



| | |
|---|---|
| Grant Agreement number | 17NRM04 |
| Project short name | nPSize |
| Project full title | Improved traceability chain of nanoparticle size measurements |
| Deliverable Reference number | **Deliverable 5** |
| Deliverable Title | **Report on full algorithm sequences for nanoparticle detection and size measurement as developed on both a physical basis and by machine learning** |
| Lead partner for the Deliverable | Pollen |
| Due Date for Deliverable | May 2020 (+ 3 M project extension) |
| Actual Date of Submission | November 2020 |
| Coordinator | **Dr Vasile-Dan Hodoroaba**<br>BAM<br>Tel: +40 30 8104 3144<br>E-mail: Dan.Hodoroaba@bam.de |
| Project website address | https://www.bam.de/Content/EN/Projects/nPSize/npsize.html |
| Partners | **POLLEN**, BAM, LNE, PTB, SMD, VSL, CEA |

# 1. Introduction

The main objective of the nPSize project is to improve the measurement capabilities for nanoparticle size based on both measurement methods traceable to SI units and new reference materials. In WP 3 two basic approaches have been used in order to develop measurement procedures resulting in traceable results of the nanoparticle size distribution: physical modelling for the methods used in the project (TSEM, SEM, AFM and SAXS) and machine learning.

### Physical modelling

In this part, the physical models associated with different shape measurements for the techniques TSEM, SEM, AFM and SAXS have been collected and further developed with the aim to simulate the resulting signal as measured by the individual methods. Uncertainties and traceability associated with each model were investigated and evaluated. In the following, the progress on these physical models is reported for each individual method.

### Machine Learning modelling

The aim of this part is to use machine learning to enable automatic measurement of nanoparticle shape from expert a-priori information only. No physical model will be used as a-priori information in this task so that Task 3.1 and Task 3.2 can be carried out in parallel without interfering with each other.

The accuracy and traceability of the size results obtained by each technique will be analyzed and compared with the physical modelling (A3.1.5). A machine learning database will then be used to create automatic detection algorithms.

# 2. Scanning Electron Microscopy in Transmission Mode (TSEM)

To characterize the size of a nanoparticle, the decisive parameter is the threshold of the TSEM signal at the boundary of the particle. Simulated relative threshold values are required to be able to set the threshold values of experimental grey value micrographs. TSEM images have been modelled for spherical, cylindrical, cubic, and bipyramidal nanoparticles of various materials under well-known measurement conditions.

The Monte-Carlo simulations that are used to generate artificial TSEM micrographs are based on the modeling of electron scattering in solid matter. Elastic scattering processes are modeled by Mott cross sections, which are provided as tables by ELSEPA. Inelastic scattering processes are dealt with in the framework of dielectric function theory. For the physics of electrons interacting in the three materials, gold, silica, carbon, PTB has used the most recent JMONSEL tables and for TiO2, PTB is currently developing the inelastic scatter tables. Furthermore, all accessible experimental parameters such as electron beam divergence and width, geometry and material of the nanoparticle, and detector geometry are considered in the simulations. The Monte Carlo simulation for TSEM has been implemented into the Geant4 framework using its Monte Carlo engine and its geometry/material library to take into account the variety of differently shaped nanoparticles. This implementation has been validated for spherical particles by comparisons with JMONSEL, developed by NIST, for which PTB thanks John Villarrubia, NIST.

Obtaining a complete scan for a micrograph by running the Monte Carlo for each pixel is very costly in terms of computation time. Therefore, for each set of experimental parameters, the transmission yield as function of thickness is simulated by running the Monte Carlo simulation without scanning. Transmission yields are recorded whilst stepping through the material for each thickness by counting the electron trajectories passing the different thickness lines. In a subsequent and independent software, such a yield curve is input for a program (implemented in python) that generates three-dimensional particle geometries from which height maps are derived and which converts the heightmap into a yield-map using the yield curve. The finite width of the beam is included by convolving the yield-map with a

Gaussian distribution representing the widened beam finally resulting in the micrograph as TSEM signal. Fig. 2.1 illustrates this efficient simulation strategy.

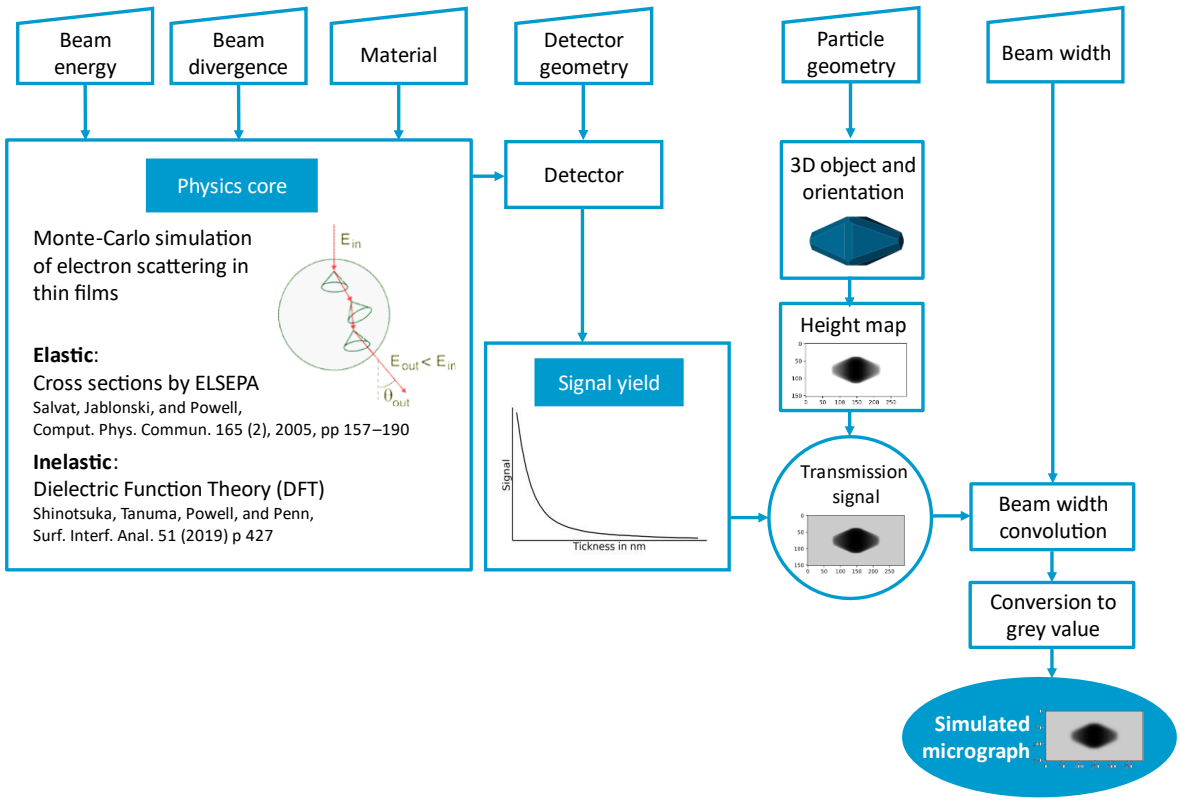



*Fig. 2.1: Efficient simulation of TSEM micrographs by running a Monte Carlo simulation to obtain a yield curve and subsequently scanning a complex geometry of a nanoparticle*

To validate the fast efficient simulation method, relative threshold values characterizing the particle boundaries obtained by the Monte Carlo running on the center and at the boundary of the particle (for which we assume an uncertainty of about $u(S_{rel})=0.025$) are compared with the relative threshold values obtained by the efficient method as described in Fig. 2.1.

If the finitely wide and divergent beam falls on the boundary of the particle, a small fraction of those trajectories that are outside the particle enters into the particle side wall due to the divergence of the beam. This effect is visible in case of particles with straight vertical sidewalls, if the complete Monte Carlo simulation is employed, but omitted by the efficient simulation method. This effect, however, is negligible for curved particle boundaries as can be seen in Fig. 2.2 (left) showing relative threshold values for cubic particles with a chamfer size of 20% of the particle size in comparison of cubes with sharp edges. Fig. 2.2 (right) shows an example of nanocubes and nanorods of gold simulated for a TSEM detector of 10.5 mrad acceptance angle.

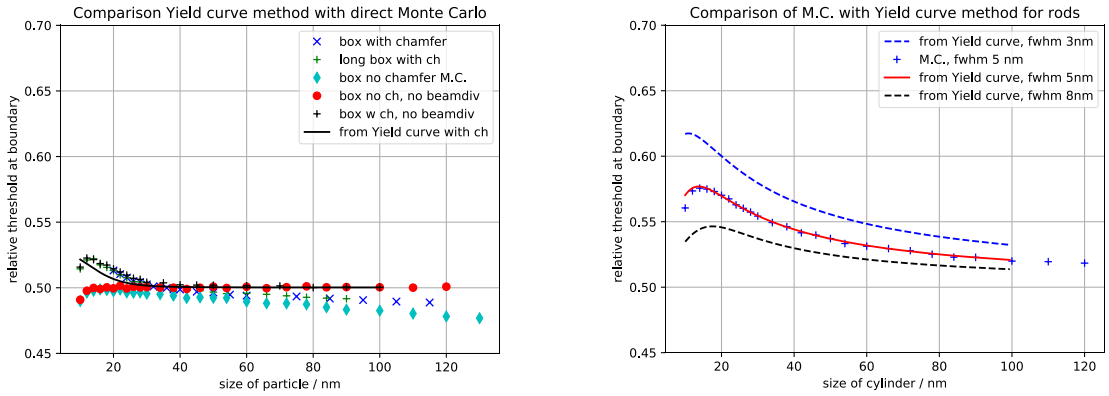



*Fig. 2.2: Simulation of relative yield threshold values of gold nanoparticles for a detector of 10.5 mrad acceptance angle. Left: cubic nanoparticles, right: nanorods*

At vertical sidewalls, as with cubic nanoparticles, the value for the relative yield threshold is 0.5 if there no beam divergence effect causes trajectories outside the particle to enter the particle for both with and without chamfer (black crosses, red circles) in Fig. 2.2 *left*, otherwise the threshold value decreases with increasing size of sidewall giving more divergent trajectories the possibility to enter the particle sidewall. For curved sidewalls, the red curve in Fig. 2.2 *right* being obtained by the efficient method agrees very well with the blue crosses (Fig. 2.2 *right*) which are obtained by the complete Monte Carlo. Furthermore, the left diagram of Fig. 2.2 shows that the threshold values of rods depend on both the diameter of the nanorod and the beam width, see dashed curves in comparison with red curve.

If the value of the nanoparticle size is approximately the value of the beam width or smaller, no reasonable boundary threshold can be determined. This limit is the structure resolution of the measurement process and can be stated to be at the particle size where the relative threshold curves have their maximum.

## 3. Physical modelling for Scanning Electron Microscopy (SEM)

To determine the equivalent diameter $D_{eq}$ of the particles by SEM technique, different segmentation methods can be used. However, the lack of reference particles and knowledge on dimensional properties of the electron beam introduces a high uncertainty to this measurement. These unknowns then make the modeling of the experimental beam difficult. An inverse method is thus proposed. This consists of simulating several profiles for different sets of experimental input parameters and then proceeding by identification between the experimental measurements (obtained profiles) and the library of simulated profiles.

A Monte Carlo algorithm (JMONSEL) was implemented. This includes various physical models of electron-sample interaction that can be used to model the secondary electron signal as a function of the beam position on the sample: i) Mott cross section for elastic scattering, ii) input tables provided with JMONSEL for inelastic scattering of electrons in gold, silica and silicon. In this way, the SEM signal can be simulated (cross-sectional profile over a spherical particle) by varying various parameters such as sample geometry (particle size), chemical composition (gold or silica), and size (standard deviation) and energy of the incident Gaussian electron beam.

For all the JMONSEL simulations carried out here, a schematic view of the studied configuration is given in Figure 3.1 : a sphere (silica or gold), with radius *R* and center coordinates (0, 0, -R) is placed on a silicon substrate assumed to be infinite.

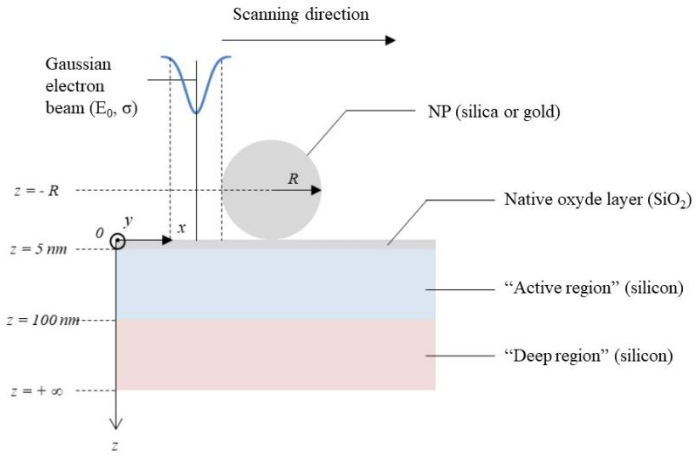



*Figure 3.1: Representation of the sample geometry used for all JMONSEL simulations.*

A 5 nm thick native oxide layer is modeled on the surface of the substrate. The substrate, consisting exclusively of silicon, is decomposed into two parts:

- A "deep" region, beyond 100 nm from the surface, where all electrons with an energy lower than 50 eV are immediately excluded from the simulation. Indeed, at this depth, these low energy electrons cannot come out of the material and be collected by a detector.

- An "active" region, between 0 and 100 nm, where all electrons are included into the simulation.

A perfectly Gaussian electron beam, with average energy $E_0$ and standard deviation σ is applied on the sample surface from an initial position ($-x_0$, 0, -h) with h >> R, h being the height of the starting point of the electron beam and x0 > R. N trajectories are modeled at this position, then the electron beam is shifted with $\begin{bmatrix} +\Delta x \\ 0 \\ 0 \end{bmatrix}$. As a result, the new beam coordinates are *($-x_0$ +Δx, 0, -h).* For each position, the secondary electron yield δ is calculated according to the number of electrons of energy < 50 eV having struck the detector(s). This operation is repeated several times in order to simulate the electron trajectories over the interval [$-x_0$, $x_0$] and thus to build up the secondary electron intensity profile along the particle.

For each set of parameters, a secondary electron profile, plotted along the NP, was generated by the Monte Carlo algorithm and stored in a database. For silica particles, three accelerating voltages (corresponding to the energy of primary electrons) have been simulated: 2 kV, 3 kV and 5 kV. For each energy, three series of measurements for three different beam sizes (standard deviation of the electron beam equals to 1 nm, 3 nm and 5 nm) were performed. Each series of measurements corresponds to the simulation of the cross-sectional profile of a nanoparticle with a radius ranging from 5 nm to 50 nm, in steps of 0.5 nm (91 profiles). Cross-sectional profiles are carried out by simulating on 121 points the yield of secondary electrons along the particle in steps of 1 nm (Figure 3.2).

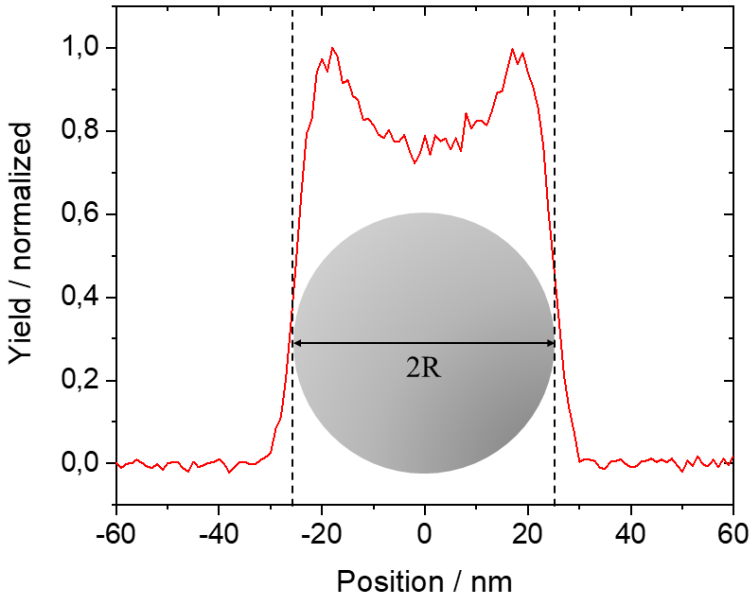



*Figure 3.2 : Cross-sectional profile (normalized) of a silica particle with a radius of 20 nm. Simulation performed with a beam size of 3 nm and an incident electron energy of 3 keV.*

Thus, 819 profiles were simulated and stored in a database for silica particles. Regarding gold particles, a series of measurements (91 profiles) with incident energy of electrons set at 3 kV and standard deviation of electron beam equal to 3 nm was simulated.

From these profiles, it is possible to determine the theoretical position of the threshold to be applied on the SEM images to obtain R, the particle radius. For this, in order to match the SEM signal, expressed in gray levels, with the signal obtained on JMONSEL defined as a secondary electron yield, all the profiles were normalized. The position of the threshold as a function of the particle radius for silica and gold with a beam size (σ) of 3 nm and an accelerating voltage of 3 kV is shown in Figure 3.3-a.

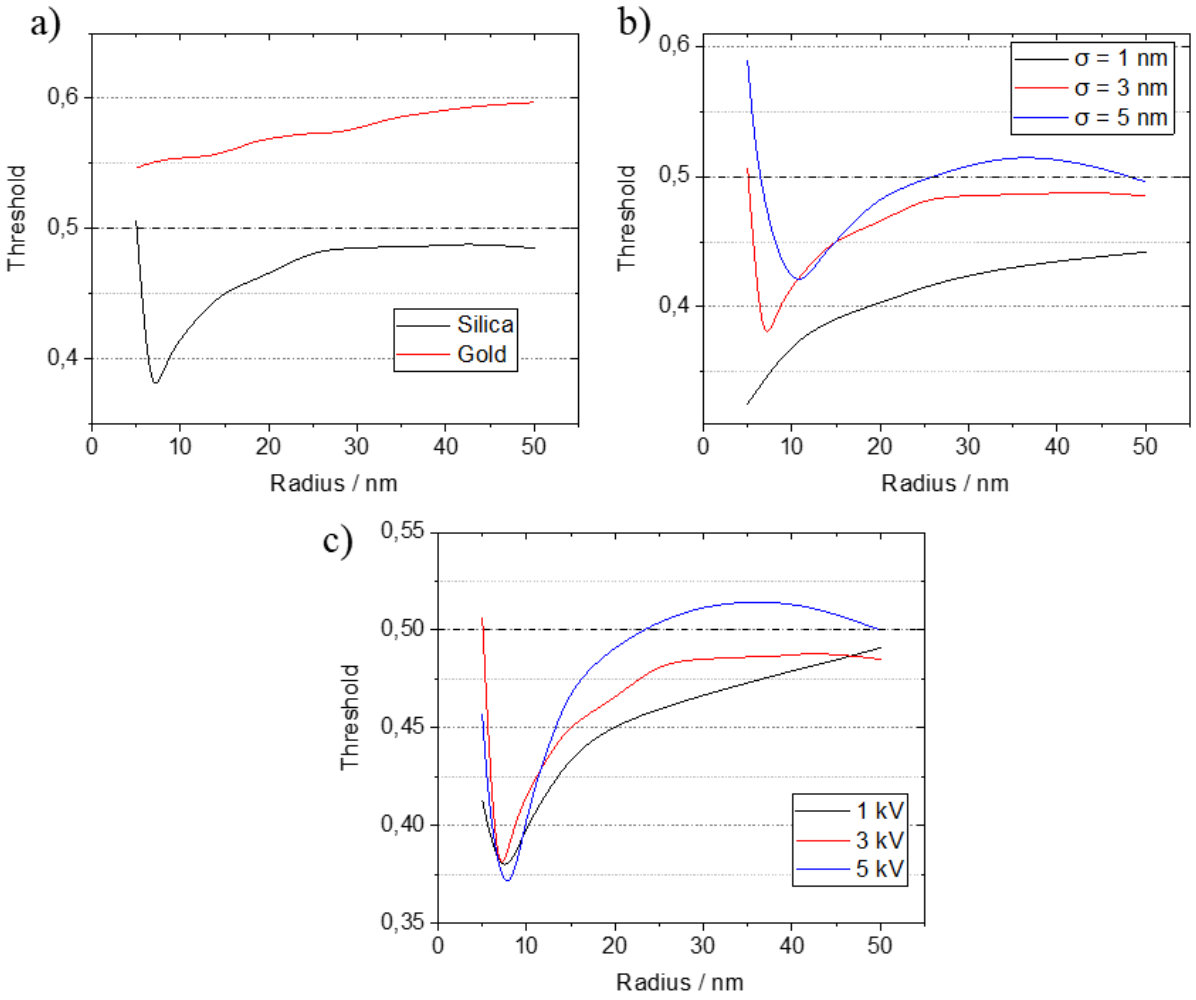



*Figure 3.3: a) Position of the threshold to be applied as a function of the particle radius for silica and gold particles with an electron energy of 3 keV and a beam size of 3 nm.*
*b) Position of the threshold to be applied as a function of the particle radius and beam size for silica particles with an incident electron energy of 3 keV.*
*c) Position of the threshold to be applied as a function of the particle radius and the energy of the incident electrons for silica particles with a beam size of 3 nm.*

We can see in Figure 3.3-a that the position of the threshold to be applied depends on the chemical composition of the particles. For example, for a particle with a radius of 30 nm, the threshold to be assigned is close to 0.5 (mid-height) for silica, 0.6 (above mid-height) for gold. Moreover, for silica particles, the position of the threshold depends strongly on the size of the particle. Moreover, for the same particle size, the position of the threshold varies according to the energy of the incident electrons (Figure 3.3-b) and the beam size (Figure 3.3-c).

A second algorithm was then developed on the Matlab software to determine which profile in the database best fits the measured SEM profile. With this method, two measurands are determined. The first one is the size of the assumed Gaussian electron beam represented by its standard deviation σ. The second is the radius R of the particle (output parameter) resulting from the comparison between the experimental profile and the library of simulated profiles (Figure 3.4).

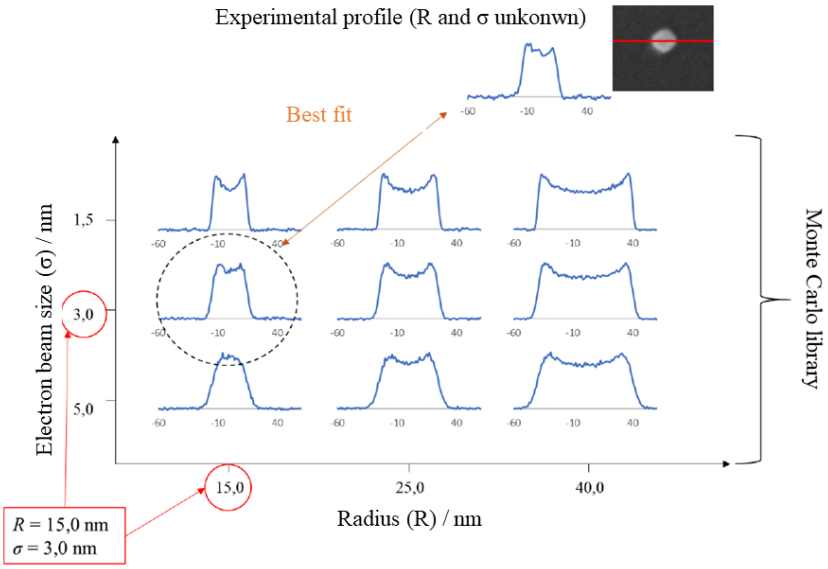



*Figure 3.4: General principle of the method. A profile in secondary electrons collected along a scan line passing through the center of the particle is compared with a series of profiles generated by the Monte Carlo method by varying the couple R (radius of the particle) and σ (beam size). The unknown parameters R and σ are then deduced according to the simulated profile showing the best match with the one obtained experimentally.*

However, to save time, the experimental profile corresponding to signal measured at the center of the particle on the SEM image is only handled to determine the parameters σ and R associated with this profile from the library. For a refined identification of R and regarding near spherical nanoparticles, a thresholding of the image is then performed. The threshold position is then evaluated from the set of parameters (accelerating voltage, chemical composition of particle, σ and R) in the database giving the best correspondence with the experimental profile (Figure 3.5).

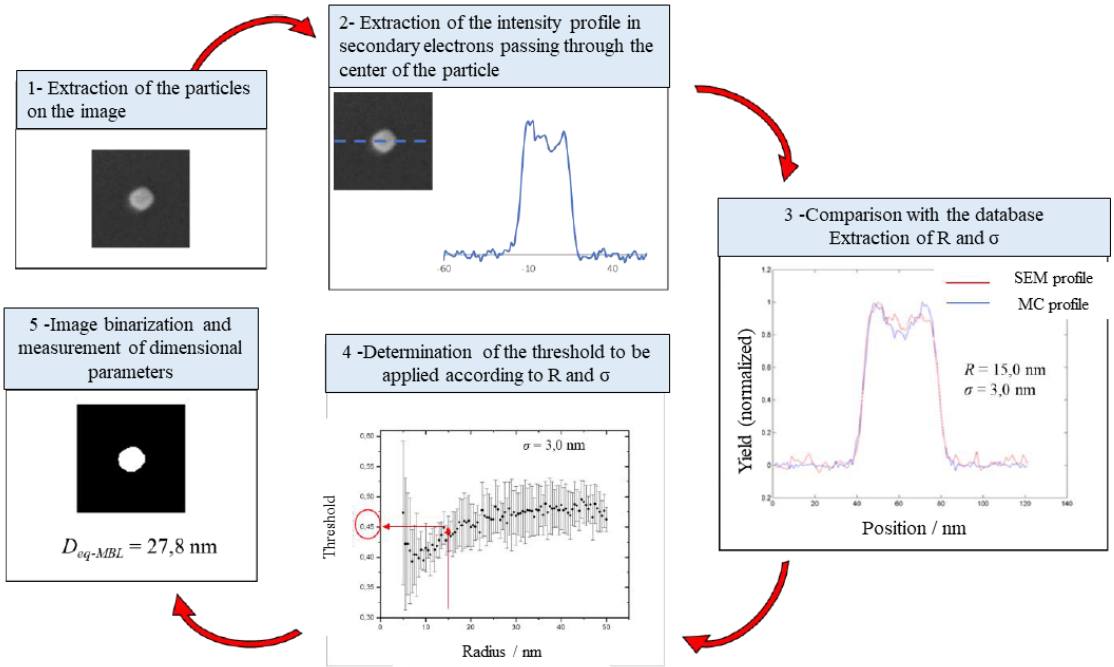



*Figure 3.5: Approach applied to measure the dimensional properties of a nanoparticle with Monte Carlo simulation:*

*1 - A particle is extracted from the SEM image and the associated thumbnail is created.*
*2 - On this particle, the secondary electron intensity profile passing through the center of the particle is extracted.*
*3 - The extracted profile is compared with the whole library of simulated profiles to determine R and σ.*
*4- These two parameters are used to determine the threshold to be applied.*
*5- The dimensional parameters are calculated from the binary image.*

This method has an advantage, as usually a single threshold is applied to the entire image. In our case, a more suitable threshold is determined for each thumbnail. In a first step, this makes it possible to get rid of the intensity variations corresponding to the substrate. Moreover, the threshold is fitted to the dimensions of the particle and the imaging parameters used.

This method was implemented to determine the dimensional properties of the silica samples produced by the CEA teams in the nPSize project (samples nPSize 8, 9, 10 and 11). The results from this method were compared with the automatic segmentation method "Active Contour", widely used whose principle is to search for inflection points around the particle to extract its contour. Results are presented in Table 1.

*Table 1: Area equivalent modal diameter evaluated using Monte Carlo library and Active Contour methods for silica samples nPSize 8, 9, 10 and 11. The uncertainty for Active Contour method is calculated following method describes in [Crouzier et al. Meas. Sci. Technol. 30(8), 2019]. Since beam width is a measurand in the Monte Carlo library method, the same procedure was used to calculate the uncertainty associated with this method without adding the uncertainty associated with the beam size.*

| Sample | Nominal diameter / nm | Area Equivalent modal diameter "Monte Carlo library" / nm | Area Equivalent modal diameter "Active Contour" / nm |
|---|---|---|---|
| nPSize 8 1st mode | 30 | 20.4 ± 0.4 *(k = 1)* | 21.0 ± 1.8 *(k = 1)* |
| nPSize 8 2nd mode | 60 | 63.9 ± 0.7 *(k = 1)* | 64.4 ± 1.8 *(k = 1)* |
| nPSize 9 1st mode | 30 | 34.1 ± 0.5 *(k = 1)* | 34.8 ± 1.8 *(k = 1)* |
| nPSize 9 2nd mode | 60 | 62.4 ± 0.9 *(k = 1)* | 63.2 ± 1.8 *(k = 1)* |
| nPSize 10 | 50 | 47.7 ± 0.7 *(k = 1)* | 48.4 ± 1.8 *(k = 1)* |
| nPSize 11 | 30 | 60.5 ± 0.9 *(k = 1)* | 61.7 ± 1.8 *(k = 1)* |

## 4. Physical modelling for Atomic Force Microscopy (AFM)

Because the selected nPSize particles have nominal dimensions that are comparable to the AFM probe size, the geometry of the features in the raw measurement data are partly resulting from the probe shape. Especially the lateral dimensions are dilated by the probe shape, so particle width and shape properties in general, cannot be accurately extracted from raw AFM data without correcting for effects from the probe shape. The accuracy of the correction process has been investigated by first simulating the measurement process of a nano particle array with a spherical probe, fig. 4.1, and subsequently correcting the dilated measured data with the same probe, fig. 4.2.

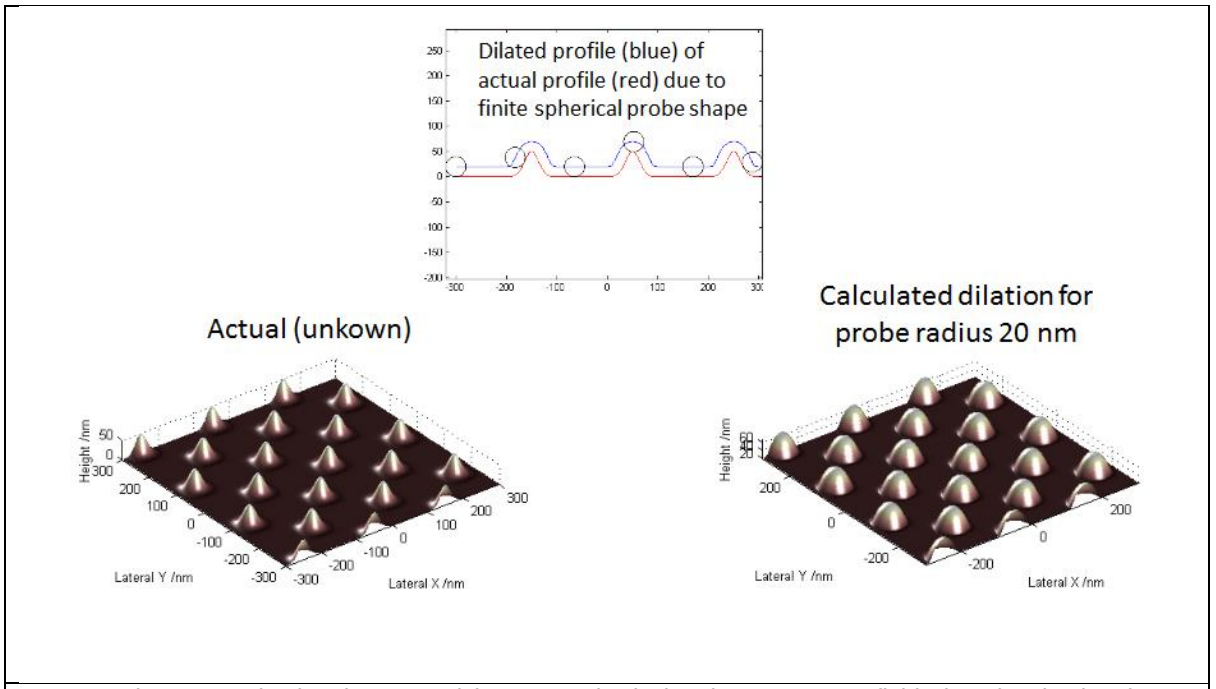



*Figure 4.1. Simulated nano particle array and calculated measurement field when the simulated array is scanned by a spherical AFM probe.*

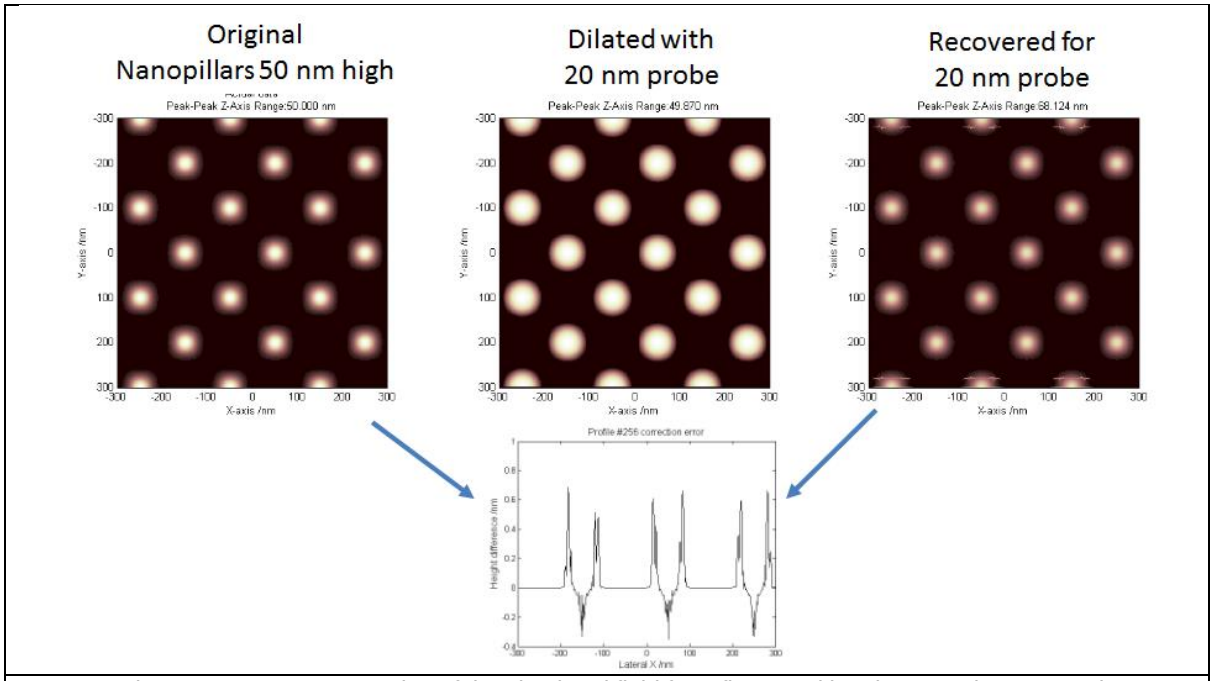



*Figure 4.2. 2D representation of the simulated field from fig. 4.1 with subsequently recovered field after correction for the probe shape. The bottom graph shows the difference between the original and recovered field.*

The difference between the original and recovered field are at the sub-nanometer level demonstrating the validity of the approach. However, the selected system is nearly ideal without measurement noise and with well-known particles and probe shape. Further investigations are ongoing to study the effect of more complex particle and probe shapes.

**Practical implementation and validation of probe shape correction**

In supporting the normative activities within the nPSize project, VSL and SMD have participated in the VAMAS/TWA 2 project 24: ”Guidelines for Shape and Size Analysis of Nano-particles by Atomic Force Microscopy“. Within this project, samples of spherical silica nano particles with a nominal height of 100 nm and two tip characterizers were distributed for a round robin test. The participants were asked to provide the particle height, particle width and a cross-sectional profile of the probe shape to allow for the correction of the raw nanoparticle profile data. In order to optimize the comparability of the measurement process between the participants to types of probes were included. In effect participating this activity was welcomed as it could be used to further develop, test and validate AFM probe-sample models within the nPSize project and additionally providing input for normative activities.

Since the probe-sample interaction results in dilation of the actual nano particle as illustrated above, accurate reconstruction of the particle width requires detailed knowledge of the probe shape. The measurement process for spherical particles and the obtained profile is illustrated below.

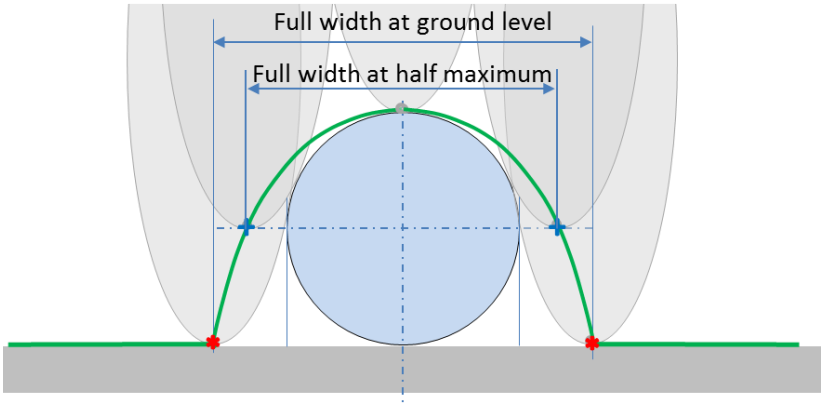



*Figure 4.3 Schematic view of the probe-sample interaction resulting in dilation of the particle profile (green line). The profile full width at ground level (red indicators) is dilated by the probe width at approximately half the nanoparticle height. The profile full width at half maximum (blue indicators) is not clearly related to a specific probe characteristic.*

The analysis of profiles to extract particle width is usually based on the measured full width a half maximum (FWHM) of the profile. However, as can be seen in figure 4.3, the FWHM of the profile is not related to a clear property of the probe shape. Correction of the measured curve for the probe shape in order to determine the particle width is therefore not clear. In contrast, the full width at ground level provides a more accurate measurand to enable correction for the probe shape: at ground level the apparent width of the particle is dilated by the full width of the probe, estimated at approximately half the height of the particle. Therefore, correction of the measured full width at ground level with the probe width estimated at half the particle height should result in the best possible estimate for the actual particle width and should be similar to the particle height for a spherical particle. Our probe shape correction will therefore be based on the full width at ground level and the estimation of the probe width at half the height of the nanoparticle.

The measurements on the silica particles revealed a height of about 116 nm, figure 4.4, This implies that the probe width has to be determined at about 60 nm from the top of the probe in order to enable reconstruction of the particle width.

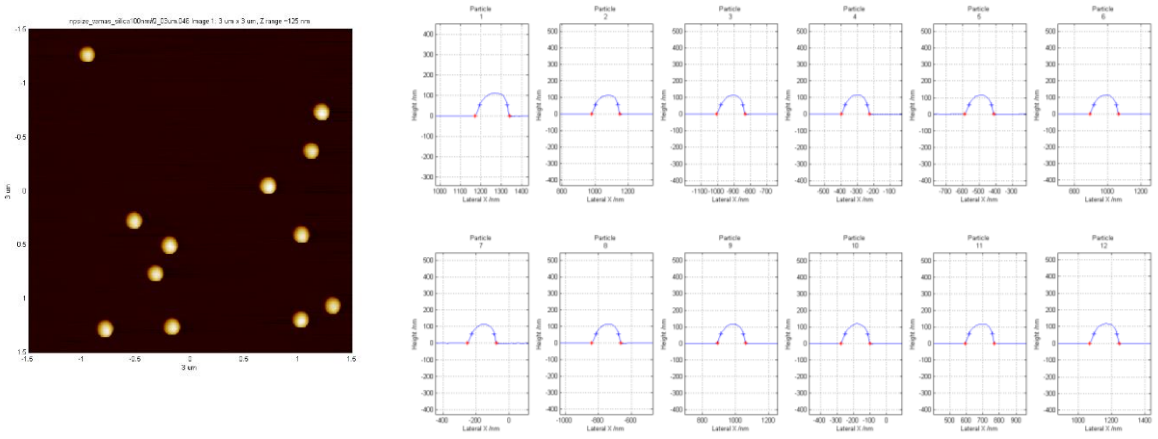


*Figure 4.6 AFM image of the silica 100 nm reference sample measured and the cross-section profiles for each particle (right).*

Measurements were performed on two types of tip characterizers on order to reconstruct the relevant part of the probe shape to enable correction of the raw nanoparticle data. The first characterizer was a sample with randomly oriented sharp structures. The measurements on this tip characterizer were analyzed by a blind reconstruction model to extract the probe shape for the two probes, Probe1 and Probe2, see Figures 4.5 and 4.6.

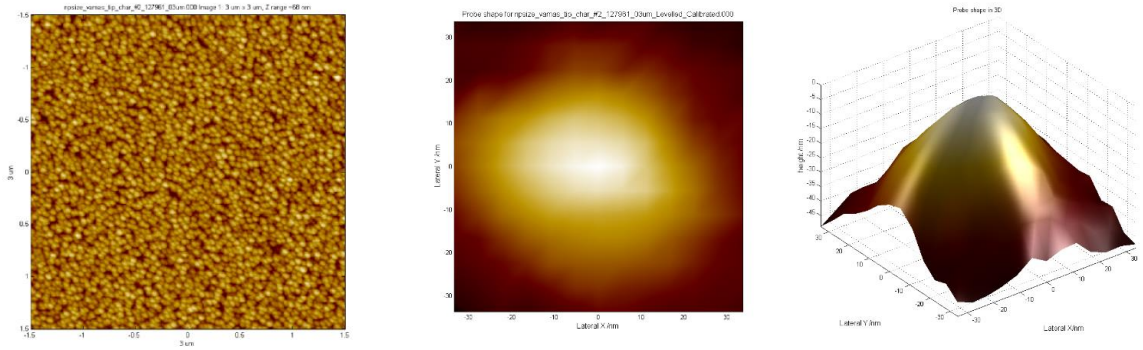


*Figure 4.5 AFM measurement (left) on the tip characterizer with Probe1 and reconstructed shape.*

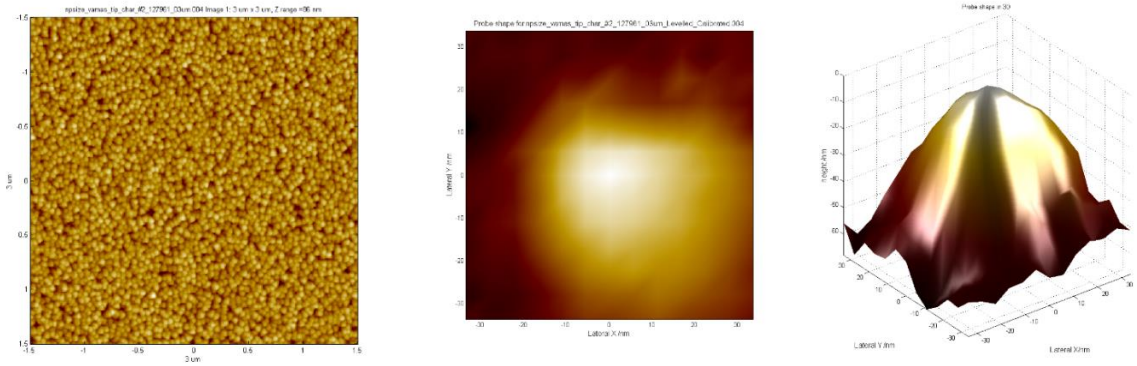


*Figure 4.6 AFM measurement (left) on the tip characterizer with Probe2 and reconstructed shape.*

From these reconstructed 3D shapes, the probe profiles along the horizontal scan direction through the maxima were calculated, as shown in figure 4.7. Note that the reconstructed profiles are not sufficiently high to accurately determine the probe width at half the height of the nanoparticle, i.e. at about -60 nm. Extrapolation is possible as indicated by the red lines, but this does not seem to be an accurate method.

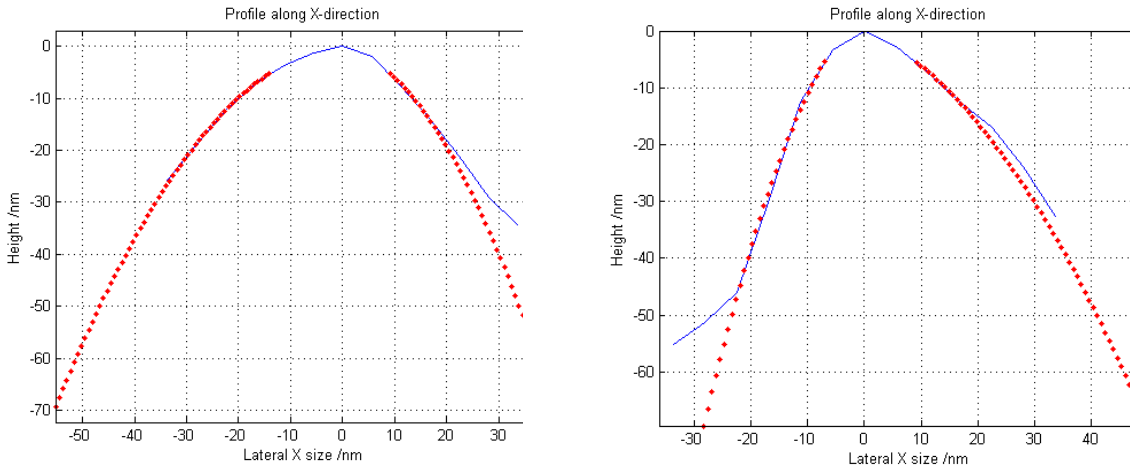



*Figure 4.7 Reconstructed profile of the probe along the fast scan direction for Probe1 (left) and Probe2 (right). The probe width at half the nanoparticle height was calculated by extrapolating the profiles (dotted red lines).*

In order to more accurately determine the probe width at about 60 nm from the top of the probe, measurements on a 1D line structure with straight edges and sufficient height were performed, see figure 4.8. As with all structures measured by AFM, the measured profiles are again dilated with the probe shape. Since these line structures are assumed to have almost straight edges, the edge regions contain information about the probe shape, whereas the details in the upper and lower levels mostly contains information of the line sample itself. In order to extract the probe shape from the edge regions it has to be determined what regions of the upper level data has to be excluded.

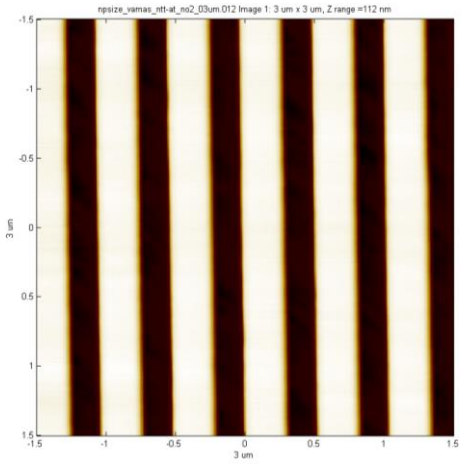



*Figure 4.8 AFM image on the line structure.*

The difficulty in the probe shape extraction on a 1D line structure is finding the exact regions that have to be excluded from the top of the profiles in order obtain an accurate reconstruction of the probe shape. In addition of just eliminating the regions of maximum values we also used the angle of the profile normal, i.e. the elevation angle, to determine which points to exclude from the profiles. However, this is still a somewhat arbitrary process since there is no clear criterion for when to exclude certain values. We do, however, already have some information about the probe shape that was reconstructed from the first tip characterizer measurements, see figure 4.7. Since the top region of this tip characterizer reconstruction is more accurate compared to the rest, it was decided to use the width at 10 nm below the top, see figure 4.7, as a criterion for the reconstruction of the probe shape from the line structure measurements. The width at 10 nm below the top of the probes was estimated as 34 nm for Probe1 and 24 nm for Probe2 resulting in the following reconstructions, see Figures 4.9 and 4.10.

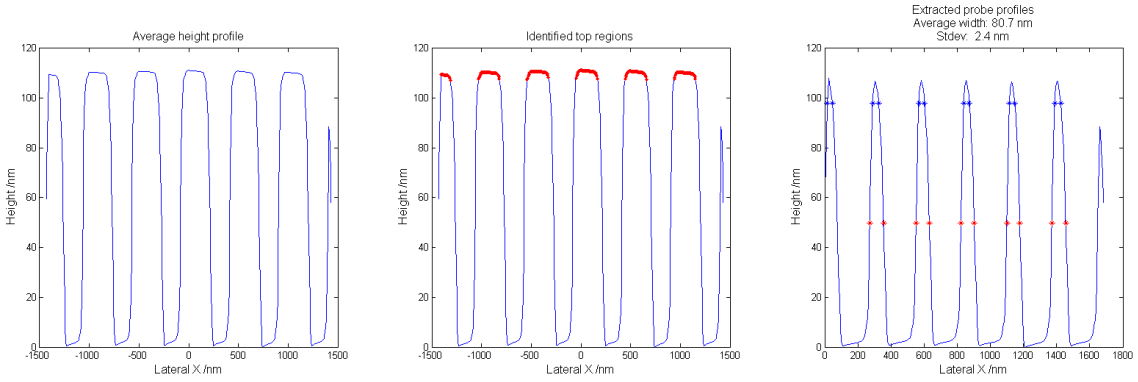



*Figure 4.9 Processing of the measurement on the line sample for the Probe1.*
*The average profile (left) was diluted with the red points shown in the center graph to provide a probe profile (right) with a width of 34 nm at 10 nm below the top (blue markers). The red markers on the right show the probe width at half the height of the silica nanoparticles.*

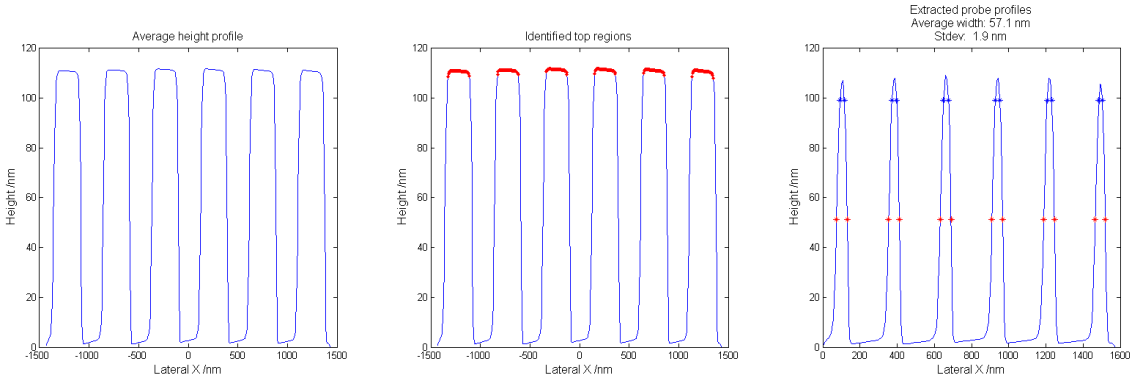



*Figure 4.10 Processing of the measurement on the line sample for Probe2. The average profile (left) was diluted with the red points shown in the center graph to provide a probe profile (right) with a width of 24 nm at 10 nm below the top (blue markers). The red markers on the right show the probe width at half the height of the silica nanoparticles.*

Since the line structures are much higher than the corrugation of the first tip characterizer, the reconstruction of the width of the probe at half the particle height is more accurate using data from the line structures compared to the values from the extrapolated profiles using the tip characterizer only.

The analysis on the line sample for the two probes resulted in an average probe width at half the nanoparticle height of 81.0 nm for Probe1 and 57.3 nm for the Probe2 and these values were used to correct the dilated profiles.

In Table 1 the results of the measurements and probe shape correction are summarized. The uncertainty of the analysis is based on the standard deviations resulting from the spread in the particle heights and the spread in the determination of the probe width. Firstly, it can be observed that the measurements with both probes result in a consistent average particle height. Secondly, the correction of the raw profiles with the calculated probe width at half the nanoparticle height results in a particle width that is consistent with the measured particle height as should be expected for spherical particles. Finally, the correction of the full width a half maximum is included to indicate that this result is not representative for the actual particle width.

*Table 1 Results for the correction of the probe shape for the values of the full width at ground level and the full width at half maximum, compared to the measured average height of the particles.*

| Probe | Silica height | Probe width at half nanoparticle height | Silica full width at ground level | | Silica full width at half maximum | |
|---|---|---|---|---|---|---|
| | | | **uncorrected** | **corrected** | **uncorrected** | **corrected** |
| | /nm | /nm | /nm | /nm | /nm | /nm |
| Probe1 | 116.0 ± 2.7 | 81.0 ± 1.0 | 194.5 ± 4.4 | 113.5 ± 4.4 | 151.9 ± 3.9 | 70.9 ± 3.9 |
| Probe2 | 115.8 ± 2.5 | 57.3 ± 3.1 | 174.3 ± 2.7 | 116.9 ± 2.7 | 135.4 ± 1.8 | 78.1 ± 1.8 |

## 5. Physical modelling for Small Angle X-Ray Scattering (SAXS)

A software tool was created to simulate the scattering from arbitrary form factors on the basis of the Debye scattering equation. This tool was then used to conduct a study to develop a simple protocol for the analysis of scattering curves for nanoparticle with complex shapes with a narrow size distribution. The chosen simulated shapes with increasing complexity were spheres, rods, cubes, octahedra and a hypothetical highly complex structure in the shape of a smurf, see figure 5.3.

The resulting scattering curves were then fitted with an ensemble of spheres using the Monte Carlo fitting algorithm implemented in McSAS. The resulting size distribution was then manually analyzed by a SAXS expert to guess the shape of the simulated ensemble. The expert was able to work out the simple shapes (spheres, rods, discs) from the size distribution, and correctly guessed the highly complex particles as a “complex shape,” but was unable to identify the cubes and octahedra. Based on this experience, a procedure for identifying the scattering from complex shapes was suggested. The whole procedure is depicted in the flow chart in the figure below.

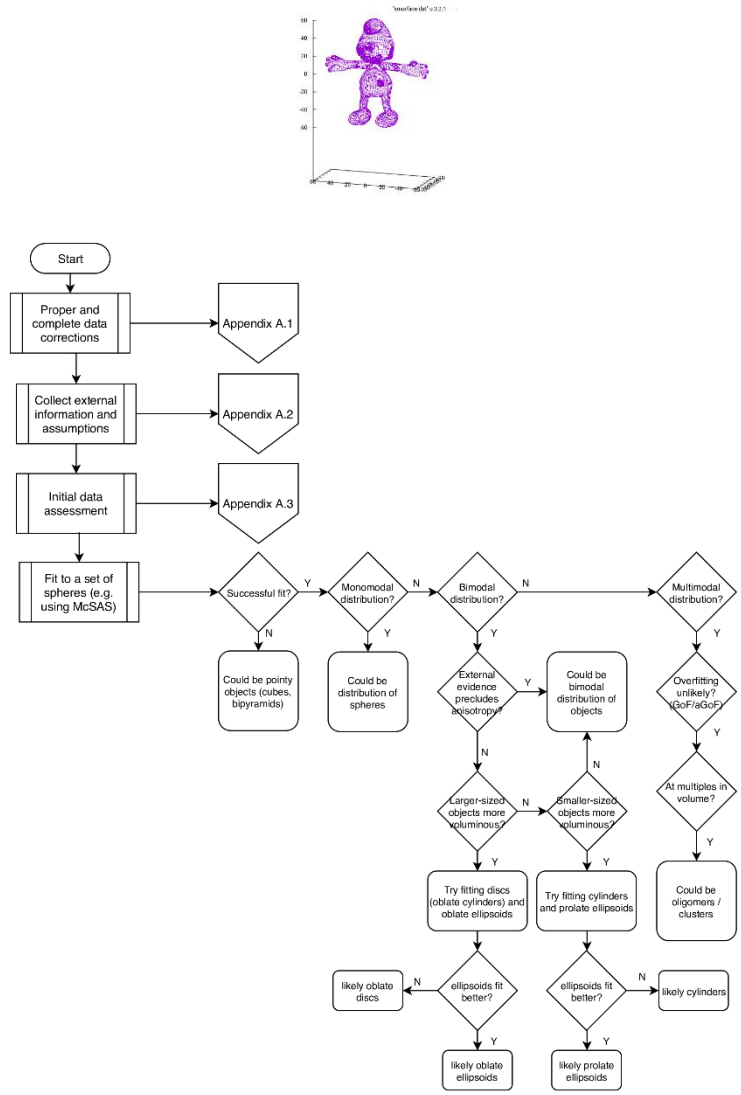



*Figure 5.3: Flow chart to determine the nanoparticle shape from the SAXS scattering curve, assuming narrow size distributions*

# 6. Machine Learning

## Background

Pollen's technologies aim to improve and fasten metrology for nanomaterials. As the design and shape of these objects varies at lot from customers and users, we propose a set of tools based on machine learning and deep learning. These solutions allow to automatize the metrology based on the expertise of the user. For the scope of this project, we develop deep learning techniques for metrology of nanoparticles.
We propose two complementary approaches:

- one for the detection of objects in the image
- a second for the analysis of the detected objects.

Before presenting deep learning techniques, it is useful to have in mind the classic techniques, or “non” deep learning techniques, to understand the added value of this latter: classic developments of new algorithms are driven by new use cases. I.e. if we want to analyze spherical particles of a certain type with a certain acquisition procedure, we need to develop an algorithm dedicated to this case. With a different case, a new algorithm needs to be developed as well. And so on.

Then, machine learning techniques began to gain interest as the core idea is to create an algorithm that is able to adapt itself to new data. The algorithm uses annotations: **image annotation is the human-powered task of annotating an image with labels**. Once the algorithm is set up, new cases can be managed thanks to new annotations. The main advantage of annotations is that they can be provided by users that are not expert in image processing and machine learning. Machine learning is working by extracting features from the data which are used then to perform the task (object detection, segmentation, ...). These features can be as simple as lines or edges, up to more complex such as wavelets or textures. In classic Machine Learning, the features are designed by humans which can be subjectivity depending on the variability of use cases we want to cover. Indeed, the designed features may contain bias and thus work better in some use cases.

Consequently, researchers started to develop more flexible machine learning with the deep learning principles. The idea behind deep learning techniques is to let the algorithm learn how to extract the features by itself.

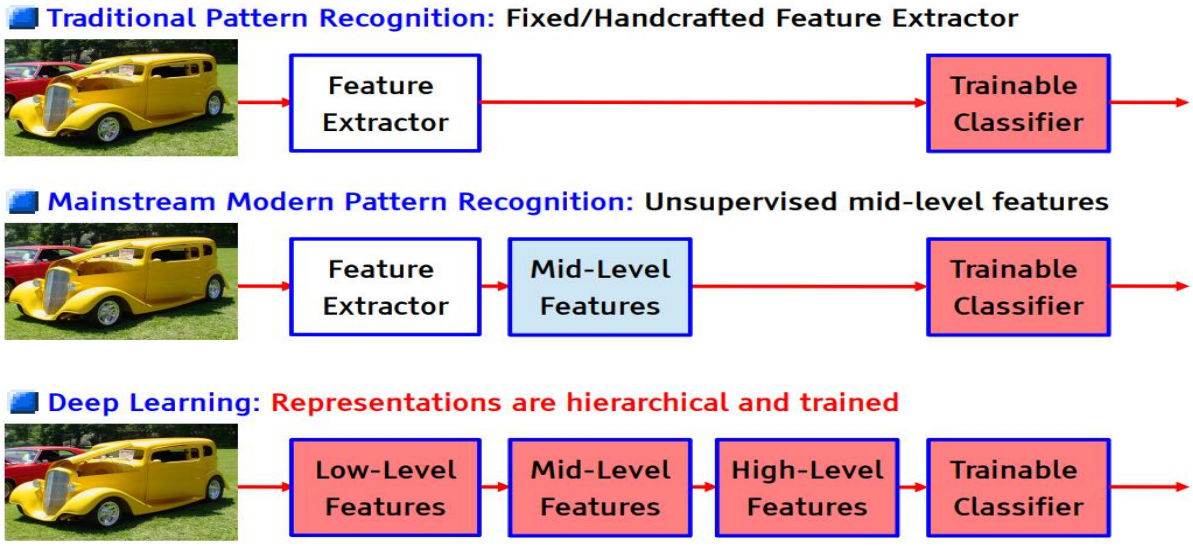



*Figure 6.1: Differences between classic machine learning and deep learning. Red boxes are part of the pipeline that are learnt[1].*

[1] Ref: https://www.college-de-france.fr/media/yann-lecun/UPL7915574462521283497_lecun_20160204_college_de_france_lecon_inaugurale.pdf

Deep learning aims to propose a set of operations that can be ordered as wanted to provide any kind of output. The list of challenges that can be solved by this way are:

- image processing with segmentation,
- object detection,
- classification,
- and also natural language processing, time series analysis...

The main common characteristic is the large quantity of data needed for the training of these models.

The first challenge of these models is the design of an architecture. The large variety of operations available is both an advantage (allowing to design any kind of architecture), but this large space of possibilities makes it difficult to create functional architectures.

The second challenge concerns the training of such models. Those have on average millions of parameters that can be tuned to provide a result. As any model, optimization is performed to find the optimal ones. But performing optimization in a such large space of parameters is difficult as the chances to fall into a local minimum are important.

In addition, the current theory around all this kind of optimization is still new and under construction as the field is moving fast.

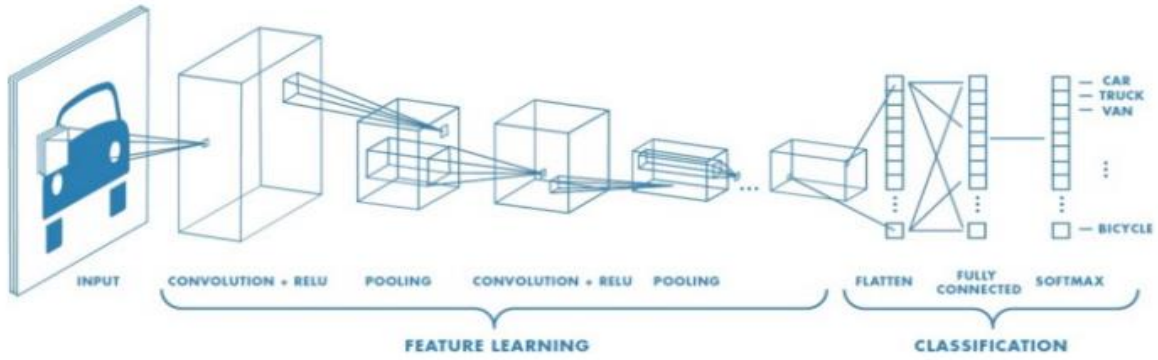



*Figure 6.2: Presentation of a deep learning pipeline.*[2]

To understand a better the working of neural networks, we will present some classic layers and how to assemble them. First there are two main concepts, neurons and layers.

- A layer is composed of a certain number of neurons
- and neurons are responsible to perform different operations.

Layers are then linked together to form a network. The first classic layer is the fully connected layer. The idea of this layer is to link each neuron from a layer to each other neuron to the next layer.
Another type of widely used layer is the convolutional layer. This time each neuron is a filter which is learnt such that convolutes the output of the previous layer to generate a set of filtered maps.

[2] Ref: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

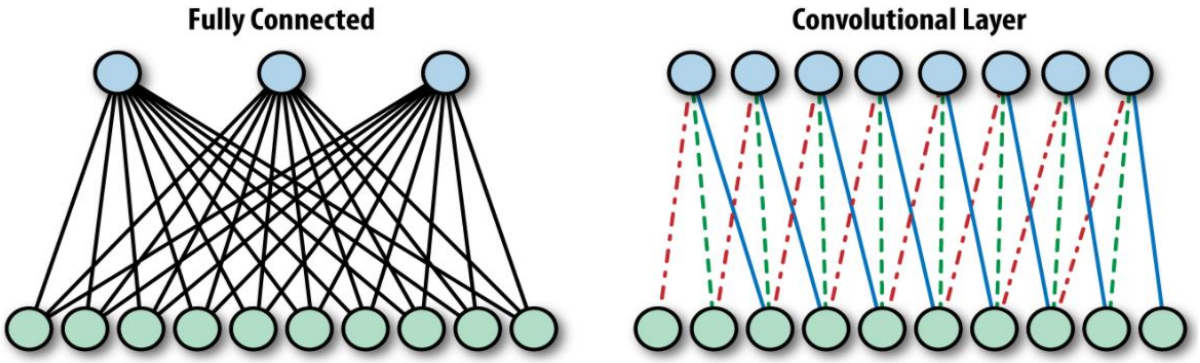


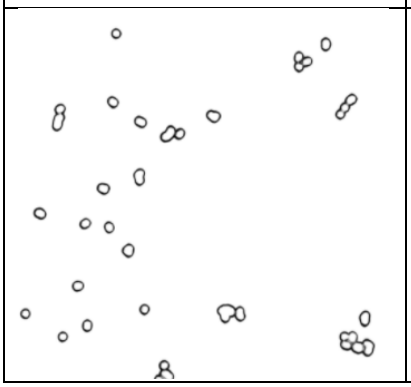
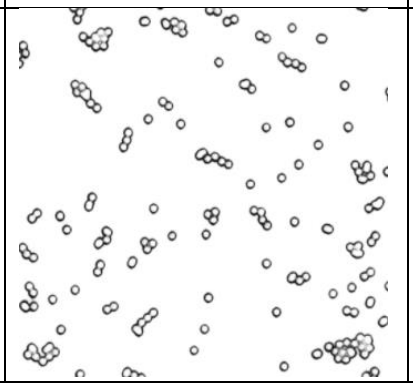
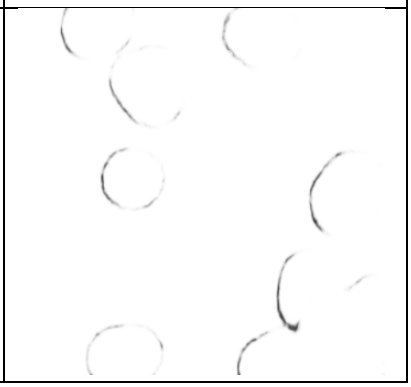

*Figure 6.3: Differences between fully connected and convolutional layers.*[3]

### Presentation of the algorithm developed by Pollen for nPSize

The goal of the object detection part is to provide the user a way to define his/her objects and be able to find it back in images. For this purpose, we developed a system where the user builds a database by selecting examples of objects to train a model. Classic approaches require at least thousands of examples to train a model, which is limiting for a user that cannot afford to annotate this amount of data. On the other hand, the user is not an expert in image processing thus (s)he doesn't want to tune a large number of parameters that have unknown effect (for her/him). For these reasons, our proposal lies on the need for the algorithm to be robust with a few examples or with a simple way to increase the database.

Our approach considers discrimination of the objects from background based on features extracted on the image. To build such a classifier, it is required to have both examples of objects and background.
Positive examples are easy to define for a user, but examples of background can be misleading. We propose the user an automatic way to retrieve negative examples in the image based on his annotation of the objects.
Then, regarding deep learning features, as most users have small annotated database, we propose a method that relies on deep learning features already trained instead of training them with the data of the user.

In the literature, it is widely admitted that such features are robust enough to propose good performances. Indeed, it is understandable that edges from cars and dogs images are not that far from edges in microscopy images. In addition, a line or a corner in natural images are similar to lines and corners from nanostructures. Then based on these features we perform the detection of the objects at different scales to capture objects with different sizes.

| nPSize11_Monomodal_silica_20%_SEM | nPSize10_Monomodal_silica_10%_SEM | nPSize10_Monomodal_silica_10%_TEM |
|---|---|---|
|  |  |  |

[3] Ref: https://www.oreilly.com/library/view/learning-tensorflow/9781491978504/ch04.html

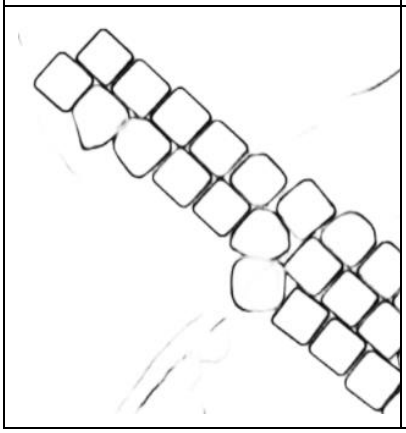
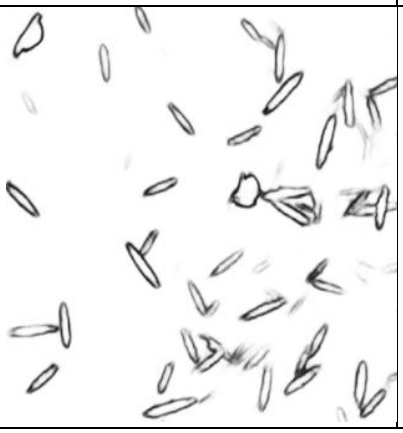
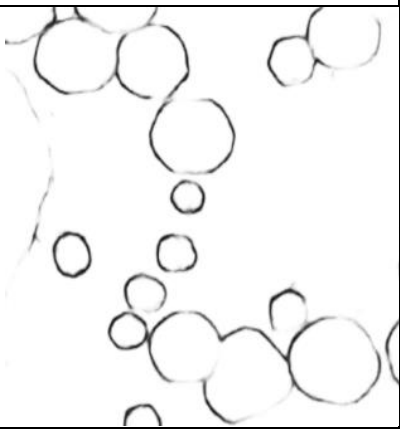
| nPSize4_Au_cube | nPSize6_Acicular_titania | nPSize1_BiModal_Gold |
|---|---|---|
|  |  |  |

*Figure 6.4: Use of already trained deep learning algorithm,*
*with detection of objects at different scales with different sizes*

Traditional approaches for the detection of edges are based on the computation of gradient or other transition functions. The issue with such approaches is that some edges will activate this definition but are not of interest. This is usually balanced by applying some priors on shapes.
With machine learning and more specifically deep learning, we can select the edges of interest through the annotation, so that only the edges that carry information on objects would be retrieved. In practice, this advantage is balanced by the fuzziness of the proposed contour. Because of the different steps of convolution, the response of the network to edges it not as clear as it is with mathematical models. We are going to apply post-processing steps in order to retrieve thinner contours corresponding to the real needs of metrology.
In the previous examples (see figure 6.4), the thickness of the edges returned by deep learning is more than 1 pixel, with the consequence that measurements can be degraded or variable.

## Conclusions and Outlook

All the components are in place for the next steps:

- use physical models to annotate (cf 6. Machine Learning) databases
- use annotated databases to run deep learning algorithms
- benchmark the different methods:
    - manual measurements
    - physical model measurements,
    - deep learning measurements

## References


1) Ref: https://www.college-de-france.fr/media/yann-lecun/UPL7915574462521283497_lecun_20160204_college_de_france_lecon_inaugurale.pdf
2) Ref: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148
3) Ref: https://www.oreilly.com/library/view/learning-tensorflow/9781491978504/ch04.html