Deep Wavelet Scattering for Quantum Energy Regression

2016 March APS Meeting
Predicting and Classifying Materials via High-Throughput Databases and Machine Learning I

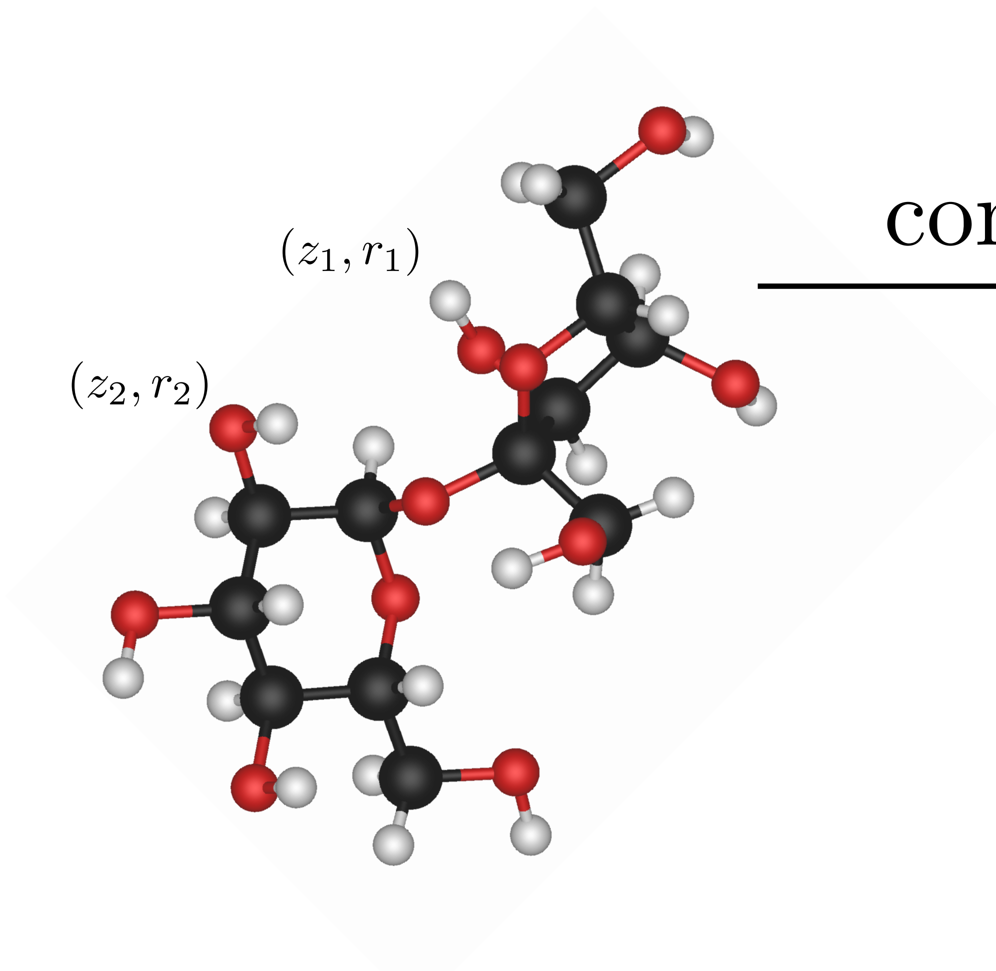
Matthew Hirn
Michigan State University
Department of Computational Mathematics, Science & Engineering
Department of Mathematics

Collaborators

Stéphane Mallat

Nicolas Poilvert

What do we want do?



Molecule

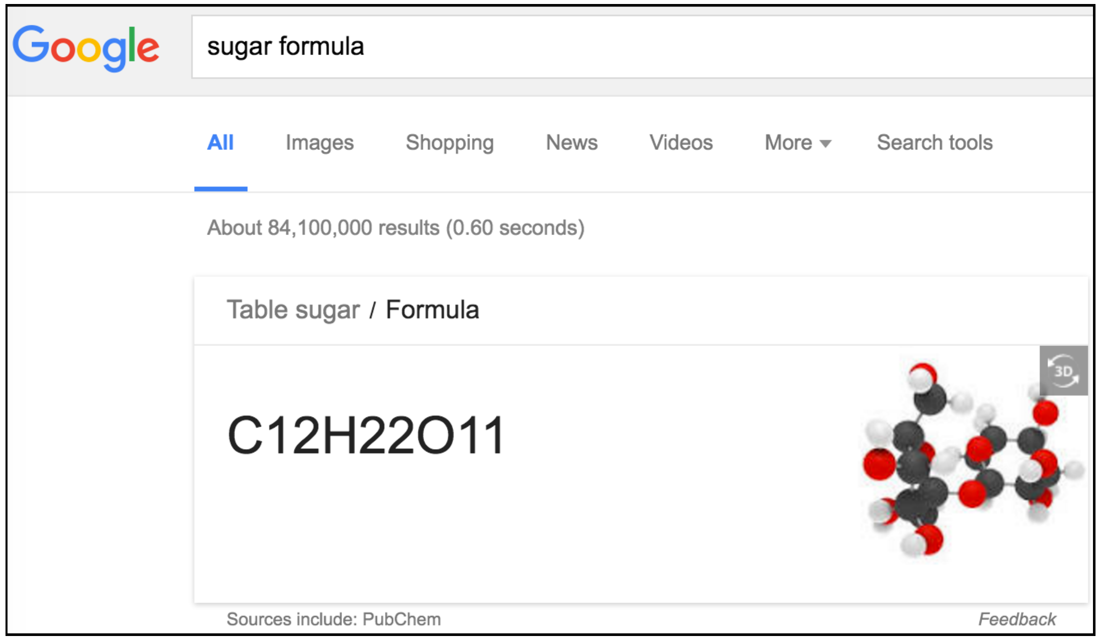
$$x = \{\text{charges, positions}\}$$
$$= \{(z_k, r_k)\}_k$$

$\xrightarrow{\text{compute}}$ Ground state energy

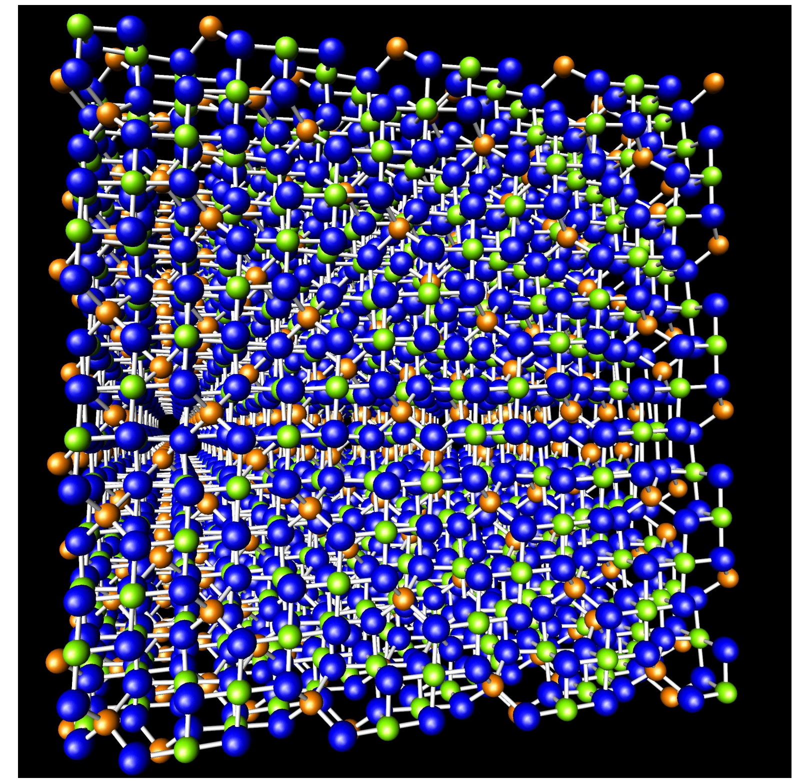
$$f(x) = E_0(x) + \sum_{k>l} \frac{z_k z_l}{|r_k - r_l|}$$

Energy from quantum electron interactions

Why do we want to do it?



Google of molecules



Materials design

Drug discovery

So what's the problem?

$$f(x) = E_0(x) + \sum_{k>l} \frac{z_k z_l}{|r_k - r_l|}$$

Schrödinger equation

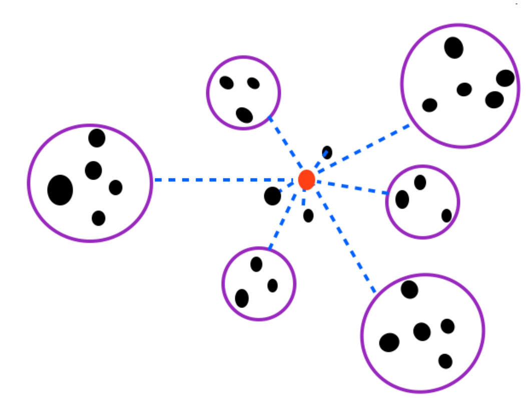
$$H[x]\Psi_0[x] = E_0(x)\Psi_0[x]$$

Not this

Fast multipole methods

(*Greengard, Rokhlin*)

Quantum mechanical (QM) approaches:



- Direct attacks (very small systems)
- Wave-function methods (small systems)
- Density functional theory (larger systems)

$O(N^\beta), \;\; \beta \geq 3$

computational cost

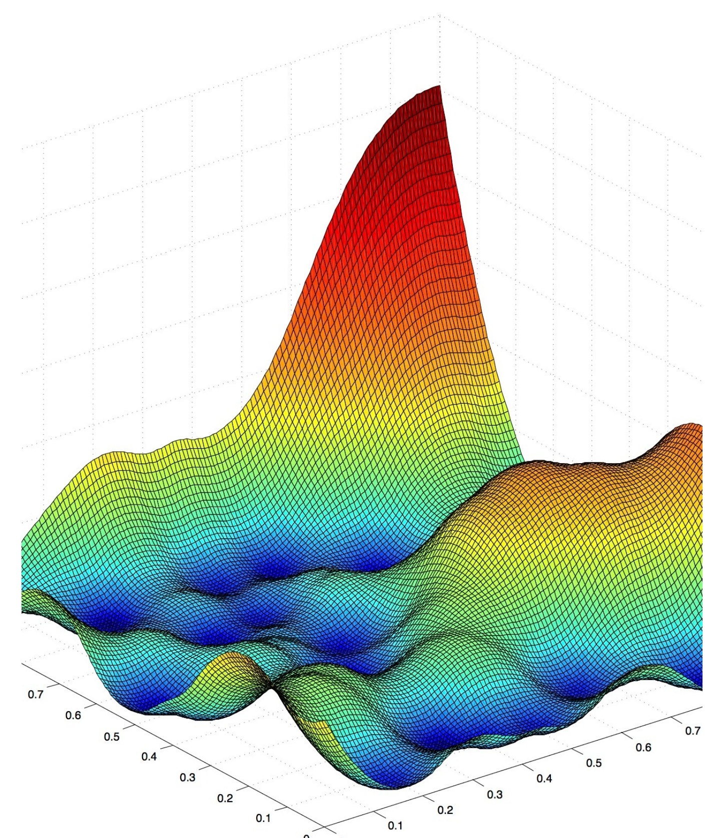
$N = \#$ electrons

Recent idea: Interpolation from known samples

- Use QM to compute training samples $\{(x_i, f(x_i))\}_{i \leq n}$
- Interpolate $f(x)$ from the training samples

<u>Issue:</u> Curse of dimensionality

- ϵ accuracy requires $O(\epsilon^{-d})$ samples
- $d = \text{dimension} = O(\#\,\text{atoms})$

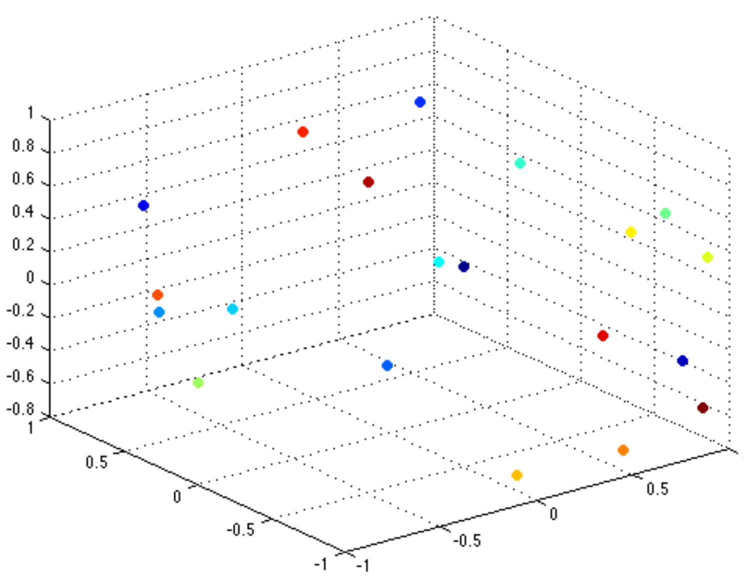


Recent idea:
Interpolation from known samples

- Use QM to compute training samples $\{(x_i, f(x_i))\}_{i \leq n}$
- Interpolate $f(x)$ from the training samples

Issue: Curse of dimensionality

- ϵ accuracy requires $O(\epsilon^{-d})$ samples
- $d = \text{dimension} = O(\#\,\text{atoms})$



Need to take advantage of ***invariants*** *and* ***regularity*** *of f*

Energy
invariants and regularity

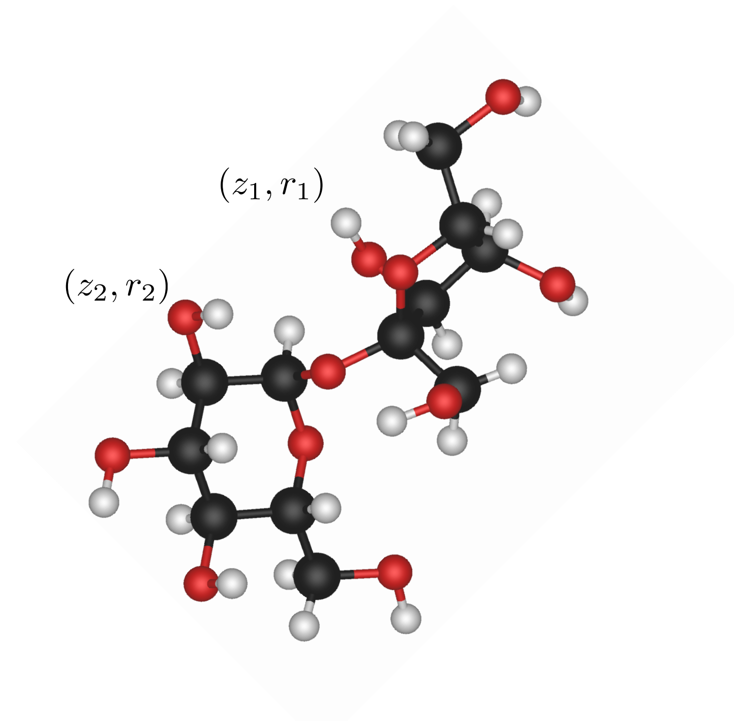
- <u>**Permutation invariance**</u>
  Invariant to permutations of the atom index k

- <u>**Isometry invariance**</u>
  Invariant to translations, rotations and reflections

- <u>**Deformation stability**</u>
  Lipschitz stable to diffeomorphisms of the molecule

(Bartók, Kondor, Csányi)

$$f(x) = E_0(x) + \sum_{k>l} \frac{z_k z_l}{|r_k - r_l|}$$



Molecule

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

Regression over a dictionary

- New representation of the molecule x:
$$\Phi(x) = (\phi_k(x))_k$$

- Linear regression (interpolation) of f using Φ:
$$\widetilde{f}(x) = \langle w, \Phi(x) \rangle = \sum_k w_k \phi_k(x)$$

- Weights $(w_k)_k$ learned from the training data

(*Bartók, Kondor, Csányi: SOAP kernels*)
(*Rupp, Tkatchenko, Müller, von Lilienfeld: Coulomb matrices*)
(*And many others...*)

Dictionary properties

$$\widetilde{f}(x) = \langle w, \Phi(x) \rangle = \sum_k w_k \phi_k(x)$$

The dictionary Φ should:

- Have the same invariants and regularity as f
  - permutation invariance
  - isometry invariance
  - deformation stability

- Span a large enough space to approximate f to high accuracy, with as few terms as possible

$$\|w\|_0 \leq M$$

Dictionary structure

Decompose Φ as:

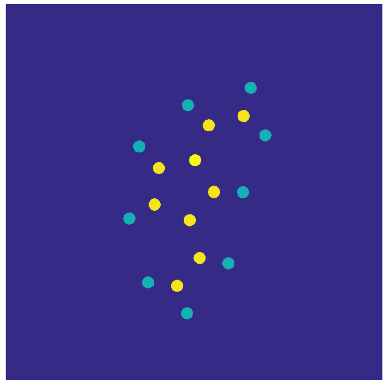
$$\Phi(x) = \Theta\rho[x], \quad x = \{(z_k, r_k)\}_k$$

- $\rho[x] : \mathbb{R}^3 \to \mathbb{R}$
- Permutation invariant
- Isometry covariant

- Isometry invariant
- Lipschitz stable to diffeomorphisms
- *This part is the main difficulty*

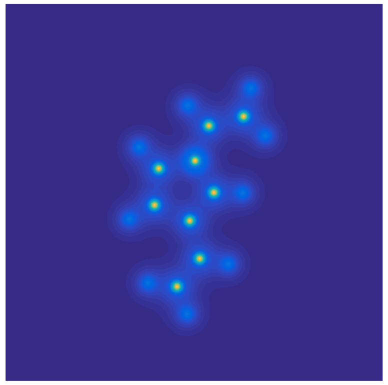
Non-interacting density

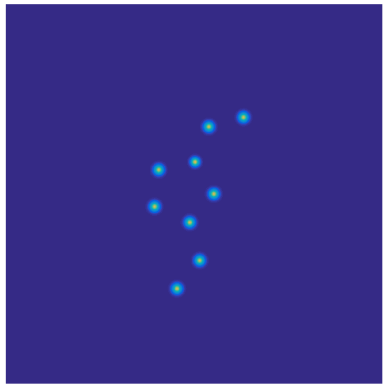
$$\rho[x](u) = \sum_k \rho[z_k](u - r_k), \quad x = \{(z_k, r_k)\}_k$$



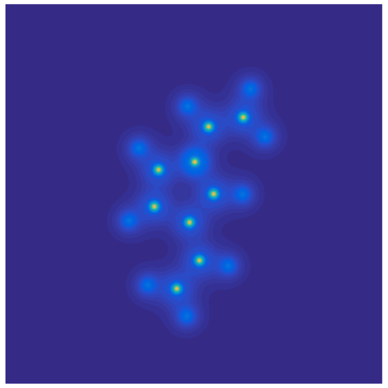
Candidates for $\rho[z_k]$:

- $\rho[z_k] = z_k \delta$
- $\rho[z_k] = \text{Isolated atomic density}$
  $= \rho_{\text{cor}}[z_k] + \rho_{\text{val}}[z_k]$





core



valence

- $\rho[z_k] = (\rho_{\text{cor}}[z_k], \rho_{\text{val}}[z_k])$

Fourier and autocorrelation

Dictionary $\Phi = \Theta\rho$

Now we focus on Θ; recall goals:

- Isometry invariant (translations, rotations, reflections)
- Stable to deformations

"Classic" translation invariant representations:

- Autocorrelation: $\Theta\rho(\tau) = \int \rho(u)\rho(u-\tau)\,du$
- Fourier modulus: $\widehat{\Theta}\rho(\omega) = |\widehat{\rho}(\omega)|^2$

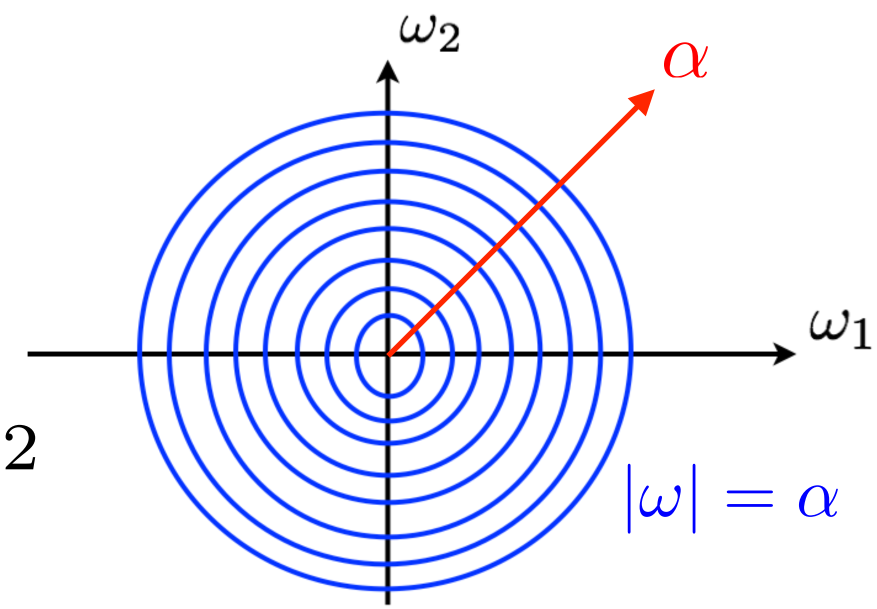
Integrate over rotations to obtain isometry invariance

Fourier and autocorrelation

Invariant Fourier operator:

$$\widehat{\Theta}\rho(\alpha) = \|\widehat{\rho}_\alpha\|_2^2 = \int_{S^2} |\widehat{\rho}(\alpha\eta)|^2 \, d\eta$$

$$\omega = \alpha\eta, \; (\alpha, \eta) \in \mathbb{R}^+ \times S^2$$



Pros:

- Isometry invariant
- Diagonalizes Coulomb:

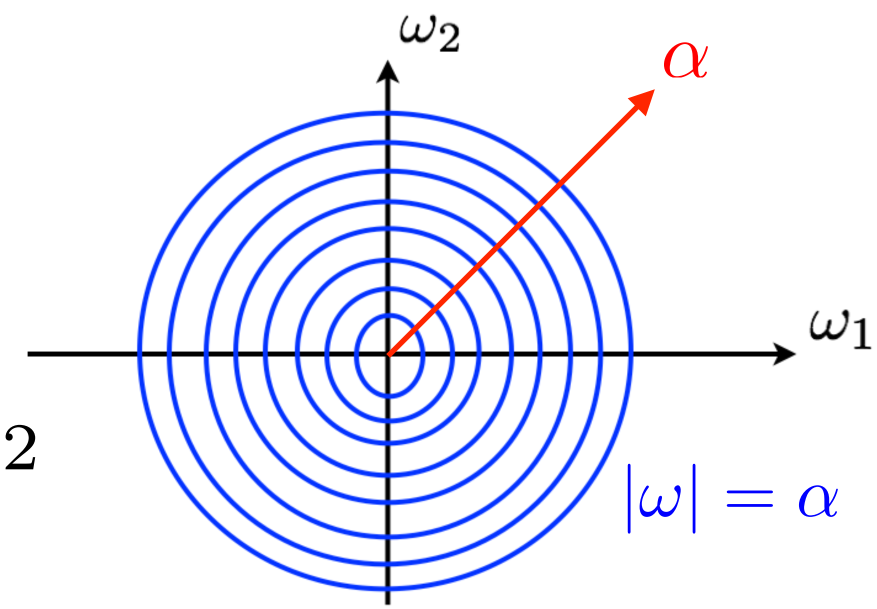
$$U(\rho) = \iint \frac{\rho(u)\rho(v)}{|u-v|} \, du \, dv = \frac{1}{2\pi^2} \int_0^\infty \alpha^{-2} \|\widehat{\rho}_\alpha\|_2^2 \, d\alpha$$

Fourier and autocorrelation

<u>Invariant Fourier operator:</u>

$$\widehat{\Theta}\rho(\alpha) = \|\widehat{\rho}_\alpha\|_2^2 = \int_{S^2} |\widehat{\rho}(\alpha\eta)|^2 \, d\eta$$

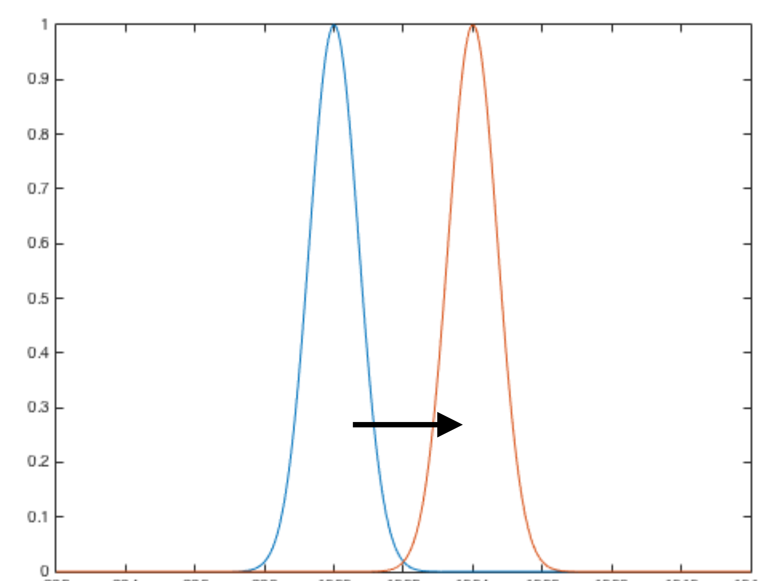
$$\omega = \alpha\eta, \; (\alpha, \eta) \in \mathbb{R}^+ \times S^2$$



<u>Cons:</u>

- Not sparse: $U(\rho) = \sum_{k=1}^{\epsilon^{-2}} w_k \|\widehat{\rho}_{k\epsilon}\|_2^2 + O(\epsilon)$
- Not stable to deformations:

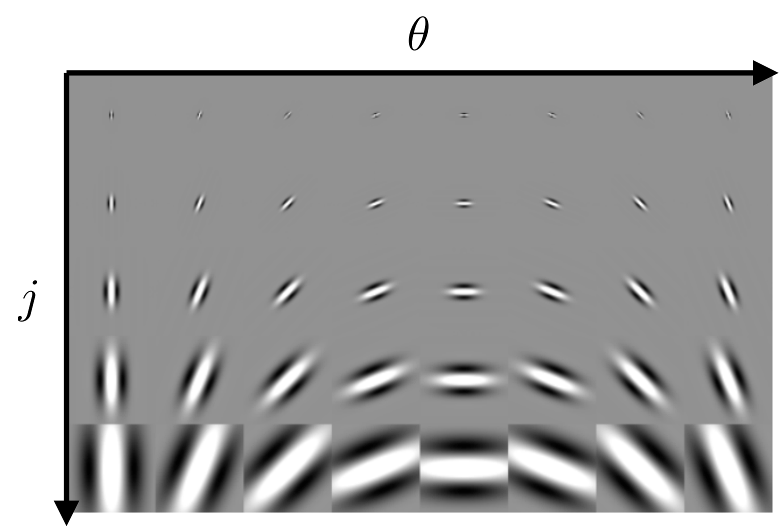
  Invariant autocorrelation:
  – Bumps located at $\{|r_k - r_l|\}_{k,l}$
  – Δ sized diffeomorphism
  moves them $\Delta|r_k - r_l|$ distance



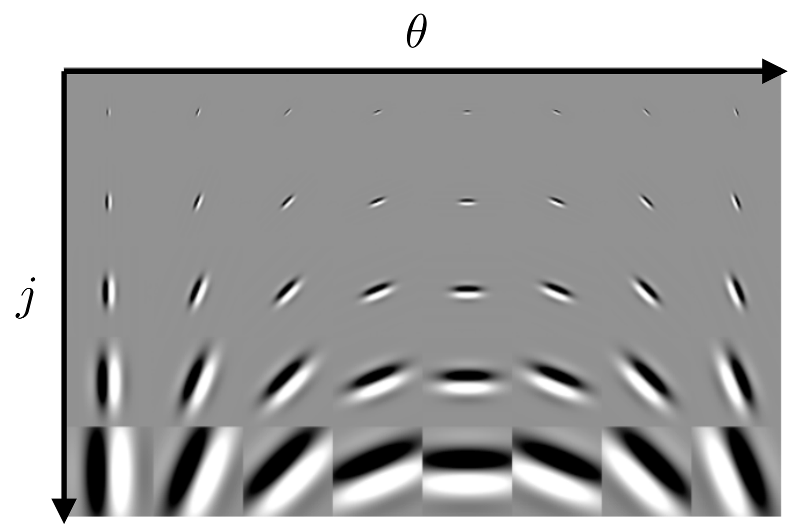
$|r_k - r_l| \gg 0$

Wavelets

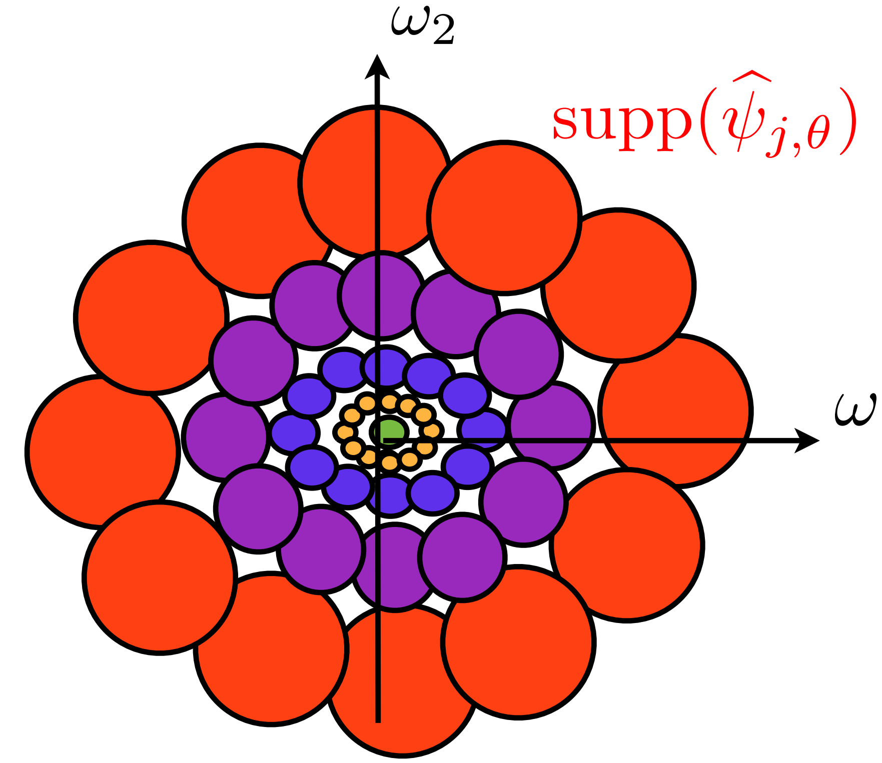
- Wavelet $\psi(u) = e^{-|u|^2}(e^{i\eta_0 \cdot u} - C), \qquad \int \psi = 0$
- Dilated and rotated:
  $\psi_{j,\theta} = 2^{-3j}\psi(2^{-j}R_\theta^{-1}u), \quad j \in \mathbb{Z}, \;\; R_\theta \in \mathrm{O}(3)$



real($\psi_{j,\theta}$)



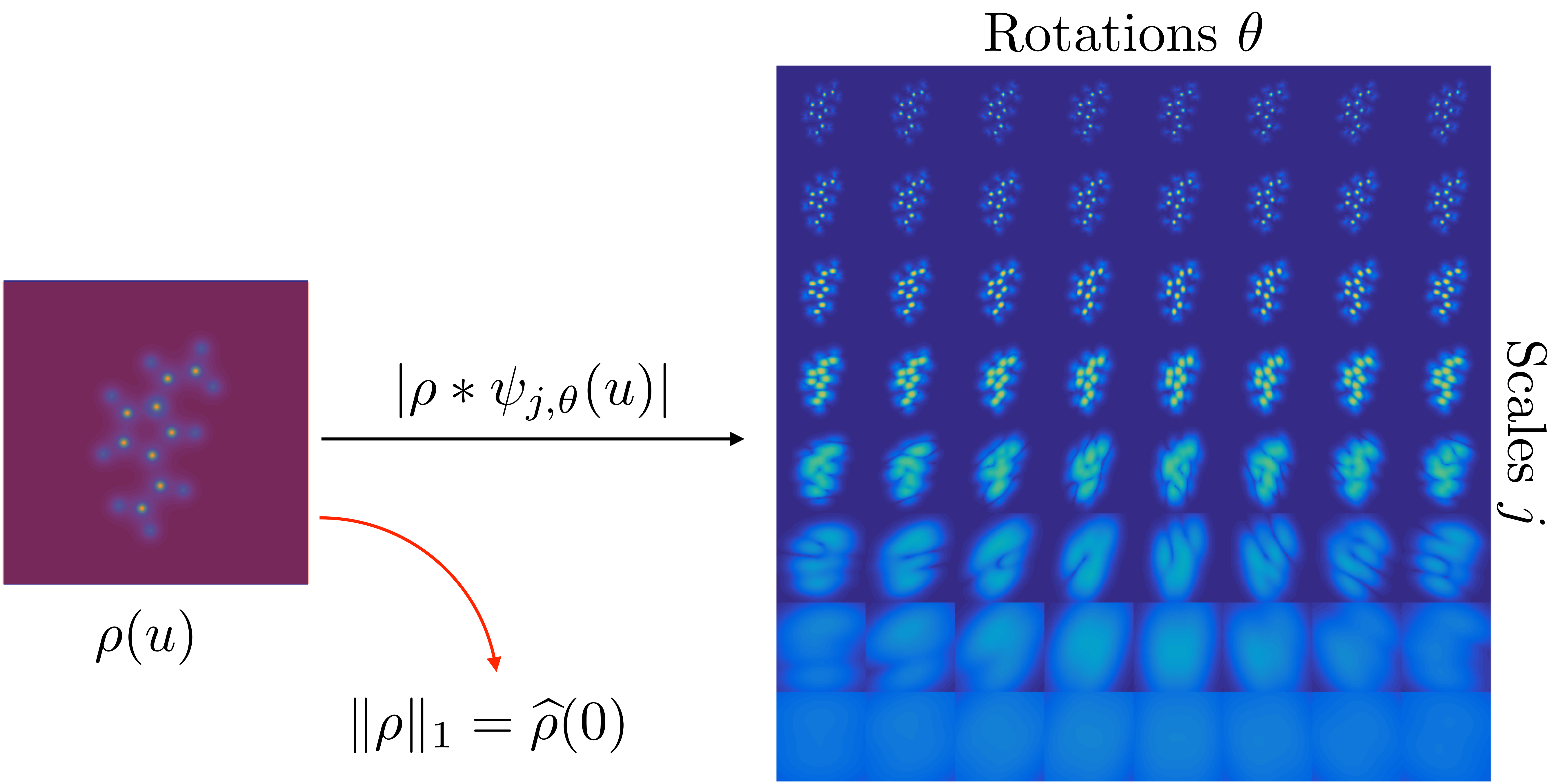
imag($\psi_{j,\theta}$)



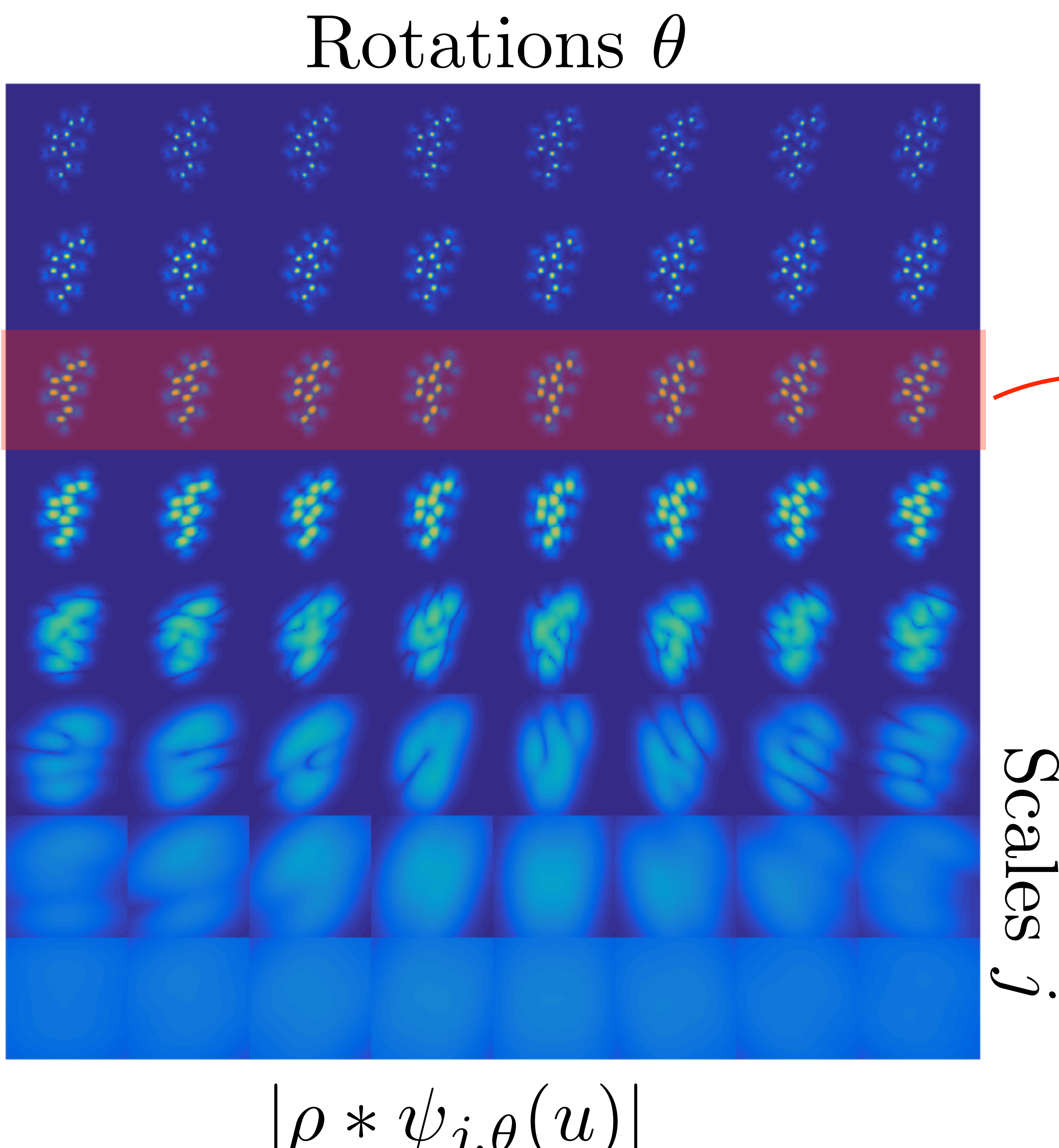
<u>Wavelet Transform:</u> $\rho \mapsto \{\int \rho, \;\; \rho * \psi_{j,\theta}(u)\}$

Interactions at scale 2^j in direction θ

Wavelet modulus

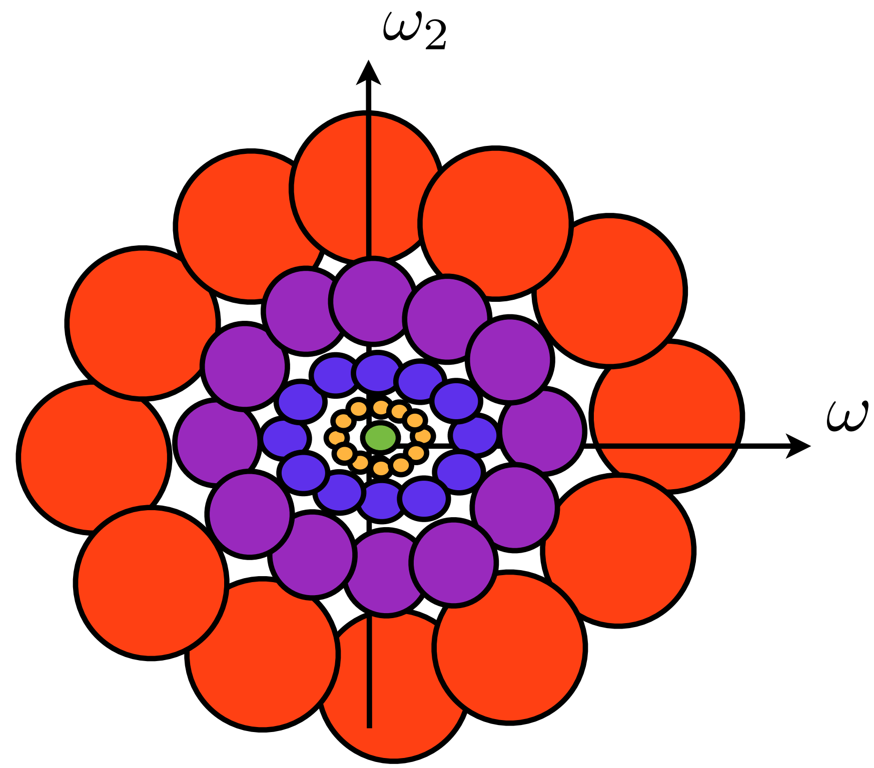


Wavelet invariants



$$\|\rho * \psi_{j,\cdot}\|_1 = \int_{\mathrm{O}(3)} \int_{\mathbb{R}^3} |\rho * \psi_{j,\theta}(u)| \, du \, d\theta$$

Invariant wavelet operator



Invariant wavelet operator:

$$\Theta\rho(j) = \|\rho * \psi_{j,\cdot}\|_1$$

Pros:

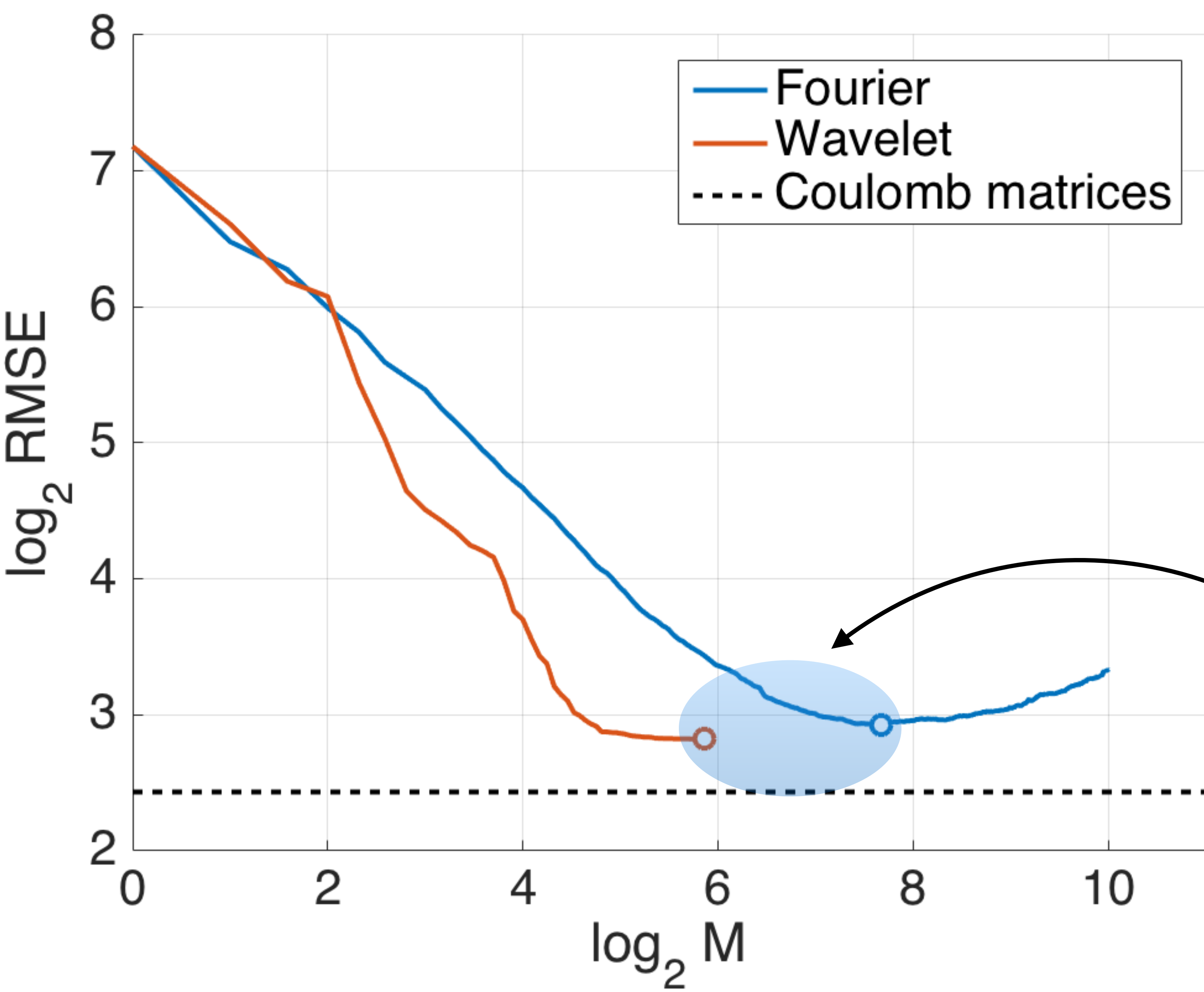
- Isometry invariant
- Stable to deformations (*Mallat*)
- Diagonalizes Coulomb and is sparse:

$$U(\rho) = \sum_{j=\log\epsilon}^{-\log\epsilon} w_j \|\rho * \psi_{j,\cdot}\|_2^2 + O(\epsilon)$$

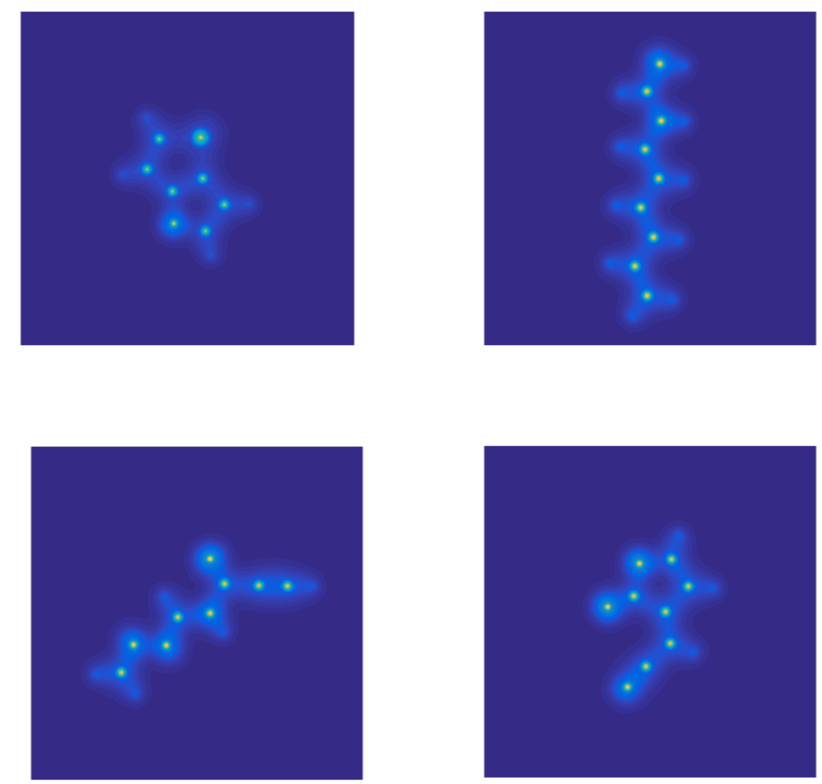
Cons: Encoding the invariants removes a lot of information

M-term regression error

$$\widetilde{f}_M(x) = \sum_{m=1}^{M} w_{k_m} \phi_{k_m}(x)$$



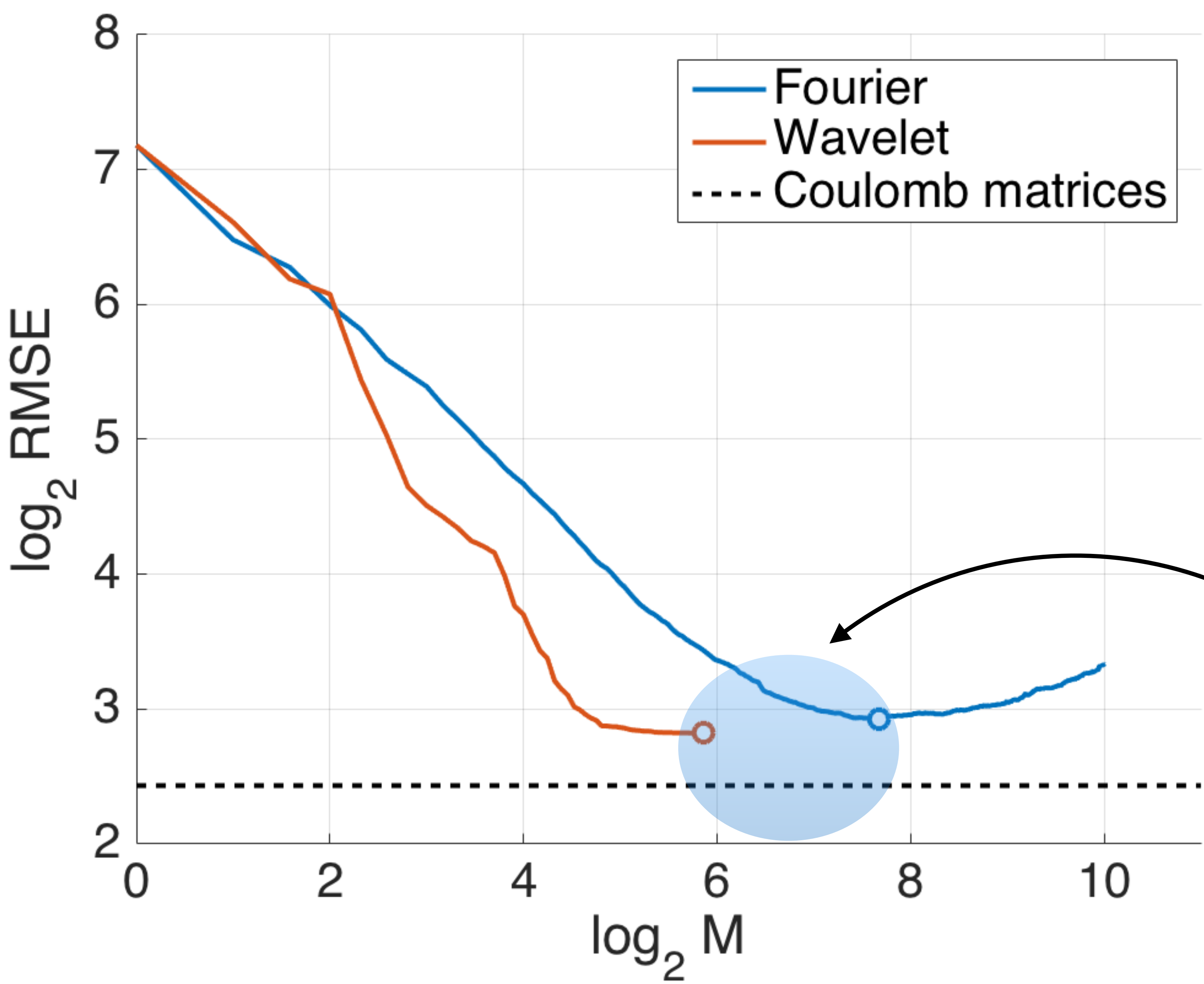
Data set:
Planar organic molecules



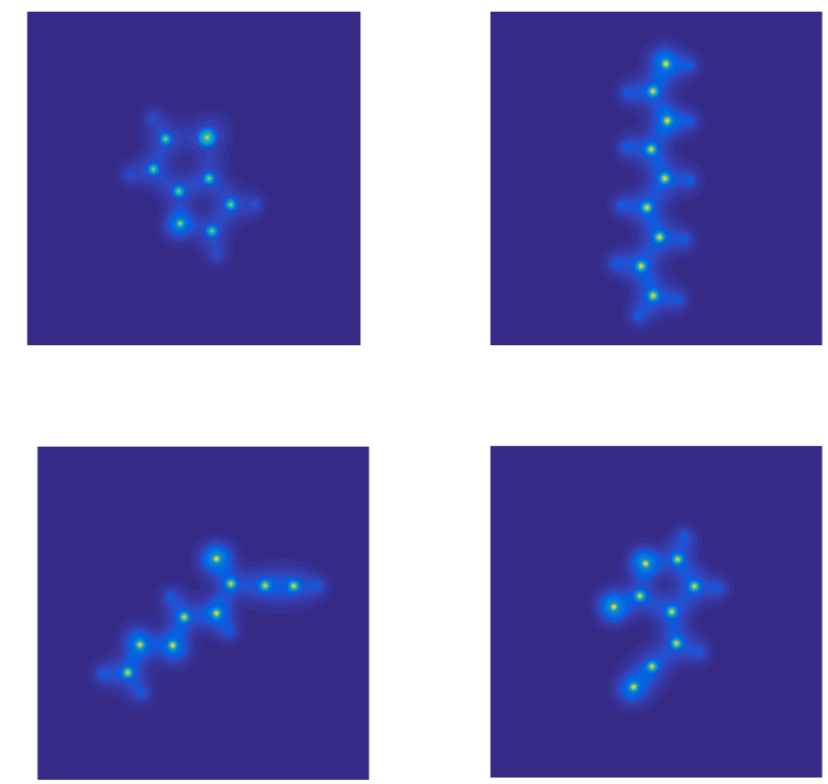
Multiscale wavelet dictionary
is sparser than Fourier dictionary

M-term regression error

$$\widetilde{f}_M(x) = \sum_{m=1}^{M} w_{k_m} \phi_{k_m}(x)$$

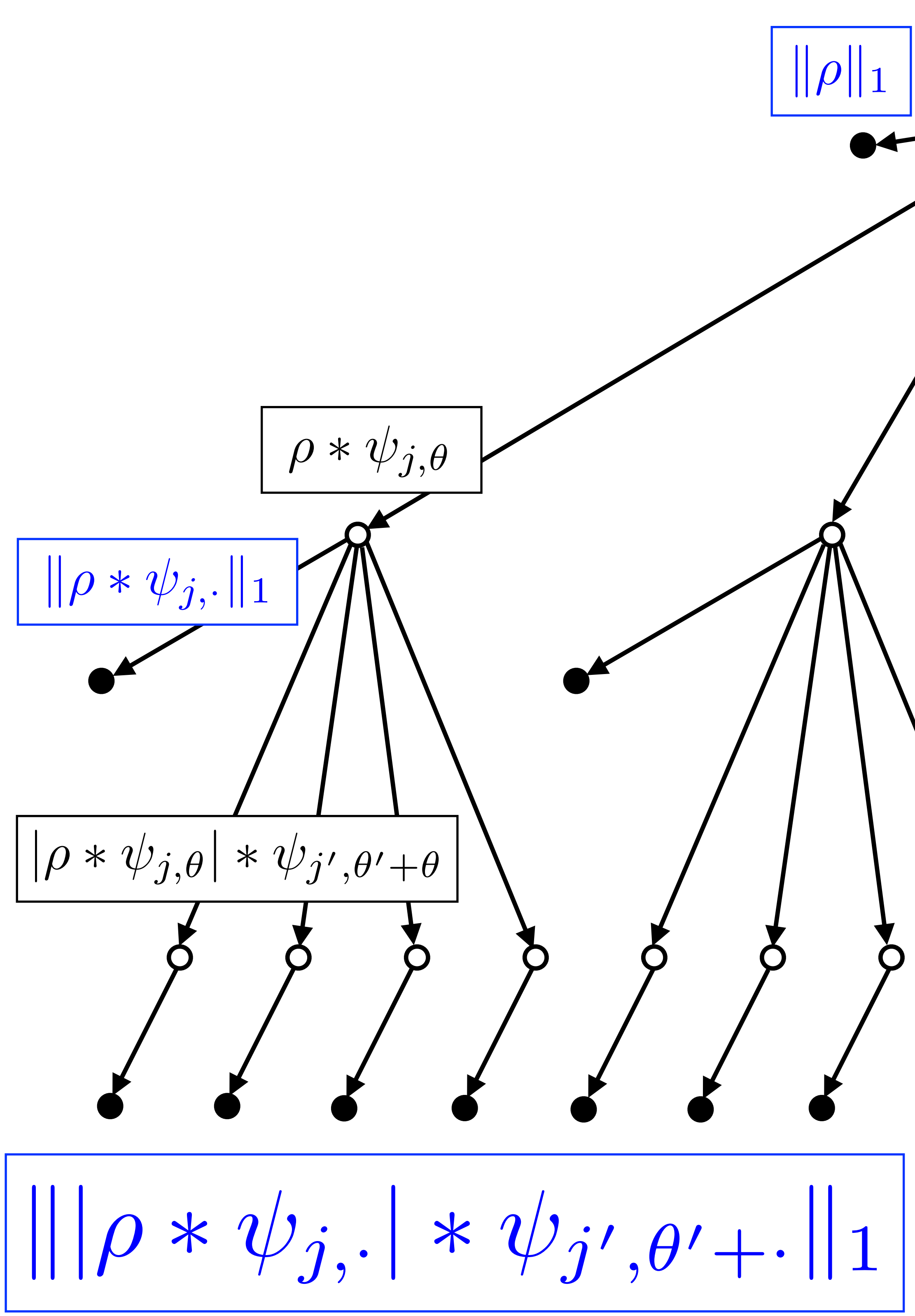


Data set:
Planar organic molecules



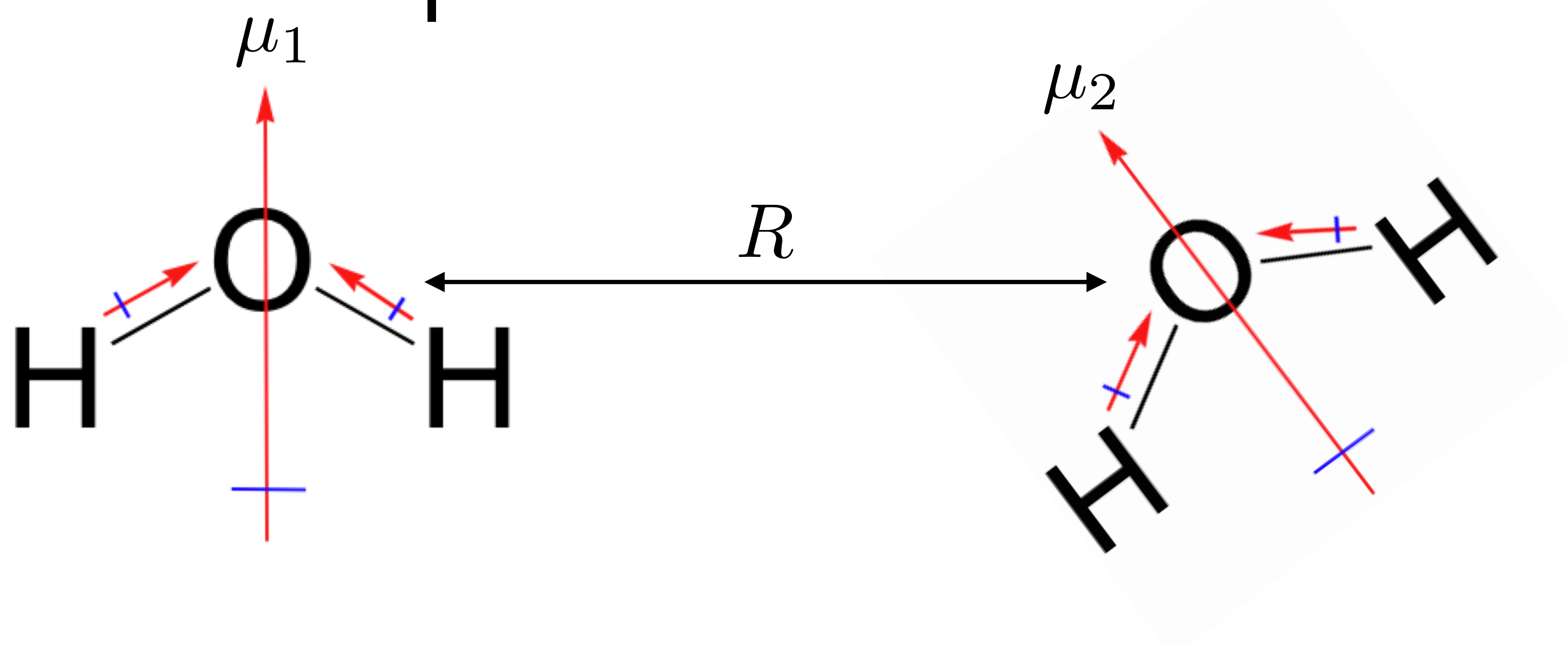
Coulomb matrices outperform both Fourier and wavelets

Invariant scattering operator



Couples wavelet features at two scales to encode complex interactions that move across scales

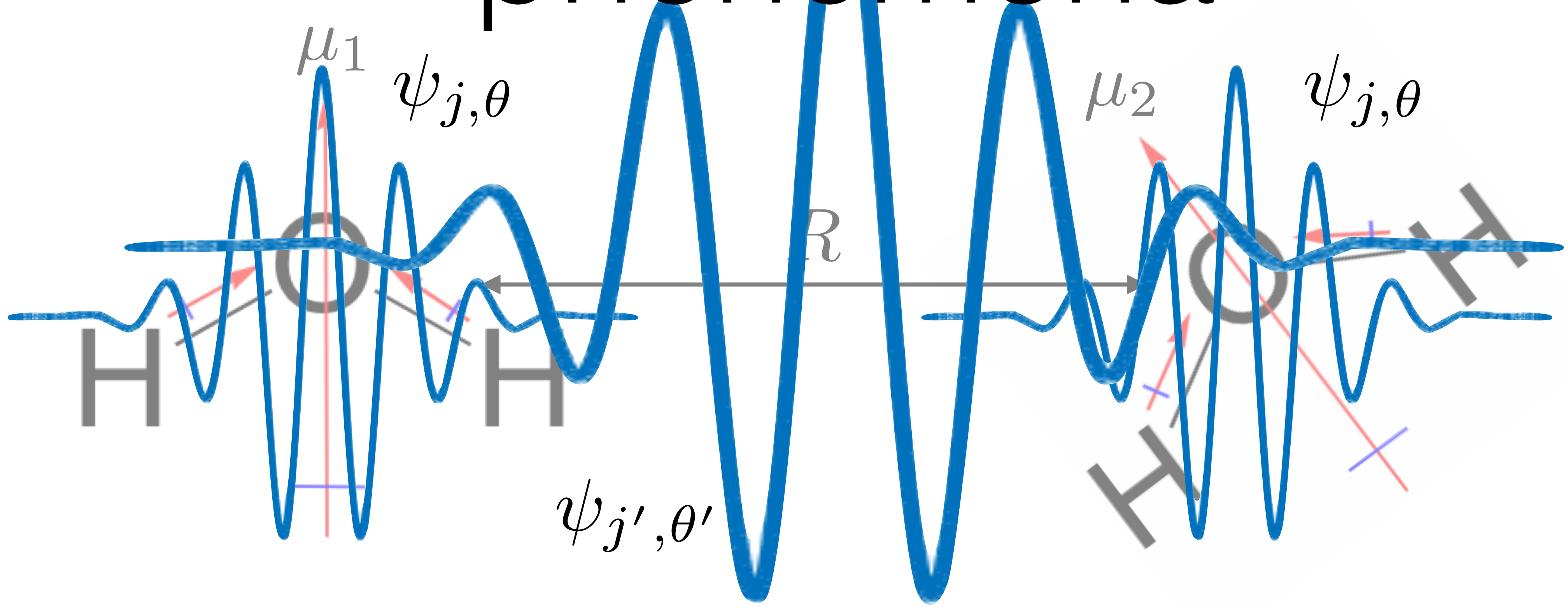
Coupled multiscale phenomena



Multipole expansion dipole-dipole moment:

$$\frac{\mu_1 \cdot \mu_2}{R^3}$$

Coupled multiscale phenomena



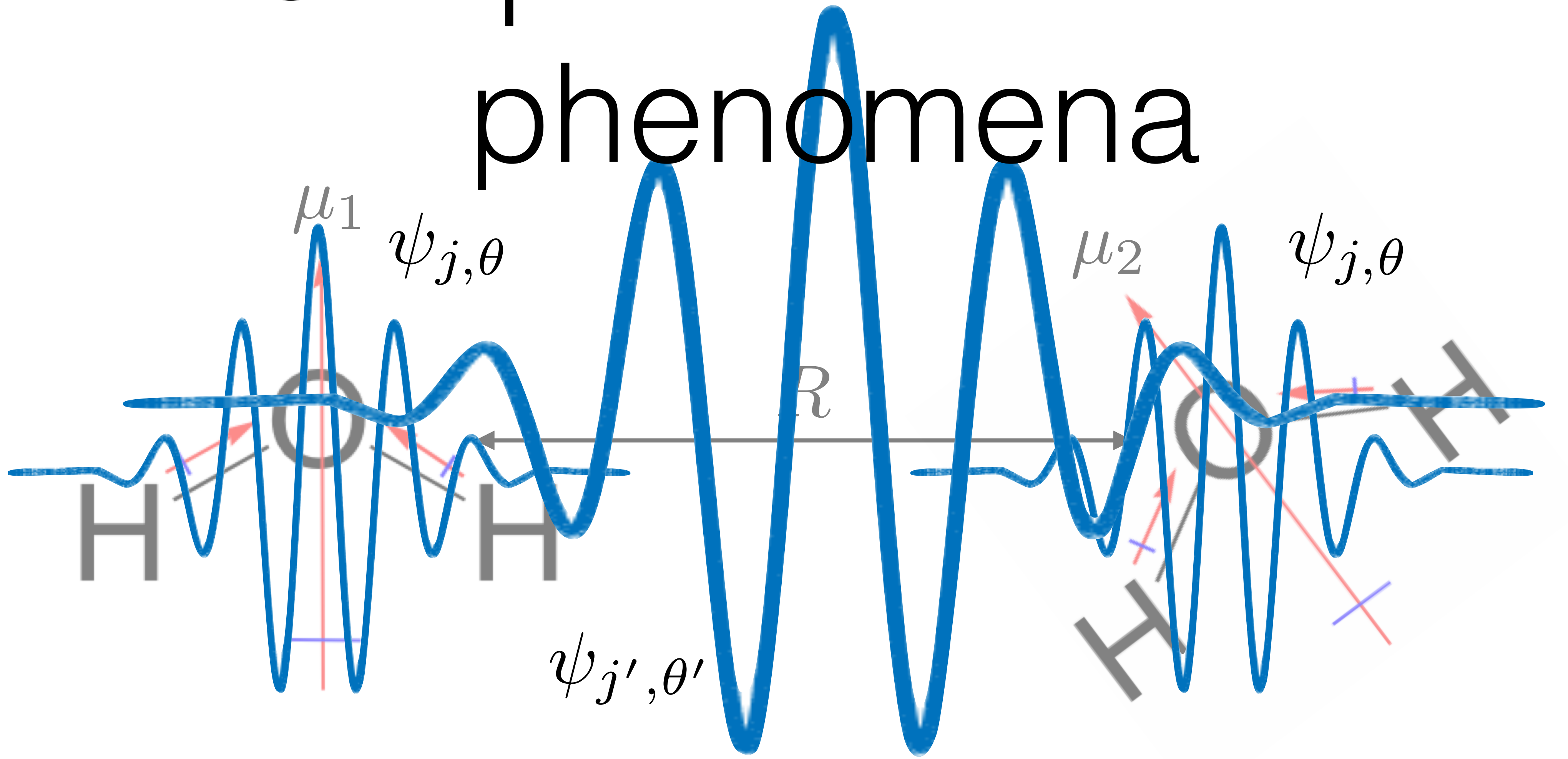
Small scale wavelets $\psi_{j,\theta}$ can learn dipole orientations

Large wavelets $\psi_{j',\theta'}$ can learn the distance between molecules

But there is no linear expansion of the dipole-dipole moment as:

$$\frac{\mu_1 \cdot \mu_2}{R^3} \neq \sum_{j \text{ small}} w_j \|\rho * \psi_{j,.}\|_1 + \sum_{j' \text{ large}} w_{j'} \|\rho * \psi_{j',.}\|_1$$

Coupled multiscale phenomena



2^{nd} order scattering features on the other hand:

First apply the small scale wavelet transform $\rho * \psi_{j,\theta} \ldots$

…then apply the large scale wavelet transform $|\rho * \psi_{j,\theta}| * \psi_{j',\theta'+\theta}$

Intuition and numerical evidence supports (theory in progress):

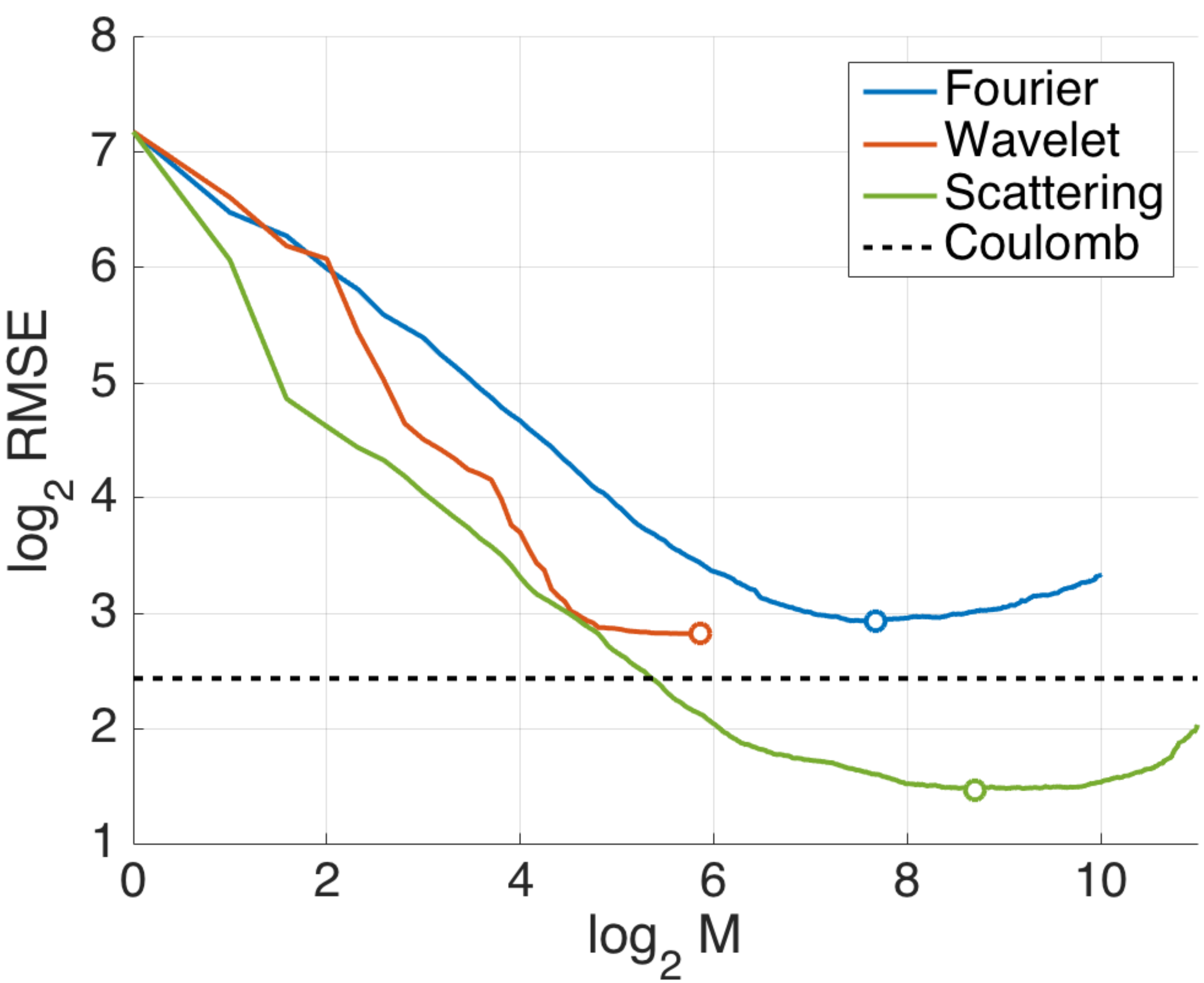
$$\frac{\mu_1 \cdot \mu_2}{R^3} \approx \sum_{j \text{ small}} \sum_{j' \text{ large}} \sum_{\theta'} w_{j,j',\theta'} \| |\rho * \psi_{j,\cdot}| * \psi_{j',\theta'+\cdot} \|_1$$

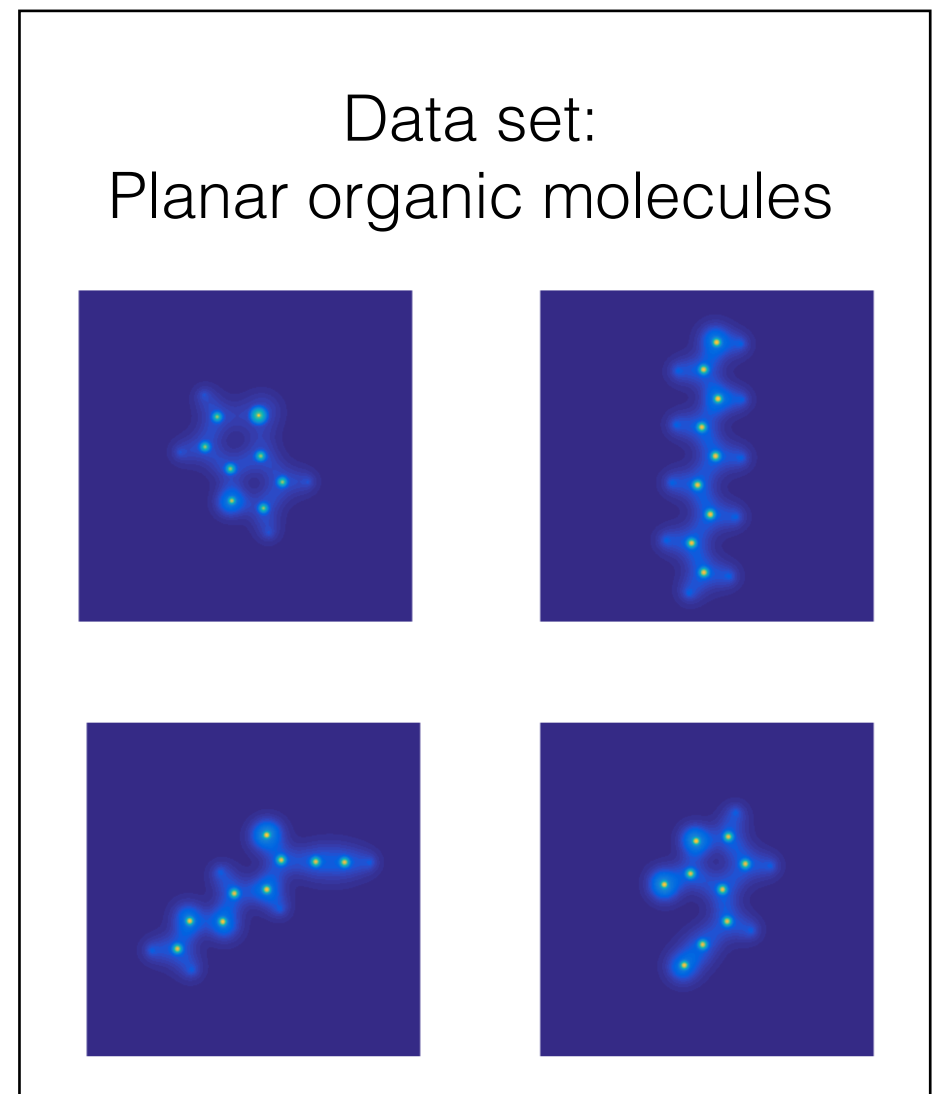
M-term regression error

$$\widetilde{f}_M(x) = \sum_{m=1}^{M} w_{k_m} \phi_{k_m}(x)$$

Errors (kcal/mol)

	Coulomb	Fourier	Wavelet	Scattering
MAE	2.4	5.3	5.4	1.7
RMSE	5.4	7.2	7.1	2.6

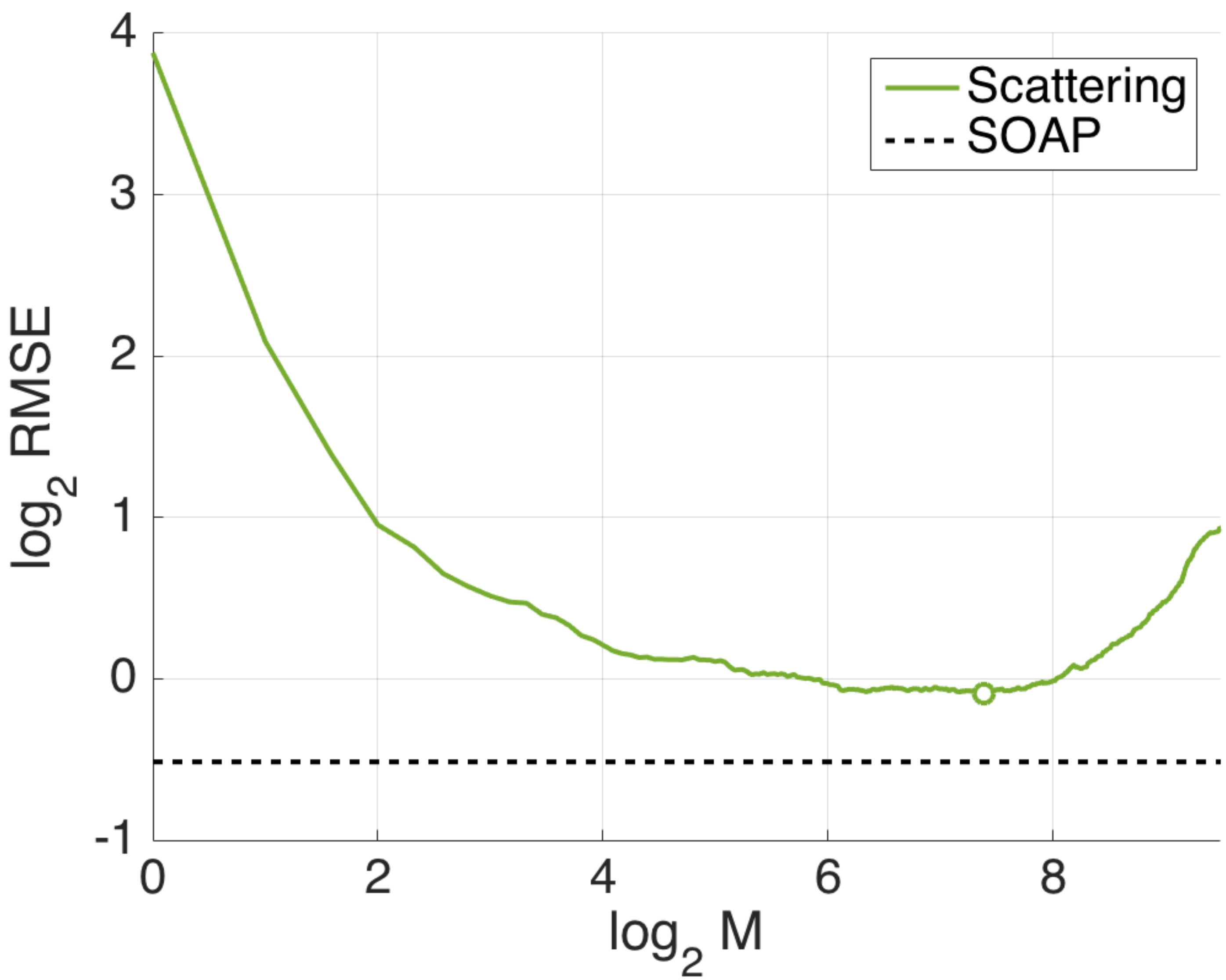




Water Molecules

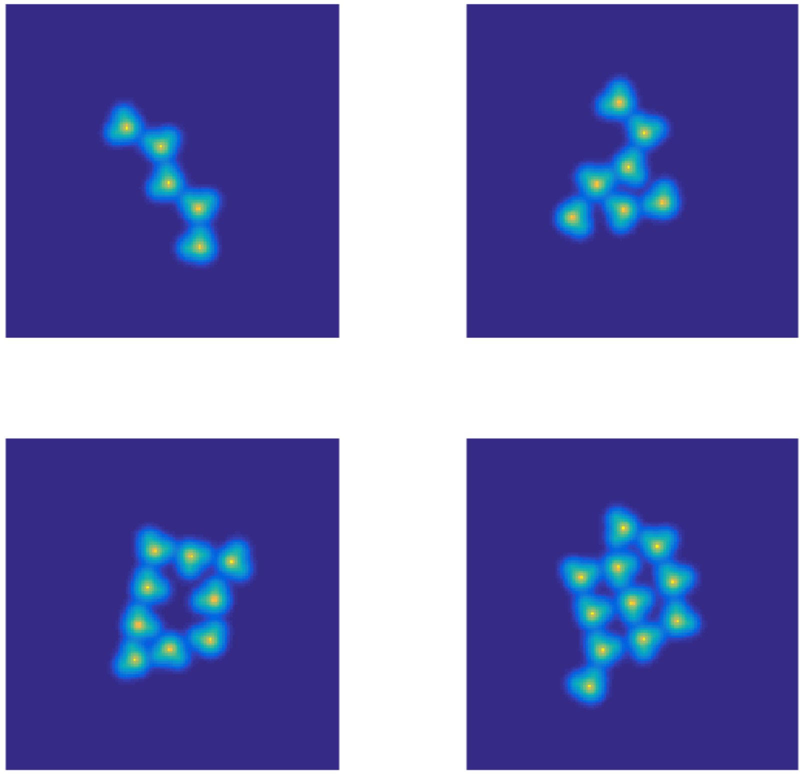
Root mean square error (meV/atom)

	SOAP	Scattering
700 Training	0.70	0.94



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

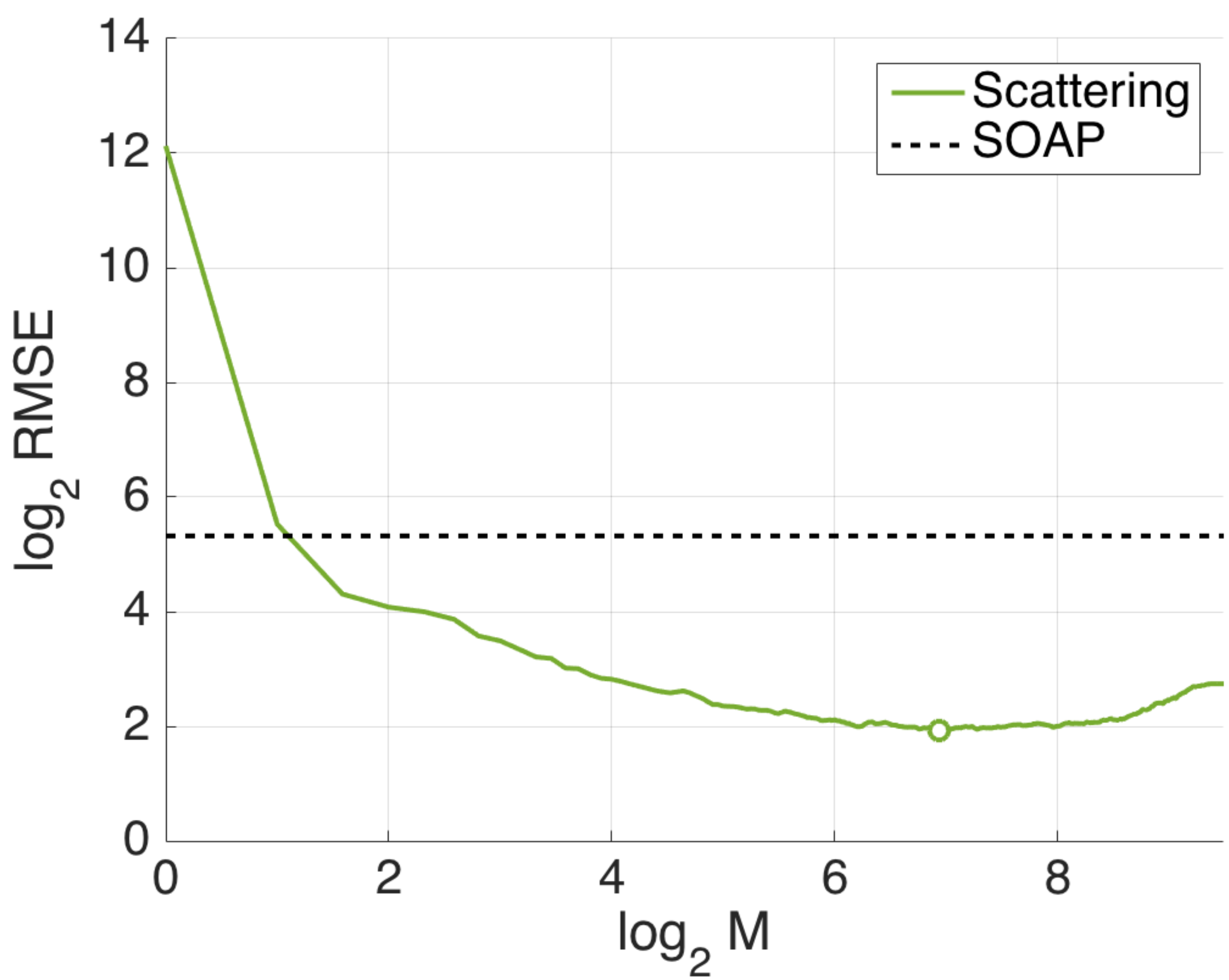
Data set:
Water molecules



Water Lithium Ion Systems

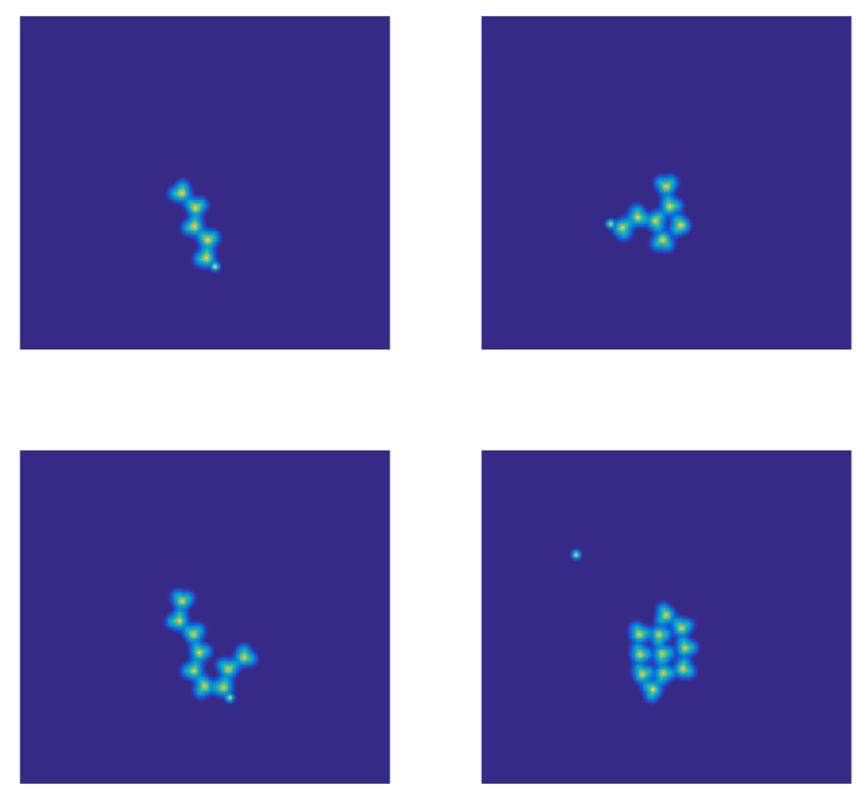
Root mean square error (meV/atom)

	SOAP	Scattering
700 Training	40	3.8



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

Data set:
Water & Lithium Ion



Future directions

- Efficient 3D software
- Rigorous link of scattering second layer with multipole methods
- Theoretical analysis of kinetic and exchange-correlation energies
- Utilization of forces
- Extensions to solid state physics, drug discovery, other many body problems in physics, other physical systems exhibiting complex multiscale behavior
- Connections with deep learning

Conclusions

To apply machine learning to many body problems, we must attack the curse of dimensionality. To do so we need to:

- Interpolate in low dimensional approximation spaces by…
- …nonlinear 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 deformations
- Cascade of operators (wavelet modulus) with the previous three properties to recover lost information