PAPER

Obtaining 3D atomic reconstructions from electron microscopy images using a Bayesian genetic algorithm: Possibilities, insights, and limitations

Tom Stoops[iD],[1,2] Annick De Backer[iD],[1,2] Ivan Lobato[iD][3] and Sandra Van Aert[iD][1,2,*]

[1]EMAT, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium, [2]NANOlab Center of Excellence, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium and [3]Correlated Imaging, The Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, OX11 0FA, United Kingdom

[*]Corresponding author. Sandra.vanaert@uantwerpen.be

Abstract

The Bayesian genetic algorithm is a powerful tool to reconstruct the three-dimensional structure of mono-atomic single-crystalline metallic nanoparticles imaged using annular dark field scanning transmission electron microscopy. The number of atoms in a projected atomic column in the image is used as input to obtain an accurate and atomically precise reconstruction of the nanoparticle, taking prior knowledge and the finite precision of atom-counting into account. However, as the number of parameters required to describe a nanoparticle with atomic detail rises quickly with the size of the studied particle, the computational costs of the Bayesian genetic algorithm rise to prohibitively expensive levels. In this study, we investigate these computational costs and propose methods and control parameters for efficient application of the algorithm to nanoparticles of at least up to 10 nm in size.

Key words: Quantitative electron microscopy, scanning transmission electron microscopy, statistical parameter estimation, 3D characterisation

Introduction

It is widely accepted that an accurate description of the morphology and structure of metallic nanoparticles is necessary to understand the unique properties that nanoparticles exhibit (Miyazaki et al. [2003], Narayanan and El-Sayed [2004], Hvolbæk et al. [2007], Long et al. [2009], Calle-Vallejo et al. [2014], McCrum et al. [2017], Altantzis et al. [2019], Yang et al. [2019], Liu et al. [2021], Irmak et al. [2021]). Through atomic resolution annular dark field scanning transmission electron microscopy (ADF-STEM) it becomes possible to extract structural information from the nanoparticles. For small or beam-sensitive particles 3D imaging procedures such as electron tomography (Bals et al. [2014], Miao et al. [2016], Altantzis et al. [2019]) require electron doses that are too large. The Z-contrast allows for the extraction of elemental and thickness information through atom counting using a single 2D projection. The atom counting results from the ADF-STEM image may be used in combination with prior knowledge of the crystalline structure to estimate the three-dimensional shape of the imaged nanoparticle.

In past studies, various methods have been proposed to refine the initial model into an accurate representation of the nanoparticle. The structure may be refined through energy minimisation schemes such as ab initio calculations (Bals et al. [2012]) or molecular dynamics-based approaches (Backer et al.

[2017], Altantzis et al. [2019]). These methods have been shown to result in qualitatively good reconstructions by comparison with electron tomography (Backer et al. [2017], Altantzis et al. [2019]). Though, in some cases these results may deviate from the experimental observations by trending unrestrained towards a global minimum.

Alternatively, more structural information from the ADF-STEM image can be taken into account to constrain the reconstruction process. Jones et al. [2014] proposed a Monte Carlo method which optimises the energy through shifting the atomic columns along the beam direction, thereby conserving the atom-counts. Irmak et al. [2021] constructed a local minima search algorithm which combines Monte Carlo with molecular dynamics to scan the energy landscape for suitable meta-stable states. Yu et al. [2016] used a genetic algorithm to optimise the structure of a nanoparticle by matching forward modelling of the experimental images and simultaneously minimising the energy. De Backer et al. [2022] used a genetic algorithm and incorporated the prior knowledge and finite precision of the atom-counting results to balance the energy optimisation and correspondence to the experimental observations.

In this study, we focus on the Bayesian genetic algorithm (BGA) proposed by De Backer et al. [2022], as this method was reported to accurately reconstruct the surface of small nanoparticles at low doses. Although genetic algorithms are used
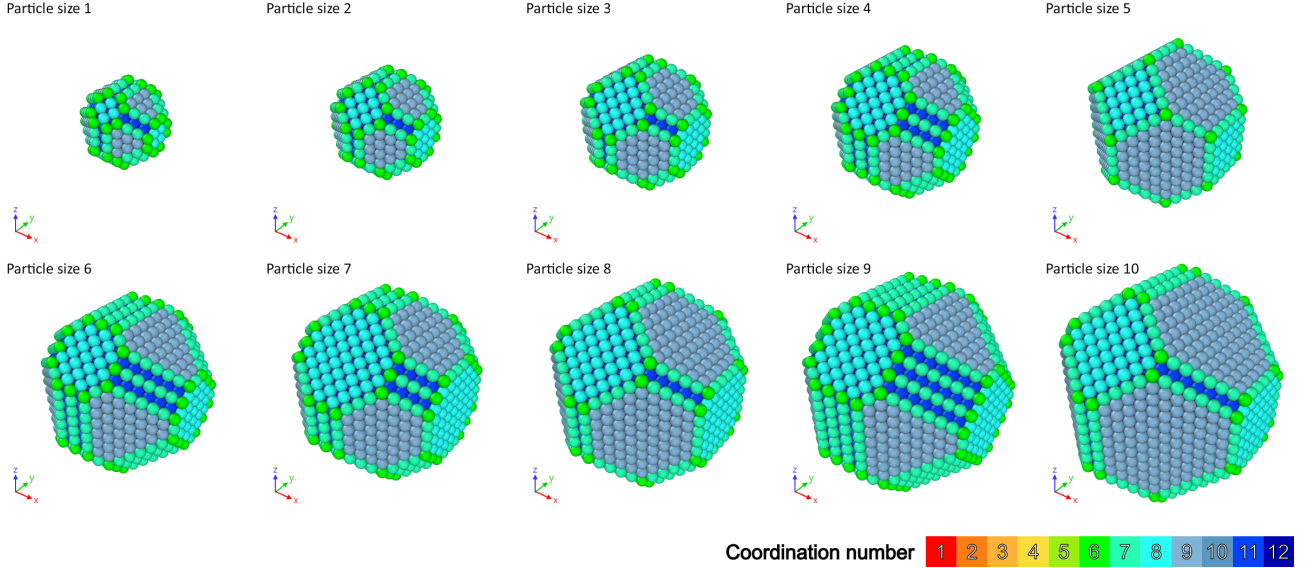
Figure 1. Ten nanoparticles generated for an approximately linearly increasing number of atomic columns along the [110] direction to study the behaviour of the computational costs of the Bayesian genetic algorithm. The colouring of the atoms corresponds to the coordination number ranging from 1 in red to 12 in dark blue. The smallest nanoparticle contains 369 atoms in 67 columns and measures approximately 2 nm, the largest nanoparticle depicted contains 5257 atoms in 383 columns and measures approximately 5 nm.

to solve large and complex optimisation problems where direct solutions cannot be found, the number of parameters needed to accurately describe the three-dimensional shape of a nanoparticle rises quickly and thus an upper limit to the size of the nanoparticle which can be studied through the BGA is set by the computational requirements. Here we will investigate the computational costs of the BGA and propose control parameters which allow us to scale up the capacity of the algorithm. This enables us to reconstruct a nanoparticle of at least up to 10 nanometres in size.

To investigate the computational cost of the BGA, we use 10 single-crystal single-element convex metallic nanoparticles of different sizes as presented in Figure 1, in the range of 2-5 nm in size. Through reconstructing these particles, we can gauge the effectiveness of the proposed genetic control parameters and the computational effort required to reconstruct a nanoparticle. In the last section, we will demonstrate the reconstruction of a larger nanoparticle of 10 nm, made possible by the optimisations introduced in this work. Note that the methodology may be applied to a broader range of particles than studied in this paper. Grain boundaries parallel to the beam can be accounted for, and particles do not need to be strictly convex in order to apply the Bayesian genetic algorithm. It suffices that all atomic columns consist of a single unbroken sequence of atoms, without cavities inside those columns.

In this paper, we revisit dose-dependent atom counting, expand on the current techniques of representing atom-counting uncertainty, improve the Bayesian genetic algorithm, and study the computational cost of the reconstruction method in more detail.

Methods

STEM simulation

Platinum Wulff structures were chosen for the 11 nanoparticles and were generated using a specialized Python package designed

to create Wulff structures in an atomistic representation (Rahm and Erhart [2020]). This shape was chosen for the simulated nanoparticles to create a representative sample of nanoparticles with sufficient thickness variation.

ADF-STEM image simulations were performed using the MULTEM package (Lobato and Van Dyck [2015], Lobato et al. [2016]). The full simulation settings can be found in the Appendix. To reflect realistic experimental conditions, averaging over 30 frozen phonon configurations was performed and a source size with Gaussian profile with a FWHM of 1Å was used to mimic general experimental conditions. Additionally, shot noise on the ADF detector was simulated through 30 noise realisations for a dose of 1e3 e/Å^2.

Dose-dependent atom counting

In this section, we will introduce the atom counting procedure as illustrated in Figure 2, including the dose-dependent contribution as demonstrated by De Backer et al. [2023].

First, a parametric model (Figure 2b) is fitted to the simulated ADF-STEM image (Figure 2a) using the StatSTEM software (De Backer et al. [2016]). This model describes the intensity at pixel (k, l) at location (x_k, y_l) in the image containing N atomic columns using a superposition of N 2D-Gaussians

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

where ρ represents the width of the 2D-Gaussians, η_n the intensity of the nth peak, β_{x_n} and β_{y_n} the x- and y-positions of the nth atomic column, and ζ a constant background which is set to zero in this simulation study. The unknown parameters to be estimated using a least-squares procedure are given by the parameter vector

$$\boldsymbol{\theta} = \left(\beta_{x_1}, \ldots, \beta_{x_N}, \beta_{y_1}, \ldots, \beta_{y_N}, \eta_1, \ldots, \eta_N, \rho\right). \quad (2)$$

From the obtained estimated parameters $\hat{\boldsymbol{\theta}}$, the estimated scattering cross-sections (SCSs) can be calculated from the volumes
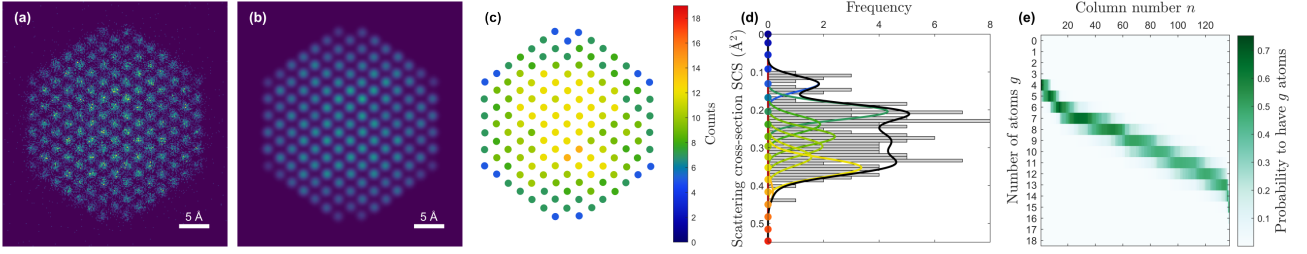
Figure 2. Dose dependent atom counting and the construction of the probability matrix for particle 3 (shown in Figure 1). **(a)** Simulated ADF-STEM image of a Pt nanoparticle along the [110] zone axis for a dose of 1e3 electrons per squared Ångström. **(b)** The fitted parametric model of the image shown in (a). **(c)** Estimated number of atoms in each atomic column for the image shown in (a). **(d)** Histograms of the observed scattering cross-sections of the image shown in (a). The estimated Gaussian mixture model is shown in black, the decomposition into the normal components is shown in colours corresponding to the number of atoms in the column. **(e)** The probability matrix indicating the probability for a column n to contain g atoms. The column labels n are assigned by sorting the scattering cross-sections in ascending order.

under the 2D-Gaussian peaks above the background (De Backer et al. [2016])

$$\text{SCS}_n = 2\pi\hat{\eta}_n\hat{\rho}^2 \ . \tag{3}$$

Figure 2d shows the histogram of the calculated SCSs, which are interpreted as independent statistical draws from a Gaussian mixture model (GMM). This model represents a linear combination of Gaussian components which describe the probability of observing a specific value of the SCS. The probability density function of the GMM is described by

$$f_{mix}(\text{SCS}_n; \boldsymbol{\psi}_G) = \sum_{g=1}^{G} \frac{\pi_g}{\sigma_g \sqrt{2\pi}} \exp\left(-\frac{(\text{SCS}_n - \mu_g)^2}{2\sigma_g^2}\right) \ , \tag{4}$$

wherein μ_g describes the locations of the normal distributions corresponding to the expected value for the SCS of a column containing g atoms. These locations were determined by simulating a Pt crystal in the [110] orientation up to a thickness of 60 atoms. The simulation settings determining the detector, probe, and frozen phonon configurations are chosen identically to the simulation settings of the simulated images, summarised in Table 3. The width of the Gaussian components in the mixture model σ_g is determined by both dose-dependent ($\sigma_{dd,g}$) and dose-independent (σ_{di}) contributions (Van Aert et al. [2019], De Backer et al. [2023])

$$\sigma_g = \sqrt{\sigma_{di}^2 + \sigma_{dd,g}^2} \quad \text{with} \quad \sigma_{dd,g} = \sqrt{\frac{\mu_g}{d}} \ . \tag{5}$$

The object $\boldsymbol{\psi}_G$ represents a vector containing all unknown parameters in the mixture model and contains G elements

$$\boldsymbol{\psi}_G = \left(\pi_1, \ldots, \pi_{G-1}, \sigma_{di}\right) \ . \tag{6}$$

Note that the locations μ_g are known and that the additional constraint on the mixing proportions $\sum_{g=1}^{G} \pi_g = 1$ ensures that only $G-1$ proportions need to be estimated.

Inherent uncertainty in the Gaussian mixture model

As described by De Backer et al. [2022] and De wael et al. [2023], the finite precision of the atom counting procedure is determined by the width of the Gaussian components, as described in equation (5). Each component represents the probability that the scattering cross-section of column n, SCS_n, is generated by the gth component $p(\text{SCS}_n|g)$. Using Bayes' theorem, the probability that the observed scattering cross-section has a thickness

of g atoms can be inferred from

$$p(g|\text{SCS}_n) = \frac{p(\text{SCS}_n|g)p(g)}{p(\text{SCS}_n)} = \frac{p(\text{SCS}_n|g)p(g)}{\sum_k p(\text{SCS}_n|k)p(k)} \ . \tag{7}$$

Under the assumption that all column thicknesses are equally likely, equal probabilities are assigned to the probability $p(g)$ of a column having g atoms. The probabilities $p(g|\text{SCS}_n)$ are visualised in the probability matrix in Figure 2e. The relationship to the width of components in the Gaussian mixture models in Figure 2d, is clearly visible. The ambiguity in assigning a number of atoms to a given column is shown in the vertical spread. When the precision of the atom counting is high, the probability matrix will have sharp and defined features. Conversely, at low doses where the precision of the atom counting is low, the probability matrix will appear more diffused.

Determining the uncertainty by sampling the distribution of the atom counts

For the larger nanoparticle (which will be discussed in section "Reconstruction of a larger nanoparticle"), a comparison of the atom counting results on the noise realised image to the ground truth thickness of the nanoparticle shows that the probability matrix constructed using equation (7) might be too narrow and thus fails to account for the true uncertainty on the counting results. Additional sources of uncertainty must be taken into account.

De wael et al. [2023] suggests an alternative approach to complement the probability matrix in order to quantify the uncertainty by sampling the actual distribution of atom counts through noise realisations of the observed GMM. Here we extend this method to capture the uncertainty in the estimation of the fitted parametric model of the image as well. If the source of the noise on the images is understood, noise realisations can be performed on the fitted parametric model, thereby generating a new image which can be analysed. By performing this bootstrapping procedure, the probability matrix can be determined for each noise realisation of the fitted parametric model using equation (7). Averaging these probability matrices for each noise realisation, gives a more accurate representation of the atom counting uncertainty.

Figure 3 shows the probability matrices for the larger nanoparticle for calculated using the inherent uncertainty in the GMM and the bootstrapping procedure which samples the distribution of the atom counts through noise realisations of the parametric model. We see that performing 50 noise realisations
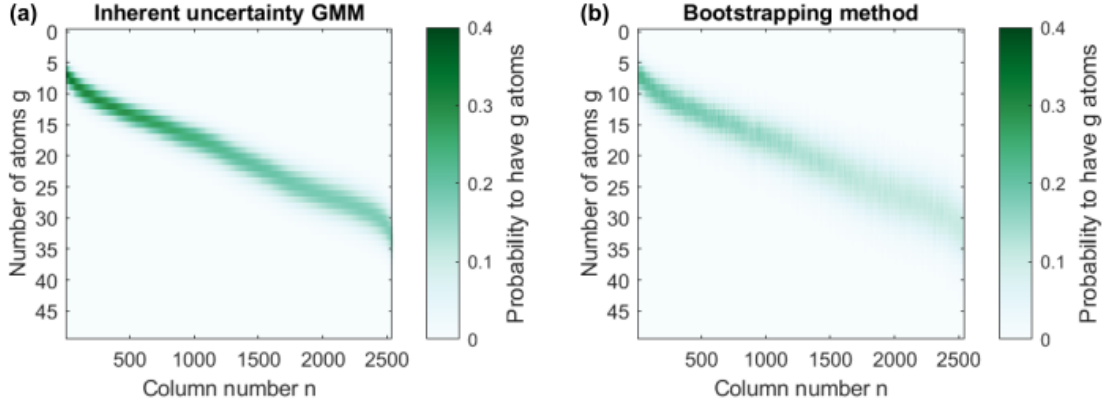
Figure 3. The probability matrix determined through the first approach discussed in section "Inherent uncertainty in the Gaussian mixture model." Determining the probability matrix through equation (7) based on the Gaussian mixture model results in the matrix shown in panel (a). On the right, Panel (b) shows the probability matrix is determined through extending the methodology and performing 50 noise realisations on the fitted parametric model of the ADF-STEM image.

of shot noise for a dose of 1e3 e/Å^2 on the fitted parametric model results in a more diffuse probability matrix, which accounts for the additional variance on the counting results caused by the finite precision of the estimated parameters of the parametric model of the image (equation (2)) and the GMM (equation (6)).
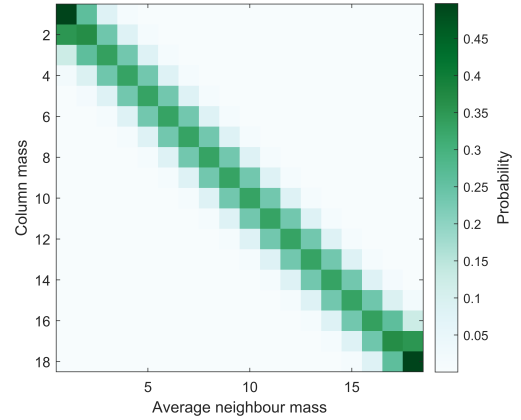

Figure 4. A visualisation of the neighbour-mass probability matrix for particle 3 (shown in Figure 1). The probability of the nth column containing g atoms is determined by the average mass of neighbouring columns NB_n.

Taking neighbouring columns into account

At low electron doses, when the overlap of Gaussian components is large, and thus the precision is small, De Backer et al. [2022] propose incorporating more prior knowledge in the form of neighbour-mass relations to further improve the quality of the reconstructions. For the convex-shaped particles considered in this work, abrupt changes in column thicknesses are unlikely. The mass or number of atoms in the nth column can be predicted by the average mass of neighbouring columns (NB_n). Therefore, a diagonal neighbour-mass matrix is proposed where the probabilities are described by a Gaussian profile such that the interval ± 1 atom corresponds to 80% of the probability in the matrix column, shown in Figure 4. The neighbour-mass probability can then be combined with the probability matrix to obtain $p(g|\text{SCS}_n \cap NB_n)$, the probability that column n contains g atoms given SCS_n and the average neighbour-mass NB_n

The Bayesian genetic algorithm

The Bayesian genetic algorithm (BGA), proposed by De Backer et al. [2022], is employed to reconstruct the three-dimensional shape of the imaged nanoparticle with atomic precision.

Using prior knowledge of the composition and the structure of the material, the atom counting results are converted into the initial model for the 3D structure of the nanoparticle by placing the columns symmetrically around a central plane at appropriate lattice positions. From this reference, we create a gene sequence consisting of two major sections describing the difference for each column in

1. the height offset from the central plane,
2. the number of atoms in the column.

These strings of genetic information span the entire solution space for the 3D structures that can generate the observed image. By construction, this requires the use of signed integers as bases for the genetic traits. This large span of bases may affect the convergence rate of the genetic algorithm negatively.

The structure of the BGA is shown in Figure 5. For every initialisation of the BGA, the algorithm will generate a population

$$p(g|\text{SCS}_n \cap NB_n) = \frac{p(\text{SCS}_n|g)p(g|NB_n)}{\sum_k p(\text{SCS}_n|k)p(k|NB_n)}. \quad (8)$$
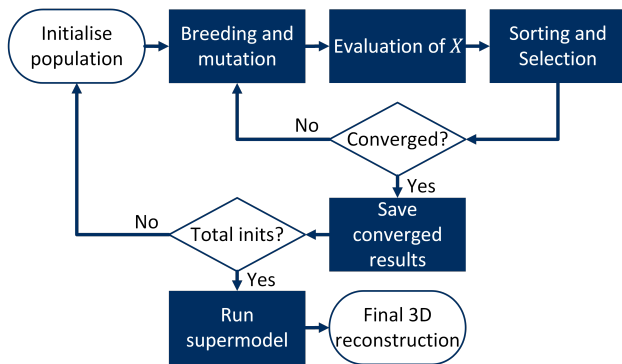
Figure 5. A Flowchart representing the order of processes in the BGA. Starting at 'Initialise population' and ending in a final solution in 'Final 3D reconstruction.'

Figure 6. Probable members are generated by considering the best members of each initialisation and calculating the frequency at which each atom is reconstructed. These probabilities are visualised in a colour scale for a noise realisation of particle 3 (shown in Figure 1). The population of probable members is generated by setting thresholds for the minimal probability in the construction shown in this figure.

Parameter	Value
Population size	$5N$
Recombination fraction	50%
Mutation density	$1/N$
Height mutation range	± 1 lattice step
Count mutation range	± 1 atom

Table 1. Control parameters chosen for the BGA as a function of the number of columns N.

of *members*, where each member represents a 3D reconstruction of the nanoparticle. Next, members undergo mutation using random number generation to create genetic diversity. In generations after the first, a crossover will occur where the *parents* randomly mix traits to generate a population of *children*, also known as bitwise independent mating (Rabinovich and Wigderson [1999]). After the population is established, the fitness is evaluated for each new member in the current population by calculating the cost-function. This function, X, balances the energy per atom with the column probabilities (De Backer et al. [2022]) and should be minimal for optimal fitness

$$X = \frac{\sum_{\ell=1}^{N} \sum_{k=1}^{g_\ell} E_k}{\sum_{m=1}^{N} g_m} \cdot \left(1 + \sqrt[N]{\prod_{n=1}^{N} p\left(g_n | \mathrm{SCS}_n \cap \mathrm{NB}_n\right)}\right). \quad (9)$$

In the next step the population is sorted according to descending fitness (ascending cost) and selects the fittest fraction (recombination fraction) to become the parents in the next generation. When the population converges, which is interpreted as the stage where for a duration of generations no significant improvement in the cost-function is made, the simulation stops and the best member (with the lowest cost-function) is saved.

However, a population converging on a solution may not result in the best possible reconstruction due to the complex nature of the cost-function landscape and the dependence on random events for the creation of members. To treat this, we run the so-called *initialisations* multiple times. Afterwards, the best members of all initialisations are put into a 'super'-population as the start of a new initialisation which is called the *supermodel*.

Additionally, effective use of genetic algorithms requires well-chosen control parameters. Although genetic algorithms have been used for several decades since the primary investigation by De Jong [1975], there is no consensus on effective control
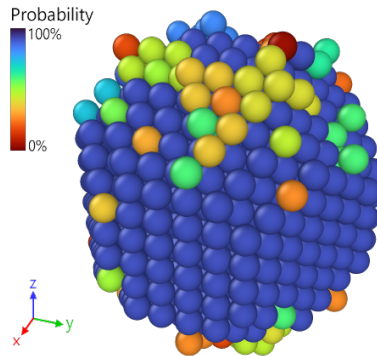
parameter choice in genetic algorithms (Mills et al. [2015]). De Jong [2007] suggests researchers should explore different control settings to find the most effective ones for particular classes of problems rather than relying on general-case control parameters. Arenas et al. [2010] and Kapoor [2011] suggest that both mutation and crossover are needed to converge on good results. Mutation creates diversity in the population, preventing premature convergence (Pandey et al. [2014]). However, too much mutation will lead to a random search rather than a structured exploration of the fitness landscape. Similarly, a high crossover rate tends to disrupt the structures selected for reproduction (Grefenstette [1986]). Furthermore, the population size should be sufficiently large since a large population increases the parallelism and helps in solving complex problems, though with diminishing returns (De Jong [2007], Kapoor [2011]) and at increased computational cost. A balance must be found in these parameters for which the Bayesian genetic algorithm will run efficiently.

For this study, we tuned the control parameters for the BGA as a function of the number of columns in the image, these are listed in Table 1. A larger particle represents a more complex problem with a larger population. Mutation scales inversely, such that the total number of changes made in a generation will scale linearly with the population size (and thus linear with the number of columns). The mutation range is limited to a small range as we assume that the final solution of the BGA will not deviate strongly from the initial symmetrical guess.

Probable solutions

To further improve the initial population of the supermodel, we propose the inclusion of probable reconstructions. After all initialisations are complete and the super-population is generated, we can make more inferences about what constitutes a 'good' member. If a trait is present in many good members, such as the reconstruction of specific atoms in the nanoparticle, it is likely to be a good trait. Using this information, we may construct additional *probable*-members for the initial super-population to compete in the supermodel. If these newly constructed members are indeed good, they compete with the other generated solutions and may pass on their traits to the final solution. If a constructed member is bad, it will be filtered out at

the selection point and will have no significant impact on the convergence of the algorithm.

We construct these probable members by taking the top members of each initialisation, and comparing the frequency at which each atom is reconstructed in 3D. A representation of the atomic reconstruction frequency is shown in Figure 6. By setting multiple thresholds for this probability, we can create a new member for the superpopulation at each threshold value. In this study we chose 20 threshold values uniformly spaced in the interval [5%, 100%].

Parallelisation

The BGA lends itself to parallelisation on two distinct levels, which we implemented in this work. Firstly, the cost-function X needs to be evaluated for all new members in a population, which is a costly endeavour for large nanoparticles and large populations. This task can be distributed efficiently over multiple cores in a CPU. Additionally, the initialisations running before the supermodel are independent of one another. These can be run in parallel on several machines or nodes in a cluster, which reduces the total real-time duration of the BGA to the duration of the longest initialisation plus the duration of the supermodel.

Naively, this parallelisation can speed up the calculations by a factor proportional to the number of CPU cores and the number of compute nodes (or machines) if we ignore the contributions of parallel overhead and neglect the time the algorithm spends outside of cost-function evaluation (the parallel component of the code). More practically, while each initialisation is accelerated by the parallelism of the cost-function evaluation, a serial part remains. We can quantify the performance gains by considering the speedup (Amdahl [1967], Gustafson [1988], Eager et al. [1989]). The time a process runs on a single thread or core T_1 can be written as

$$T_1 = T_S + T_P , \qquad (10)$$

where T_S denotes the time spent on code that is executed in series, and T_P denotes the time spent on executing parallel code. It follows that for the use of n cores in parallel, we expect the total compute time to be

$$T_n = T_S + \frac{T_P}{n} . \qquad (11)$$

The speedup gained by using n cores is then defined by

$$S(n) = \frac{T_1}{T_n} = \frac{T_S + T_P}{T_S + \frac{T_P}{n}} . \qquad (12)$$

In the limit for an infinite number of cores, the speedup approaches a constant defined by the proportion of parallel to serial code; also known as Amdahl's law (Gustafson [1988])

$$\lim_{n \to \infty} S(n) = 1 + \frac{T_P}{T_S} . \qquad (13)$$

In reality, however, communication bottlenecks will take over and decrease the speedup for increasing number of cores n. The shortest execution time can thus be reached by selecting the number of cores n that maximises the speedup S.

Nevertheless, larger numbers of cores will also result in less effective use of the cores during the serial component of the code. In high performance computing (HPC) environments, it is important to consider efficient use of resources rather

Parameter	Value
Cluster name	Vaughan
Processors	dual 2.35GHz 32-core AMD Zen 2 Rome (Epyc 7452)
Number of cores (hyperthreading disabled)	64
Available RAM	256 GB

Table 2. Configuration of HPC nodes used for the BGA simulations.

than solely focusing on the speedup and execution time. The efficiency (Eager et al. [1989]) can be defined as

$$E(n) = \frac{T_1}{nT_n} = \frac{T_S + T_P}{nT_S + T_P} , \qquad (14)$$

which should be kept as close as possible to unity for efficient use of resources. To probe the speedup and efficiency of the BGA, we have empirically calculated these quantities for a single noise realisation and single initialisation for the 10 nanoparticles. Figure 7 shows the calculated speedup and efficiency for the different sizes of nanoparticles as a function of the number of cores used. For this calculation, 100 generations were run for each nanoparticle and number of cores, thereby ensuring the simulations would not prematurely converge and influence the measurement. However, it is important to note that the results shown in Figure 7 are based on single measurements of the times T_n for each nanoparticle and number of cores which depend heavily on the stochastic nature of the BGA. Therefore, these results only serve to give an indication of the trends.

Results and discussion

Convergence rate

The computational costs of genetic algorithms can be studied by two metrics: the convergence rate of the genetic algorithm, and the computational cost of the cost-function evaluation. The first has been a topic of research for several decades and is sensitive to the control parameters (Table 1) which determine the genetic algorithm. The second is determined by the type of problem considered. In this case, the computational cost is predominantly inherited from the potential energy calculation of the nanoparticle described by the member and the calculation of the geometric mean of column probabilities. The execution time also couples with the efficiency of the used computational resource for a complex behaviour. Therefore we restrict ourselves to the discussion of the convergence rate in this section.

The convergence rate can be studied as a function of the number of columns N in the particle along the [110] zone axis, which serves as the parameter determining the values of the control parameters (Table 1). Grouping the data for all 30 noise realisations for the shot noise determines a statistical 95% confidence interval on the mean number of generations required to converge on the final solution proposed by the BGA. The conditions for linear regression are met, and the normality of residuals is confirmed by the Lilliefors test ($p = 0.36$). For the linear regression, the 95% confidence interval is calculated and overlaps well with the acquired data points and their respective confidence intervals. Therefore, we conclude that for this optimisation problem with the control parameters set by Table 1, the number of generations before convergence is a linear function of the number of columns. The data and the linear regression are shown in Figure 8.
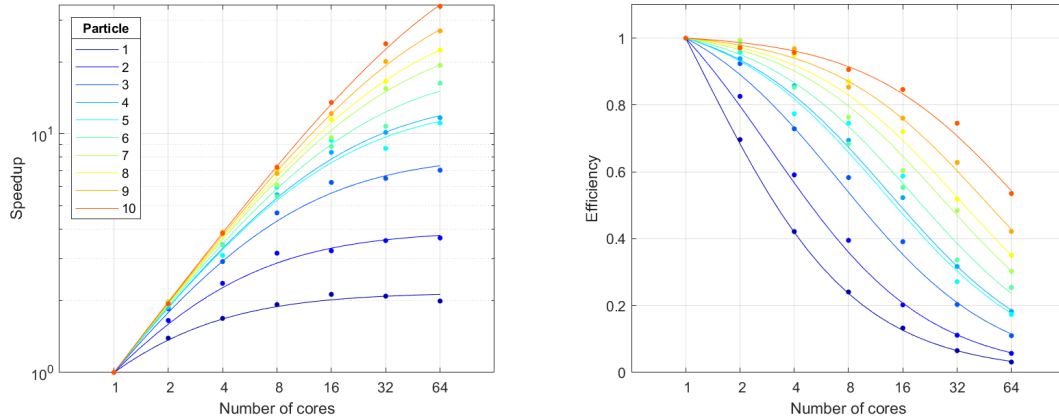
Figure 7. The speedup and CPU efficiency calculated for a single initialisation and noise realisation per nanoparticle, show the generally expected trend of increasing speedup and decreasing efficiency. In these graphs, a fitted model of the ideal case of the speedup in equation (12) and the corresponding efficiencies is shown, obtained from fitting the fraction of parallel execution time to serial execution time (T_P/T_S) in equations (12) and (14).
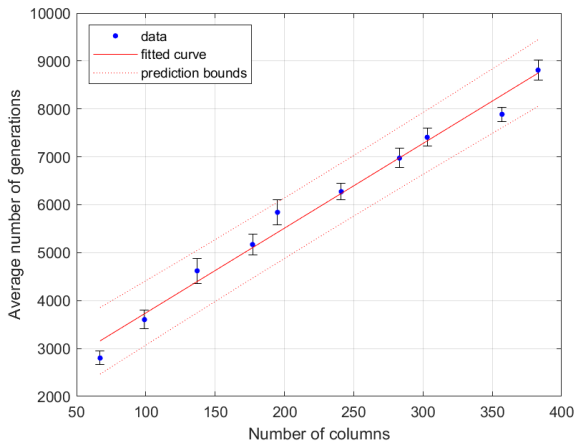

Figure 8. Convergence rate as a linear function of the number of columns in the observed image.

Benefits of parallelisation

The speedup and efficiency were calculated for the 10 smaller nanoparticles to probe the speedup gained by parallelisation of the cost-function evaluation and shown in Figure 7. From the data represented in this figure, it is evident that we must use caution when increasing the number of cores assigned to a job. Firstly, because for small nanoparticles the communication overhead forms a bottleneck reducing performance gains. For the smallest nanoparticle, it decreased the speedup after a maximum found at 16 cores. Secondly, efficient use of the requested resources is important in HPC environments. The CPU efficiency shown in Figure 7 drops off rapidly with increasing number of cores; suggesting that the compute resources were inefficiently used. Only for sufficiently large nanoparticles one should consider requesting more cores than are typically available on a desktop system.

Reconstruction of a larger nanoparticle

The BGA was shown to be a promising reconstruction method which allows for accurate three-dimensional reconstructions of nanoparticles at low doses, capable of resolving the particle with atomic detail (De Backer et al. [2022]). However, the computational requirements prohibited the reconstruction of large nanoparticles. As a result of the parallelisation and optimisations discussed in this paper, we are now able to accurately reconstruct nanoparticles of larger size. To demonstrate this, we simulated a nanoparticle of 10 nm in all three dimensions. Shot noise was added to the image to emulate a dose of 1e3 e/Å^2 in a single noise realisation, which was then counted using the dose dependent atom counting method.

To determine the probability matrix, a bootstrapping procedure is applied in which 50 noise realisations for the shot noise are used on the parametric model fitted to the initial input image, as discussed in section "Determining the uncertainty by sampling the distribution of the atom counts." Thereby allowing us to account for the additional variance in the determination of the scattering cross-sections caused by variance in the fitted parameters of the Gaussian peaks in equation (1).

The particle was reconstructed using the BGA with the parameters described in Table 1, resulting in a surface fraction reconstruction of 84%, which is in good agreement with the expected surface reconstructions (De Backer et al. [2022]) for this dose. The reconstruction is shown in Figure 9.

Conclusions

In this work we have optimised the Bayesian genetic algorithm previously presented by De Backer et al. [2022] for computational efficiency through parallelisation within and between multiple machines. A set of control parameters are suggested in Table 1 which are effective for the reconstruction of nanoparticles. Furthermore, the addition of probable members uses inference from the observed reconstructions in the initialisations to further improve the initial population of the supermodel. Through these efforts, the reconstruction of larger nanoparticles of approximately 10 nm has become possible using HPC infrastructure.

Competing interests
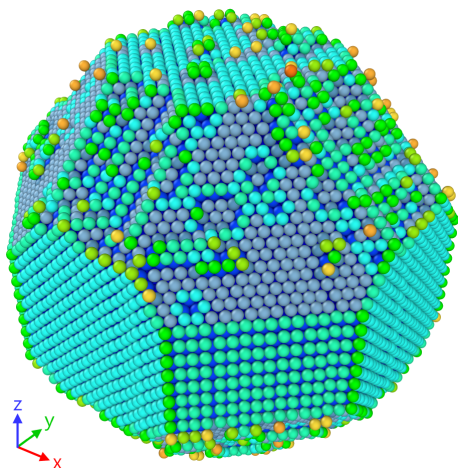
No competing interest is declared.

Figure 9. An accurate atomic resolution 3D reconstruction of a particle approximately 10 nm in size demonstrates the ability of the BGA to reconstruct larger nanoparticles. After the reconstructed particle has a recovered surface fraction of 84% compared to the ground truth.

Acknowledgements

Appendix

Multislice simulation settings

The simulation of the ADF-STEM images was performed with the settings described in Table 3, which were selected to conform to realistic settings of a probe-aberration corrected JEOL ARM200CF used in related work by De Backer et al. [2022]. The pixel size varies slightly for the different nanoparticles around the value of 0.15Å due to the choice of the scan field for each particle. In Table 4 the values for the pixel sizes are listed for each particle simulation. The slice thickness was set using MULTEM's autoslice option, which automatically determines the slice thickness based on the periodicity along the beam direction.

Reconstruction accuracy

The particles considered in this study and their reconstructions satisfy the expected fraction of correctly identified surface atoms, indicating successful reconstruction of the nanoparticles' surfaces (De Backer et al. [2022]). For the ten nanoparticles shown in Figure 1, the recovered surface fractions and their 95% confidence intervals are shown in Figure 10. The recovered surface fraction is obtained by considering all atoms with coordination number less than 12 for the ground truth nanoparticle and the reconstructed nanoparticle. The total number

Parameter	Value
Acceleration voltage	200 kV
Defocus	0 nm
Spherical abberation	0 nm
Convergence angle	22.48 mrad
Inner detector angle (ADF)	51.73 mrad
Outer detector angle (ADF)	248.41 mrad or 9.9 Å^{-1}
FWHM of the source image	1.0 Å
Root mean squared displacement	0.085 Å
Real space sampling	0.05 Å $\times$ 0.05 Å
g_{max}	10 Å^{-1}

Table 3. Parameters used for the ADF-STEM simulations of the nanoparticles and the Pt crystal in [110] and [001] zone axis up to 40 atoms thick.
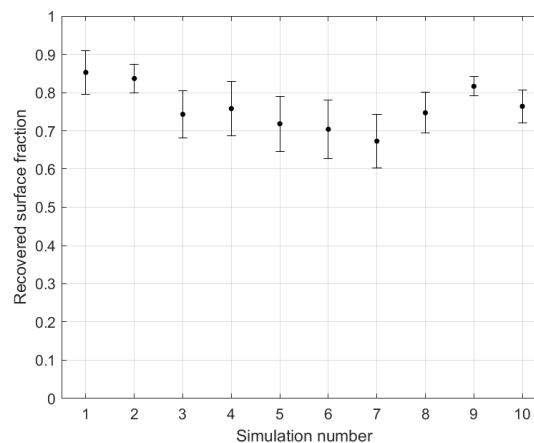

Figure 10. The fraction of recovered surface atoms for the first 10 nanoparticles used to study the computational costs of the BGA.

of coordinate-matched atoms divided by the maximum number of atoms in the surfaces of the nanoparticles is the recovered surface fraction.

References

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, and S. Bals. Three-Dimensional Quantification of the Facet Evolution of Pt Nanoparticles in a Variable Gaseous Environment. Nano Letters, 19(1):477–481, Jan. 2019. ISSN 1530-6984. doi: 10.1021/acs.nanolett.8b04303. URL https://doi.org/10.1021/acs.nanolett.8b04303. Publisher: American Chemical Society.

G. M. Amdahl. Validity of the single processor approach to achieving large scale computing capabilities. In Proceedings of the April 18-20, 1967, spring joint computer conference, AFIPS '67 (Spring), pages 483–485, New York, NY, USA, Apr. 1967. Association for Computing Machinery. ISBN 978-1-4503-7895-6. doi: 10.1145/1465482.1465560. URL https://dl.acm.org/doi/10.1145/1465482.1465560.

M. I. G. Arenas, P. A. C. Valdivieso, A. M. M. Garcia, J. J. M. Guervos, J. L. J. Laredo, and P. Garcia-Sanchez. Statistical Analysis of Parameter Setting in Real-Coded Evolutionary Algorithms. In R. Schaefer, C. Cotta, J. Kolodziej, and G. Rudolph, editors, Parallel Problem Solving from Nature,

Particle	Pixel size (Å)	Simulation box (Å×Å)	Potential sampling points
1	0.1495	40×40	800×800
2	0.1560	50×50	1000×1000
3	0.1620	50×50	1000×1000
4	0.1576	50×50	1000×1000
5	0.1616	60×60	1200×1200
6	0.1495	60×60	1200×1200
7	0.1652	60×60	1200×1200
8	0.1609	70×70	1400×1400
9	0.1499	70×70	1400×1400
10	0.1540	70×70	1400×1400
11	0.1499	140×140	2800×2800

Table 4. Pixel sizes, simulation box sizes, and the potential sampling used for the ADF-STEM simulations of the nanoparticles.

PPSN XI, pages 452–461, Berlin, Heidelberg, 2010. Springer. ISBN 978-3-642-15871-1. doi: 10.1007/978-3-642-15871-1_46.

A. D. Backer, L. Jones, I. Lobato, T. Altantzis, B. Goris, P. D. Nellist, S. Bals, and S. V. Aert. Three-dimensional atomic models from a single projection using Z-contrast imaging: verification by electron tomography and opportunities. Nanoscale, 9(25):8791–8798, June 2017. ISSN 2040-3372. doi: 10.1039/C7NR02656K. URL https://pubs.rsc.org/en/content/articlelanding/2017/nr/c7nr02656k. Publisher: The Royal Society of Chemistry.

S. Bals, S. Van Aert, C. P. Romero, K. Lauwaet, M. J. Van Bael, B. Schoeters, B. Partoens, E. Yücelen, P. Lievens, and G. Van Tendeloo. Atomic scale dynamics of ultrasmall germanium clusters. Nature Communications, 3(1):897, June 2012. ISSN 2041-1723. doi: 10.1038/ncomms1887. URL https://www.nature.com/articles/ncomms1887. Publisher: Nature Publishing Group.

S. Bals, B. Goris, L. M. Liz-Marzán, and G. Van Tendeloo. Three-Dimensional Characterization of Noble-Metal Nanoparticles and their Assemblies by Electron Tomography. Angewandte Chemie International Edition, 53(40):10600–10610, 2014. ISSN 1521-3773. doi: 10.1002/anie.201401059.

F. Calle-Vallejo, P. Sautet, and D. Loffreda. Understanding Adsorption-Induced Effects on Platinum Nanoparticles: An Energy-Decomposition Analysis. The Journal of Physical Chemistry Letters, 5(18):3120–3124, Sept. 2014. doi: 10.1021/jz501263e. URL https://doi.org/10.1021/jz501263e. Publisher: American Chemical Society.

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, Dec. 2016. ISSN 0304-3991. doi: 10.1016/j.ultramic.2016.08.018. URL https://www.sciencedirect.com/science/article/pii/S0304399116301747.

A. De Backer, S. Van Aert, C. Faes, E. Arslan Irmak, P. D. Nellist, and L. Jones. Experimental reconstructions of 3D atomic structures from electron microscopy images using a Bayesian genetic algorithm. npj Computational Materials, 8(1):1–8, Oct. 2022. ISSN 2057-3960. doi: 10.1038/s41524-022-00900-w. URL https://www.nature.com/articles/s41524-022-00900-w. Number: 1 Publisher: Nature Publishing Group.

A. De Backer, S. Bals, and S. Van Aert. A decade of atom-counting in STEM: From the first results toward reliable 3D atomic models from a single projection. Ultramicroscopy, 247:113702, May 2023. ISSN 0304-3991. doi: 10.1016/j.ultramic.2023.113702. URL https://www.sciencedirect.com/science/article/pii/S0304399123000190.

K. De Jong. Parameter Setting in EAs: a 30 Year Perspective. In F. G. Lobo, C. F. Lima, and Z. Michalewicz, editors, Parameter Setting in Evolutionary Algorithms, pages 1–18. Springer, Berlin, Heidelberg, 2007. ISBN 978-3-540-69432-8. doi: 10.1007/978-3-540-69432-8_1. URL https://doi.org/10.1007/978-3-540-69432-8_1.

K. A. De Jong. An analysis of the behavior of a class of genetic adaptive systems. PhD thesis, University of Michigan, 1975.

A. De wael, A. De Backer, C.-P. Yu, D. G. Sentürk, I. Lobato, C. Faes, and S. Van Aert. Three Approaches for Representing the Statistical Uncertainty on Atom-Counting Results in Quantitative ADF STEM. Microscopy and Microanalysis, 29(1):374–383, Feb. 2023. ISSN 1431-9276. doi: 10.1017/S1431927622012284. URL https://doi.org/10.1017/S1431927622012284.

D. Eager, J. Zahorjan, and E. Lazowska. Speedup versus efficiency in parallel systems. IEEE Transactions on Computers, 38(3):408–423, Mar. 1989. ISSN 1557-9956. doi: 10.1109/12.21127. URL https://ieeexplore.ieee.org/abstract/document/21127. Conference Name: IEEE Transactions on Computers.

J. J. Grefenstette. Optimization of Control Parameters for Genetic Algorithms. IEEE Transactions on Systems, Man, and Cybernetics, 16(1):122–128, Jan. 1986. ISSN 2168-2909. doi: 10.1109/TSMC.1986.289288. URL https://ieeexplore.ieee.org/abstract/document/4075583. Conference Name: IEEE Transactions on Systems, Man, and Cybernetics.

J. L. Gustafson. Reevaluating Amdahl's law. Communications of the ACM, 31(5):532–533, May 1988. ISSN 0001-0782. doi: 10.1145/42411.42415. URL https://dl.acm.org/doi/10.1145/42411.42415.

B. Hvolbæk, T. V. W. Janssens, B. S. Clausen, H. Falsig, C. H. Christensen, and J. K. Nørskov. Catalytic activity of Au nanoparticles. Nano Today, 2(4):14–18, Aug. 2007. ISSN 1748-0132. doi: 10.1016/S1748-0132(07)70113-5. URL https://www.sciencedirect.com/science/article/pii/S1748013207701135.

E. A. Irmak, P. Liu, S. Bals, and S. V. Aert. 3D Atomic Structure of Supported Metallic Nanoparticles Estimated from 2D ADF STEM Images: A Combination of Atom-Counting and a Local Minima Search Algorithm. Small Methods, page 9, 2021.

L. Jones, K. E. MacArthur, V. T. Fauske, A. T. J. van Helvoort, and P. D. Nellist. Rapid Estimation of Catalyst Nanoparticle Morphology and Atomic-Coordination by High-Resolution Z-Contrast Electron Microscopy. Nano Letters, 14(11):6336–6341, Nov. 2014. ISSN 1530-6984. doi: 10.

1021/nl502762m. URL https://doi.org/10.1021/nl502762m. Publisher: American Chemical Society.

V. Kapoor. An Empirical Study of the Role of Control Parameters of Genetic Algorithms in Function Optimization Problems. International Journal of Computer Applications, 31, 2011.

P. Liu, E. Arslan Irmak, A. De Backer, A. De wael, I. Lobato, A. Béché, S. Van Aert, and S. Bals. Three-dimensional atomic structure of supported Au nanoparticles at high temperature. Nanoscale, 13(3):1770–1776, 2021. ISSN 2040-3364, 2040-3372. doi: 10.1039/D0NR08664A. URL http://xlink.rsc.org/?DOI=D0NR08664A.

I. Lobato and D. Van Dyck. MULTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA. Ultramicroscopy, 156:9–17, Sept. 2015. ISSN 0304-3991. doi: 10.1016/j.ultramic.2015.04.016. URL https://www.sciencedirect.com/science/article/pii/S030439911500100X.

I. Lobato, S. Van Aert, and J. Verbeeck. Progress and new advances in simulating electron microscopy datasets using MULTEM. Ultramicroscopy, 168:17–27, Sept. 2016. ISSN 0304-3991. doi: 10.1016/j.ultramic.2016.06.003. URL https://www.sciencedirect.com/science/article/pii/S030439911630081X.

N. V. Long, N. D. Chien, T. Hayakawa, H. Hirata, G. Lakshminarayana, and M. Nogami. The synthesis and characterization of platinum nanoparticles: a method of controlling the size and morphology. Nanotechnology, 21(3):035605, Dec. 2009. ISSN 0957-4484. doi: 10.1088/0957-4484/21/3/035605. URL https://dx.doi.org/10.1088/0957-4484/21/3/035605.

I. T. McCrum, M. A. Hickner, and M. J. Janik. First-Principles Calculation of Pt Surface Energies in an Electrochemical Environment: Thermodynamic Driving Forces for Surface Faceting and Nanoparticle Reconstruction. Langmuir, 33(28):7043–7052, July 2017. ISSN 0743-7463. doi: 10.1021/acs.langmuir.7b01530. URL https://doi.org/10.1021/acs.langmuir.7b01530. Publisher: American Chemical Society.

J. Miao, P. Ercius, and S. J. L. Billinge. Atomic electron tomography: 3D structures without crystals. Science, 353(6306):aaf2157, Sept. 2016. doi: 10.1126/science.aaf2157. URL https://www.science.org/doi/full/10.1126/science.aaf2157. Publisher: American Association for the Advancement of Science.

K. L. Mills, J. J. Filliben, and A. L. Haines. Determining Relative Importance and Effective Settings for Genetic Algorithm Control Parameters. Evolutionary Computation, 23(2):309–342, June 2015. ISSN 1063-6560. doi: 10.1162/EVCO_a_00137. URL https://doi.org/10.1162/EVCO_a_00137.

A. Miyazaki, I. Balint, and Y. Nakano. Morphology Control of Platinum Nanoparticles and their Catalytic Properties. Journal of Nanoparticle Research, 5(1):69–80, Apr. 2003. ISSN 1572-896X. doi: 10.1023/A:1024451600613. URL https://doi.org/10.1023/A:1024451600613.

R. Narayanan and M. A. El-Sayed. Shape-Dependent Catalytic Activity of Platinum Nanoparticles in Colloidal Solution. Nano Letters, 4(7):1343–1348, July 2004. ISSN 1530-6984. doi: 10.1021/nl0495256. URL https://doi.org/10.1021/nl0495256. Publisher: American Chemical Society.

H. M. Pandey, A. Chaudhary, and D. Mehrotra. A comparative review of approaches to prevent premature convergence in GA. Applied Soft Computing, 24:1047–1077, Nov. 2014. ISSN 1568-4946. doi: 10.1016/j.asoc.2014.08.025. URL https://www.sciencedirect.com/science/article/pii/S1568494614003901.

Y. Rabinovich and A. Wigderson. Techniques for bounding the convergence rate of genetic algorithms. Random Structures & Algorithms, 14(2):111–138, 1999. ISSN 1098-2418. doi: 10.1002/(SICI)1098-2418(199903)14:2⟨111::AID-RSA1⟩3.0.CO;2-6. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291098-2418%28199903%2914%3A2%3C111%3A%3AAID-RSA1%3E3.0.CO%3B2-6.

J. Rahm and P. Erhart. WulffPack: A Python package for Wulff constructions. Journal of Open Source Software, 5(45):1944, Jan. 2020. ISSN 2475-9066. doi: 10.21105/joss.01944. URL https://joss.theoj.org/papers/10.21105/joss.01944.

S. Van Aert, A. De Backer, L. Jones, G. T. Martinez, A. Béché, and 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(6):066101, Feb. 2019. doi: 10.1103/PhysRevLett.122.066101. URL https://link.aps.org/doi/10.1103/PhysRevLett.122.066101. Publisher: American Physical Society.

L. Yang, Z. Zhou, J. Song, and X. Chen. Anisotropic nanomaterials for shape-dependent physicochemical and biomedical applications. Chemical Society Reviews, 48(19):5140–5176, Sept. 2019. ISSN 1460-4744. doi: 10.1039/C9CS00011A. URL https://pubs.rsc.org/en/content/articlelanding/2019/cs/c9cs00011a. Publisher: The Royal Society of Chemistry.

M. Yu, A. B. Yankovich, A. Kaczmarowski, D. Morgan, and P. M. Voyles. Integrated Computational and Experimental Structure Refinement for Nanoparticles. ACS Nano, 10(4):4031–4038, Apr. 2016. ISSN 1936-0851. doi: 10.1021/acsnano.5b05722. URL https://doi.org/10.1021/acsnano.5b05722. Publisher: American Chemical Society.