Longitudinal Bunch Diagnostics using Coherent Transition Radiation Spectroscopy

Physical Principles, Multichannel Spectrometer, Experimental Results, Mathematical Methods

Bernhard Schmidt[1], Stephan Wesch[1], Toke Kövener[2,3], Christopher Behrens[1], Eugen Hass[2], Sara Casalbuoni[4], and Peter Schmüser[1,2].

1. Deutsches Elektronen-Synchrotron DESY Hamburg, 2. Universität Hamburg, 3. CERN, Geneva, 4. Institute for Beam Physics and Technology, Karlsruhe Institute of Technology.

Corresponding authors: Bernhard.Schmidt@desy.de , Peter.Schmueser@desy.de

Abstract

This report summarizes the work on electron bunch diagnostics using coherent transition radiation spectroscopy which our group has carried out over the past 13 years and which is still ongoing.

The generation and properties of transition radiation (TR) are thoroughly treated. The spectral energy density, as described by the Ginzburg-Frank formula, is computed analytically, and the modifications caused by the finite size of the TR screen and near-field diffraction effects are carefully analyzed. The principles of electron bunch shape reconstruction using coherent transition radiation (CTR) are outlined. The three-dimensional form factor is defined and its separation into a transverse and a longitudinal part. Spectroscopic measurements yield only the absolute magnitude of the form factor but not its phase, which however is needed for computing the bunch shape via the inverse Fourier transformation. Two phase retrieval methods are investigated and illustrated with model calculations: analytic phase computation by means of the Kramers-Kronig dispersion relation, and iterative phase retrieval. Particular attention is paid to the ambiguities which are unavoidable in the reconstruction of longitudinal charge density profiles from spectroscopic measurements. The origin of these ambiguities has been identified and a thorough mathematical analysis is presented. The experimental part of the paper comprises a description of our multichannel infrared and THz spectrometer and a selection of measurements at FLASH (Free-electron LASer in Hamburg), comparing the bunch profiles derived from the spectroscopic data with the profiles determined with a transversely deflecting microwave structure.

The appendices are devoted to the mathematical methods. A rigorous derivation of the Kramers-Kronig phase formula is presented in Appendix A. Numerous analytic model calculations can be found in Appendix B. The differences between normal and truncated Gaussians are discussed in Appendix C. Finally, Appendix D contains a short description of the propagation of an electromagnetic wave front by two-dimensional fast Fourier transformation. This is the basis of a powerful numerical Mathematica™ code *THzTransport*, developed in our group, which permits the propagation of electromagnetic wave fronts (visible light, infrared or THz radiation) through an optical beam line consisting of drift spaces, lenses, mirrors and apertures.

1 Introduction

The electron bunches in the high-gain free-electron laser FLASH[1] are longitudinally compressed to achieve peak currents in the kA range which are necessary to drive the high-gain FEL process in the undulator magnets. Bunch compression is accomplished by a two-stage process: first an *energy chirp* (energy-position relationship) is imprinted onto the typically 10 ps long bunches emerging from the electron gun, and then the chirped bunches are passed through magnetic chicanes where the length is reduced to about 100 fs or less. A linearization of the accelerating voltage is achieved by superimposing the 1.3 GHz accelerating field with its third harmonic. A superconducting 3.9 GHz cavity [2] permits optimization of the bunch compression process.

Magnetic compression of intense electron bunches is strongly affected by collective effects in the chicanes and cannot be adequately described by linear beam transfer theory. Space charge forces, coherent synchrotron radiation and wake fields have a profound influence on the time profile and internal energy distribution of the compressed bunches. The collective effects have been studied by various numerical simulations (see [3] and the references quoted therein) but the parameter uncertainties are large and experimental data are indispensable for determining the length and the longitudinal density profile of the bunches before they enter the undulator.

---

[1]The physics of high-gain free-electron lasers and the technology of the soft X-ray FEL FLASH is described in [1].

Our group has applied two time-domain techniques permitting a direct visualization of longitudinal electron bunch profiles with very high resolution: (1) a transversely deflecting microwave structure TDS, and (2) electro-optic (EO) detection systems (see [4] and the references quoted therein), which will not be discussed here. In addition, a high-resolution frequency-domain technique has been developed based on a multichannel single-shot spectrometer for recording coherent transition radiation in the infrared and THz regime.

Transversely deflecting microwave structure TDS

In the TDS the temporal profile of the electron bunch is transferred to a spatial profile on a view screen by a rapidly varying electromagnetic field [5, 6, 7]. The TDS used at FLASH is a 3.6 m long traveling-wave structure operating at 2.856 GHz in which a combination of electric and magnetic fields produces a transverse force for the electrons.
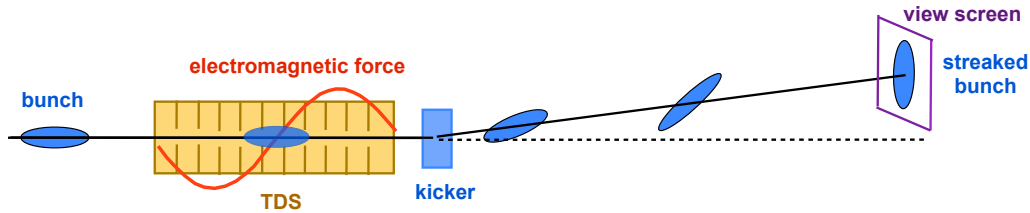

Figure 1: Principle of longitudinal charge density measurement using a transversely deflecting microwave structure. For optimum resolution the radio-frequency (RF) phase is chosen such that the bunch center coincides with the zero-crossing of the RF wave. This condition holds along the entire axis of the traveling-wave structure since electron bunch and RF wave move synchronously with a speed very close to c (light velocity in vacuum).

The bunches pass the TDS near zero crossing of the RF field, and the electrons receive a vertical kick which depends on their longitudinal position inside the bunch. The longitudinal bunch profile is thereby transformed into a streak image on the observation screen. A single bunch out of a train can be streaked. With a fast kicker magnet, this bunch is deflected towards the view screen and recorded by a digital camera. The principle of the TDS is explained in Fig. 1. The time resolution of the TDS installed at FLASH may be as good as 10 fs (rms), depending on the beam optics chosen. An essential prerequisite for good resolution is a large beta function at the position of the TDS. An image of a streaked electron bunch is shown in Fig. 2. An important application was
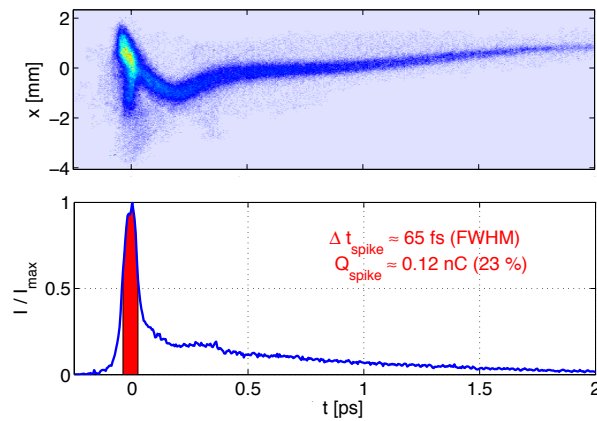

Figure 2: *Top*: Two-dimensional image of a single electron bunch whose time profile is translated into a spatial profile on an observation screen. The bunch head is at the left side. *Bottom*: Current as a function of time. The maximum current is $I_{\max} = 1.8\,\text{kA}$ in this measurement [9].

time-resolved phase space tomography [8, 9] to determine the so-called *slice emittance*. The TDS is routinely used as a diagnostic tool at FLASH, the XFEL and other accelerators.

Infrared and THz spectroscopy

Complementary to the above-mentioned time-domain techniques is spectroscopy in the frequency domain. In particular, coherent transition radiation (CTR) in the infrared and far-infrared (THz) regime has a long tradition as a tool for the longitudinal diagnostics of short electron bunches. The intensity of coherent radiation is proportional to $N^2|F(\omega)|^2$, where N is the number of electrons in the bunch and $F(\omega)$ is the longitudinal form factor (written here as a function of circular frequency $\omega = 2\pi c/\lambda$). We will discuss in detail various methods for determining the bunch length and the internal bunch structure from the measured CTR spectra.

In the past, our group has carried out numerous CTR autocorrelation studies with Martin-Puplett interferometers (see e.g. [10]). In the interferometer the optical delay between the two arms is varied in small steps by moving a mirror. Each bunch makes just one entry in the autocorrelation plot, hence many successive bunches are needed to obtain an average longitudinal shape. To open the way to CTR spectroscopy on single electron bunches a novel multichannel infrared and THz spectrometer with fast readout was jointly developed at DESY and the University of Hamburg [11]. This spectrometer, named CRISP (Coherent transition Radiation Intensity Spectrometer), and its experimental applications are described in this report.

2 Production and Properties of Transition Radiation

When a relativistic electron crosses the boundary between two media of different permittivity, the electromagnetic field carried by the particle changes abruptly upon the transition from one medium to the other. To satisfy the boundary conditions for the electric and magnetic field vectors one has add two radiation fields, one propagating in forward direction, the other in backward direction. This radiation is called transition radiation. The boundary-condition method is straightforward for an infinite planar boundary, it can be generalized to describe the radiation from screens of simple other shapes: a circular disc, a circular hole in an infinite plane, or a semi-infinite half plane [12, 13]. This will not be discussed here because the mathematical effort is considerable and the results apply only for the far-field diffraction regime. Radiation from a screen of arbitrary shape cannot be calculated analytically.

An alternative approach to compute the radiation by a relativistic charged particle at the transition from vacuum into a metal is based on the Weizsäcker-Williams method of virtual quanta, see e.g. [14]. The assumption is made that the virtual photons, constituting the self-field of the particle, are converted into real photons by reflection at the metallic interface. Effectively this means that the Fourier components of the transverse electric field of the electron are reflected at the metal surface. Then the Huygens-Fresnel principle is applied to compute the outgoing electromagnetic wave. An important prerequisite is the fact that the electromagnetic field of an ultrarelativistic electron is concentrated in a flat disc perpendicular to the direction of motion and is thus essentially transverse.

2.1 Electromagnetic field of a relativistic point charge

The electromagnetic field of a point charge q moving with constant speed $v = \beta c$ along the z direction can be determined by starting with the 4-vector potential in the particle rest frame

$$A'^{\mu} = \left(\frac{\Phi'}{c}, \, \boldsymbol{A'} \right).$$

Here Φ' is the scalar potential and $\boldsymbol{A'}$ is the vector potential. In the rest frame there is only a scalar potential, the vector potential vanishes

$$\Phi' = \frac{q}{4\pi\varepsilon_0 r'} \, , \quad \boldsymbol{A'} = (0, \, 0, \, 0) \;\; \Rightarrow \;\; A'^{\mu} = \left(\frac{\Phi'}{c}, \, 0, \, 0, \, 0 \right). \tag{1}$$

Now we carry out a Lorentz transformation into the laboratory system.

$$\Phi = \gamma(\Phi' + v\,A'_z) = \gamma\Phi' \, , \quad A_z = \gamma\left(A'_z + \frac{v}{c^2}\Phi' \right) = \frac{\gamma\,v}{c^2}\Phi' \quad \text{with} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \, . \tag{2}$$

The transverse components remain invariant: $A_x = A'_x = 0$, $A_y = A'_y = 0$. For a charged particle, moving with constant speed on a straight line, there is a simple connection between scalar and vector potential

$$\boldsymbol{A} = \frac{\boldsymbol{v}}{c^2}\,\Phi \, . \tag{3}$$

The electric and magnetic fields are computed using the Maxwell equations.

$$\boldsymbol{E}(r,\theta) = \frac{q}{4\pi\varepsilon_0 r^2} \cdot \frac{(1-\beta^2)}{(1-\beta^2\sin^2\theta)^{3/2}} \cdot \frac{\boldsymbol{r}}{r}, \quad \boldsymbol{B}(r,\theta) = \frac{1}{c^2}\,\boldsymbol{v}\times\boldsymbol{E}(r,\theta) \tag{4}$$

where θ is the angle between the z axis and the vector $\boldsymbol{r}=(x,y,z)$ and $\beta = v/c$. Details of the computation can be found in Refs. [15, 16]. The polar angle distributions of the electric field vector of a positron which is either at rest or is moving with $v = 0.87\,c$ ($\gamma = 2$) are plotted in Fig. 3. (For convenience we plot here the electric field lines of a positive charge. For an electron the field vectors point inwards and the arrow heads would be hardly visible).
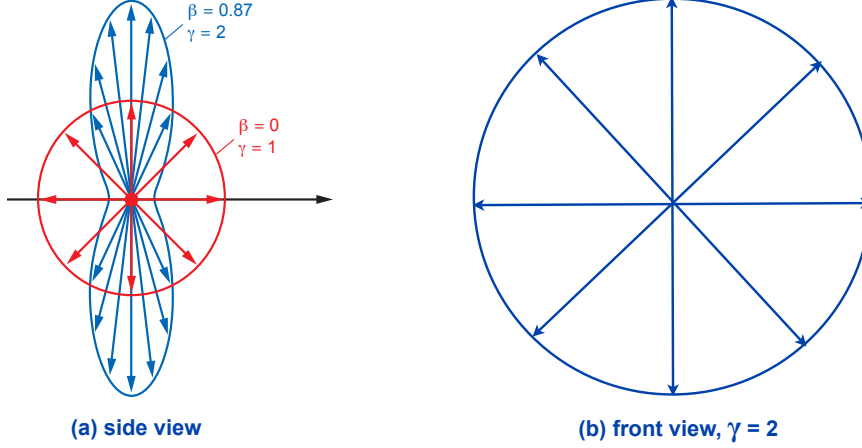

(a) side view   **(b) front view, γ = 2**

Figure 3: **(a)** Polar angle distributions of the electric field of a positron at rest (red) and a positron moving with $v = 0.87\,c$ (blue). **(b)** Field line pattern viewed along the direction of motion for the case $v = 0.87\,c$, showing the radial polarization.

Specifically, the field components parallel and perpendicular to the particle velocity $\boldsymbol{v}$ are

$$\begin{aligned}
\left|\boldsymbol{E}_\parallel\right| &= \frac{q}{4\pi\varepsilon_0 r^2}\cdot(1-\beta^2) = \frac{q}{4\pi\varepsilon_0 r^2}\cdot\frac{1}{\gamma^2} \quad \text{for } (\theta = 0)\,, \\
\left|\boldsymbol{E}_\perp\right| &= \frac{q}{4\pi\varepsilon_0 r^2}\cdot\frac{1}{\sqrt{1-\beta^2}} = \frac{q}{4\pi\varepsilon_0 r^2}\cdot\gamma \quad \text{for } (\theta = \pi/2)\,.
\end{aligned} \tag{5}$$

With increasing Lorentz factor γ there is a rapidly increasing anisotropy, the transverse field component grows with γ while the longitudinal component drops as $1/\gamma^2$. For electron energies in the GeV range the field is almost completely transverse.

2.2 Spectral energy of transition radiation in the backward hemisphere

For the special case that an electron passes from vacuum into a metal, only backward radiation is emitted at the interface since electromagnetic waves cannot propagate inside the metal. The generation of backward TR is shown schematically in Fig. 4a. A characteristic feature of transition radiation is its radial polarization (see Fig. 3b) which is very different from the well-known linear, circular or elliptic polarization of a laser beam.
In case of an infinite planar boundary, the spectral energy density of transition radiation, emitted into the backward hemisphere, is given by the Ginzburg-Frank formula

$$\left[\frac{d^2 U}{d\omega d\Omega}\right]_{\text{GF}} = \frac{e^2}{4\pi^3\varepsilon_0 c} \cdot \frac{\beta^2\sin^2\theta}{(1-\beta^2\cos^2\theta)^2} \tag{6}$$

with $\beta = v/c$ and θ the angle against the backward direction. For a derivation of this formula we refer to Landau-Lifshitz [17], see also [12]. Note that the Ginzburg-Frank formula is only valid if the radiation is
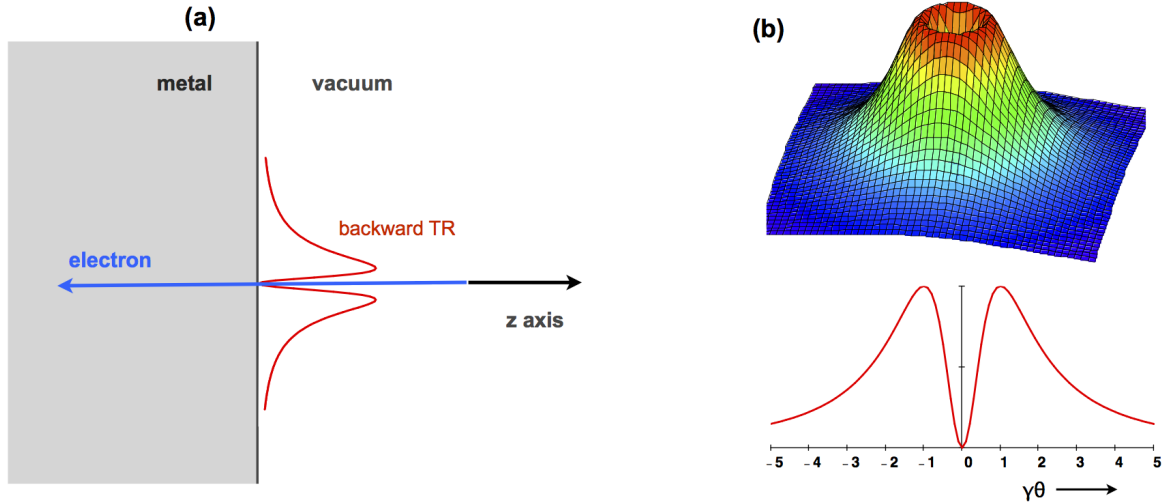
Figure 4: **(a)** Schematic view of the generation of backward transition radiation. A relativistic electron moves in negative z direction and crosses the interface between vacuum and a metal. Transition radiation is emitted into the positive z hemisphere. **(b)** The spectral energy density of transition radiation as a function of the scaled angle $\gamma\theta$ according to the Ginzburg-Frank formula. The maximum occurs at $\gamma\theta = 1$. Note that the intensity vanishes at $\theta = 0$ (exact backward direction with respect to the electron velocity). This is due to the radial polarization of TR, depicted in Fig. 3.

observed in the far-field (Fraunhofer) diffraction regime. The angular distribution is shown in Fig. 4b. The intensity vanishes in the exact backward direction at $\theta = 0$, which is a consequence of the radial polarization. The angular distribution has its maximum at the angle

$$\theta_{\mathrm{m}} = \frac{1}{\gamma} = \sqrt{1 - \beta^2} = \frac{m_e c^2}{E_e} \tag{7}$$

but extends to significantly larger angles. For an infinite screen, the spectral TR energy density (6) does not depend on the circular frequency[2] $\omega = 2\pi f$, provided one measures in the far-field and stays well below the plasma frequency $\omega_p = \sqrt{n_e e^2/(\varepsilon_0 m_e)}$ of the metal, which is in the ultraviolet. In the next section we will show that for a finite TR screen the radiation energy acquires an ω dependence and its angular distribution is widened.

2.3 Generalizations of the Ginzburg-Frank formula

The Ginzburg-Frank formula is not applicable in most practical cases because two basic conditions of the analytic derivation may not be fulfilled: (a) the radiation screens used in an accelerator are of limited size, and (b) the radiation is usually observed in the near-field and not in the far-field diffraction regime. To construct a generalization of the Ginzburg-Frank formula we make use of the fact that the electric field of an ultrarelativistic particle is predominantly perpendicular to the direction of motion, see Eq. (5). The field resembles closely that of an electromagnetic wave propagating in vacuum. This is the reason why the Weizsäcker-Williams method of virtual quanta can be utilized for computing backward TR from a metallic screen: by reflection at the screen, the virtual photons of the particle's self-field are converted to the real photons of a backward-moving electromagnetic wave. The virtual-photon method would completely fail for non-relativistic particles which have large longitudinal field components.

The virtual-photon method has been used by us [15] to compute the wave propagating in backward direction for radiation screens of arbitrary size, both in the far-field and in the near-field diffraction regimes. The transverse

---

[2]In the following ω is called "frequency" for short.

electric field component of a highly relativistic electron $(q = -e)$ moving along the z axis is [15], [16]

$$E_\perp(\rho, z, t) = -\gamma \frac{e}{4\pi\varepsilon_0} \frac{\rho}{(\rho^2 + \gamma^2(z - \beta ct)^2)^{3/2}} \quad \text{with} \quad \rho = \sqrt{x^2 + y^2}. \tag{8}$$

The field depends on the distance ρ from the axis but not on the azimuthal angle. When the relativistic electron passes by an observer at a small distance its transverse electric field appears as a very short time pulse. The Fourier transform of the transient field is derived in [15], see also Jackson [14]:

$$\tilde{E}_\perp(\rho, \omega) = \frac{-e\,\omega}{(2\pi)^{3/2}\varepsilon_0\gamma\beta^2 c^2} K_1\left(\frac{\omega\rho}{\gamma\beta c}\right). \tag{9}$$

The function K_1 is a modified Bessel function.

To preserve cylindrical symmetry the TR screen is chosen to be a circular disc of radius a which is centered with respect to the z axis. On this screen we use cylindrical coordinates (ρ, φ). The radiation is detected on a remote observation screen. Because of the cylindrical symmetry of the emitted transition radiation we can restrict
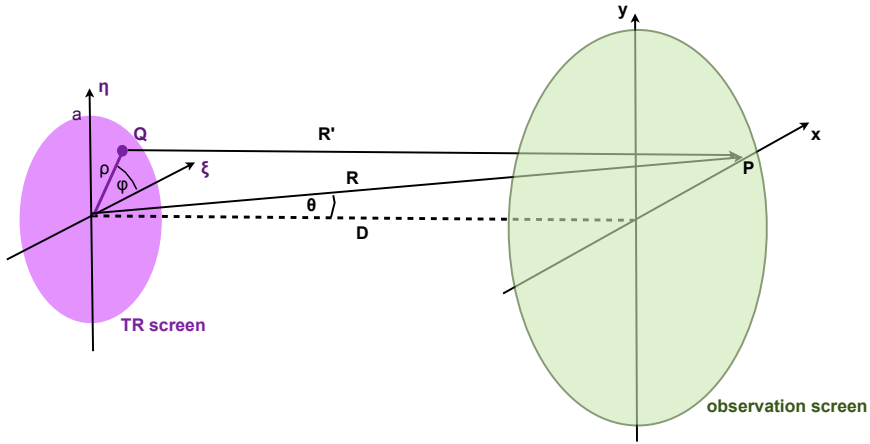

Figure 5: Diffraction geometry for a circular TR screen of radius a and an observation screen at large distance $D \gg a$.

ourselves to points on the x axis of the observation screen, hence the coordinates of our observation point are $P = (x, 0, D)$. The distance between P and the center of the TR screen is $R = \sqrt{D^2 + x^2}$, while the distance between P and an arbitrary point $Q = (\rho, \phi, 0)$ on the TR screen is $R' = \sqrt{D^2 + (x - \rho\cos\phi)^2 + (\rho\sin\phi)^2}$. The radius a of the TR screen is assumed to be much smaller than R, hence $\rho \leq a \ll R$ and the square root can be expanded into a Taylor series:

$$R' = \sqrt{D^2 + (x - \rho\cos\phi)^2 + (\rho\sin\phi)^2} \approx R\left(1 - \left(\frac{\rho}{R}\right)\frac{x\cos\phi}{R} + \frac{1}{2}\left(\frac{\rho}{R}\right)^2\right) \quad \text{with} \quad R = \sqrt{D^2 + x^2}. \tag{10}$$

The term which is linear in (ρ/R) describes far-field diffraction, the quadratic term $(\rho/R)^2$ accounts for the additional near-field diffraction effects.

Far-field diffraction
In the far-field regime, the Fourier-transformed electric field on the observation screen can be computed analytically [15]. We express it here as a function of the wave number $k = \omega/c$.

$$\left[\tilde{E}(\theta, k)\right]_{\text{far}} = \frac{e}{(2\pi)^{3/2}\varepsilon_0\, c} \frac{\exp(ikR)}{R} \frac{\beta\sin\theta}{1 - \beta^2\cos^2\theta}[1 - T(\theta, k)] \tag{11}$$

with a correction term that accounts for the finite TR screen radius a:

$$T(\theta, k) = \frac{ka}{\beta\gamma} J_0(ka\sin\theta)K_1\left(\frac{ka}{\beta\gamma}\right) + \frac{ka}{\beta^2\gamma^2\sin\theta} J_1(ka\sin\theta) K_0\left(\frac{ka}{\beta\gamma}\right). \tag{12}$$

6

The resulting spectral energy as a function of the angle θ is [15]

$$\left[\frac{d^2U}{d\omega d\Omega}\right]_{\text{far}} = \frac{e^2 k^4}{4\pi^3 \varepsilon_0 c\beta^4\gamma^2} \left|\int_0^a J_1(k\rho\sin\theta)K_1\left(\frac{k\rho}{\beta\gamma}\right)\rho\,d\rho\right|^2. \tag{13}$$

The integration can be done analytically and yields the far-field generalization of the Ginzburg-Frank formula for a circular radiation screen of finite radius a

$$\left[\frac{d^2U}{d\omega d\Omega}\right]_{\text{far}} = \left[\frac{d^2U}{d\omega d\Omega}\right]_{\text{GF}}\cdot[1-T(\theta,k)]^2 \qquad (\omega = k\,c). \tag{14}$$

The correction term $T(\theta,k)$ vanishes for $a \to \infty$, so this formula reduces to the standard Ginzburg-Frank formula (6) for a sufficiently large TR screen. Rectangular or other screen shapes can be treated with the Fourier-transform algorithm explained in Appendix D.

Near-field diffraction

In order to cover also the near-field we have to include the second-order term in Eq. (10). The angular dependence of the spectral energy is then given by

$$\left[\frac{d^2U}{d\omega d\Omega}\right]_{\text{near}} = \frac{e^2 k^4}{4\pi^3 \varepsilon_0 c\beta^4\gamma^2} \left|\int_0^a J_1(k\rho\sin\theta)K_1\left(\frac{k\rho}{\beta\gamma}\right)\exp\left(\frac{ik\,\rho^2}{2R}\right)\rho\,d\rho\right|^2 \tag{15}$$

This is the near-field generalization of the Ginzburg-Frank formula for a circular radiation screen. The difference to (13) is the extra phase factor $\exp(ik\rho^2/(2R))$. The integral in (15) must be evaluated numerically.

Effective source size and far-field condition

When the disc radius a is large and the observation screen very far away one should expect that the Ginzburg-Frank angular distribution is recovered. This is indeed the case. The question is, how large the radius has to be. It turns out that the answer depends on the wavelength and the Lorentz factor. Following Castellano et al. [18] we define an *effective source radius* by

$$r_{\text{eff}} = \gamma\lambda\,. \tag{16}$$

The first condition for obtaining the Ginzburg-Frank angular distribution is that the TR source radius has to exceed the effective source radius

$$r_{\text{source}} \equiv a \geq r_{\text{eff}} = \gamma\lambda\,. \tag{17}$$

Quantitatively, we can understand the effective source size condition as follows. We rewrite the correction term (12) using scaled variables $\xi = a/r_{\text{eff}}$ and $\theta_s = \gamma\theta$ and restricting ourselves to small angles:

$$T(\theta_s,\xi) = 2\pi\,\xi\left(\frac{1}{\theta_s}K_0(2\pi\,\xi)J_1(2\pi\,\xi\,\theta_s) + K_1(2\pi\,\xi)J_0(2\pi\,\xi\,\theta_s)\right) \tag{18}$$

The factor $(1-T(\theta_s,\xi))^2$ is plotted in Fig. 6 as a function of the scaled angle $\theta_s = \gamma\theta$ for two screen radii: $a/r_{\text{eff}} = 1$ and $a/r_{\text{eff}} = 0.5$. It is obvious that the reduction of the spectral energy caused by the finite TR screen size is small for $a/r_{\text{eff}} \geq 1$ but becomes very significant for $a/r_{\text{eff}} \leq 0.5$.

The second condition for obtaining the Ginzburg-Frank angular distribution is to have far-field diffraction, which requires

$$D \gg \gamma\,r_{\text{eff}} = \gamma^2\lambda\,. \tag{19}$$

This inequality follows from formula (10). In the far-field, the quadratic term $(\rho/R)^2$ must be much smaller than the term which is linear in (ρ/R). Assuming that the source size criterion is fulfilled, then $\rho_{\max} = a = r_{\text{eff}} = \gamma\lambda$, and with $R \approx D$ and $x \approx D/\gamma$ for a typical point on the observation screen one gets the inequality (19). In fact, when (19) is fulfilled, a numerical evaluation yields an almost perfect agreement between the formulas (14) and (15).

Significant differences arise, however, when one of these conditions is violated. In Fig. 7 we show several examples. When the far-field condition is satisfied but the source-size condition is violated, the formulas (14)
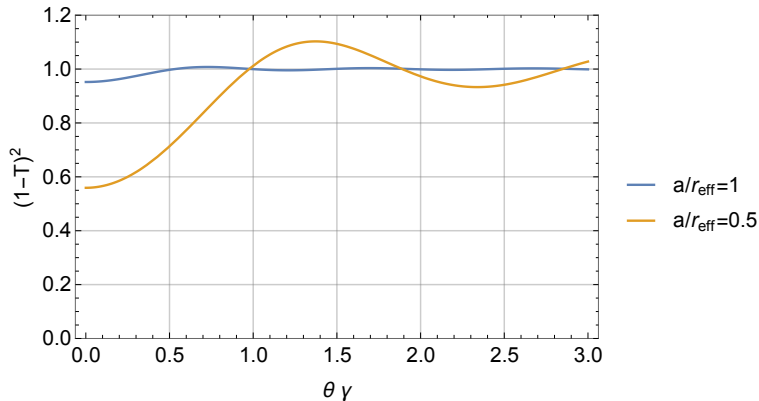
Figure 6: The factor $(1 - T(\theta_s, \xi))^2$, plotted as a function of the scaled angle $\theta_s = \gamma\theta$, for two values of the scaled disc radius $\xi = a/r_{\text{eff}}$.
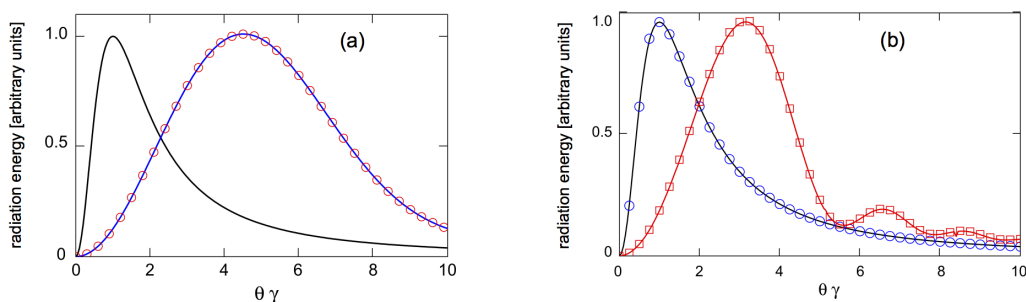

Figure 7: **(a)** Far-field TR from a circular disc of radius $a = 3\,\text{mm}$ for $\lambda = 0.3$ mm, $\gamma = 100$, $r_{\text{eff}} = 30\,\text{mm}$, $D = 4$ m. The far-field criterion (19) is satisfied: $D > \gamma r_{\text{eff}} = 3$ m. However, the effective source-size criterion is badly violated, $a \ll r_{\text{eff}}$. Black curve: Ginzburg-Frank formula; blue curve: far-field computation using Eq. (14); red circles: numerical near-field computation using Eq. (15). The distributions are individually normalized to a maximum value of 1.
(b) Near-field TR from a circular disc of radius $a = 30\,\text{mm}$ for $\lambda = 0.3$ mm, $\gamma = 100$, $r_{\text{eff}} = 30\,\text{mm}$, $D = 0.2\,\text{m}$. The effective source-size criterion is satisfied but the far-field condition is strongly violated, $D \ll \gamma r_{\text{eff}} = 3$ m. Black curve: Ginzburg-Frank formula; blue circles: far-field prediction (14); red curve: near-field prediction (15); red squares: numerical calculation using the exact square root expression for R'.

and (15) are in agreement but both predict a wider angular distribution than the Ginzburg-Frank formula (6). When the source-size condition is satisfied but the far-field condition is violated, the far-field formula (14) yields the same angular distribution as the Ginzburg-Frank formula (6) but the near-field formula (15) yields a wider distribution.

For the typical electron Lorentz factors at FLASH of $\gamma > 1000$ the effective source-size and the far-field conditions are both violated except at very small wavelengths in the few μm range. Hence Eq. (15) must be used to compute the spectral energy entering a detector with aperture angle θ_{ap}

$$U_{\text{det}}(\omega) = \int_0^{\theta_{\text{ap}}} \left[\frac{d^2U}{d\omega d\Omega}\right]_{\text{near}} 2\pi \sin\theta\, d\theta\,. \tag{20}$$

It is very important to realize that only a small fraction of the TR energy is emitted at very small angles, $\theta \leq \theta_{\text{m}}$. Hence for intensity reasons the aperture angle of a TR detector is often chosen to be much larger than $\theta_{\text{m}} = 1/\gamma$. Thereby one accepts near-field transition radiation which has a wide tail towards larger angles and, more importantly, the intensity observed at these larger angles is enhanced by the growing solid angle $d\Omega = 2\pi \sin\theta\, d\theta$. The radiation energy entering the detector continues to grow with increasing aperture angle even beyond $\theta_{\text{ap}} \geq 100\,\theta_{\text{m}}$.

8

3 Electron Bunch Shape Reconstruction using Coherent Transition Radiation

Consider an electron bunch as sketched in Fig. 8. We want to determine the transition radiation produced by the $N \gg 1$ particles in the bunch as a function of frequency ω and emission angle θ. The spectral energy receives contributions from incoherent and coherent radiation.

Incoherence means that interference terms average to zero because of random phase relations, and hence it is permitted to add probabilities. The incoherent spectral energy density produced by a bunch of N electrons is simply N times the spectral energy density produced by one electron:

$$\left[\frac{d^2 U}{d\omega d\Omega} \right]_{\text{incoh}} = N \left[\frac{d^2 U}{d\omega d\Omega} \right]_1. \tag{21}$$

Incoherent transition radiation in the visible range is very useful for transverse beam diagnostics (e.g. emittance measurements) but is not suited for determining details of the longitudinal bunch structure.

Coherence means that one has to add complex amplitudes. The absolute square of the sum amplitude yields the probability, and when multiplying this amplitude with its complex conjugate, interference terms come in. Full coherence means that the radiation fields of all N electrons add constructively. The intensity I_N is then N^2 times the intensity I_1 emitted by a single electron. Usually, however, there is partly constructive and partly destructive interference, and in that case $I_N < N^2 I_1$. The form factor, or more accurately its absolute square, is a measure of the degree of constructive interference.

3.1 Form factor

3.1.1 Three-dimensional form factor

As said above, coherence means that complex amplitudes have to be added. This will be done now. To compute coherent transition radiation (CTR) we place a "reference electron" at the bunch center and label it with the index "1". The radiation field of this single electron is given by Eq. (11), we rewrite it in the form

$$\tilde{E}_1(\boldsymbol{k}) = \frac{e}{(2\pi)^{3/2}\varepsilon_0 c} \frac{\exp(ikR)}{R} \frac{\beta \sin\theta}{1 - \beta^2 \cos^2\theta} [1 - T(\theta, k)] \tag{22}$$

with the wave vector

$$\boldsymbol{k} = \frac{\omega}{c}(\sin\theta, 0, \cos\theta), \quad k = |\boldsymbol{k}| = \omega c = \frac{2\pi}{\lambda}. \tag{23}$$

An arbitrary electron "n" at a position $\boldsymbol{r}_n$ inside the bunch produces a field of the same mathematical form, but since it crosses the TR screen at a different time and a different position, there will be a phase shift $\Delta\varphi_n = \boldsymbol{k} \cdot \boldsymbol{r}_n$ with respect to the reference electron, see Fig. 8. The total field in $\boldsymbol{k}$ direction is obtained by summing over all
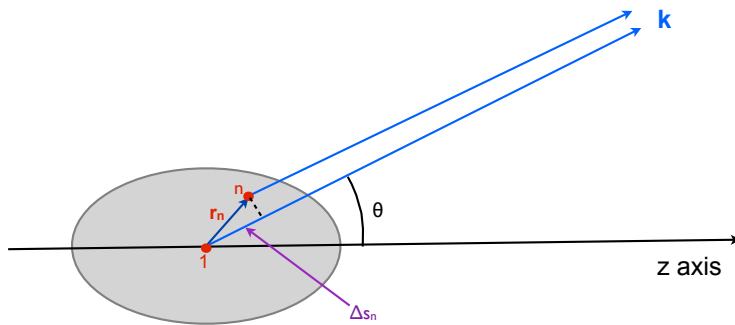

Figure 8: Computation of the phase shift between the waves produced by an arbitrary electron at $\boldsymbol{r}_n$ and the reference electron at $\boldsymbol{r}_1 = 0$. The distances to a far remote observation point differ by $\Delta s_n = s_1 - s_n = \boldsymbol{k} \cdot \boldsymbol{r}_n / k$, the phases differ by $\Delta\varphi_n = \Delta s_n (2\pi/\lambda) = \boldsymbol{k} \cdot \boldsymbol{r}_n$.

electrons in the bunch

$$\tilde{E}_{\text{tot}}(\boldsymbol{k}) = \tilde{E}_1(\boldsymbol{k}) \sum_{n=1}^{N} \exp(i\,\boldsymbol{k} \cdot \boldsymbol{r}_n). $$

A typical electron bunch consists of some 10^9 electrons, and it is useful to replace the discrete distribution of N point particles by a continuous particle density $\rho(\boldsymbol{r})$ which we normalize to 1. Hence the total field can be written as

$$\tilde{E}_{\text{tot}}(\boldsymbol{k}) = N\,\tilde{E}_1(\boldsymbol{k}) \int \rho(\boldsymbol{r})\exp(\,i\,\boldsymbol{k}\cdot\boldsymbol{r})\,d^3r \equiv N\,\tilde{E}_1(\boldsymbol{k})\tilde{F}_{\text{3D}}(\boldsymbol{k}) \tag{24}$$

where we have defined the three-dimensional bunch form factor

$$\tilde{F}_{\text{3D}}(\boldsymbol{k}) = \int \rho(\boldsymbol{r})\exp(\,i\,\boldsymbol{k}\cdot\boldsymbol{r})\,d^3r \quad\text{with}\quad \int \rho(\boldsymbol{r})\,d^3r = 1\,. \tag{25}$$

From this equation follows immediately

$$\tilde{F}_{\text{3D}}(0) = 1\,.$$

The Fourier back transformation reads

$$\rho(\boldsymbol{r}) = \frac{1}{(2\pi)^3}\int \tilde{F}_{3D}(\boldsymbol{k})\exp(-i\,\boldsymbol{k}\cdot\boldsymbol{r})\,d^3k\,. \tag{26}$$

The coherent spectral energy density produced by a bunch of N electrons is the product of the spectral energy density produced by a single electron, the number of electrons squared and the absolute square of the form factor:

$$\left[\frac{d^2U}{d\omega d\Omega}\right]_{\text{coh}} = N^2\left[\frac{d^2U}{d\omega d\Omega}\right]_1 |\tilde{F}_{\text{3D}}(\boldsymbol{k})|^2. \tag{27}$$

In the following treatment we make the simplifying assumption that the transverse density distribution of the electron bunch is independent of the longitudinal position in the bunch. In other words, we assume that the slice emittance and other beam parameters are constant along the bunch. Then the 3D particle density can be factorized:

$$\rho(x,y,z) = \rho_{\text{trans}}(x,y)\,\rho_{\text{long}}(z) \tag{28}$$

and the 3D form factor is the product of the transverse and the longitudinal form factors:

$$\tilde{F}_{\text{3D}}(\boldsymbol{k}) = \tilde{F}_{\text{trans}}(k_x,k_y)\,\tilde{F}_{\text{long}}(k_z)\,. \tag{29}$$

In reality, the electron bunches in FLASH are affected by nonlinear effects in the magnetic bunch compressor and acquire a slice emittance that varies along the bunch axis. The considerations in the next section show that a variable slice emittance has a rather small impact on the observable CTR spectra. The effect will be ignored here.

3.1.2 Transverse form factor

The transverse charge density distribution is assumed to be a cylindrically symmetric Gaussian. The normalized distribution is written in the form

$$\rho_{\text{trans}}(x,y) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{x^2}{2\,\sigma^2}\right)\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{y^2}{2\,\sigma^2}\right).$$

The transverse form factor of the two-dimensional charge distribution is defined by the equation

$$\tilde{F}_{\text{trans}}(k_x,k_y) = \iint \rho_{\text{trans}}(x,y)\,\exp(i\,[k_x x + k_y y])\,dxdy\,.$$

Because of the cylindrical symmetry we can assume without loss of generality that the wave vector $\boldsymbol{k}$ is located in the (x,z) plane. Then

$$\boldsymbol{k} = k(\sin\theta, 0, \cos\theta)\,,\quad k_x x + k_y y = k\,x\,\sin\theta\,.$$

Using these relations the transverse form factor can be written as a function of wave number k and emission angle θ

$$\tilde{F}_{\text{trans}}(k,\theta) = \exp\left(-\frac{k^2\sigma^2\sin^2\theta}{2}\right). \tag{30}$$

Backward TR is confined to a fairly narrow cone around the z axis (a typical aperture angle of the detector is 100 mrad). CTR spectroscopy is therefore not suited for determining the transverse density distribution in the bunch. At large wavelengths (small wave numbers) the product $k\,\sigma\sin\theta = 2\pi\sin\theta\,\sigma/\lambda$ is close to zero, and the transverse form factor is close to 1. However, at small wavelengths ($\lambda < \sigma$) there is considerable destructive interference and the transverse form factor may drop to small values. This reduction depends on the aperture angle of the spectrometer. To get an impression we use the following typical electron beam parameters at the position of the CRISP spectrometer in the FLASH linac:

Lorentz factor $\gamma = 1500$, normalized emittance $\varepsilon_n \approx 2\,\mu$m, beta functions $\beta_x = \beta_y = 7$ m, $\sigma \approx 100\,\mu$m.

The square of the transverse form factor must be averaged over the aperture of the spectrometer, with a weight factor given by the angular-dependent radiation energy density. This average depends on the wave number $k = 2\pi/\lambda$, the rms electron beam radius σ, and the aperture angle θ_{ap}:

$$\langle|\tilde{F}_{\mathrm{trans}}|^2\rangle(k,\sigma,\theta_{\mathrm{ap}}) = \frac{\int_0^{\theta_{\mathrm{ap}}} \mathcal{U}(\theta,k)\,\exp\left(-k^2\sigma^2\sin^2\theta\right)\,\sin\theta\,d\theta}{\int_0^{\theta_{\mathrm{ap}}} \mathcal{U}(\theta,k)\,\sin\theta\,d\theta} \tag{31}$$

where $\mathcal{U}(\theta,k)$ stands for the near-field radiation energy density (15), which is evaluated here for the case of a single aperture at a distance of $D = 1$ m. A more accurate treatment, based on a *THzTransport* simulation of the radiation transport from the TR screen through the CTR beamline to the multichannel spectrometer, will be presented in Section 4.

The root-mean-square value of the transverse form factor is the square root of expression (31):

$$\langle\tilde{F}_{\mathrm{trans}}\rangle(k) = \sqrt{\langle|\tilde{F}_{\mathrm{trans}}|^2\rangle(k,\sigma,\theta_{\mathrm{ap}})} \ . \tag{32}$$

The rms transverse form factor $\langle\tilde{F}_{\mathrm{trans}}\rangle(k)$ depends implicitly on the parameters σ and θ_{ap}, which are fixed quantities in a given experimental setup, but these dependencies are not written down here to simplify the notation.

The impact of the spectrometer aperture angle θ_{ap} is depicted in Fig. 9a where $\langle\tilde{F}_{\mathrm{trans}}\rangle(k)$ is plotted versus $\lambda = 2\pi/k$ for a typical rms beam radius of $\sigma = 100\,\mu$m and various aperture angles. The impact of the rms beam radius is shown in Fig. 9b for an aperture angle of 100 mrad (this is roughly the aperture of our spectrometer setup). The suppression of small wavelengths by the transverse form factor is substantial. The
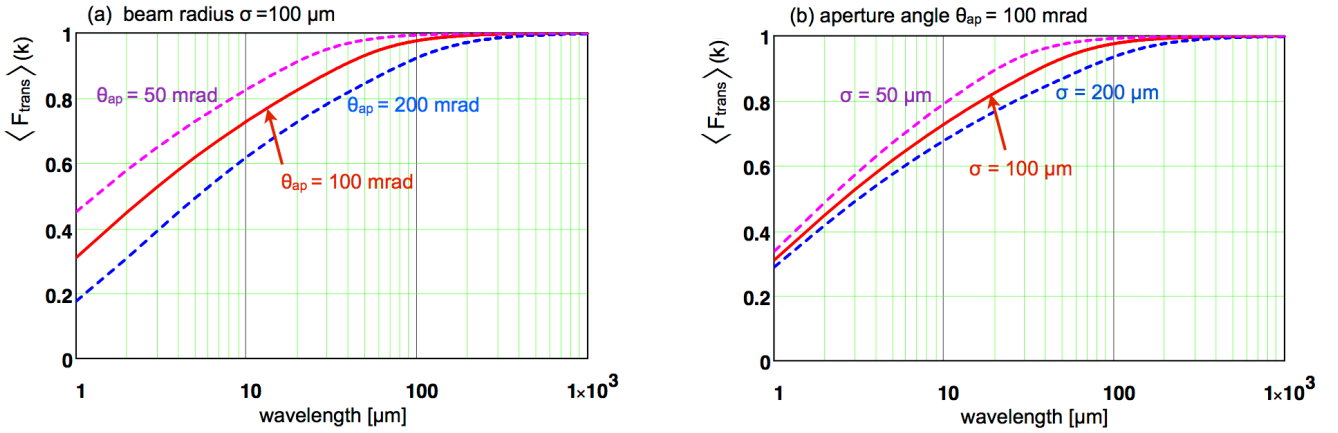

Figure 9: **(a)** The rms transverse form factor $\langle\tilde{F}_{\mathrm{trans}}\rangle(k)$, plotted as a function of wavelength $\lambda = 2\pi/k$, for aperture angles of $\theta_{\mathrm{ap}} = 50, 100, 200$ mrad and a fixed rms electron beam radius $\sigma = 100\,\mu$m.
(b) The rms transverse form factor, plotted versus $\lambda = 2\pi/k$, for a fixed aperture angle of $\theta_{\mathrm{ap}} = 100$ mrad and rms electron beam radii of $\sigma = 50, 100, 200\,\mu$m.

measured spectra have to be corrected for these losses. This will be done with the help of the response function defined in Section 4.

3.1.3 Longitudinal form factor

In the following we assume that the measured spectra have been corrected properly for the above-mentioned suppression effects at small wavelengths. The spectrometer aperture angle $\theta_{\mathrm{ap}} \approx 100\,\mathrm{mrad}$ is so small that $\cos\theta$ is almost equal to 1 in the angular range $0 \le \theta \le \theta_{\mathrm{ap}}$, hence $k_z = k\cos\theta$ can be replaced by $k = \omega/c$.

When an electron bunch crosses the TR source screen it generates a radiation pulse whose time duration τ is related to the bunch length ℓ by $\ell = v\tau = \beta c\tau$. For comparison with time-domain diagnostic instruments such as the TDS it is advantageous to express the longitudinal density distribution $\rho_{\mathrm{long}}(z)$ of the bunch as a function of time:

$$\rho(t) = \rho_{\mathrm{long}}(\beta c\, t)\,.$$

The longitudinal form factor can be written as a function of ω

$$\mathcal{F}(\omega) = \tilde{F}_{\mathrm{long}}(\omega/c)\,.$$

The subscript "long" is not needed anymore and will be dropped in the following.

The Fourier transformation relations between normalized longitudinal density distribution and longitudinal form factor are[3]

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} \rho(t)\exp(i\,\omega\,t)dt\,,\quad \rho(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathcal{F}(\omega)\exp(-i\,\omega\,t)d\omega \quad\text{with}\quad \int_{-\infty}^{\infty}\rho(t)dt = 1\,. \tag{33}$$

At low frequencies, namely when the wavelength of the radiation is long compared to the bunch length, the form factor of a bunch centered at $t = 0$ is a real number close to 1. All electrons radiate coherently which means that there is constructive interference among their radiation fields. Spectral measurements in this range yield no information on the internal charge distribution in the bunch. To gain such information, measurements at wavelengths significantly shorter than the bunch length have to be carried out. In that case the form factor becomes a complex-valued function

$$\mathcal{F}(\omega) = F(\omega)\,e^{i\,\Phi(\omega)} \tag{34}$$

whose magnitude $F(\omega) = |\mathcal{F}(\omega)|$ is generally less than 1. If both $F(\omega)$ and $\Phi(\omega)$ were known, a unique reconstruction of the charge distribution $\rho(t)$ could be achieved by the inverse Fourier transformation. Unfortunately only the spectral intensity is accessible in spectroscopic experiments at accelerators. Hence the modulus $F(\omega) = |\mathcal{F}(\omega)|$ of the longitudinal form factor can be determined while the phase $\Phi(\omega)$ remains unknown.

Some limited information is provided by the autocorrelation function which is the inverse Fourier transform of $|\mathcal{F}(\omega)|^2 = |F(\omega)|^2$ and thus a measurable quantity:

$$A(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}|\mathcal{F}(\omega)|^2\exp(-i\,\omega\,t)d\omega\,. \tag{35}$$

The autocorrelation function provides no information on the internal structure.

3.2 Ambiguities in bunch shape reconstruction from spectroscopic data

The normalized particle density $\rho(t)$ is a real function. Decomposing the form factor and its complex conjugate into real and imaginary part:

$$\mathcal{F}(\omega) = \int\rho(t)\cos(\omega t)\,dt + i\int\rho(t)\sin(\omega t)\,dt\,,\quad \mathcal{F}^*(\omega) = \int\rho(t)\cos(\omega t)\,dt - i\int\rho(t)\sin(\omega t)\,dt$$

reveals two important properties.
a) There exists a relation between the form factor and its complex conjugate: $\mathcal{F}(\omega) = \mathcal{F}^*(-\omega)$. In the next section we will introduce complex frequencies $\hat{\omega} = \omega_{\mathrm{r}} + i\,\omega_{\mathrm{i}}$. Then the relation between $\mathcal{F}$ and $\mathcal{F}^*$ can be written as

$$\mathcal{F}(\hat{\omega}) = \mathcal{F}^*(-\hat{\omega}^*)\,. \tag{36}$$

---

[3]The factors "1" in front of the Fourier integral and "$1/(2\pi)$" in front of the inverse Fourier integral are chosen such that the basic requirement $\mathcal{F}(0) = 1$ is fulfilled: at very low frequency (very large wavelength) the bunch acts as a point charge whose form factor must be unity.

b) The form factor is a real function if $\rho(t)$ is symmetric with respect to $t = 0$.

The fact that only the magnitude $F(\omega) = |\mathcal{F}(\omega)|$ of the form factor is measured but not its phase has undesirable (but unavoidable) consequences.

- Time reversal does not change $F(\omega)$. The time profiles $\rho(t)$ and $\rho(-t)$ yield the same spectrum: *Head and tail of the bunch cannot be distinguished.*

- A time shift $\rho(t) \to \rho(t + t_0)$ results in an extra phase factor $e^{i\omega t_0}$ but leaves $F(\omega)$ invariant: *The arrival time of the bunch at the TR screen cannot be measured.*

- The magnitude $F(\omega)$ does not uniquely specify the internal bunch structure (see below): *There exist many different bunch structures yielding exactly the same spectrum.*
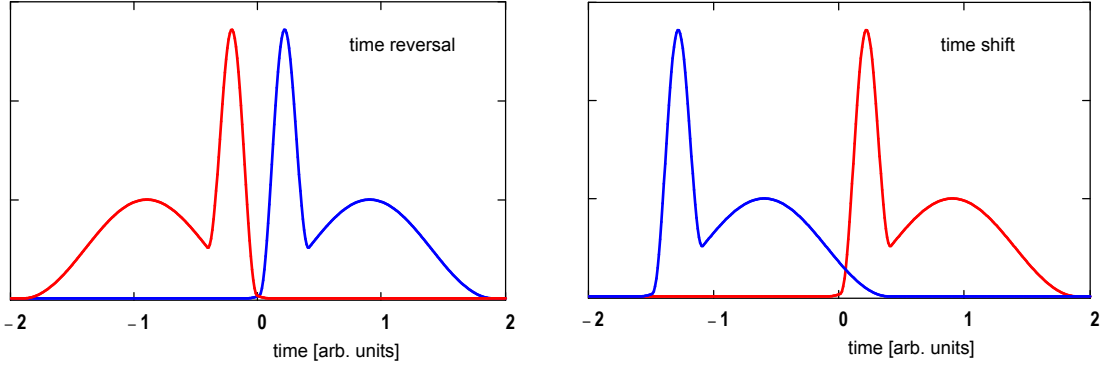
Figure 10: Time reversal or time shift of a bunch leave $F(\omega) = |\mathcal{F}(\omega)|$ invariant.

The determination of the phase $\Phi(\omega)$ is obviously of great importance. Two phase retrieval methods will be discussed in the following sections, but one has to be aware of a fundamental limitation: the unique reconstruction of a function from the magnitude of its Fourier transform is mathematically impossible in the one-dimensional case[4]. This was nicely demonstrated by Akutowicz [19]. He considered two functions $f_1(t)$ and $f_2(t)$ which vanish for $t < 0$ and which for $t \geq 0$ are given by

$$
\begin{aligned}
f_1(t) &= e^{-\beta t} \\
f_2(t) &= e^{-\beta t}\left(1 + \frac{4\beta^2(1 - \cos(\alpha t))}{\alpha^2} - \frac{4\beta \sin(\alpha t)}{\alpha}\right)
\end{aligned}
\tag{37}
$$

with real parameters $\alpha, \beta > 0$. The first function has an infinitely steep rise at $t = 0$, followed by an exponential decay. The second function has the same steep rise but the decay is superimposed with an oscillatory pattern, see Fig. 11. The complex Fourier transforms $\mathcal{F}_1(\hat{\omega})$ and $\mathcal{F}_2(\hat{\omega})$ differ considerably

$$
\mathcal{F}_1(\hat{\omega}) = \frac{1}{\beta - i\hat{\omega}}, \quad \mathcal{F}_2(\hat{\omega}) = \frac{\alpha^2 + (\beta + i\hat{\omega})^2}{[\alpha^2 + (\beta - i\hat{\omega})^2](\beta - i\hat{\omega})}
\tag{38}
$$

but their absolute magnitudes are identical on the real ω axis:

$$
|\mathcal{F}_1(\omega)| = |\mathcal{F}_2(\omega)| = \frac{1}{\sqrt{\beta^2 + \omega^2}} \quad \text{for real } \omega .
\tag{39}
$$

Therefore all three curves in Fig. 11 are permitted reconstructions. This example shows very clearly that nontrivial ambiguities in the bunch shape reconstruction are unavoidable and will occur in any phase retrieval method.

---

[4]In two or more dimensions the reconstruction is unique in the sense that the set of false reconstructions is a set of measure zero. For a proof see M.H. Hayes [20].
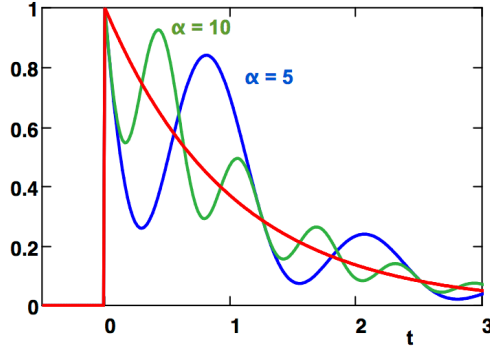
Figure 11: The functions $f_1(t)$ (red) and $f_2(t)$ for $\beta = 1$ and $\alpha = 5$ (blue) resp. $\alpha = 10$ (green).

3.3 Analytic phase retrieval

Kramers-Kronig phase

The problem that only the magnitude of a complex-valued quantity of interest is measurable but not its phase arises also for the optical reflection properties of solids. As shown in [21] it is possible to compute the phase of the complex reflectivity amplitude from the measured reflectivity by making use of the powerful mathematical theory of analytic (or holomorphic) functions, in particular by applying the Kramers-Kronig dispersion relation. The method has been adopted for the phase reconstruction of the complex bunch form factor, see [22, 23] and the references quoted therein. The *Kramers-Kronig phase* $\Phi_{KK}(\omega)$ can be computed from the real function $F(\omega) = |\mathcal{F}(\omega)|$ by means of the following principal-value integral[5]

$$\Phi_{KK}(\omega) = \frac{2\omega}{\pi} \, \mathcal{P} \int_0^\infty \frac{\ln(|\mathcal{F}(\omega')|) - \ln(|\mathcal{F}(\omega)|)}{\omega^2 - \omega'^2} \, d\omega' \,. \tag{40}$$

A rigorous mathematical derivation of formula (40), based on the theory of analytic functions, especially the Cauchy Integral Formula and the Residue Theorem, can be found in Appendix A. Here we indicate only a few steps. In analogy to the standard notation of complex numbers, $z = x + i\,y$, one defines a complex frequency by $\hat{\omega} = \omega_r + i\,\omega_i$. The form factor $\mathcal{F}(\omega)$, which is *a priori* a function of the real variable ω, is continued into the complex $\hat{\omega}$ plane, and $\mathcal{F}(\hat{\omega})$ can be shown to be an analytic function. The standard Kramers-Kronig dispersion relation between real and imaginary part (Eq. (66) in Appendix A) is not applicable here since only the magnitude $|\mathcal{F}(\omega)|$ is known from measurement. To separate magnitude and phase we compute the logarithm of expression (34) and insert the complex frequency

$$\ln(\mathcal{F}(\hat{\omega})) = \ln(F(\hat{\omega})) + i\,\Phi(\hat{\omega})\,.$$

Our aim is to find a relation between $\ln(F(\hat{\omega}))$ and $\Phi(\hat{\omega})$. This involves an integration around the closed loop depicted in Fig. 45 (see Appendix A) and the application of the Residue Theorem. A severe problem, however, is that the form factor drops to zero at infinite frequency, hence $\ln(\mathcal{F}(\hat{\omega}))$ diverges on the large semicircle Γ_1 in Fig. 45. To circumvent this difficulty, an auxiliary function, containing $\ln(\mathcal{F}(\hat{\omega}))$ as a factor, is constructed which can be integrated along the semicircle Γ_1. After many computational steps one arrives at Eq. (40).

As a first application of formula (40) we try to reconstruct the function $f_1(t)$. The KK method reproduces $f_1(t)$ indeed accurately, see Fig. 12, and also the phase is reproduced. However, any attempt to reconstruct the oscillatory function $f_2(t)$ will fail, no matter what the value of α is.

Blaschke phase

An essential prerequisite for Eq. (40) to hold is that the complex form factor does not have any zeros in the upper half of the complex $\hat{\omega}$ plane because otherwise $\ln(\mathcal{F}(\hat{\omega}))$ would have essential singularities at these points. In such a case the Residue Theorem is not applicable and formula (40) cannot be used to compute $\Phi(\omega)$. It was

---

[5]The principal value means that the singularity of the integrand at $\omega' = \omega$ is approached symmetrically from below and above, see Eq. (59) in Appendix A.
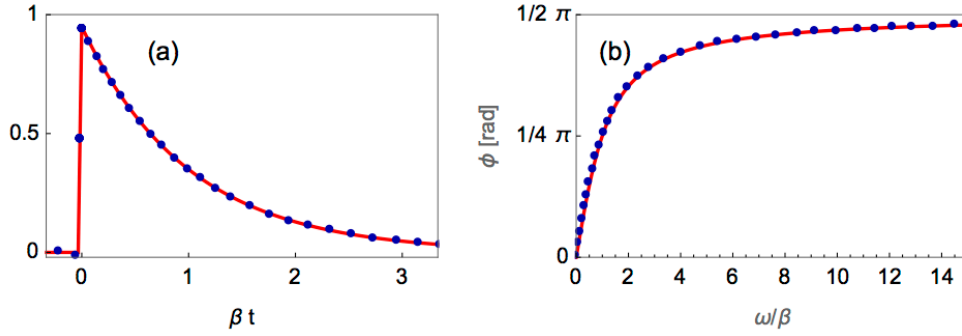
14

Figure 12: **(a)** The function $f_1(t)$ (red curve) and its Kramers-Kronig reconstruction (blue dots). **(b)** The phase $\Phi_1(\omega)$ of the complex form factor $\mathcal{F}_1(\omega)$ (red) and the Kramers-Kronig phase $\Phi_{\mathrm{KK}}(\omega)$ (blue dots).

proved by Blaschke [24] that another phase has to be taken into consideration. Suppose $\mathcal{F}(\hat{\omega})$ has a zero at the point $\hat{\omega}_1 = a_1 + i\,b_1$ in the right upper quarter of the complex plane, i.e. $\Re(\hat{\omega}_1) = a_1 > 0$ and $\Im(\hat{\omega}_1) = b_1 > 0$. Formula (36) shows that there is another zero in the left upper quarter at $\hat{\omega}'_1 = -a_1 + ib_1$. This pair of zeros can be removed by modifying $\mathcal{F}(\hat{\omega})$

$$\mathcal{F}_{\mathrm{mod}}(\hat{\omega}) = \mathcal{F}(\hat{\omega})\mathcal{B}(\hat{\omega}) \quad \text{with} \quad \mathcal{B}(\hat{\omega}) = \frac{\hat{\omega} - (a_1 - ib_1)}{\hat{\omega} - (a_1 + ib_1)} \cdot \frac{\hat{\omega} - (-a_1 - ib_1)}{\hat{\omega} - (-a_1 + ib_1)}. \tag{41}$$

Here $\mathcal{B}(\hat{\omega})$ is the so-called Blaschke factor.
On the real ω axis the absolute magnitude of the Blaschke factor is 1, hence

$$|\mathcal{F}_{\mathrm{mod}}(\omega)| = |\mathcal{F}(\omega)| \quad \text{for real} \quad \omega. \tag{42}$$

This is a very important equation. It means that the form factor $\mathcal{F}(\omega)$ and the modified form factor $\mathcal{F}_{\mathrm{mod}}(\omega)$ describe exactly the same radiation spectrum. The phase of $\mathcal{B}(\omega)$ is computed by the equation

$$\Phi_{\mathrm{B}}(\omega) = \arg(\mathcal{B}(\omega)). \tag{43}$$

This procedure is repeated for every zero of $\mathcal{F}(\hat{\omega})$ until $\mathcal{F}_{\mathrm{mod}}(\hat{\omega})$ is free from any zeros. Then Eq. (40) is valid for the modified form factor, so $\Phi_{\mathrm{mod}}(\omega) = \Phi_{\mathrm{KK}}(\omega)$. The reconstruction phase is given by the difference between KK phase and Blaschke phase

$$\Phi_{\mathrm{rec}}(\omega) = \Phi_{\mathrm{KK}}(\omega) - \Phi_{\mathrm{B}}(\omega). \tag{44}$$

Now we demonstrate that the Blaschke phase in combination with the KK phase enables a faithful reconstruction of $f_2(t)$. To this end we have to find the complex zeros of the form factor $\mathcal{F}_2(\hat{\omega})$, see Eq. (38). This is easy: the numerator of $\mathcal{F}_2(\hat{\omega})$ must vanish if we insert the complex frequency $\hat{\omega}_1 = a + ib$:

$$0 = \alpha^2 + (\beta + i\hat{\omega}_1)^2 = \alpha^2 + (\beta + ia - b)^2 = \alpha^2 + \beta^2 - a^2 + b^2 - 2b\beta + i\,2a(\beta - b).$$

Both real and imaginary part of this equation must be zero. From this condition follows immediately

$$b = \beta, \quad a = \pm\alpha.$$

The form factor $\mathcal{F}_2(\hat{\omega})$ has just one pair of zeros in the upper half plane: $\hat{\omega}_1 = \alpha + i\beta$ in the right upper quarter of the complex $\hat{\omega}$ plane and its mirror image $\hat{\omega}'_1 = -\alpha + i\beta$ in the left upper quarter. We consider the function $f_2(t)$ with the parameters $\alpha = 5$ and $\beta = 1$. Using the Eqs. (41) and (43) we compute the Blaschke phase $\Phi_{\mathrm{B}}(\omega)$. The reconstruction phase $\Phi_{\mathrm{rec}}(\omega) = \Phi_{\mathrm{KK}}(\omega) - \Phi_{\mathrm{B}}(\omega)$ is found to be identical with the analytic phase $\Phi_2(\omega)$ of the form factor $\mathcal{F}_2(\omega)$, see Fig. 13b. When we use this reconstruction phase to compute the function $f_2(t)$ from the magnitude $|\mathcal{F}_2(\omega)|$ of the form factor we find perfect agreement with the original function (Fig. 13a). This result is a remarkable success of the dispersion relation theory.
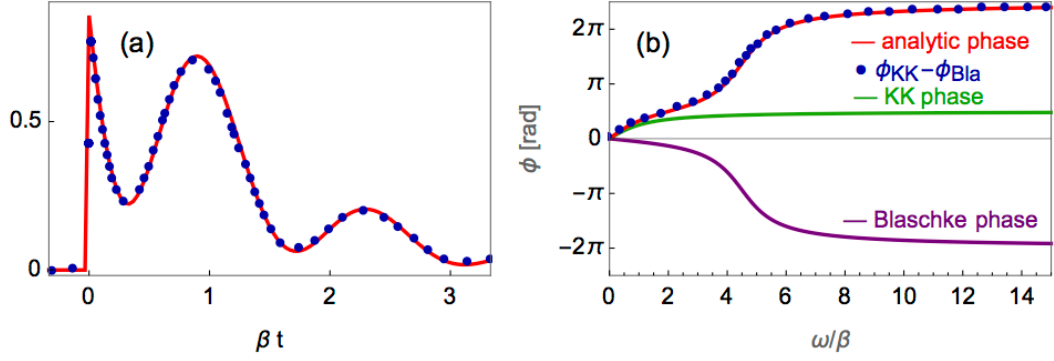
15

Figure 13: **(a)** The function $f_2(t)$ (red curve) and its reconstruction (blue dots) using the reconstruction phase $\Phi_{\text{rec}}(\omega) = \Phi_{\text{KK}}(\omega) - \Phi_{\text{B}}(\omega)$. **(b)** The analytic phase $\Phi_2(\omega)$ of the complex form factor $\mathcal{F}_2(\omega)$ (red), the KK phase $\Phi_{\text{KK}}(\omega)$ (green), the Blaschke phase $\Phi_{\text{B}}(\omega)$ (purple) and the reconstruction phase $\Phi_{\text{rec}}(\omega) = \Phi_{\text{KK}}(\omega) - \Phi_{\text{B}}(\omega)$ (blue dots).

Examples of analytic bunch shape reconstruction

Extensive model calculations for bunch shape reconstruction will be presented in Appendix B. Here we show only a few examples.

(1) If the time profile $\rho(t)$ features a single peak, such as the function $f_1(t)$, a Gaussian or one period of a cosine-squared wave, the Kramers-Kronig (KK) phase permits a perfect reconstruction as demonstrated in Fig. 14. A cosine-squared pulse of width $2b$, which is centered at t_c, is described by

$$\rho(t) = \frac{1}{b} \cos^2\left(\frac{\pi(t - t_c)}{2b}\right) \quad \text{for} \quad (t_c - b) \leq t \leq (t_c + b), \qquad \rho(t) = 0 \quad \text{otherwise} . \tag{45}$$

The Fourier transform can be computed analytically:

$$\mathcal{F}(\omega) = \frac{\pi^2 \sin(\omega b) \exp(i\,\omega t_c)}{\omega b(\pi^2 - \omega^2 b^2)} . \tag{46}$$

It is important to note that subtle mathematical problems arise with bunches of truly Gaussian shape. A Gaussian function violates causality because it extends over the full time range $-\infty < t < +\infty$. The unfortunate consequence is that the Gaussian form factor does not fulfill all requirements that are needed in the derivation of the Kramers-Kronig phase formula. A detailed study will be presented in Appendix A and Appendix C.
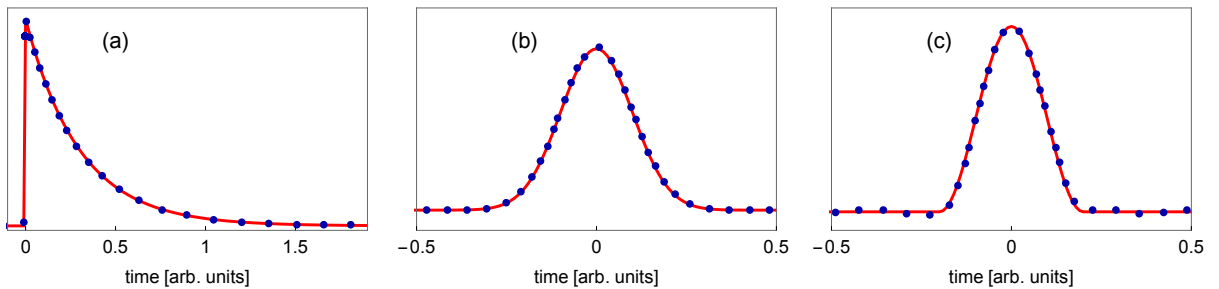

Figure 14: Reconstruction of bunch shape using the Kramers-Kronig phase for (a) the function $f_1(t)$ with an infinitely steep rise and an exponential decay, (b) a single truncated Gaussian, (c) one period of a cosine-squared wave. Red curves: input $\rho(t)$, blue dots: reconstructed $\rho(t)$.

(2) For bunch profiles featuring several peaks the situation is confusing at first sight. In some cases the input charge distribution is faithfully reconstructed using the KK phase, in other cases significant differences are found. We consider a bunch consisting of two cosine-squared pulses of different width. The Kramers-Kronig method yields a precise reproduction of the original bunch shape when the narrow peak is at the front, but it completely
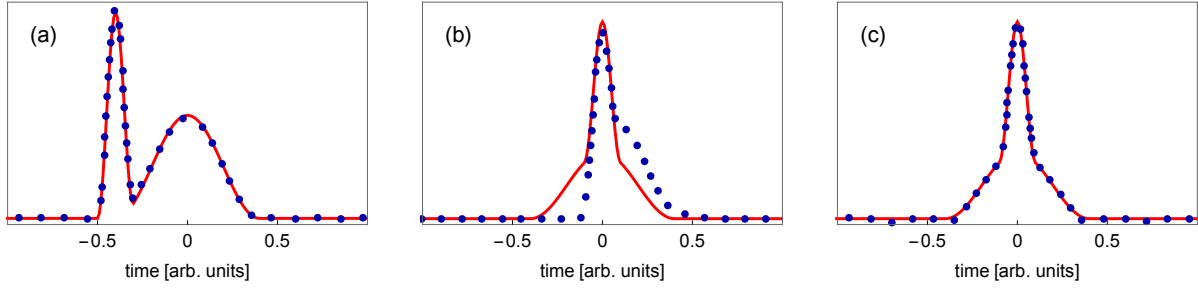
Figure 15: Reconstruction of a bunch consisting of two cosine-squared pulses of different width.
(a) Narrow peak at the front: the Kramers-Kronig (KK) phase yields a perfect reproduction of the original bunch shape. **(b)** Narrow peak at the center: the KK phase yields a bad reproduction. **(c)** Narrow peak at the center: the KK phase combined with the Blaschke phase yields an excellent reproduction of the original bunch shape. The mathematical details are presented in Appendix B.

fails when the narrow peak is centered with respect to the wide one (see Fig. 15). An excellent reproduction of the original bunch shape is achieved if both KK phase and Blaschke phase are taken into account.

(3) Our third example is a bunch consisting of three cosine-squared pulses of equal width and with uniform spacing. The amplitude ratios are $A_2/A_1 = 2/3$, $A_3/A_1 = 1/3$. The KK reconstruction agrees perfectly with the input distribution, the highest peak may be at the front or in the center of the bunch, see Fig. 16. But a slight change of the parameters may lead to different results. For example, when the amplitude ratios are $A_2/A_1 = 0.5$, $A_3/A_1 = 0.3$, the KK reconstruction works if the highest peak is at the front but it fails if it is in the center of the bunch.

Another case are three cosine-squared pulses of equal width but with non-uniform spacing. Again the KK reconstruction agrees perfectly with the input distribution if the highest peak is at the head of the bunch but fails if it is in the center, see Fig. 17.
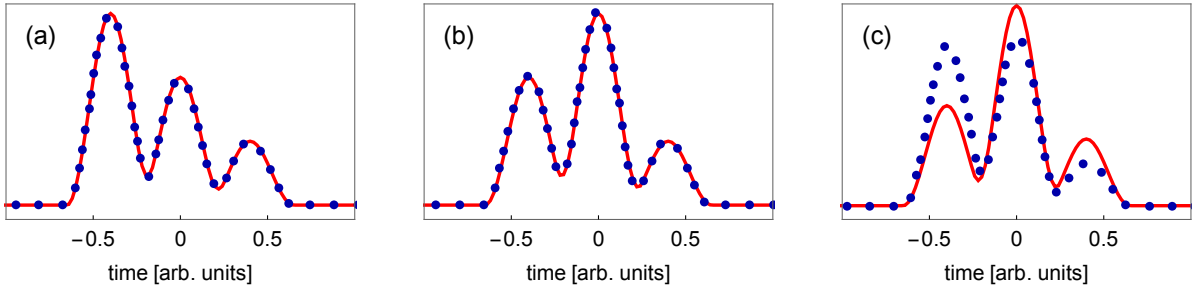

Figure 16: Three cosine-squared pulses of equal width and with uniform spacing. The amplitude ratios are $A_2/A_1 = 2/3$, $A_3/A_1 = 1/3$. **(a)** The largest peak is at the front. **(b)** The largest peak is in the center. The KK phase yields a perfect reconstruction in both cases. **(c)** For amplitude ratios of $A_2/A_1 = 1/2$, $A_3/A_1 = 1/3$, the KK reconstruction is perfect when the highest peak is at the front but fails if it is in the center.

It is instructive to look at the form factors of the bunches composed of three cosine-squared pulses. These are shown in Fig. 18. The blue curves (the KK reconstruction works) and the yellow curves (the KK reconstruction fails) are very similar, and there is no hint at all, why the KK reconstruction should work in one case but fail in the other.

The model profiles presented in this section demonstrate that a correct reconstruction of time profile with the help of the Kramers-Kronig dispersion relation cannot be guaranteed. The KK phase reconstruction fails whenever the form factor has zeros in the upper half of the complex frequency plane. This is a specific illustration of the more general theorem that a unique bunch shape reconstruction from the magnitude of the form factor is mathematically impossible.
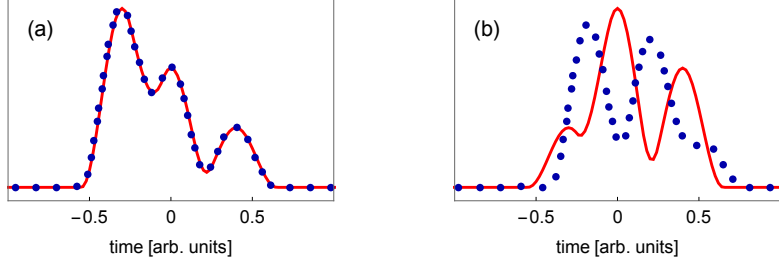
17

Figure 17: Three cosine-squared pulses of equal width and with non-uniform spacing. The ratio of the distances is $d_1/d_2 = 4/3$. **(a)** The largest peak is at the front. The KK phase yields a perfect reconstruction.
(b) The largest peak is in the center. The KK reconstruction disagrees with the input distribution. A faithful reconstruction is achieved by taking the Blaschke phase into account, see Appendix B.
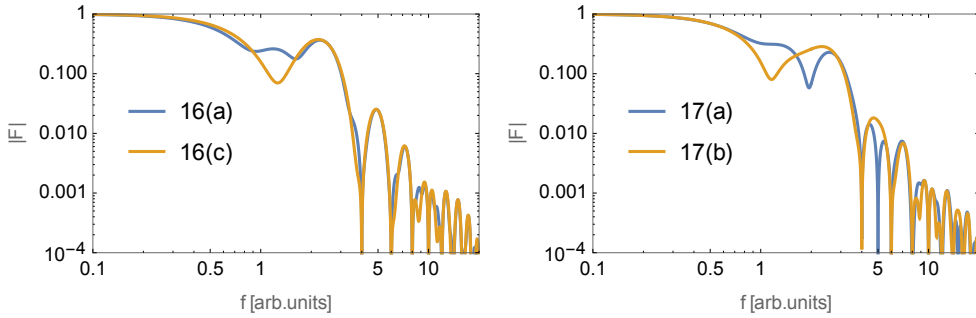

Figure 18: The computed form factor magnitudes $|\mathcal{F}(\omega)|$ of a superposition of three cosine-squared pulses of equal width, plotted versus $f = \omega/2\pi$. Left: uniform spacing, right: non-uniform spacing. The blue curves refer to the cases where the KK reconstruction works while the yellow curves curves refer to the cases where the KK reconstruction fails.

The Blaschke correction does not work for real data

The Blaschke phase is a known quantity in our model calculations where we choose a mathematically well-defined input distribution $\rho(t)$ and compute the complex form factor by Fourier transformation. Whenever this form factor has one or more zeros in the upper half of the complex plane, the KK phase alone is insufficient but the combination of Kramers-Kronig phase and Blaschke phase enables a faithful reconstruction.

In spectroscopic experiments at accelerators, however, the situation is much less favorable. There exists simply no information on such zeros of the form factor. The unfortunate consequence is that even the most precise determination of $|\mathcal{F}(\omega)| = F(\omega)$ does not allow a unique bunch shape reconstruction, there will always be ambiguities. The only phase which can be derived by analytical methods from the measured modulus of the form factor is the Kramers-Kronig phase, however the KK phase leads to wrong reconstructions if an unknown Blaschke phase should be present.

Criticism of the dispersion relation method

Computing the phase of the form factor via the Kramers Kronig dispersion relation is only justified if the form factor is an analytic function of the complex frequency. This requirement is fulfilled in the model calculations presented in this section and in Appendix A, B and C, however it may not be the case for an experimentally determined form factor which is measured at a finite number of discrete frequencies ω_j and in a limited range. Extrapolations towards very small and very large frequencies are needed, and the data have errors. One has to make the implicit assumption that an analytic function exists whose magnitude agrees (within errors) with the measured values $F(\omega_j)$, and for this function the dispersion-theoretical approach can be applied. Obviously it is desirable to have an alternative phase retrieval method at hand which does not rely on such deep lying mathematical prerequisites. The iterative phase retrieval method offers this alternative.

3.4 Iterative phase retrieval

3.4.1 Gerchberg-Saxton algorithm

Iterative algorithms for phase retrieval from intensity data are used in many research areas such as electron microscopy, X ray diffraction and astronomy. These are usually two-dimensional problems. An overview can be found in [25]. Iterative phase retrieval in the one-dimensional case of longitudinal electron bunch reconstruction from spectroscopic data has been applied recently [26, 27], and it has been claimed that thereby the restrictions of the KK method can be overcome, the main argument being that the KK phase is computed by an integral over all frequencies and thus depends on extrapolations into regimes where $F(\omega)$ has not been measured. This argument is misleading as it misses the main point, namely that a unique reconstruction of a function from the magnitude of its Fourier transform is mathematically impossible in the one-dimensional case, see Fig. 11 and the discussion in Appendix A. It is obvious that *any* phase retrieval method will suffer from this fundamental limitation. Our motivation to study the iterative method in parallel to the KK method is more of a practical nature: by comparison with time domain measurements we want to explore which of the two methods yields the most likely bunch shape.

For the iterative phase retrieval we use the Gerchberg-Saxton algorithm [28]. A block diagram is shown in Fig. 19. The basic idea is as follows. Let $\mathcal{F}(\omega) = F(\omega) \exp(i\,\Phi(\omega))$ be the form factor (Fourier transform) of
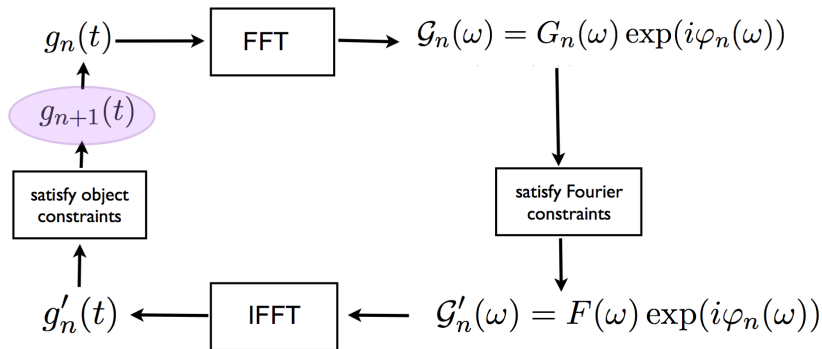

Figure 19: Block diagram of the Gerchberg-Saxton algorithm. FFT stands for Fast Fourier Transformation, IFFT for Inverse Fast Fourier Transformation.

the unknown longitudinal particle density distribution $\rho(t)$. The magnitude of the form factor $F(\omega) = |\mathcal{F}(\omega)|$ is known from measurement but the phase $\Phi(\omega)$ is unknown. The iterative loop may be started by making a first guess $g_1(t)$ of the particle density distribution and Fourier transforming it to obtain a first estimate $\mathcal{G}_1(\omega) = G_1(\omega) \exp(i\,\varphi_1(\omega))$ of the complex form factor. Then the computed modulus $G_1(\omega)$ is replaced with the measured modulus $F(\omega)$ but the computed phase $\varphi_1(\omega)$ is retained. Next an inverse Fourier transformation is carried out leading to a modified time profile $g_1'(t)$. This profile is subjected to several constraints, the most important one being: the particle density is not allowed to assume negative values. These constraints lead to a modified time profile $g_2(t)$ which is then used as starting distribution in the next iteration. Usually many iterations are needed until convergence is achieved, meaning that all constraints are fulfilled. Then one has arrived at a solution of the bunch shape reconstruction problem, but as stated above, this solution is not unique.

Alternatively the loop may be started with a guess of the complex form factor, taking the magnitude $F(\omega)$ from measurement and choosing the initial phase function either to be a constant, the Kramers-Kronig phase or a randomly varying function. In the following examples we choose random start phases.

3.4.2 Examples of iterative bunch shape reconstruction

General remarks on iterative phase retrieval with random initial phases
Without constraints on the time-domain profile, the form factor modulus can be combined with an arbitrary set of phases, yielding infinitely many different temporal profiles. In general, these profiles will not fulfill the mandatory constraint that the longitudinal particle density has to be non-negative for all times. Using this

constraint in the Gerchberg-Saxton-Loop is sufficient to reduce the number of possible solutions considerably. To investigate the uncertainties in the resulting time profile, which are caused by the randomness of the start phases, we follow a procedure proposed in [27]. The iterative loop is started 100 times, each time with a new set of random phases, and the resulting profiles are averaged. This averaging has to be done with care since the reconstructed time profiles $\rho_j(t)$ ($j = 1 \ldots 100$) will have arbitrary time shifts with respect to each other and sign-reversals of the time direction will happen (see Fig. 10). These ambiguities must be removed before averaging. To this end one optimizes the correlation coefficient between any two $\rho_i(t)$, $\rho_j(t + \delta t)$ by varying the time offset δt and by trying if time reversal $\rho_j(t) \rightarrow \rho_j(-t)$ improves the agreement. The 2σ band of the properly adjusted profiles is shown as a grey band in Figs. 21, 22, 23 and 24.
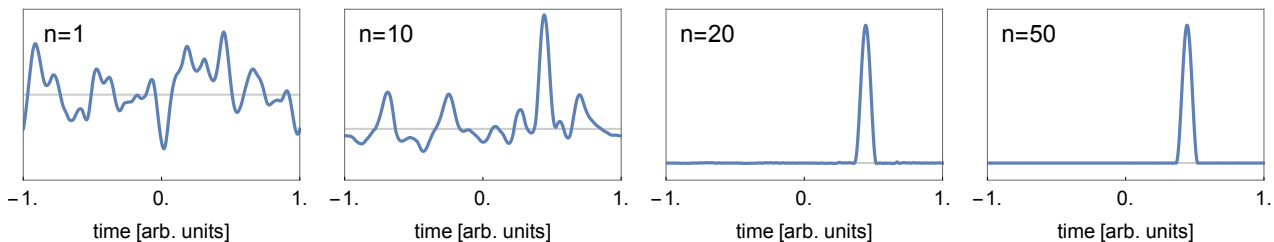

Figure 20: Iterative reconstruction of a cosine-squared charge distribution. Shown are the iteration steps $n = 1$, $n = 10$, $n = 20$ and $n = 50$.

A nice demonstration of the progressing improvement in iterative bunch shape reconstruction is presented in Fig. 20. The original bunch shape is a cosine-squared pulse. Starting with random phases the time signal $g_n(t)$ is initially very spiky and has large undershoots. The negative values are quickly eliminated by the time-domain constraints, and the input distribution $\rho(t)$ is well reproduced after about 20 iterations. The speed of convergence is found to depend on the initial conditions and on the complexity of the profile, typically some 100 iterations are needed.
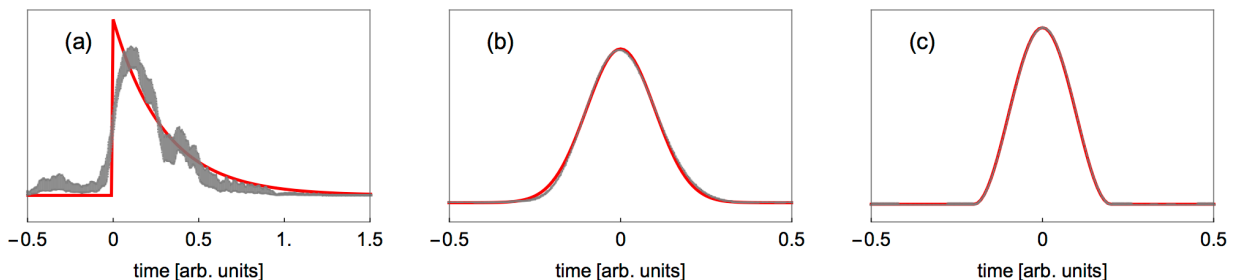

Figure 21: Iterative reconstruction of bunch shape for (a) the function $f_1(t)$ with an infinitely steep rise and an exponential decay, (b) a single truncated Gaussian, (c) one period of a cosine-squared wave. Red curves: input $\rho(t)$, shaded gray area: reconstructed $\rho(t)$. In the cases (b) and (c) the reconstruction is so good that it overlaps the original curve.

It turns out that three classes of solutions can be observed.

Class A: A unique solution with basically no variation of the resulting time profile.
Class B: One distinct shape of the profile but with a more or less pronounced uncertainty band.
Class C: Several distinct profiles with their respective uncertainty bands.

We study now the same bunch shapes as in the previous section.

(1) A single Gaussian or a truncated cosine-squared profile lead to class A solutions. Here the iterative phase reconstruction yields a unique result in very good agreement with the actual profile, as shown in Fig. 21b and Fig. 21c. However, quite a different result is obtained for the step-exponential function $f_1(t)$ which yields a class B solution. The steep initial rise is badly reproduced and a series of randomly fluctuating time profiles is
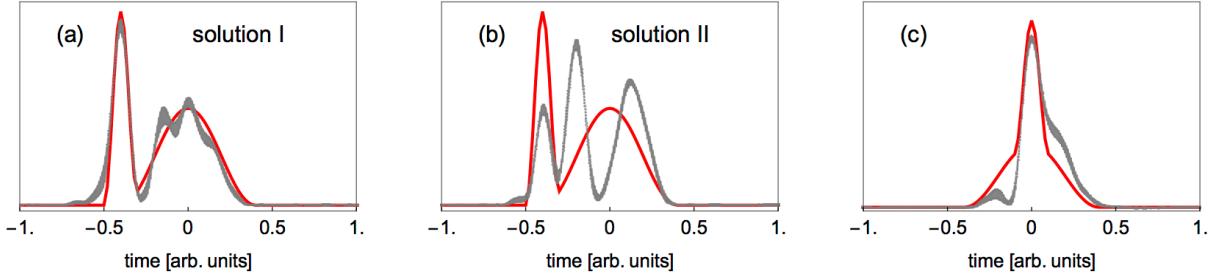
Figure 22: Iterative reconstruction of a bunch consisting of two superimposed cosine-squared peaks of different width. **(a,b)** If the narrow peak is at the front there are two types of solutions. Solution I is a time profile resembling the input profile, solution II is a profile featuring three peaks. **(c)** If the narrow peak is in the center, the reconstruction yields always the same profile which however is different from the original.
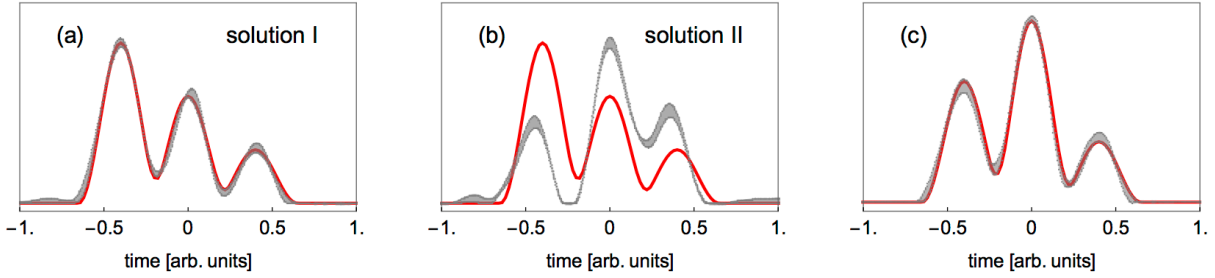

Figure 23: Iterative reconstruction of a bunch consisting of three cosine-squared pulses of equal width and with uniform spacing. The amplitude ratios are $A_2/A_1 = 2/3$, $A_3/A_1 = 1/3$. **(a, b)** If the largest peak is at the front there are two types of solutions. Solution I: In about 2/3 of the 100 iteration cycles a faithful reconstruction is achieved. Solution II: In about 1/3 of the 100 iteration cycles the Gerchberg-Saxton algorithm converges to a different profile. **(c)** If the largest peak is in the center there is a unique solution: In all iteration cycles a faithful reconstruction is achieved.

observed (Fig. 21a). The average profile is superimposed with artificial structures which appear even in front of the step. The analytic KK reconstruction is far superior in this case.

(2) Now we consider bunches with two superimposed cosine-squared peaks of different width. When the narrow peak is at the front, solutions of class C are obtained, as demonstrated by figures 22a and 22b. Two distinct sets of phases with very narrow variability are found with equal probability. One of them corresponds to a time profile resembling the input profile, but with artificial wiggles, while the second set leads to a completely different profile featuring three peaks. Notice that both solutions have the same form factor modulus and fulfill the constraint of positive charge density. Again, in this case the analytic KK reconstruction is far superior since it reproduces exactly the original time profile (see Fig. 15a). When the narrow peak and the broad peak are centered with respect to each other (Fig. 22c), the solution belongs to class A, but unfortunately it is wrong, just like the KK reconstruction depicted in Fig. 15b. So neither the KK method nor the iterative method is capable of reconstructing this shape.

(3) For "triple-peak" structures we find different behaviors depending on the details of the structure. We have seen in Fig. 16 that a bunch consisting of three cosine-squared pulses of equal width and with uniform spacing is faithfully reconstructed by the KK method. The iterative reconstruction leads to curious results. If the large peak is in the center one gets a class B solution with a good reproduction of the original shape, see Fig. 23c. However, when the large peak is at the front, one finds a class C solution: in about 2/3 of the 100 iteration cycles a faithful reconstruction is achieved (Fig. 23a) while in the remaining cycles the reconstruction is wrong (Fig. 23b).

Next we study a bunch consisting of three cosine-squared pulses of equal width and with non-uniform spacing, compare Fig. 17. When the largest peak is at the front, the iterative method yields a perfect reconstruction (Fig. 24a). When the largest peak is in the center one finds a class C solution. In 60 out of 100 iteration cycles a faithful reconstruction is achieved (Fig. 24b) but in the remaining cycles the reconstruction is wrong (Fig. 24c).
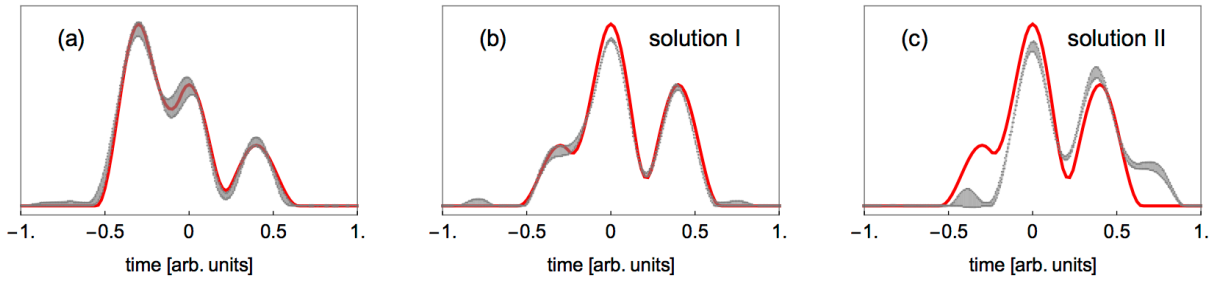
21

Figure 24: Iterative reconstruction of a bunch consisting of three cosine-squared pulses of equal width and with non-uniform spacing. **(a)** The largest peak is at the front. The iterative method yields a perfect reconstruction. **(b, c)** The largest peak is in the center. In 60 out of 100 iteration cycles a faithful reconstruction is achieved (solution I), in 40 out of 100 cycles the reconstruction is wrong (solution II).

The above examples demonstrate explicitly that the iterative method suffers from the same ambiguities as the dispersion-relation method. Moreover, it becomes evident that the intrinsic ambiguity of phase reconstruction from the magnitude of the form factor cannot be resolved by combining different phase retrieval methods.

4 Infrared and THz Spectrometer

In the description of the multichannel infrared and THz spectrometer CRISP (Coherent transition Radiation Intensity SPectrometer) we follow a previous publication [11] but address also more recent developments. An important step was the calibration [29] of the completely assembled multichannel-spectrometer which was carried out in 2016 at the infrared free-electron laser FELIX in The Netherlands.

4.1 Blazed reflection gratings

Coherent radiation from short electron bunches extends over a wide range in wavelength, from a few micrometers up to about $1\,\text{mm}$. Gratings are useful to disperse the polychromatic radiation into its spectral components. The free spectral range of a grating is defined by the requirement that different diffraction orders do not overlap. Since light of wavelength λ, diffracted in first order, will coincide with light of wavelength $\lambda/2$, diffracted in second order, the ratio of the longest and the shortest wavelength in the free spectral range is close to two. Hence many different gratings are needed to cover the full spectral range of coherent transition radiation. Overlap of different orders can be avoided by passing the radiation through a bandwidth-limiting device before it impinges on a grating. It will be shown below that this bandwidth limitation can be accomplished by a preceding grating.

A transmission grating with a large number of narrow slits distributes the radiation power almost evenly among many diffraction orders. Much superior are blazed reflection grating with triangular grooves as shown in Fig. 25a. They obey the grating equation

$$d\,(\sin\alpha + \sin\beta_m) = m\,\lambda \tag{47}$$

where d is the distance between adjacent grooves, m is the diffraction order and α the angle between the incident ray and the grating normal. To optimize the intensity for first-order diffraction ($m = 1$), the angle α
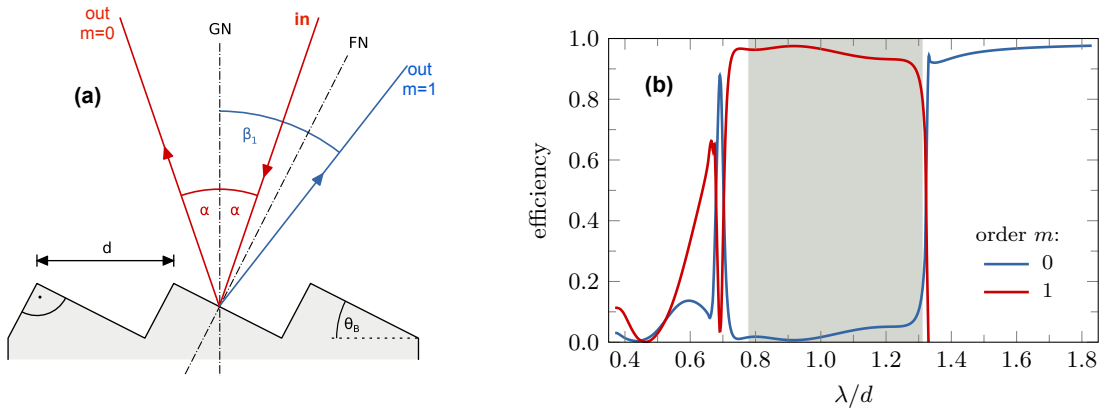

Figure 25: **(a)** Principle of a blazed reflection grating. For optimum efficiency, the incident ray and the first-order diffracted ray (for $\lambda = 1.045\,d$) have to obey the law of reflection at each facet.
(b) Efficiency curve of a gold-plated reflection grating for radiation polarized perpendicular to the grooves, computed with the code *PCGrate* (solid red curve) for first-order diffraction ($m = 1$). The wavelength range of first-order diffraction is $0.78\,d < \lambda < 1.31\,d$, it is marked by the shaded area. The computed efficiency is about 90% and almost flat. The blue curve shows the computed efficiency for zero-order diffraction. For wavelengths $\lambda > 1.33\,d$ the grating acts as a plane mirror with a reflectivity of about 95%.

is chosen such that the incident ray and the first-order diffracted ray (with a wavelength of $\lambda = 1.045\,d$, the center wavelength of the shaded area shown in Fig. 25b) enclose equal angles with respect to the facet normal FN [30]. This implies

$$\theta_B - \alpha = \beta_1 - \theta_B \quad\Rightarrow\quad \alpha = 2\,\theta_B - \beta_1$$

where θ_B is the blaze angle ($\theta_B = 27°$ in our case).

Diffraction effects vanish if the wavelength becomes too large. The incidence angle is $\alpha = 19°$ in our spectrometer setup, hence the largest possible value of $\sin\alpha + \sin\beta_m$ is 1.33. This implies that for wavelengths $\lambda > 1.33\,d$

the grating equation (47) can only be satisfied with $m = 0$ which means that no diffracted wave exists. The grating acts then as a simple plane mirror:

$$\sin \alpha + \sin \beta_0 = 0 \quad \Rightarrow \quad \beta_0 = -\alpha \,.$$

This "specular reflection" of long wavelengths is utilized in the multistage spectrometer described below.

The efficiency of a grating in a given diffraction order m is defined as the ratio of diffracted light energy to incident energy. It was computed with the commercial code *PCGrate-S6.1* by I.I.G. Inc. In Fig. 25b, the efficiency as a function of wavelength is shown for the diffraction orders $m = 1$ and $m = 0$. Short-wavelength radiation with $\lambda < 0.78\,d$ must be removed by a preceding grating stage to avoid overlap of different diffraction orders.

4.2 Multiple grating configuration

The spectrometer is equipped with five consecutive reflection gratings, G0 to G4 (see Fig. 26). Each grating exists in two variants, one for the infrared (IR) regime, the other for the THz regime. The parameters are summarized in Table 1. The IR and THz gratings are mounted on top of each other in vertical translation stages (Fig. 27a). Between each grating pair there is either a mirror (for G1, G2 and G3) or a pyroelectric detector (for G0 and G4), these are needed for alignment.

Table 1: Parameters of the gratings. The triangular grooves have a blaze angle of $\theta_B = 27°$. The distance between two grooves is called d. The minimum and maximum wavelengths of the free spectral range for first-order diffraction are called $\lambda_{\min}$ and $\lambda_{\max}$. The wavelength above which the grating acts as a plane mirror is called λ_0. All dimensions are quoted in μm. The coarse gratings with $d \geq 58.82\,\mu$m are gold-plated and were custom-made by Kugler Precision, the fine gratings with $d \leq 33.33\,\mu$m are aluminum-plated and were purchased from Newport Corporation.

IR mode $5.1 - 43.5\,\mu$m						THz mode $45.3 - 434.5\,\mu$m				
grating	d	$\lambda_{\min}$	$\lambda_{\max}$	λ_0		grating	d	$\lambda_{\min}$	$\lambda_{\max}$	λ_0
G0	4.17	-	-	5.5		G0	33.33	-	-	44
G1	6.67	5.13	8.77	8.8		G1	58.82	45.3	77.4	77.6
G2	11.11	8.56	14.6	14.7		G2	100.0	77.0	131.5	132
G3	20.0	15.4	26.3	26.4		G3	181.8	140.0	239.1	240
G4	33.33	27.5	43.5	-		G4	333.3	256.7	434.5	440

In the following we describe the THz configuration, the infrared configuration works correspondingly. The incident radiation is passed through a polarization filter (HDPE thin film THz polarizer by TYDEX) to select the polarization component perpendicular to the grooves of the gratings, and is then directed towards grating G0 which acts as a bandwidth-limiting device: short-wavelength radiation ($\lambda < \lambda_0 = 44\,\mu$m) is dispersed by G0 and guided to an absorber, long-wavelength radiation ($\lambda > \lambda_0$) is specularly reflected towards G1 which is the first grating stage of the spectrometer. Radiation in the range $\lambda_{\min} = 45.3\,\mu$m$< \lambda < \lambda_{\max} = 77.4\,\mu$m is dispersed by G1 in first-order and focused by a ring mirror onto a multi-channel detector array, while radiation with $\lambda > \lambda_0 = 77.6\,\mu$m is specularly reflected and sent to G2. The subsequent gratings work similarly and disperse the wavelength intervals $[77.0, 131.5]\,\mu$m (G2), $[140.0, 239.1]\,\mu$m (G3), and $[256.7, 434.5]\,\mu$m (G4).

For each of the gratings G1 to G4, the first-order diffracted radiation is recorded in an array of 30 pyroelectric detectors which are arranged on a circular arc covering $57°$. A ring-shaped parabolic mirror focuses the light onto this arc. The computed light dispersion and focusing is shown schematically in Fig. 27b for 5 of the 30 wavelength channels.
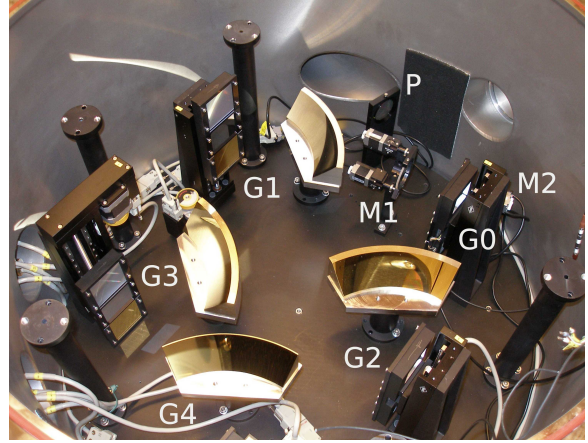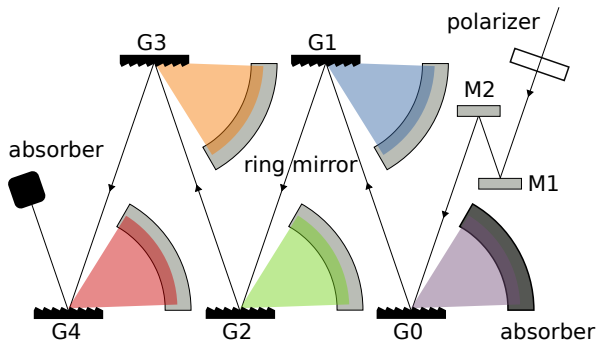
Figure 26: Schematic view and photo of the staged spectrometer equipped with five reflection gratings. The spectrometer is mounted in a vacuum vessel to avoid the absorption of THz waves in air of normal humidity. The detector arrays are not yet mounted above the focusing mirrors. Grating G4 is just outside the photo but its mirror can be seen. P is the polarizer, and M1, M2 are the input alignment mirrors.
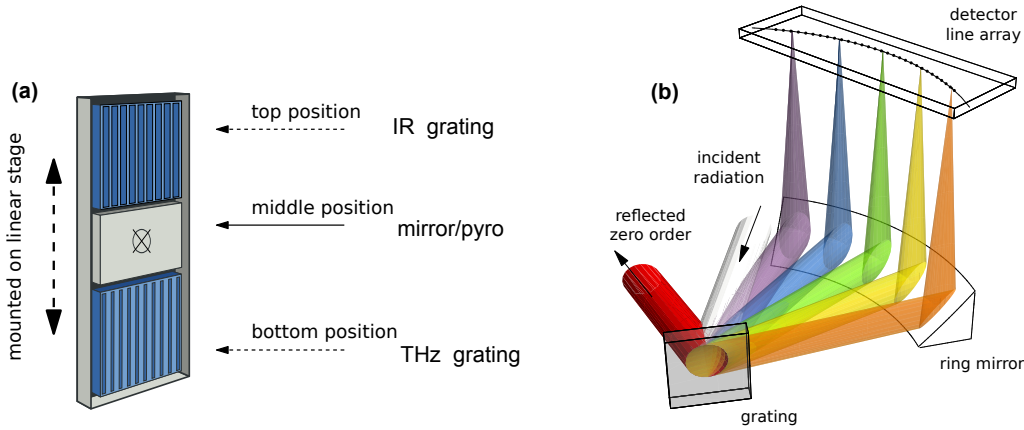

Figure 27: **(a)** Corresponding pairs of gratings are mounted on vertical translation stages with the aluminum-plated infrared grating in the upper position and the gold-plated THz (far-infrared) grating in the lower position. Between each pair there is either a plane mirror (for G1, G2, G3) or a pyroelectric detector (for G0 and G4) which are used for alignment purposes. **(b)** Arrangement of the grating, the ring mirror and the array of 30 pyroelectric detectors. The light dispersion and focusing have been computed with a ray tracing code, for clarity this is shown for only 5 of the 30 wavelength channels.

4.3 Pyroelectric detectors

A critical component of the broadband single-shot spectrometer is a detector featuring high sensitivity over the entire infrared and far-infrared (THz) regime, from μm to mm wavelengths. Bolometric devices, responding to the deposited radiation energy through a temperature rise, are capable of covering such a wide wavelength range. A special pyroelectric detector has been developed to our specification by an industrial company (InfraTec). This sensor possesses sufficient sensitivity for the application in a coherent transition radiation spectrometer and has a fast thermal response. The layout of the detector is shown in Fig. 28a. It consists of a 27 μm thick lithium tantalate (LiTaO$_3$) crystal with an active area of $2 \times 2\,\mathrm{mm}^2$. The front surface is covered with a NiCr electrode of 20 nm thickness instead of the more conventional 5 nm. The backside electrode is a 5 nm NiCr layer instead of the conventional thick gold electrode. The combination of a comparatively thick front electrode and a thin backside metallization suppresses internal reflections which are the origin of the strong wavelength-dependent

25

efficiency oscillations observed in conventional pyroelectric detectors. The beneficial effect of the novel surface layer structure is illustrated in Fig. 28b. To enhance absorption below $100\,\mu$m the front electrode is covered with a black polymer layer which is transparent above $100\,\mu$m.
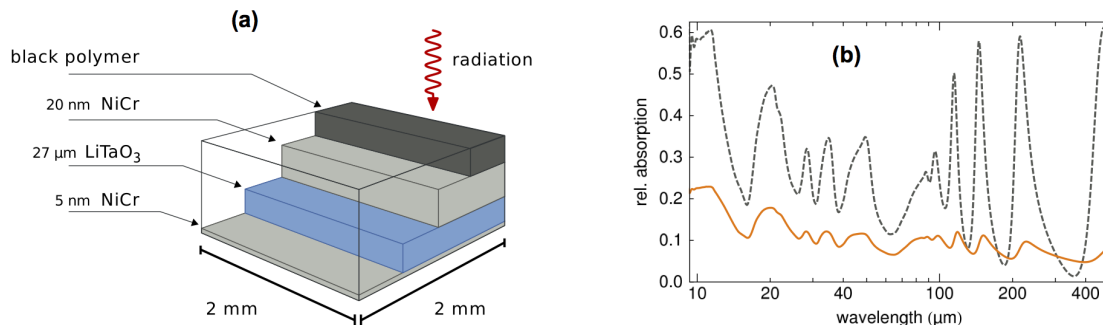

Figure 28: **(a)** Layout of the pyroelectric detector LIM-107-X005. **(b)** Computed infrared absorption as a function of wavelength. Solid yellow curve: $27\,\mu$m LiTaO$_3$ detector with optimized coatings for minimum internal reflections (20 nm NiCr at front surface and 5 nm NiCr at back surface). Dotted black curve: $27\,\mu$m LiTaO$_3$ detector with standard coatings: a 5 nm NiCr layer at the front surface and a thick gold layer at the back surface.

The thermal expansion of the pyroelectric crystal, due to the absorption of radiation, creates a surface charge which is converted into a voltage signal by the charge-sensitive preamplifier (Cremat CR110). The preamplifier and the twisted-pair line driver amplifier are mounted on the electronics board inside the vacuum vessel. Line receiver, Gaussian shaping amplifier and ADC (analog-to-digital converter) are located outside the vacuum vessel. The commercial preamplifier Cremat CR110 generates pulses with a rise time of 10 ns and a decay time of $140\,\mu$s. This is adequate for repetition rates of 10 Hz or less. A Gaussian shaping amplifier (Cremat CR200 with $4\,\mu$s shaping time) is used to optimize the signal-to-noise ratio. The shaped signals are digitized with 120 parallel ADCs with 9 MHz clock rate, 14 bit resolution and 50 MHz analog bandwidth.

4.4 Transition radiation beamline and spectrometer response function

CTR beamline
Transition radiation is produced on a screen inside the ultrahigh vacuum beam pipe of the FLASH linac at the 202 m position. The screen is tilted by 45° hence backward TR is emitted perpendicular to the electron beam axis. The screen has a rectangular shape ($16 \times 25\,\text{mm}^2$) and consists of a $380\,\mu$m thick polished silicon wafer which is coated with a 150 nm aluminum layer at the front surface. It is a so-called "off-axis" screen, positioned outside the nominal electron beam axis. Selected electron bunches can be steered onto the TR screen using a fast kicker magnet. This permits high-resolution diagnostics on a single bunch out of a long train without impeding the FEL gain process for the unkicked bunches.

The radiation is coupled out through a 0.5 mm thick window made from chemical vapor deposition (CVD) diamond. In contrast to standard window materials such as glass, quartz or polyethylene, CVD diamond has almost negligible absorption in the entire spectral range of transition radiation, from visible light up to millimeter waves, except for a narrow absorption band around $5\,\mu$m where lattice vibrations are excited. The radiation is transported to the spectrometer by an optical system consisting of four focusing and four plane mirrors. The optical layout is shown in Fig. 29, a three-dimensional technical drawing in Fig. 30. The design criteria are identical to those of the CTR beamline at the 141 m position of the FLASH linac and have been documented in Ref. [31], they need not be repeated here. Beamline and spectrometer are mounted in vacuum vessels to avoid the strong absorption of THz waves in air of normal humidity.

26

| f (mm) | | - | - | 245 | 1200 | 1200 | | 1040 | - | - | | - |
| ø (mm) | 15x25 | 20 | 86 | 86 | 86 | 86 | | 86 | 50.8 | 58x58 | | 58x58 |

(labels: nominal beam axis · kicked beam · diamond window · parabolic mirror · toroidal mirror · toroidal mirror · toroidal mirror · polarizer · grating G0 · grating G4)

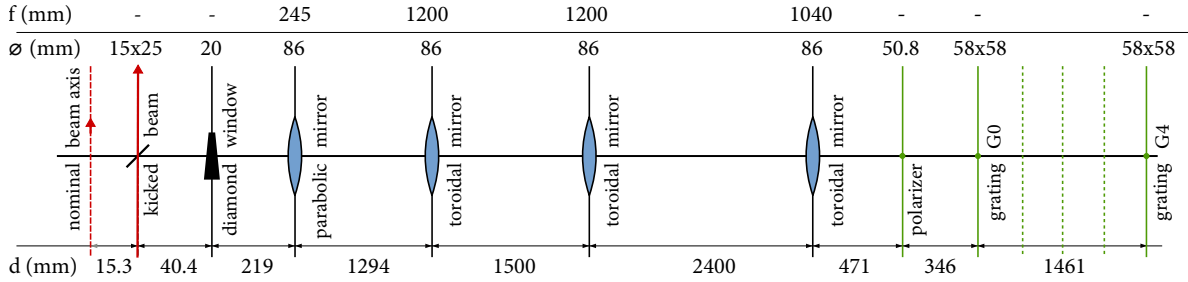| d (mm) | 15.3 | 40.4 | 219 | 1294 | 1500 | 2400 | 471 | 346 | 1461 |

Figure 29: Optical design of the CTR beamline. The focusing parabolic or toroidal mirrors M1 to M4 are shown as lenses with their respective positions, focal lengths, and diameters. Before entering the staged grating arrangement, the radially polarized transition radiation is passed through a linear polarizer to select the field component perpendicular to the grooves of the blazed gratings.
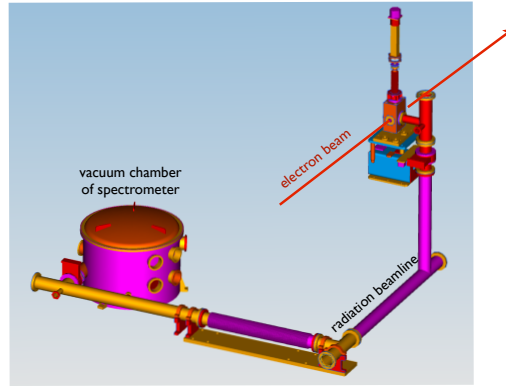

Figure 30: Technical layout of the CTR beamline at the 202 m position.

Response function

The calibration of a broadband spectrometer, covering the wavelength range from $4\,\mu$m to $400\,\mu$m, is a demanding task. There exists no "table-top" radiation source which can be tuned over such wide a range. A broadband accelerator-based source is the free-electron laser FELIX in The Netherlands, it provides monochromatic infrared radiation between a few μm and $135\,\mu$m. Our spectrometer CRISP was shipped to FELIX and calibrated with FEL radiation in 2016, see below.

The other essential component of the spectrometer setup at the 202 m position of the FLASH linac is the CTR beamline which guides the transition radiation from the TR screen to the spectrometer. This beamline is rigidly mounted in the linac tunnel and cannot be moved to an outside FEL laboratory to determine its wavelength-dependent transmission properties. Moreover, such a test would be pretty meaningless since FEL radiation and transition radiation have very different angular characteristics: the FEL beam is well-collimated and has little divergence while the TR beam has a fairly wide angular divergence and requires large-aperture mirrors in the beamline. A performance test of the entire system - CTR beamline plus CRISP spectrometer - has to be done *in situ*.

An ideal test scenario would be to have a beam of pointlike electron bunches with precisely known charge Q, which generate transition radiation of well-known emission characteristics. A realistic test scenario has to take into account that the beam optics in the FLASH linac is optimized for high-gain FEL operation and that an extremely small beam radius cannot be realized. The 202 m position of the linac is a good location for the spectrometer because here the electron beam is round and the horizontal and vertical beta functions are reasonably small ($\beta_x \approx \beta_y \approx 7$ m). The rms beam radius is $\sigma \approx 100\,\mu$m.

To investigate the performance of the multichannel spectrometer we consider therefore a *reference bunch* of length zero with a charge of $Q_{\mathrm{ref}} = 100\,$pC and a cylindrically symmetric Gaussian transverse density distribution with $\sigma_{\mathrm{ref}} = 100\,\mu$m. The transition radiation produced by the bunch upon crossing the TR screen can be

accurately computed. The radiation passes through the beamline, where losses occur due to diffraction and aperture limitations, and enters the spectrometer. Here it is decomposed into 120 spectral components (either in the IR regime or in the THz regime) that impinge on the corresponding pyroelectric detectors. The amplifiers produce 120 voltage signals v_m which are digitized by ADCs. The voltages v_m are taken as the components of a *voltage-signal vector* $\boldsymbol{v} = (v_1, v_2 \ldots v_{120})$. There are two such voltage-signal vectors, one for the IR regime, the other for the THz regime. These two vectors are the system response to the reference bunch. The same set of 120 pyro-detectors is used in the two operation modes, only the gratings are changed when switching from IR mode to THz mode or vice versa.

In each of the operation modes (IR or THz) the spectrometer defines 120 wavelength bins $\lambda_m \pm \delta\lambda_m$. The primary transition radiation energies within these bins are called U_m. Only a certain fraction u_m of the primary TR energy U_m passes through the CTR beamline and arrives at the pyro-detector m:

$$u_m = P_{\text{trans}}(\lambda_m)U_m$$

where $P_{\text{trans}}(\lambda_m)$ is the wavelength-dependent transfer function (transmission) of the CTR beamline.

Thus the overall response function of the multichannel spectrometer, as mounted at the rear end of the CTR beamline, consists of two parts: (1) the generation of transition radiation, the transmission through the CTR beamline and the grating stages as a function of wavelength, and (2) the wavelength-dependent response function of the pyroelectric detectors.

Part 1: The generation of transition radiation and its transmission through the CTR beamline (Figs. 29 and 30) have been determined by an elaborate "start-to-end" simulation using the code *THzTransport*. The mathematical formalism is explained in Appendix D. In the program, an infinitesimally short reference bunch with a charge of $Q_{\text{ref}} = 100\,\text{pC}$ and a cylindrically symmetric Gaussian transverse density distribution with $\sigma_{\text{ref}} = 100\,\mu\text{m}$ is used. The code *THzTransport* computes the transition radiation produced by the reference bunch by applying the Weizsäcker-Williams method of virtual photons (see Section 2). The electromagnetic field of a radially extended charged disc is used, hence the suppression of short wavelengths by the transverse form factor (see Fig. 9) is automatically taken into consideration. The spectral radiation components are propagated from the TR source screen through the CTR beamline to the corresponding pyro-detector. This is done for all wavelengths λ_m. The propagation proceeds in a stepwise fashion:

TR screen → diamond window → M1 → M2 → M3 → M4 → spectrometer.

All apertures and near-field diffraction effects are taken into consideration. Within the staged-grating spectrometer, the simulation distinguishes which grating guides the radiation contained in the wavelength bin $\lambda_m \pm \delta\lambda_m$ to the corresponding pyro-detector m, taking the focusing by the ring mirror into account. The computed spectral energy u_m impinging onto the pyro-detector is converted into a voltage signal v_m using the calibration described in Part 2.

Part 2: The calibration of a prototype pyroelectric detector was carried out in 2008 [32]. Combining this calibration with the numerical simulation described in *Part1* yields the preliminary spectrometer response function plotted in Fig. 31 (blue dots). The spectral response of the completely assembled CRISP spectrometer was determined in 2016 [29] with tunable monochromatic infrared radiation from the free-electron laser FELIX in the range from $6\,\mu\text{m}$ to $135\,\mu\text{m}$. Starting from the measured FEL power at the entrance aperture of the spectrometer and the known transverse intensity distribution of the FEL radiation, the monochromatic FEL beam was propagated by a *THzTransport* simulation through the polarizer, the grating stages and the focusing ring mirror up to the pyro-detector corresponding to the selected FEL wavelength. In this way the response of each of the 120 pyro-detectors was calibrated by establishing the relation between the known incident radiation energy u_m and the measured output voltage v_m, both in the IR configuration and in the THz configuration. The efficiency of the gratings was part of this calibration.

The overall spectrometer calibration with FEL radiation cannot be simply taken over to the TR diagnostic station at FLASH. The input of the well-collimated FEL radiation into the spectrometer is very different from the transition radiation input. The quite divergent transition radiation wave has to propagate through the CTR beamline before entering the spectrometer. For this reason, the pyro-detector calibration at FELIX must be combined with the above-mentioned "start-to-end" simulation (for details see Refs. [29, 33]). Having computed the TR energies u_m impinging onto the pyroelectric detectors, the detector calibration at FELIX is utilized to convert these energies u_m into voltages v_m. The improved spectrometer response function based on this overall
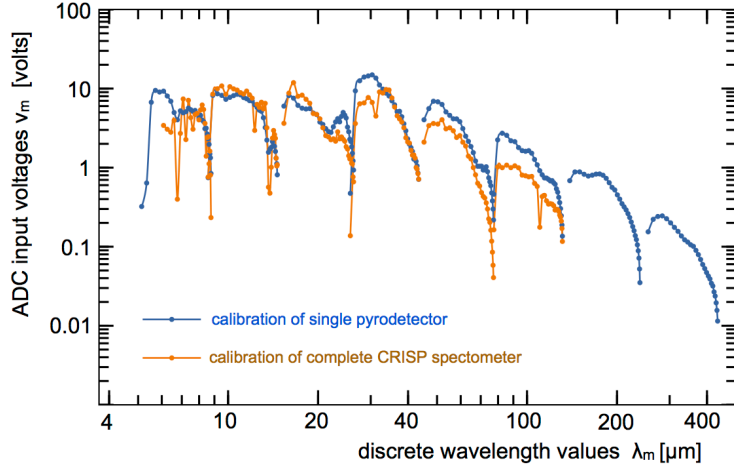
Figure 31: Blue dots: First determination of the spectrometer response function, derived from the comprehensive *THzTransport* simulation described in *Part 1* and the measured response of a single prototype pyro-detector [32]. The grating efficiency was taken from theory, see Fig. 25b. Yellow dots: Improved response function after calibration of the completely assembled CRISP spectrometer at the infrared free-electron laser FELIX (for details see Ref. [29]). This overall calibration determines the responses of all 120 pyroelectric detectors as well as the grating efficiencies. The sawtooth-like structure is caused by the variation of acceptance within each detector array.

calibration is also shown in Fig. 31 (yellow dots). The rather similar wavelength-dependencies of the blue and yellow curves allows to extrapolate the pyro-detector calibration into the wavelength range from $135 \, \mu$m to $420 \, \mu$m where no FEL radiation was available at the FELIX facility.
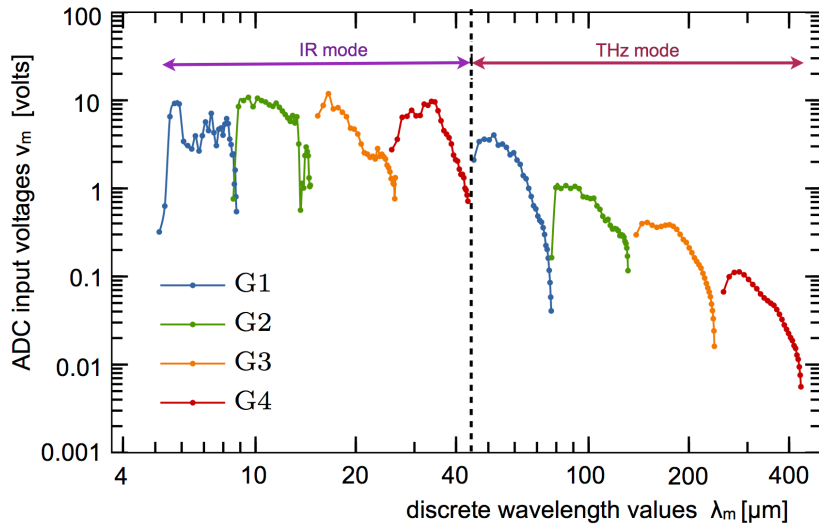

Figure 32: The overall response function $R_{ref}(\lambda)$ of the multichannel spectrometer as installed at the rear end of the CTR beamline (all corrections applied). The discrete wavelengths $\lambda_1, \lambda_2, \ldots \lambda_{120}$ (in either the IR mode or the THz mode) are the center values of the wavelength bins $\lambda_m \pm \delta\lambda_m$ associated with the 120 pyro-detectors (see text). The response function has been computed using the parameters of the reference bunch: length zero, rms radius $\sigma_{ref} = 100 \, \mu$m, charge $Q_{ref} = 100 \, $pC.

Meanwhile the 2016 data have been critically re-evaluated and a number of additional subtle corrections have been applied. The resulting overall response function $R_{ref}(\lambda)$ is shown in Fig. 32. This response function of the entire *System* (defined as the combination of TR screen, diamond window, CTR beamline and CRISP

29

spectrometer) has been computed using the parameters of the reference bunch. It relates indirectly the spectral transition radiation energy U_m inside the wavelength bin $\lambda_m \pm \delta\lambda_m$ and the ADC input voltage v_m of the corresponding pyroelectric detector. The defining equation

$$v_m = R_{\text{ref}}(\lambda_m) \quad \text{with} \quad m = 1, 2 \ldots 120 \qquad \text{(either IR or THz mode)} \tag{48}$$

looks deceivingly simple, but one has to keep in mind that all complications - the elaborate mathematical computations and the difficult calibration procedure - are hidden in the response function.

The *System* response to an arbitrary bunch with the same rms radius $\sigma = \sigma_{\text{ref}}$ but with a different charge Q and a finite length (longitudinal form factor $F(\omega) = |\mathcal{F}(\omega)|$) can be easily written down. The radiation energy U_m' in the wavelength bin $\lambda_m \pm \delta\lambda_m$, generated by this bunch, and the ADC input voltage v_m' of pyro-detector number m are related to the same quantities of the reference bunch by

$$U_m' = \frac{Q^2}{Q_{\text{ref}}^2} |F(\omega_m)|^2 U_m, \quad v_m' = \frac{Q^2}{Q_{\text{ref}}^2} |F(\omega_m)|^2 v_m \quad \text{with} \quad \omega_m = \frac{2\pi c}{\lambda_m}.$$

Inserting for v_m the above relation (48) we get the important equation

$$\boxed{|F(\omega_m)|^2 \equiv \left| F\left(\frac{2\pi c}{\lambda_m}\right) \right|^2 = \frac{Q_{\text{ref}}^2}{Q^2} \frac{v_m'}{R_{\text{ref}}(\lambda_m)} \quad (m = 1, 2 \ldots 120) \qquad \text{(either IR or THz mode)}} \tag{49}$$

which allows us to derive the longitudinal form factor from the measured voltage-signal vector $\boldsymbol{v'} = (v_1', v_2' \ldots v_{120}')$ with the help of the response function $R_{\text{ref}}(\lambda)$. Note that this equation must be used twice, in the IR mode and in the THz mode, in order to cover the full spectral range.

Impact of electron beam radius

As said above, the response function $R_{\text{ref}}(\lambda)$ has been computed for the reference rms electron beam radius
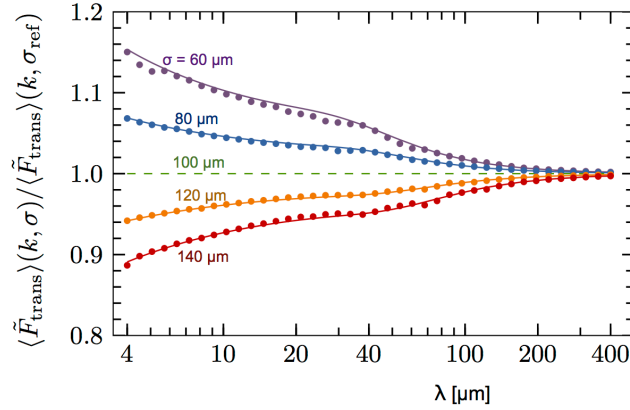

Figure 33: Influence of the electron beam radius on the rms transverse form factor (32). The ratio $\langle \tilde{F}_{\text{trans}} \rangle (k, \sigma) / \langle \tilde{F}_{\text{trans}} \rangle (k, \sigma_{\text{ref}})$ is plotted as a function of wavelength ($\lambda = 2\pi/k$). The reference radius is $\sigma_{\text{ref}} = 100\,\mu\text{m}$. The continuous curves result from computations based on the near-field formula (15) and an aperture angle of $100\,\text{mrad}$ (compare Fig. 9b). The dots come from a numerical modeling of the radiation transport through the CTR beamline to the spectrometer.

of $\sigma_{\text{ref}} = 100\,\mu\text{m}$. When the electron energy or the optics of the accelerator is changed, the beam radius at the spectrometer position will change as well. It his hence of interest to know how critically the rms transverse form factor (32) depends on σ. To get an impression, we plot in Fig. 33 the ratio

$$\frac{\langle \tilde{F}_{\text{trans}} \rangle (k, \sigma)}{\langle \tilde{F}_{\text{trans}} \rangle (k, \sigma_{\text{ref}})} \qquad \text{with} \quad k = \frac{2\pi}{\lambda}$$

as a function of wavelength for rms radii between $60\,\mu\text{m}$ and $140\,\mu\text{m}$. This figure shows that the impact of the electron beam radius is weak: a 20% uncertainty in the knowledge of the beam radius changes the rms transverse form factor by just 5% at the shortest wavelength of $4\,\mu\text{m}$ and by much less at larger wavelengths.

5 Results on bunch shape reconstruction at FLASH by CTR spectroscopy

The parallel readout of the CRISP spectrometer permits the measurement of the CTR spectrum generated by a single bunch, either in the infrared mode or in the THz mode. To cover the full spectral range from $4\,\mu$m to $420\,\mu$m, measurements with the two grating sets are needed (see Section 4). The stability of the accelerator is generally quite high and the fluctuations in the CTR spectrum within the required data taking time of a few minutes are smaller than the uncertainties caused by the preamplifier noise. For the results shown here, the form factors have been derived as averages of 200 single bunches, kicked from consecutive bunch trains at 10 Hz repetition rate. The measurement series with the two grating sets are carried out consecutively and they are individually averaged. The averaging procedure reduces detector and amplifier noise as well as statistical fluctuations, and it extends the applicability of spectroscopic bunch shape analysis to low bunch charges and large bunch lengths.

5.1 Computation of form factor

Consider now bunches with an rms radius of $\sigma = \sigma_{\mathrm{ref}} = 100\,\mu$m but with unknown charge Q and unknown longitudinal charge density profile. The spectrometer measures the deposited transition radiation energies u'_m in its 240 discrete wavelengths bins $\lambda_m \pm \delta\lambda_m$ (120 in the IR regime and 120 in the THz regime). The pyroelectric LaTiO$_3$ crystals and the amplifiers convert these energies into voltages v'_m which are digitized and recorded. Knowing the bunch charge Q, the absolute square $|F(\omega)|^2$ of the longitudinal form factor is computed at the discrete frequencies $\omega_m = 2\pi c/\lambda_m$ with the help of Eq. (49).
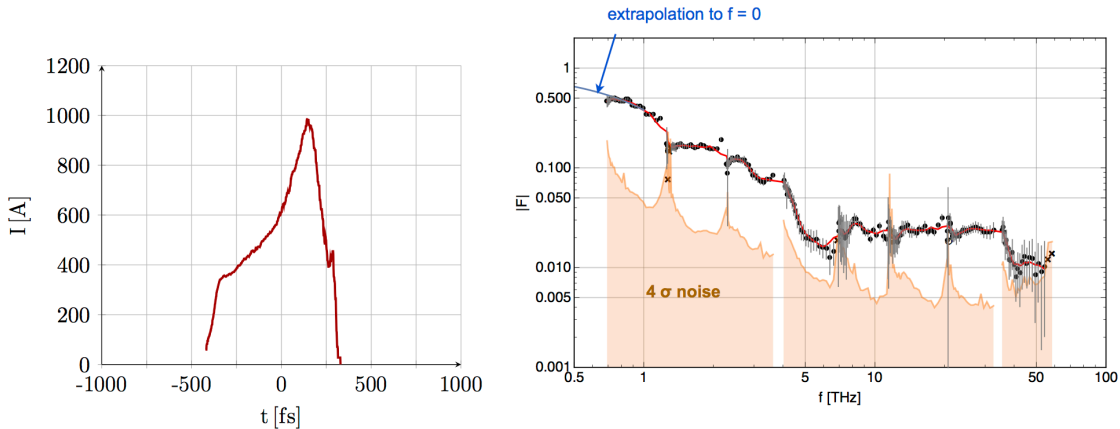

Figure 34: A measured TDS time profile and the corresponding form factor as determined from the spectroscopic data. $F(\omega) = |\mathcal{F}(\omega)|$ is plotted as a function of $f = \omega/2\pi$. The smooth red curve is a spline-interpolation which suppresses strong fluctuations in the 5-10 THz region and above 40 THz. The grey vertical lines indicate the shot-to-shot fluctuations in the sequence of 200 bunches. The shaded area shows the $4\,\sigma$ limit of the noise signal. The blue curve is the extrapolation to $f = 0$. Note that the frequency axis is logarithmic, hence the condition $F(0) = 1$ cannot be seen in this graph, but it is indeed fulfilled.

Several data processing steps have to be applied which are illustrated in Fig. 34 where an experimentally determined form factor is shown as a function of frequency.
a) Only "significant" data points are retained with a voltage signal at least 4 standard deviations above noise. The noise level is determined by measurements without electron beam.
b) Points with a strong excursion are removed.
c) The data are extrapolated into the frequency range not covered by the spectrometer. The low-frequency extrapolation is made with a Chebyshev polynomial of second order (parabola) and the basic condition $F(\omega) \to 1$ for $f = \omega/2\pi \to 0$ is imposed as a constraint. It is very reassuring that the form factors can be smoothly extrapolated to $F(0) = 1$ without any scaling factors. This means that the absolute calibration of the *System* (TR screen - CTR beamline - spectrometer) is known to a level of about 20%.

The extrapolation to high frequencies is uncritical since the measured form factor at the highest frequencies is

normally so small that it has very little influence on the bunch shape reconstruction. It can be done with an inverse power law.

5.2 Bunch shape reconstruction methods

The Kramers-Kronig phase is computed by numerical integration of formula (40) with a cutoff frequency of $\omega_{\text{cut}} = 2\pi \cdot 300\,\text{THz}$. For this purpose, the measured form factor values are interpolated by a spline function and extrapolated to high and low frequencies as mentioned above. The inverse Fourier integral is also evaluated by numerical integration. In both cases, we use a Gauss-Kronrod integration scheme with locally adaptive subintervals.

In the iterative phase reconstruction algorithm, the experimentally determined magnitude $F(\omega)$ of the form factor, including the extrapolations to low and high frequency, is evaluated at the discrete points of a frequency grid with 4000 grid points. The limiting parameter sets used are

Set 1: $-100\,\text{THz} \leq f \leq +100\,\text{THz}$, step width 50 GHz, time range 10 ps with 5 fs resolution
Set 2: $-600\,\text{THz} \leq f \leq +600\,\text{THz}$, step width 300 GHz, time range 1.6 ps with 0.8 fs resolution

or a suitable intermediate set. The frequency range and step width of the discretization are adapted to the expected width of the time profile. The condition is imposed that there be at least 200 points within the FWHM region, otherwise the maximum frequency and bin width of the grid are adjusted appropriately. The Fourier transformations of the iterative loop are done by an FFT algorithm, the inverse Fourier transformation by IFFT.

The following steps depend on the choice of the initial phases. If one starts with either a constant phase or the KK phase, the procedure is straightforward. The phase factors $\exp(-i\,\Phi(\omega_j)t)$ are evaluated at the discrete grid frequencies $\omega_j = 2\pi f_j$ and then the IFFT is carried out. For the case of "random" start phases some more effort is needed. Following a proposal by Fienup [25], the computational steps are as follows:
1) In step 1 all phases are put to zero and the IFFT is applied to the $F_j = F(\omega_j)$. This yields a symmetric time profile whose maximum is then normalized to 1.
2) In step 2 a threshold of 0.2 is imposed. All points of the time distribution whose values are below this threshold are put to zero. The points above threshold are replaced by random numbers between zero and one.
3) In step 3 an FFT is applied to the discrete time distribution generated in step 2. The resulting phases Φ_j are "quasi-random" (since their origin is a randomized time distribution), and these phases are used for starting the Gerchberg-Saxton loop. This method of generating randomized start-phases avoids the unphysical procedure of assigning a completely random start phase to each frequency point, and moreover it improves the speed of convergence.

The progress of convergence of the iteration loop is monitored by comparing the modulus of the reconstructed form factor with the measured values. The iteration is assumed to have converged if the Pearson correlation coefficient deviates from 1 by less than 10^{-4} (or if the change between subsequent iterations is below 10^{-4}).

The iterative phase retrieval procedure has several variants which can be used to improve the convergence, speed and stability of the result [25]. We have studied these variants in great detail but this is outside the scope of this paper and will be presented in a dedicated publication. We just mention that procedures like "shrink wrapping" or "bubble wrapping" , promoted in the past [34], did not improve the results.

To mitigate fluctuations in the time profile, resulting from the randomness of the start phases, we follow a procedure proposed in [27]. The iterative loop is started about 50 times with new quasi-random phases and the resulting profiles are averaged. This averaging has to be done with care since the reconstructed time profiles $\rho_j(t)$ $(j = 1 \ldots 50)$ will have arbitrary time shifts with respect to each other and sign-reversals of the time direction will happen (see Fig. 10). These ambiguities must be removed before averaging. To this end one optimizes the correlation coefficient between any two $\rho_i(t)$, $\rho_j(t + \delta t)$ by varying the time offset δt and by trying if time reversal $\rho_j(t) \to \rho_j(-t)$ improves the agreement.

Averaging is a means to identify significant structures. Structures with a high likelihood will appear repeatedly in many iteration cycles and survive the averaging while those with a low likelihood will fluctuate from one iteration cycle to the next and average to zero.

5.3 Experimental results

Many different bunch shapes can be realized at FLASH by varying the off-crest phase of the 1.3 GHz RF field in the accelerating cavities preceding the bunch compressor, and by choosing appropriate values for the amplitude and phase of the 3.9 GHz RF field in the third-harmonic cavity. Here we present four examples of longitudinal bunch shape reconstruction from spectroscopic measurements. The computed time profiles are compared with the time profiles recorded with the transversely deflecting microwave structure TDS [7], whenever available. The TDS is basically an ultrafast oscilloscope tube with a resolution in the 10 fs regime. In principle it is a single-shot device that should be suited to faithfully record the longitudinal particle density $\rho(t)$ of a single electron bunch. In practice this is not the case for the bunches having passed the magnetic bunch compressor chicanes since these might have acquired a nonvanishing average transverse momentum and other internal correlations that vary along the bunch axis. As a consequence, the streak image on the TDS view screen depends on the streak direction, see Fig. 35.

This effect can be taken care of by streaking a first bunch in positive direction and a second bunch in negative direction. The most likely bunch profile is obtained from these data by a tomographic reconstruction algorithm developed at SLAC (unpublished). The TDS time profiles shown in the following figures result from the tomographic reconstruction. The difference between the two streak directions is small for wide bunches with little structure but becomes pronounced for strongly compressed bunches featuring sharp structures [35]. Presently systematic studies are being carried out which will be reported elsewhere.
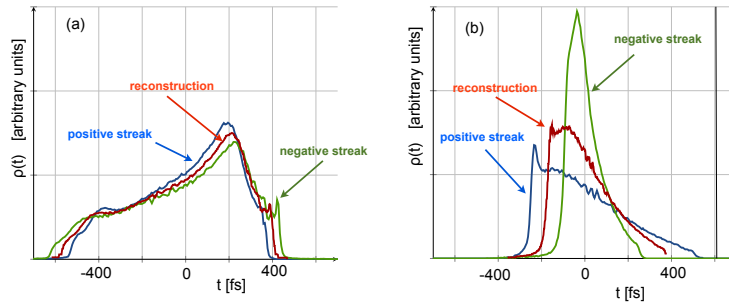

Figure 35: Time profiles measured with the TDS. The positive streak direction is shown by the blue curve, the negative streak direction by the green curve, and the reconstructed time profile by the red curve. **(a)** Bunch with moderate compression. **(b)** Bunch with strong compression and a steep initial rise.

Example 1
In this example bunches were generated with a steep decay in the tail region. The TDS profile and the
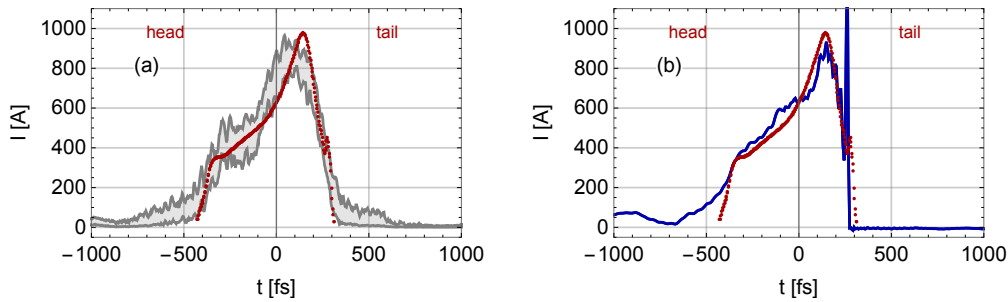

Figure 36: Reconstructed shapes of the type-1 bunches. Parameters: electron energy $E_e = 700$ MeV, bunch charge $Q = 390$ pC, TDS resolution $\sigma_{\mathrm{TDS}} = 29$ fs. Red curves: TDS measurement. **(a)** Iterative phase retrieval with quasi-random initial phases. **(b)** Analytic phase retrieval using the Kramers-Kronig phase.

longitudinal form factor have already been shown in Fig. 34. The iterative reconstruction with quasi-random start phases yields time profiles which are in good agreement with the TDS profile, see Fig. 36. However, the

analytic KK reconstruction generates a profile with an extremely sharp spike in the tail region. The TDS resolution is insufficient to decide whether this spike is real or an artefact.

Example 2

In another measurement series bunches were produced with a steep rise and a roughly exponential decay, resembling Akutowicz's function $f_1(t)$. The TDS time profile and the form factor are shown in Fig. 37. On a logarithmic scale, $F(\omega)$ exhibits a smooth drop over 1.5 orders of magnitude. The form factor is well above the 4σ limit of the noise and is measurable up to about 40 THz. The data in the 40 - 50 THz region have large point-to-point fluctuations and are of little use in signal reconstruction. The reconstructed time profiles $I(t)$
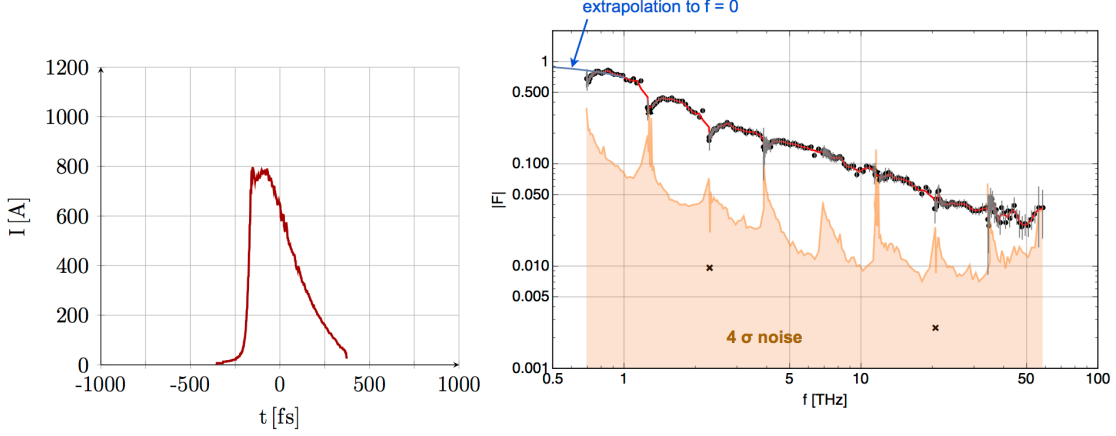

Figure 37: TDS time profile and measured form factor $F(\omega)$ of type-2 bunches (gray dots), plotted as a function of $f = \omega/2\pi$, and the smooth spline-interpolation (red curve) which suppresses strong local fluctuations around 50 THz. The shaded area shows the 4σ limit of the noise signal. Also shown is the extrapolation to $f = 0$ (blue line).
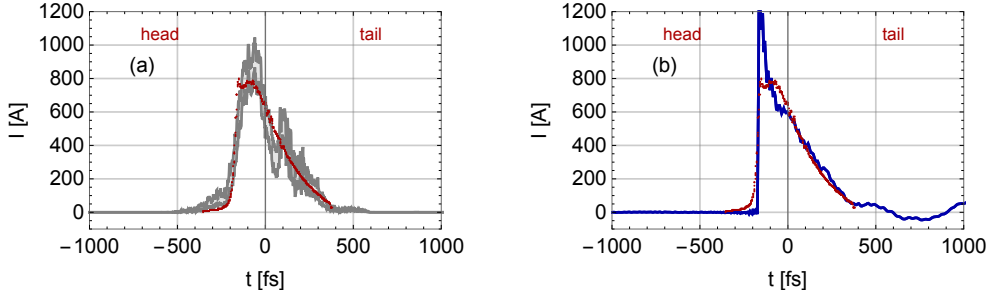

Figure 38: Reconstructed shapes of bunches with a steep rise and a roughly exponential decay. Parameters: electron energy $E_e = 540\,\text{MeV}$, bunch charge $Q = 240\,\text{pC}$, TDS resolution $\sigma_{\text{TDS}} = 49\,\text{fs}$. The bunch current $I(t)$ is plotted as a function of time. The bunch head is at the left side (early times). Red curves: TDS measurement (compare Fig. 35b). **(a)** Iterative phase retrieval with quasi-random initial phases. **(b)** Analytic phase retrieval using the Kramers-Kronig phase.

of the bunch current are presented in Fig. 38. The TDS profile (red curve) reveals that the steep rise occurs at the bunch head (early times) while the exponential drop is in the tail region. The iterative reconstruction with quasi-random start phases yields time profiles with a fairly large uncertainty band, and the steep rise is washed out. This is a general observation: the iterative method with random start-phases tends to wash out sharp structures, especially after averaging over many repetitions. The analytic reconstruction with the KK phase works very well, which may not be a surprise in view of the excellent KK reconstruction of the Akutowic function $f_1(t)$ (Fig. 14a). The initial rise is even steeper than in the TDS time profile.

Example 3

Example 3 illustrates the response to a rather long bunch. The measured form factor (Fig. 39) drops rapidly with frequency, falling below 0.1 slightly above 1 THz. In the IR regime (above 5 THz) it oscillates rather strongly from shot to shot and hardly exceeds the sensitivity limit. Nevertheless, the reconstruction of the bunch profile is possible with both the Kramers-Kronig and the iterative method. The overall shape and bunch length are again in very good agreement with the TDS measurement. As expected, the Kramers-Kronig phase compiles the high frequency content to sharp structures at the end of the bunch tail while the iterative reconstruction results in a smoother average profile. The total bunch length of about 1 ps reaches the upper limit for this bunch charge.
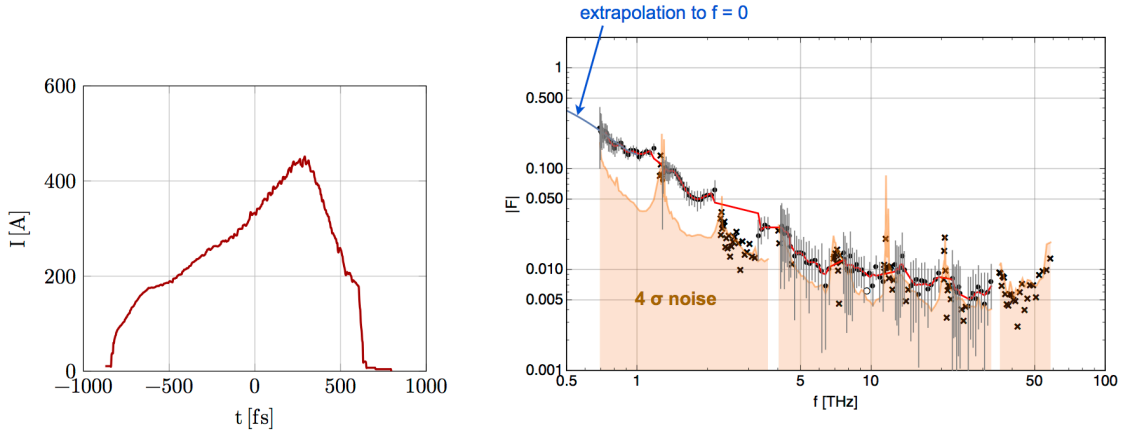

Figure 39: TDS time profile and form factor of bunch type 3. Remark on the extrapolation to $f = 0$ (blue curve): the frequency axis is logarithmic and the minimum frequency in this figure is $f_{\min} = 0.5$ THz. The grey vertical lines, which become very pronounced above 5 THz, indicate the shot-to-shot fluctuations of the measurement.
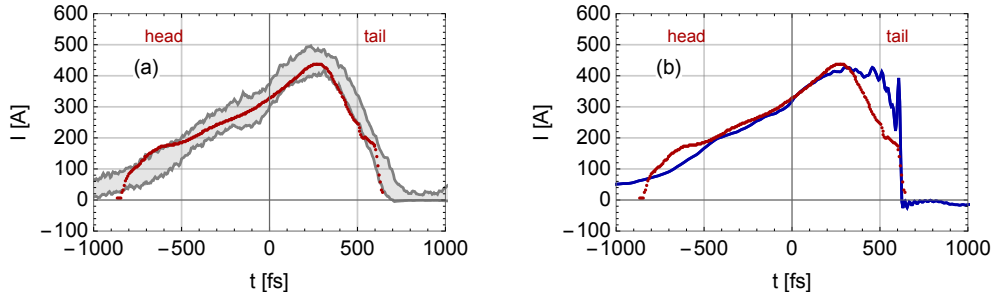

Figure 40: Reconstructed shapes of bunch type 3. Parameters: electron energy $E_e = 710$ MeV, bunch charge $Q = 400$ pC, TDS resolution $\sigma_{\mathrm{TDS}} = 45$ fs. Red curves: TDS measurement. **(a)** Iterative phase retrieval with quasi-random initial phases. **(b)** Analytic phase retrieval using the Kramers-Kronig phase.

Example 4

Example 4 finally demonstrates the possibility to measure extremely short bunches of very low charge. The total charge was only 14 pC but the bunch was highly compressed. The electron energy was $E_e = 706$ MeV. The form factor (Fig. 42) falls off very slowly, by just a factor of two between 1 THz and 40 THz. It is still well above 0.1 for the highest measured frequency of 60 THz. The reconstructed bunch profiles (Fig. 41) show a narrow peak with a FWHM of about 8 fs. The two reconstruction algorithms are in good agreement. A comparison with the TDS is meaningless in this case since the bunch length is less than the resolution limit of the TDS (about 10 fs) even with optimized accelerator optics.
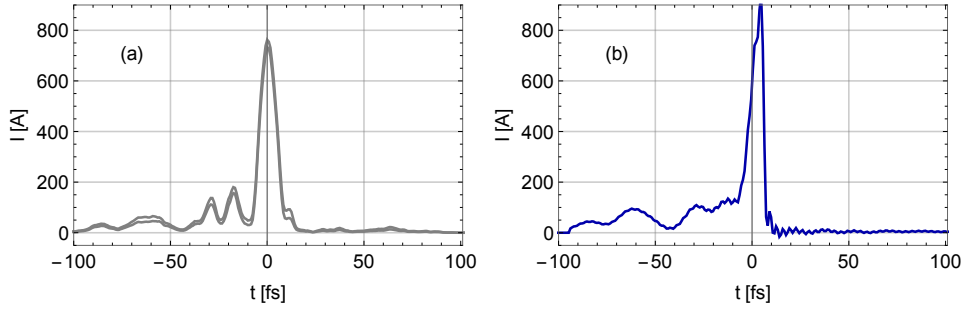
Figure 41: Reconstructed shapes of an ultrashort bunch (type 4). The full width at half maximum is 8 fs. This narrow peak cannot be resolved by the 2.86 GHz TDS that is presently installed at FLASH. **(a)** Iterative phase retrieval with quasi-random initial phases. **(b)** Analytic phase retrieval using the Kramers-Kronig phase.
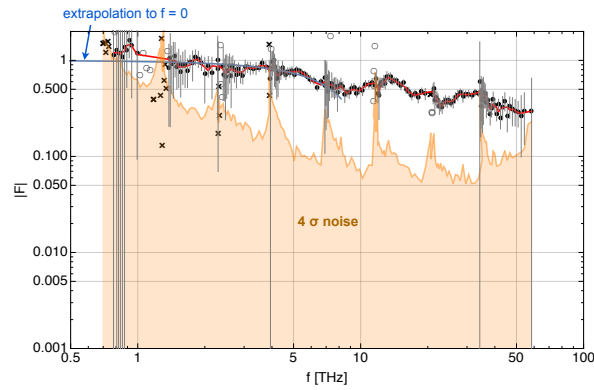

Figure 42: The form factor of the ultrashort bunch drops very slowly with frequency.

Impact of averaging

The averaging method, applied by us in the iterative reconstruction starting with quasi-random phases, has a tendency to smear out steep slopes. This is illustrated in Fig. 43 where we compare the averaged profile of type-1 bunches with two single-cycle iterations.
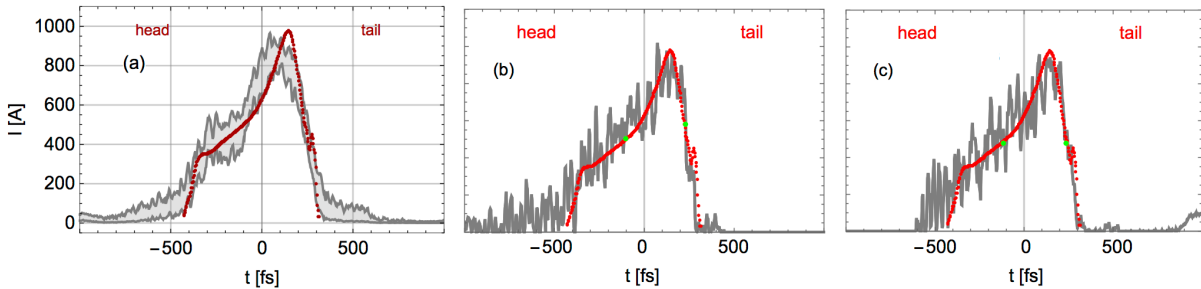

Figure 43: Reconstructed shapes of the type-1 bunches, starting with quasi-random phases. Red curves: TDS measurement. **(a)** Average profile of 100 iteration cycles. **(b, c)** Two single-cycle profiles. The rear slope is definitely steeper in the single-cycle cases and agrees well with the slope of the TDS profile.

36

5.4 Summary and conclusions

Our four examples demonstrate the great potential of broadband CTR spectroscopy for longitudinal bunch diagnostics. The bunch shapes that have been reconstructed from the measured form factors are in good agreement with the shapes determined by the time-domain instrument TDS. The inverse Fourier transformation using the Kramers-Kronig phase yields good results but the same is true for the iterative method starting either from the KK phase or from randomized phases. The ambiguities which occur frequently in the model calculations are not observed here, probably because the investigated bunches feature essentially only a single peak. A remarkable result is that the spectroscopic method enables the resolution of very short time structures (below 10 fs) where the present 2.86 GHz TDS meets its resolution limit. This is evident in Fig. 41.

We have shown that the frequency-domain technique of coherent transition radiation spectroscopy is highly competitive with high-resolution time-domain techniques, provided a broadband multi-channel spectrometer with single-shot capability is used. Both methods complement each other and their combination is vital for obtaining faithful results. Dedicated studies are underway to exploit these possibilities.

6 Appendix A: Dispersion Relations

A dispersion relation is an integral formula relating a dispersive process to an absorptive process. An example is the relation between the refractive index $n(\omega)$ of an optical medium and its extinction coefficient $k(\omega)$, or the relation between real part and imaginary part of the dielectric function of a solid. Dispersion relations follow rigorously from causality. In this appendix we follow closely the book *Optical Properties of Solids* by F. Wooten [21] but we go into more detail and present several mathematical proofs that are missing in [21]. A comprehensive treatment can be found in [36].

6.1 Basics of complex analysis

Cauchy-Riemann equations
The set of complex numbers $z = x + iy$ is called **C**. These numbers can be depicted as points in a plane (x, iy). A function $f(z)$ is called analytic (or holomorphic) in an open subset U of **C** if the differential quotient exists

$$\frac{df}{dz} = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \quad \text{for all} \ \ z \in U. \tag{50}$$

The differential quotient is defined in the same way as in real analysis, but the limit can be approached from many different directions. This has far-reaching consequences. Separating the complex function into its real and imaginary parts

$$f(z) = f(x + iy) = u(x, y) + iv(x, y) \tag{51}$$

one can prove that $f(z)$ is analytic if and only if the **Cauchy-Riemann differential equations** are fulfilled

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{52}$$

Analytic functions have many remarkable properties that are not valid for real functions. For example the derivative of an analytic function is again analytic which means that analytic functions are infinitely often differentiable.

Cauchy Integral Theorem and Residue Theorem
The Cauchy integral theorem is an important statement about line integrals in the complex plane **C**. If two different paths connect the same start and end points, and if the function is analytic in an open set containing the two paths, then these two path integrals of the function yield the same value. The theorem is usually formulated for closed paths:
Let U be an open subset of **C** which is simply connected, let $f : U \to \mathbf{C}$ be an analytic function, and let Γ be a closed loop. Then the line integral over the closed loop vanishes

$$\oint_\Gamma f(z)dz = 0 \qquad \qquad \textit{Cauchy Integral Theorem}. \tag{53}$$

Consider now the function $g(z) = f(z)/(z - z_0)$ where z_0 is an arbitrary point inside the closed loop Γ. The function $g(z)$ is analytic except for a small vicinity around the pole at z_0. The Residue Theorem states

$$\oint_\Gamma \frac{f(z)}{z - z_0}dz = 2\pi\, if(z_0) \equiv 2\pi\, i \operatorname{Res}(f, z_0) \quad \textit{Residue Theorem}. \tag{54}$$

This is easy to verify for a small circle centered at z_0 whose radius a tends to zero. On the circle we have

$$z = z_0 + ae^{i\theta}, \ \ dz = i\,ae^{i\theta}d\theta$$

$$\lim_{a \to 0} \oint_{\text{circ}} \frac{f(z)}{z - z_0}dz = f(z_0) \lim_{a \to 0} \int_0^{2\pi} \frac{1}{ae^{i\theta}} i\,ae^{i\theta}d\theta = 2\pi\, if(z_0).$$

Computation of an important integral
In the next section 6.2 an integral of the type

$$\int_{-\infty}^{\infty} \frac{e^{ix}}{x - x_0}dx$$

38

appears. The integrand has a singularity at $x = x_0$. We want to show how such integrals along the real axis can be evaluated by going into the complex plane and using the Cauchy and Residue Theorems. For this purpose we consider a closed integration path Γ consisting of three parts (see Fig. 44):

(1) A large semicircle Γ_1 of radius R which is centered at the origin $z = 0$,
(2) a straight line Γ_2 along the real axis from $x = -R$ to $x = x_0 - \varepsilon$ and from $x = x_0 + \varepsilon$ to $x = +R$,
(3) a small semicircle Γ_3 of radius ε which is centered at x_0.

Step 1 Consider a function $f(z)$ which is analytic in the upper half plane $y \geq 0$, except for a finite number of poles, and which vanishes asymptotically

$$\lim_{|z|\to\infty} f(z) = 0 \quad \text{for} \quad y \geq 0 \,. \tag{55}$$

Statement: The line integral of the function $f(z)e^{iz}$ along the semicircle Γ_1 tends to zero in the limit $R \to \infty$:

$$\lim_{R\to\infty} \int_{\Gamma_1} f(z)e^{iz}dz = 0 \,. \tag{56}$$

This statement is by no means obvious. Because of (55), the integrand tends to zero with increasing radius of the semicircle, $|f(z)e^{iz}| \to 0$ for $|z| \to \infty$ and $y > 0$, but at the same time the path length of the semicircle tends to infinity.

Proof: The proof of (56) goes as follows (see H. Cartan [37]). On the semicircle we have

$$z = Re^{i\theta} \,, \quad dz = iRe^{i\theta}d\theta \,, \quad |e^{iz}| = e^{-R\sin\theta} \,.$$

An upper limit for the integral is

$$\left| \int_{\Gamma_1} f(z)e^{iz}dz \right| \leq \int_0^\pi |f(Re^{i\theta})|e^{-R\sin\theta}Rd\theta \leq M(R)\int_0^\pi e^{-R\sin\theta}Rd\theta = 2M(R)\int_0^{\pi/2} e^{-R\sin\theta}Rd\theta$$

with

$$M(R) = \max_{0\leq\theta\leq\pi} (|f(Re^{i\theta})|) \,.$$

In the interval $0 \leq \theta \leq \pi/2$ one has $\sin\theta \geq 2\theta/\pi$, hence

$$\int_0^{\pi/2} e^{-R\sin\theta}Rd\theta \leq \int_0^{\pi/2} e^{-2\theta R/\pi}Rd\theta = \pi/2 \,.$$

The result is

$$\left| \int_{\Gamma_1} f(z)e^{iz}dz \right| \leq \pi M(R) \quad \text{and} \quad \lim_{R\to\infty} M(R) = 0 \,.$$

Therefore

$$\lim_{R\to\infty} \int_{\Gamma_1} f(z)e^{iz}dz = 0 \qquad \text{qed} \,.$$

Specifically, the function $f(z) = 1/(z - x_0)$ fulfills the condition (55). Hence

$$\lim_{R\to\infty} \int_{\Gamma_1} \frac{e^{iz}}{z - x_0}dz = 0 \,. \tag{57}$$

Step 2 Next we want to evaluate the closed-loop integral

$$\oint_\Gamma \frac{e^{iz}}{z - x_0}dz \,.$$

The function $1/(z - x_0)$ has a pole at $z = x_0$ but is analytic elsewhere. Next we show that e^{iz} is analytic in the entire complex plane. We prove this for the more general case $e^{iz\tau}$ where τ is a real number. For this purpose we write
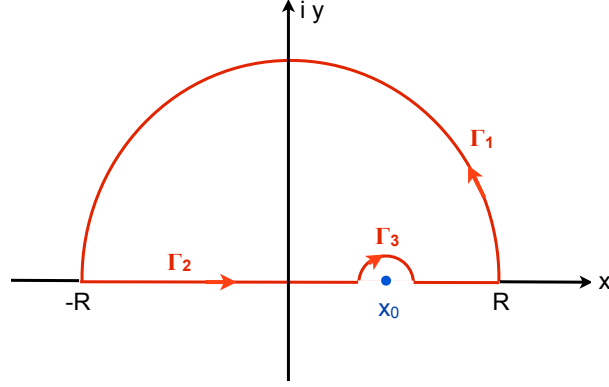
$$e^{iz\tau} = e^{i(x+iy)\tau} = u(x, y) + i\,v(x, y) \,.$$

Figure 44: The closed integration path Γ.

The real functions $u(x,y)$ and $v(x,y)$ are

$$u(x,y) = \cos(x\tau)\, e^{-y\tau}, \qquad v(x,y) = \sin(x\tau)\, e^{-y\tau}\,.$$

It is easy to verify that they fulfill the Cauchy-Riemann differential equations

$$\frac{\partial u}{\partial x} = -\tau\,\sin(x\tau)\,e^{-y\tau} = \frac{\partial v}{\partial y}\,, \qquad \frac{\partial u}{\partial y} = -\tau\,\cos(x\tau)\,e^{-y\tau} = -\frac{\partial v}{\partial x}\,.$$

This proves that $f(z) = e^{iz\tau} = e^{i(x+iy)\tau} = u(x,y) + i\,v(x,y)$ is an analytic function of the complex variable $z = x + i\,y$.

The pole at x_0 is avoided by choosing the closed integration path Γ shown in Fig. 44. There is no singularity inside the loop, hence from Cauchy's formula (53)

$$\oint_\Gamma \frac{e^{iz}}{z-x_0}dz = \int_{\Gamma_1} \frac{e^{iz}}{z-x_0}dz + \int_{\Gamma_2} \frac{e^{iz}}{z-x_0}dz + \int_{\Gamma_3} \frac{e^{iz}}{z-x_0}dz = 0\,.$$

The three path integrals are evaluated in the limit $\varepsilon \to 0$ and $R \to \infty$.
Because of (57) we get

$$\int_{\Gamma_2} \frac{e^{iz}}{z-x_0}dz = -\int_{\Gamma_3} \frac{e^{iz}}{z-x_0}dz\,.$$

The integral over the small semicircle is readily evaluated

$$z = x_0 + \varepsilon\,e^{i\theta}, \quad dz = i\,\varepsilon\,e^{i\theta}d\theta$$

$$\lim_{\varepsilon\to 0}\int_{\Gamma_3} \frac{e^{iz}}{z-x_0}dz = e^{ix_0}\lim_{\varepsilon\to 0}\int_{2\pi}^{\pi} \frac{1}{\varepsilon\,e^{i\theta}}\,i\,\varepsilon e^{i\theta}d\theta = -\pi\,i\,e^{ix_0}\,. \tag{58}$$

The integral over the path Γ_2 is therefore in the limit $R \to \infty$ and $\varepsilon \to 0$

$$\int_{\Gamma_2} \frac{e^{iz}}{z-x_0}dz = \lim_{\varepsilon\to 0}\left[\int_{-\infty}^{x_0-\varepsilon} \frac{e^{iz}}{z-x_0}dz + \int_{x_0+\varepsilon}^{\infty} \frac{e^{iz}}{z-x_0}dz\right] = i\,\pi\,e^{ix_0}\,.$$

The path Γ_2 is along the real axis, hence $z = x$ on Γ_2. It is customary to define the principal value of an integral, denoted by the letter $\mathcal{P}$, by approaching the singularity at x_0 symmetrically from both sides:

$$\mathcal{P}\int_{-\infty}^{\infty} \frac{e^{ix}}{x-x_0}dx = \lim_{\varepsilon\to 0}\left[\int_{-\infty}^{x_0-\varepsilon} \frac{e^{ix}}{x-x_0}dx + \int_{x_0+\varepsilon}^{\infty} \frac{e^{ix}}{x-x_0}dx\right] \tag{59}$$

The important result is

$$\mathcal{P}\int_{-\infty}^{\infty} \frac{e^{ix}}{x-x_0}dx = i\,\pi\,e^{ix_0}\,. \tag{60}$$

6.2 Dispersion relations as a consequence of causality

Linear response functions

The response X of a *linear* system depends linearly on the stimulus S. The response can be calculated by a convolution integral

$$X(t) = \int_{-\infty}^{+\infty} G(t - t')S(t')dt' . \tag{61}$$

$G(t - t')$ is called the response function. An example of a linear system is a dielectric medium in which the induced polarization (the response) depends linearly on the applied electric field (the stimulus):

$$\boldsymbol{P} = \chi_e \varepsilon_0 \boldsymbol{E} .$$

The electric susceptibility χ_e is the response function in this case. (Nonlinear systems, for example nonlinear crystals for frequency doubling of laser light, are not treated here).

Causality puts an important constraint on the response function. Consider for example an electromagnetic wave pulse hitting the surface of a dielectric slab where it is reflected. The stimulus is the incident pulse, the response is the reflected pulse, and both are related by (61). There cannot be a reflected pulse before the incident pulse arrives, hence we must request

$$G(t - t') = G(\tau) = 0 \qquad \text{for} \quad \tau = t - t' < 0 . \tag{62}$$

The Fourier transform of the response function

$$\tilde{G}(\omega) = \int_0^\infty G(\tau)e^{i\omega\tau}d\tau . \tag{63}$$

is in general a complex-valued function.

Complex frequency plane

Using the theory of analytic functions we want to derive a *dispersion relation* between real part and imaginary part of $\tilde{G}(\omega)$. For that purpose we define complex frequencies

$$\hat{\omega} = \omega_\mathrm{r} + i\,\omega_\mathrm{i}$$

The Fourier integral (63) defines $\tilde{G}(\omega)$ as a function of the real variable ω. We generalize this definition to comprise complex frequencies as well:

$$\tilde{G}(\hat{\omega}) = \tilde{G}(\omega_\mathrm{r} + i\,\omega_\mathrm{i}) = \int_0^\infty G(\tau)e^{i\omega_\mathrm{r}\tau}e^{-\omega_\mathrm{i}\tau}d\tau . \tag{64}$$

We know from the previous section that $\exp(i\hat{\omega}\tau)$ is an analytic function of the complex variable $\hat{\omega} = \omega_\mathrm{r} + i\,\omega_\mathrm{i}$, and hence the response function $\tilde{G}(\hat{\omega})$ is analytic. Moreover, because of $\tau \geq 0$, $|\tilde{G}(\hat{\omega})|$ is bounded in the upper half of the complex $\hat{\omega}$ plane and tends to zero for $\omega_\mathrm{i} \to \infty$. Therefore the line integral along the semicircle Γ_1 tends to zero in the limit $R \to \infty$, and we are allowed to use Eq. (60):

$$\mathcal{P} \int_{-\infty}^\infty \frac{\tilde{G}(\omega)}{\omega - \omega_0}d\omega = i\pi\tilde{G}(\omega_0) . \tag{65}$$

Separating real and imaginary part we obtain the dispersion relations

$$\Re(\tilde{G}(\omega_0)) = \frac{1}{\pi}\mathcal{P} \int_{-\infty}^\infty \frac{\Im(\tilde{G}(\omega))}{\omega - \omega_0}d\omega , \quad \Im(\tilde{G}(\omega_0)) = -\frac{1}{\pi}\mathcal{P} \int_{-\infty}^\infty \frac{\Re(\tilde{G}(\omega))}{\omega - \omega_0}d\omega \tag{66}$$

which show that real part and imaginary part of the response function are intimately connected. The real part can be computed by a principal value integral over the imaginary part, and the imaginary part can be computed by a principal value integral over the real part.

An important application is the relation between real part and imaginary part of the electric susceptibility $\chi_e(\omega) = \varepsilon(\omega) - 1 = \varepsilon_1(\omega) - 1 + i\,\varepsilon_2(\omega)$ of a solid (see [21])

$$\varepsilon_1(\omega) - 1 = \frac{1}{\pi}\mathcal{P} \int_{-\infty}^\infty \frac{\varepsilon_2(\omega')}{\omega' - \omega}d\omega' , \quad \varepsilon_2(\omega) = -\frac{1}{\pi}\mathcal{P} \int_{-\infty}^\infty \frac{\varepsilon_1(\omega') - 1}{\omega' - \omega}d\omega' . \tag{67}$$

These are the famous Kramers-Kronig dispersion relations for the dielectric function of a solid body. In the above formulas we have replaced ω_0 by ω and renamed the integration variable from ω to ω'.

Truncated bunch shapes

In order to satisfy the causality condition in our model calculations on bunch shape reconstruction, we restrict ourselves to truncated functions by requiring that the particle density in the bunch vanishes identically below a time threshold: $\rho(t) \equiv 0$ for all times $t < t_{\min}$. An example is the cosine-squared wave (45) shown in Fig. 14. Simple other choices are truncated rectangular or triangular functions.

Note that a Gaussian function violates causality because it extends over the full range $-\infty < t < +\infty$. Hence it may not be a surprise that we encounter problems with truly Gaussian-shaped time profiles. A truncated Gaussian preserves causality, this important example will be studied in sect. 6.5.2 and Appendix C.

6.3   Alternative proof of the dispersion relations between real and imaginary part

We discuss now an alternative proof of the dispersion relations between real and imaginary part which emphasizes the fundamental role of causality. Any real function $f(t)$ can be expressed as the sum of an even and an odd function:

$$f(t) = f_{\text{even}}(t) + f_{\text{odd}}(t) \quad \text{with} \quad f_{\text{even}}(t) = (f(t) + f(-t))/2, \quad f_{\text{odd}}(t) = (f(t) - f(-t))/2.$$

Now we impose the requirement of causality and request that $f(t)$ vanishes for all $t < 0$. Then the following relations hold

$$f_{\text{even}}(t) = S(t)\, f_{\text{odd}}(t), \quad f_{\text{odd}}(t) = S(t)\, f_{\text{even}}(t) \tag{68}$$

where $S(t)$ is the signum function

$$S(t) = -1 \quad \text{for} \quad t < 0, \quad S(t) = +1 \quad \text{for} \quad t > 0.$$

The signum function is the limiting case of a quadratically integrable function

$$S(t) = \lim_{a \to 0} S_a(t) \quad \text{with} \quad S_a(t) = -e^{at} \text{ for } t < 0, \quad S_a(t) = e^{-at} \text{ for } t > 0.$$

The Fourier transform of $S_a(t)$ can be easily computed

$$\tilde{S}_a(\omega) = \int_{-\infty}^{\infty} S_a(t) e^{i\omega t} dt = -\int_{-\infty}^{0} e^{at} e^{i\omega t} dt + \int_{0}^{+\infty} e^{-at} e^{i\omega t} dt = -\frac{1}{a + i\omega} - \frac{1}{-a + i\omega},$$

so the Fourier transform of the signum function is

$$\tilde{S}(\omega) = \lim_{a \to 0} \tilde{S}_a(\omega) = \frac{2\,i}{\omega}. \tag{69}$$

From $e^{i\omega t} = \cos \omega t + i \sin \omega t$ follows that an even real function $f_{\text{even}}(t)$ has a real Fourier transform, and an odd real function $f_{\text{odd}}(t)$ has a purely imaginary Fourier transform $\tilde{f}_{\text{odd}}(\omega)$. Therefore we obtain, using (68)

$$\Re(\tilde{f}(\omega)) = \tilde{f}_{\text{even}}(\omega) = \mathcal{FT}\{S(t)\, f_{\text{odd}}(t)\}, \quad \Im(\tilde{f}(\omega)) = \frac{1}{i}\, \tilde{f}_{\text{odd}}(\omega) = \frac{1}{i}\, \mathcal{FT}\{S(t)\, f_{\text{even}}(t)\} \tag{70}$$

where $\mathcal{FT}\{S(t)\, f(t)\}$ denotes the Fourier transform of the product function $S(t)\, f_{\text{odd}}(t)$, and $\mathcal{FT}\{S(t)\, f_{\text{even}}(t)\}$ the Fourier transform of $S(t)\, f_{\text{even}}(t)$.

The Fourier transform of the product of two time-domain functions is evaluated using the following mathematical theorem:

Theorem: Let $f(t)$ and $g(t)$ be quadratically integrable functions with Fourier transforms $\tilde{f}(\omega)$ and $\tilde{g}(\omega)$. The Fourier transform of the product function $h(t) = f(t)\, g(t)$ is given by the convolution integral of $\tilde{f}(\omega)$ and $\tilde{g}(\omega)$:

$$\tilde{h}(\omega_0) = \frac{1}{2\pi} \int \tilde{f}(\omega_0 - \omega) \tilde{g}(\omega) d\omega. \tag{71}$$

42

Proof: The Fourier transform of $h(t) = f(t)\,g(t)$ is defined by

$$\tilde{h}(\omega_0) = \int f(t)g(t)\,e^{i\omega_0 t}dt\,.$$

Inserting for $g(t)$ its inverse Fourier transform

$$g(t) = \frac{1}{2\pi}\int \tilde{g}(\omega)\,e^{-i\omega t}d\omega$$

one finds

$$\tilde{h}(\omega_0) = \frac{1}{2\pi}\int\left\{\int f(t)e^{i(\omega_0-\omega)t}dt\cdot\tilde{g}(\omega)\right\}d\omega = \frac{1}{2\pi}\int\tilde{f}(\omega_0-\omega)\tilde{g}(\omega)d\omega \quad\text{qed.}$$

Applying this theorem in the equations (70) we get

$$\Re(\tilde{f}(\omega_0)) = \tilde{f}_{\text{even}}(\omega_0) = \frac{1}{2\pi}\int\tilde{S}(\omega_0-\omega)\tilde{f}_{\text{odd}}(\omega)d\omega = \frac{i}{2\pi}\int\tilde{S}(\omega_0-\omega)\Im(\tilde{f}(\omega))d\omega\,,$$

$$\Im(\tilde{f}(\omega_0)) = \frac{1}{i}\tilde{f}_{\text{odd}}(\omega_0) = -\frac{i}{2\pi}\int\tilde{S}(\omega_0-\omega)\tilde{f}_{\text{even}}(\omega)d\omega = -\frac{i}{2\pi}\int\tilde{S}(\omega_0-\omega)\Re(\tilde{f}(\omega))d\omega\,.$$

Inserting the Fourier transform of the signum function, $\tilde{S}(\omega_0-\omega) = 2i/(\omega_0-\omega)$, we finally obtain the dispersion relations (66)

$$\Re(\tilde{f}(\omega_0)) = \frac{1}{\pi}\int\frac{\Im(\tilde{f}(\omega))}{\omega-\omega_0}d\omega\,, \quad \Im(\tilde{f}(\omega_0)) = -\frac{1}{\pi}\int\frac{\Re(\tilde{f}(\omega))}{\omega-\omega_0}d\omega\,. \tag{72}$$

6.4 Dispersion relation for absolute magnitude and phase

The dispersion relation (66) requires the knowledge of either the real part or the imaginary part of the response function. However, in many cases only the absolute magnitude is known, so it is desirable to derive a relation allowing to compute the phase. Wooten [21] discusses the mathematical procedures for the reflection of an electromagnetic wave from a solid. A similar task arises for the complex form factor

$$\mathcal{F}(\omega) = F(\omega)\exp(i\Phi(\omega))\,.$$

Here ω is a real variable. The spectral intensity emitted by an electron bunch via the processes of transition, diffraction or synchrotron radiation is recorded with a spectrometer or another device measuring just the intensity. From these data one derives the absolute square of the bunch form factor, $|\mathcal{F}(\omega)|^2$. To determine the phase $\Phi(\omega)$, we adopt Wooten's treatment of the complex reflectivity and derive a dispersion relation for

$$\ln(\mathcal{F}(\omega)) = \ln(F(\omega)) + i\Phi(\omega)\,. \tag{73}$$

Now we continue this expression into the complex $\hat{\omega}$ plane. We want to apply the Residue Theorem for the closed loop shown in Fig. 45. This needs a lot of work, and some problems arise.

Prerequisites for the applicability of the Residue Theorem
(1) $\ln(\mathcal{F}(\hat{\omega}))$ is not allowed to have a singularity in the upper half plane. Therefore we must require $\mathcal{F}(\hat{\omega}) \neq 0$ for any $\hat{\omega}$ with a positive imaginary part. This condition, however, is violated in many cases. Complex zeros in the upper half plane may indeed exist, and they have a strong impact on the phase determination. We show in Appendix A, sect. 6.5.3 how these zeros can be handled and demonstrate the mathematical procedure with many examples in Appendix B.

(2) A severe problem is that the form factor drops to zero at infinite frequency, hence $\ln(\mathcal{F}(\hat{\omega}))$ diverges for $|\hat{\omega}| \to \infty$. This means that the line integral of $\ln(\mathcal{F}(\hat{\omega}))$ over the large semicircle Γ_1 diverges as well. To circumvent this difficulty one defines the following auxiliary function

$$f_{\text{aux}}(\hat{\omega}) = \frac{(1+\omega_0\hat{\omega})\ln(\mathcal{F}(\hat{\omega}))}{(1+\hat{\omega}^2)(\omega_0-\hat{\omega})} = \frac{1}{\hat{\omega}-i}\frac{(1+\omega_0\hat{\omega})\ln(\mathcal{F}(\hat{\omega}))}{(\hat{\omega}+i)(\omega_0-\hat{\omega})} \tag{74}$$
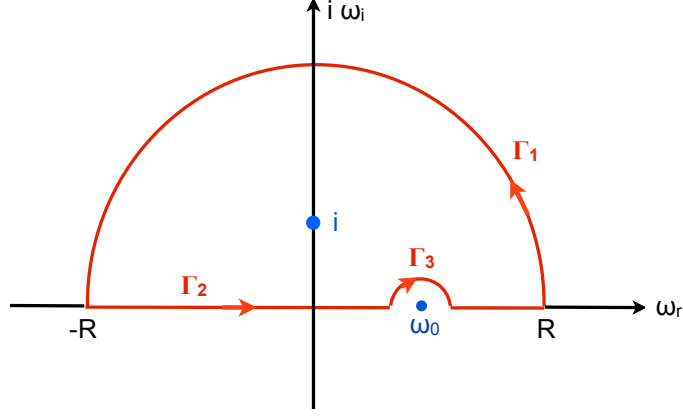
Figure 45: The closed integration path Γ for the auxiliary function (74). The difference to Fig. 44 is the pole on the imaginary axis at $\hat{\omega} = i$.

with a real positive frequency ω_0. Following [21] we furthermore make the assumption that asymptotically $\mathcal{F}(\hat{\omega})$ obeys an inverse power law:

$$\mathcal{F}(\hat{\omega}) = b\,\hat{\omega}^{-s} \quad \text{with a positive number } s \quad (|\hat{\omega}| \to \infty). \tag{75}$$

Then the function f_{aux} behaves asymptotically as

$$f_{\text{aux}}(\hat{\omega}) \approx \frac{(1 + \omega_0\hat{\omega})(-s\ln(\hat{\omega}) + \ln(b))}{(1 + \hat{\omega}^2)(\omega_0 - \hat{\omega})}\,.$$

Condition (75) is met by the form factor (46) of the cosine-squared wave cycle and the form factors of other truncated time-domain functions. However, it is violated for Gaussian form factors, as will be shown later.

If (75) is satisfied one gets on the large semicircle Γ_1

$$\hat{\omega} = Re^{i\theta}, \quad |f_{\text{aux}}(\hat{\omega})| \approx s\,\omega_0 \left|\frac{\ln(R)}{R^2}\right|\,.$$

The pathlength of Γ_1 is πR, hence the integral of $f_{\text{aux}}(\hat{\omega})$ over this semicircle vanishes if R tends to infinity:

$$\int_{\Gamma_1} f_{\text{aux}}(\hat{\omega})d\hat{\omega} \approx \pi\,s\,\omega_0 \left|\frac{\ln(R)}{R}\right| \to 0\,. \tag{76}$$

Next we compute the integral over the small semicircle Γ_3 which is centered at the real frequency ω_0 and has an infinitesimal radius ε. On the semicircle we have

$$\hat{\omega} = \omega_0 + \varepsilon\,e^{i\theta}, \quad d\hat{\omega} = i\,\varepsilon\,e^{i\theta}d\theta, \quad \lim_{\varepsilon \to 0} f_{\text{aux}}(\hat{\omega}) = \frac{\ln(\mathcal{F}(\omega_0))}{\omega_0 - \hat{\omega}}\,.$$

In analogy to Eq. (58) we find therefore

$$\begin{aligned}
\lim_{\varepsilon \to 0}\int_{\Gamma_3} f_{\text{aux}}(\hat{\omega})d\hat{\omega} &= \ln(\mathcal{F}(\omega_0))\lim_{\varepsilon \to 0}\int_{\pi}^{0} \frac{1}{(-\varepsilon\,e^{i\theta})}\,i\varepsilon e^{i\theta}d\theta = i\,\pi\,\ln(\mathcal{F}(\omega_0)) \\
&= i\,\pi\,\ln(F(\omega_0)) - \pi\,\Phi(\omega_0)\,. \tag{77}
\end{aligned}$$

Performing the integration of $f_{\text{aux}}(\hat{\omega})$ along the closed loop shown in Fig. 45 we have to keep in mind that the auxiliary function has a pole at $\hat{\omega} = i$ inside the loop. The Residue Theorem yields

$$\oint_{\Gamma} f_{\text{aux}}(\hat{\omega})d\hat{\omega} = \int_{\Gamma_1} f_{\text{aux}}(\hat{\omega})d\hat{\omega} + \int_{\Gamma_2} f_{\text{aux}}(\hat{\omega})d\hat{\omega} + \int_{\Gamma_3} f_{\text{aux}}(\hat{\omega})d\hat{\omega} = 2\pi i\frac{1 + i\omega_0}{2i(\omega_0 - i)}\ln(\mathcal{F}(i)) = i\pi\ln(\mathcal{F}(i))\,.$$

44

What is $\mathcal{F}(i)$? From the definition (33) of the form factor follows that $\mathcal{F}(i)$ is a real number which we call α:

$$\mathcal{F}(i) = \int_{-\infty}^{\infty} \rho(t)\exp(-t)dt = \alpha.$$

Putting things together, using $\ln(\mathcal{F}(\omega)) = \ln(F(\omega)) + i\Phi(\omega)$, and separating real and imaginary parts we obtain for the integral along the path Γ_2 (this is the real axis, hence we write ω instead of $\hat\omega$):

$$\int_{\Gamma_2} f_{\text{aux}}(\omega)d\omega \equiv \mathcal{P}\int_{-\infty}^{\infty}\left[\frac{(1+\omega_0\omega)\ln(F(\omega))}{(1+\omega^2)(\omega_0-\omega)} + i\,\frac{(1+\omega_0\omega)\Phi(\omega)}{(1+\omega^2)(\omega_0-\omega)}\right]d\omega = -i\pi\ln(F(\omega_0)) + \pi\Phi(\omega_0) + i\pi\ln(\alpha).$$

The real part of this equation yields a first form of the desired dispersion relation

$$\Phi(\omega_0) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{(1+\omega_0\omega)\ln(F(\omega))}{(1+\omega^2)(\omega_0-\omega)}d\omega. \tag{78}$$

Restriction to positive frequencies
The particle density $\rho(t)$ is a real function while the Fourier transform $\mathcal{F}(\omega)$ is a complex-valued function obeying the rule (36):

$$\mathcal{F}^*(\omega) = \mathcal{F}(-\omega) \quad\Rightarrow\quad F(-\omega) = F(\omega),\quad \Phi(-\omega) = -\Phi(\omega).$$

The principal value integral is split up in two parts

$$\mathcal{P}\int_{-\infty}^{\infty} f_{\text{aux}}(\omega)d\omega = \mathcal{P}\int_{-\infty}^{0} f_{\text{aux}}(\omega)d\omega + \mathcal{P}\int_{0}^{\infty} f_{\text{aux}}(\omega)d\omega.$$

In the integration over negative frequencies we make the replacement $u = -\omega$ and use $F(-u) = F(u)$

$$\mathcal{P}\int_{-\infty}^{0}\frac{(1+\omega_0\omega)\ln(F(\omega))}{(1+\omega^2)(\omega_0-\omega)}d\omega = \mathcal{P}\int_{0}^{\infty}\frac{(1-\omega_0 u)\ln(F(-u))}{(1+u^2)(\omega_0+u)}du = \mathcal{P}\int_{0}^{\infty}\frac{(1-\omega_0 u)\ln(F(u))}{(1+u^2)(\omega_0+u)}du.$$

Now we rename the integration variable u back into ω. Using

$$\frac{1+\omega_0\omega}{\omega_0-\omega} + \frac{1-\omega_0\omega}{\omega_0+\omega} = \frac{2\omega_0(1+\omega^2)}{\omega_0^2-\omega^2}$$

one can combine the two integrals and obtains

$$\Phi(\omega_0) = \frac{2\omega_0}{\pi}\mathcal{P}\int_{0}^{\infty}\frac{\ln(F(\omega))}{\omega_0^2-\omega^2}d\omega.$$

The integrand has a singularity at $\omega = \omega_0$. To cancel it one subtracts the expression

$$\frac{2\omega_0}{\pi}\mathcal{P}\int_{0}^{\infty}\frac{\ln(F(\omega_0))}{\omega_0^2-\omega^2}d\omega = \frac{2\omega_0\ln(F(\omega_0))}{\pi}\mathcal{P}\int_{0}^{\infty}\frac{1}{\omega_0^2-\omega^2}d\omega = 0.$$

It is easy to verify that this integral vanishes. With $x = \omega/\omega_0$ the integral is of the type

$$\mathcal{P}\int_{0}^{\infty}\frac{1}{1-x^2}dx = \lim_{\varepsilon\to 0}\left[\int_{0}^{1-\varepsilon} + \int_{1+\varepsilon}^{\infty}\right]\left(\frac{1}{1+x} + \frac{1}{1-x}\right)dx = \lim_{\varepsilon\to 0}\left[\,\ln(2-\varepsilon) - \ln(2+\varepsilon)\,\right] = 0.$$

So finally we arrive at the important dispersion relation

$$\Phi(\omega_0) = \frac{2\omega_0}{\pi}\mathcal{P}\int_{0}^{\infty}\frac{\ln(F(\omega)) - \ln(F(\omega_0))}{\omega_0^2-\omega^2}d\omega. \tag{79}$$

6.5 Computation of form factor phase via dispersion relation

6.5.1 Kramers-Kronig phase

As said before, the longitudinal form factor is the Fourier transform of the normalized longitudinal charge density distribution

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} \rho(t) e^{i\omega t} dt \quad \text{with} \quad \int_{-\infty}^{\infty} \rho(t) dt = 1 \, . \tag{80}$$

Spectroscopic experiments at particle accelerators yield only the absolute magnitude of the form factor, $F(\omega) = |\mathcal{F}(\omega)|$, but neither its real part nor its imaginary part. Writing

$$\mathcal{F}(\omega) = F(\omega) \exp(i\Phi(\omega)) \tag{81}$$

formula (79) can be applied to compute the phase $\Phi(\omega)$. The *Kramers-Kronig phase* is given by

$$\Phi_{\mathrm{KK}}(\omega_0) = \frac{2\omega_0}{\pi} \, \mathcal{P} \int_0^{\infty} \frac{\ln(|\mathcal{F}(\omega)|) - \ln(|\mathcal{F}(\omega_0)|)}{\omega_0^2 - \omega^2} \, d\omega \tag{82}$$

provided the prerequisites made in the derivation of this formula are fulfilled[6]. The phase (82) is also called the canonical phase or the minimal phase, the latter expression being somewhat misleading since $\Phi_{\mathrm{KK}}(\omega)$ does not correspond to a minimum in the mathematical sense. One has to be aware that there are commonly used test functions which do not obey all prerequisites made in the derivation of Eq. (82) and require a special treatment. We will study two such problematic cases: nonvanishing integral of the auxiliary function over the large semicircle, and form factors having zeros in upper half of the complex frequency plane.

6.5.2 Problem 1: Integral over large semicircle maybe nonzero

An essential prerequisite for the validity of formula (82) is that the integral of the auxiliary function along the large semicircle in Fig. 45 must vanish. This requirement is not met by Gaussian form factors. However, since it is very convenient to represent the time profile of a short electron bunch by a Gaussian function or a linear combination of Gaussians, it is worthwhile to investigate if it is possible to preserve the validity of the KK phase formula (82) by taking into account the line integral along the large semicircle as an extra contribution. We will show now that this is indeed possible.

(a) Single Gaussian
Consider a single Gaussian which for simplicity is centered at $t = 0$. The form factor is also a Gaussian.

$$\rho(t) = \frac{1}{\sqrt{2\pi}\sigma} \, \exp\left(-t^2/(2\sigma^2)\right) \, , \quad \mathcal{F}(\omega) = \exp\left(-\sigma^2\omega^2/2\right) \, .$$

With this form factor we are facing a mathematical problem. The logarithm is proportional to the square of frequency, $\ln(\mathcal{F}(\omega)) = -\sigma^2\omega^2/2$, which implies that the integral of the auxiliary function (74) over the large semicircle Γ_1 in Fig. 45 does not vanish. Is it still permitted to use the Kramers-Kronig formula for a Gaussian bunch? To decide this question, it is obviously necessary to compute explicitly the line integral along the half circle Γ_1. To this end we go into the complex frequency plane ($\hat{\omega} = \omega_{\mathrm{r}} + i\,\omega_{\mathrm{i}}$). The auxiliary function has a simple form

$$f_{\mathrm{aux}}(\hat{\omega}) = \frac{(1 + \omega_0\hat{\omega}) \ln(\mathcal{F}(\hat{\omega}))}{(1 + \hat{\omega}^2)(\omega_0 - \hat{\omega})} = \frac{(1 + \omega_0\hat{\omega})(-\sigma^2\hat{\omega}^2/2)}{(1 + \hat{\omega}^2)(\omega_0 - \hat{\omega})} \, . \tag{83}$$

Now we compute the line integral over the semicircle Γ_1. The radius R is chosen to be very large, $R \gg \omega_0$, but we do not let R tend to infinity. On the semicircle we can simplify f_{aux}:

$$\hat{\omega} = Re^{i\theta}, \quad R \gg \omega_0 \quad \Rightarrow \quad f_{\mathrm{aux}}(\hat{\omega}) \approx \omega_0\sigma^2/2 \, .$$

---

[6]Physics requires that we make a slight modification in this formula. Backward transition radiation does not extend to infinite frequencies. High-energy X or gamma rays will never be observed in backward direction. Therefore the upper integration limit is far from infinity but rather a suitable cutoff frequency ω_{cut} in the ultraviolet regime which corresponds to the radius R of the large semicircle if we choose a large but finite value for R.

Thus we get in very good approximation

$$\int_{\Gamma_1} f_{\text{aux}}(\hat{\omega})d\hat{\omega} = \omega_0\sigma^2/2 \int_0^\pi i\,R\,e^{i\theta}d\theta = +\omega_0 R\sigma^2 \tag{84}$$

so the line integral over the semicircle Γ_1 is indeed different from zero.

Next we evaluate the line integral over the straight path Γ_2, the real axis. Making use of the computational steps in section 6.4 we obtain

$$\int_{\Gamma_2} f_{\text{aux}}(\omega)d\omega = 2\omega_0\,\mathcal{P}\int_0^R \frac{\ln(|\mathcal{F}(\omega)|) - \ln(|\mathcal{F}(\omega_0)|)}{\omega_0^2 - \omega^2}\,d\omega\,.$$

Inserting the Gaussian form factor this expression yields:

$$\int_{\Gamma_2} f_{\text{aux}}(\omega)d\omega = -\omega_0 R\sigma^2\,. \tag{85}$$

The most remarkable result is: in case of a Gaussian form factor the line integral along the real axis is cancelled by the line integral along the large semicircle. The phase of the Gaussian form factor is identical to zero, as it should be.

(b) Truncated Gaussian

As stated before, a truncated Gaussian preserves causality. Hence one might expect that the integral of the auxiliary function over large semicircle vanishes in this case. We prove now that this is indeed the case. Consider the function

$$\begin{aligned} \rho(t) &= 0 \quad \text{for} \quad t < 0\,, \\ \rho(t) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(t-\tau)^2}{2\sigma^2}\right) \quad \text{for} \quad t \geq 0\,. \end{aligned} \tag{86}$$

If we choose for example $\tau = \alpha\,\sigma$ with $\alpha = 5$, the truncated Gaussian and a normal Gaussian (centered at $t = \tau$) are hardly distinguishable.
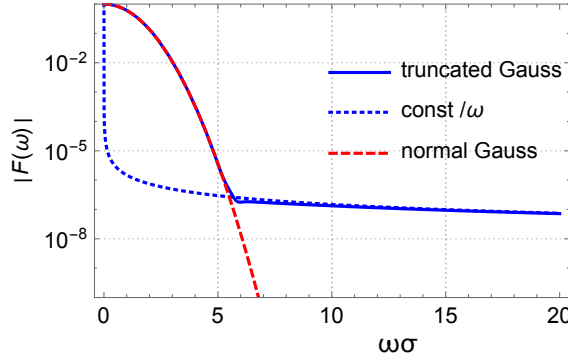

Figure 46: The magnitude $F_{\text{trunc}}(\omega)$ of the form factor of a truncated Gaussian (solid blue curve), plotted as a function of the dimensionless variable $\omega\sigma$, in comparison with a normal Gaussian form factor (dashed red curve). At large values of $\omega\sigma$, the form factor obeys an inverse power law $F_{\text{trunc}}(\omega) \approx \text{const}/\omega$ (dotted blue curve).

Their form factors are for real ω

$$\begin{aligned} \mathcal{F}_{\text{trunc}}(\omega) &= \frac{1}{2}\exp\left(-\frac{\sigma^2\omega^2}{2} + i\,\omega\tau\right)\left(1 + \text{erf}\left[\frac{\alpha + i\,\sigma\omega}{\sqrt{2}}\right]\right) \quad (\text{erf} = \text{error function})\,, \tag{87} \\ \mathcal{F}_{\text{norm}}(\omega) &= \exp\left(-\frac{\sigma^2\omega^2}{2} + i\,\omega\tau\right) \quad \text{with} \quad \tau = \alpha\,\sigma\,. \end{aligned}$$

47

The magnitudes are plotted in Fig. 46. Both form factors agree in the range $0 \leq \omega\sigma \leq \alpha$, however at large frequencies the form factor of the normal Gaussian drops to tiny values while that of the truncated Gaussian levels off. By series expansion of Eq. (87) one finds that asymptotically $F_{\text{trunc}}(\omega)$ obeys an inverse power law:

$$F_{\text{trunc}}(\omega) \approx \frac{\exp(-\alpha^2/2)}{\sqrt{2\pi}\sigma} \cdot \frac{1}{\omega} \quad \text{for} \quad \omega\sigma \gg \alpha. \tag{88}$$

Thus the condition (75) is fulfilled, here with $s = 1$, and this means that the integral of the auxiliary function over large semicircle Γ_1 vanishes. The important conclusion is: the Kramers-Kronig phase formula (82) is perfectly valid for truncated Gaussian time-domain functions.

6.5.3   Problem 2: Complex zeros of the form factor

The most important prerequisite for Eq. (82) to hold is that the form factor must be different from zero in the entire upper half plane ($\Im(\omega_{\text{r}} + i\omega_{\text{i}}) = \omega_{\text{i}} > 0$). If the form factor vanishes at some point inside the closed integration loop Γ its logarithm has an essential singularity here (not a simple pole), and the Residue Theorem can no longer be applied. The Kramers-Kronig formula becomes obsolete. Is there a way out of this difficulty?

Suppose now that the form factor has a zero of first order at some complex frequency $\hat{\omega}_1 = a + ib$ in the right upper quarter of the $\hat{\omega}$ plane ($a, b > 0$). An interesting observation is that zeros occur always in pairs, at $\hat{\omega}_1 = a + ib$ in the right upper quarter and at $\hat{\omega}_1' = -a + ib$ in the left upper quarter. This follows from Eq. (36). This pair of zeros in the upper half plane can be removed by defining a modified form factor

$$\mathcal{F}_{\text{mod}}(\hat{\omega}) = \mathcal{F}(\hat{\omega})\mathcal{B}(\hat{\omega}) \tag{89}$$

with the so-called Blaschke factor

$$\mathcal{B}(\hat{\omega}) = \frac{\hat{\omega} - \hat{\omega}_1^*}{\hat{\omega} - \hat{\omega}_1} \cdot \frac{\hat{\omega} - \hat{\omega}_1'^*}{\hat{\omega} - \hat{\omega}_1'} = \frac{\hat{\omega} - (a - ib)}{\hat{\omega} - (a + ib)} \cdot \frac{\hat{\omega} - (-a - ib)}{\hat{\omega} - (-a + ib)}. \tag{90}$$

Behavior on real axis
On the real ω axis, i.e. for $\hat{\omega} = \hat{\omega}^* = \omega$, the absolute magnitude of the Blaschke factor is 1:

$$|\mathcal{B}(\omega)| = \left|\frac{\omega - (a - ib)}{\omega - (a + ib)}\right| \cdot \left|\frac{\omega - (-a - ib)}{\omega - (-a + ib)}\right| = 1 \quad \Rightarrow \quad |\mathcal{F}_{\text{mod}}(\omega)| = |\mathcal{F}(\omega)| \quad \text{for real} \quad \omega. \tag{91}$$

This is a very important equation. It means that the form factor $\mathcal{F}(\omega)$ and the modified form factor $\mathcal{F}_{\text{mod}}(\omega)$ describe exactly the same radiation spectrum. Knowing just the absolute magnitude from experiment one cannot decide which of the two form factors (or even another expression) is the correct one.

However, the phases of $\mathcal{F}(\omega)$ and $\mathcal{F}_{\text{mod}}(\omega)$ are different because the Blaschke factor is a complex, frequency-dependent number. Real and imaginary part of $\mathcal{B}(\omega)$ are

$$\Re(\mathcal{B}(\omega)) = \frac{\omega^4 - 2\omega^2(a^2 + 3b^2) + (a^2 + b^2)^2}{[(\omega - a)^2 + b^2][(\omega + a)^2 + b^2]}, \quad \Im(\mathcal{B}(\omega)) = \frac{4\omega b(\omega^2 - a^2 - b^2)}{[(\omega - a)^2 + b^2][(\omega + a)^2 + b^2]}. \tag{92}$$

The phase of $\mathcal{B}(\omega)$ is computed by the equation

$$\Phi_{\text{B}}(\omega) = \arg(\mathcal{B}(\omega)). \tag{93}$$

This phase is not a continuous function of ω but exhibits a discontinuous jump by 2π at the frequency $\omega_{\text{jump}} = \sqrt{a^2 + b^2}$ where the imaginary part of the Blaschke factor vanishes. It is plotted in Fig. 47 as a function of the real scaled frequency ω/a. The phase jump does not present any problem[7], the equalities $\Re(\mathcal{B}(\omega)) = |\mathcal{B}(\omega)|\cos\Phi_{\text{B}}(\omega)$ and $\Im(\mathcal{B}(\omega)) = |\mathcal{B}(\omega)|\sin\Phi_{\text{B}}(\omega)$ are perfectly well fulfilled.
In Fig. 48 the function $\Phi_{\text{B}}(\omega)$ is shown for various ratios b/a. In case of a very small imaginary part of the complex zero $\hat{\omega}_1 = a + ib$, i.e. for $b/a \ll 1$, the phase excursion $0 \rightarrow -\pi \rightarrow +\pi \rightarrow 0$ is very steep and affects only a narrow frequency range. This means that complex zeros in the immediate vicinity of the real axis have a negligible influence on the bunch shape reconstruction.

---

[7]The situation is rather more complicated if one uses the arctan function instead of the arg function to compute the phase of the Blaschke factor. Then two phase jumps, each by π, are obtained near ω_{jump} which lead to errors when one computes the sine or cosine of the Blaschke phase. These deficiencies can be cured but the mathematical procedure is clumsy.
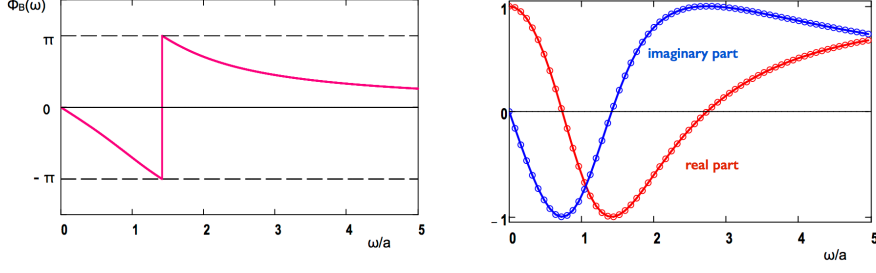
Figure 47: *Left*: The Blaschke phase $\Phi_{\mathrm{B}}(\omega)$ of a pair of zeros (at $\hat{\omega}_1 = a + ib$ and $\hat{\omega}_1' = -a + ib$) for a ratio $b/a = 1$. *Right*: Comparison of $\Re(\mathcal{B}(\omega))$ (red curve) with $\cos\Phi_{\mathrm{B}}(\omega)$ (red circles) and comparison of $\Im(\mathcal{B}(\omega))$ (blue curve) with $\sin\Phi_{\mathrm{B}}(\omega)$ (blue circles).
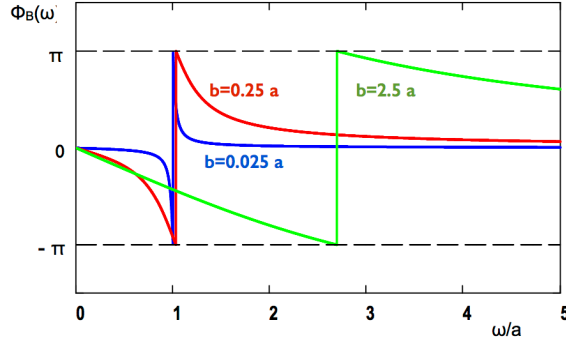

Figure 48: The Blaschke phase $\Phi_{\mathrm{B}}(\omega)$ for various ratios b/a.

On the other hand, zeros with a very large imaginary part ($b \gg a$) lead to an almost linearly rising phase at $\omega \approx a$. They do not change the internal structure of the reconstructed bunch but merely shift its position on the time axis.

Suppose now that $\mathcal{F}(\hat{\omega})$ has a finite number of pairs of zeros in the upper half of the complex $\hat{\omega}$ plane, at the frequencies $\hat{\omega}_n = a_n + ib_n$ and $\hat{\omega}_n' = -a_n + ib_n$. The *Blaschke product* is defined as the product of the Blaschke factors $\mathcal{B}_n(\omega)$ of these pairs. The phases are defined for real ω, they are additive.

$$\mathcal{B}(\hat{\omega}) = \prod_n \mathcal{B}_n(\hat{\omega}) = \prod_n \frac{\hat{\omega} - (a_n - ib_n)}{\hat{\omega} - (a_n + ib_n)} \cdot \frac{\hat{\omega} - (-a_n - ib_n)}{\hat{\omega} - (-a_n + ib_n)}, \qquad \Phi_{\mathrm{B}}(\omega) = \sum_n \Phi_{\mathrm{B},n}(\omega). \qquad (94)$$

In analogy with Eq. (89) the modified form factor is given by

$$\mathcal{F}_{\mathrm{mod}}(\hat{\omega}) = \mathcal{F}(\hat{\omega})\mathcal{B}(\hat{\omega}). \qquad (95)$$

This form factor has no zero in the upper half plane. It fulfills the requirements made in the derivation of formula (82), and hence the phase of the modified form factor is identical with the Kramers-Kronig phase, $\Phi_{\mathrm{mod}}(\omega) = \Phi_{\mathrm{KK}}(\omega)$. The phase of our original form factor, which we call *reconstruction phase* from now on, follows immediately from Eq. (95)

$$\Phi_{\mathrm{rec}}(\omega) = \Phi_{\mathrm{KK}}(\omega) - \Phi_{\mathrm{B}}(\omega). \qquad (96)$$

49

7   Appendix B: Analytic Model Calculations with Causal Functions

To assess the capabilities and the limitations of the phase reconstruction of the bunch form factor by applying the theorems of complex analysis, it is useful to employ the method for simple test functions $\rho(t)$ whose complex form factor is either known analytically or can be calculated by Fourier transformation.

We have seen already that the Kramers-Kronig phase reconstruction works well for bunches with a single peak but leads to puzzling results when two or more peaks are present: in some cases the original is perfectly reproduced, in other cases the reconstructed charge distribution is very different from the original. This will be illustrated with a number of examples, and we show the benefit of the Blaschke phase correction. In this appendix we restrict ourselves to test functions $\rho(t)$ which obey the basic requirement of causality. The same superpositions of cosine-squared pulses as in section 3.3 will be considered, and we will compute the Blaschke correction for these profiles. Another important topic is the relation between time reversal ($\rho(t) \to \rho(-t)$) and the location of the form factor zeros in the complex frequency plane.

Gaussian time profiles, violating causality, are the subject of Appendix C.

7.1   Two cosine-squared pulses of different width

Consider the superposition of two cosine-squared pulses of different amplitudes A_1, A_2, different widths $b_1 < b_2$ and center times t_1, t_2 which may be equal or different. The Kramers-Kronig reconstruction fails if the two pulses are centered with respect to each other, as shown in Fig. 15. To determine the Blaschke phase we have to find the complex zeros of the form factor which according to Eq. (46) is given by

$$\mathcal{F}(\omega) = A_1 \frac{\pi^2 \sin(\omega b_1) \exp(i\,\omega t_1)}{\omega b_1 (\pi^2 - \omega^2 b_1^2)} + A_2 \frac{\pi^2 \sin(\omega b_2) \exp(i\,\omega t_2)}{\omega b_2 (\pi^2 - \omega^2 b_2^2)} \,. \tag{97}$$

We consider the case that the two pulses are centered with respect to each other and choose $t_1 = t_2 = b_2$. Unlike the Akutowicz case, treated in section 3.3, the complex frequencies $\hat{\omega}_n$ where $\mathcal{F}(\hat{\omega})$ vanishes cannot be computed analytically but are found by a numerical search procedure. To this end, the contour lines $\Re(\mathcal{F}(\hat{\omega})) = 0$ and
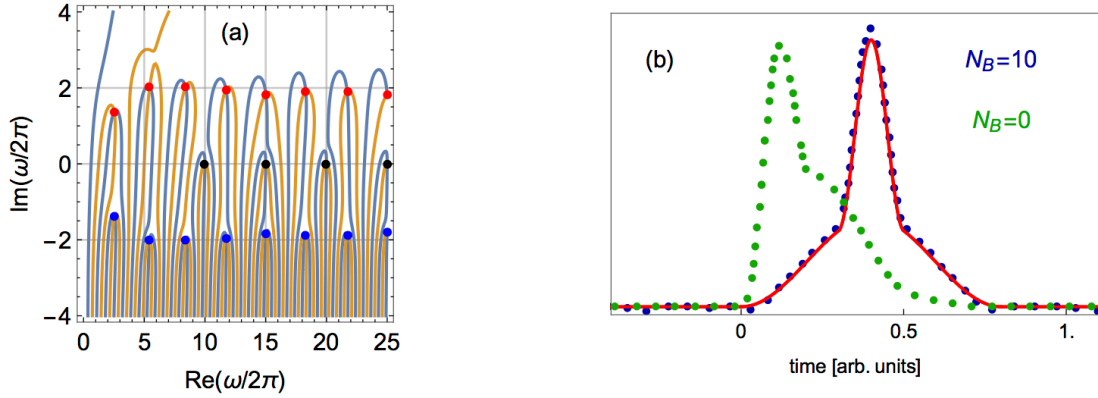

Figure 49: Reconstruction of a bunch consisting of two cosine-squared pulses of different width ($b_1 < b_2$) which are both centered at $t = b_2$. **(a)** Plot of the contour lines $\Re(\mathcal{F}(\hat{\omega})) = 0$ (blue) and $\Im(\mathcal{F}(\hat{\omega})) = 0$ (yellow) in the complex frequency plane. The dots indicate the crossing points where the form factor $\mathcal{F}(\hat{\omega}))$ vanishes.
(b) The original charge distribution $\rho(t)$ (red curve) and the reconstruction (blue dots) using the reconstruction phase Eq. (96), i.e. the difference between Kramers-Kronig phase and Blaschke phase (computed for $N_B = 10$ zeros). As a reminder, the Kramers-Kronig reconstruction without Blaschke phase is also shown here by the green dotted curve (labelled $N_B = 0$).

$\Im(\mathcal{F}(\hat{\omega})) = 0$ are plotted and their intersection points in the upper half of the complex $\hat{\omega}$ plane are determined. The numerical results are shown in Fig. 49a. The crossing points give us the complex zeros $\hat{\omega}_n = a_n + ib_n$ which are needed to compute the Blaschke phase $\Phi_B(\omega)$. In this example we take $N_B = 10$ complex zeros $\hat{\omega}_n = a_n + ib_n$ into account which are located in the right upper quarter of the $\hat{\omega}$ plane, and in addition their

mirror images $\hat{\omega}'_n = -a_n + ib_n$ in the left upper quarter. The total Blaschke phase and the reconstruction phase become

$$\Phi_{\mathrm{B}}(\omega) = \sum_{n=1}^{N_B} \Phi_{B,n}(\omega)\,, \quad \Phi_{\mathrm{rec}}(\omega) = \Phi_{\mathrm{KK}}(\omega) - \Phi_{\mathrm{B}}(\omega)\,.$$

Using the phase $\Phi_{\mathrm{rec}}(\omega)$ an almost perfect reconstruction is achieved, see Fig. 49b. An interesting question is how many zeros of the form factor have to be taken into consideration. In Fig. 50 the Kramers-Kronig-Blaschke (KKB) reconstruction is shown for $N_B = 1, 2, 5$ zeros.
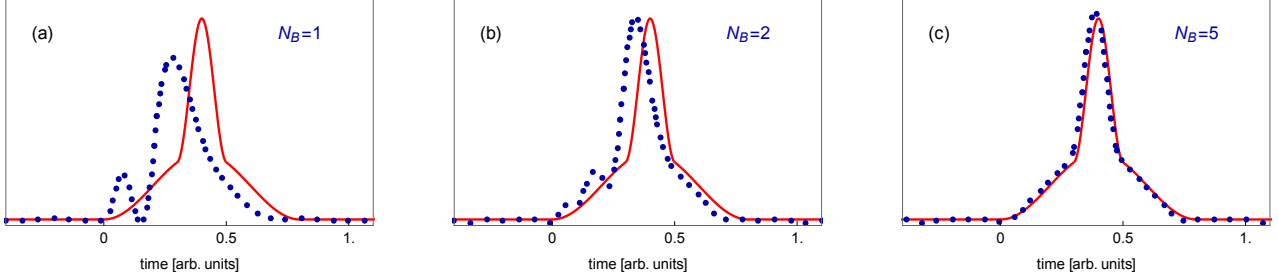

Figure 50: The Kramers-Kronig-Blaschke (KKB) reconstruction for $N_B = 1, 2, 5$ zeros.

7.2 Three cosine-squared pulses of equal width

We consider the superposition of three cosine-squared pulses of equal width and with non-uniform spacing which has already been shown in Fig. 17. First we take the case that the largest peak is at the front. Then the KK phase yields a perfect reconstruction, see Figs. 17a and 51a. This can be understood as follows: The complex form factor has no zeros in the upper half of the complex frequency plane, hence the Blaschke phase is zero and the reconstruction phase is identical with the KK phase.
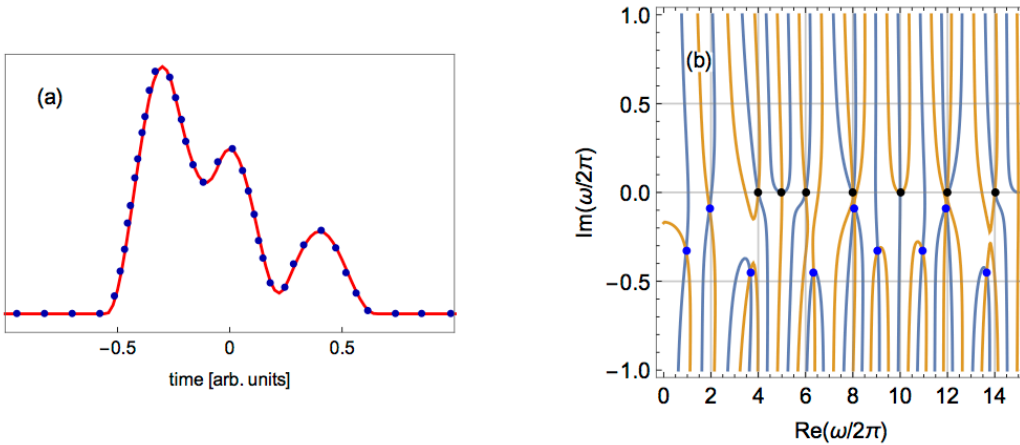

Figure 51: Three cosine-squared pulses of equal width and with non-uniform spacing. The largest peak is at the front. **(a)** The original curve (red) and the KK reconstruction (blue dots). **(b)** The contour lines $\Re(\mathcal{F}(\hat{\omega})) = 0$ (blue) and $\Im(\mathcal{F}(\hat{\omega})) = 0$ (yellow) in the complex frequency plane. The complex form factor has zeros on the real axis (black dots) and in the lower half of the complex frequency plane (blue dots), none of which contribute to the Blaschke phase. Hence the reconstruction phase is equal to the KK phase: $\Phi_{\mathrm{rec}}(\omega) \equiv \Phi_{\mathrm{KK}}(\omega)$.

When the largest peak is in the center the KK reconstruction disagrees with the input distribution, see Fig. 17b. The explanation is given in Fig. 52. The complex form factor has zeros in the upper half of the complex frequency plane, so there is a nonvanishing Blaschke phase contributing to the reconstruction phase.
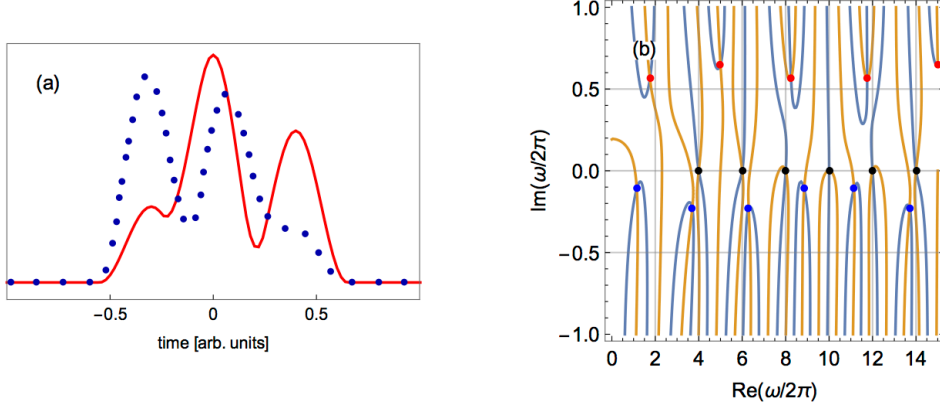
Figure 52: Three cosine-squared pulses of equal width and with non-uniform spacing. The largest peak is in the center. **(a)** The original curve (red) and the KK reconstruction (blue dots). **(b)** The complex form factor has zeros in the upper half of the complex frequency plane (red dots) which contribute to the Blaschke phase.

The shape reconstruction using the phase $\Phi_{\text{rec}}(\omega) = \Phi_{\text{KK}}(\omega) - \Phi_{\text{B}}(\omega)$ is depicted in Fig. 53. It agrees very well with the original bunch shape.
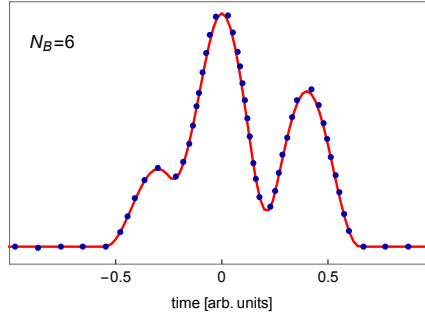

Figure 53: The Kramers-Kronig-Blaschke (KKB) reconstruction of the triple cosine-squared structure with the largest peak in the center, using $\Phi_{\text{rec}}(\omega) = \Phi_{\text{KK}}(\omega) - \Phi_{\text{B}}(\omega)$ with $N_B = 6$ pairs of zeros.

7.3 Time reversal and form factor zeros

Let $\rho_1(t)$ be a causal function and $\mathcal{F}_1(\hat{\omega})$ its complex form factor. The time-reversed function $\rho_2(t) = \rho_1(-t)$ has the form factor

$$\mathcal{F}_2(\hat{\omega}) = \mathcal{F}_1(-\hat{\omega}). \tag{98}$$

Proof:

$$\mathcal{F}_1(\hat{\omega}) = \int_{t=-\infty}^{t=+\infty} \rho_1(t) \exp(+i\,\hat{\omega}t)dt \quad \Rightarrow \quad \mathcal{F}_1(-\hat{\omega}) = \int_{t=-\infty}^{t=+\infty} \rho_1(t) \exp(-i\,\hat{\omega}t)dt = \int_{t=-\infty}^{t=+\infty} \rho_2(-t) \exp(-i\,\hat{\omega}t)dt.$$

In the last integral we substitute $t' = -t$, $dt' = -dt$ and get

$$\mathcal{F}_1(-\hat{\omega}) = \int_{t'=+\infty}^{t'=-\infty} \rho_2(t') \exp(+i\,\hat{\omega}t')(-dt') = \int_{t'=-\infty}^{t'=+\infty} \rho_2(t') \exp(+i\,\hat{\omega}t')dt' \equiv \mathcal{F}_2(\hat{\omega}) \quad \text{qed.}$$

Equation (98), combined with Eq. (36), has an important consequence: The complex zeros of $\mathcal{F}_1(\hat{\omega})$ and the zeros of $\mathcal{F}_2(\hat{\omega})$ are mirror-symmetric with respect to the real frequency axis.
Suppose $\mathcal{F}_1(\hat{\omega})$ has only zeros in the upper half of the complex $\hat{\omega}$ plane but no zeros in the lower half. Then $\mathcal{F}_2(\hat{\omega})$ behaves just in the opposite way, it has zeros in the lower half plane but no zeros in the upper half plane.

This implies that the KK reconstruction works perfectly well for $\rho_2(t)$ while it fails for $\rho_1(t)$. In fact, if one uses the KK phase to reconstruct the time profile from the modulus $F_1(\omega) = |\mathcal{F}_1(\omega)|$, the resulting time profile coincides with $\rho_2(t)$.

This behavior is demonstrated in Fig. 54 for a bunch consisting of two cosine-squared pulses of different width where the narrow peak is either at the rear (time profile $\rho_1(t)$) or at the front (time profile $\rho_2(t)$). The form factor magnitudes are identical on the real axis: $F_1(\omega) = |\mathcal{F}_1(\omega)| = F_2(\omega) = |\mathcal{F}_2(\omega)|$, and the shape reconstruction with the KK phase (40) reproduces the time profile $\rho_2(t)$ in either case.
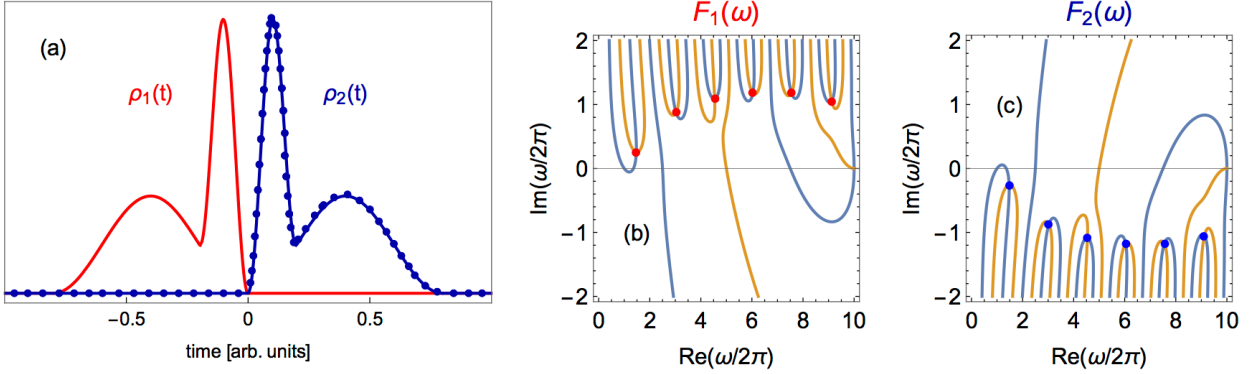

Figure 54: **(a)** Time profile of a bunch consisting of two cosine-squared pulses of different width, the narrow peak is either at the rear end (red curve $\rho_1(t)$) or at the front (blue curve $\rho_2(t)$). In both cases the KK reconstruction, indicated by the blue dots, yields the same time profile which coincides with the blue curve. **(b)** Contour plot for the case that the narrow peak is at the rear. The red dots are the complex form factor zeros. They are in the upper half of the complex plane. **(c)** Contour plot for the case that the narrow peak is at the front. The complex form factor zeros (blue dots) are in the lower half of the complex plane.
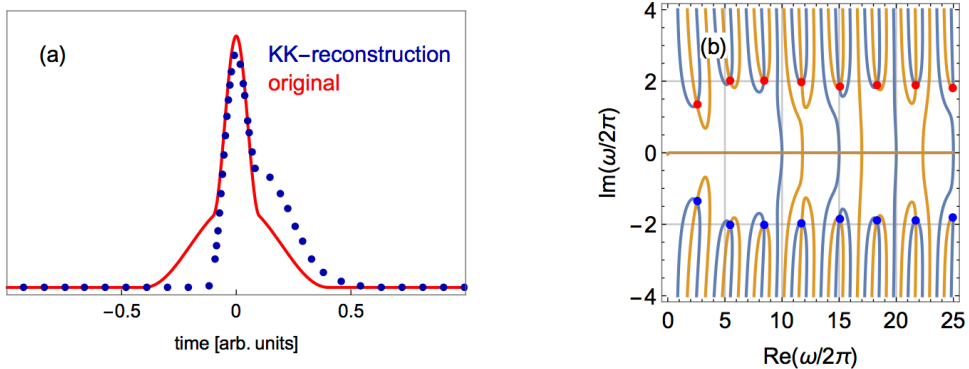

Figure 55: **(a)** A bunch consisting of two cosine-squared pulses of different width which are both centered at $t = 0$. **(b)** Contour plot. The complex form factor zeros are both in the upper half (red dots) and in the lower half (blue dots) of the complex plane.

However, if the narrow peak is centered with respect to the wide one (Fig. 55) then time reversal does not change anything. As expected, the form factor zeros in the complex $\hat{\omega}$ plane are mirror-symmetric with respect to the real axis. The KK reconstruction must necessarily fail.

53

8 Appendix C: Gaussians and truncated Gaussians

Gaussian functions are very convenient since their Fourier transforms are also Gaussians, and hence they have often been used in model calculations on analytic bunch shape reconstruction. We have shown in Appendix A that a Gaussian function in time-domain violates the fundamental requirement of causality, so the use of genuine Gaussians is mathematically questionable. Much preferable are Gaussian-like time profiles obeying causality. These can be generated by truncating a Gaussian at a sufficiently large distance from the peak.

It is the purpose of this appendix to explore to what extent normal (non-truncated) Gaussians are permitted in analytic model calculations, in spite of their mathematical deficiencies. We will demonstrate that phase retrieval based on the Kramers-Kronig dispersion relation can indeed be done with any superposition of Gaussians provided the integration in the basic KK phase formula (82) is restricted to a finite frequency range. In that case the Gaussians describing the time profile of the bunch are in reality truncated Gaussians. Quite a different result is obtained if one wants to determine the complex zeros of the form factor. We show that there are significant differences between truncated and non-truncated Gaussians. In the computation of the Blaschke phase it is mandatory to use truncated functions.

8.1 Impact of truncation on the Kramers-Kronig phase

In section 6.5.2 we have proved that the truncated Gaussian defined in Eq. (86) fulfills all requirements needed in the derivation of the Kramers-Kronig phase formula (82). For completeness we repeat this definition:

$$\rho_{\text{trunc}}(t) = 0 \text{ for } t < 0 , \qquad \rho_{\text{trunc}}(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(t-\tau)^2}{2\sigma^2}\right) \text{ for } t \geq 0$$

where τ is the "cutoff time" and $\alpha = \tau/\sigma$ the dimensionless cutoff parameter.
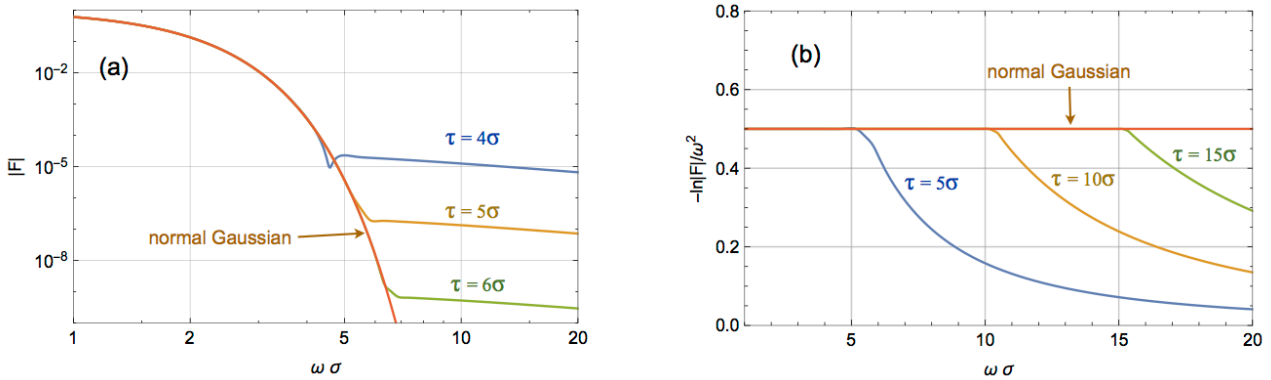

Figure 56: **(a)** The magnitude $|\mathcal{F}_{\text{trunc}}(\omega)|$ of the form factor of a truncated Gaussian is plotted as a function of the dimensionless variable $\omega\sigma$ for three cutoff times $\tau = \alpha\,\sigma$. For comparison the form factor of a normal Gaussian is also shown. **(b)** The integrand $I_{\text{trunc}}(\omega)$ of the KK phase integral (99) for truncated Gaussians with different cutoff times and the integrand $I_{\text{Gauss}}(\omega) = \sigma^2/2$ for a normal Gaussian. In this graph we put $\sigma = 1$ in arbitrary units.

The form factor $\mathcal{F}_{\text{trunc}}(\omega)$ is given by Eq. (87). In Fig. 56a we plot its absolute magnitude $|\mathcal{F}_{\text{trunc}}(\omega)|$ for several values of the cutoff time τ and compare it with the form factor of a normal Gaussian. Both form factors agree in the range $0 \leq \omega\sigma \leq \alpha$, however at larger frequencies the form factor of the normal Gaussian drops to tiny values while that of the truncated Gaussian levels off and obeys an inverse power law according to Eq. (88).

To understand the impact of truncation on the KK phase we rewrite the KK phase integral (82) for a truncated Gaussian:

$$\Phi_{\text{KK}}^{\text{trunc}}(\omega_0) = \frac{2\omega_0}{\pi} \mathcal{P} \int_0^{\omega_{\text{cut}}} \frac{\ln(|\mathcal{F}_{\text{trunc}}(\omega)|) - \ln(|\mathcal{F}_{\text{trunc}}(\omega_0)|)}{\omega_0^2 - \omega^2} \, d\omega . \tag{99}$$

Since the KK phase integral is usually evaluated by numerical integration over a limited frequency range, the upper integration limit in formula (82) has been replaced by a suitable cutoff frequency $\omega_{\mathrm{cut}} < \infty$. For the special case $\omega_0 = 0$ we have $\mathcal{F}_{\mathrm{trunc}}(\omega_0) = 1$, and the integrand takes the simple form

$$I_{\mathrm{trunc}}(\omega) = -\frac{\ln(|\mathcal{F}_{\mathrm{trunc}}(\omega)|)}{\omega^2} .$$

In case of a normal Gaussian the integrand is a constant:

$$I_{\mathrm{Gauss}}(\omega) = -\frac{\ln(|\mathcal{F}_{\mathrm{Gauss}}(\omega)|)}{\omega^2} = -\frac{-\omega^2\sigma^2}{2\omega^2} = \frac{\sigma^2}{2} .$$

The function $I_{\mathrm{trunc}}(\omega)$ is plotted in Fig. 56b for three different cutoff times, together with $I_{\mathrm{Gauss}}(\omega) = \sigma^2/2$. The functions $I_{\mathrm{trunc}}(\omega)$ and $I_{\mathrm{Gauss}}(\omega) = \sigma^2/2$ agree in the range $0 \leq \omega\sigma \leq \alpha$, but at larger frequencies $I_{\mathrm{trunc}}(\omega)$ drops rapidly to zero. This implies that the upper limit of the phase integral (99) is $\omega_{\mathrm{cut}} \approx \alpha/\sigma = \tau/\sigma^2$.

Now we turn the argumentation around. Suppose we describe the charge distribution inside a bunch by a superposition of Gaussians and want to evaluate the KK phase formula. The numerical integration is necessarily restricted to a finite frequency range $0 \leq \omega \leq \omega_{\mathrm{cut}}$. This however is equivalent to the KK phase computation for the same superposition of truncated Gaussians where the cutoff parameter is given by $\alpha \approx \sigma\,\omega_{\mathrm{cut}}$. In other words: The KK phase formula (82) is applicable for any superposition of Gaussian time-domain functions, provided the integration is restricted to a finite frequency range, which happens automatically in a numerical integration.

8.2 Impact of truncation on the Blaschke zeros

To investigate the impact of truncation on the zeros of the complex form factor we study a special case, the superposition of two Gaussian-shaped pulses of different width. In section 3.3 and Appendix B we have analyzed bunches consisting of two cosine-squared pulses of different width, see Fig. 15 and Fig. 49, and we have observed that the KK phase yields a good reconstruction when the narrow peak is at the front of the wide one, but fails when the narrow peak is centered with respect to the wide peak. The same puzzling result is obtained for two Gaussian-shaped pulses of different width. This is illustrated in Fig. 57a.
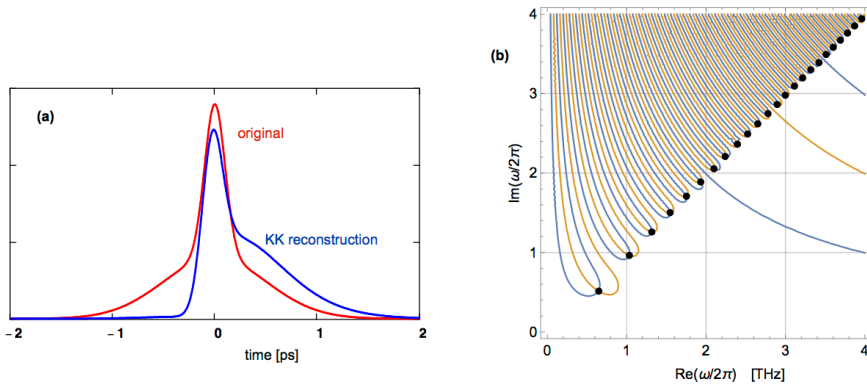

Figure 57: **(a)** Superposition of two Gaussians of different width ($\sigma_1 = 0.5\,\mathrm{ps}$, $\sigma_2 = 0.1\,\mathrm{ps}$) which are centered with respect to each other (red curve) and the KK reconstruction (blue curve). **(b)** Plot of the contour lines $\Re(\mathcal{F}(\hat{\omega})) = 0$ (blue) and $\Im(\mathcal{F}(\hat{\omega})) = 0$ (yellow) in the complex $\hat{\omega}$ plane. There are a huge number of densely spaced complex zeros (black dots).

In Appendix B we have shown that a perfect reconstruction of the two-cosine-squared bunch is achieved by combining the KK and Blaschke phases. For Gaussian pulses this turns out to be quite problematic. As an example we consider the following function

$$\rho(t) = \frac{A_1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{t^2}{2\sigma_1^2}\right) + \frac{A_2}{\sqrt{2\pi}\,\sigma_2} \exp\left(-\frac{t^2}{2\sigma_2^2}\right)$$

with the parameters $A_1 = 2/3$, $A_2 = 1/3$, $\sigma_1 = 0.5\,\mathrm{ps}$ and $\sigma_1 = 0.1\,\mathrm{ps}$. The form factor is

$$\mathcal{F}(\omega) = A_1 \exp\left(-\frac{\sigma_1^2 \omega^2}{2}\right) + A_2 \exp\left(-\frac{\sigma_2^2 \omega^2}{2}\right).$$

The contour plot of the curves $\Re(\mathcal{F}(\hat{\omega})) = 0$ and $\Im(\mathcal{F}(\hat{\omega})) = 0$ in Fig. 57b reveals there are an enormous number of densely spaced complex zeros which moreover are not close to the real axis but oriented along a diagonal in the complex plane. As a consequence the Blaschke correction becomes very tedious, one has to consider a very large number of zeros to achieve a reconstruction which might be called "acceptable" although it is far inferior to the reconstruction of the two-cosine-squared bunch.
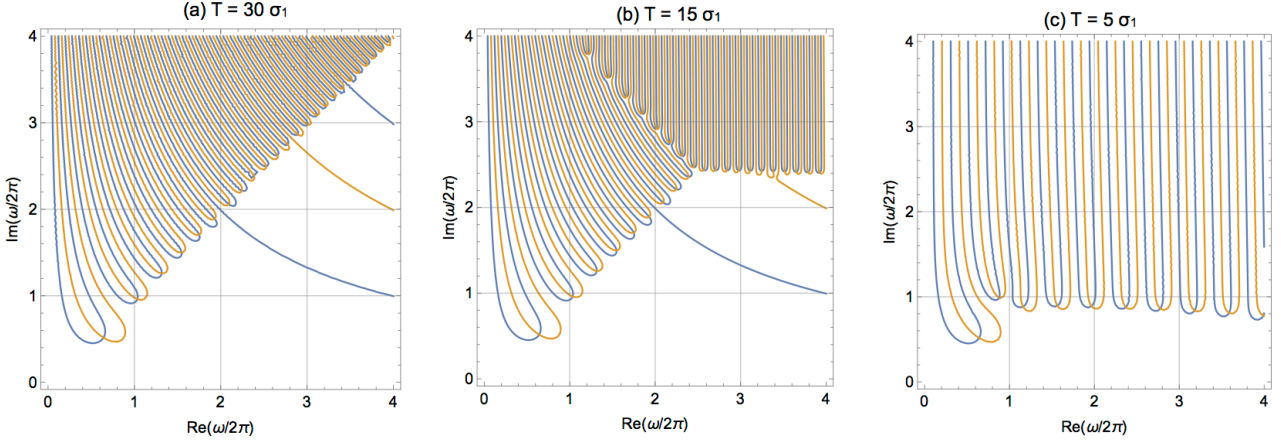

Figure 58: Superposition of two truncated Gaussians of different width ($\sigma_1 = 0.5\,\mathrm{ps}$, $\sigma_2 = 0.1\,\mathrm{ps}$), which are centered with respect to each other. The contour lines $\Re(\mathcal{F}_{\mathrm{tr}}(\hat{\omega})) = 0$ and $\Im(\mathcal{F}_{\mathrm{tr}}(\hat{\omega})) = 0$ in the complex $\hat{\omega}$ plane are plotted for three different cutoff times: **(a)** $T = 30\,\sigma_1$, **(b)** $T = 15\,\sigma_1$ and **(c)** $T = 5\,\sigma_1$.

Now we investigate whether this unpleasant property of the complex form factor is related to the non-causality of Gaussian functions. To test this idea we truncate the Gaussians in time domain and write the particle density distribution in the form[8]

$$
\begin{aligned}
\rho_{\mathrm{trunc}}(t) &= \frac{A_1}{\sqrt{2\pi}\,\sigma_1} \exp\left(-\frac{t^2}{2\sigma_1^2}\right) + \frac{A_2}{\sqrt{2\pi}\,\sigma_2} \exp\left(-\frac{t^2}{2\sigma_2^2}\right) \quad \text{for } |t| \le T, \quad (100)\\
\rho_{\mathrm{trunc}}(t) &= 0 \quad \text{for } |t| > T
\end{aligned}
$$

with a sufficiently large cutoff time T such that both Gaussians have dropped to tiny values. The Fourier transform of a single truncated Gaussian can be calculated analytically

$$
\begin{aligned}
\mathcal{F}_j(\omega) &= \frac{1}{\sqrt{2\pi}\,\sigma_j} \int_{-T}^{T} \exp\left(-\frac{t^2}{2\sigma_j^2}\right) \exp(i\,\omega t)\, dt \\
\mathcal{F}_j(\omega) &= \frac{\sqrt{\pi}}{2} \exp\left(-\frac{\sigma_j^2 \omega^2}{2}\right) \left[\mathrm{erf}\left(\frac{T - i\,\sigma_j^2 \omega}{\sqrt{2}\,\sigma_j}\right) + \mathrm{erf}\left(\frac{T + i\,\sigma_j^2 \omega}{\sqrt{2}\,\sigma_j}\right) \right]. \quad (101)
\end{aligned}
$$

The form factor of the distribution $\rho_{\mathrm{trunc}}(t)$ is

$$\mathcal{F}_{\mathrm{trunc}}(\omega) = A_1 \mathcal{F}_1(\omega) + A_2 \mathcal{F}_2(\omega).$$

It is very instructive to study the contour line pattern in the complex $\hat{\omega}$ plane for various values of the cutoff time T.

(a) Our first choice is a very large value, $T = 30\,\sigma_1$. The left picture in Fig. 58 looks exactly like Fig. 57b: if the cutoff time is much larger than the standard deviation σ_1 of the wider Gaussian, the form factor $\mathcal{F}_{\mathrm{tr}}(\omega)$ of

---

[8]To simplify the computation the Gaussians are truncated symmetrically at $t = -T$ and $t = +T$.

the truncated particle density distribution is almost identical with the form factor $\mathcal{F}(\omega)$ of the non-truncated distribution. The big theoretical advantage of the truncated function is however that it obeys causality.

(b) When the cutoff time is reduced the contour line pattern changes dramatically. For $T = 15\,\sigma_1$ the complex zeros follow the diagonal up to a certain frequency, and beyond that they are located a horizontal line.

(c) At $T = 5\,\sigma_1$ all zeros all arranged parallel to the real axis and their spacing is less dense. In fact, Fig. 59a shows great similarity with contour line pattern of the two cosine-squared pulses in Fig. 49. When the reconstruction phase $\Phi_{\mathrm{rec}}(\omega) = \Phi_{\mathrm{KK}}(\omega) - \Phi_{\mathrm{B}}(\omega)$ is used one obtains a good reconstruction of the original time profile, see Fig. 59b.
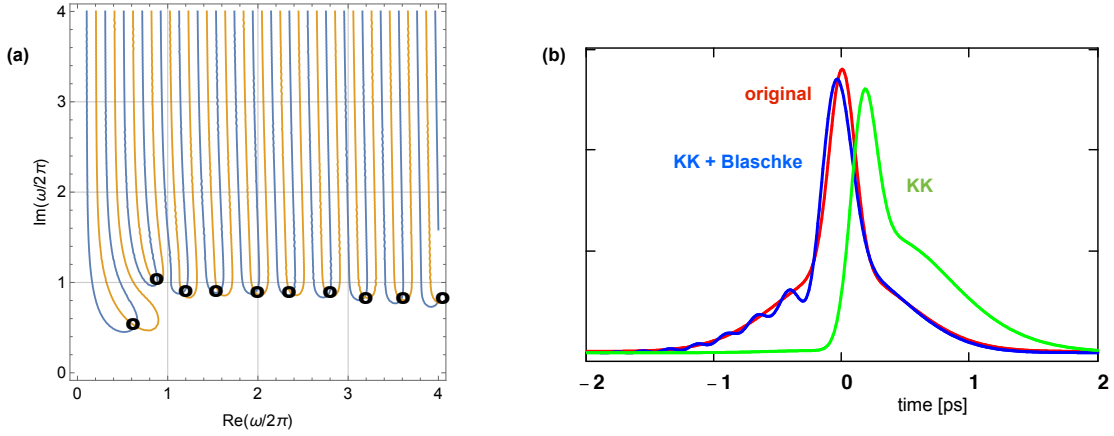

Figure 59: Reconstruction of a bunch consisting of two truncated Gaussian pulses with $\sigma_1 = 0.5\,\mathrm{ps}$ and $\sigma_2 = 0.1\,\mathrm{ps}$, which are both centered at $t = 0$. The cutoff time is $T = 5\,\sigma_1$. **(a)** Plot of the contour lines $\Re(\mathcal{F}(\hat{\omega})) = 0$ (blue) and $\Im(\mathcal{F}(\hat{\omega})) = 0$ (yellow) in the complex frequency plane. The black circles indicate the crossing points where the form factor $\mathcal{F}(\hat{\omega}))$ vanishes. **(b)** The reconstruction based on the Kramers-Kronig phase and the Blaschke phase (blue curve) is in rather good agreement with the original charge distribution $\rho(t)$ (red curve). Shown is also the Kramers-Kronig reconstruction without Blaschke phase (green curve) which fails to reproduce the original.

This example shows convincingly that the Blaschke correction should not be done with normal Gaussians but rather with truncated Gaussians.

9 Appendix D: Propagation of Radiation by Fourier Transformation

The analytical and semi-analytical computations in Section 2 are based on cylindrical symmetry. This symmetry is not present in many practical cases, for example when one wants to compute transition radiation from a rectangular or asymmetric screen. The numerical two-dimensional Fourier transformation permits to deal with such situations and offers the additional advantage that the radiation can be propagated through a whole optical system consisting of drift spaces, apertures and focusing elements such as parabolic or elliptic mirrors. The method is explained at length in Ref. [15], see also Refs. [38], [39]. Here we present the main ideas and formulas. A Mathematica™ code *THzTransport* [40] was developed for carrying out this analysis.

To explain the principle of electromagnetic wave propagation by Fourier transformation we consider a TR source without cylindrical symmetry. It is then appropriate to work in Cartesian coordinates. Position space is described by the position vectors $\boldsymbol{r} = (x, y, z)$ and Fourier space is described by the wave vectors $\boldsymbol{k} = (k_x, k_y, k_z)$ with $k = \sqrt{k_x^2 + k_y^2 + k_z^2} = \omega/c$. We choose general points $Q = (\xi, \eta, 0)$ on the source screen and $P = (x, y, D)$ on the observation screen (see Fig. 5).

When the distance D between source screen and observation screen is larger than the size of the image it is appropriate to apply the Huygens principle: each point on a wave front acts as origin of a small spherical wave, and the new wave front is formed by the superposition of these spherical waves. Using this principle, the horizontal field component at a general point $P = (x, y, D)$ is computed by a double integral over the source screen:

$$\tilde{E}_x(P, k) = -\frac{ik}{2\pi} \underbrace{\iint}_{\text{source}} \tilde{E}_x(Q, k) \frac{\exp(ikR')}{R'} \, d\xi d\eta \quad \text{with} \quad R' = \overline{QP}. \tag{102}$$

A corresponding expression holds for the vertical field component. The distance R' between Q and P is expanded up to second order

$$R' = \overline{QP} = \sqrt{D^2 + (x - \xi)^2 + (y - \eta)^2} \approx D + \frac{x^2 + y^2}{2D} - \frac{x\xi + y\eta}{D} + \frac{\xi^2 + \eta^2}{2D} \tag{103}$$

for $\xi^2 + \eta^2 \ll D^2$ and $x^2 + y^2 \ll D^2$. The horizontal component of the Fourier-transformed electric field is thus

$$\begin{aligned}
\tilde{E}_x(P, k) &= -\frac{ik}{2\pi} \frac{\exp(ikD)}{D} \exp\left(\frac{ik(x^2 + y^2)}{2D}\right) \\
&\quad \cdot \underbrace{\iint}_{\text{source}} \tilde{E}_x(Q, k) \exp\left(\frac{ik(\xi^2 + \eta^2)}{2D}\right) \exp\left(-i(k_x\xi + k_y\eta)\right) \, d\xi d\eta
\end{aligned}$$

with the transverse components of the wave vector $k_x = k\,x/D$, $k_y = k\,y/D$. The double integral can be written as a two-dimensional Fourier transformation

$$F_x(k_x, k_y) = \frac{1}{2\pi} \underbrace{\iint}_{\text{source}} G_x(\xi, \eta) \exp[-i(k_x\xi + k_y\eta)] \, d\xi d\eta \tag{104}$$

of the function

$$G_x(\xi, \eta) = \tilde{E}_x(\xi, \eta, k) \exp\left(\frac{ik(\xi^2 + \eta^2)}{2D}\right). \tag{105}$$

Corresponding expressions are obtained for the vertical component.

It is important to note that the Fourier integral (104) can describe both far-field and near-field diffraction since it contains the second-order phase factor $\exp\left(ik(\xi^2 + \eta^2)/(2D)\right)$.

The Fourier transform method is quite general and not restricted to the propagation of transition radiation from the TR source to an observation screen. In fact, the method permits the propagation of an arbitrary radiation field through a whole beamline in a stepwise procedure, going from one screen to the next. In each step, the complex electric field vector (amplitude, direction and phase) must be known in a fine grid on the "source screen". Using the above algorithm, the complex electric field vector (amplitude, direction and phase)

on the "observation screen" is then computed by FFT, again in a fine grid. In the next step, the previous "observation screen" is treated as the new "source screen". A screen can for example be an aperture, but it may also represent a lens or a focusing mirror of a given size and shape. In that case, position-dependent phase factors are applied to account for the focusing action. This is explained in Ref. [15].

References

[1] P. Schmüser, M. Dohlus, J. Rossbach, C. Behrens, *Free-Electron Lasers in the Ultraviolet and X-Ray Regime*, Springer 2014.

[2] H. Edwards, C. Behrens and E. Harms, *3.9 GHz Cavity Module for Linear Bunch Compression at FLASH*, Proc. Lin. Acc. Conf. 2010, Tsukuba, Japan.

[3] I. Zagorodnov and M. Dohlus, *Semianalytical modeling of multistage bunch compression with collective effects*, Phys. Rev. ST Accel. Beams **14**, 014403 (2011).

[4] B. Steffen *et al.*, *Electro-optic time profile monitors for femtosecond electron bunches at the soft x-ray free-electron laser FLASH*, Phys. Rev. ST Accel. Beams **12**, 032802 (2009).

[5] O. H. Altenmueller, R. R. Larsen und G. A. Loew, *Investigations of Traveling-Wave Separators for the Stanford Two-Mile Linear Accelerator*, Rev. Sci. Instrum. 35 (1964), 438.

[6] P. Krejcik *et al.*, *Commissioning of the SSPS linac bunch compressor*, Proc. Part. Acc. Conf. PAC 2003, p. 423, Portland, Oregon 2003.

[7] R. Akre *et al.*, *Bunch length measurements using a transverse RF deflecting structure in the SLAC linac*, Proc. Eur. Part. Acc. Conf. EPAC 2002, Paris 2002.

[8] M. Röhrs *et al.*, *Time-resolved electron beam phase space tomography at a soft x-ray free-electron laser*, Phys. Rev. ST Accel. Beams **12**, 050704 (2009).

[9] M. Röhrs, *Investigations of the phase space distributions of electron bunches at the FLASH-linac using a transverse deflecting structure*, Ph.D. thesis, University of Hamburg 2008, THESIS-2008-12.

[10] M. Geitz, *Investigations of the transverse and longitudinal beam parameters at the TESLA Test Facility Linac*, Ph.D. thesis Universität Hamburg 1999, DESY-THESIS-1999-033.

[11] S. Wesch *et al.*, *A Multi-Channel THz and Infrared Spectrometer for Femtosecond Electron Bunch Diagnostics by Single-Shot Spectroscopy of Coherent Radiation*, Nucl. Instrum. Meth. A **665**, 40 (2011).

[12] M.L. Ter-Mikaelian, *High-energy Electromagnetic Processes in Condensed Media*, John Wiley & Sons Inc. (1972).

[13] B. M. Bolotovskii and E. A. Galst′yan, *Diffraction and diffraction radiation*, Phys. Usp. **43**, 755 (2000).

[14] J.D. Jackson, *Classical Electrodynamics*, Third Edition, chapter 13, John Wiley 1999.

[15] S. Casalbuoni, B. Schmidt and P. Schmüser, *Far-Infrared Transition and Diffraction Radiation. Part I: Production, Diffraction and Optical Propagation*, DESY TESLA Report 2005-15 (2005).

[16] P. Schmüser, *Theoretische Physik 2 für Studierende des Lehramts: Elektrodynamik und Spezielle Relativitätstheorie*, Springer 2013.

[17] L.D. Landau, E.M. Lifshitz, *Course of Theoretical Physics, Vol. 8, Electrodynamics of Continuous Media*.

[18] M. Castellano et al., *Effects of diffraction and target finite size on coherent transition radiation spectra in bunch length measurements*, Nucl. Instr. Meth. **A 435**, 297 (1999).

[19] E. J. Akutowicz, *On the Determination of the Phase of a Fourier Integral*, Trans. Amer. Math. Soc. **83**, 179 (1956).

[20] M. H. Hayes, *The Reconstruction of a Multidimensional Sequence from the Phase or Magnitude of Its Fourier Transform*, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASP-30, No. 2, April 1982.

[21] F. Wooten, *Optical Properties of Solids*, Academic Press, New York 1972.

[22] R. Lai, U. Happek, and A.J. Sievers, *Measurement of the longitudinal asymmetry of a charged particle bunch from the coherent synchrotron or transition radiation spectrum*, Phys. Rev. E **50**, R4294 (1994).

[23] R. Lai, A.J. Sievers, *On using the coherent far IR radiation produced by a charged-particle bunch to determine its shape: I Analysis*, Nucl. Instrum. Meth. **A 397**, 221 (1997).

[24] W. Blaschke, *Eine Erweiterung des Satzes von Vitali über Folgen analytischer Funktionen*, Berichte Math.-Phys., Kl. Sächs. Gesell. der Wiss. Leipzig, 67 (1915), pp. 194 - 200.

[25] J.R. Fienup, *Phase retrieval algorithms: a comparison*, Appl. Optics **21**, 2758 (1982).

[26] S.I. Bajlekov et al., *Longitudinal electron bunch profile reconstruction by performing phase retrieval on coherent transition radiation spectra*, Phys. Rev. ST Accel.Beams **16**, 040701 (2013).

[27] D. Pelliccia and T. Sen, *A two-step method for retrieving the longitudinal profile of an electron bunch from its coherent radiation*, Nucl. Instr. Meth. **A 764**, 206 (2014).

[28] R.W. Gerchberg and W.O. Saxton, *A practical algorithm for the determination of phase from image and diffraction plane pictures*, Optik **35**, 227 (1972).

[29] T. Kövener, *THz Spectrometer Calibration at FELIX*, Master thesis, University of Hamburg 2016.

[30] C. Palmer and E. Loewen, *Diffraction grating handbook*, sixth edition, Newport Corporation 2005.

[31] S. Casalbuoni, B. Schmidt, P. Schmüser, V. Arsov and S. Wesch, *Ultrabroadband terahertz source and beamline based on coherent transition radiation*, Phys. Rev. ST Accel. Beams **12**, 030705 (2009).

[32] C. Behrens, *Detection and Spectral Measurements of Coherent Synchrotron Radiation at FLASH*, Diploma Thesis, University of Hamburg, Report DESY- THESIS-2010-02, 2008.

[33] S. Wesch, *Echtzeitbestimmung longitudinaler Elektronenstrahlparameter mittels absoluter Intensitäts- und Spektralmessung einzelner kohärenter THz Strahlungspulse*, Ph.D. thesis, University of Hamburg 2012.

[34] S. Marchesini et al., *X-ray image reconstruction from a diffraction pattern alone*, Phys. Rev. B **68**, 140101 (2003).

[35] M. Yan, *Online diagnostics of time-resolved electron beam properties with femtosecond resolution for X-ray FELs*, Ph.D. Thesis, University of Hamburg 2015, DESY-THESIS-2016-017.

[36] J.S. Toll, *Causality and the Dispersion Relation: Logical Foundations*, Phys. Rev. **104**, 1760 (1956).

[37] H. Cartan, *Elementare Theorie der analytischen Funktionen einer oder mehrerer komplexen Variablen*, BI Hochschultaschenbücher 112/112a, Bibliographisches Institut Mannheim (1966).

[38] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.

[39] G. N. Lawrence, *Optical modeling*, Vol. 11 of *Applied Optics and Optical Engineering*, Academic, San Diego, Calif., 1992.

[40] For information on the Mathematica code *THzTransport* please contact: Bernhard.Schmidt@Desy.de