FERMILAB-PUB-14-040-APC

A two-step method for retrieving the longitudinal profile of an electron bunch from its coherent radiation

Daniele Pelliccia[a], Tanaji Sen[b,*]

[a] *School of Physics, Monash University, Victoria 3800, Australia*
[b] *Accelerator Physics Center, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*

Abstract

The coherent radiation emitted by an electron bunch provides a diagnostic signal that can be used to estimate its longitudinal distribution. Commonly only the amplitude of the intensity spectrum can be measured and the associated phase must be calculated to obtain the bunch profile. Very recently an iterative method was proposed to retrieve this phase. However ambiguities associated with non-uniqueness of the solution are always present in the phase retrieval procedure. Here we present a method to overcome the ambiguity problem by first performing multiple independent runs of the phase retrieval procedure and then second, sorting the good solutions by mean of cross-correlation analysis. Results obtained with simulated bunches of various shapes and experimental measured spectra are presented, discussed and compared with the established Kramers-Kronig method. It is shown that even when the effect of the ambiguities is strong, as is the case for a double peak in the profile, the cross-correlation post-processing is able to filter out unwanted solutions. We show that, unlike the Kramers-Kronig method, the combined approach presented is able to faithfully reconstruct complicated bunch profiles.

Keywords: Phase retrieval, Electron bunch longitudinal profile, Kramers-Kronig, Iterative method, Coherent transition radiation, Photoinjector

*Corresponding author

Operated by Fermi Research Alliance, LLC under Contract No. De-AC02-07CH11359 with the United States Department of Energy.

1. Introduction

Short bunches are generated at photoinjectors such as the A0 photoinjector at Fermilab [1, 2], at X-ray Free Electron Lasers such as the Linac Coherent Light Source (LCLS) at SLAC [3] and will be a key requirement for high luminosity in the proposed International Linear Collider (ILC) [4]. At the new ASTA photoinjector at Fermilab, now being commissioned, bunches shorter than a picosecond will be created using a two stage compression scheme [5]. Streak cameras are used to measure bunch lengths down to the picosecond scale but shorter bunches require either electro-optical methods [6], transversely deflecting rf cavities [7] or methods which use the coherent radiation at infrared wavelengths emitted by the bunch. Such radiation can have different sources such as synchrotron radiation, transition radiation, and diffraction radiation to name a few. The radiation is coherent when the wavelength is comparable or longer than the bunch length.

At the A0 photoinjector the coherent transition radiation (CTR) emitted when the electron beam traverses a thin metallic foil has been used to determine the longitudinal distribution of the electron beam. The distribution is reconstructed from the measured amplitude and calculated phase of the intensity spectrum. The phase calculation uses the Kramers-Kronig (KK) method which relates the real and imaginary parts of the spectrum via a Hilbert transform. While the KK method is relatively straightforward and has been used extensively, see e.g. [8, 9, 10, 11, 12, 13], it has a few drawbacks which will be discussed below. It is therefore useful to develop an alternative technique to reconstruct the longitudinal distributions of short electron bunches.

Very recently a different approach was proposed [14]. It is based on the *iterative* retrieval of the phase from the knowledge of the measured amplitude and imposing suitable constraints which must be satisfied by the longitudinal bunch profile. This approach belongs to a much wider class of iterative methods that aim to indirectly measure a signal from the magnitude of its Fourier transform [15, 16, 17]. By retrieving the missing phase in Fourier space, the signal can be reconstructed unambiguously. In [14] such a procedure was employed to retrieve the longitudinal profile of simulated and measured CTR spectra and a heuristic approach to optimize the iterative reconstruction was given.

A fundamental issue with these iterative algorithms is that the underlying mathematical problem is inherently ill-posed [18] meaning that a number 2N

of unknown quantities (for instance amplitude and phase of the target function) must be reconstructed starting from only N measured values (i.e. the amplitude of the spectrum). In practice though, under reasonable assumptions on the object function to be retrieved, the problem can be uniquely solved in two or more dimensions [19] while fundamental ambiguities are still present in the one-dimensional case [16, 18, 19].
This aspect was briefly discussed in [14], showing that in some cases of experimental significance, the solution is in fact unique. Nevertheless a more general approach, able to deal with the ambiguities is desirable. This is the subject of this paper, which aims to extend the results obtained in [14], discussing in greater detail the problem of uniqueness of the solution. We shall show that this is extremely relevant when the longitudinal bunch distribution has more than a single peak. In this case the fundamental ambiguities inherent to the 1D problem can play a non-trivial role.
We proposed an approach inspired from a similar procedure derived for x-ray optics [20]. This method is based on the *a posteriori* selection of the results of the reconstruction of different algorithm runs, starting from independent random estimates. The selection is made on the basis of the cross-correlation that each reconstruction displays with a reference one.
The paper is organized as follows: in Sec. 2 the basis of the KK method is reviewed and the limitations associated with it are discussed. Sec. 3 deals with the problem of the ambiguities in iterative phase retrieval, while Sec. 4 is devoted to the detailed description of our method. Sec. 5 reports the results obtained with our method applied to simulated bunch profiles of characteristic shape: Gaussian, Lorentzian and combinations of the two. The effect of noise is also discussed here. The results related to measured bunch profiles at the A0 photoinjector are shown in Sec. 6 and discussed in Sec. 7. Conclusions are drawn in Sec. 8.

2. The Kramers-Kronig method and its limitations

We briefly discuss the KK method here, more complete discussions can be found in [9, 11]. Let $s(z)$ denote the longitudinal distribution of an electron bunch, then its Fourier transform (FT) is

$$S(\omega) \equiv \rho(\omega)e^{i\phi(\omega)} = \int s(z)\exp[i\frac{\omega}{c}z]\,\mathrm{d}z. \tag{1}$$

Here $\rho(\omega)$ and $\phi(\omega)$ are the amplitude and phase respectively of the complex function $S(\omega)$. The bunch form factor $F(\omega)$ is defined as

$$F(\omega) \equiv \rho^2(\omega) = S(\omega)S^*(\omega). \tag{2}$$

The form factor is related to the bunch intensity spectrum $I(\omega)$ via

$$I(\omega) = I_1(\omega)[N + N(N-1)F(\omega)], \tag{3}$$

where I_1 is the spectrum of a single particle and N is the number of particles in the bunch. In the limit of a large bunch population the intensity spectrum of coherent radiation $I(\omega) \propto F(\omega)$ or equivalently the amplitude $\rho(\omega) \propto \sqrt{I(\omega)}$.

Since $S(z)$ is a causal function, the real and imaginary parts of its FT are related by the Hilbert transform. Using this relation, the phase can be written as the sum of two terms $\phi(\omega) = \phi_m(\omega) + \phi_B(\omega)$ where ϕ_m is known as the minimal phase and ϕ_B is known as the Blaschke phase. They can be written as:

$$\phi_m(\omega) = -\frac{2\omega}{\pi}\mathcal{P}\int_0^\infty \mathrm{d}x \frac{\ln(\rho(x))}{x^2-\omega^2}, \quad \phi_B(\omega) = \sum_j \arg(\frac{\omega-\omega_j}{\omega-\omega_j^*}). \tag{4}$$

In the expression for ϕ_m, $\mathcal{P}$ denotes the Cauchy principal value while in the formula for ϕ_B, ω_j is the jth zero of $S(\omega)$ in the upper half of the complex plane. Only the zeros nearby to the real axis have a meaningful impact on the phase. When the zeros are far away, the Blaschke phase is linear in the frequency but a linear frequency shift corresponds to a spatial translation of the profile and does not affect the shape of the bunch. If there are no nearby zeros, then the minimal phase is a useful approximation to the total phase. It can be written in a form without the apparent pole-like singularity

$$\phi_m(\omega) = -\frac{2\omega}{\pi}\int_0^\infty \mathrm{d}x \frac{\ln(\rho(x)/\rho(\omega))}{x^2-\omega^2}. \tag{5}$$

The bunch profile can be obtained from the minimal phase by an inverse Fourier transform,

$$s(z) = \frac{1}{\pi c}\int_0^\infty \rho(\omega)\cos[\phi_m(\omega) - \frac{\omega}{c}z]\mathrm{d}\omega. \tag{6}$$

For a Gaussian bunch, the minimal phase is linear in ω while for a general asymmetric bunch profile, ϕ_m is nonlinear in ω.

The KK method suffers from two main limitations: (i) the modulus of the spectrum must be known in principle for all frequencies and (ii) the minimal phase must be a good approximation to the total phase [11]. In practice one needs to extrapolate the measured spectrum to very low frequencies [11] and calculate the phase from Eq. (5).
The assumption of no nearby zeros is valid for Gaussian functions and combinations of Gaussians but it is not valid for other shapes with long tails such as Lorentzians. For these shapes the Blaschke phase cannot be ignored and approximating the phase by the minimal phase does not result in an accurate reconstruction, as will be seen later. In a photoinjector with a bunch compressor, the longitudinal profile can be far from Gaussian and it not clear *a priori* that the minimal phase approximation will be accurate.
From the description above, the need for an alternative method is clear. The iterative phase retrieval approach can offer such an alternative: it is valid even in the presence of nearby zeros of $S(\omega)$ and can be more robust against missing data. Nonetheless the iterative method also suffers from limitations, which will be discussed in the next section.

3. Ambiguities in 1D phase retrieval procedures

Iterative phase retrieval methods have been applied to numerous 2D problems in optics and crystallography [21], electron microscopy [22] and x-ray imaging [23]. They are well suited when the direct imaging of a sample is not possible due to the lack of a suitable objective lens. For example, aberrations of the objective may prevent imaging at the diffraction limit or the efficiency of the objective lens makes a direct imaging procedure impractical.
The iterative method, in its original form is based on the fast Fourier transform (FFT) routine to numerically propagate the wave front from the sample plane to the detector plane and *vice versa*. The algorithm is constrained by the measured data in the Fourier space, and by any available *a priori* information in real space. Typically one imposes a "support" constraint in real space, namely imposing that the solution must be zero outside a certain region. Support information is in many cases sufficient to retrieve the phase and the algorithm is very robust against noise and also missing information [24]. In most applications, phase retrieval algorithms aim at solving some imaging problem and therefore are naturally implemented with 2D data. When a 1D problem is to be treated, fundamental ambiguities in the solution do exist [19, 16], which we will briefly describe in the following.

Let $s(t) = s(z/c)$ denote the bunch distribution in the time domain. It is straightforward to prove that for any given real numbers t_0 and α, the amplitudes of the FT of the functions $s(t)$, $s(t + t_0)$, $\exp(i\alpha)s(t)$ and $s^*(-t)$ are the same (the symbol * denotes complex conjugation). In other words if the distribution is changed with any of the above transformations , the measured spectrum $\rho^{meas}(\omega)$ will be unchanged. While a constant shift and a constant phase are actually trivial ambiguities, the ambiguity between $s(t)$ and $s^*(-t)$ is more complicated to handle. In imaging it represents the so called "twin-image" problem (see [25] for a recent review). This problem corresponds to the stagnation of the iterative algorithm which is unable to converge on either $s(t)$ or $s^*(-t)$ and keeps bouncing between a combination of the two.

Fienup [16] reported iterative retrieval of 1D functions pointing out that depending on the constraints applied, the 1D iterative procedure can have a unique solution or multiple yet correlated solutions. The presence of correlated solutions and the poor convergence due to the presence of twins is very significant for the reconstruction of longitudinal bunch profiles, as the set of possible solutions becomes very large. For instance when the distribution is asymmetric and composed of multiple, partially overlapping peaks, the relative heights and widths of the peaks can be incorrectly reconstructed due to the presence of twins of possible solutions.

Recently, a method was proposed to improve the reconstruction of a 1D signal from a single measurement of the magnitude of its FT [20]. The method works in two steps. The first step is an iterative phase retrieval algorithm employing known numerical strategies such as hybrid input-output or error reduction [17]. This procedure is repeated many times with different independent random choices of the initial phase $\phi(\omega)$ and the corresponding set of independent reconstructions is stored. At this point the second step is taken, consisting in a post-selection of the reconstructions on the basis of their cross-correlations with a reference reconstruction. The selected profiles are then averaged to produce the final result. Therefore, while ambiguities and poor reconstructions can still affect each individual solution, the post-selection ensures that only the significant reconstructions are considered. In the next section we describe in detail both steps of the method.

4. Description of the method

In this section we describe the strategies to implement the iterative phase retrieval procedure optimized for the measurement of an electron bunch profile. As anticipated, the method is divided in two steps, the first consisting of many independent runs of a suitable iterative phase retrieval algorithm, followed by result post-selection.

4.1. Iterative phase retrieval algorithm

The aim is to reconstruct the electron bunch longitudinal distribution $s(t)$ from the knowledge of the measured spectrum amplitude $\rho^{meas}(\omega)$. In reality we deal with discrete arrays of data, therefore in the following we shall use the discrete notation s_j to denote the j-th element of the array s. Lower-case symbols will denote arrays in real space (time domain) while capital letters will indicate arrays in Fourier space (frequency domain) The iterations start by defining a random array $\phi^{(0)}$ and assign it as phase to the measured amplitude, thus building the complex-valued spectrum at the zero-th iterate: $S^{(0)} = \rho^{meas} \exp(i\phi^{(0)})$. Then the following sequence is iterated:

1. Inverse FT the function $S^{(0)}$ to get the zero-th estimate of the bunch longitudinal distribution $s^{(0)}$;
2. Impose the time domain constraints, by applying any *a priori* information on the distribution, to get a new iterate of the time domain function $s^{(1)}$;
3. Calculate its FT, to get a subsequent estimate in the frequency domain $S'^{(1)}$ and calculate its phase $\phi^{(1)} = \arg\left(S'^{(1)}\right)$;
4. Update the frequency domain iterate by imposing the frequency domain constraint (i.e. the measured data), $S^{(1)} = \rho^{meas} \exp(i\phi^{(1)})$;

The procedure is iterated until convergence is attained, i.e. the solution does not appreciably change with further iterates. To monitor the quality of the convergence two different figures of merit (FOMs) are used. The first accounts for the variation of the time domain function between subsequent iterates

$$\Delta_k = \sum_j \frac{|s_j^{(k+1)}| - |s_j^{(k)}|}{|s_j^{(k)}|} \tag{7}$$

We assume that the convergence is reached whenever Δ_k becomes less than some threshold value. In addition, a second figure of merit is employed

defined as

$$\epsilon_k = \sum_j \frac{|S_j^{(k)}| - \rho_j^{meas}}{\rho_j^{meas}} \tag{8}$$

ϵ_k is used to monitor how close the kth estimate of the amplitude of the frequency domain function – before applying the frequency domain constraint – is to the measured data. By minimizing both figures of merit we ensure that the algorithm is both converging and the solution is close to the actual measured data.

The constraint in the time domain is imposed by using all types of *a priori* information that are available for the problem under study. For the results reported here we used a geometrical constraint, that means imposing the bunch length cannot exceed a certain duration. Such duration – which we call "support" – is especially important to guarantee the convergence of the algorithm. If it is too small the algorithm fails to converge because it cannot find a suitable solution fitting within the support size. If the support is too large it is likely that two or more ambiguous solutions are present at the same time, resulting in a poor convergence [25]. The support can be approximately estimated by the auto-correlation function associated with the measured spectrum [26] but care must be used if the longitudinal bunch distribution has multiple adjacent peaks or long tails, as we shall discuss in the next section.

An additional simplification (that may be included as support) comes from the nature of the function to be retrieved. The longitudinal profile of the electron bunch has to be necessarily described by a real and positive function. This fact reduces the ambiguity of Hermitian conjugation to a simpler time reversal.

To update the solution at each step using the time domain constraints, we used several cycles, each composed of a combination of hybrid input-output (HIO) followed by error reduction (ER) algorithms [17]. The HIO is known to ensure a faster convergence, being less prone to stagnations, while the ER steps are beneficial in reducing the reconstruction errors.

A last remark concerns the possibility of dynamically updating the support to closely follow the object reconstruction. This corresponds to the so-called "shrink-wrap" algorithm [27] and it has been successfully adapted to the phase retrieval applied to CTR [14] where the authors were able to reduce the effects of ambiguities using this feature. This is indeed very beneficial when the distribution is composed of a single pulse with short tails (such as

a Gaussian) or multiple well separated pulses. In this case the procedure will reduce the support size to the approximate size of the pulse (or the multiple pulses) and the algorithm will converge easily to the right shape. As we will discuss in the following, problems arise when the profile has long tails (as in a Lorentzian profile) or it is composed of partially overlapping pulses. In this case the support constraint cannot be very strict and the convergence of the iterative procedure is generally poorer. Therefore in this study we did not adopt a shrink-wrap option but introduced a solution-sorting procedure as a second step.

4.2. Post-selection of the reconstruction results

This procedure closely follows the one described in [20] which we shall summarize here. From the first step of the method we obtain a set of N independent results, each result being contained in an array with size J. Let us denote by s_j^n the pixel j of the n-th result, where $j = 1, ..., J$ and $n = 1, ..., N$. We shall denote with s^n the array n.

The aim of the post-selection procedure is to select the best results and average them, considering the figures of merit defined in the previous sub-section. Associated with each result are the values Δ and ϵ calculated with Eqs. (7) and (8) respectively at the last step of each iteration run. The specific values of these FOMs vary depending on the profile to be reconstructed, typically values of $\Delta, \epsilon \lesssim 10^{-3}$ indicate an acceptable convergence. If the values of the FOMs fail to meet this criterion the results are discarded, so that the set of N results obtained at the end will meet the required level of convergence.

The N results will be affected by ambiguities described above, the most important in our case being translation shift, time reversal or a mild problem related to the presence of twins.

The key step at this point is to select the best solution obtained, seeking the lowest value of ϵ, indicating that the modulus of the reconstructed function in Fourier space is very close to the measured data. We then check that, associated with this solution, also that Δ has a low value, indicating that the algorithm did effectively converge. Typically when ϵ is small, Δ is also small, while the converse is not always true, i.e. it is possible for the algorithm to converge to the wrong solution. The solution having the smallest value of ϵ is chosen as the reference reconstruction, denoted by s^R in the following. Every other solution is quantitatively compared to the reference by calculating the

cross-correlation:

$$c_h^n = \frac{\sum_j s_j^R s_{j+h}^n}{\left(\sum_j |s_j^R|^2\right)^{1/2} \left(\sum_j |s_{j+h}^n|^2\right)^{1/2}}. \tag{9}$$

The cross correlation is calculated for each solution as a function of the relative shift h. This procedure is necessary to account for any shift ambiguity that might be present. In practice h is a cyclic shift, i.e. one always considers $(j+h) \mod J$.
To account for the time reversal ambiguity a second cross correlation array is also computed:

$$\bar{c}_l^n = \frac{\sum_j s_j^R s_{J-j+l}^n}{\left(\sum_j |s_j^R|^2\right)^{1/2} \left(\sum_j |s_{J-j+l}^n|^2\right)^{1/2}}. \tag{10}$$

In Eq. (10) each array is reversed before computing the cross correlation. In order to choose between ambiguous solutions one compares the cross correlations in Eqs. (9) and (10). In practice, denoting with $\bar{h}$ and $\bar{l}$ the values of the shift that maximize Eqs. (9) and (10) respectively, this corresponds to seeking the maximum between $c_{\bar{h}}^n$ and $\bar{c}_{\bar{l}}^n$:

$$p_n = \max\left(c_{\bar{h}}^n, \bar{c}_{\bar{l}}^n\right). \tag{11}$$

The value of p_n indicates how well a solution is correlated with the reference. Therefore only solutions that after this procedure have a high correlation (say $p_n \geq 0.9$) are averaged.

5. Phase retrieval of simulated spectra

We performed the phase retrieval with the procedure described in the previous section using simulated spectra of three representative distributions: Gaussian, Lorentzian and different superpositions of the two. All profiles were simulated using 4096 data points.
The Gaussian case is the easiest and a very good reconstruction can be obtained both with the KK method [10] and with the phase retrieval method [14], as shown in Fig. 1a. The Gaussian has been simulated using rms width σ =70 and its maximum amplitude value was set to unity.
On the contrary, the reconstruction of a Lorentzian pulse is non-trivial in both

scenarios, but for different reasons. The complex spectrum associated with a Lorentzian distribution has zeroes in the upper half complex plane close to the real axis, therefore the KK method (based on the minimal phase) is not able to reconstruct the pulse shape properly. At the same time a Lorentzian curve has long tails, comparatively longer than a Gaussian. This fact makes the phase retrieval reconstruction based on geometrical support constraint also problematic, as the approximation of time-limited (bound) signal is not completely met.
A satisfactorily reconstruction is attained by increasing the size of the support (compared to the FWHM of the pulse). This improves the convergence even though some distortion is still present in the pulse tails, as evident from Fig. 1b. The FWHM of the simulated Lorentzian was w =50.
When a random noise is added to the pulse, the performance of the iterative algorithm is still robust. Fig. 1c shows the Gaussian case with noise. It is worth noting that the KK reconstruction, which works well for a noiseless Gaussian, has some issues when noise is added. This is not surprising considering that the spectrum amplitude is no longer a smooth analytic function, so the Hilbert transform relation between the amplitude and phase can only be approximately true. Using the iterative procedure, the retrieved profile with noise always appear smoother due to the averaging during the post-selection process. These observations are confirmed when a Lorentzian pulse with noise is simulated (Fig. 1d). The iterative algorithm is still capable of reproducing the tails of the pulse correctly, while still displaying a slightly larger width.

In Fig. 2 the retrieved shapes from simulated spectra related to double pulses are shown. The simulated bunches contain different superpositions of a Gaussian and a Lorentzian shape with different degree of overlap, which can be typical of many experimental situations. In all cases we considered a Gaussian with σ =70 and a Lorentzian with width w =50. The maximum value of the Gaussian was set to 60% of the corresponding value for the Lorentzian and the maximum height of the whole profile has been normalized to unity. We studied the effect of varying the separation between the pulses and compared the results from the iterative reconstruction to those with the KK method.

Fig. 2a shows the reconstruction obtained when the distance between the centers of the peaks is 40. In this case the profile appears rather asymmetric with different tails. The iterative reconstruction is very close to the input profile, while an evident discrepancy appears for the KK reconstruction. A

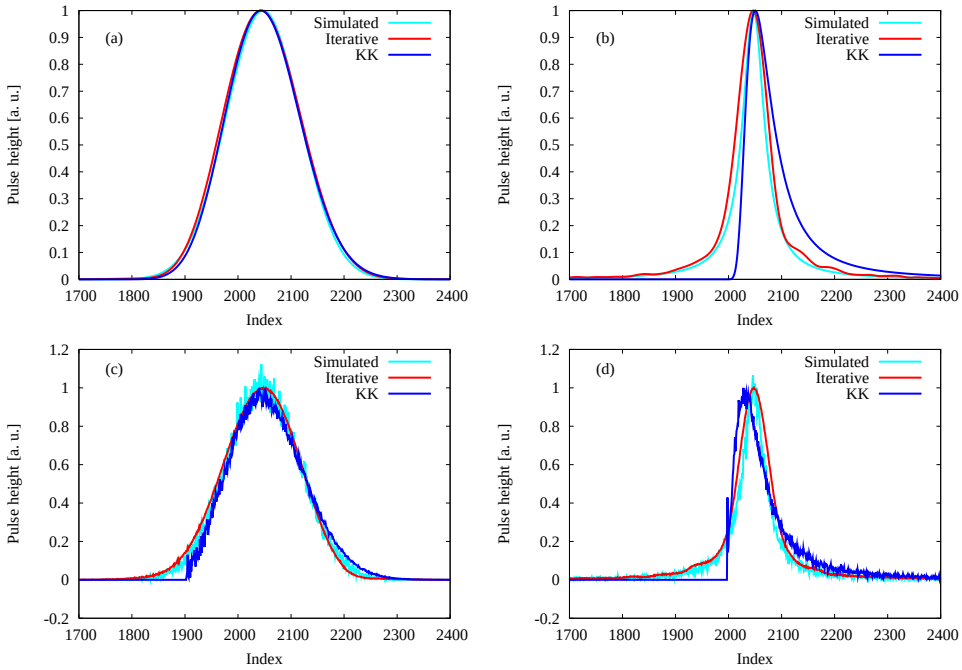

Figure 1: Plots of the simulated profile (cyan), reconstructions using the KK method (blue) and the iterative phase retrieval (red). (a) Gaussian profile. (b) Lorentzian profile. (c) Gaussian profile with noise. (d) Lorentzian profile with noise.

similar conclusion can be drawn from Fig. 2b where the separation between the peak centers is 100 but the two peaks are still partially overlapping. A value of 200 for the peak separation has been chosen for the results shown in Fig. 2c, without noise, and in Fig. 2d, with noise, respectively. In this case the overlap between the pulses is much reduced and detrimental effects of time reversal ambiguities appear to be stronger. In particular the relative heights of the peaks is not correctly reproduced and, when noise is present, distortions in the tails do appear.

6. Experimental profiles at the A0 photoinjector

The data used here were taken at Fermilab's A0 photoinjector during 2011. This was during a period when the Tevatron collider was operational and liquid helium was available for a 9-cell superconducting cavity in the beamline. We describe the experimental setup briefly here, more complete descriptions can be found in [28, 29, 30, 31]. Electron pulses are generated

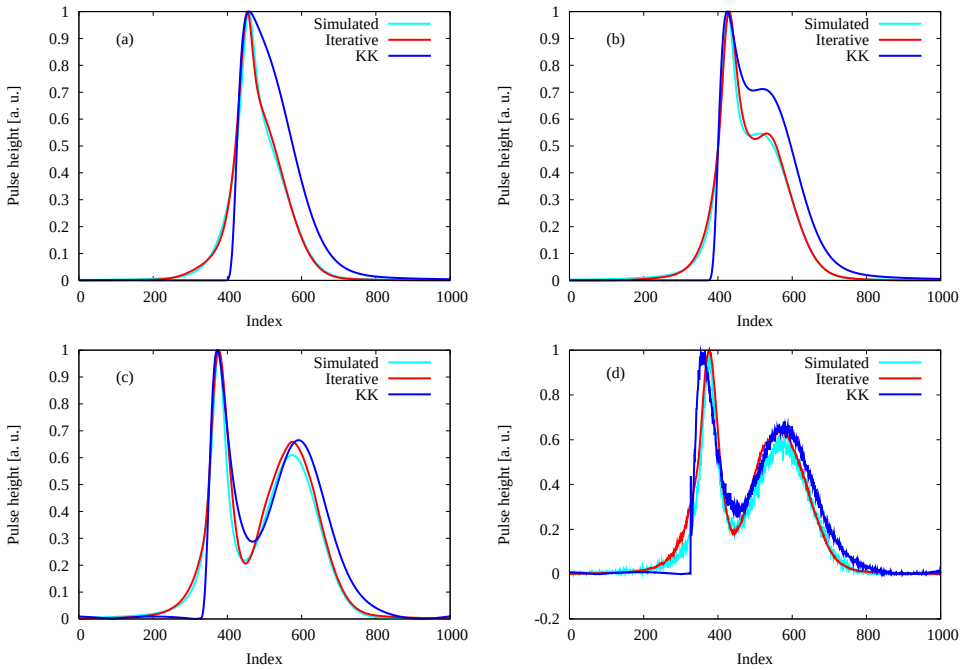

Figure 2: Plots of the simulated profile (cyan), reconstructions using the KK method (blue) and the iterative phase retrieval (red). The profiles are the sum of a Gaussian with $\sigma = 70$ and a Lorentzian with width $w = 50$ (see text). The distance d between the peak centers is varied (a) $d = 40$. (b) $d = 100$. (c) $d = 200$. (d) $d = 200$ with noise added.

with a CsTe photo-cathode within a 1.3 GHz rf gun and then accelerated with a 1.3GHz superconducting rf cavity to an energy of about 15 MeV. Bunch charge could be varied over the range 250pC - 1nC. A transverse to longitudinal emittance exchange (EEX) section consisting of a transverse deflecting mode cavity placed between two doglegs could be used to vary the bunch length by changing quadrupole settings upstream of the EEX [30]. In addition, the bunch length could also be controlled by varying the off-crest rf phase (aka energy chirp) in the 9-cell accelerating superconducting cavity which was upstream of the EEX section. Downstream of the EEX, the beam passed through a thin metallic foil (Al coating on Si substrate) generating coherent transition radiation (CTR) in the process. The radiation entered a Martin-Puplett interferometer which uses a a polarizing splitter to send the radiation along two orthogonal arms with mirrors at the ends. The mirrors change the polarizations and the beams after reflection are recombined using the same splitter and then sent to two pyroelectric detectors after being split again with another polarizing splitter. One of the arms is movable so the path length between the interfering beams can be varied. An interferogram or auto-correlation function $C(t)$ is obtained from the intensity in the detectors; measured as a function of $t = \Delta z/c$ where Δz is the path length difference from the central beam splitter to the two mirrors. The real part of the Fourier transform of the auto-correlation function yields the intensity spectrum $I(\omega)$. The detector's response is frequency dependent and is limited at low frequencies by the thickness of the pyroelectric crystal. Consequently the measured response drops to zero at frequencies below about 0.1 THz. This is corrected during post-processing with a parabolic fit which smoothly extends the response to zero frequency [29]. We use the corrected intensity curves in the following.

Figure 3 shows the reconstructed profiles from measurements taken over four different days. In these measurements, the bunch length was varied only by changing quadrupole settings upstream of the EEX and no energy chirp was applied. Both the KK and the iterative profiles are shown in Fig. 3 and they are scaled to the same peak height for ease of comparison.

On another day, the effect of an energy chirp on the bunch length was measured by varying the off-crest rf phase. The coherent synchrotron radiation (CSR) emitted in the second dogleg was monitored and at the maximum compression (17° off-crest) the CSR power reached its peak value. A discussion of this measurement and the CSR emitted by these bunches can be found in [31]. The auto-correlation function was measured with the chirp

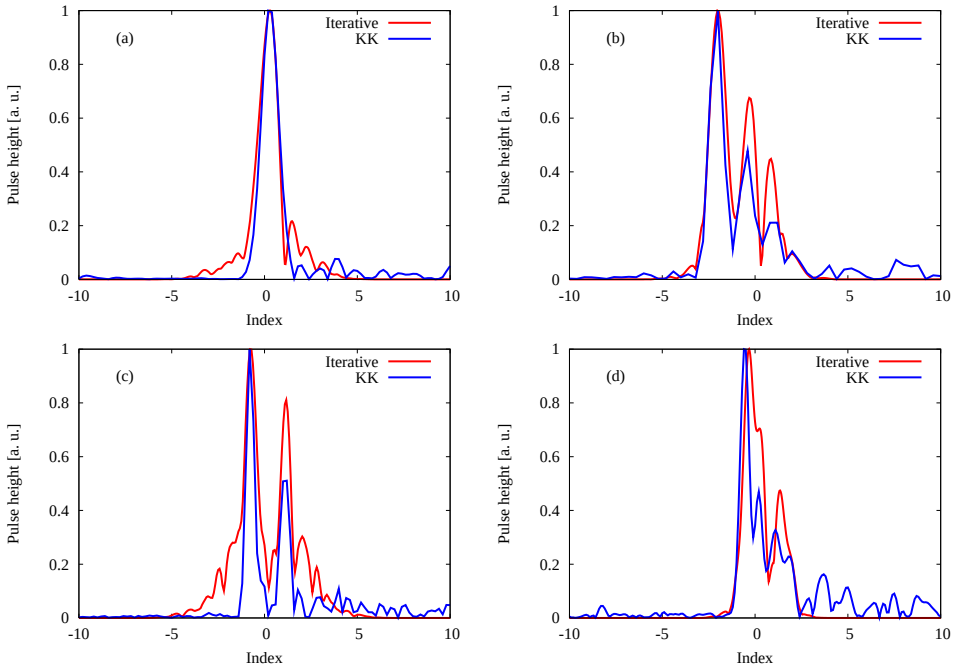

Figure 3: Reconstruction of experimental data with iterative algorithm (red) and with the KK method (blue). Data were taken in 2011 and on different dates: (a) May 9, (b) Aug 24, (c) Aug 25, (d) Oct 5

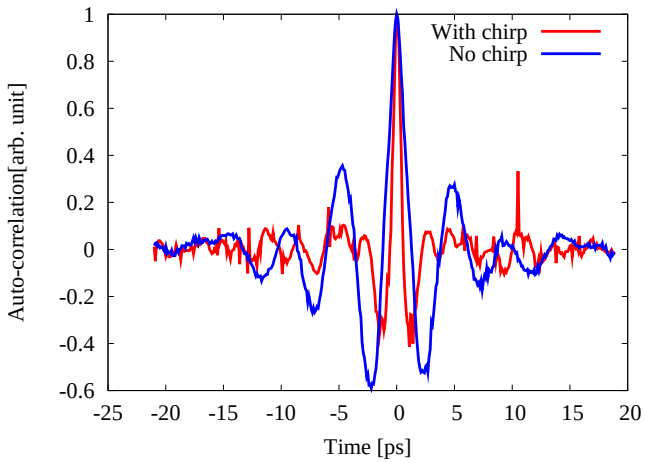

Figure 4: Measured auto-correlation function of two pulses, data taken on July 1, 2011. The data were taken with additional bunch compression using a rf chirp (red) and also without the chirp (blue). See the text for a more detailed description.

set at maximum compression and also without any chirp and are shown in Fig. 4. As expected, the auto-correlation with the chirp has a significantly smaller full width at half maximum (FWHM).
Figure 5 shows the reconstructed profiles of these two cases, both with the iterative and the KK method. Again, the profiles have been scaled to the same peak height.

7. Discussion of reconstruction results

It is instructive to compare the quality of the reconstructions of simulated profiles to understand the merits and limitations of the iterative algorithm plus post-selection, over the KK method. Both the new method and the KK method perform very well for the reconstruction of the Gaussian profile alone – Fig. 1a. The FWHMs from the two methods are shown in Table (7).
The new method presented here tends to overestimate the FWHM while a slight under-estimate result from the KK. The reason for the over-estimate lies in the post-selection procedure. While averaging the best correlated solutions, a residual translation shift between them will result in an overall larger average. Nevertheless the effect is relatively small ($\sim 3\%$) and of little importance under normal circumstances.
The same effect is more significant in the reconstruction of the Lorentzian

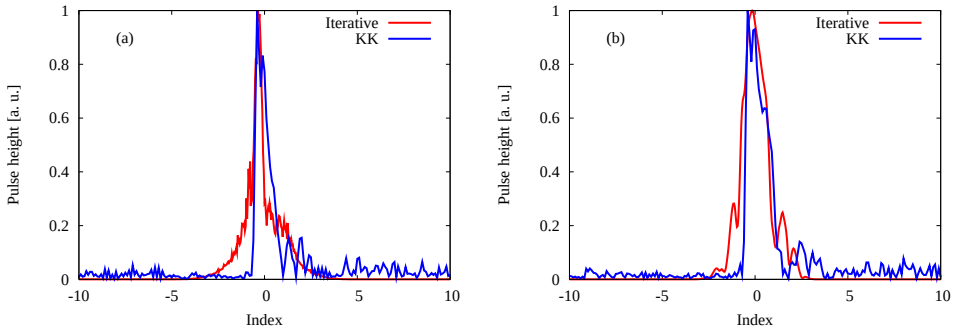

Figure 5: Reconstruction of experimental data taken on July 1, 2011 with the iterative method (red) and the KK method (blue) (a) Compressed pulse with chirp, (b) Pulse without chirp.

bunch, shown in Fig. 1b. The overestimation of the peak width is comparatively larger here because, in order to correctly reproduce the long tails typical of the Lorentzian, the geometrical support constraint must be loosened, and this in turn produces a poorer localization of the peak. Therefore when averaging the results the presence of this residual shift ambiguity is more significant and yields a larger overestimation of the width. Nevertheless the profile shape closely follows the correct profile. The KK method on the other hand produces a steep vertical slope on one side and a more slowly falling slope on the other side. These effects, arising from ignoring the Blaschke phase, seems to be characteristic of the KK method when reconstructing profiles with long tails, like Lorentzians.

From this discussion we conclude that the support size is an extremely critical parameter, when reconstructing a bunch with long tails. We chose the optimal support size by seeking the best reconstruction in preliminary runs of the algorithm. If the support is too small compared to the profile width, the algorithm fails to converge and secondary small peaks may appear. Quantitatively this is manifested by both FOMs ϵ and Δ failing to decrease. Therefore the procedure is to increase the support size, which automatically produces smaller value of Δ (i.e. increasing the support size makes the time domain constraint less and less effective). Therefore the optimal support size, on average, guarantees a low value of ϵ while avoiding stagnation of the algorithm.

The case of a double pulse is more convoluted. When two or more peaks are present, the time reversal ambiguity plays a much more critical role. Incorrect reconstructions often contain twins resulting in a larger number of

Table 1: FWHM of the simulated profiles, compared with one reconstructed by KK method and iterative retrieval.

Profile type	FWHM input	FWHM KK	FWHM iterative
Gaussian	165	163	170
Lorentzian	50	64	68
Gaussian w noise	165	160	170
Lorentzian w noise	50	68	64

peaks or in an incorrect estimation of the peak heights and widths. In this case the solution sorting procedure is a key factor in filtering out incorrect reconstructions.

The comparison in Fig. 2, between the iterative reconstruction and the KK method shows that the former always produces a result much closer to the input profile. Specifically, the peak positions are always correctly estimated. Nonetheless the quality of the reconstruction critically depends on the peak separation. The iterative reconstructions are extremely close to the simulated profiles for small peak separation (see Fig. 2 a and b) while becoming worse when the separation increases. Notably the relative peak height is not faithfully reproduced. The dip between the peaks is more correctly reproduced with the iterative reconstruction.

It is worth noting that the cases studied here – either a Lorentzian profile or the partially overlapping peaks – are generally hard to tackle with algorithms of the kind of "shrink-wrap". The double peak case is especially complicated as the shrink-wrap is able to restrict the support to the total extent of the profile, but ambiguities can still be present in reconstructing the position/width of each individual pulse.

The comparison of the experimental profiles in Fig. 3 shows similarities and differences that are consistent over the measured profiles. Table 7 shows the FWHMs from the two methods for the experimental profiles. Profile (a) from May 9 is seen to have only a single peak in both reconstructions. As before, the KK profile shows a very steep slope at the head of the pulse while the iterative profile builds up more gradually. This feature is repeated in all the measured profiles and was also seen in the simulated profiles of Figs. 1 and 2. The iterative profiles have higher secondary peaks in cases (b),(c),(d) in Fig. 3 and do not have the multiple small peaks in the tail when compared to the KK profiles. If real, these small peaks in the profile

Table 2: FWHM of the KK profiles compared with the FWHM from iterative retrieval. In the case of bunches with multiple peak, the FWHM is relative the highest peak.

Profile date	FWHM KK [ps]	FWHM iterative [ps]
May9	1.2	1.2
Aug 24	0.8	1.1
Aug 25	0.6	0.8
Oct 5	0.6	1.2
Jul 1, with chirp	0.8	0.5
Jul 1, no chirp	1.4	1.5

tails would indicate high frequency micro-structure These are unlikely to be present in the A0 photoinjector bunches given the relatively low charge and the lack of external excitations. It is therefore possible that the iterative reconstruction avoids the appearance of unphysical micro-structure, but this needs to be verified with detailed studies in the ASTA photoinjector.
Comparing the profiles in Fig. 5, we find that both methods show that the chirp reduced the pulse width relative to the unchirped case but the iterative profile is narrower. The KK profile does not show that the additional compression with the chirp reduced the FWHM compared to profiles obtained without chirp on other days, e.g. compared to the profiles seen in Figs 3c and 3d. The iterative profile on the other hand shows that the chirp results in the smallest FWHM among all profiles. This corresponds nicely to the auto-correlation function (seen in Fig 4) of the chirped pulse which also had the the smallest FWHM amongst all auto-correlations considered here. This is again suggestive, but not definitive evidence, that the iterative reconstruction may be more accurate.

8. Conclusions

We have described a two-step iterative phase retrieval method to reconstruct the phase of the spectrum from the measured spectrum amplitude. In the first step, multiple reconstructions are performed with different initial random phases. During each reconstruction, constraints in frequency space and real space are applied successively. Only those solutions are kept which are both sufficiently close to measured spectrum amplitude and have a high rate of convergence. In the second step, a post selection method using cross-

correlations is employed to remove ambiguities related to translation and reversal in real space. The final solution is obtained by averaging over the solutions surviving the post-selection step. The post-selection is crucial in removing these troublesome ambiguities that are common with complicated profiles such as those with multiple peaks.

In applying this procedure to Gaussians, Lorentzians, sums of Lorentzians and Gaussians, we found that this iterative method is able to retrieve the profiles with reasonable accuracy and much more faithfully than the Kramers-Kronig method based on the minimal phase. The latter method is known to be inaccurate for profiles with long-tails such as Lorentzians. When applied to experimental data taken at the A0 photoinjector, the iterative method yielded profiles without some of the unphysical features in the profiles of the minimal phase KK method.

The major drawback of the iterative method at present is its iterative nature, so it takes longer and requires some trial and error before the optimal solution is found. With some additional work, this procedure could be automated and made available in control room applications.

Acknowledgments

D.P. acknowledges the support of the Australian Research Council. T.S thanks Charles Thangaraj and Randy Thurman-Keup for generously sharing their data and useful discussions. Fermilab is operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy.

References

[1] J.P. Carneiro et al, Phys. Rev. ST Accel Beams, **8**, 040101 (2005)

[2] A. Lumpkin and J. Ruan, Proceedings of the Beam Instrumentation Workshop 08, p 258 (2008).

[3] Z. Huang et al, Proceedings of the 2011 Particle Accelerator Conference, p 2459 (2011)

[4] ILC Technical Design Report, arxiv:1306.6328

[5] P. Piot et al, Proceedings of the 2013 Particle Accelerator Conference, p 67 (2013)

[6] I. Wilke et al, Phys. Rev. Lett,**88**, 124801 (2002)

[7] R. Akre et al, Proceedings of the 2001 Particle Accelerator Conference, p 2353 (2001)

[8] R. Lai, U. Happek and A.J. Sievers, Phys. Rev. E 50, R4294-R4297 (1994).

[9] R. Lai and A.J. Sievers, Phys. Rev. E 50, R3342-R3344 (1994).

[10] R. Lai and A.J. Sievers, Phys. Rev. E 52, 4576-4579 (1995)

[11] R. Lai and A.J. Sievers, Nucl. Instrum. Meth. Phys. Res. A 397, 221-231 (1997).

[12] G. Schneider, R. Lai, W. Walecki and A.J. Sievers, Nucl. Instrum. Meth. Phys. Res. A 396, 283-292 (1997).

[13] D. Mihalcea, C.L. Bohn, U. Happek, and P. Piot, Phys. Rev. ST Accel. Beams 9, 082801 (2006).

[14] S.I. Bajlekov *et al.*, Phys. Rev. ST Accel. Beams 16, 040701 (2013).

[15] R.W. Gerchberg and W.O. Saxton, Optik (Stuttgart) 35, 237-250 (1972).

[16] J.R. Fienup, Opt. Lett. 3, 27-29 (1978).

[17] J.R. Fienup, Appl. Opt. 21, 2758-2769 (1982).

[18] D.M. Paganin, Coherent X-Ray Optics, Clarendon Press, Oxford, 2006.

[19] R.H.T. Bates, Optik 61, 247-262 (1982).

[20] D. Pelliccia, D.M. Paganin, A. Sorrentino, I. Bukreeva, A. Cedola, S. Lagomarsino, Opt. Lett. 37, 262-264 (2012).

[21] R.P. Millane, J. Opt. Soc. Am. A 7, 394-411 (1990).

[22] M.J. Humphry, B. Kraus, A.C. Hurst, A.M. Maiden, J.M. Rodenburg, Nature Comm. 3, 730 (2012).

[23] K.A. Nugent, Adv. Phys. 59, 1-99 (2010).

[24] J. C. H. Spence, U. Weierstall, and M. Howells. Phil. Trans. Roy. Soc. Lond. A **360**, 875-895 (2002).

[25] M. Guizar-Sicairos and J.R. Fienup, J. Opt. Soc. Am. A 29, 2367-2375 (2012).

[26] J.R. Fienup, T.R. Crimmins, and W. Holsztynski, J. Opt. Soc. Am. 72, 610-624 (1982).

[27] S. Marchesini *et al.*, Phys. Rev. B 68, 140101R (2003).

[28] L. Frohlich, DESY-Thesis-2005-011 (2005)

[29] R. Thurman-Keup et al, Proceedings of the Beam Instrumentation Workshop 08, p 153 (2008).

[30] J. Ruan et al, Phys. Rev. Lett, **106**, 244801 (2011)

[31] J.C.T. Thangaraj et al, Phys. Rev.ST-AB, **15**, 110702 (2012)