Real-time simulations of ADF STEM probe position-integrated scattering cross-sections for single element fcc crystals in zone axis orientation using a densely connected neural network

Real-time simulations of ADF STEM probe position-integrated scattering cross-sections for single element fcc crystals in zone axis orientation using a densely connected neural network

I. Lobato, A. De Backer, S. Van Aert

EMAT, University of Antwerp, Department of Physics, Groenenborgerlaan 171, B-2020 Antwerp, Belgium

NANOlab Center of Excellence, University of Antwerp, Department of Physics, Groenenborgerlaan 171, B-2020 Antwerp, Belgium

Abstract

Quantification of annular dark field (ADF) scanning transmission electron microscopy (STEM) images in terms of composition or thickness often relies on probe-position integrated scattering cross sections (PPISCS). In order to compare experimental PPISCS with theoretically predicted ones, expensive simulations are needed for a given specimen, zone axis orientation, and a variety of microscope settings. The computation time of such simulations can be in the order of hours using a single GPU card. ADF STEM simulations can be efficiently parallelised using multiple GPUs, as the calculation of each pixel is independent of other pixels. However, most research groups do not have the necessary hardware, and, in the best-case scenario, the simulation time will only be reduced proportionally to the number of GPUs used. In this manuscript, we use a learning approach and present a densely connected neural network that is able to perform real-time ADF STEM PPISCS predictions as a function of atomic column thickness for most common face-centered cubic (fcc) crystals (i.e., Al, Cu, Pd, Ag, Pt, Au and Pb) along [100] and [111] zone axis orientations, root-mean-square displacements, and microscope parameters. The proposed architecture is parameter efficient and yields accurate predictions for the PPISCS values for a wide range of input parameters that are commonly used for aberration-corrected transmission electron microscopes.

Keywords: ADF STEM simulations, Multem, Probe position integrated scattering cross section, Neural network, real-time, Tensorflow

1. Introduction

Scanning transmission electron microscopy (STEM) with an annular dark field (ADF) detector has become a popular technique for quantifying nanostructures at the atomic level due to the absence of contrast reversals in the recorded images with sample thickness and defocus. The quantification process can be performed through three-dimensional atomic resolution electron tomography [1], direct comparison of experimental data with image simulations [2], or by using statistical methods to extract quantitative information from the images [3]. Different methods have been developed for counting the number of atoms in each atomic column from a single ADF STEM image [3, 4, 5]. As a measure of performance for atom-counting, probe-position integrated scattering cross sections (PPISCS) are often used as they are highly sensitive to the number of atoms in a column and its composition [6, 4, 7]. Additionally, they are robust to probe parameters such as defocus and other aberrations [8, 9]. The PPISCS corresponds to the integrated intensity over the atomic feature and can be determined by integrating image intensities in Voronoi cells around the positions of the atomic features [8], or by estimating the volume under each atomic column by fitting a parametric model consisting of 2D overlapping Gaussian peaks to the experimental images [3]. From a set of

Email address: Ivan.Lobato@uantwerpen.be (I. Lobato)

PPISCSs, atoms can be counted using an image simulation-based approach [5], a statistics-based approach combined with prior knowledge of the sample thickness range [10], or by using a hybrid approach that includes prior knowledge from image simulations into the existing statistics-based method [11]. Furthermore, atom counts for each chemical element in alloy systems can be determined by combining prior knowledge with the so-called atomic lensing model, which enables the prediction of PPISCSs of mixed columns from the PPISCSs of atomic columns consisting of a single atomic element [12].

In order to obtain accurate results for atom-counting, ADF-STEM image simulations are required that include a quantum mechanical description of the electron-specimen interaction for the same specimen, zone axis orientation, and microscope settings as those used in the experiment. However, in practice, the absolute values of several parameters are unknown, such as defocus, spatial incoherence, root-mean-square displacement (rmsd), specimen thickness, mistilt, carbon contamination, and other experimental uncertainties [8]. This often leads to a mismatch between the experiment and simulation [13]. A common method for addressing this mismatch is to estimate the unknown parameters by matching simulations with experimental images. Although modern frozen phonon GPU multislice implementations of the electron-specimen interaction have reduced the computation time of ADF STEM simulations [14, 15, 16], these calculations still take several hours using a single GPU card. ADF STEM simulations can be efficiently parallelised using multiple GPUs, as the calculation of each pixel is independent of the other pixels. However, most research groups do not have access to the necessary hardware. Furthermore, in the best-case scenario, the simulation time can only be reduced proportionally to the number of GPUs used. Therefore, it is highly desirable to significantly speed up the calculation of the ADF STEM PPISCS in order to facilitate the quantification process.

Machine learning based on artificial neural networks has become a state-of-the-art method due to its ability to learn from data by adjusting the weight connections between neurons during the training process. Additionally, neural networks have demonstrated breakthrough performance for various tasks such as image recognition [17], image restoration [18], image super-resolution [19], natural language processing [20] and cognitive science [21]. The performance of the neural network is highly dependent on the quantity and quality of the available training data as well as on its sampling distribution. Furthermore, in order to obtain consistent predictions for physical systems, it is essential that the neural network learns the underlying physics constraints of the governing laws. This can be embedded in the network architecture itself [22, 23] or in its loss function [24, 25].

In this paper, we use a machine learning approach and present a densely connected neural network to perform real-time ADF STEM PPISCS-thickness simulations for the most common face-centered cubic (fcc) crystals along their main zone axis orientations, microscope parameters, and root-mean-square displacement (rmsd) values. In Section 2, we will explain the methodology, including the steps of data generation, neural network architecture, the choice of the loss function, and implementation details. In Section 3, we will present and discuss the results. Finally, in Section 4, we will draw conclusions.

2. Methodology

PPISCSs have been proven to be robust for various probe parameters such as defocus, aberrations, and temporal incoherence [8, 9] in aberration-corrected scanning transmission electron microscopes. Additionally, they remain invariant when dealing with spatial incoherence. Despite these advantages, it is still important to have accurate estimations of PPISCSs-thickness through image simulations in order to determine the number of atoms and composition in an atomic feature. It is noteworthy that, PPISCS-thickness values have non-linear dependence regarding the microscope parameters and the rmsd value. While it is possible to train a separate neural network for each fcc crystal for a given zone axis, this approach unnecesarily results in a large number of networks. It has been observed that PPISCS-thickness predictions share many similarities regardless of their atomic number and zone axis orientation. Additionally, the universal approximation theorem for neural networks, as stated in [26], states that any continuous function can be approximated by a multilayer neural network. Therefore, it is more efficient to train a single neural network for PPISCS-thickness predictions. Our results indicate that this approach also reduces the total number of parameters in the network without compromising its accuracy.

Our aim is therefore to train a neural network to perform real-time ADF STEM PPISCS-thickness simulations for the

most commonly used fcc crystals in material science, along their main zone axis orientations, microscope parameters, and rmsd values. The first step in achieving this is to describe the data generation process.

2.1. Data generation

In this work, use is made of the frozen atom multislice simulation approach implemented in the Multem software [14, 15], which was shown to yield results matching experiments [27]. Although our methodology can be applied to general fcc crystals with different zone axis orientations, the calculation time for ADF STEM PPISCS-thickness puts an upper limit on the number of fcc crystals and zone axis orientations that we can generate for our training data using our computational resources. Therefore, we have focused on the most commonly used fcc crystals in materials science along their most relevant zone axis orientations that are used for the atom-counting procedure. We considered Al, Ni, Cu, Pd, Ag, Pt, Au, and Pb fcc crystals along the [100] and [110] zone axis orientation (zao) up to 61 and 87 atoms, respectively. This corresponds to a maximum thickness of around 25nm for Au fcc crystals. In addition, the rmsd value for all atoms of the specimen was assumed to be the same. An x-y supercell size of $n_a \times n_b = \lfloor 50\text{Å}/a \rfloor \times \lfloor 50\text{Å}/b \rfloor$ unit cells was chosen with a and b the projected unit cell lattice parameters, where $\lfloor\rfloor$ and $\lceil\rceil$ denote the floor and ceil operators. Moreover, a numerical 2D real-space grid of $\lfloor 1536 \times n_a/n_{max} \rceil \times \lfloor 1536 \times n_b/n_{max} \rceil$ pixels was selected with $n_{max} = \max(n_a, n_b)$. The ADF STEM images were scanned over an area of one projected unit cell using pixel sizes equal to $\Delta x = a/\lfloor a/0.15 \rfloor$ and $\Delta y = b/\lfloor b/0.15 \rfloor$. The multislice frozen atom calculation was performed with 30 configurations and a slice thickness dz corresponding to the distance between consecutive crystalline planes along the beam direction. In this study, simulations were performed assuming symmetric and concentric annular detectors with an ideal detector sensitivity, which refers to a detector that has a homogeneous response to electrons. All simulation settings are summarised in Table 1.

Table 1: Microscope settings used for data generation

Atomic element Z	13, 28, 29, 46, 47, 78, 79, 82
Zone axis orientation zao	[100] and [110]
Acceleration voltage HT	[60, 80, 100, 120, 200, 300]kV
Spherical aberration C_s	$[-1.0, 1.0]\mu$m
Defocus C_{10}	$[-100.0, 100.0]$Å
Convergence angle α	[17.5, 35.5]mrad
Inner detector angle θ_0	[18.0, 250]mrad
Outer detector angle θ_e	[28.0, 250]mrad
Root mean squared displacement *rmsd*	[0.075, 0.15]Å

The range of input variables covers the most common experimental conditions for an aberration-corrected transmission electron microscope. For each specimen orientation, we simulated 5000 ADF STEM images where the input variables correspond to random draws from a uniform distribution with the ranges defined in Table 1. In order to increase our training data, 20 consecutive detector angles with a minimum angular detector size of 10mrad and with an initial random inner detector angle bigger than the convergence angle were used. In this manner, 210 consecutive detector combinations are generated, which greatly increases our training data. This results in a total set of 16.6 millions examples corresponding to the number of elements × the number of zone-axis orientations × the number of simulated ADF STEM images × the combinations of the consecutive detectors or $8 \times 2 \times 5000 \times 210$. Additionally, it is worth noting that the range of *rmsd* values shown in Table 1 represents a temperature range of around [200, 650]°C for a Au fcc crystal [28]. From the simulated images, the PPISCSs were estimated as a function of thickness by summing the image intensity values, multiplying by the product of the pixel sizes and by a scaling parameter, which is equal to the number of atoms on the column over the number of projected atoms in the image.

It is important to note that the simulation time for each example of a given specimen orientation, with randomly generated input parameters and 20 consecutive detectors, was one hour using the 12GB NVIDIA GTX Titan Volta GPU. Despite utilizing our research facility, which consists of 20 of these high-performance GPUs, it still took approximately 3.5 months to generate the full dataset. This emphasizes the significant computational resources required for this type of research.

Our next step is to design a neural network to map in real-time an input vector $x = [\mathrm{Z}, \mathrm{zao}, \mathrm{HT}, C_s, C_{10}, \alpha, \theta_0, \theta_e, rmsd]$ to an output vector y = PPISCSs.

2.2. Network architecture

Fig. 1 shows the network architecture $\mathcal{N}$ which is based on the 1D version of the densely connected network architecture DenseNet [29]. The input values of the network are denoted by x and the output equals $y_p = \mathcal{N}(x)$. Using skip connections to directly connect all layers alleviates the vanishing gradient problem, strengthens feature propagation, encourages feature reuse, and substantially reduces the number of parameters since there is no need to relearn redundant features. The most important parameter for a given number of layers n_{lay} for DenseNets is the growth rate G which regulates how much information is added to the network by each layer. To reduce the number of hyperparameters, the number of units in the input layer G_0 was set to G. The number of units of the output layer was set to 87, which represents the highest number of atoms per column in our simulations. The number of layers n_{lay} and the growth rate G are optimised and will be discussed in section 3.1. The smooth and non-monotonic Swish activation function was used for the hidden layers [30]. To fulfil the positiveness hard constraint of the PPISCS, the Softplus function was used for the activation function of the output layer.

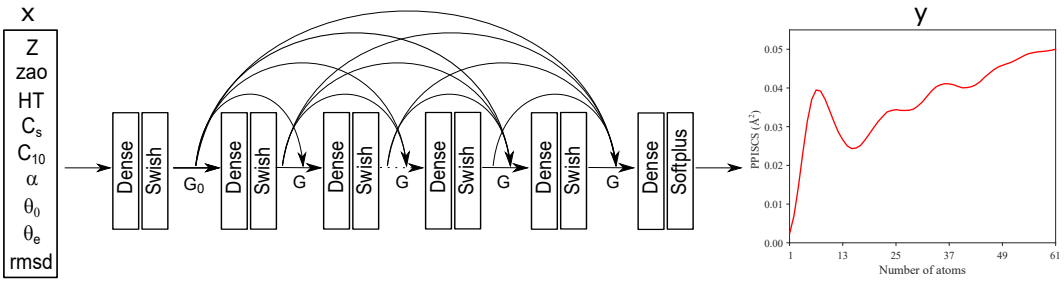


Figure 1: Densely connected network architecture for probe-position integrated scattering cross sections.

2.3. Loss function

The loss function is the effective driver of the network's learning. Its goal is to map a set of parameter values of the network onto a scalar value, which allows candidate solutions to be ranked and compared. Our loss function is composed of four terms and can be expressed as follows:

$$\mathcal{L} = \lambda_1^n \mathcal{L}_1^n + \lambda_2^n \mathcal{L}_2^n + \lambda_1^{\log} \mathcal{L}_1^{\log} + \lambda_1^{cstr} \mathcal{L}_1^{cstr}, \tag{1}$$

where λ_1^n, λ_2^n, $\lambda_1^{\log}$ and λ_1^{cstr} are the weighting parameters balancing the different loss terms, which are described in the following sections.

2.3.1. $\mathcal{L}_1^n$ loss

Fig. 2 shows the PPISCS values as a function of thickness for two different sets of specimen and microscope settings, where the figure on the left ($Z = 13$) and right ($Z = 82$) correspond to the smallest and largest values present in the training data. From this figure it is clear that there exists a large range of variation in the PPISCS values in our training data. When implementing the conventional definition for the $\mathcal{L}_1$ and also the $\mathcal{L}_2$ loss, this would mainly result in inaccurate predictions for small PPISCS values, corresponding e.g. to small detector ranges and/or low atomic numbers, and more accurate predictions for large PPISCS values. To overcome this problem, the loss function $\mathcal{L}_1$ is evaluated after normalising the ground truth y and neural network predicted y_p values with a normalization scaling factor equal to $w_{sc} = \max(y)$. This results into y^n and y_p^n, respectively, where the superscript n refers to the normalised values. The normalised $\mathcal{L}_1$, i.e. $\mathcal{L}_1^n$, is then defined as:

$$\mathcal{L}_1^n = \mathbb{E}_{y,y_p}\left\{\left\|y^n - y_p^n\right\|\right\}, \tag{2}$$

where $\mathbb{E}_{y,y_p}\{.\}$ is an operator representing the expectation value computed on variables y and y_p and not on the transformed variables y^n and y_p^n.

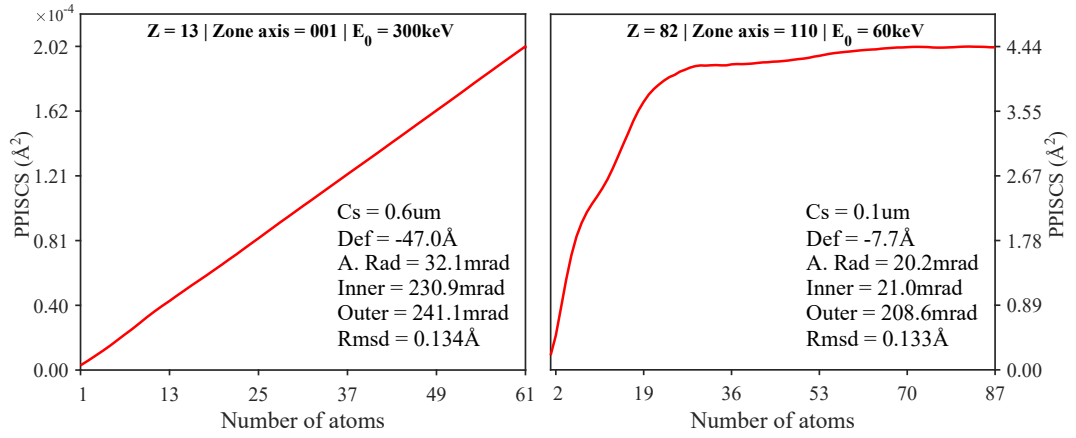


Figure 2: Probe-position integrated scattering cross sections as a function of the number of atoms for a set of specimen and microscope parameters corresponding to the smallest (left) and largest (right) values in the training data.

2.3.2. $\mathcal{L}_2^n$ loss

In the loss function in Eq. (1), the normalised $\mathcal{L}_2$ loss function, $\mathcal{L}_2^n$, is included as well since it improves predictions of the PPISCSs over the full thickness range. This can be understood since $\mathcal{L}_2$ is sensitive to large errors, which mostly occur for large thickness values of the PPISCSs. This loss function is defined as:

$$\mathcal{L}_2^n = \mathbb{E}_{y,y_p}\left\{\left\|y^n - y_p^n\right\|^2\right\} \tag{3}$$

2.3.3. $\mathcal{L}_1^{\log}$ loss

Although the $\mathcal{L}_1^n$ and $\mathcal{L}_2^n$ losses can be used to solve the problem resulting from the large variation in PPISCS values for different training examples, the training process can become unstable ending up in a local minimum. Therefore, an extra $\mathcal{L}_1$ loss function is included in which a logarithmic transformation is applied to y and y_p. The $\mathcal{L}_1^{\log}$ loss is defined as:

$$\mathcal{L}_1^{\log} = \mathbb{E}_{y,y_p}\left\{\left\|\log(y) - \log(y_p)\right\|\right\}, \tag{4}$$

This transformation reduces the difference in magnitude of the PPISCS values shown in Fig. 2 from a factor of $\sim 10^5$ to $\sim 10^1$. This stabilizes the training process and improves convergence as this transformation maintains the scale factor ratio of different training examples.

2.3.4. Constraint loss

Additivity of STEM images of contiguous detectors implies that PPISCS must also be additive. In principle, this is a hard constraint which could in principle be included in the architecture design. However, since this is not straightforward it was included in the loss function as a soft constraint [23], and is expressed as follows:

$$\mathcal{L}_1^{cstr} = \mathbb{E}_{y,y_p}\left\{\left\|y^n - y_p^{cstr}\right\|\right\}, \tag{5}$$

$$y_p^{cstr} = \left(\mathcal{N}(x_c,\theta_0,\theta_m) + \mathcal{N}(x_c,\theta_m,\theta_e)\right)/w_{sc}, \tag{6}$$

where $x_c = [\mathrm{Z}, \mathrm{zao}, \mathrm{HT}, C_s, C_{10}, \alpha, rmsd]$, $\mathcal{N}(x_c,\theta_0,\theta_m)$ and $\mathcal{N}(x_c,\theta_m,\theta_e)$ are the predicted PPISCS values of two uniform randomly generated contiguous detectors with inner and outer radius equal to $[\theta_0,\theta_m]$ and $[\theta_m,\theta_e]$, respectively.

2.4. Implementation details

In order to train our neural network, we randomly selected 16 million examples from our total dataset and used them for training. The remaining 600000 examples were used to evaluate the performance of the model. It is important to note that the validation dataset is not used during the training process, but only to evaluate the performance of the model. All models are implemented using the Keras high-level API of Tensorflow 2.10 framework [31] and are trained with 12GB NVIDIA GTX Titan Volta GPU. All network weights were initialised following reference [32]. Since the batch normalization, dropout, and weight decay hamper the model performance, they were not used in this study. Our learning policy is based on the Adam optimiser [33] with $\beta_1 = 0.9, \beta_2 = 0.999$, $\epsilon = 1 \times 10^{-7}$, and is divided in two stages. In the first part, the model is trained to minimize the loss function given by Eq. (1) with $\lambda_1^n = 1.0 \times 10^{-1}$, $\lambda_2^n = 1.0 \times 10^{-2}$, $\lambda_1^{\log} = 1.6 \times 10^1$, $\lambda_{cstr} = 1.0 \times 10^{-2}$ for 20 epochs with a learning rate of 5×10^{-5}. The weighting coefficients were chosen to give a high dominance to the $\mathcal{L}_1^{\log}$ loss while yet maintaining a small contribution from the other losses, which were found to improve the convergence. This was followed by a second training stage with $\lambda_1^n = 1.0 \times 10^2$, $\lambda_2^n = 2.0 \times 10^4$, $\lambda_1^{\log} = 1.6 \times 10^1$, $\lambda_{cstr} = 1.0 \times 10^2$ for 160 epochs with a learning rate of 2.5×10^{-5} and reduced by a factor of 0.95 every 2 epochs. During the architecture search or hyperparameter tuning, the neural networks were only trained for 5 epochs using the same parameters of the first stage of the training. In order to prevent training instability and wrong local optima, the learning rates were first warmed up for 1×10^5 steps [34, 35]. The mini-batch size of 256 was used for all experiments. The training time for each epoch was 50 minutes, giving a total training time of 5.5 days.

3. Results and discussion

3.1. Ablation study

In this subsection, we will perform a so-called ablation study to investigate the effect of the network architecture and some of its hyperparameters on the normalised $\mathcal{L}_1$ error given by Eq. (2). The learning rate, batch size, and loss weighting parameters for the first stage of the training were obtained by performing a grid search (not shown here) for a fixed densely connected architecture with $G = 160$ and $n_{lay} = 13$.

In principle, a sufficiently deep fully connected architecture should be enough to provide good PPISCS predictions. However, it is known that an optimal architecture (i.e. in terms of lower number of parameters and training time) is data dependent. For this work, we performed the ablation study for the two most common architectures: the fully connected architecture and the densely connected architecture. In addition, we also compare the computational efficient ReLU activation function against the smooth and non-monotonic Swish activation function, which has shown an improvement on the accuracy for different classification tasks [36]. Figure 3(a) summarises the performance of the densely connected architecture against the fully connected architecture for different growth rates and activation functions for a fixed number of 13 layers. The results show that for a given activation function, the densely connected architecture outperforms the fully connected architecture and requires significantly fewer parameters and computation time to achieve comparable performance. The same conclusion can be drawn for the performance of the Swish activation function against the ReLU activation function. Moreover, large G units contribute to better performance for both architectures. Figure 3(b) shows the influence of the parameter n_{lay} on the normalised $\mathcal{L}_1$ error for fixed $G = 96$. As expected, a deeper network improves the performance of the model by increasing the number of parameters, allowing the model to learn more complex features.

We can conclude from these results that the densely connected architecture is performing best in terms of the $\mathcal{L}_1^n$ error metric. Although a densely connected architecture with $(n_{lay} = 13, G = 160)$ and $(n_{lay} = 19, G = 96)$ shows similar performance with approximately the same number of parameters, deeper networks take longer for the training. Therefore, we will use the first configuration in this work. An extra advantage of this model is that its inference time is of the order of 25μs on a single thread 10th Gen Intel i7 processor 4.5Ghz.

3.2. $\mathcal{L}_1^n$ error distribution

The effect of employing a combination of loss functions capturing the relevant physical constraints of our data can be seen in Figure 4. This figure shows that the neural network produces nearly the same error distribution for all fcc crystals and zone axis orientations on the validation data. Note that loss functions based only on absolute scales of

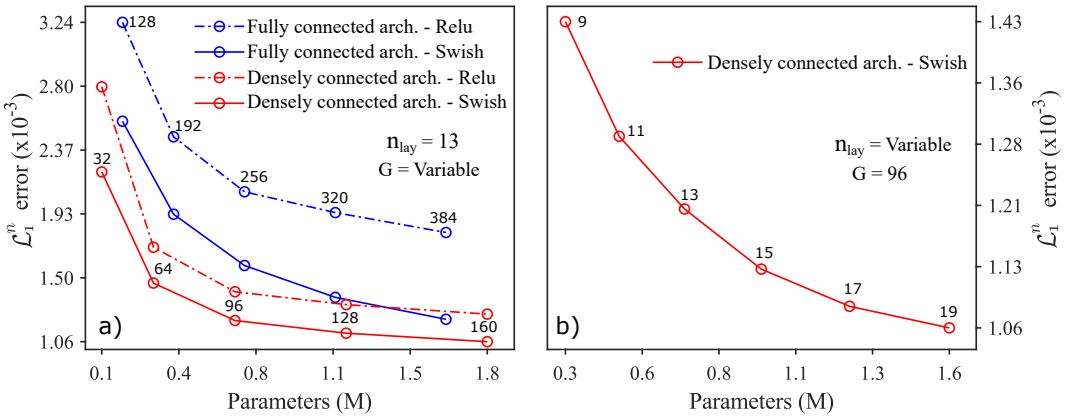
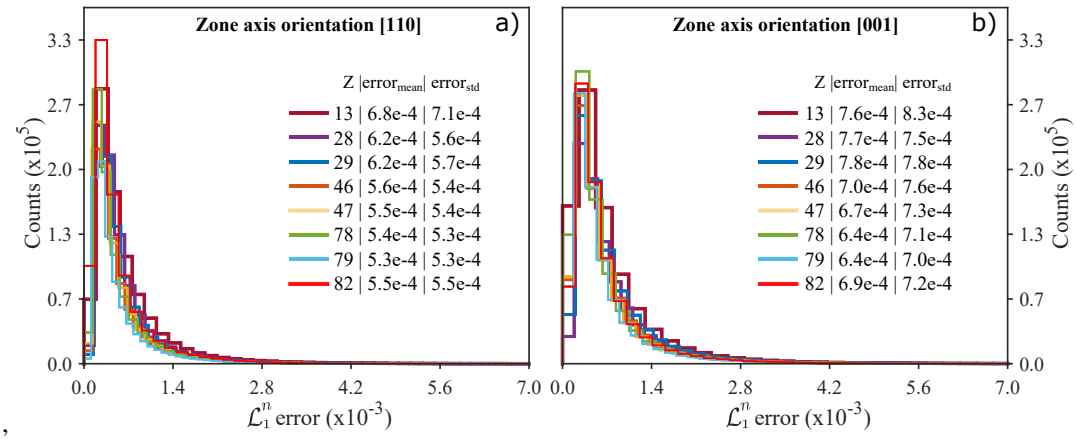


Figure 3: Ablation study of the densely connected architecture and the fully connected architecture $\mathcal{L}_1^n$ errors as a function of the size of the model. a) For different growth rates G with a fixed number of layers $n_{lay} = 13$. The growth rate G is indicated next to each data point. b) For different numbers of layers n_{lay} with a fixed growth rate $G = 96$. The number of layers n_{lay} is indicated next to each data point.

PPISCSs values such as $\mathcal{L}_1$ and $\mathcal{L}_2$ would strongly depend on the atomic number. This is due to the fact that high atomic numbers will generate higher PPISCS values than low atomic numbers for the same fcc crystal and microscope settings, and thus the loss function would be biased towards high atomic numbers.



,

Figure 4: Histogram of the $\mathcal{L}_1^n$ error of the predicted PPISCS in the validation set for different fcc crystals along the (a) [110] and (b) [001] zones axis orientation.

Although the average validation error $\mathcal{L}_1^n$ is small ($\approx 6.0 \times 10^{-4}$), all histograms show long tails independent of the atomic number or zone axis orientation. In order to perform a proper analysis of the $\mathcal{L}_1^n$ error distribution, it is necessary to show the predicted PPISCSs for the average and highest errors that can be observed in these histograms. Figures 5 and 6 show the PPISCS-thickness predictions and ground truth for the average $\mathcal{L}_1^n$ errors for all trained fcc crystal along the [001] and [110] zone axis orientation, respectively. These results show an excellent quantitative match between the ground truth and the predicted values of the PPISCSs for all cases. Moreover, the average error does not seem to be correlated with the input simulation parameters.

Figures 7 and 8 show the PPISCS-thickness predictions and ground truth for the largest $\mathcal{L}_1^n$ errors for all trained fcc crystal along the [001] and [110] zone axis orientation, respectively. These results show that even for the worst-case scenario, the neural network prediction only deviates from the ground truth by approximately 1% in terms of the $\mathcal{L}_1^n$ metric. A closer look at these figures reveals a correlation between the large values of $\mathcal{L}_1^n$ and a smaller detector size (i.e. the difference between the outer and inner detector angle) of around 10mrad. This correlation arises due to the fact that PPISCS values calculated from smaller detector sizes are for most cases highly non-linear against thickness

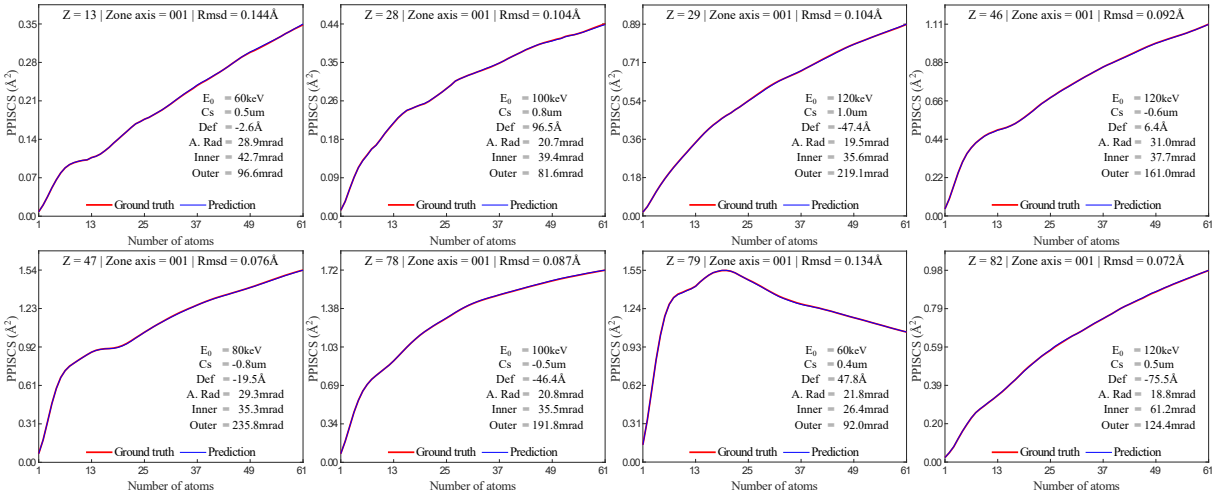


Figure 5: Probe-position integrated scattering cross sections for different fcc crystals along the [001] zone axis as a function of the number of atoms for different cases corresponding to an average error of the histogram shown in Fig. 4.

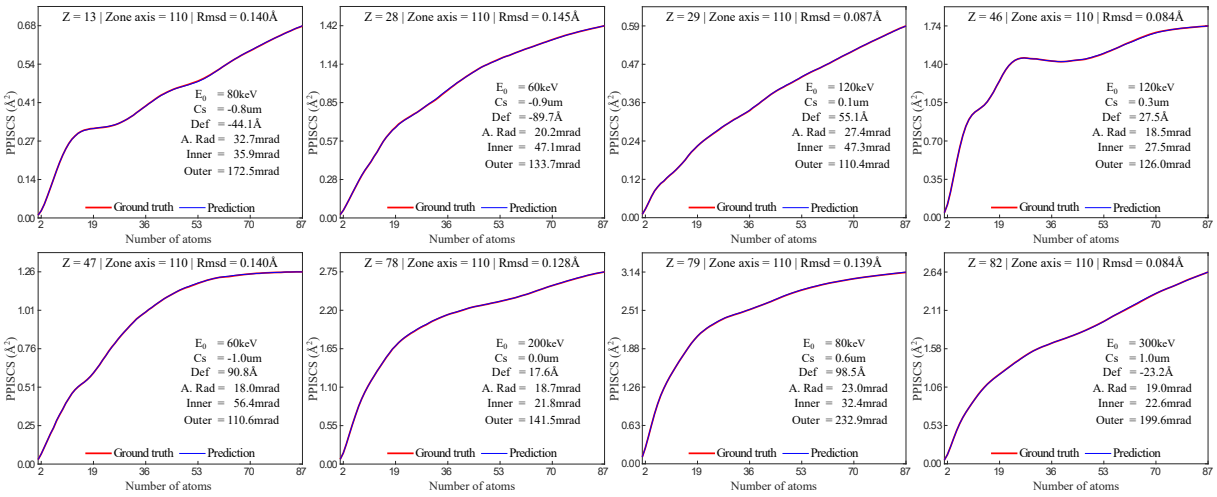


Figure 6: Probe-position integrated scattering cross sections for different fcc crystals along the [110] zone axis as a function of the number of atoms for different cases corresponding to an average error of the histogram shown in Fig. 4.

as can be seen in figure 7 and figure 8. Moreover, smaller detector sizes will also require a higher number of phonon configurations in order to get stable results. In principle, our model can be improved by generating new PPISCS values using a larger number of phonon configurations and a more powerful neural network architecture.

The results of this section show that our neural network is able to learn the complex relationship between the input parameters of our simulation and the PPISCS-thickness dependence. Moreover, it is important to notice that our neural network runs in real time on a normal desktop computer.

In order to demonstrate the power of our model, we will use the network to show some applications for which real-time PPISCS-thickness predictions are required.

3.3. Real time applications

Figure 9 shows the PPISCSs as a function of thickness for a broad range of specimen and microscope settings. In this figure, the standard settings correspond to the following input simulation parameters $x = [79, 110, 200\text{kV},$

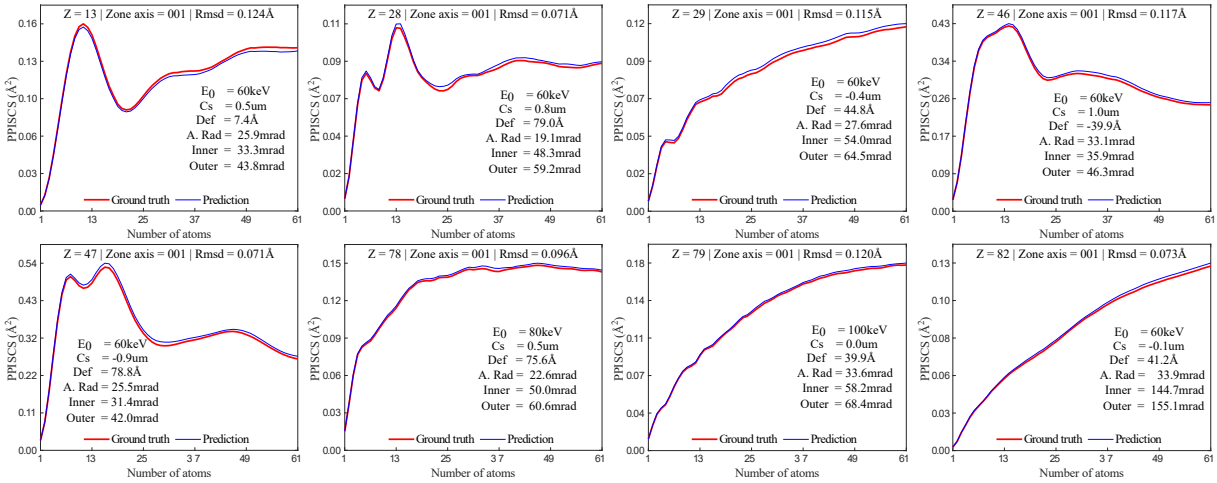


Figure 7: Probe-position integrated scattering cross sections for different fcc crystals along the [001] zone axis as a function of the number of atoms for different cases corresponding to the highest error of the histogram shown in Fig. 4.

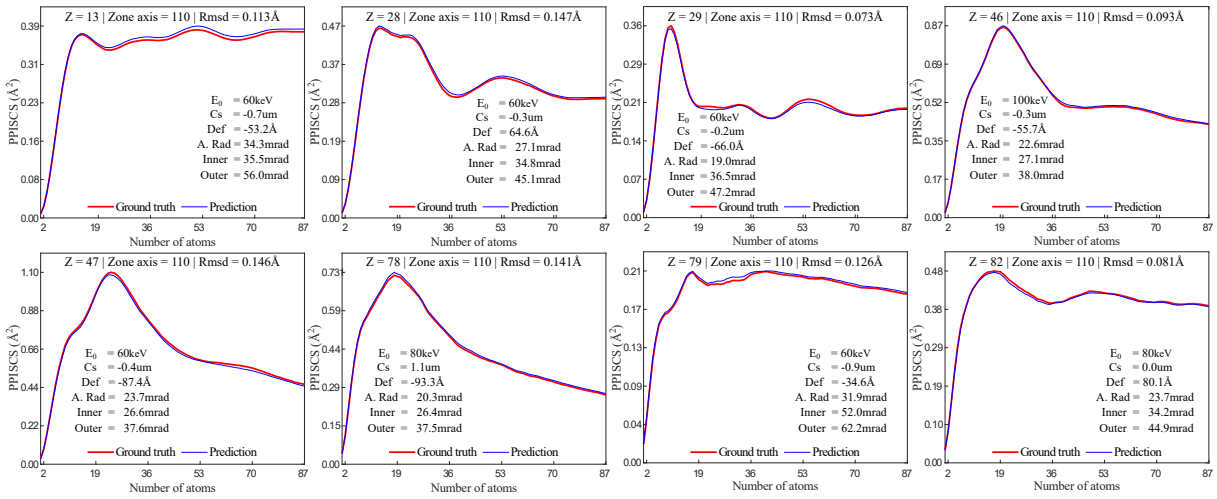


Figure 8: Probe-position integrated scattering cross sections for different fcc crystals along the [110] zone axis as a function of the number of atoms for different cases corresponding to the highest error in the histogram shown in Fig. 4.

0.001mm, −19.40Å, 24mrad, 45mrad, 160mrad, 0.085Å]. Next, each input simulation parameter has been varied independently as a function of atomic element, acceleration voltage, spherical aberration, defocus, aperture angle, inner detector angle, outer detector angle and *rmsd* in panels a-h, respectively.

The Z-contrast nature of the ADF-STEM signal for different atomic numbers can be seen in figure 9(a). Figure 9(b) shows the breakdown in the monotonically increasing relation of PPISCS with thickness and the increase of non-linearity if the acceleration voltage decreases. Furthermore, only very small variations in the PPISCS-thickness curve with spherical aberration and defocus can be seen in figure 9(c) and in figure 9(d), respectively, as expected for aberration-corrected transmission electron microscopes. Figure 9(e) shows the effect the aperture angle on the PPISCS-thickness curve. For small thickness, the PPISCS values are almost independent of the aperture angle. However, when the thicknesses increases, a non-linear dependence is observed. Figure 9(f) illustrates a well defined relationship between the PPISCS-thickness curve and inner detector angle. In particular, it is shown that the PPISCS-thickness values are inversely proportional to the inner angle. Figure 9(g) and figure 9(h) show the dependence of PPISCS-thickness curve with outer detector angle and the rmsd, respectively. Although PPISCS values mainly show

a non-linear dependence with thickness, this non-linearity can be decreased by increasing the outer angle or the rmsd.

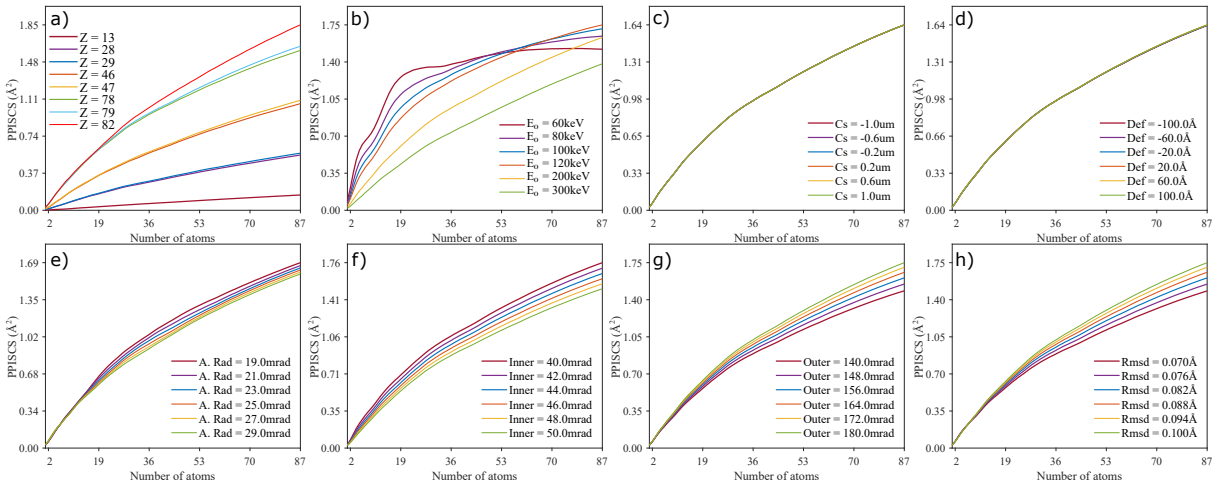


Figure 9: Probe-position integrated scattering cross sections versus the number of atoms as a function of varying independently each of the simulation input parameters as : (a) atomic element, (b) acceleration voltage, (c) spherical aberration, (d) defocus, (e) aperture angle, (f) inner detector angle, (g) outer detector angle and (h) isotropic root-mean-square displacement.

Since experimental settings are only known within a certain measurement precision, the real-time PPISCS-thickness prediction is of great value to optimize the simulation settings by matching simulated values with the experimental PPISCS-thickness curve, which can be obtained by using the statistic-based atom-counting method [3]. Therefore, the experimental PPISCS-thickness curve together with the standard deviation on the experimentally measured parameters can be used to optimize the simulation parameters. To illustrate this, PPISCS-thickness values are considered for a Ag crystal along the [001] orientation and an acceleration voltage of $300kV$. The ground truth input parameters have been chosen equal to $[47, 001, 300\text{kV}, 0.0008\text{mm}, 50.00\text{Å}, 21.00\text{mrad}, 46.00\text{mrad}, 190.00\text{mrad}, 0.092\text{Å}]$ along with their standard deviation $std(x) = [0, 0, 0\text{kV}, 0.001\text{mm}, 80\text{Å}, 0.25\text{mrad}, 1.0\text{mrad}, 1.0\text{mrad}, 0.003\text{Å}]$. The experimentally measured parameters were randomly generated from the ground truth parameters within a range of $\pm std(x)$ and are equal to $x = [47, 001, 300\text{kV}, -0.0006\text{mm}, 13.61\text{Å}, 20.78\text{mrad}, 45.09\text{mrad}, 189.56\text{mrad}, 0.090\text{Å}]$. The PPISCS-thickness curve for the ground truth and the experimentally measured curve is shown in figure 9 in red and blue, respectively. Next, the input parameters have been optimized using the derivative-free Nelder-Mead simplex method [37]. The cost function minimizes the absolute difference between the measured and predicted PPISCS-thickness curve. The optimisation process takes approximately one second on a normal desktop computer and yields the following estimated simulation parameters $x = [47, 001, 300\text{kV}, 0.0006\text{mm}, 52.79\text{Å}, 21.00\text{mrad}, 46.01\text{mrad}, 190.44\text{mrad}, 0.092\text{Å}]$. The optimized PPISCS-thickness curve is shown in figure 10 in green. This result demonstrates that the aperture angle, inner angle and rmsd can be estimated reliably. However, the values for the spherical aberration, defocus and outer angle are less accurate due to the fact that the PPISCS-thickness curve is invariant for changes in those parameters as shown in Figure 9.

It is known that the PPISCS-thickness curve can be used to estimate the number of atoms for zone axis oriented specimens. However, uncertainties in the measured microscope parameters can yield large deviations in predicted PPISCS values, especially at larger thicknesses. The standard deviation for each thickness can be estimated by taking random draws of the input parameters from a uniform distribution within their allowed measurement errors. Figure 11 shows the effect of measurement errors on the PPISCS-thickness curve for two different examples. The input parameters are shown as an inset and the standard deviations are assumed to be equal to $std(x) = [0, 0, 0\text{kV}, 0.001\text{mm}, 50\text{Å}, 0.25\text{mrad}, 0.5\text{mrad}, 0.5\text{mrad}, 0.0025\text{Å}]$. Based on 1000 random samples for the input parameters, the standard deviation on the PPISCS values has been calculated and is shown in blue. This figure shows a monotonic increase in PPISCS errors with thickness. These results demonstrate that in order to count the number of atoms based on the PPISCS-thickness curve, the quantification and inclusion of measurement errors in the microscope settings and rmsd is important.

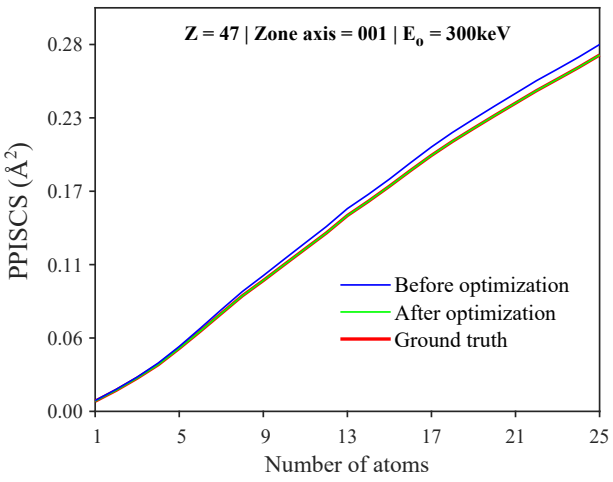


Figure 10: Probe-position integrated scattering cross sections versus number of atoms. The blue curve (before optimization) shows the simulated PPISCS-thickness values using the measured experimental parameters. The green curve illustrates the simulated PPISCS-thickness values after an optimization procedure within the experimental uncertainties. The red curve shows ground truth PPISCS-thickness values.

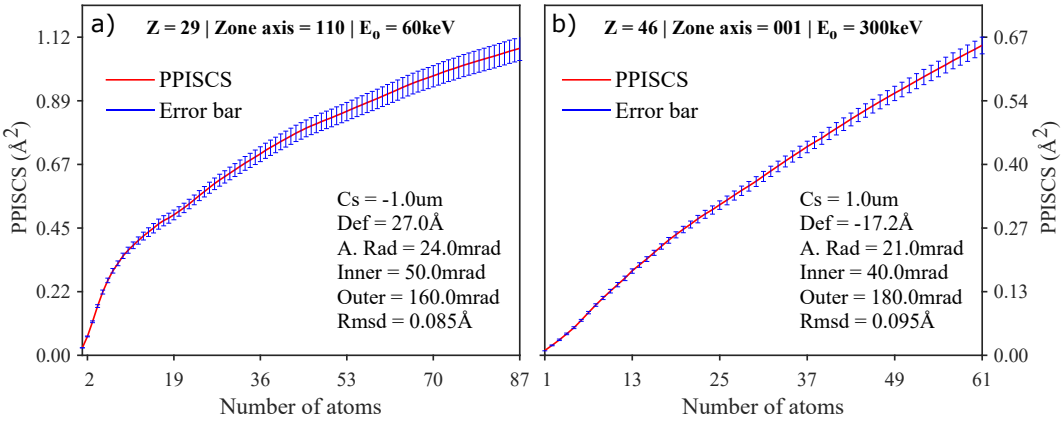


Figure 11: Influence of uncertainties in the microscope parameters and rmsd on the probe-position integrated scattering cross sections.

As the last application, we show the probability of error for atom-counting from ADF STEM images for different fcc crystals. In order to evaluate the possibilities and limitations for atom-counting, we assess the probability to miscount the number of atoms as a function of a range of microscope settings and specimen parameters [38]. The optimal experimental design corresponds to the set of parameters for which the probability of error is minimised. The availability of the neural network enables us to evaluate more quickly and for a wider range of parameters. Within statistical detection theory, the atom-counting problem is formulated as a statistical hypothesis test, where each hypothesis corresponds to a specific number of atoms. The probability of error then corresponds to the probability to choose the wrong hypothesis. The decision to assign a certain observation to a specific hypothesis is taken based on the companion set of probability functions for the hypotheses. For ADF STEM observations, the pixel intensities correspond to statistically independent electron counting results which are Poisson distributed. Then, also the probability function for the scattering cross-sections can be derived [38]. In order to illustrate the concept of the optimal experiment design, we computed here the probability of error evaluated for the outer detector angle, the acceleration voltage, the convergence angle, the thickness, the atomic number and the temperature. The probability of error for each case is also evaluated as a function of the inner detector radius θ_0. The settings which are not varied are chosen equal to $[Z = 79, zone-axis = [001], HT = 300\ \text{kV}, C_s = 0.001\text{mm}, C_{10} = -100\ \text{Å}, \alpha = 17.5\ \text{mrad}, \theta_0, \theta_e = 250\ \text{mrad}, rmsd = 0.0898\text{Å}]$ for a thickness up to 85 atoms. The incident electron dose was chosen equal to $10^4\ \text{e}^-/\text{Å}^2$.

The results are displayed in the different panels of Figure 12. It is clear from those figures that the inner angle (a-f), acceleration voltage (b), thickness (d), and atomic number (e) have the largest impact on the probability of error values. The dependence of the probability of error on the outer detector angle (a), the convergence angle (c), and the temperature (f) is much smaller. Therefore, the optimization of the first set of parameters will significantly enhance the reliability with which the number of atoms can be counted.

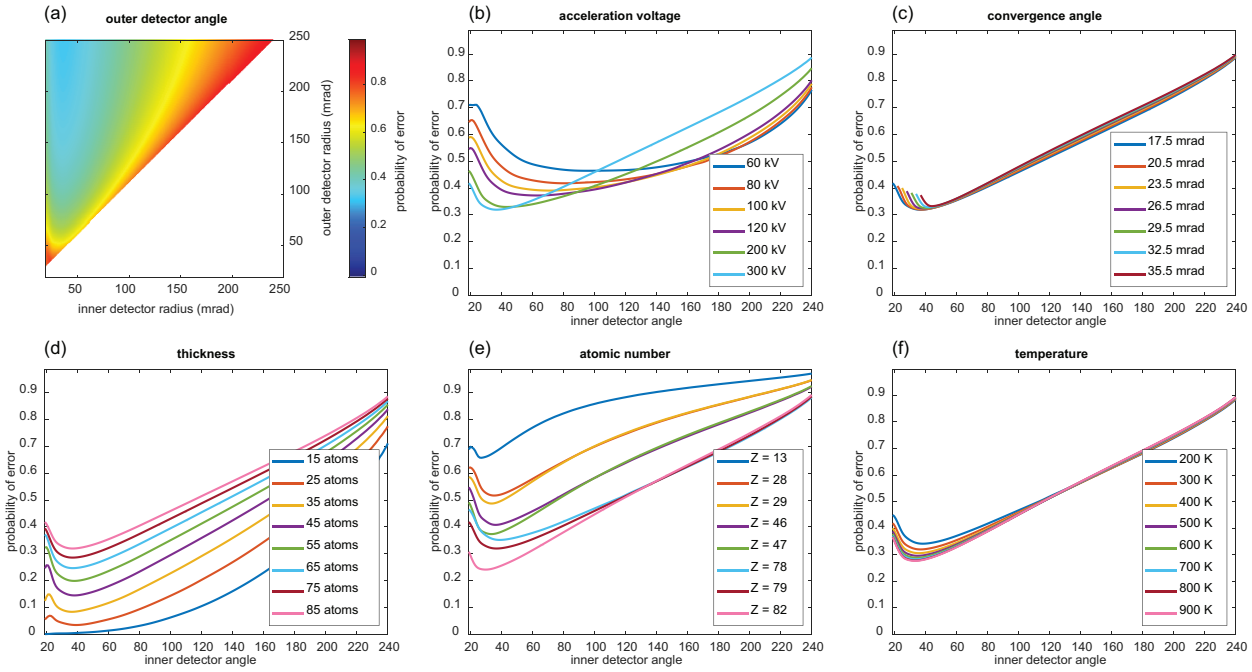


Figure 12: Probability of error for atom-counting as a function of the inner detector radius and (a) the outer detector radius, (b) the acceleration voltage, (c) the convergence angle, (d) the thickness, (e) the atomic number, (f) the temperature.

Finally, we would like to point out that our neural network model could be improved by increasing the training datasets, the number of phonons for the simulations, or by using more sophisticated architectures. Additionally, the applicability of our network can easily be extended to more fcc crystals, zone axis orientations, and a broader range of microscope parameters through the use of transfer learning. Although our network takes into account specific physical constraints, such as the positivity of PPISCS-thickness values and the additivity constraint for contiguous detectors, it is essential to emphasize that the neural network's predictions should only be trusted within the range of the input parameters used during training.

The output of our network, in principle, can be compared with a subset of experimental PPISCSs obtained from quasi-ideal direct electron detectors. However, these detectors primarily cover low and intermediate scattering angle ranges, which are not fully addressed by our current network. Consequently, as a future research direction, it is possible to enhance the capabilities of the existing PPISCS-thickness neural network by incorporating the detector sensitivity map and expanding the detector value range to include ABF-STEM PPISCS-thickness values.

Nonetheless, such an expansion poses several challenges, as it necessitates new simulations that, in principle, require the integration of 2D detector sensitivity maps. To directly incorporate this element, an architecture with convolutional layers is essential, which may hinder real-time calculations. A potential solution to this problem involves utilizing the radially averaged detector sensitivity map for each detector range. This map can be further compressed through parameterization, reducing the input parameters to the network and enabling real-time operation.

Additionally, low-angle scattering is significantly influenced by defocus, necessitating the incorporation of temporal incoherence. This factor can be quantified by a single parameter, known as defocus spread, which primarily results from the current instability of the objective lens, the overall energy spread of the incident electron beam, and the

incident electron energy. By employing numerical integration, the defocus spread can be accurately accounted for, although at the expense of increased simulation time.

It is important to recognize that the data generation time will considerably increase, as reliance on the additivity constraint for ideal contiguous detectors is no longer feasible.

4. Conclusions

In summary, we present a densely connected neural network that can predict real-time ADF STEM PPISCS-thickness values for the most common fcc crystals along their main zone axis orientations, microscope parameters, and rmsd values. We have shown that our architecture with 13 layers and a growth rate parameter equal to 160 is a parameter-efficient network and yields accurate predictions for a large range of input parameters which are commonly used for aberration-corrected transmission electron microscopes. We have also shown that our architecture can be used to estimate microscope parameters and the rmsd value of the specimen based on the PPISCS-thickness curve. Moreover, It can also be used to estimate the uncertainty of the PPISCS-thickness values resulting from experimental measurement errors. The knowledge of this uncertainty will play an important role in the proper quantification of the number of atoms based on the PPISCS-thickness curve. The inference code for MATLAB, python and the tensorflow source code for training is available in the github repository https://github.com/Ivanlh20 `https://github.com/Ivanlh20/RT_PPISCS`.

5. Acknowledgements

This work was supported by the European Research Council (Grant 770887 375 PICOMETRICS to S. Van Aert and Grant 823717 ESTEEM3). The authors acknowledge financial support from the Research Foundation Flanders (FWO, Belgium) through project fundings (G.0346.21N and EOS 30489208) and a postdoctoral grant to A. De Backer. S. Van Aert acknowledges funding from the University of Antwerp Research fund (BOF).

Author contributions

I. Lobato and S. Van Aert designed the study. I. Lobato developed the workflow for data generation, implemented, trained and evaluated neural networks models. A. De Backer evaluated the neural network model for calculating the probability of error for atom-counting. All authors participated in conceiving the research, discussing the results, and writing of the manuscript.

References

[1] Sara Bals, Marianna Casavola, Marijn A. Van Huis, Sandra Van Aert, K. Joost Batenburg, Gustaaf Van Tendeloo, and Daniël Vanmaekelbergh. Three-dimensional atomic imaging of colloidal core-shell nanocrystals. *Nano Letters*, 11(8):3420–3424, aug 2011.

[2] Dmitri O. Klenov and Susanne Stemmer. Contributions to the contrast in experimental high-angle annular dark-field images. *Ultramicroscopy*, 106(10):889–901, aug 2006.

[3] S. Van Aert, J. Verbeeck, R. Erni, S. Bals, M. Luysberg, D. Van Dyck, and G. Van Tendeloo. Quantitative atomic resolution mapping using high-angle annular dark field scanning transmission electron microscopy. *Ultramicroscopy*, 109(10):1236–1244, sep 2009.

[4] Andreas Rosenauer, Katharina Gries, Knut Müller, Angelika Pretorius, Marco Schowalter, Adrian Avramescu, Karl Engl, and Stephan Lutgen. Measurement of specimen thickness and composition in Alx Ga1 - x N / GaN using high-angle annular dark field images. *Ultramicroscopy*, 109(9):1171–1182, aug 2009.

[5] James M. Lebeau, Scott D. Findlay, Leslie J. Allen, and Susanne Stemmer. Standardless atom counting in scanning transmission electron microscopy. *Nano Letters*, 10(11):4405–4408, nov 2010.

[6] S. Van Aert, A. De Backer, G. T. Martinez, B. Goris, S. Bals, G. Van Tendeloo, and A. Rosenauer. Procedure to count atoms with trustworthy single-atom sensitivity. *Physical Review B - Condensed Matter and Materials Physics*, 87(6), feb 2013.

[7] A. De Backer, G. T. Martinez, A. Rosenauer, and S. Van Aert. Atom counting in HAADF STEM using a statistical model-based approach: Methodology, possibilities, and inherent limitations. *Ultramicroscopy*, 134:23–33, nov 2013.

[8] E H., K. E. MacArthur, T. J. Pennycook, E. Okunishi, A. J. D'Alfonso, N. R. Lugg, L. J. Allen, and P. D. Nellist. Probe integrated scattering cross sections in the analysis of atomic resolution HAADF STEM images. *Ultramicroscopy*, 133:109–119, oct 2013.

[9] G T Martinez, A. De Backer, A Rosenauer, J Verbeeck, and S. Van Aert. The effect of probe inaccuracies on the quantitative model-based analysis of high angle annular dark field scanning transmission electron microscopy images. *Micron*, 63(2014):57–63, 2014.

[10] Sandra Van Aert, Kees J. Batenburg, Marta D. Rossell, Rolf Erni, and Gustaaf Van Tendeloo. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature*, 470(7334):374–377, feb 2011.

[11] De Wael A, De Backer A, Jones L, Nellist PD, and Van Aert S. Hybrid statistics-simulations based method for atom-counting from ADF STEM images. *Ultramicroscopy*, 177:69–77, jun 2017.

[12] Karel H. W. Van Den Bos, Annick De Backer, Gerardo T Martinez, Naomi Winckelmans, Sara Bals, Peter D Nellist, and Sandra Van Aert. Unscrambling Mixed Elements using High Angle Annular Dark Field Scanning Transmission Electron Microscopy. *Physical Review Letters*, 116(24), 2016.

[13] Sergio I. Molina, Maria P. Guerrero, Pedro L. Galindo, David L. Sales, Maria Varela, and Stephen J. Pennycook. Calculation of integrated intensities in aberration-corrected Z-contrast images. *Journal of Electron Microscopy*, 60(1):29–33, feb 2011.

[14] I. Lobato and D. Van Dyck. MULTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA. *Ultramicroscopy*, 156:9–17, sep 2015.

[15] I. Lobato, S. van Aert, and J. Verbeeck. Progress and new advances in simulating electron microscopy datasets using MULTEM. *Ultramicroscopy*, 168:17–27, sep 2016.

[16] Colin Ophus. A fast image simulation algorithm for scanning transmission electron microscopy. *Advanced Structural and Chemical Imaging*, 3(1):1–11, dec 2017.

[17] Mingxing Tan and Quoc V Le. EfficientNetV2: Smaller Models and Faster Training. 2021.

[18] Ziang Cheng, Shaodi You, Viorela Ila, and Hongdong Li. Semantic Single-Image Dehazing, April 2018.

[19] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11133 LNCS, pages 63–79. Springer Verlag, sep 2019.

[20] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 2020-December, may 2020.

[21] Tyler Bonnen, Daniel L. K. Yamins, and Anthony D. Wagner. When the ventral visual stream is not enough: A deep learning account of medial temporal lobe involvement in perception. *Neuron*, 109(17):2755–2766, 2021.

[22] Arvind T Mohan, Nicholas Lubbers, Daniel Livescu, and Michael Chertkov. Embedding Hard Physical Constraints in Neural Network Coarse-Graining of 3D Turbulence. 2020.

[23] Tom Beucler, Michael Pritchard, Stephan Rasp, Jordan Ott, Pierre Baldi, and Pierre Gentine. Enforcing Analytic Constraints in Neural-Networks Emulating Physical Systems. *Physical Review Letters*, 126(9), sep 2019.

[24] Rui Wang and Rose Yu. Physics-Guided Deep Learning for Dynamical Systems: A Survey. pages 1–28, 2021.

[25] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, feb 2019.

[26] Kurt Hornik. Approximation capabilities of muitilayer feedforward networks. *Neural Networks*, 4:251–257, 1991.

[27] David A Muller, Byard Edwards, Earl J. Kirkland E, and John Silcox. Simulation of thermal diffuse scattering including a detailed phonon dispersion curve. *Ultramicroscopy*, 86(3-4):371–380, 2001.

[28] H. X. Gao and L. M. Peng. Parameterization of the temperature dependence of the Debye–Waller factors. *Acta Crystallographica Section A Foundations of Crystallography*, 55(5):926–932, 2002.

[29] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, volume 2017-Janua, pages 2261–2269. Institute of Electrical and Electronics Engineers Inc., aug 2017.

[30] Prajit Ramachandran, Barret Zoph, and Quoc V Le Google Brain. Searching for activation functions. In *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings*, 2018.

[31] `https://www.tensorflow.org`.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2015 Inter, pages 1026–1034, 2015.

[33] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv:1412. 6980*, 2014.

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 2017-Decem, pages 5999–6009, jun 2017.

[35] Martin Popel and Ondřej Bojar. Training Tips for the Transformer Model. *The Prague Bulletin of Mathematical Linguistics*, 110(1):43–70, mar 2018.

[36] Prajit Ramachandran, Barret Zoph, and Quoc V Le Google Brain. Swish: a self-gated activation function. *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings*, 10 2017.

[37] Jeffrey C. Lagarias, James A. Reeds, Margaret H. Wright, and Paul E. Wright. Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9:112–147, 1998.

[38] A. De Backer, A. De wael, J. Gonnissen, and S. Van Aert. Optimal experimental design for nano-particle atom-counting from high-resolution stem images. *Ultramicroscopy*, 151:46–55, 4 2015.