ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Alaa El-Din KK, Forte A, Kasim MF, Miniati F, Vinko SM. 2024 STEP: extraction of underlying physics with robust machine learning. *R. Soc. Open Sci.* **11**: 231374.
https://doi.org/10.1098/rsos.231374

Received: 19 September 2023
Accepted: 20 March 2024

Subject Category:
Physics and biophysics

Subject Areas:
spectroscopy, artificial intelligence, plasma physics

Keywords:
physics, machine learning, artificial intelligence, differentiable modelling, resonant inelastic X-ray scattering, spectroscopy

Author for correspondence:
Karim K. Alaa El-Din
e-mails: karim.alaael-din@physics.ox.ac.uk; karim@aedin.dev

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7247039.

THE ROYAL SOCIETY PUBLISHING

STEP: extraction of underlying physics with robust machine learning

Karim K. Alaa El-Din[1], Alessandro Forte[1], Muhammad Firmansyah Kasim[1,2], Francesco Miniati[1] and Sam M. Vinko[1,3]

[1]Department of Physics, University of Oxford, Oxford, UK
[2]Machine Discovery, Oxford OX4 4GP, UK
[3]Central Laser Facility, STFC Rutherford Appleton Laboratory, Didcot, OX11 0QX, UK

KKAE-D, 0000-0001-9140-4489

A prevalent class of challenges in modern physics are inverse problems, where physical quantities must be extracted from experimental measurements. End-to-end machine learning approaches to inverse problems typically require constructing sophisticated estimators to achieve the desired accuracy, largely because they need to learn the complex underlying physical model. Here, we discuss an alternative paradigm: by making the physical model auto-differentiable we can construct a neural surrogate to represent the unknown physical quantity sought, while avoiding having to relearn the known physics entirely. We dub this process surrogate training embedded in physics (STEP) and illustrate that it generalizes well and is robust against overfitting and significant noise in the data. We demonstrate how STEP can be applied to perform dynamic kernel deconvolution to analyse resonant inelastic X-ray scattering spectra and show that surprisingly simple estimator architectures suffice to extract the relevant physical information.

1. Introduction

In modern science, complex integrated experiments are a key tool for discovery [1–3]. They allow researchers to probe phenomena that would otherwise be inaccessible but provide only indirect or integrated measurement data. Therefore, the extraction of quantities of interest from these data constitutes a significant and important challenge in its own right. Explicitly inverting complex models for integrated experiments is often computationally prohibitive or ineffective, particularly in a low-data, high-noise regime. While machine learning (ML) tends to perform very

well for such inverse problems [4–6], it frequently struggles with accuracy when simply used as an end-to-end replacement for the inverse model. A key problem with such end-to-end approaches is the loss of physical information encoded in existing models. This information has to be captured directly by the ML estimator, leading to massively increased estimator complexity to account for lost inductive bias.

To address this challenge, we describe an approach that combines existing physical models with ML in a process we dub surrogate training embedded in physics (STEP). In STEP, we explicitly choose a separation of the components of the physical model into those assumed to be known *a priori*, and those assumed to be unknown. We keep the known components of the model and introduce an ML estimator to act as a surrogate for the unknown components. Training is performed by evaluating the loss on the total model output and propagating it back to the estimator through the known physics. We repeat this process until convergence. The estimators thus act as surrogates of components and their results have a natural interpretation as the mathematical best fit of the parameter with respect to the data. This approach sidesteps the common issue of interpretability of ML estimators in the sciences; since the inductive bias is determined by the known physical model, the ML component can be much reduced in complexity, and in most cases act as a simple regression on the space of quantities we wish to interpret. In the simplest case where the desired quantities can be parametrized as scalars, this approach simplifies the well-known parameter fitting via gradient descent [7]. For more complex cases, one can seek to extract more complex information from the data (one-dimensional functions, two-dimensional maps, functionals, etc.), in which case ML models with higher expressive power are required, for example, feed-forward neural networks (FFNNs) or convolutional neural networks (CNNs).

STEP neural networks are significantly more constrained in complexity compared with an end-to-end estimator for the same task, as they need only model a subset of the given problem—the unknown—rather than the entire process. The combination of reduced model complexity and correct (assumed) inductive bias provided by the known components of the model directly leads to robustness against data paucity and low signal-to-noise ratio (SNR). In addition, this approach minimizes the importance of hyperparameter optimization and the choice of ML architecture mode generally. This reduced complexity is a key distinction between STEP and related methods [8,9] including physics-informed neural networks (PINNs) [10] and partial differential equation (PDE) solvers [11,12], which similarly use differentiable physics but to a different end. The advantages of STEP stem from being able to identify a viable, accurate and computationally tractable model to describe at least some aspect of the inverse problem being studied. This may not always be possible, or indeed desirable, for some applications. However, for problems where small signals are buried in large integrated datasets and most core relationships between the parameters are well understood, albeit complex, in terms of physical law, it provides a way to maximize the amount of information that can be extracted from sparse data sources with poor SNR. Such problems are, unfortunately, common within the areas of nuclear fusion research, particle physics exploration and the spectroscopic probing of quantum systems, to name just a few. With this caveat, it is worth highlighting that STEP generalizes not only to many physical systems (so long as they have differentiable models) but also to different properties within each model, as we can change which components are considered known and unknown depending on the property we are interested in.

STEP and STEP-like approaches have recently gained traction in robotics [13–18] and in quantum chemistry [19–21]. In robotics, STEP can be seen as an alternative to reinforcement learning [22,23] with many similarities shared between the two approaches, whereas the application to quantum chemistry occurs primarily because of the difficulty of inverting processes in this field. Particularly in the latter case, work on this subject therefore combines highly complex physical models such as density functional theory with STEP [20,21], which leads to a requirement for subject-specific expertise to understand the role of ML within the paradigm. However, possible applications of STEP are by no means restricted to these particular fields. Instead, this method may be seen as a potential tool wherever a differentiable physical model describing the relevant integrated experiment exists. We therefore aim to illustrate how the benefits of this method, such as low computational complexity and robustness against noise and data paucity [21], may be obtained for a very different physical system.

In the following, we will illustrate these advantages in the case of artificial resonant inelastic X-ray scattering (RIXS) data generated at different SNR levels from real X-ray free electron laser (XFEL) pulses. Interestingly, the RIXS forward process includes a fairly involved convolution where the kernel varies for each data point [24], making inversion a particularly difficult deconvolutional task. Success here therefore indicates generalizability to deconvolutions more generally. We will begin by providing a general mathematical description of STEP, before introducing the specific model for

RIXS in XFEL experiments. We then proceed to the simplest application of experimental significance, which is a parameter fitting of a single scalar—in this case, the temperature—which we extract from the integrated dataset. Using the same model and set-up, we then show how this capability can be extended to fit more complex unknowns, such as the electronic structure as measured via its density of states, a continuous function in one dimension. Importantly, in moving between the two cases we make no changes in the overall underlying models: we simply change the quantities that we assume to be known and those that are unknown and thus need to be solved for in the inverse problem. Finally, we contrast this approach with an end-to-end estimator based on CNNs within a noisy, low-data regime. The STEP method outperforms the CNN significantly in our analysis. Note that it may be possible to find an end-to-end approach that matches STEP performance, but that the existence of such a model is uncertain. Furthermore, the search for such an estimator, and its ideal hyperparametrization, is expected to be highly time consuming, primarily because full inversion of the RIXS process (which constitutes a contraction map) under noise is an ill-posed problem [25]. This is also the primary reason for the dearth of non-ML solutions to this task [26]. We contrast this with the comparatively simple implementation of STEP, which is a well-defined task, as it explicitly avoids full inversion.

2. The STEP paradigm

2.1. Inverse problems with physical machine learning

Assume some known physical process described by a model P, which takes a set of N parameters (scalars, functions or functionals) B_i with $i = 1, \ldots, N$ as an input and returns another function A: $x \to A(x)$, i.e. it itself is a functional which may be written as

$$A(x) = P[B_1, \ldots, B_N](x)\,. \tag{2.1}$$

Here, $A(x)$ could, for example, be the scattered spectrum from a material sample at frequency x, while B_i may represent material properties, incoming light spectra, line shapes or other properties that affect the measured spectrum. Now consider the case where we want to extract the unknown parameter B_i: $y \to B_i(y)$ for some i and where A, P and B_j for any $j \neq i$ are known. Mathematically, we may look for an inverse functional P_i^{-1} with respect to B_i such that

$$B_i(y) = P_i^{-1}[A, B_1, \ldots, B_{i-1}, B_{i+1}, \ldots, B_N](y), \tag{2.2}$$

so long as such an inverse exists and is unique. To construct a complete inverse map, we use an estimator to approximate P_i^{-1} by $\widetilde{P}_i^{-1}$ and obtain an estimate for B_i as

$$\widetilde{B}_i(y) = \widetilde{P}_i^{-1}[A, B_1, \ldots, B_{i-1}, B_{i+1}, \ldots, B_N](y)\,. \tag{2.3}$$

While this is a highly successful approach [5,6], particularly in image analysis [4], it also relies on the approximate inversion of the potentially highly complex and generally known process P. Therefore, non-trivial design choices are often made regarding the estimator [4–6], as it has to invert a potentially highly complex process, as illustrated in figure 1 on the right-hand side. In many cases, this inversion may be mathematically ill-conditioned. Furthermore, this method does not necessarily generalize beyond the domain of the data it was trained on (i.e. it may not apply to all B_i), making predictions outside the range of training values for A and B_i non-trustworthy. Finally, we will also require many distinct data pairs A and B_i for an estimator to learn the general inverse P^{-1}, data which may not be readily available.

2.2. STEP inversion

To address all these issues (but possibly at the cost of longer computational times), we may instead apply the STEP method. Here, we do not use an estimator to approximate the highly complex object P^{-1}, but rather to estimate the generally much simpler unknown parameter B_i as $\widetilde{B}_i^{(STEP)}$. We then have

$$\widetilde{A}(x) = P[B_1, \ldots, \widetilde{B}_i^{(STEP)}, \ldots, B_N](x), \tag{2.4}$$

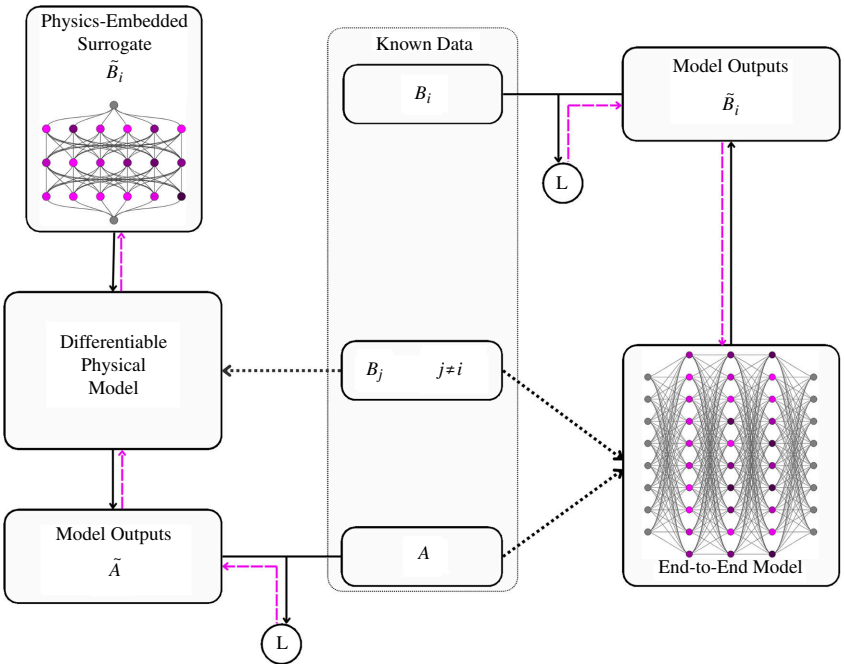

Figure 1. Comparison between STEP (left) and end-to-end ML schemes (right) with the data provided indicated in the middle. Dotted arrows represent features acting as inputs to models, black arrows signify forward passes and magenta dashed arrows show backpropagation. The two nodes labelled *L* indicate the evaluation of a loss function between the respective properties. Note that for inverse problems, the notions of feature and label differ between end-to-end and STEP paradigms, with the label from end-to-end approaches being represented as the learnable physical object in STEP instead.

and may compute some loss $\mathcal{L}(A, \tilde{A})$ to measure our estimate's performance on the data, as illustrated in figure 1 on the left. As long as the process P is known, mathematically differentiable and implemented in an automatically differentiable way, we can then use backpropagation and gradient descent to make $\tilde{B}_i^{(STEP)}$ approximate the underlying B_i with high accuracy. Additionally, in the physical case, we can define P and B_j on the one hand and B_i on the other hand to represent known and unknown processes, respectively, constituting an important split that allows us to focus on discovering physical unknowns B_i rather than inverting well-understood processes P. Finally, STEP lets us generalize to arbitrary B_i with confidence and ensures that $\tilde{B}_i^{(STEP)}$ always has an intuitive and meaningful interpretation as the optimized estimate of B_i, two important features in physical ML.

2.3. Limitations

There are limitations to consider when using STEP. First and foremost, it relies on an accurate understanding of the process P, and the ability to implement this process in an automatically differentiable way. However, for integrated experiments in physics, the former is in fact the underlying assumption in measurement routines. For automatic differentiation of P, we may leverage the plethora of tools for differentiable physical programming [19,27–33] developed in the context of supervised learning. Built within the constantly expanding ecosystems of e.g. PyTorch [34] and JAX [35], these tools enable full differentiability in physical models by reducing intractable computational graphs and non-differentiable effects into feasibly differentiable forms using automatic differentiation (e.g. autograd [36]). Notably, such frameworks allow for the use of gradient descent with backpropagation through physical systems and provide the capability to train differentiable ML surrogates directly embedded in such models, thereby enabling the use of STEP. Note that additional care has to be taken to avoid vanishing gradients in the backpropagation.

Second, the use of STEP may incur a significant increase in computational time if a very large number of distinct B_i has to be extracted, as it relies on repeated fittings rather than simple evaluations of the approximate inverse maps following pre-training. This issue has become less significant with the continuing rise in available computing power. Furthermore, in scientific exploration, accuracy nearly

always supplants fitting time as the key performance metric: finding an unreliable result quickly is often of little practical use.

3. STEP application to resonant inelastic X-ray scattering

3.1. Resonant inelastic X-ray scattering process

RIXS is among the most widely used spectroscopic techniques to study the electronic structure of materials and probe elementary excitations in complex systems by measuring their energy, momentum and polarization dependence [37]. Recent applications to high energy density physics [24] show further promise of applying this technique to matter in extreme conditions including planetary physics, astrophysics and inertial confinement fusion research [38,39]. The intensity of scattered radiation at discrete energies $\omega_{2,i}$ as measured with an energy resolution of ΔE is given by the integral

$$I(\omega_{2,i}) \propto \int_{\omega_{2,i}-\Delta E/2}^{\omega_{2,i}+\Delta E/2} d\omega_2 \;\; \partial_{\omega_2}\sigma, \tag{3.1}$$

where $\partial_{\omega_2}\sigma$ is the RIXS scattering cross-section, indicative of the amount of light scattered from the material to a particular energy ω_2. Under certain conditions [40], the RIXS process dominates the various scattering channels, and its cross-section can be written as a sum over shifted and weighted one-dimensional convolutions [24],

$$\partial_{\omega_2}\sigma = \sum_f L_f(\omega_2) \int_{-\infty}^{\infty} d\omega_1 \Phi(\omega_1)\rho_{\text{eff}}(\omega_1 - \omega_2 + \epsilon_f)\,. \tag{3.2}$$

Here, ω_1 is the angular frequency and $\Phi(\omega_1)$ is the spectral pulse shape of the incoming X-ray pulse (i.e. the kernel), and L_f, ϵ_f are known material-dependent parameters, while ρ_{eff} is the effective density of states (DoS) for the material, defined as

$$\rho_{\text{eff}}(\Delta) = \rho(\Delta)|M|^2 f_{FD}(\Delta; T)\,. \tag{3.3}$$

Both the DoS and the temperature T which enters equation (3.3) through f_{FD} are notoriously difficult to extract, as RIXS does not constitute a pure convolution with incoming pulses. Furthermore, this process is generally studied at XFEL facilities, which generate incoming pulses (kernels) $\Phi(\omega_1)$ from noise, and thus have a large shot-to-shot variance and irregular, spiky profiles [41]. Finally, as RIXS cross-sections are small, the measurements typically have low signal-to-noise. These complications make non-ML methods for deconvolution, such as the Richardson–Lucy method, and tools such as TomoPy [42], or MANTiS [43], intractable for RIXS analysis, constituting a bottleneck for spectroscopy in high energy density physics applications [26]. To implement STEP for RIXS, we first ensure that the forward model is programmed in a completely differentiable manner. We furthermore modify backwards passes to avoid the vanishing gradient problem by omitting exponentially small factors, specifically the Lorentzian and thermal suppression factors (seen in electronic supplementary material, appendix B).

3.2. Data and objectives

To test model performances on RIXS, we generated artificial noisy data using the following forward model:

- Generate artificial modulated DoS ($\rho' = \rho|M|^2$) designed to resemble the real DoS (Gaussian energy bands and an optional square-root continuum).
- Weigh DoS contributions by thermal factor f_{FD} obtained from temperature T.
- Evaluate $I(\omega_{2,i})$ using real XFEL pulses as $\Phi(\omega_1)$, real material parameters from iron (see electronic supplementary material, SM) and the weighted artificial DoS.
- Add Gaussian noise to the obtained spectra $I(\boldsymbol{\omega_2})$, where $\boldsymbol{\omega_2} = (\omega_{2,1}, \ldots, \omega_{2,M})$ is the set of measurement energies and M is the number of sampling points. The noise is distributed with a standard deviation of $\sigma = \epsilon \cdot \max(I(\boldsymbol{\omega_2}))$ for $\epsilon = 0, 0.1, 0.2, 0.3$.

Resultant spectra for different combinations of DoS and XFEL pulses are shown in figure 2. We evaluate model performances with the mean-squared error loss between a vector quantity $\mathbf{A}$ and its estimate $\tilde{\mathbf{A}}$

$$\mathcal{L}(\mathbf{A}, \tilde{\mathbf{A}}) = \frac{1}{N}\sum_{i=1}^{N}\left(A_i - \tilde{A}_i\right)^2. \tag{3.4}$$

3.3. Extracting the temperature

The simplest application of STEP to the RIXS process is the extraction of the scalar temperature T, which enters into $\rho_{\text{eff}}(\Delta)$ through the Fermi–Dirac distribution evaluated at energy Δ,

$$f_{FD}(\Delta; T) = \frac{1}{e^{(\Delta-\mu)/k_B T}+1}. \tag{3.5}$$

Here, μ denotes the chemical potential, which depends on the temperature T as described in electronic supplementary material, appendix D, while k_B is the Boltzmann constant. We can now characterize all parameters of the model except for T as known parameters, i.e. they have a numerical value or functional form which we may assume to be exact. In the case of temperature extraction, we assume that all other parameters are known, including the DoS at 0K ρ, which can be obtained e.g. by density functional theory calculations.

Explicitly using the notation defined above, we may then write

$$I_k(\omega_{2,i}) = P'_{\text{RIXS}}[\Phi_k, \rho', T](\omega_{2,i}). \tag{3.6}$$

This expression can be differentiated with respect to either $\rho' = \rho|M|^2$ or T, therefore admitting the STEP scheme.

Applying this method to the set of six different synthetic DoS generated as described in §3.2 with different levels of noise and 50 different XFEL pulses, we found excellent predictions of temperature independent of noise as illustrated in figure 3. Convergence was achieved after less than 2000 epochs each, depending on the initial random value of temperature. This highlights basic functioning of STEP for scalars, the low computational complexity of STEP and robustness against noise.

3.4. Extracting the density of states

Let us now consider the more complex extraction of the DoS function from RIXS data using STEP. Note that we here assume that all other parameters including the temperature are known. The modified process functional P_{RIXS} now takes ρ_{eff} and the kernel Φ_k for XFEL pulse k as inputs to yield.

$$I_k(\omega_{2,i}) = P_{\text{RIXS}}[\Phi_k, \rho_{\text{eff}}](\omega_{2,i}). \tag{3.7}$$

This overall framework holds for any dynamic kernel convolution, and to invert it we would have to find an inverse process P^{-1}_{RIXS} such that

$$\rho_{\text{eff}}(\Delta) = P^{-1}_{\text{RIXS}}[I_k, \Phi_k](\Delta), \qquad \Delta = \omega_1 - \omega_2 + \epsilon_f, \tag{3.8}$$

for any k and ρ_{eff}. Instead, we can use a feed-forward neural surrogate with four hidden layers and 40 nodes each and softplus activation function to directly generate an estimate $\tilde{\rho}_{\text{eff}}$ using STEP. Note that we use a neural network rather than scalar fittings to ensure continuity and smoothness of $\tilde{\rho}_{\text{eff}}$. This estimate is then used in place of ρ_{eff} and trained iteratively using gradient descent and backpropagation via PyTorch's autograd [36]. The loss used for training is the mean squared error (MSE) loss between artificially measured and estimated intensities $\mathcal{L}(I(\boldsymbol{\omega_2}), \tilde{I}(\boldsymbol{\omega_2}))$. We train the neural surrogate on individual DoS using 50 XFEL pulses each, using batches of eight samples as well as the ADAM [44] optimizer, and achieve convergence after 10 000 epochs (see electronic supplementary material, SM). This computation takes 17 min per DoS on an AMD Ryzen 5 3500 u CPU. As can be seen in figure 4, the STEP reconstructions closely match the true DoS even at significant noise levels of $\varepsilon = 1$.

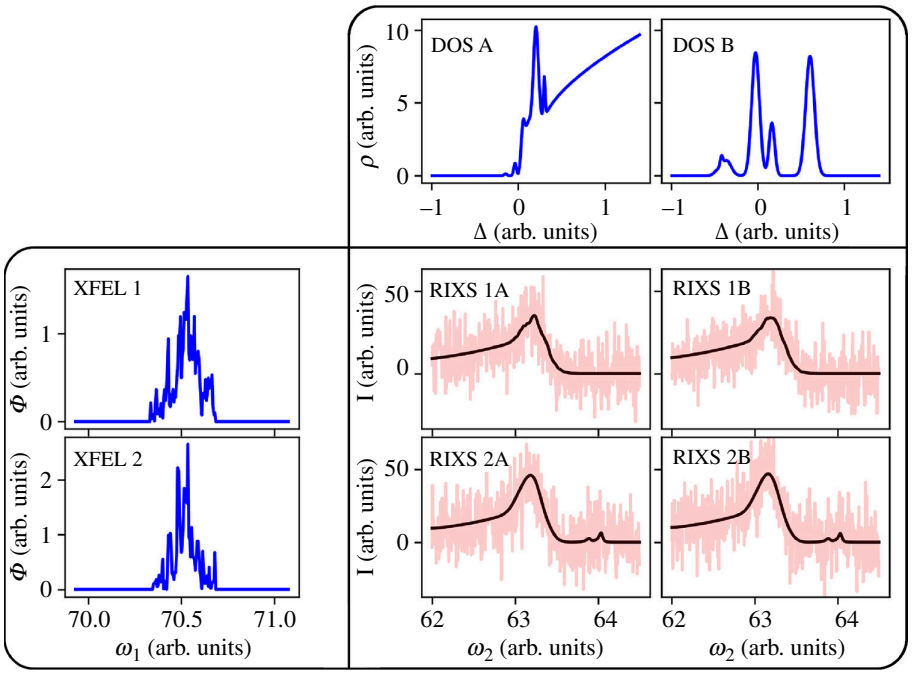

Figure 2. An illustration of the artificial data generated. The real XFEL pulses 1 and 2 with intensities Φ (spectra shown in left column) are used as incoming light into simulated RIXS scattering processes in materials with DoS A and B (top row). The four panels in the bottom right show the simulated scattering spectra with intensities I before (black) and after (red) noise is applied.

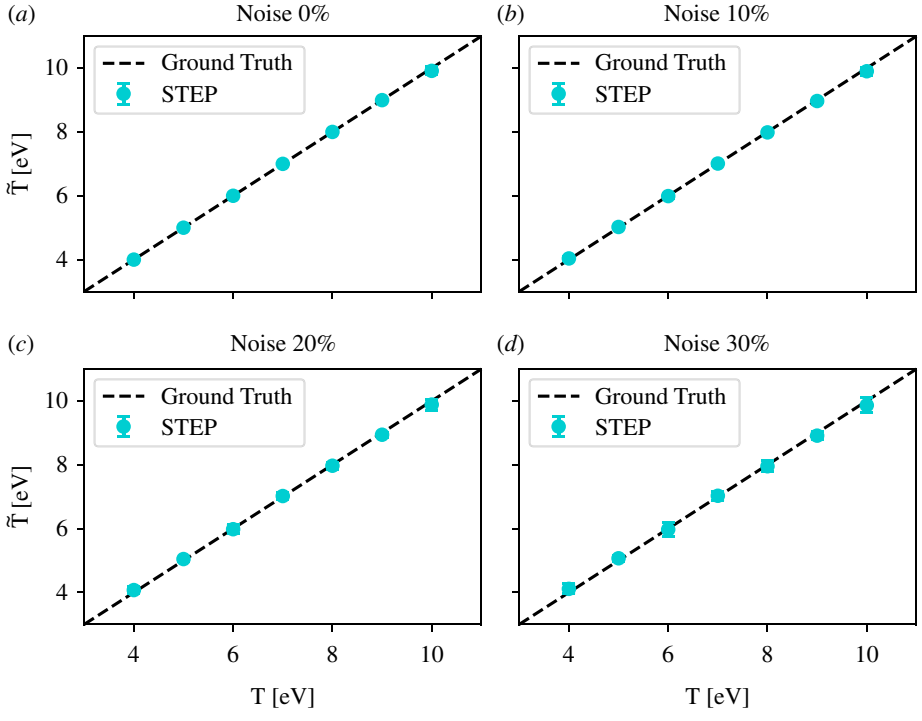

Figure 3. Temperatures extracted from RIXS measurements plotted against true temperatures. Panels (*a*) to (*d*) correspond to noise levels of 0%, 10%, 20% and 30%, respectively. Note that all values are plotted with error bars, most of which are too small to be seen.

4. Comparison against an end-to-end convolutional neural network

Instead of using STEP, we can estimate the process function defined in equation (3.8) using a custom CNN, similar to state-of-the-art architectures for deconvolution [4]. The CNN is designed to predict different DoS from pairs of XFEL and RIXS spectra and is trained on a correspondingly large dataset. The advantage of this process is the additional speed gained by only training the network once, and

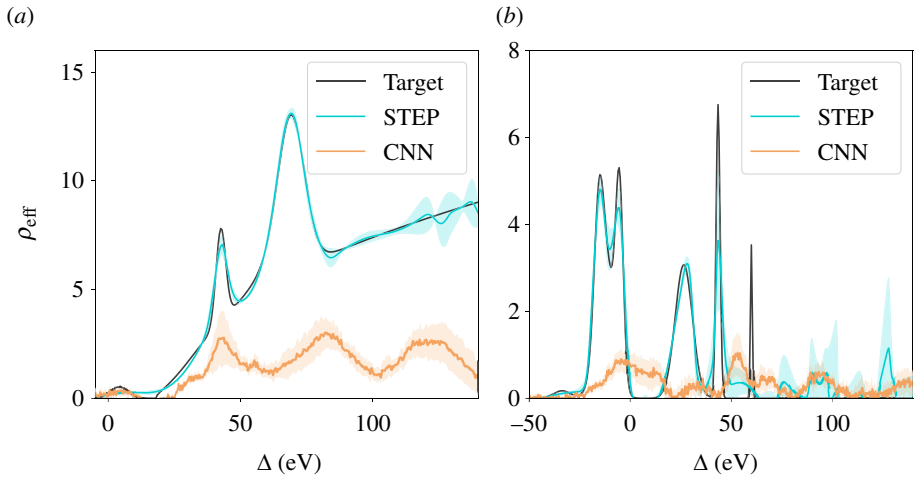

Figure 4. The results of both STEP and CNN methods applied to two different DoS in the test set. Shaded areas indicate standard deviations with respect to different pulses and different random fitting seeds for CNN and STEP, respectively. Panel (*a*) shows the best and panel (*b*) the worst *qualitative* reconstruction by the CNN. Quantitively this relation is inverted, as the underestimation of the continuum contribution constitutes a substantial error in panel (*a*), while the CNN's overall tendency to underpredict features allows for relatively low error for the sparser DoS shown in panel (*b*). Notice that the STEP method is able to reconstruct features much narrower than the XFEL bandwidth (around 25 eV).

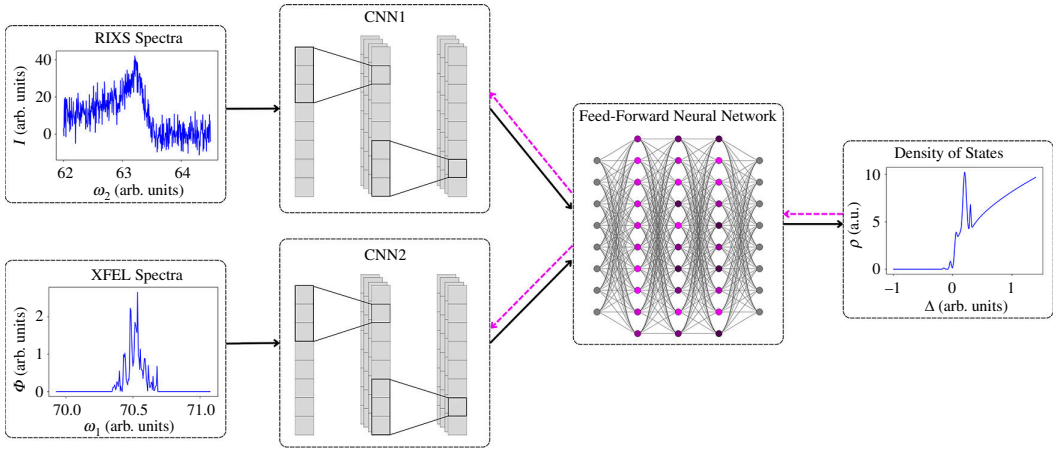

Figure 5. The end-to-end CNN-based architecture is described in §4. Note that black arrows indicate forward passes, while magenta dashed arrows indicate backpropagation. The RIXS and XFEL spectra are used as features, while the density of states acts as a label.

the ability to apply it to data for any given density of states without further training. However, the data requirements are also correspondingly substantial, and we trained the CNN on a set consisting of 140 artificial DoS, and 50 XFEL pulses each (7000 samples). Notably, the DoS span a large space of possible functions, and half of them (70) have a square-root continuum contribution while the other half does not. Training was performed for 5000 epochs, taking 1 h on an AMD Ryzen 5 3500 u CPU. Longer training times lead to overfitting and were therefore avoided.

The end-to-end network used in our research is designed to capture information from two correlated one-dimensional signals, which have internal spatial ordering, but exist on different axes. The chosen architecture is illustrated in figure 5 and consists of two CNNs whose outputs are fed into a joint FFNN. Crucially, a change in the particular CNN architecture is not expected to improve performance, as it does not address the core problem of combining the two data signals in a natural manner. The model was trained using the MSE loss $\mathcal{L}(\rho_{\text{eff}}, \tilde{\rho}_{\text{eff}})$ as well as the ADAM [44] optimizer, and for more reliable evaluation, the mean across all XFEL-RIXS pairs for a given DoS is used for evaluation on the test set. Hyperparameters were found using 100 iterations of random search, yielding no L2 regularization, our convolutional layers with eight channels each per CNN component and a four-layer FFNN with 200 and 100 node layers to merge the two signals. The interested reader may find details in electronic supplementary material, appendix C.

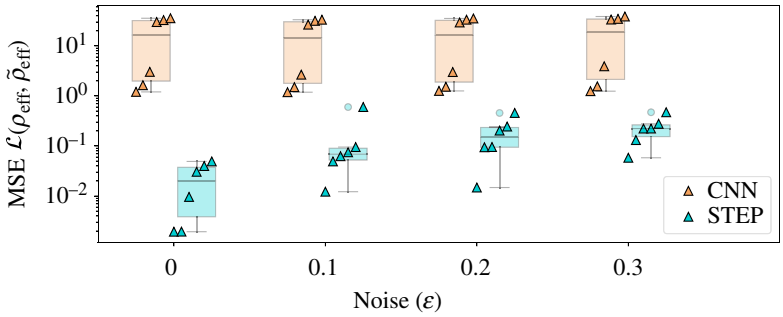

Figure 6. Quantile and box plots of the difference in MSE loss between true and predicted ρ_{eff} for both methods (CNN, STEP) and different levels of noise (ε). Note the log-scale on the *y*-axis, indicative of the large difference between the two methods, as well as the difference between different noise levels for STEP and the two clusters of varying performance for the CNN MSE.

A note is in order: we do not claim that the CNN architecture used here is necessarily the best non-STEP architecture possible. However, we stress that the choice of estimator is not obvious and is generally challenging and time-consuming to find. Additionally, we may note that the RIXS process constitutes a contraction map, and that a full inversion therefore constitutes an ill-posed problem. Particularly in the presence of noise, this makes a full inversion ineffective, no matter which particular ML architecture or indeed non-ML method is chosen [25]. There are no such challenges within the STEP approach, where the architecture needs only to be able to reproduce the function we seek to represent. The required complexity is so low that hyperparametrization becomes trivial. Furthermore, because there is a one-to-one correspondence between ρ_{eff} and the material under investigation, and we only need to represent ρ_{eff} and not the entire physics-based model, there is also no need to construct a workable inductive bias into the NN architecture, nor is there a need for large training datasets.

4.1. Results

We evaluate the performance of both STEP and CNN approaches across six distinct ρ_{eff}, three with and three without square-root contributions. While STEP is trained on each DoS individually, the CNN was trained on a distinct set of 140 DoS and then evaluated on the test set as indicated above. We first did this for a noise level of $\epsilon = 0.1$. Here, STEP managed to converge to each DoS with high accuracy, only missing very narrow peaks and exhibiting growing standard deviations with respect to the random seed in the regime $\Delta > 100$ eV. The latter is expected, as the RIXS signal in this regime is suppressed by the factor L_f (see electronic supplementary material, SM), leading to weaker regularization from the physical model. The CNN approach on the other hand struggled to converge adequately, with the best and worst *qualitative* performance across the test set for these conditions shown in figure 4. As seen in figure 4*a*, the CNN manages to qualitatively identify peaks for some DoS and even identifies a bulk of the DoS corresponding to the continuum. However, it also significantly underestimates the amplitude of any peak and misinterprets the continuum to consist of another, broader spike. In figure 4*b*, it clearly struggles to identify any of the peaks with any reliability, instead predicting the majority of the DoS to hover near zero, in order to minimize penalties for incorrectly predicted Gaussian peaks. This seems to indicate that the estimator struggles with more complex and narrower structures in the DoS, representing a failure to generalize.

The difference in performance becomes even more evident when investigated across all six test DoS and different noise levels, as seen in figure 6. Note the log-scale chosen for the loss plot in this figure. Interestingly, there is also a clear split in the loss of the CNN on the different DoS of the test set. This can be explained when considering the different shapes of DoS in the test set. While the CNN was better at qualitatively extracting the shape for DoS with continuum contributions, it quantitatively performed better on the data without square-root continuum, as it could minimize penalties by guessing near zero across all values. While the STEP method performs over an order of magnitude better for the noiseless case, its error increases with growing noise. This effect is still remarkably small, but can further be mitigated by simply including some additional data points, when available. Overall, it is clear that STEP performs much better than the CNN and exhibits remarkable noise resilience and generalizability.

5. Conclusion

We gave a formal description of the STEP paradigm, which has recently emerged in physical ML, and show how it can be applied to inverse problems for partially known physics. We have further illustrated the benefits of this technique, including that it naturally generalizes, is interpretable and robust against overfitting. We contrasted this against end-to-end approaches, which are prevalent but come with their own challenges. Additionally, we demonstrate how the results such as those by Kasim and Vinko [21] and Li *et al*. [20] for DFT extend to physical problems in entirely different regimes, such as dynamic kernel deconvolution for RIXS. Crucially, little amounts of noisy data and surprisingly simple estimators suffice under the STEP paradigm for experimental analysis. We believe that this feature makes STEP a suitable tool across physical experiments, and appealingly it can be applied in post-analysis to experimental data already collected. The primary requirement remains the differentiable implementation of a physical process, the overhead for which decreases with the rapid development of better libraries for differentiable modelling. Overall, this paradigm shows great potential for application anywhere where underlying quantities are to be extracted from known measurement schemes.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All data and code are available under [45].

Supplementary material is available online [46].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. K.K.A.E.-D.: data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft, writing—review and editing; A.F.: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing—review and editing; M.F.K.: conceptualization, resources, writing—review and editing; F.M.: conceptualization, resources, writing—review and editing; S.M.V.: conceptualization, project administration, resources, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. The authors have received funding from the Royal Society, the UK EPSRC grant EP/W010097/1, and the UK STFC XFEL Hub. A per author breakdown can be found under acknowledgements.

Acknowledgements. K.K.A.E.-D., A.F., F.M. and S.M.V. acknowledge support from the Royal Society and from the UK EPSRC grant EP/W010097/1. A.F. acknowledges support from the UK STFC XFEL Hub. S.M.V. is a Royal Society University Research Fellow.

References

1. McNeil BWJ, Thompson NR. 2010 X-ray free-electron lasers. *Nat. Photon.* **4**, 814–821. (doi:10.1038/nphoton.2010.239)
2. Aad G *et al*. 2013 Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Phys. Lett. B* **726**, 88–119. (doi:10.1016/j.physletb.2013.08.010)
3. Roland Garoby AV *et al*. 2017 The European spallation source design. *Phys. Scr.* **93**, 014001. (doi:10.1088/1402-4896/aaecea)
4. Jin KH, McCann MT, Froustey E, Unser M. 2017 Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **26**, 4509–4522. (doi:10.1109/TIP.2017.2713099)
5. Stielow T, Scheel S. 2021 Reconstruction of nanoscale particles from single-shot wide-angle free-electron-laser diffraction patterns with physics-informed neural networks. *Phys. Rev. E* **103**, 053312. (doi:10.1103/PhysRevE.103.053312)
6. Montes-Campos H, Carrete J, Bichelmaier S, Varela LM, Madsen GKH. 2022 A differentiable neural-network force field for ionic liquids. *J. Chem. Inf. Model.* **62**, 88–101. (doi:10.1021/acs.jcim.1c01380)
7. Thuerey N, Holl P, Mueller M, Schnell P, Trost F, Um K. 2022 *Physics-based deep learning*. (doi:10.48550/arXiv.2109.05237)
8. Amos B, Kolter JZ. 2017 Optnet: differentiable optimization as a layer in neural networks. In *Proc. of the 34th Int. Conf. on Machine Learning*, vol. **70**, pp. 136–145, PMLR. https://proceedings.mlr.press/v70/amos17a/amos17a.pdf.
9. Kang R, Kyritsis DC, Liatsis P. 2022 Self-validated physics-embedding network: a general framework for inverse modelling. *arXiv*. See https://arxiv.org/abs/2210.06071.
10. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. 2021 Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440. (doi:10.1038/s42254-021-00314-5)
11. Um K, Brand R, Yun (Raymond) F, Holl P, Thuerey N. 2020 Solver-in-the-loop: learning from differentiable physics to interact with iterative PDE-solvers. *Adv. Neural Inf. Process. Syst.* **33**, 6111–6122.
12. Kochkov D, Smith JA, Alieva A, Wang Q, Brenner MP, Hoyer S. 2021 Machine learning-accelerated computational fluid dynamics. *Proc. Natl Acad. Sci. USA* **118**, e2101784118. (doi:10.1073/pnas.2101784118)

13. Toussaint MA, Allen KR, Smith KA, Tenenbaum JB. 2018 Differentiable physics and stable modes for tool-use and manipulation planning. In *Proc. of Robotics: Science and Systems, Pittsburgh, PA, June*. (doi:10.15607/RSS.2018.XIV.001)
14. Degrave J, Hermans M, Dambre J, Wyffels F. 2019 A differentiable physics engine for deep learning in robotics. *Front. Neurorobot.* **13**, 6. (doi:10.3389/fnbot.2019.00006)
15. de Avila Belbute-Peres F, Smith K, Allen K, Tenenbaum J, Kolter JZ. 2018 End-to-end differentiable physics for learning and control. In *Advances in neural information processing systems* (eds S Bengio, H Wallach, H Larochelle, K Grauman, N Cesa-Bianchi, R Garnett), vol. **31**. New York, NY: Curran Associates, Inc.
16. Ma P, Du T, Zhang JZ, Wu K, Spielberg A, Katzschmann RK, Matusik W. 2021 DiffAqua. *ACM Trans. Graph* **40**, 1–14. (doi:10.1145/3450626.3459832)
17. Qiao YL, Liang J, Koltun V, Lin MC. 2020 Scalable differentiable physics for learning and control. In *37th Int. Conf. on Machine Learning, ICML, 13–18 July 2020*, Virtual, vol. **119**, pp. 7847–7856, PMLR.
18. Suh HJT, Simchowitz M, Zhang K, Tedrake R. 2022 Do differentiable simulators give better policy gradients? In *39th Int. Conf. on Machine Learning, ICML*, 17–23 July 2022, vol. **162**, pp. 20668–20696, PMLR.
19. Schoenholz S, Cubuk ED. 2020 JAX MD: a framework for differentiable physics. *Adv. Neural Inf. Process. Syst.* **33**, 11 428–11 441.
20. Li L, Hoyer S, Pederson R, Sun R, Cubuk ED, Riley P, Burke K. 2021 Kohn-Sham equations as regularizer: building prior knowledge into machine-learned physics. *Phys. Rev. Lett.* **126**, 036401. (doi:10.1103/PhysRevLett.126.036401)
21. Kasim MF, Vinko SM. 2021 Learning the exchange-correlation functional from nature with fully differentiable density functional theory. *Phys. Rev. Lett.* **127**, **9**. (doi:10.1103/PhysRevLett.127.126403)
22. Kaelbling LP, Littman ML, Moore AW. 1996 Reinforcement learning: a survey. *J. Artif. Intell. Res.* **4**, 237–285. (doi:10.1613/jair.301)
23. Li Y. 2017 Deep reinforcement learning: an overview. *arXiv*. See https://arxiv.org/abs/1701.07274.
24. Humphries OS *et al*. 2020 Probing the electronic structure of warm dense nickel via resonant inelastic X-ray scattering. *Phys. Rev. Lett.* **125**, 195001. (doi:10.1103/PhysRevLett.125.195001)
25. Forte A, et al. 2024 Resonant Inelastic X-ray scattering in warm-dense Fe compounds beyond the SASE FEL resolution limit. *arXiv*. (doi:10.48550/arXiv.2402.00039)
26. Kayser Y *et al*. 2019 Core-level nonlinear spectroscopy triggered by stochastic X-ray pulses. *Nat. Commun.* **10**, 4761. (doi:10.1038/s41467-019-12717-1)
27. Innes M, Edelman A, Fischer K, Rackauckas C, Saba E, Shah VB, Tebbutt W. 2019 A differentiable programming system to bridge machine learning and scientific computing. *arXiv*. (doi:10.48550/arXiv.1907.07587)
28. Freeman CD, Frey E, Raichuk A, Girgin S, Mordatch I, Bachem O. 2021 A differentiable physics engine for large scale rigid body simulation. *arXiv*. (doi:10.48550/arXiv.2106.13281)
29. Hu Y, Anderson L, Li TM, Sun Q, Carr N, Ragan-Kelley J, Durand F. 2020 DiffTaichi: differentiable programming for physical simulation. In *Int. Conf. on Learning Representations, ICLR*, Virtual, 26–31 April 2020. (doi:10.48550/arXiv.1910.00935)
30. Zhang X, Chan GKL. 2022 Differentiable quantum chemistry with PySCF for molecules and materials at the mean-field level and beyond. *J. Chem. Phys.* **157**, 7. (doi:10.1063/5.0118200)
31. Kasim MF, Vinko SM. Ξ-Torch: differentiable scientific computing library. *arXiv*. (doi:10.48550/arXiv.2010.01921)
32. Kasim MF, Lehtola S, Vinko SM. 2022 DQC: A Python program package for differentiable quantum chemistry. *J. Chem. Phys.* **156**, 084801. (doi:10.1063/5.0076202)
33. Geilinger M, Hahn D, Zehnder J, Bächer M, Thomaszewski B, Coros S. 2020 ADD: Analytically differentiable dynamics for multi-body systems with frictional contact. *ACM Trans. Graph* **39**, 1–15. (doi:10.1145/3414685.3417766)
34. Paszke A et al. 2019 Pytorch: an imperative style, high-performance deep learning library (eds HM Wallach, H Larochelle, A Beygelzimer, F d'Alché-Buc, EB Fox). In *Advances in neural information processing systems*, vol. **32**, pp. 8026–8037, 57 Morehouse Lane, Red Hook, NY, United States: Curran Associates Inc.
35. Bradbury J *et al*. 2018 JAX: Composable transformations of Python+Numpy programs. See http://github.com/google/jax.
36. Paszke A *et al*. 2017 Automatic differentiation in PyTorch. In *NIPS Workshop Autodiff, 9 December 2017*, Long Beach, California, USA.
37. Ament LJP, van Veenendaal M, Devereaux TP, Hill JP, van den Brink J. 2011 Resonant inelastic X-ray scattering studies of elementary excitations. *Rev. Mod. Phys.* **83**, 705–767. (doi:10.1103/RevModPhys.83.705)
38. Eggert JH, Hicks DG, Celliers PM, Bradley DK, McWilliams RS, Jeanloz R, Miller JE, Boehly TR, Collins GW. 2010 Melting temperature of diamond at ultrahigh pressure. *Nat. Phys.* **6**, 40–43. (doi:10.1038/nphys1438)
39. Gaffney JA *et al*. 2018 A review of equation-of-state models for inertial confinement fusion materials. *High Energ. Dens. Phys.* **28**, 7–24. (doi:10.1016/j.hedp.2018.08.001)
40. Sturm K. 1993 Dynamic structure factor: an introduction. *Z. Nat.* **48**, 233–242. (doi:10.1515/zna-1993-1-244)
41. Brau CA. 1990 Free-electron lasers. In Physics of particle accelerators, Fermilab/Cornell University, pp. 1615–1706, vol. **184**. San Diego, California, USA: AIP Publishing. (doi:10.1063/1.38022)
42. Gürsoy D, De Carlo F, Xiao X, Jacobsen C. 2014 TomoPy: a framework for the analysis of synchrotron tomographic data. *J. Synchrotron Radiat.* **21**, 1188–1193. (doi:10.1107/S1600577514013939)
43. Lerotic M, Mak R, Wirick S, Meirer F, Jacobsen C. 2014 MANTiS: a program for the analysis of X-ray spectromicroscopy data. *J. Synchrotron Radiat.* **21**, 1206–1212. (doi:10.1107/S1600577514013964)

44. Kingma DP, Ba JL. 2015 Adam: a method for stochastic optimization. In *3rd Int. Conf. on Learning Representations, ICLR 2015 - Conference Track Proceedings, 7–9 May 2015*. San Diego, California, USA.
45. OxfordHED. *University of Oxford High Energy Density group*. GitHub. See https://github.com/OxfordHED/rixs_nn_analysis.
46. El-Din A, Kacper K, Forte A, Kasim MF, Miniati F, Vinko S. 2024 Supplementary material from: STEP: Extraction of underlying Physics with robust Machine Learning. FigShare (doi:10.6084/m9.figshare.c.7247039)