

Extending the Length and Time Scales of Atomistic Simulations Using Neural Network Potentials

Jörg Behler

Lehrstuhl für Theoretische Chemie
Ruhr-Universität Bochum
D-44780 Bochum, Germany

joerg.behler@theochem.ruhr-uni-bochum.de

- **electronic structure methods are too expensive for many problems**
- **force fields are often not reactive \Rightarrow no "chemistry"**
- **accuracy of many empirical potentials/force fields is not satisfactory**
- **often system-specific terms, no general functional form
 \Rightarrow different types of potentials for molecules, metals, semiconductors ...**
- **systematic improvement of conventional potentials very difficult**
 - **limitations imposed by functional form**
 - **range of validity cannot be extended systematically**
(applicability often restricted to small part of configuration space)
- **construction of atomistic potentials is often very cumbersome and time demanding**

Physical Approach: "Conventional" Potentials

Typically fixed functional form \Rightarrow intrinsically **limited accuracy**

Example: Lennard-Jones potential
(classical force fields)

$$E_{ij} = 4\epsilon \left[\left(\frac{\sigma}{R_{ij}} \right)^{12} - \left(\frac{\sigma}{R_{ij}} \right)^6 \right]$$

Mathematical Approach ("Machine Learning"):

Use of extremely flexible functions adapting to high-level reference data.

Example: Artificial Neural Networks

Formal Proof: Neural Networks Enable an Exact Representation

*“Neural networks with sufficiently smooth activation functions are capable of **arbitrarily accurate** approximation to a function and its derivatives.”*

K. Hornik, *Neural Networks* 4 (1991) 251.

1 Hamiltonian depends on

- nuclear positions R
- nuclear charges Z (chemical elements)
- system charge Q (usually neutral)

⇒ Structure contains **all information required** to fully define the energy

2 ⇒ An **unknown function** $E(R,Z,Q)$ exists, but we know only a set of discrete function values

⇒ We need to find the best possible continuous analytic representation

3 **Neural Networks enable in principle an exact representation**

Proof e.g.: K. Hornik, *Neural Networks* 4 (1991) 251.

⇒ Possible to combine the best of both worlds:

⇒ correct chemistry: **ab initio** calculations

⇒ numerical accuracy and efficiency: Neural Network

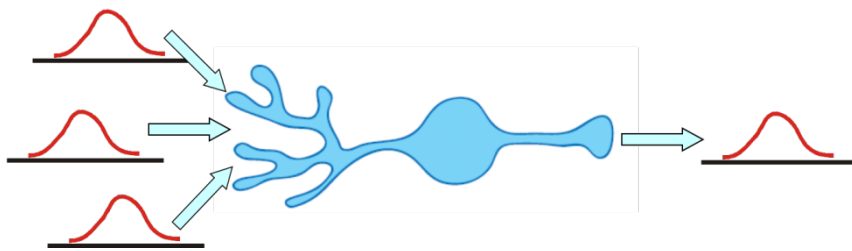
4 Efficient molecular dynamics simulations, Monte Carlo etc.

Neural Networks (NN)

“Artificial neural networks are massively parallel interconnected networks of simple (usually adaptive) elements and their hierarchical organizations, which are intended to interact with the objects of the real world in the same way as **biological nervous systems** do.”

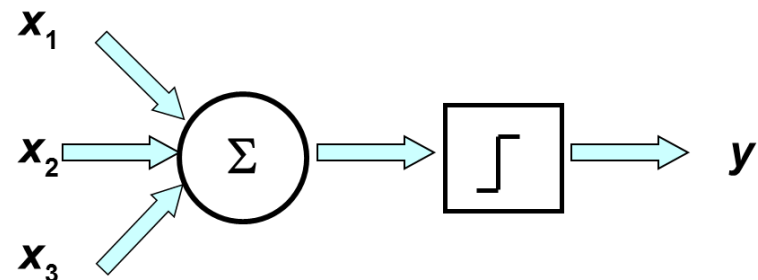
T. Kohonen, *Neural Networks* 1, 3 (1988).

“Bio-Neurons”



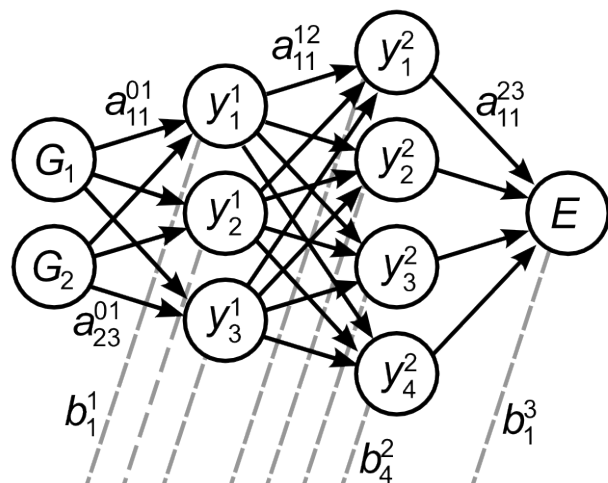
W. McCulloch and W. Pitts, *Bull. Math. Biophys.* 5 (1943) 115.

Artificial Neuron



⇒ artificial NNs approximate the signal processing in the nervous system

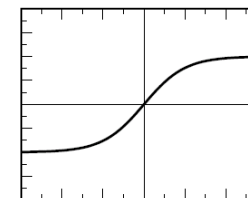
Input Layer Hidden Layer 1 Hidden Layer 2 Output Layer



a_{ij}^{kl} ← layer superscripts
 ← connecting weight
 ← node subscripts
 b_i^j bias weight

optimized using high-level training data

activation functions:



Bias Node

analytic expression for this example NN:

$$E = f_1^3 \left(b_1^3 + \sum_{k=1}^4 a_{k1}^{23} \cdot f_k^2 \left(b_k^2 + \sum_{j=1}^3 a_{jk}^{12} \cdot f_j^1 \left(b_j^1 + \sum_{i=1}^2 a_{ij}^{01} \cdot G_i \right) \right) \right)$$

T.B. Blank, S.D. Brown, A.W. Calhoun, and D.J. Doren, *J. Chem. Phys.* **103** (1995) 4129.

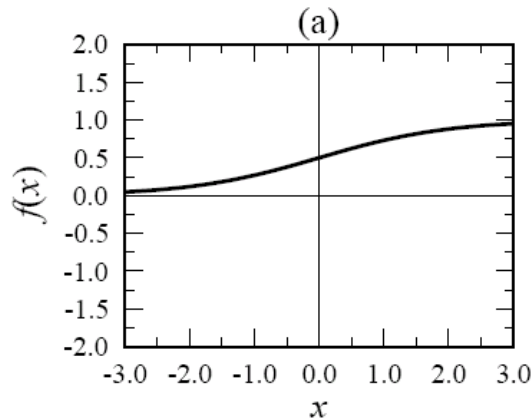
S. Lorenz, A. Groß, and M. Scheffler, *Chem. Phys. Lett.* **395** (2004) 210.

J. Behler, S. Lorenz, and K. Reuter, *J. Chem. Phys.* **127** (2007) 014705.

Activation functions enable the representation of **general nonlinear functions**.

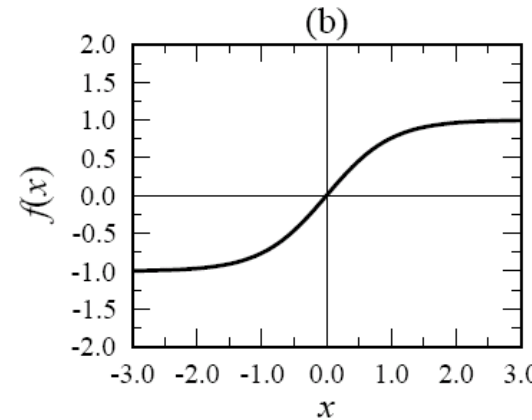
**sigmoid
function**

$$f(x) = \frac{1}{1 + e^{-x}}$$



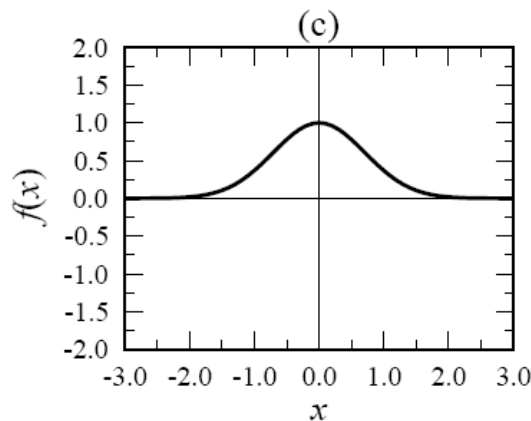
**hyperbolic
tangent**

$$f(x) = \tanh(x)$$



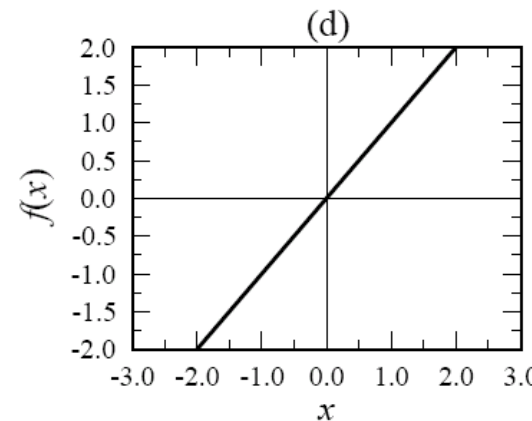
**Gaussian
function**

$$f(x) = e^{-\eta x^2}$$



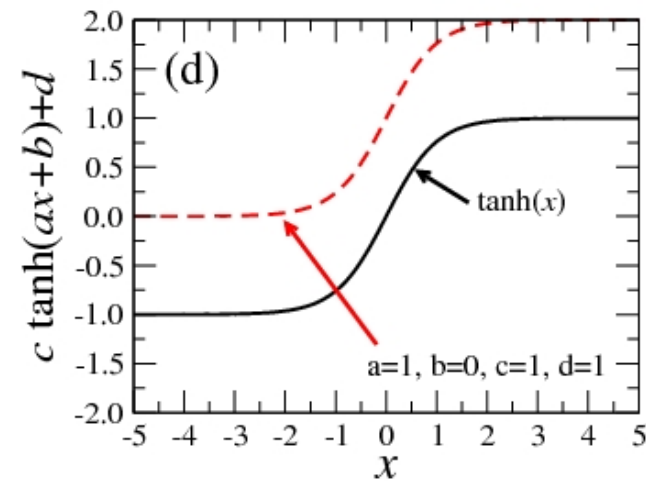
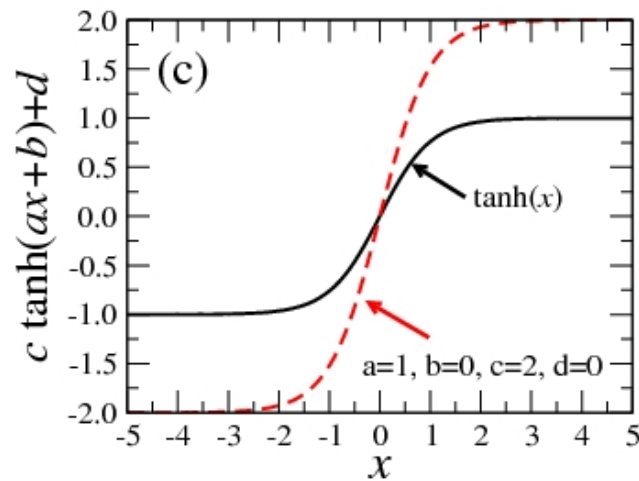
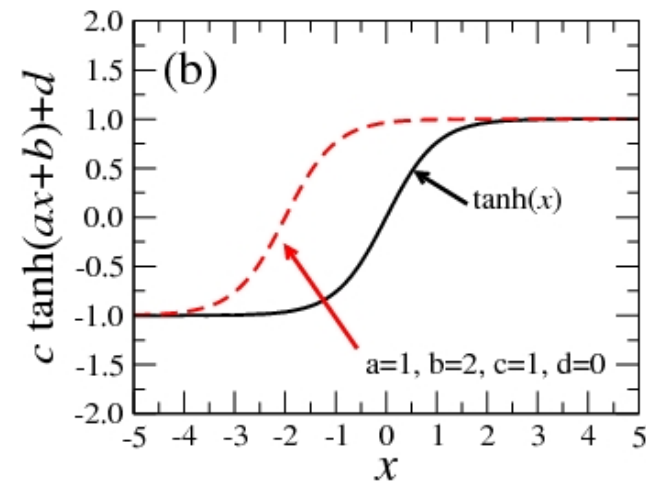
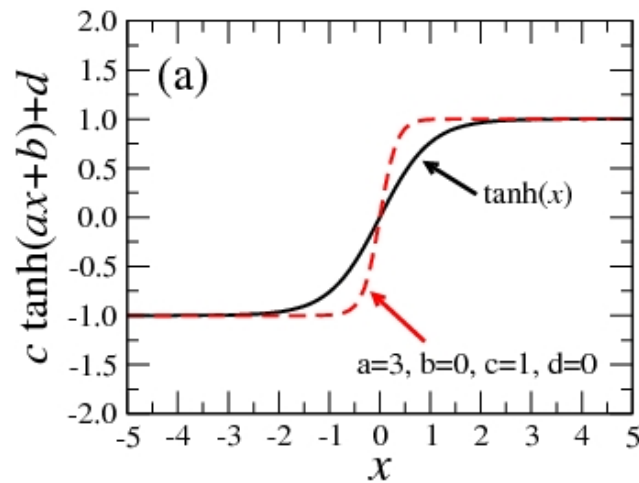
**linear
function**

$$f(x) = x$$



- converge for very small and very large arguments
- have a nonlinear shape for intermediate values

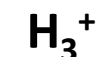
Basic functional element of the NN: $f(x) = c \cdot \tanh(ax + b) + d$



⇒ very simple but flexible

Neural Network Potentials for Small Molecules:

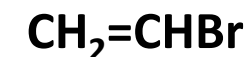
F.V. Prudente, and J.J.S. Neto, *Chem. Phys. Lett.* **287**, 585 (1998).



H. Gassner et al., *J. Phys. Chem. A* **102**, 4596 (1998).



L.M. Raff et al., *J. Chem. Phys.* **122**, 084104 (2005).



S. Manzhos, and T. Carrington, Jr., *J. Chem. Phys.* **125**, 084109 (2006).



Molecule – (Frozen) Surface Interactions:

T.B. Blank et al., *J. Chem. Phys.* **103**, 4129 (1995).



S. Lorenz, A. Groß, and M. Scheffler, *Chem. Phys. Lett.* **395**, 210 (2004).



J. Behler, S. Lorenz, and K. Reuter, *J. Chem. Phys.* **127**, 014705 (2007).



J. Ludwig, and D.G. Vlachos, *J. Chem. Phys.* **127**, 154716 (2007).



D.A.R.S. Latino et al., *J. Electroanal. Chem.* **624**, 109 (2008).



Reviews:

C.M. Handley, and P.L.A. Popelier, *J. Phys. Chem. A* **114**, 3371 (2011).

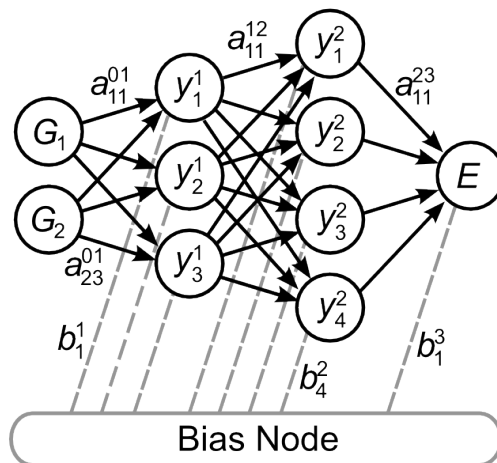
J. Behler, *Phys. Chem. Chem. Phys.* **13**, 17930 (2011).

High-Dimensional Neural Network Potentials

Is it possible to use Neural Network potentials to construct high-dimensional potential energy surfaces?

Conceptual problems for a direct application of Neural Networks to large systems:

Input Layer Hidden Layer 1 Hidden Layer 2 Output Layer



- limited number of dimensions (up to ≈ 12)
- permutation symmetry of the system is not included (exchange of atoms changes the energy)
- energy generally depends on rotation and translation
- potential is valid only for a given system size

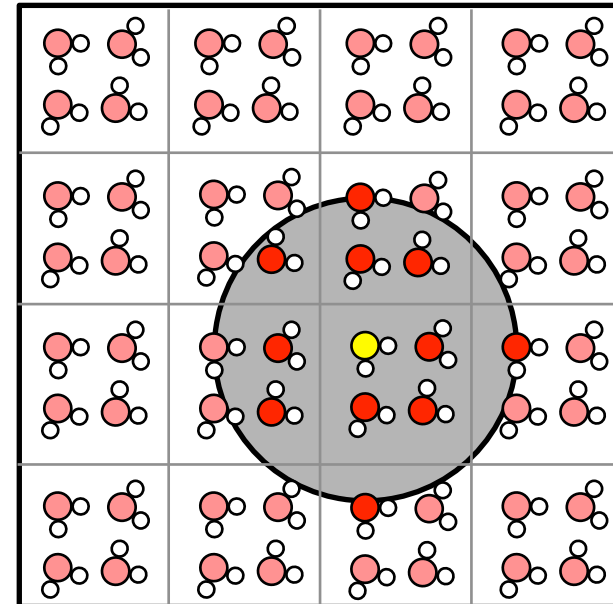
⇒ A new Neural Network scheme is required to deal with high-dimensional systems

Total energy represented as a sum of atomic energies E_i :

$$E = \sum_i E_i$$

- separate NN for each atom
- E_i depend on chemical environment
- environments defined by sufficiently large cutoff $R_c \approx 6 \text{ \AA} - 10 \text{ \AA}$
- chemical environment is described by many-body „symmetry functions“

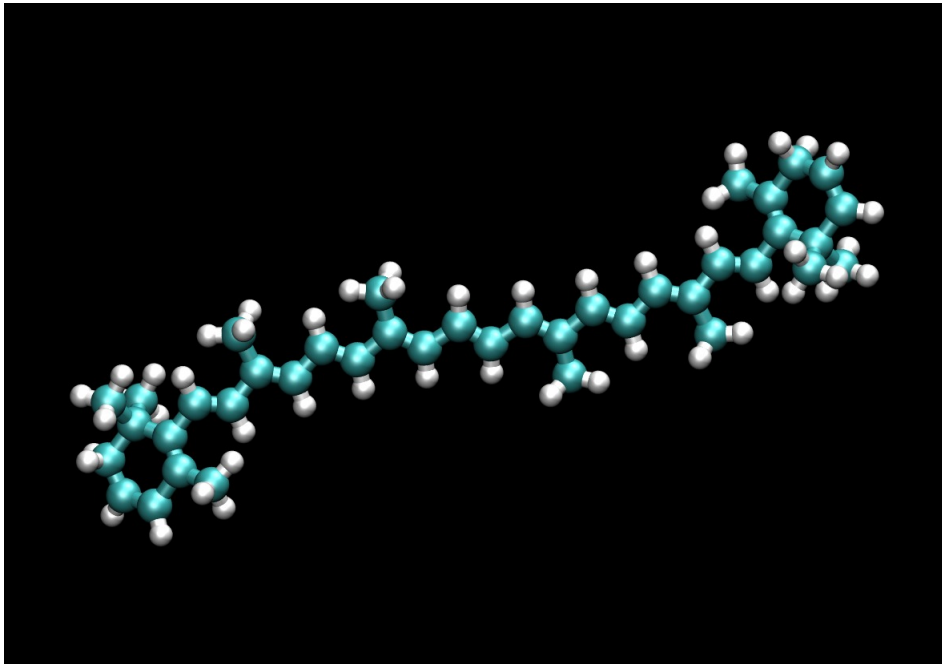
⇒ all requirements formally fulfilled



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

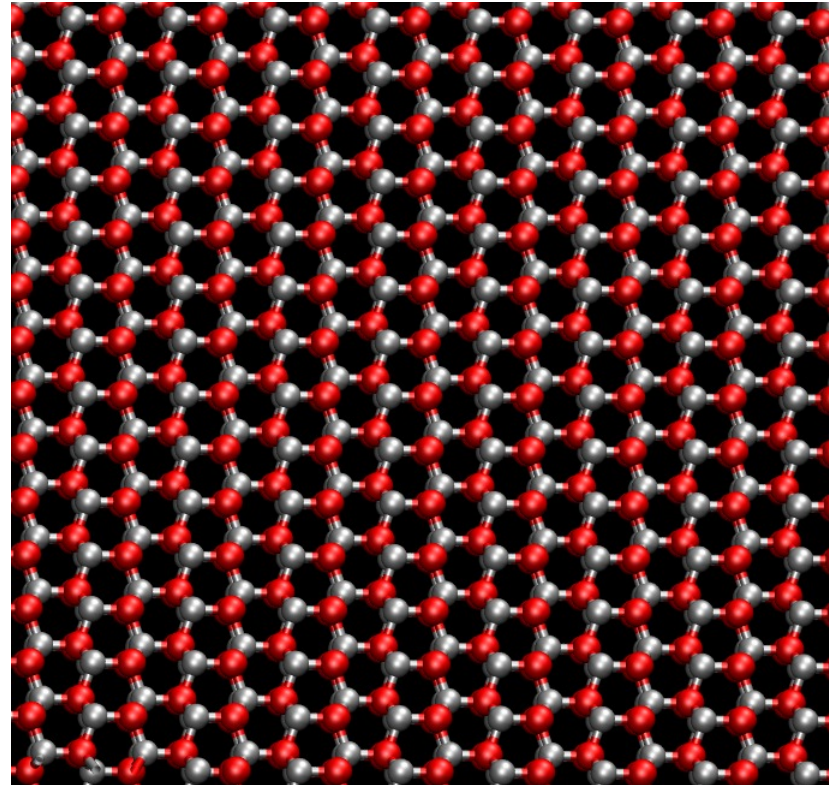
J. Behler, *J. Chem. Phys.* **134** (2011) 074106.

Molecular Case:



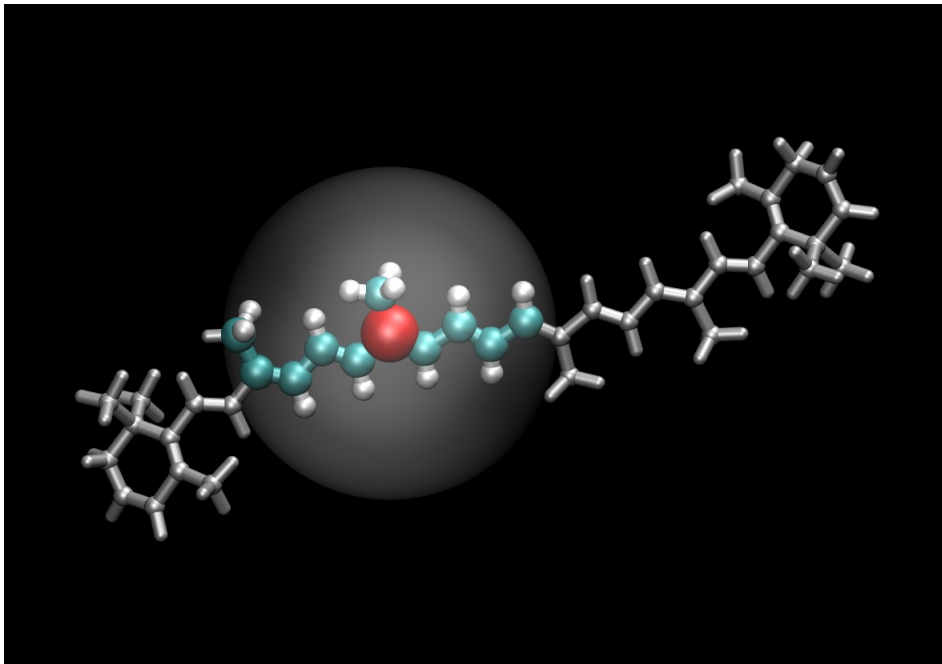
β -carotene

Periodic Case:



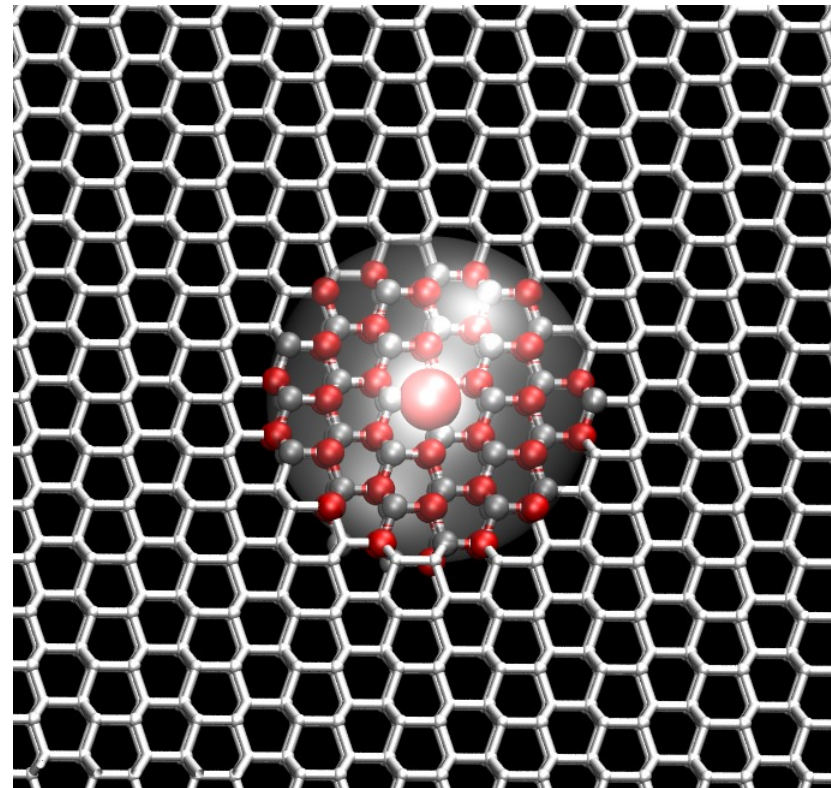
zinc oxide (wurtzite structure)

Molecular Case:



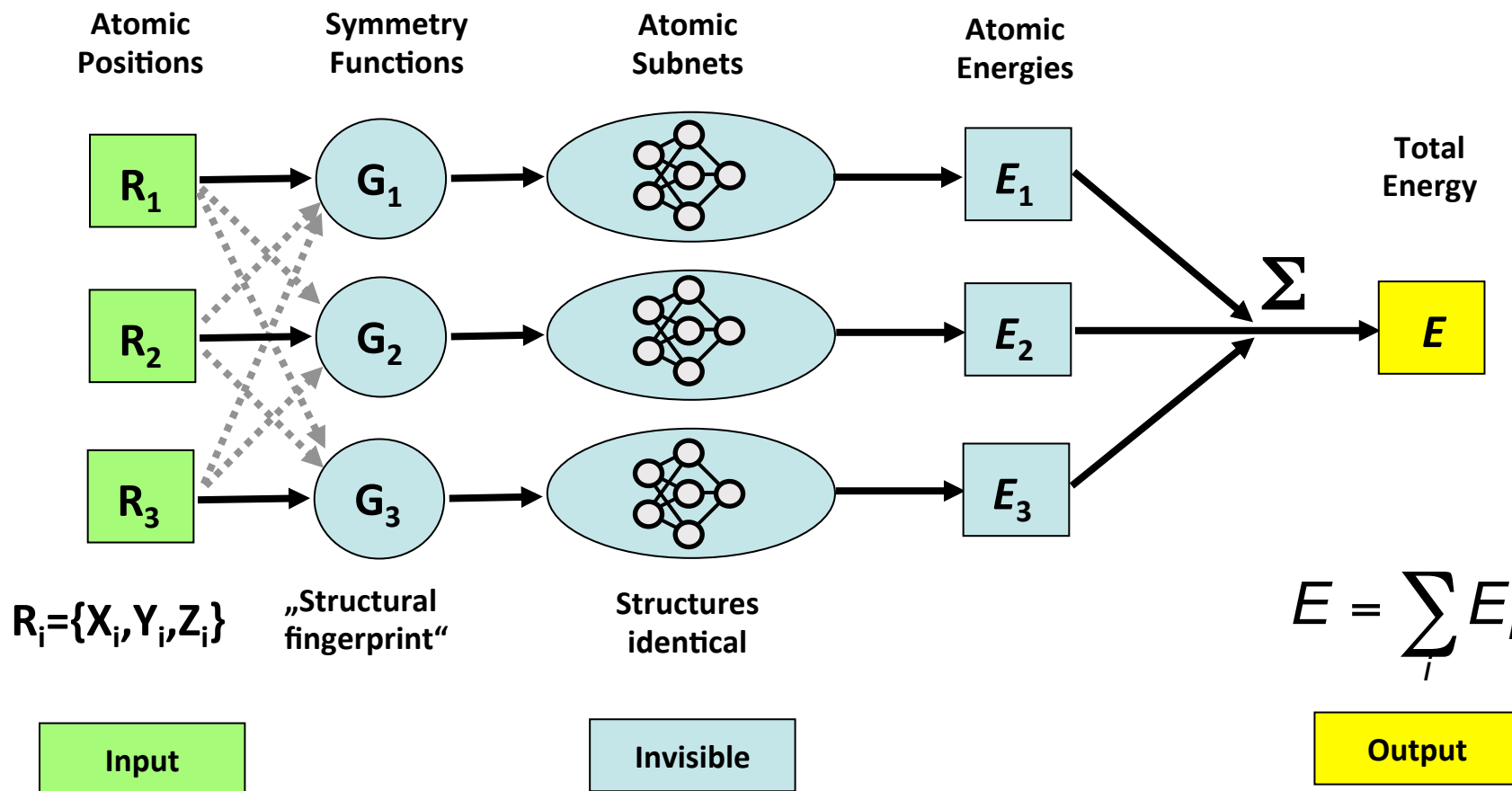
β -carotene

Periodic Case:



zinc oxide (wurtzite structure)

Example: 3-Atom System



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

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

Problem: Interactions are not necessarily short-ranged

Example: Multicomponent-systems: oxides, water

- significant charge-transfer
- long-ranged electrostatic interactions must be included
- modification of the Neural Network scheme is needed

Generalized total energy expression:

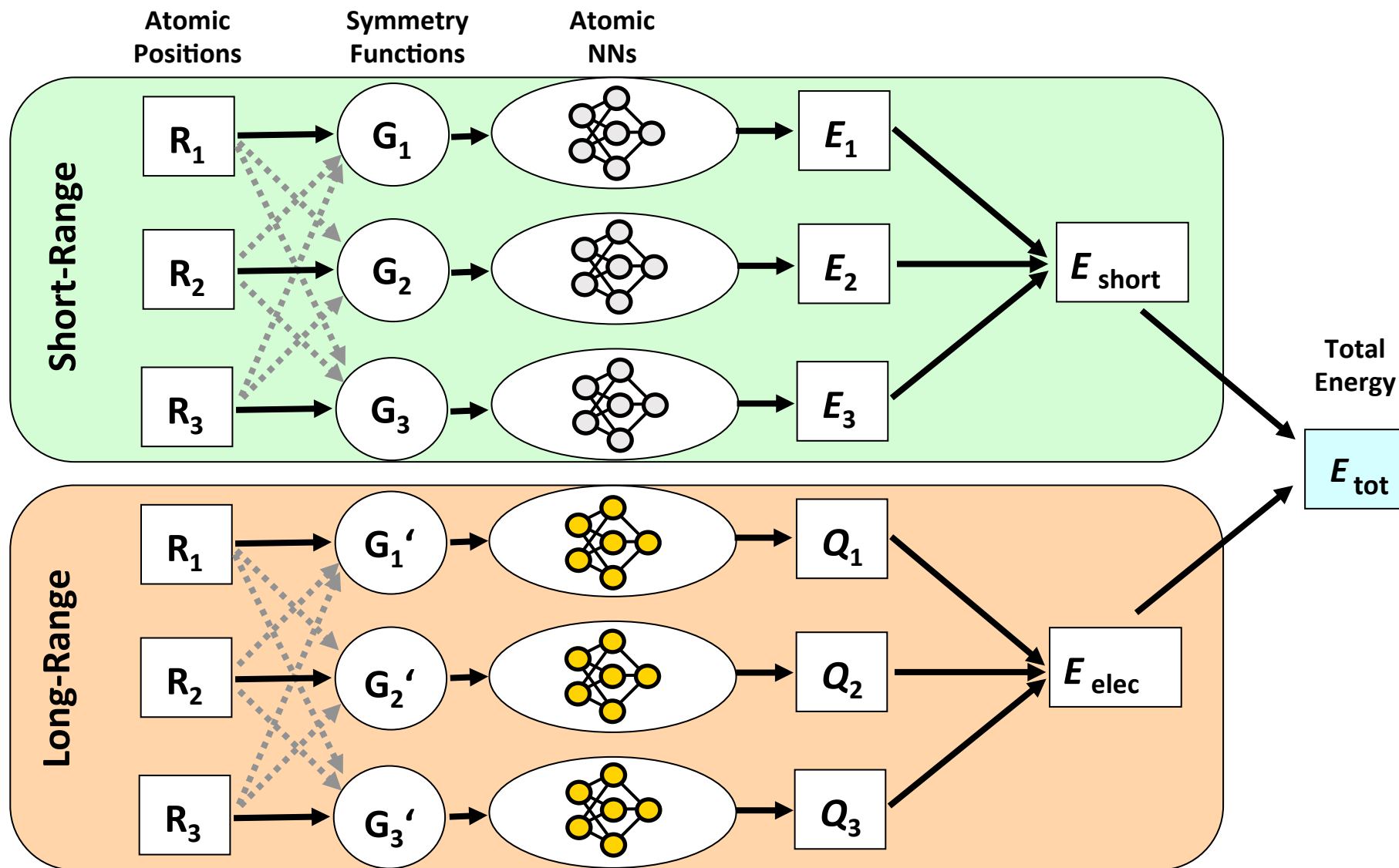
$$E_{tot} = E_{short} + E_{elec}$$

“Short-range” part:
atomic energy contributions
depending on local environment

Long-range part:
atomic charges
depending on local environment

- electrostatic interactions are not truncated
 - standard methods (Ewald sum) can be used
- ⇒ Neural Network potentials applicable to multicomponent systems

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



Symmetry Functions

Starting Point: Atomic Structure (Cartesian Coordinates)

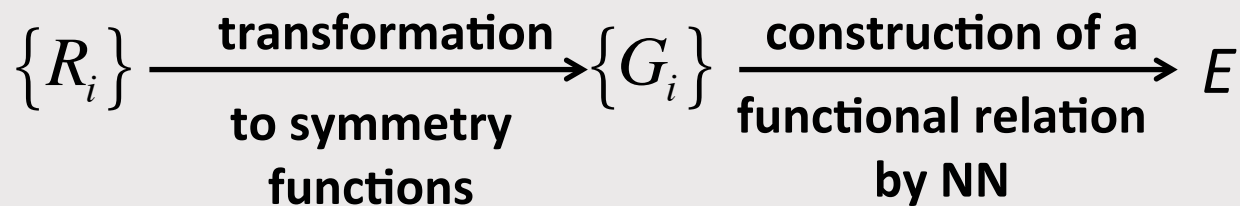
Goal: Vector of coordinates $\{G_i\}$ (symmetry functions) with the properties:

- structural description
- translational and rotational invariance
- atomic permutation symmetry included
- number independent of coordination
- zero value and derivative at cutoff radius
- continuous in value and slope

S. Lorenz, A. Groß, and M. Scheffler, *Chem. Phys. Lett.* **395**, 210 (2004).

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

Two-Step Approach



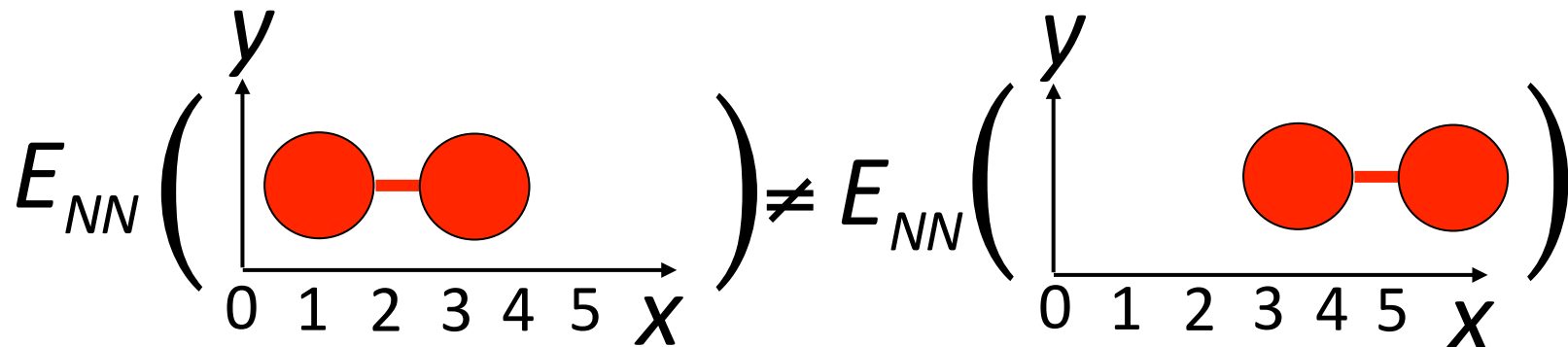
Problem 1: Cartesian Coordinates

- Neural Networks just process numbers
- absolute values of Cartesian coordinate have no meaning
- only relative atomic positions are important

⇒ Cartesian coordinates cannot be used

Example:

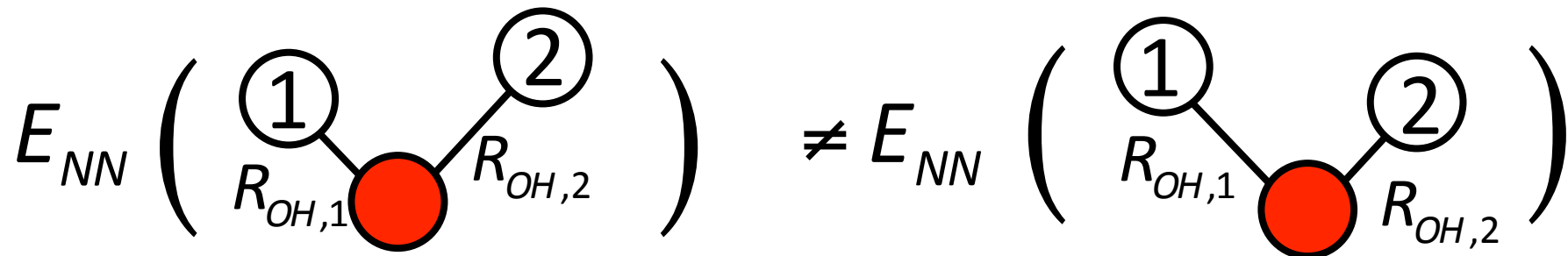
Translation:



Solution: Use internal coordinates: distances, angles, ...

Problem 2: Permutation Symmetry

Example: Water monomer at finite temperature



⇒ Both molecules have a different set of coordinate values

Solution: Symmetrization step

⇒ invariant with respect to order of H atoms

$$G_1 = (R_{OH,1} + R_{OH,2})^2 \quad G_3 = R_{HH}$$

$$G_2 = (R_{OH,1} - R_{OH,2})^2$$

⇒ Coordinate vector invariant with respect to H-atom exchange

Problem 3: Internal Coordinates

- symmetrization is limited to very small molecules (4-5 atoms)
- internal coordinates are not unique
- number of coordinates depends on system size

⇒ **not applicable to large systems**

Solution for High-Dimensional Neural Network Potentials:

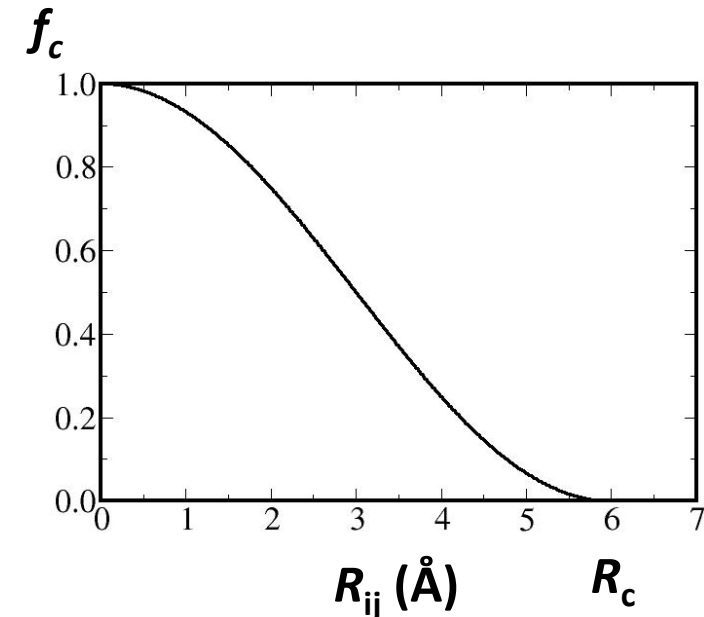
Local Motifs and Many-Body Symmetry Functions

- describe only close (chemically interacting) atoms
⇒ cutoff
- use many-body functions depending on all neighbors
⇒ independent of number of neighbors
- translationally and rotationally invariant like internal coordinates
- continuous in value and slope
⇒ forces

Cutoff Function

- decays to zero in value and slope at R_c
- reflects decreasing chemical interaction
- central component of all symmetry functions
- R_c is increased until potential converges

$$f_c(R_{ij}) = \begin{cases} \frac{1}{2} \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1 \right] & \text{for } R_{ij} < R_c \\ 0 & \text{for } R_{ij} > R_c \end{cases}$$



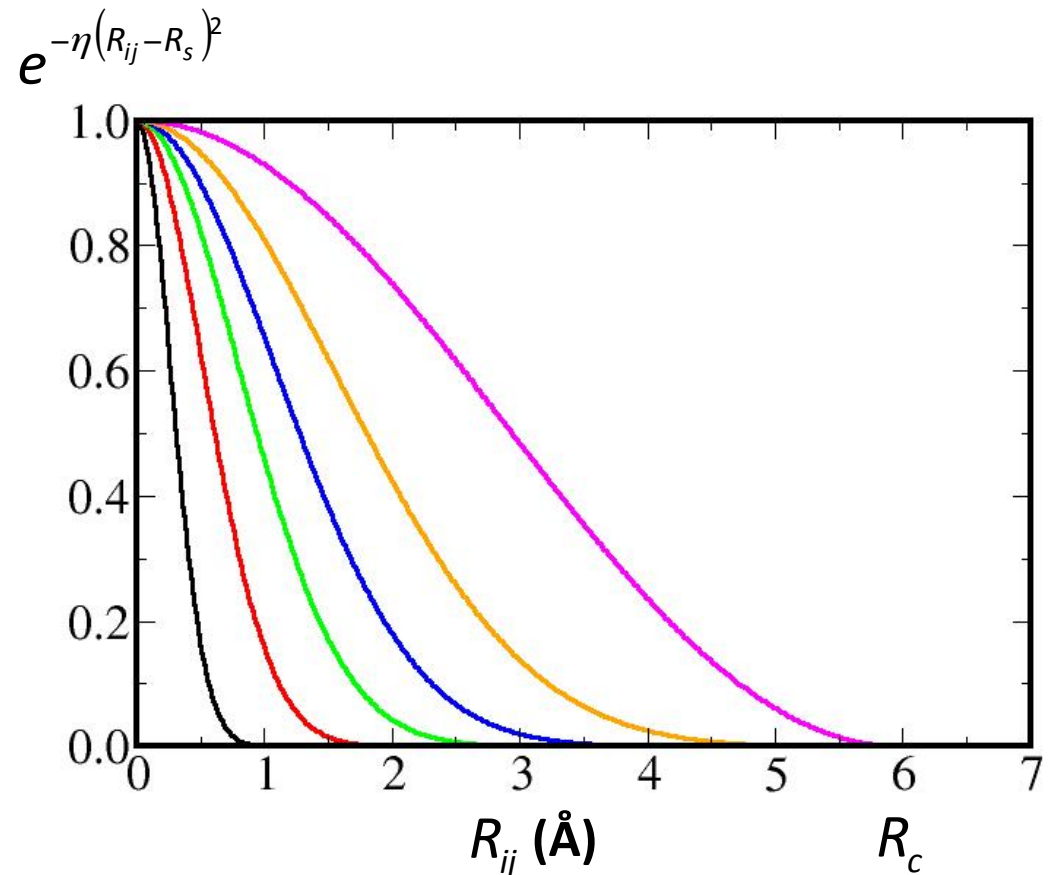
Typical cutoff radius: $R_c = 6 - 8 \text{ \AA}$

Convergence parameter!

- decay with increasing distance \Rightarrow Gaussians
- summation over all neighbors
- many-body term, interpretation as coordination number
- one-to-one correspondence between function value and R_{ij}

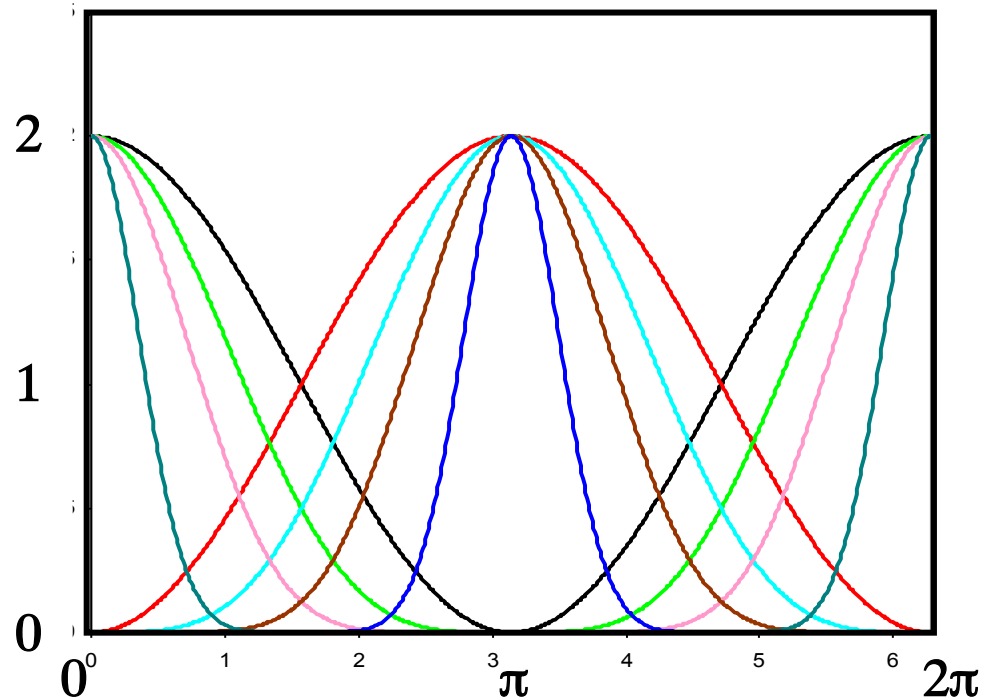
$$G_i^{rad} = \sum_j e^{-\eta(R_{ij}-R_s)^2} f_c(R_{ij})$$

Set of radial functions:
„Radial Fingerprint“



Purpose: Description of angular distribution of neighbors

$$G_i^\mu = 2^{1-\xi} \sum_{j,k \neq i} (1 + \lambda \cos \theta_{ijk})^\xi \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk})$$



⇒ Several angular functions are used

- all symmetry functions are many-body terms
- typical number: about 10 radial and 40 angular functions

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

Properties of Neural Network Potentials

- **Large training sets are needed**

Reason: NN does not know about physics

Solution: Use more information (forces)

Problems of forces: - Noise
- Still coverage of configuration space needed

BUT: - Pays off quickly
- Provides access to systems beyond reach for DFT

- **Limited to the accuracy of the reference method**

Problem: required number of points accessible only with GGA DFT

- **Poor extrapolation capabilities**

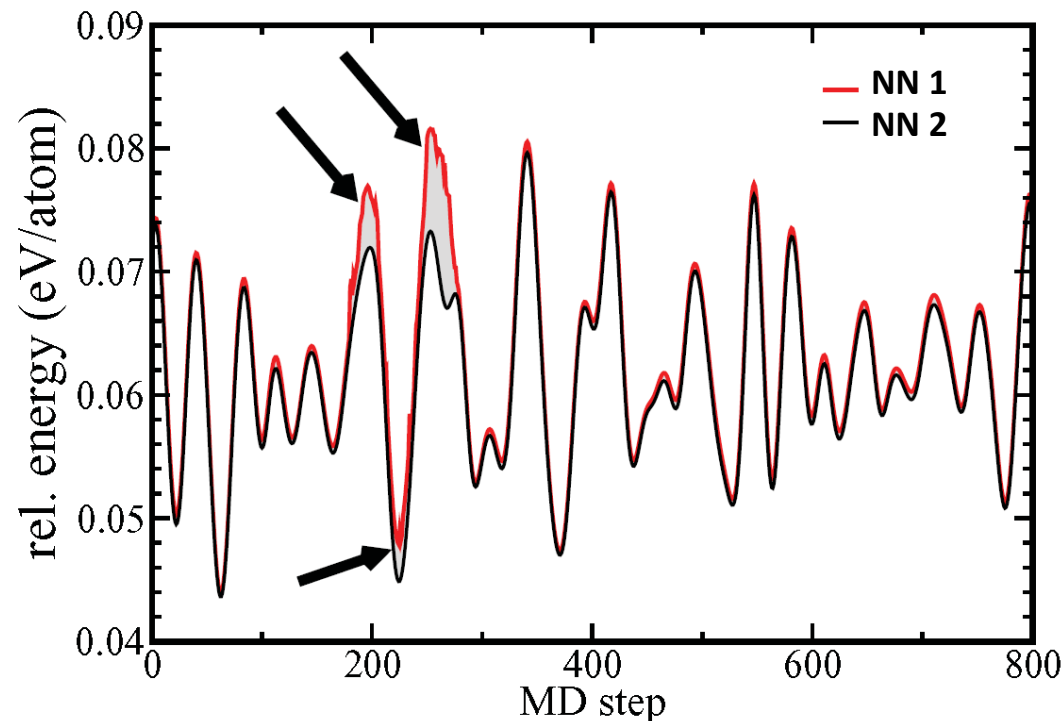
Reason: Non-physical functional form

Solution: Compare predictions of different fits
⇒ **ideal for improving training set**

Neural Network Potentials must be reliable for structures **NOT** included in the training set. **How do we know if a prediction is reliable?**

1. Independent test set not used for training

2. Systematic identification of missing parts of configuration space



⇒ We know if there is a problem without DFT calculations!

N. Artrith and J. Behler, *Phys. Rev. B* **85** (2012) 045439.

- **Limited applicability:**

- Reason: - Complexity of configuration space
- Description by symmetry functions
- Solution: either
- many atoms but few (3-4) elements
or
- many elements but few atoms
or
- restriction of chemical generality

- **Empirical components:**

- Selection of - NN architecture
- Initial weights
- Optimization algorithm
- BUT: You know what you have got!

- **Training is (CPU) time consuming**

There are **unique advantages** of Neural Network Potentials

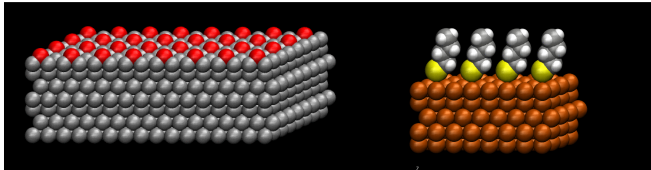
- numerically very accurate (few meV/atom \ll DFT XC error bar)
for all types of systems
- reasonably fast (> 200 atoms per second and core, almost linear scaling)
- general, unbiased, no system specific terms in the functional form
- arbitrary atomic configurations (no grid)
- can be (and have been) used to solve complex physical / chemical problems!

⇒ NN Potentials are tools for difficult systems!

Applicability and Applications

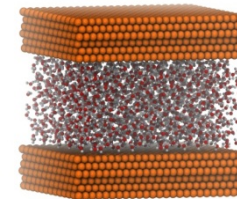
Surface Science:

- heterogeneous catalysis
- corrosion
- self-assembled monolayers



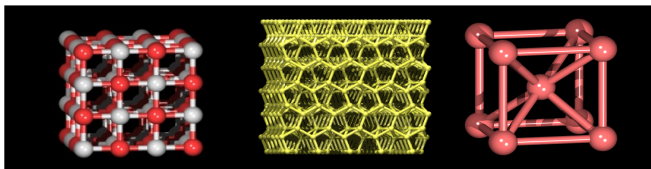
Solid-Liquid Interface:

- heterogeneous catalysis
- electrochemistry



Materials Science:

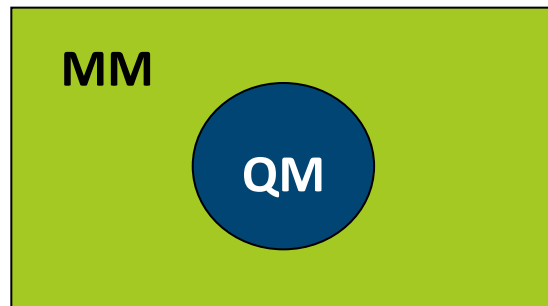
- crystal structure prediction
- phase diagrams
- properties of materials



⇒ **Reliable potentials for large-scale simulations often not available!**

Localized vs. Non-Localized Chemical Reactions

Localized reactions: QM/MM



Example:

- enzymes

Delocalized reactions: NN



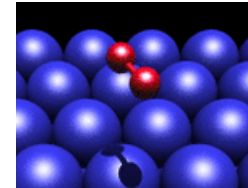
Example:

- structural phase transitions
- chemistry at surfaces and interfaces

- **Oxygen Dissociation at the Al(111) surface**

with K. Reuter, S. Lorenz, M. Scheffler, A. Groß, C. Carbogno

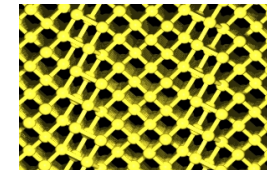
PRL **94** (2005) 036104; *PRB* **75** (2007) 115409; *PRB* **77** (2008) 115421; *PRL* **101** (2008) 096104; *PRB* **81** (2010) 035410.



- **High-Pressure Phase Diagram of Silicon**

with D. Donadio, R. Martoňák, and M. Parrinello

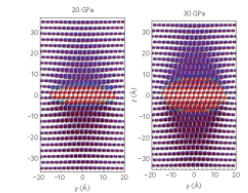
PRL **98** (2007) 146401; *PRL* **100** (2008) 185501; *Phys. Stat. Sol. b* **245** (2008) 2618.



- **Graphite-to-Diamond Nucleation Mechanism**

with R. Khaliullin, H. Eshet, T.D. Kühne and M. Parrinello

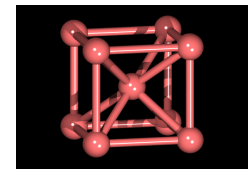
PRB **81** (2010) 100103; *Nature Materials* **10** (2011) 693.



- **Phase Diagram of Sodium**

with H. Eshet, R. Khaliullin, T.D. Kühne and M. Parrinello

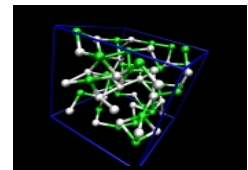
PRB **81** (2010) 184107; *PRL* **108** (2012) 115701.



- **Phase Change Materials (GeTe)**

with G.C. Sosso and M. Bernasconi

PRB **85** (2012) 174103; *PRB* **86** (2012) 104301; *Phys. Stat. Sol. B* **249** (2012) 1880.



Ongoing Work:
**1. Heterogeneous
Catalysis**

Are Neural Network Potentials applicable to metal surfaces?

Data Set for Copper:

	Training	Test
Bulk	13,903	1,545
Slabs	12,483	1,413
Clusters	7,577	842
Total	33,963	3,800

⇒ 617,475 atomic environments

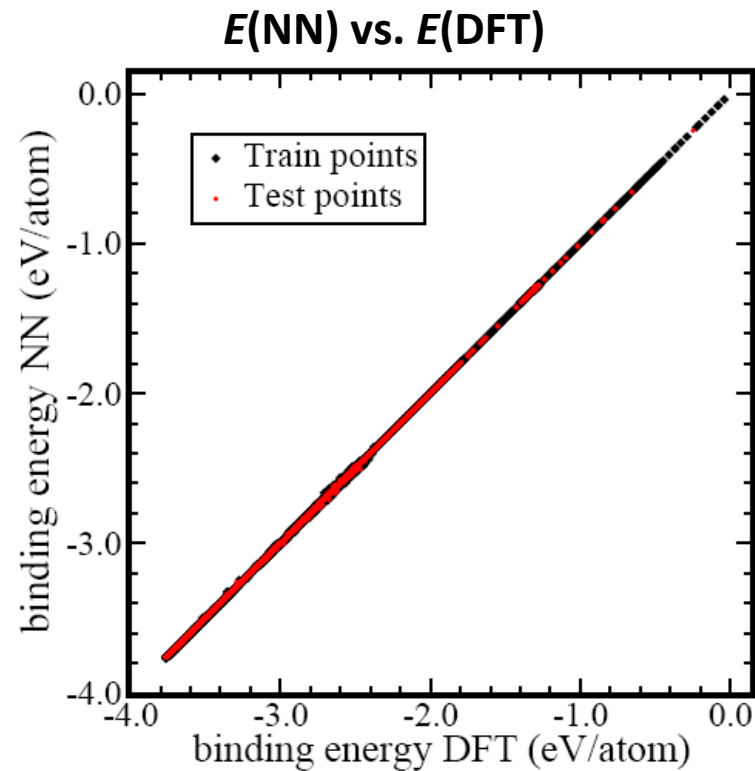
⇒ 1,852,425 pieces of information

RMSEs:

	E (eV/atom)	F (eV/Bohr)
Training	0.0036	0.0428
Test	0.0039	0.0420

DFT Code: FHI-aims, PBE XC

V. Blum *et al.*, *Comp. Phys. Comm.* **180** (2009) 2175.

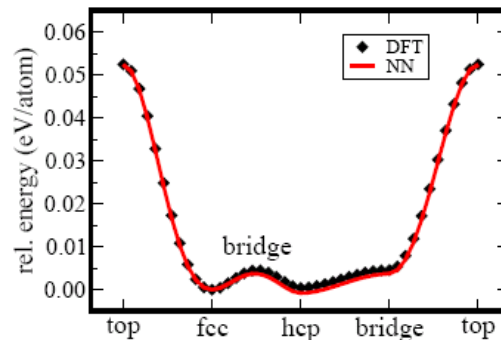
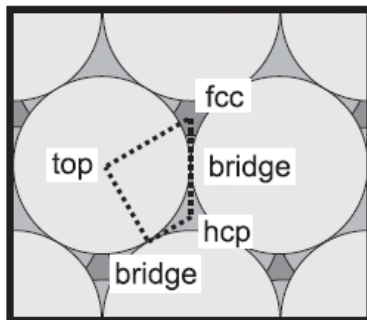


NN architecture: 51-30-30-1 *ttl*

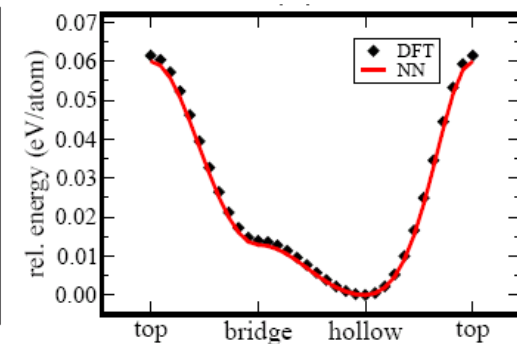
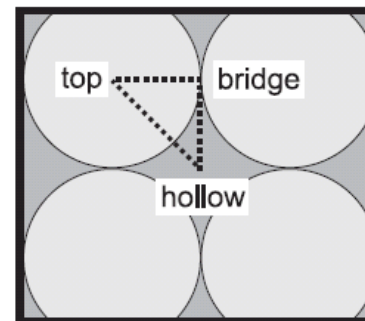
Bulk Copper:

Structure	E_{coh} eV/atom		Lattice parameters Å		B GPa	
	DFT	NN	DFT	NN	DFT	NN
fcc	3.763	3.756	a=3.630	a=3.630	140	138
bcc	3.719	3.716	a=2.885	a=2.887	137	135
sc	3.281	3.282	a=2.407	a=2.407	103	108
hcp	3.740	3.740	a=4.862 c/a=1.63	a=4.856 c/a=1.63	-	-

Cu(111):

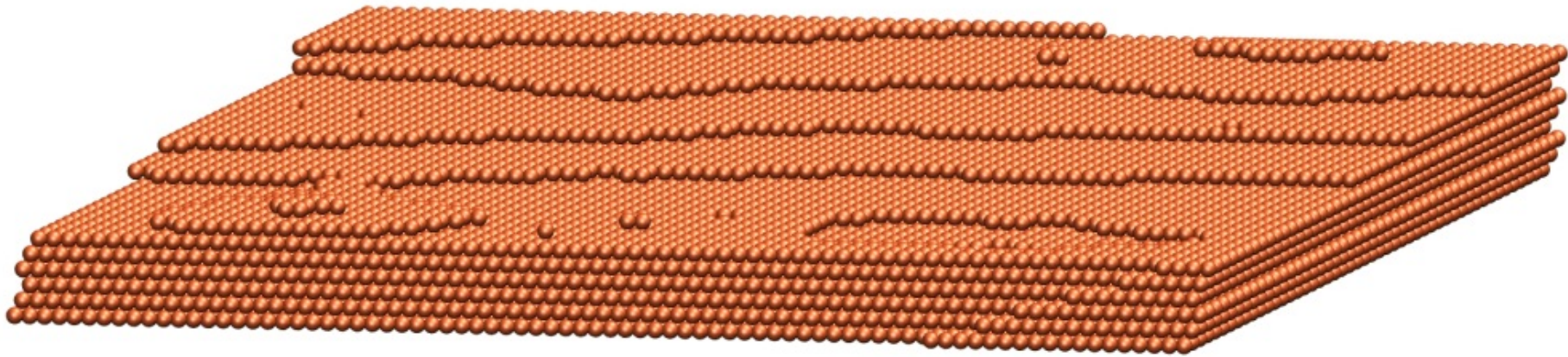


Cu(100):



N. Artrith and J. Behler, *Phys. Rev. B* **85** (2012) 045439.

Model of a “real” surface with steps, kinks, and adatoms ($\approx 29\,000$ atoms).



DFT \Rightarrow impossible

NN $\Rightarrow \approx 2$ mins (single core)

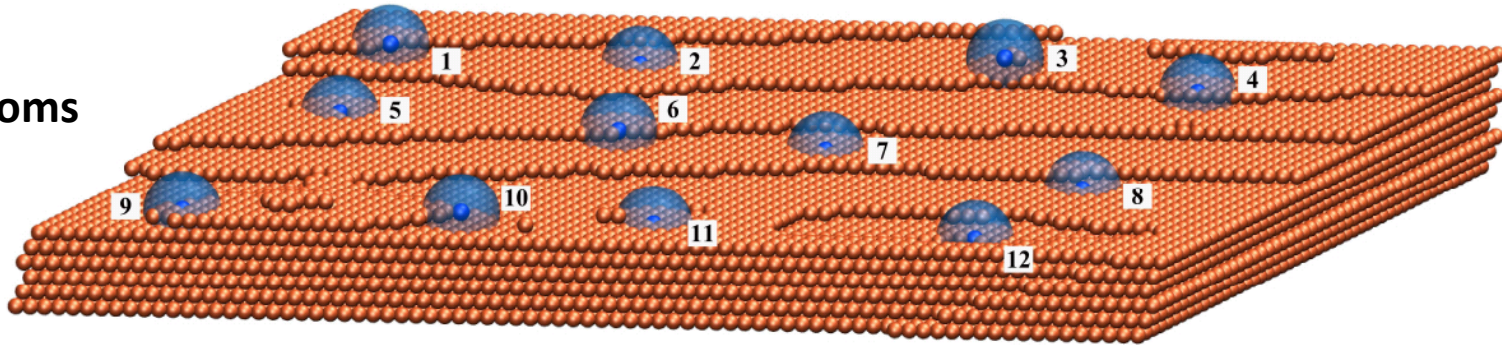
\Rightarrow NN potentials can be used to study systems of this size.

How do we know if the potential is correct?

N. Artrith and J. Behler, *Phys. Rev. B* **85** (2012) 045439.

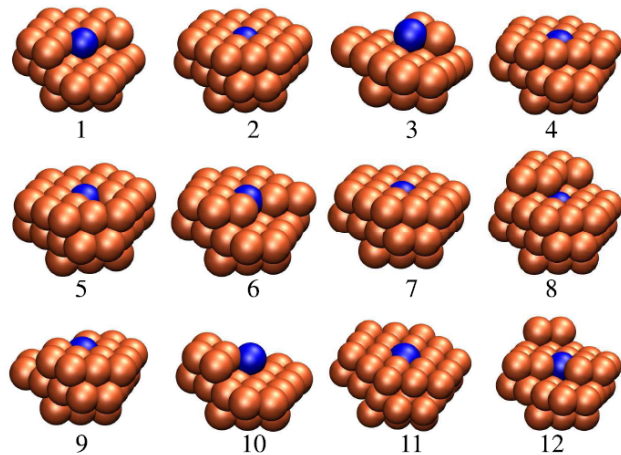
A "Real" Copper Surface:

29 000 atoms

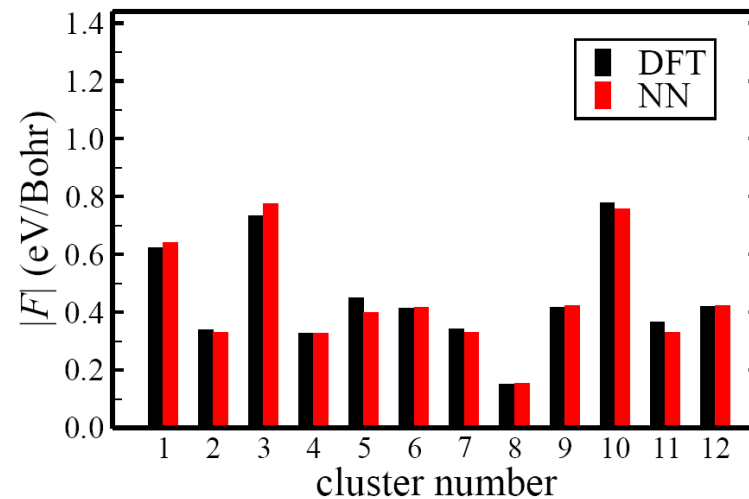


We need local properties related to the potential energy surface \Rightarrow forces

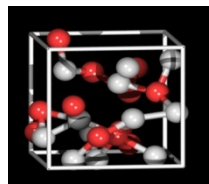
Clusters:



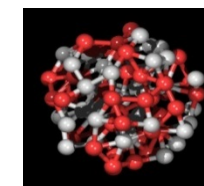
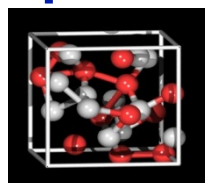
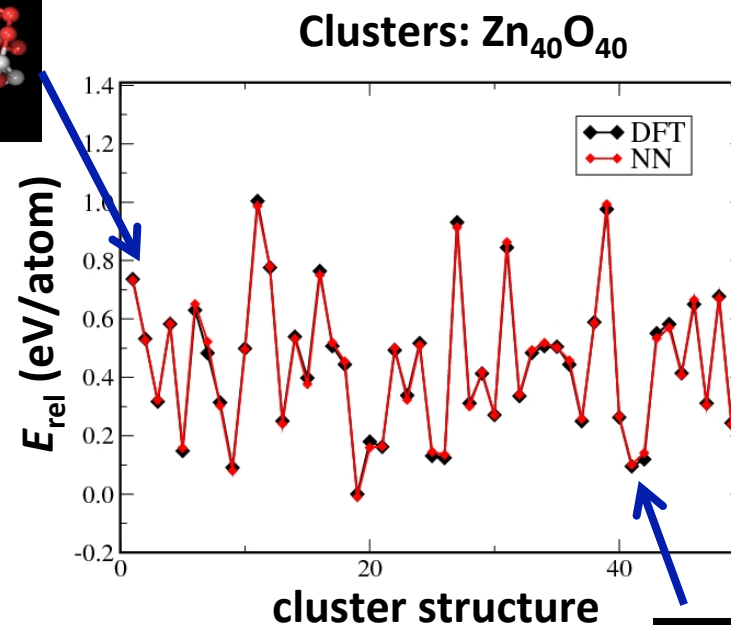
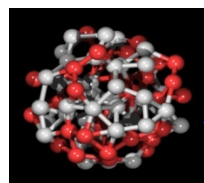
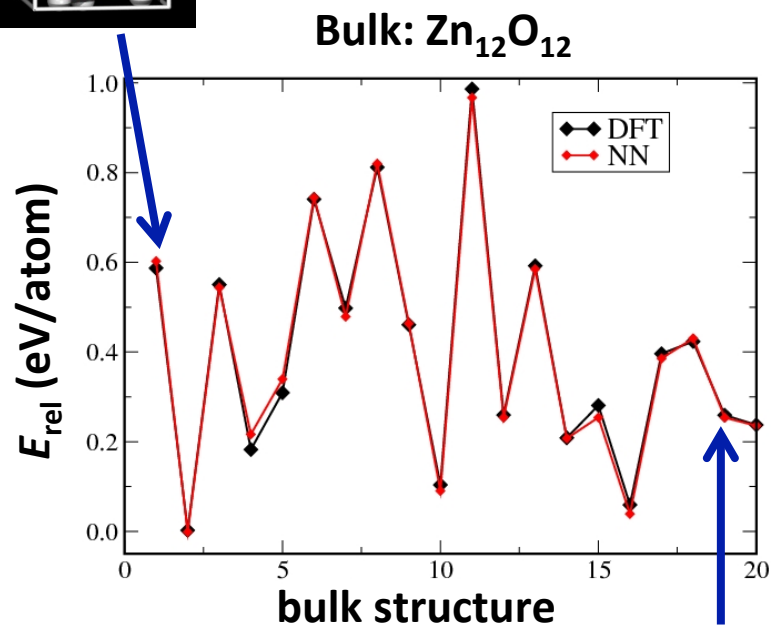
Forces at central atoms in clusters:



N. Artrith and J. Behler, *Phys. Rev. B* **85** (2012) 045439.



Comparison DFT - NN:



Neural Network Performance:

Example: 1000 atom system: 10 s CPU time on standard-PC (energy + forces)

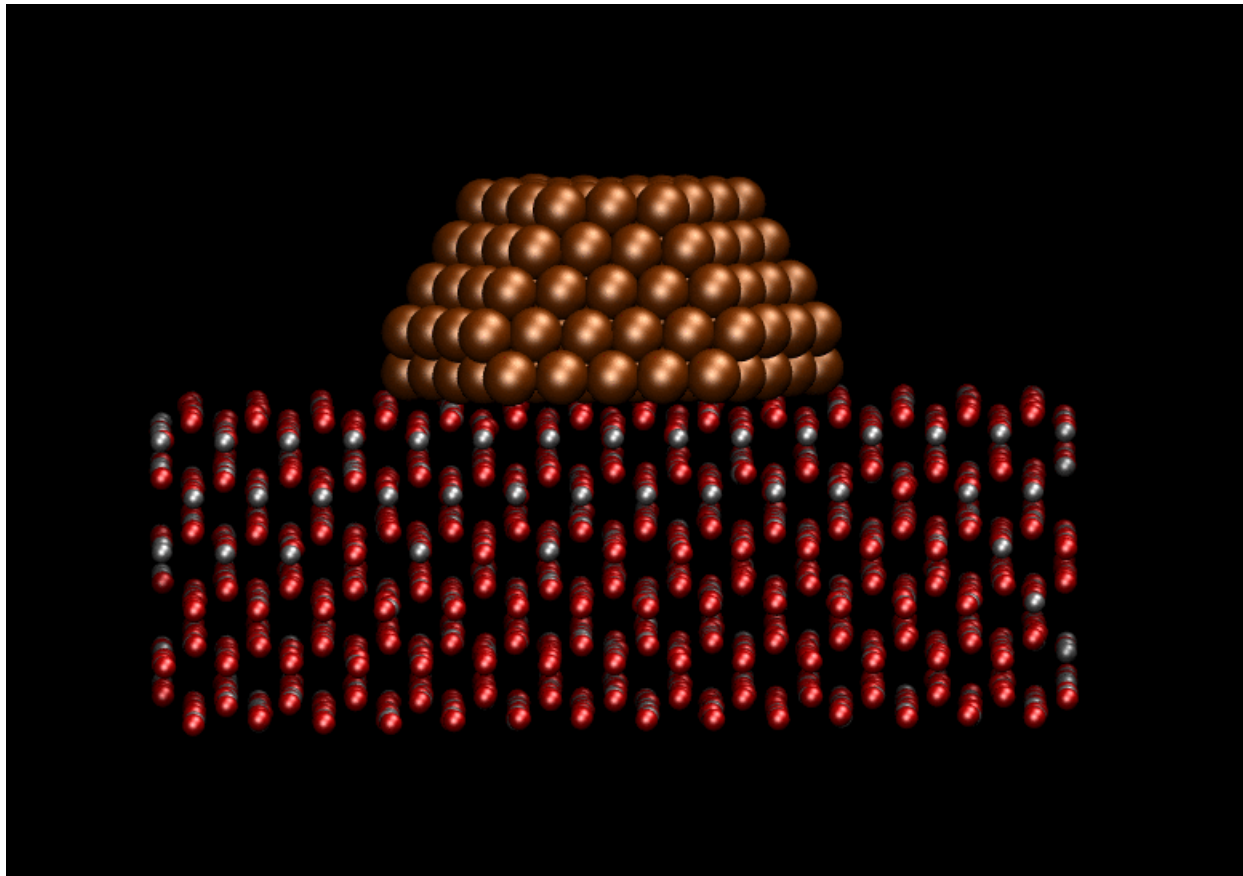
From monocomponent and binary systems to CuZnO:

System	Structures	Energy eV / atom		Force eV / Bohr	
		train	test	train	test
copper	40 000	0.004	0.004	0.043	0.042
ZnO	40 000	0.002	0.003	0.161	0.162
CuZnO	100 000	0.005	0.005	0.094	0.089

Ternary NN architecture: 156-15-15-15-1

- ⇒ NN potential for **ternary system has the same quality** as for pure Cu or ZnO!
- ⇒ Limit of NNs has not yet been reached

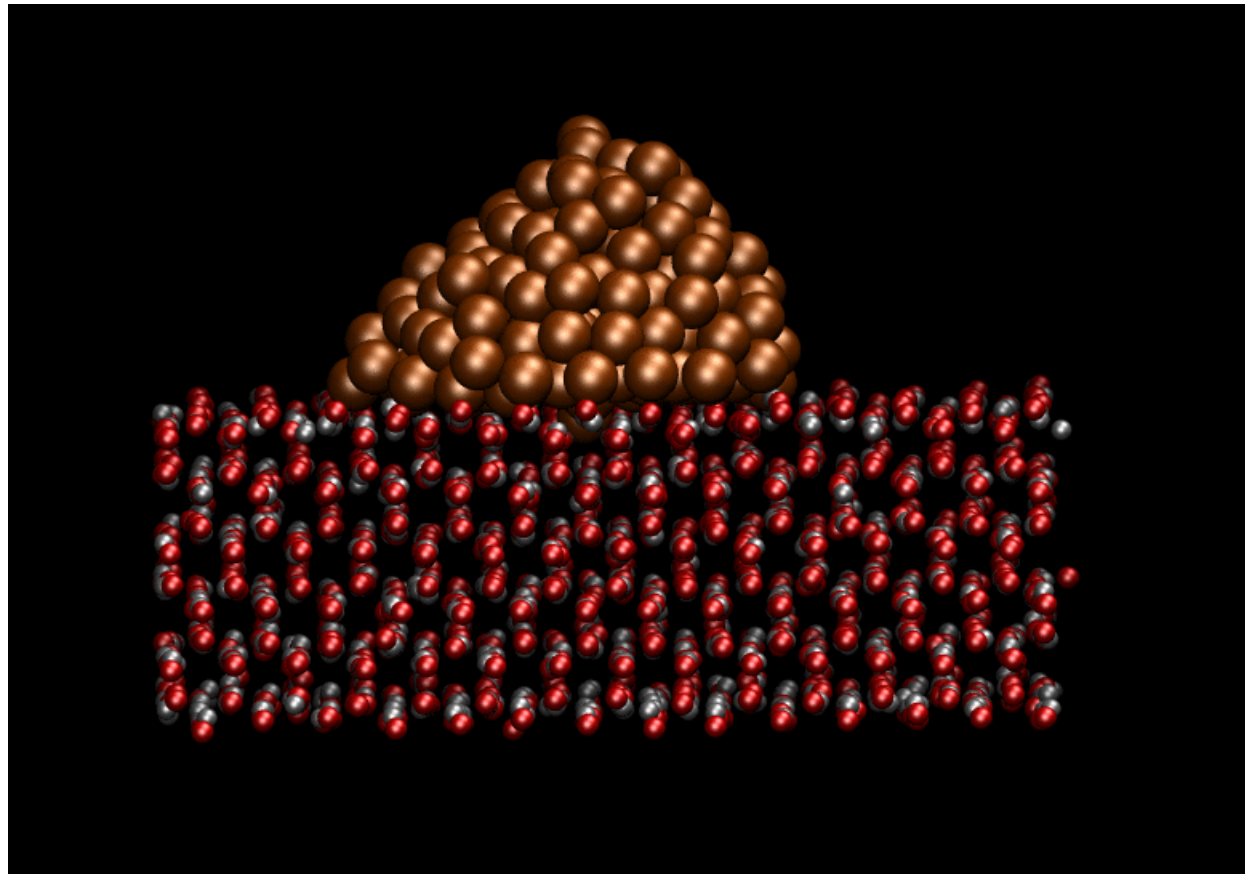
Cu(111)/ZnO(10-10)



233 copper atoms

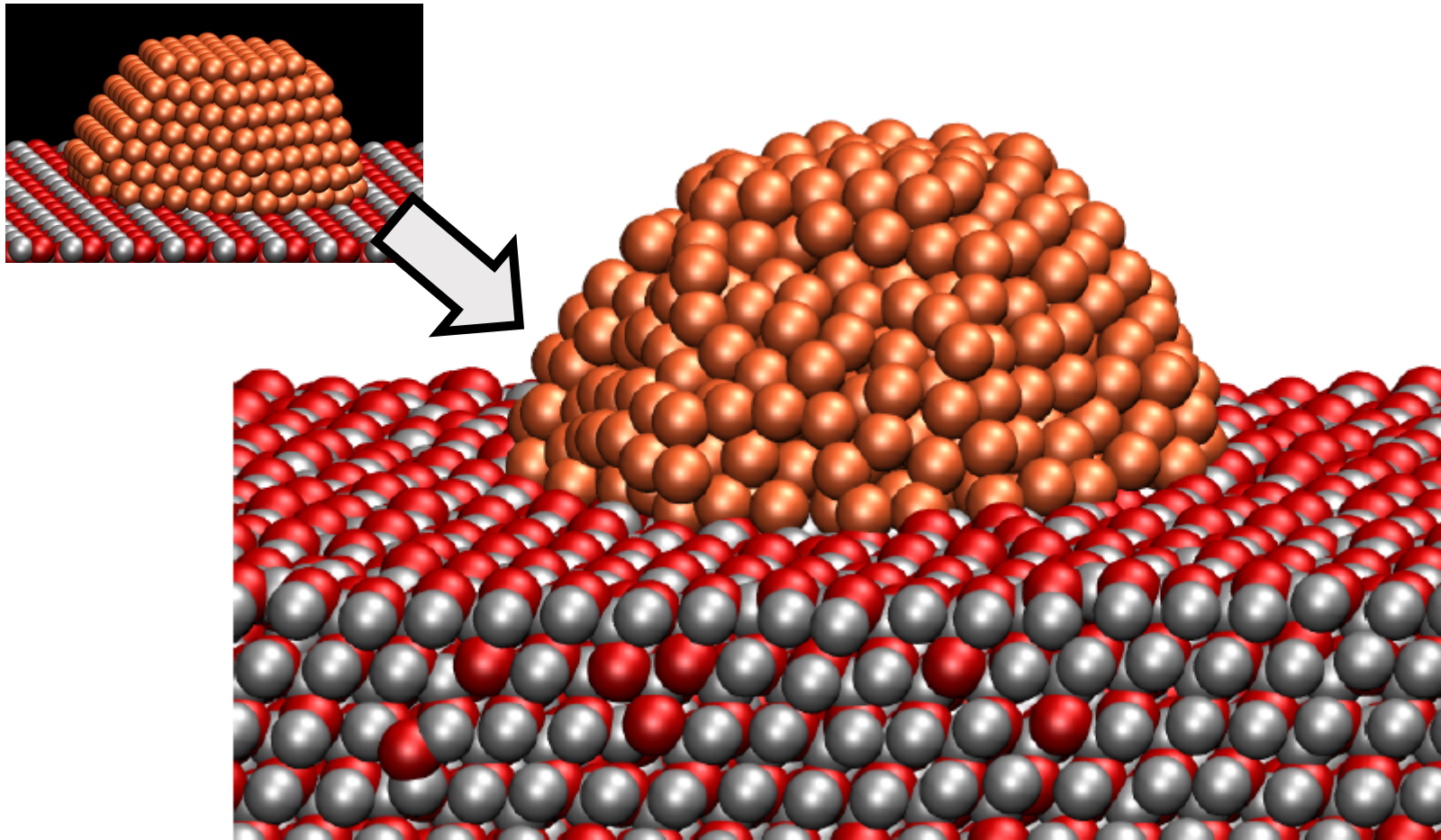
1344 ZnO formula units

⇒ 2921 atoms



⇒ MD simulations allow to improve the potential systematically

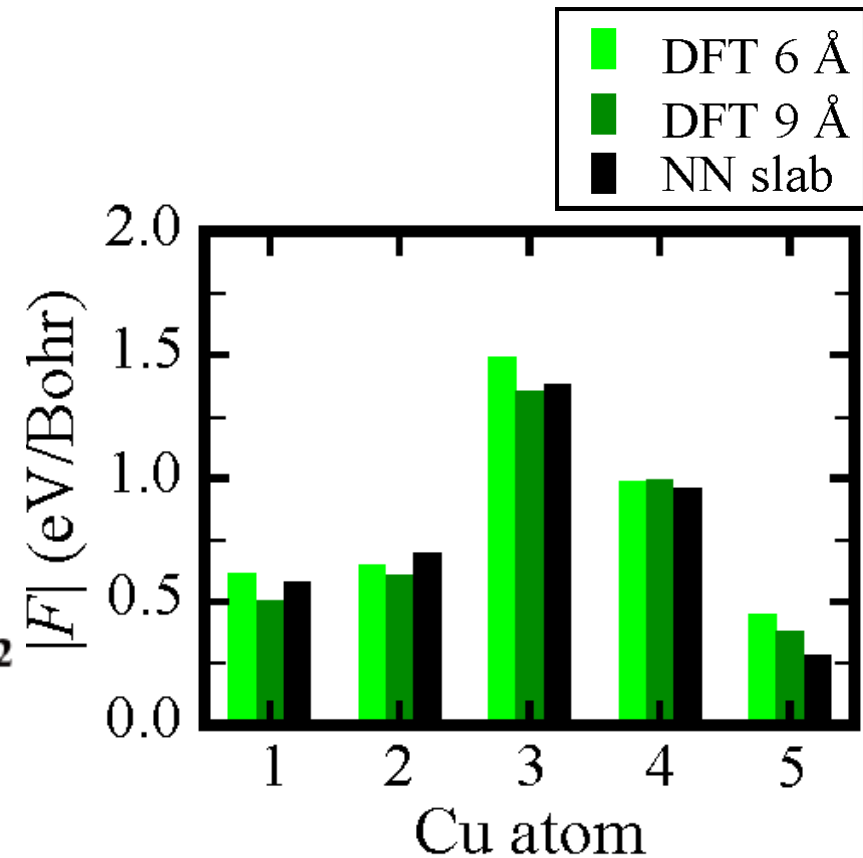
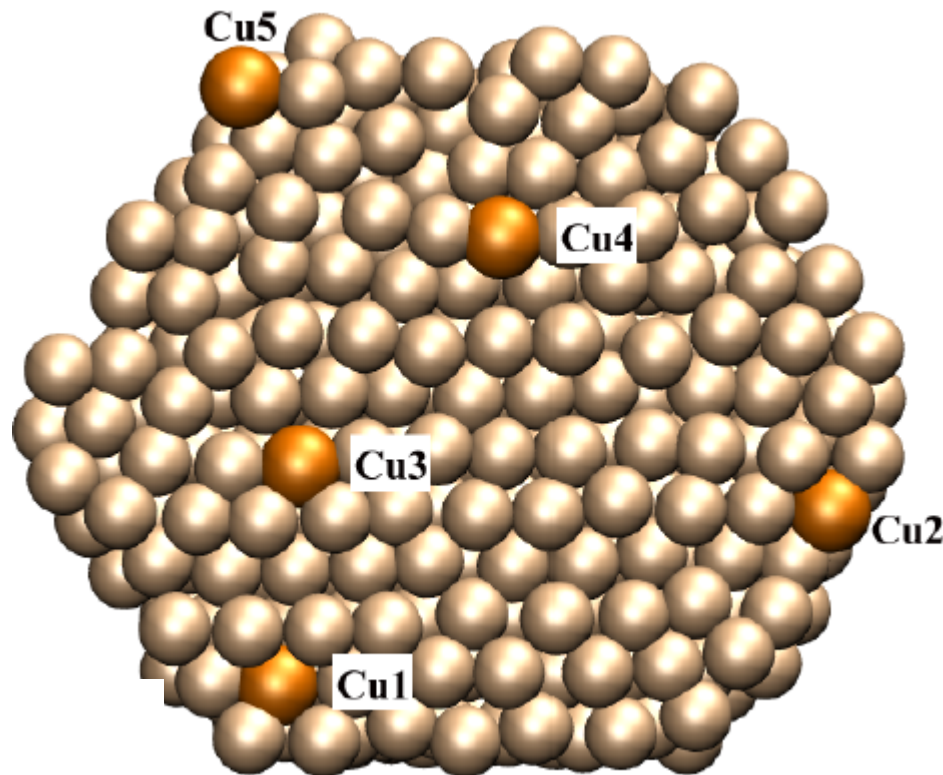
Random Snapshot from MD Simulation at 1000 K



N. Artrith, B. Hiller, and J. Behler, *Phys. Status Solidi B*, 250, 1191 (2012).

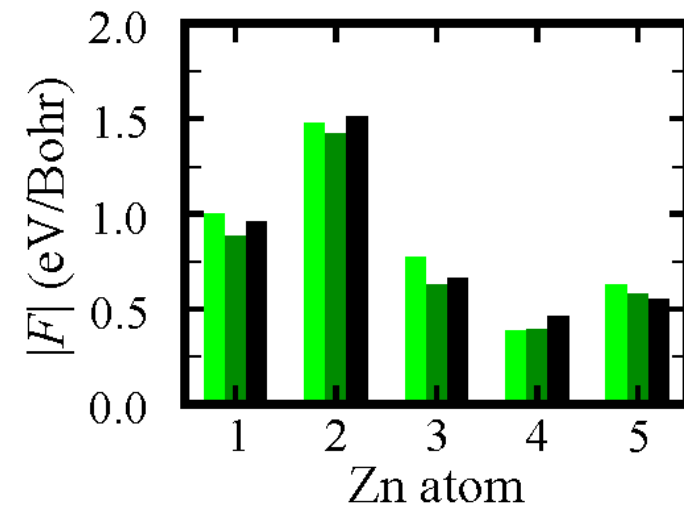
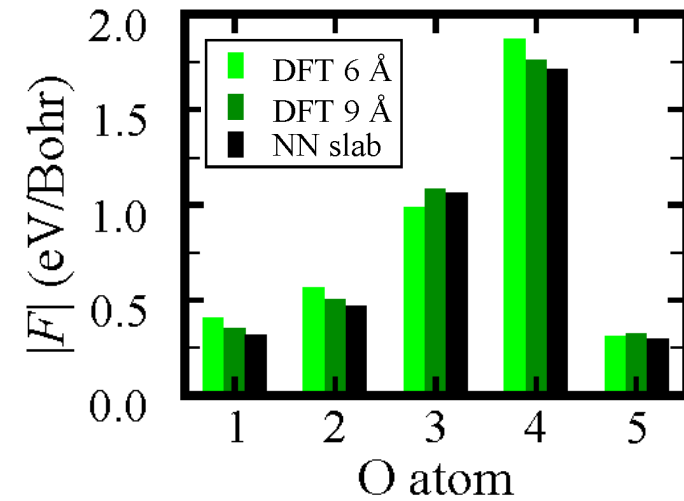
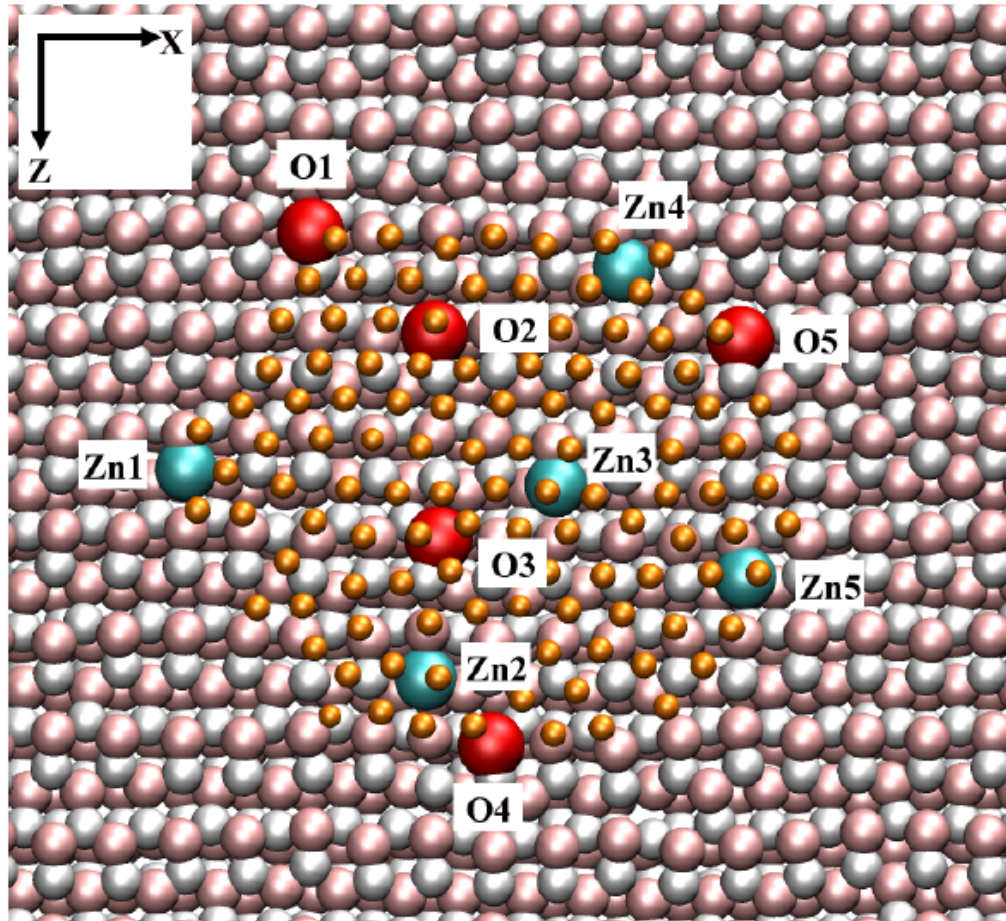
Forces at Selected Copper Atoms of the Interface

Bottom View:

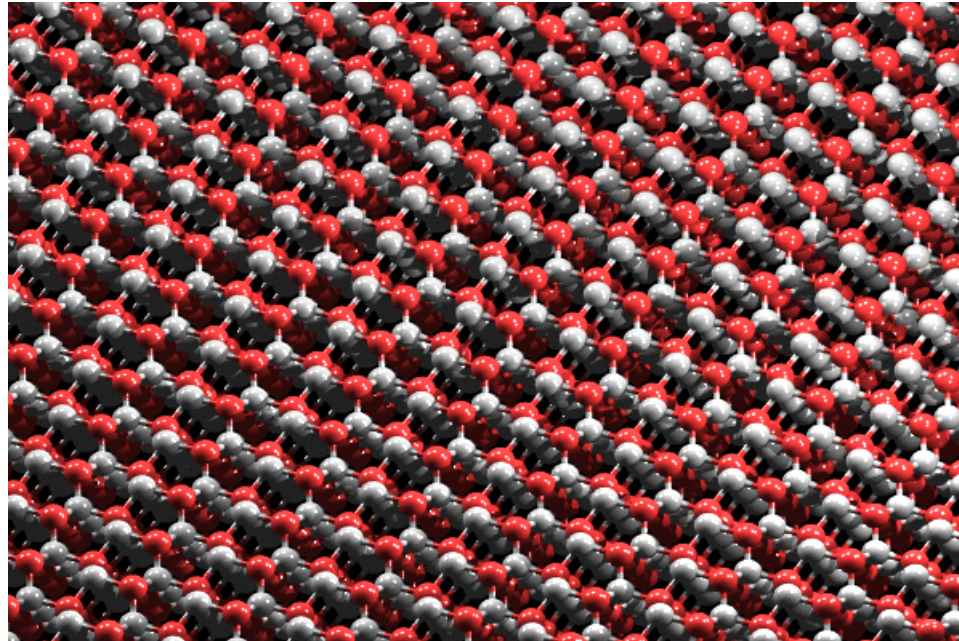


N. Artrith, B. Hiller, and J. Behler, *Phys. Status Solidi B*, 250, 1191 (2012).

Forces at Selected Zinc and Oxygen Atoms at the Interface

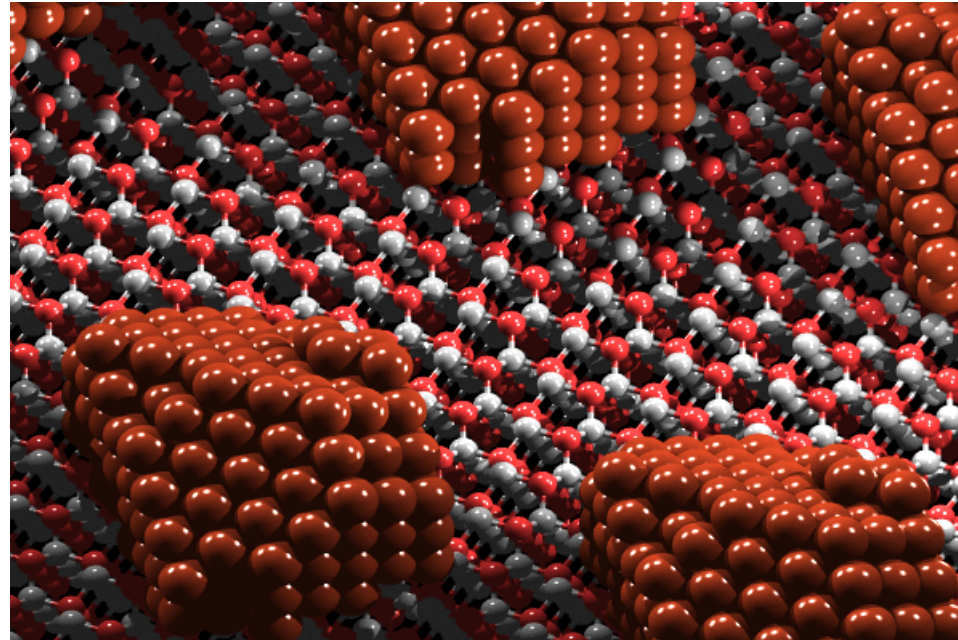


Preliminary Monte Carlo results:



- (5 x 10) ZnO(1 0 -1 0) supercell, 6 layers \Rightarrow 600 atoms
- growth of 200 Cu atoms
- 1000 Metropolis equilibration steps per added Cu atom

Preliminary Monte Carlo results:



⇒ cluster shape close to Wulff constructions

Next steps: lattice-free simulations

Conclusions

Advantages of Neural Network Potentials

- provide very **accurate energies** (error ≈ 5 meV/atom)
- only atomic positions needed \Rightarrow **“reactive”**
- **fast** (100 atoms per second per core)
- can be combined with any electronic structure method
- **universal** functional form, no system-specific terms
 \Rightarrow applicable to solids, molecules and surfaces
- **systematic** improvement possible

Disadvantages of Neural Network Potentials

- large training sets needed \Rightarrow **costly**
- currently **restricted** to 3-4 chemical elements
- “non-physical” (unbiased) functional form,
 \Rightarrow **limited extrapolation** capabilities
 \Rightarrow construction and application needs to be done with care



PCCP Vol. 13, Issue 40, 2011

The Group@RUB

Christopher Handley
Björn Hiller
Felix Hoffmann
Eriko Hojo
Philipp Schienbein
Sinja Klees
Suresh Kondati Natarajan
Tobias Morawietz
Porntip Seema

Alumni:

Nongnuch Artrith (MIT)
Jovan Jose K.V. (Indiana University)
Marcus Maschke (RUB)
Vikas Sharma (University of Delhi)



Fritz-Haber Institut Berlin

Sönke Lorenz

Matthias Scheffler

Volker Blum + FHI-aims team

Karsten Reuter (now TU München)

University of Ulm

Christian Carbogno (now FHI)

Axel Groß

ETH Zürich

Michele Parrinello

Roman Martoňák (now Bratislava)

Thomas Kühne (now University of Mainz)

Hagai Eshet (now Tel Aviv University)

Rustam Khaliullin (now University of Mainz)

Ruhr-Universität Bochum

Dominik Marx + Theoretische Chemie

Christof Wöll (now KIT), Martin Muhler + SFB 558

Martina Havenith + RESOLV

Universität Wien

Andreas Singraber

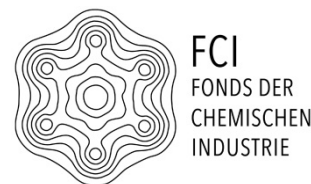
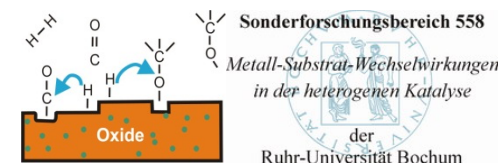
Christoph Dellago

Università di Milano Bicocca

Gabriele Soso

Marco Bernasconi

Funding:



RUB

RESEARCH DEPARTMENT
Interfacial Systems Chemistry

Thank you!

