# Learning Music, Images and Physics with Deep Neural Networks

*Joan Bruna, Matthew Hirn, Stéphane Mallat*
*Vincent Lostanlen, Edouard Oyallon, Nicolas Poilvert,*
*Laurent Sifre, Irène Waldspurger*

**École Normale Supérieure**
www.di.ens.fr/data

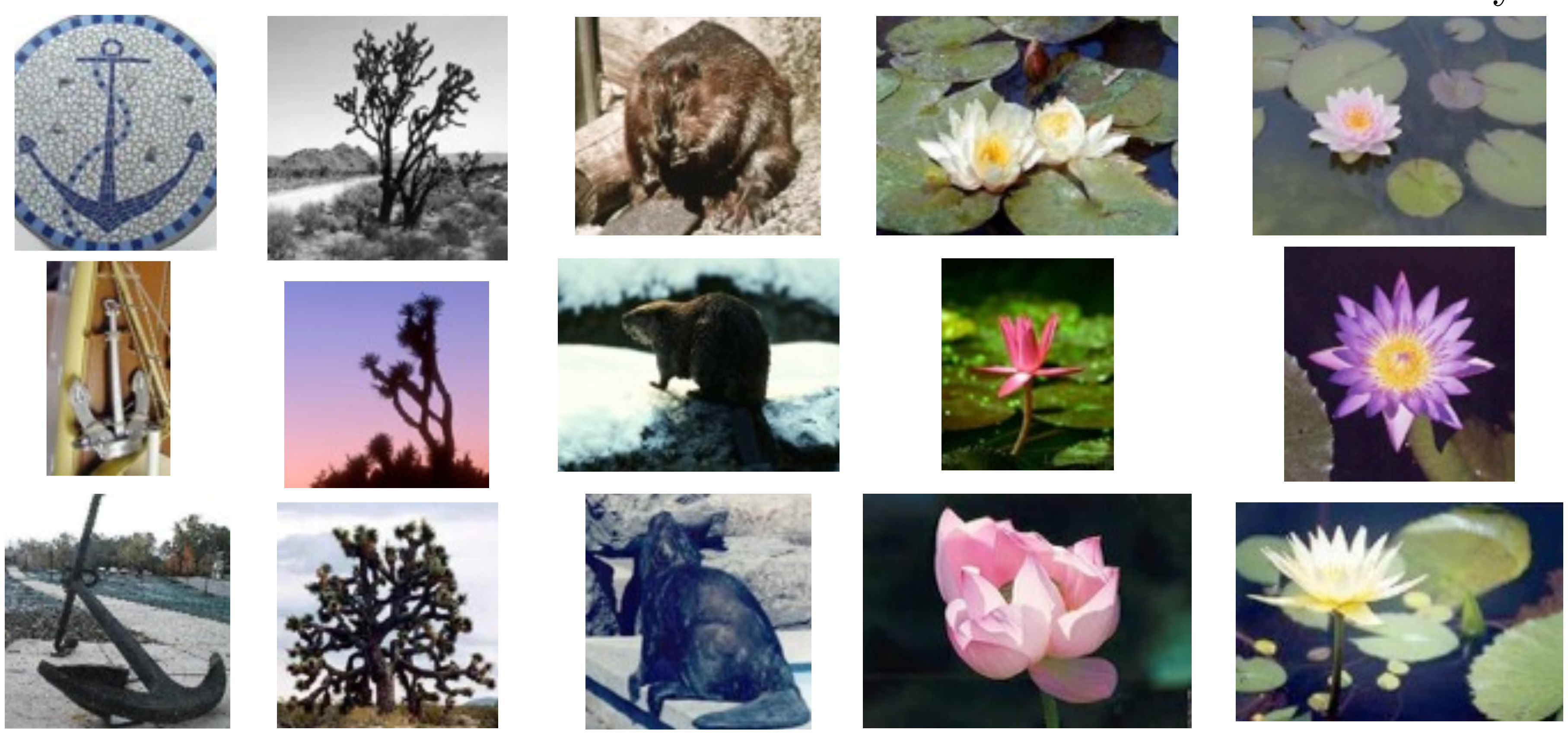
# High Dimensional Learning

- High-dimensional $x = (x(1), ..., x(d)) \in \mathbb{R}^d$:

- **Classification:** estimate a class label $f(x)$ given $n$ sample values $\{x_i \, , \, y_i = f(x_i)\}_{i \leq n}$

Image Classification $d = 10^6$

Anchor Joshua Tree Beaver Lotus Water Lily



Huge variability inside classes

# High Dimensional Learning

- High-dimensional $x = (x(1), ..., x(d)) \in \mathbb{R}^d$:

- **Classification:** estimate a class label $f(x)$ given $n$ sample values $\{x_i \, , \, y_i = f(x_i)\}_{i \leq n}$

Audio: instrument recognition

Huge variability inside classes

# High Dimensional Learning

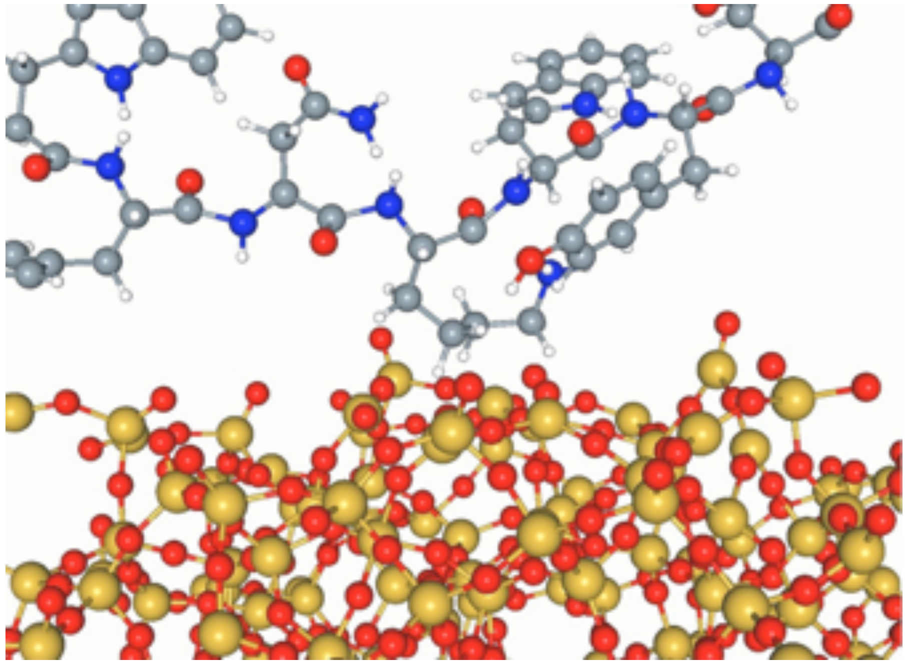
- High-dimensional $x = (x(1), ..., x(d)) \in \mathbb{R}^d$:

- **Regression:** approximate a *functional* $f(x)$
  given $n$ sample values $\{x_i \, , \, y_i = f(x_i) \in \mathbb{R}\}_{i \leq n}$

Physics: energy $f(x)$ of a state vector $x$

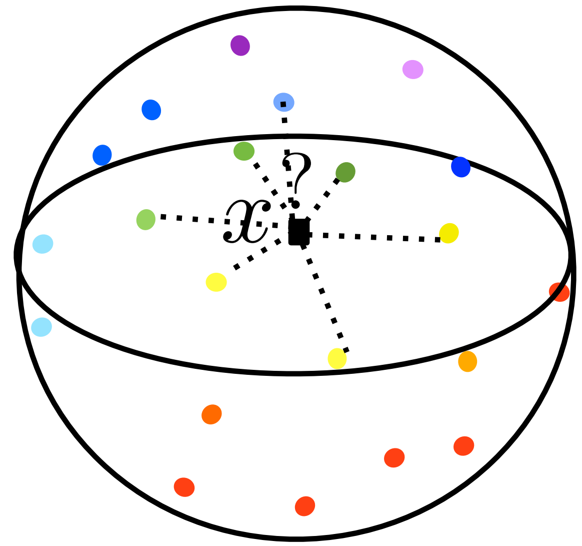
Astronomy

Quantum Chemistry



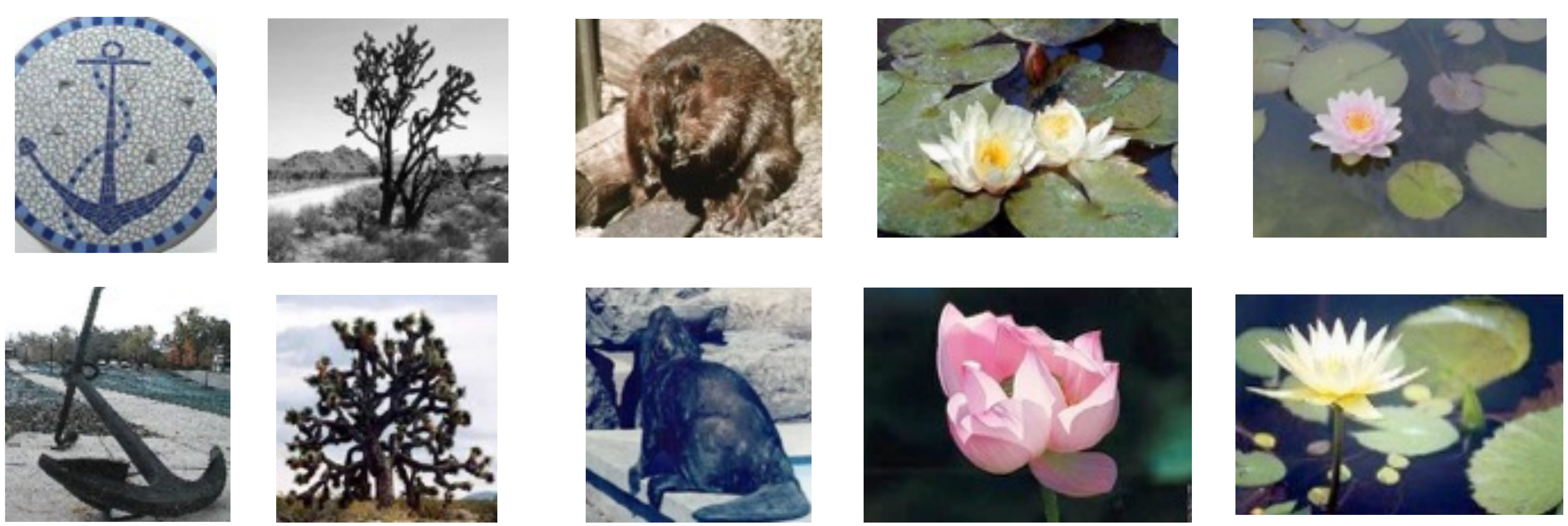
# Curse of Dimensionality

- $f(x)$ can be approximated from examples $\{x_i \,,\, f(x_i)\}_i$ by local interpolation if $f$ is regular and there are close examples:




- Need $\epsilon^{-d}$ points to cover $[0,1]^d$ at a Euclidean distance $\epsilon$

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



Huge variability inside classes

# Learning by Euclidean Embedding

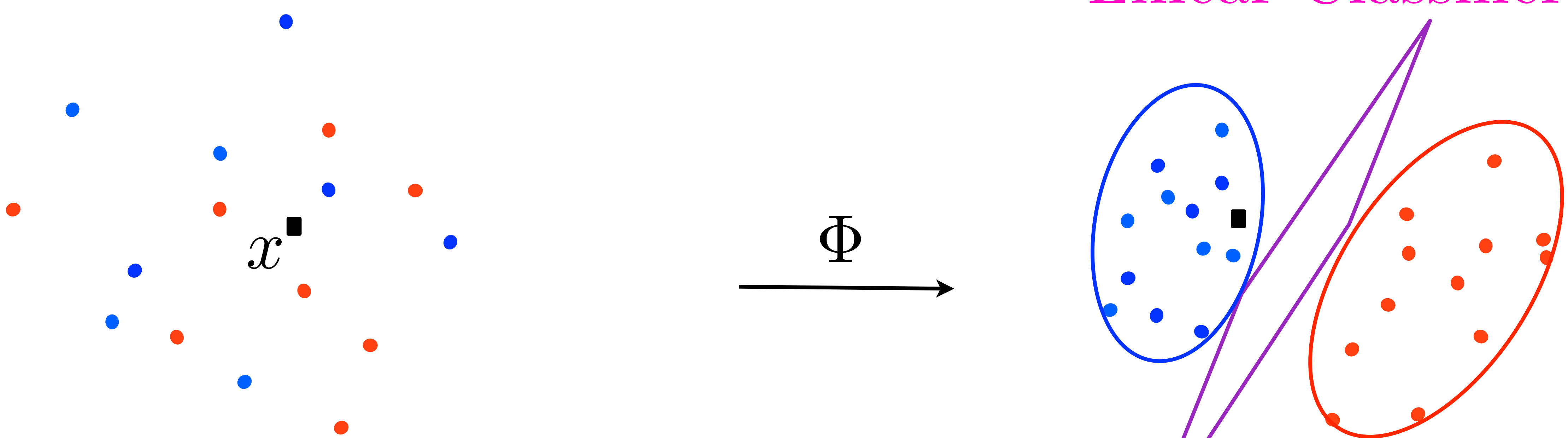
Data: $x \in \mathbb{R}^d$

$\|x - x'\|$: non-informative

Representation

$\Phi x \in \mathcal{H}$

Linear Classifier




Gaussian & Separated

"Similarity" metric: $\Delta(x, x')$ $\longleftrightarrow$ $\|\Phi x - \Phi x'\|$

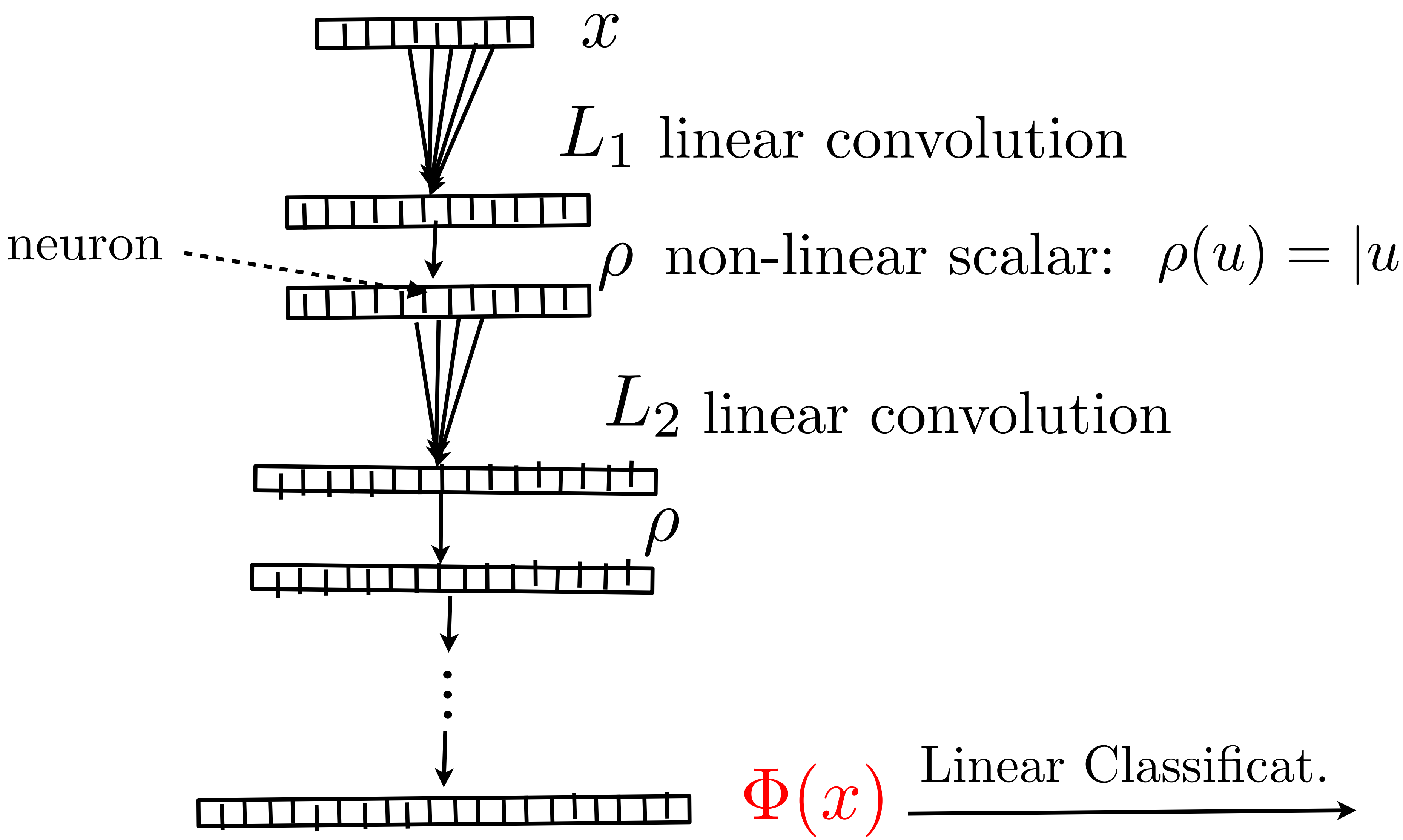
Equivalent Euclidean metric:

$$C_1 \|\Phi x - \Phi x'\| \leq \Delta(x, x') \leq C_2 \|\Phi x - \Phi x'\|$$

How to define $\Phi$ ?

# Deep Convolution Neworks

- The revival of an old (1950) idea: *Y. LeCun, G. Hinton*




Optimize the $L_k$ with support constraints: over $10^9$ parameters

Exceptional results for *images*, *speech*, *bio-data* classification.

Products by FaceBook, IBM, Google, Microsoft, Yahoo...

Why does it work so well ?

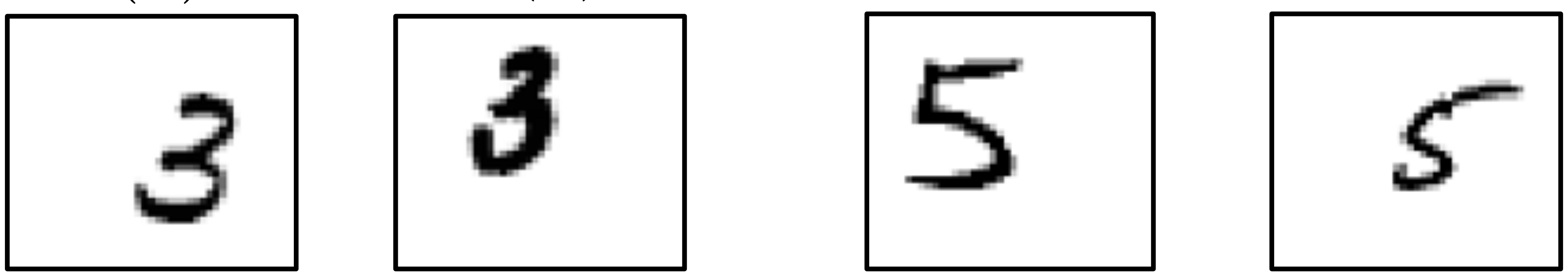
# Overview

- Deep multiscale networks: invariant and stable metrics on groups
- Image classification
- Models of audio and image textures: information theory
- Learning physics: quantum chemistry energy regression

# Image Metrics

- Low-dimensional "geometric shapes"

$x(u)$ $x'(u)$



Deformation metric: (classic mechanics) *Grenander*

Diffeomorphism action: $D_\tau x(u) = x(u - \tau(u))$

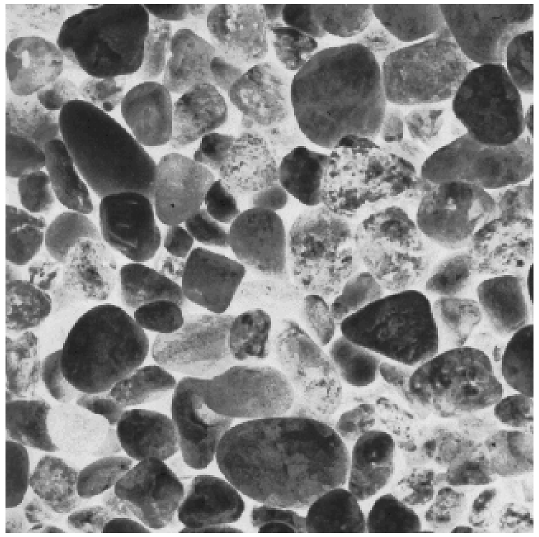
$$\Delta(x, x') \sim \min_{\tau} \|D_\tau x - x'\| + \|\nabla \tau\|_\infty \, \|x\|$$

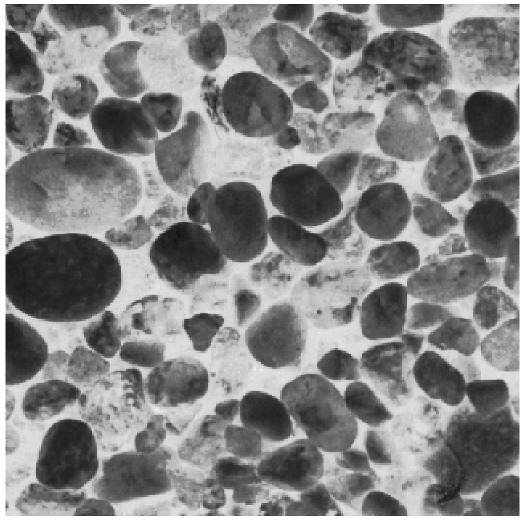
Invariant to translations

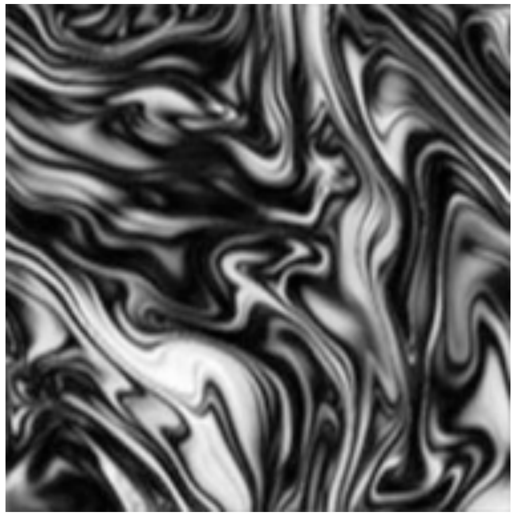
diffeomorphism amplitude

# Image Metrics

- High dimensional textures: $X(u)$ ergodic stationary processes







2D Turbulence

Highly non-Gaussian processes

- A Euclidean metric is a Maximum Likelihood on Gaussian models.

- Can we find $\Phi$ so that $\Phi(X)$ is nearly Gaussian, without loosing information ?

# Euclidean Metric Embedding

- Stability to additive perturbations:

$$\|\Phi x - \Phi x'\| \leq C \, \|x - x'\|$$

- Invariance to translations:

$$x_c(u) = x(u-c) \;\Rightarrow\; \Phi(x_c) = \Phi(x)$$

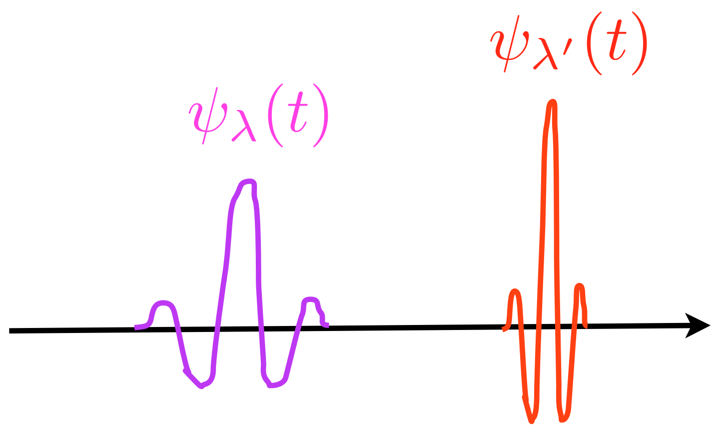
- Stability to deformations:

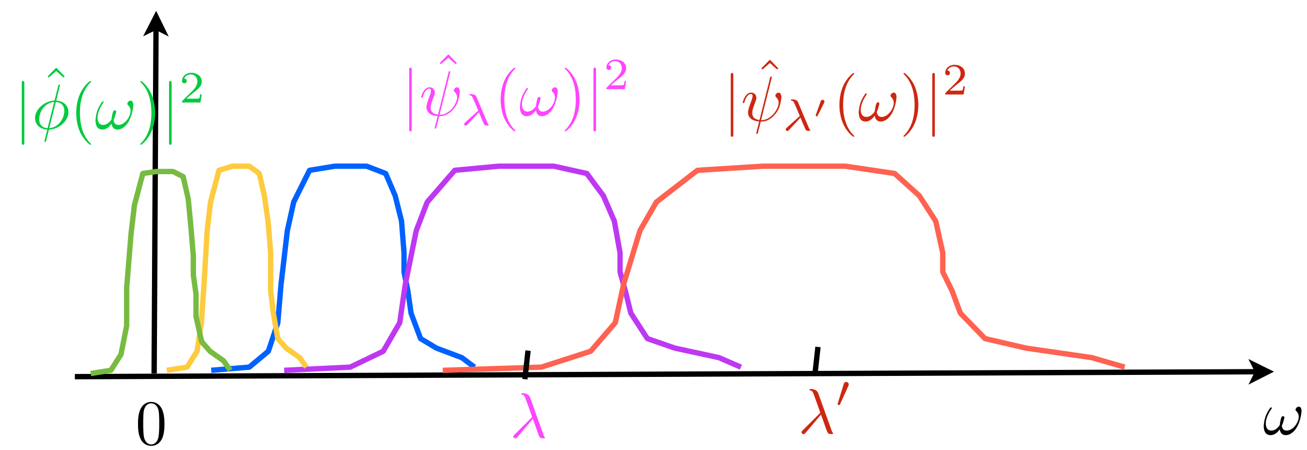
$$x_\tau(u) = x(u-\tau(u)) \;\Rightarrow\; \|\Phi x - \Phi x_\tau\| \leq C \, \|\nabla \tau\|_\infty \, \|x\|$$

Failure of Fourier and classic invariants

# Wavelet Transform

- Dilated wavelets: $\psi_\lambda(t) = 2^{-j/Q}\, \psi(2^{-j/Q} t)$ with $\lambda = 2^{-j/Q}$




Q-constant band-pass filters $\hat\psi_\lambda$

- Wavelet transform: $W x = \begin{pmatrix} x \star \phi_{2^J}(t) \\ x \star \psi_\lambda(t) \end{pmatrix}_{\lambda \leq 2^J}$ : average ; : higher frequencies

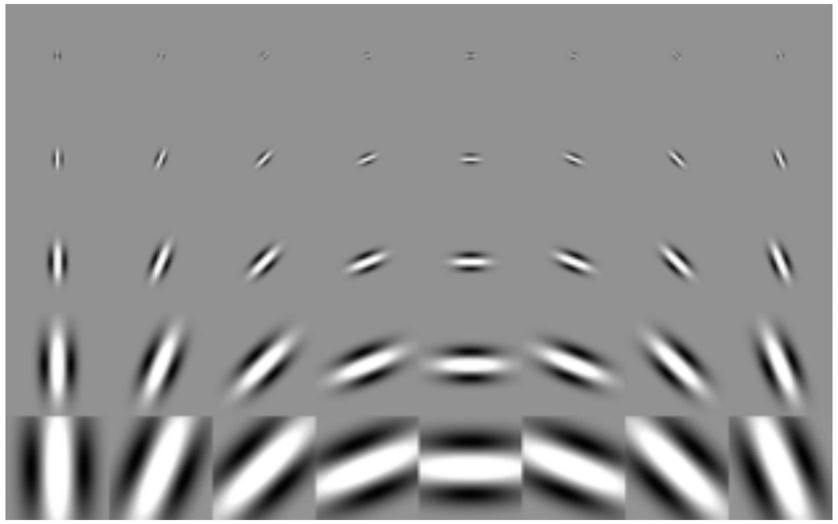
Preserves norm: $\|Wx\|^2 = \|x\|^2$ .

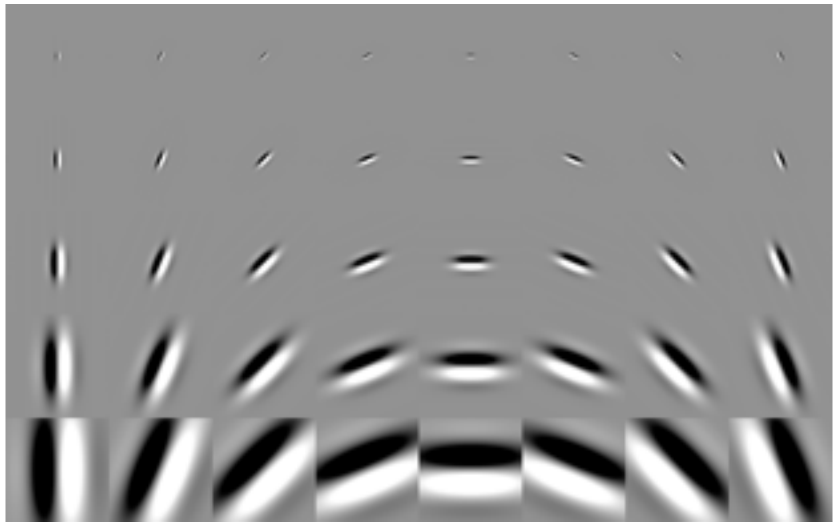
# Scale separation with Wavelets

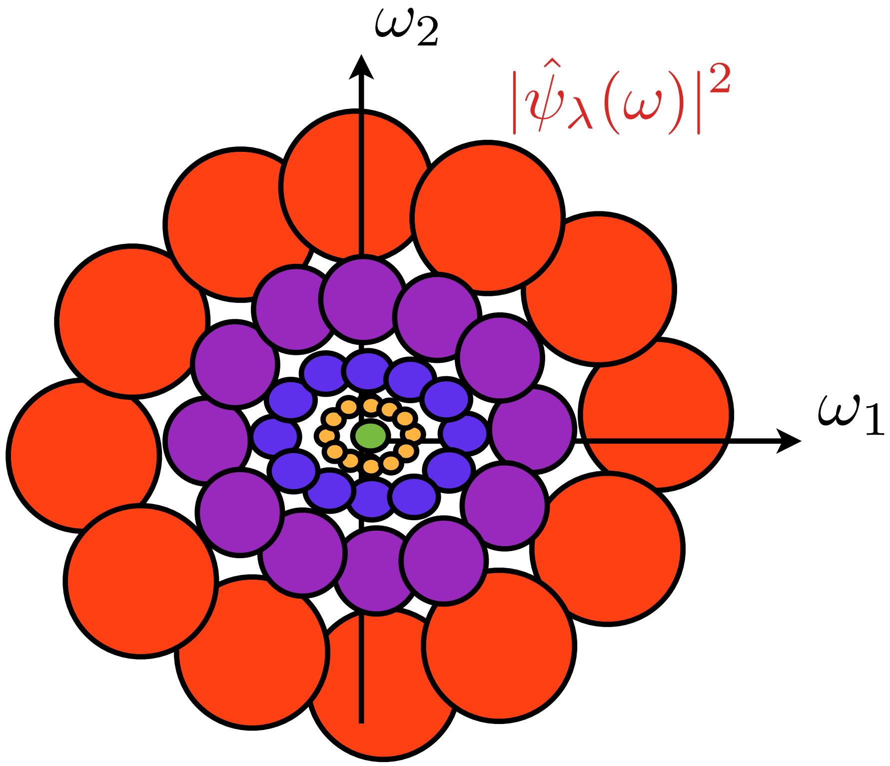
- Complex wavelet: $\psi(t) = g(t) \exp i\xi t \quad , \quad t = (t_1, t_2)$

  rotated and dilated: $\psi_\lambda(t) = 2^{-j}\, \psi(2^{-j} r_\theta t)$ with $\lambda = (2^j, \theta)$

real parts imaginary parts



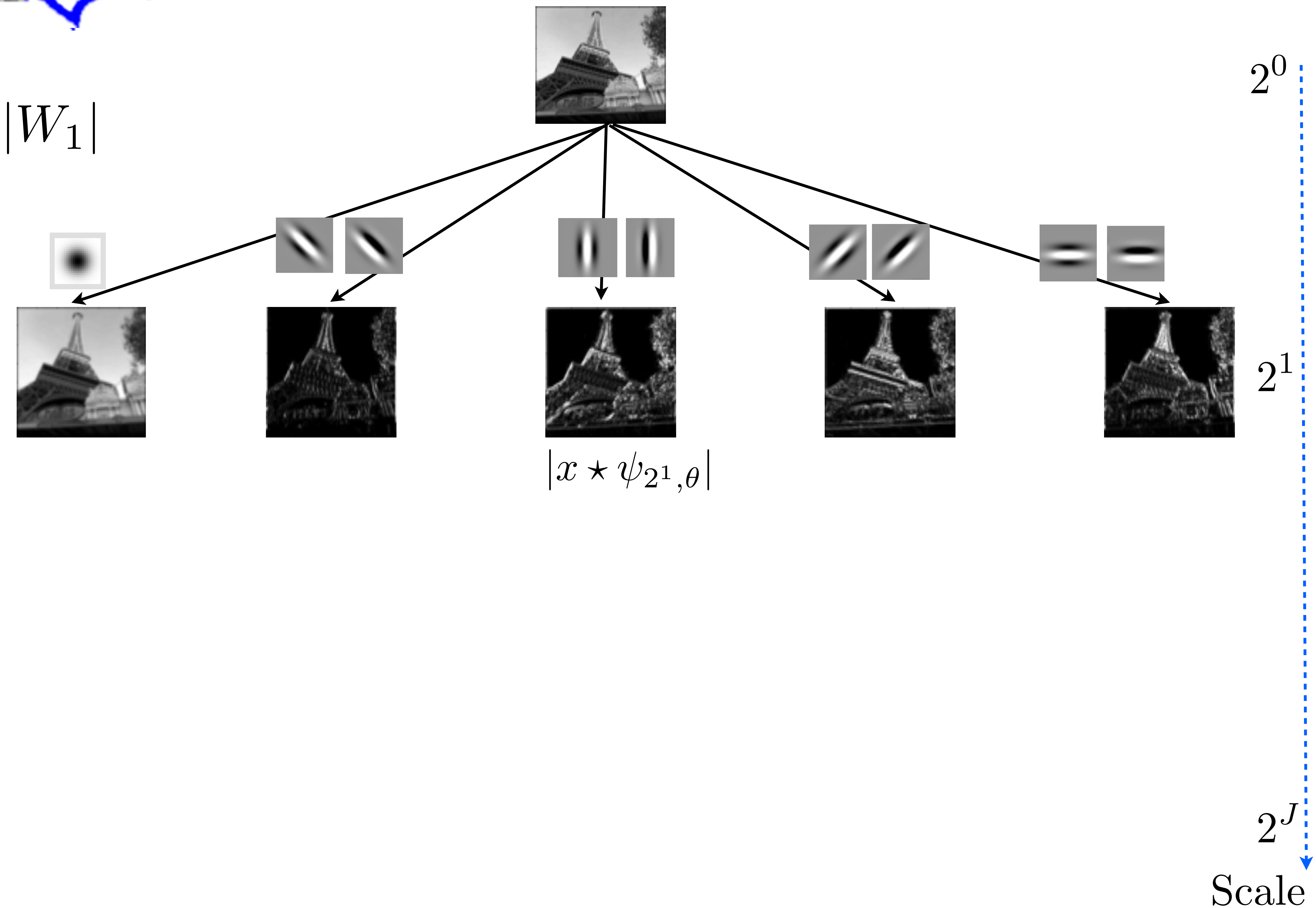




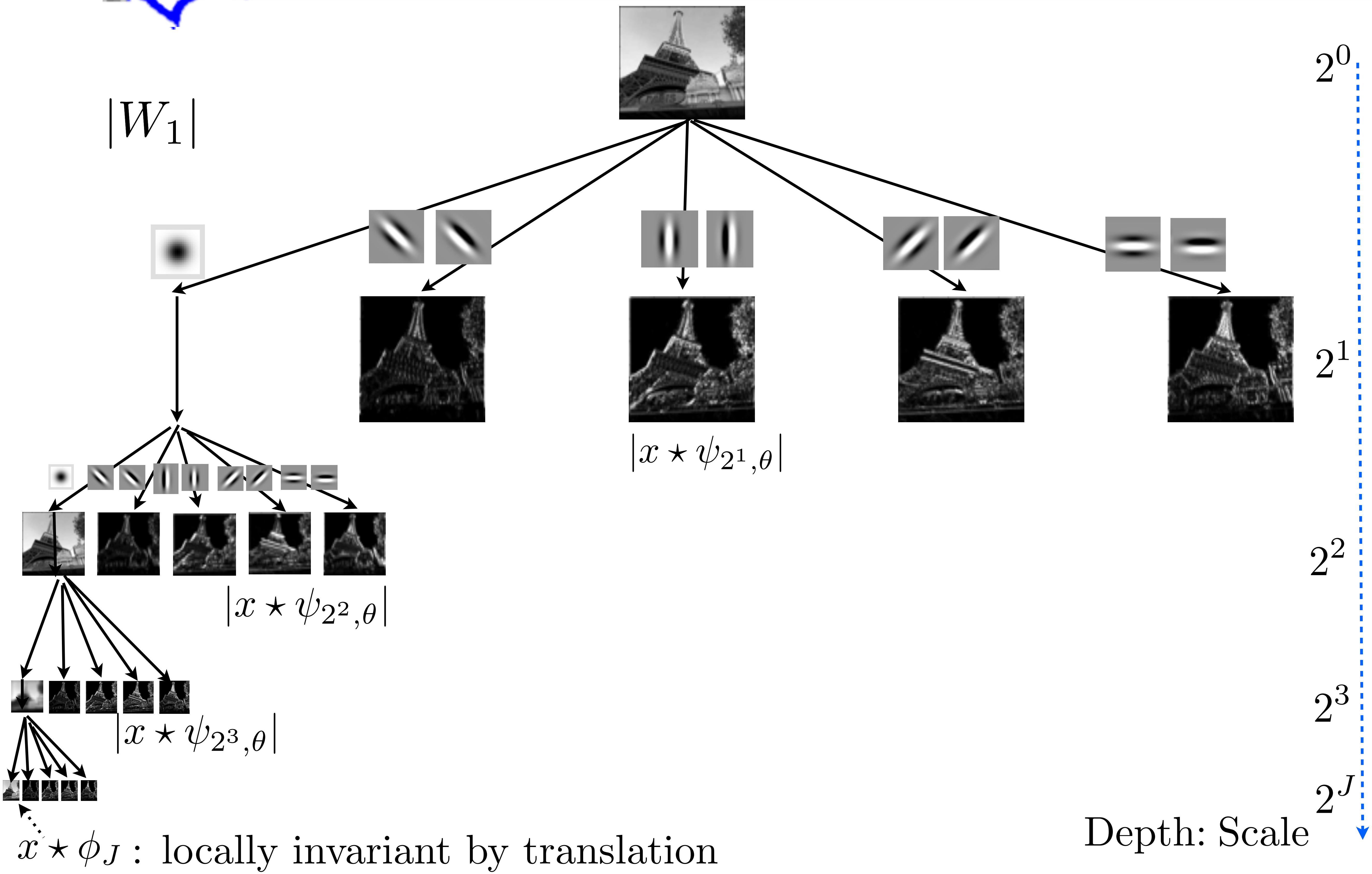

- Wavelet transform: $$W x = \begin{pmatrix} x \star \phi_{2^J}(t) \\ x \star \psi_\lambda(t) \end{pmatrix}_{\lambda \leq 2^J} \begin{array}{l} : \text{ average} \\ : \text{ higher frequencies} \end{array}$$

  Preserves norm: $\|Wx\|^2 = \|x\|^2$ .

# Fast Wavelet Transform



# Wavelet Transform



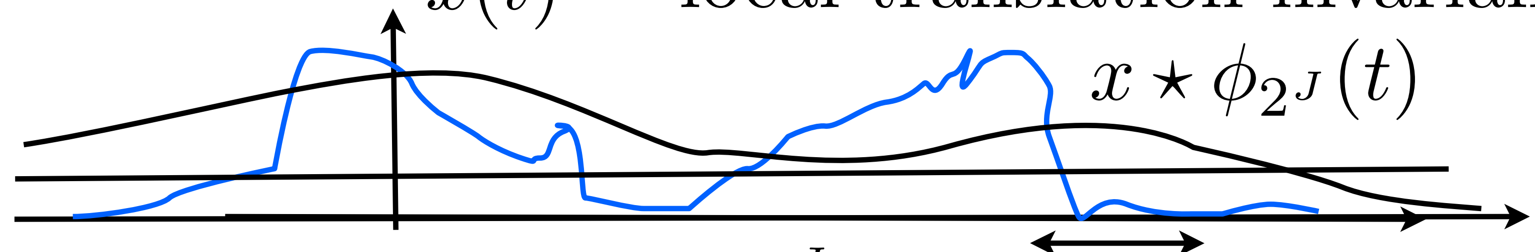

How to make everything invariant to translation ?

# Wavelet Translation Invariance

First wavelet transform

$$W_1 x = \begin{pmatrix} x \star \phi_{2^J} \\ x \star \psi_{\lambda_1} \end{pmatrix}_{\lambda_1}$$

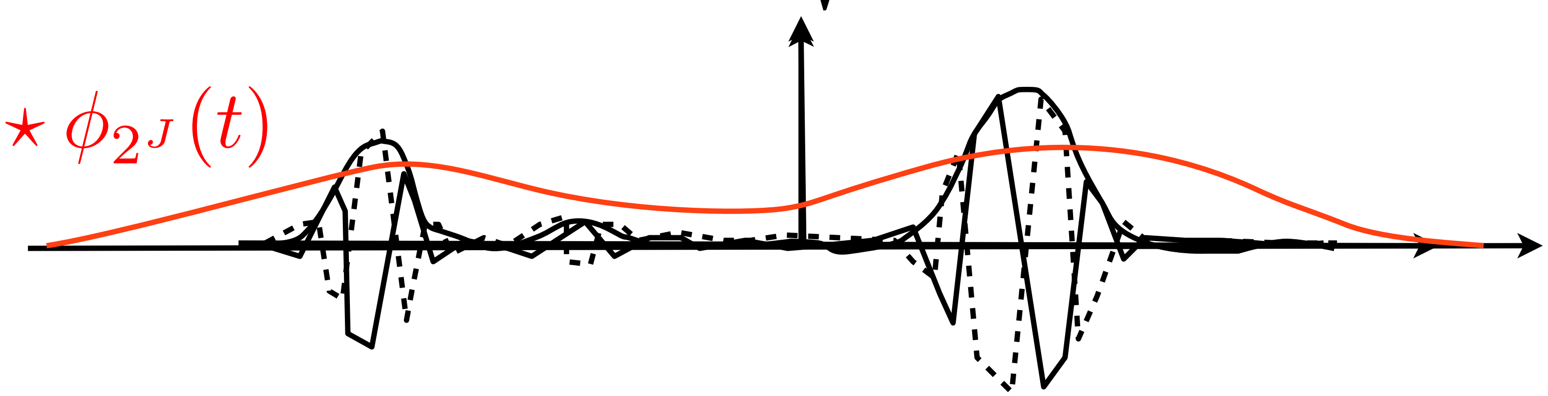
$$|W_1| x = \begin{pmatrix} x \star \phi_{2^J} \\ |x \star \psi_{\lambda_1}| \end{pmatrix}_{\lambda_1}$$




Modulus improves invariance: $|x \star \psi_{\lambda_1}(t)| = \sqrt{|x \star \psi^a_{\lambda_1}(t)|^2 + |x \star \psi^b_{\lambda_1}(t)|^2}$

$x \star \psi_{\lambda_1}(t) = x \star \psi^a_{\lambda_1}(t) + i\, x \star \psi^b_{\lambda_1}(t)$

but covariant

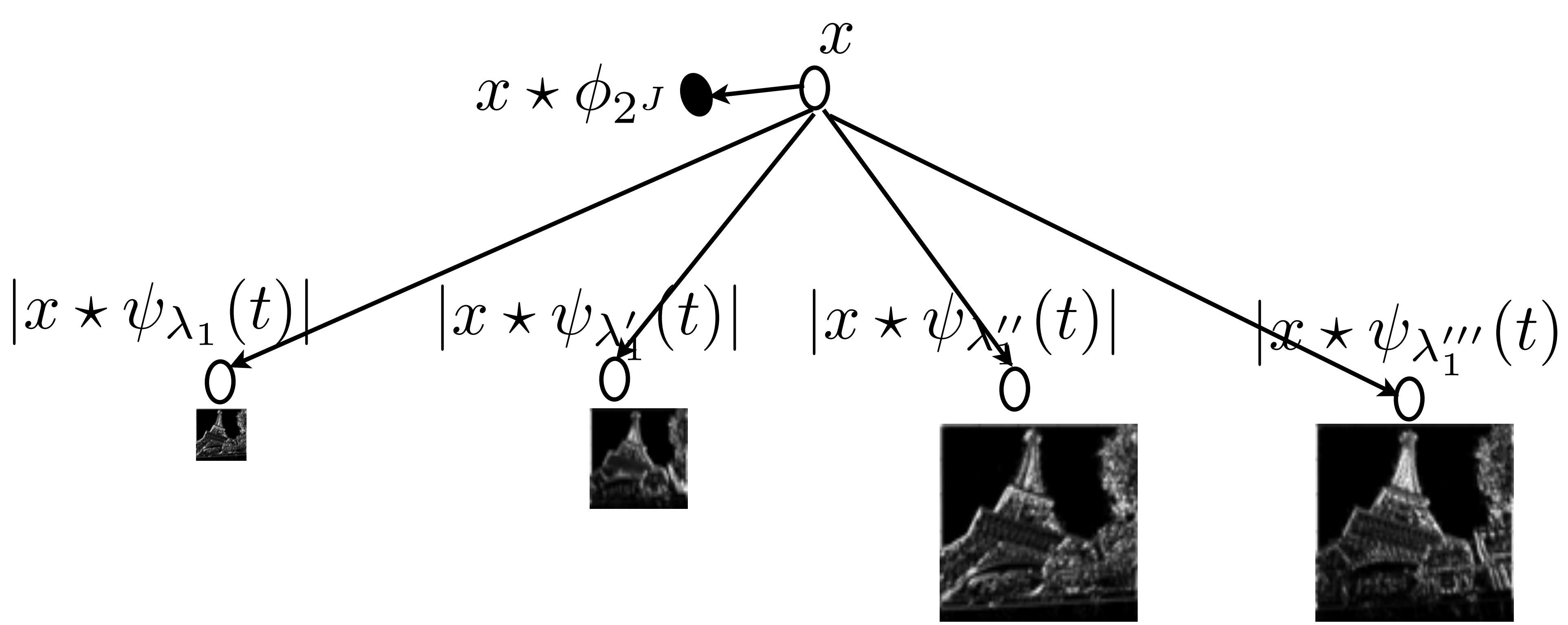



Second wavelet transform modulus

$$|W_2|\ |x \star \psi_{\lambda_1}| = \begin{pmatrix} |x \star \psi_{\lambda_1}| \star \phi_{2^J}(t) \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}(t)| \end{pmatrix}_{\lambda_2}$$

# Scattering Transform

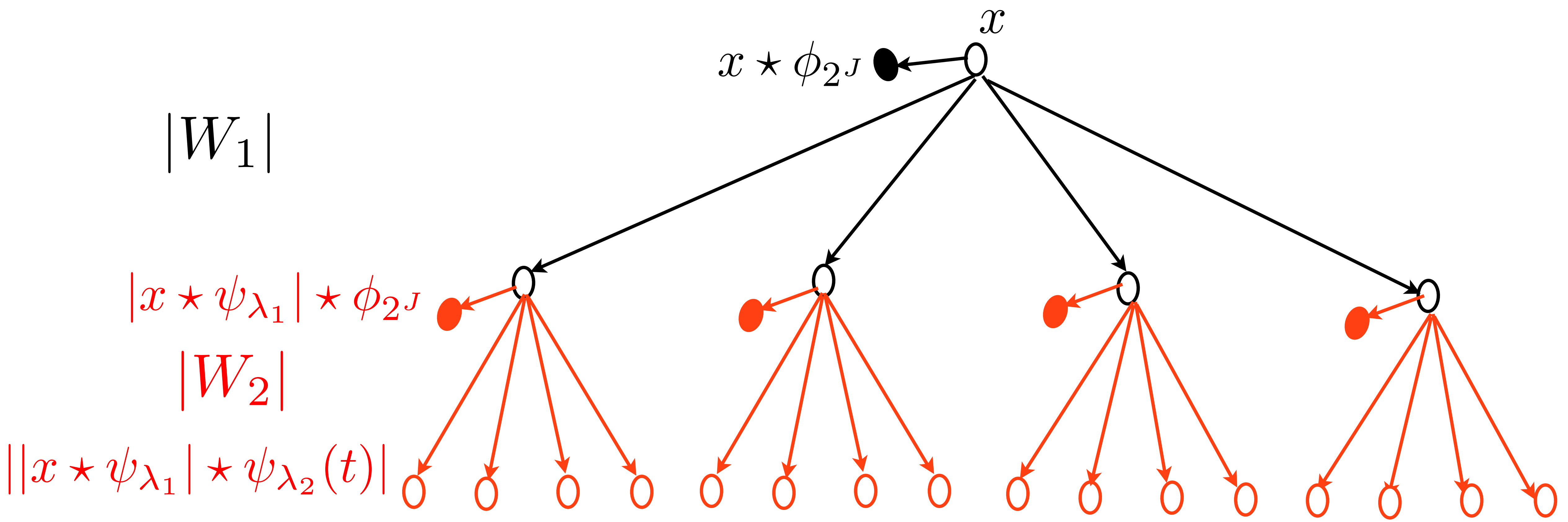
$|W_1|$



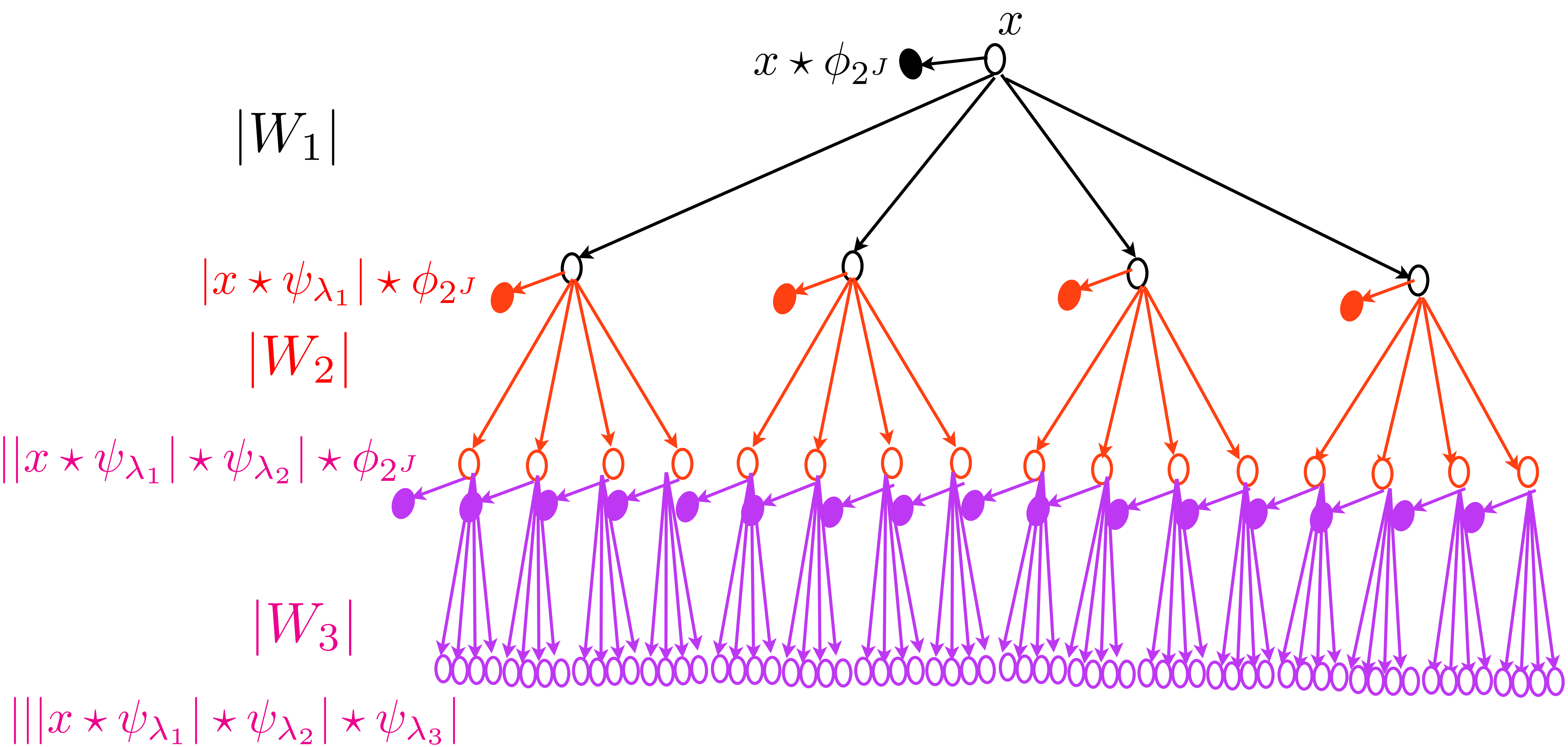
# Scattering Transform

$|W_1|$

$|W_2|$



# Scattering Neural Network



# Scattering Properties

$$S_J x = \left( \begin{array}{c} x \star \phi_{2^J} \\ |x \star \psi_{\lambda_1}| \star \phi_{2^J} \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_{2^J} \\ |||x \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \star \phi_{2^J} \\ \dots \end{array} \right)_{\lambda_1,\lambda_2,\lambda_3,\dots} = \dots |W_3|\,|W_2|\,|W_1|\,x$$

$W_k$ is unitary $\Rightarrow |W_k|$ is contractive

**Theorem**: *For appropriate wavelets, a scattering is*

*contractive* $\|S_J x - S_J y\| \leq \|x - y\|$ ($\mathbf{L^2}$ *stability*)

*preserves norms* $\|S_J x\| = \|x\|$

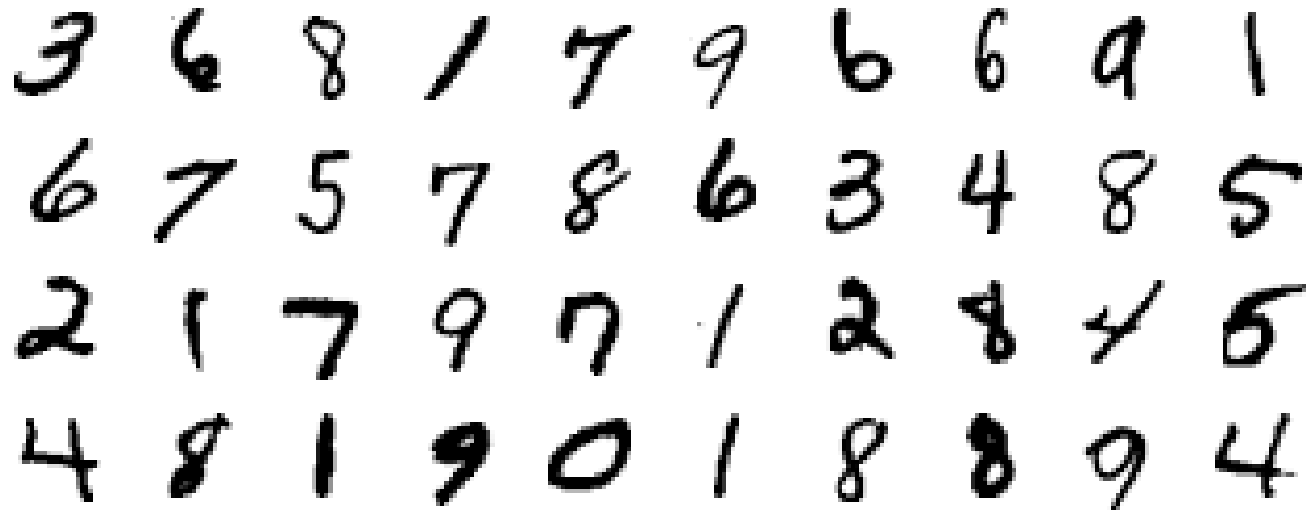
*translations invariance and deformation stability:*

*if* $x_\tau(u) = x(u - \tau(u))$ *then*

$$\lim_{J\to\infty} \|S_J x_\tau - S_J x\| \leq C\,\|\nabla\tau\|_\infty\,\|x\|$$

# Digit Classification: MNIST

*Joan Bruna*



$$x \longrightarrow \boxed{S_J x} \longrightarrow \boxed{\text{Linear Classifier}} \longrightarrow y = f(x)$$

Classification Errors

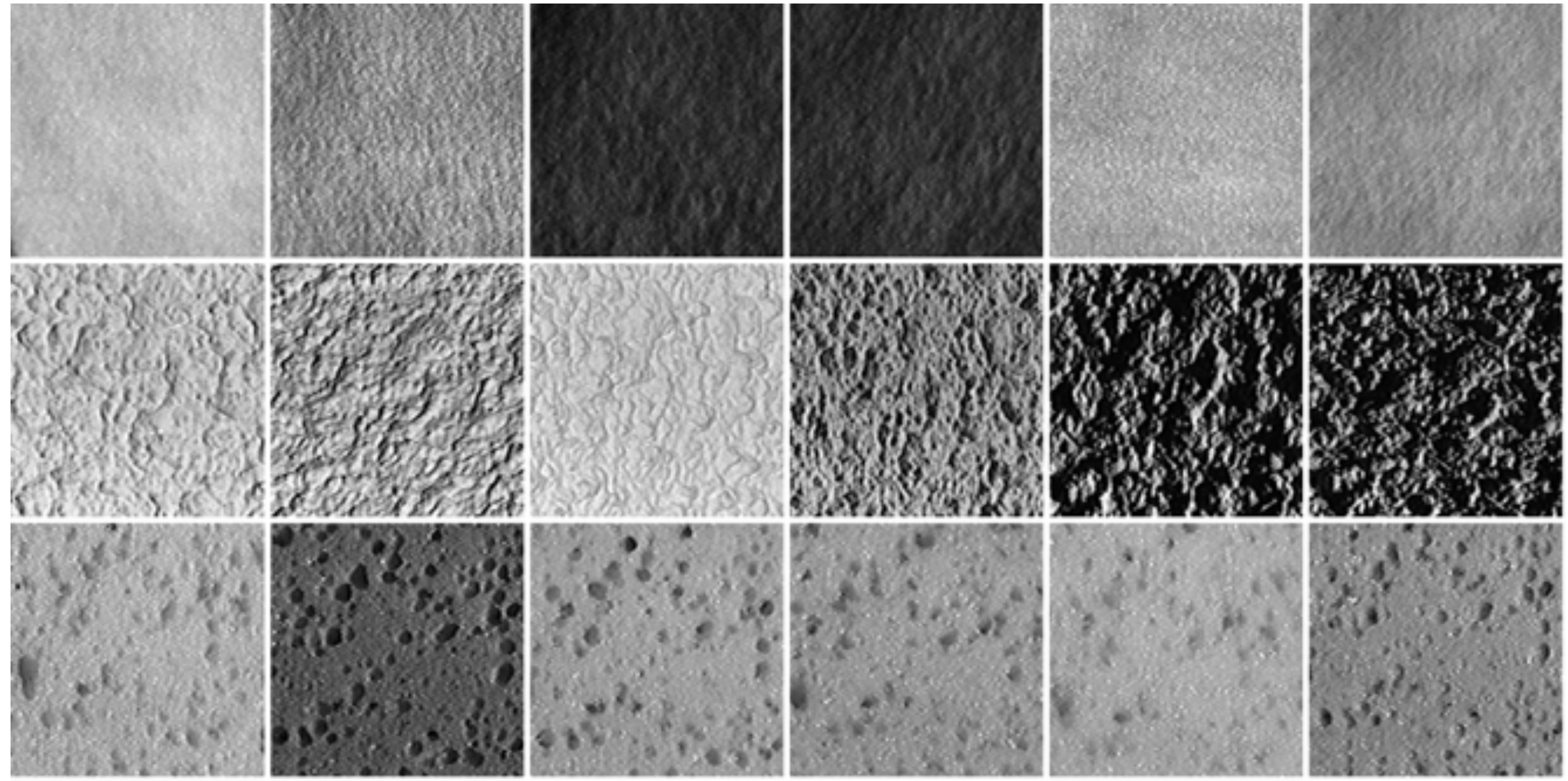
| Training size | Conv. Net. | Scattering |
|---|---|---|
| 50000 | 0.5% | **0.4%** |

LeCun et. al.

# Classification of Textures

*J. Bruna*

CUREt database
61 classes



Scat. Moments

$x \longrightarrow \boxed{S_J x} \longrightarrow$ Linear Classifier $\longrightarrow y = f(x)$

Classification Errors $\quad 2^J =$ image size

| Training per class | Fourier Spectr. | Histogr. Features | Scattering |
|---|---|---|---|
| 46 | 1% | 1% | **0.2** % |

# Scattering Moments of Processes

The scattering transform of a stationary process $X(t)$

$$S_J X = \begin{pmatrix} X \\ |X \star \psi_{\lambda_1}| \\ ||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \\ |||X \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \\ \dots \end{pmatrix} \star \phi_{2^J}$$

: Gaussian for $2^J$ large if $X$ is ergodic

$$\Big\downarrow\; J \to \infty$$

$$\mathbb{E}(SX) = \begin{pmatrix} \mathbb{E}(X) \\ \mathbb{E}(|X \star \psi_{\lambda_1}|) \\ \mathbb{E}(||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|) \\ \mathbb{E}(|||X \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}|) \\ \dots \end{pmatrix}_{\lambda_1, \lambda_2, \lambda_3, \dots}$$

# Representation of Random Processes

$$\mathbb{E}(SX) = \left( \begin{array}{rcl} \mathbb{E}(X) & = & \mathbb{E}(U_0 X) \\ \mathbb{E}(|X \star \psi_{\lambda_1}|) & = & \mathbb{E}(U_1 X) \\ \mathbb{E}(||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|) & = & \mathbb{E}(U_2 X) \\ \mathbb{E}(|||X \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}|) & = & \mathbb{E}(U_3 X) \\ \dots & & \end{array} \right)_{\lambda_1,\lambda_2,\lambda_3,\dots}$$

**Theorem** *(Boltzmann)* The distribution $p(x)$ which satisfies

$$\int_{\mathbb{R}^N} U_m x \; p(x)\, dx = E(U_m X)$$

with a maximum entropy $H_{\max} = -\int p(x) \log p(x)\, dx$ is

$$p(x) = \frac{1}{Z} \exp\Big( \sum_{m=1}^{\infty} \lambda_m \,.\, U_m x \Big)$$

$H_{\max} \geq H(X)$ (entropie of X)

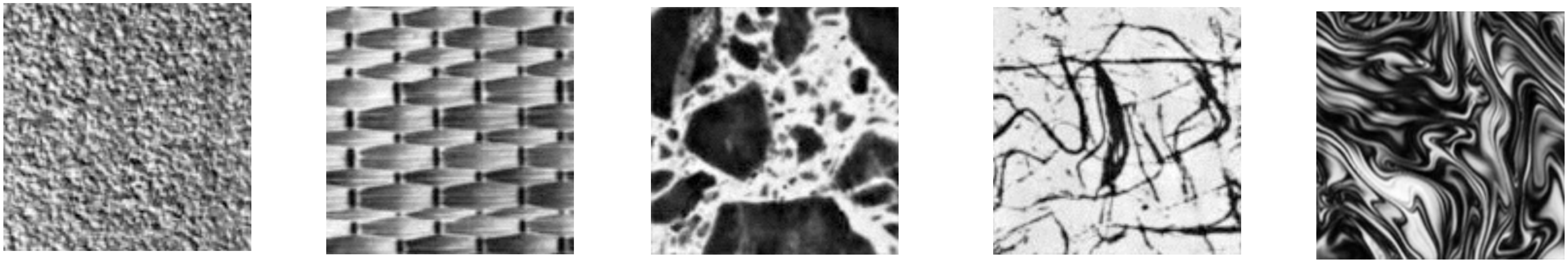
Little loss of information: $H_{\max} \approx H(X)$

# Ergodic Texture Reconstructions

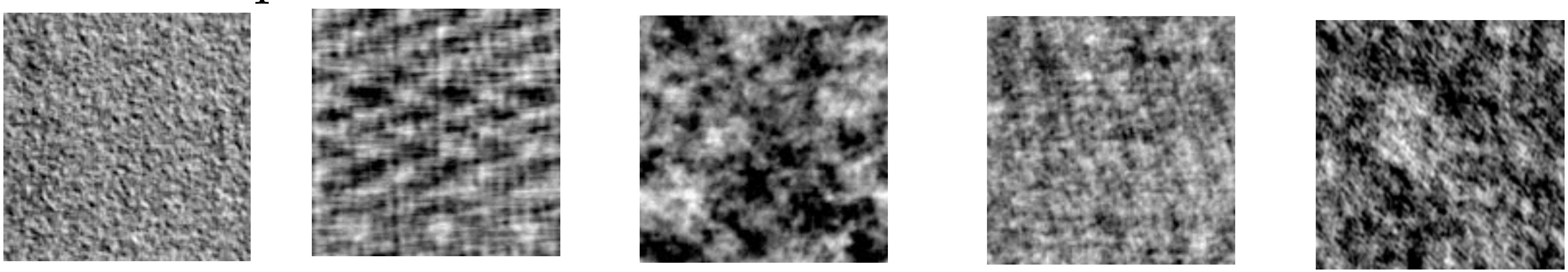
*Joan Bruna*

Original Textures

2D Turbulence

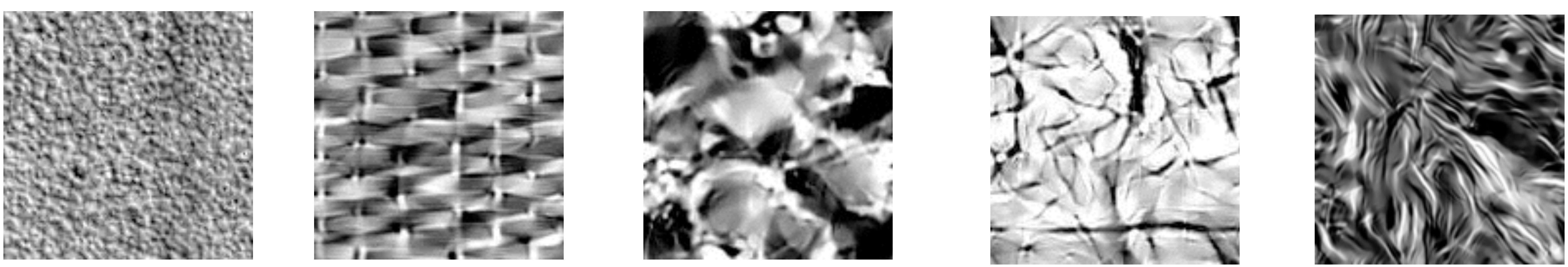


Gaussian process model with same second order moments



Second order Gaussian Scattering: $O(\log N^2)$ moments

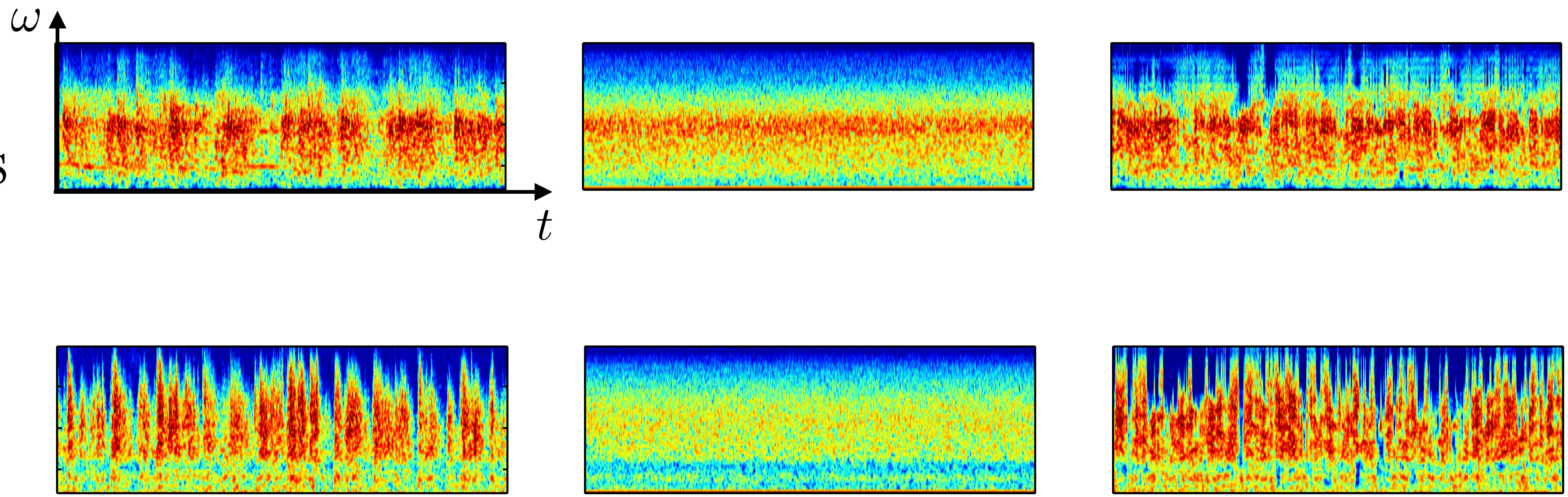
$\mathbb{E}(|x \star \psi_{\lambda_1}|)$ , $\mathbb{E}(||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|)$



# Representation of Audio Textures

*Joan Bruna*

Original | Gaussian in time | Gaussian in scattering




Applauds

Paper

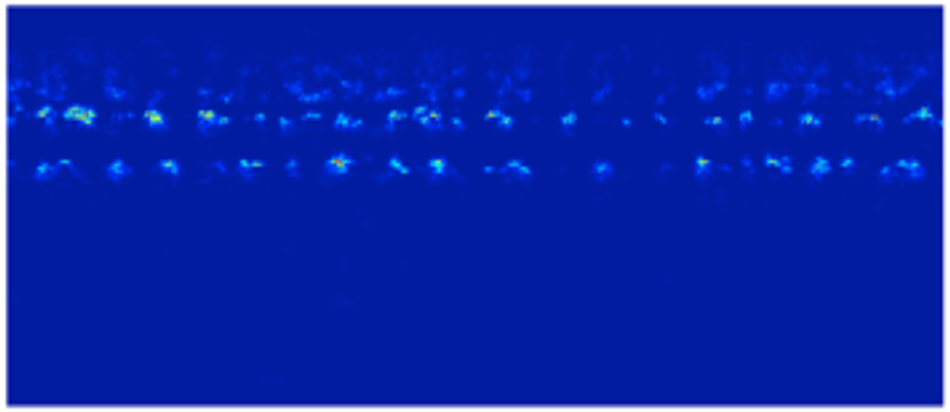
Cocktail Party

# Failures: Harmonic Sounds

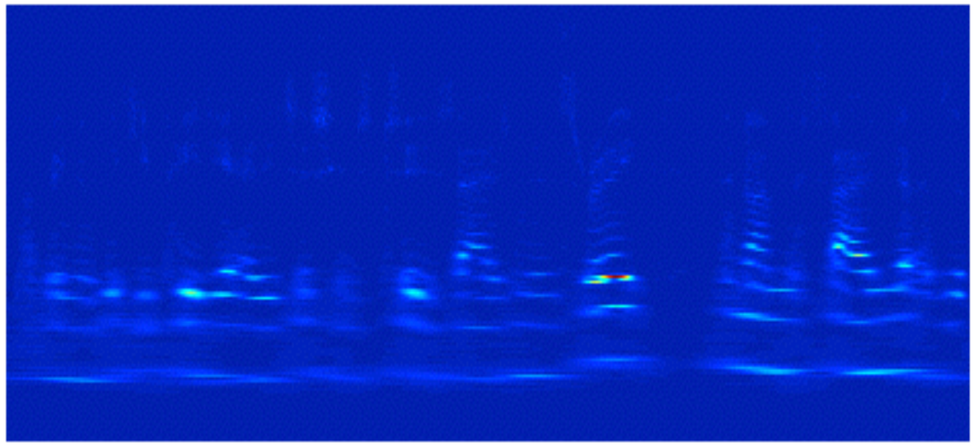


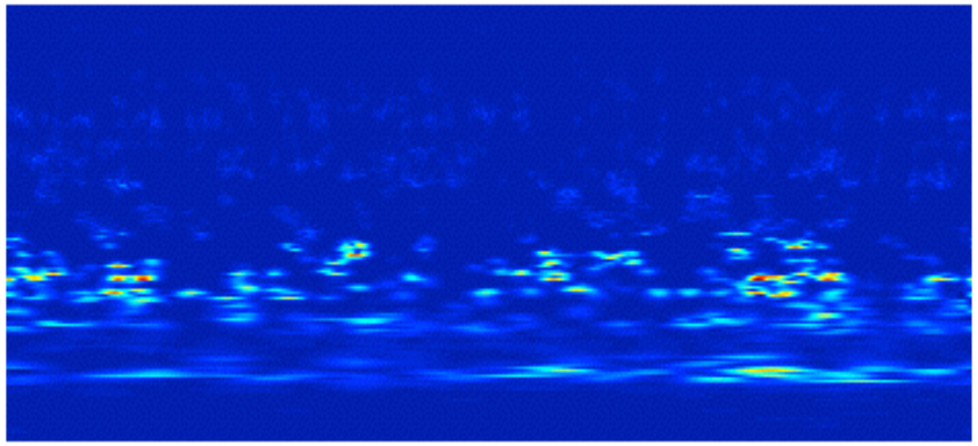
Need to express frequency channel interactions: time-frequency image

Bird

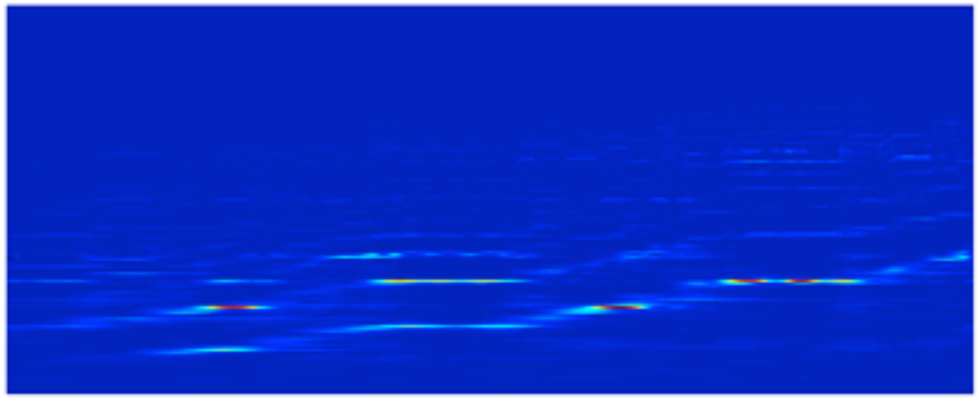


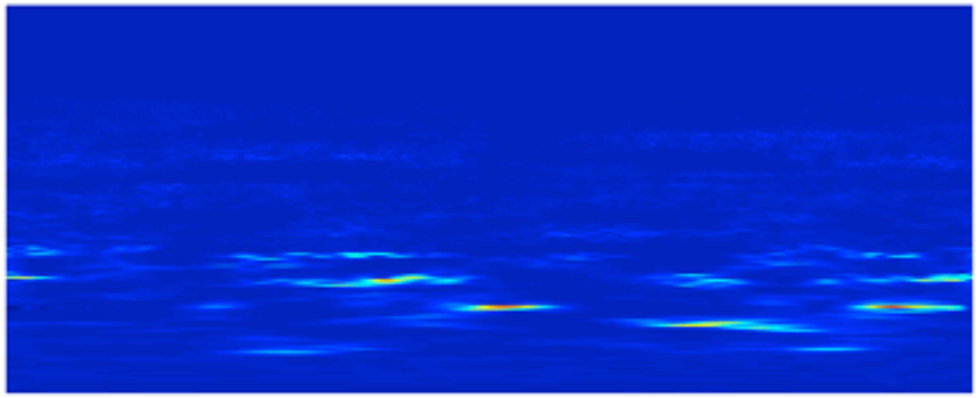
Speech



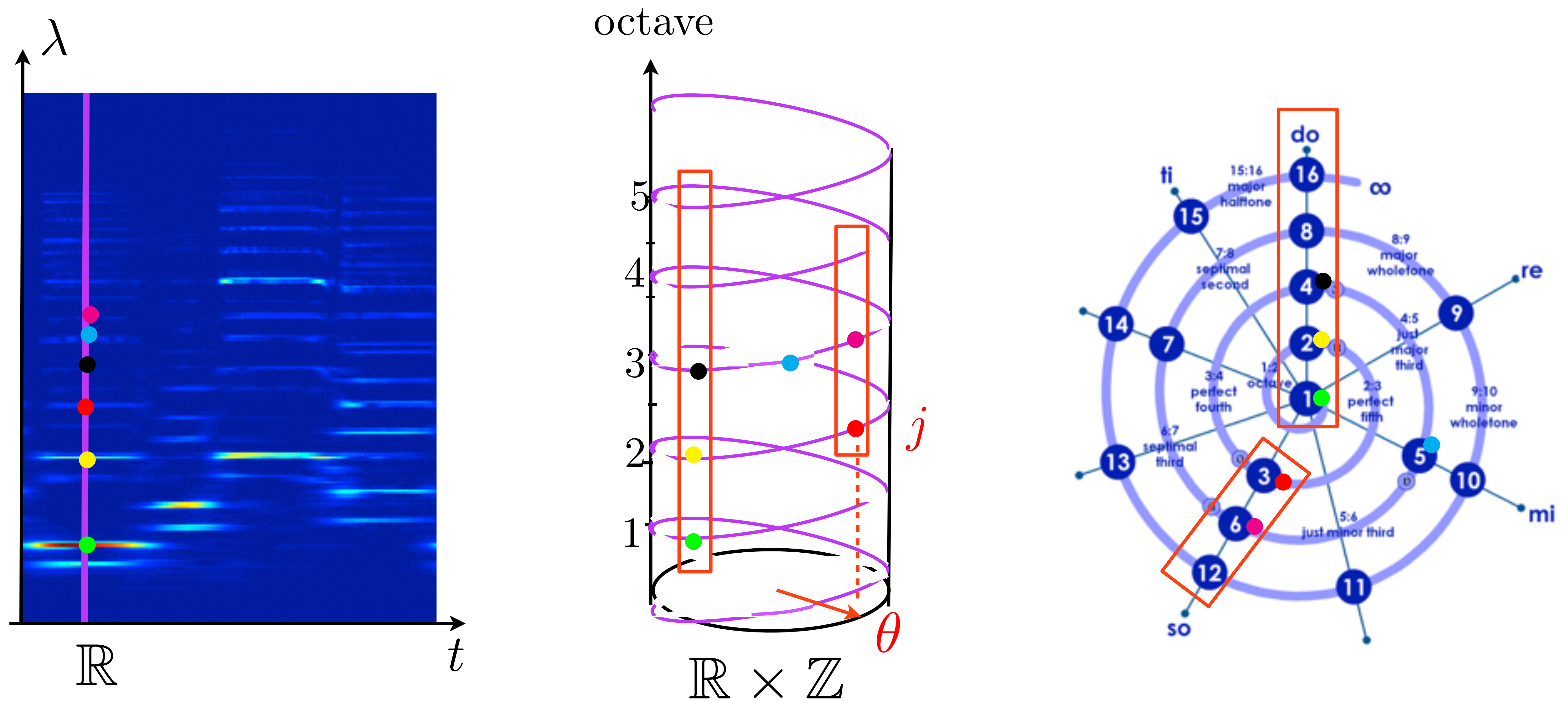


Cello



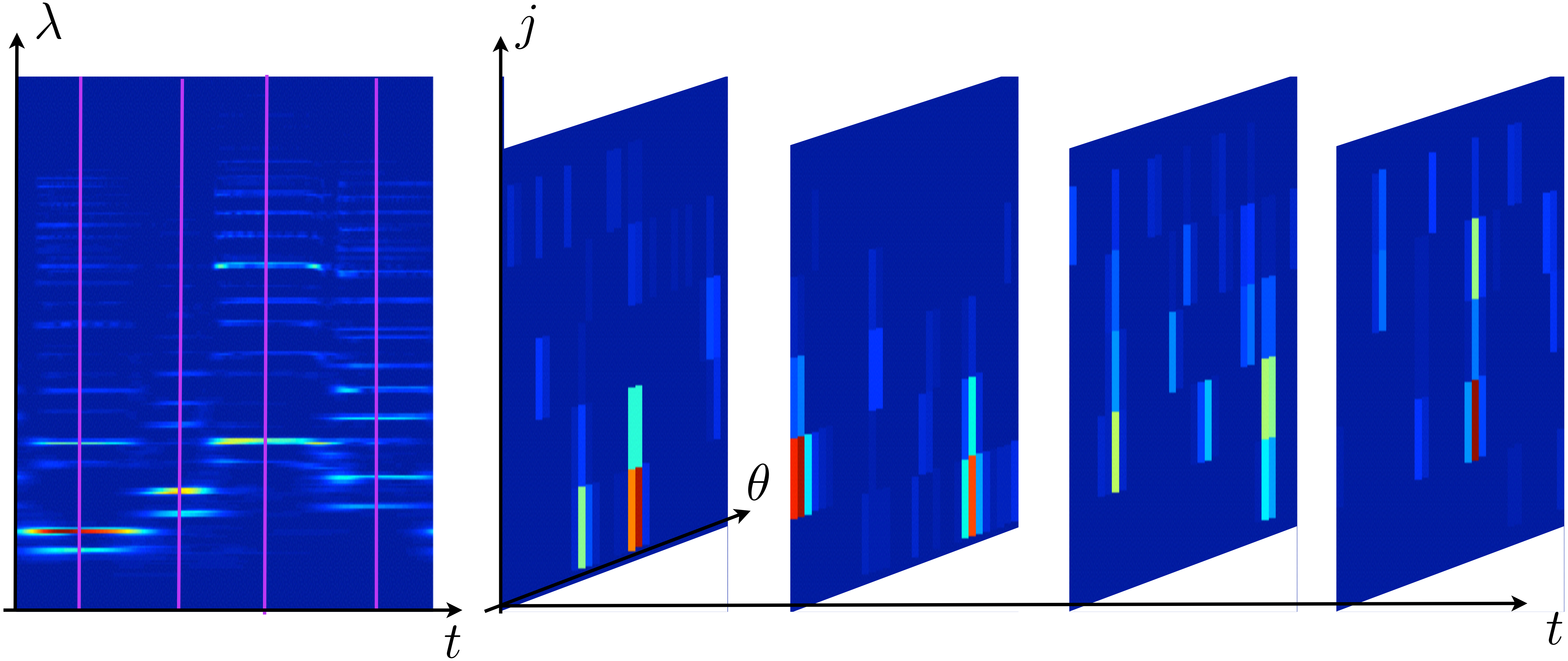


Need to capture frequency variability and structures.

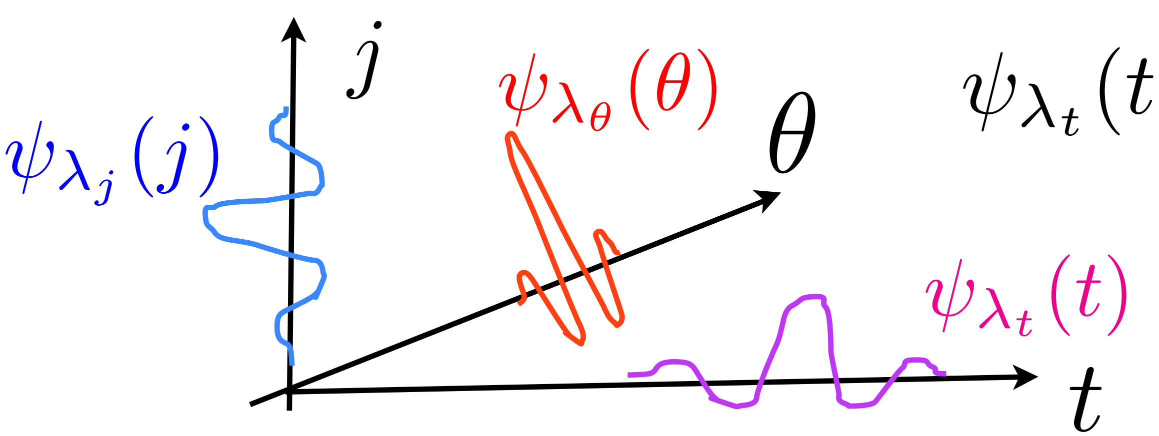



- Alignment of harmonics in two main groups.

  More regular variations along $(\theta, j)$ than $\lambda$



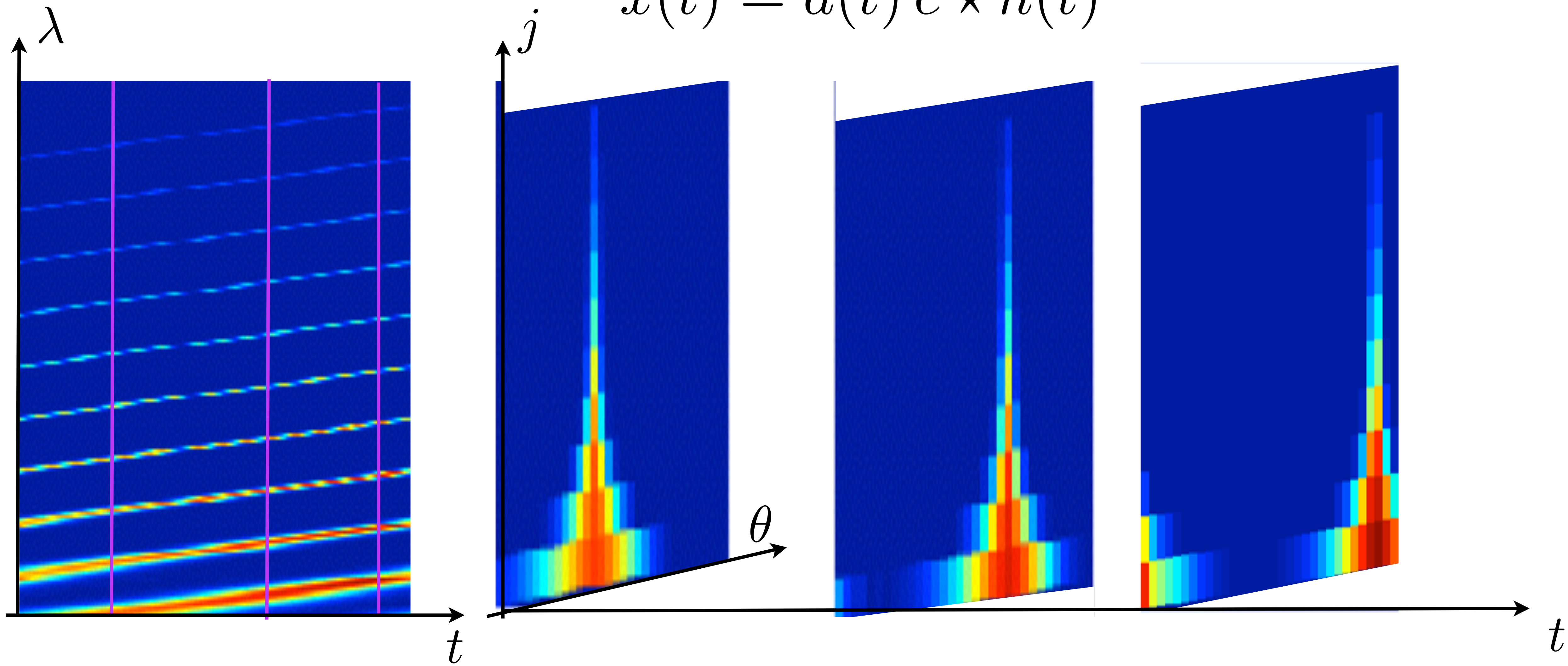

3D separable Spiral wavelet transform $W_2$



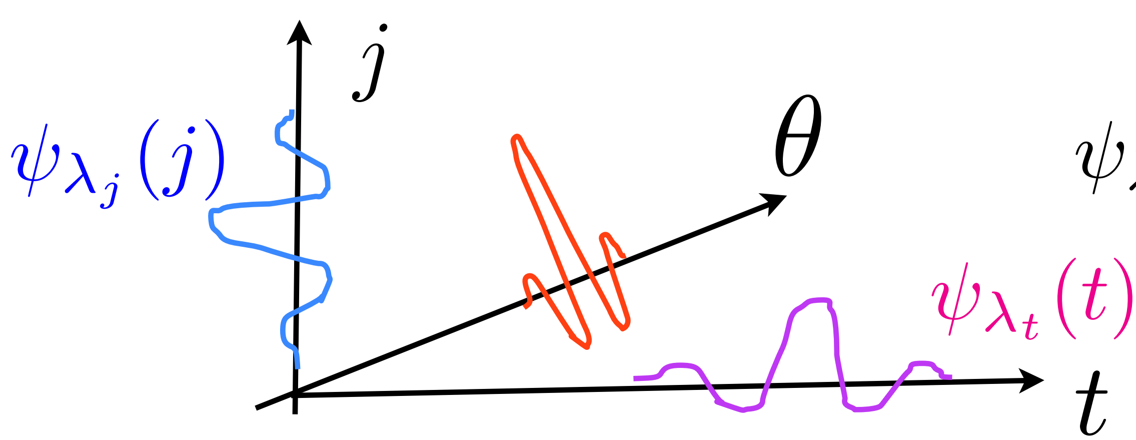

$$\psi_{\lambda_t}(t)\,\psi_{\lambda_\theta}(\theta)\,\psi_{\lambda_j}(j)$$

# Shepard-Risset Glissando

$$x(t) = a(t)\, e \star h(t)$$




3D separable Spiral wavelet transform $W_2$

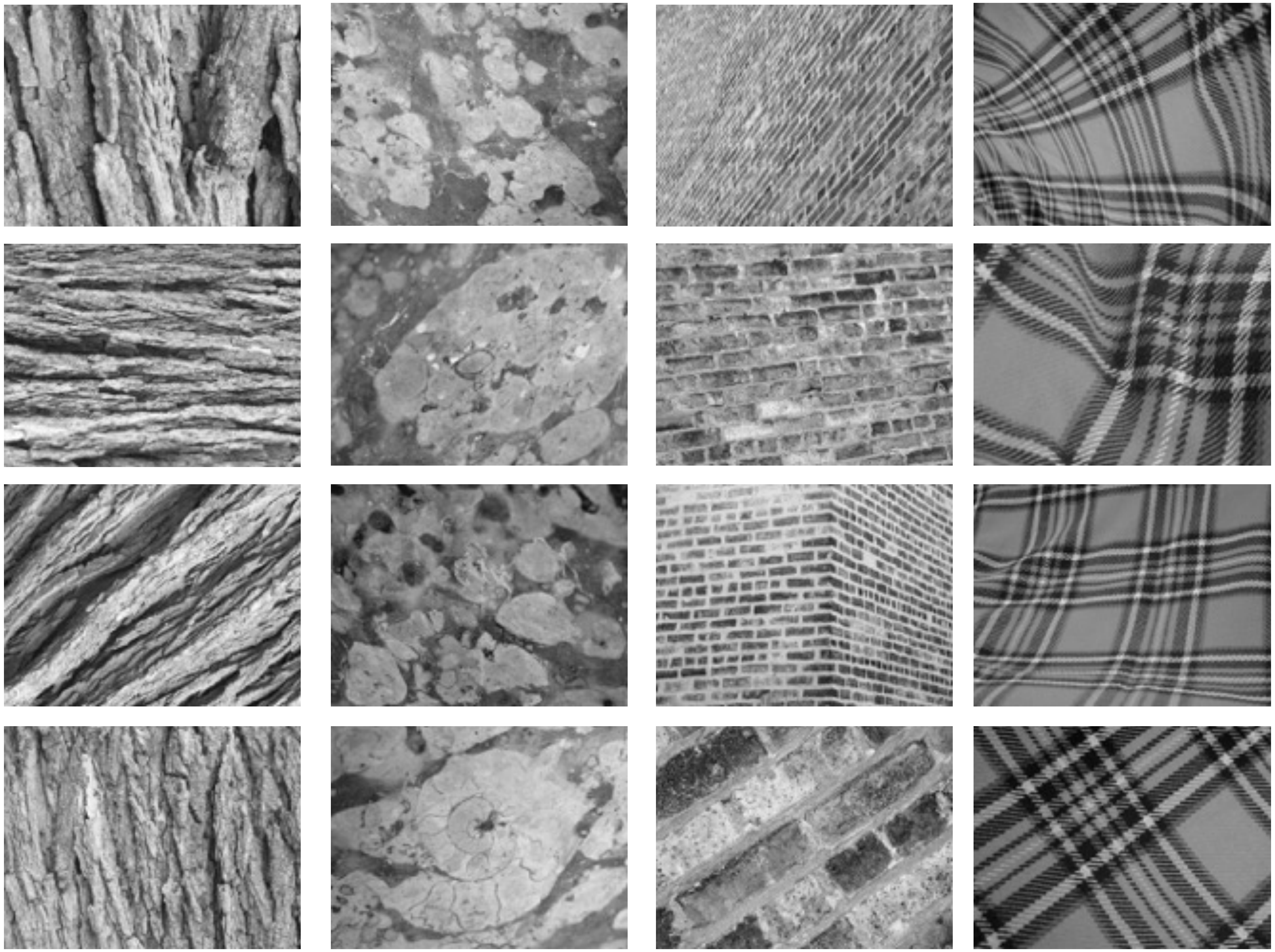



$$\psi_{\lambda_t}(t)\, \psi_{\lambda_\theta}(\theta)\, \psi_{\lambda_j}(j)$$

# Rotation and Scaling Invariance

*Laurent Sifre*

UIUC database:
25 classes



Scattering classification errors

| Training | Scat. Translation |
|---|---|
| 20 | 20 % |

# Extension to Rigid Mouvements

*Laurent Sifre*

Need to capture the variability of spatial directions.

- Group of rigid displacements: translations and rotations

- Action on wavelet coefficients:

rotation & translation | rotation & translation , angle translation

$$x(u) \qquad x(r_\alpha(u)) \longrightarrow \boxed{|W_1|} \longrightarrow x_j(u,\theta) = |x \star \psi_{2^j,\theta}(u)| \qquad x_j(r_\alpha u, \theta - \alpha)$$

$$\downarrow$$

$$\int x(u)du$$

# Extension to Rigid Mouvements

*Laurent Sifre*

- To build invariants: second wavelet transform on $\mathbf{L^2}(G)$: convolutions of $x_j(u,\theta)$ with wavelets $\psi_{\lambda_2}(u,\theta)$

$$x_j \circledast \psi_{\lambda_2}(u,\theta) = \int_{\mathbb{R}^2} \int_0^{2\pi} x_j(v,\alpha)\, \psi_{\lambda_2}(u-v, \theta-\alpha)\, dv\, d\alpha$$

- Scattering on rigid mouvements:

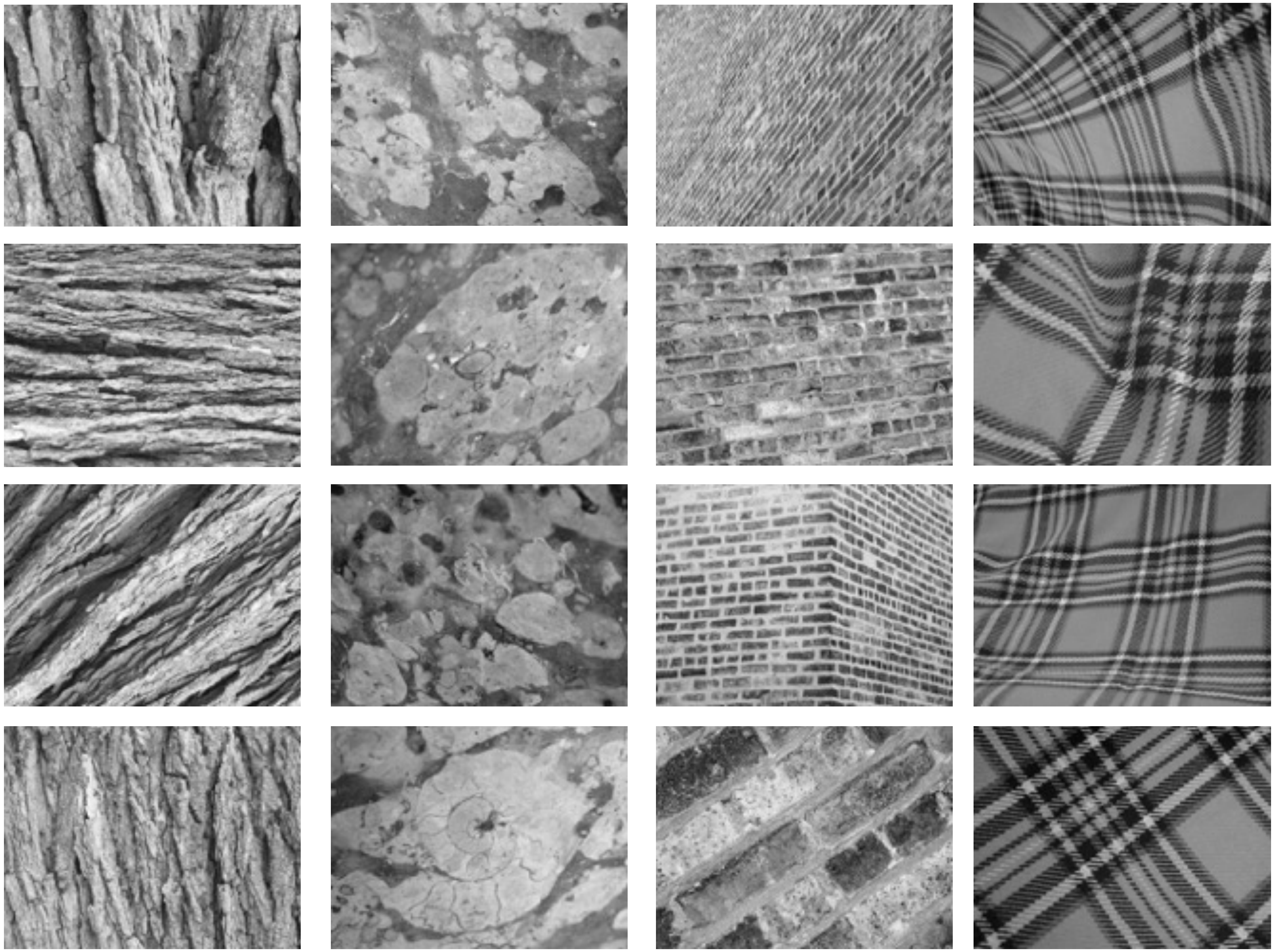
Wavelets on Translations  Wavelets on Rigid Mvt.  Wavelets on Rigid Mvt.

$$\begin{tikzcd}
x(u) \arrow[r] & \boxed{|W_1|} \arrow[r] \arrow[d] & x_j(u\,,\,\theta) \arrow[r] & \boxed{|W_2|} \arrow[r] \arrow[d] & |x_j \circledast \psi_{\lambda_2}(v,\theta)| \arrow[r] & \boxed{|W_3|} \arrow[r] \arrow[d] & {} \\
& \int x(u) du & & \int x_j(u,\theta)\, dud\theta & & \int |x_j \circledast \psi_{\lambda_2}(v,\theta)| dud\theta &
\end{tikzcd}$$

# Rotation and Scaling Invariance

*Laurent Sifre*

UIUC database:
25 classes



Scattering classification errors

| Training | Scat. Translation | Scat. Rigid Mouvt. |
|---|---|---|
| 20 | 20 % | **0.6**% |

# 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
variable selection : 2000

## Classification Accuracy

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

State of the art
Unsupervised

# Learning Physics: N-Body Problem

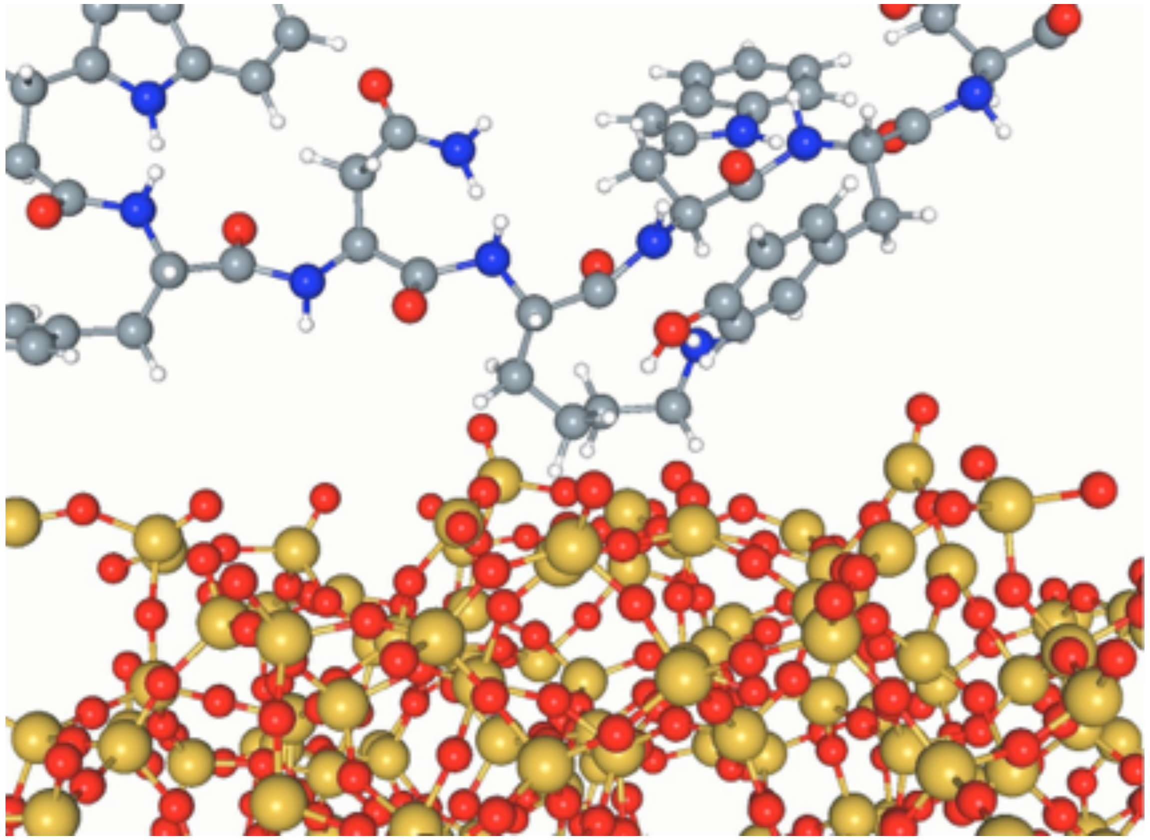
*N. Poilvert*
*Matthew Hirn*

- Energy of $d$ interacting bodies:

Can we learn the interaction energy $f(x)$ of a system with $x = \Big\{ \text{positions}, \text{values} \Big\}$ ?

Astronomy

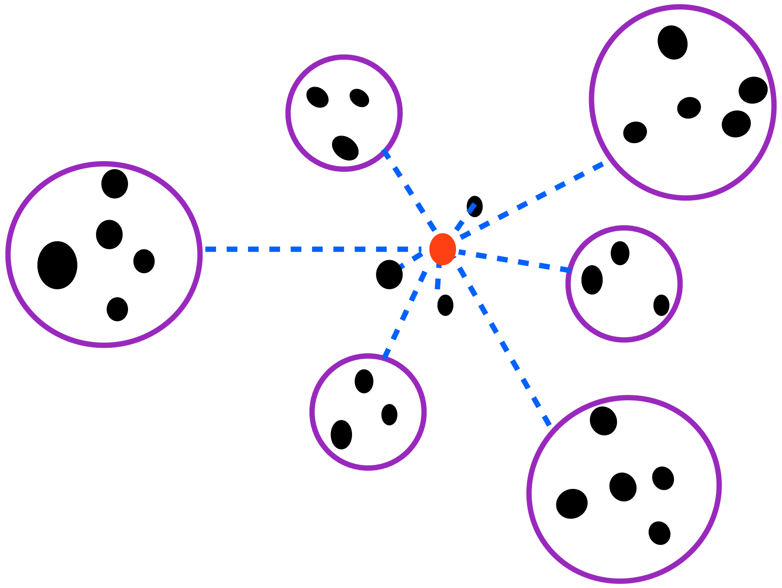
Quantum Chemistry



# Multiscale Interactions

- A system of $d$ particles involves $d^2$ interactions

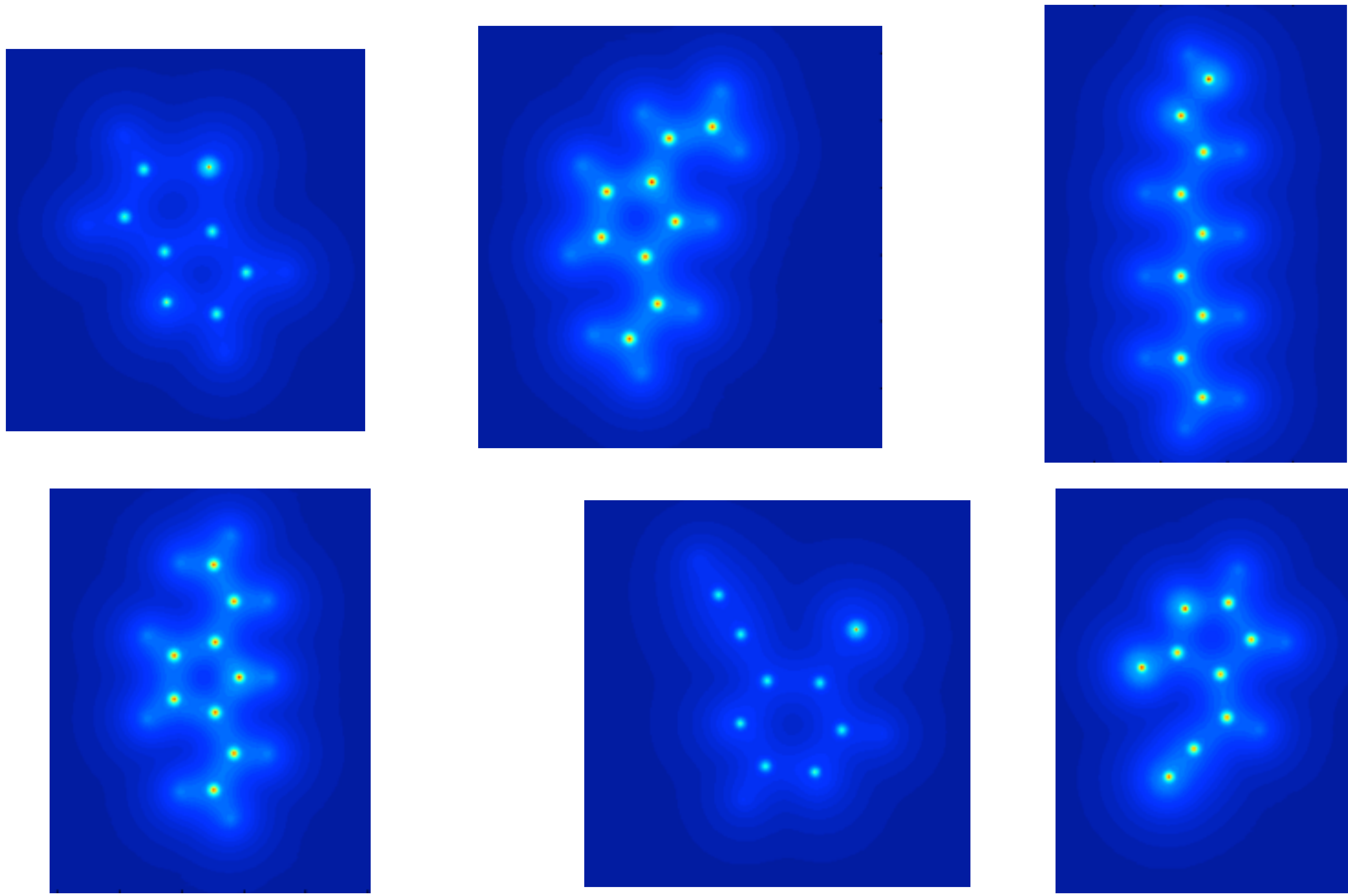
- Multiscale separation into $O(\log^2 d)$ interactions



# Quantum Chemistry

Electronic density $\rho_x(u)$: computed by solving Schrodinger

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



# Density Functional Theory

*Kohn-Sham* model:

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

Molecular energy

Kinetic energy

electron-nuclei attraction

electron-electron Coulomb repulsion

Exchange correlat. energy

At equilibrium:

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

- $f(x)$ is invariant to isometries and is deformation stable

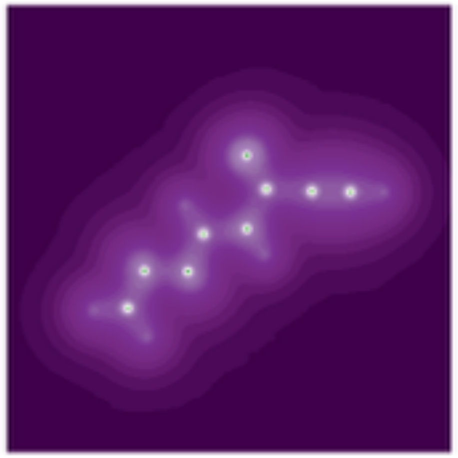
# Atomization Density

- We do not know the electronic density $\rho_x$ at equilibrium.

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

# Quantum Regression

*N. Poilvert*
*Matthew Hirn*

- Sparse regression computed over a representation invariant to action of isometries in $\mathbb{R}^3$:

$$\Phi x = \{\phi_n(\tilde{\rho}_x)\}_n \; : \left| \begin{array}{c} \text{Fourier modulus coefficients and squared} \\ \text{or} \\ \text{scattering coefficients and squared} \end{array} \right.$$

Partial Least Square regression on the training set:

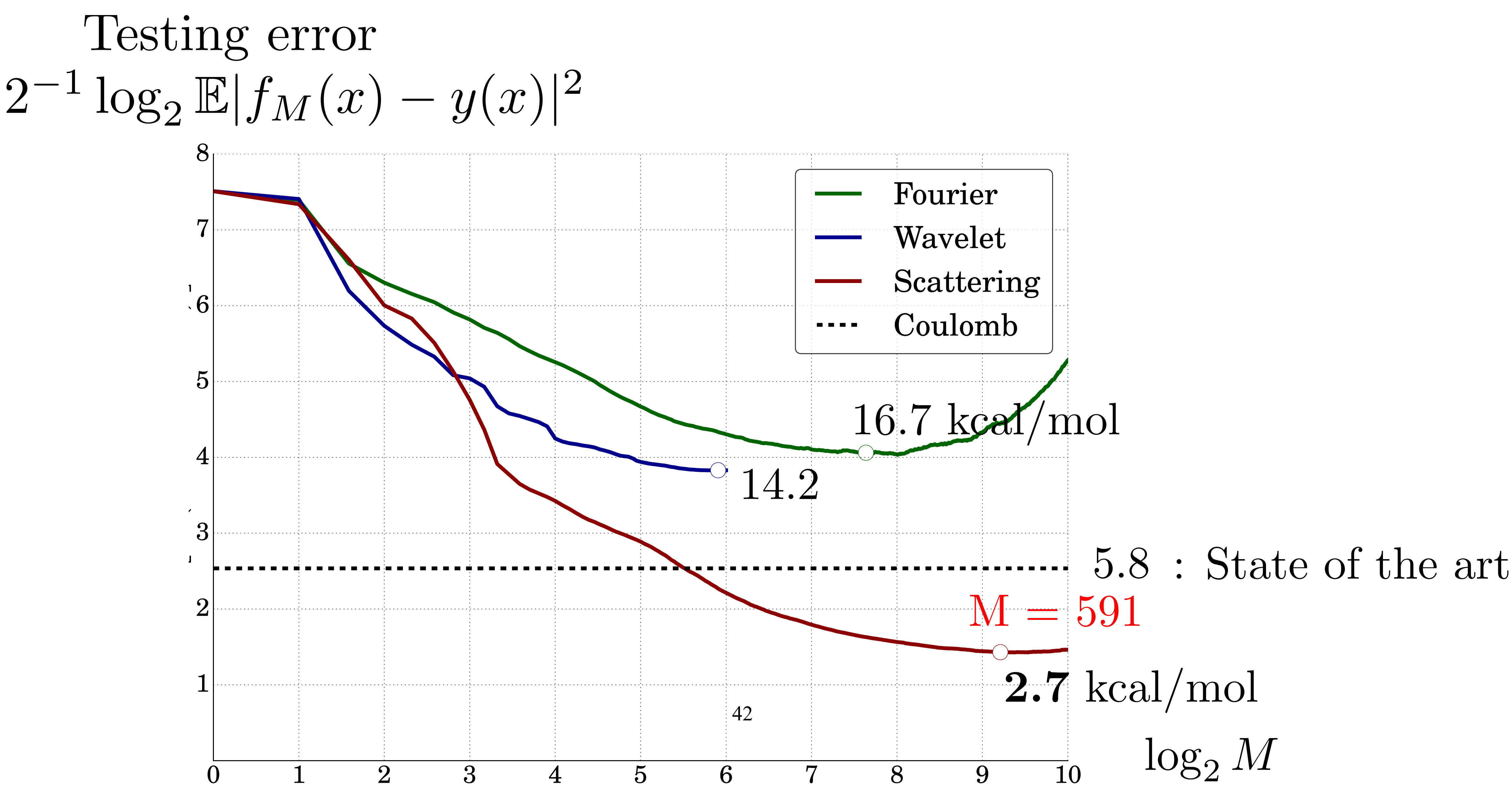
$$f_M(x) = \sum_{k=1}^{M} w_k \, \phi_{n_k}(\tilde{\rho}_x)$$

M: number of variables

# Scattering Regression

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

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



# Conclusion

- A major challenge of data analysis is to find Euclidean embeddings of metrics $\Leftrightarrow$ build Gaussian models

- Continuity to action of diffeomorphisms $\Rightarrow$ wavelets

- Known geometry $\Rightarrow$ no need to learn.
  Unknown geometry: learn wavelets on appropriate groups.

- Can learn physics from prior on geometry and invariants.

- Applications to images, audio and natural languages

www.di.ens.fr/data/scattering