Recovering the phase and amplitude of X-ray FEL pulses using neural networks and differentiable models

D. Ratner,1* F. Christie,2 J. P. Ryan,1,3 A. Edelean,1 A. Lutman,1 and X. Zhang1

1SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
2Deutsches Elektronen-Synchrotron, Hamburg 22001-22769, Germany
3Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
*dratner@slac.stanford.edu

Abstract: Dynamics experiments are an important use-case for X-ray free-electron lasers (XFELs), but time-domain measurements of the X-ray pulses themselves remain a challenge. Shot-by-shot X-ray diagnostics could enable a new class of simpler and potentially higher-resolution pump-probe experiments. Here, we report training neural networks to combine low-resolution measurements in both the time and frequency domains to recover X-ray pulses at high-resolution. Critically, we also recover the phase, opening the door to coherent-control experiments with XFELs. The model-based generative neural-network architecture can be trained directly on unlabeled experimental data and is fast enough for real-time analysis on the new generation of MHz XFELs.

1. Motivation

X-ray free-electron lasers (XFELs) produce femtosecond pulses of X-rays for a wide variety of photon science applications. The Linac coherent light source (LCLS) [1] uses the self-amplified spontaneous emission (SASE) process to generate a train of 120 independent pulses per second, with each pulse itself composed of statistically independent spikes. While most current XFEL applications assume only average properties of the pulses, in principle new science could emerge with pulse-to-pulse measurements of the individual X-ray spikes as a function of time. For example, probing ultrafast dynamics of linear X-ray-matter interactions requires only knowledge of the X-ray arrival times (delays) and possibly the X-ray spectrum. To this end, we previously investigated applying time-domain ghost imaging to linear pump-probe experiments, using the fluctuating separation of SASE spikes to recover dynamics of a sample [2]. However, if one hopes to exploit more powerful X-ray analogues of optical nonlinear spectroscopies [3–6] precise knowledge of the full electric field of the X-ray pulse is necessary, i.e., both the amplitude and the phase. For example, in the recently-demonstrated stimulated X-ray Raman scattering technique [7–9], the relative phase between the field at the pump and Stokes transitions can alter the population transfer dynamics. Beyond this basic non-linearity, higher-dimensional measurements can provide detailed information of time-dependent electronic structure with atomic site-specificity [10].

At present, direct measurement of full X-ray fields is not feasible in general, and even time-domain measurement of the power remains a challenge. Recent single-shot full-field measurements at FERMI used a variation of FROG [11], but this technique has yet to be demonstrated in the X-ray regime or for the higher number of modes present in SASE XFELs and requires a challenging experimental setup. At the shortest pulse lengths (sub-10fs) the power profile can be retrieved by THz or infrared streaking of photoelectrons generated by the X-ray interaction with a gas [12–16]. At longer pulse lengths, indirect measurement is possible by...
observing the pattern of energy loss in the electron beam that generated the X-rays using an X-band transverse cavity (XTCAV) [17–19]. The XTCAV is the dominant method used for pulse measurement at LCLS, but the resolution is not sufficient to observe individual SASE spikes except at the longest wavelengths, and the XTCAV measurement contains no phase information. On the other hand, frequency-domain measurements are relatively straightforward [20,21], and there is a long history in optics of applying iterative phase retrieval to spectral measurements to reconstruct ultrafast pulses (see e.g., FROG, SPIDER, and more recent variations [22–28]). The same concept can be applied to XFELS, combining corrupted measurements of the X-ray power in both the time and frequency domains to improve the resolution of the time-domain measurement [29,30]. In this paper we extend the work of [30] to reconstruct the full field, giving not only higher resolution of the time-domain power, but also revealing the relative phase between spikes. We do so with an efficient machine-learning implementation that can run at the full beam-rate of future MHz XFELs.

2. Pulse reconstruction task and proposed method

Formally, our pulse reconstruction scheme seeks to recover the X-ray field, $f(t) \in \mathbb{C}^p$, given the time domain power, $P(t) \in \mathbb{R}^{m_t}$, and the spectral power, $S(\nu) \in \mathbb{R}^{m_\nu}$, where each measurement is divided into $m_t$ and $m_\nu$ discrete time and frequency measurements respectively and the reconstructed field has $p$ discrete points. In practice, we are only given a corrupted measurement of both domains, which we will denote $\tilde{P}(t)$, $\tilde{S}(\nu)$, and our goal is to find a mapping

$$\tilde{P}(t), \tilde{S}(\nu) \rightarrow f(t).$$

(1)

Note that recovery of a unique phase is not strictly guaranteed for one-dimensional functions. However, the correct solution can often be found given sufficient constraints from the time and frequency properties of the function, and it has been shown [31] that a one-dimensional function is nearly uniquely determined from its modulus and the modulus of its Fourier transform. For the case of pulse reconstruction, we must accept certain unavoidable ambiguities, for example the overall phase of the field or time-reversal of the phase for time-symmetric pulses. In the special case of a multi-spike SASE FEL pulse, which can be modeled as a sum of finite, independent Gaussian pulses, we will see that recovery is often feasible and unique.

Equation (1) is an inverse problem: while the mapping from power to field is difficult to learn, we can write down a ‘forward’ mapping $\Phi : f(t) \rightarrow \tilde{P}(t), \tilde{S}(\nu)$ that, if given the true field, can predict the associated corrupted measurements:

$$\tilde{P}(t) = |f(t)|^2 \star R_t + N_t$$

$$\tilde{S}(\nu) = |\mathcal{F} \mathcal{T} \{f(t)\}|^2 \star R_\nu + N_\nu,$$

(2)

where $R_t$ and $R_\nu$ are the measurement resolution functions in the time and frequency domains respectively, $N_t$ and $N_\nu$ are measurement Gaussian noise in the time and frequency domains respectively, $\mathcal{F} \mathcal{T}$ denotes a Fourier transform, and ‘$\star$’ denotes a convolution. Inverting Eq. (2) is difficult due to the loss of phase information. The existing state-of-art for X-ray FEL pulse recovery [30] works by guessing a solution for $f(t)$, and iteratively updating that guess to improve consistency with Eq. (2). The problem is that the iterative task must be repeated for every pulse; with the LCLS-II upgrade operating at up to a MHz (generating 10 s of billions of pulses for a single experiment) and each iterative solution taking seconds to converge (see Appendix, Section C), recovering the field for every pulse in an experiment is unrealistic.

Instead, we will train a neural network (NN) inverse mapping, an approach that has found success for inverse problems in other areas of physics [32], including recent works for two-dimensional phase retrieval [33–36] and accelerators [37]. Though NN training is computationally expensive, inference requires only a single forward-pass through the network, potentially many
orders of magnitude faster than the existing iterative solution. Neural networks also can leverage prior information contained in the training set (e.g., the FEL coherence length) which could both recover a higher resolution amplitude and also reveal the phase of the field. In our case, the NN’s input features are the power measurements, $x$: $\tilde{P}(t), \tilde{S}(\nu)$, and the labels are the complex field, $y$: $f(t)$. The complex field labels, $y$, consist of either the field’s real and imaginary components, $\mathcal{R}\{f(t)\}, \mathcal{I}\{f(t)\}$, or by writing the field as $f(t) = A(t) \exp(i\phi)$, the amplitude and phase, $A, \phi$; in the first case we form $y$ by concatenating $\mathcal{R}\{f(t)\}$ and $\mathcal{I}\{f(t)\}$, and in the second case by concatenating $A$ and $\phi$, in either representation making a vector of $2p$ labels for each example.

We then train the network by minimizing the difference between the predicted labels, $y_{\text{pred}}$, and the ground truth labels, $y_{\text{GT}}$, with a mean absolute error loss function

$$L_{\text{mae}} = \frac{1}{2np} \sum_{i}^{n} \sum_{j}^{2p} |y_{\text{pred}}^i(j) - y_{\text{GT}}^i(j)|,$$

(3)

where the sum is over the $n$ training examples and the $2p$ labels in each example. For the amplitude/phase representation we again use mean absolute error for the amplitudes, but use a modified loss function on the phase portion of the labels

$$L_{\text{amp}} = \frac{1}{np} \sum_{i}^{n} \sum_{j}^{p} |A_{\text{pred}}^i(j) - A_{\text{GT}}^i(j)|$$

$$L_{\text{sine}} = \frac{1}{np} \sum_{i}^{n} \sum_{j}^{p} |A_{\text{GT}}^i(j) \sin \left(\frac{\phi_{\text{pred}}^i(j) - \phi_{\text{GT}}^i(j)}{2}\right)|.$$  

(4)

The sine loss was designed so that a phase error of $\pi$ has maximal penalty, and an offset of $2\pi$ is not penalized. Including $A_{\text{GT}}$ ensures that phase errors are ignored in regions with zero amplitude where the phase is undefined. In the end, we will train with the real/imaginary representation, but the phase representation is useful for evaluating the performance separately on amplitude and phase components.

The difficulty in training a NN with Eq. (3) again lies in the phase recovery: because the absolute phase cannot be measured, each pulse corresponds to a family of solutions

$$\phi \in \{\phi_0 + \xi\} \text{ with } \xi \text{ mod } 2\pi = \text{const.},$$

(5)

with $\xi$ a manifold of solutions that are all equally valid. During training, there is no way for the network to know which particular solution from the manifold was chosen as the label, and the loss between the given label and another equally valid solution on the same manifold could be as large as that of an incorrect solution on a different manifold. Figures 1(c) and 1(d) illustrate an example: given the solid lines as labels, Eq. (3) would strongly penalize a NN for producing the equally valid dashed lines.

In this paper, we employ two solutions to handle the arbitrary absolute phase. First, we select a single solution on the phase manifold so that each example corresponds to a well-defined label. We chose to define $\phi = 0$ (or equivalently purely real field) in the center of the pulse, and then train using Eq. (3). Though the definition collapses the phase manifold to a single point, if the center of the pulse is a waist between spikes, the phase may be changing rapidly, which will amplify errors. We will call this first approach the ‘$L_{\text{mae}}$ NN.’

Second, rather than restricting the labels to a single point on the manifold, we instead map the entire manifold back onto the well-defined input space, and define the loss function on the input...
Fig. 1. We simulate power measurements in the time domain (a) and spectral domain (b) including noise and measurement resolution. Our goal is to recover the complex field, represented either as the real and imaginary components (c) or equivalently the phase and amplitude (d). The solid and dashed lines show fields with different phases that map to the identical power; both fields should be considered correct solutions to the inverse problem.

The loss function then becomes

\[ L_\Phi = \sum_i \sum_j \left| \Phi(y_{\text{pred}}(j)) - x'(j) \right|^2, \]  

where \( \Phi \) is the forward model defined by Eq. (2). This second approach is functionally equivalent to the physics-informed neural network (PINN), which was previously introduced to solve inverse problems in the context of partial-differential equations [38]. We will use the PINN terminology going forward.

We can understand the PINN approach through the concept of generative NNs: rather than trying to predict a specific label, the NN learns to draw samples from the correct manifold. We start by considering a generative adversarial network (GAN) [39], which consists of two networks: a generator that creates new examples given a random 'latent' variable, and a discriminator that classifies examples as from the training set or from the generator; the training process rewards the discriminator for correctly classifying examples, and rewards the generator when it tricks the discriminator. In our case, the first network would need to be a “conditional” generator [40], i.e. it would take power measurements \( P(t), \tilde{S}(\nu) \) as inputs, and then generate corresponding example fields \( R\{f(t)\}, I\{f(t)\} \) as outputs. The generator's task is to learn to produce outputs drawn from the manifold corresponding to the inputs.
In a typical GAN, the discriminator’s role is to identify examples drawn from an incorrect manifold. In our setup we can replace the learned NN discriminator with the known, physical forward model, Eq. (2), which maps the generator’s output back onto the measured input power space. This forward model loss function plays the same role as the discriminator, penalizing solutions from the generator that are drawn from the wrong manifold. Crucially, by writing the forward model in a differentiable language (in this case PyTorch), we can directly apply the forward model to the NN output and train with standard NN platforms.

Before continuing, we note that similar forward-model approaches have been implemented over the last few years. For example, automatic differentiation has been used to incorporate physical models into inverse problems [34,41–43], and is expected to play a growing role in machine learning for the sciences broadly [44]. We also note similarity of our approach to the recently published phaseGAN [36], which is in turn derived from the cycleGAN architecture [45]; both make use of untrained models to build loss functions without labels. In a similar spirit, recent works have used another type of generative model known as a variational autoencoder to resolve ambiguities in signal reconstruction [46,47]. Finally, we note that in our example we do not want to map out the full manifold, so there is no need to input a random latent variable. As such our architecture is not strictly-speaking a generative network, and instead becomes equivalent to a PINN.

We present both training architectures in Fig. 2. We highlight that Eq. (6) does not include $y_{GT}$, meaning that the PINN is trained without labeled data; the only constraint is that the predictions must be consistent with the measurements, as defined by $\Phi$. We can further extend the PINN concept to also force the predictions to obey statistical properties. For example, FEL pulses have a correlation length given by the Pierce $\rho$-parameter, which can be observed through the first-order correlation

**Fig. 2.** Schematic of the training flow. At left, the input to the NN inverse model is the power and spectrum from either measured or simulated pulses. The NN outputs are the real and imaginary field components (center, green). If using simulations, a standard labeled loss function (e.g., Eq. (3)) updates the NN parameters to minimize the difference between predicted and simulated fields. Alternatively, a physics forward model, $\Phi$, maps the predicted outputs into predicted inputs (right, green), and an unlabeled loss function (Eq. (6)) updates the NN parameters by comparison to the true inputs. Additional loss terms can be imposed directly on the statistical properties of the predictions, for example the first-order correlation, $g_1$ (Eq. (7)). Combinations of all three loss functions are also possible.
order correlation function [48]

\[ g_1(\tau) = \frac{\langle f(t)f(t-\tau)^* \rangle}{\sqrt{\langle |f(t)|^2 \rangle \langle |f(t-\tau)|^2 \rangle}}, \tag{7} \]

where the expectation values are taken over time, \( t \), and \( \tau \) is the time-domain delay. Just as Eq. (6) forces the solution to be consistent with the measurements, the \( L_{g_1} \) loss penalizes solutions for not matching the expected \( g_1(\tau) \) function:

\[ L_{g_1} = \sum_{\tau} \left| g_{1\text{pred}}(\tau) - g_{1\text{GT}}(\tau) \right|^2, \tag{8} \]

where \( d \) is the number of delays in the comparison. The ground truth correlation function, \( g_{1\text{GT}} \), can be approximated analytically from the measured FEL gain length (see e.g., Eq. 6.67 in [48]) or calculated numerically from a small number of simulations, so as in Eq. (6) there is no need for a full labeled dataset. Note that \( g_1 \) is a statistical property, and the expectation value in Eq. (7) is taken over all examples in each training batch, so there is no explicit summation over examples in Eq. (8).

3. Results and discussion

We have implemented NNs using both a standard labeled loss function with \( L_{\text{mae}} \), as well as a PINN trained with \( L_{\Phi} \). The simulation parameters are taken from typical short-pulse, soft X-ray mode at LCLS, with 2.5 nm wavelength, 8-12 fs pulse length, and 2-5 spikes per pulse. The training set consists of 120k 1D FEL simulations [49], with a split of 100k/10k/10k for training, validation, and testing. The \( L_{\text{mae}} \) was trained with both real/imaginary and amplitude/phase labels and both had similar performance, so we only present the real/imaginary-trained NN here. We run simulations with various diagnostic resolutions \( R_t \) and \( R_\nu \), and apply measurement noise of 2% during training and testing. Because the forward model may not be known perfectly, particularly for the measured \( R_t \) resolution function of the time-domain measurements, we introduce errors during training to make the network less susceptible to \( R_t \) errors during inference. In the examples given, we train with an average \( R_t \) 10% higher than the value used in the test set. We then randomly vary the value of \( R_t \) by 15% rms during creation of the training dataset for both methods, and additionally in the loss function (Eq. (2)) while training the PINN. Note that robustness to systematic errors has not been studied and requires more assessment.

We start by assuming FWHM resolutions of \( R_t = 3 \) fs for the XTCAV and \( R_\nu = 80 \) meV for the spectrometer, with results shown in Figs. 3 and 4. Table 1 gives summary statistics for other resolution values. To give a sense for the quality of predictions, we show additional pulse examples in the appendix, Figs. 5 and 6. We find that training with \( L_{\text{mae}} \) gives the best performance. However, some users may still prefer the PINN because it takes only unlabeled inputs for training, and thus is trainable directly on experimental data; there is no need to simulate a new training dataset matching the exact experimental conditions, and the parameters of the training set are guaranteed to match those of the test data. The PINN is also more resistant to over-training, with equal losses for training and test sets. A combined training, using both \( L_{\Phi} \) and \( L_{g_1} \) loss functions and with \( L_{\text{mae}} \) applied to the amplitude portion of the outputs, had the best performance predicting field amplitudes (see Appendix, Section A). The higher accuracy of the labeled NNs compared to the PINN is expected; because the labeled NNs are trained on pulses drawn from the same distribution as the test data, these models effectively have a strong prior influencing the field reconstruction. The poor temporal resolution of the power measurement increases the impact of the \( L_{\text{mae}} \) NN’s prior. While using the \( L_{g_1} \) loss gives the PINN weak prior information, additional physical constraints may be needed for the PINN’s performance to equal that of the \( L_{\text{mae}} \) NN.
Fig. 3. Pulse example assuming $R_t = 3$ fs and $R_\nu = 80$ meV with error near the median score: $L_{\text{max}}$ NN error = 0.16 (median 0.16) and PINN error = 0.52 (median 0.51). Top row shows the power and spectral power features used as inputs to the model (black dotted line), as well as the result of applying the forward model to the predicted fields (blue and orange). Middle and bottom rows show the ground truth and predictions in real/imaginary representation (middle row) or amplitude and phase (bottom row). The phases have no meaning at the edges of the pulse where the amplitude goes to zero. For visualization purposes, we run an optimization on the prediction to find the solution on the predicted manifold that is closest to the label; this would not be necessary during inference.
Fig. 4. Error distribution for model with $R_t = 3$ fs and $R_\nu = 80$ meV. (A) Comparison of errors for the $L_{\text{mae}}$ and PINN models. Selecting shots with $L_{\Phi}$ below a threshold (red points) reduces prediction errors during inference. (b) $L_{\text{mae}}$ NN, field error as a function of the full-width pulse length shows a slight increase at longer pulses. Below, errors are shown separately for the amplitude (left) and phase (right). Black stars denote reconstruction errors for the pulse in Fig. 3.

Table 1. Summary of Performance for Different Values of Temporal and Spectral Resolution$^a$  

| $R_t$ (fs) | $R_\nu$ (meV) | Field err $L_{\text{mae}}/L_{\Phi}$ | Amplitude err $L_{\text{mae}}/L_{\Phi}$ | Phase err $L_{\text{mae}}/L_{\Phi}$ |
|------------|---------------|------------------------------------|----------------------------------------|-----------------------------------|
| 1.5        | 80            | 0.13/0.40                          | 0.90/2.66                              | 0.69/2.14                         |
| 3.0        | 80            | 0.16/0.51                          | 1.27/3.71                              | 0.79/2.75                         |
| 5.0        | 80            | 0.18/0.60                          | 1.49/4.90                              | 0.86/2.91                         |
| 0.6        | 400           | 0.14/0.46                          | 0.76/1.62                              | 0.82/3.04                         |
| 0.0        | 0             | 0.10/0.14                          | 0.65/0.65                              | 0.57/0.84                         |

$^a$Line 4 swaps poor temporal resolution for equivalent poor spectral resolution (compared to Line 2). Line 5 shows performance in the absence of experimental resolution limits.
Fig. 5. Example plots of reconstructions using the $\mathcal{L}_{\text{mae}}$ NN, with field errors of 0.11 for row 1 (median score for 5% cut on forward model error), 0.31 for row 2 (worst case for 5% cut on forward model error), and 0.55 for row 3 ($99^{\text{th}}$ percentile worst score over all). All plots use $R_t = 3$ fs and $R_\nu = 80$ meV.
Fig. 6. Example pulse reconstructions using the PINN, with field errors of 0.14 for row 1 (median score for $R_t = R_v = 0$), 0.28 for row 2 (median score for 5% cut on forward model error with $R_t = 3$ fs, $R_v = 80$ meV), and 1.03 for row 3 (90$^{th}$ percentile worst score over all with $R_t = 3$ fs, $R_v = 80$ meV).
The PINN’s predicted fields have larger errors than the \( L_{\text{mae}} \) NN’s, but in many examples the corresponding forward-model errors are approximately the same. This observation could indicate inherent phase ambiguity, which may be present for fields with a high degree of symmetry. However, for SASE FEL pulses with at least a few spikes, the phase errors appear dominated by the poor time-resolution of the XTCAV diagnostic. As a check, we trained a PINN under the assumption of perfect XTCAV resolution (see final line Table 1). We find field errors decrease by a factor of 3.7 and phase errors decrease by a factor of 3.3 compared to the standard 3 fs case, confirming the phase reconstruction errors in Figs. 3 and 4 are dominated by diagnostic resolution rather than inherent ambiguities in the phase. (See Figs. 6(a) and (b) for an example pulse with median error.) Indeed in this idealized case, the PINN nearly matches the \( L_{\text{mae}} \) NN performance. Even though the PINN is not able to fully resolve features at the time-scale of single spikes with realistic diagnostics, the predictions still provide more information about the amplitude and phase than is currently available to users.

With any machine learning algorithm, it is important to understand when the method fails. The PINN has the advantage that it is trained by the same metric as used in [30], so a user can easily compare the results through the residuals of the predicted inputs. On the other hand, in Fig. 4 the \( L_{\text{mae}} \) NN’s field errors are correlated to \( L_\Phi \), which can be calculated even for test data. We can use this correlation to throw away points that are more likely to have large field errors. For LCLS-II, when an experiment can collect a million shots per second, it may be feasible to throw away questionable data aggressively. For example, a cut retaining just 5% of the data reduces the \( L_{\text{mae}} \) NN’s median field label error by 30% and the maximum error by more than a factor of three (see red points in Fig. 4). Similarly for the PINN, a 5% cut reduces the median field label error by 40% and the maximum field error by a factor of 2. Future work could also estimate the uncertainty of the networks themselves on a shot-by-shot basis, for example using Bayesian dropout during inference [50].

While the iterative method used in Refs. [29,30] does not provide estimates of the phases, we can compare the NN and iterative performance on the field amplitudes. The NN approaches achieve higher accuracy than the iterative method, with the \( L_{\text{mae}} \) NN’s median amplitude errors more than a factor of 3 smaller, and the PINN’s errors 20% smaller than the iterative method. As expected, the NN is also many orders of magnitude faster, capable of processing millions of shots in a second (see Appendix, Section C). It may also be possible to combine the NN and iterative methods, using the NN as a starting guess for the iterative process; this should allow faster convergence than starting from a random guess, while providing strictly equal or better performance compared to the NN. Similar ideas have been implemented in other phase retrieval problems [35,51].

4. Conclusion and future outlook

We have presented a method to recover the full field, phase and amplitude, shot-by-shot for XFEL pulses, opening a path towards coherent-control experiments. We describe two possible architectures, one trained on standard labeled data, and one trained solely using a physical model, allowing that version to be learned directly from unlabeled experimental data. Both methods recover the amplitude with higher accuracy—by a factor of 3 for the labeled NN—than the existing iterative method. With sufficient diagnostic resolution, the proposed NNs also have the ability to recover the phase, which has yet to be demonstrated for multi-spike SASE XFEL pulses by any method. Finally, the NNs have potential to run in real time for MHz repetition rate machines.

Additional applications of the concept may be found for advanced operating modes, for example to resolve microbunching-instability sidebands in seeded FELs [52–56], which can then be accounted for during data analysis [57]. The examples given in this work are for soft X-ray parameters, and though there is no hard limit, the performance is expected to degrade as the number of modes increases and the XTCAV resolution decreases at shorter X-ray wavelengths.
On the other hand, seeded FELs have long coherence lengths, and an XTCAV likely would have sufficient resolution for application to hard X-ray self-seeding [58–61].

One advantage of the NN approach is the computational efficiency, which makes it well-suited to high repetition-rate XFELs. However, LCLS time-domain measurements currently rely on an XTCAV which runs at only 120 Hz, limited in rate both by the need for strong X-band streaking and the damage threshold of the electron YAG screen. Taking full advantage of the NN speed will require development of new diagnostics. For example, replacing the X-band cavity with a passive dechirper would avoid the need for a high-rate, high-voltage X-band source [62]. Alternatively, bending radiation can reveal microbunching imprinted on the beam by the FEL process [63], avoiding need for an electron screen. As described in the introduction, direct measurement of X-rays is possible for short pulses [12–16]. Nearly all of these methods could also benefit from including the diagnostics analysis itself as part of the machine learning pipeline; for example, the NN could take XTCAV image data directly as input and treat the XTCAV analysis as part of the forward model [19], or for attosecond pulses the NN could take as inputs the X-ray diagnostics [15,16]. Ultimately, verification of any XFEL phase prediction will require development of phase-sensitive experiments.

Appendix A. Additional examples of reconstructions

To give a sense of the quality of reconstructions associated with various levels of field errors, we show additional examples using both models. Figure 5 shows three additional examples using the $L_{\text{mae}}$ NN: Row 1 (field error = 0.11) and row 2 (error = 0.31) give the median and worst case values respectively for the 5% of shots with the lowest forward model error [i.e., the red points in Fig. 4(a)]. Row 2 is also equivalent to the 90th percentile worst case for all pulses. Row 3 (error = 0.6) is the 99th percentile worst case for all pulses.

Figure 6 shows three examples using the PINN: Row 1 (error = 0.14) is the median value for a model trained with perfect diagnostic resolution ($R_t = R_v = 0$). Row 2 (error = 0.3) is the median error for the top 5% of shots with the lowest forward model error in Fig. 4(a). Row 3 (error = 1.0) is the 90th percentile worst case for all pulses. The high-quality predictions for the model trained with perfect diagnostics [Table 1 and Fig. 6(b)] suggest that the phase errors in models with realistic diagnostic resolution [e.g., Fig. 6(f)] are driven by the resolution and not inherent phase ambiguity.

Appendix B. Neural network details

The models presented in the paper use the same inputs and outputs, as well as the same architecture, with the only differences in the loss function and regularization. Each input consists of the time-domain power ($n_p = 180$) concatenated with the spectral power ($n_s = 200$), for a total length of 380 points per example. The outputs consist of the real and imaginary components of field in the time domain, with a total 200 points ($p = 100$) for each example. To improve the spectral resolution we zero-pad the simulated time-domain field to make the time window 5-fold larger.

The networks contain four convolutional layers, each with 20 filters and a stride of two, and nine fully-connected layers, each with 200 neurons. All activations are rectified linear units (ReLU) except for the output layer, which has linear activation. The losses described in the main text (including the forward model) are implemented as custom loss functions written in PyTorch. See Table 2 for architecture details.

To reduce over-fitting and match typical experimental noise, both inputs include a Gaussian noise layer of 2% rms noise. The standard NN also uses dropout of 0.05 on the first fully-connected layer. The PINN did not benefit from additional regularization. The training consisted of 3000 epochs with a mini-batch size of 1000 and learning rate of $3 \times 10^{-4}$, followed by 3000 epochs with a mini-batch size of 5000 and learning rate of $1 \times 10^{-4}$. All training used the Adam optimizer with default PyTorch settings for beta of 0.9 and 0.999. We did only rough optimization of
hyper-parameters (regularization, network width and depth, mini-batch size), and improved performance should be possible from additional optimization. The PINN loss improved by 10% after doubling samples from 50k to 100k, suggesting additional gains are possible from a larger test set.

For the PINN, we apply additional constraints on the coherence length \((g_1)\) and smoothness of the solutions. The smoothness constraint penalizes the gradient of the prediction,

\[
L_s = \frac{1}{2np} \sum_{t} \sum_{\tau} \left| y'(t) - y'(t-1) \right|^2 ,
\]

where \(t\) is the time index and the sum is over all \(n\) examples in the minibatch. The total loss function for the PINN is then

\[
L_{\text{PINN}} = L_\Phi + \lambda_{g1} L_{g1} + \lambda_s L_s .
\]

The PINN results in the paper use \(\lambda_{g1} = 0.1\) and \(\lambda_s = 1\) for the first 3000 epochs, and either \(\lambda_s = 0.1\) for the final 3000 epochs or \(\lambda_s = 0.01\) (for the perfect resolution case). As with other hyperparameters, we expect performance gains are possible with careful optimization. For the combined model we use

\[
L_{\text{comb}} = (1 - \lambda_{\text{amp}}) L_\Phi + \lambda_{g1} L_{g1} + \lambda_s L_s + \lambda_{\text{amp}} L_{\text{amp}} .
\]

Using \(\lambda_{\text{mae}} = 0.5\), we find the combined model has smaller errors on amplitudes (1.0 compared to 1.3 amplitude error), but worse performance on phase (2.1 compared to 0.8 sine error). The combined model outperforms the PINN on both metrics, but loses the benefit of unlabeled training.

---

**Table 2. Description of the NN Architecture Used for all Models**

| Layer | Output Shape | Parameters |
|-------|--------------|------------|
| Conv1d | [-1, 20, 366] | 320        |
| Conv1d | [-1, 20, 176] | 6,020      |
| Conv1d | [-1, 20, 81] | 6,020      |
| Conv1d | [-1, 20, 34] | 6,020      |
| Flatten | [-1, 680] | 0          |
| Dense | [-1, 200] | 136,200    |
| Dropout | [-1, 200] | 0          |
| Dense | [-1, 200] | 40,200     |
| Dense | [-1, 200] | 40,200     |
| Dense | [-1, 200] | 40,200     |
| Dense | [-1, 200] | 40,200     |
| Dense | [-1, 200] | 40,200     |
| Dense | [-1, 200] | 40,200     |
| Dense | [-1, 200] | 40,200     |

\(^a\) As per convention, ‘-1’ in the middle column refers to batch size. In total, the model has 476,180 parameters.
Appendix C. Computation time comparison

The computation speed was calculated from the total time required for processing a batch of 10,000 examples on an Intel E5-2640 v3 CPU. The iterative approach required on average 20 sec/pulse, while the NN models required on average 20 µsec/pulse for inference plus 30 µsec/pulse data load time, nearly six orders of magnitude faster. Furthermore, a principal advantage of NNs is the ability to run efficiently on GPUs, and running on a GeForce RTX 2080 we found an average inference time of just 0.2 µsec/pulse. While we expect the iterative method could be sped-up through both algorithm and hardware improvements, we do not see a path to handling the full beam rate of a MHz FEL with iterative solvers.

Funding. U.S. Department of Energy (DE-AC02-76SF00515); Office of Science; Basic Energy Sciences.

Acknowledgments. We would like to thank Andre Al Haddad, T.J. Lane, and Gabriel Marcus for helpful discussions and assistance with simulations. This work was supported by the U.S. Department of Energy, under DOE Contract No. DE-AC02-76SF00515 and the Office of Science, Office of Basic Energy Sciences.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References
1. P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H.-D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky, and J. Galayda, “First lasing and operation of an angstrom-wavelength free-electron laser,” Nat. Photonics 4(9), 641–647 (2010).
2. D. Ratner, J. Cryan, T. Lane, S. Li, and G. Stupakov, “Pump-probe ghost imaging with sase fels,” Phys. Rev. X 9(1), 011045 (2019).
3. J. D. Biggs, Y. Zhang, D. Healion, and S. Mukamel, “Two-dimensional stimulated resonance raman spectroscopy of molecules with broadband x-ray pulses,” J. Chem. Phys. 136(17), 174117 (2012).
4. J. D. Biggs, Y. Zhang, D. Healion, and S. Mukamel, “Watching energy transfer in metalloporphyrin heterodimers using stimulated x-ray raman spectroscopy,” Proc. Natl. Acad. Sci. 110(39), 15597–15601 (2013).
5. S. Mukamel, D. Healion, Y. Zhang, and J. D. Biggs, “Multidimensional attosecond resonant x-ray spectroscopy of molecules: Lessons from the optical regime,” Annu. Rev. Phys. Chem. 64(1), 101–127 (2013).
6. I. V. Schweigert and S. Mukamel, “Coherent ultrafast core-hole correlation spectroscopy: X-ray analogues of multidimensional nmr,” Phys. Rev. Lett. 99(16), 163001 (2007).
7. C. Weninger, M. Purvis, D. Ryan, R. A. London, J. D. Bozek, C. Bostedt, A. Graf, G. Brown, J. J. Rocca, and N. Rohringer, “Stimulated electronic x-ray raman scattering,” Phys. Rev. Lett. 111(23), 233902 (2013).
8. J. T. O’Neal, E. G. Champenois, S. Oherli, R. Obaid, A. Al-Haddad, J. Barnard, N. Berrah, R. Coffee, J. Duris, G. Galinis, D. Garratt, J. M. Glownia, D. Haxton, P. Ho, S. Li, X. Li, J. MacArthur, J. P. Marangos, A. Natan, N. Shvaram, D. S. Slaughter, P. Walter, S. Wandel, L. Young, C. Bostedt, P. H. Bucksbaum, A. Picón, A. Marinielli, and J. P. Cryan, “Electronic population transfer via impulsive stimulated x-ray raman scattering with attosecond soft-x-ray pulses,” Phys. Rev. Lett. 125(7), 073203 (2020).
9. U. Eichmann, H. Röttke, S. Meise, J.-E. Rubensson, J. Söderström, M. Meyer, T. M. Baumann, R. Bolli, A. De Fanis, P. Gerychtol, M. Ilchen, T. Mazza, J. Montano, V. Music, Y. Ovcharenko, D. E. Rivas, S. Serkez, R. Wagner, and S. Eisebitt, “Photon-recoil imaging: Expanding the view of nonlinear x-ray physics,” Science 369(6511), 1630–1633 (2020).
10. I. V. Schweigert and S. Mukamel, “Probing interactions between core-electron transitions by ultrafast two-dimensional x-ray coherent correlation spectroscopy,” J. Chem. Phys. 128(18), 184307 (2008).
11. W. K. Peters, T. Jones, A. Elmov, E. Pedersoli, L. Foglia, R. Mincigrucci, I. Nikolov, R. Trebino, M. B. Danailov, F. Capotondi, F. Bencivenga, and P. Bowlan, “All-optical single-shot complete electric field measurement of extreme ultraviolet free electron laser,” Optica 8(4), 545–550 (2021).
12. U. Frühling, M. Wieland, M. Gensch, T. Gebert, B. Schütte, M. Kričkunova, R. Kalm, F. Budzyn, O. Grimm, J. Rossbach, E. Plöninger, and M. Drescher, “Single-shot terahertz-field-driven x-ray streak camera,” Nat. Photonics 3(9), 523–528 (2009).
13. I. Griguraš, A. R. Maier, C. Behrens, T. Mazza, T. J. Kelly, P. Radcliffe, S. Dürer, A. K. Kazansky, N. M. Kabachnik, T. Tsheitentscher, J. T. Costello, M. Meyer, M. C. Hoffmann, H. Schlärb, and A. L. Cavalieri, “Ultrafast x-ray pulse characterization at free-electron lasers,” Nat. Photonics 6(12), 852–857 (2012).
14. S. Li, E. G. Champenois, R. Coffee, Z. Guo, K. Hegazy, A. Kamalov, A. Natan, J. O’Neal, T. Osipov, M. Owens, D. Ray, D. Rich, P. Walter, A. Marinielli, and J. P. Cryan, “A co-axial velocity map imaging spectrometer for electrons,” AIP Adv. 8(11), 115308 (2018).
15. N. Hartmann, G. Hartmann, R. Heider, M. S. Wagner, M. Ilchen, J. Buck, A. O. Lindahl, C. Benko, J. Grüner, J. Krzywinski, J. Liu, A. A. Lutman, A. Marinelli, T. Maxwell, A. A. Miahnahri, S. P. Moeller, M. Planas, J. Robinson, A. K. Kazansky, N. M. Kabachnik, J. Viehwiss, T. Feurer, R. Kienberger, R. N. Coffee, and W. Helml, “Attosecond time–energy structure of x-ray free-electron laser pulses,” Nat. Photonics 12(4), 215–220 (2018).

16. S. Li, Z. Guo, R. N. Coffee, K. Hegazy, Z. Huang, A. Natan, T. Osipov, D. Ray, A. Marinelli, and J. P. Cryan, “Characterizing isolated attosecond pulses with angular streaking,” Opt. Express 26(4), 4531–4547 (2018).

17. Y. Ding, C. Behrens, P. Emma, J. Frisch, Z. Huang, H. Loos, P. Krejcik, and M. Wang, “Femtosecond x-ray pulse temporal characterization in free-electron lasers using a transverse deflector,” Phys. Rev. Accel. Beams 14(12), 120701 (2011).

18. C. Behrens, F.-J. Decker, Y. Ding, V. A. Dolgashev, J. Frisch, Z. Huang, P. Krejcik, H. Loos, A. Lutman, T. J. Maxwell, J. Turner, J. Wang, M.-H. Wang, J. Welch, and J. Wu, “Few-femtosecond time-resolved measurements of x-ray free-electron lasers,” Nat. Commun. 5(1), 3762 (2014).

19. X. Ren, A. Edelen, A. Lutman, G. Marcus, T. Maxwell, and D. Ratner, “Temporal power reconstruction for an x-ray free-electron laser using convolutional neural networks,” Phys. Rev. Accel. Beams 23(4), 040701 (2020).

20. P. Heimann, O. Krupin, W. F. Schlott, J. Turner, J. Krzywinski, F. Sorgenfrei, M. Messerschmidt, D. Bernstein, J. Chalupsýk, V. Hájíková, S. Hau-Riege, M. Holmes, L. Juha, N. Kelez, J. Lünig, D. Nordlund, M. Fernandez Perea, A. Scherz, R. Soufu, W. Wurth, and M. Rowen, “Linac coherent light source soft x-ray materials science instrument optical design and monochromator commissioning,” Rev. Sci. Instrum. 82(9), 093104 (2011).

21. D. Zhu, M. Cammarata, J. M. Feldkamp, D. M. Fritz, J. B. Hastings, S. Lee, H. T. Lemke, A. Robert, J. L. Turner, and Y. Feng, “A single-shot transmissive spectrometer for hard x-ray free electron lasers,” Appl. Phys. Lett. 101(3), 034103 (2012).

22. K. W. DeLong, R. Trebino, J. Hunter, and W. E. White, “Frequency-resolved optical gating with the use of second-harmonic generation,” J. Opt. Soc. Am. B 11(11), 2206–2215 (1994).

23. C. Iaconis and I. A. Walmsey, “Spectral phase interferometry for direct electric field reconstruction of ultrashort optical pulses,” Opt. Lett. 23(10), 792–794 (1998).

24. Y. Mairesse and F. Quéré, “Frequency-resolved optical gating for complete reconstruction of attosecond bursts,” Phys. Rev. A 71(4), 014101 (2005).

25. O. Riz, B. Leshem, J. Miao, B. Nadler, D. Oron, and N. Dudovich, “Direct phase retrieval in double blind fourier holography,” Opt. Express 22(21), 24935–24950 (2014).

26. P. Isidronenko, O. Lahav, Z. Avnur, and O. Cohen, “Ptychographic reconstruction algorithm for frequency-resolved optical gating: super-resolution and supreme robustness,” Optica 3(12), 1320–1330 (2016).

27. T. Gaumannz, A. Jain, and H. J. Wörner, “Complete reconstruction of ultra-broadband isolated attosecond pulses including partial averaging over the angular distribution,” Opt. Express 26(11), 14719–14740 (2018).

28. T. Schweizer, M. H. Brügmann, W. Helml, N. Hartmann, R. Coffee, and T. Feurer, “Attoclock ptychography,” Appl. Sci. 8(7), 1039 (2018).

29. F. Christie, Y. Ding, Z. Huang, V. A. Jhalani, J. Krzywinski, A. A. Lutman, T. J. Maxwell, D. Ratner, J. Rönsch-Schulenburg, and M. Vogt, “Temporal X-ray reconstruction using temporal and spectral measurements,” Sci. Rep. 10(1), 9799 (2020).

30. F. Christie, A. A. Lutman, Y. Ding, Z. Huang, V. A. Jhalani, J. Krzywinski, T. J. Maxwell, D. Ratner, J. Rönsch-Schulenburg, and M. Vogt, “Temporal x-ray reconstruction using temporal and spectral measurements at ICh,” J. Phys.: Conf. Ser. 1067, 032011 (2020).

31. J.-H. Chung and A. M. Weiner, “Ambiguity of ultrashort pulse shapes retrieved from the intensity autocorrelation and the power spectrum,” IEEE J. Sel. Top. Quantum Electron. 7(4), 656–666 (2001).

32. Y. D. Hezaveh, L. P. Levasseur, and P. J. Marshall, “Fast automated analysis of strong gravitational lenses with convolutional neural networks,” Nature 548, 555–557 (2017).

33. M. Cherukara, Y. Nashed, and R. Harder, “Real-time coherent diffraction inversion using deep generative networks,” Sci. Rep. 8(1), 16520 (2018).

34. M. J. Cherukara, T. Zhou, Y. Nashed, P. Enfedaque, A. Hexemer, R. J. Harder, and M. V. Holt, “Ai-enabled high-resolution scanning coherent diffraction imaging,” Appl. Phys. Lett. 117(4), 044103 (2020).

35. H. Chan, Y. S. G. Nashed, S. Kandel, S. Hruszkewycz, S. Sankaranarayanan, R. J. Harder, and M. J. Cherukara, “Real-time 3d nanoscale coherent imaging via physics-aware deep learning.” (2020).

36. Y. Zhang, M. A. Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, and P. Villanueva-Perez, “Phasegan: A deep-learning phase-retrieval approach for unpaired datasets,” (2020).

37. A. L. Edelen, J. P. Edelen, S. G. Biedron, S. V. Milton, and P. J. van der Slot, “Using neural network control policies for rapid switching between beam parameters in a free electron laser,” in Workshop on Deep Learning for Physical Sciences (DLPS 2017), NeurIPS 2017, (Long Beach, CA, USA, 2017).

38. M. Raisoi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” J. Comput. Phys. 378, 686–707 (2019).

39. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS 2014) p. 2672 (2014).

40. M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv:1411.1784 [cs.LG] (2014).
62. A. A. Lutman, T. J. Maxwell, J. P. MacArthur, M. W. Guetg, N. Berrah, R. N. Coffee, Y. Ding, Z. Huang, A. Marinelli, S. Moeller, and J. C. U. Zemella, “Fresh-slice multicolour x-ray free-electron lasers,” Nat. Photonics 10(11), 745–750 (2016).

63. E. L. Saldin, E. A. Schneidmiller, and M. V. Yurkov, “Optical afterburner for an x-ray free electron laser as a tool for pump-probe experiments,” Phys. Rev. Spec. Top.–Accel. Beams 13(3), 030701 (2010).