Machine-learning of density functionals for applications in molecules and materials

Kieron Burke and friends
UC Irvine
Chemistry & Physics

http://dft.uci.edu

Outline

A. Background in DFT

B. Background in ML-DFT

C. Machine-learned KS kinetic energy of molecules (3D)

D. Machine-learning of XC for strongly correlated solids (1D).

E. Can exact conditions help us learn?

The electronic structure problem

- Use atomic units
- Born-Oppenheimer approximation
- Wavefunctions antisymmetric and normalized
- Only discuss ground-state electronic problem here, but many variations.
- All non-relativistic, non-magnetic here

Hamiltonian for N electrons in the presence of external potential $v(\mathbf{r})$:

$$\hat{H} = \hat{T} + \hat{V}_{\rm ee} + \hat{V},$$

where the kinetic and elec-elec repulsion energies are

$$\hat{T} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2, \qquad \hat{V}_{\rm ee} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|},$$

and difference between systems is N and the one-body potential

$$\hat{V} = \sum_{i=1}^{N} v(\mathbf{r}_i)$$

Often $v(\mathbf{r})$ is electron-nucleus attraction

$$v(\mathbf{r}) = -\sum_{\alpha}\frac{Z_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|}$$

where α runs over all nuclei, plus weak applied **E** and **B** fields.

$$\{\hat{T} + \hat{V}_{\rm ee} + \hat{V}\}\,\Psi = E\,\Psi, \qquad E = \min_{\Psi}\langle\Psi|\hat{T} + \hat{V}_{\rm ee} + \hat{V}|\Psi\rangle$$

Constrained search proof

1. Rewrite variational principle (Levy 79):

$$\begin{aligned} E &= \min_{\Psi} \langle \Psi | \hat{T} + \hat{V}_{\rm ee} + \hat{V} | \Psi \rangle \\ &= \min_{n} \left\{ F[n] + \int d^3r \; v(\mathbf{r}) n(\mathbf{r}) \right\} \end{aligned}$$

   where

$$F[n] = \min_{\Psi \to n} \langle \Psi | \hat{T} + \hat{V}_{\rm ee} | \Psi \rangle$$

   - The minimum is taken over all positive $n(\mathbf{r})$ such that $\int d^3r \; n(\mathbf{r}) = N$

2. The external potential $v(\mathbf{r})$ and the hamiltonian $\hat{H}$ are determined to within an additive constant by $n(\mathbf{r})$

- P. Hohenberg and W. Kohn, Phys. Rev. **136**, B 864 (1964).
- M. Levy, Proc. Natl. Acad. Sci. (U.S.A.) **76**, 6062 (1979).

KS equations (1965)

Define *fictitious* non-interacting electrons satisfying:

$$\left\{-\frac{1}{2}\nabla^2 + v_{\rm S}(\mathbf{r})\right\}\phi_j(\mathbf{r}) = \epsilon_j\phi_j(\mathbf{r}), \qquad \sum_{j=1}^{N}|\phi_j(\mathbf{r})|^2 = n(\mathbf{r}).$$

where $v_{\rm S}(\mathbf{r})$ is *defined* to yield $n(\mathbf{r})$.

Define $T_{\rm S}$ as the kinetic energy of the KS electrons, U as their Hartree energy and

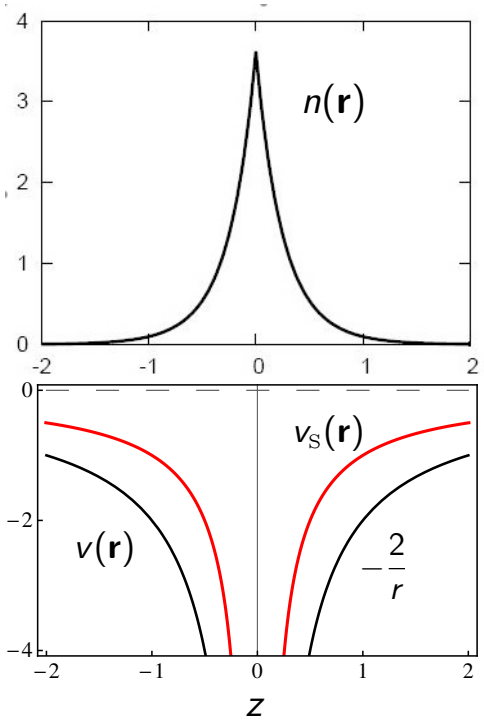
$$F = T + V_{\rm ee} = T_{\rm S} + U + E_{\rm XC}$$

the remainder is the exchange-correlation energy.

Most important result of exact DFT:

$$v_{\rm S}(\mathbf{r}) = v(\mathbf{r}) + \int d^3r \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + v_{\rm XC}[n](\mathbf{r}), \qquad v_{\rm XC}(\mathbf{r}) = \frac{\delta E_{\rm XC}}{\delta n(\mathbf{r})}$$

Knowing $E_{\rm XC}[n]$ gives closed set of self-consistent equations.



Today's commonly-used functionals

- Local density approximation (LDA) $E_{\rm X}^{\rm LDA}[n] = A_{\rm X} \int d^3r \; n^{4/3}(\mathbf{r})$
  - Uses only n(**r**) at a point. $A_{\rm X} = -(3/4)(3/\pi)^{1/3} = -0.738.$
- Generalized gradient approx (GGA)
  - Uses both n(**r**) and $|\nabla n(\mathbf{r})|$
  - Should be more accurate, corrects overbinding of LDA
  - Examples are PBE and BLYP
- Hybrid:
  - Mixes some fraction of HF
  - Examples are B3LYP and PBE0

A few recent applications

- Computers, codes, algorithms always improving
- Making bona fide predictions
- E.g., a new better catalyst for Haber-Bosch process ('fixing' ammonia from air) was predicted after about 25,000 failed experiments (Jens Norskov's group)
- Now scanning chemical and materials spaces using big data methods for materials design (materials genome project).
- World's hottest superconductor (203K) is hydrogen sulfide, predicted by DFT calculations, then made.

Breadth of applications

Computational exfoliation

nature nanotechnology

ARTICLES

https://doi.org/10.1038/s41565-017-0035-5

Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds

Nicolas Mounet[1*], Marco Gibertini[1], Philippe Schwaller[1], Davide Campi[1], Andrius Merkys[1,2], Antimo Marrazzo[1], Thibault Sohier[1], Ivano Eligio Castelli[1], Andrea Cepellotti[1], Giovanni Pizzi[1] and Nicola Marzari[1*]

Two-dimensional (2D) materials have emerged as promising candidates for next-generation electronic and optoelectronic applications. Yet, only a few dozen 2D materials have been successfully synthesized or exfoliated. Here, we search for 2D materials that can be easily exfoliated from their parent compounds. Starting from 108,423 unique, experimentally known 3D compounds, we identify a subset of 5,619 compounds that appear layered according to robust geometric and bonding criteria. High-throughput calculations using van der Waals density functional theory, validated against experimental structural data and calculated random phase approximation binding energies, further allowed the identification of 1,825 compounds that are either easily or potentially exfoliable. In particular, the subset of 1,036 easily exfoliable cases provides novel structural prototypes and simple ternary compounds as well as a large portfolio of materials to search from for optimal properties. For a subset of 258 compounds, we explore vibrational, electronic, magnetic and topological properties, identifying 56 ferromagnetic and antiferromagnetic systems, including half-metals and half-semiconductors.

Prediction of room-temperature spintronics

Prediction of a room-temperature and switchable Kane-Mele quantum spin Hall insulator

Antimo Marrazzo,[1,*] Marco Gibertini,[1] Davide Campi,[1] Nicolas Mounet,[1] and Nicola Marzari[1,†]

[1] *Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, 1015, Switzerland*

(Dated: December 12, 2017)

Fundamental research and technological applications of topological insulators are hindered by the rarity of materials exhibiting a robust topologically non-trivial phase, especially in two dimensions. Here, by means of extensive first-principles calculations, we propose a novel quantum spin Hall insulator with a sizeable band gap of ∼0.5 eV that is a monolayer of Jacutingaite, a naturally occurring layered mineral first discovered in 2008 in Brazil and recently synthesised. This system realises the paradigmatic Kane-Mele model for quantum spin Hall insulators in a potentially exfoliable two-dimensional monolayer, with helical edge states that are robust even beyond room temperature and that can be manipulated exploiting a unique strong interplay between spin-orbit coupling, crystal-symmetry breaking and dielectric response.

Italian opera?

Merkys *et al. J Cheminform (2017) 9:56*
DOI 10.1186/s13321-017-0242-y

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

A posteriori metadata from automated provenance tracking: integration of AiiDA and TCOD

Andrius Merkys[1,2*], Nicolas Mounet[1], Andrea Cepellotti[1], Nicola Marzari[1], Saulius Gražulis[2,3] and Giovanni Pizzi[1]

Abstract

In order to make results of computational scientific research findable, accessible, interoperable and re-usable, it is necessary to decorate them with standardised metadata. However, there are a number of technical and practical challenges that make this process difficult to achieve in practice. Here the implementation of a protocol is presented to tag crystal structures with their computed properties, without the need of human intervention to curate the data. This protocol leverages the capabilities of AiiDA, an open-source platform to manage and automate scientific computational workflows, and the TCOD, an open-access database storing computed materials properties using a well-defined and exhaustive ontology. Based on these, the complete procedure to deposit computed data in the TCOD database is automated. All relevant metadata are extracted from the full provenance information that AiiDA tracks and stores automatically while managing the calculations. Such a protocol also enables reproducibility of scientific data in the field of computational materials science. As a proof of concept, the AiiDA–TCOD interface is used to deposit 170 theoretical structures together with their computed properties and their full provenance graphs, consisting in over 4600 AiiDA nodes.

Keywords: DFT, Reproducibility, Provenance, Open data, Ontology, Materials science

Electrochemistry

PHYSICAL REVIEW MATERIALS **1**, 025402 (2017)

Ionic correlations and failure of Nernst-Einstein relation in solid-state electrolytes

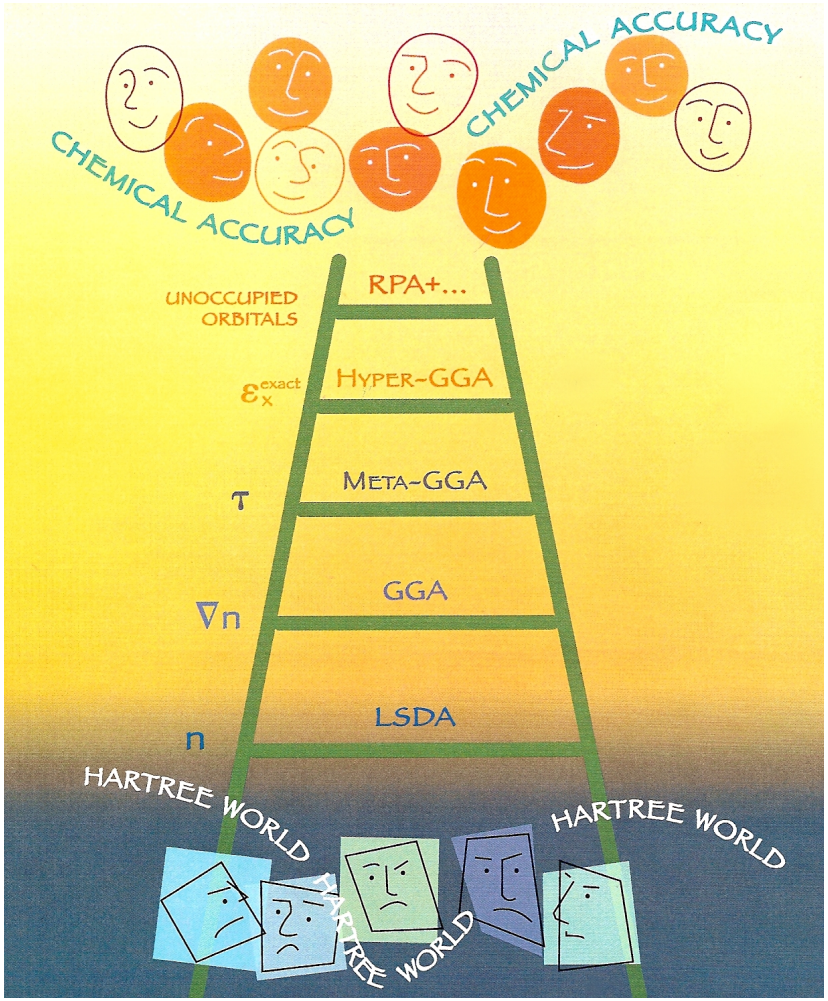
Aris Marcolongo and Nicola Marzari
Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
(Received 16 January 2017; published 5 July 2017)

A microscopic understanding of fast ionic transport is fundamental to design novel solid-state electrolytes. We address the role of correlations in these systems and study in detail the tracer and charge diffusion coefficients, deriving a novel inequality between these two quantities. We investigate the failure of the Nernst-Einstein and the physical consequences of a nontrivial Haven ratio with extensive first-principles molecular dynamics in the fast ion conductor $Li_{10}GeP_2S_{12}$. Last, we show that the approximate tracer diffusion still provides accurate activation free energies.

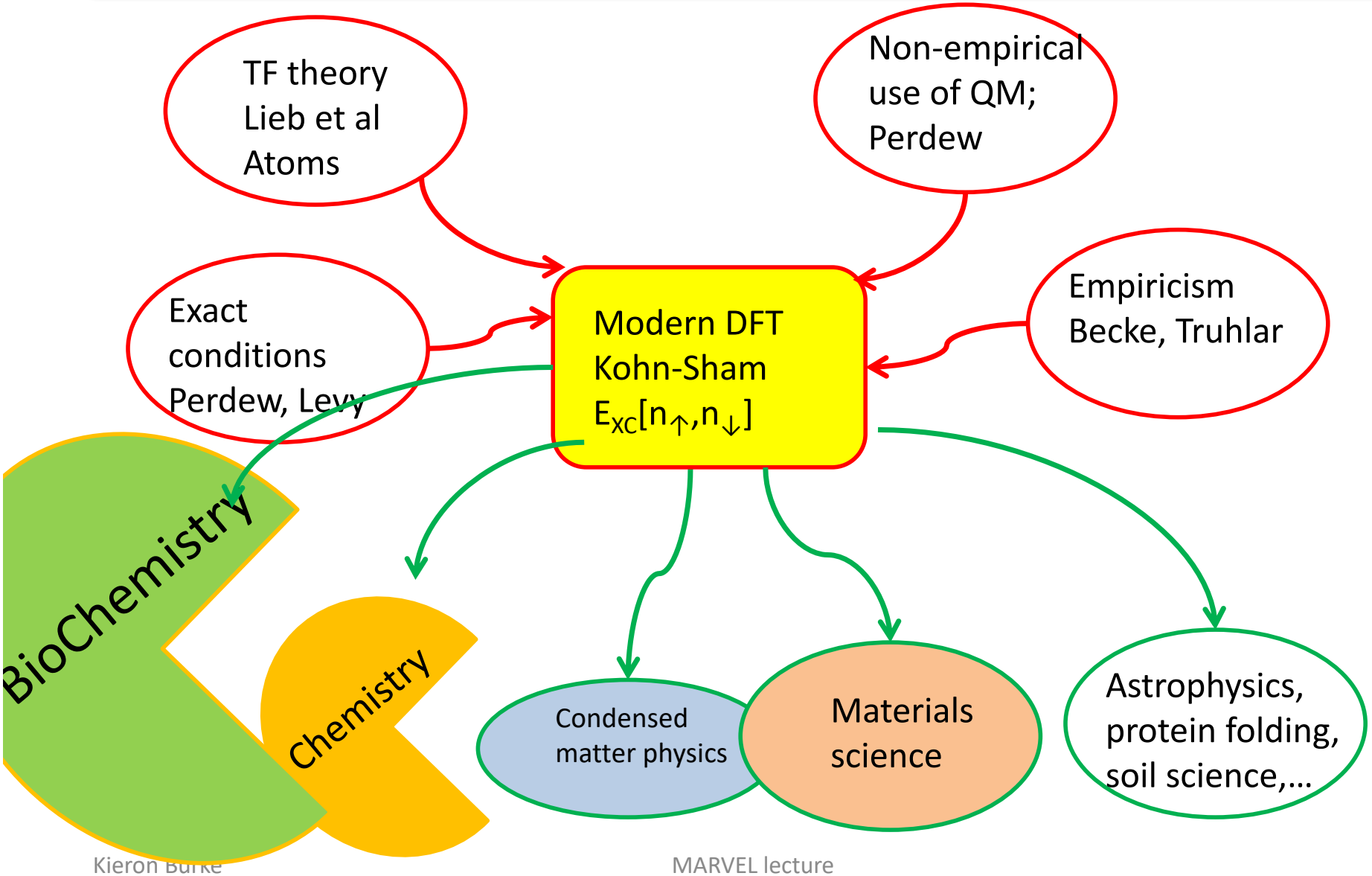
DOI: 10.1103/PhysRevMaterials.1.025402

Perdew's systematic approach to XC

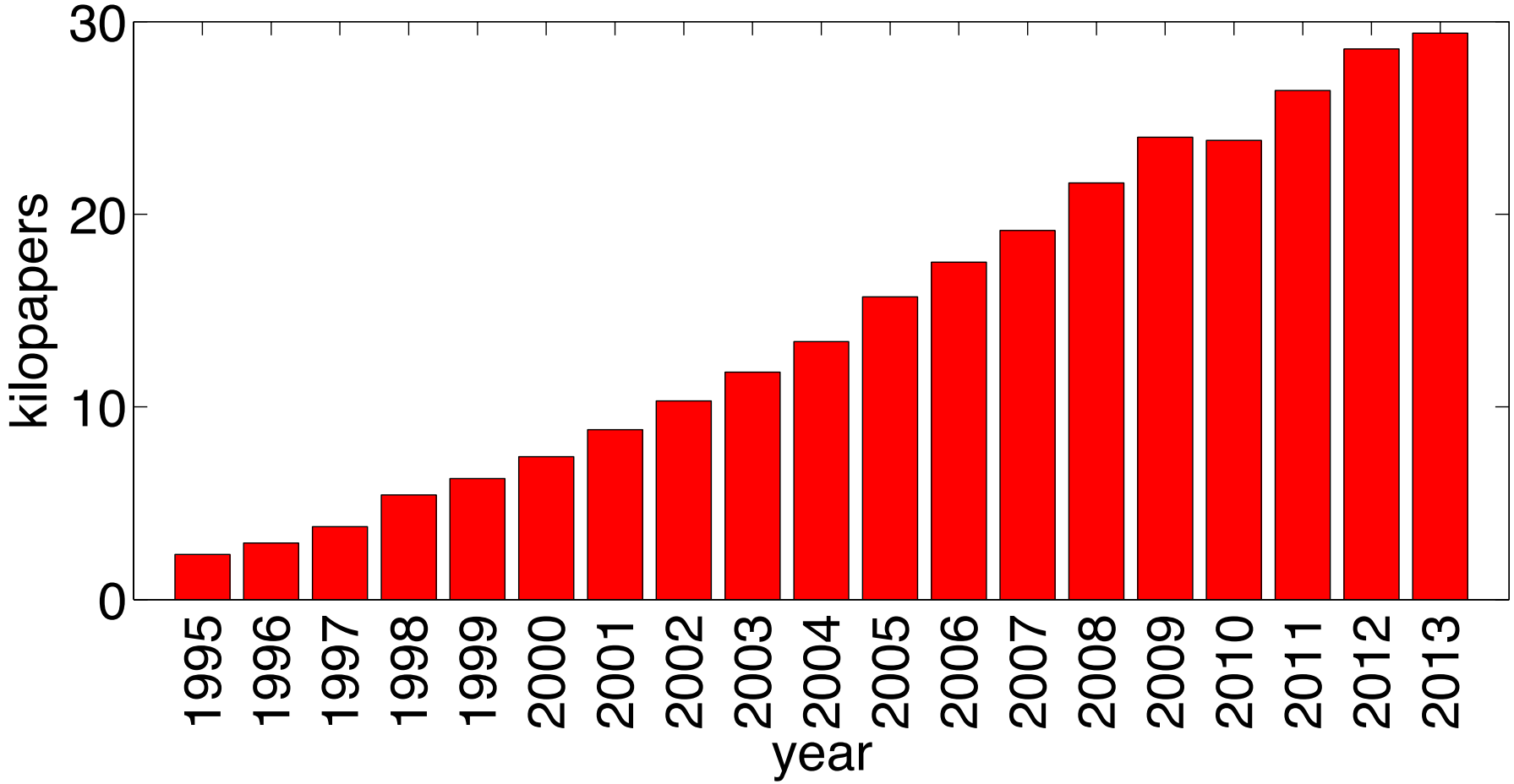
- Idea: Successively refine approximations
- Use exact conditions
- Avoid fitting of parameters to data sets
- Each rung is more sophisticated, but costs more



Big picture


TF theory
Lieb et al
Atoms
Non-empirical
use of QM;
Perdew
Exact
conditions
Perdew, Levy
Modern DFT
Kohn-Sham
$E_{XC}[n_\uparrow,n_\downarrow]$
Empiricism
Becke, Truhlar
BioChemistry
Chemistry
Condensed
matter physics
Materials
science
Astrophysics,
protein folding,
soil science,…

DFT papers



DFT: A Theory Full of Holes, Aurora Pribram-Jones, David A. Gross, Kieron Burke, Annual Review of Physical Chemistry (2014).

In reality…


wB97
LM81
X3LYP
PW92
SAOP
OBPW
BNL05
PW91
EB08
M06-L
KMLYP
TLYP
PBE
LYP
V5LYP
LB9483
VP86
WC06
PBEsol
wB97XD
wB97X
MCY06
B2PLYP
BMK04
FT97
BOP
B97
PBEOP
WI
GL76
B3LYP
EXX
rPBE
B97-2
CS1
Bh3LYP
rB86
DK87
vBH
M05-2X
BR89
mPW
WL
TPSSh
GV09
HSE
OPTX
KT2
xPBE
muB88
M06-2X
HL71
M06-HF
M05
revPBE
HCTH
PKZB99
PBE0
muPBE
LRC-wPBE0
PL81
BP86
LG93
KCIS
LSDA
KT1
lhf
wigner
VWN5
O3LYP
SOGGA
RPA
LC-wPBE
B88
B95
BRC89
BJ06
B97-D
PZ81
TPSS
PK06
B86-mgc
VS98
EDF1
B94
AM05
GG99
PW86
B97-1
EDF2
mPBE
G96
CAM-B3LYP
wPBE
KT3
VWN1

Machine learning in physical sciences

- Explosion of interest in last 5 years
- Machine learning/big data/data science very broad terms
- Some examples:
  - Searching databases of materials calculations to find optimal functionality
  - Searching chemical compound space
  - Accelerating sampling
  - Designing interatomic potentials

ML interatomic potentials

PHYSICAL REVIEW MATERIALS **2**, 013808 (2018)

Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron

Daniele Dragoni,[1,2] Thomas D. Daff,[3] Gábor Csányi,[3] and Nicola Marzari[1]

[1]*Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

[2]*Dipartimento di Scienza dei Materiali, Università di Milano-Bicocca, Via R. Cozzi 55, I-20125 Milano, Italy*

[3]*Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom*

(Received 23 August 2017; published 30 January 2018)

We show that the Gaussian Approximation Potential (GAP) machine-learning framework can describe complex magnetic potential energy surfaces, taking ferromagnetic iron as a paradigmatic challenging case. The training database includes total energies, forces, and stresses obtained from density-functional theory in the generalized-gradient approximation, and comprises approximately 150,000 local atomic environments, ranging from pristine and defected bulk configurations to surfaces and generalized stacking faults with different crystallographic orientations. We find the structural, vibrational, and thermodynamic properties of the GAP model to be in excellent agreement with those obtained directly from first-principles electronic-structure calculations. There is good transferability to quantities, such as Peierls energy barriers, which are determined to a large extent by atomic configurations that were not part of the training set. We observe the benefit and the need of using highly converged electronic-structure calculations to sample a target potential energy surface. The end result is a systematically improvable potential that can achieve the same accuracy of density-functional theory calculations, but at a fraction of the computational cost.

DOI: 10.1103/PhysRevMaterials.2.013808

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

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

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

length scale

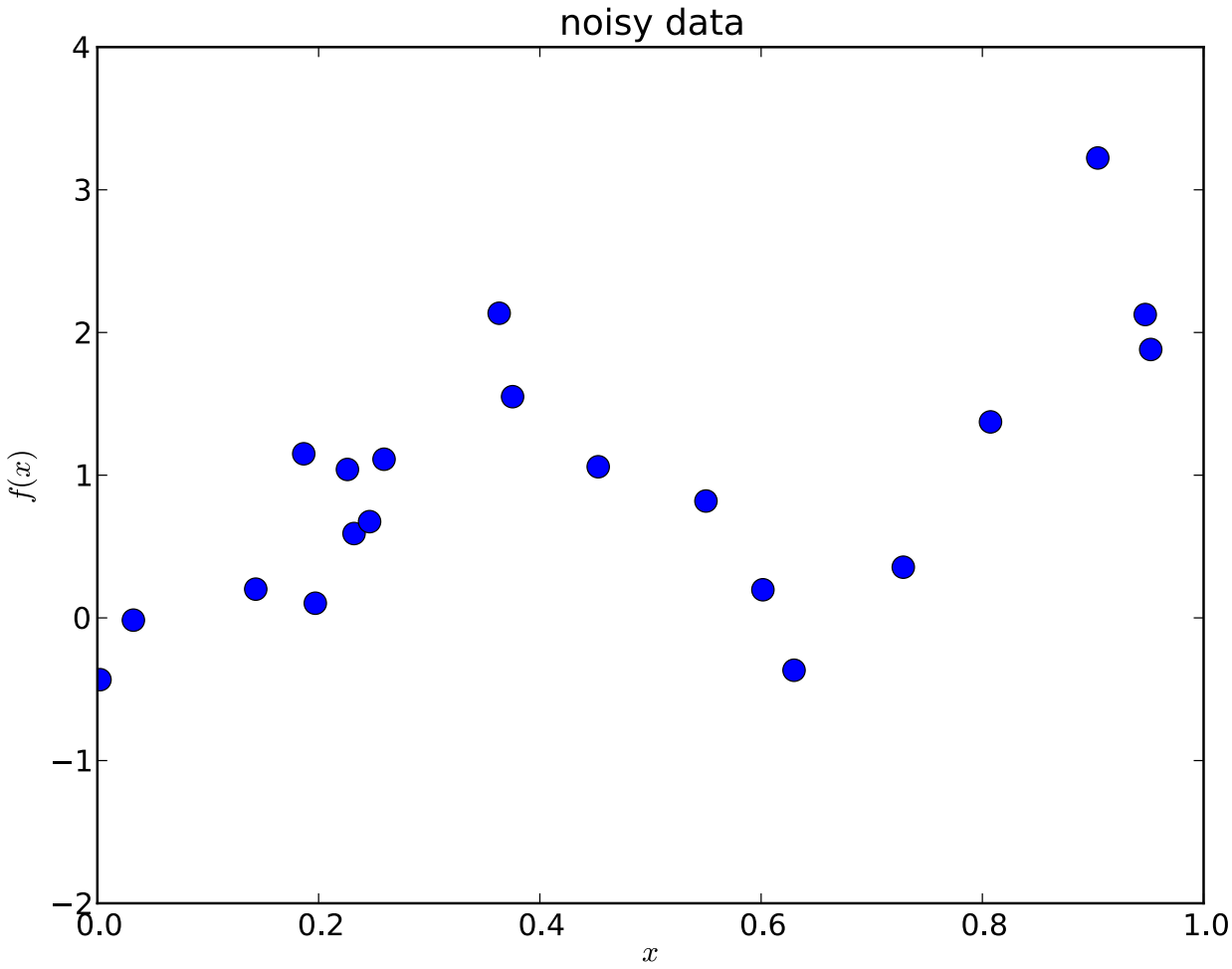
- Minimize:

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

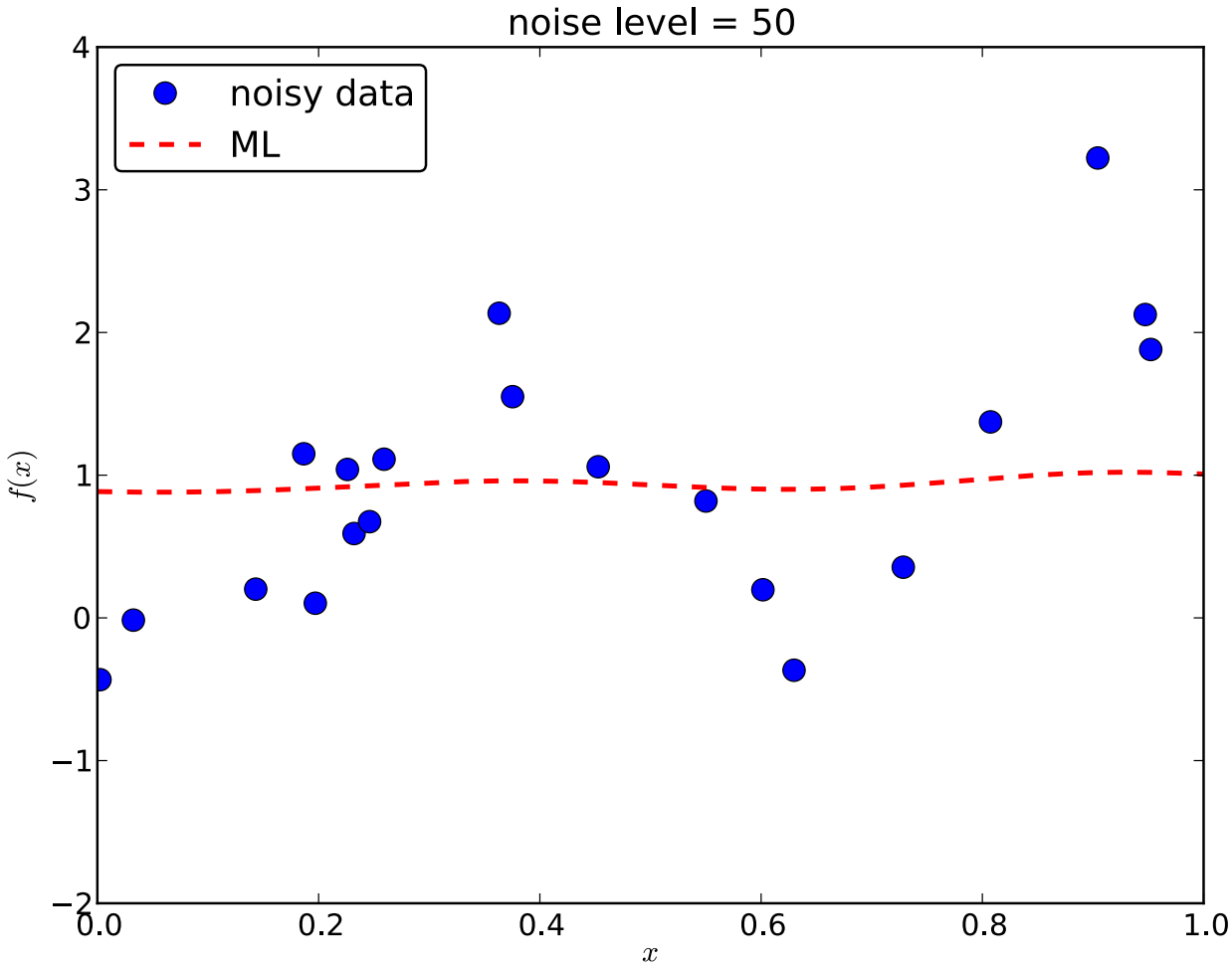
noise level

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

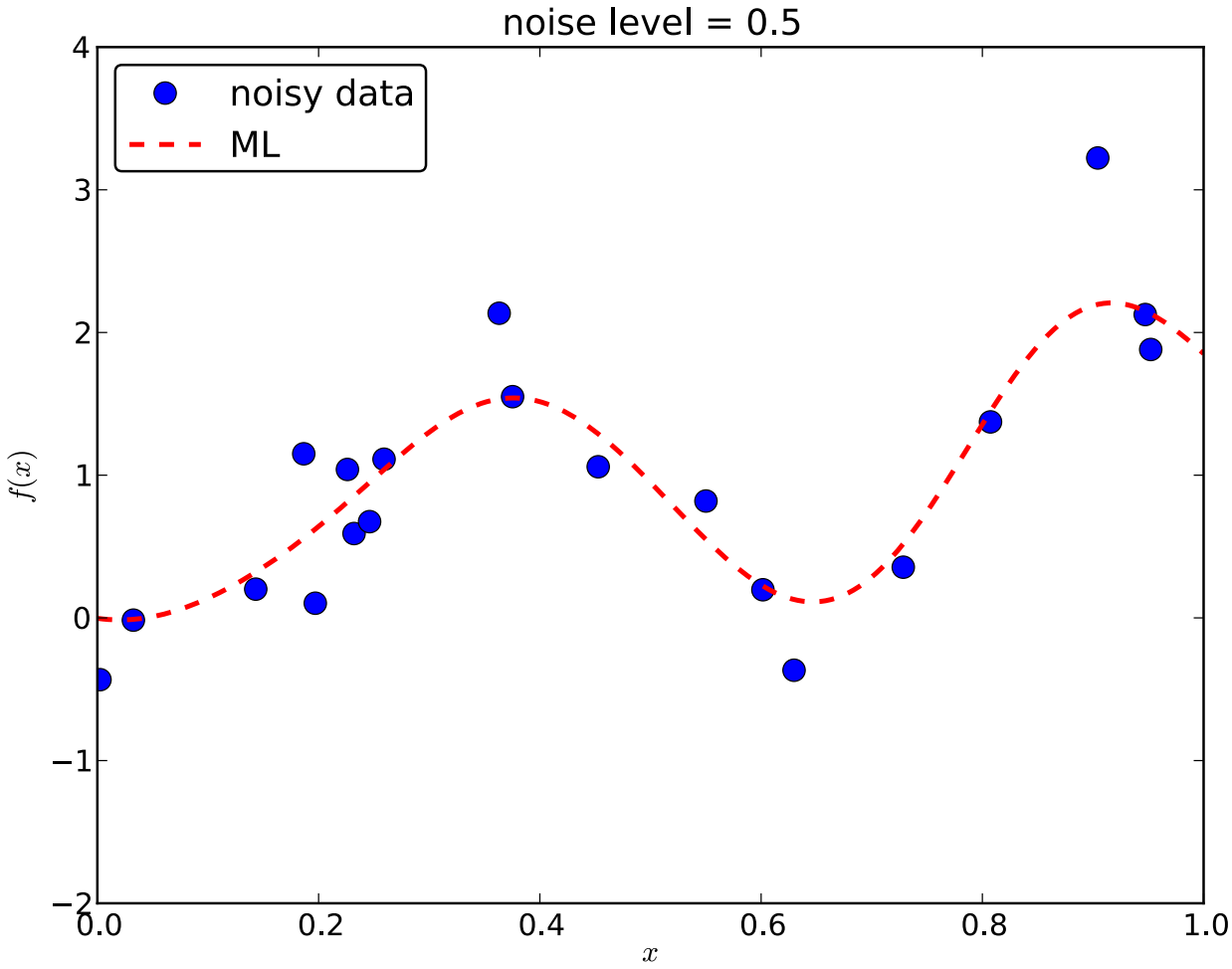
Fitting a simple function



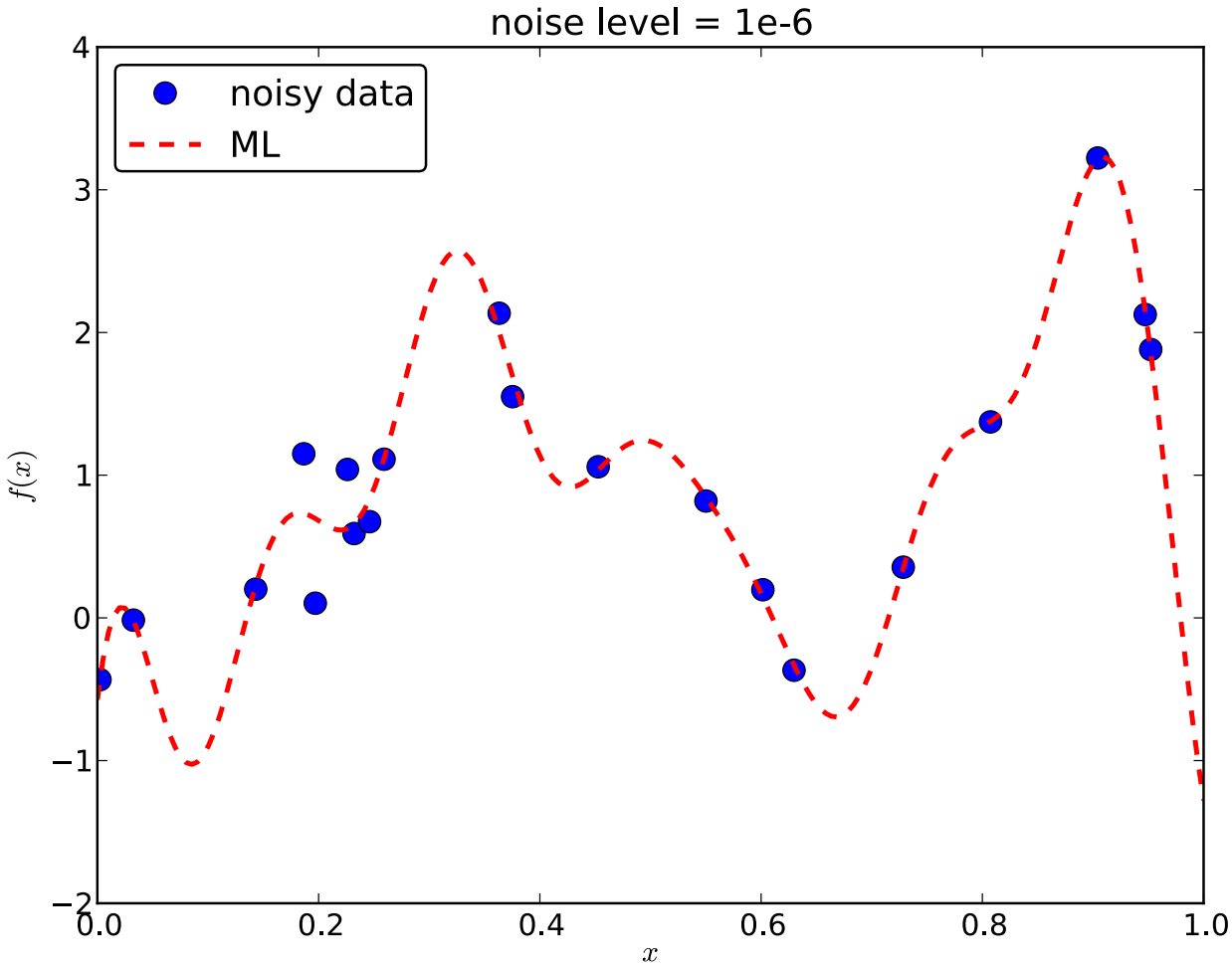
Too high noise level: underfit



Medium noise level

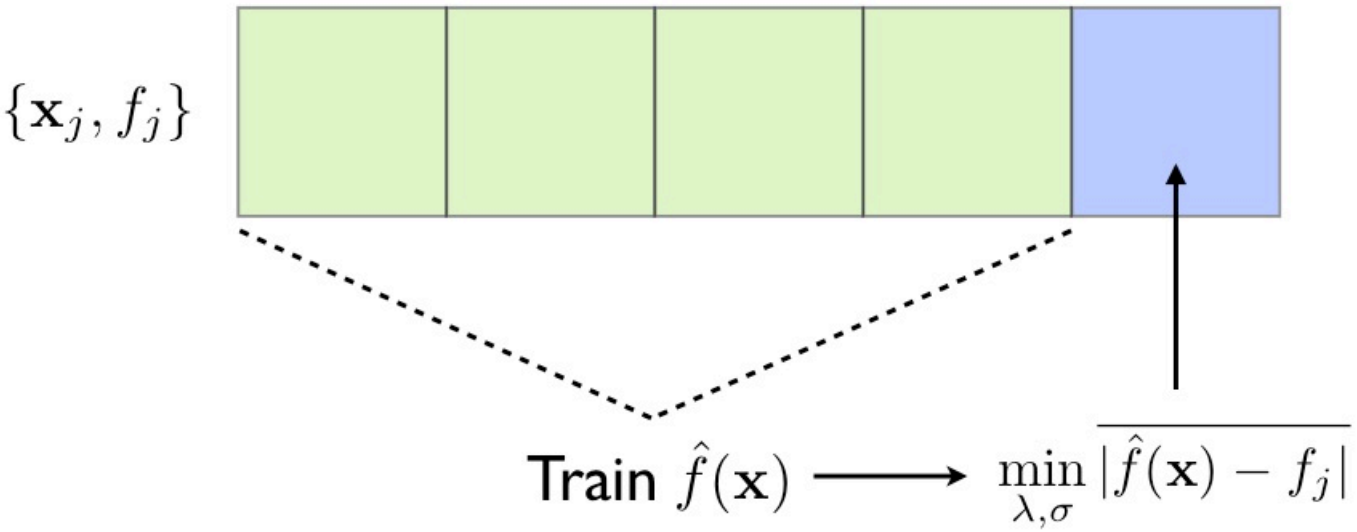


Small noise level: overfit



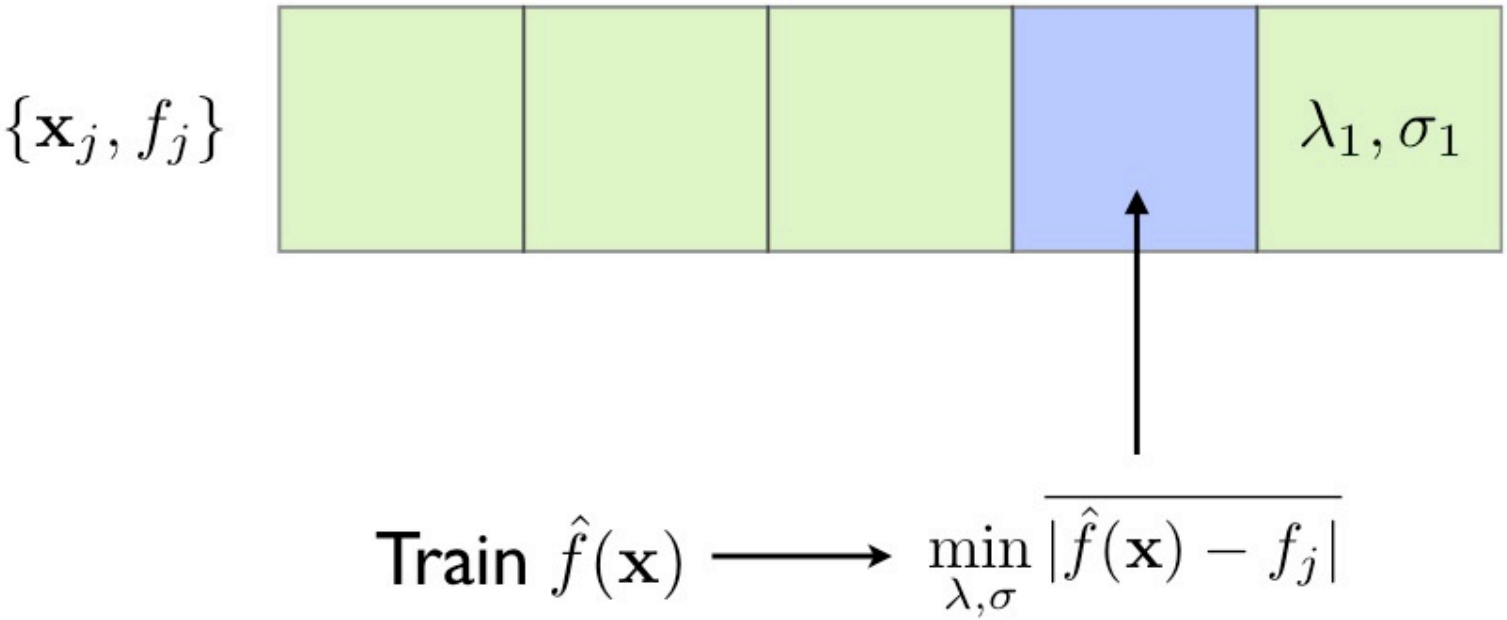
Cross validation

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



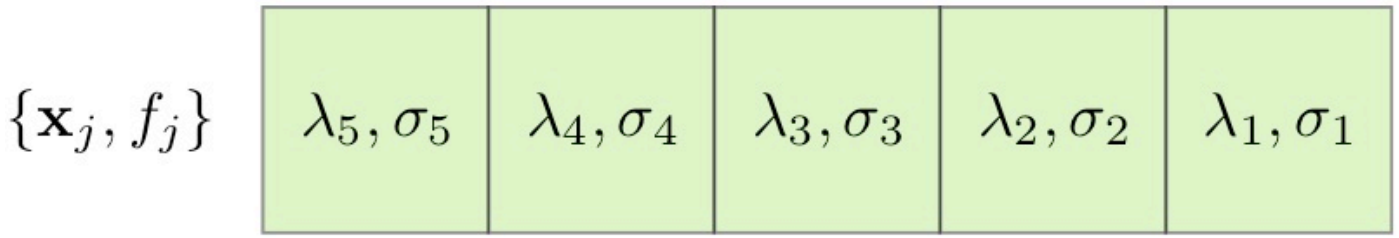
Cross validation

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



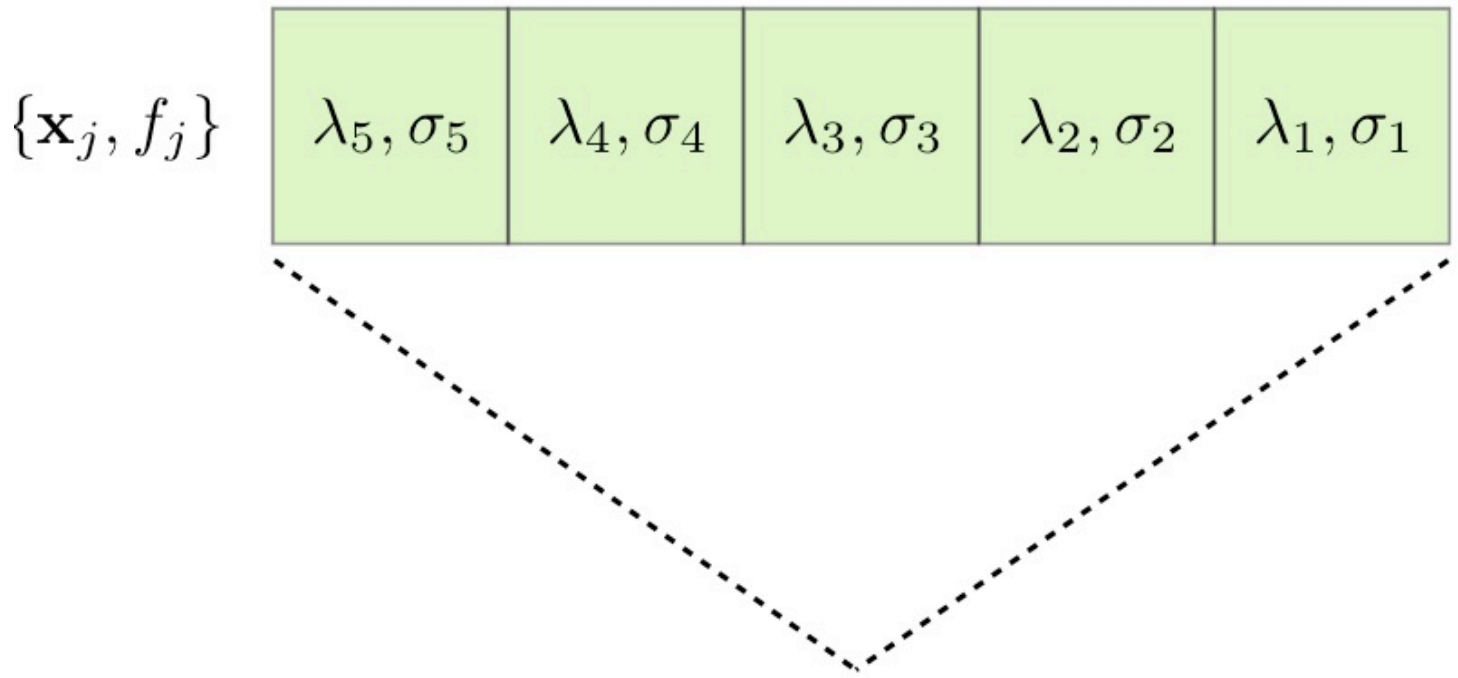
More cross validation

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



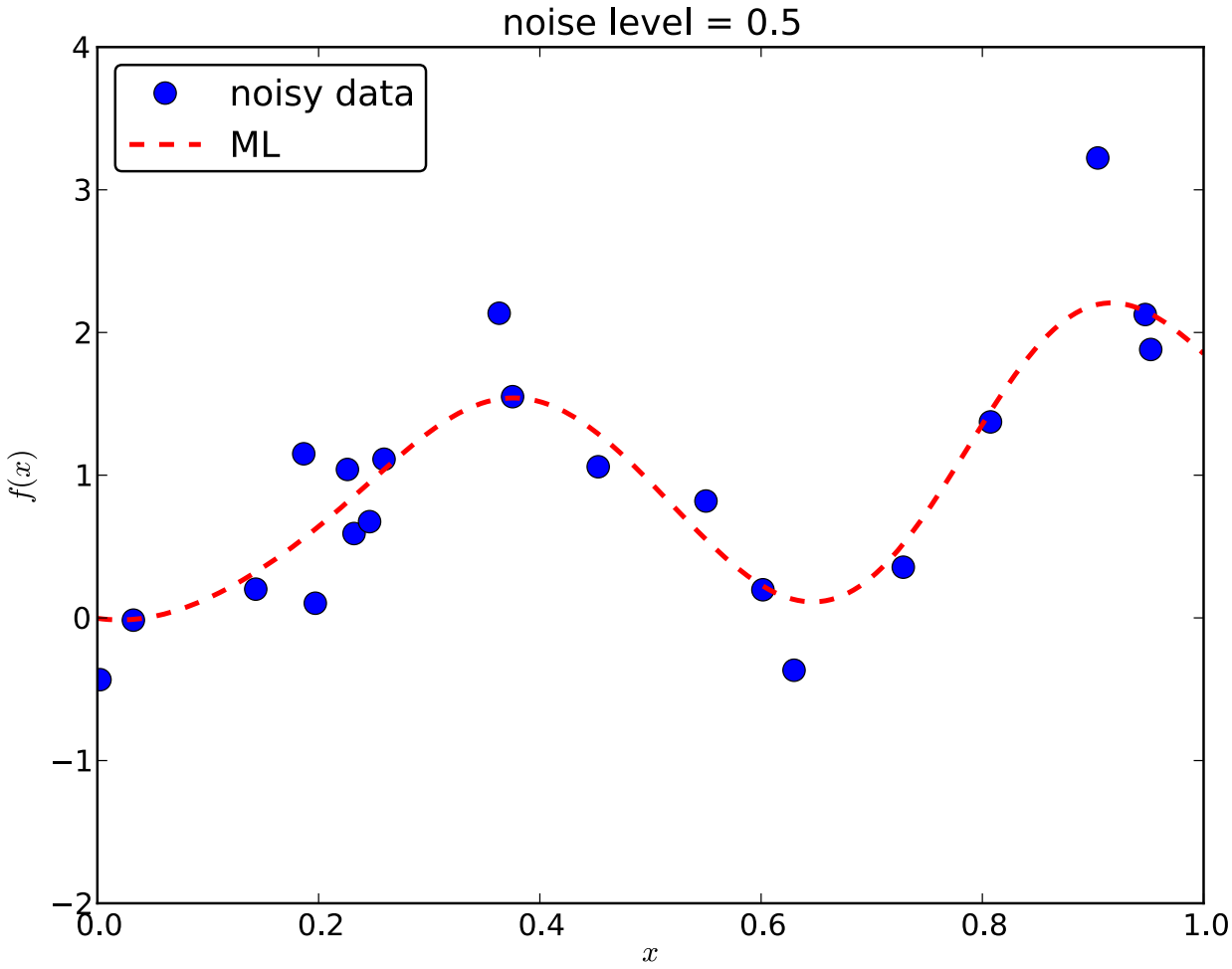
Average over samples

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

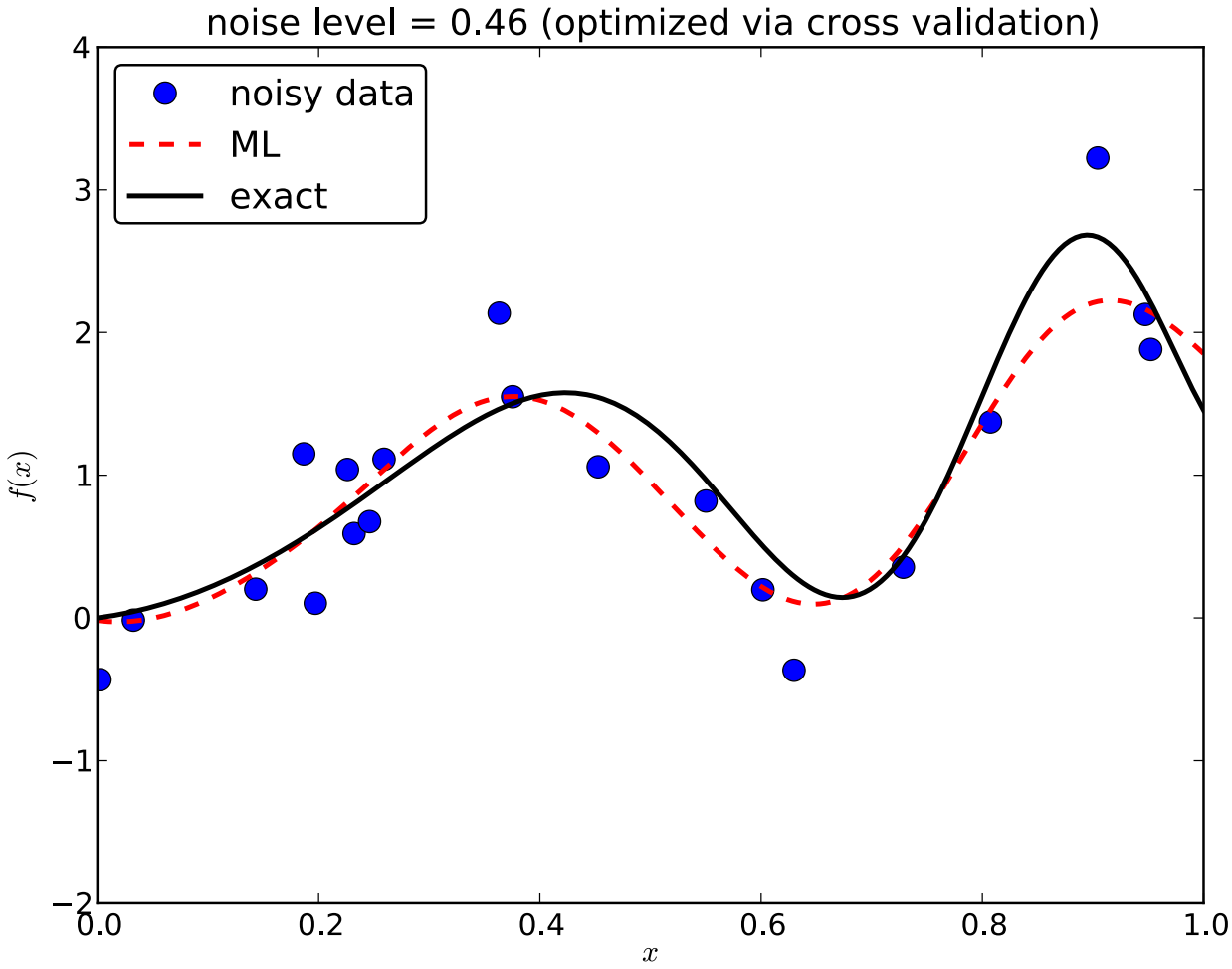


$$\lambda, \sigma = \text{median}_j \{\lambda_j, \sigma_j\}$$

Medium noise level



Exact function and best fit



ML applications in electronic structure

- Most with Klaus Mueller of TU Berlin, computer science.


Klaus-Robert Müller

- ML now being applied directly to, e.g., molecular energies from geometries for drug design, many by Matthias Rupp (FHI Berlin).


Matthias Rupp

- Our efforts are focused on finding $T_s[n]$ from examples, work by John Snyder (Humboldt fellow at TU Berlin/MPI Halle)


John Snyder

If only we knew the kinetic energy as a density functional

- The KS equations are solving the following equation for us:

$$\frac{\delta T_{\mathrm{S}}}{\delta n(\mathbf{r})} = -v(\mathbf{r}) - v_{\mathrm{H}}[n](\mathbf{r}) - v_{\mathrm{XC}}[n](\mathbf{r})$$

- If we had an explicit approximation for $T_S[n]$, we could solve this directly.

Basic idea: Orbital-free DFT

- For a limited variety of one-body potentials, construct kinetic energy functionals via ML that are sufficiently accurate to do the job.
- Typically, require accuracy of 1 kcal/mol =1.6 mHa=0.05 eV
- For kinetic energy, also need functional derivative.
- Most useful when multiple slightly different uses of DFT, eg during an MD run.
- Functional could be disposable, i.e., thrown away at end of run.
- Is completely non-local in general, so best advantage when bonds break, as local functionals fail.

Demo problem in DFT

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

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

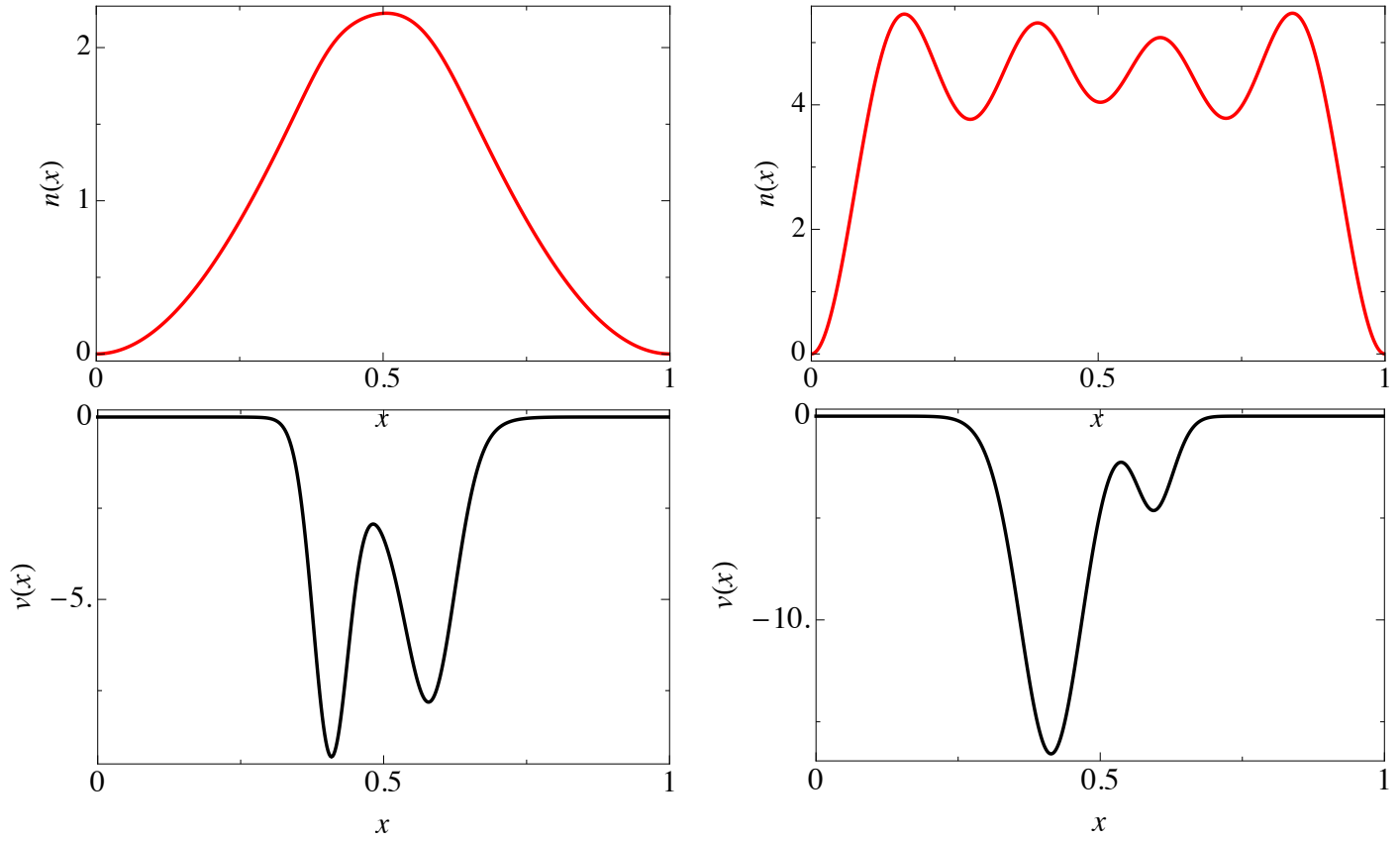
- Represent the density on a grid with spacing $\Delta x = 1/(G-1)$
- ML-DFA for KE:

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

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

Dataset

Generate 2000 potentials. Solve for up to 4 electrons.



Performance for T_s

kcal/mol

N	M	λ	σ	$\overline{\lvert\Delta T\rvert}$	$\lvert\Delta T\rvert^{\mathrm{std}}$	$\lvert\Delta T\rvert^{\mathrm{max}}$
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

functional derivative?

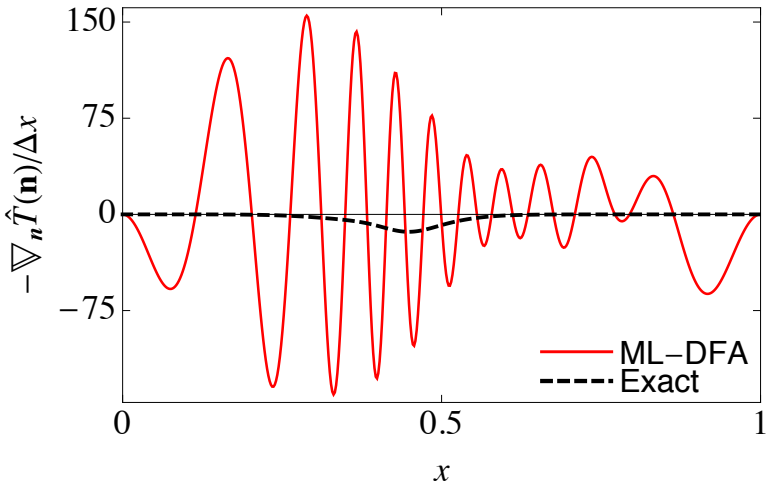
Exact

$$\frac{\delta T[n]}{\delta n(x)} = \mu - v(x)$$

ML-DFA

$$\frac{1}{\Delta x}\nabla_{\mathbf{n}}\hat{T}(\mathbf{n}) = \sum_{j=1}^{M} \alpha'_j(\mathbf{n}_j - \mathbf{n})k(\mathbf{n}_j, \mathbf{n})$$

$$\alpha'_j = \alpha_j/(\sigma^2\Delta x)$$



- Functionals are defined on infinite-dimensional spaces
- With finite interpolation, can always find bad directions
- Can we make a cruder definition that will work for our purposes?

Principal component analysis

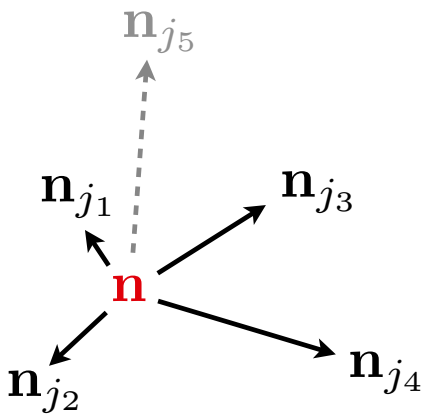
$$X = (\mathbf{n}_{j_1} - \mathbf{n}, \ldots, \mathbf{n}_{j_m} - \mathbf{n})^\top$$

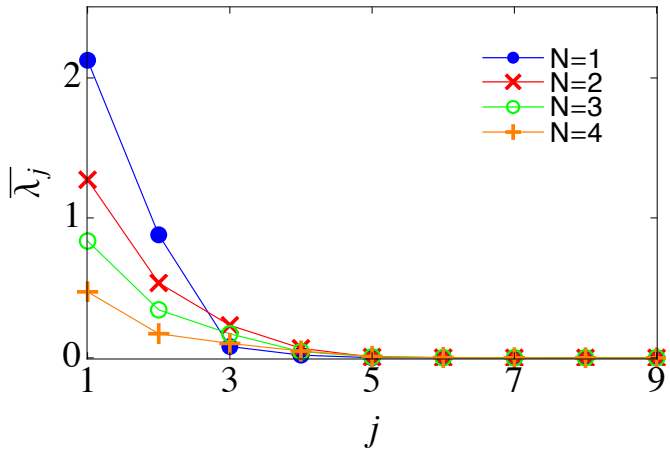
$$C = \frac{1}{m} X^\top X$$

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

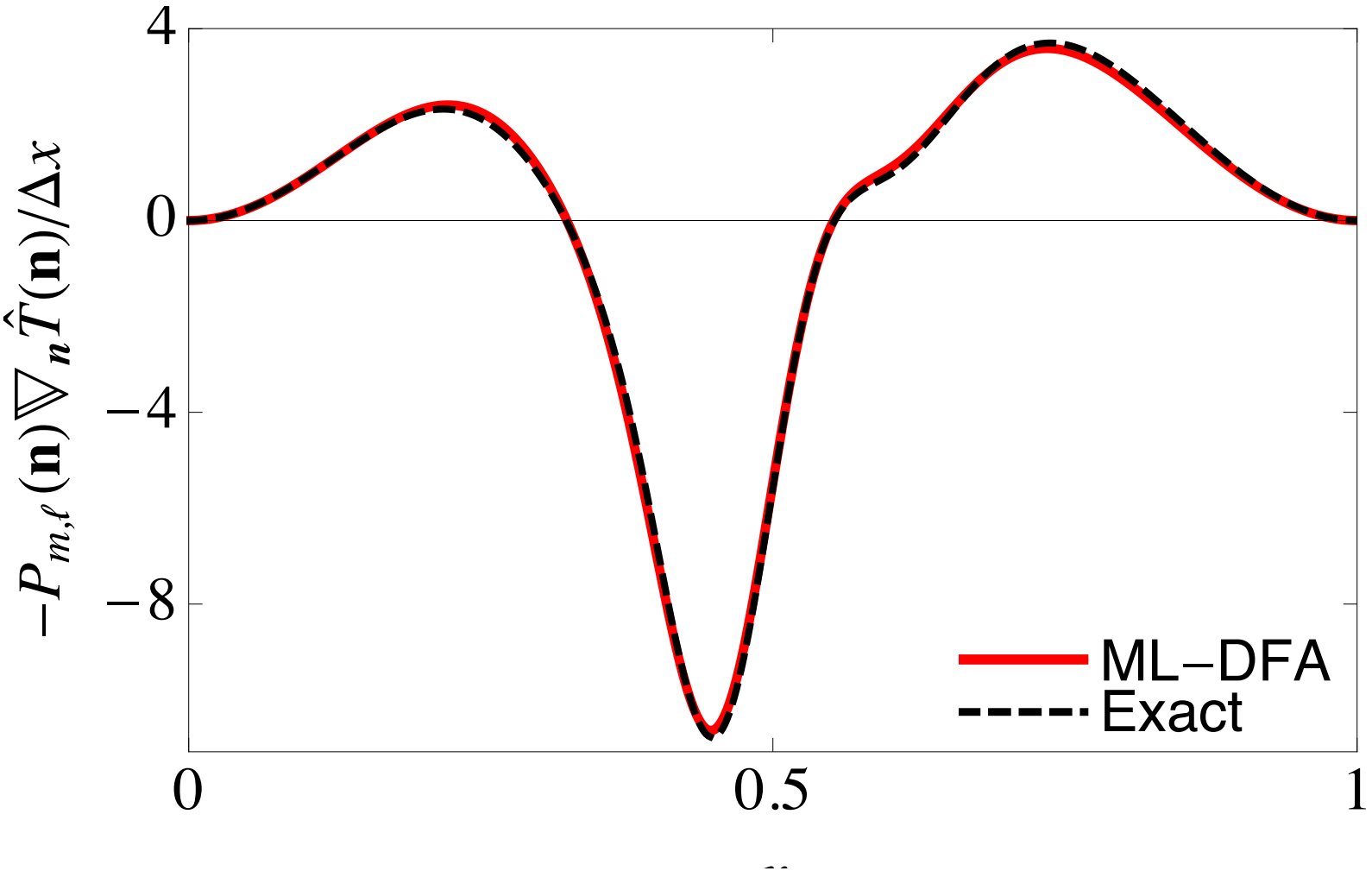
$$P_{m,\ell}(\mathbf{n}) = V^\top V$$

$$V = (\mathbf{x}_1, \ldots, \mathbf{x}_\ell)^\top$$





Projected functional derivative



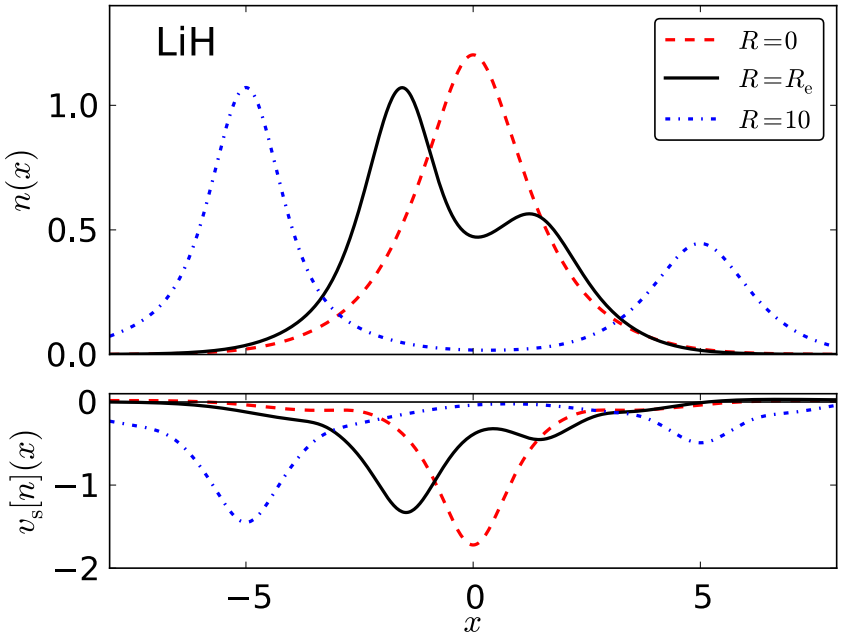
Lessons

- Exact noise-free data infinitely available for $T_s[n]$, every cycle of every KS calculation in the world provides examples.
- Need very accurate derivatives to get accurate density from Euler equation.
- Can find ways to bypass this.
- Functionals can be made arbitrarily accurate with sufficient data.

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

Bond-breaking with ML

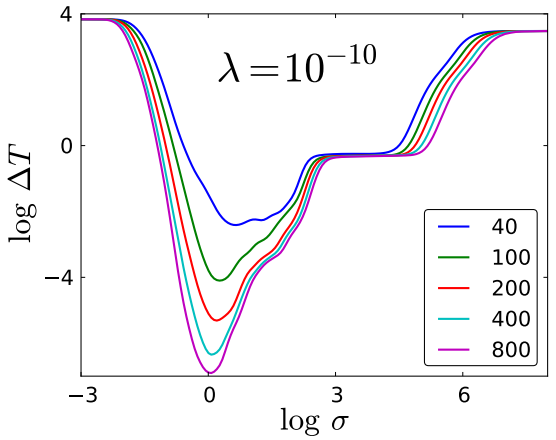
- Performed many 1d KS calculations of diatomics as function of bond length, using LDA with soft-Coulomb repulsion, including several with more than 2 electrons



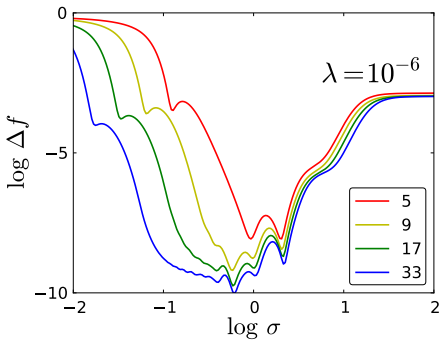
Orbital-free Bond Breaking via Machine Learning John C. Snyder, Matthias Rupp, Katja Hansen, Leo Blooston, Klaus-Robert Müller, Kieron Burke, J. Chem. Phys. **139**, 224104 (2013)

Kevin's paper: from functions to functionals

- Plot error as a function of hyperparameters



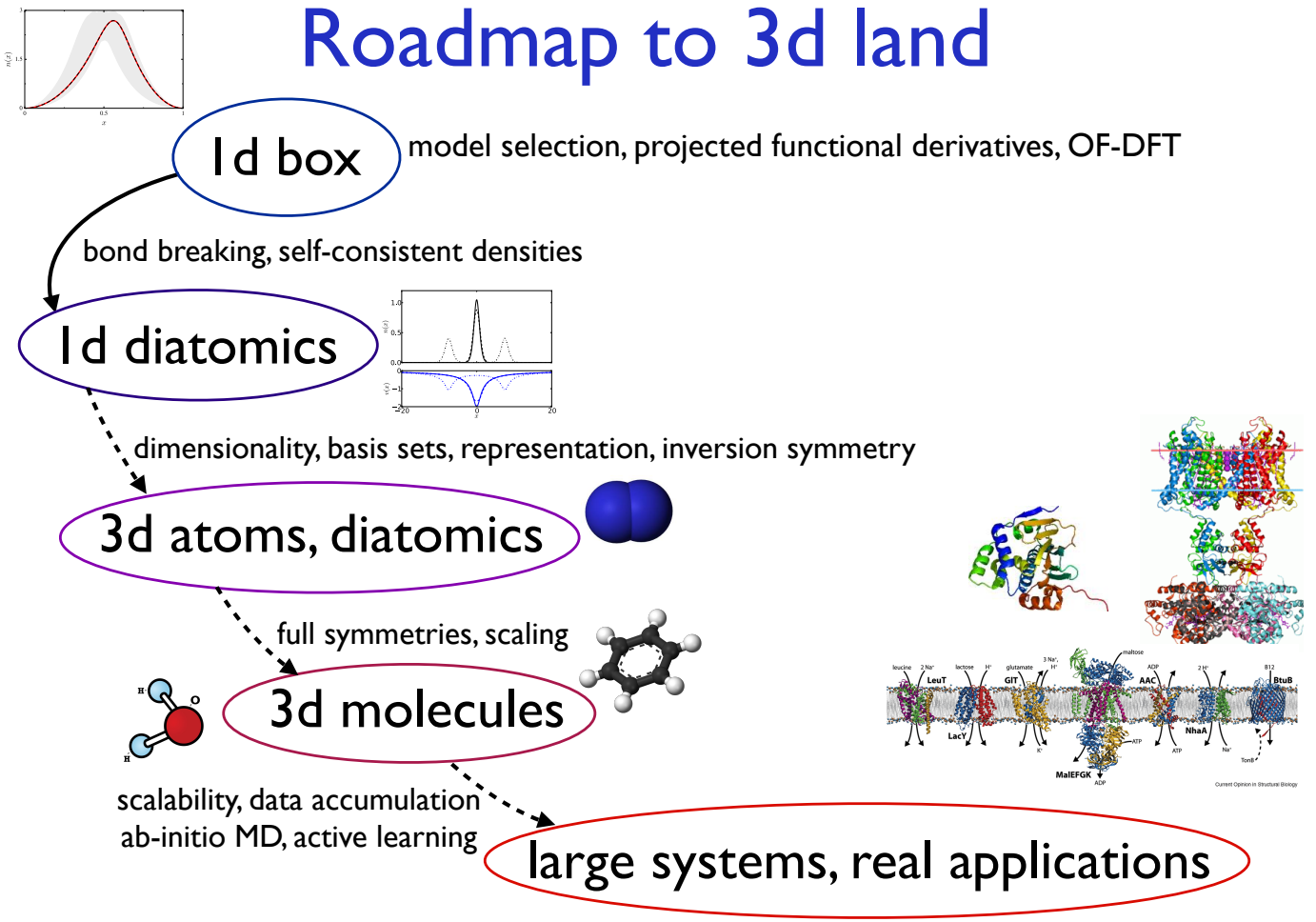
- Repeat for fitting f(x)= cos x



- 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



D. Recent results

Felix Brockherde

Li Li

Thomas E. Baker

2 new papers

- *By-passing the KS equations with ML (acc Nat Comm)*
  - Felix Brockherde, Li Li, Klaus Muller, KB,…
  - Avoids functional derivative
  - Applied in 3D
  - Still doing KS problem, $T_s[n]$
- *Pure Density Functional for Strong Correlations and the Thermodynamic Limit Using Machine Learning. in Phys Rev B.*
  - Li Li, Thomas E. Baker, Steven R. White and KB
  - Do interacting functional (ie. Exact E_{XC})
  - Do strong correlation
  - Do thermodynamic limit
  - Still in 1d

Machine-learned approximations

By-passing the Kohn-Sham equations with machine learning

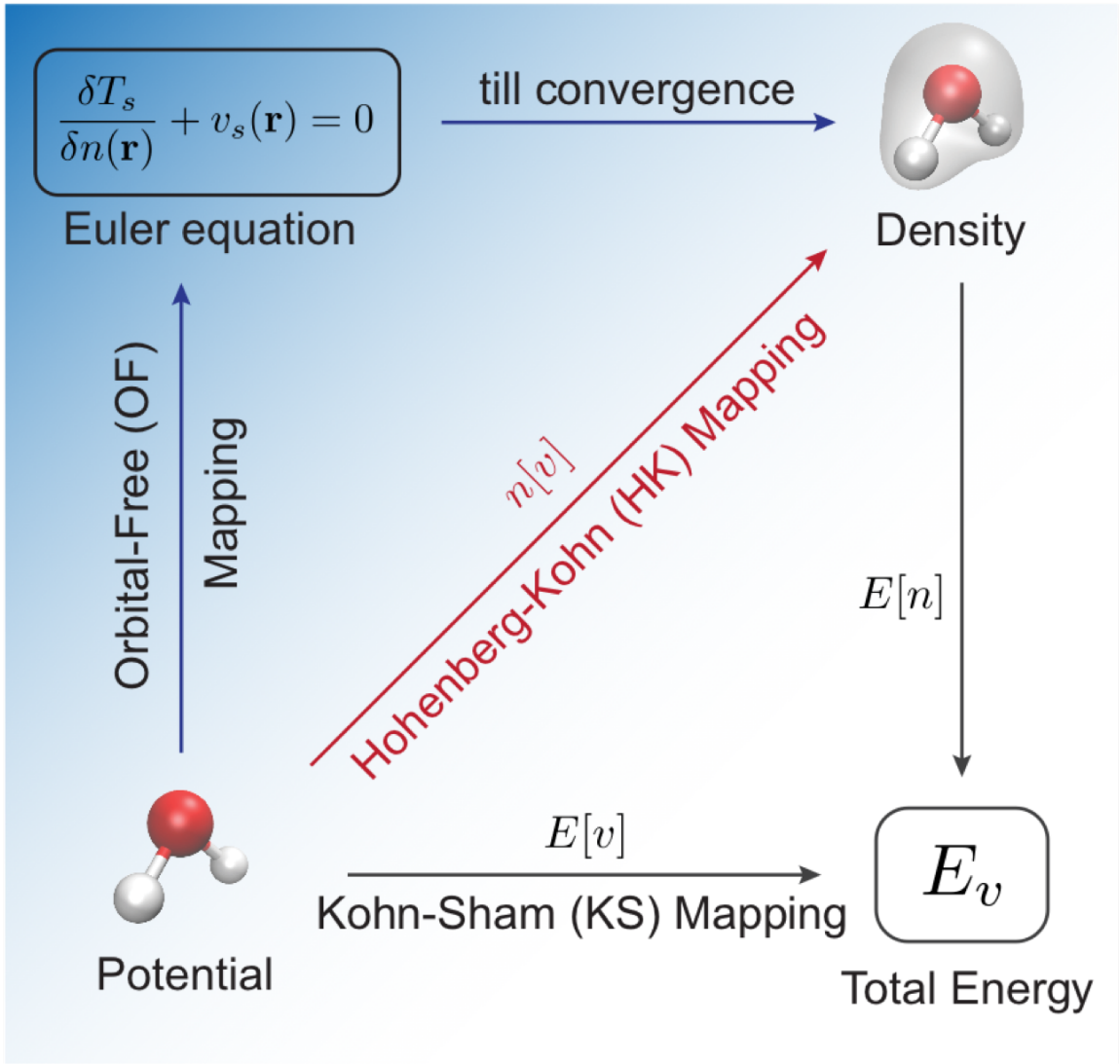
Felix Brockherde,[1,2] Leslie Vogt,[3] Li Li,[4] Mark E. Tuckerman,[3,5,6] Kieron Burke,[7,4,*] and Klaus-Robert Müller[1,8,*]

[1] *Machine Learning Group, Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany*
[2] *Max-Planck-Institut für Mikrostrukturphysik, Weinberg 2, 06120 Halle, Germany*
[3] *Department of Chemistry, New York University, New York, NY 10003, USA*
[4] *Departments of Physics and Astronomy, University of California, Irvine, CA 92697, USA*
[5] *Courant Institute of Mathematical Science, New York University, New York, NY 10003, USA*
[6] *NYU-ECNU Center for Computational Chemistry at NYU Shanghai, 3663 Zhongshan Road North, Shanghai 200062, China*
[7] *Departments of Chemistry, University of California, Irvine, CA 92697, USA*
[8] *Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Republic of Korea*

(Dated: January 31, 2017)

Last year, at least 30,000 scientific papers used the Kohn-Sham scheme of density functional theory to solve electronic structure problems in a wide variety of scientific fields, ranging from materials science to biochemistry to astrophysics. Machine learning holds the promise of learning the kinetic energy functional via examples, by-passing the need to solve the Kohn-Sham equations. This should yield substantial savings in computer time, allowing either larger systems or longer time-scales to be tackled. However, attempts to machine-learn this functional have been limited by the need to find its derivative. The present work overcomes this difficulty by directly learning the density-potential and energy-density maps for test systems and various molecules. Both improved accuracy and lower computational cost with this method is demonstrated by reproducing DFT energies for a range of molecular geometries generated during molecular dynamics simulations. Moreover, the methodology could be applied directly to quantum chemical calculations, allowing construction of density functionals of quantum-chemical accuracy.

By-passing KS



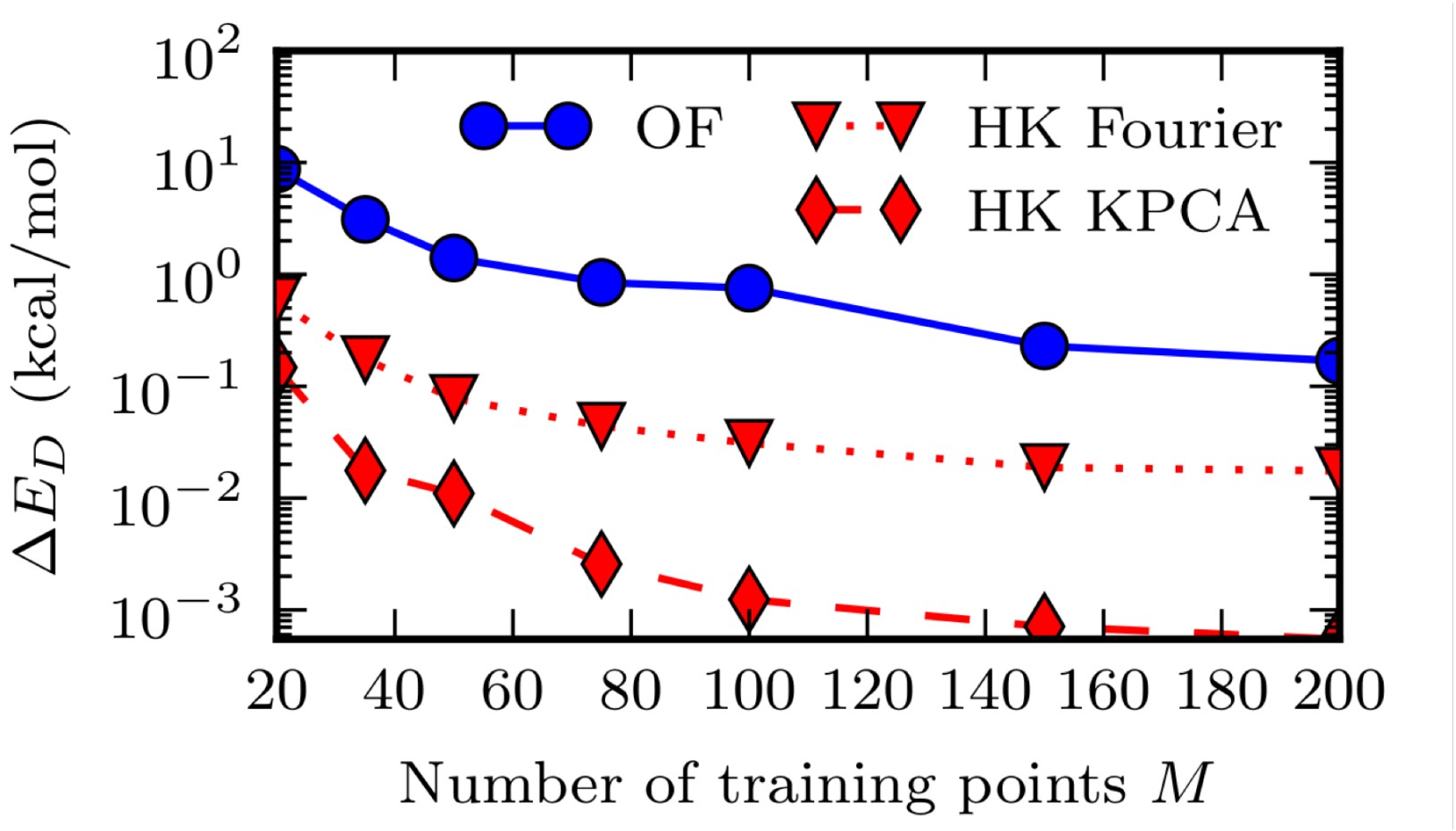
Performance of ML for HK map: Box problem

	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

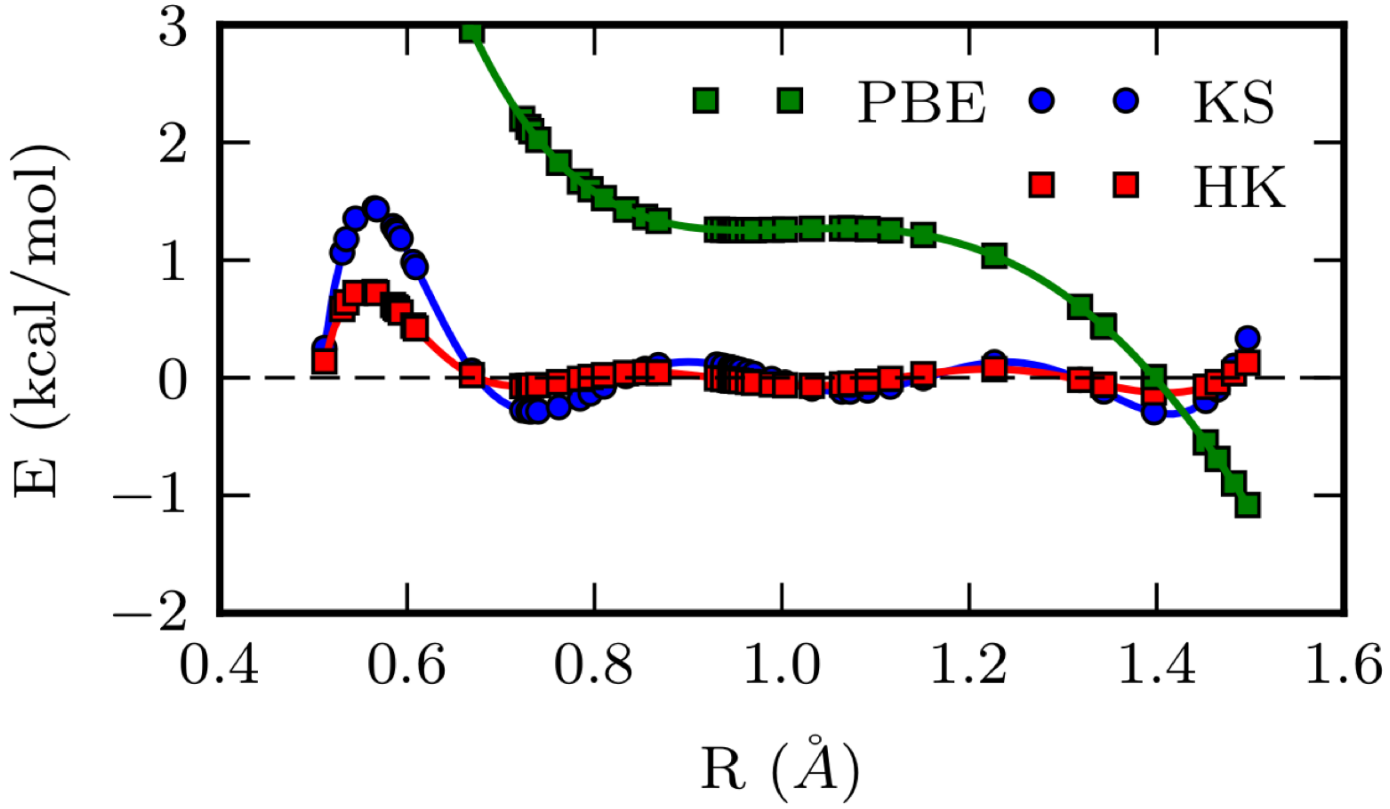
Table I. Energy errors in kcal/mol for the 1-D data set for various M, the number of training points. For definitions, see text.

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

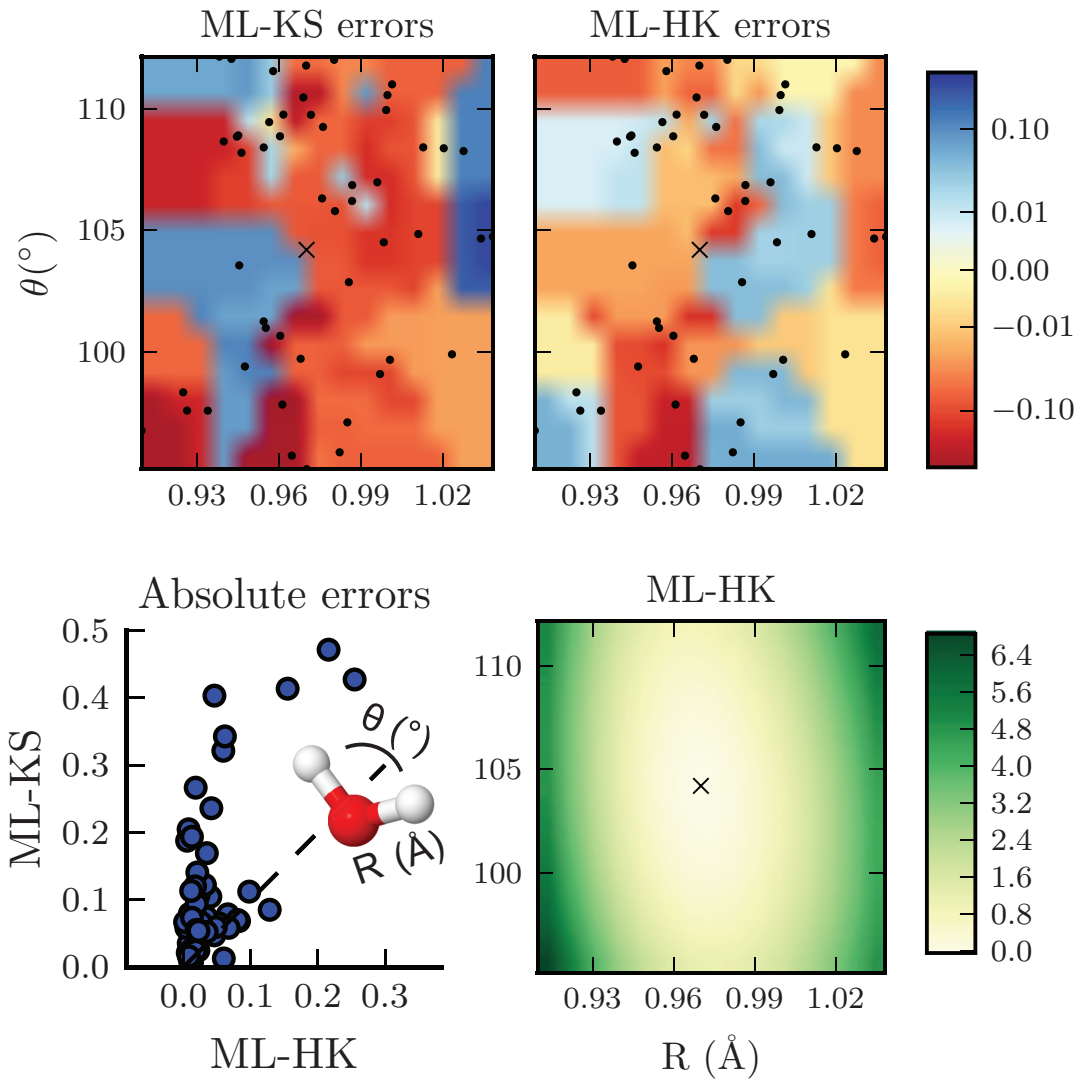
Convergence of different HK maps



Error for H_2



H_2O


ML-KS errors
ML-HK errors
110
105
100
θ(°)
0.93 0.96 0.99 1.02
0.93 0.96 0.99 1.02
0.10
0.01
0.00
−0.01
−0.10
Absolute errors
ML-HK
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)
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

MD simulations testing ML method

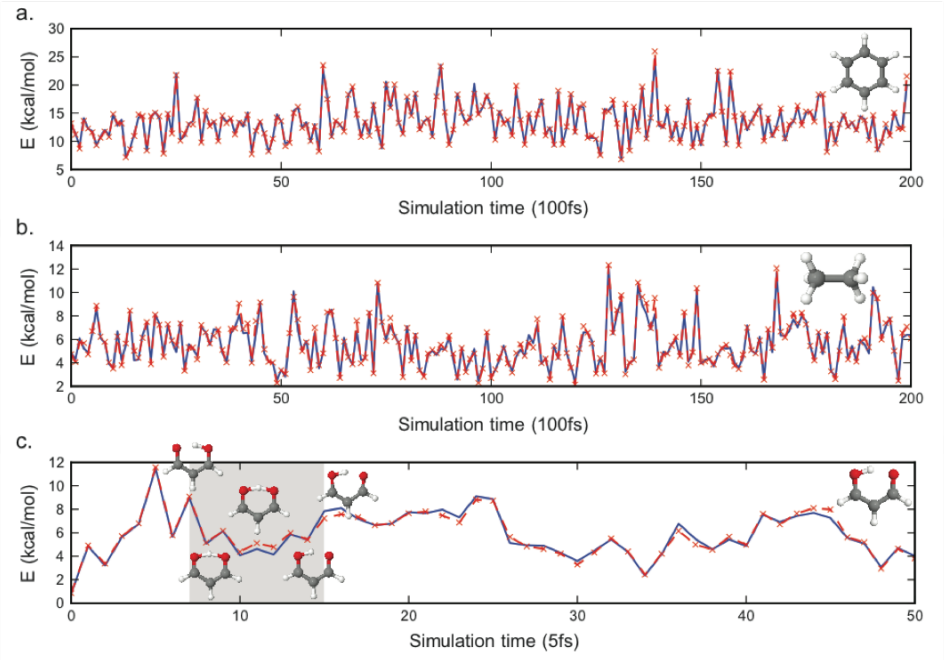


Figure 3. Energy errors of ML-HK along MD trajectories. PBE values in blue, ML-HK values in red. **a.** A 2 ps classical trajectory of benzene. **b.** A 2 ps classical trajectory of ethane. **c.** A 0.25 ps ab-initio trajectory of malonaldehyde. The ML model correctly predicts energies during a proton 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.

Accuracy of densities

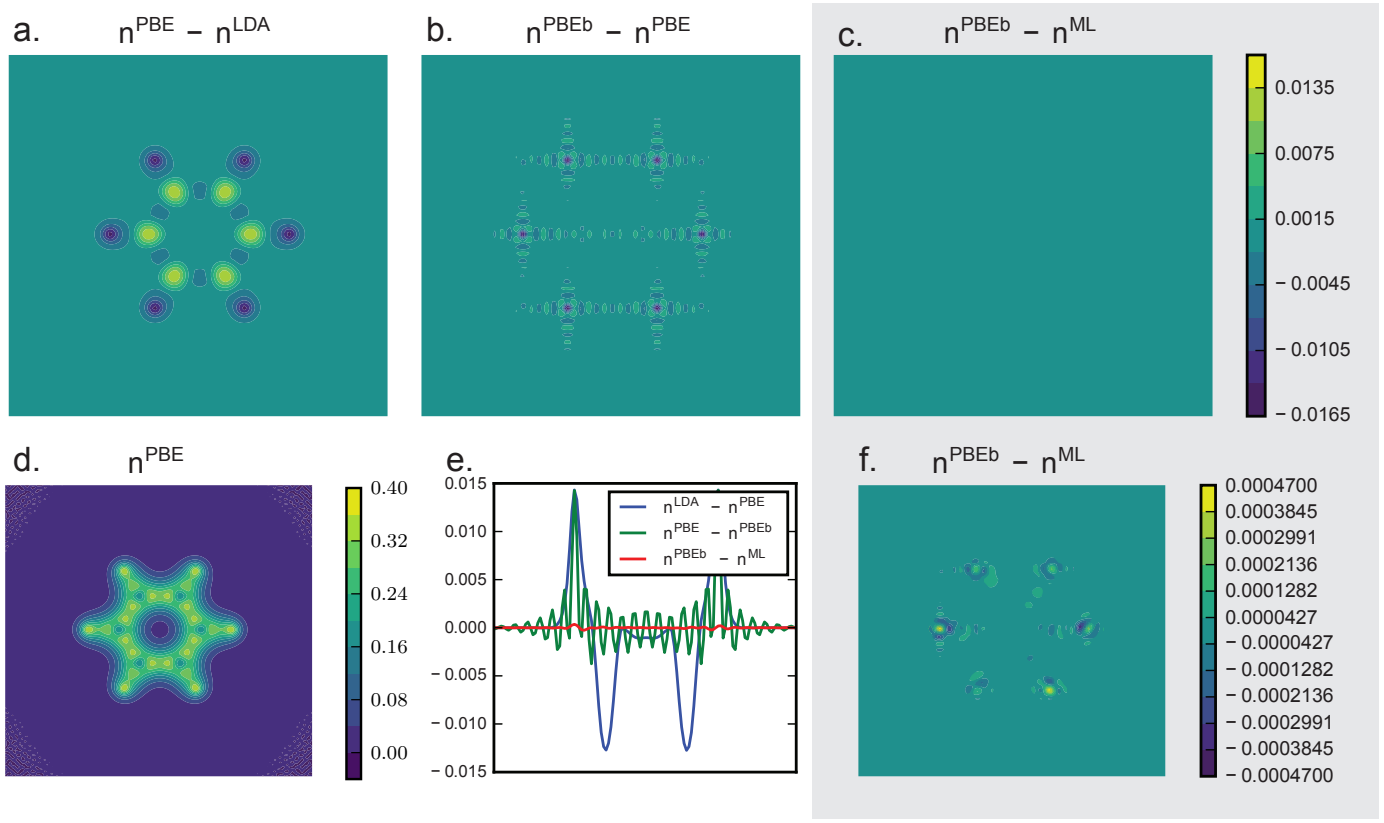
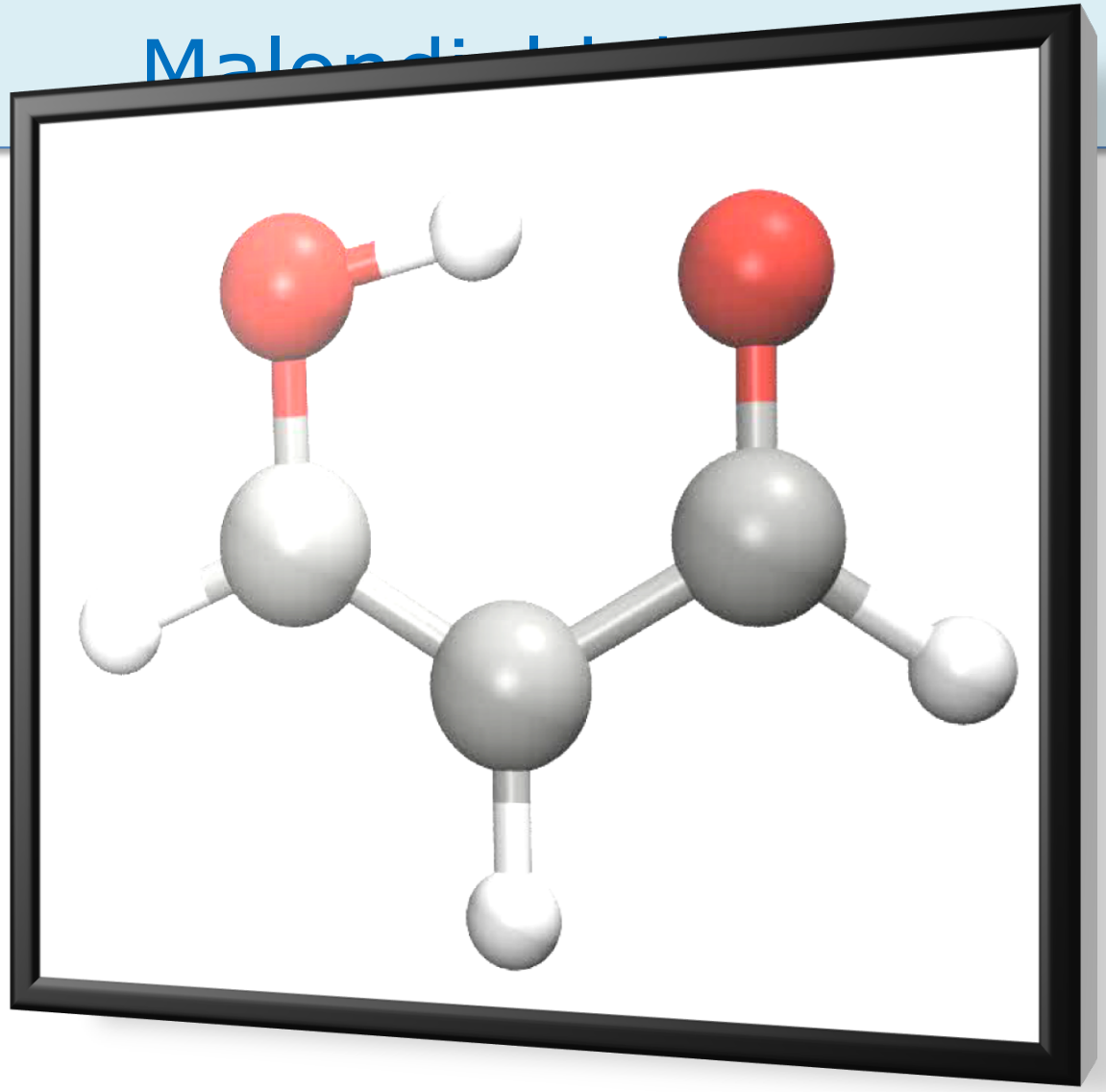


Figure 4. The precision of our density predictions using the Fourier basis for ML-HK for the molecular plane of benzene. The plots show **a**. the difference between the valence density of benzene when using PBE and LDA functionals at the PBE optimized geometry. **b**. error introduced by using the Fourier 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.).

Malondialdehyde



B175: By-passing the Kohn-Sham equations with machine learning Felix Brockherde, Leslie Vogt, Li Li, Mark E Tuckerman, Kieron Burke, Klaus-Robert M□uller, *(to appear in Nature Communications)* (2017).

Lessons

- Our 1d gradient methods become prohibitively expensive in 3d.
- Instead of using $T_s[n]$, learn n[v](r).
- Much smarter than learning $E[v_s]$
- Works for H_2 and H_2O and …
- ..MD of malonaldehyde using ML forces with Leslie Vogt and Mark Tuckerman.

C. ML for the exact functional

1d electronic structure

- Use DMRG to solve continuum problems in 1d.
- Much success in past, showing failures of DFT approximations for strong correlation.
- Here we use DMRG to generate much data of exact densities and energies
- All restricted to 1d.

Guaranteed Convergence of the Kohn-Sham Equations Lucas O. Wagner, E. M. Stoudenmire, Kieron Burke, Steven R. White, Phys. Rev. Lett. **111**, 093003 (2013).

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).

ML on exact chains of 1d H

- We train and test a machine learning F[n], the universal part of the electronic density functional, to within quantum chemical accuracy. We
  - bypass the standard Kohn-Sham approach
  - include the strong correlation of highly-stretched bonds
  - create a model for the infinite chain limit.

Convergence for H_2

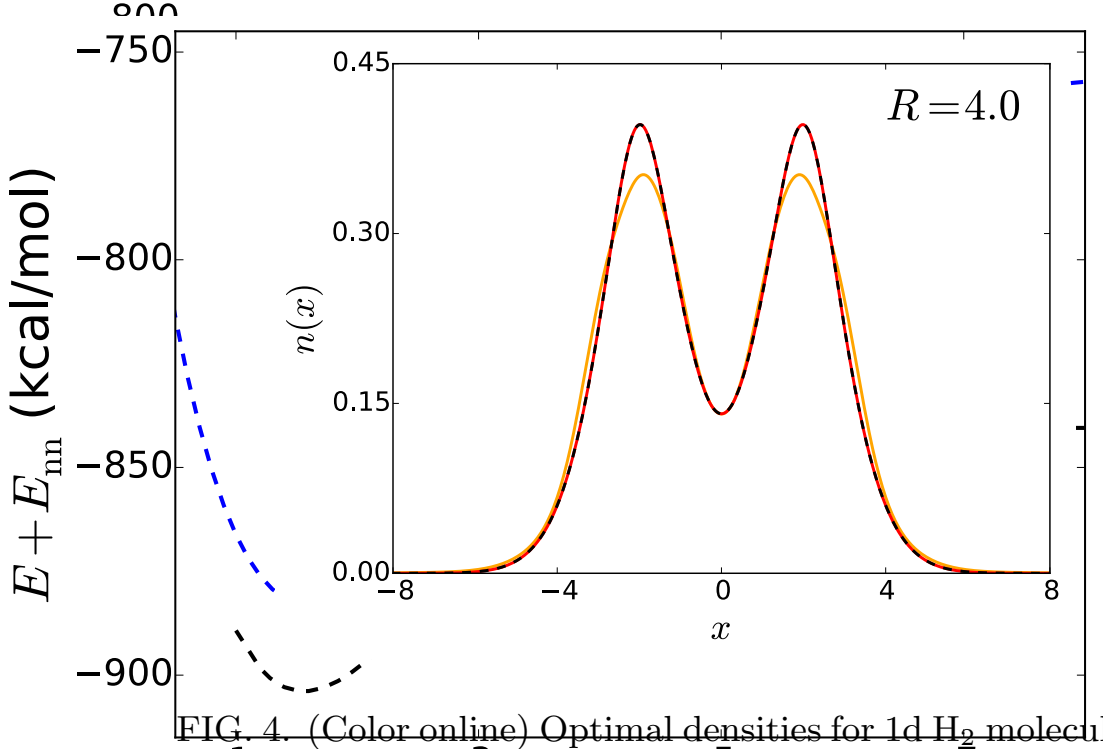


FIG. 4. (Color online) Optimal densities for 1d H_2 molecule in the test set: DMRG (black dashed), ML with N_T=5 (orange solid), ML with N_T=2 (red solid).

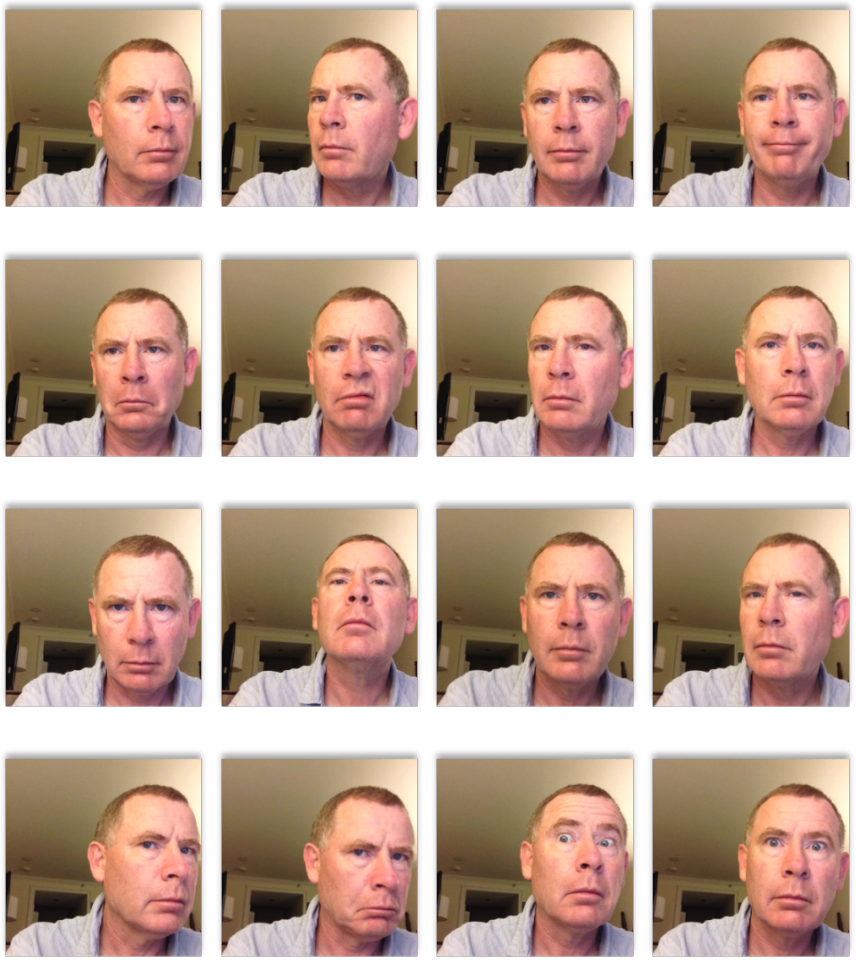
FIG. 3. (Color online) Same as Fig. 2. The green curves are ML with $N_T = 5$ on both the exact (dashed) and ML-optimized (solid) densities. The red solid curve is the ML

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

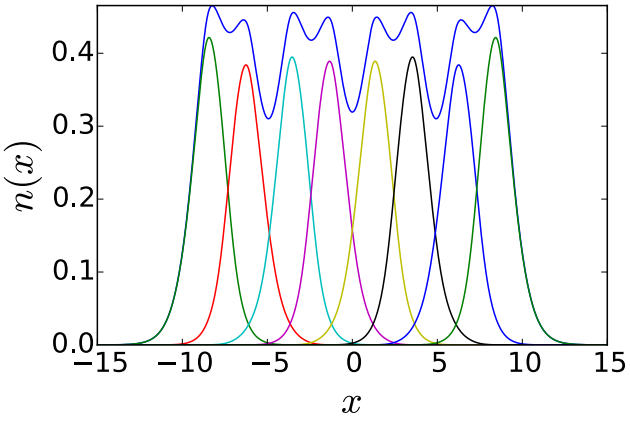


FIG. 5. Partition density of each H atom in H_8.

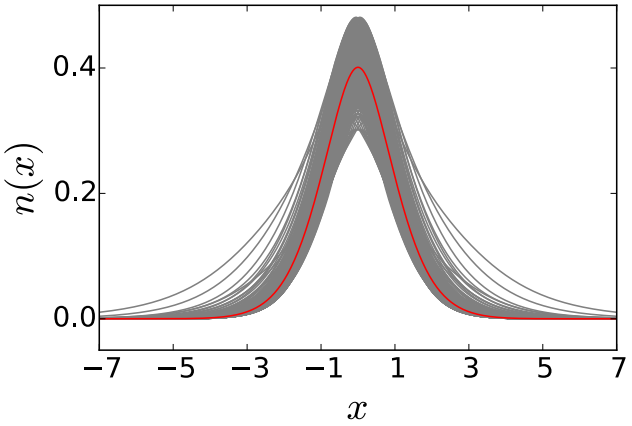


FIG. 6. Single H atom densities for H atoms in different chains and atomic distance (gray). The average density is plotted in red.

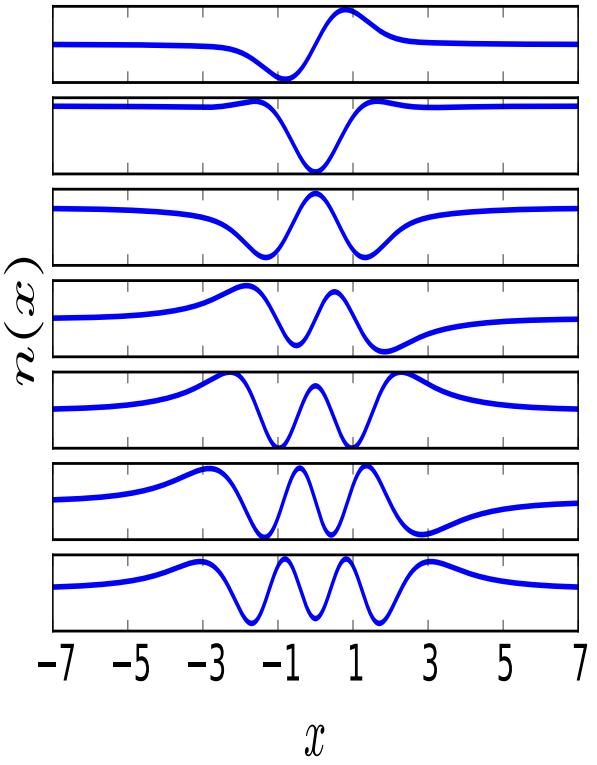


FIG. 7. First 7 principal components of the densities shown in Fig. 6, from top to bottom.

Improved convergence from basis

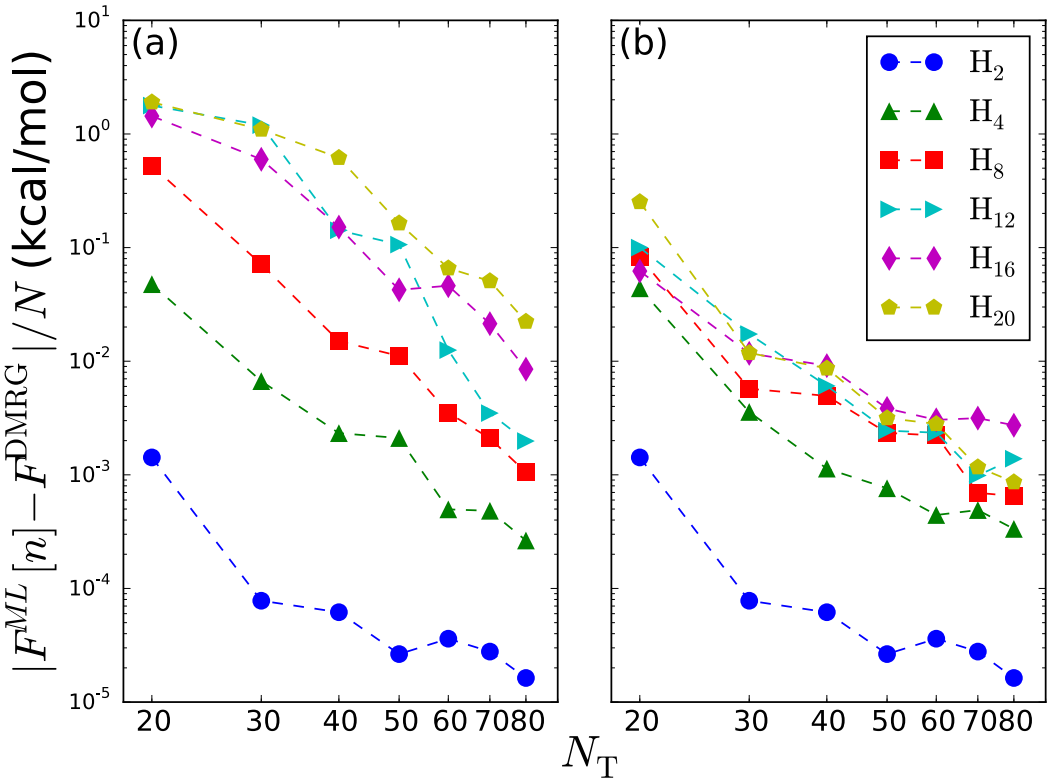
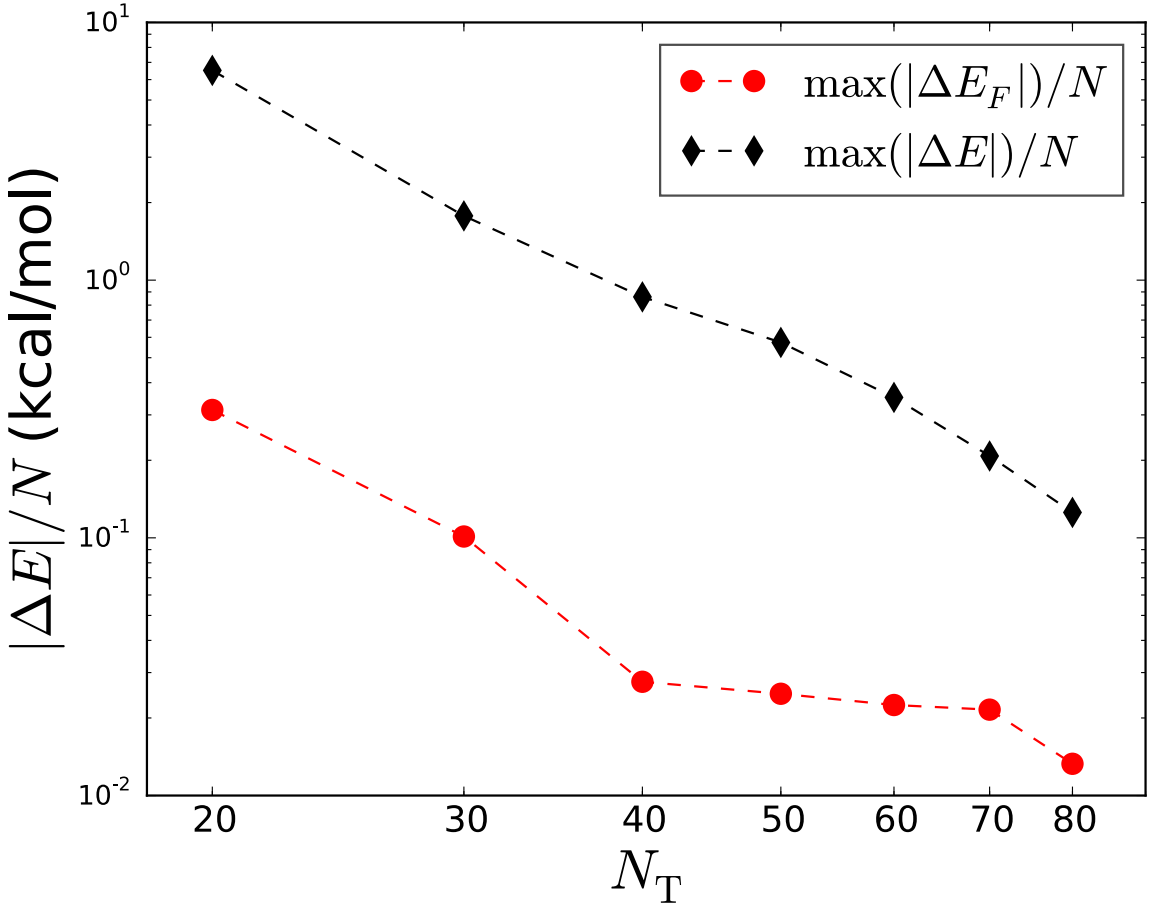
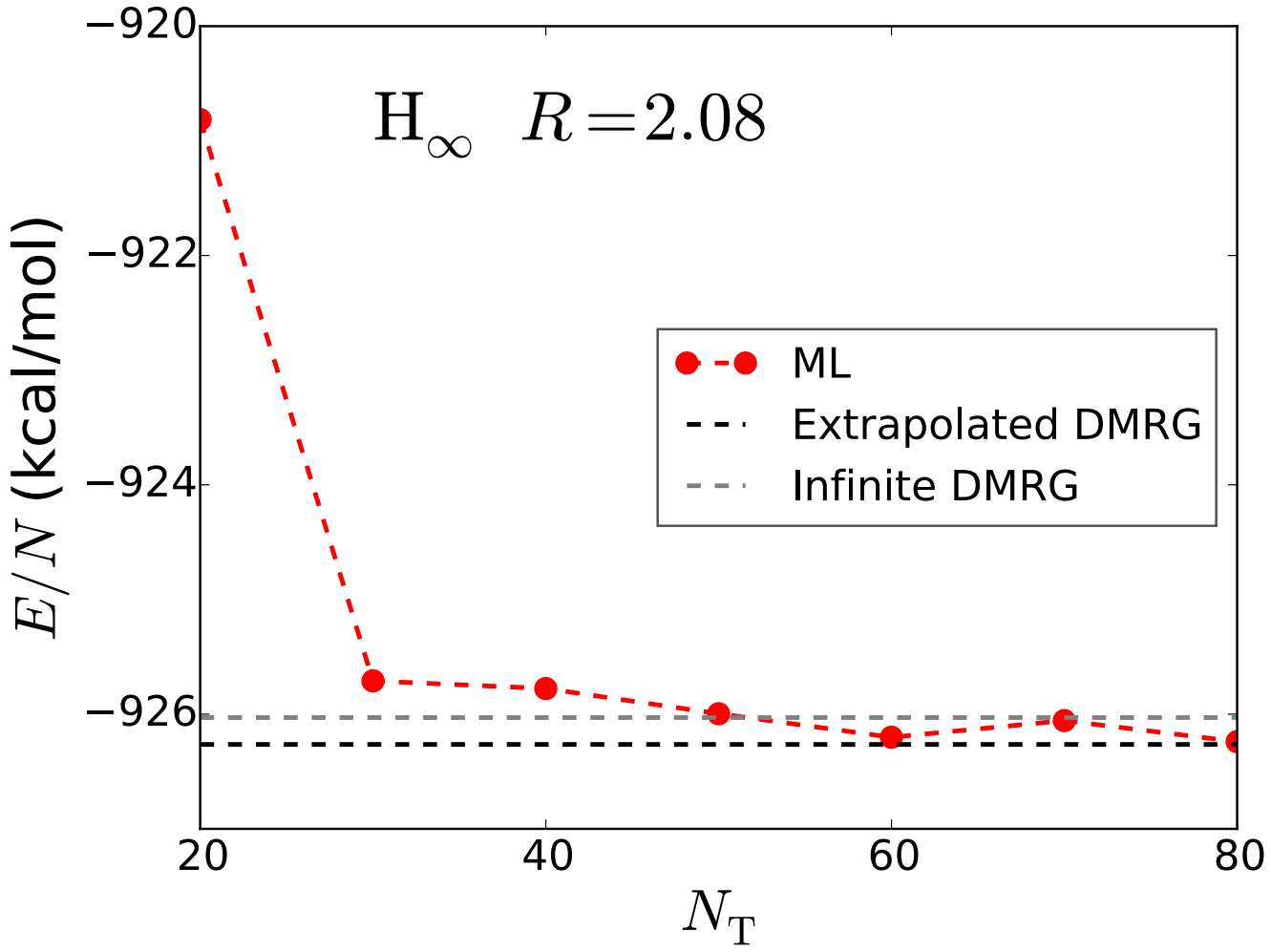


FIG. 8. (Color online) Learning curves for several 1d H chains. (a) ML using the total density. (b) ML using the bulk partition densities (see text).

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.

Papers (all on dft.uci.edu)

Nonlinear gradient denoising: Finding accurate extrema from inaccurate functional derivatives John C. Snyder, Matthias Rupp, Klaus-Robert Müller, Kieron Burke, *International Journal of Quantum Chemistry* **115**, 1102--1114 (2015).

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).

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* n/a--n/a (2015).

Kernels, Pre-Images and Optimization John C. Snyder, Sebastian Mika, Kieron Burke, Klaus-Robert Müller, *Chapter in Empirical Inference - Festschrift in Honor of Vladimir N. Vapnik* (2013).

Orbital-free Bond Breaking via Machine Learning John C. Snyder, Matthias Rupp, Katja Hansen, Leo Blooston, Klaus-Robert Müller, Kieron Burke, *J. Chem. Phys.* **139**, 224104 (2013).

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

[175] By-passing the Kohn-Sham equations with machine learning Felix Brockherde, Li Li, Kieron Burke, Klaus-Robert M□uller, (to appear, *Nature Communications*, 2017).

[176] 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 (2016).

Abomination or breakthrough?

- ML functionals of this type are completely different from those we are used to.
- No way to check even simple conditions, such as positivity in domain of application.
- Working on incorporating exact conditions into ML functionals.
- They compliment the existing human-functionals, but do not replace them.
- Real test: Generality

Exact conditions in ML-DFT?

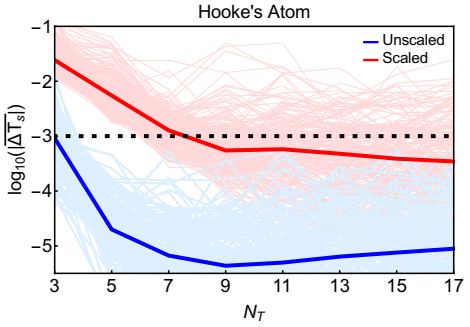


FIG. 2. (color online) The learning curves for functionals trained on scaled (blue) and unscaled (red) densities for the 1D Hooke's atom. Accuracy of 1 mH is denoted by the dashed line (black).

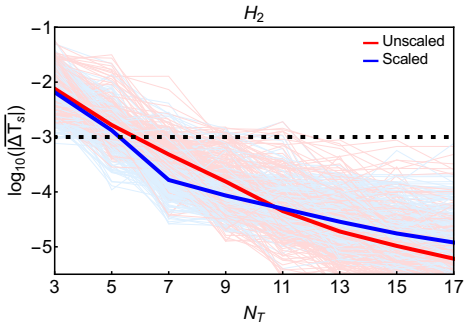


FIG. 3. (color online) Same as Fig. 2, but for H_2 densities.

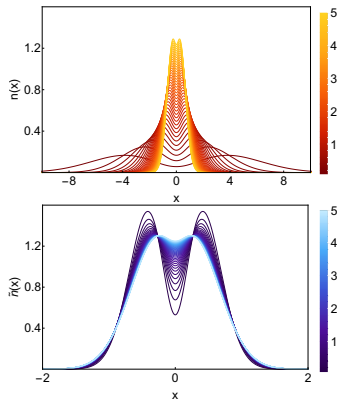


FIG. 4. (color online) The Hooke's atom densities with different ω are plotted before scaling (top) and after scaling (bottom). The color of each line indicates the value of ω. Notably, the densities appear "more similar" to each other after scaling.

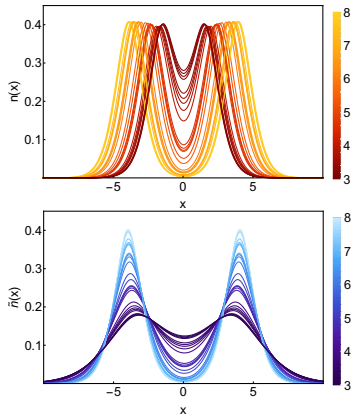


FIG. 7. (color online) The H_2 densities considered are plotted before scaling (top) and after scaling (bottom). The color each line indicates the separation, R. The densities do not appear to be made more similar by scaling.

$$T_S[n_\gamma] = \gamma^2\, T_S[n], \qquad n_\gamma(\mathbf{r}) = \gamma^3\, n(\gamma \mathbf{r})$$

Can exact conditions improve machine-learned density functionals?

Jacob Hollingsworth[1], Li Li (李力)[1], Thomas E. Baker[1,2], Kieron Burke[1,3]
[1] *Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA*
[2] *Institut quantique & Département de physique, Université de Sherbrooke, Québec, Canada J1K 2R1 and*
[3] *Department of Chemistry, University of California, Irvine, CA 92697, USA*

(Dated: Saturday 10th February, 2018)

See dft.uci.edu

Summary

- ML functionals can
  - find accurate densities
  - break bonds
  - Do the full functional for strongly correlated solids (in 1D)
  - Can now do MD of small molecules in 3D
- Thanks to
  - Students: Tom Baker, Li Li, John Snyder, Kevin Vu, Isabelle Pelaschier
  - Collaborators: Klaus Mueller, Matthias Rupp, Katia Hansen, Felix Brockherde, Leslie Vogt, Mark Tuckerman
  - Institute of Pure and Applied Math, UCLA
  - Funders: NSF from chem, DMR, math