



ELSEVIER

Contents lists available at ScienceDirect

## Ultramicroscopy

journal homepage: [www.elsevier.com/locate/ultramic](http://www.elsevier.com/locate/ultramic)

# Atom counting in HAADF STEM using a statistical model-based approach: Methodology, possibilities, and inherent limitations

A. De Backer<sup>a,\*</sup>, G.T. Martinez<sup>a</sup>, A. Rosenauer<sup>b</sup>, S. Van Aert<sup>a</sup><sup>a</sup> Electron Microscopy for Materials Science (EMAT), University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium<sup>b</sup> Institut für Festkörperphysik, Universität Bremen, Otto-Hahn-Alle 1, D-28359 Bremen, Germany

## ARTICLE INFO

Available online 17 May 2013

## Keywords:

High-resolution scanning transmission electron microscopy (HR-STEM)  
Atom counting  
Statistical parameter estimation theory

## ABSTRACT

In the present paper, a statistical model-based method to count the number of atoms of monotype crystalline nanostructures from high resolution high-angle annular dark-field (HAADF) scanning transmission electron microscopy (STEM) images is discussed in detail together with a thorough study on the possibilities and inherent limitations. In order to count the number of atoms, it is assumed that the total scattered intensity scales with the number of atoms per atom column. These intensities are quantitatively determined using model-based statistical parameter estimation theory. The distribution describing the probability that intensity values are generated by atomic columns containing a specific number of atoms is inferred on the basis of the experimental scattered intensities. Finally, the number of atoms per atom column is quantified using this estimated probability distribution. The number of atom columns available in the observed STEM image, the number of components in the estimated probability distribution, the width of the components of the probability distribution, and the typical shape of a criterion to assess the number of components in the probability distribution directly affect the accuracy and precision with which the number of atoms in a particular atom column can be estimated. It is shown that single atom sensitivity is feasible taking the latter aspects into consideration.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The physical properties of nanostructures are controlled by the exact chemical and structural composition. Therefore, it is important to reliably quantify parameters, such as the position of the atoms, the type of the atoms, and the number of atoms, in order to completely understand the structure. In the last few years, atomic resolution high-angle annular dark-field (HAADF) scanning transmission electron microscopy (STEM) has become a popular technique for this type of analysis because it provides direct interpretable images. As generally known, the HAADF STEM image intensities show Z-contrast and are highly sensitive to the number of atoms in the atom columns [1–6].

In the past, different approaches have been proposed to count the number of atoms of a nanostructure in HAADF STEM. Erni et al. [5] attempted to count the number of atoms by measuring the absolute value of the differences in intensity of neighbouring atom columns. LeBeau et al. [6] proposed to solve the counting problem by comparing simulated atom column intensities with normalised experimentally measured atom column intensities [7–11]. Recently, Van Aert et al. [12] introduced a new statistical model-based approach to count the number of atoms from HAADF STEM images

of a structure viewed along a zone-axis. This method may lead to promising results for the determination of the three-dimensional (3D) arrangement of atoms in crystalline nanoparticles [13–15]. By repeating the counting method for different viewing directions, the 3D atomic structure can be reconstructed using discrete tomography [16,17]. Moreover, this statistical model-based method promises counting results with single atom sensitivity for each atom column. In the present paper, the methodology presented in [12] to count atoms for structures containing one atom type will be discussed in detail together with a thorough study on the possibilities and inherent limitations imposed by unavoidable statistical uncertainties.

Although HAADF STEM images are highly sensitive to the number of projected atoms, it is impossible to directly count the atoms by means of visual interpretation only. Hence, the starting point of the method will be the use of statistical model-based estimation procedures to extract quantitative information from the HAADF STEM images. In order to detect individual atoms, it is assumed that the total scattered intensity per atom column in HAADF STEM images scales with the number of atoms in the column. This statistical model-based approach proceeds in two steps.

In the first step of the counting procedure, the HAADF STEM image will be considered as a data plane from which structure parameters have to be determined with the highest possible precision and accuracy [18]. Therefore, a parametric model is used which can be fitted to the experimental data using a criterion of goodness of fit, quantifying the similarity between the experimental

\* Corresponding author. Tel.: +32 3 265 33 31.

E-mail address: [Annick.DeBacker@ua.ac.be](mailto:Annick.DeBacker@ua.ac.be) (A. De Backer).

images and the model. The model parameters are estimated using an iterative optimisation scheme. In the past, this statistical parameter estimation approach has been applied to various electron microscopy data including TEM images [18–22], reconstructed exit waves [23–25], electron energy loss spectroscopy data [26,27], and STEM images [28,29]. For a quantitative analysis, an appropriate model that accurately describes all effects of the image formation, like dynamical electron diffraction effects and thermal diffuse scattering in HAADF STEM, should be employed. However, it would be very time consuming to use this type of models in an iterative optimisation scheme. As a result, an empirical incoherent imaging model, parametric in the width, the height, and the position of the atom columns will be used [29]. Estimates for the total scattered intensity per atom column can be obtained from the estimates of the parameters of this incoherent imaging model.

These estimates can be used as an input for the second step in the procedure to count the number of atoms per atom column. Therefore, the estimated total scattered intensities of all atom columns can be visualised in a histogram. Ideally, this histogram would consist of isolated components, where each component is generated by a set of columns having the same number of atoms projected on top of each other, that is in the direction perpendicular to the recorded image. In practice, however, the components are smeared out which makes a visual determination of their locations impossible. This is due to a combination of different effects, such as experimental detection noise, instabilities of the material under the electron beam, different vertical onset of columns of the same height, vacancies, relaxation at the boundaries, contamination, intensity transfer between columns, the influence of neighbouring columns of different heights. In order to solve this problem, the estimated total scattered intensities will be regarded as a statistical draw from an unknown probability distribution consisting of a superposition of Gaussian components. On the basis of the measured scattered intensities, the locations and the number of components of this distribution will be estimated. This will require a combination of an estimation procedure to determine the locations of the components and the use of order selection criteria to determine the number of statistically significant components. Once the distribution is known, the number of atoms in a particular projected column can be identified at atomic scale by assigning its scattered intensity to the component of the estimated probability distribution with the largest probability to generate this scattered intensity. As an example, the full method will be applied to a multislice simulation of an HAADF STEM image of a gold wedge.

By means of simulations the possibilities and inherent limitations of the method to estimate the number of atoms will be studied. For this purpose, the performance of different order selection criteria will be compared and the interpretation of the order selection criterion which performs best will be studied in more detail. Furthermore, the precision with which the number of atoms in a structure can be determined using this statistical model-based method will be investigated.

The remainder of this article will be organised as follows. In Section 2, the methodology to estimate the number of atoms of a nanostructure using statistical methods is described in detail. This is followed by an example in Section 3 illustrating the successive steps to count the number of atoms. In Section 4, the possibilities and inherent limitations of the method are investigated. Finally, in Section 5, conclusions are drawn.

## 2. Methodology

### 2.1. Model-based parameter estimation

The first step of the proposed method is to obtain precise and accurate measurements for the total intensity of the scattered

electrons for each atom column from the HAADF STEM image. Therefore, use is made of statistical parameter estimation theory [18,30]. The details of this approach are described in [29]. This approach requires a model-based method, in which the HAADF STEM image is considered as a data plane from which unknown parameters have to be estimated in a statistical way. For a successful application, a parametric model describing the expectations of the pixel values of the HAADF STEM images should be available. Assuming a HAADF detector, an incoherent STEM image will be formed, which can be written as a convolution between an object function and the probe intensity [31–33]. The model describing the expectations of the image intensities at the pixels  $(k,l)$ , corresponding to the STEM probe at position  $\mathbf{r}_{kl} = (x_k, y_l)^T$  is given by the following incoherent expression:

$$f_{kl}(\theta) = f(\mathbf{r}_{kl}; \theta) = O(\mathbf{r}_{kl}; \theta) * P(\mathbf{r}_{kl}) \quad (1)$$

with  $O(\mathbf{r}; \theta)$  the object function depending on a set of unknown structure parameters  $\theta$  and  $P(\mathbf{r})$  the probe intensity depending on a set of probe parameters including the acceleration voltage, the objective aperture semi-angle, defocus, spherical aberration constant, and higher order aberration coefficients.

The object function contains information about the specimen scattering, which is related to the projected potential, and the HAADF detector [33]. This function is sharply peaked at the atom column positions and can therefore be modelled as a superposition of Gaussian peaks. Therefore, the expectation of the object function at the pixel  $(k,l)$  at the position  $(x_k, y_l)$  of the HAADF STEM image can be described as

$$O(\mathbf{r}_{kl}; \theta) = \zeta + \sum_{n=1}^N \eta_n \exp\left(-\frac{(x_k - \beta_{x_n})^2 + (y_l - \beta_{y_n})^2}{2\rho^2}\right) \quad (2)$$

where  $\zeta$  is a constant background,  $\rho$  is the width of the Gaussian peak,  $\eta_n$  is the height of the  $n$ th Gaussian peak,  $\beta_{x_n}$  and  $\beta_{y_n}$  are the  $x$ - and  $y$ -coordinate of the  $n$ th atom column, respectively, and  $N$  is the total number of analysed atom columns. The unknown parameters are thus given by the parameter vector

$$\theta = (\beta_{x_1}, \dots, \beta_{x_N}, \beta_{y_1}, \dots, \beta_{y_N}, \rho, \eta_1, \dots, \eta_N, \zeta)^T \quad (3)$$

From the parametrised object function the volume under a peak above the background can be calculated as follows:

$$V_n = 2\pi\eta_n\rho^2 \quad (4)$$

These numbers reflect the total intensity of electrons scattered toward the HAADF detector for every atom column and therefore correspond to scattering cross sections.

The unknown parameters  $\theta$  of the model are estimated in the least squares sense. The uniformly weighted least squares estimates  $\hat{\theta}$  are given by the values of  $\mathbf{t}$  that minimise the uniformly weighted least squares criterion:

$$\hat{\theta} = \arg \min_{\mathbf{t}} \sum_{k=1}^K \sum_{l=1}^L (w_{kl} - f_{kl}(\mathbf{t}))^2 \quad (5)$$

with  $w_{kl}$  the value of the recorded HAADF STEM image at the pixel  $(k,l)$  and the function  $f_{kl}$  given by Eq. (1). Estimates of the total scattered intensities  $\hat{V}_n$  directly follow when using Eq. (4). In the next paragraph, it will be described how these estimates can be used to determine the number of atoms in each atom column.

### 2.2. Probability distribution of the estimated peak volumes

The estimated scattered intensities or peak volumes calculated using statistical parameter estimation can be visualised in a histogram. Ideally, this histogram would consist of isolated components, where each component is generated by a set of atom columns having the same number of atoms projected on top of each other. In practice, however, the components are smeared out

such that it is impossible to visually determine their locations. This is due to a combination of different effect, such as experimental detection noise, instabilities of the material under the electron beam, different vertical onset of columns of the same height, vacancies, relaxation at the boundaries, contamination, intensity transfer between columns, the influence of neighbouring columns of different heights. Therefore, instead of visually interpreting histograms, the location and the number of components of the unknown distribution underlying the histogram are estimated on the basis of the measured total scattered intensities. A combination of an algorithm to estimate the locations of the components and the use of order selection criteria to evaluate the number of statistically significant components is proposed [12,13,34]. This procedure, of which the result only depends on the values of the estimated peak volumes, leads to an accurate and precise estimate of the probability distribution as will be discussed in Section 4. Another advantage is that the result is independent of the subjective choice of bins in the histogram.

The estimated scattered intensities are regarded as independent statistical draws from a so-called Gaussian mixture model. In a sense, the assumption of independent statistical draws implies that cross-talk between neighbouring atomic columns is not significantly contributing [18]. The model is defined as a superposition of Gaussian components and describes the probability that a specific intensity value is observed. The probability density function of a mixture model with  $G$  components can parametrically be written as

$$f_{\text{mix}}(\hat{\mathbf{V}}; \Psi_G) = \sum_{g=1}^G \pi_g \phi_g(\hat{\mathbf{V}}; \mu_g, \sigma) \quad (6)$$

The symbol  $\Psi_G$  represents the vector containing all unknown parameters in the mixture model with  $G$  components:

$$\Psi_G = (\pi_1, \dots, \pi_{G-1}, \mu_1, \dots, \mu_G, \sigma)^T \quad (7)$$

The parameters  $\pi_g$ ,  $\mu_g$  and  $\sigma$  denote the mixing proportion of the  $g$ th component, the location of the  $g$ th component and the width of the components, respectively. The mixing proportions  $\pi_g$  define the probability that a column contains  $g$  atoms with  $g$  ranging between 1 and a maximum number of components  $G$ , which in general is not known beforehand. The mixing proportions  $\pi_g$  sum up to unity. Therefore, the  $G$ th mixing proportion is omitted in the parameter vector  $\Psi_G$ . Each component  $\phi_g(\hat{\mathbf{V}}; \mu_g, \sigma)$  of the mixture model is generated by a set of atom columns having the same number of atoms. Generally, if there is an increase of one atom between subsequent components, component  $g$  corresponds to atom columns with  $g$  atoms, and  $G$  is the number of atoms found in the thickest atom column. Depending on the nanostructure under study, it is possible that the first component does not correspond to atom columns having one atom, but correspond to columns with a higher number of atoms. In that case, an offset should be taken into account when assigning number of atoms to atom columns of the different components.

The components  $\phi_g(\hat{\mathbf{V}}; \mu_g, \sigma)$  are modelled as homoscedastic normal components, i.e. with common variance  $\sigma^2$ :

$$\phi_g(\hat{\mathbf{V}}; \mu_g, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(\hat{\mathbf{V}} - \mu_g)^2}{2\sigma^2}\right) \quad (8)$$

The parameters  $\Psi_G$  of the Gaussian mixture model, Eq. (6), can be computed using the maximum likelihood estimator [34–37] for which the estimates are found by maximising the log likelihood function of the unknown parameter vector  $\Psi_G$ , which is defined as

$$\log L(\Psi_G) = \sum_{n=1}^N \log\left(\sum_{g=1}^G \pi_g \phi_g(\hat{V}_n; \mu_g, \sigma)\right) \quad (9)$$

The maximum likelihood estimates of  $\Psi_G$  are computed by solving the likelihood equation:

$$\frac{\partial \log L(\Psi_G)}{\partial \Psi_G} = 0 \quad (10)$$

This equation can only be solved iteratively. A very useful method is the so-called expectation maximisation algorithm [38]. Using this algorithm, a solution for the parameters  $\Psi_G$  can be obtained for a fixed number of  $G$  components. The algorithm proceeds in two steps: the E-step for expectation and the M-step for maximisation. In practice, the posterior probability that the  $n$ th peak volume with observed value  $\hat{V}_n$  belongs to the  $g$ th component of the mixture is calculated in the E-step, i.e.

$$\tau_g(\hat{V}_n; \Psi_G^{(k)}) = \pi_g^{(k)} \phi_g(\hat{V}_n; \mu_g^{(k)}, \sigma^{(k)}) \Big/ \sum_{h=1}^G \pi_h^{(k)} \phi_h(\hat{V}_n; \mu_h^{(k)}, \sigma^{(k)}) \quad (11)$$

for  $g = 1, \dots, G$  and  $n = 1, \dots, N$ . The iteration step is denoted by  $k$ .

The M-step calculates the updated estimate  $\Psi_G^{(k+1)}$ . The updated estimates for the mixing proportions,  $\pi_g^{(k+1)}$ , can be calculated as follows:

$$\pi_g^{(k+1)} = \sum_{n=1}^N \tau_g(\hat{V}_n; \Psi_G^{(k)}) / N \quad (12)$$

for  $g = 1, \dots, G$ . Updates for the unknown parameters  $\mu_g^{(k+1)}$  and  $\sigma^{(k+1)}$  of the component densities are given by

$$\mu_g^{(k+1)} = \sum_{n=1}^N \tau_g(\hat{V}_n; \Psi_G^{(k)}) \hat{V}_n \Big/ \sum_{n=1}^N \tau_g(\hat{V}_n; \Psi_G^{(k)}) \quad (13)$$

$$\sigma^{(k+1)} = \sum_{g=1}^G \sum_{n=1}^N \tau_g(\hat{V}_n; \Psi_G^{(k)}) (\hat{V}_n - \mu_g^{(k+1)})^2 / N \quad (14)$$

By means of simulations of which the design is discussed in Section 4.3, it has been found that unbiased estimates for the parameters of the mixture model are obtained with following input values for the algorithm. Reliable starting values for the mixing proportions  $\pi_g$  are given by equal probabilities:

$$\pi_g^{(0)} = \frac{1}{G} \quad (15)$$

The starting value for the variance  $\sigma$  of the components can be chosen as

$$\sigma^{(0)} = \frac{\max(\hat{\mathbf{V}}) - \min(\hat{\mathbf{V}})}{2 \cdot G} \quad (16)$$

For the starting values of the locations, different sets are used with the locations ranging between the minimum and maximum of the data set. The E- and M-step are iterated until convergence is attained and estimates are obtained for each set of starting parameters. The estimates  $\Psi_G$  of the parameters in the mixture model are then given by the set of estimated parameters having the maximal log likelihood value. For the true number of components  $G$ , the obtained locations  $\mu_g$  of the components correspond to the average total scattered intensities for sets of atom columns having the same number of atoms.

However, so far, only estimation of the probability distribution of the scattered intensities is considered presuming a specific number of components. Section 2.3 deals with determining the number of components  $G$  in a mixture model using order selection criteria.

### 2.3. Assessing the number of components in the mixture model using order selection criteria

In Section 2.2, it is explained how we can obtain the probability density for a mixture model presuming a certain number of components in the mixture model. However, determining the

order of a mixture model, i.e. the number of components in the distribution of the total scattered intensities using so-called order selection criteria, is a difficult task. Different approaches are available to assess the number of components in a mixture model. A brief overview is given here. For other approaches and references, we refer to McLachlan and Peel [34].

The information criteria for model selection calculate a measure for the goodness of fit taking the complexity of the model into account. Therefore, the criteria include a criterion specific penalty term which increases with the number of parameters in the Gaussian mixture model, i.e. with the number of components in the distribution. As such, a good tradeoff between the goodness of fit and the complexity of the model arises. The general expression for information criteria is twice the negative penalised log likelihood criterion:

$$-2 \log L(\hat{\Psi}) + 2C \quad (17)$$

where the first term accounts for the lack of fit and the second term for the complexity of the model. The model that minimises the negative penalised log likelihood is selected as the probability density describing the observed peak volumes. In practice, the values of the criterion will be plotted as a function of the number of components. In the ideal case, the values of the criterion will show a sharp minimum for the true number of components of the underlying probability density function.

Different criteria which all account for the complexity of the model in a different way have been proposed in the literature.

In Akaike's Information Criterion (AIC) [39], the penalty term simply equals the total number of parameters of the model:

$$AIC(G) = -2 \log L(\hat{\Psi}) + 2d \quad (18)$$

with  $d$  the number of parameters in the mixture model where  $d=2G$  assuming Eq. (6). A very similar criterion is the Generalised Information Criterion (GIC) [40], where the contribution of the penalty term can be modified by the parameter  $r$ :

$$GIC(G) = -2 \log L(\hat{\Psi}) + r \cdot d \quad (19)$$

In the present paper the parameter  $r$  for the GIC is chosen equal to 3 [41], penalising more heavily for the complexity of the model as compared to AIC.

Another criterion considered here is the Bayesian Information Criterion (BIC) [42]:

$$BIC(G) = -2 \log L(\hat{\Psi}) + d \log(N) \quad (20)$$

Here, the penalty term depends on the sample size  $N$ . If the sample size  $N > 8$ , then BIC accounts more for the complexity of the model as compared to AIC.

In the Classification Likelihood information Criterion (CLC) [43], the entropy  $EN(\hat{\tau})$  associated to the model with  $G$  components is used as the term that penalises the complexity of the model:

$$CLC(G) = -2 \log L(\hat{\Psi}) + 2EN(\hat{\tau}) \quad (21)$$

The entropy associated to the model with  $G$  components is given by

$$EN(\hat{\tau}) = - \sum_{g=1}^G \sum_{n=1}^N \tau_g(\hat{V}_n; \hat{\Psi}) \log \tau_g(\hat{V}_n; \hat{\Psi}) \quad (22)$$

where  $\tau_g(\hat{V}_n; \hat{\Psi})$  is given by expression (11). The entropy of the model expresses how well the components of the mixture model are separated, i.e. how many overlap there is between neighbouring components. If the components are well separated, then  $EN(\hat{\tau})$  is close to zero. In other words, the separation of the components in the mixture model defines how drastically the log likelihood is penalised.

The Approximate Weight of Evidence criterion (AWE) depends on this entropy term, the number of parameters and the sample

size  $N$  [44]:

$$AWE(G) = -2 \log L(\hat{\Psi}) + 2EN(\hat{\tau}) + 2d(3/2 + \log N) \quad (23)$$

If the components are well separated, the performance of the AWE criterion resembles BIC.

Since BIC and CLC tend to underestimate and overestimate the number of components of the probability distribution respectively [45,46], the Integrated Classification Likelihood criterion (ICL) is introduced to overcome these shortcomings [47]:

$$ICL(G) = -2 \log L(\hat{\Psi}) + 2EN(\hat{\tau}) + d \log N \quad (24)$$

In Section 4.1, the performance of these criteria is compared by means of simulations.

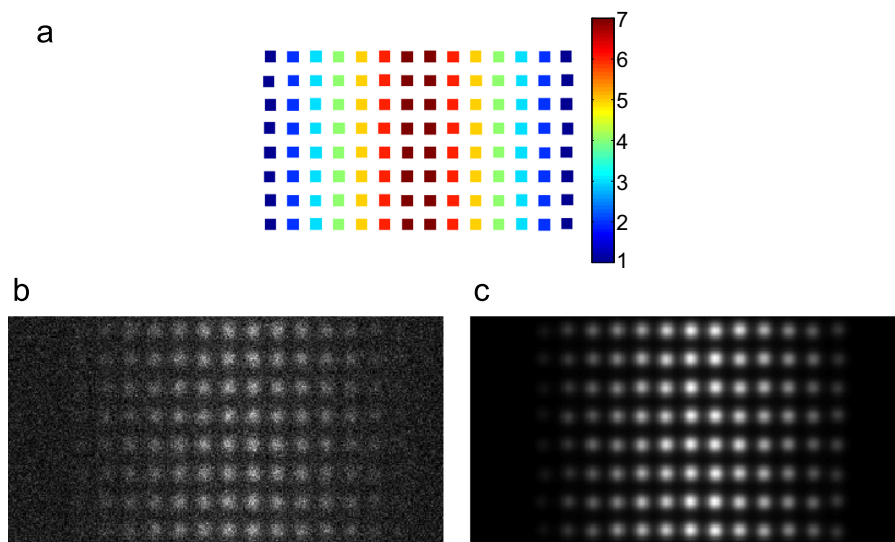
### 3. Counting atoms based on a simulated Au wedge

An example illustrating the procedure to count the number of atoms is worked out in this section. The STEMsim software [48] has been used to simulate a HAADF STEM image of a Au wedge. The parameters of the multislice simulation are listed in Table 1. The input structure for the STEM simulation is shown in Fig. 1(a). The noise in the simulated image, shown in Fig. 1(b), has been assumed to be Poisson distributed. First, estimates for the unknown object parameters are obtained from the simulated image in the least squares sense using the expectation model given by Eq. (1). The thus obtained model for the HAADF STEM image evaluated at the estimated parameters is shown in Fig. 1(c). The volume under the peak above the background  $\hat{V}$ , i.e. the total scattered intensity, is calculated from the estimated parameters for each atom column. These estimated peak volumes  $\hat{V}$  are used to determine the number of atoms in each atom column. The estimated peak volumes  $\hat{V}$  are analysed through the combination of the expectation maximisation algorithm to determine the component locations of the Gaussian mixture model and the ICL criterion to assess the number of significant components in the mixture model. Fig. 2(b) shows the ICL. The criterion shows a sharp minimum at 7, corresponding to the estimated number of components in the probability distribution. Fig. 2(a) shows the histogram of the estimated peak volumes together with the estimated mixture model. From Fig. 2(a), it is clear that the Gaussian components are approximately equidistantly separated meaning that the estimated peak volumes increase linearly with the number of atoms in an atom column. This knowledge can be used to determine the number of atoms in the atom columns assigned to the first component. The mean of the first component equals  $0.25 \text{ e}^{-\text{\AA}^2}$ . The average increment between the successive peak volumes equals  $0.23 \text{ e}^{-\text{\AA}^2}$ . From this ratio between both values, it can be concluded that the number of atoms in the atom

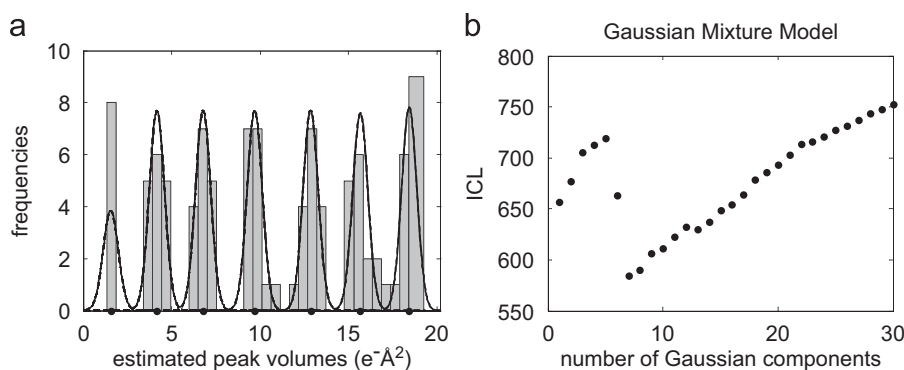
**Table 1**

Parameters for the multislice simulation of an HAADF STEM image of a Au wedge using the STEMsim software.

Parameter	Symbol	Value
Slice thickness	$Z_{\text{slice}}$ (Å)	2.04
Debye-Waller factor	$B$ (Å <sup>2</sup> )	0.6373
Acceleration voltage	$V_0$ (kV)	300
Defocus	$e$ (Å)	-83.01
Spherical aberration	$C_s$ (mm)	0.035
Spherical aberration of 5th order	$C_5$ (mm)	0
Convergence angle	$\alpha$ (mrad)	21.7881
FWHM of the source image	$\text{FWHM}_S$ (Å)	0.7
Pixel size	$\Delta x = \Delta y$ (Å)	0.09955
Number of pixels $x$ -direction	$K$	320
Number of pixels $y$ -direction	$L$	170
Total number of incident electrons	$N_i$	$8.3 \times 10^6$



**Fig. 1.** Input structure, simulated HAADF STEM image and corresponding model of a Au wedge. (a) Input structure, (b) simulation and (c) model.



**Fig. 2.** Analysis of the estimated peak volumes: (a) histogram of the estimated peak volumes  $\hat{V}$  with the curve indicating the estimated mixture model and the points indicating the estimated locations of the components, (b) ICL criterion evaluated as a function of the number of Gaussian components in the Gaussian mixture model. (a) Histogram and (b) ICL.

columns assigned to the first component equals one. Finally, based on the analysis shown in Fig. 2(a) and (b), a figure can be constructed that shows the number of atoms per atom column by assigning each total scattered intensity to the component of the estimated probability distribution with the largest probability for this scattered intensity. This counting result equals the original input used for the STEMsim simulation shown in Fig. 1(a). The total number of atoms in the Au wedge equals 448.

#### 4. Possibilities and inherent limitations

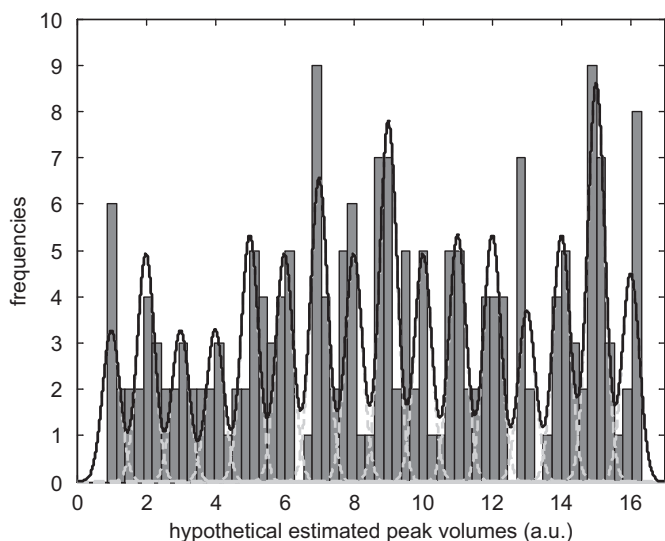
In Sections 2 and 3, it is described and shown how values for the total scattered intensities per atom column can be obtained from HAADF STEM images and next how the distribution of these scattered intensities is analysed to determine the number of atoms per atom column using statistical model-based methods. Other experimental examples including counting results for an embedded nanoparticle [13], a free-standing core-shell nanocrystal [14], ultrasmall clusters [15] and a larger Au nanorod [12] have been described elsewhere. In this section, we will discuss the possibilities and inherent limitations of this statistical model-based atom counting method.

The ultimate goal is to determine the number of atoms in an atom column with the highest possible accuracy and precision. However, modelling errors in the Gaussian mixture model for the

probability distribution limit the accuracy resulting in systematic errors for the number of atoms. In Sections 4.1 and 4.2 it is studied under which conditions unbiased measurements can be obtained for the number of atoms in an atom column by using information criteria to assess the order of the mixture model. On the other hand, the precision on the estimated number of atoms is limited due to the unavoidable presence of noise in the estimated scattered intensities. In Sections 4.3 and 4.4, the variance on the parameters of the Gaussian mixture model and subsequently the precision with which the number of atoms can be determined is analysed.

##### 4.1. Performance of the order selection criteria

In Section 2.3, different order selection criteria which are present in the literature are briefly discussed. In this section, the performance of these criteria is studied by simulating random draws from probability distributions as a function of the relative width of the Gaussian components. This relative width expresses the ratio between the width  $\sigma$  of the Gaussian components and the increment  $\delta$  between successive components. The histograms are simulated using realistic input values based on experimental images. The locations of the different components in the histogram are equidistantly chosen as it is assumed that the peak volumes almost linearly increase with the number of atoms in a column for thin specimens. Histograms with 16 components



**Fig. 3.** Example of a histogram (random draw) together with the underlying simulated probability distribution with 16 components and relative width  $\sigma/\delta = 0.25$ .

**Table 2**

Performance of different order selection criteria in terms of the percentage of correctly determined number of components as a function of the relative width  $\sigma/\delta$  of the components. The average number of data points per component  $N/G$  equals 12.5. The largest value  $\sigma/\delta$ , for which a percentage above 50 is still observed, is highlighted for each criterion.

$\sigma/\delta$	CLC	AIC	GIC	AWE	BIC	ICL
0.01	77	81	91	100	100	100
0.05	45	45	80	100	97	98
0.10	14	21	60	100	93	98
0.15	19	27	77	92	99	100
0.20	27	38	74	0	68	99
0.25	8	45	44	0	0	78
0.30	1	30	1	0	0	8

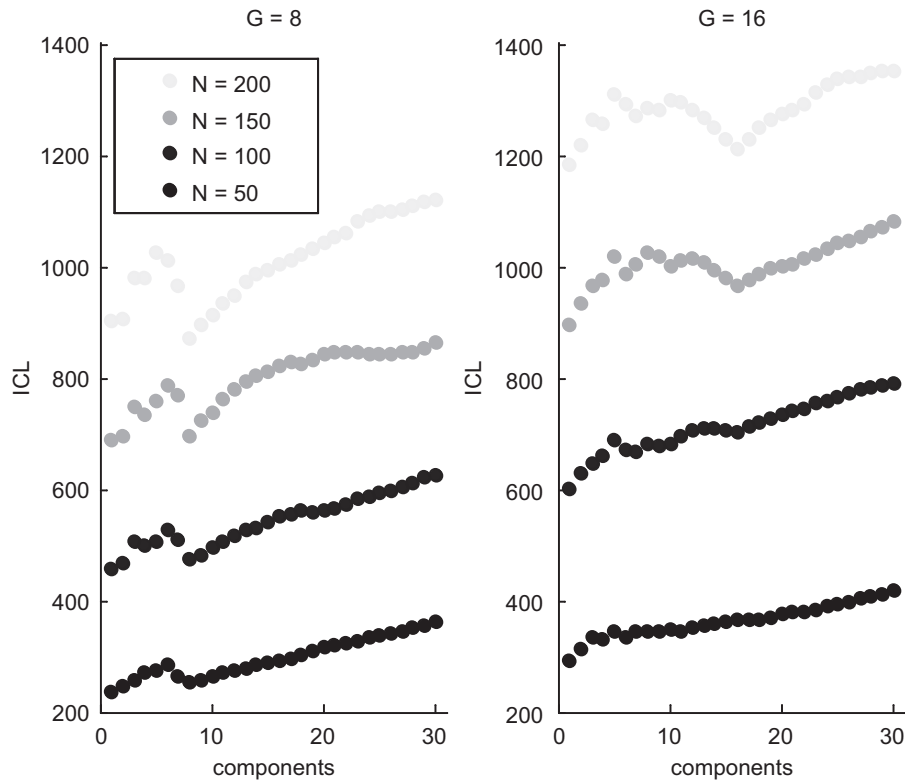
located at  $\mu = 1, 2, \dots, 16$  are simulated using the probability distribution given by Eq. (6). The relative width  $\sigma/\delta$  of these components is varied between  $0.01 \leq \sigma/\delta \leq 0.30$  and describes how well the different components in the histogram are separated. The proportions  $\pi_g$  for the different components are multinomially distributed with equal probability of  $1/16$ . The number of observations  $N$ , i.e. the number of atom columns, equals 200. For each value of the relative width  $\sigma/\delta$ , 100 random draws from the Gaussian mixture model are simulated. An example of such a random draw of a probability distribution is shown in Fig. 3. The information criteria, which are discussed in Section 2.3, are evaluated for these simulated random draws from the distributions assuming that the number of components may vary between 5 and a maximum value of 30. This minimum of 5 is set based on the characteristic shape of the criterion discussed in more detail in Section 4.2. In Table 2, the performance of different order selection criteria is shown in terms of the percentage of correctly determined number of components as a function of the relative width of the Gaussian components. The number of components is chosen based on the minimum of the criterion without visual inspection of the criterion. From the results in Table 2, it can be concluded that the ICL outperforms the other information criteria for assessing the number of components in a Gaussian mixture model for atom counting applications. In combination with visual inspection of the criterion and a higher number of peak volumes  $N$ , it is expected that the percentage of correctly determined number of components will increase. Consequently also more reliable results

will be possible for higher relative widths  $\sigma/\delta$ , as will be shown in Section 4.2.

#### 4.2. Interpretation of the Integration Classification Likelihood criterion

The purpose of this paragraph is to get more insight in the ICL since it turns out that this criterion has optimal properties. When evaluating ICL as a function of the number of components of the Gaussian mixture model, it has been observed that the true number of components often corresponds to a local optimum. This knowledge may help interpreting ICL in a correct way and may improve the reliability of the selected number of components. Fig. 4 shows the shape of ICL for an underlying distribution consisting of 8 and 16 components as a function of an increasing number of observations corresponding to the number of atom columns  $N$ . The random draws from the distributions are simulated under the same conditions as described in Section 4.1. The relative width of the components equals  $\sigma/\delta = 0.25$ . Two main conclusions can be drawn from Fig. 4. First, it is observed that the shape of the criterion is comparable for the different cases. It appears that ICL is relatively small for a small number of components and shows an extra minimum at the correct number of components, especially when the number of observations is smaller. This behaviour hampers finding the correct number of components of the underlying distribution by using ICL without visual inspection and explains the sudden drop down in the performance of the ICL in Table 2. Therefore, small values corresponding to a small number of components are neglected during visual interpretation as it can be assumed that in the atom counting analysis at least a certain number of components will be available in the distribution of estimated peak volumes. This number can be determined based on some prior knowledge about the size of the particle under study. It should be noticed that if the number of components of the underlying distribution is small, no local minima are found for a higher number of components in ICL. Secondly, it has been observed from Fig. 4 that the minimum of interest in the ICL criterion corresponding to the correct number of components becomes more pronounced if the number of observations  $N$  increases. In addition, if the underlying distribution consists of many components, the number of observations  $N$  should be larger than for a distribution with a small number of components to determine the correct order of the mixture model. This conclusion follows from Fig. 4 where a minimum in ICL for  $G=16$  components for  $N=50$  and  $N=100$  is rather unclear whereas a more pronounced minimum for  $G=8$  components is already observed for  $N=50$ . This means that the quality of the ICL increases for an increasing ratio of  $N/G$ . While the number of observations corresponds to the number of atom columns in an experimental image when a single image is recorded, this number can be increased by analysing larger fields of view, if experimental conditions are favourable.

Based on these conclusions, the number of observations, i.e. atom columns, needed to obtain reliable results from the evaluation of ICL can be derived as a function of the number of components in the probability distribution. Therefore, simulations similar to those discussed in Section 4.1 are used. The relative width of the components equals  $\sigma/\delta = 0.25$ . The percentage of correctly determined number of components is calculated for random draws from probability distributions consisting of a certain number of observations and components. This calculation is repeated increasing the number of observations till this percentage equals 95. The results of this analysis are shown in Fig. 5. From this figure, it can be concluded that the necessary number of observations increases almost linearly with the number of components of the underlying Gaussian mixture model. It should be



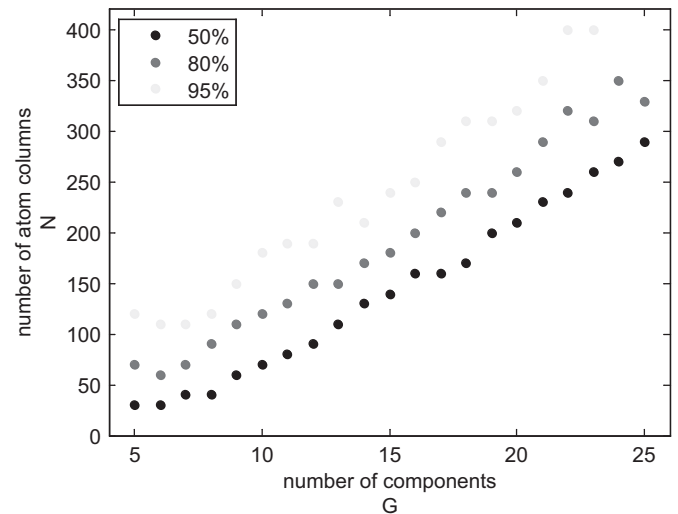
**Fig. 4.** ICL as a function of an increasing number of atom columns  $N$  for  $G=8$  components at the left side and  $G=16$  components at the right side.

mentioned that the results in Fig. 5 are obtained without any visual inspection of ICL. This means that the necessary number of observations needed to obtain reliable results can slightly be reduced when accompanying this analysis with a visual inspection of the ICL criterion.

Since the necessary number of observations increases almost linearly with the number of components, the number of observations per component  $N/G$  can be considered as a measure to extend this analysis for higher relative widths. Then, the reliability of the ICL can be derived more generally as a function of the relative width  $\sigma/\delta$ . The results of these simulations are shown in Fig. 6 for relative widths  $\sigma/\delta$  between 0.2 and 0.4. For these simulations, probability distributions with 10 components were used under the same simulation conditions as those of Section 4.1. The ICL is evaluated from 8 to 12 components. From the figure, it can be concluded that the number of observations per component  $N/G$  should be increased for higher relative widths  $\sigma/\delta$  in order to retain the high reliability of the ICL. A reliability of 50% is still possible for high overlap between neighbouring components.

In general, it can be concluded that the use of the ICL criterion leads to unbiased measurements for the number of components in a mixture model provided that enough observations are available. Since the number of components determines the number of atoms that will be assigned to each column, the analysis also reflects the accuracy of the method to count the number of atoms. In Section 4.3, it will be mentioned that unbiased estimates for the parameters of the probability distribution are obtained which are also required for accurate atom counting results. The precision of the atom counting method will be discussed in Sections 4.3 and 4.4.

It has experimentally been confirmed that the ICL criterion leads to unbiased measurements. For a Au nanorod which has been discussed in [12] where the average  $\sigma/\delta$  equals 0.38 and the average  $N/G$  equals 87, the number of atoms has been counted and the accuracy of the counting result has been tested by means of a comparison with detailed STEM image simulations. For this nanorod

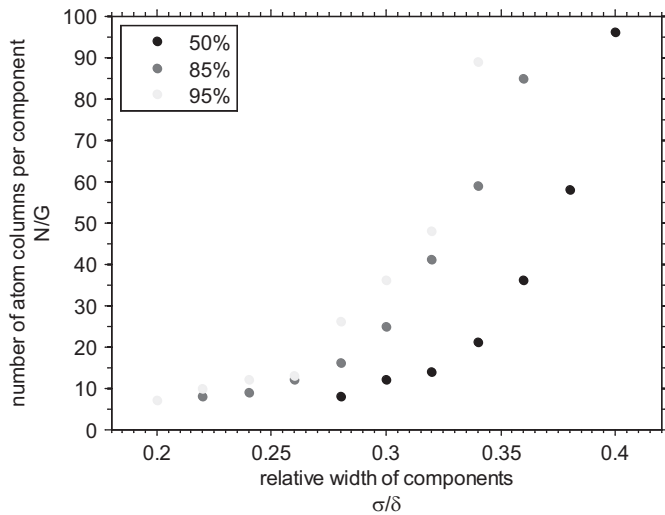


**Fig. 5.** The number of atom columns needed to obtain reliable results from the ICL criterion for a given number of components. Different levels have been chosen to express the reliability in terms of the percentage of correctly assigned number of components. The black, dark grey and light grey dots correspond to 50%, 80% and 95% respectively.

an excellent match of the experimental and simulated intensities within the expected 5–10% error range has been found which validates the accuracy of the counting procedure.

#### 4.3. Precision of the estimated parameters of the Gaussian mixture model

The precision of the number of atoms is mainly determined by the width  $\sigma$  of the Gaussian components in the mixture model, and the precision of the locations  $\mu_g$  of the components in the



**Fig. 6.** The number of atom columns per component needed to obtain reliable results from the ICL criterion for a certain relative width. Different levels have been chosen to express the reliability in terms of the percentage of correctly assigned number of components. The black, dark grey and light grey dots correspond to 50%, 80% and 95% respectively.

Gaussian mixture model. The parameter  $\mu_g$  can be estimated by maximising the likelihood function using the expectation maximisation algorithm discussed in Section 2.2. Statistical parameter estimation theory provides an expression for the highest precision, or in other words, the minimum variance, with which the parameters of the Gaussian mixture model can be determined. This is the so-called Cramér–Rao lower bound (CRLB) [18,30,49], which is a theoretical lower bound on the variance of any unbiased estimator of unknown model parameters. One of the properties of the maximum likelihood estimator is that it achieves the Cramér–Rao lower bound asymptotically, i.e. for an infinite number of observations. Consequently, if the Cramér–Rao lower bound is attained using the expectation maximisation algorithm, the Cramér–Rao lower bound provides a useful measure for the error bar on the estimated parameters of the Gaussian mixture model. In Appendix A, it is summarised how the Cramér–Rao lower bound is computed for the estimated parameters of the Gaussian mixture model.

In this paragraph, it is investigated if the lower bound on the variance of the estimated parameters is attained by means of simulations. Therefore, 100 different noise realisations of a Gaussian mixture model with 5 components are simulated. The values for the true parameters in the simulation are set as follows: the proportions are the same for each component, i.e.  $\pi_g = 0.2$ , and the relative width equals  $\sigma/\delta = 0.25$ . The number of observations  $N$  equals 100. In this analysis unbiased estimates have been obtained for the parameters of the Gaussian mixture model, which is fundamental for obtaining an accurate estimate for the number of atoms. The results for the attainability of the Cramér–Rao lower bound on the variance of the estimated parameters of the mixture model are presented in Table 3. This table shows the computed value for the lower bound on the variance of a particular parameter in the second column. The third column shows the sample variance obtained from the parameter estimates from the 100 noise realisations using the expectation maximisation algorithm; in the last column the corresponding 95% confidence intervals on the sample variances are given. Since the 95% confidence interval of the sample variance encloses the Cramér–Rao lower bound, it can be concluded from Table 3 that the Cramér–Rao lower bound on the variance of the parameters of the Gaussian mixture model is attained using the expectation

**Table 3**

Attainability of the CRLB on the estimated parameters of a Gaussian mixture model with 5 components. The sample variances are computed from parameter estimates obtained from 100 noise realisations of the Gaussian mixture model and compared with the CRLB on the variance of the model parameters.

Parameter	CRLB	Sample variance	95% confidence interval
$\pi_1$	0.0017	0.0022	[0.0017; 0.0030]
$\pi_2$	0.0018	0.0026	[0.0020; 0.0035]
$\pi_3$	0.0018	0.0019	[0.0014; 0.0025]
$\pi_4$	0.0018	0.0017	[0.0013; 0.0029]
$\pi_5$	0.0017	0.0017	[0.0013; 0.0024]
$\mu_1$	0.0038	0.0041	[0.0032; 0.0056]
$\mu_2$	0.0048	0.0048	[0.0037; 0.0065]
$\mu_3$	0.0048	0.0055	[0.0042; 0.0074]
$\mu_4$	0.0048	0.0045	[0.0035; 0.0061]
$\mu_5$	0.0038	0.0031	[0.0024; 0.0041]
$\sigma$	0.00044	0.00043	[0.00033; 0.00057]

maximisation algorithm. This means that the Cramér–Rao lower bound can be used as a measure for the variance on the estimated parameters of the Gaussian mixture model. This variance will be used to determine the precision with which the number of atoms in a particular atom column can be estimated and will be the subject of the following paragraph.

#### 4.4. Single atom sensitivity

The number of atoms in a particular atom column is determined by assigning each estimated total scattered intensity to the component of the estimated probability distribution with the largest probability to generate this scattered intensity. In this way, the precision of the number of atoms is mainly defined by the overlap between neighbouring components in the Gaussian mixture model. This concept has been used in [12] to experimentally illustrate that this statistical atom counting procedure enables us to count atoms with single atom sensitivity. Single atom sensitivity is defined as an error bar of  $\pm 0$  atoms: if there is no overlap between neighbouring components in the Gaussian mixture model single atom sensitivity is attained as in this case an estimated total scattered intensity can unambiguously be assigned to a certain location  $\mu_g$ . However, overlap between neighbouring components in the Gaussian mixture model is unavoidable and is given by a combination of the precision with which the location parameter  $\mu_g$  in the Gaussian mixture model can be determined, the width  $\sigma$  and the proportions  $\pi_g$  of the components in the Gaussian mixture model. Nevertheless, the probability to assign the correct number of atoms, i.e. the fraction of the analysed atom columns measured without error, can be computed.

The precision of the location parameter  $\mu_g$  can be quantified using the Cramér–Rao lower bound on the variance of the estimated parameters of the Gaussian mixture model as discussed in Section 4.3 and will be denoted as  $CRLB_{\mu}$ . The width  $\sigma$  of the components is non-zero due to a combination of experimental detection noise and instabilities. These two parts define an effective width  $\sigma_{\text{eff}}$  for the normal components:

$$\sigma_{\text{eff}} = \sqrt{CRLB_{\mu} + \sigma^2} \quad (25)$$

The amount of overlap between neighbouring components is furthermore influenced by the proportions  $\pi_g$  and is also used to define the probability to assign the correct number of atoms to a specific component. This procedure on how to compute these probabilities is illustrated in Fig. 7. The component shown in colour will be used for this purpose and will be referred to as component  $j$ . This distribution  $j$  will be generated by a set of atom



columns having the same number of atoms in it. Atom columns having scattered intensities belonging to the central dark grey part, interval  $I_j$ , of this component will be assigned to component  $j$ . This central part  $I_j$  is bounded by the intersections of component  $j$  with  $j-1$ , and  $j$  with  $j+1$ . The location of these intersections can be expressed as a function of  $\sigma_{\text{eff}}$ ,  $\pi_g$ , and  $\mu_g$ :

$$S_{j(j+1)} = \frac{2\sigma_{\text{eff}} \ln\left(\frac{\pi_j}{\pi_{j+1}}\right) - \mu_j^2 + \mu_{j+1}^2}{2(\mu_{j+1} - \mu_j)} \quad (26)$$

Using these locations of the intersections which define the interval  $I_j$ , the posterior probability with which an intensity belonging to component  $j$  resides in  $I_j$  can be computed as follows:

$$p(I_j|C_j) = F(S_{j(j+1)}, \mu_j, \sigma_{\text{eff}}) - F(S_{j(j-1)}, \mu_j, \sigma_{\text{eff}}) \quad (27)$$

where  $F(S, \mu_j, \sigma_{\text{eff}})$  corresponds to the cumulative distribution function of the normal distribution at location  $\mu_j$  having a width  $\sigma_{\text{eff}}$ . The posterior probability that an intensity belonging to the central dark grey part stems from component  $j$  is then given by Bayes' rule:

$$p(C_j|I_j) = p(I_j|C_j) \cdot \frac{p(C_j)}{p(I_j)} \quad (28)$$

where  $p(C_j)$  equals the mixing proportion  $\pi_j$  of the  $j$ th component and  $p(I_j)$  corresponds to the area below the Gaussian mixture model between  $S_{j(j-1)}$  and  $S_{j(j+1)}$ , hatched in Fig. 7:

$$p(I_j) = \int_{S_{j(j-1)}}^{S_{j(j+1)}} f_{\text{mix}}(V; \Psi) dV \quad (29)$$

Formula (28) expresses the probability that a total scattered intensity of interval  $I_j$  originates from the correct component  $j$  and thus identifies the percentage of atom columns that has been determined with single atom sensitivity. For the whole structure of interest, the probability to count the number of atoms in a column without error can then be obtained by averaging the probabilities of Eq. (28) for all the components in the Gaussian mixture model.

Similarly, the probability to miscount the number of atoms with  $\pm 1$  can be computed. This probability is related to the outermost light grey parts in Fig. 7. Atom column intensities whose intensities belong to these parts will be assigned to component  $j-1$  or  $j+1$ . In practice, the probability to miscount the number of atoms with  $\pm 2$  will generally equal zero, but can be computed in the same way taking the overlap of distribution  $j$  with second neighbouring components into account.

If equal proportions are assumed for the components in the Gaussian mixture model, the intersection between neighbouring components is located in the middle between the locations  $\mu_g$  of the considered components. The probability of counting the number of atoms correctly, i.e. with an error of  $\pm 0$  atoms, is then

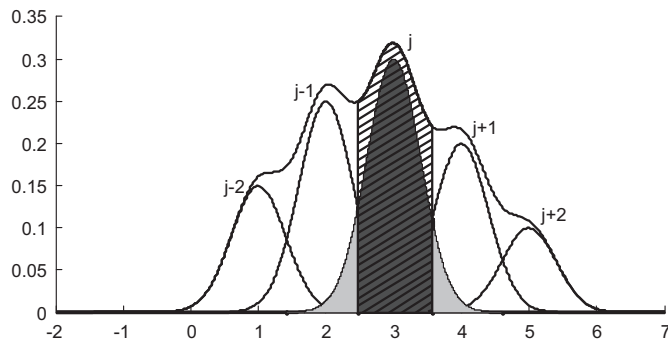


Fig. 7. Overlap between neighbouring components in the Gaussian mixture model to illustrate the concept of single atom sensitivity.

defined only by  $\sigma$  and  $CRLB_\mu$ . In order to illustrate how this probability relates to the latter parameters, tolerance levels for single atom sensitivity are drawn on a plot of  $CRLB_\mu$  as a function of the relative width  $\sigma/\delta$  in Fig. 8. For this figure, the Cramér–Rao lower bound on the variance of the location parameter  $\mu_g$  has been calculated for different values of the parameters  $\pi_g$ ,  $\mu_g$ , and  $\sigma$  of the Gaussian mixture model based on the formulae in Appendix A. In addition, the number of components  $G$  and the number of observations  $N$  have been varied. Based on the computed values, it was observed that the value of  $CRLB_\mu$  is determined by the relative width  $\sigma/\delta$  of the components and the number of observations per component  $N/G$ . Therefore,  $CRLB_\mu$  has been calculated for the central component of a probability distribution with equal proportions for different values of  $\sigma/\delta$  and  $N/G$ . The values of  $CRLB_\mu$  increase for a decreasing ratio of  $N/G$  and an increasing relative width  $\sigma/\delta$ . Based on formula (27), lines are drawn in Fig. 8 for the 68% and 95% tolerance levels. These lines show that the precision with which the number of atoms can be counted in a particular atom column is limited by the relative width of the components. Indeed, if the relative width goes beyond a certain value, then neighbouring components will strongly overlap. Fig. 8 can be used to give an indication of the probability to count the correct number of atoms for a specific case.

The precision with which the number of atoms is determined in the example of the Au wedge described in Section 3 can be derived using formula (27). For this example, it can be concluded that single atom sensitivity is attained for 99.9% of the analysed atom columns. The fact that this probability is larger than 95% can also be derived from Fig. 8, since the number of observations per component  $N/G$  equals 16 and the average relative width  $\sigma/\delta$  equals 0.15.

### 5. Discussion and conclusions

In the present paper, the method introduced in [13] to count the number of atoms in a nanostructure containing one atom type from HAADF STEM images has been explained in detail together with a thorough study on the possibilities and inherent limitations imposed by unavoidable statistical uncertainties. The method largely proceeds in two steps. First, an empirical incoherent imaging model is fitted to the experimental HAADF STEM images

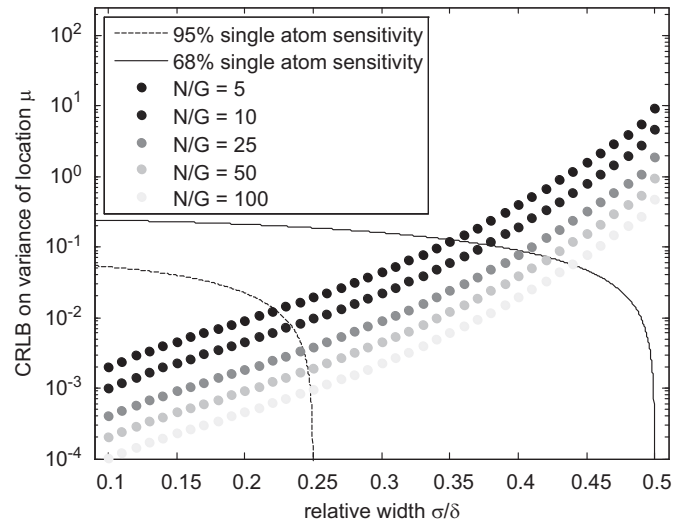


Fig. 8. Precision on the location  $\mu$  for a given relative width  $\sigma/\delta$  and number of observations per component  $N/G$ . The dashed line indicates the 95% tolerance level for single atom sensitivity; the full line indicates the 68% tolerance level for single atom sensitivity.

to extract quantitative values for the total scattered intensity per atom column. These values can be used to determine the number of atoms per atom column, if it is assumed that the total scattered intensity per atom column scales with the number of atoms. In the second step, the number of atoms per atom column is determined by fitting a Gaussian finite mixture model to the probability distribution of the estimated total scattered intensities using the expectation maximisation algorithm. This probability distribution consists of components of sets of atom columns having the same number of atoms. The number of components in the probability distribution is determined by the use of an order selection criterion.

By means of simulations, the possibilities and inherent limitations of this atom counting method have been investigated. The performance of different order selection criteria has been investigated. It turns out that the integrated classification likelihood criterion (ICL) outperforms other selection criteria when one is evaluating probability distributions for counting the number of atoms in a nanostructure. In addition, it has been shown that different aspects need to be taken into account for a correct interpretation of the ICL criterion. The number of observations, the number of components in the probability distribution, the width of the components, and the typical shape of the criterion directly affect the accuracy with which the number of atoms can be determined. Also the precision with which the number of atoms in a particular atom column can be estimated has been studied. It has been shown that the variance of the expectation maximisation estimates for the parameters of the Gaussian mixture model attains the Cramér–Rao lower bound, i.e. the highest precision or in other words the minimum variance with which the parameters of the model can be determined. Moreover, it has been shown that single atom sensitivity is feasible.

In conclusion, the presented statistical model-based approach to count the number of atoms can be used without prior knowledge of the atomic structure and is independent from image simulations. This powerful method enables one to identify the percentage of atom columns that has been determined with single atom sensitivity.

## Acknowledgements

The authors gratefully acknowledge financial support from the Research Foundation Flanders (FWO, Belgium) through project fundings (G.0393.11, G.0064.10 and G.0374.13) and a PhD research grant to A.D.B. A.R. thanks the DFG under contract number RO2057/8-1. The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007–2013] under grant agreement no 312483 (ESTEEM2).

## Appendix A. Cramér–Rao lower bound for the parameters of the Gaussian mixture model: Fisher information and attainable precision

The concept of Fisher information and attainable precision is briefly discussed in this appendix; the details of these concepts are described in [30]. The usual way to describe the fluctuating behaviour of data in the presence of noise is by modelling the observations as stochastic variables. By definition, each set of observations modelled as stochastic variables is characterised by a joint probability density function. The joint probability density function can usually be written as a function of the quantities to be estimated, i.e. the unknown parameters. For the Gaussian mixture model, the observations, i.e. the estimated scattered intensities, are assumed to be statistically independent having the following

joint probability density function:

$$p(\mathbf{V}, \Psi) = \prod_{n=1}^N f_{\text{mix}}(V_n, \Psi) \quad (\text{A.1})$$

where  $f_{\text{mix}}(V_n; \Psi)$  is the probability density function of a mixture model given by Eq. (6). The dependence of the joint probability density function on the unknown parameters  $\Psi$  defines the Fisher information matrix and the Cramér–Rao lower bound. The concept of Fisher information is a measure to quantitatively express the amount of information on the unknown parameters included in the observations. This Fisher information matrix is defined as

$$F_{\Psi} = -\mathbb{E} \left[ \frac{\partial^2 \ln p(\mathbf{V}, \Psi)}{\partial \Psi \partial \Psi^T} \right] \quad (\text{A.2})$$

In practice, the following integral is numerically evaluated:

$$F_{\Psi} = N \int_{-\infty}^{\infty} \frac{\partial^2 \ln f_{\text{mix}}(V; \Psi)}{\partial \Psi^2} f_{\text{mix}}(V; \Psi) dV \quad (\text{A.3})$$

Using the concept of Fisher information, it is possible to determine the attainable precision, i.e. the lowest variance with which a parameter can be estimated unbiasedly. An expression can be derived for the lower bound on the variance and this lower bound is called the Cramér–Rao lower bound. This lower bound is independent of the estimation method, but depends on the statistical properties of the observations and mostly on the hypothetical true values of the parameters. The Cramér–Rao inequality, with  $\hat{\Psi}$  an unbiased estimator of  $\Psi$ , states

$$\text{cov}(\hat{\Psi}, \hat{\Psi}) \geq F_{\Psi}^{-1} \quad (\text{A.4})$$

The matrix  $F_{\Psi}^{-1}$  is the Cramér–Rao lower bound on the variance of  $\hat{\Psi}$ .

This inequality expresses that the difference between the covariance matrix of any unbiased estimator and the Cramér–Rao lower bound is positive semi-definite. A property of a positive semi-definite matrix is that its diagonal elements cannot be negative. This means that the diagonal elements of  $\text{cov}(\hat{\Psi}, \hat{\Psi})$  will always be larger than or equal to the corresponding diagonal elements of the inverse of the Fisher information matrix. Therefore, the diagonal elements of  $F_{\Psi}^{-1}$  define lower bounds on the variances of the elements of  $\hat{\Psi}$ :

$$\text{var}(\hat{\Psi}_R) \geq [F_{\Psi}^{-1}]_{RR} \quad (\text{A.5})$$

where  $\hat{\Psi}_R$  is the  $R$ th element of  $\hat{\Psi}$ , and  $[F_{\Psi}^{-1}]_{RR}$  is the  $(R,R)$ th element of the inverse of the Fisher information matrix.

## References

- [1] S.J. Pennycook, L.A. Boatner, Chemically sensitive structure-imaging with a scanning transmission electron microscope, *Nature* 336 (1988) 565–567.
- [2] P. Hartel, D. Rose, C. Dinges, Conditions and reasons for incoherent imaging in STEM, *Ultramicroscopy* 63 (1996) 63–114.
- [3] A. Singhal, J.C. Yang, J.M. Gibson, STEM-based mass spectroscopy of supported Re clusters, *Ultramicroscopy* 67 (1997) 191–206.
- [4] P.M. Voyles, D.A. Muller, J.L. Grazul, P.H. Citrin, H.-J.L. Gossmann, Atomic-scale imaging of individual dopant atoms and clusters in highly n-type bulk Si, *Nature* 416 (2002) 826–829.
- [5] R. Erni, H. Heinrich, G. Kostorz, Quantitative characterisation of chemical inhomogeneities in Al–Ag using high-resolution Z-contrast STEM, *Ultramicroscopy* 94 (2003) 125–133.
- [6] J.M. LeBeau, S.D. Findlay, L.J. Allen, S. Stemmer, Standardless atom counting in scanning transmission electron microscopy, *Nanoletters* 10 (2010) 4405–4408.
- [7] J.M. LeBeau, S.D. Findlay, L.J. Allen, S. Stemmer, Quantitative atomic resolution scanning transmission electron microscopy, *Physical Review Letters* 100 (2008) 206101.
- [8] J.M. LeBeau, S. Stemmer, Experimental quantification of annular dark-field images in scanning transmission electron microscopy, *Ultramicroscopy* 108 (2008) 1653–1658.
- [9] J.M. LeBeau, A.J. D’Alfonso, S.D. Findlay, S. Stemmer, L.J. Allen, Quantitative comparisons of contrast in experimental and simulated bright-field scanning transmission electron microscopy images, *Physical Review B* 80 (2009) 174106.

- [10] J.M. LeBeau, S.D. Findlay, X. Wang, A.J. Jacobson, L.J. Allen, S. Stemmer, High-angle scattering of fast electrons from crystals containing heavy elements: simulation and experiment, *Physical Review B* 79 (2009) 214110.
- [11] A. Rosenauer, K. Gries, K. Müller, A. Pretorius, M. Schowalter, A. Avramescu, K. Engl, S. Lutgen, Measurement of specimen thickness and composition in  $\text{Al}_x\text{Ga}_{1-x}\text{N}/\text{GaN}$  using high-angle annular dark field images, *Ultramicroscopy* 109 (2009) 1171–1182.
- [12] S. Van Aert, A. De Backer, G.T. Martinez, B. Goris, S. Bals, G. Van Tendeloo, A. Rosenauer, Procedure to count atoms with trustworthy single-atom sensitivity, *Physical Review B* 87 (2013) 064107.
- [13] S. Van Aert, K.J. Batenburg, M.D. Rossell, R. Erni, G. Van Tendeloo, Three-dimensional atomic imaging of crystalline nanoparticles, *Nature* 470 (2011) 374–377.
- [14] S. Bals, M. Casavola, M.A. van Huis, S. Van Aert, K.J. Batenburg, G. Van Tendeloo, D. Vanmaekelbergh, Three-dimensional atomic imaging of colloidal core-shell nanocrystals, *Nano Letters* 11 (8) (2011) 3420–3424.
- [15] S. Bals, S. Van Aert, C.P. Romero, K. Lauwaet, M.J. Van Bael, B. Schoeters, A. Partoens, E. Yücelen, P. Lievens, G. Van Tendeloo, Atomic scale dynamics of ultrasmall germanium clusters, *Nature Communications* 3 (2012) 897.
- [16] K.J. Batenburg, A network flow algorithm for reconstructing binary images from discrete X-rays, *Journal of Mathematical Imaging and Vision* 27 (2007) 175–191.
- [17] J.R. Jinschek, K.J. Batenburg, H.A. Calderon, R. Kilaas, V. Radmilovic, C. Kisielowski, 3-D reconstruction of the atomic positions in a simulated gold nanocrystal based on discrete tomography: prospects of atomic resolution electron tomography, *Ultramicroscopy* 108 (2008) 589–604.
- [18] A.J. den Dekker, S. Van Aert, A. van den Bos, D. Van Dyck, Maximum likelihood estimation of structure parameters from high resolution electron microscopy images. Part I: a theoretical framework, *Ultramicroscopy* 104 (2005) 83–106.
- [19] A.J. den Dekker, S. Van Aert, D. Van Dyck, A. van den Bos, P. Geuens, Does a monochromator improve the precision in quantitative HRTEM? *Ultramicroscopy* 89 (2001) 275–290.
- [20] S. Van Aert, A.J. den Dekker, D. Van Dyck, A. van den Bos, High-resolution electron microscopy and electron tomography: resolution versus precision, *Journal of Structural Biology* 138 (2002) 21–33.
- [21] S. Van Aert, A.J. den Dekker, A. van den Bos, D. Van Dyck, J.H. Chen, Maximum likelihood estimation of structure parameters from high resolution electron microscopy images: part II: a practical example, *Ultramicroscopy* 104 (2005) 107–125.
- [22] S. Van Aert, L.Y. Chang, S. Bals, A.I. Kirkland, G. Van Tendeloo, Effect of amorphous layers on the interpretation of restored exit waves, *Ultramicroscopy* 109 (2009) 237–246.
- [23] S. Bals, S. Van Aert, G. Van Tendeloo, D. Ávila-Brandé, Statistical estimation of atomic positions from exit wave reconstruction with a precision in the picometer range, *Physical Review Letters* 96 (2006) 096106.
- [24] A. De Backer, S. Van Aert, D. Van Dyck, High precision measurements of atom column positions using model-based exit wave reconstruction, *Ultramicroscopy* 111 (2011) 1475–1482.
- [25] M. Klingstedt, M. Sundberg, L. Eriksson, S. Haigh, A. Kirkland, D. Grüner, A. De Backer, S. Van Aert, O. Terasaki, Exit wave reconstruction from focal series of HRTEM images, single crystal XRD and total energy studies on  $\text{Sb}_x\text{WO}_{3+y}$  ( $x \sim 0.11$ ), *Zeitschrift für Kristallographie* 227 (2012) 341–349.
- [26] J. Verbeeck, S. Van Aert, Model based quantification of EELS spectra, *Ultramicroscopy* 101 (2004) 207–224.
- [27] J. Verbeeck, S. Van Aert, G. Bertoni, Model-based quantification of EELS spectra: including the fine structure, *Ultramicroscopy* 106 (2006) 976–980.
- [28] M. Huijben, R.G., D.H.A. Blank, S. Bals, S. Van Aert, J. Verbeeck, G. Van Tendeloo, A. Brinkman, H. Hilgenschamp, Electronically coupled complementary interfaces between perovskite band insulators, *Nature Materials* 5 (2006) 556–560.
- [29] S. Van Aert, J. Verbeeck, R. Erni, S. Bals, M. Luysberg, D. Van Dyck, G. Van Tendeloo, Quantitative atomic resolution mapping using high-angle annular dark field scanning transmission electron microscopy, *Ultramicroscopy* 109 (2009) 1236–1244.
- [30] A. van den Bos, *Parameter Estimation for Scientists and Engineers*, Wiley, 2007.
- [31] S.J. Pennycook, D.E. Jesson, High-resolution Z-contrast imaging of crystals, *Ultramicroscopy* 37 (1991) 14–38.
- [32] J. Broeckx, M. Op de Beeck, D. Van Dyck, A useful approximation of the exit wave function in coherent STEM, *Ultramicroscopy* 60 (1995) 71–80.
- [33] P.D. Nellist, Scanning transmission electron microscopy, in: P.W. Hawkes, J.H. C. Spence (Eds.), *Science of Microscopy*, vol. 1, Springer, New York, 2007, pp. 65–132. (Chapter 2).
- [34] G. McLachlan, D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics, John Wiley and Sons, Inc, 2000.
- [35] V. Hasselblad, Estimation of finite mixture of distributions from the exponential family, *Journal of the American Statistical Association* 64 (1969) 1459–1471.
- [36] N.E. Day, Estimating the components of a mixture of normal distributions, *Biometrika* 56 (1969) 463–474.
- [37] J.H. Wolfe, Pattern clustering by multivariate mixture analysis, *Multivariate Behavioral Research* 5 (1970) 329–350.
- [38] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B* 39 (1977) 1–38. (with discussion).
- [39] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19 (6) (1974) 716–723.
- [40] P.M.T. Broersen, H.E. Wensink, On finite sample theory for autoregressive model order selection, *IEEE Transactions on Signal Processing* 41 (1) (1993) 194–204.
- [41] P.M.T. Broersen, H.E. Wensink, On the penalty factor for autoregressive order selection in finite samples, *IEEE Transactions on Signal Processing* 44 (1996) 748–752.
- [42] G. Schwarz, Estimating the dimension of a model, *Annals of Statistics* 6 (2) (1978) 461–464.
- [43] C. Biernacki, G. Govaert, Using the classification likelihood to choose the number of clusters, *Computing Science and Statistics* 29 (1997) 451–457.
- [44] J.D. Banfield, A.E. Raftery, Model-based Gaussian and non-Gaussian clustering, *Biometrics* 49 (1993) 803.
- [45] G. Celeux, G. Soromenho, An entropy criterion for assessing the number of clusters in a mixture model, *Journal of Classification* 13 (2) (1996) 195–212.
- [46] C. Biernacki, G. Celeux, G. Govaert, An improvement of the NEC criterion for assessing the number of clusters in a mixture model, *Pattern Recognition Letters* 20 (3) (1999) 267–272.
- [47] C. Biernacki, G. Celeux, G. Govaert, Assessing a mixture model for clustering with the integrated classification likelihood, Technical Report 3521, INRIA, Rhône-Alpes, 1998.
- [48] A. Rosenauer, M. Schowalter, STEMSIM—a new software tool for simulation of STEM HAADF Z-contrast imaging, in: A.G. Cullis, P.A. Midgley (Eds.), *Springer Proceedings in Physics: Microscopy of Semiconducting Materials 2007*, Springer Proceedings in Physics, vol. 120, Springer, Netherlands, 2008, pp. 169–172.
- [49] S. Van Aert, A.J. den Dekker, D. Van Dyck, A. van den Bos, Optimal experimental design of STEM measurement of atom column positions, *Ultramicroscopy* 90 (2002) 273–289.