Abstract: Density functional theory (DFT) is an extremely popular approach to electronic structure problems in both materials science and chemistry and many other fields. Over the past several years, often in collaboration with Klaus Mueller at TU Berlin, we have explored using machine-learning to find the density functionals that must be approximated in DFT calculations. I will summarize our results so far, and report on two new works.
Finding density functionals with ML

Kieron Burke and friends
UC Irvine Physics & Chemistry

http://dft.uci.edu
Outline

- Review of density functional theory (DFT)
- ML for finding functionals for box problems
- ML for bond breaking
- Latest results in 3D and for $E_{XC}$
- Summary
The electronic structure problem

- Use atomic units
- Born-Oppenheimer approximation
- All non-relativistic (but can be added back in)
- Wavefunctions antisymmetric and normalized
- Only discuss ground-state electronic problem here, but many variations.

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

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

where the kinetic and elec-elec repulsion energies are

$$\hat{T} = -\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2,$$

$$\hat{V}_{ee} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{1}{|r_i - r_j|},$$

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

$$\hat{V} = \sum_{i=1}^{N} v(r_i)$$

Often $v(r)$ is electron-nucleus attraction

$$v(r) = -\sum_{\alpha} \frac{Z_{\alpha}}{|r - R_{\alpha}|}$$

where $\alpha$ runs over all nuclei, plus weak applied $E$ and $B$ fields.

$$\{ \hat{T} + \hat{V}_{ee} + \hat{V} \} \psi = E \psi, \quad E = \min_{\psi} \langle \psi | \hat{T} + \hat{V}_{ee} + \hat{V} | \psi \rangle$$
HK theorem (1964)

- Makes TF an approximation to an exact theory
- Can find both ground-state density and energy via Euler equation

\[ E = \min_\psi \langle \psi | \hat{T} + \hat{V}_{ee} + \hat{V} | \psi \rangle \]
\[ = \min_n \left\{ F[n] + \int d^3r \, v(r)n(r) \right\} \]

where
\[ F[n] = \min_{\psi \to n} \langle \psi | \hat{T} + \hat{V}_{ee} | \psi \rangle \]

- The minimum is taken over all positive \( n(r) \) such that \( \int d^3r \, n(r) = N \)
- The external potential \( v(r) \) and the Hamiltonian \( \hat{H} \) are determined to within an additive constant by \( n(r) \)

KS DFT method (1965)

1964: HK theorem: There exists $F[n]$

Define *fictitious* non-interacting electrons satisfying:

\[
\left\{ -\frac{1}{2} \nabla^2 + v_s(r) \right\} \phi_j(r) = \epsilon_j \phi_j(r), \quad \sum_{j=1}^{N} |\phi_j(r)|^2 = n(r).
\]

where $v_s(r)$ is defined to yield $n(r)$.

Define $T_s$ as the kinetic energy of the KS electrons, $U$ as their Hartree energy and

\[F = T + V_{ee} = T_s + U + E_{XC}\]

the remainder is the exchange-correlation energy.

Most important result of exact DFT:

\[v_s(r) = v(r) + \int d^3r \frac{n(r')}{|r - r'|} + v_{xc}[n](r), \quad v_{xc}(r) = \frac{\delta E_{xc}}{\delta n(r)}\]

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

**Orbital-free DFT:** Approximate $T_s[n]$ directly, and go much, much faster.
Today’s commonly-used functionals

- **Local density approximation (LDA)**
  - Uses only $n(r)$ at a point.
  
  \[
  E_{X}^{\text{LDA}}[n] = A_{X} \int d^{3}r \ n^{4/3}(r)
  \]
  \[
  A_{X} = -(3/4)(3/\pi)^{1/3} = -0.738.
  \]

- **Generalized gradient approx (GGA)**
  - Uses both $n(r)$ and $|\nabla n(r)|$
  - Should be more accurate, corrects overbinding of LDA
  - Examples are PBE and BLYP

- **Hybrid:**
  - Mixes some fraction of HF with GGA
  - Examples are B3LYP and PBE0

\[
E_{XC}^{\text{hyb}} = a \left( E_{X} - E_{X}^{\text{GGA}} \right) + E_{XC}^{\text{GGA}}
\]
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,...
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,..
- Gigo principle: Results only as good as $E_{XC}[n]$ approximation used.
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,..
- Gigo principle: Results only as good as $E_{XC}[n]$ approximation used.
- Much faster than wavefunction/green’s function methods for large systems.
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,..
- Gigo principle: Results only as good as $E_{\text{XC}}[\rho]$ approximation used.
- Much faster than wavefunction/green’s function methods for large systems.
- Never fast enough for e.g., biomolecular MD simulations.
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,..
- Gigo principle: Results only as good as $E_{XC}[n]$ approximation used.
- Much faster than wavefunction/green's function methods for large systems.
- Never fast enough for e.g., biomolecular MD simulations.
- Often inaccurate for strongly correlated systems, including many energy-related materials.
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,..
- Gigo principle: Results only as good as $E_{\text{XC}}[n]$ approximation used.
- Much faster than wavefunction/green’s function methods for large systems.
- Never fast enough for e.g., biomolecular MD simulations.
- Often inaccurate for strongly correlated systems, including many energy-related materials.
- Catalysts found by DFT being patented and used.
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,..
- Gigo principle: Results only as good as $E_{XC}[n]$ approximation used.
- Much faster than wavefunction/green’s function methods for large systems.
- Never fast enough for e.g., biomolecular MD simulations.
- Often inaccurate for strongly correlated systems, including many energy-related materials.
- Catalysts found by DFT being patented and used.
- High pressure 200K superconductors predicted from DFT calculations.
Importance of DFT

- Spans many fields: chemistry, materials science, condensed-matter physics,..
- Gigo principle: Results only as good as $E_{XC}[n]$ approximation used.
- Much faster than wavefunction/green’s function methods for large systems.
- Never fast enough for e.g., biomolecular MD simulations.
- Often inaccurate for strongly correlated systems, including many energy-related materials.
- Catalysts found by DFT being patented and used.
- High pressure 200K superconductors predicted from DFT calculations.
- Last year, 30,000 scientific papers published using DFT.

Axel Becke: 2016 Killam Prize in Natural Sciences
Systematic approach to DFT

- Lieb and Simon proved TF is leading order in an unusual semiclassical limit.
- For atoms, this is same as keeping neutral and taking N to infinity.
- With collaborators, I have shown LDA is leading term in same limit.
- In model cases, can find leading corrections, which are uniform asymptotic expansions in \( \hbar \) and far more accurate than present-day DFT approximations.
Orbital-free DFT

- Long-time dream of electronic structure
- If you know $T_s[n]$ sufficiently accurately, you avoid computational cost of solving KS equations.
- Like TF, only accurate enough for prediction.
- Go from hundreds to millions of atoms
- See work by Emily Carter and Sam Trickey
- Also, extremely important in plasma physics simulations at million K scale
Strong correlation: Where DFT fails

- Ongoing project with Steve White at UCI
- Apply DMRG to continuum problems
- Understand limitations and failures of standard DFT approximations
B. Machine learning- demo
ML applications in electronic structure

- Most with Klaus Mueller of TU Berlin, computer science.
ML applications in electronic structure

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

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

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

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

- Our efforts are focused on finding $T_s[n]$ from examples, work by John Snyder (Humboldt fellow at TU Berlin/MPI Halle)
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:
  \[ \hat{T}(n) = \hat{T} \sum_{j=1}^{M} \alpha_j k(n_j, n) \]
Dataset

Generate 2000 potentials. Solve for up to 4 electrons.
### Performance for $T_s$

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

LDA $\sim$ 223 kcal/mol, Gradient correction $\sim$ 159 kcal/mol
**Functional derivative?**

- 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?

\[
\frac{\delta T[n]}{\delta n(x)} = \mu - n(x) \quad \leftrightarrow \quad \frac{1}{\Delta x} \nabla_n \hat{T}(n) = \sum_{j=1}^{M} \alpha'_j (n_j - n) k(n_j, n)
\]

\[
\alpha'_j = \frac{\alpha_j}{(\sigma^2 \Delta x)}
\]

Aug 11, 2016

Perimeter Institute
Principal component analysis

\[ X = (n_{j_1} - n, \ldots, n_{j_m} - n)^T \]

\[ C = \frac{1}{m} X^\top X \]

\[ \lambda_j, x_j \]

\[ P_{m,\ell}(n) = V^\top V \]

\[ V = (x_1, \ldots, x_{\ell})^T \]
Projected functional derivative

\[-P_{m,l}(n) \nabla_n \hat{T}(n)/\Delta x\]

Aug 11, 2016
Perimeter Institute
Constrained optimized density

- Gradient descent search:

\[ n^{(j+1)} = n^{(j)} - \epsilon P_{m,\ell}(n^{(j)}) \left( v + \nabla_n \hat{T}(n^{(j)}) / \Delta x \right) \]

\begin{align*}
N &= 1 \\
M &= 100
\end{align*}

\begin{align*}
\Delta T_j &= 0.154 \text{ kcal/mol} \\
\Delta T_j^{\text{sc}} &= 6.53 \text{ kcal/mol} \\
\text{Ratio} &= 43
\end{align*}

\begin{align*}
m &= 15, \ell = 5
\end{align*}
Bond-breaking with ML

- Standard $T_s[n]$ approximations like TF go bad when bonds break.
- 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.

Constrained optimal density

- Convergence of constrained optimal density with # of training points.

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

FIG. 7. Difference between the constrained optimal density $\tilde{n}(x)$ and the KS density $n(x)$ for various numbers of training densities $N_T$. The error decreases uniformly for all $x$. The system is $H_2$ at equilibrium bond length. The inset shows the KS density.

Aug 11, 2016

Perimeter Institute
Bond-breaking with ML

- Standard $T_s[n]$ approximations like TF go bad when bonds break.
- 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

Constrained optimal density

- Convergence of constrained optimal density with # of training points.

FIG. 7. Difference between the constrained optimal density $\tilde{n}(x)$ and the KS density $n(x)$ for various numbers of training densities $N_T$. The error decreases uniformly for all $x$. The system is H$_2$ at equilibrium bond length. The inset shows the KS density.
Types of errors in DFT

- $\Delta E_F = \tilde{E}_{xc}[n] - E_{xc}[n]$
- $\Delta E_D = \tilde{E}_{xc}[^\bar{n}] - \tilde{E}_{xc}[n]$
- $\Delta E = \Delta E_F + \Delta E_D$

Error analysis of energies in kcal/mol as a function of $R$ with different numbers of training data, on constrained optimal densities.

FIG. 8. The total error of the model and the functional- and density-driven errors $\Delta E_F$ and $\Delta E_D$ for $H_2$ with (a) 10 and (b) 20 training densities.
Functional derivatives and densities

- How can we get accurate densities from lousy derivatives?
- Once solution density is within interpolation manifold, simply constrain derivative to stay on that manifold
- Analogy:
  - Problem: find global minimum of 2D surface, given exact data along a 1D curve in that surface that passes through the minimum.
  - Solution: Make sure you stay on the path.
- PS: Inspired density-corrected DFT, which corrects many self-interaction errors!


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
- scalability, data accumulation
- ab-initio MD, active learning

3d molecules

large systems, real applications

Aug 11, 2016
Perimeter Institute
D. Recent results

- Either submitted or about to be.

Felix Brockherde

Li Li

Thomas E. Baker
2 new papers

• **By-passing the KS equations with ML**
  – 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**
  – Li Li, Thomas E. Baker, Steven R. White and KB
  – Do interacting functional (ie. Exact Exc)
  – Do strong correlation
  – Do thermodynamic limit
  – Still in 1d
By-passing KS

\[ \frac{\partial T_s}{\partial n(r)} + v_s(r) = 0 \]  
Euler equation
till convergence

Density

Hohenberg-Kohn (HK) Mapping

Orbital-Free (OF) Mapping

Kohn-Sham (KS) Mapping

Potential

Total Energy

Aug 11, 2016
Perimeter Institute
Convergence of different HK maps

\[
\Delta E_D \text{ (kcal/mol)}
\]

Number of training points \( M \)

- OF
- HK Grid
- HK Fourier
- HK KPCA
Non-interacting HK map

Potential → Potential as Gaussian blobs → Independent ML models → Data-driven and physically motivated basis representations → Density

Aug 11, 2016

Perimeter Institute
Error for $H_2$
ML of exact functionals

- Use DMRG to solve continuum problems in 1d.
- Much success in past, showing failures of DFT approximations for strong correlation.
- Here use DMRG to generate much data of exact densities and energies
- All restricted to 1d.
- We train and test a machine learning $F[n]$, the universal part of the electronic density functional, to within quantum chemical accuracy. We (a) bypass the standard Kohn-Sham approach, (b) include the strong correlation of highly-stretched bonds and (c) create a model for the infinite chain limit.

Guaranteed Convergence of the Kohn-Sham Equations

One-Dimensional Continuum Electronic Structure with the Density-Matrix Renormalization Group and Its Implication for Density-Functional Theory
Convergence for $H_2$

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. Red curves are the same with $N_T = 20$. 

Aug 11, 2016

Perimeter Institute
FIG. 4. (Color online) Optimal densities for 1d H$_2$ molecule in the test set: DMRG (black), ML with $N_T = 5$ (orange), ML with $N_T = 20$ (red).
FIG. 5. Partition density of each H atom in H₈.
PCA basis for atomic densities

\[ n_i(R, x) = f_0(x) + \sum_{p=1}^{7} c_{i,p}(R)f_p(x). \]
Improved convergence from basis
FIG. 9. (Color online) Electronic energy per atom in the thermodynamic limit, both via DMRG chains (extrapolated to infinity) and using machine learning with 50 data points per chain.
Convergence for infinite chain

FIG. 1. (Color online) Electronic energy of infinite chain from model learned from extrapolated chain densities and energies.


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).


Summary

- ML of functionals works in model cases to produce highly accurate approximate functionals
Summary

- ML of functionals works in model cases to produce highly accurate approximate functionals
- Totally different approach from anything in DFT before
Summary

- ML of functionals works in model cases to produce highly accurate approximate functionals
- Totally different approach from anything in DFT before
- ML can even
  - find accurate densities
  - say when it will work within tolerance (makes Klaus nervous)
  - break bonds
  - Do the full functional
Summary

- ML of functionals works in model cases to produce highly accurate approximate functionals
- Totally different approach from anything in DFT before
- ML can even
  - find accurate densities
  - say when it will work within tolerance (makes Klaus nervous)
  - break bonds
  - Do the full functional
- But
  - only demonstrated in 1d
  - Need to do arbitrary-sized system (representation question)
Summary

- ML of functionals works in model cases to produce highly accurate approximate functionals
- Totally different approach from anything in DFT before
- ML can even
  - find accurate densities
  - say when it will work within tolerance (makes Klaus nervous)
  - break bonds
  - Do the full functional
- But
  - only demonstrated in 1d
  - Need to do arbitrary-sized system (representation question)
- Ongoing projects:
  - ML exact interacting functional in thermodynamic limit
  - Full 3D calculation of water molecule, using ML for $n[v](r)$
Summary

- ML of functionals works in model cases to produce highly accurate approximate functionals
- Totally different approach from anything in DFT before
- ML can even
  - find accurate densities
  - say when it will work within tolerance (makes Klaus nervous)
  - break bonds
  - Do the full functional
- But
  - only demonstrated in 1d
  - Need to do arbitrary-sized system (representation question)
- Ongoing projects:
  - ML exact interacting functional in thermodynamic limit
  - Full 3D calculation of water molecule, using ML for n[v](r)
- Thanks to
  - Students: Li Li, John Snyder, Kevin Vu, Isabelle Pelaschier
  - Collaborators: Klaus Mueller, Matthias Rupp, Katia Hansen
  - Funders: NSF from chem, DMR, math