Supervised Machine Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles

Janis Timoshenko[1], Deyu Lu[2], Yuewei Lin[3] and Anatoly I. Frenkel[1,4*]*

[1] Department of Material Science and Chemical Engineering, Stony Brook University, Stony Brook, NY 11794, USA

[2] Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA

[3] Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973, USA

[4] Division of Chemistry, Brookhaven National Laboratory, Upton, NY 11973, USA

[*] Corresponding authors. E-mails: Janis.timosenko@stonybrook.edu, anatoly.frenkel@stonybrook.edu

Supporting Information

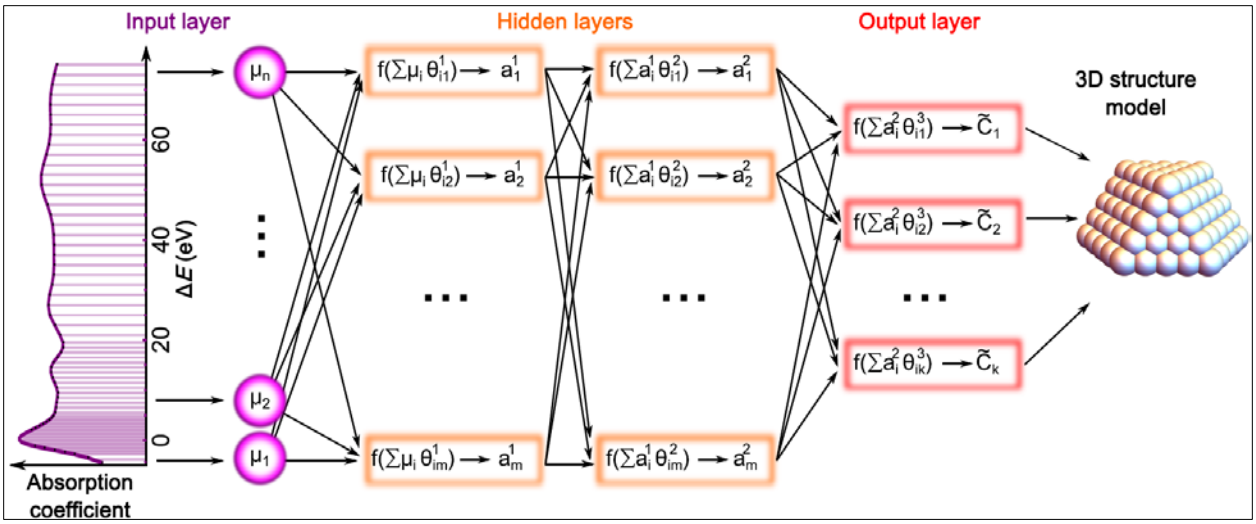

Figure S1 | Schematic representation of artificial neural network –based method for prediction of nanoparticle size and shape. Preprocessed XANES spectrum for a nanoparticle is discretized and used to set the values of neural network (NN) nodes in the *input layer*. The input is then processed in the NN *hidden layers*. Each hidden node represents a function, which adds together the node inputs, weighted with θ, applies to the sum the *activation function f*, and returns the result as a real number a to be used by NN nodes in the subsequent layers. Hidden layers map the input to the *output* – set of predicted average coordination numbers $\tilde{\boldsymbol{C}}_{\boldsymbol{i}}$. Average coordination numbers are then used to determine the size and shape of the nanoparticle.

Note 1. Details of sample preparation and experimental characterization

Experimental X-ray absorption spectroscopy (XAS) data for Pt nanoparticles (NPs) of different sizes and shapes supported on γ-Al_2O_3 were taken from Refs.[1-2] Pt NPs were prepared via inverse micelle encapsulation method[2] or by the incipient wetness method.[29] Narrow NPs size and shape distributions were confirmed with TEM, as explained in Refs.[1-2] XAS measurements were performed at National Synchrotron Light Source, Brookhaven National Laboratory, and Advanced Photon Source, Argonne National Laboratory. All XAS measurements were performed in He or H_2/He atmosphere at room temperature in transmission mode. Conventional analysis of EXAFS data for investigated samples was presented in Refs.[1-2]

Note 2. Details of ab initio calculations of XANES spectra

Pt L_3-edge XANES simulations were performed with two different codes: FEFF[3] and FDMNES.[4] Parameters for ab initio XANES simulations were chosen to ensure as good as possible agreements with experimental Pt foil XANES data.

FEFF version 9.6.4 was used for self-consistent calculations within full multiple scattering (FMS) approach and muffin-tin (MT) approximation. FMS clusters size was chosen at a large enough value so that the whole cluster is included in the FMS calculations. Default value for MT radius (1.5 Å) for Pt atoms was used. Complex exchange–correlation Hedin–Lundqvist potential was used. To model core-hole, random phase approximation (RPA) was used. No broadening, except due to the core-hole, was included in XANES calculations.

For FDMNES calculations we used FDMNES II program, revision 9th. Similarly as for FEFF calculations, we relied on FMS and MT approximations, since the finite-difference method, also implemented in FMDNES, did not result in a significantly better agreement with experimental data in

this case, and was more computationally expensive. Real Hedin–Lundqvist exchange–correlation potential, and no core hole approximation were used for FDMNES simulations.

Regular Pt NPs structure models with fcc-type structure and with octahedral, truncated octahedral, cuboctahedral and cubic shapes were constructed by cutting with (100) and (111) planes fcc-type Pt lattice with lattice constant a_0 =3.924 Å.[5] These regular structures were further truncated with an additional (100) or (111) plane to model the interface with substrate. Regular icosahedral and hcp-type clusters were constructed using cluster coordinates generator from Ref.,[6] maintaining the same nearest-neighbor distance 2.775 Å as in fcc-type clusters.

These cluster models were used for XANES simulations. Both for FDMNES and FEFF calculations site-specific XANES calculations for all non-equivalent sites in cluster models were performed. The total XANES spectrum was obtained as an average of site-specific spectra.[7-8] The calculated absorption spectra were shifted in energy by ΔE to align the energy scale used in theoretical calculations with the energy scale of experimental data. The values of ΔE for FEFF and FDMNES spectra were chosen as the ones that gave the best agreement between experimental and calculated Pt L_3-edge XANES for Pt foil, and were fixed for all further calculations. After the alignment, the theoretical and experimental spectra were re-interpolated on a non-uniform grid that spanned energies from E_{min} = 11561.5 eV to E_{max} = 11641.0 eV, with the step size of 0.2 eV for data points near absorption edge, which gradually increased up to 1.0 eV for points at $E = E_{max}$.

Note 3. Details of supervised machine learning implementation and training

In this study for analysis of XANES spectra we used regression with an off-the-shelf implementation of artificial neural network, provided by *Wolfram Mathematica* software (version 10).[9] SML hyperparameters (number of NN nodes, number of training examples, etc.)

were optimized to ensure optimal performance on validation data set (theoretical XANES data for NPs with different sizes and shapes). We have found that good results can be obtained with artificial neural network with 129 nodes (corresponding to the number of data points in the XANES spectra) in the input layer and two hidden layers with 339 - 387 nodes in each. Hyperbolic tangent function was used as an activation function, and default values for the regularization coefficients were used.

To establish relation between the features in averaged XANES $\mu(E)$ and average coordination numbers $\{C_1, C_2, C_3, \ldots\}$, we train artificial NN with theoretical XANES data, calculated with FEFF and FDMNES codes for Pt particles of different sizes and shapes. Considering that XANES calculations for each particle may require several CPU hours, it is impractical to use directly for NN training particle-averaged XANES data that correspond to specific particles. Note also that the difference between most of the particle-averaged coordination numbers (CNs), and, hence, also particle-averaged XANES spectra is small, especially for bigger particles. Therefore, to prevent NN overfitting, very large data set of particle-averaged spectra would be required. On the other hand, the sets of average coordination numbers $\{C_1, C_2, C_3, \ldots\}$ do not necessarily need to correspond to any realistic particle model during the training. Therefore, we use for NN training an artificial dataset, created from site-specific theoretical XANES spectra for a small ensemble of 18 particles of different sizes and shapes with a total of 218 non-equivalent Pt sites as shown in **Fig. S2a**. For each of the sites we perform both FEFF and FDMNES calculations, resulting in 436 site-specific spectra. Sets of site-specific coordination numbers $\{c_1, c_2, c_3, \ldots\}$ for each of those sites are known. To construct one training example, we pick randomly *n* of these sites, and create corresponding average spectrum as $\mu^i(E) = \sum_{j=1}^{n} \mu_j(E)/n$, where $\mu_j(E)$ are site-specific spectra calculated either with FEFF or

FDMNES for j-th of the randomly chosen sites. The corresponding average CNs set can be obtained as $\{C_1, C_2, C_3, \ ...\}^i = \sum_{j=1}^{n}\{c_1, c_2, c_3, \ ...\}_j \ /n$. Note that the selected n sites do not need to correspond to the same NP. We repeat this process N_t times to generate as many new training examples as required. Clearly, such approach allows us to obtain quickly an almost unlimited number of training examples, and requires much less computational efforts than the direct usage of particle-averaged XANES data for NN training. We also believe that such approach allows one to explore the configurational space much more efficiently than when particle-averaged XANES data are used for NN training, and allows us to achieve good performance with fewer number of training examples. For our purposes, we have found that a good performance of NN can be achieved with n = 3 and N_t in the range between 200,000 and 600,000.

Importantly, unlike it was for NN training, for the validation of our NN we use **particle-averaged** XANES spectra, corresponding to Pt NPs of specific size and geometry. For validation we used structure models that were indirectly used for generation of training data set (**Fig.S2a**), as well as particles of different sizes and shapes (**Fig.2a** in the main text). Good accuracy yielded by our NN in the determination of CNs for such validation dataset proves the validity of our training approach.

In addition, we have found that the systematic errors of NN predictions can be reduced significantly, if instead of using XANES spectra $\mu(E)$ directly as input for NN we provide $\Delta\mu(E) = \mu(E) - \mu_{bulk}(E)$, where μ_{bulk} is XANES spectrum for bulk material. Of course, we subtract bulk data both from the spectra used for NN training as well as from the spectra that are later analyzed with the trained NN. For theoretical spectra calculated with FDMNES and FEFF $\mu_{bulk}(E)$ is obtained in, correspondingly, FDMNES and FEFF calculations for bulk Pt. For experimental XANES spectra as μ_{bulk} we use experimental XANES spectrum for Pt foil. A

small subset of $\Delta\mu^i(E)$ spectra, used for NN training, is shown in **Fig. S2b** along with the corresponding average CNs $\{C_1, C_2, C_3, C_4\}^i$.

We used 4 separate neural networks to predict the 1st, 2nd, 3rd and 4th coordination numbers. These 4 NNs were completely independent, except that they were trained on the same sets of training data. 200 000 – 600 000 training examples were used for training. We repeated the NNs training several (five) times with different sets of training examples and different numbers of hidden nodes to estimate the uncertainties of predicted coordination numbers. The coordination numbers, reported here and in the main text, correspond to average prediction of all five neural networks.

We have also tried other supervised machine learning (SML) approaches, as supported vector regression, ridge regression and artificial neural networks with different architectures as implemented in *TensorFlow*[10] package. All 4 CNs were predicted simultaneously in these cases within one SML routine. All these methods yielded results comparable with the ones obtained with neural networks, implemented in *Mathematica*.

We have found that reasonable results can be obtained using even the simplest SML method - multiple linear regression.[11] The accuracy of CNs prediction for validation data set, as demonstrated by linear regression method implemented in *Mathematica*, is just slightly worse than that for NN method. Training data sets of the same size were used to train both NN and linear regression routines. The advantage of NN can be demonstrated by comparing root-mean square errors (RMSE), yielded by NN (0.7, 0.5, 2.1, 2.2 for the 1st, 2nd, 3rd and 4th CNs, respectively) and linear regression method (0.8, 1.0, 5.3, 4.1, respectively). Better performance of NN is a result of enhanced sensitivity to the non-linear relations between features in XANES

spectra and coordination numbers, ensured by nodes in the hidden layers of NN that are not present in linear regression method.

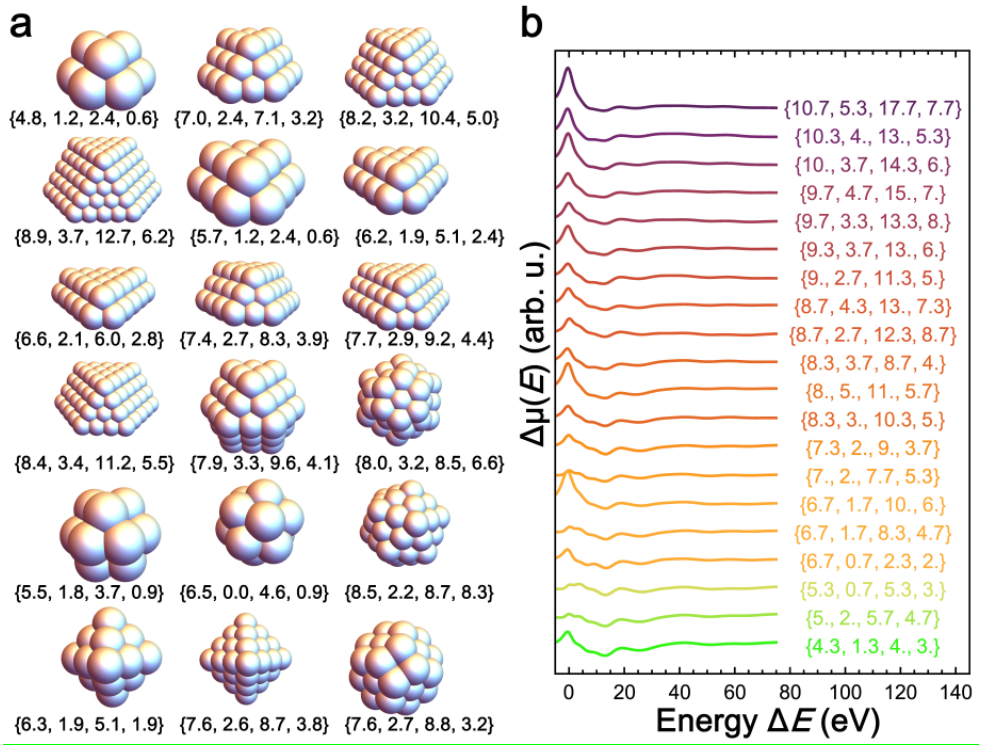

Figure S2. | **Cluster models and theoretical XANES data for neural network training.** Pt NPs used to construct training data set are shown in (**a**). Linear combinations of site-specific spectra, calculated for these NPs with FEFF[3] or FDMNES[4] codes, were used to construct training data set, a small subset of which is shown in (**b**) together with the corresponding averaged CNs. Bulk Pt XANES is subtracted from each spectrum, and spectra are shifted vertically for clarity.

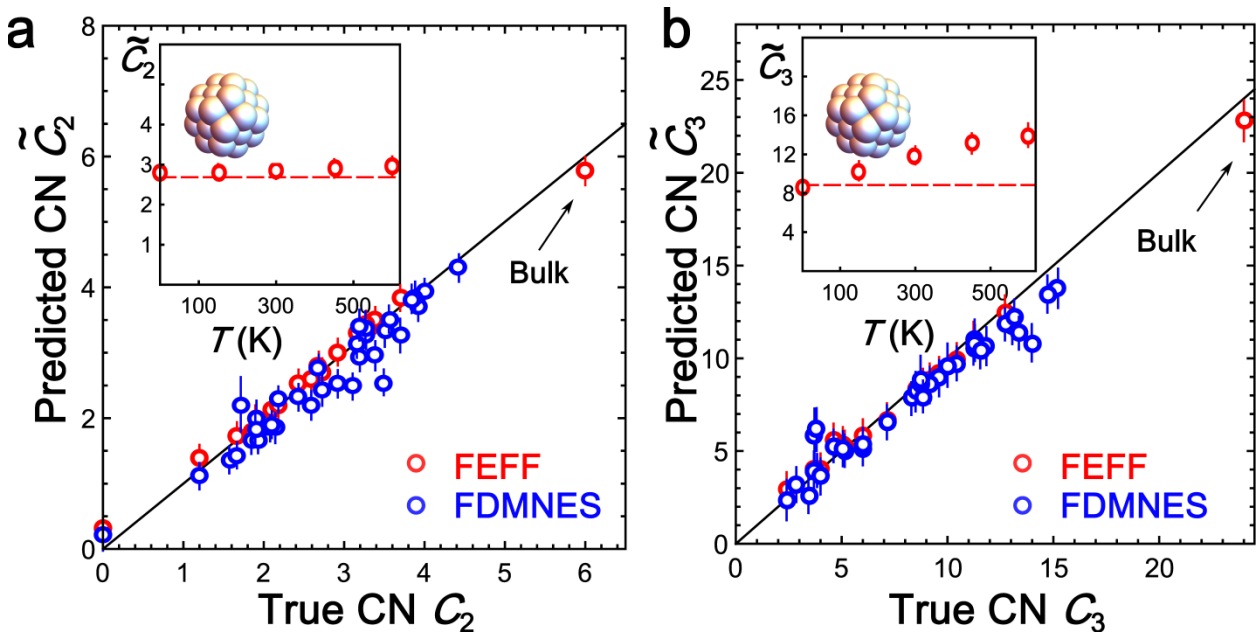

Figure S3 | Validation of neural network using theoretical XANES data (2nd and 3rd coordination shells). True coordination numbers for the 2nd **(a)** and 3rd **(b)** coordination shell for particles, shown in **Fig.S2a** and **Fig.2a (**in the main text), are compared with the CNs, predicted by neural network from XANES data, generated by FEFF[3] or FDMNES[4] codes. Solid lines – guides for eye. In the insets we show the coordination numbers, predicted by NN for truncated octahedral particle with 38 atoms from FEFF-generated XANES data, where thermal disorder was introduced using Debye model[12] in the temperature range between 0 and 600 K. Debye temperature was set to bulk Pt value 244 K.[1] Horizontal dashed lines show the true value of corresponding CN.

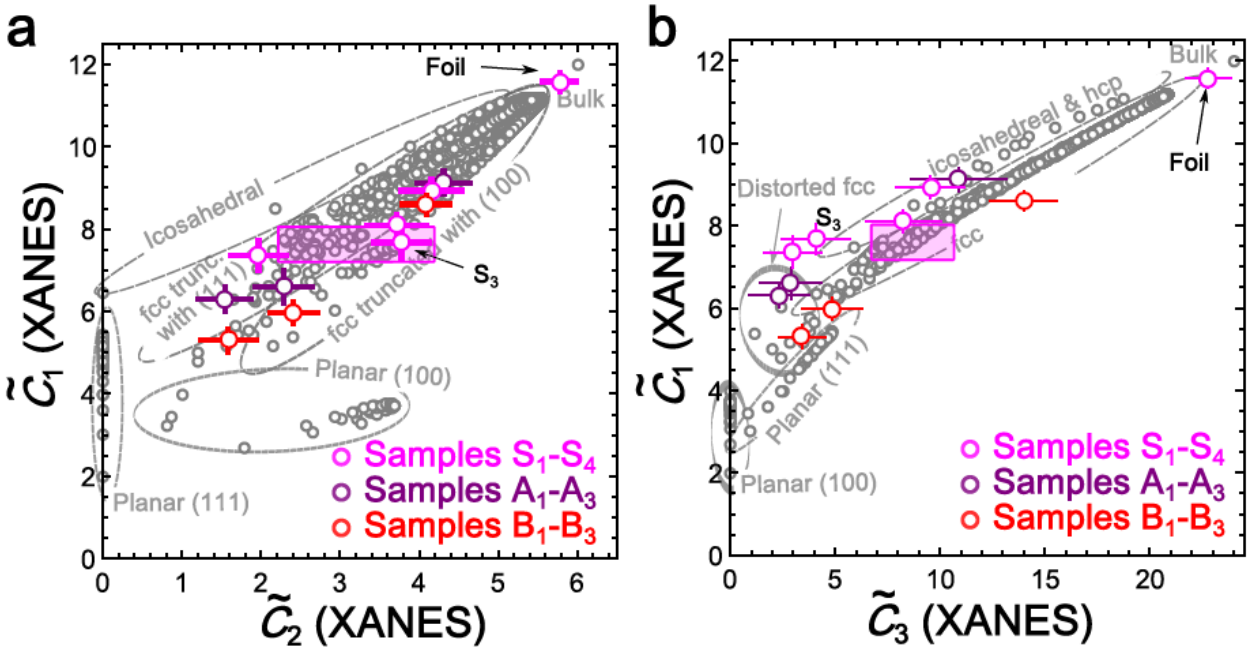

Figure S4 | Coordination numbers predicted by neural network from experimental XANES[1-2] (2nd and 3rd coordination shells). Coordination numbers, predicted by NN method for the 1st and 2nd **(a)** and 1st and 3rd **(b)** coordination shells are shown. Gray empty circles correspond to CNs for Pt model clusters with different sizes and shapes, obtained from fcc-type Pt structure, truncated along (100) or (111) planes, as well for clusters with icosahedral and hcp-type structures. Magenta rectangles show the confidence regions for CNs, obtained for sample S_3 from multiple-scattering EXAFS analysis in Ref.[1]

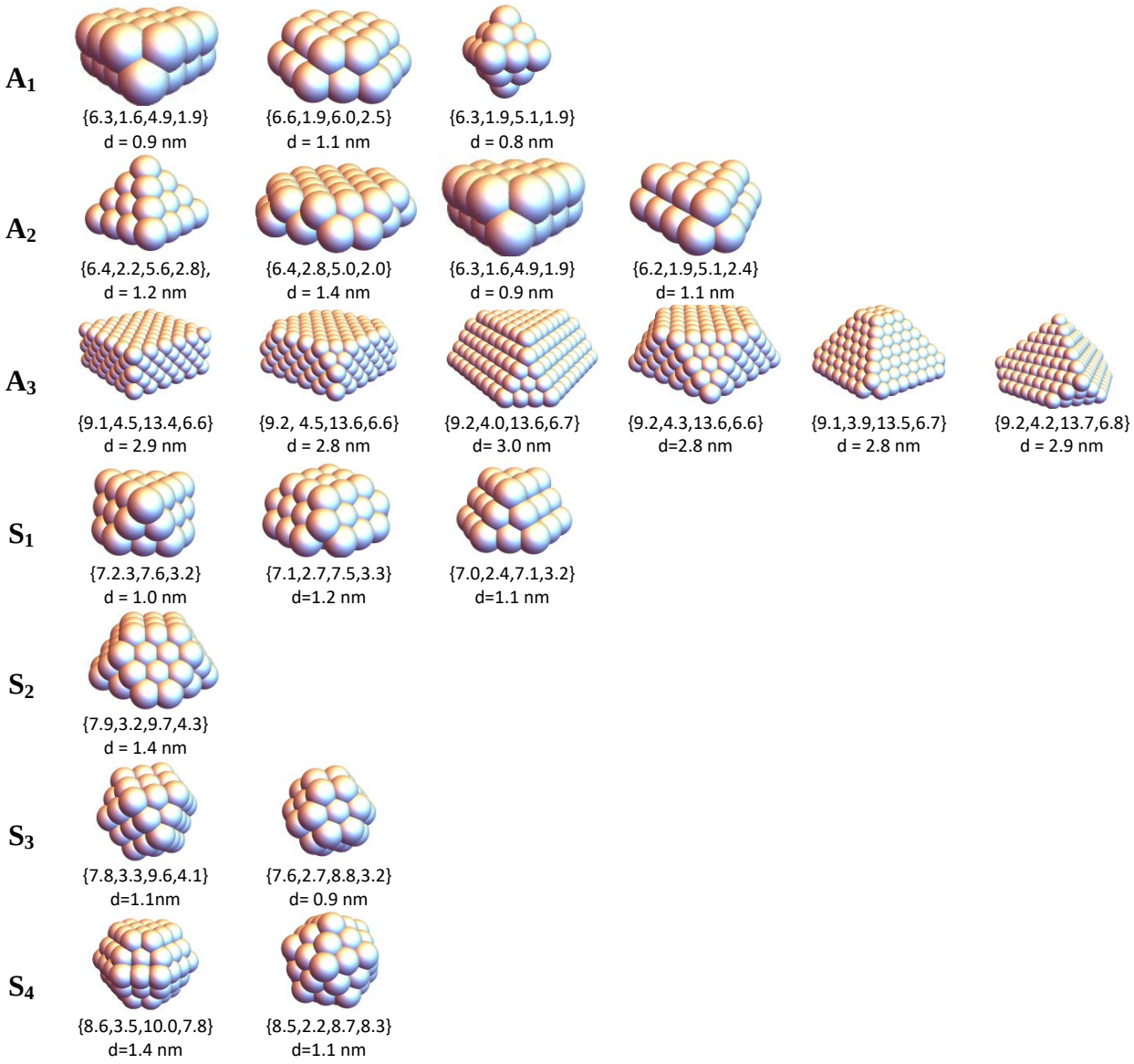

Figure S5 | Additional examples of structure models that have CNs and sizes close to those, estimated by neural network-based analysis of experimental XANES data and TEM analysis for Pt NPs on γ-Al_2O_3. CNs for the first 4 coordination shells (given in curly brackets) and particle size *d* for each of the shown models are close to those for structure models in **Fig.3c** in the main text. For sample A_3 only a small subset of possible (similar) structure models with close sizes and coordination numbers is shown.

Note 4. Details of RMC-EXAFS analysis

For validation of structure models, obtained in NN/XANES analysis, we used RMC-EXAFS method with evolutionary algorithm (EA), as implemented in EvAX code.[13] We represent each sample as a small ensemble of NPs with the shape corresponding to NN/XANES results. Number of particles in each ensemble is chosen to have in total 300 – 500 atoms in the model. During RMC analysis, such composite model evolves via small random displacements of all atoms, with the aim to improve the agreement between experiment and theoretical EXAFS data. We constrain the maximal allowed displacements for atoms from their initial positions to be 0.4 Å to preserve the overall shape of particles.

For calculations of theoretical EXAFS spectra we use FEFF8.5L code.[14] To compare the calculated and experimental EXAFS data, we use Morlet wavelet transform[15-16] (WT): WTs of experimental and theoretical EXAFS are calculated, and the disagreement between experiment and theory is characterized as Euclidean distance between obtained WTs in the *k*-range between 3 and 15 $Å^{-1}$ and in *R*-range between 1.5 and 6 Å, thus including in the analysis contributions from the first 6 coordination shells and multiple-scattering contributions with up to 4^{th} order. In details RMC-EXAFS procedure is described in Ref.[17]

If the initial structure model is close to the structure of investigated material, a good agreement between experimental EXAFS and RMC simulations results can be expected. If, in turn, the chosen structure model does not represent the structure of the material, agreement between experimental data and theory will be noticeably worse.[17-18] In **Fig. S6** we demonstrate the results of such RMC-EXAFS simulations for structure models that were established based on our NN/XANES method. In addition, for sample S_2 we carried out additional RMC-EXAFS

simulation with structure models that had the 1st shell CN close to the one, predicted by NN/XANES method, but significantly different 2nd, 3rd and 4th CNs. As shown in **Fig.S6**, while the models that agreed with the predictions of NN/XANES method resulted in a convincing agreement with the experimental EXAFS data, significantly worse agreement was obtained with the structure model that does not agree with the NN/XANES results, especially in the *R* range between 4 and 6 Å, corresponding to the contributions of distant coordination shells and multiple-scattering effects.

A good agreement obtained between experimental and theoretical EXAFS data after RMC simulations indicates that the structure models, yielded by NN/XANES method, indeed agree well with the available experimental EXAFS data. Besides structure model validation, RMC-EXAFS simulations with the structure models derived from NN/XANES analysis can be used to reconstruct the bond length distribution functions for the investigated samples and, hence, to extract information on average interatomic distances and disorder effects,[17] which is not readily accessible from our NN/XANES method.

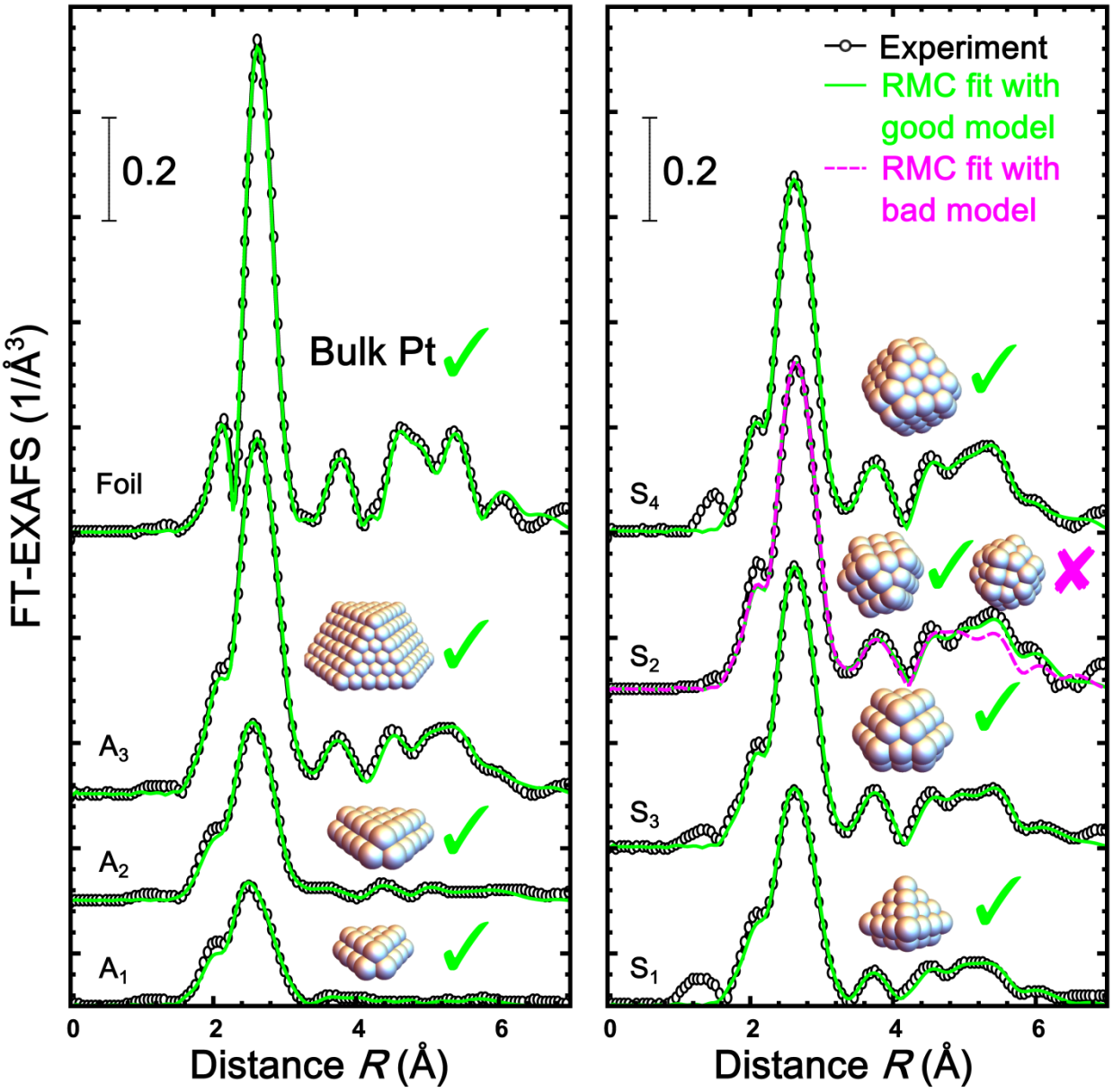

Figure S6 | Validation of Pt NPs 3D shapes, determined by the NN/XANES method, with RMC-EXAFS simulations. 3D shapes of particles ("good models"), corresponding to the coordination numbers (CNs), obtained in NN/XANES analysis were used as initial structure models for RMC-EXAFS simulations. Structure models, distorted by RMC method, are shown in the insets (note that they are almost indistinguishable from the original structure models, shown in **Fig. 3c** in the main text). Agreement between Fourier-transformed (FT) experimental EXAFS data and theoretical EXAFS, obtained in RMC simulations, is shown. In addition, for sample S_2 we carried out RMC simulations with the structure models ("bad model") that had the 1st shell CN as predicted by NN/XANES method, but significantly different 2nd, 3rd and 4th CNs (the used hcp Pt_{57} particle has CNs $\{\mathbf{8.0}, \mathbf{3.2}, \mathbf{8.5}, \mathbf{6.6}\}$). As shown with the dashed magenta lines, RMC-EXAFS spectrum for this model differs noticeably from experimental spectrum in the *R* range between 4 and 6 Å. FT-EXAFS spectra are shifted vertically for clarity.

References

(1) Cuenya, B. R.; Frenkel, A.; Mostafa, S.; Behafarid, F.; Croy, J.; Ono, L.; Wang, Q. Anomalous lattice dynamics and thermal properties of supported size-and shape-selected Pt nanoparticles. *Phys. Rev. B* **2010,** *82*, 155450.
(2) Sanchez, S. I.; Menard, L. D.; Bram, A.; Kang, J. H.; Small, M. W.; Nuzzo, R. G.; Frenkel, A. I. The emergence of nonbulk properties in supported metal clusters: negative thermal expansion and atomic disorder in Pt nanoclusters supported on γ-Al_2O_3. *J. Am. Chem. Soc.* **2009,** *131*, 7040-7054.
(3) Rehr, J. J.; Kas, J. J.; Vila, F. D.; Prange, M. P.; Jorissen, K. Parameter-free calculations of X-ray spectra with FEFF9. *Phys. Chem. Chem. Phys.* **2010,** *12*, 5503-5513.
(4) Bunău, O.; Joly, Y. Self-consistent aspects of X-ray absorption calculations. *J. Phys.: Condens. Matter* **2009,** *21*, 345501.
(5) Edwards, J. W.; Speiser, R.; Johnston, H. L. High temperature structure and thermal expansion of some metals as determined by X-ray diffraction data. I. Platinum, tantalum, niobium, and molybdenum. *J. Appl. Phys.* **1951,** *22*, 424-428.
(6) Glasner, D.; Frenkel, A. I.; Hedman, B.; Pianetta, P. Geometrical characteristics of regular polyhedra: application to EXAFS studies of nanoclusters. *AIP Conf. Proc.* **2007,** *882*, 746-748.
(7) Ankudinov, A.; Rehr, J.; Low, J. J.; Bare, S. R. Sensitivity of Pt X-ray absorption near edge structure to the morphology of small Pt clusters. *J. Chem. Phys.* **2002,** *116*, 1911-1919.
(8) Bazin, D.; Sayers, D.; Rehr, J.; Mottet, C. Numerical simulation of the platinum LIII edge white line relative to nanometer scale clusters. *J. Phys. Chem. B* **1997,** *101*, 5332-5336.
(9) *Mathematica*, Wolfram Research, Inc.: Champaign, Illinois, 2014.
(10) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M. *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*, 2015, http://tensorflow.org/.
(11) Bhadeshia, H. K. D. H. Neural networks in materials science. *ISIJ Int.* **1999,** *39*, 966-979.
(12) Rehr, J.; Ankudinov, A. Progress in the theory and interpretation of XANES. *Coord. Chem. Rev.* **2005,** *249*, 131-140.
(13) Timoshenko, J.; Kuzmin, A.; Purans, J. EXAFS study of hydrogen intercalation into ReO_3 using the evolutionary algorithm. *J. Phys.: Condens. Matter* **2014,** *26*.
(14) Ankudinov, A. L.; Ravel, B.; Rehr, J. J.; Conradson, S. D. Real-space multiple-scattering calculation and interpretation of X-ray-absorption near-edge structure. *Phys. Rev. B* **1998,** *58*, 7565-7576.
(15) Timoshenko, J.; Kuzmin, A. Wavelet data analysis of EXAFS spectra. *Comp. Phys. Commun.* **2009,** *180*, 920-925.
(16) Timoshenko, J.; Kuzmin, A.; Purans, J. Reverse Monte Carlo modeling of thermal disorder in crystalline materials from EXAFS spectra. *Comp. Phys. Commun.* **2012,** *183*, 1237-1245.
(17) Timoshenko, J.; Frenkel, A. I. Probing structural relaxation in nanosized catalysts by combining EXAFS and reverse Monte Carlo methods. *Catal. Today* **2017,** *280*, 274-282.
(18) Timoshenko, J.; Anspoks, A.; Kalinko, A.; Kuzmin, A. Local structure of nanosized tungstates revealed by evolutionary algorithm. *Phys. Status Solidi A* **2015,** *212*, 265-273.