

Machine learning based longitudinal virtual diagnostics at SwissFEL

S. Bettoni,^{a)} G. L. Orlandi,^{b)} F. Salomone,^{c)} R. Boiger, R. Ischebeck, and R. Xue^{d)}

Paul Scherrer Institut, 5232 Villigen, Switzerland

A. Mostacci

Sapienza University of Rome, 00161 Rome, Italy

(Dated: 7 December 2023)

The bunch length in linac driven Free Electron Laser (FEL) is a major parameter to be characterized to optimize the final accelerator performances. In linear machines this observable is typically determined from the beam imaged on a screen located downstream of a Transverse Deflecting Structure (TDS) used to impinge a time dependent kick along the longitudinal coordinate of the beam. This measurement is typically performed during the machine setup and only sporadically to check the beam duration, but it cannot be continuously repeated, because time consuming and invasive. A non-invasive method to determine the electron bunch length was already presented in the past. This method is based on the analysis of the synchrotron radiation light spot emitted by the bunch passing through a magnetic chicane, provided that the energy chirp impinged on the bunch by the upstream radiofrequency structures is known. In order to overcome a systematic discrepancy affecting the synchrotron radiation monitor based results compared to the absolute TDS based ones, we implemented and optimized a Machine Learning approach to predict the bunch length downstream of the two SwissFEL compression stages - from about 10 fs up to about 2 ps - as well as the beam longitudinal profile at the first one.

I. INTRODUCTION

In the latest decades several Free Electron Laser (FEL) facilities^{1–4} are available all around the world to investigate material properties on the atomic scale^{5–7}, for instance, to explore chemical^{8,9} and biological^{10,11} mechanisms. One of the main advantages of the FELs over storage ring synchrotrons is the possibility to perform time resolved measurements exploiting the very short pulse length achievable in FELs (sub-fs and even shorter pulse modes). For these measurements the knowledge of FEL pulse length hitting the samples is crucial not only for the setup of the lasing but also for the correct interpretation of the results as well. The length of the electron bunch represents an upper limit to the duration of the photon pulse. A preliminary calibration of the photon length versus the electron bunch at the machine set-up for a photon beam time session can be used as a reference during the entire users' experiment session.

Several techniques are used in a FEL facility to characterize the bunch length such as streak cameras, electro-optic sampling¹², and active^{13–15} or passive streaking^{16,17} devices. The most used technique is based on the Transverse Deflecting Structures (TDSs). The RF field resonating in a TDS introduces a time-dependent transverse kick (streak) along the electron bunch longitudinal coordinate. The TDS streaked beam can be imaged by a view-screen. Provided that the calibration of the centroid of the streaked beam versus TDS phase has been

performed, from the knowledge of the TDS frequency an absolute measurement of the the time duration of the electron pulse as well as the longitudinal profile can be determined. Such a measurement is absolute, very reliable, but fully destructive for the beam, and it is time consuming since it may require the adjustment of the strength of some quadrupoles to optimize the measurement resolution. Furthermore, at some special locations of the machine, this measurement may also create losses due to the beam itself and secondary particle showers generated by the beam hitting the screen used to image the streaked downstream of the TDS. For all these reasons a non-invasive and faster method to determine the electron bunch length is very attractive.

In this paper we present the results of a Machine Learning (ML) approach to determine the electron bunch length at SwissFEL¹⁸, the FEL facility in operation since 2019 at Paul Scherrer Institut. In consideration of the limited amount of data available from a TDS measurement because of the related machine protection issues, we started implementing, testing and optimizing a ML algorithm taking advantage of the large amount of bunch length measurements which were obtainable from a Synchrotron Radiation Monitor (SRM) based and fully non-invasive method¹⁹ as described in Sec. III. We applied hence the same ML algorithm to the absolute measurements of the bunch length obtained by the TDS measurement technique. Besides bunch length prediction, we adapted the ML algorithm to reconstruct the beam longitudinal profile as well.

We determined the bunch length at different compression stages to maximize the range of variation of the beam duration, and to have a way to better characterize the beam parameters as well. Some experiments may require more monochromatic or large bandwidth beam. A very effective way to tune the latter parameter is to vary

^{a)}Electronic mail: simona.bettoni@psi.ch

^{b)}Electronic mail: gianluca.orlandi@psi.ch

^{c)}Also with Sapienza University of Rome, 00161 Rome, Italy

^{d)}Presently at II. Institute of Physics C, RWTH-Aachen University, 52074 Aachen, Germany

the bunch length and the current profile at the exit of the first compression stage. Different beam properties at this location have indeed a strong impact on the wake-field excited by the beam propagating along the downstream linac, which dechirp the beam, varying the final energy variation along the bunch and consequently the bandwidth of the radiation. This is the reason why the knowledge of the bunch length and current longitudinal profile also at the first compression stage is important for the full beam characterization.

ML is taking more and more space in the beam diagnostics in all the major accelerators all around the world^{20–23}. This is the first time that such a technique has been successfully applied at SwissFEL to determine the bunch longitudinal properties.

II. SwissFEL LINAC

SwissFEL is the FEL facility in operation since 2019 at the Paul Scherrer Institut in Switzerland, able to provide radiation in a wavelength ranging from 0.1 nm up to 7 nm generated in two branches (hard x-ray and soft x-ray). The schematic layout of the machine is shown in Fig. 1. The electron bunches with a charge from 10 pC up to 200 pC are generated by a semi-conductor Cs₂Te photocathode illuminated by ultraviolet laser pulses at a repetition rate of 100 Hz according to a 28 ns long two-bunch macro-pulse structure. Since we performed all the measurements presented in this paper at 200 pC, we will focus on the parameters referring to this case. A relatively long bunch is emitted from the photocathode (few ps rms, corresponding to about 20 A peak current) to counteract the space charge force, which may induce emittance degradation. Downstream of the gun section, there is the injector, where four radio frequency (rf) S-band (f , frequency = 2.9988 GHz) and one X-band (f = 11.9920 GHz) rf structures are installed. The first two structures (operated on-crest) boost the bunch energy up to about 140 MeV, whereas the other two (operated off-crest) accelerate up to about 300 MeV and impinge the energy-time dependence along the electron bunch to allow the longitudinal compression of the beam to be performed in a downstream magnetic chicane. In the rest of the paper we will indicate with 90 degrees the phase corresponding to the on-crest operation. A fourth harmonic X-band cavity (operated in anti-crest, i.e. at 270 degrees) linearizes the beam longitudinal phase space of the S-band energy-chirped bunch. With the energy correlation induced by the injector's structures the beam is compressed traveling through a four dipole chicane (Bunch Compressor 1, BC1) at an energy of 300 MeV. The following C-band (f = 5.7120 GHz) linac 1 (operated off-crest) boosts the bunch up to a second four dipoles chicane (Bunch Compressor 2, BC2) where the beam is further compressed at an energy of 2.1 GeV. After that, the beam is accelerated up to the nominal energies by the downstream linacs before being injected into the soft

and hard x-ray undulator lines.

The absolute bunch length measurement can be performed in the linac by means of S-band²⁴ and C-band²⁵ TDSs, located downstream of BC1 and at the entrance of the high energy undulators line, respectively. Table I shows the parameters relevant for the compression stages used for the measurements described in this paper.

TABLE I. Most relevant beam and machine parameters at the two SwissFEL compression stages for the 200 pC operating mode. R_{56} is the element of the transport matrix relating the beam momentum and longitudinal coordinate.

	BC1	BC2
Energy (GeV)	0.3	2.1
Initial bunch length rms (ps)	3.7	450
Final bunch length rms (fs)	500	20
R_{56} (mm)	64.3	20.7

As aforementioned, we performed measurements at BC1 using the SR-based method to test the ML approach, using the large available number of data we could have with this method and at this location. After that, we applied the same ML algorithm to retrieve the bunch length. After having successfully tested the ML approach to retrieve bunch lengths from few hundreds of fs up to few ps, we applied the same method to the second SwissFEL bunch compressor, where we retrieved bunch lengths of the order of 10 fs. With a small modification of the ML code, we determined not only the bunch length, but also the beam longitudinal profile at BC1.

III. SRM-BASED BUNCH-LENGTH MEASUREMENTS

In this section we briefly recall the method to determine the electron bunch length based on the Synchrotron Radiation Monitor (SRM) we already presented in the past¹⁹. We compared the SRM-predicted bunch-duration with the outcome of the bunch-length measurements performed using the standard technique based on the TDS streaking. We calibrated the SRM-based predictions with the TDS-based measured bunch length changing the longitudinal chirp of the bunch at fixed charge, i.e., according to the usual procedure of machine set-up for users' operations. Due to the non-invasiveness of the SRM, we could retrieve many data to test the ML approach, verify the robustness of the results varying the model parameters, and determine the minimum number of images necessary to converge to a good solution.

A. The method

The two SwissFEL magnetic chicanes are equipped with SRMs for the shot-to-shot and non-invasive determination of the mean and spread of the electron beam energy distribution at the maximum machine repetition

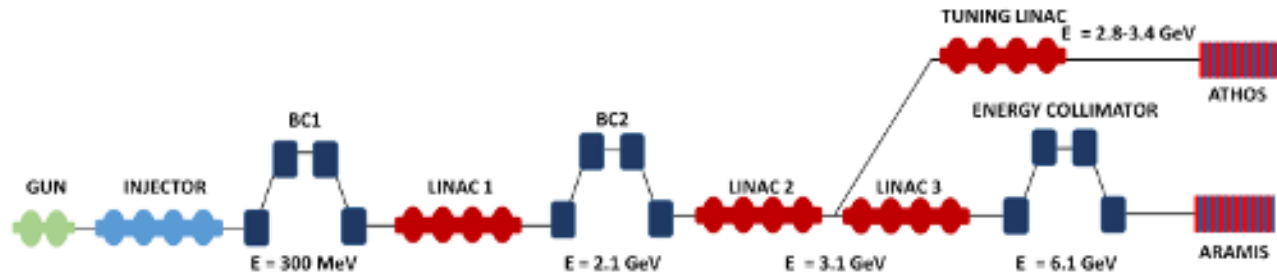


FIG. 1. SwissFEL schematic layout. The full length of the machine is about 740 m.

rate of 100 Hz. Thanks to a periscope mirror set-up and a 300 mm focal length lens, the sCMOS camera of each SRM can acquire the transverse profile of the electron beam at the middle plane of the chicane by imaging the edge-SR light emitted by the electron beam while crossing the front edge of the third dipole of the magnetic chicane. The mean and rms parameters of the energy distribution of the electron bunch can be determined from the analysis of the horizontal projection of the imaged SR light spot for the given value of the dispersion η at the location where the SR light is emitted. The SRM in operation at BC1 is equipped with a fast gated MCP sCMOS camera that is able to resolve either bunch-1 or bunch-2. The SRM at BC2 is instead equipped with a standard sCMOS camera which is not able to discriminate the single bunch in the 28 ns long 2-bunch macro-structure of the SwissFEL beam. With a projected pixel size of 99 μm and 45 μm respectively, the two SRMs can ensure a relative energy spread resolution of 2.21×10^{-4} and 1.61×10^{-4} in the two chicanes at the nominal dispersion of 448 mm and 280 mm, respectively.

Recently, besides the characterization of the energy distribution of the electron beam, a new diagnostics application of the SRM has been developed and proposed¹⁹. Such a new diagnostics application of a SRM in a magnetic chicane permits a non-invasive and shot-to-shot determination of the electron bunch length from the measurement of the SRM horizontal rms size σ_x of the electron beam, provided that a suitable algorithm processing the RF settings (phase and amplitude) of the upstream accelerating structures is applied. Figure 2 shows a schematic view of the SwissFEL injector, where the relevant components are sketched.

The formal development of the SRM based measurement method of the electron bunch length starts from a first order Taylor series expansion - around the nominal setting phase of each accelerating structure - of the energy gain that the electron beam experiences. This permits to express the relative variation of the beam energy ΔE with respect to the central energy of the beam E_0 as a function of the longitudinal bunch duration, Δz , via the proportionality constant defined by the linear chirp

T_1 , as:

$$\frac{\Delta E}{E_0} = \sum_{j=1}^M \left[\frac{A_j k_j}{E_0} \cos(\phi_j) \right] \Delta z = T_1 \Delta z, \quad (1)$$

where M is the number of structures up to the compressor, and the coefficient T_1 , defined as:

$$T_1 = \sum_{j=1}^M \left[\frac{A_j k_j}{E_0} \cos(\phi_j) \right], \quad (2)$$

can be determined shot-to-shot from the beam synchronous acquisition of the injector rf parameters, i.e. the rf voltage amplitude, A_j , and phases, ϕ_j , of the accelerating structures installed along the injector, and the wave number of each cavity k_j . On the other hand, the relative energy spread of the electron beam can be expressed as a function of the spread Δx of the distribution of the horizontal coordinates of the electron beam measured by the SRM monitor in the magnetic chicane:

$$\frac{\Delta E}{E_0} = \frac{\Delta x}{\eta}, \quad (3)$$

where η is the dispersion of the magnetic chicane at the location where the light is emitted. From the equalization of the two formulas - eqs.(1,3) - the sigma rms (σ_z , bunch length) of the distribution of the longitudinal coordinates of the electron beam at the entrance of the magnetic chicane can be expressed as a function of the SRM measurement (σ_x) of the horizontal distribution of the electron in the magnetic chicane:

$$\sigma_z = \frac{1}{T_1} \frac{\sigma_x}{\eta}. \quad (4)$$

Finally, after completion of the beam compression, the bunch length at the exit of the magnetic chicane, rms value σ_t , as a function of the parameter R_{56} reads¹⁹:

$$\sigma_t = \frac{1}{c} |1 + T_1 R_{56}| \frac{1}{T_1} \frac{\sigma_x}{\eta}, \quad (5)$$

where c is the in vacuum speed of the light.

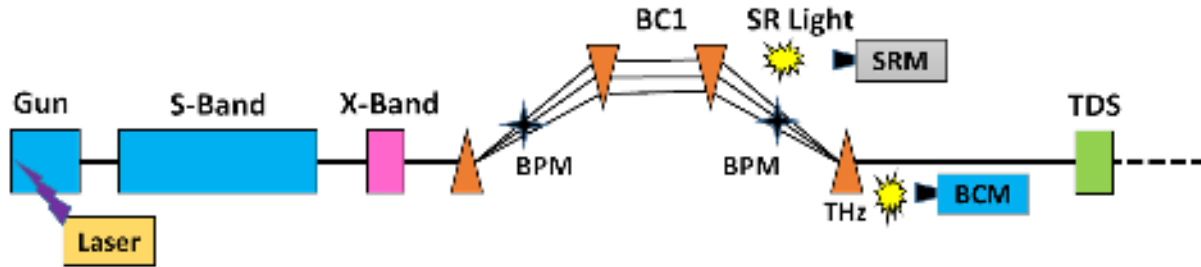


FIG. 2. Sketch of the SwissFEL injector: rf structures, magnetic chicane and main beam instrumentation are reported.

B. The measurements

We applied the method described in the previous subsection to the SRM installed at BC1 - just downstream of the SwissFEL injector - to characterize the rms σ_z of the longitudinal distribution of the electron bunch by determining the rms σ_x of the horizontal profile of the electron beam at the middle point of the magnetic chicane. We varied the bunch length at BC1 by more than a factor 5 changing the off-crest phase of the rf accelerating structures upstream of BC1, keeping constant the other parameters (bunch charge, R_{56} , X-band phase and amplitude, etc.). The phase of off-crest S-Band RF structures is indeed the parameter typically adjusted to set the bunch length during the machine preparation for users' operation and to tune it during the photon delivery runs.

Figure 3 shows the obtained comparison of the bunch length determined using the SRM-based method and the TDS²⁴ in two measurement sessions one month apart.

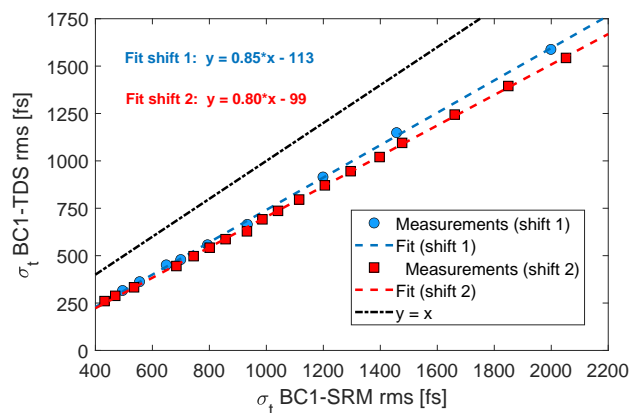


FIG. 3. Comparison of the bunch length obtained using the SRM-based method and the TDS at BC1. The result of the linear fit of the measurements for two shifts happening at one month difference in time is reported.

The experimental data shown in the plot indicate a good reproducibility of the measurements as well as a systematic discrepancy between the absolute TDS measurements and the SRM predictions of the bunch length.

The origin of the systematic discrepancy is not understood yet. Nevertheless, the calibration curves plotted in Fig. 3 permit to determine the corrections to be applied at the SRM based measurements to get them aligned to the absolute characterization of the TDS. The calibration curve from SRM to TDS has been obtained under the assumption that only the rf phases are tuned to set the compression according to standard procedure for the setting and tuning of the machine for users' operations. More calibration curves analogous to that reported in Fig. 3 may be determined varying other machine parameters.

As aforementioned, we used the SRM method predictions to exploit its shot-to-shot and non-invasive features which were very useful for the acquisition of a large amount of data and fundamental for the implementation and optimization of the ML algorithm, as in the following described.

IV. MACHINE LEARNING APPROACH

Neural networks (NN), a special type of machine learning methods²⁶ are powerful computational instruments to model the functional dependence between input (x_i) and output (y_i) data. Due to the universal approximation theorem²⁷, it is possible to represent any function with NN. Many different neural network architectures evolved in the last decades²⁸. Within this work we consider densely connected deep forward neural networks. This means that our network consists of several layers (deep), each of them containing several neurons. All the neurons from one layer are connected with every single neuron of the following layer (densely connected) through specific activation functions. In other words, the NN maps the inputs x_i to the outputs y_i by using activation functions g :

$$y_i = g \left(\sum_{i=1}^N w_i x_i + b \right), \quad (6)$$

where the weights w_i and bias b are learnt through optimization, and N is the number of data. Training the neural network means that the weights and biases applied to the inputs of each layer are adjusted according

to an optimization procedure aiming to minimize a loss function defined accordingly to the problem, so that at the end of this procedure the outputs predicted by the NN match with those not used to determine the NN. In some cases a Gaussian noise layer²⁹ may be also added to the data to make the neural network more stable.

The main steps of ML are typically the training and the validation, which use different subsets randomly selected among the full available dataset. The dataset devoted to the training is used by the ML algorithm to determine the w_i and b of eq. 6. After this, the validation compares the expectation of the model prediction using the x_i of the data excluded from the training process. The split parameter defines the fraction of the data going to the training and to the validation. For example a split parameter equal to 10 means that 10 % of the data is used for the validation and the remaining 90 % is used for the training of the model. The data for the validation are randomly extracted from the complete dataset (considered as a block of data). A further verification of the ability of the model to describe the system is done by the testing. In this case all the data referring to particular set-points are excluded from the training and the validation as well. After that the training is performed on the remaining data, the obtained model is used to foresee the NN expectation for the data excluded from the training and the validation. Some pre-processing may be applied to the data depending on the nature of the problem to be described or the values of the data. We applied a scaling of the data because of their large relative variation (bunch length) or alignment of them (bunch profiles) to compensate for some jitter intrinsic in the measurement.

To set up this NN architecture, the following hyperparameters need to be chosen: number of layers and neurons, activation functions, noise level of Gaussian noise layer, loss function (minimization function, which may be defined according to the model), optimization algorithm and its learning rate (proportional to the number of iterations necessary to obtain the minimum and stable loss function), number of epochs (number of iterations the optimization algorithm should perform).

The ML framework we selected is TensorFlow³⁰ and Keras³¹, open source Python libraries. In the following section the details on how the number of layers and neurons are chosen in our application are described. For all the layers, a REctified Linear Unit (RELU)³² was chosen as activation function. As optimization algorithm we used the Adam algorithm³³.

The application of ML to determine the electron bunch length in a large range of values using both the SwissFEL compressors and the bunch longitudinal profile are described in the following subsections.

A. The measurements

We applied ML to the exit of both the SwissFEL compressors to verify the possibility of the NN to correctly

retrieve very different values of bunch lengths.

We determined the electron bunch length using SRM based method - as described in the previous section - as well as the TDS-based technique. In the first case we could take a larger number of shots for each machine setting, since we could operate SwissFEL at its maximum frequency of 100 Hz. This gave us the opportunity to test the model in a large range of final bunch lengths, and study the minimum number of data to have a convergence of the results. After that, we used the TDS measurements of the bunch length to implement the ML model under bunch-length conditions around the SwissFEL design value. In the case of the TDS measurement the number of shots available is heavily reduced compared to the SRM case, since the repetition rate of the facility is limited to 1 Hz for machine protection and safety issues.

After applying the ML to determine the electron bunch length, we also applied the ML to the vertical profiles imaged by the view-screen downstream the TDS to determine not only the bunch length, but also the longitudinal profile of the bunch.

1. Bunch length reconstruction at BC1

As the first step of the ML we built the matrix containing the machine parameters which play a role on the bunch compression, i. e. the phase and amplitude of the rf field in the structures upstream of BC1. We recorded 3000 shots for each of the considered 18 set-points of the phases. These are the input data x_i for the NN. Here and in the following we label each set-point with a phase, but, due to the feed-back acting to maintain the beam energy constant along the machine, at each phase corresponds also an rf amplitude. We used all these data to build up a matrix whose dimensions are the number of shots of each subset multiplied by the number of compression phases, and the number of channels (5 phases and 5 amplitudes), respectively. As a test, we added also other channels (bunch charge, position of the beam in dispersive beam position monitors, ...) without significantly improving the accuracy of the model prediction. As a second step we determined the electron bunch length for each of the shots. This provides the output data y_i . We did this in two ways: we determined the bunch length applying the SRM method, and from the profile of the TDS-streaked beam.

We used the results of the validation phase to determine the optimal values for the number of layers and neurons. As a loss function to train the NN we used the Mean Squared Error (MSE), defined as:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_{pi} - y_{ti})^2 \quad (7)$$

where N is the number of data considered, and y_{pi} and y_{ti} the output prediction and the training values, respec-

tively. To evaluate the goodness of the NN prediction, considering the large variation of the y_i for the different compression settings, we defined a relative quantity, that we called Relative Mean Squared Relative Error (we will refer to this as *RMSRE* in the following) as:

$$RMSRE = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{(y_{pi} - y_{ti})^2}{y_{pi}^2}} \quad (8)$$

where N is the number of data. We found a nearly constant *RMSRE* for a number of layers and neurons between 5 and 10. For less complex models (smaller number of layers and/or neurons) we did not obtain a satisfactory prediction of the measurements, and for more complex structures we do not have a convergence of the model, realistically due to an over-fitting effect (we excluded these cases because we want the model as simple as possible for a satisfactory characterization of our problem). Based on these considerations, we worked using 5 layers and 5 neurons for the following analyses. With these values we obtained in quite a short time (around 1 minute considering all the data corresponding to the 18 compression phases on a standard laptop) a very good *RMSRE* of the order of a few percent.

Figure 4 shows the comparison of the measured data with the prediction of the model using the SRM method to determine the bunch length for values ranging from about 300 fs up to few ps. We obtained a *RMSRE* of

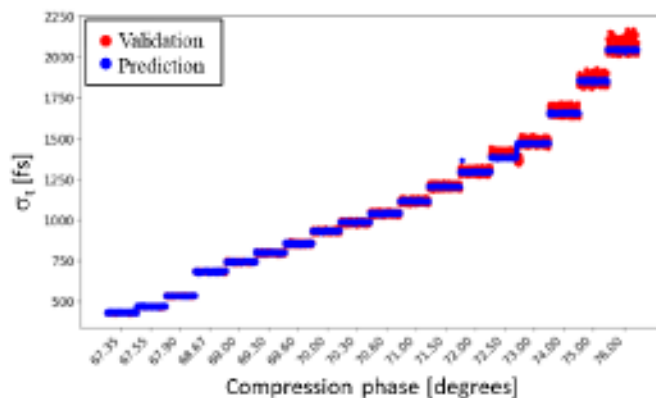


FIG. 4. Validation ML applied to SRM-method: comparison of the ML predicted bunch length as a function of the compression phases and the measurements excluded from the training phase. This corresponds to a *RMSRE* of 1.2 %.

1.2 %. In particular, Fig. 5 shows the histogram of the relative difference of the data used for the validation and the prediction of the model. There are only few points where the relative difference of the measurements and the ML predictions differ by 10 %, whereas for the majority of the cases we obtained a discrepancy in the range of about ± 2.5 %.

To perform the testing, we applied the training to 7 of the 18 set-point phases, and we used the NN obtained

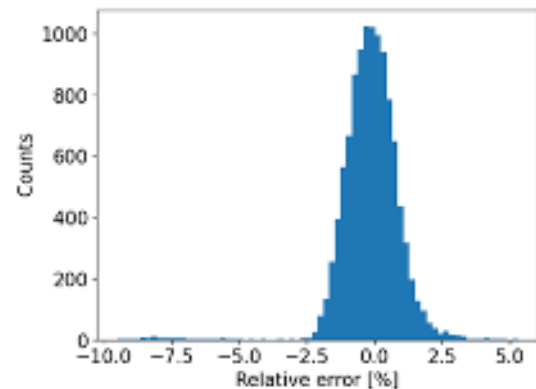


FIG. 5. Validation ML applied to SRM-method: histogram of the relative difference between the predicted values and those extracted for the validation.

with this limited dataset to predict the value of the bunch length for the other 11 phases. Figure 6 shows the obtained result. As expected, in this case we obtained a

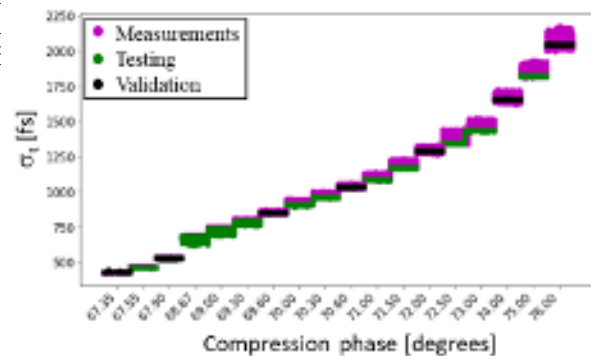


FIG. 6. Testing ML applied to SRM-method: comparison of the ML predicted bunch length as a function of the compression phases and the measurements excluded from the training phase. The *RMSRE* computed only on the cases used for the testing is 3 %.

RMSRE worse than the previous case, but still of 3 %, giving some confidence of the ability of the NN trained to describe the system.

We used the full data set to determine the minimum number of shots necessary to obtain a convergence of the ML model: this in view of applying the ML model to the bunch length determined using the vertical profile of the streaked beam instead of the SRM data. To do this, we randomly extracted a subset of data that we used for the training and validation by varying the split parameter. We applied then the ML analysis to the different cases. Figure 7 shows the mean, the standard deviation, and the minimum of the *RMSRE* obtained considering 10 runs

of the ML model for each case. The minimum gives the

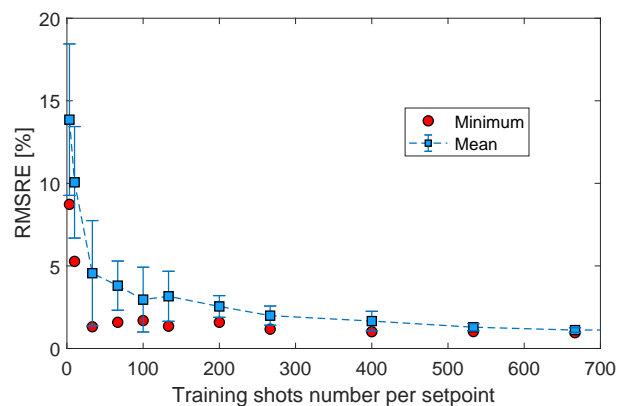


FIG. 7. Accuracy of the NN prediction using the SRM data as a function of the number of shots considered for the training. The calculations have been repeated for 10 runs assuming 5 layers and neurons.

information on how good the ML model is able to predict the bunch length repeating the run for 10 seeds, whereas the mean and the standard deviation provide information on the expectation for the prediction for a single ML run. As evident from Fig. 7 the model accuracy is quite constant for a number of shots as small as 30. The mean value is constant for a minimum of 200 shots with the error bars increasing toward the origin (more unstable for less data). This latter value corresponds to something less than 1 minute for an acquisition of data for a single compression phase using the TDS to determine the bunch length. This seems to be a reasonable number of shots that we need to acquire to train the NN assuming a reasonable number of seeds.

We finally applied the ML to determine the bunch length obtained using the standard technique to measure the electron bunch length based on the streaking of the beam by the TDS. We considered values of the compression around the SwissFEL design, corresponding to a final rms bunch length around 500 fs. For each set phase we took a number of shots a factor 3 larger than that necessary (600 shots per set-point phase) to repeat the previous analysis on the number of data using this different diagnostics. Figure 8 shows the very good agreement obtained in the validation phase for this case. In this case we obtained a *RMSRE* of 0.84 %. This number is even better than that obtained using the SRM. This is due to the fact that in this case we have less noise, since any time jitter (rf and/or beam arrival time) is compressed more in this case, since the beam travels through the entire BC1 and not only for 3/4 of its full length. The distribution of the relative differences of the bunch length obtained by the TDS-based measurements and the expectations from the NN is shown in Fig. 9. With only three data sets available, we applied the testing to estimate the intermediate bunch length using the measurements with the

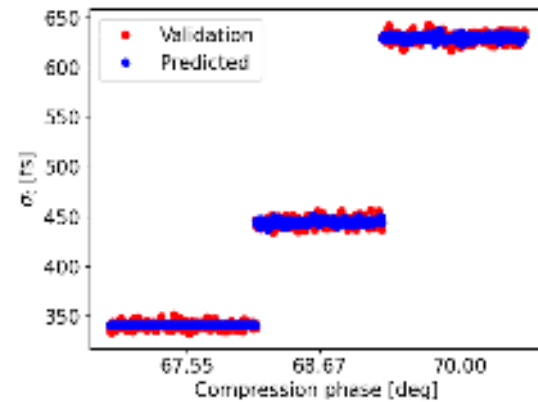


FIG. 8. Validation ML applied to the TDS measurements: comparison of the ML predicted bunch length as a function of the compression phases and the measurements excluded from the training phase. This corresponds to a *RMSRE* of 0.84 %.

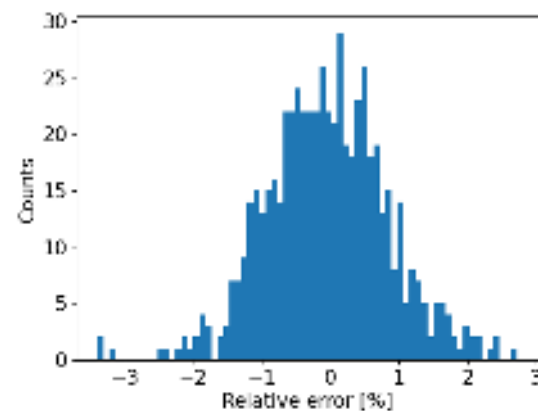


FIG. 9. Validation ML applied to TDS: histogram of the relative difference between the predicted values and those extracted for the validation.

TDS. In this case we obtained a quite satisfactory result, as shown in Fig. 10.

The model converges at a *RMSRE* below 2 % for a number of shots around 60, as shown in Fig. 11. This will allow in the future to have a realistic invasive training phase of the ML model using the TDS-based method to determine the y_i in a reasonable time, even at the reduced frequency of 1 Hz, i.e., the maximum possible machine repetition rate when the beam is intercepted by a screen.

In summary, we obtained a prediction of the bunch length at the level of few percent applying a ML algorithm to bunch length values ranging from few hundreds of fs up to about 2 ps. We have also performed studies on the convergence of the model varying its complexity and the number of data provided to it. We used this

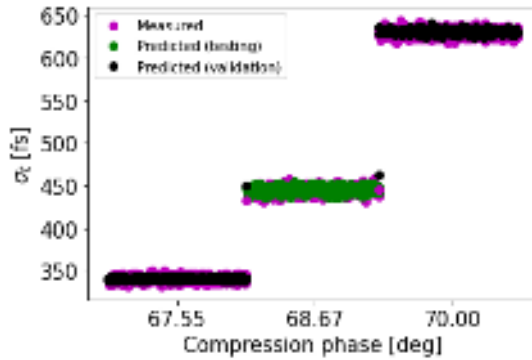


FIG. 10. Testing ML applied to the TDS measurements: comparison of the ML predicted bunch length as a function of the compression phases and the measurements excluded from the training phase. The *RMSRE* computed only on the cases used for the testing is 1.2 %.

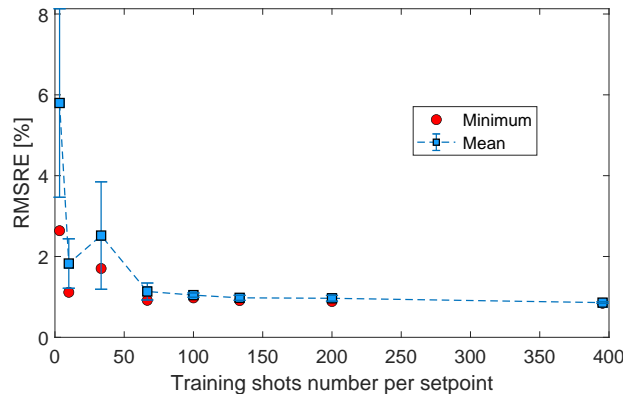


FIG. 11. Determination of the accuracy of the NN prediction using the TDS screen data as a function of the number of shots considered for the training. The calculations have been repeated for 10 seeds assuming 5 layers and 5 neurons.

latter study to determine the minimum number of shots necessary to obtain satisfactory predictions using a non-invasive diagnostics (SRM) and a more precise but invasive one (TDS). In both cases we determined that already with few tens of shots per set-point the NN provides an estimation of the bunch length precise at the level of a few percent. We verified with the testing that the NN well represents the system also when a smaller number of data points is provided. We repeated this investigation also using the two diagnostics, even if, due to the larger number of data initially available from the SRM, this latter test was more meaningful.

2. Bunch current profile reconstruction at BC1

As anticipated, we applied ML to determine not only the bunch length but also the beam longitudinal profile. To do this we substituted the value of the bunch length with the profile of the streaked beam imaged on the screen downstream of the TDS. The y_i in this case is not a scalar anymore for each shot, but an array with a size given by the number of the camera pixels of the profile in the streaking direction. Apart from this, all the rest is analogous to what was described for the bunch length reconstruction.

Figure 12 shows the comparison of one of the profiles measured with the prediction of the trained NN model without applying any pre-processing to the measured data (phase = 68.67 degrees, $\sigma_t = 700$ fs). As

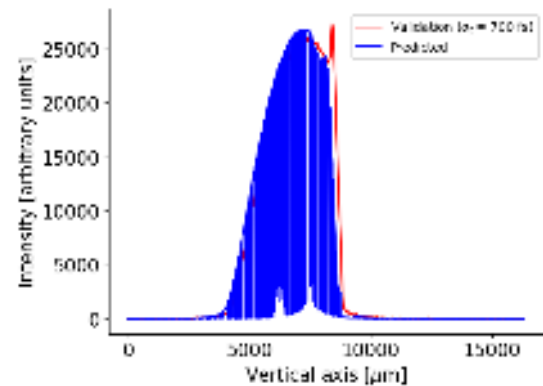


FIG. 12. Comparison of one of the profiles expected by the trained NN and the measured one at an intermediate compression phase.

shown in Fig. 12 we obtained a predicted profile characterized by a very large ringing compared to the measured one. The so predicted beam profile was obtained after tuning the number of layers, neurons and split, as well as applying different Gaussian noises. The spikes visible in the predicted profile are clearly an artifact of the NN, since the TDS-based method has enough resolution to spot these structures. We identified as the source of these not physical modulations the jitter of the beam profile in the direction of the streaking on the view-screen where it is imaged. The TDS-streaked beam profile suffers from several sources of time jitter (laser arrival time, TDS phase and amplitude jitter), which translates to a transverse jitter of the centroid position in the direction of the streaking. This jitter must be added to the transverse orbit jitter of the beam. Differently from the case of the bunch length, these jitters may be very harmful during the reconstruction of the beam longitudinal profile. As a simplified example Fig. 13 shows a series of jittering square profiles artificially created to simulate the data of the training phase of the model. In presence of jitter the intensity associated to each pixel along the central region where the signal is flat for all the shots may

the beam longitudinal profiles at SwissFEL using ML.

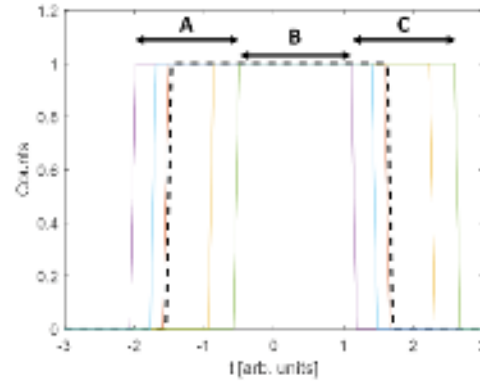


FIG. 13. Example of a series of jittering square profiles artificially created (*counts* represents the amplitude of the profile, and *t* the time coordinate along the bunch. The region of the profile B may be correctly reconstructed by the ML, whereas in the sections corresponding to A and C will very probably be not correctly reconstructed, because very contradictory (very different intensity for the same set-point of the compression phase in our case) distributions are given to the model.

be correctly reconstructed. On the contrary, in the region where the shot-to-shot intensity difference from one profile to another one is large (region A and C) the ML model will get some contradictory data, and it will very probably fail. Based on this considerations, we applied a pre-processing to the profiles before using them in the training phase. We align them to the same centroid position (same intensity, same rising edge pixel, or other criteria may be considered). This pre-processing is essential, since it heavily reduces the large ringing present in the reconstructed profile shown in Fig. 13, as reported in Fig. 14. The ML determined with a very satisfactory

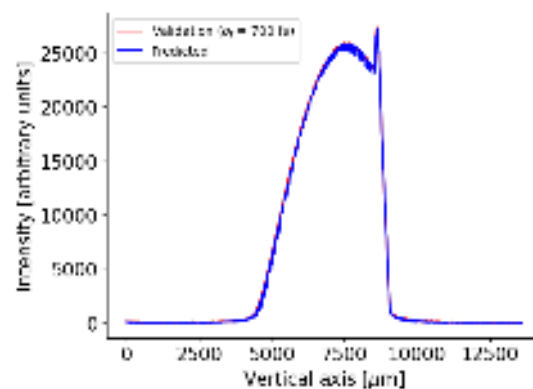


FIG. 14. Comparison of the electron beam longitudinal profile measured streaking the beam with a TDS and the ML prediction. We obtained this result applying a pre-processing to the measured profiles to align them using the centroid.

agreement the beam longitudinal profile determined intercepting the bunch with a screen downstream of the TDS. This represents only the first attempt to determine

3. Bunch length reconstruction at BC2

We applied the ML at BC2, similarly to how we did for the low energy compressor. This is useful to extend the range of bunch length values where the ML can successfully predict the electron bunch length. In particular we verify that NN can successfully determine the beam duration for few fs bunch lengths, values in the range those at the undulator lines.

In this case we directly applied the ML to the bunch length measured using the TDS due to some more analyses and developments necessary to correctly apply the SRM-based method at this location of the machine. Another important issue which prevents to perform many measurements of the bunch length at this location is the beam-loss control at high energy (from 2.1 GeV) when the beam is intercepted by a screen for the TDS measurement. All these arguments make the use of the ML at SwissFEL really useful for the machine tuning.

We restricted ourselves to some of the shortest values of the bunch length achieved at the maximum SwissFEL charge of 200 pC to explore a different range of beam duration compared to those measured at BC1. We took 300 shots and 150 shots for the lowest and the two other compression phases, respectively.

In the ML model the only differences compared to the BC1 analysis is that the input data are the phase and the amplitude of all the RF cavities upstream of BC2 (gun, injector, and Linac 1 as well) and not only at the injector, and the bunch length is measured using a C-band TDS at a screen downstream of BC2. Except for this, the procedure is exactly the same one we followed for BC1.

Figure 15 shows the very good agreement of the bunch length foreseen by ML and the measurements kept for the validation phase. We obtained a value of RMRSE of 6.0 % with an absolute difference of the predicted and the measured bunch lengths of the order of few fs, as shown in 15. The distribution of the relative error is of the order of 10 % with a large fraction of the data in the ± 5 %, as shown in Fig. 16. Also in this case we applied the testing to determine the intermediate bunch length. Figure 17 shows the result.

The ML allowed to predict the bunch length with an absolute error of the order of only a few fs and a *RMSRE* smaller than 10 %. This result is remarkable, especially considering the relatively limited amount of available data due to technical limitations (low tolerance to beam losses in the high energy session of the machine when an intercepting device is used to measure), which makes the use of a ML approach even more attracting.

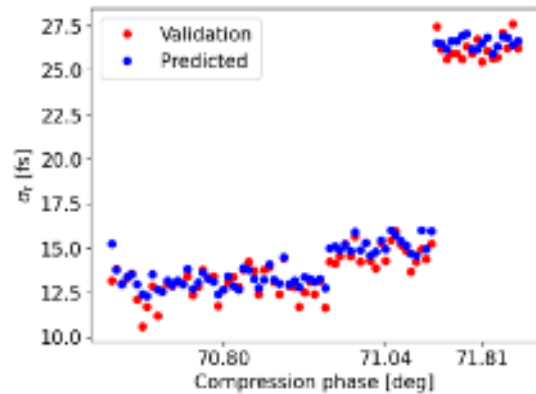


FIG. 15. Validation ML applied to TDS measurements at BC2: comparison of the ML predicted bunch length as a function of the compression phases and the measurements excluded from the training phase. The bunch length is computed using the vertical profile of the streaked beam. This corresponds to a *RMSRE* of 6.0 %.

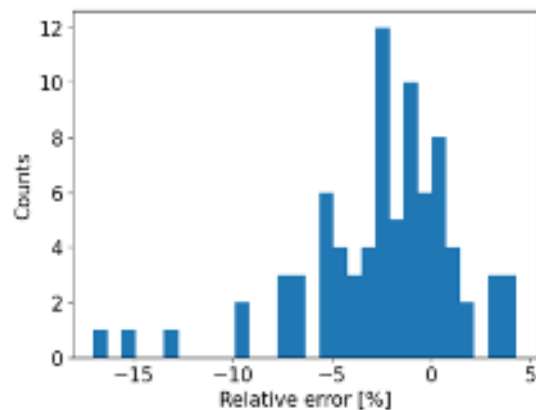


FIG. 16. Validation ML applied to TDS: histogram of the relative difference between the predicted values and those extracted for the validation.

V. CONCLUSIONS

We presented the implementation and optimization of a ML algorithm to determine the electron bunch length at SwissFEL. We performed the studies at both the magnetic compressors to cover a very large range of bunch lengths varying by a factor 200 (from about 2 ps down to about 10 fs). We achieved a relative difference of the order of few percents between the ML predictions and the measured values for the entire bunch lengths range. The combined knowledge of the bunch lengths at the exit of both the compressors allows to determine not only the beam final bunch length (given by the bunch length at BC2) but also the bandwidth, strongly dependent on the wakefields differently excited by different electron bunch lengths coming from the first compressor, and to better characterize the electron bunch.

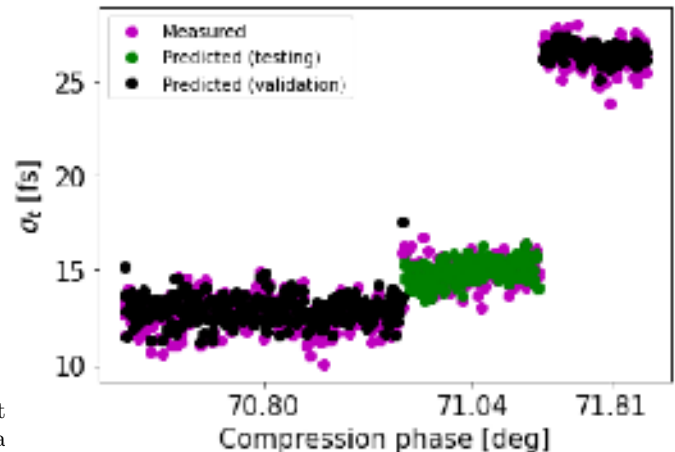


FIG. 17. Testing ML applied to the TDS measurements at BC2: comparison of the ML predicted bunch length as a function of the compression phases and the measurements excluded from the training phase. The *RMSRE* computed only on the cases used for the testing is 4 %, corresponding to a maximum absolute difference between predictions and measurements of few fs.

We used the non-invasiveness of a previously published method based on the analysis of the SR emitted by the bunch at a chicane (BC1) to acquire a very large number of data. This was used to test, optimize, and analyze the robustness of the ML algorithm in view of its application to predict the bunch duration obtained by the TDS, since the latter approach is destructive and much slower than the SRM-based method.

With an opportune pre-processing of the data and a minor modification of the ML code, we determined not only the bunch length but also the beam longitudinal profiles in a very good agreement to those measured using the TDS.

Last but not least from the point of view of the improvement and optimization of the diagnostics of the electron bunch length in SwissFEL, the so obtained ML results permit us to complement the systematic discrepancy affecting the SRM-based predictions of the bunch length. In this way we have an alternative and reliable shot-to-shot and non-invasive approach to measure the electron bunch length.

In conclusions, we applied ML to determine the bunch length and the longitudinal profile in a non-invasive way, reaching a very good level of accuracy in the prediction even using a limited number of data (less than 100 per machine set-points), and a relatively small number of channels (RF amplitudes and phases of the accelerating structures upstream of the compressor).

ACKNOWLEDGMENTS

The authors would like to thank the controls and operation groups for the support during the measurements shifts, and the extensive contributions of all PSI expert groups and the SwissFEL team to the construction and operation of the facility.

- ¹P. Emma, et al., "First lasing and operation of an ångström-wavelength free-electron laser", *Nat. Photon.* 4, 641–647 (2010).
- ²T. Ishikawa, et al., "A compact X-ray free-electron laser emitting in the sub-ångström region", *Nat. Photon.* 6, 540–544 (2012).
- ³H.-S. Kang, et al., "Hard X-ray free-electron laser with femtosecond-scale timing jitter", *Nat. Photon.* 11, 708–713 (2017).
- ⁴M. Altarelli, "The European X-ray free-electron laser facility in Hamburg", *Nucl. Instrum. Methods B* 269, 2845–2849 (2011).
- ⁵C. Bostedt, "Linac Coherent Light Source: the first five years", *Rev. Mod. Phys.* 88, 015007 (2016).
- ⁶P. Skopintsev et al., "Femtosecond-to-Millisecond Structural Changes in a Light-Driven Sodium Pump", *Nature* 583, 314 (2020).
- ⁷C. Bacellar et al., "Spin Cascade and Doming in Ferric Hemes: Femtosecond X-Ray Absorption and X-Ray Emission Studies", *Proc. Natl. Acad. Sci. U.S.A.* 117, 21914 (2020).
- ⁸M. Dell'Angela, et al., "Real-time observation of surface bond breaking with an X-ray laser", *Science* 339, 1302–1305 (2013).
- ⁹H. Oeström, et al., "Probing the transition state region in catalytic CO oxidation on Ru", *Science* 347, 978–982 (2015).
- ¹⁰C. Bacellar et al., "Femtosecond X-Ray Spectroscopy of Haem Proteins", *Faraday Discuss.* 228, 312 (2021).
- ¹¹P. Bächtel et al., "Lipidic Cubic Phase Serial Femtosecond Crystallography Structure of a Photosynthetic Reaction Centre", *Acta Crystallogr D Struct Biol* 78, 698 (2022).
- ¹²B. Steffen, S. Casalbuoni, E.-A. Knabbe, H. Schlarb, B. Schmidt, P. Schmuser, and A. Winter, "Electro optic bunch length measurements at the VUV-FEL at DESY", in *Proceeding at the 2005 Particle Accelerator Conference*, Knoxville, (USA).
- ¹³R. Akre, L. Bentson, P. Emma, and P. Krejcik, in *Proceedings of the Particle Accelerator Conference*, Chicago, IL, 2001 (IEEE, New York, 2001).
- ¹⁴D. Alesini, G. Di Pirro, L. Ficcadenti, A. Mostacci, L. Palumbo, J. Rosenzweig, and C. Vaccarezza, *Nucl. Instrum. Methods Phys. Res., Sect. A* 568, 488 (2006).
- ¹⁵M. Röhrs, C. Gerth, H. Schlarb, B. Schmidt, and P. Schmüser, *Phys. Rev. ST Accel. Beams* 12, 050704 (2009).
- ¹⁶S. Bettoni, P. Craievich, A. A. Lutman, and M. Pedrozzi, *Phys. Rev. Accel. Beams* 19, 021304 (2016).
- ¹⁷P. Dijkstal, A. Malyzhenkov, P. Craievich, E. Ferrari, R. Ganter, S. Reiche, T. Schietinger, P. Juranić, E. Prat, *Phys. Rev. Research* 4, 013017 (2022).

- ¹⁸E. Prat, R. Abela, M. Aiba, et al., "A compact and cost-effective hard X-ray free-electron laser driven by a high-brightness and low-energy electron beam", *Nat. Photonics* 14, 748–754 (2020).
- ¹⁹G. L. Orlandi, R. Xue, H. Brands, F. Frei, Z. Geng, V. Thominet, and S. Bettoni, "Bunch length and energy measurements in the bunch compressor of a free-electron laser", *Phys. Rev. Accel. Beams* 22, 072803 (2019).
- ²⁰F. Bakkali Taheri, I. V. Konoplev, G. Doucas, P. Baddoo, R. Bartolini, J. Cowley, and S. M. Hooker, "Electron bunch profile reconstruction based on phase-constrained iterative algorithm", *Phys. Rev. Accel. Beams* 19, 032801 (2016).
- ²¹A. Scheinker and S. Gessner, "Adaptive method for electron bunch profile prediction", *Phys. Rev. ST Accel. Beams* 18, 102801 (2015).
- ²²C. Emma, A. Edelen, M. J. Hogan, B. O'Shea, G. White, and V. Yakimenko, "Machine learning-based longitudinal phase space prediction of particle accelerators", *Phys. Rev. Accel. Beams* 21, 112802 (2018).
- ²³J. Zhu, Y. Chen, F. Brinker, W. Decking, S. Tomin, and H. Schlarb, "High-Fidelity Prediction of Megapixel Longitudinal Phase-Space Images of Electron Beams Using Encoder-Decoder Neural Networks", *Physical Review Applied* 16, 024005 (2021).
- ²⁴P. Craievich, R. Ischebeck, F. Löhl, G.L. Orlandi, E. Prat, *Transverse Deflecting Structures for Bunch Length and Slice Emittance Measurements on SwissFEL*, *Proceedings of FEL2013*, New York, NY, USA, TUPSO14 236-241 (2013).
- ²⁵E. Prat, A. Malyzhenkov, P. Craievich, "Sub-femtosecond time-resolved measurements of electron bunches with a C-band radio-frequency deflector in x-ray free-electron lasers", *Rev Sci Instrum* 94 (4): 043103 (2023).
- ²⁶T. Mitchell, "Machine Learning", McGraw Hill, ISBN 0-07-042807-7.
- ²⁷K. Hornik, M. Stinchcombe, and H. White "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks Vol. 2*, Pergamon Press. pp. 359–366.
- ²⁸L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions", *Journal of Big Data* volume 8, Article number: 53 (2021).
- ²⁹R. Reed, S. Oh, and R. Marks, "Regularization using jittered training data", *International Joint Conference on Neural Networks*, vol. 3, pp. 147–152, (1992).
- ³⁰<https://www.tensorflow.org/>.
- ³¹<https://keras.io>.
- ³²X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks", *proceeding at the International Conference on Artificial Intelligence and Statistics* (2011).
- ³³D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization", *proceeding at the 3rd International Conference on Learning* (2015).