Hidden Markov model for atom-counting from sequential ADF STEM images : methodology, possibilities and limitations

Hidden Markov model for atom-counting from sequential ADF STEM images: methodology, possibilities and limitations

Annelies De wael[a,b], Annick De Backer[a,b], Sandra Van Aert[a,b,*]

[a]*EMAT, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium*
[b]*NANOlab Center of Excellence, University of Antwerp, Belgium*

Abstract

We present a quantitative method which allows us to reliably measure dynamic changes in the atomic structure of monatomic crystalline nanomaterials from a time series of atomic resolution annular dark field scanning transmission electron microscopy images. The approach is based on the so-called hidden Markov model and estimates the number of atoms in each atomic column of the nanomaterial in each frame of the time series. We discuss the origin of the improved performance for time series atom-counting as compared to the current state-of-the-art atom-counting procedures, and show that the so-called transition probabilities that describe the probability for an atomic column to lose or gain one or more atoms from frame to frame are particularly important. Using these transition probabilities, we show that the method can also be used to estimate the probability and cross section related to structural changes. Furthermore, we explore the possibilities for applying the method to time series recorded under variable environmental conditions. The method is shown to be promising for a reliable quantitative analysis of dynamic processes such as surface diffusion, adatom dynamics, beam effects, or *in situ* experiments.

Keywords: atom-counting, dynamic structural changes, quantitative electron microscopy, scanning transmission electron microscopy

1. Introduction

Nanomaterials exhibit unique properties depending on their size and shape, as well as on the type of atoms [1, 2, 3]. Particularly interesting are the changes in the nanomaterials through time, which can occur via adatom dynamics [4, 5], surface diffusion and reconstruction [6, 7, 8, 9, 10], due to electron beam irradiation [11, 12, 13, 14], or during *in situ* experiments such as heating or gas flow experiments [15, 16]. In order to study the properties of such time varying processes, it is important to correctly quantify the 3D atomic structure of the nanomaterials, and the structural changes that occur over time. In order to unravel the structure-properties relationship of nanomaterials, atomic resolution annular dark field (ADF) scanning transmission electron microscopy (STEM) is an ideal technique [17, 18]. When the inner angle of the annular detector is chosen sufficiently high - about three times larger than the probe convergence angle, the ADF STEM imaging is incoherent. In such high angle annular dark field (HAADF) STEM images, the intensity increases monotonically with the atomic number Z and the thickness of the material [18, 19, 20, 21]. The HAADF STEM is therefore very suitable for quantitative analysis. However, depending on the specimen thickness, quantitative analysis of ADF STEM image intensities can also be done using smaller detector inner angles, balancing the detector collection area against the coherence of the image intensity in order to obtain the optimal experiment design [22]. For a single atomic resolution ADF STEM image, the 3D atomic structure of the nanomaterial can be retrieved using 2D atom-counting results. An initial 3D configuration is therefore generated starting from the counting results, using prior knowledge about the material's crystal structure. Next, an energy minimization using ab-initio calculations or a Monte Carlo approach is performed to relax the 3D atomic structure [23, 24, 25, 26].

Different methods have been developed that can reliably count the number of atoms in each atomic column from a single ADF STEM image [23, 27, 28, 29, 30, 31]. These approaches are all based on the so-called scattering cross sections, a measure for the total intensity of electrons scattered from the atomic columns [32, 33]. The scattering cross section of an atomic column increases with increasing atomic mass number Z or increasing thickness of the atomic column, and can be estimated from an ADF STEM image by integrating image intensities in Voronoi cells around the atomic column positions [33], or by fitting a parametric model consisting of 2D overlapping Gaussian peaks [29, 32, 34]. Once the scattering cross sections for all atomic columns are estimated, atoms can be counted by using an image simulations-based approach [23, 27], a statistics-based approach [28, 29, 30], or a hybrid approach, combining both techniques [31]. In case of the im-

*sandra.vanaert@uantwerpen.be

age simulations-based approach, normalised experimental scattering cross sections from the ADF STEM image are directly compared to scattering cross sections obtained from simulated images with the same microscope settings of the same material in a bulk crystal at different thicknesses. The statistics-based approach on the other hand estimates a Gaussian mixture model based on the set of scattering cross sections, where each Gaussian component corresponds to a set of atomic columns with a given thickness. Then, for each atomic column the most likely number of atoms is assigned. The parameters of the Gaussian mixture model even provide a measure for the precision of the counting results. Finally, the hybrid statistics-simulations based method for atom-counting includes prior knowledge from image simulations to estimate the locations of the components of the Gaussian mixture model. This approach allows for more reliable atom-counting, particularly at low electron doses.

In this paper, we propose a method for atom-counting, specifically designed for the analysis of sequential ADF STEM images, using a so-called hidden Markov model. The method was introduced in [35] to quantitatively analyse possible changes of the atomic structure from time series of ADF STEM images. Here, its methodology, possibilities and limitations will be discussed in more detail. In Section 2, the theoretical framework of the hidden Markov model is introduced and an explicit link between this new method and the existing procedure for atom-counting from a single image is made [28, 29, 30, 31]. In this section, the so-called transition probabilities are introduced as key parameters of the hidden Markov model, as they quantify the probabilities for an atomic column to lose or gain one or more atoms from frame to frame during the time series. Next, in Section 3, we discuss the possibilities and inherent limitations of this method for atom-counting from time series of ADF STEM images. It will be shown that the transition probabilities estimated from this analysis can be linked to physical cross sections related to structural changes. Furthermore, we discuss how to deal with variable environmental conditions during the time series.

2. Time series atom-counting methodology

In this section, we discuss in detail how to count the number of atoms in each atomic column of a nanostructure from a time series of ADF STEM images [Figure 1(a)]. First, the ADF STEM image intensities need to be normalised to the incident electron beam [36, 37, 38, 39]. Then, in order to reliably quantify the changes for each atomic column of the nanostructure, the time series needs to be aligned. This is achieved using rigid registration to account for sample drift and/or rotation [Figure 1(b)]. One can also apply non-rigid registration of the images to compensate for scan distortions in order to improve the precision of the quantification of the ADF STEM images performed during the next steps [40]. Next, the atomic column positions need to be reliably quantified in each frame. In case of high electron dose recordings, a simple peak finding routine is sufficient. However, when the electron dose is low, and the noise level in the images is high, it becomes difficult to distinguish atomic columns from noise fluctuations. It is however crucial to correctly select all atomic columns. This can be done using the maximum *a posteriori* (MAP) probability rule [Figure 1(c)] [41, 42]. Using these atomic column positions, a parametric imaging model is fitted to each ADF STEM image using the StatSTEM software [34]. From this model, which takes overlapping intensities of neighbouring atomic columns into account, scattering cross sections can be estimated, corresponding to the total scattered intensity for each atomic column [Figure 1(d)]. Although crosstalk is not estimated by this incoherent imaging model, it was shown that this effect does not hamper reliable quantification [43, 44]. In order to apply the hidden Markov model to the ADF STEM time series, the atomic columns of the nanostructure are tracked in each image, and the scattering cross sections of each atomic column are sorted accordingly [Figure 1(e)]. Together, the scattering cross sections of all atomic columns in each ADF STEM image of the time series are the input for the hidden Markov model analysis [Figure 1(f)]. The output of the hidden Markov model is the set of atom-counting results for each ADF STEM image of the time series [Figure 1(g)].

2.1. Scattering cross sections for time series analysis

In order to analyse the ADF STEM time series, the scattering cross sections of the atomic columns in each image are estimated [Figure 1(d)] [34]. The ADF STEM image intensities are peaked at the atomic column positions [18]. Therefore, a parametric model consisting of overlapping 2D Gaussian peaks is fitted to the image. The expectation of the image intensity at pixel (k,l), corresponding to position (x_k, y_l) in the ADF STEM image, can be described by the expectation model $f_{kl}(\theta)$:

$$f_{kl}(\theta) = \zeta + \sum_{n=1}^{N} \eta_n \exp\left(-\frac{\left(x_k-\beta_{x_k}\right)^2+\left(y_l-\beta_{y_l}\right)^2}{2\rho^2}\right), \quad (1)$$

with θ the vector of unknown structure parameters: a constant background ζ, the width of the Gaussian peaks ρ, the height of each Gaussian peak η_n, and the position coordinates of each atomic column β_{x_n} and β_{y_n}. The total number of atomic columns analysed using this model corresponds to N. The volumes under the 2D Gaussian peaks quantify the scattering cross sections of each atomic column n in the image:

$$o^{(n)} = 2\pi\eta_n\rho^2. \quad (2)$$

This procedure is applied to each individual ADF STEM image of the time series. Each atomic column present in at least one of the images of the time series therefore receives a unique index n. The scattering cross sections are ranked accordingly [Figure 1(e)]. When an atomic column n is absent in a frame t according to the MAP rule [41, 42], it is not included in the fitting procedure for that image, but

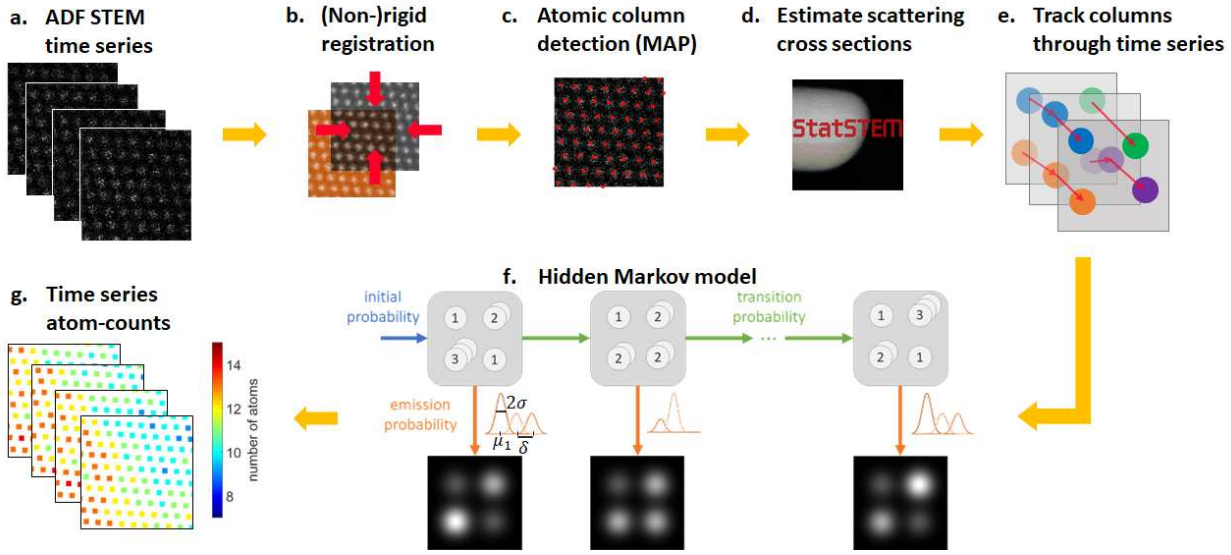

Figure 1: Workflow for atom-counting from a time series of ADF STEM images using hidden Markov models.

its scattering cross section $o_t^{(n)}$ is simply set to zero. By sorting the set of scattering cross sections in this manner, we can construct the so-called observed sequence:

$$\mathbf{O} = \begin{pmatrix} o_1^{(1)} & o_1^{(2)} & \cdots & o_1^{(N)} \\ o_2^{(1)} & o_2^{(2)} & \cdots & o_2^{(N)} \\ \cdots & & & \\ o_T^{(1)} & o_T^{(2)} & \cdots & o_T^{(N)} \end{pmatrix}. \tag{3}$$

This corresponds to the set of scattering cross sections of an ADF STEM time series with N unique atomic columns and T frames, and it serves as an input for the hidden Markov model that will be used to count the number of atoms.

2.2. Probability distribution of estimated scattering cross sections

A hidden Markov model consists of two layers [Figure 1(f)]. The first layer is "hidden", and is observed only indirectly through the observed sequence that constitutes the second layer of the hidden Markov model. The first layer is a first order Markov chain. This implies that the state at a given time t depends only on the previous state, at time $t-1$, and not on the state of the system before that. In the context of atom-counting, the hidden states of the first layer correspond to the number of atoms in each atomic column at each image of the time series. The number of atoms in each atomic column of the nanoparticle in each frame can only be observed indirectly through the scattering cross sections. Therefore, in the context of atom-counting, the set of all scattering cross sections estimated from the ADF STEM images of the time series $\mathbf{O} = \{o_t^{(n)}\}$, introduced in Equation (3), constitutes the second layer. The set of all hidden states is summarised by the hidden state sequence $\mathbf{H} = \{\mathbf{h}_t^{(n)}\}$. In this expression, $\mathbf{h}_t^{(n)}$ is a so-called latent variable that describes the hidden state (number of atoms) at time t for atomic column n. In the context of hidden Markov models, this is commonly chosen as a binary vector, with binary elements $h_{tg}^{(n)} = 1$ if and only if the number of atoms in atomic column n in frame t equals g, otherwise $h_{tg}^{(n)} = 0$.

The joint probability density function of the hidden state sequence $\mathbf{H}$ and the observed sequence $\mathbf{O}$ is described as:

$$p(\mathbf{O}, \mathbf{H}|\mathbf{\Omega}) = \prod_{n=1}^{N}\prod_{g=0}^{G} (\iota_g)^{h_{1g}^{(n)}} \left(\mathcal{N}(o_1^{(n)}|a\mathcal{M}_g, \sigma)\right)^{h_{1g}^{(n)}} \tag{4a}$$

$$\times \prod_{t=2}^{T}\prod_{n=1}^{N}\prod_{j=0}^{G}\prod_{g=0}^{G} (A_{jg})^{h_{t-1,j}^{(n)} h_{tg}^{(n)}} \tag{4b}$$

$$\times \prod_{t=2}^{T}\prod_{g=0}^{G}\prod_{n=1}^{N} \left(\mathcal{N}(o_t^{(n)}|a\mathcal{M}_g, \sigma)\right)^{h_{tg}^{(n)}}. \tag{4c}$$

This expresses the likelihood that a specific state sequence $\mathbf{H}$ corresponds to the observed sequence $\mathbf{O}$, given a set of model parameters $\mathbf{\Omega}$. In this expression, G is the maximum number of atoms in any atomic column of the time series and T is the number of frames in the time series. Note that the products over the number of atoms start from 0, in order to allow atomic columns to be absent in some of the frames. The unknown model parameters are given by the parameter vector:

$$\mathbf{\Omega} = (\iota_0, \iota_1, \cdots, \iota_{G-1}, A_{00}, \cdots, A_{0,G-1}, A_{10}, \cdots A_{G,G-1}, a, \sigma). \tag{5}$$

The notation used for the hidden Markov model throughout this paper is summarised in Table A.1 of Appendix A. The initial probability ι_g expresses the probability for an atomic column to contain g atoms in the first frame [blue arrow in Figure 1(f)]. Only G initial probabilities

are estimated, since $\sum_{g=0}^{G} \iota_g = 1$. The transition probability A_{jg} expresses the probability that an atomic column has j atoms in one frame and g atoms in the next frame [green arrows in Figure 1(f)]. All transition probabilities, for $0 \leq j \leq G$ and $0 \leq g \leq G$ are summarised in the $(G+1) \times (G+1)$ transition matrix $\mathbf{A} = \{A_{jg}\}$. Only $(G+1)G$ transition probabilities are estimated, since $\sum_{g=0}^{G} A_{jg} = 1$, $\forall 0 \leq j \leq G$. Note that we implicitly assume that atomic columns with the same thickness will undergo structural changes in a similar manner by modelling all structural changes using only one transition matrix. The probability to observe a scattering cross section with value $o_t^{(n)}$ when g atoms are in the atomic column, is modelled by a Gaussian emission probability $\mathcal{N}\left(o_t^{(n)}|\mu_g, \sigma\right)$ [orange arrows in Figure 1(f)]. In this expression, the average scattering cross section for an atomic column with g atoms is determined by $\mu_g = a\mathcal{M}_g$, with $\mathcal{M}_g$ the library value, i.e. the scattering cross section determined from image simulations, for an atomic column with g atoms, as in [31]. In this manner, we include prior knowledge from image simulations, but allow for small deviations between the parameters used for the image simulations and the actual experimental imaging conditions by estimating the linear scaling parameter a. The width of the Gaussian emission probability is σ, as indicated schematically in Figure 1(f). The average distance between subsequent average scattering cross sections μ_g and μ_{g+1}, with $0 \leq g < G$, is called δ, and is also indicated in Figure 1(f). This allows us to express the relative width of the Gaussian emission probability σ/δ, which is a measure for the amount of overlap between neighbouring emission probabilities, and therefore also a measure for the noise level of the ADF STEM images from which the scattering cross sections are estimated. Equation (4) expresses the joint probability density function of a factorial hidden Markov model [45], since the number of atoms in each atomic column is described as a separate hidden state. In a conventional hidden Markov model, the hidden states would be defined as the set of atom-counts at each time corresponding to a specific 3D atomic model, which makes unravelling the physical meaning of each hidden state unnecessarily complex. In addition, also the observations are factorised in Equation (4), as each atomic column has a separate scattering cross section, which makes it possible to derive analytical update formulas for the parameters, contrary to most factorial hidden Markov models, where such an approach becomes intractable [45]. The parameters of the factorial hidden Markov model for atom-counting are estimated using an Expectation-Maximisation algorithm, often called a Baum-Welch algorithm [46]. The counts are retrieved using a Viterbi decoding algorithm that determines the most likely state sequence $\mathbf{H}$ [47, 48]. This state sequence contains the evolution of the number of atoms in each of the atomic columns in the nanomaterial during the time series [Figure 1(g)].

When only 1 frame is considered ($T = 1$), the joint probability density function reduces to Equation (4a). This is the joint probability density function of a Gaussian mixture model with latent variables $h_{1g}^{(n)}$, observed variables $o_1^{(n)}$, and mixing proportions ι_g [49]. In order to understand this, we can derive the marginal probability distribution for the observed sequence $\mathbf{O}$ when $T = 1$, given the model parameters $\mathbf{\Omega}$:

$$\begin{aligned} p(\mathbf{O}|\mathbf{\Omega}) &= \sum_{\mathbf{H}} p(\mathbf{O}, \mathbf{H}|\mathbf{\Omega}) \\ &= \sum_{\mathbf{H}} \left[\prod_{n=1}^{N} \prod_{g=0}^{G} (\iota_g)^{h_{1g}^{(n)}} \left(\mathcal{N}(o_1^{(n)}|a\mathcal{M}_g, \sigma) \right)^{h_{1g}^{(n)}} \right] \\ &= \prod_{n=1}^{N} \sum_{g=0}^{G} \iota_g \mathcal{N}(o_1^{(n)}|a\mathcal{M}_g, \sigma), \end{aligned} \tag{6}$$

where the sum over $\mathbf{H}$ implies summing over all possible values of $h_{1g}^{(n)}$, or in other words, over all possible numbers of atoms g for each atomic column n. In this expression, we indeed recognise a Gaussian mixture model for N atomic columns, with mixing proportions ι_g. In other words, for a single frame, the proposed approach in this paper for atom-counting from a time series of ADF STEM images reduces to the existing approach for atom-counting [28, 29, 30, 31]. The mixing proportions correspond to the probability of having g atoms in an atomic column. During the time series, these probabilities are modified by the transition probabilities A_{jg}. We can therefore derive mixing proportions at each frame t, to quantify how the probability of having g atoms in an atomic column is changed after each transition from one frame to the next. The mixing proportion at frame $t = 1$ is given by

$$\pi_g^{(1)} = \iota_g. \tag{7}$$

In a frame $t > 1$, the probability of having g atoms in an atomic column depends on the probabilities for all possible transitions $j \rightarrow g$ from frame $t-1$ to frame t, and on the probability distribution for the different thicknesses j in the previous frame, such that:

$$\pi_g^{(t)} = \sum_{j=0}^{G} \pi_j^{(t-1)} A_{jg}, \tag{8}$$

In Section 3.2, we will use this expression to link the transition probabilities to a physical cross section related to structural changes.

2.3. Parameter estimation

The estimates for the unknown model parameters $\mathbf{\Omega}$ of the factorial hidden Markov model [Equation (5)] are obtained by maximising the likelihood [Equation (4)], or equivalently maximising the expectation value of the log likelihood [46, 50]. The expectation value of the log likelihood is evaluated during the E-step of the Expectation-Maximisation algorithm, and used to determine updates

for the model parameter estimates in the M-step of the Expectation-Maximisation algorithm. A more detailed derivation is given in Appendix A. The update formulas for the model parameters are calculated in the M-step and are expressed by the following equations:

$$\iota_g = \frac{\sum_{n=1}^{N} \mathbb{E}\left[h_{1g}^{(n)}\right]}{\sum_{n=1}^{N}\sum_{j=0}^{G} \mathbb{E}\left[h_{1j}^{(n)}\right]}, \tag{9}$$

$$A_{jg} = \frac{\sum_{t=2}^{T}\sum_{n=1}^{N} \mathbb{E}\left[h_{tg}^{(n)} h_{t-1,j}^{(n)}\right]}{\sum_{t=2}^{T}\sum_{n=1}^{N}\sum_{g'=0}^{G} \mathbb{E}\left[h_{tg'}^{(n)} h_{t-1,j}^{(n)}\right]}, \tag{10}$$

$$a = \frac{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right] o_t^{(n)} \mathcal{M}_g}{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right] \mathcal{M}_g^2}, \tag{11}$$

$$\sigma = \sqrt{\frac{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right] (o_t^{(n)} - a\mathcal{M}_g)^2}{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right]}}. \tag{12}$$

The expected values $\mathbb{E}\left[h_{tg}^{(n)}\right]$ and $\mathbb{E}\left[h_{tg}^{(n)} h_{t-1,j}^{(n)}\right]$ of the hidden states are first determined in the E-step of the Expectation-Maximisation algorithm, and are derived in Appendix A.1. The update formulas are derived in Appendix A.2. These iterative parameter updates are initialised by starting values. For the initial probabilities and transition probabilities, uniform starting values are used: $\iota_g = \frac{1}{G+1}$ and $A_{jg} = \frac{1}{G+1}$, where $G+1$ is the number of different thicknesses considered during the analysis, as the counting is allowed to start from 0 atoms. In this manner, no prior knowledge on the thickness distribution of the nanostructure or the amount of structural changes that might occur during the time series is imposed during the parameter estimation procedure. The scaling parameter is initialised by $a = 1$, which implies that we expect no or only small deviations between the parameters used to perform the image simulations and the actual experimental settings. Estimating the scaling parameter will nonetheless allow for some small variations, enabling us to account for small mismatches between experiment and image simulation. The width is initialised by $\sigma = \frac{\max(\mathbf{O})-\min(\mathbf{O})}{2(G+1)}$. The maximum number of atoms in an atomic column G that is considered during the analysis should be chosen larger than the maximum number of atoms one might expect from prior knowledge about the material such as sample preparation or additional information resulting from extra viewing directions.

2.4. Retrieving reliable counting results

In order to obtain atom-counting results for the time series, the hidden states $\mathbf{H}$ need to be estimated from the factorial hidden Markov model, once its parameters $\mathbf{\Omega}$ are estimated as described in Section 2.3. An intuitive approach might be to select the number of atoms g that is most likely for each atomic column n in a frame t, based on the likelihood defined by Equation (4). This corresponds to the approach used to determine the number of atoms from a frame by frame analysis using the existing approach for atom-counting [28, 29, 30, 31]. However, this does not suffice. Suppose that the most likely number of atoms in atomic column n at time t is g, and was j at time $t-1$. The transition probability A_{jg} may be zero or close to zero, causing this transition to be invalid or very unlikely. Therefore, an algorithm is required to consider the entire state sequence: the Viterbi algorithm [47, 48]. This path backtracking algorithm determines the most likely state sequence $\mathbf{H}$ given the observed sequence $\mathbf{O}$ and the estimated model parameters $\mathbf{\Omega}$, including the transition probabilities A_{jg}. An elaborate derivation is provided in Appendix A.3.

3. Possibilities and limitations

3.1. Performance for time series atom-counting

In order to study the performance for time series atom-counting of the proposed hidden Markov model, a set of hypothetical time series of scattering cross sections with different lengths was generated by creating hidden state sequences and observed sequences. The parameters used to generate the hypothetical hidden Markov models are summarised in Table B.1 of Appendix B. The state sequences correspond to $N = 100$ atomic columns with up to $G = 10$ atoms in an atomic column. Uniform initial probabilities ι_g were used, which represents a uniform distribution of thicknesses ranging between 1 and the maximum thickness G in the first frame. No missing atomic columns were included in the state sequences. The number of atoms in each state sequence is then changed according to a transition matrix $\mathbf{A}$ with a Gaussian spread around the diagonal with full widh at half maximum (FWHM) equal to 1.5, shown in Figure B.1 of Appendix B. This implies that some structural changes will occur during the time series, but the probability that large jumps in the number of atoms, i.e. ± 2 or more, are generated from frame to frame is small. A unitary transition matrix would imply no structural changes. The observed sequences of scattering cross sections are created using a Gaussian emission probability $\mathcal{N}(o_t^{(n)}|a\mathcal{M}_g, \sigma)$, with average scattering cross section equal to the scattering cross sections $\mathcal{M}_g$ obtained from HAADF STEM image simulations of Pt(110) as described in Appendix B. The relative width of the Gaussian emission probability considered in the simulation to study the performance is $\sigma/\delta = 0.3$ and 0.7. This width accounts for fluctuations of the scattering cross sections around the average value, caused by various effects such as scan noise, counting statistics or neighbouring atomic columns. Note that this simulation set-up implicitly assumes that all structural changes are described by the transition matrix, regardless of the process that causes the changes.

At each length of the time series T and noise level, i.e. relative width of the Gaussian emission probability σ/δ, 100 noise realisations of the hypothetical time series with the above mentioned settings were analysed. Using these hypothetical time series, the performance of the proposed hidden Markov model for time series atom-counting is compared to the existing state-of-the-art atom-counting procedure, the so-called hybrid method for atom-counting [31]. In order to count the number of atoms from a time series using the hybrid method, the scattering cross sections of all frames of the time series are jointly analysed by estimating the parameters of one Gaussian mixture model. The counting results are then extracted per frame. We will call this approach the collective hybrid method. In Figure 2, we evaluate the average percentage of correctly counted atomic columns from the hypothetical time series, with a 95% confidence interval for both methods. From these results, it is clear that the hidden Markov model counts the number of atoms more reliably than the collective hybrid method, especially for longer time series. Note that in the case of $T = 1$, both methods indeed yield exactly the same result, as mentioned before in Section 2.2. Adding even only a few extra frames allows the hidden Markov model to quickly exceed the performance of the collective hybrid method. One could argue that this improvement of the counting performance with an increasing time series length T is simply the result of the better statistics obtained by analysing the increasing set of $N \times T$ scattering cross sections. This is however not the main reason for the improved performance. We demonstrate this by analysing hypothetical hidden Markov models with the same parameters used for the analysis shown in Figure 2, as summarised in Table B.2 of Appendix B, but with a constant total number of scattering cross sections $T \times N = 1000$ considered during the analysis. Figure 3 shows the average percentage of correctly counted atomic columns with 95% confidence intervals for the analysis of these hypothetical hidden Markov models as a function of the number of frames in the time series T and the number of atomic columns in each frame N. For the collective hybrid method, increasing the number of atomic columns N or the number of frames T would both result in a larger set of scattering cross sections that are jointly analysed. For the hidden Markov model on the other hand, there is a difference, and increasing the number of frames improves the atom-counting performance, as shown in Figure 3. In other words, it is beneficial to increase the length of the time series, rather than increasing the number of atomic columns in each frame of the time series.

The improved performance of the hidden Markov model over the collective hybrid method for the same noise level (i.e. the same relative width of the Gaussian emission probability σ/δ) can therefore be attributed to the inclusion of transition probabilities to explicitly model structural changes over time. By increasing the number of frames T, the amount of frame transitions increases. Moreover, the reliability with which the transition probabilities that model the changes in the number of atoms in an atomic column through time can be estimated also increases. Instead of estimating one set of mixing proportions for the entire time series, as is the case for the collective hybrid method, initial probabilities and transition probabilities are estimated by the hidden Markov model analysis, to allow for changes in the distribution of different thicknesses in the nanostructure from frame to frame. This allows us to use the Viterbi path backtracking algorithm to determine the number of atoms in each atomic column in each frame of the time series by considering the most likely sequence, rather than the most likely number of atoms to fit the scattering cross section values in each frame separately. In this manner, we exceed the limitations imposed by the overlapping Gaussian distributions. To demonstrate this, a hypothetical time series with up to $G = 6$ atoms thickness was created using parameters summarised in Table B.3 of Appendix B. Figure 4(a) shows the set of scattering cross sections of

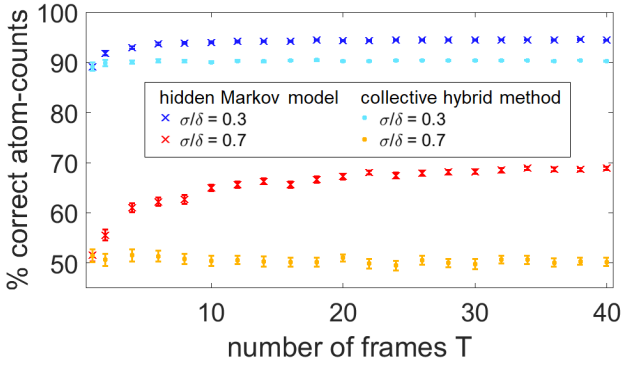

Figure 2: Comparison of the performance for atom-counting from a time series of the hidden Markov model and the collective hybrid method for atom-counting, at different noise levels σ/δ, evaluated as a function of the length of the time series T.

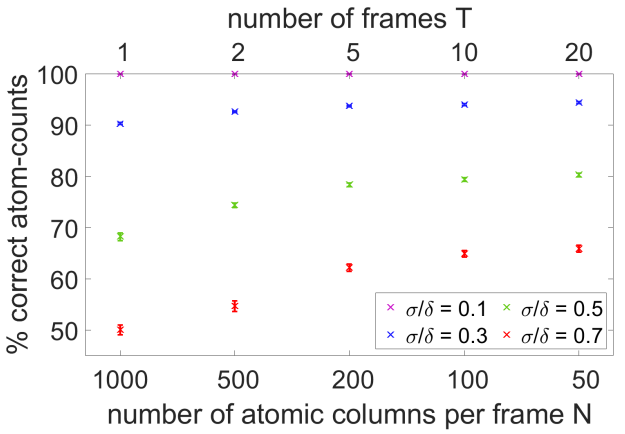

Figure 3: Performance for atom-counting from a time series of the hidden Markov model at different noise levels σ/δ, evaluated as a function of the length of the time series T and the number of atomic columns in each frame N, while the product $N \times T$ was kept constant.

the entire time series. The scattering cross sections corresponding to atomic columns with 3, 4 and 5 atoms are colour-coded in blue, green and red respectively. The scattering cross sections that correspond to atomic columns with 1, 2 or 6 atoms are all in grey. Figures 4(b) and 4(c) show the same colour-codings for the atomic columns that are counted as 3, 4 and 5 atoms respectively, using the collective hybrid method for atom-counting and the hidden Markov model for atom-counting. In the collective hybrid method, atomic columns in the tails of the Gaussian components are miscounted when components overlap [Figure 4(b)]. This limitation is no longer present for the hidden Markov model [Figure 4(c)], thanks to the Viterbi path backtracking algorithm which exploits the information on the structural changes captured by the transition probabilities. This means that the hidden Markov model for atom-counting from a time series of ADF STEM images yields more accurate counting results, as compared to the existing atom-counting approach.

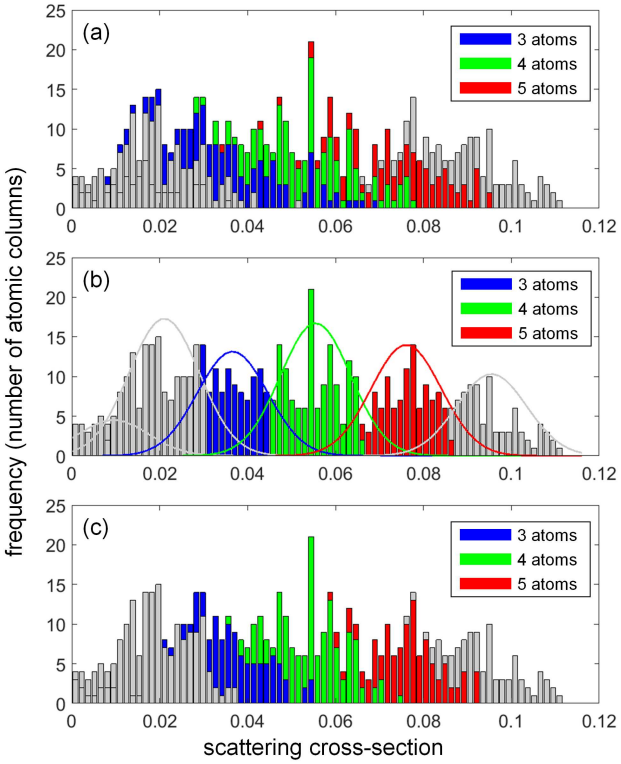

Figure 4: Histogram showing all scattering cross sections of a hypothetical time series, colour coded according to the number of atoms in an atomic column: blue for 3 atoms, green for 4 atoms, red for 5 atoms and grey for any different number of atoms. (a) Ground truth. (b) Atoms counted using the collective hybrid method, with the estimated Gaussian mixture model. (c) Atoms counted using the hidden Markov model.

3.2. Physical interpretation of transition probabilities

So far, we have shown that the hidden Markov model yields more reliable counting results for the analysis of a time series of a changing nanostructure. In order to obtain these atom-counts, the parameters of the hidden Markov model are estimated as discussed in Section 2. An important parameter of the hidden Markov model is the transition matrix $\mathbf{A}$, as this is used to model the probability that the number of atoms in an atomic column changes, and thus the probability of structural changes. In this section, we will discuss the link between the transition matrix and the cross section related to the process that drives the structural changes throughout the time series. Therefore, we have to assess the accuracy of the estimated transition probabilities, and their physical interpretation. Intensity variations during a time series can be caused either by noise fluctuations, or by actual structural changes. The hidden Markov model estimates two separate parameters to model both contributions. The noise is modelled by the width σ of the Gaussian emission probability, and the structural changes are modelled by the transition matrix $\mathbf{A}$. We will assess the accuracy of the estimated probability of structural changes using a combined measure for the $G(G+1)$ estimated transition probabilities and the G estimated initial probabilities:

$$P = \frac{\sum_{g=0}^{G} \pi_g (1 - A_{gg})}{G+1}. \tag{13}$$

This expresses the weighted average probability that the number of atoms in an atomic column changes during the time series, as $1 - A_{gg}$ corresponds to the probability that the number of atoms in an atomic column will not stay the same from frame to frame. The weights π_g follow from the estimated transition and initial probabilities:

$$\pi_g = \frac{1}{T-1} \sum_{t=1}^{T-1} \pi_g^{(t)}, \tag{14}$$

with $\pi_g^{(t)}$ defined in Equation (8).

In order to validate the physical interpretation of the estimated transition matrix in terms of atomic structural changes, we created 50 noise realisations of a hypothetical time series with the parameters of the hidden Markov model estimated from the experimental time series of the Pt wedge that will be analysed further in this section, shown in Figure 5. The width of the Gaussian emission probability σ is varied to obtain different noise levels. The width σ can be related to the electron dose d through $\sigma = \sqrt{\frac{\mu_g}{d}}$ when only Poisson noise is present, with μ_g the scattering cross section for g atoms in an atomic column [51]. In this manner, at the average sample thickness, the noise levels σ/δ used during the following analysis can be translated to electron doses ranging between 7×10^4 e$^-/\mathring{A}^2$ and 9×10^2 e$^-/\mathring{A}^2$. Table 1 summarises the 95% confidence intervals for the mean estimated values of the probability of structural changes P from Equation (13) and the relative width of the Gaussian emission probability σ/δ, at the different noise levels, together with the ground truth of these parameters. The percentage of correctly counted

atomic columns is also summarised using a 95% confidence interval. Furthermore, results are shown for an analysis of 50 noise realisations of a hypothetical time series with the same settings, but with a more diffuse transition matrix, as to allow for more structural changes during the time series. This transition matrix is shown in Figure B.2 of Appendix B. In this manner, we can assess the accuracy of the estimated parameters that quantify noise and structural changes. For a small underlying probability of structural changes P, the estimated values of P and σ/δ are accurate up to high noise levels. This indicates that intensity variations due to structural changes are indeed quantified by the transition matrix $\mathbf{A}$, while noise fluctuations are quantified by the width of the Gaussian emission probability σ. When the underlying probability of structural changes P is larger, the estimated values of P and σ/δ still remain accurate up to a reasonably high noise level, but are slightly more inaccurate, as well as less precise, at high noise levels. This also leads to a lower percentage of correct atom-counts. We conclude that the separation of noise fluctuations from atomic structural changes is reliably done by the hidden Markov model analysis for small values of P, even at a low electron dose. When the probability of structural changes P is larger, the separation of noise and structural changes can still be performed reasonably well, although slight inaccuracies arise when the electron dose is lowered, making it more difficult to discern the origin of the fluctuations in the scattering cross sections from frame to frame.

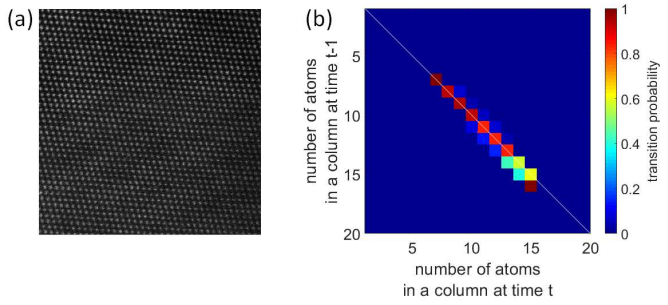

Figure 5: (a) First frame of Pt wedge time series. (b) Estimated transition matrix.

Therefore, when the electron dose is sufficiently high, Equation (13) reliably describes the probability of structural changes. In this manner, we can now estimate the probability of structural changes for an experimental ADF STEM time series of a Pt wedge using the hidden Markov model analysis. The time series consists of $T = 6$ frames and was previously also analysed in [51]. The first frame of the time series is shown in Figure 5(a), and all frames are shown in Figure C.1 of Appendix C. The consecutive ADF STEM images of the Pt wedge were recorded using a beam current of 37 pA, corresponding to an electron dose of 2.8 $\times 10^4$ $e^-/\mathring{A}^2$. The counting results obtained by the hidden Markov model analysis for each image are shown in Figure C.2 of Appendix C. The transition matrix estimated during this hidden Markov model analysis is shown in Figure 5(b). The white diagonal line plotted on top of the transition matrix indicates the transitions where the number of atoms in an atomic column stays the same. The upper and lower triangles contain the probabilities for an atomic column to respectively gain or lose one or more atoms. Using Equation (13), this yields an estimated probability of structural changes $P = 12.6\%$. We do not expect structural changes to be caused by sputtering of atoms from the surface, only by surface diffusion, since the threshold energy for sputtering Pt atoms from a convex surface with step sites is 379 keV, well above the incident electron energy of 300 keV [13, 51]. In order to relate the probability of structural changes to the probability of surface displacement, a factor $\frac{1}{2}$ should be added in Equation (13), to account for the top and bottom surface of the nanostructure:

$$P_{\mathrm{sd}} = \frac{1}{2}\frac{\sum_{g=0}^{G}\pi_g(1-A_{gg})}{G+1}. \tag{15}$$

The cross section σ_{sd} and probability P_{sd} for surface displacement are related as follows [51]:

$$P_{\mathrm{sd}} = \sigma_{\mathrm{sd}} d \frac{N_{\mathrm{ad}}}{N}, \tag{16}$$

with d the electron dose, N the total number of atomic columns, and N_{ad} the number of adatoms, taken equal to the number of atoms with coordination number less than or equal to 6 in the symmetrical 3D atom configuration based on the atom-counting. In this manner, we estimate the cross section related to surface diffusion for this Pt wedge equal to $\sigma_{sd} = (5.60 \pm 0.05) \times 10^{-6}\mathring{A}^2$ from the probability of surface displacement $P_{sd} = 6.3\%$ and the estimated fraction of adatoms $\frac{N_{\mathrm{ad}}}{N} = 0.4$. The error bar in this expression is equal to the standard deviation of the mean estimated cross section from 100 hypothetical time series using the same parameters estimated for the experimental Pt wedge. This value for the cross section related to surface displacement corresponds to a surface diffusion threshold energy of 1.09 eV [52]. This is in close agreement with the theoretical threshold energy for surface diffusion for Pt(110), which is calculated as 1.07 eV, corresponding to a cross section of $\sigma_{sd} = 16 \times 10^{-6}\mathring{A}^2$ [53]. The underestimation of the cross section by the hidden Markov model analysis can be understood since the estimated transition probabilities describe net structural changes from frame to frame. These net structural changes can however be the result of multiple hops of atoms, potentially in opposite directions, yielding a smaller amount of net changes after the frame is acquired as compared to the actual amount of structural changes. In this manner, we unavoidably underestimate the cross section related to surface diffusion. However, this methodology gets a closer match with the theory as compared to the current state-of-the-art analysis of variance method, where a cross section

P (%)			σ/δ			correct atom-counts (%)	
ground truth	95% CI		ground truth	95% CI		95% CI	
	lower bound	upper bound		lower bound	upper bound	lower bound	upper bound
12.55	12.42	12.69	0.1	0.0998	0.1002	99.99	100.00
	12.35	12.70	0.3	0.2990	0.3010	97.16	97.28
	12.38	12.81	0.5	0.4968	0.5003	88.69	89.00
	11.67	12.44	0.7	0.6945	0.6998	79.02	79.53
	11.55	12.74	0.9	0.8908	0.8972	69.83	71.12
59.52	59.01	59.38	0.1	0.1000	0.1004	99.99	100.00
	59.01	59.47	0.3	0.2974	0.2992	92.89	93.05
	59.60	60.34	0.5	0.4818	0.4888	75.88	76.16
	61.50	63.26	0.7	0.6603	0.6762	61.54	62.07
	63.72	66.70	0.9	0.8363	0.8525	50.07	51.35

Table 1: Accuracy of the probability of structural changes P, the relative width of the Gaussian emission probability σ/δ and the atom-counts estimated by the hidden Markov model.

of $\sigma_{sd} = (0.74 \pm 0.20) \times 10^{-6} \text{Å}^2$ was estimated [51], and is therefore a promising approach for the reliable quantification of physical cross sections from experimental ADF STEM time series.

3.3. Multiple transition matrices

The hidden Markov model presented in this paper models a set of transition probabilities, summarised by one transition matrix $\mathbf{A}$. In other words, so far, we have assumed that the experimental conditions driving the structural changes, such as electron irradiation, temperature or gas flow, are time-invariant. In order to incorporate nanomaterials in real applications, a fundamental understanding of the mechanisms behind atomic structure evolutions under variable environmental conditions is of crucial importance. We therefore need to test whether the method presented here can be used to reliably quantify, for example, changes during reduction and oxidation reactions. To this purpose, we will look at the atomic structure of a Pt catalyst nanoparticle, shown in Figure 6(a), previously studied in an alternating hydrogen (H_2) and oxygen (O_2) flow using the existing state-of-the-art atom-counting procedure for single ADF STEM images [16]. During the gas flow experiment, the Pt nanoparticle changes from a facetted to a round morphology and vice versa. Therefore, the underlying transition probabilities are not the same for each frame transition. In fact, two alternating transition matrices can be recognised, as illustrated in Figure 6(a). In this section, we examine the effect on the quantification when we nonetheless assume one transition matrix during the estimation procedure.

To this purpose, we created a hypothetical time series with $T = 10$ frames, based on the atom-counts that were previously quantified from the single ADF STEM images of the catalyst Pt nanoparticle in an H_2 and O_2 environment [16]. The initial probabilities ι_g are set equal to the mixing proportions quantified previously for the Pt nanoparticle in an H_2 environment. A direct Markov chain analysis of a transition from the atom-counts previously quantified for the Pt nanoparticle in an H_2 environment to the atom-counts quantified in the O_2 environment yields the transition matrix for the first process $\mathbf{A}_{H_2 \to O_2}$, shown in Figure 6(b). Figure 6(c) shows the transition matrix for the inverse process $\mathbf{A}_{O_2 \to H_2}$. Note that Figure 6(b) clearly shows non-zero probabilities for transitions of 0 atoms to multiple atoms, which indicate the appearance of extra atomic columns during the transition from a facetted to a round morphology. The frame transitions of the state sequence are then generated alternately from the two transition matrices $\mathbf{A}_{H_2 \to O_2}$ and $\mathbf{A}_{O_2 \to H_2}$. The number of atoms in an atomic column ranges between 0 and $G = 26$, and the total number of atomic columns analysed in each frame is $N = 2457$. The observed sequence of scattering cross sections is created using a Gaussian emission probability with average scattering cross sections equal to the Pt library values $\mathcal{M}_g$ previously also used in the simulations in Sections 3.1 and 3.2. The relative width of the Gaussian emission probability is $\sigma/\delta = 0.1$ and 0.7, which corresponds approximately to an electron dose of respectively 10^5 e$^-/\text{Å}^2$ and 2×10^3 e$^-/\text{Å}^2$ per frame, assuming Poisson noise only. These parameters are also summarised in Table B.4 of Appendix B. At each noise level σ/δ, 50 noise realisations of the observed sequence have been analysed. The hidden Markov model analysis is performed assuming a maximum thickness of $G = 30$ atoms. Figure 6(d) shows the transition matrix estimated from a time series with $\sigma/\delta = 0.1$ by the hidden Markov model which assumes only one transition matrix during the time series. This estimated transition matrix for the time series with alternating structural changes is close to the average of both underlying transition matrices, weighted with the number of times a frame transition occurred driven by each type of transition matrix. The percentages of correctly counted atomic columns for the different time series are summarised in the first two columns of Table 2 using 95% confidence intervals obtained from the 50 noise realisations. Next, we compare these results to the performance discussed in Section 3.1. The analysis shown in

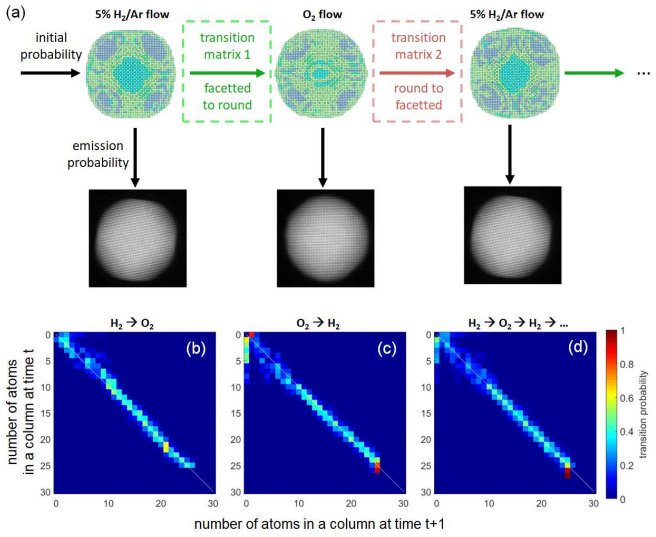

Figure 6: (a) A Pt catalyst nanoparticle in an alternating hydrogen and oxygen flow changes morphology from facetted to round and vice versa, analysed previously using single frame atom-counting [16]. These structural changes can be described by alternating transition matrices. (b-c) Transition matrices 1 & 2, estimated from the atom-counts previously quantified in [16] using a direct Markov chain analysis for the changes from facetted to round morphology and vice versa respectively. (d) Transition matrix estimated using a hidden Markov model assuming only one transition matrix from a hypothetical time series corresponding to a Pt nanoparticle with alternating facetted and round morphology with $\sigma/\delta = 0.1$.

Figures 2 and 3 yields $(99.99 \pm 0.01)\%$ and $(64.66 \pm 0.67)\%$ correctly counted atomic columns for $T = 10$ at $\sigma/\delta = 0.1$ and 0.7 respectively. From Table 2, it follows that the performance for atom-counting from a time series with two underlying transition matrices using the hidden Markov model as described in this paper remains unaffected for low enough σ/δ, corresponding to high enough electron doses. However, at higher σ/δ, or equivalently at lower electron doses, the performance is slightly worse as compared to atom-counting from a time series with only one underlying transition matrix, as discerning between noise and structural changes becomes more challenging. Nonetheless, the hidden Markov model, with $(58.00 \pm 0.78)\%$ correctly counted atomic columns as summarised in Table 2, still outperforms the collective hybrid analysis, which reached only $(51.44 \pm 0.82)\%$ correctly counted atomic columns for $T = 10$ and $\sigma/\delta = 0.7$, as shown in Figure 2. We conclude that at high electron doses the assumption of a constant transition matrix does not pose any restrictions for applying the hidden Markov model for counting the number of atoms in a time series of ADF STEM images of an *in situ* experiment with strongly alternating environmental conditions. When the electron dose is lower, the performance for atom-counting decreases, but nonetheless reasonably reliable counting results can still be obtained.

One should however be careful when interpreting the transition probabilities in terms of physical cross sections in the same manner as discussed in Section 3.2. In order to estimate physical cross sections related to the two separate processes, we analyse the estimated atom-counts of the time series as a Markov chain with two transition matrices. The probabilities of structural changes for both processes can then be estimated from these two transition matrices, similarly to the approach discussed in Section 3.2. This is described in more detail in Appendix D.1. The probabilities of structural changes $P_{H_2 \to O_2}$ and $P_{O_2 \to H_2}$ estimated in this manner from the hypothetical time series discussed above are summarised in the last two columns of Table 2 using 95% confidence intervals on the mean estimated value, together with the expected values. The expected probabilities of structural changes $P_{H_2 \to O_2}$ and $P_{O_2 \to H_2}$ are estimated using Equation (13) from the transition matrices $\mathbf{A}_{\mathrm{H_2 \to O_2}}$ and $\mathbf{A}_{\mathrm{O_2 \to H_2}}$ obtained from the previously quantified counting results, shown in Figure 6(b) and 6(c). At a low noise level σ/δ, the hidden Markov model accurately estimates the probabilities of structural changes for both processes. When the noise increases, the estimated probabilities are slightly inaccurate, as it becomes more difficult to separate structural changes from noise fluctuations. These results show that through post-processing of the results obtained from the hidden Markov model analysis, probabilities of structural changes and even cross sections can still be quantified for the different physical processes causing structural changes during a time series with alternating environmental conditions.

We can now go a step further, and consider more than two transition matrices. In order to test the performance for atom-counting of the hidden Markov model analysis for, for example, a heating experiment where the temperature is increased gradually, we simulated a hypothetical observed sequence generated from a state sequence obtained by changing the transition matrix at each frame transition, according to the parameters summarised in Table B.5 of Appendix B. At each transition from frame to frame, a transition matrix with a Gaussian spread around the diagonal is used, each time with an increasing value for the FWHM, in order to mimic the increasing mobility of the atoms in the nanostructure when the temperature increases. The transition matrices used for this simulation are shown in Figure D.1 of Appendix D.2. At a low and high noise level ($\sigma/\delta = 0.3$ and 0.7), we obtain a percentage of respectively 93.6% and 60.8% correctly counted atomic columns during the time series. This performance is close to the results shown previously in Figure 2, where at $T = 22$ the respective obtained percentages of correct atom-counts are $(94.34 \pm 0.13)\%$ and $(68.00 \pm 0.41)\%$. This implies that we can still obtain reasonably reliable counting results from a time series with non-constant underlying transition matrix.

In conclusion, it is possible to use the existing framework of hidden Markov model for atom-counting when the underlying driving process is non-constant. This opens up possibilities for the analysis of *in situ* heating or gas flow

	correct atom-counts (%)	$P_{H_2 \to O_2}$ (%)	$P_{O_2 \to H_2}$ (%)
expected value	100	70.09	64.20
$\sigma/\delta = 0.1$ (95% CI)	[99.99; 100.00]	[69.65; 70.08]	[63.92; 64.29]
$\sigma/\delta = 0.7$ (95% CI)	[57.28; 58.72]	[65.87; 68.37]	[64.85; 67.17]

Table 2: Performance for atom-counting and accuracy of the estimated probability of structural changes P for the two processes of the variable gas flow illustrated in Figure 6.

experiments.

4. Conclusions

We have presented a methodology for atom-counting from a time series of ADF STEM images using a hidden Markov model. For single frames, this approach for atom-counting is identical to the existing state-of-the-art atom-counting procedure, but it quickly outperforms the state-of-the-art when one or more frames are added to the time series, both at high and low electron doses. Adding extra frames of the nanostructure under study to the time series further increases the performance of the hidden Markov model. We have shown that this improvement is largely due to the increased number of frames in the time series, which leads to more reliable estimates for the transition probabilities. Therefore, the Viterbi path backtracking algorithm yields a higher percentage of correctly counted atomic columns, by retrieving the most likely state sequence.

Furthermore, we have shown how the transition matrix can be related to the probability of structural changes, and how this probability of structural changes can be related to a physical cross section. We have applied this to an experimental time series of a Pt wedge, for which we calculated the cross section related to surface displacement. The corresponding surface diffusion threshold energy is in close agreement with the theoretical value. The estimated cross section value underestimates the theoretical value, as multiple hopping events can occur during one transition from frame to frame. We have discussed the estimation of a cross section related to adatom hopping from the transition matrix, but in general cross sections for any mechanism driving the structural changes can be estimated from the transition matrix in a similar manner.

Finally, we have discussed whether it is possible to use the hidden Markov model which assumes a constant transition matrix throughout the time series, when in fact the structural changes are driven by multiple transition matrices, changing over the course of the time series. We show that the percentage of correctly counted atomic columns is not strongly affected, especially at high electron doses, but one should be careful when interpreting the transition probabilities in terms of physical probabilities of structural changes or even cross sections, as an average transition matrix is estimated for the time series. When the changes in the environmental conditions throughout the time series are known, it is however possible to extract probabilities of structural changes for the different conditions using a post-processing based on the counting results obtained by the hidden Markov model analysis.

In summary, the hidden Markov model opens up possibilities for the reliable quantification of the atomic structure of nanoparticles that undergo structural changes, for example during *in situ* gas flow or heating experiments, as a consequence of the electron beam irradiation, or resulting from adatom hopping and surface diffusion. The hidden Markov model for atom-counting from time series of ADF STEM images has been implemented in the freely available StatSTEM software [34].

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 770887 and No. 823717 ESTEEM3). The authors acknowledge financial support from the Research Foundation Flanders (FWO, Belgium) through grants to A.D.w. and A.D.B. and projects G.0502.18N and EOS 30489208.

A. Parameter estimation

Table A.1 summarises the notation used for the factorial hidden Markov model for atom-counting.

The parameters of the factorial hidden Markov model are estimated using an Expectation-Maximisation algorithm, typically called a Baum-Welch algorithm in the context of hidden Markov models. This is an iterative updating algorithm, consisting of two steps: an E-step, and an M-step. During the E-step, the likelihood is evaluated, and this quantity is maximised during the M-step. The parameter estimates are in practice obtained by maximising the expectation value of the log likelihood [46, 50]:

$$\begin{aligned}\mathbb{E}\left[\ln p(\mathbf{O},\mathbf{H}|\mathbf{\Omega})\right] &= \sum_{n=1}^{N}\sum_{g=0}^{G}\mathbb{E}\left[h_{1g}^{(n)}\right]\ln \iota_g \\ &+ \sum_{t=2}^{T}\sum_{n=1}^{N}\sum_{j=0}^{G}\sum_{g=0}^{G}\mathbb{E}\left[h_{t-1,j}^{(n)}h_{tg}^{(n)}\right]\ln A_{jg} \\ &+ \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{g=0}^{G}\mathbb{E}\left[h_{tg}^{(n)}\right]\ln \mathcal{N}\left(o_t^{(n)}|a\mathcal{M}_g,\sigma\right). \qquad \text{(A.1)}\end{aligned}$$

A.1. E-step

In this step, the likelihood is evaluated, and the expected values $\mathbb{E}\left[h_{tg}^{(n)}\right]$ and $\mathbb{E}\left[h_{tg}^{(n)}h_{t-1,j}^{(n)}\right]$ that occur in the update

Symbol	Explanation
T	Number of frames in the time series
N	Number of atomic columns in each frame
G	Maximum number of atoms in an atomic column during the time series
$\mathbf{H}$	$T \times N \times (G+1)$ tensor with hidden states: $\mathbf{H} = \left\{\mathbf{h}_t^{(n)}\right\}$
$\mathbf{h}_t^{(n)}$	$1 \times (G+1)$ binary state vector at time t: $\mathbf{h}_t^{(n)} = \left\{h_{tg}^{(n)}\right\}$
$h_{tg}^{(n)}$	Binary state variable: $h_{tg}^{(n)} = 1$ if atomic column n has g atoms at time t
$\mathbf{O}$	$T \times N$ matrix with observed states: $\mathbf{O} = \left\{o_t^{(n)}\right\}$
$o_t^{(n)}$	Scattering cross section of atomic column n at time t
ι_g	Initial probability for an atomic column to have g atoms in frame 1
$\mathbf{A}$	$(G+1) \times (G+1)$ transition matrix: $\mathbf{A} = \{A_{jg}\}$
A_{jg}	Transition probability from j to g atoms between two frames
μ_g	Average scattering cross section corresponding to g atoms in an atomic column
$\mathcal{M}_g$	Library value for an atomic column with g atoms
a	Scaling parameter relating the average scattering cross section to the library: $\mu_g = a\mathcal{M}_g$
σ	Width of the Gaussian distribution around the average scattering cross section
δ	Average distance between subsequent scattering cross sections μ_g and μ_{g+1}
$\mathbf{\Omega}$	Parameter vector of the hidden Markov model, defined in Equation (5)

Table A.1: Notation overview

formulas in the M-step are determined. The derivation is similar to the E-step of the general Baum-Welch algorithm, but factorised over the atomic columns. The expectation value of a binary variable, equals the probability for this binary variable to equal 1. Therefore, the expected values $\mathbb{E}\left[h_{tg}^{(n)}\right]$ and $\mathbb{E}\left[h_{tg}^{(n)} h_{t-1,j}^{(n)}\right]$ can be written as the following probabilities:

$$\mathbb{E}\left[h_{tg}^{(n)}\right] = p\left(h_{tg}^{(n)} = 1|\mathbf{O}, \mathbf{\Omega}\right), \tag{A.2}$$

often called the state occupation probability, and

$$\mathbb{E}\left[h_{t-1,j}^{(n)} h_{tg}^{(n)}\right] = p\left(h_{t-1,j}^{(n)} = 1, h_{tg}^{(n)} = 1|\mathbf{O}, \mathbf{\Omega}\right), \tag{A.3}$$

since the product of two binary variables only equals 1 if both variables are equal to 1. Equations (A.2) and (A.3) can be rewritten as follows [54]:

$$p\left(h_{tg}^{(n)} = 1|\mathbf{O}\right) = \frac{p\left(\mathbf{O}|h_{tg}^{(n)} = 1\right) p\left(h_{tg}^{(n)} = 1\right)}{p\left(\mathbf{O}\right)} \tag{A.4}$$

$$= \frac{p\left(o_1^{(n)}, \cdots, o_t^{(n)}, h_{tg}^{(n)} = 1\right) p\left(o_{t+1}^{(n)}, \cdots, o_T^{(n)}|h_{tg}^{(n)} = 1\right)}{p\left(\mathbf{O}\right)}, \tag{A.5}$$

and

$$p\left(h_{t-1,j}^{(n)} = 1, h_{tg}^{(n)} = 1|\mathbf{O}\right)$$

$$= \frac{p\left(\mathbf{O}|h_{t-1,j}^{(n)} = 1, h_{tg}^{(n)} = 1\right) p\left(h_{t-1,j}^{(n)} = 1, h_{tg}^{(n)} = 1\right)}{p\left(\mathbf{O}\right)} \tag{A.6}$$

$$= \frac{1}{p\left(\mathbf{O}\right)}\Bigg[p\left(o_1^{(n)}, \cdots, o_{t-1}^{(n)}, h_{t-1,j}^{(n)} = 1\right)$$

$$\times p\left(h_{t-1,j}^{(n)} = 1|h_{tg}^{(n)} = 1\right) p\left(o_t^{(n)}|h_{tg}^{(n)} = 1\right)$$

$$\times p\left(o_{t+1}^{(n)}, \cdots, o_T^{(n)}|h_{tg}^{(n)} = 1\right)\Bigg]. \tag{A.7}$$

Note that each probability in the expression above is implicitly conditional on the parameter vector $\mathbf{\Omega}$, which was omitted in order to reduce the complexity of the notation. We can now introduce the (scaled) forward variables $\hat{\alpha}(h_{tg}^{(n)})$ and (scaled) backward variables $\hat{\beta}(h_{tg}^{(n)})$:

$$\hat{\alpha}(h_{tg}^{(n)}) = \frac{p\left(o_1^{(n)}, \cdots, o_t^{(n)}, h_{tg}^{(n)} = 1|\mathbf{\Omega}\right)}{p\left(o_1^{(n)}, \cdots, o_t^{(n)}|\mathbf{\Omega}\right)}$$

$$= \frac{p\left(o_1^{(n)}, \cdots, o_t^{(n)}, h_{tg}^{(n)} = 1|\mathbf{\Omega}\right)}{\prod_{r=1}^{t} c_t^{(n)}}, \tag{A.8}$$

and

$$\hat{\beta}(h_{tg}^{(n)}) = \frac{p\left(o_{t+1}^{(n)}, \cdots, o_T^{(n)}|h_{tg}^{(n)} = 1, \mathbf{\Omega}\right)}{p\left(o_{t+1}^{(n)}, \cdots, o_T^{(n)}|\mathbf{\Omega}\right)}$$

$$= \frac{p\left(o_{t+1}^{(n)}, \cdots, o_T^{(n)}|h_{tg}^{(n)} = 1, \mathbf{\Omega}\right)}{\prod_{r=t+1}^{T} c_t^{(n)}}, \tag{A.9}$$

with

$$c_t^{(n)} = p(o_t^{(n)}|o_1^{(n)}, \cdots, o_{t-1}^{(n)}|\mathbf{\Omega}). \tag{A.10}$$

The expected values required during the M-step can therefore be determined using the so-called forward-backward algorithm:

$$\mathbb{E}\left[h_{tg}^{(n)}\right] = \hat{\alpha}(h_{tg}^{(n)})\hat{\beta}(h_{tg}^{(n)}) \tag{A.11}$$

$$\mathbb{E}\left[h_{tg}^{(n)} h_{t-1,j}^{(n)}\right] = \hat{\alpha}(h_{t-1,j}^{(n)}) A_{jg} \mathcal{N}(o_t^{(n)}|a\mathcal{M}_g, \sigma)\hat{\beta}(h_{tg}^{(n)}), \tag{A.12}$$

with the forward and backward variables determined recursively. For computational reasons, an additional normalisation of $\mathbb{E}\left[h_{tg}^{(n)} h_{t-1,j}^{(n)}\right]$ is performed. Initialisation of the normalised forward variable is done by

$$\hat{\alpha}(h_{1g}^{(n)}) = \frac{\iota_g \mathcal{N}(o_1^{(n)}|a\mathcal{M}_g, \sigma)}{c_1^{(n)}}. \tag{A.13}$$

Then we forwards propagate through time (hence the name), in order to recursively obtain the next forward variables:

$$\hat{\alpha}(h_{tg}^{(n)}) = \frac{1}{c_t^{(n)}} \mathcal{N}(o_t^{(n)}|a\mathcal{M}_g, \sigma) \sum_{j=1}^{G} \hat{\alpha}(h_{t-1,j}^{(n)}) A_{jg}.$$

Initialisation of the backward variable is done by

$$\hat{\beta}(h_{Tg}^{(n)}) = 1. \tag{A.14}$$

Then we backwards propagate through time (hence the name), in order to recursively obtain the other backward variables:

$$\hat{\beta}(h_{tg}^{(n)}) = \frac{1}{c_{t+1}^{(n)}} \sum_{j=1}^{G} \mathcal{N}(o_{t+1}^{(n)}|a\mathcal{M}_j, \sigma) \hat{\beta}(h_{t+1,j}^{(n)}) A_{gj}. \tag{A.15}$$

It can furthermore be shown that the likelihood is obtained as the product of the scaling parameters $c_t^{(n)}$ that are used to avoid numerical issues with the forward-backward algorithm, as follows:

$$\begin{aligned} c_t^{(n)} &= p(o_t^{(n)}|o_1^{(n)}, ..., o_{t-1}^{(n)}, \mathbf{\Omega}) && \text{(A.16)} \\ \Rightarrow\ & p(o_1^{(n)}, ..., o_t^{(n)}|\mathbf{\Omega}) = \prod_{q=1}^{t} c_q^{(n)} && \text{(A.17)} \\ \Rightarrow\ & p(\mathbf{O}|\mathbf{\Omega}) = \prod_{n=1}^{N} \prod_{q=1}^{T} c_q^{(n)}. && \text{(A.18)} \end{aligned}$$

A.2. M-step

The update formulas for the model parameters are calculated in the M-step using the expected values $\mathbb{E}\left[h_{tg}^{(n)}\right]$ and $\mathbb{E}\left[h_{tg}^{(n)} h_{t-1,j}^{(n)}\right]$ calculated during the E-step. These expressions are obtained by maximising the expectation value of the log likelihood from Equation (A.1), while taking into account that ι_G and A_{jG} $\forall 0 \leq j \leq G$ need not be estimated, since $\sum_{g=0}^{G} \iota_g = 1$ and $\sum_{g=0}^{G} A_{jg} = 1$. This results in the following update formulas:

$$\iota_g = \frac{\sum_{n=1}^{N} \mathbb{E}\left[h_{1g}^{(n)}\right]}{\sum_{n=1}^{N} \sum_{j=0}^{G} \mathbb{E}\left[h_{1j}^{(n)}\right]}, \tag{A.19}$$

$$A_{jg} = \frac{\sum_{t=2}^{T} \sum_{n=1}^{N} \mathbb{E}\left[h_{tg}^{(n)} h_{t-1,j}^{(n)}\right]}{\sum_{t=2}^{T} \sum_{n=1}^{N} \sum_{g'=0}^{G} \mathbb{E}\left[h_{tg'}^{(n)} h_{t-1,j}^{(n)}\right]}, \tag{A.20}$$

$$a = \frac{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right] o_t^{(n)} \mathcal{M}_g}{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right] \mathcal{M}_g^2}, \tag{A.21}$$

$$\sigma = \sqrt{\frac{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right] (o_t^{(n)} - a\mathcal{M}_g)^2}{\sum_{t=1}^{T} \sum_{n=1}^{N} \sum_{g=0}^{G} \mathbb{E}\left[h_{tg}^{(n)}\right]}}. \tag{A.22}$$

A.3. Viterbi algorithm

The goal of the Viterbi algorithm for hidden Markov models is to determine the most likely hidden state sequence $\mathbf{H}$. We could determine the individually most likely number of atoms in atomic column n at time t as

$$q_t^{(n)} = \underset{0 \leqslant g \leqslant G}{\arg\max} \left(\mathbb{E}\left[h_{tg}^{(n)}\right] \right), \text{ for } 1 \leqslant t \leqslant T, \tag{A.23}$$

but the transition probability from $t-1$ to this state at time t may be zero, causing this state to be invalid. Therefore, an algorithm is required to consider the entire state sequence: the Viterbi algorithm.

In order to retrieve the state sequence, the so-called best score $\delta^{(n)}(t, g)$ is introduced. This considers the probability for each possible state sequence $\mathbf{h}_1^{(n)}, \cdots, \mathbf{h}_{t-1}^{(n)}$ that could have proceeded state g at time t, and that could account for the first t observations $o_1^{(n)}, \cdots, o_t^{(n)}$:

$$\delta^{(n)}(t, g) = \max_{\mathbf{h}_1^{(n)}, \cdots, \mathbf{h}_{t-1}^{(n)}} P\Big(\mathbf{h}_1^{(n)}, \cdots, \mathbf{h}_{t-1}^{(n)}, h_{tg}^{(n)} = 1, o_1^{(n)}, \cdots, o_t^{(n)}|\mathbf{\Omega} \Big). \tag{A.24}$$

The best score at the next time $t+1$ can be retrieved using a recursion relation:

$$\delta^{(n)}(t+1, j) = \max_{g} \left[\delta^{(n)}(t, g) \cdot A_{gj} \right] \mathcal{N}(o_{t+1}^{(n)}|a\mathcal{M}_j, \sigma). \tag{A.25}$$

In order to keep the numerical range of the best score computationally feasible, in practice this expression is normalised for each time t and each atomic column n. In order to retrieve the most likely state sequence, we need to keep track of the argument that maximises this expression. Note that this is not affected by the additional normalisation. Therefore, an argument array $\phi^{(n)}(t, g)$ is introduced.

We can now summarise the Viterbi algorithm as follows. First, we initialise the best score and the argument array:

$$\delta^{(n)}(1, g) = \iota_g \mathcal{N}(o_1^{(n)}|a\mathcal{M}_g, \sigma), \tag{A.26}$$

$$\phi^{(n)}(1, g) = 0. \tag{A.27}$$

Then the recursion formula is applied for each next time t:

$$\delta^{(n)}(t, g) = \max_{0 \leqslant j \leqslant G} \left[\delta^{(n)}(t-1, j) A_{jg} \right] \mathcal{N}(o_t^{(n)}|a\mathcal{M}_g, \sigma), \tag{A.28}$$

$$\phi^{(n)}(t,g) = \underset{0\leqslant j\leqslant G}{\arg\max}\left[\delta^{(n)}(t-1,j)A_{jg}\right]. \tag{A.29}$$

Next, we perform a termination step at time T:

$$\hat{P} = \max_{0\leqslant g\leqslant G}\left[\delta^{(n)}(T,g)\right], \tag{A.30}$$

$$\hat{\mathbf{h}}_T^{(n)} = \underset{0\leqslant g\leqslant G}{\arg\max}\left[\delta^{(n)}(T,g)\right]. \tag{A.31}$$

Finally, path backtracking is performed in order to determine the most likely hidden state sequence:

$$\hat{\mathbf{h}}_t^{(n)} = \phi^{(n)}(t+1,\hat{h}_{t+1}^{(n)}) \tag{A.32}$$

B. Simulation parameters

A hypothetical hidden Markov model is simulated by creating hypothetical hidden state sequences (number of atoms) and corresponding observed sequences (scattering cross sections). We summarise the parameters used in the simulations of Section 3 in Tables B.1-B.5.

Each simulation discussed in this paper uses the same library of scattering cross sections obtained from image simulations. These scattering cross sections, also called library values $\mathcal{M}_g$, were obtained from HAADF STEM image simulations using the MULTEM software [55, 56] for Pt(110) at 300 keV, using a 21 mrad semi-convergence angle, 58-190 mrad detector collection range, and a pixel size of 9.1 pm. This library corresponds to the experimental images of a Pt wedge that will be discussed in Section 3.2, and which have previously been analysed in [51] as well.

The state sequences of such hypothetical time series correspond to T frames with N atomic columns in each frame with up to G atoms in an atomic column, created using a set of initial probabilities ι_g for $0 \leq g \leq G$. The number of atoms in each state sequence is then changed according to a transition matrix $\mathbf{A}$. The transition matrix can be generated using a Gaussian spread around the diagonal, such as shown in Figure B.1. The observed sequences of scattering cross sections are generated using a Gaussian emission probability $\mathcal{N}(o_t^{(n)}|a\mathcal{M}_\}, \sigma)$. The library values $\mathcal{M}_g$ were scaled using a scaling parameter $a = 1$ in all simulations. The relative width of the Gaussians considered in the simulation to study the performance is σ/δ, where δ is the average difference between subsequent average scattering cross sections $a\mathcal{M}_g$.

C. Pt wedge

Figure C.1 shows the 6 ADF STEM images of the time series of a Pt wedge analysed in Section 3.2. The counting results obtained for these ADF STEM images using a hidden Markov model are shown in Figure C.2.

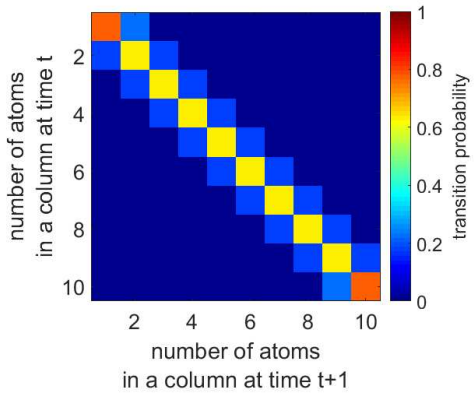

Figure B.1: Transition matrix with a Gaussian spread around the diagonal with FWHM = 1.5.

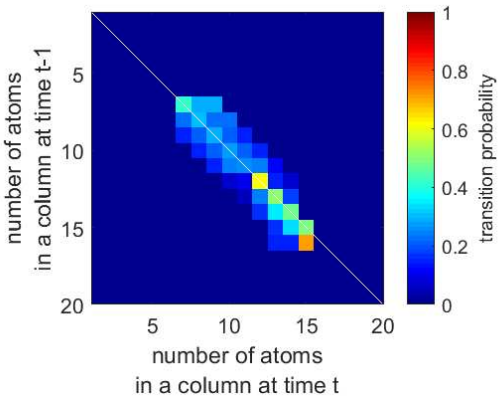

Figure B.2: More diffuse transition matrix, based on the estimated transition matrix for the Pt wedge time series from Figure 5, used for the results in Table 1.

D. Multiple transition matrices

D.1. Probability of structural changes in case of two transition matrices

In Section 3.3, a time series with 2 different underlying transition matrices was analysed using the hidden Markov model presented in this paper. The counting results obtained from this analysis can be regarded as a Markov chain. As we know at which transitions, the process that drives the structural changes, we are able to estimate a Markov chain with two transition matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, instead of one. We can again derive a probability of structural changes from each transition matrix, according to Equations (D.1) and (D.2):

$$P_1 = \frac{\sum_{g=0}^{G}\pi_g^{[1]}(1-[\mathbf{A}_1]_{gg})}{G+1} \tag{D.1}$$

$$P_2 = \frac{\sum_{g=0}^{G}\pi_g^{[2]}(1-[\mathbf{A}_2]_{gg})}{G+1}. \tag{D.2}$$

In this expression, the weights $\pi_g^{[x]}$, with $x = 1, 2$, are determined by which type of transition occurs from frame

Parameter	Value
T	1 to 40
N	100
G	10
ι_g	Uniform
A_{jg}	Gaussian spread around diagonal with FWHM = 1.5
a	1
$\mathcal{M}_g$	Pt(110)
σ/δ	0.3 and 0.7

Table B.1: Parameters of the hypothetical hidden Markov models used to obtain the results from Figure 2.

Parameter	Value
T	1, 2, 5, 10, 20
N	$1000/T$
G	10
ι_g	Uniform
A_{jg}	Gaussian spread around diagonal with FWHM = 1.5
a	1
$\mathcal{M}_g$	Pt(110)
σ/δ	0.3 and 0.7

Table B.2: Parameters of the hypothetical hidden Markov models used to obtain the results from Figure 3.

to frame:

$$\pi_g^{[x]} = \frac{1}{T_x} \sum_{t \in \mathcal{T}_x} \pi_g^{(t)}, \tag{D.3}$$

with $\mathcal{T}_x$ the set of T_x frames that change to the next frame according to transition matrix $\mathbf{A}_x$, with $x = 1, 2$. In this expression, $\pi_g^{(t)}$ remains the same, as defined by Equation (8).

D.2. Gradually increasing probability of structural changes

Figure D.1 shows the gradually changing transition matrices that are used in the time series analysed in Section 3.3. The FWHM of the Gaussian spread of the diagonal is increased from 1 to 3 in steps of 0.1 throughout the time series, and therefore the probability of structural changes P calculated using Equation (13) is also increased throughout the time series.

Parameter	Value
T	10
N	60
G	6
ι_g	Uniform
A_{jg}	Gaussian spread around diagonal with FWHM = 1.5
a	1
$\mathcal{M}_g$	Pt(110)
σ/δ	0.7

Table B.3: Parameters of the hypothetical hidden Markov model used to obtain the results from Figure 4.

Parameter	Value
T	10
N	2457
G	26
ι_g	Mixing proportions π_g previously estimated from Pt nanoparticle in H_2 environment in [16].
A_{jg}	Alternating transition matrices for $H_2 \rightarrow O_2$ and $O_2 \rightarrow H_2$ transitions, obtained from atom-counts previously quantified in [16].
a	1
$\mathcal{M}_g$	Pt(110)
σ/δ	0.1 and 0.7

Table B.4: Parameters of the hypothetical hidden Markov models used to study the effect of two transition matrices in Section 3.3.

Parameter	Value
T	22
N	100
G	10
ι_g	Uniform
A_{jg}	Gaussian spread around diagonal with FWHM = 1 to 3
a	1
$\mathcal{M}_g$	Pt(110)
σ/δ	0.3 and 0.7

Table B.5: Parameters of the hypothetical hidden Markov models used to study the effect of gradually changing transition matrices in Section 3.3.

References

[1] T. J. Park, G. C. Papaefthymiou, A. J. Viescas, A. R. Moodenbaugh, S. S. Wong, Size-Dependent Magnetic Properties of Single-Crystalline Multiferroic $BiFeO_3$ Nanoparticles, Nano Letters 7 (2007) 766–772.

[2] H. J. Chen, X. S. Kou, Z. Yang, W. H. Ni, J. F. Wang, Shape- and Size-Dependent Refractive Index Sensitivity of Gold Nanoparticles, Langmuir 24 (2008) 5233–5237.

[3] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, R. P. Van Duyne, Biosensing with plasmonic nanosensors, Nature Materials 7 (2008) 442–453.

[4] H. Chang, M. Saito, T. Nagai, Y. Liang, Y. Kawazoe, Z. Wang, H. Wu, K. Kimoto, Y. Ikuhara, Single adatom dynamics at monatomic steps of free-standing few-layer reduced graphene, Sci. Rep. 4 (2015) 6037.

[5] T. Furnival, D. Knez, E. Schmidt, R. K. Leary, G. Kothleitner, F. Hofer, P. D. Bristowe, P. A. Midgley, Adatom dynamics and the surface reconstruction of Si(110) revealed using time-resolved electron microscopy, Applied Physics Letters 113 (2018) 183104.

[6] A. Surrey, D. Pohl, L. Schultz, B. Rellinghaus, Quantitative Measurement of the Surface Self-Diffusion on Au Nano-

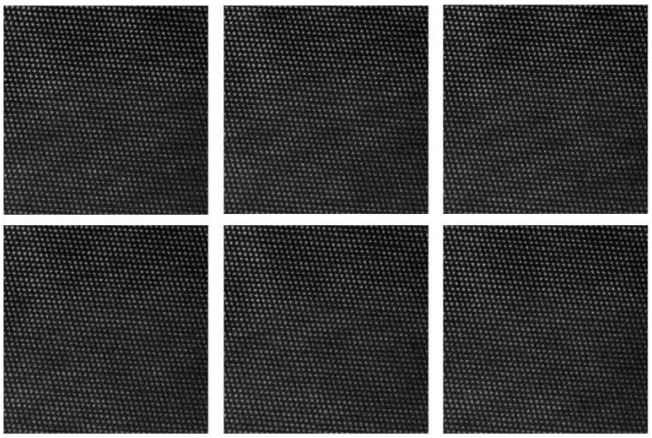

Figure C.1: ADF STEM images of the time series of the Pt wedge discussed in Section 3.2.

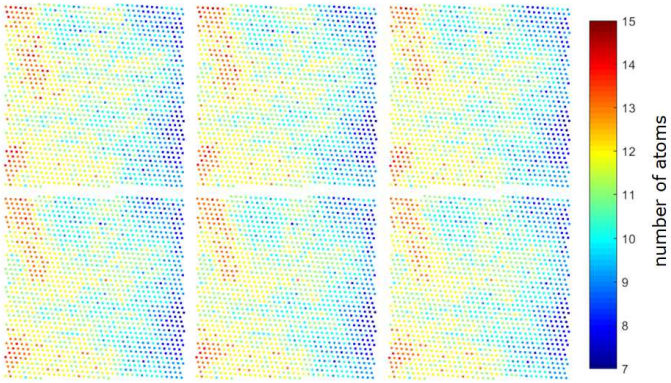

Figure C.2: Counting results obtained by the hidden Markov model analysis of the the time series of Figure C.1.

particles by Aberration-Corrected Transmission Electron Microscopy, Nano Letters 12 (2012) 6071–6077.

[7] S. Schneider, A. Surrey, D. Pohl, L. Schultz, B. Rellinghaus, Atomic surface diffusion on Pt nanoparticles quantified by high-resolution transmission electron microscopy, Micron 63 (2014) 52–56.

[8] V. Cherepanov, B. Voigtländer, Influence of material, surface reconstruction, and strain on diffusion at the Ge(111) surface, Physical Review B 69 (2004) 125331.

[9] P. A. Buffat, Electron diffraction and HRTEM studies of multiply-twinned structures and dynamical events in metal nanoparticles: facts and artefacts, Materials Chemistry and Physics 81 (2003) 368–375.

[10] J. Yu, X. Li, J. Miao, W. Yuan, S. Zhou, B. Zhu, Y. Gao, H. Yang, Z. Zhang, Y. Wang, Atomic Mechanism in Layer-by-Layer Growth via Surface Reconstruction, Nano Letters 19 (2019) 4205–4210.

[11] S. Iijima, T. Ichihashi, Structural Instability of Ultrafine Particles of Metals, Physical Review Letters 56 (1986) 616–619.

[12] P. E. Batson, Motion of Gold Atoms on Carbon in the Aberration-Corrected STEM, Microscopy and Microanalysis 14 (2008) 89–97.

[13] R. F. Egerton, R. McLeod, F. Wang, M. Malac, Basic questions related to electron-induced sputtering in the TEM, Ultramicroscopy 110 (2010) 991–997.

[14] J. Lee, W. Zhou, J. Pennycook, J.-C. Idrobo, S. T. Pantelides, Direct visualization of reversible dynamics in a Si6 cluster embedded in a graphene pore, Nature Communications 4 (2013) 1650.

[15] M. L. Taheri, E. A. Stach, I. Arslan, P. A. Crozier, B. C. Kabius, T. LaGrange, A. M. Minor, S. Takeda, M. Tanase, J. B. Wagner, R. Sharma, Current status and future directions for in situ transmission electron microscopy, Ultramicroscopy 170 (2016) 86–95.

[16] T. Altantzis, I. Lobato, A. De Backer, A. Béché, Y. Zhang, S. Basak, M. Porcu, Q. Xu, A. Sánchez-Iglesias, L. M. Liz-Marzán, G. Van Tendeloo, S. Van Aert, S. Bals, Three-Dimensional Quantification of the Facet Evolution of Pt Nanoparticles in a Variable Gaseous Environment, Nano Letters 19 (2019) 447–481.

[17] A. V. Crewe, J. Wall, L. M. Welter, A High-Resolution Scanning Transmission Electron Microscope, Journal of Applied Physics 39 (1968) 5861–5868.

[18] P. D. Nellist, S. J. Pennycook, The Principles and Interpretation of Annular Dark-Field Z-Contrast Imaging, Advances in Imaging and Electron Physics 113 (2000) 147.

[19] S. J. Pennycook, L. A. Boatner, Chemically sensitive structure-imaging with a scanning transmission electron microscope, Nature 336 (1988) 565–567.

[20] J. M. Cowley, M. S. Hansen, S. Y. Wang, Imaging modes with an annular detector in STEM, Ultramicroscopy 58 (1995) 18–24.

[21] P. Hartel, D. Rose, C. Dinges, Conditions and reasons for incoherent imaging in STEM, Ultramicroscopy 63 (1996) 63–114.

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

[23] L. Jones, K. E. MacArthur, V. T. Fauske, A. T. J. van Helvoort, P. D. Nellist, Rapid estimation of catalyst nanoparticle morphology and atomic-coordination by high-resolution Z-contrast electron microscopy, Nano Letters 14 (2014) 6336–6341.

[24] J. J. Geuchies, C. van Overbeek, W. H. Evers, B. Goris, A. De Backer, G. P. Gantapara, F. T. Rabouw, J. Hilhorst, J. L. Peters, O. Konovalov, A. V. Petukhov, M. Dijkstra, L. D. A. Siebbeles, S. Van Aert, S. Bals, D. Vanmaekelbergh, In situ study of consecutive phase transitions in the formation of atomatomic coherent two-dimensional supperlattices from nanocrystals, Nature Materials 15 (2016) 1248–1254.

[25] M. Yu, A. B. Yankovich, A. Kaczamarowski, D. Morgan, P. M. Voyles, Integrated Computational and Experimental Structure Refinement for Nanoparticles, ACS Nano 10 (2016) 4031–4038.

[26] A. De Backer, L. Jones, I. Lobato, T. Altantzis, B. Goris, P. D. Nellist, S. Bals, S. Van Aert, Three-dimensional atomic models from a single projection using Z-contrast imaging: verification by electron tomography and opportunities, Nanoscale 9 (2017) 8791–8798.

[27] J. M. LeBeau, S. D. Findlay, L. J. Allen, S. Stemmer, Standardless atom counting in scanning transmission electron microscopy, Nano Letters 10 (2010) 4405–4408.

[28] 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.

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

[30] 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.

[31] A. De wael, A. De Backer, L. Jones, P. D. Nellist, S. Van Aert, Hybrid statistics-simulations based method for atom-counting using scanning transmission electron microscopy, Ultramicroscopy 177 (2017) 69–77.

[32] 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.

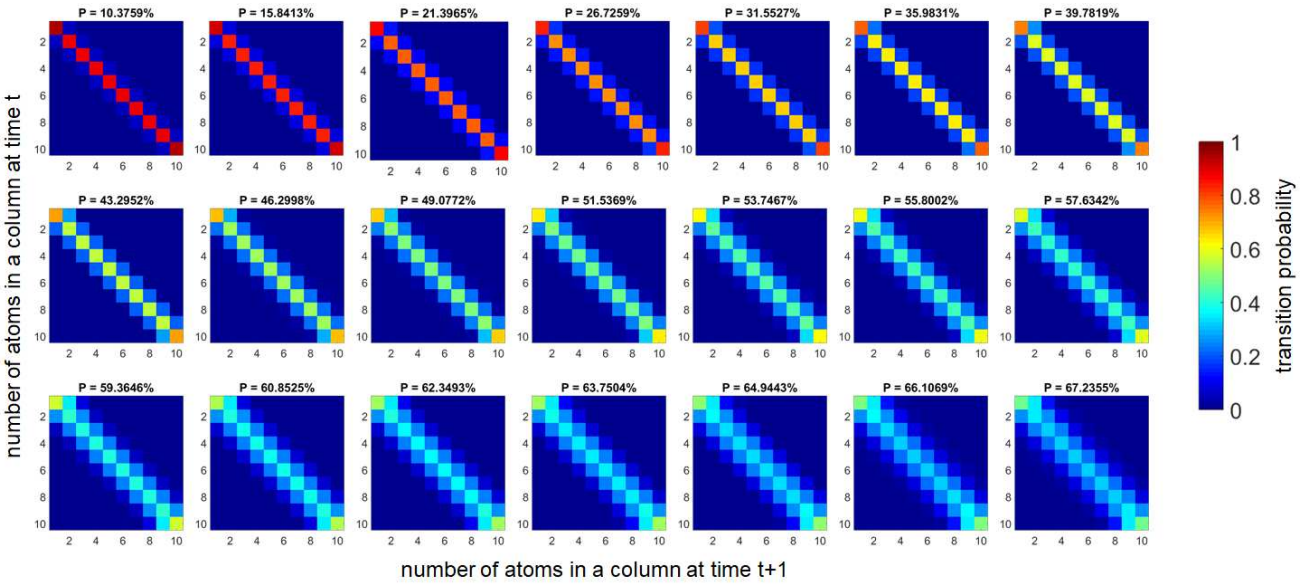

Figure D.1: Transition matrices with a Gaussian spread around the diagonal with an increasing value for the FWHM (1 to 3, in steps of 0.1), and therefore an increasing probability of structural changes P.

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

[34] A. De Backer, K. H. W. van den Bos, W. Van den Broek, J. Sijbers, S. Van Aert, StatSTEM: An efficient approach for accurate and precise model-based quantification of atomic resolution electron microscopy images, Ultramicroscopy 171 (2016) 104–116.

[35] A. De wael, A. De Backer, L. Jones, A. Varambhia, P. D. Nellist, S. Van Aert, Measuring Dynamic Structural Changes of Nanoparticles at the Atomic Scale Using Scanning Transmission Electron Microscopy, Physical Review Letters 124 (2020) 106105.

[36] J. M. LeBeau, S. Stemmer, Experimental Quantification of Annular Dark-Field Images in Scanning Transmission Electron Microscopy, Ultramicroscopy 108 (2008) 1653–1658.

[37] A. Rosenauer, K. Gries, K. Müller, A. Pretorius, M. Schowalter, A. Avramescu, K. Engl, S. Lutgen, Measurement of specimen thickness and composition in $Al_xGa_{1-x}N/GaN$ using high-angle annular dark field images, Ultramicroscopy 109 (2009) 1171–1182.

[38] G. T. Martinez, L. Jones, A. De Backer, A. Béché, J. Verbeeck, S. Van Aert, P. Nellist, Quantitative STEM normalisation: The importance of the electron flux, Ultramicroscopy 159 (2015) 46–58.

[39] F. Krause, M. Schowalter, T. Grieb, K. Müller-Caspary, T. Mehrtens, A. Rosenauer, Effects of instrument imperfections on quantitative scanning transmission electron microscopy, Ultramicroscopy 161 (2016) 146–160.

[40] L. Jones, H. Yang, T. J. Pennycook, M. S. J. Marshall, S. Van Aert, N. D. Browning, M. R. Castell, P. D. Nellist, Smart Align - a new tool for robust non-rigid registration of scanning microscope data, Advanced Structural and Chemical Imaging 1 (2015).

[41] J. Fatermans, A. J. den Dekker, K. Müller-Caspary, I. Lobato, C. M. O'Leary, P. D. Nellist, S. Van Aert, Single Atom Detection from Low Contrast-to-Noise Ratio Electron Microscopy Images, Physical Review Letters 121 (2018) 056101.

[42] J. Fatermans, S. Van Aert, A. J. den Dekker, The maximum a posteriori probability rule for atom column detection from HAADF STEM images, Ultramicroscopy 201 (2019) 81–91.

[43] G. T. Martinez, A. Rosenauer, A. De Backer, J. Verbeeck, S. Van Aert, Quantitative composition determination at the atomic level using model-based high-angle annular dark field scanning transmission electron microscopy, Ultramicroscopy 137 (2014) 12–19.

[44] G. T. Martinez, K. H. W. van den Bos, M. Alania, P. D. Nellist, S. Van Aert, Thickness dependence of scattering cross-sections in quantitative scanning transmission electron microscopy, Ultramicroscopy 187 (2018) 84–92.

[45] Z. Ghahramani, M. I. Jordan, Factorial Hidden Markov Models, Machine Learning 29 (1997) 245–273.

[46] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (1989) 257–286.

[47] A. J. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, IEEE Transactions on Information Theory 13 (1967) 260–269.

[48] G. D. Forney, The Viterbi algorithm, IEEE Transactions on Information Theory 61 (1973) 268–278.

[49] G. McLachlan, D. Peel, Finite Mixture Models, Wiley series in probability and statistics. John Wiley and Sons, inc., 2000.

[50] L. E. Baum, G. R. Sell, Growth functions for transformations of manifolds, Pacific Journal of Mathematics 27 (1968) 211–227.

[51] S. Van Aert, A. De Backer, L. Jones, G. T. Martinez, A. Béché, P. D. Nellist, Control of Knock-On Damage for 3D Atomic Scale Quantification of Nanostructures: Making Every Electron Count in Scanning Transmission Electron Microscopy, Physical Review Letters 122 (2019) 066101.

[52] R. F. Egerton, Beam-Induced Motion of Adatoms in the Transmission Electron Microscope, Microscopy and Microanalysis 19 (2013) 479–486.

[53] T. Halicioglu, G. M. Pound, A calculation of the diffusion energies for adatoms on surfaces of F.C.C. metals, Thin Solid Films 57 (1979) 241–245.

[54] C. Bishop, Pattern Recognition and Machine Learning, New York: Springer-Verlag, 2006.

[55] I. Lobato, D. Van Dyck, An accurate parameterization for the scattering factors, electron densities and electrostatic potentials for neutral atoms that obey all physical constraints, Acta Crystallographica Section A 70 (2014) 636–649.

[56] I. Lobato, S. Van Aert, J. Verbeeck, Progress and new advances

in simulating electron microscopy datasets using MULTEM, Ultramicroscopy 168 (2016) 17–27.