High Dimensional Learning rather than Computing in Quantum Chemistry

Matthew Hirn*, Stéphane Mallat*, Nicolas Poilvert**

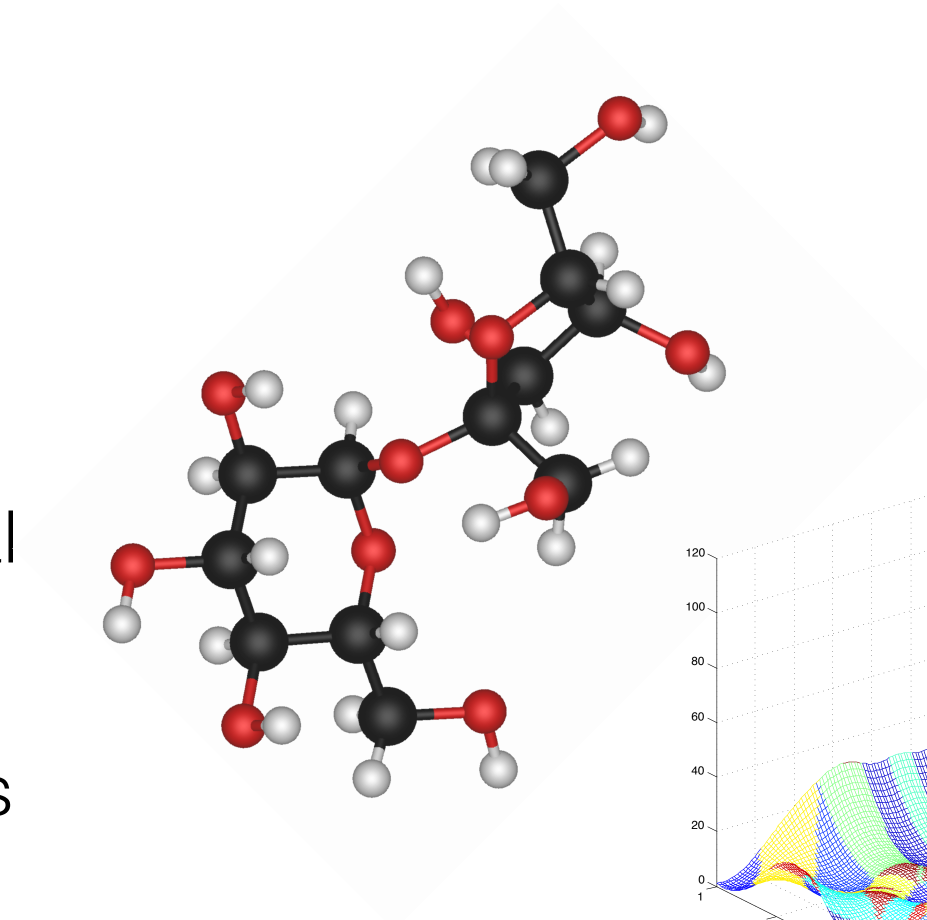
*École normale supérieure
**Pennsylvania State University

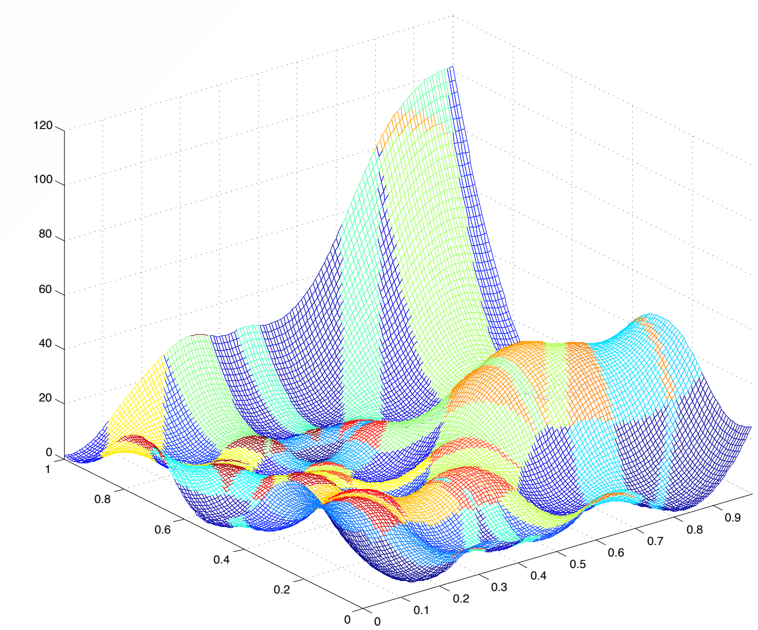
Yale University
Applied Mathematics Seminar
February 4, 2015

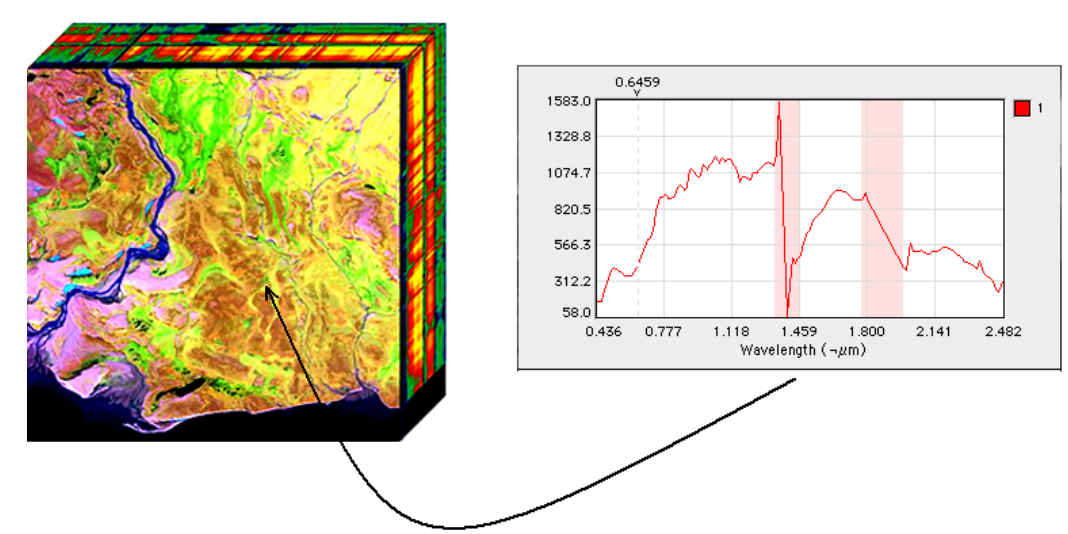
Introduction

Broad Motivation

- Big Data: Massive amounts of high dimensional data
- Audio, medical, images, hyperspectral, video, dynamical systems, quantum chemistry
- Want to learn important features of new data fast
- Interplay between discrete and continuous at the interface of analysis, geometry, computer science, statistics, chemistry, physics

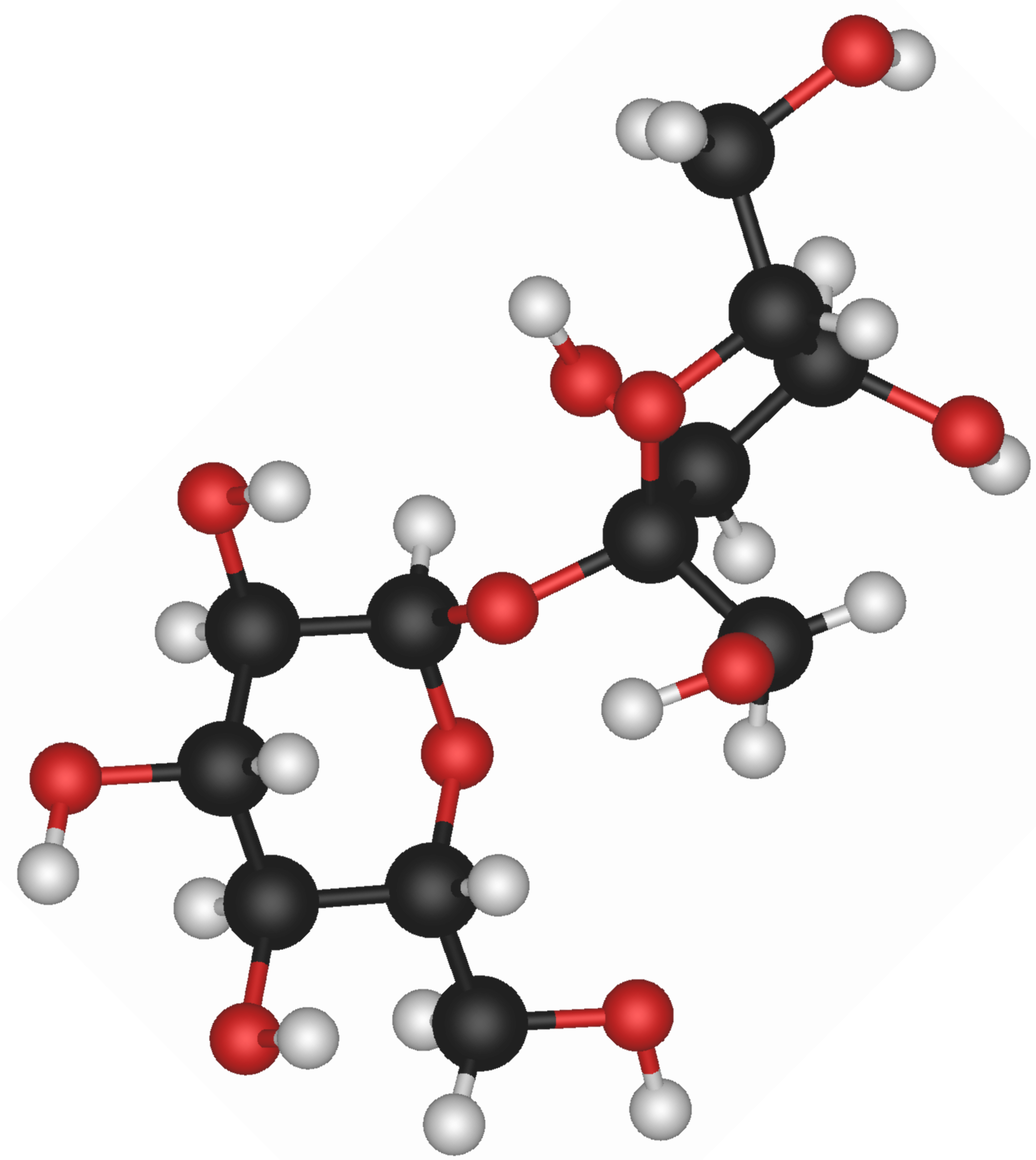






Quantum Chemistry Motivation

- Chemists want to build "Google of molecules"
- Applications in pharmaceutical industry, materials science, among others
- Need to compute potential energy of each molecule
- Billions of molecules
- Complex, time consuming computation



Energy Computation

- <u>Exact:</u>
  Schrödinger's Equation: $\widehat{H}\Psi = E\Psi$
  Extremely high dimensional eigenvalue problem
  Example: Alcohol $\mathrm{C_2H_6O}$ is ~2^{300} dimensional

- <u>Approximate:</u>
  ab-initio methods:
  - Coupled cluster methods
  - Density functional theory
  Scales as $O(N^{\alpha})$ where $4 \leq \alpha \leq 7$

Number of electrons

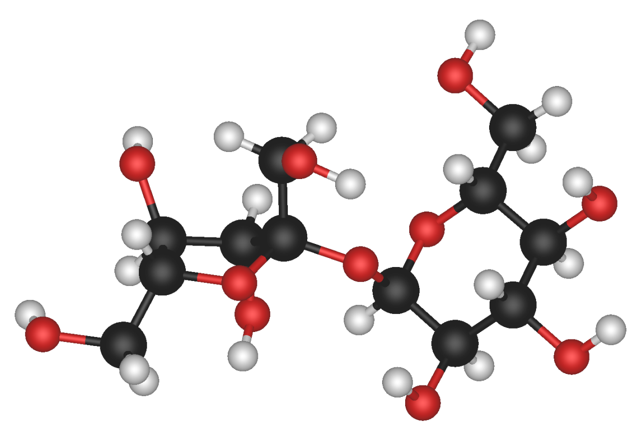
Regression

- High dimensional $x \in \mathbb{R}^d$
- Approximate a functional $f(x)$ given n sample values $\{x_i, f(x_i)\}_i$
- Many body problems:
  Energy $f(x)$ of a state $x = \{(p_k, q_k)\}_k$

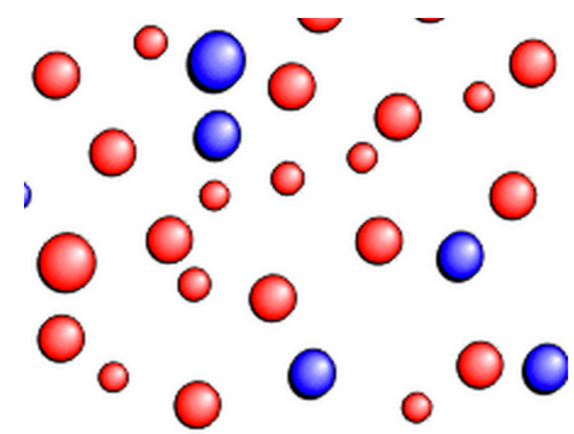
Position

Celestial mechanics: mass of body
Classical electrostatics: charge of particle
Quantum chemistry: total protonic charge of atom

Astronomy



Quantum Chemistry



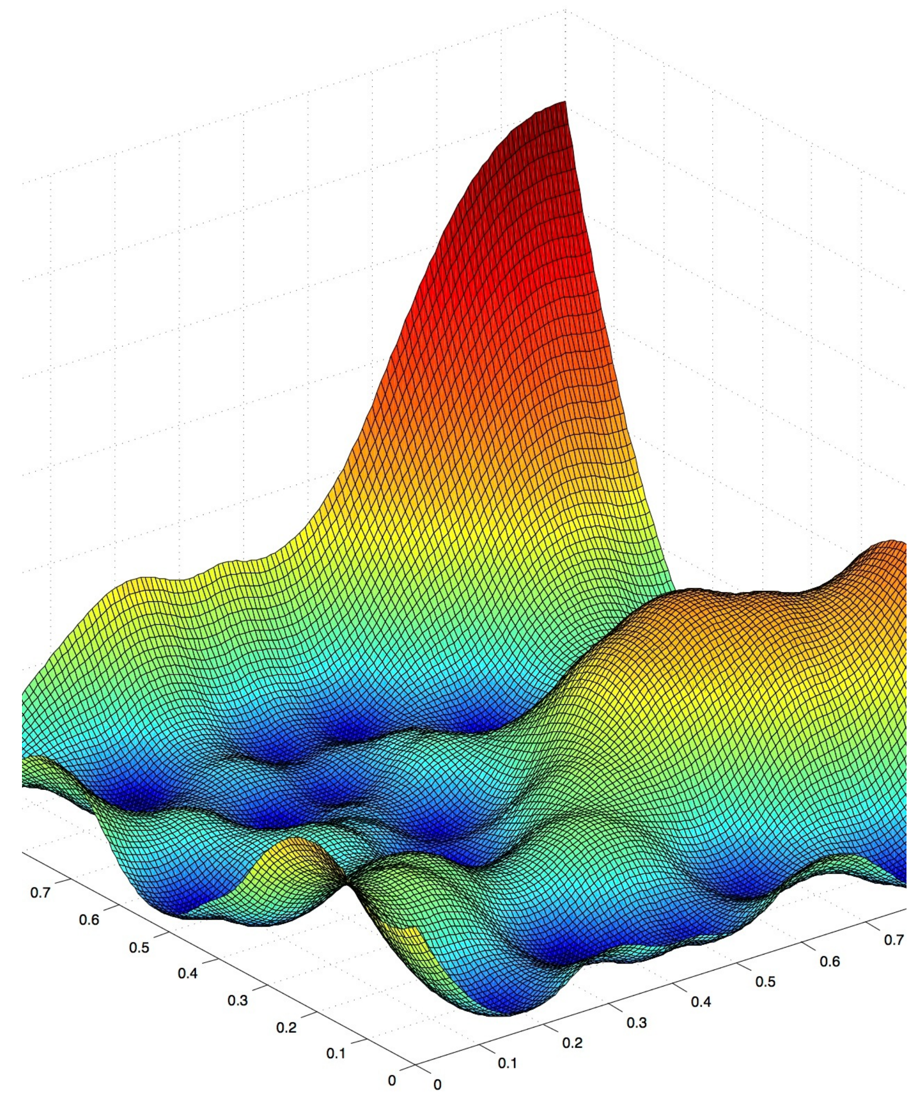
Classical Electrostatics

Curse of Dimensionality

- High dimensional $x \in X \subset \mathbb{R}^d$

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

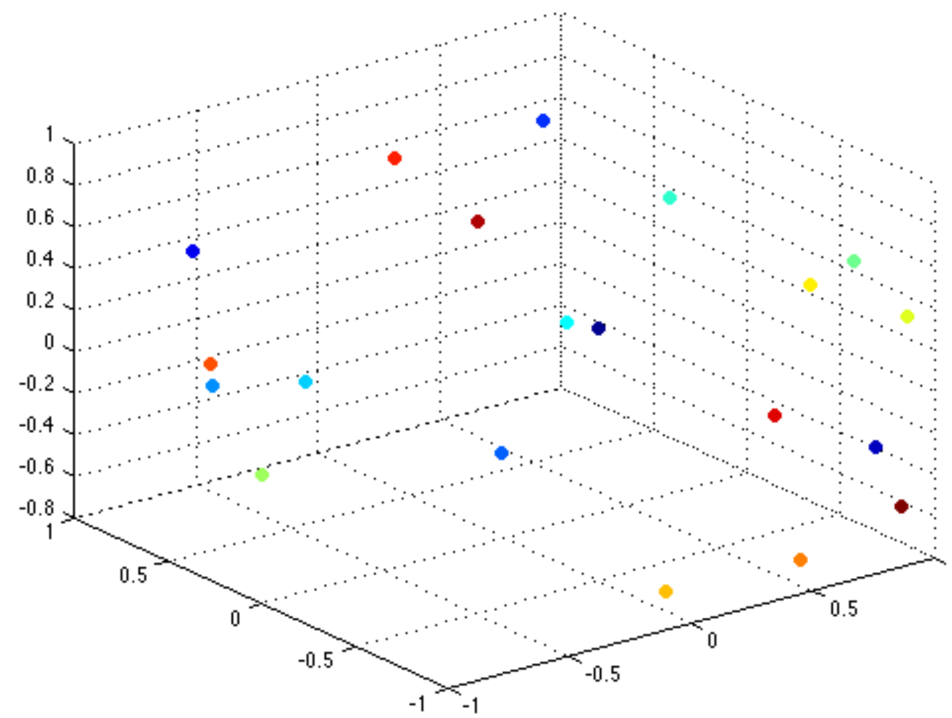
- $f(x)$ can be approximated from the samples by local interpolation if f is regular and there are close examples



- Need $n = \epsilon^{-d}$ points to cover $[0,1]^d$ with an ϵ-net

Curse of Dimensionality

- High dimensional $x \in X \subset \mathbb{R}^d$
- 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 ϵ-net

$$\Longrightarrow \|x - x_i\| \text{ is always large}$$

Sparse Linear Regression

- Representation of x : $\Phi(x) = \{\phi_n(x)\}_n$

- Regression $\tilde{f}(x)$ of $f(x)$ linear in $\Phi(x)$:
$$\tilde{f}(x) = \sum_n \alpha_n \phi_n(x)$$

- Few samples $\{x_i, f(x_i)\}_i$ so want a low dimensional approximation of f to avoid curse of dimensionality

- Find regression functions $\{\phi_n\}_n$ with similar properties as f to allow us to compute a sparse regression

Finding a Good Representation

Energy Properties

- State: $x = \{(p_k, q_k)\}_k$ positions of atoms and number of protons
- Energy: $f(x)$

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

2. **Isometry Invariance:**
   Invariant to actions of the isometry group $\mathrm{E}(3) = \mathbb{R}^3 \rtimes \mathrm{O}(3)$ on the molecular state

3. **Deformation Stability:**
   The energy is differentiable with respect to the distances between atoms

4. **Multiscale Interactions:**
   - Highly energetic covalent bonds between neighboring atoms
   - Weaker energetic exchanges at larger distances (Van der Waals interactions)

- Want a representation $\Phi(x)$ that satisfies these properties

Current State of the Art

- The set of pairwise distances between atoms defines a set of isometry invariant descriptors that is stable to deformations

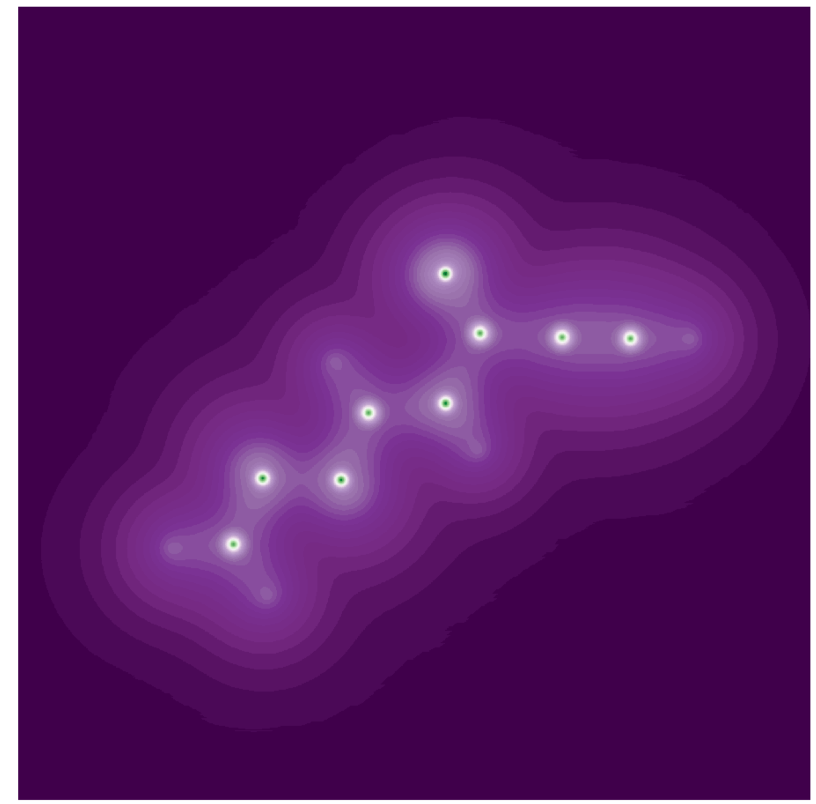
- Coulomb matrices [Rupp, et al 2012] refine this idea:

$$C_{k,l}(x) = \begin{cases} \frac{1}{2} q_k^{2.4} & k = l \\ \dfrac{q_k q_l}{|p_k - p_l|} & k \neq l \end{cases}$$

- Issues:
  - Not permutation invariant (sorted random matrices)
  - Different size matrices (zero padding)
  - All length scales are treated equally (nonlinear kernel)

Density Functional Theory

- State: $x = \{(p_k, q_k)\}_k$
- Energy: $f(x)$
- Electronic density: $x \mapsto \rho_x(u)$
- Hohenberg and Kohn 1964:



$$\rho_x = \arg\min_{\rho} E(\rho) \text{ and } f(x) = E(\rho_x)$$

$$E(\rho) = \underbrace{T(\rho)}_{\text{Kinetic energy}} + \underbrace{\int_{\mathbb{R}^3} \rho(u) V_{\mathrm{e}}(u)\, du}_{\substack{\text{External energy} \\ \text{(electron-nuclei attraction)}}} + \underbrace{\frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(u)\rho(v)}{|u-v|}\, du\, dv}_{\substack{\text{Coulomb energy} \\ \text{(electron-electron repulsion)}}} + \underbrace{E_{\mathrm{xc}}(\rho)}_{\substack{\text{Exchange} \\ \text{correlation} \\ \text{energy}}}$$

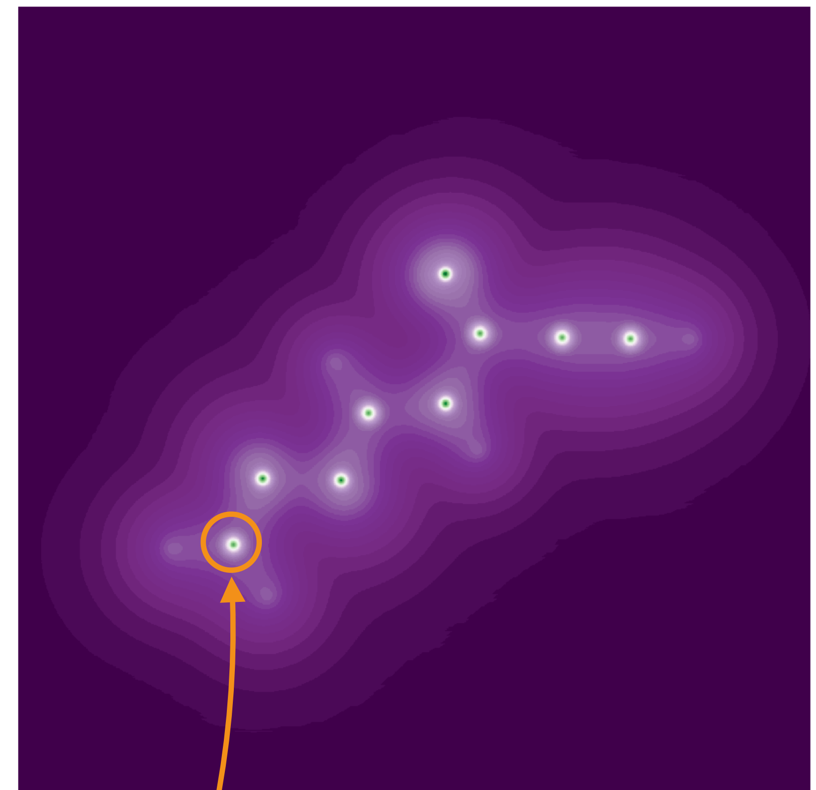
Density Functional Theoretic Learning rather than Computing

- Construct a representation $\Phi(\rho) = \{\phi_n(\rho)\}_n$ and compute a linear regression:

$$\tilde{f}(x) = \widetilde{E}(\tilde{\rho}_x) = \sum_n \alpha_n \phi_n(\tilde{\rho}_x)$$

- To avoid computing ρ_x, we take $\tilde{\rho}_x$ to be an approximation of ρ_x

- Local behavior near the nucleus of an atom is the same as the isolated electronic density of that atom



$\rho_x(u)$

Density Functional Theoretic Learning rather than Computing

- Construct a representation $\Phi(\rho) = \{\phi_n(\rho)\}_n$ and compute a linear regression:

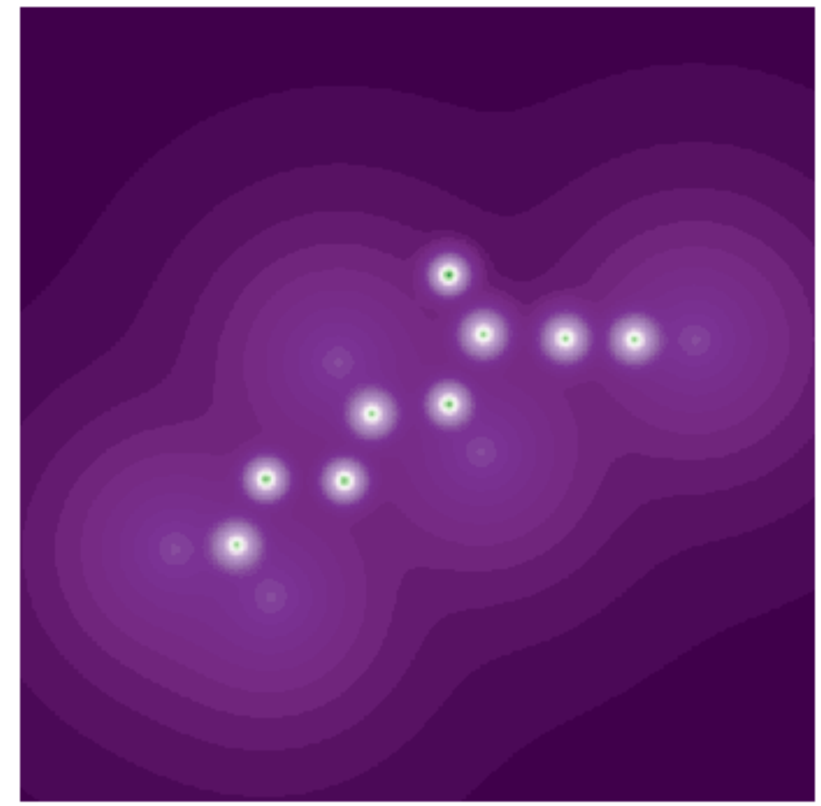
$$\tilde{f}(x) = \widetilde{E}(\tilde{\rho}_x) = \sum_n \alpha_n \phi_n(\tilde{\rho}_x)$$

- To avoid computing ρ_x, we take $\tilde{\rho}_x$ to be an approximation of ρ_x

- Atomic density approximation:

$$\tilde{\rho}_x(u) = \sum_k \rho_{a(k)}(u - p_k)$$

$$\rho_a = \text{ density of atom } a \text{ centered at zero}$$



$\tilde{\rho}_x(u)$

Stability to Deformations

- Deformation operator: $\rho_x = D\rho$
- Then $f(x) = E(\rho_x) = ED(\rho)$
- Want to linearly expand $ED(\rho)$ in $\Phi(\rho)$
- Diffeomorphism model: $\rho_x(u) = D_\tau \rho(u) = \rho(u - \tau(u))$
- Want Φ to be Lipschitz continuous to diffeomorphisms:
$$\|\Phi(\rho) - \Phi(D_\tau \rho)\| \leq C \cdot \sup_{u \in \mathbb{R}^3} \|\nabla \tau(u)\| \cdot \|\rho\|_2$$
- If $E(\rho)$ is well approximated by a linear regression in $\Phi(\rho)$, and $\Phi(\rho)$ is Lipschitz continuous over D_τ, then we can still linearly expand $ED_\tau(\rho)$ in $\Phi(\rho)$ with small error

Fourier and Wavelet Invariant Representations

Coulomb Potential Energy

- Coulomb Potential Energy:

$$U(\rho) = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \rho(u)\rho(v)V(u-v)\,du\,dv, \;\; V(u) = |u|^{-1}$$

- Convolutional formula for Coulomb energy:

$$U(\rho) = \frac{1}{2} \int_{\mathbb{R}^3} \rho * \bar{\rho}(u)V(u)\,du, \;\; \bar{\rho}(u) = \rho(-u)$$

- Fourier transform:

$$\hat{\rho}(\omega) = \int_{\mathbb{R}^3} \rho(u)e^{-iu\cdot\omega}\,du$$

- Coulomb energy in Fourier:

$$U(\rho) = \frac{1}{2(2\pi)^3} \int_{\mathbb{R}^3} |\hat{\rho}(\omega)|^2 \widehat{V}(\omega)\,d\omega$$

Fourier Regression of Coulomb Potential Energy

- Coulomb energy in Fourier:

$$U(\rho) = \frac{1}{2(2\pi)^3} \int_{\mathbb{R}^3} |\hat{\rho}(\omega)|^2 \widehat{V}(\omega)\, d\omega$$

- Isometry invariant Fourier representation:
  In polar coordinates $\omega = \gamma\eta$ with $\gamma = |\omega|$ and $\eta \in S^2$, $\widehat{V}(\omega) = \widehat{V}(\gamma)$ so

$$U(\rho) = \frac{1}{2(2\pi)^3} \int_{\mathbb{R}} \widehat{V}(\gamma)\phi_\gamma^2(\rho)\, d\gamma, \quad \phi_\gamma^2(\rho) = \int_{|\omega|=\gamma} |\hat{\rho}(\omega)|^2\, d\omega$$

- To learn discrete weights, approximate with Riemann sum:

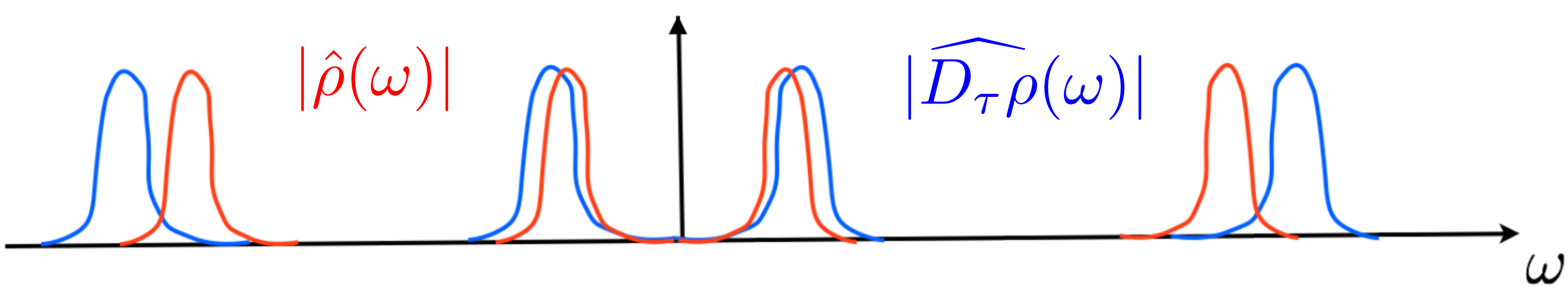
$$\widetilde{U}(\rho) = \frac{\Delta}{2(2\pi)^3} \sum_{m=1}^{M} \widehat{V}(m\Delta)\phi_{m\Delta}^2(\rho)$$

- To get $|U(\rho) - \widetilde{U}(\rho)| < \epsilon$ need $\Delta \sim \epsilon$ and so $M = O(\epsilon^{-1})$

Fourier Limitations

- The Fourier representation does not take advantage of the regularity of $\widehat{V}(\omega)$ away from $\omega = 0$. Therefore it needs $O(\epsilon^{-1})$ terms to achieve precision ϵ.

- Fourier is not stable to small diffeomorphisms at the high frequencies.



$$\| |\hat{\rho}| - |\widehat{D_\tau \rho}| \|_2 \gg \sup_{u \in \mathbb{R}^3} \|\nabla \tau(u)\| \cdot \|\rho\|_2$$

Wavelets

- Complex valued Morlet wavelet:

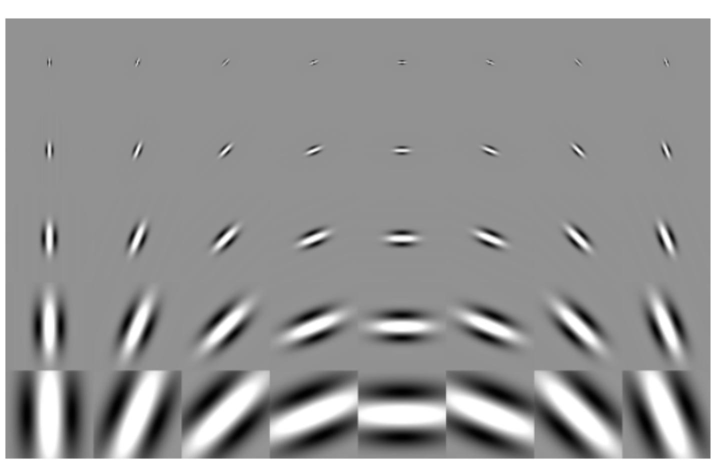
$$\psi(u) = g(u)(e^{i\eta_0 \cdot u} - C), \quad \int_{\mathbb{R}^3} \psi(u)\, du = 0$$

- Wavelet transform dilates and rotates the wavelet:

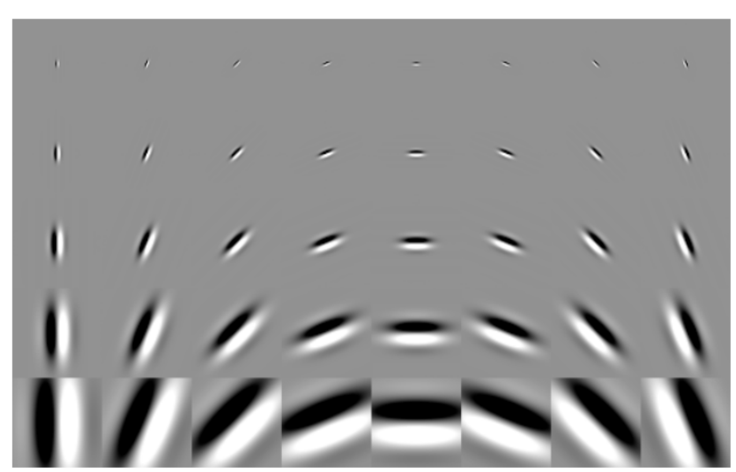
$$\psi_{j,r}(u) = 2^{-3\frac{j}{Q}} \psi(2^{-\frac{j}{Q}} r^{-1} u), \;\; (j,r) \in \mathbb{Z} \times \mathrm{O}(3)$$

$$Q \in \mathbb{N} : \;\text{Scale oversampling factor}$$

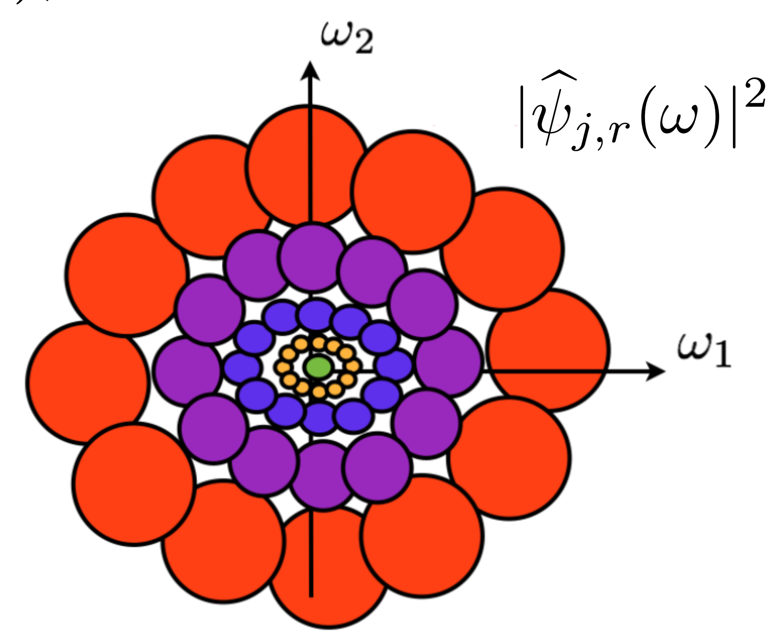
$$W[j,r]\rho(u) = \{\rho * \psi_{j,r}(u)\}_{j\in\mathbb{Z},\, r\in \mathrm{O}(3),\, u\in\mathbb{R}^3}$$



Real parts

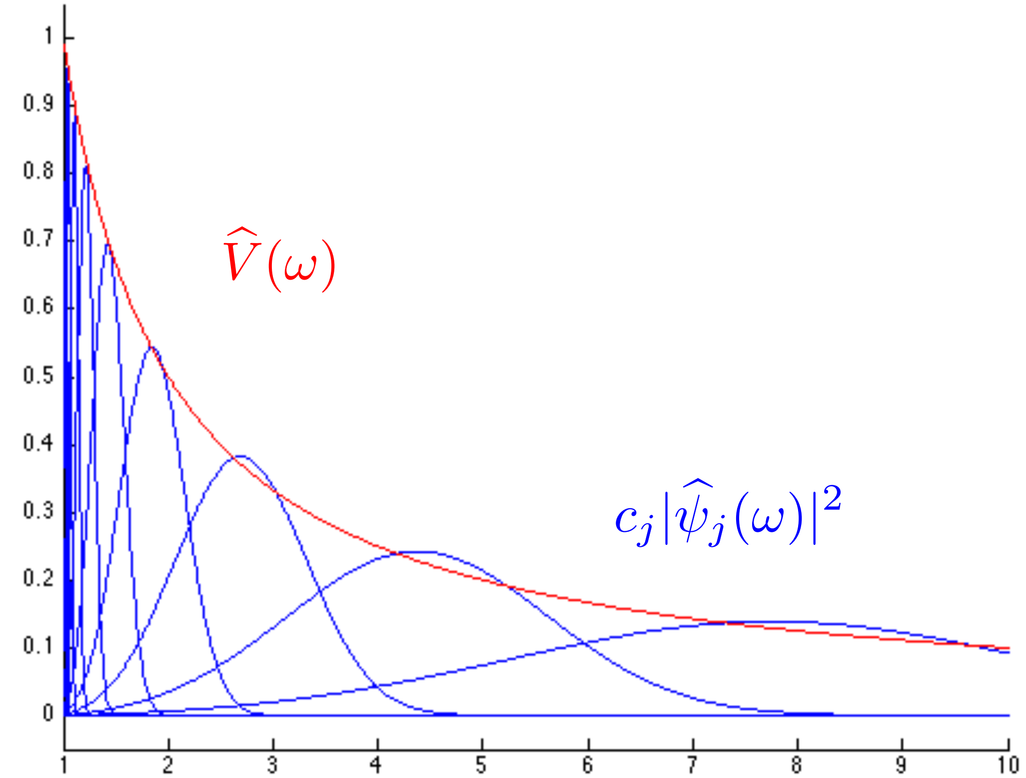


Imaginary parts



Fourier vs Wavelets

- Wavelets separate scales logarithmically and can thus take advantage of the multiscale structure of the energy. For the Coulomb potential energy, wavelets take advantage of the regularity of $\widehat{V}(\omega)$ away from $\omega = 0$.



- Mallat 2012: Wavelets are Lipschitz continuous to the action of diffeomorphisms:

$$\|[W, D_\tau]\| = \|W D_\tau - D_\tau W\| \leq C \cdot \sup_{u \in \mathbb{R}^3} \|\nabla \tau(u)\|$$

Wavelet Regression of Coulomb Potential Energy

- Regression with wavelet energy functionals:

$$\widetilde{U}(\rho) = \sum_{j=j_{\min}}^{j=j_{\max}} \alpha_j \phi_j^2(\rho), \quad \phi_j^2(\rho) = \int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} |\rho * \psi_{j,r}(u)|^2 \, dr \, du$$

- **<u>Theorem</u>** (H., Mallat, Poilvert 2015): For all $\epsilon > 0$ there exits a scale oversampling factor $Q \in \mathbb{N}$ such that

$$|U(\rho) - \widetilde{U}(\rho)| < \epsilon \cdot \max(\|\rho\|_1^2, \|\rho\|_2^2)$$

with $|j_{\min} - j_{\max}| = O(|\log \epsilon|)$.

Quantum Wavelet and Fourier Dictionaries

- Full quantum energy is not quadratic. Need linear and quadratic terms.

- Covalent bonds between atoms dominate the energy. These involve two electrons each. Thus the the majority of the energy is proportional to the sum of the charges:

$$\phi_0(\rho) = \int_{\mathbb{R}^3} \rho(u)\, du = \sum_k q_k$$

- We complement Fourier and Wavelet dictionaries by incorporating this linear term with $\mathbf{L}^1$ and $\mathbf{L}^2$ terms.

Quantum Wavelet and Fourier Dictionaries

- Fourier $\mathbf{L}^p$ terms and dictionary:

$$\phi_{\gamma,p}(\rho) = \left(\int_{|\omega|=\gamma} |\hat{\rho}(\omega)|^p \, d\omega \right)^{1/p}$$

$$\Phi_F(\rho) = \{\phi_0(\rho),\, \phi_{m\Delta,1}(\rho),\, \phi^2_{m\Delta,1}(\rho),\, \phi^2_{m\Delta,2}(\rho)\}_{m\in\mathbb{N}}$$

- Wavelet $\mathbf{L}^p$ terms and dictionary:

$$\phi_{j,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} |\rho * \psi_{j,r}(u)|^p \, dr \, du \right)^{1/p}$$

$$\Phi_W(\rho) = \{\phi_0(\rho),\, \phi_{j,1}(\rho),\, \phi^2_{j,1}(\rho),\, \phi^2_{j,2}(\rho)\}_{j\in\mathbb{Z}}$$

Orthogonal Least Squares

- Training set: $\{(x_i, f(x_i)\}_i \mapsto \{(\tilde{\rho}_{x_i}, E(\rho_{x_i})\}_i$
- Greedy algorithm to compute M-term sparse regression:

$$\tilde{f}_M(x) = \widetilde{E}_M(\tilde{\rho}_x) = \sum_{k=1}^{M} \alpha_k \phi_{n_k}(\tilde{\rho}_x)$$

- The algorithm selects the functions $\{\phi_{n_k}\}_k$ and learns the weights $\{\alpha_k\}_k$ by minimizing

$$\sum_i |E(\rho_{x_i}) - \widetilde{E}_m(\tilde{\rho}_{x_i})|^2$$

at each iteration $m = 1, \ldots, M$

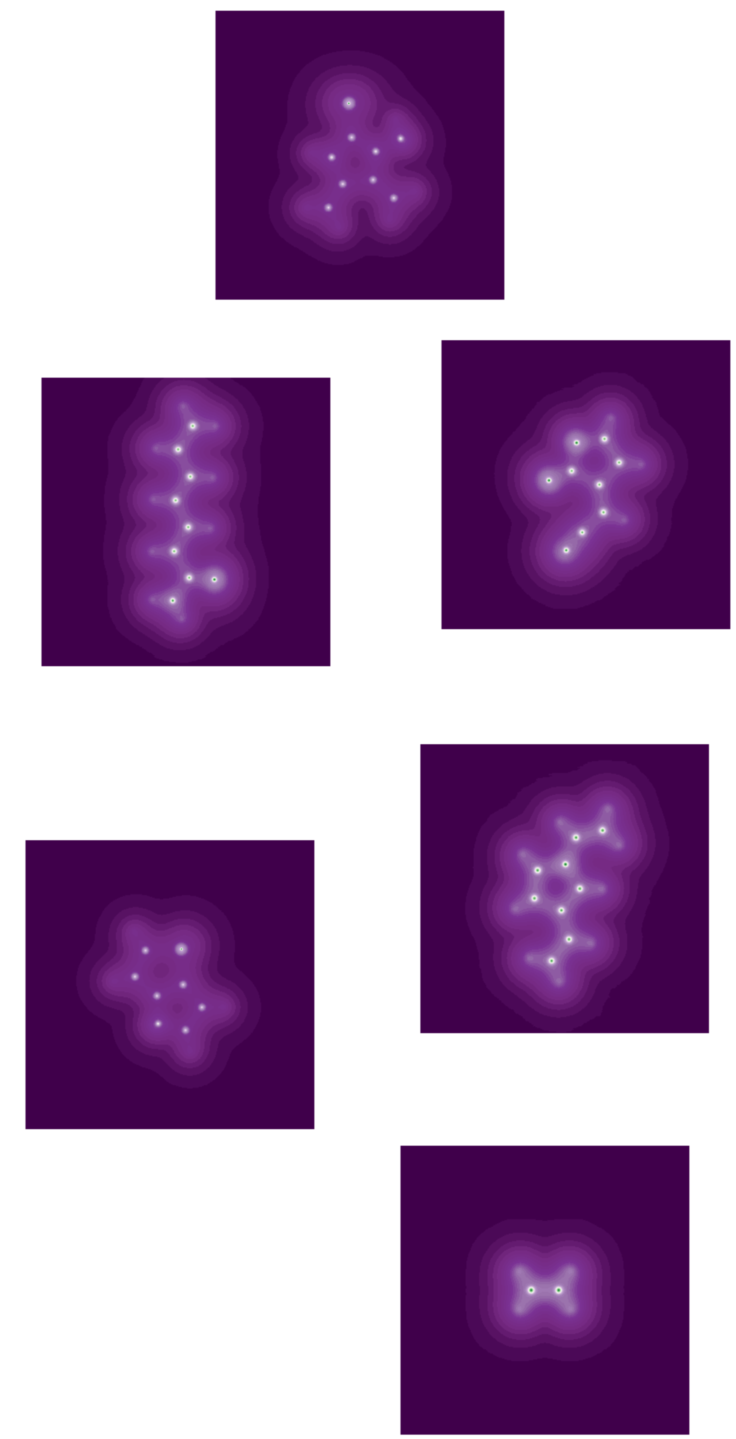
Data Set

- Data set $\{x_i, f(x_i)\}_i$ consisting of over 4000 planar organic molecules made up of hydrogen, carbon, nitrogen, oxygen, sulfur, and chlorine.

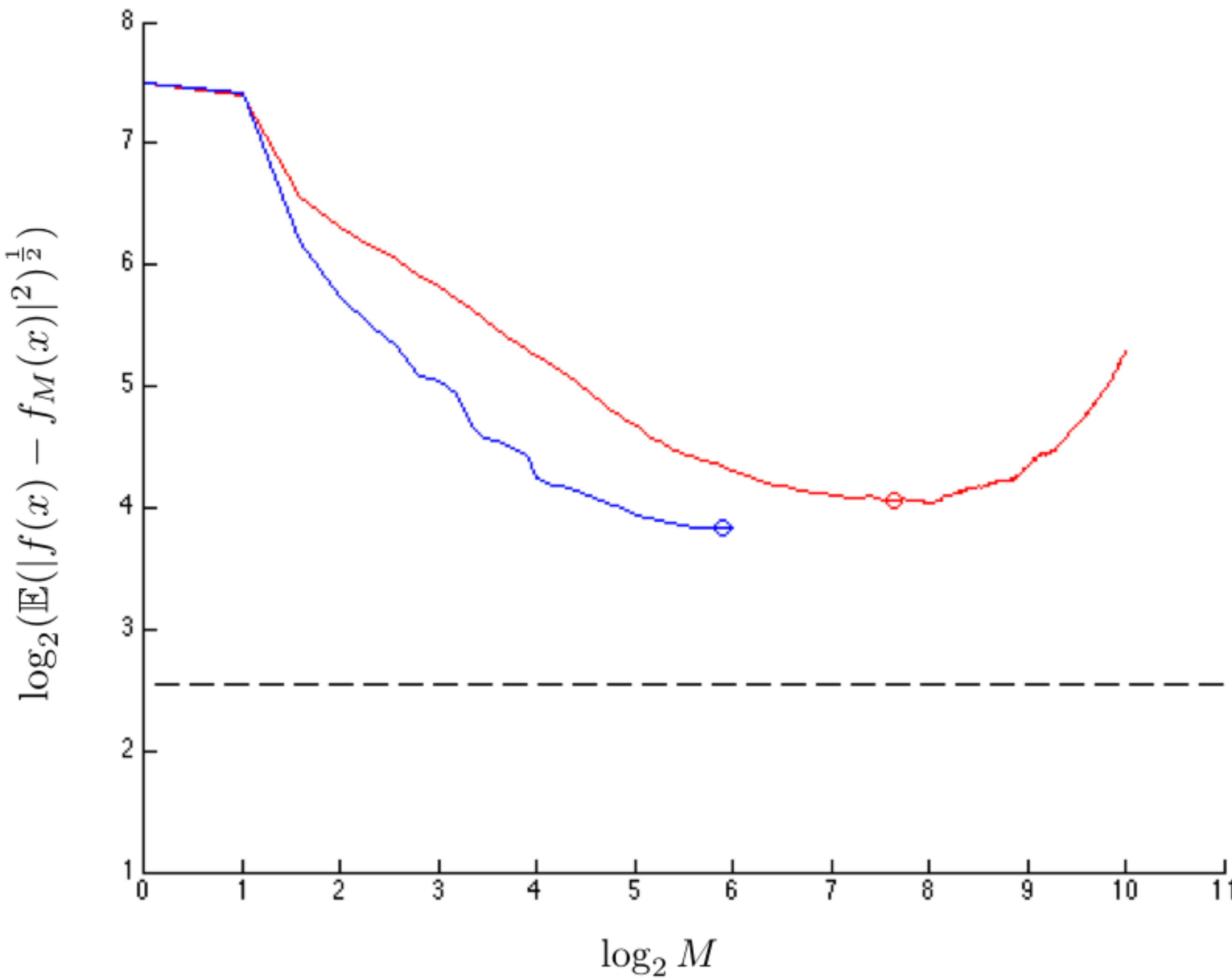
- Molecules have between 6 and 20 atoms

- Each molecule x_i is unique and in its ground state configuration (configuration that minimizes energy)

- $f(x_i)$ is the atomization energy of the molecule (energy necessary to break atomic bonds)



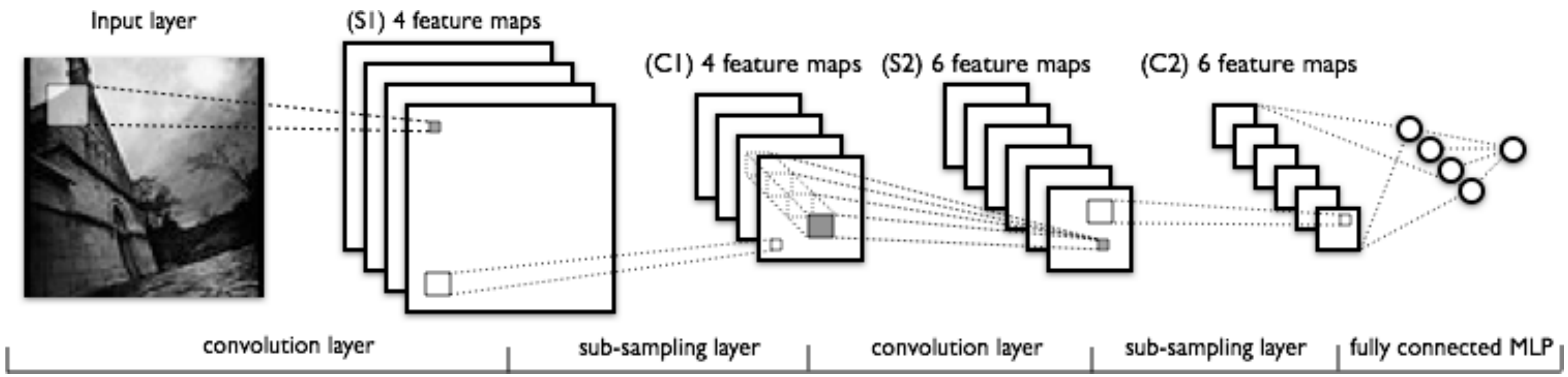
Fourier and Wavelet M-term Regression Error



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

The Scattering Transform

Deep Convolutional Networks



- Convolution layer: $h(u) = \tanh\left(g * L_k(u) + b_k\right)$
- Sub-sampling layer (nonlinear): Max pooling
- Linear filters L_k and weights b_k are learned from training data via back-propagation

Scattering Dictionary

Layer 0

$$\rho$$

$$\downarrow$$

$$\phi_0(\rho) = \int_{\mathbb{R}^3} \rho(u)\, du$$

Scattering Dictionary

Layer 0

Layer 1

$$|\rho * \psi_{j_1'',r}(u)|$$

ρ

$$|\rho * \psi_{j_1',r}(u)|$$

$$|\rho * \psi_{j_1,r}(u)|$$

$$\phi_0(\rho) = \int_{\mathbb{R}^3} \rho(u)\,du$$

$$\phi_{j_1,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} |\rho * \psi_{j_1,r}(u)|^p \, dr \, du \right)^{1/p}$$

Scattering Dictionary

Layer 0

ρ

$$\phi_0(\rho) = \int_{\mathbb{R}^3} \rho(u)\, du$$

Layer 1

$|\rho * \psi_{j_1'',r}(u)|$

$|\rho * \psi_{j_1',r}(u)|$

$|\rho * \psi_{j_1,r}(u)|$

$$\phi_{j_1,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} |\rho * \psi_{j_1,r}(u)|^p \, dr\, du \right)^{1/p}$$

Layer 2

$||\rho * \psi_{j_1'',\cdot}| \star \Psi_{\lambda_2''}(u,r)|$

$||\rho * \psi_{j_1'',\cdot}| \star \Psi_{\lambda_2'}(u,r)|$

$||\rho * \psi_{j_1'',\cdot}| \star \Psi_{\lambda_2}(u,r)|$

$||\rho * \psi_{j_1',\cdot}| \star \Psi_{\lambda_2''}(u,r)|$

$||\rho * \psi_{j_1',\cdot}| \star \Psi_{\lambda_2'}(u,r)|$

$||\rho * \psi_{j_1',\cdot}| \star \Psi_{\lambda_2}(u,r)|$

$||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2''}(u,r)|$

$||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2'}(u,r)|$

$||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2}(u,r)|$

$$\phi_{j_1,\lambda_2,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} ||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2}(u,r)|^p \, dr\, du \right)^{1/p}$$

Scattering Dictionary

Layer 0

ρ

Layer 1

$|\rho * \psi_{j_1'',r}(u)|$

$|\rho * \psi_{j_1',r}(u)|$

$|\rho * \psi_{j_1,r}(u)|$

Layer 2

$||\rho * \psi_{j_1'',\cdot}| \star \Psi_{\lambda_2''}(u,r)|$

$||\rho * \psi_{j_1'',\cdot}| \star \Psi_{\lambda_2'}(u,r)|$

$||\rho * \psi_{j_1'',\cdot}| \star \Psi_{\lambda_2}(u,r)|$

$||\rho * \psi_{j_1',\cdot}| \star \Psi_{\lambda_2''}(u,r)|$

$||\rho * \psi_{j_1',\cdot}| \star \Psi_{\lambda_2'}(u,r)|$

$||\rho * \psi_{j_1',\cdot}| \star \Psi_{\lambda_2}(u,r)|$

$||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2''}(u,r)|$

$||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2'}(u,r)|$

$||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2}(u,r)|$

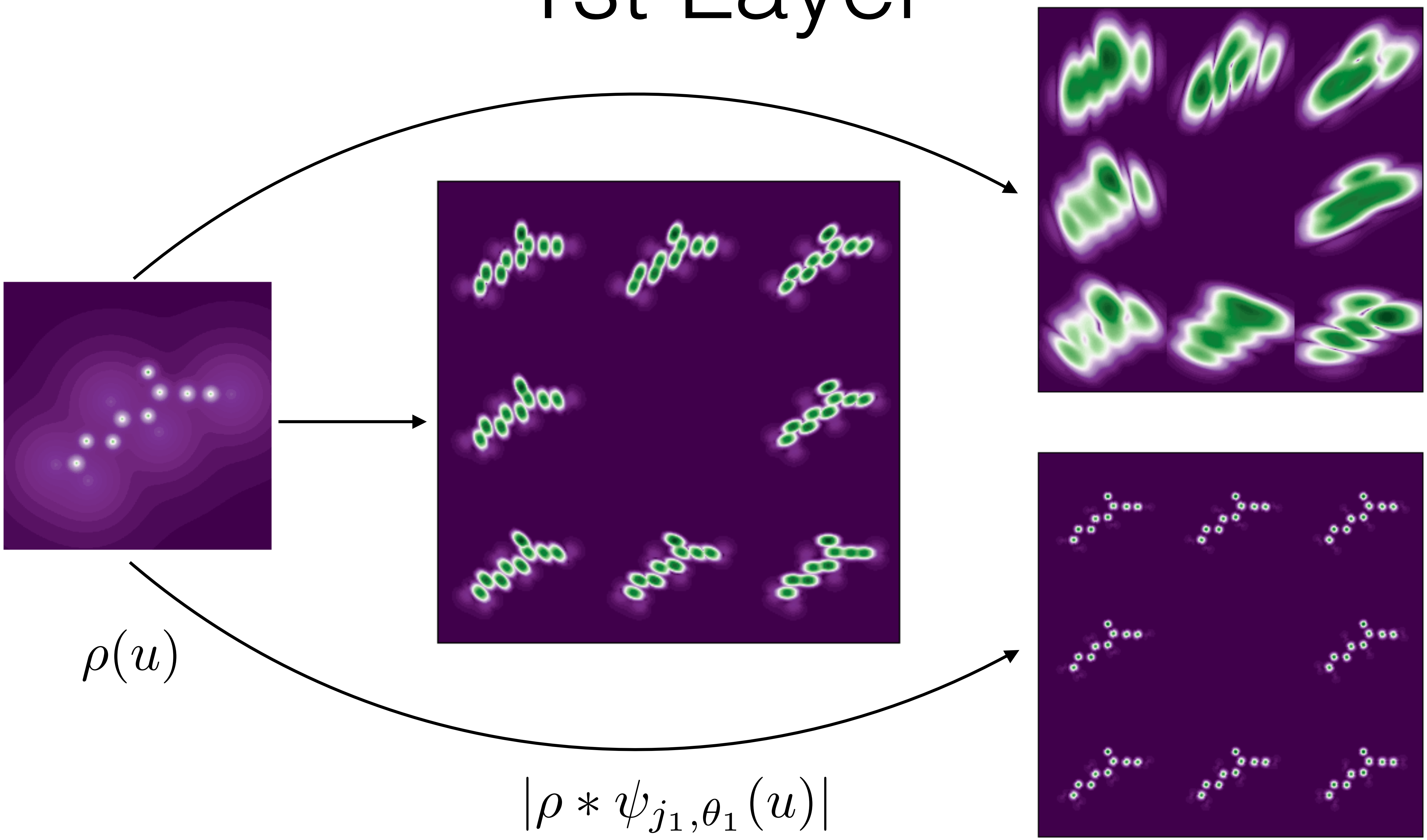
$$\phi_0(\rho) = \int_{\mathbb{R}^3} \rho(u)\, du$$

$$\phi_{j_1,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} |\rho * \psi_{j_1,r}(u)|^p \, dr\, du \right)^{1/p}$$

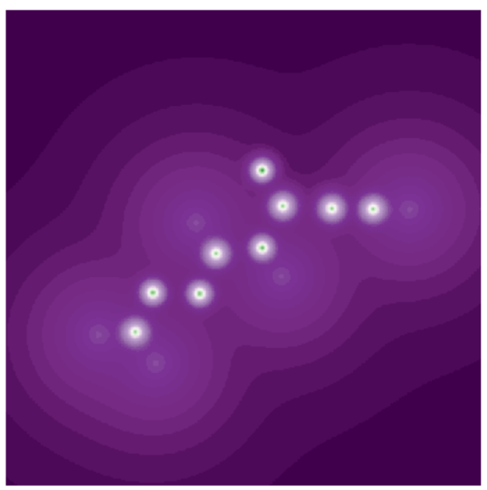
$$\phi_{j_1,\lambda_2,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} ||\rho * \psi_{j_1,\cdot}| \star \Psi_{\lambda_2}(u,r)|^p \, dr\, du \right)^{1/p}$$

$$\Phi_S(\rho) = \{\phi_0(\rho), \phi_{j_1,1}(\rho), \phi^2_{j_1,1}(\rho), \phi^2_{j_1,2}(\rho), \phi_{j_1,\lambda_2,1}(\rho), \phi^2_{j_1,\lambda_2,1}(\rho), \phi^2_{j_1,\lambda_2,2}(\rho)\}_{j_1,\lambda_2}$$

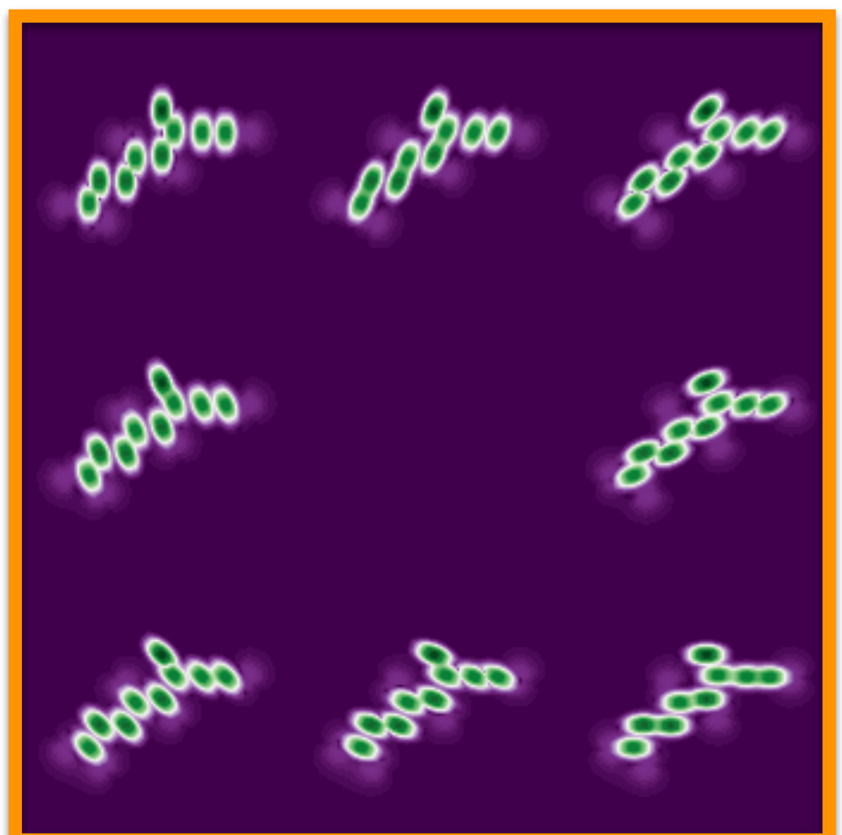
Scattering in 2D: 1st Layer



Scattering in 2D: 1st Layer



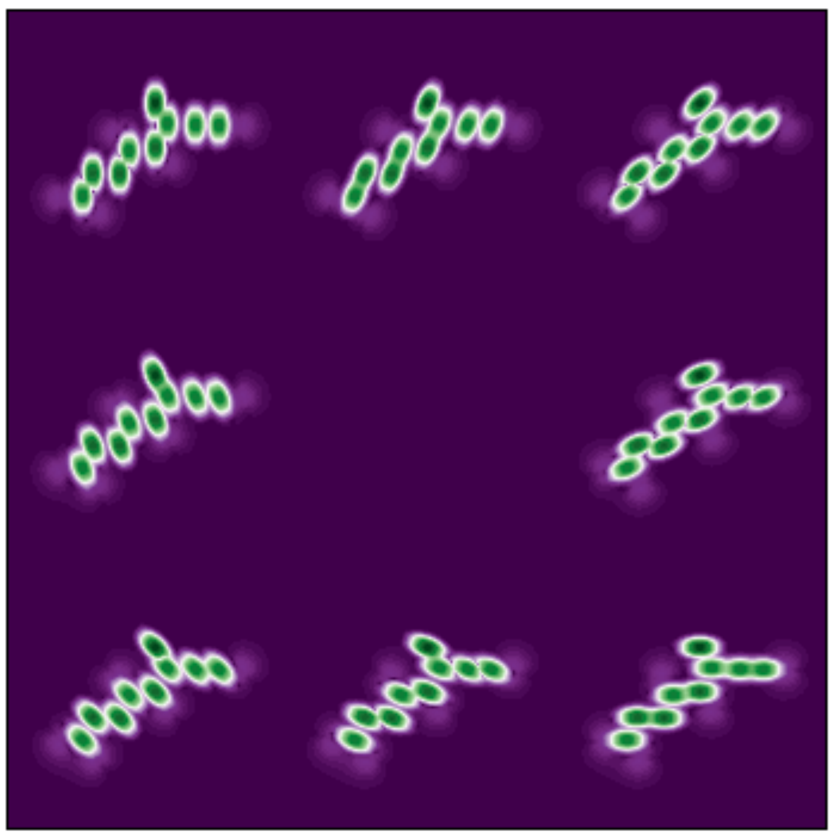
$\rho(u)$



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

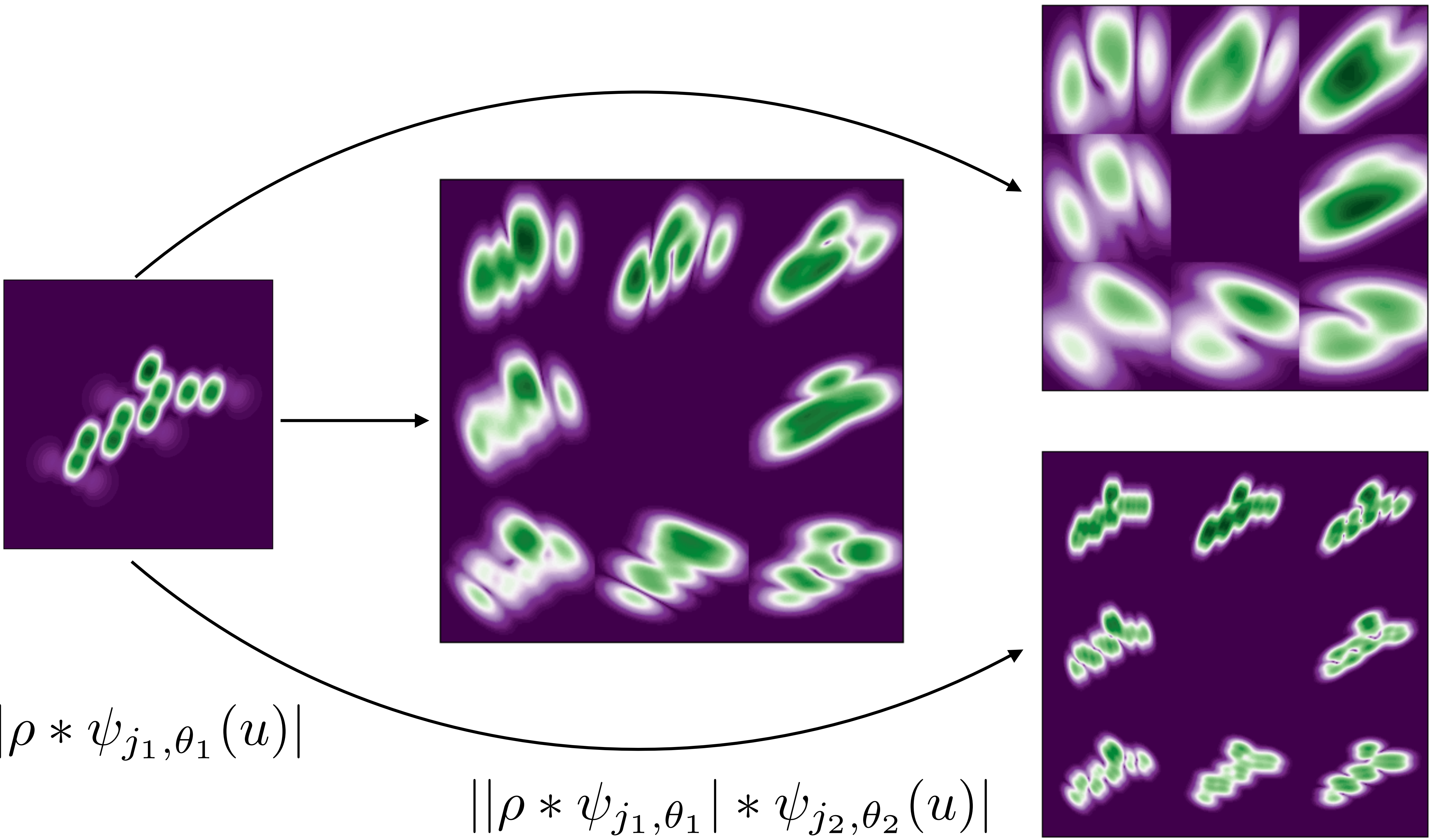
$$\phi^{p}_{j_1,p}(\rho) =$$
$$= \int_{\mathbb{R}^2} \int_0^{2\pi} |\rho * \psi_{j_1,\theta_1}(u)|^p \, d\theta_1 \, du$$

Scattering in 2D: 2nd Layer Translation Variability



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

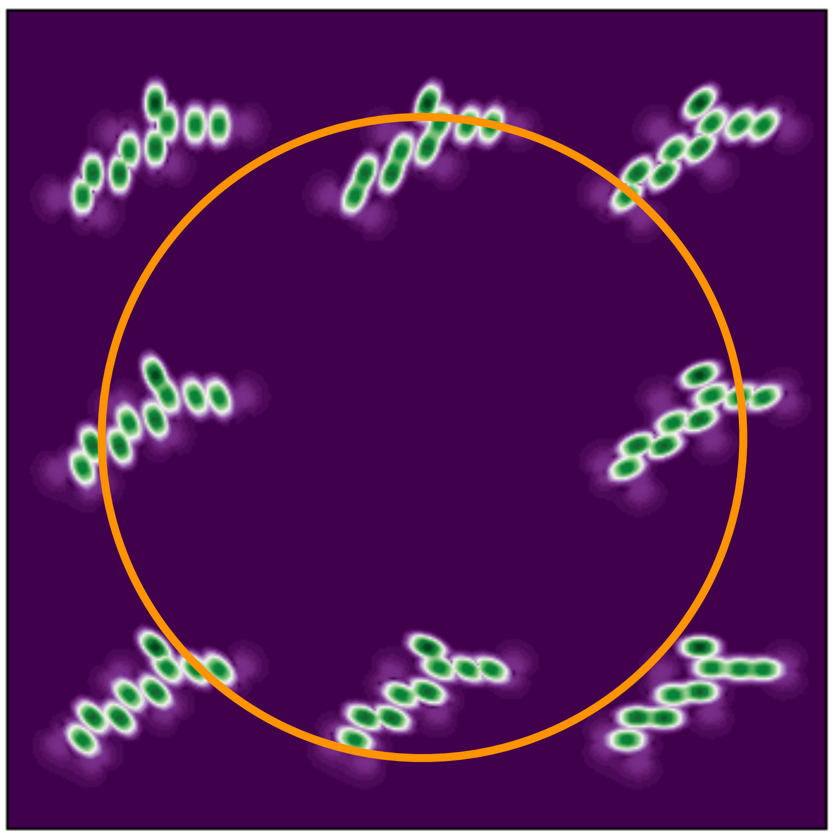
Scattering in 2D: 2nd Layer Translation Variability



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

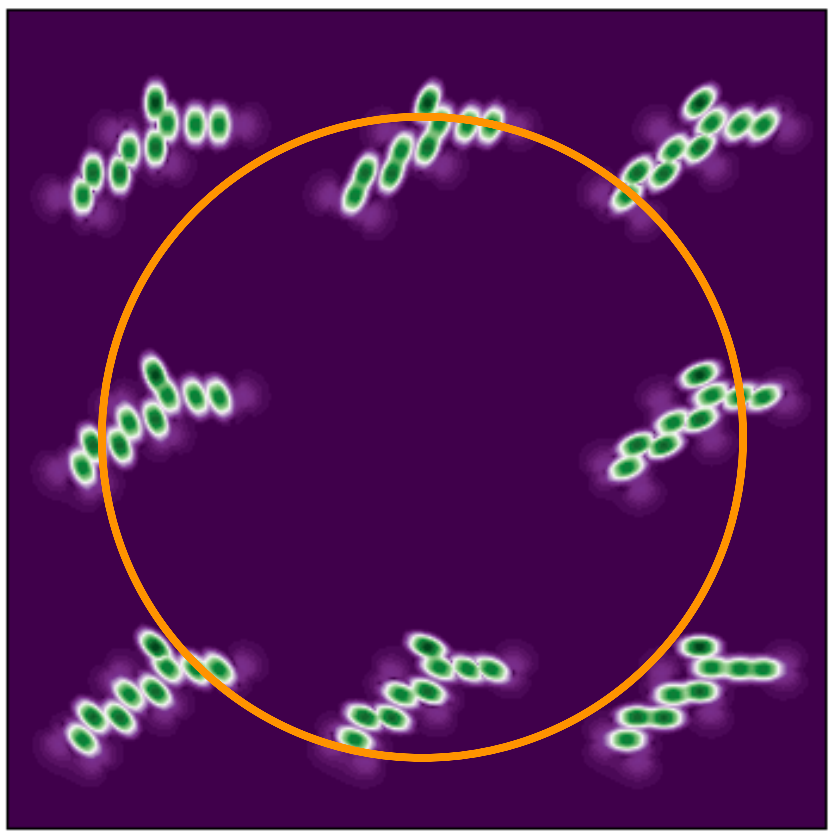
$||\rho * \psi_{j_1,\theta_1}| * \psi_{j_2,\theta_2}(u)|$

Scattering in 2D:
2nd Layer Rotation Variability



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

Scattering in 2D: 2nd Layer Rotation Variability



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

- 1D wavelet $\psi^{1\mathrm{D}}$ periodized over $[0, 2\pi)$:

$$\overline{\psi}_l(\theta) = \sum_{k \in \mathbb{Z}} \psi_l^{1\mathrm{D}}(\theta - 2\pi k)$$

- 2nd layer wavelet transform over angles defined in terms of circular convolution:

$$|\rho * \psi_{j_1,\cdot}(u)| \circledast \overline{\psi}_{l_2}(\theta_1)$$

Scattering in 2D: Roto-Translation 2nd Layer

- Spatial 2D convolution to recover translation variability:

$$|\rho * \psi_{j_1,\theta_1}| * \psi_{j_2,\theta_2}(u)$$

- Circular 1D convolution to recover rotation variability:

$$|\rho * \psi_{j_1,\cdot}(u)| \circledast \overline{\psi}_{l_2}(\theta_1)$$

- Combining yields a 3D convolution to recover roto-translation variability:

$$||\rho * \psi_{j_1,\cdot}(u)| \star \Psi_{j_2,\theta_2,l_2}(u,\theta_1)| = ||\rho * \psi_{j_1,\cdot}| * \psi_{j_2,\theta_2}(u) \circledast \overline{\psi}_{l_2}(\theta_1)|$$

$$\text{where:}$$

$$\Psi_{j_2,\theta_2,l_2}(u,\theta) = \psi_{j_2,\theta_2}(u)\overline{\psi}_{l_2}(\theta)$$

$$\star = (*, \circledast)$$

Scattering in 3D: 1st Layer

- $\mathrm{E}(3) = \mathbb{R}^3 \rtimes \mathrm{O}(3)$ and $\mathrm{O}(3) = S^2 \rtimes \mathrm{O}(2)$

- If we use a wavelet ψ that is radially symmetric about an axis η_0, then we can ignore the $\mathrm{O}(2)$ component since ψ will not vary over $\mathrm{O}(2)$

$$\text{if } r\eta_0 = \eta_0 \text{ then } \psi(ru) = \psi(u), \;\; r \in \mathrm{O}(3)$$

$$\psi(u) = g(u)(e^{i\eta_0 \cdot u} - C)$$

- For the first layer wavelet transform, this means we can index the rotation by $\eta \in S^2$:

$$\psi_{j,r}(u) = \psi_{j,\eta}(u) = 2^{-3\frac{j}{Q}} \psi(2^{-\frac{j}{Q}} r^{-1} u), \;\; \eta = r\eta_0 \in S^2, \;\; j \in \mathbb{Z}$$

$$\rho(u) \mapsto |\rho * \psi_{j,\eta}(u)|$$

$$\phi_{j,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{S^2} |\rho * \psi_{j,\eta}(u)|^p \, d\eta \, du \right)^{1/p}$$

Scattering in 3D: 2nd Layer

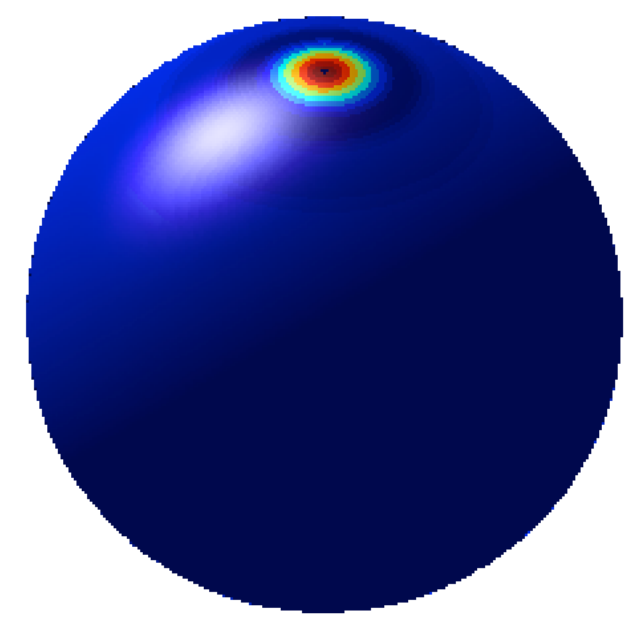
- The second layer can be computed as two separable wavelet transforms, one over translations ($\mathbb{R}^3$) and one over rotations (S^2).

- Isotropic wavelet over S^2:

$$\overline{\psi}_{l,\nu} : S^2 \to \mathbb{R}, \text{ scale } 2^l \text{ and translation } \nu \in S^2$$

- Wavelet transform over $\mathbb{R}^3$ with the same Morlet wavelet:

$$|\rho * \psi_{j_1,\eta}| * \psi_{j_2,\eta_2}(u)$$



- Followed by the wavelet transform over S^2:

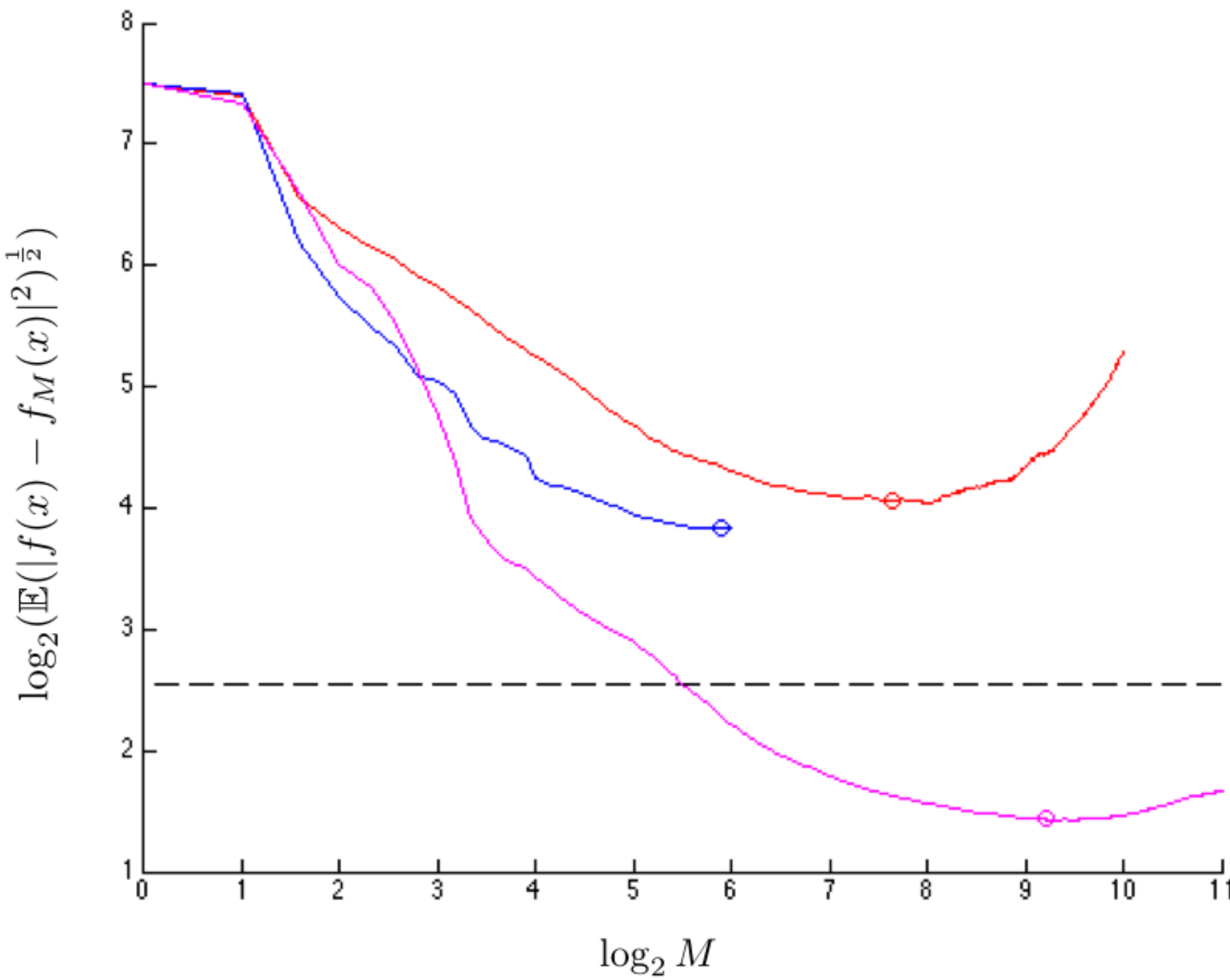
$$\int_{S^2} |\rho * \psi_{j_1,\eta}| * \psi_{j_2,\eta_2}(u) \overline{\psi}_{l_2,\nu}(\eta)\, d\eta$$

- Second layer functionals:

$$\phi_{j_1,j_2,\eta_2,l_2,p}(\rho) = \left(\int_{\mathbb{R}^3} \int_{S^2} \left| \int_{S^2} |\rho * \psi_{j_1,\eta}| * \psi_{j_2,\eta_2}(u) \overline{\psi}_{l_2,\nu}(\eta)\, d\eta \right|^p d\nu\, du \right)^{1/p}$$

Numerical Results

Scattering
M-term Regression Error



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

Numerical Results

	Coulomb	Fourier	Wavelet	Scattering	Chemical Accuracy
ℓ^1: MAE	2.4	11	11	**1.8**	1.0
ℓ^2: RMSE	5.8	17	14	**2.7**	
ℓ^∞: Max	224	272	97	**42**	

Error in kcal/mol

- Scattering terms:
  - First term is total charge: $\phi_0(\rho) = \int_{\mathbb{R}^3} \rho(u)\, du = \sum_k q_k$
  - Other selected terms correspond to important geometric scales that range over the distance between two neighbouring atoms and the diameter of the molecule

Conclusion

- The scattering transform defines a representation that captures the fundamental properties of the quantum molecular energy, and which is sufficiently rich to achieve highly accurate energy estimates.
- One can learn physics through data and compute fast.
- Can we learn other physical functionals?

http://www.di.ens.fr/~hirn/