Faraday Discussions

Cite this: *Faraday Discuss.*, 2021, **228**, 502

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Perspectives for analyzing non-linear photo-ionization spectra with deep neural networks trained with synthetic Hamilton matrices

Sajal Kumar Giri,* Lazaro Alonso,* Ulf Saalmann * and Jan Michael Rost *

Received 11th October 2020, Accepted 15th December 2020
DOI: 10.1039/d0fd00117a

We have constructed deep neural networks, which can map fluctuating photo-electron spectra obtained from noisy pulses to spectra from noise-free pulses. The network is trained on spectra from noisy pulses in combination with random Hamilton matrices, representing systems which could exist but do not necessarily exist. In [Giri *et al.*, *Phys. Rev. Lett.*, 2020, **124**, 113201] we performed a purification of fluctuating spectra, that is, mapping them to those from Fourier-limited Gaussian pulses. Here, we investigate the performance of such neural-network-based maps for predicting spectra of double pulses, pulses with a chirp and even partially-coherent pulses from fluctuating spectra generated by noisy pulses. Secondly, we demonstrate that along with purification of a fluctuating double-pulse spectrum, one can estimate the time-delay of the underlying double pulse, an attractive feature for single-shot spectra from SASE FELs. We demonstrate our approach with resonant two-photon ionization, a non-linear process, sensitive to details of the laser pulse.

1 Introduction

Machine learning (ML) has recently been applied not only in physics,[1–3] but more specifically in strong-field physics.[4–6] One of the most abundant topics has been the reconstruction of the temporal shape of an ultrashort laser pulse, aided by ML techniques.[7–9] The most popular techniques for this reconstruction have been different variants of streaking techniques which normally require considerable additional experimental effort, namely a Terahertz laser light source. With its help, one can generate a large amount of data—the streaking traces—which can be processed with ML to extract the attosecond pulse shape.[7,8] However, also a direct method from single-shot spectra has been introduced.[9]

Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, Dresden, 01187, Germany. E-mail: sajal@pks.mpg.de; lalonso@pks.mpg.de; us@pks.mpg.de; rost@pks.mpg.de

In a different vein, a trained neural network has been proposed to represent a (semi-)classical path integral for strong-field physics,[10] replacing the need to explicitly calculate a large number of classical trajectories to eventually determine the photo-ionization cross section, which is, however, still an approximation as it is constructed semi-classically. To supply training data for a network which can represent the full quantum path integral, implies most likely use of a numerical effort that would be higher than calculating observables directly.

In general, training of a deep neural network needs a very large amount of non-trivial training data. To generate them experimentally requires substantial additional effort (see the streaking example above). To obtain such data without serious approximations within theory is often prohibitively expensive, as in the second example.

Acknowledging this situation, we have invented another approach: to calculate exactly and explicitly (with the time-dependent Schrödinger equation) photo-electron spectra with a large number of pulses and artificial systems, for which the calculation can be done very quickly. In this way we are able to supply learning data consisting of about 10^7 spectra. A network trained with these synthetic systems, is not only able to purify noisy test spectra unknown to the network, but from the same class of synthetic systems the training was performed with. Also "real" spectra can be purified, which could come from experiment, or for this work, from a realistic full calculation with parameters for the helium atom. Moreover, noise is, in the context of machine learning, helpful when applied to non-linear photo-ionization: photo-excitation and ionization processes are subject to strict angular-momentum selection rules, thereby limiting the coupling of light to matter. If a light pulse contains noise and operates in a non-linear (at least two-photon absorption) regime it will couple to a much larger part of the electron dynamics of the target. This helps to train the mapping better and enlarges the pool of training spectra naturally.

In general, all trained networks we will present, map one type of spectrum into another (desired) one for a photo-ionization scenario of which only a few key elements need to be specified: the target system should have an excited state around the photon energy ω_*, above the ground state, and intensities of the light pulse should be such that two-photon processes dominate. It is not necessary to know more about the target system, as ideally all target systems accessible by the light as specified are covered by the learning space for the synthetic systems, represented by synthetic Hamilton matrices (SHMs). Therefore, one can apply a trained network also to an experimental spectrum from noisy pulses without detailed knowledge of the target system.

Once the design for training such networks with SHMs is set up, that is, the spectra for learning have been computed, it is not difficult to construct other maps with new networks, as the major effort is to supply the learning data which do not have to be changed, while training new networks is computationally relatively cheap. This allows us to provide several mappings in the following to predict spectra for ideal double, chirped and even highly structured partially coherent pulses from noisy spectra. Finally, we will introduce network based mapping for a typical SASE FEL situation: there, single-shot noisy spectra are recorded which depend on further, not explicitly known parameters, *e.g.*, the geometrical orientation of the sample or the time-delay of double pulses used. Considering the latter situation, we reconstruct from noisy spectra

simultaneously the noise-free spectra and the time-delay of the double pulse. While we cannot do this with the accuracy of the designated algorithms as described in the context of streaking above, we do not need any additional information but the spectrum itself.

The paper is organized as follows: in Section 2 we give details on the representation of the noisy pulses, explain how to construct the SHMs and describe our fast propagation scheme to solve the electronic Schrödinger equation to obtain the photo-ionization spectra. Section 3 details how the network is trained and set up, including measures on how to quantify errors in the reconstruction of spectra and a convenient way to parameterize them. In Section 4 we present the predictions of the photo-ionization spectra for various pulse forms. Section 5 discusses the single-shot FEL scenario. The paper ends with conclusions in Section 6.

2 Prerequisites

To determine the photo-ionization dynamics we need two elements, the noisy pulses and an efficient way to describe the electron dynamics. In the end we will specify the process we are interested in, namely two-photon absorption.

2.1 Pulses

We distinguish between the "noisy pulses" which lead to fluctuating spectra and the "reference pulses" for which we want to predict spectra.

There are many different possibilities for incorporating noise into a signal. We choose the partial-coherence method.[11,12] With this method one can create noisy pulses whose average over an ensemble has a well-defined pulse shape. As experimentally demonstrated,[12] these kinds of pulses represent pulses from SASE FELs well. In the following, we will use the pulse parameterisation

$$f(t) = NG_T(t)F_\tau(t), \tag{1a}$$

$$G_T(t) = \mathrm{e}^{-2\ln 2 t^2/T^2}, \quad F_\tau(t) = \mathscr{F}^{-1}\left[\mathrm{e}^{i\phi(\omega)}\mathscr{F}\left[\mathrm{e}^{-t^2/\tau^2}\cos(\omega_* t)\right]\right], \tag{1b}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ are the Fourier transform and its inverse, and ω_* is the carrier frequency. Noise is introduced through random spectral phases ϕ, uniformly distributed in the interval $-\pi \le \phi \le +\pi$. The time scale of the fluctuations is given by the coherence time τ, while the Gaussian $G_T(t)$ limits the typical pulse duration to T. Otherwise, the pulse duration could grow beyond all limits due to the presence of random spectral phases. A specific (deterministic) noise realization we will label with $\phi_l(\omega)$. If not stated otherwise, we use $T = 3$ fs and $\tau = 0.5$ fs in the following. In order to deal with comparable pulses, we use the normalisation constant N to fix the pulse energy E_p, which would otherwise fluctuate from realisation to realisation.

Any reasonable pulse can serve as a reference pulse, for which the map created by the network can predict the spectrum. *Reasonable* means in the present context that the reference pulse's frequency spectrum is covered by the learning space of fluctuating spectra. The simplest choice is the Gaussian $G_T(t)$ in eqn (1) itself rendering the prediction equivalent to removing the fluctuations from the spectrum. Therefore, we call this type of map "purification".[6] In Section 5 we will purify fluctuating spectra from double pulses.

2.2 Paradigmatic 1-dimensional strong-field electron dynamics

Although the subsequent scheme to construct SHMs is general, for the sake of clarity we will describe it for the processes we will consider as an example, namely two-photon absorption in a helium atom. Thereby, the carrier frequency ω_* of the laser is chosen to be quasi-resonant with the transition energy to the first optically allowed excited state.

A simple and convenient way to realize this concept is to consider 1D dynamics with a soft-core potential. The corresponding active one-electron Hamiltonian for helium is given by

$$H_0 = \frac{\hat{p}^2}{2} + V(x) = \frac{1}{2}\left(i\frac{\mathrm{d}}{\mathrm{d}x}\right)^2 - \frac{1}{\sqrt{x^2+a^2}}, \tag{2}$$

with the soft-core parameter $a = 1/\sqrt{2}$ which gives a ground-state energy $E_0 = -24.2$ eV, close to the ionization potential of real helium (24.6 eV). We represent the Hamiltonian on a grid $x_j = j\Delta x$, with $\Delta x = 0.067$ a.u. and $x_{\max} = 500$ a.u., and determine by diagonalization the eigenenergies $H_0|\alpha\rangle = |\alpha\rangle\tilde{E}_\alpha$, from the ground state up to $\tilde{E}_\alpha \leq E_{\max} \approx 48$ eV, resulting in 600 eigenstates.

With these eigenstates we calculate the matrix of the time-dependent Hamiltonian $H(t) = H_0 + A(t)\hat{p}$ in velocity gauge

$$\tilde{H}_{\alpha\beta}(t) = \tilde{E}_\alpha\delta_{\alpha\beta} + A(t)\tilde{V}_{\alpha\beta} \quad \text{with} \quad \tilde{V}_{\alpha\beta} \equiv \left\langle \alpha \middle| i\frac{\mathrm{d}}{\mathrm{d}x}\middle| \beta \right\rangle \tag{3}$$

with the vector potential $A(t) = A\,f(t)$, A being the field amplitude.

2.3 Synthetic Hamilton matrices (SHMs)

Since we want to train our network such that it recognizes almost arbitrary systems, which only need to have a (quasi-)resonant transition energy for the first absorbed photon, we create SHMs by randomly changing energies E_α, and matrix elements $V_{\alpha\beta}$, about the 1D example system defined in eqn (2) and (3) through the variation of four parameters in

$$E_\alpha = 3^{[\xi_1-\gamma]}\tilde{E}_\alpha \text{ for } \tilde{E}_\alpha < 0,\ \alpha > 0, \tag{4a}$$

$$V_{0\alpha} = 3^{\xi_2}\tilde{V}_{0\alpha} \text{ for } \tilde{E}_\alpha < 0, \tag{4b}$$

$$V_{\alpha\beta} = 3^{\xi_3}\tilde{V}_{\alpha\beta} \text{ for } \tilde{E}_\alpha < 0,\ \tilde{E}_\beta > 0, \tag{4c}$$

$$V_{\alpha\beta} = 3^{\xi_4}\tilde{V}_{\alpha\beta} \text{ for } \tilde{E}\alpha > 0, \tilde{E}_\beta > 0. \tag{4d}$$

Here, $\xi_{i=1\ldots4} = [-1, +1]$ are four uniform random numbers which lead to a large variety of artificial systems with different bound-state energies eqn (4a), and couplings between ground and bound states eqn (4b), as well as between bound and free states eqn (4c), and among free states eqn (4d). Finally, with the parameter γ the condition of resonant first-photon absorption can be met. In the present case the energy difference between ground and the excited state is equal to the central laser frequency ω_*, *i.e.*, $E_1 - E_0 = \omega_*$ if $\gamma = 0.891$ and $\xi_1 = 0$. Note, that γ does not normally hamper the application to experimental situations, as one typically knows the binding energy and the central photon frequency. Finally, we construct SHM $H_{\alpha\beta}(t)$ inserting E_α, and $V_{\alpha\beta}$ into eqn (3).

The idea of SHM is an essential part of our approach which serves two purposes: (i) it allows us to supply a sufficient number of theoretical learning data for the network and (ii) it represents a large variety of systems which could exist in nature but do not necessarily. The SHM should be "dense enough" in the parameter space such that always the Hamilton matrix of a real system one is interested in can be interpolated between SHMs, as interpolation capability is a strength of neural networks (in contrast to extrapolation). Of course, one can formulate more sophisticated SHMs with more parameters, but for the present case the four random parameters are sufficient.

Yet, we need to overcome one final obstacle, and that is the calculation of the spectra based on the SHMs. To obtain those spectra for arbitrary pulse forms $A(t)$ requires us to solve the time-dependent Schrödinger equation (TDSE) which in turn implies that we need an extremely fast propagation scheme to be able to solve the order of 10^7 TDSEs in a reasonable time.

2.4 Fast solution of the TDSE with SHMs

To achieve high propagation efficiency, we make use of the fact that the Hamilton matrix eqn (3) depends explicitly on time only through the vector potential $A = A(t)$. Hence, instead of discretizing the time equidistantly, we discretize the vector potential $A_{\min} \leq A \leq A_{\max}$ in $j_{\max}$ steps with $A_j = j\delta A$ with $\delta A = (A_{\max} - A_{\min})/(j_{\max} - 1)$.

With the time-independent Hamiltonian $H_j = H_0 + A_j \hat{p}$, we can construct a short-time propagator which is valid over a time span δt_j short enough such that a fixed A_j is a reasonable approximation. Therefore, the unitary short-time propagator can be obtained by direct integration,

$$U^j = \mathrm{e}^{-iH^j\delta t_j}. \tag{5}$$

The full propagator $U(t_\mathrm{f}, t_\mathrm{i}) = \prod_k U^{j_k}$, is now simply a concatenation of the short-time propagators over respective time spans δt_k (with $k = 1, \ldots, k_{\max}$) over which the discretized A_j hold, where $\delta t_1 = t_1 - t_\mathrm{i}$ and $\delta t_{k_{\max}} = t_\mathrm{f} - t_{k_{\max}-1}$.

To make efficient use of the SHMs, it is imperative that we use the matrix elements from eqn (4) as they do not require explicit integration over wave functions. Hence, we diagonalize $\langle\alpha|H^j|\beta\rangle = E_\alpha\delta_{\alpha\beta} + j\delta A V_{\alpha\beta}$ in the basis of H_0 to give its eigenenergies E^j_γ and eigenfunctions $\phi^j_\gamma = \sum_\alpha W^j_{\gamma\alpha}\phi_\alpha$ leading to the short-time propagator

$$U^j_{\alpha\beta} = \sum_\gamma W^j_{\alpha\gamma}\mathrm{e}^{-iE^j_\gamma\delta t_j} W^j_{\gamma\beta} \tag{6}$$

for fixed vector potential A_j. With $\delta A = 0.008$ we reach convergence in the solution of the TDSE which has been checked against converged propagation results by conventional solution for the cases studied here.

Note, that over the entire pulse $A(t)$ certain A_j may occur more than once with different time intervals over which they are valid (if the local derivative $\mathrm{d}A(t)/\mathrm{d}t|_{A_j}$ is large, the time interval will be small and *vice versa*). Therefore it is worthwhile to compute the $U^j_{\alpha\beta}$ beforehand and keep them stored. They can be used for all pulses (the fluctuating ones as well as the reference one) for a Hamilton matrix specified by the elements in eqn (4). Furthermore, we do not calculate the full

matrix of the propagator which would involve many matrix products. It is sufficient to propagate the vector $|0\rangle$ of the initial state (the ground state of the system) which requires only the computation of matrix-vector products. Only in this way were we able to calculate the millions of spectra, necessary to train the network.

3 Training the network

Through training with fluctuating spectra from the SHMs, the deep neural network encodes the dynamics of two-photon absorption spectra with the central photon frequency ω_* for all target systems covered by the SHMs. If the network "sees" during training a specific class of spectra much more often than representatives of other classes, it will be biased towards those often found spectra once trained. Hence, we have to fill the learning space of spectra (available for training, validating and testing the network) as homogeneously as possible.

3.1 Generating spectra

Synthetic Hamilton matrices which nearly satisfy the resonance condition, *i.e.*, $\xi_1 = 0$ in eqn (4), are particularly sensitive to the pulse shape and therefore generate more structured and diverse spectra through nonlinear processes, here resonant two-photon ionization, than SHMs with $\xi_1 \neq 0$. To sample the space of input spectra as homogeneously as possible, 50% of the spectra come from SHMs with $\xi_1 \approx 0$ and the other 50% of spectra are from SHMs with uniform ξ_1, randomly selected in the range $[-1, +1]$. After training on these spectra the network is not biased for ξ_1 around zero but works equally well for all ξ_1 in the specified range.

We calculate $n_{\mathrm{mat}} = 40\,000$ reference spectra from the same number of SHMs. For each reference spectrum, we calculate $n_{\mathrm{pul}} = 200$ spectra ("fluctuating spectra") from noisy pulses obtained with the partial-coherence method[11] using a different noise realization for each SHM. Since solving the TDSE for a single spectrum takes only a few seconds thanks to the highly-optimized propagation scheme outlined in Section 2.4, this procedure can be executed despite the need to solve about 10^7 TDSEs.

For each SHM, we average over all fluctuating spectra $\overline{P}_k(E) = \frac{1}{n_{\mathrm{pul}}}\sum_l P_{kl}(E)$ instead of using the individual fluctuating spectra $P_{kl}(E)$ computed from $H_{kl}(t)$, where k labels the Hamilton matrix and l the noisy pulse. We normalize all averaged fluctuating and reference spectra, *i.e.*, $\int \mathrm{d}E\, P(E) = 1$.

The resulting set of 40 000 averaged fluctuating spectra constitutes a major part of the learning space to train the networks in Section 4 for the prediction of spectra from different pulse shapes.

3.2 Parameterization of spectra and cost functions

For efficient representation we parameterize each spectrum $\bar{P}_k(E)$ in a basis of harmonic oscillator eigenfunctions $\{\chi_\kappa\}$,

$$\overline{P}_k(E) = \left| \sum_{\kappa=1}^{n_{\mathrm{bas}}} \overline{C}_k^{\kappa} \chi_{\kappa-1}(E) \right|^2, \tag{7}$$

with the vector $\bar{\mathbf{C}} \equiv \{\bar{C}_1 \ldots \bar{C}_{n_{\text{bas}}}\}$ of coefficients. A basis size of $n_{\text{bas}} = 100$ is required for the averaged fluctuating spectra, while for the noise-free spectra $n_{\text{bas}} = 60$ is sufficient.

The network maps the coefficients of the fluctuating spectra to those of the predicted underlying noise-free spectrum, $\{\bar{\mathbf{C}}_k\} \rightarrow \{\mathbf{C}_k\}$. The goal of the training is to minimize the difference between the predicted vector $\mathbf{C}_k$ for the noise-free spectrum and $\mathbf{C}_k^{\text{ref}}$ of the expected reference spectrum. The coefficients allow us to define a difference familiar from vector spaces as

$$\delta_\Omega \equiv \frac{1}{n_\Omega} \sum_{k=1}^{n_\Omega} \left[\mathbf{C}_k - \mathbf{C}_k^{\text{ref}}\right]^2, \tag{8a}$$

which we use for the cost function in the network training. As a measure for the difference of two (normalized) spectra i and j we define their "distance"

$$D_{ij} = \int \mathrm{d}E \left|P_i(E) - P_j(E)\right| \tag{8b}$$

and the average mutual distance

$$\overline{D}_\Omega = \frac{2}{n_\Omega(n_\Omega - 1)} \sum_{i>j} D_{ij} \tag{8c}$$

within a set of n_Ω spectra. With

$$\varepsilon_\Omega \equiv \frac{1}{n_\Omega} \sum_{k=1}^{n_\Omega} D_{k,k_{\text{ref}}}, \tag{8d}$$

one can quantify the error in terms of the distance eqn (8b), of the spectrum k from the reference spectrum k_{ref}, where $\varepsilon \leq 2$. The label Ω stands for the set of data the error is calculated for and can assume the values "train", "val", or "test" for training, validation or test data, respectively.

3.3 The training setup

The full set of learning data contains $n_{\text{mat}} = 40\,000$ pairs of spectra. Each pair consists of an averaged noisy spectrum with its respective reference spectrum for the same SHM. The full learning data set with n_{mat} pairs is split into training

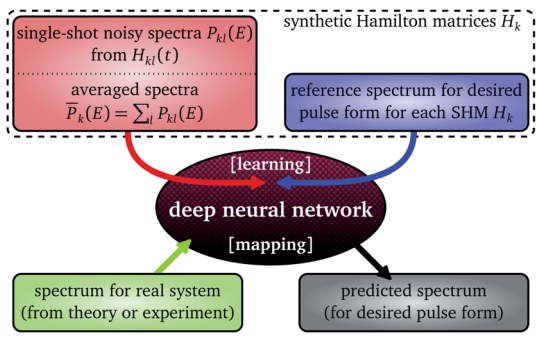

Fig. 1 Sketch of training and use of a deep neural network with synthetic Hamilton matrices and noisy spectra.

(80%), validation (10%) and test (10%) data. Training corresponds mathematically to minimizing the cost function eqn (8a), with Ω = train. Fig. 1 provides a sketch of what goes into training and prediction.

Implemented with the deep-learning library KERAS,[13] a fully connected feed-forward artificial neural network is used to establish the mapping. It contains 5 layers with 60 neurons on each and was trained at a learning rate of 0.001 with 100 epochs, a batch size of 200 and a learning patience of 25. Each hidden layer neuron contains a ReLU activation function.[14] The Adam optimizer[15] is used to minimize the cost function, eqn (8a). The training success is quantified with the error functions eqn (8a) and (8d), which both decay logarithmically with the size of the learning data, typical for deep learning.[16,17]

4 Prediction of spectra for different pulse shapes

To assess the quality of the mapping achieved with the trained networks on the basis of the SHM learning data, we will discuss scenarios with three different reference pulses for which we predict spectra: (i) double pulses with different time delays T_d and peak amplitude ratios $A_1 : A_2$, (ii) chirped pulses with chirp parameter β, and (iii) partially coherent reference pulses with different coherence times τ according to eqn (1). We have used the network setup for all three scenarios as described in the previous section with the *same* set of fluctuating spectra for training, but paired for each SHM with reference spectra which differ to the corresponding above reference pulses. The fluctuating spectra used as the input of the network have been generated with the pulses from eqn (1) with a pulse length of $T = 3$ fs, a coherence time of $\tau = 0.5$ fs, central photon frequency of $\omega_* = 21$ eV and spectral intensities between 8×10^{15} W fs cm^{-2} and 8×10^{16} W fs cm^{-2}.

For further reference and to give an overview how successfully the trained networks can predict spectra for the different pulse shapes from the fluctuating spectra, we show to begin with in Fig. 2 the absolute distance errors ($\varepsilon \leq 2$) of all predicted spectra. Note that for double pulses, the error decreases with increasing time-delay which is probably to be expected since it is easier to identify the time delay if it is larger. The smallest one $T_d = 4$ fs, basically corresponds to a single pulse (recall that the width of each individual pulse is $T = 3$ fs). Interestingly, the sensitivity to the amplitude ratios of the double pulses is even larger than to the time delay: the spectrum from a 1st pulse which is stronger than the 2nd one is

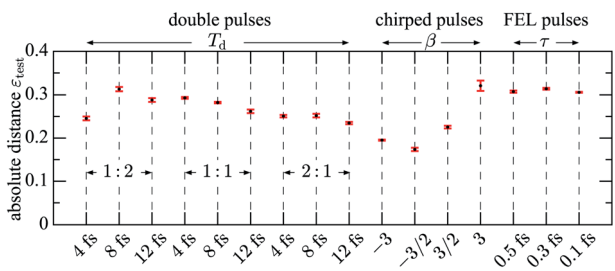

Fig. 2 Absolute distance error ε_{test} eqn (8d), of the different predicted spectra for test data: *double pulses* with time delay T_d and amplitude ratios $A_1 : A_2$ as indicated, *chirped pules* with chirp parameter β, and *partially coherent pulses* with coherence time τ.

easier to predict than *vice versa*, with pulses of equal strength taking the middle position in terms of the error.

The strongest sensitivity occurs for spectra from chirped pulses where the ones with the most positive chirp ($\beta = +3$) are twice as difficult to predict than ones with $\beta = -3$. We will come back to this point later. Finally, it is surprising that a spectrum from a partially-coherent pulse, which is naturally very "busy", can be identified and therefore predicted from the (averaged) fluctuating spectra, even if the coherence time is shorter than that of the noise ($\tau = 0.5$ fs) with similar accuracy as for longer coherence times of the reference spectrum. We will discuss the spectra from the different pulse forms now in detail.

4.1 Prediction of spectra from double pulses

The reference pulse is here given by

$$f_{\mathrm{d}}(t) = N_{\mathrm{d}}[A_1 G_T(t + T_{\mathrm{d}}/2) + A_2 G_T(t - T_{\mathrm{d}}/2)]\cos(\omega_* t), \tag{9}$$

where T_{d} is the delay between the maxima of the two pulses with shape G_T from eqn (1b), and respective amplitudes A_i. The normalization constant N_{d} is used in the same manner as in eqn (1).

Fig. 3 shows predicted spectra for exemplary double pulses with pulse shapes indicated in gray. Comparison of black and blue curves also helps to develop a sense for what the quantitative distance errors in Fig. 2 mean for the quality of the predictions. The generally good agreement proves that the training of the network was successful and has generated an accurate map.

However, the test data, although not used for training, belong to the same class of SHM that are used for training. A more realistic test is the prediction of a 3D helium spectrum as shown in Fig. 3 (middle), as this is similar to predicting spectra from experimental fluctuating pulses. In general, the prediction works very well, as one can see—only small details of the spectral structures are

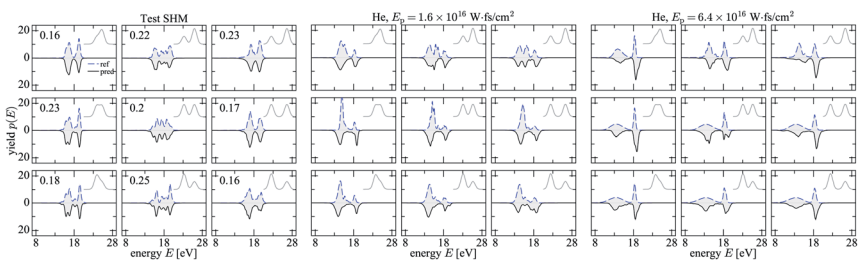

Fig. 3 Predicted photoelectron spectra (black) are compared to reference spectra (blue, dashed). All spectra are normalized. The corresponding reference pulses (gray) are shown in each panel. In each of the three figure matrices with 3 × 3 panels, time delays T_{d} are 4 fs, 8 fs, and 12 fs, respectively, from left to right; and from top to bottom, pulse amplitude ratios $A_1 : A_2$ are 1 : 2, 1 : 1, and 2 : 1, respectively, see eqn (9). *Left matrix*: prediction for a SHM from test data. The SHM is chosen such that $E_{\mathrm{p}} \approx 3.84 \times 10^{16}$ W fs cm^{-2} and each prediction returns an absolute distance (numbers in the panels) $D_{k_{\mathrm{ref}}}$, *cf.* eqn (8b), within the range of 30–70% in the error distribution. *Middle matrix*: prediction of noisy 3D helium spectra (composed of the sum of the two relevant angular-momentum channels s and d) through the trained network for pulses of pulse energy $E_{\mathrm{p}} = 1.6 \times 10^{16}$ W fs cm^{-2}. *Right matrix*: same as middle matrix but for an energy of $E_{\mathrm{p}} = 6.4 \times 10^{16}$ W fs cm^{-2}.

sometimes not resolved. This is remarkable, as the shapes of the spectra from the same reference pulses are quite different for the 1D system used for training and the 3D helium (compare the individual equivalent panels on the left and in the middle Fig. 3). This confirms the transferability of the network and underlines its interpolation capability.

Predictions become worse for increasing pulse energy as shown in the right part of Fig. 3. This is also true for the test data (not shown) but to a slightly lesser extent. While features are still reproduced, the predicted spectra are in general slightly too wide compared to the reference spectra.

4.2 Prediction of spectra from chirped pulses

The chirped reference pulses are parameterized by β and read

$$f_\beta(t) = N_\beta G_\beta(t)\cos(\varphi_\beta(t)), \tag{10a}$$

$$\varphi_\beta(t) = \omega_* t + \frac{2\ln 2}{\beta + 1/\beta}\frac{t^2}{T^2}, \quad T_\beta = \sqrt{1+\beta^2}\,T, \tag{10b}$$

with the Gaussian from eqn (1b) and $T = 3$ fs. Again we normalize the pulse energy, here by means of N_β, as before in eqn (1) and (9). The predicted spectra are shown in Fig. 4. They do not exhibit detailed structure, mostly a single peak with different form of the shoulders and reconstruction seems to work well with the exception of large positive chirp, where the position of the spectral peak is systematically red shifted in the predicted spectrum consistent with the largest error (see Fig. 2) the positively chirped spectra have.

4.3 Prediction of spectra from partially-coherent pulses

We finally will predict spectra from pulses which are themselves "noisy", *i.e.*, partially coherent and generated according to eqn (1) but for different coherence

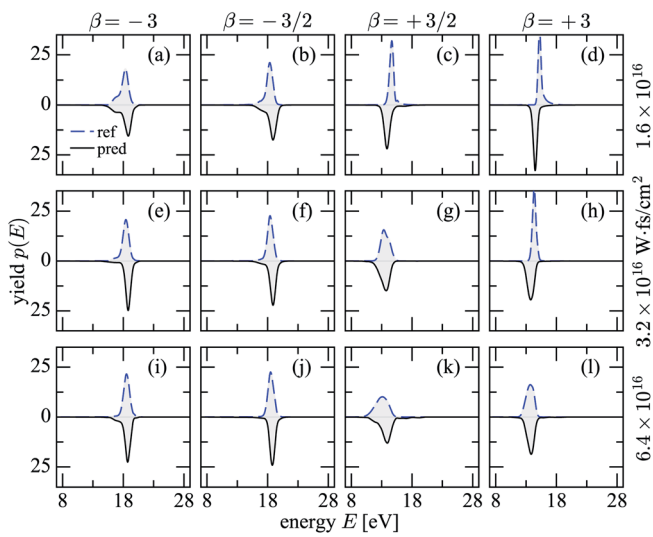

Fig. 4 Prediction of 3D helium spectra (black, solid) for chirped pulses, eqn (10). The reference 3D helium spectra are shown with blue dashed lines.

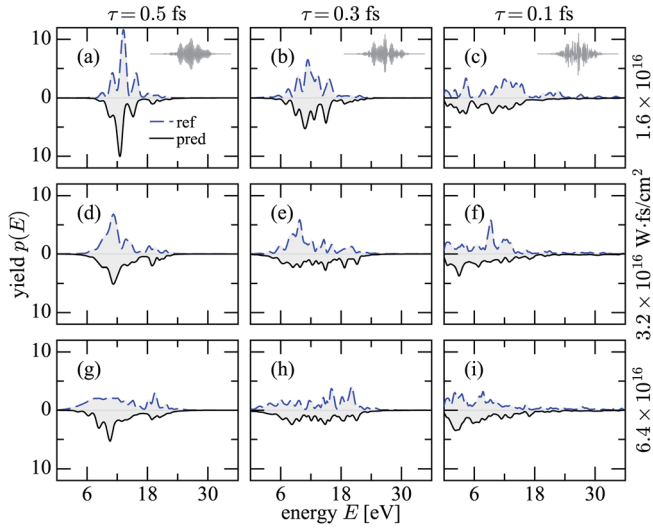

Fig. 5 Prediction of 3D helium spectra (black) for partially coherent pulses (1). The reference 3D helium spectra are shown with blue dashed lines.

times τ, typical for SASE FELs. The motivation for such reference spectra was to see where the prediction breaks down since we had the expectation that, at least for spectra from pulses with coherence times much shorter than the ones used for the learning space of fluctuating spectra, the trained network would lose its predictive capability, even more so as the spectra have quite detailed features, see Fig. 5. However, to our surprise this is not the case, as also revealed by the errors given in Fig. 2.

4.4 Prediction errors for different pulse shapes

Now, we are in the position to understand details of the distance errors $\varepsilon_{\text{test}}$ in Fig. 2 for reference spectra from different pulse shapes. As one can see from Fig. 6 as a rule of thumb, the smaller the ionization probability P_{ion} (shown in Fig. 6 with blue points), the smaller the diversity of spectra the pulses generate, including reference spectra. All spectra in this section have been analyzed with networks trained with a learning data set of the same size and a common set of input

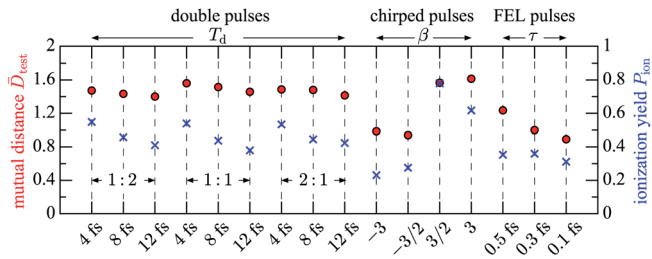

Fig. 6 Properties of test reference spectra for the pulses from Fig. 2. Average ionization yield P_{ion} (blue, right axis) and average mutual distance $\bar{D}_{\text{test}}$ (red, left axis).

averaged fluctuating spectra. Therefore, one would expect that the average mutual distance $\bar{D}_\Omega$, defined in eqn (8c), of the reference spectra is larger for a more extended space of highly diverse spectra as compared to a smaller space of less diverse spectra. This is indeed the case as $\bar{D}_{\text{test}}$ shown with red points in Fig. 6 reveal: they follow the trend of P_{ion} for the test data. Since it is more difficult for the network to interpolate if the available reference spectra are more distant, one would expect larger errors, which explains the trend of the distance errors in Fig. 2. Particularly striking is the change for chirped pulses:[18] negative chirp produces small P_{ion} and in turn a moderate diversity of spectra with relatively small $\bar{D}_{\text{test}}$, and therefore also the smallest $\varepsilon_{\text{test}}$. For positive chirp, the exact opposite holds. One cannot expect that ionization yield, distance of spectra and errors are directly proportional, as the physical process leading from the pulses to the spectra is still non-linear. For instance, long time-delays in double pulses give rise to more diverse spectra than short time-delays. Moreover, the $\varepsilon_{\text{test}}$ are for predictions from noisy spectra. Yet, the causal chain of $P_{\text{ion}} \rightarrow \bar{D}_{\text{test}} \rightarrow \varepsilon_{\text{test}}$ holds.

This section has shown that the trained networks can predict spectra from widely varying pulse forms well. The effort one has to invest into the deep neural networks for the prediction of the spectra depends on the diversity of spectra a certain pulse form is capable of generating.

5 Single-shot noisy double pulses: simultaneous purification of spectra and reconstruction of time-delay

The analysis of the previous section has prepared us for the final goal of this work, namely purifying the spectra while simultaneously extracting the correct time-delay from spectra recorded with noisy double pulses which have an unknown time-delay within a certain interval. This scenario is motivated by SASE XFEL pulses,[19] where the pulse is either split by a chicane for the relativistic electron bunch, which creates the light pulse, or by situations where an XFEL pulse and a time-delayed strong laser pulse are used together, whereby the delay between the two pulses is characterized by a jitter from shot to shot.

We model fluctuating double pulses with noise-free double pulses and admixture of noisy double pulses,

$$f_{\text{dq}}(t) = N_{\text{dq}}[G_T(t + T_{\text{d}}/2) + G_T(t - T_{\text{d}}/2)][\cos(\omega_* t) + qF_\tau(t)], \tag{11}$$

where $q = 0.32$, $\tau = 0.3$ fs, G_T and F_τ are from eqn (1) and the time-delays T_{d} vary between 2 fs and 14 fs. Hence, for this task we have to create a new learning space of fluctuating spectra as input for the network based on fluctuating double pulses. And again, the normalization factor N_{dq} ensures the required pulse energy.

Since so far we have not extracted the time-delay of the pulses from the spectra, we verify in Section 5.1, that it is possible to identify the time-delay of double pulses from noise-free spectra generated by those pulses. In Section 5.2 we will address fluctuating spectra. We first determine the pulses' time-delay T_{d} encoded in single-shot spectra generated with noisy double pulses. Subsequently, we average the single-shot spectra with identified T_{d} over small intervals of time-delay (1 fs) and purify these averaged spectra. Recall, that purifying means that

we remove the fluctuations from spectra by predicting the spectra generated from the respective noise-free pulse forms, in the present case from the noise-free double pulses.

5.1 Extraction of time-delay from spectra generated with double pulses

Here, we aim at constructing a network-based map to extract the time-delays T_d of double pulses from the (noise-free) spectra the pulses f_{d0} from eqn (11) generate. To this end we have generated a learning data set of spectra from 20 000 SHMs, each paired with a single double pulse $f_{d0}(t)$ with delays between 2 and 14 fs. The learning data is distributed into training, validation and test data as before (see Section 3.3), and the network is also that of Section 3.3, but the number of neurons on each layer is 50, the learning rate is 0.008 and the number of epochs is 200.

Fig. 7 shows the training success with the SHMs as well as the transfer of the network to unknown 3D helium spectra. The trained network reproduces well the delays (results scatter along the ideal red line with an error given in the inset). For short T_d the results deviate from the ideal line since the individual pulses in the double pulse have a width of $T = 3$ fs which limits the resolution towards small time-delays. Results for the reconstructed time-delay for full 3D helium spectra

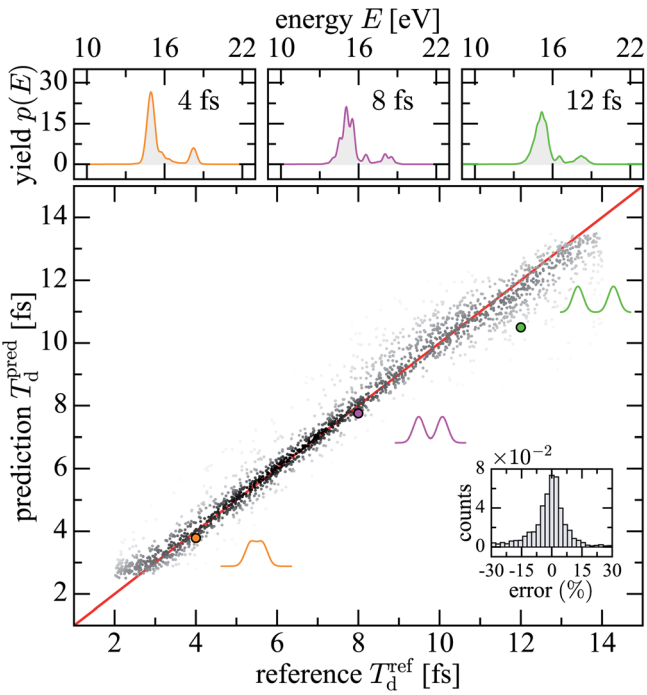

Fig. 7 Predicted time-delays against reference time-delays for the test data. The pulse energy is $E_p = 1.6 \times 10^{16}$ W fs cm^{-2}. The error distribution of the time-delays for the test data is shown in the lower inset. The red line represents error-free prediction. The trained network is transferred to the 3D helium spectra for three time delays: 4 fs, 8 fs, and 12 fs with the reconstructed time delays are shown as circles, and the double pulse shapes sketched. The upper inset gives the corresponding photoelectron spectra.

are given for T_d of 4, 8, and 12 fs, respectively, and demonstrate the transferability of the network. The upper row shows the corresponding 3D helium spectra. Given the similarity of these spectra for different time-delays it is remarkable that the trained network can reliably extract the time-delays. We may conclude that we can map out the delay of the pulse from the spectrum it has produced with the help of the trained network.

5.2 Purification of single-shot spectra and simultaneous extraction of the time-delay of the generating double pulse

Finally, we analyze noisy single-shot spectra with the goal to purify them as in Section 4 and to extract the time delay of the generating double pulse as in Section 5.1, simultaneously. In order to have reasonable statistics for the map and also reasonably different spectra for different time delays, we reconstruct from each noisy single-shot spectrum (all for the same SHM) the time-delay but average the spectra afterwards over small intervals (1 fs) of time-delays. Subsequently, the averaged spectra are passed through another trained network to purify them. The result is shown in Fig. 8. The scattered points are reconstructed time-delays coloured with the reference time-delays. The even change in color demonstrates that the reconstruction of time-delays for the test data has been successful. The spectra within 1 fs intervals of reconstructed time-delays are averaged and subsequently purified. They are shown on the right in red along with reference spectra (black),

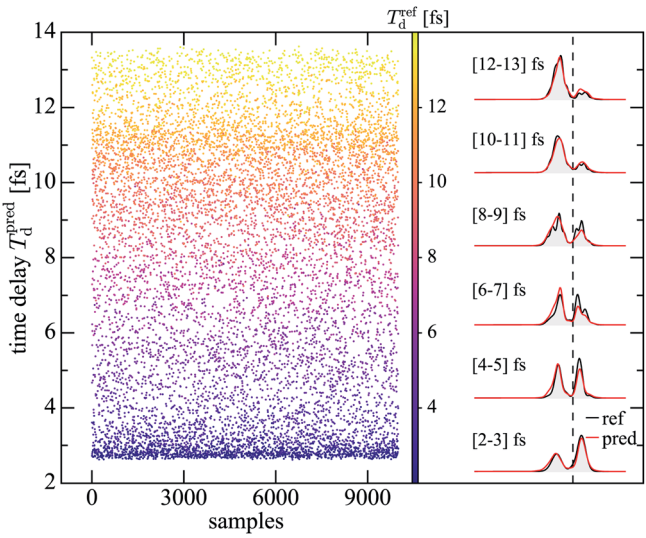

Fig. 8 Simultaneous reconstruction of time-delay and purification of noisy spectra for a single Hamilton matrix taken from test data. Single-shot fluctuating spectra for random time-delays are passed through a network to reconstruct the underlying time delays which are shown as scattered points where the color represents the reference time-delay. We consider 12 intervals of time delay in the range 2–14 fs with interval length of 1 fs. All single-shot spectra which fall into interval of time-delay are averaged. The averaged spectra are passed through another network which maps averaged noisy spectra to purified ones. The predicted purified spectra (red) are compared to reference spectra (black).

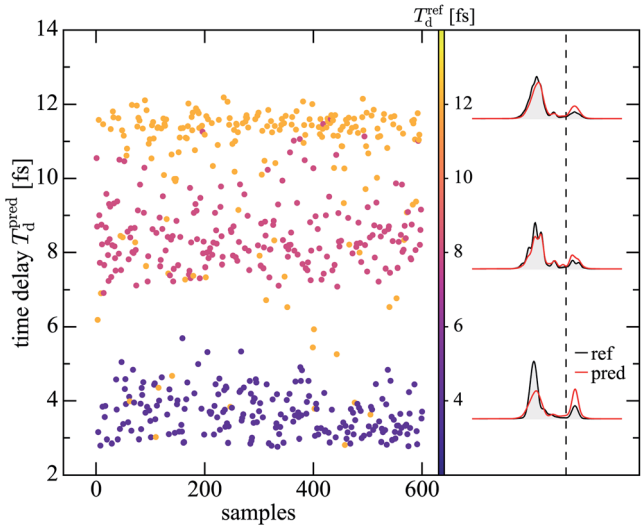

Fig. 9 Same as Fig. 8 but for 3D helium for which the network was not trained. The distribution of predicted time-delays shows three main peaks at 4, 8, and 12 fs. The single-shot spectra are averaged over all spectra with time-delays in an interval of 1 fs about the three peaks. The averaged spectra are passed through the trained network to obtain the corresponding purified spectra shown on the right (red). The three averaged reference spectra (black) are obtained in the same way.

averaged over the same interval of time-delays. The generally good agreement demonstrates that reconstruction of time-delays and purification of the single-shot spectra is possible without additional information over the single-shot spectra.

The last step is to prove that the reconstruction and purification can be transferred to spectra unknown to the networks. To this end we take noisy single-shot spectra of 3D helium with three well-defined time-delays and pass them through the trained network for reconstruction of the time-delay. The scattered points in Fig. 9 show the reconstructed time-delays. We average the corresponding spectra over 1 fs around the three peak time-delays in the scattered points and pass the averaged spectra through the purification network to arrive at the three spectra on the right in red. They agree well with the corresponding reference spectra, averaged over the same intervals of time-delay (black). Hence, the trained networks should be able to reconstruct the time-delay and purify the corresponding fluctuating experimental spectra as they are produced by SASE FELs.

6 Conclusions

To summarize, we have devised a strategy to create maps through deep neural networks between fluctuating nonlinear photo-ionization spectra and noise-free spectra, and between fluctuating single-shot spectra and pulse properties. A crucial part of this strategy is the formulation of synthetic Hamilton matrices which describe artificial systems, similar to ones existing in reality. We use the SHM to generate a sufficient amount of spectra for training the network. In a first

application[6] we purified fluctuating spectra as typically produced by SASE FELs through a neural-network-based map.

Here, we have taken this mapping capability to a new level by predicting from fluctuating spectra—which should come ultimately from experiment—the spectra which would be obtained with specific noise-free pulses, namely double pulses, chirped pulses and chaotic (partially-coherent) pulses. While generally the prediction works as well as the purification (prediction) for simple Gaussian pulses, the error analysis has revealed interesting differences for the different pulse shapes.

In a second application we have constructed a neural-network-based map which can extract the time-delay of double pulses from fluctuating single-shot spectra generated by those noisy double pulses. Finally, we could demonstrate that suitably trained networks can achieve both, purification and extraction of the time-delay, from fluctuating single-shot spectra as typically produced by SASE FELs. Clearly, neural networks open promising new ways to analyze particular noisy data with a potential which has been by far not exhausted.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge financial support by "BiGmax", the Max Planck Society's research network on big-data-driven materials science, and by "QUTIF", a priority program (no. 1840) of the Deutsche Forschungsgemeinschaft. Open Access funding was provided by the Max Planck Society.

Notes and references

1 V. Dunjko and H. J. Briegel, *Rep. Prog. Phys.*, 2018, **81**, 074001.
2 P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher and D. J. Schwab, *Phys. Rep.*, 2019, **810**, 1.
3 G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborová, 2019, arXiv:1903.10563.
4 R. Selle, T. Brixner, T. Bayer, M. Wollenhaupt and T. Baumert, *J. Phys. B: At., Mol. Opt. Phys.*, 2008, **41**, 074019.
5 A. Sanchez-Gonzalez, P. Micaelli, C. Olivier, T. R. Barillot, M. Ilchen, A. A. Lutman, A. Marinelli, T. Maxwell, A. Achner, M. Agåker, N. Berrah, C. Bostedt, J. D. Bozek, J. Buck, P. H. Bucksbaum, S. C. Montero, B. Cooper, J. P. Cryan, M. Dong, R. Feifel, L. J. Frasinski, H. Fukuzawa, A. Galler, G. Hartmann, N. Hartmann, W. Helml, A. S. Johnson, A. Knie, A. O. Lindahl, J. Liu, K. Motomura, M. Mucke, C. O'Grady, J.-E. Rubensson, E. R. Simpson, R. J. Squibb, C. Såthe, K. Ueda, M. Vacher, D. J. Walke, V. Zhaunerchyk, R. N. Coffee and J. P. Marangos, *Nat. Commun.*, 2017, **8**, ncomms15461.
6 S. K. Giri, U. Saalmann and J. M. Rost, *Phys. Rev. Lett.*, 2020, **124**, 113201.
7 J. White and Z. Chang, *Opt. Express*, 2019, **27**, 4799.
8 Z. Zhu, J. White, Z. Chang and S. Pang, *Sci. Rep.*, 2020, **10**, 1.

9 R. Ziv, A. Dikopoltsev, T. Zahavy, I. Rubinstein, P. Sidorenko, O. Cohen and M. Segev, *Opt. Express*, 2020, **28**, 7528.
10 X. Liu, G. Zhang, J. Li, G. Shi, M. Zhou, B. Huang, Y. Tang, X. Song and W. Yang, *Phys. Rev. Lett.*, 2020, **124**, 113202.
11 T. Pfeifer, Y. Jiang, S. Düsterer, R. Moshammer and J. Ullrich, *Opt. Lett.*, 2010, **35**, 3441.
12 R. Moshammer, T. Pfeifer, A. Rudenko, Y. H. Jiang, L. Foucar, M. Kurka, K. U. Kühnel, C. D. Schröter, J. Ullrich, O. Herrwerth, M. F. Kling, X.-J. Liu, K. Motomura, H. Fukuzawa, A. Yamada, K. Ueda, K. L. Ishikawa, K. Nagaya, H. Iwayama, A. Sugishima, Y. Mizoguchi, S. Yase, M. Yao, N. Saito, A. Belkacem, M. Nagasono, A. Higashiya, M. Yabashi, T. Ishikawa, H. Ohashi, H. Kimura and T. Togashi, *Opt. Express*, 2011, **19**, 21698.
13 F. Chollet, *Keras: The Python deep learning library*, https://keras.io, 2015.
14 B. Hanin, *Mathematics*, 2019, **7**, 992.
15 D. P. Kingma and J. L. Ba, 2017, arXiv:1412.6980 [cs].
16 S.-i. Amari, N. Fujita and S. Shinomoto, *Neural Comput.*, 1992, **4**, 605.
17 J. Hestness, S. Narang, N. Ardalani, G. Diamos, H. Jun, H. Kianinejad, M. M. A. Patwary, Y. Yang and Y. Zhou, 2017, arXiv:1712.00409 [cs, stat].
18 U. Saalmann, S. K. Giri and J. M. Rost, *Phys. Rev. Lett.*, 2018, **121**, 153203.
19 A. Marinelli, D. Ratner, A. A. Lutman, J. Turner, J. Welch, F.-J. Decker, H. Loos, C. Behrens, S. Gilevich, A. A. Miahnahri, S. Vetter, T. J. Maxwell, Y. Ding, R. Coffee, S. Wakatsuki and Z. Huang, *Nat. Commun.*, 2015, **6**, 6369.