Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deep neural networks for high harmonic spectroscopy in solids

Nikolai Klimkin (nd.klimkin@physics.msu.ru)
Moscow State University https://orcid.org/0000-0002-6576-3602
Misha (Mikhail) Ivanov
Max-Born Institute https://orcid.org/0000-0002-8817-2469

Article

Keywords: Ultrafast Dynamic Decoding, Quantum Systems, Strong Laser Fields, Nonlinear Optical Response, Ultrashort Laser Pulses, Attosecond Spectroscopy

Posted Date: November 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-94390/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Deep neural networks for high harmonic spectroscopy in solids

Nikolai D Klimkin[1,2] and Misha Ivanov[3,4,5]

[1]*Russian Quantum Center, Bolshoy Bulvar 30, bld. 1, Moscow, Russia 121205*

[2]*Faculty of Physics, Lomonosov Moscow State University, Leninskie Gory 1-2, Moscow 119991*

[3]*Max-Born-Institute, Max-Born Straße 2A, D-12489 Berlin, Germany.*

[4]*Department of Physics, Humboldt University, Newtonstraße 15, D-12489 Berlin, Germany.*

[5]*Blackett Laboratory, Imperial College London, South Kensington Campus, SW7 2AZ London, United Kingdom.*

Neural networks are a prominent tool for identifying and modeling complex patterns, which are otherwise hard to detect and analyse. While machine learning and neural networks have been finding applications across many areas of science and technology, their use in decoding ultrafast dynamics of quantum systems driven by strong laser fields has been limited so far. Here we use deep neural networks to analyze spectra of highly nonlinear optical response of a crystal to intense few-cycle laser pulses. We construct a deep neural network that can efficiently utilize such spectra to resolve both the complex spectral phase of ultrashort laser pulses and simultaneously reconstruct the band structure of the crystal. Our results offer a new tool for attosecond spectroscopy of quantum dynamics and also open a route to developing all-solid-state devices for complete characterization of few-cycle pulses, including their nonlinear chirp and the carrier envelope phase.

Electrons provide the fundamental first step in response of matter to light. The feasibility of shaping light pulses at the scale of individual oscillations, from mid-IR to UV [1,2], offers rich opportunities for controlling electronic response to light on sub-cycle timescale (e.g. [3–13]), leading to a variety of fascinating phenomena such as optically induced anomalous Hall effect [14–16], topological phase transitions with polarization-tailored light [13], or the topological resonance [7]. Over multiple laser cycles, control of electron dynamics with light also enables the so-called Floquet engineering – the tailored modification of the cycle-average properties of a light-dressed system, see e.g. [17] for a recent review.

In this context, starting with the pioneering work [18], high harmonic spectroscopy has developed into a powerful tool for exploring laser-driven electron dynamics in solids, see e.g. recent reviews [19–21]. Examples include identification of the common physical mechanisms underlying high harmonic generation in atoms, molecules and solids (e.g. [10,22]), observation of Bloch oscillations [23], resolving interfering pathways of electrons in crystals with about 1-fsec precision [24], inducing [13] and monitoring topological [25–27] and Mott insulator-to-metal [28] phase transitions, resolving coherent oscillation of electronic charge density order [29], identifying the van Hove singularities in the conduction bands [30], and reconstructing effective potentials seen by the light-riven electrons with picometer accuracy [9].

Here we apply machine learning to the analysis of high harmonic generation from a crystal, which allows us to 'kill two birds with one stone': reconstruct the band structure of the crystal and fully characterize incident few-cycle laser pulses, including both their nonlinear chirp and the

phase of the carrier oscillations under the envelope (the carrier-envelope phase, CEP).

The fundamental role of the CEP in nonlinear light-matter interaction has been understood theoretically over two decades ago [31–34], stimulating first experiments in the gas phase [35]. Powerful gas-phase methods for characterizing few-cycle pulses have been developed, including stereo-above-threshold ionization (stereo-ATI) [36–38], attosecond streak camera and its modifications [39–43], and half-cycle high harmonic cutoffs [44]. Using nonlinear response of solids for characterizing the CEP has also been pursued [45–47]. Yet, all-optical, all-solid-state characterization of few-cycle pulses, including their CEP, remains a major challenge. We hope to change this situation. Particularly relevant to our work are the earlier proposal on using interference patterns in spectrally overlapping regions of even- and odd-order harmonics in solids [48] and the use of two-color high harmonic spectroscopy for all-optical reconstruction of the band structure [49] from the two-dimensional high harmonic spectra, recorded as a function of the harmonic frequency and the two-color delay.

Two-dimensional spectra of the nonlinear-optical response may provide sufficient information to recover the pulse. One prominent example is frequency-resolved optical gating, which uses the second-order optical response recorded as a function of the time-delay between the two incident pulses, the target pulses and the auxiliary gate pulse (e.g. [41,50,51].) Extending this analysis to highly nonlinear optical response remains an open problem. The crucial importance of addressing this problem stems from the fact that such analysis would allow one to characterize the laser pulse directly in the interaction region.

In special cases, such as the case of the two-color high harmonic spectroscopy of attosecond

pulses using fundamental and the second harmonic (e.g. [52]), the 2D harmonic spectra recorded as a function of the two-color phase and the harmonic frequency may carry sufficient information for reconstructing attosecond pulses produced by highly nonlinear optical response directly in the interaction region. Such reconstruction does, however, require detailed knowledge about the physics and dynamics of the microscopic quantum response. The possibility of solving a full reconstruction problem in a general case, recovering both the pulse and the quantum system, remains completely unexplored.

In the absence of simple and/or well known functional dependence between the response data and the parameters one wishes to reconstruct the problem is well suited for neural networks. Such networks aim to find a smooth analytical function $f_\theta(x)$ which connects the input x_i and the desired output y_i. If it is successful, one can conclude that the data x indeed does contain the information y. Pertinent examples include applications to solving the Schroedinger equation, where neural networks can output highly accurate results, see e.g. Ref. [53]. Our results show that, given sufficient training set, the 2D spectra of the high harmonic response as a function of the nonlinear response frequency and the CEP of an unknown driving pulse allow for simultaneous reconstruction of both the pulse and the unknown crystal band structure.

To demonstrate the method, we assume no apriori knowledge about the incident pulse and use rather limited knowledge about the nonlinear medium. In our simulations, we use one-dimensional lattice with two kinds of sites. The sites are characterized by their on-site energies and nearest neighbor, next-nearest neighbor, etc couplings. Both the on-site energies and the couplings are

assumed to be unknown for the analysis, and the reconstruction procedure is expected to output both the parameters of the pulse and of the lattice. We are thus faced with a nonlinear optimization problem with a very large input dimension which requires us to find the optimal interpolation between the existing trial samples.

Such a regression tool is readily provided by deep neural networks (DNNs) [54], already used for such diverse applications as boosting the signal-to-noise ratio in LHC collision data [55], establishing a fast mapping between galaxy and dark matter distribution [56], and constructing efficient representations of many-body quantum states [57]. The inherent resilience of neural networks to noise is an important asset for pulse shape characterization. The emergence of photonic implementations of feed-forward neural networks [58] outlines a perspective of implementing this regression scheme in an all-optical way.

For the quantum system, we use a modified Rice-Mele model [59] with nearest neighbor, next nearest neighbor, etc. hoppings, see the cartoon in Figure 1(a). The system is described by the Hamiltonian

$$\hat{H} = \left[\frac{1}{2} \sum_{j=0}^{N} t_j \sum_{\alpha \in \{A,B\}} s_\alpha \left| i\alpha \right\rangle \left\langle i+j, \alpha \right| + h_1 \sum_{i=-\infty}^{+\infty} \left| i, A \right\rangle \left\langle i+1, B \right| \right] + \text{h.c.} \tag{1}$$

Here $s_\alpha = -1$ for $\alpha = A$ and $s_\alpha = +1$ for $\alpha = B$, N is the maximum hopping order, which was varied from 2 to 6. The onsite energies are thus $-t_0$ for site A and $+t_0$ for site B, while the hopping terms $t_{j \geq 1}$ connect the sites of the same kind. In the simulations, t_j vary randomly between training samples and are unknown to the neural network. These hopping terms are selected in such a way that the band gap always lies between 8.0 and 16.0 eV. The higher hoppings $\{t_1, t_2, \ldots\}$ are ran-

domly generated within the ranges: $[-6.0, -1.0]$, $[-1.5, 1.5]$, $[-1.0, 1.0]$, $[-1.0, 1.0]$, $[-0.5, 0.5]$, respectively. In all our computations, we assume $\hbar = a = 1$, where a is the lattice constant.

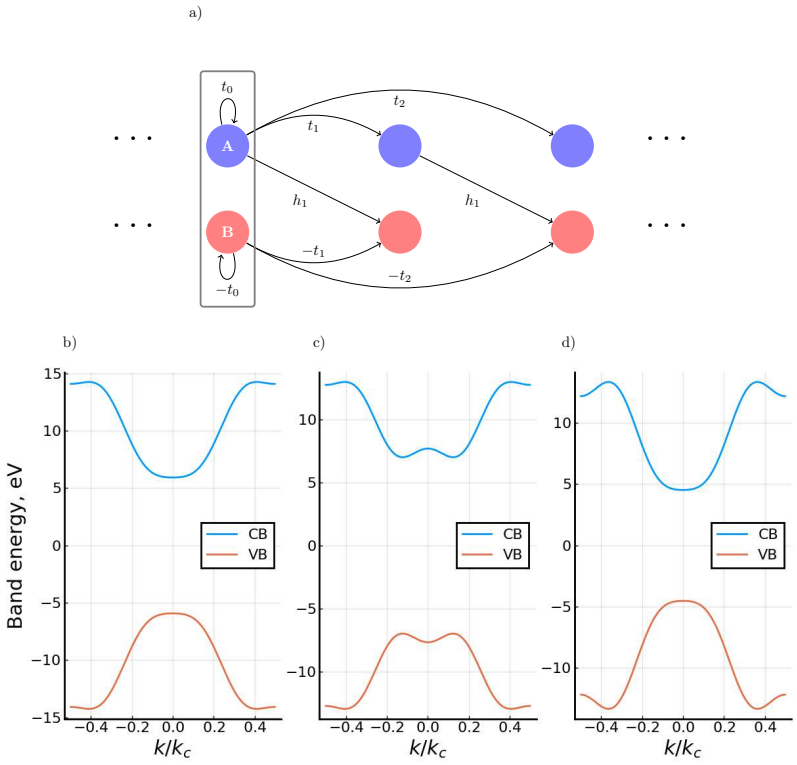


Figure 1: Model system used for pulse reconstruction and band structure spectroscopy via nonlinear-optical response. (a) The two types of sites present, A and B, are connected by hopping constants t_j between similar sites and h_1 between the next-neighbor sites of the different type. (b-d) Examples of band structures that can be generated with this model system.

The Hamiltonian also includes the term proportional to h_1, which describes hopping be-

tween sites A and B. This term explicitly breaks the inversion symmetry of the model, so that its nonlinear-optical response allows us to distinguish between pulses with CEPs differing by π. We set h_1 to be much smaller than the typical values of t_j; we used $h_1 = 0.01$ eV and $h_1 = 0.05$ eV. This simplifies modelling relaxation. This system allows one to generate a rich variety of band structures and band-gaps $\epsilon_g(k)$, see Fig. 1(b-d) for some examples.

When we set the lattice constant $a = 1, k_c = 2\pi$, in the k space, the field-free Hamiltonian transforms to:

$$\hat{H}(k) = \epsilon(k)\sigma_z + h_1 \left(\sigma_x \cos k + \sigma_y \sin k\right) \tag{2}$$

$$\epsilon(k) = \sum_{j=0}^{N} t_j \cos(jk) \tag{3}$$

The field is introduced via the Peierls substitution, which transforms the laser-driven Hamiltonian in the momentum space as:

$$\hat{H} = \int dk \, |k\rangle \, \hat{H}(k + A(t)) \, \langle k| \tag{4}$$

We evolve the system using the master equation with a decoherence term:

$$\dot{\rho} = i[\rho, H] - \rho_{od}/T_2 \tag{5}$$

Here ρ_{od} denotes the off-diagonal part of the density matrix in the Hamiltonian eigenstate basis, T_2 is the dephasing time, set to $T_0/2$, where T_0 is a single fundamental laser period.

The vector potential of the incident laser field, $A(t)$, is generated in the frequency domain with the unknown to the neural network quadratic and the cubic phases:

$$A(\omega) \propto \exp(i\phi(\omega)) = \exp\left(i\lambda(\omega - \omega_0)^3/6 - \mu(\omega - \omega_0)^2/2 + i\varphi\right) \tag{6}$$

The complex parameter μ is defined in such a way that, in the absence of the cubic chirp, the pulse has a temporal width $\sigma \equiv T_0$: $\mu \equiv \dfrac{\sigma^2 - i\alpha}{1 + \alpha^2/\sigma^4}$, where α is the chirp parameter. The chirp parameters α and λ are expressed via dimensionless quantities β and ϵ, $\alpha = \pi(\sigma/2)^2 \times \beta$, $\lambda = \pi\sigma^3 \times \epsilon$. The dimensionless parameters vary in the ranges $\beta \in [-2.0, 2.0]\,, \epsilon \in [-1.0, 1.0]$. Finally, φ sets the CEP of the pulse, also unknown to the neural network and chosen randomly. The system evolution is simulated for 40 laser cycles; an additional Gaussian cutoff is introduced at the edges of the leading and the rear tails of the pulse for numerical stability, see Methods.

The typical pulses we have used for reconstruction are shown in Fig. 2, both in time and frequency domain. The latter shows the spectral phases (red curves) alongside the spectral amplitudes (blue curves), as a function of ω/ω_0, where ω_0 is the carrier.

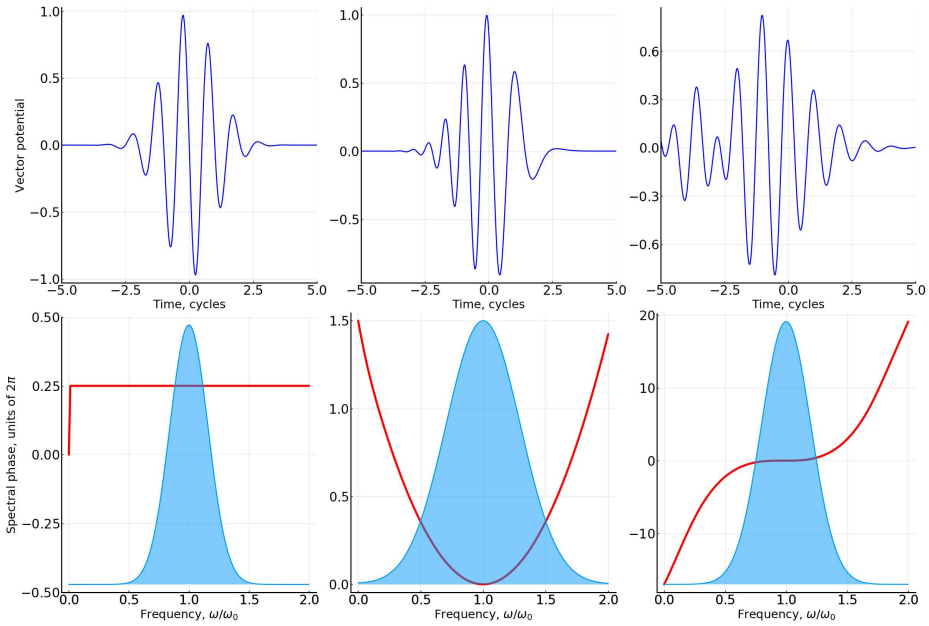


Figure 2: Sample pulse shapes used in the problem. The top row shows the time-domain, and the middle and bottom rows the frequency-domain representations. In the bottom row, the spectral phase (red line) and spectral amplitude normalized to the central frequency amplitude (blue area), are plotted with respect to frequency. The pulse parameters are $(\varphi, \beta, \epsilon)$, left to right: $(\pi/2, 0.0, 0.0), (0.0, 2.0, 0.0), (0.0, 1.0, 1.0)$.

Our typical simulated "measurement" assumes that one can systematically vary the (unknown) CEP. Thus, we perform calculations by varying $\varphi + \Delta\varphi$, with $\Delta\varphi$ spanning the full range $\Delta\varphi \in [0, 2\pi)$. For each $\Delta\varphi$, we measure the absolute value of the spectral amplitude of the laser-induced current $|j(\omega)|$, which is given by the Fourier transform of the calculated current $j(t)$. The input data is composed of the absolute values of the integer harmonic amplitudes $|j(N\omega, \varphi + \Delta\varphi)|$. The resulting 2D map as a function of $\Delta\varphi$ and ω is used as the input into the neural network. The network must then infer the (randomly chosen in each trial) intraband hoppings $\{t_j\}$, the unknown

initial CEP φ, and the pulse parameters α and λ. The values of σ, frequency ω, and h_1 are kept constant throughout.

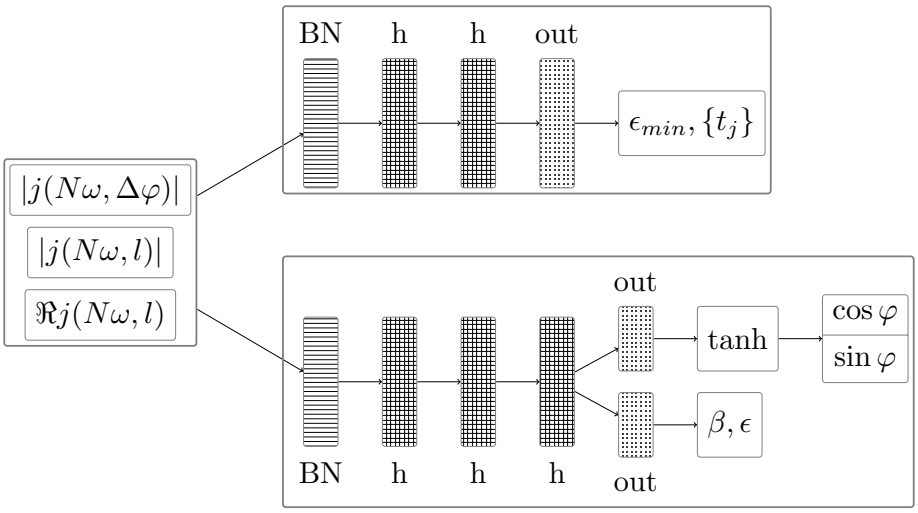
The phase $\Delta\varphi$ is discretized into 64 steps. Per each laser cycle, 80 time points are recorded, allowing us to recover integer harmonics up to $N = 40$, thus a single 2D plot has 41x64 points.

As the neural network depends on rotation-invariant features to infer the band and the non-CEP pulse parameters, the raw input data is supplemented by its own Fourier transform in the φ direction. The 2D spectra are therefore first transformed, then the absolute value and the real part of their Fourier amplitudes are concatenated with the initial image and input into the network.

We have generated data sets that consist of 65536 samples for pulses with no cubic phase, and 262144 samples for pulses with cubic phase. Next, 4096 samples have been set apart to form the test set. The input data consisted of the basic plots $|j(N\omega, \Delta\varphi)|$, concatenated with the absolute values of their Fourier amplitudes along the φ axis, and the real values of the same amplitudes (see Results). The required output were the band parameters, the CEP in the form of $(\cos\varphi, \sin\varphi)$, and the dimensionless chirp and cubic phase.

We have also expanded the training set without computing additional training samples. The 2D plots $|j(N\omega, \Delta\varphi)|$ have been transformed by a circular shift along the $\Delta\varphi$ axis, after which they, complete with their Fourier transforms, have been added to the original training set with a shifted $\Delta\varphi$. With the training sets with no cubic phase, we have augmented the data by 15 additional angles (i.e. expanded the training set by a factor of 16), for those with the cubic phase

by 2 angles. The augmented data have been separated into batches of 512 samples, with the batches being reshuffled each epoch. The test data have not been augmented.



[1]

Figure 3: Networks used in the paper. The output layers for the CEP and the pulse parameters are separated since the activation function for the CEP neurons is set to tanh. All other activation functions are identity. The initial (striped) layer is the batch normalization, intended to force the NN to identify nonlinear features between the data. The intermediate (checkered) layers are the hidden layers. The final (dotted) is the output layer. $j(N\omega, l)$ designates the Fourier transform of $|j(N\omega, \Delta\varphi)|$ along the $\Delta\varphi$ axis.

In the reconstruction procedure, we have used two separate neural networks, one for recovering the band parameters and another one for the pulse parameters. Both were constructed as feed-forward neural networks with different layouts, see Fig. 3. The division of responsibilities between the two networks allowed us to optimize solutions of both tasks independently. A very brief tutorial description of the key concepts behind neural networks is given in the supplementary section. Both networks used one layer (striped in Fig. 3) to normalize the data, which emphasizes nonlinear dependencies. The normalized data where then fed into several hidden layers (shown in blue), ending with the output layer (yellow). We have found empirically that two hidden layers (with the number of neurons per layer $(n_1, n_2) = (800, 400)$) were sufficient for band reconstruction in our case, while three hidden layers ($(n_1, n_2, n_3) = (800, 400, 400)$) were needed to reconstruct the pulse. In all hidden layers, we used the Swish activation function [60]. They were trained with the ADAM optimizer [61] for 40 epochs with learning rate 10^{-3}, then 20 epochs with 10^{-4}, then 10 epochs with $5 \cdot 10^{-5}$. The code was written in the Julia language [62] using the Flux.jl [63,64] library for implementing neural networks and is available on GitHub [1].

The typical recovery results are demonstrated in Fig. 4. Here, the hopping order is limited to 4. The agreement between the actual and the reconstructed data, both for the system and for the pulse, is excellent. Tables with detailed analysis of the average performance are given in the supplementary material.

[1] `https://github.com/KlimkinND/PulseReconstruction`

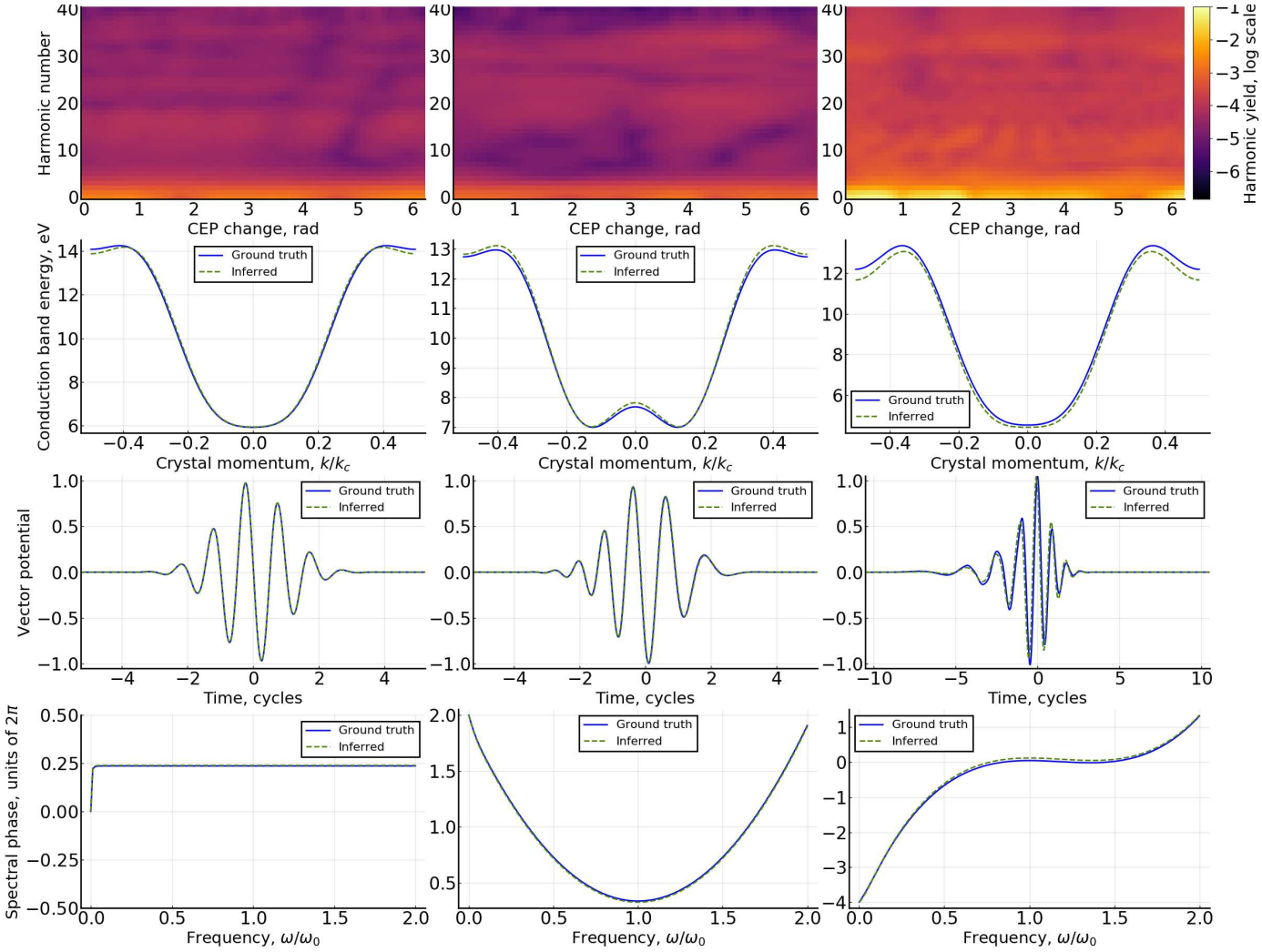


Figure 4: (color online) Bands and pulses recovered by the neural network. Left column shows the case of no chirp, the central column shows the case when only quadratic phase is used for input, the right column shows cases with the cubic phase. 1st row: raw HHG spectra which are input into the NN. 2nd row: comparison of the true and reconstructed corresponding crystal bands. Comparison of the true and reconstructed pulses in time (3rd row) and frequency (4th row) domains. The blue solid curves depict correct results, while the dashed green curves show the reconstructed values.

Our results demonstrate that solid-state HHG spectra contain more information than one extracts with conventional methods: not only the pulse parameters, but also the parameters of the quantum system are robustly reconstructed. Our approach replicates the advantages of gas-phase HHG spectroscopy, namely, the ability to resolve the CEP of laser pulses and the complex pulse shapes with polynomial spectral phase nonlinearities. At the same time, it requires neither the XUV pulses nor the photoelectron spectroscopy, such as the stereo-ATI. Moreover, it allows for all-optical solid-state implementation. In terms of required observables, it is closest to the method based on measuring the half-cycle cutoffs in gas-phase high harmonic generation [44]. However, it also allows one to deal with very strong chirps. Applying neural networks to the analysis of half-cycle cutoffs and high harmonic generation spectra in the gas phase could be very interesting, especially in molecules, where multiple coupled harmonic generation channels present challenges for unravelling the underlying laser-driven multi-electron dynamics [65]

One could apply the developed neural network design to standard gas-phase experiments, to analyse the possibility of resolving the spectral phase and the CEP for short pulses by processing HHG spectra generated by known inert gases (Ar, Ne, etc.). In this case the neural network can be trained using TDSE simulations of the necessary responses before being applied to real experimental data.

The key difficulty of using high harmonic spectroscopy in solids is that, without apriori knowledge of the band structure, one lacks closed-form solutions for electron dynamics, similar to those available in the gas-phase. Our method circumvents this difficulty. Pulse characterization

device implementing our principles could be tabletop, all solid-state, and capable of operating at ambient conditions.

Another interesting direction to pursue would be to apply novel physics-informed neural network architectures[66,67] to resolve Hamiltonians of systems with many degrees of freedom (such as molecules) using time-resolved HHG spectra, such as those obtained from solids driven by mid-IR fields [24]. Neural networks can also be used for processing the sets of harmonic spectra connected by other relations, such as being measured for different angles between the crystal axes and the driving field, to uncover effective laser-modified potentials for the charge motion, extending the pioneering work in Ref. [9] to recover effective potentials of active band electrons. Here, once again, one can take advantage of our idea of using neural networks to extend a method of processing analytically-tractable systems to intractable ones, recovering effective structures.

Methods

Driving pulses in time domain. In time domain, the pulse shape corresponding to the frequency-domain representation in the main text can be written in closed analytical form:

$$A(t) = \frac{F_0}{\omega_0} \operatorname{Im}\left(\mathcal{E}(t) \exp(i(\omega_0 t + \varphi))\right) \tag{7}$$

The envelope function in the above equation is:

$$\mathcal{E}(t) = 2\sqrt{\frac{\pi|\mu|}{2}} \left|\frac{2}{\lambda}\right|^{1/3} \operatorname{Ai}\left(\left(\frac{2}{\lambda}\right)^{1/3} \left(t + \frac{\mu^2}{2\lambda}\right)\right) \exp\left(\frac{\mu}{\lambda}\left(\frac{\mu^2}{3\lambda} + t\right)\right) \tag{8}$$

The normalizing factor is introduced to keep the peak absolute value of the envelope function at 1 when there's no cubic phase, and conserve the total pulse power at all parameter values. For numerical stability, we introduce a cutoff function that ensures that $A(t) = 0$ at the beginning and end of the pulse:

$$\text{cutf}(t) = \begin{cases} 1.0, & \text{if } t > 2T \\ \exp\left(-\dfrac{(t-2T)^2}{2(T/2)^2}\right), & \text{if } t \leq 2T \end{cases} \tag{9}$$

with the numerical vector potential taking the form

$$A_{num}(t) = A(t) * \text{cutf}(t) * \text{cutf}(T_{max} - t) \tag{10}$$

Evaluation of the Airy function. We have approximated the Airy function required for computing the pulse shape in (8) with Maclaurin series [68, Eq. 9.4.1] and asymptotic series [68, Eq. 9.7.5, 9.7.9] for small and large arguments, respectively. For $|z| < 3.5$ we used the Maclaurin series up to $N = 25$; for $|z| \geq 3.5$, the asymptotic expansion for $N = 5$ was used. This approximation allows for a maximum relative error below $\delta = 3 \cdot 10^{-4}$.

Solution of the Time dependent Schroedinger Equation. We discretize the 1D Brillouin zone of the crystal into 32 k points (finer discretization does not worsen the NN's performance). For each point, we initialize the system in the pure ground state of the Hamiltonian (2) for $A(t_0) = 0$, then evolve this state according to (5). The dephasing time $\tau_2 = T_0/2$ is selected phenomenologically. For all computations, the field $F_0 = 4.0$ eV, and the central frequency $\omega_0 = 0.8$ eV are kept constant.

80 times per cycle, we record the current, computed as:

$$\hat{\mathcal{J}}(k) = \frac{\partial \hat{H}(k)}{\partial k} = -\sum_j j t_j \sin(jk) + h_1(-\sigma_x \sin(k) + \sigma_y \cos(k)) \tag{11}$$

$$\langle j \rangle_k (t) = \text{tr}(\rho_k(t)\mathcal{J}(k + A(t))) \tag{12}$$

$$\langle j \rangle (t) = \sum_k \langle j \rangle_k (t) \tag{13}$$

In the above equation, we neglect the $2\pi/N_k$ ($N_k = 32$) factor.

Each "recording" interval is separated into 2 integration steps. Each integration step consists of $\nu = 4$ iterations of the following procedure:

$$\rho \to \exp(-iH(k + A(t))dt/\nu)\rho \exp(iH(k + A(t))dt/\nu) \tag{14}$$

$$\rho \to \rho - (dt/\nu) * \rho_{od}/T_2 \tag{15}$$

Within each integration step, the field $A(t)$ is kept constant. The matrix exponent is computed numerically exactly in the assumption that $H(k)$ is traceless.

Supplementary information

Neural networks Here we first give a very brief general introduction into neural networks, and then discuss applications to our particular problem.

Suppose we search for a model that best describes a known experimental data set $(\mathbf{x}_i, \mathbf{y}_i)$ with K total samples. Generally, the workflow is:

1. Define the model as a trial function $f_\theta(\mathbf{x})$ that depends on a vector of parameters θ.

2. Define a measure of divergence between the results predicted by the model $\{f_\theta(\mathbf{x}_i)\}$ and experimental results $\{\mathbf{y}_i\}$, expressed as $D = \frac{1}{K}\sum_{i=1}^{K} \mathcal{L}(f_\theta(\mathbf{x}_i), \mathbf{y}_i)$. The most common measure for D is the mean squared error (MSE), obtained for $\mathcal{L} = ||f_\theta(\mathbf{x}_i) - \mathbf{y}_i||^2$.

3. Optimize the parameters θ to minimize the divergence D.

The most basic example of such an approach is the linear regression: we define the model as $f(\mathbf{x}) = \hat{W}\mathbf{x} + \mathbf{b}$ with $\theta \equiv \{\hat{W}, \mathbf{b}\}$, and use the MSE divergence. Minimizing the divergence with respect to $\hat{W}$ and $\mathbf{b}$, we obtain what's known as the linear least squares method. For this case, we have two options available. First, we could find the optimal θ analytically by and solving the overdefined system of linear equations [2]. However, this may not be optimal for high-dimensional data, and so we can choose an alternative way where we optimize the same parameters with gradient descent, updating θ against the gradient of the divergence function: $\delta\theta \propto -\nabla_\theta \sum_i \mathcal{L}(f_\theta(\mathbf{x}_i), \mathbf{y}_i)$. Obviously, the result must be the same, which can be verified analytically.

Evidently, linear regression cannot model most data sets. The simple nonlinear regression methods generally rely on fitting the experimental data to an ansatz selected by hand. Such methods include logistical regression, polynomial regression, and the more exotic expansions over various bases such as the Legendre polynomials and harmonic functions. However, selecting the correct

[2] https://en.wikipedia.org/wiki/Ordinary_least_squares#Matrix/vector_formulation

nonlinear regression requires extensive trial-and-error testing of various ansätze, which is generally not possible except for simple cases with known symmetries and properties of the experimental data set. A convenient solution to this complication is using neural networks, which serves as a universal ansatz by being capable of approximating any continuous finite function to an arbitrary precision [69]. This property is the key advantage of neural networks, which allows them to be successfully applied even to data sets where the very existence of a connection between the input and the output is not evident.

We will now discuss the core principles of a neural network. A neural network is constructed of building blocks called neurons. The neuron, demonstrated schematically on Fig. S1 is defined as follows.

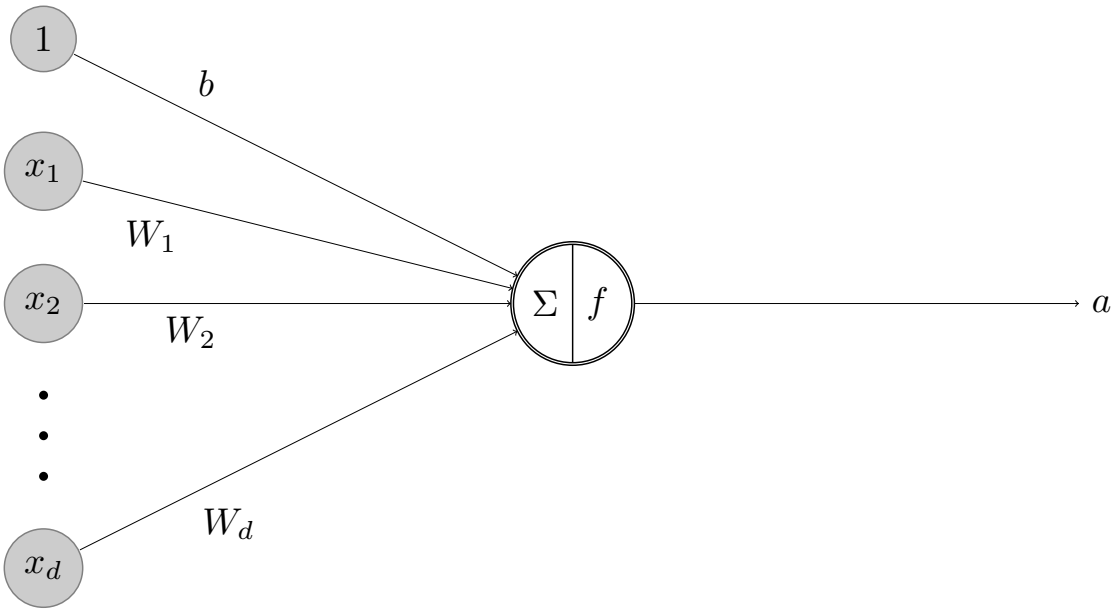


Figure S1: The vector $\mathbf{x}$ is the d-dimensional input vector, $\mathbf{W}$ is the d-dimensional weight vector, b is the bias. The output is computed as $a = f(\mathbf{W}\mathbf{x} + b)$, commonly rewritten as $a = f(\mathbf{W}^{[1]}\mathbf{x}^{[1]})$ with $W_0^{[1]} = b$, $x_0^{[1]} = 1$.

The neuron is a function with n-dimensional input $\mathbf{x}$ and one-dimensional output $a \equiv f(\mathbf{W}\mathbf{x} + b)$. Here $\mathbf{W}$ is the weight vector of the neuron, b is the bias, and f is the activation

function. The typically chosen activation functions include the logistical function, ReLU, Soft-
plus, Swish, etc. [3] A neuron encodes a single nonlinearity, such as a smooth step for the logistical
($\equiv$ sigmoid) activation function. Neural networks use compositions of these nonlinearities to ap-
proximate arbitrary smooth functions.

In a neural network, the neurons are arranged in layers where each neuron has the same input vector $\mathbf{x}$ and the same activation function f. For this layer, the bias $\mathbf{b}$ is expressed by a vector, the weights $\hat{W}$ by a matrix, and the output is given by $\mathbf{a} = f(\hat{W}\mathbf{x}+\mathbf{b})$, where f is applied individually to each argument. The simplest neural network, called the single-layer perceptron, consists of only two such layers, and is schematically demonstrated on Fig. S2. The first (hidden) layer contains all the nonlinearities, and gets the raw data as the input. The second (output) layer may or may not be linear, i.e. have an identity activation function, and in the linear case simply applies a learnable linear transform to the outputs of the hidden layer.

[3] `https://en.wikipedia.org/wiki/Activation_function`

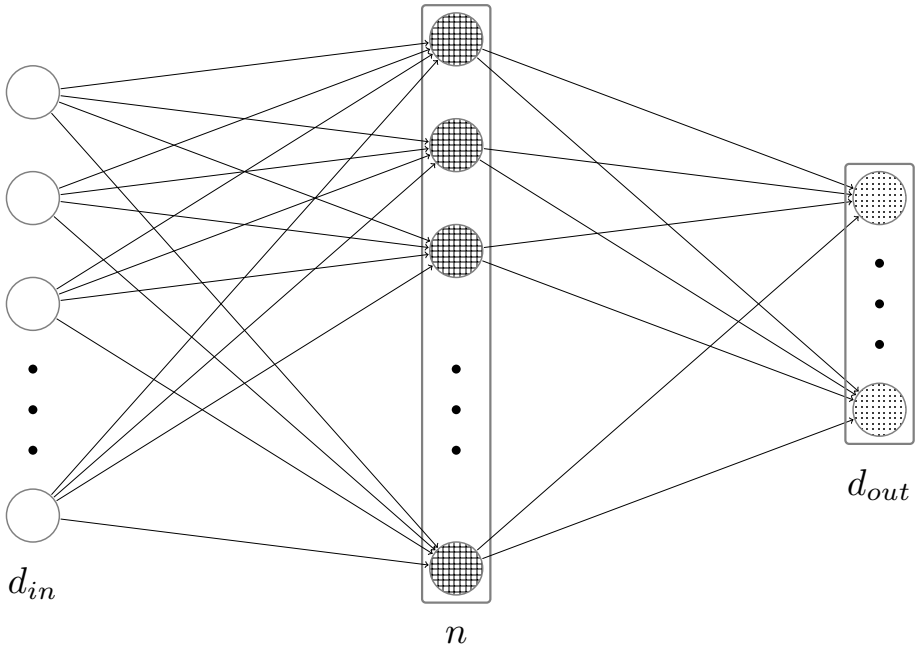


Figure S2: Neuron layout of a single-layer perceptron. The empty circles represent elements of the input data, the checkered circles the hidden layer neurons, and the dotted circles the output neurons.

The universal approximation theorem [69] states that any continuous finite function can be approximated by a perceptron with a sufficiently large hidden layer with a non-polynomial acti-

vation function. However, for many cases this turns out to be computationally non-optimal. A more efficient representation that requires less parameters can be achieved by using stacked layers, with the output of k-th layer being the input for $k+1$-th. For deep networks, the equation $\mathbf{a} = f(\hat{W}\mathbf{x} + \mathbf{b})$ is commonly rewritten as $\mathbf{a}^{[k]} = f^{[k]}(\hat{W}^{[k]}\mathbf{x}^{[k]})$, where k is the layer number, $W_{i0}^{[k]} = b_i$, $x_0^{[k]} = 1$. The output function is therefore given by: $f_\theta(\mathbf{x}) = f^{[N]}(\hat{W}^{[N]}\mathbf{x}^{[N]}) = f^{[N]}(\hat{W}^{[N]}f^{[N-1]}(\hat{W}^{[N-1]}f^{[N-2]}(\ldots)))$.

The general heuristic states that the higher layers process higher-level features, and thus can learn highly nonlinear and nonlocal patterns. A 'dual' version of the universal approximation theorem [70,71] demonstrates that an arbitrary continuous function can also be approximated by a neural network with a fixed (but not arbitrarily small) number of neurons per layer and sufficiently large number of layers. Shown on figure S3 is the general scheme of a deep neural network.

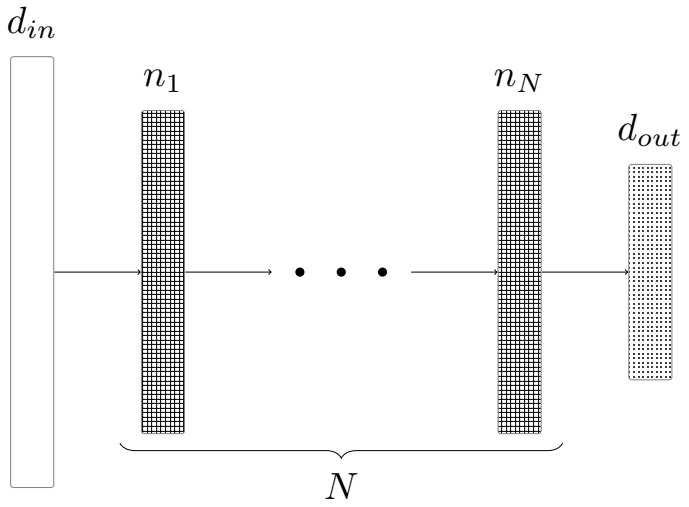


Figure S3: In a deep feed-forward neural network, the output of each layer is used as the input of the next one, with the network output being evaluated at the last layer.

Deep neural networks can be trained in an optimal way thanks to a procedure called 'back-propagation' [72] that allows to efficiently differentiate the loss function with respect to each of its parameters $\theta \equiv \{\hat{W}^{[N]} \dots \hat{W}^{[1]}\}$. This enables optimizing them with the gradient descent procedure, which consists of updating θ against the gradient of the divergence function, called 'loss' for neural networks: $\delta\theta \propto -\nabla_\theta \sum_i \mathcal{L}(f_\theta(\mathbf{x}_i), \mathbf{y}_i)$, ideally yielding a function that minimizes the

error for each sample simultaneously. However, the 'naive' gradient descent is vulnerable to noise and local minima, and in practice, more sophisticated methods are employed, such as ADAM [61]. ADAM is based on correcting the gradient descent update by using an adapted momentum of the parameters which prevents them from getting stuck in a local minimum. For the same goal, the input data is separated into 'batches' that consist of a defined number of samples. The parameters are updated after processing each batch. The process of optimizing the FFNN parameters, or 'training' the FFNN, therefore consists of the following steps.

1. Select a batch of data as input.
2. (forward pass) Evaluate the output of each consecutive layer to obtain the NN output.
3. Compute the loss function between the produced output and the desired output.
4. (backward pass) Calculate the gradient of the loss function with respect to each parameter by reverse propagating the loss gradient from the last to the first layer.
5. Compute the update to the parameters using the parameter update procedure (gradient descent, ADAM, ADAGrad, ADAMax, etc) and apply it.

The above steps are repeated for the entire data set. Processing the entire data set according to the training procedure is called a 'training epoch'. A neural network can be trained for any number of epochs, ranging from tens to thousands.

Therefore, deep neural networks can be used to model complex dependencies in a compu-

tationally optimal way, often requiring several orders of magnitude less computational time than exact methods. However, their application is not limited to regression problems.

Average performance of the selected network The average performance of the selected neural network for different unknown parameters is given below. For parameters that do not change sign, the average relative error is also indicated.

CEP + 4 parameters:

	$\delta\epsilon_{min}$	δt_1	δt_2	δt_3	$\delta\varphi$
abs	0.031	0.044	0.052	0.038	0.031
rel	0.005	0.021			

CEP+chirp+4 parameters:

	$\delta\epsilon_{min}$	δt_1	δt_2	δt_3	$\delta\varphi$	$\delta\beta$
abs	0.073	0.110	0.140	0.097	0.093	0.067
rel	0.013	0.049				

CEP + chirp + cubic phase + 4 parameters:

	$\delta\epsilon_{min}$	δt_1	δt_2	δt_3	$\delta\varphi$	$\delta\beta$	$\delta\epsilon$
abs	0.099	0.115	0.153	0.121	0.375	0.148	0.060
rel	0.022	0.042					

Figure S4: Performance of the 2 neural networks for different unknown parameter sets

CEP + chirp + cubic phase + 2 parameters:

	$\delta\epsilon_{min}$	δt_1	$\delta\varphi$	$\delta\beta$	$\delta\epsilon$
abs	0.041	0.033	0.189	0.064	0.027
rel	0.010	0.019			

CEP + chirp + cubic phase + 4 parameters:

	$\delta\epsilon_{min}$	δt_1	δt_2	δt_3	$\delta\varphi$	$\delta\beta$	$\delta\epsilon$
abs	0.099	0.115	0.153	0.121	0.375	0.148	0.060
rel	0.022	0.042					

Figure S5: Performance of the NNs for different maximum hopping orders

Acknowledgements We thank Vera V. Tiunova for her useful feedback. M.I. and N.K. acknowledge funding of the DFG QUTIF program. N.K. acknowledges funding by the UMNIK foundation.

Data availability All data supporting the figures is available upon request. The codes used are available publicly on GitHub [4].

Competing interests The authors declare no competing interests.

Authors contribution N.K. and M.I. developed the idea behind the paper. N.K. performed the calculations. Both authors wrote the manuscript.

[4] https://github.com/KlimkinND/PulseReconstruction

References

1. Wirth, A. *et al.* Synthesized light transients. *Science* **334**, 195–200 (2011). URL https://science.sciencemag.org/content/334/6053/195. https://science.sciencemag.org/content/334/6053/195.full.pdf.

2. Hassan, M. T. *et al.* Optical attosecond pulses and tracking the nonlinear response of bound electrons. *Nature* **530**, 66–70 (2016). URL https://doi.org/10.1038/nature16528.

3. Schultze, M. *et al.* Controlling dielectrics with the electric field of light. *Nature* **493**, 75–78 (2013).

4. Schiffrin, A. *et al.* Optical-field-induced current in dielectrics. *Nature* **493**, 70–74 (2013).

5. Kelardeh, H. K., Apalkov, V. & Stockman, M. I. Attosecond strong-field interferometry in graphene: Chirality, singularity, and berry phase. *Physical Review B* **93**, 155434 (2016).

6. Kelardeh, H. K., Apalkov, V. & Stockman, M. I. Graphene in ultrafast and superstrong laser fields. *Physical Review B* **91**, 045439 (2015).

7. Motlagh, S. A. O., Nematollahi, F., Apalkov, V. & Stockman, M. I. Topological resonance and single-optical-cycle valley polarization in gapped graphene. *Physical Review B* **100**, 115431 (2019).

8. Garg, M. *et al.* Multi-petahertz electronic metrology. *Nature* **538**, 359–363 (2016).

9. Lakhotia, H. *et al.* Laser picoscopy of valence electrons in solids. *Nature* **583**, 55–59 (2020).

10. Vampa, G. *et al.* Linking high harmonics from gases and solids. *Nature* **522**, 462–464 (2015).

11. Reimann, J. *et al.* Subcycle observation of lightwave-driven dirac currents in a topological surface band. *Nature* **562**, 396–400 (2018).

12. Langer, F. *et al.* Lightwave valleytronics in a monolayer of tungsten diselenide. *Nature* **557**, 76–80 (2018).

13. Jimenez-Galan, A., Silva, R. E. F., Smirnova, O. & Ivanov, M. Lightwave topology for strong-field valleytronics (2019). 1910.07398.

14. Motlagh, S. A. O., Apalkov, V. & Stockman, M. I. Anomalous ultrafast all-optical hall effect in gapped graphene (2020). 2007.12757.

15. Sato, S. *et al.* Light-induced anomalous hall effect in massless dirac fermion systems and topological insulators with dissipation. *New Journal of Physics* **21**, 093005 (2019).

16. McIver, J. W. *et al.* Light-induced anomalous hall effect in graphene. *Nature physics* **16**, 38–41 (2020).

17. Oka, T. & Kitamura, S. Floquet engineering of quantum materials. *Annual Review of Condensed Matter Physics* **10**, 387–408 (2019).

18. Ghimire, S. *et al.* Observation of high-order harmonic generation in a bulk crystal. *Nature physics* **7**, 138–141 (2011).

19. Vampa, G. & Brabec, T. Merge of high harmonic generation from gases and solids and its implications for attosecond science. *Journal of Physics B: Atomic, Molecular and Optical Physics* **50**, 083001 (2017).

20. Kruchinin, S. Y., Krausz, F. & Yakovlev, V. S. Colloquium: Strong-field phenomena in periodic systems. *Reviews of Modern Physics* **90**, 021002 (2018).

21. Ghimire, S. & Reis, D. A. High-harmonic generation from solids. *Nature Physics* **15**, 10–16 (2019).

22. Vampa, G., McDonald, C., Orlando, G., Corkum, P. & Brabec, T. Semiclassical analysis of high harmonic generation in bulk crystals. *Physical Review B* **91**, 064302 (2015).

23. Schubert, O. *et al.* Sub-cycle control of terahertz high-harmonic generation by dynamical bloch oscillations. *Nature Photonics* **8**, 119–123 (2014).

24. Hohenleutner, M. *et al.* Real-time observation of interfering crystal electrons in high-harmonic generation. *Nature* **523**, 572–575 (2015).

25. Silva, R., Jiménez-Galán, Á., Amorim, B., Smirnova, O. & Ivanov, M. Topological strong-field physics on sub-laser-cycle timescale. *Nature Photonics* **13**, 849–854 (2019).

26. Chacón, A. *et al.* Circular dichroism in high-order harmonic generation: Heralding topological phases and transitions in chern insulators (2018). 1807.01616.

27. Bauer, D. & Hansen, K. K. High-harmonic generation in solids with and without topological edge states. *Physical Review Letters* **120**, 177401 (2018).

384 28. Silva, R., Blinov, I. V., Rubtsov, A. N., Smirnova, O. & Ivanov, M. High-harmonic spec-
385 troscopy of ultrafast many-body dynamics in strongly correlated systems. *Nature Photonics*
386 **12**, 266–270 (2018).

387 29. Nag, T., Slager, R.-J., Higuchi, T. & Oka, T. Dynamical synchronization transition in interact-
388 ing electron systems. *Physical Review B* **100**, 134301 (2019).

389 30. Uzan, A. J. *et al.* Attosecond spectral singularities in solid-state high-harmonic generation.
390 *Nature Photonics* **14**, 183–187 (2020).

391 31. de Bohan, A., Antoine, P., Milošević, D. B. & Piraux, B. Phase-dependent harmonic emission
392 with ultrashort laser pulses. *Physical review letters* **81**, 1837 (1998).

393 32. Cormier, E. & Lambropoulos, P. Effect of the initial phase of the field in ionization by ultra-
394 short laser pulses. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma*
395 *Physics* **2**, 15–20 (1998).

396 33. Tempea, G., Geissler, M. & Brabec, T. Phase sensitivity of high-order harmonic generation
397 with few-cycle laser pulses. *JOSA B* **16**, 669–673 (1999).

398 34. Dietrich, P., Krausz, F. & Corkum, P. Determining the absolute carrier phase of a few-cycle
399 laser pulse. *Optics letters* **25**, 16–18 (2000).

400 35. Paulus, G. *et al.* Absolute-phase phenomena in photoionization with few-cycle laser pulses.
401 *Nature* **414**, 182–184 (2001).

402 36. Zhang, Y. *et al.* Single-shot, real-time carrier-envelope phase measurement and tag-
403 ging based on stereographic above-threshold ionization at short-wave infrared wavelengths.
404 *Opt. Lett.* **42**, 5150–5153 (2017). URL `http://ol.osa.org/abstract.cfm?URI=`
405 `ol-42-24-5150`.

406 37. Paulus, G. G. *et al.* Measurement of the phase of few-cycle laser pulses. *Physical review*
407 *letters* **91**, 253004 (2003).

408 38. Milošević, D., Paulus, G., Bauer, D. & Becker, W. Above-threshold ionization by few-cycle
409 pulses. *Journal of Physics B: Atomic, Molecular and Optical Physics* **39**, R203 (2006).

410 39. Goulielmakis, E. *et al.* Direct measurement of light waves. *Science (New York, N.Y.)* **305**,
411 1267–9 (2004).

412 40. Itatani, J. *et al.* Attosecond streak camera. *Physical review letters* **88**, 173903 (2002).

413 41. Mairesse, Y. & Quéré, F. Frequency-resolved optical gating for complete reconstruction of
414 attosecond bursts. *Physical Review A* **71**, 011401 (2005).

415 42. Baltuška, A. *et al.* Attosecond control of electronic processes by intense light fields. *Nature*
416 **421**, 611–615 (2003).

417 43. Kienberger, R. *et al.* Atomic transient recorder. *Nature* **427**, 817–821 (2004).

418 44. Haworth, C. *et al.* Half-cycle cutoffs in harmonic spectra and robust carrier-envelope phase
419 retrieval. *Nature Physics* **3** (2007).

45. Dombi, P. *et al.* Direct measurement and analysis of the carrier-envelope phase in light pulses approaching the single-cycle regime. *New Journal of Physics* **6**, 39–39 (2004). URL https://doi.org/10.1088%2F1367-2630%2F6%2F1%2F039.

46. Apolonski, A. *et al.* Observation of light-phase-sensitive photoemission from a metal. *Physical review letters* **92**, 073902 (2004).

47. Paasch-Colberg, T. *et al.* Solid-state light-phase detector. *Nature Photonics* **8** (2013).

48. Mehendale, M., Mitchell, S., Likforman, J.-P., Villeneuve, D. & Corkum, P. Method for single-shot measurement of the carrier envelope phase of a few-cycle laser pulse. *Optics letters* **25**, 1672–1674 (2000).

49. Vampa, G. *et al.* All-optical reconstruction of crystal band structure. *Phys. Rev. Lett.* **115**, 193603 (2015). URL https://link.aps.org/doi/10.1103/PhysRevLett.115.193603.

50. Trebino, R. & Kane, D. J. Using phase retrieval to measure the intensity and phase of ultrashort pulses: frequency-resolved optical gating. *J. Opt. Soc. Am. A* **10**, 1101–1111 (1993). URL http://josaa.osa.org/abstract.cfm?URI=josaa-10-5-1101.

51. Miranda, M. *et al.* Characterization of broadband few-cycle laser pulses with the d-scan technique. *Opt. Express* **20**, 18732–18743 (2012). URL http://www.opticsexpress.org/abstract.cfm?URI=oe-20-17-18732.

52. Dudovich, N. *et al.* Measuring and controlling the birth of attosecond xuv pulses. *Nature physics* **2**, 781–786 (2006).

440 53. Mills, K., Spanner, M. & Tamblyn, I. Deep learning and the schrödinger equation. *Physical*
441 *Review A* **96**, 042113 (2017).

442 54. Carleo, G. *et al.* Machine learning and the physical sciences. *Rev. Mod. Phys.* **91**, 045002
443 (2019). URL https://link.aps.org/doi/10.1103/RevModPhys.91.045002.

444 55. Baldi, P., Sadowski, P. & Whiteson, D. Searching for exotic particles in high-energy physics
445 with deep learning. *Nature Communications* **5**, 4308 (2014). URL https://doi.org/
446 10.1038/ncomms5308.

447 56. Zhang, X. *et al.* From dark matter to galaxies with convolutional networks (2019). 1902.
448 05965.

449 57. Sharir, O., Levine, Y., Wies, N., Carleo, G. & Shashua, A. Deep autoregressive models for the
450 efficient variational simulation of many-body quantum systems. *Physical Review Letters* **124**
451 (2020). URL http://dx.doi.org/10.1103/PhysRevLett.124.020503.

452 58. Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nature Photonics* **11**,
453 441–446 (2017). URL https://doi.org/10.1038/nphoton.2017.93.

454 59. Rice, M. & Mele, E. Elementary excitations of a linearly conjugated diatomic polymer. *Phys-*
455 *ical Review Letters* **49**, 1455 (1982).

456 60. Ramachandran, P., Zoph, B. & Le, Q. V. Searching for activation functions. *arXiv preprint*
457 *arXiv:1710.05941* (2017).

61. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

62. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A fresh approach to numerical computing. *SIAM review* **59**, 65–98 (2017). URL https://doi.org/10.1137/141000671.

63. Innes, M. *et al.* Fashionable modelling with flux. *CoRR* **abs/1811.01457** (2018). URL https://arxiv.org/abs/1811.01457. 1811.01457.

64. Innes, M. Flux: Elegant machine learning with julia. *Journal of Open Source Software* (2018).

65. Bruner, B. D. *et al.* Multidimensional high harmonic spectroscopy of polyatomic molecules: detecting sub-cycle laser-driven hole dynamics upon ionization in strong mid-ir laser fields. *Faraday discussions* **194**, 369–405 (2016).

66. Greydanus, S., Dzamba, M. & Yosinski, J. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, 15379–15389 (2019).

67. Toth, P. *et al.* Hamiltonian generative networks. *arXiv preprint arXiv:1909.13789* (2019).

68. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.27 of 2020-06-15. URL http://dlmf.nist.gov/. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

69. Pinkus, A. Approximation theory of the mlp model in neural networks. *Acta numerica* **8**, 143–195 (1999).

70. Lu, Z., Pu, H., Wang, F., Hu, Z. & Wang, L. The expressive power of neural networks: A view from the width. In *Advances in neural information processing systems*, 6231–6239 (2017).

71. Kidger, P. & Lyons, T. Universal approximation with deep narrow networks. In *Conference on Learning Theory*, 2306–2327 (2020).

72. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *nature* **323**, 533–536 (1986).

Figures

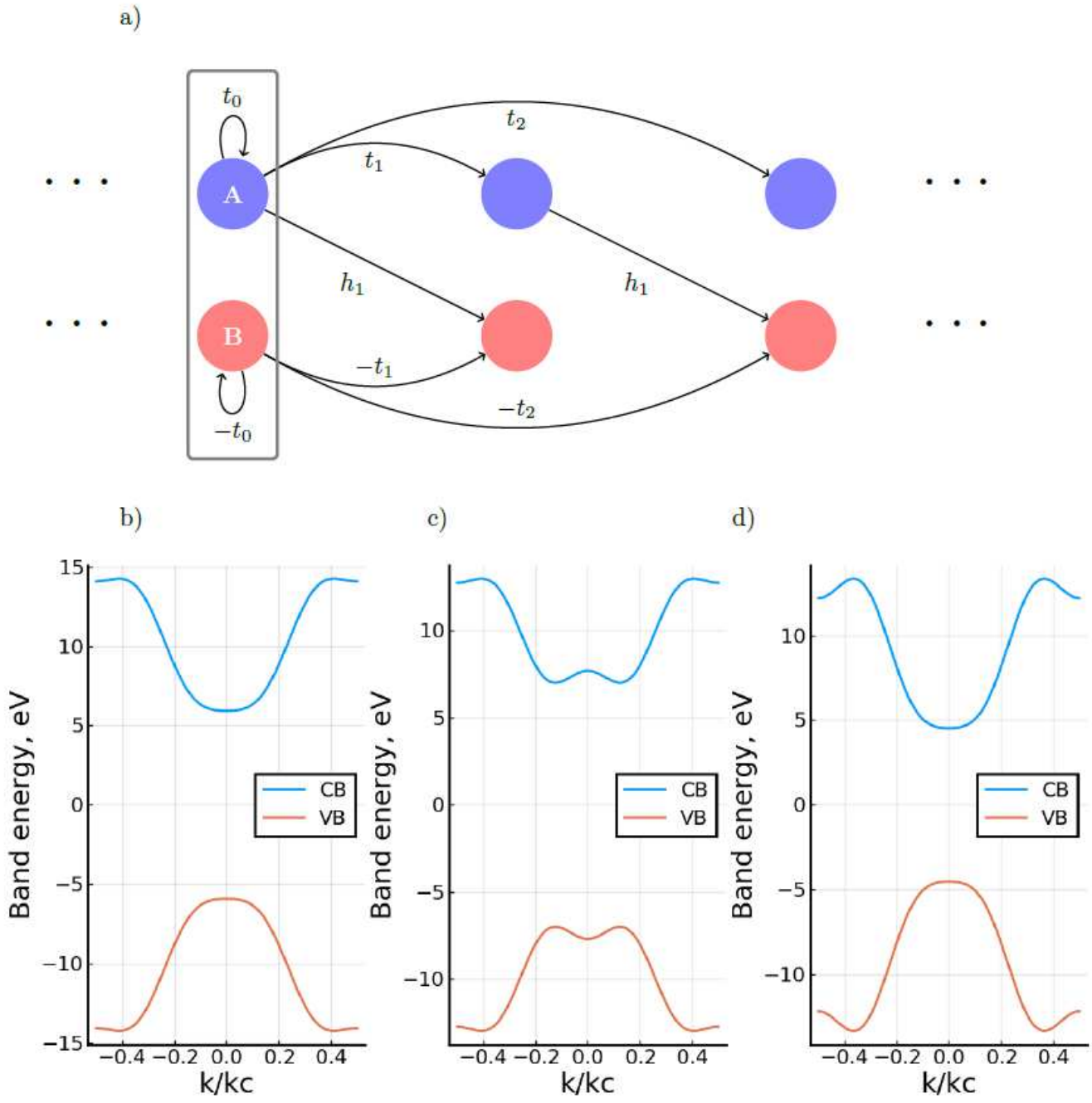


Figure 1

Model system used for pulse reconstruction and band structure spectroscopy via nonlinear-optical response. (a) The two types of sites present, A and B, are connected by hopping constants tj between similar sites and h1 between the next-neighbor sites of the different type. (b-d) Examples of band structures that can be generated with this model system.

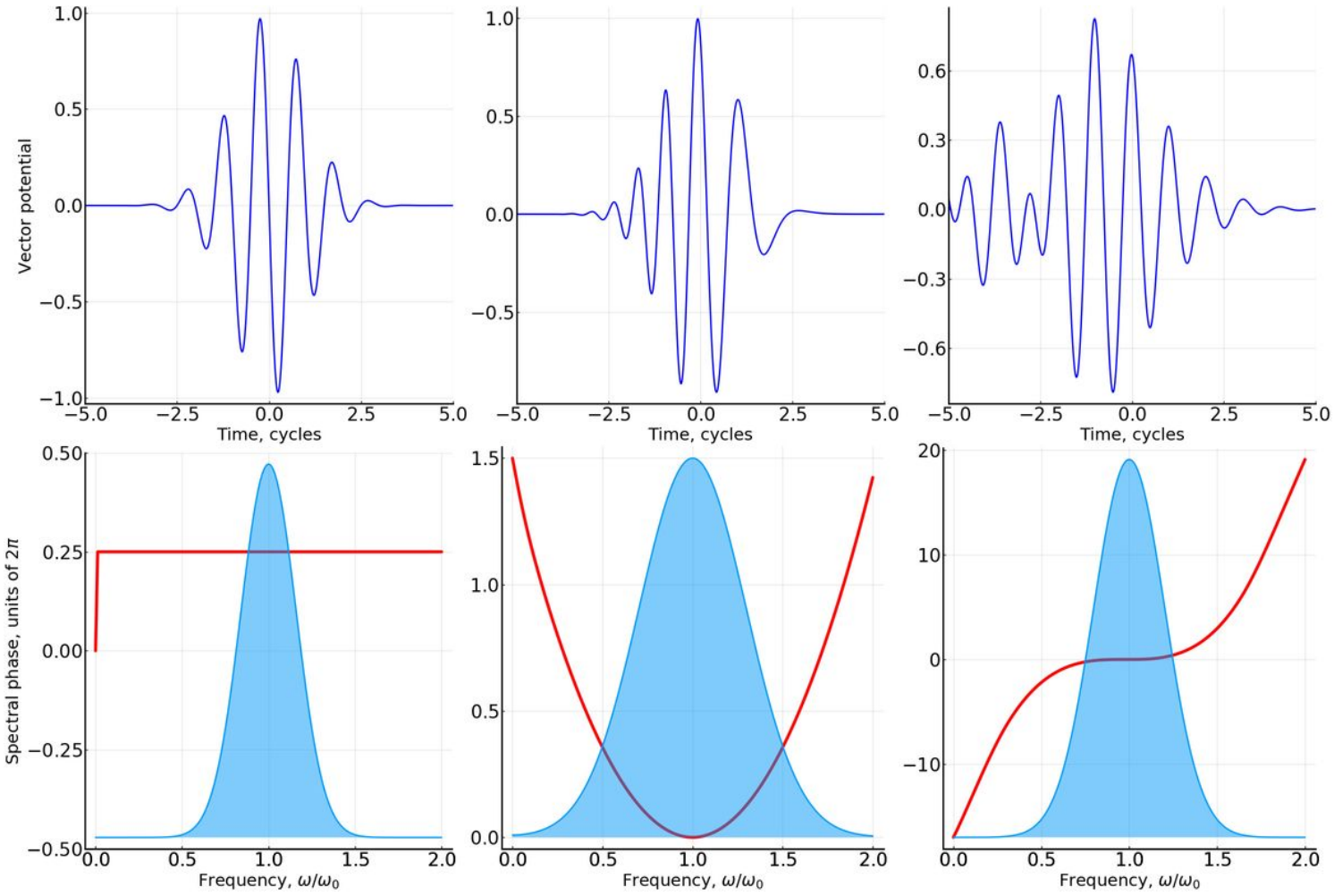


Figure 2

Sample pulse shapes used in the problem. The top row shows the time-domain, and the middle and bottom rows the frequency-domain representations. In the bottom row, the spectral phase (red line) and spectral amplitude normalized to the central frequency amplitude (blue area), are plotted with respect to frequency. The pulse parameters are (φ, β, ε), left to right: (π=2; 0:0; 0:0); (0:0; 2:0; 0:0); (0:0; 1:0; 1:0).

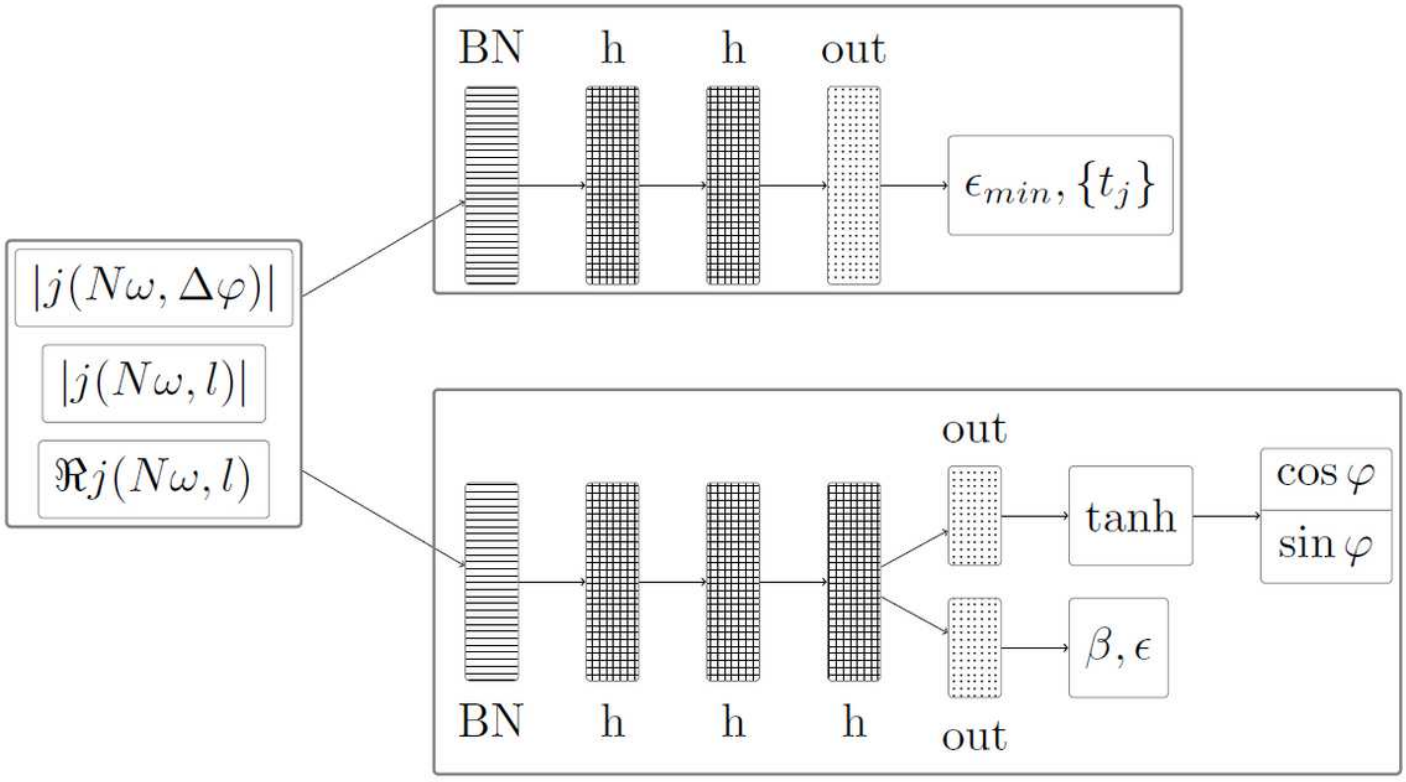


Figure 3

Networks used in the paper. The output layers for the CEP and the pulse parameters are separated since the activation function for the CEP neurons is set to tanh. All other activation functions are identity. The initial (striped) layer is the batch normalization, intended to force the NN to identify nonlinear features between the data. The intermediate (checkered) layers are the hidden layers. The final (dotted) is the output layer. j(Nω,l) designates the Fourier transform of ljNω,Δφl along the Δφ axis.

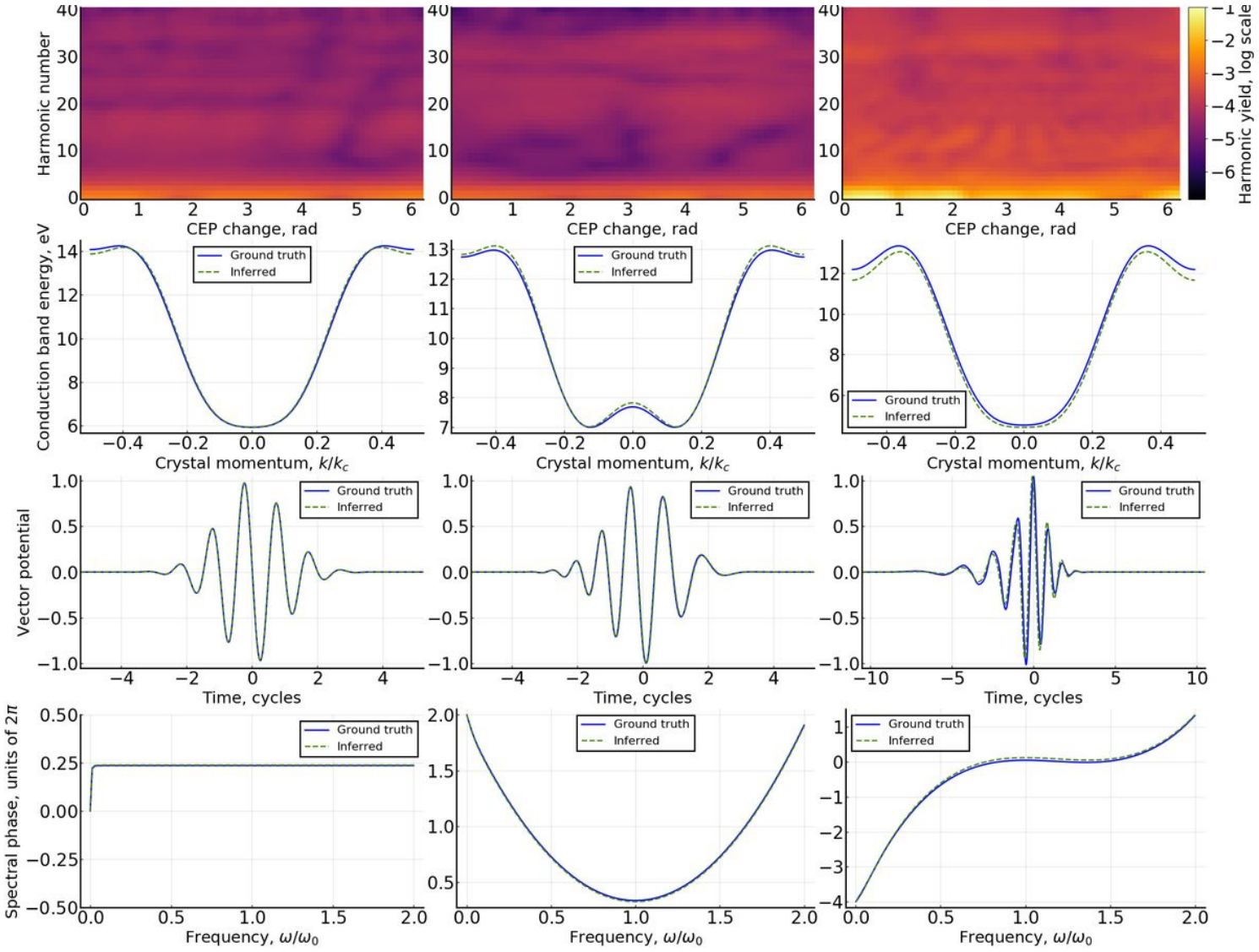


Figure 4

Bands and pulses recovered by the neural network. Left column shows the case of no chirp, the central column shows the case when only quadratic phase is used for input, the right column shows cases with the cubic phase. 1st row: raw HHG spectra which are input into the NN. 2nd row: comparison of the true and reconstructed corresponding crystal bands. Comparison of the true and reconstructed pulses in time (3rd row) and frequency (4th row) domains. The blue solid curves depict correct results, while the dashed green curves show the reconstructed values.