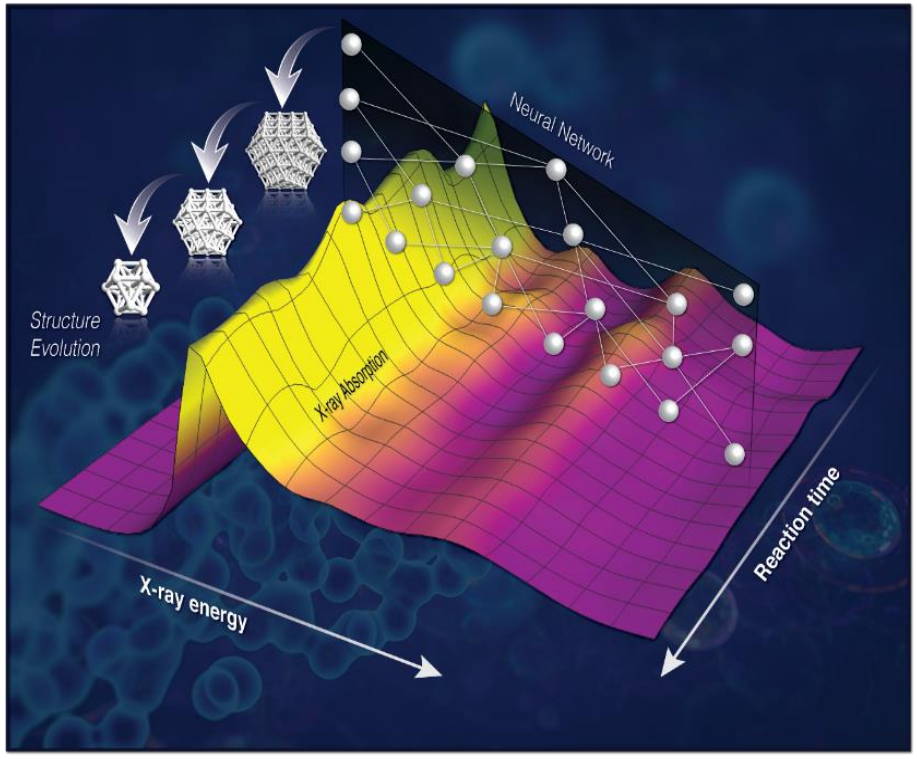


Machine Learning Methods

(for XAS data interpretation)

Department of Interface Science

Fritz-Haber Institute of the Max Planck Society, Berlin

Instrumentation:

1964 first users of hard X-ray synchrotron radiation (DESY in Hamburg)

Data analysis:

1965 Fast Fouier Transform (Cooley & Tukey)

Computing:

1970s First personal computers

EXAFS as a practical tool:

1970-1975 works by Sayers, Lyttle, and Stern: EXAFS equation, FT-EXAFS

Instrumentation:

1987 First experiments in QXAFS mode (DESY in Hamburg)

Data analysis:

1989 Backpropagation algorithm, deep machine learning (Yann LeCun)

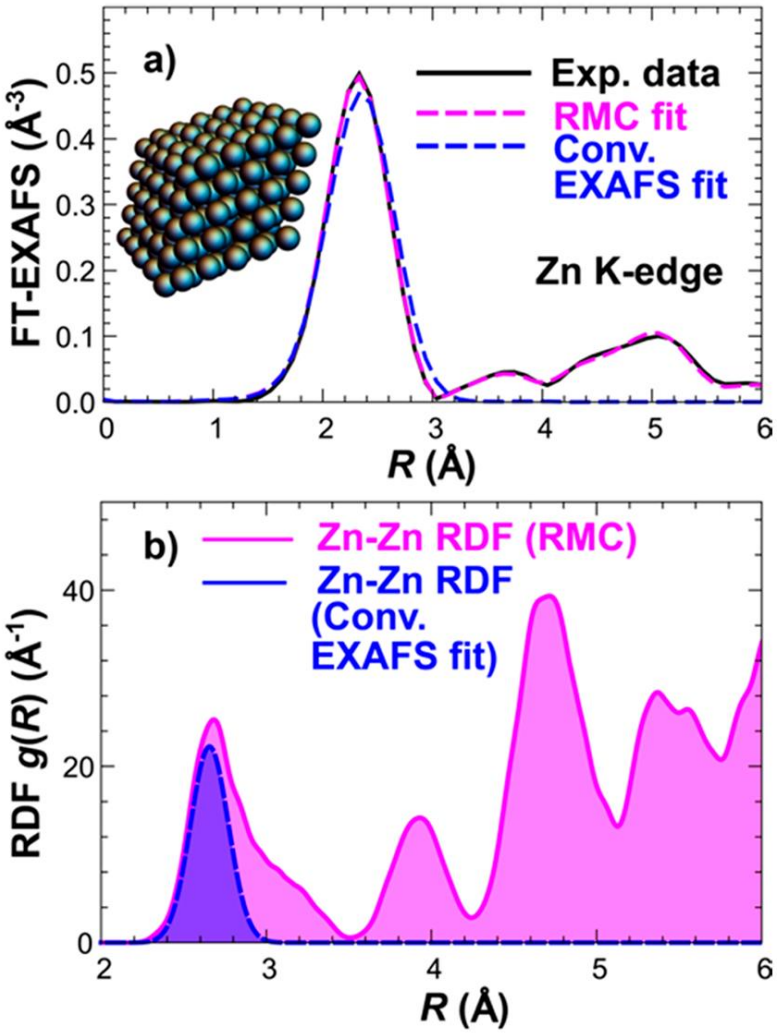
Computing:

1990s GPUs, high performance computing systems

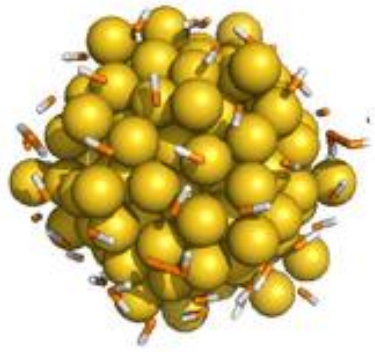
Machine-learning-driven transformation of XAS research field?

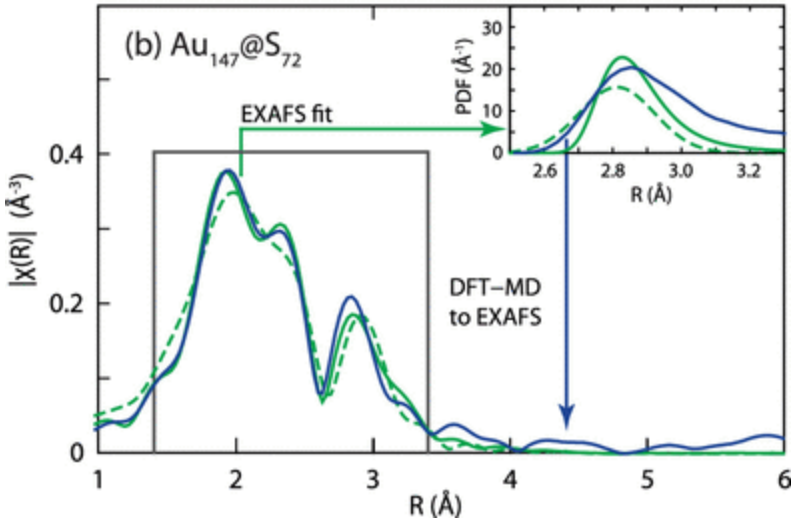
Why do we need advanced XAS analysis?

1: Materials are disordered



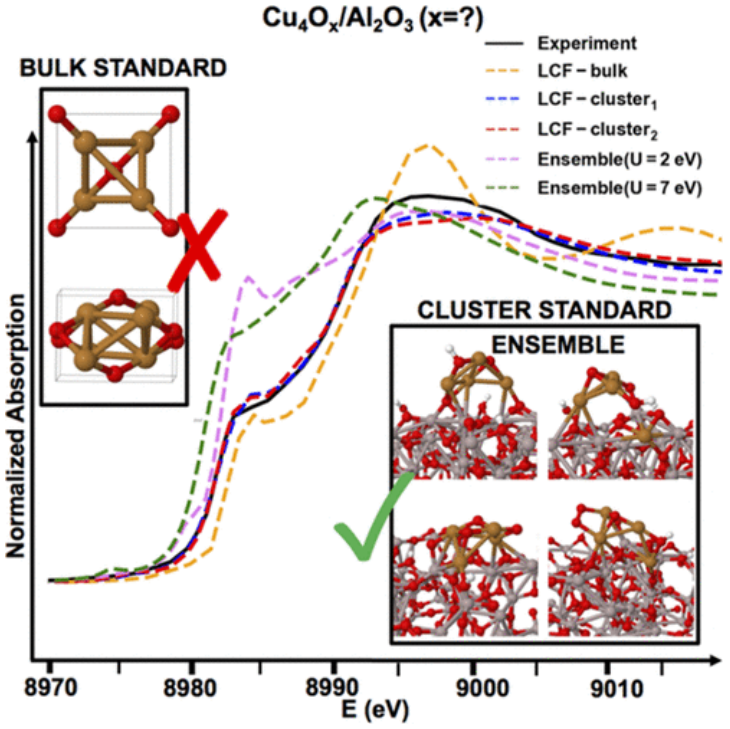
Timoshenko & Roldan
Chem. Rev. 121 (2021)





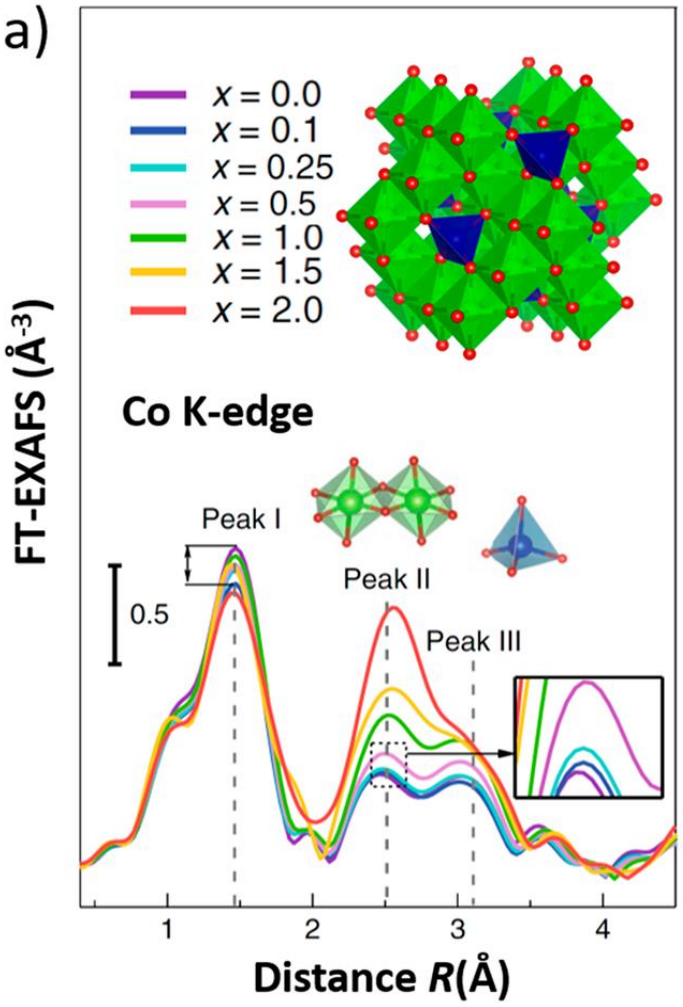
Chill, Frenkel et al
ACS Nano 9 (2015)

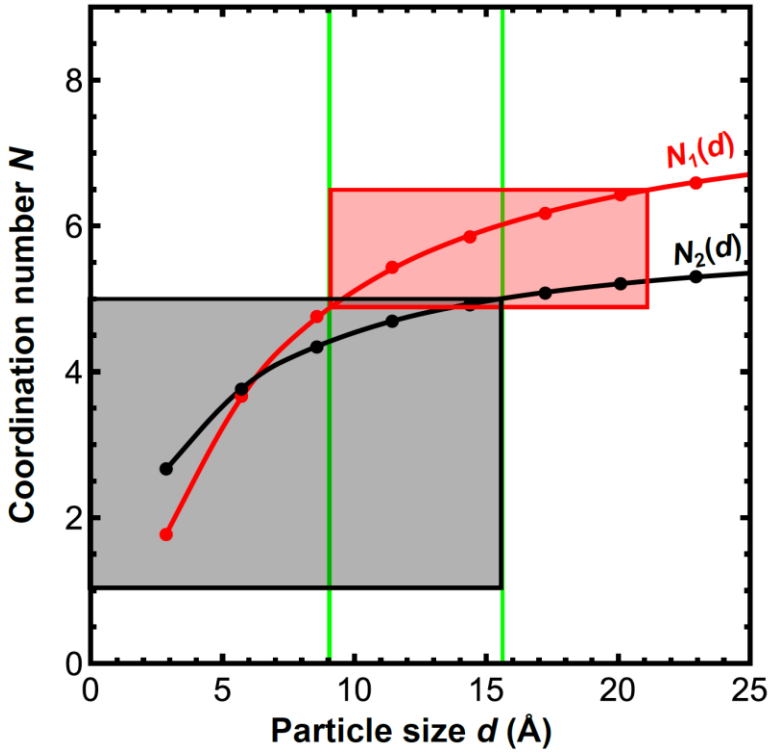
EXAFS equation is not accurate for many realistic materials with static and thermal disorder

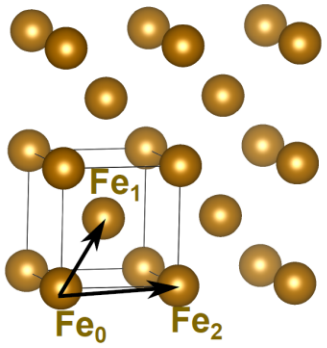


Zandkarimi, Alexandrova et al
J. Phys. Chem. C 124 (2020)

2: There is relevant information beyond the first FT peak





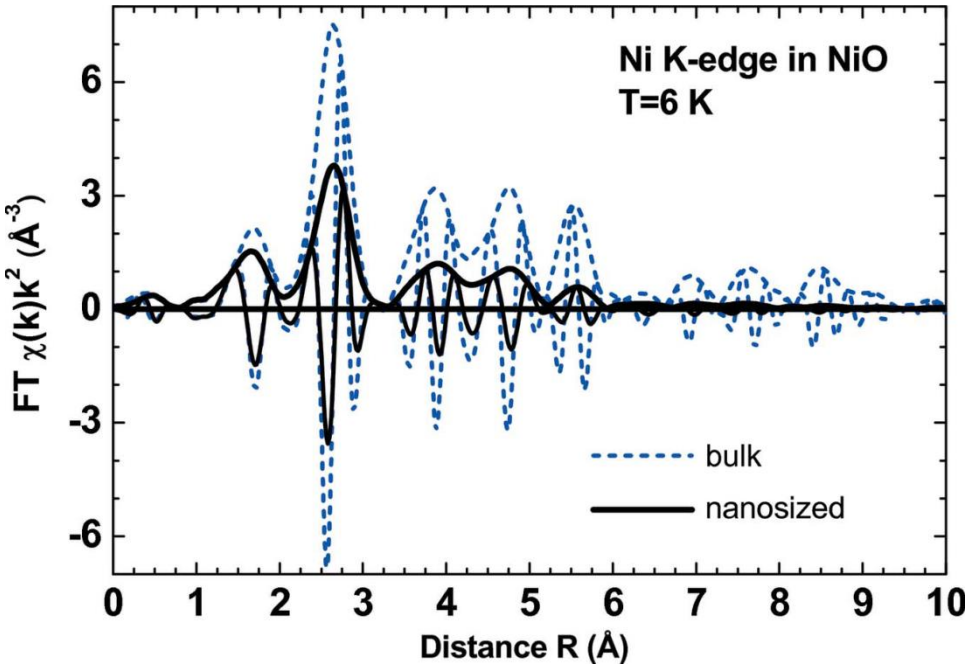


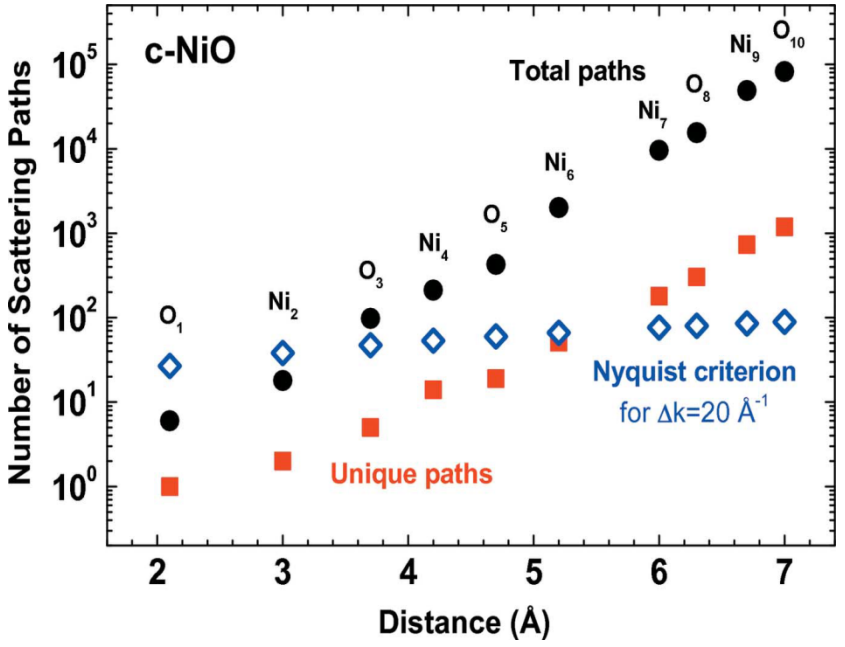
EXAFS fitting beyond the 1st/2nd coordination shell is very difficult

Wu et al Nat. Catal. 2 (2019)

Varnell, Timoshenko et al. Nat. Commun. 7 (2016)

2: There is relevant information beyond the first FT peak

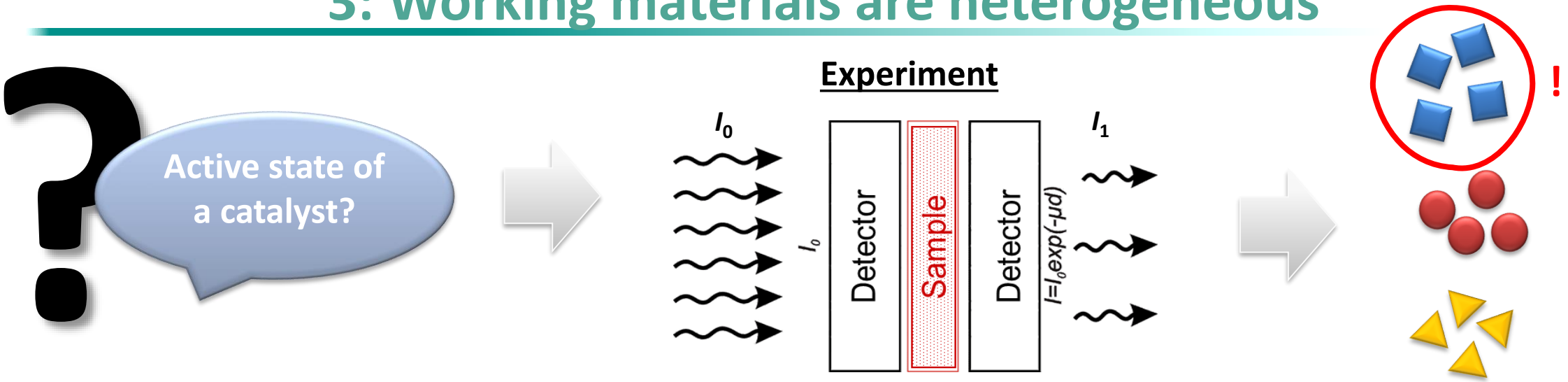




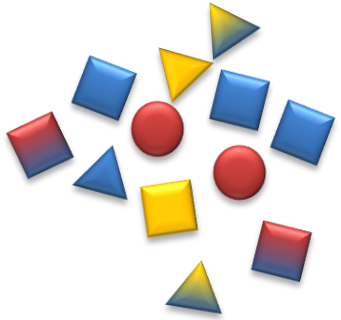
A. Kuzmin and J. Chaboy, IUCrJ 1 (2014) 571-589.

EXAFS fitting beyond the 1st/2nd coordination shell is very difficult

3: Working materials are heterogeneous



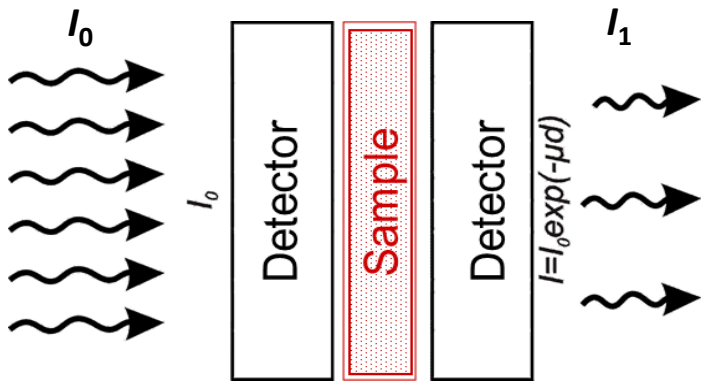
Reality:

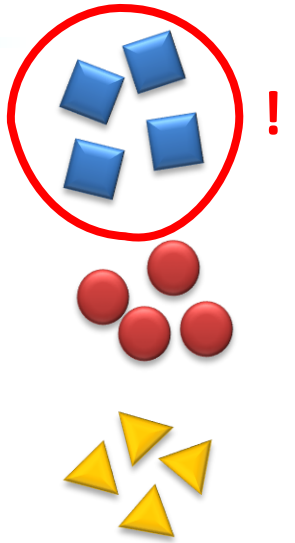


3: Working materials are heterogeneous

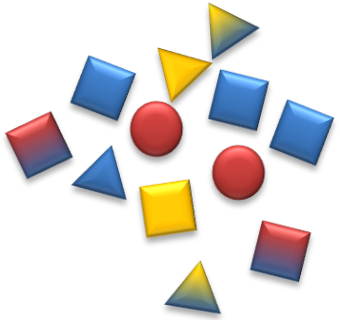
Active state of a catalyst?

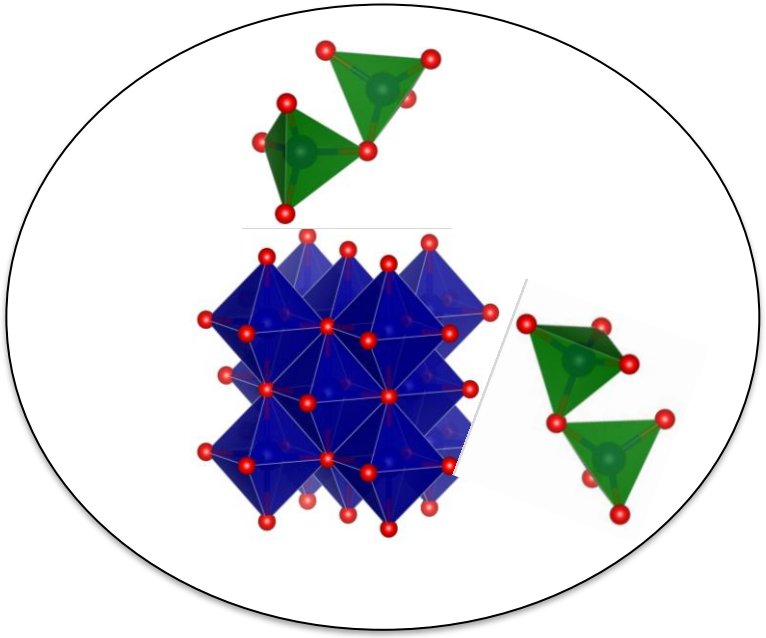
Experiment

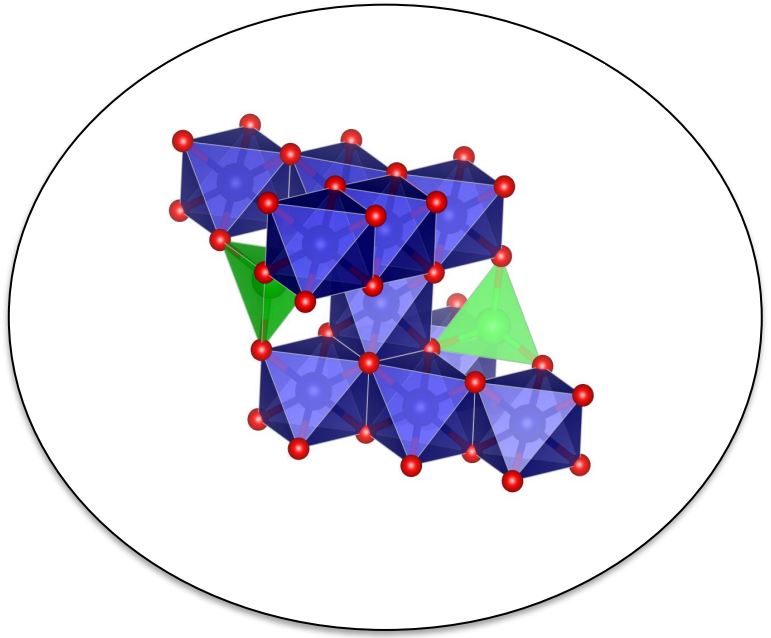




Reality:



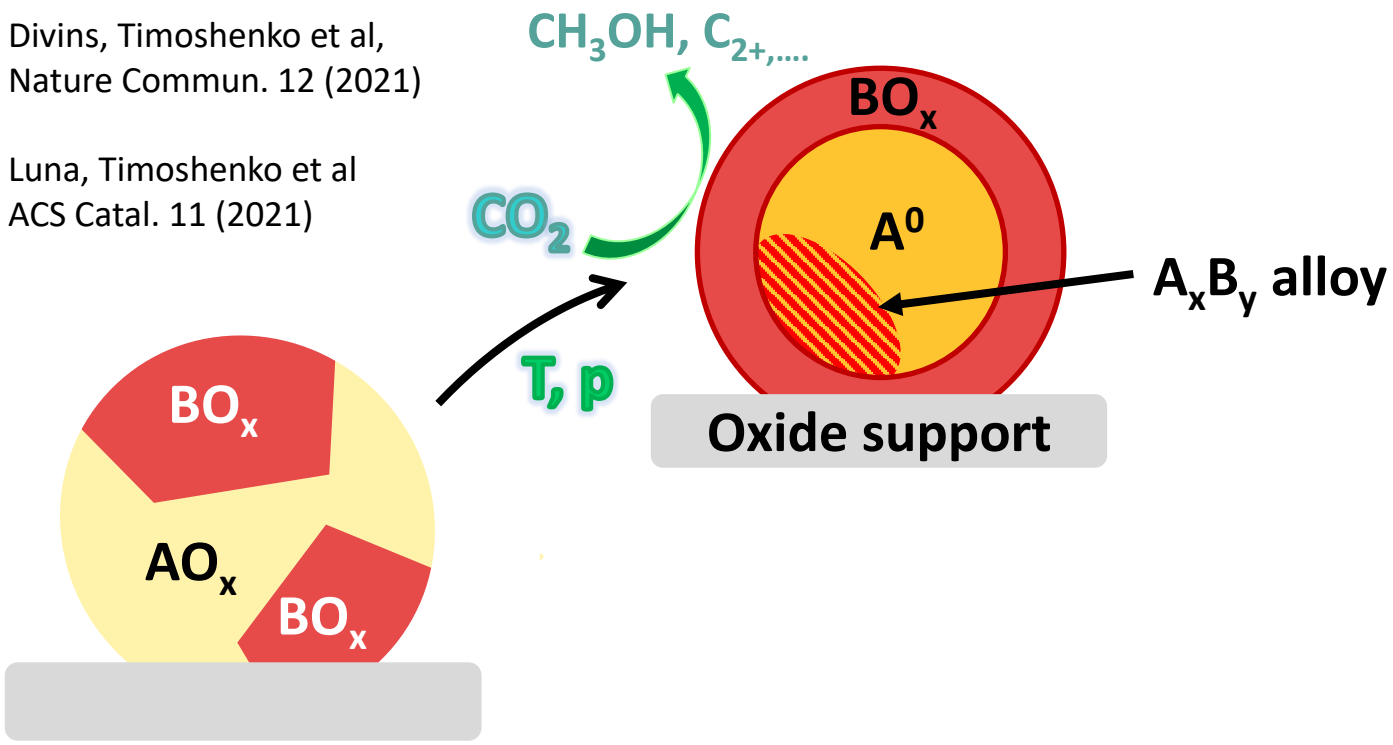




3: Working materials are heterogeneous

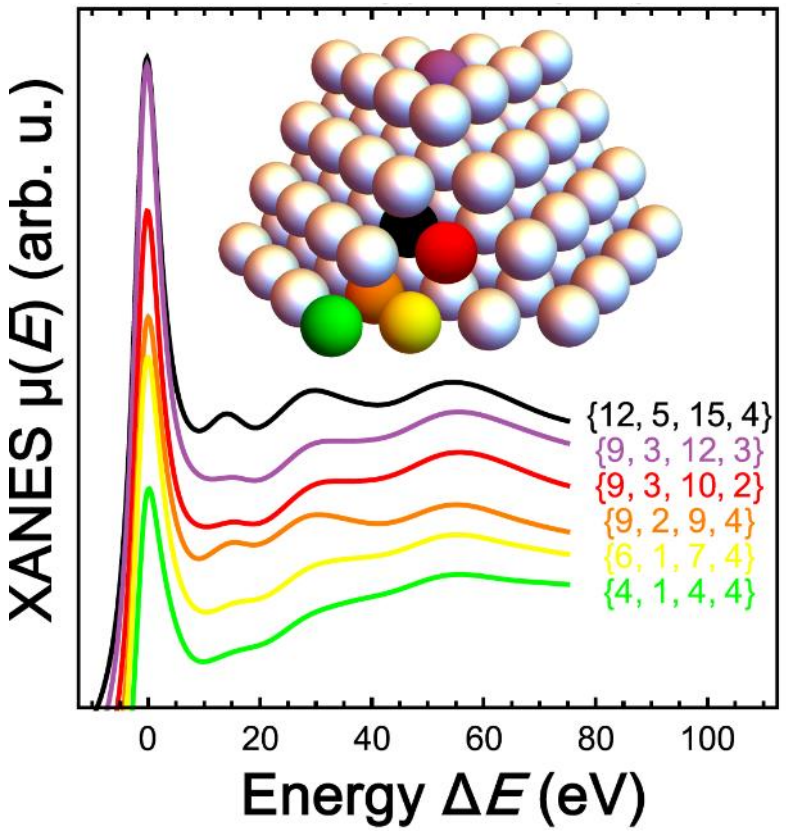
Divins, Timoshenko et al,
Nature Commun. 12 (2021)

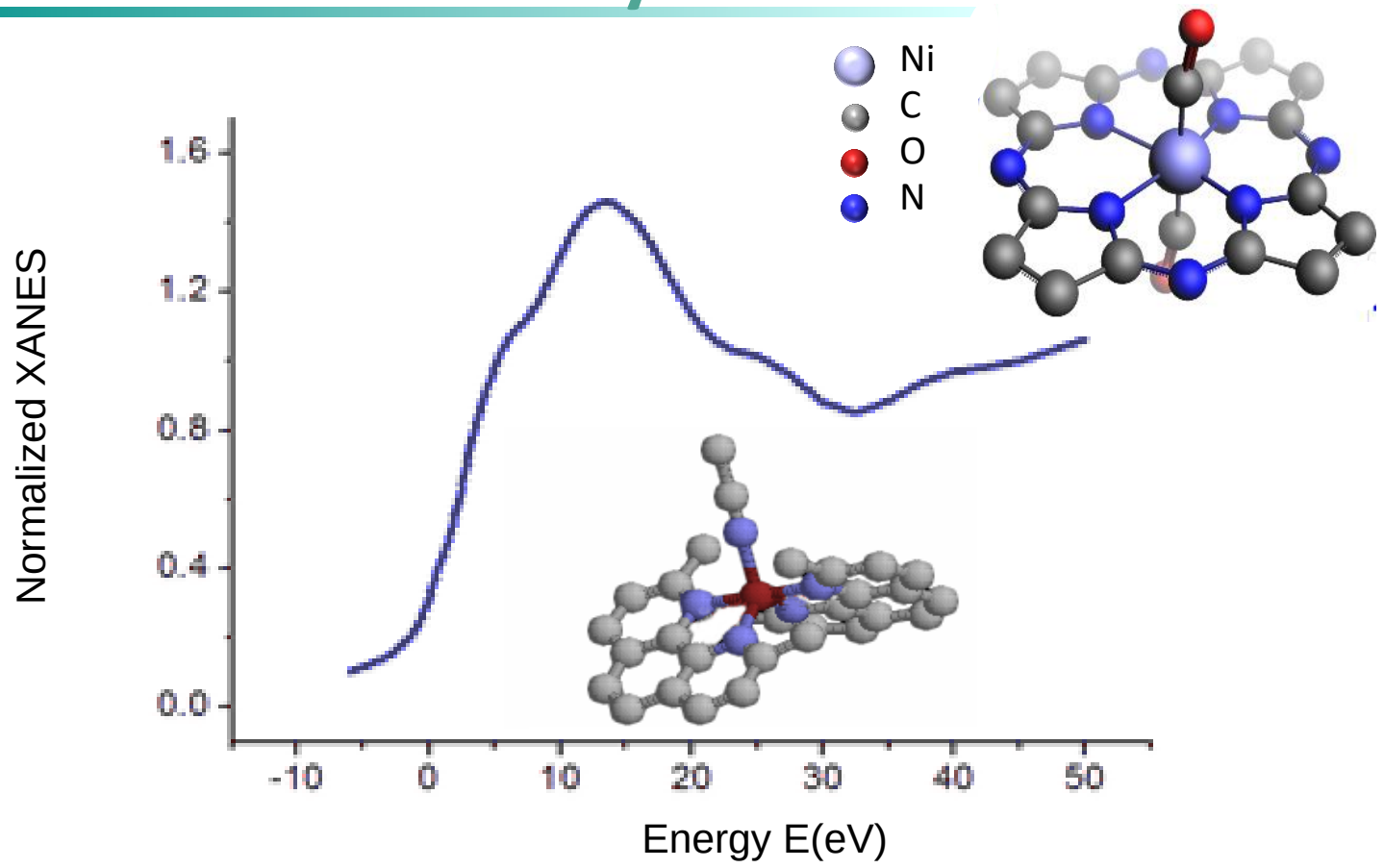
Luna, Timoshenko et al
ACS Catal. 11 (2021)



- **Multishell EXAFS analysis of heterogeneous materials by conventional methods is practically impossible**
- **Methods like RMC and MD are also typically not suitable for such materials**

4: Quantitative XANES analysis is diffucult





Timoshenko, Frenkel et al, J. Phys. Chem. Lett. 8 5091 (2017)

Martini, Timoshenko, et al, J. Am. Chem. Soc. 145 (2023)

- Local structure of nanomaterials - very different from bulk
- Interactions with support, ligands and adsorbates
- New species appear as intermediate states, with no stable analogues
- Different absorbing sites (e.g., core sites and surface sites in a particle) have distinct XANES spectra

Machine learning for XAS analysis

Janis Timoshenko, janis@fhi-berlin.mpg.de - Department of Interface Science, Fritz Haber Institute

Machine learning

Unsupervised ML:

- Clustering
  - (e.g., spatially resolved XAS)
- Dimensionality reduction
  - PCA
  - MCR-ALS

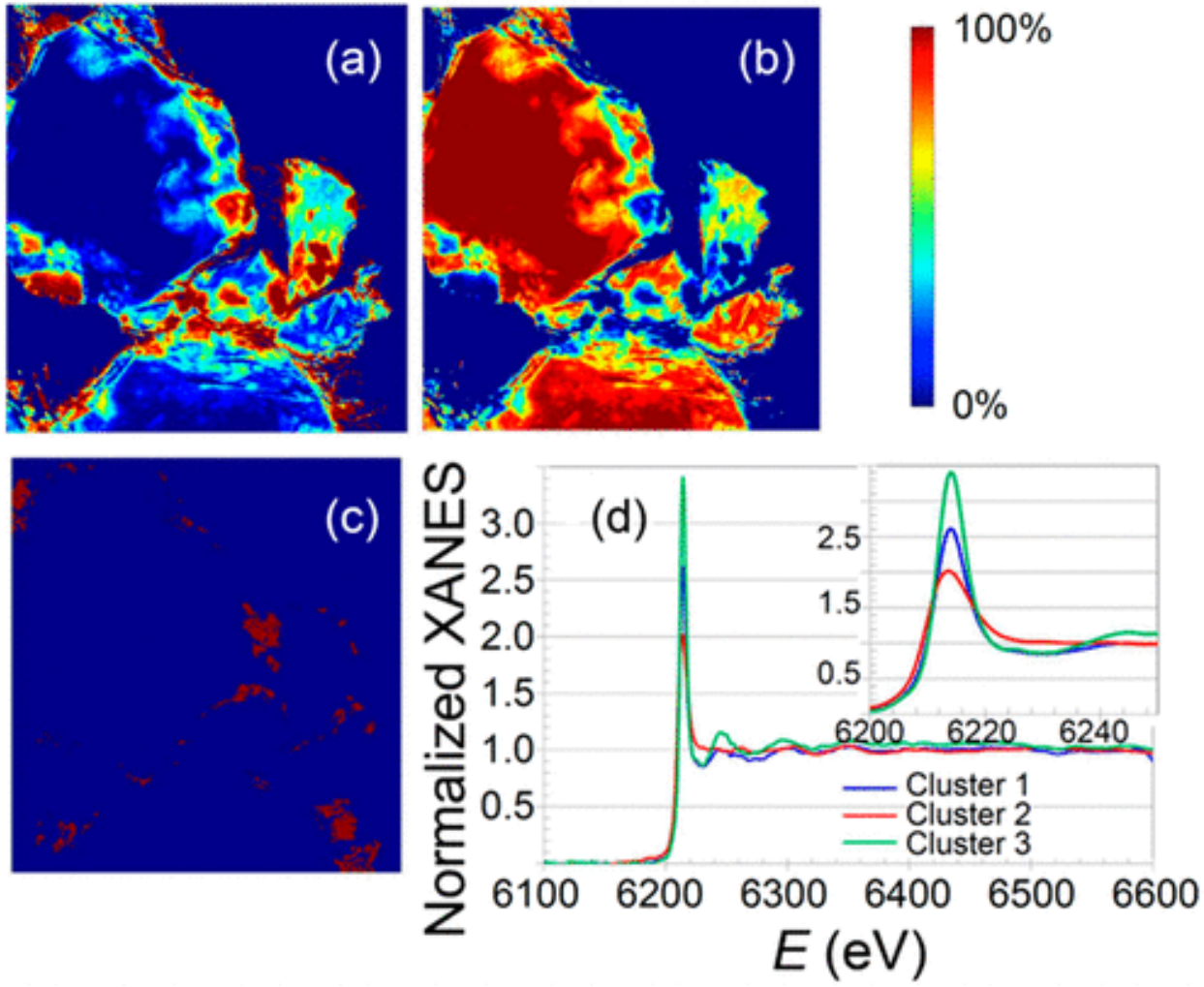
Supervised classification:

- Fingerprinting

Supervised regression:

- Quantitative structural analysis

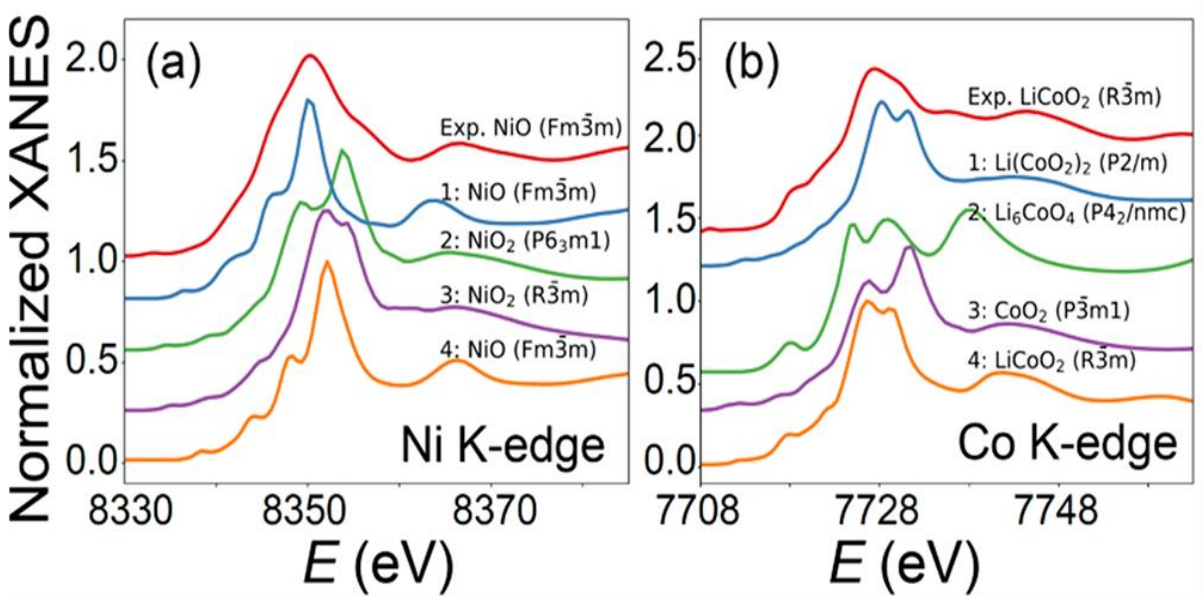
Clustering



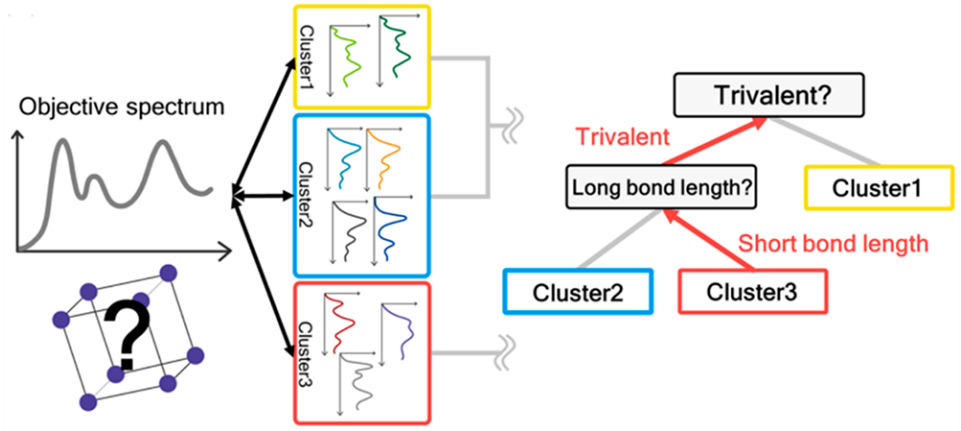
Nd L_3-edge EXAFS for permanent magnet material, $Nd_2Fe_{14}B$

Duan, Sci. Rep. 6 (2016)
Timoshenko & Frenkel, ACS Catal.9 (2018)

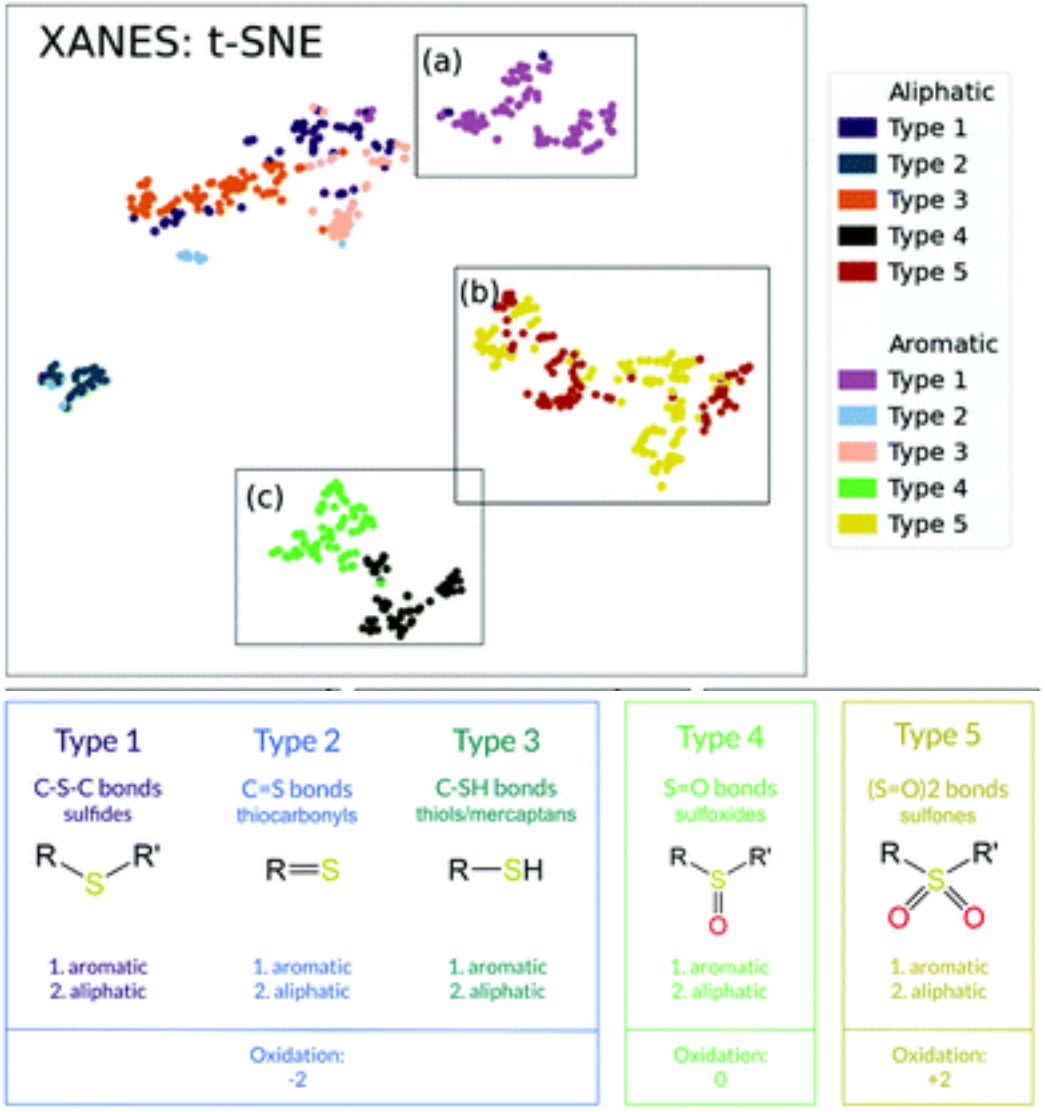
Classification and automatic spectra recognition



Zheng et al Npj Computational Materials 4 (2018)



Kiyohara et al Sci. Rep. 8 (2018)



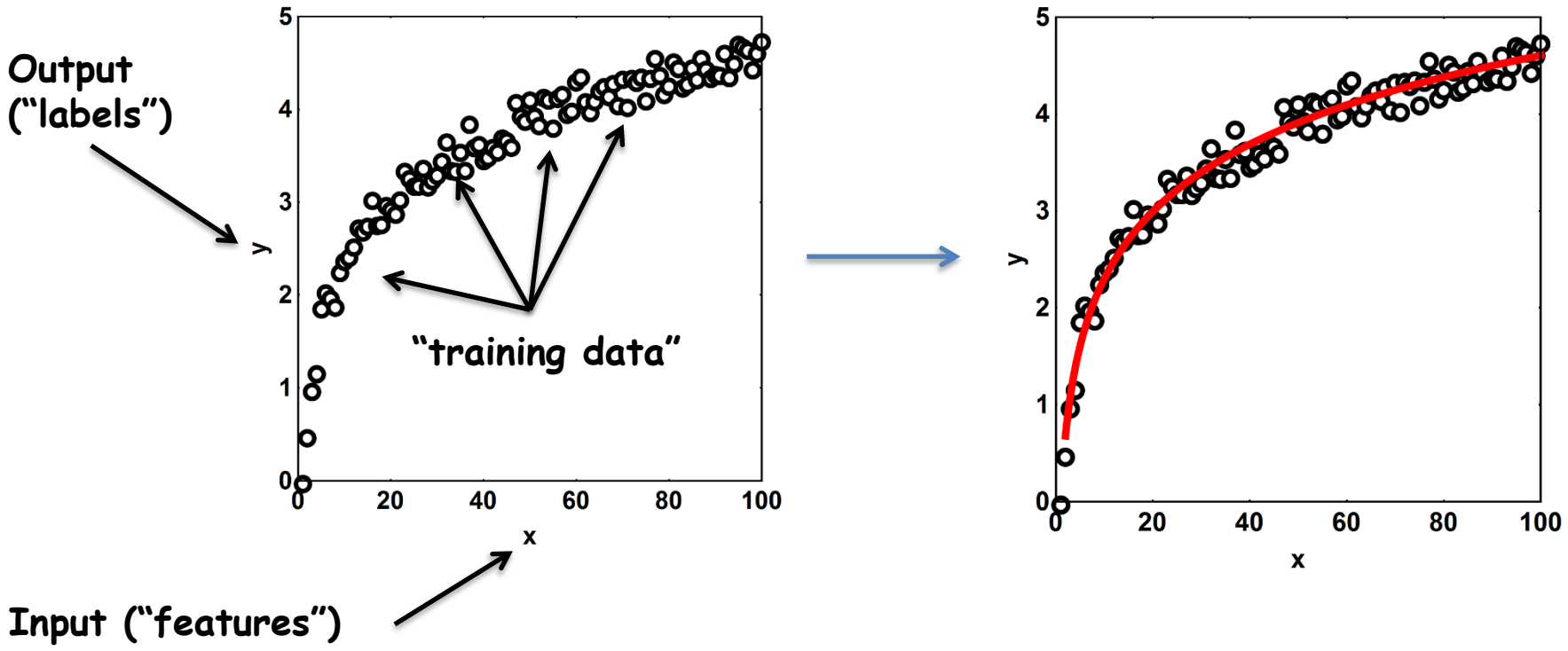
Tetef, Govind, Seidler, Phys. Chem. Chem. Phys. 23 (2021)

Supervised machine learning: regression

Regression:

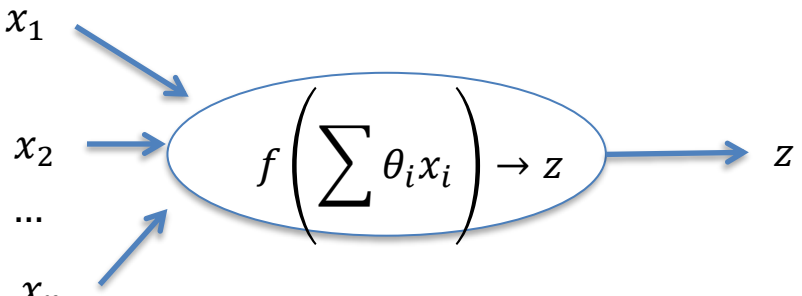
Find such function $y = h\left(x, \vec{\theta}\right) \rightarrow \mathbb{R}^n$ that for given *x* we can "predict" corresponding *y* value

Parameters



Neural networks

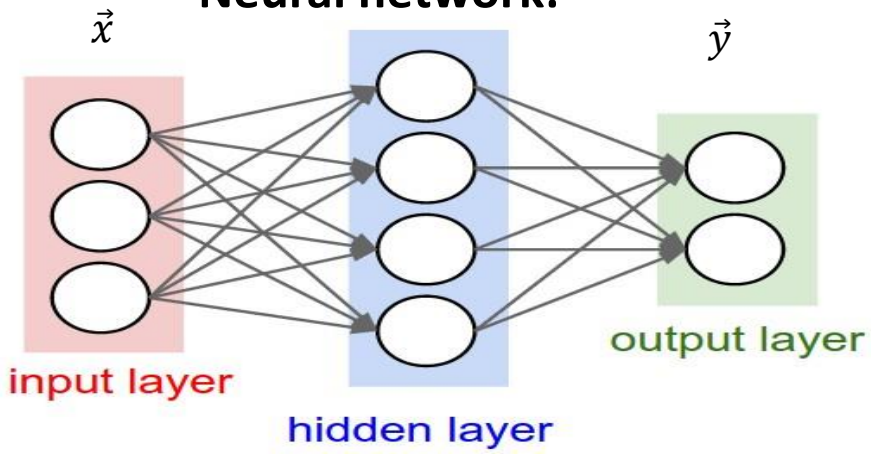
Neural network node:



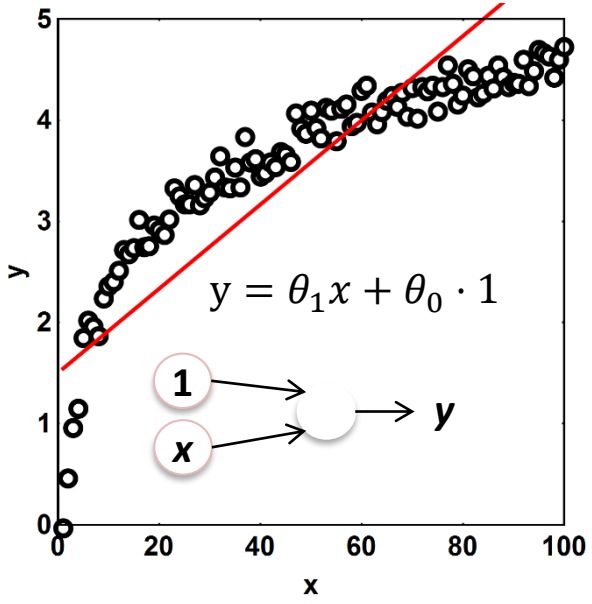
$f(a) = \tanh(a)$ **(for hidden layers)**

$f(a) = a$ **(for output layer)**

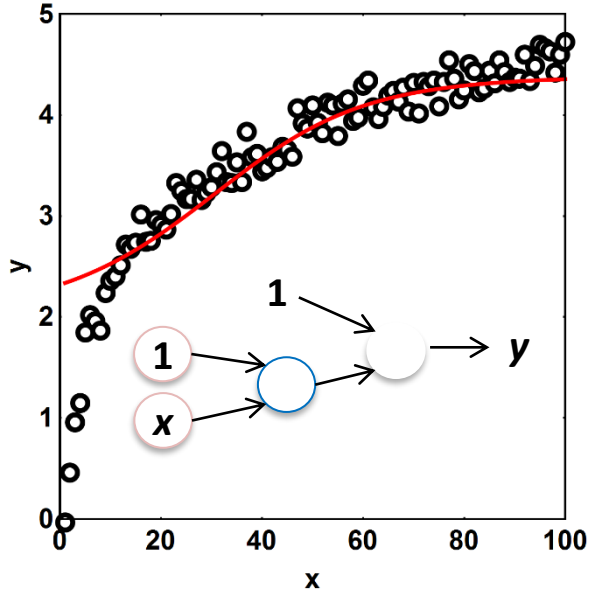
Neural network:



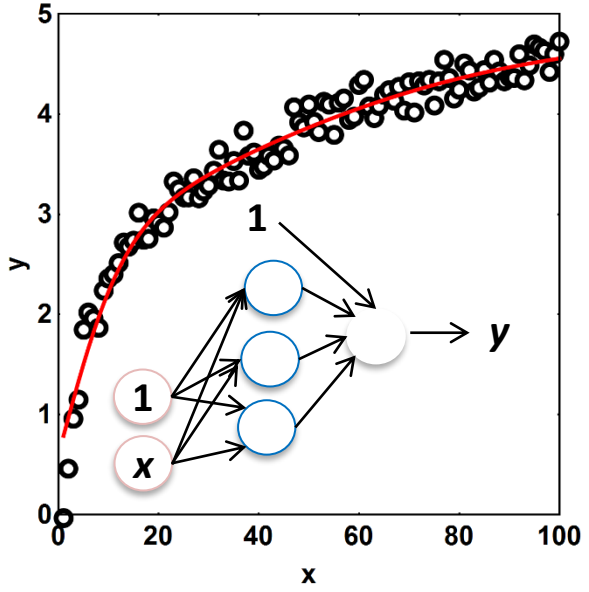
No hidden nodes:



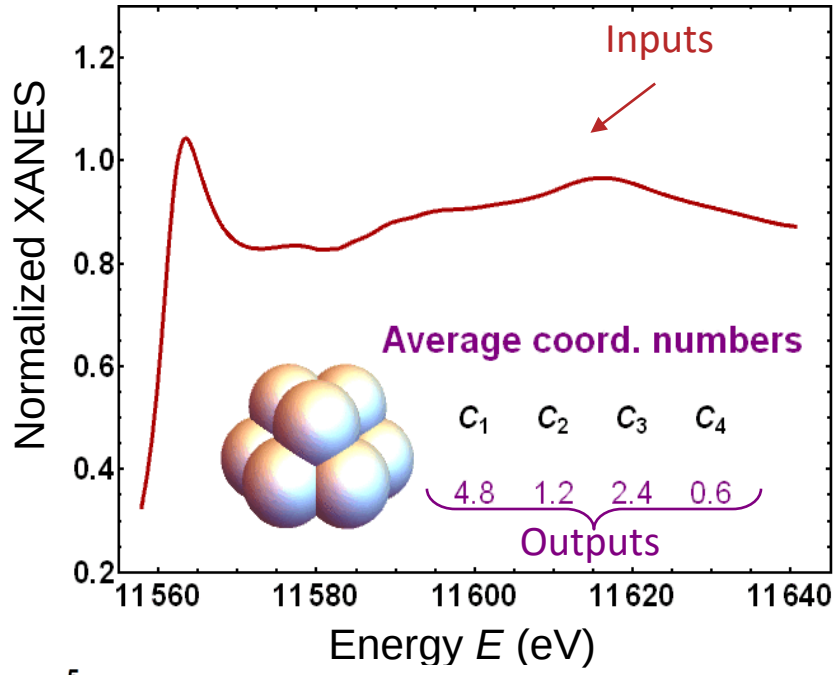
1 hidden node:

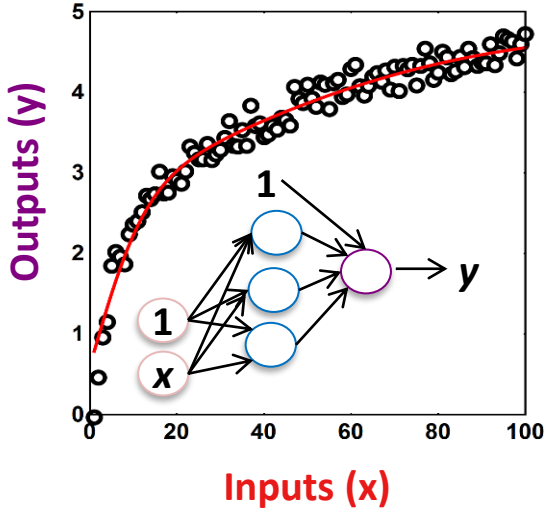


3 hidden nodes:



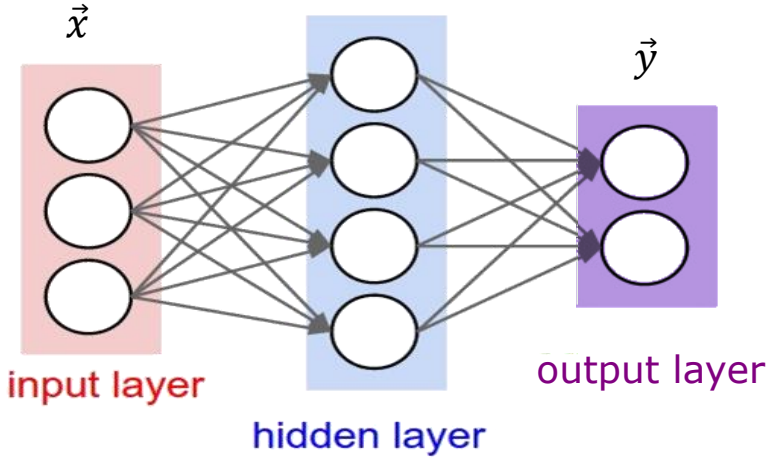
Machine learning for analysis of XAS data





Large set of theoretical spectra to **establish relation between XAS features and NP structure**

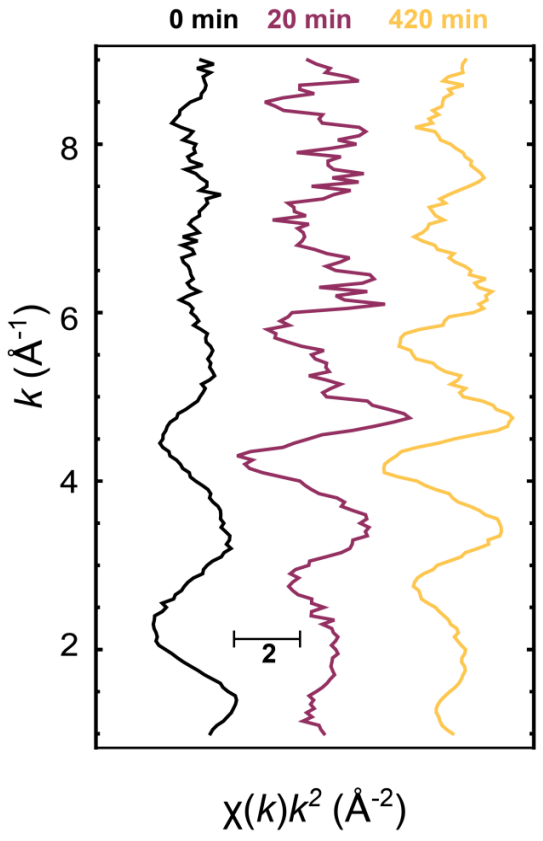
Artificial neural network



Timoshenko et al, J. Phys. Chem. Lett. 8 (2017)
Timoshenko et al., ACS Catal. 9 (2019)

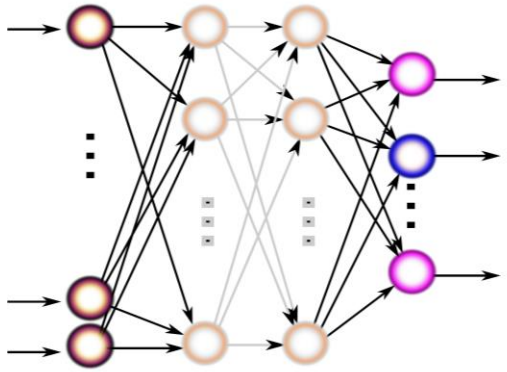
Machine learning for EXAFS analysis

In situ EXAFS



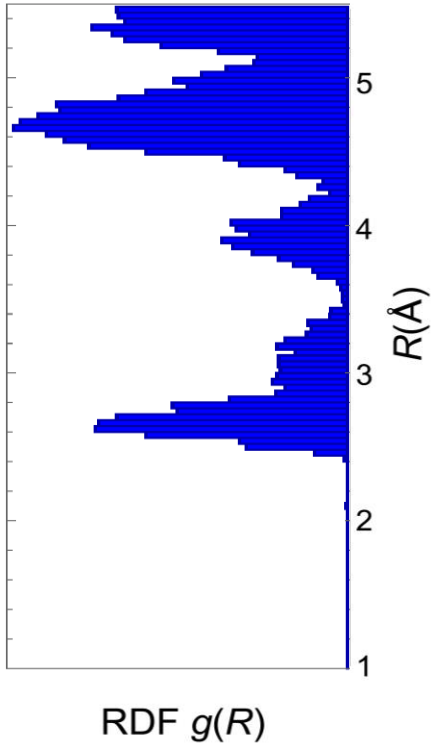
$Cu_{50}Zn_{50}$, Cu K-edge

Neural network

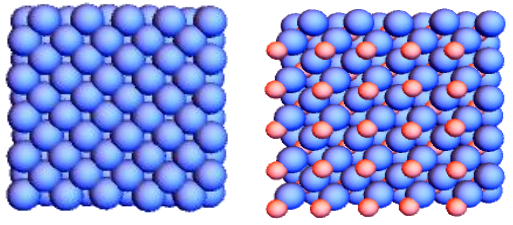


Simulated MD-EXAFS for training

Radial distribution functions



RDF $g(R)$



Cu **Cu_2O**

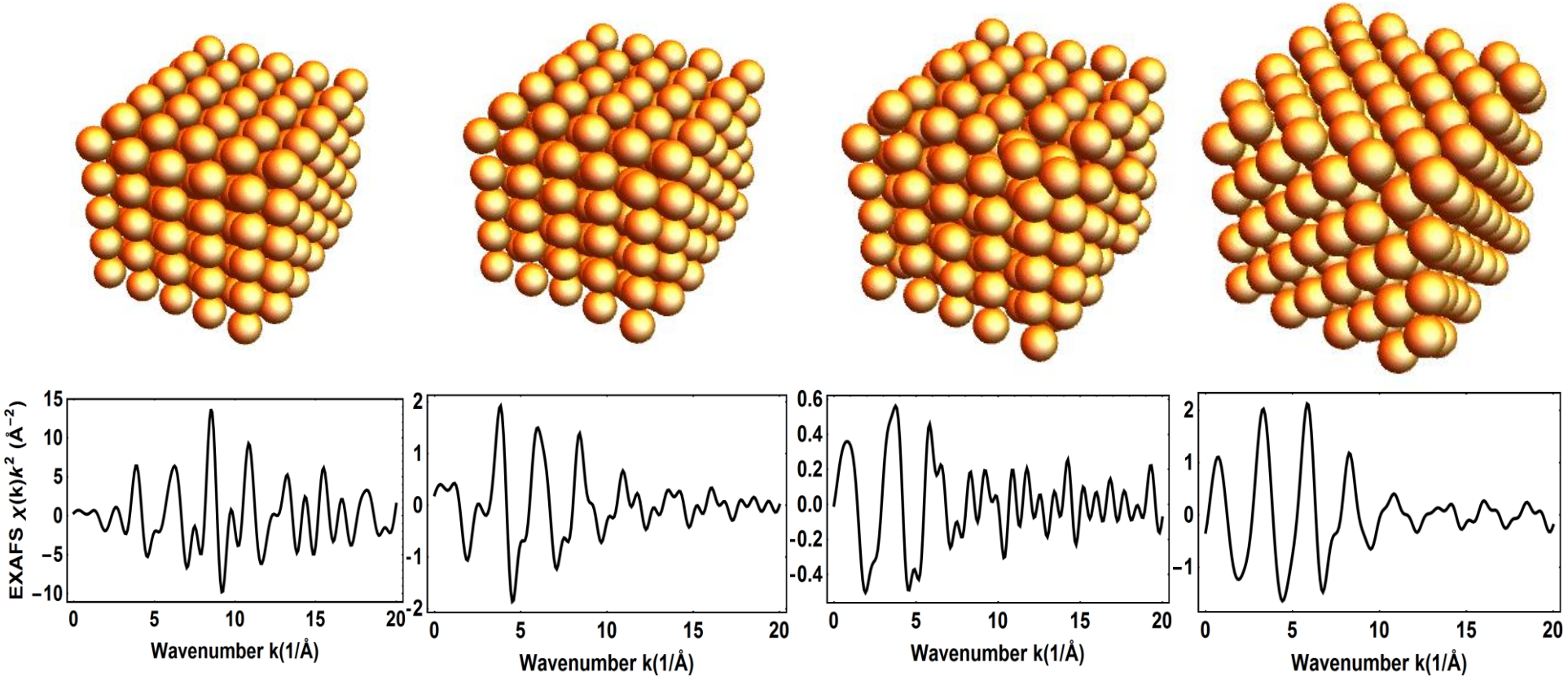
Timoshenko et al, Phys. Rev. Lett. 120 (2018) , Nano Lett. 19 (2019), ACS Catal. 9 (2019)

Timoshenko, Roldan et al, J. Phys. Chem. C 123 (2019)

MD-EXAFS for metals and oxides

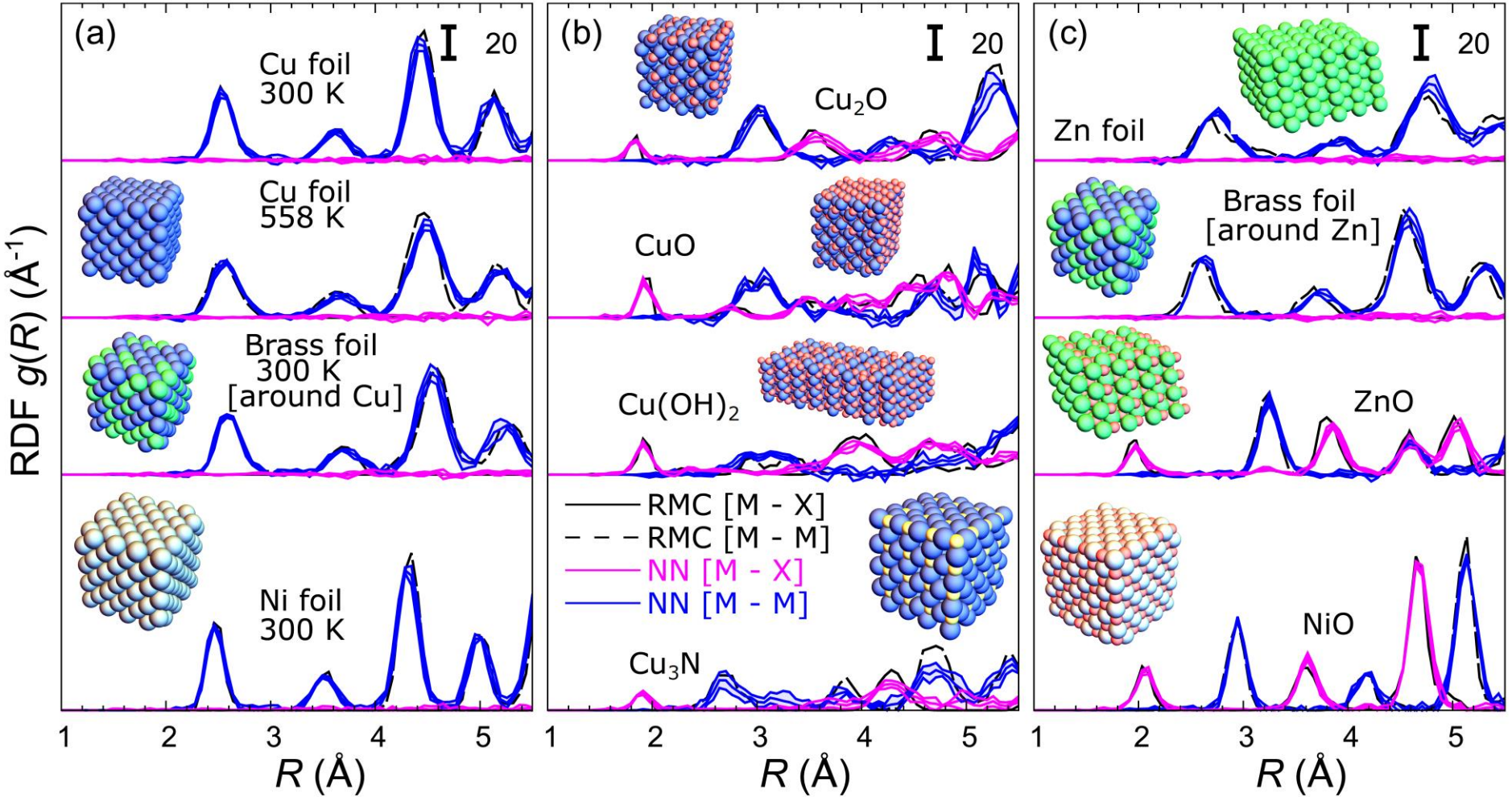
Challenge: disorder effects need to be included in the training data set

For this purpose we can rely on molecular dynamics simulations



Validation of NN accuracy with RMC method

Timoshenko, Jeon, Roldan Cuenya
Chem. Sci., 2020, 11, 3727-3736

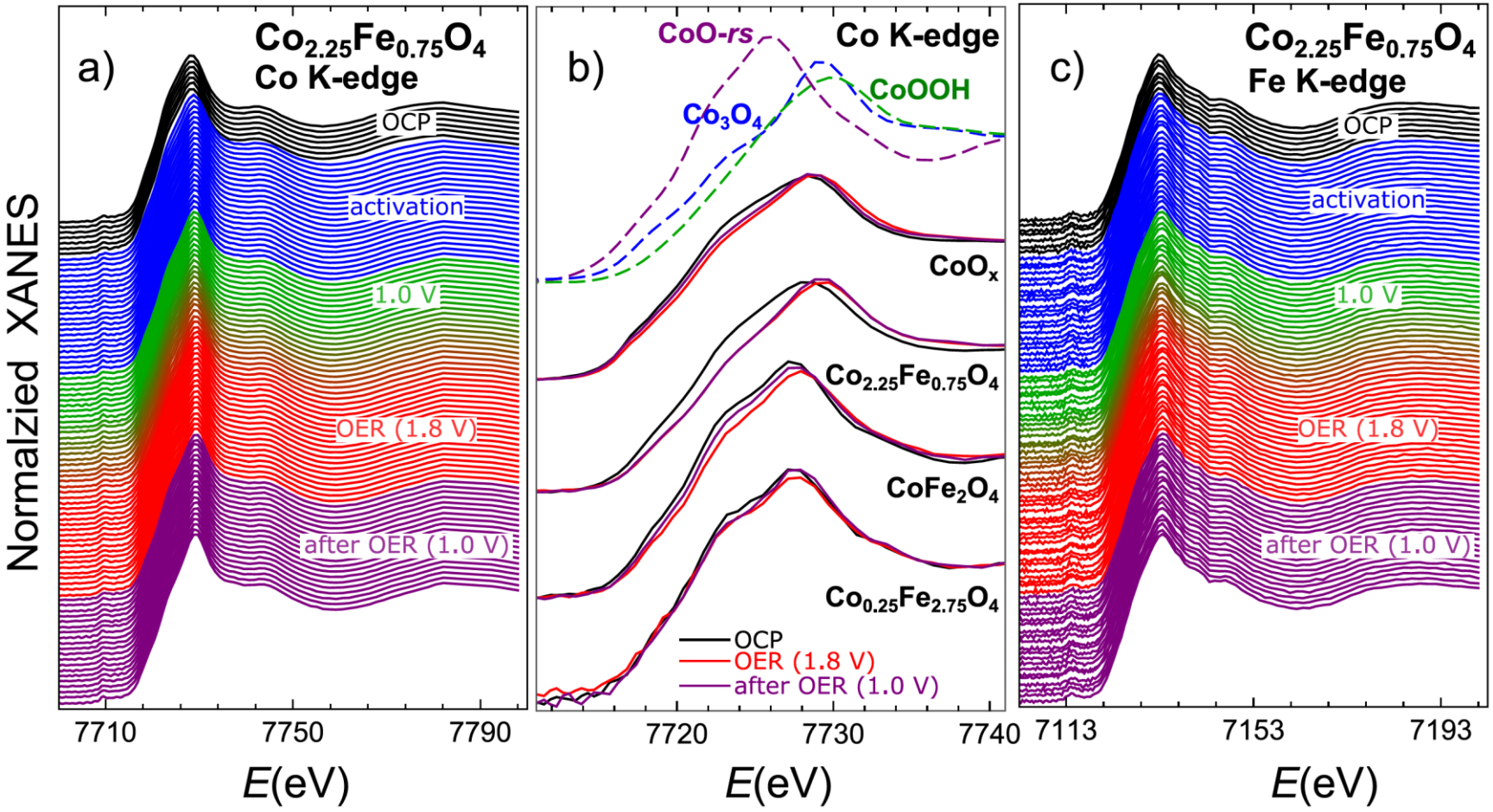


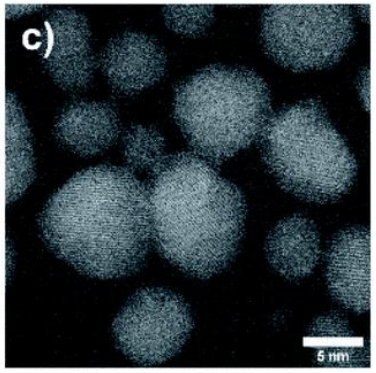
- ca. 300 MD models
- ca. 13 000 EXAFS spectra, further linearly combined to construct ca. 40000 training examples

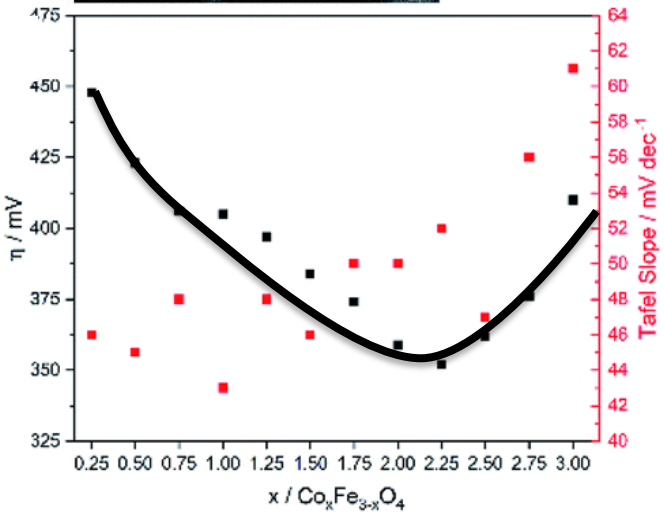
Example 1:
$Co_xFe_{3-x}O_4$ catalysts for Oxygen Evolution Reaction

- **Unsupervised ML: Principal Component Analysis**
- **Supervised regression: Neural Networks for EXAFS analysis**

QXAFS measurements for $Co_xFe_{3-x}O_4$ catalysts





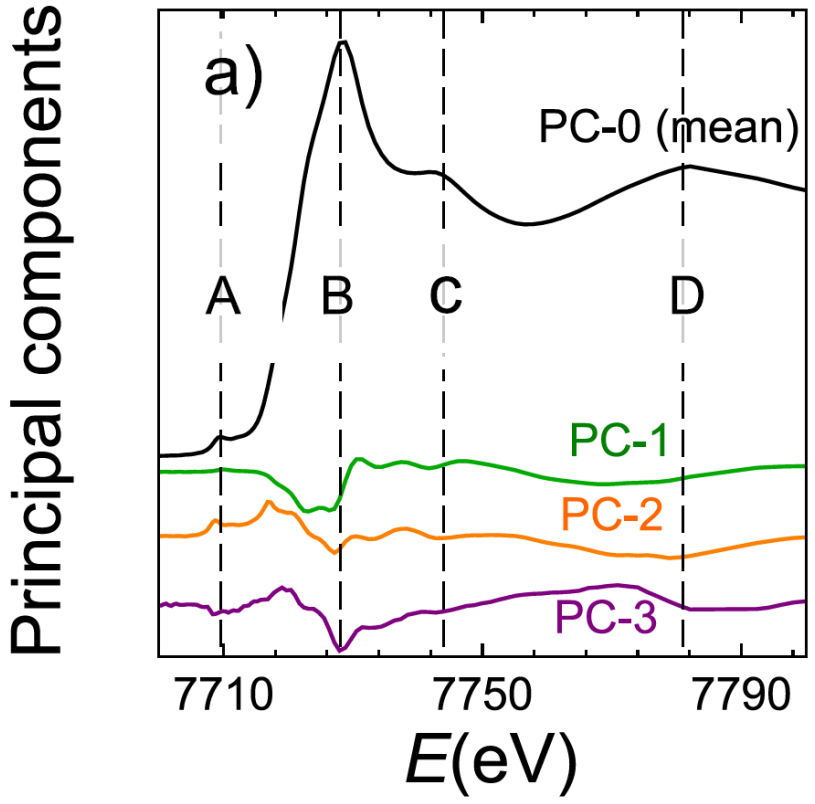


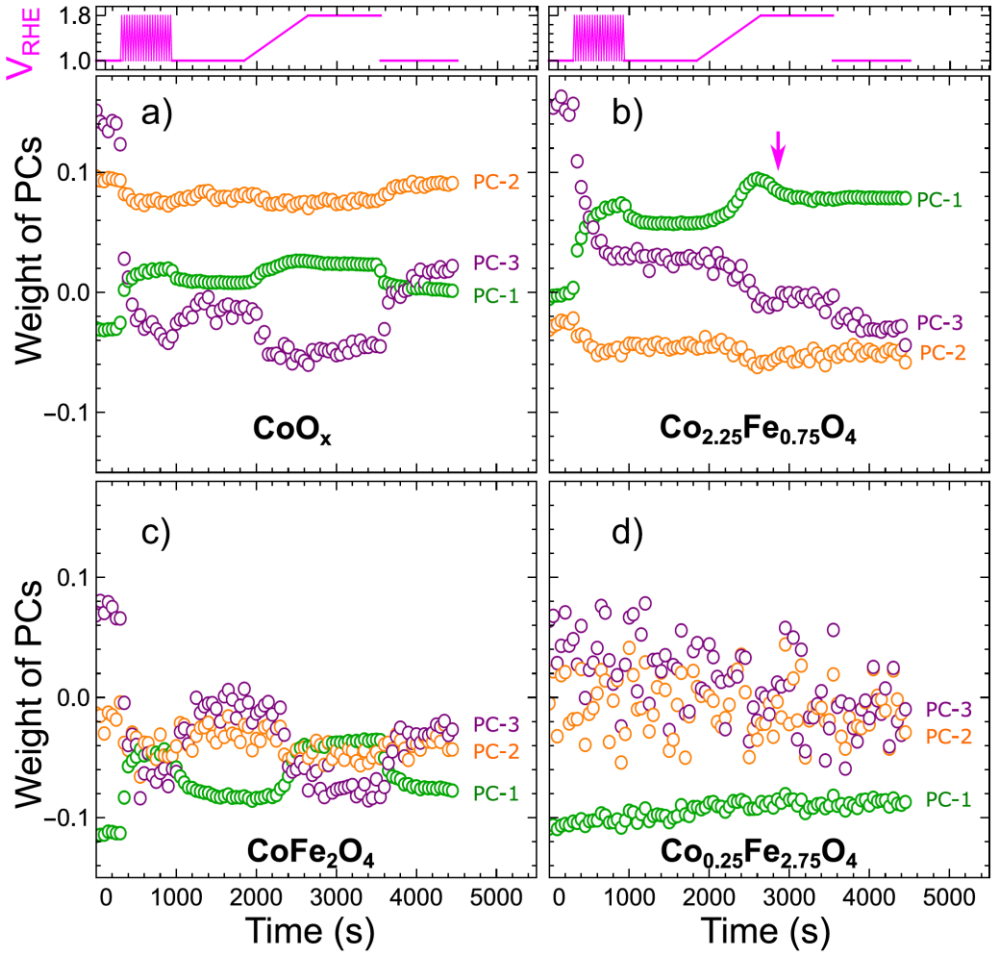
QXAFS data with 1s per spectrum!

- Spherical $Co_xFe_{3-x}O_4$ NPs, size decreases with increasing Co content from 9 nm to 3 nm
- OER activity increases with Co content up to *x* = 2.25, decreases for more Co-rich catalysts

Saddeler, Haase, Timoshenko, Roldan Cuenya, Shulz et al, J. Mater. Chem. 9 (2021) Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023

Interpretation of XANES: Principal Component Analysis (Co K-edge)



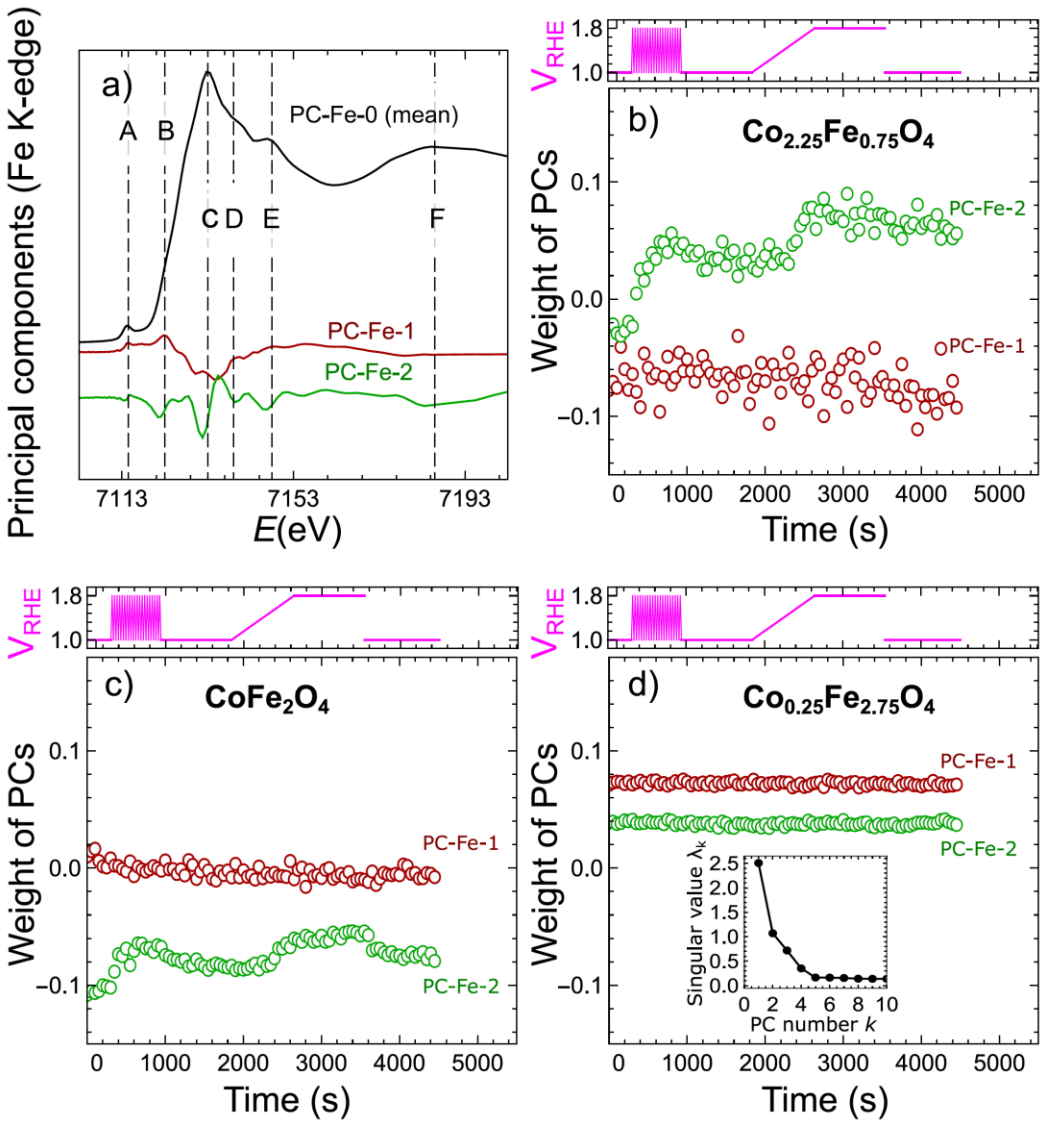


Data-driven analysis: only 4 spectroscopically distinct Co species:

- PC-1: Co oxidation from Co^{2+} to Co^{3+} state
- PC-2: tetrahedrally coordinated Co^{2+} species
- PC-3: octahedrally coordinated Co^{2+} species

Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023)

Interpretation of XANES: Principal Component Analysis (Fe K-edge)



Much less variation in Fe K-edge XAS

All XANES spectra resemble γ-Fe_2O_3

Only 3 spectroscopically distinct Fe species:

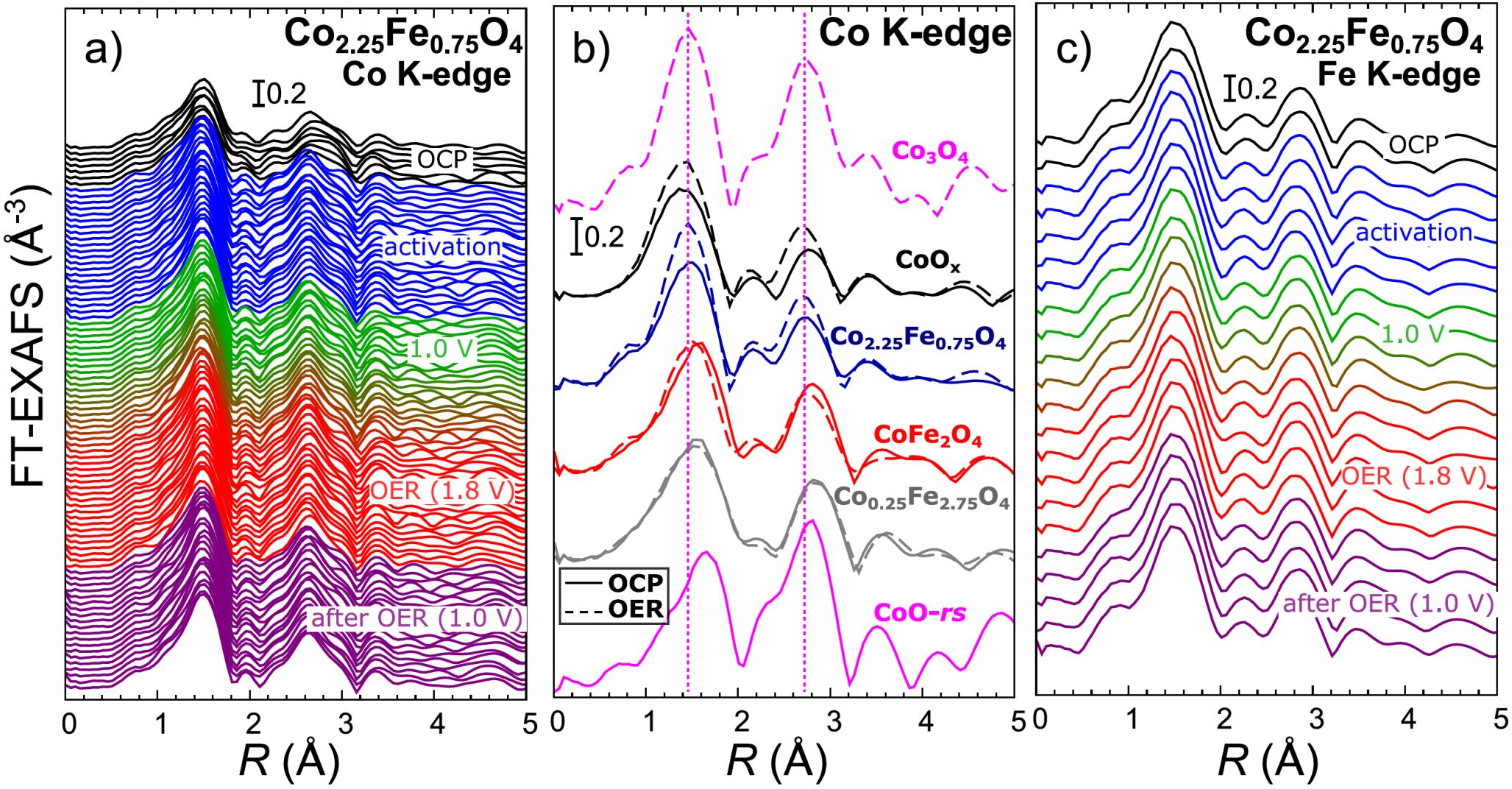
- Fe-PC-1 tracks change from octehadrally coordinated Fe to tetrahedrally coordinated
- Fe-PC-2 tracks further Fe oxidation

Subtle, but systematic increase in Fe oxidation for Co-rich samples: importance of Fe for OER!

Co-poor sample is OER inactive, and no change in Fe or Co oxidation state there

Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023)

EXAFS data

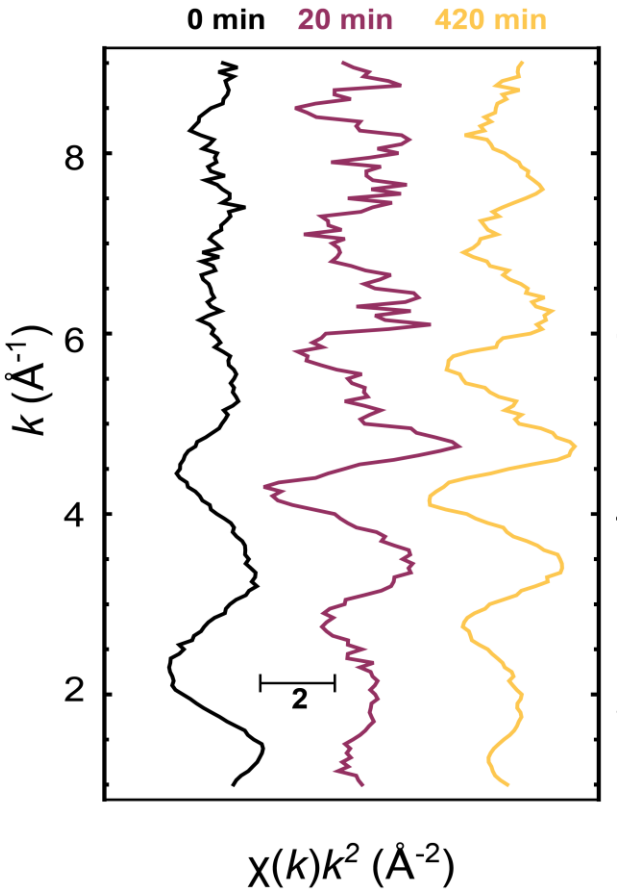


Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023

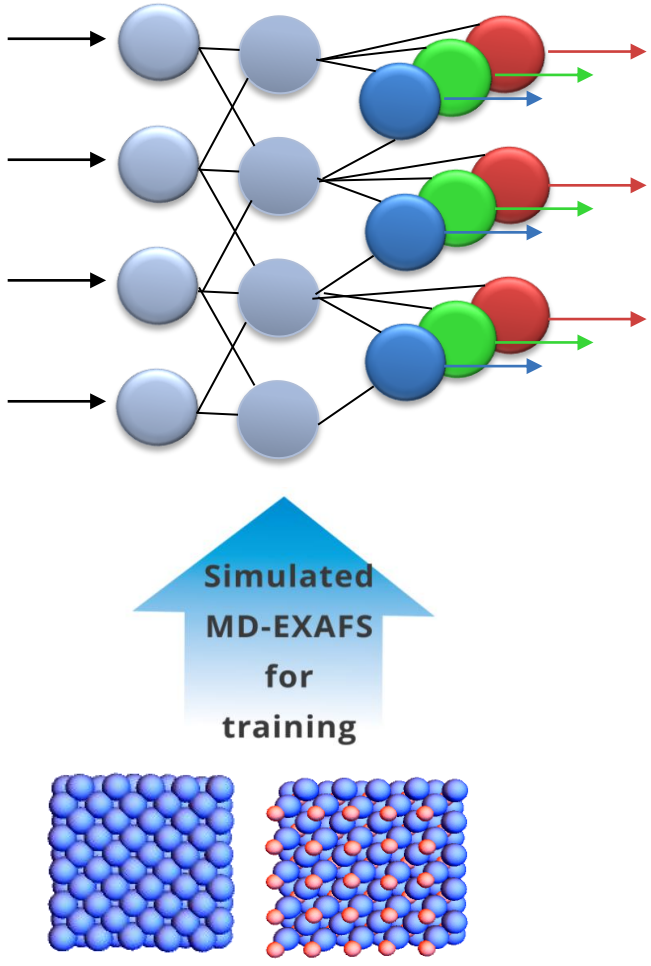
Interpretation of EXAFS by supervised machine learning

Artificial neural network approach allows «inversion» of EXAFS spectra and extraction of RDFs

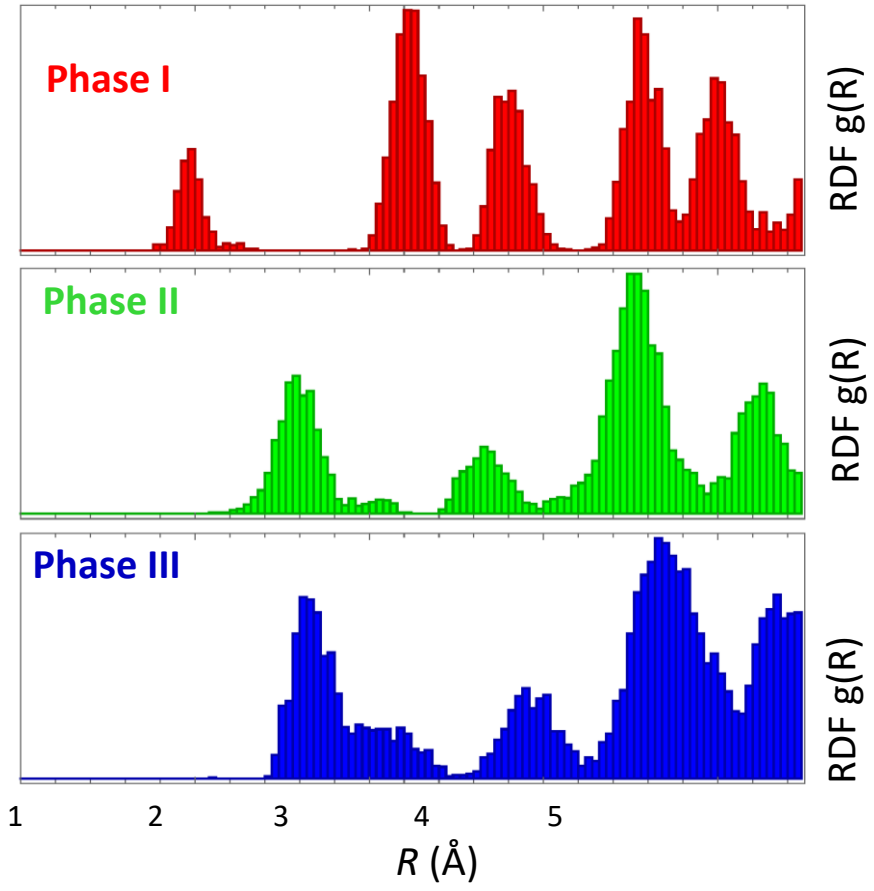
In situ EXAFS



Neural network

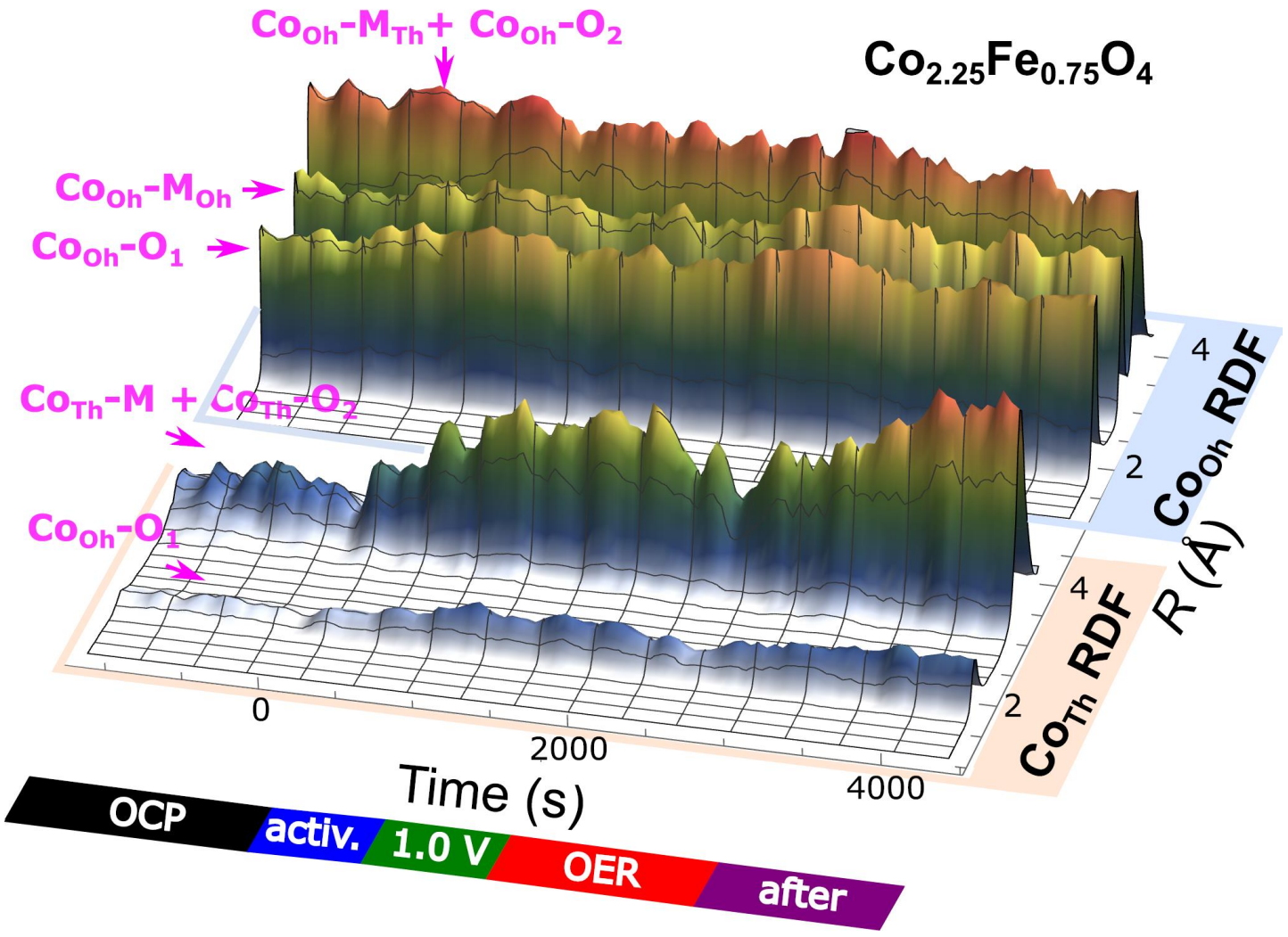


Partial radial distribution functions

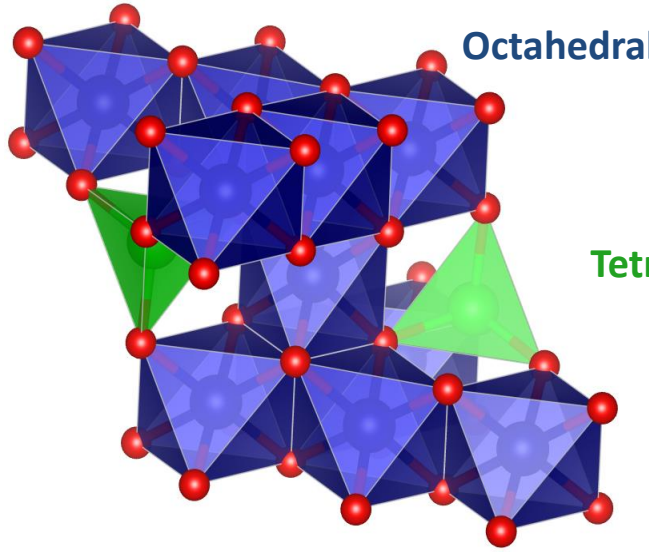


Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023)

Neural networks for analysis EXAFS data in heterogeneous mixtures

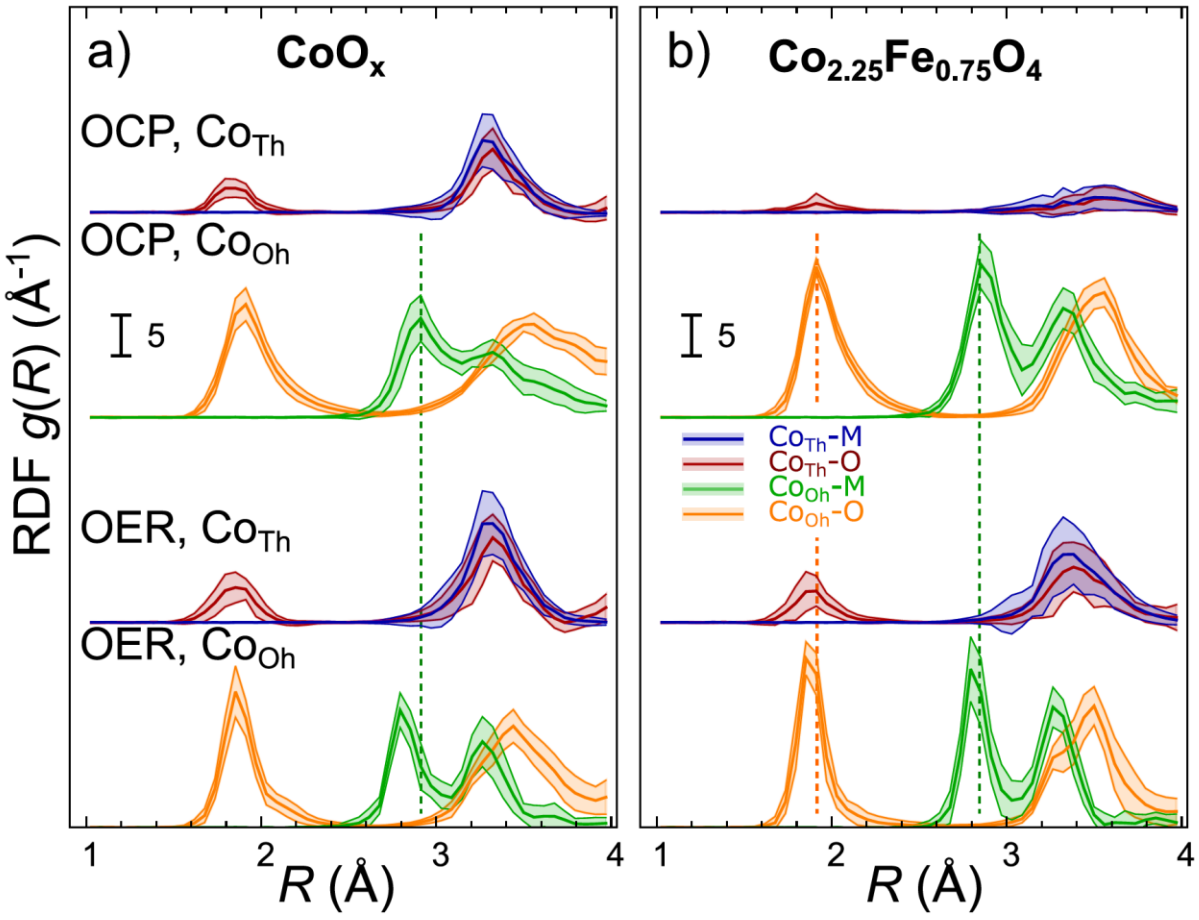


In oxides, neural network allows to separatelly quantify and track the evolution of tetrahedrally and octahedrally coordinated species!



Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023)

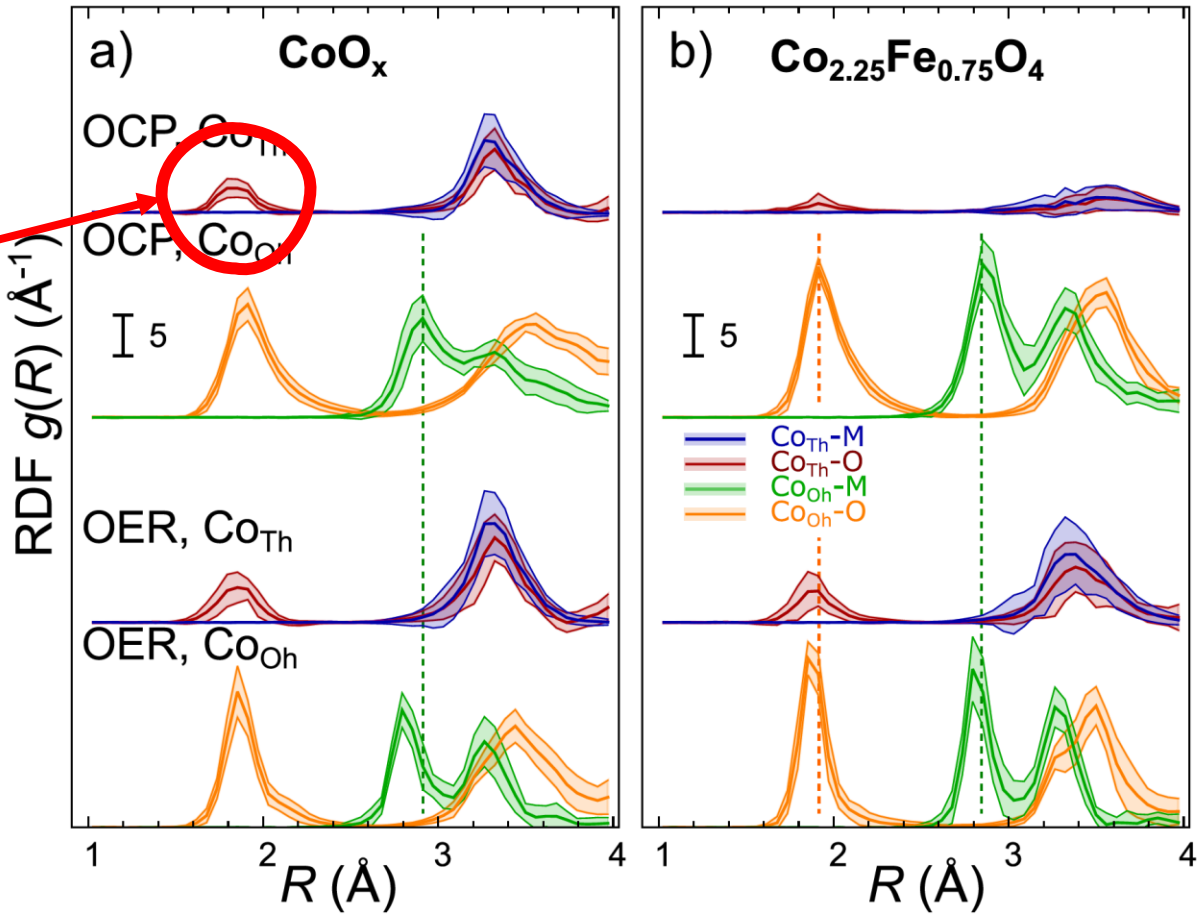
Evolution of local structure in $Co_xFe_{3-x}O_4$



Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 2023 (*in press*)

Evolution of local structure in $Co_xFe_{3-x}O_4$

Area under the 1st peak – concentrations of tetrahedrally and octahedrally coordinated species

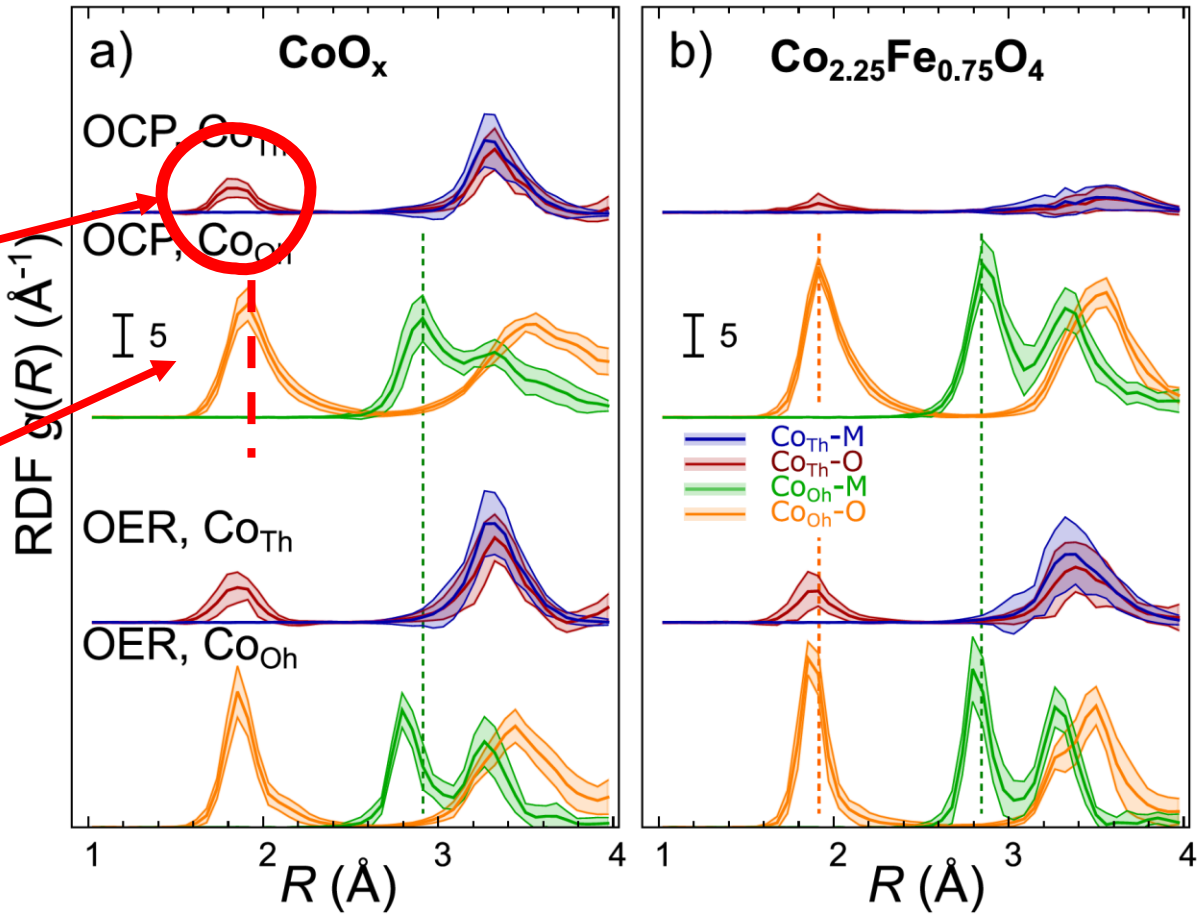


Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 2023 (*in press*)

Evolution of local structure in $Co_xFe_{3-x}O_4$

Area under the 1st peak – concentrations of tetrahedrally and octahedrally coordinated species

Positions of RDF peaks – oxidation state



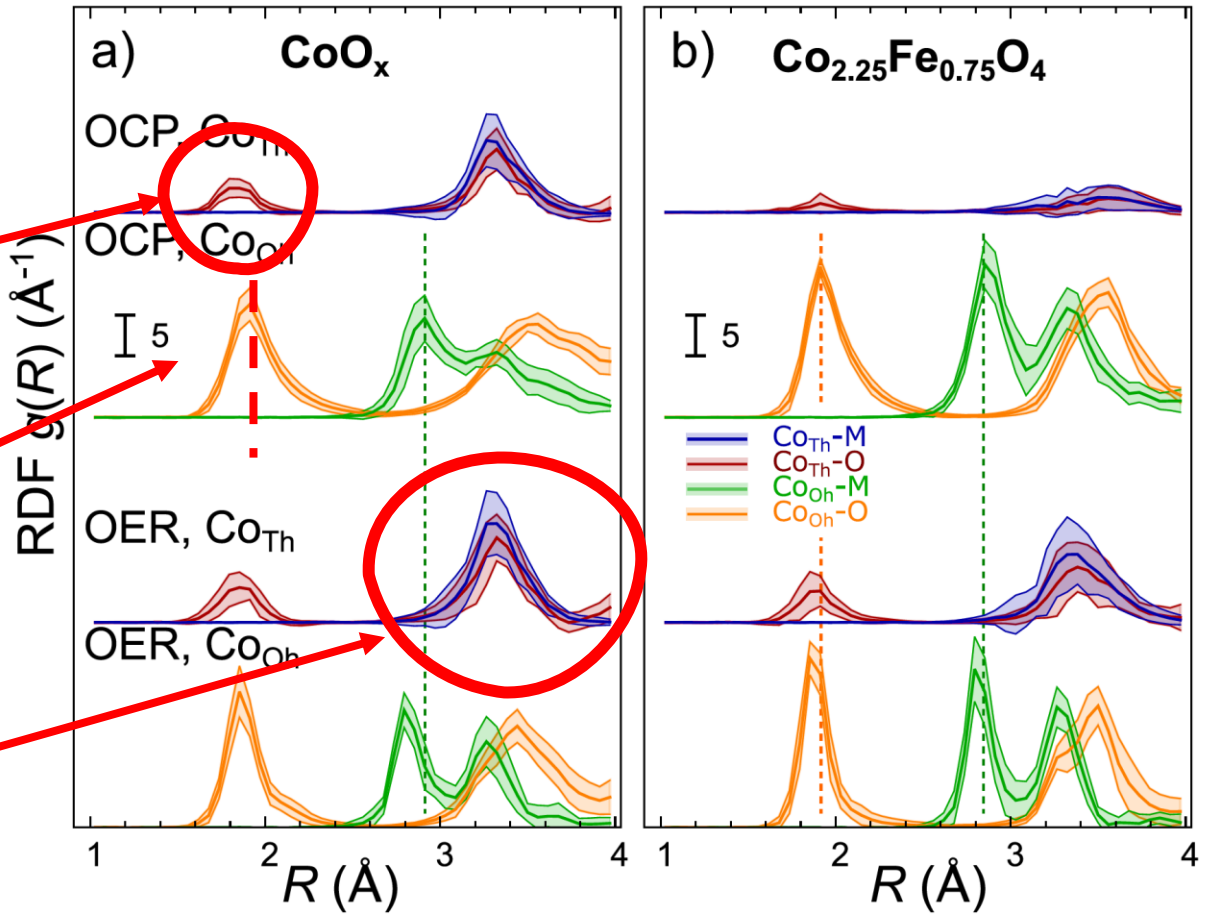
Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 2023 (*in press*)

Evolution of local structure in $Co_xFe_{3-x}O_4$

Area under the 1st peak – concentrations of tetrahedrally and octahedrally coordinated species

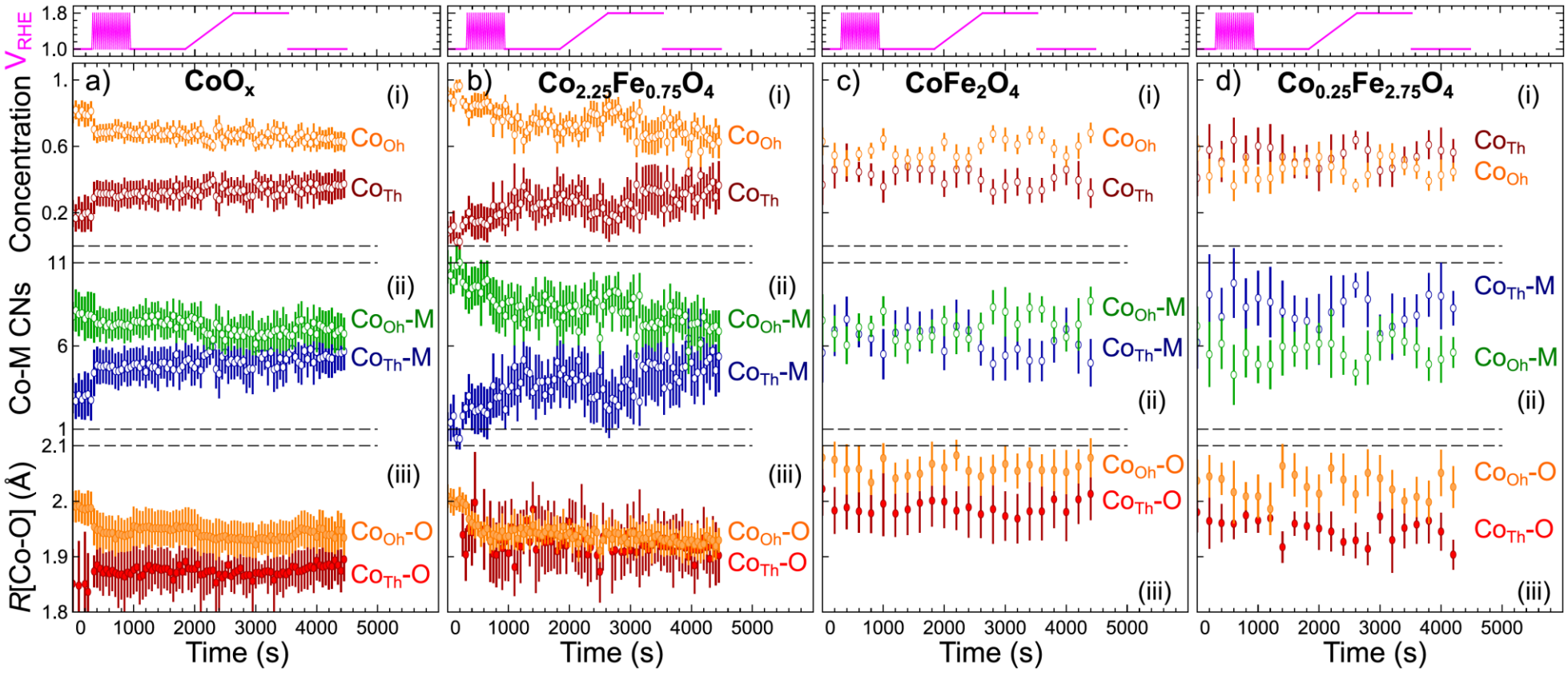
Positions of RDF peaks – oxidation state

Distant coordination shells– 3D structure, crystallographic structure



Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 2023 (*in press*)

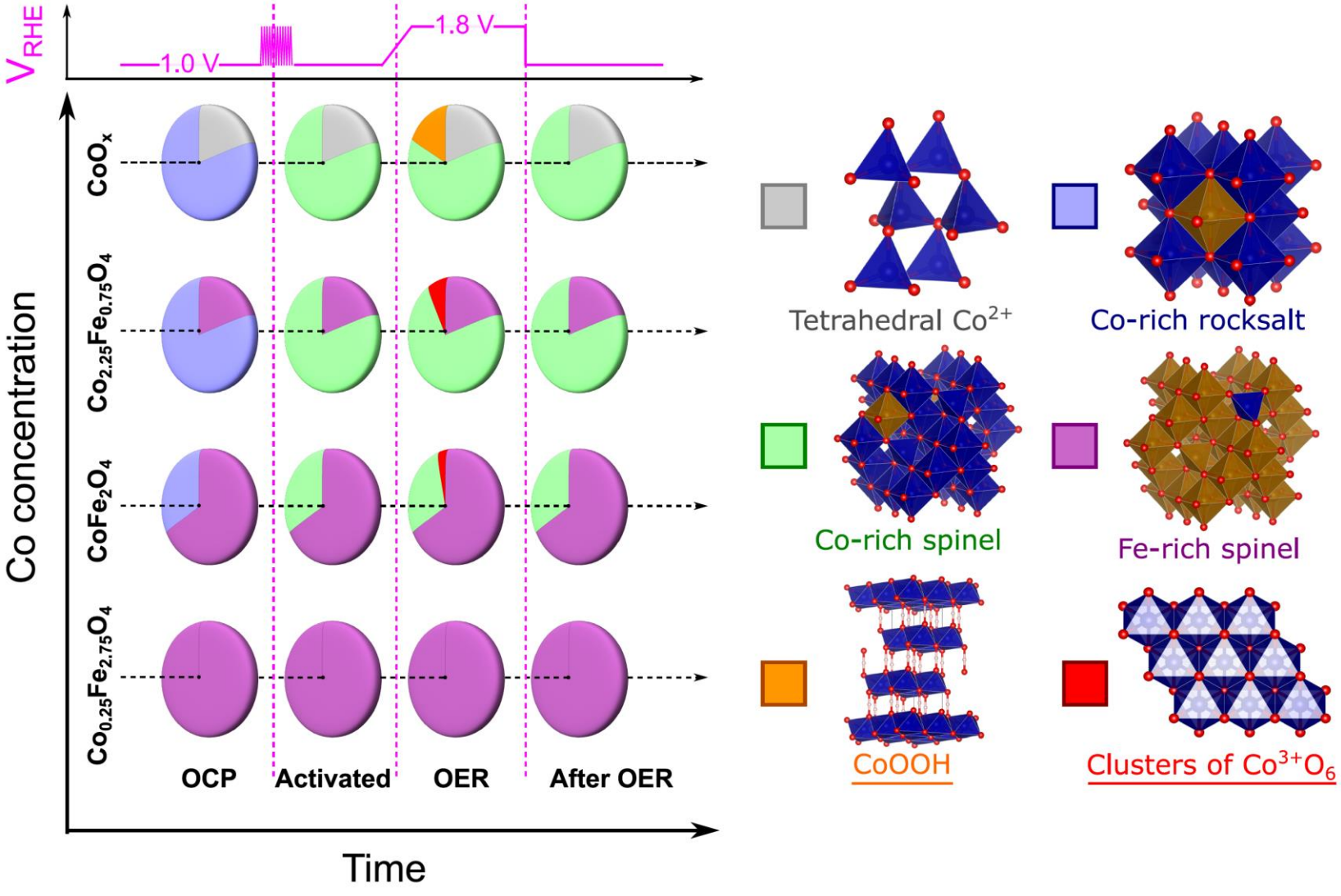
Evolution of local structure in $Co_xFe_{3-x}O_4$



- **Activation: irreversible transformation from rocksalt-like to spinel-like Co species**
- **OER: formation of octahedral Co^{3+} species**
- **Distant coordination shells: formation of layered CoOOH-like structures in pure CoO_x, and of more compact 3D arrangements of $Co^{3+}O_6$ octahedra in $Co_xFe_{3-x}O_4$**
- **Fe K-edge EXAFS: incorporation of Fe into rocksalt-like Co-rich structures**

Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023)

Summary: evolution of local structure and chemical state in $Co_xFe_{3-x}O_4$



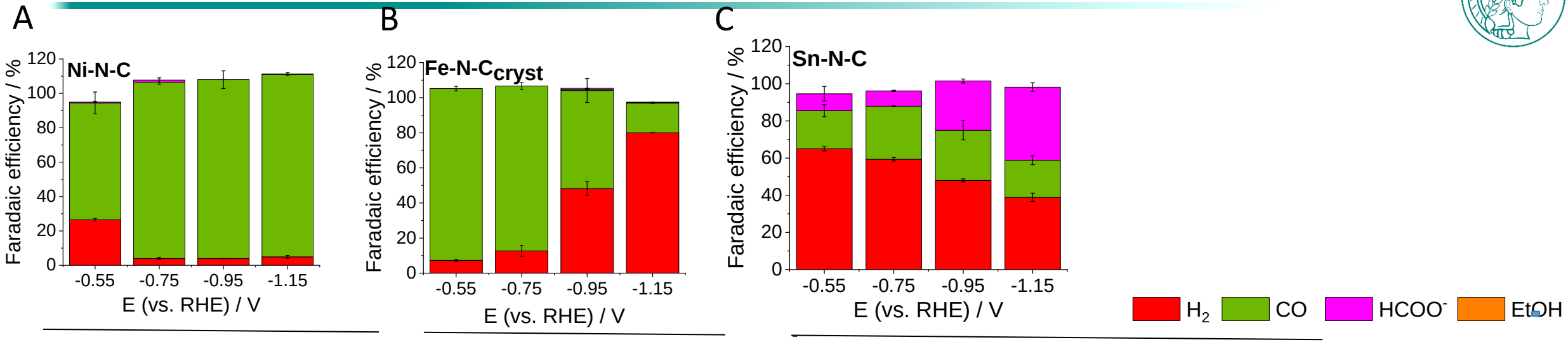
- **Strongly heterogeneous, segregated structure that evolves under reaction conditions**
- **Evolution of all structures under activation and OER**
- **Interplay between Fe and Co species in Co-rich bimetallic structures (unique active species)**

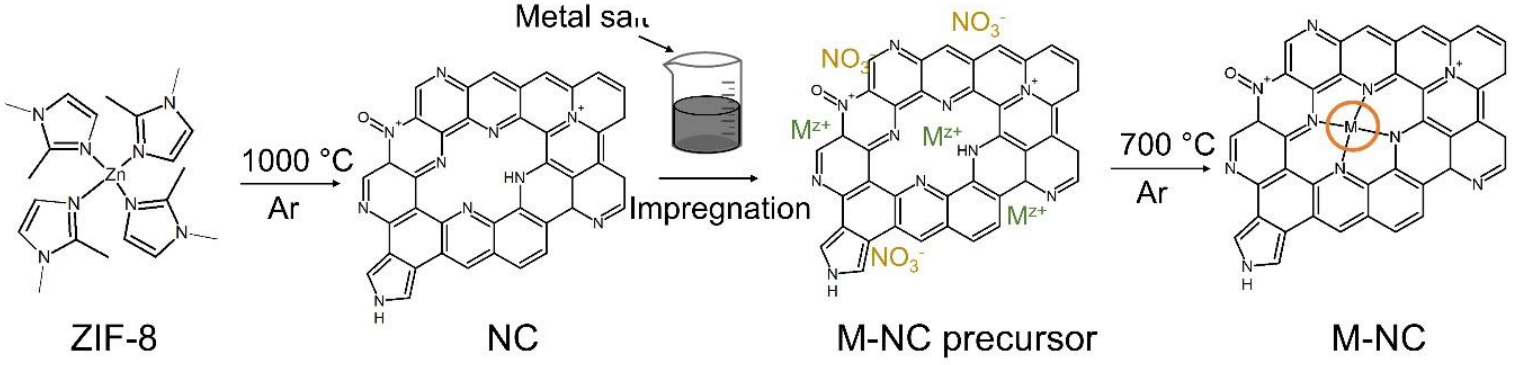
Timoshenko, Haase, Roldan Cuenya, J. Am. Chem. Soc. 145 (2023)

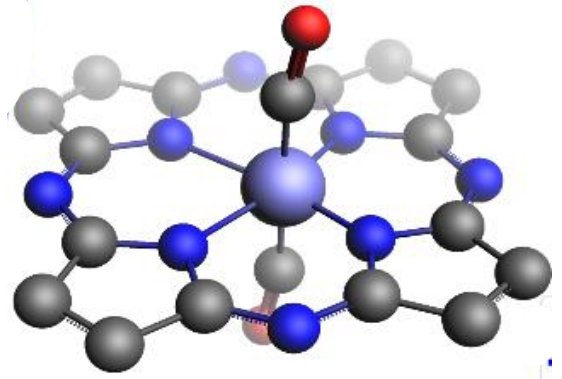
Example 2:
Ni single atoms for electrocatalytic CO_2 reduction

- **Unsupervised ML: multivariate curve resolution**
- **Inverse supervised ML for XANES analysis**

Transition metal nitrogen co-doped carbon catalysts



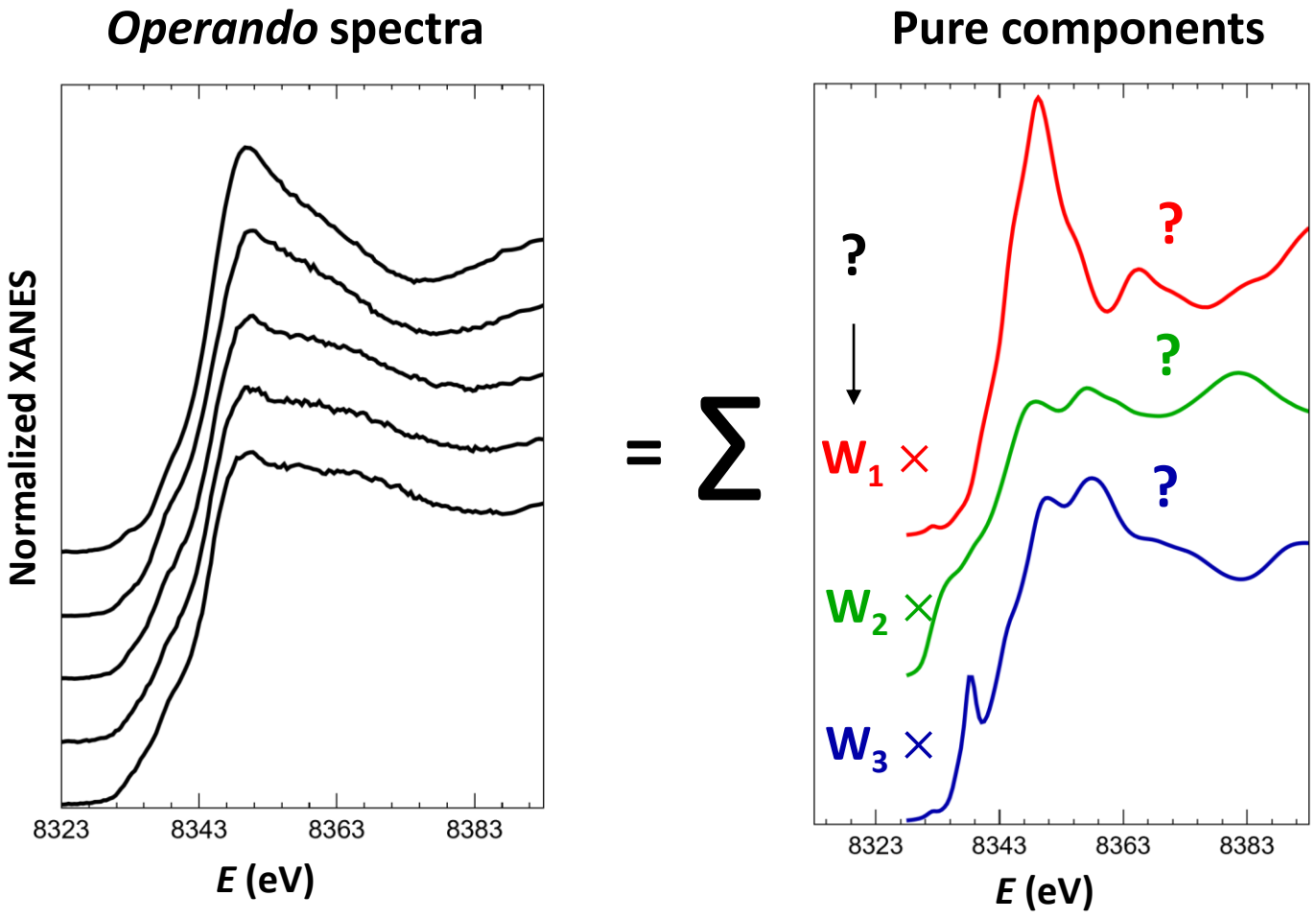




M-N-C catalysts for CO_2RR: promising catalytic properties, but the active species/active states are not clear

- Active species transiently present under working conditions (*operando* methods needed)
- Activity associated with disordered, ultradispersed phases (no good XRD)
- Many species coexisting, co-contributing to spectroscopic data

Unsupervised machine learning: principal components



Collected Ni K-edge XANES spectra – mixtures of contributions of different species.

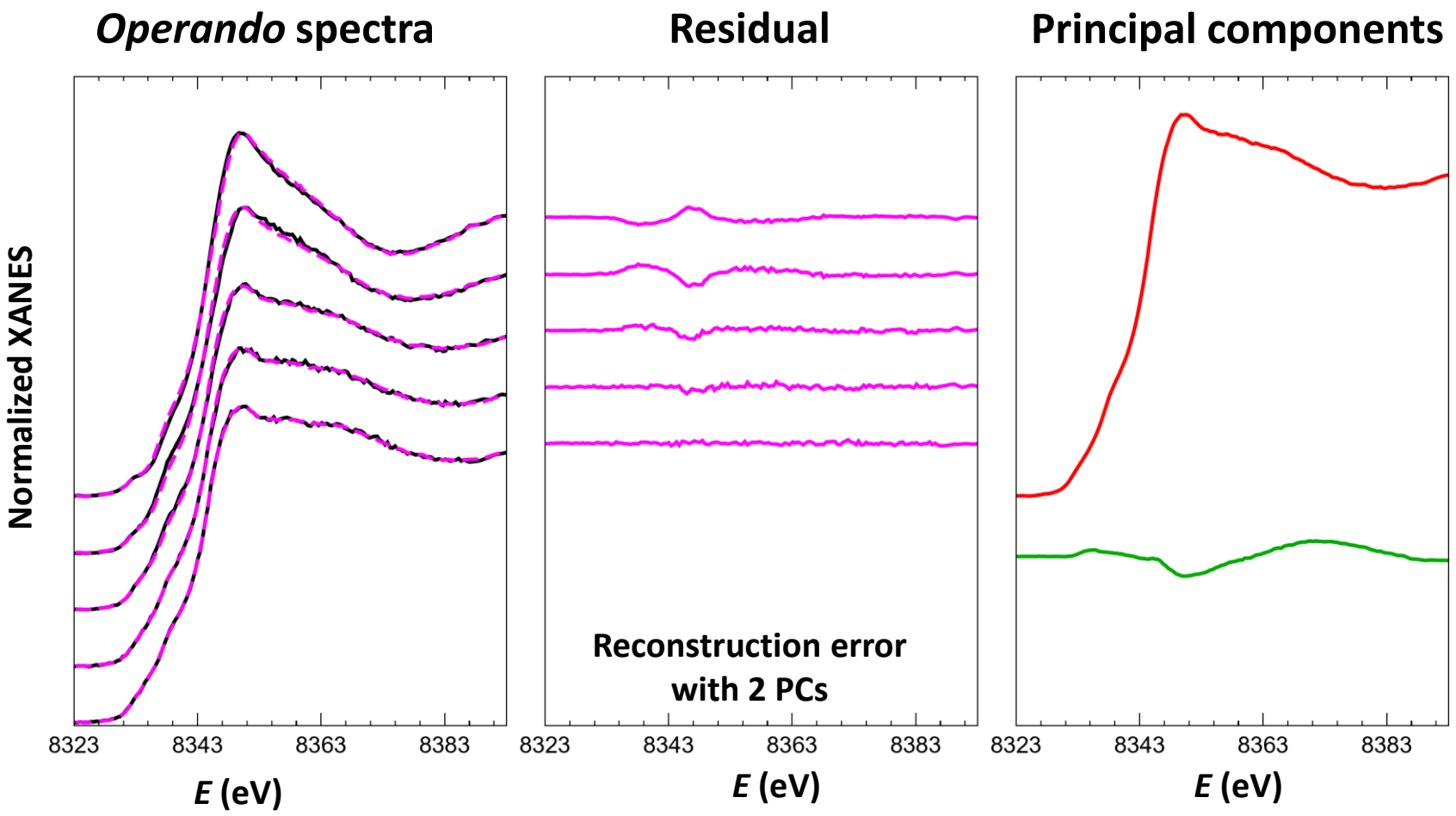
Spectral decomposition problem:

- What are the spectra for pure species?
- What are these species?
- How their concentrations change with time?

Not an unambiguos decomposition: additional constraints needed!

Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

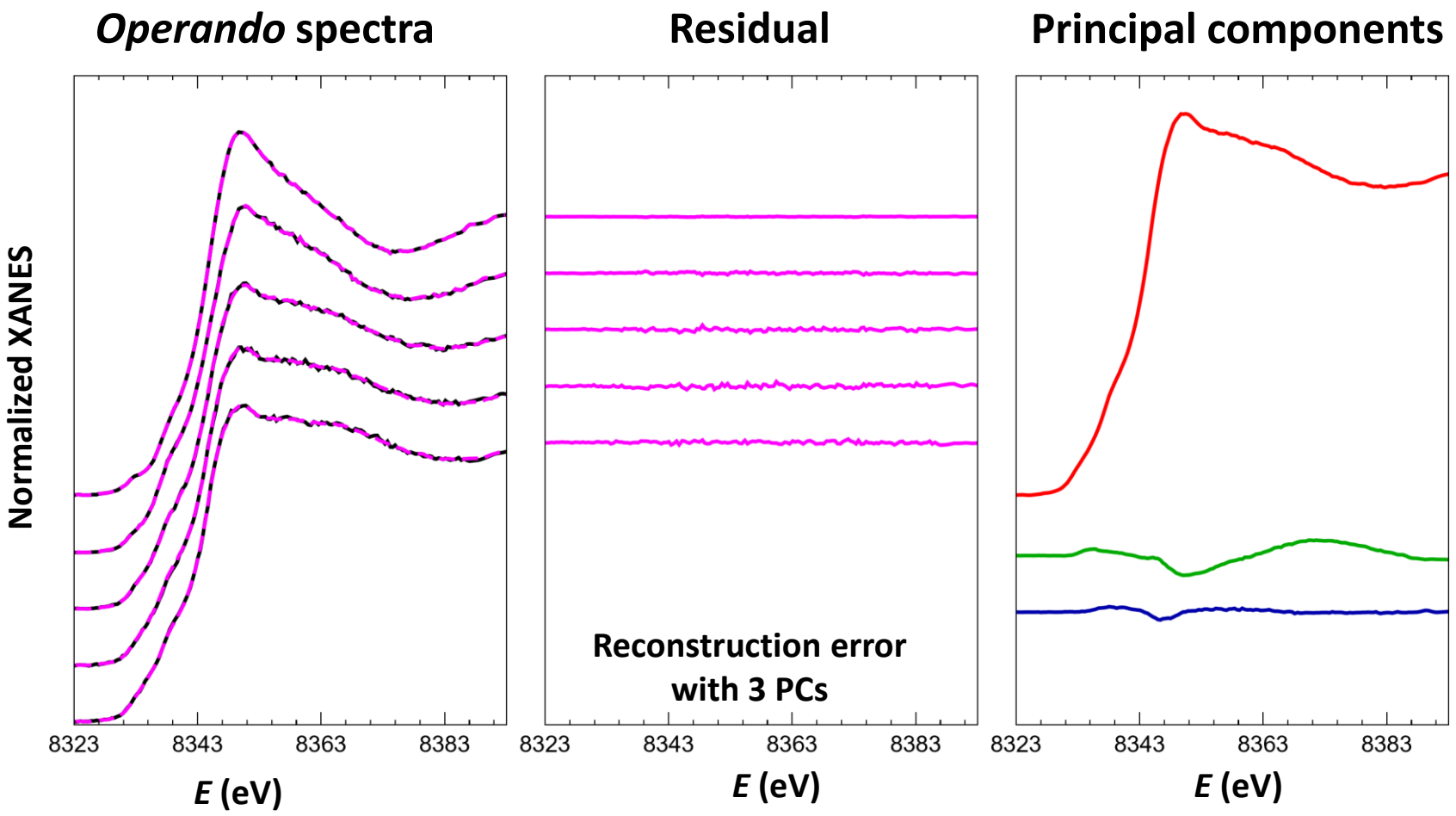
Unsupervised machine learning: principal components



PCA: constraint – spectral components are linearly independent

Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

Unsupervised machine learning: principal components

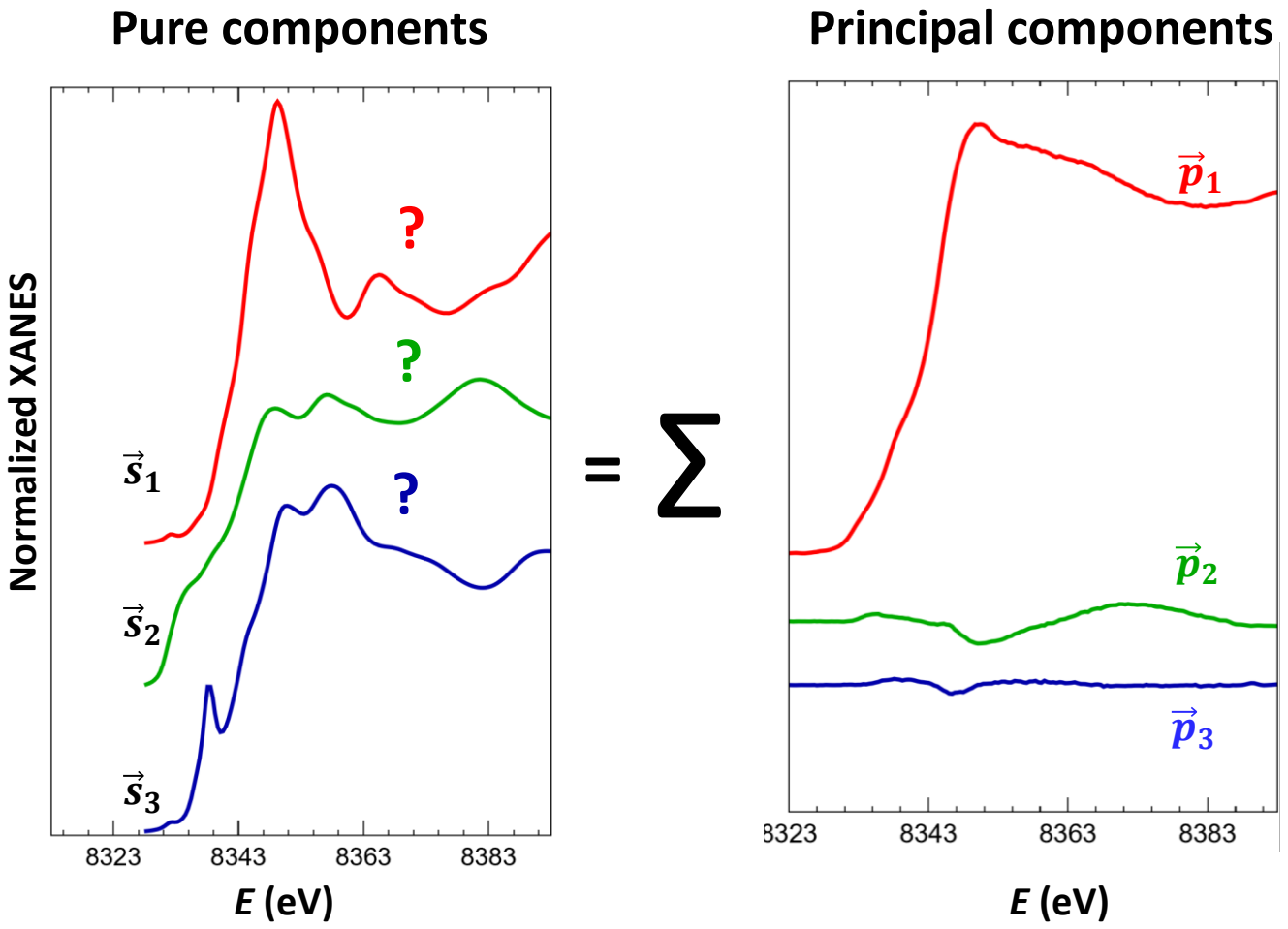


PCA: constraint – spectral components are linearly independent

Three components sufficient to represent variations in spectra → 3 different species coexisiting

Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

From principal components to spectra: transformation matrix



Spectra for pure species: some linear combinations of principal components with unknown weights

$$\vec{s}_1 = t_{11}\vec{p}_1 + t_{12}\vec{p}_2 + t_{13}\vec{p}_3$$

$$\vec{s}_2 = t_{21}\vec{p}_1 + t_{22}\vec{p}_2 + t_{23}\vec{p}_3$$

$$\vec{s}_3 = t_{31}\vec{p}_1 + t_{32}\vec{p}_2 + t_{32}\vec{p}_3$$

Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

From principal components to spectra: transformation matrix

$$\vec{s}_1 = t_{11}\vec{p}_1 + t_{12}\vec{p}_2 + t_{13}\vec{p}_3$$

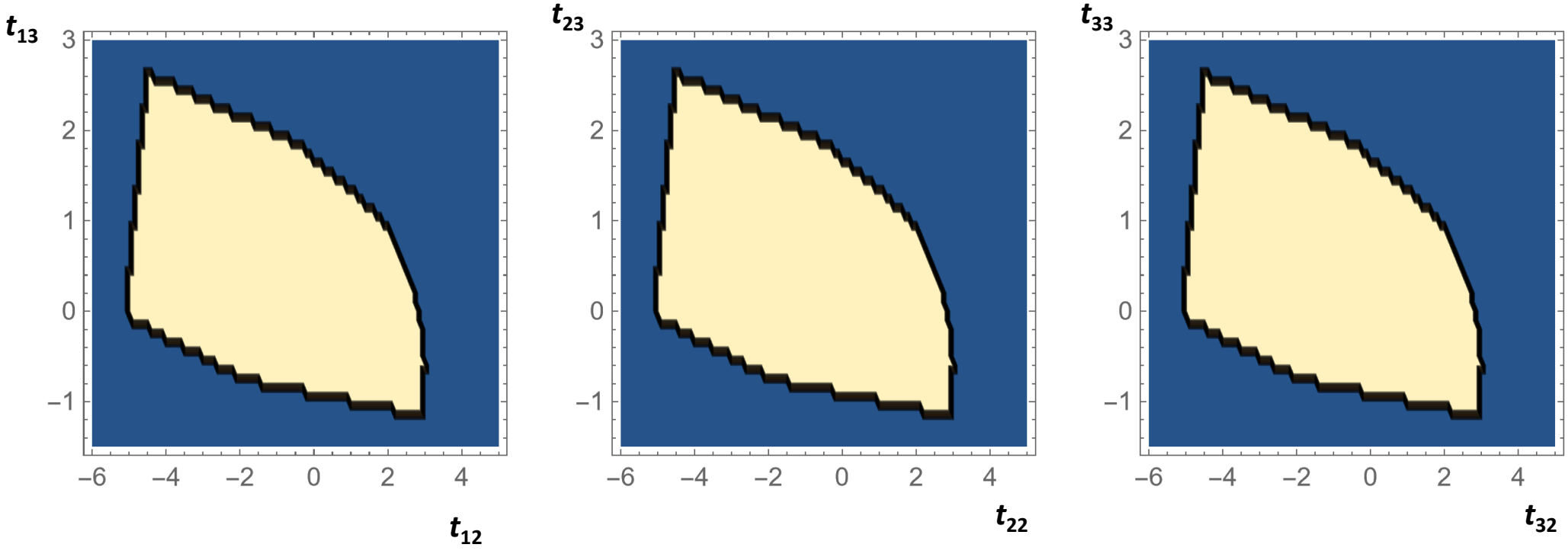
$$\vec{s}_2 = t_{21}\vec{p}_1 + t_{22}\vec{p}_2 + t_{23}\vec{p}_3$$

$$\vec{s}_3 = t_{31}\vec{p}_1 + t_{32}\vec{p}_2 + t_{32}\vec{p}_3$$

Constraints:

- Concentrations of all species adds to 1
- Spectra non-negative, normalized, with limited maximal intensity

Areas of feasible solutions



Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

From principal components to spectra: transformation matrix

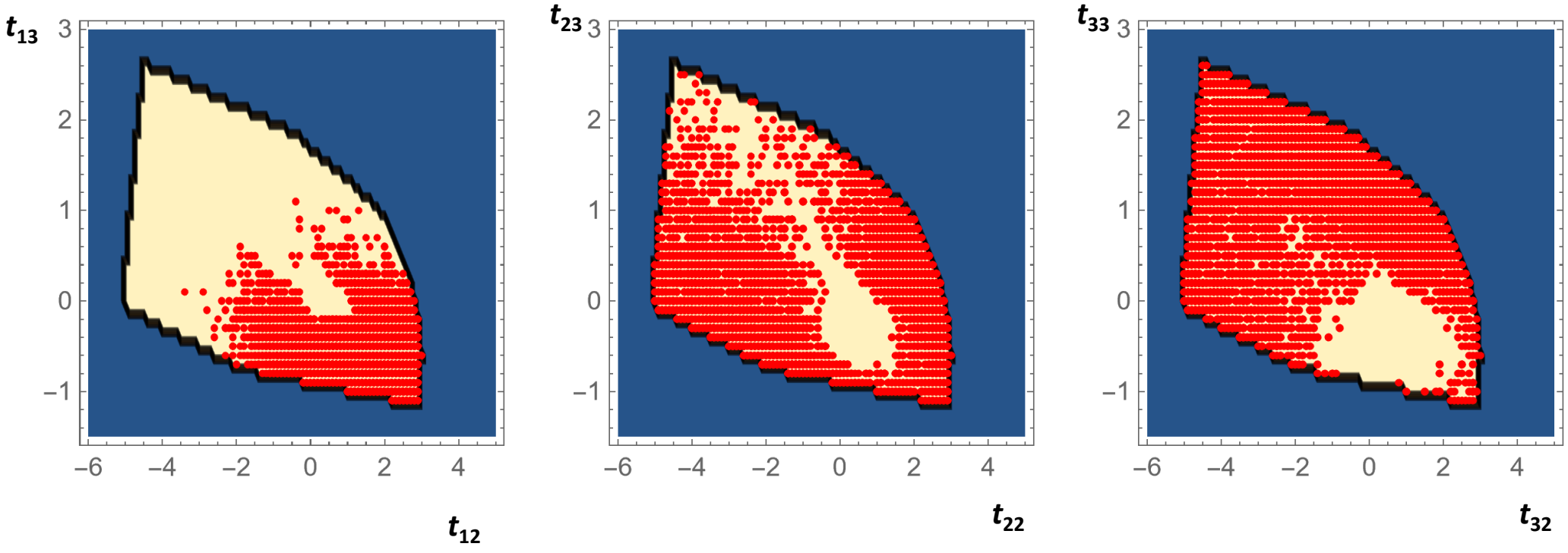
$$\vec{s}_1 = t_{11}\vec{p}_1 + t_{12}\vec{p}_2 + t_{13}\vec{p}_3$$

$$\vec{s}_2 = t_{21}\vec{p}_1 + t_{22}\vec{p}_2 + t_{23}\vec{p}_3$$

$$\vec{s}_3 = t_{31}\vec{p}_1 + t_{32}\vec{p}_2 + t_{32}\vec{p}_3$$

Constraints:

- Concentrations of all species adds to 1
- Spectra non-negative, normalized, with limited maximal intensity
- + Concentrations are between 0 and 1



Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

From principal components to spectra: transformation matrix

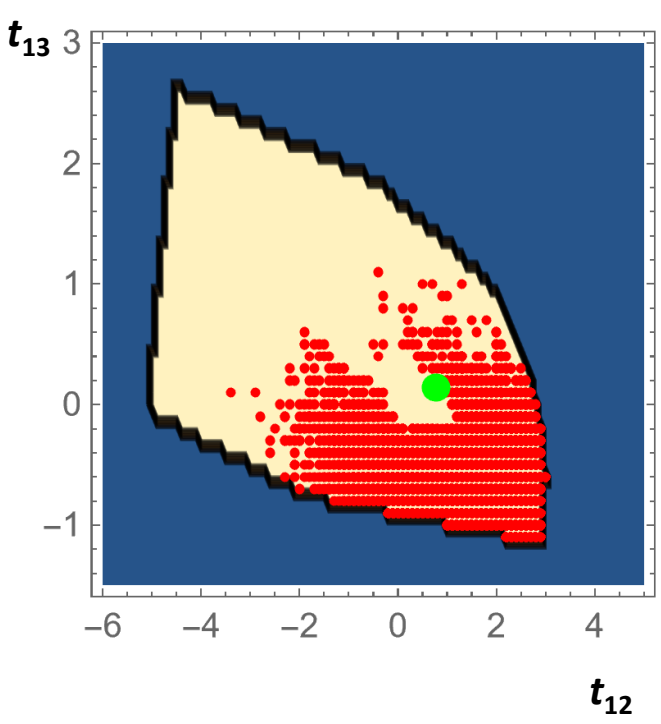
$$\vec{s}_1 = t_{11}\vec{p}_1 + t_{12}\vec{p}_2 + t_{13}\vec{p}_3$$

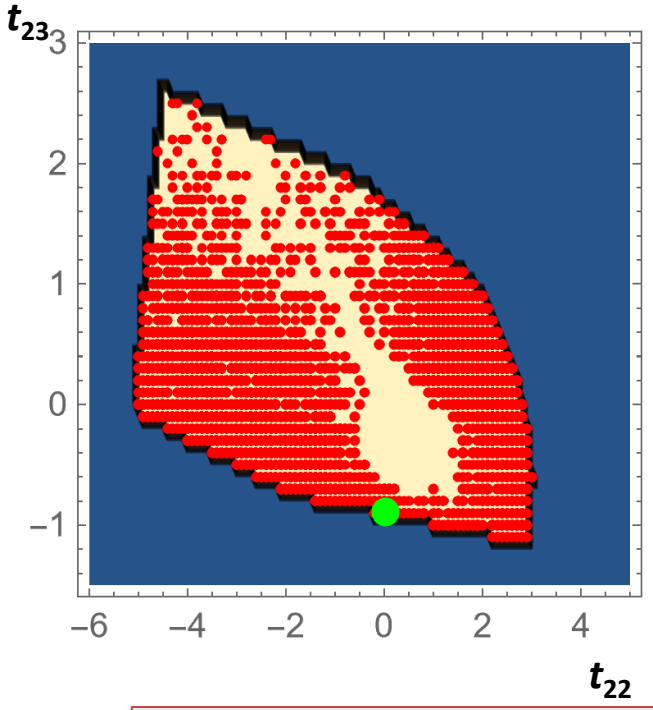
$$\vec{s}_2 = t_{21}\vec{p}_1 + t_{22}\vec{p}_2 + t_{23}\vec{p}_3$$

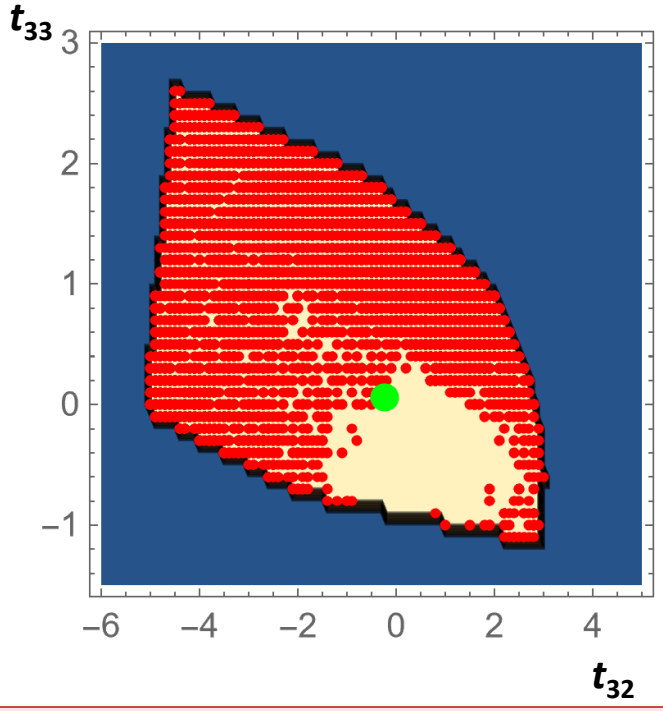
$$\vec{s}_3 = t_{31}\vec{p}_1 + t_{32}\vec{p}_2 + t_{32}\vec{p}_3$$

Constraints:

- Concentrations of all species adds to 1
- Spectra non-negative, normalized, with limited maximal intensity
- + Concentrations are between 0 and 1
- + 1st and last spectra correspond to pure species



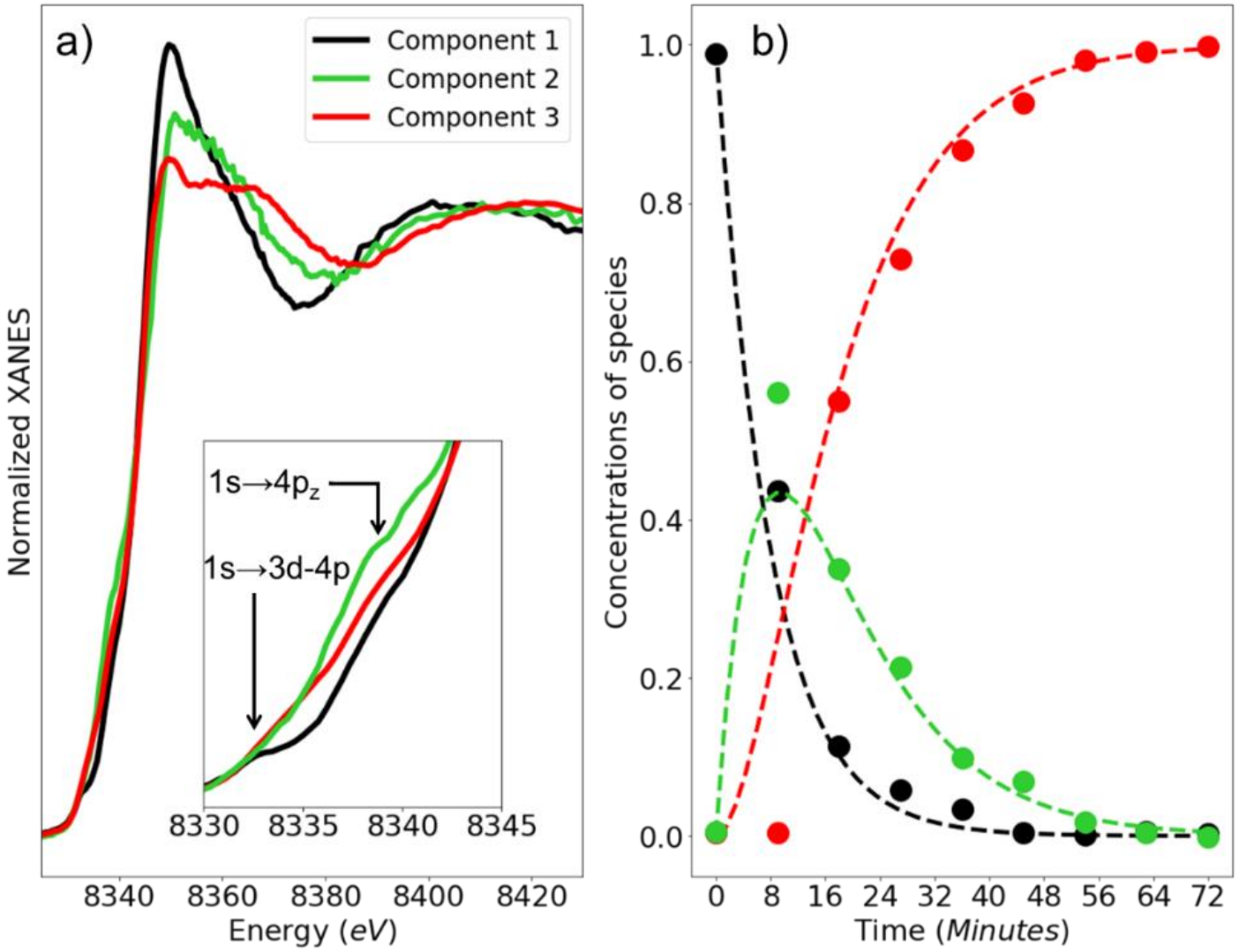




Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

- **It is not sufficient to show one solution to problem – one needs to show ALL solutions consistent with the model!**

From principal components to spectra: transformation matrix



Obtained spectra for pure species, and their concentration profiles.

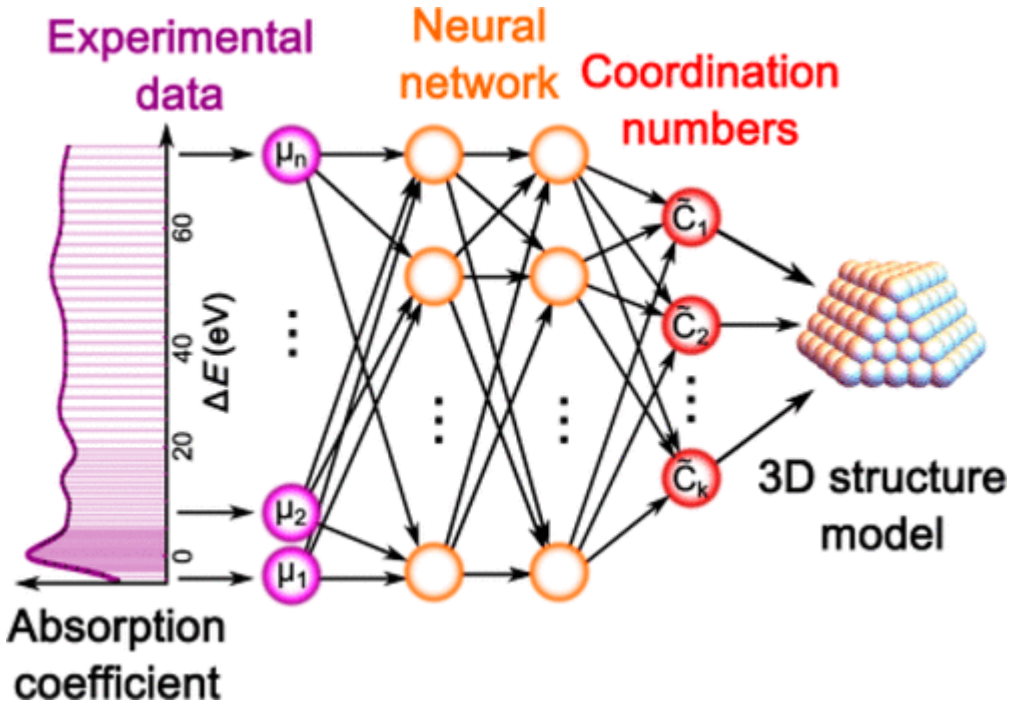
- What are these species? → XANES fitting

… but direct XANES fitting is computationally too expensive!

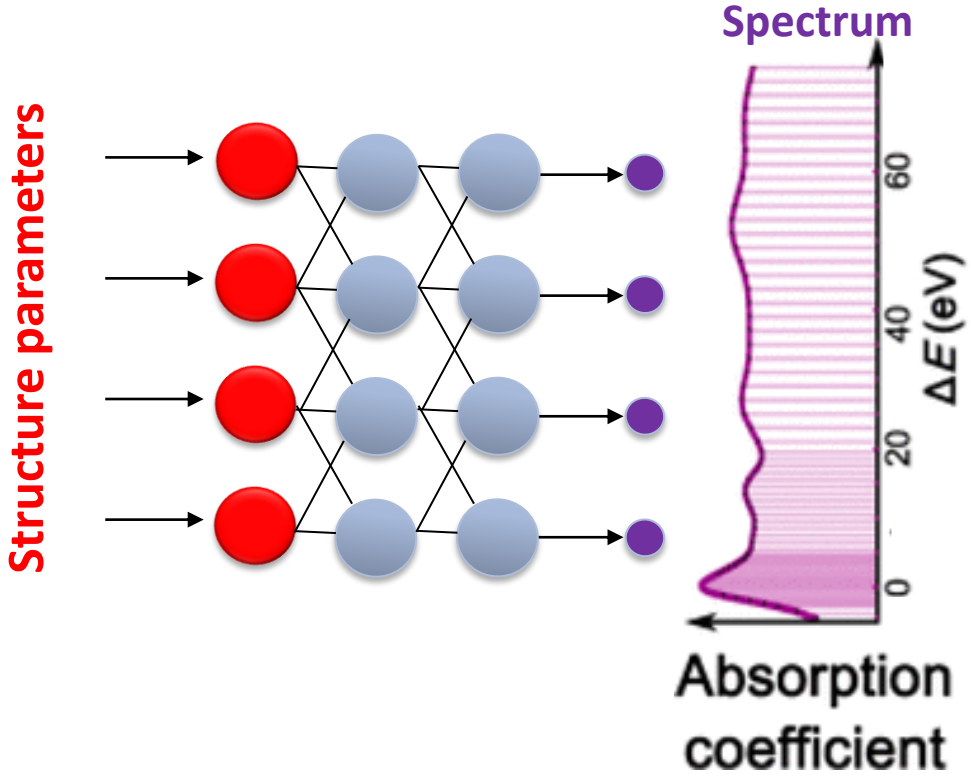
Martini, Roldan et al, J. Am. Chem. Soc. 145 (2023)

Direct vs. Inverse ML approach

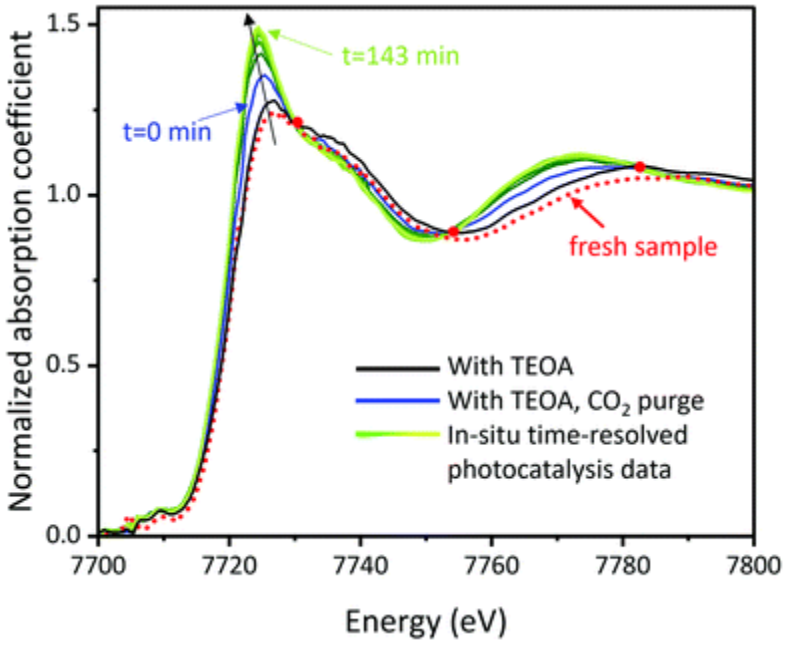
Direct approach

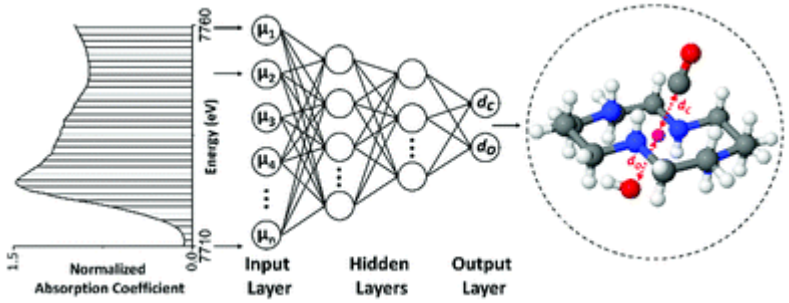


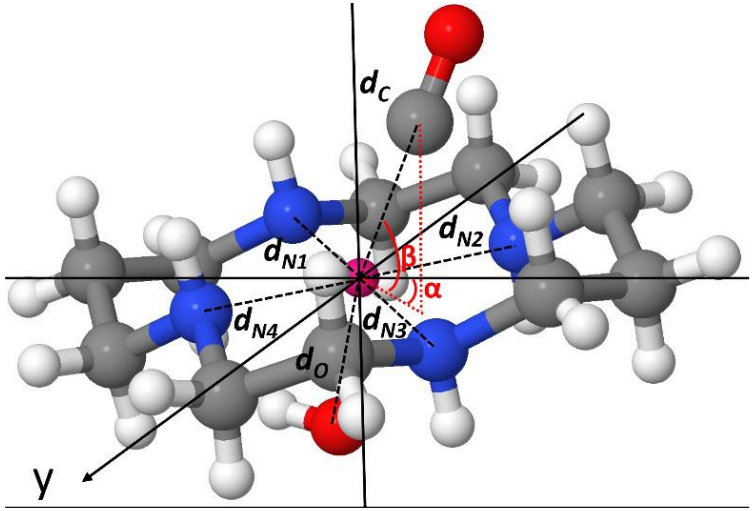
Inverse approach

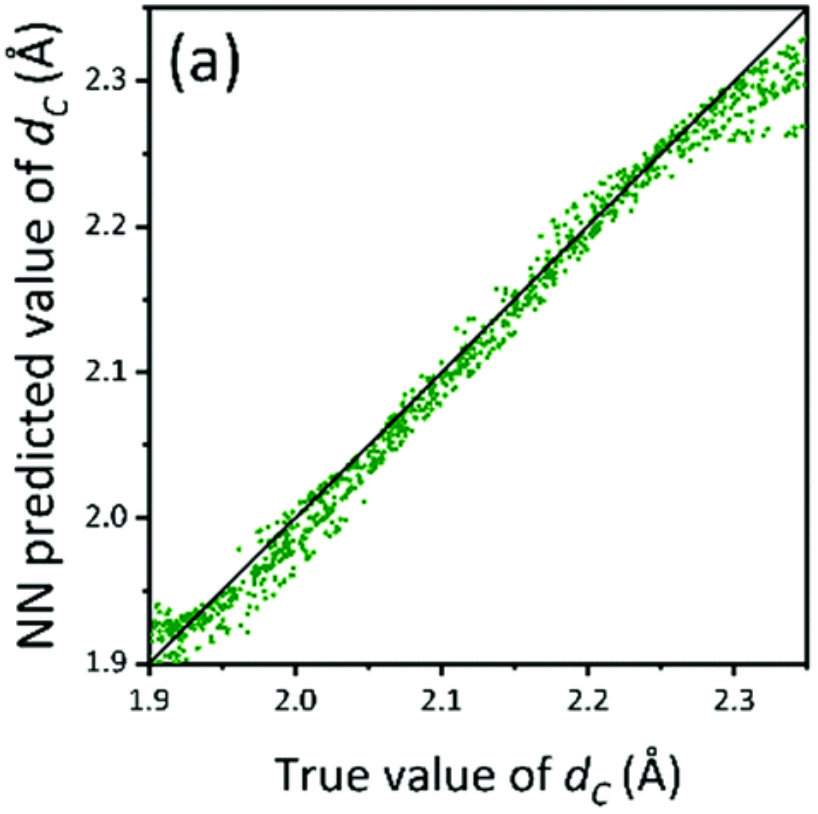


Direct ML approach for single atom catalysts



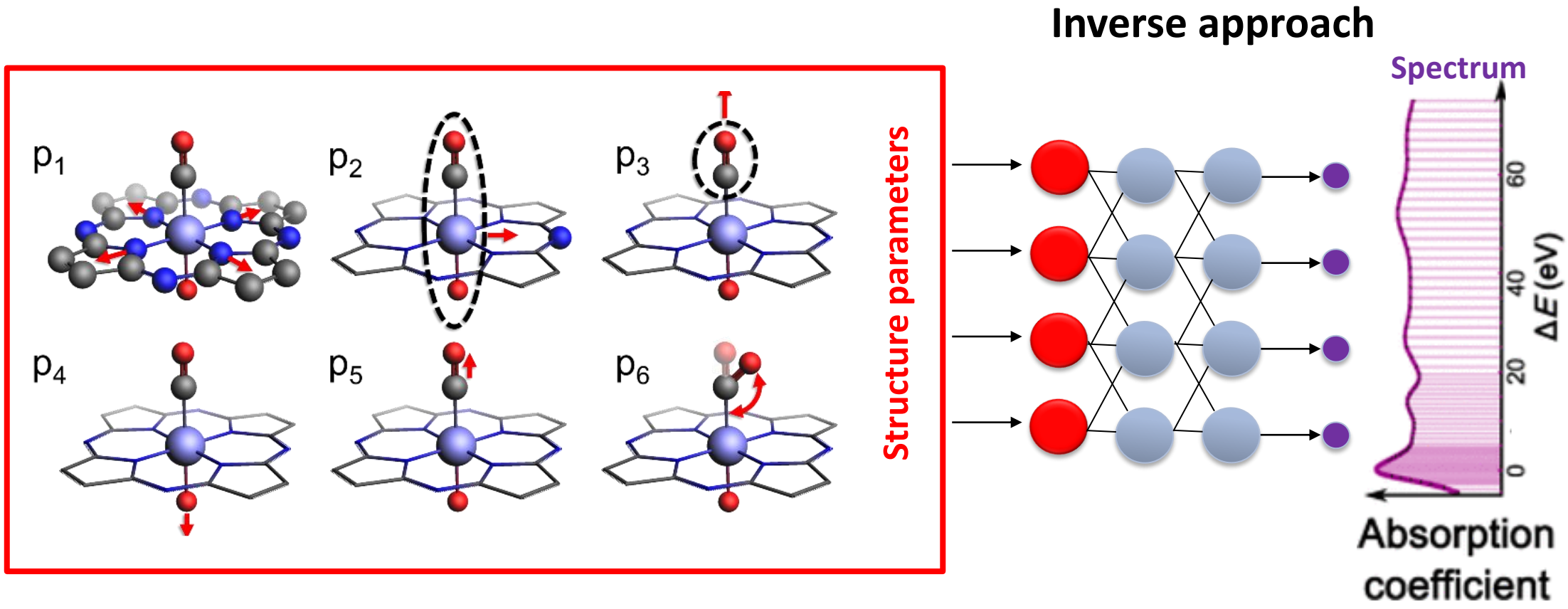






Xiang, Frenkel et al, Phys. Chem. Chem. Phys., 24 (2022)

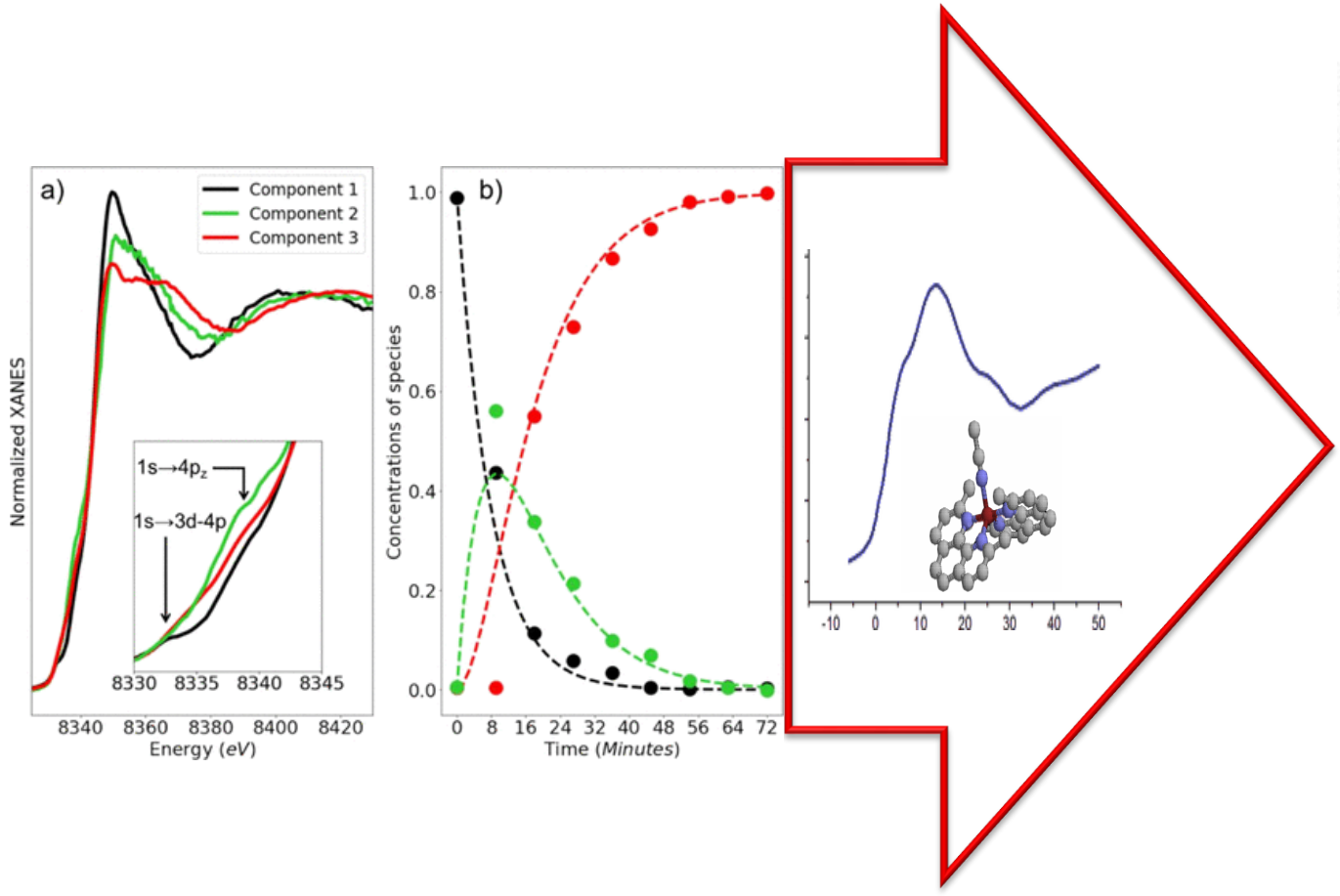
Inverse ML approach: analysis of XANES data for Ni single atom catalysts

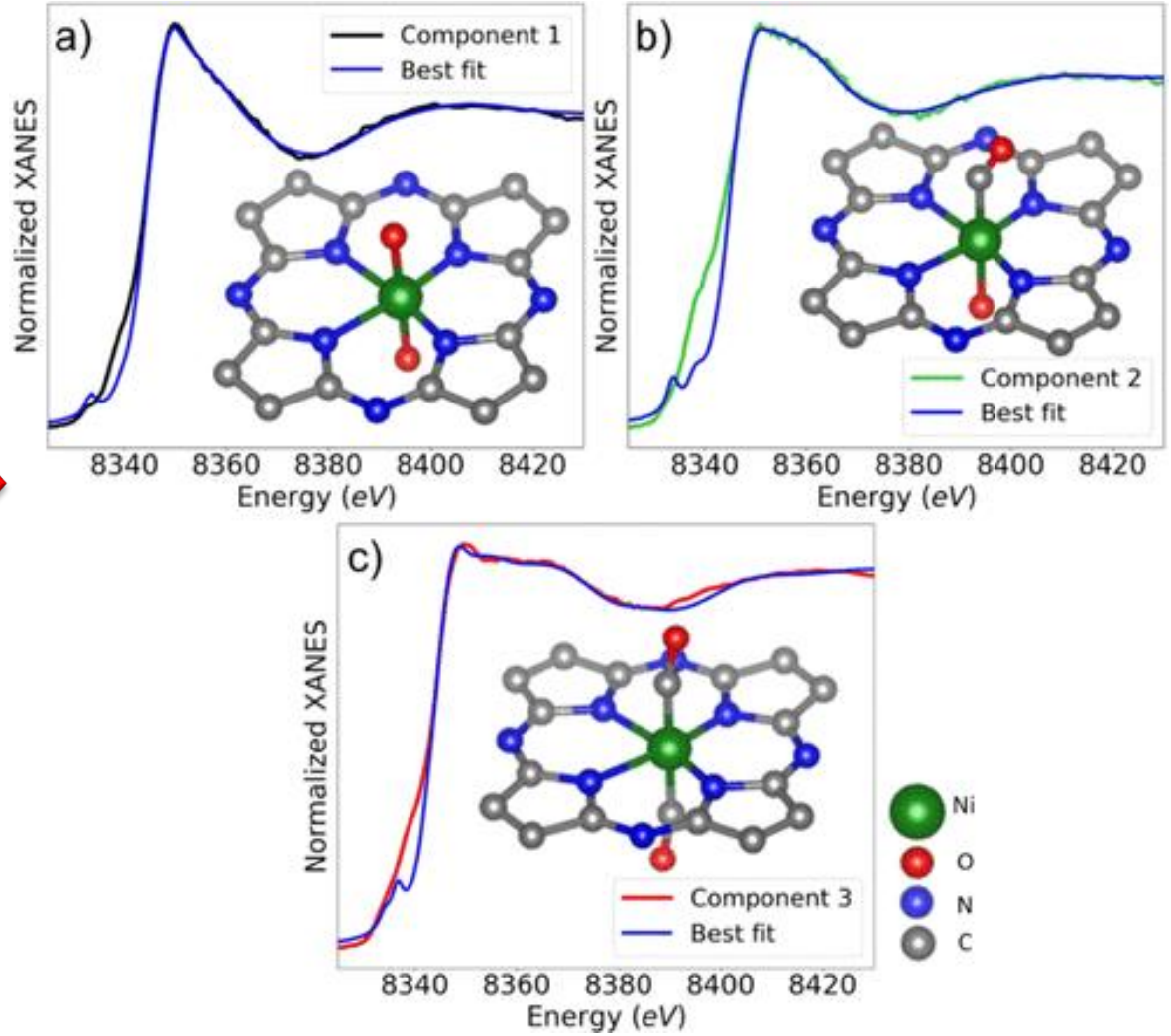


Martini, Timoshenko, et al, J. Am. Chem. Soc. 145 (2023)

Inverse ML approach: analysis of XANES data for Ni single atom catalysts

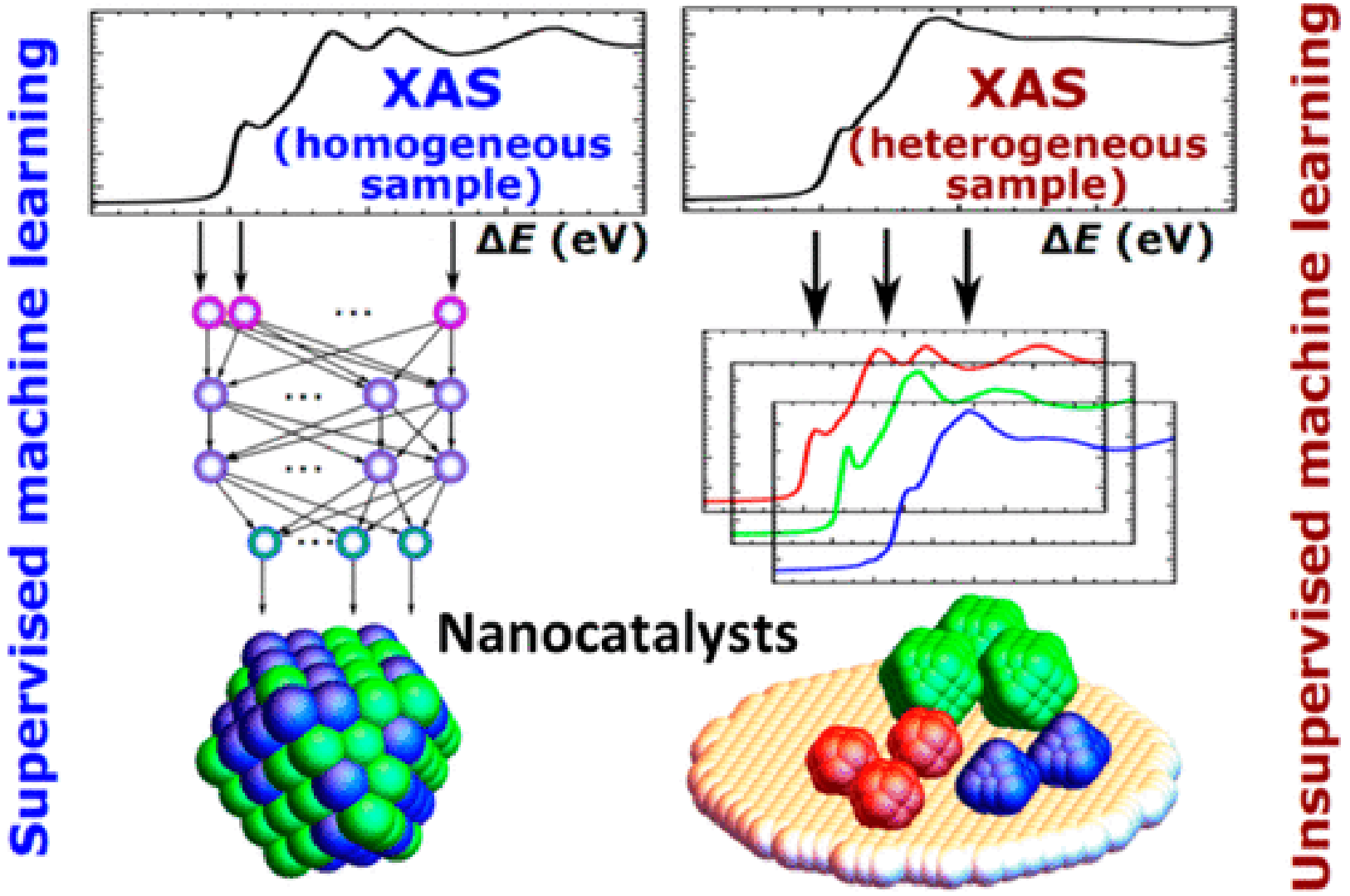
ML-assisted XANES fitting





Martini, Timoshenko, et al, J. Am. Chem. Soc. 145 (2023)

Summary

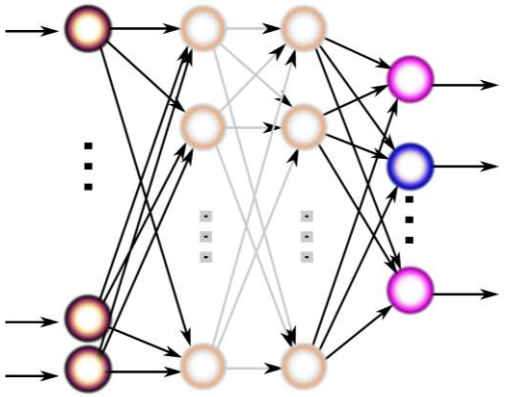

Supervised machine learning
XAS
(homogeneous sample)
ΔE (eV)
XAS
(heterogeneous sample)
ΔE (eV)
Nanocatalysts
Unsupervised machine learning

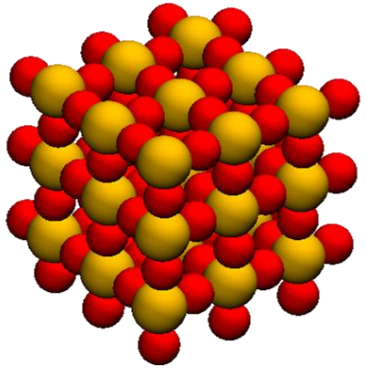
Summary

Experimental XANES/EXAFS data

Theoretical XAS data for training

Supervised Machine learning





Structure descriptors of

- **(nano)crystalline material**
- **Disordered materials**
- **Mixtures**

Accounts for

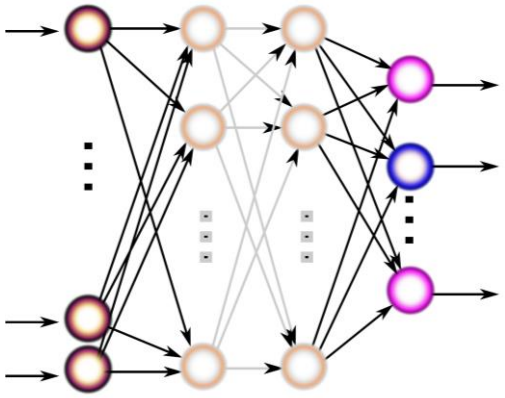
- **multiple-scattering effects**
- **contributions of further coordination shells**
- **thermal and static disorder of the system**

Summary

Experimental XANES/EXAFS data

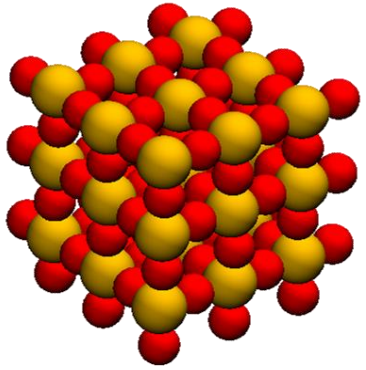
Theoretical XAS data for training

Supervised Machine learning



Structure descriptors of

- (nano)crystalline material
- Disordered materials
- Mixtures



Accounts for

- multiple-scattering effects
- contributions of further coordination shells
- thermal and static disorder of the system

Additonal advantages:

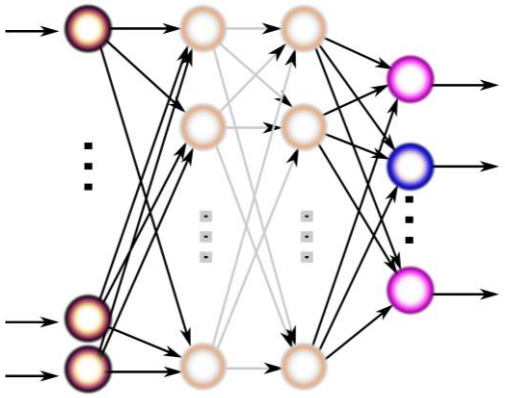
- Analysis can be done quickly (within seconds)
- Large data sets can be analyzed
- Insight from theory is helping in the analysis of experimental data (additional constraints)

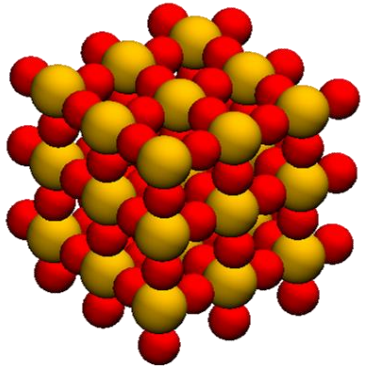
Summary

Experimental XANES/EXAFS data

Theoretical XAS data for training

Supervised Machine learning





Structure descriptors of

- **(nano)crystalline material**
- **Disordered materials**
- **Mixtures**

Additonal advantages:

- **Analysis can be done quickly (within seconds)**
- **Large data sets can be analyzed**
- **Insight from theory is helping in the analysis of experimental data (additional constraints)**

Accounts for

- **multiple-scattering effects**
- **contributions of further coordination shells**
- **thermal and static disorder of the system**

- **Supervised ML is bad at extrapolation: structure too different from those used for training → meaningless result**

Acknowledgements

Key collaborators:

Institute of Solid State Physics, University of Latvia **:**

Alexei Kuzmin, Inga Jonane-Pudza, Arturs Cintins, Andris Anspoks, Juris Purans

Stony Brook University/ Brookhaven National Lab:

Prof. Anatoly I. Frenkel, Nick Marcella, Yang Liu

Fritz-Haber Instiute of Max-Planck Society:

Prof. Beatriz Roldan Cuenya, Hyo Sang Jeon, Felix T. Haase

Thank you for your attention!