Quantum Chemistry Energy Regression with Scattering Transforms

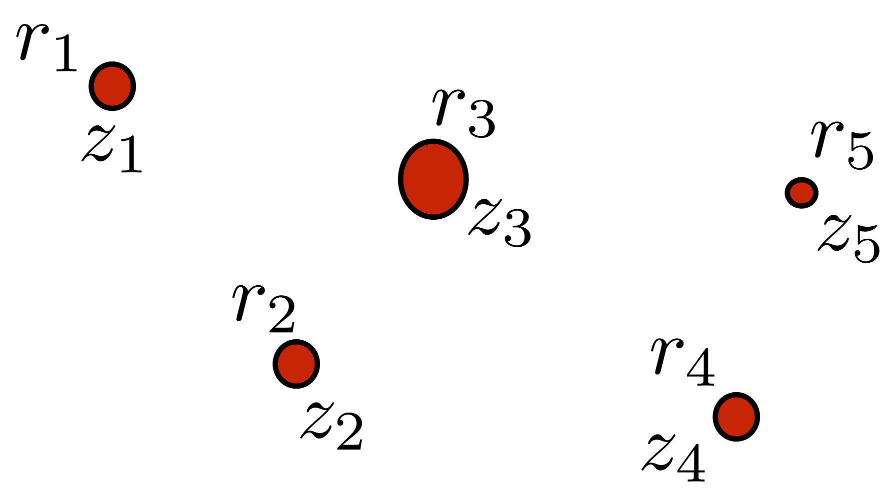
Matthew Hirn, Stéphane Mallat
École Normale Supérieure

Nicolas Poilvert
Penn-State

Quantum Chemistry Regression

Compute the atomisation energy $y(x)$ of molecules given

$$x = \{z_k(\text{charge})\,,\, r_k(\text{position})\}_{k \le d} \in \mathbb{R}^{2d}$$



- **Regression**: estimate $y(x)$ given N examples $\{x_i\,,\, y_i\}_{i \le N}$

- *Curse of dimensionality*: since $N \ll 2^d$, a priori hopeless unless dimensionality can be reduced.

- How to address this problem mathematically ?

Linear Regression

- Linear approximation of $y(x)$ in a dictionary $\Phi(x) = \{\phi_k(x)\}_k$

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

which minimizes the training error $\sum_{i=1}^N |f(x_i) - y_i|^2$

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

- Kernel $K(x,x') = \langle \Phi(x), \Phi(x')\rangle = \sum_k \phi_k(x)\,\phi_k(x')$

$$\Rightarrow\;\; f(x) = \sum_{i=1}^N \alpha_i\, K(x,x_i)$$

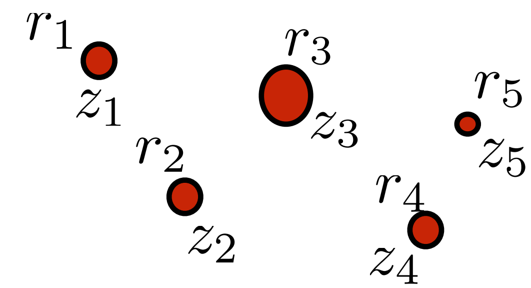
Problem: design $\Phi(x)$ to minimize $\mathbb{E}(|f(x) - y(x)|^2)$,

$\Phi(x)$ or $K(x,x')$ have same regularity properties as $y(x)$.

Energy Properties

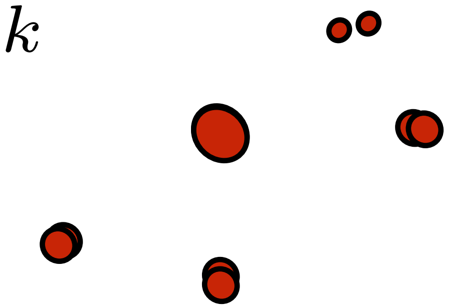
(Bartok, Csanyi, Kondor)

Atomisation energy $y(x)$ of a molecule $x = \{z_k, r_k\}_{k \leq d}$:



- Invariant to permutations of the index k.

- Invariant to rigid movements of positions $\{r_k\}_k$

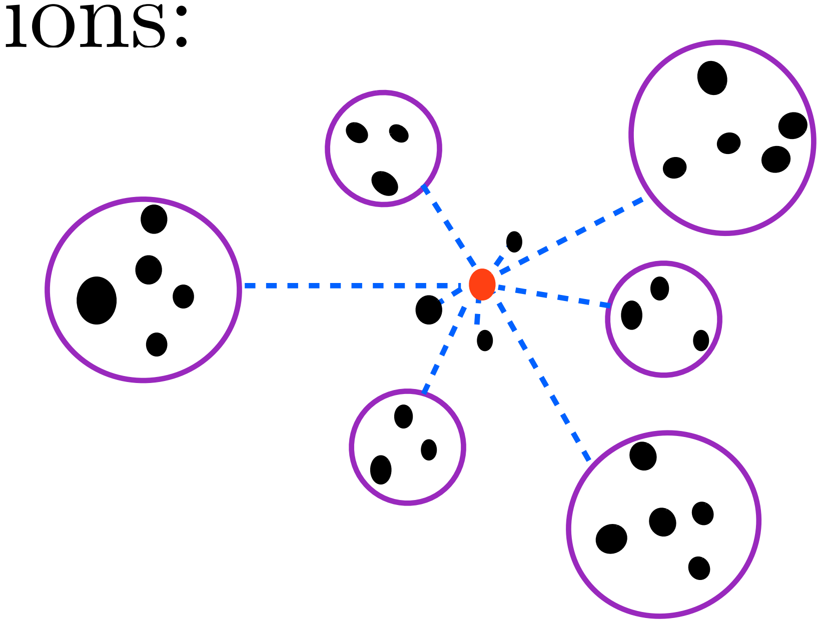


- Regular variations relatively to deformations

- Factorization in multiscale local interactions:

  covalent bounds, Van Der Waal forces...

  Can reduce d into $O(\log d)$ interactions



Generic properties: same in images

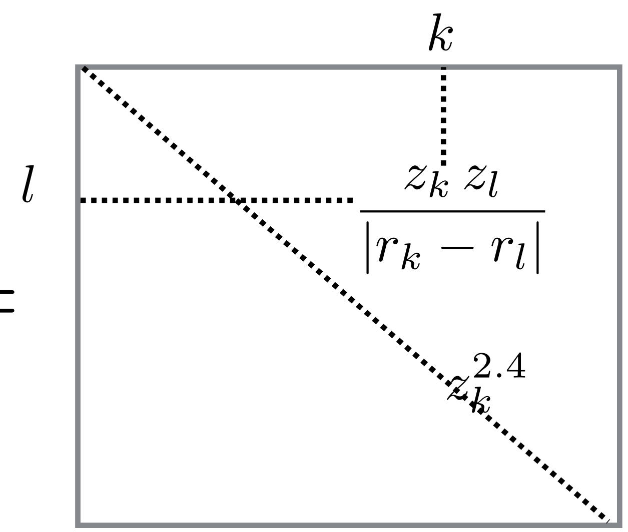
Overview

- Coulomb kernel representations
- Density functional approach to representation
- From Fourier to wavelet energy regressions
- Wavelet scattering dictionaries: deep networks without learning
- Numerical energy regression results
- Relations with image classification and deep networks

Coulomb Kernel

(Rupp, Tkatchenko, Muller, von Lilienfeld)

- Coulomb kernel: $K(x, x') = e^{-\lambda \, \|M(x) - M(x')\|}$

with $M(x) =$ 

: invariant to rigid mouvements

stable to deformations

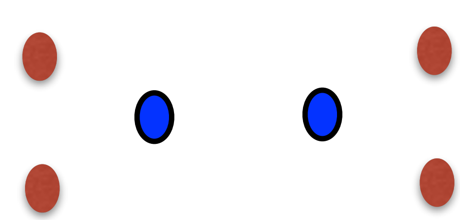
but $K(x, x')$ is not invariant to the molecule indexing

partly fixed with orderings based on column energy

but creates deformation instabilities and non differentiability

for symmetric molecules:

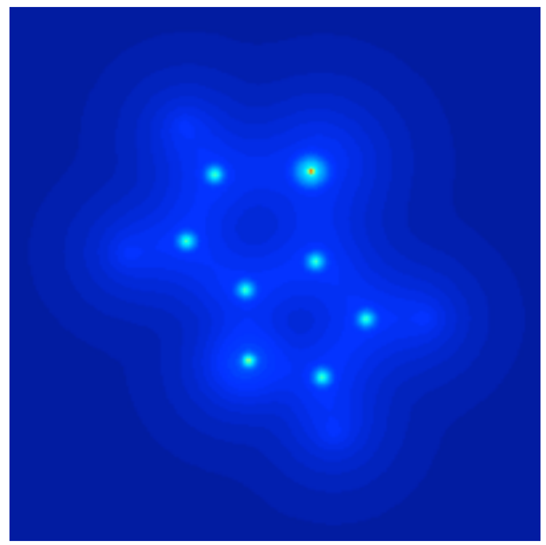
$\Rightarrow$ no accurate force fields



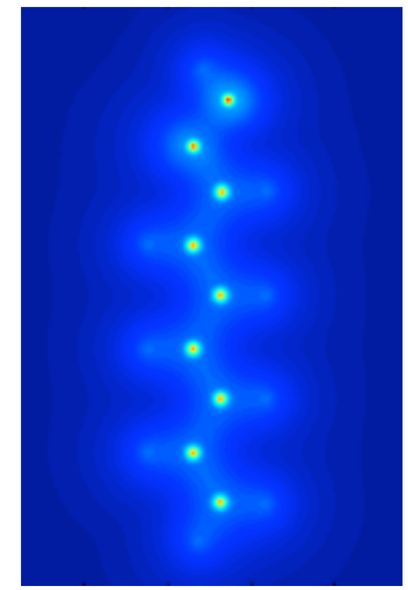
Density Functional Theory

- Computes the energy of a molecule x from its electronic probability density $\rho_x(u)$ for $u \in \mathbb{R}^3$

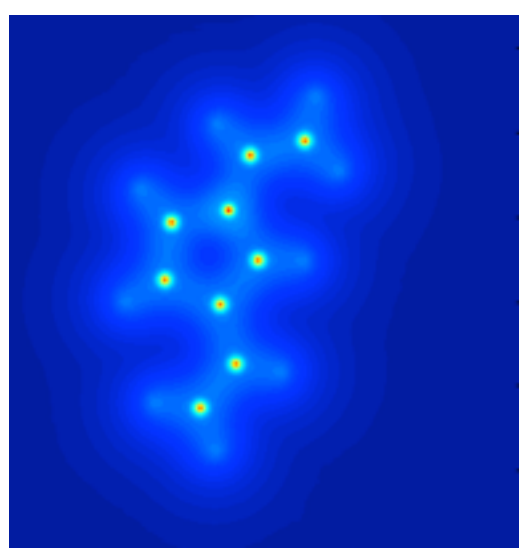
Organic molecules with Hydrogne, Carbon Nitrogen, Oxygen Sulfur, Chlorine



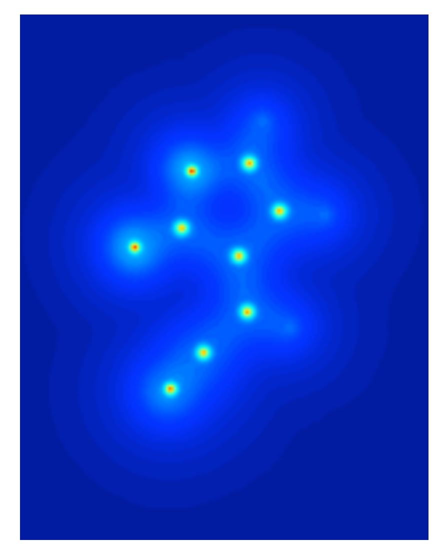
H_4C_6OS



H_9C_7NO



H_9C_8N



$H_3C_6NO_2$

Density Functional Theory

Kohn-Sham model:

$$E(\rho) = T(\rho) + \int \rho(u)\, V(u) + \frac{1}{2}\int \frac{\rho(u)\rho(v)}{|u-v|} du dv + E_{xc}(\rho)$$

$E(\rho)$	$T(\rho)$	$\int \rho(u)\, V(u)$	$\frac{1}{2}\int \frac{\rho(u)\rho(v)}{\vert u-v\vert} du dv$	$E_{xc}(\rho)$
Molecular energy	Kinetic energy	electron-nuclei attraction	electron-electron Coulomb repulsion	Exchange correlat. energy

At equilibrium:

$$y(x) = E(\rho_x) = \min_{\rho} E(\rho)$$

- ρ_x is independent of the atom indexing in x.
- Given ρ_x, we could regress $E(\rho_x)$ in a dictionary $\Phi(\rho_x)$
- Which dictionary ? How to approximate ρ_x ?

Coulomb Interactions in Fourier

- Coulomb potential energy:

$$E_C(\rho) = \iint_{\mathbb{R}^6} \rho(u)\rho(v)V(u-v)\,dudv$$

with $V(u) = |u|^{-1}$: singular

Diagonalized in Fourier: $\hat{\rho}(\omega) = \int_{\mathbb{R}^3} \rho(u)\, e^{i\omega.u}\, du$

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

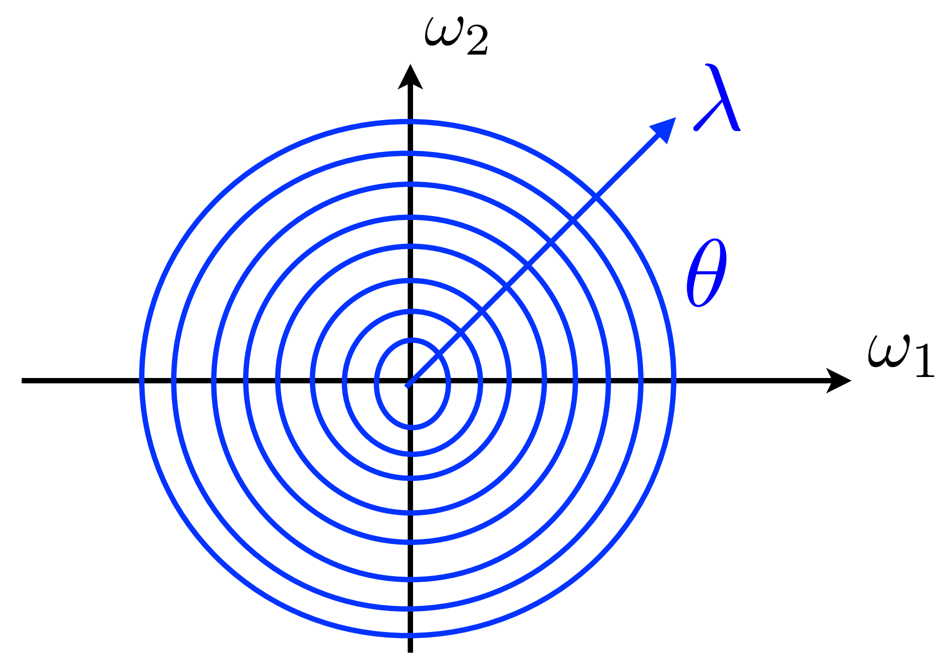
with $\widehat{V}(\omega) = 4\pi\,|\omega|^{-2}$

Coulomb in Fourier Dictionary

$$E_C(\rho) = C \int |\omega|^{-2}\, |\hat\rho(\omega)|^2\, d\omega$$

In polar coordinates $\omega = (\lambda, \theta)$:

$$E_C(\rho) = C \int \lambda^{-2} \left(\int |\hat\rho(\lambda,\theta)|^2\, d\theta \right) d\lambda$$



Fourier dictionary: $\left(\phi_k(\rho) = \int |\hat\rho(k\epsilon,\theta)|^2\, d\theta \right)_k$

Invariant to translations and rotations

$$\Rightarrow \quad E_C(\rho) = \sum_{k=1}^{\epsilon^{-2}} w_k\, \phi_k(\rho)\, (1 + O(\epsilon)) \quad \text{with } w_k = C\, k^{-2}$$

Problems: needs $M = \epsilon^{-2}$ terms: not sparse

the $\phi_k(\rho)$ are not stable to deformations

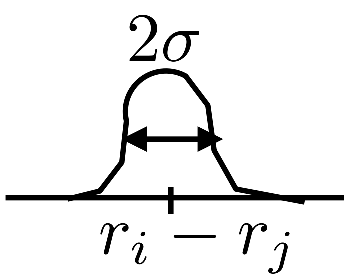
Large Scale Instabilities

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

- **Deformations produce instabilities at large distances:**

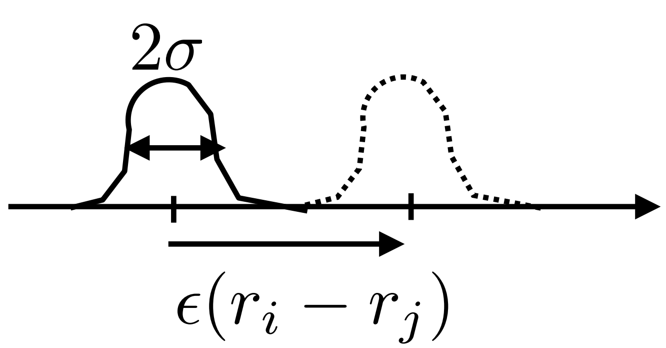
  $\rho(u)$: bumps of width σ at positions $\{r_i\}_i$

  $\Phi\rho(\tau)$: bumps of width 2σ at positions $\{r_i - r_j\}_{i,j}$



  A small deformation changes distances by $\epsilon(r_i - r_j)$

  Unstable if $|r_i - r_j| \geq \sigma/\epsilon$



(SOAP: Bartok, Csanyi Kondor)

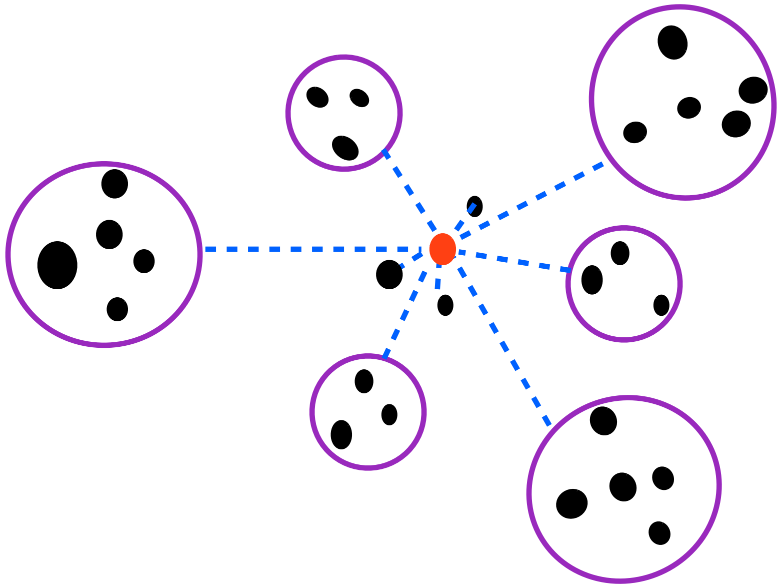
Coulomb Multiscale Factorizations

- Multiscale regroupment of interactions:

For an error ϵ, interactions can be reduced to $O(\log \epsilon)$ groups

Fast multipoles (*Rocklin, Greengard*)

Potential $V(u) = |u|^{-1} \quad \Rightarrow$

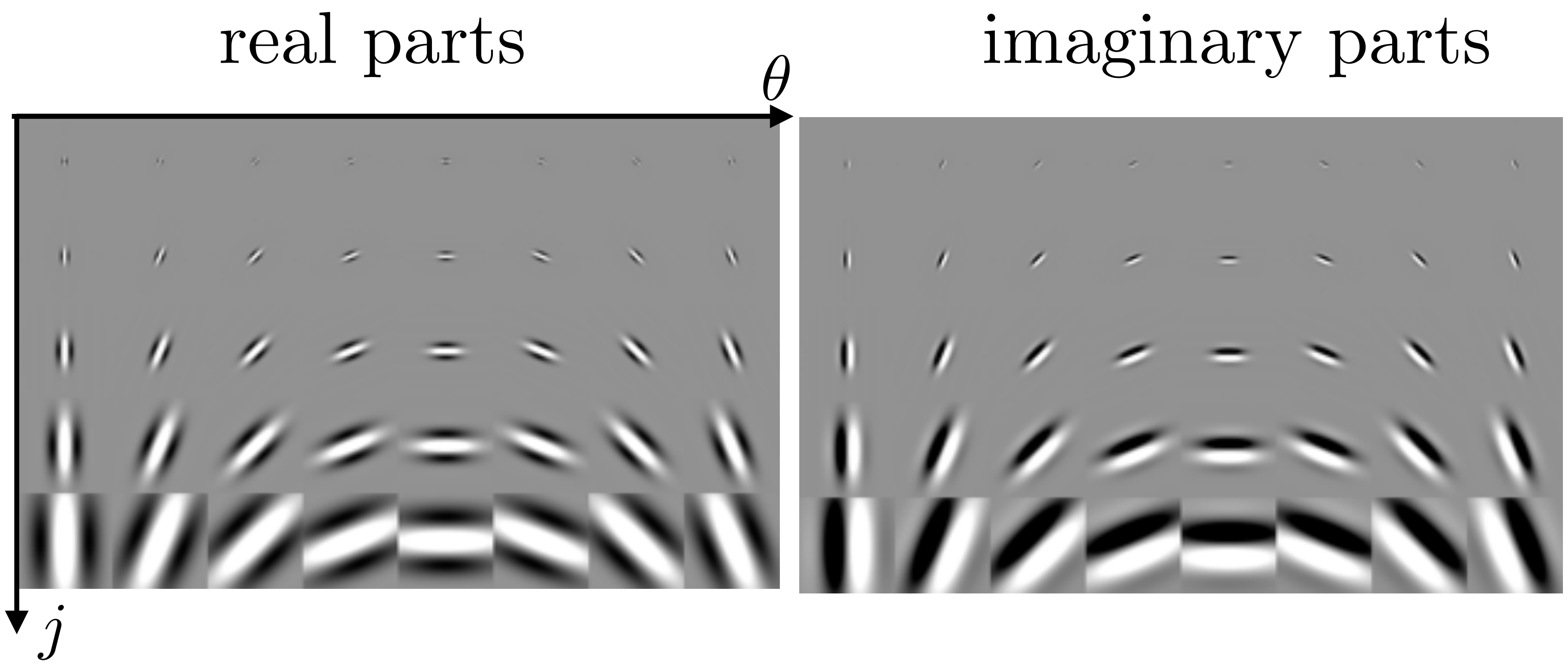


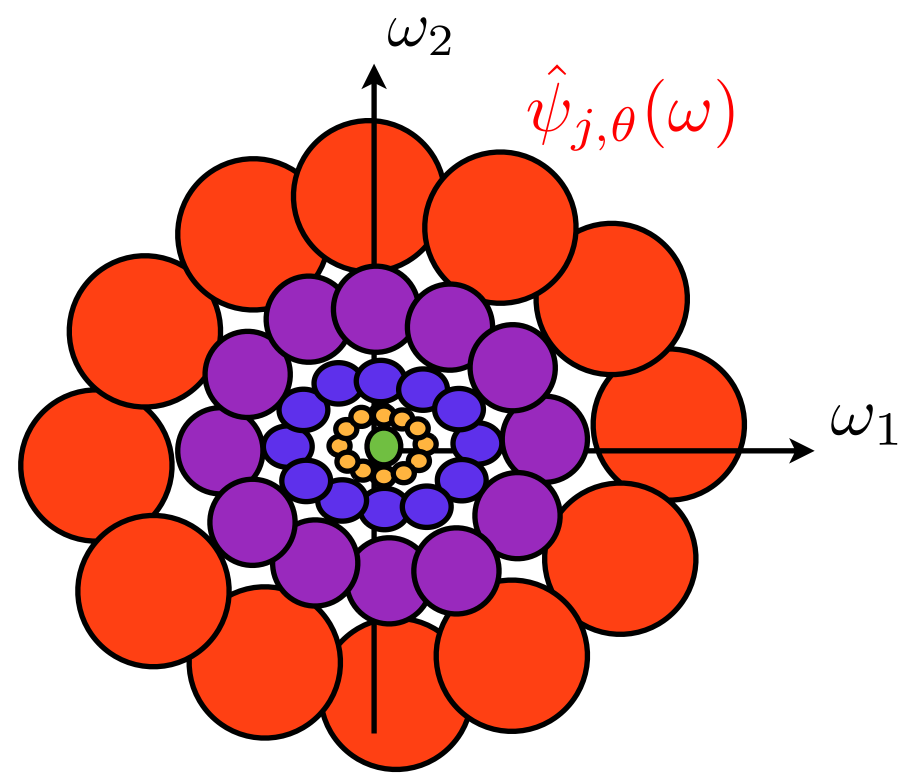
$\Rightarrow \; E_C(\rho)$ can be computed with $O(|\log \epsilon|^2)$ operations.

Scale separation with Wavelets

- Complex wavelet: $\psi(u) = g(u) e^{i\xi . u}$, $u = (u_1, u_2)$

  rotated and dilated: $\psi_{j,\theta}(u) = 2^{-2j}\, \psi(2^{-j} R_\theta u)$





Total charge: $\int \rho(v) dv$

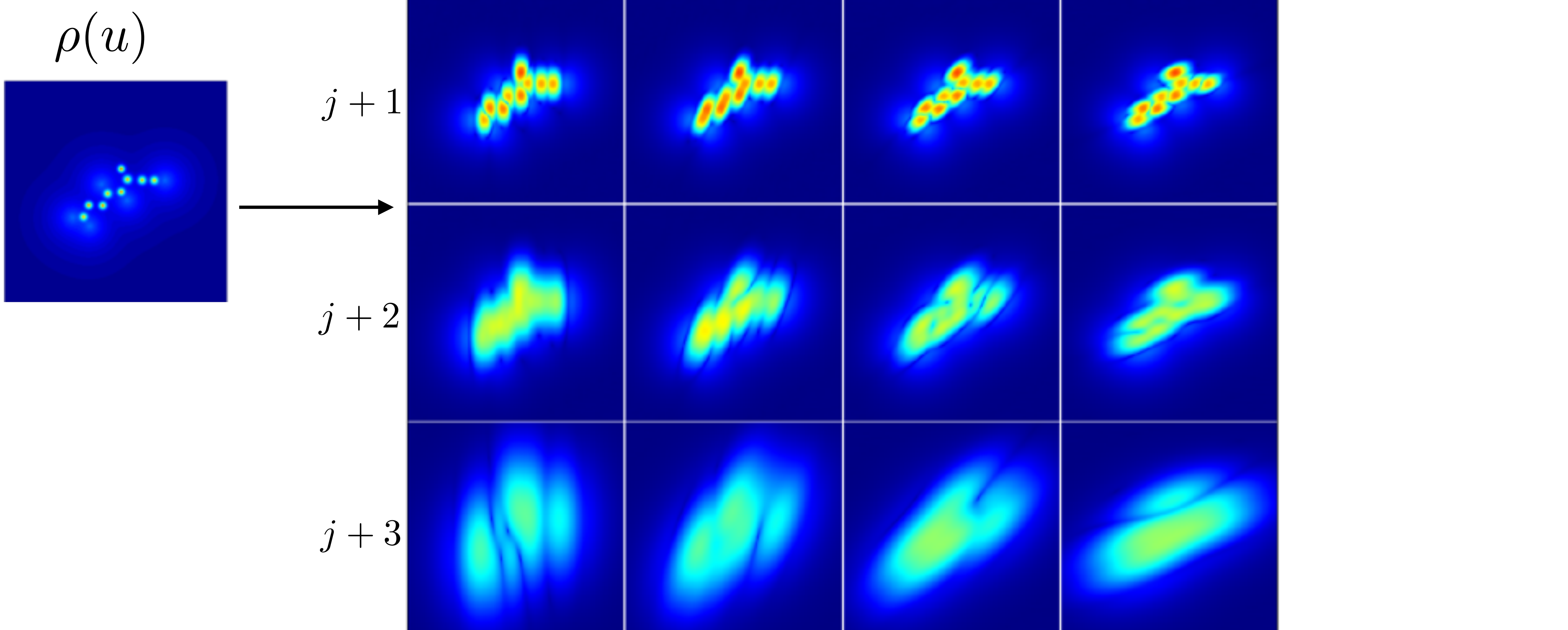
Wavelet coefficients: $\rho \star \psi_{j,\theta}(u) = \int \rho(v)\, \psi_{j,\theta}(u - v)\, dv$ $\mapsto \rho$

interaction at a scale 2^j along a direction θ

stable to deformations

Wavelet Interference for Densities

$$\rho = \sum_k z_k \delta(u - r_k) \Rightarrow |\rho \star \psi_{j,\theta}(u)| = \Big| \sum_k z_k \, \psi_{j,\theta}(u - r_k) \Big|$$



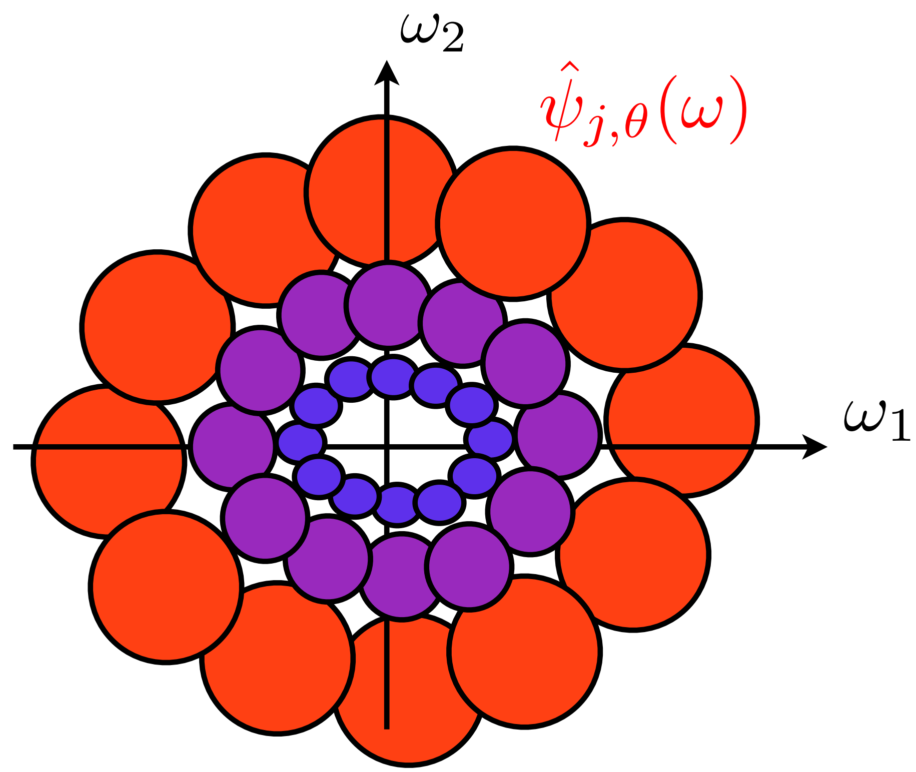
Sparse Wavelet Regression

Wavelet dictionary:

$$\left\{ \phi_j(\rho) = \int_0^{2\pi} \int_{\mathbb{R}^2} |\rho \star \psi_{j,\theta}(u)|^2 \, du d\theta \right\}_j$$

Translation and rotation invariant

Stable to deformations



Theorem: For any $\epsilon > 0$ there exists wavelets with

$$E_C(\rho) = \sum_j v_j \, \phi_j(\rho) \Big(1 + O(\epsilon) \Big)$$

with a sparse regression of $M = O(|\log \epsilon|^2)$ terms.

Dictionaries for Quantum Energies

- The coulomb energy term $E_C(\rho)$ is quadratic in ρ
- Chemical bound energy rather increase linearly with ρ

- Fourier dictionary:

$$\phi_k^1(\rho) = \int |\hat{\rho}(k\epsilon, \theta)|\, d\theta \;\text{ and }\; \phi_k^2(\rho) = \int |\hat{\rho}(k\epsilon, \theta)|^2\, d\theta$$

In numerics: 1500 vectors

- Wavelet dictionary:

$$\phi_j^1(\rho) = \iint |\rho \star \psi_{j,\theta}|\, du d\theta \;\text{ and }\; \phi_j^2(\rho) = \iint |\rho \star \psi_{j,\theta}|^2\, du d\theta$$

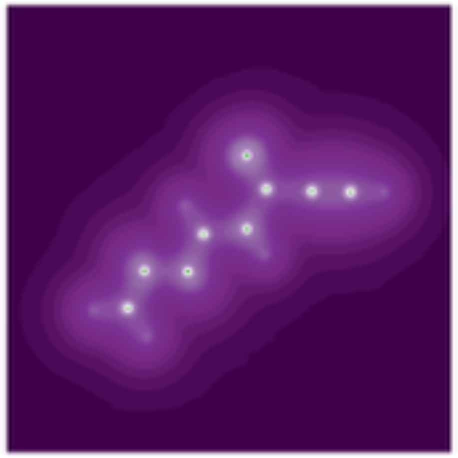
In numerics: 60 vectors

Atomization Density

- We do not know the electronic density ρ_x at equilibrium.
- The electronic density ρ_x of $x = \{z_k \,,\, r_k\}_{k \leq d}$ is approximated by the sum of the densities of all atoms:

$$\tilde{\rho}_x(u) = \sum_{k=1}^{d} \rho_{z_k}(u - r_k)$$

Electronic density $\rho_x(u)$



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

- Fourier and Wavelet dictionaries: $\{\phi_k^1(\tilde{\rho}_x) \,,\, \phi_k^2(\tilde{\rho}_x)\}_k$

Sparse Linear Regressions

- Sparse regression in a dictionary $\{\phi_k(\tilde{\rho}_x)\}_k$ by selecting M dictionary vectors:

$$f_M(x) = \sum_{m=1}^{M} w_m \, \phi_{k_m}(\tilde{\rho}_x)$$

which minimise the error $\sum_{i=1}^{N} |f_M(x_i) - y_i|^2$

- Greedy selection of the $\{\phi_{k_m}(\tilde{\rho}_x)\}_{m \leq M}$, one at a time, with an orthogonal least square pursuit which decorelate vectors.

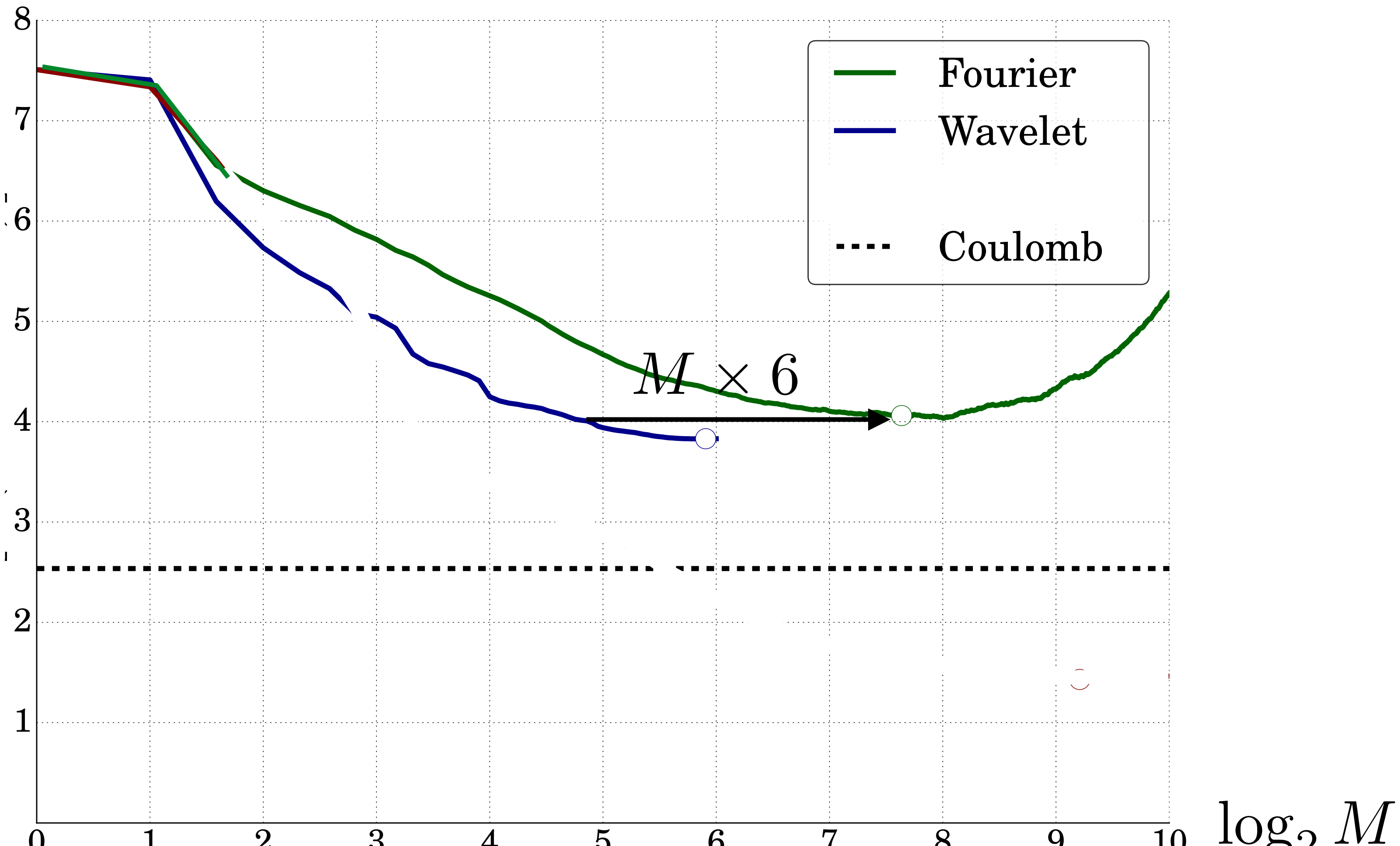
- Cross-validation on M.

Fourier and Wavelets Regressions

Data basis $\{x_i \, , \, y_i = E(\rho_{x_i})\}_{i \leq N}$ of 4357 planar molecules

$$\text{Regression:} \quad f_M(x) = \sum_{m=1}^{M} w_m \, \phi_{k_m}(\tilde{\rho}_x)$$

Testing error
$2^{-1} \log_2 \mathbb{E}|f_M(x) - y(x)|^2$



Energy Regression Results

$$f_M(x) = \sum_{m=1}^{M} w_m \, \phi_{k_m}(\tilde{\rho}_x)$$

RMS testing error $(\mathbb{E}|f_M(x) - y(x)|^2)^{1/2}$ in kcal/mol:

Training size	Fourier	Wavelet	Coulomb
4357	16.7	14.2	**5.8**
454 (QM7)	16.1	15.4	20.5

$$MSE = bias + variance$$

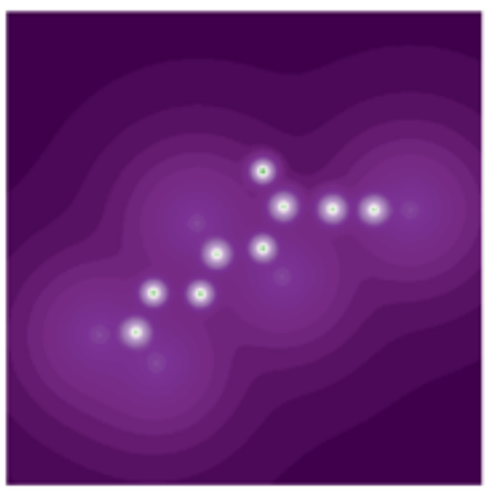
For Fourier and Wavelet the bias dominates the error:

- Fourier dictionary not stable to deformations, not multiscale

- Wavelet dictionary too small : 60 vectors.

Wavelet Dictionary

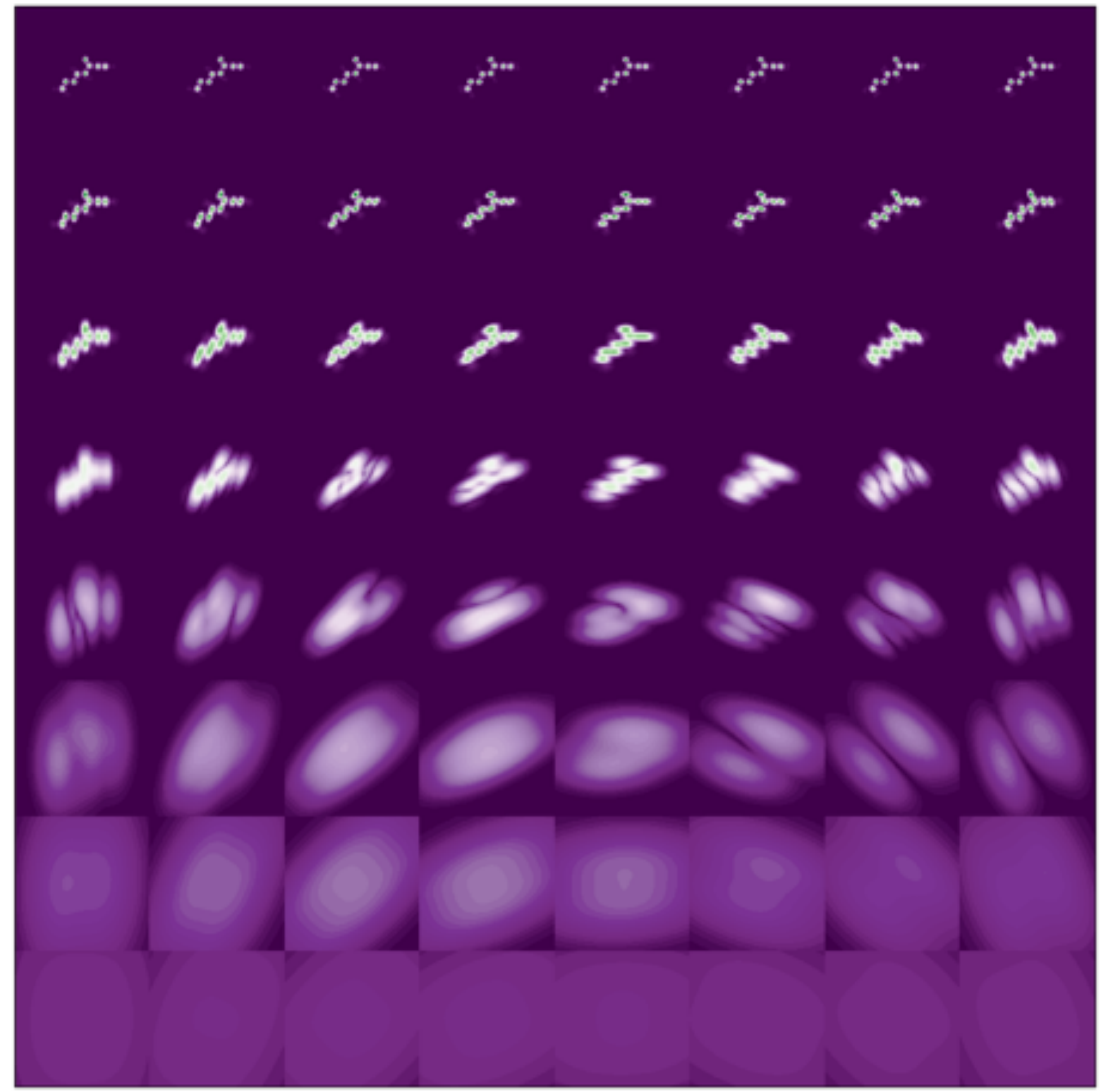
1st network layer



$\rho(u)$

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

Rotations θ_1

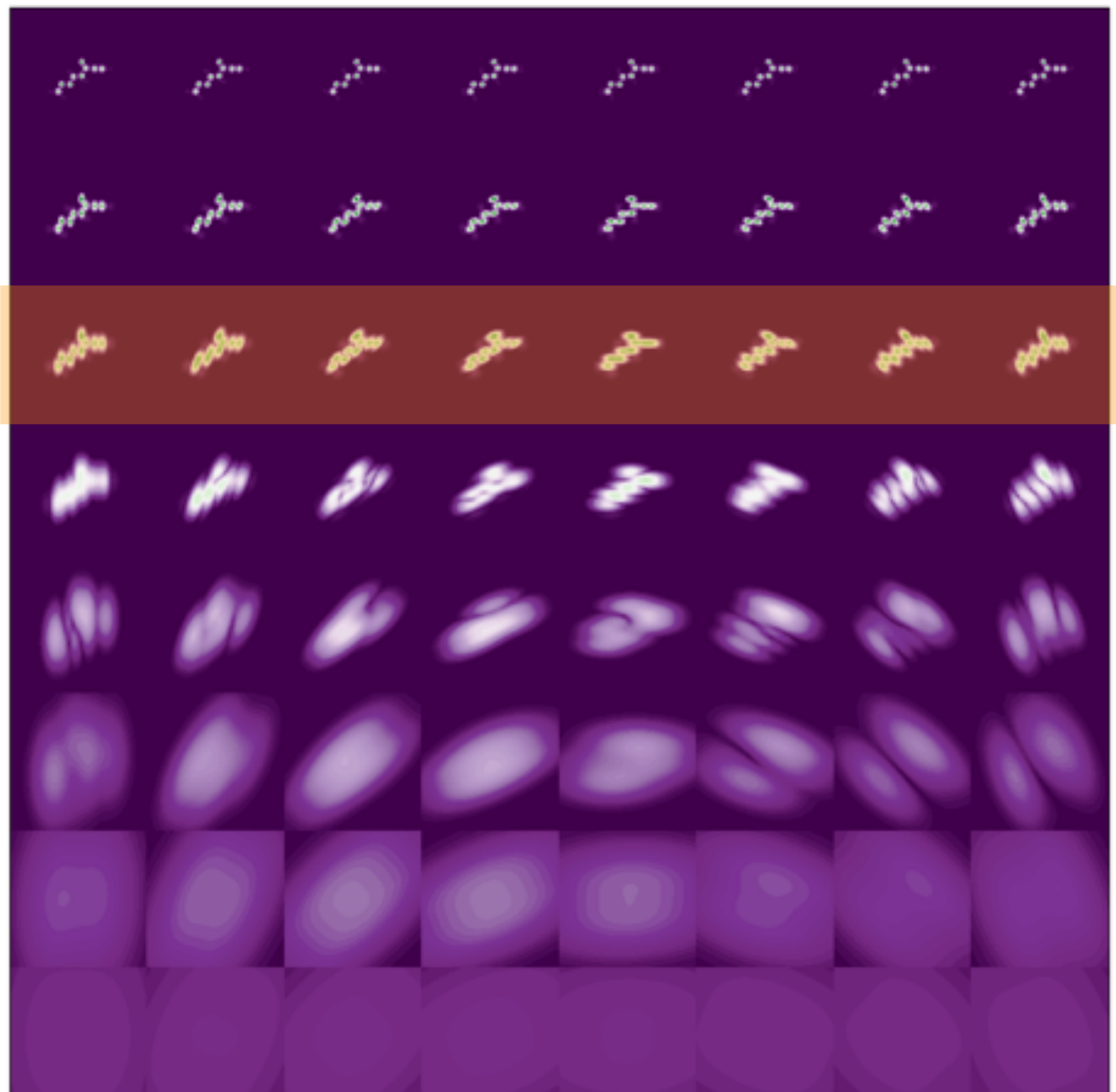


Scales j_1

Wavelet Dictionary

1st network layer

Rotations θ_1



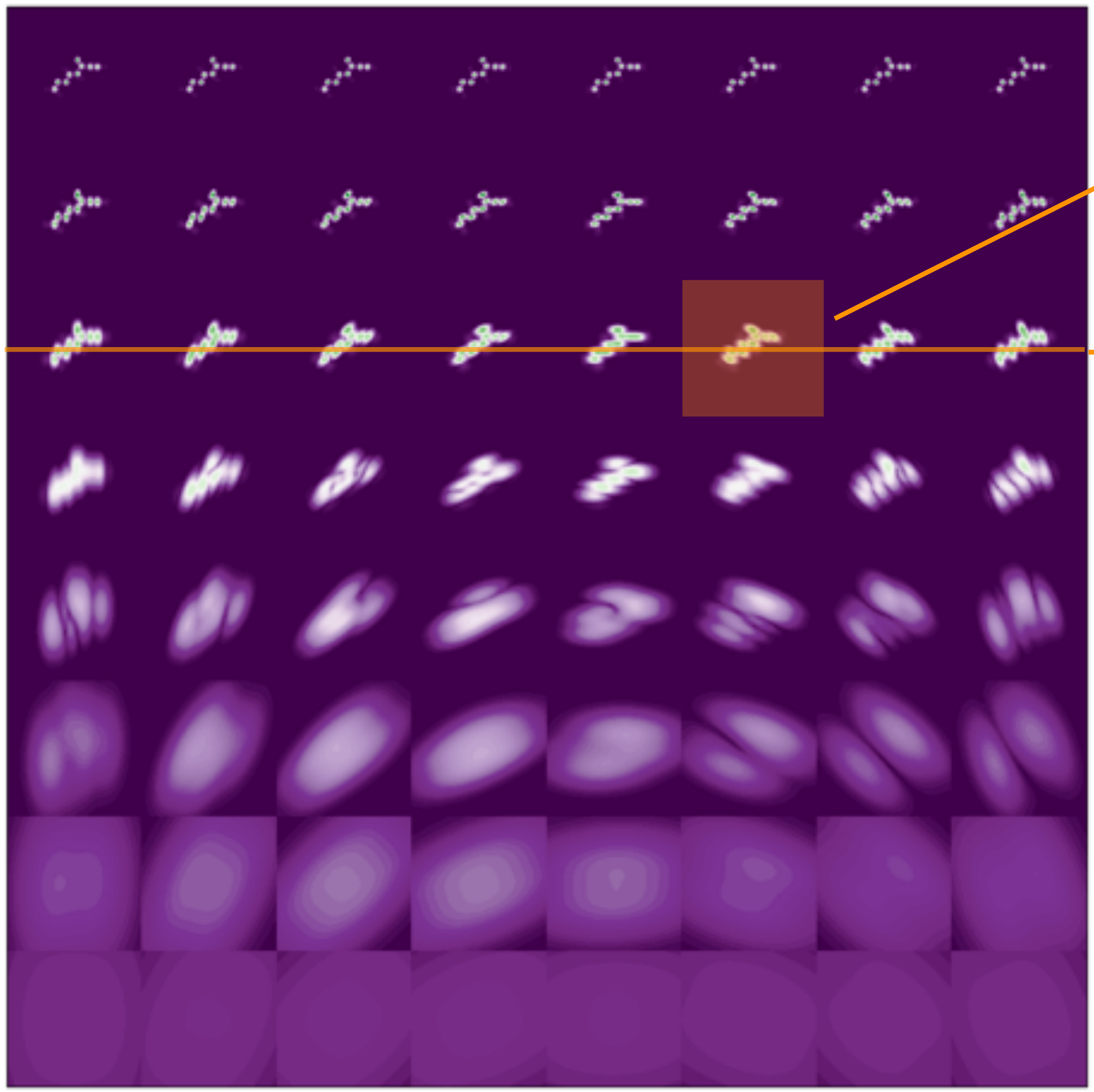
Scales j_1

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

$$\phi^1_{j_1}(\rho) = \int_{\mathbb{R}^2} \int_0^{2\pi} |\rho * \psi_{j_1,\theta_1}(u)| \, d\theta_1 \, du$$

Scattering Dictionary

Rotations θ_1



Scales j_1

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

2nd Order Interferences

Recover translation variability:

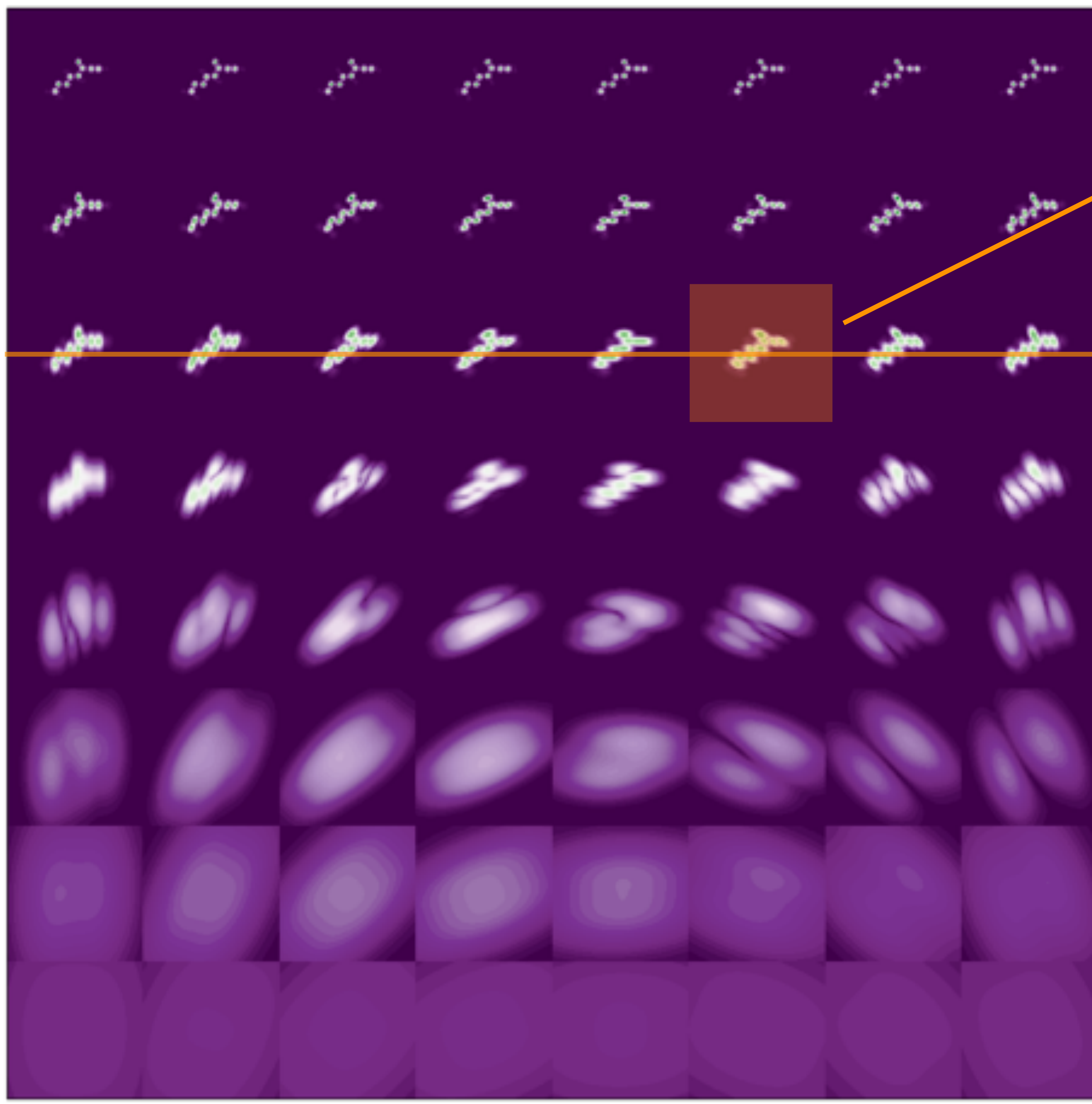
$$|\rho * \psi_{j_1,\theta_1}| * \psi_{j_2,\theta_2}(u)$$

Recover rotation variability:

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

Scattering Dictionary

Rotations θ_1



Scales j_1

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

2nd Order Interferences

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:

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

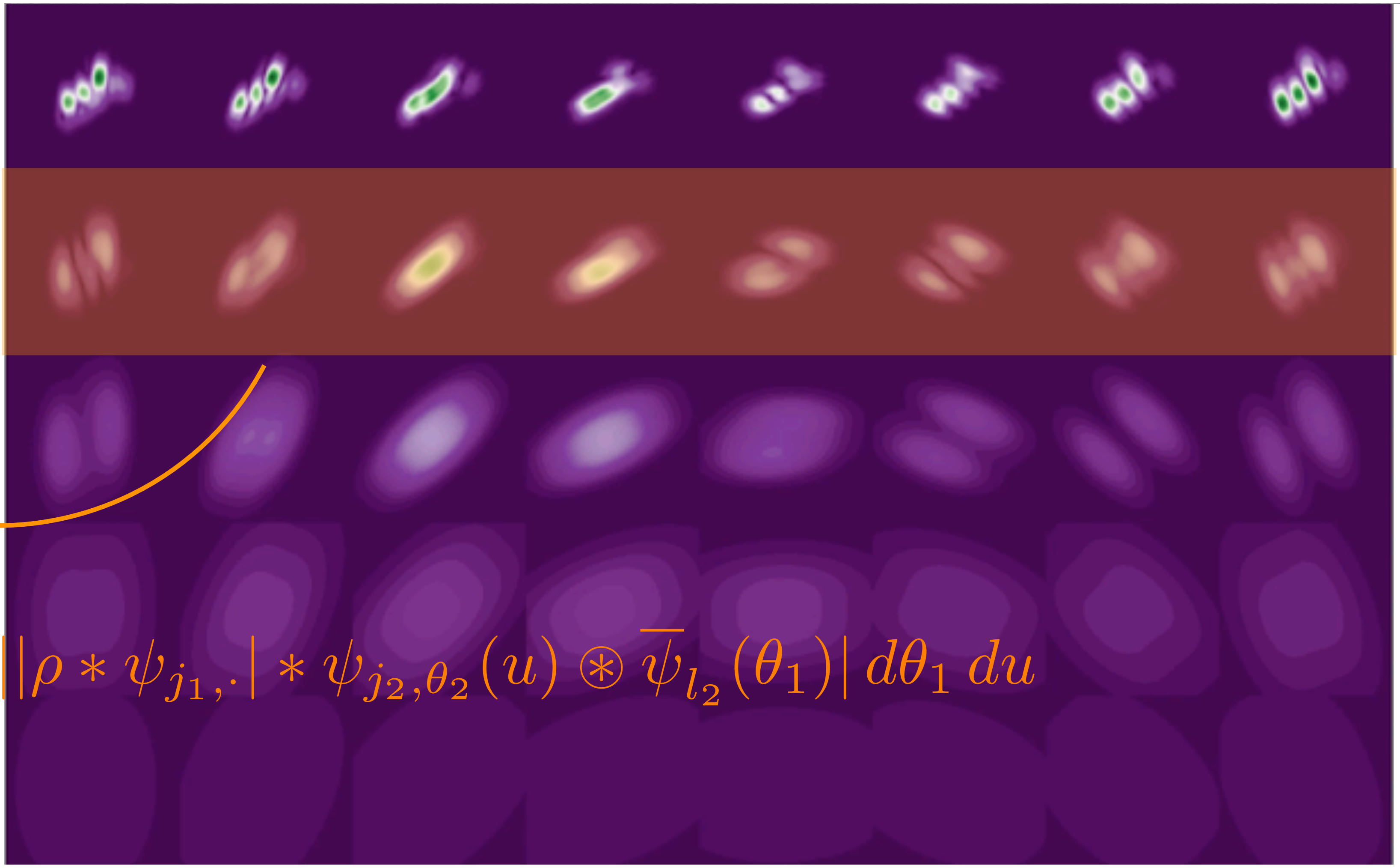
Scattering Second Order

1st network layer

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

2nd network layer

θ_1



Scales j_2

Rotations θ_2

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

j_1, l_2 fixed

$$\phi_{j_1,j_2,\theta_2,l_2}(\rho) = \int_{\mathbb{R}^2} \int_0^{2\pi} ||\rho * \psi_{j_1,\cdot}| * \psi_{j_2,\theta_2}(u) \circledast \overline{\psi}_{l_2}(\theta_1)| \, d\theta_1 \, du$$

Scattering Dictionary

- Sparse regression of $f(x)$ computed in a scattering dictionary computed for the atomization density: $\rho = \tilde{\rho}_x$.

$$\underbrace{\Big\{ \phi^1_j(\tilde{\rho}_x)\,,\, \phi^2_j(\tilde{\rho}_x) \Big\}_j}_{\text{60 vectors}} \cup \underbrace{\Big\{ \phi^1_{j,j_2,\theta_2,\ell_2}(\tilde{\rho}_x)\,,\, \phi^2_{j,j_2,\theta_2,\ell_2}(\tilde{\rho}_x) \Big\}_{j,j_2,\theta_2,\ell_2}}_{10^4 \text{ vectors}}$$

Invariant by translations and rotations

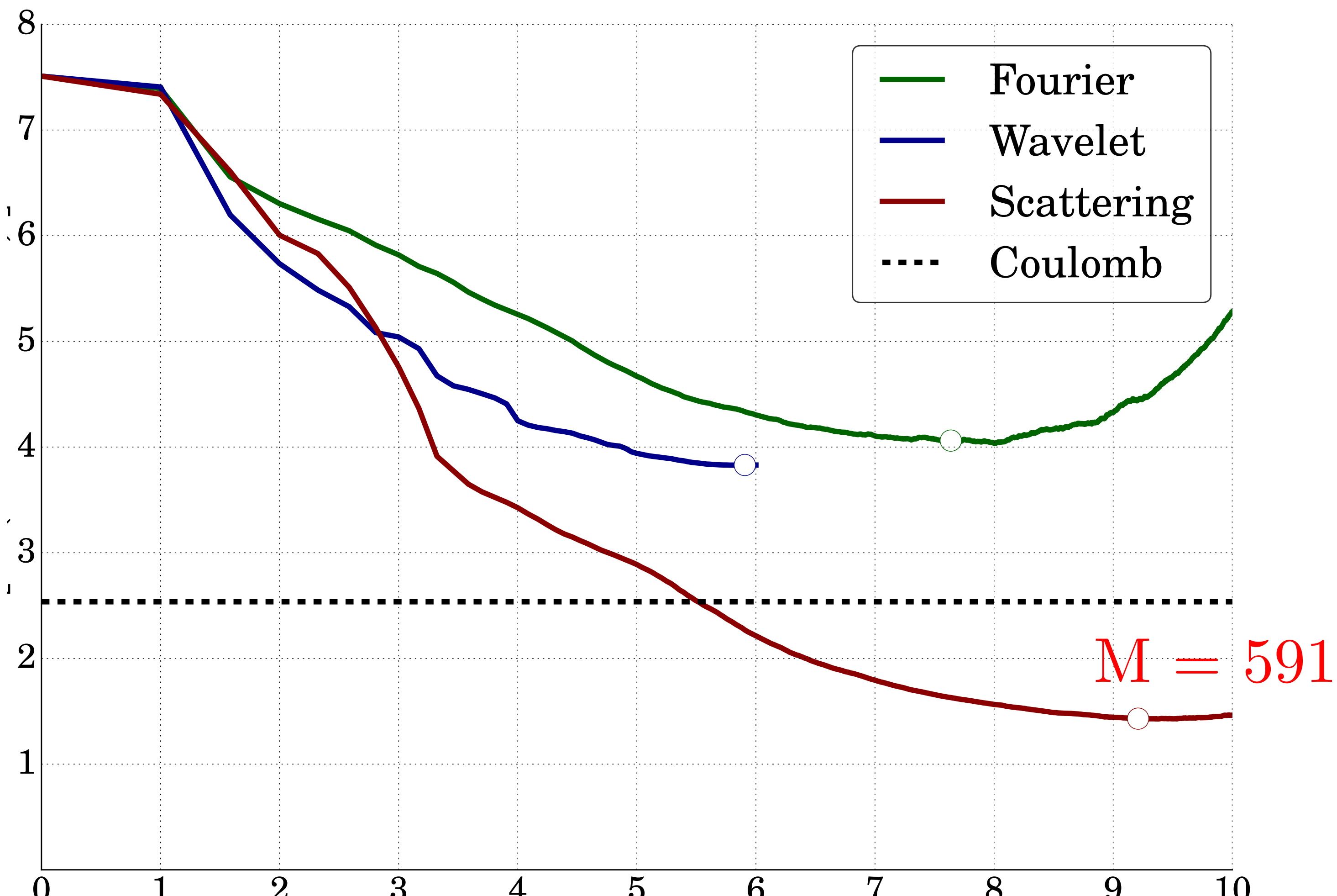
Stable to deformations

Scattering Regression

Data basis $\{x_i \, , \, f(x_i) = E(\rho_{x_i})\}_{i \leq N}$ of 4357 planar molecules

$$\text{Regression:} \quad f_M(x) = \sum_{m=1}^{M} w_m \, \phi_{k_m}(\tilde\rho_x)$$

Testing error
$2^{-1} \log_2 \mathbb{E}|f_M(x) - y(x)|^2$



$\log_2 M$

Energy Regression Results

$$f_M(x) = \sum_{m=1}^{M} w_m \, \phi_{k_m}(\tilde{\rho}_x)$$

RMS testing error $(\mathbb{E}|f_M(x) - y(x)|^2)^{1/2}$ in kcal/mol:

Training size	Fourier	Wavelet	Coulomb	Scattering
4357	16.7	14.2	5.8	**2.7**
454 (QM7)	16.1	15.4	20.5	**9.0**

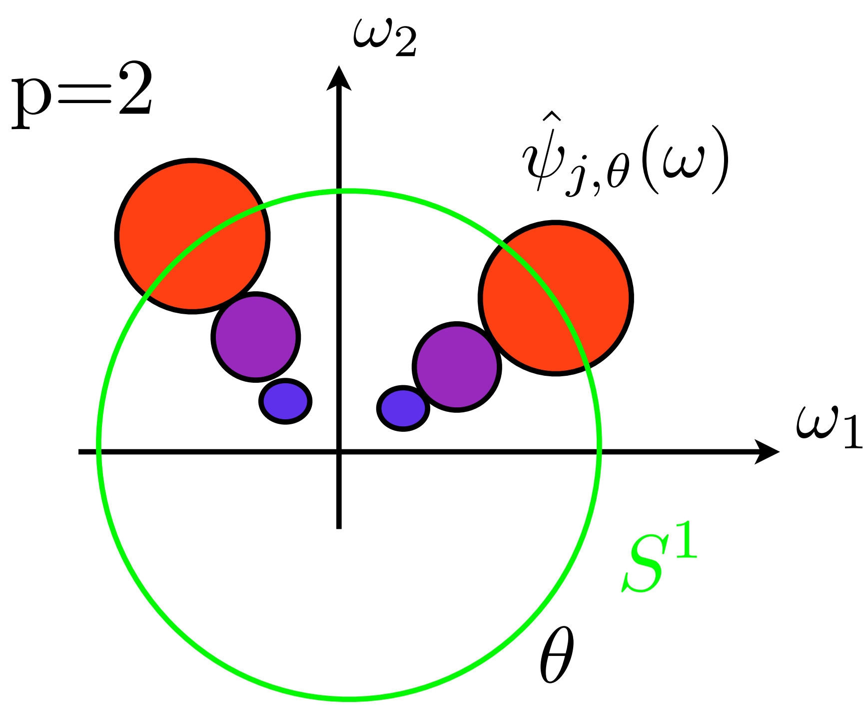
- For field calculations:

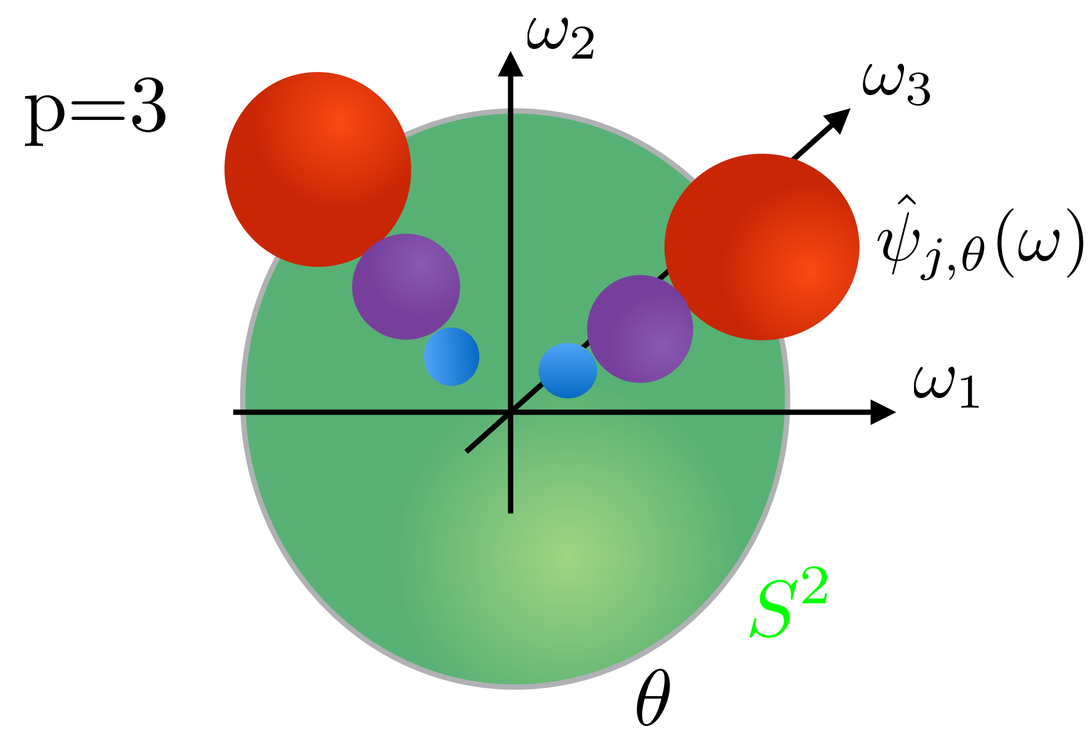
$$\nabla f_M(x) = \sum_{m=1}^{M} w_m \, \nabla \phi_{k_m}(\tilde{\rho}_x)$$

From 2D to 3D Scattering

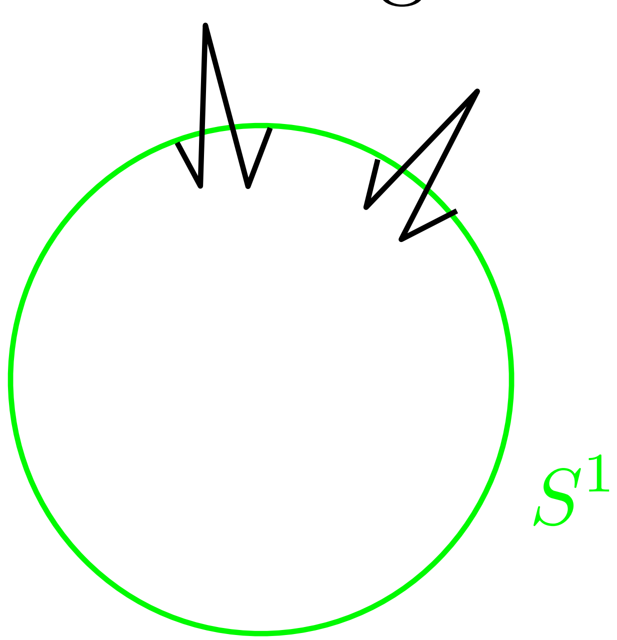
Translation wavelet: $\psi_{j,\theta}(u) = 2^{-pj}\,\psi(2^{-j} R_\theta^{-1} u)$

$$\theta \in S^{p-1} \;,\; u \in \mathbb{R}^p$$



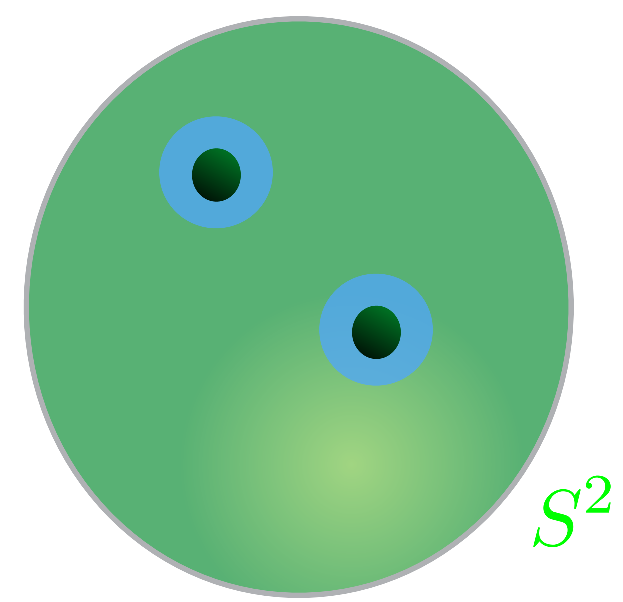


- Scattering:



Angular wavelet:

$$\overline{\psi}_\ell(\theta) = 2^{(-p+1)\ell}\,\overline{\psi}(2^{-\ell}\theta)$$



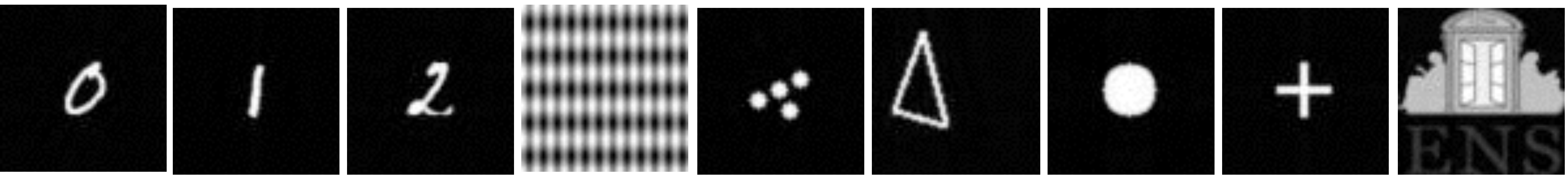
To be programmed...

Reconstruction from Scattering

Joan Bruna

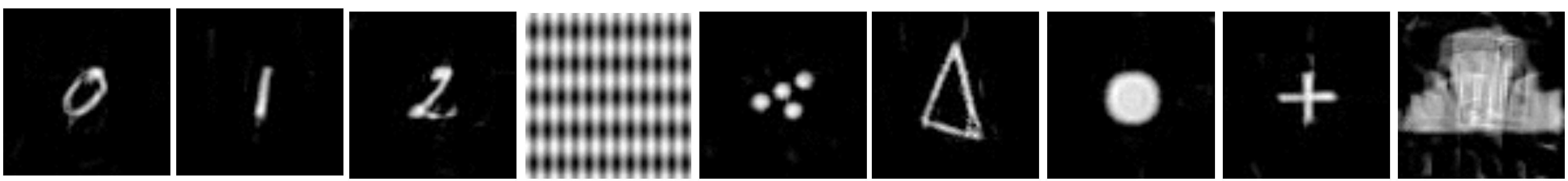
- Scattering without rotation invariance: no angle averaging.

Original images of N^2 pixels:



Order $m = 2$

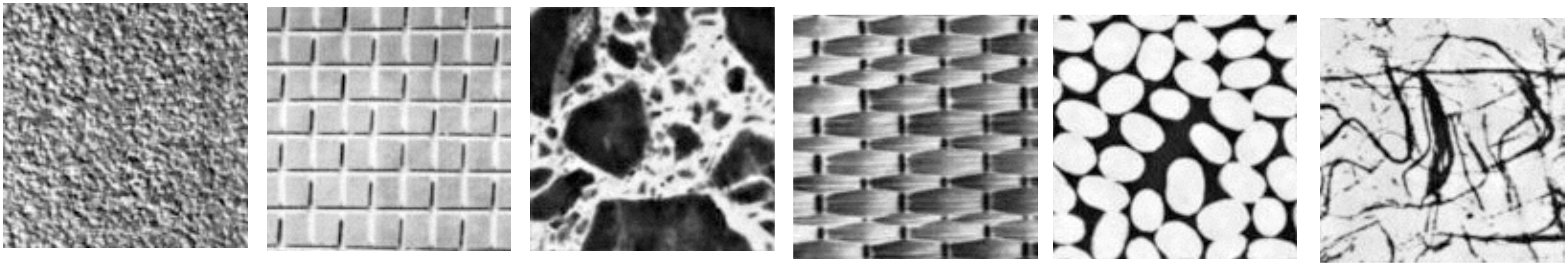
Reconstruction from $\{\|x\|_1\,,\ \|x \star \psi_{\lambda_1}\|_1\,,\ \||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1\}$: $O(\log_2^2 N)$ coeff.



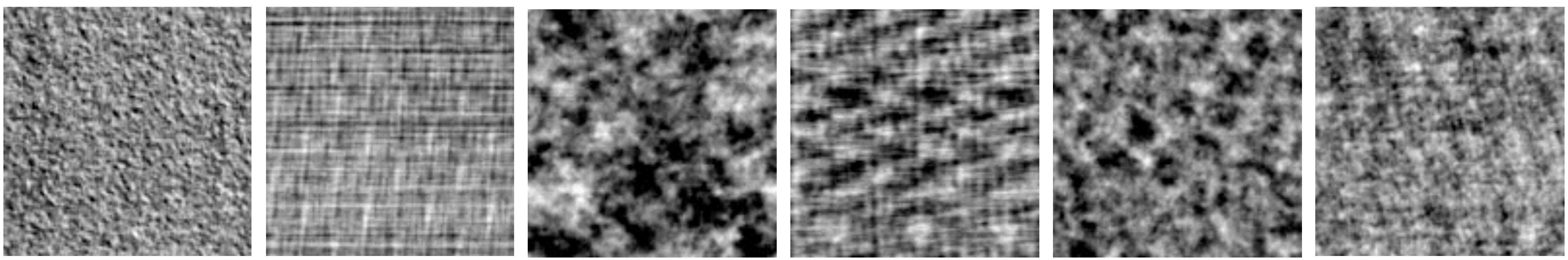
Ergodic Texture Reconstructions

Joan Bruna

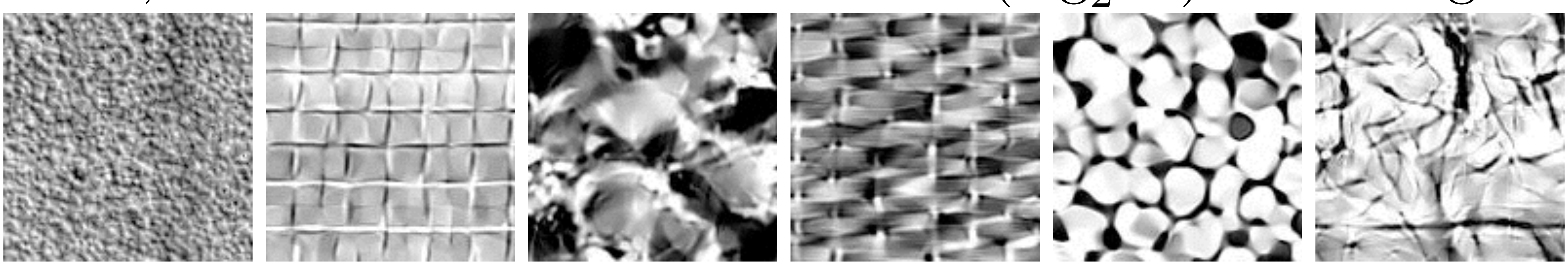
Original Textures



Gaussian stationary process: recovered from autocorrelation

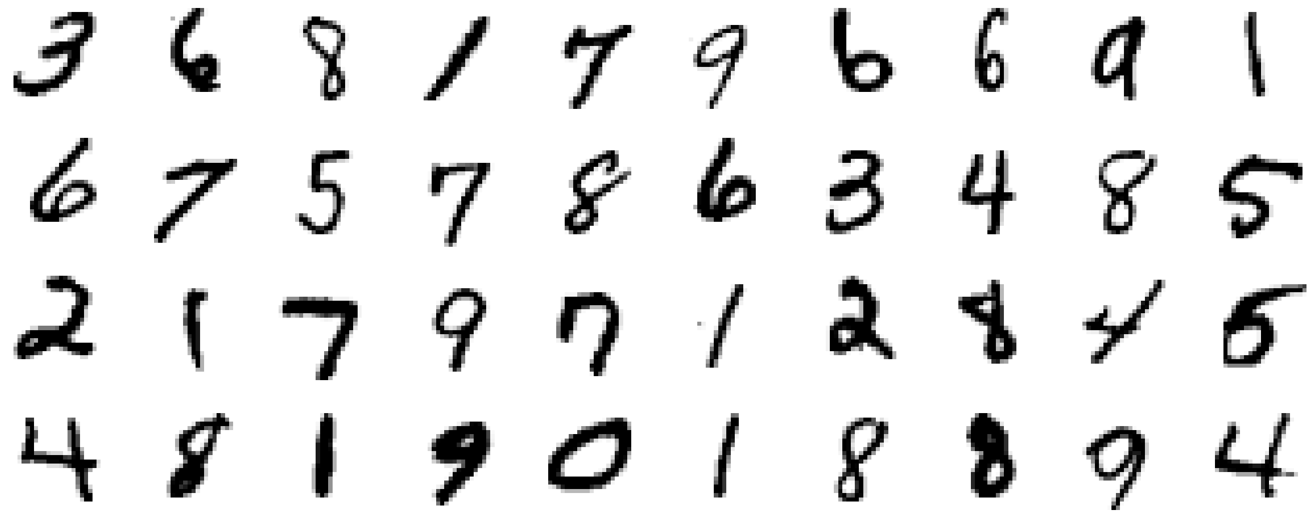


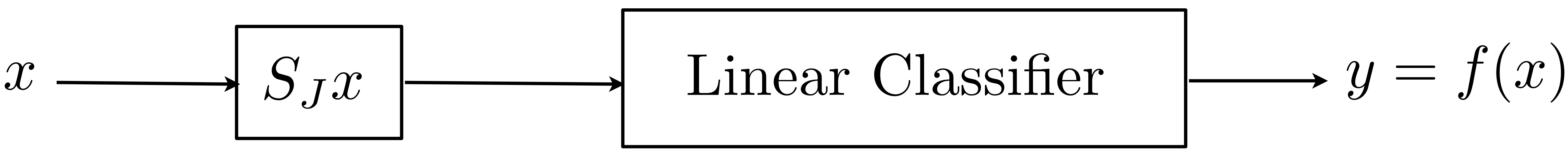
$m = 2$, $2^J = N$: reconstruction from $O(\log_2^2 N)$ scattering coeff.



Digit Classification: MNIST

Joan Bruna





Classification Errors

Training size	Conv. Net.	Scattering
60000	0.5%	**0.4**%

LeCun et. al.

- Know source of variability: translations, deformations.

Complex Image Classification

CalTech 101 data-basis:

Edouard Oyallon

$$x \longrightarrow \boxed{\begin{array}{c} S_J x \\ \text{Rigid Mvt.} \end{array}} \longrightarrow \boxed{\text{Linear Classif.}} \longrightarrow y$$

computes invariants

Classification Accuracy

Data Basis	Deep-Net	Scat.-2
CalTech-101	**85%**	80%
CIFAR-10	**90%**	80%

Trained on 10^6 images

Conclusion

- Quantum energy regression involves generic invariants to rigid movements, stability to deformations, multiscale interactions
- These properties require scale separations, hence wavelets.
- Multilayer wavelet scattering create large number of invariants
- Equivalent to deep networks with predefined wavelet filters
- Knowing physics provides the invariants: can avoid learning representations

Looking for a Post-Doc !