TSINGHUA-PRINCETON-COMBUSTION INSTITUTE

2026 SUMMER SCHOOL ON COMBUSTION

Quantum Mechanics, Statistical Mechanics, and Machine Learning for Molecular Simulations

Alexandre Tkatchenko

University of Luxembourg, Luxembourg

July 06-07, 2026

TSINGHUA-PRINCETON-COMBUSTION INSTITUTE

2026 SUMMER SCHOOL ON COMBUSTION

Key Activities / 重要活动			
July 5 (Sunday) /7 月 5 日 （周日）	10:00-17:30	Registration 注册	Northeast Gate, Lee Shau Kee Sci. and Tech. Building 李兆基科技大楼东北门
	18:30	Welcome Reception 开班仪式	A-278, Multifunction Room, Lee Shau Kee Sci. and Tech. Building 李兆基科技大楼多功能厅
Class Schedule / 课程安排			
Schedule 时间	Location 地点	Morning 上午 (9:00-9:50/10:00-10:50/11:00-11:50)	Afternoon 下午 (14:00-14:50/15:00-15:50/16:00-16:50)
July 6-10 (Monday-Friday) /7 月 6-10 日 （周一至周五）	Zone A, 6th Teaching Building 第六教学楼 A 区	**Theoretical and Numerical Combustion** **classroom: 6A018 (0th floor)** Thierry Poinsot	**Combustion Chemistry: From Fundamentals to Kinetic Modelling for Low-Carbon Technologies** **classroom: 6A018 (0th floor)** Alison Tomlin
July 6-7 (Monday-Tuesday) /7 月 6-7 日 （周一至周二）	Zone A, 6th Teaching Building 第六教学楼 A 区	**Spectroscopic Diagnostics for Combustion Chemistry** **classroom: 6A203 (2nd floor)** Pascale Desgroux	**Quantum Mechanics, Statistical Mechanics,and Machine Learning for Molecular Simulations** **classroom: 6A203 (2nd floor)** Alexandre Tkatchenko
July 8-10 (Wednesday-Friday) /7 月 8-10 日 （周三至周五）	Zone A, 6th Teaching Building 第六教学楼 A 区	**Introduction to Plasma-assisted Combustion** **classroom: 6A203 (2nd floor)** Deanna Lacoste	**AI for Combustion** **classroom: 6A016 (0th floor)** Matthias Ihme
Special Activities / 特殊活动			
July 5 (Sunday) /7 月 5 日 （周日）	13:30-17:30	Art Museum Visit / 艺术博物馆参观	Tsinghua University Art Museum 清华大学艺术博物馆
July 7 (Tuesday) /7 月 7 日 （周二）	17:00-18:00	Campus Tour / 校园游览	Tsinghua University 清华大学

July 8 (Wednesday) /7月8日（周三）	17:00-17:30	Group Picture Taking / 暑期学校合影	Mong Man Wai Concert Hall 蒙民伟音乐厅
July 8 (Wednesday) /7月8日（周三）	18:30-19:30 19:30-21:00	Poster Presentation / 海报展示 Career Panel / 职业发展论坛	B-518, Lee Shau Kee Sci. and Tech. Building 李兆基科技大楼B-518会议室
July 9 (Thursday) /7月9日（周四）	18:00	Farewell Reception / 欢送会	Guan Chou Yuan Restaurant 观畴园餐厅
July 10 (Friday) /7月10日（周五）	8:00-18:00	Program Certificate Distribution / 学习证书发放	6th Teaching Building 第六教学楼
July 11 (Saturday) /7月11日（周六）	9:30-11:30	CCE Laboratory Tour / 燃烧能源中心实验室参观	Northeast Gate, Lee Shau Kee Sci. and Tech. Building 李兆基科技大楼东北门

Electronic version of all lecture materials are available at the summer school website

FITMOL

Machine Learning in Chemistry and Physics

Alexandre Tkatchenko *et al.*

Chair for Theoretical Chemical Physics (TCP), *Dept of Physics and Materials Science* (DPhyMS), *University of Luxembourg*

www.tcpunilu.com
alexandre.tkatchenko@uni.lu

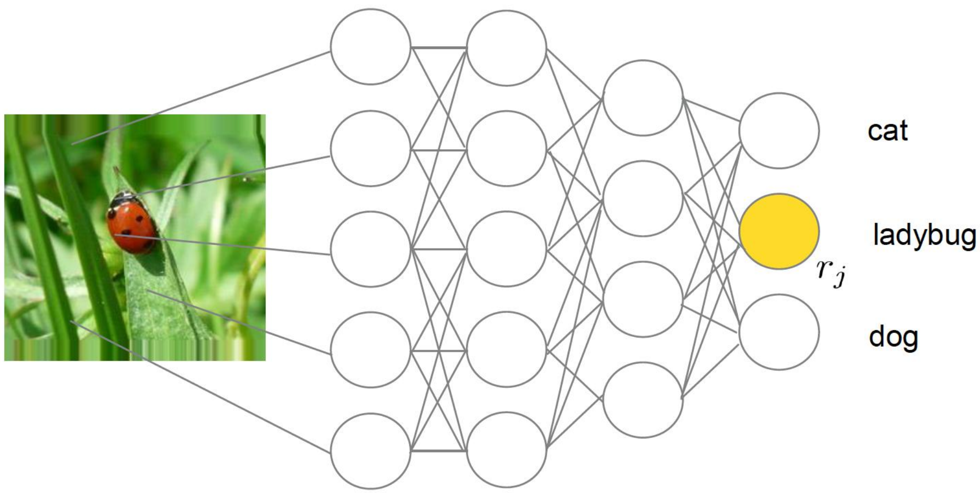
Tsinghua University, Lecture 1

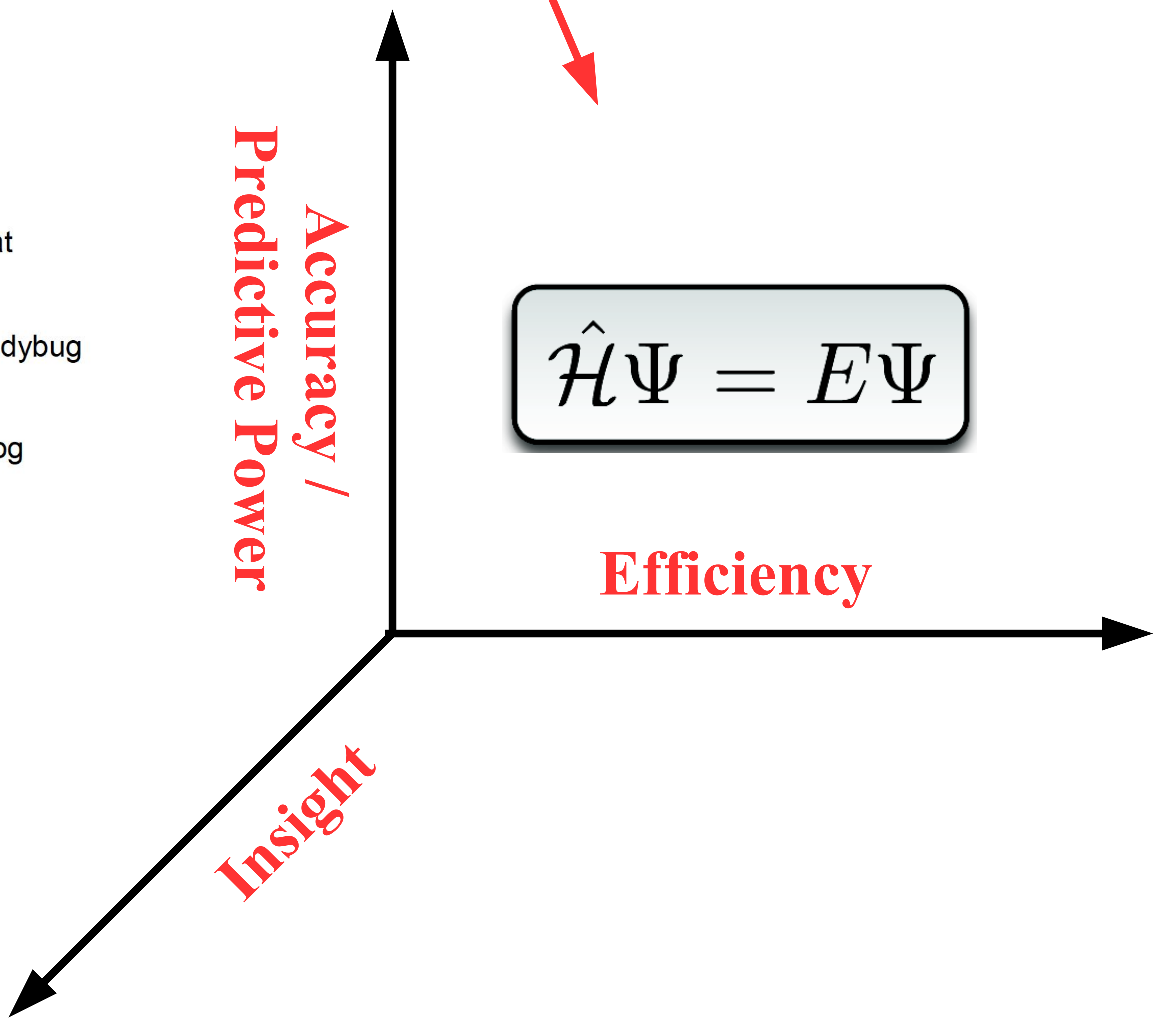
Comment | Published: 31 May 2021

Best practices in machine learning for chemistry

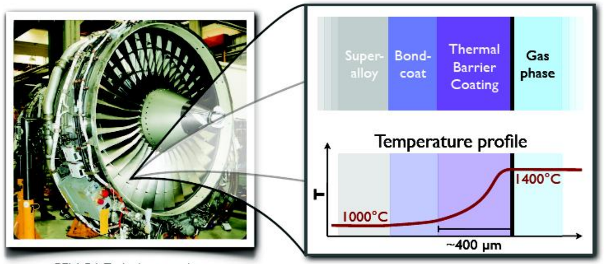
Nongnuch Artrith, Keith T. Butler, François-Xavier Coudert, Seungwu Han, Olexandr Isayev, Anubhav Jain & Aron Walsh

Machine Learning (ML) and Natural Sciences





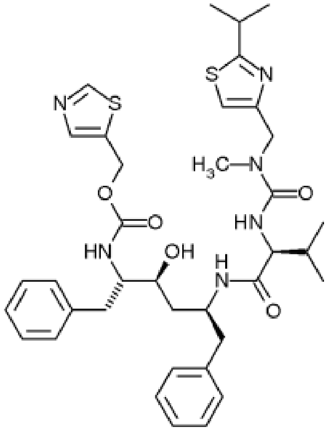
Physics (+ Chemistry and Biology)

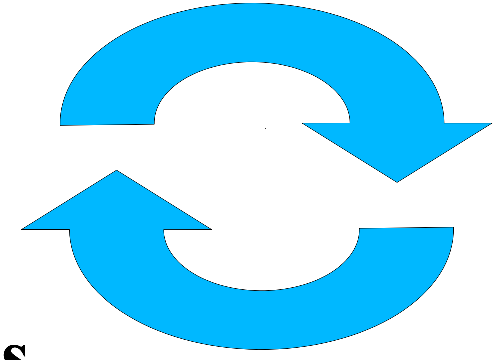


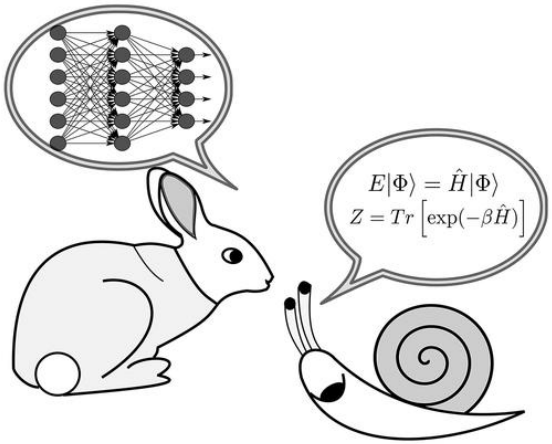
CFM 56-7 airplane engine

$$\hat{\mathcal{H}}\Psi = E\Psi$$

Quantum Mechanics







Statistical Mechanics

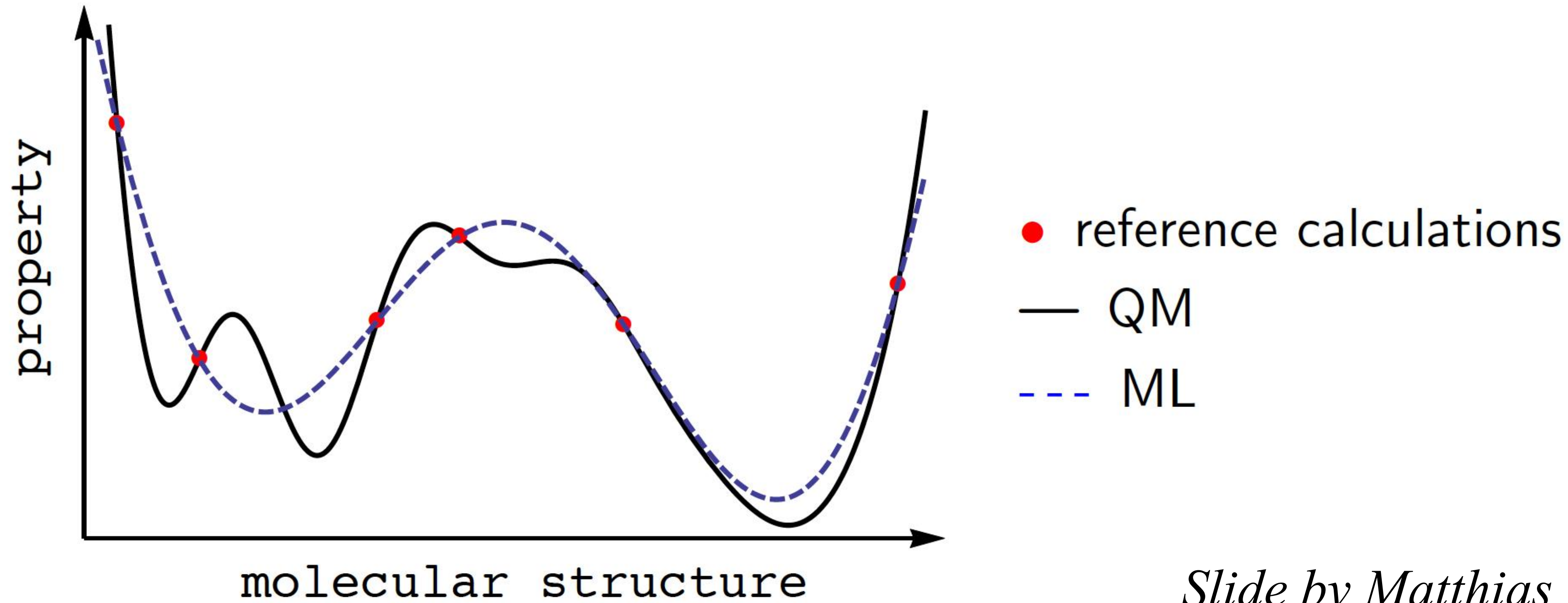
Machine Learning

$$Z = \mathrm{Tr}(e^{-\beta\hat{H}})$$

Quantum Mechanics / ML models

Exploit redundancy in a series of QM calculations

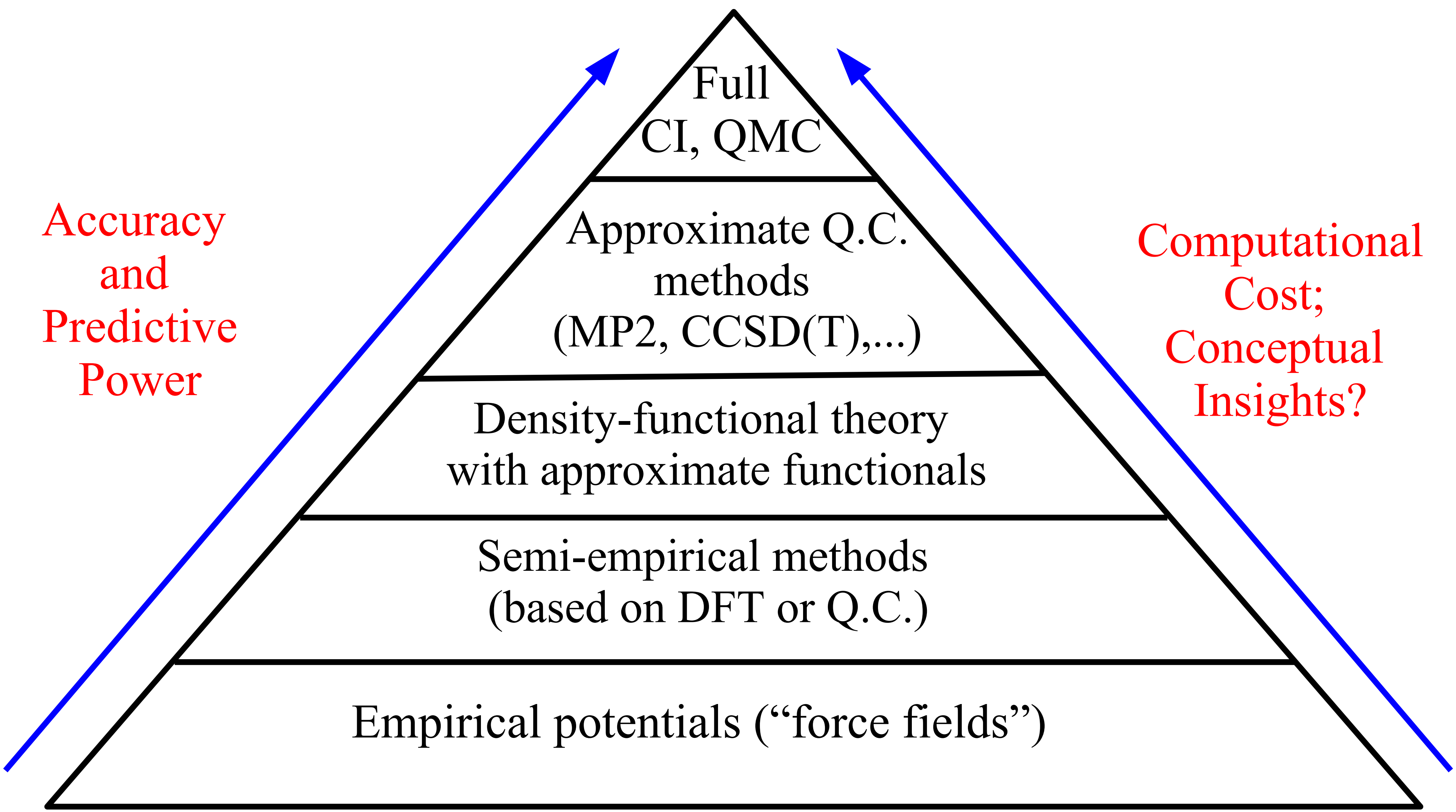
- QM/ML = quantum mechanics + machine learning
- Interpolate between QM calculations using ML
- Smoothness assumption (regularization)



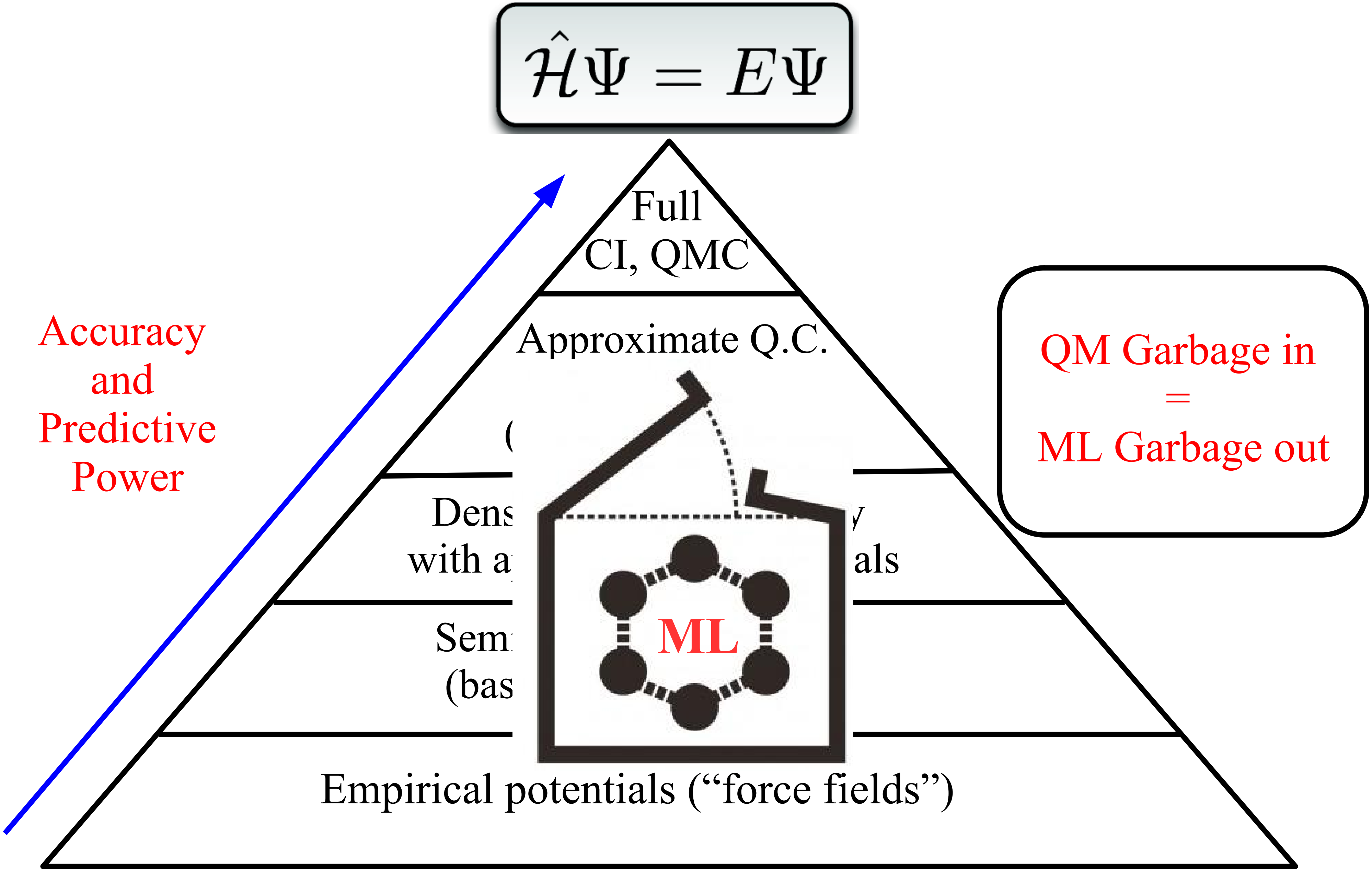
Slide by Matthias Rupp

Current state-of-the-art of QM methods

$$\hat{\mathcal{H}}\Psi = E\Psi$$

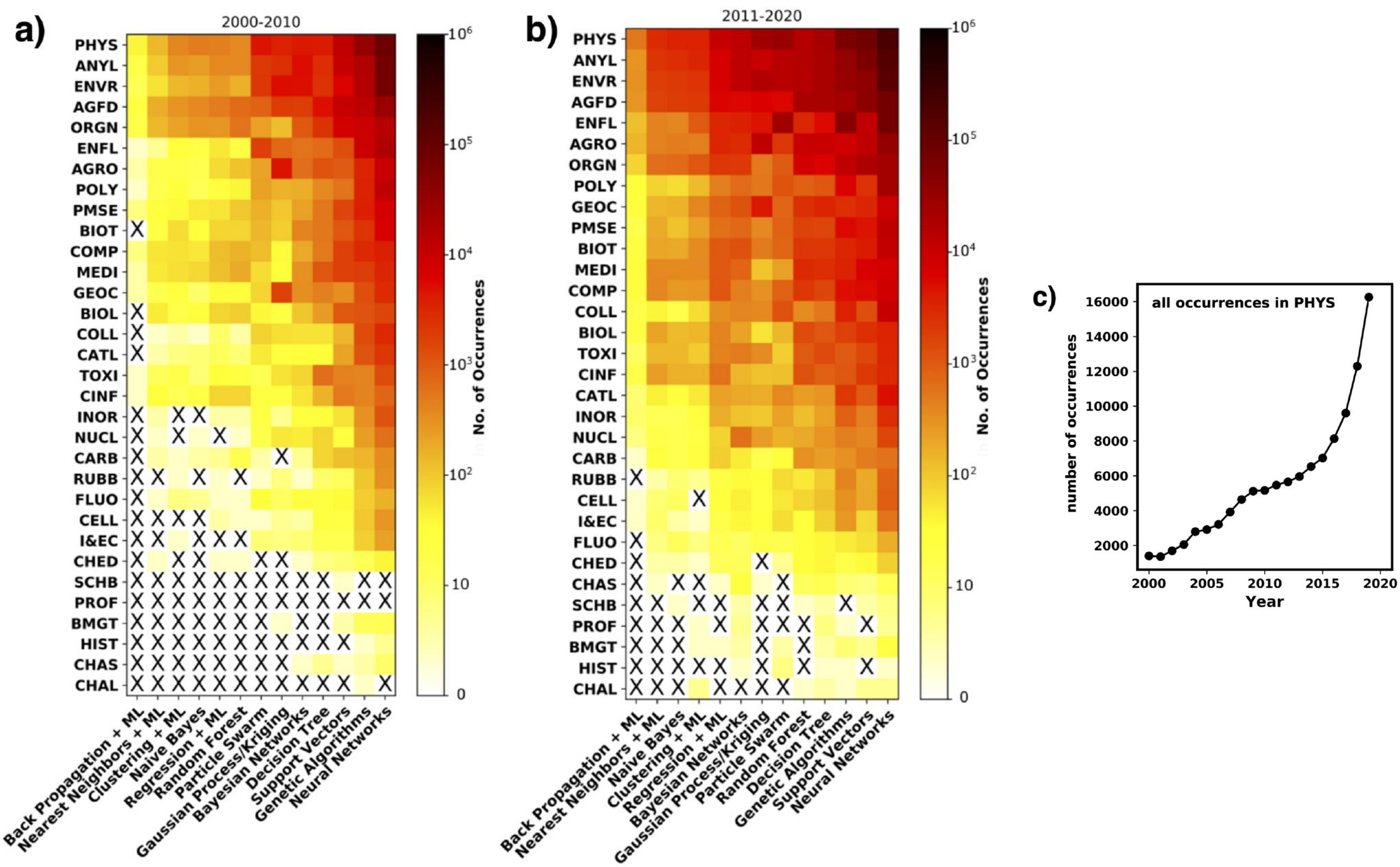


Current state-of-the-art of QM/ML methods



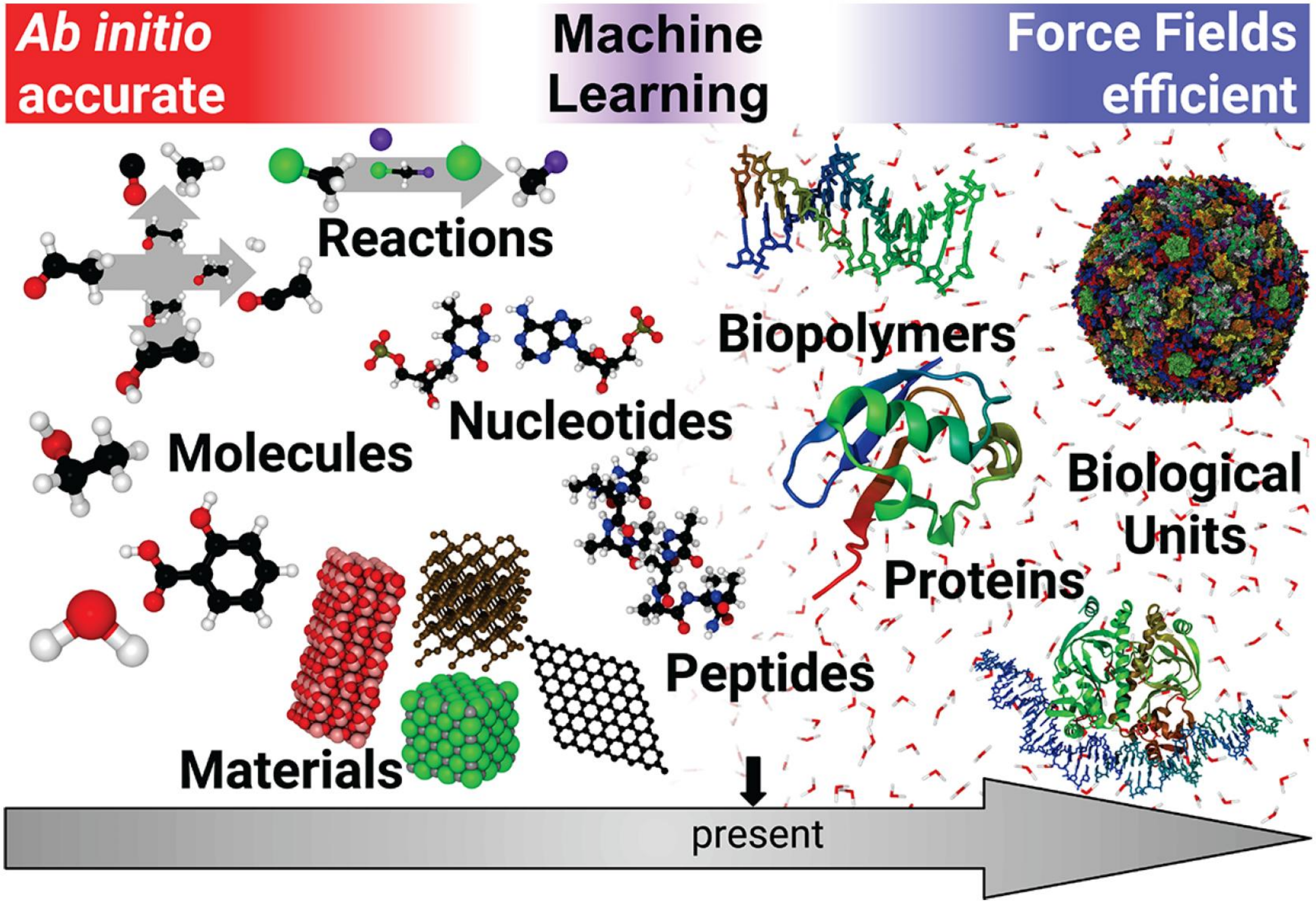
A. Tkatchenko, ***Nature Commun.*** 11, 4125 (2020).

AI/ML Methods in Chemistry by ACS Division



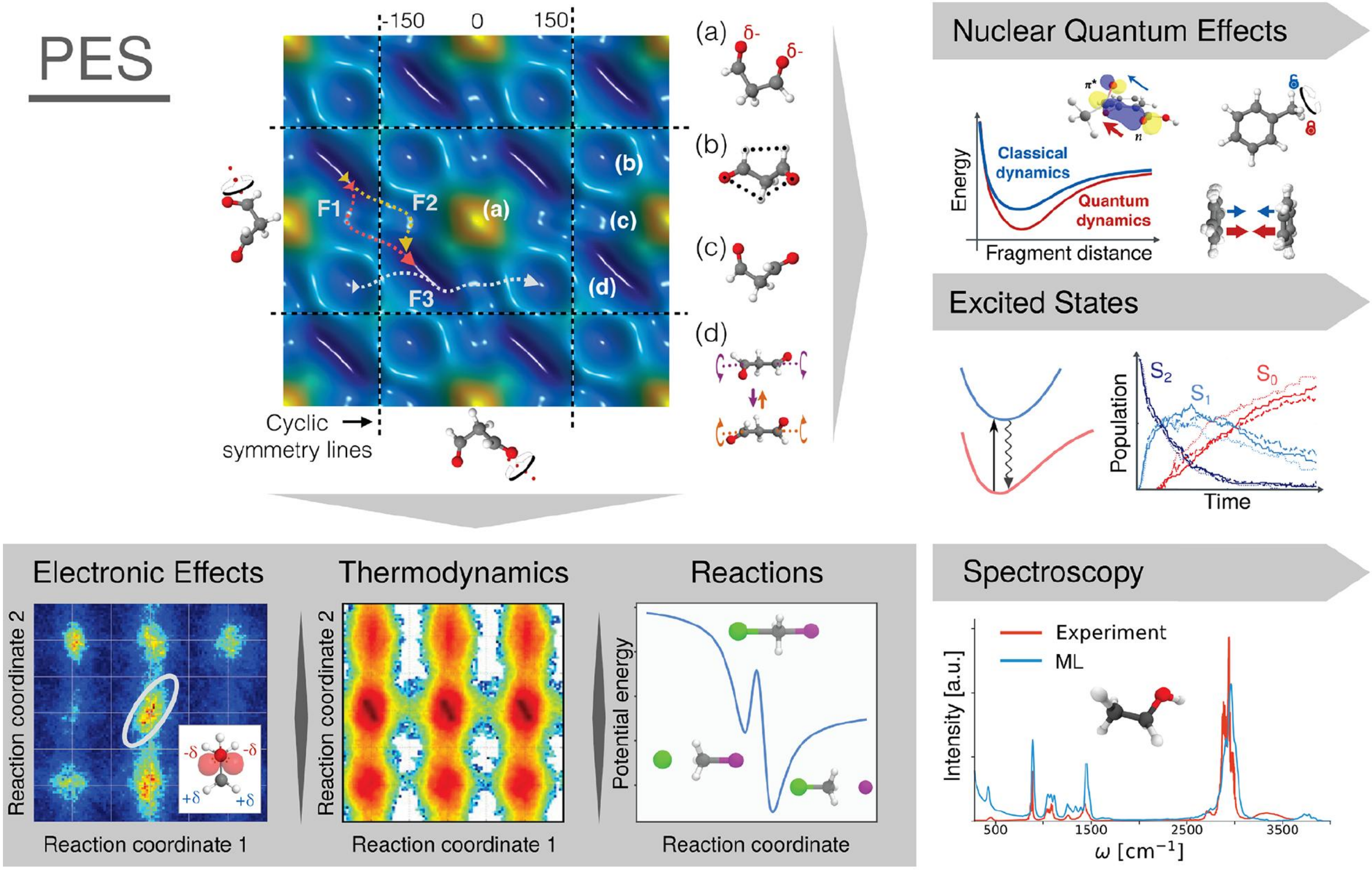
J. Keith et al., ***Chem. Rev.*** 121, 9816 (2021)

Current state-of-the-art of Machine Learning Force Fields (MLFF)



O. Unke *et al.*, ***Chem. Rev.*** 121, 10142 (2021).
I. Poltavsky and A. Tkatchenko, ***J. Phys. Chem. Lett.*** 12, 6551 (2021).

Current state-of-the-art of Machine Learning Force Fields (MLFF)



O. Unke *et al.*, ***Chem. Rev.*** 121, 10142 (2021).

Big Data for Molecules and Materials

nomad-coe.eu

max-centre.eu

e cam

e-cam2020.eu

MARVEL

NATIONAL CENTRE OF COMPETENCE IN RESEARCH

nccr-marvel.ch

Big Data for Molecules and Materials

Table 5. ML Databases for CompChem

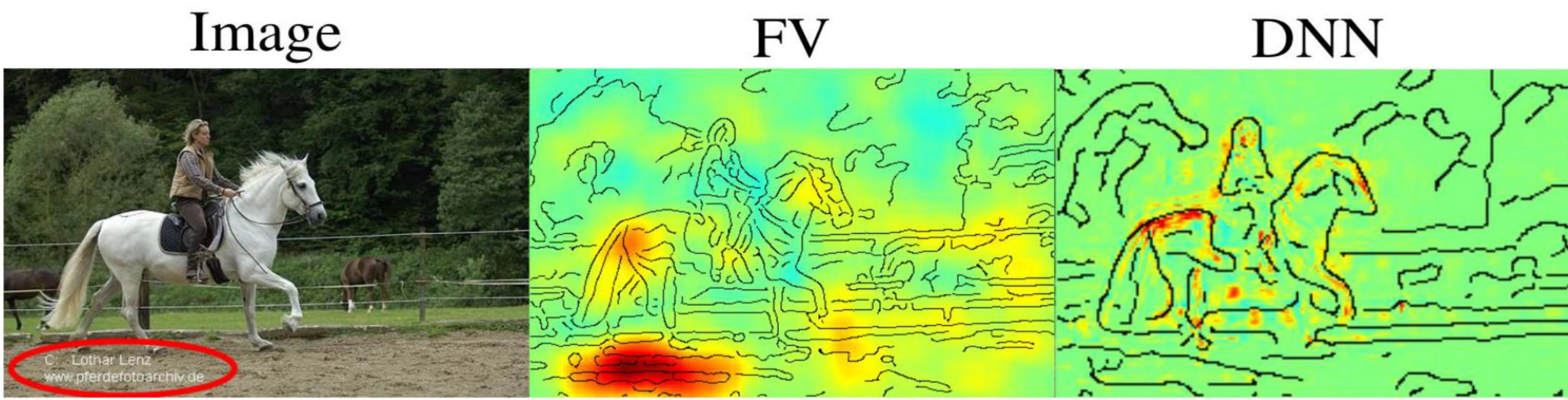
database	description	location
AFLOWLIB	databases containing calculated properties of over 625k materials[510]	http://www.aflowlib.org
ANI-1	large computational DFT database, which consists of more than 20 M off equilibrium conformations for 57.5k small organic molecules[511,512]	https://github.com/isayev/ANI1_dataset
ANI-1x/ANI-1ccx	ANI-1x contains multiple QM properties from 5 M DFT calculations, while ANI-1ccx contains 500k data points obtained with an accurate CCSD(T)/CBS extrapolation[513]	https://github.com/aiqm/ANI1x_datasets
BindingDB	measured binding affinities focusing on interactions of proteins considered to be candidates as drug-targets; 1 200 000 binding data for 5500 proteins and over 520 000 drug-like molecules[514]	http://www.bindingdb.org
Clean Energy Project	contains ~10 000 000 molecular motifs of potential interest which cover small molecule organic photovoltaics and oligomer sequences for polymeric materials[515]	http://cepdb.molecularspace.org
CoRE MOF	database containing over 4700 porous structures of metal−organic frameworks with publicly available atomic coordinates; includes important physical and chemical properties[516]	10.11578/1118280
FreeSolv	experimental and calculated hydration free energies for neutral molecules in water[517]	http://www.escholarship.org/uc/item/6sd403pz
GDB	GDB-11, GDB-13, and GDB-17; together these databases contain billions of small organic molecules following simple chemical stability and synthetic feasibility rules[518]	http://gdb.unibe.ch/downloads/
Hypothetical Zeolites	contains approximately 1 M zeolite structures[519]	http://www.hypotheticalzeolites.net/
Materials Project	contains computed structural, electronic, and energetic data for over 500k compounds[520]	https://www.materialsproject.org
MD17	data sets in this package range in size from 150k to nearly 1 M conformational geometries; all trajectories are calculated at a temperature of 500 K and a resolution of 0.5 fs[372]	http://www.sgdml.org
MoleculeNet	contains data on the properties of over 700k compounds[521]	http://moleculenet.ai
Open Catalyst Project	1.2 M molecular relaxations with results from over 250 M DFT calculations relevant for renewable energy storage[522]	https://opencatalystproject.org/index.html
OQMD	consists of DFT predicted crystallographic parameters and formation energies for over 200k experimentally observed crystal structures[523]	http://oqmd.org
PubChemQC PM6	provides 221 million molecular structures optimized with the PM6 method and several electronic properties computed at the same level of theory[524]	http://pubchemqc.riken.jp/pm6_datasets.html
PubChemQC	provides ~3 million molecular structures optimized by DFT and excited states for over 2 million molecules using TD-DFT[525]	http://pubchemqc.riken.jp/
QM7-X	comprehensive data set of 42 physicochemical properties for ~4.2 M equilibrium and nonequilibrium structures of small organic molecules with up to seven non-hydrogen (C, N, O, S, Cl) atoms[526]	https://zenodo.org/record/4288677#.X9jHNC2ZNTY
QM9	geometric, energetic, electronic, and thermodynamic properties for 134k stable small organic molecules out of GDB-17[527]	https://figshare.com/collections/Quantum_chemistry_structures_and_properties_of_134_kilo_molecules/978904
Synthesis Project	collection of aggregated synthesis parameters computed using the text contained within over 640 000 journal articles[528]	www.synthesisproject.org
quantum-machine.org	a repository of diverse data sets, including valence electron densities, chemical reactions, solvated protein fragments, and molecular Hamiltonians	http://quantum-machine.org/datasets/

J. Keith et al., ***Chem. Rev.*** 121, 9816 (2021)

The Curse of Data:
"ML is the belief in the ignorance of experts"

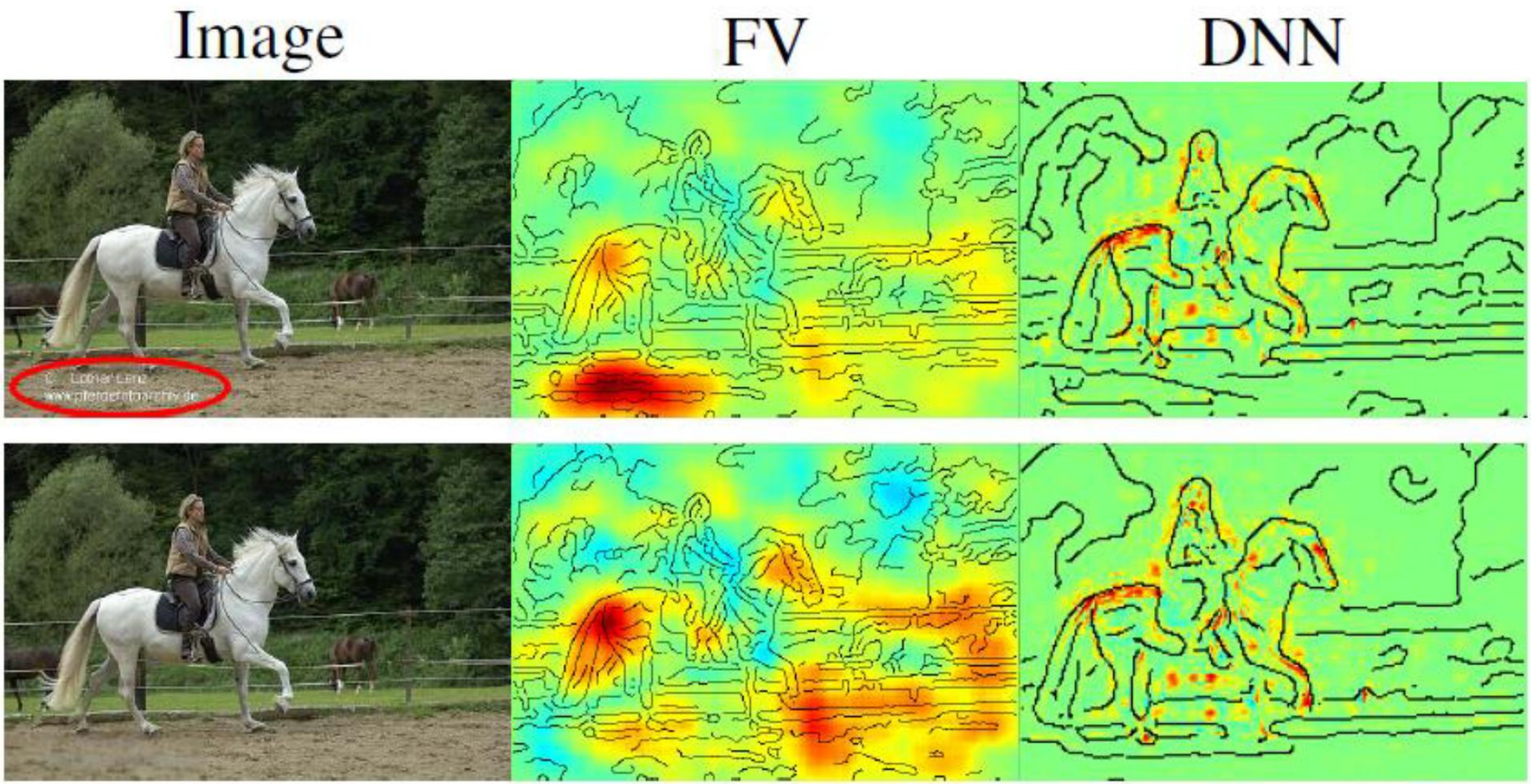
Test error for various classes:

	aeroplane	bicycle	bird	boat	bottle	bus	car
Fisher	79.08%	66.44%	45.90%	70.88%	27.64%	69.67%	80.96%
DeepNet	88.08%	79.69%	80.77%	77.20%	35.48%	72.71%	86.30%
	cat	**chair**	**cow**	**diningtable**	**dog**	**horse**	**motorbike**
Fisher	59.92%	51.92%	47.60%	58.06%	42.28%	80.45%	69.34%
DeepNet	81.10%	51.04%	61.10%	64.62%	76.17%	81.60%	79.33%
	person	**pottedplant**	**sheep**	**sofa**	**train**	**tvmonitor**	**mAP**
Fisher	85.10%	28.62%	49.58%	49.31%	82.71%	54.33%	59.99%
DeepNet	92.43%	49.99%	74.04%	49.48%	87.07%	67.08%	72.12%



Slide by Klaus Mueller – TU Berlin

The Curse of Data: Understanding is Key



Fisher

Neural networks

Slide by Klaus Mueller – TU Berlin

ML for Physics; Physics for ML

- Balancing between descriptor (prior knowledge) and data. Physics is deductive, ML is inductive

- Physicists are great at postulating laws from restricted empirical data

- Physics challenges ML: No noise, high-dimensional data: scalars, vectors, tensors, …; Data points can be chosen at will!

- Abundance of reliable and verifiable data from experiments and calculations

- Behavior of many systems in physics, chemistry, and biology is highly non-linear: Perfect for ML applications

ML for Physics; Physics for ML

- Balancing between descriptor (prior knowledge) and data. Physics is deductive, ML is inductive

- Physicists are great at postulating laws from restricted empirical data

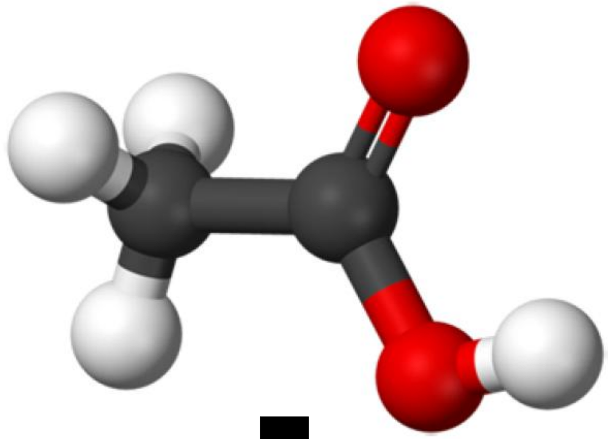
- Physics challenges ML: No noise, high-dimensional data: scalars, vectors, tensors, …; Data points can be chosen at will!

- Abundance of reliable and verifiable data from experiments and calculations

- Behavior of many systems in physics, chemistry, and biology is highly non-linear: Perfect for ML applications

> *It is nice to know that the computer understands the problem.*
> *But I would like to understand it too.*
> ***E. Wigner***

Quantum physics/chemistry today

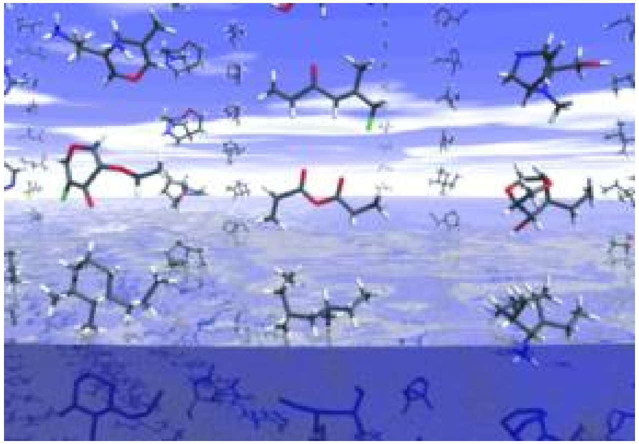


DFT
MP2
CCSD(T)
...

$$\hat{\mathcal{H}}(R_1, Z_1, ..., R_N, Z_N)\widetilde{\Psi} = E\widetilde{\Psi}$$

Properties: Energy, polarizability, HOMO, LUMO, …
Dynamics: Thermal properties, spectroscopy, ...

Quantum physics/chemistry tomorrow?



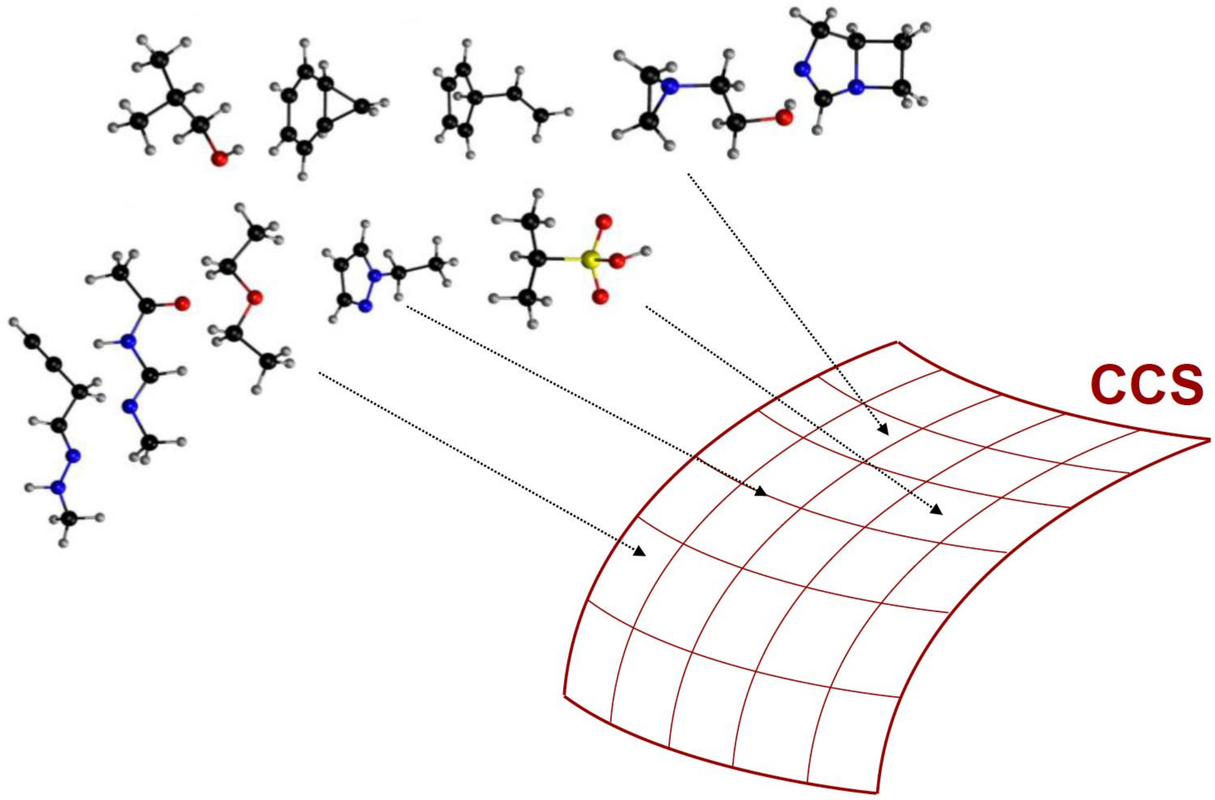
Training data:
molecular properties

ML

Insights:

- Structure of chemical space
- Reactivity trends, aromaticity, “new” chemistry
- Molecular design through multi-property optimization
- ...

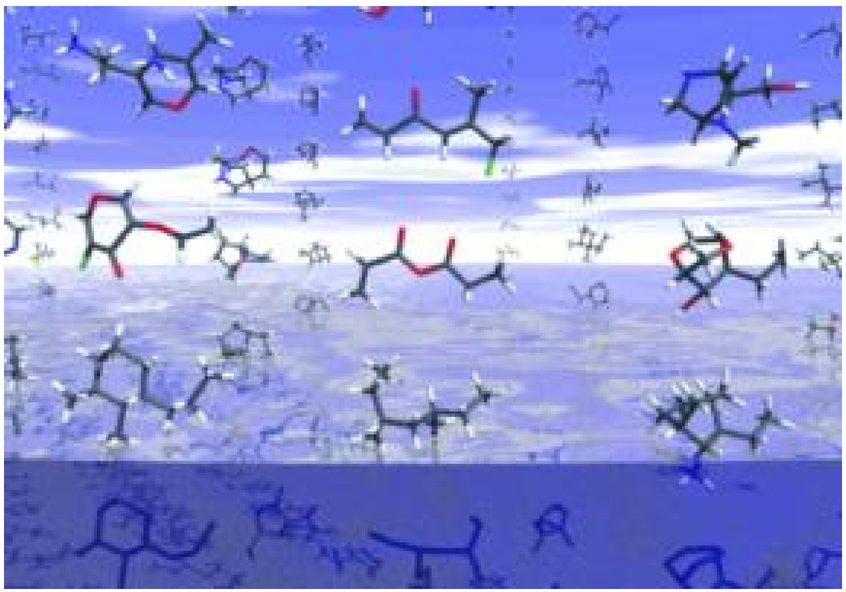
Chemical Space: Molecular big data



- Graph theory: combinatorial explosion
- At least 10^{60} small drug candidate molecules
- Finding needles in a haystack

$\{\boldsymbol{R}_i, Z_i\}$ maps to $\{P_1, P_2, P_3, P_4, ...\}$

Machine learning for molecular big data

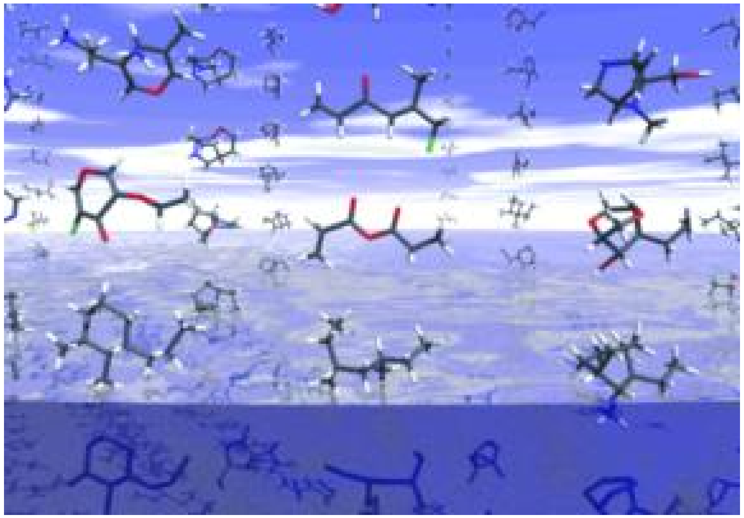


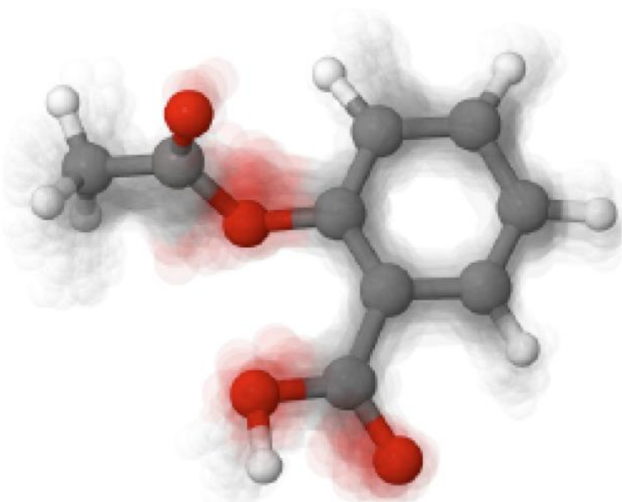
$\{\boldsymbol{R}_i, Z_i\}$ maps to $\{P_1, P_2, P_3, P_4, \ldots\}$

- Descriptor: what's a good representation of a molecule?
- Metric: how to define distance between two molecules?
- Data selection: Which molecules to use for training?
- Properties: which set of properties uniquely defines a molecule?

Can we obtain insights into Chemical Compound Space (CCS) ?

Molecular Data in this Talk





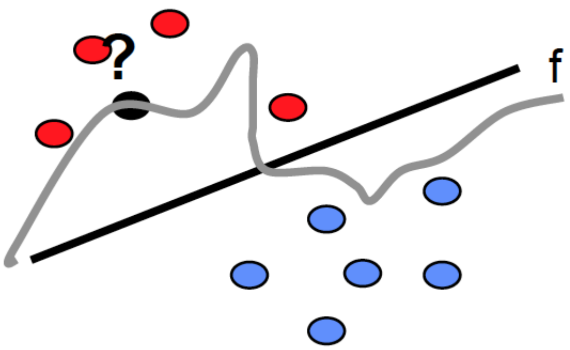
GDB mol graphs: J. L. Reymond (U. Bern)
http://gdb.unibe.ch/downloads/

QM7/QM9 datasets: Hybrid DFT calculations by von Lilienfeld's group (Sci. Data 2014) and my group (PRL 2012).

MD17/MD22/ISO17 datasets: Molecular dynamics trajectories from my group (DFT and CCSD(T) levels)

www.quantum-machine.org

Machine Learning in a nutshell



Typical scenario: learning from data

• given data set **X** and labels **Y** (generated by some joint probabilty distribution p(x,y))

• **LEARN/INFER** underlying **unknown** mapping

Y = f(X)

Example: ~~understand~~ **fit** chemical compound space, distinguish brain states …

BUT: how to do this optimally with good performance on **unseen** data?

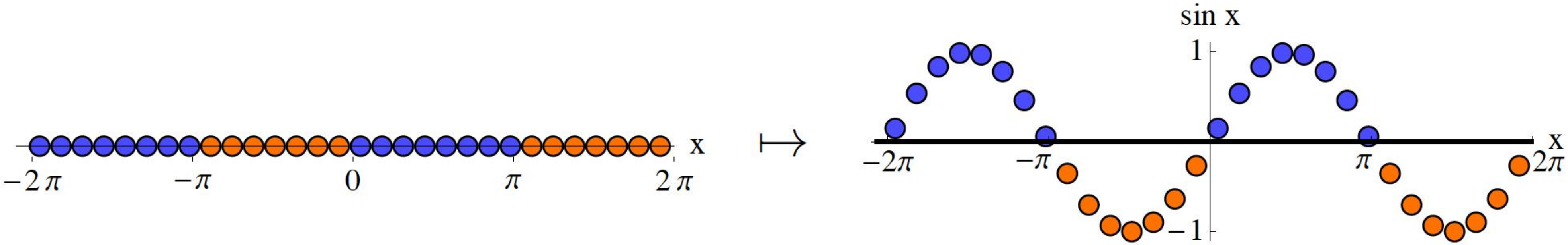
Most popular techniques: **kernel methods** and (deep) **neural networks**.

Slide by Klaus Mueller – TU Berlin

Kernel Learning

Idea:

- Transform samples into higher-dimensional space
- *Implicitly* compute inner products there
- Rewrite linear algorithm to use only inner products



Input space $\mathcal{X}$ $\xrightarrow{\phi}$ Feature space $\mathcal{H}$

$$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}, \qquad k(x, z) = \langle \phi(x), \phi(z) \rangle$$

Slide by Matthias Rupp

Kernel Ridge Regression

- Regularized form of ordinary regression
- Regularization prevents over-fitting by penalizing large coefficients
- Use of kernels for non-linearity

Solution has form

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_i, \mathbf{x})$$

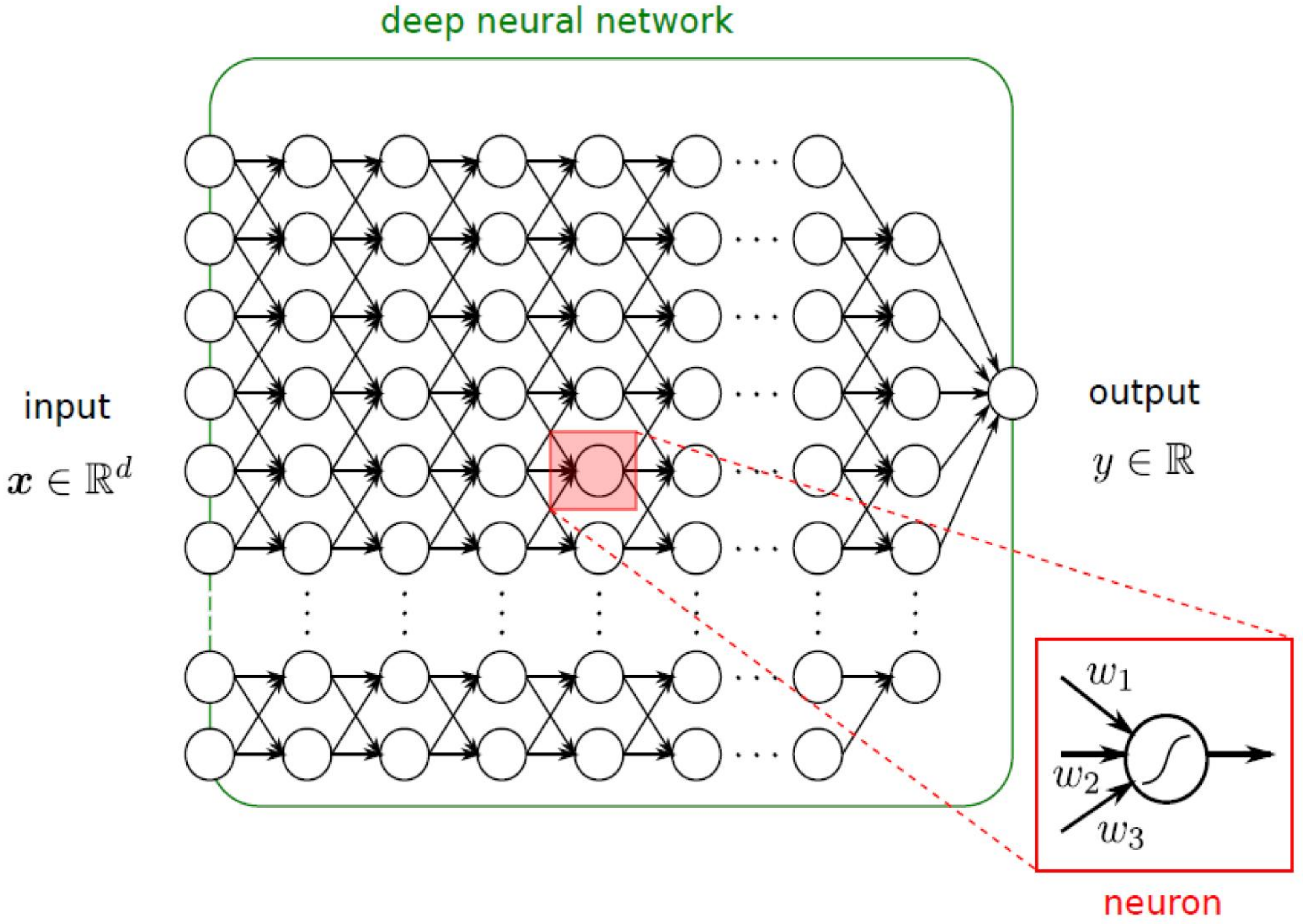
Coefficients α are obtained by solving

$$\sum_{i=1}^{n} (f(\mathbf{x}_i) - y_i)^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha},$$

which has solution

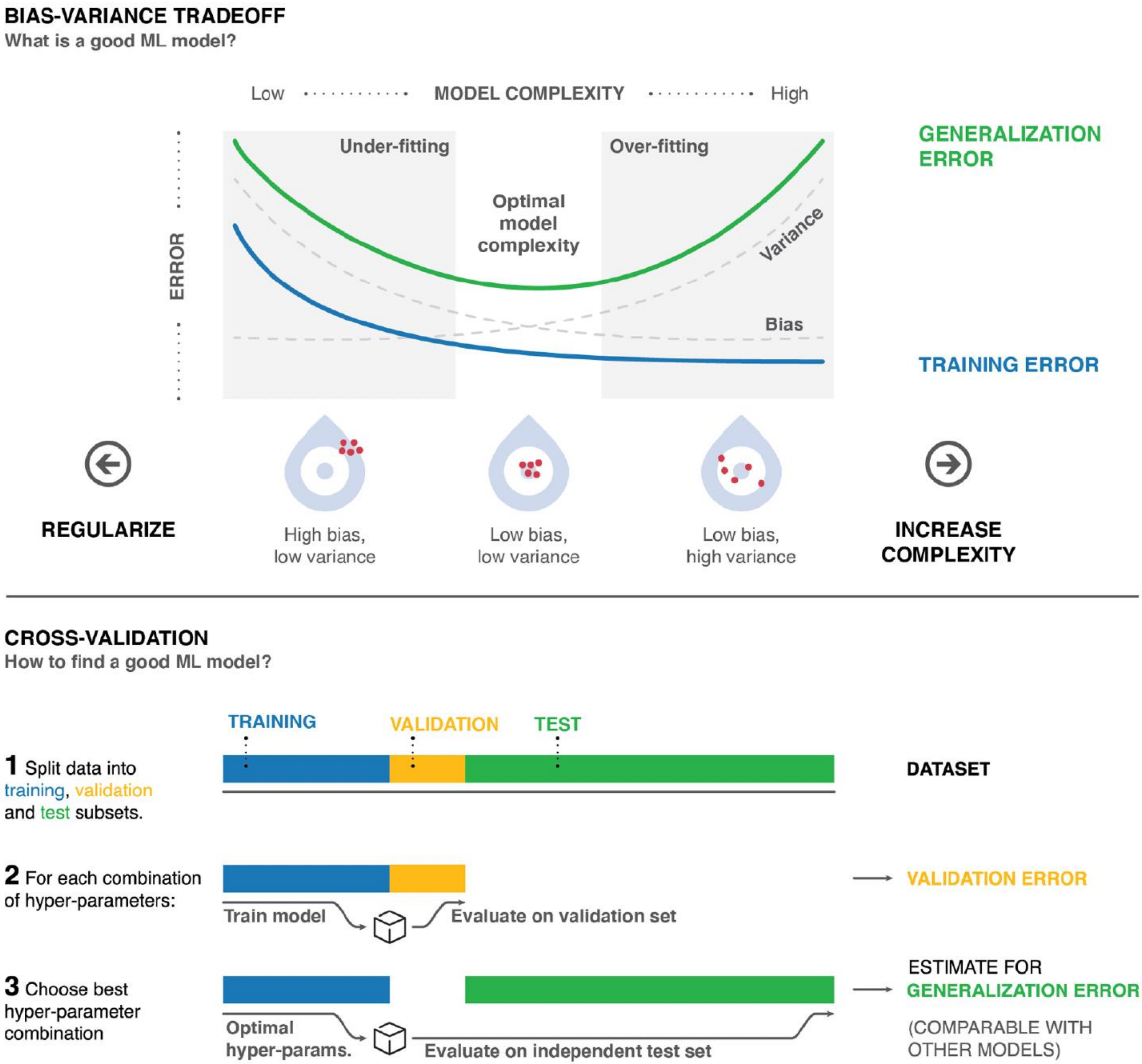
$$\alpha = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}.$$

Neural Networks



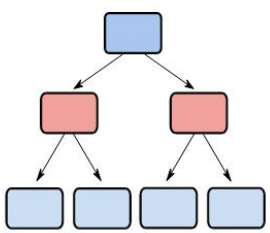
- Neuron applies a nonlinear function to its input.
- Examples of functions: hyperbolic tangent, rectification.

How to train ML models

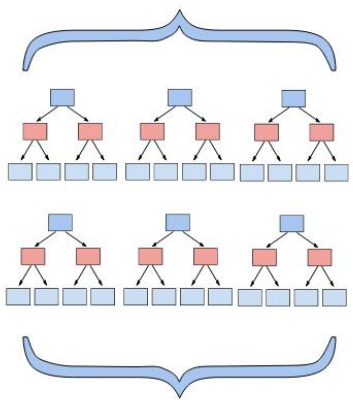


Other Examples of (Non)Linear ML Methods

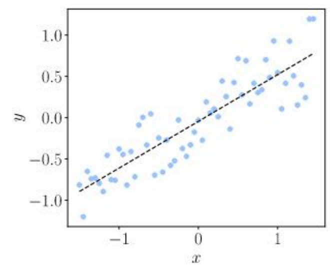
a) Decision Trees



b) Random Forest

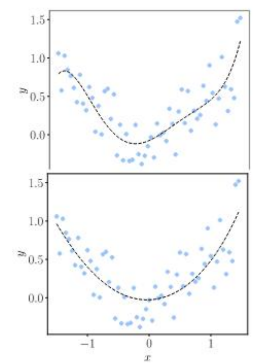


c) Linear Regression



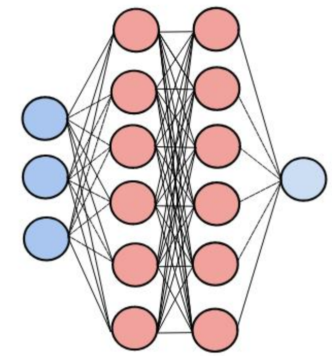
$$Loss = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij} W_j)^2$$

d) Regularized Linear Regression



$$Loss = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij} W_j)^2 + \lambda \sum_{j=0}^{M} |W_j|$$

e) Neural Networks



“First” papers on ML for Molecules/Materials: Importance of Baselines

PHYSICAL REVIEW LETTERS

Highlights Recent Accepted Collections Authors Referees Search Press About Editorial Team

Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces

Jörg Behler and Michele Parrinello
Phys. Rev. Lett. **98**, 146401 – Published 2 April 2007

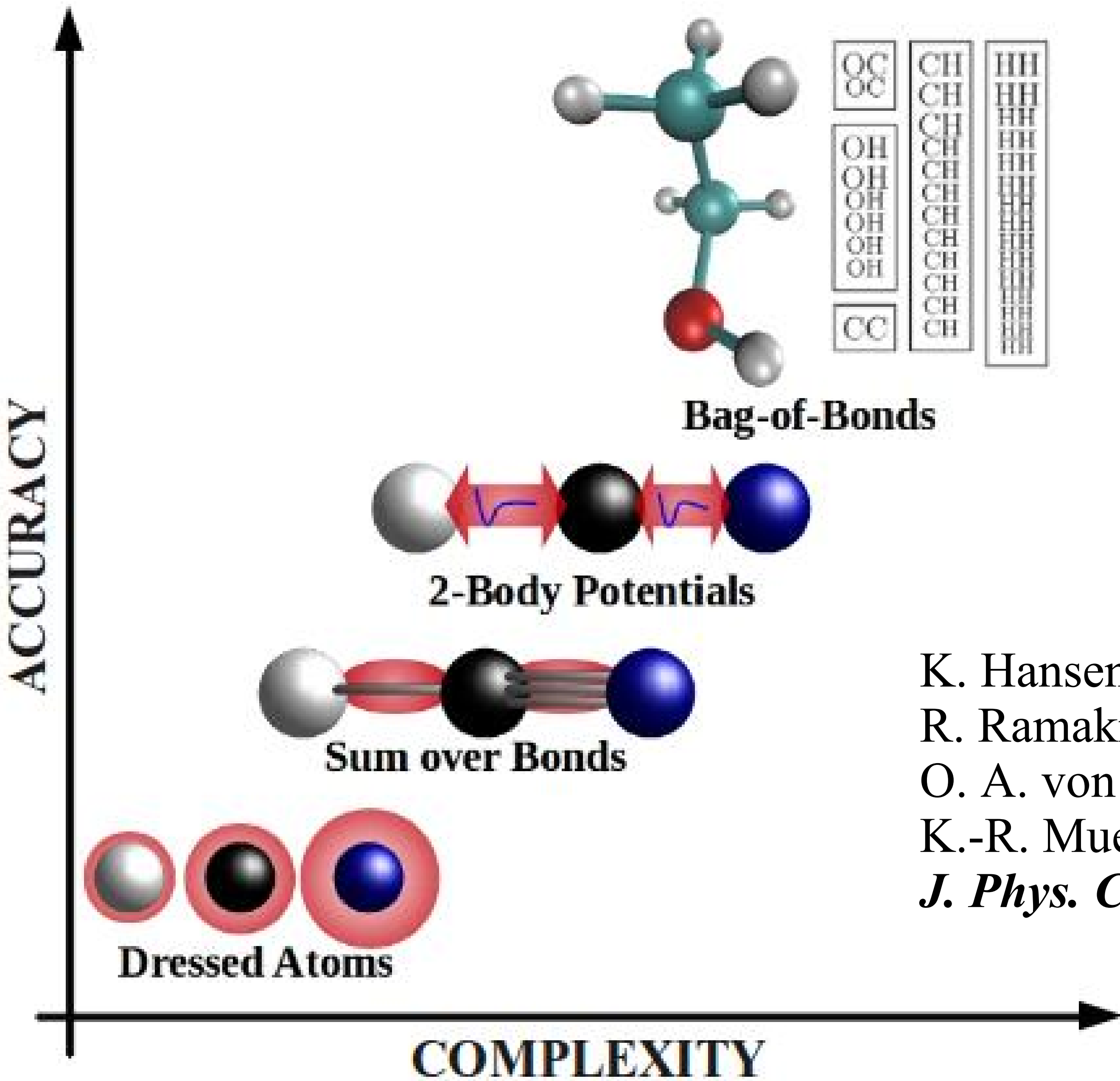
Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons

Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi
Phys. Rev. Lett. **104**, 136403 – Published 1 April 2010

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

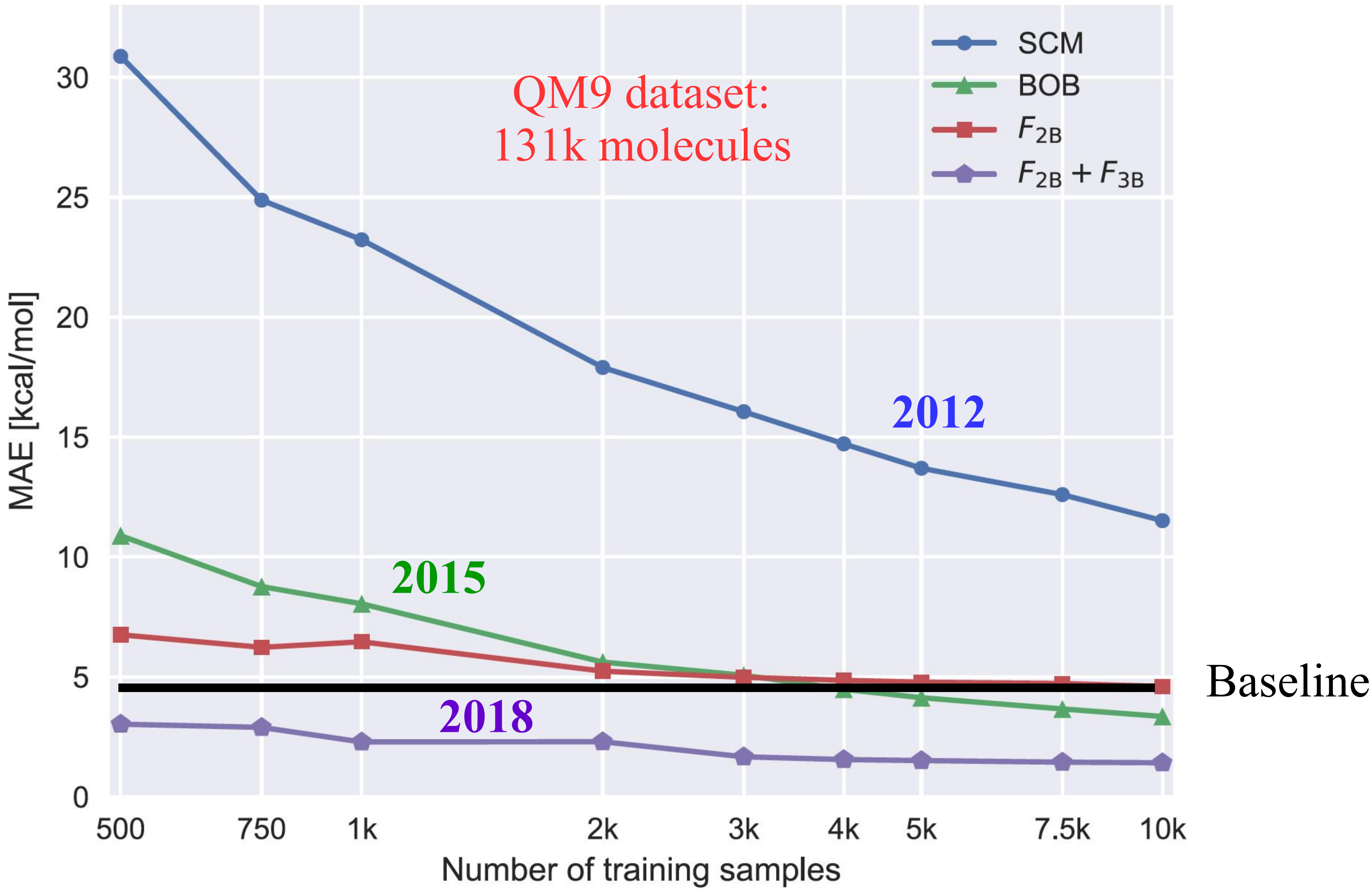
Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld
Phys. Rev. Lett. **108**, 058301 – Published 31 January 2012

Baselines for Molecular Property Prediction: From Sum-Over-Atoms to Bag-of-Bonds



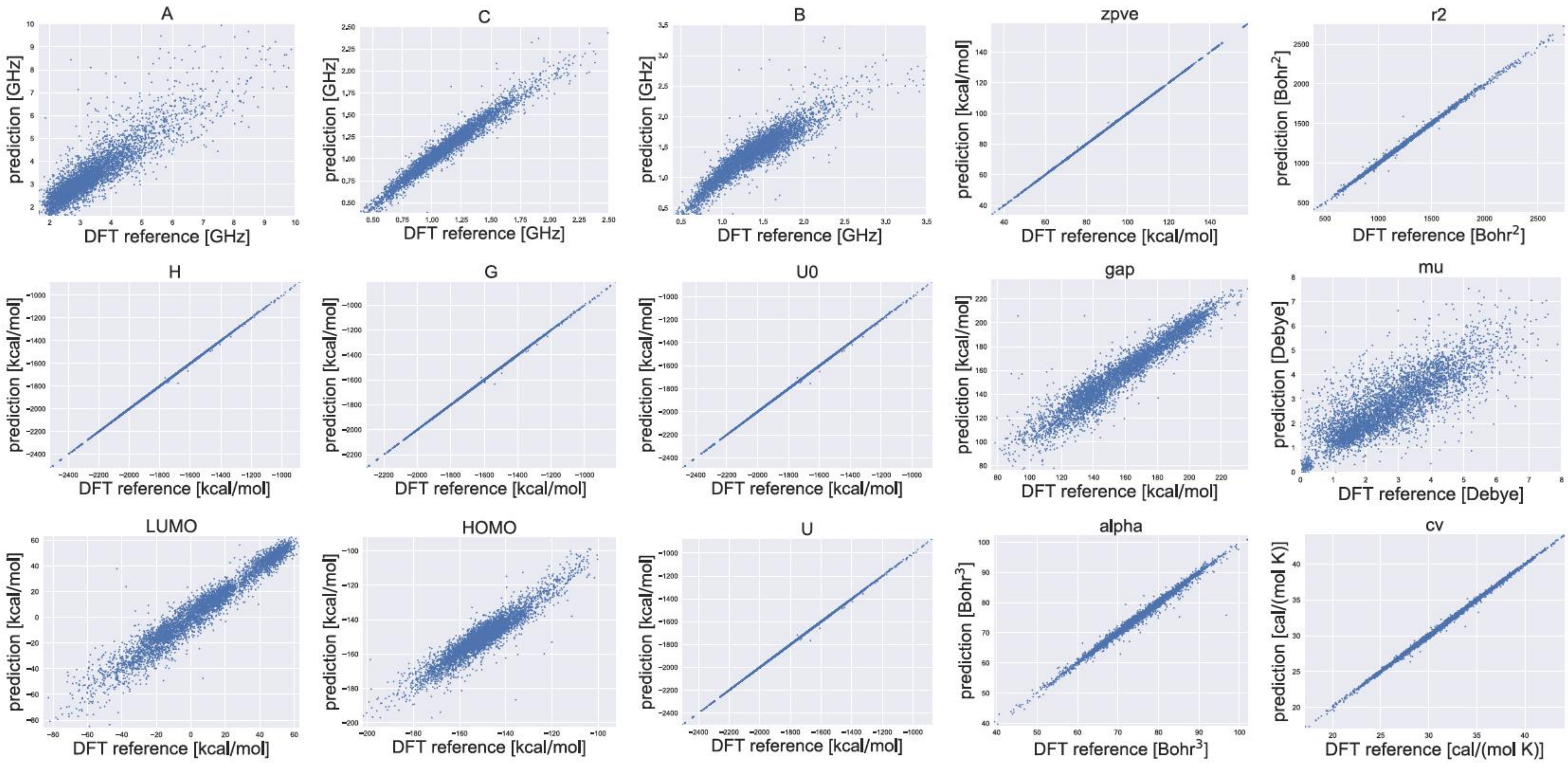
K. Hansen, F. Biegler,
R. Ramakrishnan, W. Pronobis,
O. A. von Lilienfeld,
K.-R. Mueller, and A. Tkatchenko,
J. Phys. Chem. Lett. 6, 2326 (2015).

Molecular Datasets:
Evolution of Physics in Representations



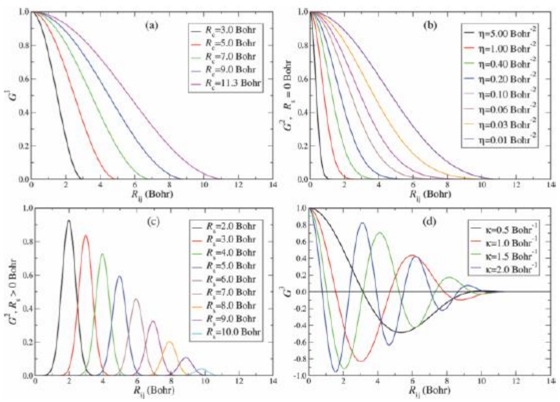
W. Pronobis, A. Tkatchenko, and K.-R. Mueller, ***J. Chem. Theory Comput.*** (2018).

QM9 dataset: Extensive and Intensive Properties



W. Pronobis, A. Tkatchenko, and K.-R. Mueller, ***J. Chem. Theory Comput.*** (2018).

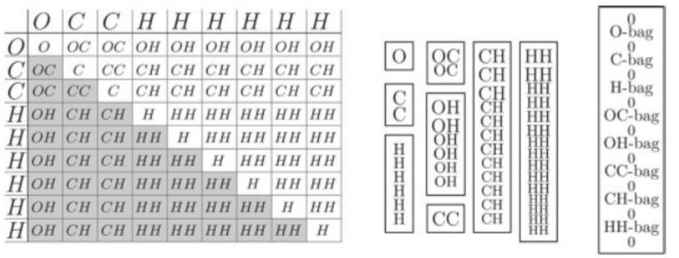
Zoo of Descriptors for Molecules and Solids



Atom-centered symmetry functions (Behler et al. 2007)

$$M_{ij} = \begin{cases} 0.5 Z_i^{2.4} & \text{for } i = j \\ \frac{Z_i Z_j}{d_{ij}} & \text{for } i \neq j \end{cases}$$

Coulomb matrix (Rupp et al. 2012)



Bag of bonds (Hansen et al. 2015)

$$\{Z_i, \mathbf{R}_i\}$$

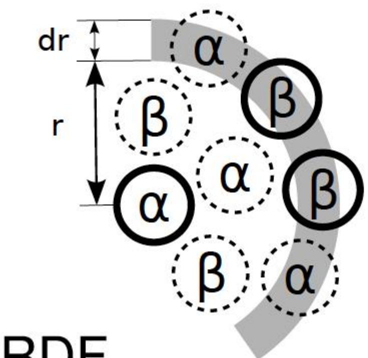
$$\{Z_i, d_{ij}\}$$

$$k(\rho, \rho') = \int d\hat{R} \left| \rho(\mathbf{r}) \rho'(\hat{R}\mathbf{r}) \right|^n$$

SOAP (Bartók et al. 2013)

$$x_{ij} = \begin{cases} 0.5 Z_i^{2.4} & \text{if } i = j \\ \frac{Z_i Z_j}{\phi(\mathbf{r}_i, \mathbf{r}_j)} & \text{if } i \neq j \end{cases}$$

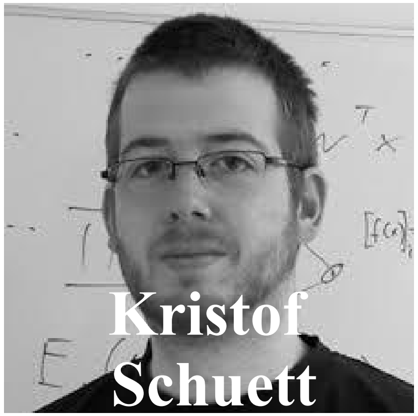
Sine matrix (Faber et al. 2015)



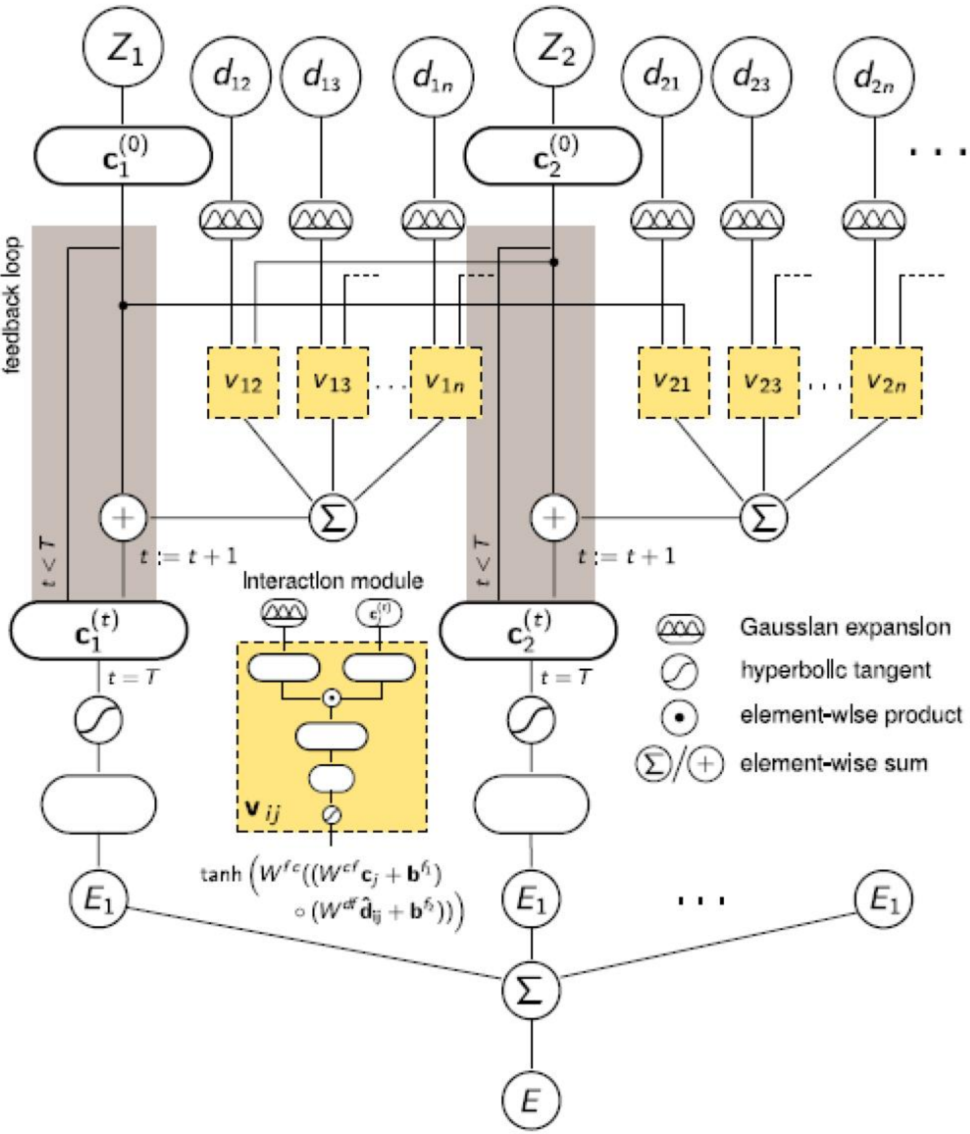
PRDF (Schütt et al, 2014)

Slide from K. Schuett

Learning the Representation: Molecular Deep Tensor Neural Networks (DTNN)



Molecular Deep Tensor Neural Networks (DTNN)



$$\hat{\mathcal{H}}\Psi = E\Psi$$

Atom descriptors:

$$\mathbf{c}_i^{(0)} = \mathbf{c}_{Z_i} \in \mathbb{R}^d$$

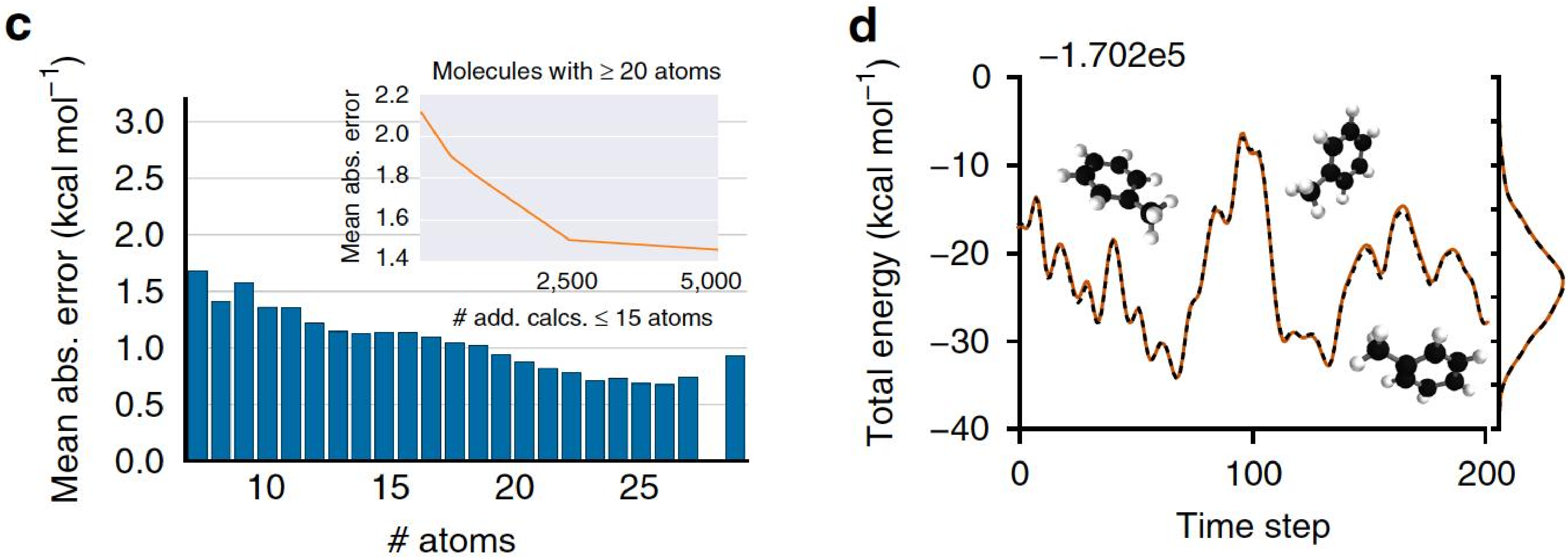
$$\mathbf{c}_i^{(t+1)} = \mathbf{c}_i^{(t)} + \sum_{j \neq i} \tanh(\mathbf{v}_{ij})$$

Interaction correction:

$$\mathbf{v}_{ij} = \mathbf{c}_j^{(t)} V_k \hat{\mathbf{d}}_{ij} + (W^c \mathbf{c}_j^{(t)})_k + (W^d \hat{\mathbf{d}}_{ij})_k + b_k$$

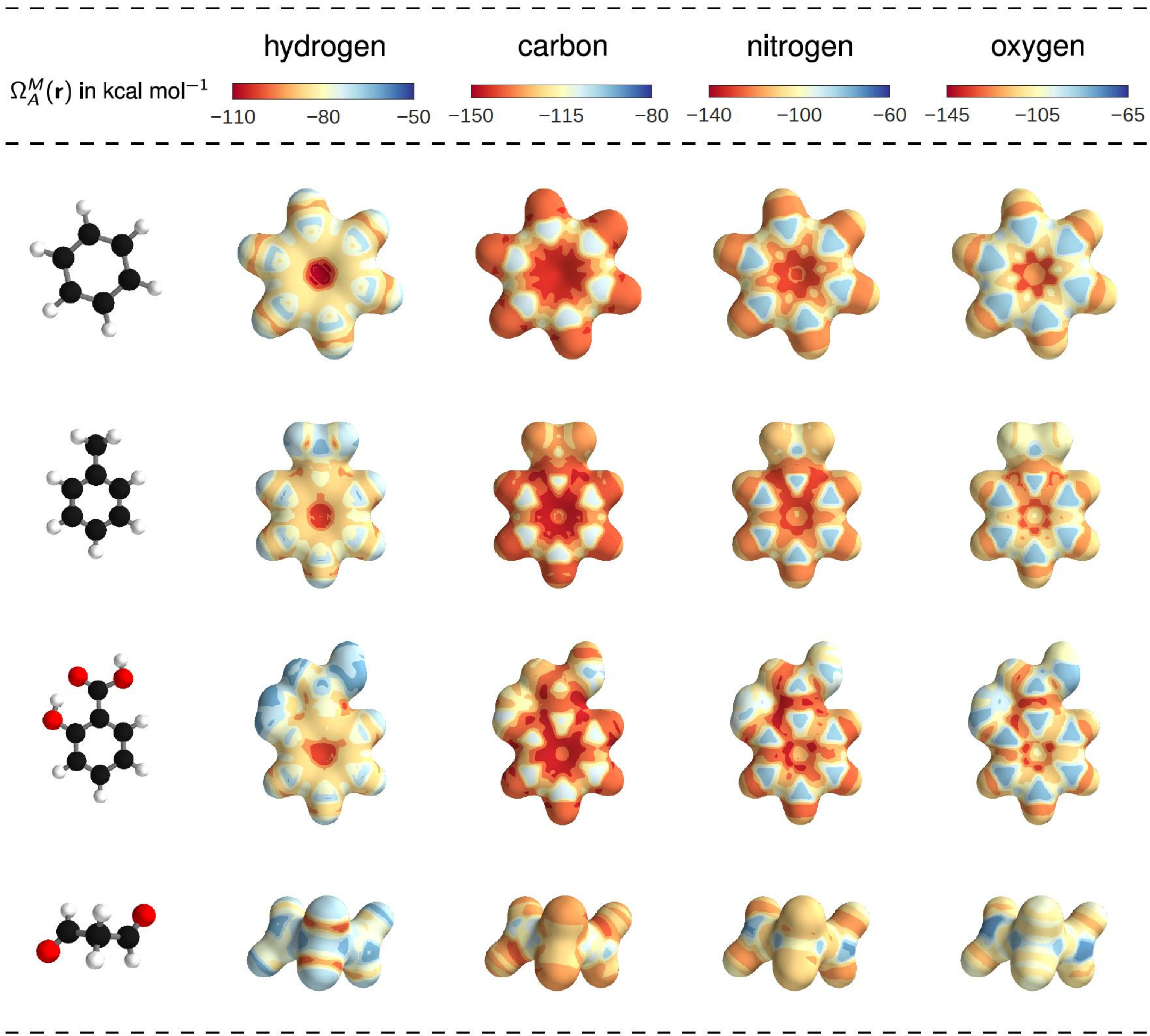
K. T. Schuett, F. Arbabzadah, S. Chmiela, K.-R. Mueller, and A. Tkatchenko, ***Nature Commun.*** 8, 13890 (2017).

Molecular DTNN: Performance on QM9 and MD

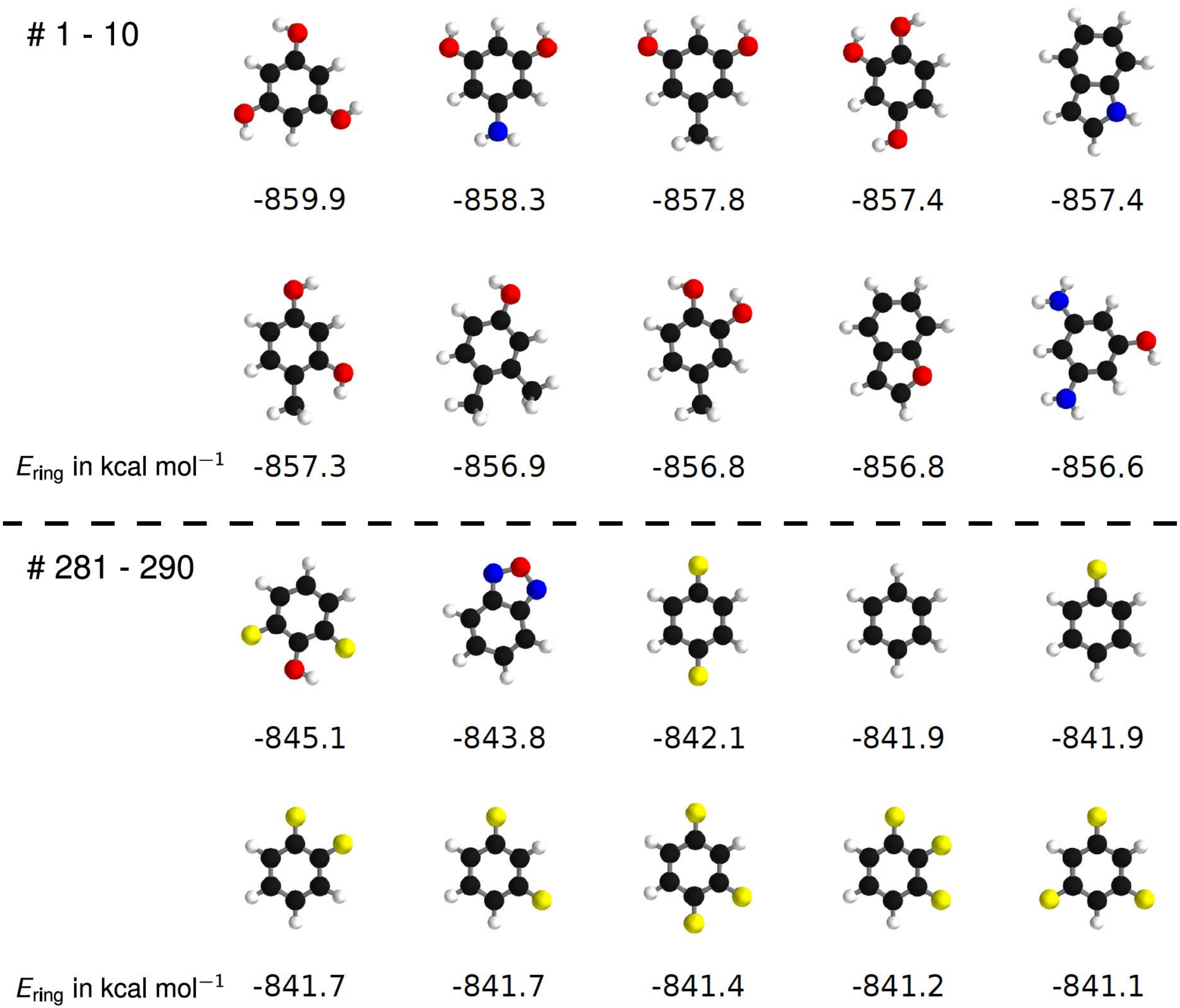


K. T. Schuett, F. Arbabzadah, S. Chmiela, K.-R. Mueller, and A. Tkatchenko, ***Nature Commun.*** 8, 13890 (2017).

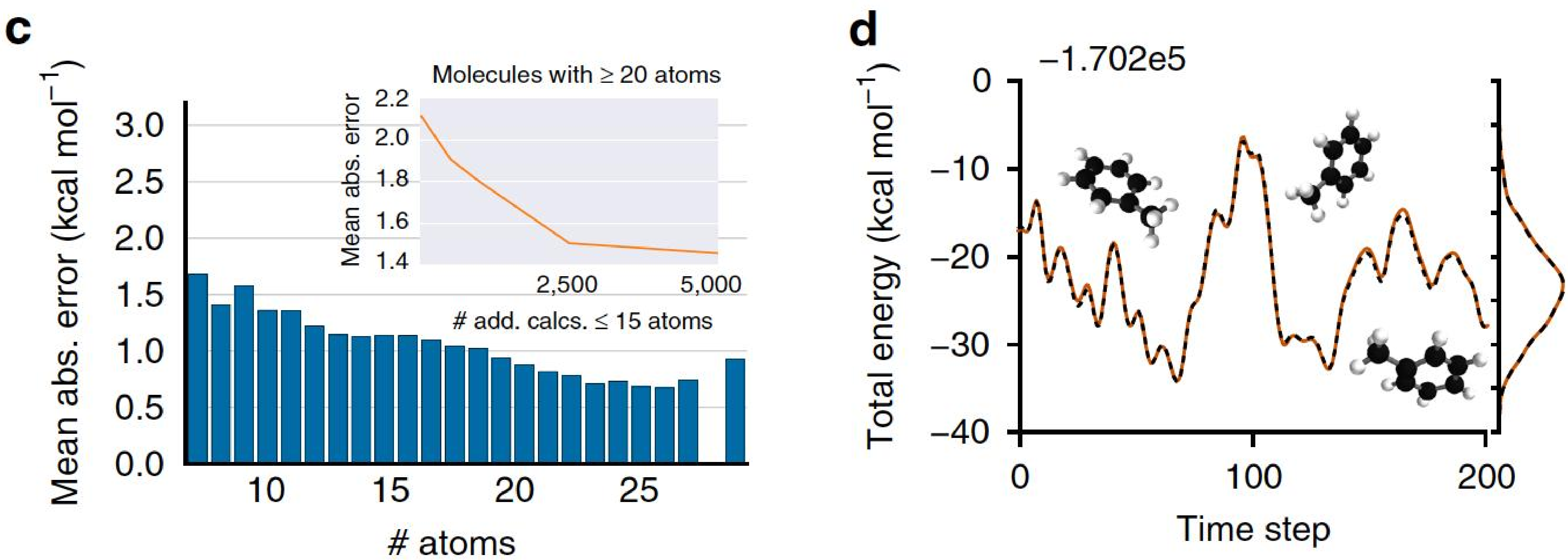
Molecular DTNN: What Did it Learn ?



Quantum Chemical Insights: Aromaticity



Challenge: Learning Full Chemical Space with ML

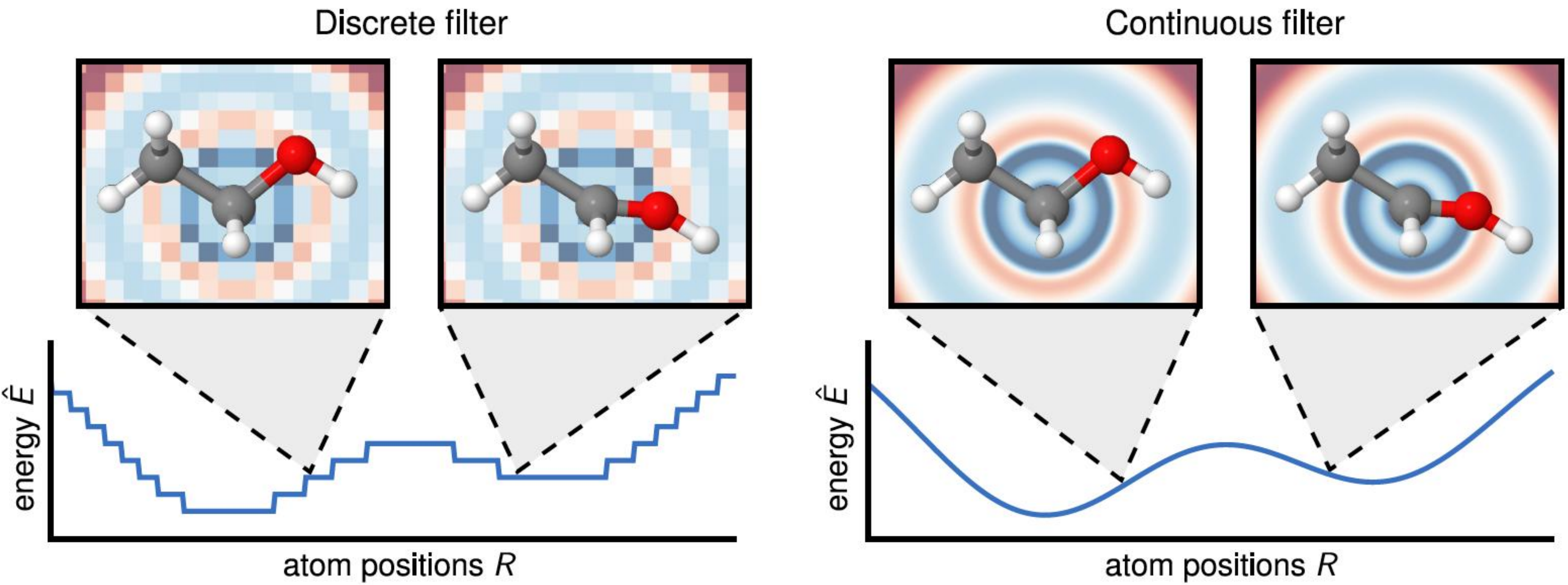


Accurately representing BOTH compositional and conformational degrees of freedom is very difficult.

For C7O2H10 isomer and MD data, the error grows to 1.7 kcal/mol.

K. T. Schuett, F. Arbabzadah, S. Chmiela, K.-R. Mueller, and A. Tkatchenko, ***Nature Commun.*** 8, 13890 (2017).

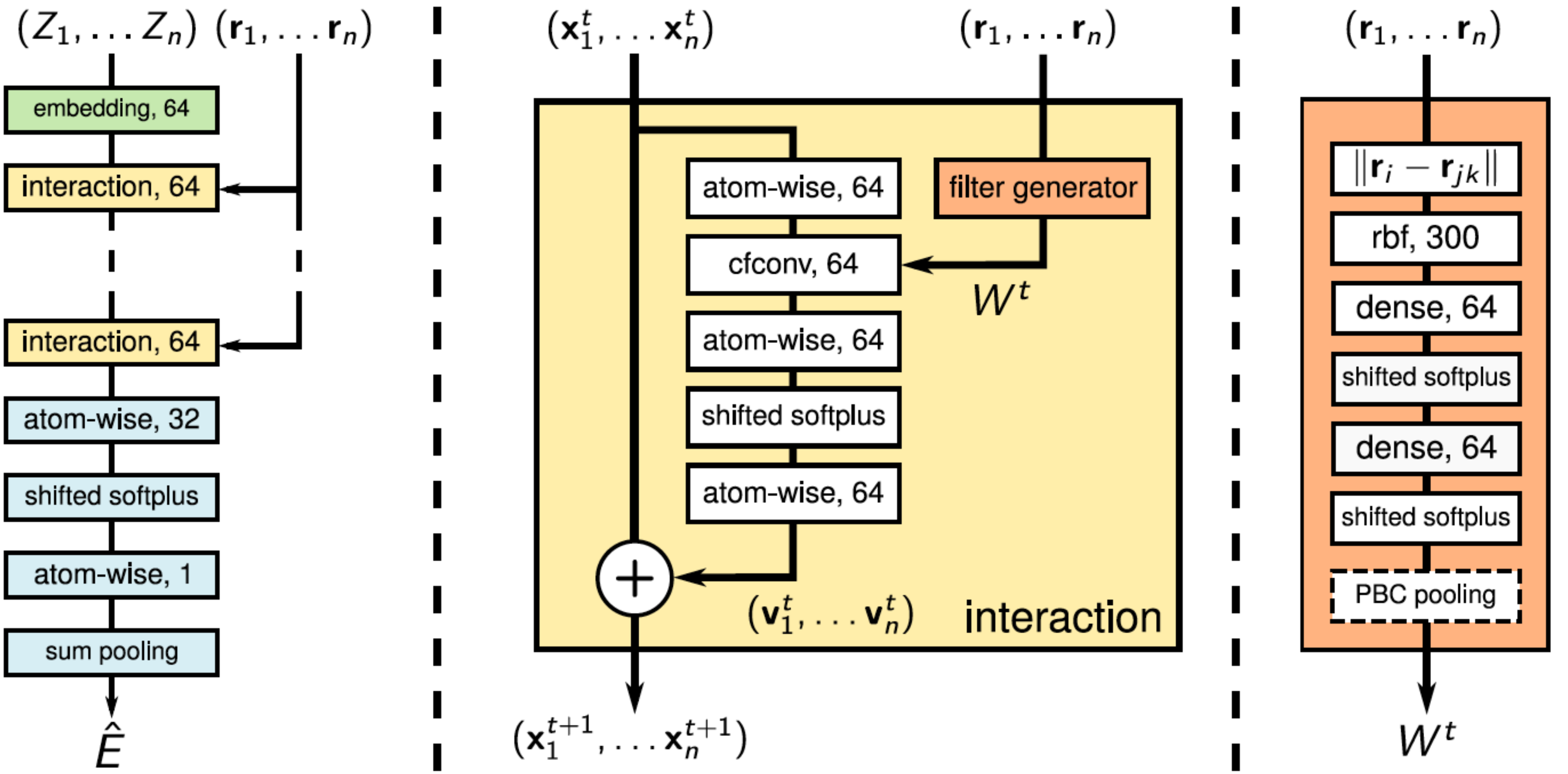
From DTNN to SchNet architecture



$$\mathbf{v}_i^{(t)} = \sum_{j=1}^{N_{\text{atom}}} \mathbf{x}_j^{(t)} \circ \underbrace{W_{[d_{ij}]}^{(t)}}_{\text{parameter tensor}} \qquad \mathbf{v}_i^{(t)} = \sum_{j=1}^{N_{\text{atom}}} \mathbf{x}_j^{(t)} \circ \underbrace{W^{(t)}(d_{ij})}_{\text{neural network}}$$

K.T. Schuett, P.J. Kindermans, H.E. Sauceda, S. Chmiela, A. Tkatchenko, K.-R. Mueller (2017). *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*. NeurIPS.

From DTNN to SchNet architecture



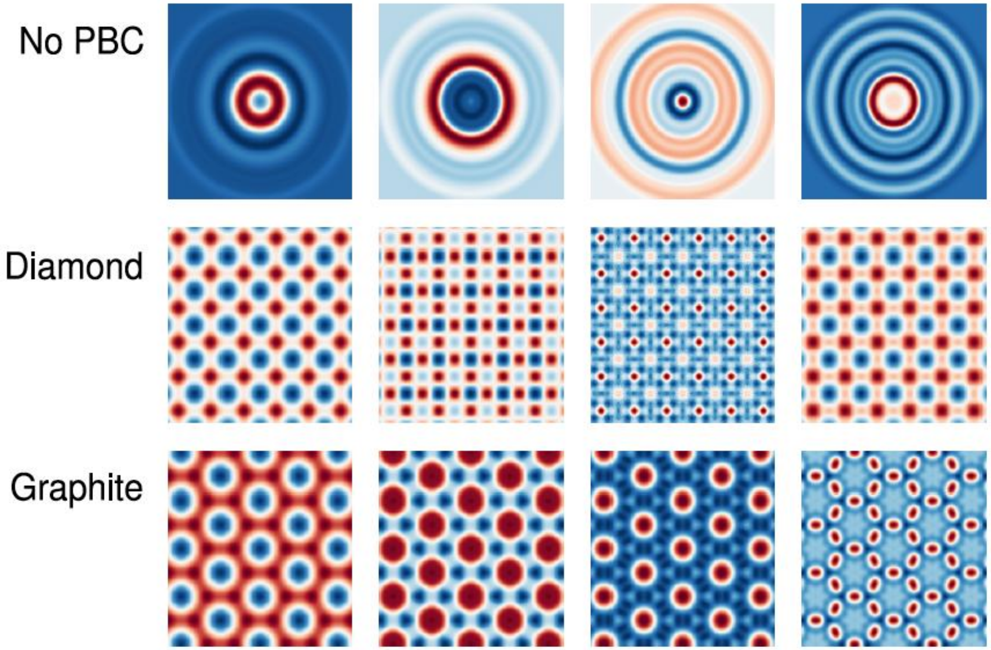
K.T. Schuett, P.J. Kindermans, H.E. Sauceda, S. Chmiela, A. Tkatchenko, K.-R. Mueller (2017). *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*. NeurIPS.

SchNet architecture: Application to Materials

$$\mathbf{x}_i^{l+1} = \sum_{j=0}^{n_{\text{atoms}}} \mathbf{x}_j^l \circ \underbrace{\left(\sum_{b=0}^{n_{\text{cells}}} \tilde{W}^l (\mathbf{r}_{jb} - \mathbf{r}_{ia}) \right)}_{\text{periodic filter}}$$

Model	$N = 3,000$	$N = 60,000$
ext. Coulomb matrix[1]	0.640	–
Ewald sum matrix[1]	0.490	–
sine matrix[1]	0.370	–
SchNet ($T = 6$)	**0.127**	**0.035**

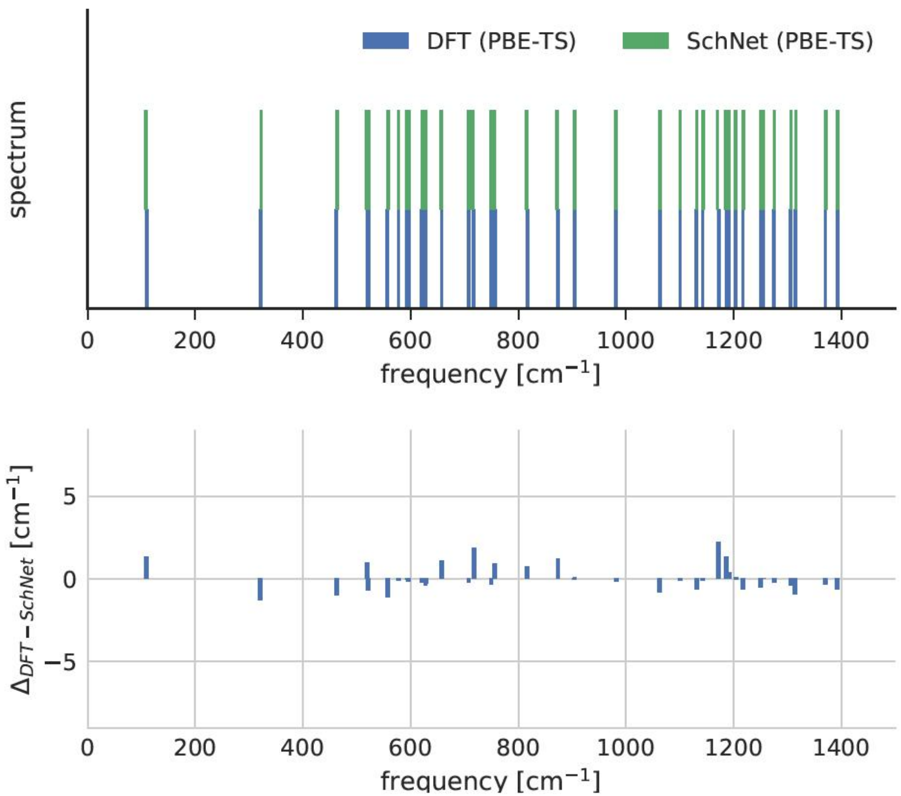
[1] Faber et al. Crystal structure representations for machine learning models of formation energies (2015). IJQC **115**(16).



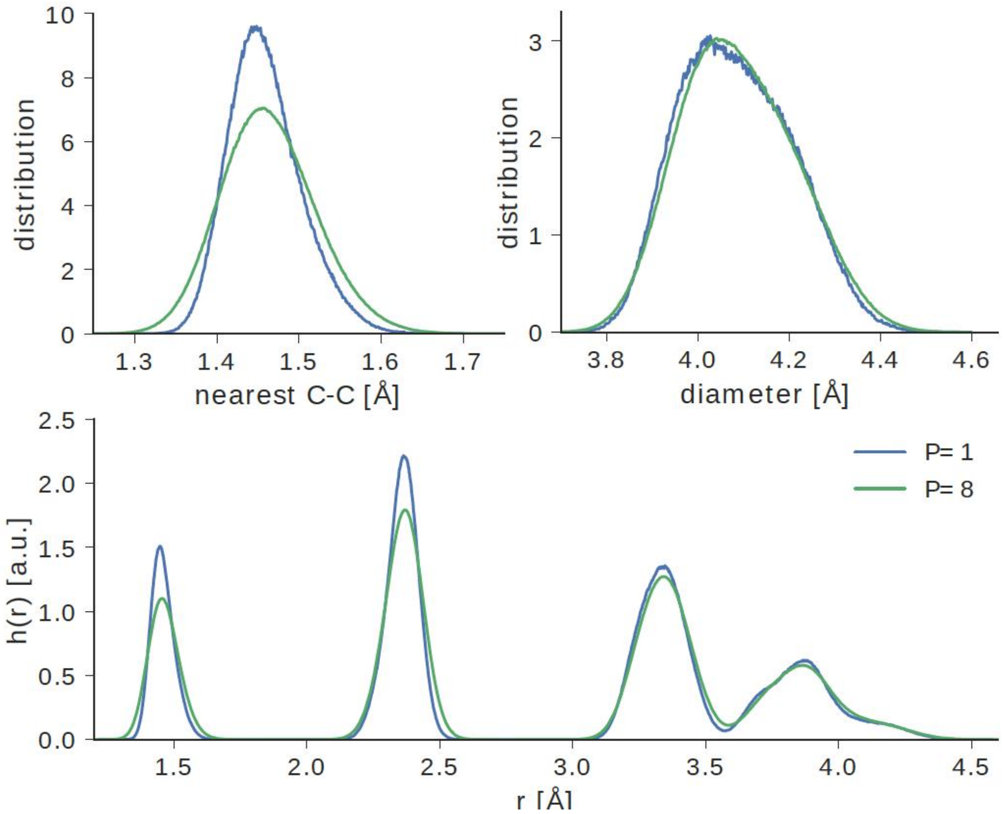
K.T. Schuett, P.J. Kindermans, H.E. Sauceda, S. Chmiela, A. Tkatchenko, K.-R. Mueller (2017). *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*. NeurIPS.

SchNet architecture: Application to Molecular Dynamics

Accurate prediction of vibrational frequencies

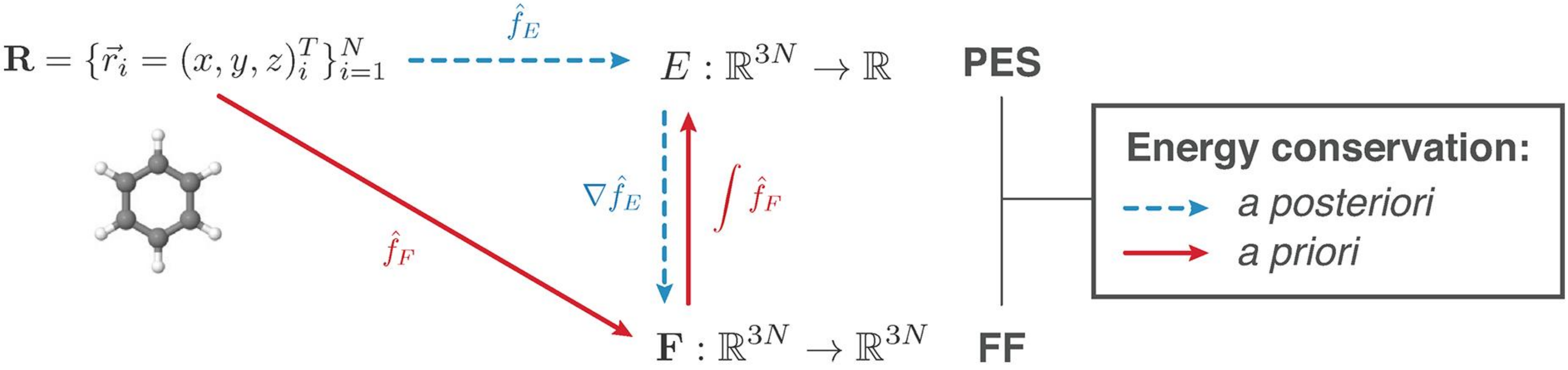


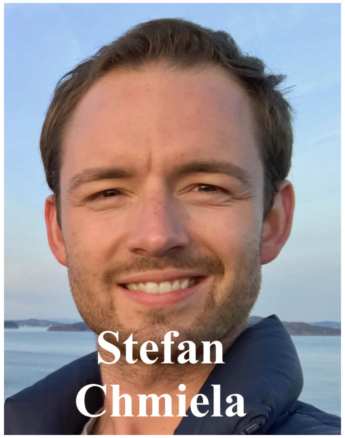
PIMD@SchNet shows delocalization of bonds



K.T. Schuett, H.E. Sauceda, P. J. Kindermans, S. Chmiela, A. Tkatchenko, K.-R. Mueller, *J. Chem. Phys.* **148**, 241722 (2018).

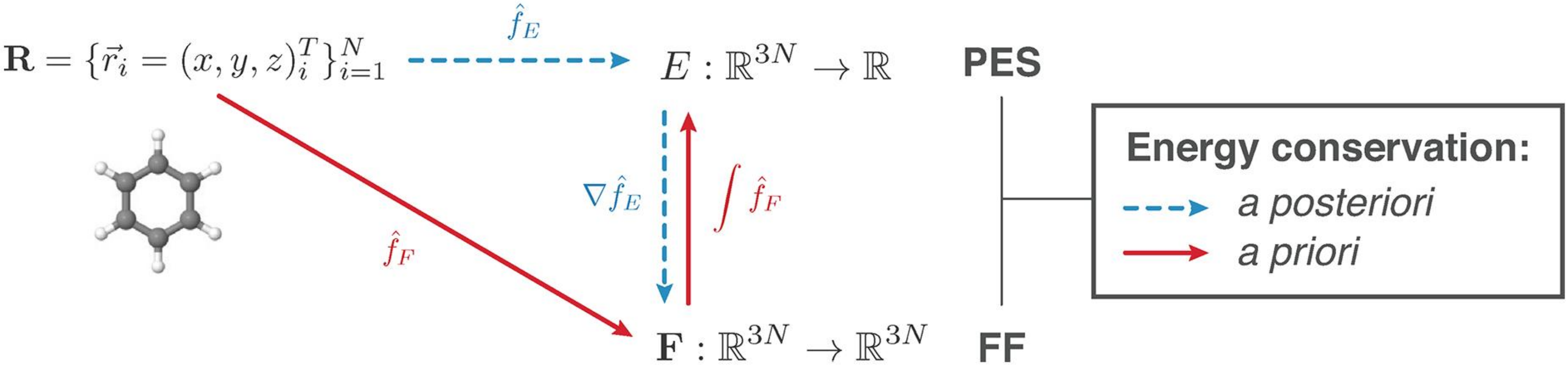
MLFFs based on Physical Principles

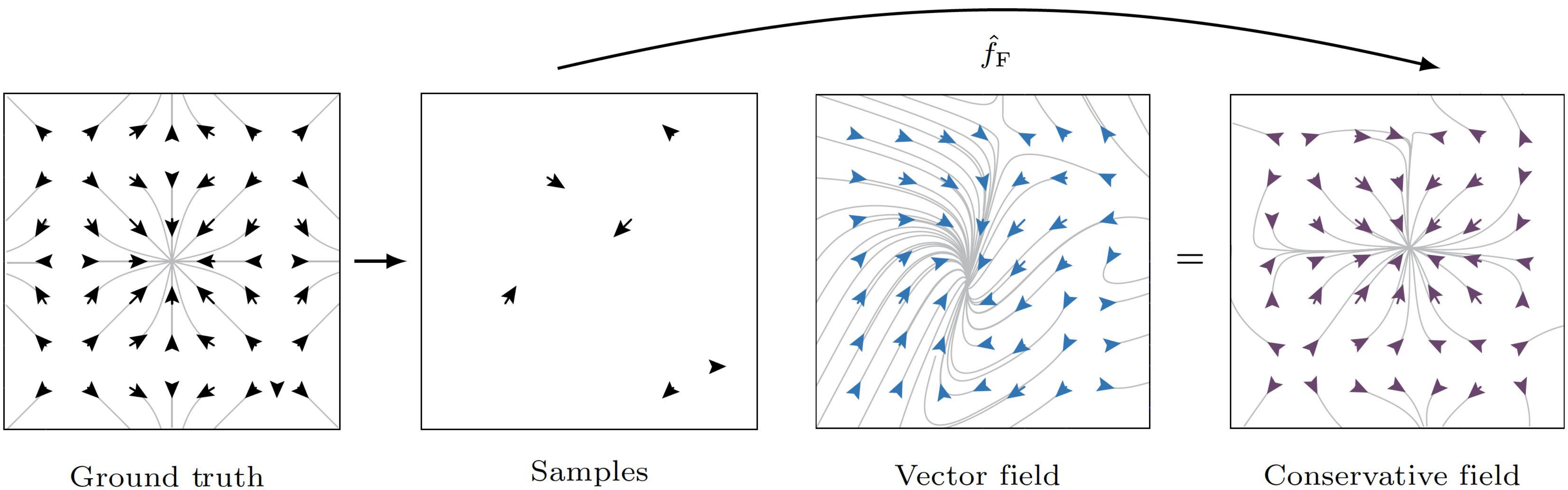




S. Chmiela *et al.*, ***Science Adv.* 3**, e1603015 (2017); ***Nature Commun.* 9**, 3887 (2018).

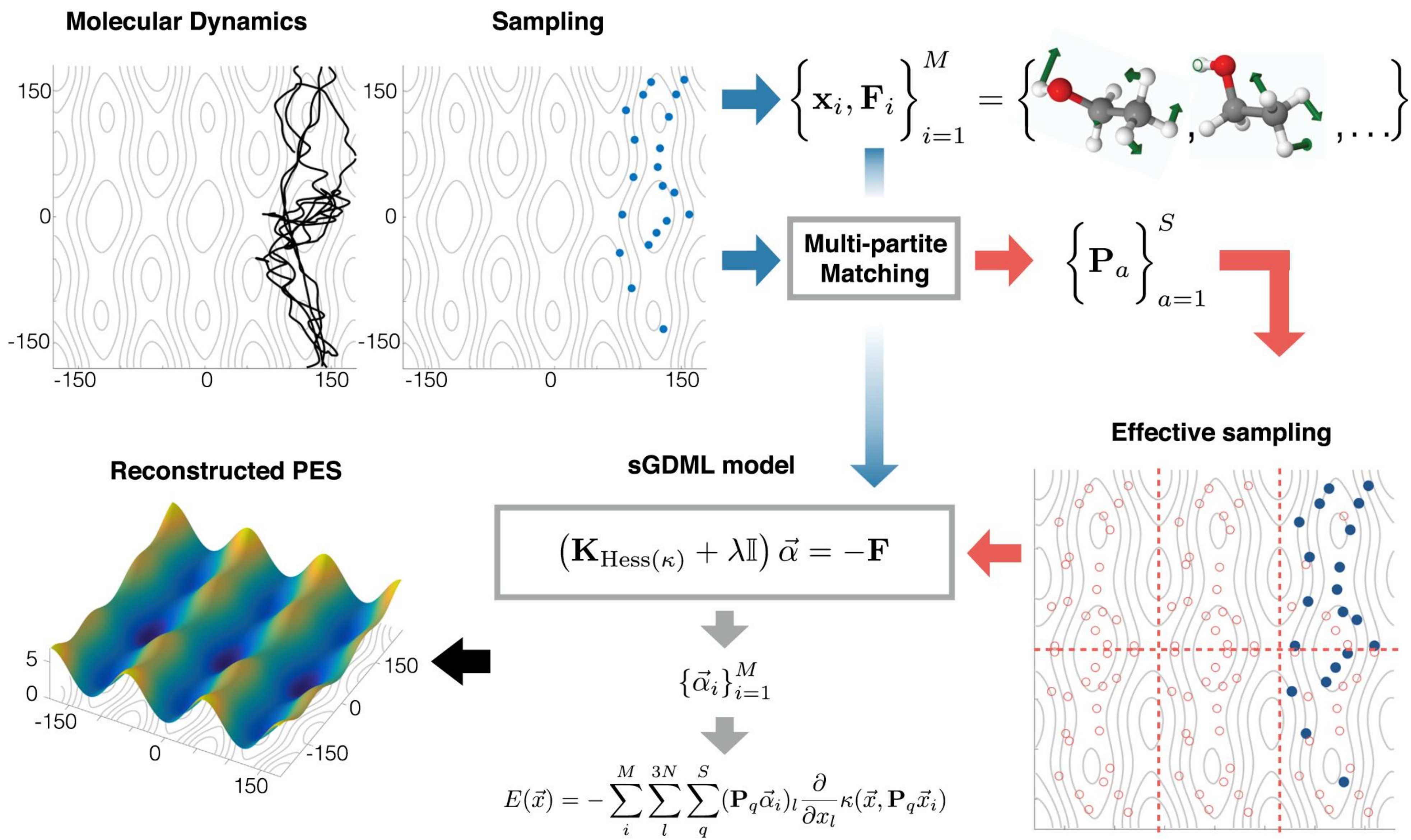
MLFFs based on Physical Principles





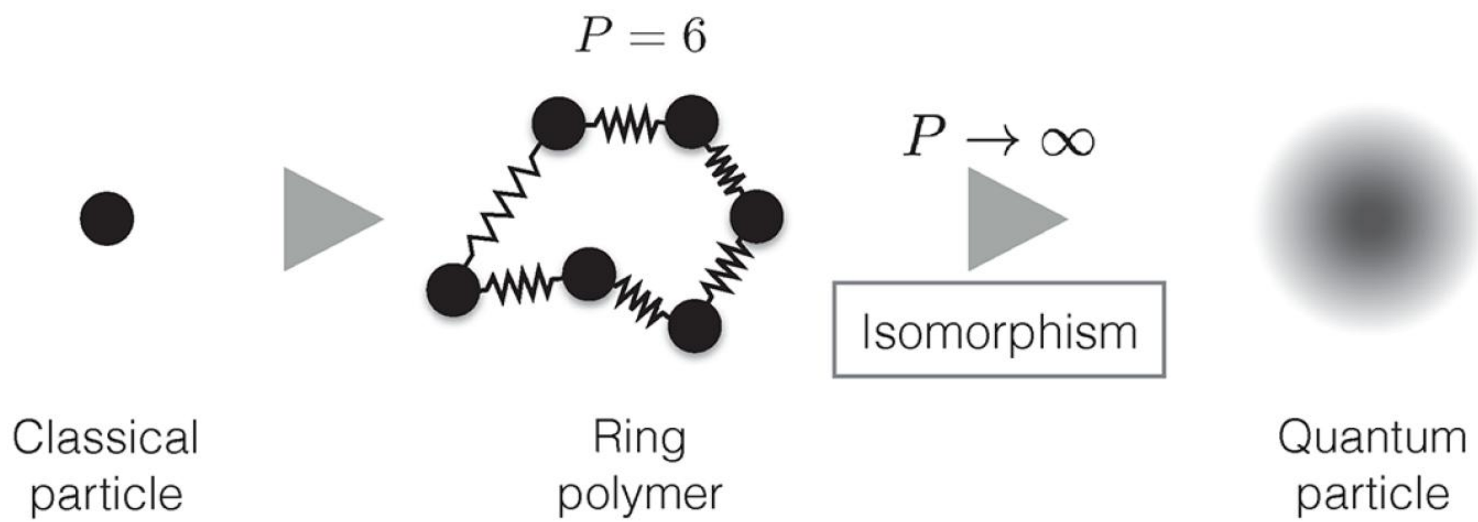
S. Chmiela *et al.*, ***Science Adv.* 3**, e1603015 (2017); ***Nature Commun.* 9**, 3887 (2018).

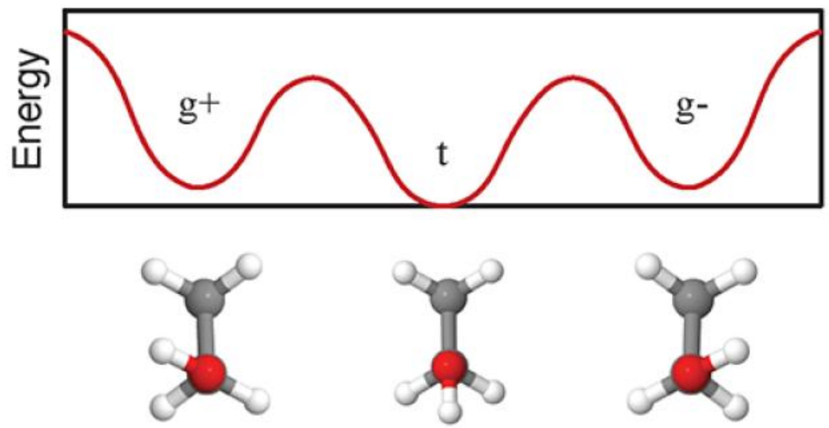
sGDML: MLFF based on Physical Laws and Symmetries



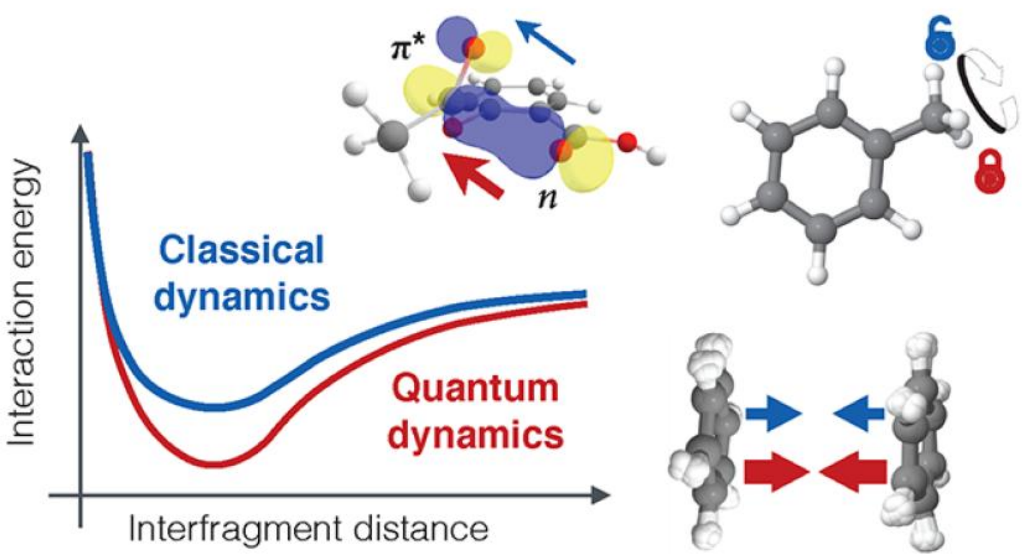
S. Chmiela, H. Sauceda, K.-R. Mueller, and A. Tkatchenko, ***Nature Commun.*** 9, 3887 (2018).

sGDML: Quantum Simulations with Exact Electrons and Nuclei



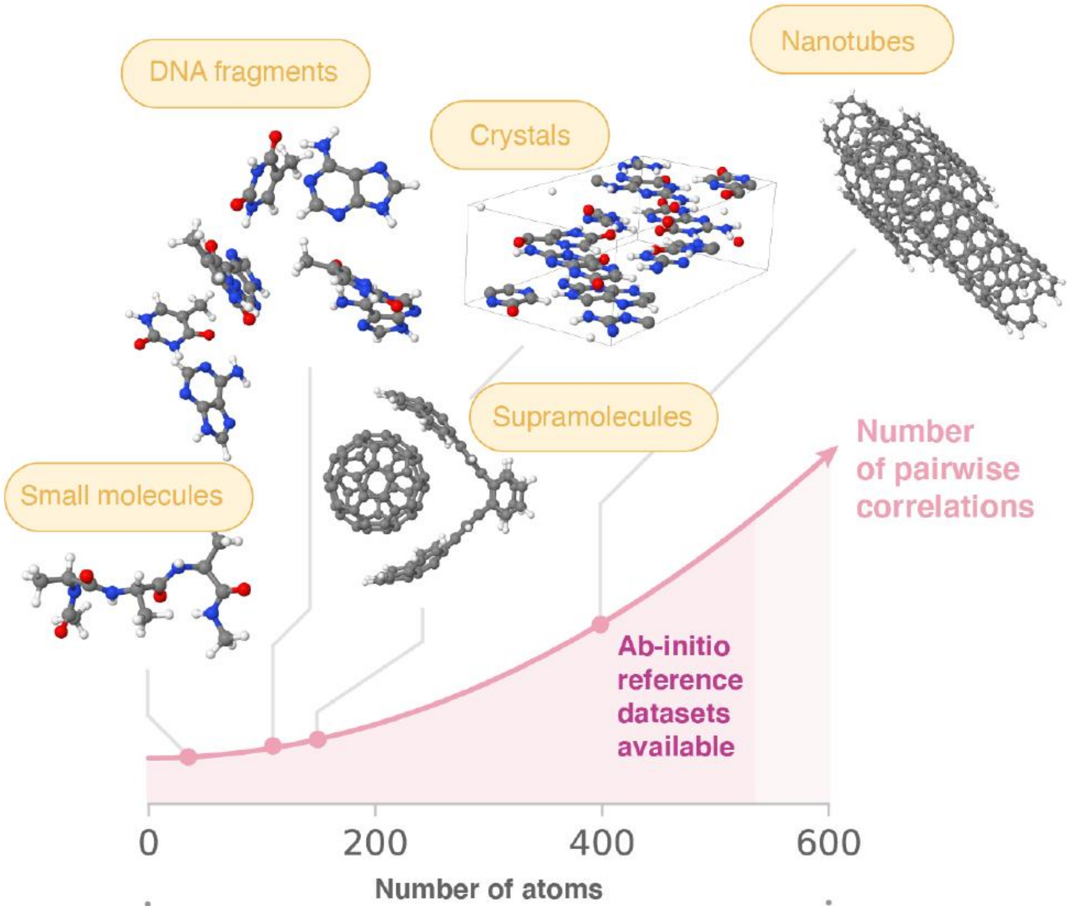


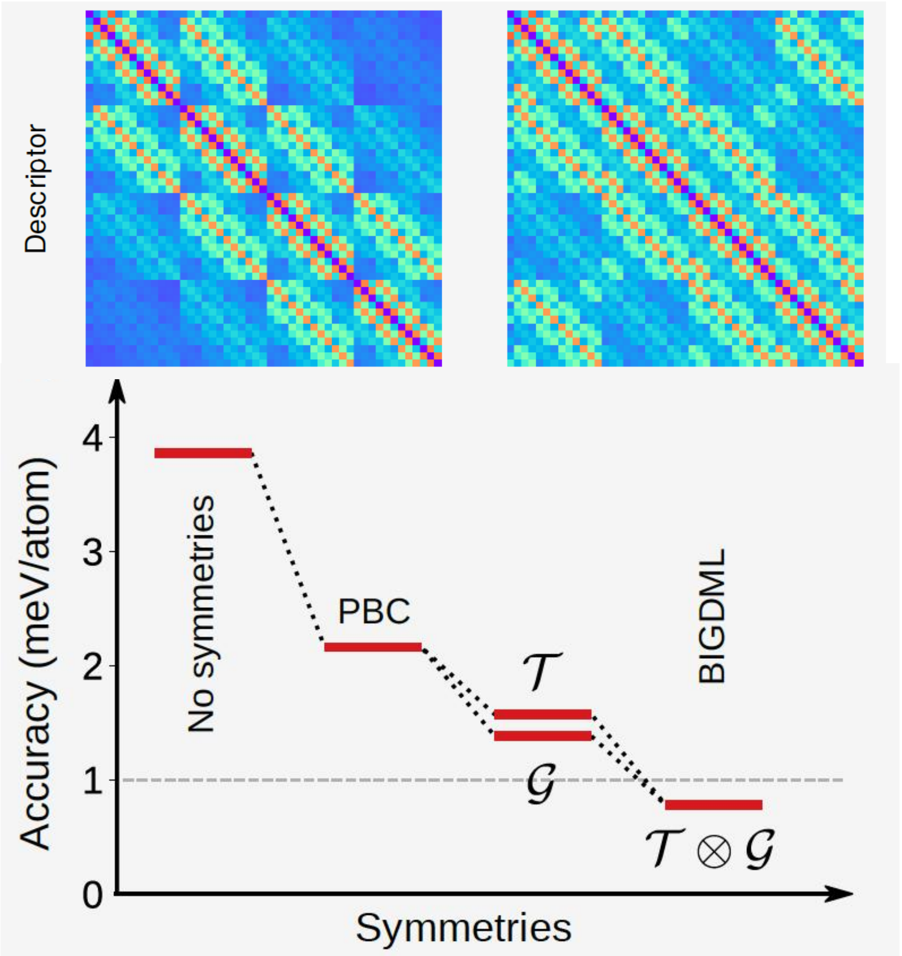
PIMD	30.3 %	39.4 %	30.3 %
Experiment	30 %	40 %	30 %



H. Sauceda, V. Vassilev-Galindo, S. Chmiela, K.-R. Mueller, and A. Tkatchenko, ***Nature Commun.*** **12**, 442 (2021).

sGDML: One Kernel for Molecules and Materials



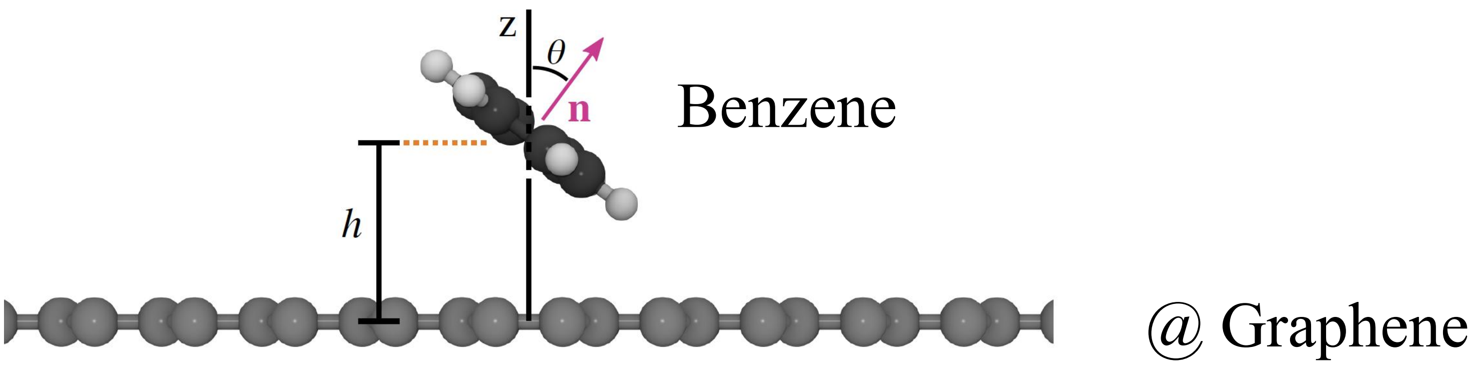


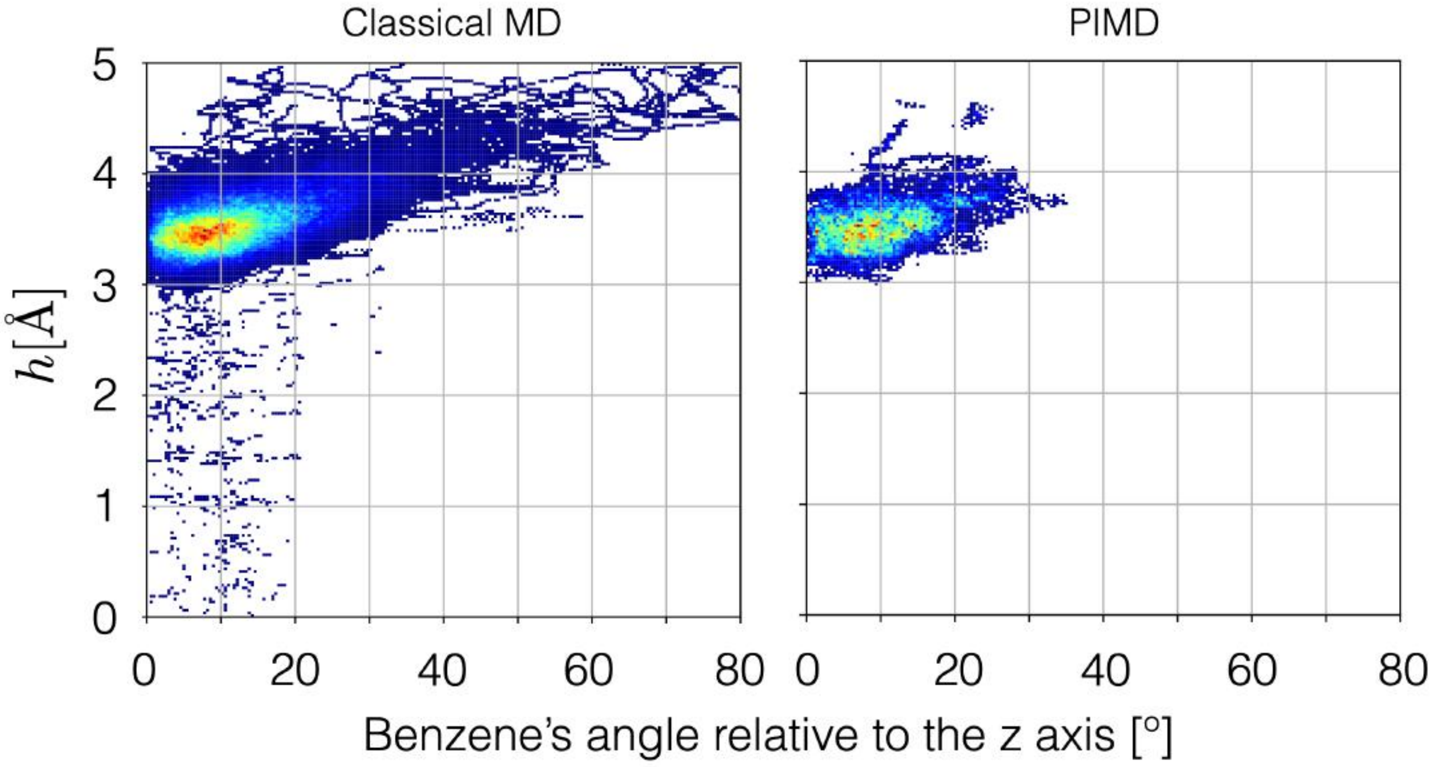
Preconditioned Large-scale sGDML

sGDML for Materials: BIGDML

S. Chmiela *et al.*, to be submitted ; H. E. Sauceda *et al.*, ***Nature Commun. (2022)***.

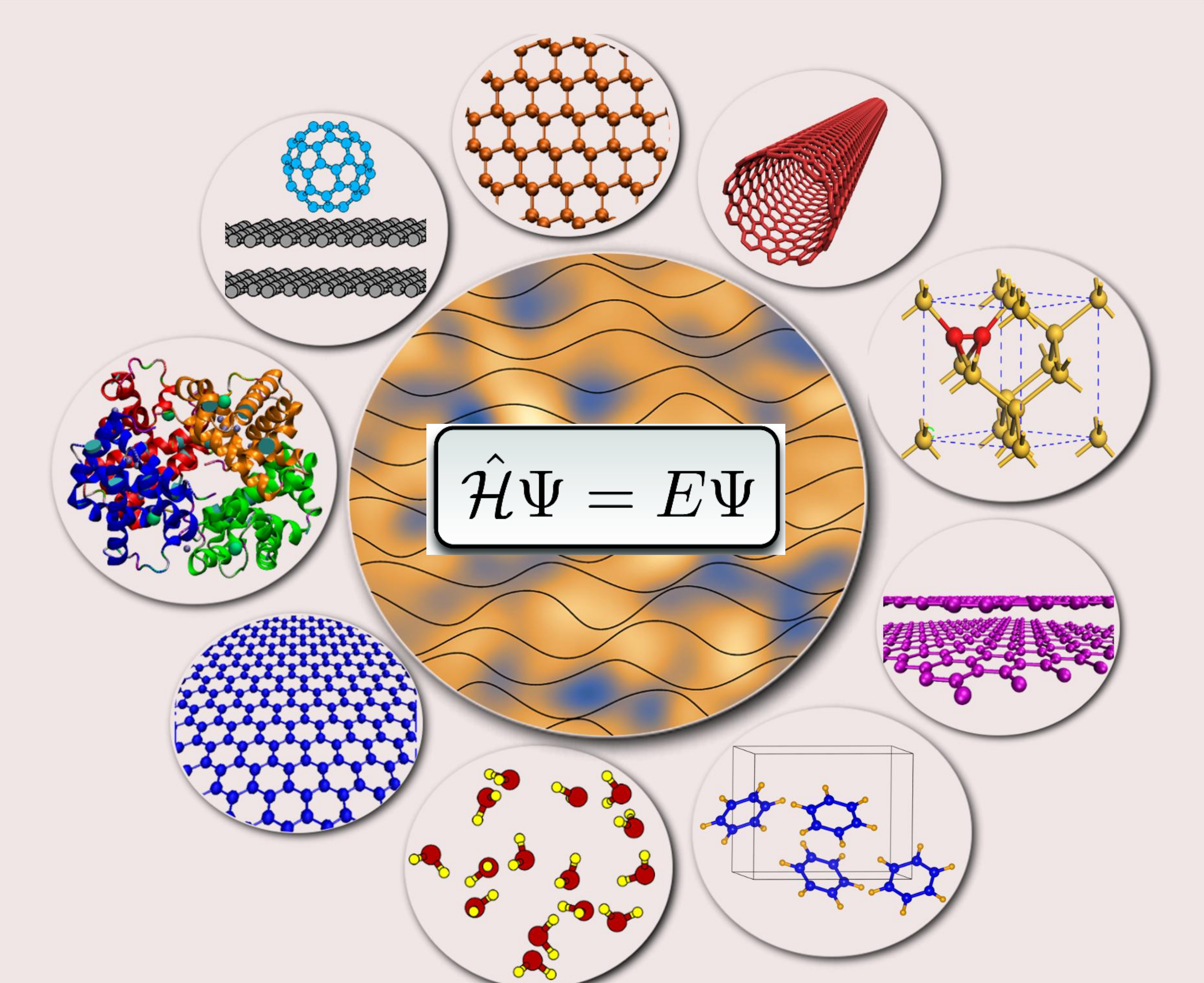
sGDML: One Kernel for Molecules and Materials





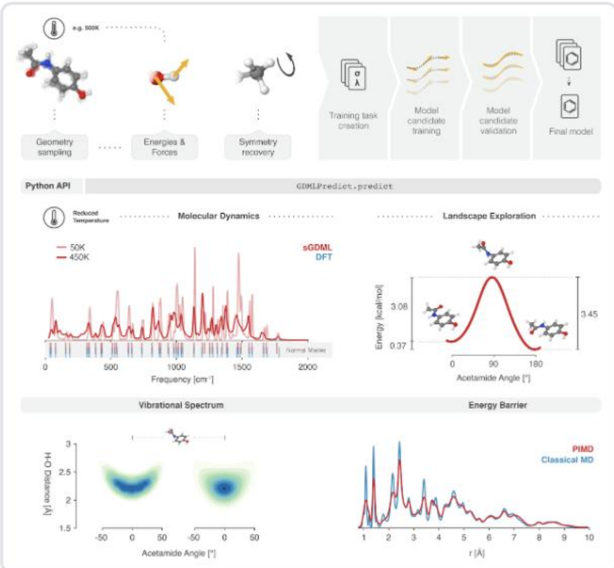
H. E. Sauceda *et al.*, ***Nature Commun. (2022)***.

Challenge: Integrating ML and Physics for Complex Molecules and Materials



Science 351, 1171 (2016); *Annu. Rev. Mater. Res.* 49, 30 (2019).

Software Available: sGDML and SchNetPack

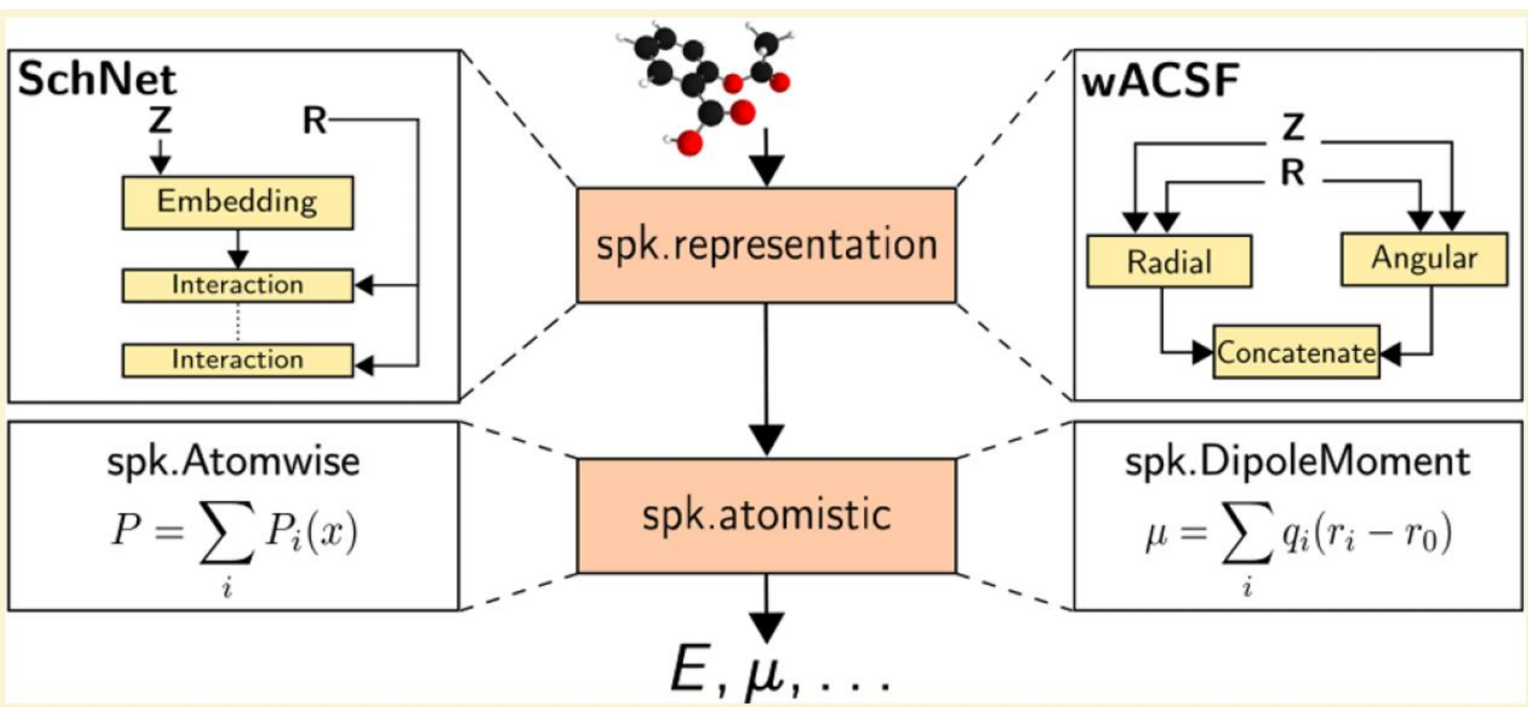


sGDML: Constructing Accurate and Data Efficient Molecular Force Fields Using Machine Learning

Chmiela, S., Sauceda, H. E., Poltavsky, I., Müller, K.-R., Tkatchenko, A., *Computer Physics Communications*, **240**, 2019, 38-45.

Article | Supplement | BibTex

www.sgdml.org



GitHub: SchNetPack

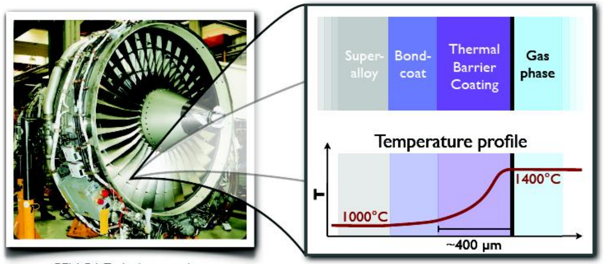
Distilling Chemistry and Physics out of ML

Alexandre Tkatchenko *et al.*

Chair for Theoretical Chemical Physics (TCP), *Dept of Physics and Materials Science* (DPhyMS), *University of Luxembourg*

www.tcpunilu.com
alexandre.tkatchenko@uni.lu

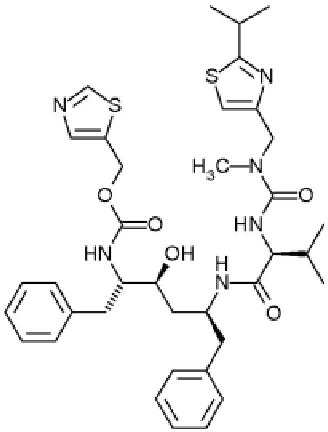
Tsinghua University, Lecture 2

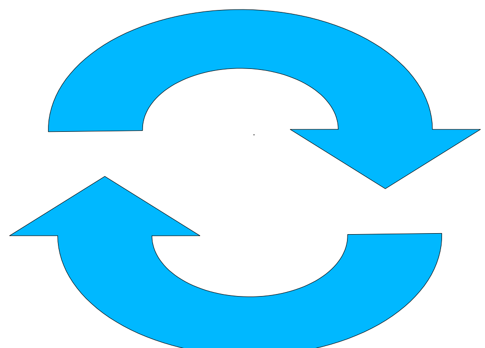


CFM 56-7 airplane engine

$$\hat{\mathcal{H}}\Psi = E\Psi$$

Quantum Mechanics





Statistical Mechanics

Machine Learning

$$Z = \mathrm{Tr}(e^{-\beta \hat{H}})$$

"Nearsightedness" of Electronic Matter

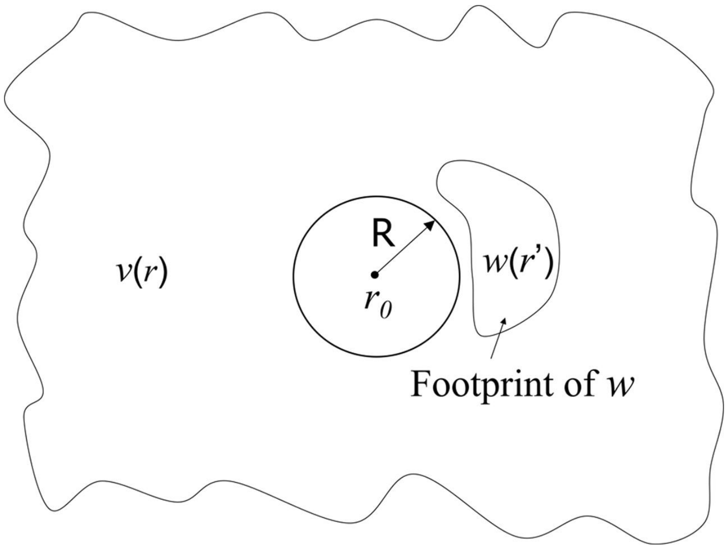
Nearsightedness of electronic matter

E. Prodan and W. Kohn

+ See all authors and affiliations

PNAS August 16, 2005 102 (33) 11635-11638; https://doi.org/10.1073/pnas.0505436102

Contributed by W. Kohn, June 28, 2005



"Nearsightedness" of Electronic Matter

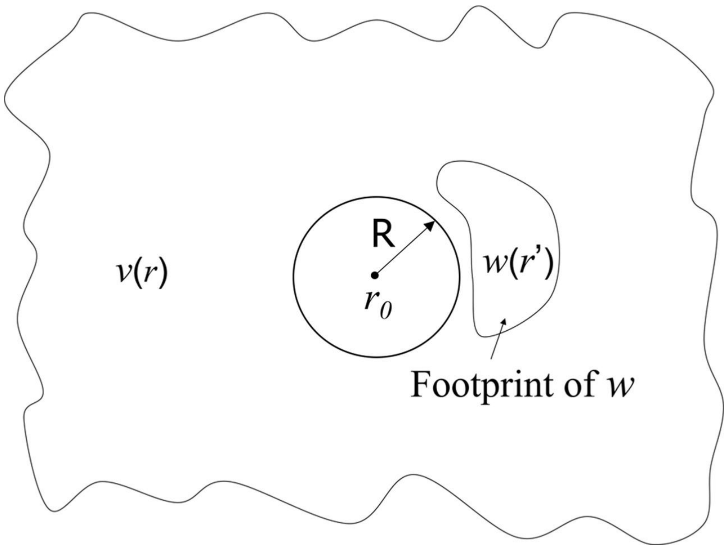
Nearsightedness of electronic matter

E. Prodan and W. Kohn

+ See all authors and affiliations

PNAS August 16, 2005 102 (33) 11635-11638; https://doi.org/10.1073/pnas.0505436102

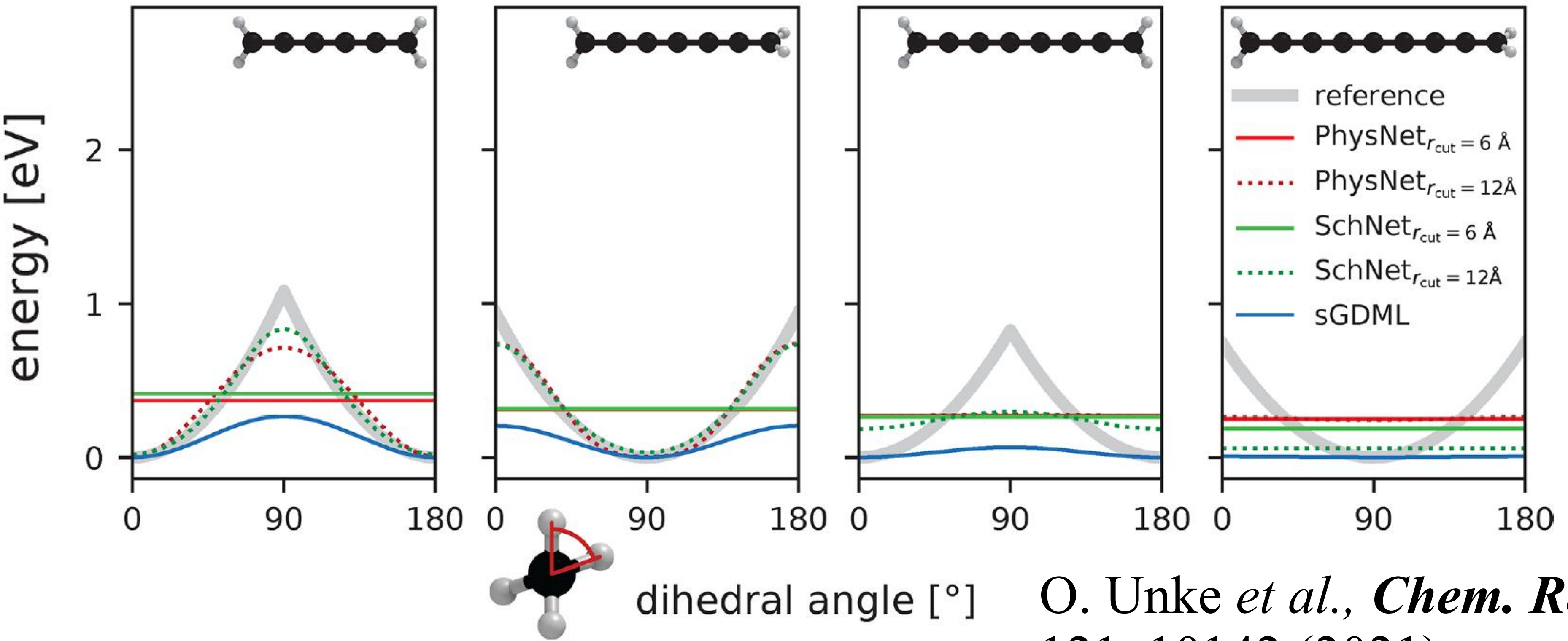
Contributed by W. Kohn, June 28, 2005



Interacting Fermions

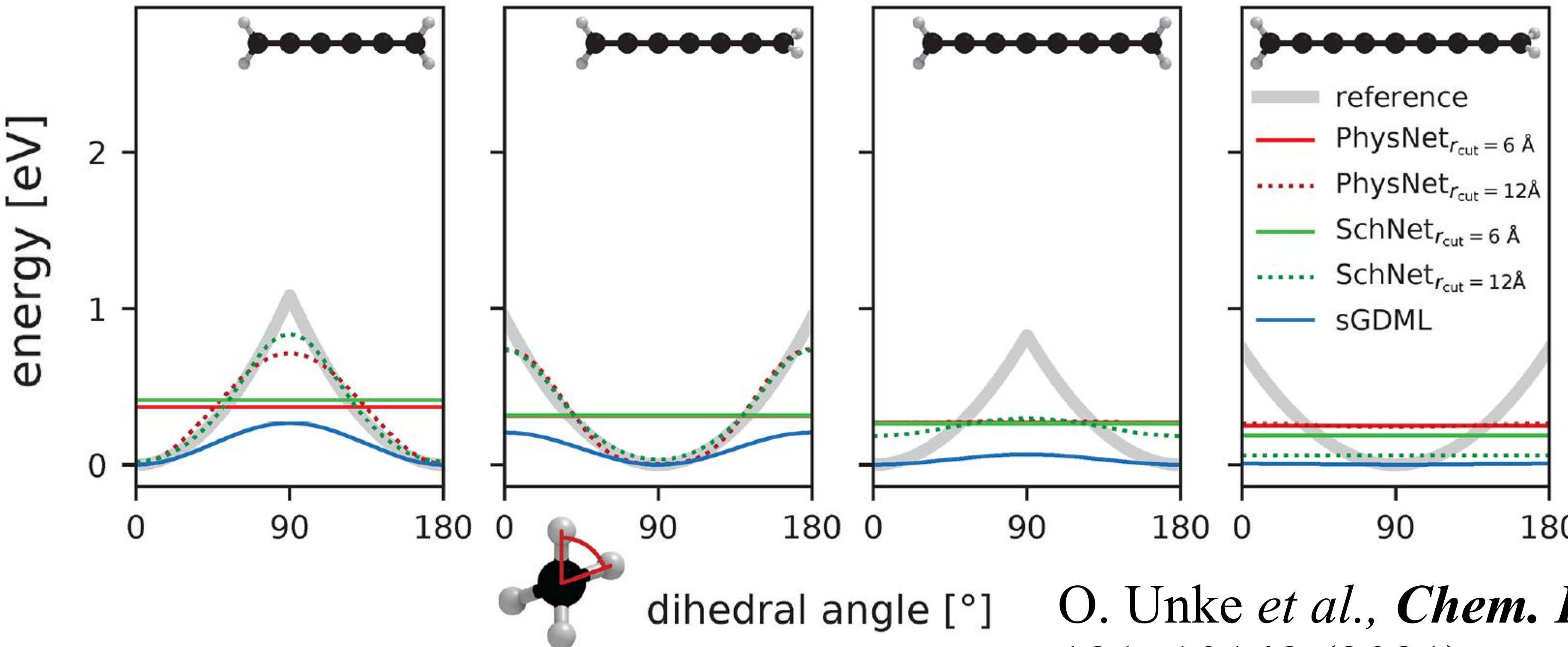
As is well known, the long-range Coulomb potential, because of perturbing electric charges, is screened out by metallic electrons. Preliminary model calculations for metallic electrons, in the Thomas–Fermi approximation, indicate that they are charge-nearsighted, i.e., have a finite R_c. However, charged *insulating* fermions are "classically farsighted," in the sense that, at sufficiently large distances, the fermions "see" the classical long-range total potential $\int \rho_t(r')/|r_0 - r'|dr'$, where ρ_t is the total perturbing charge density, including depolarization. Thus, for example, in metals, replacing a neutral atom or ion by another atom or ion always has short-range electronic consequences, whereas in an insulator ions lead to classical long-range electronic effects.

"Nearsightedness" of Electronic Matter

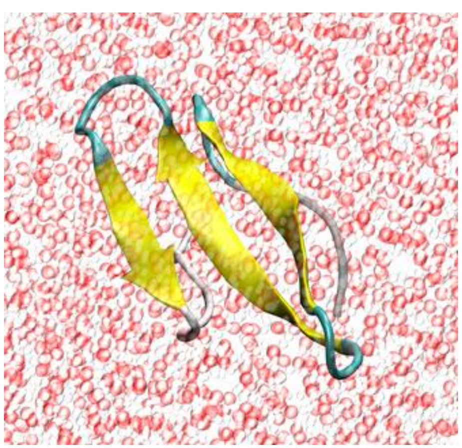


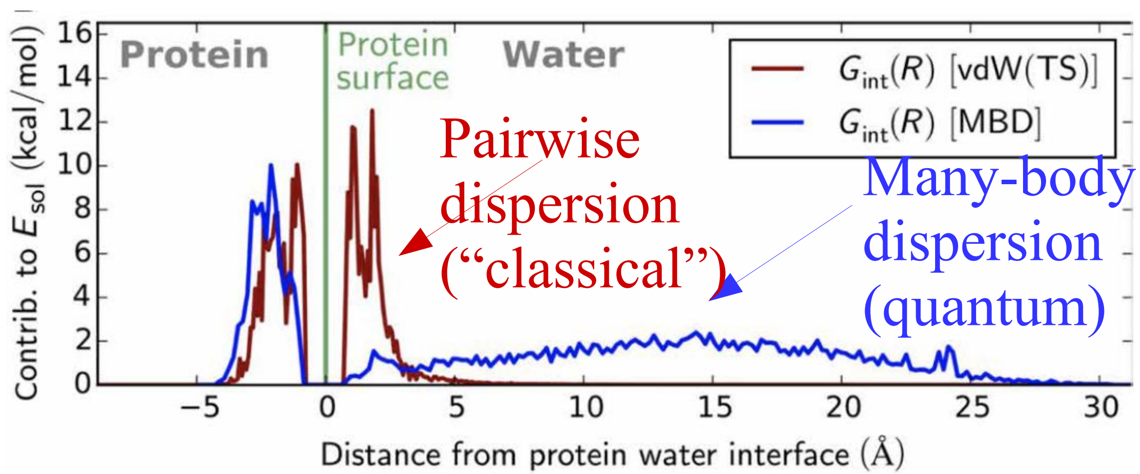
O. Unke *et al.*, ***Chem. Rev.*** 121, 10142 (2021).

“Nearsightedness” of Electronic Matter



O. Unke *et al.*, ***Chem. Rev.*** 121, 10142 (2021).



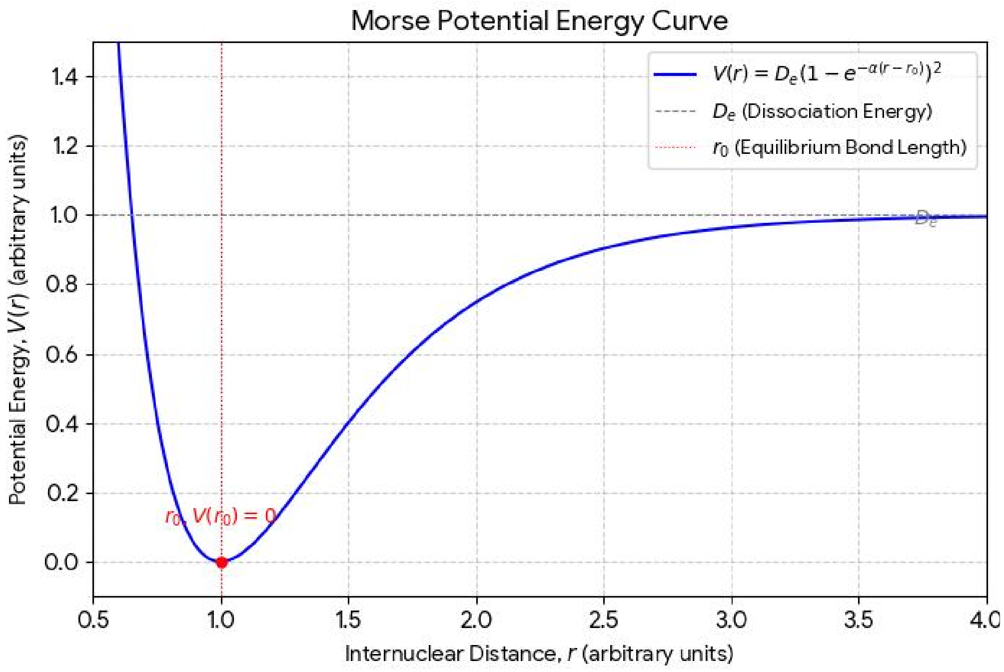


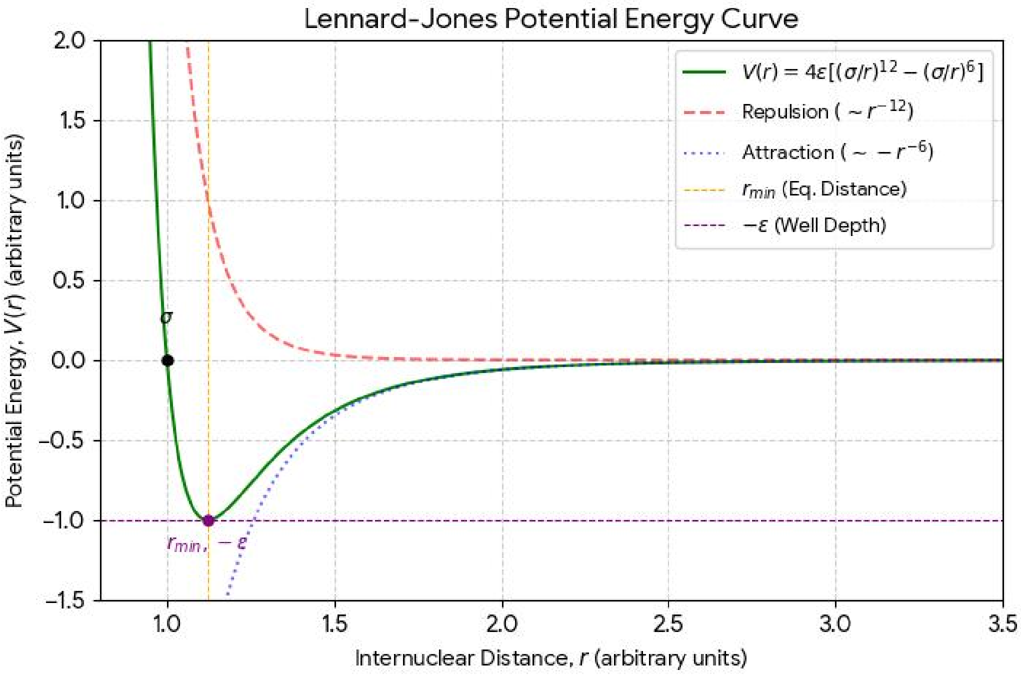
M. Stoehr and A. Tkatchenko, ***Science Adv.*** 5, eaax0024 (2019); ***Science*** 351, 1171 (2016).

Molecular Mechanics *in Vacuo*

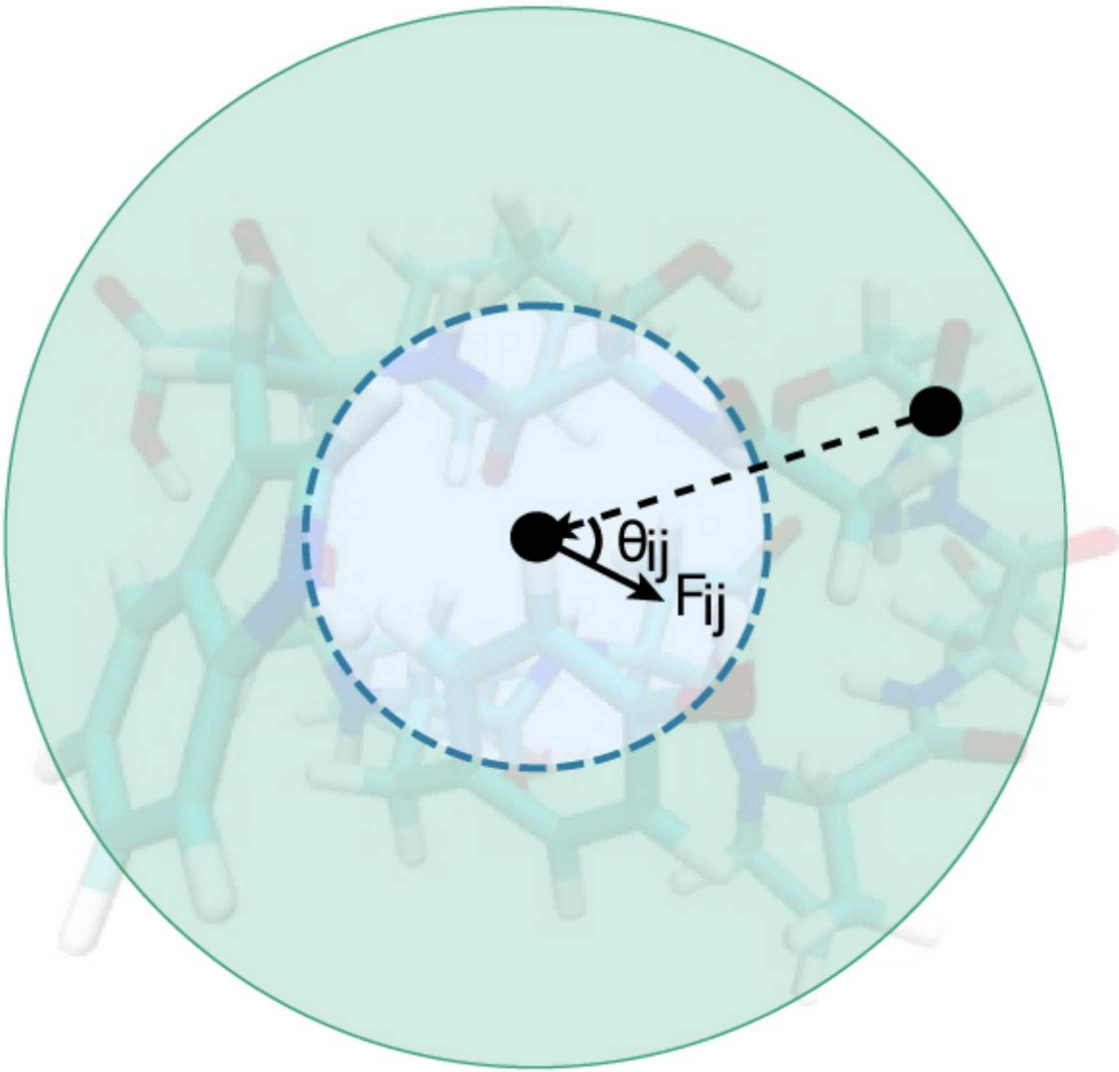
Bonded potentials:
bonds, angles, and torsions

Non-Bonded potentials:
van der Waals and electrostatics





Interatomic Interactions: Range and (An)Isotropy



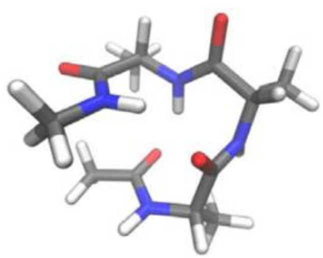
Short Range (< 5 Å) - Anisotropic
Long Range (> 5 Å) - Assumed Isotropic

Efficient interatomic descriptors for accurate machine learning force fields of extended molecules

Adil Kabylda, Valentin Vassilev-Galindo, Stefan Chmiela, Igor Poltavsky & Alexandre Tkatchenko

Nature Communications **14**, Article number: 3562 (2023)

Ac-Ala3-NHMe



Reference Data:
DFT-PBE+MBD

SCIENCE ADVANCES | RESEARCH ARTICLE

CHEMICAL PHYSICS

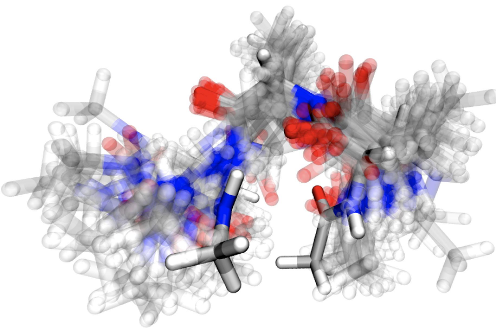
Accurate global machine learning force fields for molecules with hundreds of atoms

Stefan Chmiela[1,2]*, Valentin Vassilev-Galindo[3], Oliver T. Unke[4,1], Adil Kabylda[3], Huziel E. Sauceda[5,1,6,2], Alexandre Tkatchenko[3]*, Klaus-Robert Müller[1,2,4,7,8]*

Efficient interatomic descriptors for accurate machine learning force fields of extended molecules

Adil Kabylda, Valentin Vassilev-Galindo, Stefan Chmiela, Igor Poltavsky & Alexandre Tkatchenko

Nature Communications **14**, Article number: 3562 (2023)



Reference Data:
DFT-PBE+MBD

SCIENCE ADVANCES | RESEARCH ARTICLE

CHEMICAL PHYSICS

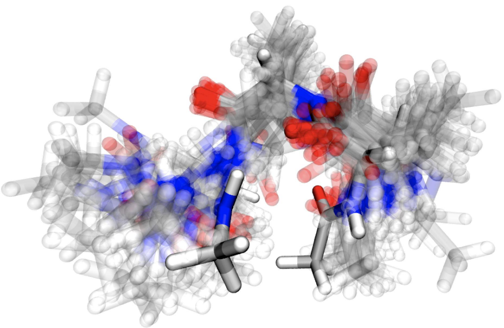
Accurate global machine learning force fields for molecules with hundreds of atoms

Stefan Chmiela[1,2]*, Valentin Vassilev-Galindo[3], Oliver T. Unke[4,1], Adil Kabylda[3], Huziel E. Sauceda[5,1,6,2], Alexandre Tkatchenko[3]*, Klaus-Robert Müller[1,2,4,7,8]*

Efficient interatomic descriptors for accurate machine learning force fields of extended molecules

Adil Kabylda, Valentin Vassilev-Galindo, Stefan Chmiela, Igor Poltavsky & Alexandre Tkatchenko

Nature Communications **14**, Article number: 3562 (2023)

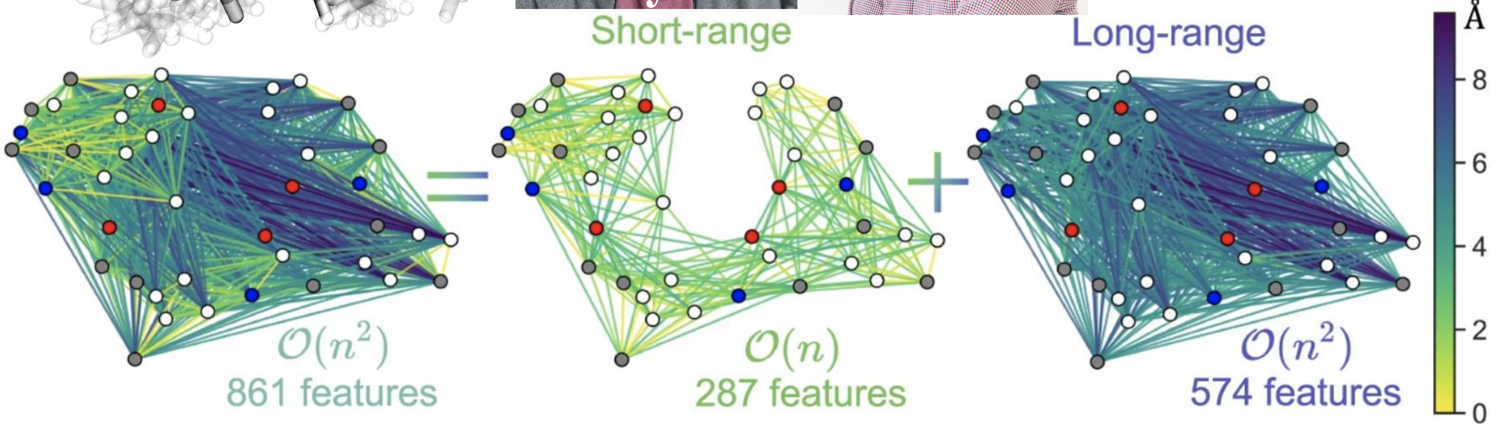


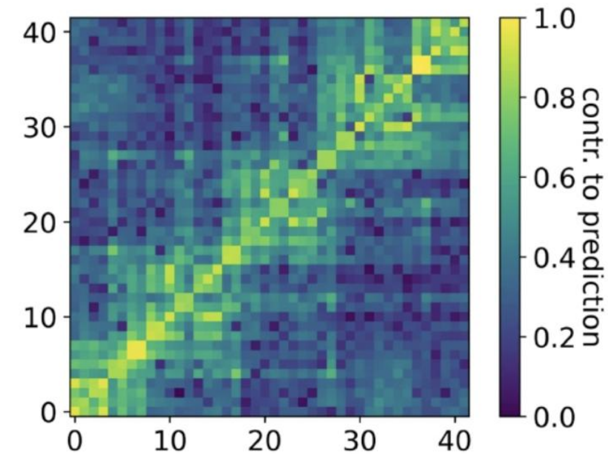
SCIENCE ADVANCES | RESEARCH ARTICLE

CHEMICAL PHYSICS

Accurate global machine learning force fields for molecules with hundreds of atoms

Stefan Chmiela[1,2]*, Valentin Vassilev-Galindo[3], Oliver T. Unke[4,1], Adil Kabylda[3], Huziel E. Sauceda[5,1,6,2], Alexandre Tkatchenko[3]*, Klaus-Robert Müller[1,2,4,7,8]*





Efficient interatomic descriptors for accurate machine learning force fields of extended molecules

Adil Kabylda, Valentin Vassilev-Galindo, Stefan Chmiela, Igor Poltavsky & Alexandre Tkatchenko

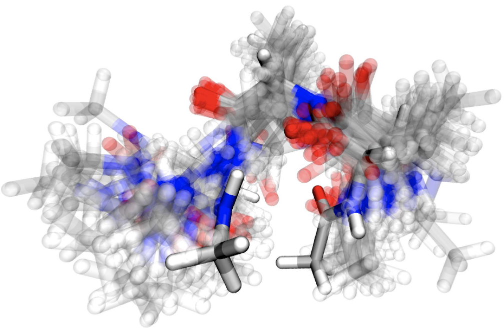
Nature Communications 14, Article number: 3562 (2023)

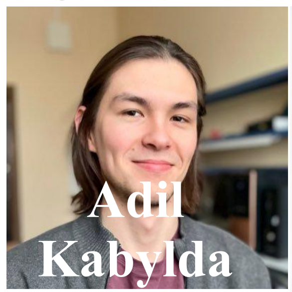
SCIENCE ADVANCES | RESEARCH ARTICLE

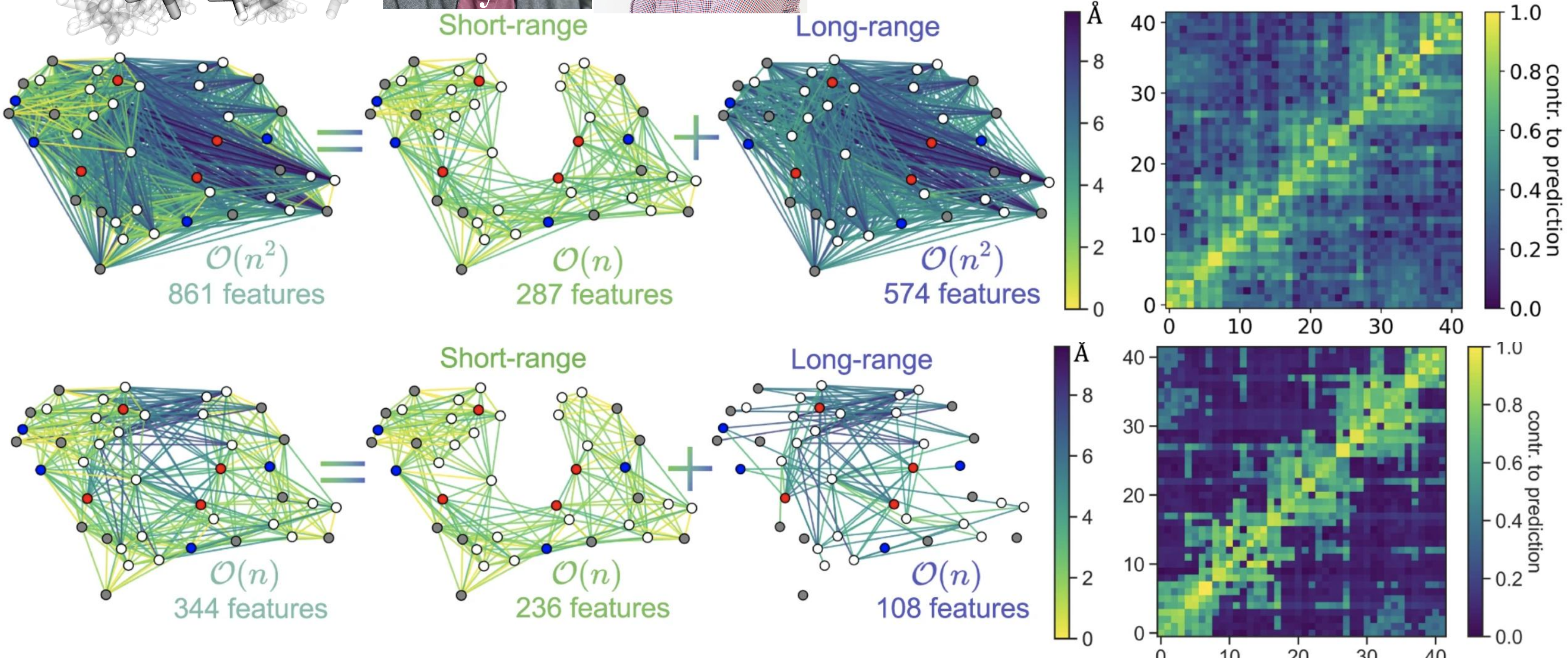
CHEMICAL PHYSICS

Accurate global machine learning force fields for molecules with hundreds of atoms

Stefan Chmiela[1,2]*, Valentin Vassilev-Galindo[3], Oliver T. Unke[4,1], Adil Kabylda[3], Huziel E. Sauceda[5,1,6,2], Alexandre Tkatchenko[3]*, Klaus-Robert Müller[1,2,4,7,8]*



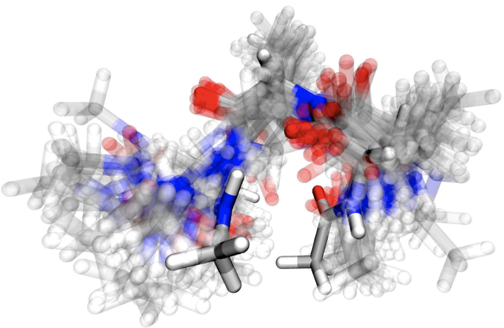




Efficient interatomic descriptors for accurate machine learning force fields of extended molecules

Adil Kabylda, Valentin Vassilev-Galindo, Stefan Chmiela, Igor Poltavsky & Alexandre Tkatchenko

Nature Communications **14**, Article number: 3562 (2023)

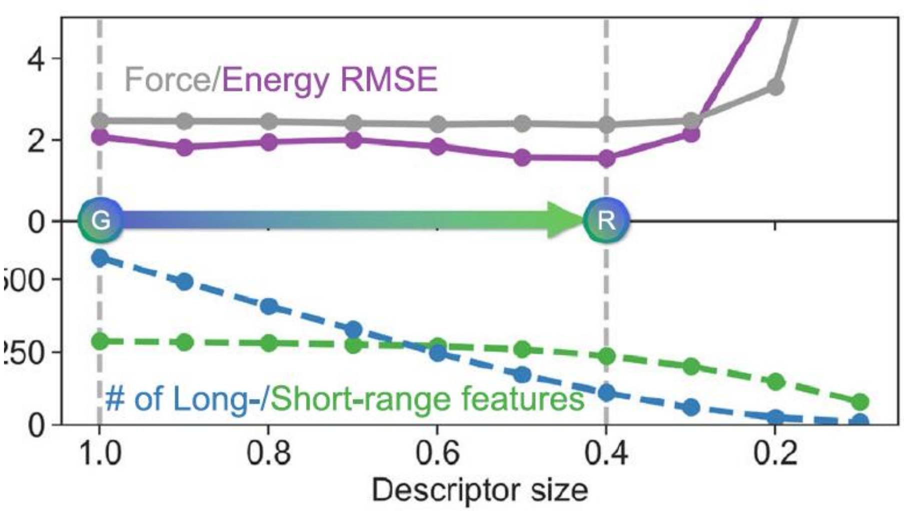


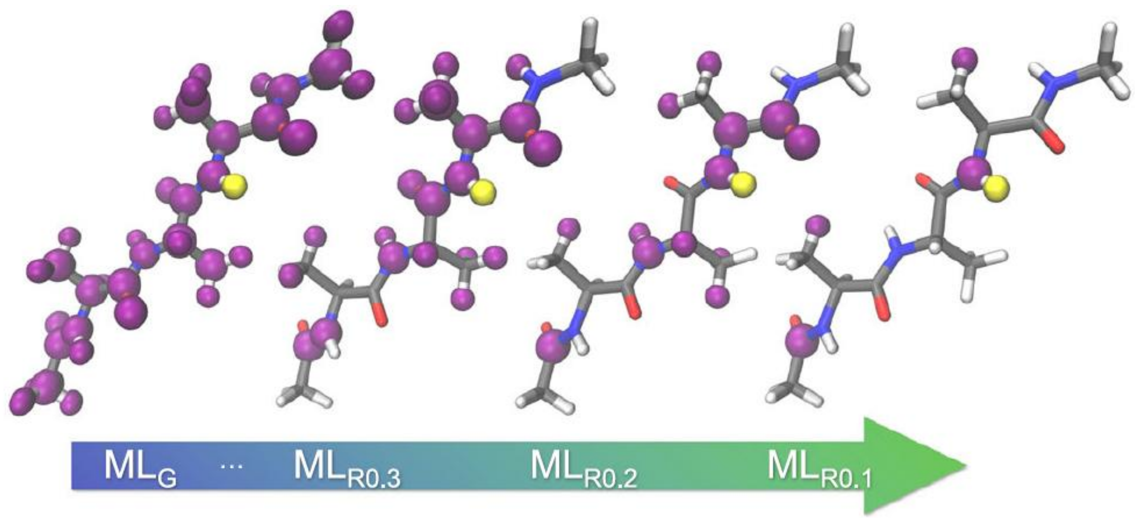
SCIENCE ADVANCES | RESEARCH ARTICLE

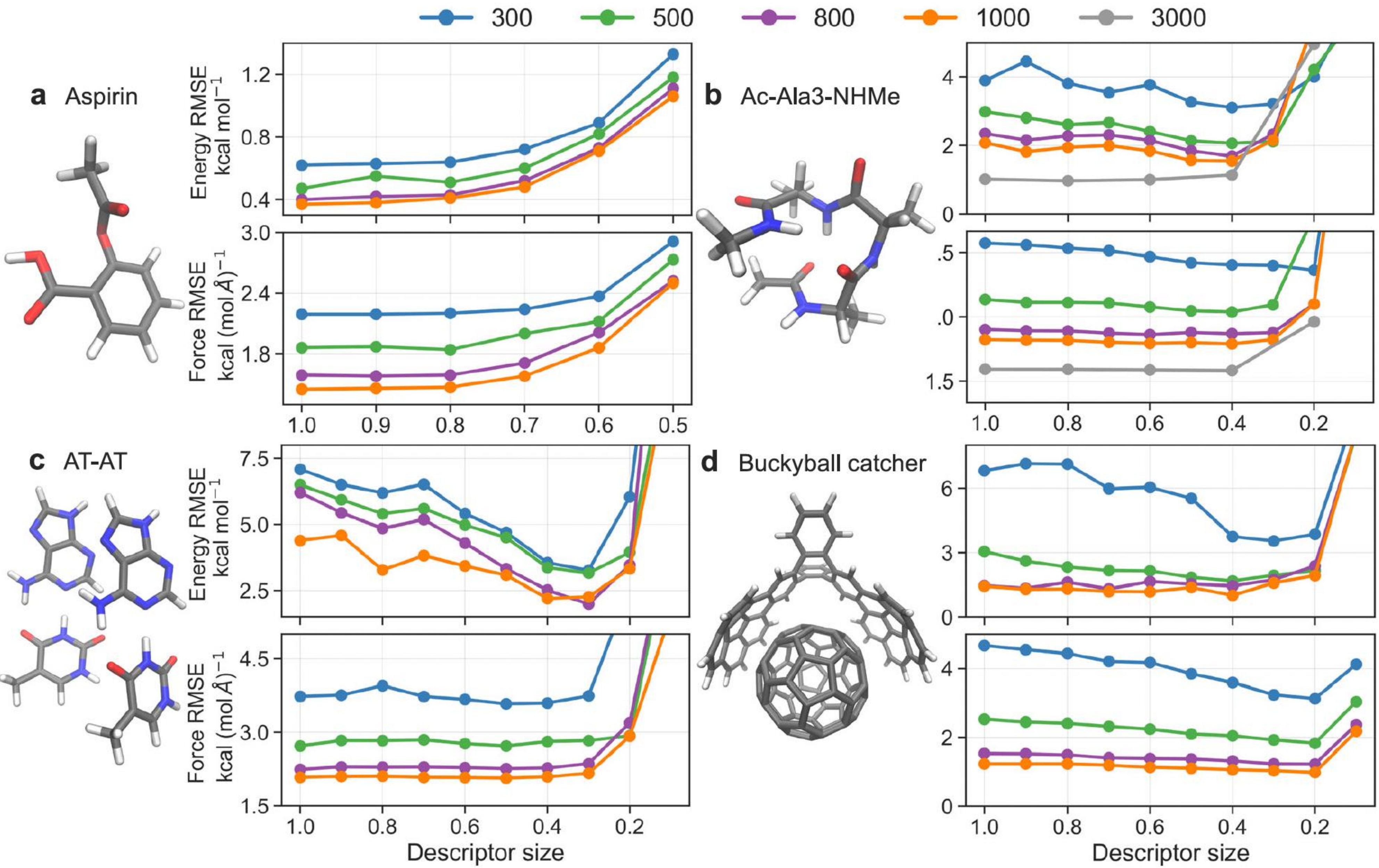
CHEMICAL PHYSICS

Accurate global machine learning force fields for molecules with hundreds of atoms

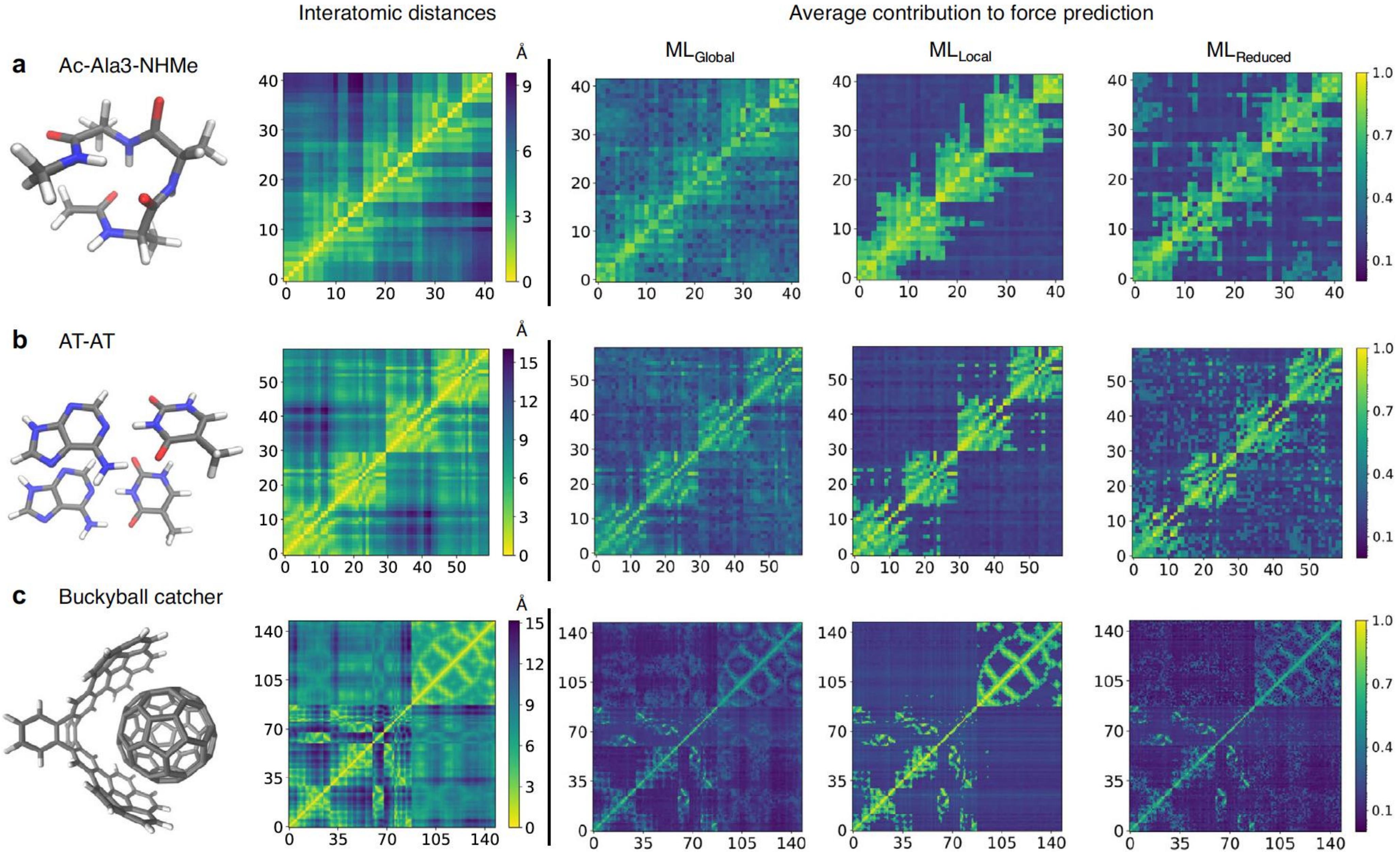
Stefan Chmiela[1,2]*, Valentin Vassilev-Galindo[3], Oliver T. Unke[4,1], Adil Kabylda[3], Huziel E. Sauceda[5,1,6,2], Alexandre Tkatchenko[3]*, Klaus-Robert Müller[1,2,4,7,8]*







300
500
800
1000
3000
a Aspirin
b Ac-Ala3-NHMe
c AT-AT
d Buckyball catcher
Energy RMSE kcal mol⁻¹
Force RMSE kcal (mol Å)⁻¹
Descriptor size



Interatomic distances
Average contribution to force prediction
a
Ac-Ala3-NHMe
b
AT-AT
c
Buckyball catcher
Å
ML_Global
ML_Local
ML_Reduced

Efficient interatomic descriptors for accurate machine learning force fields of extended molecules

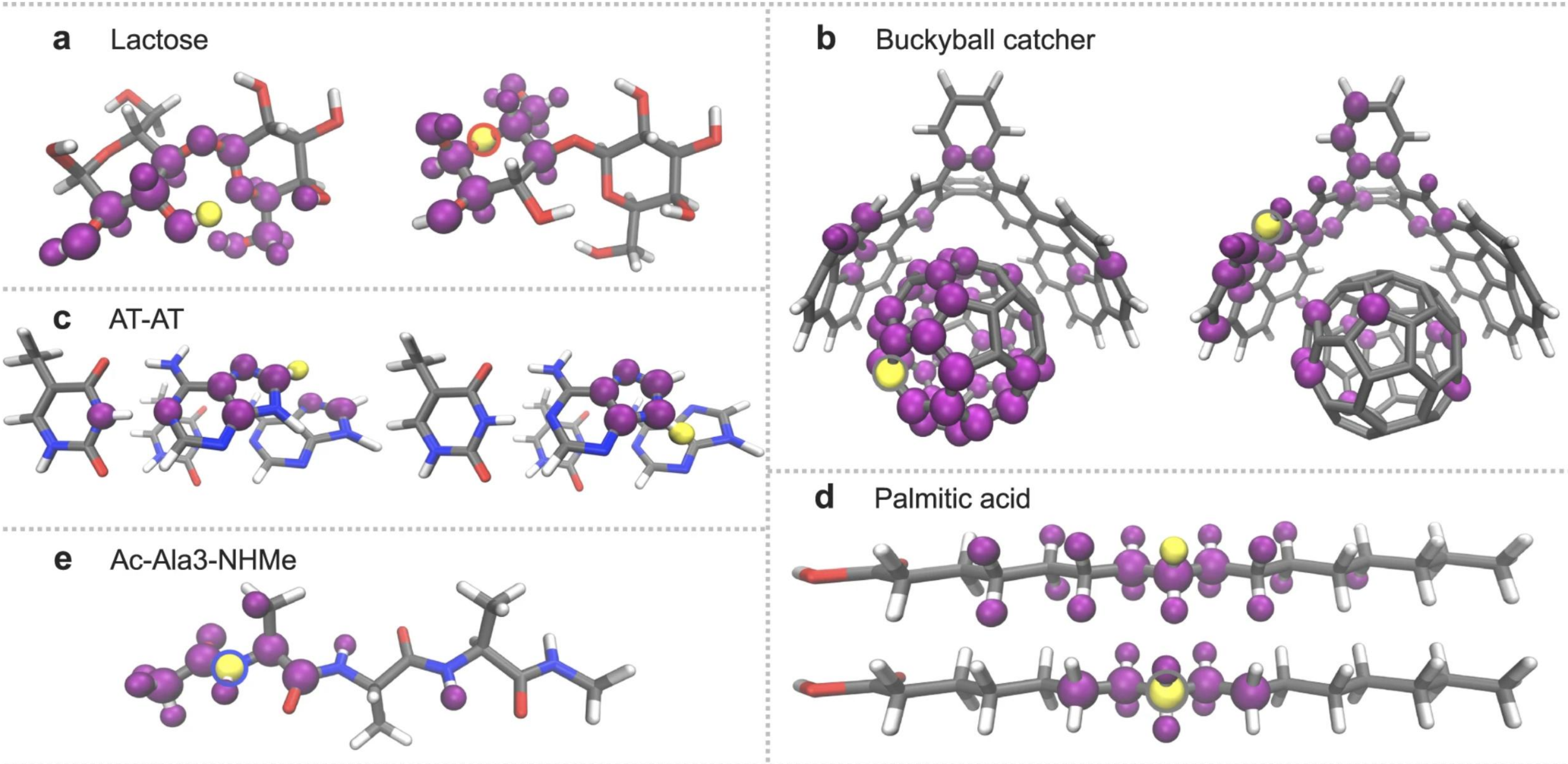
Adil Kabylda, Valentin Vassilev-Galindo, Stefan Chmiela, Igor Poltavsky & Alexandre Tkatchenko

Nature Communications **14**, Article number: 3562 (2023)

SCIENCE ADVANCES | RESEARCH ARTICLE

CHEMICAL PHYSICS

Accurate global machine learning force fields for molecules with hundreds of atoms



Efficient interatomic descriptors for accurate machine learning force fields of extended molecules

Adil Kabylda, Valentin Vassilev-Galindo, Stefan Chmiela, Igor Poltavsky & Alexandre Tkatchenko

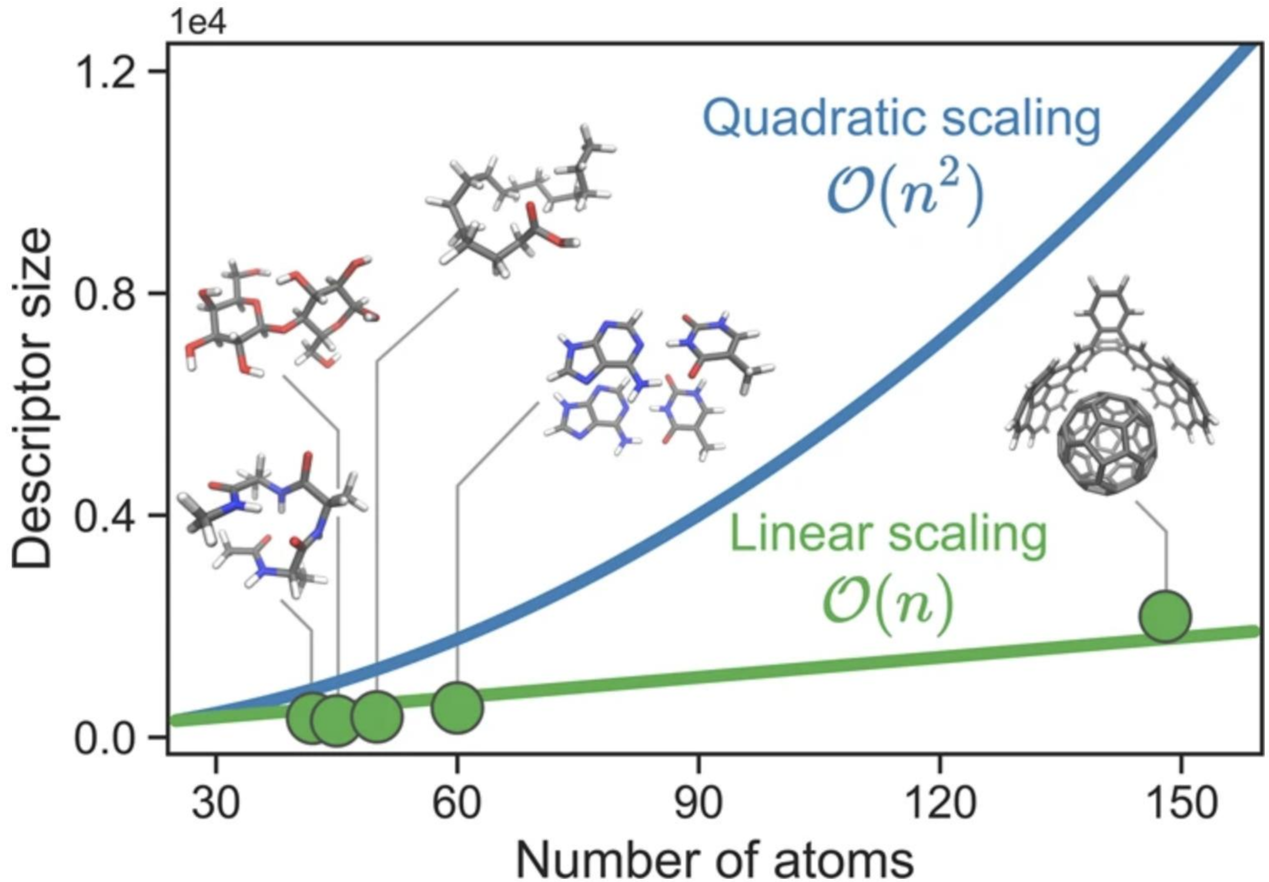
Nature Communications 14, Article number: 3562 (2023)

SCIENCE ADVANCES | RESEARCH ARTICLE

CHEMICAL PHYSICS

Accurate global machine learning force fields for molecules with hundreds of atoms

Stefan Chmiela[1,2]*, Valentin Vassilev-Galindo[3], Oliver T. Unke[4,1], Adil Kabylda[3], Huziel E. Sauceda[5,1,6,2], Alexandre Tkatchenko[3]*, Klaus-Robert Müller[1,2,4,7,8]*



Analyzing Atomic Interactions in Molecules as Learned by Neural Networks

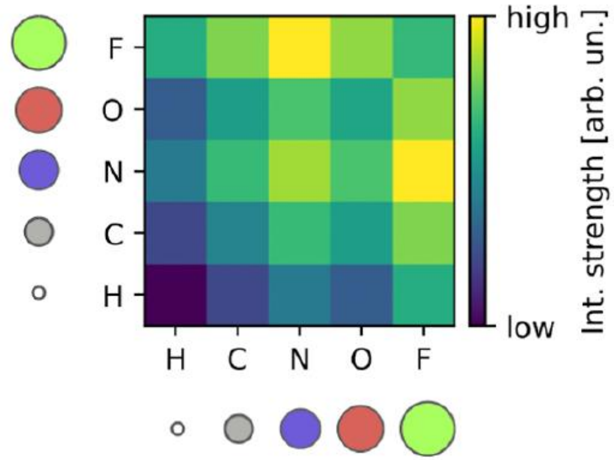
Malte Esders,* Thomas Schnake,[▽] Jonas Lederer,[▽] Adil Kabylda, Grégoire Montavon, Alexandre Tkatchenko,* and Klaus-Robert Müller*

Cite This: https://doi.org/10.1021/acs.jctc.4c01424

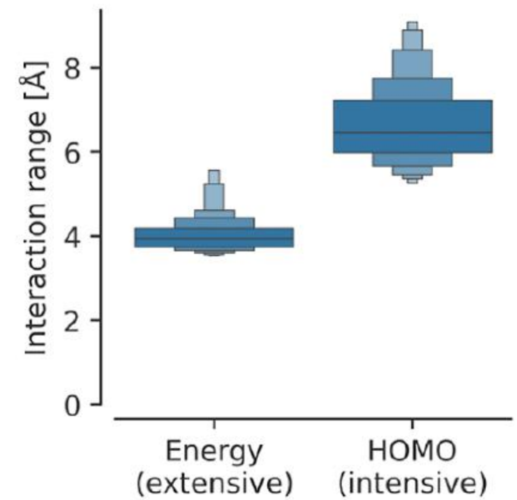
Analyzing Atomic Interactions in Molecules as Learned by Neural Networks

Malte Esders,* Thomas Schnake,∇ Jonas Lederer,∇ Adil Kabylda, Grégoire Montavon, Alexandre Tkatchenko,* and Klaus-Robert Müller*

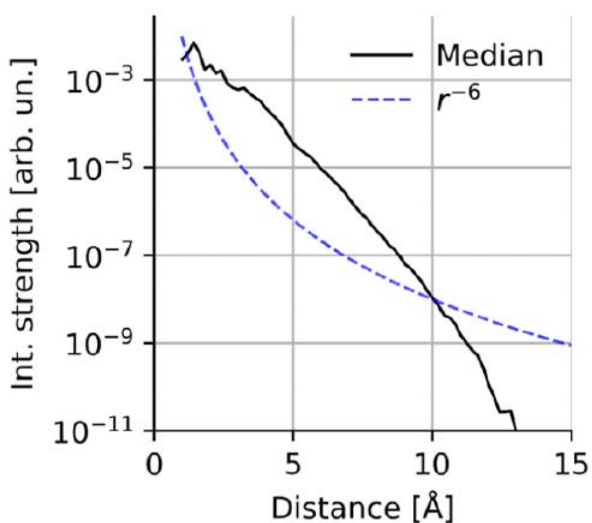
1. Interaction strength is atom-type dependent ✓



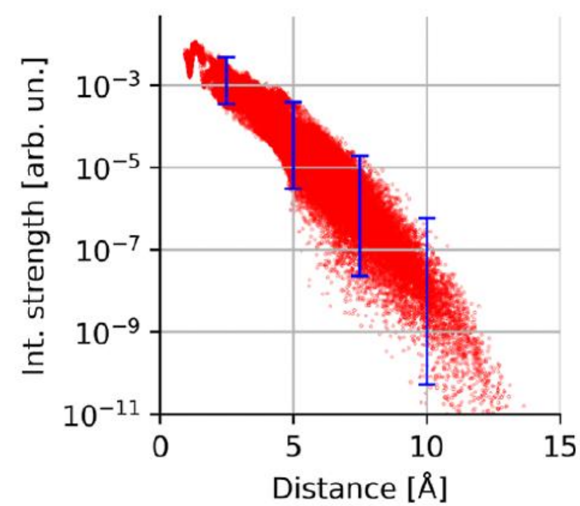
2. Intensive/Extensive property: range differs ✓

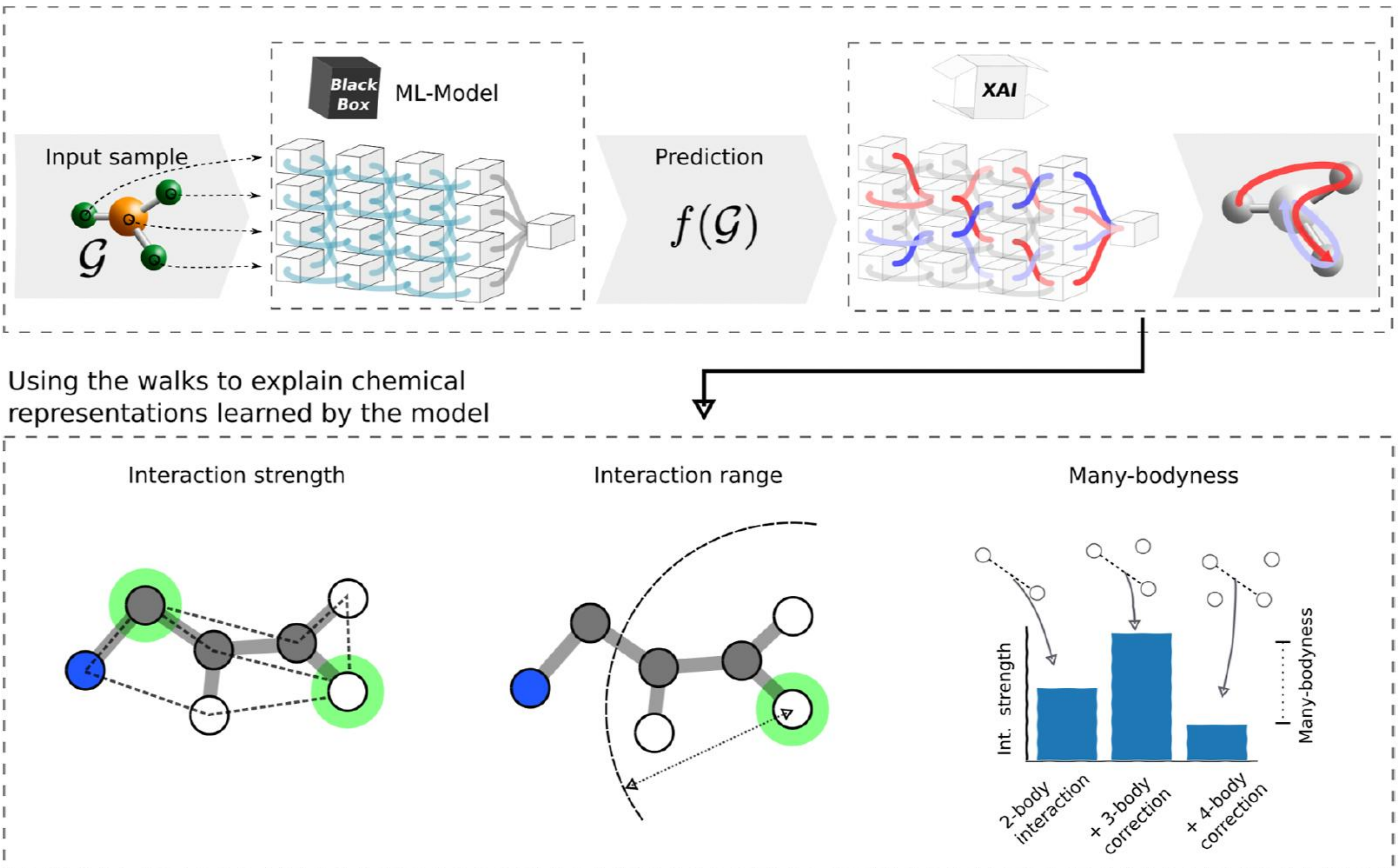


3. Interaction strength follows power law ✗

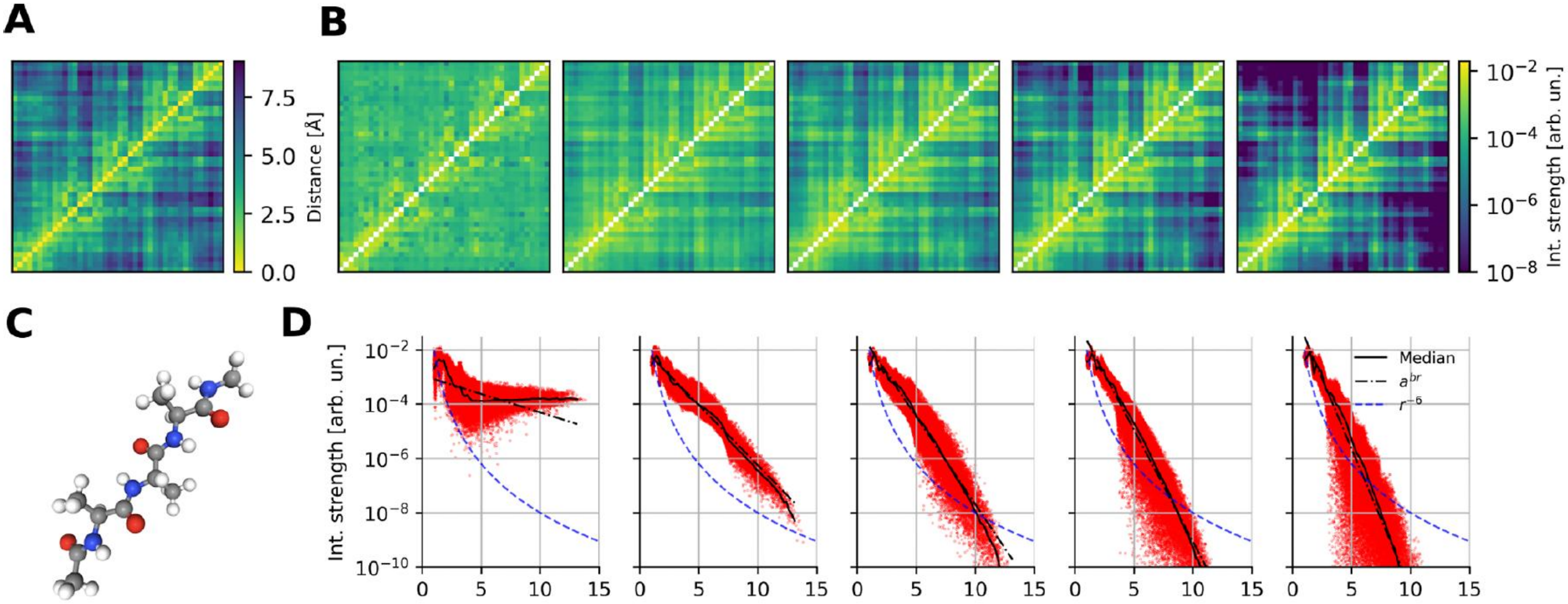


4. Many-bodyness of the interaction ✓

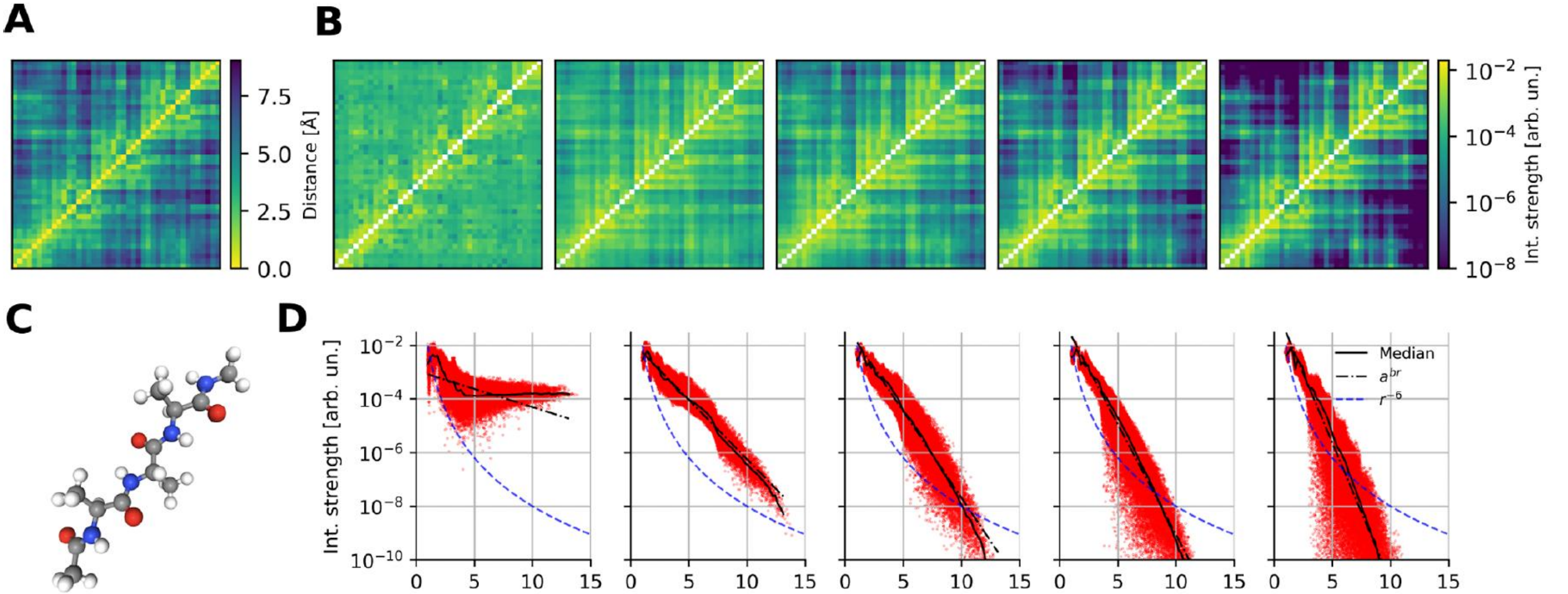




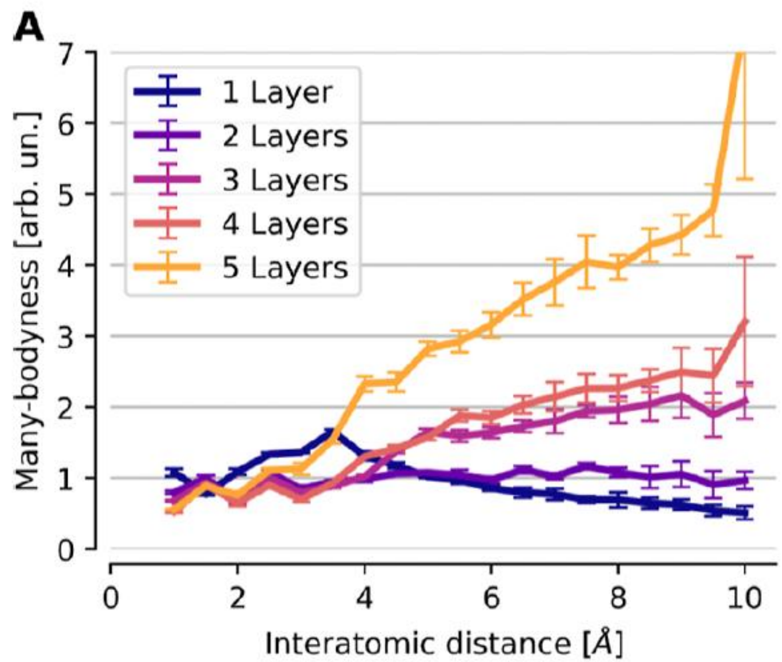
Input sample
$\mathcal{G}$
Black Box
ML-Model
Prediction
$f(\mathcal{G})$
XAI
Using the walks to explain chemical
representations learned by the model
Interaction strength
Interaction range
Many-bodyness
Int. strength
2-body interaction
+ 3-body correction
+ 4-body correction
Many-bodyness



A
B
C
D
Distance [Å]
7.5
5.0
2.5
0.0
Int. strength [arb. un.]
10^-2
10^-4
10^-6
10^-8
10^-10
0
5
10
15
Median
a^br
r^-6



A
B
C
D
Distance [Å]
7.5
5.0
2.5
0.0
Int. strength [arb. un.]
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
Median
a^{br}
r^{-6}
0
5
10
15



A
1 Layer
2 Layers
3 Layers
4 Layers
5 Layers
Many-bodyness [arb. un.]
Interatomic distance [Å]
0
1
2
3
4
5
6
7
8
10

How Atoms Interact in Molecules

Adil Kabylda,[1] Malte Esders,[2,3] Matteo Gori,[1] Stefan Chmiela,[2,3]
Klaus-Robert Müller,[2,3,4,5,6,*] and Alexandre Tkatchenko[1,†]

How Atoms Interact in Molecules

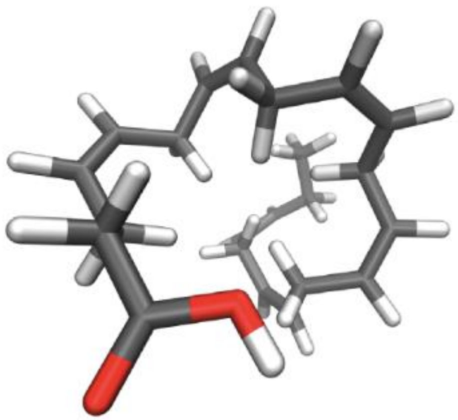
Adil Kabylda,[1] Malte Esders,[2,3] Matteo Gori,[1] Stefan Chmiela,[2,3]
Klaus-Robert Müller,[2,3,4,5,6,*] and Alexandre Tkatchenko[1,†]

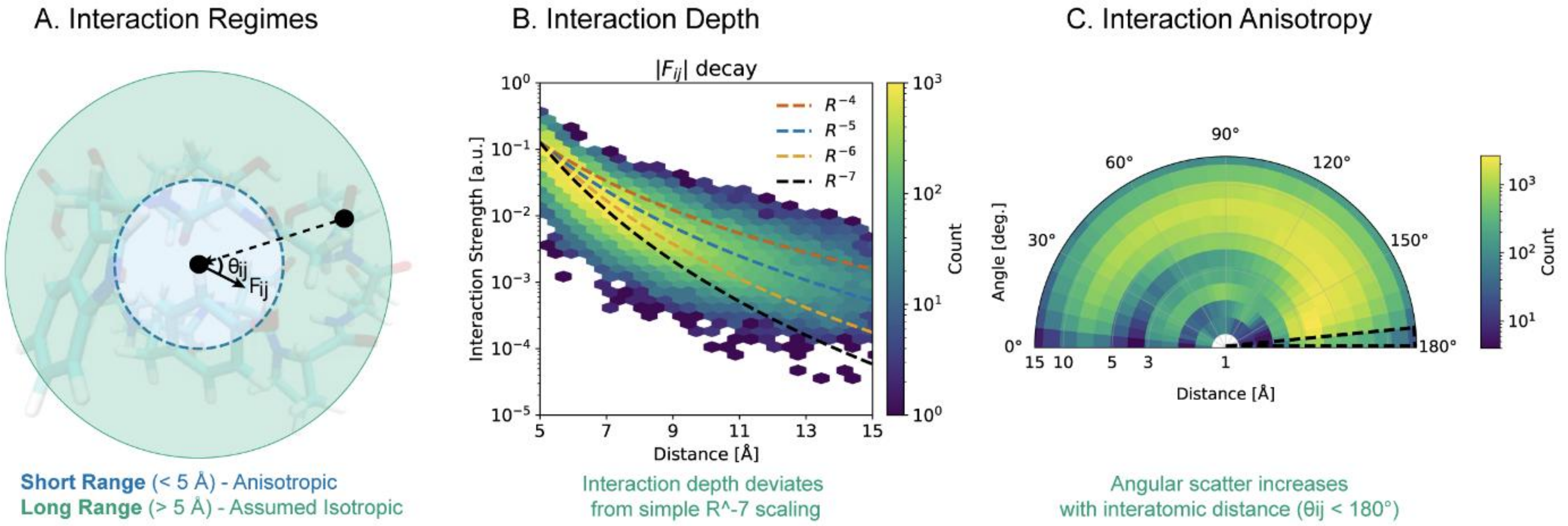
$$\hat{H} = -\sum_i \frac{\nabla_i^2}{2} - \sum_{i,A} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}.$$

$$\mathbf{F}_i = -\frac{\partial E}{\partial \mathbf{R}_i} = -\left\langle \Psi \middle| \frac{\partial \hat{H}}{\partial \mathbf{R}_i} \middle| \Psi \right\rangle,$$

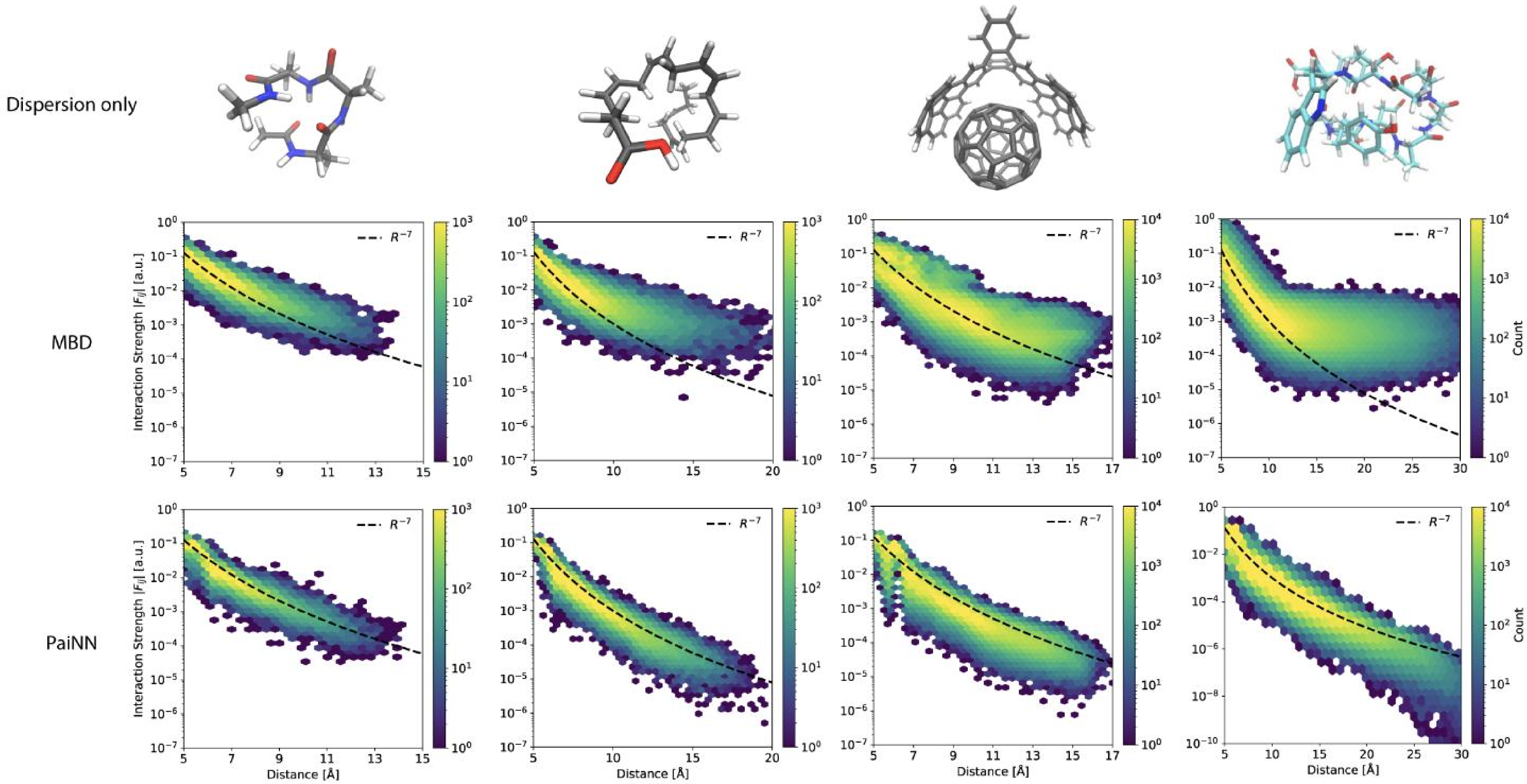
$$\underbrace{\{Z_i, \mathbf{R}_i\}}_{\dim = 4N} \xrightarrow{\hat{H}} \underbrace{|\Psi\rangle \in \mathcal{H}}_{\dim = \infty} \xrightarrow{\langle \partial_R \hat{H} \rangle} \underbrace{\{\mathbf{F}_i\}}_{\dim = 3N}$$

Range and Anisotropy of Quantum Interactions

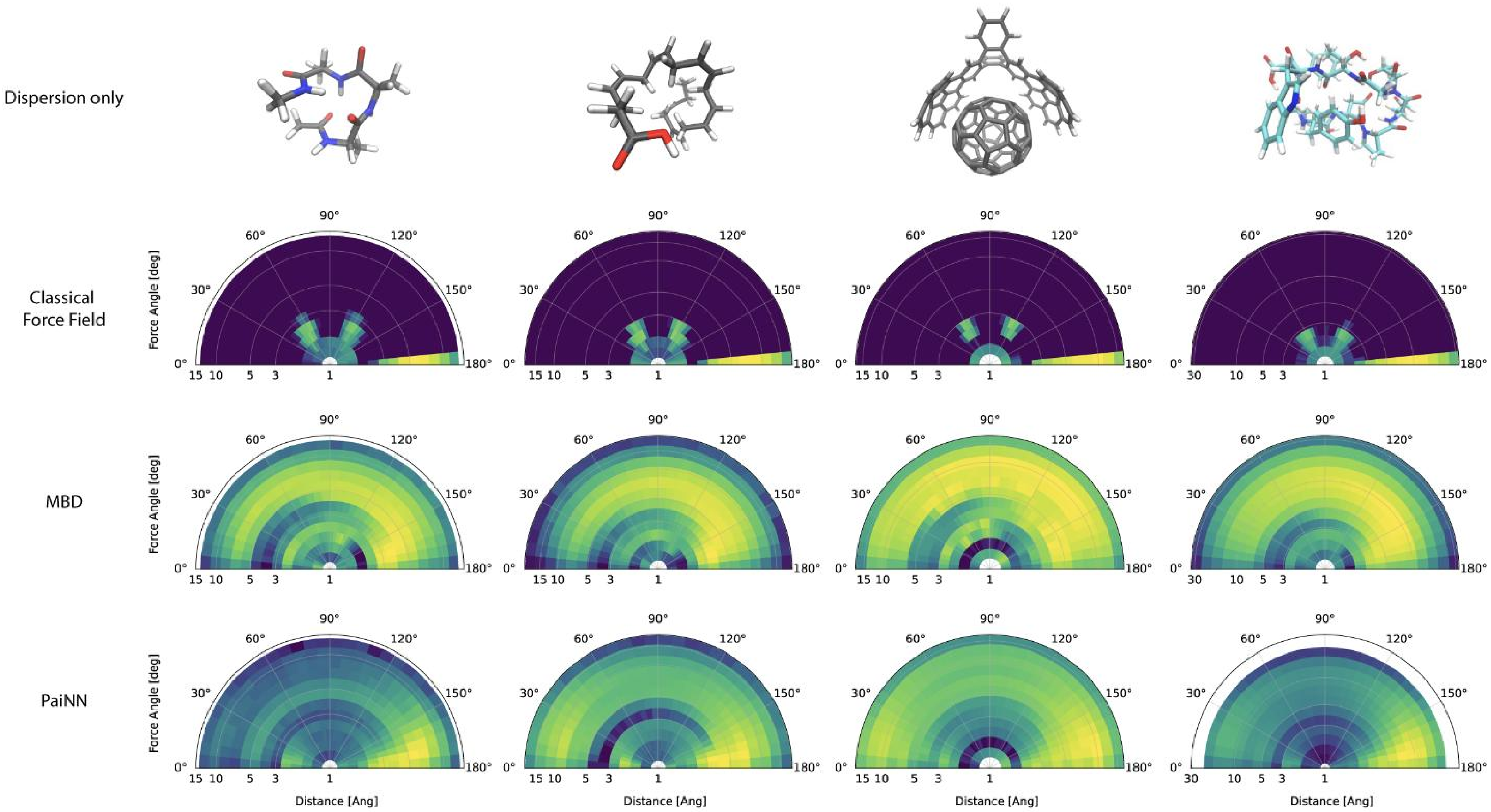




Scaling with Molecular Shape and Size



Anisotropy of Interatomic Interactions



Do We Need Non-Linear ML? Not Always ...

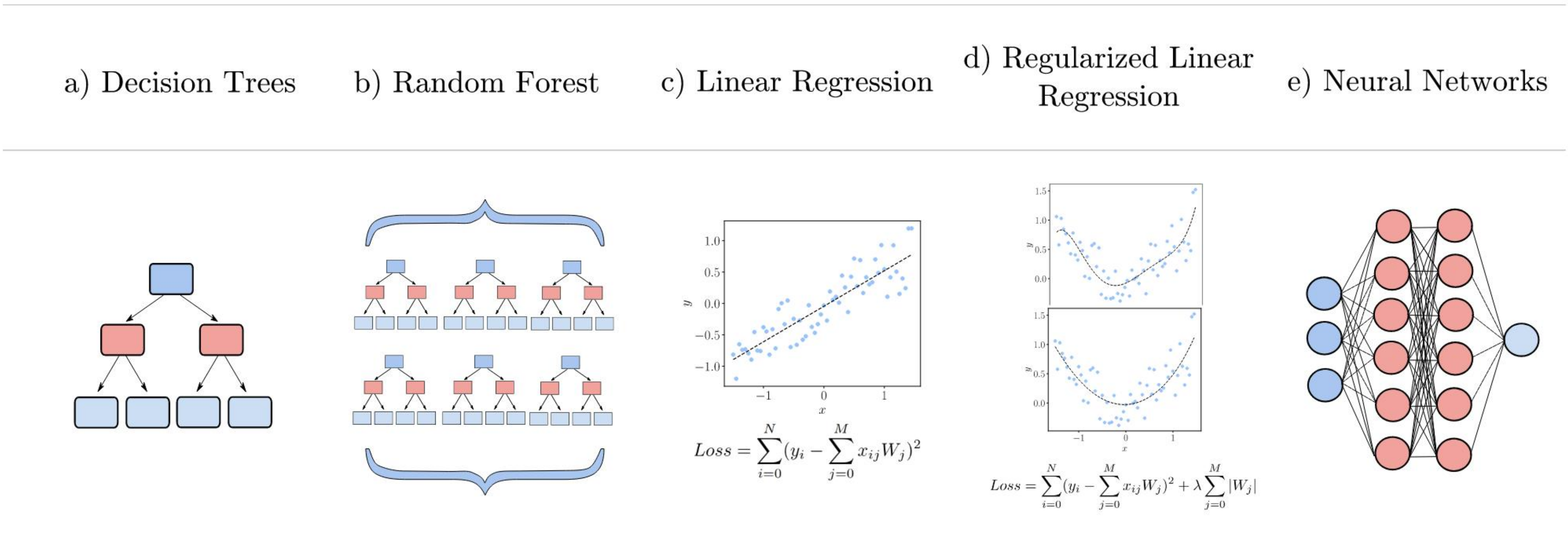
SCIENCE ADVANCES | RESEARCH ARTICLE

CONDENSED MATTER PHYSICS

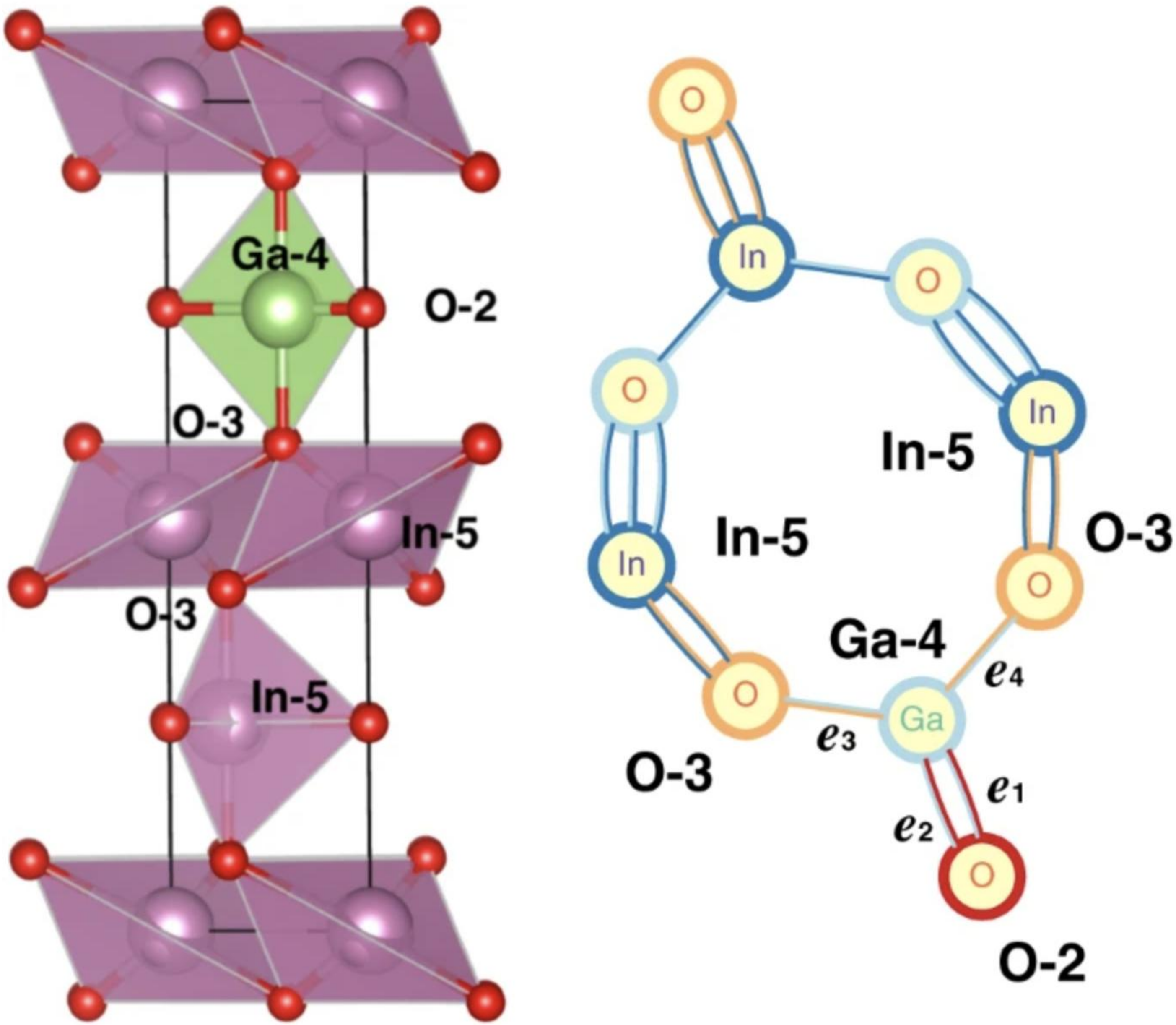
Machine learning of material properties: Predictive and interpretable multilinear models

Alice E. A. Allen* and Alexandre Tkatchenko*

Do We Need Non-Linear ML? Not Always ...

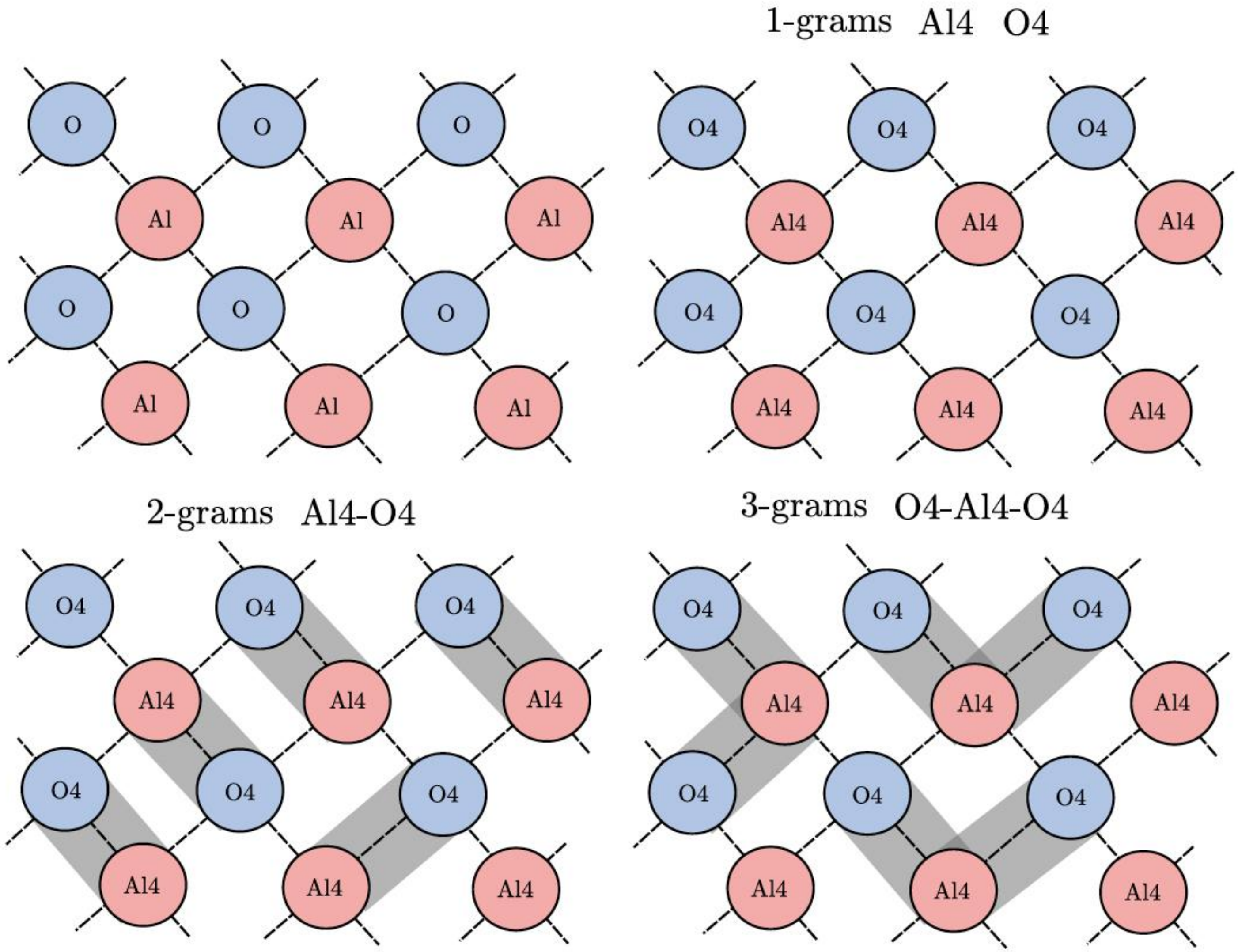


Transparent Conducting Oxides: Kaggle Competition



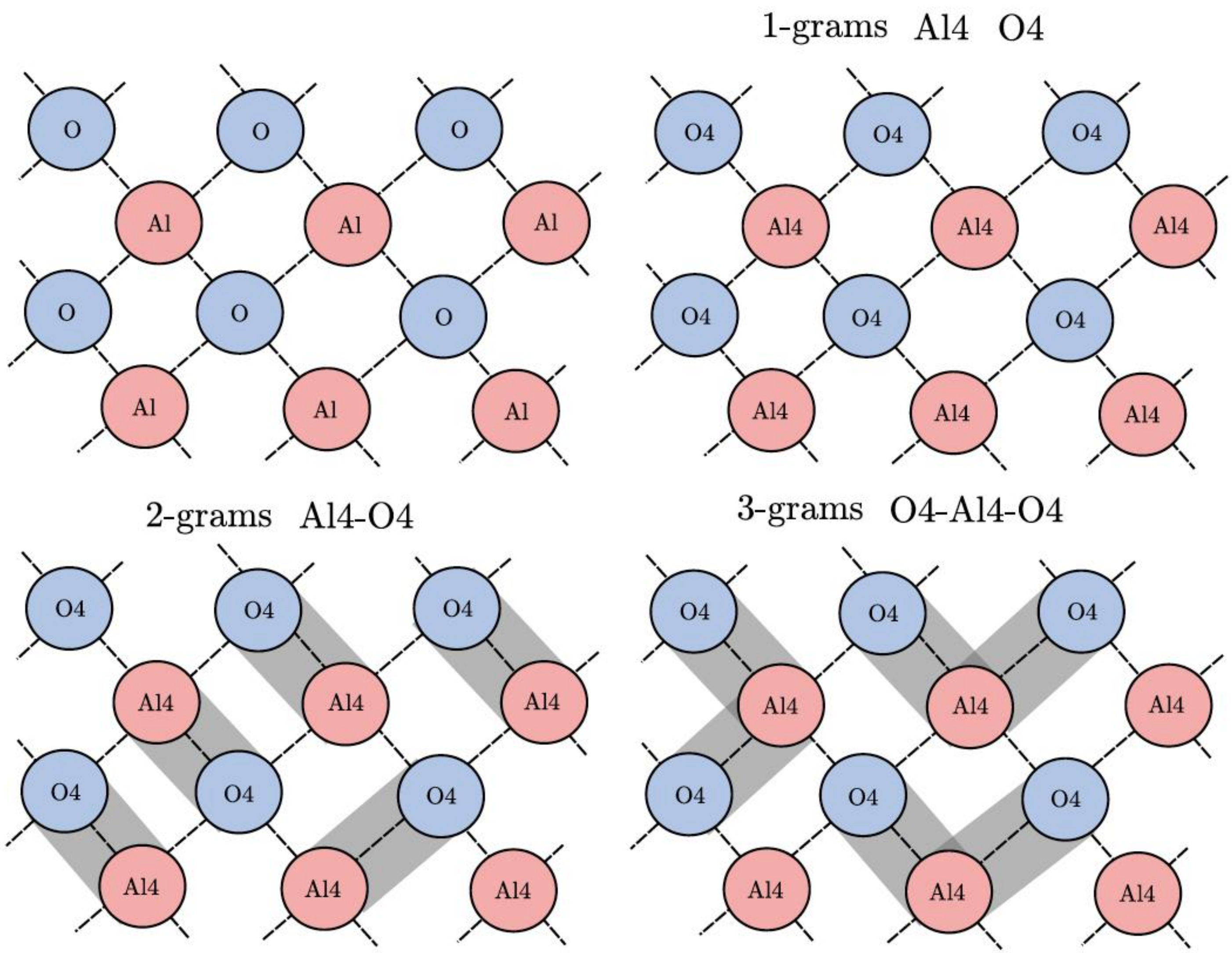
C. Sutton et al., ***npj Comp. Mater.*** *5*, 111 (2019).

Transparent Conducting Oxides: Winning KRR model



C. Sutton et al., ***npj Comp. Mater.*** *5*, 111 (2019).

Transparent Conducting Oxides: Multilinear Model



$$E(\chi) = \sum_i \alpha_i \frac{\chi_i}{V} + \sum_{j<i} \beta_{ij} \frac{\chi_i \chi_j}{V^2} + c$$

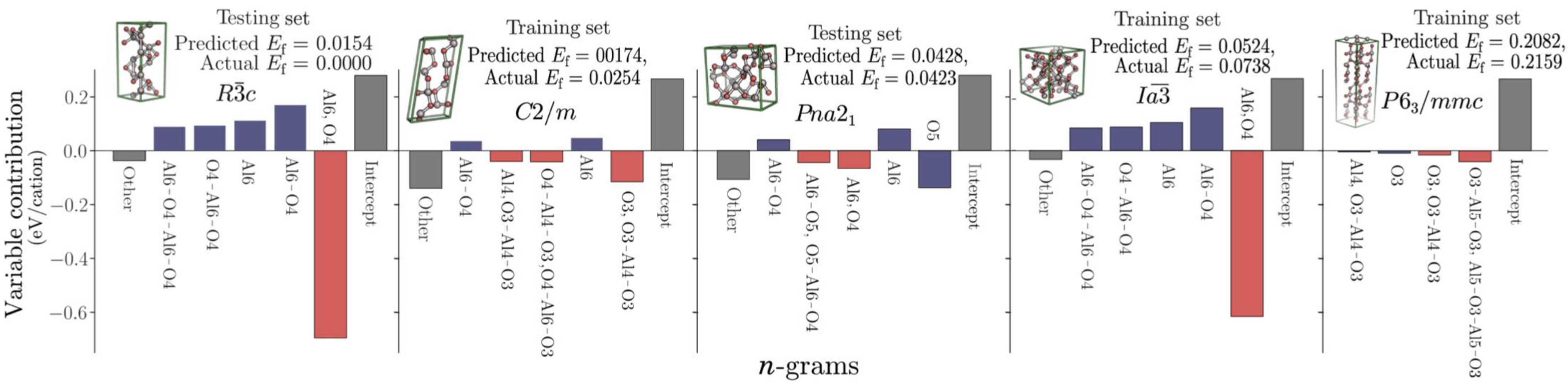
Transparent Conducting Oxides: Multilinear Model

	MAE (eV per cation)			
	E_f		E_{bg}	
Model	**Train**	**Test**	**Train**	**Test**
KRR	0.011	0.015	0.088	0.107
Linear	0.022	0.022	0.143	0.143
Bilinear	0.013	0.015	0.085	0.105

$$E(\chi) = \sum_i \alpha_i \frac{\chi_i}{V} + \sum_{j<i} \beta_{ij} \frac{\chi_i \chi_j}{V^2} + c$$

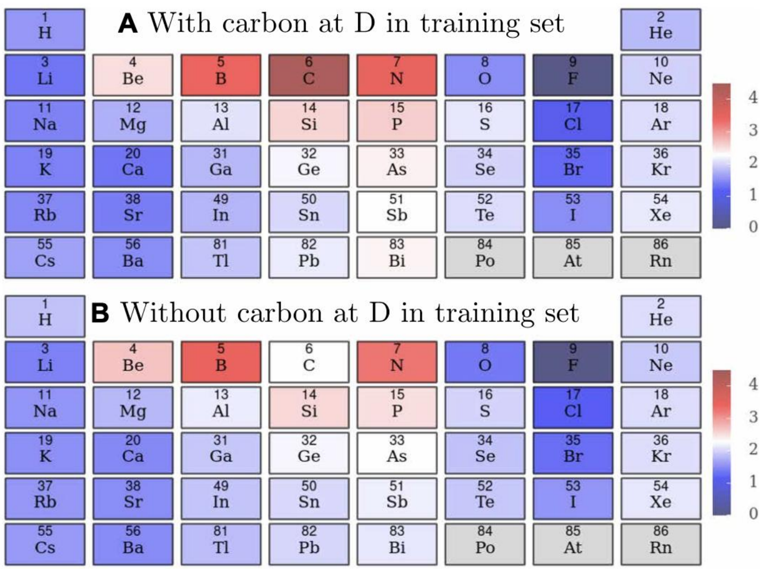
A.E.A. Allen and A. Tkatchenko, ***Science Adv.*** 8, eabm7185 (2022).

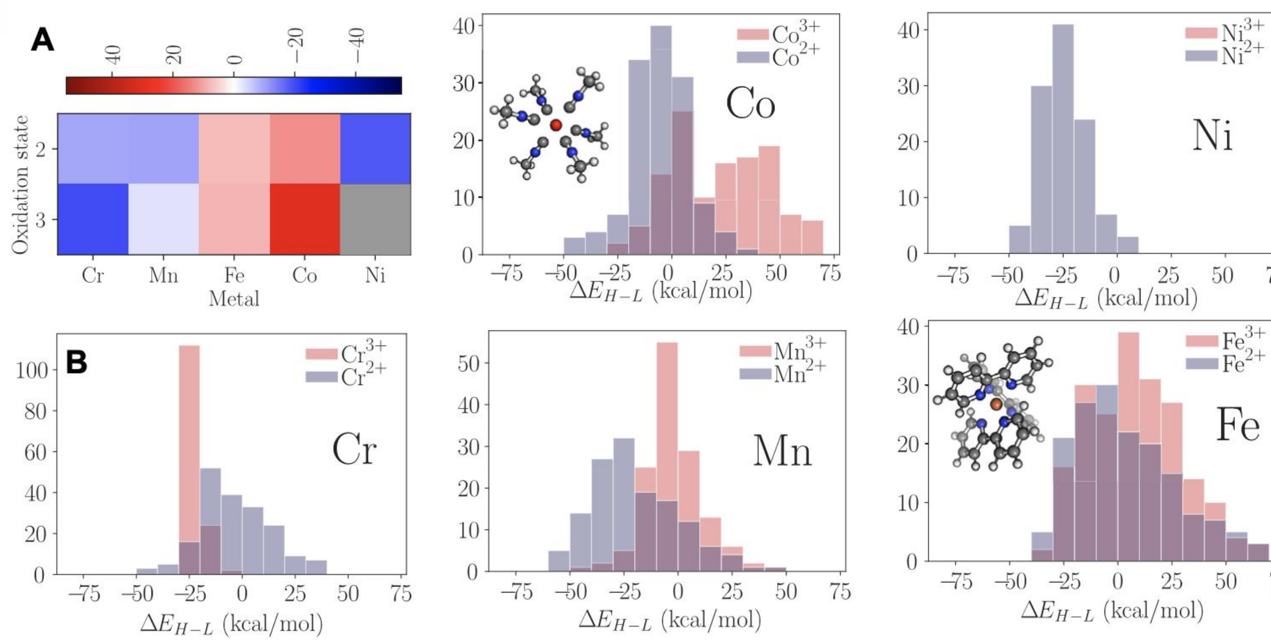
Transparent Conducting Oxides: Multilinear Model



A.E.A. Allen and A. Tkatchenko, ***Science Adv.*** 8, eabm7185 (2022).

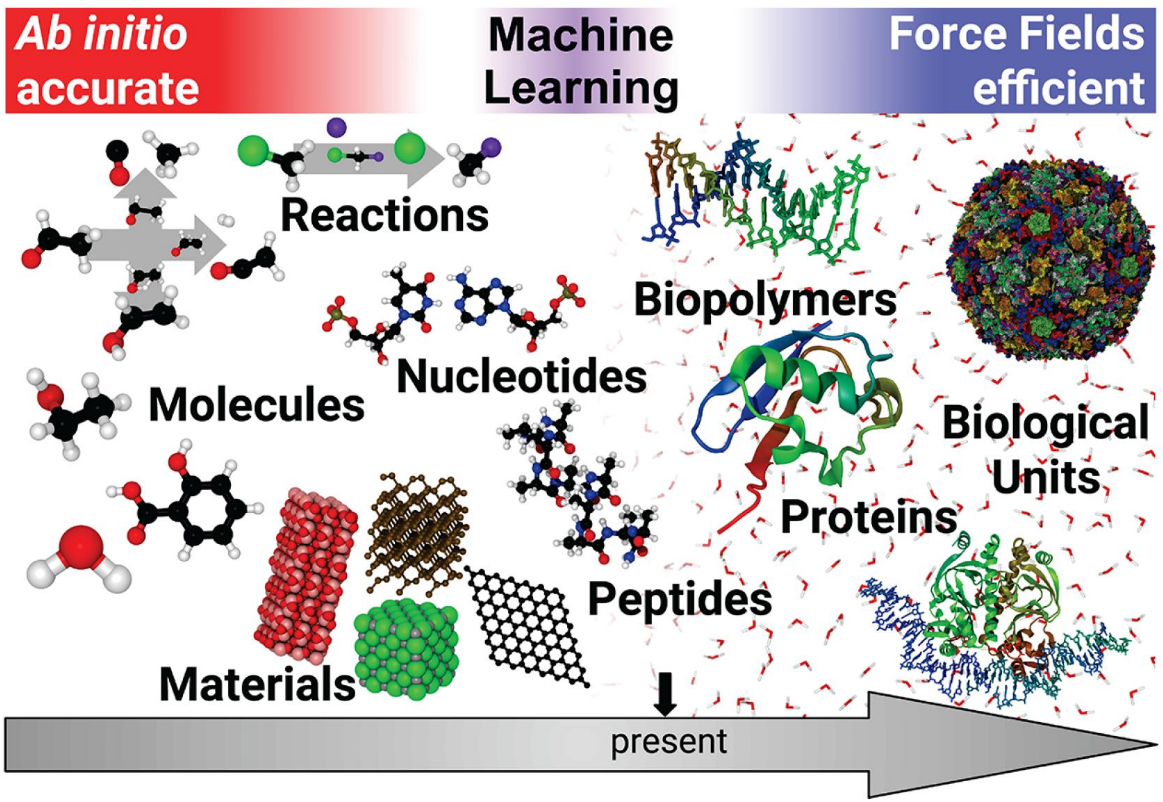
Elpasolite Universe (1.2M crystals) and Spin-Crossover Molecules: Multilinear Models

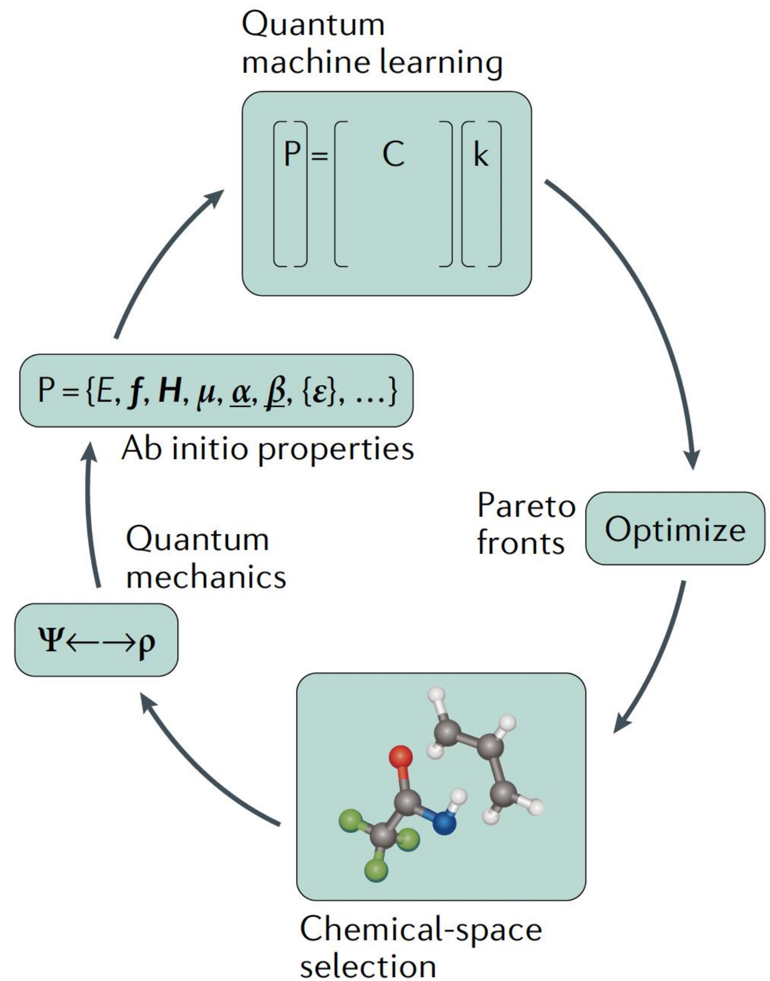




A.E.A. Allen and A. Tkatchenko, ***Science Adv.*** 8, eabm7185 (2022).

Machine Learning for Chemistry and Physics





清华大学燃烧能源中心
Center for Combustion Energy