Machine Learning in Electronic Structure: Finding Better Density Functionals than Humans do

Kieron Burke and friends

UC Irvine

Chemistry & Physics
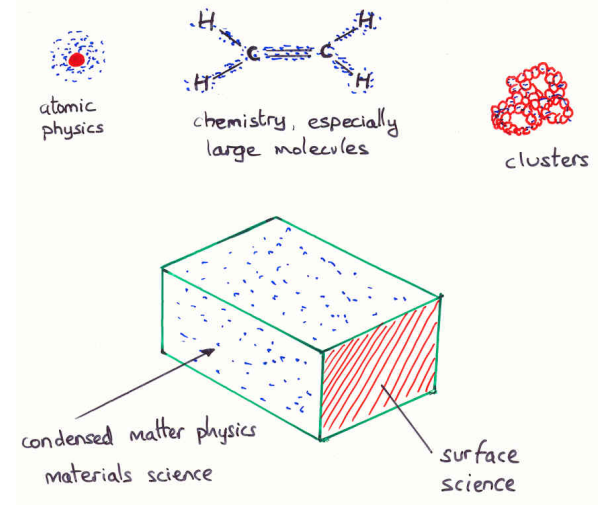
http://dft.uci.edu

Outline

1. Electronic structure of everyday matter

2. Proof of principle: Particle in a box

1. Two examples:
   a. Machine-learning of XC for strongly correlated solids (1D).
   b. Machine-learned KS kinetic energy of molecules (3D)

KS equations (1965)

Define *fictitious* non-interacting electrons satisfying:

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

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

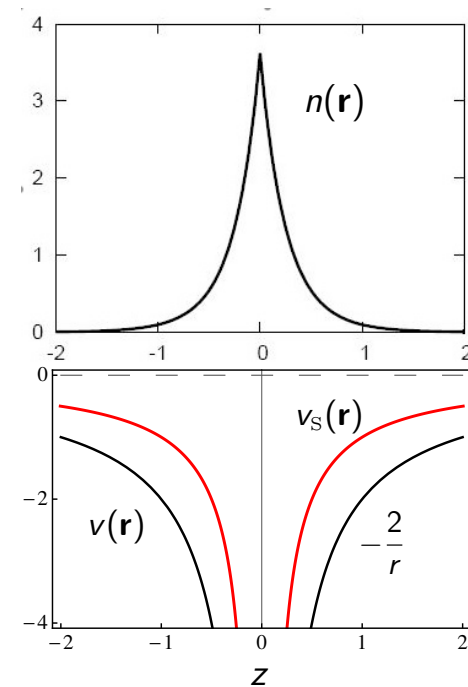Define T_{S} as the kinetic energy of the KS electrons, U as their Hartree energy and

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

the remainder is the exchange-correlation energy.

Most important result of exact DFT:

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

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

DFT in a nutshell, Kieron Burke, Lucas O. Wagner, *Int. J. Quant. Chem.* **113**, 96-101 (2013).

The Hubbard dimer: a density functional case study of a many-body problem D J Carrascal, J Ferrer, J C Smith, K Burke, *J Physics: Cond Mat* **27**, 393001 (2015)

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
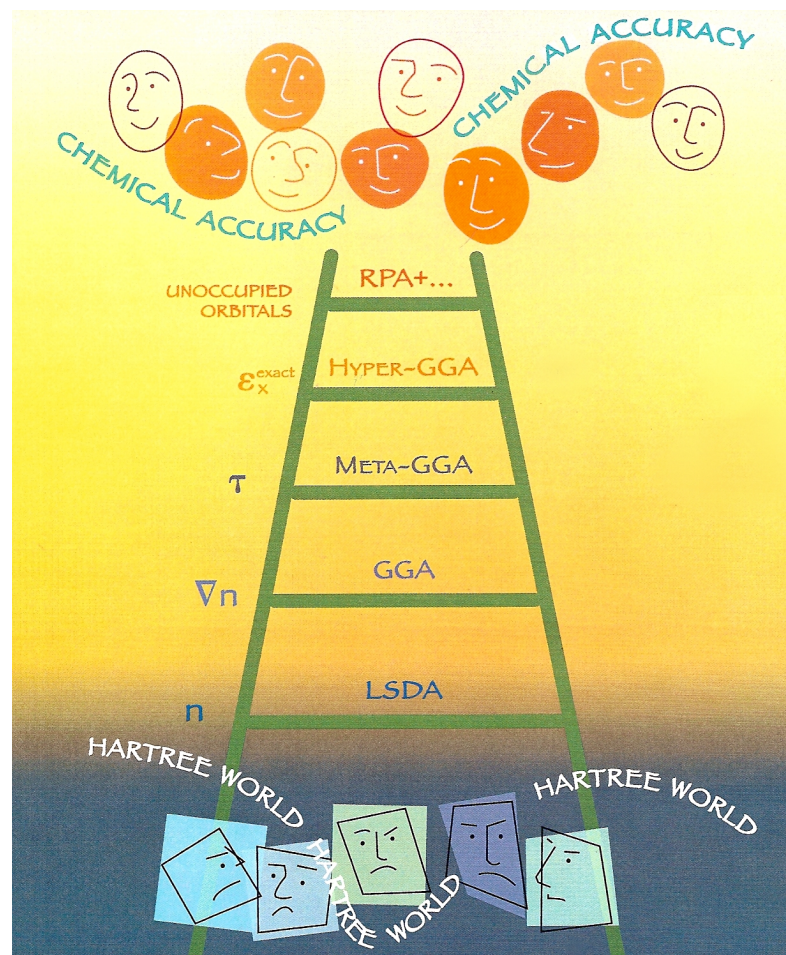
Fitting a round peg into a round hole: asymptotically correcting the generalized gradient approximation for correlation Antonio Cancio, Guo P. Chen, Brandon T. Krull and Kieron Burke, *The Journal of Chemical Physics* **149**, 084116 (2018)

Kier

A few recent applications

- Fertilizer
- Materials genome
- Hydrogen sulfide
- Airbook
- Juno
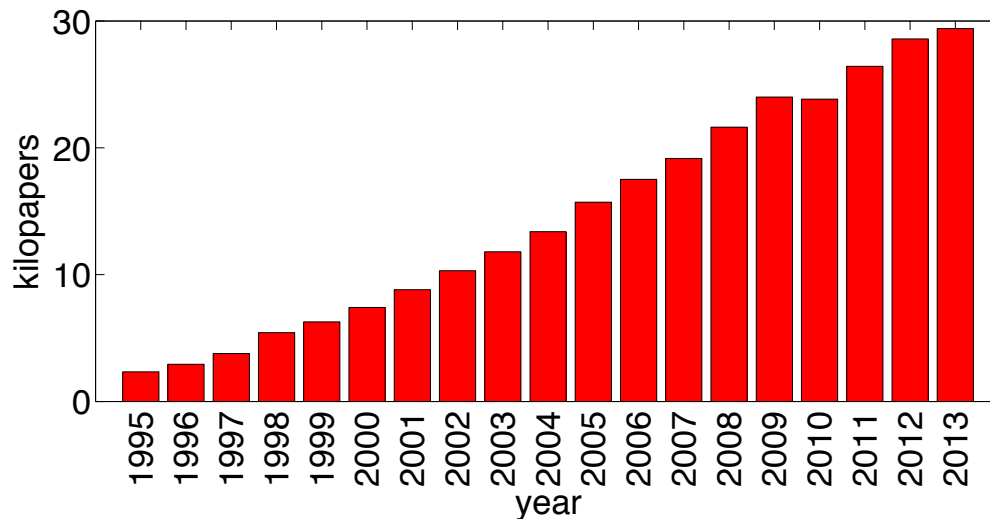
DFT: A Theory Full of Holes, Aurora Pribram-Jones, David A. Gross, Kieron Burke, Annual Review of Physical Chemistry (2014).

T_S versus E_{XC}

- Most DFT research focusses on XC

- But if we knew $T_S[n]$, we could bypass solving the KS equations.

- Known as orbital-free DFT.

- Semilocal approximations fail when electrons localized on more than one site.
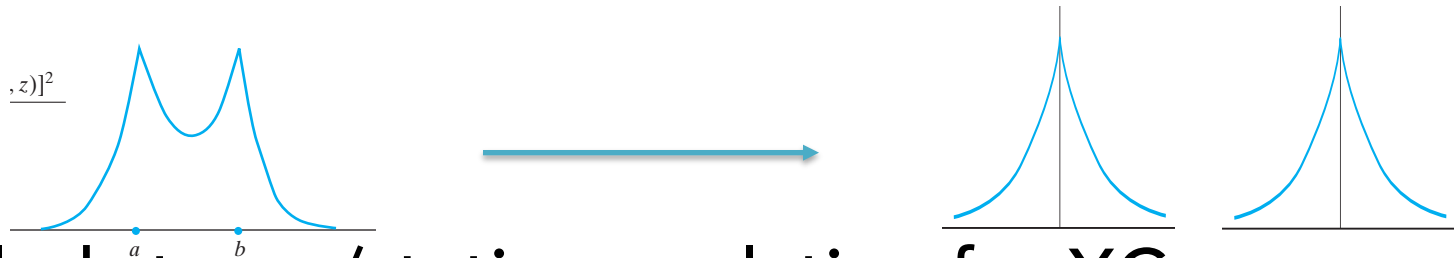
- Called strong/static correlation for XC

- Called self-interaction error for H_2^+

- Also happens for T_S

Machine learning in electronic structure

- 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
  - Accelerated sampling
  - Designing interatomic potentials

]**Editorial: Special Topic on Data-enabled Theoretical Chemistry** Matthias Rupp, O. Anatole von Lilienfeld, Kieron Burke, Journal of Chemical Physics Guest Editorial: Special Topic on Data-Enabled Theoretical Chemistry 148, 241401 (2018)

Original team for ML DFT (2010)

- 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
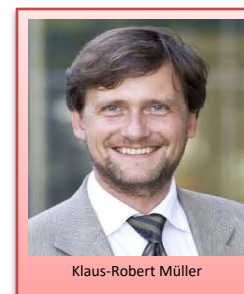
Kernel ridge regression

- 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

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

Demo problem in DFT

- *N* non-interacting same-spin fermions confined to 1d box

- Define class of potential:

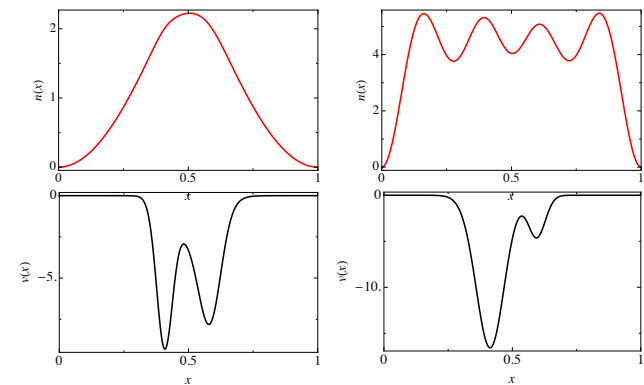$$v(x) = -\sum_{i=1}^{3} a_i \exp(-(x - b_i)^2/(2c_i^2))$$

- 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 (-∫ dx (n(x)-n'(x))²)/(2σ²))

Generate 2000 potentials. Solve for up to 4 electrons.

Performance for T_s

N	M	λ	σ	$\overline{|\Delta T|}$	$|\Delta T|^{\mathrm{std}}$	$|\Delta T|^{\mathrm{max}}$
				kcal/mol		
	40	2.4×10^{-5}	238	3.3	3.0	23.
	60	1.0×10^{-5}	95	1.2	1.2	10.
1	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

We don't just need the energy

- 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.
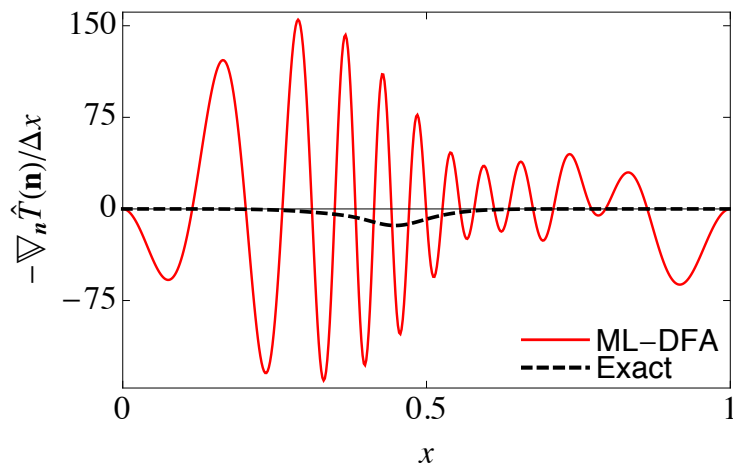
functional derivative?

Exact   ML-DFA

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

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

- 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

$$X = \left(\mathbf{n}_{j_1} - \mathbf{n}, \ldots, \mathbf{n}_{j_m} - \mathbf{n} \right)^{\top}$$
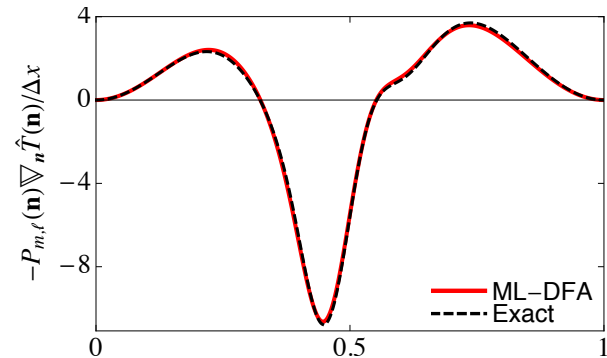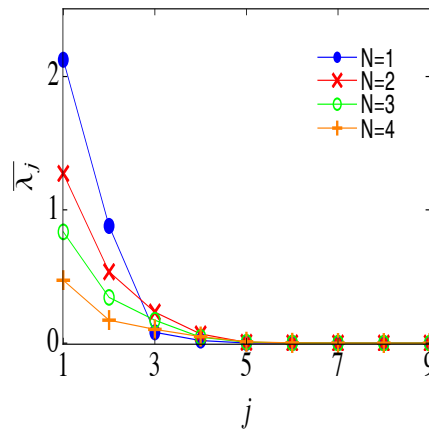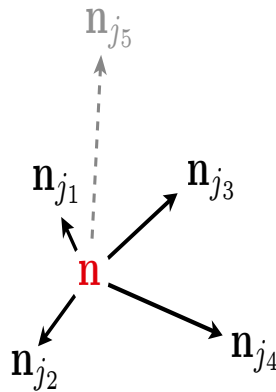
$$C = \frac{1}{m} X^{\top} X$$

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

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

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

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)

Strong correlation and 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.

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).

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

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.

Facial recognition via PCA
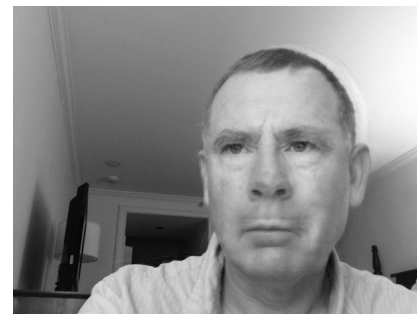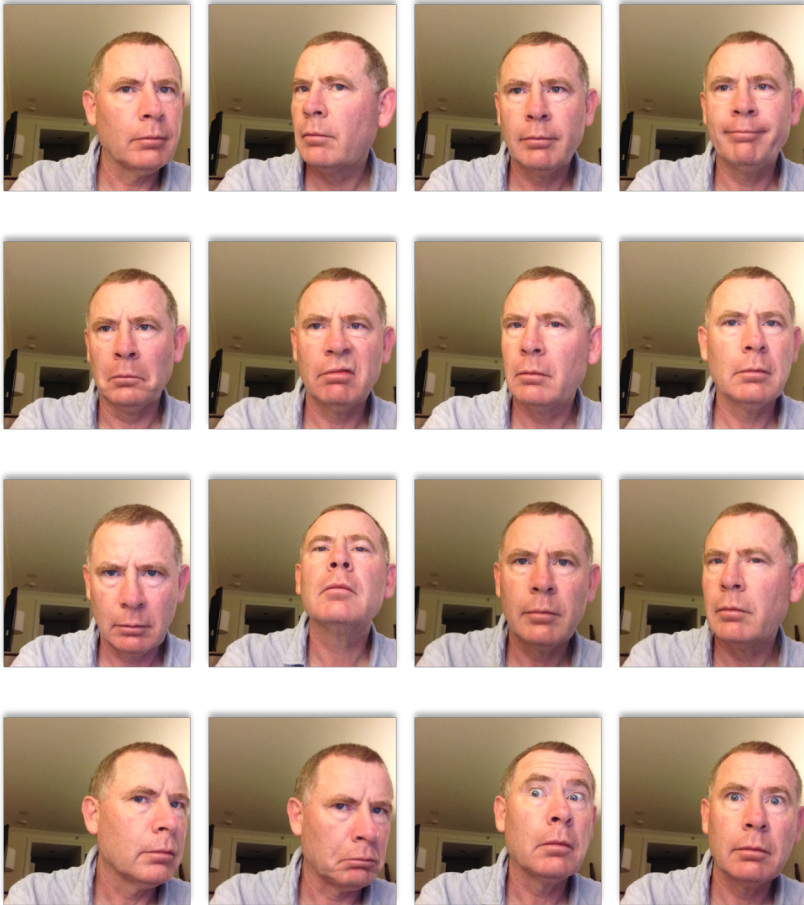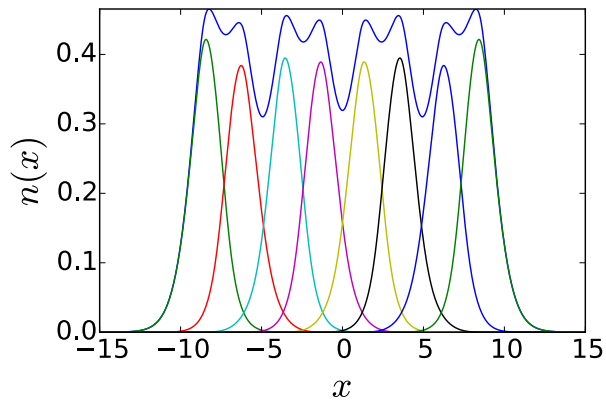
PCA basis for atomic densities

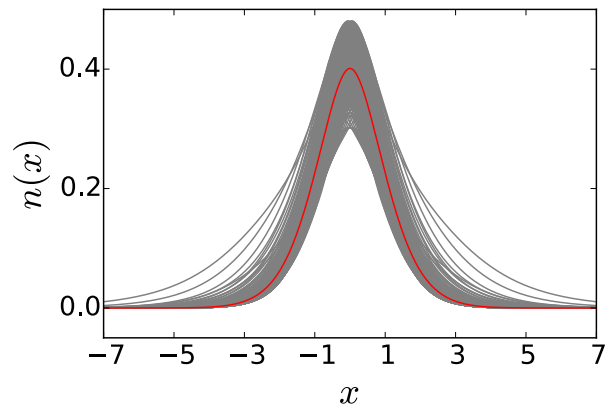FIG. 5. Partition density of each H atom in H_8.

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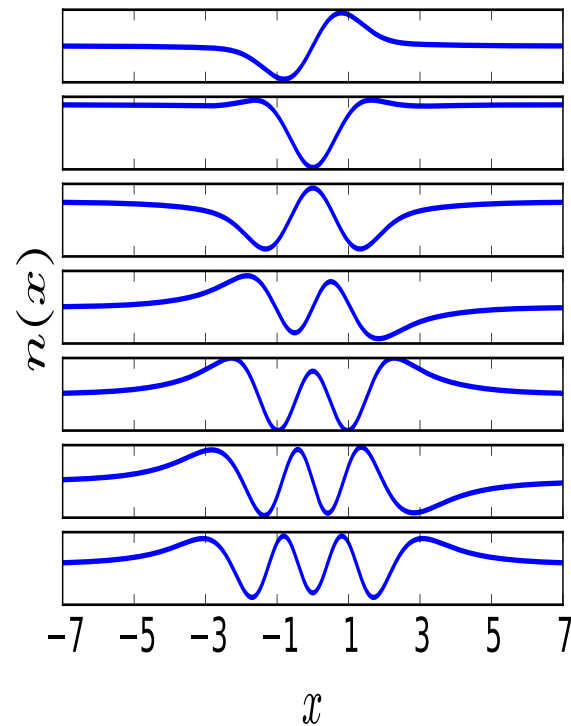

FIG. 7. First 7 principal components of the densities shown in Fig. 6, from top to bottom.
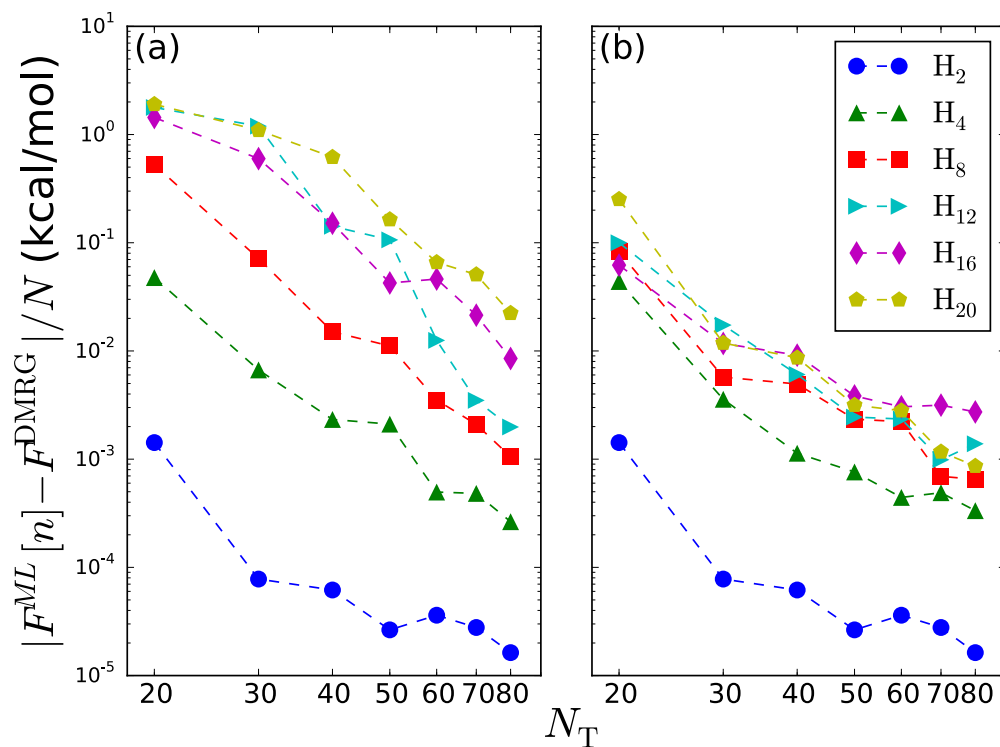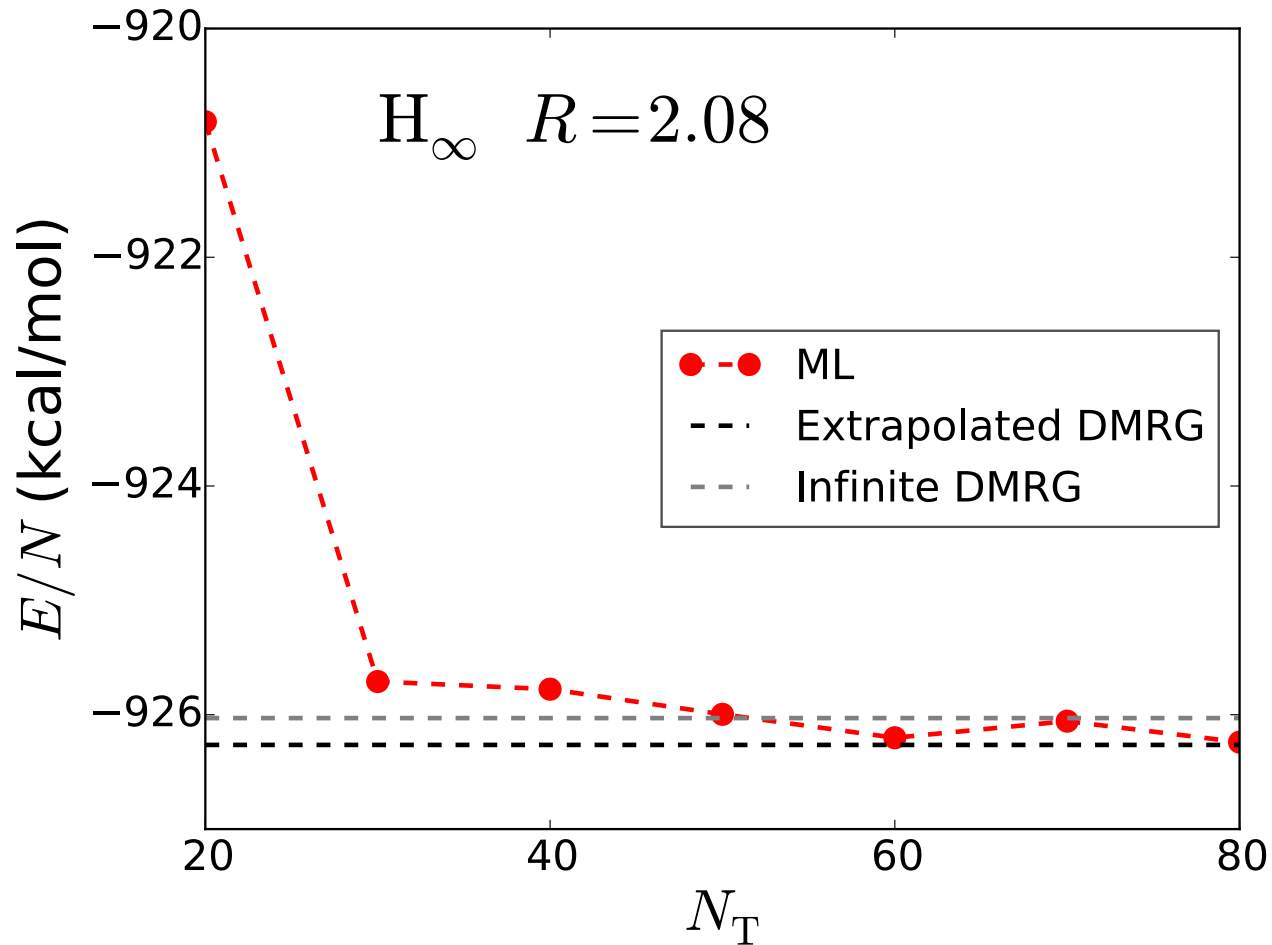
Improved convergence from basis

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).

Convergence for infinite chain

$H_\infty \quad R = 2.08$

Legend:
- ML
- Extrapolated DMRG
- Infinite DMRG

Axes: E/N (kcal/mol) vs N_T
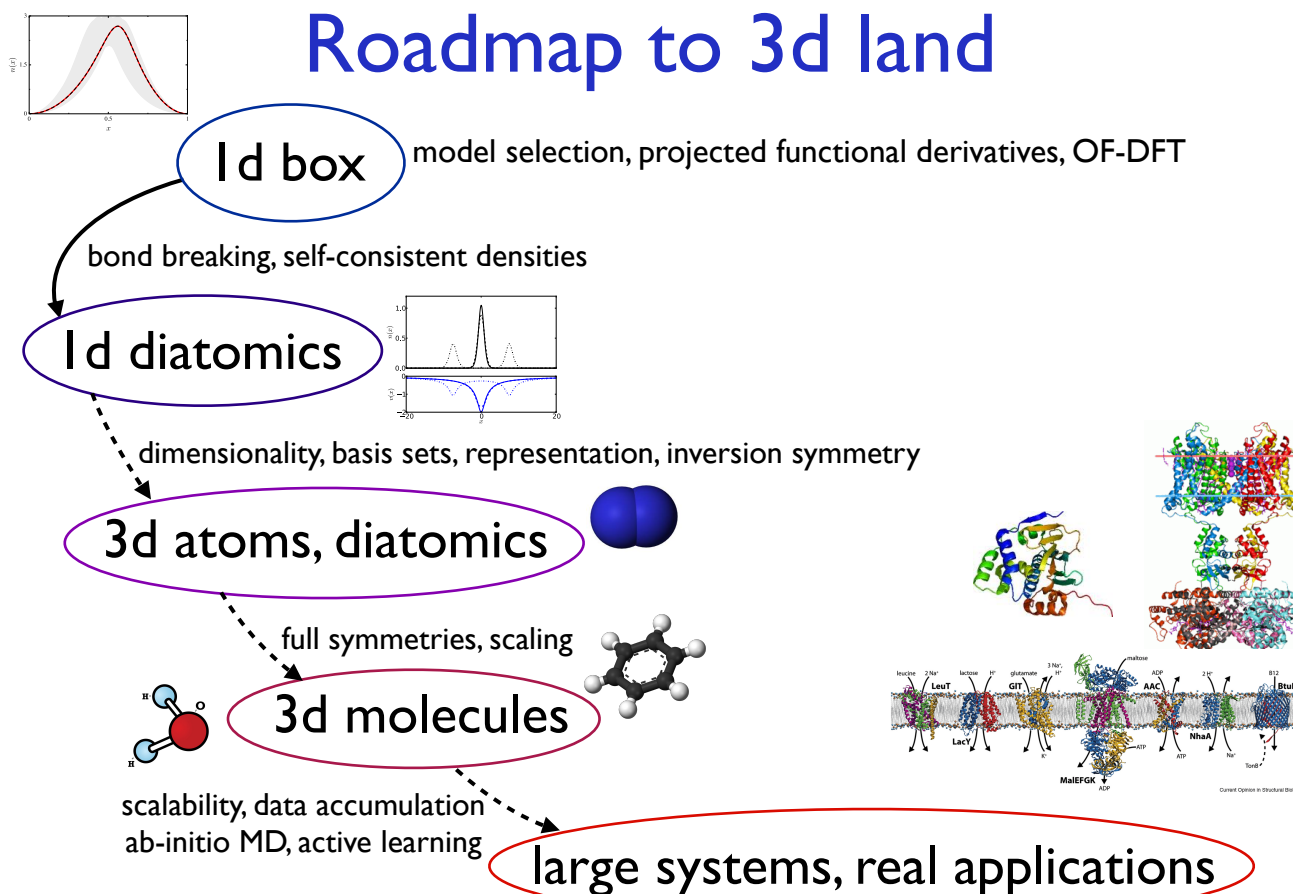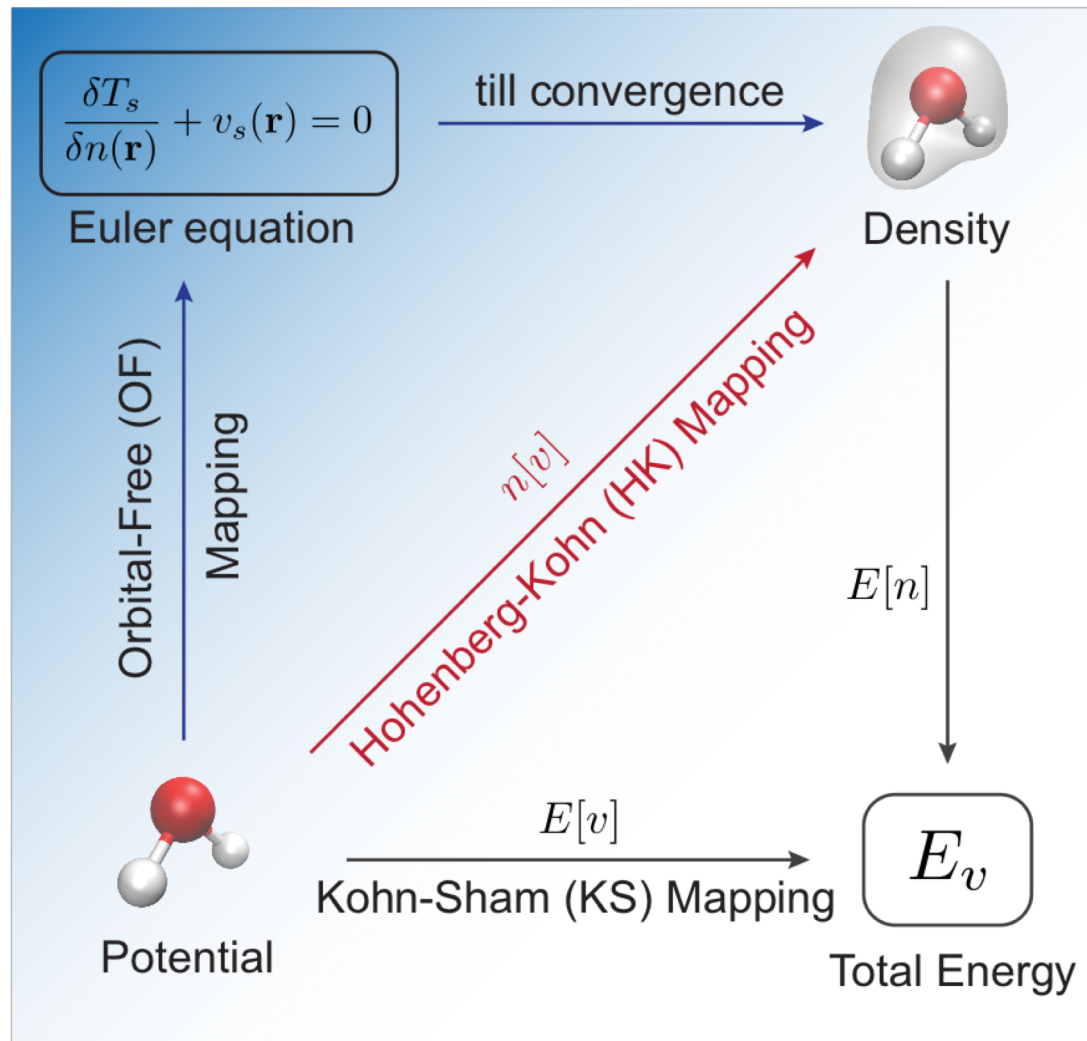
Lessons from this part

- 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.

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

By-passing KS
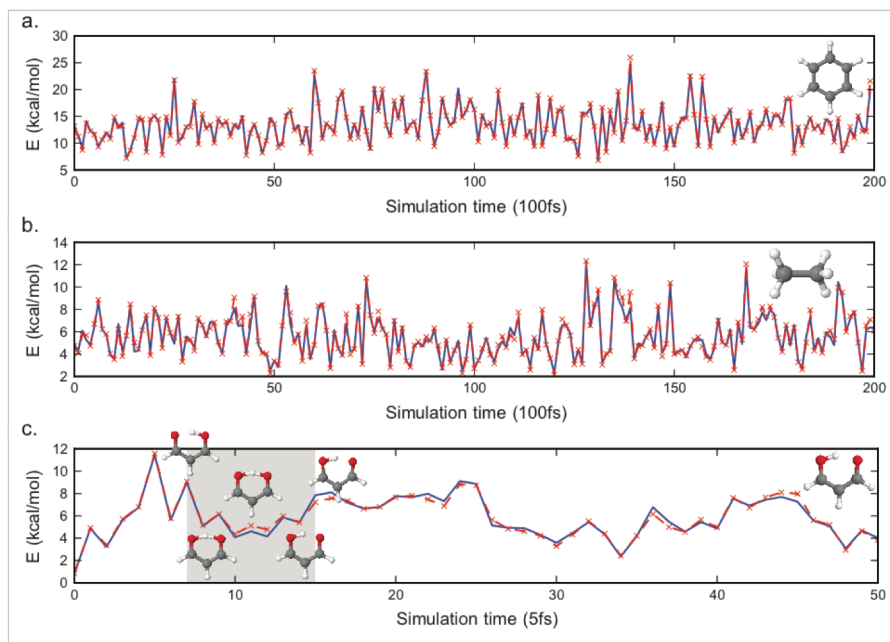
MD simulations testing ML method

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.

Training trajectories	Benzene		Ethane		Malonaldehyde	
	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.

Malonaldehyde
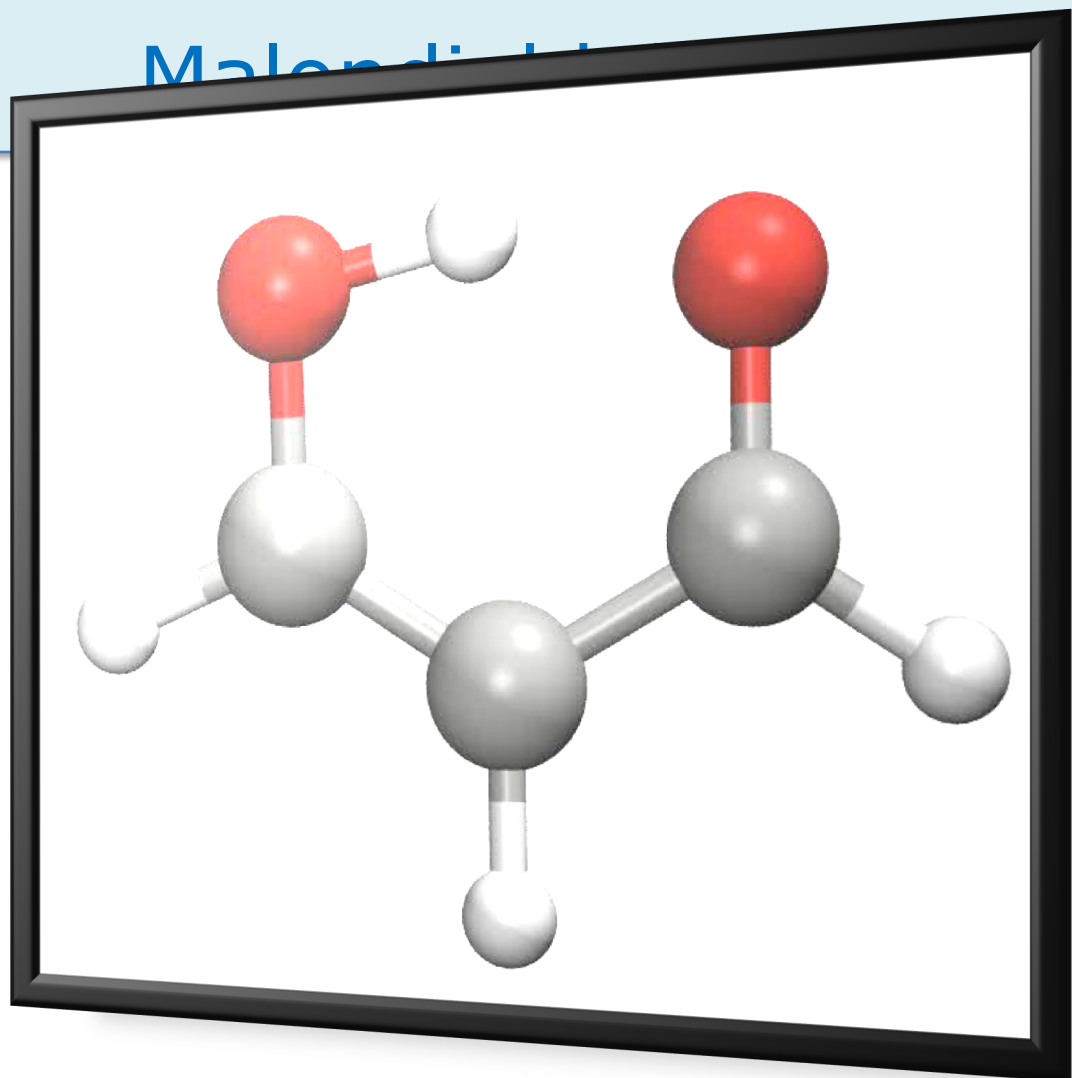
[174] **By-passing the Kohn-Sham equations with machine learning** Felix Brockherde, Leslie Vogt, Li Li, Mark E Tuckerman, Kieron Burke, Klaus-Robert Müller, Nature Communications 8, 872 (2017).

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.

Our 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).

Bypassing the Kohn-Sham equations with machine learning Felix Brockherde, Leslie Vogt, Li Li, Mark E Tuckerman, Kieron Burke, Klaus-Robert Müller, Nature Communications 8, 872 (2017).

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).

Can exact conditions improve machine-learned density functionals? Jacob Hollingsworth, Li Li, Thomas E. Baker, Kieron Burke, The Journal of Chemical Physics 148, 241743 (2018).

Points for this conference

- Electronic structure is (largely) deterministic and based on a Hamiltonian

- DFT is not mean-field

- HK theorem is a statement about the minimal information needed to identify the system

- How do very simple formulas 'solve' a quantum many-fermion problem?

- Mermin theorem

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