scientific reports

OPEN

Artificial intelligence for online characterization of ultrashort X-ray free-electron laser pulses

Kristina Dingel [1,2,8✉], Thorsten Otto[1,8], Lutz Marder[2,3], Lars Funke[4], Arne Held[4], Sara Savio[4], Andreas Hans[2,3], Gregor Hartmann[2,5], David Meier[1,2,5], Jens Viefhaus[2,5], Bernhard Sick[1,2], Arno Ehresmann [2,3], Markus Ilchen[3,6,7] & Wolfram Helml[4]

X-ray free-electron lasers (XFELs) as the world's brightest light sources provide ultrashort X-ray pulses with a duration typically in the order of femtoseconds. Recently, they have approached and entered the attosecond regime, which holds new promises for single-molecule imaging and studying nonlinear and ultrafast phenomena such as localized electron dynamics. The technological evolution of XFELs toward well-controllable light sources for precise metrology of ultrafast processes has been, however, hampered by the diagnostic capabilities for characterizing X-ray pulses at the attosecond frontier. In this regard, the spectroscopic technique of photoelectron angular streaking has successfully proven how to non-destructively retrieve the exact time–energy structure of XFEL pulses on a single-shot basis. By using artificial intelligence techniques, in particular convolutional neural networks, we here show how this technique can be leveraged from its proof-of-principle stage toward routine diagnostics even at high-repetition-rate XFELs, thus enhancing and refining their scientific accessibility in all related disciplines.

X-ray free-electron lasers (XFELs) are the world's fastest X-ray cameras, providing ultrashort exposure times in combination with a spatial resolution limit down to the sub-nanometer range, which allows for time-resolved experiments '*freezing*' the motion of atoms and molecules. In fact, XFELs have revolutionized several fields of science enabling us to observe the role of transient structures and resonances in atoms[1] as well as single-molecule or cluster imaging[2], investigations of ultrafast processes at element-specific observer sites[3], and the study of nonlinear light–matter interaction in the X-ray regime[4].

Over the past decade, further development of the underlying machine operation techniques has enabled increasingly sophisticated control over the photon pulse parameters. One of the most recent major upgrades is the increased repetition rate of XFELs that is anticipated to initiate a leap from proof-of-principle experiments to advanced applications of interdisciplinary importance, thus representing a cornerstone of modern XFEL science[5].

Most of the FELs and in fact all XFELs worldwide are currently based on the principle of *self-amplification of spontaneous emission* (SASE)[6]. More precisely, their pulses are formed stochastically through the interplay between the relativistically accelerated electron bunches themselves and the spontaneously emitted synchrotron radiation, caused by their sinusoidal trajectories inside magnetic structures with periodically changing polarity, so-called *undulators*. This feedback interaction leads to a subsequent density modulation of the undulating electrons, resulting in bursts of ultrashort X-ray pulses with a peak brightness up to and exceeding 10^{32} $\frac{\text{photons}}{\text{sec·mrad}^2\text{·mm}^2\text{·0.1\%BW}}$ (with bandwidth BW). The amplification process generates a non-predictable time–energy structure for every single pulse, constituting one of the biggest limitations for XFEL science so far. There are currently no control mechanisms and no routinely available diagnostics in place to *directly* measure the temporal properties of these X-ray pulses. This is due to the stochastic nature of each individual XFEL pulse, making their

[1]Intelligent Embedded Systems, University of Kassel, Wilhelmshöher Allee 73, 34121 Kassel, Germany. [2]Artificial Intelligence Methods for Experiment Design (AIM-ED), Joint Lab Helmholtzzentrum für Materialien und Energie, Berlin (HZB) and University of Kassel, Berlin, Germany. [3]Institute for Physics and CINSaT, University of Kassel, Heinrich-Plett-Straße 40, 34132 Kassel, Germany. [4]Fakultät Physik, Technische Universität Dortmund, Maria-Goeppert-Mayer-Straße 2, 44227 Dortmund, Germany. [5]Present address: Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin, Germany. [6]European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany. [7]Present address: Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany. [8]These authors contributed equally: Kristina Dingel and Thorsten Otto. ✉email: kristina.dingel@uni-kassel.de

single-shot characterization necessary and precluding standard integrated methods developed for attosecond pulses from lab-based sources[7–9]. Hence, the bulk of dynamics on attosecond to femtosecond time scales occurring during the exposure to the X-rays is unfortunately only inferred via indirect pulse measurements such as spectral analysis[10] or electron beam diagnostics[11].

Recently, we have demonstrated a new technique termed *angular streaking* that is capable of retrieving the time–energy structure of all incoming SASE X-ray pulses non-destructively with attosecond resolution[12]. Besides the major diagnostic breakthrough, this generally paves the way for time-resolved and nonlinear attoscience in the X-ray regime. In fact, the onset of all structural dynamics in matter can now be studied in detail even from specific observer sites through strongly localized electrons. Fast and reliable feedback of this novel diagnostic regarding the experiment and the machine itself is of utmost importance for upcoming scientific applications at XFELs[13,14]. For high-repetition-rate XFELs such as the European XFEL near Hamburg, Germany, conventional analysis approaches fail to accommodate the enormous amount and complexity of angular-streaking data in full depth. Especially for online analysis and, ultimately, (re)active control and pulse shaping during beam times, conventional data processing methods are not suited. Therefore, several core challenges of XFELs are anticipated to be tackled by artificial intelligence (AI), in particular machine-learning (ML) techniques.

In this article, we present a machine-learning-based proof-of-concept on retrieving the full and detailed XFEL pulse temporal profile, including the pulse duration and its intensity substructure. In addition, we show that it is possible to extract temporal information on the electronic processes after photoionization initiated by the X-rays, that is a subsequent Auger decay, via the method of angular streaking paired with analysis through neural networks (NNs). Moreover, by using simulated streaking data with various degrees of instrument noise and different electron emission signatures, we demonstrate the flexibility of the NN-based online diagnostic tool for XFELs. It is thus robust against detector noise and machine fluctuations, and covering the vast majority of current and future operation modes.

Application case: Angular streaking

A long-standing goal in laser and X-ray research is to enable measurements providing both temporal and spatial real-time information about electronic and consequential structural changes on a molecular level with element-site specificity—a so-called *molecular movie*. For this, a suitable ultrashort X-ray pulse duration is one of the key parameters, which is both hard to facilitate and difficult to measure. Yet, a reliable time-resolving experimental method is essential for determining parameters such as the detailed intensity profile of the SASE FEL pulse[15], corresponding damaging thresholds for materials under investigation[16], the nanoscale interpretation of ultrafast single-shot diffraction imaging[17], and the probabilities for multi-photon processes[18,19], to name a few. Applying the *angular-streaking technique*[20] to the field of XFELs leverages a versatile approach for the temporal and spectral characterization of individual (X)FEL pulses[12].

The applied scientific instruments for this new method are *angle-resolving electron spectrometers*[12,13]. In case of the first demonstration of angular streaking at XFELs and also in the present case, 16 individually working time-of-flight (TOF) spectrometers are arranged in a ring-like structure around the target region, perpendicular to the propagation direction of the incoming X-rays[12]. Together with a co-propagating circularly polarized infrared laser, spatially and temporally overlapped with the XFEL at the target region, this setup enables angular streaking. Atoms from a target gas are ionized with the XFEL pulse and the emitted electrons are swept, i.e., streaked, in energy *and* angle by the concomitant rotating electric field vector of the infrared laser. In a simplified picture, the streaking field vector can be understood as the hand of a clock that encodes the parameter time via the angles at which electrons are detected with accordingly shifted energies (see illustration in Fig. 1). Given sufficiently many electrons to "report" on their ionization time within the SASE pulse, the measured electron emission patterns contain the information of the full time–energy structure of the ionizing XFEL pulse with attosecond resolution. The method can be adapted for pulses with different photon energy by selecting target gases of suitable electron binding energies and photoionization cross sections. The mechanism and experimental setup for the angular-streaking technique as applied to SASE X-ray pulses is described by Hartmann et al.[12] and the general principle can be found here[21–23].

In the experiment under consideration, a time trace is measured for photoelectrons emitted by each X-ray shot and in each TOF spectrometer, hence, generating 16 traces at a rate set by the repetition frequency of the XFEL and of the overlapped streaking laser. For single-shot spectroscopy, a trace represents the number of electrons arriving after specific flight times. These time-domain traces can be converted to the energy domain (*spectra*) by taking into account the length of the flight path and the actions of additional electric fields along their paths, which are routinely used for enhancing the achievable energy resolution and the electron collection efficiency. The combined representation of a full angle-resolved streaking measurement forms an image with 16 columns, representing the respective detector angles, and several rows corresponding to the range of electron energies detected in the specific measurement (*detector image*, cf. Fig. 2a). Time-dependent electron spectra (*spectrograms*) are then generated by converting the emission angles to times using the known rotation period of the electric field vector for the circularly polarized streaking laser. (cf. Fig. 2c and d). In the present ML case study, we have simulated the spectrograms and the according detector images based on the equations for streaking previously derived[22], following the procedure established and described in detail in the SI of Hartmann et al.[12]. Thus, we can at will generate both a huge "data" set for training the ML algorithms, as well as an additional set of completely known target shots for testing the developed NN predictions.

For an X-ray pulse with no temporal and spatial overlap in the interaction region with an external streaking field (*unstreaked shot*), the spectra in all detectors are showing the characteristic electron energy distributions (*spectral lines*) for the target under investigation. Typically each line also shows an angular dependence in signal intensity. In Fig. 2f and g one can see this variation in the low intensity regions around 0° and 180° in contrast
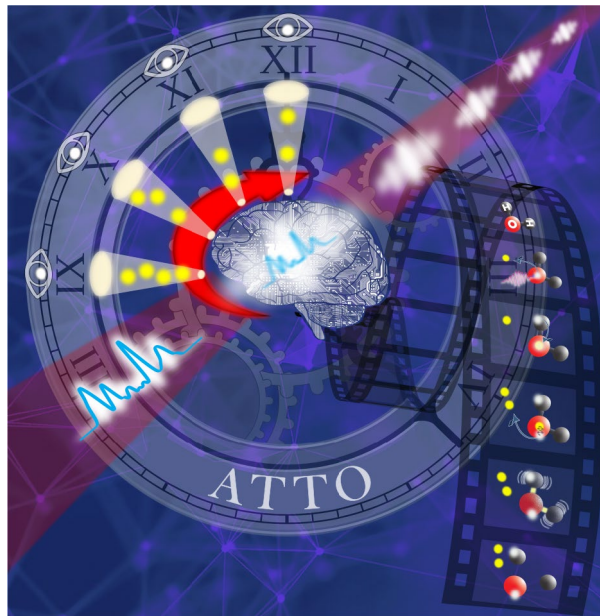
Figure 1. Illustration of this work's topical orientation. A series of previously unknown XFEL pulses releases electrons (indicated in yellow) from a gas target with characteristic kinetic energies. Their kinetic energy shifts caused by a circularly polarized infrared streaking field are recorded in an angle-resolving spectrometer (indicated by the cones and eyes). In the artificial brain, the thus obtained information about the electrons is processed and reveals the underlying attosecond substructure of the X-ray pulses. The precise knowledge of the time–energy substructure can then be used to shape the X-ray pulses via a feedback loop to the XFEL machine settings or to enable access to ultrafast electron dynamics at the attosecond frontier, as here indicated by a molecular movie of an interatomic Auger decay in water with subsequent dissociation.

to the high-intensity parts at 90° and 270°, with intermediate intensities in the columns at angles in between. If a *circularly polarized streaking laser* is present, the detector image is modulated according to the instantaneous streaking laser vector potential, leading to a sinusoidal variation of the spectral lines along the angle axis.
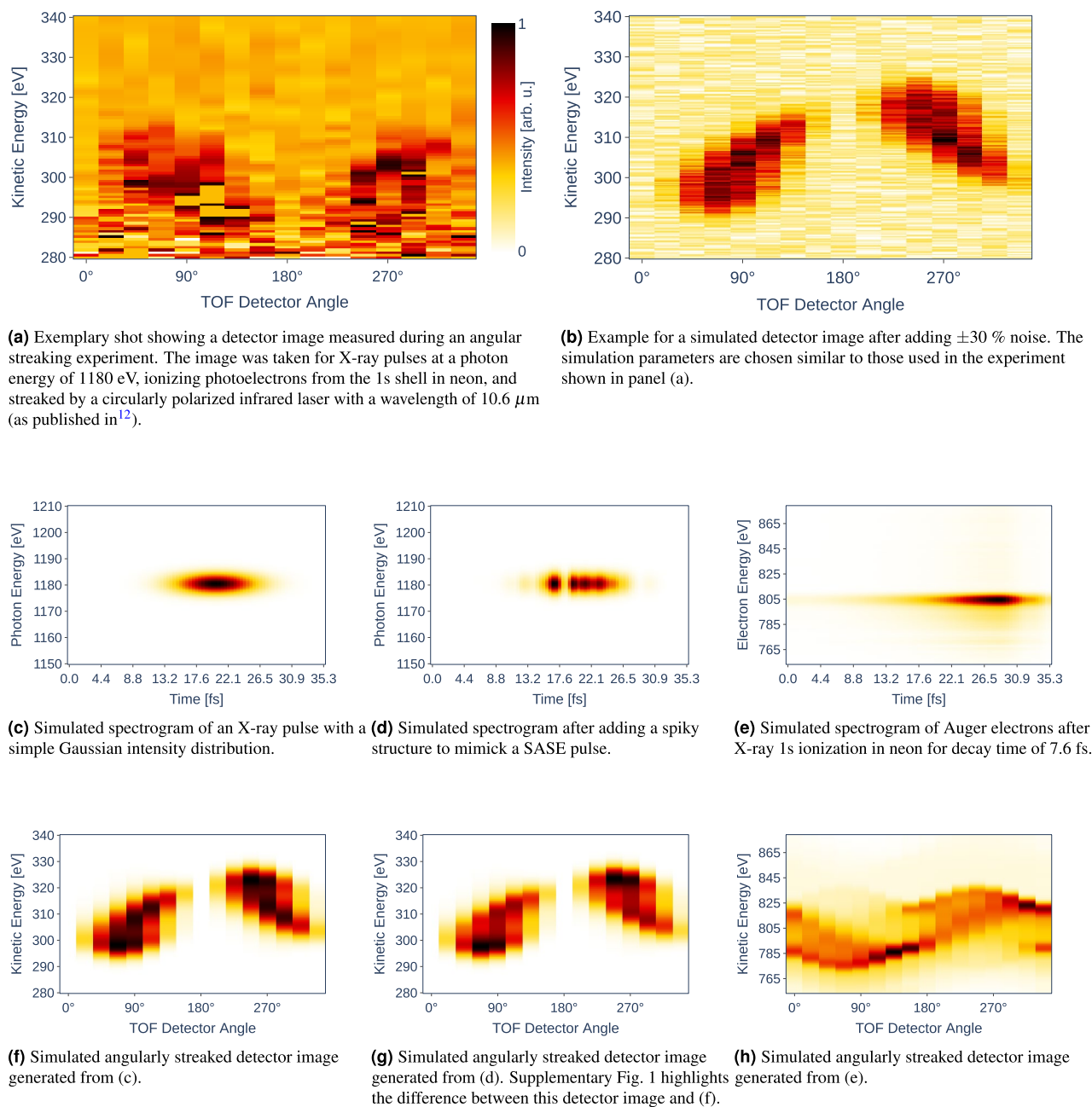
The goal regarding SASE FEL X-ray pulse characterization and their potential control is to reconstruct the spectrogram from a measured detector image, which gives the full information about the X-ray time–energy structure (cf. Fig. 3, dotted line). In many experimental situations, however, it is sufficient to restrict our analysis to some of the *most relevant* SASE X-ray parameters (cf. Fig. 3, red line). In the subsequent discussion, we have, therefore, focused on NN predictions of the temporal aspects of ultrashort FEL pulses. Details about the NNs' framework conditions, the chosen architecture, and hyperparameter optimization can be found in the "Methods" below.

We picked the following pulse characteristics for a comparison of their reconstruction by the NNs (*prediction*) with the originally simulated data (*target*):

Kick. The kick is the maximum streaking shift in electron kinetic energies for each X-ray shot, and thus for a given temporal delay and phase relation between the X-ray pulse and the infrared streaking laser. There are two main reasons for a change in the kick from shot to shot. The first is the *relative timing jitter* between the X-ray and the streaking pulse[24–26], which is unavoidable due to the stochastic generation process of the SASE mechanism and additional fluctuations in arrival time caused by air fluctuations, thermal expansion in optomechanical components, and general synchronization errors between the two separate laser pulses. The second reason for variations of the kick is the *random change of the carrier–envelope phase* of the streaking laser from shot to shot. One can solve this by stabilizing the carrier–envelope phase[27], which is a rather difficult technical requirement, or by using the technique of angular streaking[12], which is the basis for the simulations studied in this article.

Since we translate a temporal distribution (X-ray intensity structure) into an energy distribution (kinetic energy of the streaked photoelectrons), a lower kick value means a more shallow gradient of the streaking ramp, which is given by the kick over the cycle period of the electric field corresponding to the streaking laser wavelength. Thus, the resolution of the measurement is directly degrading with decreasing kick. In this sense, the determination of the kick strength is not so much interesting in itself, but is a measure for the quality of the reconstruction and can be used as a filtering handle of the data. It is also a good consistency check for the functioning of the applied NNs, since the kick is a parameter that can also be readily assessed with other analysis methods.

Pulse duration. The pulse duration is the most important parameter for many ultrafast free-electron laser experiments, e.g., a variety of pump/probe measurements of electronic state changes or investigations of nonlinear excitation dynamics [19,29–31], albeit it is one of the most difficult to measure directly. Especially for XFEL SASE

(a) Exemplary shot showing a detector image measured during an angular streaking experiment. The image was taken for X-ray pulses at a photon energy of 1180 eV, ionizing photoelectrons from the 1s shell in neon, and streaked by a circularly polarized infrared laser with a wavelength of 10.6 μm (as published in[12]).

(b) Example for a simulated detector image after adding ±30 % noise. The simulation parameters are chosen similar to those used in the experiment shown in panel (a).

(c) Simulated spectrogram of an X-ray pulse with a simple Gaussian intensity distribution.

(d) Simulated spectrogram after adding a spiky structure to mimic a SASE pulse.

(e) Simulated spectrogram of Auger electrons after X-ray 1s ionization in neon for decay time of 7.6 fs.

(f) Simulated angularly streaked detector image generated from (c).

(g) Simulated angularly streaked detector image generated from (d). Supplementary Fig. 1 highlights the difference between this detector image and (f).

(h) Simulated angularly streaked detector image generated from (e).

Figure 2. Experimental (**a**) and simulated (**b**)–(**h**) data for an angular streaking measurement. (**a**) shows real data from Hartmann et al.[12] in neon. (**b**) displays a simulated detector image under similar conditions with the addition of ±30% noise. In [(**c**) and (**f**)] a simulated spectrogram and the corresponding detector image, respectively, are shown before and after [(**d**) and (**g**)] adding a pulse structure to a Gaussian intensity distribution. Note the very slight differences between (**f**) and (**g**), which are caused by the additional SASE pulse structure and give rise to an accordingly spiky pulse reconstruction. (**e**) and (**h**) display a simulated Auger electron spectrogram and the respective angularly streaked detector image, after ionization of the Ne 1s shell with an arbitrarily chosen decay time of 7.6 fs. All displayed figures are normalized to the interval [0, 1].

pulses, each pulse has a different duration and erratic intensity structure that even complicates the definition of the term *pulse duration*. In this article, we use the *root-mean-square (RMS) duration*, i.e., the square root of the time variance of the temporal intensity profile [32],

$$t_{\mathrm{p,RMS}} = \sqrt{\langle t^2 \rangle - \langle t \rangle^2}, \tag{1}$$

where

4
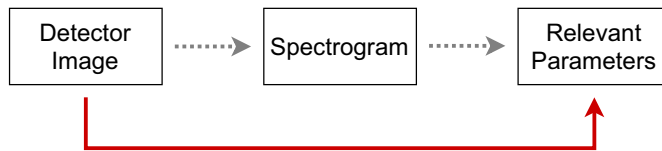
Figure 3. There are two ways to extract important pulse characteristics from the detector image. The dotted line refers to a full reconstruction (spectrogram) and the extraction of relevant parameters out of this reconstruction. The other way is to skip the full reconstruction and to concentrate on typically most relevant parameters only. In this article, we choose the second approach.
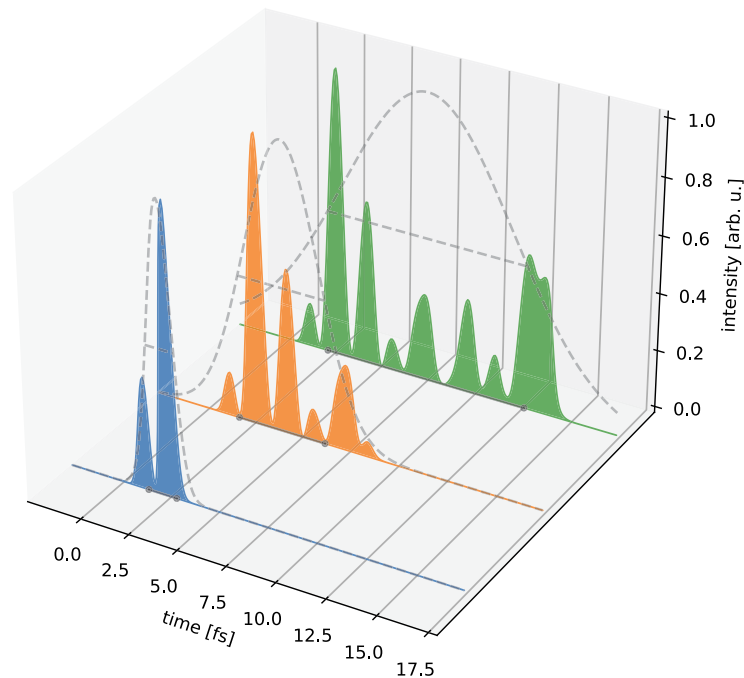
Figure 4. Three examples of different SASE XFEL pulse intensity structures showing varying total duration and complexity generated with OCELOT[28]. The pulse durations given as FWHM/RMS, respectively, for the exemplary shots are 1.5 fs/640 as (blue), 4.5 fs/1.9 fs (orange) und 10.4 fs/4.4 fs (green). The corresponding Gaussian pulse envelopes are shown as dashed grey curves, including the FWHM durations, and their projections onto the time axes as black lines.

$$\langle t^n \rangle = \frac{1}{N} \int_{-\infty}^{\infty} t^n I(t) \; dt \quad \text{and} \quad N = \int_{-\infty}^{\infty} I(t) \; dt \tag{2}$$

are the n-th moment and the normalization constant, respectively, as the basic definition of the *pulse duration*.

A common choice for more well-behaved Gaussian-like laser pulses from table-top systems is the *full width at half-maximum (FWHM)*. Given a Gaussian distribution with standard deviation σ, corresponding to the RMS duration in this case, the FWHM is calculated as follows:

$$FWHM = 2\sqrt{2 \ln 2}\sigma \approx 2.35 \cdot \sigma. \tag{3}$$

As SASE pulses are generally spiky and irregular (cf. Fig. 4), this metric is not fully applicable. In our simulation case, however, this quantity is nevertheless of interest due to the fact that we use the FWHM to generate (and in fact define) the Gaussian distribution envelopes in Fig. 4 and because the FWHM better relates to the intuitive concept of a full-length pulse duration. The RMS duration, however, gives a more complete measure of the temporal distribution of the pulse energy including possible pulse wings[32] or substructure (see also the next paragraph).

Pulse structure. Due to the microbunching in the FEL each SASE pulse has an individual intensity profile, made up of several shorter 'spikes' with random intensities (cf. Fig. 4). The average number of spikes per pulse is determined by the specific operation parameters of the XFEL. It can be expressed in a statistical treatment as the number of individual energy modes contributing to the XFEL pulse[33]. The ensuing pulse shape can be arbitrarily complex. The shorter the overall pulse duration in relation to the single-spike length, i.e., the fewer spikes per complete pulse, the more important individual spikes are becoming (Fig. 4). Especially for estimating the damage thresholds of investigated probes as well as for experiments sensitive to the instantaneous X-ray intensity, or for ultrafast pump/probe measurements, the XFEL pulse structure needs to be known exactly to interpret the observed data on a shot-to-shot basis unambiguously.

Auger decay time. Many of the scientifically interesting processes of non-equilibrium physics and structure-changing chemistry are not directly triggered by the exciting X-ray pulse but are the result of subsequent complex relaxation dynamics. These dynamics are determined by the time-dependent, i.e., transient electronic structure of the system under study. One of the most fundamental electronic processes after inner-shell ionization of matter by X-rays is the *Auger decay*, whereby a second electron from an outer shell fills the generated core hole and transfers the excess energy to a third electron (*Auger electron*), which is then emitted from the ion. This process is specific to the contributing discrete electronic states of an atomic or molecular system and has a characteristic time constant for the emission of the third electron (*Auger decay time*). In our simulations, we assume that one Auger decay channel dominates for neon (Ne) after 1s ionization. The corresponding Auger decay time on the order of 2–3 fs[34] can serve as a fundamental benchmark for demonstrating the capability of the method to retrieve ultrafast timing information from recorded data.

Results

All of the above described SASE XFEL pulse characteristics can be predicted with varying degrees of accuracy by utilizing convolutional NNs. For each pulse characteristic, we will examine the results of the trained models in more detail.
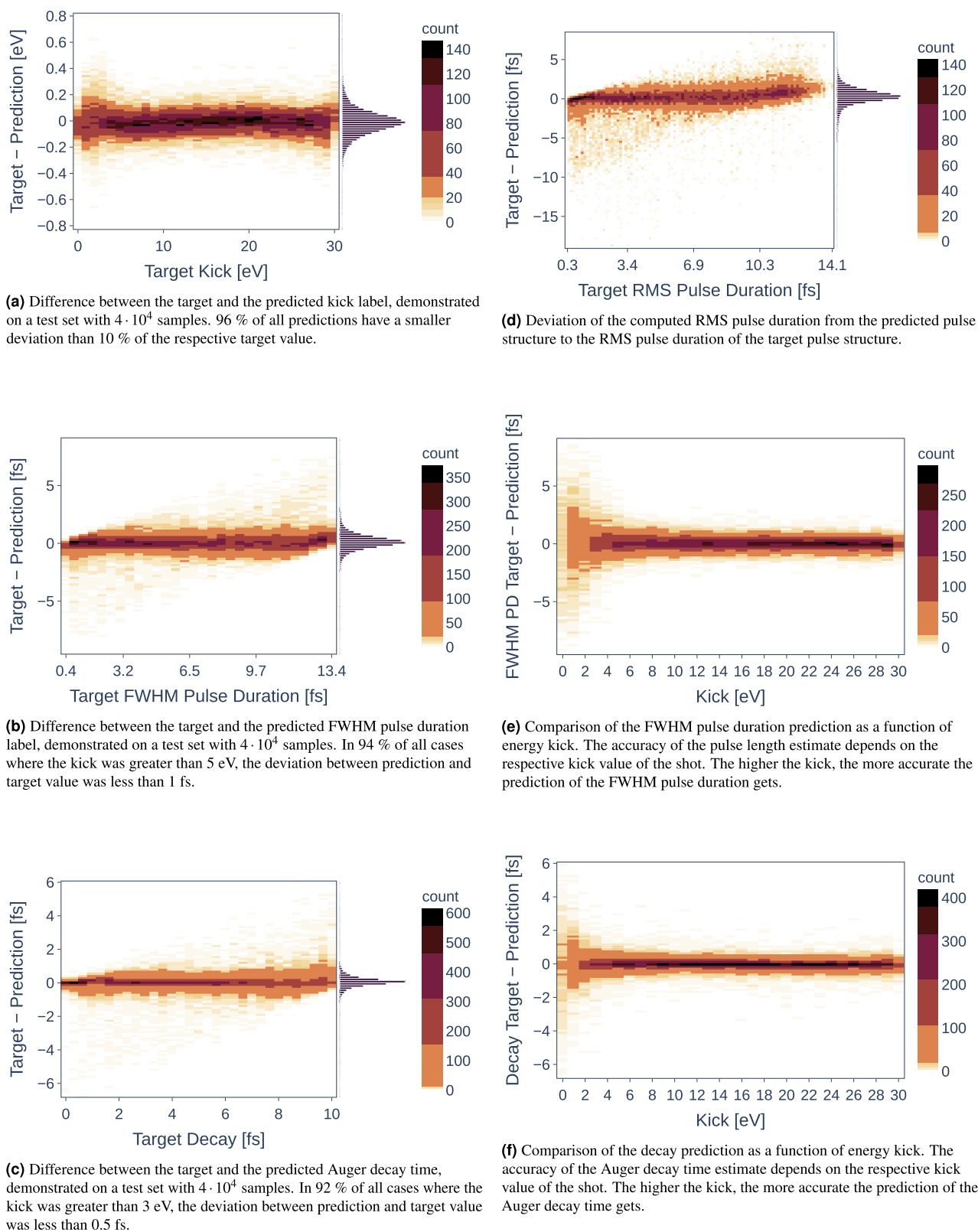
Kick. Of all the characteristics studied, the kick turned out to be the easiest to predict. Figure 5a shows that most of the predictions only slightly deviate from the respective targets. It turns out that 96% of all predictions have a smaller deviation than 10% of the respective target value. Though the kick can easily be derived from the detector images, an accurate estimate of this parameter is necessary for better judging the reliability of the reconstruction for the *FWHM pulse duration*, *pulse structure* and *Auger decay time*. This will become apparent in the following paragraphs.

FWHM pulse duration. The comparison of predicted and target FWHM pulse durations is shown in Fig. 5b. As for the kick estimates, the majority of the values are well reconstructed. However, it is evident that a few of the predictions deviate strongly from the target values. One hypothesis explaining this behavior is that for smaller kicks the resolution of the measurement degrades and predicting a pulse duration can become arbitrarily difficult. That is why we have investigated the accuracy of the FWHM pulse duration estimate against the true kick value.

Figure 5e confirms the previously stated hypothesis. Above a (true) kick value of approximately 5 eV, estimating the FWHM pulse duration becomes feasible. In fact, in 94% of all cases where the kick was greater than 5 eV, the deviation between prediction and target value was less than 1 fs. This is due to the fact that small kick values on the order of the nominal SASE bandwidth correlate to unsuccessful angular streaking shots that need to be discarded anyway. Supplementary Fig. 2a and b display exemplary shots with a large and a small kick, respectively.

Auger decay time. The auger decay time can also be well approximated by the respective NN (cf. Fig. 5c). Most of the estimates hardly deviate from the zero line of the difference between the target value and the prediction. However, as for the FWHM pulse duration estimates, there are some outliers strongly deviating from the true Auger decay time value. Using the same reasoning as for the FWHM pulse duration estimates, we have compared the prediction of the Auger decay time value with the true kick value. Here, the same behavior can be observed as for the FWHM pulse duration (cf. Fig. 5f). It is evident that a reasonable determination of the Auger decay time is only possible for a kick value of 3 eV or higher. In fact, in 92% of all cases where the kick was greater than 3 eV, the deviation between prediction and target value was less than 0.5 fs. It follows that shots with small kick values should be discarded in advance to appropriately approximate the true Auger decay time. As before, Supplementary Figs. 2c and d display exemplary shots with a small and a large kick for the decay reconstruction, respectively.

Pulse structure and RMS pulse duration. The full temporal pulse structure of the SASE pulse is probably the most difficult property to predict in our study. This is not surprising, as it is also the most complex one, being represented by a vector instead of a single value in case of the kick or the pulse duration, which holds the information of the intensity distribution over time. Altogether, the trained network works relatively well in its objective to predict the trend, i.e., peak positions and their relative intensities of the pulse structure. However, as has been expected, these predictions get less accurate for more complex pulse structures. This behavior can be seen for two different exemplary simulated SASE pulses in Fig. 6, one relatively simple (Fig. 6a) and one more

(a) Difference between the target and the predicted kick label, demonstrated on a test set with $4 \cdot 10^4$ samples. 96 % of all predictions have a smaller deviation than 10 % of the respective target value.

(d) Deviation of the computed RMS pulse duration from the predicted pulse structure to the RMS pulse duration of the target pulse structure.

(b) Difference between the target and the predicted FWHM pulse duration label, demonstrated on a test set with $4 \cdot 10^4$ samples. In 94 % of all cases where the kick was greater than 5 eV, the deviation between prediction and target value was less than 1 fs.

(e) Comparison of the FWHM pulse duration prediction as a function of energy kick. The accuracy of the pulse length estimate depends on the respective kick value of the shot. The higher the kick, the more accurate the prediction of the FWHM pulse duration gets.

(c) Difference between the target and the predicted Auger decay time, demonstrated on a test set with $4 \cdot 10^4$ samples. In 92 % of all cases where the kick was greater than 3 eV, the deviation between prediction and target value was less than 0.5 fs.

(f) Comparison of the decay prediction as a function of energy kick. The accuracy of the Auger decay time estimate depends on the respective kick value of the shot. The higher the kick, the more accurate the prediction of the Auger decay time gets.

Figure 5. Prediction accuracies and dependencies of the labels *kick*, *RMS pulse duration*, *FWHM pulse duration*, and *decay* using data sets containing all levels of noise [$\pm 0\%, \pm 10\%, \pm 20\%, \pm 30\%$]. The slight skew of the distributions in Figs. (**b**), (**c**) & (**d**) stems from the inherent tendency of NNs to predict values closer to the mean of the learned parameter space for unclear inputs, i.e., from detector images with too low kicks.

(a) Simple SASE pulse structure.

(b) Complex SASE pulse structure.
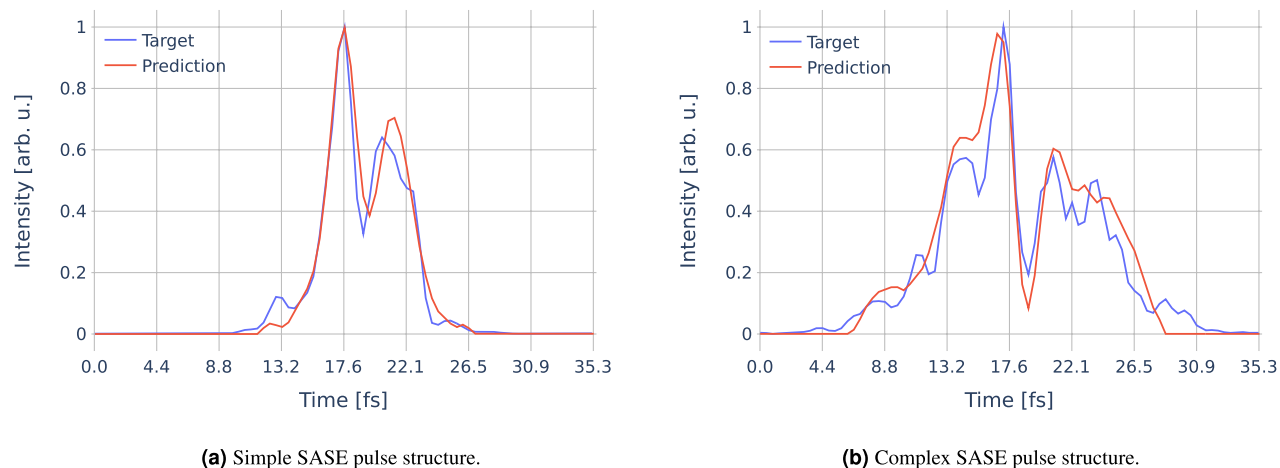
Figure 6. Examples for simple (**a**) and more complex (**b**) simulated and reconstructed SASE pulse structures with RMS pulse durations of 6.2 fs and 11.7 fs, respectively.

Noise	Kick	FWHM	RMS	Decay
±0%	0.19 eV	0.52 fs	0.67 fs	0.08 fs
±10%	0.24 eV	0.88 fs	0.65 fs	0.17 fs
±20%	0.34 eV	1.05 fs	0.65 fs	0.26 fs
±30%	0.43 eV	1.2 fs	0.73 fs	0.34 fs

Table 1. Standard deviations of the predicted labels with regard to the respective targets kick, pulse duration, and Auger decay computed on 1000 samples, respectively. Settings for the kick, FWHM/RMS pulse duration, Auger decay sample data: *kick* = 22.5 eV, *pulse duration* = 4.85 fs, *Auger decay* = 7.0 fs.

complex (Fig. 6b), in which not all of the finer structures could be reliably reproduced. Nevertheless, the main features including the larger peaks can always be predicted.

Remarkably, the quality of the predicted pulse structure does not show a significant dependency on the value of the kick; except for a kick very close to or equalling zero, which is not surprising, as this again corresponds to an unsuccessful event where basically no streaking occurred. There is also no significant dependency on the duration of the pulse, as one might have expected. An absolute value for the mean squared error (MSE) does indeed increase with the pulse duration; normalized to it, however, the average 'MSE per time step' is more or less constant (cf. Supplementary Fig. 3). Now that our model is capable of extracting the pulse structure, it is possible to compute the RMS pulse duration by using Eq. (1). Figure 5d shows the deviation of this computed RMS pulse duration to the RMS pulse duration of the target pulse structure. The average deviation lies below 1 fs. Only for very long pulse durations there is a trend toward a slight underestimation, although the error remains just around 10% in most cases. We note that an additional NN for the direct prediction of the RMS pulse duration could serve as a comparative measure for the RMS value calculated from the pulse structure, allowing a coarse estimation of the quality of the reconstructions.

Influence of noise on the NN performance. In order to investigate the influence of noise on the predictions, we generated a test set comprising of detector images with several different noise levels ($p = [0.0, 0.1, 0.2, 0.3]$) as shown in Eq. (4) in the "Methods" section and fed it into the NNs.

As expected, additional noise influences the result of the NN's prediction (cf. Table 1). An example for a noisy simulated detector image is given in Fig. 2b, more expressive examples for different simulation settings are shown in Supplementary Fig. 4. The prediction for non-noisy data is nearly perfect, whereas the prediction for noisy data slightly differs from the target. Despite the decreased prediction accuracies, it is evident that the NNs can handle noise robustly.

Discussion and outlook: Online SASE-pulse characterization and shaping

So far, we have shown that several characteristics of XFEL pulses are predictable with varying degrees of accuracy. To investigate how close the current status comes to real-time pulse characterization during experimental campaigns, we need to address a number of different issues, which can each be tackled exploiting the specific analytical strength made accessible by the methods of NNs:

Output speed. For evaluating the input images at full speed of the XFEL repetition rate in the kHz–MHz regime, an efficient analysis is inevitable. NNs are known for delivering outputs quickly. We have used several

BS	1	64	128	256	512	1024	2048	4096
Dur [ms]	1.38	1.44	1.48	1.54	1.62	1.54	1.38	1.44

Table 2. Time measurements for predictions of the trained model on a GeForce RTX 2070 card with different batch sizes (BS). We ran hundred experiments and averaged the results.

batch sizes, starting from one image up to 4096 as input. Investigating this is important as such a comparison determines whether a batch-wise and therefore highly parallelized analysis is performing better than analyzing image-wise. Batch-wise evaluation is specifically suited for the European XFEL facility, since a train with a very fast succession of pulses (maximum 2700 per 600 µs) is followed by a pause of several milliseconds that can be used for analysis purposes[5,35]. We have tested how fast the NN output is generated on a GeForce RTX 2070 GPU using single precision floating point format (Table 2).

The model is able to reach quick predictions mostly independent of the batch size as the computation on a GPU runs all tasks, i.e., computes a prediction for each image within the batch, in parallel. In general, the number of input images in one batch is only limited by the RAM of the used GPU. Thus, it is apparent that it is advantageous to analyze a larger batch of data than individual images. With a batch size of 4096 our current model is already able to keep up with the European XFEL in high-repetition mode for online predictions.

Reliability estimation. Next to fast evaluation, a degree of certainty in the NN predictions must be ensured. As shown in the results, the NN prediction may deviate quite substantially from the target. Some of the difficulties can be directly circumvented. By determining the kick, for example, we can already filter whether a prediction regarding the labels *Auger decay time* or *pulse duration* is reasonable. But this still does not give us a direct statement specifying how certain the NN's prediction is. Optimally, we would want to have a reliable measure of how good the predictions of the trained models are, even for unknown shots without a target.

There are several ways to determine the prediction uncertainties of NNs. The epistemic uncertainty determines the uncertainty due to insufficient knowledge. This can be implemented, for example, via Monte Carlo dropout[36] or Monte Carlo batch normalization[37]. The aleatoric uncertainty determines the uncertainty due to the complexity of the problem, and can be investigated by creating a fitted cost function[38]. This topic is currently a very active area of ML research. We are in the process of developing our own approach to benchmark the specific task of stochastic X-ray pulse reconstruction, ideally combining both uncertainty determinations in one procedure as previously shown[38].

Gap between simulation and reality. So far, our NNs are suited only for data that look exactly like the input shown in Fig. 2. Whether the NNs are suitable for predictions on experimental data (cf. Fig. 2a) is not easy to validate, especially since we do not have true labels for the experimental data. In addition, we need to identify to what extent our modeled noise replicates real noise of the spectrometer, e.g., electronic ringing of the detector readout or background signals from undesired processes. There are two ways to tackle the gap between simulation and experiment. Either the real data must be denoised before the analysis (e.g., Denoising Autoencoders[39]) or the simulation data must be provided with additional, appropriately modeled noise. Simultaneous approaches in both directions should give a more complete understanding for mitigating this issue in future efforts.

Responding to changes. We have shown that our developed NNs work on data with several levels of noise. However, when utilizing real-life TOF spectrometers, it may occur that TOF sensors fail or produce unrealistic results. In such cases NN re-training or knowledge extension is inevitable. Here, online learning[40] is a helpful tool. In this case, the model is trained continuously on newly generated data. Thus, the training can be quickly adapted to new environments. To circumvent the catastrophic forgetting[41] of NNs, continual learning[42,43] may be utilized.

Pulse shaping. The term pulse shaping can be interpreted in two technically different ways. With the online analysis methods already demonstrated in the current manuscript, we can establish an X-ray sorting scheme based on the full characterization of every single pulse and the possibility to filter for desired pulse shapes and durations. Especially with high-repetition rate XFELs, this can be a vital form of "passive shaping". The second, and more exciting, route refers to an actual pulse shaping in terms of intelligent experimentation schemes and dynamical interaction of the machine with simultaneous photon-based measurements. Thus, the live updates on X-ray pulse changes may be used for a more detailed control of the parameters and for actual SASE pulse shaping. First steps for this interaction have been identified to entail a feedback loop to the accelerator that provides an online data stream of, e.g., X-ray pulse duration and spectrum. The machine operators can choose to engage this loop into their electron bunch compression algorithms with a preset goal of optimization to be pursued. Further steps towards intelligent and active experimentation are in development and will be presented in future studies.

Conclusion. In this article, we demonstrated a path toward online characterization of free-electron laser pulses by applying NNs on detector images captured with angular streaking. In addition to several predictable characteristics, we have been able to identify and confirm dependencies between the respective characteristics that can be used to control the machine settings during experimental campaigns. This way, the angular streaking

Figure 7. We use a simulated spectrogram to construct according detector images. These images are used for training the NNs and to extract several different pulse characteristics afterward.

technique has the potential to be leveraged from the proof-of-principle stage to a robust and highly advanced diagnostic tool for all free-electron laser facilities, including high-repetition rate operation. In addition, these novel ML reconstruction procedures may also be used for better online X-ray pulse control and future FEL pulse shaping on demand. Further steps to a successful implementation of these advanced methods involve closing the gap between simulation and experimental data through an instrument-specific treatment of the measurement noise and a reliable concept for error and reliability estimation, which we will investigate in future work.

Methods: Machine learning procedure design

In real-world XFEL experiments, the spectrograms or pulse characteristics of individual SASE pulses have to be reconstructed from the detector image. There are first approaches for deriving single-shot characteristics of rapid pulse sequences from high-repetition rate XFELs[44]. Unfortunately, they are only suitable to a limited degree for providing detailed insights via real-time processing during experimental campaigns.

Here, we apply specifically developed NNs on the angular streaking approach to demonstrate the possibility of a fast online pulse characterization, as NNs, particularly convolutional NNs, have proven to be suitable for similar challenges[45,46].

General machine learning problem formulation. For each pulse characteristic, we need to train a NN that takes detector images as inputs (cf. Fig. 7). The outputs for each of the NNs vary and are listed below:

Kick. The kick is the amplitude of the wave-like intensity distribution within the detector image (cf. Fig. 2). When changing the kick, the spectrogram stays as is as the kick only affects the streaking signal captured in the detector image. That is the reason why the kick is easily extractable from the detector images. The NN has to solve a regression task, where the output is one number in the unit eV.

FWHM pulse duration. The FWHM pulse duration is well extractable from the spectrogram, as it can be seen as $2.35 \cdot \sigma$ (cf. Eq. 3) in the direction of x (time scale), with σ being the standard deviation of the 2D gaussian distribution in x direction. The longer the FWHM pulse duration, the longer the distribution stretches in x direction. Within the detector image, a change of the pulse duration mostly affects the width and the peculiarity of the wave form. Here again, the NN has to solve a regression task, where the output is one number in the unit fs.

Auger decay time. The Auger decay is visible in the spectrograms as well as the detector images. Within the spectrogram, the length of the tail after the 2D Gaussian distribution indicates the decay time (cf. Fig. 2e). The longer the tail, the larger the decay. Within the detector image, a larger decay affects the distortion of the wave (cf. Fig. 2h). Once more, the NN has to solve a regression task, where the output is one number in the unit fs.

Pulse structure. The pulse structure is the most challenging feature to extract, as the output itself consists of several values indicating the intensities of multiple spikes within the SASE pulses. By looking at the spectrogram in Fig. 2c, one can see that the pulse structure can be derived by summing up the intensities at each point in time along the vertical axis. The pulse characteristic will be determined here as the intensity as a function of arrival time, where the intensity is integrated over all photon energies within the 6 eV spectral bandwidth. This leads to an output similar to Fig. 4. In this case, the NN has to solve a regression task, where the output consists of several time steps in arbitrary intensity units.

A note regarding the *RMS Pulse Duration*: As the RMS pulse duration can be directly derived from the pulse structure, there is no need to train an independent NN for this pulse characteristic.

After the general examination of the ML problem, the next sections will look at how the ML pipeline looks in detail and how to successively address the individual ML problems above.

Framework conditions. In order to train NNs in a supervised manner, we require training data $\mathscr{D}_K$ of size $K \in \mathbb{N}$, which comprises K simulated detector images $\mathscr{X} = \{\mathbf{X}_i \in \mathbf{M}^{m \times n}(\mathbb{R}), \ i = 1, \ldots, K\}$ and K corresponding pulse characteristics $\mathscr{L} = \{L_i \in \mathbb{R}^j_+, \ i = 1, \ldots, K\}$. Here, m is the number of TOF detectors used within the complete spectrometer setup and n displays the electron kinetic energy in intervals. The size of j changes according to the pulse characteristic that has to be predicted. In the following, we will refer to pulse characteristics as *labels*. To verify the performance of a NN, we split $\mathscr{D}_K$ into two distinct sets, $\mathscr{D}_{train}$ and $\mathscr{D}_{test}$, such that $\mathscr{D}_K = \mathscr{D}_{train} \cup \mathscr{D}_{test}$. The accuracy of the NN is determined by $\mathscr{D}_{test}$. We train the NNs with n distinct batches of detector images $\mathscr{B}_n \in \mathscr{D}_{train}$. Similarly, we test the performance of the NNs with m batches of detector images $\mathscr{B}_m \in \mathscr{D}_{test}$. To avoid overfitting of the NN, we utilize *cross-validation*[47].

Although we work in a simulation environment, it is reasonable to choose values that would correspond to real experimental data (cf. Fig. 2a). Therefore, we take previously acquired data from earlier experimental campaigns[12] as an example. In particular this means:

- For each of the two use cases, Ne 1s and KLL Auger electrons data, we generate a size of $K = 4.4 \cdot 10^6$ samples to be predicted. Of these, $4 \cdot 10^6$ are used for training and $4 \cdot 10^5$ for testing.
- Our angle-resolved spectrometer consists of $m = 16$ TOF detectors.
- We fix the intervals in the TOF detectors to $n = 200$, with a varying energy bin size.

More particularly, this means that the following NN architecture depends on the chosen parameters, though it can be adapted easily if, e.g., more TOF detectors are added.

Preparing the simulation data.   We derive artificial detector images for Ne 1s and KLL Auger electrons from the simulation environment as introduced by Hartmann et al.[12]. The kinetic energy of the 1s photoelectrons depends on the ionizing X-ray photon energy, which, in this case, is set to 1180 eV. We include a spectral bandwidth of 6 eV for the X-ray pulses, but omit the effect of a potential chirp in these simulations, which would only have a marginal effect on the parameters reconstructed in this study. A photon energy of 1180 eV results in Ne 1s photoelectron kinetic energies centered at ~ 310 eV. The Auger electron kinetic energies are independent of the X-ray photon energy and bandwidth, with the main peak lying at ~ 804 eV and a standard deviation determined by the detector resolution. The angular streaking maps the TOF measurements of 16 detectors distributed over 360° to a window of 35.3 fs, as this is the duration of one optical cycle for the chosen streaking wavelength $\lambda = 10.6$ μm of the circularly polarized laser in Hartmann et al.[12].

We chose the range and the precision based on the experimental implementation expected for real streaking measurements at XFELs. We want our models to estimate kicks in the range of 0–30 eV, FWHM pulse durations in the range of 0.4–1.34 fs, and decays in the range of 0–10 fs. The temporal resolution of the pulse structure was equally chosen in accordance with the expected duration of the shortest features in the X-ray pulse intensity structures (the SASE spikes), leading to a grid size along the time axis for the pulse structure reconstruction of 441 as.

Figures 2c and f are simulated without artifacts. In Fig. 2c, there is only one randomly placed Gaussian distribution present in the spectrogram. The underlying pulse structure is neglected so far. To get closer to the real data (cf. Fig. 2a), we implement three steps. We add a pulse structure to the spectrogram and noise to the simulated detector images, and prepare the data for NN training by utilizing data normalization.

Step 1: Adding a Pulse Structure to the Spectrogram: To achieve a SASE-like temporal structure in the spectrogram, we modulate the original Gaussian time distribution with a spiky intensity profile (cf. Fig. 4). We obtain the latter by generating a comb of Gaussian spikes with randomized amplitudes and spike durations as predicted by theory for a typical setting of an XFEL in ultrashort-pulse mode[33].

Step 2: Adding Noise to the Detector Image: Additional noise is added to $\mathbf{X} \in \mathscr{X}$, representing the intensity values for each pixel in the detector image, during training and testing as shown in Eq. (4). A given percentage p of the maximum intensity value $x_{\max}$ of $\mathbf{X} \in \mathscr{X}$ is used as an upper and lower bound of an equal distribution $\mathscr{G}$ to draw w from:

$$\mathbf{X}_{\text{noisy}} = \left(x_{i,j} + (x_{\max} \cdot w)\right)_{i=1,\ldots,m,j=1,\ldots,n}, w \sim \mathscr{G}\ (-p,p), \tag{4}$$

Figure 2b displays a detector image with added noise ($p = 0.3$).

Step 3: Normalizing the Data: It is evident, that the range of the intensity values differs from case to case. To counteract this, we perform a min-max-normalization for each $\mathbf{X} \in \mathscr{X}$. Therefore, the minimum ($x_{\min}$) and maximum ($x_{\max}$) intensity value of $\mathbf{X}$ are used to perform the transformation for each pixel value $x_{k,l}$:

$$\mathbf{X}_{\text{norm}} = \left(\frac{x_{k,l} - x_{\min}}{x_{\max} - x_{\min}}\right)_{k=1,\ldots,m,l=1,\ldots,n}, \tag{5}$$

After normalization, all values in $\mathbf{X}_{\text{norm}}$ lie within the interval [0, 1].

Designing the machine learning models.   As we want to extract information from images, the most intuitive solution is to use a convolutional NN[48,49], which uses convolutions and pooling to extract low- and high-level features such as edges and predicts an estimate of those features using fully-connected layers. In our case, the estimates ideally should correspond to the target pulse characteristics.

Architecture.   One key problem regarding the choice of a proper NN architecture is the dimensionality of the detector images, which are not equal in size and therefore not symmetrical. This fact needed to be taken into account in the design of the NN. Furthermore, we wanted our NN architecture to be as dense as possible to ensure generalization and performance for reliable online operation. After testing several architecture configurations with different numbers of layers and neurons, the most suitable network architecture for our problem is an NN with three convolutional blocks (cf. Fig. 8). Each block contains a convolutional layer, followed by an activation function and a max-pooling layer. The convolutional layers use a 3×3 kernel, stride of 1×1 and 1×1 zero-padding. The pooling layers use a 3×3 kernel, stride of 2×2 and 1×1 zero-padding. Architectures with less than three convolutional blocks could not grasp all required features necessary to derive the underlying mapping from detector image to the desired label. The NN is specifically designed to cut both dimensions, i.e.,
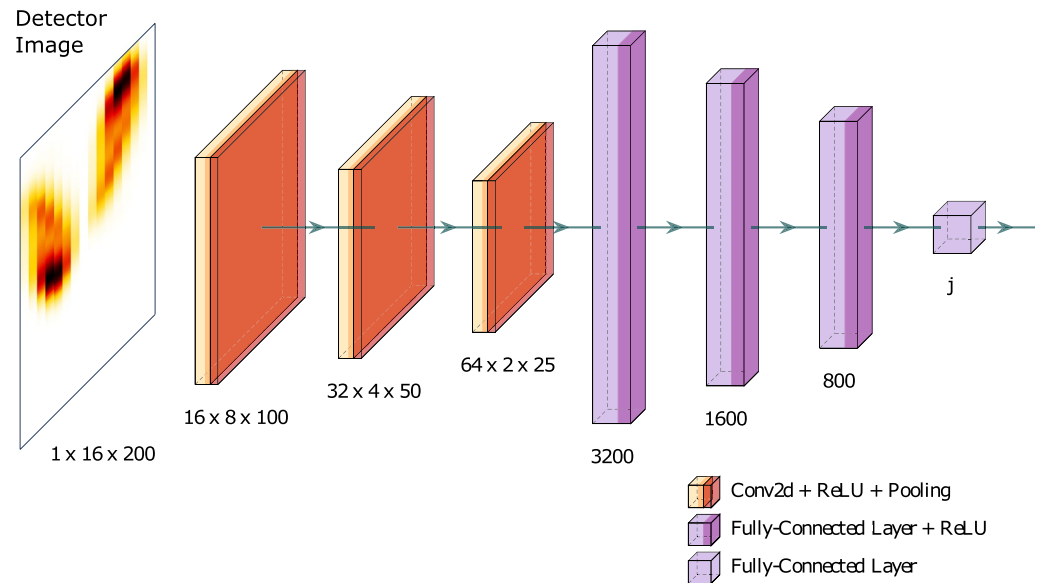
Figure 8. The convolutional neural network architecture used in the present approach. The dimensions of the first layers (orange) are displayed in [channel, height, width]. The dimensions of the fully-connected layers (violett) show the number of used neurons.

width and height of the image, in half after each block. A filter size of [16, 32, 64] for the respective convolutional layers has proven to be sufficient. For the fully-connected stage (except the last layer), we use three layers with [3200, 1600, 800] neurons, respectively. Architectures with less than three layers within the fully-connected part did not transmit enough information to be able to solve the problem decently. The size of the last layer depends on the label to predict, i.e., the size of j in $\mathscr{L}$.

When predicting the *kick*, *FWHM pulse duration*, or *decay time*, $j = 1$. When predicting the *pulse structure*, j corresponds to the dimension of the spectrogram's x-axis. We utilize the mean squared error loss function to train and optimize the network as the prediction of the pulse characteristics is a regression task in all cases.

Hyperparameter optimization. The NN architecture is not the only choice to be considered. Especially when training the NNs, appropriately chosen hyperparameters are important to achieve efficient and goal-oriented training. Important parameters in this context are the *batch size*, *type of activation function*, *optimizer*, and *learning rate*. To find the best suitable combination of hyperparameters, we performed a grid-search on distinct data sets using the approach from before with the following values:

- *Batch size*: [64, 128, 256, 512, 1024].
- *Activation function*: [ReLU, Sigmoid].
- *Optimizer*: [Adam, SGD (with Momentum)].
- *Learning rate*: [0.01, 0.001, 0.0001, 0.00001].

After NN training, we evaluated the respective parameter combinations according to the following criteria:

- *Criterion 1*: The test loss (after inputting $\mathscr{D}_{\text{test}}$ into the trained NN) should be minimal.
- *Criterion 2*: The standard deviation of the test loss curve should be minimal to penalize slow convergence and overfitting.

In general, it should be noted that there is not only one combination of hyperparameters that achieves good results during training. Nevertheless, there has been an evident leader. The best hyperparameter configuration for all labels is a batch size of 64, a ReLU activation function, a learning rate of 0.0001, and Adam as optimizer.

Data availability

A repository containing the detector image analysis software and according simulation data will be provided on request. Therefore, please contact the corresponding author Kristina Dingel (kristina.dingel@uni-kassel.de).

References

1. Mazza, T. *et al.* Mapping resonance structures in transient core-ionized atoms. *Phys. Rev. X* **10**, 041056 (2020).
2. Ho, P. J. *et al.* The role of transient resonances for ultra-fast imaging of single sucrose nanoclusters. *Nat. Commun.* **11**, 1–9 (2020).
3. Young, L. *et al.* Roadmap of ultrafast x-ray atomic and molecular physics. *J. Phys. B At. Mol. Opt. Phys.* **51**, 032003 (2018).
4. Eichmann, U. *et al.* Photon-recoil imaging: Expanding the view of nonlinear x-ray physics. *Science* **369**, 1630–1633 (2020).
5. Decking, W. *et al.* A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator. *Nat. Photonics* **14**, 391–397 (2020).
6. Milton, S. *et al.* Exponential gain and saturation of a self-amplified spontaneous emission free-electron laser. *Science* **292**, 2037–2041 (2001).
7. Tzallas, P., Charalambidis, D., Papadogiannis, N. A., Witte, K. & Tsakiris, G. D. Direct observation of attosecond light bunching. *Nature* **426**, 267–271 (2003).
8. Paul, P. M. *et al.* Observation of a train of attosecond pulses from high harmonic generation. *Science* **292**, 1689–92 (2001).
9. Kienberger, R. *et al.* Atomic transient recorder. *Nature* **427**, 817–821 (2004).
10. Serkez, S. *et al.* Opportunities for two-color experiments in the soft X-ray regime at the European XFEL. *Appl. Sci.* **10**, 2728 (2020).
11. Behrens, C. *et al.* Few-femtosecond time-resolved measurements of X-ray free-electron lasers. *Nat. Commun.* **5**, 1–7 (2014).
12. Hartmann, N. *et al.* Attosecond time-energy structure of X-ray free-electron laser pulses. *Nat. Photonics* **12**, 215 (2018).
13. Duris, J. *et al.* Tunable isolated attosecond X-ray pulses with gigawatt peak power from a free-electron laser. *Nat. Photonics* **14**, 30–36 (2020).
14. Driver, T. *et al.* Attosecond transient absorption spooktroscopy: A ghost imaging approach to ultrafast absorption spectroscopy. *Phys. Chem. Chem. Phys.* **22**, 2704–2712 (2020).
15. Bonifacio, R., De Salvo, L., Pierini, P., Piovella, N. & Pellegrini, C. Spectrum, temporal structure, and fluctuations in a high-gain free-electron laser starting from noise. *Phys. Rev. Lett.* **73**, 70–73 (1994).
16. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–7 (2000).
17. Barty, A. *et al.* Ultrafast single-shot diffraction imaging of nanoscale dynamics. *Nat. Photonics* **2**, 415–419 (2008).
18. Young, L. *et al.* Femtosecond electronic response of atoms to ultra-intense X-rays. *Nature* **466**, 56–61 (2010).
19. Rudenko, A. *et al.* Femtosecond response of polyatomic molecules to ultra-intense hard X-rays. *Nature* **546**, 129–132 (2017).
20. Eckle, P. *et al.* Attosecond angular streaking. *Nat. Phys.* **4**, 565–570 (2008).
21. Constant, E., Taranukhin, V., Stolow, A. & Corkum, P. Methods for the measurement of the duration of high-harmonic pulses. *Phys. Rev. A* **56**, 3870–3878 (1997).
22. Itatani, J. *et al.* Attosecond streak camera. *Phys. Rev. Lett.* **88**, 173903 (2002).
23. Kazansky, A. K., Bozhevolnov, A. V., Sazhina, I. P. & Kabachnik, N. M. Interference effects in angular streaking with a rotating terahertz field. *Phys. Rev. A* **93**, 013407 (2016).
24. Bionta, M. R. *et al.* Spectral encoding of x-ray/optical relative delay. *Opt. Express* **19**, 21855–65 (2011).
25. Hartmann, N. *et al.* Sub-femtosecond precision measurement of relative X-ray arrival time for free-electron lasers. *Nat. Photonics* **8**, 706–709 (2014).
26. Diez, M. *et al.* A self-referenced in-situ arrival time monitor for X-ray free-electron lasers. *Sci. Rep.* **11**, 3562 (2021).
27. Fuji, T. *et al.* Monolithic carrier-envelope phase-stabilization scheme. *Opt. Lett.* **30**, 332 (2005).
28. Agapov, I., Geloni, G., Tomin, S. & Zagorodnov, I. OCELOT: A software framework for synchrotron light source and FEL studies. *Nucl. Instrum. Methods Phys. Res. Sect. A* **768**, 151–156 (2014).
29. Rudek, B. *et al.* Ultra-efficient ionization of heavy atoms by intense X-ray free-electron laser pulses. *Nat. Photonics* **6**, 858–865 (2012).
30. Picón, A. *et al.* Hetero-site-specific X-ray pump-probe spectroscopy for femtosecond intramolecular dynamics. *Nat. Commun.* **7**, 11652 (2016).
31. Li, X. *et al.* Electron-ion coincidence measurements of molecular dynamics with intense X-ray pulses. *Sci. Rep.* **11**, 1–12 (2021).
32. Sorokin, E., Tempea, G. & Brabec, T. Measurement of the root-mean-square width and the root-mean-square chirp in ultrafast optics. *JOSA B* **17**, 146–150 (2000).
33. Krinsky, S. & Gluckstern, R. Analysis of statistical correlations and intensity spiking in the self-amplified spontaneous-emission free-electron laser. *Phys. Rev. Spec. Top. Accel. Beams* **6**, 050701 (2003).
34. Haynes, D. C. *et al.* Clocking Auger electrons. *Nat. Phys.* **17**, 512–518 (2021).
35. Pergament, M. *et al.* Versatile optical laser system for experiments at the European X-ray free-electron laser facility. *Opt. Express* **24**, 29349–29359 (2016).
36. Gal, Y. & Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of The 33rd International Conference on Machine Learning, vol. 48 of Proceedings of Machine Learning Research* (eds Balcan, M. F. & Weinberger, K. Q.) 1050–1059 (PMLR, New York, New York, USA, 2016).
37. Teye, M., Azizpour, H. & Smith, K. Bayesian uncertainty estimation for batch normalized deep networks. In *Proceedings of the 35th International Conference on Machine Learning vol. 80 of Proceedings of Machine Learning Research* (eds Dy, J. & Krause, A.) 4907–4916 (PMLR, 2018).
38. Kendall, A. & Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision?. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* 5580–5590 (2017).
39. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. & Manzagol, P.-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**, 3371–3408 (2010).
40. Fontenla-Romero, Ó., Martinez-Rego, D., Guijarro-Berdiñas, B., Pérez-Sánchez, B. & Peteiro-Barral, D. Online machine learning. In *Efficiency and Scalability Methods for Computational Intellect* 27–54 (IGI Global, 2013).
41. Kirkpatrick, J. *et al.* Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci.* **114**, 3521–3526 (2017).
42. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C. & Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Netw.* **113**, 54–71 (2019).
43. He, Y., Sick, B. & CLea, R. An adaptive continual learning framework for regression tasks. *AI Perspect.* **3**, 1–16 (2021).
44. Heider, R. *et al.* Megahertz-compatible angular streaking with few-femtosecond resolution at x-ray free-electron lasers. *Phys. Rev. A* **100**, 053420 (2019).
45. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, Vol. 28 (eds Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. & Garnett, R.) (Curran Associates, Inc., 2015).
46. Kyrkou, C., Plastiras, G., Theocharides, T., Venieris, S. I. & Bouganis, C.- S. DroNet: Efficient convolutional neural network detector for real-time UAV applications. In *2018 Design, Automation Test in Europe Conference Exhibition* 967–972 (2018).
47. Berrar, D. Cross-validation. In *Encyclopedia of Bioinformatics and Computational Biology* (eds Ranganathan, S. *et al.*) 542–545 (Academic Press, Oxford, 2019).

48. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* **25**, 1097–1105 (2012).
49. Gu, J. *et al.* Recent advances in convolutional neural networks. *Pattern Recogn.* **77**, 354–377 (2018).

Acknowledgements

Author contributions

K.D.: Conceptualization, Methodology, Software, Investigation, Writing—Original Draft, Visualization. T.O.: Conceptualization, Methodology, Software, Investigation, Writing—Original Draft, Visualization. L.M.: Methodology, Software, Investigation, Writing—Original Draft, Visualization. L.F.: Writing—Review & Editing, Visualization. A.H.: Writing—Review & Editing. S.S.: Writing—Review & Editing. A.H.: Writing—Review & Editing. G.H.: Software, Writing—Review & Editing. D.M.: Writing—Review & Editing. J.V.: Writing—Review & Editing. B.S.: Writing—Review & Editing, Supervision, Funding acquisition. A.E.: Writing—Review & Editing. M.I.: Conceptualization, Writing—Original Draft, Supervision. W.H.: Conceptualization, Writing—Original Draft, Supervision, Project administration, Funding acquisition.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-21646-x.

Correspondence and requests for materials should be addressed to K.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.