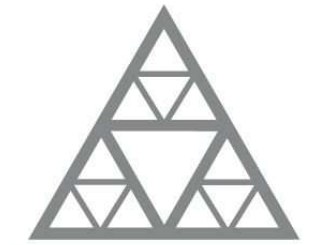
Séminaire de Calcul Scientifique du CERMICS



Solid Harmonic Wavelet Scattering for Predictions of Molecule Properties

Louis Thiry (ENS Paris)

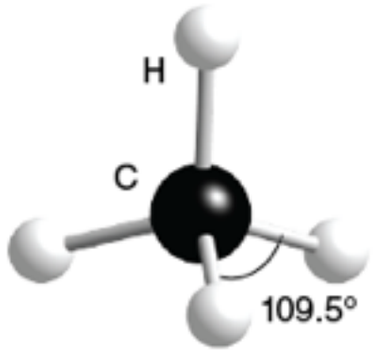
17 mai 2018

Solid harmonic scattering for quantum energy regressions

Louis THIRY, Data Team, ENS

Molecular property regression

Methane molecule



Numerical representation :

$$CH_4 \xrightarrow{\Phi} \begin{cases} -\text{positions + charges: } \{r_k, z_k\}_k \\ -\text{graph: } \quad \text{H}-\text{C}(\text{H})(\text{H})-\text{H} \\ \dots \end{cases}$$

Learn a regressor for the target quantity (energy):

$$\Phi(CH_4) \xrightarrow{\tilde{f}} E(CH_4)$$

Database

Source:

- Experimental measures
- Numerical simulations

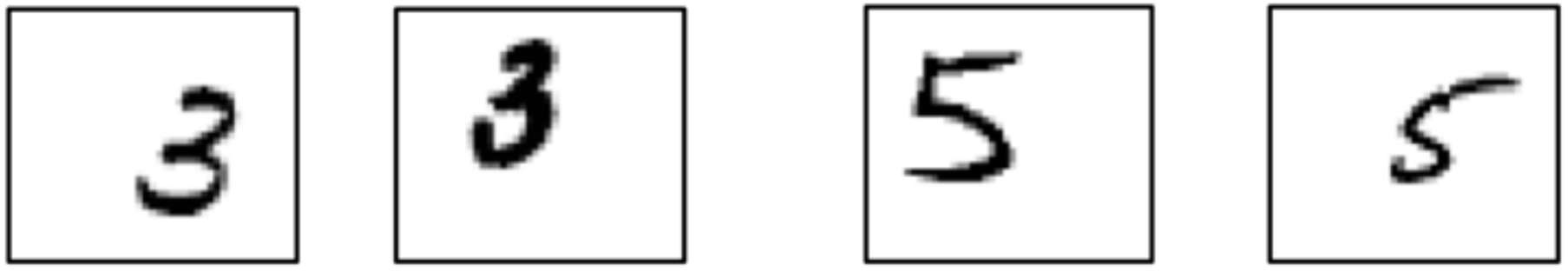
QM9 database

- **130.000** organic molecules
- up to 9 atoms of type C / O / N / Fl
- 13 properties : atomization energies, polarizability, dipole moment, thermal capacity …
- computed using density functional theory (B3LYP)
- Error in energies ~5 $kcal.mol^{-1}$

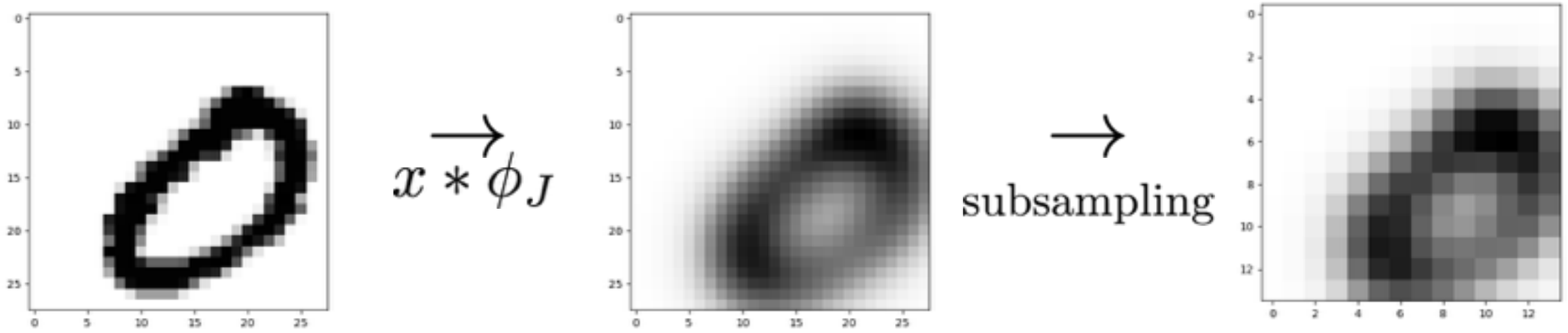
Digits classification

Handwritten digits classification

Invariance to translation, stability to deformation

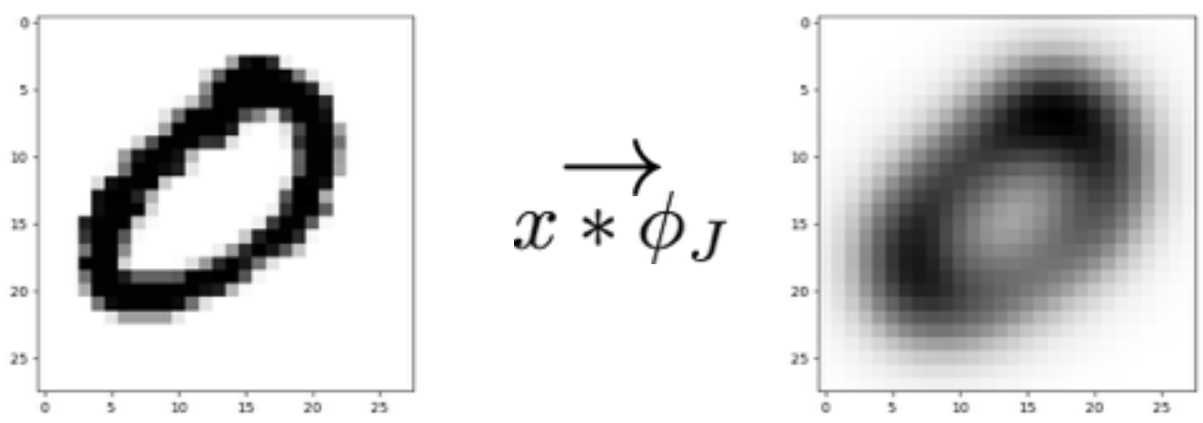


Create invariants and stability to translation



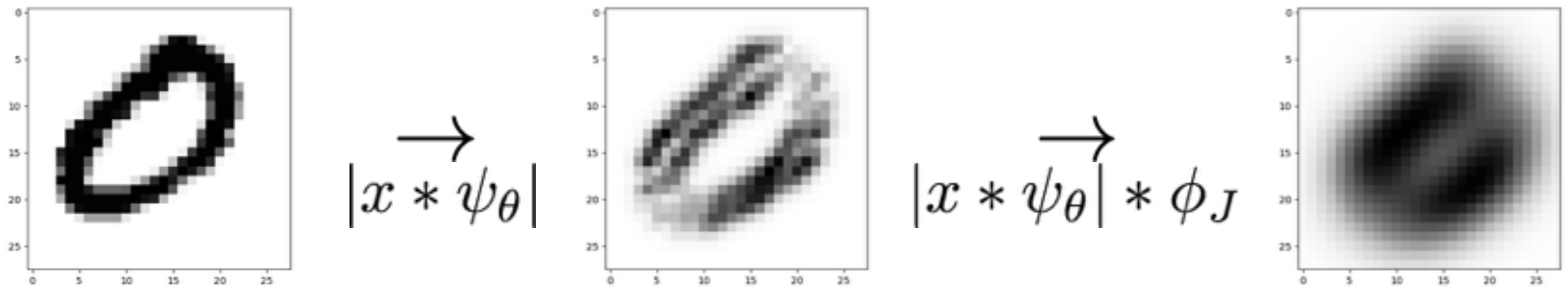
local averaging of pixel values with Gaussian kernel ϕ_J :

- locally translation invariant
- stable to additive and geometric deformations
- dimensionality reduction (subsampling)
- loss of high frequency information

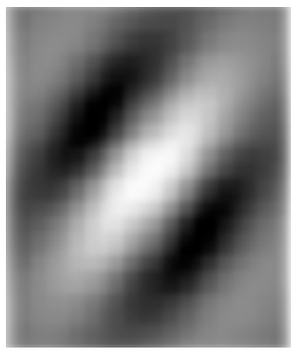


Recover lost information

- convolution and modulus with oriented wavelets recover high frequencies
- separate different orientations
- local averaging to create translation invariance
- subsampling to reduce dimension

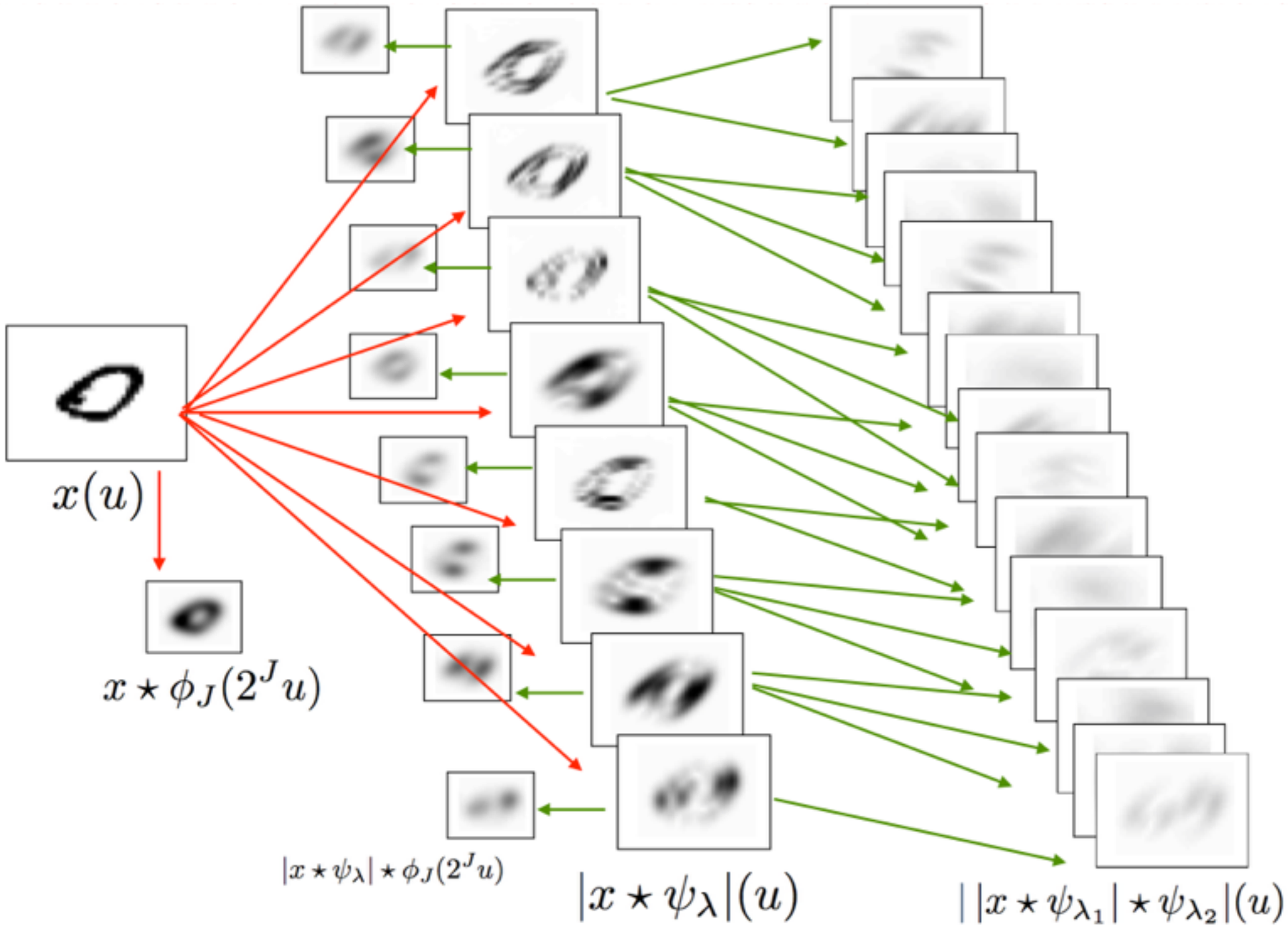


Oriented Gabor wavelet ψ_θ



Scattering transform

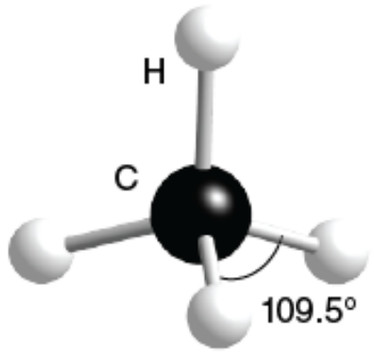
Mallat (2011), Mallat, Bruna (2012)



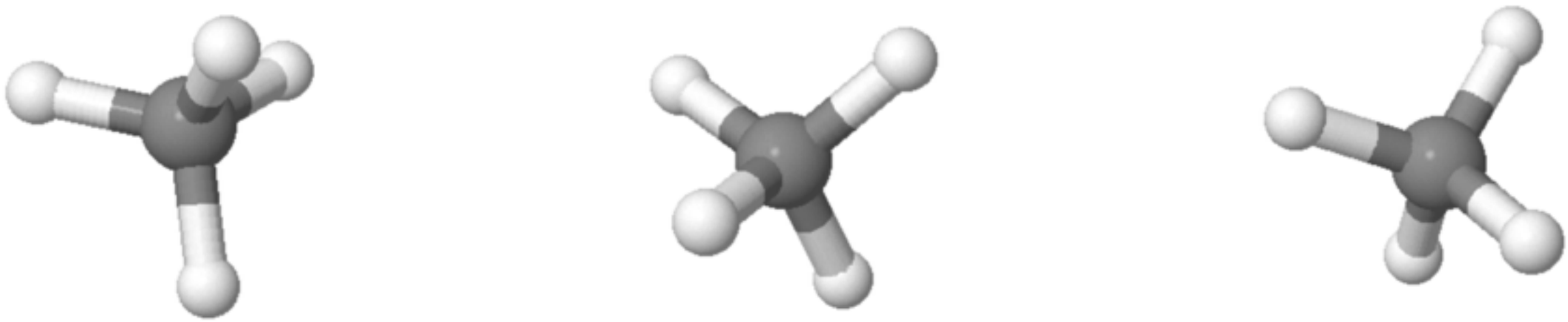
(from Joan Bruna UCB course)

Quantum energy regressions

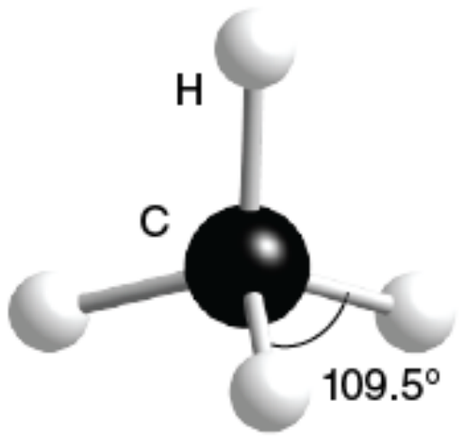
Methane molecule

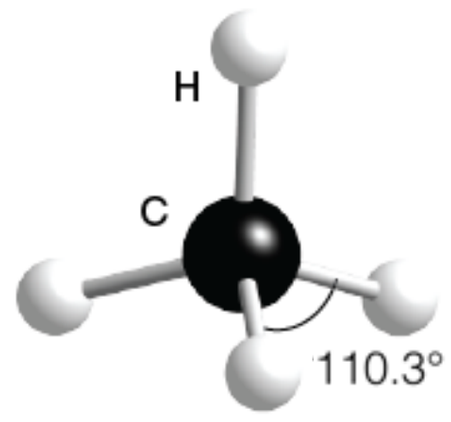


Invariance to **translation** and **rotation** :



Stable to **deformation** :



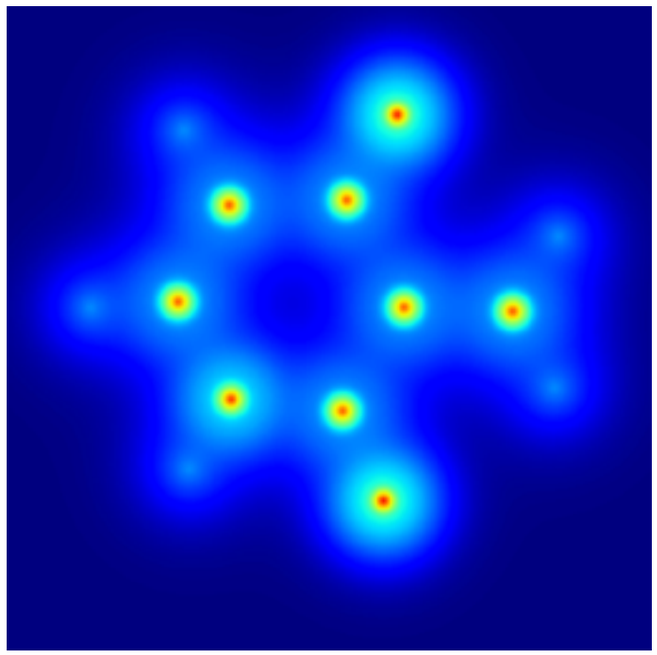


Quantum Energy Regression using Scattering Transforms

M. Hirn, N. Poilvert, S. Mallat (2016)

- Only planar molecule to have 2D images
- Create a fictious image of the molecule

$$\tilde{\rho}_x(u) = \sum_{k=1}^{K} \rho_{\text{at}}^{a(k)}(u - p_k)$$

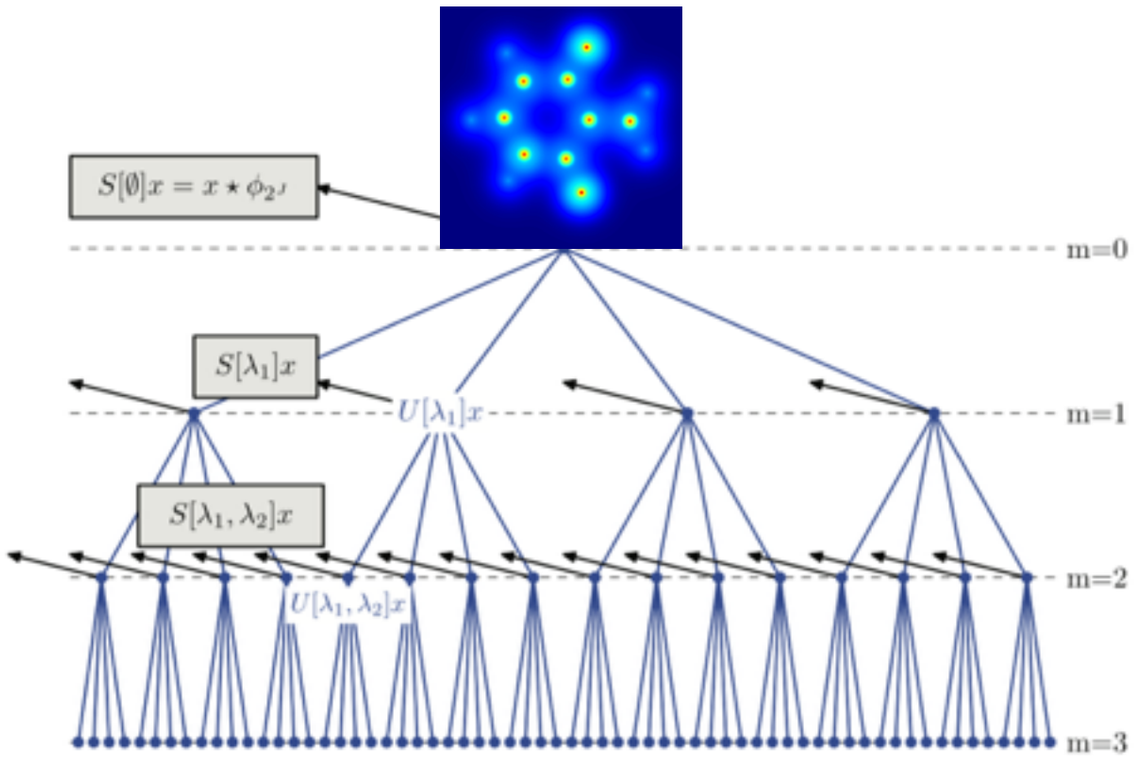


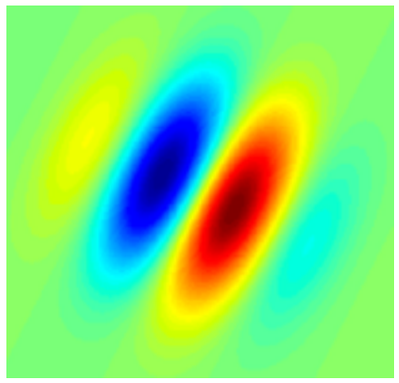
- Apply the standard scattering pipeline

Quantum Energy Regression using Scattering Transforms

M. Hirn, N. Poilvert, S. Mallat (2016)

- Compute scattering coefficients of the image





Gabor filters

- Perform linear regression on scattering coefficients
- Database QM2D : energies of 4357 molecules (DFT)
- MAE : 1.8 kcal/mol, RMSE : 2.7 kcal/mol

Solid harmonic wavelet scattering for predictions of molecule properties

G. Exarchakis, M. Eickenberg, M. Hirn, S. Mallat, L. Thiry (2018)

<u>2D to 3D</u>

- most molecules are 3D

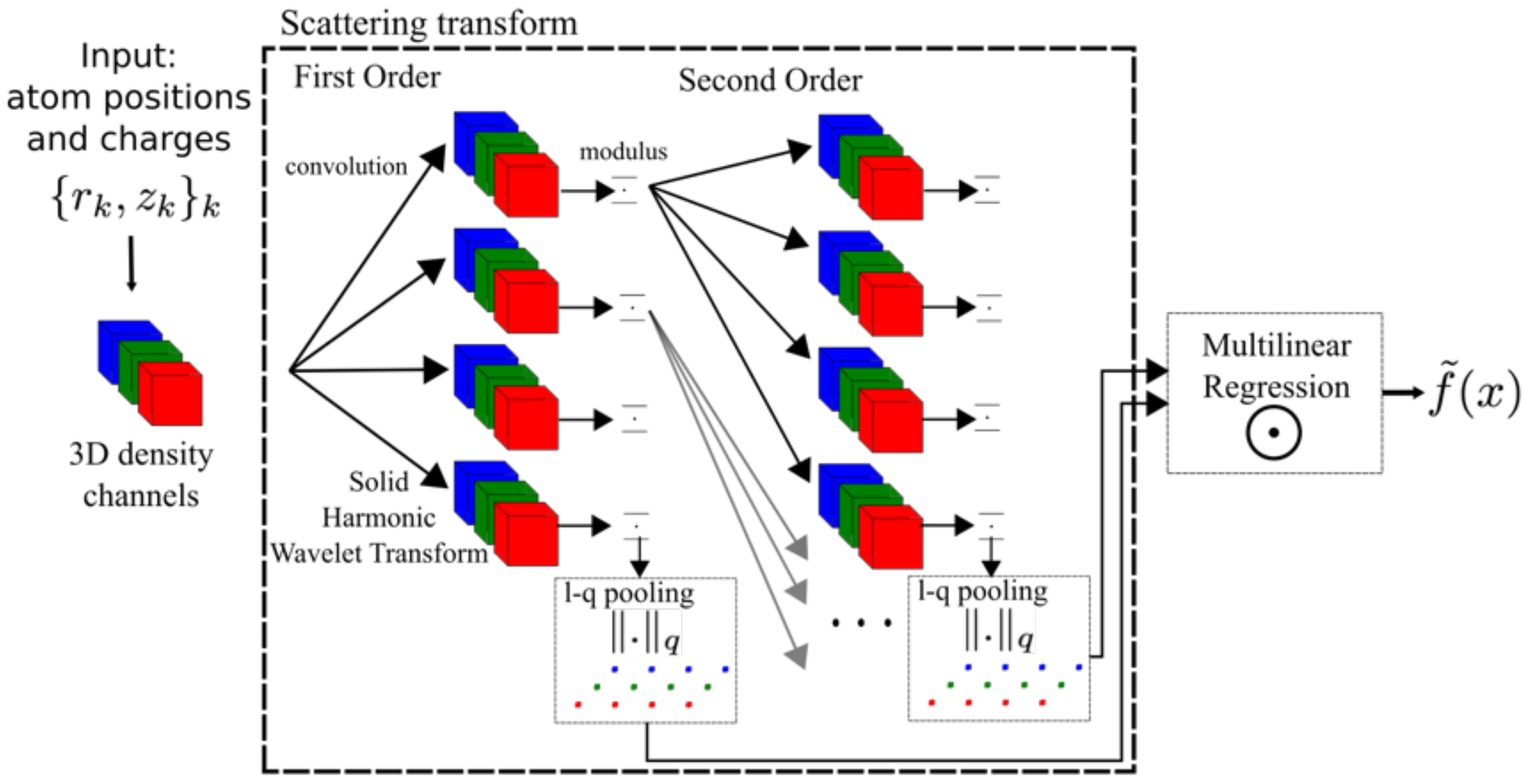
- implementation issues

- Gabor filters do not seem to be relevant

- inspire from chemistry

Solid harmonic wavelet scattering for predictions of molecule properties

G. Exarchakis, M. Eickenberg, M. Hirn, S. Mallat, L. Thiry (2018)

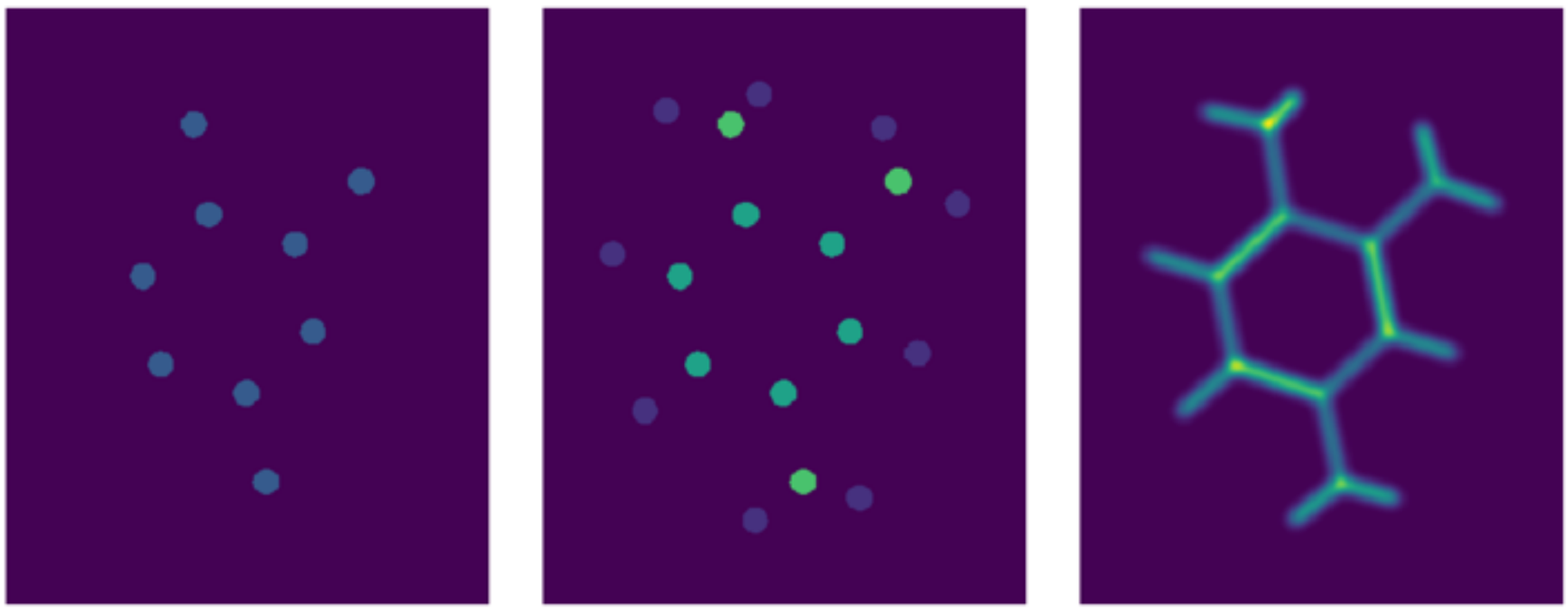


3D densities

- Replace isolated atomic densities by Gaussians for aliasing reasons :

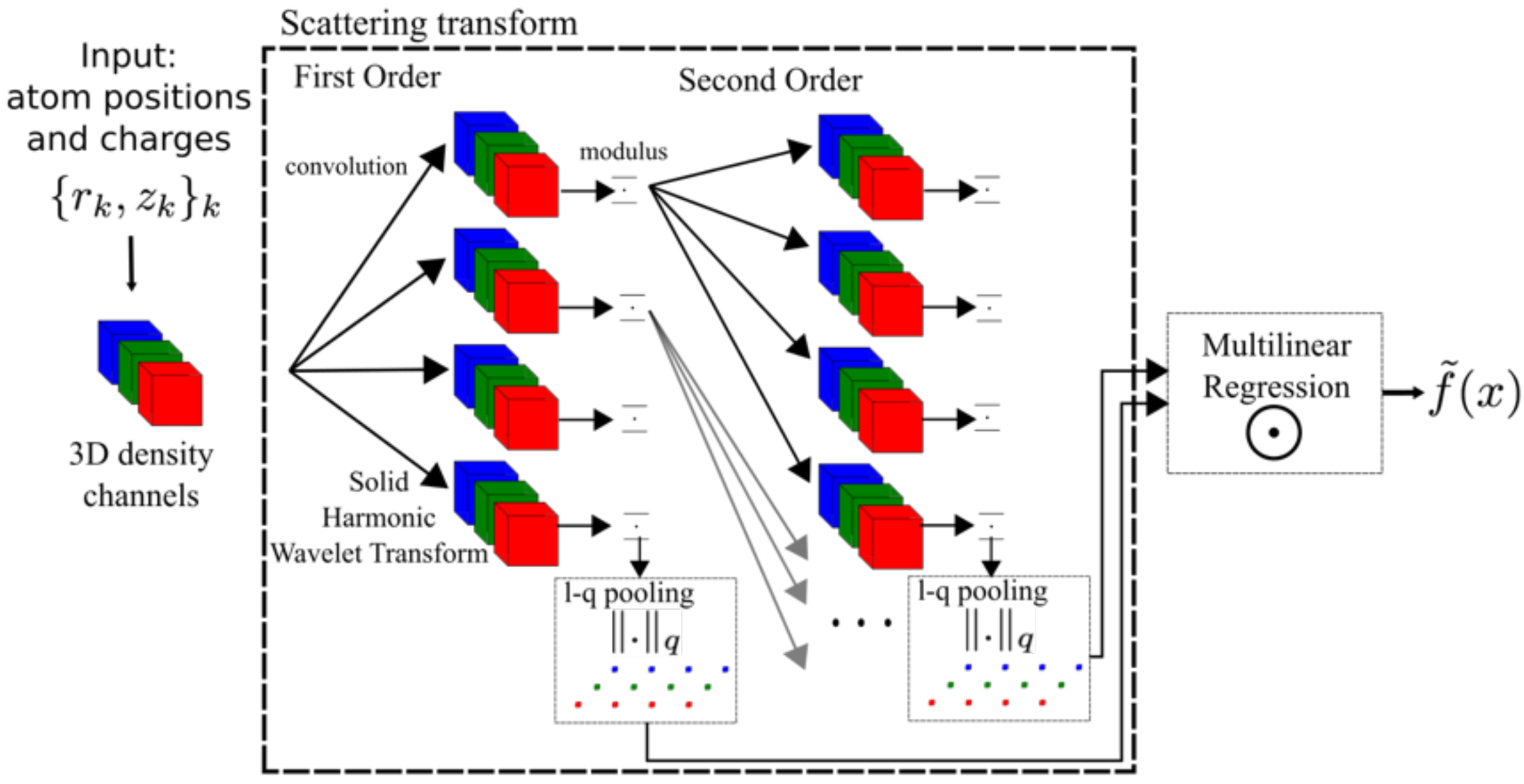
$$\rho(u) = \sum_n g(u - r_n)$$

- Include bond information



Solid harmonic wavelet scattering for predictions of molecule properties

G. Exarchakis, M. Eickenberg, M. Hirn, S. Mallat, L. Thiry (2018)



3D Solid harmonic wavelets

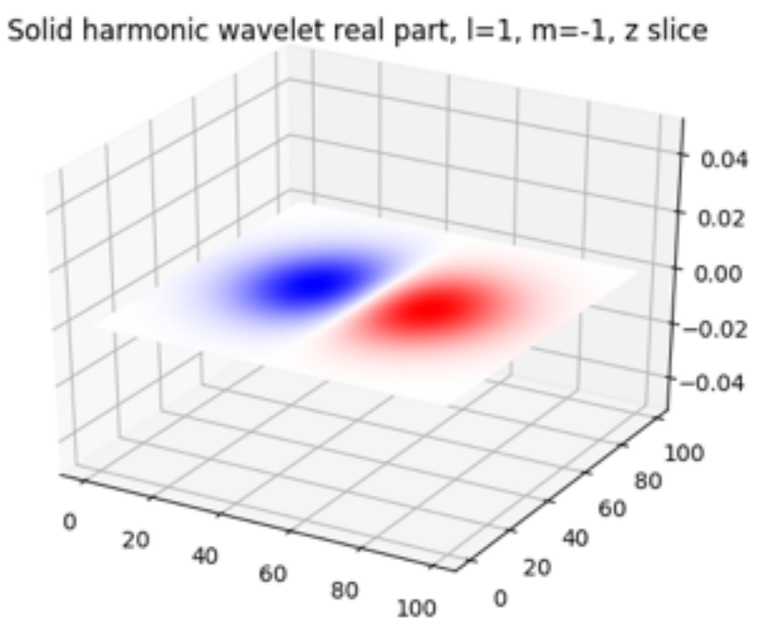
$$\psi_{l,m}(r,\theta,\phi) = K_l \underbrace{e^{-r^2/2} r^l}_{\substack{\text{polynomial} \\ \text{gaussian}}} \underbrace{Y_l^m(\theta,\phi)}_{\substack{\text{spherical} \\ \text{harmonic}}}$$

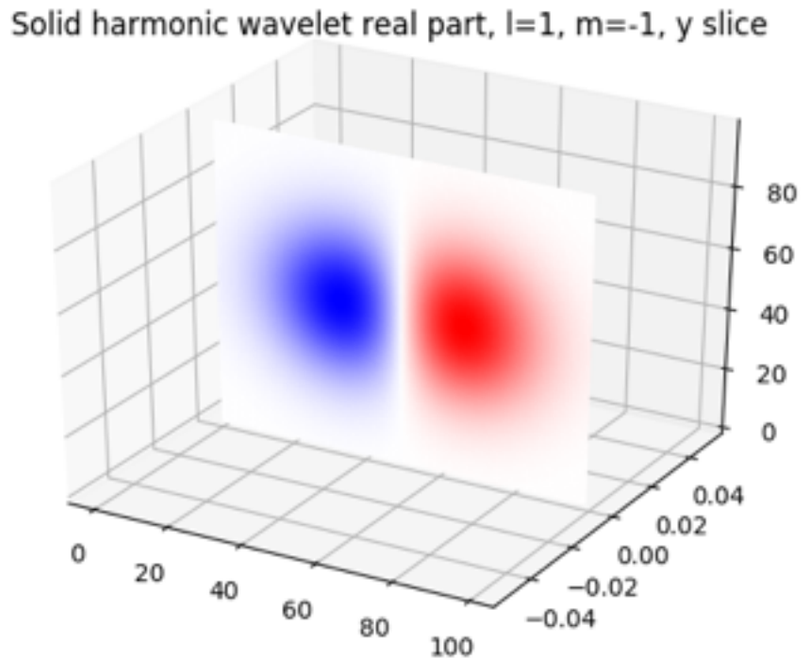
$$\psi_0(r,\theta,\phi) = K e^{-r^2/2}, \text{ Gaussian}$$

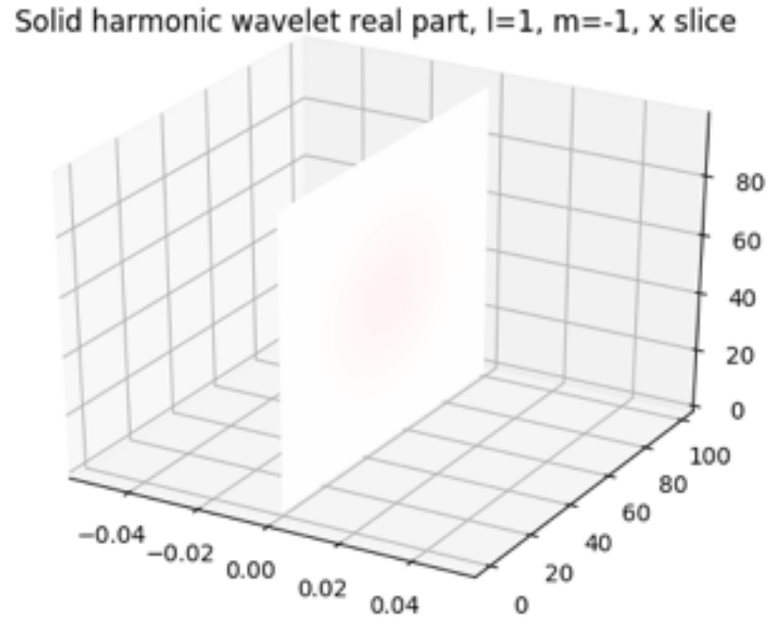
Spherical harmonics:

- $\Delta_{\mathbb{S}^2} Y_l^m = -l(l+1) Y_l^m$
- $\langle Y_l^m, Y_{l'}^{m'} \rangle_{\mathbb{S}^2} = \int_{\mathbb{S}^2} Y_l^m(\Omega) \overline{Y_{l'}^{m'}(\Omega)} d\Omega = \delta_{l,l',m,m'}$
- $(Y_l^m)_{m=-l \ldots l}$ is stable by rotation

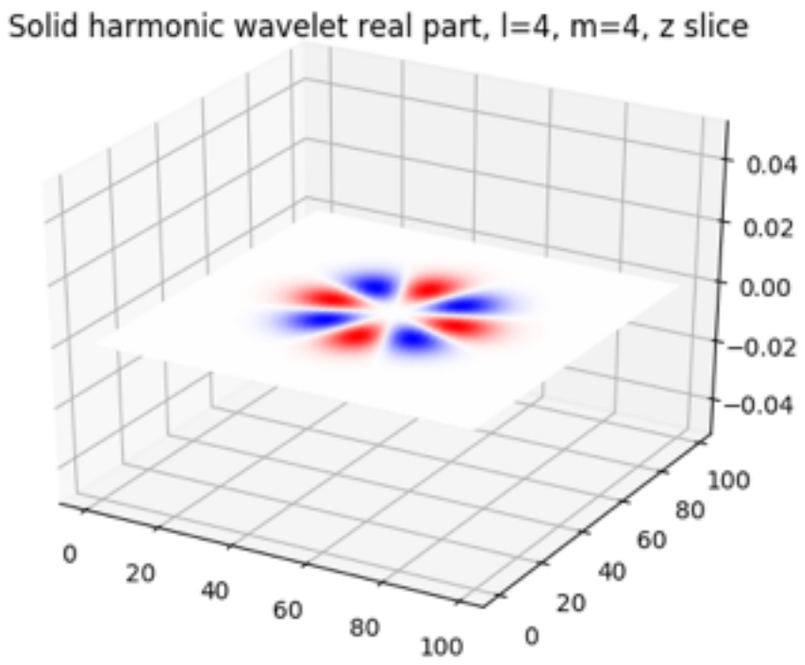
3D Solid harmonic wavelets

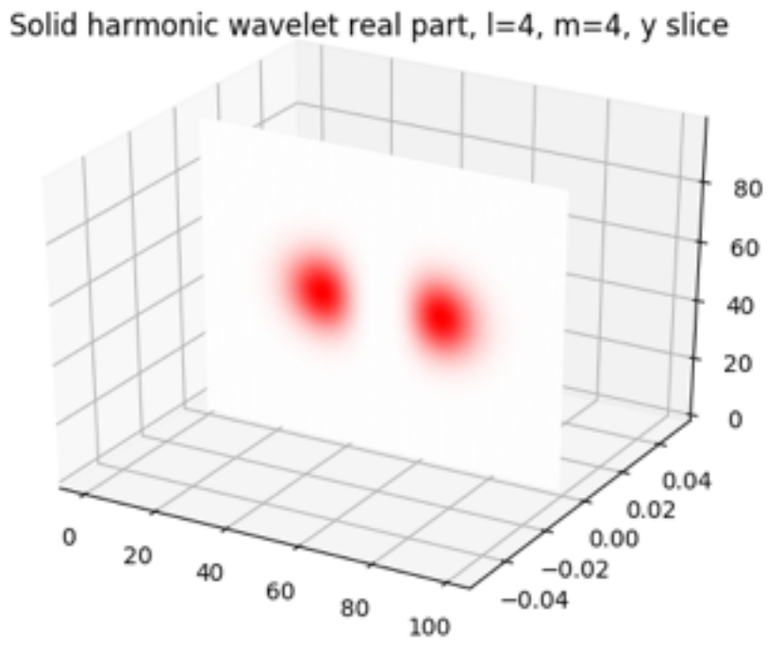


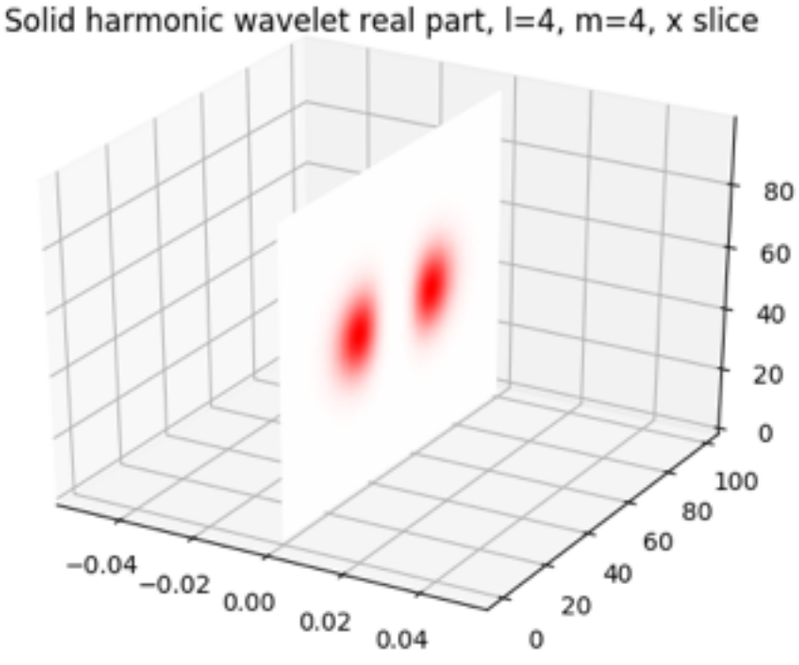




$$l = 1, m = -1$$







$$l = 4, m = 4$$

Rotation-translation invariance

Special « convolution and modulus » :

$$|\rho * \psi_l| = \left(\sum_{m=-l}^{l} |\rho * \psi_{l,m}|^2 \right)^{1/2}$$

Translation and rotation covariant :

$$|\tau.\rho * \psi_l| = \tau.|\rho * \psi_l|$$

$$|R.\rho * \psi_l| = R.|\rho * \psi_l|$$

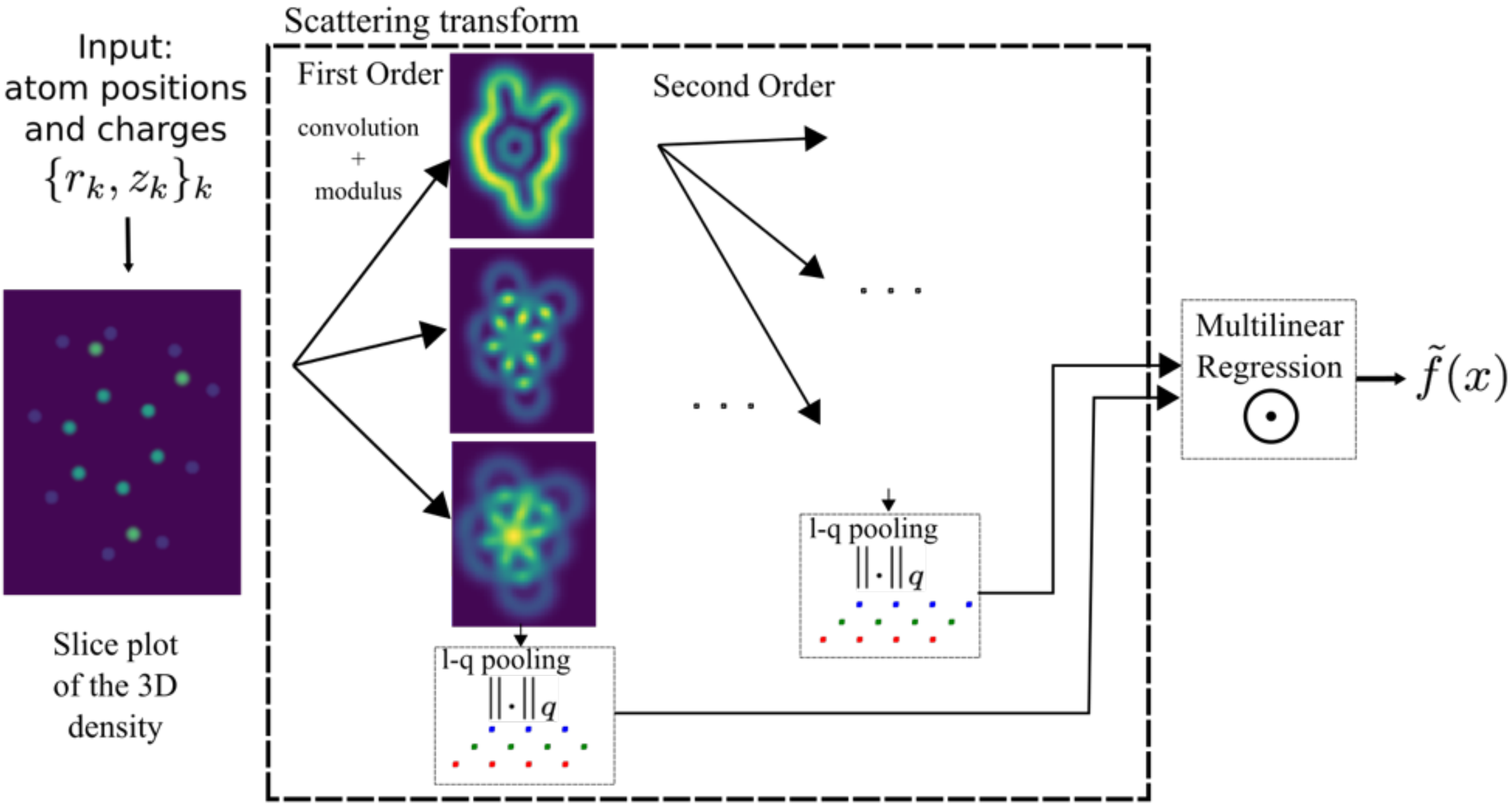
l-q pooling :

$$c_{l,q} = \int_{\mathbb{R}^3} |\rho * \psi_l|^q$$

coefficient c in invariant :

$$\int_{\mathbb{R}^3} |R.\rho * \psi_l|^q = \int_{\mathbb{R}^3} R.|\rho * \psi_l|^q = \int_{\mathbb{R}^3} |\rho * \psi_l|^q$$

Regression with scattering transform



Multi-linear regression

Linear regression

$$E = \sum_{k=1}^{K} \alpha_k c_k + b \ : \ K \text{ coefficients to learn}$$

Bilinear regression

$$E = \sum_{k,l=1}^{K} \alpha_{k,l} c_k c_l + b \ : \ K^2 \text{ coefficients to learn}$$

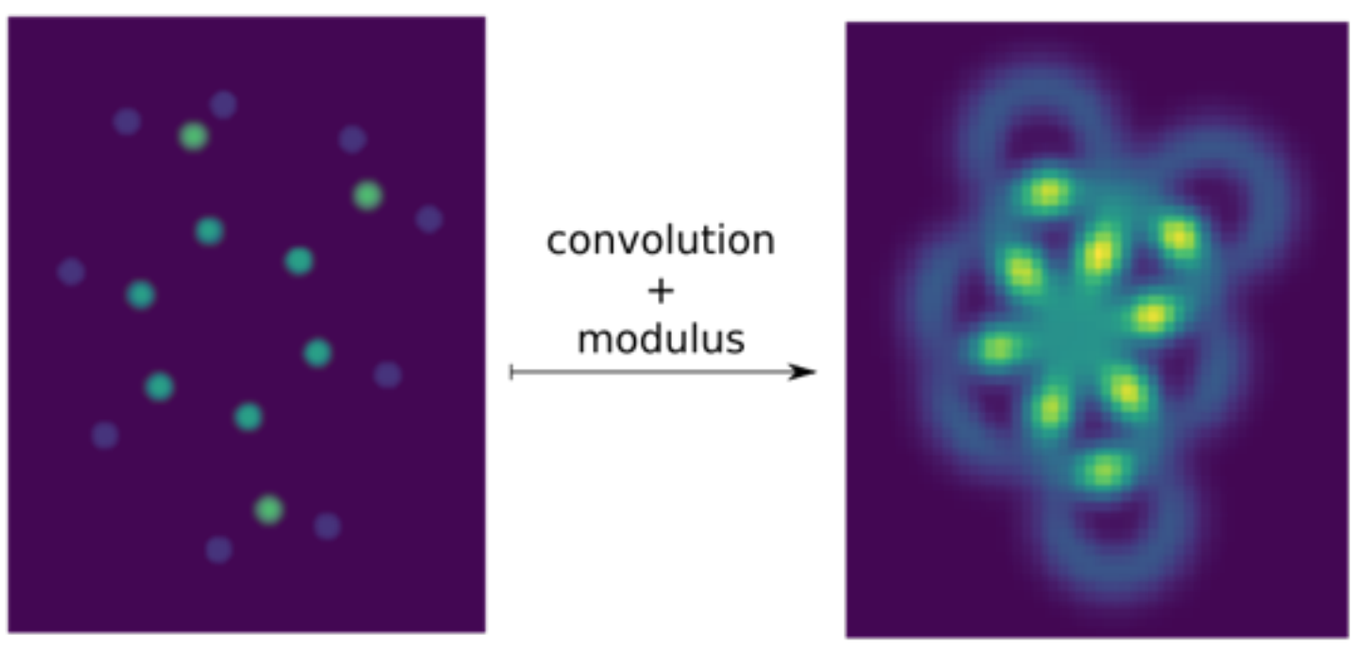
Factorized bilinear regression

$$E = \left(\sum_k \alpha_k c_k \right) \cdot \left(\sum_k \beta_k c_k \right) + b \ : \ 2K \text{ coefficients to learn}$$

Interpretation

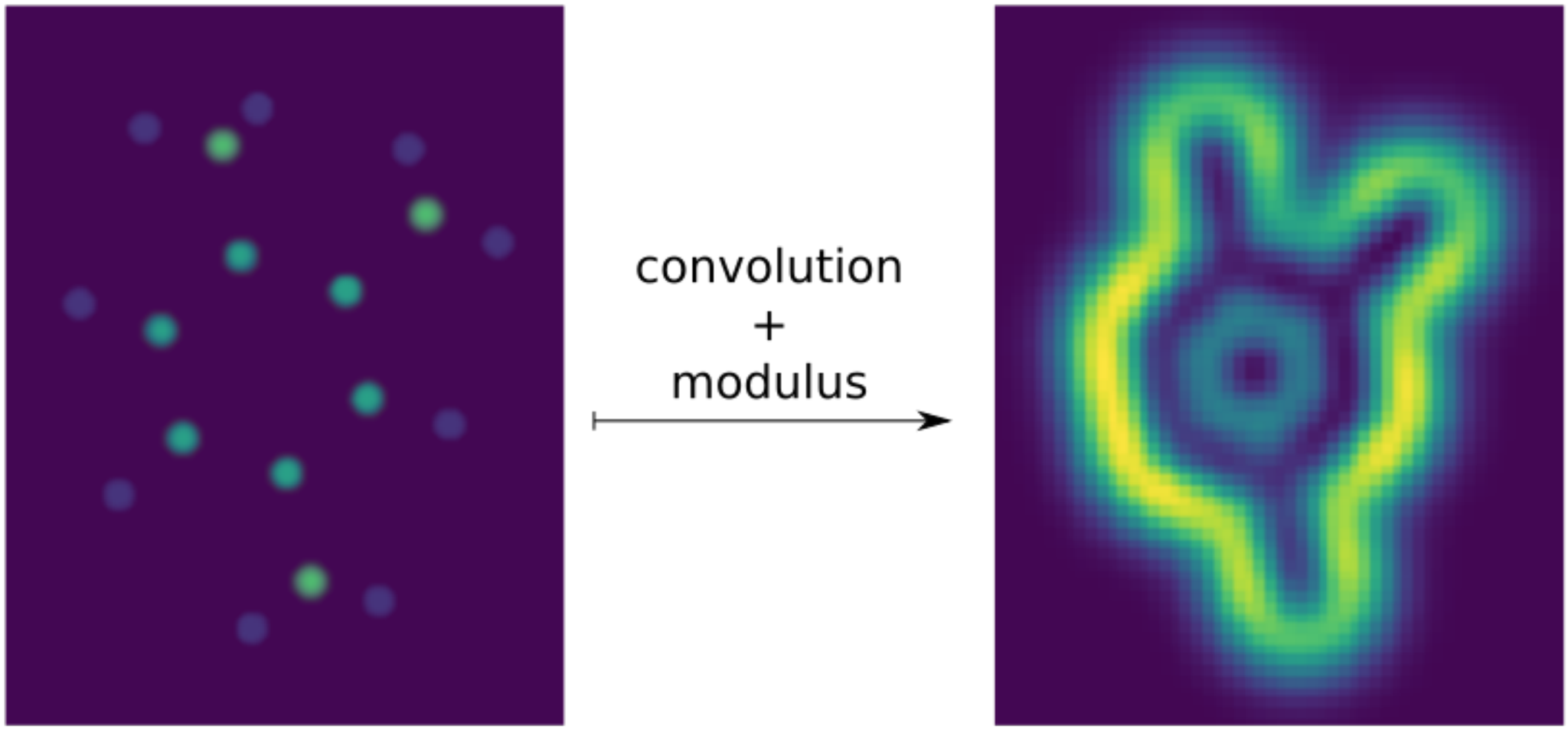
$$g * \psi_{l,m}(u) = k\ \psi_{l,m}(\frac{u}{s})$$

$$\rho(u) = \sum_n g(u - r_n) \xrightarrow[\text{modulus}]{\text{convolution}} |\rho * \psi_l|(u) = \left(\sum_{m=-l}^{l} \left| \sum_n \psi_{l,m}(u - r_n) \right|^2 \right)^{1/2}$$

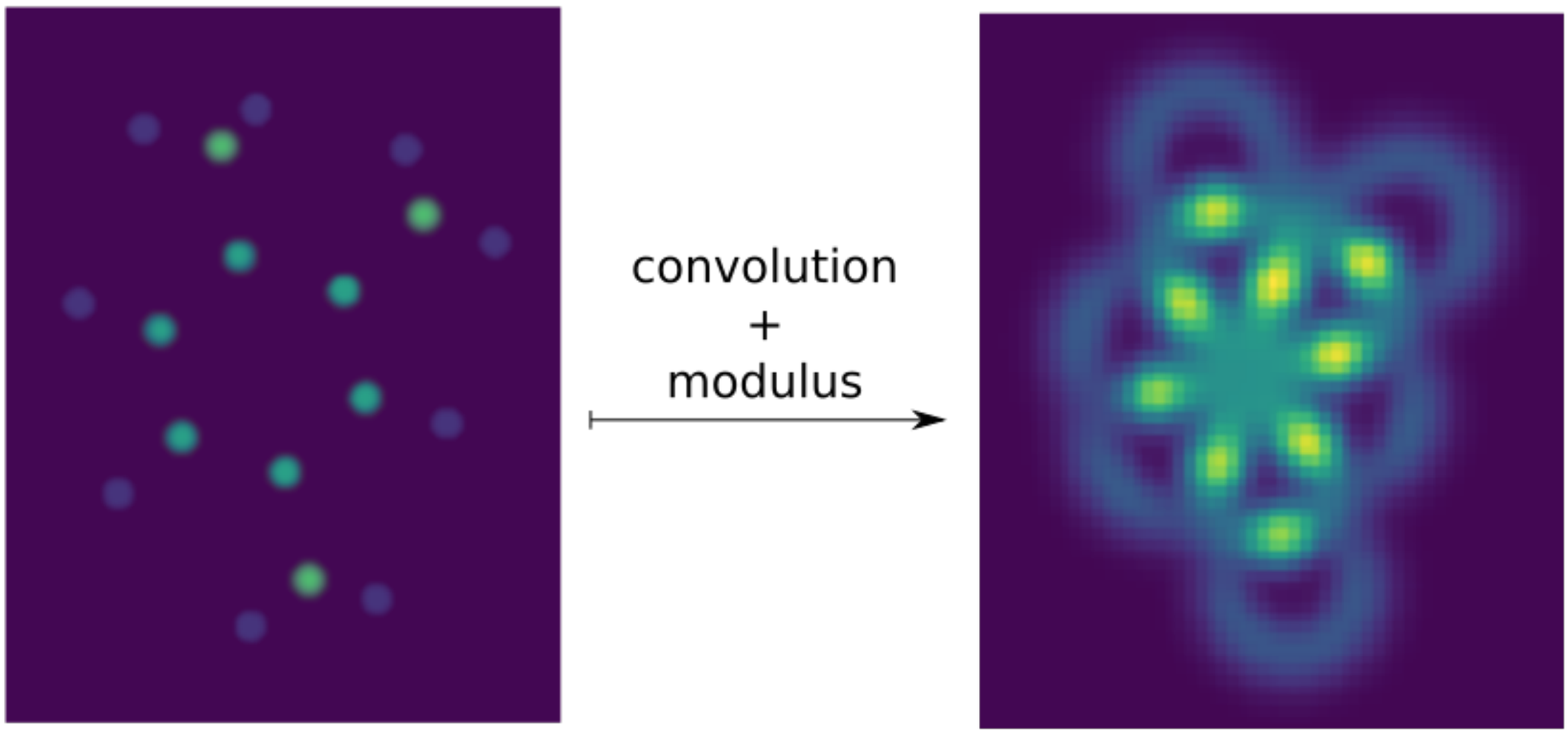


Interference patterns of solid harmonic wavelets

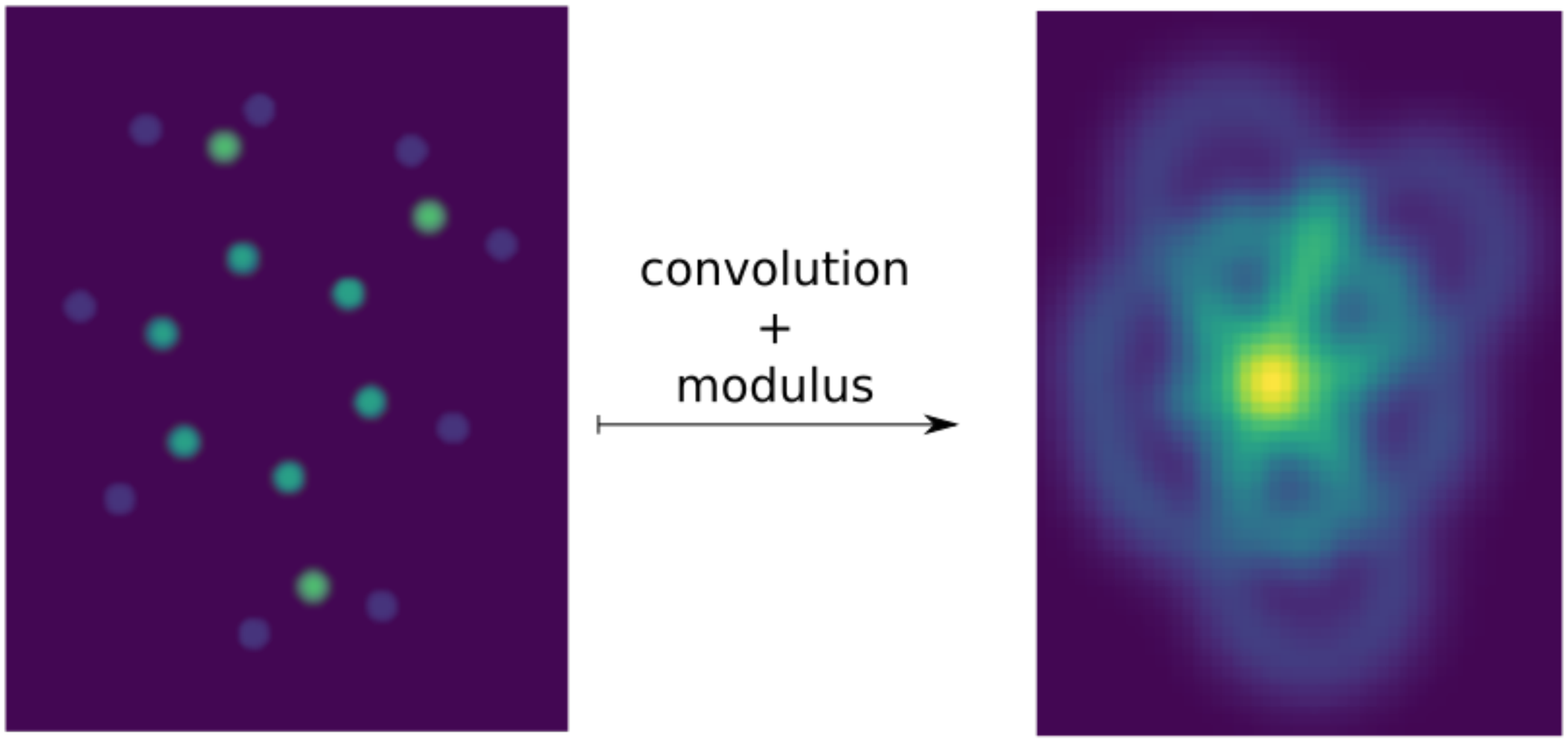
Interference patterns



Interference patterns



Interference patterns



Results

Property	Scattering linear reg.	Scattering tri-linear reg.	State of the art	DFT precision
Energy at 0 K ($kcal.mol^{-1}$)	1.89	0.5	0.45	2.3
Thermal capacity ($cal.mol^{-1}.K^{-1}$)	0.10	0.049	0.04	0.34

Next step: forces regression

Direct regressions take away

- « classical » regression problem
- largely inspired by the image classification method
- based on rotation/translation invariance of the properties of the molecules
- regression error is 1 OM lower than database errors —> need more precise database
- do not really generalizes to bigger molecules