

StatSTEM

User Guide

StatSTEM is freely available under the GNU public license
(GPLv3)

Contents

1	Introduction	3
1.1	Overview	4
1.1.1	Load and save files	5
1.1.2	Select image	6
1.1.3	Closing StatSTEM	6
2	Preparation	7
2.1	Get peak locations	7
2.1.1	Peak-finder routine	7
2.1.2	Import locations from file	7
2.1.3	Get locations from MAP	9
2.2	Add/Remove peaks	9
2.2.1	Type	9
2.2.2	Add and Remove	9
2.2.3	Select and Remove region	9
2.3	Assign column types	10
2.3.1	Projected unit cell	10
2.3.2	Auto assign	11
2.3.3	Add missing types	11
2.3.4	Change type to	11
2.4	Image parameters/actions	11
2.4.1	Pixel size	11
2.4.2	Cut part from image	11
2.4.3	Flip image contrast	12
2.4.4	Replace image background	12
2.4.5	Normalise image	13
3	Model fitting	15
3.1	Standard procedure	15
3.1.1	Background	15
3.1.2	Column width	16
3.1.3	Test for convergence	17
3.1.4	Parallel computing	17
3.2	Model selection (MAP)	17
3.2.1	General info	18
3.2.2	Electron counts	18
3.2.3	Running MAP	19
3.2.4	Options	20
3.2.5	Guidelines	21

4 Analysis	23
4.1 General options	23
4.1.1 Select new model from MAP	23
4.1.2 Show models from MAP	23
4.1.3 Select columns in image	23
4.1.4 Select columns in histogram	23
4.1.5 Select columns on type	24
4.2 Index columns	24
4.2.1 Projected unit cell	24
4.2.2 Start indexing	25
4.3 Atom counting - Statistical	26
4.3.1 Pre-analysis	27
4.3.2 Post-analysis	27
4.3.3 Create 3D model	28
4.4 Atom counting - Simulation	29
4.4.1 Match with simulations	29
4.4.2 Create 3D model	30
4.5 Atom counting - Time series	30
4.5.1 Run HMM	31
4.6 3D model	32
4.6.1 Save model as XYZ	34
4.6.2 Coordination number	34
4.6.3 Save coor number as XYZ	35
4.7 Strain and more	35
4.7.1 Lattice of type	36
4.7.2 Show shift central atom	36
4.7.3 Calculate octahedral tilt	36
4.7.4 Make displacement map	37
4.7.5 Make strain map	38
5 Remarks and suggestions	39
6 References	40
Appendix	42
A Indexing/strain mapping	42

1 Introduction

The StatSTEM software package is aimed on providing a framework for quantifying scanning transmission electron microscopy (STEM) images. For this, StatSTEM makes use of model-based parameter fitting and statistical techniques, providing accurate and precise quantitative measurements. The quantification process proceeds in three steps:

- Preparation (see section 2)
- Model fitting (see section 3)
- Analysis (see section 4).

The StatSTEM software package can be downloaded from the StatSTEM website. The program is started by opening MATLAB and running the main script *StatSTEM.m*, entering the graphical user interface (GUI).

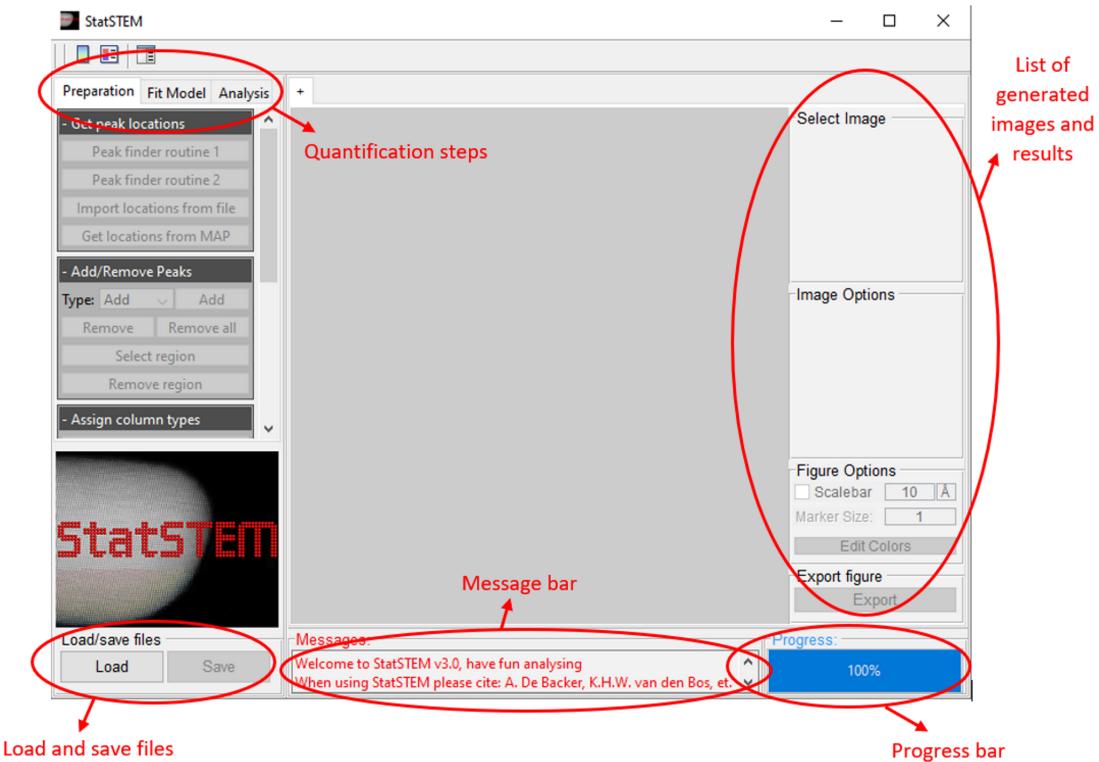


Figure 1.1: Overview of the GUI of StatSTEM, highlighting the main sections.

1.1 Overview

Figure 1.1 shows the main sections of the GUI of StatSTEM. The three steps of the image-quantification process are displayed as separate tab panels in the upper corner on the left-hand side. Each of these panels and their underlying functions are thoroughly discussed further in this manual. The right-hand side of the GUI keeps track of images and results that are generated during the different steps of the quantification process. At the bottom left of the GUI, the user has the possibility to load and store files. Next to this, an information screen is located which contains messages indicating when computations are finished, when changes are made, or when an error occurred. At the bottom right, there is a progress bar indicating the status of the current computation. As computations are often iterative, it is noted that the progress bar is not always a reliable indicator for estimating how long a computation will take.

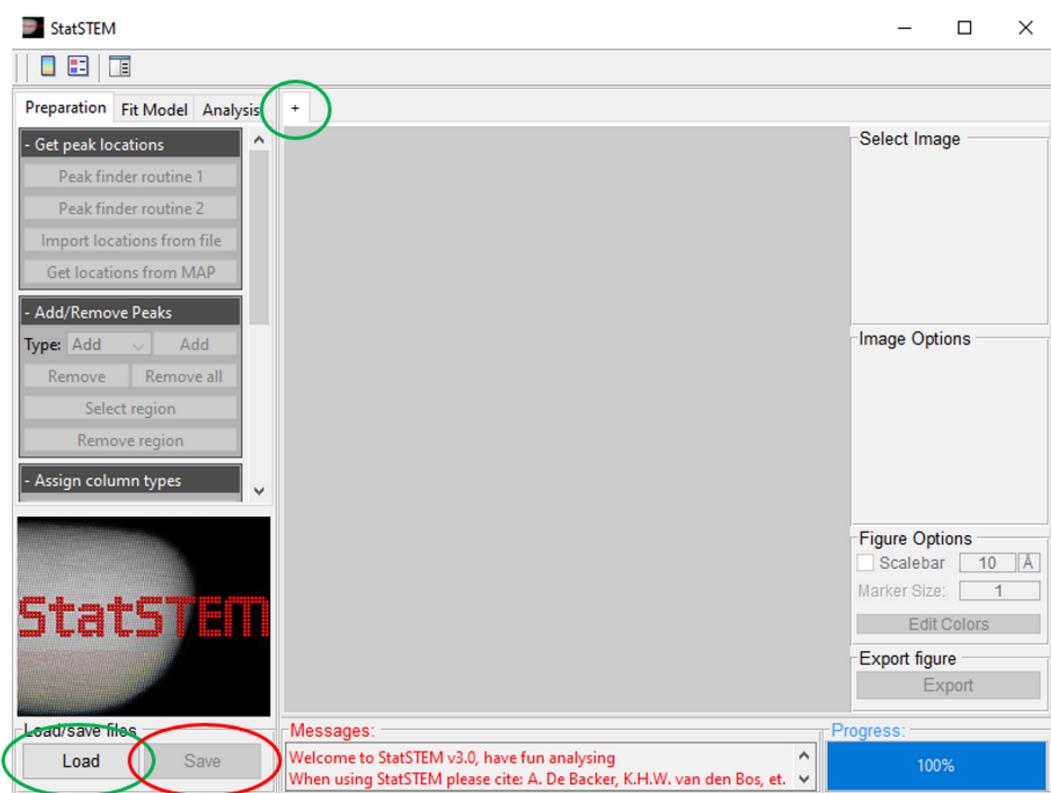


Figure 1.2: Buttons for loading and saving files indicated by the green and red circles, respectively.

1.1.1 Load and save files

The first step of using StatSTEM consists of loading the file you want to investigate into the GUI. You can do this by clicking on the *Load* button in the bottom left corner or on the addition tab in the top, as indicated by the green circles in Figure 1.2. The loaded file may be a single image or a by StatSTEM previously analysed dataset. The preferred file format is that of MATLAB (.mat). When you are satisfied with the analysis of your image data, you can save all the generated results in the MATLAB format by clicking on the *Save* button next to the *Load* button (see red circle in Figure 1.2). In the saved file, all variables are stored as StatSTEM class files. When working with these files, make sure that the StatSTEM folder is always loaded to the path in MATLAB. In the class files, a description

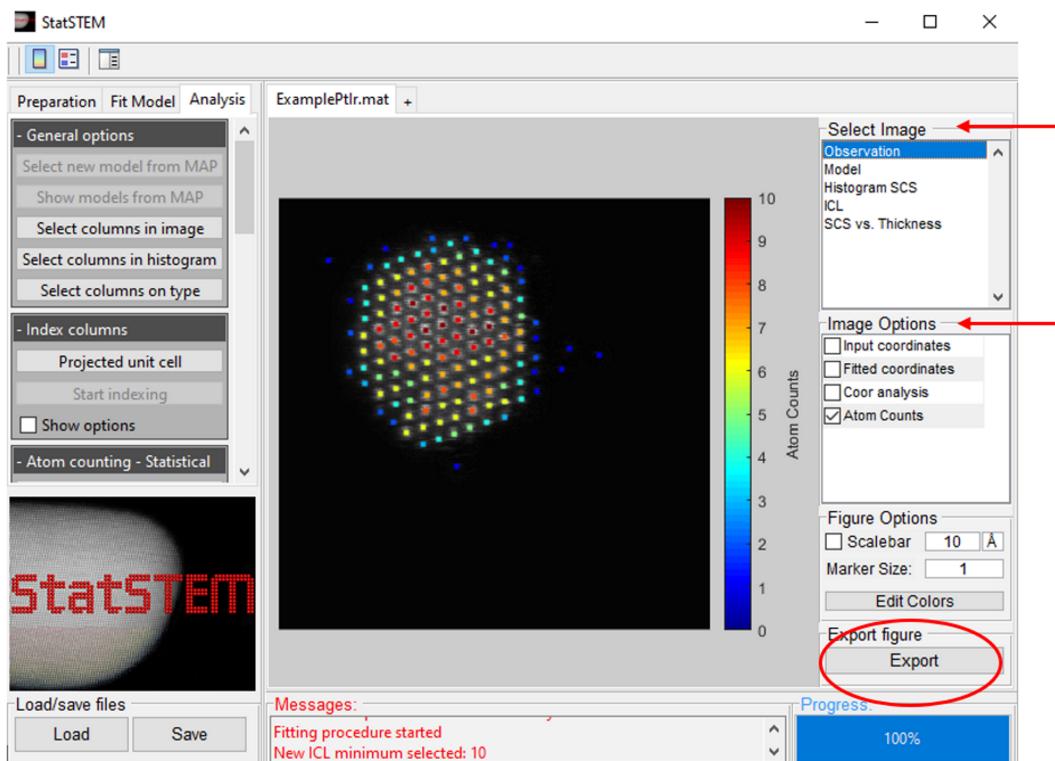


Figure 1.3: On the right-hand side of the GUI of StatSTEM, the desired image or plot can be displayed from the *Select Image* panel, in combination with the indicated desired parameters from *Image Options* (both indicated by the red arrows). The *Export* button, indicated by the red circle, can be used to open the image in a new MATLAB figure.

is given on what information the parameters hold and what their purpose is.

1.1.2 Select image

In StatSTEM, the generated images and parameters during the quantification process can easily be displayed. The *Select Image* panel at the right-hand side allows to select the image or plot of interest. Hereby, the different parameters in *Image Options* below *Select Image* can be ticked off or on. By hitting the *Export* button, the currently displayed image is opened in a new MATLAB figure which can then be saved in the desired format.

1.1.3 Closing StatSTEM

StatSTEM can be closed by hitting the cross in the top-right corner. When closing StatSTEM, make sure that all results are saved (see section 1.1.1), as no warnings of unsaved results are given.

2 Preparation

In StatSTEM, model-based parameter fitting is applied for investigating STEM images. The parametric model that is used consists of a superposition of Gaussian peaks, describing the projected atomic column intensities. For the parameter-optimisation procedure, starting coordinates for the atomic column positions need to be defined. Different options for choosing these initial coordinates adequately are available in the *Preparation* panel.

2.1 Get peak locations

2.1.1 Peak-finder routine

In StatSTEM, two peak-finder routines are available (*Peak-finder routine 1* and *Peak-finder routine 2*) which search for local maxima in the image. When using these peak finders, a new window opens where the parameters of the peak-finder routine can be tuned. These routines work by using filters for smoothing the image. In *Peak-finder routine 1*, there is the option of adding three different filters: an *average*, *disk*, and *gaussian* filter. In *Peak-finder routine 2*, there is no explicit way of altering the filter, but this is implicitly achievable by altering the *Estimated Radius* of the atomic columns. For both peak-finder routines, a threshold value can be defined for removing nuisance pixel intensities from the background. Lastly, *Peak-finder routine 2* offers an extra option, which is defining the *Minimum Distance* between the projected atomic columns in the image. In the peak-finder window, the routine can be tested for a variety of settings. When satisfactory peak locations are found, these coordinates can be exported to StatSTEM by hitting *Use values* or *Confirm values* for *Peak-finder routine 1* or *Peak-finder routine 2*, respectively.

2.1.2 Import locations from file

Atom column coordinates can also be loaded manually into StatSTEM by clicking *Import locations from file*. For this, a MATLAB (.mat) or text (.txt) file needs to be used with the x-coordinates in the first column and the y-coordinates in the second column expressed in Ångström. It is noted that when a previous analysis of StatSTEM is loaded through the *Load* button (see section 1.1.1),

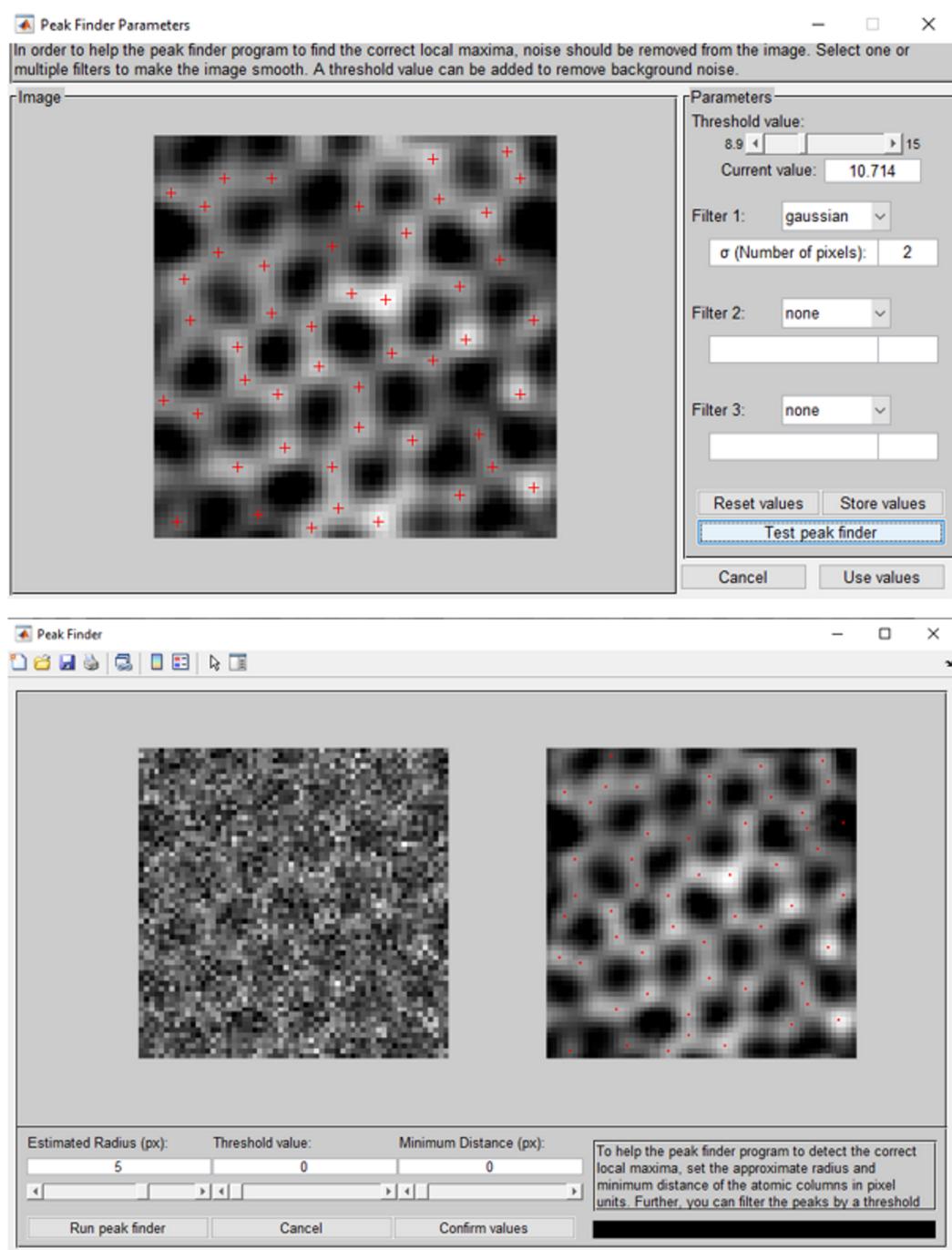


Figure 2.1: Examples of applying *Peak-finder routine 1* (top), and *Peak-finder routine 2* (bottom) to an image of graphene.

also the starting coordinates of this analysis are loaded and can be used as initial coordinates for the new analysis as well.

2.1.3 Get locations from MAP

By pressing the *Get locations from MAP* button, the coordinates of the atomic columns detected by the maximum a posteriori (MAP) probability rule are loaded. This option is only available when the MAP rule has run previously in StatSTEM, or when the loaded StatSTEM file contains an outputStatSTEM_MAP class. More details on the use and working principle of the MAP rule for atom column detection follows in section 3.2.

2.2 Add/Remove peaks

In this panel, different routines are available to define, remove, or change starting coordinates manually.

2.2.1 Type

With this option, it is possible to specify the types of atomic columns that are present in the image. For example, it can be used when the image consists of atomic columns of different elements. In the drop-down menu, you can choose to *Add* or *Remove*, or specify the *Names* of the atom types, which are otherwise labelled as numbers. In section 2.3, further options are discussed for handling different atom types.

2.2.2 Add and Remove

By clicking the *Add* button, you can manually indicate the positions of the atomic columns in the image. By pressing the *escape* key or clicking outside the image, you exit the peak-adding mode. The types of the added peaks need to be specified by the drop-down menu left from the *Add* button. Coordinates labelled by different types are shown in different colours. Similarly, you can also manually remove peaks by clicking the *Remove* button. This allows you to remove individual coordinates from the image. For removing all coordinates, you can simply press *Remove all*.

2.2.3 Select and Remove region

By pressing *Select region*, the user can select a region from the image in which the starting coordinates should be maintained. Outside the selected region, all starting coordinates are removed. In this routine, the user defines the corner points of the selected area one by one. Right clicking connects the last defined point with the starting point. By pressing the *escape* key, the routine is aborted. The button *Remove region* works analogously, but the opposite happens.

Here, a region from the image is selected from which the starting coordinates should be removed. The coordinates outside this region are maintained.

2.3 Assign column types

In this panel, different options are listed to deal with an image where different atomic column types are present.

2.3.1 Projected unit cell

In StatSTEM, an automatic routine is available that can identify the different column types in an image from a given projected unit cell. Here, the relative location of each column, together with the lattice parameters, should be given. You can use the buttons *New* and *Delete* to add and remove an atomic column, respectively. All columns, except the first one, can be removed by using the button *Clear*. The user can also provide information on the depth location of each atom in a column by using the button *z-information*. Its functionality is explained in section 4.2.1. Make sure that in StatSTEM you fill in the correct pixel size of the image (see section 2.4.1), as the lattice parameters in the projected unit cell should be close to the experimental values. It is noted that StatSTEM contains a database with projected unit cells for some common materials and viewing directions.

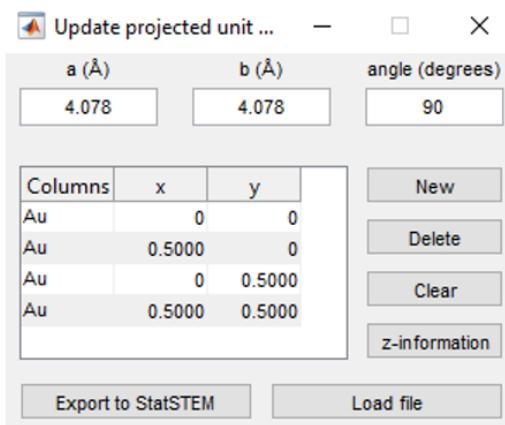


Figure 2.2: The atomic column locations in a projected unit cell of Au, viewed along the [100]-direction.

2.3.2 Auto assign

By using this tool, StatSTEM identifies different column types which are present in the image. In this procedure, lattice directions are first determined by comparing the most central coordinate with its neighbouring coordinates. For this, input on the projected unit cell is required (see section 2.3.1). Then, columns are indexed with respect to the central coordinate. In this manner, the positions of all columns in the projected unit cell are identified and different column types can be assigned.

2.3.3 Add missing types

In this procedure, StatSTEM uses the projected unit cell to find and add the locations of missing column types, following a similar procedure as for *Auto assign*.

2.3.4 Change type to

With this function, a region in the image can be selected where all the contained atom column types are changed to the selected type label, indicated by the drop-down menu to the right of the *Change type to* button. The routine works in the same way as described in section 2.2.3 where the user defines the corner points of the selected area one by one. Right clicking connects the last defined point with the starting point. By pressing the *escape* key, the routine is aborted.

2.4 Image parameters/actions

In this panel, general image parameters can be changed and image operations can be executed.

2.4.1 Pixel size

When an image is loaded to StatSTEM, the pixel size is standard put to 1 Å. This value can easily be changed, which results in a rescaling of the starting coordinates and fitted parameters. ***Note that it is important to provide an accurate value for the image pixel size since the reliability of certain calculations in StatSTEM depends on this value!***

2.4.2 Cut part from image

This option allows one to cut out a rectangular region of interest from the loaded image in order to perform the image quantification on a smaller section. This is done by dragging the left-mouse button over the image, selecting the area of interest. By clicking, the selected

region is confirmed and the image is cropped. By pressing the *escape* key, the cropping process can be cancelled. Note that once the image cropping has been performed, the original image cannot be retrieved. In order to retrieve the original image, it needs to be reloaded in StatSTEM using the *Load* button, as described in section 1.1.1.

2.4.3 Flip image contrast

Typically, atomic columns in annular dark-field (ADF) STEM images are depicted as bright spots on a dark background. By hitting the *Flip contrast image* button, the image contrast can be reversed, meaning that the atomic columns are displayed as dark spots on a bright background. This function can also be used to flip the contrast of annular bright-field (ABF) STEM images.

2.4.4 Replace image background

By using this function, part of the image intensities can be replaced by a certain background intensity. For example, this function may be useful for removing image contributions from neighbouring nanoparticles. When pressing *Replace image background*, the user is asked on how the new background value will be provided, as shown in Figure 2.3. When *Select region in current image* is selected, an area needs to be indicated on the current figure by drawing its corner points one by one. By right clicking, the last corner point is automatically connected to the starting point. The new background value is equal to the average of the pixel intensities within the drawn region. Then, the region of the image to which the new background value should be applied should be drawn in a similar way. The routine can be aborted by pressing the *escape* key. When selecting *Select region in other image*, the procedure works in the same way as described before, but this time the new background value is selected

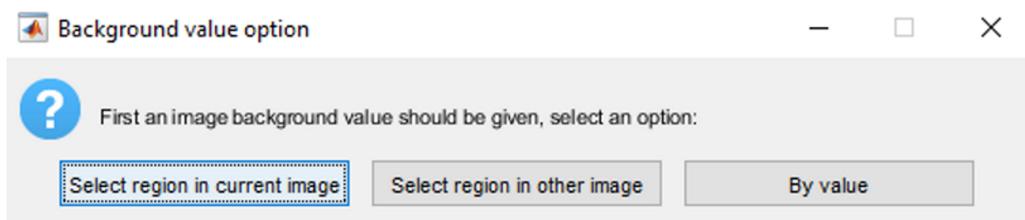


Figure 2.3: Question popping up on how to replace the background value after hitting *Replace image background*.

by indicating an area on another image. Lastly, by clicking *By value*, the new background value needs to be provided as a number. Similarly as described in section 2.4.2 for cropping the image, once the background has been replaced by another value, the original image can only be retrieved by reloading it through the *Load* button at the bottom left of the GUI.

2.4.5 Normalise image

Normalising STEM image data is important for directly comparing experimental images with simulated ones. Typically, image normalisation is achieved by a detector scan from which the averaged image intensity in vacuum (I_{vac}), corresponding to the area next to the detector, and the averaged detector intensity (I_{det}) can be measured. A normalised image (I_{norm}) from the raw image (I_{raw}) is then calculated by the following equation:

$$I_{norm} = \frac{I_{raw} - I_{vac}}{I_{det} - I_{vac}}. \quad (2.1)$$

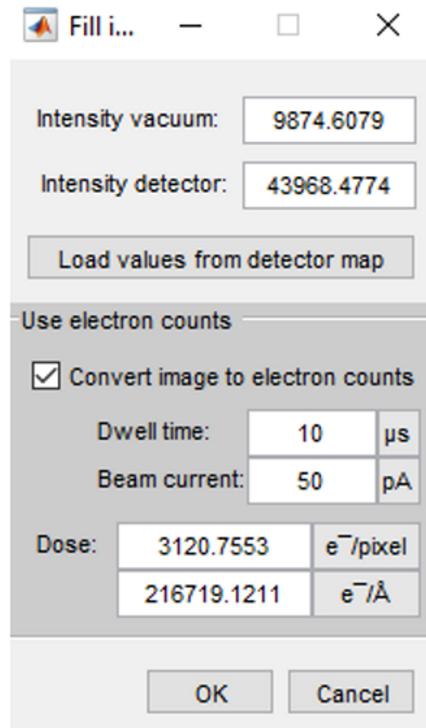


Figure 2.4: Example showing typical values to normalise an image. Standard I_{vac} and I_{det} are put to 0 and 1, respectively.

It is important to stress that the raw image data and the detector scan should be recorded under the same imaging conditions. More information on this topic can be found in Ref. [1]. When clicking the *Normalise image* button, the user can change the *Intensity vacuum* (I_{vac}) and *Intensity detector* (I_{det}) values. Alternatively, these values can be directly loaded from a detector scan by pressing *Load values from detector map*. When the detector file consists of multiple detector scans, StatSTEM opens a window where it can be indicated which detector scan should be used or which scans should be averaged. When proper normalisation values have been chosen, the normalisation is performed when *Ok* is selected. If the box *Convert image to electron counts* is ticked on, the user can specify the pixel dwell time (μs) and beam current (pA) which allow to calculate the incident electron dose. In this case, pressing *Ok* converts the raw image into electron counts. This conversion is necessary for correctly applying the maximum a posteriori (MAP) probability rule for atom column detection, which is further explained in section 3.2.

3 Model fitting

The *Fit Model* panel allows to model the loaded image by a superposition of Gaussian peaks, describing the projected atomic columns. StatSTEM offers to do this in two ways. There is the so-called standard procedure, where the model-fitting routine is performed based on the starting atom column coordinates provided in the *Preparation* panel. Another, more general, approach is also available. This method starts from the provided initial atom column coordinates and automatically searches for more possible column locations. As such, the model-fitting procedure can be performed without a priori specifying where the atomic columns are expected to be located. This method can be especially useful for quantifying images exhibiting low signal-to-noise ratio (SNR) and low contrast.

3.1 Standard procedure

The standard way of modelling each atomic column in the image as a Gaussian peak is to optimise the parameters of the provided atomic columns indicated in the *Preparation* panel. The parameters of atomic columns that are not initially indicated cannot be optimised by the standard procedure. In order to start the fitting procedure, at least one starting coordinate needs to be defined. Detailed information about the fitting routine can be found in Ref. [2]. By hitting the *Run fitting routine* button, the fitting process is initiated. When the routine is done, the optimised model is shown (see Figure 3.1). The fitting procedure can be aborted by hitting the *Abort fitting routine* button. When the procedure is aborted, StatSTEM shows a message in the message bar at the bottom of the GUI. Along with fitting the model, also the total intensity of electrons scattered by each atomic column, the so-called scattering cross-section, is calculated. These scattering cross-sections can be visualised by a histogram, available at the right-hand side in *Select Image* (see Figure 3.2). The fitting routine can be adjusted by changing its options which are available by ticking the *Show options* box. In the following sections, the different options are discussed.

3.1.1 Background

Here, the user can choose to fit a constant background by ticking the *Fit background* box. If the background is chosen not to be fitted,

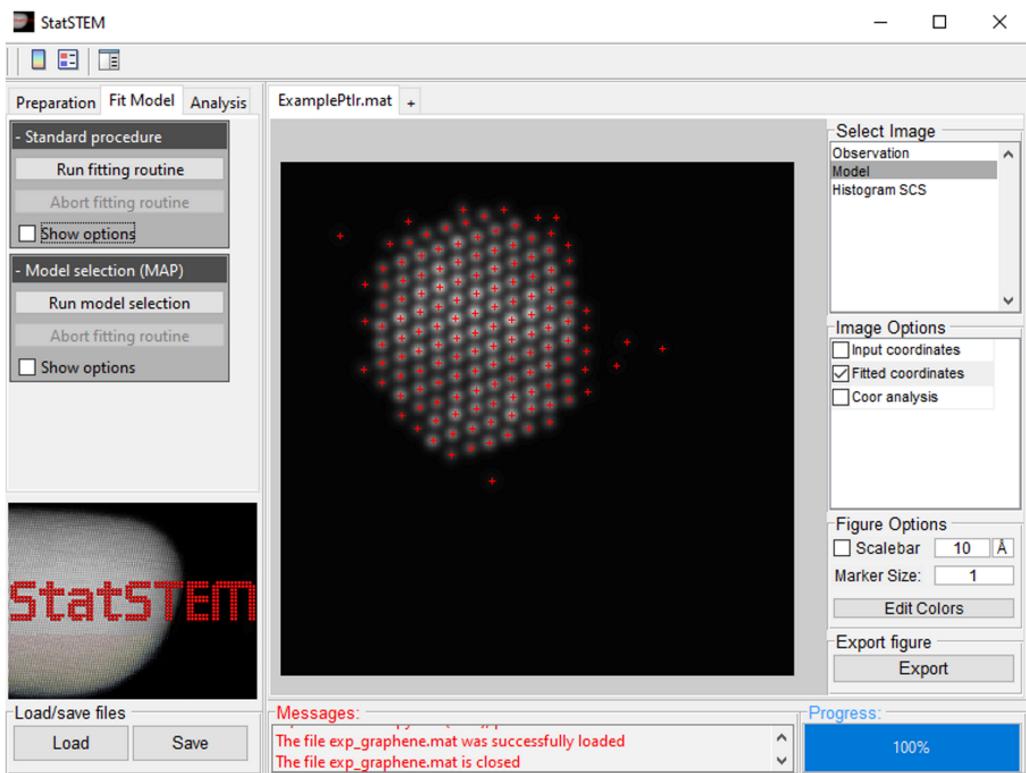


Figure 3.1: The optimised model of an experimental high-angle ADF (HAADF) STEM image of a Pt/Ir sample, overlain with the fitted atomic column coordinates.

a constant background value should be given. This value can be directly entered or obtained from the average pixel values of a region of the image that can be indicated by using the *Select* button.

3.1.2 Column width

In the fitting procedure, there is the possibility to fit the Gaussian peaks to the atomic columns with equal or different widths. In the *Same width* option, the estimated Gaussian peaks have the same width for columns of the same atom type, as labelled in the *Preparation* panel discussed in section 2. In the *Different width* option, a different width is estimated for each Gaussian peak. In the *User defined* option, the user can define a value for the width of each column type (Å). By default, the *Same width* option is used as this is computationally less demanding as compared to fitting all the peaks with a different width and provides reliable results.

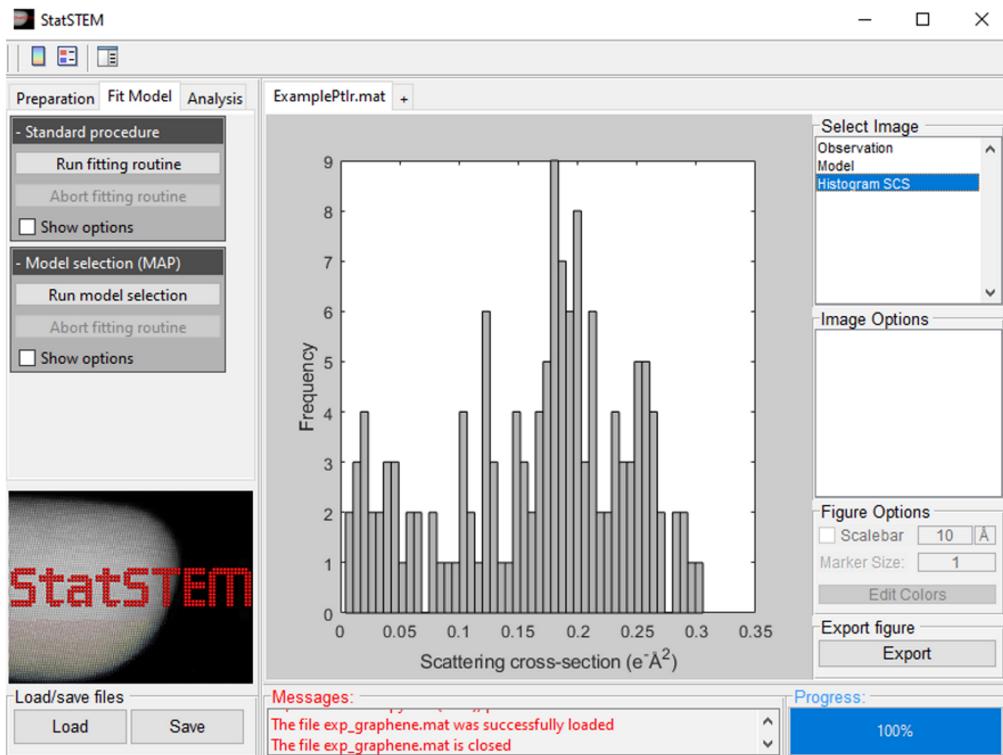


Figure 3.2: Histogram of scattering cross-sections of the atomic columns in an experimental HAADF STEM image of a Pt/Ir sample.

3.1.3 Test for convergence

This option may be used for quickly testing the correctness of the starting coordinates and fitting parameters. In this case, the number of iterations is limited to 4. After a test is done, the newly obtained coordinates may be preferred to use as starting coordinates. By hitting the *Re-use fitted coordinates* button, the quickly fitted coordinates can be used as new starting positions for the atomic columns.

3.1.4 Parallel computing

For improving computational speed, the fitting procedure uses parallel computing in which the calculations are divided over the different CPU cores of the computer. The *Number of CPU cores* used for parallel computing may be reduced to lower the CPU usage during the fitting procedure. Be aware that then total calculation time possibly increases.

3.2 Model selection (MAP)

Here, a more general way of modelling electron microscopy images, as compared to the standard procedure described in section 3.1, is

offered. This method makes use of the concept of model selection in combination with a Bayesian framework for detecting the number of atomic columns present in the image data. For this, use is made of the so-called maximum a posteriori (MAP) probability rule which has been proposed for single atom detection [3]. The routine searches for possible column locations and compares the probabilities of candidate models to each other. As such, the most probable number of atomic columns and corresponding parametric model can be automatically derived from the available image data. More details on the working principles of this methodology can be found in Ref. [4].

3.2.1 General info

Similarly as for the standard procedure of model fitting mentioned in section 3.1, the parametric model consists of Gaussian peaks, describing the projected atomic columns, superposed on a constant background. Here, there is no option though to change the values of the background or column width. By default, a constant background and equal column widths are fitted. Best results are obtained when the background in the image can be adequately modelled as a constant. When this is not the case, consider cropping the image (see section 2.4.2) so that the background can be sufficiently described to be constant. As opposed to the standard procedure, it is optional for the model-selection method to have starting coordinates provided in the *Preparation* panel. When no initial coordinates are given, the routine starts from fitting zero atomic columns (so only background) and continues to search for more columns from there. When input has been given, the routine starts from the provided input coordinates. Note that, in order to apply the model-selection method, there is no use in providing different column types for the input coordinates (see section 2.3), since in the detection mechanism no differentiation can be made between different column types.

3.2.2 Electron counts

For correctly applying atom column detection by MAP probability, it is crucial that the image data is properly converted to electron counts. As mentioned earlier in section 2.4.5, StatSTEM offers this possibility when beam current and pixel dwell time are known and when the image can be normalised by using a detector scan. There are also alternative ways for converting image data to electron counts [1]. When images are recorded using direct-electron detectors, con-

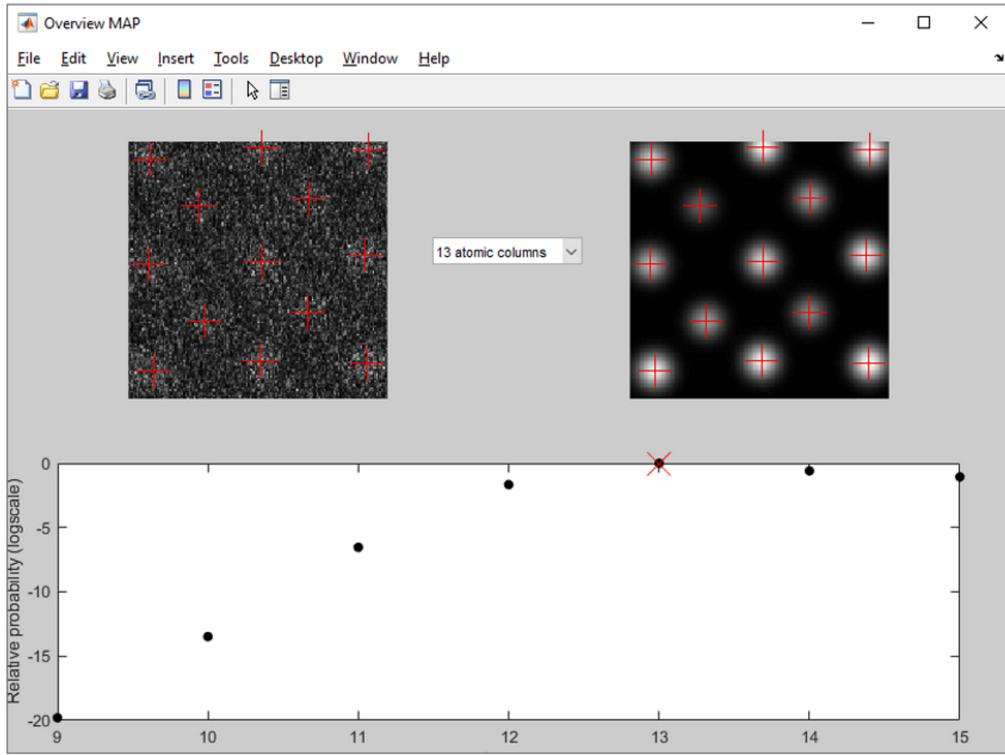


Figure 3.3: Extra screen opening when MAP routine is finished and showing an overview of the fitted number of atomic columns and optimised models for HAADF STEM image data of SrTiO_3 .

version to electron counts might not be necessary anymore, as the image is directly recorded in this format. ***Always be cautious whether conversion to electron counts is still needed when applying the MAP methodology!*** In case of doubt that the electron conversion was done correctly, you can apply Poisson noise to the fitted model. This should lead to a similar looking image as the original image with a similar range of pixel intensities.

3.2.3 Running MAP

For optimal results, the model-selection (or atom detection) routine should be applied to images exhibiting low SNR and low contrast, and thus to low contrast-to-noise ratio (CNR) images. When there are regions in the image where atomic columns are easily recognisable (for example the middle section of a nanoparticle), it is better to avoid these areas and focus on the more challenging sections only (for example the border region of a nanoparticle). Cropping images can be done directly in StatSTEM, as explained in section 2.4.2.

By hitting the *Run model selection* button, the generalised fitting routine is started. The routine can be aborted by pressing *Abort fitting routine*, similarly as for the standard procedure. While the calculations are being performed, progress can be followed through the progress bar at the bottom right of the GUI and through the updating of the relative probability curve and detected atom column coordinates shown on the raw image data and the optimised model. When the procedure is finished, the user is asked to indicate the desired number of columns (and associated optimised model) from the relative probability curve for further analysis in StatSTEM by using the cursor. For this, the user can rely on an extra screen (*Overview MAP*) that opens up when the calculation is finished (see Figure 3.3). Here, an overview is given of the different models and corresponding number of atomic columns that have been fitted during the procedure. By default, the most probable number of atomic columns is initially highlighted. It is noted that the extra screen is merely a helping tool for deciding what number of atomic columns should be chosen. Closing this screen does not affect StatSTEM. Once the desired number of atomic columns is chosen by using the cursor from the relative probability curve in the main screen of StatSTEM, the extra screen is automatically closed and an extra plot option in *Select Image* at the right-hand side appears for visualising the probability curve as a function of the number of atomic columns with the chosen number of columns highlighted (see Figure 3.4). If one is not satisfied with the selected number of atomic columns, another number can be chosen through the *Analysis* panel. More information on this follows in section 4.1.1.

3.2.4 Options

Several settings are available when running the model-selection procedure by the MAP probability rule. The most important option is the *Max. # columns* setting. When this is specified to be 0 or any number smaller than or equal to the number of provided starting coordinates, specified in the *Preparation* panel (see sections 2.1 and 2.2), the model-selection method continues until the probabilities of newly added column coordinates drop. Thus, in this case, the routine is automatically stopped when maximum probability is reached. Since it is unknown when the routine will terminate exactly, the progress bar may fluctuate. On the other hand, when the *Max. # columns* is put to a number greater than the number of

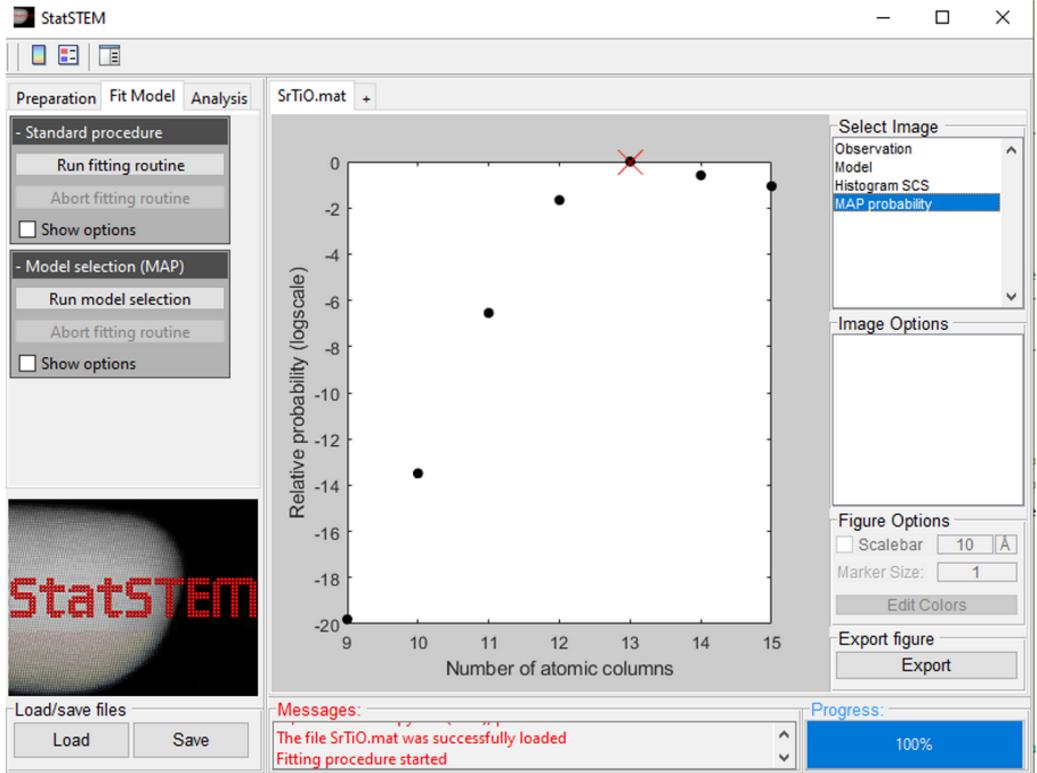


Figure 3.4: Relative probability curve and indicated number of atomic columns by using MAP probability for an experimental HAADF STEM image of a SrTiO_3 sample.

initial coordinates, the calculation runs until the provided number of columns. The remaining options are more advanced settings and provide the ranges between which the parameters of the model (background, column width, column intensities and column coordinates) are expected to lie a priori. More information on this can be found in Ref. [4]. By default, the ranges for the background and column intensities are defined between 0 and the maximum pixel intensity, whereas the values for the column width and coordinates are defined within the field of view of the image. Note that it is not possible to define individual ranges on column width, intensity and position for different atomic columns.

3.2.5 Guidelines

Sometimes, running MAP under the default options may not be optimal which can lead to unrealistic model fits. For stabilising the fitting routine in the MAP procedure, consider providing more physical values for *Min. width* and *Max. width* related to a realistic range

for the typical width of an atomic column in the image. Also, when considering small nanoparticles on a background, it is general good practice to provide a value for *Min. intensity* which indicates the minimum expected intensity of a single atom in the image. If this value is not specified, MAP can become sensitive to small variations in the background which leads to overdetection and the indication of possible atomic columns which are physically unrealistic. A value for *Min. intensity* can, for example, be determined from an image simulation of a single atom. Typically, the ranges related to the positions of the atomic columns should not be changed. In addition, MAP should not be applied in a situation where background has been altered or removed from an image (see section 2.4.4). The MAP procedure is designed to recognise atomic columns from noise and removing background hinders its judgement. For fitting with the standard procedure (see section 3.1), background removal can be applied. Lastly, it should be stressed that the use of the MAP methodology in StatSTEM should be applied to images containing a limited number of atomic columns (<100). ***Apply MAP only on images of small nanoparticles or images of bulk materials with a limited field of view!***

4 Analysis

In this panel, several procedures are available for using the fitted model parameters for further analysis, such as atom counting, column coordinate analysis, and creating (simple) three-dimensional (3D) structural models.

4.1 General options

4.1.1 Select new model from MAP

When the model-selection method by maximum a posteriori (MAP) probability has been applied previously, one has the option to choose a different model and corresponding number of atomic columns for further analysis by pressing *Select new model from MAP*. A new model can be selected by using the cursor for indicating the desired number of atomic columns from the relative probability curve. More details can be found in section 3.2.3.

4.1.2 Show models from MAP

This option allows to open an extra screen (*Overview MAP*), showing an overview of the optimised models and corresponding number of atomic columns. This screen also opens automatically when pressing the *Select new model from MAP* button, as discussed in the previous section 4.1.1. This overview screen serves as a helping tool for deciding what model should be chosen for further analysis.

4.1.3 Select columns in image

By hitting the *Select columns in image* button, one can indicate a region of the image to take into account for further analysis. Columns located outside this region are neglected. The user defines the corner points of the selected area one by one. Right clicking connects the last defined point with the starting point. By pressing the *escape* key, the routine is aborted.

4.1.4 Select columns in histogram

This option allows to exclude outliers in the histogram of scattering cross-sections. First, the lower limit must be defined, then the upper limit. The scattering cross-sections of the columns that fall outside the limit are neglected for further analysis. By pressing the *escape* key, the routine can be aborted.

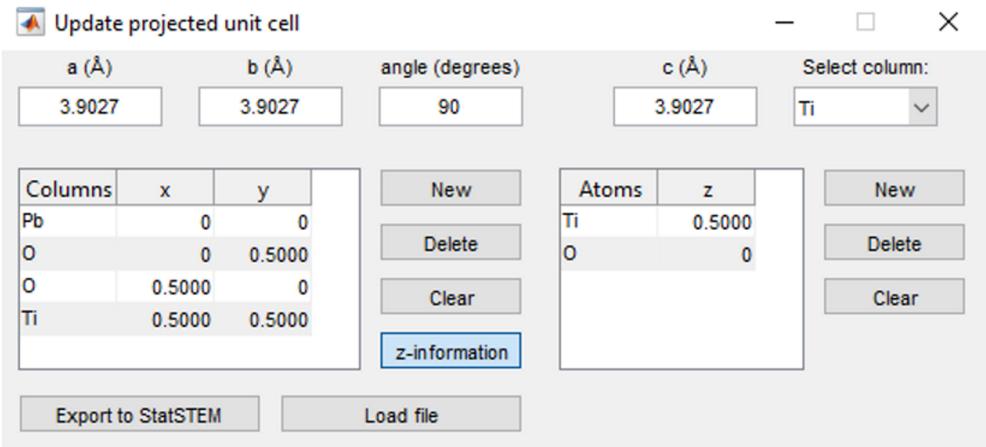


Figure 4.1: A projected unit cell can be made by defining the projected lattice constants (a and b) and the relative positions of the columns. Additional information can be given on the depth location (z) of the atoms in the columns.

4.1.5 Select columns on type

By pressing *Select columns on type*, columns can be selected for further analysis based on their column type labels, given in the *Preparation* panel (see section 2).

4.2 Index columns

Here, one can index the columns in terms of distance from a reference point. This operation should be done before performing a column coordinate based analysis (see section 4.7) or creating a 3D model (see sections 4.3 and 4.4).

4.2.1 Projected unit cell

A necessary input to index columns is the location of the different column positions in a projected unit cell, mentioned in section 2.3.1. If one intends to make a 3D model later on, z -information must be given, as shown in Figure 4.1. Then, it becomes possible to define the lattice parameter in c -direction and the depth locations of the atoms in each column. Atoms in a column can be added and removed by the buttons *New* and *Delete*, respectively. All atoms, except the first one, can be deleted by using the *Clear* button. Each column can be selected from the drop-down menu in the top-right corner. Make sure that in StatSTEM you provide the correct pixel size of the

image (see section 2.4.1) as the lattice parameters in the projected unit cell should be close to the experimental values. It is noted that a database is available in StatSTEM for some common materials and viewing directions.

4.2.2 Start indexing

With *Start indexing*, all coordinates are indexed as a function of distance in unit cells from a reference coordinate. For this procedure, a reference coordinate must first be chosen. Then, lattice directions are searched and the indexing procedure starts. If the automatic routines in StatSTEM fail, you can guide StatSTEM by using the advanced options listed below. The details of the automatic routine are described in Appendix A.

Reference coordinate

With this option, a reference coordinate can be chosen and a displacement map is made from this coordinate. Furthermore, the reference coordinate gets the index (0,0) during the creation of the strain map. One can choose between different column types for selecting a reference coordinate. By default, StatSTEM uses the *Most central* coordinate as a reference, this can be changed by manually selecting another coordinate by indicating *User defined*.

Direction a lattice

For finding the direction of the lattice, an *Automatic* routine can be used or a manual (*User defined*) input can be given. The automatic routine uses the projected unit cell parameters to identify the lattice direction. Here, the distance of the neighbouring coordinates with respect to the reference coordinate is compared to the given lattice parameters in the projected unit cell. Once the direction is found, the lattice parameters are automatically improved by fitting. This option can be disabled by ticking off *Improve values by fitting*. When this setting is enabled, a box of N unit cells (standard 3 unit cells (UCs) to each side) around the reference coordinate is used for finding the values of the a (and b) lattice parameter in the image. This option is advised to be used, as the pixel size recorded by an electron microscope is not always very accurate. Be, however, aware that this option changes the values of a and b and should only be turned off when you are 100 % sure about the pixel size.

4.3 Atom counting - Statistical

In HAADF STEM imaging, a statistics-based method has been developed to count the number of atoms based on the scattering cross-sections (the total intensities of electrons scattered by the atomic columns), which increase monotonically with thickness [5]. ***Be aware that this method is only reliable when only one column type is present!*** In the statistics-based method, the scattering cross-sections are presented in a histogram. Owing to a combination of experimental noise and residual instabilities, broadened - rather than discrete - components are observed in such a histogram. Therefore, these results cannot directly be interpreted in terms of number of atoms. By evaluation of the so-called integration classification likelihood (ICL) criterion in combination with Gaussian-mixture model estimation, the number of components and their locations can be found. From the estimated locations of the components, the number of atoms in the columns can be quantified. More information on this methodology can be found in Refs. [6, 7, 8].

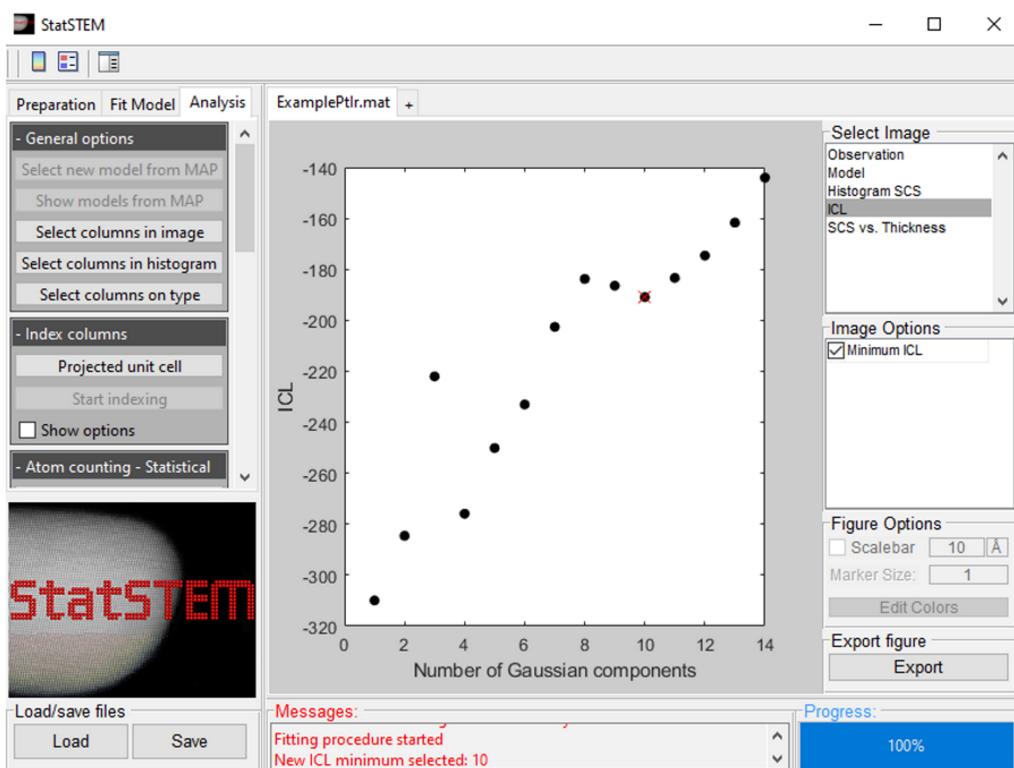


Figure 4.2: The ICL criterion evaluated for an experimental image. A local minimum appears at 10 Gaussian components.

4.3.1 Pre-analysis

In order to evaluate the ICL criterion, an upper limit on the number of components must be given, which can be specified by *Max components*. Up to the provided number of components, Gaussian mixture models are fitted to the histogram of scattering cross-sections and the ICL criterion is determined. A rough estimate for this upper limit can be obtained by using the shape of the particle under study.

4.3.2 Post-analysis

In StatSTEM, the statistical atom-counting analysis can be performed by hitting *Run ICL*. When the Gaussian-mixture model has been fitted to the histogram of the scattering cross-sections and the ICL calculation has been performed for each number of components, a suitable number of components needs to be selected by using the cursor. In the statistics-based atom-counting procedure, one searches for local minima in the ICL curve (see Figure 4.2). After a local minimum is selected, the experimental image with the correspond-

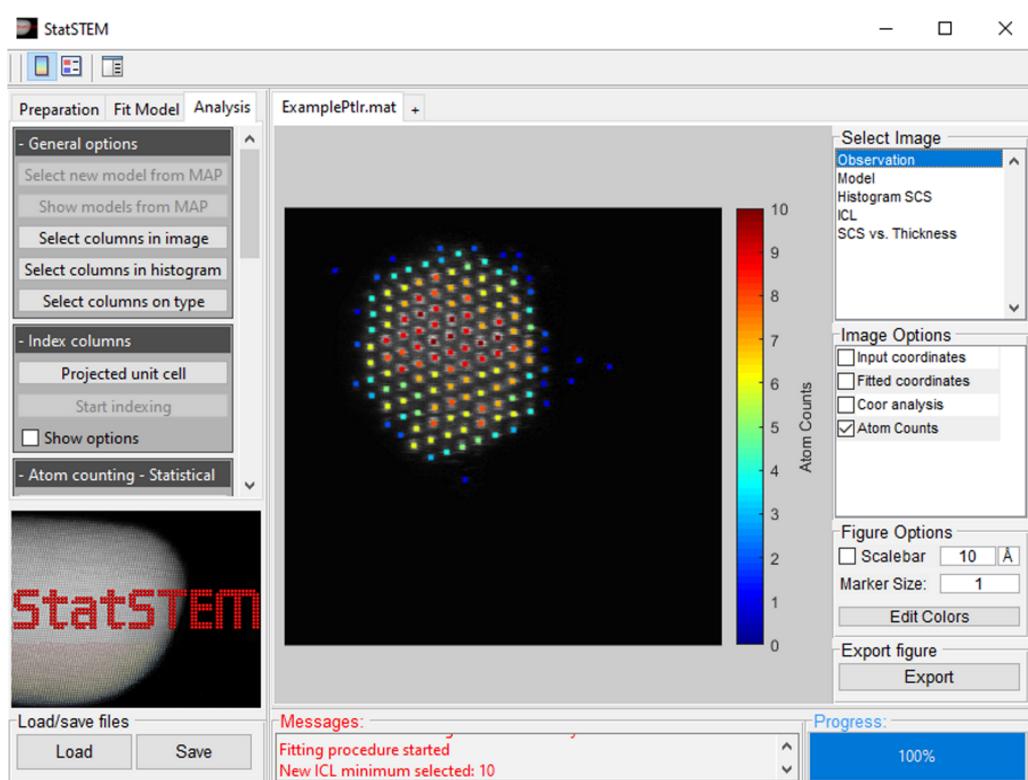


Figure 4.3: The atom counts of the atomic columns determined from an experimental HAADF STEM image of a Pt/Ir particle.

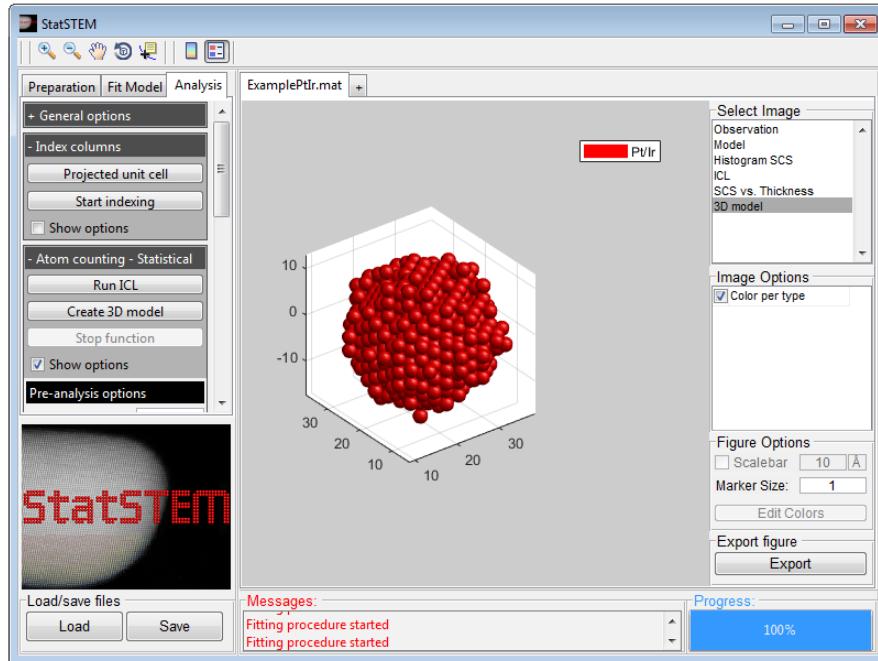


Figure 4.4: A 3D model of a Pt/Ir particle.

ing atom counts is shown (see Figure 4.3). The calculation can be aborted by pressing *Stop function*. Once the procedure is aborted, StatSTEM shows a message asking whether the user wants to select a local minimum in the current ICL graph. Counting results can be rescaled by providing a *Counting offset*. This is particularly useful for thick particles in which no thin columns are present. Another local minimum from the ICL curve can be selected by pressing the *Select new ICL minimum* button.

4.3.3 Create 3D model

When the atoms are counted per atomic column and the columns are indexed (see section 4.2), a 3D model can be made. Here, the atoms are distributed symmetrically along the z-direction. ***Note that this 3D model is only a simple model for visualising the 3D shape of the particle. It should not be used as a final result!*** Note that for this procedure the projected unit cell should contain z-information (see section 4.2.1). The 3D model that is shown is colour coded, meaning that each atom type is displayed in a different colour.

4.4 Atom counting - Simulation

Another way of performing atom counting is by directly comparing libraries of simulated scattering cross-sections with experimental cross-sections.

4.4.1 Match with simulations

By clicking on this button, a library containing simulated scattering cross-sections is asked to be loaded. The atom counts in the image under investigation are computed by comparing the measured cross-sections from the fitted model to the loaded library of simulated cross-sectional values. A MATLAB (.mat) or text (.txt) file can be loaded containing the simulated values of the scattering cross-sections. The cross-sections must be provided in function of column thickness, expressed as a column vector. The atom counts following from the simulation-based method can be visualised by indicating *Lib Counts* in the *Image Options* panel at the right-hand side of the GUI, as opposed to selecting *Atom Counts* for the statistics-based

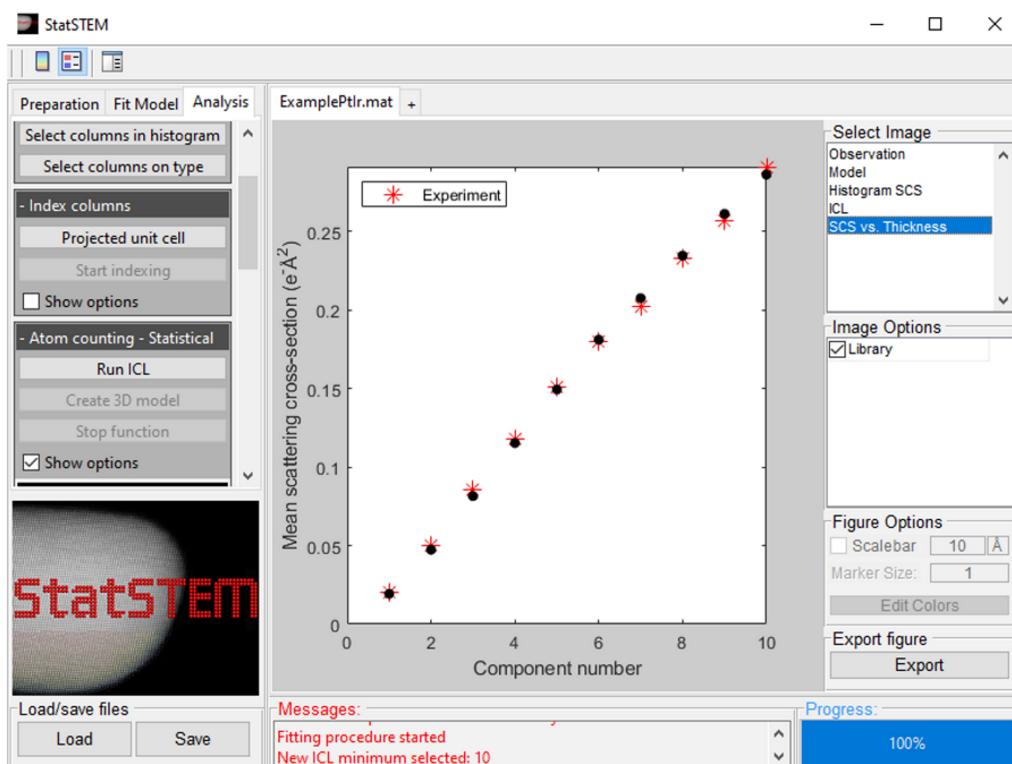


Figure 4.5: Comparison between the scattering cross-sections obtained by the statistics-based method (Experiment) and image simulations (Library).

method. By selecting *SCS vs. Thickness* from the *Select Image* panel, the scattering cross-sections obtained by the statistics-based method are compared to the simulated cross-sectional values (see Figure 4.5).

4.4.2 Create 3D model

This button has the exact same functionality as explained before in section 4.3.3 and is used for quickly building a 3D visualisation of the material structure based on the atom counts and indexed columns.

4.5 Atom counting - Time series

When the goal is to perform atom counting from a time series of ADF STEM images, this functionality can be used. The methodology makes use of a hidden Markov model (HMM) to determine the number of atoms in each time frame by explicitly taking into account the possibility of structural changes. More information on this methodology can be found in Refs. [9] and [10]. At the moment, preprocessing still happens outside StatSTEM, but in a later release this will be fully included. The required format is an inputStat-

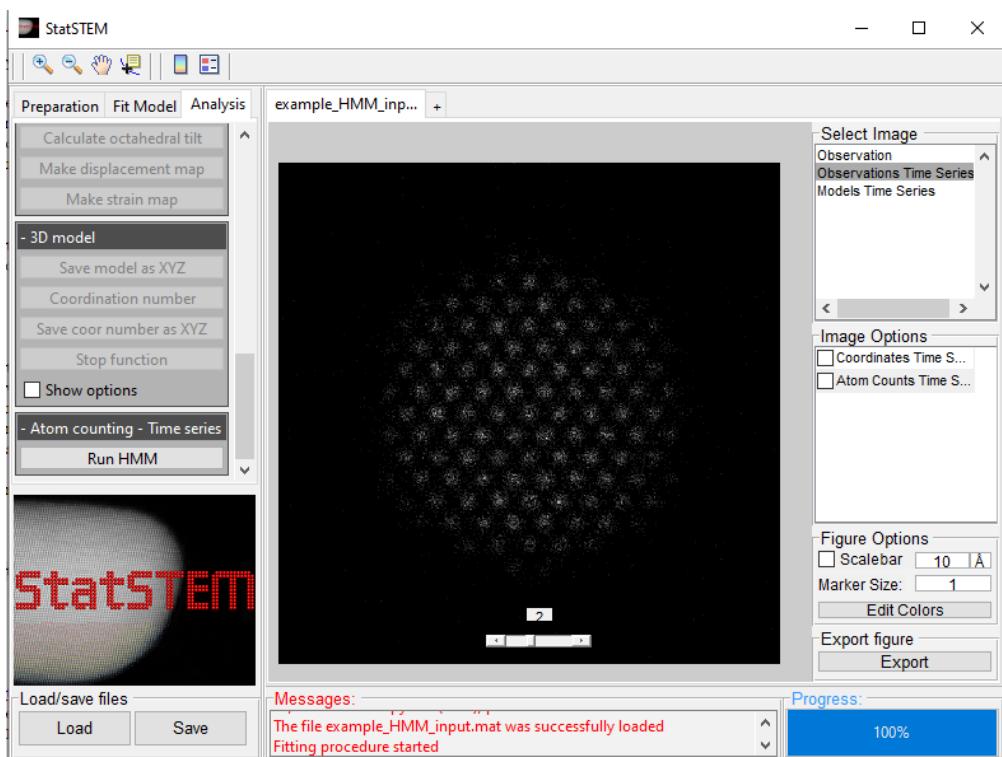


Figure 4.6: An example of a time series of ADF STEM images of a Pt particle.

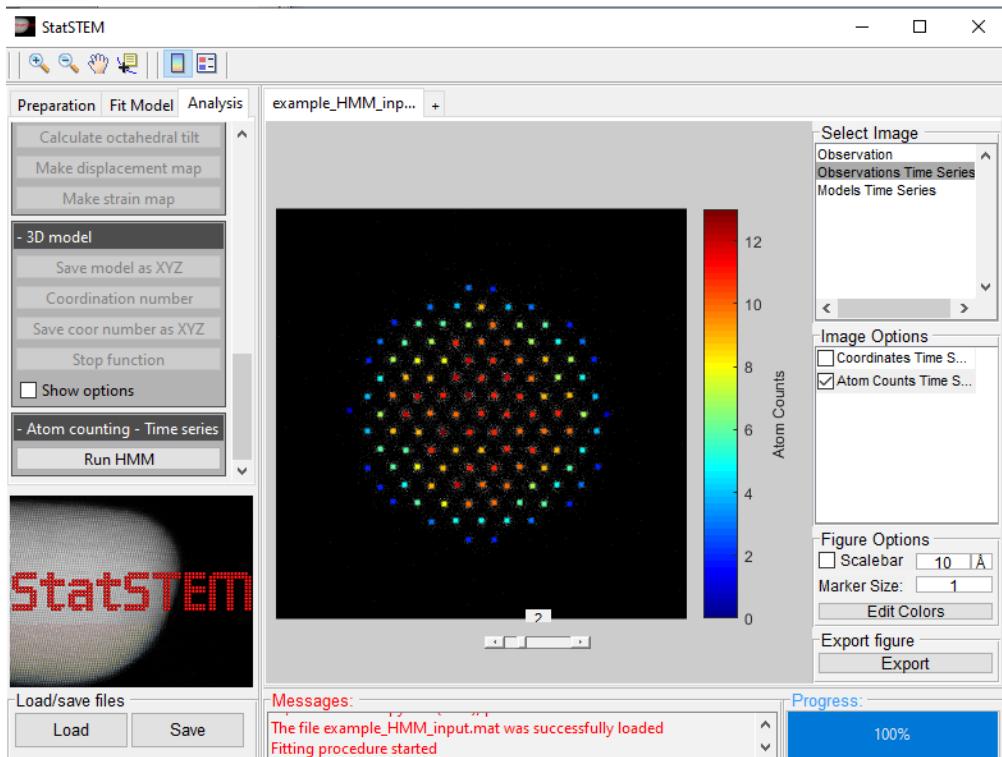


Figure 4.7: Atom counts for a time series of ADF STEM images of a Pt particle obtained by using the hidden Markov model.

STEM_HMM structure containing all images, all coordinates, and all models of the images in the time series. Using a slider, different images of the time series can be displayed. Most importantly, the scattering cross-sections determined from each image are added as a row in a matrix O , which contains the observations. It is important to sort all scattering cross-sections in the same order throughout the time series, and in the same order as the input coordinates.

4.5.1 Run HMM

By clicking this button, the hidden Markov model is estimated from the matrix O containing the observed sequence of scattering cross-sections for all columns in the ADF STEM images of the time series. The output is a state sequence, obtained by the Viterbi algorithm, and therefore called H_{viterbi} in the output structure, which can be saved from StatSTEM. This is a matrix with the same dimensions as the observed sequence, which contains the corresponding number of atoms in the columns. Atom counts are displayed in StatSTEM, and can be viewed for different images using a slider.

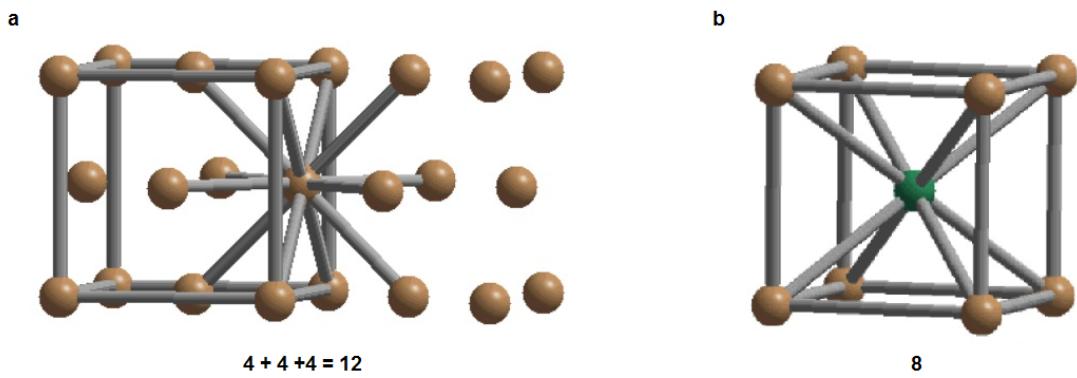


Figure 4.8: The coordination number of (a) an FCC and (b) a BCC crystal can be calculated by searching for neighbouring atoms within a radius of $a/\sqrt{2}$ or $2 \times a/\sqrt{3}$, respectively.

4.6 3D model

In sections 4.3 and 4.4, it has been mentioned that a 3D model can be made from atom-counting results. Once the model is constructed,

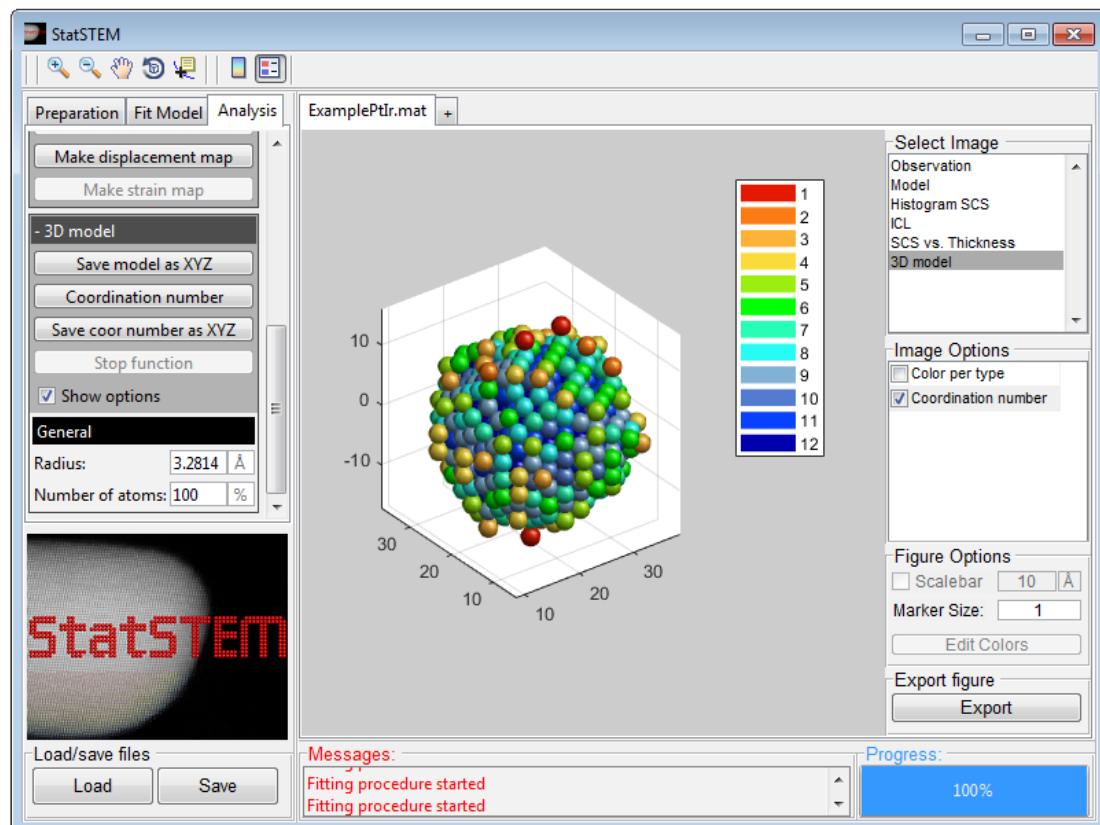


Figure 4.9: A 3D model of a Pt/Ir particle, indicating the coordination number per atom.

Coordination number	Atom type	Coordination number	Atom type
1	V	7	Lu
2	Mg	8	Yb
3	Au	9	Al
4	Na	10	Np
5	Se	11	Ho
6	Zr	12	Co

Table 4.1: Atom type per coordination number that is used when storing the 3D model with coordination numbers as an XYZ file.

this panel provides options to export the coordinates as an XYZ file or to calculate the coordination number.

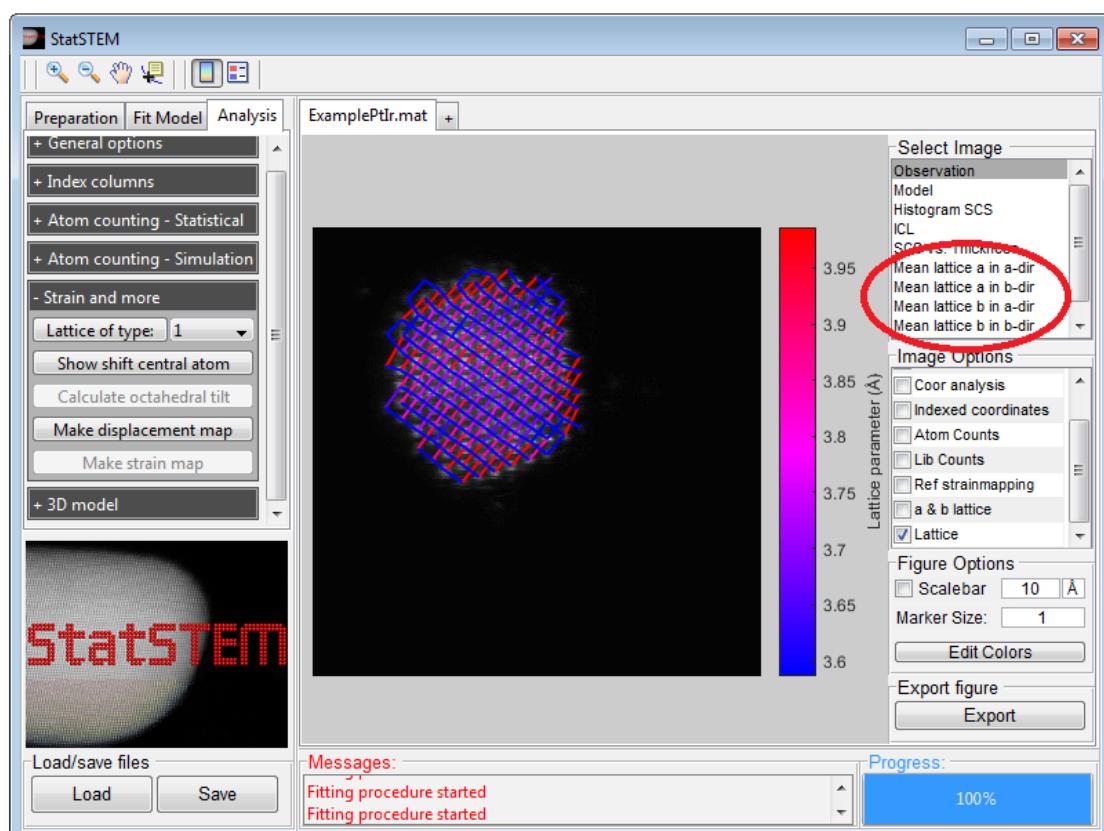


Figure 4.10: The lattice parameters a and b of a Pt/Ir nanoparticle. Plotting the a and b lattice parameters as a function of distance in the a - or b -direction is possible, as indicated by the red circle.

4.6.1 Save model as XYZ

With this option, the constructed 3D model can be saved as an XYZ file that can be loaded into other software packages such as Vesta or Visual Molecular Dynamics (VMD).

4.6.2 Coordination number

This function can be used to determine the coordination number of each atom in the 3D model. The coordination number is determined by calculating the number of neighbours of each atom within a specific radius.

Radius

In StatSTEM, the standard radius that is used is $a \times 0.8$, which is a little bit larger than $a/\sqrt{2}$ to compensate for small fluctuations in the atom positions.

Number of atoms

By default, the coordination number is determined for all atoms (100 %). As this is a demanding calculation, the user can decide to

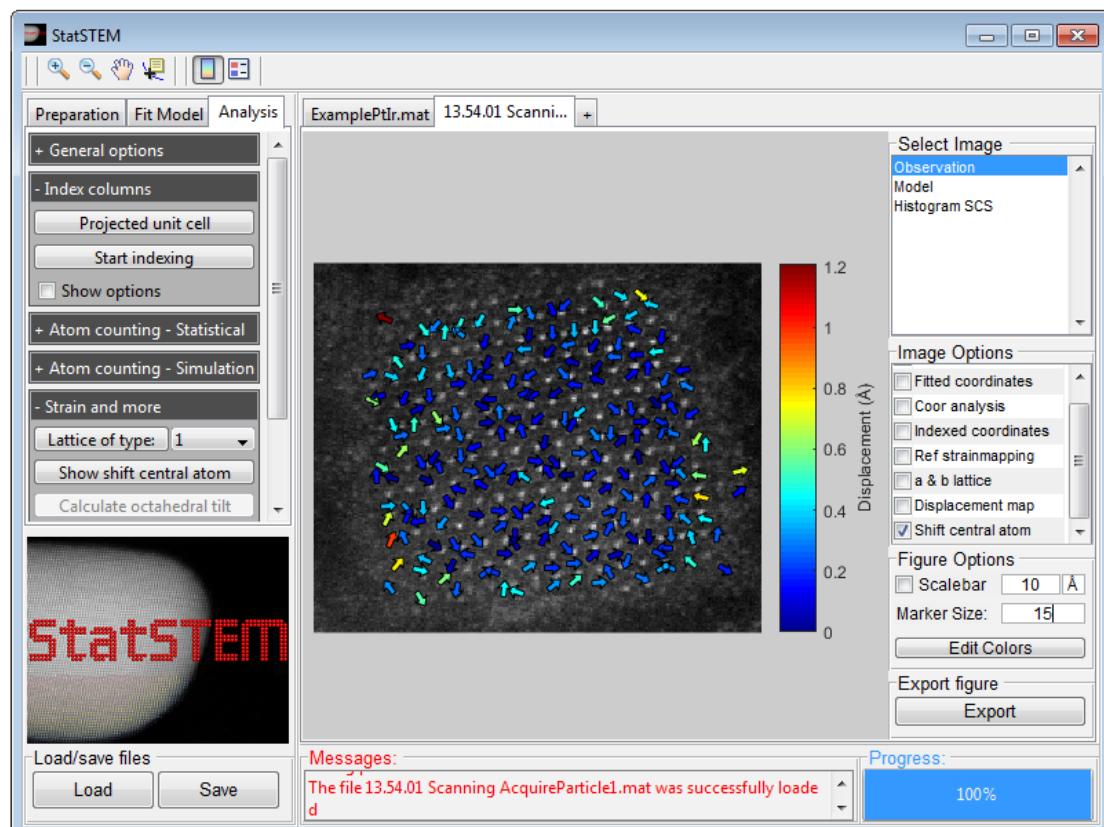


Figure 4.11: Displacement map of the central atom in PbCsBr_3 .

leave out the most central atoms for computational purposes. The coordination numbers of these atoms are determined based on the distance from the centre of the particle. In this manner, one can calculate the coordination number only for a fraction of the atoms in the particle.

4.6.3 Save coor number as XYZ

Once the coordination numbers are determined, they can be saved as an XYZ file by hitting *Save coor number as XYZ*. Hereby, a specific atom type is given in function of the coordination number. The types as a function of coordination number are listed in Table 4.1.

4.7 Strain and more

In this panel, one can use the fitted atom column coordinates of the model for determining the lattice parameters, measuring displacement of atoms, analysing octahedral tilt, and constructing strain maps.

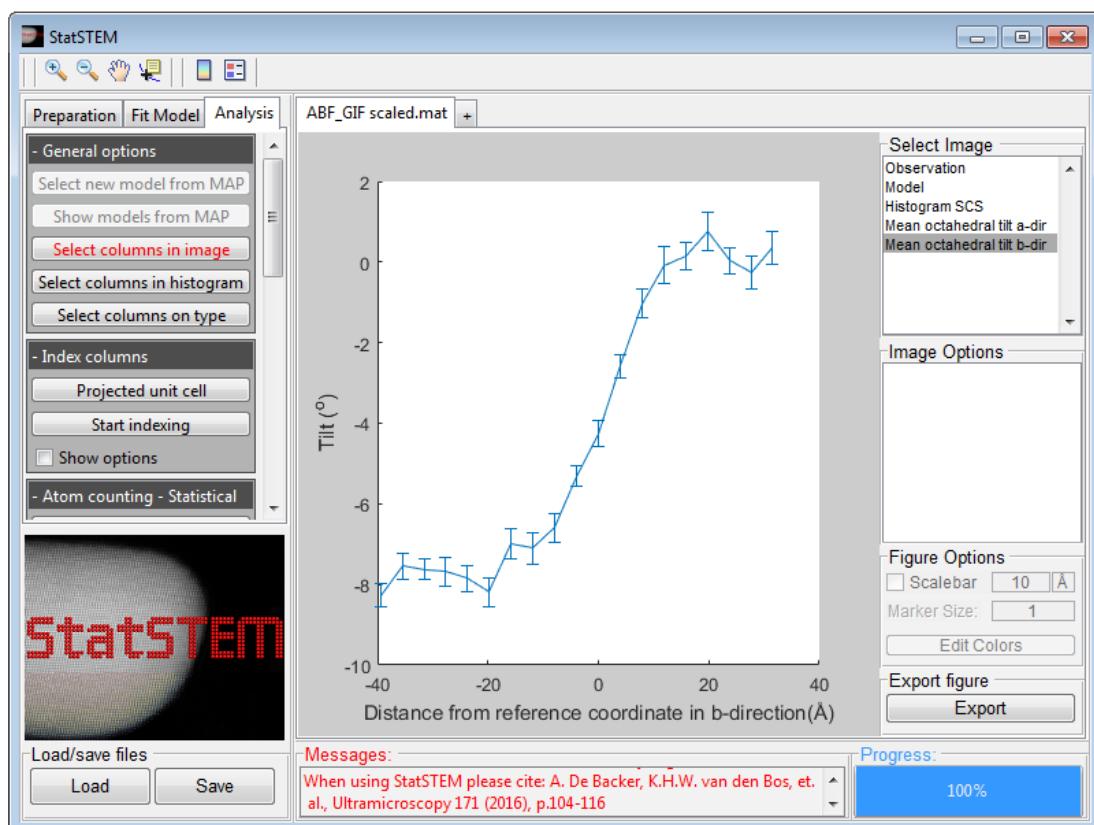


Figure 4.12: Octahedral tilt measured along the b-direction in PtTiO_3 .

4.7.1 Lattice of type

This function determines the lattice parameter per column type. Automatically, a coloured plot is made that shows the lattice parameters, as illustrated in Figure 4.10. StatSTEM also enables in the *Select Image* panel at the right-hand side of the GUI to visualise plots where the lattice parameters as a function of distance from the reference coordinate are shown, in both a- and b-directions.

4.7.2 Show shift central atom

By clicking on this button, a displacement map is made of the central atom in a unit cell, based on the column indexing described in section 4.2. By using the values of the projected unit cell, the expected coordinates are calculated. The displacement map is generated by comparing the expected coordinates with the measured coordinates.

4.7.3 Calculate octahedral tilt

In perovskite materials, there are oxygen octahedra present surrounding the B-cation (or A-cation). Due to internal strain, the oxygen octahedra can rotate. From the indexed columns, octahedral

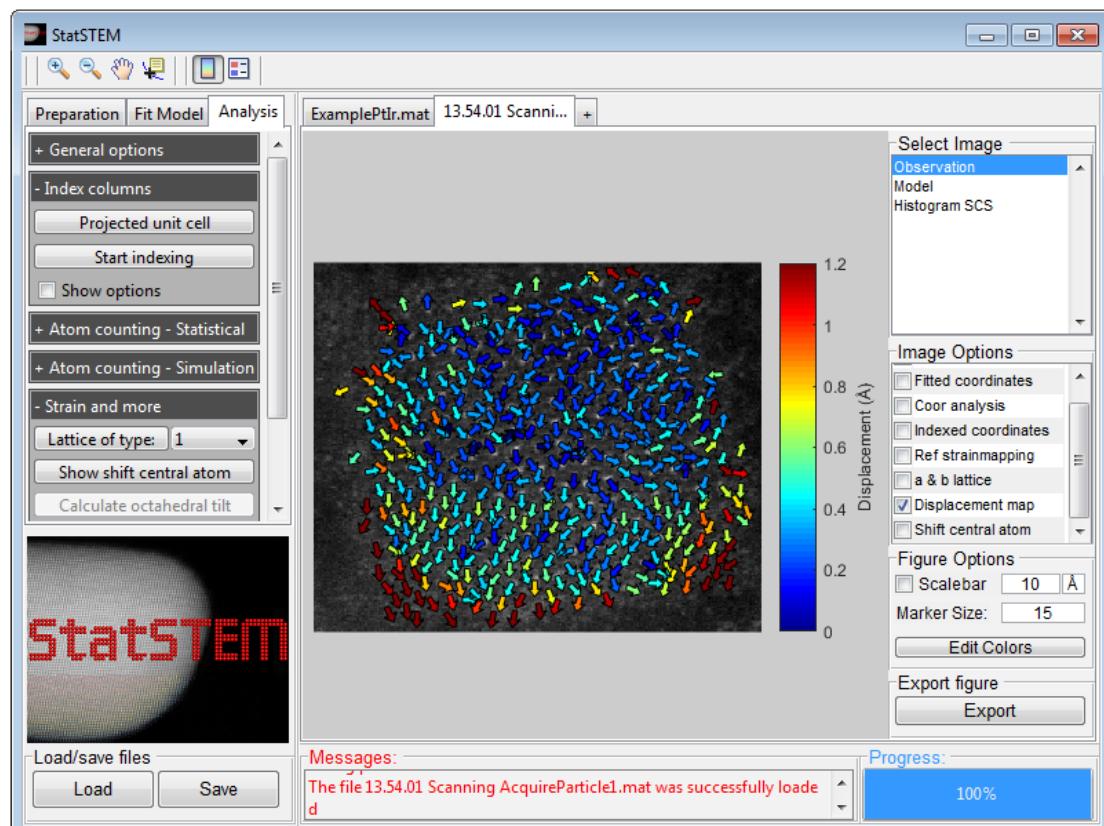


Figure 4.13: Displacement map of a PbCsBr_3 particle.

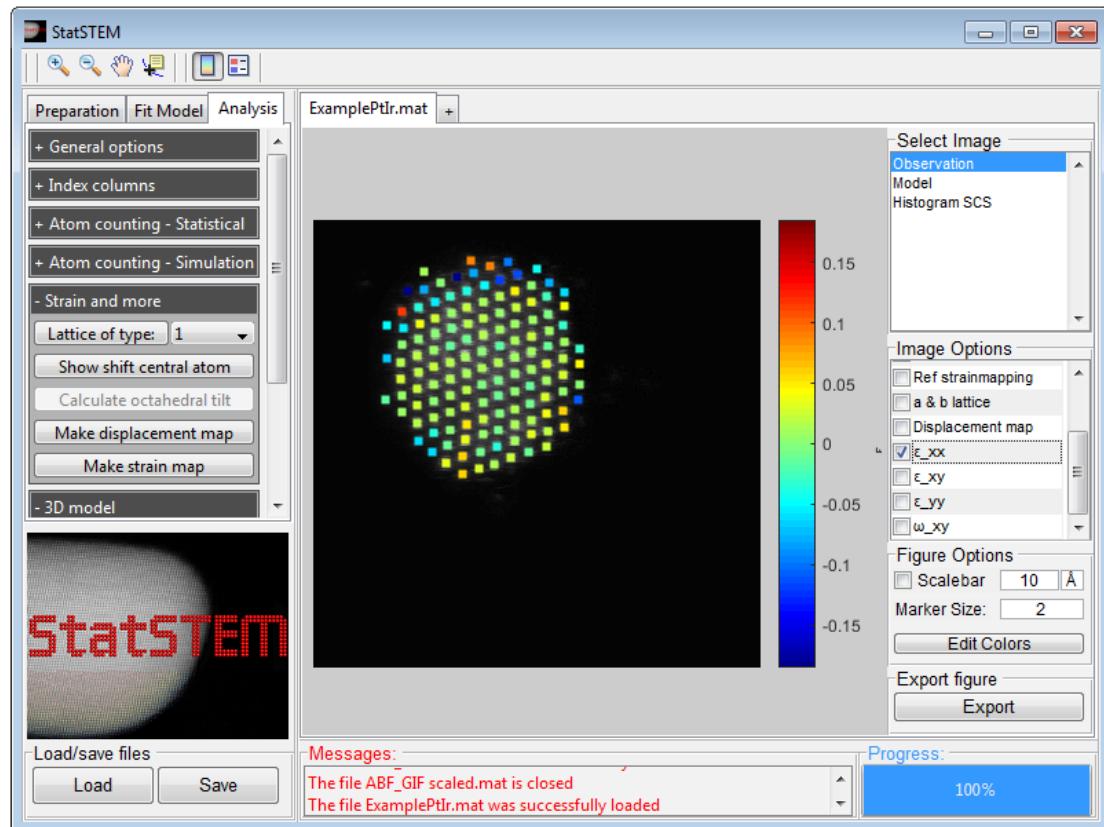


Figure 4.14: An ϵ_{xx} strain map of a Pt/Ir particle.

tilt can be determined when there are atoms present at the relative positions in the unit cell: (0,0.5) and (0.5,0). If this condition is satisfied, the *Calculate octahedral tilt* button becomes available to determine the octahedral tilt as a function of distance in the a- and b-direction. The distance is measured from the reference coordinate selected when indexing the columns, as described in section 4.2. In the plot, the octahedral tilt is calculated in regions where it is assumed that the octahedral tilt is alternating between a clockwise and an anti-clockwise rotation.

4.7.4 Make displacement map

By hitting this button, a displacement map is constructed based on the column indexing described in section 4.2. By using the values of the projected unit cell, the expected coordinates are calculated. The displacement map is generated by comparing the expected coordinates with the measured coordinates.

4.7.5 Make strain map

By clicking *Make strain map*, the ϵ_{xx} , ϵ_{xy} , ϵ_y and ω_{xy} strain maps are generated. Hereby, the derivative of the displacement map is used [11]. An example is shown in Fig. 4.14.

5 Remarks and suggestions

When downloading StatSTEM, a folder with examples is included. In this folder, MATLAB scripts are also available to use StatSTEM without the GUI. The StatSTEM website offers a forum in case you have any questions, remarks or suggestions. Here, also new releases of StatSTEM with new functionalities are announced.

6 References

- [1] F.F. Krause, M. Schowalter, T. Grieb, K. Müller-Caspary, T. Mehrtens, and A. Rosenauer. Effects of instrument imperfections on quantitative scanning transmission electron microscopy. *Ultramicroscopy*, 161:146–160, 2016.
- [2] A. De Backer, K.H.W. van den Bos, W. Van den Broek, J. Sijbers, and S. Van Aert. StatSTEM: An efficient approach for accurate and precise model-based quantification of atomic resolution electron microscopy images. *Ultramicroscopy*, 171:104–116, 2016.
- [3] J. Fatermans, A.J. den Dekker, K. Müller-Caspary, I. Lobato, C.M. O’Leary, P.D. Nellist, and S. Van Aert. Single atom detection from low contrast-to-noise ratio electron microscopy images. *Physical Review Letters*, 121(5), 2018.
- [4] J. Fatermans, S. Van Aert, and A.J. den Dekker. The maximum a posteriori probability rule for atom column detection from haadf stem images. *Ultramicroscopy*, 201:81–91, 2019.
- [5] G.T. Martinez, L. Jones, A. De Backer, A. Béché, J. Verbeeck, S. Van Aert, and P.D. Nellist. Quantitative STEM normalisation: The importance of the electron flux. *Ultramicroscopy*, 159:46–58, 2015.
- [6] S. Van Aert, K.J. Batenburg, M.D. Rossell, R. Erni, and G. Van Tendeloo. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature*, 470:374–377, 2011.
- [7] S. Van Aert, A. De Backer, G.T. Martinez, B. Goris, S. Bals, and G. Van Tendeloo. Procedure to count atoms with trustworthy single-atom sensitivity. *Physical Review B*, 87(064107), 2013.
- [8] A. De Backer, G.T. Martinez, A. Rosenauer, and S. Van Aert. Atom counting in HAADF STEM using a statistical model-based approach: methodology, possibilities, and inherent limitations. *Ultramicroscopy*, 134:23–33, 2013.

- [9] A. De wael, A. De Backer, L. Jones, A. Varambhia, P.D. Nellist, and S. Van Aert. Measuring dynamic structural changes of nanoparticles at the atomic scale using scanning transmission electron microscopy. *Physical Review Letters*, 124(10), 2020.
- [10] A. De wael, A. De Backer, and S. Van Aert. Hidden markov model for atom-counting from sequential adf stem images: Methodology, possibilities and limitations. *Ultramicroscopy*, page 113131, 2020.
- [11] P.L. Galindo, S. Kret, A.M. Sanchez, J.-Y. Laval, A. Yáñez, J. Pizarro, E. Guerrero, T. Ben, and S.I. Molina. The peak pairs algorithm for strain mapping from hrtem images. *Ultramicroscopy*, 107(12):1186–1193, 2007.

Appendix A Indexing/strain mapping

Statistical parameter estimation theory is capable of extracting atomic column positions with high accuracy and precision from STEM images. A direct comparison of the measured column positions with the expected column positions of an ideal crystal lattice gives the displacement of the atomic columns (see Figure A.1(b)). By using first derivatives, these measured displacement vectors can be used to compute atomically resolved strain maps (see Figures A.1(c) - A.1(f)). For creating a displacement map, a projected unit cell is used which describes the projected lattice parameters and the relative positions of the different columns. Next, a reference coordinate is given in an unstrained area. Here, the distances of the nearest neighbouring atomic columns are compared to the projected lattice parameters to find the lattice directions in the image (see Figure A.1(a)). These lattice directions are used to predict the column positions of an ideal structure. Since the pixel size given by the microscope may contain inaccuracies, the dimensions of the lattice parameters are refined by fitting the ideal column positions to the measured column positions in a selected area of $N \times N$ unit cells around the reference coordinate. Standard, an area of 3×3 unit cells is used, as is indicated by the blue region in Figure A.1(a). After the lattice parameters are found, the column positions of an ideal crystal are predicted. As the crystal structure under study may be heavily strained, measured column positions can be shifted by distances larger than the lattice parameter, hampering a direct identification of the predicted and measured column positions. Therefore, starting from the reference coordinate, the neighbouring atomic columns are identified by using the refined lattice parameters. Next, the identified columns serve as new starting points to identify their neighbouring columns. This procedure continues until the boundary of the image or the edge of the particle is reached. During this process, lattice parameters and their directions are continuously updated as strain is usually non-uniform throughout a particle. Here, the lattice parameters $\mathbf{a} = (a_x, a_y)$ and $\mathbf{b} = (b_x, b_y)$ in the \mathbf{a} - and \mathbf{b} -lattice directions are updated by using a scaling factor ν :

$$\mathbf{a}^{(new)} = \mathbf{a}^{(old)} + (\mathbf{a}^{(meas)} - \mathbf{a}^{(old)}) \cdot \nu, \quad (\text{A.1})$$

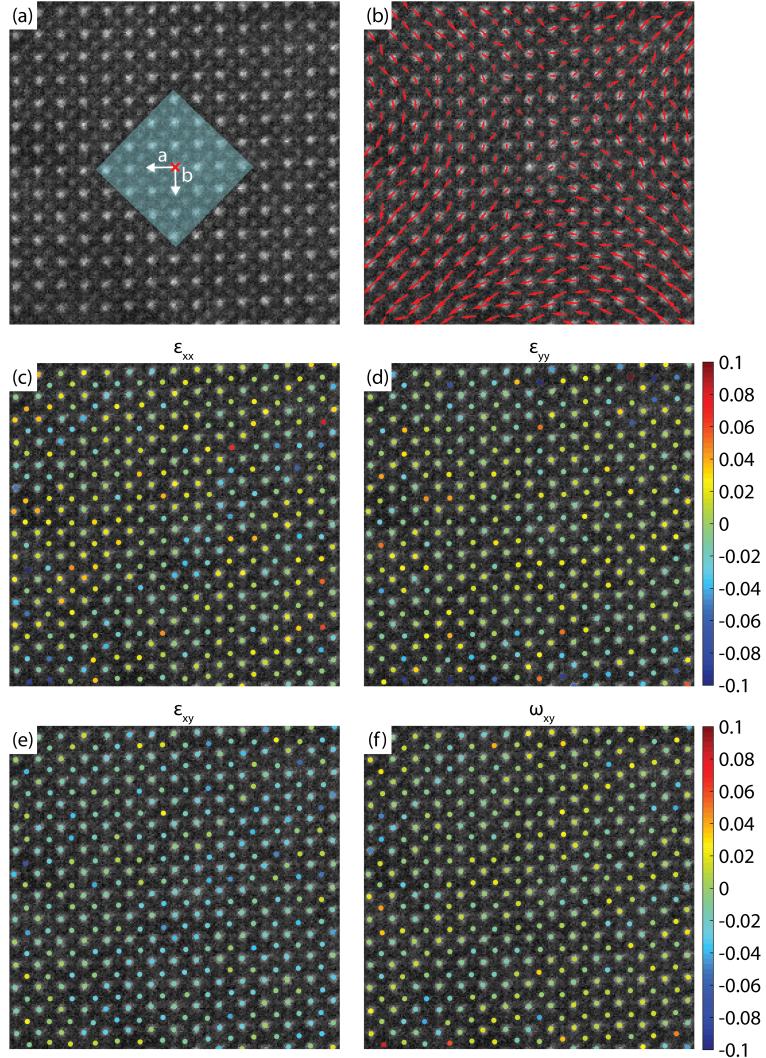


Figure A.1: Strain mapping procedure on an experimental HAADF STEM image of a CsPbBr_3 nanoparticle. (a) The procedure starts by selecting a reference coordinate, indicated by the red cross. Next, the dimensions of a projected unit cell of the ideal structure are used to find the lattice directions \mathbf{a} and \mathbf{b} . As the pixel size in the image might be inaccurate, a region of 3×3 unit cells, indicated in blue, is selected around the reference coordinate to refine the lattice parameters by using a fitting procedure. (b) A displacement map is created by comparing the measured column positions to the positions in an ideal unstrained structure. The first derivatives of the displacement map to the different lattice directions are used to determine the (c) ϵ_{xx} , (d) ϵ_{yy} , (e) ϵ_{xy} strain maps and (f) ω_{xy} rotation map.

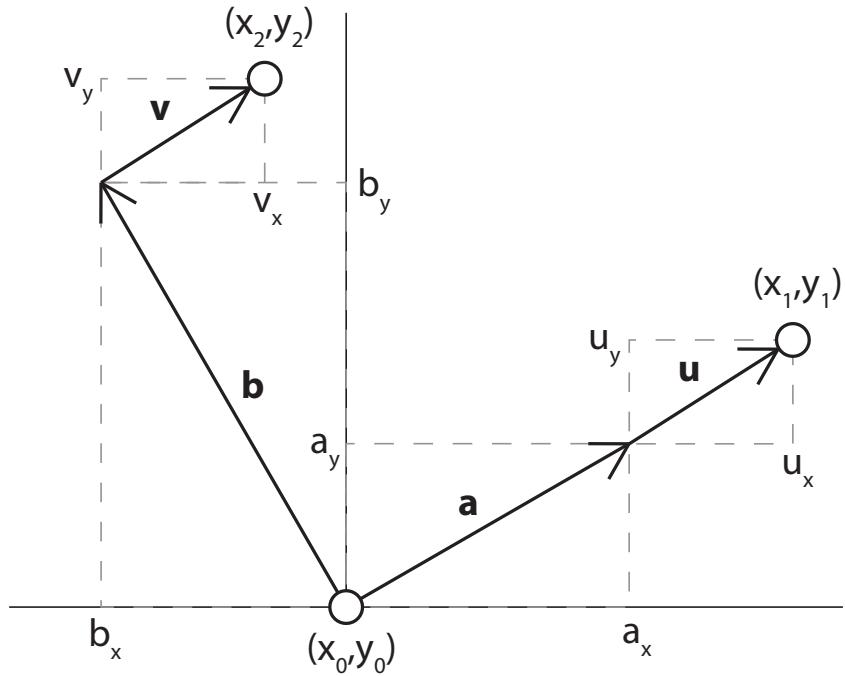


Figure A.2: Diagram of the calculation of displacement vectors \mathbf{u} and \mathbf{v} from two positions in the \mathbf{a} and \mathbf{b} lattice direction.

where $\mathbf{a}^{(new)}$ is the new lattice parameter, $\mathbf{a}^{(old)}$ is the old lattice parameter and $\mathbf{a}^{(meas)}$ is the measured lattice parameter. When all the measured atomic columns are identified, the displacement map is generated by determining the distances between the measured column positions from the image and the predicted column positions for an ideal structure (see Figure A.1(b)). In the next step, the measured displacements of the atomic column positions with respect to an ideal structure are used to calculate strain maps. Here, the first derivatives of the displacement vectors with respect to the lattice parameters give the different strain components [11]:

$$\begin{aligned}\epsilon_{xx} &= \frac{\partial \mathbf{u}}{\partial x}, \quad \epsilon_{yy} = \frac{\partial \mathbf{v}}{\partial y}, \\ \epsilon_{xy} &= \frac{1}{2} \left(\frac{\partial \mathbf{u}}{\partial y} + \frac{\partial \mathbf{v}}{\partial x} \right) = \frac{1}{2} (\epsilon_1 + \epsilon_2), \quad \omega_{xy} = \frac{1}{2} (\epsilon_1 - \epsilon_2),\end{aligned}\tag{A.2}$$

where $\mathbf{u} = (u_x, u_y)$ and $\mathbf{v} = (v_x, v_y)$ are the displacement vectors of an atomic column in the \mathbf{a} - and \mathbf{b} -lattice directions, respectively (see Figure A.2). In the StatSTEM software, the displacement vectors per atomic column are determined by using the average lattice parameters that are obtained by measuring the distance between the

selected column and its neighbouring columns. The equations of the different strain components can be rewritten in a set of linear equations [11]:

$$\left. \begin{array}{l} u_x = a_x \epsilon_{xx} + a_y \epsilon_1 \\ u_y = a_y \epsilon_{yy} + a_x \epsilon_2 \\ v_x = b_x \epsilon_{xx} + b_y \epsilon_1 \\ v_y = b_y \epsilon_{yy} + b_x \epsilon_2 \end{array} \right\} \quad \begin{bmatrix} \epsilon_{xx} & \epsilon_2 \\ \epsilon_1 & \epsilon_{yy} \end{bmatrix} = \begin{bmatrix} a_x & a_y \\ b_x & b_y \end{bmatrix}^{-1} \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}. \quad (\text{A.3})$$

From these linear equations, the values of the different strain components at each atomic column position can be computed (see Figures A.1(c)-A.1(f)). Furthermore, by using the standard deviation of the measured column positions, the error on the measured strain values can be computed, which is at maximum 0.008 in the strain maps presented in Figures A.1(c)-A.1(f).