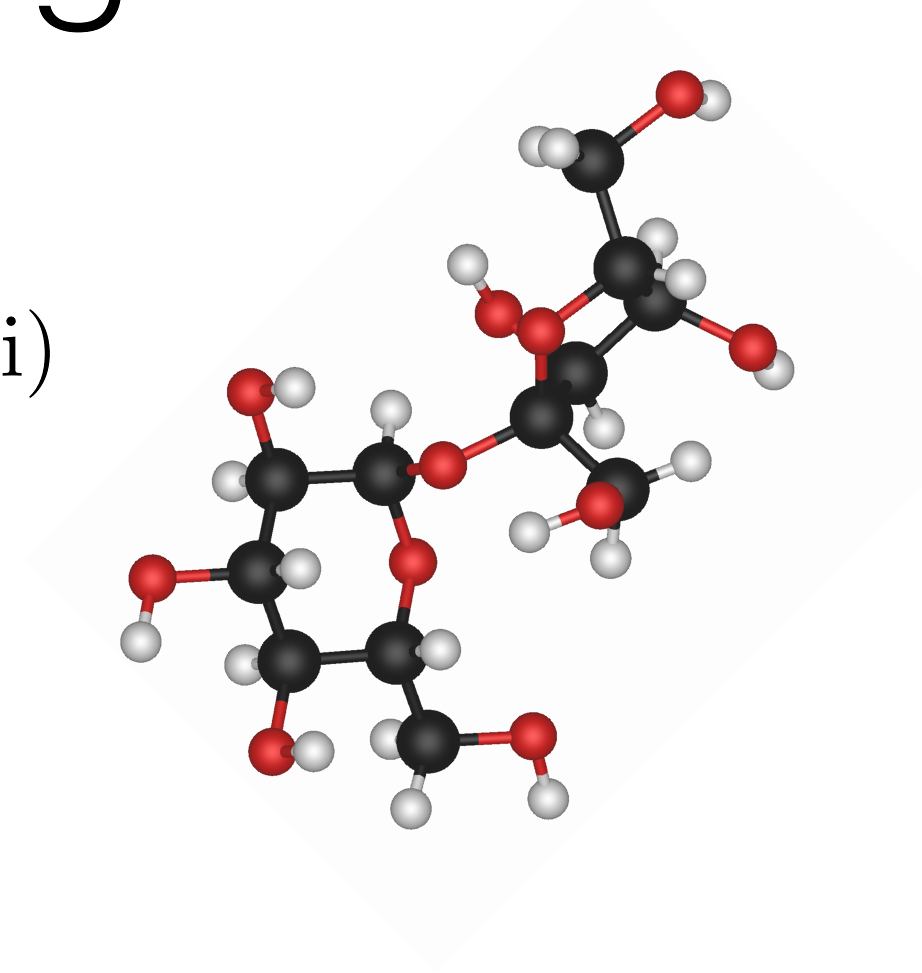
Regression of Quantum Energies by Scattering

Matthew Hirn, Stéphane Mallat, Nicolas Poilvert

May 7, 2015

Quantum Energy Regression



- $x = \text{state of molecule}$

  $= \text{(nuclear charges, positions of nuclei)}$

  $= \{(z_k, r_k) \in \mathbb{R} \times \mathbb{R}^3\}_k$
- Energy of the molecule:
  Schrödinger's Equation

$$\widehat{H}\Psi = E\Psi, \quad E = E(x)$$

- Instead of using quantum mechanical computations to compute the energy, want to interpolate the energy from a set of know examples.

- Interpolation is faster, but can we make it accurate?

Quantum Energy Regression

- Given a *small* training set of known examples:

$$\text{Training Set} = \{(x_i, E(x_i)\}_i$$

- Using knowledge of the physics, construct a new nonlinear representation of the molecular state

$$\Phi(x) = \{\phi_p(x)\}_p$$

- Linearly regress the energy over this new dictionary:

$$\widetilde{E}(x) = \sum_p \alpha_p \phi_p(x)$$

$$\text{Learn the weights } \{\alpha_p\}_p \text{ from the training set}$$

- Computational cost is the cost of computing $\Phi(x)$

Representation Properties

$$\text{Regression: } \widetilde{E}(x) = \sum_p \alpha_p \phi_p(x), \quad x = \{(z_k, r_k)\}_k$$

1. **Sparse Regression:**
   Can only learn a few weights from limited training

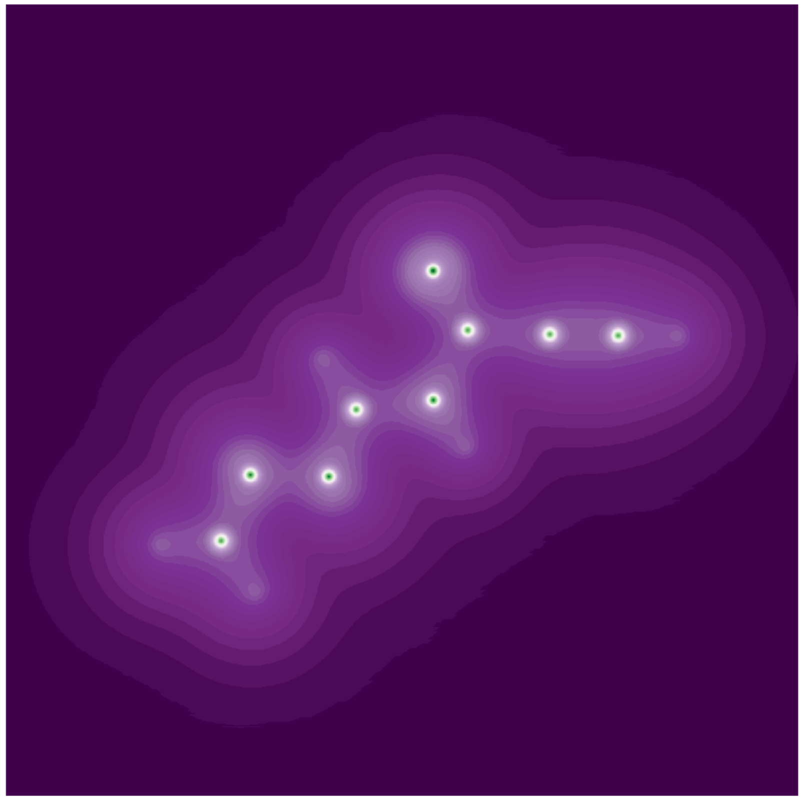
2. **Permutation Invariance:**
   Invariant to permutations of the indexation of the atoms in each molecule

3. **Isometry Invariance:**
   Invariant to actions of the isometry group on the molecular state

4. **Deformation Stability:**
   Differentiable with respect to the positions of the atoms

5. **Multi-scale Interactions:**
   - Highly energetic covalent bonds between neighboring atoms
   - Weaker energetic exchanges at longer distances

Permutation Invariance: Density Functional Theory



- Molecular state:

$$x = \{(z_k, r_k)\}_k$$

- Electronic density:

$$x \mapsto \rho_x(u)$$

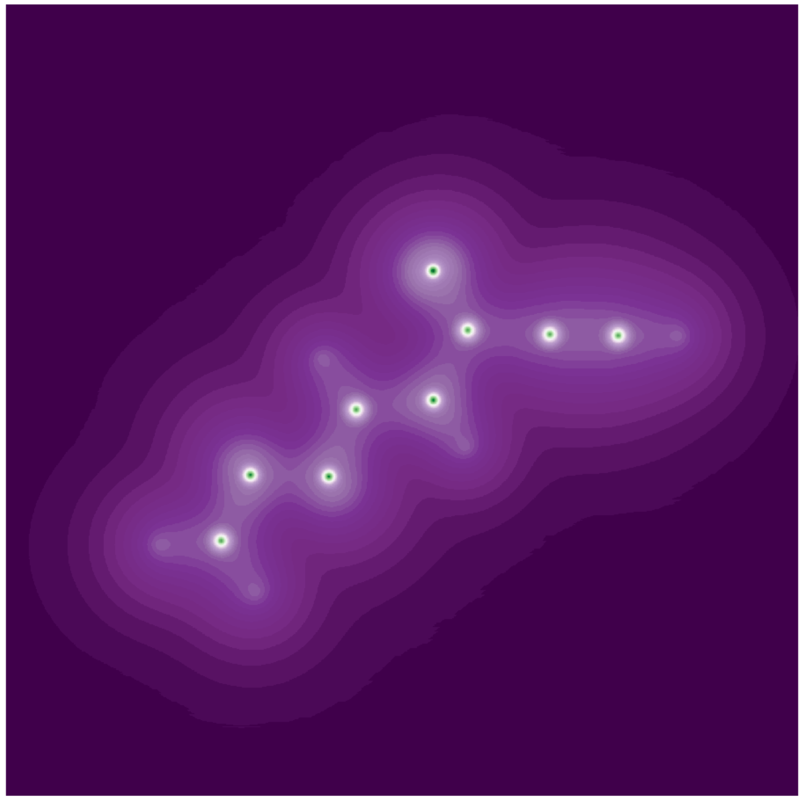
- Hohenberg and Kohn 1964:

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

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

Permutation Invariance: Density Functional Theory

- ρ_x is permutation invariant, but still need other properties
- Given ρ_x, we could interpolate $E(\rho_x)$ in a dictionary $\Phi(\rho_x) = \{\phi_p(\rho_x)\}_p$
- Computing ρ_x requires complex quantum mechanical calculations. Need a substitute that we can compute fast.



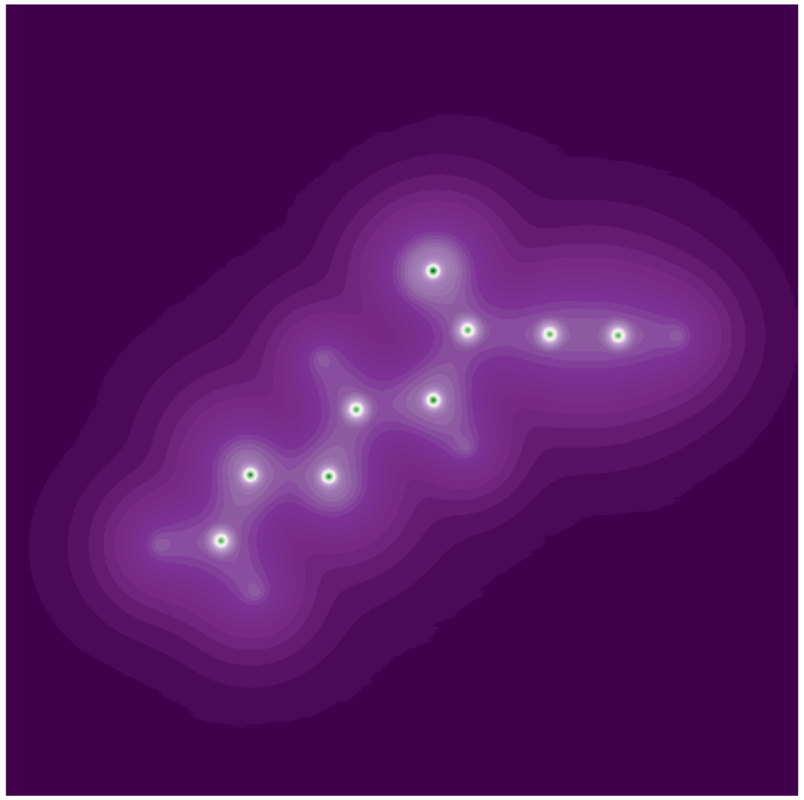
$$E(\rho) = \underbrace{T(\rho)}_{\text{Kinetic energy}} + \underbrace{\int_{\mathbb{R}^3} \rho(u) V_{\mathrm{e}}(u)\, du}_{\substack{\text{External energy} \\ \text{(electron-nuclei attraction)}}} + \underbrace{\frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho(u)\rho(v)}{|u-v|}\, du\, dv}_{\substack{\text{Coulomb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

Approximate Electronic Density

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

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

$$\rho_a = \text{exact electronic density of atom } a$$



$\rho_x(u)$

- The regression is then computed as:

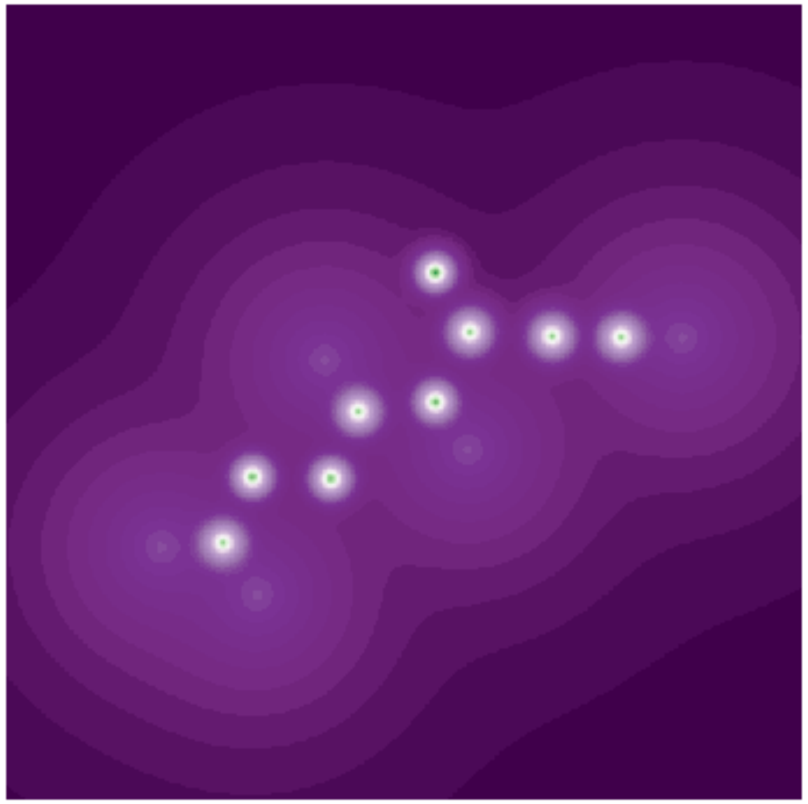
$$E(x) = E(\rho_x) \approx \widetilde{E}(\tilde{\rho}_x) = \sum_p \alpha_p \phi_p(\tilde{\rho}_x)$$

Approximate Electronic Density

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

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

$$\rho_a = \text{exact electronic density of atom } a$$



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

- The regression is then computed as:

$$E(x) = E(\rho_x) \approx \widetilde{E}(\tilde{\rho}_x) = \sum_p \alpha_p \phi_p(\tilde{\rho}_x)$$

$$\text{What is the dictionary } \Phi = \{\phi_p\}_p?$$

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 frequency:

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

Fourier Regression of Coulomb Potential Energy

- Coulomb energy diagonalized in Fourier:

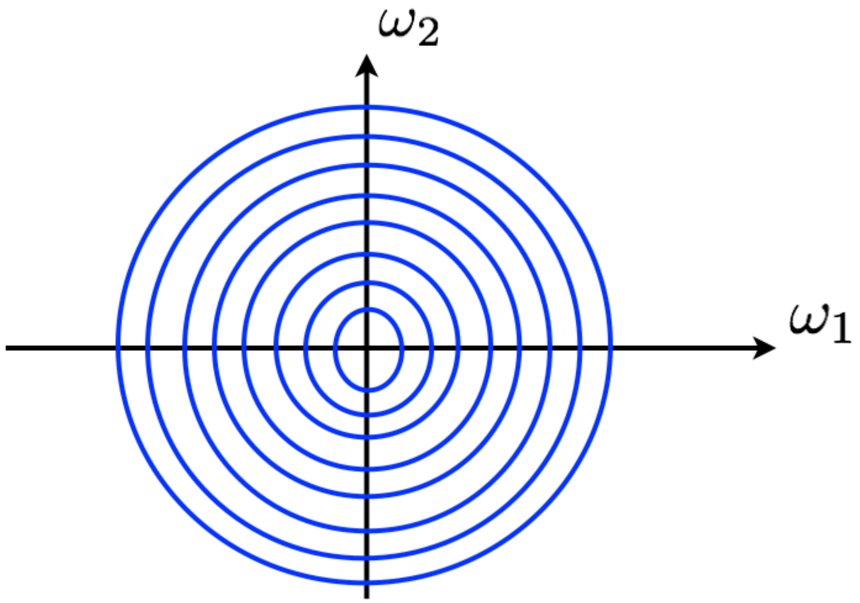
$$U(\rho) = \frac{1}{2(2\pi)^3} \int_{\mathbb{R}^3} |\hat{\rho}(\omega)|^2 \widehat{V}(\omega)\, d\omega \qquad \widehat{V}(\omega) = C|\omega|^{-2}$$

- **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:

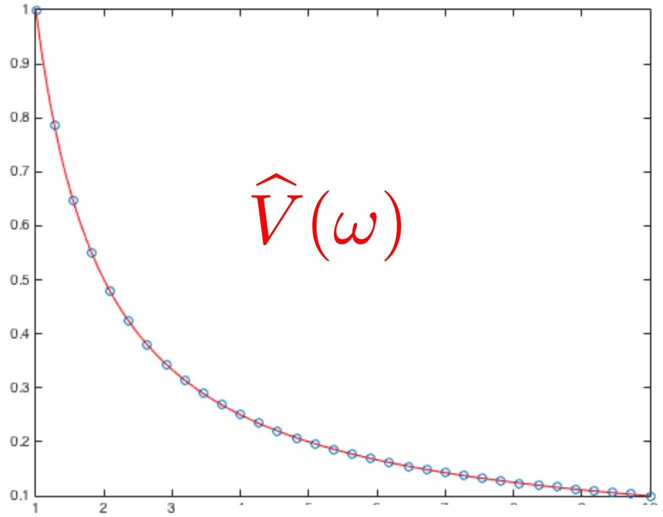
$$\widetilde{U}(\rho) = \frac{\Delta}{2(2\pi)^3} \sum_{m=1}^{M} \widehat{V}(m\Delta)\phi_{m\Delta}^2(\rho)$$



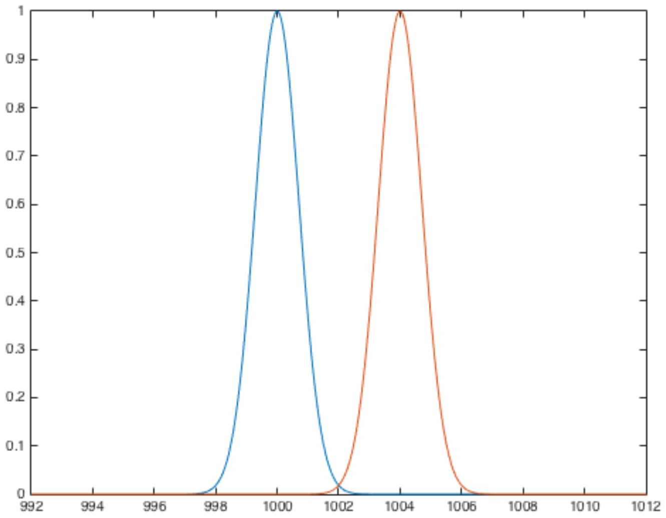
- $U(\rho) = (1 + O(\epsilon))\widetilde{U}(\rho) \Rightarrow M = O(\epsilon^{-2})$

Fourier Limitations

- The Fourier representation does not take advantage of the regularity of $\widehat{V}(\omega)$ away from $\omega = 0$. Therefore it is not sparse.



- Deformations produce instabilities at large distances, so it cannot model long range interactions:



$r_k - r_l = 1000$

$\sigma = 2$

$\epsilon = 0.004$

Dictionary $\phi_\omega(\rho) = |\hat{\rho}(\omega)|^2$ is equivalent to $\phi_u(\rho) = \rho * \bar{\rho}(u)$

$\rho(u) =$ bumps of width σ at positions $\{r_k\}_k$

$\phi_u(\rho) =$ bumps of width 2σ at positions $\{r_k - r_l\}_{k,l}$

A small diffeomorphism changes distances by $\epsilon(r_k - r_l)$

Unstable if $|r_k - r_l| \geq 2\sigma/\epsilon$

Wavelets

- Complex valued Morlet wavelet:

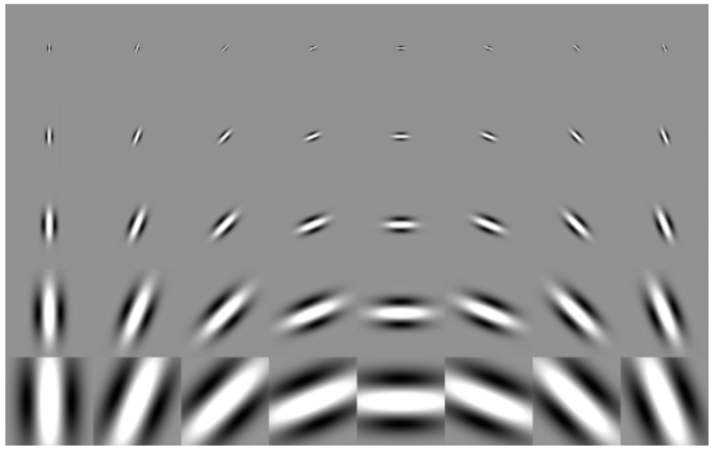
$$\psi(u) = g(u)(e^{i\eta_0 \cdot u} - C), \quad \int_{\mathbb{R}^3} \psi(u)\, du = 0$$

- Wavelet transform dilates and rotates the wavelet:

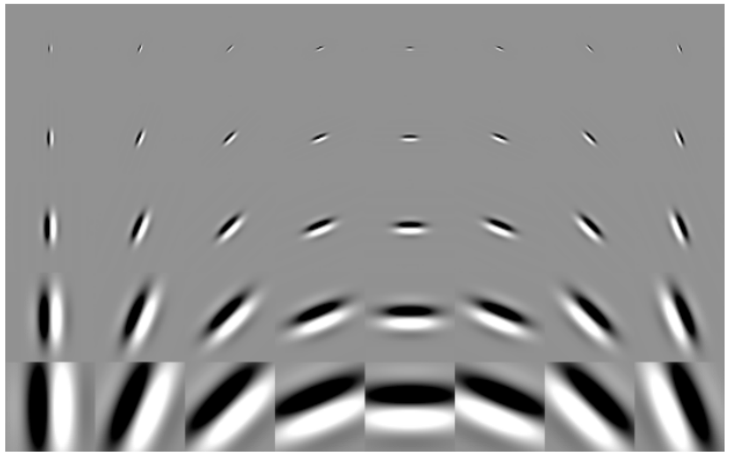
$$\psi_{j,r}(u) = 2^{-3\frac{j}{Q}} \psi(2^{-\frac{j}{Q}} r^{-1} u), \quad (j, r) \in \mathbb{Z} \times \mathrm{O}(3)$$

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

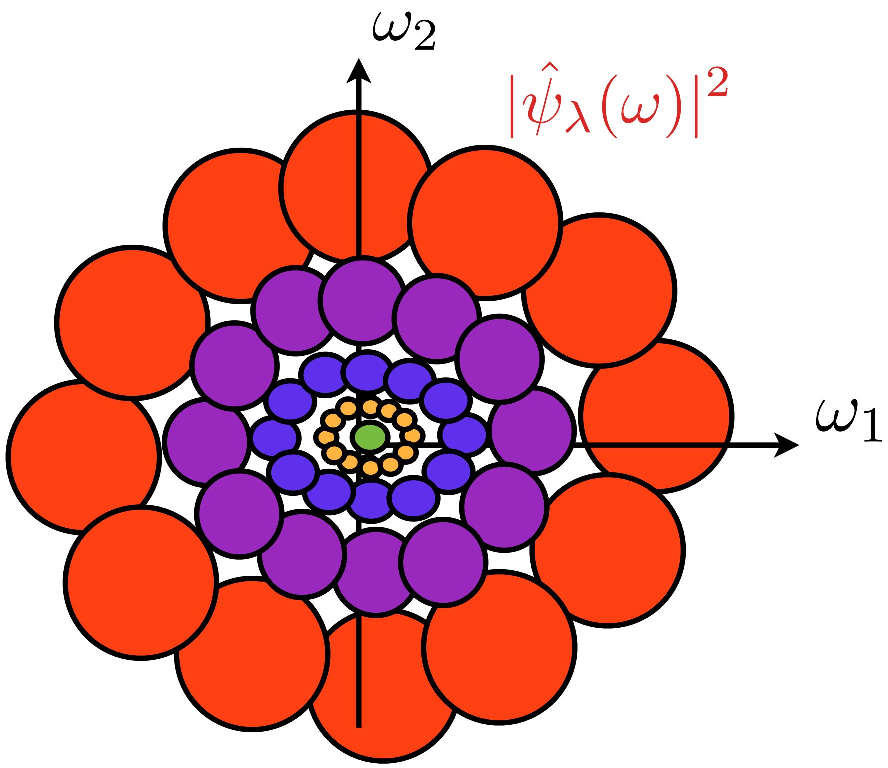
$$W[j,r]\rho(u) = \{\rho * \psi_{j,r}(u)\}_{j \in \mathbb{Z},\, r \in \mathrm{O}(3),\, u \in \mathbb{R}^3}$$



Real parts

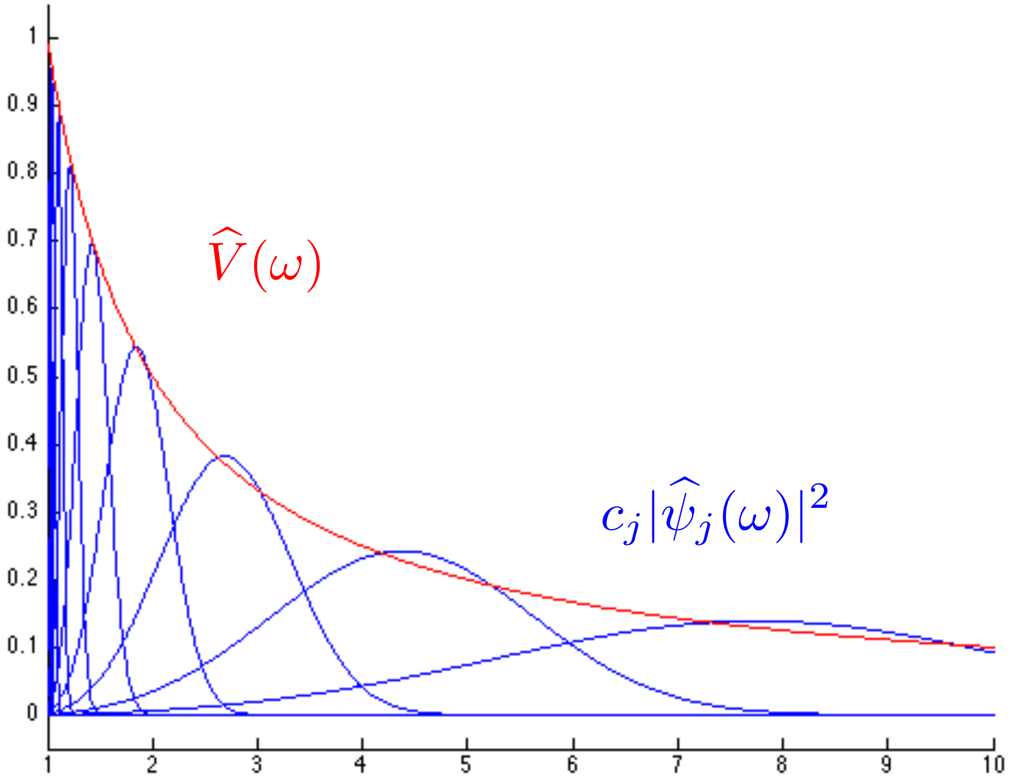


Imaginary parts



Fourier vs Wavelets

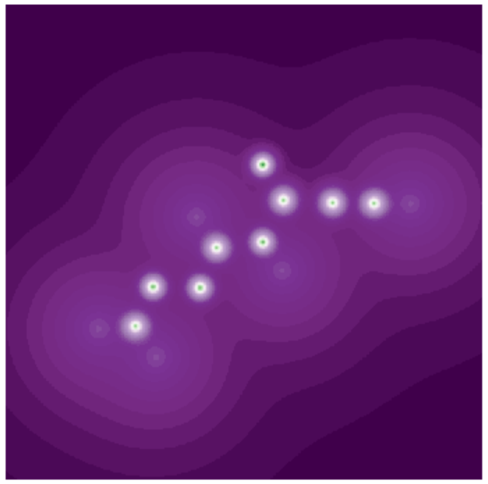
- Wavelets separate scales logarithmically and can thus take advantage of the multi-scale structure of the energy. For the Coulomb potential energy, wavelets take advantage of the regularity of $\widehat{V}(\omega)$ away from $\omega = 0$.



- Mallat 2012: Wavelets are stable. They 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 Transform: Interactions of Waves

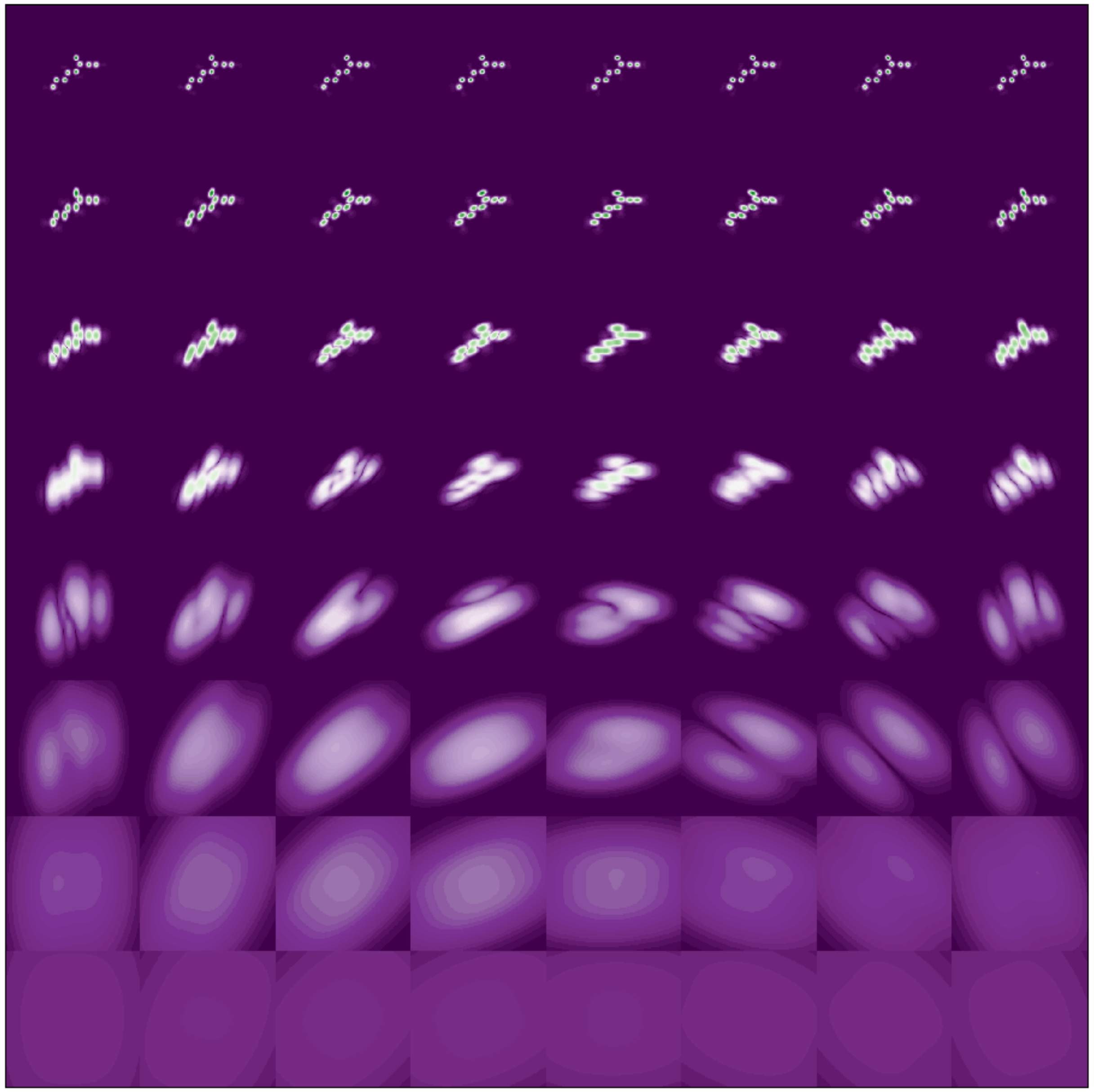


$\rho(u)$

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

$\longrightarrow$

Rotations θ_1



Scales j_1

Wavelet Regression of Coulomb Potential Energy

Definition: Isometry Invariant Wavelet Representation

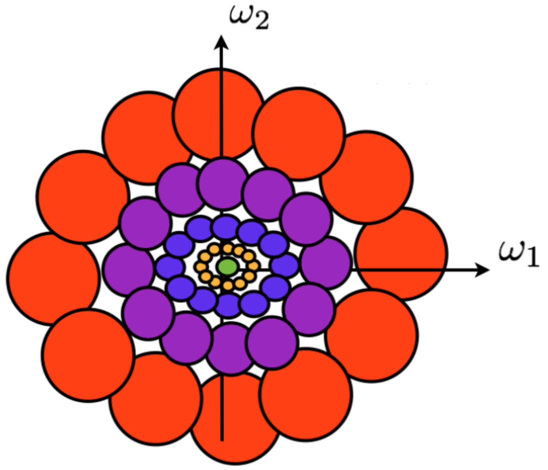
Define the $\mathbf{L}^2$ *Isometry Invariant Wavelet Representation* $\Phi = \{\phi_j^2\}_{j \in \mathbb{Z}}$ as

$$\phi_j^2(\rho) = \int_{\mathbb{R}^3} \int_{\mathrm{O}(3)} |\rho * \psi_{j,r}(u)|^2 \, dr \, du$$

Theorem: Wavelet Coulomb Regression (H., Mallat, Poilvert 2015)

For a scale oversampling factor $Q \in \mathbb{N}$, a minimum scale $j_{\text{min}} \in \mathbb{Z}$, and a maximum scale $j_{\text{max}} \in \mathbb{Z}$, define the Wavelet Regression of the Coulomb potential energy as:

$$\widetilde{U}(\rho) = \sum_{j=j_{\text{min}}}^{j_{\text{max}}} (c_1 2^{2j/Q} + c_2)\phi_j^2(\rho)$$



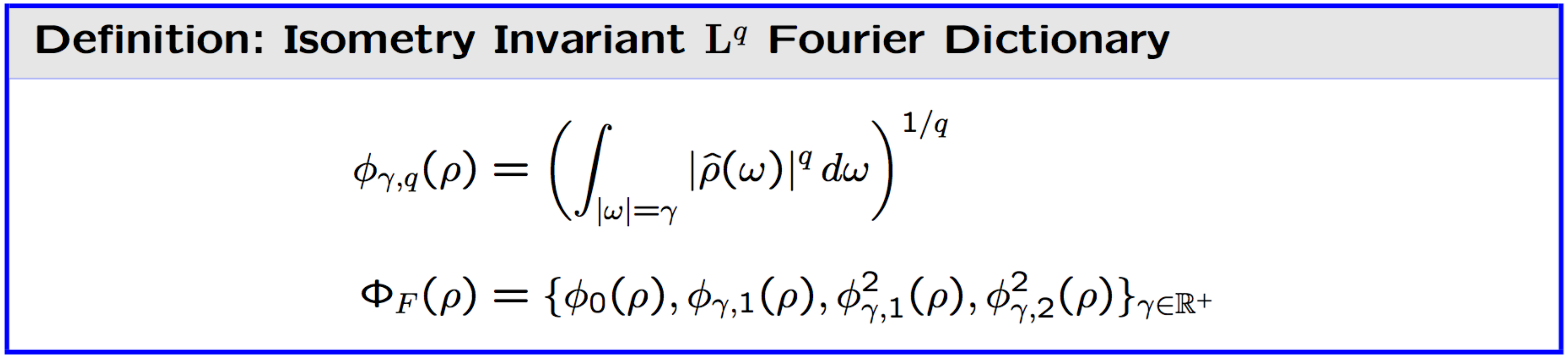
Then, for all $\epsilon > 0$, there exists Q, j_{min} and j_{max} such that

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

The number of terms grows as $|j_{\text{min}} - j_{\text{max}}| = O(|\log \epsilon|)$.

Quantum Wavelet and Fourier Dictionaries

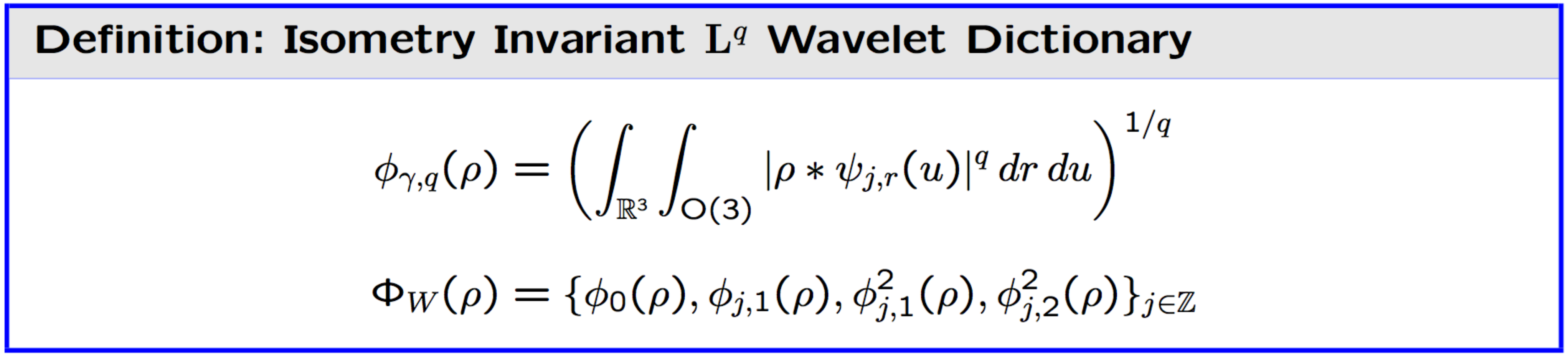
- Full quantum energy is not solely quadratic!
  - Coulomb energy is quadratic in ρ
  - Chemical (covalent) bond energy grows linearly in ρ



Definition: Isometry Invariant L^q Fourier Dictionary

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

$$\Phi_F(\rho) = \{\phi_0(\rho), \phi_{\gamma,1}(\rho), \phi^2_{\gamma,1}(\rho), \phi^2_{\gamma,2}(\rho)\}_{\gamma \in \mathbb{R}^+}$$



Definition: Isometry Invariant L^q Wavelet Dictionary

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

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

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

Learning the Weights

- Training set: $\{(x_i, E(x_i)\}_i \mapsto \{(\tilde{\rho}_{x_i}, E(\rho_{x_i})\}_i$
- Want an optimal M-term sparse regression that will minimize the error over the training set (NP hard!):

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

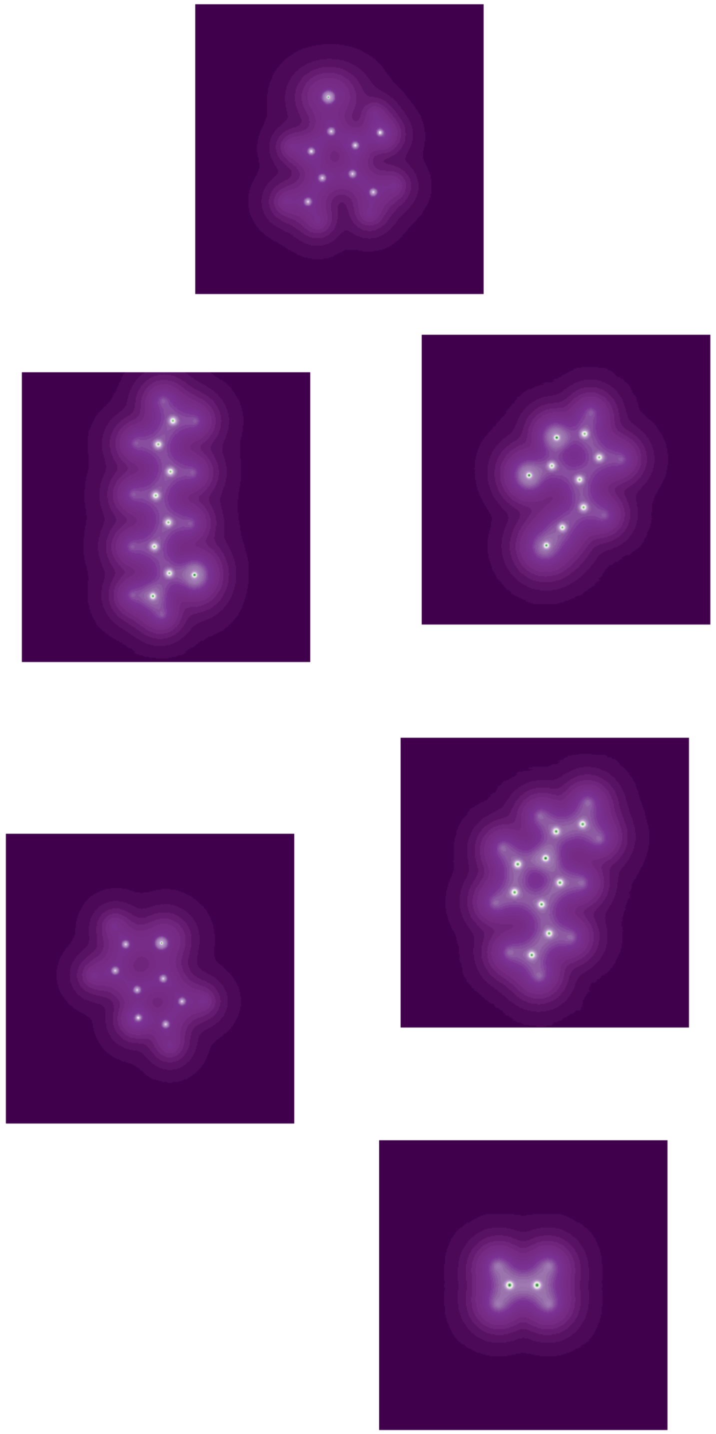
- Orthogonal least squares greedy algorithm selects the functionals and learns the weights one at a time by minimizing

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

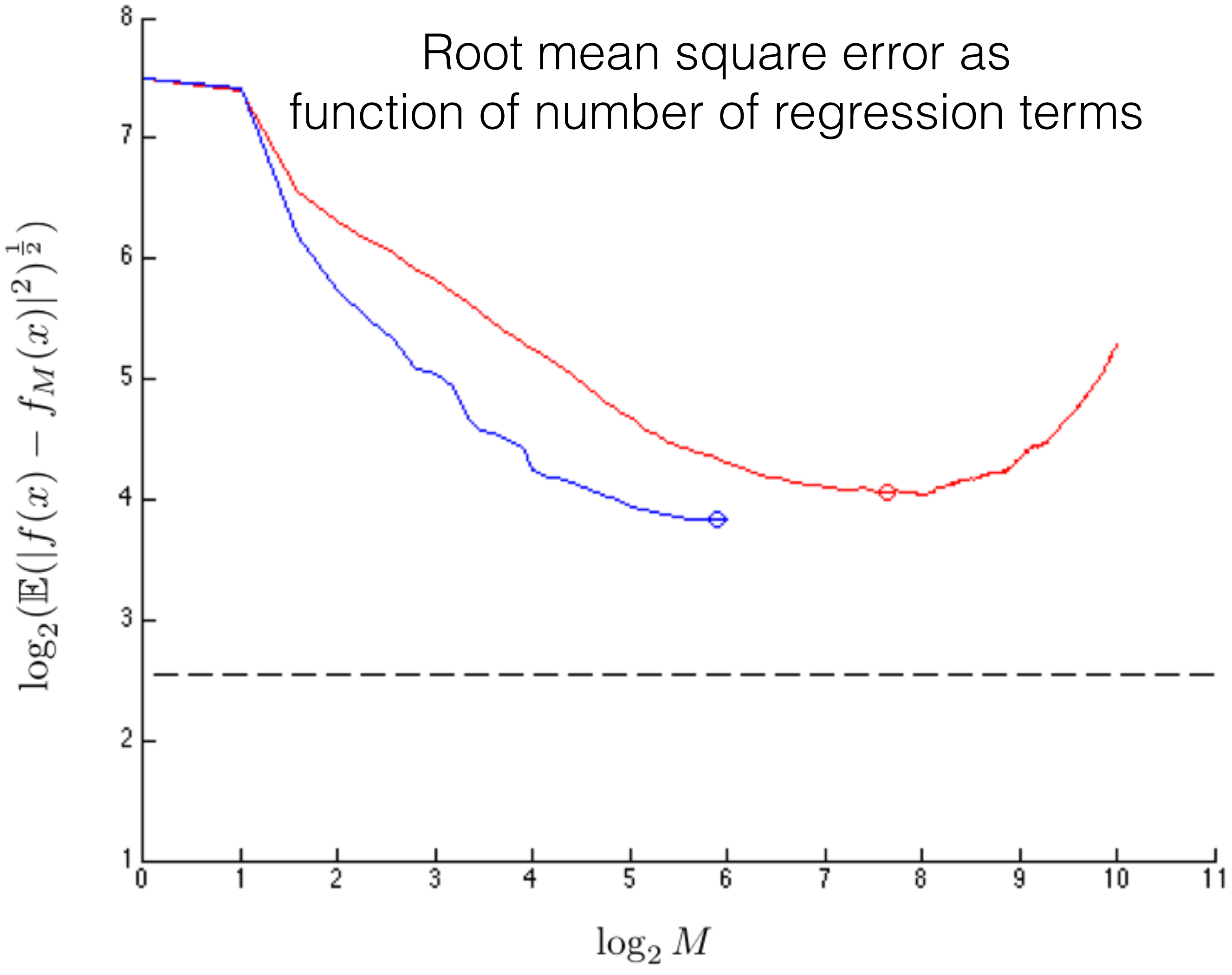
  at each iteration $m = 1, \ldots, M$
- The optimal value of M (to avoid overfit) is learned through cross validation
- The variance error is reduced by “bagging”

Data Set

- Data set $\{x_i, E(x_i)\}_i$ 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)
- $E(x_i)$ is the atomization energy of the molecule (energy necessary to break atomic bonds)



Fourier and Wavelet
M-term Regression Error


Root mean square error as function of number of regression terms

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

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

Root Mean Square Error (kcal/mol)

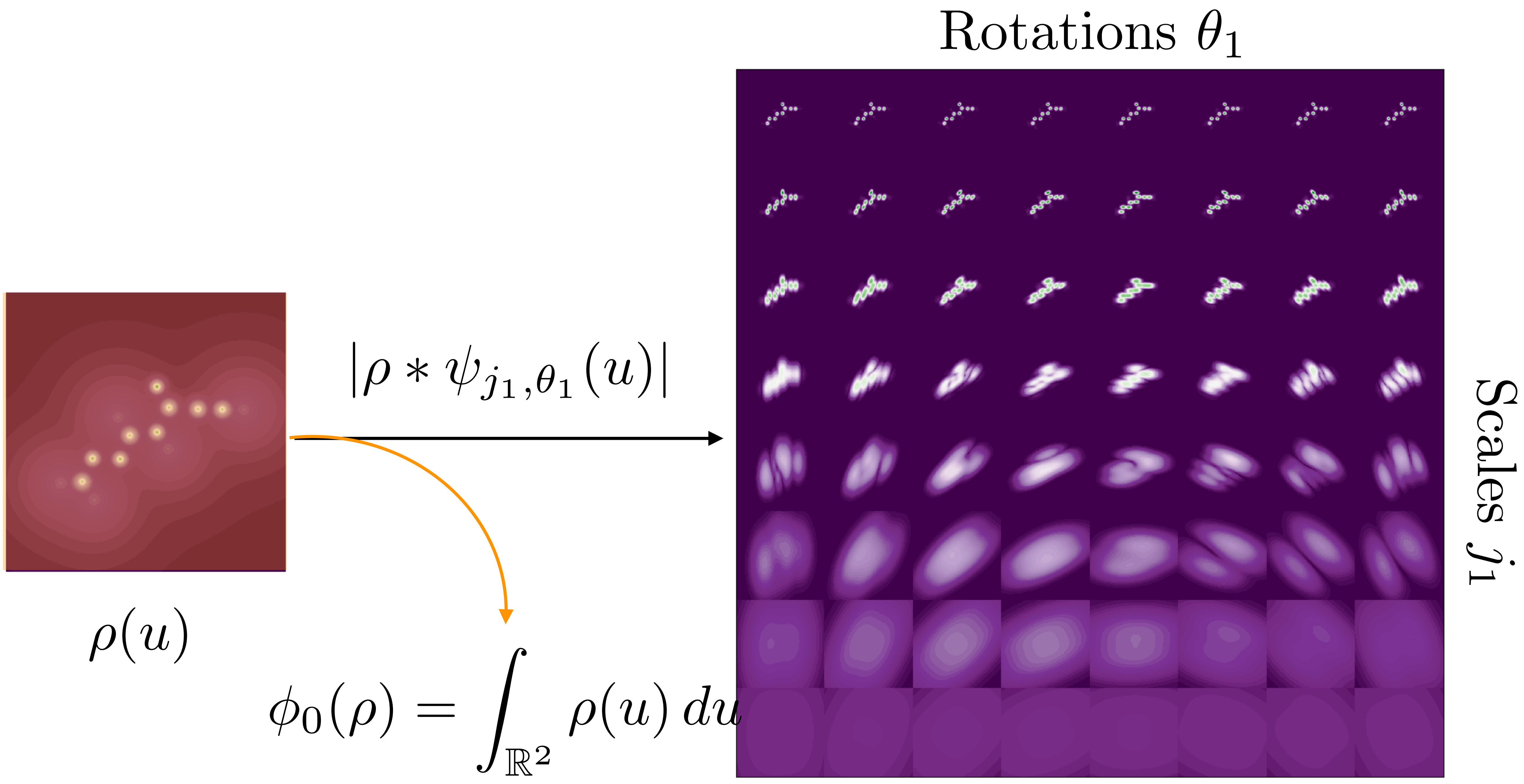
	Coulomb	Fourier	Wavelet
4000 Training	5.8	17	14
400 Training	21	16	16

Error = Bias + Variance

↑

Dominating

Scattering in 2D:
Layer 1



Rotations θ_1

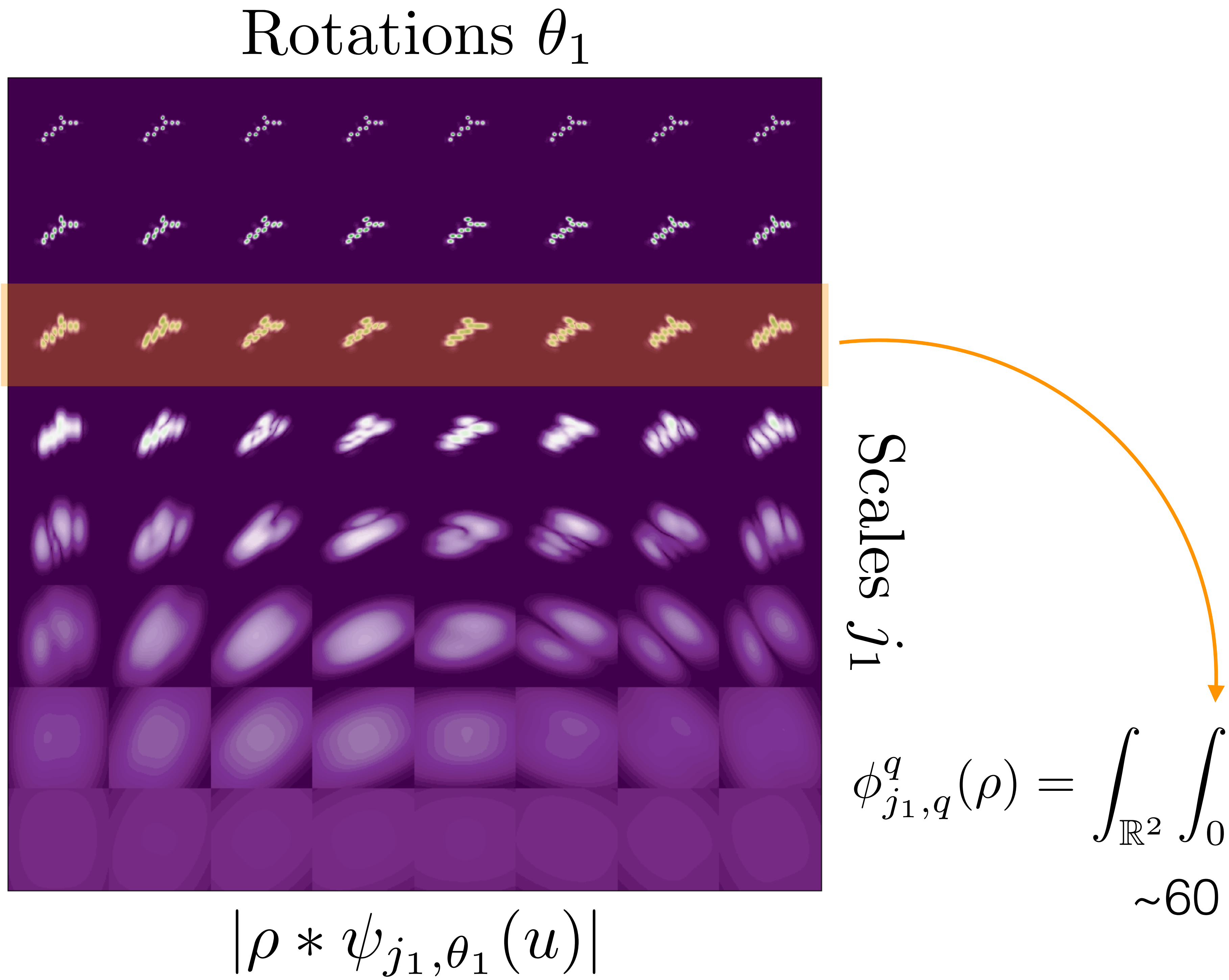
Scales j_1

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

$\rho(u)$

$$\phi_0(\rho) = \int_{\mathbb{R}^2} \rho(u)\, du$$

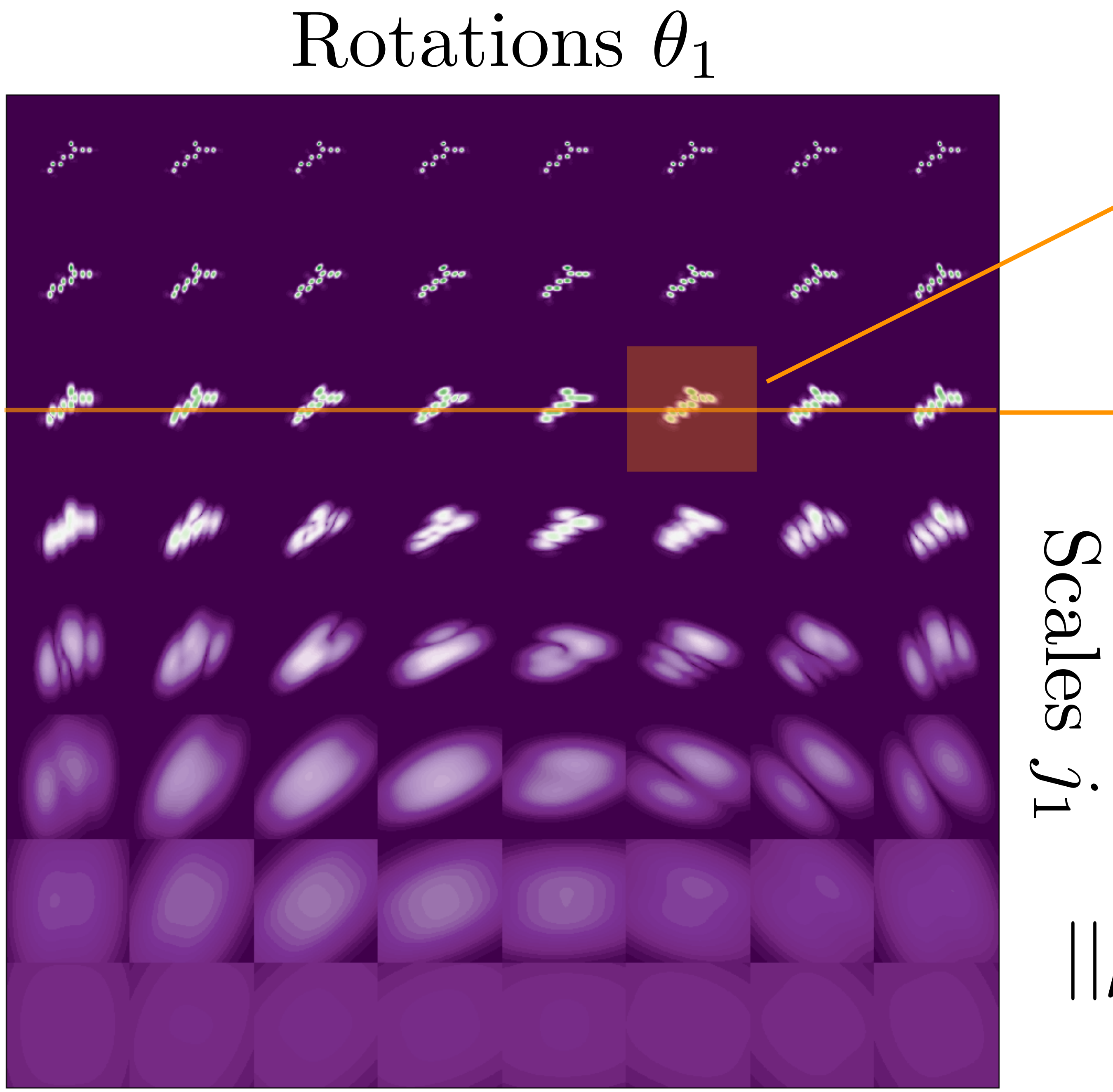
Scattering in 2D:
Layer 1



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

~60 functionals

Scattering in 2D: Layer 2



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

Recover translation variability:

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

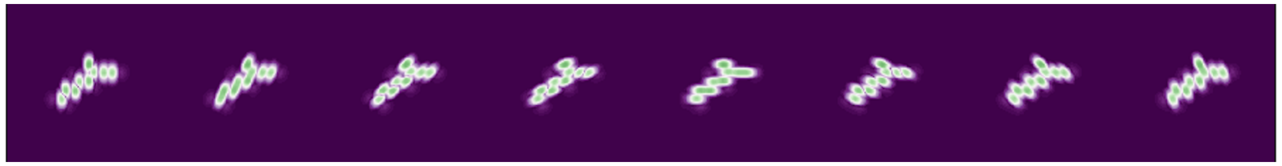
Recover rotation variability:

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

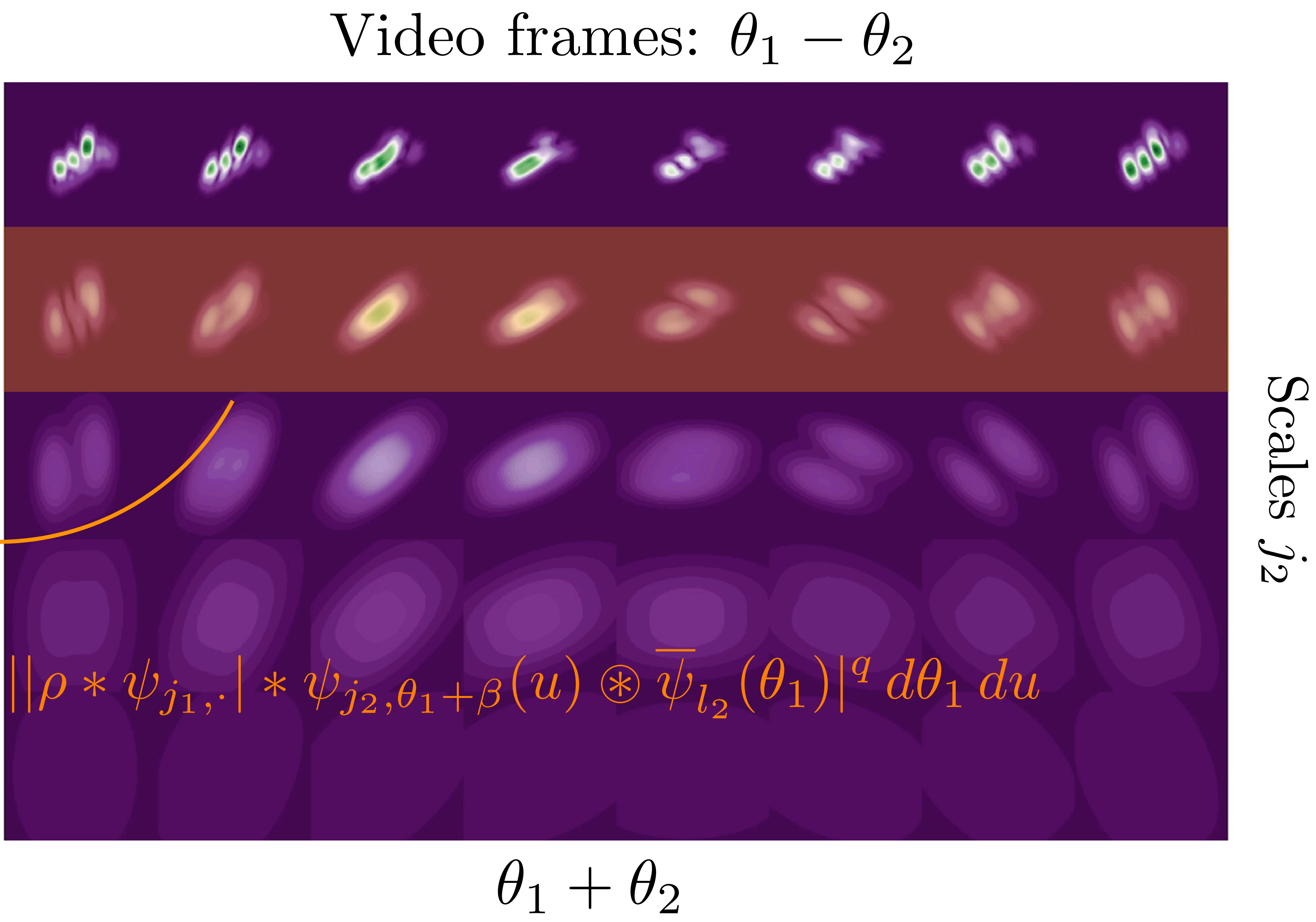
Combine to recover
roto-translation variabiltiy:

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

Scattering in 2D: Layer 2



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



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

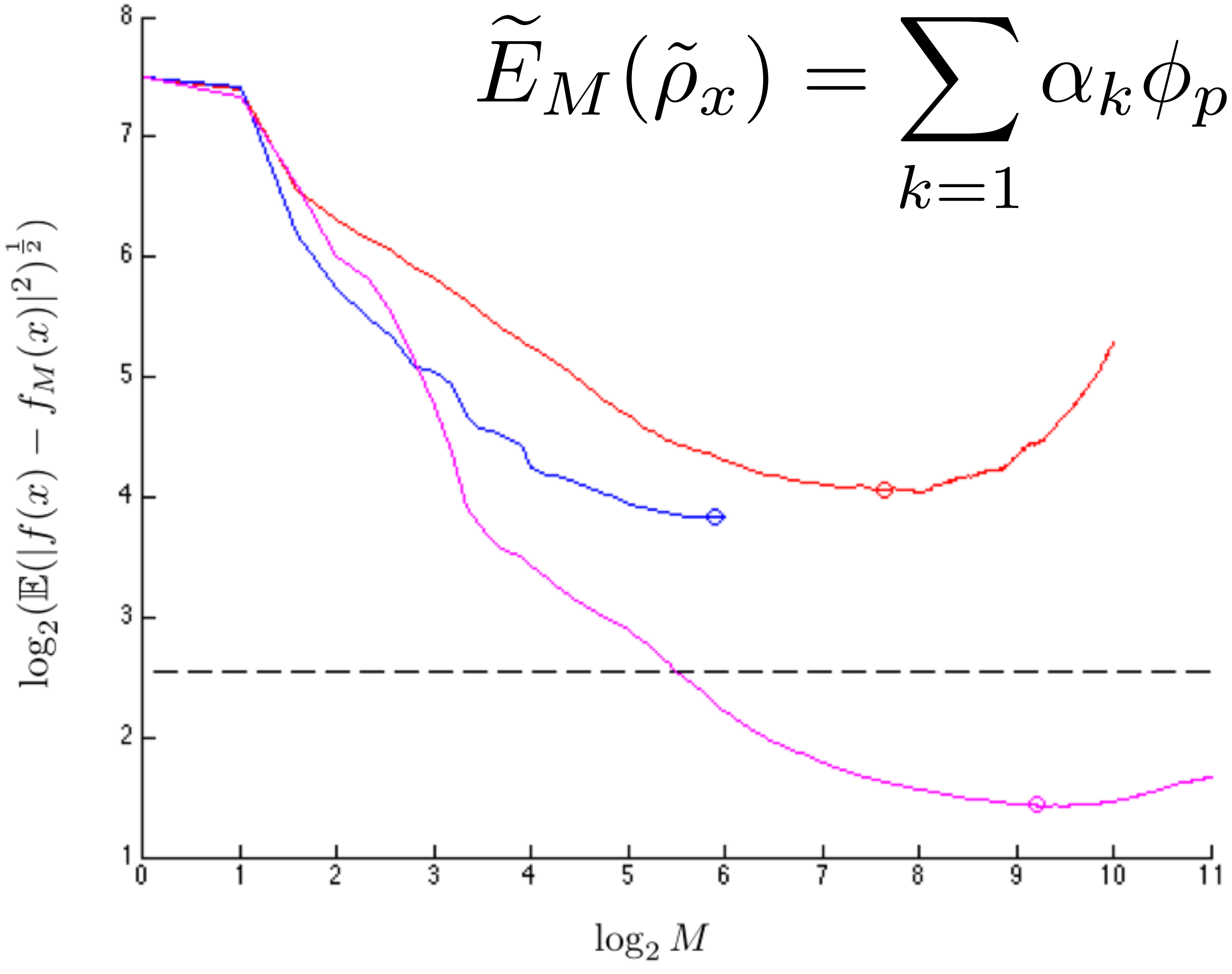
j_1, l_2 fixed

$$\phi^q_{j_1,j_2,\beta,l_2,q}(\rho) = \int_{\mathbb{R}^2}\int_0^{2\pi} ||\rho * \psi_{j_1,\cdot}| * \psi_{j_2,\theta_1+\beta}(u) \circledast \overline{\psi}_{l_2}(\theta_1)|^q \, d\theta_1 \, du$$

~10000 functionals

Scattering
M-term Regression Error

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



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

Numerical Results

Root Mean Square Error (kcal/mol)

	Coulomb	Fourier	Wavelet	Scattering
4000 Training	5.8	17	14	**2.7**
400 Training	21	16	16	**9.0**

- Cost in real time: minutes per molecule

- Some open questions and future directions:
  - More training -> Better scattering result?
  - Utilize forces: Scattering functionals are differentiable
  - Mathematical explanation for the 2nd layer functionals?
  - Can we learn chemistry?
  - What other physical functionals can we learn?

Scattering in 3D:
1st Layer

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

- 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

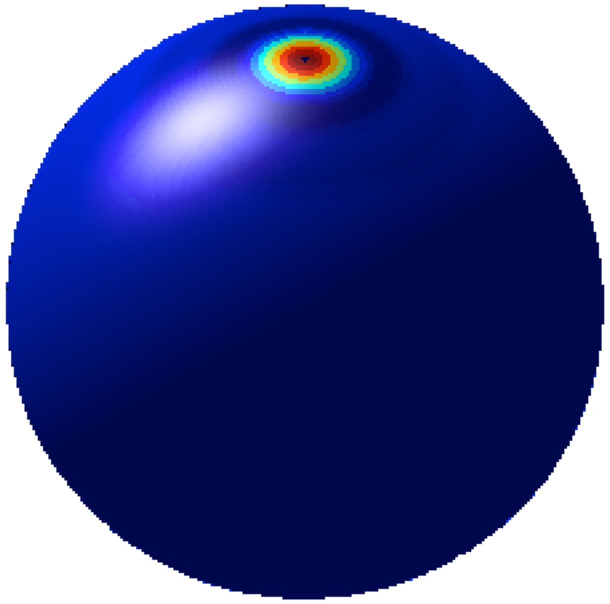
- The second layer can be computed as two separable wavelet transforms, one over translations ($\mathbb{R}^3$) and one over rotations (S^2).

- 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:

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

Thank you