The Cambrian Explosion of Machine Learning Potentials

Jörg Behler

Lehrstuhl für Theoretische Chemie II
Ruhr-Universität Bochum, Germany
and
Research Center
Chemical Sciences and Sustainability
Research Alliance Ruhr

1

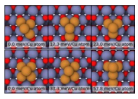
Introduction

Goal:

Atomic-level understanding of complex systems in chemistry and materials science

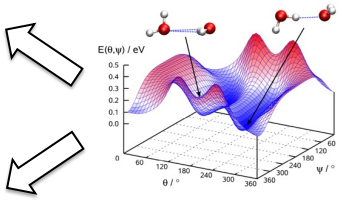
⇒ Predictive computer simulations with first-principles quality

Energy
global and local minima

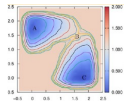


Forces
dynamics, free energies

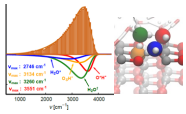
Central Role:
Potential Energy Surface



Reactions
barriers / transition states



Vibrations
properties, analysis



⇒ The accuracy of the obtained results depends on the quality of the PES

2

Definition of Machine Learning Potentials

RUB

There is no universally accepted definition of machine learning potentials

Challenge: The construction essentially all atomistic potentials involves fitting/parameter optimization.

⇒ Where is the border between fitted potentials and MLPs?

A pragmatic definition of Machine Learning Potentials

- A machine learning potential uses a flexible machine learning method to represent the potential energy surface (total energy and its analytic derivatives) as a function of the atomic coordinates.

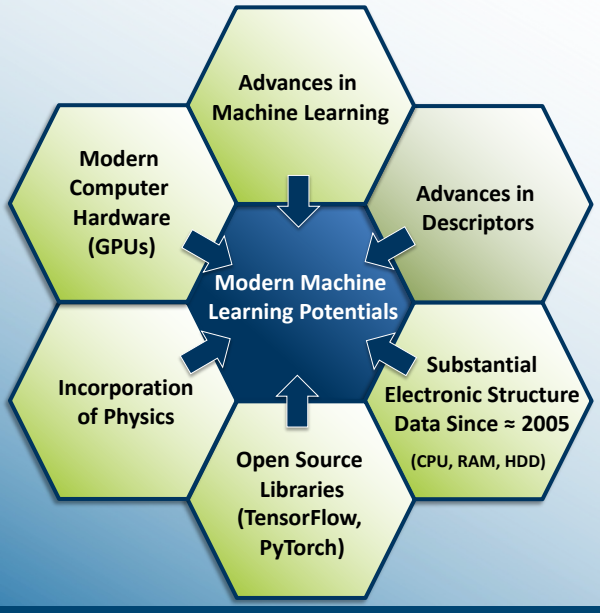
  ⇒ in principle no physical functional form needed
- A machine learning potential is constructed using a consistent set of reference electronic structure calculations.

⇒ No "artificial intelligence", just brute force fitting (?)

3

The Cambrian Explosion of Machine Learning Potentials: Reasons

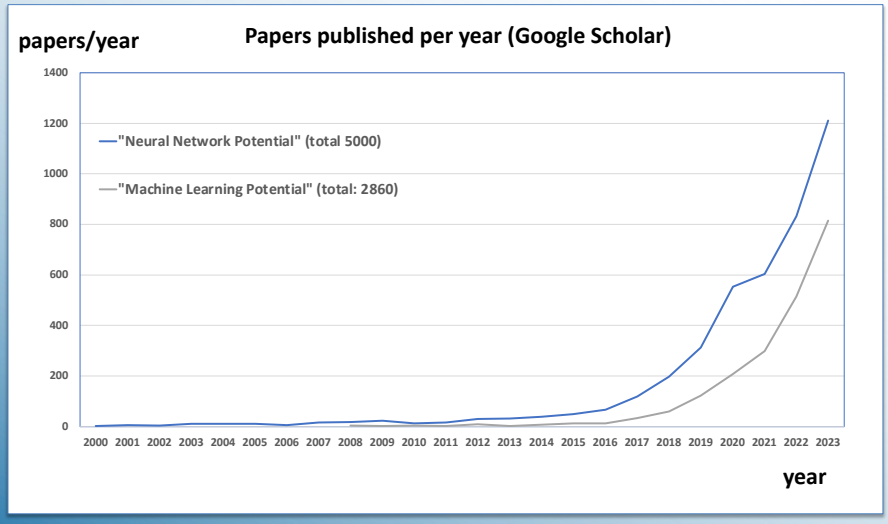
RUB



4

The Cambrian Explosion of Machine Learning Potentials

RUB



5

Machine Learning Potentials

RUB

How to get an overview about the huge amount of papers?

We need a classification scheme for machine learning potentials
What are the criteria?

Historical?

Physics?

Applicability?

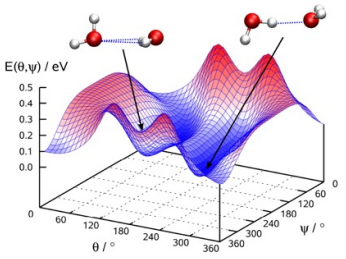
ML Method?

Descriptors?

⇒ There is no unique way

6

First-Generation Machine Learning Potentials 1995 – 2007



7

Machine Learning Potentials

RUB

1995 2000 2005 2010 2015 2020

First-Generation MLPs

- about 30 papers (1995 – 2007) from about 10 groups
- all methods in the first decade are based on neural networks

Examples:

T.B. Blank, S.D. Brown, A.W. Calhoun, and D.J. Doren, *J. Chem. Phys.* **103** (1995) 4129.
H. Gassner, M. Probst, A. Lauenstein, K. Hermansson, J. Phys. Chem. A **102** (1998) 4596.
S. Lorenz, A. Groß, M. Scheffler, Chem. Phys. Lett. **395** (2004) 210.
S. Manzhos, T. Carrington, Jr., J. Chem. Phys. **125** (2006) 194105.
J. Behler, S. Lorenz, K. Reuter, J. Chem. Phys. **127** (2007) 014705.

Focus:

Training
Symmetry
Surfaces
Spectroscopy
Surface Symmetry

⇒ **Basic ideas and key concepts are well established**

Jörg Behler Grenoble 2024

8

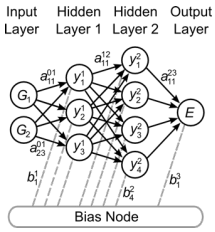
Machine Learning Potentials

RUB

1995 2000 2005 2010 2015 2020

First-Generation MLPs

Limitation: Applicable to low-dimensional systems only

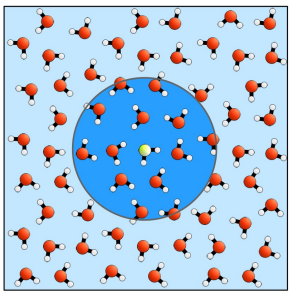


Challenges:

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

⇒ **Another approach is required for high-dimensional systems**

Second-Generation Machine Learning Potentials: Locality



High-Dimensional Neural Network Potentials

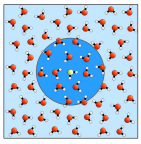
RUB

3 Steps:

1. Total energy is the sum of atomic energies

$$E = \sum_i E_i$$

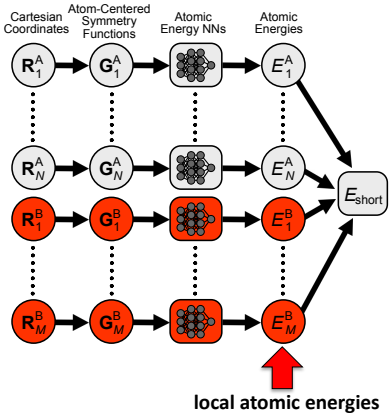
2. Atomic energies depend on local environments
   ⇒ cutoff radius



3. Description of local atomic environments by many-body atom-centered symmetry functions
   ⇒ structural fingerprints
   (invariances: rotation, translation, permutation)

J. Behler, M. Parrinello, *Phys. Rev. Lett.* 98 (2007) 146401.
J. Behler, *J. Chem. Phys.* 134 (2011) 074106.
J. Behler, *Angew. Chem. Int. Ed.* 56 (2017) 12828.

Structure for a binary system A_NB_M

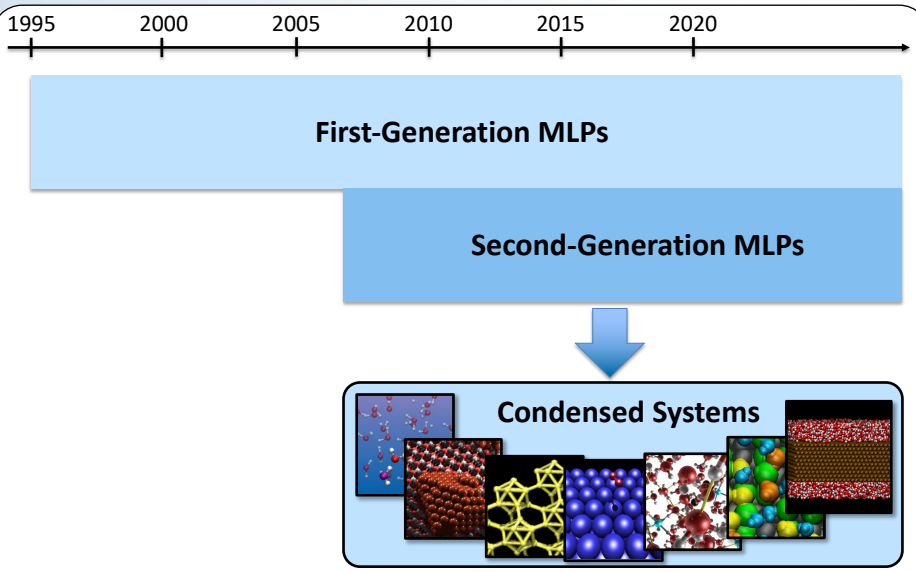


⇒ applicable to thousands of atoms

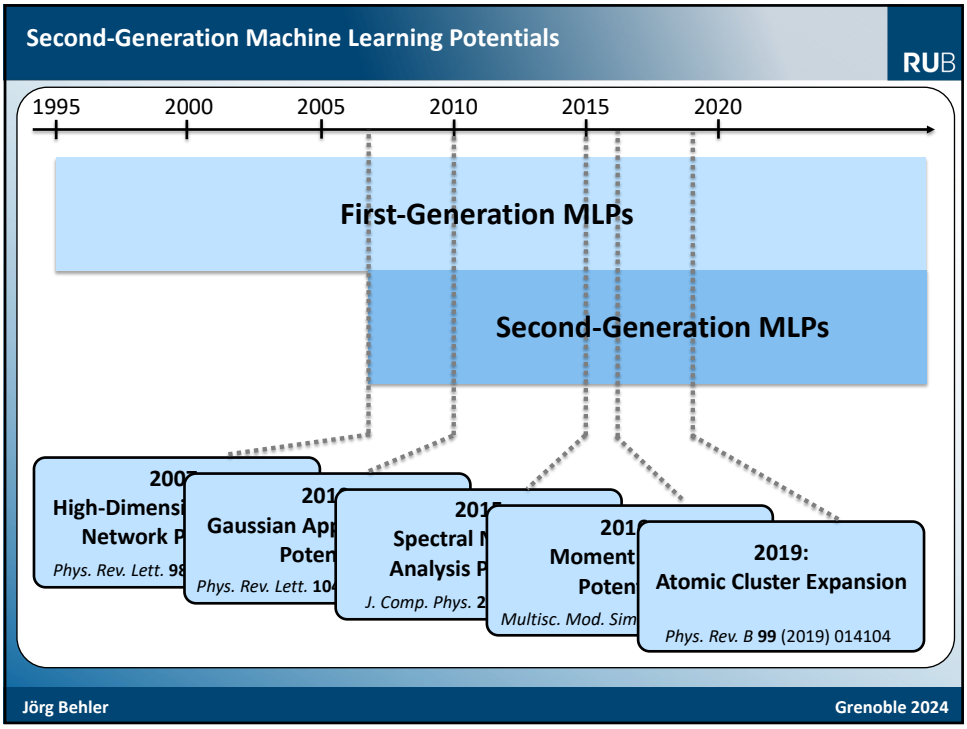
11

Second-Generation Machine Learning Potentials

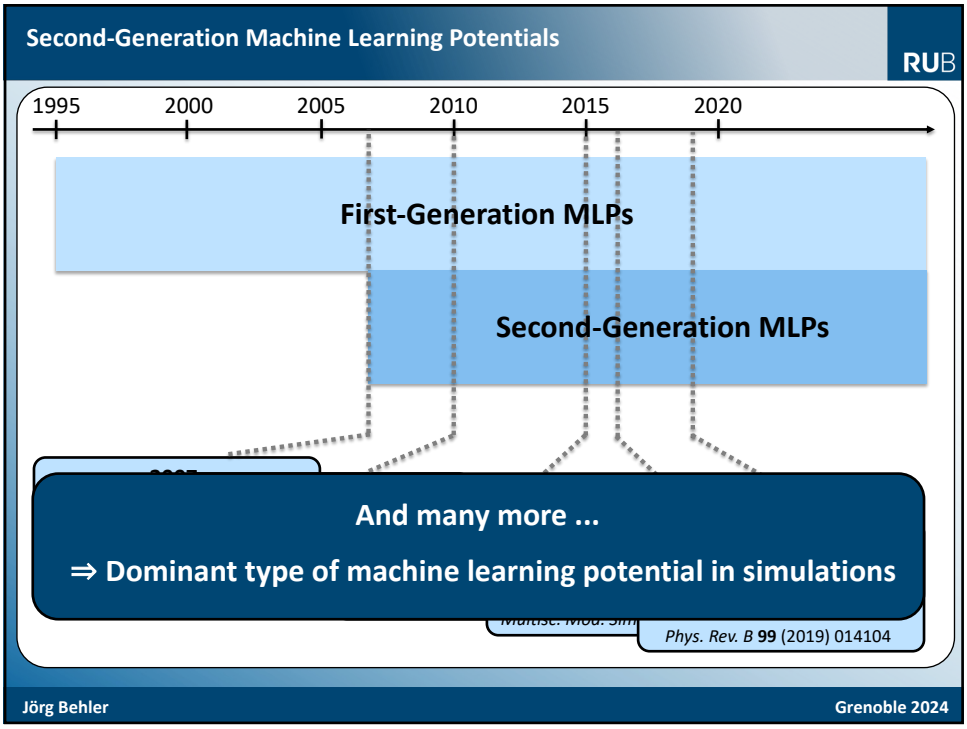
RUB



12


Second-Generation Machine Learning Potentials
RUB
1995
2000
2005
2010
2015
2020
First-Generation MLPs
Second-Generation MLPs
2019:
Atomic Cluster Expansion
Phys. Rev. B 99 (2019) 014104
Jörg Behler
Grenoble 2024

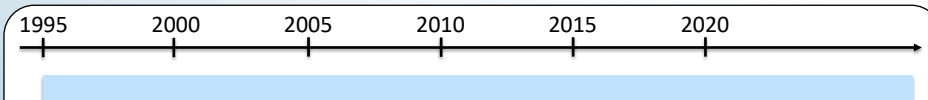
13


Second-Generation Machine Learning Potentials
RUB
1995
2000
2005
2010
2015
2020
First-Generation MLPs
Second-Generation MLPs
And many more ...
⇒ Dominant type of machine learning potential in simulations
Phys. Rev. B 99 (2019) 014104
Jörg Behler
Grenoble 2024

14

Second-Generation Machine Learning Potentials

RUB



First-Generation MLPs

Second-Generation MLPs

Limitation: Truncation of the interactions at the cutoff radius

15

Hessian-based Locality Test

RUB

A Hessian-Based Analytic Locality Test

Challenge: Atomic forces depend on the entire environment

- different neighbors interact differently
  ⇒ distance and type of bonding are important
  ⇒ ***influence of individual neighbors*** is of interest
- forces are very small or even zero in symmetric environments
  ⇒ ***forces are not a good measure*** for interaction strength
- force convergence of central atom measures only total interaction
  ⇒ ***forces can cancel each other*** also in non-symmetric environments

Solution:

Hessian: $H_{A_\alpha B_\beta} = \frac{\partial^2 E}{\partial A_\alpha \partial B_\beta} = -\frac{\partial f_{B_\beta}}{\partial A_\alpha} = -\frac{\partial f_{A_\alpha}}{\partial B_\beta}$ with $\alpha, \beta = \{x, y, z\}$

⇒ information about dependence of an atomic force on each atom in the system

⇒ spatial cutoff can be quantified!

M. Herbold and J. Behler, J. Chem. Phys. **156** (2022) 114106.

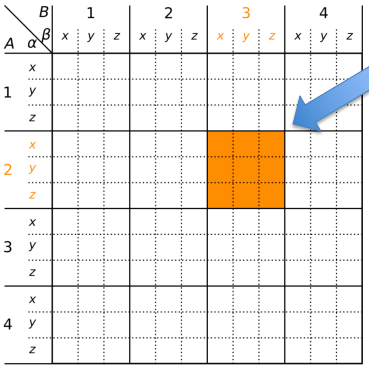
16

Hessian-based Locality Test

RUB

DFT Hessian: $H_{A_\alpha B_\beta} = \frac{\partial^2 E}{\partial A_\alpha \partial B_\beta} = -\frac{\partial f_{B_\beta}}{\partial A_\alpha} = -\frac{\partial f_{A_\alpha}}{\partial B_\beta}$ with $\alpha, \beta = \{x, y, z\}$

Hessian: (3*N* x 3*N*) matrix



Dependence of atomic force vector on atomic position of a neighbor: (3x3) matrix

⇒ inconvenient, scalar property needed

⇒ Hessian submatrix norm

$$||\mathbf{h}_{AB}|| = \sqrt{\sum_{\alpha=x,y,z} \sum_{\beta=x,y,z} h^2_{A_\alpha B_\beta}}$$

Dependence of force on each individual atom in the system can be quantified

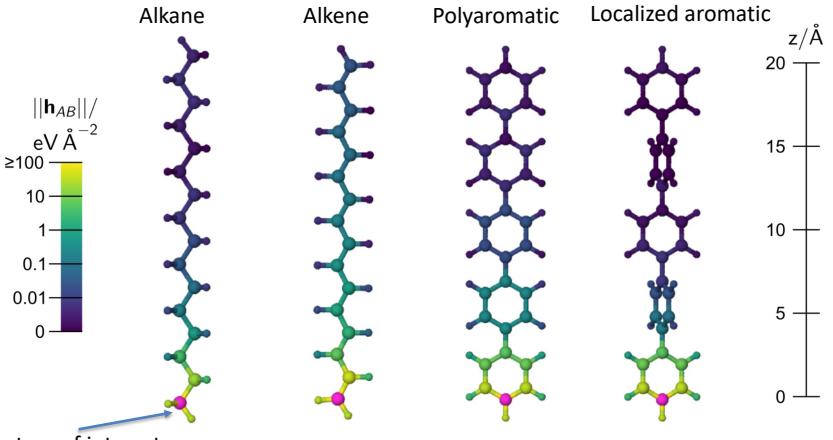
M. Herbold and J. Behler, J. Chem. Phys. **156** (2022) 114106.

17

Hessian-based Locality Test

RUB

1D Model Systems



atom of interest: terminal **carbon**

- contribution of each atom in the system can be quantified
- no cancellation of contributions possible

M. Herbold and J. Behler, J. Chem. Phys. **156** (2022) 114106.

18

Second-Generation Machine Learning Potentials

RUB

2G MLPs are the main production methods for MLP-driven simulations

There are now two classes:

Predefined Descriptors (HDNNP, SOAP, DeePMD,...)

- simple ⇒ fast
- most applications so far
- more difficult for many elements

Message Passing Neural Networks (DTNN, SchNet, MACE, ...)

- iterative extension of environment
- learnable features
  ⇒ more demanding

⇒ for a given system the accuracy is very similar

19

Second-Generation Machine Learning Potentials

RUB

If the training of the MLP for a system does not work, there is a physical reason ⇒ must be understood

1) Data is not accurate enough

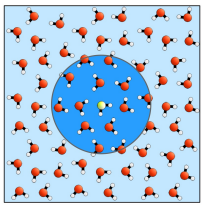
2) Descriptors are insufficient

⇒ properties of different structures are averaged ⇒ large errors

3) Physics is missing

The cutoff used in second-generation MLPs is an approximation!

⇒ Convergence tests are needed



Locality tests:

Statistical analysis:

V. L. Deringer and G. Csányi, Phys. Rev. B 95 (2017) 094203.

Hessian-based:

M. Herbold and J. Behler, J. Chem. Phys. **156** (2022) 114106.

⇒ For some systems, long-range interactions may be important!

20

Third-Generation Machine Learning Potentials: Long-Range Interactions

What about electrostatics?

21

Learning Electrostatic Multipoles

RUB

Goal:

Use ML to represent flexible electrostatic charges and multipoles as a function of the molecular structure.

⇒ improved description of electrostatics in classical force fields for small molecules

Main methods: Kriging and Neural Networks

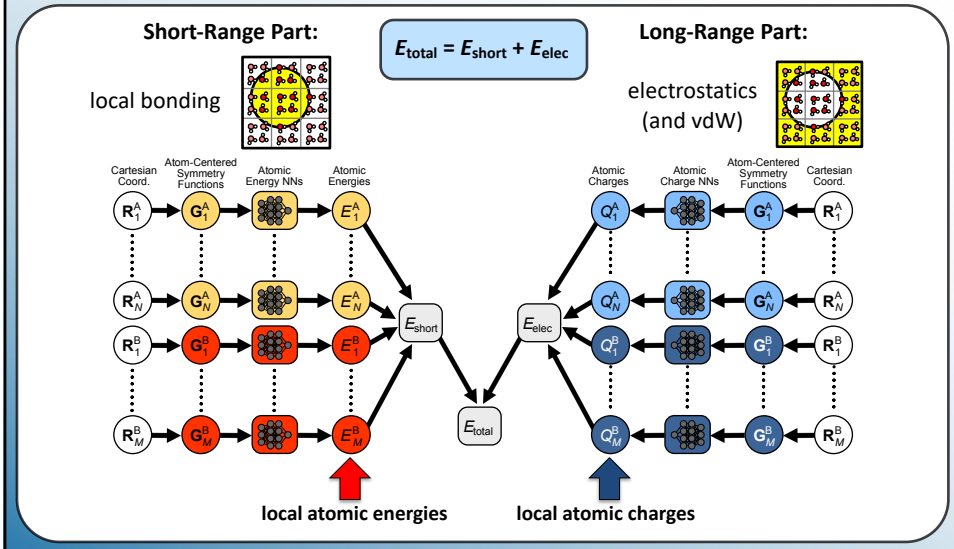
⇒ not a full MLP, since only the electrostatic component is constructed using ML

S. Houlding, S. Y. Liem and P. L. A. Popelier, Int. J. Quant. Chem. **107** (2007) 2817.

22

Third-Generation Neural Network Potentials: Method

RUB



N. Artrith, T. Morawietz and J. Behler, *Phys. Rev. B* 83 (2011) 153101.
T. Morawietz and J. Behler, *J. Chem. Phys.* 136 (2012) 064103.

23

Machine Learning Potentials

RUB

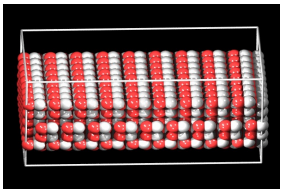
1995 2000 2005 2010 2015 2020

First-generation MLPs

Second-generation MLPs

Third-generation MLPs

Zinc Oxide Surface

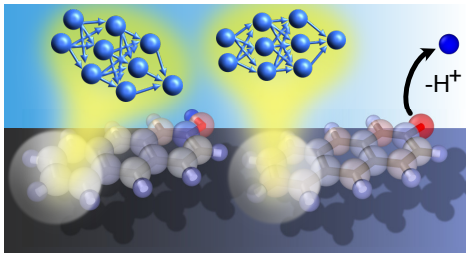


Examples:

N. Artrith, T. Morawietz and J. Behler, *Phys. Rev. B* 83 (2011) 153101.
T. Morawietz and J. Behler, *J. Chem. Phys.* 136 (2012) 064103.
K Yao et al., *Chem. Sci.* 9 (2018) 2261.
O. T. Unke and M. Meuwly, *J. Chem. Theory Comput.* 15 (2019) 3678.

24

Fourth-Generation Machine Learning Potentials: Global Charge Distribution



25

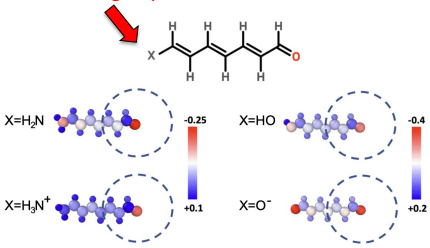
Limitations of Local Machine Learning Potentials

RUB

Effect of distance structural changes in molecular systems:

Example 1: C_7H_7OX

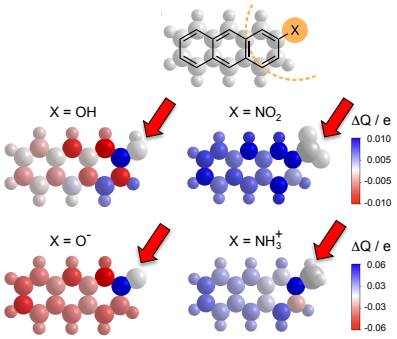
functional group



DFT Hirshfeld charges

⇒ different local properties for the same local geometry
⇒ contradictory training data

Example 2: anthracene



⇒ charge transfer determines reactivity (e.g. mesomeric effect)

T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Nature Commun. 12 (2021) 398.

Jörg Behler Grenoble 2024

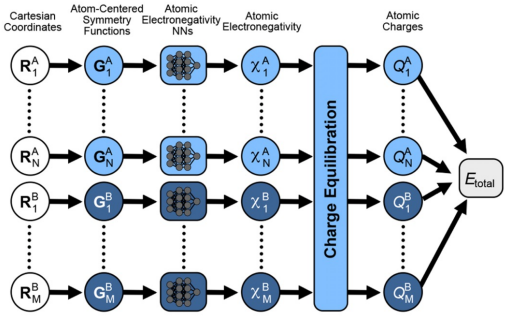
26

CENT – Charge Equilibration Neural Network Technique

RUB

$$E_{\mathrm{tot}}(\{Q_i\}) = \sum_{i=1}^{N_{\mathrm{atom}}} \left(E_i^0 + \chi_i Q_i + \frac{1}{2} J_{ii} Q_i^2 \right) + \frac{1}{2} \iint \frac{\rho(\mathbf{R})\,\rho(\mathbf{R}')}{|\mathbf{R} - \mathbf{R}'|} \mathrm{d}\mathbf{R}\,\mathrm{d}\mathbf{R}'$$

electronegativities — hardness — Coulomb energy



⇒ global electronic structure included (non-local charge transfer)

⇒ applications: systems with primarily ionic bonding

S. A. Ghasemi, A. Hofstetter, S. Saha, S. Goedecker, Phys. Rev. B **92** (2015) 045131.

27

Fourth-Generation Neural Network Potentials: Method

RUB

Goal: Combination of the advantages of CENT and HDNNPs

CENT

global charge distribution
⇒ electrostatics

2G-HDNNP

atomic energies
⇒ local bonding

modified training: atomic charge

additional atomic descriptor: atomic charge

4G-HDNNP

$E_{\mathrm{total}} = E_{\mathrm{short}} + E_{\mathrm{elec}}$

atomic energies + global charges
⇒ local bonding + electrostatics

T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Nature Commun. 12 (2021) 398.
T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Acc. Chem. Res. 54 (2021) 808.

28

Fourth-Generation Neural Network Potentials: Method

RUB

Goal: Combination of the advantages of CENT and HDNNPs

glob
⇒ e

criptor:

Advantages:

- applicable to all types of bonding and systems (ionic, covalent, metallic, ...)
- long-range electrostatic interactions (flexible charges)
- description of non-local charge transfer
- applicable to multiple global charge states

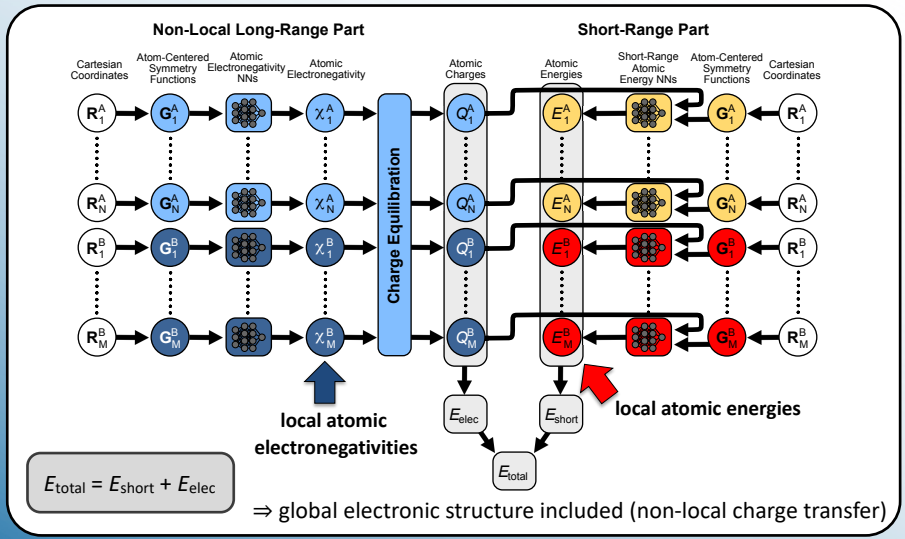
atomic energies + global charges
⇒ local bonding + electrostatics

T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Nature Commun. 12 (2021) 398.
T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Acc. Chem. Res. 54 (2021) 808.

29

Fourth-Generation Neural Network Potentials: Method

RUB



⇒ global electronic structure included (non-local charge transfer)

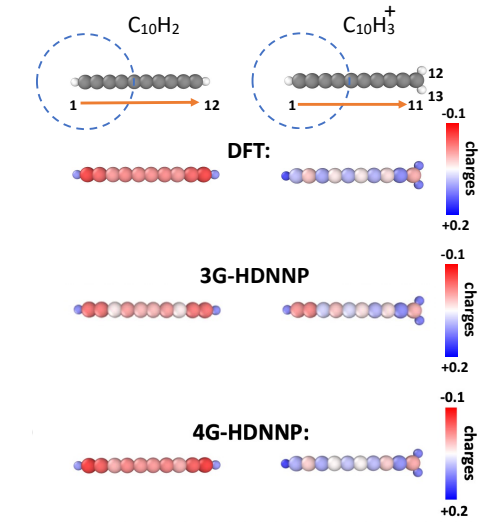
T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Nature Commun. 12 (2021) 398.
T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Acc. Chem. Res. 54 (2021) 808.

30

Fourth-Generation Neural Network Potentials: Examples

RUB

Test Case: Organic Molecules



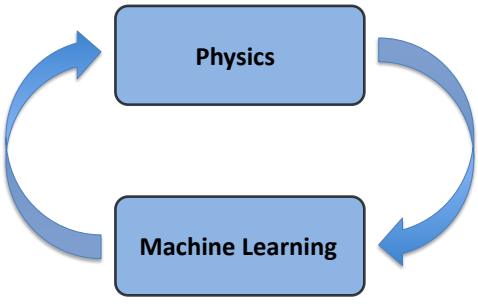
Root Mean Squared Errors

Method		*Q* (e)	*E* (meV/ atom)	*F* (eV/Å)
2G	train	---	1.6	0.131
	test	---	1.6	0.130
3G	train	0.027	3.2	0.653
	test	0.027	3.2	0.658
4G	train	0.006	1.1	0.078
	test	0.007	1.2	0.078

Jörg Behler Grenoble 2024

31

Machine Learning Potentials: Where are we now?



32

Challenge: How to measure the accuracy?

To date strong focus on *E* and *F* RMSEs: "the lower the better..."

But we need to know:

- accuracy for unknown (relevant) structures
- stability in simulations
- are the trajectories correct

We need to check the right properties

⇒ a lot of effort, requires physical knowledge about the system

What is our benchmark? What is the truth?

- Theory? Experiment?

Coupled cluster calculations for large systems are impossible!
⇒ Often we do not know the truth

Good news: In DFT we have accepted this dilemma and it works!

33

What to do?

- **The MLP must have a low RMSE for *E* and *F***

  Mandatory but not a sufficient criterion
  ⇒ large remaining errors indicate problems (data, descriptor, physics...)

- **Molecular dynamics must be long-term stable**

  But: long-term stability is no proof for the right physics (*"Seeing is believing"?*)
  "Well, you can run it..."

- **Understand the physics of the system before training a potential**

  ⇒ check the critical properties

  Comparison to experiment is often difficult

Validation of MLPs is become a main task

34

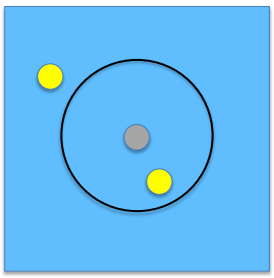
Example:
Multiple Charge States

35

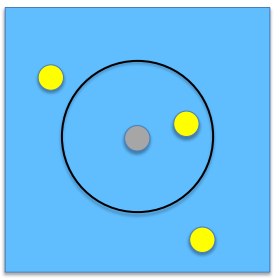
The Fe^{2+}/Fe^{3+} redox system in water

RUB

$FeCl_2$ in water



$FeCl_3$ in water



Expectation:

- A local second-generation MLP must fail for this combined system
- A global fourth-generation MLP should work for this system

How can we investigate/validate this?

36

The Fe^{2+}/Fe^{3+} redox system in water

RUB

Construction of 6 HDNNPs

HDNNP	Dataset	*E* RMSE (meV/atom)	*F* RMSE (eV/Å)	*Q* RMSE (me)
2G(Fe^{2+})	$FeCl_2$/water	0.203 (0.213)	0.032 (0.032)	---
2G(Fe^{3+})	$FeCl_3$/water	0.232 (0.245)	0.037 (0.037)	---
2G(Fe^{2}+/Fe^{3+})	all	0.262 (0.271)	0.034 (0.035)	---
4G(Fe^{2+})	$FeCl_2$/water	0.203 (0.258)	0.031(0.036)	2.896 (2.896)
4G(Fe^{3+})	$FeCl_3$/water	0.237 (0.258)	0.036 (0.036)	3.130 (3.124)
4G(Fe^{2+}/Fe^{3+})	all	0.258 (0.273)	0.033 (0.033)	3.105 (3.108)

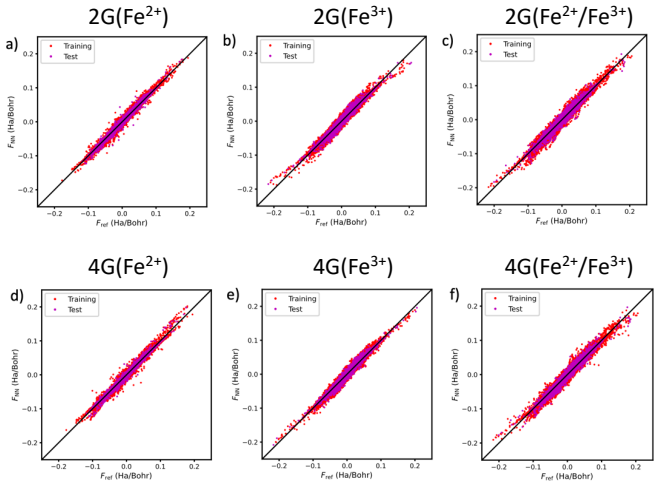
⇒ no significant difference, all very accurate (?)

37

The Fe^{2+}/Fe^{3+} redox system in water

RUB

What about outliers? Force correlation plots

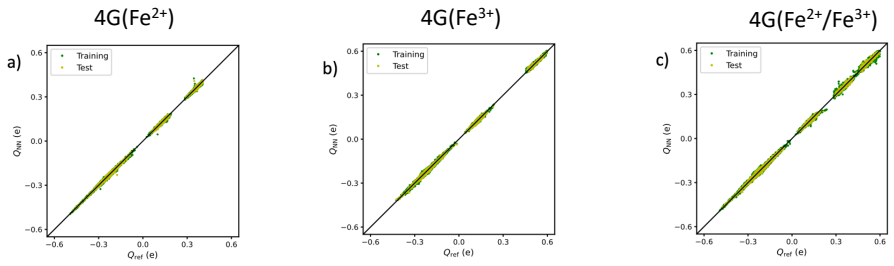


38

The Fe^{2+}/Fe^{3+} redox system in water

RUB

4G-HDNNP Charge Correlation



⇒ Charges of both oxidation states are well represented

39

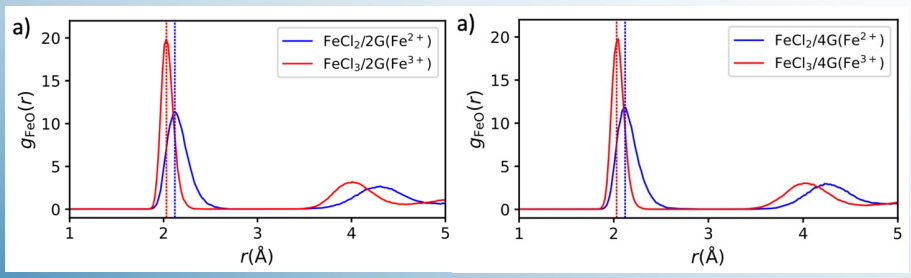
The Fe^{2+}/Fe^{3+} redox system in water

RUB

Fe-O Radial Distribution Functions:

Test 1: HDNNPs trained only to $FeCl_2$ *OR* $FeCl_3$ (15 Å box)

2G-HDNNP **4G-HDNNP**



⇒ All trajectories are stable

⇒ For separate datasets all HDNNPs work fine

40

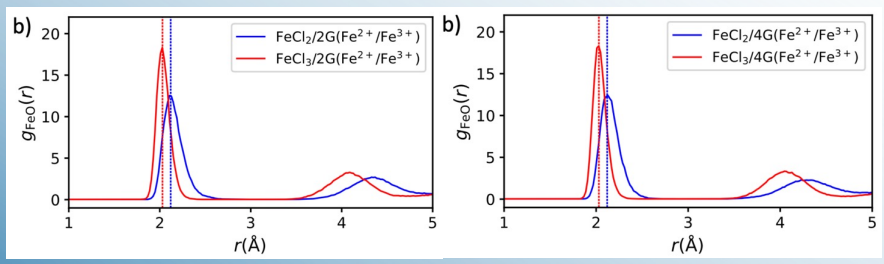
The Fe^{2+}/Fe^{3+} redox system in water

RUB

Fe-O Radial Distribution Functions:

Test 2: HDNNPs trained to $FeCl_2$ *AND* $FeCl_3$ (15 Å box)

2G-HDNNP **4G-HDNNP**



⇒ All trajectories are stable
⇒ For combined datasets all HDNNPs *SEEM* to work fine

41

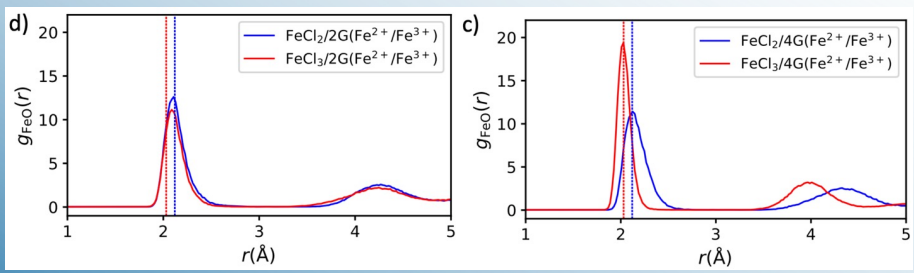
The Fe^{2+}/Fe^{3+} redox system in water

RUB

Fe-O Radial Distribution Functions:

Test 3: HDNNPs trained to $FeCl_2$ *AND* $FeCl_3$ (30 Å box)

2G-HDNNP **4G-HDNNP**



⇒ 2G-HDNNPs fail, 4G-HDNNPs are correct

42

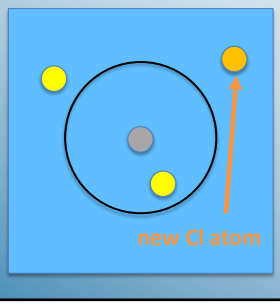
The Fe^{2+}/Fe^{3+} redox system in water

RUB

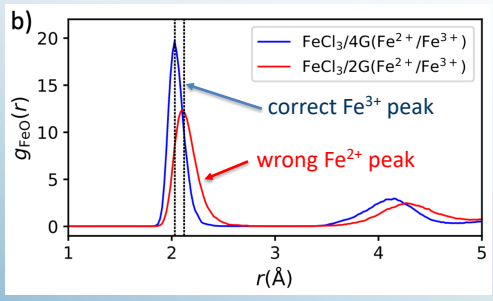
Fe-O Radial Distribution Functions:

Test 4: HDNNPs trained to $FeCl_2$ *AND* $FeCl_3$ – Addition of Cl atom

Initial structure:
$FeCl_2$ in water



Newly equilibrated system:



⇒ 2G-HDNNPs fail, 4G-HDNNPs are correct

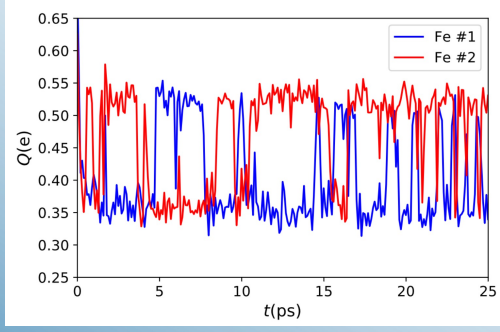
43

The Fe^{2+}/Fe^{3+} redox system in water

RUB

Extrapolation Test: Electron Transfer Reaction

Test: Fe_2Cl_5 in a large box of water



⇒ 4G-HDNNP can predict electron transfer and charge conservation

44

Conclusions

45

Conclusions

RUB

Physics remains important	⇒	Right model for the system
Real applications are important	⇒	We must go beyond model toy systems
ML not good for extrapolation	⇒	Range of validity is essential information
Reference methods are important	⇒	MLP error is smaller than DFT error
Reporting RMSEs is not enough	⇒	Stability in simulations is essential

⇒ Using Machine Learning Potentials requires physical understanding and validation

⇒ There is still a lot of room for further methodical developments

46