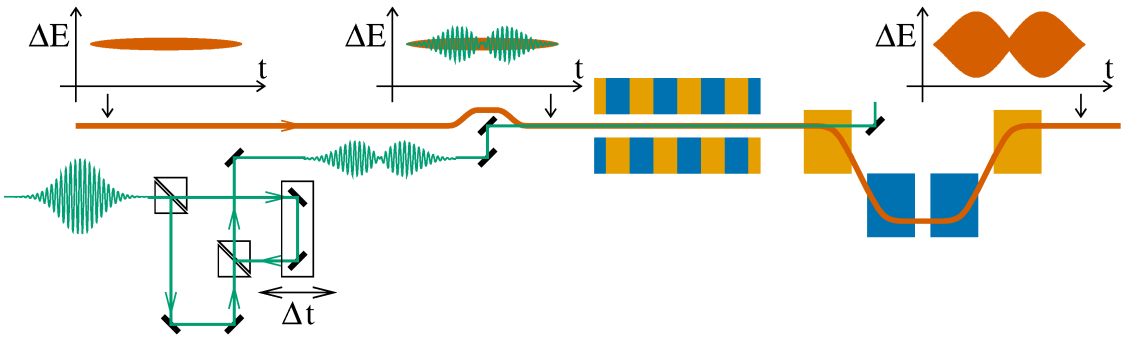

Figure 4.2: Schematics of the FLARE interferometer setup. A shaped laser pulse (green) from an interferometer imparts a corresponding energy modulation (ΔE) onto the electron bunch (orange) in the laser heater. This modulated bunch then generates a temporally shaped FEL pulse from the modulated electron distribution (top right) in the main undulator. Figure taken from Amstutz et al. [118]

pulses interfere coherently, resulting in a flat field envelope at the center. However, if they have an odd integer multiple of π phase difference, the destructive interference results in a bimodal field envelope with zero field amplitude at the center. Heating the beam with this non-uniform field profile allows for increasing the energy spread in certain longitudinal regions of the bunch while leaving certain regions unaltered. Using a sufficiently high pulse energy, the bunch can be overheated around the location of the two lobes of the envelope, which would then suppress the FEL process in these regions. Meanwhile, the minimum of the envelope at the center is small enough that it does not overheat the beam. Within this gap, the FEL process is not suppressed but rather enhanced due to the suppression of MBI. Hence, the resulting FEL pulse is given by this gap, which can be tuned in the order of femtoseconds by controlling the delay and the laser heater pulse energy.

Figure 4.3 shows the phase space distribution of a 300 pC electron bunch, measured using a transverse deflection structure (TDS, see Section 4.2) downstream of the FLASH2 undulator section at a beam energy of 860 MeV and an FEL wavelength of 25 nm. The left plot shows the case where the laser heater pulses are blocked, i.e., the beam is unheated. The bunch shows the characteristic jagged structures in the phase space distribution. These structures result from a combination of the FEL process and MBI, which are not suppressed in the absence of laser heating. corresponding to the phase space distribution resulting from the FEL process. When the laser heater pulses are overlaid with the FLARE interferometer in the destructive interference mode, the FEL process is essentially limited to the gap between the two lobes of the laser pulse amplitude profile. This is visible in the middle plot, where the lasing is limited to a small section around $t \approx -80$ fs. The right plot shows the longitudinal phase space distribution of the bunch when the FEL process is suppressed completely by laser overheating.

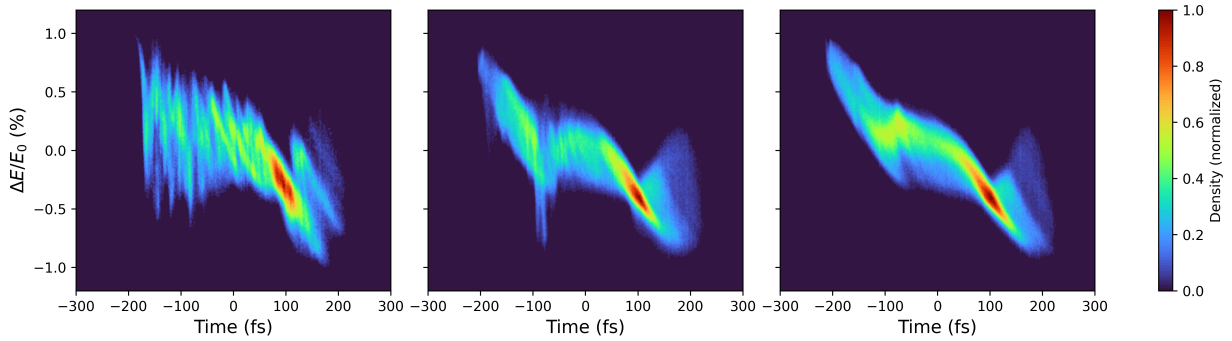

Figure 4.3: TDS measurements of the longitudinal phase space distribution downstream of the FLASH2 undulator. Left: An unheated bunch, showing the typical jagged features associated with lasing and MBI. Center: A partially overheated bunch showing a short-pulse structure at around -80 fs. Right: An overheated bunch with lasing suppressed.

4.2 Transverse Deflecting Structure

Transverse deflecting structures (TDS) are diagnostic tools in modern electron linear accelerators for investigating the temporal characteristics of relativistic electron bunches [23]. These radiofrequency (RF) devices operate by imparting a time-dependent transverse electromagnetic field, typically using a TM_{110} mode, which applies a transverse Lorentz force onto a particle bunch as it passes through the structure. The fundamental principle behind TDS operation is the correlation between the longitudinal position of particles within the bunch and the transverse momentum they acquire due to the time-varying fields. Typically, the TDS is driven at or near the zero-crossing phase of the RF field, where the transverse deflecting force changes sign. At this phase, particles at the head and tail of the bunch experience equal and opposite transverse kicks, resulting in a time-to-space mapping across the transverse angle. This process is often referred to as 'streaking' the bunch. When such a deflected bunch encounters a downstream observation screen, its longitudinal profile is effectively translated into a measurable transverse distribution [123, 124].

An energy spectrometer, such as a dipole magnet, can be employed with the TDS to gain insight into the longitudinal phase space distribution of the bunch, as shown in Fig. 4.4. The spectrometer introduces an energy-dependent transverse deflection in the plane orthogonal to the streaking direction of the TDS. In this configuration, one transverse axis of the observation screen corresponds to time (via the TDS streaking), while the orthogonal axis encodes energy (via the dipole dispersion). As a result, the two-dimensional distribution observed on the screen provides a direct mapping of the longitudinal phase space density of the electron bunch.

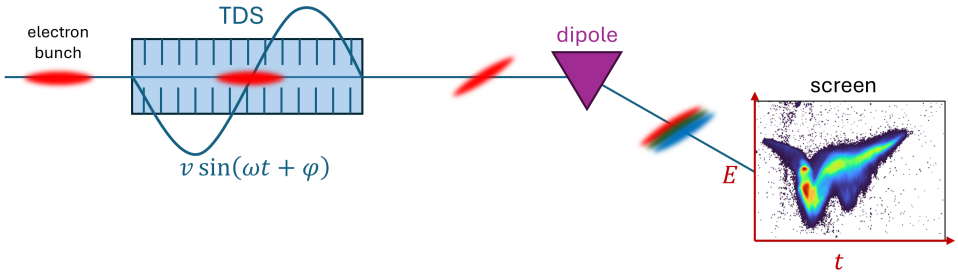

Figure 4.4: Schematic of a longitudinal phase space diagnostic setup. The electron bunch passes through a TDS, which imparts a time-dependent transverse kick. A subsequent dipole magnet introduces energy dispersion and directs the bunch towards the screen. The resulting distribution on the screen creates a mapping of the longitudinal phase space distribution.

At FLASH2, the implementation of such diagnostics is realized through the PolariX-TDS [125], an advanced X-band transverse deflecting structure specifically designed for high-resolution temporal characterization of electron bunches. The PolariX system offers a unique capability of adjustable polarization, which allows changing the orientation of the streaking field of the TDS to an arbitrary azimuthal angle. It provides sub-10 femtosecond temporal resolution, which is essential for resolving ultrashort bunches produced in high-compression modes. In practice, PolariX streaks the bunch along a chosen direction, and, when combined with a downstream dipole, reconstructs time-energy phase space or even 6-dimensional phase-space maps through multi-angle projections. Since January 2021, two PolariX cavities have been operational in the FLASH2 and FLASHForward beamlines [126, 127].

4.3 Methods for Photon Profile Reconstruction

Accurate characterization of FEL pulses is essential for the reliable interpretation of high-resolution time-resolved photon science experiments. A common approach is to infer the FEL pulse duration from the temporal profile of the electron bunch that drives the lasing process. However, this method often fails to yield an accurate representation of the actual photon pulse duration. Discrepancies between the length of the electron bunch and the resulting FEL pulse arise due to the complex and nonlinear nature of the FEL amplification process. This process is highly sensitive to temporal variations in key beam parameters including current, emittance, and energy spread [24, 128]. Intra-bunch non-uniformities can lead to spatially and temporally uneven lasing, thereby distorting the intensity and temporal structure of the emitted radiation. Moreover, shot-to-shot fluctuations caused by jitter introduce significant variations in photon pulse characteristics such as energy, temporal shape, and intensity profile. These intrinsic fluctuations underscore the need for robust and accurate photon pulse reconstruction techniques.

Machine learning approaches have been explored for FEL pulse reconstruction. Sanchez-Gonzalez et al. [129] proposed models such as artificial neural networks (ANN) [130] and support vector regressors (SVR) [131] to predict FEL pulse properties, including photon spectra and central photon energies, directly from machine and beam parameters. More recently, Korten et al. [132] introduced a machine learning method employing a multi-layer perceptron (MLP) [133] to estimate the FEL power profile. Their approach predicts the lasing-suppressed electron beam power profile (referred to as "lasing off") from machine parameters, which is then subtracted from the measured electron beam power profile during lasing ("lasing-on") to isolate the FEL pulse power profile. This method was successfully demonstrated at FLASH [132].

Recent machine learning applications for FEL pulse characterization have focused on predicting pulse properties directly from machine and beam parameters. Sanchez-Gonzalez et al. [129], for instance, used models such as artificial neural networks and support vector regressors [131] to predict properties including photon spectra and central photon energies. More recently, Korten et al. [132] introduced an multi-layer perceptron (MLP)-based [133] method to estimate the FEL power profile. Their approach, which was successfully demonstrated at FLASH, predicts the lasing-suppressed electron beam power profile (referred to as "lasing off") from machine parameters, which is then subtracted from the measured electron beam power profile during lasing ("lasing-on") to isolate the FEL pulse power profile.

Another approach to reconstructing the temporal profile of the FEL pulse is to infer it from the phase space distribution of the electron bunch after the lasing process, exploiting the intrinsic correlation between electron beam properties and FEL radiation characteristics. This method relies on measuring the "lasing footprint" imprinted on the electron bunch's longitudinal phase space distribution during the FEL process. As electrons interact with the undulator radiation field, energy is transferred from the electron beam to the photon field. This interaction manifests as both an energy loss and an increase in energy spread within the electron bunch, producing a distinctive signature in the longitudinal phase space distribution that directly encodes the temporal structure of the X-ray pulse [24, 134].

To reconstruct the time-resolved X-ray power profile, the electron phase space distribution from the TDS image is temporally sliced to extract parameters like centroid energy $E(t_i)$ and current $I(t_i)$ at each time slice t_i. By comparing the centroid energies for lasing-on ($E_{\text{on}}(t_i)$) and lasing-off ($E_{\text{off}}(t_i)$) conditions, the mean energy loss $\Delta E(t_i) = E_{\text{off}}(t_i) - E_{\text{on}}(t_i)$ can be determined, enabling the calculation of the instantaneous FEL power as [24]

$$P(t_i) = \Delta E(t_i) \times \frac{I_{\text{on}}(t_i)}{e}, \tag{4.1}$$

where e is the elementary charge. Shot-to-shot fluctuations in beam energy can be corrected by calibrating with simultaneous pulse energy measurements from a gas monitor detector

(GMD) [135].

A complementary method estimates the FEL power profile from the increase in energy spread induced by lasing. This is done by comparing the slice energy spreads with lasing on and off, such that [24]

$$P(t_i) \propto \left(\sigma^2_{E,\mathrm{on}}(t_i) - \sigma^2_{E,\mathrm{off}}(t_i)\right) \times I^{2/3}_{\mathrm{on}}, \tag{4.2}$$

where $\sigma_{E,\mathrm{on}}(t_i)$ and $\sigma_{E,\mathrm{off}}(t_i)$ are the RMS slice energy spread with lasing on and off, respectively. Due to the proportionality in the relation, this method requires normalization using independent pulse energy measurements. Figure 4.5 presents the FEL pulse power profiles reconstructed using both the energy loss and energy spread methods, demonstrating good agreement between the two approaches.

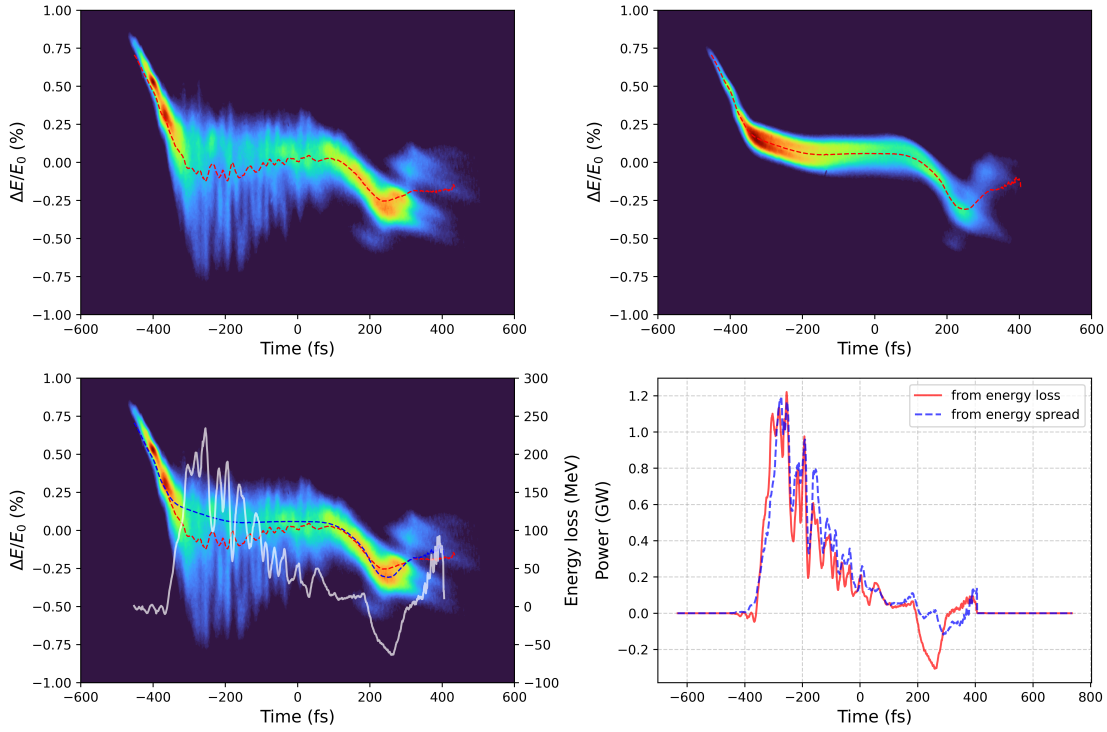

Figure 4.5: Top left: Single-shot longitudinal phase space distribution of a lasing electron bunch ('lasing-on' image) recorded by the PolariX-TDS. Top right: Longitudinal phase space distribution of a non-lasing electron bunch ('lasing-off' image). The red dotted curve in both images represents the centroid energy for each temporal slice. Bottom left: Lasing-on image with the centroids plotted together. The white curve shows the energy loss per slice due to lasing. Bottom right: Reconstructed FEL power profile. The solid red curve corresponds to the power derived from the energy-loss method, while the dashed blue curve shows the power calculated from the energy-spread method. The data correspond to an electron bunch with a total charge of 320 pC and beam energy of 1050 MeV.

The reconstruction techniques discussed above rely on sampling lasing-off images by intentionally suppressing FEL gain, typically achieved by manipulating the orbit in the undulator.

While this calibration step can be performed prior to experiments under stable conditions, maintaining its validity throughout extended operation is challenging. Fluctuations in beam properties, such as shot-to-shot variations or energy jitter, can lead to discrepancies between the recorded lasing-off references and the actual lasing-on conditions. As a result, mismatches between these datasets can cause significant reconstruction errors. For instance, in Fig. 4.5, the appearance of an unphysical 'negative power' region around 250 fs comes from subtle differences in the phase space distribution of the bunch tail. Here, a small offset in the energy centroid, combined with high current density, results in an exaggerated error during power calculation, particularly in the energy-loss-based approach. These limitations underscore the sensitivity of the method to beam stability and highlight the need for periodically updating reference datasets. Even so, there remains a considerable risk of introducing artifacts, compromising the reliability of the reconstructed photon pulse profile.

To address the limitations associated with relying on recording frequent lasing-off references during runtime, several machine learning approaches have been developed that aim to reconstruct the FEL power profile directly from lasing-on measurements [136, 137]. One such approach by Ren et al. [136] leverages numerical simulations of the FEL process using *GENESIS* [138], followed by beam tracking to the TDS using *ELEGANT* [139], to generate synthetic longitudinal phase space images. These simulated images, paired with the corresponding photon power profiles, are used to train a convolutional neural network (CNN) that maps lasing-on phase space images to power profiles. Although effective, this method depends on the availability of large-scale simulation datasets, which require significant computational resources to produce. In a different strategy, Goetzke et al. [137] propose a self-supervised learning framework based on a U-Net architecture [105]. Their model generates a lasing-off reference image by masking the lasing-affected region within the bunch and predicting the corresponding non-lasing equivalent. This technique is particularly advantageous for short FEL pulses, where the masked region is temporally localized, enhancing the accuracy of the reconstructed reference. On the other hand, this method presumes knowledge of the lasing region in the first place.

The subject of this chapter is another machine learning approach to the same problem - extracting the FEL temporal power profile from the longitudinal phase space images without the need for a reference lasing-off image. A U-Net architecture with a simpler image augmentation algorithm is used for the preparation of training samples. The coming sections describe this method in detail.

4.4 U-Net-Based Reconstruction of FEL Pulse Profiles

This thesis proposes a novel method for reconstructing the FEL's temporal pulse profile using a U-Net-based convolutional neural network (CNN) [1]. The network is trained to generate a 'lasing-off' image from a measured 'lasing-on' image. The difference between these two images is then used to estimate the FEL-induced energy loss, which in turn allows for the reconstruction of the X-ray pulse power profile.

4.4.1 Outline of the Method

Unlike the previously described approaches that rely on extensive start-to-end simulations or precise knowledge of machine parameters, this method leverages simple image augmentations to emulate the effect of lasing on the electron beam. Specifically, synthetic lasing-on images are generated by applying augmentations to experimentally measured lasing-off images, aiming to replicate the characteristic signatures observed in actual lasing-on phase space measurements such as localized energy loss and increased energy spread. These image pairs, measured lasing-off images and their corresponding synthetically augmented 'lasing-on' counterparts, serve as the training dataset for a U-Net architecture. The network is trained to perform the inverse transformation: given a synthetic lasing-on image as input, the network learns to reconstruct the corresponding lasing-off image. Because the data generation relies on basic augmentations rather than physics-based simulations, training samples can be prepared efficiently on a standard desktop machine without the need for high-performance computing resources.

Once trained, the network can be applied to measured lasing-on phase space images. The network attempts to predict the corresponding lasing-off image, effectively removing the FEL-induced modifications. By comparing the measured lasing-on image with the predicted lasing-off image, both the energy loss and energy spread at each temporal slice is estimated, from which the FEL power profile is constructed using Eq. (4.1) and Eq. (4.2), respectively.

4.4.2 Training Sample Preparation

The process of generating the training dataset from raw experimental data is divided into two main stages. First, a preprocessing procedure is applied to the raw images to clean, isolate, and normalize the electron bunch. Second, a multi-step data augmentation pipeline is used to synthetically generate the 'lasing-on' counterparts from the cleaned 'lasing-off' images, creating the final input-label pairs for training.

[1] For details regarding U-Net, refer to Section 4.4.3

Preprocessing

To improve signal fidelity and isolate the electron bunch from background noise, the raw image I_{raw} undergoes a three-step preprocessing procedure.

1. Median Filtering: A median filter with kernel size $k \times k$ is applied to suppress high-frequency noise while preserving the edges.

$$I_{\text{med}}(x, y) = \text{median}\{\, I_{\text{raw}}(i, j) \mid (i, j) \in \mathcal{W}_k(x, y) \,\} \tag{4.3}$$

where $\mathcal{W}_k(x, y)$ denotes the $k \times k$ window centered at pixel (x, y).

2. Region of Interest (ROI) Extraction: The filtered image I_{med} is then processed by an ROI extraction algorithm, which isolates the region corresponding to the electron bunch. First, the mean of the image I_{med} is calculated and is subtracted from the image, giving $I_{\text{sub}} = I_{\text{med}} - \mu_0$, where μ_0 is the mean of I_{med}. I_{sub} is then smoothed using a 7×7 Gaussian kernel G, producing

$$I_{\text{filt}} = I_{\text{sub}} * G. \tag{4.4}$$

Here, $*$ indicate the convolution operation.

The mean and standard deviation (μ_1, σ_1) of I_{filt} are computed, and a threshold is defined as

$$T = \mu_1 + \lambda\sigma_1, \tag{4.5}$$

where λ is a threshold factor that is tuned empirically for each dataset (typically $\lambda = 0.001$).

A binary mask $M(x, y)$ is created by thresholding I_{filt} as

$$M(x, y) = \begin{cases} 1 & \text{if } I_{\text{filt}}(x, y) \geq T \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

The largest connected component of this mask would correspond to the actual electron bunch. This region is selected, and the corresponding mask M_{ROI} is applied to the original image via an element-wise multiplication denotedby $\odot$ as

$$I_{\text{ROI}} = \max(I_{\text{med}} \odot M_{\text{ROI}}, 0). \tag{4.7}$$

This isolates the most prominent ROI while suppressing the background pixels.

3. Intensity Normalization: Finally, the ROI image is normalized to its maximum value,

$$I_{\text{norm}}(x, y) = \frac{I_{\text{ROI}}(x, y)}{\max(I_{\text{ROI}})} \tag{4.8}$$

This ensures all input images fall within a normalized range [0,1], which is critical for stable neural network training.

Data Augmentation

Preparing realistic training data is a decisive step for the success of the U-Net-based pulse reconstruction method. The main challenge lies in simulating the effect of the FEL lasing process on the longitudinal phase space images, particularly how the electron energy distribution changes during lasing.

Physically, not all electrons within a given temporal slice contribute equally to the lasing process. Only a subset of electrons undergoes energy loss due to coherent interaction with the FEL radiation field. This selective energy loss modifies the original energy distribution within each slice, producing a distinctive "lasing footprint" in the longitudinal phase space distribution.

To simulate this effect synthetically, the initial electron energy distribution for each time slice is modified by a two-Gaussian augmentation method:

Energy loss simulation: A Gaussian-shaped depletion is subtracted near the centroid of the initial energy distribution, representing electrons that lose energy to the FEL radiation.

Energy spread simulation: A broader Gaussian distribution with the same number of electrons is added at a lower energy offset, modeling the group of electrons shifted to lower energies and the increase in local energy spread resulting from lasing.

This two-Gaussian operation approximates the redistribution of electrons caused by energy transfer during the lasing process. Figure 4.6 illustrates this two-Gaussian image augmentation process.

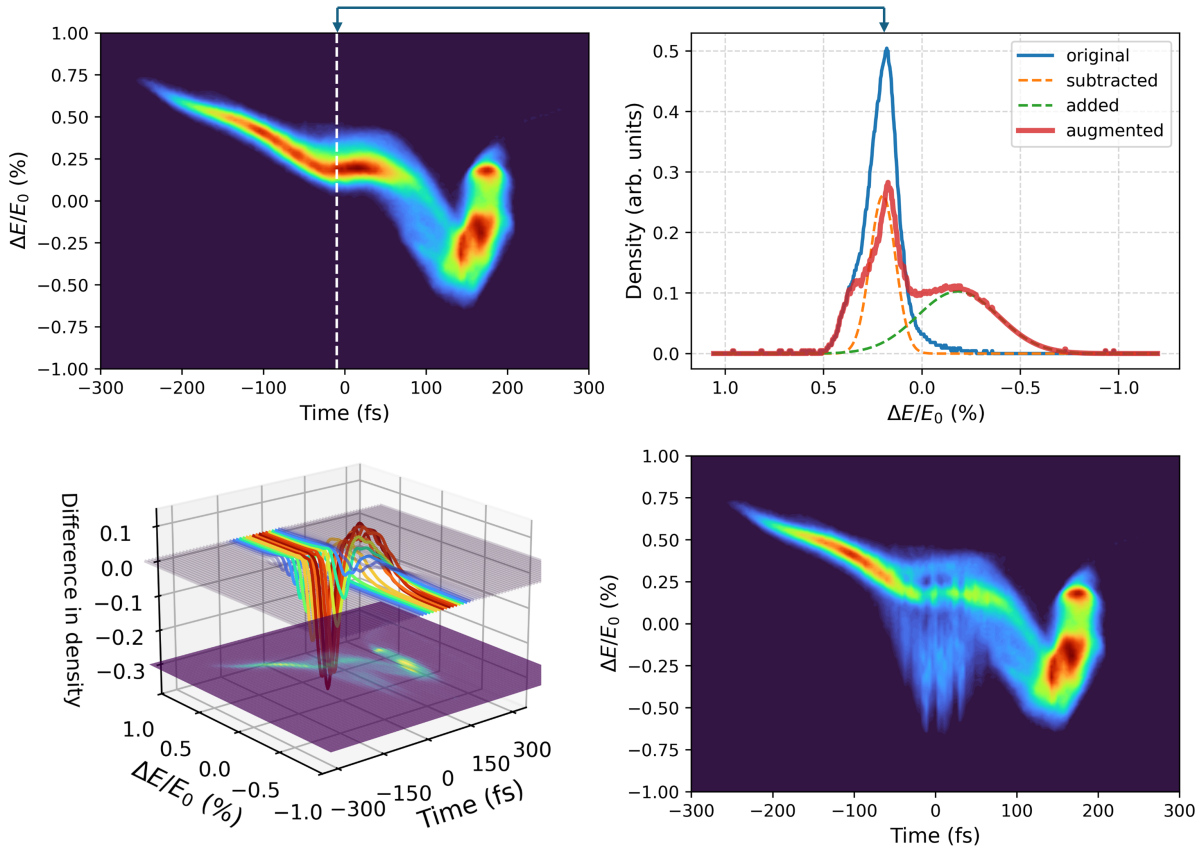

Figure 4.6: Visualization of the two-Gaussian augmentation method. Top left: measured lasing-off longitudinal phase space image. Top right: Energy distribution (blue curve) for the time slice indicated by the white dashed line in the top left image. A narrow Gaussian (orange dashed) is subtracted near the centroid to simulate energy-depleted electrons, while a broader Gaussian (green dashed) is added at lower energies to mimic the energy loss and increased spread due to lasing. The resulting augmented distribution is shown in red. Bottom left: Difference between the original and augmented phase space images across all slices. This 3D plot visualizes the subtracted density as the negative-going red 'trenches' and the added density as the positive-going 'ridges'. Augmentation is restricted to a specific time window, indicated by the colored curves. Bottom right: Final augmented image representing the synthetic lasing-on phase space distribution, produced by applying the augmentation slice-by-slice within the window.

To ensure the trained network is robust and generalizes well to real experimental data, a highly diverse set of training samples is required. This is achieved by randomly varying the parameters of the two Gaussians, including their peak amplitudes, widths, and the energy shift of the added Gaussian. Rather than attempting to model the complex correlations of the underlying FEL physics, which would be computationally intensive and require a precise machine model, a dataset with diversity is prioritized. The parameters are sampled from uniform distributions with bounds determined empirically.

The primary goal of this approach is not to perfectly replicate the complex FEL process, but to generate a dataset with sufficient diversity to train a robust model. By exposing the U-Net to a broad spectrum of plausible lasing footprints, the network learns to generalize effectively. This makes the final model less sensitive to the inherent shot-to-shot fluctuations of the accelerator and robust enough to handle real experimental data.

Importantly, lasing effects are localized in time and do not uniformly affect the entire electron bunch. To reflect this, the augmentation is restricted to a selected temporal window within the longitudinal phase space image. Outside this window, the image remains unchanged. The width, shape, and position of this augmentation window are also randomized to simulate the variety of pulse durations, shapes, and locations. By applying these controlled, physically motivated augmentations to experimentally measured lasing-off images, a large synthetic dataset of paired lasing-off and corresponding 'fake' lasing-on images is generated.

In addition to the two-Gaussian model described above, a second augmentation technique to capture the asymmetry observed in energy distributions after lasing is employed. This method convolves the original slice energy distribution with an asymmetric gamma function. The convolution shifts the centroid of the energy distribution toward lower energies while creating a pronounced long tail on the lower energy side, effectively replicating the increased energy spread and skewness introduced by the FEL interaction. Figure 4.7 describes the convolution-based image augmentation method.

The training dataset is constructed as a mixture of samples generated using both augmentation methods, the two-Gaussian subtraction/addition approach and the gamma-function convolution technique. Randomly applying these different methods during training enriches the diversity of simulated lasing effects and enhances the network's ability to generalize to real experimental data.

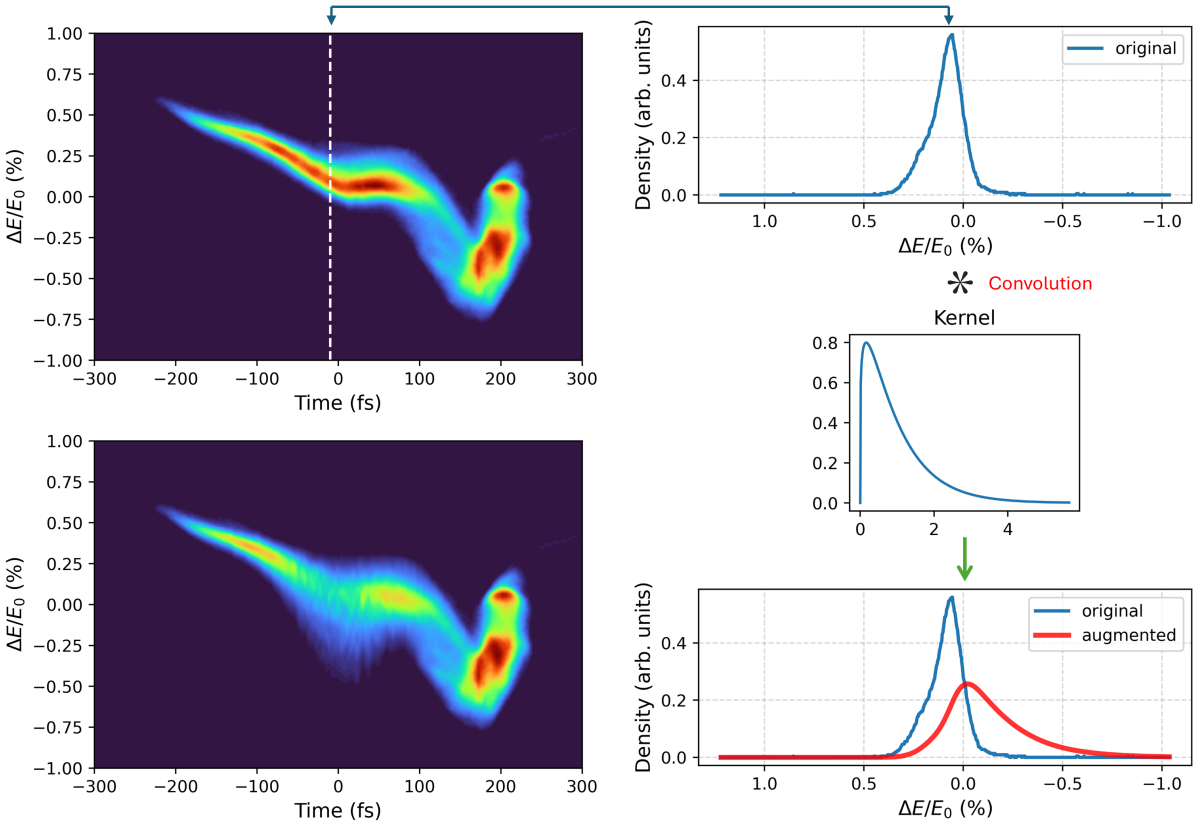

Figure 4.7: Visualization of the convolution-based augmentation method using an asymmetric kernel. Top left: Original lasing-off longitudinal phase space image. The vertical dashed white line marks the time slice used for demonstration. Top right: Energy distribution (blue curve) of the selected slice. Center right: Asymmetric convolution kernel (a decaying gamma-like function) designed to shift the centroid to lower energies and introduce a long tail, mimicking the energy spread from lasing. Bottom right: Resulting augmented energy distribution (red curve) after convolution, showing centroid shift and broadened spread relative to the original (blue). Bottom left: Final augmented phase space image after applying the convolution-based augmentation to slices within a selected time window, replicating the FEL-induced modulation.

To simulate realistic lasing-on images and expand the dataset, a multi-step data augmentation process is applied to each normalized lasing-off image I_{norm}.

1. Spatial Translation: First, to introduce variation and account for alignment uncertainties, a spatial translation is applied as

$$I_{\text{trans}}(x, y) = I_{\text{norm}}(x - t_x, y - t_y), \tag{4.9}$$

where t_x and t_y are chosen uniformly from a range of values appropriate for each dataset.

2. Simulating FEL-induced Energy Modulation: Next, a dynamic modulation is applied to simulate the localized energy change from FEL lasing. This is a column-wise operation, with a unique scaling factor $S(x)$ for each time slice (column) x. This factor is composed of a windowing function $W(x)$ and a random modulation waveform $R(x)$. Here, $W(x)$ acts as an envelope function, localizing the lasing process, while $R(x)$ defines the stochastic fluctuations within this window. Two types of windows are randomly chosen:

Smoothed Heaviside Window: This smooth approximation of a rectangular window is defined as

$$W_{\mathrm{H}}(x) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{\pi} \arctan\left(\dfrac{x - x_1}{\epsilon}\right), & x < c_0 \\ \\ 1 - \left(\dfrac{1}{2} + \dfrac{1}{\pi} \arctan\left(\dfrac{x - x_2}{\epsilon}\right)\right), & x \geq c_0 \end{cases}, \tag{4.10}$$

with $x_1 = c_0 - \frac{w}{2}$, $x_2 = c_0 + \frac{w}{2}$, where c_0 is the window center, w its width, and ϵ controls the smoothness of the transition.

Gaussian Window: Alternatively, a Gaussian envelope is used.

$$W_{\mathrm{G}}(x) = A \exp\left(-\frac{(x - c_0)^2}{2\sigma^2}\right), \tag{4.11}$$

where A is the amplitude, c_0 is the center, and $\sigma = \frac{w}{2.355}$, chosen so that w corresponds to the full width at half maximum (FWHM).

To make the augmentation more dynamic, a random modulation waveform is applied multiplicatively along the columns given by,

$$R(x) = \sum_{i=1}^{N} \left[a_i \sin(2\pi f_i x + \phi_i) + b_i \cos(2\pi f_i x + \phi_i)\right], \tag{4.12}$$

where a_i, b_i are the amplitudes, f_i the frequency, and ϕ_i the phase which are chosen uniformly random from a range of values suitable for each dataset. A window function $W(x)$ is randomly selected to be either $W_H(x)$ or $W_G(x)$. The final scaling factor per column is given by,

$$S(x) = W(x)(1 + R(x)). \tag{4.13}$$

Column-wise Augmentation: Each vertical column (energy slice) of the translated image, which we denote as $I_x(y) = I_{\mathrm{trans}}(x, y)$, is then modified using $S(x)$ and one of two techniques.

a) Gaussian Subtraction and Addition: This approach simulates the shift and broadening of the energy distribution by subtracting and adding Gaussian distributions. First, it

fits a Gaussian $G(y)$ to the original slice $I_x(y)$.

$$I_x(y) \approx a \cdot G(y), \quad \text{where } G(y) = \exp\left(-\frac{(y-y_0)^2}{2\sigma^2}\right). \tag{4.14}$$

A scaled version of this Gaussian is subtracted from the slice distribution, and a shifted Gaussian is added as

$$I_{\text{aug}}(y) = I_x(y) - a_{\text{sub}} G(y) + a_{\text{add}} \exp\left(-\frac{(y-(y_0+\Delta y_{\text{add}}))^2}{2\,(\sigma_{\text{add}})^2}\right) \tag{4.15}$$

where $a_{\text{sub}} = S(x) \cdot a$ is the amplitude of the subtracted Gaussian, a_{add} is the amplitude of the added Gaussian, $\Delta y_{\text{add}} = S(x) \cdot \Delta y$, where Δy is a randomly sampled shift, and σ_{add} is the standard deviation of the added Gaussian. The parameters a_{add} and σ_{add} are chosen so that the subtracted and added Gaussians have the same area under the curve.

b) Gamma Convolution: Alternatively, a convolution with an asymmetric kernel based on the gamma distribution is applied. The kernel is given by

$$K(y) = \frac{1}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-y/\theta}, \tag{4.16}$$

where α is fixed at a suitable value for the dataset (e.g. 1.2) and $\theta = S(x)$ is the scaling factor. The augmented column is given by

$$I_{\text{aug}}(y) = (I_x * K)(y). \tag{4.17}$$

These augmentation steps are repeated for each column x to produce the final augmented image, which mimics a lasing-on image.

3. Downsampling: After augmentation, each image is downsampled by a factor of 1/5 in both dimensions to reduce the computational load while training, while keeping a high enough resolution to resolve fs-level features. Each input–output pair (X, Y) corresponds to an augmented lasing-on-like image X and its original lasing-off counterpart Y.

From a dataset of N measured lasing-off images, each image undergoes M random augmentations, resulting in a training set of $N \times M$ pairs. For the results discussed in the following sections, the training set contained a total of 9380 training pairs based on 9 sets of lasing-off images. Some examples of the training pairs are shown in Fig. 4.8.

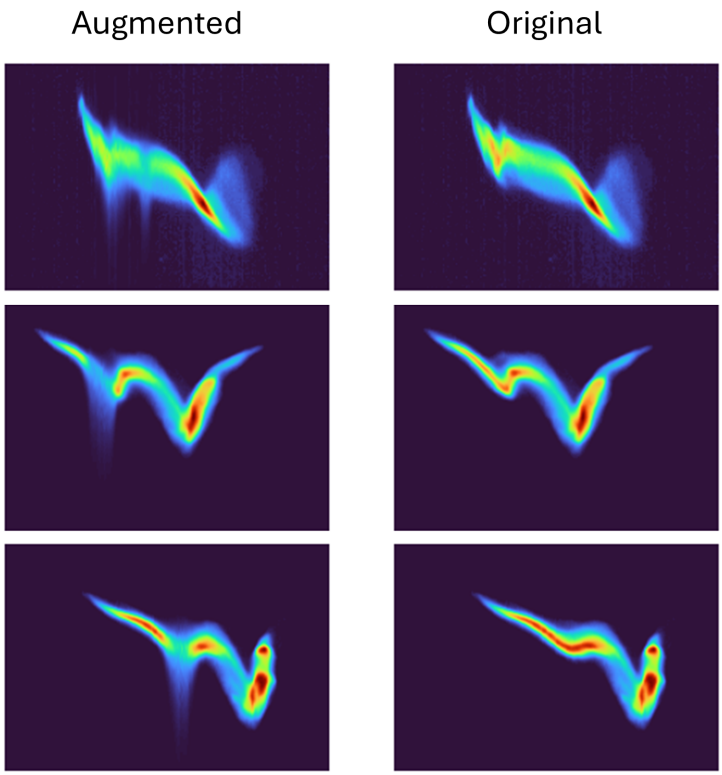

Figure 4.8: Three example training data pairs. The left column shows the augmented lasing-on image, which goes as the input to the network, and the right column shows the original lasing-off images, which act as the target outputs (labels) for the supervised training of the network.

4.4.3 Details of the Neural Network

To process and reconstruct images of the electron bunch from noisy input data, a modified U-Net architecture [105] was utilized. The U-Net is a fully convolutional network based on an encoder-decoder structure widely used for for image-to-image tasks such as image restoration and segmentation. The encoder gradually reduces spatial dimensions while increasing the number of represented features, allowing the network to learn hierarchical feature representations. The decoder then progressively reconstructs the spatial resolution of the output while combining low-level features from the encoder via concatenation. Figure 4.9 provides a visual schematic of the U-Net architecture used.

The network is constructed from a repeated "convolutional blocks" consisting of three main operations. A 2D convolution, Batch normalization, and ReLU activation [140]. The purple arrow in Fig. 4.9 represents this convolutional block.

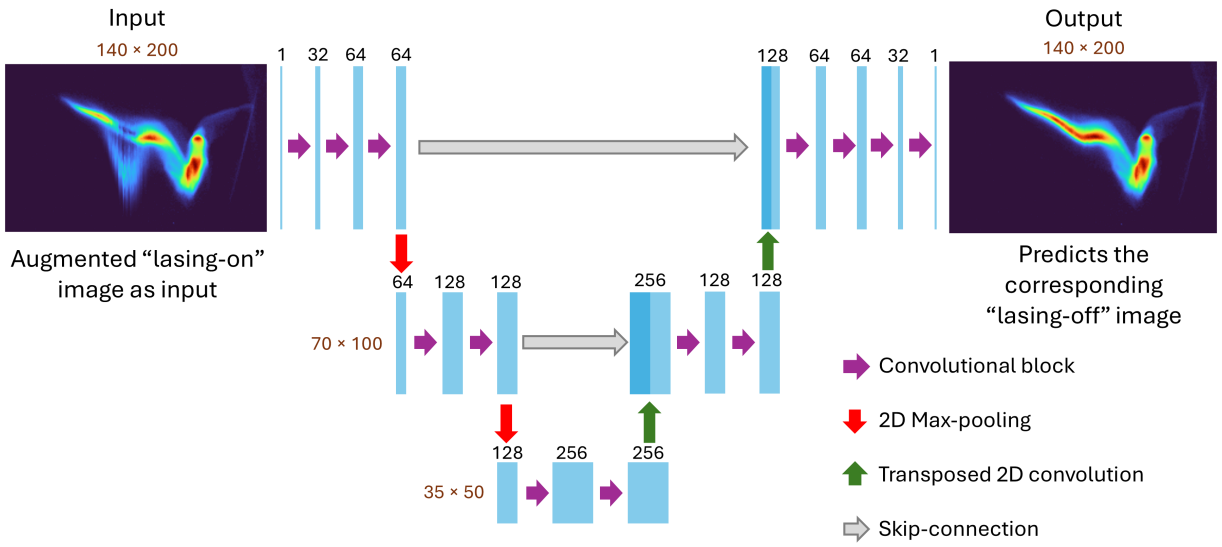

Figure 4.9: Detailed architecture of the modified U-Net model used. The encoder processes an augmented "lasing-on" phase space image (140 × 200 pixels) to a compact representation, and the decoder reconstructs the corresponding "lasing-off" image by progressively upsampling features and integrating low-level spatial information via skip connections.

Encoder

The encoder path, or contracting path, is shown on the left side of Fig. 4.9. Its purpose is to capture the context of the input image. The input, an augmented 140 × 200 image, is treated as having a single channel (i.e. grayscale). This input first passes through two convolutional blocks (as defined above). A convolutional layer applies a set of learnable filters to the image, where each filter is designed to detect a specific feature, like an edge or a texture. The output of one filter across the entire image is called a feature map. As shown in the diagram, the first convolutional layer expands the input from 1 channel to 32, and the second expands it from 32 to 64. In this context, the number of channels corresponds to the number of feature maps generated by the layer. This block, therefore, outputs 64 distinct feature maps, each representing different learned features, but retains the original 140 × 200 spatial dimensions. Next, a 2D max pooling operation (red arrow) downsamples the spatial resolution by a factor of 2, resulting in 70 × 100 feature maps. This step helps to reduce computation and make the feature representations more robust. This process is repeated where a second block of two convolutional layers is followed by another 2D max pooling operation that reduces the size to 35 × 50.

Bottleneck

At the bottom of the U-Net, two more convolutional layers with 256 channels each extract deep, high-level features. These layers form the bottleneck, where the network has the most compact and informative representation of the input.

Decoder

The decoder path, or expansive path (right side of Fig. 4.9), works to reconstruct the full-resolution output image. It begins with the 256-channel feature map from the bottleneck. The first decoder block uses a transposed 2D convolution [141] (green arrow) to upsample the data, doubling its spatial dimensions ($35 \times 50 \rightarrow 70 \times 100$) and halving its number of channels from 256 to 128.

A key feature of the U-Net is the skip connection [105] (grey arrow). This connection copies the 128-channel feature map from the corresponding encoder layer and concatenates (stacks) it with the 128-channel upsampled map. This merging, resulting in a 256-channel map, allows the decoder to directly access low-level spatial information like precise edge locations from the encoder, which would otherwise be lost. The resulting combined map is then processed by two sequential convolutional blocks. This entire upsampling block is repeated a second time as shown in the diagram.

Finally, to produce the output image, the feature map is passed through a few more convolutional blocks to reduce the feature map down to the single-channel, 140×200 grayscale image, which represents the network's prediction for the lasing-off image.

Training Configuration

The network was trained to minimize the pixel-wise mean squared error (MSE) loss between its prediction and the ground truth lasing-off image. The ground truth is the corresponding target image that the network is trying to learn to reproduce. To optimize the model, the Adam optimizer [92] was used, which is a popular and efficient stochastic optimization algorithm. The initial learning rate was set to 10^{-4} and a small weight decay of 1×10^{-5} was used to regularize the model and prevent overfitting by penalizing large weights [142]. Training was conducted for a total of 150 epochs, where one epoch represents a complete pass of the entire training dataset through the network.

The learning rate was managed by a learning rate scheduler, which dynamically adjusts the rate during training. The method used here monitors the validation loss and reduces the learning rate by a factor of 10 if the loss stops improving for 5 consecutive epochs. This helps to fine-tune the model's weights more precisely once it is close to a good solution.

4.4.4 Results

The neural network was trained exclusively on a synthetic dataset generated from experimentally measured lasing-off longitudinal phase space images. The central objective was to enable the model to generalize this image reconstruction process to real lasing-on images. To evaluate this generalization, a subset of measured lasing-on images was selected. These were drawn from datasets corresponding to machine conditions similar to those used in generating the synthetic training data. The selected lasing-on images were fed into the trained network, which predicted their corresponding lasing-off representations. From the difference in slice mean energies between the two, the power profiles of the FEL pulses were computed.

Examples of these reconstructed images and their resulting power profiles are shown in Fig. 4.10 and Fig. 4.11. The results clearly demonstrate that the network successfully identifies and removes lasing-induced features, effectively reconstructing a proxy lasing-off image. Notably, the model is able to detect multiple lasing peaks at different temporal positions within the bunch, and the shape of the resulting power profile aligns well with intuitive expectations based on the visible structures in the phase space images.

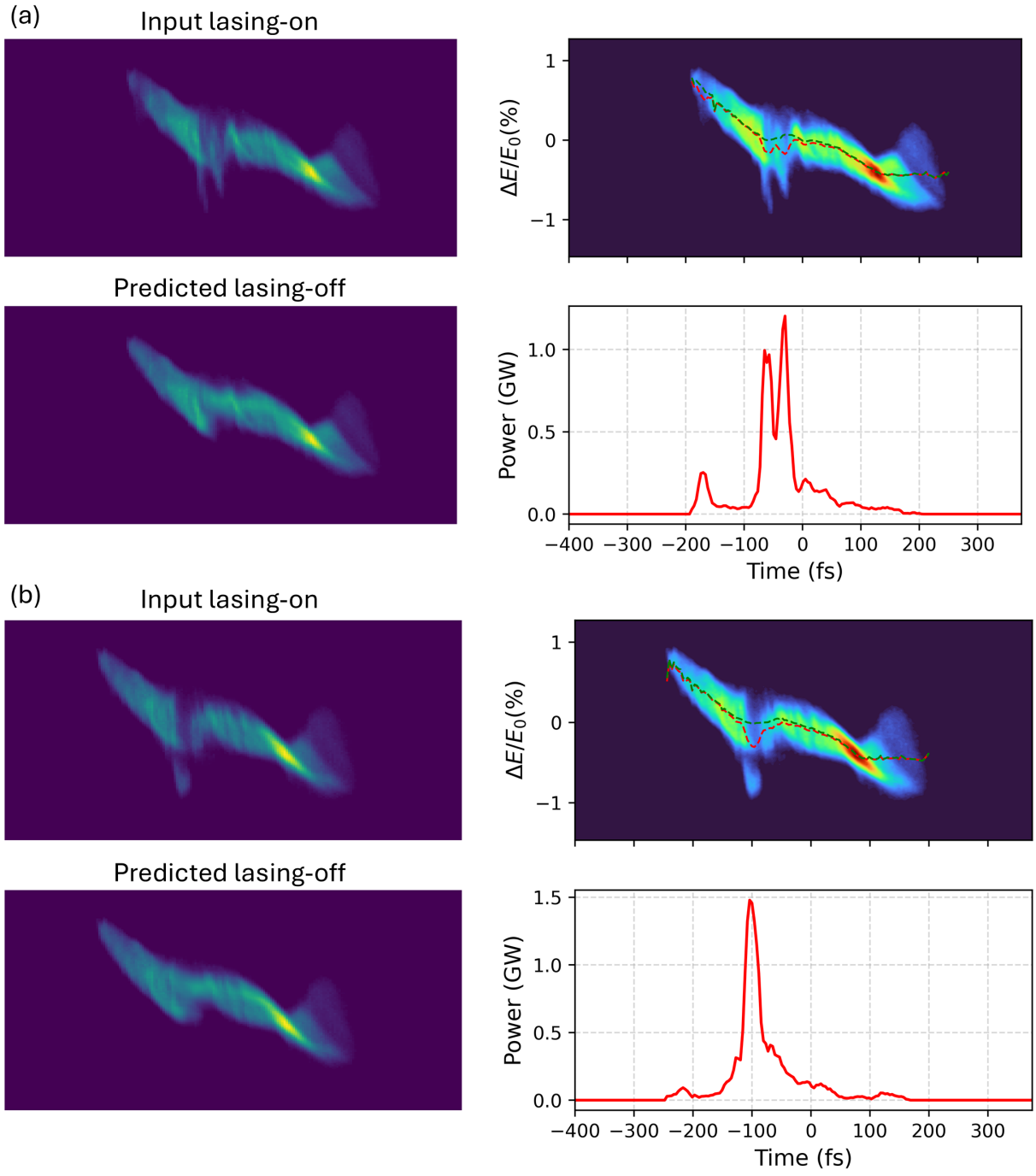

Figure 4.10: Two example results, (a) and (b), of the time-profile reconstruction of the FEL pulses using the U-Net architecture with real measured lasing-on images given as input. Top left: The measured lasing-on image. Bottom left: Predicted lasing-off image from the neural network. Top right: Lasing-on image overlaid with the slice centroids plotted for the lasing-on (red dashed curve) and predicted lasing-off (blue dashed curve) images. Bottom right: Calculated power profile from the difference between the energy centroids.

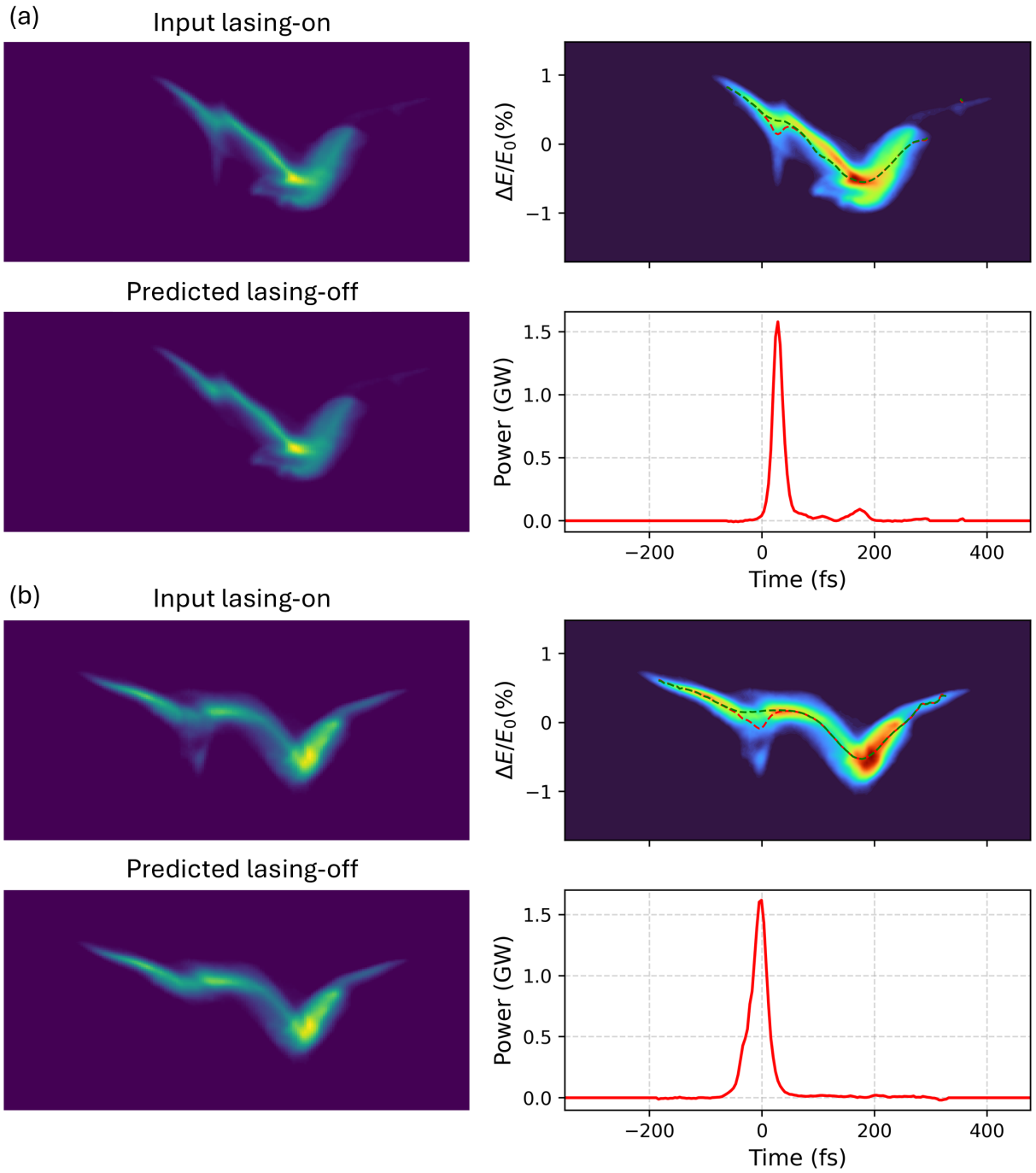

Figure 4.11: Two example results, (a) and (b), of the time-profile reconstruction of the FEL pulses using the U-Net architecture with real measured lasing-on images given as input. Refer to Fig. 4.10 caption for more details.

To quantitatively assess the model's performance, the pulse energy was calculated by integrating the area under the predicted power profile. These model-derived pulse energies were then compared to independent pulse energy measurements from the Gas Monitor Detector (GMD). Figure 4.12 shows this comparison. The predicted pulse energies exhibit a high degree of correlation with the GMD measurements. However, a systematic deviation from the identity line is observed in the top-left dataset. This indicates that while the network accurately captures the relative fluctuations, the absolute energy values may be subject to a systematic offset. This could be due to differences in the machine conditions between the training and test data, leading to minor artifacts in the results. However, the high correlation validates the model's ability to accurately reconstruct the temporal power distribution and capture the shot-to-shot variations in the total energy of the FEL pulse.

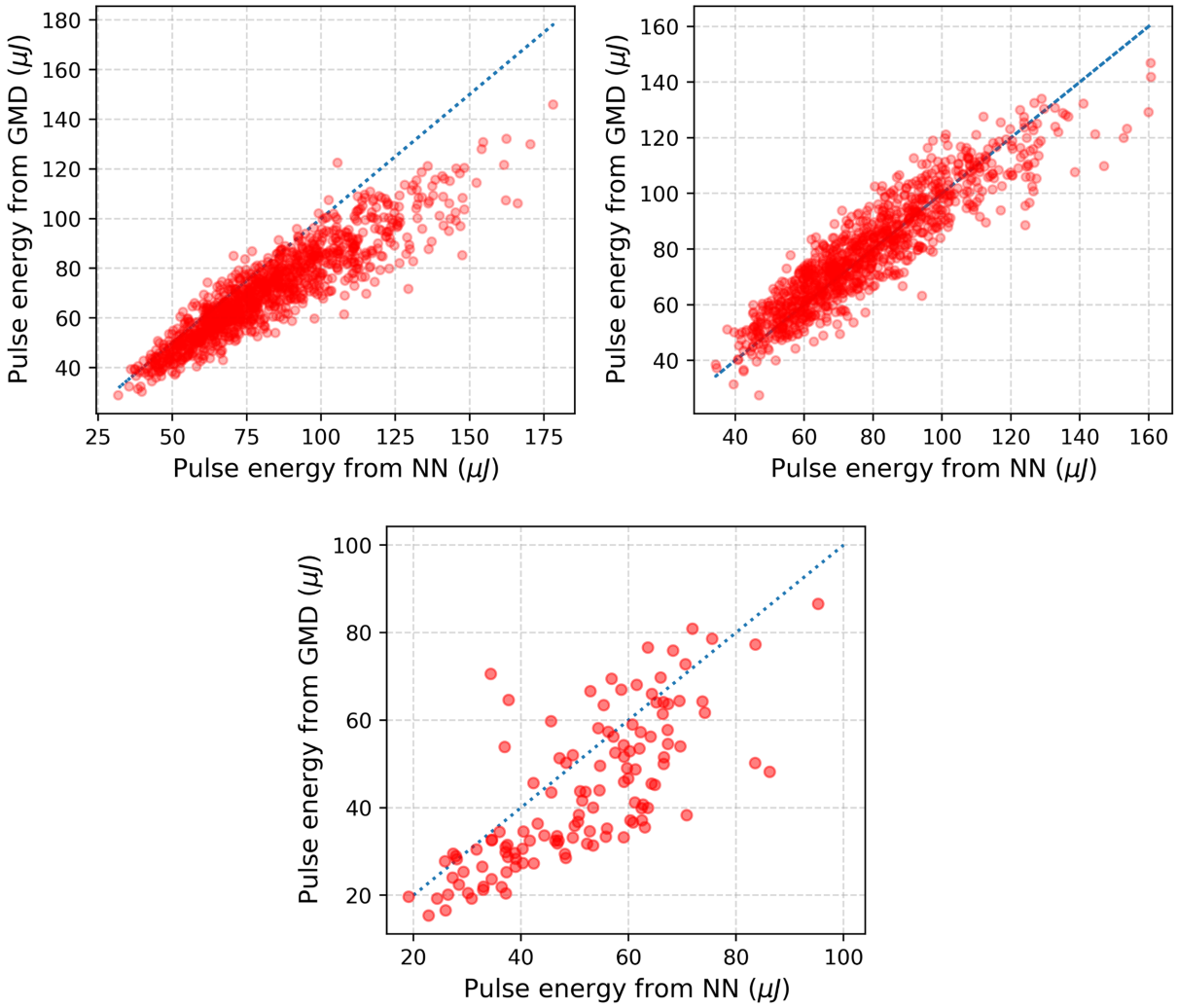

Figure 4.12: Comparison of pulse energies estimated using the neural network (NN) and measured using the GMD for three different datasets from November 8, 2023 (bottom), and January 25, 2024 (top). The dashed line marks the identity line.

Comparison with the Standard Method

To evaluate the performance of the U-Net–based reconstruction relative to the standard approach based on Eq. 4.1, five datasets recorded during two shifts at FLASH (November 2023 and January 25, 2024) were analyzed. For the standard method, several lasing-off reference images were acquired by suppressing the lasing process (achieved by steering the electron beam). The mean of these lasing-off images was then used as the reference for reconstructing the pulse power profiles.

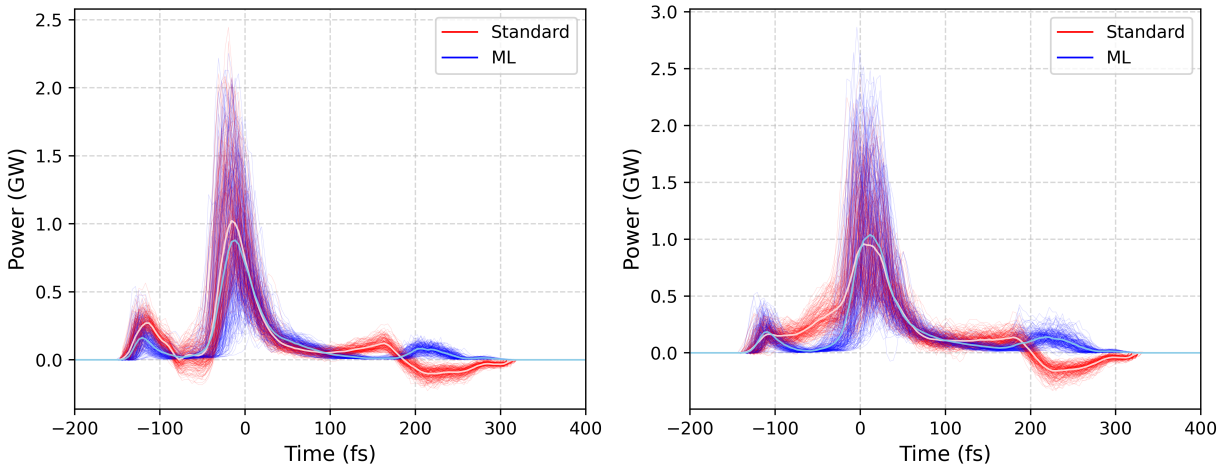

Figure 4.13: Examples of the calculated FEL pulse power profiles using both the standard (red curves) and machine learning-based (blue curves) methods. The plot shows results from two example datasets from January 25, 2024. Only 200 examples are plotted here for ease of visualization. The thicker, solid lines represent the average profile for each method, while the thin, semi-transparent lines show the individual shots.

Figure 4.13 presents comparisons of the pulse profiles obtained with both methods on two datasets recorded on January 25, 2024. To facilitate an objective comparison, the power profiles were normalized using pulse energy measurements from the GMD. Specifically, the raw power profile for each method was rescaled by the ratio of the GMD-measured energy to the area under its curve as

$$P_{\text{norm}}(t) = P(t) \times \frac{E_{\text{GMD}}}{\int P(t)dt}. \tag{4.18}$$

Inspection of the results reveals that both methods reliably capture the main intense peak, and they also identify a secondary peak around −100 fs. However, the standard method exhibits noticeable artifacts, in particular, a negative power region beyond 200 fs. These artifacts likely arise from mismatches between the lasing-on images and the fixed lasing-off reference images. In contrast, the U-Net–based approach generates a lasing-off image dynamically from each lasing-on input, more effectively compensating for such mismatches.

This is confirmed by the absence of negative-power artifacts in the ML-derived power profiles. While the ML method avoids these impossible values, it produces a small positive feature in the same region. This is likely a residual artifact, but even then it represents a significant improvement over the unphysical result of the standard method.

Furthermore, because the negative artifacts in the standard method lead to an underestimation of the integrated pulse energy, normalization by the GMD energy consequently overestimates the peak power. Thus, the deep learning–based reconstruction not only provides cleaner power profiles but also more accurate quantitative measurements of both the pulse energy and peak power.

To further assess the consistency between the two methods, key characteristics of the reconstructed power profiles such as the peak power (after normalization using GMD pulse energy) and the full width at half maximum (FWHM) of the central peak were compared. Since many of the pulses exhibit multiple peaks or substructures, the FWHM was defined as the total width of the pulse at half of its maximum value, encompassing all prominent features (see Fig. 4.14).

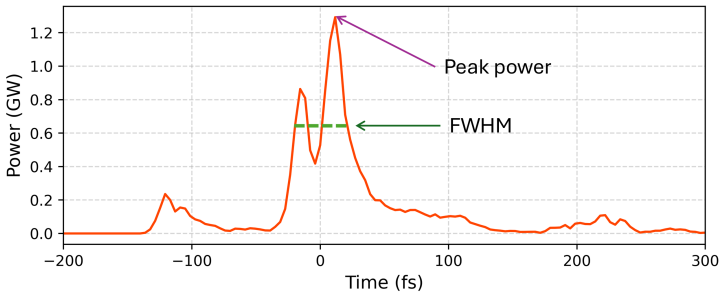

Figure 4.14: Temporal power profile of a measured FEL pulse. The plot illustrates the definition of peak power and FWHM pulse duration.

Figure 4.15 presents a comparison of these quantities between the standard and U-Net–based approaches. The peak power estimates from both methods agree well, showing a strong linear correlation. The FWHM values also exhibit a linear trend; however, the standard method generally yields slightly broader pulses. This could be attributed to the averaging involved in creating a static lasing-off reference, which may smooth out fine temporal features.

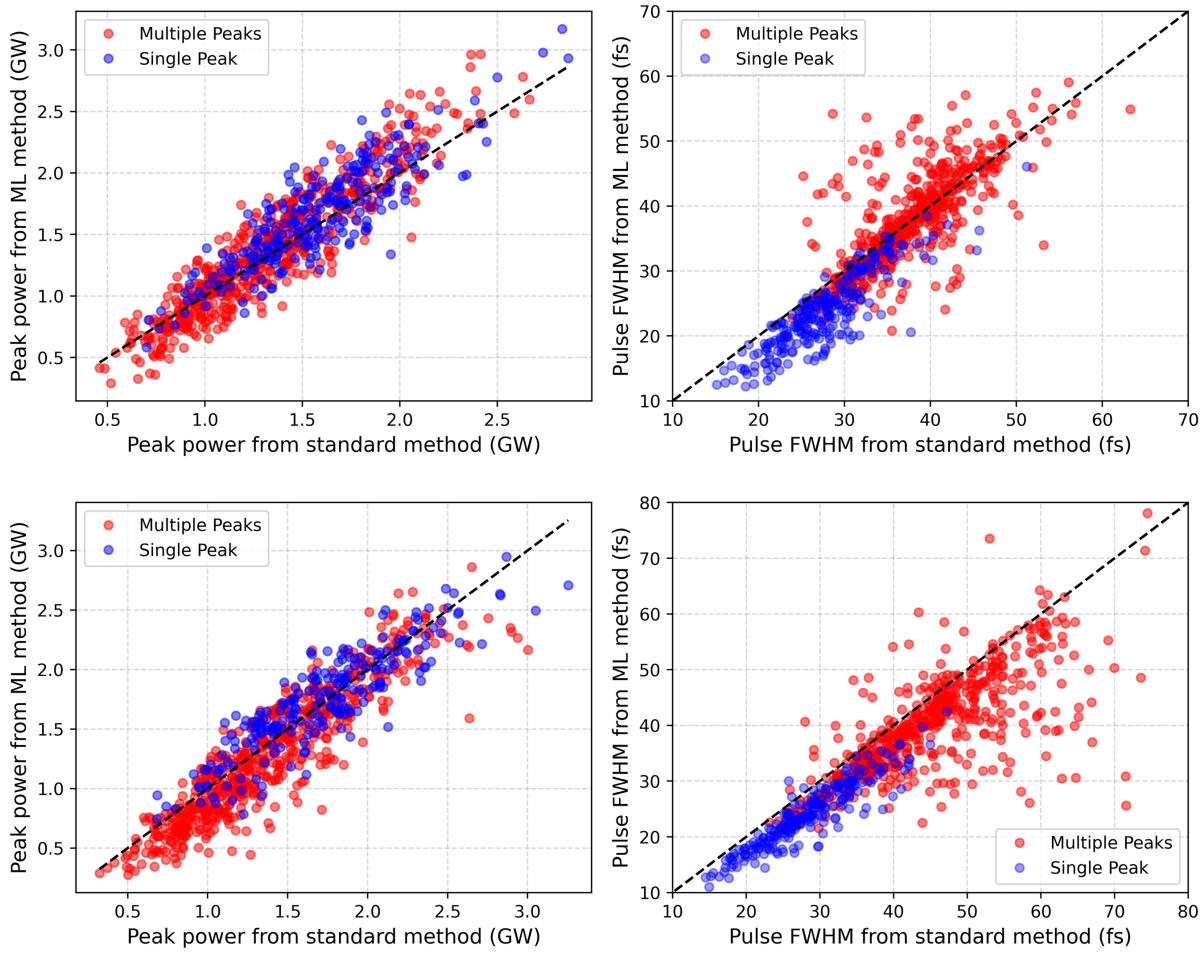

Figure 4.15: Comparison of pulse characteristics extracted using the standard and ML-based methods for two datasets from January 25, 2024. Left: Peak power after GMD normalization. Right: FWHM of the reconstructed pulse. Blue dots indicate single-peaked pulses, red dots indicate pulses with multiple peaks. The dashed lines mark identity.

Similarly, Fig. 4.16 presents the reconstructed FEL pulse power profiles for three datasets recorded during a shift on November 8, 2023. In these cases, the standard method consistently introduces artifacts in the reconstructed profiles, particularly beyond 100 fs. The presence of such artifacts inflates the total integrated energy. Since the profiles are normalized to match the GMD-measured pulse energy, the normalization process suppresses the actual peak amplitude, leading to a systematic underestimation of the peak power. In contrast, the machine learning–based method avoids this issue entirely. The resulting reconstructed profiles are clean, with single well-defined peaks and no artificial structures.

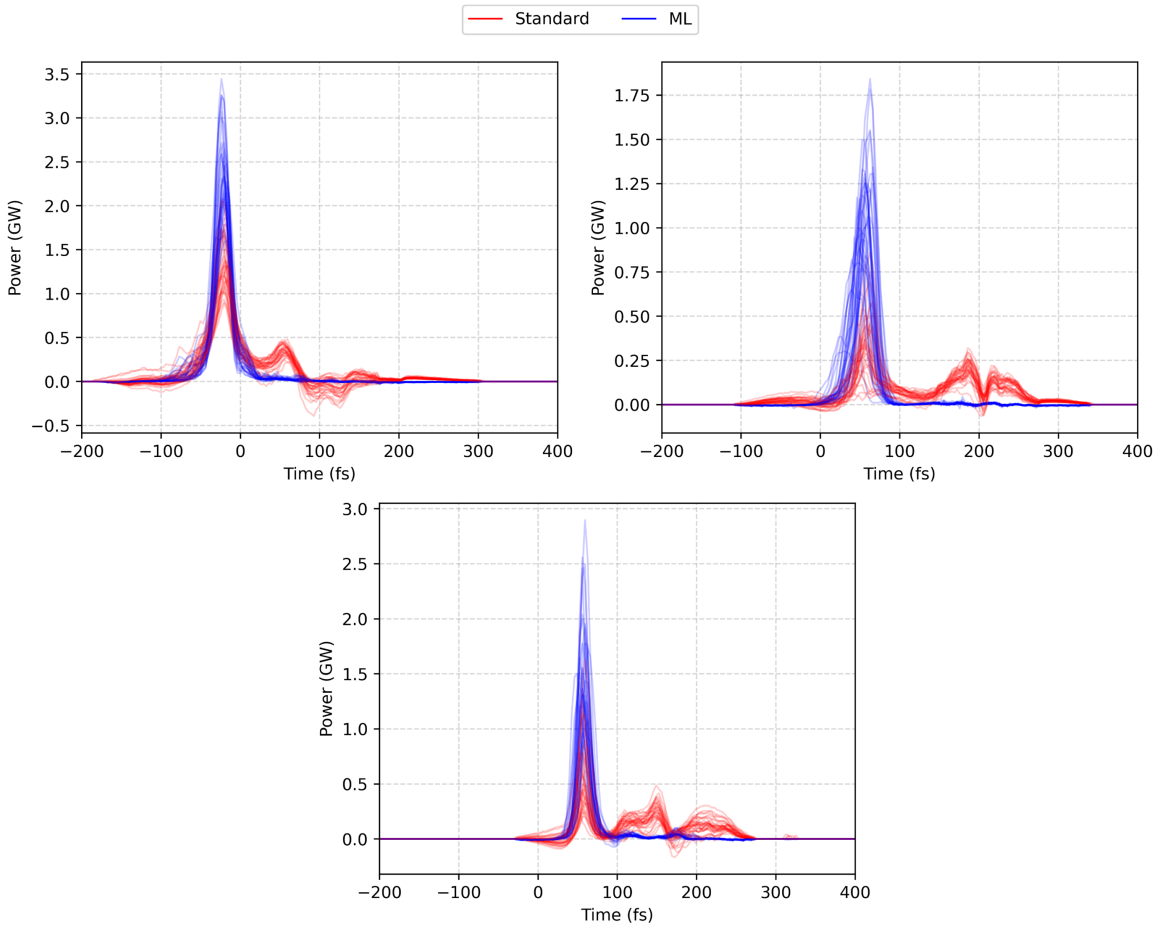

Figure 4.16: Examples of the calculated FEL pulse power profiles using both the standard (red curves) and machine learning based (blue curves) methods. The plot shows results from three datasets from November 08, 2023.

Figure 4.17 provides a quantitative comparison of the peak power and full width at half maximum (FWHM) derived from the reconstructed power profiles using both methods. Although the results from the two approaches exhibit a general linear trend, there is a clear bias. The standard method generally yields a lower peak power and a higher FWHM compared to the ML-based method. This discrepancy can be attributed to the presence of non-physical features in the standard reconstructions (as shown in Fig. 4.16), which distort the shape of the power profile. Such artifacts, for example, can artificially broaden the pulse and suppress the main peak, which would explain the apparent underestimation of peak power and overestimation of pulse width. The ML-based method, by contrast, avoids these artifacts and is therefore believed to offer a more physically realistic representation of the actual pulse structure.

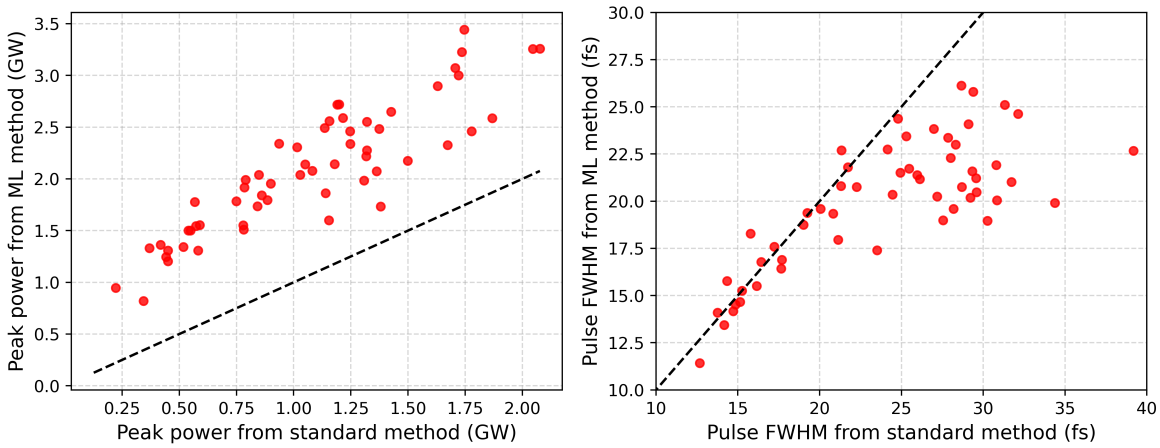

Figure 4.17: Comparison of peak power (left) and FWHM (right) extracted using the standard and ML-based methods for three datasets from November 8, 2023. The dashed line marks the identity line.

4.4.5 Generalization and Transfer Learning

A key goal of the method discussed above was to enable the trained model to generalize the lasing-off image reconstruction to arbitrary electron bunch shapes. Ideally, once trained, the model should be capable of predicting the lasing-off counterpart for any given lasing-on image, regardless of the underlying bunch structure. However, this generalization turned out to be challenging in practice. The training data consisted of a limited number of lasing-off image sets, which could not fully represent the diversity of possible bunch configurations encountered in real operation.

As a result, the trained network may become biased toward the specific features present in the training samples. While this behavior is advantageous when the data for analysis shares similar characteristics, it poses a problem when analyzing images from significantly different bunch shapes. In such cases, the network may misinterpret unfamiliar structures or produce an inaccurate lasing-off prediction, leading to errors in the final power profile.

To address this limitation, transfer learning offers a practical solution. After the model has learned to remove lasing-related structures, it can be further fine-tuned using a small, targeted dataset that better represents the new bunch conditions. This approach involves collecting a set of real lasing-off images from the new bunch conditions, generating a corresponding synthetic lasing-on dataset, and then continuing the training process using this new data.

By fine-tuning on a dataset that closely matches the conditions of the new machine setup and the new bunch shape, the network adapts its internal representation to better handle the specific new features. This allows it to generate more accurate lasing-off images and more reliable power profiles under varying machine conditions.

For online diagnostics, a small set of lasing-off images can be recorded at the start of the experiment. These can be used to create training samples tailored to the specific electron bunch shape. It was observed that a few thousand samples are usually sufficient. The pretrained neural network can then be fine-tuned on this new dataset and used for shot-to-shot diagnostics during the rest of the experiment. In practice, generating the training data and fine-tuning the model typically takes only 20 to 30 minutes.

To test this, the previously trained U-Net model was fine-tuned using data recorded at FLASH on March 02, 2023. Figure 4.18 shows an example of the rough bunch shape for that configuration and the corresponding generated lasing-on image. For the fine-tuning, a set of approximately 2900 training pairs was used, generated through image augmentation from measured lasing-off images.

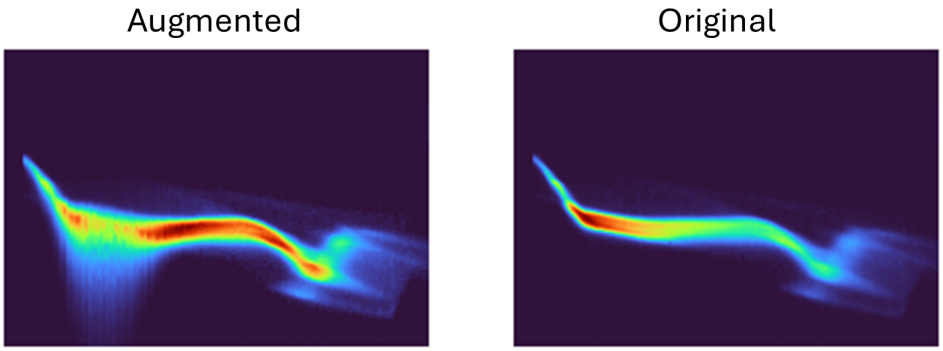

Figure 4.18: An example of the augmented lasing-on image (left) and the original lasing-off image (right) used for fine-tuning.

Some example results are shown in Fig. 4.19, which presents the reconstructed FEL pulse profiles for two different datasets. These results demonstrate that the fine-tuned model can adapt to new bunch conditions and accurately remove lasing features without introducing unwanted artifacts.

Figure 4.20 compares the pulse energies calculated from the neural network's reconstructed power profiles with the pulse energies measured by the GMD. The results show a general linear correlation, indicating that the model predictions are in good agreement with the actual measurements. However, a larger spread in the predicted values is observed at higher pulse energies. This may be due to the longer pulse lengths, which can make it more challenging for the network to accurately reconstruct the corresponding lasing-off images.

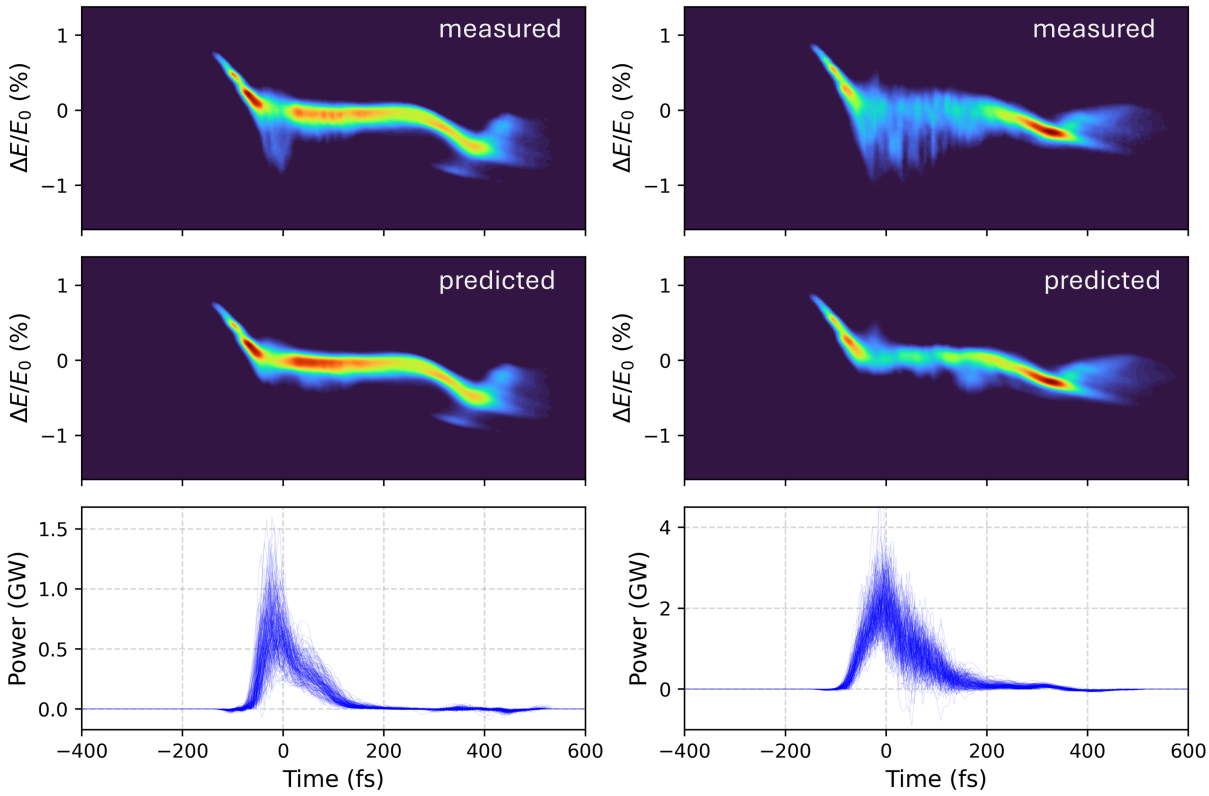

Figure 4.19: Measured longitudinal phase space distribution of the electron bunches (top row) and the predicted lasing-off images (middle row), and the reconstructed pulse power profiles for 200 shots using the U-Net-based method (bottom row).

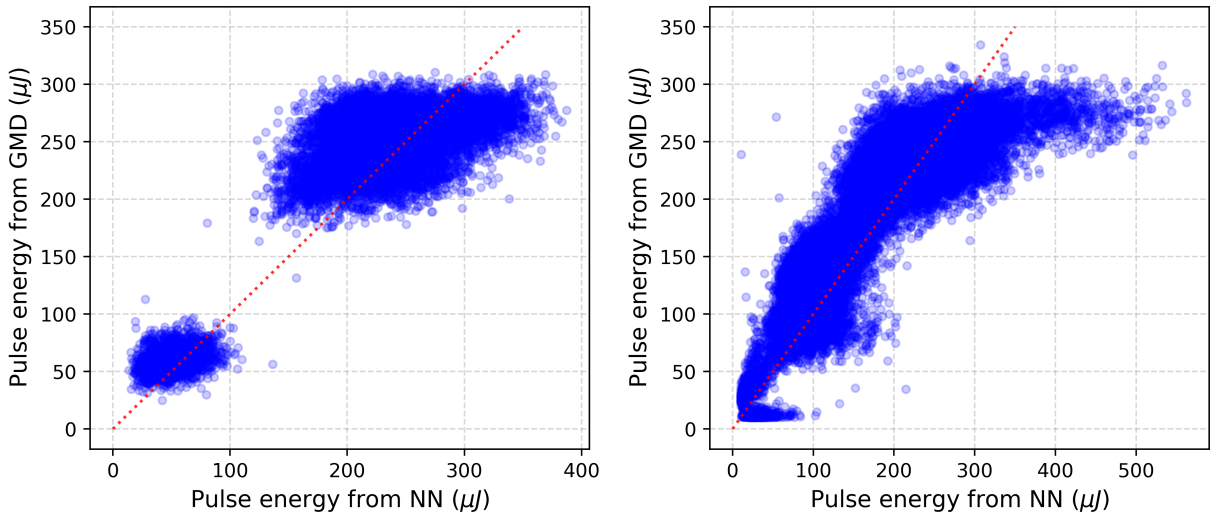

Figure 4.20: Comparison of pulse energies extracted using the ML-based method and measured using the GMD for two datasets from March 2, 2023. The dashed lines mark identity.

4.4.6 Drawbacks and Limitations

As seen in Fig. 4.20, the predictions from the neural network begin to show a noticeable spread as the FEL pulse energy increases. Inspection of the samples indicates that these high-energy shots correspond to longer pulses (a representative example is shown in Fig. 4.19, right). This implies that the network struggles to accurately reconstruct lasing-off references for long pulses, where most or all of the electron bunch contributes to lasing. A few such failure cases are shown in Fig. 4.21. In these examples, the network either overcorrects the lasing signal or fails to fully remove lasing features, leading to a faulty estimation of the reconstructed power profile. A comparison of the pulse energies calculated from the predicted pulse profiles for a set of long and short FEL pulses in relation to the GMD pulse energy measurement is given in Fig. 4.22. It is observed that the samples where most of the bunch is lasing instead of a localized FEL pulse, the calculated pulse energy doesn't match with the GMD measurements but instead is underestimated.

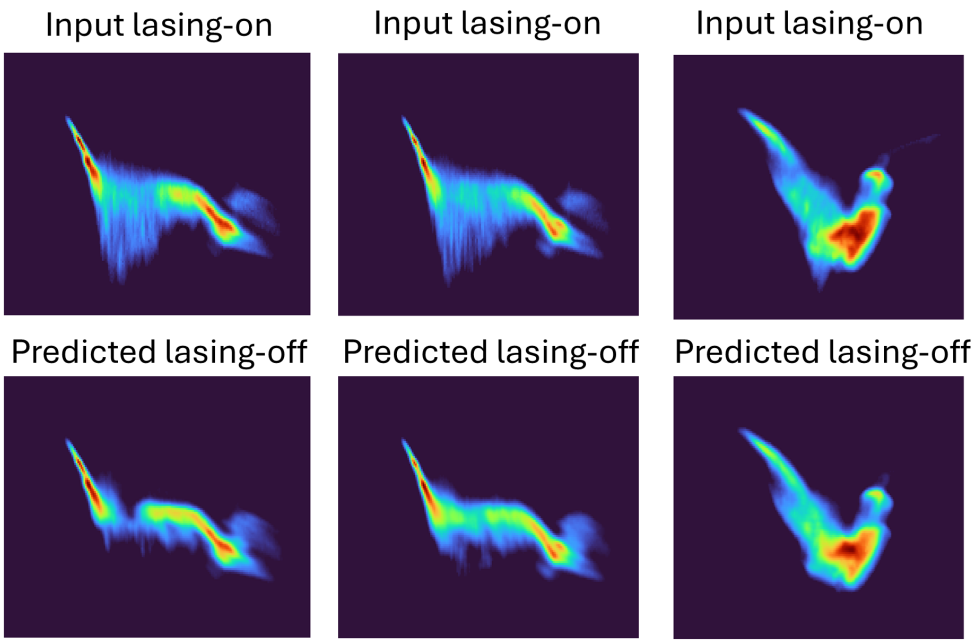

Figure 4.21: Three examples of the raw input lasing-on images (top row) and the predicted lasing-off images (bottom row) where the network fails to generate a good lasing-off reference. The failure is visible in the predicted images where the network creates a gap in the bunch profile (left) or fails to fully remove the lasing features (middle and right), leaving behind significant residual structures that should not be present in a lasing-off reference.

The network typically relies on the presence of non-lasing regions within the bunch as a reference to help reconstruct the lasing-off image. When the entire bunch is lasing, there is no clear baseline for comparison, making it difficult for the network to distinguish and remove the lasing contributions. This challenge is not unique to the network, as human

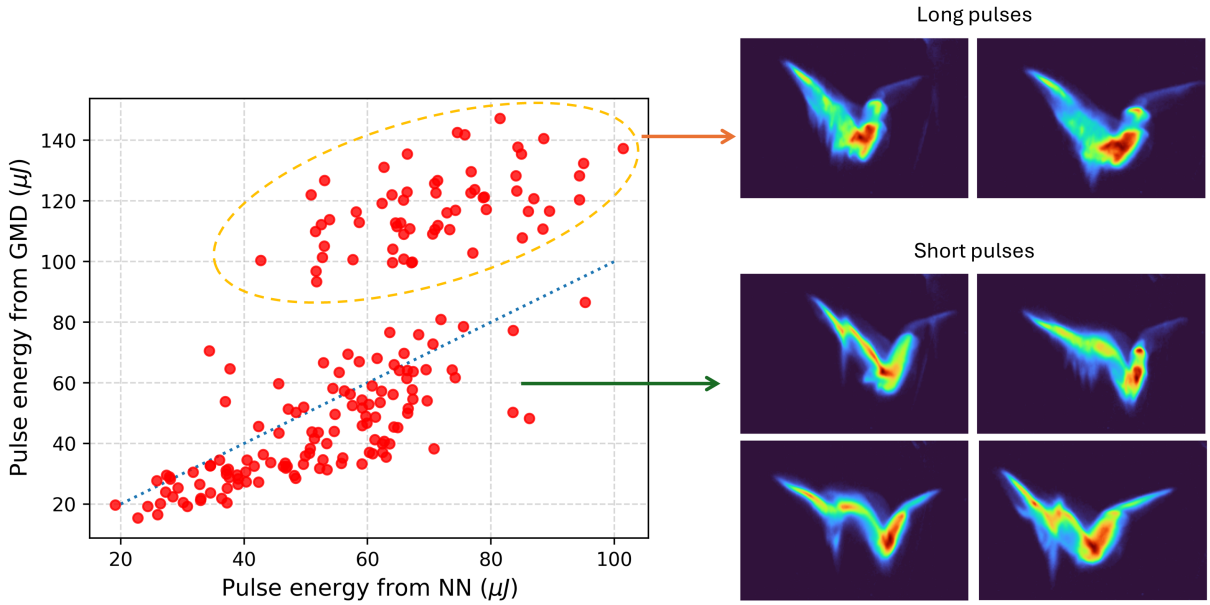

Figure 4.22: Comparison of the pulse energy calculated from the predicted power profiles and the GMD measurements (left) for six different datasets. The measured lasing-on images are shown on the right. Cases where the lasing is localized show better agreement with the GMD values, while datasets where the entire bunch is lasing (highlighted by the dashed orange circle) show larger deviations. The network underestimates the pulse energies in this case.

interpretation would face similar difficulties without a known lasing-off reference. This limits the method's applicability for diagnosing long FEL pulses, but remains highly effective for the experiments where the goal is to produce short, femtosecond-scale pulses. In this regime, the method provides a reliable and practical tool for reconstructing pulse profiles.

Another limitation of the network is that it can sometimes misinterpret sharp features in the phase space density as lasing signatures, even when those features are also present in the lasing-off image. An example of this is shown in Fig. 4.23, where the network identifies two sharp features around -200 fs and 100 fs as lasing, resulting in two artificially strong peaks in the reconstructed temporal power profile. However, by inspecting the corresponding lasing-off phase space image (Fig. 4.24), it becomes clear that these features are intrinsic to the electron bunch and not caused by the lasing process. In this case, the network misses the actual lasing region between -200 fs and 0 fs, while incorrectly flagging non-lasing features as lasing. This misclassification comes from a bias in the training data, which is heavily skewed toward examples with short, localized FEL pulses. Even with fine-tuning, it is difficult to overcome this problem unless a significantly larger and more diverse dataset is used, or the training is continued for a longer period to help the network learn that such features can occur in both lasing-on and lasing-off conditions. This creates a problem for using the method in real-time diagnostics. If the network needs to be retrained for longer every time

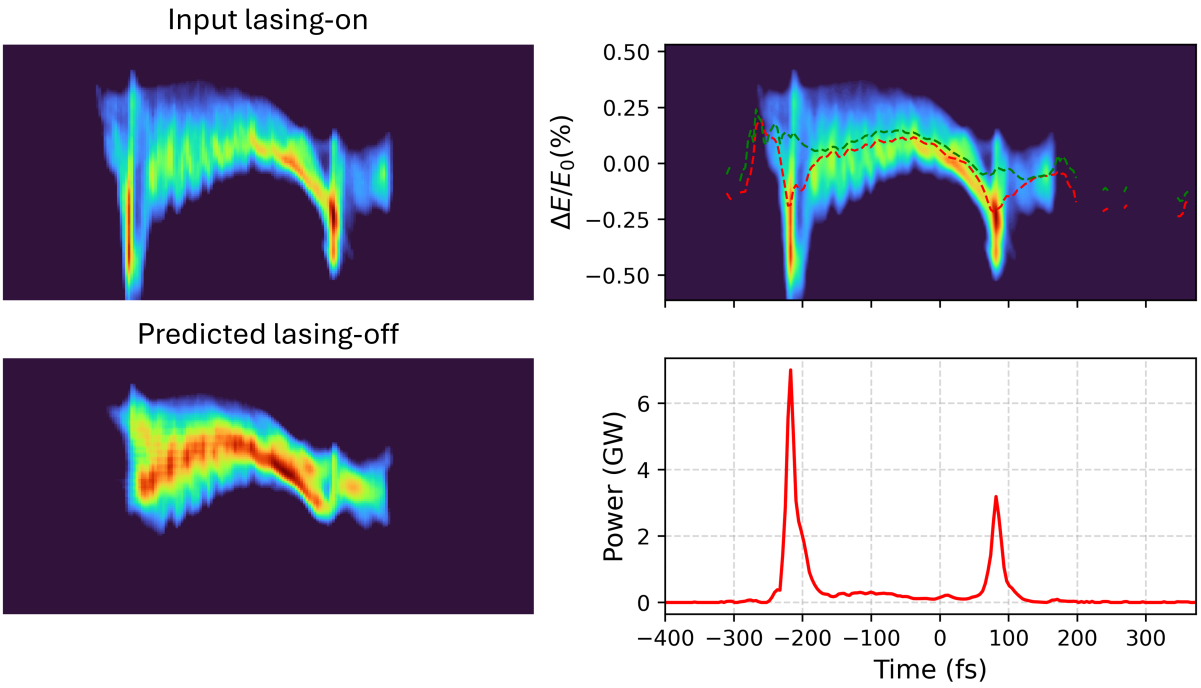

Figure 4.23: An example of the phase space reconstruction with the U-Net-based model where it fails. The sharp power peaks at around -200 and 100 fs are artificial and not part of lasing.

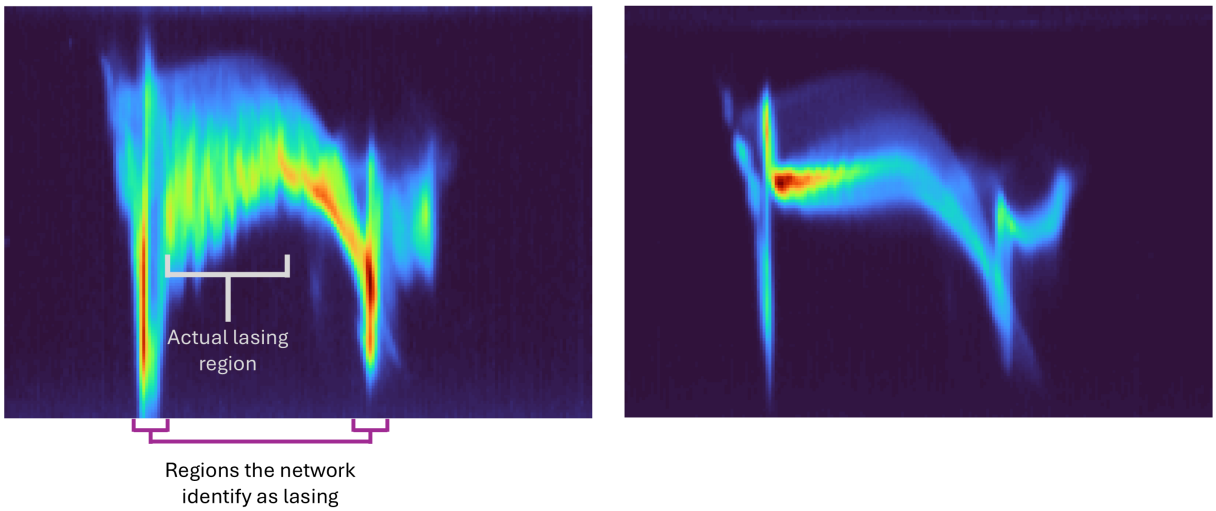

Figure 4.24: A comparison between the measured lasing-on (left) and lasing-off (right) images. It is clear that the two sharp features are inherent to the electron bunch, and not coming from lasing. The real lasing region appears between the two peaks, which the network fails to recognize.

the electron bunch shape changes, it becomes too slow and impractical. The time and effort needed to prepare new training data and retrain the model take away the advantage of this method. These limitations mainly arise because the model has no access to the actual electron bunch shape or the corresponding lasing-off reference image. If the network could somehow incorporate a measured lasing-off image, essentially learning the features present in it, it could produce a more accurate reconstruction of the phase space distribution, which closely resembles the measured lasing-off images. With this goal in mind, a new neural network architecture was developed that allows the measured lasing-off image to be used as an additional input.

4.5 Y-Net-Based Reconstruction of FEL Pulse Profiles

The 'Y-Net' model represents an advancement over the U-Net approaches by utilizing both a lasing-on image and a lasing-off reference image to predict the corresponding lasing-off image. This dual-input strategy eliminates the need for dataset-specific network fine-tuning, as the reference image provides explicit information about the expected lasing-off image structure. Additionally, this approach enables the network to better distinguish between features inherent to the electron bunch and those associated with lasing, thereby reducing reconstruction artifacts.

4.5.1 Outline of the Method

The Y-Net method builds upon the previously described U-Net approach, which relies solely on lasing-on images as input. The image augmentation procedures remain identical to those outlined in the previous section. The difference here is that, for each augmented lasing-on and lasing-off pair, a reference lasing-off image was randomly selected from the lasing-off dataset. This random selection prevents the network from learning image-specific artifacts and ensures better generalization. The randomly selected reference image was also constrained to be a different image from the target (ground-truth) lasing-off image. This constraint reflects real-world measurement conditions where perfect alignment between reference and target images is not achievable. So the network must learn to extract shape and distribution information from the lasing-off reference while maintaining spatial correspondence with the lasing-on input.

Due to inherent shot-to-shot timing jitter and other experimental variations, the lasing-off reference needs to be aligned with the lasing-on input before training and inference. This alignment makes it easier for the network to learn spatial correlations. The alignment is performed by calculating the cross-correlation between the images and applying a shift based on the coordinates of the maximum correlation value.

Let $I_{\text{on}}(x, y)$ and $I_{\text{ref}}(x, y)$ denote the lasing-on and reference lasing-off images, respectively. The goal is to align I_{ref} to I_{on} by finding the shift $(\Delta x, \Delta y)$ that maximizes their similarity. The cross-correlation $C(i, j)$ between the two images is calculated as

$$C(i, j) = (I_{\text{on}} \star I_{\text{ref}})(i, j), \tag{4.19}$$

where $\star$ denotes the 2D cross-correlation operation, defined as a sliding dot-product. This operation measures the similarity between I_{on} and I_{ref} at every offset (i, j). An example of the cross-correlation matrix calculated this way is shown in Fig. 4.25.

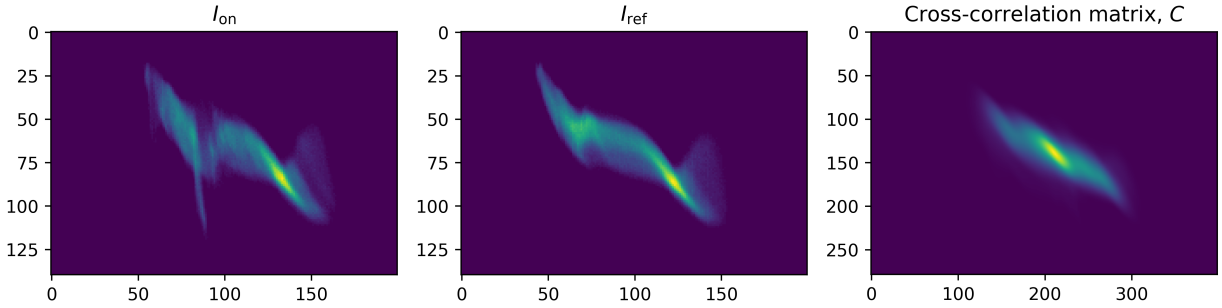

Figure 4.25: An example of the cross-correlation matrix (right) between a lasing-on image (left) and a lasing-off reference (middle). The shift required to align the two images can be calculated from the coordinates of the maximum of the cross-correlation matrix.

The spatial position (i^*, j^*) of the maximum value in the correlation matrix $C(i, j)$ corresponds to the offset between the two images, which is calculated as

$$(i^*, j^*) = \arg\max_{(i,j)} C(i, j). \tag{4.20}$$

To compute the actual displacement relative to the original image coordinates, the offset of this peak from the center of the correlation matrix is calculated as

$$\Delta y = i^* - (M - 1), \quad \Delta x = j^* - (N - 1), \tag{4.21}$$

where M and N are the vertical and horizontal dimensions of the image, respectively. The values Δy and Δx represent the number of pixels by which I_{ref} must be shifted vertically and horizontally, respectively, to align with I_{on} . The aligned version of I_{ref} , denoted $\widetilde{I}_{\text{ref}}(x, y)$, is computed by applying a translation given by

$$\widetilde{I}_{\text{ref}}(x, y) = I_{\text{ref}}(x - \Delta x, y - \Delta y). \tag{4.22}$$

The network receives I_{on} and $\widetilde{I}_{\text{ref}}$ as inputs, while the original lasing-off image I_{off} (from which I_{on} was generated through augmentation) serves as the target output. This configuration transforms the traditional training pairs into training triplets, enabling the network to leverage both spatial and contextual information for improved reconstruction accuracy.

4.5.2 Details of the Neural Network

The Y-Net model is a fully convolutional network consisting of two parallel encoder branches and a single decoder branch. The two encoders process the lasing-on and reference images independently. Their outputs are then merged at a central bottleneck before being passed

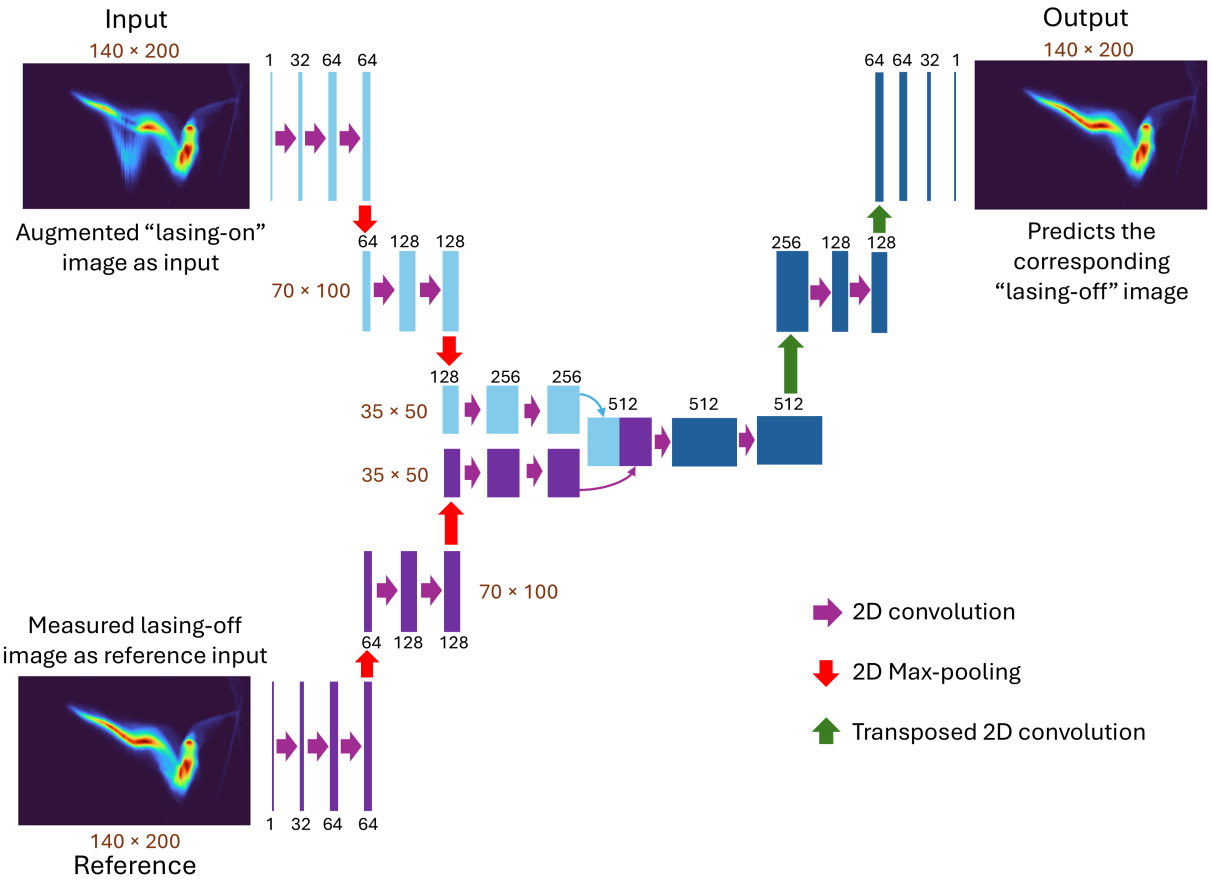

Figure 4.26: Architecture of the Y-Net model used for the phase space reconstruction task. The network utilizes a dual-encoder structure: the upper branch processes the "lasing-on" input, while the lower branch extracts features from the reference "lasing-off" image. These features are fused at the bottleneck and upsampled via the decoder path to predict the final output. The channel depth and spatial dimensions are indicated at each stage.

to the decoder, which reconstructs the final image. This two-input, one-output structure resembles the letter 'Y', as depicted in Fig. 4.26. The network is constructed from the same repeated "convolutional blocks" (2D Convolution with 3×3 filter, Batch Normalization, and ReLU) as defined in the U-Net description.

Encoder Branches The two encoder paths are shown on the left side of Fig. 4.26. Their purpose is to capture the context of both the "lasing-on" input and the "reference" input independently. Both branches have an identical structure. Each 140×200 single-channel input image passes through two convolutional blocks (purple arrow), expanding the channels from 1 to 32, and then to 64. Next, a 2D max pooling operation (red arrow) downsamples the resolution to 70×100. This is followed by two more convolutional blocks and a second max pooling operation, reducing the size to 35×50. A final block of two convolutional layers produces the 256-channel feature maps.

Feature Fusion Feature fusion occurs at the bottleneck level, where the 256-channel feature maps from both encoder branches are concatenated along the channel dimension, resulting in a 512-channel representation. This is then processed through two convolutional blocks to integrate the information from both input streams and produce the final bottleneck representation.

Decoder The decoder branch (right side of Fig. 4.26) reconstructs the output image through a series of upsampling and convolution operations. It begins with the 512-channel feature map from the bottleneck. The first decoder block uses a transposed 2D convolution [141] (green arrow) to upsample the data to 70×100 and reduce its channels to 256. This is then processed by two convolutional blocks, further reducing the channels to 128. This upsampling block (transposed convolution followed by two convolutional blocks) is repeated a second time, as shown in the diagram, increasing the size to 140×200 and reducing features from 128 to 64, and then to 32 channels. Finally, a 1×1 convolution reduces the feature map down to the single-channel, 140×200 grayscale image, which represents the network's prediction for the "lasing-off" image. A primary difference from the U-Net here is the absence of skip connections. In this architecture, the encoder branch processing the reference image is intended to provide the necessary spatial and contextual information, replacing the role of the U-Net's skip connections.

The network was trained on 8000 training triplets made using datasets from 02/03/2023 and 25/01/2024. The training used the same configurations as for the U-Net architecture described in Section 4.4.3.

4.5.3 Results

For inference, a randomly selected lasing-off image was aligned to the corresponding lasing-on input image using the cross-correlation method and provided as the reference input. Predictions were generated for different datasets under various operating conditions. The objective was to evaluate whether this model offers improved generalization for reconstruction across arbitrary datasets without requiring fine-tuning.

Figures 4.27–4.31 present several examples of the lasing-off predictions along with the corresponding power profile calculations. The results demonstrate that despite being trained on only two datasets, the network can generate more accurate lasing-off predictions compared to the U-Net model, especially when provided with a suitable reference image. The network appears to have learned how to transform the lasing-on image to resemble the given lasing-off reference while preserving the underlying physical structure. Particularly noteworthy is the case shown in Fig. 4.29, where the U-Net model incorrectly identified the two sharp features

at −250 fs and 80 fs as originating from lasing. In contrast, the Y-Net model correctly attributes these features to inherent characteristics of the electron bunch, utilizing information provided by the reference image. Although the reconstruction is not perfect and some artifacts remain at these temporal positions, the result represents a significant improvement over the U-Net predictions and demonstrates better contextual awareness.

An additional observation is that the energy spread method (Eq. (4.2)) yields better reconstructed profiles than the energy loss method (Eq. (4.1)) when applied to Y-Net predictions. This may be attributed to the absence of skip connections, which typically transfer spatial information to the later decoder layers and maintain direct correspondence between input and output. Without these connections, the Y-Net produces reconstructions that preserve overall shape and features while reducing pixel-level correspondence, making the energy spread method more effective for profile reconstruction.

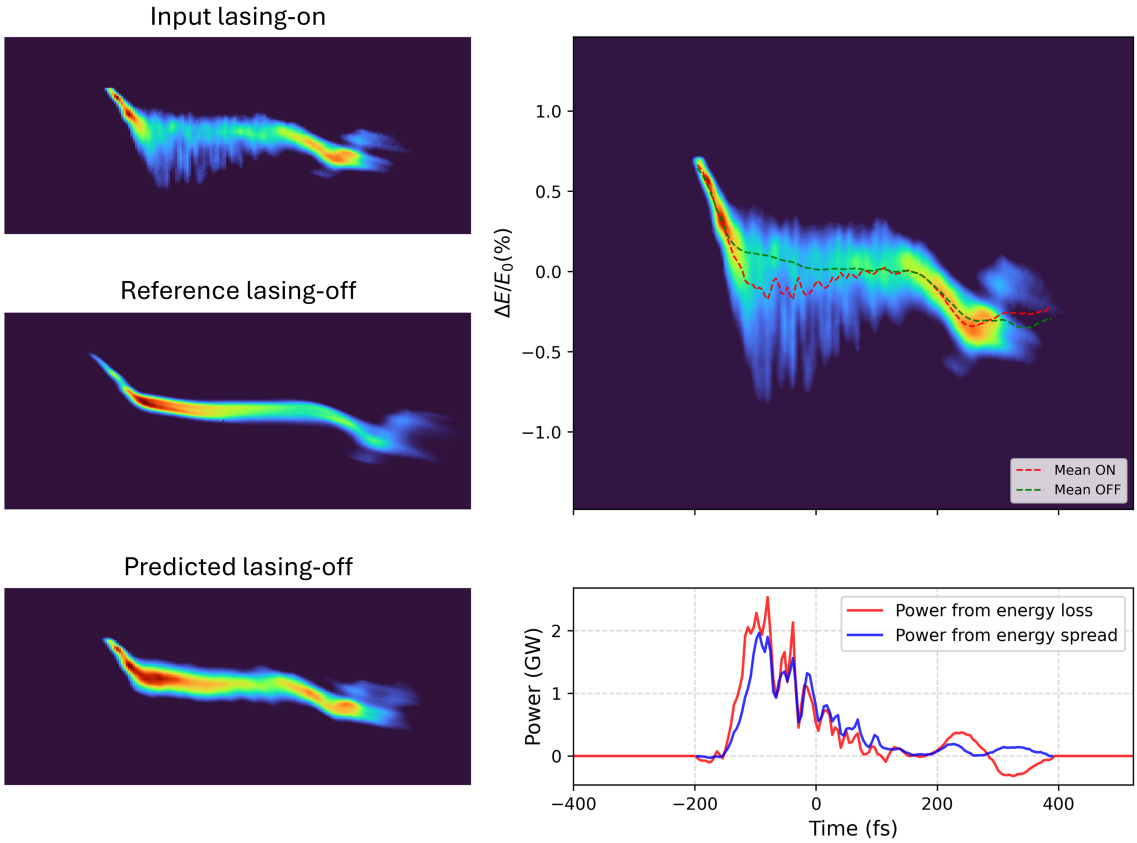

Figure 4.27: Pulse reconstruction results from the Y-Net model for the dataset from March 2, 2023. Top left: Measured lasing-on image given as input. Middle left: Measured lasing-off reference input. Bottom left: Predicted lasing-off image from the neural network. Top right: Lasing-on image overlaid with the slice centroids plotted for the lasing-on (red dashed curve) and predicted lasing-off (green dashed curve) images. Bottom right: Calculated power profile using the energy loss and energy spread methods (Eq. (4.1) and (4.2)), scaled by GMD pulse energy measurement.

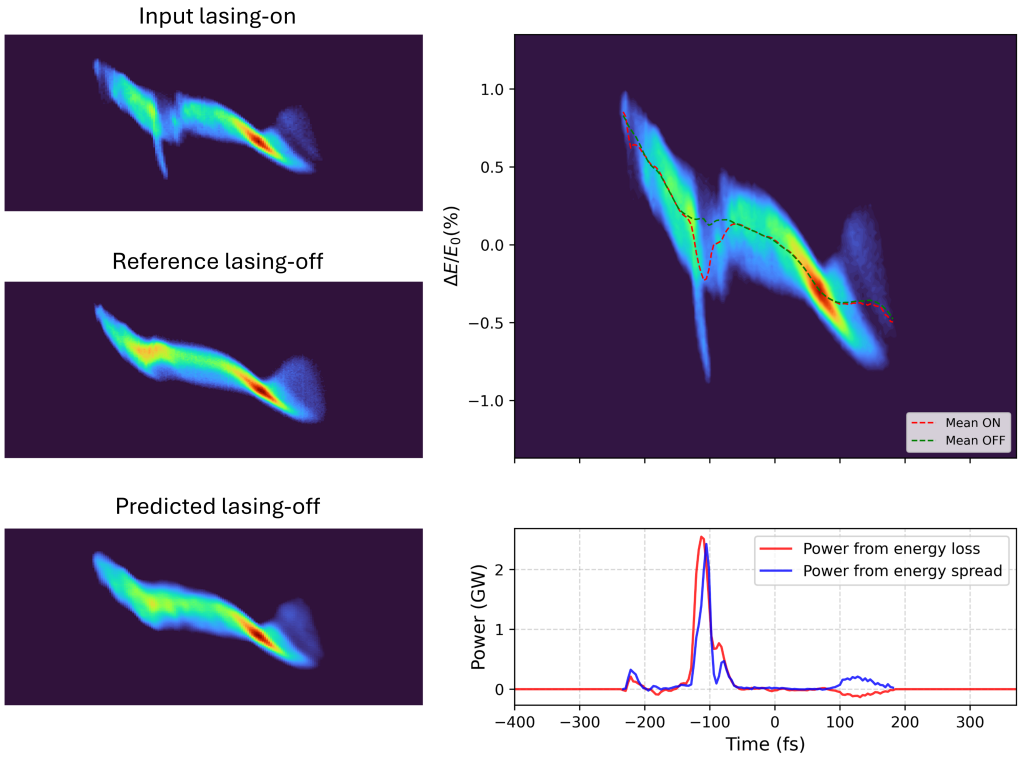

Figure 4.28: Pulse reconstruction results with the Y-Net model for a dataset from January 25, 2024.

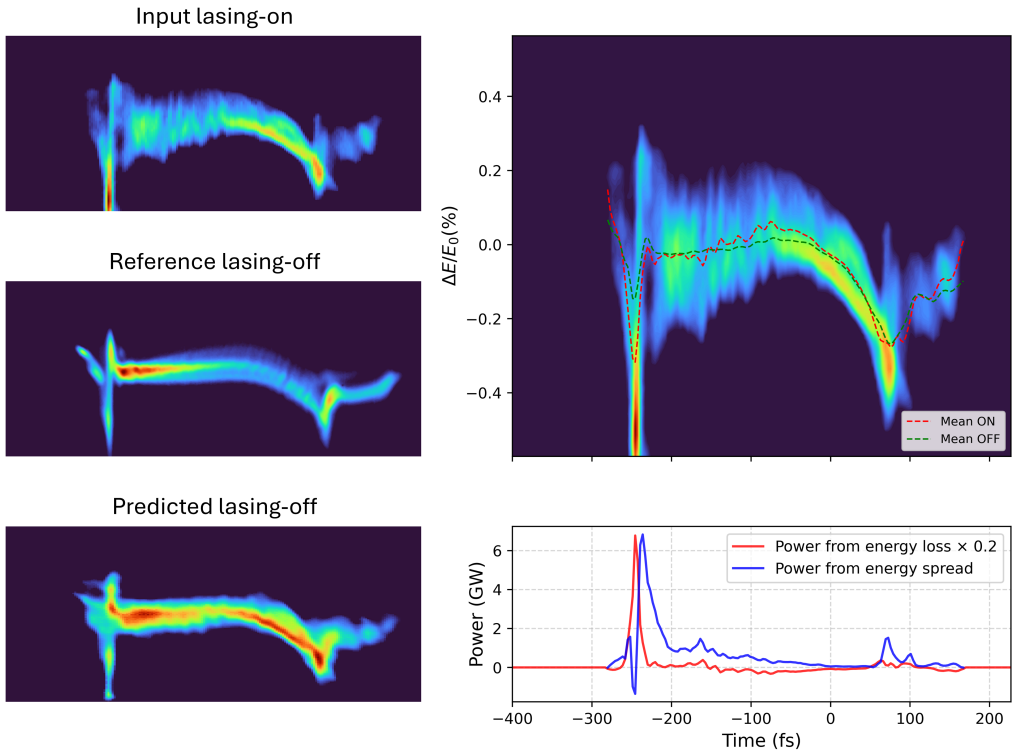

Figure 4.29: Pulse reconstruction results with the Y-Net model for a dataset from November 08, 2023.

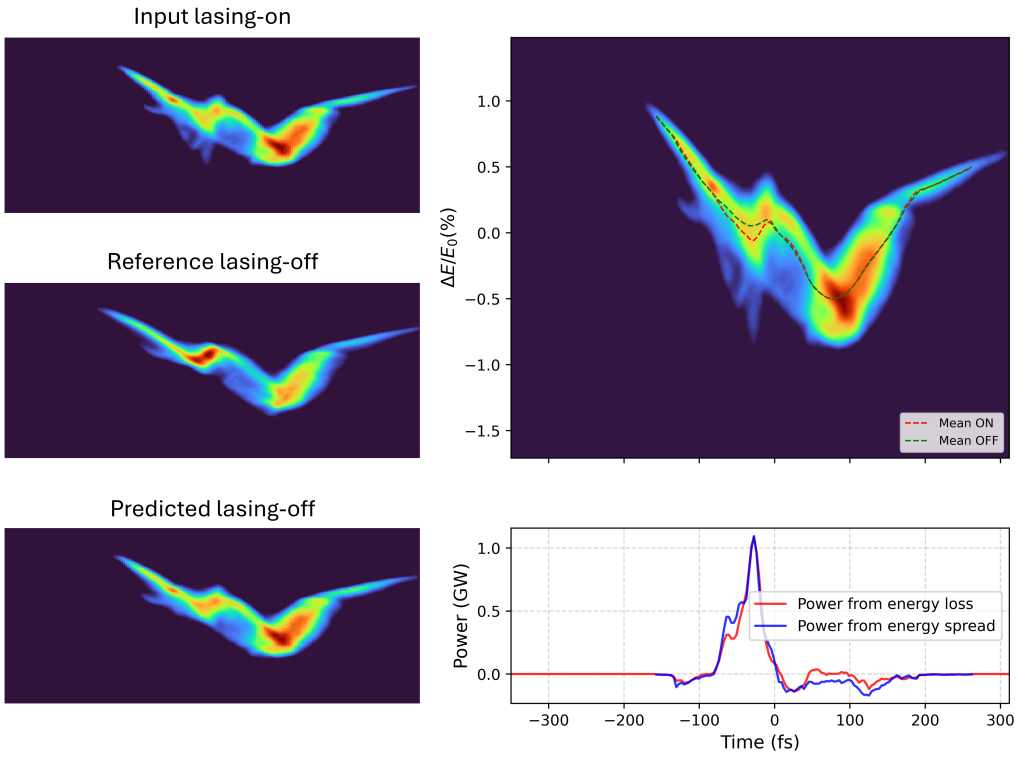

Figure 4.30: Pulse reconstruction results with the Y-Net model for a dataset from November 21, 2023.

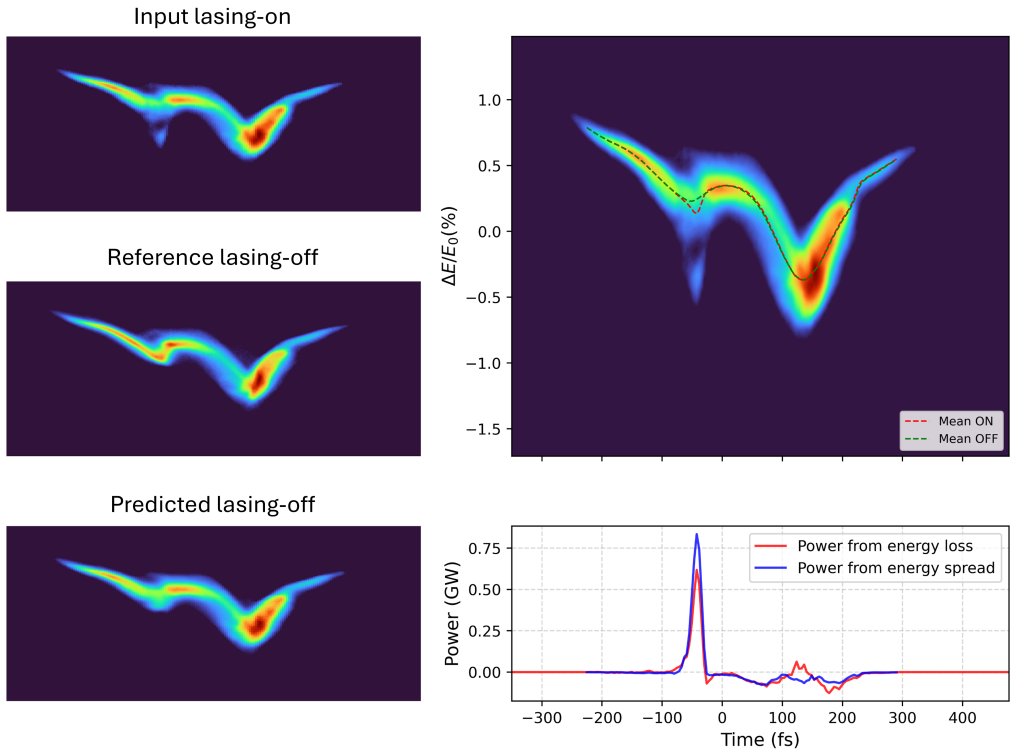

Figure 4.31: Pulse reconstruction results with the Y-Net model for a dataset from November 21, 2023.

Comparing the pulse energies calculated from the power profiles extracted using the energy loss method with the pulse energy measurements from the GMD, as depicted in Fig. 4.32, reveals that although they show a linear correlation (though with different slopes for different datasets), the pulse energies generally do not agree.

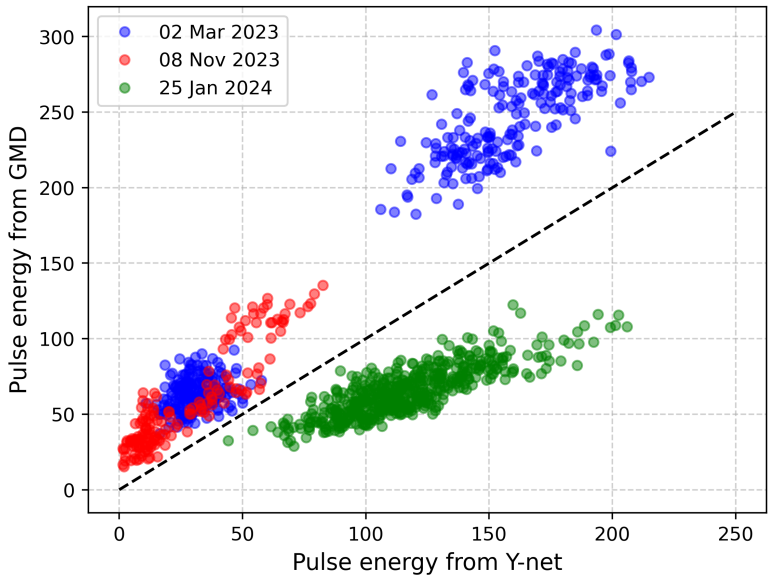

Figure 4.32: Comparison of pulse energies extracted using the Y-Net-based method and GMD pulse energy measurements for three different datasets. The dashed line marks the identity line.

While the Y-Net-based method shares some drawbacks with the standard approach, particularly the diminished one-to-one correspondence between input and output images that can give rise to undesired artifacts, it offers certain improvements. Most notably, the lasing-off output is still derived from the given lasing-on image while incorporating contextual information from the reference. The energy loss method appears to miscalculate peak powers and pulse energies, making the energy spread method a more suitable approach when combined with scaling to GMD pulse energy measurements.

For practical implementation, several lasing-off reference images should be acquired before each experimental run. A lasing-off reference can then be randomly sampled from this set, or alternatively, the mean of all lasing-off images can be used as the reference. With this approach, the network can predict the lasing-off image from the lasing-on image and the reference image without requiring fine-tuning, making it particularly valuable for real-time applications and datasets with varying experimental conditions.

4.6 Summary

This chapter presented two deep-learning methods for the estimation of the FEL pulse profiles by reconstructing the lasing-off phase space distribution from lasing-on images captured by the PolariX-TDS device at FLASH. The standard experimental technique requires acquiring several lasing-off reference images, which are then compared with measured lasing-on images to calculate the FEL power profile through established methods.

The first approach uses a U-Net architecture to predict lasing-off images directly from lasing-on inputs, removing the need for reference images. This method demonstrates excellent performance for short pulses where lasing features remain spatially localized, making it particularly well-suited for experiments that target short (few-fs to sub-fs) FEL pulses. However, the network encounters challenges when processing long pulses, where most or all of the electron bunch exhibits lasing. In these cases, the absence of unlased regions in the input image removes the reference structure that the network requires to construct an accurate lasing-off representation, limiting its usability for general-purpose applications. Additionally, while the method performs well on datasets used during training, it requires fine-tuning when applied to data collected under different experimental conditions. This limitation necessitates model retraining before each experimental run, thereby constraining its utility for real-time diagnostics.

The second approach, termed the Y-Net model, addresses the limitations by utilizing both the lasing-on image and a reference lasing-off image as inputs to predict the corresponding lasing-off image. By providing explicit information about the expected bunch structure through the reference image, this approach eliminates the need for dataset-specific fine-tuning. The Y-Net shows improved ability to distinguish between features inherent to the electron bunch and those from lasing. This is particularly evident in cases where the U-Net incorrectly identified sharp temporal features as lasing artifacts, while the Y-Net correctly attributed them to bunch characteristics.

The ability of the Y-Net method to operate without fine-tuning makes it particularly suitable for real-time applications and datasets with varying experimental conditions. Future improvements could be achieved through deeper architectures trained on more diverse datasets, or by incorporating advanced modules such as cross-attention mechanisms [143] that can learn spatial associations between different image regions, thereby enhancing the network's contextual awareness.

Chapter 5

Implementation of EEHG at DELTA

5.1 Overview of the EEHG Implementation

The short-pulse facility at DELTA has been operational since 2011, utilizing the CHG seeding scheme discussed in Chapter 3. Earlier efforts were focused on ambitious plans to rebuild a 21-meter section around the present U250 undulator to achieve the world's first implementation of the echo-enabled harmonic generation (EEHG) seeding scheme in a storage ring [62]. This large-scale plan included the purchase of two new electromagnetic undulators and designs for the required magnetic chicanes and a modified vacuum chamber [144] with the existing U250 undulator serving as the radiator. A schematic comparison of the CHG setup and the proposed EEHG configurations is presented in Fig. 5.1, while Table 5.1 summarizes the associated parameters.

As part of this thesis, extensive simulation studies were conducted to optimize the waist size and focal point of the two lasers, as well as to refine the design of the chicanes. The particle tracking simulations were performed using *ELEGANT* [139], while the design and optimization of the chicane magnets were carried out using *CST Studio* [145]. The finalized designs for the chicane magnets are shown in Fig. 5.2 and Fig. 5.3.

Due to rising costs of components and limited manpower, the project was discontinued in 2022. As an alternative, a simpler proof-of-principle implementation of the EEHG scheme was carried out as the project SPEED (Short-Pulse Emission via Echo at DELTA) [146]. This approach reconfigured the U250 undulator itself to integrate the necessary components (two modulators, two chicanes, and a radiator) within a compact 4.75-meter setup without requiring any additional components. The next chapter discusses the redesigned configuration of the U250 undulator for SPEED.

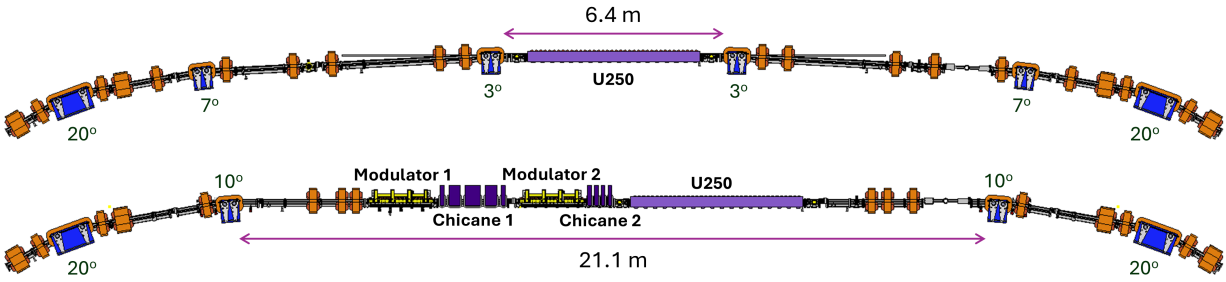

Figure 5.1: Schematics of the previous CHG setup (top) and the proposed EEHG setup with two new undulators (modulator 1 and 2), chicanes (bottom) and the U250 as radiator. Dipole magnets are shown in blue with the bending angle annotated. Quadrupole magnets are shown in orange.

CHG Configuration		
Component	**Parameter**	**Value**
U250 (full)	Total length	4.75 m
	Period length	0.25 m
	Total periods	17
	Peak magnetic field	0.76 T
CHG Setup (within U250)	No. of periods	7 (modulator) + 3 (chicane) + 7 (radiator)
	Max. R_{56} (chicane)	135 μm
Proposed Large-Scale EEHG Configuration (Canceled Plan)		
Component	**Parameter**	**Value**
Modulators (×2)	Total length	1.85 m
	Period length	0.20 m
	No. of periods	7
	Peak magnetic field	0.62 T
Chicanes (×2)	Max. R_{56} (chicane 1)	1600 μm
	Max. R_{56} (chicane 2)	200 μm
Radiator (U250)	Total length	4.75 m
	Period length	0.25 m
	Total periods	17
	Peak magnetic field	0.76 T

Table 5.1: Comparison of the existing CHG setup parameters and the parameters for the initially proposed, large-scale EEHG configuration [147].

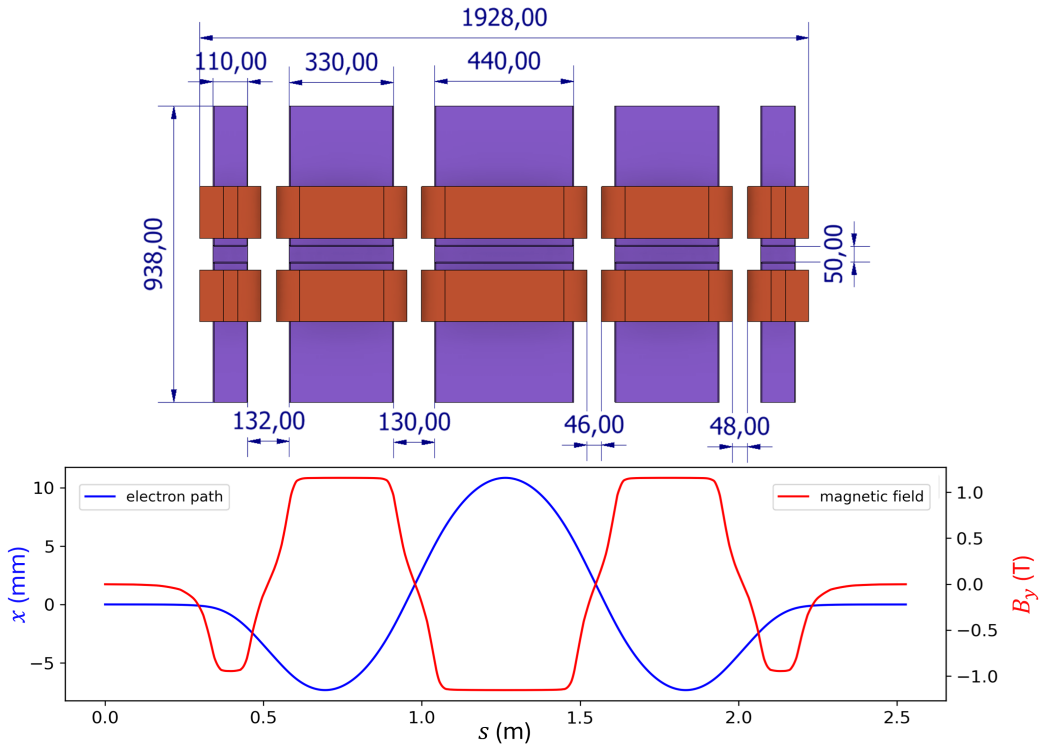

Figure 5.2: Proposed design of the first chicane (top) and the vertical magnetic field at a current of 500 A (bottom, red) and the corresponding simulated horizontal trajectory of an electron at an energy of 1.5 GeV (bottom, blue).

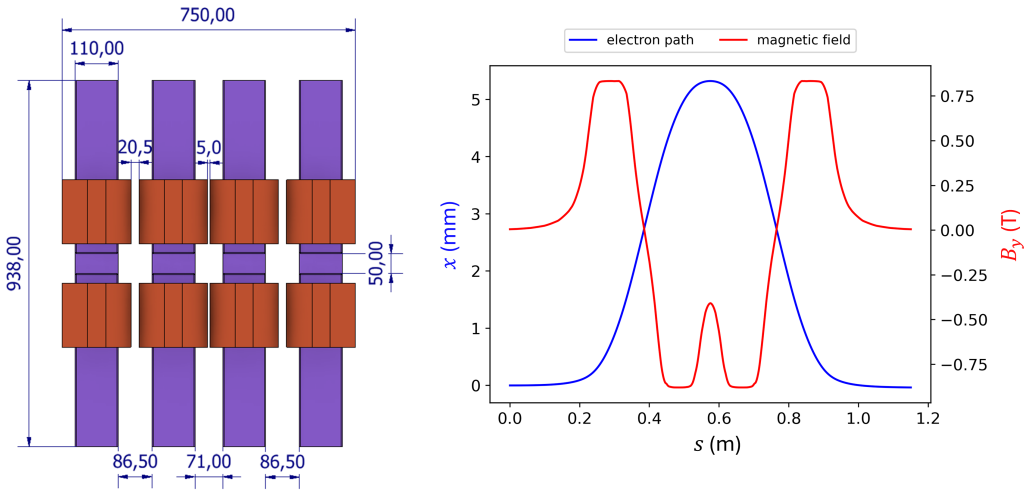

Figure 5.3: Proposed design of the second chicane (left) and the vertical magnetic field at a current of 500 A (right, red) and the corresponding simulated horizontal trajectory of an electron at an energy of 1.5 GeV (right, blue).

5.2 Outline of the SPEED Configuration

In the summer of 2022, the U250 undulator was rewired to implement the EEHG configuration [146]. This reconfiguration enabled all the essential components for the EEHG scheme - two modulators, two chicanes, and a radiator - to be integrated within the existing U250 undulator. To allow for switching between the EEHG configuration and standard user operations, where the entire device operates as a single undulator, a previously constructed board with copper bars was replaced by 3 such boards [146, 148]. This enables switching between the two modes within about half an hour, ensuring operational flexibility without requiring extensive hardware modifications.

The magnetic field distribution of the modified U250 undulator in the EEHG configuration is shown in Fig. 5.4. In this setup, the two modulators are designed with only four undulator periods each, while the radiator consists of three-and-a-half periods. This limitation imposes constraints on both the achievable energy modulation amplitude and the intensity of the generated EEHG radiation.

The U250 undulator comprises a total of 38 poles, with split coils installed at poles 01, 02, 17, 18, 21, 22, 37, and 38 [149]. These split coils allow the generation of either 1/4 or 3/4 of the full magnetic field strength by selectively powering one of the partial coils. By exciting both coils either in opposite or in parallel directions, 1/2 or the full magnetic field can be achieved, respectively. In the EEHG configuration, the end poles are configured to provide 1/2 of the full field strength. Because some of these undulator end poles are

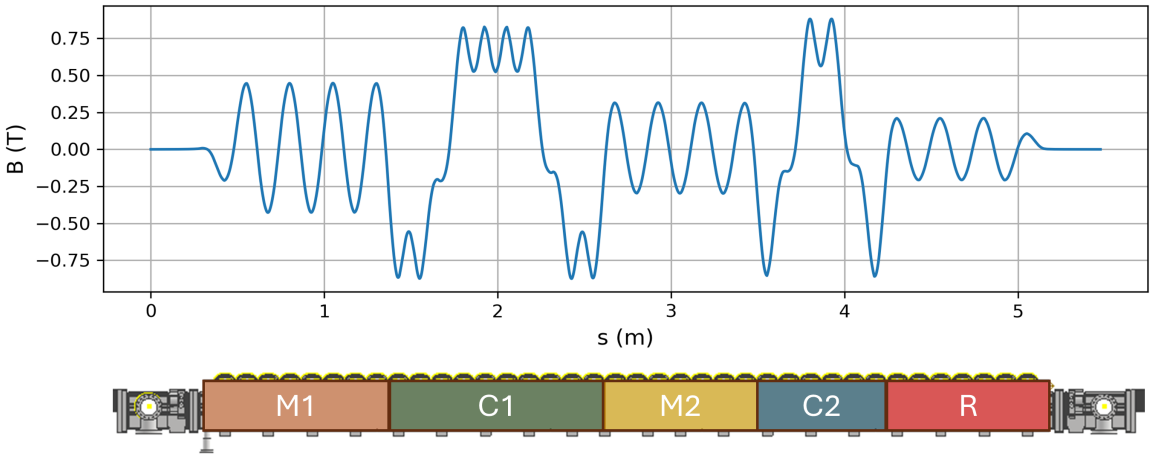

Figure 5.4: Modified SPEED configuration of the U250 undulator. The abbreviations M1,C1, M2, C2, R correspond to modulator 1, chicane 1, modulator 2, chicane 2 and radiator, respectively. The top plot shows the magnetic field as a function of longitudinal position for a configuration with M1 tuned to 800 nm, M2 tuned to 400 nm and R at 200 nm, while the current for the chicanes is set to 800 A.

integrated into the chicane magnets, an asymmetric chicane configuration is created when the adjacent undulator settings are different. This choice enables an additional undulator period, but also introduces a field-dependent shift in the beam axis from the undulator itself. In addition, adjustments to the chicane strengths can also result in horizontal shifts of the particle trajectory. Figure 5.5 shows the simulated particle trajectories under different configurations of the first modulator and chicane. The simulations use a custom code detailed in Section 5.4. As the magnetic field strength of either component is altered, a corresponding horizontal displacement of the beam axis occurs. This effect is not limited to the first modulator and chicane but is also observed in the second modulator, the second chicane, and the radiator.

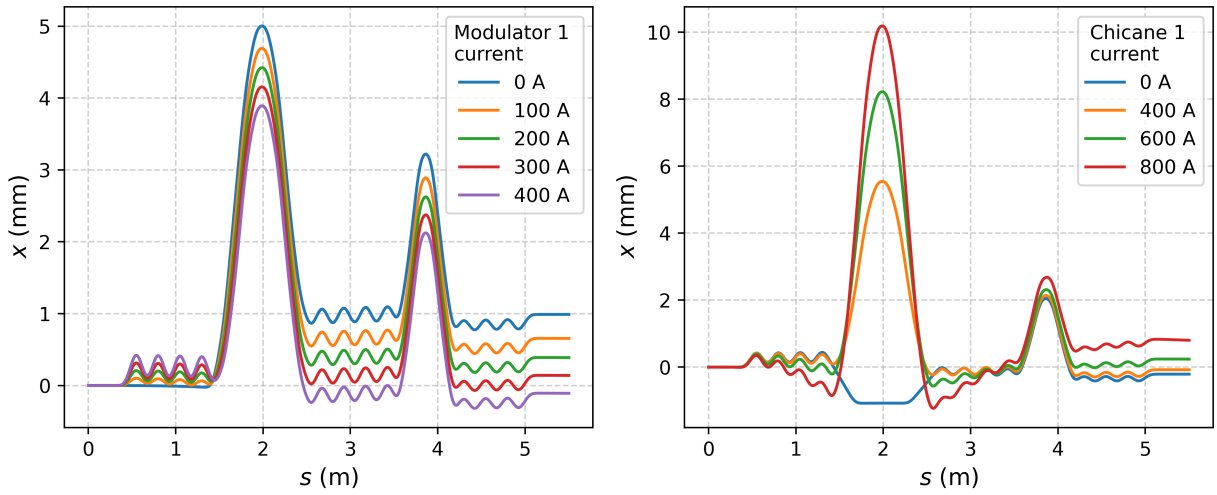

Figure 5.5: Simulated horizontal particle trajectories through the SPEED setup for different modulator 1 currents (left) and different chicane 1 currents (right). For the left plot, the fixed settings are $I_{C1} = 300$ A, $I_{C2} = 500$ A, $\lambda_{M2} = 400$ nm (280 A), and $\lambda_R = 200$ nm (190 A). For the right plot, λ_{M1} is set to 800 nm (398 A), while the other fixed settings remain the same.

For the chicanes, the horizontal shift in the particle axis remains minimal under moderate excitation strengths. However, as the excitation current exceeds 700 A, the magnetic yokes begin to saturate, resulting in a significant deviation from the desired beam axis. This can lead to an unintended exit angle of the electrons from the undulator, as shown in Fig. 5.6.

Figure 5.7 presents the horizontal position and angle of an electron as it exits the U250 undulator for varying excitation currents across different sections of the EEHG setup. Both the modulators and the radiator exhibit a consistent, linear shift in the particle axis as the current is increased. In contrast, the chicanes demonstrate a more rapid shift once the excitation current exceeds 600 A, as the magnets approach saturation. This significant divergence of the trajectory introduces stricter limits to the operational parameter range for the setup, as such misalignment can degrade the energy modulation and the microbunch formation.

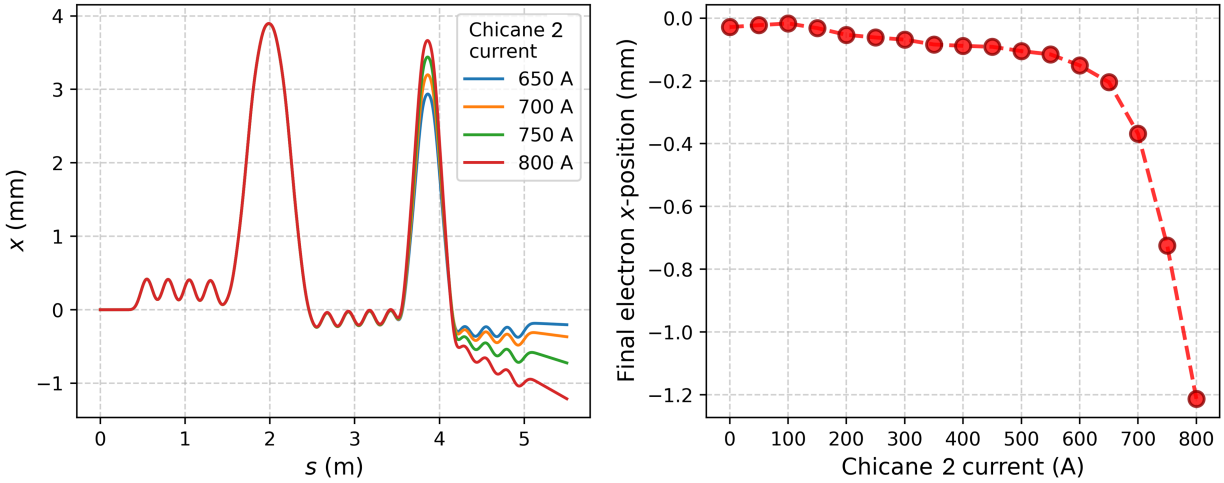

Figure 5.6: Simulated horizontal particle trajectories through the SPEED setup for different chicane 2 currents (left). Fixed settings are $\lambda_{M1} = 800$ nm, $\lambda_{M2} = 400$ nm, $I_{C1} = 300$ A, and $\lambda_R = 200$ nm. The final horizontal position of the particle for different C2 currents (right, red circles). The trajectory starts to shift significantly for high currents beyond 700 A.

The maximum K value for all undulators in this configuration is 10.3, which corresponds to a wavelength of 800 nm at a beam energy of 1.5 GeV. The maximum R_{56} value achievable with this setup is 550 μm and 105 μm for the first and second chicane, respectively, at a current of 800 A. However, as seen in Fig. 5.7, operating the chicanes at their maximum strength introduces additional unwanted effects. Notably, the non-zero R_{52} values inherent to the asymmetric chicanes and induced by the saturation effects can lead to the smearing of the microbunching, further complicating the optimization of the setup. A method to counteract this effect and minimize the R_{52} value was proposed. The approach and the associated simulation studies to address this challenge will be discussed in Section 5.5.

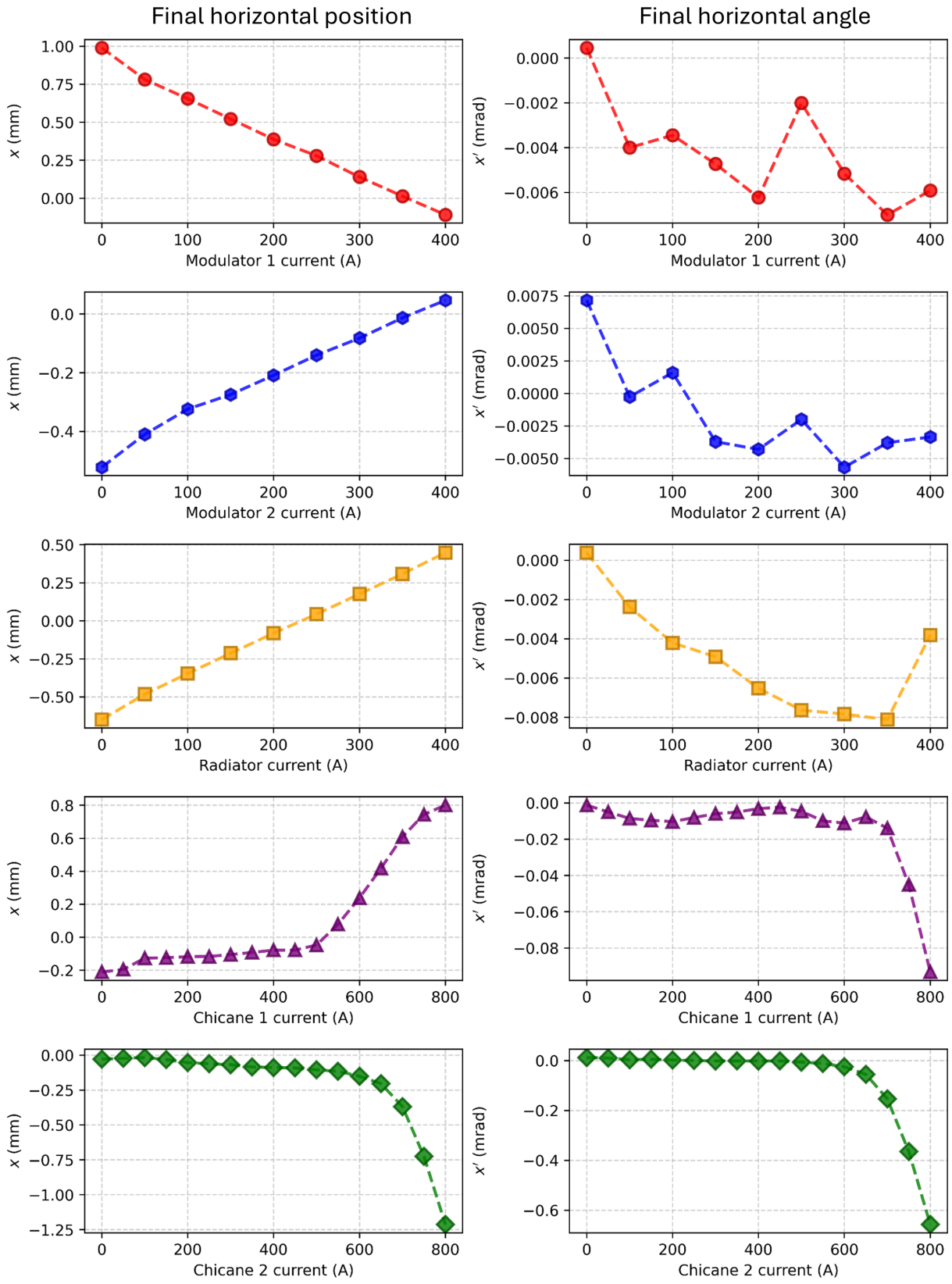

Figure 5.7: Simulated horizontal position (left column) and angle (right column) of the electron exiting the U250 undulator. Each row plots the exit coordinates as a function of one component's current, while all other components are held at their nominal values. The nominal settings are: $\lambda_{M1} = 800$ nm (398 A), $\lambda_{M2} = 400$ nm (280 A), $\lambda_R = 200$ nm (190 A), $I_{C1} = 300$ A, and $I_{C2} = 500$ A. Magnetic fields were calculated using *CST Studio* simulations incorporating the measured permeability data of the U250 yoke material.

5.3 Estimation of Chicane Strength

In order to determine the current-to-R_{56} calibration for the chicanes, two particles with a given energy difference δE were tracked through the simulated magnetic fields derived from *CST Microwave Studio* [145]. The longitudinal separation Δz between the particles was then used to calculate the R_{56} value of the chicane from the relationship from Eq. (2.54)

$$R_{56} = \frac{\Delta z}{\Delta E / E_0}. \tag{5.1}$$

The values of R_{56} were calculated for various currents of both the first and second chicanes.

Further simulations [1] were performed with different configurations of modulators and radiator to assess their impact on the chicane strength. Figure 5.8 demonstrates the significant impact of modulator and radiator settings on the R_{56} value of the first chicane ($R_{56}^{(1)}$). A difference of approximately 80 μm in $R_{56}^{(1)}$ is observed at the maximum current of 800 A, depending on whether the modulators are powered.

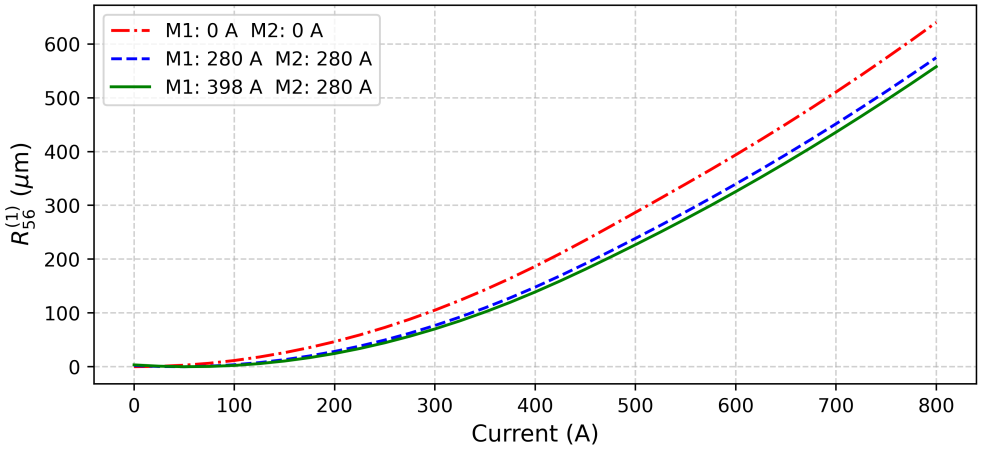

Figure 5.8: R_{56} value of the first chicane as a function of current for different settings of the first and second modulator.

For the EEHG setup, a fixed configuration for the first and second modulator will be used, specifically at currents of 398 A and 280 A, corresponding to wavelengths of 800 nm and 400 nm, respectively. Consequently, the current-to-R_{56} calibration was performed for this particular setup, which is shown in Fig. 5.9 (left). Under these conditions, the first chicane achieves a maximum $R_{56}^{(1)}$ value of 558 μm at a current of 800 A.

The strength of the second chicane, similarly to the first, is also influenced by the modulator 2 and radiator settings. However, simulations (Fig. 5.9, right) show that its influence on the

[1] See Section 5.4

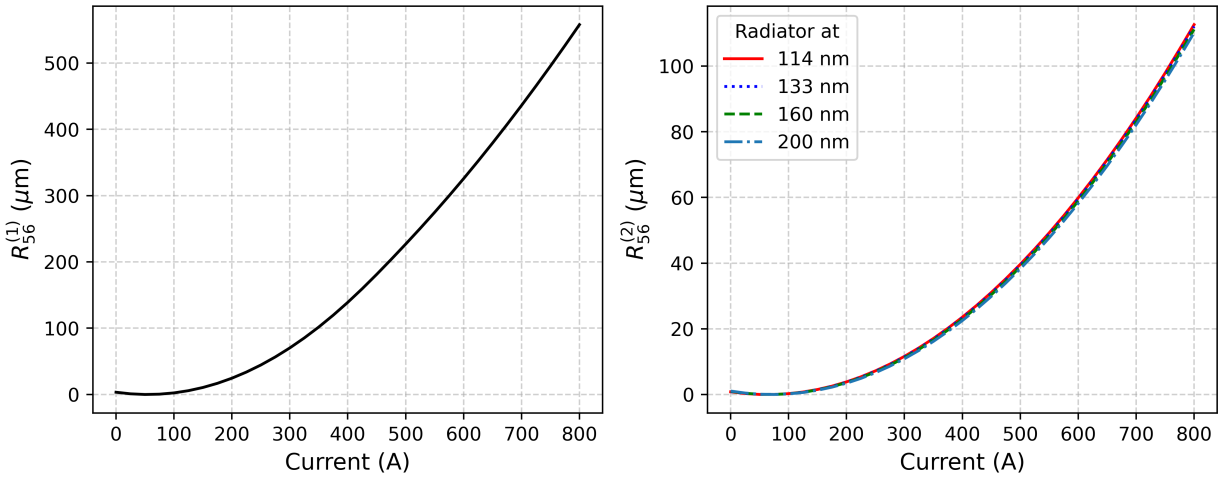

Figure 5.9: The R_{56} value of the first chicane as a function of current (left) when the modulators are set to 800 nm and 400 nm, respectively, and the R_{56} value of the second chicane as a function of current (right) for different radiator settings.

R_{56} of the second chicane ($R_{56}^{(2)}$) is negligible. Therefore, an approximation from any single calibration curve is sufficient for practical purposes.

It is also noted that, due to the interleaved field profile of the chicanes, the R_{56} values do not vanish at zero chicane currents. Instead, the minimum R_{56} is observed at a positive offset current (around 60 A), depending on the undulator configurations.

To experimentally validate the chicane strengths obtained from the numerical simulations, the modulators and radiator were arranged in an optical-klystron configuration [150], allowing the extraction of R_{56} values from measured spectra by comparing them with theoretical predictions. In an optical-klystron arrangement, two undulators are tuned to the same wavelength λ_r, with a chicane between them, which can be used to control the relative delay of the synchrotron light emission. The interference between the radiation emitted by the two undulators results in a characteristic spectral modulation. The intensity I_{ok} of the optical-klystron spectrum is described by [77]

$$I_{\mathrm{ok}}(\lambda) = 2I_{\mathrm{u}}\left(1 + f\cos\left(2\pi\left(N_{\mathrm{u}} + N_d\right)\frac{\lambda - \lambda_r}{\lambda}\right)\right), \tag{5.2}$$

where N_{u} is the number of undulator periods and N_d characterizes the strength of the chicane via

$$N_d = \frac{R_{56}}{2\lambda_r}. \tag{5.3}$$

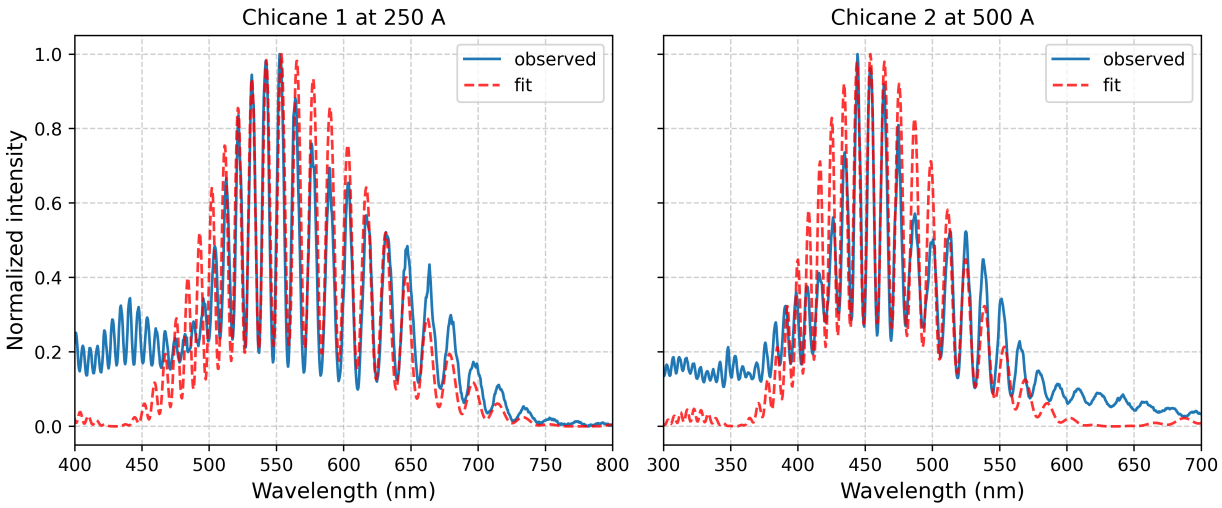

Figure 5.10: Representative spectral fitting for the first and second chicanes in the optical-klystron configuration. The fit prioritizes matching the fringe width and separation due to deviations from the idealized theory.

The intensity spectrum I_u of a single undulator is given by

$$I_\mathrm{u}(\lambda) \propto \left(\frac{\sin\left(\pi N_\mathrm{u} \frac{\lambda-\lambda_r}{\lambda}\right)}{\pi N_\mathrm{u} \frac{\lambda-\lambda_r}{\lambda}} \right)^2 . \tag{5.4}$$

The modulation depth of the spectrum, denoted as f, is determined by

$$f = f_0 \exp\left(-8\pi^2 \left(\left(N_\mathrm{u} + N_d\right)\sigma_E\right)^2\right), \tag{5.5}$$

where σ_E is the relative energy spread of the electron beam and $f_0 \approx 1$ accounts for the resolution of the spectrometer used in the measurement.

To determine the longitudinal dispersion of the first EEHG chicane experimentally, the first and second modulators were set to a current of 340 A, and the resulting spectra were recorded using a fiber spectrometer [151] for various first-chicane current values. A least-squares fitting procedure was employed to optimize the parameters λ_r, f_0, and N_d. The extracted value of N_d, along with the fitted λ_r, was then used in Eq. (5.3) to estimate R_{56}. The same procedure was repeated for the second chicane, with the second modulator and radiator set to 300 A.

Figure 5.10 presents representative examples of the spectral fitting procedure for both chicanes. It is worth noting that due to the specific characteristics of the SPEED magnetic configuration, including the non-overlapping emission axes of the undulators, strong influence of endpoles, and uncertainty of the energy spread, the observed spectra deviate slightly from the idealized theoretical predictions. Consequently, the fitting approach was adjusted to prioritize matching the fringe width and separation rather than the overall spectral shape

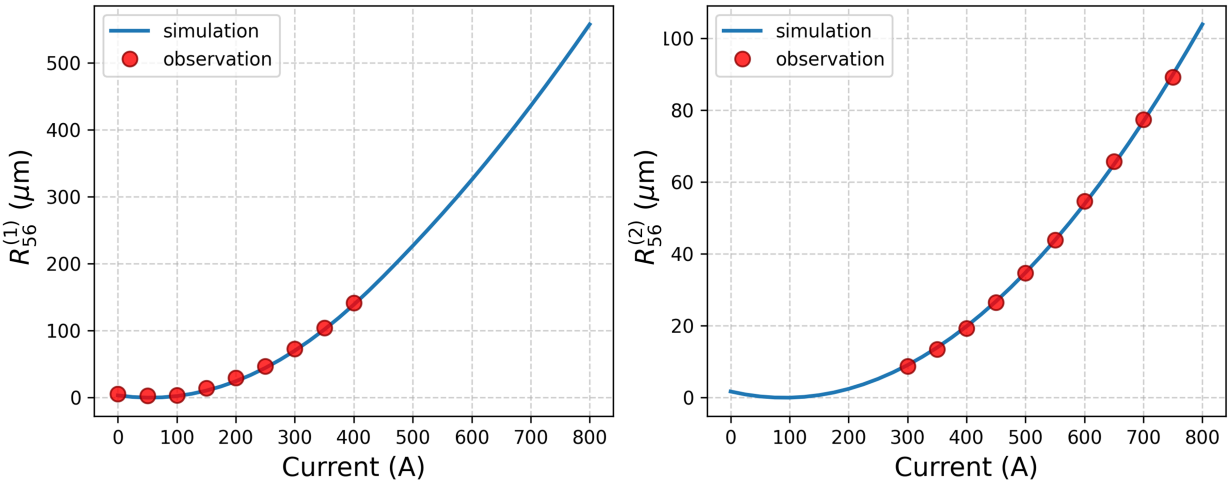

Figure 5.11: Comparison between the simulated and experimentally obtained R_{56} values for both EEHG chicanes.

to ensure an accurate extraction of R_{56}.

The experimentally extracted R_{56} values were then compared with the values obtained from simulations based on the *CST*-generated magnetic fields. Figure 5.11 shows this comparison. The measurements for chicane 1 were limited to currents below 400 A, as interference fringes became too narrow for reliable fitting at higher currents, while chicane 2 measurements were restricted to currents above 300 A due to broad fringes and poor fit at lower settings. Within these ranges, the comparison shows a strong agreement, verifying the reliability of the computational model used in the chicane calibration.

A polynomial fit was performed over the R_{56} values estimated from the simulations. The current-to-R_{56} relation for the chicanes can be best described by a cubic polynomial as follows, where I is the respective chicane current in Amperes.

$$R_{56}^{(1)}(I)\,[\mu\text{m}] = -5.27 \times 10^{-7}\, I^3 + 1.52 \times 10^{-3}\, I^2 - 1.86 \times 10^{-1}\, I + 5.15 \tag{5.6}$$

$$R_{56}^{(2)}(I)\,[\mu\text{m}] = -2.79 \times 10^{-9}\, I^3 + 2.11 \times 10^{-4}\, I^2 - 2.70 \times 10^{-2}\, I + 0.83 \tag{5.7}$$

5.4 Simulation Studies for SPEED

Since the proposed design of SPEED used modulators with only four periods and limited chicane strengths, a simulation study was necessary to evaluate whether this setup could effectively generate high harmonics. Unlike a set of conventional undulators, the SPEED setup had a more complex structure. This made it necessary to develop a dedicated simulation to model how the laser interacts with the electron beam and how microbunching develops as

a result. To achieve this, a custom *Python* code was written to simulate the laser-electron interaction and analyze the resulting microbunching. The program uses a magnetic field profile generated in *CST Studio* and calculates the motion of electrons by considering both the forces from the laser's electric field and the Lorentz force due to the magnetic field. The laser field is parametrized by its fundamental properties, including wavelength, pulse energy, pulse duration, focal position, and beam waist at the focus etc. This is discussed in detail in the next subsection.

At the end of each simulation, the bunching factor is calculated using Eq. (2.55) at different wavelengths by analyzing the phase space distribution of the modulated electron beam. This helps to determine the efficiency of harmonic generation in the SPEED setup and provides insight into how the system parameters could be optimized. The complete code, along with documentation and example configurations, was made publicly available on GitHub[2]. Following is the detailed description of the simulation.

5.4.1 Modeling the Magnetic Field

The magnetic field used in the laser-electron interaction can be obtained via two distinct methods, depending on the available input. For final particle tracking and detailed analysis, precomputed field maps from *CST Microwave Studio* are used. These maps are read from file and interpolated to create a continuous function along the beam path, capturing realistic effects such as iron non-linearity and fringe fields.

However, for parameter optimization and scanning large parameter spaces, *CST* simulations are computationally expensive. In these cases, an analytical model is used to synthesize the field via superposition of individual magnet contributions. In the analytical mode, the magnetic field for modulators and radiator is determined by their fundamental wavelength λ_0 defined by Eq. (2.24). For a relativistic electron beam with the Lorentz factor γ, the undulator parameter K is calculated as [26]

$$K = \sqrt{\frac{4\lambda\gamma^2}{\lambda_u} - 2}, \tag{5.8}$$

where $\lambda_u = 0.25$ m is the undulator period. The corresponding on-axis peak magnetic field B in the undulator gap is then [26]

$$B = \frac{2\pi m_e c K}{e\lambda_u}, \tag{5.9}$$

where m_e is the electron mass, c is the speed of light, and e is the elementary charge.

[2] https://github.com/arjun-krishnan/Laser-modulation

The coil current required to produce B is derived using the relationship [26]

$$I = \frac{B \cdot h}{\mu_0 N}, \tag{5.10}$$

where $h = 0.05$ m is the gap between magnet poles, $N = 48$ is the total number of windings per coil and μ_0 is the vacuum permeability.

The U250 undulator is modelled as a sequence of 38 dipole magnets, each assigned a current value based on the currents in the corresponding section. The lattice structure alternates between modulators, chicanes, and radiator, with each magnet's polarity following a predefined sequence to generate the desired field orientation.

The total magnetic field $B(z)$ along the beamline is constructed by superimposing contributions from all magnets. Each magnet's field is modeled as a rectangular profile with softened edges to avoid discontinuities, approximating realistic fringe fields. The field contribution from a single magnet centered at position z_m is expressed using a logistic function

$$B_m(z) = \frac{B_{0,m}}{1 + \exp\left(\frac{z-(z_m-\Delta/2)}{\epsilon}\right)} + \frac{B_{0,m}}{1 + \exp\left(\frac{(z_m+\Delta/2)-z}{\epsilon}\right)}, \tag{5.11}$$

where $B_{0,m}$ is the peak field of the m-th magnet, $\Delta = 0.08$ m is the effective yoke length, and $\epsilon = 0.02$ m controls the edge softness. The total field $B(z)$ is the sum of contributions from all 38 magnets in the lattice. A global scaling factor of 1.19 adjusts the peak field $B_{0,m}$ to match results from more accurate finite-elements simulations such as those from *CST Microwave Studio* which also incorporates the measured magnetization curve of the iron used for the U250 magnets. The synthesized $B(z)$ is interpolated using a cubic spline to provide a continuous function for particle tracking simulations.

5.4.2 Modeling the Laser Pulse

The following description and modelling of the laser pulse are based on textbooks like [40] and [152].

The laser pulse is modeled within a lab-frame Cartesian coordinate system (x, y, s), where s is the propagation axis and t is the time in the lab-frame. The laser focus is located at $s = s_0$. The temporal pulse envelope is defined in a co-moving time coordinate $t' = t - s/c$. The laser electric field is modeled by a Gaussian distribution both in space and time. The maximum power of the laser pulse is determined using the relationship

$$P_{\max} = \frac{0.94\ E_{\mathrm{L}}}{T_0}, \tag{5.12}$$

where E_L is the laser pulse energy and T_0 is the full-width at half-maximum pulse duration. The factor 0.94 arises from the Gaussian time integral. The peak intensity at the laser focus is then calculated using the waist sizes of the Gaussian beam profile at the focus (as defined in Eq. (2.35)) in the horizontal and vertical directions, w_{0x} and w_{0y}, as

$$I_0 = \frac{2\,P_\mathrm{max}}{\pi w_{0x} w_{0y}}. \tag{5.13}$$

The factor of $2/\pi$ arises from the two-dimensional spatial integration of the Gaussian beam profile. The electric field amplitude is derived from the intensity as

$$\mathcal{E}_0 = \sqrt{2 Z_0 I_0}, \tag{5.14}$$

where $Z_0 = 377\ \Omega$ is the impedance of free space. A deviation from an ideal Gaussian beam profile is incorporated into the simulations by introducing the beam quality factor or M^2, which modifies the horizontal and vertical Rayleigh length as (see Section 2.4)

$$s_{R_x} = \frac{\pi w_{0x}^2}{M^2 \lambda}, \qquad s_{R_y} = \frac{\pi w_{0y}^2}{M^2 \lambda}. \tag{5.15}$$

The beam width at any position s is then given by

$$w_x(s) = w_{0x}\sqrt{1 + \frac{(\Delta s)^2}{s_{R_x}^2}}, \qquad w_y(s) = w_{0y}\sqrt{1 + \frac{(\Delta s)^2}{s_{R_y}^2}}, \tag{5.16}$$

where $\Delta s = s - s_0$.

The temporal profile incorporates group delay dispersion (GDD) and third-order dispersion (TOD) through spectral phase modulation. The pulse is first defined in the co-moving time frame t' with a Gaussian envelope

$$\mathcal{E}(t') = \exp\left(-\frac{t'^2}{2\sigma_{t'}^2}\right)\exp(i\omega_0 t'), \tag{5.17}$$

where $\sigma_{t'} = T_0/(2\sqrt{\ln 2})$, and ω_0 is the central angular frequency. A Fourier transform converts this to the frequency domain where quadratic and cubic spectral phase terms are applied

$$\phi(\omega) = \frac{1}{2}D_2(\omega - \omega_0)^2 + \frac{1}{6}D_3(\omega - \omega_0)^3 \tag{5.18}$$

with D_2(GDD) and D_3(TOD) characterizing the dispersion, while the higher-order terms are explicitly set to zero. An inverse Fourier transform then yields the dispersed temporal profile $\mathcal{E}_d(t')$.

The total electric field at position (x, y, s) and time t combines a spatial Gaussian enve-

lope, the dispersed temporal profile, and oscillating phase terms. The spatial electric field amplitude is given by

$$\tilde{\mathcal{E}}(x,y,s) = E_0\sqrt{\frac{w_{0x}}{w_x(s)}}\sqrt{\frac{w_{0y}}{w_y(s)}}\exp\left(-\frac{y^2}{w_y^2(s)} - \frac{x^2}{w_x^2(s)}\right) \tag{5.19}$$

The spatial phase terms account for wavefront curvature and Gouy phase shift

$$\Phi_s(x,y,s) = -\frac{k}{2}\frac{x^2}{R_x} - \frac{k}{2}\frac{y^2}{R_y} + \frac{1}{2}\arctan\left(\frac{\Delta s}{s_{R_x}}\right) + \frac{1}{2}\arctan\left(\frac{\Delta s}{s_{R_y}}\right) \tag{5.20}$$

where the wavefront curvatures are

$$R_x(\Delta s) = \Delta s\left(1 + \left(\frac{s_{R_x}}{\Delta s}\right)^2\right), \qquad R_y(\Delta s) = \Delta s\left(1 + \left(\frac{s_{R_y}}{\Delta s}\right)^2\right). \tag{5.21}$$

The complete expression for the scalar electric field becomes

$$\mathcal{E}_{\text{total}}(x,y,s,t) = \tilde{\mathcal{E}}(x,y,s)\cdot\mathcal{E}_d\left(t - \frac{s}{c}\right)\cdot\exp\left(i\Phi_s(x,y,s)\right) \tag{5.22}$$

5.4.3 Simulating Laser-Electron Interaction

The simulation tracks the interaction of a relativistic electron bunch with the laser field by numerically solving the equations of motion for individual electrons in the presence of the laser's electromagnetic field and the magnetic field of the undulator. Each particle in the simulation represents a single electron, and collective effects are neglected for this simulation. Since simulating $\sim 10^7$ electrons in a realistic scenario would be too intensive and time-consuming, representative 10^5 electrons were chosen for the simulations. The electron bunch is initialized with its 6D phase distribution. The particles are distributed uniformly over a length of 100 fs longitudinally. The relative energy deviation follows a Gaussian distribution with an rms spread $\sigma_E = 7 \times 10^{-4}$ with a mean energy of 1.492 GeV. The transverse distribution is generated from the accelerator lattice parameters including the optical functions $(\alpha_{x,y}, \beta_{x,y})$, emittances ($\epsilon_x = 16$ nm rad, $\epsilon_y = 1.6$ nm rad) and dispersion (D, D'). At each time step, the evolution of the electron distribution is computed. The electric field of the laser is evaluated at the position of each electron using Eq. 5.22 for each time step. If an additional laser field component is present, such as in a two-laser interaction scenario in the case of the EEHG implementation at DELTA, the contributions from both fields are summed.

The momentum update is performed using the relativistic Lorentz force equation

$$\frac{d\mathbf{p}}{dt} = e\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right), \tag{5.23}$$

where $\mathbf{E}$ is the total electric field at the electron position, $\mathbf{v}$ is its velocity, and $\mathbf{B}$ is the corresponding magnetic field. The equation is solved numerically at each time step to update the electron's momentum. The default time step used for the simulation is 1 ps.

The new position of each electron is computed using

$$\frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{\gamma m_e}, \tag{5.24}$$

where $\mathbf{p}$ is the updated momentum. The relativistic Lorentz factor γ is calculated from the magnitude of the updated momentum $|\mathbf{p}|$ as

$$\gamma = \sqrt{1 + \frac{|\mathbf{p}|^2}{m_e^2 c^2}}. \tag{5.25}$$

The simulation continues iterating over successive time steps until the electron bunch has propagated through the entire interaction region. The bunching factor is computed from the distribution of electrons at the end of the interaction given by Eq. 2.55.

5.5 Simulation Results

Extensive simulation studies were carried out to optimize the parameters of the laser and the different sections of the U250 undulator. This section discusses the results in detail.

5.5.1 Choice of Modulation Wavelength

The laser system at DELTA is equipped with second-harmonic generation (SHG) and third-harmonic generation (THG) units, enabling the conversion of the fundamental 800 nm laser wavelength to 400 nm and 266 nm radiation, respectively. This enables multiple wavelength configurations for the first and second EEHG energy modulation stages.

One possible approach involves splitting the 800 nm beam with a 50-50 beam splitter to create two identical copies of the laser pulse for sequential modulations. An alternative approach utilizes the SHG unit, employing both the generated 400 nm radiation and the unconverted residual 800 nm beam for the two-stage modulation. Extending this scheme further, the THG unit can be used to generate 266 nm radiation, which would serve as the

U250 configuration		Pulse energy (mJ)		Energy modulation amplitude (%)	
		L1	L2	M1	M2
M1: 800 nm	M2: 800 nm	4.0	4.0	0.32	0.32
M1: 800 nm	M2: 400 nm	2.5	2.5	0.20	0.15
M1: 800 nm	M2: 266 nm	1.8	0.6	0.18	0.06
M1: 400 nm	M2: 266 nm	1.5	0.6	0.12	0.06

Table 5.2: Pulse energy and energy modulation amplitude for different modulator configurations. The FWHM pulse length is considered to be 50 fs and the waist size w_0 to be 1.6 mm for all cases.

wavelength for the second modulation. In such a scenario, the first modulation stage could be implemented using either the fundamental 800 nm wavelength or the frequency-doubled 400 nm beam.

Selecting a longer wavelength for the first modulation (λ_1) and a shorter wavelength for the second (λ_2) is generally more favorable for the bunching efficiency. The harmonic number is given by $h = Km + n$, where $K = \lambda_1/\lambda_2$ and m, n are integers. The configuration $\lambda_1 < \lambda_2$ results in $K < 1$, which necessitates a higher integer m to reach the same harmonic h. A larger m generally leads to a reduction in the achievable bunching factor.

To systematically compare the efficacy of the EEHG scheme under different choices of modulation wavelengths, the laser-electron interaction was simulated by tracking an ensemble of on-axis, on-momentum test particles to compute the induced energy modulation amplitude. The waist sizes w_{0x} and w_{0y} were assumed to be 1.6 mm, while the M^2 given a value of 2. The pulse energies of the laser pulses used in the simulations for various wavelength configurations, along with the resulting energy modulation amplitudes, are presented in Table 5.2. With the modulation amplitudes determined, the theoretical bunching factor was subsequently calculated for a range of chicane parameters. Specifically, a systematic scan was performed over a grid of R_{56} values for both chicanes, covering their entire operational range from the minimum to the maximum setting. For each combination of chicane strengths, the bunching factor was computed at multiple harmonics of the seed wavelength to identify the conditions that yield the highest achievable bunching factor at a given harmonic. The maximum attainable bunching factor across all feasible configurations was extracted for each harmonic order.

Figure 5.12 presents the theoretical bunching factors at various harmonics of 800 nm for the four different modulation wavelength combinations discussed earlier. The results demonstrate that the optimal configuration, within the constraints imposed by the limited chicane strengths and achievable modulation amplitudes, is obtained when the first modulation is

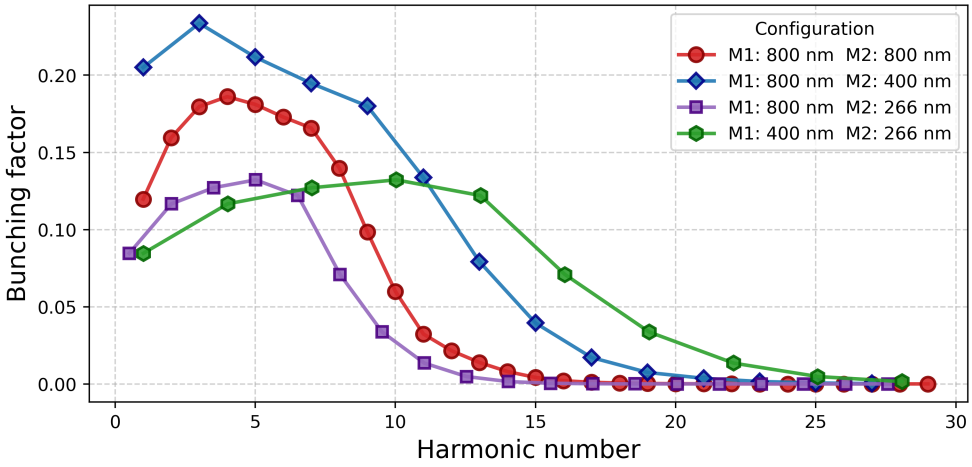

Figure 5.12: Theoretical bunching factor at different harmonics of 800 nm for different combinations of the modulation wavelengths specified in the legend.

performed at 800 nm and the second at 400 nm. The subsequent sections of this chapter focus exclusively on this wavelength combination for more detailed simulations.

5.5.2 Choice of Laser Waist Size

The transverse profile of the laser waist critically influences the microbunching process in EEHG, owing to the finite spatial overlap between the electron beam and the laser field. When the laser waist is comparable to the electron beam dimensions, particles at distinct transverse coordinates experience non-uniform electric field. The electrons near the optical axis undergo stronger energy modulation compared to those at the beam periphery, leading to a transverse-dependent variation in the modulation amplitude. This variation smears out the microbunches and degrades the efficacy of the method. One method to address this undesirable effect is to increase the laser waist size significantly beyond the transverse size of the electron beam. By doing so, the radial intensity gradient of the laser field is reduced. Furthermore, a larger waist provides a longer Rayleigh range, minimizing laser divergence through the modulator. This ensures that electrons across the entire beam experience a more uniform energy modulation. On the other hand, for a given laser pulse energy, increasing the waist size reduces the peak intensity. This introduces a critical trade-off, requiring a search for an optimum waist size.

To systematically determine the optimal laser waist sizes for both energy modulation stages, a series of numerical simulations were performed using the custom simulation procedure described in Section 5.4. In these simulations, the first energy modulation was modeled using an 800 nm laser pulse with a pulse energy of 2.5 mJ and a FWHM pulse duration of 50 fs, while the second used a 400 nm pulse with identical energy and duration parameters. The

transverse laser profiles were modeled as Gaussian beams with $M^2 = 2$ and variable waist sizes $w_0^{(1)}$ and $w_0^{(2)}$ for the first and second modulators, respectively. For each parameter combination, the electrons were tracked through the U250 undulator, and the resultant bunching factor b_n at several harmonics were computed from the final longitudinal phase space distribution.

Since the microbunching process is highly sensitive to both the energy modulation amplitudes and the dispersive strength of the chicanes, selecting the appropriate chicane settings was crucial for optimizing the bunching factor. The optimal chicane parameters had to be recalculated for each set of laser waist sizes, as the modulation amplitudes directly influence the required dispersion. To achieve this, the laser-electron interaction was first simulated for a reference case involving on-axis, on-momentum electrons, from which the energy modulation amplitude was extracted based on the maximum energy deviation of the test particles.

Once the modulation amplitudes were determined, the corresponding optimum chicane strengths were computed theoretically using Eq. 2.69 and 2.65. The required chicane currents were then obtained by converting the dispersion parameter R_{56} into the corresponding current values via the established calibration formula. Given the deviations from the theoretical predictions due to practical constraints and nonlinear effects, additional simulations were conducted by scanning the chicane settings within a predefined range around the theoretical optimum. This range was chosen to lie within the operational limits of the SPEED chicanes, here assumed to span from 0 A to 700 A. For each laser waist configuration, the maximum bunching factor obtained within this range of chicane settings was recorded.

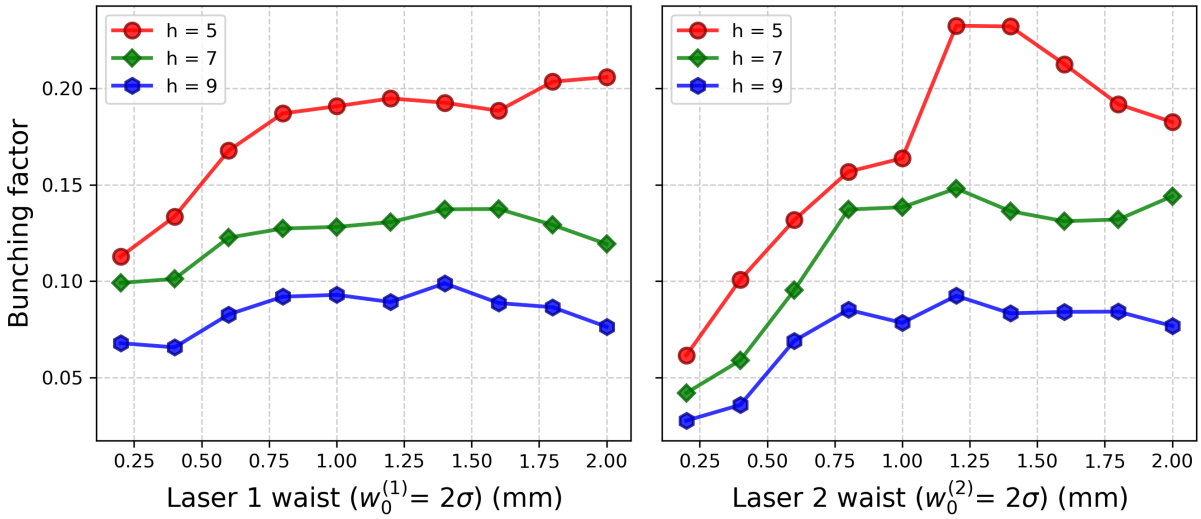

Figure 5.13: Simulated bunching factor at 3 different harmonics of 800 nm for different laser waist sizes of the first (left) and second (right) laser pulse.

Figure 5.13 presents the maximum simulated bunching factor at the 5th, 7th, and 9th harmonics of 800 nm as a function of the laser waist sizes for both the first and second laser

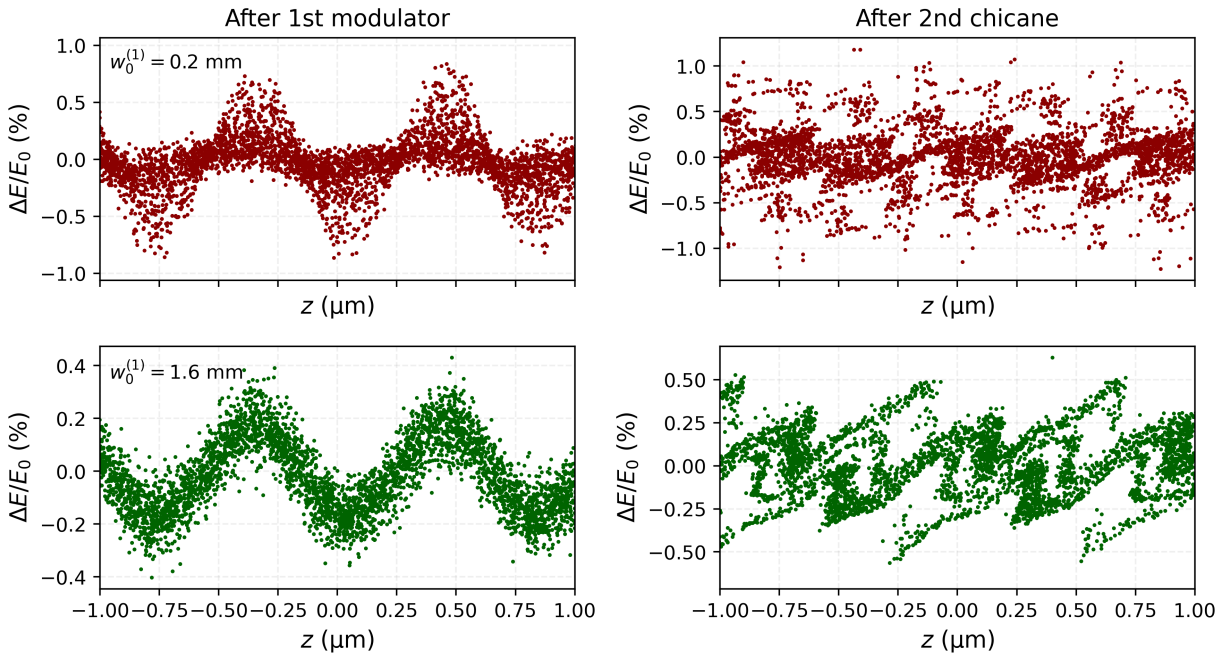

Figure 5.14: Longitudinal phase space distribution of electrons for two different first laser beam waist sizes: 0.2 mm (top row) and 1.6 mm (bottom row) after the first modulator (left column) and after the second chicane (right column).

pulse. In the simulations where the waist of the first laser beam was varied, the waist of the second laser beam was held constant at $w_0 = 1.6$ mm, and vice versa. The results clearly indicate that the waist size of the second laser has a more pronounced effect on the bunching factor, whereas variations in the first laser beam waist lead to a more gradual change.

A closer examination of the data reveals that while a tighter laser focus introduces greater inhomogeneity in the energy modulation across the electron beam. This is primarily because the laser beam becomes smaller than the transverse size of the electron beam, leading to non-uniform interaction across the particle distribution. This effect is particularly pronounced in the case of the second laser, where a tightly focused beam generates highly localized energy modulation. As a result, the energy bands formed after the first chicane are smeared out, leading to the degradation of microbunching.

Figure 5.14 and 5.15 show the longitudinal phase space distribution of the electron beam two different laser waist sizes: 0.2 mm and 1.6 mm. A smaller laser waist leads to a highly inhomogeneous energy modulation which destroys the narrow energy stripes that are essential for the formation of higher harmonics in EEHG, while a larger waist size preserves them.

Conversely, as the laser waist increases beyond the transverse size of the electron beam, the modulation becomes more uniform across the particle distribution, effectively preserving the microstructures necessary for the formation of well-defined microbunches. However, as the waist continues to expand, the energy modulation amplitude gradually decreases. A

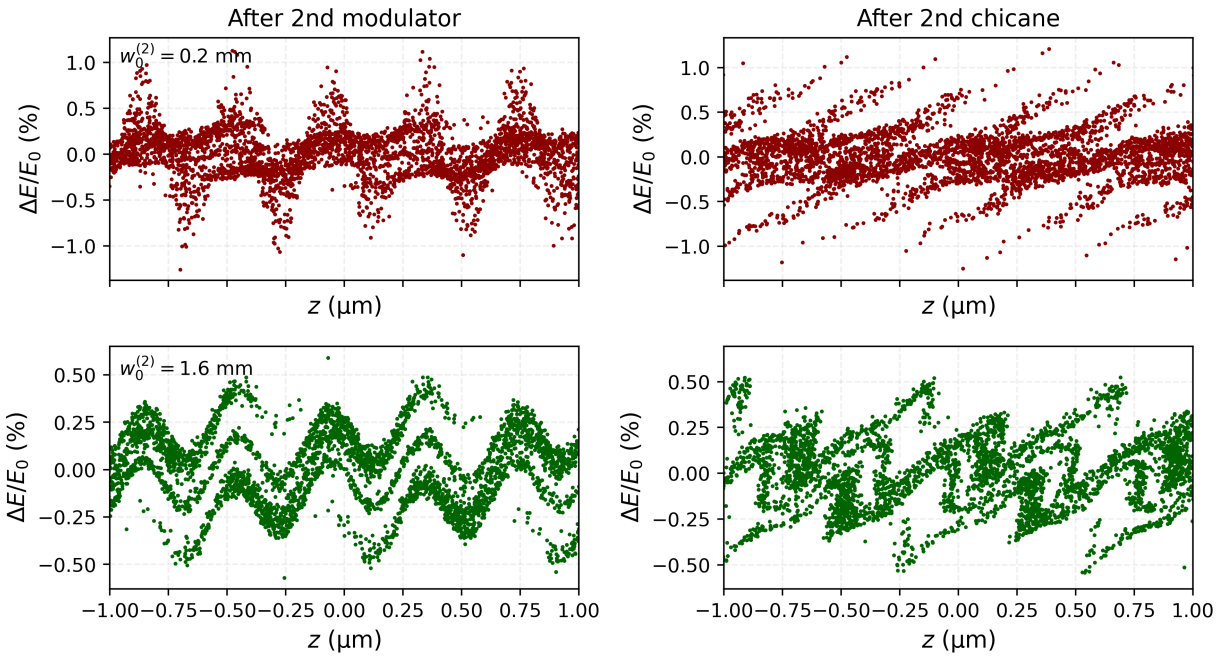

Figure 5.15: Longitudinal phase space distribution of electrons for two different second laser beam waist sizes: 0.2 mm (top row) and 1.6 mm (bottom row) after the second modulator (left column) and after the second chicane (right column).

lower modulation amplitude necessitates a stronger chicane to achieve the optimal longitudinal phase space compression, but beyond a certain point, the chicane strength reaches its practical limits, leading to a decline in the bunching factor.

Based on these simulations, an optimal choice for the laser waist size appears to be in the range of 1.2 to 1.5 mm. This range represents a trade-off between achieving a sufficiently homogeneous energy modulation while maintaining a high enough modulation amplitude for efficient microbunching formation.

5.5.3 Minimizing the R_{52} Value

Another critical issue that requires attention is the presence of nonzero R_{52} values introduced by the asymmetric chicanes (see Section 5.2). The R_{52} term introduces a transverse-longitudinal coupling that smears the microbunching structure and degrades EEHG performance. To correct this, a method was implemented to fine-tune the R_{52} values by independently adjusting the currents in the coil pairs at poles 11 and 16 for the first chicane, and poles 27 and 30 for the second chicane (as shown in Fig. 5.16). The required independent power supplies were already part of the hardware configuration. This allowed the R_{52} correction to be implemented conveniently, avoiding the much larger effort of physically modifying the coils.

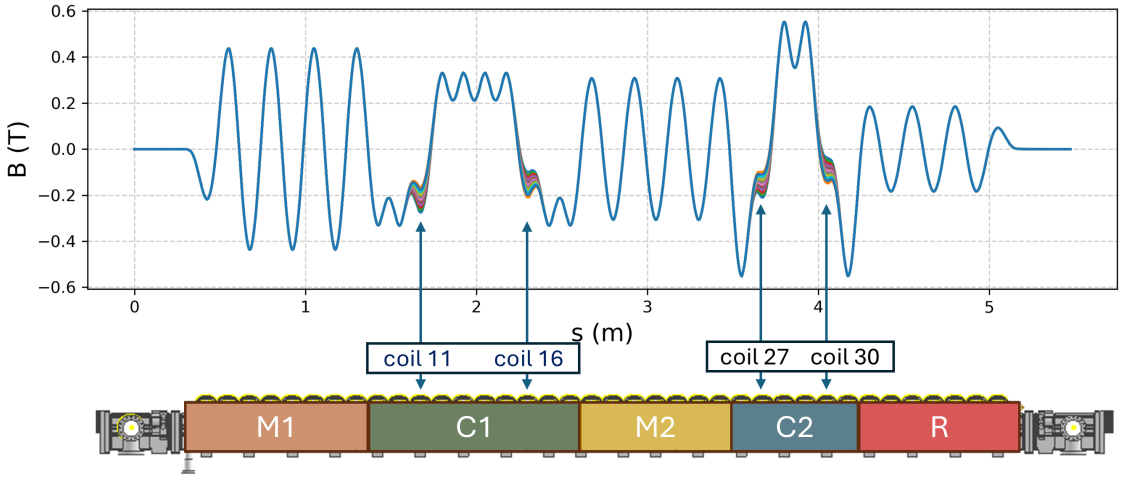

Figure 5.16: Schematic of the SPEED setup (bottom) with its corresponding on-axis vertical magnetic field profile $B(s)$ (top). The highlighted coils (11, 16, 27, 30) are used to fine-tune the R_{52} values of the chicanes. The superimposed traces on the plot show the local magnetic field changes generated by different corrector current settings.

To quantify the impact of the R_{52} correction, simulations were performed to analyze the energy modulation and microbunching for different correction currents applied to coils 11, 16, 27, and 30. The magnetic field configurations for various correction scenarios were simulated using *CST Studio*, while the laser-electron interaction and microbunching evolution were modeled using the custom-developed simulation program. The program calculates the R_{52} coefficient by tracking a distribution of electrons, two of which having identical initial conditions except for a small difference in their horizontal entrance angle $\Delta x'$. The longitudinal displacement Δz between these two electrons at the end of the undulator is then used to compute R_{52} value as $R_{52} = \Delta z / \Delta x'$. Additionally, the program extracts the final electron distribution to estimate the bunching factor.

For the study of R_{52} correction in the first chicane, a specific configuration was chosen where the first modulator was tuned to 800 nm, the second modulator to 400 nm, and the first chicane operated at an excitation current of 350 A. To introduce the R_{52} correction, an additional current increment of dI was applied to coil 11 while the same amount was subtracted from coil 16. Magnetic field configurations were simulated for correction currents ranging from -50 A to +80 A in steps of 10 A. Additionally, the program extracts the final electron distribution to estimate the bunching factor.

Figure 5.17 presents the dependence of R_{52} value and the bunching factor at the exit of the U250 undulator on the correction current applied to coils 11 and 16 for the first chicane. The results indicate a linear relationship between R_{52} value and the correction current, with R_{52} value reaching zero at approximately +50.9 A. Notably, the bunching factor achieves its maximum value at approximately the same correction current.

Similarly, Figure 5.18 illustrates the correction study for the second chicane. Here the configuration is chosen such that the first modulator and chicane are turned off, with the second modulator tuned to 400 nm and the radiator tuned to 200 nm, while the second chicane is set to 500 A. The trend observed in this case is consistent with that of the first chicane, with the bunching factor peaking at the point where R_{52} value is close to zero.

It is worth noting that in both chicanes, the maximum bunching factor does not coincide exactly with the zero-crossing of the R_{52} value. This offset could be due to the dependence of R_{56} on the correction current. As the EEHG bunching mechanism is highly sensitive to the precise tuning of R_{56}, the variations induced by the correction coils can shift the condition for maximum bunching slightly away from the point of perfect R_{52} cancellation.

To further analyze the impact of R_{52} value on the EEHG configuration, simulations were conducted by independently varying the correction currents applied to the first and second chicanes. In these simulations, the first modulator was tuned to 800 nm, the second modulator to 400 nm, and the radiator to 114 nm, corresponding to the 7th harmonic of 800 nm. The laser and chicane parameters were optimized to maximize the bunching factor at this wavelength. This optimization yielded excitation currents of 550 A for the first chicane and 500 A for the second chicane, with energy modulation amplitudes of 0.20% and 0.18% for the first and second modulators, respectively.

Figure 5.19 shows the calculated bunching factor at the 6th, 7th, and 8th harmonics of 800 nm as a function of the correction currents applied to both chicanes. For the second chicane, the bunching factor follows a pattern similar to the one observed in Fig. 5.18, where it reaches a peak around a correction current of 30 A.

In contrast, the effect of the correction current on the first chicane exhibits a more intricate relationship. This complexity arises because modifying the correction current, in addition to influencing the R_{52} value, also modifies the R_{56} value of the chicane. Since the optimal

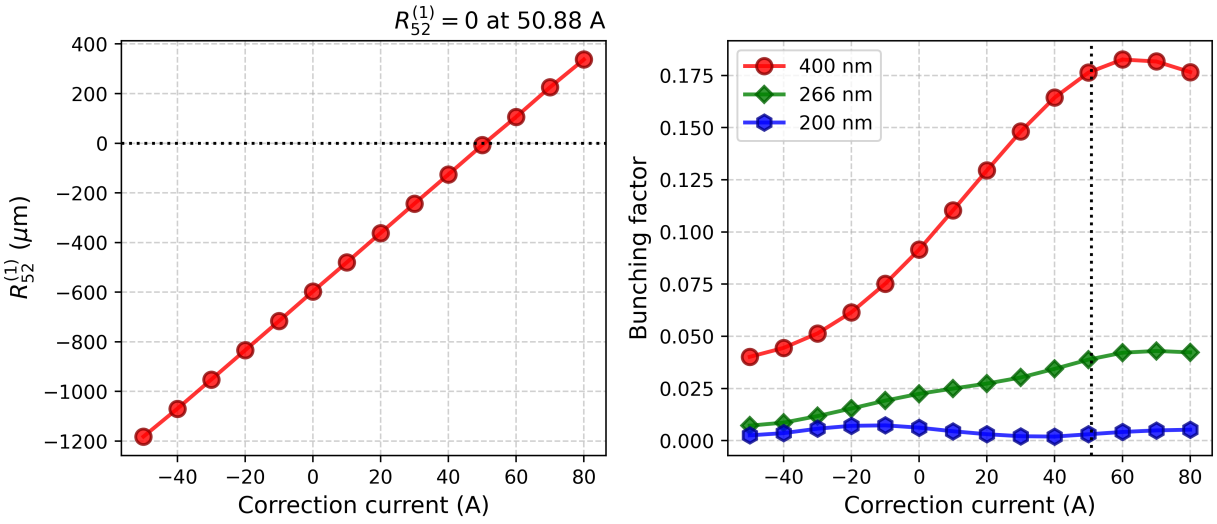

Figure 5.17: Left: Calculated R_{52} value of the first chicane at the U250 exit for different correction currents applied to coils 11 and 16, with modulator 1 at 800 nm, modulator 2 at 400 nm, and chicane 1 at 350 A (while the radiator and second chicane are off). The dotted black line marks $R_{52} = 0$ at 50.9 A. Right: Bunching factor at three harmonics of 800 nm.

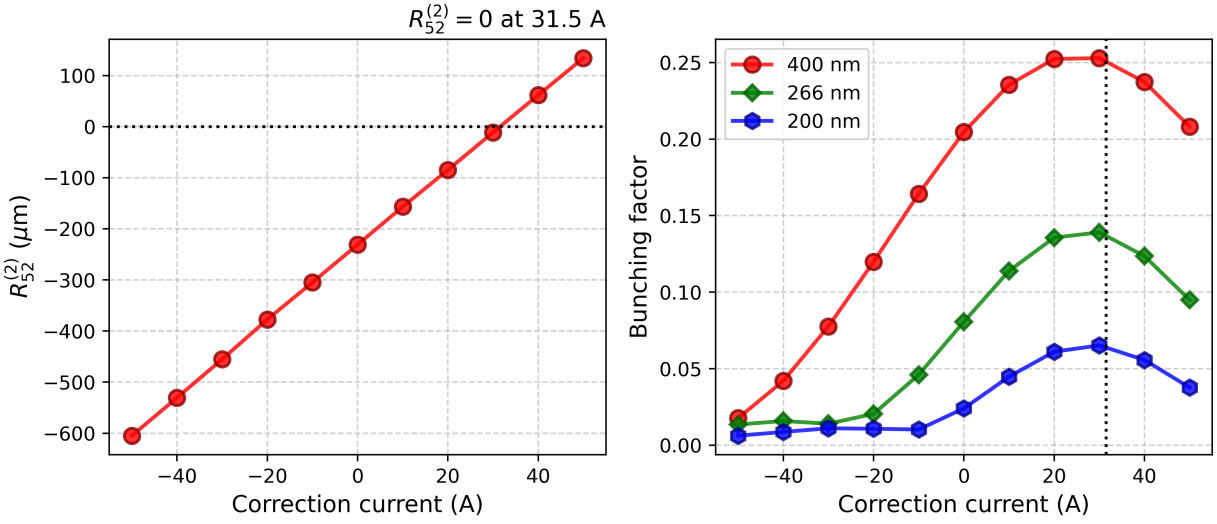

Figure 5.18: Left: Calculated R_{52} value of the second chicane at the U250 exit for different correction currents applied to coils 27 and 30, with modulator 2 at 400 nm, radiator at 200 nm, and chicane 2 at 500 A (while the first modulator and chicane are off). The dotted black line marks $R_{52} = 0$ at 31.5 A. Right: Bunching factor at three harmonics of 800 nm, peaking where R_{52} value is zero.

R_{56} is different for each harmonic, bunching factors of specific harmonics peak at specific R_{56} values, as seen in the case of the 133 nm bunching factor peaking around a correction current of 50 A in Fig. 5.19.

Another observation is that the bunching factor at the target wavelength (114 nm) exhibits

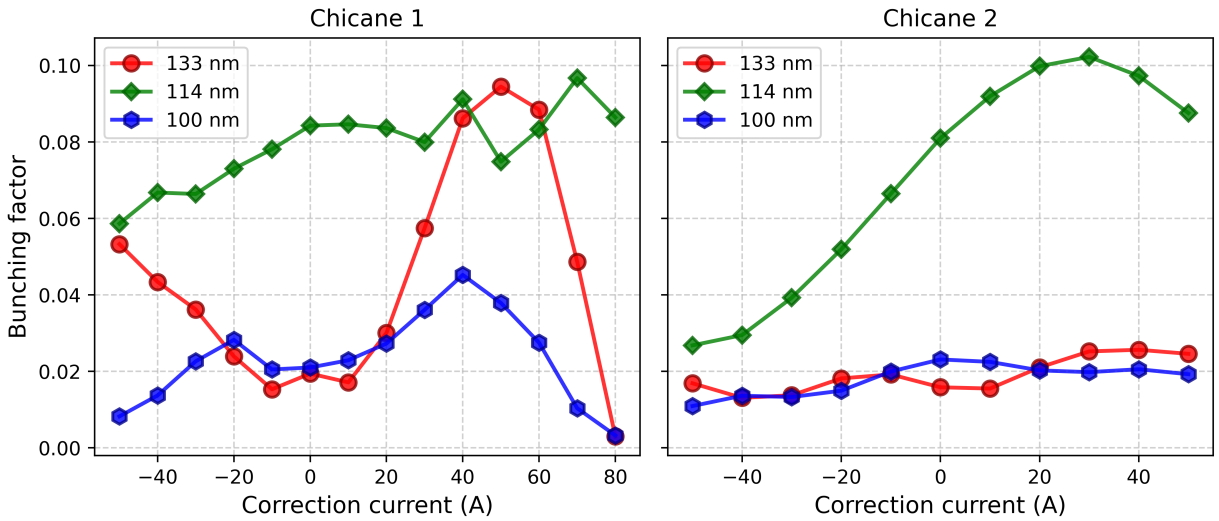

Figure 5.19: Bunching factor at different harmonics of 800 nm for different correction currents of the first chicane (left) and for different correction currents of the second chicane (right).

less variation with the first chicane correction current compared to the second chicane. After the first chicane, the energy modulation imparted by the first modulator has undergone overshearing, meaning that the fine microbunching structures required for coherent emission have not yet formed. At this stage, even if the R_{52} value is nonzero, its impact is less critical. In contrast, after the second chicane, where microbunching has already developed, a large residual R_{52} value can introduce significant smearing, effectively disrupting the fine structures necessary for high-harmonic generation. Consequently, correcting R_{52} value in the second chicane has a much more pronounced effect on the overall performance of the EEHG process compared to the first chicane.

To minimize the effect of a nonzero R_{52} value, a procedure for systematically controlling the currents in the correction coils was implemented and tested. One major challenge comes from the power supplies, which limit the maximum current in the half-coils to 200 A. This limitation is particularly severe for the first chicane due to its high excitation current requirements. Specifically, to set the first modulator to 800 nm, the excitation current needs to be 398 A, meaning that half-coil number 11 must operate at 199 A, nearly reaching the upper current limit of the power supply. As a result, the ability to fine-tune R_{52} value for the first chicane is significantly constrained. In contrast, the second modulator and radiator operate at lower excitation currents, meaning the required currents for half coils of the second chicane are substantially lower, which allows for greater flexibility in R_{52} correction.

It should be noted that unlike R_{52}, the transverse-longitudinal coupling term R_{51} does not require correction here. Since R_{51} describes the dependence of longitudinal position on transverse offset, a parallel displacement of the electron beam does not change the path

length through the chicane, making the R_{51} effect negligible for microbunching.

5.5.4 Optimizing the Chicanes

In the EEHG seeding scheme, precise tuning of the chicane strengths is critical, as the bunching factor is highly sensitive to their configuration. A 2D parameter scan of the bunching factor as a function of the first and second chicane strengths, $R_{56}^{(1)}$ and $R_{56}^{(2)}$ respectively, typically reveals optimal regions, which appear as localized islands in the bunching map. Figure 5.20 illustrates two examples of theoretical bunching maps for different harmonics calculated using Eq. (2.62). In these calculations, the first seed pulse has a wavelength of 800 nm and induces an energy modulation amplitude of 0.20%, while the second seed pulse at 400 nm imparts an energy modulation amplitude of 0.15%. The optimal harmonic generation via EEHG occurs within the identified islands in the bunching map. As the target harmonic number increases, these islands shift towards higher first chicane strengths and slightly lower second chicane strengths. For clarity, all harmonic numbers discussed in this section are relative to the 800 nm fundamental wavelength of the first seed laser.

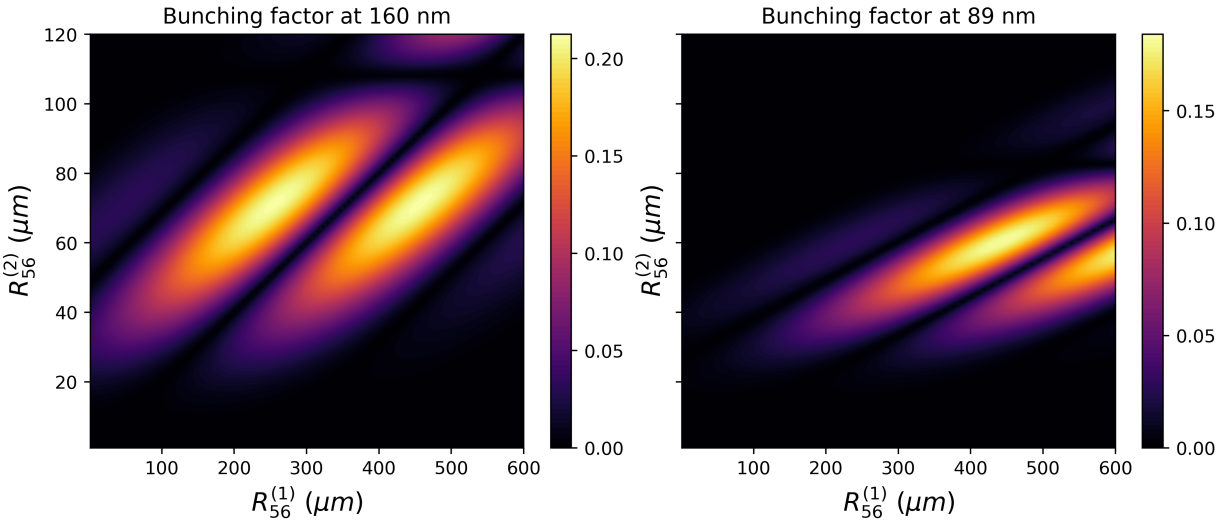

Figure 5.20: Theoretical bunching factor at 160 nm (left) and 89 nm (right) shown as a 2D map for different first and second chicane strengths (referred to here as "bunching maps"). The seed pulses have wavelengths of 800 nm and 400 nm, with energy modulation amplitudes of 0.20% and 0.15%, respectively.

However, the theoretical formula does not account for additional effects that could influence the bunching factor. For instance, CHG-like structures may form in the electron distribution within either modulator, and intricate phase space structures resulting from two-fold modulation could contribute to higher-order harmonics not predicted by the simple model. Furthermore, the unique SPEED configuration at DELTA introduces additional complexity,

as the modulators are interleaved with the chicanes, potentially leading to non-standard phase space structures.

To thoroughly investigate the impact of chicane tuning on the bunching factor, detailed particle tracking simulations using the *Python* code described in Section 5.4 were performed for various chicane strength combinations. Each simulation tracked an ensemble of 1×10^5 electrons within a 10 μm longitudinal slice up to the end of the second chicane. The seed laser pulses were modeled as transform-limited pulses with a FWHM duration of 50 fs, focused at the center of their respective modulators. Both the 800 nm and 400 nm seed pulses had pulse energies of 2.5 mJ, with a beam waist size of 1.0 mm. For this baseline study, an ideal magnetic field profile was constructed using the method described in Section 5.4.1. These parameters resulted in energy modulation amplitudes $\Delta E/E_0$ of approximately 0.20% and 0.15% for the first and second modulators, respectively. At the end of the simulations, the bunching factor was computed from the resulting phase space distribution of the electrons. The primary goal was to understand the qualitative structure of the bunching factor landscape to guide the experimental search, as the true experimental conditions might be different.

Third Harmonic (267 nm)

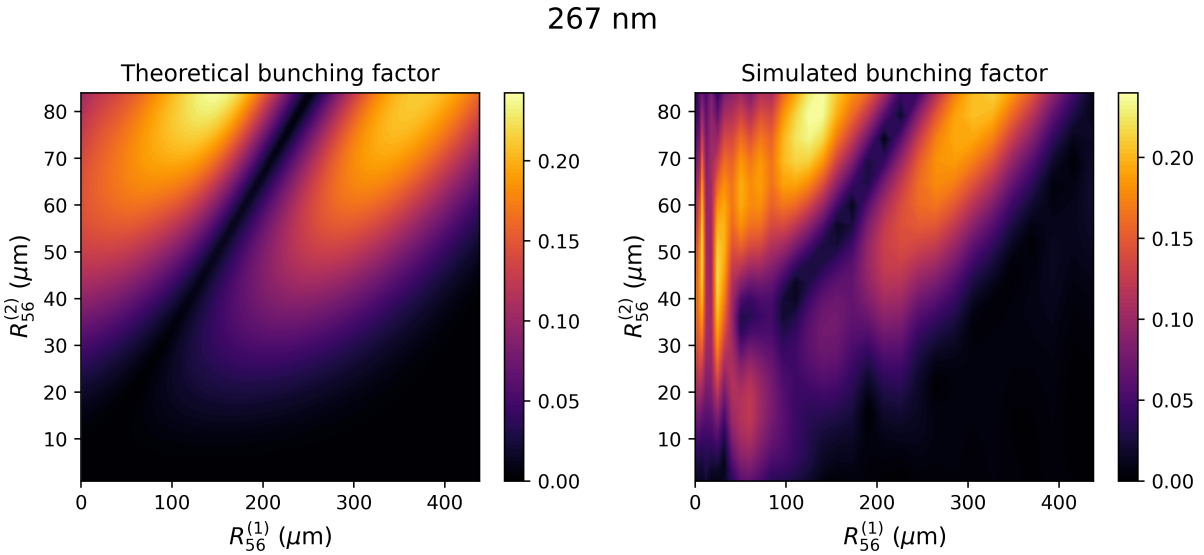

Figure 5.21: Theoretical (left) and simulated (right) bunching maps for the third harmonic (267 nm) generated via the EEHG process.

Figure 5.21 presents the theoretical (Eq. (2.62)) and simulated bunching maps corresponding to the third harmonic (267 nm) of an 800 nm seed laser. The bunching maps exhibit broadly similar features, including two distinct regions of high bunching factor (referred to as "EEHG islands" here). However, a notable discrepancy arises in the simulated map at low values of

$R_{56}^{(1)}$, where vertical fringe patterns are observed. These fringes originate from the sensitivity of the bunching factor to the relative phase difference between the two seed laser pulses, a parameter not explicitly included in the theoretical formalism.

The theoretical model assumes ideal conditions where the first chicane sufficiently shears the beam phase space, effectively averaging out phase-dependent interference effects. At low $R_{56}^{(1)}$ values, however, the energy modulation imparted by the first seed laser is not sufficiently over-sheared. Consequently, the phase space distribution after the second modulation retains sensitivity to the relative phase of the two seed pulses. This sensitivity to the phase difference, which dictates whether the second modulation favorably or unfavorably shapes the phase space, appear as periodic fringes in the simulated bunching map. As $R_{56}^{(1)}$increases, the enhanced shearing decouples the phase relationship between the modulations, suppressing the fringes and aligning the simulation more closely with theory.

Further examination of the phase space structures at the centers of the two "EEHG islands" reveals a crucial distinction between them. The phase space distribution at higher $R_{56}^{(1)}$ exhibits the characteristic EEHG structure with fine energy stripes, which effectively contributes to the desired harmonic generation. However, the phase space distribution at lower $R_{56}^{(1)}$ does not show this fine striping and instead exhibits a complex two-fold modulation while still possessing a strong third harmonic component. Figure 5.22 compares these two distinct phase space distributions.

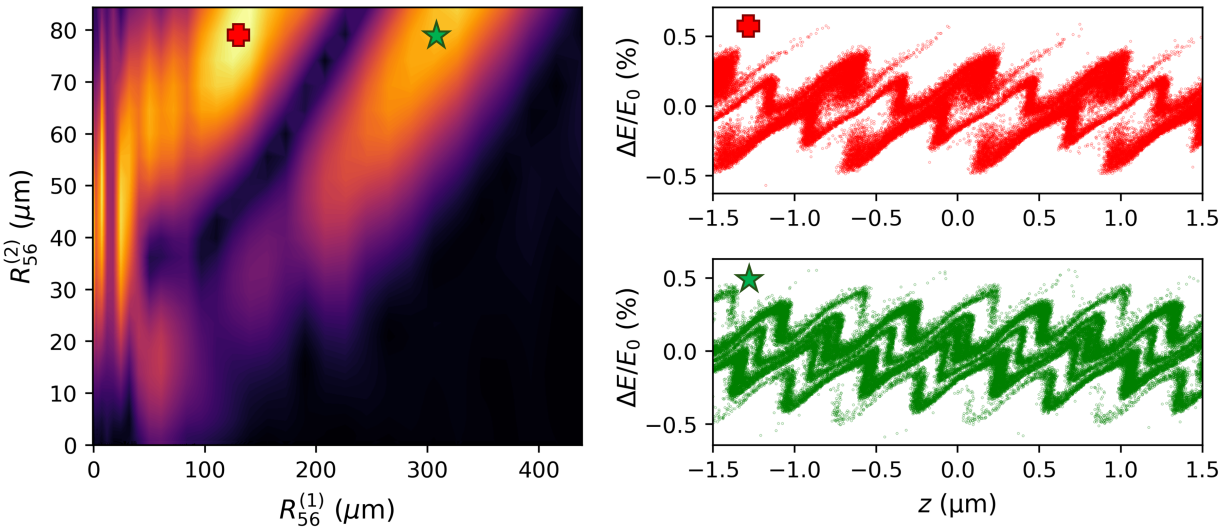

Figure 5.22: Bunching map (left) and longitudinal phase space distributions (right) for the third harmonic (267 nm). The two points marked on the bunching map correspond to the phase space plots. The distribution for the green star (bottom right) exhibits the characteristic structure of EEHG, defined by fine energy stripes, while the distribution for the red cross (top right) lacks these fine stripes.

Fourth Harmonic (200 nm)

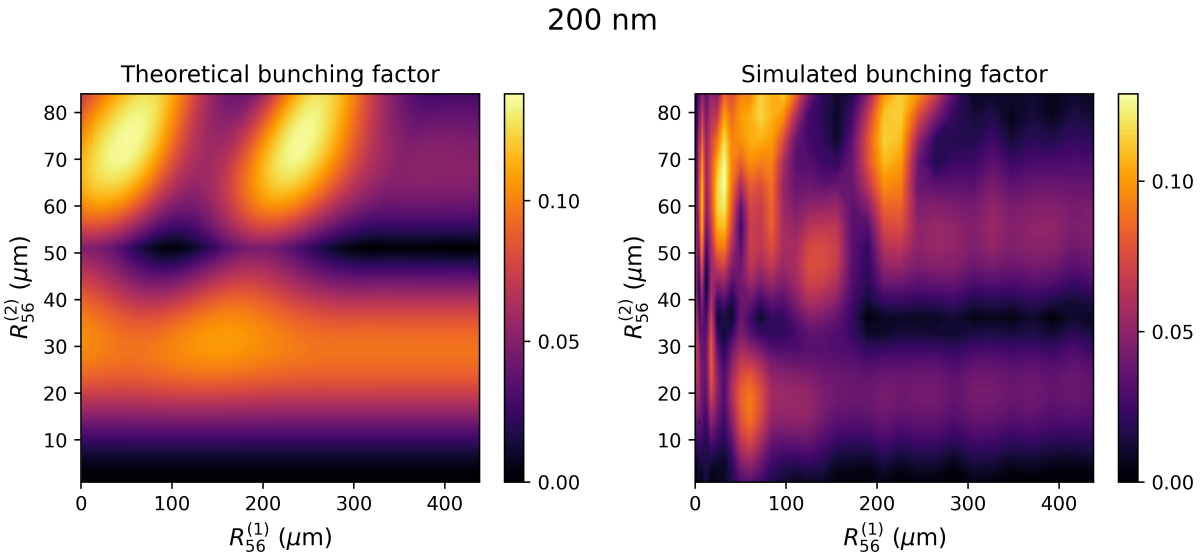

Figure 5.23: Theoretical (left) and simulated (right) bunching maps for the fourth harmonic (200 nm) generated via the EEHG process.

Figure 5.23 presents the theoretical and simulated bunching maps for harmonic generation at 200 nm, corresponding to the fourth harmonic of the 800 nm seed laser. Both maps exhibit qualitative similarities, including the presence of two distinct EEHG islands where the bunching factor is maximized. Vertical fringe patterns, analogous to those observed in the 267 nm case, reappear here as well.

A notable additional feature in this case is the appearance of bands at fixed $R_{56}^{(2)}$ values ($\approx 20\,\mu$m and $50\,\mu$m), where the simulated bunching factor remains nearly constant across a wide range of $R_{56}^{(1)}$. Their insensitivity to the first chicane's parameter suggests that an alternative harmonic generation mechanism dominates in these regions. A likely explanation for these bands is Coherent Harmonic Generation (CHG), where the 200 nm radiation arises from a single-stage process driven only by the second seed laser (400 nm) and the second chicane. The weaker bunching factors in these bands, compared to the EEHG islands, show the reduced efficiency of the CHG process relative to the multi-stage EEHG technique.

Fifth Harmonic (160 nm)

At 160 nm, the theoretical and simulated bunching maps look similar with the fringes still visible at low $R_{56}^{(1)}$. Figure 5.24 illustrates this.

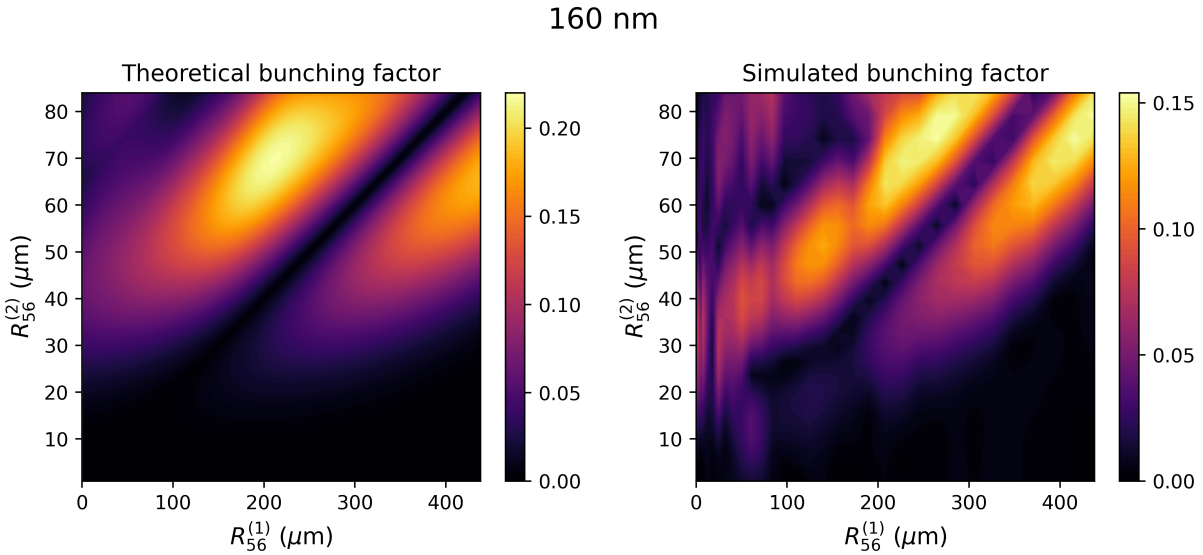

Figure 5.24: Theoretical (left) and simulated (right) bunching maps for the fifth harmonic (160 nm) generated via the EEHG process.

Sixth Harmonic (133 nm)

In Fig 5.25, in addition to the expected features, the constant $R_{56}^{(2)}$ band of non zero bunching factor is also seen since 133 nm is the 3rd harmonic of the second seed laser at 400 nm.

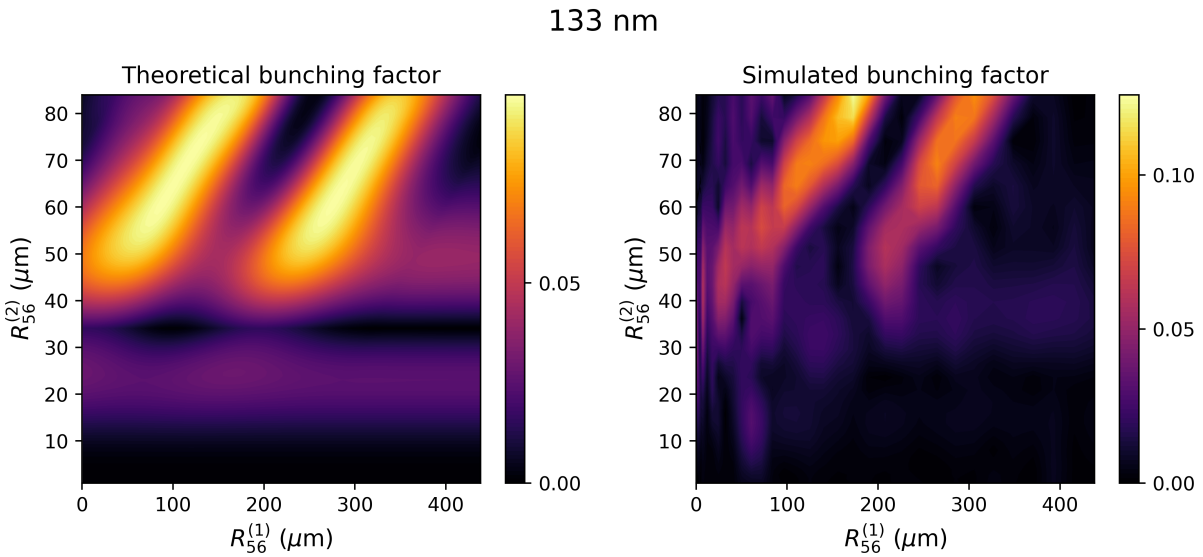

Figure 5.25: Theoretical (left) and simulated (right) bunching maps for the sixth harmonic (133 nm) generated via the EEHG process.

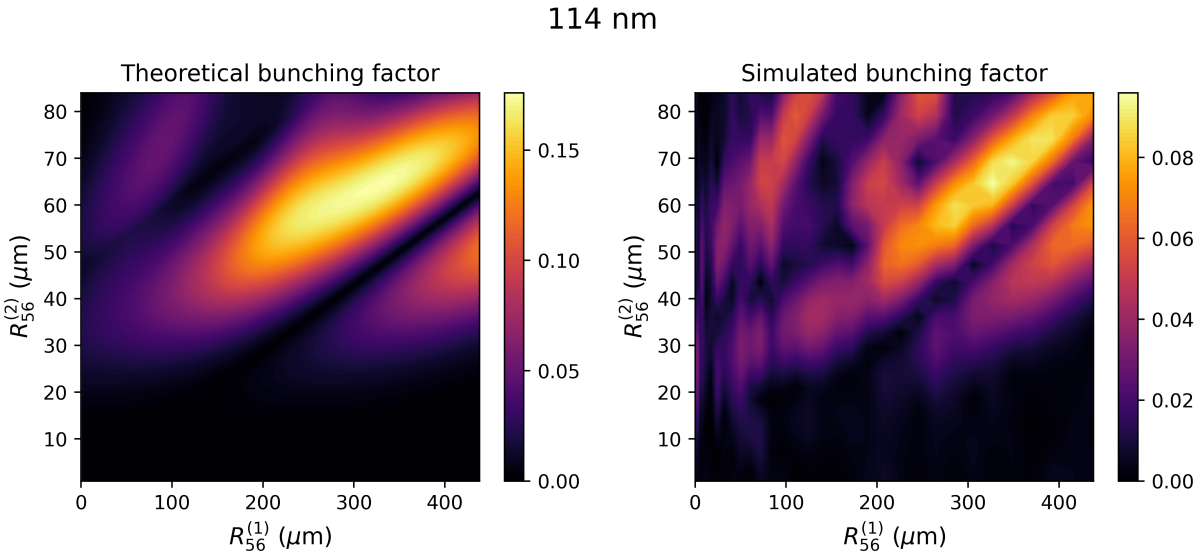

Figure 5.26: Theoretical (left) and simulated (right) bunching maps for the seventh harmonic (114 nm) generated via the EEHG process.

Seventh Harmonic (114 nm)

At the 7th harmonic at 114 nm in Fig. 5.26, more EEHG islands begin to appear (corresponding to different orders of n from Eq. (2.62)), where the second pair appears at lower $R_{56}^{(1)}$ values but with a weaker bunching factor.

Higher Harmonics

Figure 5.27 shows the theoretical and simulated bunching map at the 8th, 9th and 10th harmonic of 800 nm. The optimal first chicane settings shift to higher $R_{56}^{(1)}$ values as the target harmonic is increased. Additionally, the constant $R_{56}^{(2)}$ bands, which arise from the less efficient CHG mechanism, are visible only at even harmonics and become progressively weaker at higher harmonic numbers.

The simulated maps generally show differences in the shape and strength of the EEHG islands compared to the analytical theory. As described earlier, the theoretical maps are derived from an idealized 1D formalism, whereas the simulations track an ensemble of particles through a more realistic magnetic field model. This simulation model captures more complex physics such as the non-uniform energy modulation experienced by the particles at different transverse and longitudinal positions and field overlaps from the interleaved design of the modulators and chicanes which contribute to non-ideal particle trajectories, which are not included in the simple 1D theory. These simulation-specific effects make the bunching map slightly distorted while keeping the general features the same.

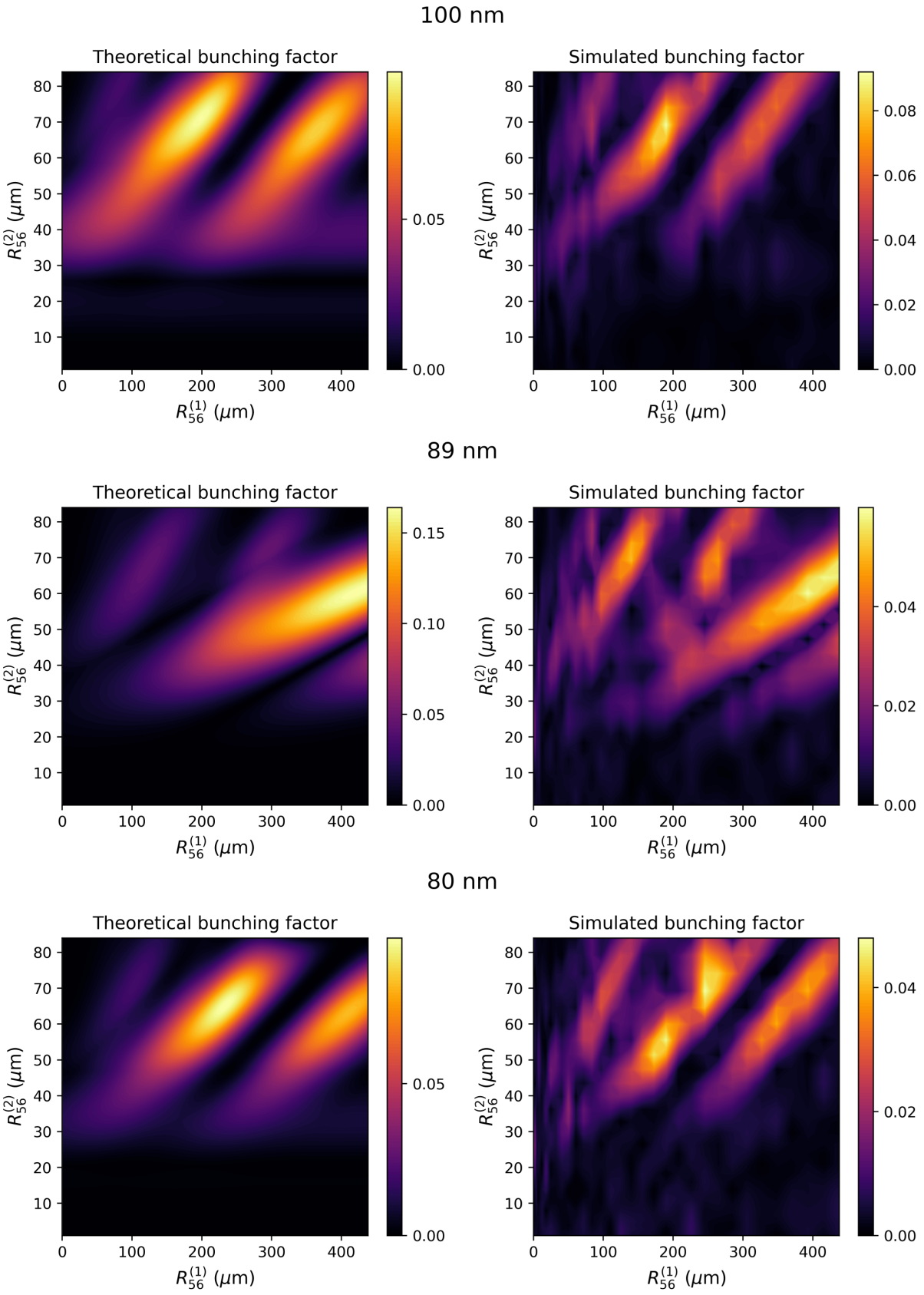

Figure 5.27: Theoretical (left) and simulated (right) bunching maps for the 8th (top), 9th (middle) and 10th (bottom) harmonic generated via the EEHG process.

5.6 Experimental Methods

The EEHG implementation at DELTA utilizes 8 mJ laser pulses at 800 nm from a Ti:sapphire laser system. These pulses are directed through a nonlinear beta barium borate (BBO) crystal for second harmonic generation (SHG). This SHG process converts a fraction of the energy of each pulse, resulting in an output containing both the fundamental 800 nm light and the generated 400 nm second harmonic. The residual 800 nm component serves as the seed for the initial energy modulation, while the 400 nm component is employed for the second modulation stage. Figure 5.28 provides a schematic overview of the laser beam path. To ensure that both modulations affect the same electron slice within the bunch, a motorized linear stage is integrated into the path of the 800 nm pulses for longitudinal timing adjustments.

The focusing optics for the two wavelengths are optimized independently to achieve efficient

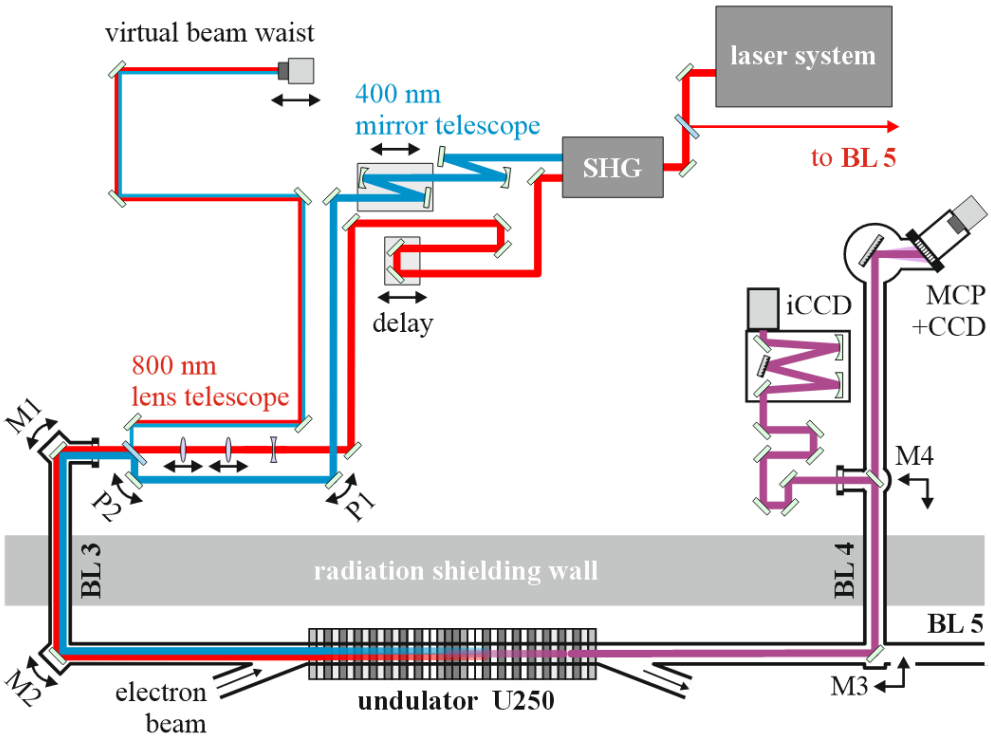

Figure 5.28: Schematic setup with laser beam (800 nm red; 400 nm blue) and EEHG radiation (magenta). Double arrows indicate movable parts. The 8 mJ pulses from the Ti:sapphire laser system undergo second harmonic generation (SHG). The generated 400 nm light is focused by a mirror telescope, while the residual 800 nm light is sent through a variable delay stage and focused by a lens telescope. Leakage from a dielectric 400-nm mirror is monitored by a camera (virtual beam waist). Two dielectric 800/400/266-nm mirrors (M1 and M2 at beamline BL 3) align the laser pulses with the electron beam in the undulator U250. EEHG pulses are sent by an Al-coated water-cooled mirror M3 to beamline BL 4 with two alternative grating spectrometers.

interaction within their respective modulators. The 800 nm beam is focused into the first modulator by a three-lens telescope, with the focus and waist size tunable via two movable lenses. In parallel, a dedicated mirror telescope focuses the 400 nm beam into the second modulator, where the focus can be adjusted by moving one curved mirror of the telescope [72]. A fraction of each beam is extracted and guided onto a rail-mounted CCD beam profiler (WinCamD-UCD12 from DataRay [153]) positioned at a "virtual waist" that replicates the optical path length to the undulator. This setup allows real-time monitoring and alignment of the beam waist.

The laser-electron beam spatial overlap is achieved and optimized using two in-vacuum dielectric mirrors (M1 and M2) installed at beamline BL 3. The CHG and EEHG pulses are extracted by an aluminum-coated, water-cooled mirror (M3) and directed to beamline BL 4, where they are analyzed using either an in-vacuum grating spectrometer [81] for wavelengths smaller than 210 nm, or an on-air Czerny-Turner type spectrometer [79] for 400, 266 and 200 nm signals.

5.6.1 Optimizing the Laser-Electron Interaction

The optimization of laser-electron interaction for the EEHG configuration involves alignment of both spatial and temporal overlap between the electron bunches and the two laser pulses in their respective modulators.

The initial configuration involves setting the U250 undulator with the first modulator tuned to a fundamental wavelength of 800 nm (coil current of 398 A) and the second modulator to 400 nm (279 A). The chicane currents are initially set to moderate values, typically around 350 A for the first chicane and 500 A for the second.

The optimization process begins by introducing the 800 nm laser pulses into the electron beam path. The temporal overlap between the laser pulses and the electron bunches is established using a vector modulator to adjust the RF signal for the laser oscillator, and is monitored using a photodiode and a streak camera. Subsequently, transverse overlap is adjusted by fine-tuning the angles of the in-vacuum mirrors M1 and M2. This procedure is described in detail in Section 2.8. The effectiveness of this alignment is monitored through the intensity of the induced THz signal at BL 5a.

Once the transverse and longitudinal overlap is optimized for the 800 nm pulses, the corresponding mirror angles and vector modulator delay settings are recorded. Following this, the 800 nm beam is blocked, and the 400 nm laser pulses are introduced. These pulses interact with the electron bunches within the second modulator, and their spatial and temporal overlap is independently optimized using the same procedure described above.

Due to differences in the alignment, the optimum angles for the M1 and M2 mirrors differ slightly between the 800 nm and 400 nm beams. This is corrected using the piezo-controlled mirrors P1 and P2 in the 400 nm beam path (see Fig. 5.28). The optimization follows a systematic iterative procedure:

1. With the 400 nm seeding active, the horizontal angle of M1 (M1-x) is gradually adjusted from its optimum position for the 400 nm beam towards its previously determined optimum for the 800 nm beam. Throughout this adjustment, the THz signal is continuously monitored.

2. When the THz signal decreases significantly, the horizontal angle of P1 is fine-tuned to restore it. This iterative adjustment is repeated until M1-x reaches the optimum position for the 800 nm beam while maintaining a strong THz signal.

3. The process is then repeated for the vertical angles of M1 and P1, ensuring that the overlap is optimized in both transverse dimensions.

4. The same optimization is conducted for M2 and P2, aligning both mirrors to maintain optimal overlap with both the 800 nm and 400 nm beams.

The result is a mirror configuration where M1 and M2 remain at their 800 nm-optimized positions, while P1 and P2 compensate for the offsets in the 400 nm path.

To ensure that both pulses interact with the same longitudinal slice of the electron bunch, an approach utilizing the THz radiation emitted by the bunch is followed [154]. The result of this method is shown in Fig. 5.29, which plots the THz signal (top) and the EEHG signal (bottom) as a function of the delay between the two seed pulses. When the relative delay between the seed pulses is in the picosecond range, the interaction produces two distinct dips in the longitudinal electron density profile, leading to the emission of two temporally separated THz pulses. The interference between these THz pulses results in an intensity oscillation of the THz signal as a function of the delay between the seed pulses, as seen in Fig. 5.29 (top). However, as the relative delay approaches zero, both laser pulses act on the same electron slice simultaneously, reducing the overall number of electrons participating in the coherent THz emission. Consequently, this results in a characteristic dip on top of the central maximum of the interference pattern, which serves as a reliable signature of two-fold energy modulation. This dip is clearly observed near zero delay in Fig. 5.29 and overlaps with the induced EEHG signal, further validating the efficacy of this method.

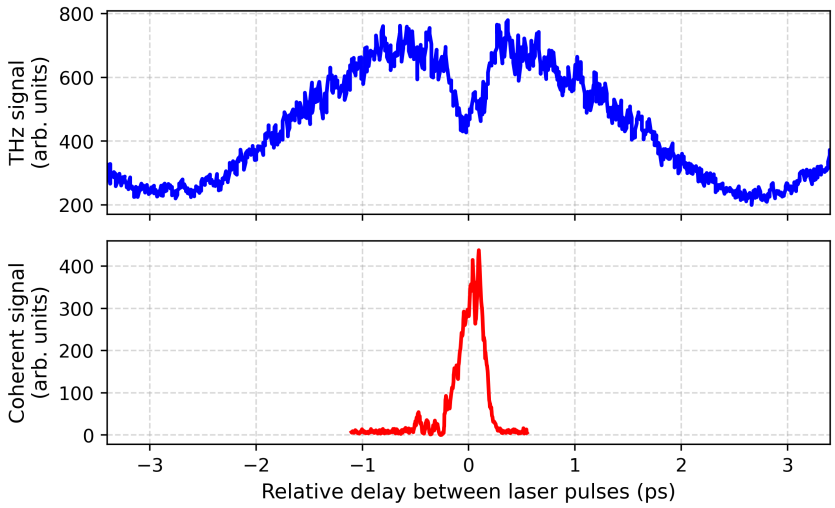

Figure 5.29: THz intensity as a function of the relative delay between the seed pulses (top, blue). The dip in the intensity indicates the zero-delay condition where both the seed pulses act on the same electron slice. The corresponding EEHG signal (bottom, red) observed at 160 nm using an in-vacuum XUV spectrometer.

5.7 Experimental Observation of EEHG Radiation

Since the wavelength used for the second energy modulation is half of the first modulation wavelength, higher bunching factors are expected for the odd harmonics of the first modulation wavelength (see Section 2.6.3). By targeting odd harmonics of the fundamental 800 nm wavelength, one can unambiguously distinguish the EEHG signal from the CHG signal of the 400 nm laser pulses.

The EEHG signal is observed in the same way as in the CHG setup with the in-vacuum XUV spectrometer for wavelengths below 200 nm. For each target harmonic, the experimental search begins by setting the radiator and spectrometer to the desired wavelength and tuning the relative delay between the seed pulses to that corresponding to the dip in THz intensity. A coarse, manual scan of the chicane strengths helps to locate an initial coherent signal. To verify the signal's origin, the timing of the seed lasers is moved by more than the bunch length using the vector modulator. This effectively removes the laser-electron interaction while allowing the laser radiation to still reach the beamline. This helps in ruling out false signals arising from stray laser light or high-order harmonics generated at the optical surfaces.

Figure 5.30 shows two example raw images captured by the XUV spectrometer. The horizontal axis represents the dispersive plane (wavelength), while the vertical axis corresponds to the spatial profile. The top panel reveals a distinct signal centered at approximately $x = 750$ pixels, corresponding to the target EEHG radiation at 160 nm. In contrast, the bottom panel shows the spectrometer output when the laser timing is detuned to remove the laser-electron overlap. The larger feature persisting in both images (around $x = 150$ pixels)

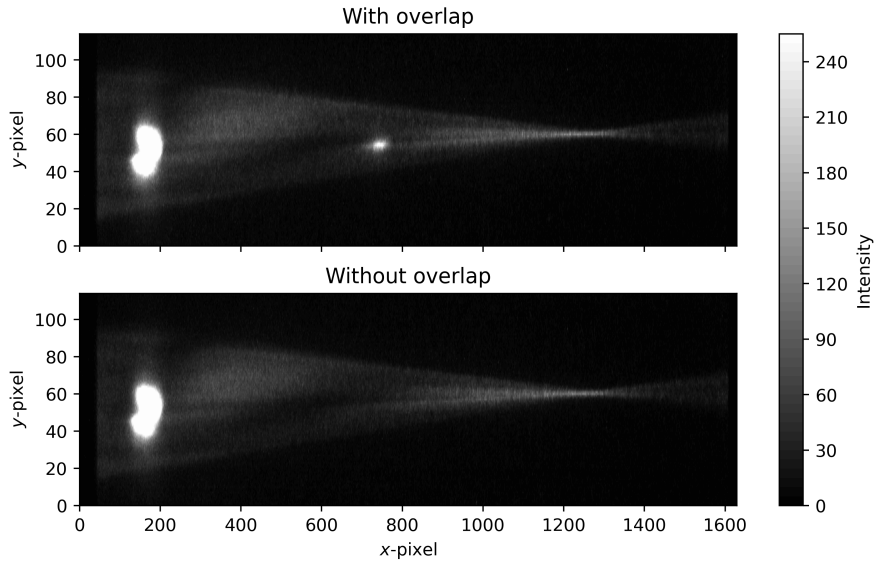

Figure 5.30: Raw images from the XUV spectrometer. Top: With laser-electron overlap, the coherent EEHG signal at 160 nm appears as a bright spot centered at $x \approx 750$. Bottom: When the overlap is removed, the central signal vanishes, confirming the laser-induced nature of the signal. The persistent feature on the left ($x \approx 150$) corresponds to the 133 nm third-harmonic generated on optical surfaces. The broad features in the background are attributed to the incoherent synchrotron radiation and stray laser light.

is identified as a 133 nm signal. This is suspected to be a third-harmonic generated at a mirror surface, as evidenced by its insensitivity to the temporal overlap. The background exhibits a conical shape due to the focusing properties of the spectrometer grating, while the broad features arise from the incoherent synchrotron radiation background and stray laser light. For the analysis of the EEHG signal, only the background-subtracted intensity of the central feature is considered, while the spectral analysis is performed by integrating the signal vertically within a region of interest around the central spot.

Further maximization of the signal intensity requires fine alignment of the spectrometer system. This involves optimizing several degrees of freedom, including its transverse position and angle, as well as the angle of the mirror M3 directing the beam onto the spectrometer (see Fig. 5.28).

A notable experimental limitation is the degradation of the harmonic signal at high chicane currents, typically above 650 A. This signal detrimentation is attributed to the effects of magnet saturation on the electron beam orbit. As the chicane magnets approach saturation, their total field integral becomes non-zero, introducing an uncompensated angular kick to the electron trajectory. This introduces an uncompensated R_{52} term as detailed in Section 5.5.3, which can shear the electron distribution in longitudinal phase space and directly suppress the microbunching. Additionally, the orbit kick steers the electron beam away from the optimized spatial overlap with the seed laser pulses. Finally, a significant deviation in the beam

trajectory can cause the coherently emitted radiation to partially or fully miss the narrow entrance slit of the spectrometer, leading to artificially low or zero intensity measurements.

To mitigate these issues, a feedforward correction was implemented. This procedure automatically adjusts pole 30 (see Fig. 5.16) of the U250 based on a pre-calculated table during the second chicane scans. While this enables measurements up to 790 A, orbit stability is not perfectly maintained due to imperfections in the correction, and measurements above 650 A still suffer from reduced signal intensity and stability.

5.7.1 Fifth Harmonic (160 nm)

The 5th harmonic at 160 nm, the longest wavelength accessible to the in-vacuum XUV spectrometer, produces the most intense EEHG signal observed. However, the signal at this wavelength is superimposed on another parasitic signal, which is also dependent on the delay between the seed pulses. Figure 5.31 shows both signal components. While the EEHG signal is present only during temporal overlap of the laser pulses and the electron bunch, the parasitic signal persists regardless of the presence of the electron beam. Its intensity depends only on the relative temporal delay between the 800 nm and 400 nm seed laser pulses. The red curve in Fig. 5.31 shows this laser-only signal independent of laser-electron overlap, suggesting it is an optical artifact generated via a third-order nonlinear process. It is hypothesized that this is sum-frequency generation, where two photons from the 400 nm beam and one from the 800 nm beam combine to produce the observed 160 nm light. The origin of this harmonic signal is currently unconfirmed, but it likely originates from a nonlinear interaction on an optical surface [155]. Since the efficiency of a third-order nonlinear process scales with the cube of the laser intensity, the most probable sources are the optical surfaces where the laser spot size is smallest, namely the second dielectric mirror (M2) and the final aluminum-coated mirror (M3).

The green curve in Fig. 5.31 shows the net EEHG signal, isolated by subtracting the background from the total signal measured during laser-electron overlap. This signal was initially found with the first chicane set to a strong dispersive strength of $R_{56}^{(1)} = 220$ μm and the second chicane set to a relatively weak strength of $R_{56}^{(2)} = 60$ μm. A notable feature of the delay scan is its slight asymmetry. This is likely due to the asymmetry introduced to the laser temporal profile from the uncompensated third-order dispersion (TOD), as discussed in Chapter 3. The observation of a signal under these specific conditions where the first chicane is strong enough to shear the initial energy modulation past the optimum bunching condition ("overshearing") is a primary indicator of a higher-order bunching mechanism like EEHG.

To confirm the mechanism as EEHG, an extensive scan of the chicane strengths was performed to map the intensity of the coherent signal for different combinations of the chicane

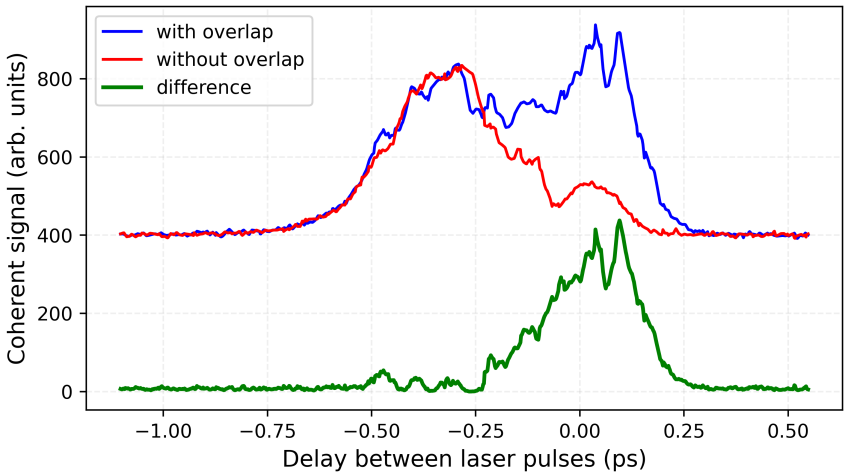

Figure 5.31: Measured intensity at 160 nm as a function of the temporal delay between the seed laser pulses. The blue curve is the raw signal, containing both the EEHG contribution and the parasitic signal. The red line indicates the parasitic signal measured without laser-electron overlap. The green curve shows the difference signal.

settings. Figure 5.32 shows the signal intensity as the second chicane was scanned from 0 μm to 110 μm for several fixed values of the first chicane strength. The current setting of the first chicane and the corresponding $R_{56}^{(1)}$ values are specified for each subplot. The red curve in Fig. 5.32 shows the background-subtracted signal. To better visualize the underlying trend free from high-frequency noise, a smoothed curve (blue) was generated using a Savitzky–Golay filter [156]. Unless otherwise specified, all intensity and spectra plots in this section are presented after subtracting the background measured without laser-electron overlap.

At low $R_{56}^{(1)}$ values, the data reveal a single broad peak. As $R_{56}^{(1)}$ is increased, this peak becomes narrower and shifts to higher $R_{56}^{(2)}$ values, while a second peak emerges at a lower $R_{56}^{(2)}$ value. This observed two-peak structure is a hallmark of the EEHG process. The two peaks correspond to the two optimal sets of chicane strengths that maximize the bunching factor for a given harmonic, a signature that cannot be reproduced by simpler processes like CHG. This observation corresponds to the theoretical and simulated bunching maps shown in Section 5.5.4 and provides unambiguous evidence of successful EEHG.

To validate the experimental results, the observed intensity profile was compared against theoretical calculations of the bunching factor squared using Eq. 2.62. The theoretical bunching map for a grid of values of energy modulation amplitudes was calculated, and the best quantitative fit to the observations was found at $\Delta E_1/E_0 = 0.11\%$ and $\Delta E_2/E_0 = 0.14\%$, for which the mean squared error (MSE) between the theoretical bunching factor squared and the measured intensity profile was minimized. These energy modulation amplitudes were lower than what the simulations predicted (0.16 % and 0.15 % for the first and second mod-

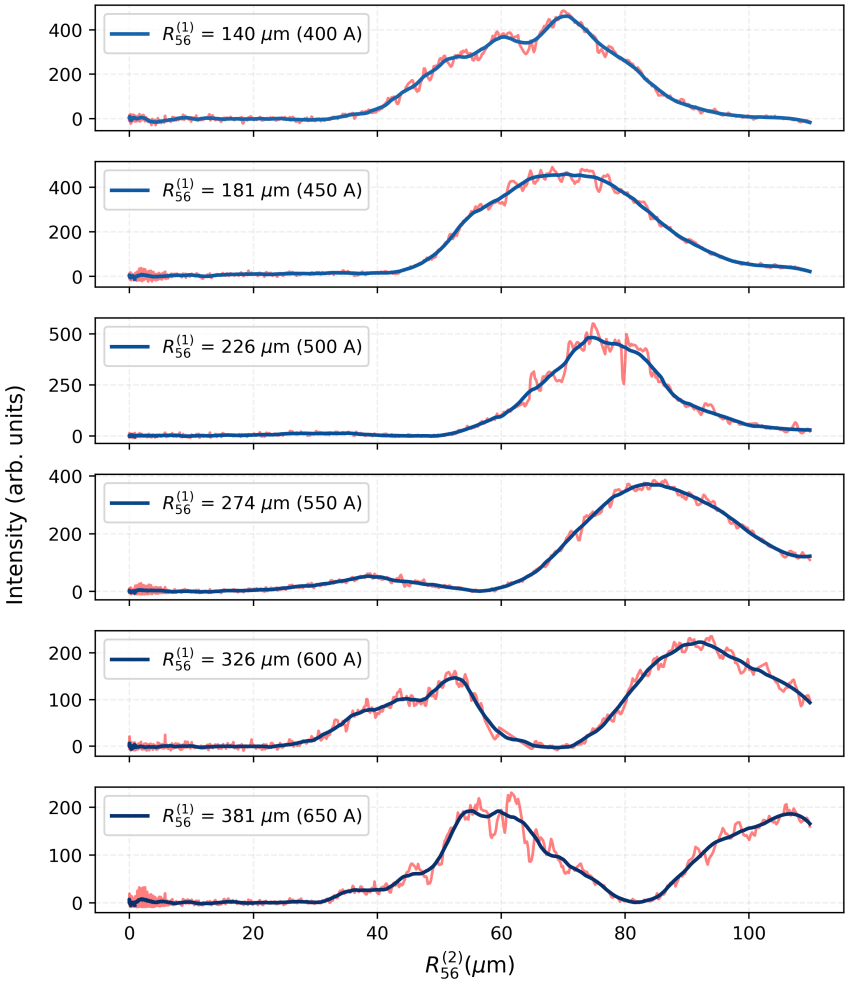

Figure 5.32: Measured 160 nm coherent signal intensity as a function of chicane strength. Each subplot shows the signal intensity (red curve) as the second chicane strength $R_{56}^{(2)}$ is scanned for a fixed value of the first chicane strength $R_{56}^{(1)}$. The overlaid blue curve is a smoothed representation using a Savitzky-Golay filter.

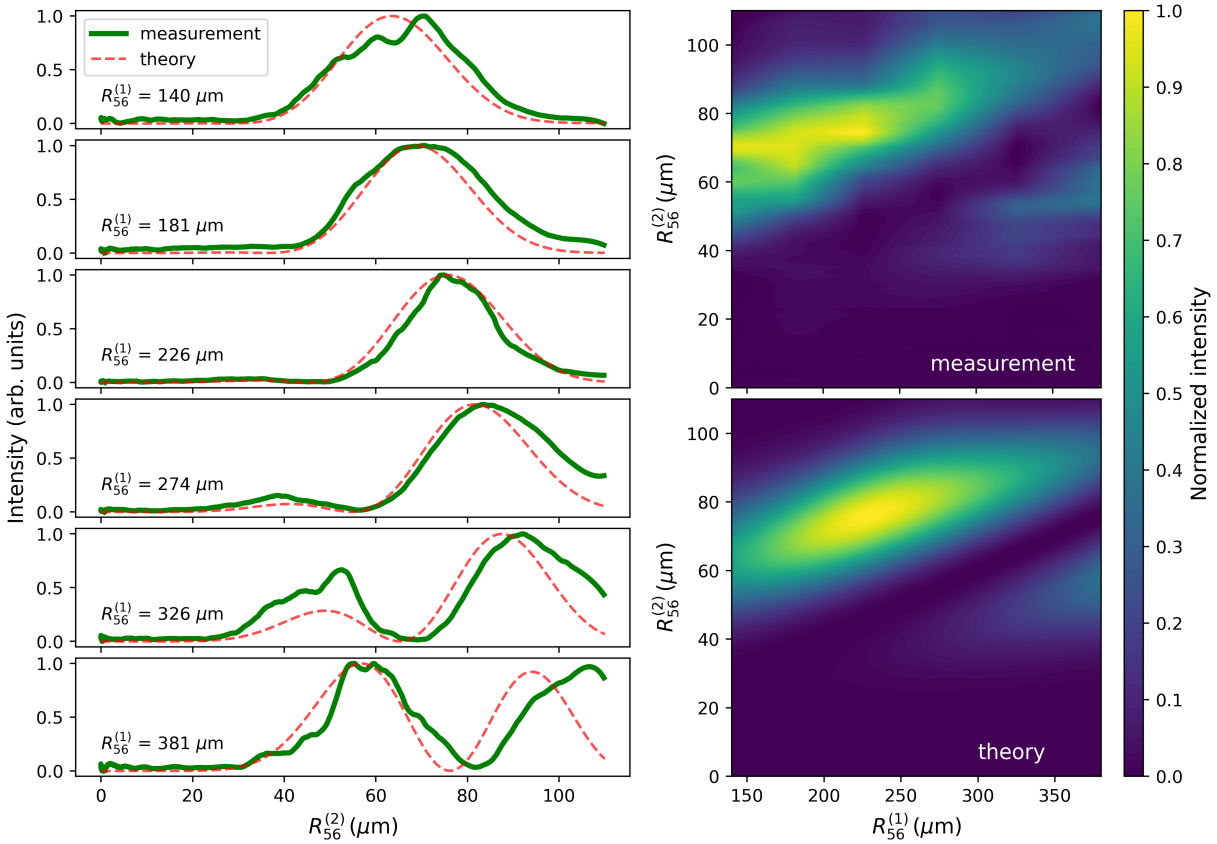

Figure 5.33: Comparison between the observed coherent signal intensity and the theoretical squared bunching factor at 160 nm. Left: Scans of the coherent signal intensity as a function of $R_{56}^{(2)}$ for six different $R_{56}^{(1)}$ values. The solid green line represents the smoothed experimental data, while the dashed red line shows the theoretical bunching factor squared. Right: 2D intensity maps constructed from experimental scans (top) and theory (bottom). The strong qualitative agreement in the positions of the two intensity maxima validates the EEHG model for the experimental observation.

ulations, respectively, for pulse energies of 1.8 mJ and 2.2 mJ, which were the measured laser power going into the seed beamline). Similar discrepancies have previously been observed and reported in [59, 157] and may arise from suboptimal laser-electron overlap, incorrect focal position, or reduced reflectivity of the in-vacuum mirrors.

Figure 5.33 presents a comparison between the experimental data and the theoretical bunching factor squared (left column, dashed red curve). The theoretical bunching map accurately reproduces several important features of the measurement (left column, solid green curve), such as the positions of the intensity peaks and the characteristic appearance of a second peak as $R_{56}^{(1)}$ is increased. To visualize the full parameter space, the experimental scans were used to generate a 2D intensity map via interpolation. This map is color-coded according to the normalized intensity. It is important to note that such interpolation from discrete data can potentially introduce visual artifacts. Despite this, the interpolated experimental

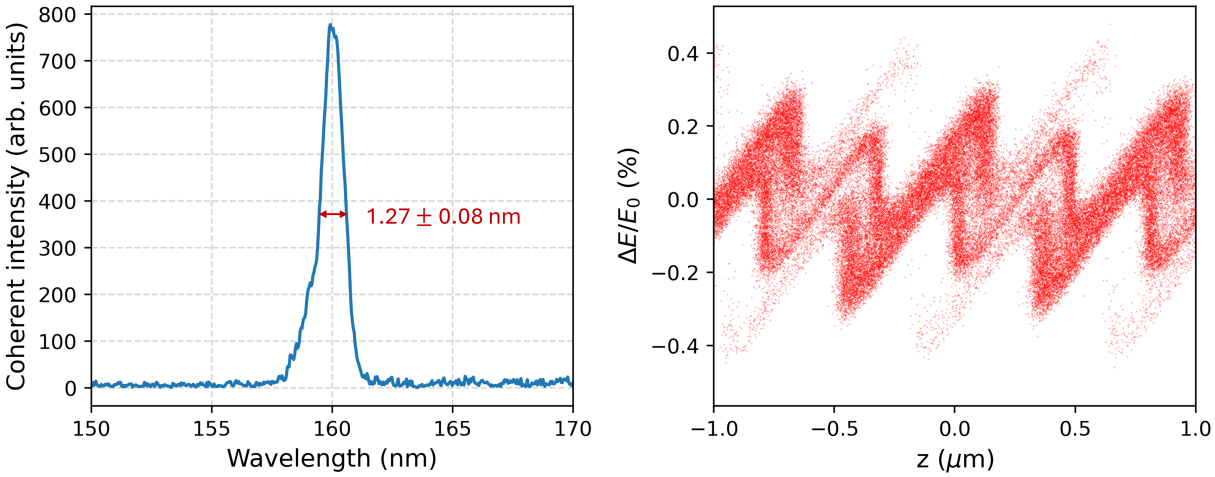

Figure 5.34: The measured spectrum of the coherent signal at 160 nm for the optimized chicane configuration ($R_{56}^{(1)} = 140\mu$m, $R_{56}^{(2)} = 72\mu$m) after background subtraction, showing an emission peak with an FWHM of 1.27 ± 0.08 nm. The figure on the right shows the corresponding simulated longitudinal phase space distribution of electrons after the second chicane for first and second energy modulation amplitudes of 0.11% and 0.14%.

map (top right) and the theoretical map (bottom right) show excellent qualitative agreement. This strong correspondence between measurement and theory provides compelling validation of the EEHG model.

Further characterization of the coherent emission was performed at the optimal chicane settings of $R_{56}^{(1)} = 226$ μm and $R_{56}^{(2)} = 60$ μm. The measured spectrum of the signal after background subtraction is shown in the left panel of Fig. 5.34. The spectrum reveals an emission line with a narrow bandwidth, having an FWHM of 1.27 ± 0.08 nm. The uncertainty here is largely determined by the spectrometer resolution of 0.07 nm, with a minor contribution from statistical fluctuations. Assuming a Fourier-limited Gaussian profile, this bandwidth corresponds to a minimum pulse duration of approximately 30 fs. The corresponding simulated longitudinal phase space distribution of the electrons under these conditions is shown on the right panel of Fig. 5.34. The distribution exhibits a fine periodic structure typical of the EEHG mechanism.

5.7.2 Seventh Harmonic (114 nm)

A significant advantage at 114 nm is the complete absence of the parasitic signal that complicated the measurement at 160 nm. This allowed for the acquisition of a clean EEHG signal, as confirmed by the temporal delay scan presented in Fig. 5.35. The signal, initially found with chicane strengths of $R_{56}^{(1)} = 226$ μm and $R_{56}^{(2)} = 50$ μm, exhibits a distinct peak only when the laser and electron bunch are in temporal overlap. Similar to the measurements at 160 nm, the signal displays an asymmetry, likely resulting from non-zero TOD.

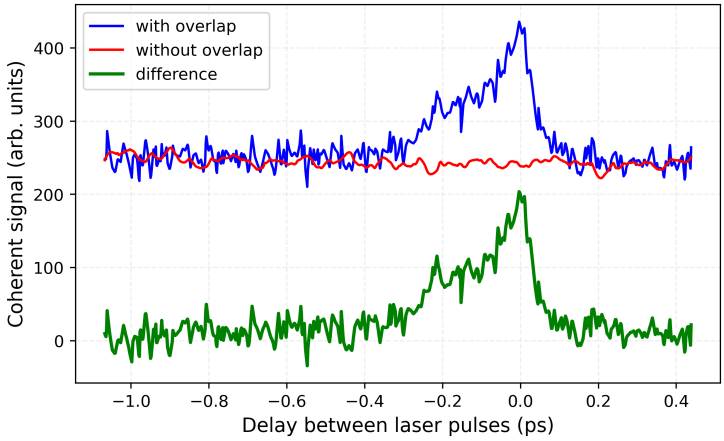

Figure 5.35: Measured intensity at 114 nm as a function of the temporal delay between the seed laser pulses. The blue curve is the signal with laser-electron overlap, and the red line indicates the background intensity measured without the overlap. The green curve shows the net EEHG signal after background subtraction.

To verify the origin of the signal, scans of the second chicane strength performed for three different $R_{56}^{(1)}$ settings are shown in Fig. 5.36. Similar to the case of the 160 nm signal, the peak shifts to higher $R_{56}^{(2)}$ values as $R_{56}^{(1)}$ is increased. On the other hand, the appearance of two peaks similar to the ones present at large $R_{56}^{(1)}$ values for 160 nm is absent at this harmonic within the scanned parameter range. This absence is consistent with theoretical expectations for higher harmonics. Figure 5.37 shows the comparison between the measured intensity and the theoretical bunching factor squared at 114 nm. Here the best fit was found at energy modulation amplitudes of $\Delta E_1/E_0 = 0.12\%$ and $\Delta E_2/E_0 = 0.14\%$. The measurements show excellent agreement with the theory. In this case, the second maximum of the bunching factor appears at a higher first chicane strength that was beyond the R_{56} range of the chicanes.

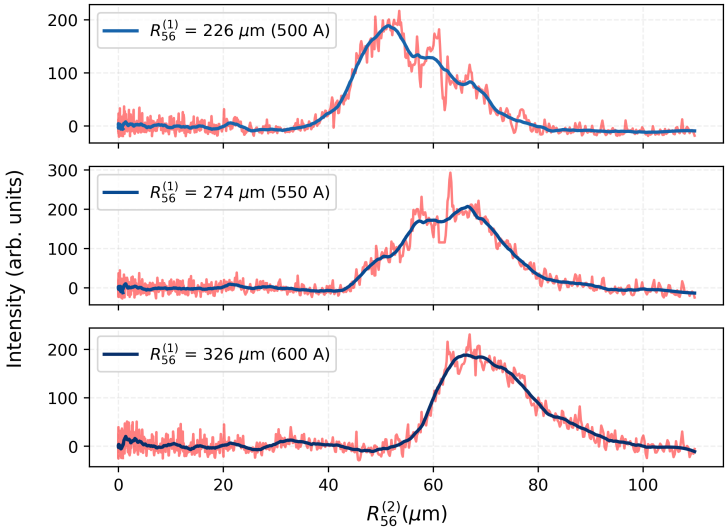

Figure 5.36: Measured 114 nm coherent signal intensity as a function of second chicane strength. Each subplot shows the signal intensity (red curve) as $R_{56}^{(2)}$ is scanned for a fixed value of $R_{56}^{(1)}$. The overlaid blue curve is a smoothed representation using a Savitzky-Golay filter.

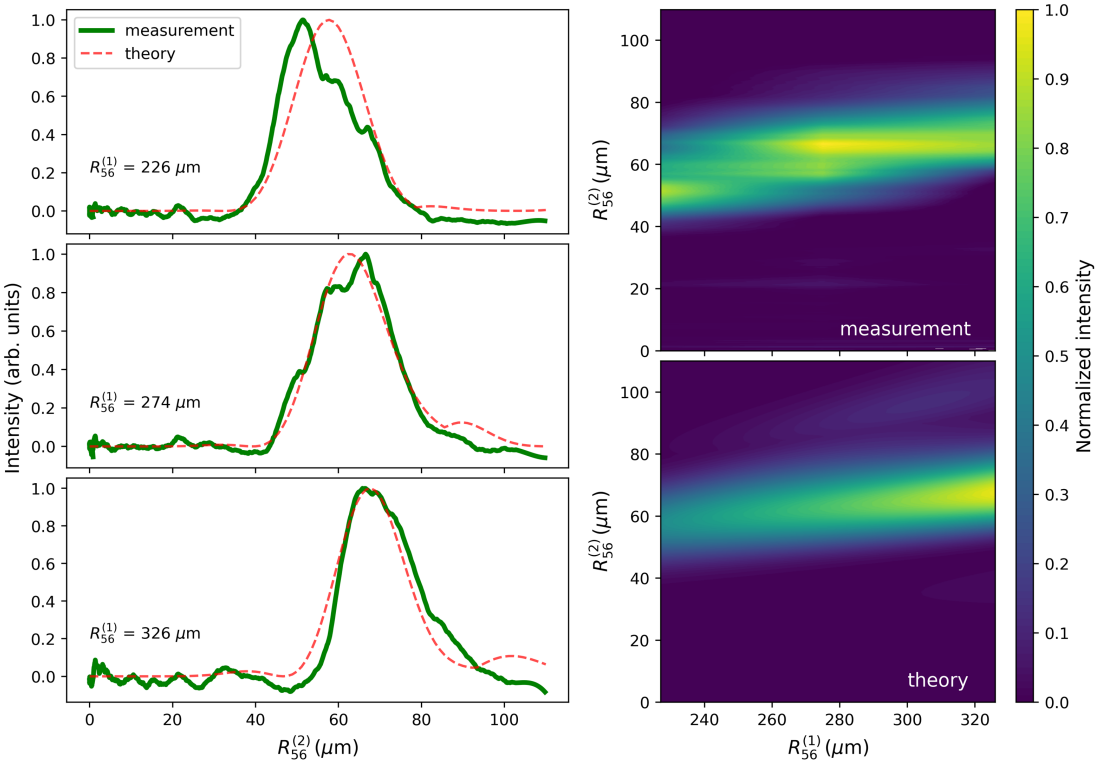

Figure 5.37: Comparison between the observed coherent signal intensity and the theoretical squared bunching factor at 114 nm. Left: Scans of the coherent signal intensity as a function of $R_{56}^{(2)}$ for three different $R_{56}^{(1)}$ values. The solid green line represents the smoothed experimental data, while the dashed red line shows the theoretical bunching factor squared. Right: 2D intensity maps constructed from experimental scans (top) and theory (bottom).

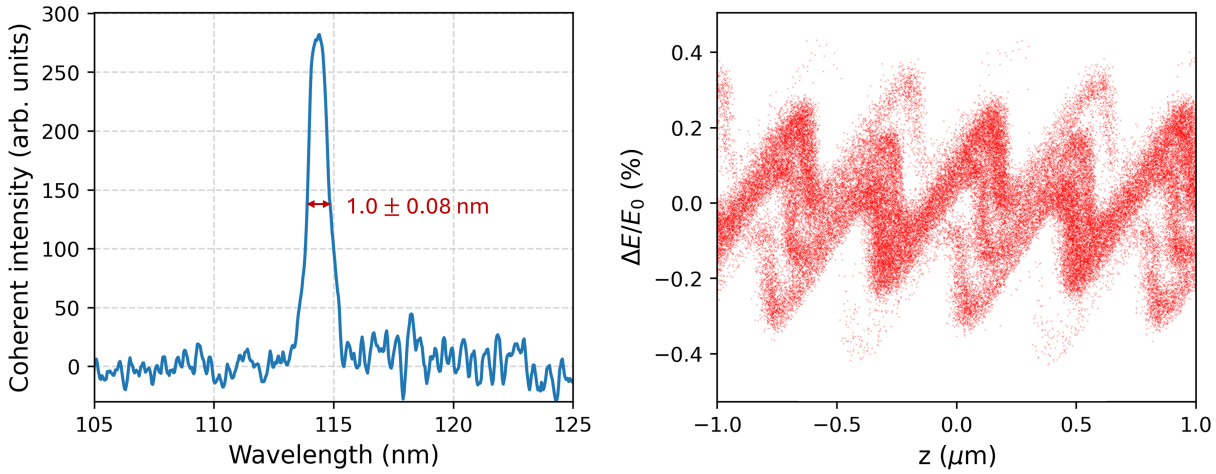

Figure 5.38: The measured spectrum of the coherent signal at 114 nm for the optimized chicane configuration ($R_{56}^{(1)} = 275\,\mu$m, $R_{56}^{(2)} = 65\,\mu$m), showing a narrow emission peak with an FWHM of 1.00 ± 0.08 nm. The figure on the right shows the corresponding simulated longitudinal phase space distribution of electrons after the second chicane for first and second energy modulation amplitudes of 0.12% and 0.14%.

Figure 5.38 presents the spectrum and the corresponding longitudinal phase space distribution of electrons after the second chicane. The spectrum shows a narrowband emission with an FWHM of 1.00 ± 0.08 nm and the phase space distribution shows substructures typical of EEHG.

5.7.3 Ninth Harmonic (89 nm)

At the 9th harmonic of 800 nm, scanning the delay between the laser pulses revealed two peaks separated by 200 fs as shown in Fig. 5.39. This feature was visible as an asymmetric shoulder in the scans of lower harmonics, but at this wavelength, it is resolved as a separate peak. This is likely due to satellite pulses trailing the main laser pulse, an artifact possibly arising from uncompensated TOD. The precise reason why the second peak is better visible at higher harmonics is not fully understood.

The EEHG signal at the 9th harmonic was found at a higher first chicane strength compared to the lower harmonics, placing it near the upper limit of the fixed parameter range for the chicane scans. Nevertheless, the chicane scans in Fig.5.40 show the same behavior observed previously, the intensity peak systematically shifting to higher $R_{56}^{(2)}$ values as $R_{56}^{(1)}$ is increased.

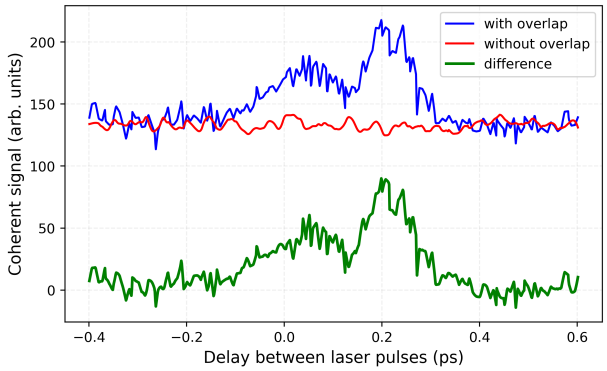

Figure 5.39: Measured intensity at 89 nm as a function of the temporal delay between the seed laser pulses. The blue curve is the signal with laser-electron overlap, and the red line indicates the background intensity measured without the overlap. The green curve shows the net EEHG signal after background subtraction.

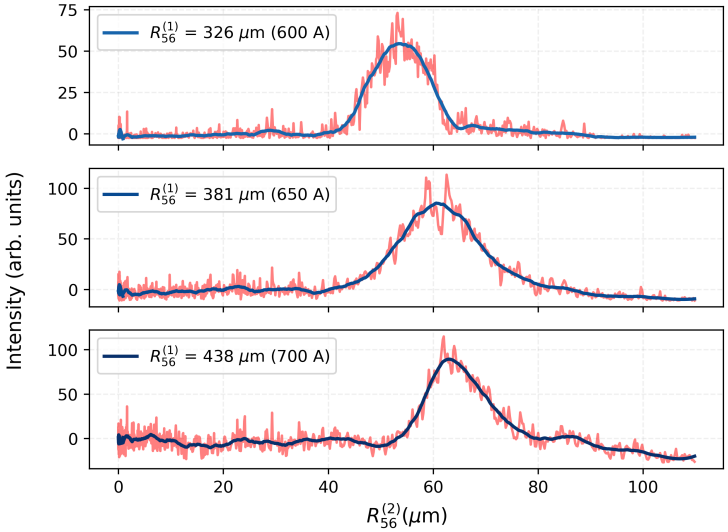

Figure 5.40: Measured 89 nm coherent signal intensity as a function of second chicane strength for different values of $R_{56}^{(1)}$. Each subplot shows the raw signal intensity (red curve) overlaid with a smoothed representation (blue).

The measurements were validated against theoretical calculations, as shown in Fig. 5.41. The calculations with $\Delta E_1/E_0 = 0.12\%$ and $\Delta E_2/E_0 = 0.14\%$ shows excellent agreement with the measured data in terms of both the peak positions and their evolution. The spectrum, shown in Fig. 5.42 (left), reveals an exceptionally narrow emission line with an FWHM of only 0.70 ± 0.08 nm.

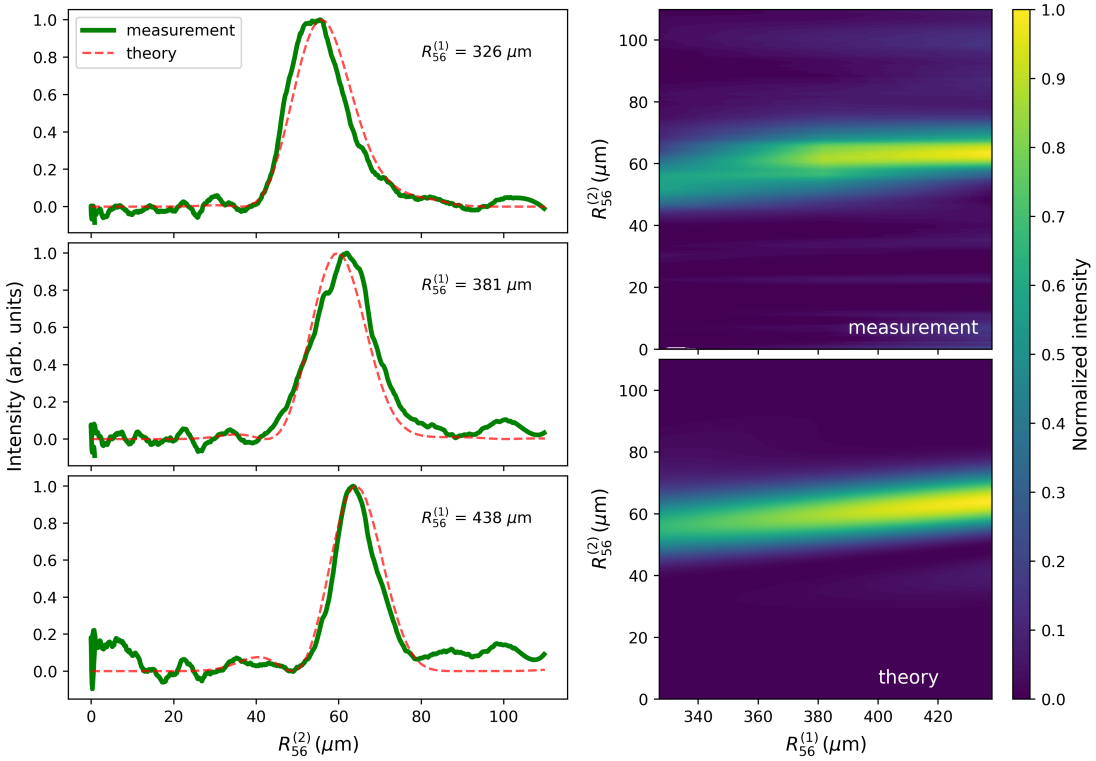

Figure 5.41: Comparison between the observed coherent signal intensity and the theoretical squared bunching factor at 89 nm. Left: Scans of the coherent signal intensity as a function of $R_{56}^{(2)}$ for three different $R_{56}^{(1)}$ values. The solid green line represents the smoothed experimental data, while the dashed red line shows the theoretical bunching factor squared. Right: 2D intensity maps constructed from experimental scans (top) and theory (bottom).

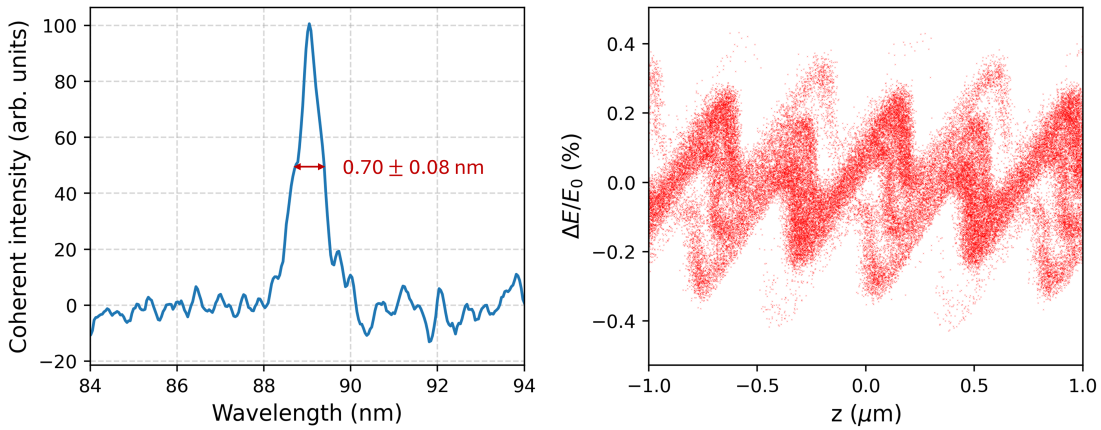

Figure 5.42: The measured spectrum of the coherent signal at 89 nm for the optimized chicane configuration ($R_{56}^{(1)} = 439\mu$m, $R_{56}^{(2)} = 60\mu$m), showing a narrow emission peak with an FWHM of 0.70 ± 0.08 nm. The figure on the right shows the corresponding simulated longitudinal phase space distribution of electrons after the second chicane for first and second energy modulation amplitudes of 0.12% and 0.14%.

5.7.4 Eleventh Harmonic (73 nm)

The 11th harmonic at 73 nm was the highest harmonic studied within the scope of this thesis. In this case, the signal was only just distinguishable from the background, resulting in a poor signal-to-noise ratio. The background-subtracted laser delay scan, shown in Fig. 5.43, again reveals an asymmetric structure associated with large TOD. However, unlike the lower harmonics, the left peak is observed to be of higher intensity than the right peak. This could be due to the high sensitivity of the bunching factor to the chicane strength at this high harmonic, which can suppress the central peak relative to the side peaks depending on the exact chicane settings.

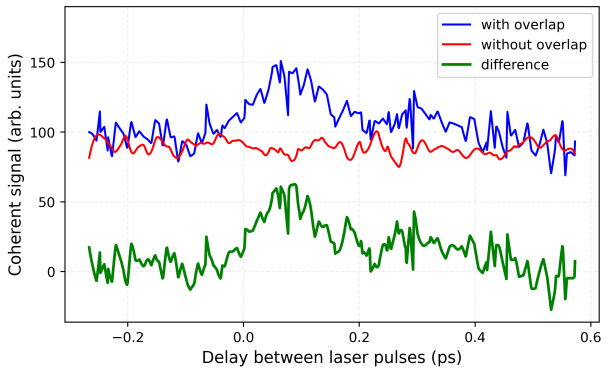

Figure 5.43: Measured intensity at 73 nm as a function of the temporal delay between the seed laser pulses. The blue curve is the signal with laser-electron overlap, and the red curve indicates the background intensity measured without the overlap. The green curve shows the net EEHG signal after background subtraction.

Scans of the second chicane (Fig. 5.44) show a noisy but distinguishable peak in all three cases. A comparison with theory was performed, as seen in Fig. 5.45. For these calculations, the energy modulation amplitudes from the 9th harmonic fit ($\Delta E_1/E_0 = 0.12\%$ and $\Delta E_2/E_0 = 0.14\%$) were used, as the high noise level in the data did not allow for a reliable automated search for the best fit. This high noise is also the likely cause of the nonzero values seen in the experimental data for $R_{56}^{(2)} < 25$ μm. Despite the noise, there is still a decent agreement between experimental data and theory. Finally, the spectrum was analyzed, which is shown in Fig. 5.46 (left). The signal is narrowband with an FWHM of 0.78 ± 0.08 nm with the complex double modulation features of EEHG visible in the corresponding longitudinal phase space distribution of electrons (Fig. 5.46 (right)).

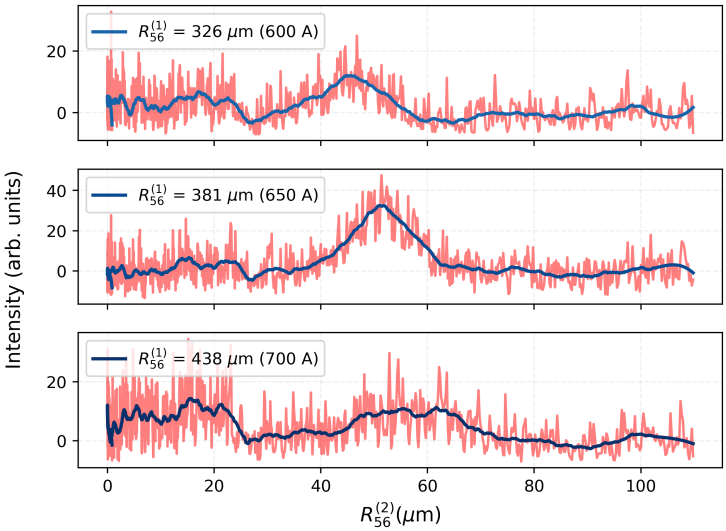

Figure 5.44: Measured 73 nm coherent signal intensity as a function of second chicane strength. Each subplot shows the raw signal intensity (red curve) overlaid with a smoothed representation (blue).

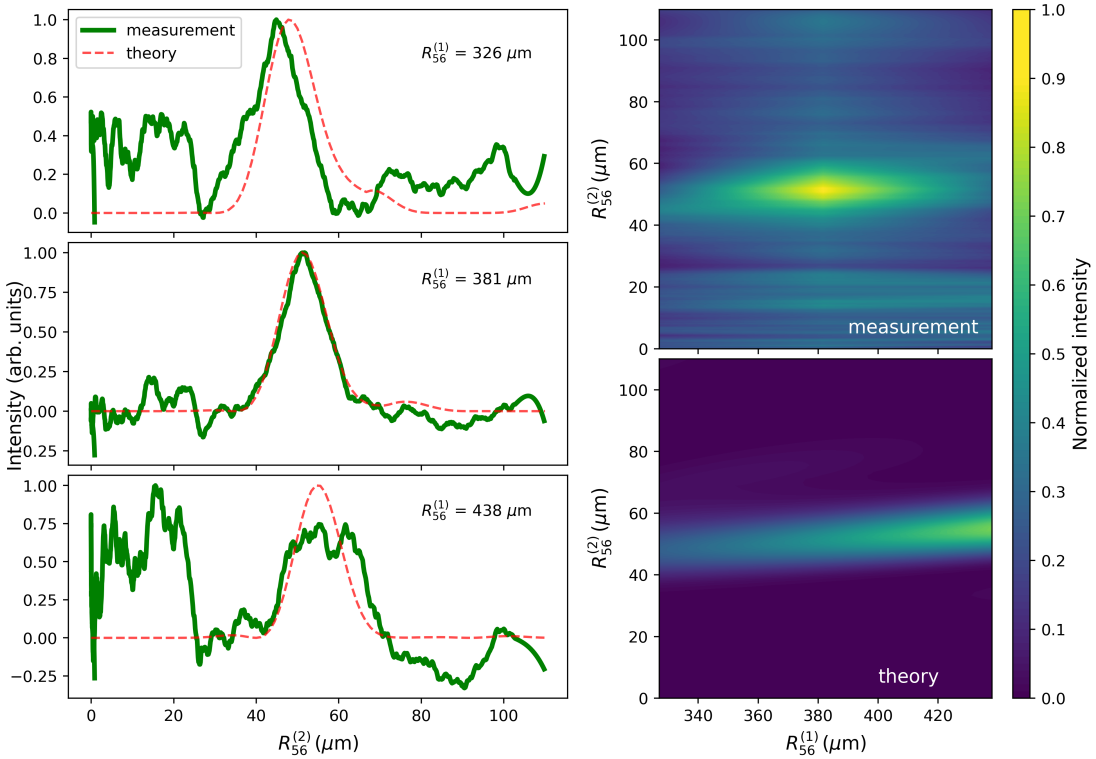

Figure 5.45: Comparison between the observed coherent signal intensity and the theoretical squared bunching factor at 73 nm. Left: Scans of the coherent signal intensity as a function of $R_{56}^{(2)}$ for three different $R_{56}^{(1)}$ values. The solid green line represents the smoothed experimental data, while the dashed red line shows the theoretical bunching factor squared. Right: 2D intensity maps constructed from experimental scans (top) and theory (bottom).

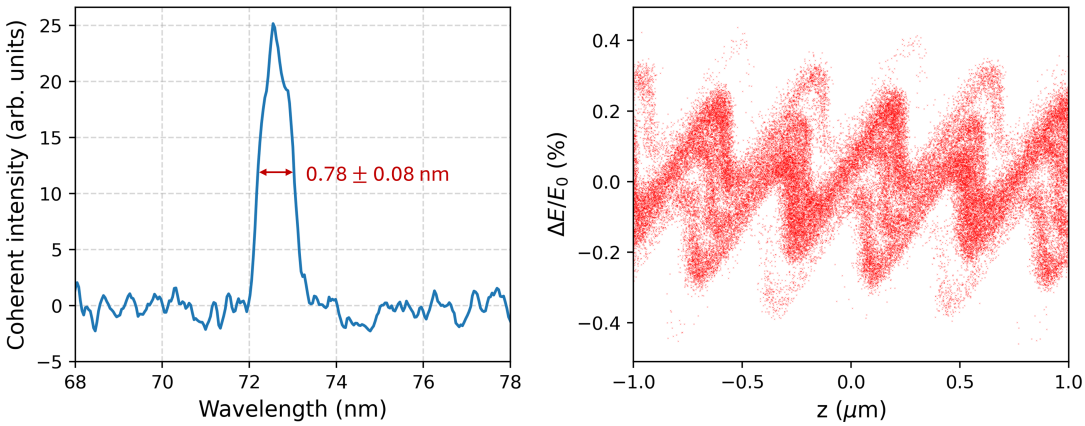

Figure 5.46: Measured spectrum of the coherent signal at 73 nm for the optimized chicane configuration ($R_{56}^{(1)} = 382$ μm, $R_{56}^{(2)} = 53$ μm), showing a narrow emission peak with an FWHM of 0.78 ± 0.08 nm. The figure on the right shows the corresponding simulated longitudinal phase space distribution of electrons after the second chicane for first and second energy modulation amplitudes of 0.12% and 0.14%.

5.7.5 Influence of Beam Current

A fundamental test of the coherent nature of the signal is its intensity scaling with beam current. The intensity of coherent emission is expected to scale quadratically with the number of contributing electrons ($I_{\text{coh}} \propto b_h^2 N_e^2$) whereas incoherent radiation scales linearly ($I_{\text{incoh}} \propto N_e$). To verify this relationship, the EEHG signal intensity at 160 nm was recorded while the beam current was gradually reduced from 10 mA to 0 mA over a span of around 45 minutes using a scraper. The coherent signal was sampled from the narrow spectral emission line, while the incoherent background was sampled from a different, broader region of the spectrum. Figure 5.47 shows the results of this measurement. The coherent signal intensity shows a good fit to a quadratic function, while the incoherent signal follows a clear linear trend. This quadratic dependence confirms the coherent origin of the signal.

To validate the simulation parameters used throughout this chapter, a global error analysis was performed. The measured intensities were compared against the theoretical bunching factor squared across a range of energy modulation amplitudes. To facilitate a direct comparison, both the experimental and simulated curves were normalized to their respective maximum values. Figure 5.48 presents the root-mean-square (RMS) error, computed across all R_{56} scans for the 5th, 7th, and 9th harmonics. The 11th harmonic was excluded from this calculation, as its high noise level would skew the comparison. This global analysis assumes that the energy modulation amplitudes remained consistent across the different measurements, neglecting the small drift that could have occurred between shifts. Under this assumption, the optimal fit corresponds to energy modulation amplitudes of $\Delta E_1/E_0 = 0.12\%$

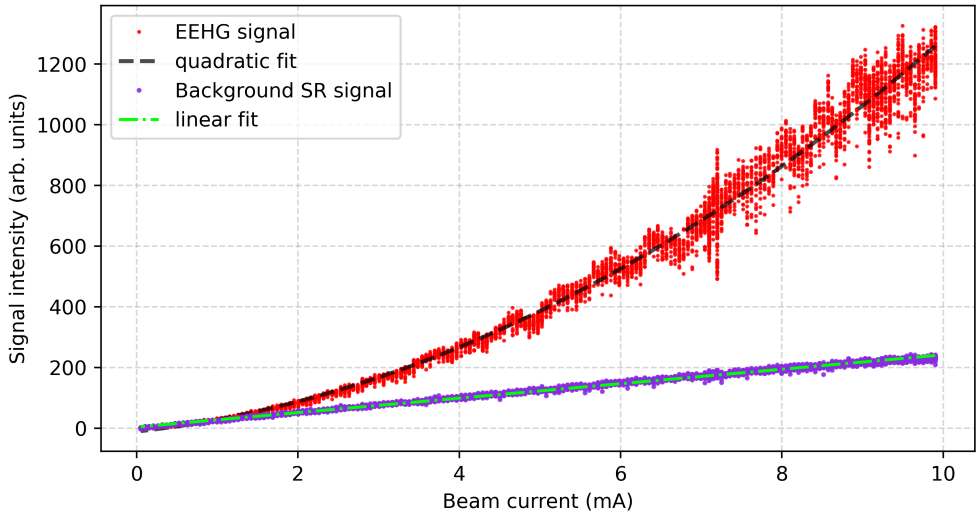

Figure 5.47: Coherent and incoherent signal intensity scaling with beam current. The intensity of the 160 nm signal (red dots), sampled from a narrow spectral line, follows a quadratic fit (dashed black line), confirming the coherent emission. The incoherent background (purple dots), sampled from a broad spectral region, scales linearly (green dash-dotted line) as expected for incoherent emission.

and $\Delta E_2/E_0 = 0.14\%$. This confirms that the parameters used in the individual sections represent the global minimum.

The EEHG implementation resulted in successful observations of the coherently emitted signal at odd harmonics of 800 nm up to the 11th harmonic. The signal-to-background ratio (SBR), which is the ratio between the coherently emitted signal and the incoherent synchrotron radiation background, are given at a bunch current of 10 mA by scaling the observed values with the coherent-to-incoherent ratio estimated at different currents from Fig. 5.47.

The key observations and parameters for each harmonic are summarized in Table 5.3. The total uncertainty in the FWHM was calculated by combining the statistical error, derived from the standard error of the mean (SEM) of the intensity profile, and the spectrometer resolution. While the statistical error was calculated to be small (< 0.04 nm), the spectrometer resolution (~ 0.07 nm) is the dominant contribution. The combined uncertainty is estimated to be ± 0.08 nm for all harmonics. For the SBR, the uncertainty was estimated by propagating the SEM of the signal and background spectral regions. Furthermore, since the SBR was normalized to a reference beam current of 10 mA, the uncertainty associated with this scaling factor, derived from the covariance matrices of the polynomial fits from Fig. 5.47, was also included.

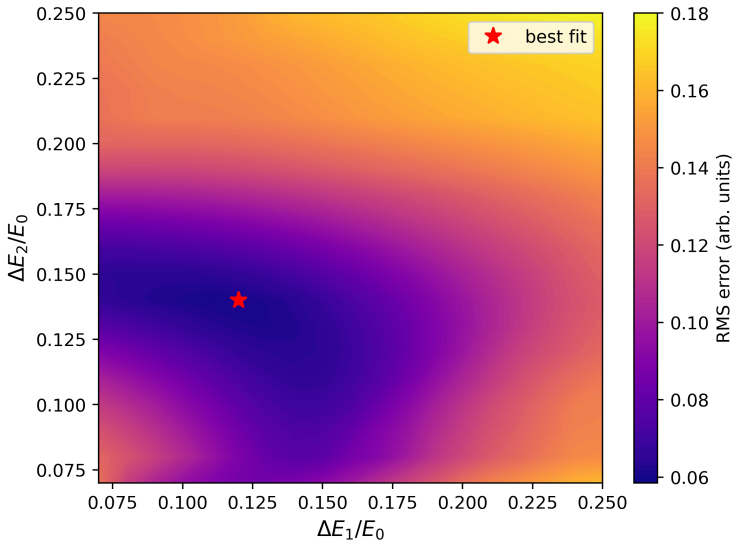

Figure 5.48: Global optimization of energy modulation amplitudes. The colormap shows the root-mean-square (RMS) error between the measured coherent intensity and the simulated bunching factor squared for harmonics 5, 7, and 9. Both datasets were normalized to their peak values before comparing across all R_{56} scans. The minimum error (red star) confirms the optimal parameters used for the analysis were $\Delta E_1/E_0 = 0.12\%$ and $\Delta E_2/E_0 = 0.14\%$.

h	λ (nm)	$\Delta E_1/E_0$, $\Delta E_2/E_0$ (%)	Optimum $R_{56}^{(1)}$, $R_{56}^{(2)}$ (μm)	FWHM bandwidth (nm)	SBR at 10 mA
5	160	0.11, 0.14	140, 72	1.27 ± 0.08	5.34 ± 0.04
7	114	0.12, 0.14	275, 65	1.00 ± 0.08	2.36 ± 0.02
9	89	0.12, 0.14	439, 60	0.70 ± 0.08	1.52 ± 0.01
11	73	0.12, 0.14	382, 53	0.78 ± 0.08	1.42 ± 0.01

Table 5.3: Summary of key parameters and results for the observed EEHG harmonics h. The estimated energy modulation amplitudes, chicane strengths that maximize the intensity, measured spectral bandwidth (FWHM), and the estimated signal-to-background ratio (SBR) at 10 mA are listed for each harmonic.

Further analysis of the experimental data shows significant deviation from the ideal EEHG theory. For an ideal Gaussian seed laser pulse, the SBR and relative spectral bandwidth ($\Delta\lambda/\lambda$) of the EEHG radiation are expected to scale proportionally to $m^{-2/3}$ [158]. Here, the harmonic is set by $h = Km + n$ where $K = 2$ and $n = -1$ as described in Section 2.6.3. Figure 5.49 shows the measured and simulated relative bandwidth and SBR plotted for different harmonics compared to the theoretical scaling laws normalized to the $h = 5$ measurement. It shows that the measured values deviate significantly from the theoretical estimates. In the case of SBR, the faster fall-off of the measurements is reproduced by the simulations. This agreement indicates that the SBR degradation is driven by physical beam parameters

modeled in the simulation, such as finite beam emittance and residual R_{52}, which smear the microbunching structure more severely at shorter wavelengths. However, the bandwidth measurements show an increase at higher harmonics that is absent in both theory and simulations. This discrepancy suggests that the broadening is not a result of beam dynamics but could be due to the wavelength-dependent focusing properties of the spectrometer, which can cause defocusing at the detector plane that artificially broadens the measured spectrum.

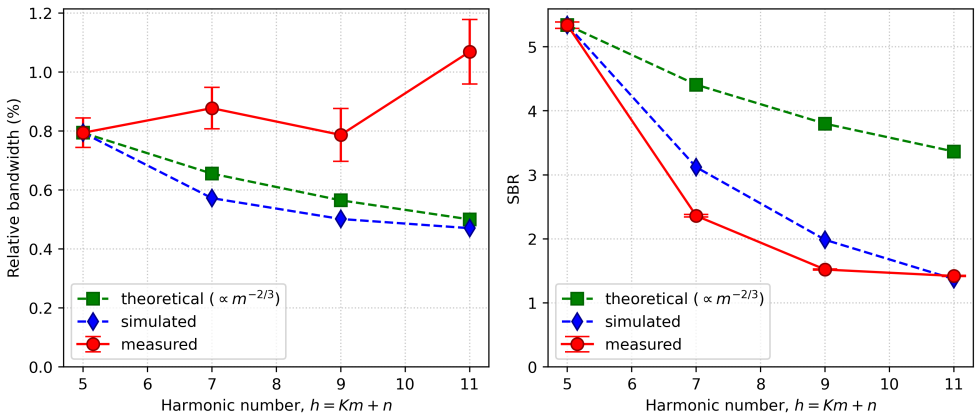

Figure 5.49: Comparison of measured (red), simulated (blue), and theoretical (green) scaling for the relative spectral bandwidth (left) and SBR (right). The theoretical values, representing ideal scaling laws ($m^{-2/3}$ for bandwidth and SBR), are normalized to the $h = 5$ measurement.

5.8 Summary

This chapter has detailed the simulation, implementation, and experimental validation of the EEHG scheme at DELTA, realized through the novel SPEED project, which is the first ever successful implementation of EEHG in an electron storage ring. The existing U250 undulator was reconfigured into a compact 4.75-meter arrangement, integrating all necessary EEHG components (two modulators, two chicanes, and a radiator) within a single device for a proof-of-principle setup.

Significant technical challenges inherent to this compact design, such as magnetic field saturation, particle trajectory shifts, and nonzero R_{52} matrix elements, were identified. Extensive simulation studies were carried out to optimize key parameters, including the laser wavelengths, waist sizes, and chicane strengths. A dedicated correction scheme for the detrimental R_{52} and saturation effects was also developed and simulated.

The experiments successfully demonstrated EEHG, yielding a narrowband coherent signal at odd harmonics of the 800 nm laser. The signal was validated up to the 11th harmonic (73 nm). Reaching this high harmonic is a distinguishing feature of the EEHG scheme. A comparable CHG setup, even with higher laser power, would be limited to lower harmonics. The origin of the signal was unambiguously confirmed as EEHG through three key observations. First, the signal dependence on the relative delay between the two seed pulses and the laser-electron overlap confirmed the laser-driven nature of the emission. Second, the measured coherent signal as a function of the two chicane strengths $R_{56}^{(1)}$ and $R_{56}^{(2)}$ revealed the distinctive peak structures predicted by EEHG theory. Finally, the signal intensity showed a clear quadratic dependence on the electron beam current, a signature of coherent emission.

The most significant outcome of this work is that a distinguishable, high-harmonic signal was achieved despite remarkably weak energy modulation amplitudes, which were calculated to be just $\Delta E_1/E_0 = 0.12\%$ and $\Delta E_2/E_0 = 0.14\%$ compared to a natural energy spread of 0.07%. This limitation was a direct consequence of the physical constraint of using modulators with only four undulator periods each as well as the limited laser power. Achieving a clear signal up to the 11th harmonic under these conditions exceeded the expectations for such a constrained setup.

The proof-of-principle experiment demonstrated that the EEHG process is robust, yielding results even with a weak energy modulation. This provides strong motivation for an upgraded system, as it is readily conceivable that dedicated and longer modulators, stronger chicanes, and more powerful laser pulses would achieve significantly higher harmonics with higher intensity and SBR. The next chapter explores the design optimization and simulation of such an implementation, using permanent magnet undulators scaled to fit within the same straight section with a length below 5 m.

Chapter 6

Conceptual Study for a Permanent-Magnet-Based EEHG Setup

In the context of the EEHG implementation within the SPEED project, several physical and technical constraints significantly limited the operational flexibility of the setup. One major limitation was the need to accommodate all five components (2 modulators, 2 chicanes, and a radiator) of the EEHG scheme within a single undulator (U250) comprising only 19 periods. The relatively large period length, 250 mm, of the undulator restricts the number of effective periods available for each component. To make full use of the limited undulator length, the modulators and chicanes had to be intertwined. However, this compromise introduced several undesired and detrimental effects, such as field-dependent horizontal beam displacement and nonzero R_{52} values of the chicanes.

Other proposals for EEHG in a storage ring consider modulators and radiator in different straight sections, with a part or the whole storage ring providing the large dispersion after the first modulation instead of a dedicated chicane. But these proposals have not been attempted in any storage ring yet.

In this thesis, an alternative permanent-magnet-based EEHG setup designed to fit within a single nominal straight section is proposed. This approach allows for shorter period lengths and a higher number of undulator periods for both the modulators and radiator, while allowing the chicanes to be spatially separated and independently operated. This chapter presents theoretical and simulation studies exploring the feasibility and performance of such a permanent-magnet-based EEHG setup. While the work is discussed here as a potential replacement for the current SPEED configuration, the design is broadly applicable as a blueprint for an EEHG-based user facility at any storage ring with a typical straight section of around 5 meters. Compared to established techniques like femtoslicing [159], this approach

promises a dramatic increase in flux, potentially delivering up to 10^5 more photons per pulse with the same laser-based repetition rate. However, this gain in intensity is balanced by operation within a more limited wavelength range. The following studies explore the performance of such a configuration to demonstrate its potential.

6.1 Design and Parameter Constraints

The objective of this study is to develop a model for a permanent-magnet-based EEHG setup that fits within the 4.875 m long straight section as an alternative to the U250 undulator. For simplicity, the same seed laser wavelengths used in the current SPEED setup are assumed: 800 nm for the first modulator and 400 nm for the second. These wavelengths were specifically chosen to enable the use of a Ti:sapphire laser system, which offers the optimal combination of high pulse energy and short pulse duration. This represents the best available choice, as alternative laser systems such as Nd:YAG (1064 nm) and fiber lasers (1030-2200 nm) typically operate at longer wavelengths, which would require longer undulator period lengths for the modulators and would also result in longer output wavelengths.

This study focuses on a conventional undulator design, as alternatives like in-vacuum undulators were excluded due to their larger spatial requirements and higher costs. A critical parameter for the design of the permanent-magnet undulators (PMUs) is the minimum magnetic pole gap. This gap must be sufficiently large to accommodate the vacuum chamber and provide adequate clearance for the electron beam to ensure that the new setup does not become the lifetime-limiting aperture of the storage ring. A suitable value is determined by calculating the required aperture through a scaling from the superconducting wiggler (SCW), which is the most restrictive element currently in the DELTA storage ring.

The SCW has an inner vertical aperture of $A_{\mathrm{SCW}} = 10$ mm at a location with a vertical beta function $\beta_{y,\mathrm{SCW}}$ of approximately 6 m. The U250 undulator section has a maximum vertical beta function of $\beta_{y,\mathrm{U250}} \approx 15$ m. To maintain an equivalent effective aperture relative to the local beam size, the required inner aperture for the new setup A_0 is scaled as

$$A_0 = A_{\mathrm{SCW}} \sqrt{\frac{\beta_{y,\ \mathrm{U250}}}{\beta_{y,\ \mathrm{SCW}}}} \approx 15.8 \text{ mm}. \tag{6.1}$$

Assuming a 1 mm chamber wall thickness on the top and bottom and a safety clearance of 1 mm above and below, the necessary magnetic gap is 19.8 mm. Thus, a value of 20 mm is chosen for the magnetic pole gap for all undulators in this study. The following calculations were done for a beam energy of $E = 1.5$ GeV. A lower beam energy would further relax the undulator period length and R_{56} requirements.

6.1.1 Undulator Period Length

Figure 6.1 shows the required magnetic field for achieving resonance at 800 nm and 400 nm for different period lengths, alongside the maximum magnetic field achievable for a hybrid planar-PMU at those period lengths. The required magnetic field is calculated from the corresponding K value for the 800 nm and 400 nm wavelengths using

$$B_{\mathrm{req}}[T] = \frac{K}{93.4\lambda_{\mathrm{u}}[m]}, \tag{6.2}$$

where K is given by

$$K = \sqrt{4\gamma^2 \frac{\lambda_0}{\lambda_{\mathrm{u}}} - 2}. \tag{6.3}$$

Here, γ is the Lorentz factor of the electrons, λ_0 is the fundamental wavelength of the undulator, and λ_{u} is the undulator period length.

The following calculations are based on hybrid PMUs [160] in which strong permanent magnets are placed between soft iron poles. This arrangement uses the magnets to generate the magnetic field and the iron poles to guide and concentrate it into the electron beam path. The achievable magnetic field for a hybrid PMU is calculated using a formula from [161]

$$B_{\mathrm{PMU}}[T] = 2.435 \exp\left[-3.503\frac{g}{\lambda_{\mathrm{u}}} - 0.05\left(\frac{g}{\lambda_{\mathrm{u}}}\right)^2\right], \tag{6.4}$$

where g represents the pole gap, taken as 20 mm in this analysis.

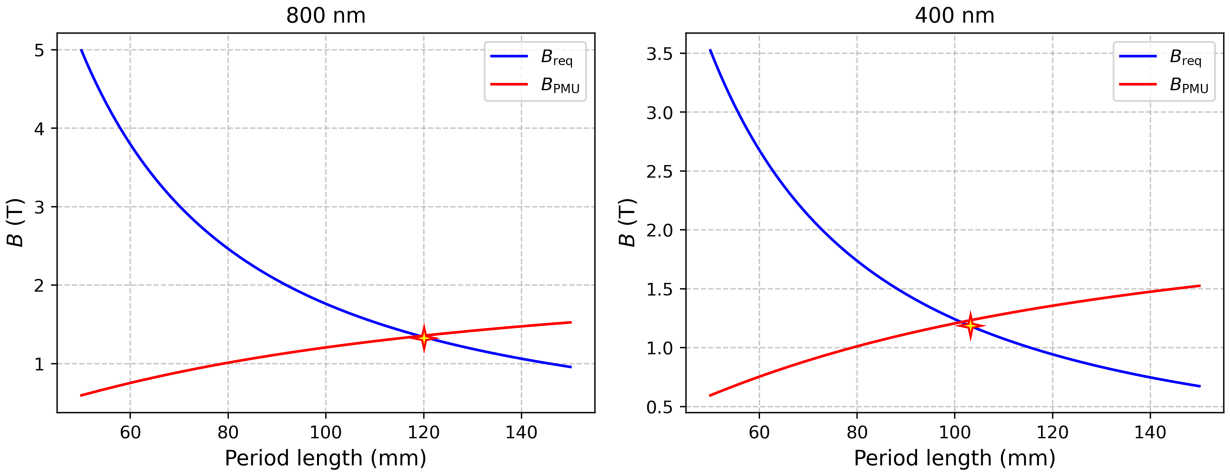

Figure 6.1: The required (blue) and achievable (red) magnetic fields for a hybrid PMU with a 20 mm pole gap for different period lengths of an undulator tuned to 800 nm (left) and 400 nm (right). The red stars indicate the chosen working point for each modulator.

Figure 6.1 reveals that for the first modulator tuned to 800 nm, period lengths shorter than

119 mm require magnetic fields exceeding what is achievable, while this crossing point occurs at approximately 101 mm for 400 nm. The period lengths were chosen to be close to this crossing point so that the available space is utilized to the maximum. Period lengths of 120 mm and 102 mm were chosen for the first and second modulators, respectively.

For the radiator, the 10th harmonic (80 nm) and beyond were targeted. This requirement means the radiator undulator must generate a sufficient magnetic field to tune the undulator to 80 nm or shorter wavelengths. Figure 6.2 compares the required magnetic field for tuning a PMU to 80 nm with the achievable magnetic field at different period lengths. Based on this comparison, a period length of 72 mm was chosen for the radiator. Higher harmonics would require weaker peak magnetic fields, which remain within the achievable range.

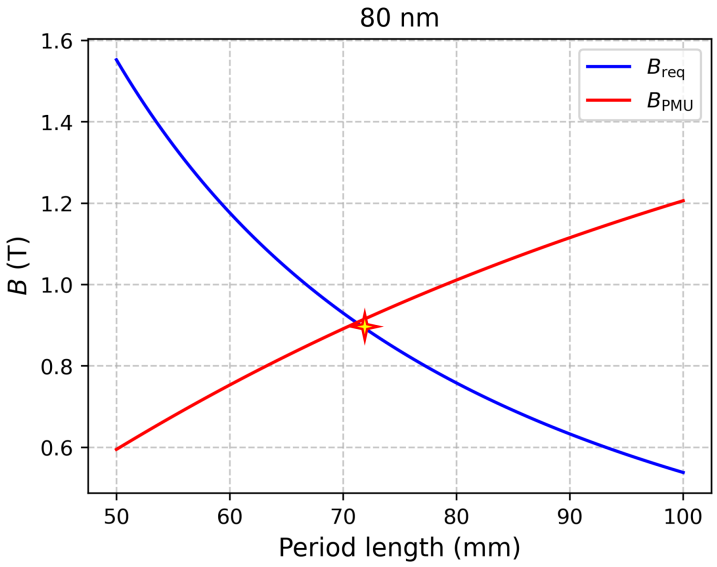

Figure 6.2: The required (blue) and achievable (red) magnetic fields for a hybrid PMU with a 20 mm pole gap for different period lengths of an undulator tuned to 80 nm. The red star indicates the working point chosen for the radiator.

6.1.2 Harmonic Range

The harmonic range was defined by the practical performance limits of the proposed setup. The lower bound of the 10th harmonic is determined by the tuning range of the radiator, while the upper bound was established by estimating the point at which the coherent signal can no longer be clearly distinguished from the incoherent background radiation. A signal-to-background ratio (SBR) greater than 10 was set as the threshold for a usable signal.

The SBR was estimated through a simplified calculation. The incoherent power, which forms the noise background, is proportional to the total number of electrons N_e in the bunch. For this calculation, a total bunch length of 80 ps and a single-bunch current of 10 mA were assumed, consistent with typical operating parameters at the DELTA facility.

The coherently emitted power P_c is generated only by the fraction of electrons within the 100 fs longitudinal window modulated by the laser pulse. Because this laser pulse has a Gaussian temporal profile, the strength of the modulation and thus the resulting bunching factor varies along this window. The bunching factor within the temporal window was modeled with a Gaussian profile, with its peak value for a given harmonic estimated using Eq. 2.63. To accurately calculate the total coherently emitted power, the modulated window is computationally divided into a series of smaller longitudinal slices. The total power is then the sum of the power coherently emitted from each individual slice. This relationship is expressed as

$$P_c \propto \sum_{i=1}^{n} N_i^2 b_{h,i}^2. \tag{6.5}$$

Here, the index i denotes an individual slice, where N_i and $b_{h,i}$ represent the number of electrons and bunching factor calculated using Eq. (2.55) for harmonic h within that slice, respectively, and n denotes the total number of slices.

The SBR between coherently and incoherently emitted radiation is then given by

$$\mathrm{SBR} = \frac{\sum_{i=1}^{n} N_i^2 b_{h,i}^2}{N_e}. \tag{6.6}$$

Figure 6.3 shows the estimated SBR based on the theoretical maximum bunching factor for different harmonics of 800 nm. It shows that the SBR is expected to fall below the threshold of 10 at approximately the 29th harmonic, which was therefore chosen as the upper limit for the optimization study. Since the modulation wavelength used in the second modulator is half of the first modulation wavelength, the bunching factor is optimized for odd harmonics (see Section 2.6.3) between 11 and 29.

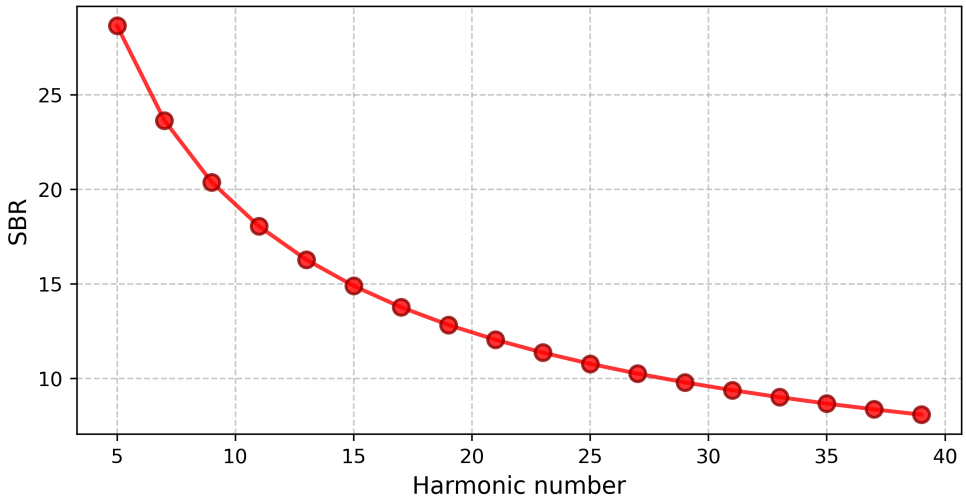

Figure 6.3: Estimated SBR based on the theoretical maximum bunching factor for different harmonics of 800 nm.

6.2 Optimization of the EEHG Configuration

6.2.1 Optimization Methodology

Choosing the optimum parameters for the two lasers and modulators is a complex task involving several interdependent variables and competing trade-offs. The laser waist size must be large enough to ensure a sufficiently uniform energy modulation across the transverse extent of the electron beam. However, with fixed pulse energy and duration, a larger spot size reduces the peak electric field strength, leading to a smaller energy modulation amplitude. As a result, the final bunching factor may be diminished. It is recommended that the laser beam size σ_r be at least 3 times larger than the electron beam size σ_e [13] to minimize the non-uniformity in energy modulation. Nevertheless, in practical implementations, this guideline must be balanced against the reduction in modulation strength with increasing spot size.

In an EEHG setup, optimizing the laser beam sizes cannot be decoupled from the rest of the parameters. The peak energy modulation amplitude depends on both the laser waist size and the number of periods in the modulator. Meanwhile, the conversion of energy modulation into density modulation, and hence the final bunching factor, depends critically on the chicane strengths. If the modulators are too long, they limit the available space for the radiator and chicanes, which reduces the maximum achievable dispersion. Conversely, if the modulators are too short, the energy modulation becomes too weak, resulting in suboptimal bunching. Therefore, an optimum must be found that balances both energy modulation amplitude and chicane strength within the constraints of the total available straight section length while leaving enough space for a sufficiently strong radiator.

To define the optimization space, the radiator length needs to be fixed first, as the radiator is not involved in the bunching process. As a starting point, a radiator with 15 periods including endpoles was selected. With a period length of 72 mm, the radiator has a total length of 1080 mm, leaving a maximum space of 3795 mm for the two modulators and chicanes.

The optimization of microbunching was carried out by systematically varying the number of periods in the two modulators (N_{M1} and N_{M2}) and the transverse laser beam sizes (σ_{L1} and σ_{L2}). The goal was to find the conditions producing the maximum bunching factor for odd harmonics in the range 11 to 29. To make this large parameter scan computationally efficient, a simplified 1D-model was utilized here. The energy modulation amplitude ΔE imparted to an electron by the laser field is modeled as a Gaussian transverse envelope combined with a

sinusoidal longitudinal modulation as

$$\Delta E(x, y, z) = \Delta E_0 \exp\left(-\frac{x^2 + y^2}{4\sigma_{\mathrm{L}}^2}\right) \sin\left(\frac{2\pi z}{\lambda}\right), \tag{6.7}$$

where ΔE_0 is the maximum on-axis modulation amplitude, λ is the laser wavelength, and σ_{L} is the transverse rms size of the laser. The use of a single transverse size parameter assumes a circular Gaussian beam profile ($\sigma_x = \sigma_y = \sigma_{\mathrm{L}}$). The variables x and y are the transverse coordinates and z is the longitudinal coordinate, all defined within the co-propagating frame of the bunch. The value ΔE_0 for each combination of modulator length and laser size was pre-calculated using the detailed simulation code referenced in Section 5.4.

With the modulation amplitudes known, the optimal R_{56} values for the two chicanes are then estimated using a semi-analytical model. The optimal $R_{56}^{(1)}$ and $R_{56}^{(2)}$ were determined based on the modulation amplitudes $A_1 = \Delta E_0^{(1)}/\sigma_E$ and $A_2 = \Delta E_0^{(2)}/\sigma_E$, where $\Delta E_0^{(1)}$ and $\Delta E_0^{(2)}$ denote the first and second energy modulation amplitudes and σ_E is the relative energy spread of the electron beam. From Eq. (2.65), the optimum $R_{56}^{(2)}$ is calculated as

$$R_{56}^{(2)} = \frac{B_2}{k\sigma_E}. \tag{6.8}$$

The optimal $R_{56}^{(1)}$ was scanned numerically within a feasible range and selected to maximize the analytical expression for the bunching factor in Eq. 2.62.

Parameter Scan Process

For a fixed pair of laser pulse energies and waist sizes, the scan proceeded as follows:

1. For each combination of number of modulator periods, the modulator lengths L_{M1} and L_{M2} and corresponding energy modulation amplitudes were calculated. From these, the optimal chicane strengths $R_{56,\mathrm{opt}}^{(1)}$ and $R_{56,\mathrm{opt}}^{(2)}$ were determined. To account for simulation discrepancies and ensure operational flexibility, the design maximum for the second chicane was set to $R_{56,\mathrm{max}}^{(2)} = R_{56,\mathrm{opt}}^{(2)} + 20\,\mu\mathrm{m}$.

2. To allow for better operational flexibility, higher fields, and lower costs, an electromagnetic design is considered for the chicanes. The design assumes a three-electromagnet chicane, as sketched in Fig. 6.4, with a 30 mm coil thickness resulting in a 60 mm magnet separation. Given the separation between magnets d and magnet length L, the chicane strength is approximately given by [2]

$$R_{56} = 4r(\tan(\alpha) - \alpha) + 2d\frac{\tan^2(\alpha)}{\cos(\alpha)}, \tag{6.9}$$

where $r = E/ecB$ represents the bending radius of one dipole magnet for a given magnetic field B, and $\alpha = \arcsin(L/r)$ is the bending angle. Based on *CST* simulations [145] for a similar magnet as depicted in Fig. 5.2 with a 20 mm pole gap, a peak field of 1.3 T was chosen.

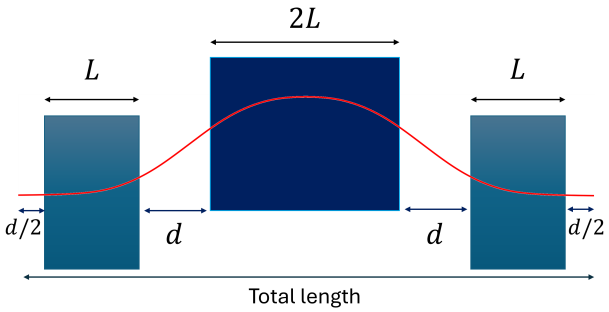

Figure 6.4: Sketch of a three-dipole-magnet chicane. In this case, the two central dipole magnets are combined to form a magnet that is twice the length of the other two. The red curve shows the electron orbit.

3. Knowing the maximum second chicane strength $R^{(2)}_{56,\max}$, Eq. 6.9 was solved numerically to find the required magnet length L. The total length of the second chicane was then calculated as

$$L_{C2} = 4L + 3d + 20\,\text{mm}. \tag{6.10}$$

this includes an additional 10 mm separation between the chicane and the adjacent undulators.

4. The space left for the first chicane is then

$$L_{C1} = 3795\,\text{mm} - (L_{M1} + L_{M2} + L_{C2}). \tag{6.11}$$

The maximum possible strength for the first chicane was then calculated based on the available length L_{C1} using Eqs. (6.10) and (6.9).

5. Finally, with the maximum possible chicane strengths known, the bunching factor was calculated for a grid of $R^{(1)}_{56}$ and $R^{(2)}_{56}$ values within their feasible ranges. By searching over all undulator and chicane configurations, the maximum bunching factor for the initial pair of laser waist sizes was identified.

6.2.2 Optimization of Laser Waist Size

The initial step in the parameter optimization involved determining the ideal rms beam sizes for the two seed lasers. The optimization was performed in two stages. First, a broad and computationally efficient scan to find the optimal laser waists, followed by a detailed

scan to pinpoint the optimal modulator periods. First, the simplified 1D-model was used to efficiently explore the large 4-dimensional parameter space of the two laser waists σ_{L1}, σ_{L2} and the two number of periods $N_{\mathrm{M1}}, N_{\mathrm{M2}}$ for the modulators.

The simulations assume a commercially available high-power laser system capable of generating 100 fs pulses with energies exceeding 14 mJ [162, 163]. Based on a second-harmonic generation efficiency of 31% and a residual fundamental power of 35% as observed with the current system at DELTA, the pulse energies for the first and second modulators were set to 5.0 mJ and 4.4 mJ, respectively.

Figure 6.5 displays the resulting maximum bunching factor at different rms laser beam sizes for four distinct harmonics. For each pair of laser waist sizes, the number of periods for both the first and second modulators was scanned in a grid from 6 to 15, and only the peak bunching factor from this grid was recorded for the heatmap shown in Fig. 6.5. Since the goal is to find a configuration that performs well across the entire harmonic range of interest rather than for a single harmonic, the bunching factor was first normalized to the theoretical maximum bunching factor and averaged over odd harmonics from 11 to 27.

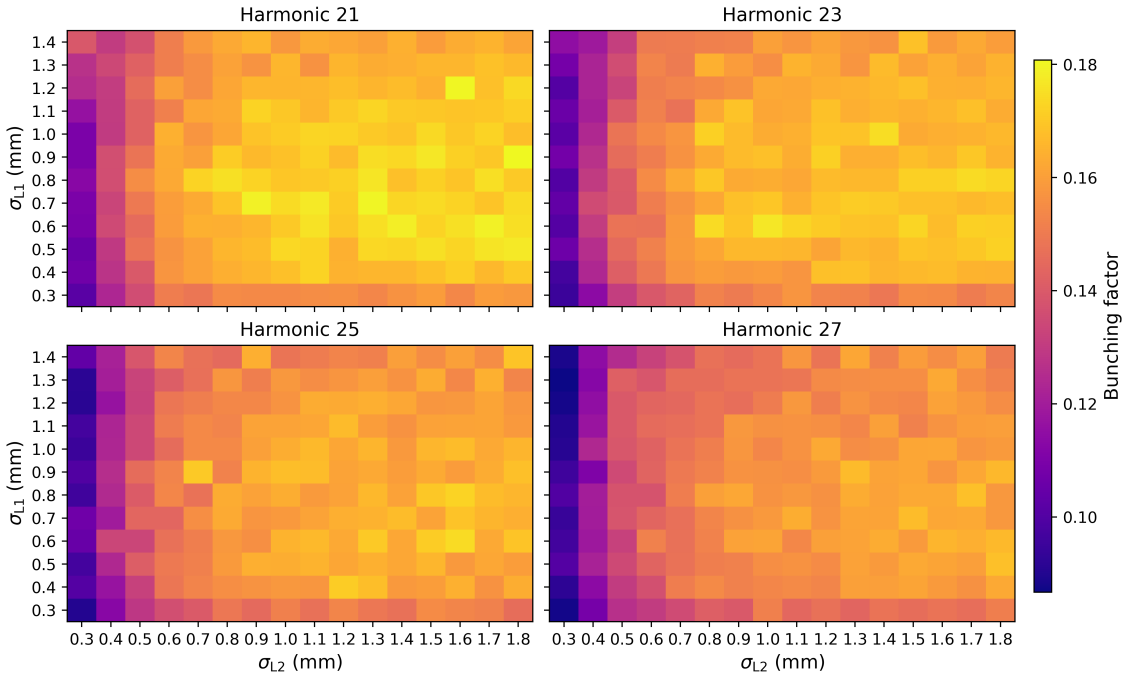

Figure 6.5: Maximum bunching factor for different rms beam sizes σ_{L1} and σ_{L2} of the first and second seed laser, respectively, shown for four different harmonics. Each pixel in the heatmaps represents the peak bunching factor found after performing a full optimization scan over the number of modulator periods for that specific pair of laser sizes.

The averaged result, plotted in Fig. 6.6, reveals distinct dependencies for each laser beam size. For the first laser, a small waist σ_{L1} introduces excessive non-uniformity in the energy modulation, whereas an overly large waist reduces the modulation amplitude to sub-optimal

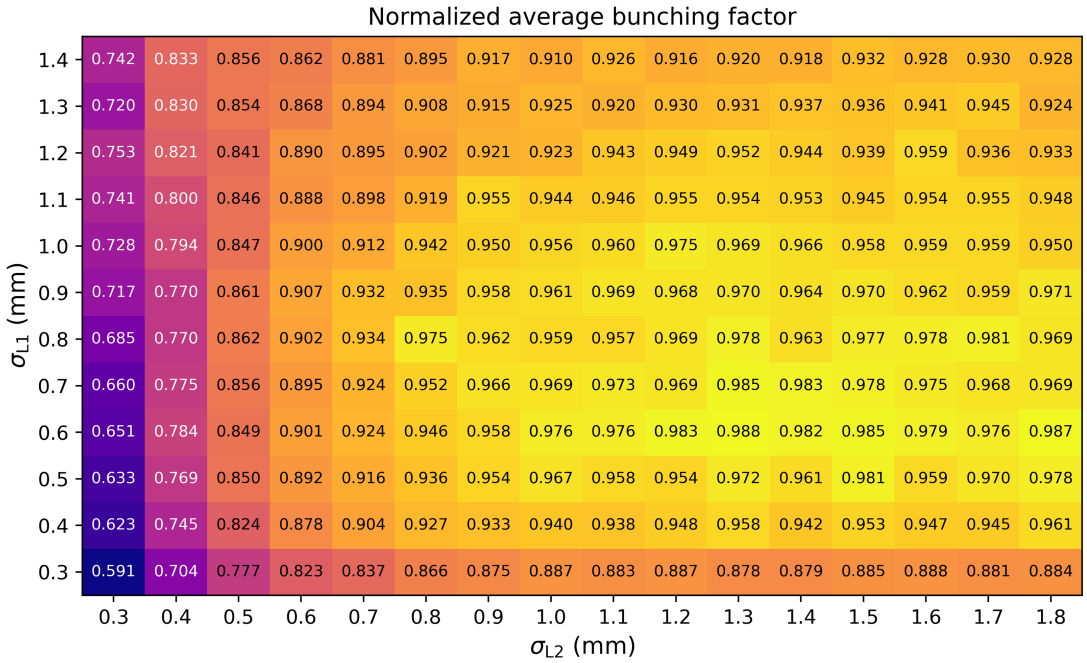

Figure 6.6: Normalized average of the bunching factors across different odd harmonics from 11 to 27. Each point in the plot represents the optimized result from a full inner scan over all modulator period combinations.

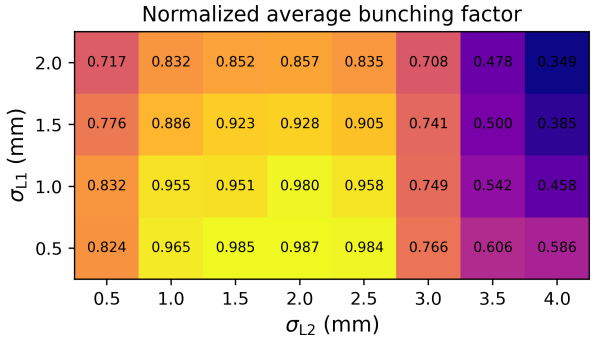

Figure 6.7: Normalized average of the bunching factors across different odd harmonics from 11 to 27 for a longer range of waist sizes.

levels. In contrast, the bunching factor generally improves with a larger beam size of the second laser σ_{L2}, as this creates a more uniform energy modulation. Although a larger waist reduces the energy modulation amplitude, this can be effectively compensated for by increasing the chicane strength. While a bunching factor drop-off was expected due to decreasing laser field intensity at larger waists, it instead forms a plateau for $\sigma_{L2} > 1.3$ mm. To investigate this trend further, a broader scan was conducted (Fig. 6.7), which shows that the bunching factor only begins to decline significantly when $\sigma_{L2} > 3.0$ mm. Because the performance is nearly identical between 1.5 mm and 2.5 mm, the waist size at the onset of the plateau was selected to maintain higher laser intensity and energy modulation without

sacrificing the bunching factor. Therefore, the configuration of $\sigma_{L1} = 0.6$ mm and $\sigma_{L2} = 1.3$ mm was chosen as the working point for subsequent simulations.

6.2.3 Optimization of Number of Modulator Periods

With the optimal laser beam sizes of $\sigma_{L1} = 0.6$ mm and $\sigma_{L2} = 1.3$ mm identified from the broad, simplified scan, the second stage of the optimization was performed. For this, the detailed simulation code (described in Section 5.4) was used to refine the choice for the number of modulator periods (N_{M1}, N_{M2}).

Figure 6.8 presents the bunching factor as a function of the number of periods in each modulator for four different harmonics. The corresponding normalized and averaged bunching

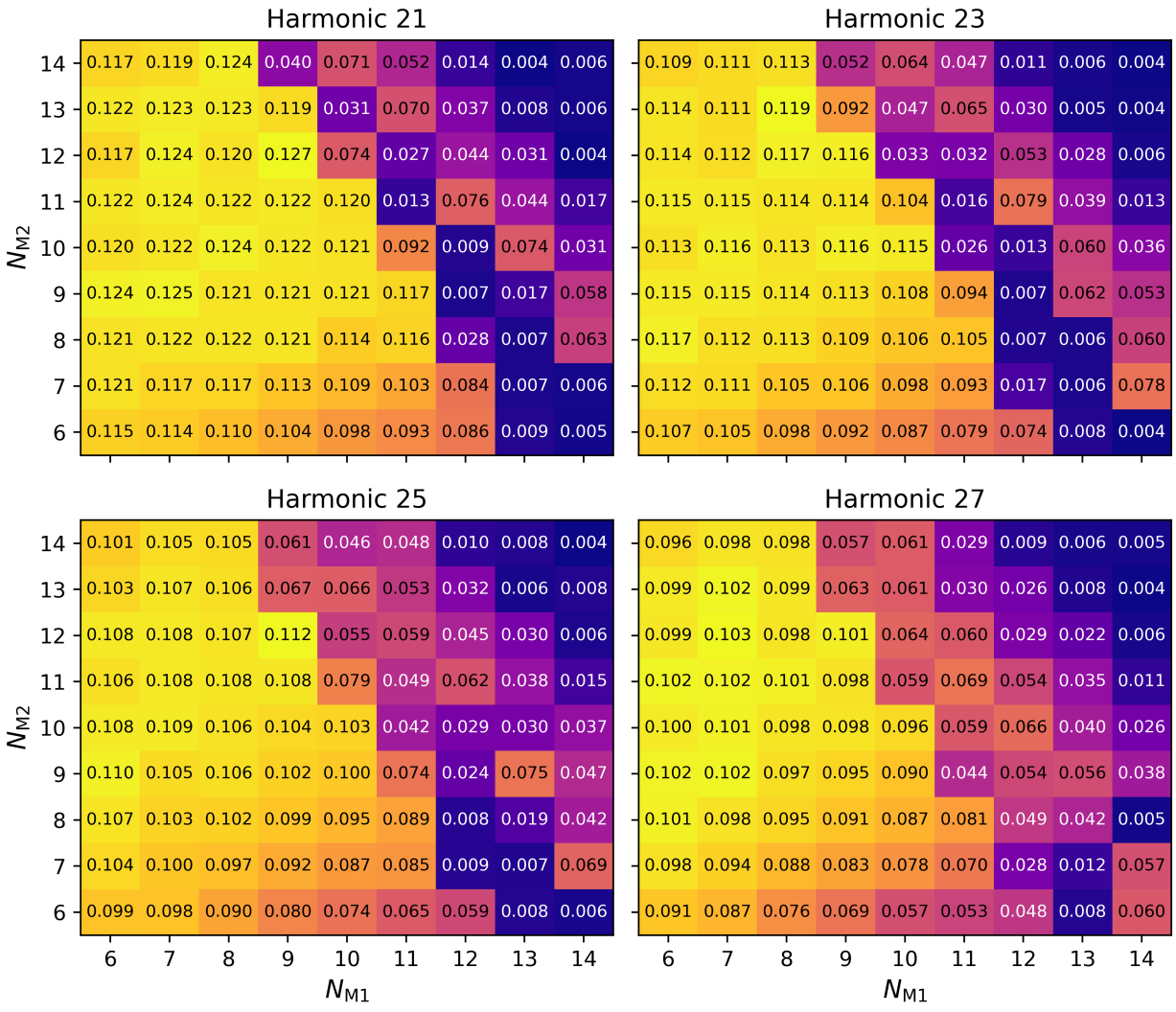

Figure 6.8: Maximum bunching factor for different numbers of periods in the first modulator N_{M1} and the second modulator N_{M2} for four distinct harmonics.

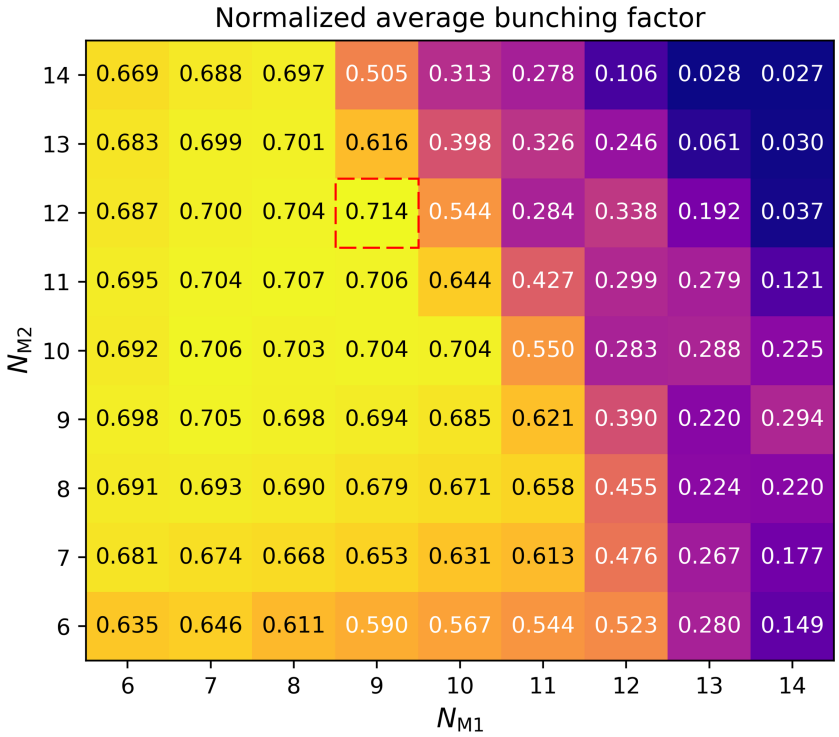

Figure 6.9: Normalized average bunching factor as a function of the number of undulator periods in each modulator. The bunching factor is averaged over all odd harmonics from 11 to 27. The red dashed outline indicates the optimal configuration of $N_{M1} = 9$ and $N_{M2} = 12$, which yields the highest average bunching factor.

factor is shown in Fig. 6.9. Averaging over all odd harmonics from 11 to 27, the maximum bunching factor occurs for 9 periods in the first modulator and 12 periods in the second, indicated by the red dashed outline.

The results clearly show that the bunching factor decreases sharply as the modulators are extended. While longer modulators enhance the energy modulation, they simultaneously restrict the space available for the chicanes. This limitation prevents the chicanes from reaching the strengths required for efficient microbunching. The boundary defined by this trade-off is evident in Fig. 6.9. The optimal configuration therefore lies directly adjacent to this boundary, striking a balance between providing sufficient energy modulation and leaving adequate space for strong chicanes.

Figure 6.10 shows the optimum chicane strengths to optimize the bunching factor for different odd harmonics of 800 nm up to the 29th harmonic, corresponding to a wavelength of 27.6 nm. With a maximum chicane strength of $R_{56}^{(1)} > 1200$ μm and $R_{56}^{(2)} > 50$ μm, all harmonics below 30 will be accessible with the optimized setup.

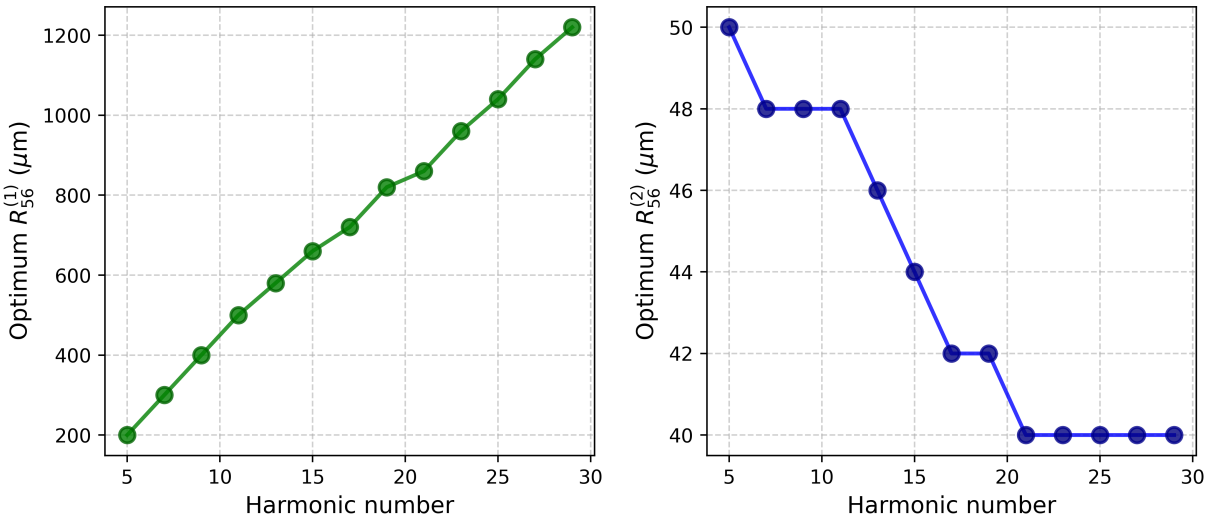

Figure 6.10: First (left) and second (right) chicane strength required to optimize the bunching factor at different harmonics of 800 nm. The setup is optimized for the odd harmonics spanning the range from 11 to 29.

6.3 Optimized Design and Performance Analysis

6.3.1 Final Design Parameters

Table 6.1 summarizes the optimized choices for the undulator and chicane parameters and a sketch of the optimized configuration is depicted in Fig. 6.11. With these choices, the total length of the section comes to be 4875 mm.

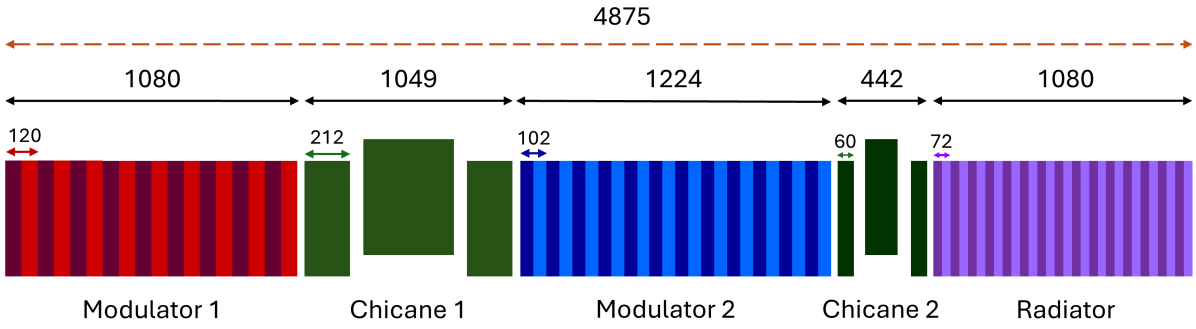

Figure 6.11: The optimized EEHG setup with hybrid PMUs and electromagnetic chicanes. The measurements are shown in millimeters.

Component	**Specifications**
1st modulator	No. of periods = 9 Period length = 120 mm Total length = 1080 mm
2nd modulator	No. of periods = 12 Period length = 102 mm Total length = 1224 mm
Radiator	No. of periods = 15 Period length = 72 mm Total length = 1080 mm
1st chicane	Total length = 1049 mm Max. R_{56} = 1244 μm
2nd chicane	Total length = 442 mm Max. R_{56} = 50 μm
1st laser	Pulse energy = 5.0 mJ Waist size = 0.6 mm
2nd laser	Pulse energy = 4.4 mJ Waist size = 1.3 mm

Table 6.1: Optimized undulator, chicane, and laser parameters.

6.3.2 Coherent-to-Incoherent Ratio

To estimate the ratio between coherent and incoherent power, their different origins must be considered. Coherent EEHG radiation is emitted only from the 100 fs temporal slice modulated by the laser, whereas the entire electron bunch contributes to the incoherent background radiation. For this analysis, a typical 80 ps single-bunch at DELTA was assumed, carrying a 10 mA beam current. This corresponds to a total bunch charge of 3.84 nC, or a total number of electrons $N_e \approx 2.4 \times 10^{10}$. The total incoherent radiation power P_{inc} scales linearly with N_e. To calculate the coherent power P_{coh} the electron bunch was simulated and divided into 800 nm-wide longitudinal slices. Assuming a uniform longitudinal density, each slice contained approximately 8×10^5 electrons. The bunching factor was then calculated for each individual slice, and the total coherently emitted power was determined using the relationship in Eq. 6.5.

Figure 6.12 shows an example of the electron phase space distribution with two 100 fs seed pulses and the corresponding bunching profile.

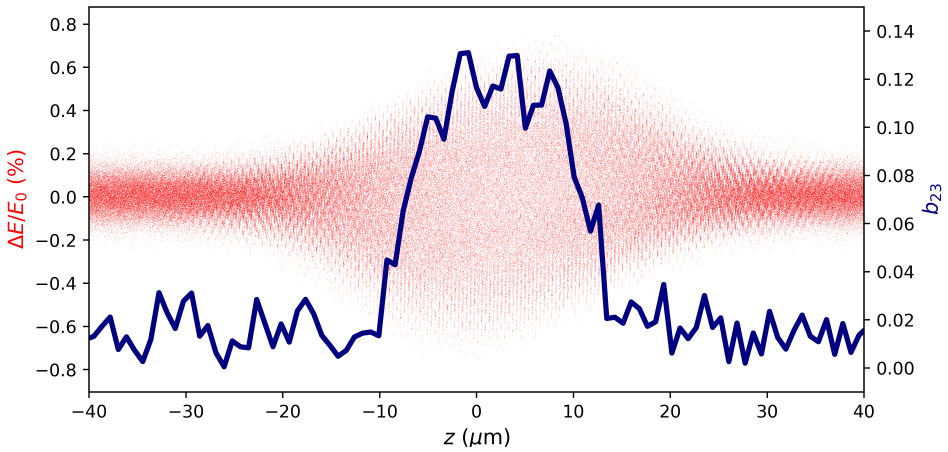

Figure 6.12: Electron phase space distribution (red) and the corresponding slice bunching factor (dark blue) at the 23rd harmonic along an 80 μm (267 fs) region of the electron bunch. The two seed pulses carry 5 mJ and 4.4 mJ pulse energy with 100 fs pulse duration. The chicanes are set to 990 μm and 42 μm, respectively.

The signal-to-background ratio (SBR) between P_{coh} and P_{inc} is then calculated using Eq. (6.6). Figure 6.13 displays the calculated maximum bunching factor and corresponding SBR for different odd harmonics of 800 nm. The results show that the setup provides substantial bunching factors above 0.1 up to the 25th harmonic, while the coherently emitted power exceeds the incoherent background by at least a factor of 10 up to the 23rd harmonic.

Considering the lower operational limit of the radiator at the 10th harmonic and the upper limit imposed by the first chicane strength, the proposed setup can operate effectively between the 11th and 29th harmonics of 800 nm. Table 6.2 summarizes the optimal chicane

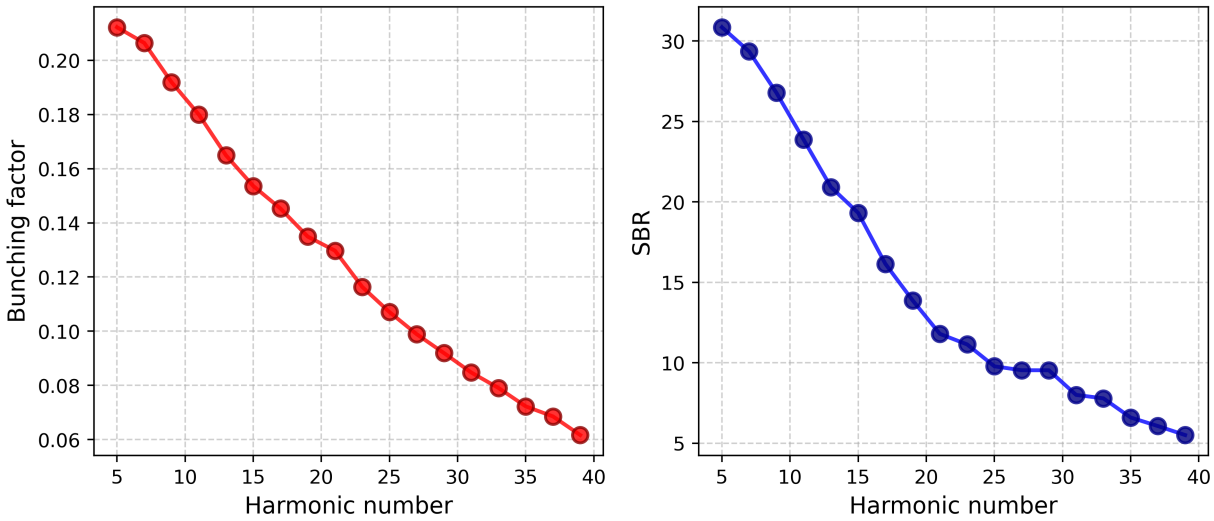

Figure 6.13: Left: The maximum bunching factor for odd harmonics up to the 39th harmonic. **Right:** The corresponding signal-to-noise ratio.

strengths, maximum bunching factors, corresponding SBR, and the coherent emission flux for several harmonics within this operational range. The coherent emission flux was estimated by scaling the incoherent flux, calculated via *SPECTRA* [113] using DELTA parameters with a beam current of 10 mA, by the estimated SBR.

h	λ [nm]	$R_{56}^{(1)}$ [μm]	$R_{56}^{(2)}$ [μm]	Max b_h	SBR	Coherent flux [$\times 10^{10}$ ph/s/0.1% BW]
11	72.7	510	48	0.180	23.9	8.3
13	61.5	590	46	0.165	20.9	7.2
15	53.3	660	44	0.154	19.3	6.6
17	47.1	730	42	0.141	16.2	5.6
19	42.1	820	42	0.135	13.9	4.8
21	38.1	910	42	0.128	11.8	4.0
23	34.8	990	42	0.114	11.1	3.8
25	32.0	1050	40	0.106	9.8	3.3
27	29.6	1130	40	0.099	9.5	3.2
29	27.6	1230	40	0.092	9.5	3.2

Table 6.2: Optimized chicane parameters, bunching factor, SBR, and the coherent emission flux at 10 mA for the odd harmonics in the operational range.

6.4 Summary

This chapter investigated the theoretical feasibility of an EEHG setup based on permanent-magnet undulators that could be accommodated within a straight section of less than 5 meters, which represents typical straight section lengths available at many synchrotron radiation sources. This investigation demonstrated that even under these tight spatial constraints, a setup capable of reaching up to the 29th harmonic of the 800 nm seed wavelength with an SBR ≥ 10 remains feasible.

In this study, a total length restriction of 4.875 m was imposed, matching the current U250-based SPEED setup at DELTA. Through systematic optimization of the undulator and chicane section lengths, an optimized configuration that operates within the harmonic range of 11 to 29 was developed while maintaining an SBR ratio above 10 up to the 23rd harmonic (34.8 nm) using the beam parameters of the DELTA storage ring.

Storage rings with shorter electron bunches, achieved through higher radiofrequency voltage and/or lower momentum compaction factor, would yield significantly improved SBR due to the larger fraction of energy-modulated electrons. Furthermore, while this study used DELTA parameters, implementation at modern light sources with lower beam emittance and energy spread would relax the constraints. Alternatively, seeding with shorter wavelengths would decrease the required modulator period length and enable access to shorter output wavelengths.

While the wavelength range achieved here does not extend to the X-ray regime accessible by state-of-the-art EEHG implementations at FELs, the proposed configuration offers the distinct advantage of compact implementation. The compactness of the design makes it practical for integration into many existing storage rings without significant modifications. To further enhance cost-efficiency, the modulators can be implemented as fixed-gap devices if the beam energy is constant, leaving only the radiator requiring a tunable gap mechanism.

While this study focused on planar undulators, implementing an elliptical undulator as the radiator would significantly increase experimental opportunities by providing variable polarization. However, this would require higher costs and an increase in spatial dimensions. This opens the possibility for developing EEHG-based user facilities that would provide enhanced accessibility and operational flexibility compared to dedicated FEL installations. Furthermore, a storage-ring-based source offers inherently higher shot-to-shot stability than a linac-based FEL, while also providing a higher repetition rate in the kHz range, compared to the 100 Hz or less for FELs that use normal-conducting linacs.

Chapter 7

Conclusion

The studies presented in this thesis focus on generating and characterizing coherently emitted, ultrashort, short-wavelength light pulses. The work broadly spans two complementary research directions: the development of deep learning frameworks for beam diagnostics and the experimental implementation of echo-enabled harmonic generation (EEHG) in a storage ring.

The first part of this work explored the application of deep learning to beam diagnostics. The application of convolutional neural networks (CNNs) to coherent harmonic generation (CHG) spectra proved capable of identifying the spectro-temporal properties of the seed laser and CHG pulses. Furthermore, for free-electron laser (FEL) diagnostics, the development of U-Net and Y-Net architectures showed that photon pulse profiles can be reconstructed from measured longitudinal phase space images without relying on reference measurements. This confirms that machine learning can provide reliable, single-shot diagnostics while eliminating the artifacts often introduced by reference-subtraction schemes.

Experimentally, the commissioning of the SPEED project at DELTA provided the first proof-of-principle that the EEHG mechanism is robust enough for electron storage rings. Despite the constraints of a standard storage ring straight section, weak energy modulation amplitudes, and large natural beam energy spread, coherent emission was successfully characterized up to the 11th harmonic (73 nm). These results validate that the intricate phase space manipulation required for EEHG is feasible in a circular accelerator. Following these experiments, a conceptual design for a permanent-magnet-based EEHG setup was developed. This design enables access to shorter wavelengths (down to 27.6 nm with a signal-to-background ratio close to 10) while remaining compact enough for typical straight sections of many storage rings around the world. Producing femtosecond pulses in this regime, comparable to the 30 nm output long-offered by the FLASH FEL, would be a major milestone for storage-ring-based sources. Unlike the femtoslicing technique, the EEHG scheme provides

high longitudinal coherence, which allows for the exploitation of phase-dependent phenomena such as coherent control [164]. While the current design is optimized for the specific constraints of the DELTA storage ring, by utilizing longer modulators, reduced beam emittance, higher seed pulse energy, or shorter seed wavelengths, future implementations could potentially reach even shorter wavelengths of high scientific importance, such as 13.5 nm for EUV lithography, 4.4 nm (within the water window) for biological imaging.

The contributions detailed here are relevant for the development of future accelerator-based light sources. While the current trend in storage ring development is directed towards low-emittance storage rings [165–169], there remains a scientific need for high-repetition-rate, ultrashort pulses that low-emittance rings cannot naturally provide. A storage ring designed specifically for short-pulse coherent emission, utilizing multiple straight sections for CHG, EEHG, or similar schemes, could address this need.

The intersection of the experimental and computational avenues offers a promising path forward for the operation of future light sources. Deep learning frameworks are well-suited to handle the complex parameter space of seeding schemes like EEHG, which require precise control over laser timing, spatial overlap, chicane strengths, and electron orbit. For instance, the neural networks trained to analyze the radiation output could be integrated into a real-time feedback system enabling automated control of the seed laser and the non-destructive optimization of the seeding process. Deep learning models can also be used for virtual diagnostics that infer unmeasurable beam properties, such as the bunching factor, based on available machine parameters. Furthermore, methods such as reinforcement learning [170] can be used to find optimal working points in the high-dimensional parameter space that are difficult for human operators to locate. By replacing the demanding manual tuning currently required for such experiments, such an autonomous and data-driven control loop could streamline the operation of future facilities.

Bibliography

[1] A. H. Zewail. *Femtochemistry: Atomic-scale dynamics of the chemical bond*. The Journal of Physical Chemistry A 104, pp. 5660–5694 (2000).

[2] A. W. Chao and M. Tigner, eds. *Handbook of Accelerator Physics and Engineering*. 3rd. World Scientific. (1999).

[3] M. Abo-Bakr et al. *Steady-state far-infrared coherent synchrotron radiation detected at BESSY II*. Phys. Rev. Lett. 88, p. 254801 (2002).

[4] A. L'Huillier and P. Balcou. *High-order harmonic generation in rare gases with a 1-ps 1053-nm laser*. Phys. Rev. Lett. 70, pp. 774–777 (1993).

[5] P. B. Corkum. *Plasma perspective on strong field multiphoton ionization*. Phys. Rev. Lett. 71, pp. 1994–1997 (1993).

[6] M. Lewenstein et al. *Theory of high-harmonic generation by low-frequency laser fields*. Phys. Rev. A 49, pp. 2117–2132 (1994).

[7] E. Constant et al. *Optimizing high harmonic generation in absorbing gases: model and experiment*. Phys. Rev. Lett. 82, pp. 1668–1671 (1999).

[8] C. Pellegrini, A. Marinelli, and S. Reiche. *The physics of X-ray free-electron lasers*. Rev. Mod. Phys. 88, p. 015006 (2016).

[9] *Lightsources.org*. Accessed: 2026-01-11. (2026).

[10] R. W. Schoenlein et al. *Generation of femtosecond pulses of synchrotron radiation*. Science 287, pp. 2237–2240 (2000).

[11] R. Coisson and F. De Martini. *Free-electron coherent relativistic scatterer for UV generation*. Free-Electron Generators of Coherent Radiation, pp. 939–960 (1982).

[12] G. Stupakov. *Using the beam-echo effect for generation of short-wavelength radiation*. Phys. Rev. Lett 102, p. 074801 (2009).

[13] D. Xiang and G. Stupakov. *Echo-enabled harmonic generation free electron laser*. Phys. Rev. Accel. Beams 12, p. 030702 (2009).

[14] B. Girard et al. *Optical frequency multiplication by an optical klystron*. Phys. Rev. Lett. 53, p. 2405 (1984).

[15] M. Labat et al. *Coherent harmonic generation on UVSOR-II storage ring*. Eur. Phys. J. D 44, p. 187 (2007).

[16] G. De Ninno et al. *Generation of ultrashort coherent vacuum ultraviolet pulses using electron storage rings: A new bright light source for experiments*. Phys. Rev. Lett. 101, p. 053902 (2008).

[17] S. Khan et al. *Coherent harmonic generation at DELTA: A new facility for ultrashort pulses in the VUV and THz regime*. Synchrotron Radiation News 24, pp. 18–23 (2011).

[18] S. Khan et al. *Generation of ultrashort and coherent synchrotron radiation pulses at DELTA*. Synchrotron Radiation News 26, pp. 25–29 (2013).

[19] D. Xiang et al. *Demonstration of the echo-enabled harmonic generation technique for short-wavelength seeded free electron lasers*. Phys. Rev. Lett. 105, p. 114801 (2010).

[20] P. Rebernik Ribič et al. *Coherent soft X-ray pulses from an echo-enabled harmonic generation free-electron laser*. Nature Photonics 13, pp. 555–561 (2019).

[21] C. Feng et al. *Coherent extreme ultraviolet free-electron laser with echo-enabled harmonic generation*. Phys. Rev. Accel. Beams 22, p. 050703 (2019).

[22] D. Gauthier et al. *Spectrotemporal shaping of seeded free-electron laser pulses*. Phys. Rev. Lett. 115, p. 114801 (2015).

[23] P. Emma, J. Frisch, and P. Krejcik. *A transverse RF deflecting structure for bunch length and phase space diagnostics*. Technical Report LCLS-TN-00-12. SLAC National Accelerator Laboratory. (2000).

[24] C. Behrens et al. *Few-femtosecond time-resolved measurements of X-ray free-electron lasers*. Nature Communications 5, p. 3762 (2014).

[25] Nobel Prize Outreach. *The Nobel prize in physics 2024*. Accessed: 2025-11-07. (2025).

[26] K. Wille. *The Physics of Particle Accelerators: An Introduction*. Clarendon Press. (2000).

[27] H. Wiedemann. *Particle Accelerator Physics, Fourth Edition*. Springer. (2015).

[28] A. Wolski. *Beam Dynamics in High Energy Particle Accelerators*. Imperial College Press. (2014).

[29] H. Mais and G. Ripken. *Theory of coupled synchro-betatron oscillation*. DESY, Hamburg, Germany. (1982).

[30] D. Edwards and L. Teng. *Parametrization of linear coupled motion in periodic systems*. IEEE Transactions on Nuclear Science 20, pp. 885–888 (1973).

[31] J. J. Thomson. *XL. Cathode rays*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 44, pp. 293–316 (1897).

[32] E. Esarey, C. B. Schroeder, and W. P. Leemans. *Physics of laser-driven plasma-based electron accelerators*. Rev. Mod. Phys. 81, pp. 1229–1285 (2009).

[33] J. Larmor. *LXIII. on the theory of the magnetic influence on spectra; and on the radiation from moving ions*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 44, pp. 503–512 (1897).

[34] A. Liénard. *Champ électrique et magnétique produit par une charge électrique concentrée en un point et animée d'un mouvement quelconque*. G. Carré et C. Naud. (1898).

[35] K.-J. Kim, Z. Huang, and R. Lindberg. *Synchrotron Radiation and Free-Electron Lasers: Principles of Coherent X-Ray Generation*. Cambridge University Press. (2017).

[36] Coherent, Inc. *Operator's manual - Coherent Legend Elite ultrafast amplifier laser systems*. (2010).

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

[38] W. E. Lamb. *Theory of an optical maser*. Phys. Rev. 134, A1429–A1450 (1964).

[39] A. M. Weiner. *Ultrafast Optics*. John Wiley & Sons. (2009).

[40] A. Siegman. *Lasers*. University Science Books. (1986).

[41] L. J. Aplet and J. W. Carson. *A Faraday effect optical isolator*. Appl. Opt. 3, pp. 544–545 (1964).

[42] P. Ungelenk. *Generation and detection schemes for laser-induced coherent terahertz radiation at the electron storage ring DELTA*. Dissertation. TU Dortmund, Germany. (2015).

[43] Z. Huang and K.-J. Kim. *Review of X-ray free-electron laser theory*. Phys. Rev. Accel. Beams 10, p. 034801 (2007).

[44] K.-J. Kim, L. Ryan, and Z. Huang. *High-gain regime: 1D*. Proc. of the CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs. Hamburg, Germany. (2018).

[45] P. Schmüser et al. *Free-Electron Lasers in the Ultraviolet and X-Ray Regime: Physical Principles, Experimental Results, Technical Realization*. Springer International Publishing, pp. 63–84 (2014).

[46] H. P. Freund and T. M. Antonsen Jr. *Principles of Free Electron Lasers*. Springer International Publishing. (2024).

[47] S. Reiche. *Overview of seeding methods for FELs*. Proc. of IPAC2013. Shanghai, China, pp. 2063–2067 (2013).

[48] S. Ackermann et al. *Generation of coherent 19- and 38-nm radiation at a free-electron laser directly seeded at 38 nm*. Phys. Rev. Lett. 111, p. 114801 (2013).

[49] H. Tomizawa et al. *Stable operation of HHG-seeded EUV-FEL at the SCSS test accelerator*. Proc. of FEL2013. New York, USA, pp. 728–733 (2013).

[50] E. Hemsing, A. Halavanau, and Z. Zhang. *Enhanced self-seeding with ultrashort electron beams*. Phys. Rev. Lett. 125, p. 044801 (2020).

[51] J. Amann et al. *Demonstration of self-seeding in a hard-X-ray free-electron laser*. Nature Photonics 6, pp. 693–698 (2012).

[52] E. Allaria et al. *Highly coherent and stable pulses from the FERMI seeded free-electron laser in the extreme ultraviolet*. Nature Photonics 6, pp. 699–704 (2012).

[53] E. Allaria et al. *Two-stage seeded soft-X-ray free-electron laser*. Nature Photonics 7, pp. 913–918 (2013).

[54] Z. Zhao et al. *Status of the SXFEL facility*. Applied Sciences 7, (2017).

[55] M. Vogt, S. Schreiber, and J. Zemella. *FLASH: Status and upgrade*. Proc. of FLS'23. 67. Lucerne, Switzerland, pp. 32–37 (2024).

[56] D. Robin et al. *Low alpha experiments at the ALS*. AIP Conference Proceedings 367, pp. 181–190 (1996).

[57] I. P. S. Martin et al. *Experience with low-alpha lattices at the Diamond Light Source*. Phys. Rev. Accel. Beams 14, p. 040705 (2011).

[58] A. A. Zholents and M. S. Zolotorev. *Femtosecond X-ray pulses of synchrotron radiation*. Phys. Rev. Lett. 76, pp. 912–915 (1996).

[59] S. Khan et al. *Femtosecond undulator radiation from sliced electron bunches*. Phys. Rev. Lett. 97, p. 074801 (2006).

[60] G. Ingold et al. *FEMTO: A sub-ps tunable hard X-ray undulator source for laser/X-ray pump-probe experiments at the SLS*. Synchrotron Radiation News 20, pp. 35–39 (2007).

[61] M. Labat et al. *Commissioning of a multi-beamline femtoslicing facility at SOLEIL*. Journal of Synchrotron Radiation 25, pp. 385–398 (2018).

[62] S. Hilbrich et al. *Upgrade plans for the short-pulse facility at DELTA*. Proc. of FEL 2014, pp. 255–259 (2014).

[63] Z. Huang et al. *Suppression of microbunching instability in the linac coherent light source*. Phys. Rev. Accel. Beams 7, p. 074401 (2004).

[64] L.-H. Yu. *Generation of intense UV radiation by subharmonically seeded single-pass free-electron lasers*. Phys. Rev. A 44, p. 5178 (1991).

[65] G. Schmidt et al. *Status of the synchrotron light source DELTA*. Proc. European Particle Accelerator Conference. Paris, France, pp. 745–747 (2002).

[66] D. Schirmer et al. *Status of the synchrotron light source DELTA*. Proc. European Particle Accelerator Conference. Lucerne, Switzerland, pp. 2296–2298 (2004).

[67] A. Jankowiak et al. *Reconstruction of the 75 MeV linac of the DELTA synchrotron-radiation-facility*. Proc. European Particle Accelerator Conference. Vienna, Austria, pp. 634–636 (2000).

[68] F. Marhauser et al. *HOM damped 500 MHz cavity design for 3rd generation SR sources*. Proc. of the 2001 Particle Accelerator Conference. Chicago, USA, pp. 846–848 (2001).

[69] N. Mezentsev et al. *Superconducting 7 Tesla wiggler for Delta synchrotron radiation source*. 26th Russian Particle Accelerator Conference. Protvino, Russia, pp. 410–413 (2018).

[70] D. Nölle et al. *First lasing of the FELICITA I FEL at DELTA*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 445, pp. 128–133 (2000).

[71] D. Schirmer. *Synchrotron radiation sources at DELTA*. Internal Report. (2009).

[72] A. M. auf der Heide. *Simulation and commissioning of a multi-wavelength telescope for the short-pulse facility at DELTA*. Masterarbeit. TU Dortmund, Germany. (2015).

[73] M. Höner. *Optical design and construction of a dedicated THz beamline at DELTA and study of laser-electron interaction*. Diplomarbeit. TU Dortmund, Germany. (2011).

[74] M. Huck. *Characterization of VUV pulses from the short-pulse facility at DELTA and steps towards pump-probe experiments*. Dissertation. TU Dortmund, Germany. (2015).

[75] S. Döring et al. *A short-pulse facility for time and angle resolved photoemission experiments at BL5*. 9th DELTA Annual Report, (2013).

[76] H. Huck. *Optimierung und Charakterisierung des Free Electron Lasers am Speicherring DELTA*. Dissertation. TU Dortmund, Germany. (2009).

[77] R. Molo et al. *Investigation of the magnetic chicane of the short-pulse facility at the DELTA storage ring*. Proc. of IPAC13. Shanghai, China, pp. 1889–1891 (2013).

[78] Optronis GmbH. *Streak camera unit SC-10*. Data Sheet. (2021).

[79] M. Czerny and A. F. Turner. *Über den astigmatismus bei spiegelspektrometern*. Zeitschrift für Physik 61, pp. 792–797 (1930).

[80] Andor Technology Ltd. *iStar CCD spectroscopy specification sheet*. Oxford Instruments Andor Ltd. (2025).

[81] HP Spectroscopy. *easyLIGHT XUV: High-efficiency flat-field XUV spectrometer*. (2025).

[82] W. S. McCulloch and W. Pitts. *A logical calculus of the ideas immanent in nervous activity*. The Bulletin of Mathematical Biophysics 5, pp. 115–133 (1943).

[83] F. Rosenblatt. *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review 65, pp. 386–408 (1958).

[84] M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press. (1969).

[85] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning representations by back-propagating errors*. Nature 323, pp. 533–536 (1986).

[86] J. J. Hopfield. *Neural networks and physical systems with emergent collective computational abilities*. Proc. of the National Academy of Sciences 79, pp. 2554–2558 (1982).

[87] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press. (2016).

[88] Y. Bengio. *Learning Deep Architectures for AI*. Foundations and Trends in Machine Learning. Now Publishers Inc., pp. 1–127 (2009).

[89] K. Hornik, M. Stinchcombe, and H. White. *Multilayer feedforward networks are universal approximators*. Neural Networks 2, pp. 359–366 (1989).

[90] H. Zhao et al. *Loss functions for neural networks for image processing*. arXiv preprint arXiv:1511.08861. (2015).

[91] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press. (1995).

[92] D. P. Kingma and J. Ba. *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980. (2017).

[93] S. Ioffe and C. Szegedy. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. arXiv preprint arXiv:1502.03167. (2015).

[94] A. Tikhonov and V. Arsenin. *Solutions of Ill-posed Problems*. Halsted Press book. Winston. (1977).

[95] N. Srivastava et al. *Dropout: A simple way to prevent neural networks from overfitting*. Journal of Machine Learning Research 15, pp. 1929–1958 (2014).

[96] M. Feurer and F. Hutter. *Automated Machine Learning: Methods, Systems, Challenges*. Ed. by F. Hutter, L. Kotthoff, and J. Vanschoren. Springer International Publishing. (2019).

[97] J. Bergstra and Y. Bengio. *Random search for hyper-parameter optimization*. Journal of Machine Learning Research 13, pp. 281–305 (2012).

[98] J. Snoek, H. Larochelle, and R. P. Adams. *Practical Bayesian optimization of machine learning algorithms*. Proc. of the 26th International Conference on Neural Information Processing Systems. Vol. 2, pp. 2951–2959 (2012).

[99] Y. LeCun, Y. Bengio, and G. Hinton. *Deep learning*. Nature 521, pp. 436–444 (2015).

[100] D. H. Hubel and T. N. Wiesel. *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*. The Journal of Physiology 160, pp. 106–154 (1962).

[101] Y. LeCun et al. *Gradient-based learning applied to document recognition*. Proc. of the IEEE 86, pp. 2278–2324 (1998).

[102] A. Krizhevsky, I. Sutskever, and G. E. Hinton. *ImageNet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems. Vol. 25, pp. 1097–1105 (2012).

[103] K. Simonyan and A. Zisserman. *Very deep convolutional networks for large-scale image recognition*. International Conference on Learning Representations (ICLR). San Diego, USA. (2015).

[104] K. He et al. *Deep residual learning for image recognition*. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA, pp. 770–778 (2016).

[105] O. Ronneberger, P. Fischer, and T. Brox. *U-Net: Convolutional networks for biomedical image segmentation*. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241 (2015).

[106] E. B. Treacy. *Optical pulse compression with diffraction gratings*. IEEE Journal of Quantum Electronics 5, pp. 454–458 (1969).

[107] P. Ungelenk et al. *Continuously tunable narrowband pulses in the THz gap from laser-modulated electron bunches in a storage ring*. Phys. Rev. Accel. Beams 20, p. 020706 (2017).

[108] D. Kane and R. Trebino. *Characterization of arbitrary femtosecond pulses using frequency-resolved optical gating*. IEEE Journal of Quantum Electronics 29, pp. 571–579 (1993).

[109] C. Iaconis and I. A. Walmsley. *Spectral phase interferometry for direct electric-field reconstruction of ultrashort optical pulses*. Opt. Lett. 23, pp. 792–794 (1998).

[110] M. Abadi et al. *TensorFlow: a system for large-scale machine learning*. 12th USENIX symposium on operating systems design and implementation (OSDI 16), pp. 265–283 (2016).

[111] P. J. Huber. *Robust estimation of a location parameter*. Breakthroughs in statistics: Methodology and distribution, pp. 492–518 (1992).

[112] Y. Gal and Z. Ghahramani. *Dropout as a Bayesian approximation: Representing model uncertainty in deep learning*. Proceedings of The 33rd International Conference on Machine Learning. Vol. 48. New York, USA, pp. 1050–1059 (2016).

[113] T. Tanaka. *Major upgrade of the synchrotron radiation calculation code SPECTRA*. Journal of Synchrotron Radiation 28, pp. 1267–1272 (2021).

[114] A. Marinelli et al. *Optical shaping of X-ray free-electron lasers*. Phys. Rev. Lett. 116, p. 254801 (2016).

[115] Z. Zhang et al. *Experimental demonstration of enhanced self-amplified spontaneous emission by photocathode temporal shaping and self-compression in a magnetic wiggler*. New Journal of Physics 22, p. 083030 (2020).

[116] J. P. Duris et al. *Controllable X-ray pulse trains from enhanced self-amplified spontaneous emission*. Phys. Rev. Lett. 126, p. 104802 (2021).

[117] S. Li et al. *"Beam à la carte": Laser heater shaping for attosecond pulses in a multiplexed X-ray free-electron laser*. Applied Physics Letters 125, p. 191101 (2024).

[118] P. Amstutz et al. *Towards short-pulse generation at FLASH via laser-assisted electron bunch manipulation*. Proc. of IPAC'24. Nashville, USA, pp. 404–407 (2024).

[119] D. Ratner et al. *Time-resolved imaging of the microbunching instability and energy spread at the Linac Coherent Light Source*. Phys. Rev. Accel. Beams 18, p. 030704 (2015).

[120] E. Saldin, E. Schneidmiller, and M. Yurkov. *Longitudinal space charge-driven microbunching instability in the TESLA Test Facility linac*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 528, pp. 355–359 (2004).

[121] J. Tang et al. *Laguerre-Gaussian mode laser heater for microbunching instability suppression in free-electron lasers*. Phys. Rev. Lett. 124, p. 134801 (2020).

[122] D. Cesar et al. *Electron beam shaping via laser heater temporal shaping*. Phys. Rev. Accel. Beams 24, p. 110703 (2021).

[123] R. Akre et al. *A transverse RF deflecting structure for bunch length and phase space diagnostics*. Proceedings of the 2001 Particle Accelerator Conference. Vol. 3. Chicago, USA, pp. 2353–2355 (2001).

[124] H. Ego et al. *Transverse C-band deflecting structure for longitudinal electron-bunch-diagnosis in XFEL "SACLA"*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 795, pp. 381–388 (2015).

[125] P. Craievich et al. *Novel X-band transverse deflection structure with variable polarization*. Phys. Rev. Accel. Beams 23, p. 112001 (2020).

[126] M. Vogt et al. *RF commissioning and first beam operation of the PolariX transverse deflecting structures in the FLASH2 beamline*. Proc. of FEL2022. Trieste, Italy, pp. 428–431 (2022).

[127] B. Marchetti et al. *Experimental demonstration of novel beam characterization using a polarizable X-band transverse deflection structure*. Scientific Reports 11, p. 3560 (2021).

[128] S. Düsterer. *Determination of temporal FEL pulse properties: challenging concepts and experiments*. SPIE Proceedings 8078, pp. 69–77 (2011).

[129] A. Sanchez-Gonzalez et al. *Accurate prediction of X-ray pulse properties from a free-electron laser using machine learning*. Nature Communications 8, p. 15461 (2017).

[130] B. Cheng and D. M. Titterington. *Neural networks: A review from a statistical perspective*. Statistical Science 9, pp. 2–30 (1994).

[131] A. J. Smola and B. Schölkopf. *A tutorial on support vector regression*. Statistics and Computing 14, pp. 199–222 (2004).

[132] T. Korten et al. *Virtual pulse reconstruction diagnostic for single-shot measurement of free electron laser radiation power*. Phys. Rev. Accel. Beams 28, p. 030703 (2025).

[133] S. Haykin. *Neural Networks: A Comprehensive Foundation*. 2nd. Prentice Hall PTR. (1998).

[134] Y. Ding et al. *Femtosecond X-ray pulse temporal characterization in free-electron lasers using a transverse deflector*. Phys. Rev. Accel. Beams 14, p. 120701 (2011).

[135] K. Tiedtke et al. *Gas detectors for X-ray lasers*. Journal of Applied Physics 103, p. 094511 (2008).

[136] X. Ren et al. *Temporal power reconstruction for an X-ray free-electron laser using convolutional neural networks*. Phys. Rev. Accel. Beams 23, p. 040701 (2020).

[137] G. Goetzke et al. *femto-PIXAR: a self-supervised neural network method for reconstructing femtosecond X-ray free electron laser pulses*. Opt. Express 33, pp. 31235–31243 (2025).

[138] S. Reiche. *GENESIS 1.3: a fully 3D time-dependent FEL simulation code*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 429, pp. 243–248 (1999).

[139] M. Borland. *ELEGANT: A flexible SDDS-compliant code for accelerator simulation*. Technical Report. Argonne National Lab., IL, US. (2000).

[140] X. Glorot, A. Bordes, and Y. Bengio. *Deep sparse rectifier neural networks*. Proc. of the Fourteenth International Conference on Artificial Intelligence and Statistics. Vol. 15. Proc. of Machine Learning Research. Fort Lauderdale, USA, pp. 315–323 (2011).

[141] J. Long, E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015).

[142] I. Loshchilov and F. Hutter. *Decoupled weight decay regularization*. arXiv preprint arXiv:1711.05101. (2019).

[143] A. Vaswani et al. *Attention is all you need*. Advances in Neural Information Processing Systems. Vol. 30. 1, pp. 261–272 (2017).

[144] B. Büsing et al. *Progress towards EEHG seeding at the DELTA storage ring*. Proc. of IPAC'22. 13. Bangkok, Thailand, pp. 1420–1422 (2022).

[145] Dassault Systèmes. *CST Microwave Studio*. (2022).

[146] S. Khan et al. *SPEED: Worldwide first EEHG implementation at a storage ring*. Proc. IPAC'23. 14. Venice, Italy, pp. 1057–1060 (2023).

[147] R. Molo. *Towards echo-enabled harmonic generation at FLASH1 and DELTA*. Dissertation. TU Dortmund, Germany. (2017).

[148] B. Büsing et al. *Preparatory experimental investigations in view of EEHG at the DELTA storage ring*. Proc. of FEL 2022. 40. Trieste, Italy, pp. 313–316 (2022).

[149] T. Schmidt. *Aufbau des FEL experiments FELICITA I im sichtbaren und ultravioletten spektralbereich am speicherring DELTA*. Dissertation. Universität Dortmund, Germany. (1997).

[150] P. Elleaume. *Theory of the optical klystron*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 250, pp. 220–227 (1986).

[151] Ocean Optics. *Maya2000 and Maya2000Pro data sheet*. (2010).

[152] R. W. Boyd. *Nonlinear Optics*. Third Edition. Academic Press. (2008).

[153] *https://www.dataray.com*. Accessed: 2025-11-15. (2025).

[154] A. M. auf der Heide et al. *Progress towards an EEHG-based short-pulse source at DELTA*. Proc. of IPAC'17, pp. 2582–2585 (2017).

[155] T. Tsang. *Reflected optical harmonics from dielectric mirrors*. Appl. Opt. 33, pp. 7720–7724 (1994).

[156] A. Savitzky and M. J. E. Golay. *Smoothing and differentiation of data by simplified least squares procedures*. Analytical Chemistry 36, pp. 1627–1639 (1964).

[157] R. W. Schoenlein et al. *Generation of femtosecond X-ray pulses via laser–electron beam interaction*. Applied Physics B 71, pp. 1–10 (2000).

[158] E. Hemsing. *Minimum spectral bandwidth in echo seeded free electron lasers*. Frontiers in Physics 7:35, (2019).

[159] K. Holldack et al. *Femtosecond terahertz radiation from femtoslicing at BESSY*. Phys. Rev. Lett. 96, p. 054801 (2006).

[160] J. A. Clarke. *The Science and Technology of Undulators and Wigglers*. Oxford University Press. (2004).

[161] J. Bahrdt. *Synchrotron Light Sources and Free-Electron Lasers: Accelerator Physics, Instrumentation and Science Applications*. Springer International Publishing, pp. 751–819 (2016).

[162] Coherent, Inc. *Legend Elite Duo HE ultrafast Ti:Sapphire amplifier datasheet*. Accessed: 2025-09-09. (2019).

[163] Spectra-Physics. *Spitfire® Ace™ advanced Ti:Sapphire regenerative amplifiers*. Accessed: 2025-09-09. (2025).

[164] K. C. Prince et al. *Coherent control with a short-wavelength free-electron laser*. Nature Photonics 10, pp. 176–179 (2016).

[165] P. Raimondi and S. M. Liuzzo. *Toward a diffraction limited light source*. Phys. Rev. Accel. Beams 26, p. 021601 (2023).

[166] P. Raimondi et al. *The Extremely Brilliant Source storage ring of the European Synchrotron Radiation Facility*. Communications Physics 6, p. 82 (2023).

[167] L. Liu et al. *Sirius: A 5BA low-emittance lattice with superbends for the new Brazilian Synchrotron Light Source*. Synchrotron Radiation News 26, pp. 34–38 (2013).

[168] J. Kerby. *The advanced photon source upgrade: A brighter future for X-ray science*. Synchrotron Radiation News 36, pp. 26–27 (2023).

[169] E. Al-Dmour et al. *MAX 4U: an upgrade of the MAX IV 3 GeV ring*. Proc. of IPAC'25. Taipei, Taiwan, pp. 752–755 (2025).

[170] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. IEEE Transactions on Neural Networks 9, pp. 1054–1054 (1998).

Acknowledgements

I started my PhD work right in the middle of the COVID-19 pandemic. Prof. Dr. Shaukat Khan was very supportive during my transition from a visitor to a part of DELTA during that difficult time. I would like to express my gratitude to him for giving me the opportunity to work on this interesting topic and for allowing me the freedom to explore other avenues on my own. He constantly provided me with new ideas, criticism, and most importantly, motivation, even when I was on the brink of giving up. His enthusiasm, even at 3 am after a long shift, was truly inspiring. I am also grateful to Prof. Dr. Metin Tolan, who kindly offered to be the second evaluator of my thesis.

Working at DELTA has been made enjoyable and productive by many of my colleagues. Dr. Carsten Mai, my go-to person for all things science or otherwise, has been immensely helpful in helping me set up, understand the workings of DELTA, and navigate the tiring German bureaucracy. Benedikt Büsing was a reliable source for all machine-related matters. For all my doubts and questions during machine operation, he was always available for an answer. I thank him for that.

My group members, Zohair Usfoor and Vivek Vijayan, have been part of almost all the experiment shifts during the past five years. We shared many frustrating moments and laughs throughout this time. Zohair has been particularly helpful in navigating administrative challenges, making many phone calls on my behalf. I thank him for all the help he has rendered, both at DELTA and outside.

My former colleague Dr. Stephan Kotter was my first contact in Dortmund. He lent me his bike and invited me for lunch when I knew nobody in the city. Those bike rides and cooking sessions made the COVID times much more bearable.

I especially thank Tanja Schulte-Eickhoff, Dr. Gerald Schmidt, Dr. Peter Hartmann, Dr. Detlev Schirmer, and Arne Held for the discussions on science, philosophy, music, politics, and many other things. In addition, I express my gratitude to all scientists and technicians at DELTA who have helped me in many ways throughout my work.

I extend my sincere thanks to my collaborators at DESY, Hamburg, who were instrumental in introducing me to the FLARE project and providing access to the data necessary for my

research. Insightful discussions with Dr. Christopher Gerth and Dr. Philipp Amstutz laid the foundation for my work with this dataset. I also thank Dr. Stefan Düsterer and Gesa Goetzke for their assistance in setting up my computational environment on the Maxwell HPC cluster and for clarifying the nuances of the data.

My personal life outside of work was made exciting by my friends Amrutha, Akhil, Alphin, Abhiram, Devika, Naseem, Deepa, Neelima, and Malavika. Thank you for all the lighthearted and fun moments that made the otherwise stressful periods of my life much brighter.

My close friends back from India, Nidhin and Abhi, have been a vital part of the last few years. Although the currents of life have drifted us apart, our spirited discussions on science, God, and philosophy were some of the best late-night conversations ever. The two-week-long Europe trip I took with Nidhin, the first trip after the pandemic settled down, remains a core memory.

Dr. Aarya Prabhakaran, my best friend and partner for the past nine years, has been the most influential person in my life during this journey. She has been with me through the best and worst times, even when facing her own difficulties. Our simple walks through forests and hikes along the mountains remain some of the best memories of my life in Europe. She motivated me to try my hand at machine learning and apply it to my research problems, which turned out to be a significant and successful chunk of my PhD thesis. Thank you for inspiring me and believing in me even when I struggled to believe in myself.

Finally, I would like to thank my parents for making me who I am today. My mother has been a constant source of love and care throughout my life. She provided my sister and me with the best education and the encouragement to follow our passions. Today, we enjoy the fruits of her labor. I thank my sister for the support she has provided to both me and our parents in my absence.

My father, who passed away during the final stages of writing this thesis, was my first science teacher. I fondly remember the countless times we watched the night sky, tracing the constellations and searching for the Milky Way. He ignited that spark of curiosity and love for science in that little child, a spark I hope to carry on. Rest in peace, Dad. I hope I have made you proud.