Neural network atomistic simulations for aqueous systems

Matti Hellström

Group of Prof. Dr. Jörg Behler
Theoretische Chemie
Institut für Physikalische Chemie
Georg-August-Universität Göttingen
D-37077 Göttingen, Germany

matti.hellstroem@chemie.uni-goettingen.de

- **Machine Learning, Artificial Intelligence, Big Data**
- Image recognition
- Speech recognition
- Self-driving vehicles
- Recommender systems (Amazon, Youtube, …)
- Game-playing
- Business intelligence
- Computational chemistry

Machine learning and “Big Data”

1959: "Field of study that gives computers the ability to learn without being explicitly programmed"

Arthur Lee Samuel (1901-1990),
Pioneer in artificial intelligence

1997: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E"

Tom Mitchell, "Machine Learning",
McGraw Hill 1997.

- Linear regression
- Logistic regression
- Kernel ridge regression
- Support vector machines
- **Neural networks**

- QSAR, “quantitative structure-activity relationship”
  - Property (e.g. melting point, solubility, enzyme activity) directly predicted from structure
- **Machine learning potentials**
  - Computationally inexpensive molecular simulations with first-principles quality
    - Molecular dynamics and Monte Carlo

Surface Science:

- heterogeneous catalysis
- surface reconstructions, defects
- self-assembly, functional materials

(Interfacial) Water:

- reactions involving proton transfer
- assembly/catalysis at interfaces
- electrochemistry

Materials:

- crystal structure prediction
- phase diagrams
- properties of materials

Organic-Inorganic Systems:

- metal organic frameworks
- biomineralization
- coordination chemistry

Surface Science:

- heterogeneous catalysis
- surface reconstructions, defects
- self-assembly, functional materials

(Interfacial) Water:

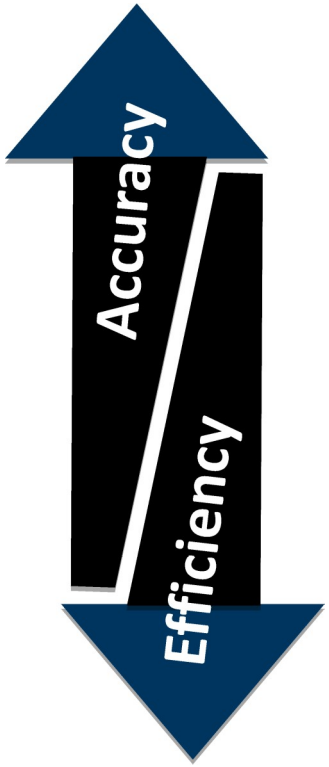
- reactions involving proton transfer
- assembly/catalysis at interfaces
- electrochemistry

Complex Bonding and Large Systems
⇒ Challenging for Theory

- crystal structure prediction
- phase diagrams
- properties of materials

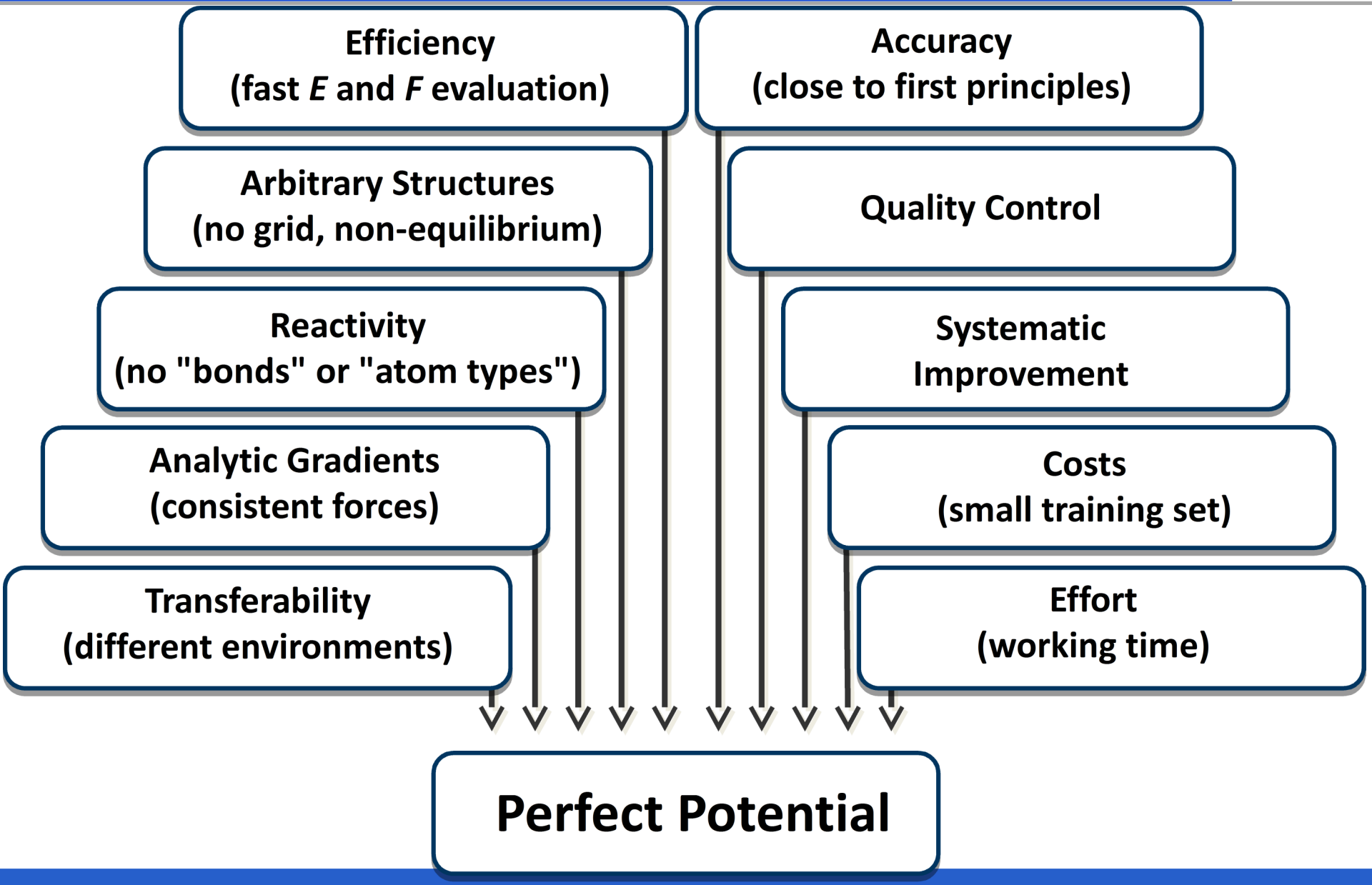
- metal organic frameworks
- biomineralization
- coordination chemistry

Potentials for Different Length and Time Scales

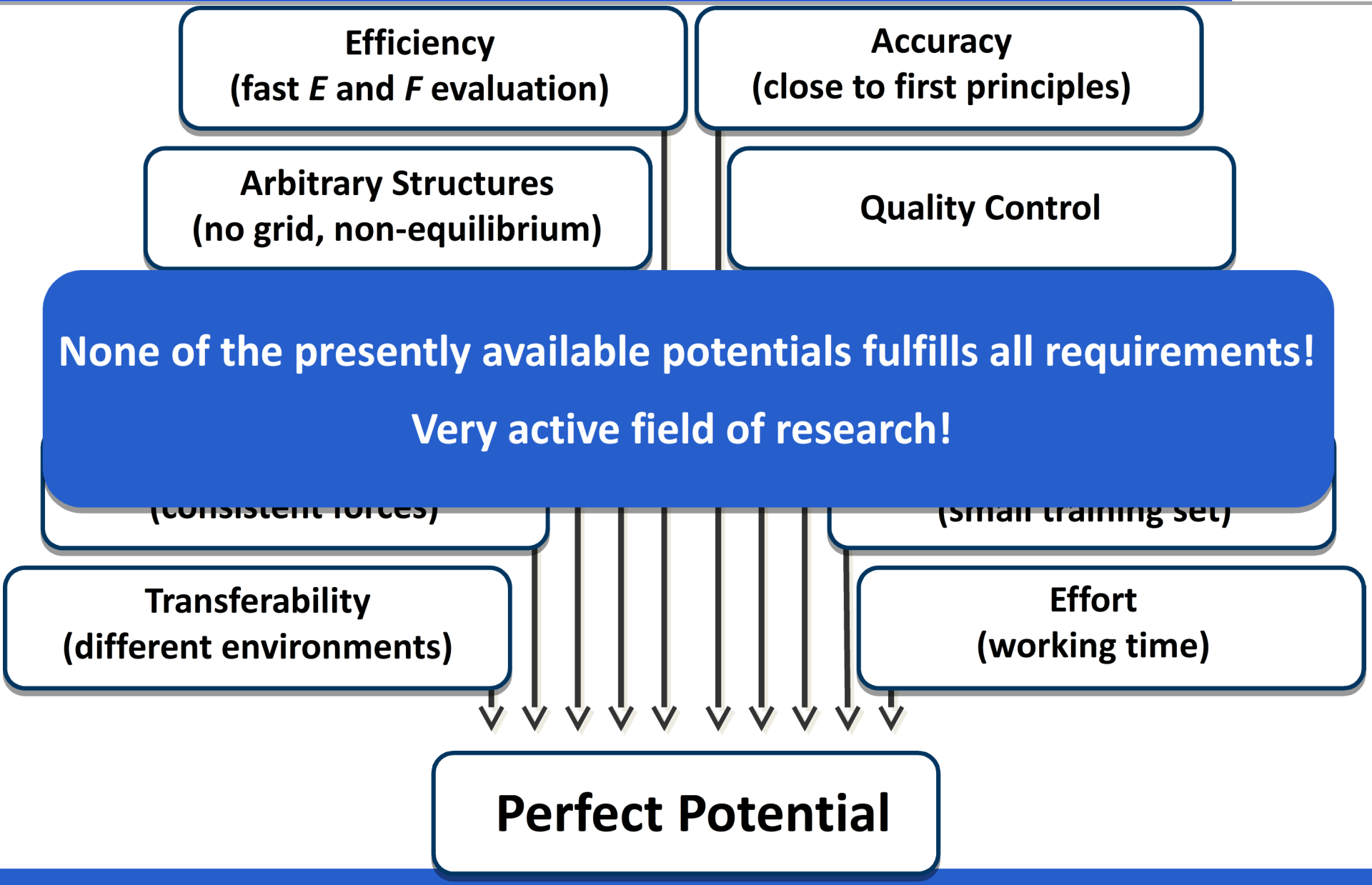


Method	Atoms	Simulation Time
CI, CC, MP2, MP4, …	10 - 100	0
Density Functional Theory	100 - 1000	100 ps
Semiempirical Methods Tight Binding	1000 - 10 000	1 ns
„Reactive Potentials“ EAM, Tersoff, …	1000 - 100 000	10 ns
Classical Force Fields (LJ, harmonic, Coulomb)	1 000 000	1 ms

Challenges for the Construction of Atomistic Potentials



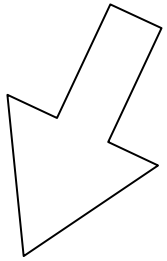
Challenges for the Construction of Atomistic Potentials



Atomistic Potentials:

Direct functional relation between the structure and the energy (and forces)

conventional approach: approximations



„Physical Potentials“ (> 99 %)

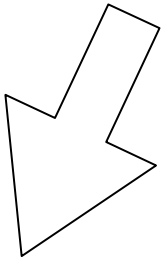
Simplification

AMBER, CHARMM, ReaxFF, ...

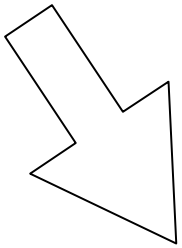
Atomistic Potentials:

Direct functional relation between
the structure and the energy (and forces)

conventional approach:
approximations



fitting



„Physical Potentials“
(> 99 %)

„Machine Learning“ or
„Mathematical Potentials“ (< 1 %)

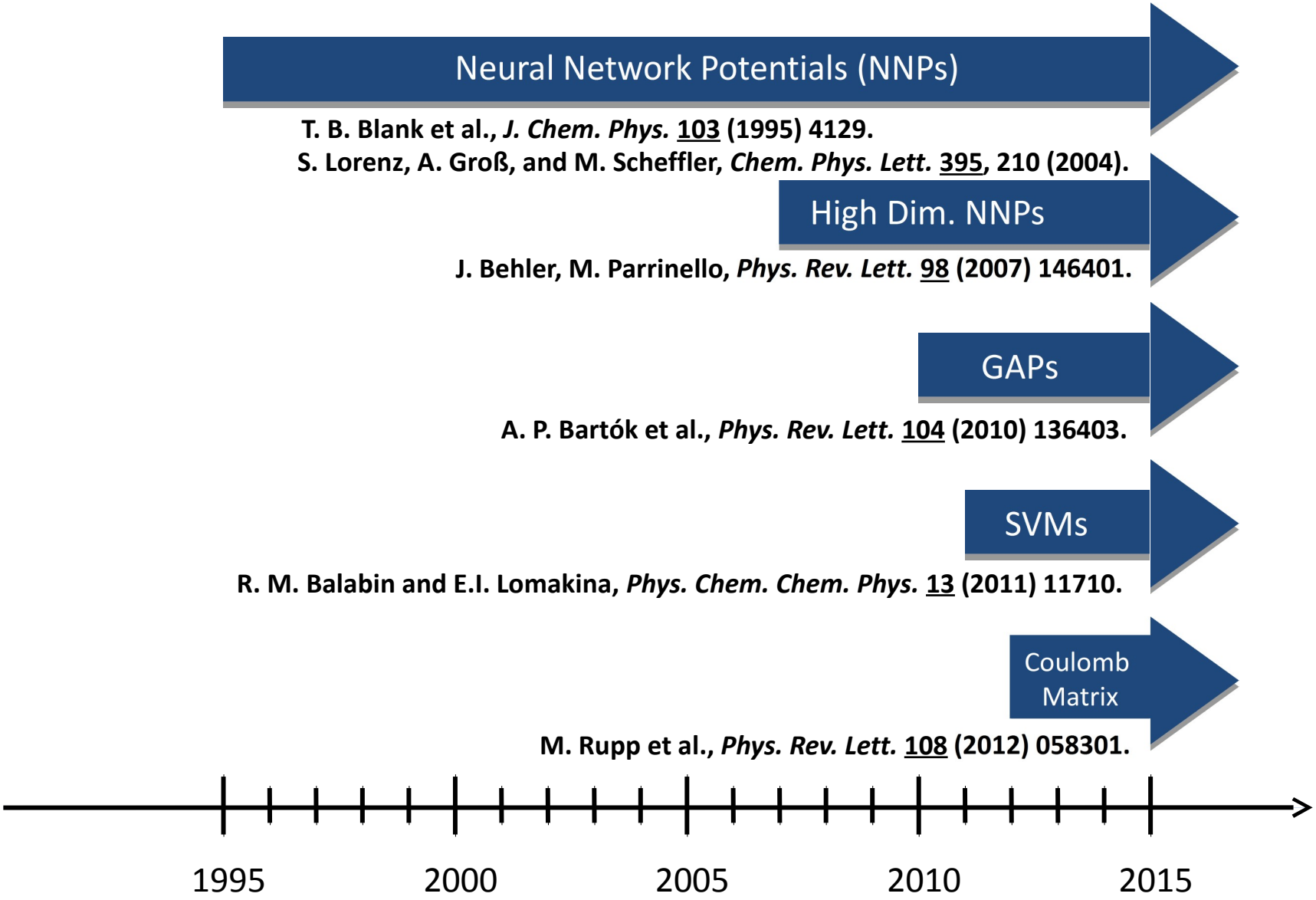
Simplification

AMBER, CHARMM, ReaxFF, ...

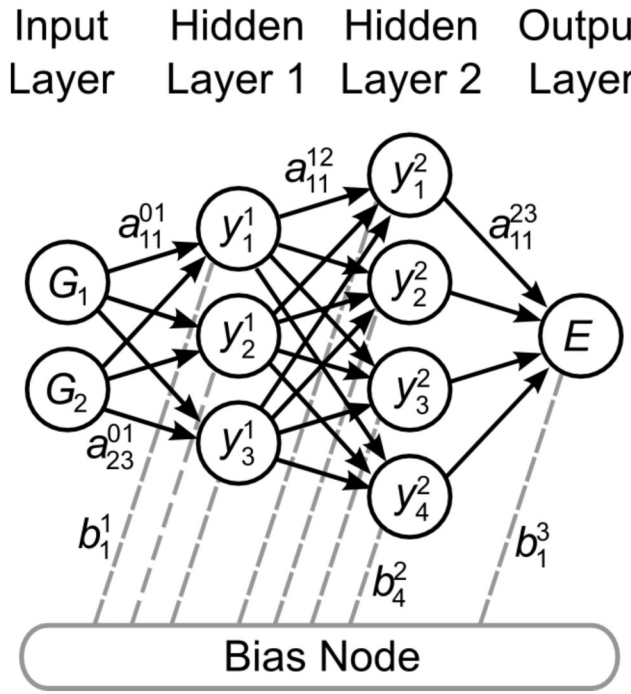
Representation

Neural networks,
Gaussian approximation potentials,
kernel ridge regression, ...

Machine Learning ("Mathematical") Potentials

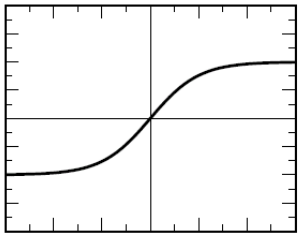


Conventional Neural Network Potentials: Analytic Form



Neural Network "*learns*"
energy surface from reference electronic structure calculations
⇒ iterative optimization

Nonlinear Activation Functions:
⇒ high flexibility

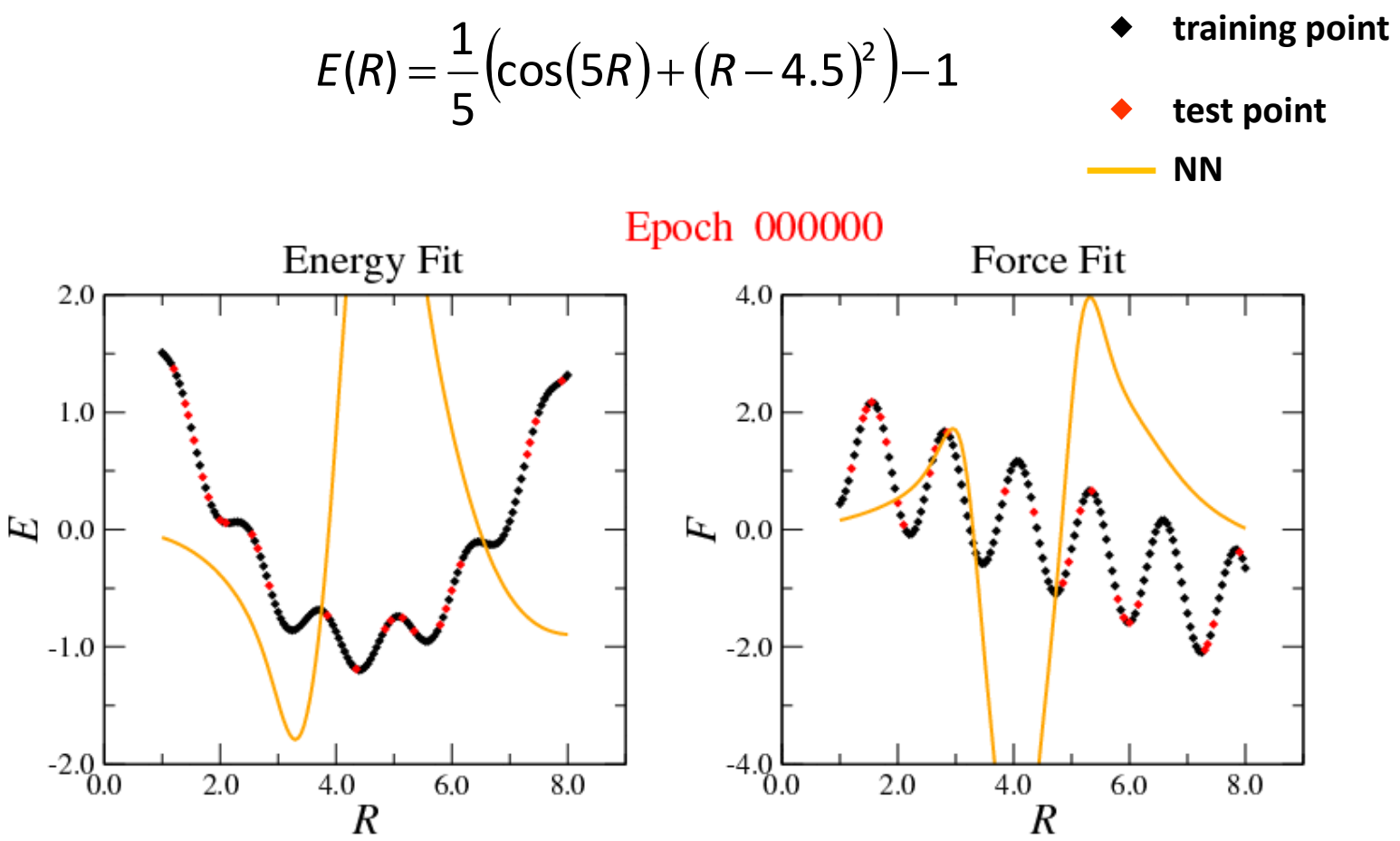


Analytic Form:

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

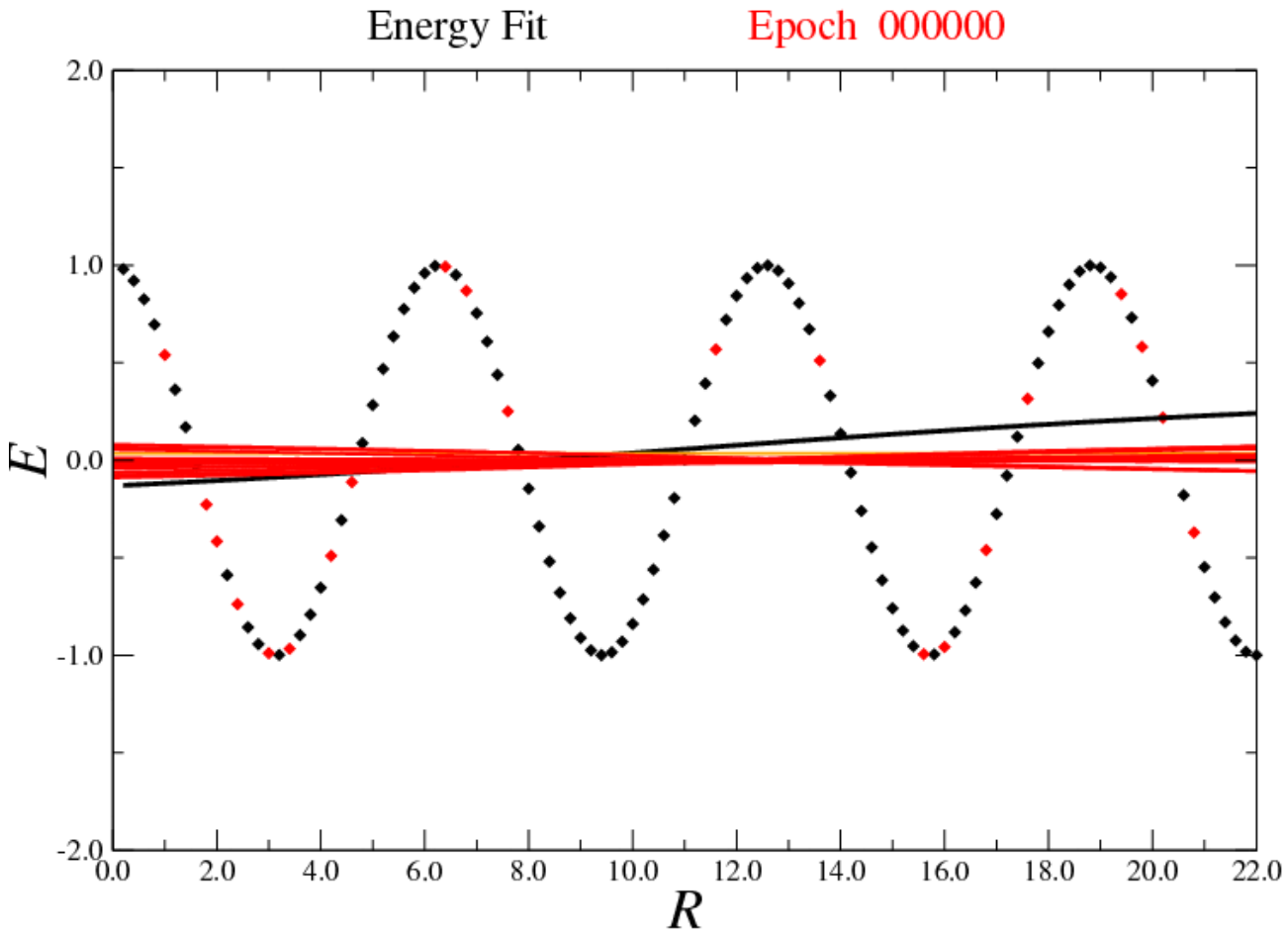
Representation of a 1-dimensional model potential function:

$$E(R) = \frac{1}{5}\left(\cos(5R) + (R-4.5)^2\right) - 1$$



NN based on initial random weight parameters!

Fitting of a 1-Dimensional Model Function: cos(x)



bias weight

individual nodes

NN fit

training point

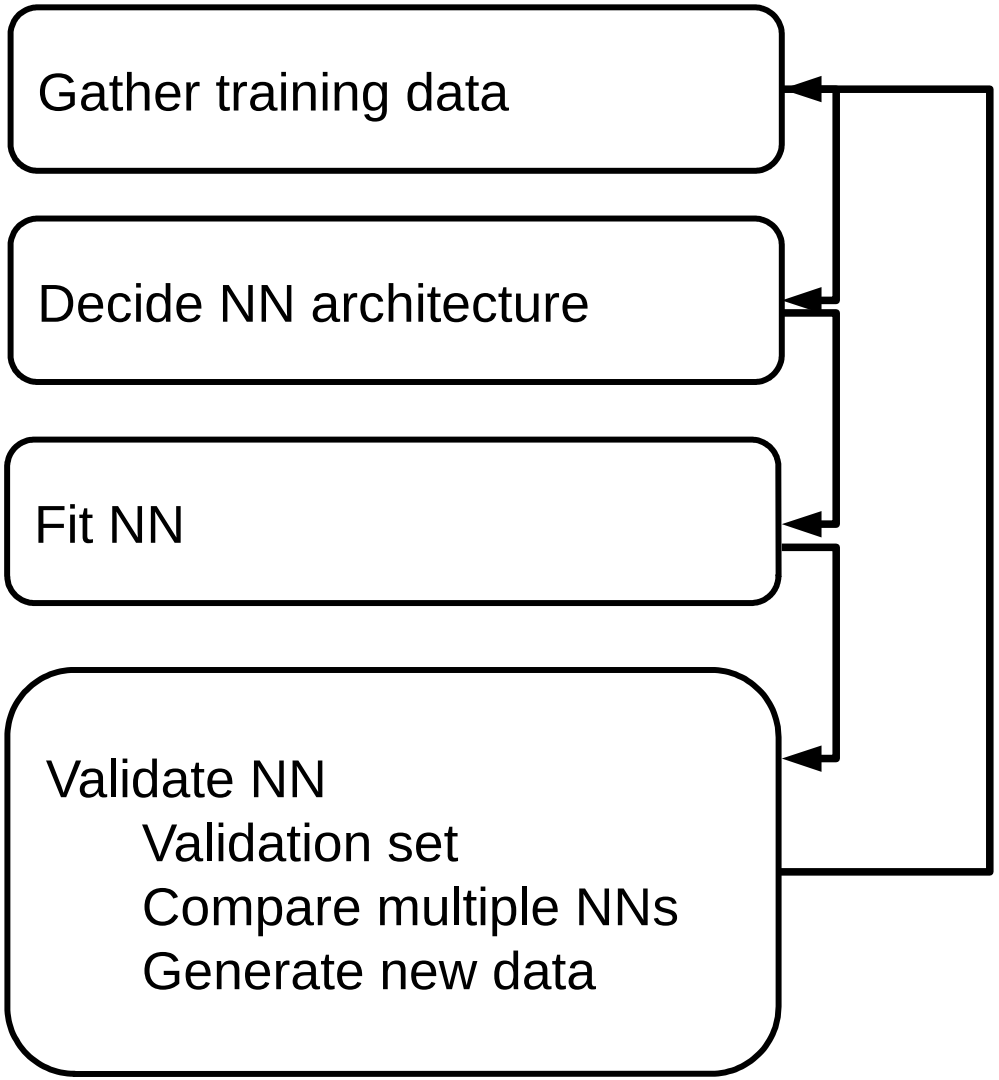
test point

NN architecture: 1-14-1 tl

Benefits of machine learning potentials

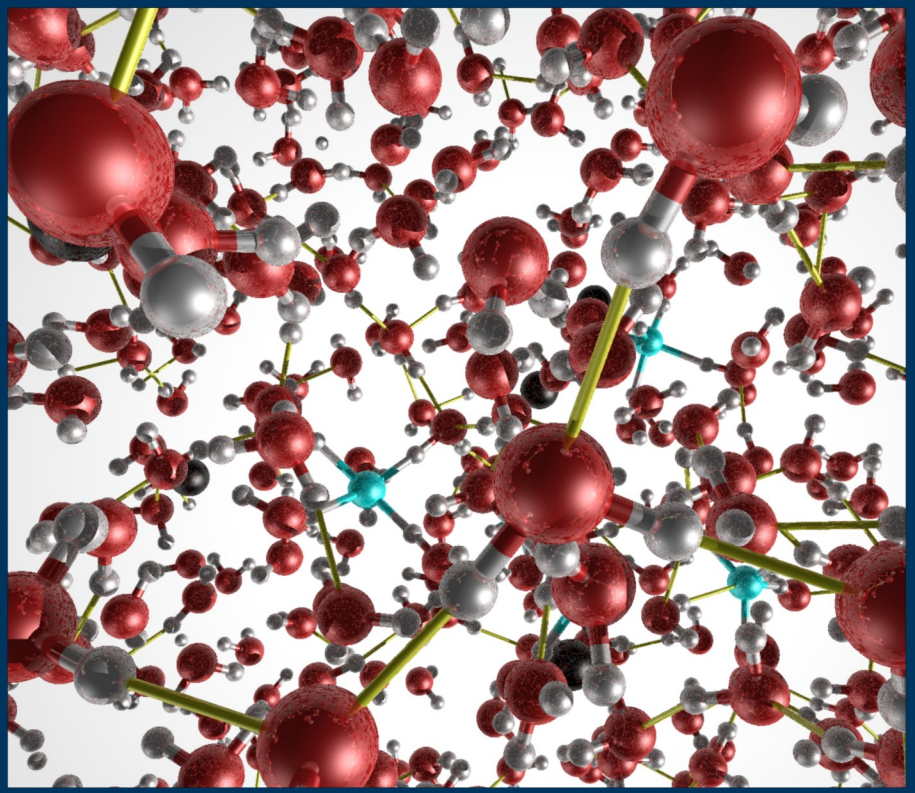
- **Any system, any type of bonding**
  (ionic, covalent, hydrogen-bonding, dispersion interactions, …)
- **Remarkably accurate**
- **Computationally inexpensive**
  Large systems and long timescales possible
- **Only atomic positions needed**
  No predefined molecules, neural network is “reactive”
- **Many reference methods possible**
- **Systematic improvement possible**

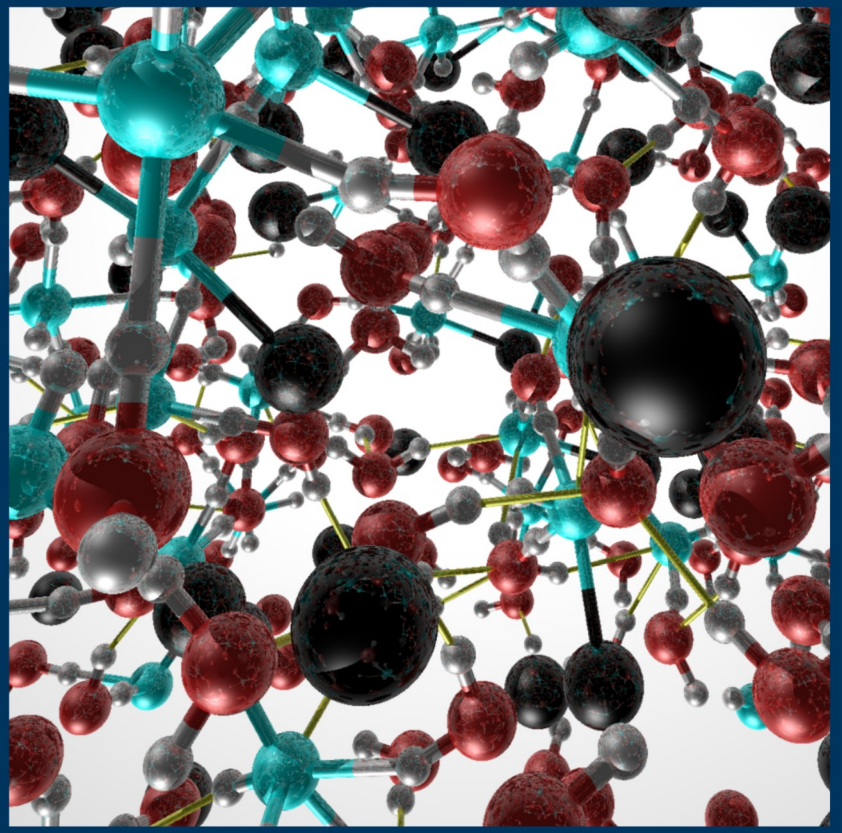
J. Behler, *Phys. Chem. Chem. Phys.* **13**, 17930 (2011).



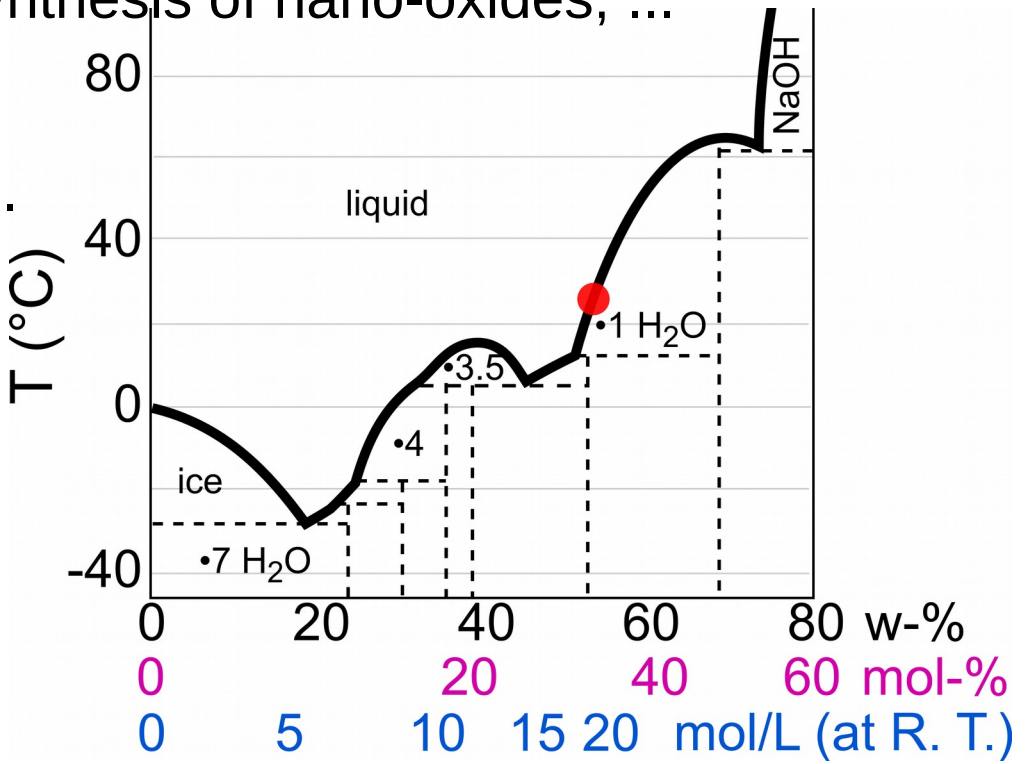
Iterative fitting procedure

NaOH Solutions



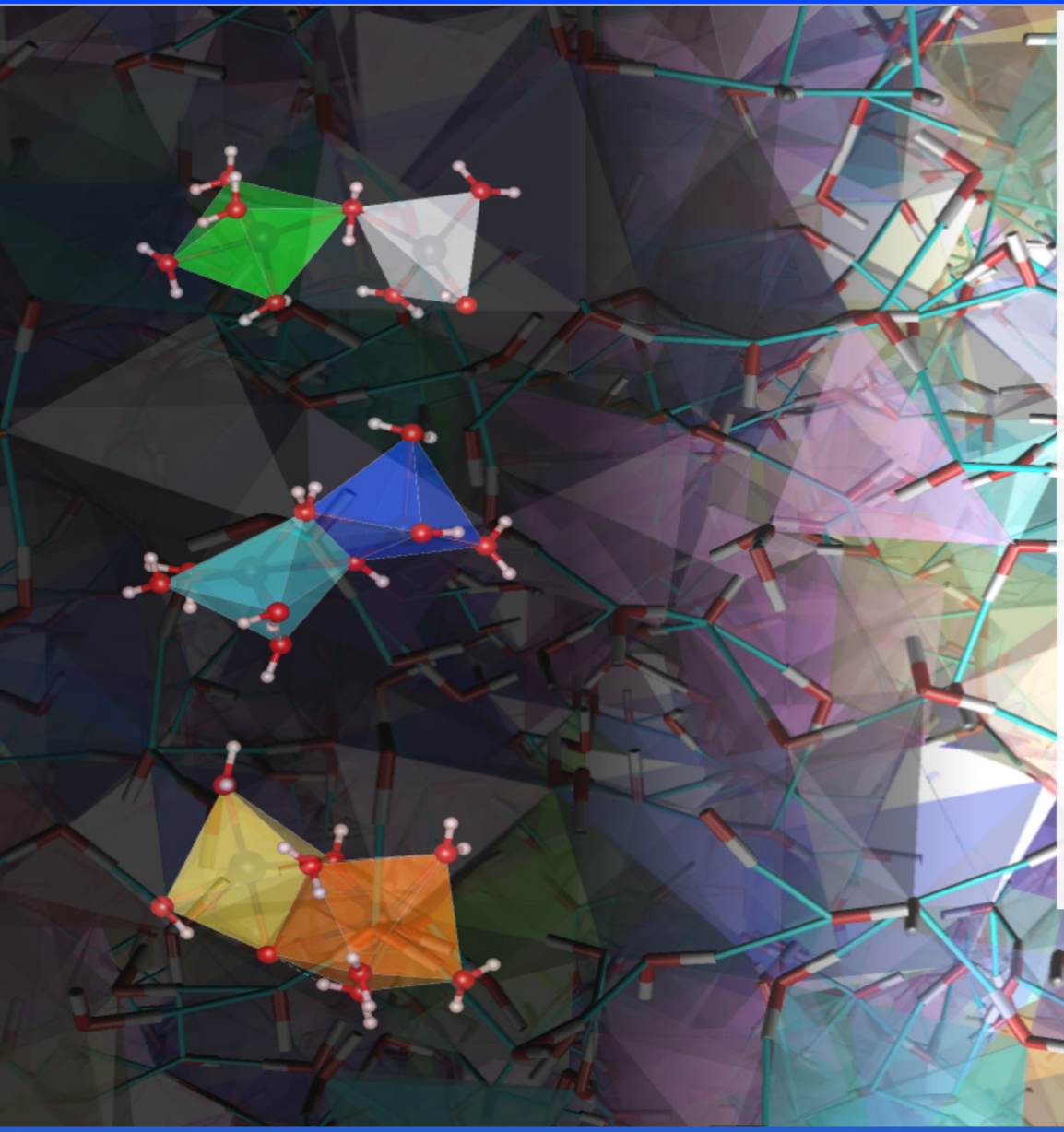


- Commodity chemical, annual production 60 million tonnes
- Chloralkali process: $2\ NaCl + 2\ H_2O \rightarrow Cl_2 + H_2 + 2\ NaOH$
- Paper and textile industries, synthesis of nano-oxides, ...
- *Very* soluble in water:
  $NaOH:H_2O$ ~ 1:2 at room temp.
  1 H_2O per ion!
  19 mol/L



From "Caustic soda solution handbook", The Dow Chemical Company, 2010

NaOH solutions (in water)



- Hydrogen bond fluctuations drive proton transfer
- Coordination polyhedra around Na^+ at different concentrations
- Ion pairing
- Nuclear quantum effects
- Proton-transfer-driven water exchange mechanism in the Na^+ solvation shell

Challenges for machine learning potentials

- **Descriptor: what does the ML method "see"?**
  - Atom-centered symmetry functions (ACSFs)
  - Smooth overlap of atomic positions (SOAP)
- **Gathering/generating (useful) training data**
  - Random, molecular dynamics, Monte Carlo, …
  - Ensuring that the whole PES is covered (avoiding "extrapolations")
- **Fitting thousands of parameters**
  - Algorithms (gradient descent, Adam, Kalman filter, …)
  - Computational cost, parallelization
  - Hyperparameters: learning rate, NN architecture, …
  - Overfitting: poor generelization
  - Underfitting: poor model!

Conclusions

Summary: Neural Network Potentials

General machine learning

QSAR

Neural network (and other machine learning) potentials

- Universal (not system-specific) functional form
- Accurate and fast
- Molecular dynamics and Monte Carlo

Reviews:

J. Behler. *Phys. Chem. Chem. Phys.* **13** (2011) 17930
J. Behler. *Angew. Chem. Int. Edit.* (2017)
M. Hellström, J. Behler. *Handbook of Materials Modeling* (Springer, 2018)

Research applications (NaOH and ZnO):

M. Hellström, J. Behler. *J. Phys. Chem. Lett.* **7** (2016) 3302
V. Quaranta, M. Hellström, J. Behler, *J. Phys. Chem. Lett.* **8** (2017) 1476
M. Hellström, J. Behler. *Phys. Chem. Chem. Phys.* **19** (2017) 82
M. Hellström, J. Behler. *J. Phys. Chem. B* **121** (2017) 4184