Ultramicroscopy 245 (2023) 113663

Contents lists available at ScienceDirect

Ultramicroscopy

journal homepage: www.elsevier.com/locate/ultramic

Near-real-time diagnosis of electron optical phase aberrations in scanning transmission electron microscopy using an artificial neural network

Giovanni Bertoni [a,*], Enzo Rotunno [a,*], Daan Marsmans [b], Peter Tiemeijer [b], Amir H. Tavabi [c], Rafal E. Dunin-Borkowski [c], Vincenzo Grillo [a]

[a] Istituto Nanoscienze, Consiglio Nazionale delle Ricerche, Via G. Campi 213/A, 41125 Modena, Italy
[b] Thermo Fisher Scientific, PO Box 80066, 5600 KA Eindhoven, the Netherlands
[c] Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, 52425 Jülich, Germany

ARTICLE INFO

Keywords:
Electron optical phase
Aberration correction
Neural network
Artificial intelligence
Spatial resolution
Scanning transmission electron microscopy

ABSTRACT

The key to optimizing spatial resolution in a state-of-the-art scanning transmission electron microscope is the ability to measure and correct for electron optical aberrations of the probe-forming lenses precisely. Several diagnostic methods for aberration measurement and correction have been proposed, albeit often at the cost of relatively long acquisition times. Here, we illustrate how artificial intelligence can be used to provide near-real-time diagnosis of aberrations from individual Ronchigrams. The demonstrated speed of aberration measurement is important because microscope conditions can change rapidly. It is also important for the operation of MEMS-based hardware correction elements, which have less intrinsic stability than conventional electromagnetic lenses.

1. Introduction

Aberration correction has been key to achieving sub-Å spatial resolution in state-of-the-art transmission electron microscopes (TEMs) [1–4]. At the same time, it has required great effort in hardware fabrication and instrument control. In order to circumvent the unavoidable spherical aberration of round electromagnetic lenses and residual non-symmetrical aberrations, many lenses and multipoles have to be combined and controlled with extreme precision [5–9]. The situation is similar to that of adaptive correction of aberrations in astronomical telescopes, for which rapid adjustments are required to account for atmospheric changes [10, 11].

The conceptual scheme that underlies control of aberration correction is based on the decomposition of the complex phase landscape that acts on an electron in a lens into well-known Zernike polynomials [12], which can be expressed in the form of a classical aberration polynomial series. Aberration correction aims to introduce a similar series of polynomials of opposite sign, in order to provide near-perfect compensation up to a well-defined order. An experimental microscopist is typically trained to recognize and manually correct lower-order aberrations, such as defocus (C_1) and astigmatism (A_1). However, in order to obtain the best possible resolution of the microscope, higher-order aberrations must also be tuned. This procedure is too difficult to complete manually and requires sophisticated automated procedures for diagnosing and correcting the aberrations. Aberration diagnosis is a key factor that affects the precision and speed of correction. Many ideas and concepts have been introduced to make aberration evaluation as reliable as possible for both the imaging system after the sample and the probe-forming system before it.

For the imaging system, the gold standard is the use of a Zemlin tableau [13], which involves the use of different inclined plane wave illuminations to map variations in lower order aberrations such as apparent defocus and astigmatism. In the case of scanning TEM (STEM), a similar approach involves measurements of beam shift [14] or changes in contrast in STEM images as a function of the beam tilt [15]. In the latter approach, beam tilt series of STEM images recorded at different defoci (usually from an Au/C cross-grating replica) are used to determine the point spread function of the source by deconvolution [2]. This "STEM tableau" approach is the built-in procedure in the microscope used in the present work. It is used as a benchmark for our work due to its precision and accuracy in the measurement of aberrations.

The work presented in this paper involves the analysis of a Ronchigram, *i.e.*, a coherent diffraction image of a convergent probe after scattering through a thin crystalline or amorphous sample [16]. A Ronchigram contains both diffraction information and a distorted image of the sample. The image has a locally-varying magnification, which

* Corresponding authors.
E-mail addresses: giovanni.bertoni@cnr.it (G. Bertoni), enzo.rotunno@cnr.it (E. Rotunno).

https://doi.org/10.1016/j.ultramic.2022.113663
Received 28 April 2022; Received in revised form 17 October 2022; Accepted 17 December 2022
Available online 18 December 2022

depends on the second derivative of the aberration phase [17]. Some methods combine information from Ronchigrams recorded at different beam tilt angles. However, the Ronchigram can itself be regarded as a superimposed diffraction map for different tilt components of the probe. In this sense, it is similar to a full Zemlin tableau. This correspondence has been used to propose an aberration measurement approach [18]. Lupini et al. demonstrated that, after segmenting a Ronchigram of an amorphous material into sub-images, each of which corresponds to a different effective beam tilt angle [19], it is possible to apply a defocus/astigmatism fitting routine just as in a Zemlin-like scheme. Alternative aberration diagnosis methods make use of other features in Ronchigrams, which are indicative of the richness of the information that it contains [20–25]. If a single Ronchigram can be used to evaluate aberrations, then it can be used to provide fast correction feedback. This capability can also be useful for MEMS-based optics [26, 27], which can even be used for spherical aberration correction [28], albeit with lower stability than conventional optics.

A recent addition to this complicated landscape of methods and concepts is the introduction of artificial intelligence (AI) and deep learning. AI makes use of a training procedure to recognize patterns and hidden recurrences in a dataset, allowing the parametric behavior of a system to be predicted [27, 29]. A recent example of the use of AI in electron optics, which also involves the use of MEMS technology, is our application of AI to tune an orbital angular momentum (OAM) sorter [30–33]. An OAM sorter is a combination of two tunable phase elements, which disperse an electron beam into a spectrum of discrete OAM states. We applied AI to a single image of an OAM spectrum to determine all of the relevant alignment parameters and to automatically tune the OAM sorter [26, 27]. Recent approaches have demonstrated the use of deep learning techniques to optimize a STEM aperture for improving resolution [34, 35], confirming that the analysis of a single Ronchigram [36] can be translated directly into aberration diagnostics.

Here, we provide the first demonstration of aberration diagnosis based on an artificial neural network (ANN). We start from the idea of segmentation discussed by Lupini et al. [19] However, instead of relying on a semi-analytical approach, we apply ANN recognition to the full set of sub-images without assuming any approximation.

2. Methods

2.1. Ronchigram modeling

As a starting point, we describe diffraction information in a Ronchigram in terms of the scattering of a convergent probe through a phase object. The probe is described in real space x by the expression $\psi_p(x) = IFT\{e^{-i\chi(q)}A(q)\}$, where q is a reciprocal space coordinate (*i.e.*, frequency) conjugated to x, while $\chi(q)$ is the aberration function of the illumination system and $A(q)$ is a top hat function that describes the use of a condenser aperture. Hereafter, *IFT* and *FT* refer to inverse and direct Fourier transforms, respectively. The sample has a multiplicative effect, which can be described by the transmission function $T(x) = e^{-i\sigma V(x)}$, where the interaction parameter $\sigma = 2\pi m_0 \gamma e\lambda/h^2$, m_0 and e are the rest mass and charge of the electron, λ is the electron wavelength, γ is the relativistic factor, h is the Planck constant and $\sigma V(x)$ is the phase shift produced by the sample. If the sample is sufficiently thin, then $V(x)$ is the effective projected atomic potential. The intensity $I(q)$ of the Ronchigram function $\psi_t(q)$ in reciprocal space can be written in the form

$$I(q) = |\psi_t(q)|^2 = \left|FT\left(IFT\left\{e^{-i\chi(q)}A(q)\right\}e^{-i\sigma V(x)}\right)\right|^2. \quad (1)$$

2.2. Sample function

The sample potential $V(x)$ should be as close as possible to the real function that describes the potential of an amorphous phase object (in our case, a thin film of amorphous C, or a-C). The way in which the sample transmits spatial frequencies is critical for ANN training. As a first approach, a random distribution $V(x) = (\overline{V}t)\mathrm{rand}(x)$ can be used, where $\mathrm{rand}(x)$ is a uniform random number in the range [0, 1]. A good choice for $\sigma\overline{V}t$ is the value $\pi/4$, as used by Schnitzer et al. [37], as it is close to the value obtained from the weak phase object (WPO) approximation by considering an interaction constant $\sigma = 0.0065\ V^{-1}\mathrm{nm}^{-1}$ at 300 kV, a sample thickness $t \sim 10$ nm and a mean inner potential $\overline{V} = 10.7$ V for a-C [38, 39]. However, in this first approach, the potential contains all frequencies with the same weight (*i.e.*, white noise), which is unrealistic. A better way in which to build the potential is to also consider the dependence on the frequency of transmission of a thin amorphous C film (or the scattering intensity $I_e(q)$) in reciprocal space [40]. A good approximation (assuming azimuthal symmetry of scattering) is given by the electron atomic scattering factor $f(q)$, which can be calculated analytically as a sum of Gaussian functions, as described by Peng et al. [41] Its angular dependence is shown in Fig. 1, in which the electron scattering factor is compared with the square root of the scattered intensity measured from a thin a-C film at 300 kV. Thermal vibrations can be considered as a further multiplicative factor with a Gaussian function $b(q)$ in reciprocal space by using a Debye-Waller factor, here taken to be 0.5 Å^2 [42]. Both terms act as low pass filters in the frequency domain for $V(x)$. We note here that the sample function is a critical part of the model and that the ANN failed to fit when a poor model of the amorphous film was used.

2.3. Aberration function

In the extended notation of Lupini et al. [19], which is derived from the well-established Krivanek notation [43], the aberration function $\chi(q)$ in reciprocal space can be written in the form

$$\chi(q) = 2\pi\sum_n \lambda^n \sum_{m=0}^{n+1}\left[\frac{C_{n,m,a}q^{n+1}\cos(m\phi) + C_{n,m,b}q^{n+1}\sin(m\phi)}{n+1}\right], \quad (2)$$

where $C_{n,m,a}$ and $C_{n,m,b}$ are aberration coefficients of order n and symmetry m, under the condition that $n+m$ is an odd number. The coefficients reduce to one when $m = 0$ (*e.g.*, for defocus and spherical aberration). In Eq. (2), q is the length of the vector q and ϕ is the azimuthal angle (with the optical axis in the center). Most software truncates $\chi(q)$ at $n = 5$. We use a similar truncation below, as explained in Section 3.2, as well as the traditional nomenclature for the lower orders (see Table 1).

2.4. Detector

The Ronchigram intensity $I(q)$ is typically recorded using a digital detector, resulting in a convolution of $I(q)$ with the point spread function (PSF) p of the camera:

$$I(\widetilde{q}) = |IFT\{FT(I(\widetilde{q}))FT(p(\widetilde{q}))\}| + c(\widetilde{q}), \quad (3)$$

where $\widetilde{q}$ indicates that the convolution is performed in the pixel dimension. The PSF introduces further attenuation of high spatial frequencies. The function $c(\widetilde{q})$ describes the counting noise of the camera. The absolute value is a reminder that the result of the convolution is real. A defocused Ronchigram (*e.g.,* with $C_{1,0} \sim -2000$ nm) of an amorphous thin object contains contrast that varies in a similar manner to that in an HRTEM image (recorded using parallel illumination). The PSF of the detector affects the frequency cutoff in the image and is comparable to scattering in the object [41]. This effect is not negligible and must be accounted for when training the ANN, otherwise its regression ability on experimental images is seriously compromised.

An example of the fidelity of the model is presented in Fig. 2, in the form of a comparison of an experimental defocused Ronchigram (Fig. 2a,b) and a simulated model from Eqs. (1)-3 (Fig. 2c,d) at 300 kV. Fig. 2e shows radial profiles of the fast Fourier transforms of the

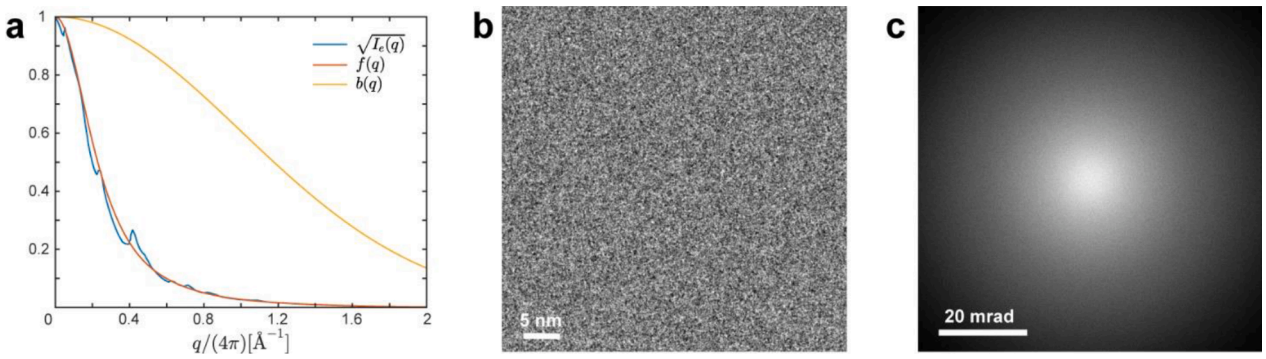

Fig. 1. (a) Plot of the square root of the intensity of elastic scattering from a thin film of a-C, as measured from electron diffraction (blue curve) [40] and compared with an analytical function for the atomic scattering factor of carbon $f(q)$ (red curve) [41] and the scattering profile expected from thermal vibrations $b(q)$ (orange curve) [42]. (b) Calculated $V(x)$ after considering the scattering described in (a). (c) Fourier transform $FT(V(x))$.

Table 1
Mean absolute error (MAE) in each of the 8 low-order aberrations after ANN training. The aberrations were generated randomly over the intervals reported in the 'Range' column. The MAE values were calculated from both the synthetic dataset (Section 4.1) and the experimental dataset (Section 4.2). The 'Symbol' column refers to the notation from Haider et al. [2], while the 'Coefficients' column refers to the notation used in the present work and in reference [19].

Aberration	Symbol	Coefficients (this work)	Range	MAE synth. data	MAE exp. data
Defocus	C_1	$C_{1,0}$	[−2200 nm, −1800 nm]	15 nm	31 nm
Spherical aberration	C_3	$C_{3,0}$	[−100 µm, 100 µm]	3.49 µm	15 µm
2-*fold astigmatism*	A_1	$C_{1,2,a}/C_{1,2,b}$	[−100 nm, 100 nm]	4.8 nm	17 nm
Coma	B_2	$C_{2,1,a}/C_{2,1b}$	[−1000 nm, 1000 nm]	48 nm	204 nm
3-*fold astigmatism*	A_2	$C_{2,3,a}/C_{2,3,b}$	[−1000 nm, 1000 nm]	39 nm	131 nm

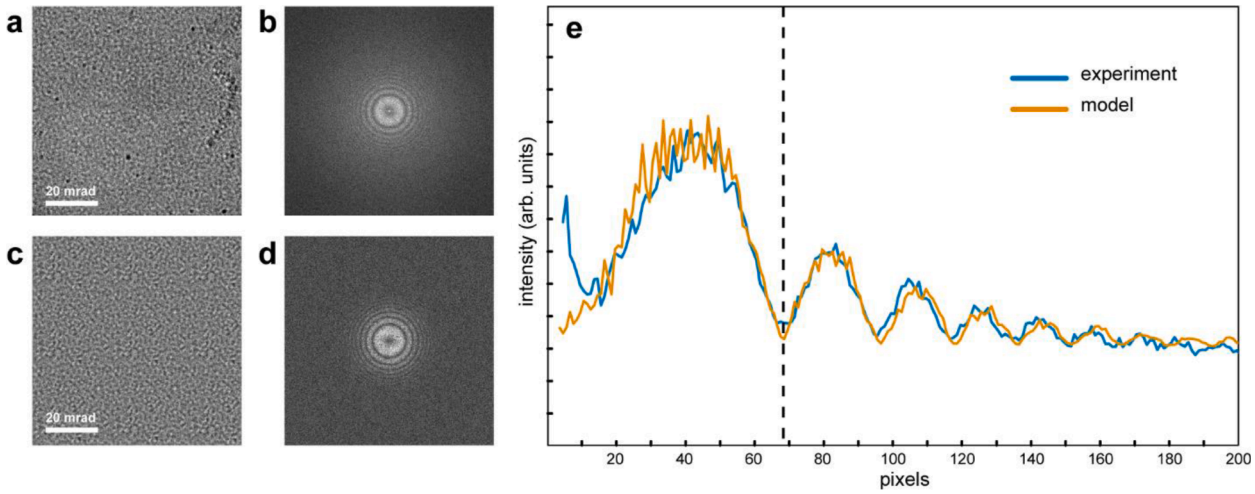

Fig. 2. Comparison between (a) an experimental defocused Ronchigram and (b) a simulated model according to Eqs. (1)-3 at 300 kV. The corresponding FFTs are shown in (b) and (d), respectively. (e) Radial profiles of the FFTs of the experimental defocused Ronchigram (blue curve) and the simulated model (orange curve). The other aberration values are set to zero. The defocus $C_{1,0}$ can be calculated from the first minimum in the profile ($C_{1,0} = -1890$ nm, dashed line).

experimental Ronchigram (blue line) and the simulated model (orange line). The intensities of the maxima and minima are well reproduced, as is the decrease in contrast at higher spatial frequencies.

3. Theory and calculations

The method involves dividing an experimental Ronchigram recorded from an amorphous C film (*e.g.*, a standard C film on a Cu mesh grid) into a set of sub-images and numerically performing a fast Fourier transform (FFT) of each sub-image. The results are stacked to form a 3D image, which is fed into the ANN for pattern recognition. Unlike the analytical method of Lupini et al. [19], the fitting procedure of the ANN analyzes the overall pattern and not the separate sub-images.

3.1. ANN model

In order to fit Ronchigram images, a convolutional neural network was implemented in Python using the TensorFlow library [44]. As ANNs are highly problem-specific mathematical tools, they need to be custom-tailored to improve their efficiency for a given task. A few distinct families of ANN are generally referenced in the literature. Among these families, some architectures excel in specific tasks and are given names. Our ANN is a custom network that belongs to the convolutional family [45]. It has similarities to other convolutional neural networks, such as AlexNet [46] and VGG16 [47]. We customized it in terms of the number and dimensions of the layers. The network consists of several sequential layers (deep structure), in which each layer has a specific functionality [48, 49]. The aim of the deep structure is to

recognize distinct features in the input images, in order to assign them to a specific class (*e.g.*, an animal, or a flower, or 'classification') or to relate them to a set of varying parameters (*e.g.*, the aberrations to be measured, or 'regression'). A sketch of our network is shown in Fig. 3a. It is made up of 5 sequential convolutional blocks (orange), each of which comprises a 2D convolutional layer, followed by a 2D average pooling layer and a dropout layer, followed by fully connected (or dense) layers (blue block). These convolutional blocks apply a filter to the input to create a feature map that summarizes the presence of detected features in the input. In each block, the convolutional layer convolves the input image (or feature map) with a set of kernels. The average pooling layer down-samples the feature map (after convolution) by dividing it into sub-regions and taking the average in each sub-region, of dimension [2 × 2] in our case. In order to avoid overfitting, which can occur when the ANN tries to memorize the training data instead of generalizing it to new data, as well as to improve robustness to noise, a dropout layer [50] is inserted after every pooling layer. In the dropout layer, a hyperparameter has the probability (20% in our case) to be set to zero before it is passed to the next convolutional block. We stacked 5 convolution blocks with increasing numbers of kernels in the 2D convolutional layers, in order to abstract features at different 'length-scales' in the input images. After the convolutional blocks, a flatten layer converts dimensionality to a single channel, in order to connect the extracted feature map to 3 dense layers that are fully connected each other (deep layers, analogous to neurons in biology), to reorganize the image features and finally to output the 8 parameters that are needed to calculate $C_{1,0}$, $C_{1,2}$, $C_{2,1}$, $C_{2,3}$ and $C_{3,0}$. A preliminary optimization of the network was performed by changing the kernel size in the convolutional layers from [3, 3] to [5, 5] and the starting set of kernels (the number of filters in the first convolutional layer) between 16, 32, 64 and 128. We found the lowest mean square error after convergence with 64 starting kernels with [3, 3] size, which results in ~2.7 M total hyperparameters in the network.

3.2. ANN training

A dataset was built with $N = 24{,}000$ sets of 25 (5 × 5) fast Fourier transforms (FFTs) from Ronchigram images (Fig. 3b). Each Ronchigram was calculated at angles up to 40 mrad according to Eq. (1) (with $C_{1,0} \sim -2000$ nm) and at 2048 × 2048 pixels to reduce artifacts in the Fourier transforms due to the rapidly varying transmission function $e^{-i\chi(q)}$. Each generated image $I(q)$ was then resized to 320 × 320 pixels and divided into a set of 5 × 5 sub-images, each of which was 64 × 64 pixels in size (Fig. 3c). The 25 FFTs from the sub-images were then calculated (Fig. 3d). Each set of 25 FFTs constitutes one of 24,000 inputs for ANN training, *i.e.*, the input matrix has dimensions (24,000, 64, 64, 25). The corresponding (24,000, j) matrix has j known aberration coefficients. For clarity, in the example shown in Fig. 3b-d the value of $C_{3,0}$ (spherical aberration) is exaggerated to show its effect on the FFTs ($C_{3,0} = 1$ mm). ANN training was limited to the $j = 8$ low-order aberration values, which were generated randomly in the intervals reported in Table 1. Residual aberrations up to $n = 5$ were generated according to a normal distribution centered around 0 (corrected values) with a standard deviation of 0.1 of their maximum absolute values, which were considered to be 10 μm for $n = 3$ (except for $C_{3,0}$), 100 μm for $n = 4$, and 10 mm for $n = 5$. The network was trained for 200 epochs (where an epoch is one complete pass of the training dataset through the algorithm) with a batch size of 40 (the number of sets of FFTs processed before the model was updated). From the full dataset, a validation dataset of 20% (*i.e.*, 4800 sets of FFTs) was initially extracted to provide a test against unseen data. The rest of the dataset (19,200 sets of FFTs) was used in the training procedure to fit the hyperparameters (training dataset). After convergence, the loss function (the function specifying the penalty for an incorrect estimate) was below 0.01 and was evaluated as the mean square error between true values and estimates of the 8 free parameters (or aberration coefficients).

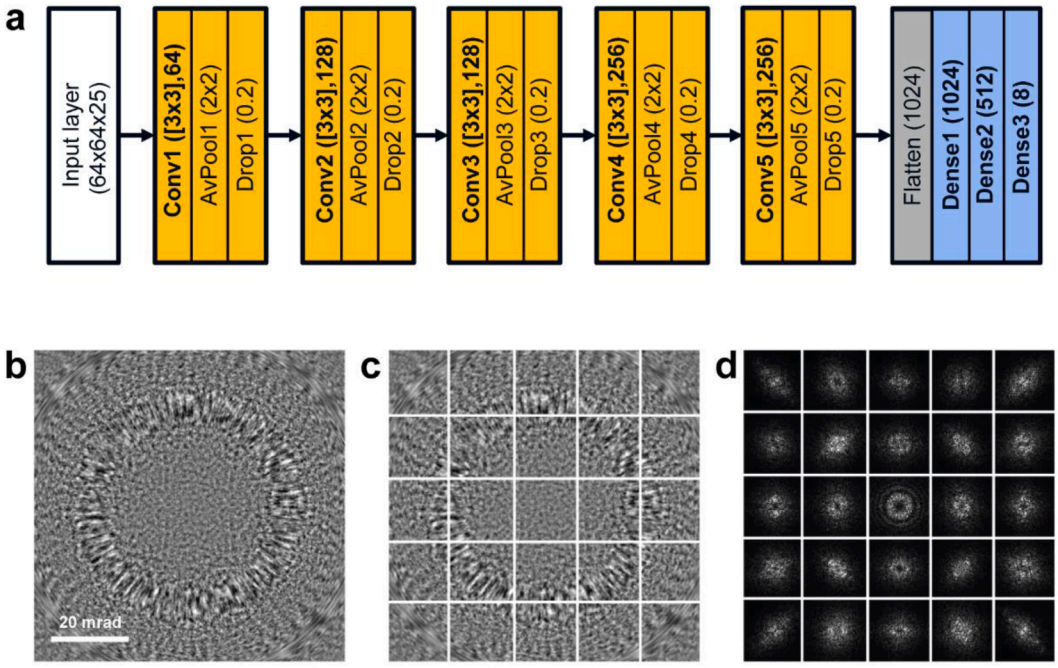

Fig. 3. (a) Sketch of the artificial neural network (ANN) used to fit synthetic Ronchigrams from a 5 × 5 set of FFTs. The ANN consists of 5 sequential convolutional blocks comprising a 2D convolutional layer with an increasing number of filters (64, 128, 128, 256, 256) and kernel size [3, 3], followed by a 2D average pooling layer with kernel size [2, 2] and a dropping layer with probability 0.2. After these blocks, a flatten layer connects the last convolutional block to the sequence of 3 fully connected (dense) layers with dimension 1024, 512 and 8, to finally output the 8 aberration values. The layers with hyperparameters are indicated in bold. (b-d) Example of a simulated Ronchigram used as a single input image for the ANN. (b) 2048 × 2048 defocused Ronchigram generated according to Eq. (3). (c) 320 × 320 resized Ronchigram divided into 5 × 5 sub-images of 64 × 64 pixels. (d) Corresponding FFTs from the sub-images in (c) constituting one (64 × 64 × 25) input for the ANN. For clarity, in this example the value of $C_{3,0}$ (spherical aberration) was exaggerated to show its effect on the FFTs ($C_{3,0} = 1$ mm).

4. Results and discussion

4.1. ANN results on synthetic data

Table 1 shows the mean absolute error (MAE) for the 8 fitted aberration parameters after ANN training. The errors are ~5% of the maximum values of the aberrations in their respective ranges. The ANN is therefore capable of reducing the aberrations by a factor of ~20. This result is promising, considering that only a single image acquisition is needed for the ANN to measure the aberration coefficients.

The precision achieved by the ANN in the estimations of the aberrations is limited (see Table 1). The dominant value limiting the resolution is two-fold astigmatism $C_{1,2}$, while the values for $C_{2,1}$, $C_{2,3}$ and $C_{3,0}$ are in the range expected for sub-Å resolution [51, 52]. In order to reach optimal resolution (*e.g.*, $d < 80$ pm or $\alpha > 15$ mrad at 300 kV), the residual value of $C_{1,2}$ should be kept below 1 nm.

The present AI procedure cannot routinely achieve this level of precision from a single Ronchigram. The observed lack of precision may result partly from the fact that the amount of information in a single Ronchigram is limited. However, a comparison with existing methods based on measurements from a single Ronchigram indicates a similar level of uncertainty [18, 19]. Possible solutions may be a refinement directly from a series of STEM images [53], or a refinement based on an iterative procedure [27].

The strength of the ANN-based diagnosis resides in its speed of ~80 ms on a machine with one Intel Core i7–7700 (4 cores with hyper-threading, meaning 8 threads), including image pre-processing (image sub-division and FFTs calculation). We expect that this time will not scale with the number of parameters to fit, if higher-order aberrations are included, considering that ANNs can fit thousands of parameters simultaneously [54]. The time is much shorter than the acquisition time of the image (~1 s), leaving the acquisition time as the limiting factor and making our method suitable for real-time diagnosis. For comparison, the standard STEM method [2] requires several images to build a tableau (*e.g.*, 7 images for diagnosis up to 2nd order and 17 images for full diagnosis up to 4th order, to be repeated at two defoci), resulting in precise aberration measurements at the expense of a relatively long total acquisition time.

We can exploit the ability of the ANN to measure all low-order aberrations from a single image in an iterative approach to further increase the resolution. In the present procedure, only an optimal fraction R of the ANN-calculated correction was applied each time. R was updated at each iteration by measuring the optimal STEM aperture α according to the quarter-wave criterion [55, 56], which is the radius of the region of frequencies q with aberration phase $|\chi(q)| \leq \pi/4$. Based on the aperture α, the resolution of the STEM probe can be derived, according to the Rayleigh criterion, to be $d = 0.61(\lambda/\alpha)$. We chose a correction factor $R = 1 - \alpha/\alpha_0$, where α_0 is the target optimal convergence, set for instance to $\alpha_0 = 12.1$ mrad, which corresponds to a target resolution $d_0 < 100$ pm at 300 kV. By using this criterion, the ANN typically reached convergence for correction in 1–5 iterations (*i.e.*, 5–10 s including acquisition time).

An example set of Ronchigram images for a defocus of $C_{1,0} = 0$ with corresponding probe images before and after aberration correction using iterative ANN correction is shown in Fig. 4, confirming that the optimal aperture α and resolution d improve after ANN correction, resulting in a probe that provides atomic spatial resolution.

Future improvements could be achieved, for instance, by increasing the training dataset (by at least a factor 10) or by increasing the complexity and topology of the ANN. Another possibility would be to test different segmentation strategies of the Ronchigrams into sub-images, such as increasing the number of sub-images (*e.g.*, from 5 × 5 to 7 × 7) for better sampling in the radial direction q, or by mixing different segmentations (*e.g.*, 5 × 5, 4 × 4, …). However, these

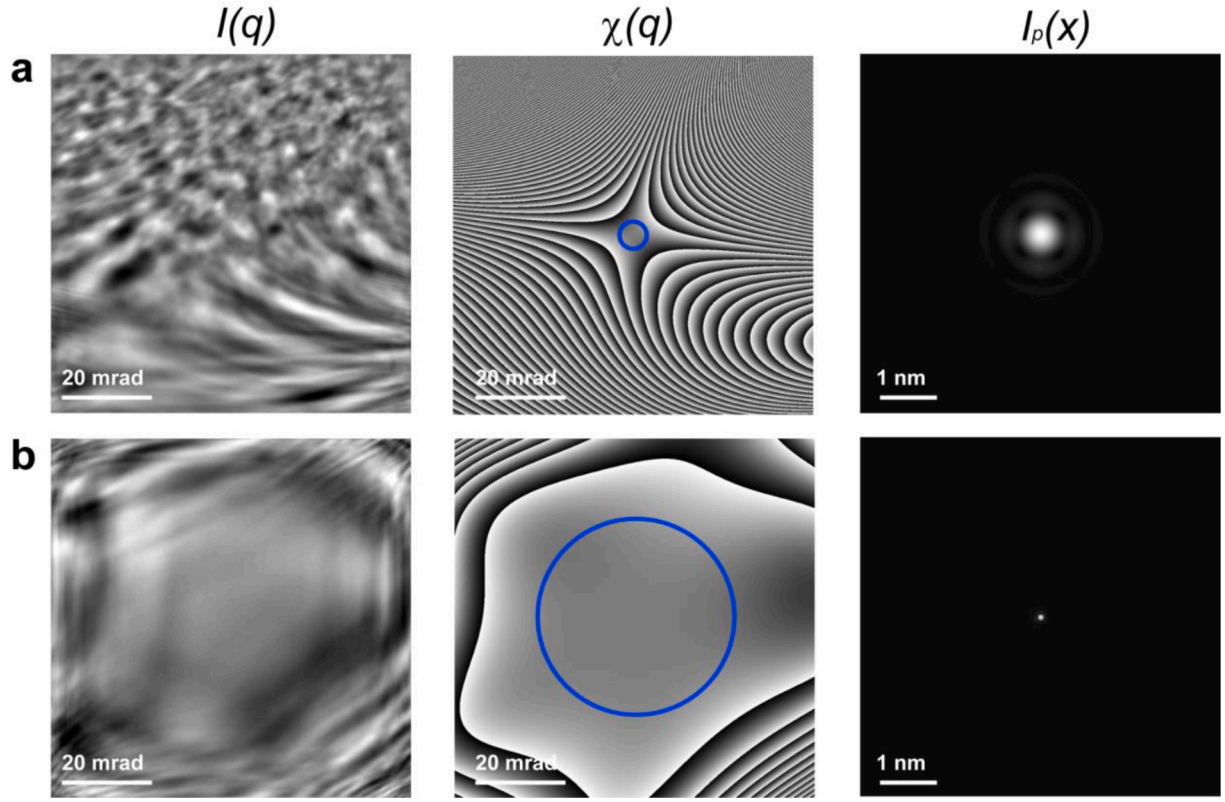

Fig. 4. Simulated images illustrating ANN correction. (a) Uncorrected probe. (b) Corrected probe after subtraction of the aberration values measured using the ANN. (Left) Ronchigram intensity $I(q)$ and (center) corresponding aberration phase $\chi(q)$ calculated at a defocus of $C_{1,0} = 0$ nm. The blue circle indicates the optimal probe aperture α used to estimate the resolution according to the Rayleigh criterion. (Right) Corresponding calculated probe intensity $I_p(x)$ in vacuum.

techniques require an increase in hardware memory requirements and computation time for generating the training dataset and for achieving ANN convergence.

4.2. ANN results on experimental data

Motivated by the results obtained using synthetic data, we tested the ANN on a limited set of experimental data. Since the network has never seen real data during training, effects due to the sample, environmental conditions and noise, or systematic errors such as the centering of the Ronchigram, will result in errors in estimates of the aberration values. Such errors can only be reduced with direct training on a large set of experimental data, which we cannot acquire automatically at present (and it would be too long a procedure to perform manually). Several series of Ronchigrams were recorded at 300 kV on a Thermo Fisher Scientific Spectra TEM equipped with a probe aberration corrector from CEOS GmbH. A convergence aperture of 64 mrad was used and 2048 × 2048 images were acquired with a Ceta camera (Thermo Fisher Scientific). The aberrations $C_{1,2}$, $C_{2,1}$, $C_{2,3}$ and $C_{3,0}$ were mis-tuned intentionally over a broad range of combinations. A defocused Ronchigram image was acquired for every mistuned value (9 images for $C_{1,2}$, $C_{2,1}$, $C_{2,3}$ and 2 images for $C_{3,0}$). $C_{1,0}$ was set to −2000 nm for all of the acquisitions. The mis-tuned aberrations were quantified from the Ronchigram images using the ANN. Before passing the image to the ANN, the Ronchigram was centered based on the center of expansion/contraction in a quick focal series (4–5 images). Fig. 5 shows ANN results obtained for $C_{1,2}$, $C_{2,1}$, $C_{2,3}$ and $C_{3,0}$ (blue points). The vertical bars are errors in the ANN estimations, calculated as 95% confidence intervals from the MAE values measured from the synthetic data (see Table 1). The horizontal bars are 95% confidence intervals from the STEM tableau built-in procedure. As expected, the performance on experimental data is worse than on synthetic data. By assuming that the estimation from the built-in method is the ground truth, the MAE values measured from the experimental data were found to be ~2 times larger for $C_{1,0}$, ~3 times larger for $C_{1,2}$, and $C_{2,3}$ and ~4 times larger for $C_{2,1}$ and $C_{3,0}$. The experimental MAE values are reported in Table 1.

The best results were seen for two-fold astigmatism ($C_{1,2}$) and three-fold astigmatism ($C_{2,3}$), with most values measured by the ANN close to STEM tableau values based on the 95% confidence intervals. Higher deviations were found for coma ($C_{2,1}$) and spherical aberration ($C_{3,0}$). In the case of coma, the higher error may result from an apparent beam tilt caused by a systematic error in the center of the Ronchigram on the camera. In the case of spherical aberration, the ANN measured a bias of approximately +15 µm with respect to the STEM tableau. This difference may result from a misinterpretation (or cross-talk) with defocus ($C_{1,0}$) due to the limited sampling in q from the 5 × 5 sub-division. A solution to this problem may be to increase the sub-division. However, we cannot exclude some drift in the aberration values due to instrumental instabilities during a long experimental session.

These first results are promising, demonstrating that it is possible to use an ANN to fit aberrations in real experiments. The results can be improved by refining the training dataset on experimental Ronchigram images at the microscope. In this way, the ANN can be trained against instabilities, such as drift of aberration values, shift of the image on the detector and noise. The synthetic model cannot reproduce the experimental data completely. The model can also be improved, for instance by accounting for small deviations seen at low frequencies in experimental FFTs, defects in the amorphous sample (*e.g.*, contamination during acquisition or irregularities in the film), or by building a better model for the (probably correlated) noise in images recorded using a real detector.

Possible future implementations can also involve training the ANN on higher-order aberrations (initially up to 5th order) and coupling it with miniaturized hardware (MEMS) that can be mounted on apertures or holders, in order to realize affordable hardware for beam shaping or aberration correction. It would be particularly interesting to test a similar ANN on crystalline materials [20–25] to verify if it can discern aberrations directly from a sample under investigation. It has been demonstrated that the complex structure of a diffraction spot in defocused Ronchigrams from very thin crystalline samples encodes

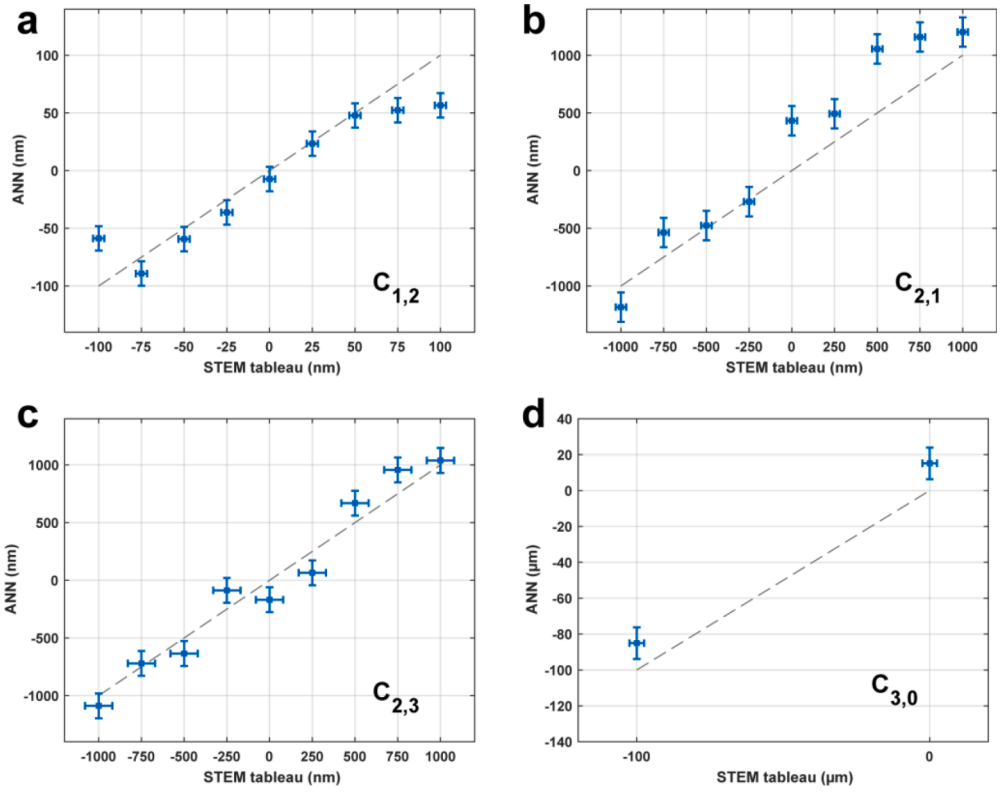

Fig. 5. (a-d) Experimental aberration values measured using the ANN (blue squares) plotted as a function of values obtained using the built-in STEM tableau method of the probe corrector of the microscope, which are taken as a reference. (a) Comparison for 2-fold astigmatism ($C_{1,2}$). (b) Comparison for coma ($C_{2,1}$). (c) Comparison for 3-fold astigmatism ($C_{2,3}$). (d) Comparison for spherical aberration ($C_{3,0}$). The error bars are calculated as 95% confidence intervals from MAE values from the synthetic data in Table 1 (vertical bars) and as 95% confidence intervals from the STEM tableau built-in procedure (horizontal bars).

information about aberration phase [23,24]. The use of ANN to extract this information would remove the need to move the probe onto an amorphous film, thereby reducing the acquisition time and the risk of drift of aberration values for fully automated experiments at the microscope [57–59].

5. Conclusions

We have demonstrated that an ANN can be used to measure aberrations from single Ronchigrams for fast online tuning, offering an approach that can be used to keep a STEM tuned and to assist an operator. Iterative correction, resulting in a spatial resolution better than 100 pm, has been demonstrated on synthetic data from a series of 5 consecutively acquired Ronchigrams. Despite the currently lower precision and accuracy of the ANN with respect to the established STEM tableau method when applied to experimental data, the speed of the ANN makes it promising for tuning and continuous diagnosis of lower-order aberrations during high-resolution STEM directly at the microscope.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Vincenzo Grillo reports financial support was provided by Horizon 2020. Rafal E. Dunin-Borkowski reports financial support was provided by Horizon 2020.

Data Availability

Data will be made available on request.

Acknowledgments

The authors acknowledge funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 964591 'SMART-electron'. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme (Grant no. 823717, project 'ESTEEM3').

References

[1] D.A. Muller, L.F. Kourkoutis, M. Murfitt, J.H. Song, H.Y. Hwang, J. Silcox, N. Dellby, O.L. Krivanek, Atomic-scale chemical imaging of composition and bonding by aberration-corrected microscopy, Science 319 (2008) 1073–1076.
[2] M. Haider, P. Hartel, H. Müller, S. Uhlemann, J. Zach, Current and future aberration correctors for the improvement of resolution in electron microscopy, Philos. Trans. Royal Soc. A 367 (2009) 3665–3682.
[3] S.J. Pennycook, The impact of STEM aberration correction on materials science, Ultramicroscopy 180 (2017) 22–33.
[4] M. Linck, H. Müller, P. Hartel, S. Perl, S. Uhlemann, M. Haider, On the benefit of aberration correction in cryo electron microscopy, Microsc. Microanal. 26 (2020) 2156–2157.
[5] M. Haider, H. Rose, S. Uhlemann, B. Kabius, K. Urban, Towards 0.1 nm resolution with the first spherically corrected transmission electron microscope, J. Electron Microsc. 47 (1998) 395–405.
[6] M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius, K. Urban, Electron microscopy image enhanced, Nature 392 (1998) 768–769.
[7] P.E. Batson, N. Dellby, O.L. Krivanek, Sub-ångstrom resolution using aberration corrected electron optics, Nature 418 (2002) 617–620.
[8] O.L. Krivanek, P.D. Nellist, N. Dellby, M.F. Murfitt, Z. Szilagyi, Towards sub-0.5 Å electron beams, Ultramicroscopy 96 (2003) 229–237.
[9] F. Hosokawa, H. Sawada, Y. Kondo, K. Takayanagi, K. Suenaga, Development of Cs and Cc correctors for transmission electron microscopy, Microscopy 62 (2013) 23–41.
[10] D.J. Schroeder, Astronom. Opt. (2000).
[11] R. Paxman, S. Crippen, Aberration correction for phased-array telescopes using phase diversity, SPIE (1990).
[12] E.J. Kirkland, The Transmission Electron Microscope. Advanced Computing in Electron Microscopy, Springer US, Boston, MA, 2010, pp. 5–27.
[13] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, K.H. Herrmann, Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms, Ultramicroscopy 3 (1978) 49–60.
[14] C.T. Koch, Aberration-compensated large-angle rocking-beam electron diffraction, Ultramicroscopy 111 (2011) 828–840.
[15] R.F. Klie, C. Johnson, Y. Zhu, Atomic-resolution STEM in the aberration-corrected JEOL JEM2200FS, Microsc. Microanal. 14 (2008) 104–112.
[16] J.M. Cowley, Electron diffraction phenomena observed with a high resolution STEM instrument, J. Electron Microsc. Tech. 3 (1986) 25–44.
[17] P.E. Batson, Aberration correction results in the IBM STEM instrument, Ultramicroscopy 96 (2003) 239–249.
[18] H. Sawada, T. Sannomiya, F. Hosokawa, T. Nakamichi, T. Kaneyama, T. Tomita, Y. Kondo, T. Tanaka, Y. Oshima, Y. Tanishiro, K. Takayanagi, Measurement method of aberration from Ronchigram by autocorrelation function, Ultramicroscopy 108 (2008) 1467–1475.
[19] A.R. Lupini, P. Wang, P.D. Nellist, A.I. Kirkland, S.J. Pennycook, Aberration measurement using the Ronchigram contrast transfer function, Ultramicroscopy 110 (2010) 891–898.
[20] Q.M. Ramasse, A.L. Bleloch, Diagnosis of aberrations from crystalline samples in scanning transmission electron microscopy, Ultramicroscopy 106 (2005) 37–56.
[21] T. Yamazaki, Y. Kotaka, Y. Kikuchi, K. Watanabe, Precise measurement of third-order spherical aberration using low-order zone-axis Ronchigrams, Ultramicroscopy 106 (2006) 153–163.
[22] K. Kuramochi, T. Yamazaki, Y. Kotaka, Y. Kikuchi, I. Hashimoto, K. Watanabe, Measurement of twofold astigmatism of probe-forming lens using low-order zone-axis ronchigram, Ultramicroscopy 108 (2008) 339–345.
[23] A.R. Lupini, S.J. Pennycook, Rapid autotuning for crystalline specimens from an inline hologram, J. Electron Microsc. 57 (2008) 195–201.
[24] K. Kimoto, K. Ishizuka, Rapid measurement of low-order aberrations using Fourier transforms of crystalline Ronchigrams, Ultramicroscopy 180 (2017) 59–65.
[25] H. Sawada, C.S. Allen, S. Wang, J.H. Warner, A.I. Kirkland, Aberration measurement of the probe-forming system of an electron microscope using two-dimensional materials, Ultramicroscopy 182 (2017) 195–204.
[26] E. Rotunno, A.H. Tavabi, P. Rosi, S. Frabboni, P. Tiemeijer, R.E. Dunin-Borkowski, V. Grillo, Alignment of electron optical beam shaping elements using a convolutional neural network, Ultramicroscopy 228 (2021), 113338.
[27] P. Rosi, A. Clausen, D. Weber, A.H. Tavabi, S. Frabboni, P. Tiemeijer, R.E. Dunin-Borkowski, E. Rotunno, V. Grillo, Automatic alignment of an orbital angular momentum sorter in a transmission electron microscope using a convolution neural network, Microsc. Microanal. (2022) 1–9, https://doi.org/10.1017/S143192762201248X.
[28] J. Verbeeck, A. Béché, K. Müller-Caspary, G. Guzzinati, M.A. Luong, M.Den Hertog, Demonstration of a 2 × 2 programmable phase plate for electrons, Ultramicroscopy 190 (2018) 58–65.
[29] S.V. Kalinin, C. Ophus, P.M. Voyles, R. Erni, D. Kepaptsoglou, V. Grillo, A. R. Lupini, M.P. Oxley, E. Schwenker, M.K.Y. Chan, J. Etheridge, X. Li, G.G.D. Han, M. Ziatdinov, N. Shibata, S.J. Pennycook, Machine learning in scanning transmission electron microscopy, Nat. Rev. Method. Primer. 2 (2022) 11.
[30] G.C.G. Berkhout, M.P.J. Lavery, J. Courtial, M.W. Beijersbergen, M.J. Padgett, Efficient sorting of orbital angular momentum states of light, Phys. Rev. Lett. 105 (2010), 153601.
[31] B.J. McMorran, T.R. Harvey, M.P.J. Lavery, Efficient sorting of free electron orbital angular momentum, New J. Phys. 19 (2017), 023053.
[32] A.H. Tavabi, P. Rosi, E. Rotunno, A. Roncaglia, L. Belsito, S. Frabboni, G. Pozzi, G. C. Gazzadi, P.-.H. Lu, R. Nijland, M. Ghosh, P. Tiemeijer, E. Karimi, R.E. Dunin-Borkowski, V. Grillo, Experimental demonstration of an electrostatic orbital angular momentum sorter for electron beams, Phys. Rev. Lett. 126 (2021), 094802.
[33] V. Grillo, A.H. Tavabi, F. Venturi, H. Larocque, R. Balboni, G.C. Gazzadi, S. Frabboni, P.-.H. Lu, E. Mafakheri, F. Bouchard, R.E. Dunin-Borkowski, R. W. Boyd, M.P.J. Lavery, M.J. Padgett, E. Karimi, Measuring the orbital angular momentum spectrum of an electron beam, Nat. Commun. 8 (2017) 15536.
[34] C. Zhang, Z. Baraissov, C. Duncan, A. Hanuka, A. Edelen, J. Maxson, D. Muller, Aberration corrector tuning with machine-learning-based emittance measurements and bayesian optimization, Microsc. Microanal. 27 (2021) 810–812.
[35] N. Schnitzer, S.H. Sung, R. Hovden, Optimal STEM Convergence Angle selection using a convolutional neural network and the Strehl ratio, Microsc. Microanal. 26 (2020) 921–928.
[36] R. Sagawa, F. Uematsu, K. Aibara, T. Nakamichi, S. Morishita, Aberration measurement and correction in scanning transmission electron microscopy using machine learning, Microsc. Microanal. 27 (2021) 814–816.
[37] N. Schnitzer, S.H. Sung, R. Hovden, Introduction to the Ronchigram and its Calculation with Ronchigram.com, Micros. Today 27 (2019) 12–15.
[38] M. Schowalter, J.T. Titantah, D. Lamoen, P. Kruse, Ab initio computation of the mean inner Coulomb potential of amorphous carbon structures, Appl. Phys. Lett. 86 (2005), 112102.
[39] A. Harscher, H. Lichte, Inelastic mean free path and mean inner potential of carbon foil and vitrified ice measured with electron holography, in: ICEM14, Cancun, Mexico, 31 August–4 September, Institute of Physics Publishing, Cancun, Mexico, 1998, pp. 553–554.
[40] B. O'Malley, I. Snook, D. McCulloch, Reverse Monte Carlo analysis of the structure of glassy carbon using electron-microscopy data, Phys. Rev. B 57 (1998) 14148–14157.
[41] L.M. Peng, Electron atomic scattering factors and scattering potentials of crystals, Micron 30 (1999) 625–648.

[42] C.B. Boothroyd, Quantification of high-resolution electron microscope images of amorphous carbon, Ultramicroscopy 83 (2000) 159–168.
[43] O.L. Krivanek, N. Dellby, A.R. Lupini, Towards sub-Å electron beams, Ultramicroscopy 78 (1999) 1–11.
[44] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow: a system for large-scale machine learning, in: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, USENIX Association, Savannah, GA, USA, 2016, pp. 265–283.
[45] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90.
[46] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L. D. Jackel, Backpropagation applied to handwritten zip code recognition, Neur. Comput. 1 (1989) 541–551.
[47] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, ICLR 2015. https://arxiv.org/abs/1409.1556.
[48] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36 (1980) 193–202.
[49] W. Xu, J.M. LeBeau, A deep convolutional neural network to analyze position averaged convergent beam electron diffraction patterns, Ultramicroscopy 188 (2018) 59–69.
[50] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[51] S. Uhlemann, M. Haider, Residual wave aberrations in the first spherical aberration corrected transmission electron microscope, Ultramicroscopy 72 (1998) 109–119.
[52] M. Haider, S. Uhlemann, J. Zach, Upper limits for the residual aberrations of a high-resolution aberration-corrected STEM, Ultramicroscopy 81 (2000) 163–175.
[53] R. Ishikawa, R. Tanaka, S. Morishita, Y. Kohno, H. Sawada, T. Sasaki, M. Ichikawa, M. Hasegawa, N. Shibata, Y. Ikuhara, Automated geometric aberration correction for large-angle illumination STEM, Ultramicroscopy 222 (2021), 113215.
[54] J. Turner, J. Cano, V. Radu, E.J. Crowley, M. O'Boyle, A. Storkey, Characterising across-stack optimisations for deep convolutional neural networks, in: 2018 IEEE International Symposium on Workload Characterization (IISWC), 2018, pp. 101–110.
[55] E.J. Kirkland, Linear Image Approximations. Advanced Computing in Electron Microscopy, Springer US, Boston, MA, 2010, pp. 29–60.
[56] H. Müller, S. Uhlemann, P. Hartel, M. Haider, Advancing the hexapole Cs-corrector for the scanning transmission electron microscope, Microsc. Microanal. 12 (2006) 442–455.
[57] M. Ziatdinov, O. Dyck, A. Maksov, X. Li, X. Sang, K. Xiao, R.R. Unocic, R. Vasudevan, S. Jesse, S.V. Kalinin, Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: chemical Identification and Tracking Local Transformations, ACS Nano 11 (2017) 12742–12752.
[58] M. Xu, A. Kumar, J. LeBeau, Automating electron microscopy through machine learning and USETEM, Microsc. Microanal. 27 (2021) 2988–2989.
[59] K.M. Roccapriore, O. Dyck, M.P. Oxley, M. Ziatdinov, S.V. Kalinin, Automated experiment in 4D-STEM: exploring emergent physics and structural behaviors, ACS Nano 16 (2022) 7605–7614, https://doi.org/10.1021/acsnano.1c11118.