Can machines make density functionals?

Kieron Burke and friends
UC Irvine
Chemistry
Physics & Astronomy

http://dft.uci.edu

Outline

a. Machine learning in electronic structure
   a. Databases
   b. Forcefields

b. ML for DFT
   - Backround
   - Machine-learned KS kinetic energy of molecules (3D)
   - Machine-learning of XC for strongly correlated solids (1D).
   - Can exact conditions improve ML DFT?

c. Observations

A. Machine learning in electronic structure

The rise of the machines

THE FIRST INTERNATIONAL CONFERENCE ON

MACHINE LEARNING AND PHYSICS

第一届“机器学习和物理学”国际会议

July 4 - 6, 2018, Beijing, China

Insitute for Advanced Study, Tsinghua University

（清华大学高等研究院 主办）

Conference Website and Registration :

http: // mlphys2018.csp.escience.cn

INVITED SPEAKERS

(incomplete list)

Fakher F. Assaad, Wuerzburg

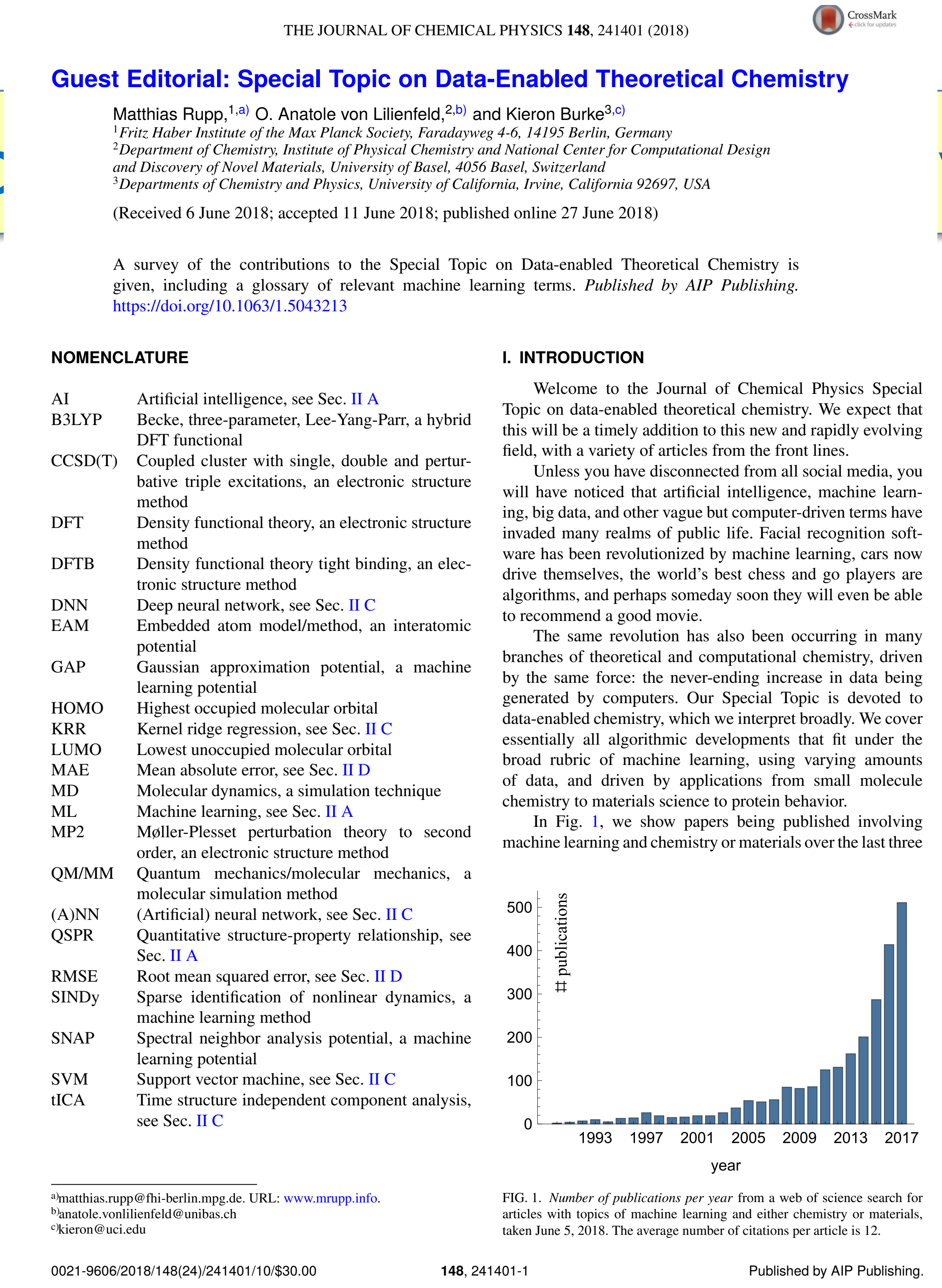
Hans J. Briegel, Innsbruck

Kieron Burke, UC Irvine

Giuseppe Carleo, Flatiron Institute

There is an increasing interest in bringing together the machine learning and physics research, and a large number of interesting works have been done in the recent few years. On one hand, the newly developed of machine

Sp ys


THE JOURNAL OF CHEMICAL PHYSICS **148**, 241401 (2018)

Guest Editorial: Special Topic on Data-Enabled Theoretical Chemistry

Matthias Rupp,[1,a)] O. Anatole von Lilienfeld,[2,b)] and Kieron Burke[3,c)]
[1] *Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, 14195 Berlin, Germany*
[2] *Department of Chemistry, Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials, University of Basel, 4056 Basel, Switzerland*
[3] *Departments of Chemistry and Physics, University of California, Irvine, California 92697, USA*

(Received 6 June 2018; accepted 11 June 2018; published online 27 June 2018)

A survey of the contributions to the Special Topic on Data-enabled Theoretical Chemistry is given, including a glossary of relevant machine learning terms. *Published by AIP Publishing.* https://doi.org/10.1063/1.5043213

NOMENCLATURE

AI	Artificial intelligence, see Sec. II A
B3LYP	Becke, three-parameter, Lee-Yang-Parr, a hybrid DFT functional
CCSD(T)	Coupled cluster with single, double and perturbative triple excitations, an electronic structure method
DFT	Density functional theory, an electronic structure method
DFTB	Density functional theory tight binding, an electronic structure method
DNN	Deep neural network, see Sec. II C
EAM	Embedded atom model/method, an interatomic potential
GAP	Gaussian approximation potential, a machine learning potential
HOMO	Highest occupied molecular orbital
KRR	Kernel ridge regression, see Sec. II C
LUMO	Lowest unoccupied molecular orbital
MAE	Mean absolute error, see Sec. II D
MD	Molecular dynamics, a simulation technique
ML	Machine learning, see Sec. II A
MP2	Møller-Plesset perturbation theory to second order, an electronic structure method
QM/MM	Quantum mechanics/molecular mechanics, a molecular simulation method
(A)NN	(Artificial) neural network, see Sec. II C
QSPR	Quantitative structure-property relationship, see Sec. II A
RMSE	Root mean squared error, see Sec. II D
SINDy	Sparse identification of nonlinear dynamics, a machine learning method
SNAP	Spectral neighbor analysis potential, a machine learning potential
SVM	Support vector machine, see Sec. II C
tICA	Time structure independent component analysis, see Sec. II C

I. INTRODUCTION

Welcome to the Journal of Chemical Physics Special Topic on data-enabled theoretical chemistry. We expect that this will be a timely addition to this new and rapidly evolving field, with a variety of articles from the front lines.

Unless you have disconnected from all social media, you will have noticed that artificial intelligence, machine learning, big data, and other vague but computer-driven terms have invaded many realms of public life. Facial recognition software has been revolutionized by machine learning, cars now drive themselves, the world's best chess and go players are algorithms, and perhaps someday soon they will even be able to recommend a good movie.

The same revolution has also been occurring in many branches of theoretical and computational chemistry, driven by the same force: the never-ending increase in data being generated by computers. Our Special Topic is devoted to data-enabled chemistry, which we interpret broadly. We cover essentially all algorithmic developments that fit under the broad rubric of machine learning, using varying amounts of data, and driven by applications from small molecule chemistry to materials science to protein behavior.

In Fig. 1, we show papers being published involving machine learning and chemistry or materials over the last three

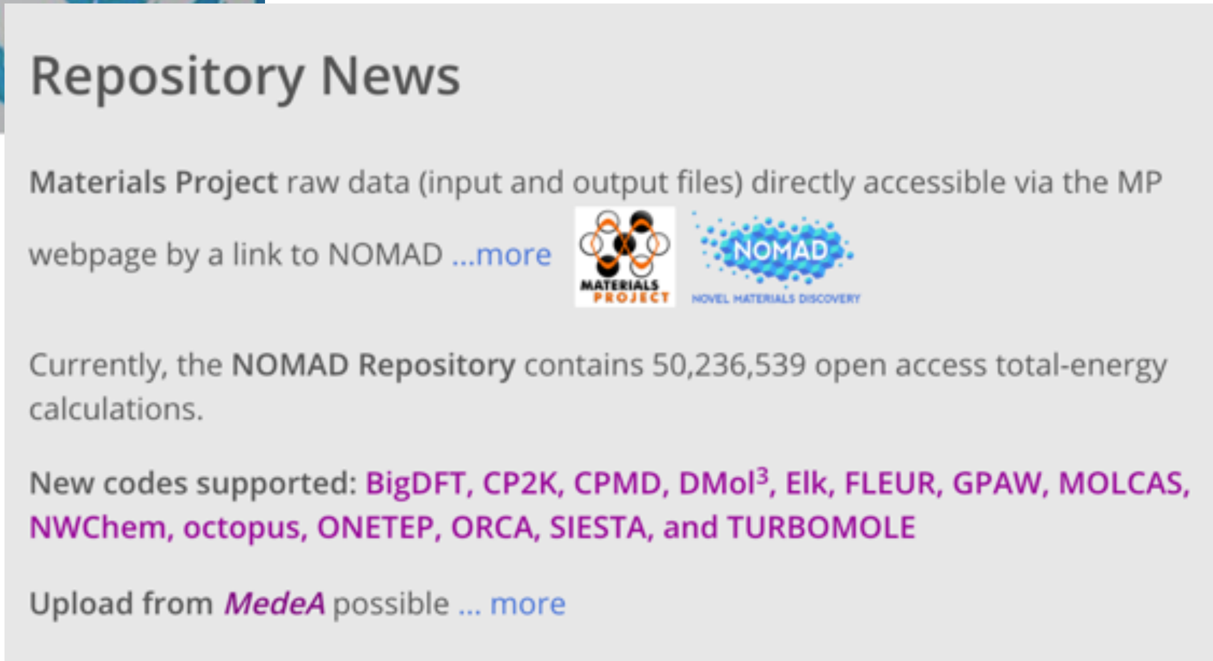
FIG. 1. *Number of publications per year* from a web of science search for articles with topics of machine learning and either chemistry or materials, taken June 5, 2018. The average number of citations per article is 12.

[a)] matthias.rupp@fhi-berlin.mpg.de. URL: www.mrupp.info.
[b)] anatole.vonlilienfeld@unibas.ch
[c)] kieron@uci.edu

ML taking off in computational chemistry/materials

- 44 papers in the field in Special Topic
- Searching chemical (and materials) compound space
- Making interatomic potentials
- Making new density functionals
- Making accurate potential energy surfaces for small molecules
- Accelerating MD by finding slow variables
- …

The NOMAD Laboratory
A European Centre of Excellence



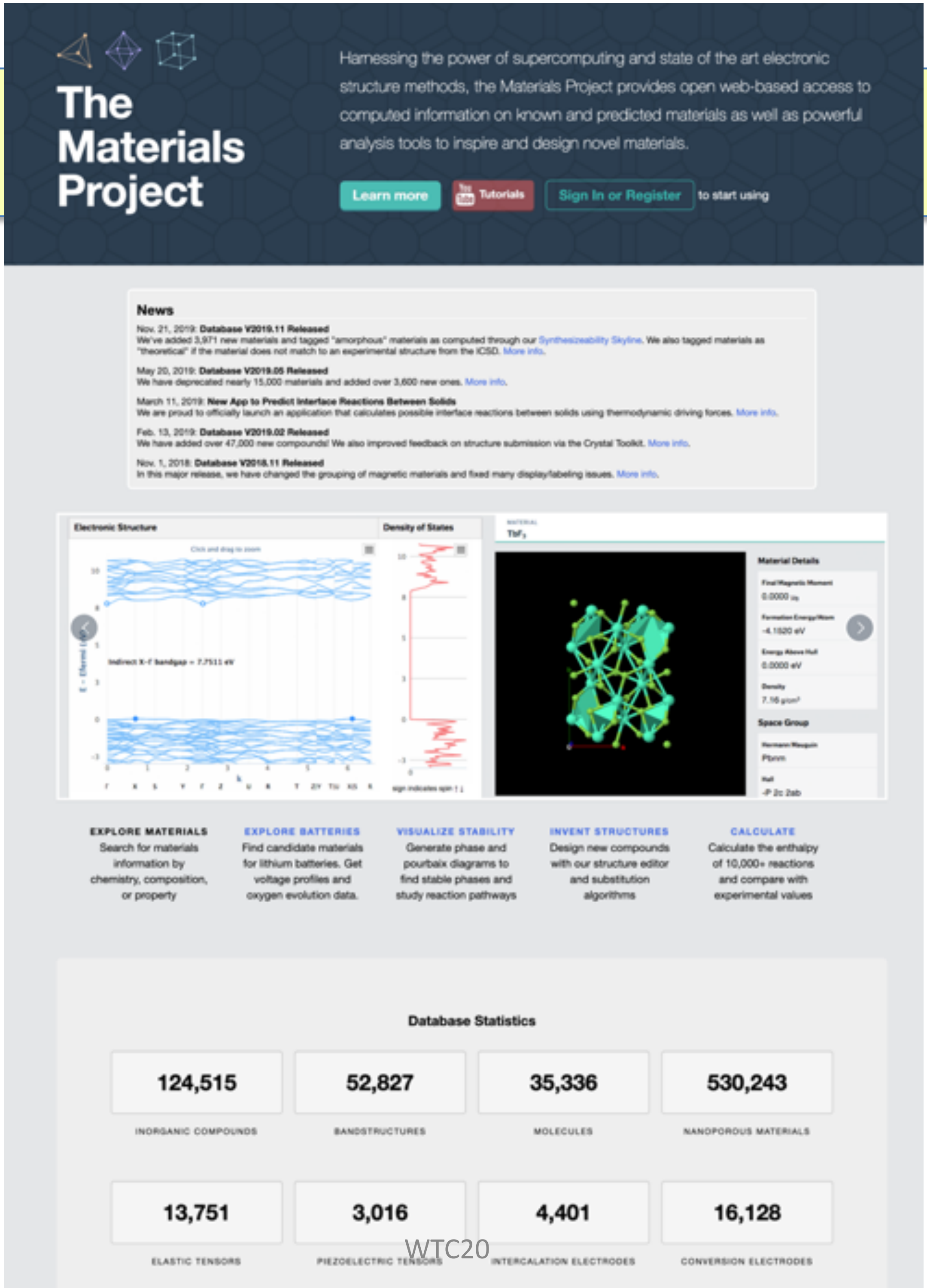
Repository News

Materials Project raw data (input and output files) directly accessible via the MP webpage by a link to NOMAD ...more

Currently, the **NOMAD Repository** contains 50,236,539 open access total-energy calculations.

New codes supported: BigDFT, CP2K, CPMD, DMol^3, Elk, FLEUR, GPAW, MOLCAS, NWChem, octopus, ONETEP, ORCA, SIESTA, and TURBOMOLE

Upload from *MedeA* possible ... more



The Materials Project

Harnessing the power of supercomputing and state of the art electronic structure methods, the Materials Project provides open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials.

to start using

News

Nov. 21, 2019: **Database V2019.11 Released**
We've added 3,971 new materials and tagged "amorphous" materials as computed through our Synthesizeability Skyline. We also tagged materials as "theoretical" if the material does not match to an experimental structure from the ICSD. More info.

May 20, 2019: **Database V2019.05 Released**
We have deprecated nearly 15,000 materials and added over 3,600 new ones. More info.

March 11, 2019: **New App to Predict Interface Reactions Between Solids**
We are proud to officially launch an application that calculates possible interface reactions between solids using thermodynamic driving forces. More info.

Feb. 13, 2019: **Database V2019.02 Released**
We have added over 47,000 new compounds! We also improved feedback on structure submission via the Crystal Toolkit. More info.

Nov. 1, 2018: **Database V2018.11 Released**
In this major release, we have changed the grouping of magnetic materials and fixed many display/labeling issues. More info.

Material Details

Final Magnetic Moment
0.0000 μB

Formation Energy/Atom
-4.1520 eV

Energy Above Hull
0.0000 eV

Density
7.16 g/cm³

Space Group

Hermann Mauguin
Pbnm

Hall
-P 2c 2ab

EXPLORE MATERIALS
Search for materials information by chemistry, composition, or property

EXPLORE BATTERIES
Find candidate materials for lithium batteries. Get voltage profiles and oxygen evolution data.

VISUALIZE STABILITY
Generate phase and pourbaix diagrams to find stable phases and study reaction pathways

INVENT STRUCTURES
Design new compounds with our structure editor and substitution algorithms

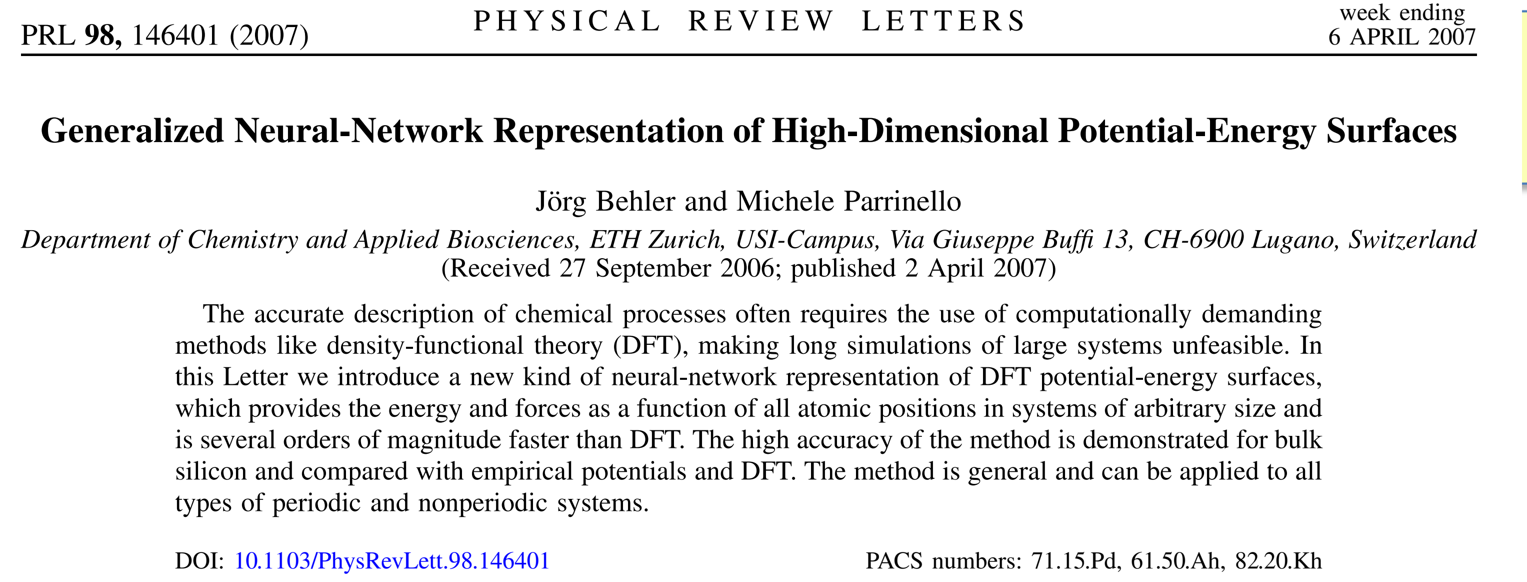
CALCULATE
Calculate the enthalpy of 10,000+ reactions and compare with experimental values

Database Statistics

124,515	52,827	35,336	530,243
INORGANIC COMPOUNDS	BANDSTRUCTURES	MOLECULES	NANOPOROUS MATERIALS
13,751	3,016	4,401	16,128
ELASTIC TENSORS	PIEZOELECTRIC TENSORS	INTERCALATION ELECTRODES	CONVERSION ELECTRODES

Force fields

- Used for classical MD simulations of, e.g.,
  - Biomolecules in drug design (CHARMM, AMBER, etc.)
  - Microscopic amounts of materials (Stillinger-Weber)
  - For water, at least 104 different respectable force fields
  - Often not good at bond-breaking
- Want improved accuracy, or greater transferability, etc.
- Two big uses:
  - Searching chemical compound space: What molecules could exist (dozen elements, all possible bonds)?
  - Searching materials configuration space: 1-2 elements, but thousands to millions of them



PRL **98,** 146401 (2007) PHYSICAL REVIEW LETTERS week ending 6 APRIL 2007

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

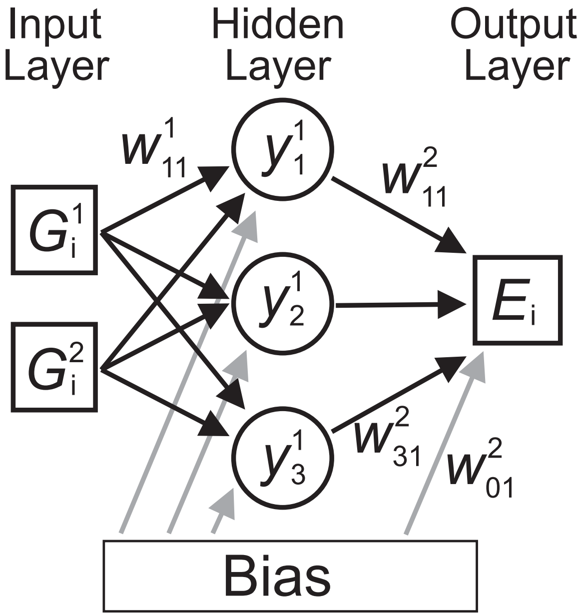
Jörg Behler and Michele Parrinello

Department of Chemistry and Applied Biosciences, ETH Zurich, USI-Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

(Received 27 September 2006; published 2 April 2007)

The accurate description of chemical processes often requires the use of computationally demanding methods like density-functional theory (DFT), making long simulations of large systems unfeasible. In this Letter we introduce a new kind of neural-network representation of DFT potential-energy surfaces, which provides the energy and forces as a function of all atomic positions in systems of arbitrary size and is several orders of magnitude faster than DFT. The high accuracy of the method is demonstrated for bulk silicon and compared with empirical potentials and DFT. The method is general and can be applied to all types of periodic and nonperiodic systems.

DOI: 10.1103/PhysRevLett.98.146401 PACS numbers: 71.15.Pd, 61.50.Ah, 82.20.Kh



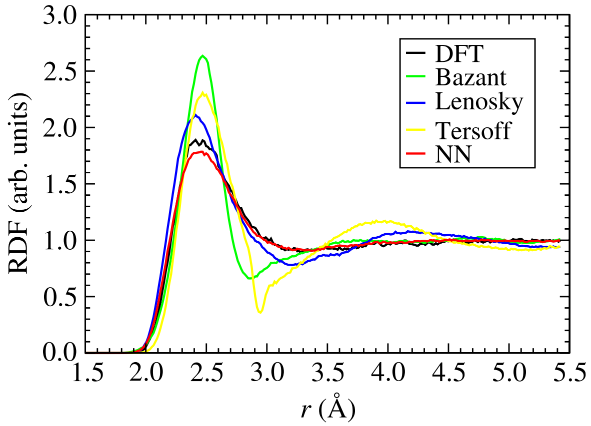
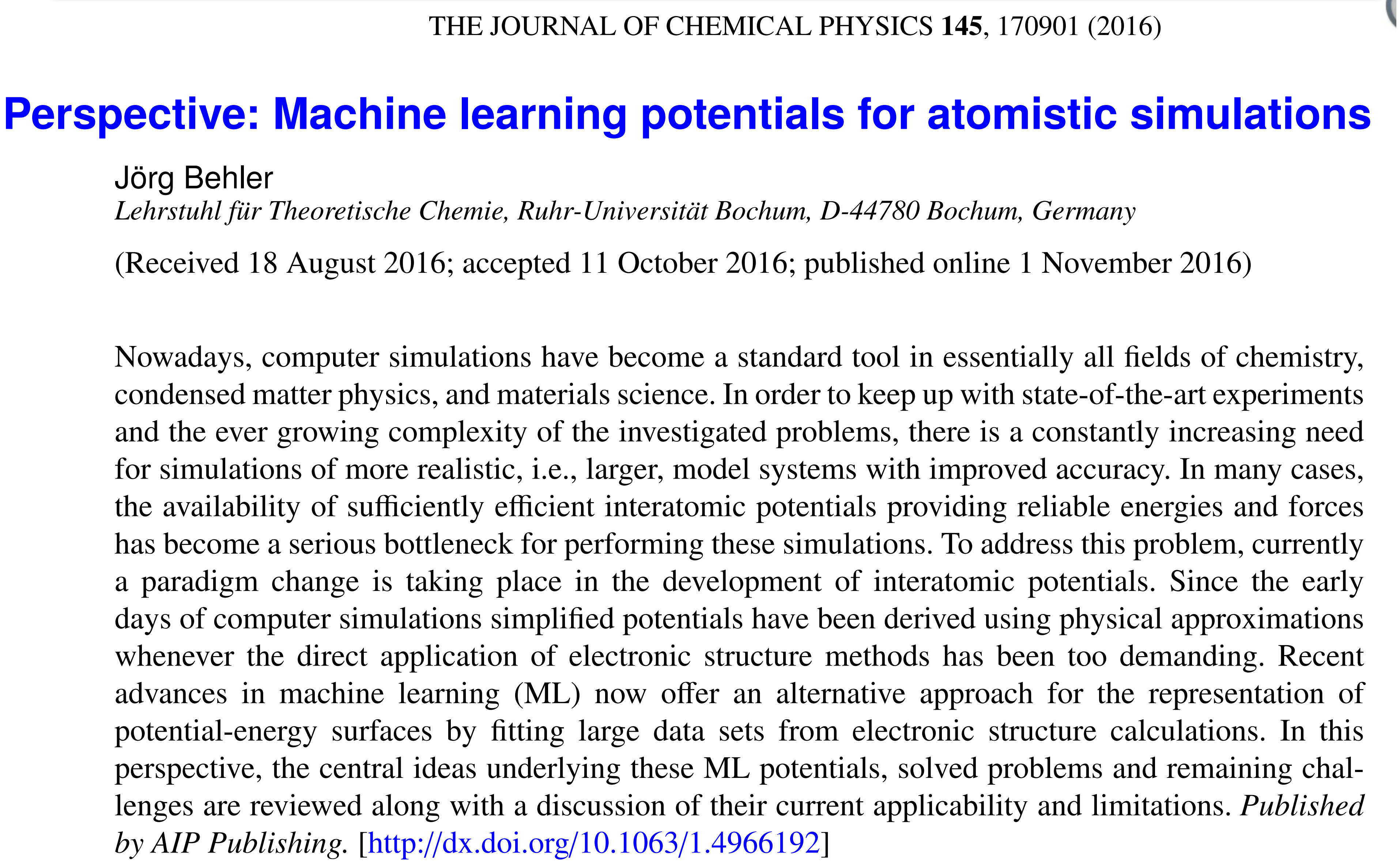


FIG. 3 (color online). Radial distribution function (RDF) of a silicon melt at 3000 K as obtained using a cubic 64 atom cell ($a = 20.526$ bohr). The curves shown were obtained from the Bazant [17,19], the Lenosky [15,19], the Tersoff [16,20], a neural network (NN) potential, and from density-functional theory (DFT) [18].

Behler-Parinello forces



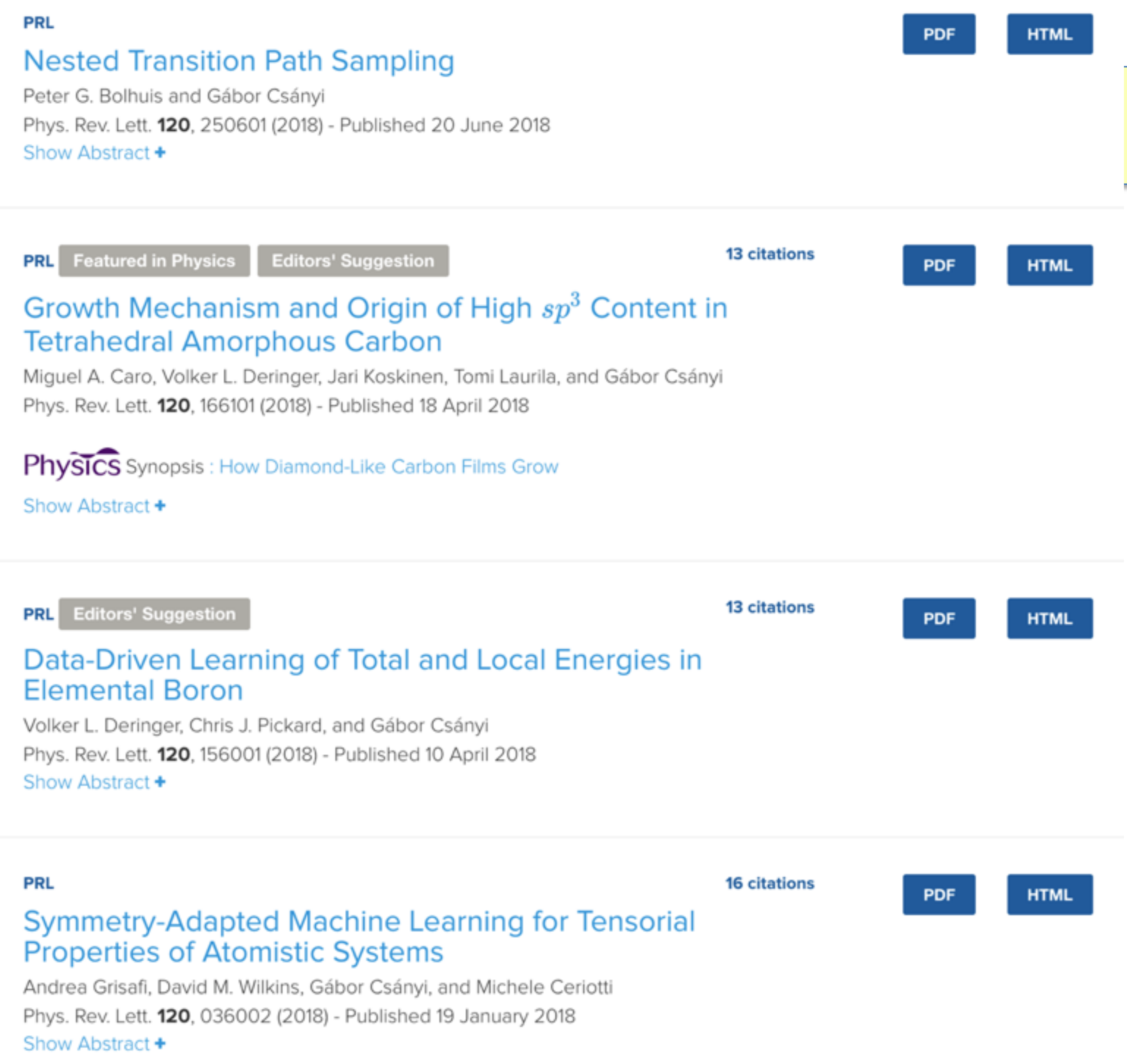
THE JOURNAL OF CHEMICAL PHYSICS **145**, 170901 (2016)

Perspective: Machine learning potentials for atomistic simulations

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

(Received 18 August 2016; accepted 11 October 2016; published online 1 November 2016)

Nowadays, computer simulations have become a standard tool in essentially all fields of chemistry, condensed matter physics, and materials science. In order to keep up with state-of-the-art experiments and the ever growing complexity of the investigated problems, there is a constantly increasing need for simulations of more realistic, i.e., larger, model systems with improved accuracy. In many cases, the availability of sufficiently efficient interatomic potentials providing reliable energies and forces has become a serious bottleneck for performing these simulations. To address this problem, currently a paradigm change is taking place in the development of interatomic potentials. Since the early days of computer simulations simplified potentials have been derived using physical approximations whenever the direct application of electronic structure methods has been too demanding. Recent advances in machine learning (ML) now offer an alternative approach for the representation of potential-energy surfaces by fitting large data sets from electronic structure calculations. In this perspective, the central ideas underlying these ML potentials, solved problems and remaining challenges are reviewed along with a discussion of their current applicability and limitations. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4966192]



PRL

Nested Transition Path Sampling

Peter G. Bolhuis and Gábor Csányi

Phys. Rev. Lett. **120**, 250601 (2018) - Published 20 June 2018

PRL Featured in Physics Editors' Suggestion

13 citations

Growth Mechanism and Origin of High sp^3 Content in Tetrahedral Amorphous Carbon

Miguel A. Caro, Volker L. Deringer, Jari Koskinen, Tomi Laurila, and Gábor Csányi

Phys. Rev. Lett. **120**, 166101 (2018) - Published 18 April 2018

Physics Synopsis : How Diamond-Like Carbon Films Grow

PRL Editors' Suggestion

13 citations

Data-Driven Learning of Total and Local Energies in Elemental Boron

Volker L. Deringer, Chris J. Pickard, and Gábor Csányi

Phys. Rev. Lett. **120**, 156001 (2018) - Published 10 April 2018

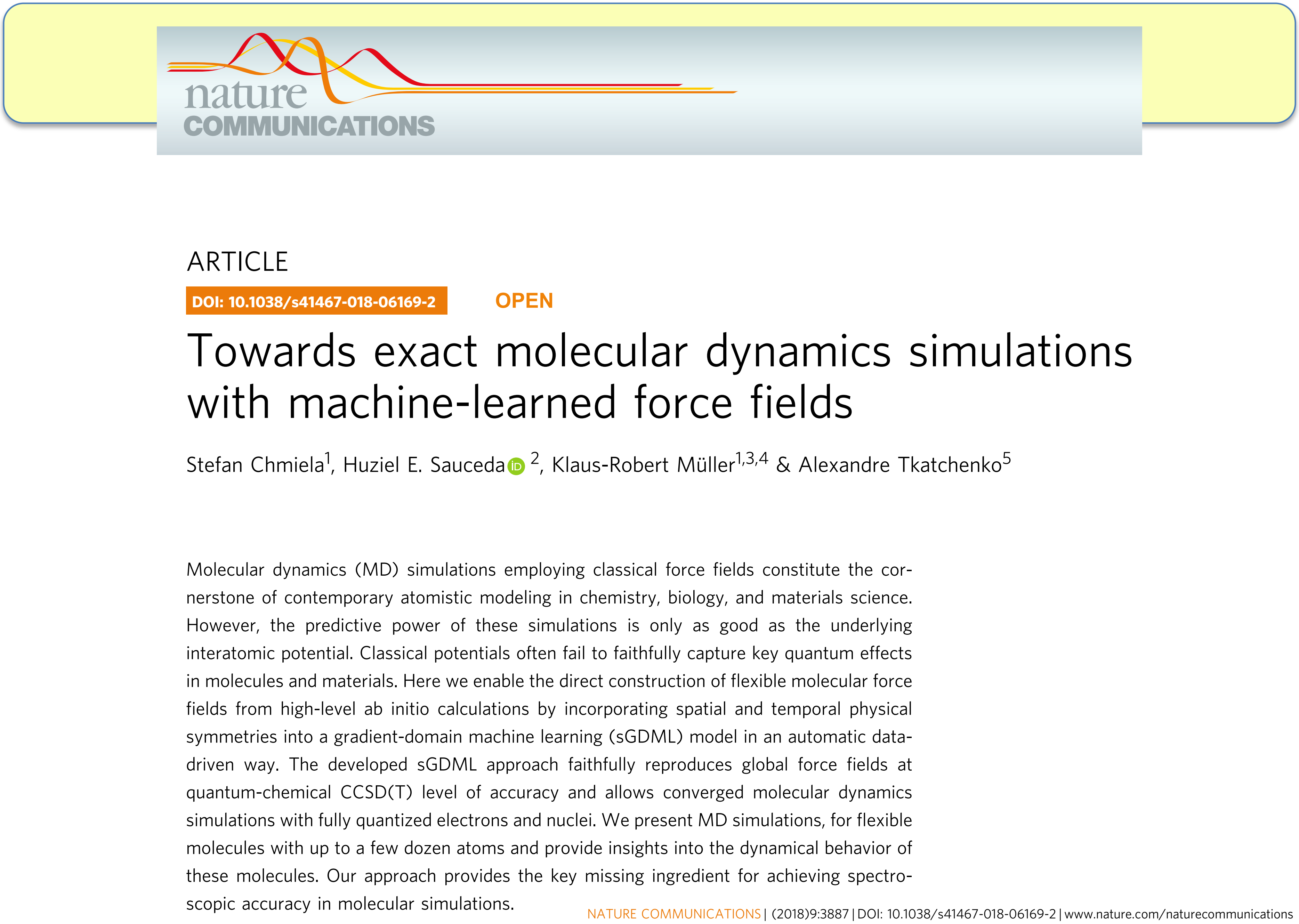
PRL

16 citations

Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems

Andrea Grisafi, David M. Wilkins, Gábor Csányi, and Michele Ceriotti

Phys. Rev. Lett. **120**, 036002 (2018) - Published 19 January 2018



nature
COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-06169-2

OPEN

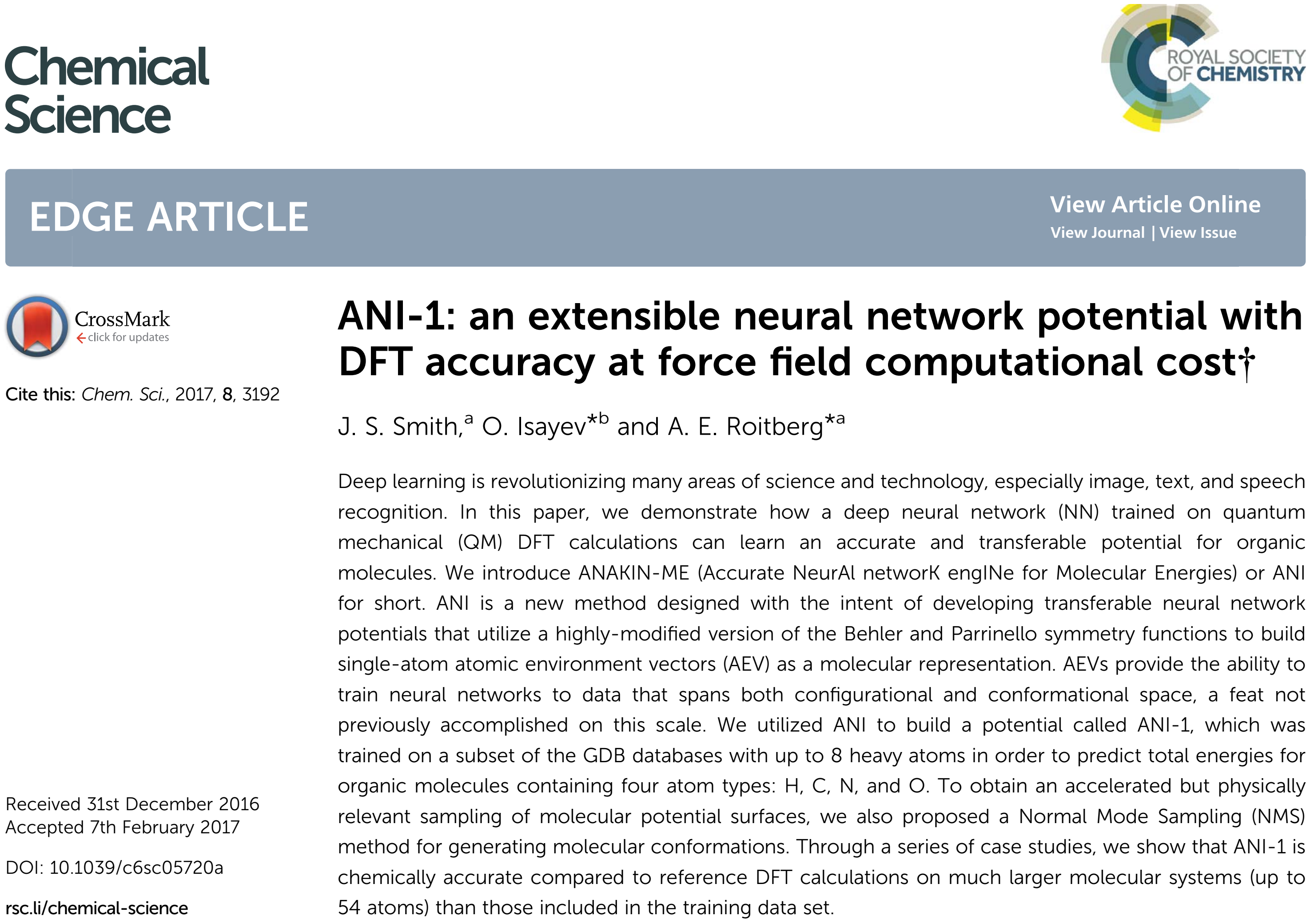
Towards exact molecular dynamics simulations with machine-learned force fields

Stefan Chmiela[1], Huziel E. Sauceda[2], Klaus-Robert Müller[1,3,4] & Alexandre Tkatchenko[5]

Molecular dynamics (MD) simulations employing classical force fields constitute the cornerstone of contemporary atomistic modeling in chemistry, biology, and materials science. However, the predictive power of these simulations is only as good as the underlying interatomic potential. Classical potentials often fail to faithfully capture key quantum effects in molecules and materials. Here we enable the direct construction of flexible molecular force fields from high-level ab initio calculations by incorporating spatial and temporal physical symmetries into a gradient-domain machine learning (sGDML) model in an automatic data-driven way. The developed sGDML approach faithfully reproduces global force fields at quantum-chemical CCSD(T) level of accuracy and allows converged molecular dynamics simulations with fully quantized electrons and nuclei. We present MD simulations, for flexible molecules with up to a few dozen atoms and provide insights into the dynamical behavior of these molecules. Our approach provides the key missing ingredient for achieving spectroscopic accuracy in molecular simulations.

NATURE COMMUNICATIONS | (2018)9:3887 | DOI: 10.1038/s41467-018-06169-2 | www.nature.com/naturecommunications

Let the force be with you



Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

CrossMark
click for updates

Cite this: *Chem. Sci.*, 2017, **8**, 3192

ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost†

J. S. Smith,[a] O. Isayev*[b] and A. E. Roitberg*[a]

Deep learning is revolutionizing many areas of science and technology, especially image, text, and speech recognition. In this paper, we demonstrate how a deep neural network (NN) trained on quantum mechanical (QM) DFT calculations can learn an accurate and transferable potential for organic molecules. We introduce ANAKIN-ME (Accurate NeurAl networK engINe for Molecular Energies) or ANI for short. ANI is a new method designed with the intent of developing transferable neural network potentials that utilize a highly-modified version of the Behler and Parrinello symmetry functions to build single-atom atomic environment vectors (AEV) as a molecular representation. AEVs provide the ability to train neural networks to data that spans both configurational and conformational space, a feat not previously accomplished on this scale. We utilized ANI to build a potential called ANI-1, which was trained on a subset of the GDB databases with up to 8 heavy atoms in order to predict total energies for organic molecules containing four atom types: H, C, N, and O. To obtain an accelerated but physically relevant sampling of molecular potential surfaces, we also proposed a Normal Mode Sampling (NMS) method for generating molecular conformations. Through a series of case studies, we show that ANI-1 is chemically accurate compared to reference DFT calculations on much larger molecular systems (up to 54 atoms) than those included in the training data set.

Received 31st December 2016
Accepted 7th February 2017

DOI: 10.1039/c6sc05720a

rsc.li/chemical-science

B. ML-DFT: An unholy alliance?

Exchange-correlation potentials

David J. Tozer, Victoria E. Ingamells, and Nicholas C. Handy
Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, United Kingdom

[redacted]

We describe our implementation of the Zhao, Morrison, and Parr method [Phys. Rev. A **50**, 2138 (1994)] for the calculation of molecular exchange-correlation potentials from high-level *ab initio* densities. The use of conventional Gaussian basis sets demands careful consideration of the value of the Lagrange multiplier associated with the constraint that reproduces the input density. Although formally infinite, we demonstrate that a finite value should be used in finite basis set calculations. The potential has been determined for Ne, HF, N_2, H_2O, and $N_2(1.5r_e)$, and compared with popular analytic potentials. We have then examined how well the Zhao, Morrison, Parr potential can be represented using a computational neural network. Assuming $v_{xc}=v_{xc}(\rho)$, we incorporate the neural network into a regular Kohn–Sham procedure [Phys. Rev. A **140**, 1133 (1965)] with encouraging results. The extension of this method to include density derivatives is briefly outlined. [redacted]

Various flavors

- For $E_{XC}[n]$: Take existing forms, possibly generalized, and improve fitting (greater accuracy or applicability)

- For either $T_S[n]$ or $E_{XC}[n]$: Create highly non-local approximations that work where semilocal functionals fail, i.e., stretched bonds and strong correlation.

Machine learning: Kernel ridge regression

- Powerful branch of artificial intelligence
- Essentially fitting and interpolating
- Maps problem into much higher-dimension feature space, using a simple kernel
- Higher-dimension often means more linear
- Perform regression in feature space
- Project back to original problem

Kernel ridge regression

- Kernel ridge regression (KRR). Given $\{\mathbf{x}_j, f_j\}$

$$\hat{f}(\mathbf{x}) = \sum_{j=1}^{M} \alpha_j k(\mathbf{x}_j, \mathbf{x})$$

length scale

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2/(2\sigma^2))$$

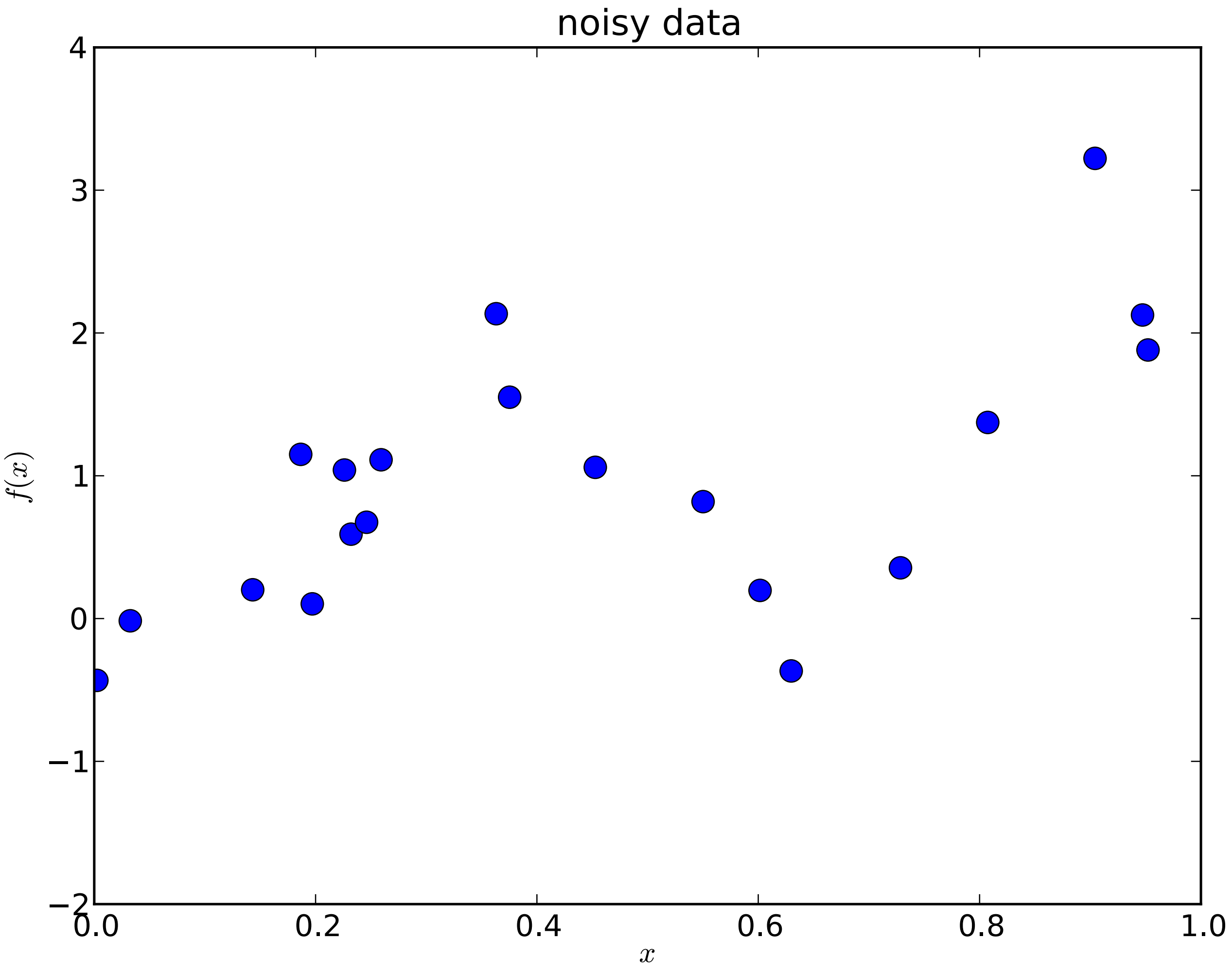
- Minimize:

$$\mathcal{C}(\boldsymbol{\alpha}) = \sum_{j=1}^{M} (\hat{f}(\mathbf{x}_j) - f_j)^2 + \lambda^2 \|\alpha\|^2$$

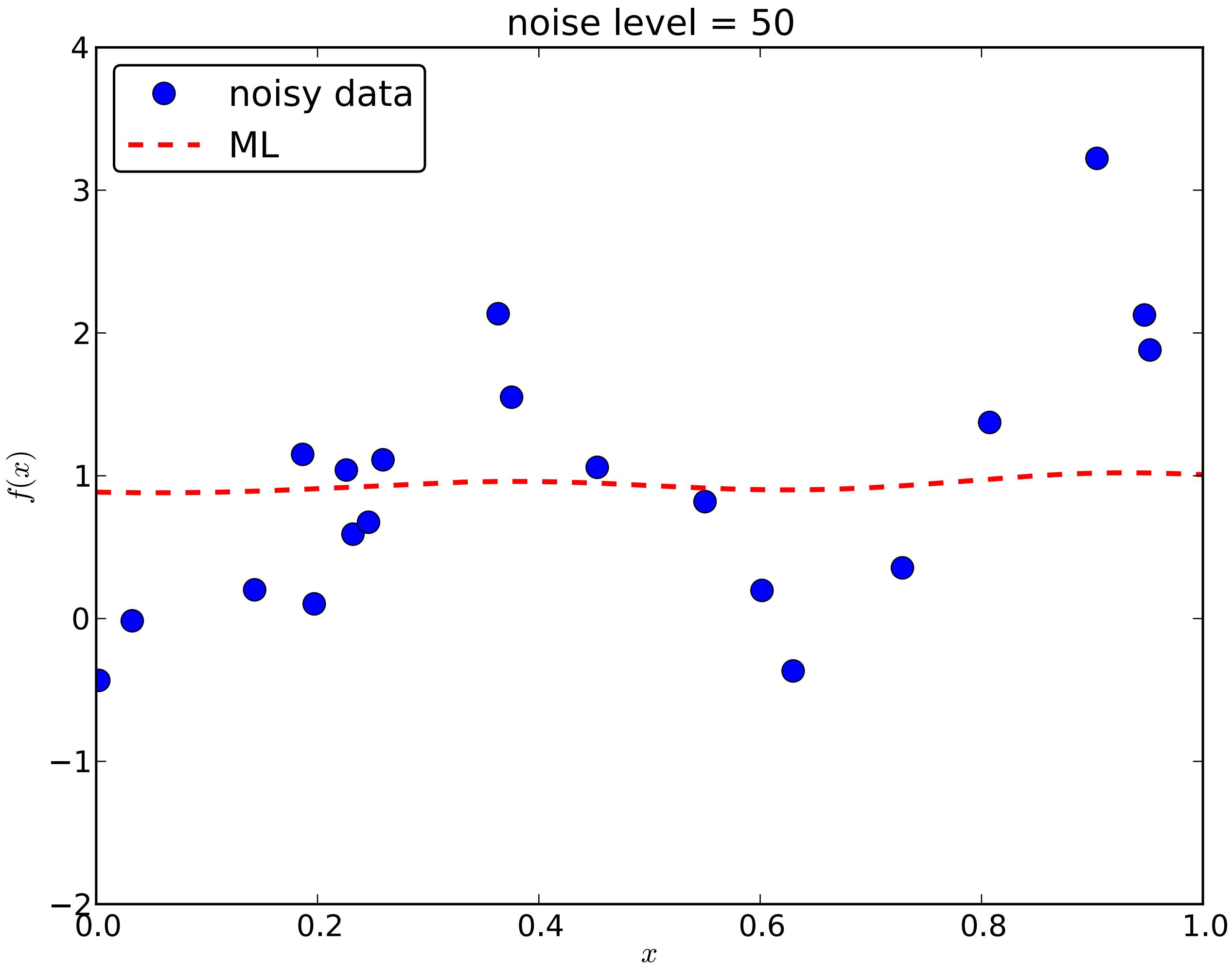
noise level

$$\boldsymbol{\alpha} = (K + \lambda^2 I)^{-1} \mathbf{f}$$

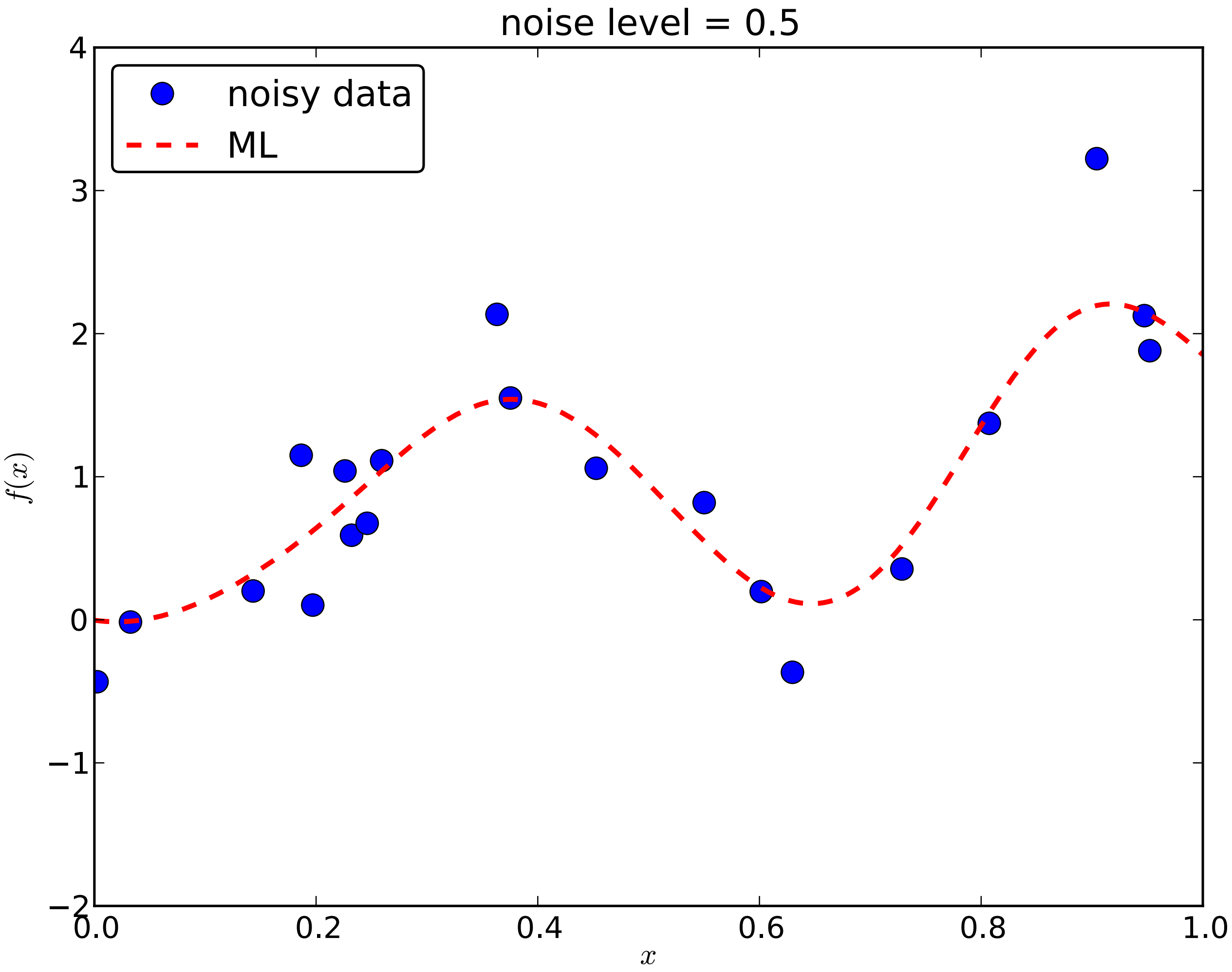
Fitting a simple function



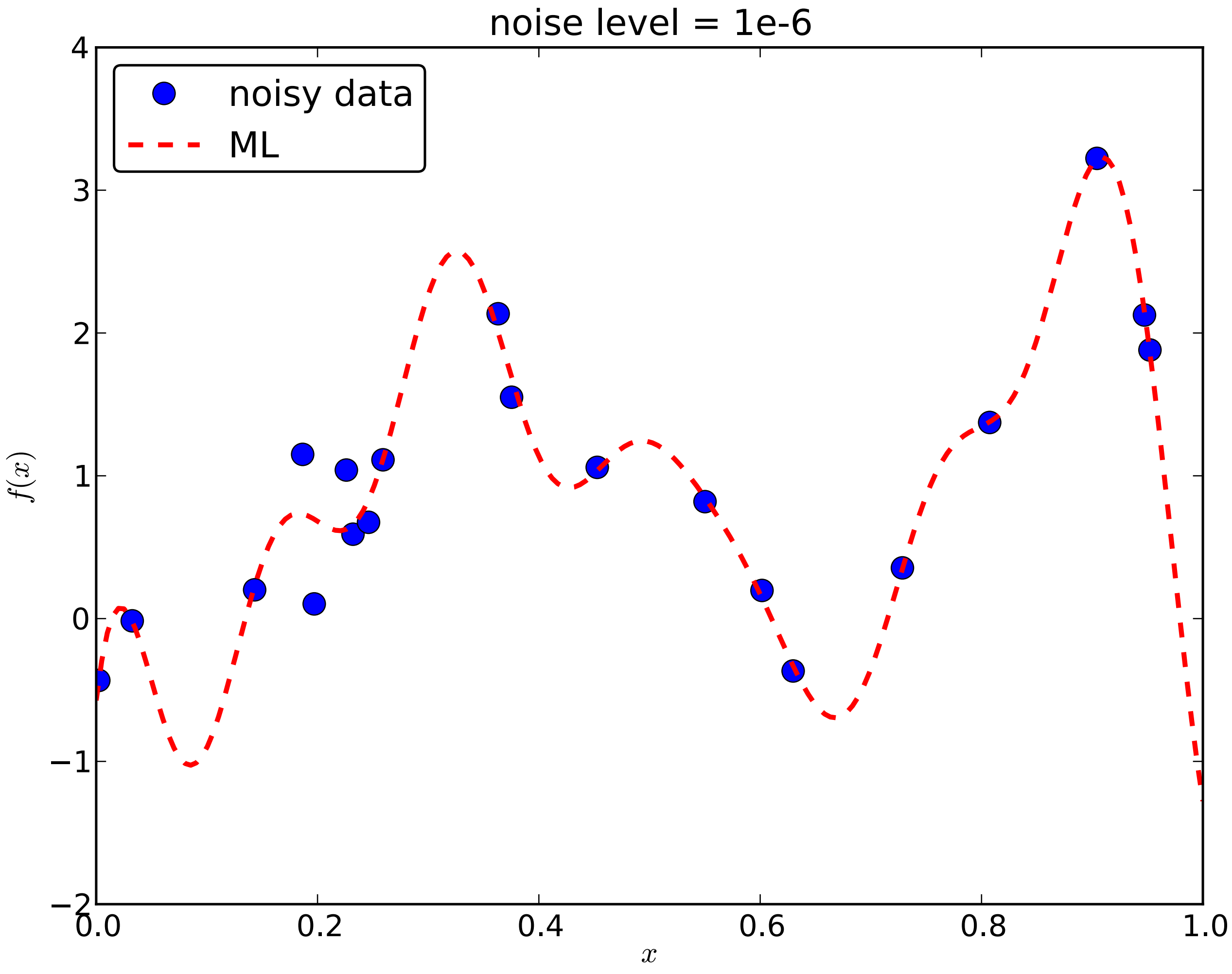
Too high noise level: underfit



Medium noise level

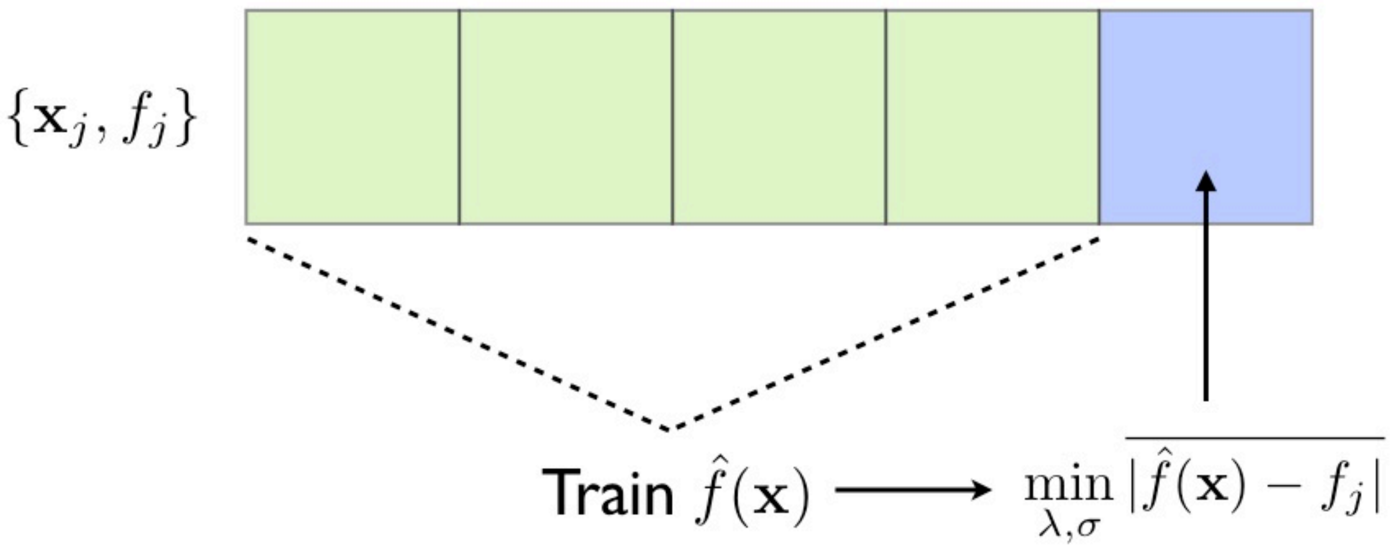


Small noise level: overfit



Cross validation

- Always optimize on samples *not* in the training set

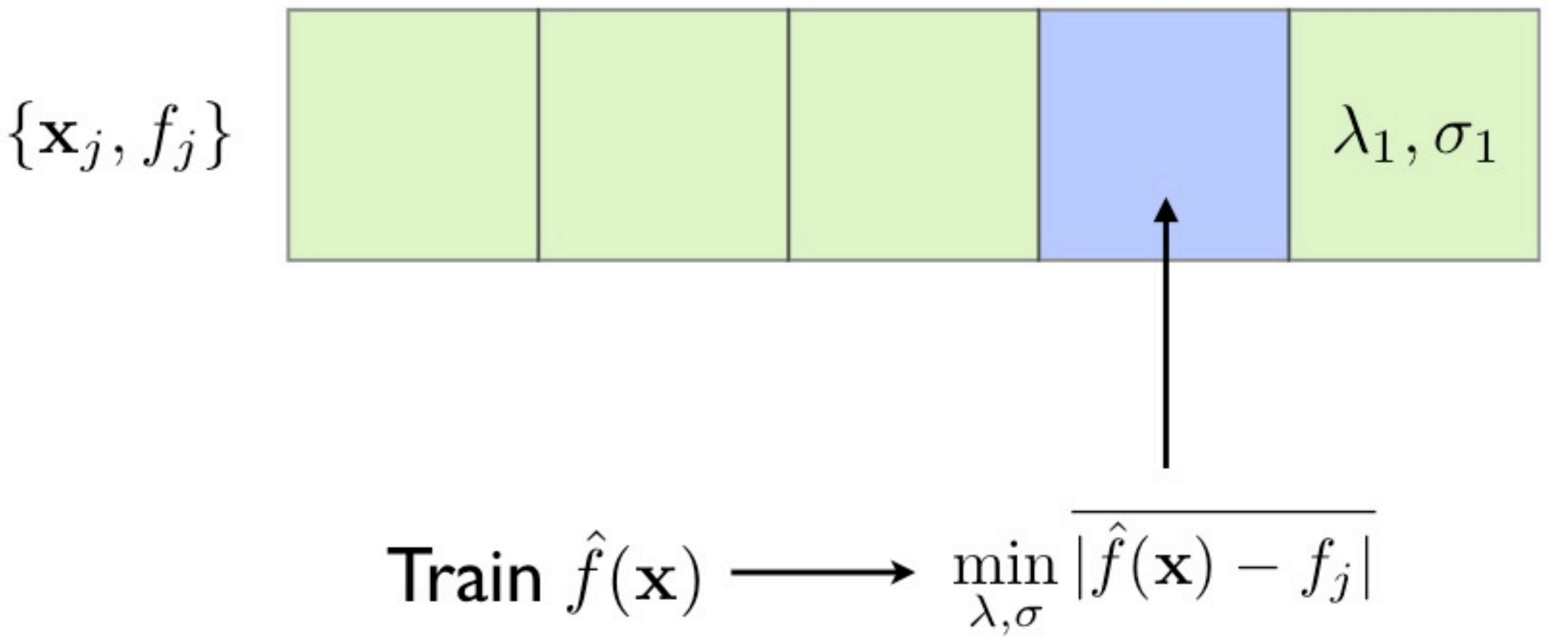


$\{\mathbf{x}_j, f_j\}$

Train $\hat{f}(\mathbf{x}) \longrightarrow \min_{\lambda,\sigma} \overline{|\hat{f}(\mathbf{x}) - f_j|}$

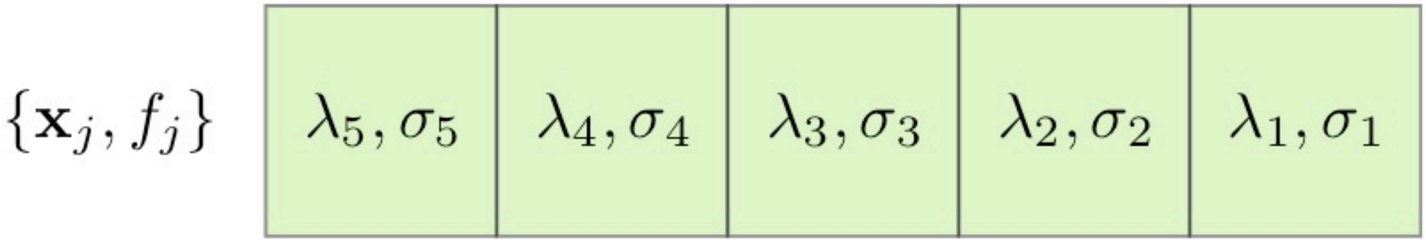
Cross validation

- Always optimize on samples *not* in the training set



More cross validation

- Always optimize on samples *not* in the training set

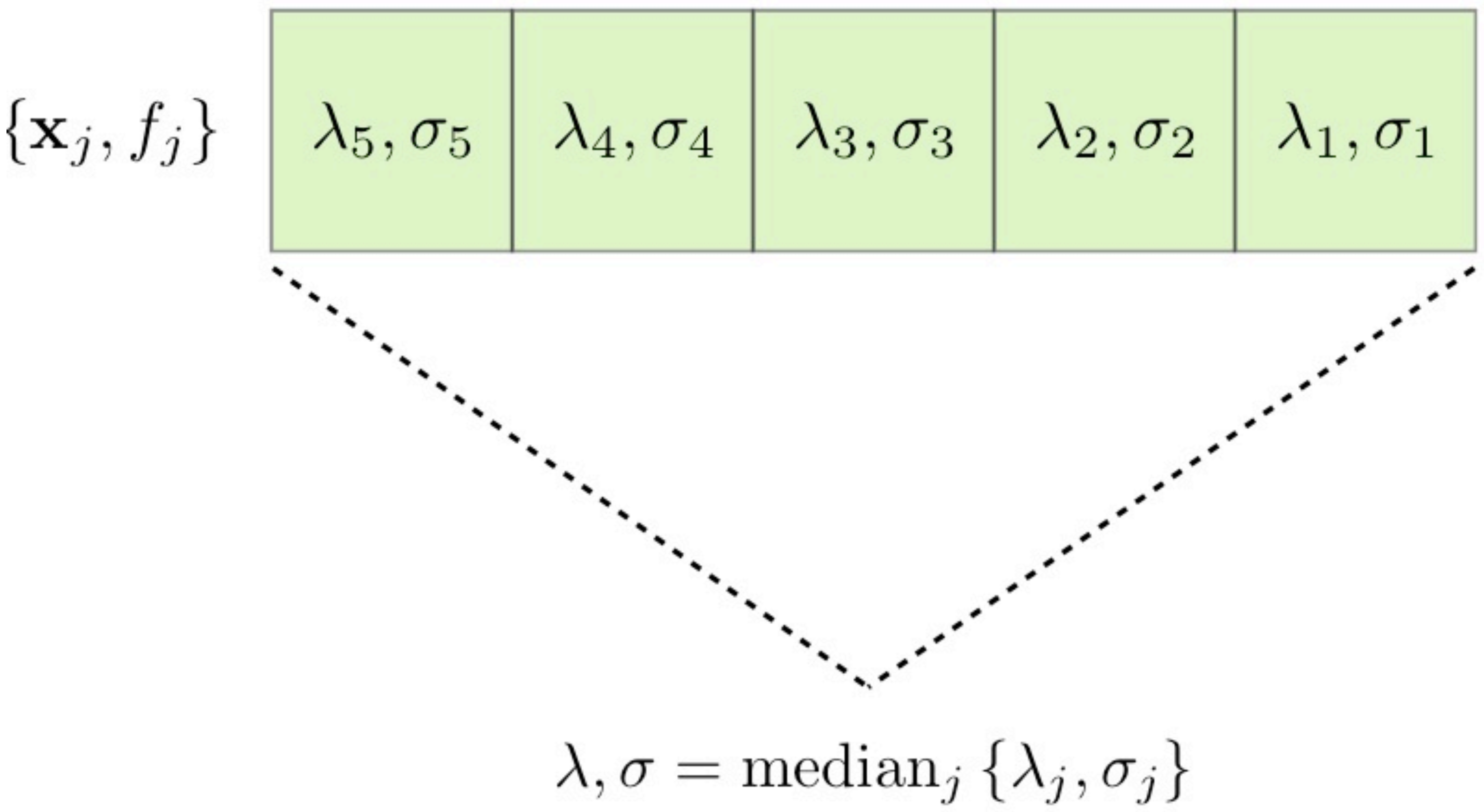


$\{\mathbf{x}_j, f_j\}$

λ_5, σ_5 | λ_4, σ_4 | λ_3, σ_3 | λ_2, σ_2 | λ_1, σ_1

Average over samples

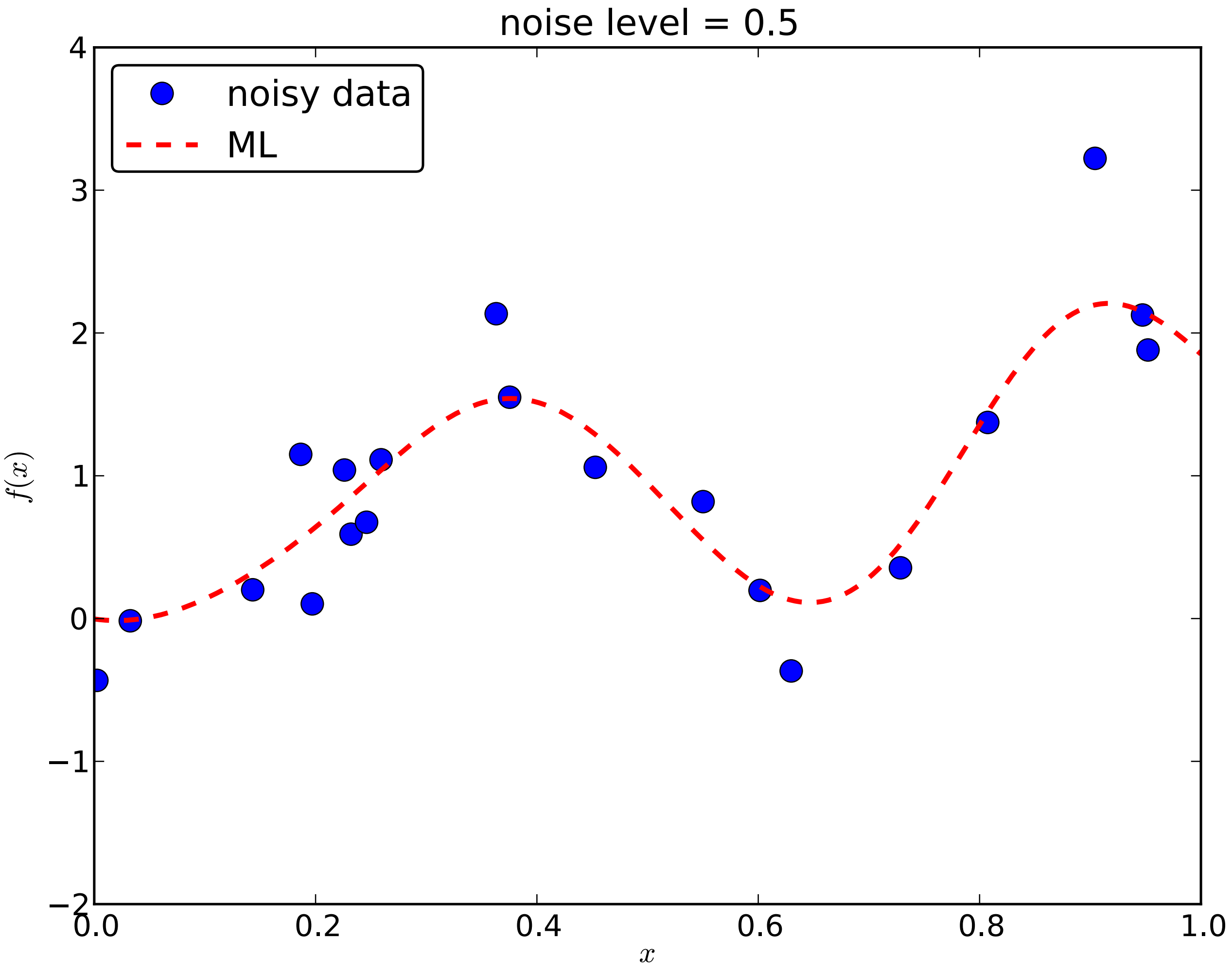
- Always optimize on samples *not* in the training set



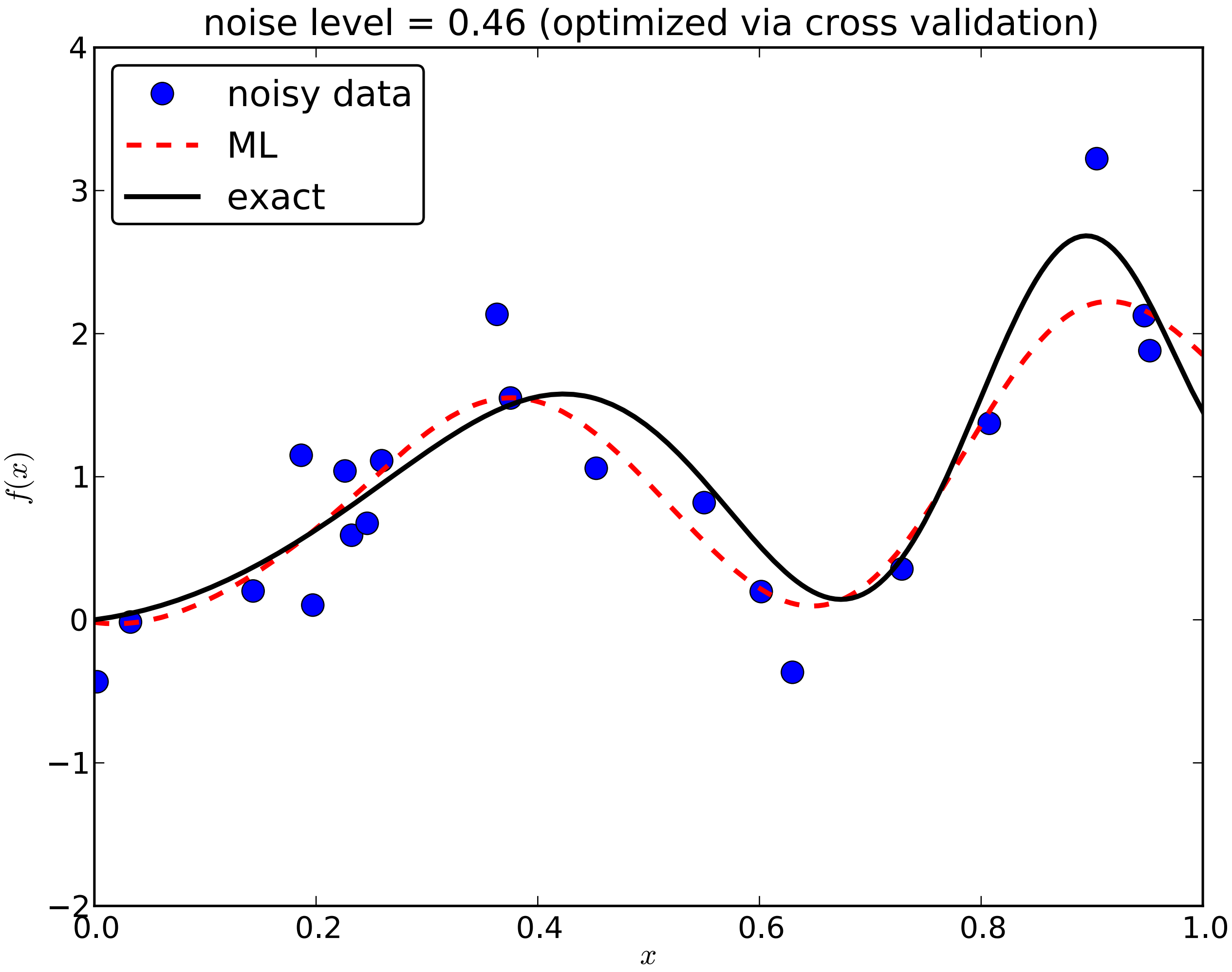
$\{\mathbf{x}_j, f_j\}$	λ_5, σ_5	λ_4, σ_4	λ_3, σ_3	λ_2, σ_2	λ_1, σ_1

$$\lambda, \sigma = \text{median}_j \, \{\lambda_j, \sigma_j\}$$

Medium noise level



Exact function and best fit



Demo problem in DFT

- *N* non-interacting same-spin fermions confined to 1d box
- Define class of potential:

$$v(x) = -\sum_{i=1}^{3} a_i \exp(-(x-b_i)^2/(2c_i^2))$$

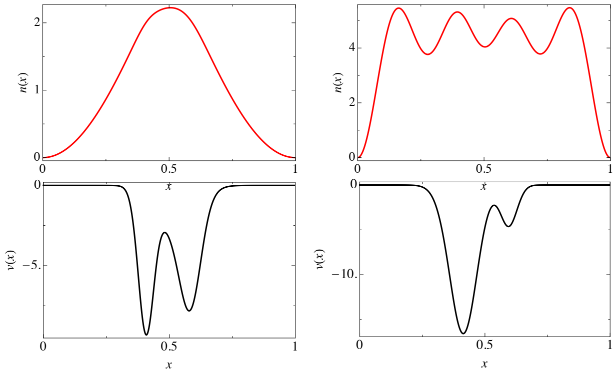
- Represent the density on a grid with spacing $\Delta x = 1/(G-1)$
- ML-DFA for KE:

$$\hat{T}(\mathbf{n}) = \bar{T}\sum_{j=1}^{M} \alpha_j k(\mathbf{n}_j, \mathbf{n})$$

$k[n,n'] = \exp\left(-\int dx\, (n(x)-n'(x))^2/(2\sigma^2)\right)$

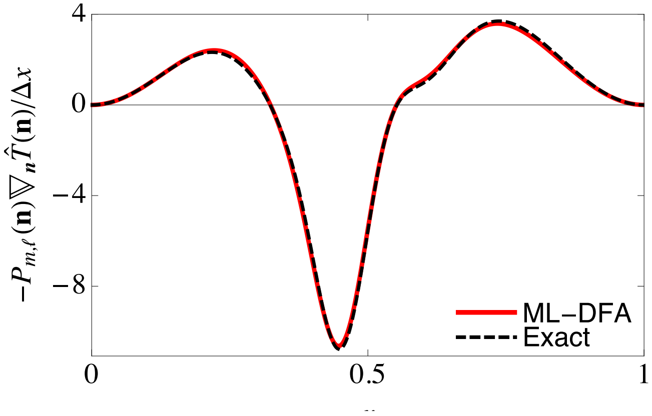
Test case: KS electrons in a box

Generate 2000 potentials. Solve for up to 4 electrons.



kcal/mol

N	M	λ	σ	$\overline{\lvert\Delta T\rvert}$	$\lvert\Delta T\rvert^{\text{std}}$	$\lvert\Delta T\rvert^{\max}$
1	40	2.4×10^{-5}	238	3.3	3.0	23.
	60	1.0×10^{-5}	95	1.2	1.2	10.
	80	6.7×10^{-6}	48	0.43	0.54	7.1
	100	3.4×10^{-7}	43	0.15	0.24	3.2
	150	2.5×10^{-7}	33	0.060	0.10	1.3
	200	1.7×10^{-7}	28	0.031	0.053	0.65
2	100	1.3×10^{-7}	52	0.13	0.20	1.8
3	100	2.0×10^{-7}	74	0.12	0.18	1.8
4	100	1.4×10^{-7}	73	0.078	0.14	2.3
1-4†	400	1.8×10^{-7}	47	0.12	0.20	3.6



LDA ~ 223 kcal/mol, Gradient correction ~ 159 kcal/mol

Finding Density Functionals with Machine Learning John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, Kieron Burke, Phys. Rev. Lett. **108**, 253002 (2012)

Accurate energies on coarse grids

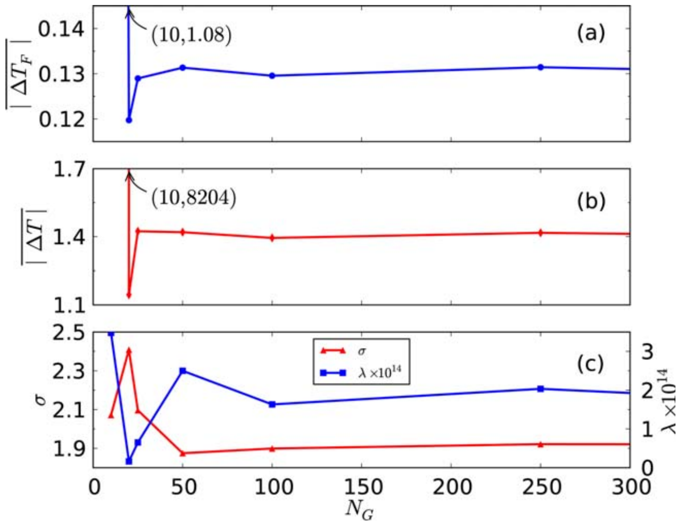
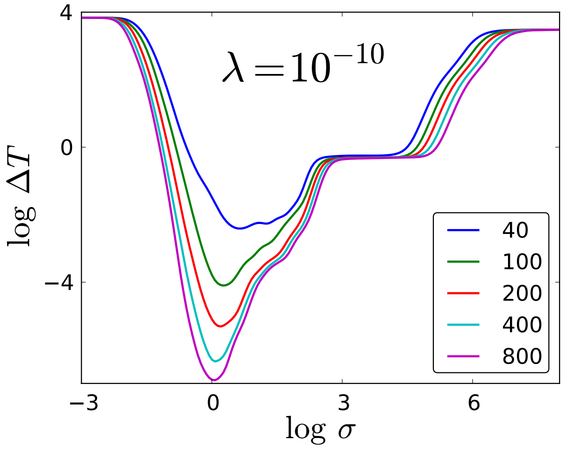


Figure 7. The effect of using a sparse grid to represent the density on the performance of the MLA, for $N=1$, $N_T=100$, with the Gaussian kernel. Here (a) $\overline{|\Delta T_F|}=\overline{T^{ML}[n]-T[n]}$ is the mean absolute functional-driven error of the MLA evaluated on the test set in kcal/mol, (b) $\overline{|\Delta T|}=\overline{T^{ML}[\tilde{n}]-T[n]}$ gives the error of KE evaluated on constrained optimal densities in kcal/mol and (c) the corresponding re-crossvalidated hyperparameters λ and σ. The MAE is completely unaffected as N_G is reduced until $\sim N_G=10$, when it jumps sharply.

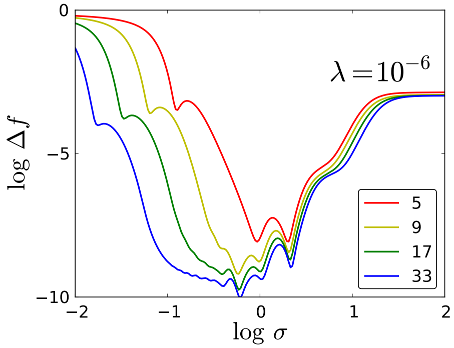
Understanding Machine-learned Density Functionals Li Li, John C. Snyder, Isabelle M. Pelaschier, Jessica Huang, Uma-Naresh Niranjan, Paul Duncan, Matthias Rupp, Klaus-Robert Müller, Kieron Burke, *International Journal of Quantum Chemistry* **116**, 819--833 (2016)

Kevin’s paper: from functions to functionals

- Plot error as a function of hyperparameters



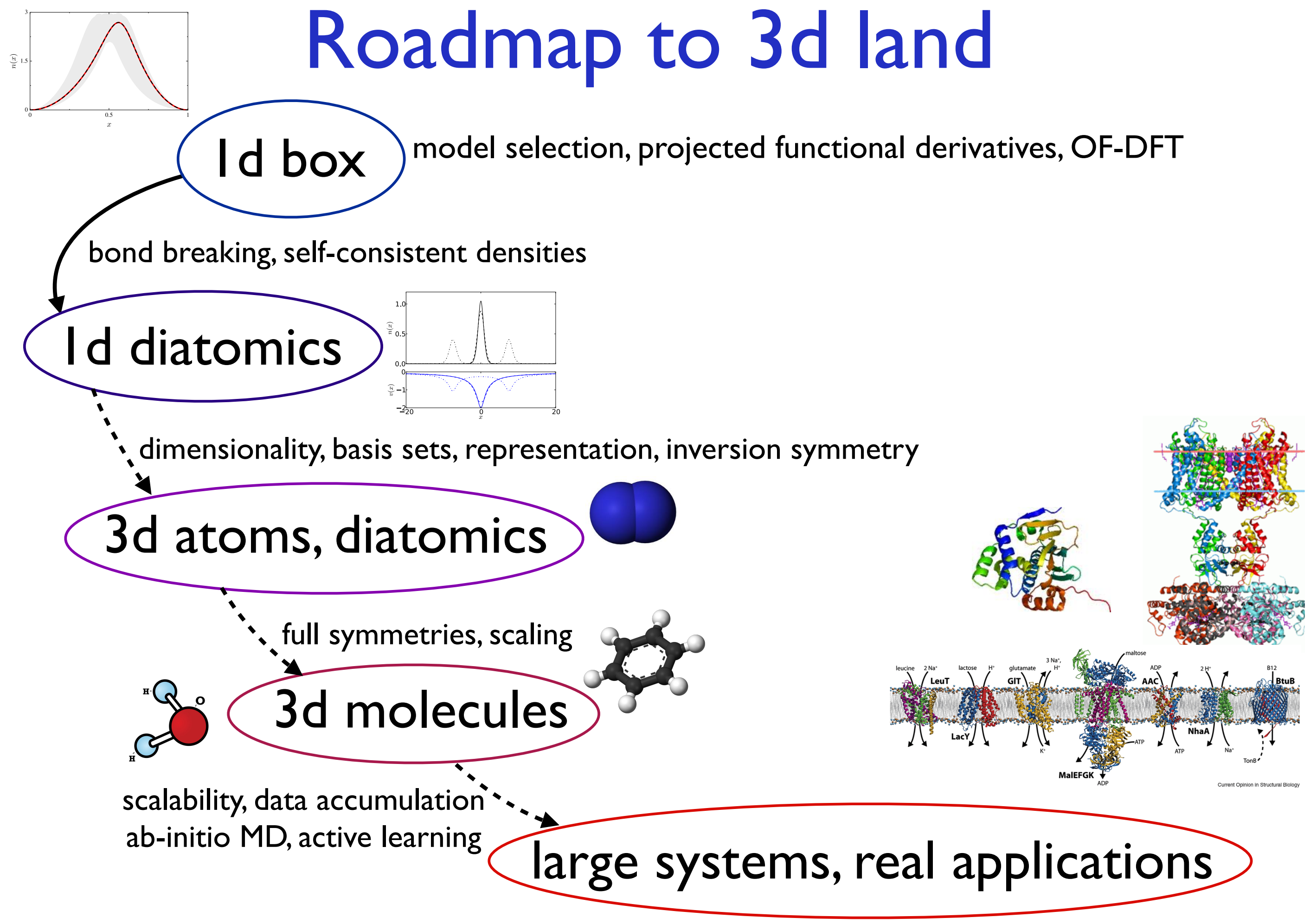
- Repeat for fitting f(x)= cos x



- Curves have roughly the same “valley” shape for all N_T
- Bottom of the valley is an order of magnitude deeper than the walls
- These valleys are nearly identical in shape for sufficiently large N_T, which indicates that this particular feature arises in a systematic manner as N_T increases

Understanding kernel ridge regression: Common behaviors from simple functions to density functionals, Kevin Vu, John C. Snyder, Li Li, Matthias Rupp, Brandon F. Chen, Tarek Khelif, Klaus-Robert Müller, Kieron Burke, *International Journal of Quantum Chemistry* **115**, 1115--1128 (2015).

Road map back to reality



Roadmap to 3d land

1d box

model selection, projected functional derivatives, OF-DFT

bond breaking, self-consistent densities

1d diatomics

dimensionality, basis sets, representation, inversion symmetry

3d atoms, diatomics

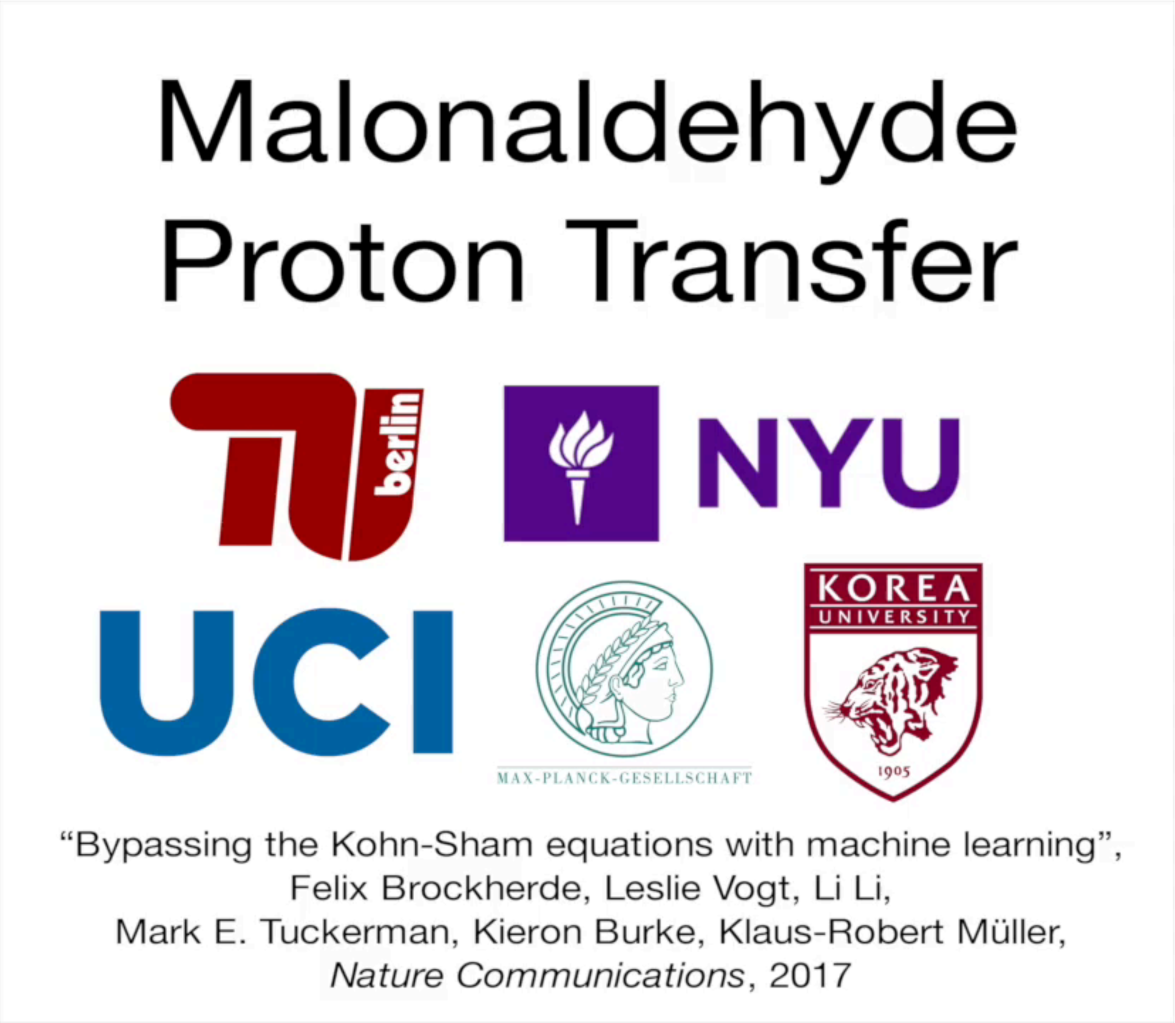
full symmetries, scaling

3d molecules

scalability, data accumulation
ab-initio MD, active learning

large systems, real applications

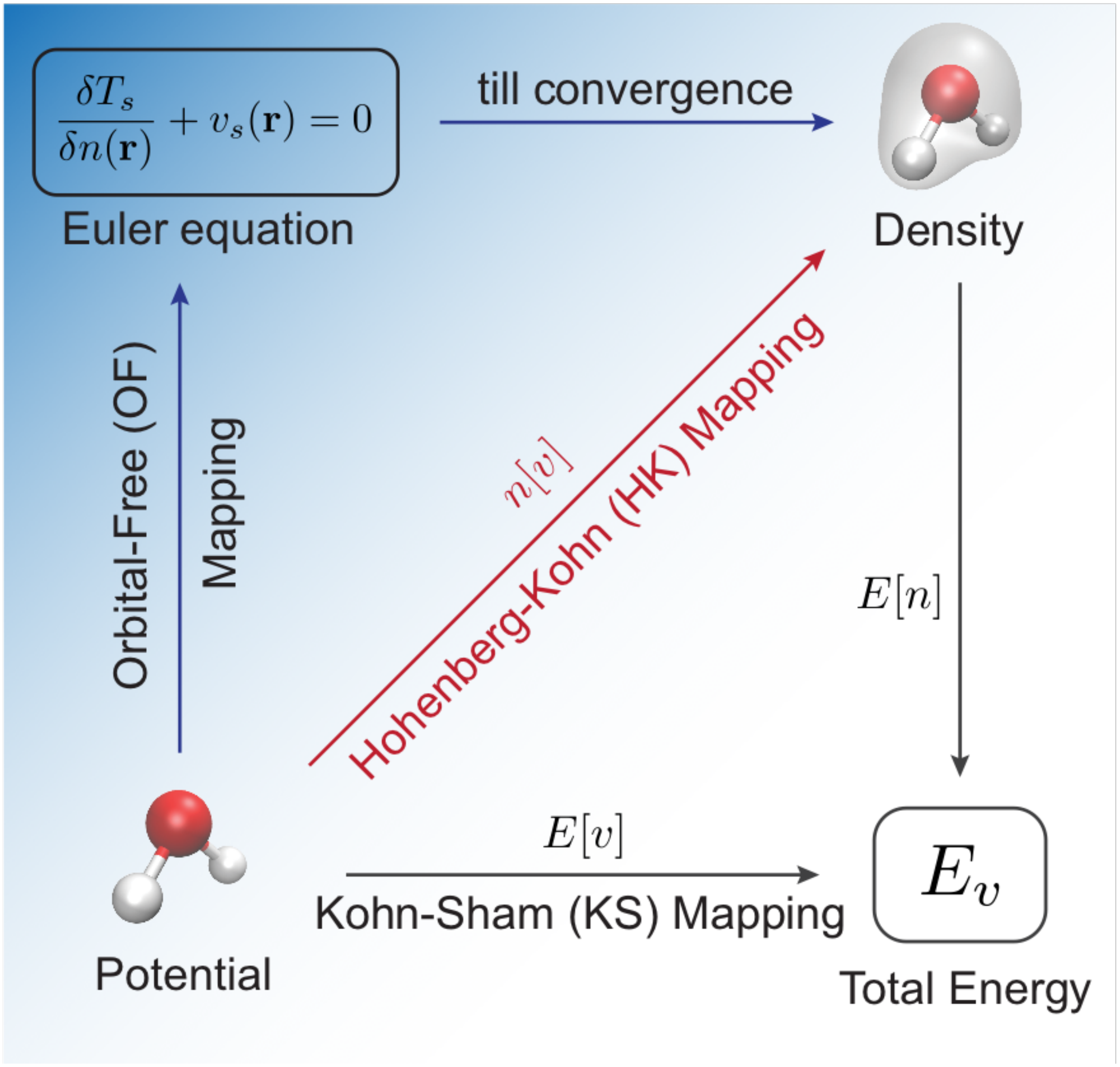
Bypassing KS equations



Malonaldehyde
Proton Transfer

"Bypassing the Kohn-Sham equations with machine learning",
Felix Brockherde, Leslie Vogt, Li Li,
Mark E. Tuckerman, Kieron Burke, Klaus-Robert Müller,
Nature Communications, 2017

By-passing KS



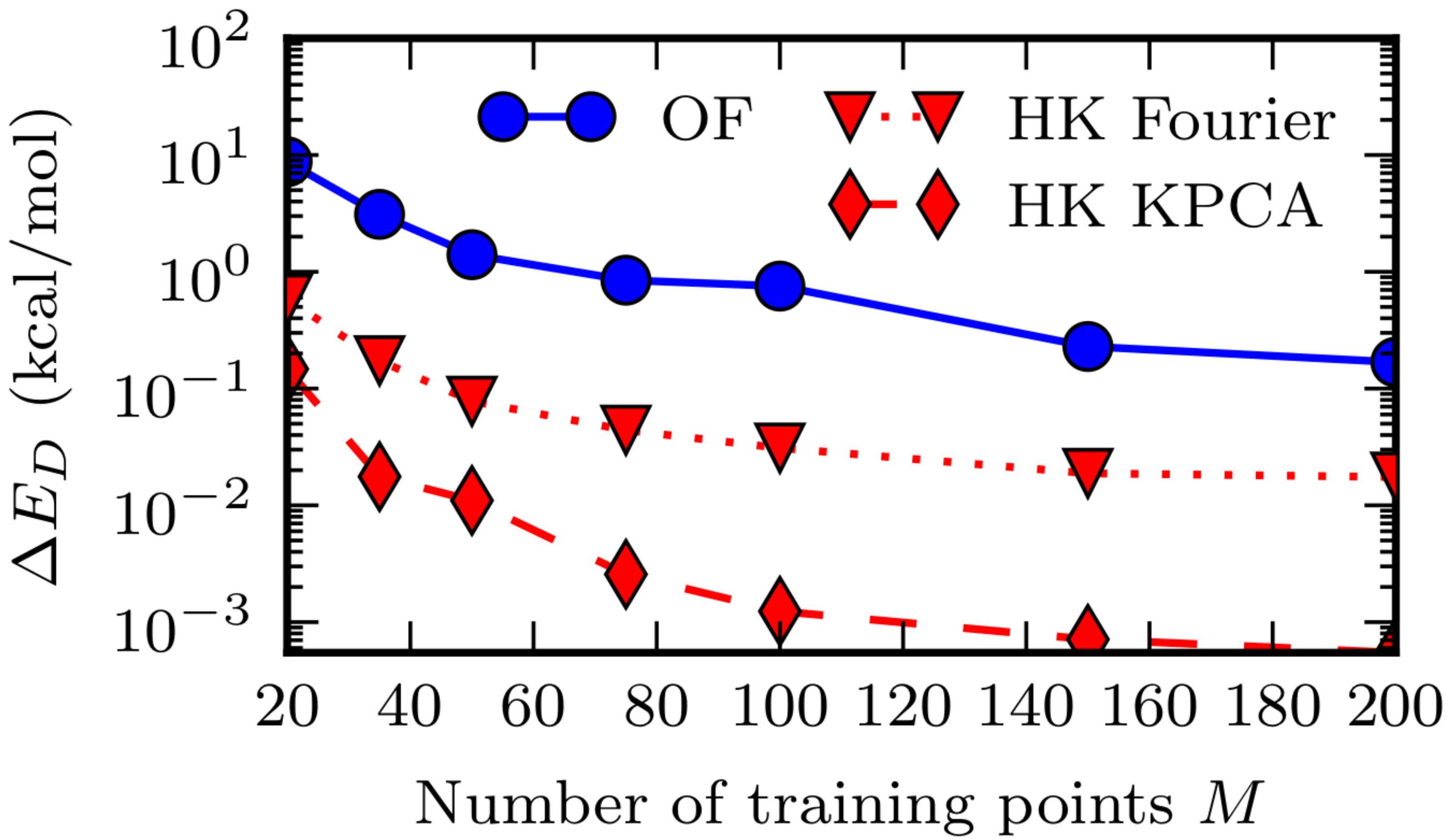
Performance of ML for HK map: Box problem

	ML-OF						ML-HK (grid)						ML-HK (other)			
	ΔE		ΔE_F		ΔE_D		ΔE		ΔE_D		ΔE_D^{ML}		ΔE_D (Fourier)		ΔE_D (KPCA)	
M	MAE	max	MAE	max	MAE	max	MAE	max	MAE	max	MAE	max	MAE	max	MAE	max
20	7.7	47	7.7	60	8.8	87	3.5	27	0.76	8.9	9.7	70	0.58	8	0.15	2.9
50	1.6	30	1.3	7.3	1.4	31	1.2	7.1	0.079	0.92	0.27	2.4	0.078	0.91	0.011	0.17
100	0.74	17	0.2	2.6	0.75	17	0.19	2.1	0.027	0.43	0.18	2.4	0.031	0.42	0.0012	0.028
200	0.17	2.9	0.039	0.6	0.17	2.9	0.042	0.59	0.0065	0.15	0.02	0.46	0.017	0.14	0.00055	0.015

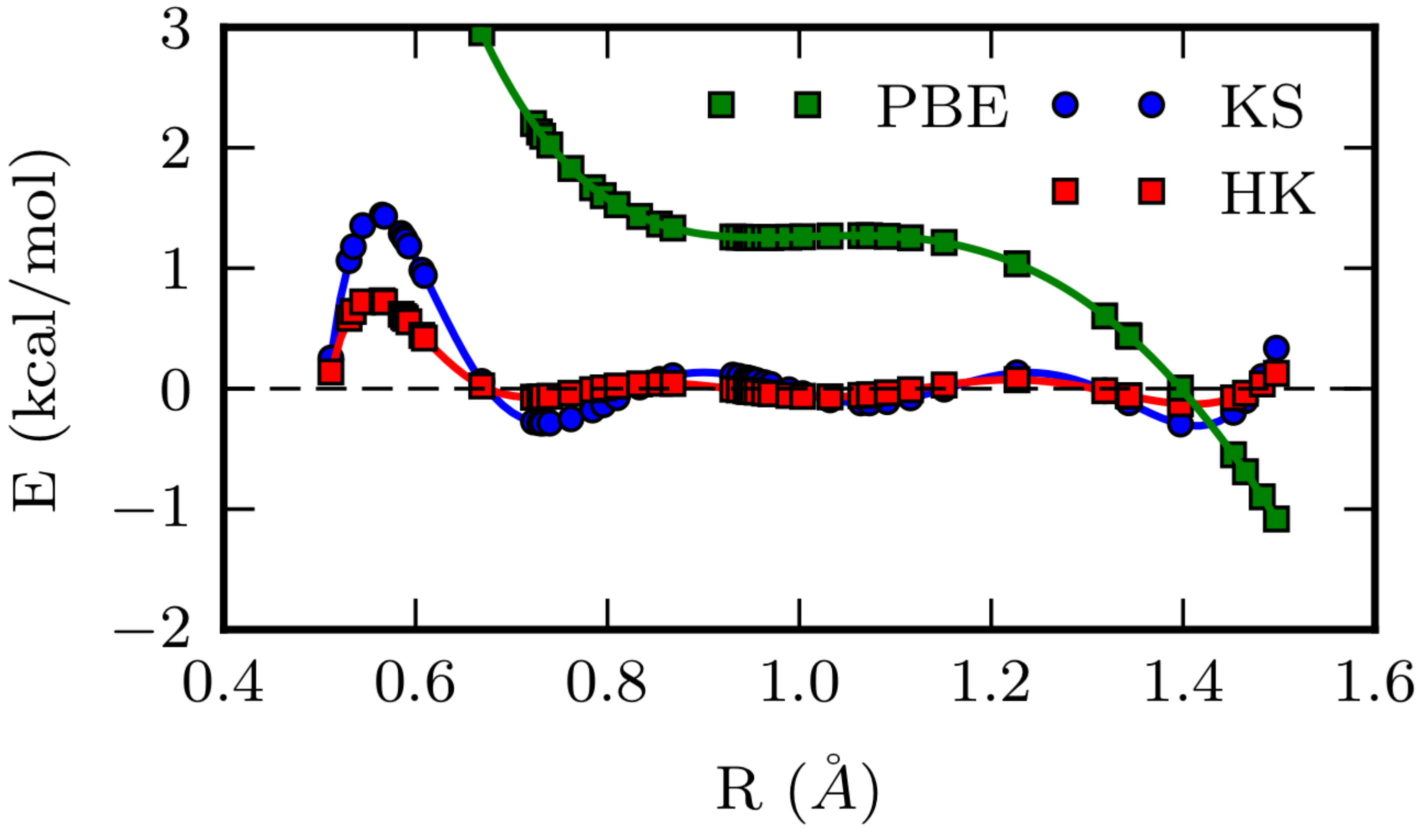
Table I. Energy errors in kcal/mol for the 1-D data set for various M, the number of training points. For definitions, see text.

Understanding and reducing errors in density functional calculations Min-Cheol Kim, Eunji Sim, Kieron Burke, *Phys. Rev. Lett.* **111**, 073003 (2013).

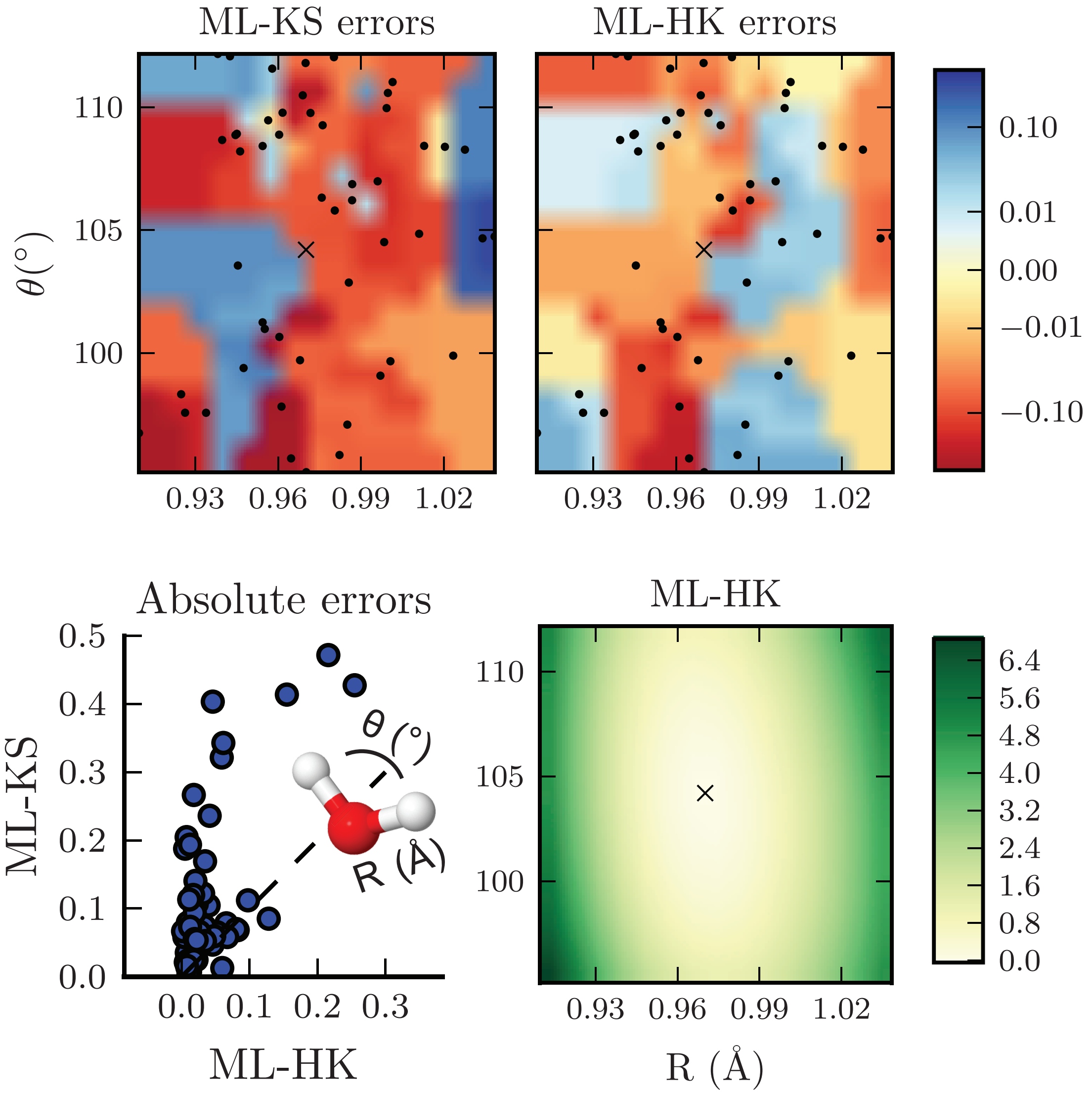
Convergence of different HK maps



Error for H_2



H_2O


ML-KS errors
ML-HK errors
110
105
100
θ(°)
0.93
0.96
0.99
1.02
0.10
0.01
0.00
−0.01
−0.10
Absolute errors
0.5
0.4
0.3
0.2
0.1
0.0
ML-KS
0.0
0.1
0.2
0.3
ML-HK
θ (°)
R (Å)
ML-HK
110
105
100
0.93
0.96
0.99
1.02
R (Å)
6.4
5.6
4.8
4.0
3.2
2.4
1.6
0.8
0.0

Accuracy of densities

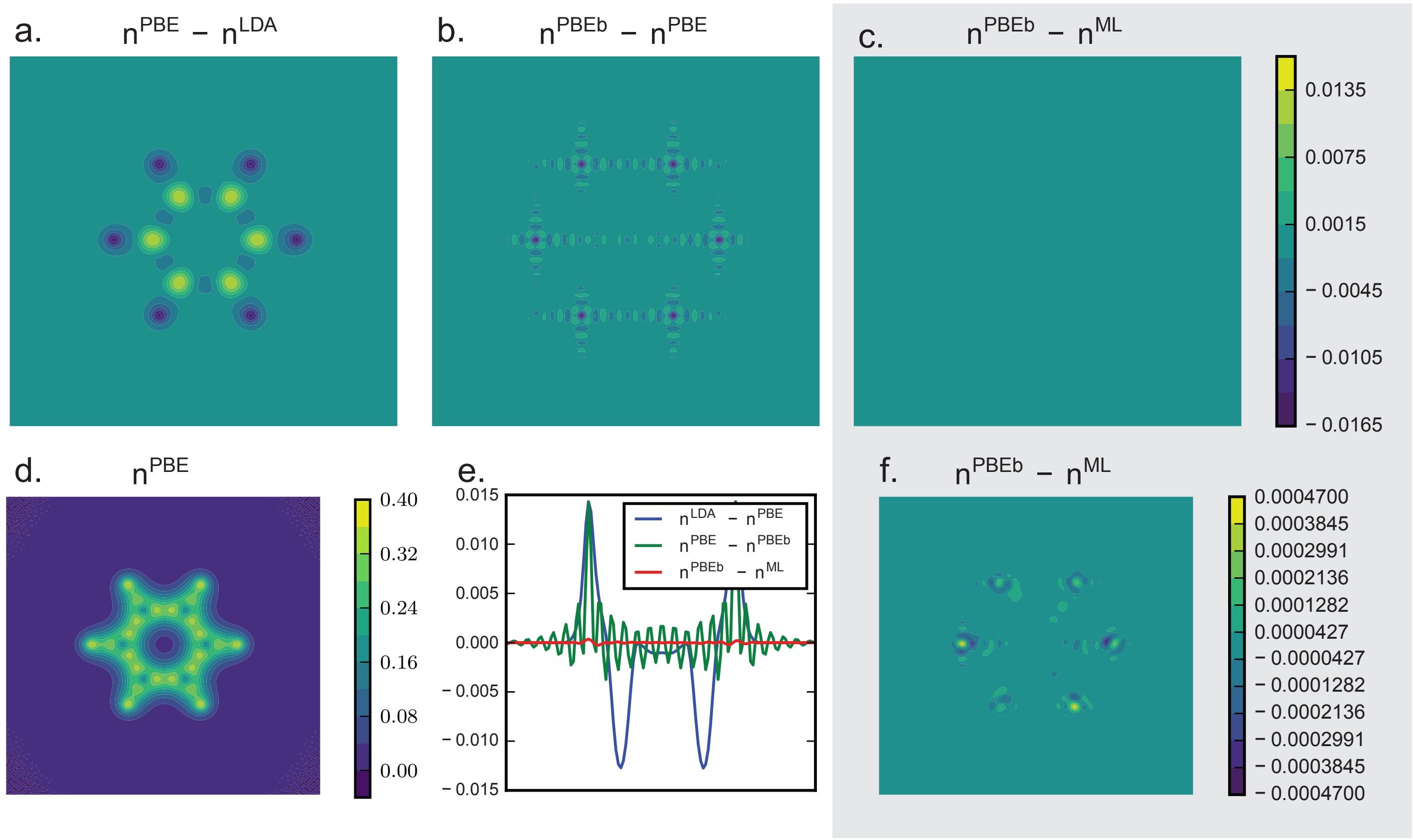


Figure 4. The precision of our density predictions using the Fourier basis for ML-HK for the molecular plane of benzene. The plots show **a**. the difference between the valence density of benzene when using PBE and LDA functionals at the PBE optimized geometry. **b**. error introduced by using the Fourier basis representation. **c**. error introduced by the $n^{\mathrm{ML}}[v]$ density fitting (a. - c. on same color scale). **d**. the total PBE valence density **e**. the density differences along a 1-D cut of a. - c. **f**. the density error introduced with the ML-HK map (same data, but different scale, as in c.).

MD simulations testing ML method

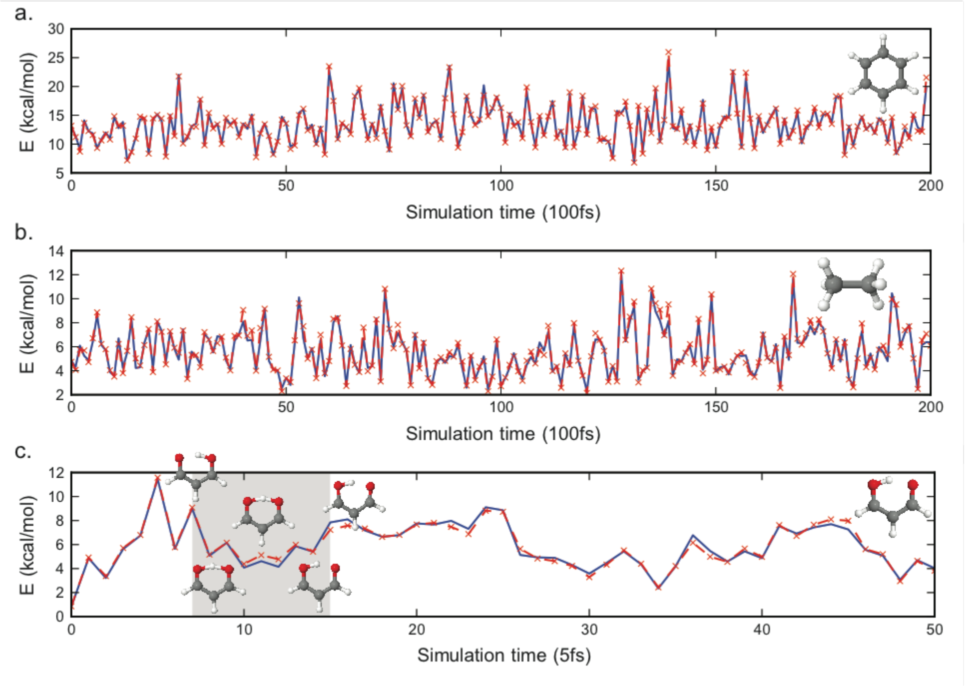


Figure 3. Energy errors of ML-HK along MD trajectories. PBE values in blue, ML-HK values in red. **a.** A 2 ps classical trajectory of benzene. **b.** A 2 ps classical trajectory of ethane. **c.** A 0.25 ps ab-initio trajectory of malonaldehyde. The ML model correctly predicts energies during a proton transfer in frames 7 to 15 without explicitly including these geometries in the training set.

	Benzene		Ethane		Malonaldehyde	
Training trajectories	MAE	max	MAE	max	MAE	max
300K	0.395742	1.92642	0.212137	1.33947		
300K + 350K	0.260517	1.76190	0.236088	1.38227	0.206795	0.725515
300K + 400K	0.370876	2.1162	0.101054	0.576107		

Table V. Errors (ΔE_D in kcal/mol) on the MD datasets for different training trajectory combinations.

XC for strong correlation in thermodynamic limit (in 1D)

Mott-Hubbard gap

More Difficult

- Classic prototype of condensed matter
- Infinite chain of H atoms
- When lattice spacing is large, must be an insulator
- But with one electron per site, always a band metal

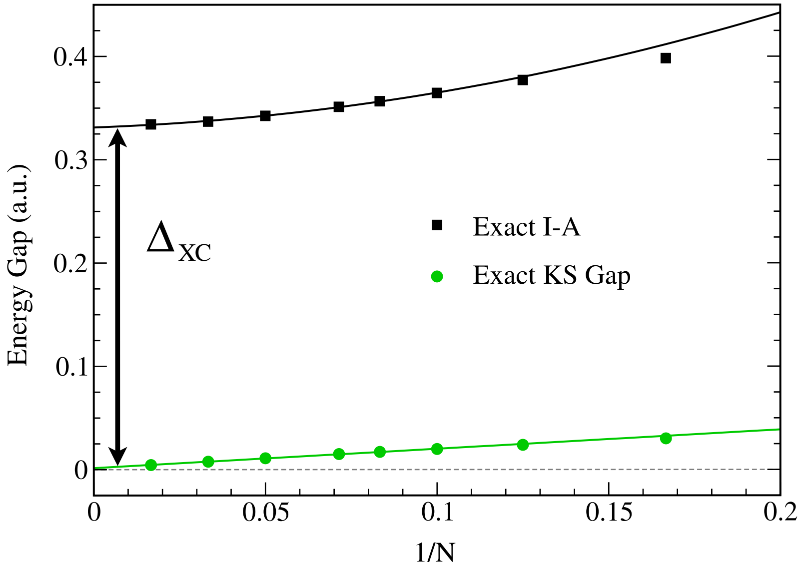


FIG. 3 (color online). Exact gaps for chains of N soft hydrogen atoms with atomic separation $b = 4$ (error bars are less than symbol sizes). The upper curve is a quadratic fit of exact gaps of the largest six systems and extrapolates to a finite value $E_g \simeq 0.33$. The exact Kohn-Sham gaps, in contrast, extrapolate to zero showing that for $N \to \infty$ the true KS system is metallic (lower curve is a linear fit of exat KS gaps of the largest six systems).

[133] *One-Dimensional Continuum Electronic Structure with the Density-Matrix Renormalization Group and Its Implications for Density-Functional Theory* E.M. Stoudenmire, Lucas O. Wagner, Steven R. White, Kieron Burke, Phys. Rev. Lett. **109**, 056402 (2012).

[177] Understanding Band Gaps of solids in generalized Kohn-Sham Theory, J P Perdew, W Yang, K Burke, Z Yang, E K U Gross, M Scheffler, G E Scuseria, T M Henderson, I Y Zhang, A Ruzsinszky, H Peng, J Sun, *Proceedings of the National Academy of Sciences,* 2801-2806 (2016).

Convergence for H_2

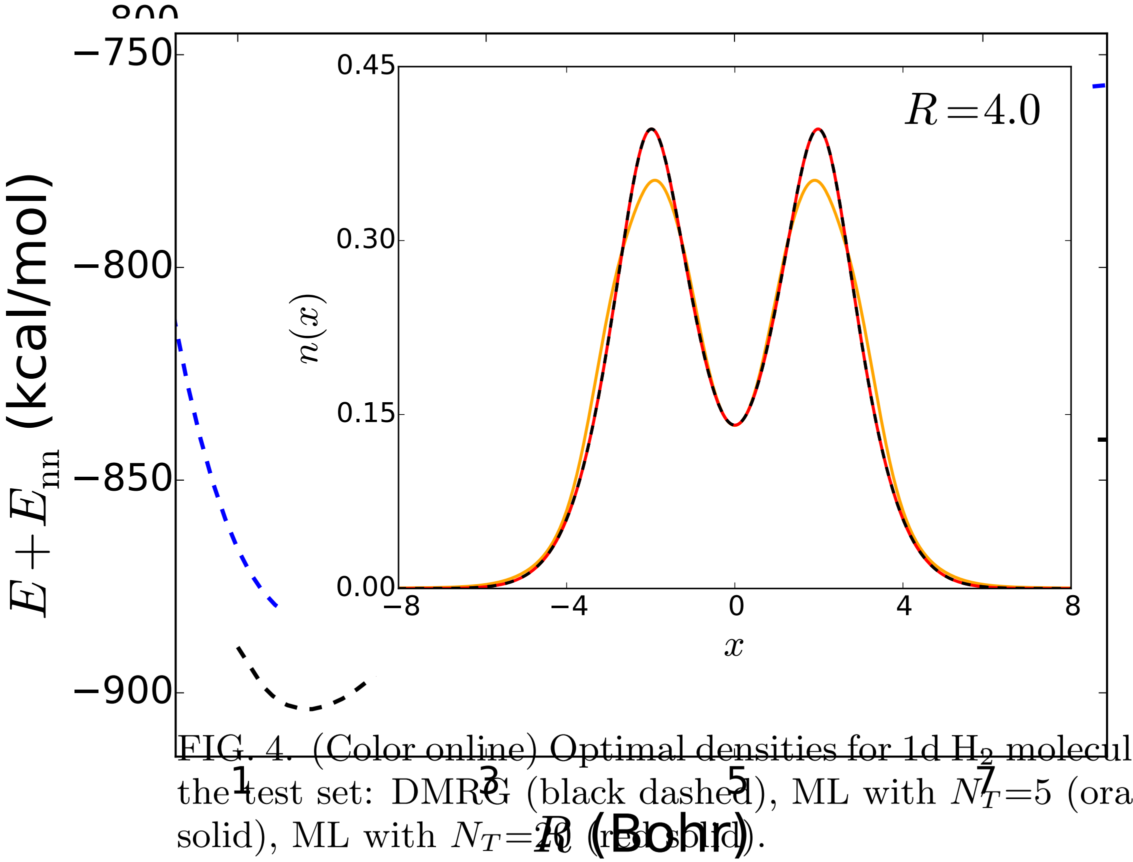


FIG. 4. (Color online) Optimal densities for 1d H_2 molecule in the test set: DMRG (black dashed), ML with N_T=5 (orange solid), ML with N_T=20 (red solid).

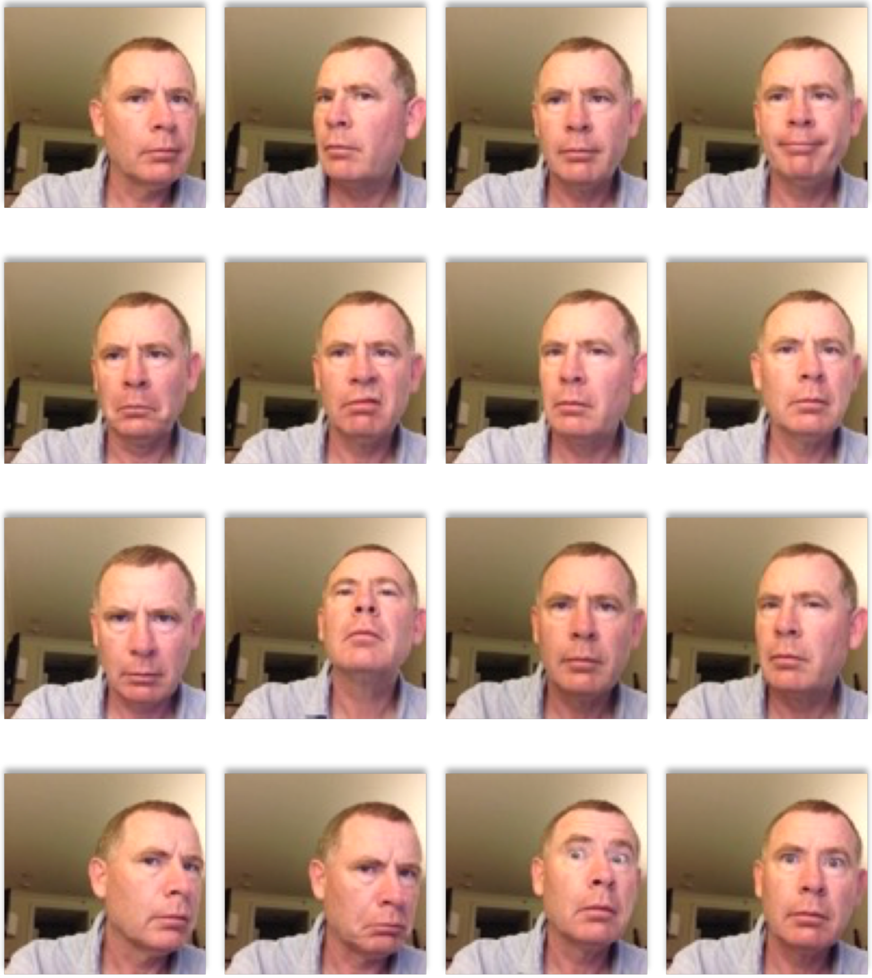
FIG. 3. (Color online) Same as Fig. 2. The green curves are ML with $N_{\rm T} = 5$ on both the exact (dashed) and ML-optimized (solid) densities. The red solid curve is the ML

FIG. 2. (Color online) Binding curve for a 1d H_2 molecule. Black: highly accurate, converged DMRG results. Blue: LDA

Vital issue in ML: Representation of data

- We want to calculate F[n] sufficiently accurately to solve Euler equation directly for the density.
- Have all those problems with functional derivative.
- Amount of data needed explodes as chain length increases.
- Need better representation for the data.
- Li's thesis problem.

Facial recognition via PCA



Mean face

Plus one principal component

Plus two

Plus three

Plus four

Plus 5

Plus 6

Plus 7

Plus 8

Original

PCA basis for atomic densities

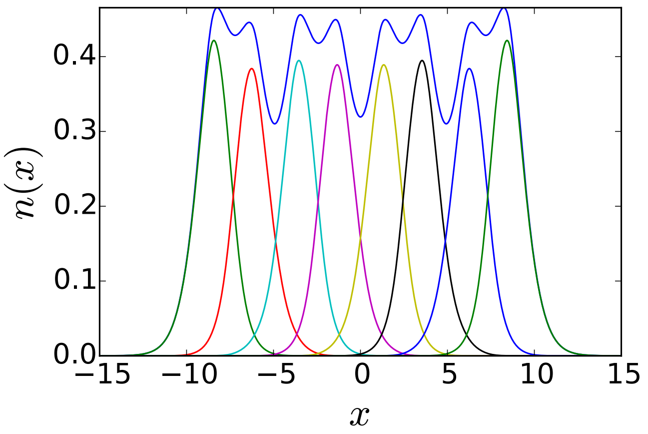


FIG. 5. Partition density of each H atom in H_8.

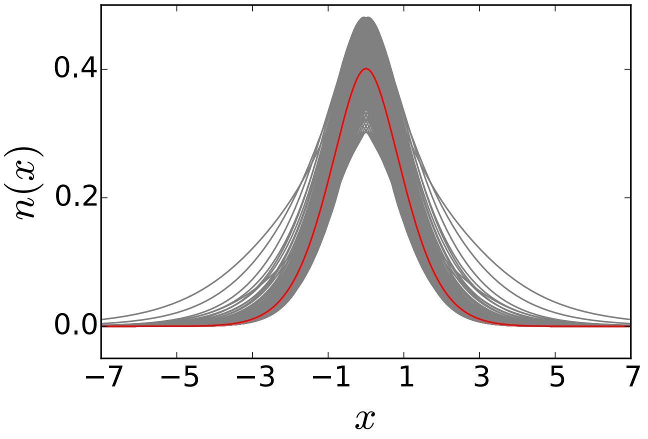


FIG. 6. Single H atom densities for H atoms in different chains and atomic distance (gray). The average density is plotted in red.

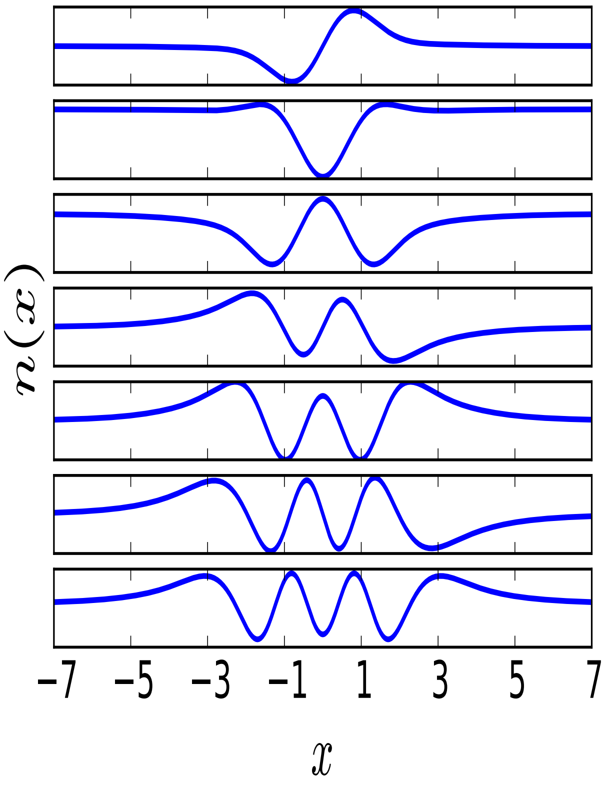


FIG. 7. First 7 principal components of the densities shown in Fig. 6, from top to bottom.

Improved convergence from basis

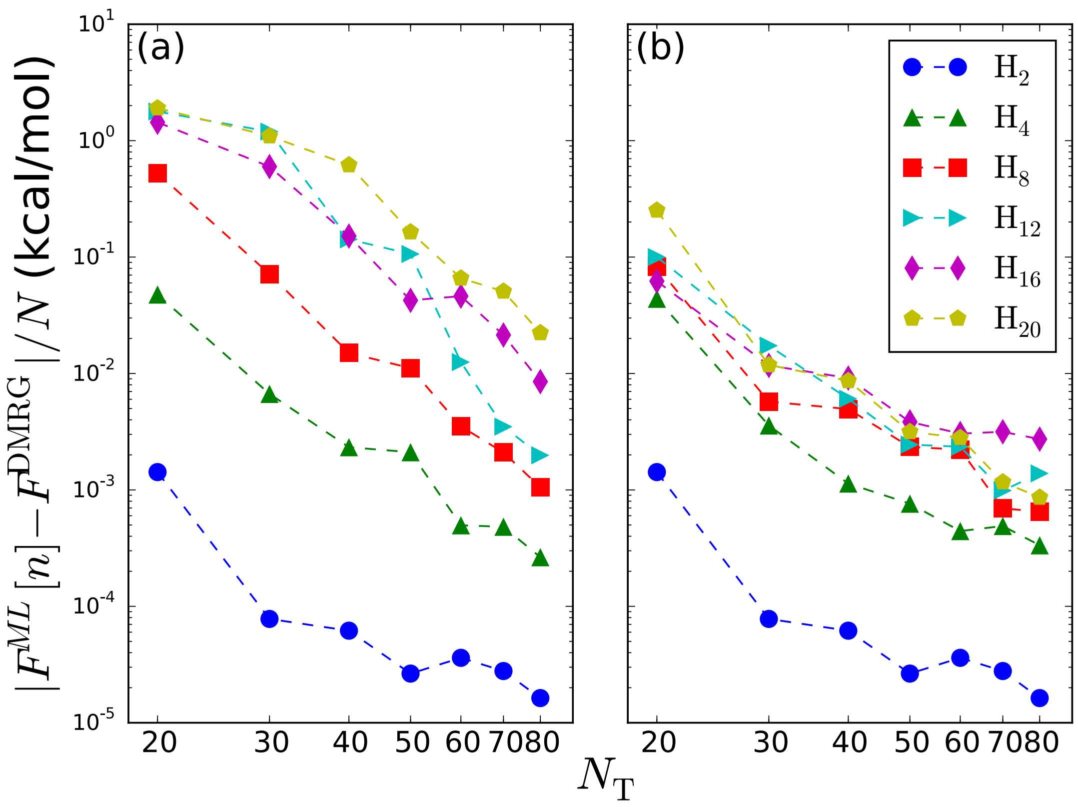
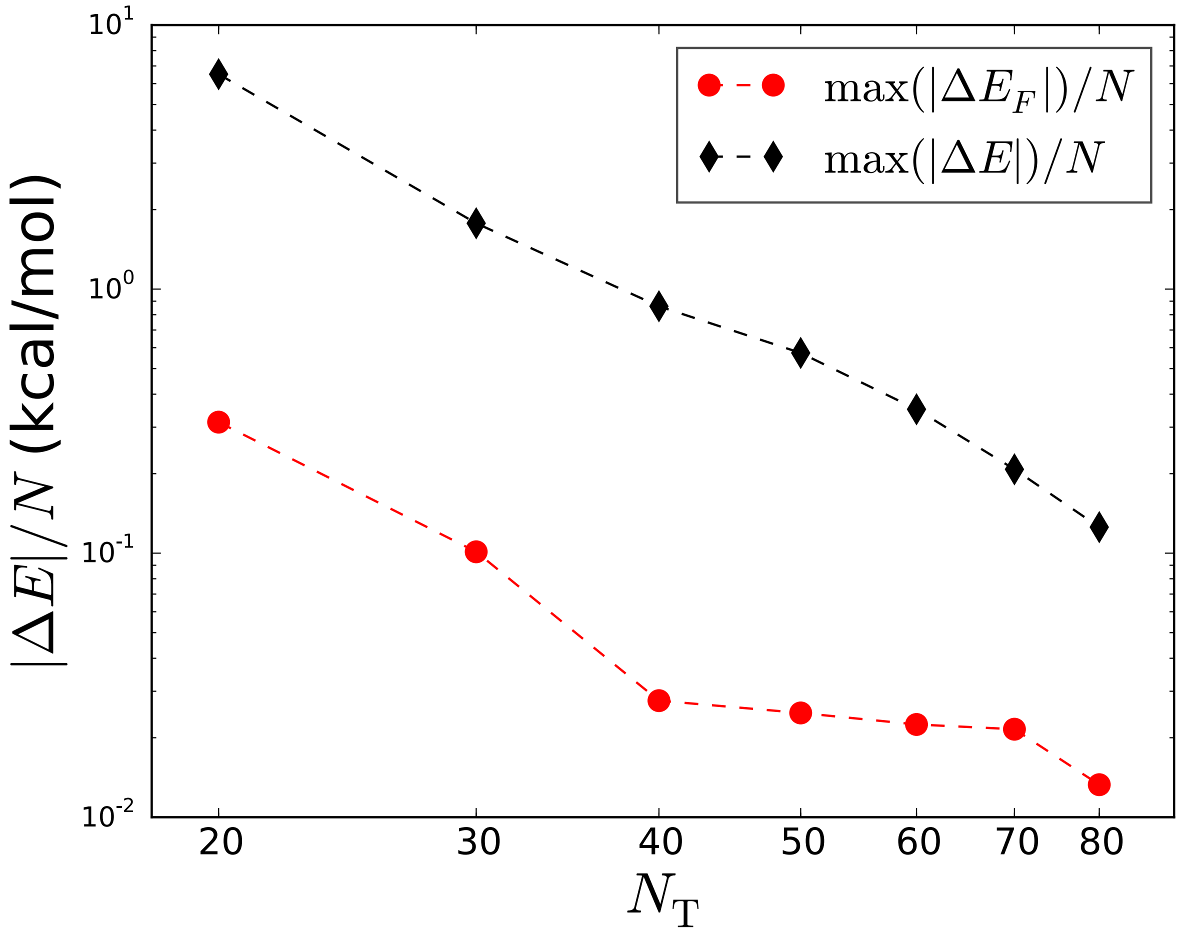
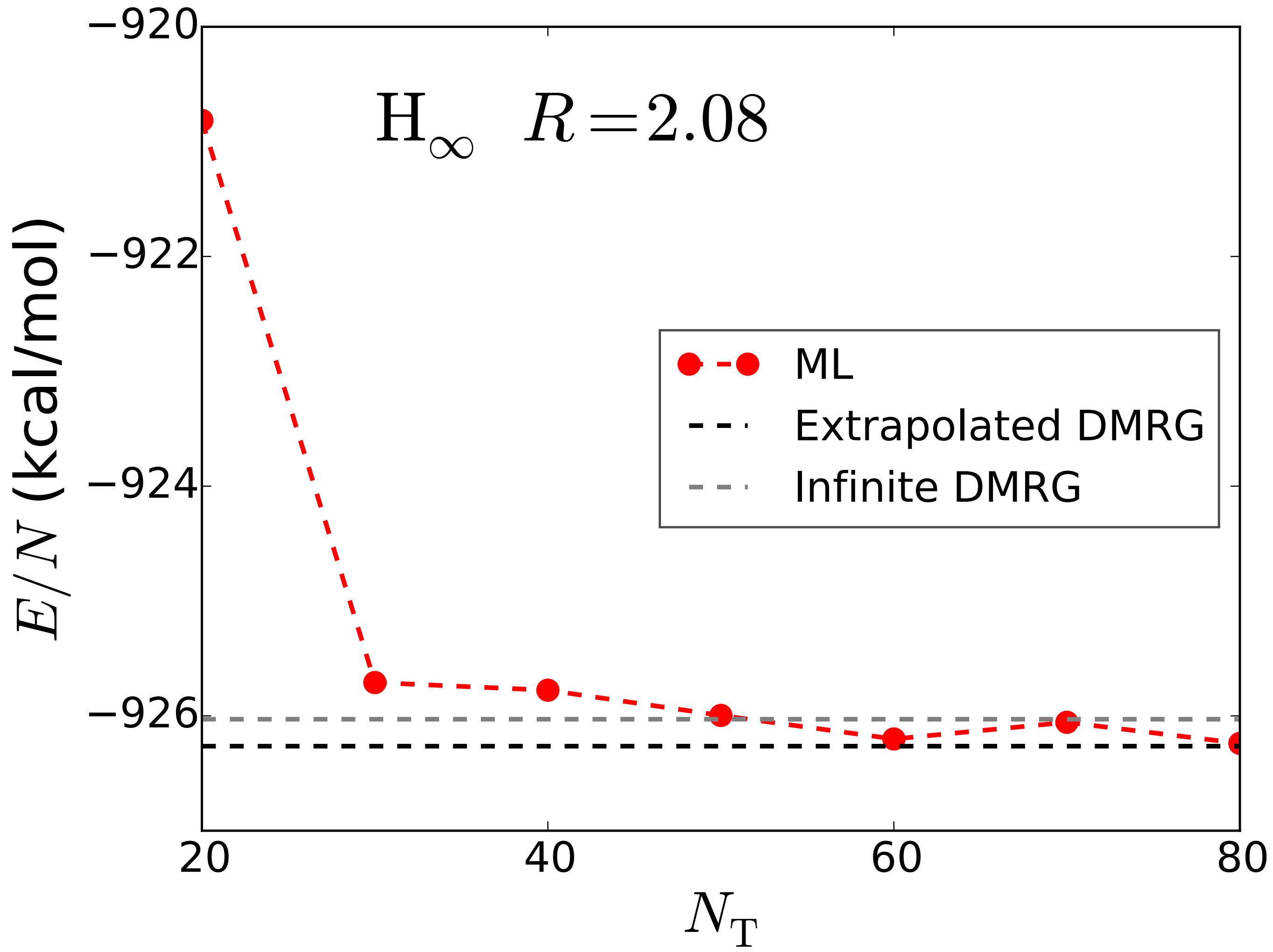


FIG. 8. (Color online) Learning curves for several 1d H chains. (a) ML using the total density. (b) ML using the bulk partition densities (see text).

Origin of error for chain



Convergence for infinite chain



Pure density functional for strong correlations and the thermodynamic limit from machine learning Li Li, Thomas E. Baker, Steven R. White, Kieron Burke, *Phys. Rev. B* **94**, 245129 (2016).

Lessons from this part

- Can learn exact functional from exact data.
- Can learn F[n] instead of $T_s[n]$ so accurately you can even get density.
- Created a new data-driven basis by using atoms in molecules; greatly reduced computational cost.
- Extrapolate to infinite chain limit to within 1 kcal/mol.
- No problem in principle to do in 3d.

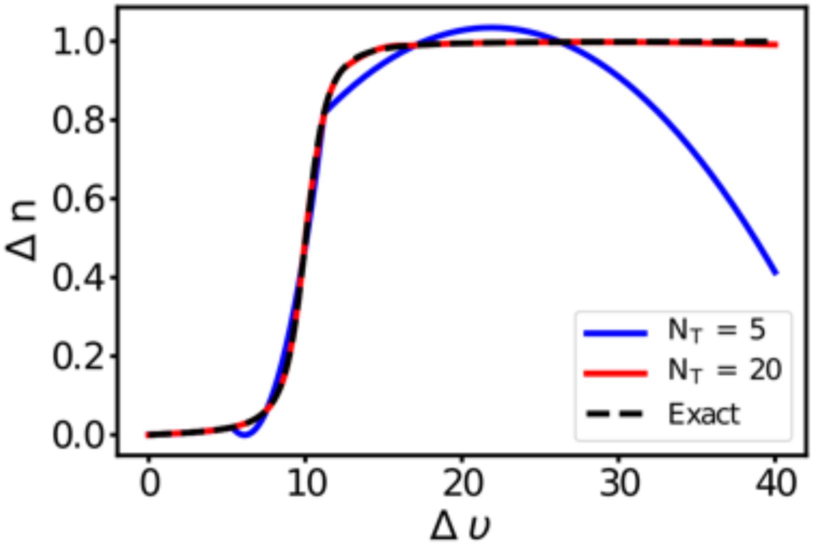
ML Density Functionals via Legendre Transformation

The Idea

- Levy-Lieb constrained search formalism defines the universal part of the functional via LT of the total energy functional[1],

$$F[n] = \sup_{v(\boldsymbol{r})}\left\{E[v(\boldsymbol{r}); N] - \int n(\boldsymbol{r})v(\boldsymbol{r})d^3r\right\}$$

- This crude ML algorithm incorporates the basic theorem of DFT within it and automatically satisfies certain exact condition, $F^{ML}[n] \leq F[n]$.
- Model is trained on $v(\boldsymbol{r})$ and $E[v(\boldsymbol{r})]$ data to determine $F[n]$ and self-consistent $n(\boldsymbol{r})$.



In collaboration with Stefan Mandt, UCI comp sci; Bhupalee Kalita

---

1. E. H. Lieb, *Int. J. Quantum Chem.* **24**, 243 (1983).

Can exact conditions improve ML functionals?

New paper (dft.uci.edu)

- Can we use exact conditions to improve ML density functional approximations?
- Given a set of densities to train on, impose quadratic scaling of kinetic energy.

- Train on scaled densities, then rescale. Do we do any better?
- Original idea of John Snyder.

[185] **Can exact conditions improve machine-learned density functionals?** Jacob Hollingsworth, Li Li, Thomas E. Baker, Kieron Burke, *to appear in J Chem Phys, Special Topic*(2018).

Imposing an exact condition

We attempt to impose $T_s[n_\gamma] = \gamma^2\, T_s[n]$ on a machine learned functional. We define:

$$l^2[n] = \int_{-\infty}^{\infty} dx\, x^2\, n(x).$$

and define the scaled density:

$$\tilde{n}(x) = n_l(x) = l\, n(lx).$$

We write the kinetic energy functional

$$T_s[n] = \frac{1}{l^2[n]}\, \tilde{T}_s[\tilde{n}].$$

We machine learn $\tilde{T}_s[\tilde{n}]$, as opposed to directly machine learning $T_s[n]$.

Learning curve on 1D Hooke's atom

$$H = -\frac{1}{2}\left(\frac{d^2}{dx_1^2} + \frac{d^2}{dx_2^2}\right) + \frac{1}{2}\omega^2(x_1^2 + x_2^2) + \delta\,(x_1 - x_2)\,.$$

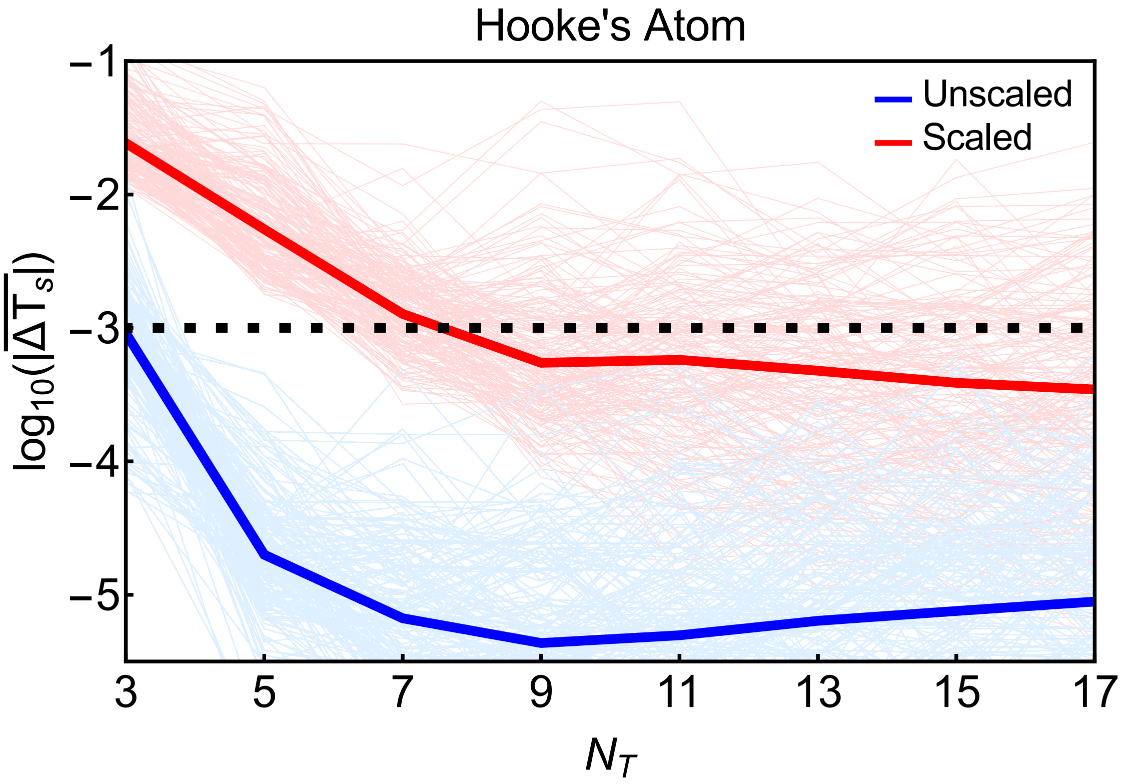


FIG. 2. (color online) The learning curves for functionals trained on scaled (blue) and unscaled (red) densities for the 1D Hooke's atom. Accuracy of 1 mH is denoted by the dashed line (black).

Learning curve on 1D H_2

$$V_{ext}(x) = -Ae^{-\kappa|x-\frac{R}{2}|} - Ae^{-\kappa|x+\frac{R}{2}|}$$

$$V_{int}(x,x') = Ae^{-\kappa|x-x'|}$$

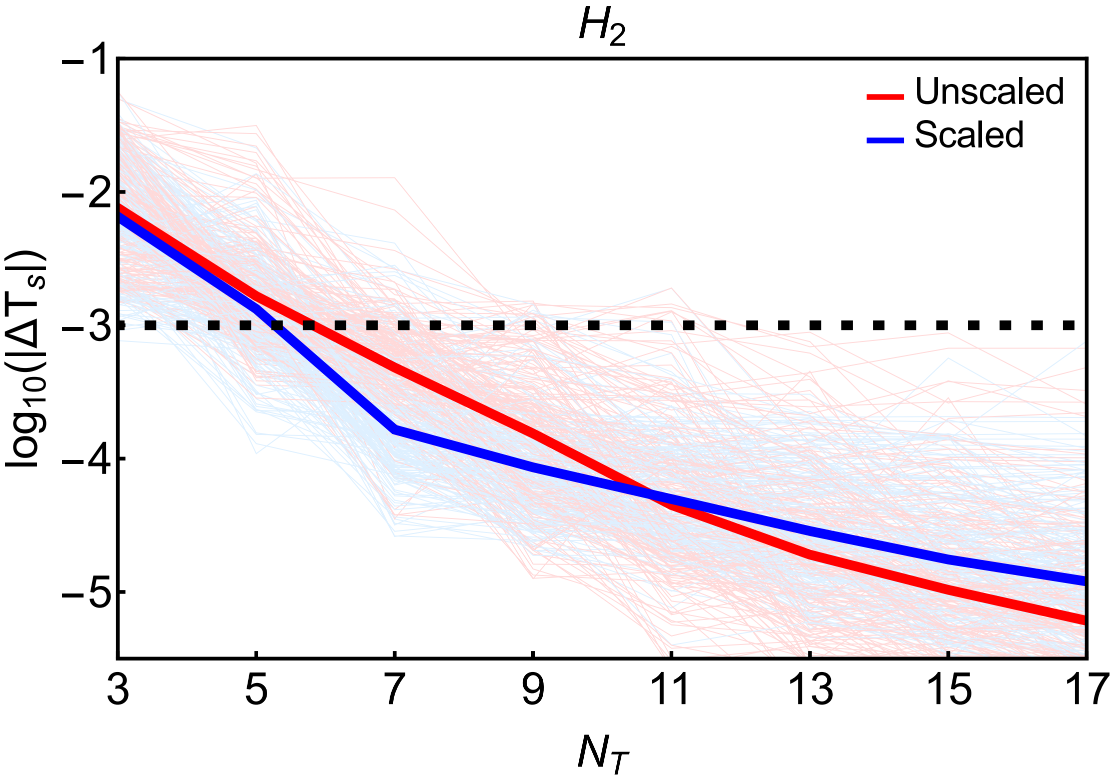
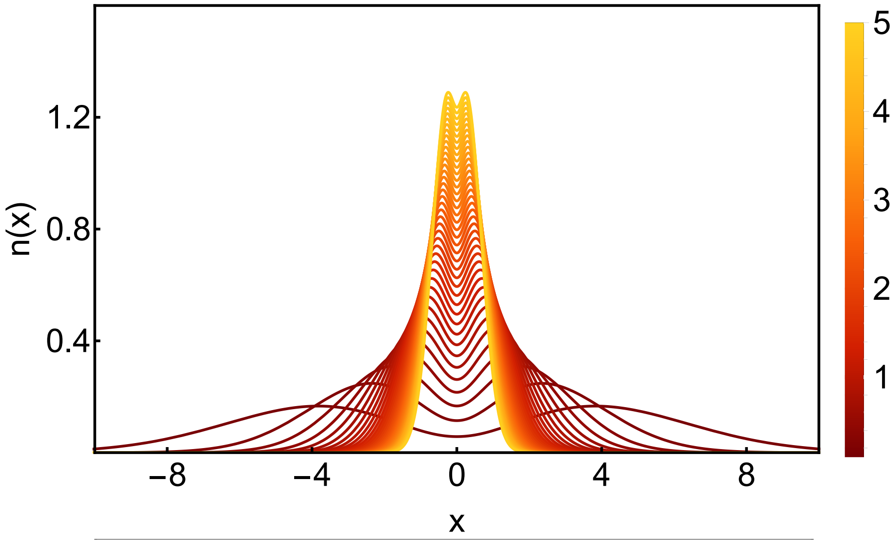
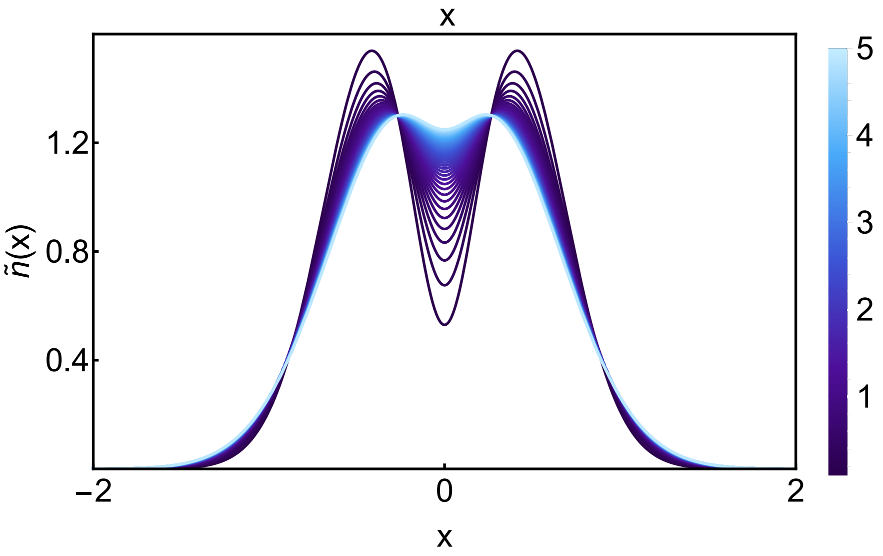


FIG. 3. (color online) Same as Fig. 2, but for H_2 densities.

Densities in 1D Hooke’s atom





Densities in 1D H_2

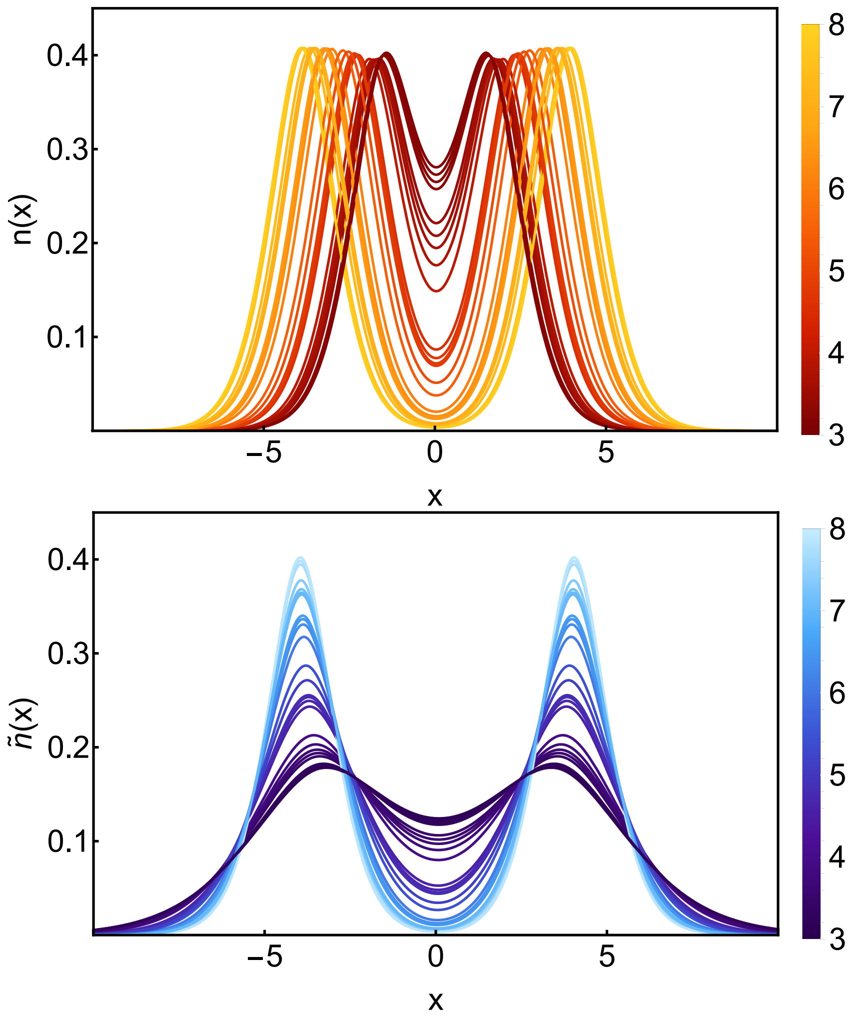
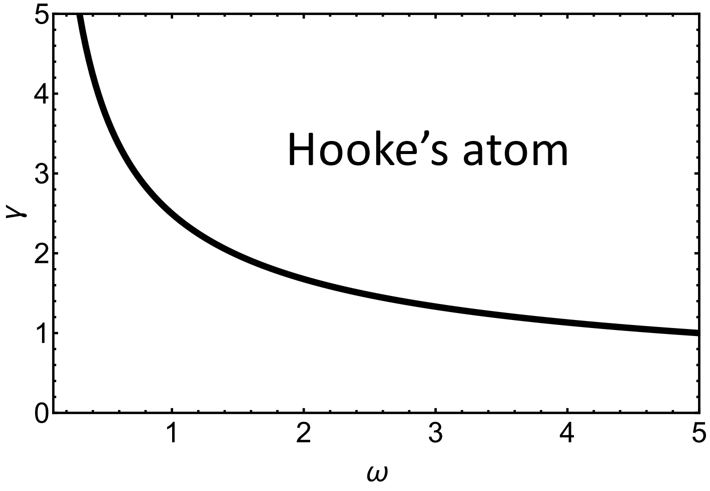
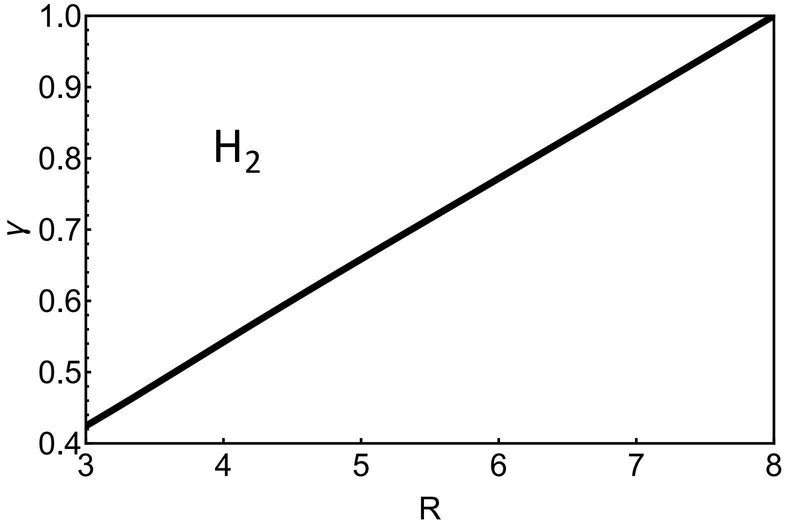


FIG. 7. (color online) The H_2 densities considered are plotted before scaling (top) and after scaling (bottom). The color each line indicates the separation, R. The densities do not appear to be made more similar by scaling.

Origin of difference





$$l[n] = \sqrt{2 + \frac{R^2}{2}}.$$

Conclusions on scaling

- Imposition of exact coordinate scaling does typically improve ML functionals
- But if scaling of densities makes them differ more among themselves, effect may be negligible.
- Speculation: Even in H_2 case, learning would be improved if we wanted more than one density at each R
- So depends on application.

Higher accuracy for weak correlation

ChemRxiv™

doi.org/10.26434/chemrxiv.8079917.v1

Density Functionals with Quantum Chemical Accuracy: From Machine Learning to Molecular Dynamics

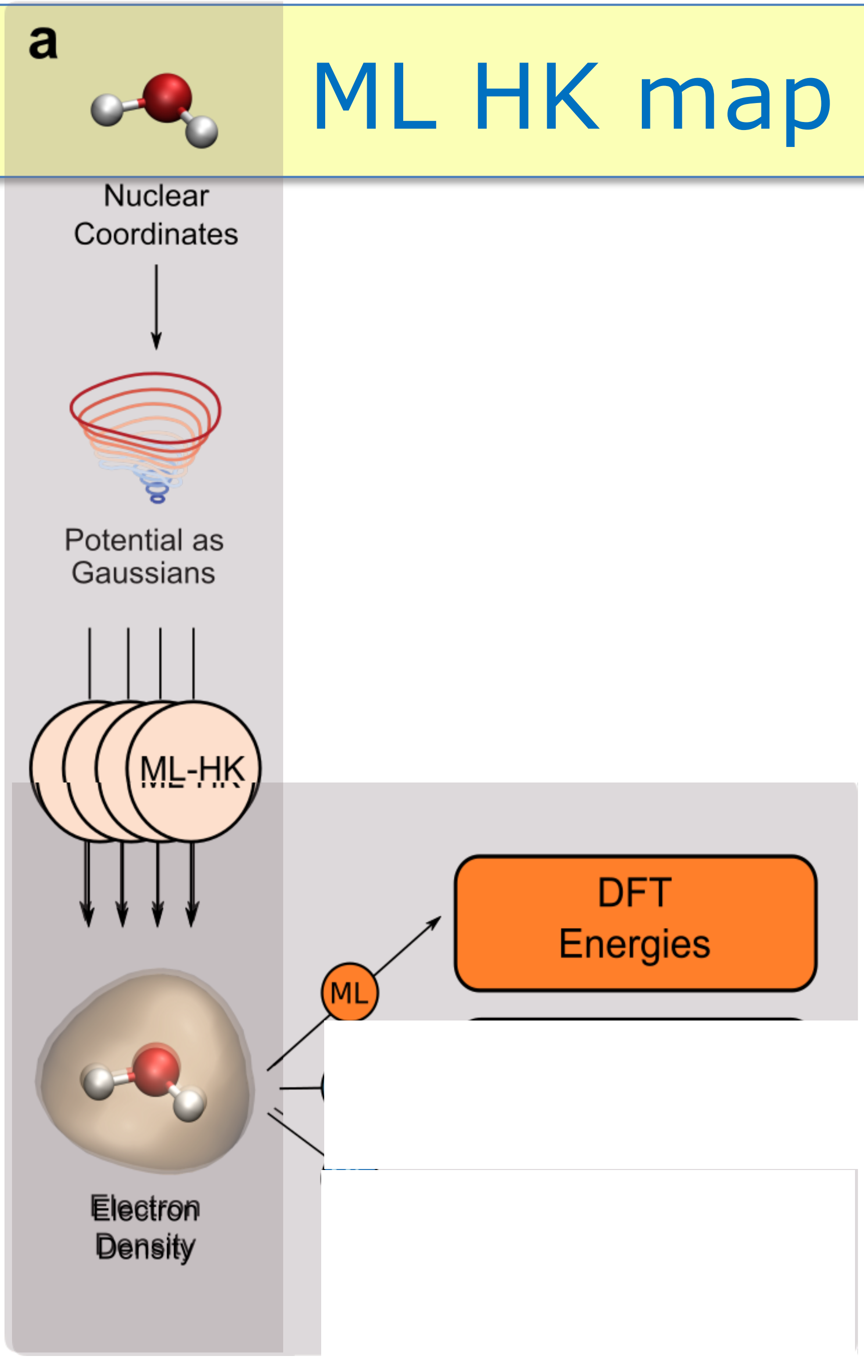
Mihail Bogojeski, Leslie Vogt-Maranto, Mark E. Tuckerman, Klaus-Robert Mueller, Kieron Burke

Submitted date: 03/05/2019 • Posted date: 06/05/2019

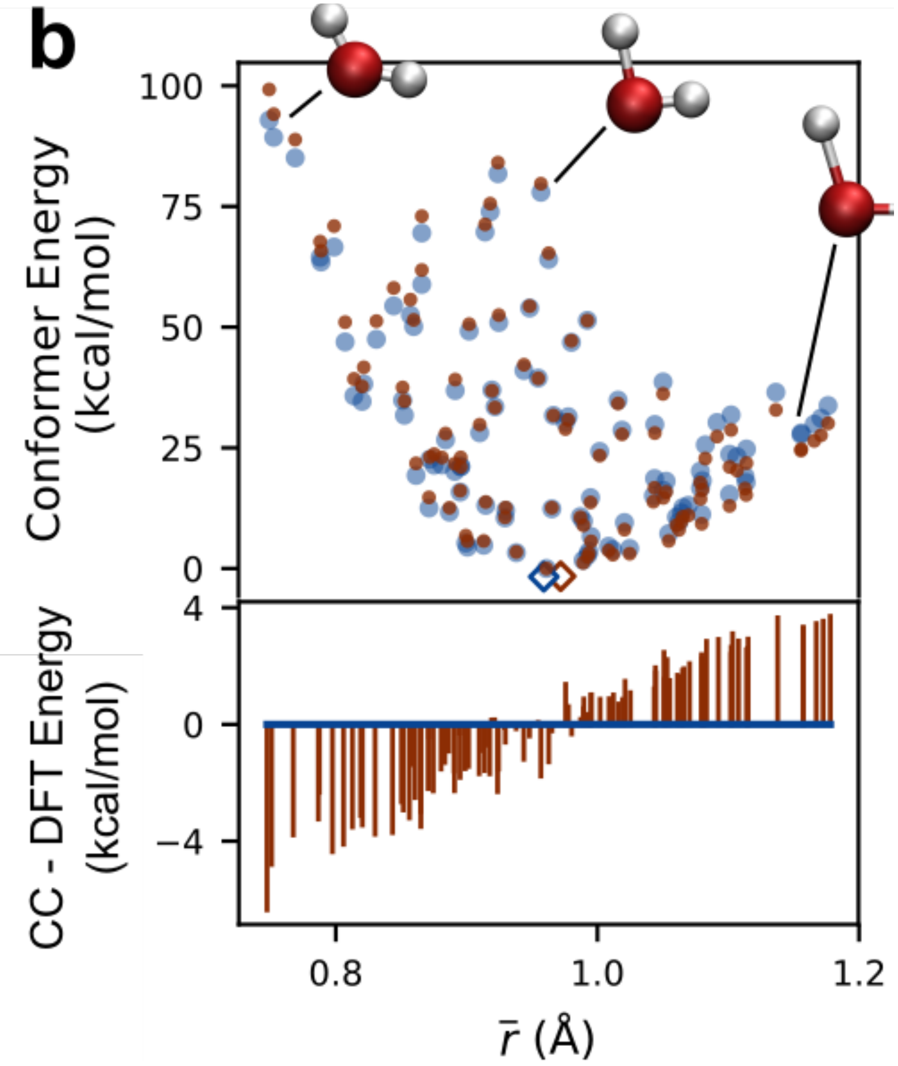
Licence: CC BY-NC-ND 4.0

Citation information: Bogojeski, Mihail; Vogt-Maranto, Leslie; Tuckerman, Mark E.; Mueller, Klaus-Robert; Burke, Kieron (2019): Density Functionals with Quantum Chemical Accuracy: From Machine Learning to Molecular Dynamics. ChemRxiv. Preprint.

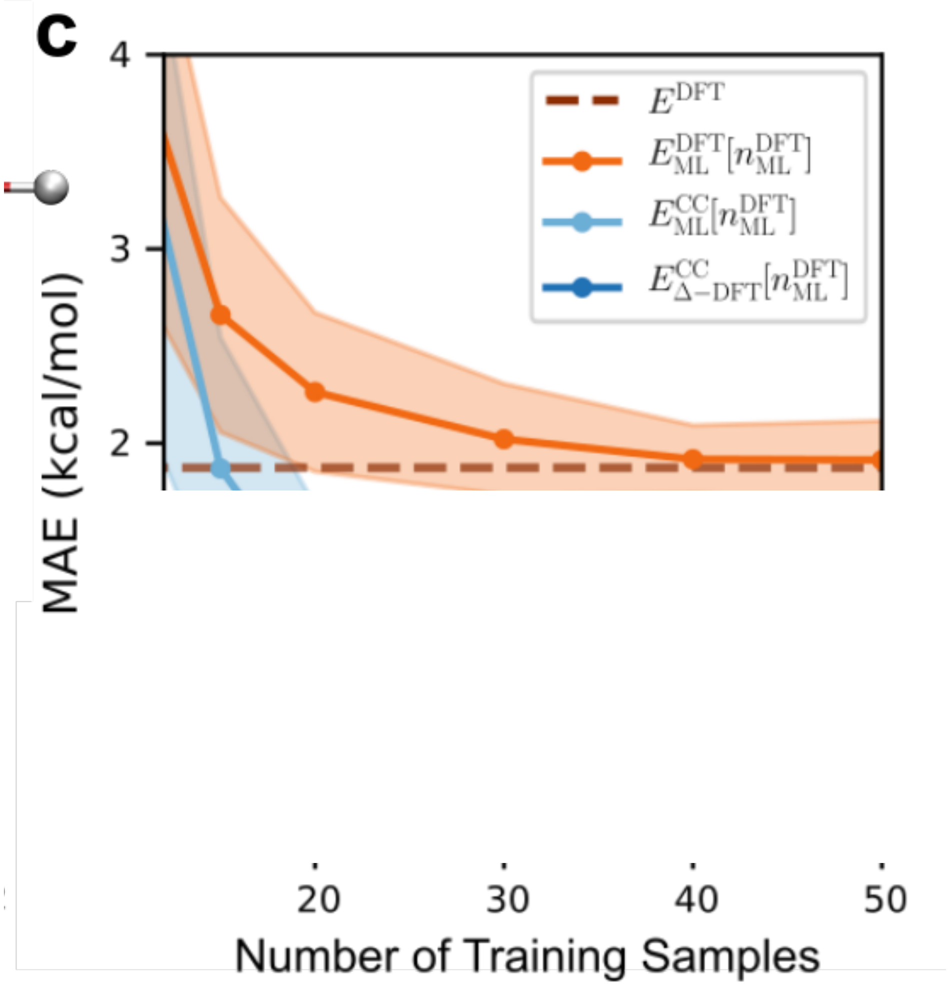
ML HK map


a
Nuclear Coordinates
Potential as Gaussians
ML-HK
Electron Density
ML
DFT Energies

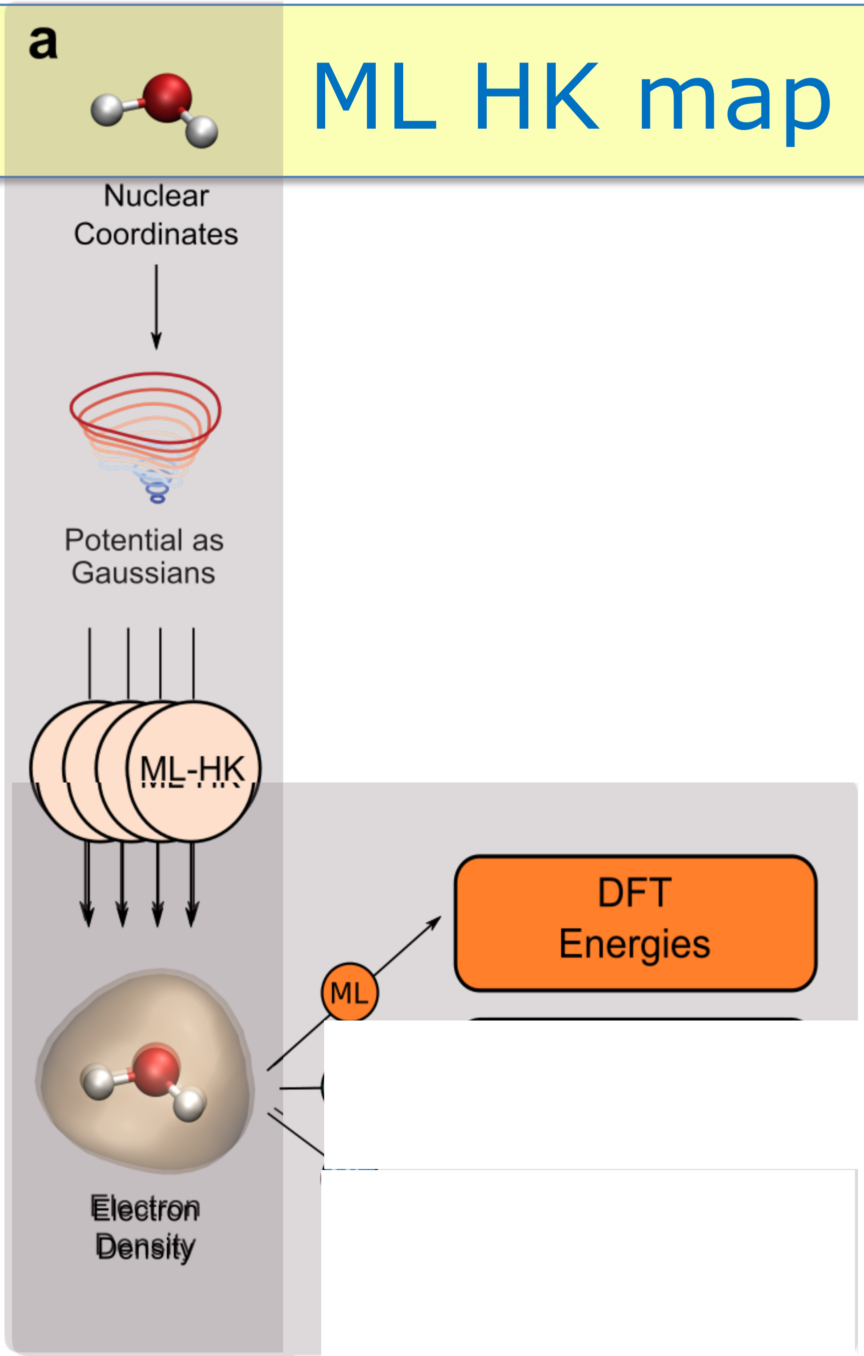
Range of PBE errors



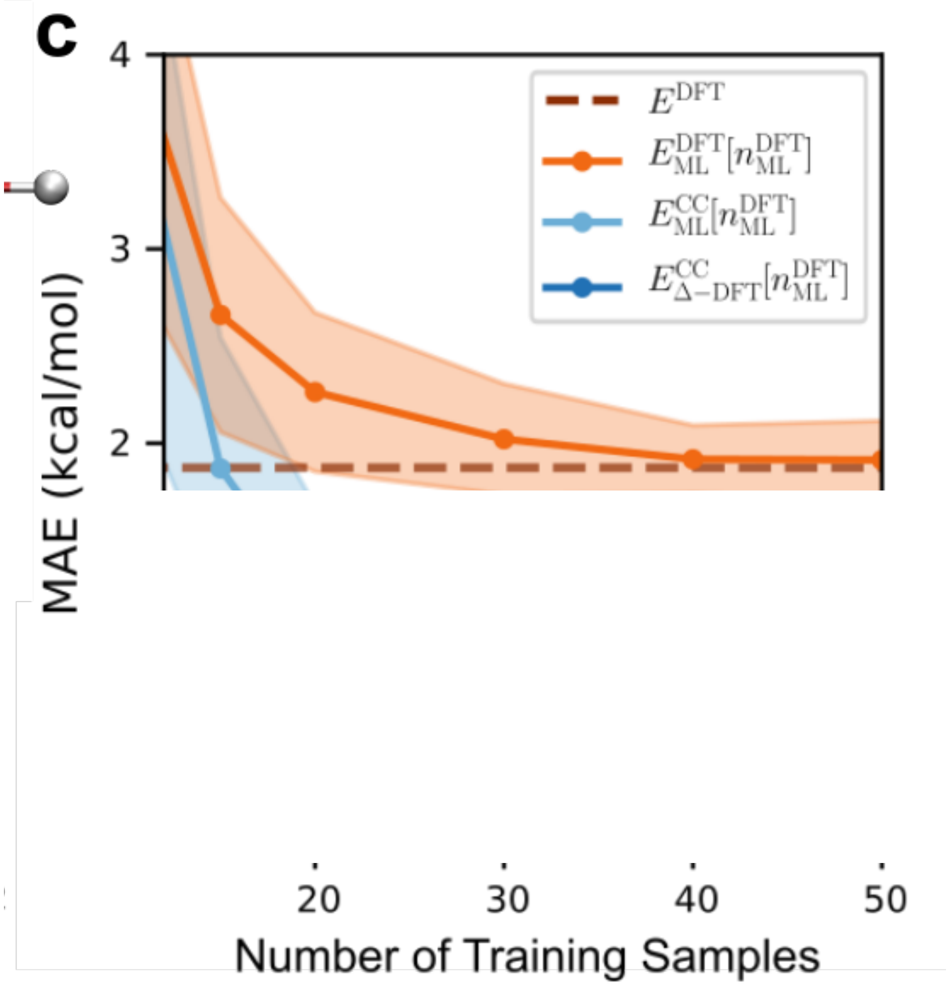
Learning curves



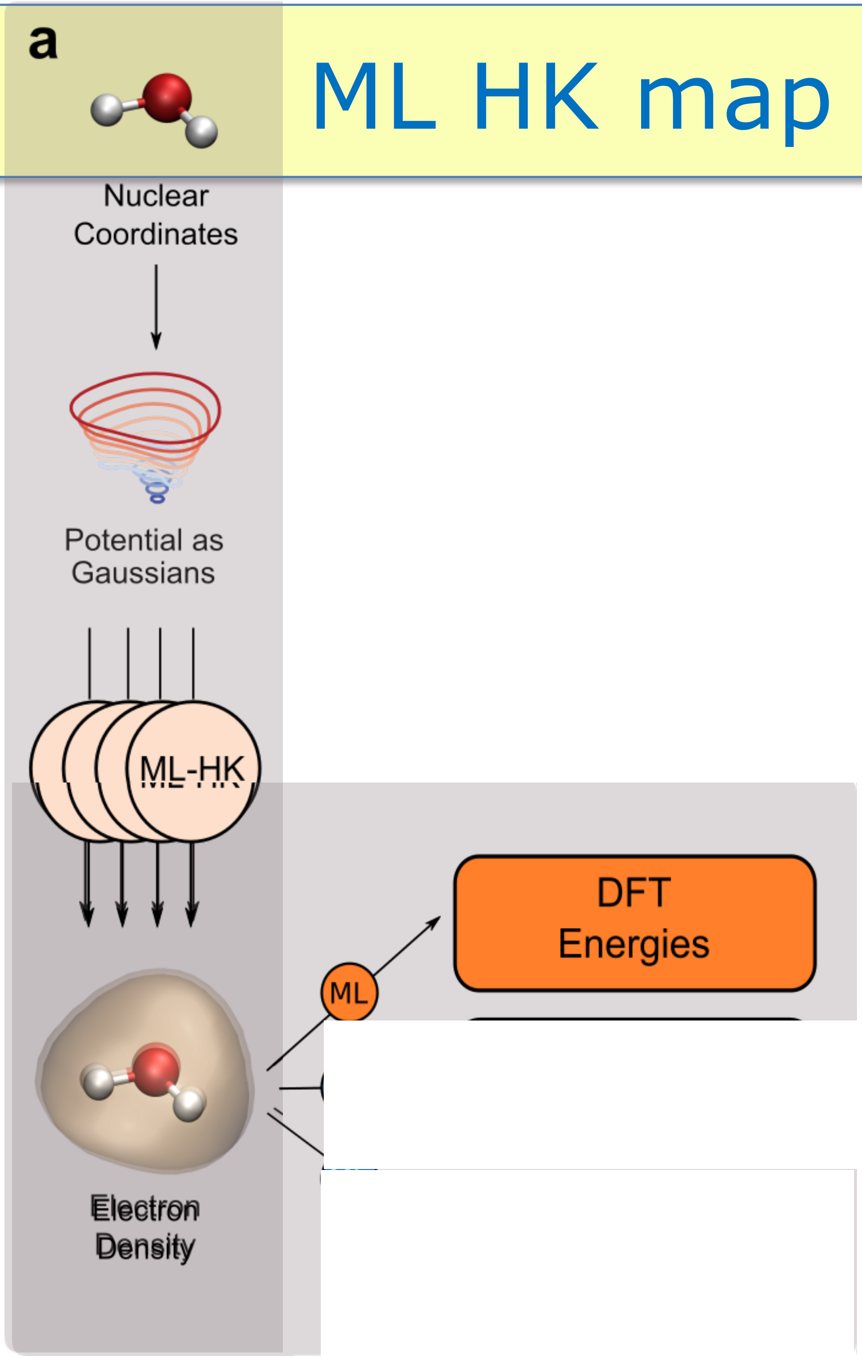
ML HK map


a
Nuclear Coordinates
Potential as Gaussians
ML-HK
Electron Density
ML
DFT Energies

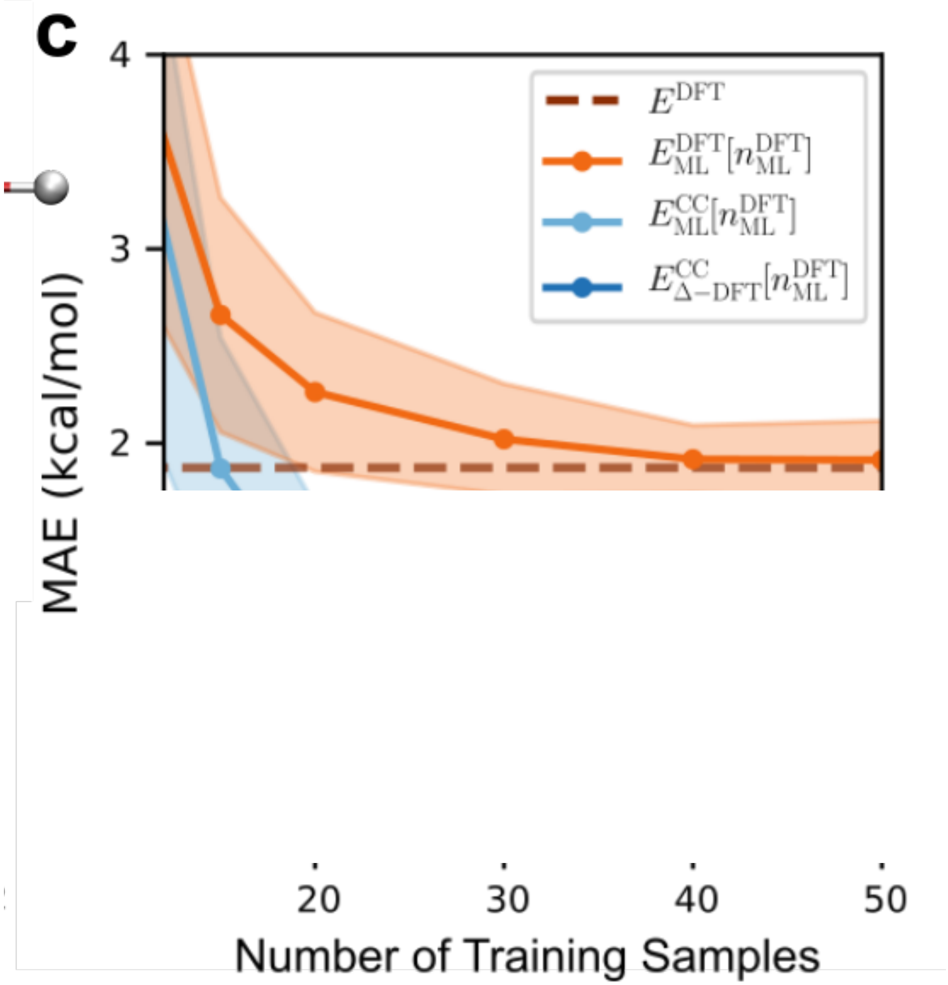
Learning curves



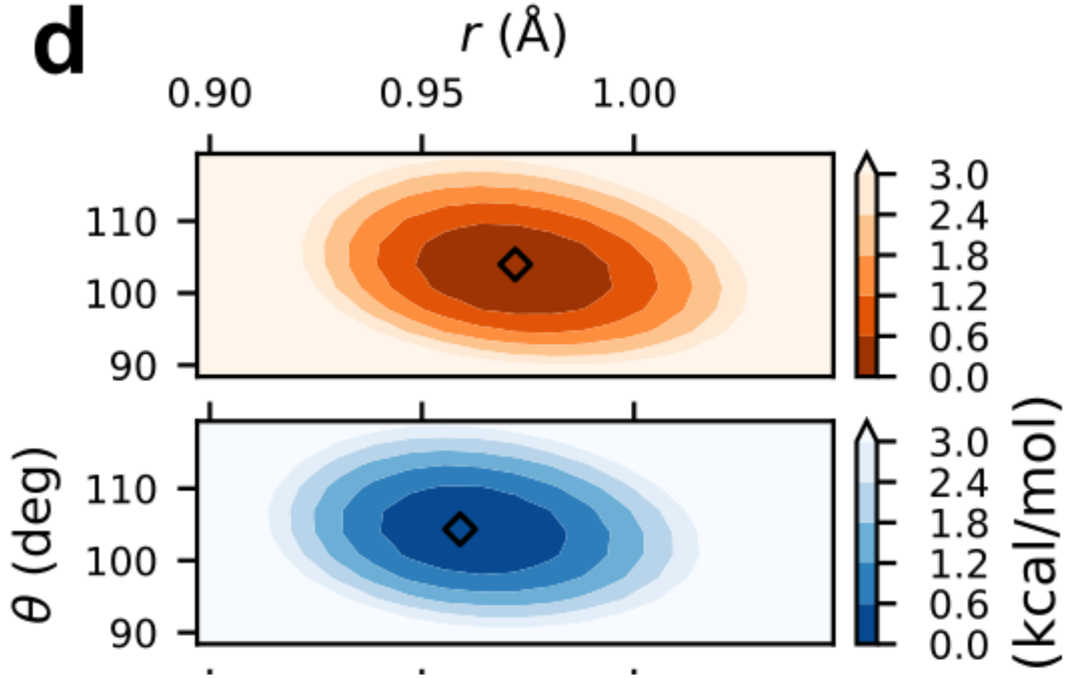
ML HK map


a
Nuclear Coordinates
Potential as Gaussians
ML-HK
Electron Density
ML
DFT Energies

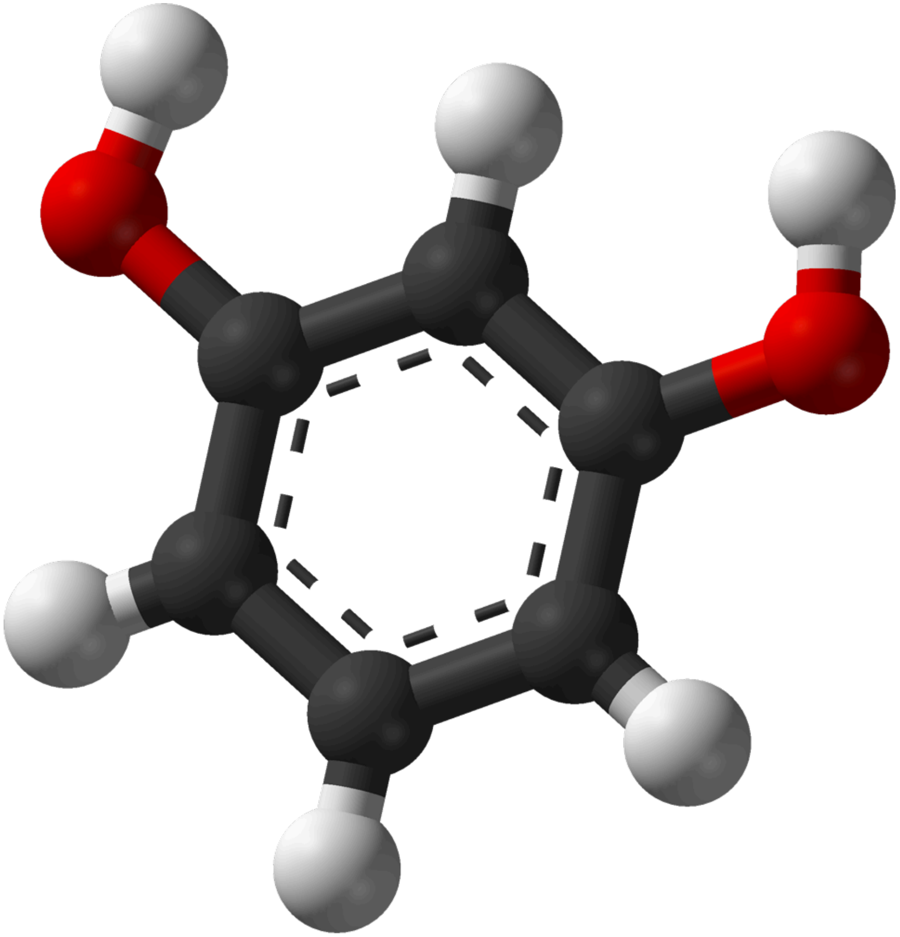
Learning curves



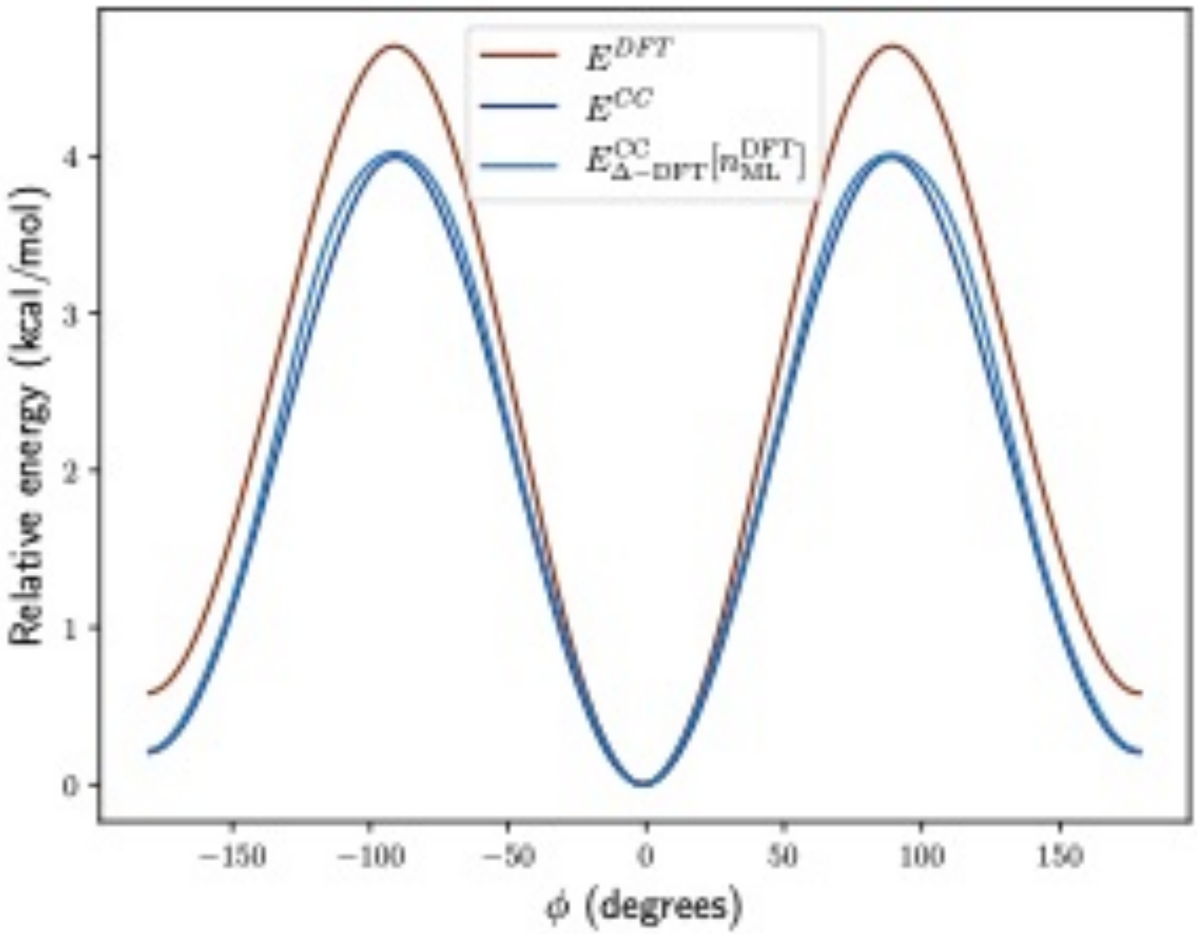
Potential energy surfaces



Resorcinol



Resorcinol barriers



Supplemental Figure 12: Relative energies for DFT and CC showing the difference in the OH rotational barrier between the two methods, along with the energy predicted by the $E^{CC}_{s\Delta\text{-DFT}}[n^{DFT}_{sML}]$ energy map.

Different molecular dynamics

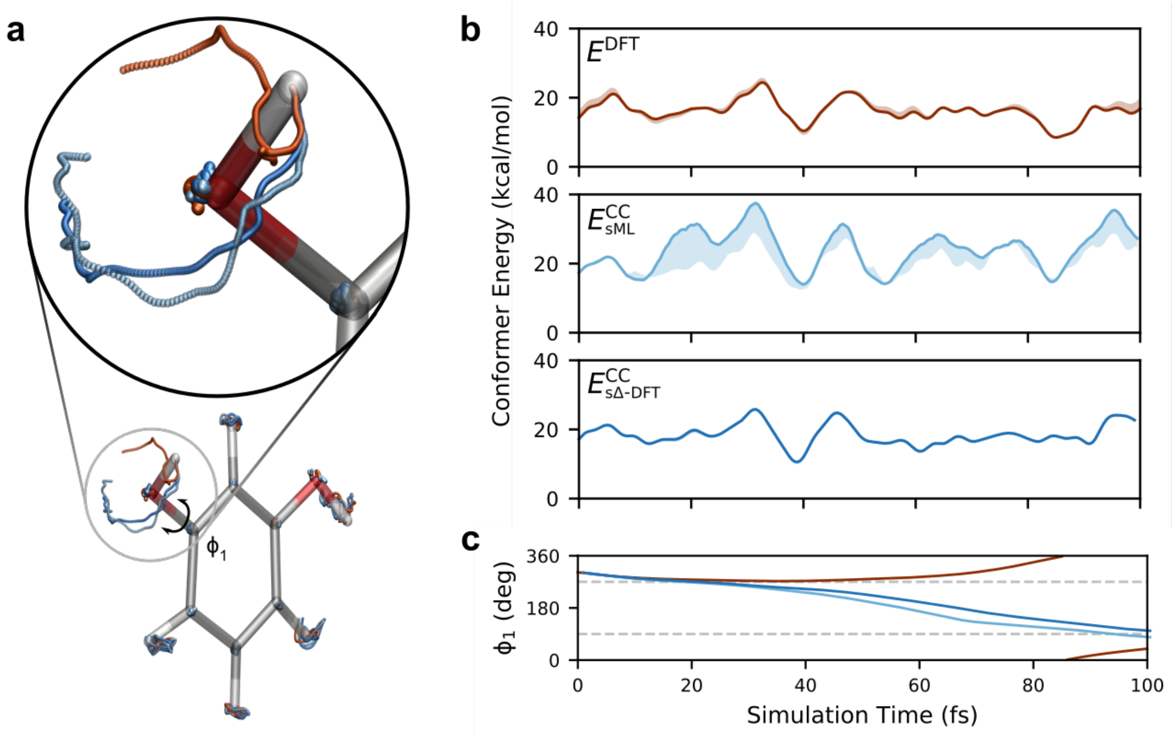


Figure 3: Resorcinol dynamics from an initial condition near a conformational change showing a) the atomic positions explored during 100 fs NVE MD trajectories run with standard DFT (dark orange), $E^{\mathrm{CC}}_{\mathrm{sML}}[n^{\mathrm{DFT}}_{\mathrm{sML}}]$ with RESPA-corrected forces (light blue), and $E^{\mathrm{CC}}_{\mathrm{s\Delta\text{-}DFT}}[n^{\mathrm{DFT}}_{\mathrm{sML}}]$ (blue), b) the conformer energy along each trajectory (solid lines), with the error relative to CC shown as a shaded line width, and c) the evolution of the C-C-O-H dihedral angle for each trajectory with dashed grey lines indicating the barrier between conformers. For this figure, all DFT calculations use PBE and all CC energies are from CCSD(T).

Coming soon

- Collobration with U. Tokyo (physics)
  - Grad student Ryo Nagai
  - Train approximate functionals on accurate densities and energies of 3 molecules
  - Test on 150 molecular atomization energies

C. Observations on ML in physical science

- Experimental data versus computer-generated data.
- Image processing or not?
- Reproducibility?
- Hamiltonians?
- Slow convergence of statistics
- Too many possibilities

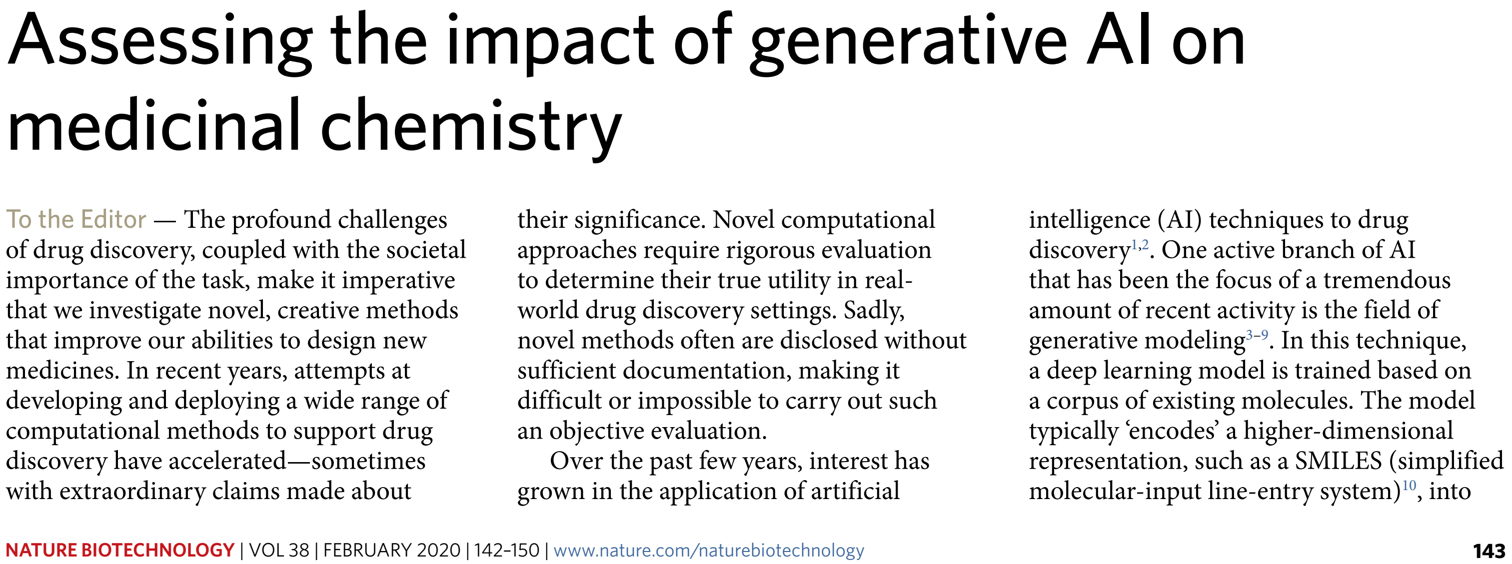
Paradigm shift?

- Term first used by Thomas Kuhn to describe advent of quantum mechanics
- Due to computing resources and the automated gathering of data, machine learning has become cost-effective.
- Will become a whole new approach to analyzing science
- Keep in mind: Planck’s constant (1900), Schroedinger equation (1926).

Bad practices

- More than 50% of papers do not meet minimal standards of publishability
- Either conscious or unconscious ignoring of existing literature
- Claims of significance because “first use of AI..”
- Cherry picking of datasets, metrics, etc.
- Generally, sloppy science

Critique of recent work



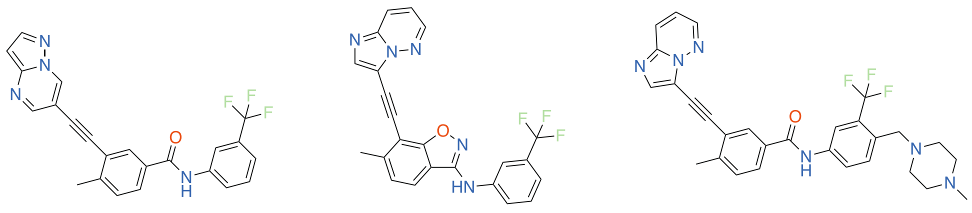
Assessing the impact of generative AI on medicinal chemistry

To the Editor — The profound challenges of drug discovery, coupled with the societal importance of the task, make it imperative that we investigate novel, creative methods that improve our abilities to design new medicines. In recent years, attempts at developing and deploying a wide range of computational methods to support drug discovery have accelerated—sometimes with extraordinary claims made about their significance. Novel computational approaches require rigorous evaluation to determine their true utility in real-world drug discovery settings. Sadly, novel methods often are disclosed without sufficient documentation, making it difficult or impossible to carry out such an objective evaluation.

Over the past few years, interest has grown in the application of artificial intelligence (AI) techniques to drug discovery[1,2]. One active branch of AI that has been the focus of a tremendous amount of recent activity is the field of generative modeling[3–9]. In this technique, a deep learning model is trained based on a corpus of existing molecules. The model typically 'encodes' a higher-dimensional representation, such as a SMILES (simplified molecular-input line-entry system)[10], into

NATURE BIOTECHNOLOGY | VOL 38 | FEBRUARY 2020 | 142–150 | www.nature.com/naturebiotechnology 143

correspondence



Gao et al.
Compound 7r
6 nM

Zhavoronkov et al.
Compound 1
10 nM

Ponatinib
9 nM

Fig. 1 | Comparison of compound 1 and two related inhibitors. The chemical structure of compound 1 from Zhavoronkov et al. compared with those of ponatinib, a marketed multi-kinase inhibitor, and a DD1 inhibitor reported by Gao et al.[14] in 2013.

Summary

- Machine learning is really popular
- ML is easy to apply to image recognition
- ML hard to apply to Hamiltonian problems
- ML can be used to make density functionals that fail differently from human functionals
- Thanks to
  - Students: Bhupalee Kalita, Tom Baker, Li Li, John Snyder, Kevin Vu, Isabelle Pelaschier, Jacob Hollingsworth
  - Collaborators: Klaus Mueller, Matthias Rupp, Katia Hansen, Felix Brockherde, Leslie Vogt, Mark Tuckerman, Mihail Bogojeski
  - Institute of Pure and Applied Math, UCLA
  - Funders: NSF from chem, DMR, math