UNIVERSITÄT DUISBURG ESSEN

Neural-network potentials for atomistic material simulations

Hamidreza Hajiyani, Rossitza Pentcheva

[a] Department of Physics and Center of Nanointegration (CENIDE), Universität Duisburg-Essen, Duisburg, Germany

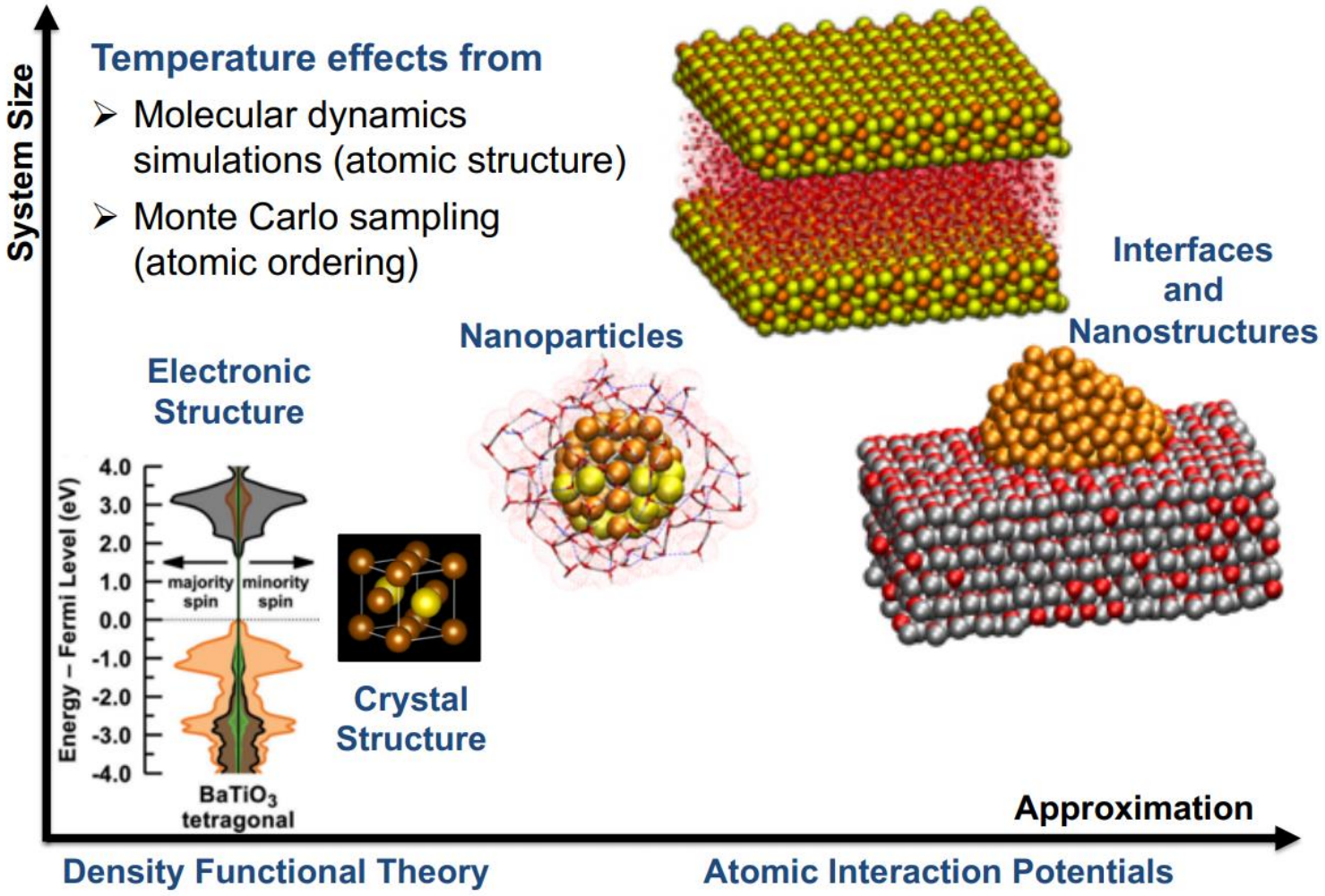
Deutsche Forschungsgemeinschaft DFG

Solar H_2 DFG SPP 1613

CRC/TRR 247 HETEROGENEOUS OXIDATION CATALYSIS IN THE LIQUID PHASE

lrz

Computational Methods and Length Scales



DFT limitations:

- System size: A few hundred atoms
- Time scale: Several hundred picoseconds
- Algorithmic scaling: O(N^3) or O($N\ ln(N)$), N is the number of electrons
  The system size accessible by DFT does not grow as fast as the computational power

DFT limitations:

- System size: A few hundred atoms
- Time scale: Several hundred picoseconds
- Algorithmic scaling: O(N^3) or O($N\ ln(N)$), N is the number of electrons
  The system size accessible by DFT does not grow as fast as the computational power

Empirical atomic potentials limitations:

- Rely on fitting parameters to reproduce the experimental or DFT results which scales with the complicity of the system (up to 30 parameters)
- Validation for each system in necessary
- Not every model is suitable for every applications:
  Embedded atom model (EAM) potentials are go for description of metallic solids but less adequate for molecules.

DFT limitations:

- System size: A few hundred atoms
- Time scale: Several hundred picoseconds
- Algorithmic scaling: O(N^3) or O($N\ ln(N)$), N is the number of electrons

The system size accessible by DFT does not grow as fast as the computational power

Empirical atomic potentials limitations:

- Rely on fitting parameters to reproduce the experimental or DFT results which scales with the complicity of the system (up to 30 parameters)
- Validation for each system in necessary
- Not every model is suitable for every applications:

Embedded atom model (EAM) potentials are go for description of metallic solids but less adequate for molecules.

Circumvent the parametrization problem by using machine learning techniques

Instead of fitting the parameters of a predefined model to reference data, an appropriate model that is able to describe the feature-space of the input data shall be automatically determined and parametrized by the machine learning method.

BP potentials for artificial neural-networks (ANNs)

Behler and Parrinello introduced one of the first machine-learning potentials which was based the artificial neural-network (1, 2).

Advantage

- Substitute the physical model with mathematical representations with hundreds to thousands of parameters which can be well automated with minimal human intervention.
- Reach the accuracy of DFT
- Designed potential can be applied into arbitrary structures.

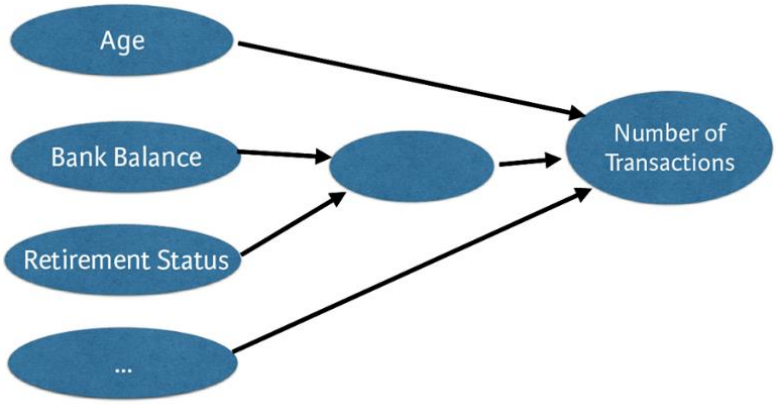
Disadvantage

- Some physical properties can not be investigated

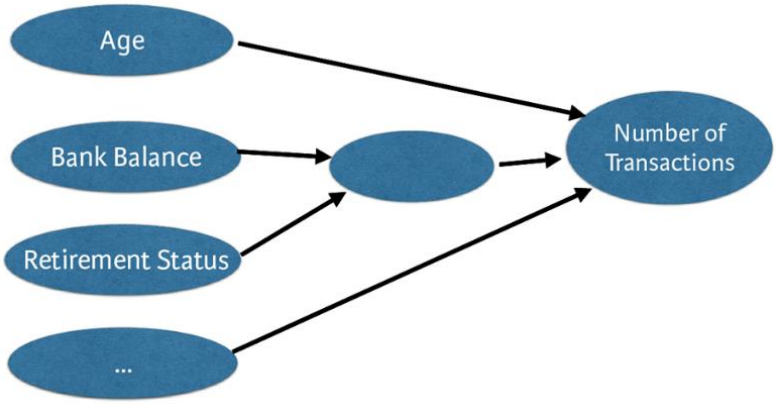
1) J. Behler, M. Parrinello, Phys. Rev. Lett., 98 (2007), p. 146401
2) J. Behler, Int. J. Quant. Chem., 115 (2015), pp. 1032-1050

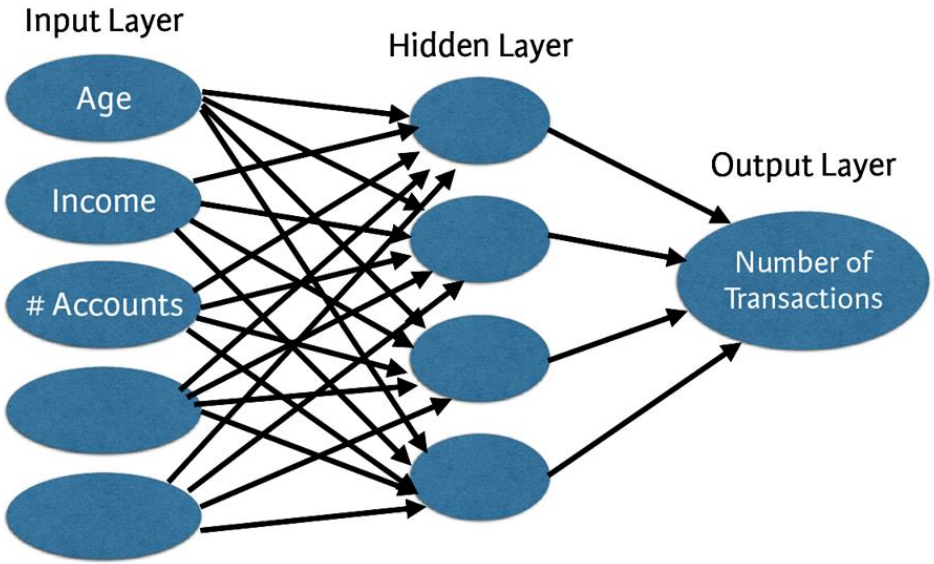
What is the neural-network potentials

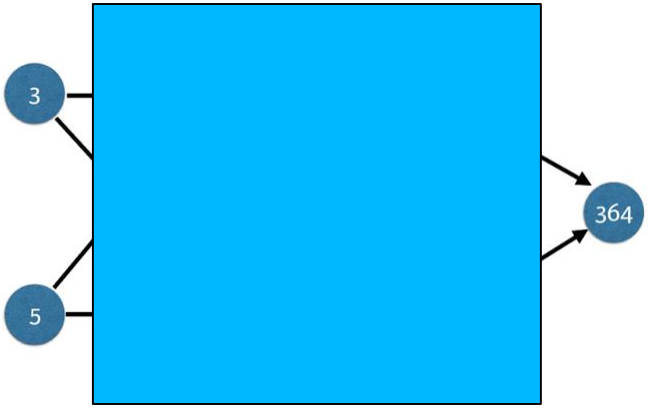
UNIVERSITÄT DUISBURG ESSEN



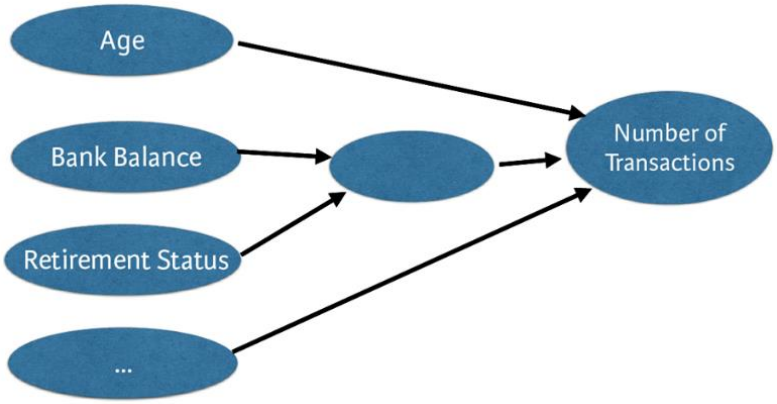
What is the neural-network potentials

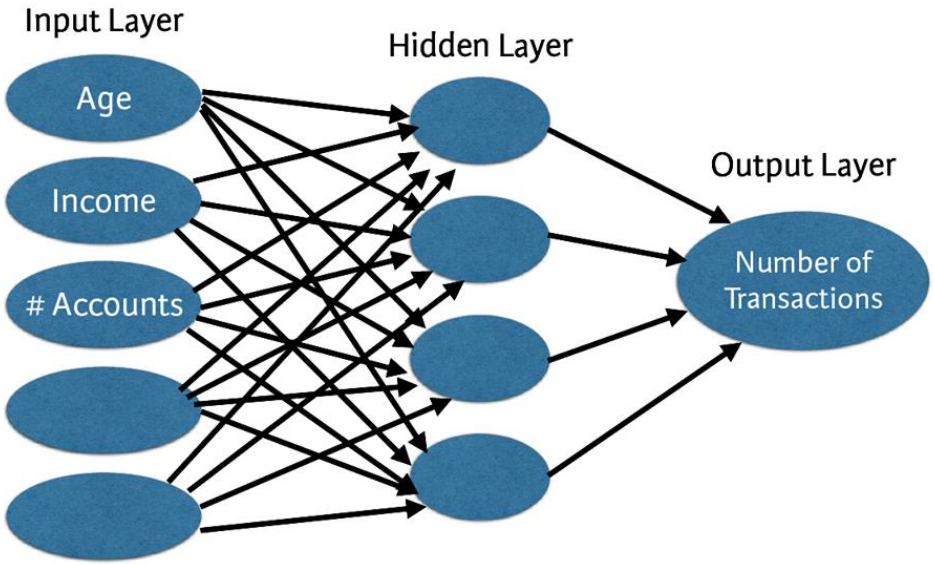


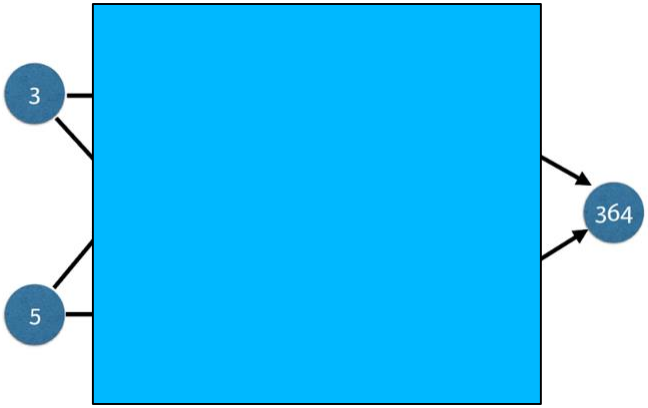


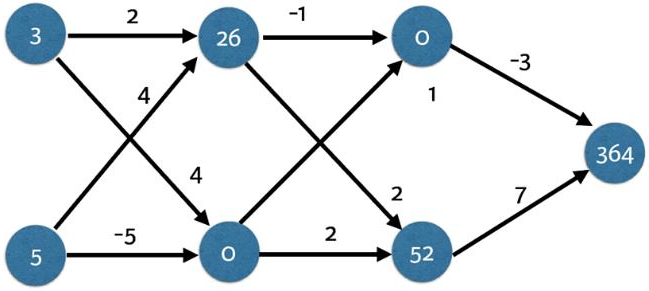


What is the neural-network potentials

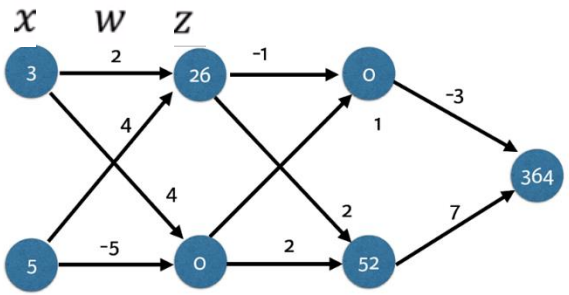




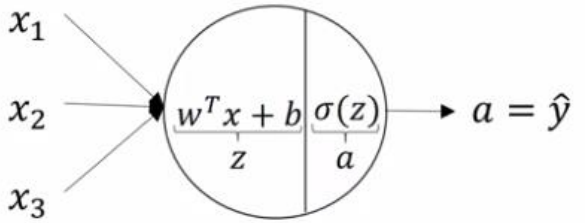




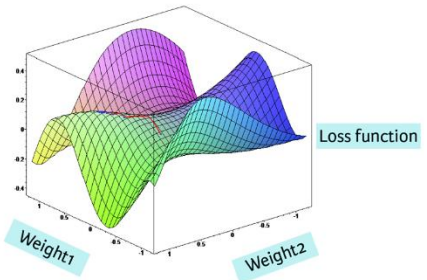
Logistic regression: Binary classification

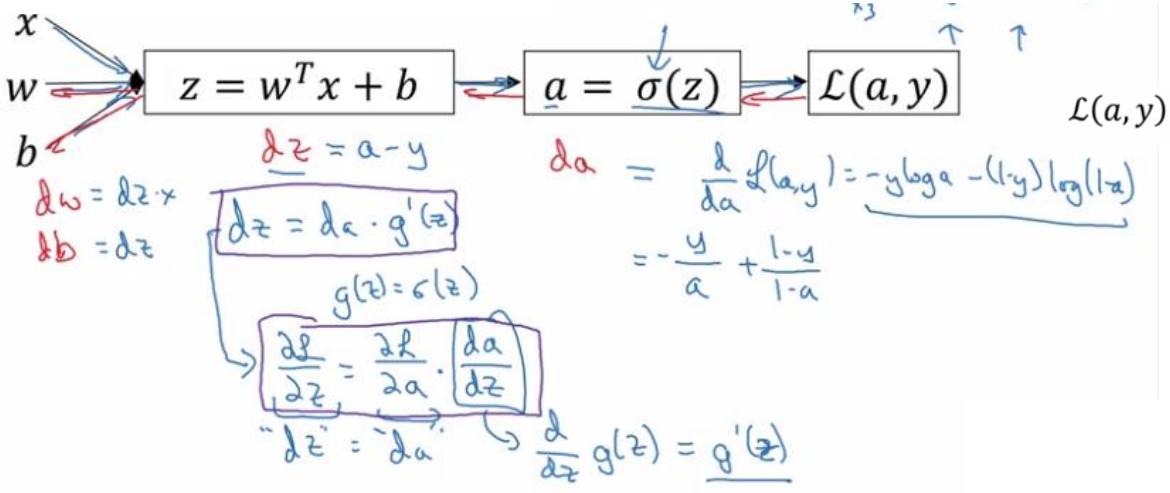


Calculate with ReLU Activation Function



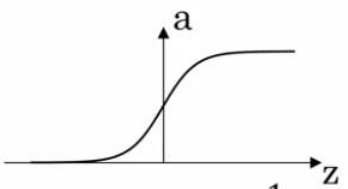
Loss function



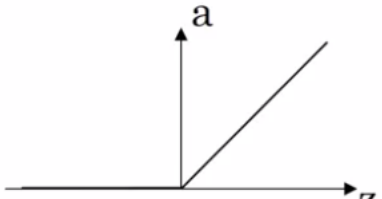


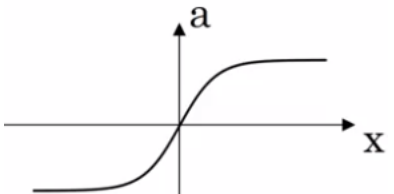
$$\mathcal{L}(a, y) = -(y \log(a) + (1 - y) \log(1 - a))$$

- Activation function:



sigmoid: $a = \frac{1}{1 + e^{-z}}$





Forward and backward propagation

UNIVERSITÄT DUISBURG ESSEN

Forward propagation

$$
\begin{aligned}
Z^{[1]} &= W^{[1]}X + b^{[1]} \\
A^{[1]} &= g^{[1]}(Z^{[1]}) \\
Z^{[2]} &= W^{[2]}A^{[1]} + b^{[2]} \\
A^{[2]} &= g^{[2]}(Z^{[2]}) \\
&\vdots \\
A^{[L]} &= g^{[L]}(Z^{[L]}) = \hat{Y}
\end{aligned}
$$

Backward propagation

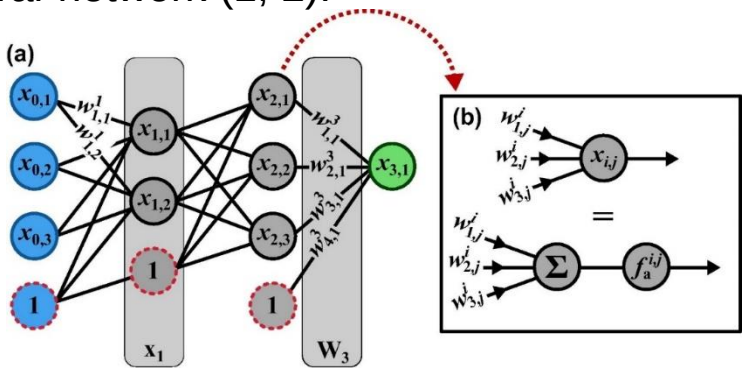
$$
\begin{aligned}
dZ^{[L]} &= A^{[L]} - Y \\
dW^{[L]} &= \frac{1}{m} dZ^{[L]} A^{[L]T} \\
db^{[L]} &= \frac{1}{m} np.\mathrm{sum}(\mathrm{d}Z^{[L]}, axis = 1, keepdims = True) \\
dZ^{[L-1]} &= dW^{[L]T} dZ^{[L]} g'^{[L]}(Z^{[L-1]}) \\
&\vdots \\
dZ^{[1]} &= dW^{[L]T} dZ^{[2]} g'^{[1]}(Z^{[1]}) \\
dW^{[1]} &= \frac{1}{m} dZ^{[1]} A^{[1]T} \\
db^{[1]} &= \frac{1}{m} np.\mathrm{sum}(\mathrm{d}Z^{[1]}, axis = 1, keepdims = True)
\end{aligned}
$$

$$
\begin{aligned}
w_1 &:= w_1 - \alpha \underline{dw_1} \\
w_2 &:= w_2 - \alpha \underline{dw_2} \\
b &:= b - \alpha \underline{db}.
\end{aligned}
$$

Multilayer perceptron (MPL) artificial neural network

UNIVERSITÄT DUISBURG ESSEN

- One of the first machine-learning potentials was introduced by Behler and Parrinello that was based the artificial neural-network (1, 2).
- 3-2-3-1 architecture



- Value in i-th layer

$$\mathbf{x}_i(\mathbf{x}_{i-1}) = f_a^i(\mathbf{W}_i\mathbf{x}_{i-1}) \quad \text{with} \quad i = 1, \ldots, (N-1),$$

- Expression of above structure

$$\mathcal{N}(\mathbf{x}_0; \{\mathbf{W}_i\}) = f_a^3\left\{\mathbf{W}_3 f_a^2\left[\mathbf{W}_2 f_a^1(\mathbf{W}_1\,\mathbf{x}_0)\right]\right\} = \mathbf{x}_3.$$

- Error for each reference sample *n*

$$e_n(\mathbf{x}_{0,n}, \mathbf{y}_n; \{\mathbf{W}_\ell\}) = \mathcal{N}(\mathbf{x}_{0,n}; \{\mathbf{W}_\ell\}) - \mathbf{y}_n$$

- Optimization process:
Minimization of error function

$$\{\mathbf{W}_\ell^{\text{opt}}\} = \arg\min_{\{\mathbf{W}_\ell\}} \mathcal{E}(\{\mathbf{W}_\ell\}) \quad \text{with} \quad \mathcal{E}(\{\mathbf{W}_\ell\}) = \frac{1}{2}\sum_n^{\text{samples}} e_n^2,$$

$$\nabla\mathcal{E} = \mathbf{J}^T\mathbf{e} \quad \text{with} \quad (\mathbf{J})_{\alpha n} = \frac{\partial\mathcal{N}(\mathbf{x}_{0,n})}{\partial w_\alpha} \quad \text{and} \quad (\mathbf{e})_n = e_n$$

$$\frac{\partial x_{i,j}}{\partial w_{m,j}^i} = f_a^{i,j\,(1)}\left(\sum_k w_{k,j}^i x_{i-1,k}\right)\cdot x_{i-1,m},$$

1) J. Behler, M. Parrinello, Phys. Rev. Lett., 98 (2007), p. 146401
2) J. Behler, Int. J. Quant. Chem., 115 (2015), pp. 1032-1050

Minimization of error function:

UNIVERSITÄT DUISBURG ESSEN

- Optimization process:
Minimization of error function

$$\{\mathbf{W}_\ell^{\text{opt}}\} = \arg\min_{\{\mathbf{W}_\ell\}} \mathcal{E}(\{\mathbf{W}_\ell\}) \quad \text{with} \quad \mathcal{E}(\{\mathbf{W}_\ell\}) = \frac{1}{2}\sum_n^{\text{samples}} e_n^2,$$

$$\nabla\mathcal{E} = \mathbf{J}^T\mathbf{e} \quad \text{with} \quad (\mathbf{J})_{\alpha n} = \frac{\partial\mathcal{N}(\mathbf{x}_{0,n})}{\partial w_\alpha} \quad \text{and} \quad (\mathbf{e})_n = e_n$$

$$\frac{\partial x_{i,j}}{\partial w^i_{m,j}} = f_a^{i,j\,(1)}\left(\sum_k w^i_{k,j} x_{i-1,k}\right)\cdot x_{i-1,m},$$

1. Gradient descent:

$$\mathbf{w}^{(I+1)} = \mathbf{w}^{(I)} + \Delta\mathbf{w}^{\text{GD},(I+1)} \quad \text{with} \quad \Delta\mathbf{w}^{\text{GD},(I+1)} = -\gamma\nabla\mathcal{E},$$

2. Limited-memory BFGS (L-BFGS):
BFGS: Broyden–Fletcher–Goldfarb–Shanno method

$$\Delta\mathbf{w}^{\text{QN},(I+1)} = -\left(\mathbf{H}^{(I)}\right)^{-1}\nabla\mathcal{E}^{(I)},$$

Hessian matrix

3. Levenberg-Marquardt (LM)

$$\Delta\mathbf{w}^{\text{LM},(I+1)} = -\left(\mathbf{J}^{T,(I)}\mathbf{J}^{(I)} + \lambda\mathbf{I}\right)^{-1}\mathbf{J}^{T,(I)}\mathbf{e}^{(I)},$$

Jacobian matrix

C. Broyden. J. Inst. Math. Appl., 6 (1970), pp. 222-231,
R. Fletcher. Comput. J., 13 (1970), pp. 317-322,
D. Goldfarb. Math. Comput., 24 (1970), pp. 23-26,
D. Shanno. Math. Comput., 24 (1970), pp. 647-656,

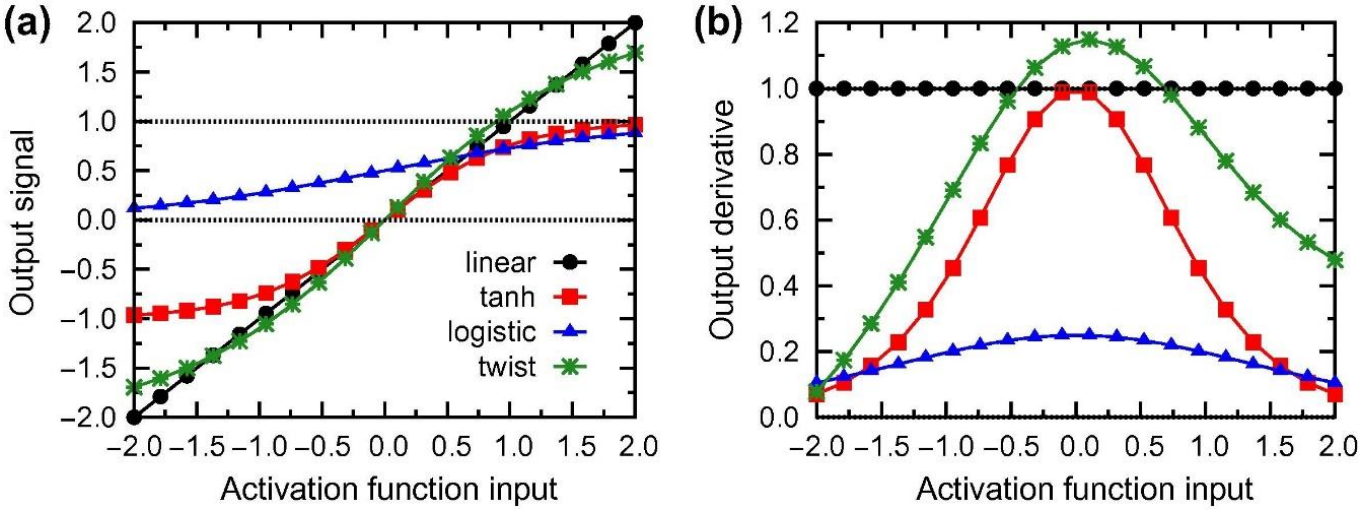
BP method for artificial neural-networks (ANNs)

UNIVERSITÄT DUISBURG ESSEN

- Activation functions

$$
\begin{aligned}
&\text{linear function} && f_a^1(x) = x, \\
&\text{hyperbolic tangent} && f_a^2(x) = \tanh(x) = \tfrac{1-e^{-2x}}{1+e^{-2x}}, \\
&\text{logistic function} && f_a^3(x) = \tfrac{1}{1+e^{-x}}, \quad \text{and} \\
&\text{tanh with linear twisting} && f_a^4(x) = 1.7159 \tanh\left(\tfrac{2}{3}x\right) + ax,
\end{aligned}
$$

- Activation function and its derivatives



1) J. Behler, M. Parrinello, Phys. Rev. Lett., 98 (2007), p. 146401
2) J. Behler, Int. J. Quant. Chem., 115 (2015), pp. 1032-1050

What need to be used as a input

UNIVERSITÄT DUISBURG ESSEN

- Early approaches used the Cartesian atomic coordinates for the atomic configurations of $\sigma = \{R_i\}$ as input to represent the structural energy $E(\sigma)$.

$$E(\sigma) \approx E^{ANN}(\sigma) = N(\{R_i\})$$

- Problem: the number of ANN input nodes depended on the number of atoms in the structure, resulting specialized potentials that were not transferable to different number of atoms.

- New methods overcome this problem by partitioning the total energy into atomic contributions

$$E(\sigma) = \sum_i^{atoms} E_i(\sigma) \approx \sum_i^{atoms} E_i(\sigma_i)$$

Where $\sigma_i \subset \sigma$ only depends on the coordinates of atoms within a cutoff radius R around atom i. σ_i shows the local structure around atom i. machine learning methods derive a model for the atomic energy instead of the total structure energy.

Atomistic potentials

UNIVERSITÄT DUISBURG ESSEN

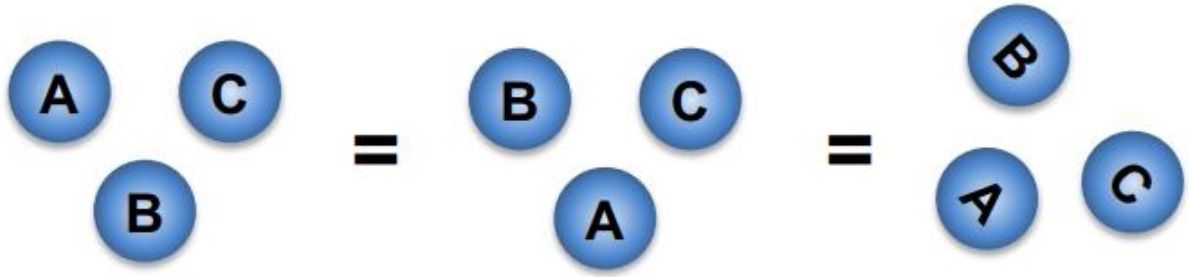
$$E(\sigma) = \sum_i^{atoms} E_i(\sigma) \approx \sum_i^{atoms} E_i(\sigma_i)$$

- The atomic energy σ_i has to be invariant with respect to:

Exchange of equivalent atoms (order of counting)

Translation/ rotation of the entire structure

Therefore, we transform the Cartesian coordinates to invariant coordinates by introducing the fingerprint functions with $\tilde{\sigma}_i = \tau(\sigma_i)$



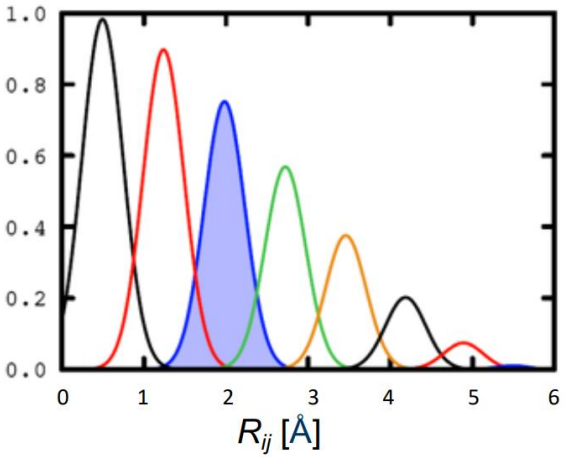
Representation of local structure environment

UNIVERSITÄT DUISBURG ESSEN

- Invariant bases set of radial and angular symmetry functions were introduced by Behler and Parrinello (BP):

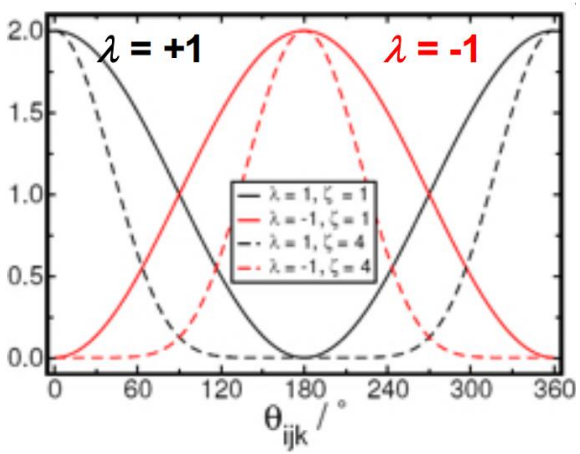
Radial symmetry functions

$$G_i^2 = \sum_j e^{-\eta\left(R_{ij}-R_s\right)^2} \cdot f_c\left(R_{ij}\right)$$

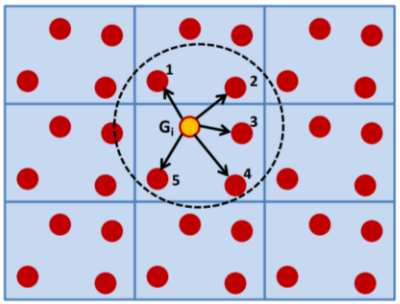


Angular symmetry functions

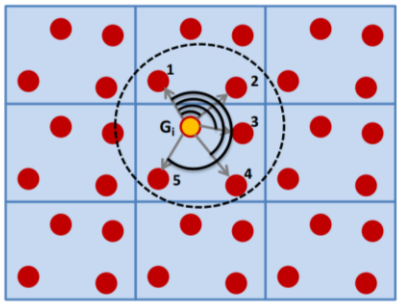
$$G_i = 2^{1-\zeta} \sum_i \sum_k \left[\left(1 + \lambda \cdot \cos\theta_{ijk}\right)^\zeta \cdot e^{-\eta\left(R_{ij}^2+R_{ik}^2+R_{jk}^2\right)} \cdot f_c\left(R_{ij}\right) \cdot f_c\left(R_{ik}\right) \cdot f_c\left(R_{jk}\right)\right.$$



Atomic Radial Distribution Function



Atomic Angular Distribution Function



J. Behler, and M. Parrinello, Phys. Rev. Lett. 98, (2007) 146401.

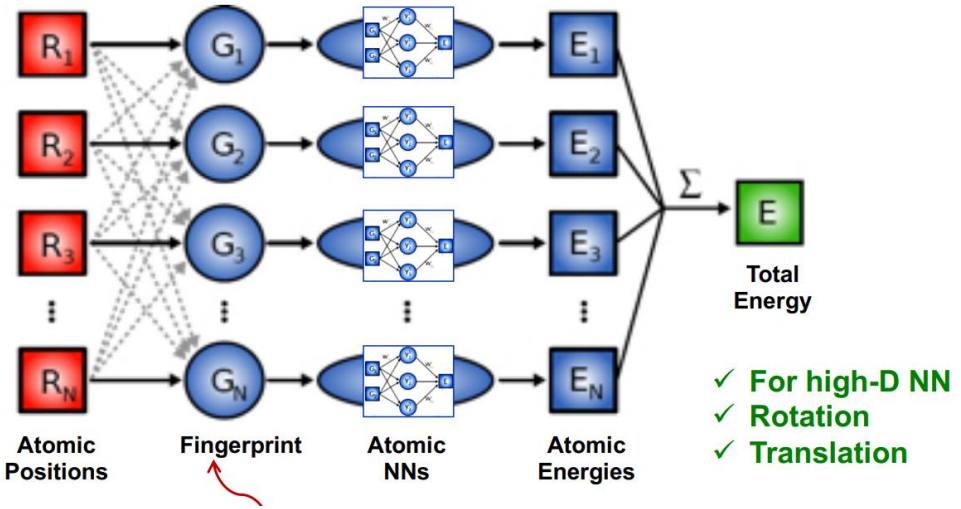
J. Behler, J. Chem. Phys. 134, (2011) 074106.

N. Artrith, T. Morawietz, J. Behler, Phys. Rev. B 83, (2011) 153101.

Representation of local structure environment

UNIVERSITÄT DUISBURG ESSEN

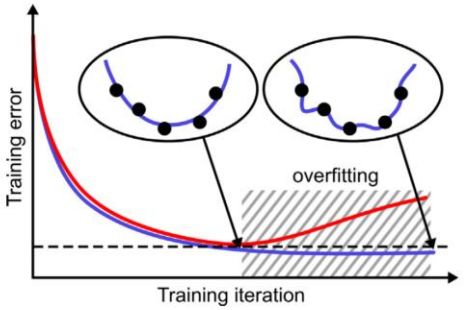
- Partitioning of the total energy into atomic contribution introduces an additional layer of complexity into the ANNs
- Training ANN potentials has to occur simultaneously for all atoms in structure



- Training process is monitored by the root mean squared error (RMSE) and mean absolute error (MAE)

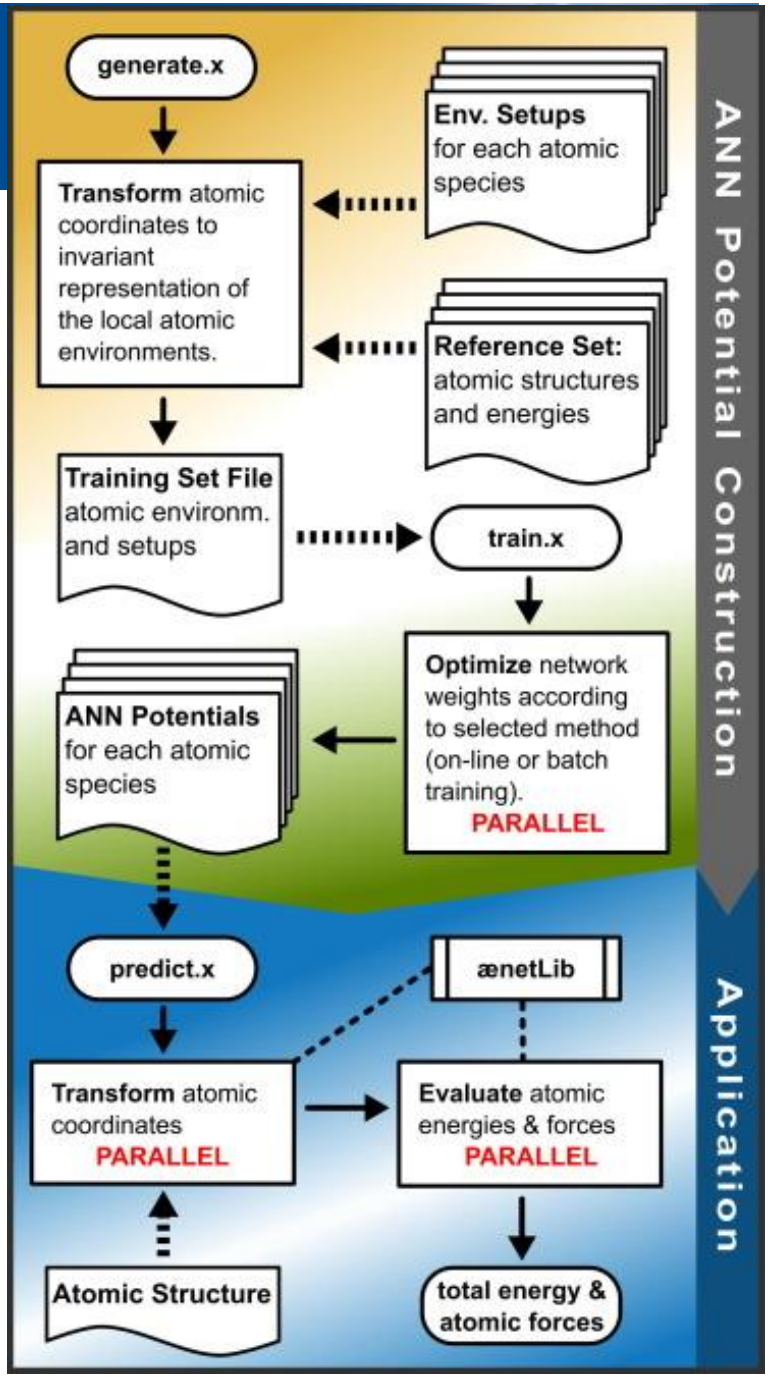
$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{\sigma}^{\text{structures}} \left[E^{\mathrm{ANN}}(\sigma) - E^{\sigma}_{\mathrm{ref}}\right]^2},$$

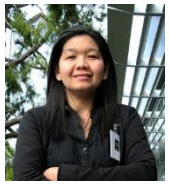
$$\mathrm{MAE} = \frac{1}{N}\sum_{\sigma}^{\text{structures}} \left|E^{\mathrm{ANN}}(\sigma) - E^{\sigma}_{\mathrm{ref}}\right|,$$



Implementation details: the aenet code

- The high-dimentional ANN potential approach in implemented in the *Atomic Energy Network* (**ænet**) software package [1,2]
- **ænet** is a modern Fortran and C code
- **ænet** provides tools for both the construction and application of ANN potentials




N. Artrith


G. Ceder

RuNNer Code


J. Behler

[1] N. Artrith and A. Urban, Comput. Mater. Sci. 114 (2016) 135-150.
[2] N. Artrith, A. Urban, and G. Ceder, Phys. Rev. B 96 (2017) 014112.

Implementation details: the aenet code

- Construction:

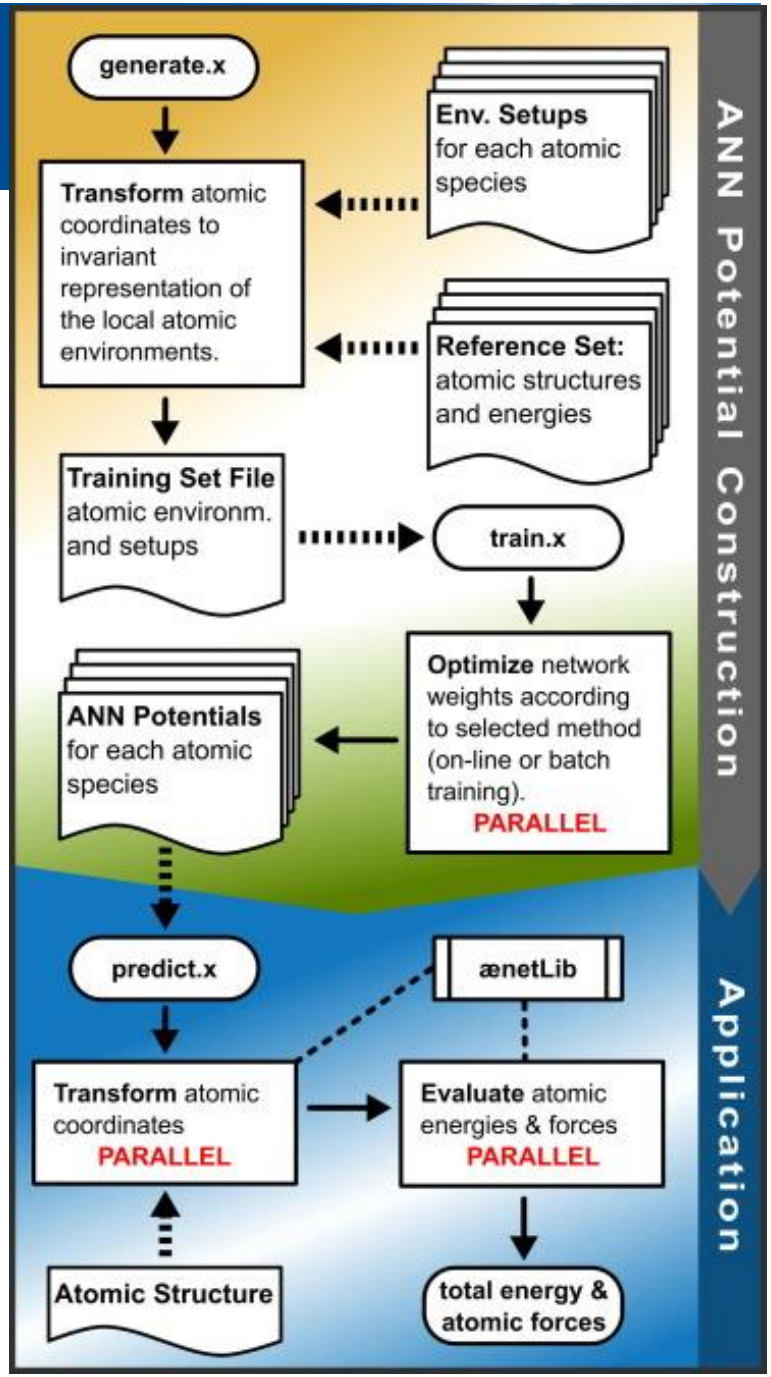
1. Selection of suitable reference structure and the computation of their energy (by DFT) with the XFS format.

2. Transformation of Cartesian coordinates of the structures to the invariant, atom-centered bases (generate.x)

- Training:

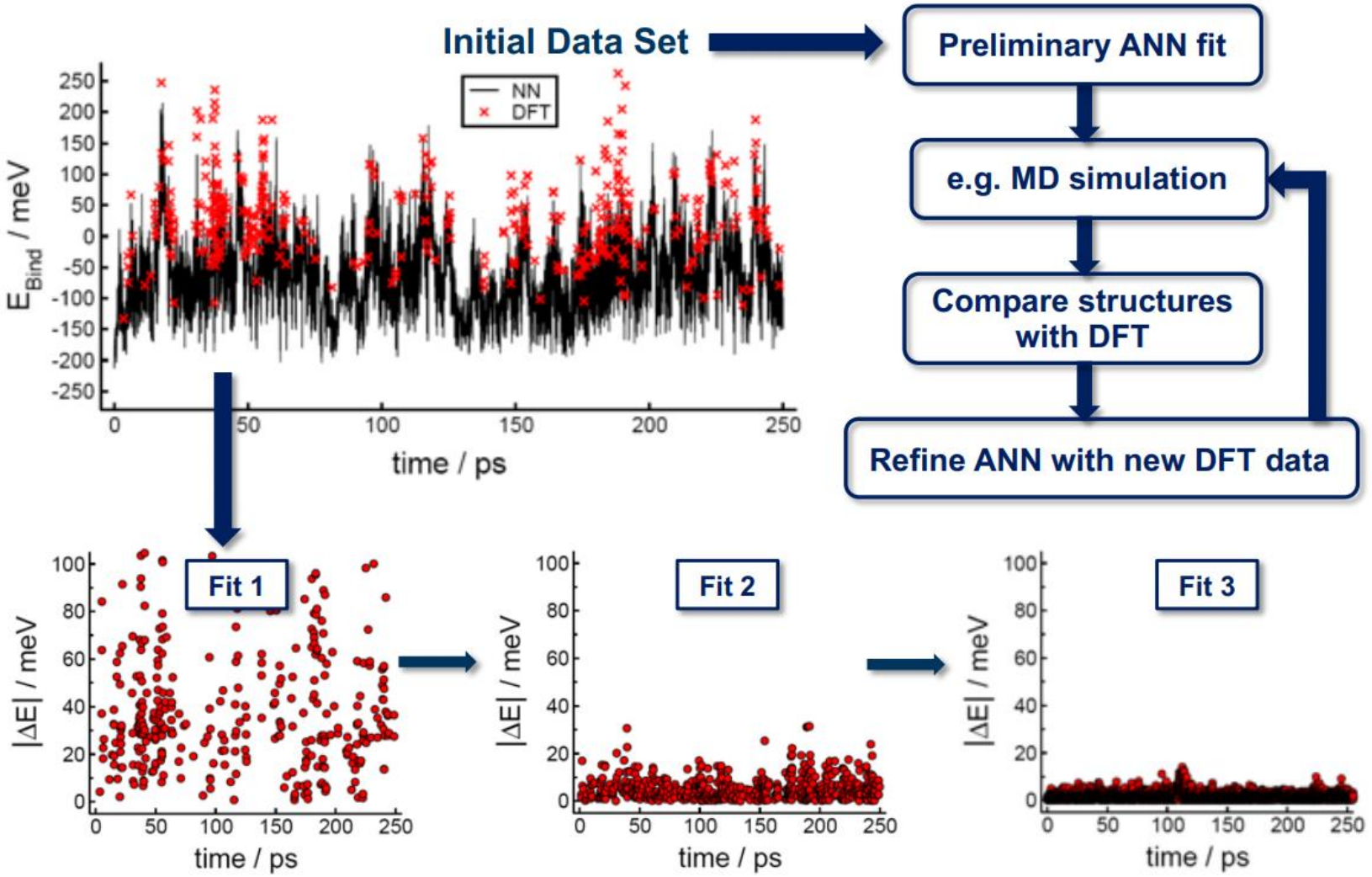
1. Specify ANN architectures for all chemical species
2. Define activation function of each hidden layer

Apply ANN potentials for MD or MC simulations



[1] N. Artrith and A. Urban, Comput. Mater. Sci. 114 (2016) 135-150.
[2] N. Artrith, A. Urban, and G. Ceder, Phys. Rev. B 96 (2017) 014112.

Construction of training dataset

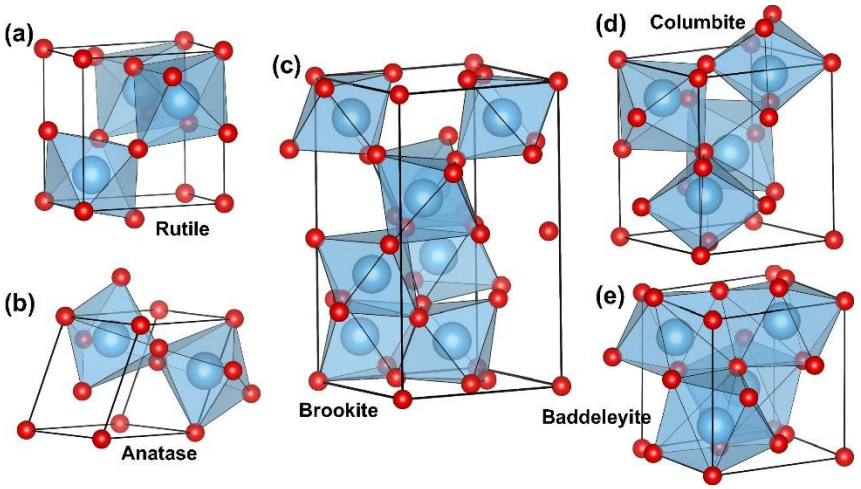


N. Artrith, T. Morawietz, and J. Behler, *Phys. Rev. B* **83**, (2011) 153101.
T Morawietz, A Singraber, C Dellago, J Behler *Proc. Natl. Acad. Sci. U. S. A.* **113**, (2016) 8368-8373.

ANN potential for bulk TiO_2

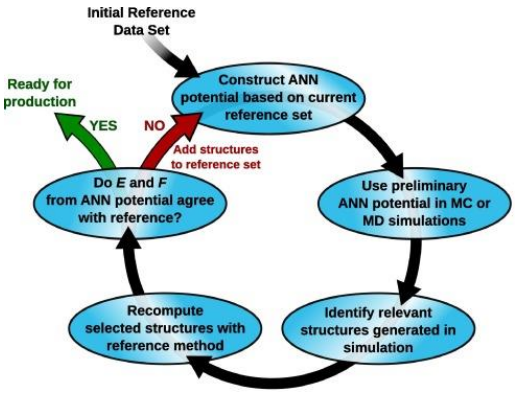
UNIVERSITÄT DUISBURG ESSEN

- An initial set of reference structures for potential construction was generated by distorting rutile, anatase and brookite structures in three different ways:
- 10 % variation of lattice constants
- Tilting the crystal cell by monoclinic strain
- Tensile strain and stretch in one dimension



- Primary ANN potentials was generated by previous dataset, then was used to generate additional reference structures by short MD simulations

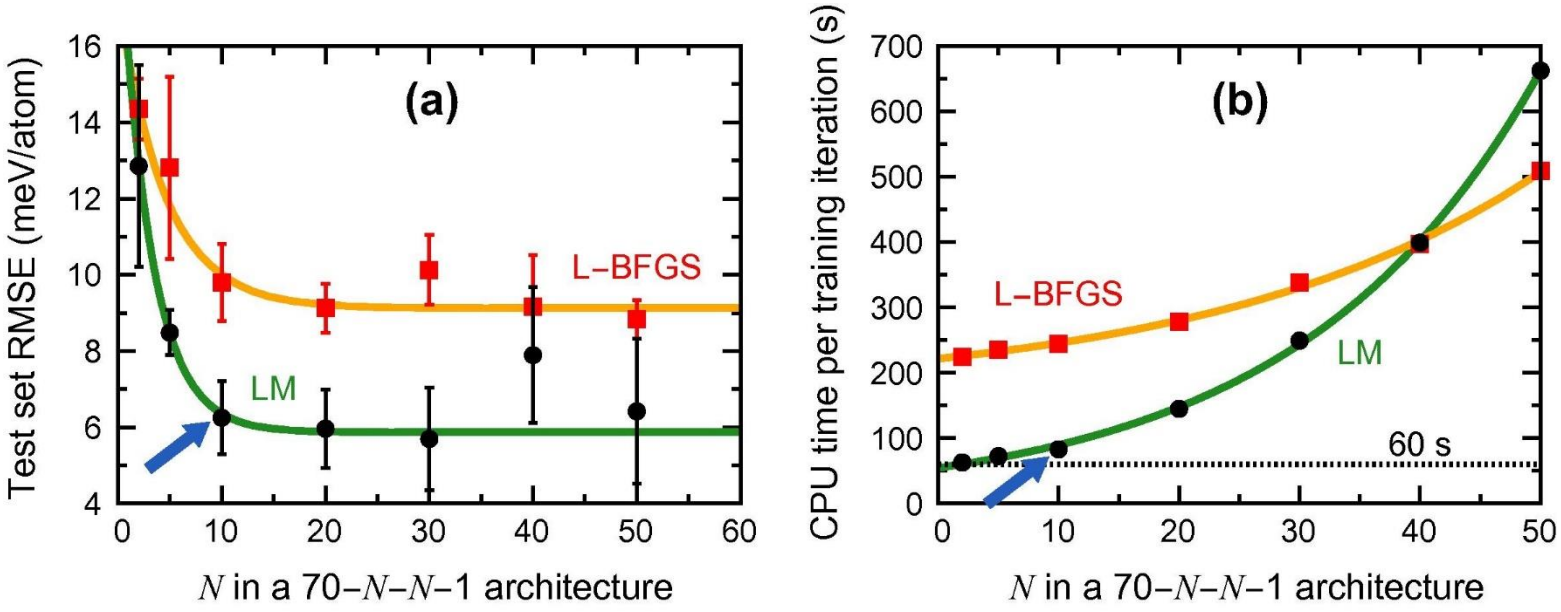
7694 structures



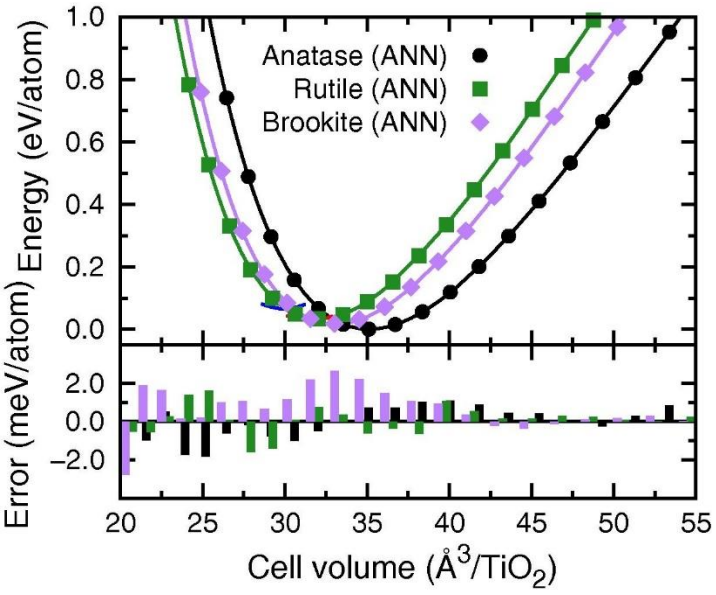
Best network architecture

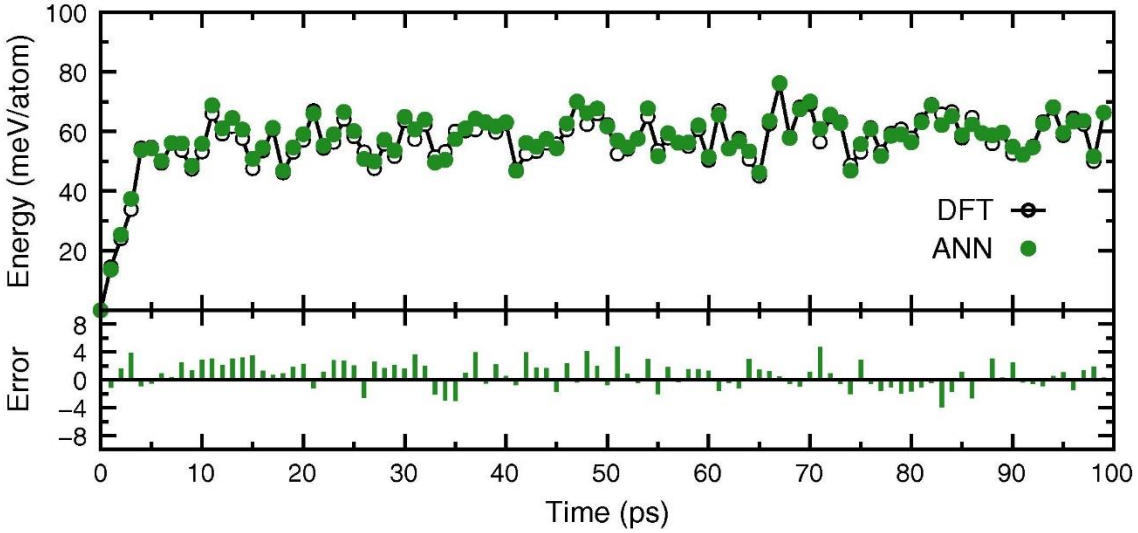
UNIVERSITÄT DUISBURG ESSEN

- Identify the balance between model complexity and transferability
- Too few model parameters: we can not fit to reference structures
- Too many model parameters: long time for training, overfitting, reduce transferability
- Simulations of 64 cores
- Optimized architecture: 70-10-10-1



ANN potential benchmark





- Energy as function of the cell volume
- MD simulation of anatase (162 atoms) at 500 K.
- Compare the corresponding energy every 1000th MD simulation with DFT
- Error less than 4 meV/atom

Catalytic activity of Au/Cu nanoparticle

UNIVERSITÄT DUISBURG ESSEN

- Gold/Copper nanoparticle is a promising candidate for CO_2 reduction reaction.
- Activity of the nanoparticles depends on the particle size and surface composition.
- Aqueous solvents can change the surface structure of nanoparticle
- Explicit water molecule on the surface is necessary to reproduce the experimental results

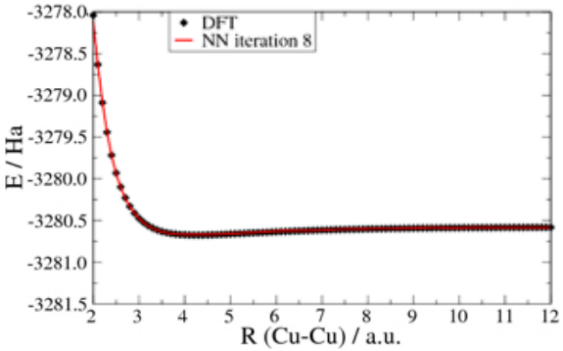
N. Artrith, A.M. Kolpak, Nano Lett., 14 (2014), pp. 2670-2676

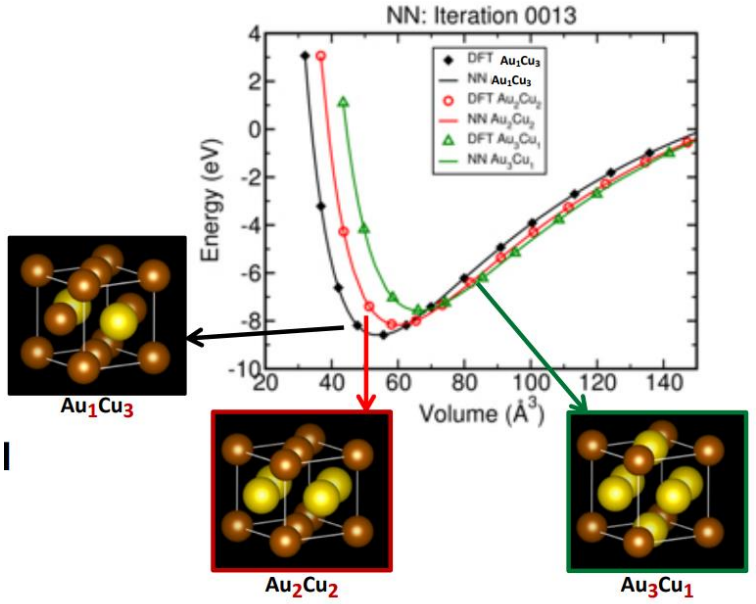
Computational details

UNIVERSITÄT DUISBURG ESSEN

- DFT reference data set comprising energies of 10000 structures (including bulk, slab, atomic cluster and molecules)
- ANN architecture: 148-5-5-1
- Bulk composition: Au_1Cu_3, Au_2Cu_2, Au_3Cu_1
- 4000 atom require 56 s on a single node

Example: Copper Dimer





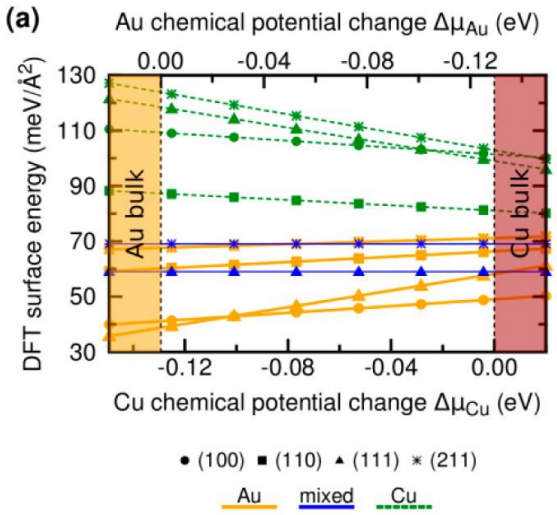
N. Artrith, A.M. Kolpak, Nano Lett., 14 (2014), pp. 2670-2676

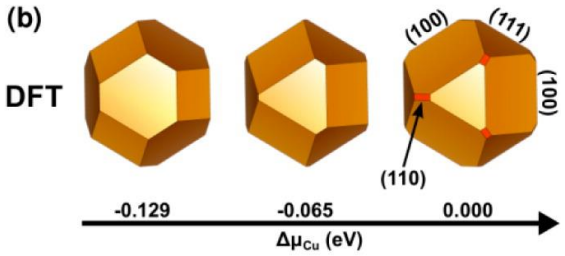
Wullf construction by DFT and ANN:

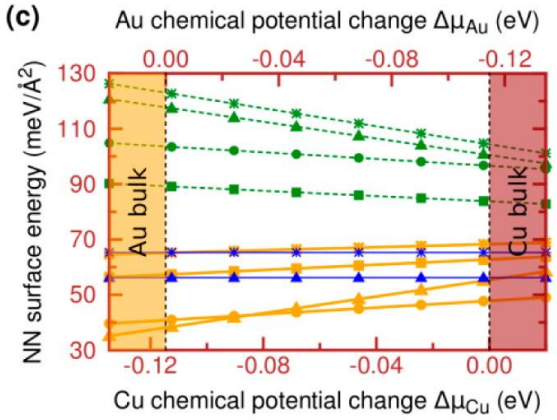
UNIVERSITÄT DUISBURG ESSEN

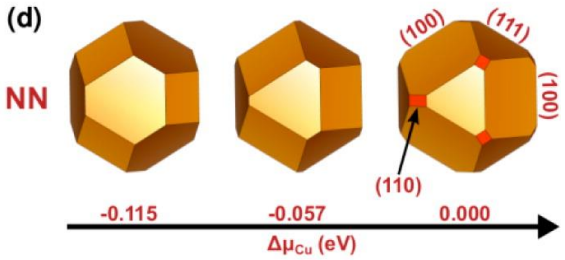
- The gold-terminated (100) and (111) surface are the most stable over entire potential energy.

- Problem: experimental particle size is around 2-4nm (300-1500 atoms)







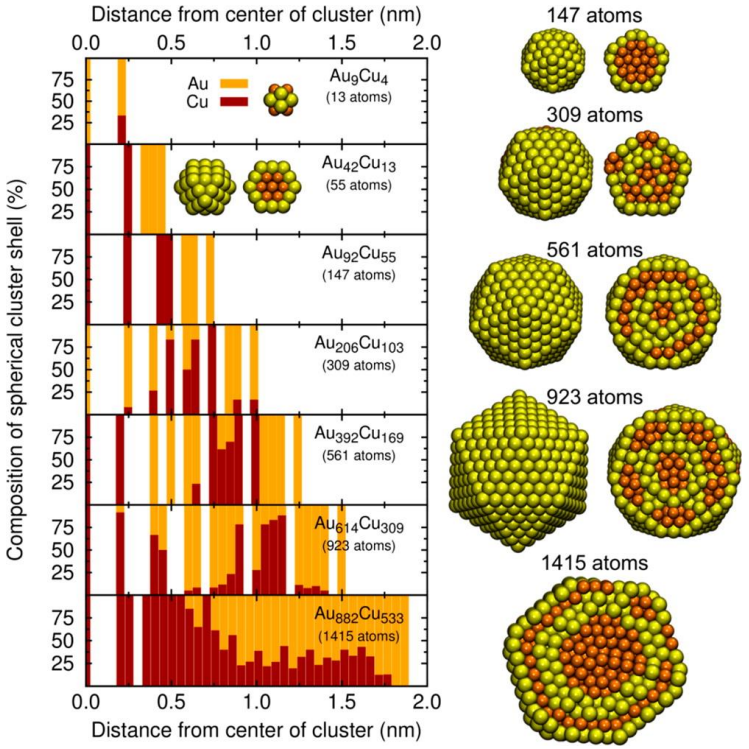


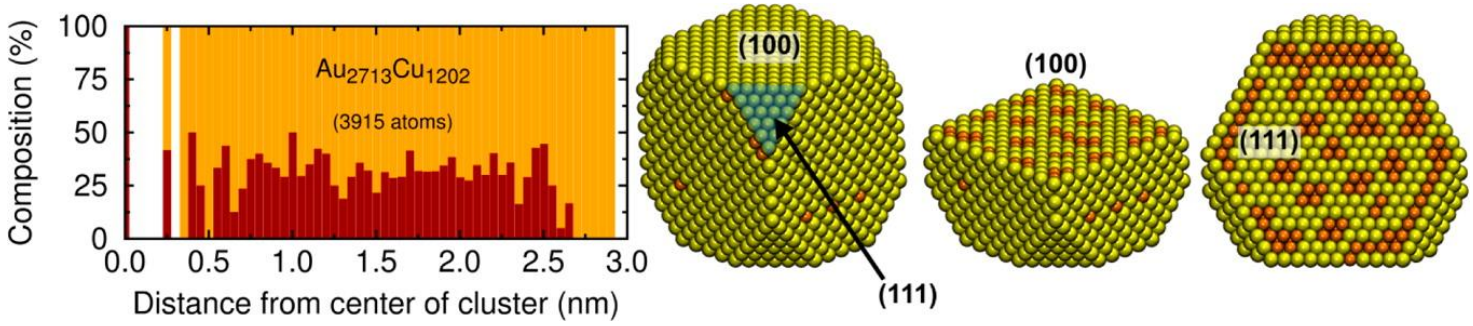
N. Artrith, A.M. Kolpak, Nano Lett., 14 (2014), pp. 2670-2676

Thermodynamic equilibrium with MC simulations

UNIVERSITÄT DUISBURG ESSEN

- Monte-Carlo simulation by using NN potential.
- Ground state of compositions at T=300 K
- Problem: experimental particle size is around 2-4nm (300-1500 atoms)
- No core-shell morphology for 6 nm particle

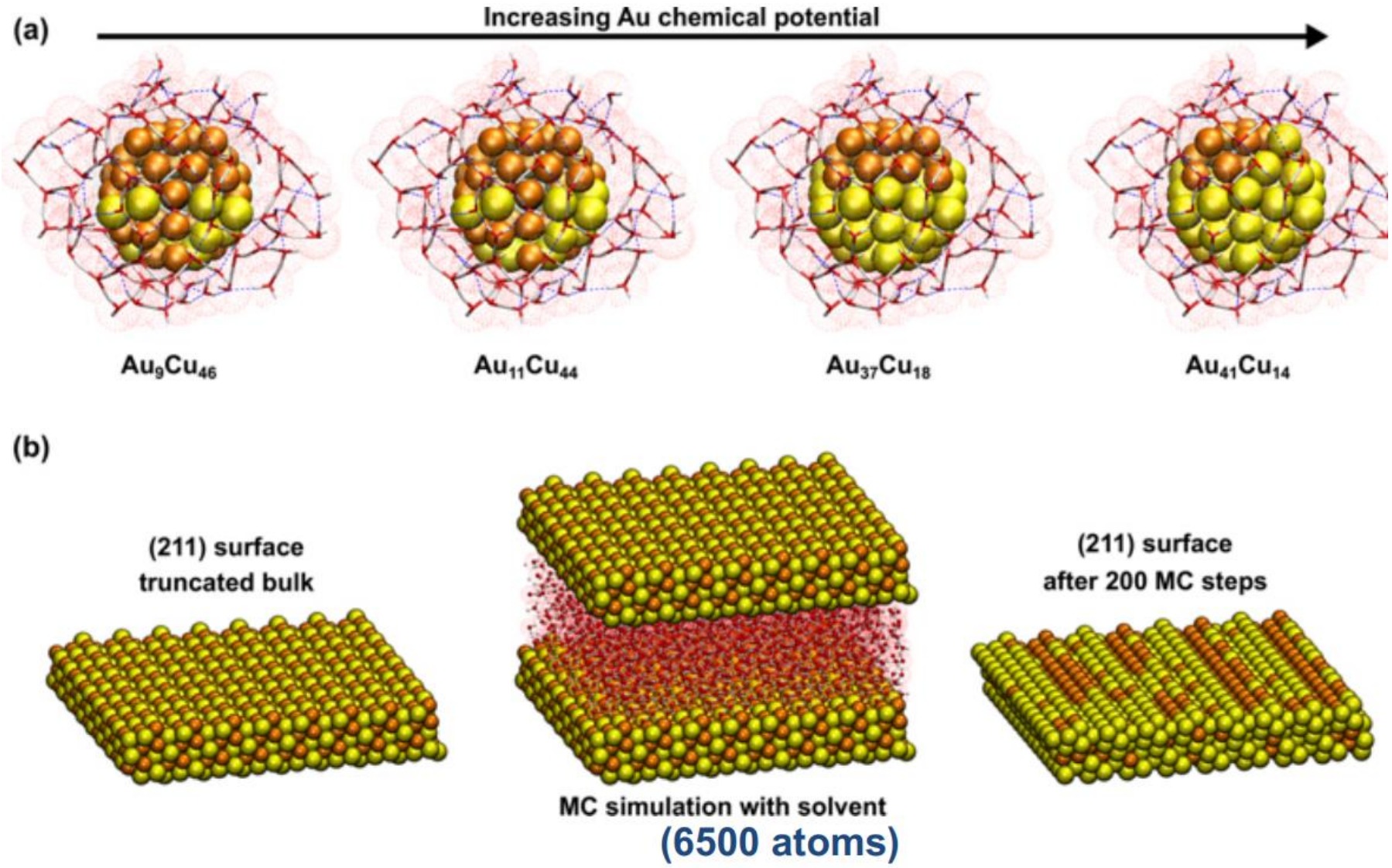




N. Artrith, A.M. Kolpak, Nano Lett., 14 (2014), pp. 2670-2676

Influence of water on the surface structure

UNIVERSITÄT DUISBURG ESSEN



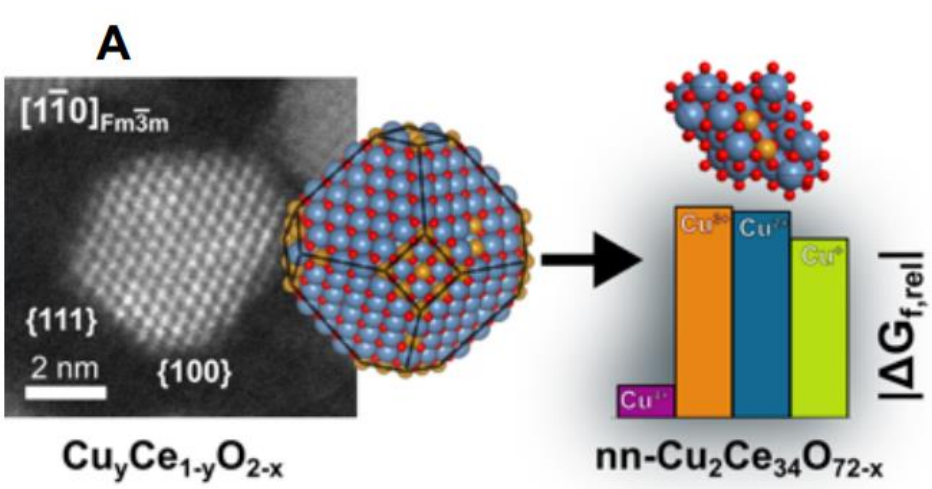
N. Artrith and A. Kolpak, *Nano Lett.*,14, (**2014**) 2670.

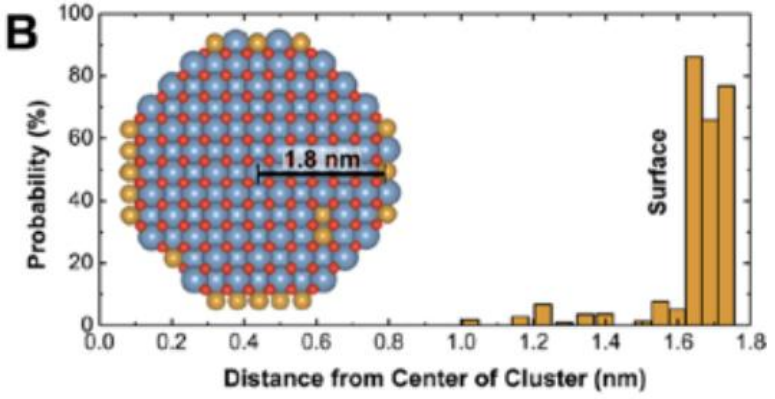
CuO/CeO_2 catalyst for CO oxidation

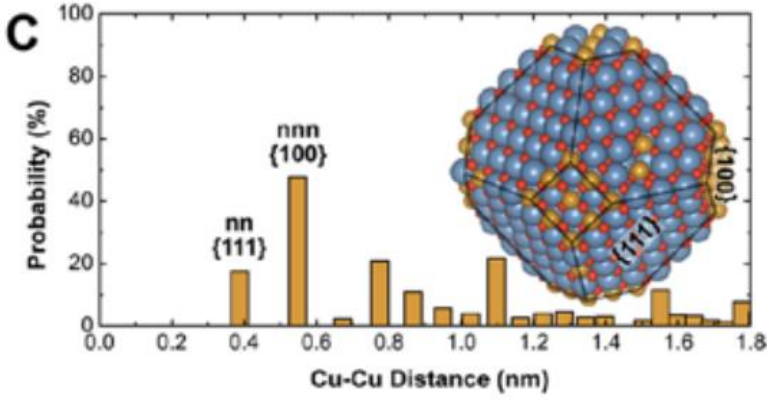
UNIVERSITÄT DUISBURG ESSEN

Cu Distribution in the particles

- MC simulations of a **3.5 nm (~1,300 atoms, $Cu_{54}Ce_{405}O_{834}$)**: Cu is most stable near surface
- Cu adsorption on (100) surface and on edges favorable





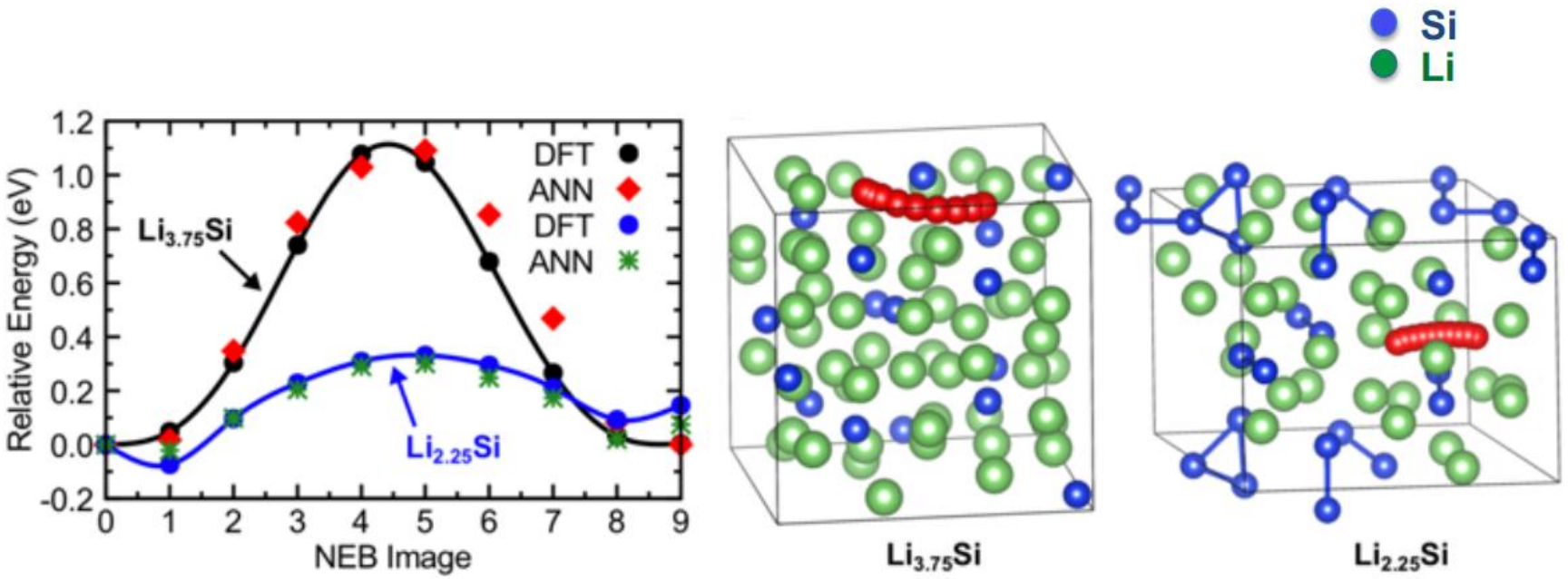


Cu-Cu pair distribution

- Cu clustering: Cu-Cu pair distribution
- Combined probability that either the nearest neighbor (nn) or next-nearest neighbor (nnn) site of a Cu defect is also a Cu defect is around 70%

J.S. Elias, **N. Artrith**, M. Bugnet, L. Giordano, G. A. Botton, A.M. Kolpak, and Y. Shao-Horn*
ACS Catalysis **6**, (2016), 1675-1679

The LiSi ANN Potential is Accurate for Diffusion



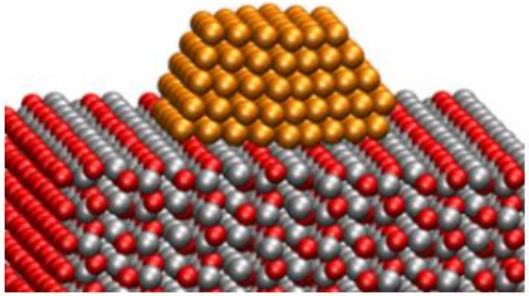
- Structures not in training set
- **Tested many different diffusion pathways in different alloy compositions**

Cu nanoparticle on ZnO

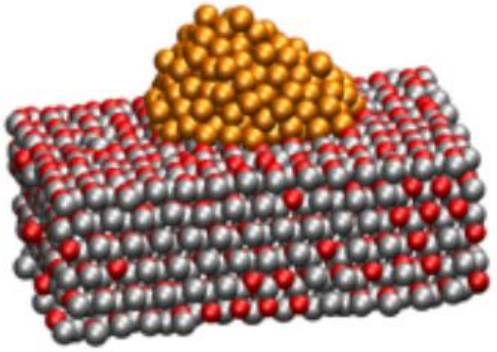
UNIVERSITÄT DUISBURG ESSEN

ANN-MD Simulation: Slab model ~8,000 atoms: NVT, MD at 1000 K

Training and testing sets for the ANN potential:
Cu/Zn/O structures: (e.g. ideal, vacancies, defects)
~100,000 structures (90% train, 10% test)
RMSEs E_{total} : **0.005** eV/atom
Forces: **0.090** eV/Bohr

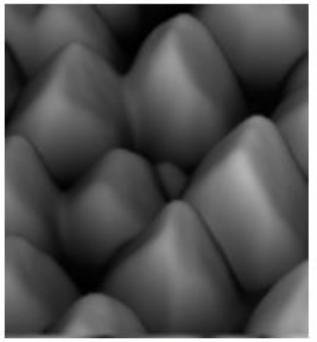


Initial configuration/ MD movie



Configuration at 300 ps

N. Artrith, B. Hiller, J. Behler
Phys. Stat. Sol. B 250 (**2013**) 1191 (invited feature article).



STM image of Cu@ZnO(1010), T= 290 K

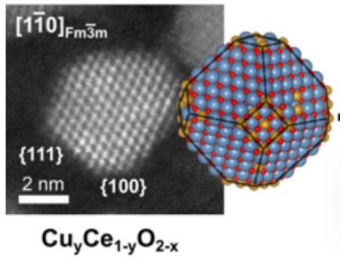
U. Köhler, et. al, *Phys. Status Solidi B* 250 (**2013**) 1122.

- **ANN potentials allow to simulate structural models with thousands of atoms while providing high accuracy close to the reference method**

Outlook:

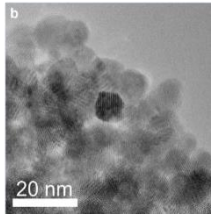
- Identification of potential energy surface by ANN for bulk Co_3O_4, $CoFe_2O_4$
- MD and MC simulations

- Influence of inversion parameter, cation ordering in bulk and surface

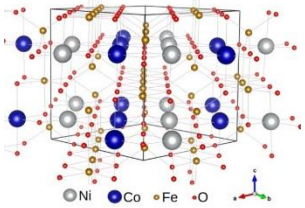


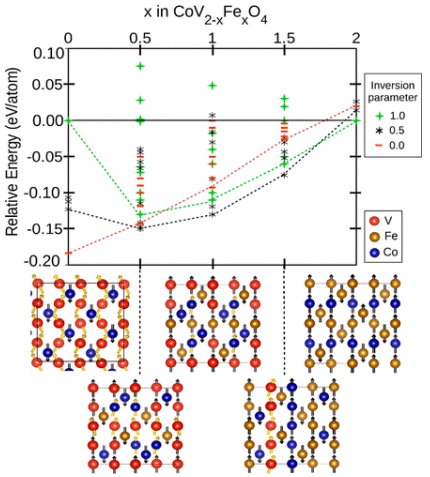
$Cu_yCe_{1-y}O_{2-x}$

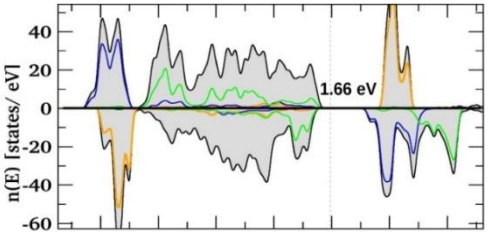
$CoFe_2O_4$



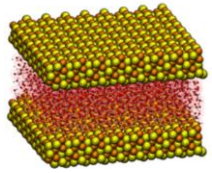
$CoFe_2O_4$

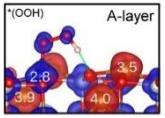


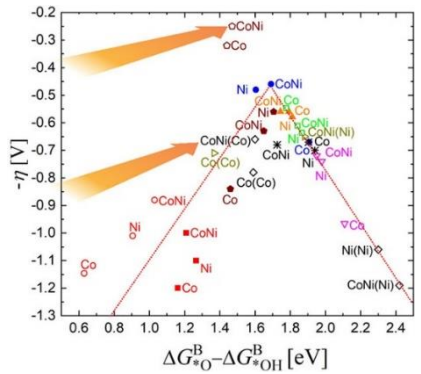




- Including explicit water
- Including temperature
- Surface reconstruction during OER
- Training DFT dataset is already exist



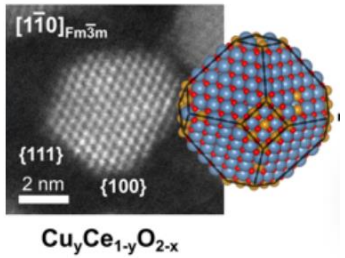




Outlook:

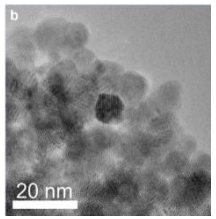
- Identification of potential energy surface by ANN for bulk Co_3O_4, $CoFe_2O_4$
- MD and MC simulations

- Influence of inversion parameter, cation ordering in bulk and surface

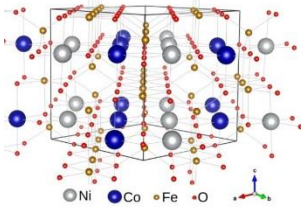


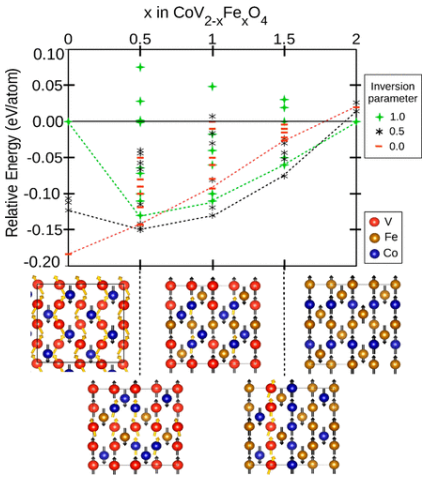
$Cu_yCe_{1-y}O_{2-x}$

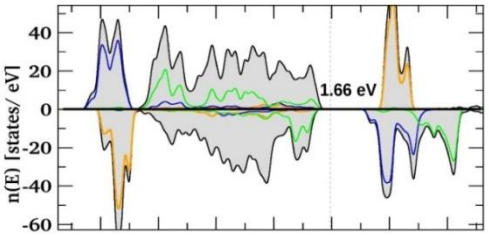
$CoFe_2O_4$



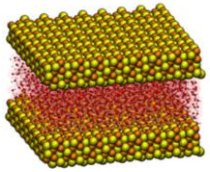
$CoFe_2O_4$

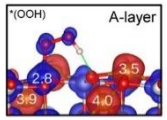


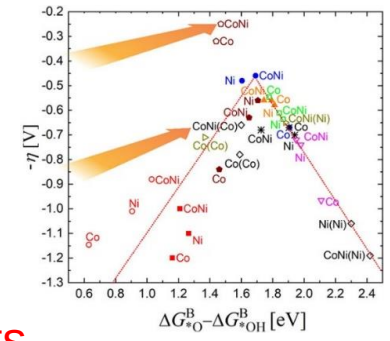




- Including explicit water
- Including temperature
- Surface reconstruction during OER
- Training DFT dataset is already exist







Non of that will be achieved with collaboration with experts

Thank you for your attention