High Dimensional Learning Rather than Computing in Quantum Chemistry

Matthew Hirn, Stéphane Mallat, Nicolas Poilvert

Michigan State University
Chemistry Seminar

November 17, 2015
Today’s Talk

1. Introduction
   • Total Energy of Molecules
   • Regression of Energy over a Dictionary

2. Dictionaries
   • Coulomb Matrices and SOAP Kernels
   • Density Functional Theory Approach
   • From Fourier to Wavelet Dictionaries
   • Wavelet Scattering Networks

3. Numerical Regression Results
   • Planar Organic Molecules
   • Water Clusters
   • Water & Li+ Clusters
Introduction: Quantum Energy Regression
Energy of a Molecule

- Born-Oppenheimer Approximation:
  - Nuclei: Classical
  - Electrons: Quantum

- \( x = \) state of molecule
  \( = (\) nuclear charges, positions of nuclei\( ) \)
  \( = \{(z_k, r_k) \in \mathbb{R} \times \mathbb{R}^3\}_{k=1}^d \)

- **Total Energy** of the molecule:

\[
\mathcal{E}(x) = E(x) + \frac{1}{2} \sum_{k \neq l} \frac{z_k z_l}{|r_k - r_l|}
\]

- Electronic Time Independent Schrödinger’s Equation:

\[
H(x)\Psi = E(x)\Psi
\]
Schrödinger’s Equation

- Electronic Time Independent Schrödinger’s Equation Ground State:
  \[ H(x)\Psi = E(x)\Psi \iff E(x) = \inf\{\langle\Psi, H(x)\Psi\rangle : \Psi \in \mathcal{H}, \|\Psi\| = 1\} \]

- Hamiltonian:
  - Kinetic Energy
  - Electron-Nuclei
  - Electron-Electron

\[
H(x) = -\frac{1}{2} \sum_{j=1}^{N} \Delta \bar{r}_j - \sum_{j=1}^{N} \sum_{k=1}^{d} \frac{z_k}{|\bar{r}_j - r_k|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\bar{r}_i - \bar{r}_j|}
\]

\[\bar{r}_j = \text{position of } j^{th} \text{ electron, } j = 1, \ldots, N\]

- Need to solve many, many times for:
  - Google of molecules
  - Molecular dynamics
  - Molecular geometry optimization
Solving Schrödinger’s Equation

• Direct Attacks

• Approximations:

  • Wavefunction Methods:
    Restrict $\Psi$ to a tractable subspace

  • Density Functional Methods:
    \[
    \rho(u) = \int_{\mathbb{R}^{3(N-1)}} |\Psi(u, \vec{r}_2, \ldots, \vec{r}_N)|^2 \, d\vec{r}_2 \cdots d\vec{r}_N
    \]

• A New Approach… Regression:
  Interpolate $E$ from a set of known examples \(\{(x_i, E(x_i))\}_i\)
Curse of Dimensionality

- High dimensional \( x \in \Omega \subset \mathbb{R}^d \)

- Approximate a function \( f(x) \) given \( n \) samples \( \{x_i, f(x_i)\}_i \)

- \( f(x) \) can be approximated from the samples by local interpolation if \( f \) is regular and there are close examples

- Need \( n = \epsilon^{-d} \) points to cover \( [0, 1]^d \) with an \( \epsilon \)-net
Curse of Dimensionality

- High dimensional $x \in \Omega \subset \mathbb{R}^d$
- Approximate a function $f(x)$ given $n$ samples $\{x_i, f(x_i)\}_i$
- $f(x)$ can be approximated from the samples by local interpolation if $f$ is regular and there are close examples
- Need $n = \epsilon^{-d}$ points to cover $[0, 1]^d$ with an $\epsilon$-net

$\Rightarrow$ Need to interpolate $f$ in a low dimensional approximation space
Quantum Energy Regression

• Linear approximation of $E$ in a dictionary $\Phi = \{\phi_p\}_p$:

$$\tilde{E}(x) = \langle \alpha, \Phi(x) \rangle = \sum_p \alpha_p \phi_p(x)$$

which minimizes the training error $\sum_i |E(x_i) - \tilde{E}(x_i)|^2$

with $\|\alpha\|_2 \leq \lambda$ or $\|\alpha\|_0 \leq M$.

• Kernel $K(x, x') = \langle \Phi(x), \Phi(x') \rangle = \sum_p \phi_p(x)\phi_p(x')$

$$\Rightarrow \tilde{E}(x) = \sum_i w_i K(x, x_i)$$

Goal: Design $\Phi$ or $K$ to minimize $\mathbb{E}(|E(x) - \tilde{E}(x)|^2)$

Problem: Curse of Dimensionality: #(Training) $\ll O(\epsilon^{-d})$

Approach: $\Phi$ or $K$ have same regularity properties as $E$

$$|\tilde{E}(x) - \tilde{E}(x')| \leq \|\alpha\| \|\Phi(x) - \Phi(x')\|$$
Energy Properties

Energy $E(x)$ of a molecule $x = \{(z_k, r_k)\}_{k=1}^d$

1. **Permutation Invariance:**
   Invariant to permutations of the indexation of the atoms in each molecule.

2. **Isometry Invariance:**
   Invariant to translations, rotations and reflections.

3. **Deformation Stability:**
   Lipschitz stable to diffeomorphisms of the molecule.

4. **Multi-scale Interactions:**
   - Highly energetic covalent bonds between neighboring atoms
   - Weaker energetic exchanges at longer distances
   - Factorization into multi-scale local interactions
   
   Can reduce $O(d)$ interactions into $O(\log d)$ interactions

Fast multipole methods (Greengard, Rocklin)

Want representation $\Phi$ to have these properties!
Non-Interacting Variables

- Suppose $E$ can be decomposed into at most $n$ interacting variables:

$$E(x) = E(r_1, \ldots, r_d) = \sum_{\ell=1}^{L} E_\ell(r_{k_{1,\ell}}, \ldots, r_{k_{n,\ell}})$$

If the $E_\ell$'s are Lipschitz, need $O(L\epsilon^{-n})$ samples for $\epsilon$ precision

- Classical: Fast Multipole Methods utilize multiscale regroupments of charges and far field expansions

- Quantum: Learn a nonlinear "change of variables"

$$x \mapsto \{\phi_p(x)\}_{p=1}^{M}$$

from $d$ interacting variables to $M$ non-interacting variables
Dictionary Construction: Distance Based Previous Results
Coulomb Matrices
*(Rupp, Tkatchenko, Müller, von Lilienfeld)*

\[ C(x)_{k,l} = \begin{cases} 
0.5z_k^2z_l^2 & \text{if } k = l \\
z_kz_l/|r_k - r_l| & \text{if } k \neq l 
\end{cases} \]

**Kernel:**
\[ K(x, x') = e^{-\|C(x) - C(x')\|_1}/\sigma \]

- Not permutation invariant.
- Fix: Sort the rows/columns by norm.

Now not continuous to diffeomorphisms
SOAP Kernels

(Bartók, Kondor, Csányi)

- \( E(x) = \sum_{k=1}^{d} \varepsilon_k(x) \) + (long range contributions)

- \( \rho_{\text{soap}}(u) = \sum_{k=1}^{d} g_{\sigma_k}(u - r_k), \quad g_\sigma(u) = e^{-|u|^2/2\sigma^2} \)

- Autocorrelation: \( a(\tau) = \int_{\mathbb{R}^3} \rho_{\text{soap}}(u) \rho_{\text{soap}}(u - \tau) \, du \)

\[
= \sum_{k,l=1}^{d} g_{\sqrt{\sigma_k + \sigma_l}}(\tau - (r_k - r_l))
\]

- Localized around each atom:

\[
a_k(\tau) = \sum_{l:|r_k - r_l| < \delta} g_{\sqrt{\sigma_k + \sigma_l}}(\tau - (r_k - r_l))
\]

- Kernel: \( K(a_k, a_l) = \int_{\text{SO}(3)} |\langle a_k, R_\theta a_l \rangle|^q \, d\theta \)

Not stable to diffeomorphisms for long range interactions (more on this later)
Dictionary Construction: DFT Inspired Permutation Invariance
Density Functional Theory

- Molecular state:
  \[ x = \{(z_k, r_k)\}_k \]

- Electronic density:
  \[ x \mapsto \rho_x(u) \]

- Hohenberg and Kohn 1964:
  \[ \rho_x = \operatorname{arg\,min}_{\rho} E_{\text{HK}}(\rho) \text{ and } E(x) = E_{\text{HK}}(\rho_x) \]

\[ E_{\text{HK}}(\rho) = \underbrace{T(\rho)}_{\text{Kinetic Energy}} - \sum_k \int_{\mathbb{R}^3} \frac{z_k \rho(u)}{|r_k - u|} \, du + \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(u)\rho(v)}{|u - v|} \, du \, dv + \underbrace{E_{\text{xc}}(\rho)}_{\text{Exchange Correlation Energy}} \]

- External Energy (electron-nuclei attraction)
- Hartree Energy (electron-electron repulsion)
Permutation Invariance

- \( \rho_x \) is permutation invariant

- Given \( \rho_x \), we could interpolate \( E_{HK}(\rho_x) \) in a dictionary
  \[
  \Phi(x) = \{ \phi_p(x) \}_p = \{ \varphi_p(\rho_x) \}_p
  \]

- Computing \( \rho_x \) requires complex quantum mechanical calculations. Need a substitute that we can compute fast.

\[
E_{HK}(\rho) = T(\rho) - \sum_k \int_{\mathbb{R}^3} \frac{z_k \rho(u)}{|r_k - u|} \, du + \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(u)\rho(v)}{|u - v|} \, du \, dv + E_{xc}(\rho)
\]

- Kinetic Energy
- External Energy (electron-nuclei attraction)
- Hartree Energy (electron-electron repulsion)
- Exchange Correlation Energy
Atomic Electronic Density

- Approximate the electronic density with a linear superposition of atomic electronic densities:

\[ \tilde{\rho}_x(u) = \sum_k \rho_a(k)(u - r_k) \]

\( \rho_a = \) exact electronic density of atom \( a \)

- The regression is computed as:

\[ E(x) = E_{HK}(\rho_x) \approx \tilde{E}(\tilde{\rho}_x) = \sum_p \alpha_p \varphi_p(\tilde{\rho}_x) \]

How to design the functionals \( \varphi_p \)?
Dictionary Construction: From Fourier to Wavelet Dictionaries
Fourier Dictionary

- Fourier Transform: \( \hat{\rho}(\omega) = \int \rho(u)e^{-iu \cdot \omega} du \)

- "Classic" translation invariant representations:
  - Autocorrelation: \( \Phi(\rho)(\tau) = \int \rho(u)\rho(u - \tau) du \) (SOAP)
  - Fourier Modulus: \( \bar{\Phi}(\rho)(\omega) = |\hat{\rho}(\omega)|^2 \)

- Isometry invariance by integrating over rotations:

\[
\Phi_F = \{ \varphi_{\gamma,q} : \gamma \geq 0, q = 1, 2 \} \\
\varphi_{0,q}(\rho) = |\hat{\rho}(0)|^q = (\text{total charge})^q \\
\varphi_{\gamma,q}(\rho) = \int_{|\omega| = \gamma} |\hat{\rho}(\omega)|^q d\omega
\]
Coulomb Potential Energy

- Coulomb Potential Energy:
  \[ U(\rho) = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \rho(u)\rho(v)V(u-v) \, du \, dv, \quad V(u) = |u|^{-1} \]

- Convolutional formula for Coulomb energy:
  \[ U(\rho) = \frac{1}{2} \int_{\mathbb{R}^3} \rho \ast \bar{\rho}(u)V(u) \, du, \quad \bar{\rho}(u) = \rho(-u) \]

- Fourier transform:
  \[ \hat{\rho}(\omega) = \int_{\mathbb{R}^3} \rho(u)e^{-iu\cdot\omega} \, du \]

- Coulomb energy in frequency:
  \[ U(\rho) = \frac{1}{2(2\pi)^3} \int_{\mathbb{R}^3} |\hat{\rho}(\omega)|^2 \hat{V}(\omega) \, d\omega \]
Fourier Regression of Coulomb Potential Energy

- Coulomb energy diagonalized in Fourier:
  \[ U(\rho) = \frac{1}{2(2\pi)^3} \int_{\mathbb{R}^3} |\hat{\rho}(\omega)|^2 \hat{V}(\omega) \, d\omega \quad \hat{V}(\omega) = C|\omega|^{-2} \]

- **Isometry Invariant Fourier Dictionary:**
  In polar coordinates \( \omega = \gamma \eta \) with \( \gamma = |\omega| \) and \( \eta \in S^2 \), \( \hat{V}(\omega) = \hat{V}(\gamma) \) so
  \[ U(\rho) = \frac{1}{2(2\pi)^3} \int_{\mathbb{R}^+} \hat{V}(\gamma) \varphi_{\gamma,2}(\rho) \, d\gamma, \quad \varphi_{\gamma,2}(\rho) = \int_{|\omega| = \gamma} |\hat{\rho}(\omega)|^2 \, d\omega \]

- To learn discrete weights, approximate with Riemann sum:
  \[ \tilde{U}(\rho) = \frac{\Delta}{2(2\pi)^3} \sum_{m=1}^{M} \hat{V}(m\Delta) \varphi_{m\Delta,2}(\rho) \]

- \( U(\rho) = \tilde{U}(\rho) + O(\epsilon) \Rightarrow M \sim O(\epsilon^{-2}) \)
Fourier is Not Sparse

\[ \Phi_F = \{ \varphi_{\gamma,q} : \gamma \geq 0, \ q = 1, 2 \} \]

\[ \varphi_{0,q}(\rho) = |\hat{\rho}(0)|^q = \text{(total charge)}^q \]

\[ \varphi_{\gamma,q}(\rho) = \int_{|\omega| = \gamma} |\hat{\rho}(\omega)|^q d\omega \]

- Fourier is not sparse!
- Requires \( M \sim O(\epsilon^{-2}) \) terms to regress Coulomb energy.

\[ \tilde{U}(\rho) = \frac{\Delta}{2(2\pi)^3} \sum_{m=1}^{M} \hat{V}(m\Delta) \phi_m^2(\rho) \]

\[ U(\rho) = \tilde{U}(\rho) + O(\epsilon) \Rightarrow M \sim O(\epsilon^{-2}) \]

\[ \hat{V}(\gamma) \]
Fourier Long Range Instabilities

• "Classic" translation invariant representations:
  - Autocorrelation: $\Phi(\rho)(\tau) = \int \rho(u)\rho(u - \tau) \, du$
  - Fourier Modulus: $\Phi(\rho)(\omega) = |\hat{\rho}(\omega)|^2$

• $\rho(u) =$ bumps of width $\sigma$ at positions $\{r_k\}_k$
• $\Phi(\rho)(\tau) =$ bumps of width $2\sigma$ at positions $\{r_k - r_l\}_{k,l}$
• Diffeomorphism $D_\tau \rho(u) = \rho(u - \tau(u))$
• A small diffeomorphism changes distances by $\varepsilon|r_k - r_l|$
• Unstable if $|r_k - r_l| \geq 2\sigma/\varepsilon$
• Equivalent to instability of high frequency Fourier modes.
Wavelets

- Complex Wavelet: \( \psi(u) = g(u)e^{i\eta_0 \cdot u} \)

- Dilated and rotated:
  \[
  \psi_{j,\theta}(u) = 2^{-nj} \psi(2^{-j} R_\theta u), \quad (j, \theta) \in \mathbb{Z} \times S^{n-1}
  \]

Wavelet Transform: \( \rho \mapsto \{ \int \rho, \rho \ast \psi_{j,\theta}(u) \} \)

Interactions at scale \( 2^j \) in direction \( \theta \)
Wavelet Transform: Interactions of Waves

\[ |\rho \ast \psi_{j_1, \theta_1}(u)| \]

Rotations \( \theta_1 \)

Scales \( j_1 \)
Wavelet Dictionary

\[ \Phi_W = \{ \varphi_q, \varphi_{j,q} : j \in \mathbb{Z}, q = 1, 2 \} \]

\[ \varphi_q(\rho) = \int_{\mathbb{R}^n} |\rho(u)|^q \, du = \|\rho\|^q_q \]

\[ \varphi_{j,q}(\rho) = \int_{S^{n-1}} \int_{\mathbb{R}^n} |\rho * \psi_{j,\theta}(u)|^q \, du \, d\theta \]
\[ = \|\rho * \psi_{j,\cdot}\|^q_q \]

- Permutation invariant \( \checkmark \)
- Isometry invariant \( \checkmark \)
- Lipschitz stable to diffeomorphisms \( \checkmark \)

\[ \| \Phi(\rho) - \Phi(D\tau \rho) \| \leq C \cdot \sup_{u \in \mathbb{R}^3} \| \nabla \tau(u) \| \cdot \| \rho \|_2 \]  \( (\text{Mallat 2012}) \)
Wavelets are Sparse

Hartree Energy

\[
U(\rho) = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(u)\rho(v)}{|u - v|} \, du \, dv
\]

\[
\Phi_W = \{\varphi_q, \varphi_{j,q} : j \in \mathbb{Z}, q = 1, 2\}
\]

\[
\varphi_q(\rho) = \int_{\mathbb{R}^n} |\rho(u)|^q \, du
\]

\[
\varphi_{j,q}(\rho) = \int_{S^{n-1}} \int_{\mathbb{R}^n} |\rho * \psi_{j,\theta}(u)|^q \, du \, d\theta
\]

Suppose that \( \rho \) is Lipschitz and \( 0 < \epsilon < 1 \).
Then there exists Littlewood-Paley wavelets \( \{\psi_{j,\theta}\}_{j,\theta} \), scales \( \{j_\ell\}_{\ell=1}^M \) and weights \( \{\alpha_\ell\}_{\ell=1}^M \) such that

\[
\sum_{\ell=1}^M \alpha_\ell \varphi_{j_\ell,2}(\rho) = (1 + O(\epsilon))U(\rho) + O(\epsilon), \quad M \sim O(|\log \epsilon|^2)
\]

(H., Mallat, Poilvert)
Wavelets are Sparse

Fourier Sampling

Wavelet Sampling
Regression and Initial Numerical Results
Sparse Linear Regression

- Sparse regression in dictionary $\Phi = \{\phi_p\}_p$
  
  by selecting $M$ functionals:

  $\tilde{E}_M(x) = \sum_{\ell=1}^{M} \alpha_\ell \phi_{p_\ell}(x)$

  which minimize the training error $\sum_i |E(x_i) - \tilde{E}_M(x_i)|^2$

- Greedy selection of the $\{\phi_{p_\ell}\}_{\ell=1}^{M}$
  
  by orthogonal least squares

- Cross-validation on $M$

- ”Bagging” to reduce variance error
Fourier and Wavelet M-term Regression Error

\[ 2^{-1} \log_2(\mathbb{E}|E(x) - \widetilde{E}_M(x)|^2) \]

Multiscale wavelet dictionary is sparser than Fourier dictionary

Data set: Planar organic molecules

Key: Fourier, Wavelets, Coulomb matrices (dashed line)
Fourier and Wavelet M-term Regression Error

\[ 2^{-1} \log_2 (\mathbb{E}|E(x) - \tilde{E}_M(x)|^2) \]

Coulomb matrices outperform both Fourier and wavelets

Root Mean Square Error (kcal/mol)

<table>
<thead>
<tr>
<th></th>
<th>Coulomb</th>
<th>Fourier</th>
<th>Wavelet</th>
</tr>
</thead>
<tbody>
<tr>
<td>4000 Training</td>
<td>5.8</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>400 Training</td>
<td>21</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>

Error = Bias + Variance

Dominating

Key: Fourier, Wavelets, Coulomb matrices (dashed line)

- Need to augment wavelet dictionary within the range of useful scales
Deep Wavelet Scattering Dictionary
Multipole expansion dipole-dipole moment:

\[ \frac{\mu_1 \cdot \mu_2}{R^3} \]
Small scale wavelets $\psi_{j_1}$ can learn dipole orientations
Large scale wavelets $\psi_{j_2}$ can learn distance between molecules
But there is no linear expansion of the dipole-dipole moment as:

$$\frac{\mu_1 \cdot \mu_2}{R^3} \neq \sum_{j_1 \text{ small}} \alpha_{j_1} \| \rho \ast \psi_{j_1} \| + \sum_{j_2 \text{ large}} \alpha_{j_2} \| \rho \ast \psi_{j_2} \|$$
A better solution is to:

- First apply the small scale wavelet transform $| \rho \ast \psi_{j_1}|$...
- ...then apply the large scale wavelet transform $||\rho \ast \psi_{j_1} \ast \psi_{j_2}||$
- It is then plausible that (numerical evidence is positive):

$$\frac{\mu_1 \cdot \mu_2}{R^3} \approx \sum_{j_1 \text{ small}} \sum_{j_2 \text{ large}} \alpha_{j_1,j_2} \||\rho \ast \psi_{j_1} \ast \psi_{j_2}||$$
Scattering Network

\[ \|\rho\|_q \]

\[ |\rho \ast \psi_{\lambda_1}| \]

\[ \|\rho \ast \psi_{\lambda_1}\|_q \]

\[ |\rho \ast \psi_{\lambda_1} \ast \psi_{\lambda_2}| \]

\[ \|\rho \ast \psi_{\lambda_1} \ast \psi_{\lambda_2}\|_q \]

\[ |\rho \ast \psi_{\lambda_1} \ast \psi_{\lambda_2} \ast \psi_{\lambda_3}| \]
Scattering Network

Similar to a deep convolutional network!

Fully learned linear filters are replaced by wavelets to guarantee regularity properties of the network and to circumvent the curse of dimensionality.
Scattering in 2D: Layer 1

Wave Interactions

\[ \varphi_q(\rho) = \int_{\mathbb{R}^2} |\rho(u)|^q \, du \]

\[ |\rho \ast \psi_{j_1, \theta_1}(u)| \]
Scattering in 2D: Layer 1

Wave Interactions

Rotations $\theta_1$

Scales $j_1$

$|\rho \ast \psi_{j_1, \theta_1}(u)|$

$\varphi_{j_1, q}(\rho) = \int_0^{2\pi} \int_{\mathbb{R}^2} |\rho \ast \psi_{j_1, \theta_1}(u)|^q du \, d\theta_1$

$\sim 60$ functionals
Scattering in 2D: Layer 2

Interactions of Interactions

Rotations $\theta_1$

Scales $j_1$

\[ |\rho \ast \psi_{j_1, \theta_1}(u)| \]

\[ |\rho \ast \psi_{j_1, \theta_2}(u)| \]

Compute interactions of interactions via a second wavelet transform!

Recovers information lost in the wavelet dictionary.
Scattering in 2D: Layer 2
Interactions of Interactions

\[ \Phi_s = \{ \varphi_q, \varphi_{j_1,q}, \varphi_{j_1,j_2,\beta,q} : j_1, j_2 \in \mathbb{Z}, \beta \in [0, 2\pi), q = 1, 2 \} \]

\[ \varphi_{j_1,j_2,\beta,q}(\rho) = \int_0^{2\pi} \int_{\mathbb{R}^2} \| \rho * \psi_{j_1,\theta_1} * \psi_{j_2,\theta_2 + \beta}(u) \|^q \, du \, d\theta_1 \sim 10000 \text{ functionals} \]
Numerical Results
Scattering
M-term Regression Error

\[ 2^{-1} \log_2 (\mathbb{E}|E(x) - \tilde{E}_M(x)|^2) \]

Scattering is sparser than even wavelets, and more accurate than Coulomb matrices!

Data set:
Planar organic molecules

Key: Fourier, Wavelets, Scattering, Coulomb matrices (dashed line)
Scattering M-term Regression Error

\[2^{-1} \log_2 (\mathbb{E}|E(x) - \tilde{E}_M(x)|^2)\]

Data set: Planar organic molecules

Root Mean Square Error (kcal/mol)

<table>
<thead>
<tr>
<th></th>
<th>Coulomb</th>
<th>Fourier</th>
<th>Wavelet</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>4000 Training</td>
<td>5.8</td>
<td>17</td>
<td>14</td>
<td>2.7</td>
</tr>
<tr>
<td>400 Training</td>
<td>21</td>
<td>16</td>
<td>16</td>
<td>9.0</td>
</tr>
</tbody>
</table>

Key: Fourier, Wavelets, Scattering, Coulomb matrices (dashed line)
Water Molecules

\[ 2^{-1} \log_2 (\mathbb{E}|E(x) - \tilde{E}_M(x)|^2) \]

Root mean square error (meV/atom)

<table>
<thead>
<tr>
<th></th>
<th>SOAP</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>700 Training</td>
<td>0.70</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Long range interactions are \( O(1/R^3) \)

Data set:
Water molecules

Key: Scattering, SOAP (dashed line)
Water Lithium Ion Systems

\[ 2^{-1} \log_2(\mathbb{E}|E(x) - \tilde{E}_M(x)|^2) \]

Root mean square error (meV/atom)

<table>
<thead>
<tr>
<th></th>
<th>SOAP</th>
<th>Scattering</th>
</tr>
</thead>
<tbody>
<tr>
<td>700 Training</td>
<td>40</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Key: Scattering, SOAP (dashed line)

Long range interactions are \( O(1/R^2) \)

Data set: Water & Lithium Ion
Future Directions

• Utilize forces for molecular dynamics and molecular geometry optimization:

\[ \nabla \tilde{E}(x) = \sum_p \alpha_p \nabla \phi_p(x) \]

• Rigorous interpretation of 2nd layer scattering functionals
  - Multipole moments
  - Other components of the energy (kinetic, etc…)

• Transferability of scattering energy models
  - Uncertainty quantification
  - Force fields

- **3D code (no mathematical difficulty, programming in progress)**
Conclusions

To interpolate quantum molecular energies, we must attack the curse of dimensionality. To do so we want to:

• Reduce the dimension by…
• …separation of original variables into learned non-interacting variables.

To achieve these goals, we utilize:

• Scale and angular separation over groups (translation, rotation)
• Symmetry and invariance properties
• Stability to diffeomorphisms
• Cascade of operators (wavelet modulus) with the previous three properties
Thank You!

www.math.msu.edu/~mhirn
mhirn@msu.edu