Purifying Electron Spectra from Noisy Pulses with Machine Learning Using Synthetic Hamilton Matrices

Sajal Kumar Giri, Ulf Saalmann, and Jan M. Rost
Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Straße 38, 01187 Dresden, Germany

(Received 9 August 2019; accepted 18 February 2020; published 17 March 2020)

Photoelectron spectra obtained with intense pulses generated by free-electron lasers through self-amplified spontaneous emission are intrinsically noisy and vary from shot to shot. We extract the purified spectrum, corresponding to a Fourier-limited pulse, with the help of a deep neural network. It is trained on a huge number of spectra, which was made possible by an extremely efficient propagation of the Schrödinger equation with synthetic Hamilton matrices and random realizations of fluctuating pulses. We show that the trained network is sufficiently generic such that it can purify atomic or molecular spectra, dominated by resonant two- or three-photon ionization, nonlinear processes which are particularly sensitive to pulse fluctuations. This is possible without training on those systems.

DOI: 10.1103/PhysRevLett.124.113201

Recent years have seen an avalanchelike increase of machine-learning applications in physics [1–3], which roughly fall into three categories: (a) applications within theory, e.g., for quantum information [1] or to elucidate intricate many-body properties [4], (b) within experiment to optimize experimental conditions, e.g., to characterize a free-electron laser (FEL) pulse [5], and (c) applications that condition learning algorithms theoretically with the goal to apply the trained model to experimental data. Our work falls in category (c). Although in principle far more general, we choose to be specific and apply the approach we develop to the purification of noisy photoelectron spectra as routinely obtained with self-amplified spontaneous emission (SASE) FELs operating in the desired frequency range.

Our goal is to train a deep neural network with sufficiently many noisy spectra and their pure counterparts, such that the trained network will be able to purify a noisy spectrum which is not contained in the training data, in particular an experimental one. With purification, we mean that upon feeding with a noisy photoelectron spectrum the network returns a reference spectrum that would be obtained if the target system would be driven by an ideal Gaussian laser pulse, which we call the reference pulse, cf. Fig. 1. This may seem straightforward. Yet, it is anything but trivial to generate a sufficient amount of suitable training data with an acceptable effort. This is in general the bottleneck for machine-learning applications in theory which requires new ways of thinking. In this vein, we introduce synthetic Hamilton matrices (SHMs). Synthetic means that we vary the matrix elements (here in a random fashion) about base values such that later on the trained network is able to purify real spectra from either experiment or theory. The SHMs are constructed to speed up the generation of training data and we also expect them to become useful for other dynamical problems for which neural networks must be trained. Since the SHMs cover a large range of possible systems we can afford to use for the base itself explicitly calculated photoionization dynamics in one dimension which is fast to compute and provides a suitable anchor point for the SHMs.

Setting up networks with SHMs.—To put our approach to a credible test we need (i) a physical process, which is sensitive to the pulse profile, (ii) a realistic way to model fluctuating pulses and we need to prepare a large set of spectra suitable for training the network. This involves (iii) a scheme to efficiently propagate millions of time-dependent Schrödinger equations, (iv) a broad and uniform

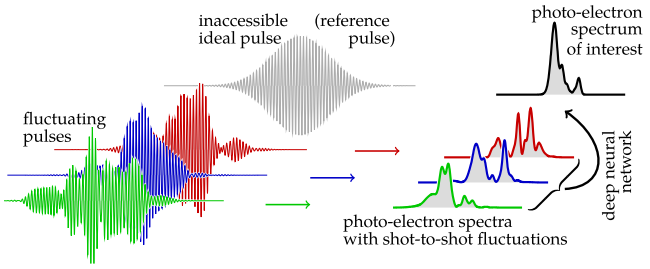

FIG. 1. Sketch of the problem: Photoelectron spectra from fluctuating pulses are purified using a deep neural network.

sampling of the generated spectra, and (v) a trainable parametrization.

(i) As a physical process which is nonlinear in the driving light and therefore very sensitive to the intensity of the light pulse and hence its profile in time, we have chosen quasiresonant few-photon ionization. It can lead to multipeak structures in the photoelectron spectrum [6–10].

(ii) Fluctuating pulses from SASE FELs can be modeled by the so-called partial-coherence method [11,12], an experimentally verified method, which allows one to create ensembles of pulses $f_l(t)$ which differ through fluctuations while the ensemble average converges to a well defined pulse shape [13]. Those pulses have a characteristic duration T and a coherence time τ, we use $T = 3$ fs and $\tau = 1/2$ fs here. Apart from the intrinsic noise the pulses additionally jitter in their pulse energy. We normalize all pulses $f_l(t)$ to unit pulse energy. This is also possible in the experiment as pulse energies can be easily measured shot to shot with gas monitoring detectors [22].

(iii) In principle numerical codes are available to propagate the time-dependent Schrödinger equation (TDSE) for one active electron in a strong laser field and calculate the resulting photoelectron spectrum $P(E)$ [23–25]. However, the creation of a training data set from millions of pulses is prohibitively expensive, yet essential for successful deep learning.

To overcome this obstacle we work with Hamilton matrices whose construction is detailed in the Supplemental Material [13]. The new element, particularly formulated for the present context, is the generation of n_{mat} Hamilton matrices with random energies E_α^k, coupling matrix elements $V_{\alpha\beta}^k$, and field strengths A_k, corresponding to intensities (referring to the Fourier-limited pulse) in the range of $5 \times 10^{15}, \ldots, 5 \times 10^{16}$ W/cm^2. Furthermore, for each Hamilton matrix the coupling to the light is augmented by n_{pul} noise realizations $f_l(t)$ with a central frequency of 21 eV to arrive at

$$\mathbf{E}_k = E_\alpha^k \delta_{\alpha\beta}, \qquad \mathbf{V}_k = V_{\alpha\beta}^k, \tag{1a}$$

$$\mathbf{H}_{kl}(t) = \mathbf{E}_k + A_k f_l(t) \mathbf{V}_k, \tag{1b}$$

whereby $k = 1, \ldots, n_{\mathrm{mat}}$ and $l = 1, \ldots, n_{\mathrm{pul}}$. Boldface symbols in Eqs. (1) describe matrices in terms of field-free states. It is only through these synthetic Hamilton matrices that we are able to create a sufficient number of nontrivial training data. The matrices have been derived varying a 1D Hamilton operator (our base system), but since the energies $\mathbf{E}_k$ and the coupling matrix elements $\mathbf{V}_k$ are chosen randomly, these SHMs can purify real (3D) spectra, as we will see subsequently.

(iv) We have to create a set of spectra for training, validation, and testing which should cover to a large extent the domain of realizable spectra. This step is crucial and most expensive numerically, particularly when compared to the (modest) resources needed to set up and train the network. To cover the domain of realizable spectra uniformly, we calculate first 4×10^4 reference spectra [26]. Among those we select the $n_{\mathrm{mat}} = 2 \times 10^4$ spectra with the largest mutual difference

$$D_{ij} = \int dE |P_i(E) - P_j(E)|. \tag{2}$$

For each member of this subset of reference spectra, we calculate $n_{\mathrm{pul}} = 500$ fluctuating spectra from noisy pulses generated with the partial-coherence method [11,13] with a different noise realization for each (synthetic) Hamilton matrix. Hence, we must propagate about $n_{\mathrm{mat}} \times n_{\mathrm{pul}} = 10^7$ TDSEs, which takes, however, only a few seconds for a single TDSE thanks to our highly optimized propagation scheme [13]. It includes prediagonalization of the Hamilton matrices which saves computing time since one and the same system is propagated for different pulse realizations. Finally, we have for each Hamilton matrix (1) one reference spectrum $P_k^{\mathrm{ref}}(E)$ and n_{pul} noisy spectra $P_{kl}(E)$, i. e., a total of $n_{\mathrm{mat}} \times [n_{\mathrm{pul}} + 1]$ spectra.

Instead of the individual $P_{kl}(E)$ we use partially averaged spectra $\overline{P}_{kj}(E) = (1/m) \sum_{l \in \{s_j\}} P_{kl}(E)$ for efficient training. To this end we draw a random subset $\{s_j\}$ containing m spectra from the n_{pul} fluctuating spectra for each SHM and repeat this procedure $\bar{n}_{\mathrm{pul}}$ times to create $j = 1, \ldots, \bar{n}_{\mathrm{pul}}$ averaged spectra. For our application $\bar{n}_{\mathrm{pul}} = 10$ and $m = 200$ is a good compromise between rugged spectra for smaller m and an increased training effort for larger m. All spectra are normalized, i.e., $\int dE\, P(E) = 1$.

(v) To complete the final step, the parametrization of the spectra for training, we represent the resulting averaged spectra $\overline{P}_{kj}(E)$ in terms of harmonic oscillator eigenfunctions χ_κ as

$$\overline{P}_{kj}(E) = \left| \sum_{\kappa=1}^{n_{\mathrm{bas}}} \overline{C}_{kj}^{\kappa} \chi_{\kappa-1}(E) \right|^2, \tag{3}$$

with the set $\overline{\mathbf{C}} \equiv \{\overline{C}_1, \ldots, \overline{C}_{n_{\mathrm{bas}}}\}$ of coefficients. A basis size of $n_{\mathrm{bas}} = 60$ was necessary for the averaged fluctuating spectra, while using a similar expression for the noise-free spectra $n_{\mathrm{bas}} = 40$ was sufficient [13]. The network consists of mapping the coefficients $\{\overline{\mathbf{C}}_{kj}\} \to \{\mathbf{C}_{kj}\}$. The training aims at minimizing the difference between the predicted $\mathbf{C}_{kj}$ for the noise-free spectrum and the expected reference spectra $\mathbf{C}_k^{\mathrm{ref}}$.

The connection between Hamilton matrices, pulses, and electron spectra just outlined is summarized schematically in Fig. 2.

Building and training the network.—With $n_{\mathrm{mat}} = 2 \times 10^4$ reference spectra and $\bar{n}_{\mathrm{pul}} = 10$ averaged noisy "copies" of each reference spectrum, we have $n \equiv n_{\mathrm{mat}} \times \bar{n}_{\mathrm{pul}} = 2 \times 10^5$ pairs available for building the network model. Each pair

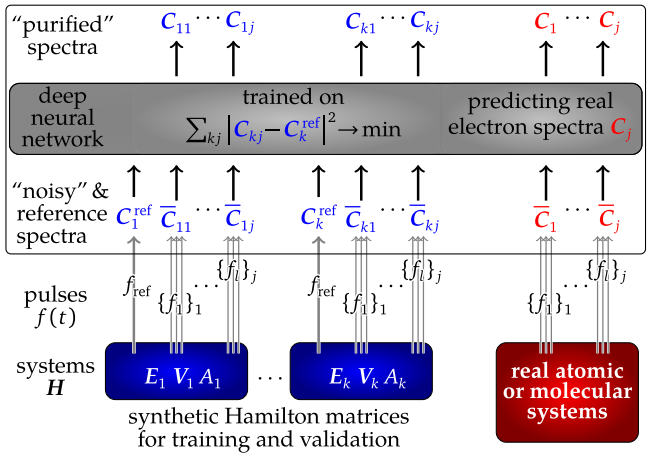

FIG. 2. Building a network with synthetic Hamilton matrices and noisy pulses (left-hand side, blue) and using it (right-hand side, red). The setup of n_{mat} SHMs is exemplified with the first ($\mathbf{H}_1$) and kth ($\mathbf{H}_k$) one. With noisy pulse shapes $f_l(t)$ or the reference pulse shape $f_{\mathrm{ref}}(t)$ the SHMs are amended to $\mathbf{H}_{1l}$ and $\mathbf{H}_{kl}$, respectively, see Eqs. (1). Noisy spectra as calculated with the $\{\mathbf{H}_{kl}\}$ are averaged over 200 realizations with the same k resulting for each $k = 1, \ldots, n_{\mathrm{mat}}$ SHM in ten different spectra with coefficients $\overline{\mathbf{C}}_{kj}$, $j = 1, \ldots, 10$. The network is trained on the predictions $\mathbf{C}_{kj}$ together with the reference $\mathbf{C}_k^{\mathrm{ref}}$, i.e., only spectra are processed by the network as emphasized by the black frame. The right-hand side (red) sketches how noisy spectra (from either experiment or theory) to be purified are also averaged over 200 realizations before fed into the trained network to retrieve the reference spectrum.

consists of an averaged noisy spectrum with its respective reference spectrum. Note that the network operates exclusively on the electron spectra, cf. Fig. 2. Once trained, it is therefore directly applicable to the experiment which has only access to spectra.

The full data set with n pairs of spectra is split in the ratio 8:1:1 between training ($n_{\mathrm{train}} = 0.8n$), validation ($n_{\mathrm{val}} = 0.1n$), and test ($n_{\mathrm{test}} = 0.1n$) data. Implemented with the deep-learning library KERAS [27], a fully connected feed-forward neural network is used [13]. The training success and resulting performance of the network as a function of the size of the training data is quantified with the cost function δ, using the basis representation (3) of the spectra, and a more intuitive error ε

$$\delta_\Omega \equiv \frac{1}{n_\Omega} \sum_{j,k=1}^{n_\Omega} [\mathbf{C}_{kj} - \mathbf{C}_k^{\mathrm{ref}}]^2, \tag{4a}$$

$$\varepsilon_\Omega \equiv \frac{1}{n_\Omega} \sum_{j,k=1}^{n_\Omega} D_{kj,k\mathrm{ref}}, \tag{4b}$$

for training ($\Omega = \mathrm{train}$), validation ($\Omega = \mathrm{val}$), and test ($\Omega = \mathrm{test}$) data set, respectively. The error ε_Ω with an upper limit $\varepsilon \leq 2$ measures the difference $D_{kj,k\mathrm{ref}}$, see Eq. (2), between a predicted spectrum $P_{kj}(E)$ and its reference spectrum $P_k^{\mathrm{ref}}(E)$. The maximal error $\varepsilon = 2$ occurs if the two normalized (i.e., unit-area) spectra are completely disjunct. Both errors (4) decay logarithmically as a function of the SHM data size n [13].

Purification of spectra from SHMs.—We are finally in a position to purify noisy spectra and do this first with the n_{test} SHM-generated spectra the network was not trained on. Typical snapshots of these spectra are shown in Fig. 3(d). To get a realistic picture we have selected spectra, cf. Figs. 3(a)–3(c), which belong to three groups purified with different residual errors in increasing order: Only 1% of the spectra have a purification error better than the one shown in Fig. 3(a), the prediction in Fig. 3(b) has a median error $\varepsilon = \varepsilon_{50\%}$ such that half of the spectra have a smaller, and half of them have a larger prediction error. Finally, only 1% of the purified spectra have a larger error than the one shown in Fig. 3(c). The gray-shaded curves provide the reference spectrum $P_k^{\mathrm{ref}}(E)$ in each case. The simple average (from the test-data set for a specific SHM and field intensity) set is shown as a dashed line.

One sees that the purification works quite well, even for a typical "worst case" as in Fig. 3(c), where all peaks including the fine structure, appear at the correct energies, despite the fact that none of the features is contained in the fully averaged spectra $\overline{P}_k(E) = \sum_j \overline{P}_{kj}(E)$. We also note that spectra of rather different shapes and details of the structure, from a smooth single peak [Fig. 3(a)] over a triple peak [Fig. 3(b)] to a fine-structured multipeak shape

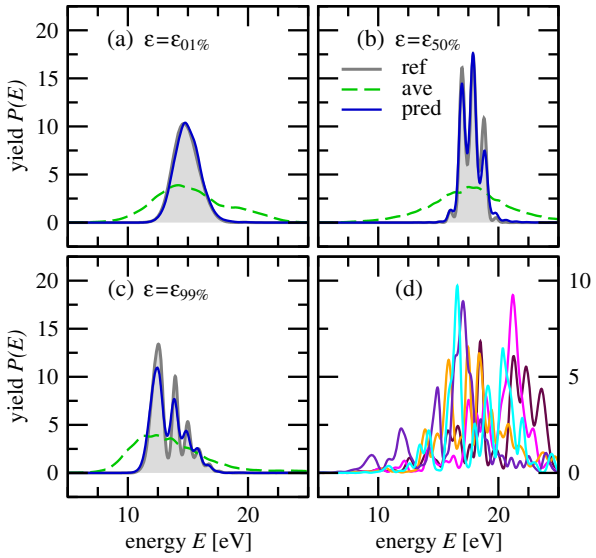

FIG. 3. Photoelectron spectra from the SHM test-data set. The full average of fluctuating spectra (green-dashed line) and prediction from the network (blue) are compared to the reference (gray and shaded). Panels (a)–(c): Examples with three prediction errors $\varepsilon = \varepsilon_p$ are shown, with p indicating the percentage of spectra having a smaller error, i.e., 99% of all spectra from the test-data set have a smaller prediction error than the one shown in panel (c). Panel (d): Five single-shot spectra for the Hamilton matrix used in panel (b).

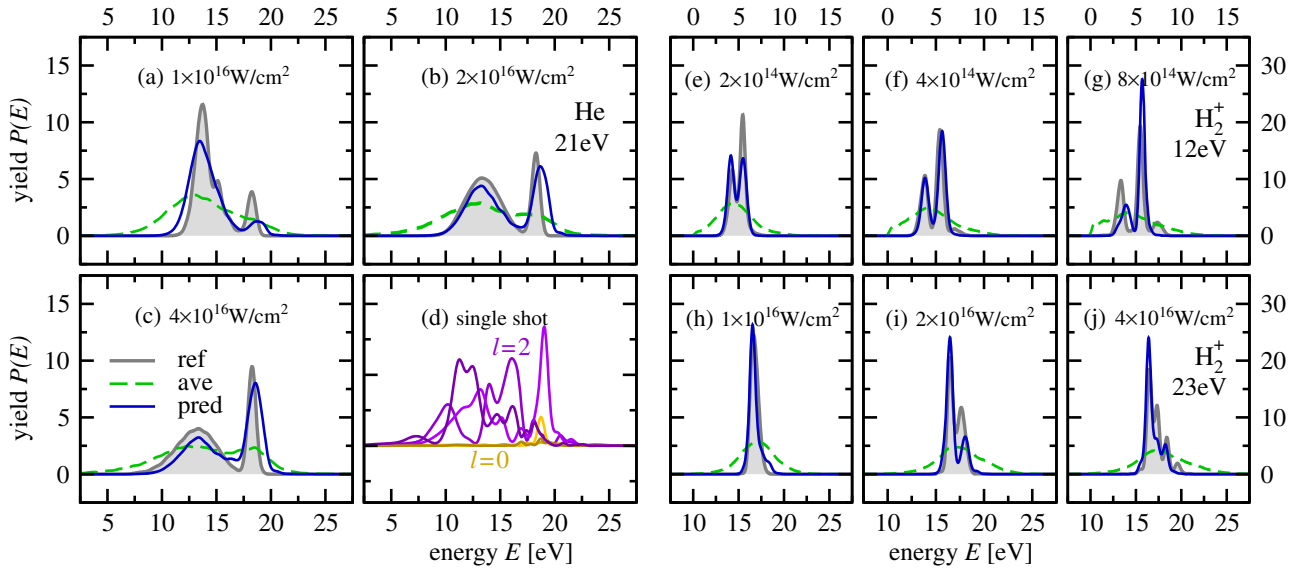

FIG. 4. Photoelectron spectra for the He atom (a)–(d) and the H_2^+ molecule (e)–(j). For He we show results for one photon-frequency ω (case α in the text), for H_2^+ for two different ω (cases β and γ in the text), whereby each combination is presented for three different intensities I (with ω and I are being specified in the panels). As in Fig. 3, fully averaged and predicted spectra are compared to the reference spectra. Three single-shot spectra for $I = 2 \times 10^{16}$ W/cm^2, as used in panel (b), are shown in panel (d), separately for the s channel ($\ell = 0$, orangelike lines) and the d channel ($\ell = 2$, violetlike lines), respectively.

[Fig. 3(c)], can be purified successfully. The rather diverse spectra $P_{kl}(E)$ from fluctuating pulses $f_l(t)$ for a specific SHM, as shown in Fig. 3(d), indicate the strong sensitivity to the pulse profile which is due to Stark shifts and Autler-Townes splittings. The complete failure of the fully averaged spectra in revealing the reference spectrum $P_k^{\mathrm{ref}}(E)$ is striking. This happens despite the fact that the reference *pulse* is retrieved by averaging a sufficient number of fluctuating pulses if created by the partial-coherence method [11]. The corresponding reference *spectrum*, however, is never obtained by averaging the fluctuating spectra since the underlying ionization dynamics is nonlinear. The consequence is an intricate mapping between fluctuating spectra and the reference spectrum, which is constructed with the deep neural network.

Purification of spectra from physical systems.—So far the successful purification referred to spectra not known to the network, but generated through SHMs which were also used to train the network, only with different parameters. In the following, we will apply the trained network to photoelectron spectra for three cases of two different physical systems: ($\boldsymbol{\alpha}$) The He atom dominated by 2-photon absorption [28] and the hydrogen molecule ion H_2^+ ionized by ($\boldsymbol{\beta}$) 2- and ($\boldsymbol{\gamma}$) 3-photon processes. These spectra have been obtained in full 3D; for technical details see [13]. In case $\boldsymbol{\alpha}$ [Figs. 4(a)–4(c)] the spectra consist of contributions from the s and d manifolds, which can be reached by a 2-photon process, whereby the d channel clearly dominates [cf. Fig. 4(d)]. For H_2^+, aligned along the laser polarization, either the gerade continuum for case $\boldsymbol{\beta}$ [Figs. 4(h)–4(j)], or the ungerade continuum for case $\boldsymbol{\gamma}$ [Figs. 4(e)–4(g)], is considered. The central frequencies for the laser pulse are chosen according to the transition energies $\omega_\alpha = E_{2p} - E_{1s} = 20.95$eV, $\omega_\beta = E_{2\sigma_u} - E_{1\sigma_g} = 23.05$eV, and $\omega_\gamma = E_{1\sigma_u} - E_{1\sigma_g} = 11.83$ eV, respectively. Fluctuating pulses $f_l(t)$ are created as before but we use new random realizations. As in the training procedure, we have created ten averaged spectra, which are fed into the trained network. Each one is composed of 200 fluctuating spectra [26]. The ten resulting purified spectra from the network are again averaged to arrive at the network's estimate of the reference spectrum.

We show in Fig. 4 results for three different intensities in the range where few-photon ionization is nonperturbative. As expected from SHM-generated spectra in Fig. 3, the fully averaged spectra (green-dashed lines) do not provide sensible information about the reference spectra. The mapping with the network (blue-solid lines), however, reveals the respective peak structure of the photoelectron spectra.

Note that the network was neither trained on the spectra of the 3D helium atom nor on those of the hydrogen molecule ion, nevertheless these spectra are purified successfully with the network mapping as shown in Fig. 4. The training of the network was performed with synthetic data derived from a representative 1D photoionization dynamics only, which allowed us to keep the size of the Hamilton matrices small enough to be able to compute the 10^7 TDSEs for a sufficient amount of training data. Apparently, although generated from the 1D derived ones, the SHMs represent dynamical systems sufficiently generic such that also realistic 3D spectra from the three rather different processes $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ could be purified with the

same network. Hence, it should also work on experimental spectra, which will be slightly different to the extent to which many-electron effects show up in photoelectron spectra as compared to the present 3D single-active-electron calculations. To measure reference spectra in a proof-of-principle experiment one could either use seeded FEL pulses [29–32] or set up an experiment at a coherent (high-harmonic) source and generate noisy pulses artificially.

To summarize, we have devised a strategy to purify noisy photoelectron spectra, typical for SASE FELs with the help of a deep neural network. While this example was chosen on purpose to be specific, through its design our approach is far more general. First, we have checked [13] that other noise models [33,34] can be used. Second, purification could be conditioned on any arbitrary reference pulse. Third, and most importantly, the systematic introduction of synthetic Hamilton matrices permits one to generate a training data set of ample size with reasonable computational effort and renders the trained network applicable for scenarios where it was not trained for. In the present example, we applied the network trained on synthetic dynamics to purify realistic 3D spectra. For future work, we would like to point out that noisy pulses driving nonlinear processes are actually advantageous, since they allow one to obtain the target response over a wide spectral and dynamic range in a single shot, provided one has tools to analyze the resulting spectra.

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) through the priority program 1840 "Quantum Dynamics in Tailored Intense Fields" and "BiGmax", the Max Planck Society's research network on big-data-driven materials science.

[1] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress, Rep. Prog. Phys. **81**, 074001 (2018).

[2] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, Phys. Rep. **810**, 1 (2019).

[3] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[4] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[5] A. Sanchez-Gonzalez *et al.*, Accurate prediction of x-ray pulse properties from a free-electron laser using machine learning, Nat. Commun. **8**, 15461 (2017).

[6] D. Rogus and M. Lewenstein, Resonant ionisation by smooth laser pulses, J. Phys. B **19**, 3051 (1986).

[7] S. H. Autler and C. H. Townes, Stark effect in rapidly varying fields, Phys. Rev. **100**, 703 (1955).

[8] C. Meier and V. Engel, Interference Structure in the Photoelectron Spectra Obtained from Multiphoton Ionization of Na_2 with a Strong Femtosecond Laser Pulse, Phys. Rev. Lett. **73**, 3207 (1994).

[9] P. V. Demekhin and L. S. Cederbaum, Coherent intense resonant laser pulses lead to interference in the time domain seen in the spectrum of the emitted particles, Phys. Rev. A **86**, 063412 (2012).

[10] M. Baghery, U. Saalmann, and J. M. Rost, Essential Conditions for Dynamic Interference, Phys. Rev. Lett. **118**, 143202 (2017).

[11] T. Pfeifer, Y. Jiang, S. Düsterer, R. Moshammer, and J. Ullrich, Partial-coherence method to model experimental free-electron laser pulse statistics, Opt. Lett. **35**, 3441 (2010).

[12] R. Moshammer *et al.*, Second-order autocorrelation of XUV FEL pulses via time-resolved two-photon single ionization of He, Opt. Express **19**, 21698 (2011).

[13] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.113201 for details on generating laser pulses, propagating the time-dependent Schrödinger equation and training of artificial neural networks, which includes additional Refs. [14–21].

[14] G. Geloni, E. Saldin, L. Samoylova, E. Schneidmiller, H. Sinn, T. Tschentscher, and M. Yurkov, Coherence properties of the European XFEL, New J. Phys. **12**, 035021 (2010).

[15] X. M. Tong and C. D. Lin, Empirical formula for static field ionization rates of atoms and molecules by lasers in the barrier-suppression regime, J. Phys. B **38**, 2593 (2005).

[16] L. D. Landau and E. M. Lifschitz, *Quantum Mechanics* (Pergamon Press, Oxford, 1989).

[17] X. Glorot, A. Bordes, and Y. Bengio, Deep sparse rectifier neural networks, Proc. Mach. Learn. Res. **15**, 315 (2011), http://proceedings.mlr.press/v15/glorot11a.html.

[18] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, Proc. Mach. Learn. Res. **9**, 249 (2010), http://proceedings.mlr.press/v9/glorot10a.html.

[19] D. P. Kingma and J. L. Ba, ADAM: A method for stochastic optimization, arXiv:1412.6980.

[20] M. Li, T. Zhang, Y. Chen, and A. J. Smola, Efficient mini-batch training for stochastic optimization, in *Proceedings of the 20th ACM SIGKDD Conference* (Association for Computing Machinery, New York, 2014), p. 661.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. **15**, 1929 (2014), http://jmlr.org/papers/v15/srivastava14a.html.

[22] K. Tiedtke *et al.*, Gas detectors for x-ray lasers, J. Appl. Phys. **103**, 094511 (2008).

[23] L. Tao and A. Scrinzi, Photo-electron momentum spectra from minimal volumes: The time-dependent surface flux method, New J. Phys. **14**, 013021 (2012); A. Scrinzi, t-SURFF: Fully differential two-electron photo-emission spectra, New J. Phys. **14**, 085008 (2012).

[24] V. Mosert and D. Bauer, Photoelectron spectra with Qprop and t-SURFF, Comput. Phys. Commun. **207**, 452 (2016).

[25] S.PatchkovskiiandH. G.Muller,Simple,accurate,andefficient implementation of 1-electron atomic time-dependent

Schrödinger equation in spherical coordinates, Comput. Phys. Commun. **199**, 153 (2016).

[26] We will use the terms "reference spectrum" and "fluctuating spectrum" to refer to a "spectrum from a reference pulse" and a "spectrum from a fluctuating pulse", respectively. We never deal with the spectral representation of the pulse itself.

[27] F. Chollet, Keras: The Python deep learning library, https://keras.io (2015).

[28] U. Saalmann, S. K. Giri, and J. M. Rost, Adiabatic Passage to the Continuum: Controlling Ionization with Chirped Laser Pulses, Phys. Rev. Lett. **121**, 153203 (2018).

[29] J. Amann *et al.*, Demonstration of self-seeding in a hard-x-ray free-electron laser, Nat. Photonics **6**, 693 (2012).

[30] E. Allaria *et al.*, Highly coherent and stable pulses from the FERMI seeded free-electron laser in the extreme ultraviolet, Nat. Photonics **6**, 699 (2012).

[31] E. Allaria *et al.*, Two-stage seeded soft-x-ray free-electron laser, Nat. Photonics **7**, 913 (2013).

[32] P. R. Ribič *et al.*, Coherent soft x-ray pulses from an echo-enabled harmonic generation free-electron laser, Nat. Photonics **13**, 555 (2019).

[33] N. Rohringer and R. Santra, X-ray nonlinear optical processes using a self-amplified spontaneous emission free-electron laser, Phys. Rev. A **76**, 033416 (2007).

[34] G. M. Nikolopoulos and P. Lambropoulos, Effects of free-electron-laser field fluctuations on the frequency response of driven atomic resonances, Phys. Rev. A **86**, 033420 (2012).

Supplement "Purifying electron spectra from noisy pulses with machine learning using synthetic Hamilton matrices"

Sajal K Giri, Ulf Saalmann, and Jan M Rost

Technical details for laser pulses, the propagation of the time-dependent Schrödinger equation and the training of artificial neural networks are provided. Parameters for the numerical calculations are specified.

1 Partial-coherence method

In order to create fluctuating pulses we apply the so-called partial-coherence method (PCM), which has been devised to simulate pulses from free-electron laser (FEL) sources [1]. Thereby a fluctuating (or noisy) pulse is given by

$$f_l(t) = N_l\, G(t)\, \mathcal{F}^{-1}\big[\mathrm{e}^{\mathrm{i}\phi_l(\omega)}\mathcal{F}[g(t)\cos(\omega_* t)]\big], \tag{S1a}$$

$$G(t) = \mathrm{e}^{-2\ln 2\, t^2/T^2} \quad \text{and} \quad g(t) = \mathrm{e}^{-t^2/\tau^2}, \tag{S1b}$$

with ω_* the carrier frequency. $\mathcal{F}$ and $\mathcal{F}^{-1}$ denotes the Fourier and the inverse Fourier transform, respectively. The actual noise realization, indicated by the index l, is given by random spectral phases ϕ_l, uniformly distributed in the range of $-\pi \ldots +\pi$. For $\phi(\omega)=0$ and $\tau\to 0$ we get an ideal Gaussian pulse — here the reference pulse $f_{\mathrm{ref}}(t)$. The two Gaussians, defined in Eq. (S1b), appear in the pulse expression (S1a) inside and outside the Fourier transforms and serve different purposes. On one hand, $g(t)$ quantifies the time scale of the fluctuations[1] by means of the coherence time τ. On the other side, $G(t)$ is a masking function in time that fixes the typical pulse duration T, which would be otherwise arbitrarily long. Thus $\tau \ll T$ implies strongly fluctuating pulses, which is the typical case at X-ray free-electron laser as, e. g., the European XFEL [2]. The pre-factor N_l guarantees that the pulse energy of the fluctuating pulse is the same as the one of the reference pulse: $\int \mathrm{d}t\, {f_l}^2(t) = \int \mathrm{d}t\, {f_{\mathrm{ref}}}^2(t)$. This "normalization" is the only difference to the standard PCM [1].

For the propagation of the time-dependent Schrödinger equation, cf. Sect. 3 below, we need the time dependent vector potential $A_l(t)$ at equidistant instants of time $t_\eta = \eta\,\delta t$ with $\eta = -\eta_{\max} \ldots + \eta_{\max}$. For obtaining $A_l(t_\eta)$ we have chosen maximum time $t_{\max} = -t_{\min} = 3{,}000$ a.u. and $\eta_{\max} = 2^{16}$ (thereby $\delta t \approx 0.046$ a.u.), did a fast Fourier transformation FFT, created $2\eta_{\max}$ random phases ϕ_η uniformly between $-\pi$ and $+\pi$, and did finally an inverse FFT.

2 Synthetic Hamilton matrices and creation of training/validation data

In order to create a sufficient amount of training data (we calculate 10^7 time-dependent Schrödinger equations) we resort to Hamilton matrices based on an 1-dimensional model system. The starting point is the Hamilton operator

$$\widehat{H}(t) = -\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} - \frac{1}{\sqrt{x^2+1/2}} + A(t)\,\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}, \tag{S2}$$

for which we define numerically on a grid ($x_j = j\Delta x$, with $\Delta x = 0.1$ a.u. and $x_{\max} = 500$ a.u.) all eigenstates $\widehat{H}\varphi_\alpha = \varphi_\alpha \widetilde{E}_\alpha$ with $\widetilde{E}_\alpha \le E_{\max} \approx 48$ eV, resulting in 600 eigenstates. By means of these states, we build the (600×600) Hamilton matrix

$$\widetilde{H}_{\alpha\beta}(t) = \widetilde{E}_\alpha\,\delta_{\alpha\beta} + A(t)\,\widetilde{V}_{\alpha\beta} \qquad \text{with} \quad \widetilde{V}_{\alpha\beta} \equiv \big\langle \varphi_\alpha \big| \mathrm{i}\tfrac{\mathrm{d}}{\mathrm{d}x} \big| \varphi_\beta \big\rangle. \tag{S3}$$

[1] One might consider other functions $g(t)$, e. g. [3], $g_{\mathrm{s}}(t) = \mathrm{sech}(\pi t/\sqrt{2}\tau)$ or $g_{\mathrm{e}}(t) = \exp(-2|t|/\tau)$. All three have the same coherence time, but different decay characteristics. We have tested all three noise types. Since they did not show any qualitative difference in terms of predictability of spectra, we present only results for the Gaussian-type noise (S1b).

This is a model of a 1-dimensional “helium atom”, since the ground-state energy $E_0 = -24.2\,\mathrm{eV}$ is close to the ionization potential of real helium (24.6 eV).

As we want to train the network for “arbitrary” quantum systems, we create synthetic Hamilton matrices by randomly changing energies $\widetilde{E}_\alpha$ and matrix elements $\widetilde{V}_{\alpha\beta}$ in the following way

$$E_\alpha = 3^{[\xi_1-\gamma]}\widetilde{E}_\alpha \qquad \text{for } \widetilde{E}_\alpha<0,\ \alpha>0, \tag{S4a}$$

$$V_{0\alpha} = 3^{\xi_2}\widetilde{V}_{0\alpha} \qquad \text{for } \widetilde{E}_\alpha<0, \tag{S4b}$$

$$V_{\alpha\beta} = 3^{\xi_3}\widetilde{V}_{\alpha\beta} \qquad \text{for } \widetilde{E}_\alpha<0,\ \widetilde{E}_\beta>0, \tag{S4c}$$

$$V_{\alpha\beta} = 3^{\xi_4}\widetilde{V}_{\alpha\beta} \qquad \text{for } \widetilde{E}_\alpha>0,\ \widetilde{E}_\beta>0, \tag{S4d}$$

with four uniform random numbers $\xi_1 = [-1,+1.1]$, $\xi_2 = [-2,+2]$, and $\xi_{3,4} = [-1,+1]$. Thereby we modify the bound-state energies (S4a) and the couplings between ground and bound states (S4b), between bound and free states (S4c) and among free states (S4d), respectively. With $\gamma = 0.88$ and $\xi_1 = 0$ the energy difference between ground and excited state is equal to the central laser frequency ω_*, i. e. $E_1 - E_0 = \omega_*$.

3 Propagation of the time-dependent Schrödinger equation (TDSE)

“1D” systems for training data

Typically the propagation of the TDSE is done in fixed time steps δt over which the Hamilton matrix is assumed to be constant, i. e. one step is taken by

$$\psi_\alpha(t+\delta t) = \sum_\beta U_{\alpha\beta}(t+\delta t, t)\psi_\beta(t), \tag{S5a}$$

$$U_{\alpha\beta}(t+\delta t, t) = \mathrm{e}^{-\mathrm{i}H_{\alpha\beta}(t+\delta t/2)\,\delta t}, \tag{S5b}$$

with $U_{\alpha\beta}$ the matrix representation of the (unitary) time-evolution operator.

In order to improve efficiency (considerably) we refrain from using a fixed time step, but rather discretize the laser pulse with a given step size in the *vector potential* $A(t)$. With δA the step height, we can find intervals of time $t = t_j^{\mathrm{beg}} \ldots t_j^{\mathrm{end}}$ for which $\lfloor A(t)/\delta A \rceil = j$, with $\lfloor \ldots \rceil$ denoting the nearest integer. Now we have variable time steps, in which the assumption of constant values of the vector potential is a good approximation by construction. Such a discretization perfectly adapts to any pulse. Figure S1 shows this discretization of the pulse.

Since there could be more than one interval for a given j we pre-diagonalize the matrices

$$H^j_{\alpha\beta} = E_\alpha\delta_{\alpha\beta} + j\,\delta A V_{\alpha\beta} \qquad \rightarrow \qquad E^j_\alpha,\ T^j_{\alpha\beta} \tag{S6}$$

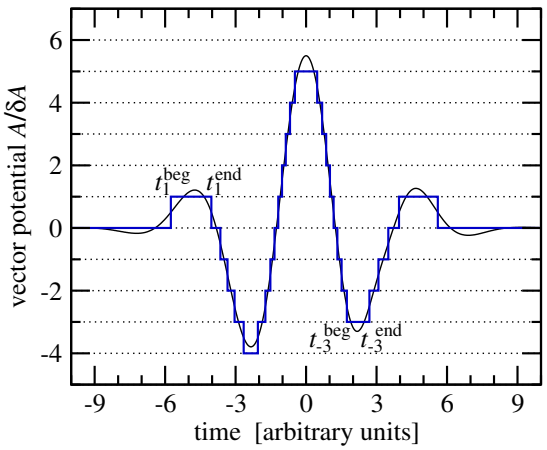

Figure S1: Sketch of the discretization of a time-dependent vector potential $A(t)$ in units of δA. Two generic time steps $t_j^{\mathrm{beg}} \ldots t_j^{\mathrm{end}}$ are indexed. Note that the step height δA here is rather large for visualization purposes.

and calculate by means of the eigenvalues E_α^j and eigenvectors $T_{\alpha\beta}^j$ the corresponding matrices for the time evolution operator

$$U_{\alpha\beta}^j = \sum_\gamma T_{\gamma\alpha}^j e^{-\mathrm{i}E_\gamma^j[t_j^{\mathrm{end}}-t_j^{\mathrm{beg}}]} T_{\gamma\beta}^j \tag{S7}$$

for all j. Additionally and similarly important in view of efficiency, the pre-calculated E_α^j and $T_{\alpha\beta}^j$ can be used for all pulses (the fluctuating ones as well as the reference one) for a specific system. Only by this method we were able to calculate millions of spectra, necessary for training the network.

Physical systems for applications

On one hand, the number of TDSEs to be solved in the 3D case are considerably lower. On the other hand, a single calculation is more expensive.

He In order to calculate the field-free eigenstates we use an effective potential [4]. For angular momenta $\ell = 0 \ldots 4$ we calculated the lowest 2250 states in a box of size $r = 0 \ldots 3{\times}10^3 \mathrm{a}_0$ with a Numerov step size of $\delta r = 0.01\,\mathrm{a}_0$.

H_2^+ Field-free states and their couplings are calculated in terms of prolate-spheroidal coordinates $\{\xi, \eta\}$, in which the problem separates [5]. The angular domain $-1 \le \xi \le +1$ is described in terms of Legendre polynomials, the radial domain $1\,\mathrm{a}_0 \le \eta \le 2000\,\mathrm{a}_0$ in terms of B-splines. For the coupling we assume that the molecule is aligned with the laser polarization, the orientation with the large dipole matrix elements. The internuclear distance is taken as $R = 2\,\mathrm{a}_0$.

The propagation in time is achieved by direct computation

$$\psi_\alpha(t+\delta t) = \sum_{k=0}^{k_{\max}} \frac{[-\mathrm{i}\,\delta t]^k}{k!} \sum_{\beta_1 \ldots \beta_k} H_{\alpha\beta_k}(t+\delta t/2, t) \ldots H_{\beta_2\beta_1}(t+\delta t/2, t)\psi_{\beta_1}(t), \tag{S8}$$

with $k_{\max}$ defined by a sufficiently small contribution, i. e. $\left|\sum_{\beta_1 \ldots \beta_k} \ldots\right|^2 < 10^{-30}$. Hereby, $k_{\max}$ may change from step to step.

4 Photo-electron spectra

From the solution of the time-dependent Schrödinger equation, i. e. by means of the field-free-state expansion amplitudes $a_\alpha(t)$ for $t \to \infty$, we get the corresponding photo-electron spectrum as

$$P(E) = \sum_\alpha \left|a_\alpha(+\infty)\right|^2 K(E - E_\alpha), \tag{S9}$$

with the kernel function $K(E) = \exp(-E^2/\Delta E^2)/\sqrt{\pi}\Delta E$. We have chosen $\Delta E = 0.2\,\mathrm{eV}$, which is slightly smaller than the typical spacing between the field-free states.

Those spectra are fitted by the k first harmonic oscillator eigenfunctions $\chi_i(E)$ according to

$$P(E) = \left|\sum_{i=0}^{k-1} C_i \chi_i(E)\right|^2. \tag{S10}$$

Those eigenfunctions (normalized by the pre-factor N_i) are given by

$$\chi_i(E) = N_i\, \mathrm{e}^{-\Omega_*[E-E_*]^2/2} H_i\big(\sqrt{\Omega_*}[E-E_*]\big), \tag{S11a}$$

$$\text{with } \Omega_* = T_0^2/4\ln 2 \text{ and } E_* = E_0 + 2\omega_*, \tag{S11b}$$

where H_i stands for the ith Hermite polynomial. Hereby, the frequency Ω_* matches the bandwidth of the pulse and the displacement E_* corresponds to the energy reached by a two-photon transition from the ground state. Thus, a weak Gaussian pulse (in perturbation regime) with a duration T_0 and a carrier frequency ω_* would give a photo-electron spectrum that is very similar to $\chi_0(E)$ as defined in Eq. (S11).

5 Artificial neural network

We use a fully connected feed-forward artificial neural network to establish a mapping from noisy $\overline{\mathbf{C}}_{kj}$ to noise-free spectra $\mathbf{C}_k^{\mathrm{ref}}$, cf. Fig. 2 in the main text. In our network $\overline{\mathbf{C}}_{kj}$ and $\mathbf{C}_{kj}$ are connected through 7 hidden layers which contain 40 nodes each. They are linked with weights $\mathbf{W}$ and have biases $\mathbf{B}$ and activation functions $\mathbf{f}$. This simple network can be described mathematically as

$$\mathbf{C} = \mathbf{W}_7\mathbf{Y}_7 + \mathbf{B}_7, \quad \mathbf{Y}_{7-k} = \mathbf{f}_{7-k}\big(\mathbf{W}_{6-k}\mathbf{Y}_{6-k} + \mathbf{B}_{6-k}\big) \text{ with } k = 0\ldots6, \quad \mathbf{Y}_0 = \overline{\mathbf{C}}. \tag{S12}$$

In the above equations $\mathbf{W}_k$ represents weights connecting kth layer neurons with $(k+1)$th layer neurons and $\mathbf{B}_k$ stands for biases linked to $(k+1)$th layer neurons. The dimensions of $\mathbf{Y}_k$, $\mathbf{W}_k$ and $\mathbf{B}_k$ are n_k, $n_{k+1}\times n_k$ and n_{k+1}, respectively, where n_k is the number of kth layer neurons. (The activation functions $\mathbf{f}_{k+1}$ are to be understood as n_k-dimensional functions.) Note that input and output layer dimensions are $n_{\mathrm{bas}} = 60$ and $n_{\mathrm{bas}} = 40$ respectively, the number of basis functions to represent the photoelectron spectra. The total number of trainable parameters in the network (S12) is 13,920. We use *ReLU* [6], i. e. $f(x) = \max(0, x)$, as activation function to introduce the non-linearity in the connection for all hidden layers. Note that this functional value does not saturate for very large value of x.

For the initial choice of random weights and biases, we use the so-called XAVIER initialization [7], where random weights are considered from a normal distribution with zero mean and a variance of $2/[n_k+n_{k+1}]$, where n_k is the number of neurons at kth layer. Such an initialization reduces the possibility to drive the network into saturated states and accelerates convergence. All biases $\mathbf{B}$ are set to 1 initially. The weights and biases of the network (S12) are obtained using the ADAM optimization algorithm [8], which is very efficient for stochastic optimization that is based on an adaptive-learning method by which it computes individual learning rates for each weight and the biases in the network. We use the same values for the parameters β_1, β_2, α, and ε as the ones given in "Algorithm 1" in the original paper on the ADAM method [8].

In order to reduce the computational cost, we applied mini-batch optimization [9] with the batch size being 500 and introduced early stopping with a patience of 35 (the number of epochs with an increasing cost function on the validation data). Typically we had to run about 100 epochs in order to train the network. On a single batch, a forward pass and a backward pass is a single iteration. Each epoch covers all training samples once. For a total sample size of 1.6×10^5 and a batch size of 500 we need 320 iterations to complete a single epoch. To reduce the complexity of the network during the training we drop some neurons with a dropping rate 0.1 [10].

Training-data size

The value of the cost function [Eq. (4) in the main text] as well as the average difference [Eq. (5) Ibid.] decrease logarithmically with the training-data size as shown for validation data set in Fig. S2. The figure shows also that along with the mean, obtained from several optimization runs for the network, the standard deviation among those runs decreases.

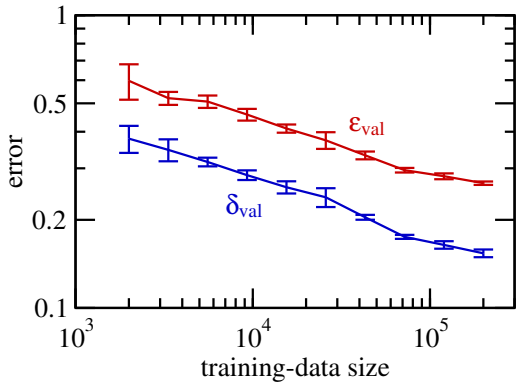

Figure S2: Predicted performance by the validation data measured with the cost function δ_{val} [Eq. (4)] and the absolute error $\varepsilon_{\mathrm{val}}$ [Eq. (5)] as a function of the size n of pairs of spectra contained in the complete SHM data set.

References

[1] T. Pfeifer, Y. Jiang, S. Düsterer, R. Moshammer, and J. Ullrich, *Partial-coherence method to model experimental free-electron laser pulse statistics*. Opt. Lett. **35**, 3441 (2010).

[2] G. Geloni, E. Saldin, L. Samoylova, E. Schneidmiller, H. Sinn, Th Tschentscher, and M. Yurkov, *Coherence properties of the European XFEL*. New J. Phys. **12**, 035021 (2010).

[3] G. M. Nikolopoulos and P. Lambropoulos, *Effects of free-electron-laser field fluctuations on the frequency response of driven atomic resonances*. Phys. Rev. A **86**, 033420 (2012).

[4] X. M. Tong and C. D. Lin, *Empirical formula for static field ionization rates of atoms and molecules by lasers in the barrier-suppression regime*. J. Phys. B **38**, 2593 (2005).

[5] L. D. Landau and E. M. Lifschitz, *Quantum mechanics*. Pergamon Press, Oxford 1989.

[6] X. Glorot, A. Bordes, and Y. Bengio, *Deep sparse rectifier neural networks*. PMLR **15**, 315 (2011).

[7] X. Glorot and Y. Bengio, *Understanding the difficulty of training deep feedforward neural networks*. PMLR **9**, 249 (2010).

[8] D. P. Kingma and J. L. Ba, *ADAM: A method for stochastic optimization*. arXiv:1412.6980 [cs] (2017).

[9] M. Li, T. Zhang, Y. Chen, and A. J. Smola, *Efficient mini-batch training for stochastic optimization* in *Proceedings of the 20th ACM SIGKDD Conference*, 661 (2014).

[10] N. Srivastava, G. Hinton, A. Krizhevsky , I. Sutskever, and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*. JMLR **15**, 1929 (2014).