High Dimensional Learning

From Images to Quantum Chemistry

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

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

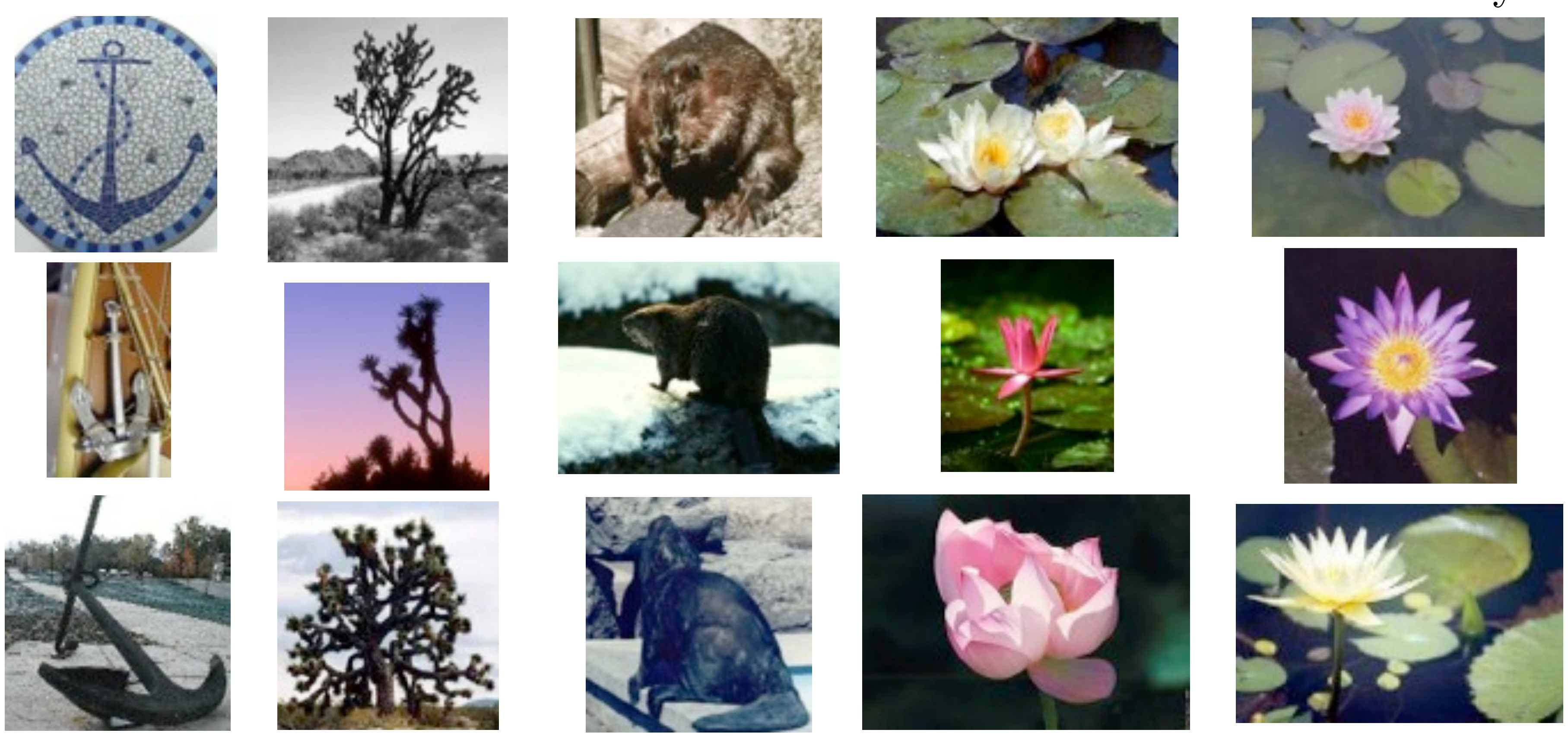
High Dimensional Learning

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

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

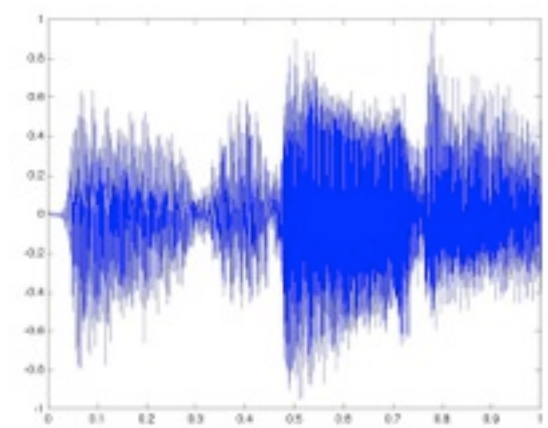
Image Classification $d = 10^6$

Anchor | Joshua Tree | Beaver | Lotus | Water Lily



Sons

$d = 10^4/s$



High Dimensional Learning

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

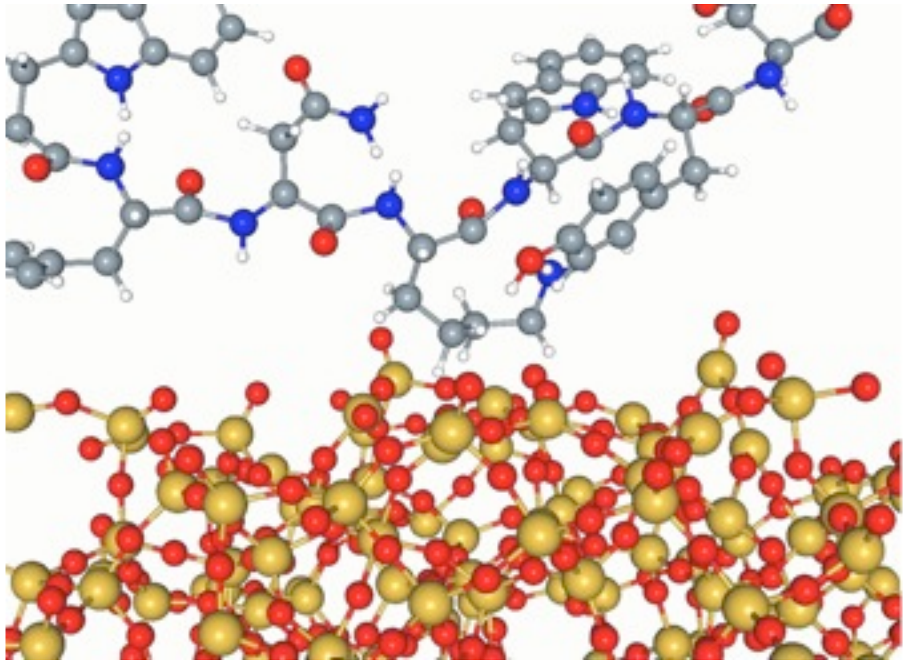
- **Regression:** approximate a *functional* $f(x)$
  given n sample values $\{x_i \, , \, y_i = f(x_i) \in \mathbb{R}\}_{i \leq n}$

Physics: Many Body Problem

Interaction energy $f(x)$ of a system: $x = \Big\{ \text{positions, values} \Big\}$

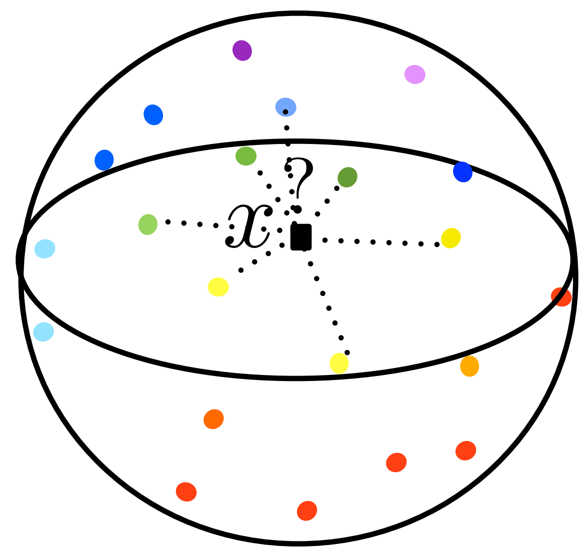
Astronomy

Quantum Chemistry



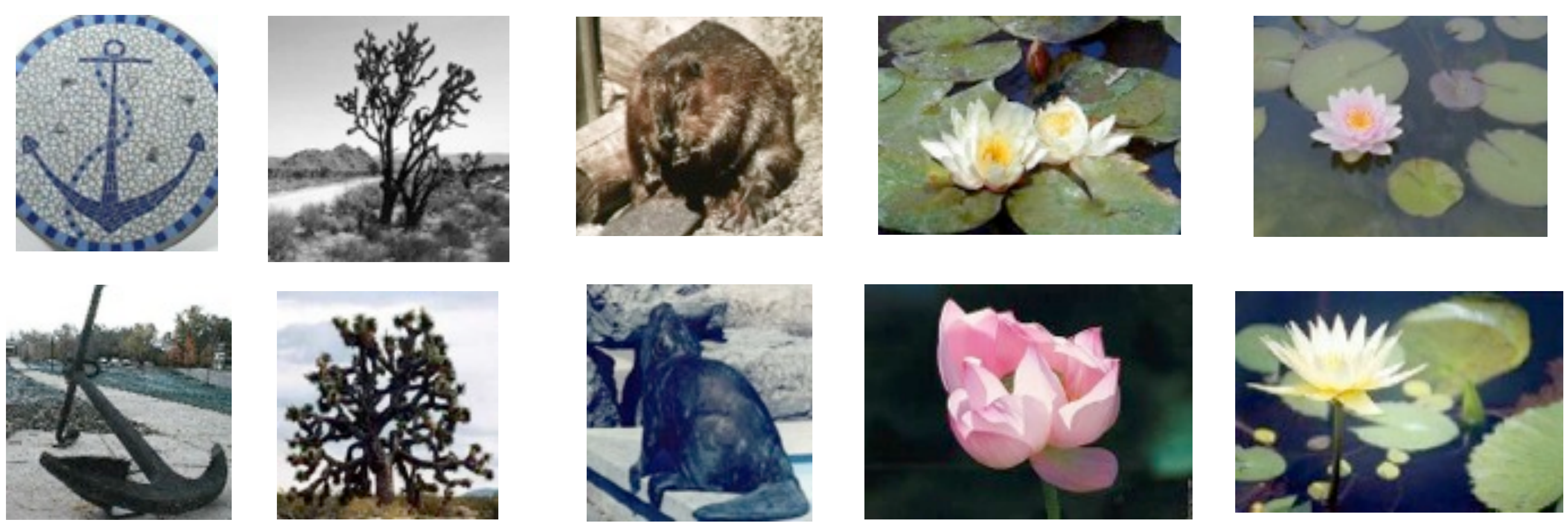
Curse of Dimensionality

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



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

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



Learning by Euclidean Embedding

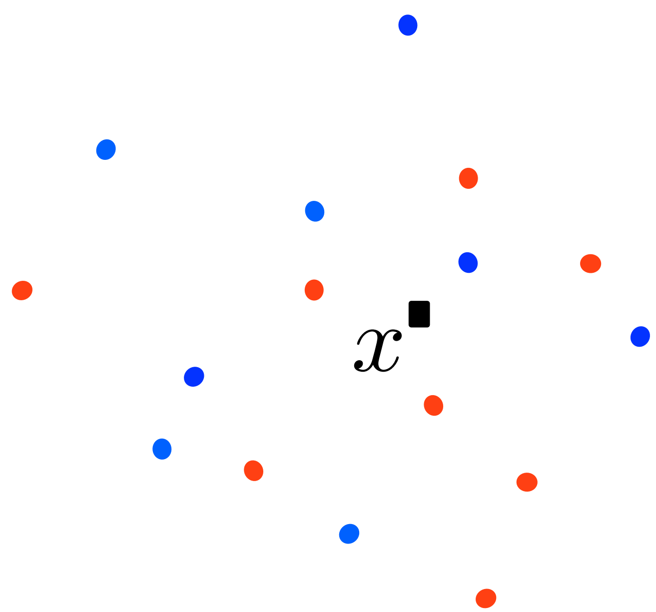
Data: $x \in \mathbb{R}^d$

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

Representation

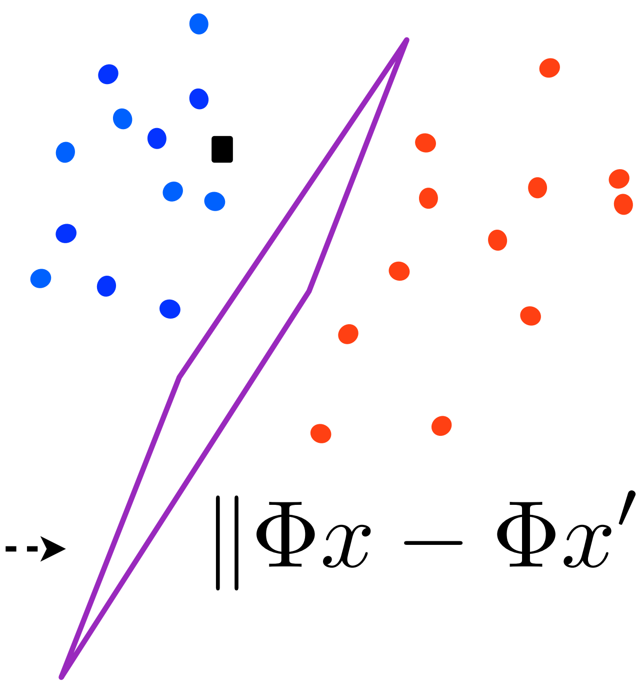
$\Phi x \in \mathcal{H}$

Linear Classifier



Intelligence

Φ



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

Equivalent Euclidean metric:

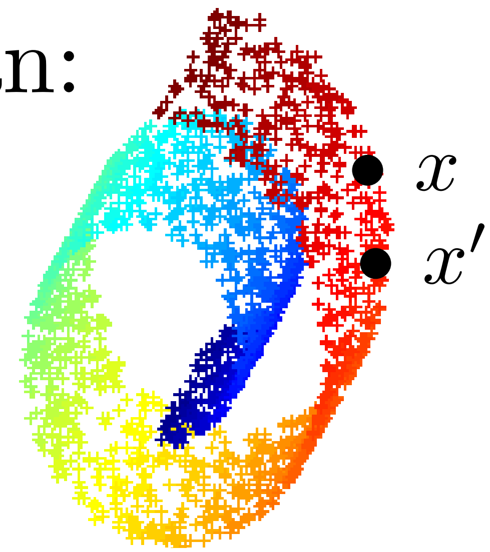
$$C_1 \|\Phi x - \Phi x'\| \leq \Delta(x, x') \leq C_2 \|\Phi x - \Phi x'\|$$

How to define Φ ?

Known Euclidean Embeddings

- If the data is in a low-dimensional manifold, no curse, the manifold metric is locally nearly Euclidean:

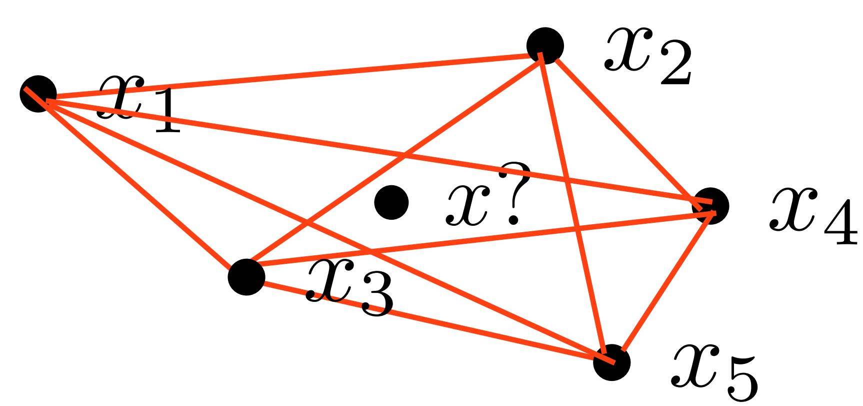
$$\langle \Phi(x), \Phi(x') \rangle = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$$



- Embedding of Banach metrics over finite set of points $\{x_i\}_i$

$\Leftrightarrow$ Euclidean embedding of graphs

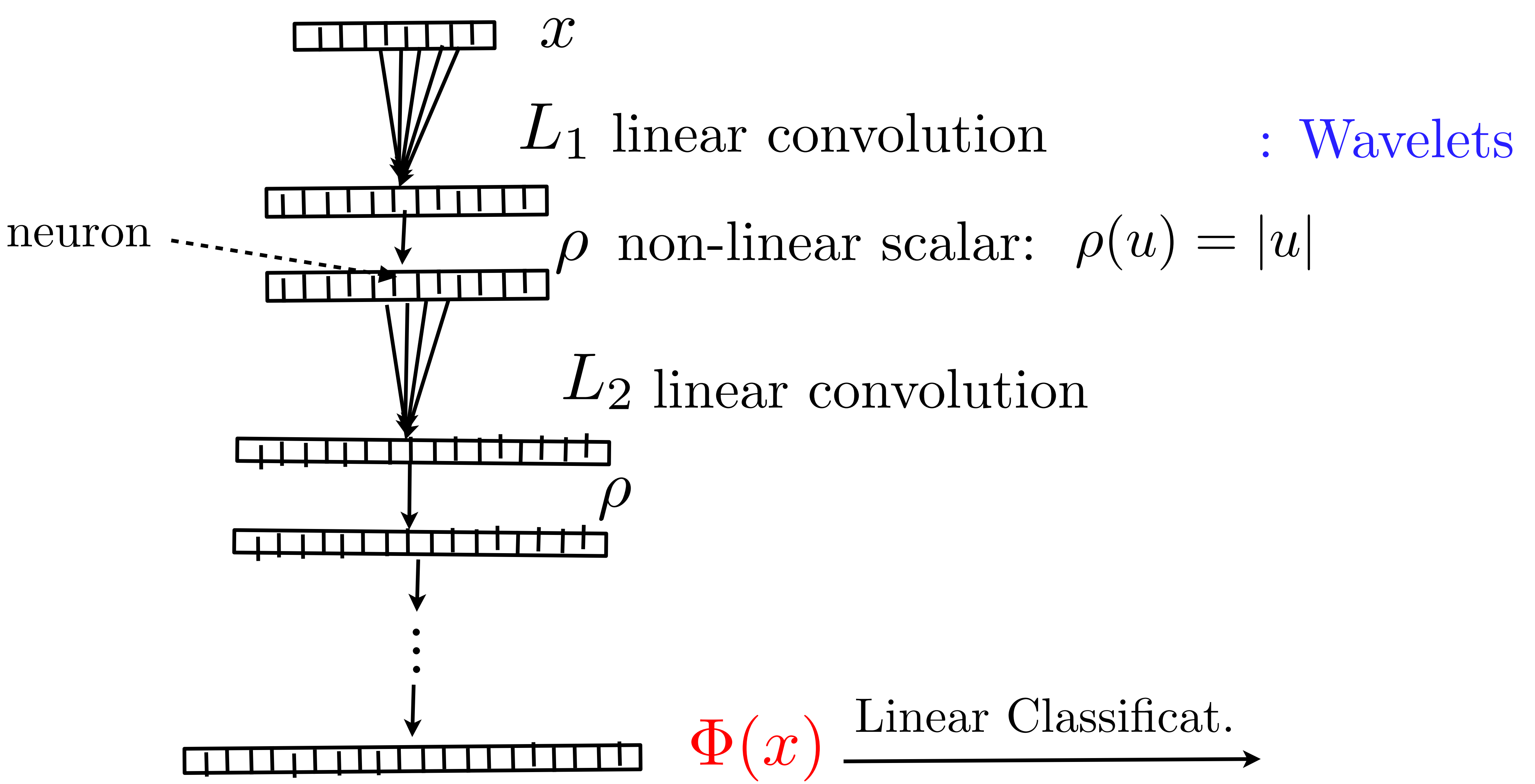
Bourgain/Johnsone Lindenstrass



- Generalisation for all x: need to embed the full space. Must impose a regularity condition on the metric.

Deep Convolution Neworks

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



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

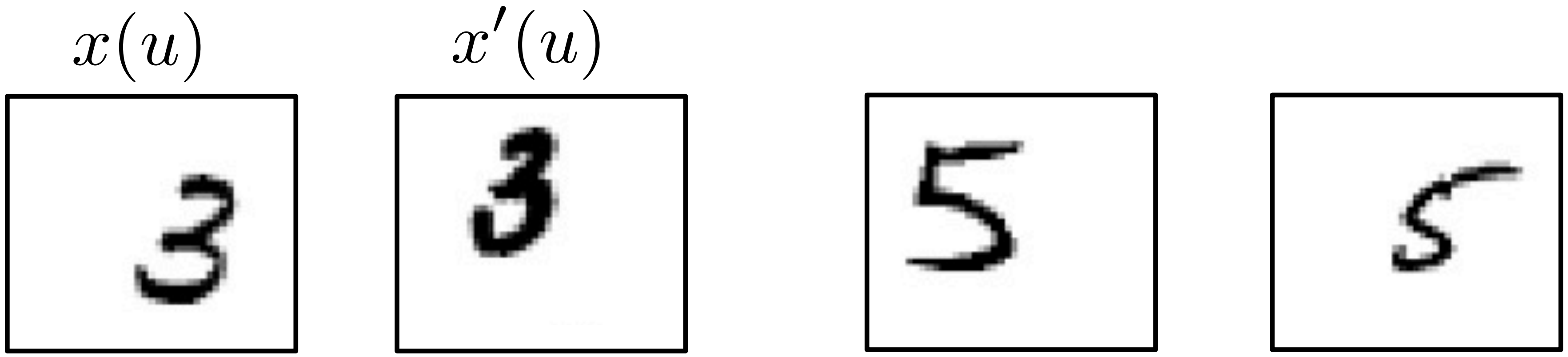
- Deep multiscale networks for embedding geometric metrics: invariance and continuity to diffeomorphisms

- Models of random processes and image classification

- Learning physics: quantum chemistry energy regression

Image Metrics

- Low-dimensional "geometric shapes"



Deformation metric: (classic mechanics) *Grenander*

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

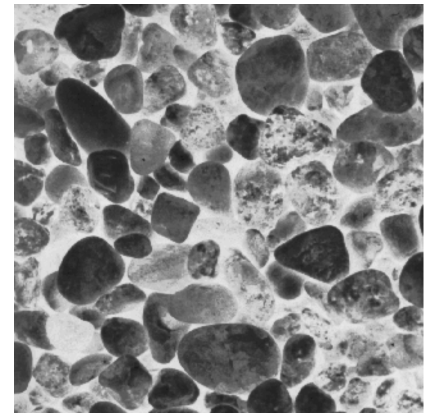
$$\Delta(x, x') \sim \min_\tau \|D_\tau x - x'\| + \|\nabla \tau\|_\infty \, \|x\|$$

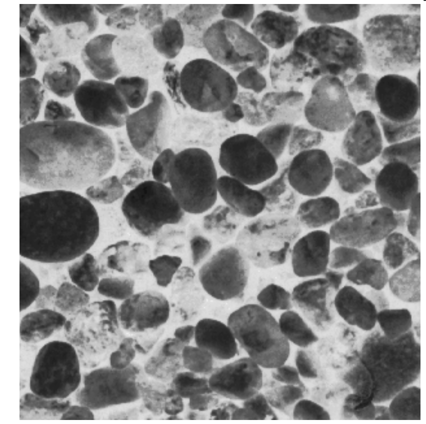
Invariant to translations

diffeomorphism amplitude

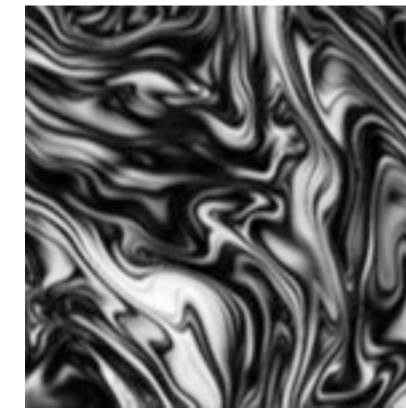
Image Metrics

- High dimensional textures: ergodic stationary processes





2D Turbulence



- What metric on stationary processes ? (statistical physics)
Bounded by a deformation metric:

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

x

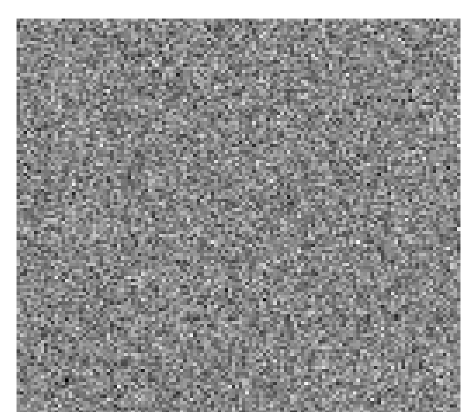
x'

But not equivalent: need that

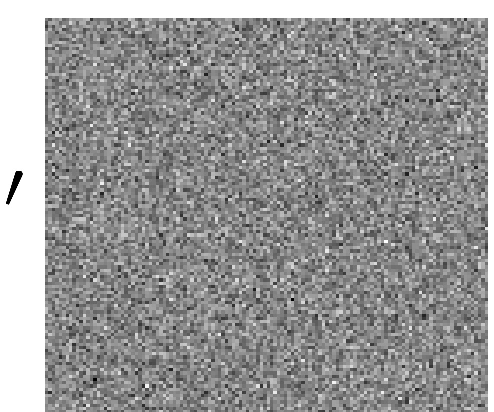
$\Delta(x', x) = 0$ if x and x'

are realisations of same process

x



x'



Euclidean Metric Embedding

- Embedding: find an equivalent Euclidean metric

$$\|\Phi x - \Phi x'\| \sim \Delta(x, x')$$

$$\text{with} \quad \Delta(x, x') \leq \min_{\tau} \|D_\tau x - x'\| + \|\nabla \tau\|_\infty \|x\|$$

- Equivalent conditions on Φ:

- **Continuous in $\mathbf{L^2}$**: $D_\tau = Id \;\Rightarrow \|\Phi x - \Phi x'\| \leq C \|x - x'\|$

- **Lipschitz continuous** to diffeomorphism actions:

$$x' = D_\tau x \;\Rightarrow\; \|\Phi x - \Phi D_\tau x\| \leq C \|\nabla \tau\|_\infty \|x\|$$

$\Rightarrow$ Invariance to translation

Fourier Deformation Instability

- Fourier transform $\hat{x}(\omega) = \int x(t)\, e^{-i\omega t}\, dt$

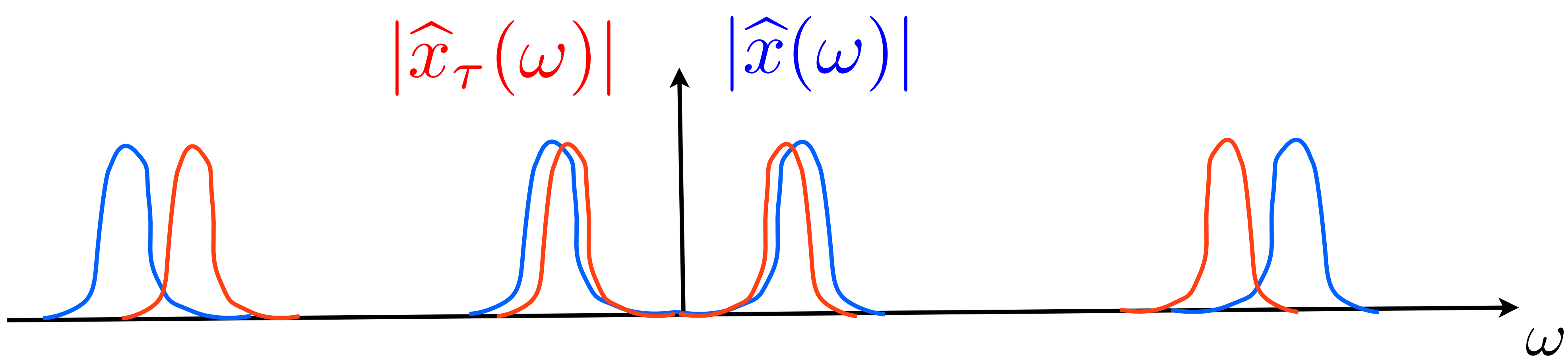
  The modulus is invariant to translations:

  $$x_c(t) = x(t-c) \quad \Rightarrow \quad \Phi(x) = |\hat{x}| = |\hat{x}_c|$$

- Continuous in $\mathbf{L^2}$: $\| |\hat{x}| - |\hat{x}'| \| \leq (2\pi)^{-1/2} \|x - x'\|$

- Instabilites to small deformations $x_\tau(t) = x(t - \tau(t))$:

  $\big|\, |\hat{x}_\tau(\omega)| - |\hat{x}(\omega)| \,\big|$ is big at high frequencies



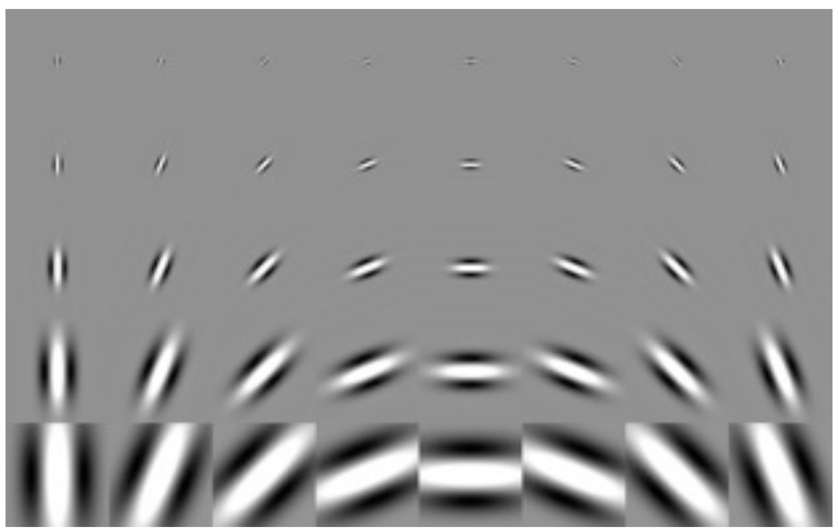
$$\Rightarrow \quad \| |\hat{x}| - |\hat{x}_\tau| \| \gg \|\nabla\tau\|_\infty \, \|x\|$$

Scale separation with Wavelets

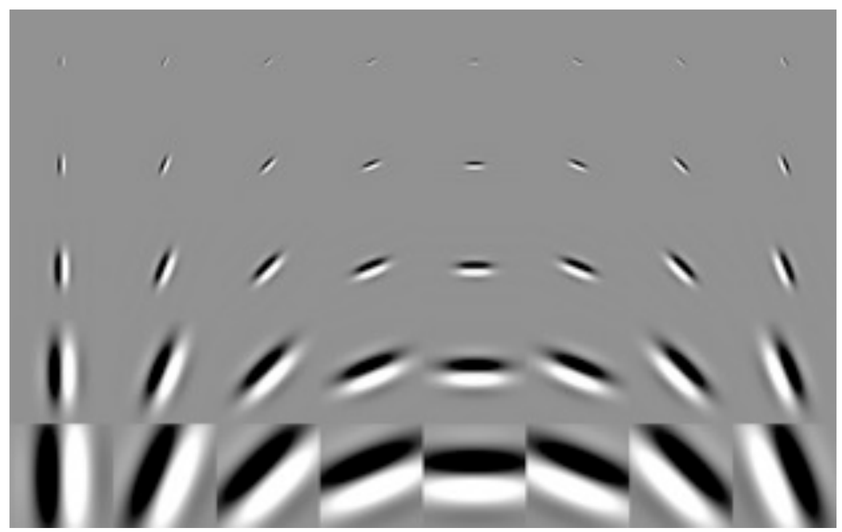
- Complex wavelet: $\psi(t) = g(t) \exp i\xi t \quad , \quad t = (t_1, t_2)$

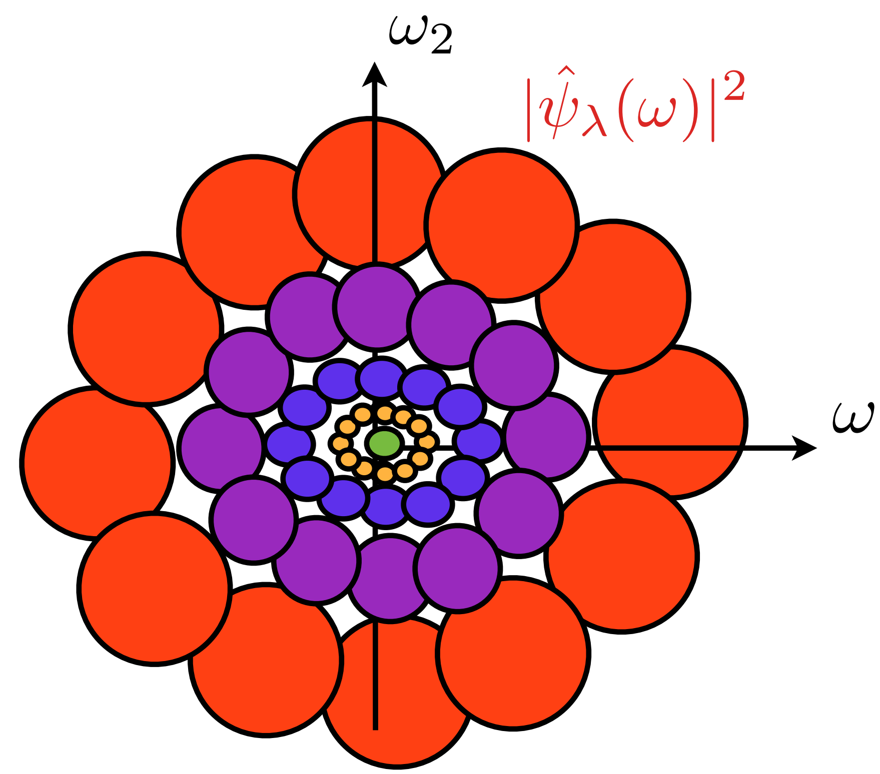
  rotated and dilated: $\psi_\lambda(t) = 2^{-j}\, \psi(2^{-j} r_\theta t)$ with $\lambda = (2^j, \theta)$

real parts



imaginary parts

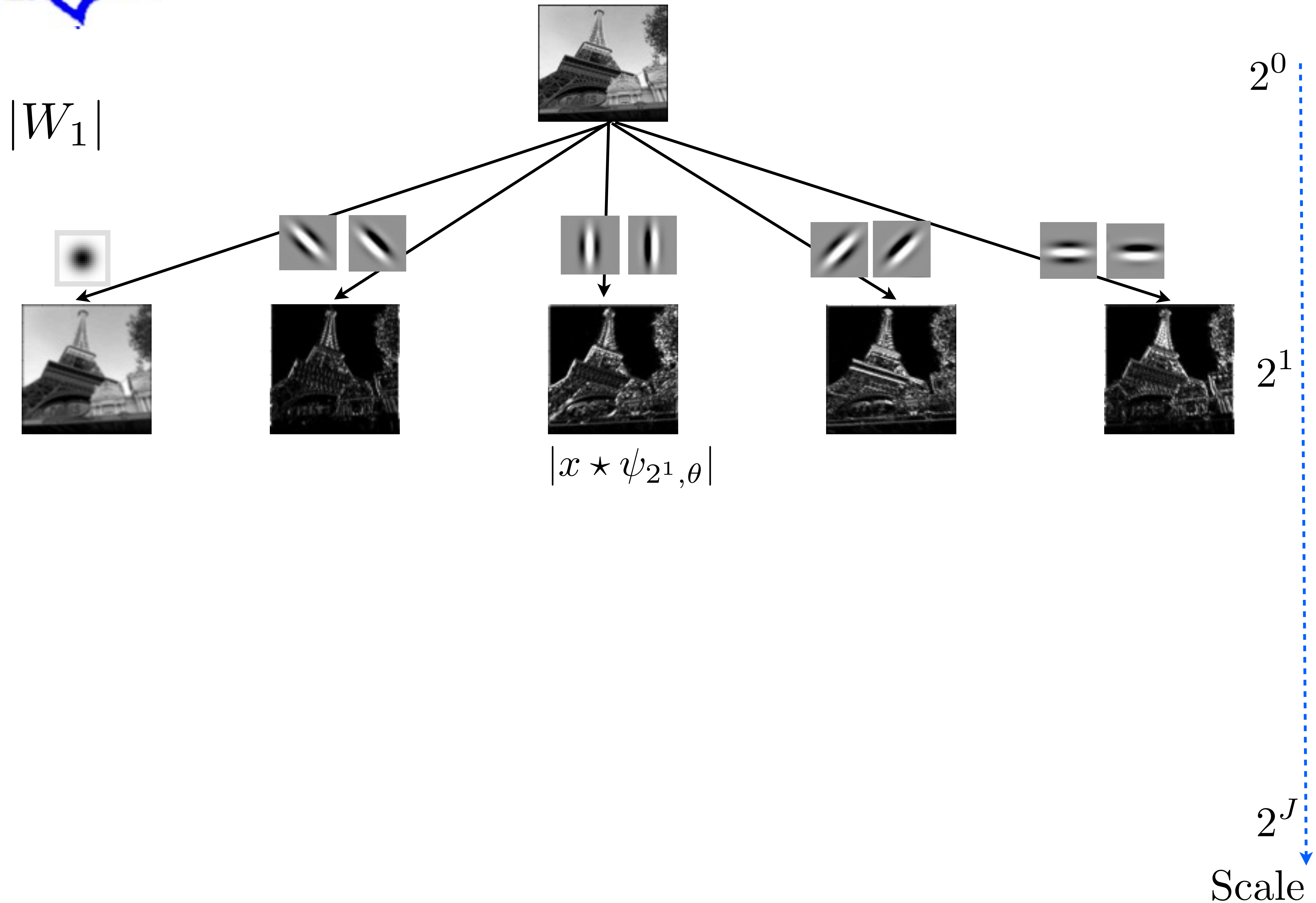




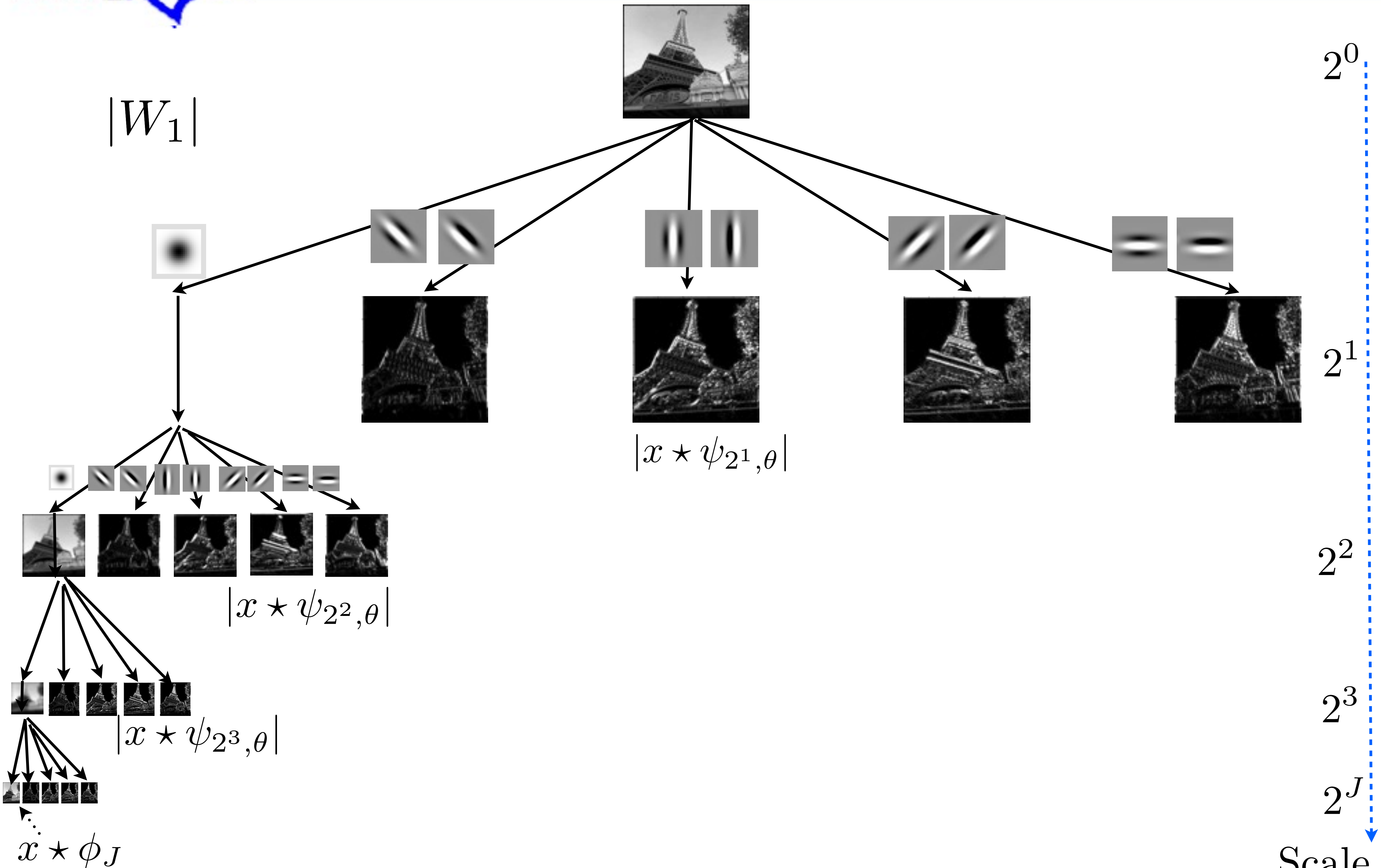
- Wavelet transform: $W x = \begin{pmatrix} x \star \phi_{2^J}(t) \\ x \star \psi_\lambda(t) \end{pmatrix}_{\lambda \leq 2^J}$

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

Fast Wavelet Transform



Wavelet Transform

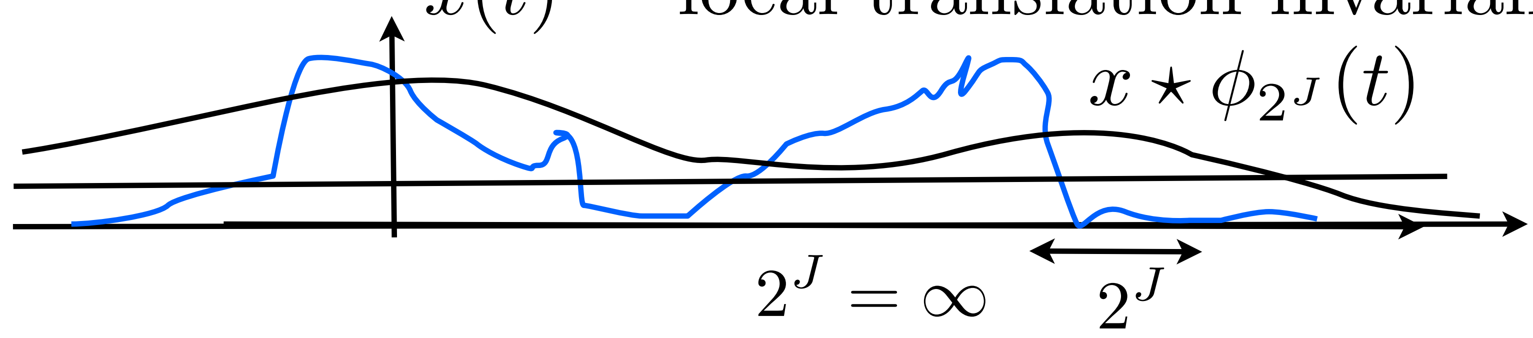


Wavelet Translation Invariance

First wavelet transform

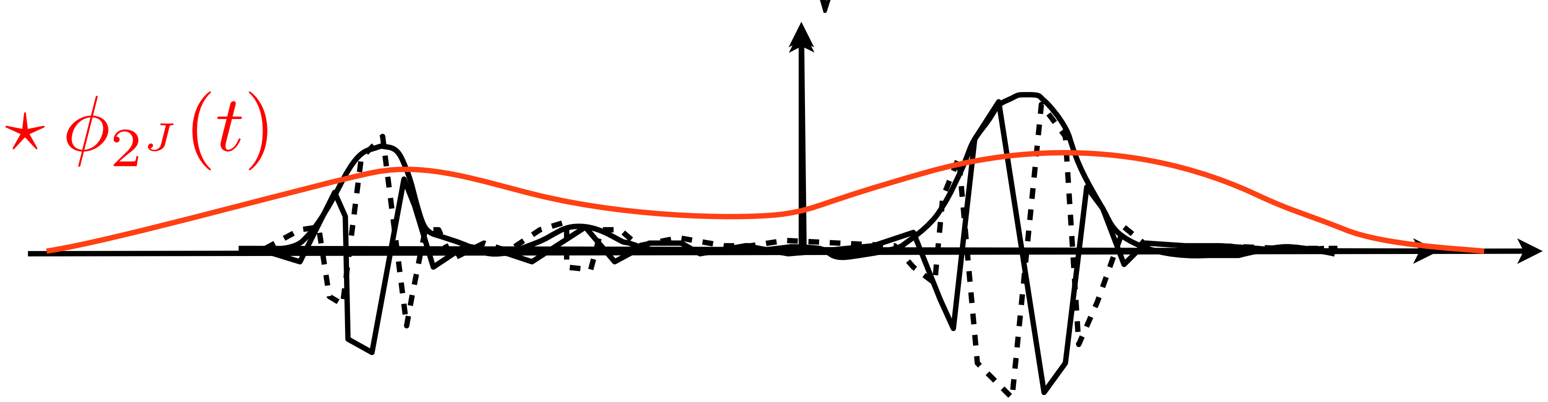
$$W_1 x = \begin{pmatrix} x \star \phi_{2^J} \\ x \star \psi_{\lambda_1} \end{pmatrix}_{\lambda_1}$$

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



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

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

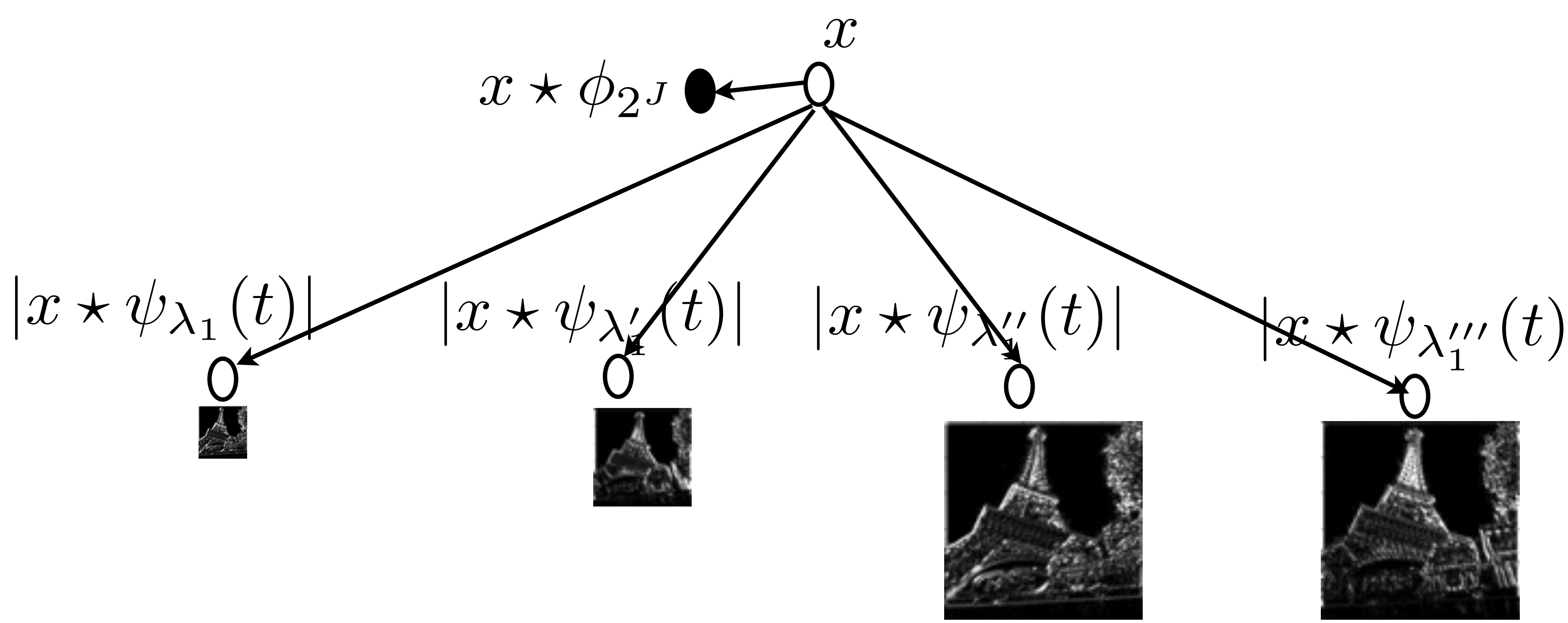


Second wavelet transform modulus

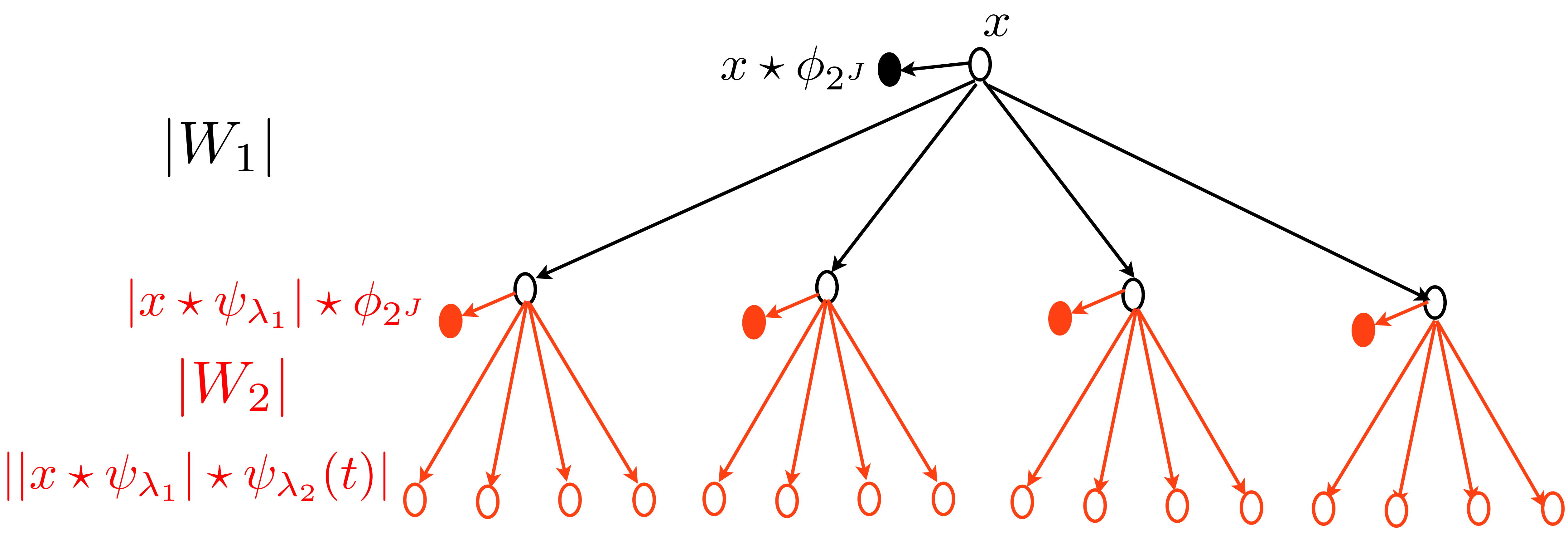
$$|W_2|\ |x \star \psi_{\lambda_1}| = \begin{pmatrix} |x \star \psi_{\lambda_1}| \star \phi_{2^J}(t) \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}(t)| \end{pmatrix}_{\lambda_2}$$

Scattering Transform

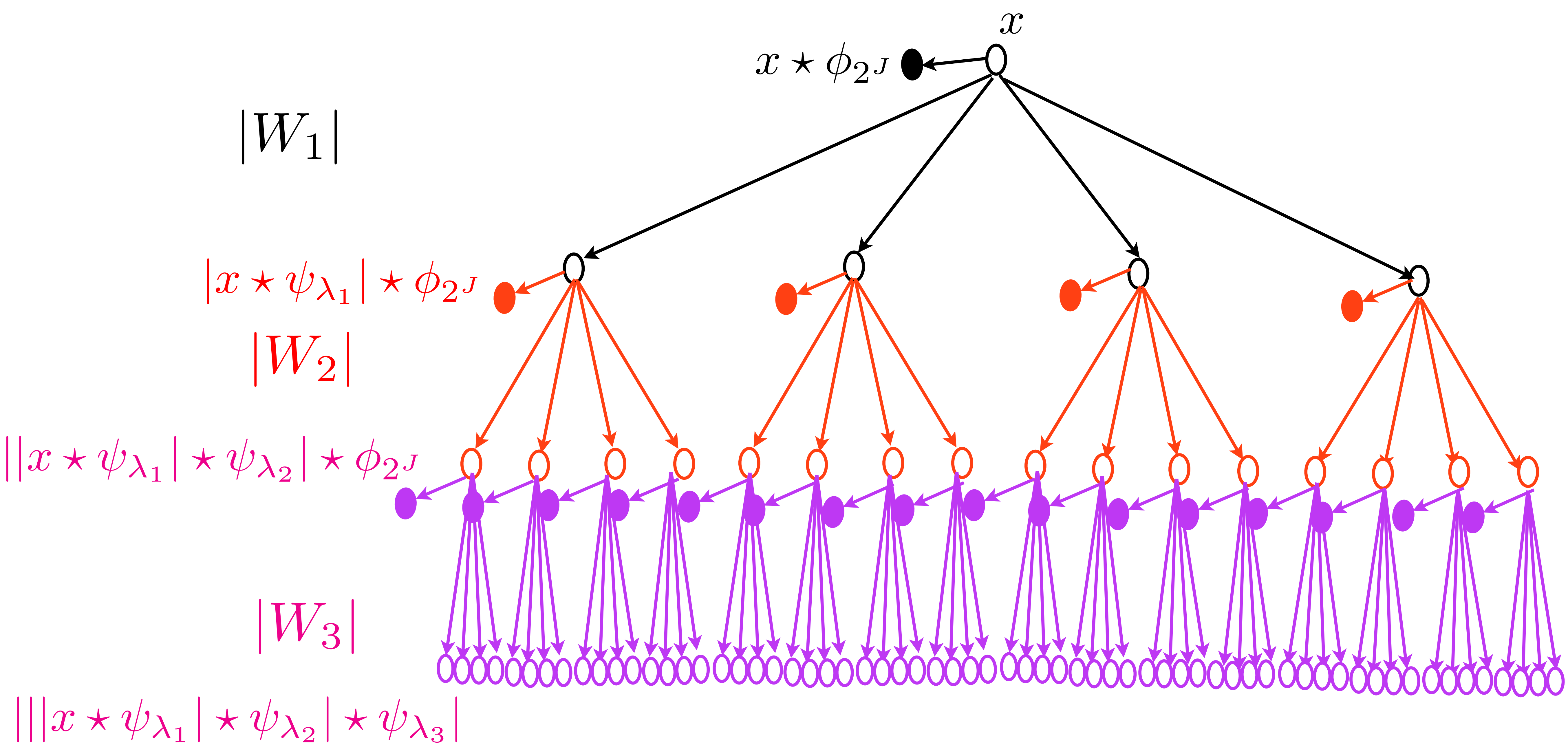
$|W_1|$



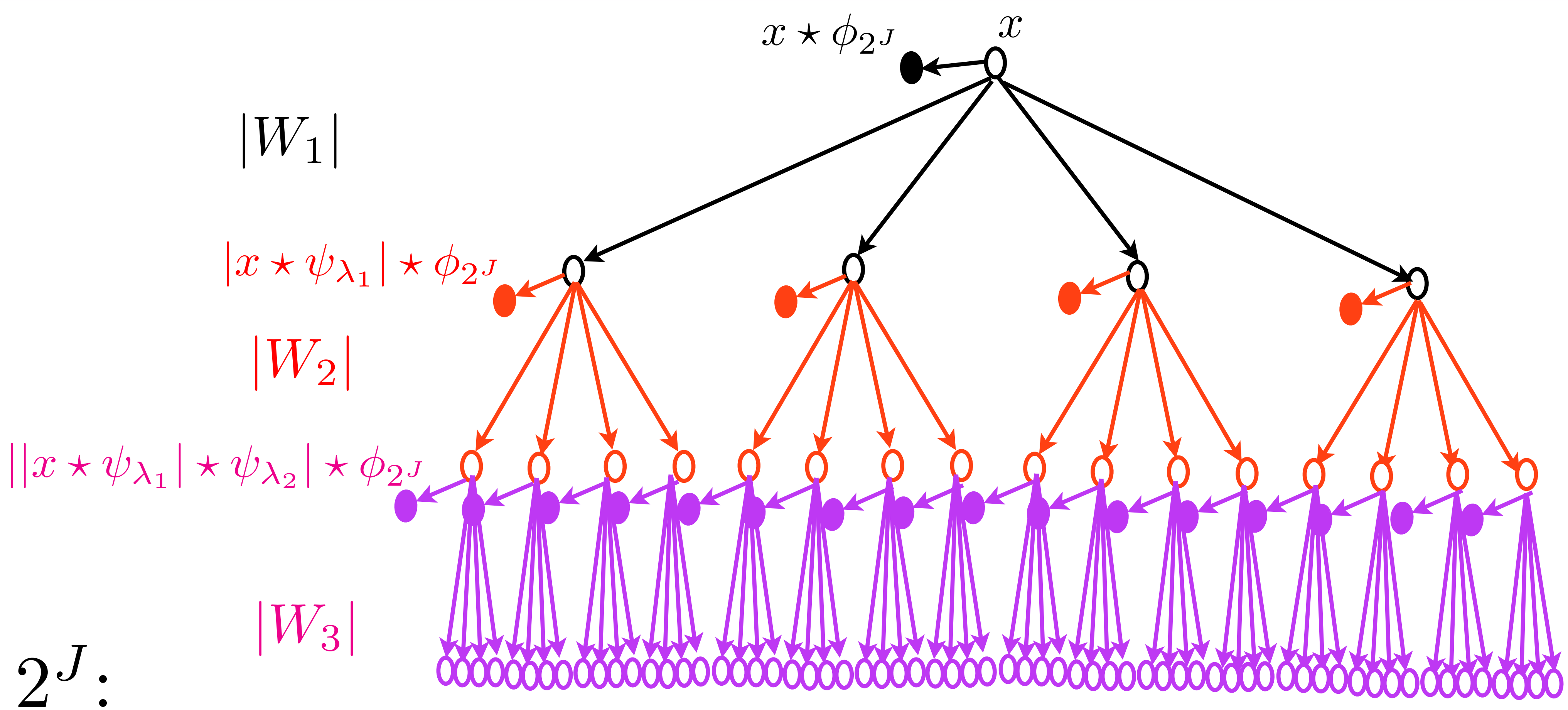
Scattering Transform



Scattering Neural Network



Wavelet Scattering



Scattering at 2^J:

$$S_J x(\lambda_1, ..., \lambda_m) = ||x \star \psi_{\lambda_1}| \star ...| \star \psi_{\lambda_m}| \star \phi_{2^J}$$

path variable

$$x \in \mathbf{L^1} \;\Rightarrow\; \lim_{J \to \infty} S_J x(\lambda_1, ..., \lambda_m) = \| ||x \star \psi_{\lambda_1}| \star ....| \star \psi_{\lambda_m}\|_1$$

Theorem: The energy of last layer coefficients converge to 0

$$\lim_{m \to \infty} \sum_{\lambda_1, ..., \lambda_m} \|S_J x(\lambda_1, ..., \lambda_m)\|^2 = 0$$

Scattering Properties

$$S_J x = \begin{pmatrix} x \star \phi_{2^J} \\ |x \star \psi_{\lambda_1}| \star \phi_{2^J} \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_{2^J} \\ |||x \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \star \phi_{2^J} \\ \dots \end{pmatrix}_{\lambda_1, \lambda_2, \lambda_3, \dots} = \dots |W_3| \, |W_2| \, |W_1| \, x$$

$\|W_k x\| = \|x\| \Rightarrow \| |W_k x| - |W_k x'| \| \leq \|x - x'\|$

Lemma: $\|[W_k, D_\tau]\| = \|W_k D_\tau - D_\tau W_k\| \leq C \, \|\nabla \tau\|_\infty$

Theorem: *For appropriate wavelets, a scattering is*

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

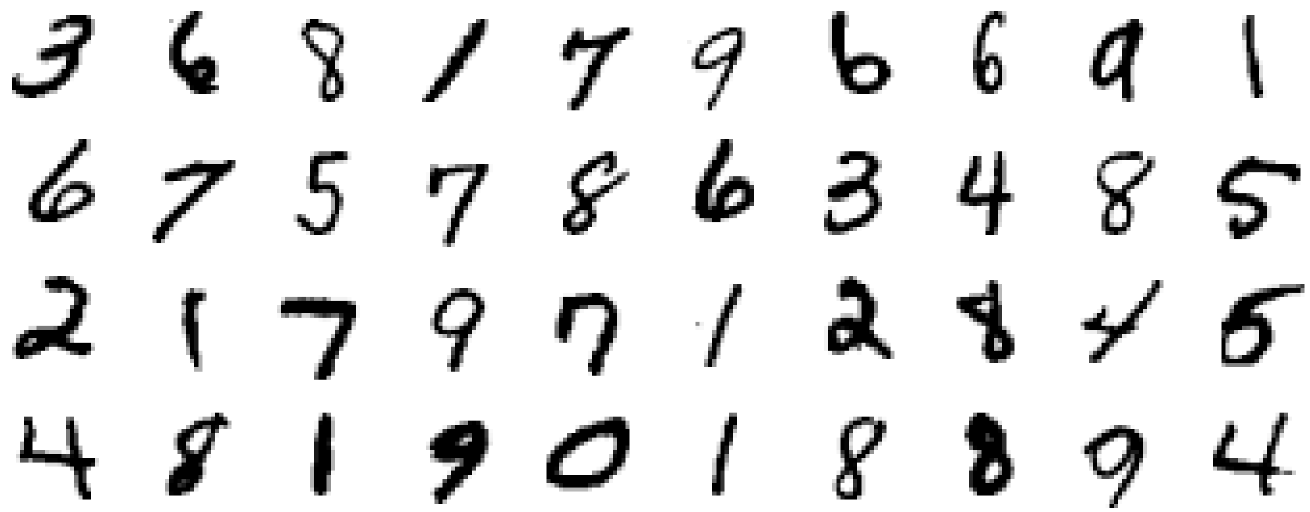
preserves norms $\|S_J x\| = \|x\|$

translations invariance and deformation stability:

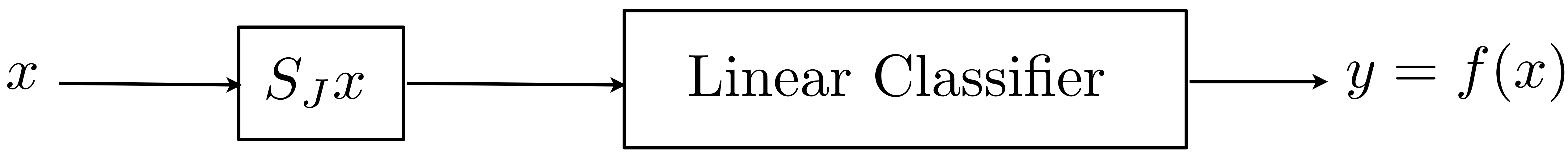
if $D_\tau x(u) = x(u - \tau(u))$ *then*

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

Digit Classification: MNIST



Joan Bruna



Classification Errors

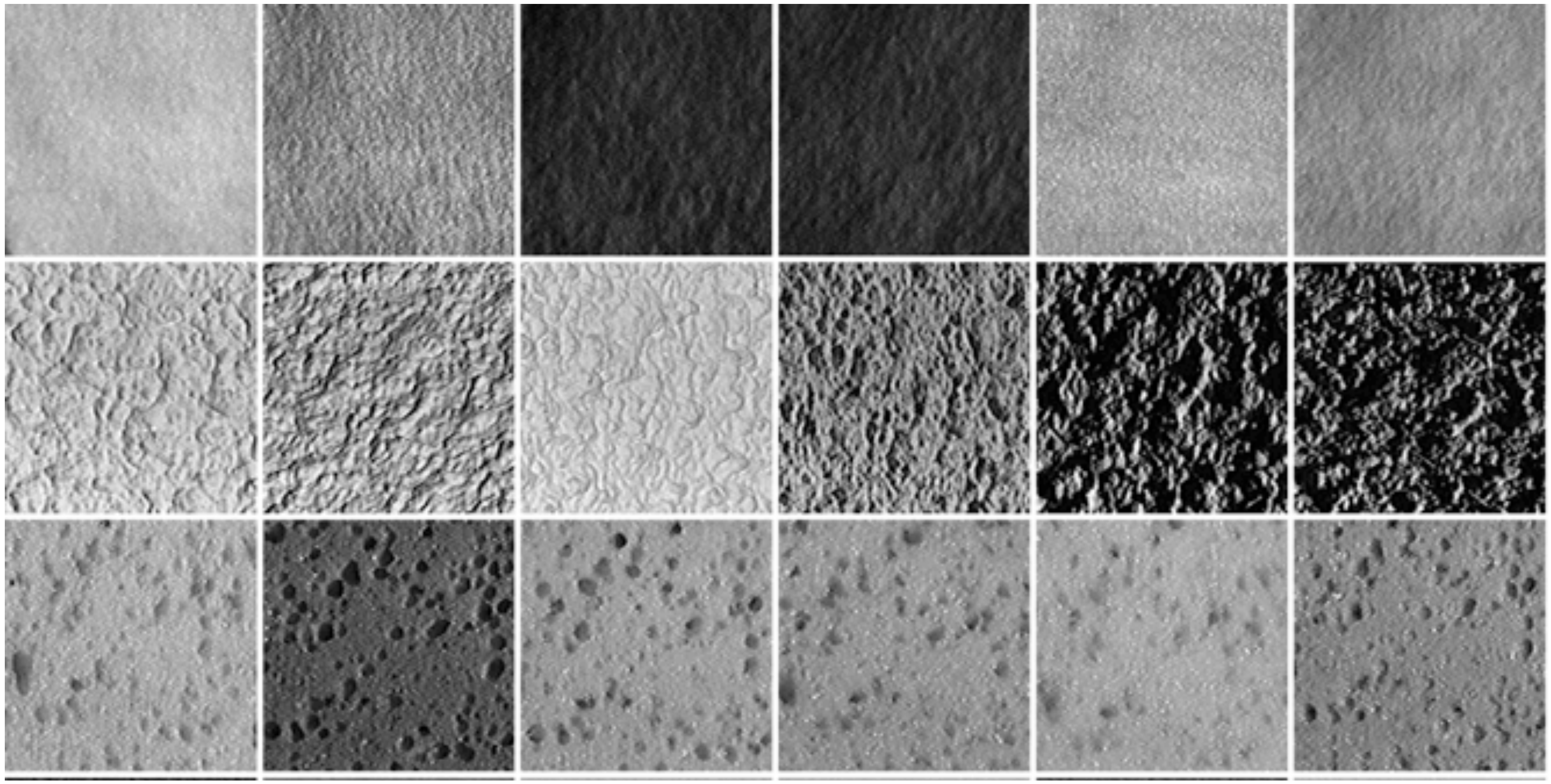
Training size	Conv. Net.	Scattering
300	7.2%	**4.4%**
5000	1.5%	**1.0%**
50000	0.5%	**0.4%**

LeCun et. al.

Classification of Textures

J. Bruna

CUREt database
61 classes



Scattering Moments of Processes

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

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

converges to moments if X is ergodic when 2^J increases

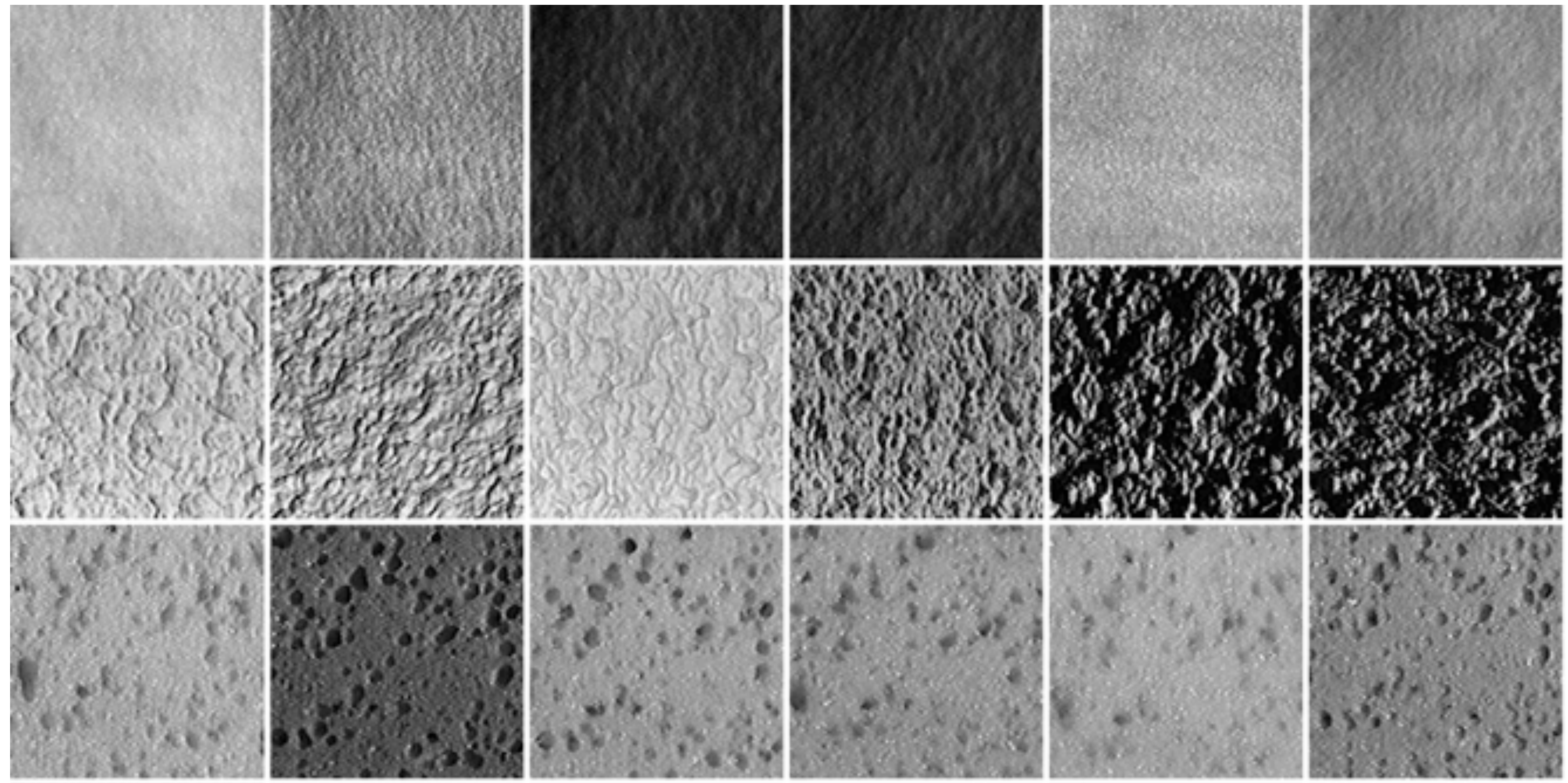
$$\mathbb{E}(SX) = \begin{pmatrix} \mathbb{E}(X) \\ \mathbb{E}(|X \star \psi_{\lambda_1}|) \\ \mathbb{E}(||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|) \\ \mathbb{E}(|||X \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}|) \\ ... \end{pmatrix}_{\lambda_1,\lambda_2,\lambda_3,...}$$

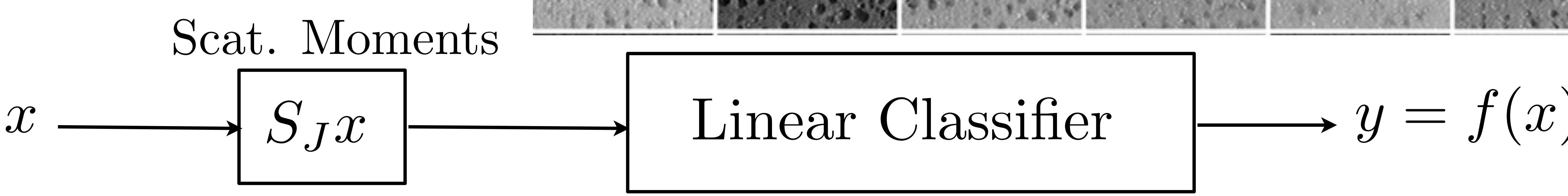
- Does $\mathbb{E}(SX)$ approximates well enough the distribution of X ?

Classification of Textures

J. Bruna

CUREt database
61 classes





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

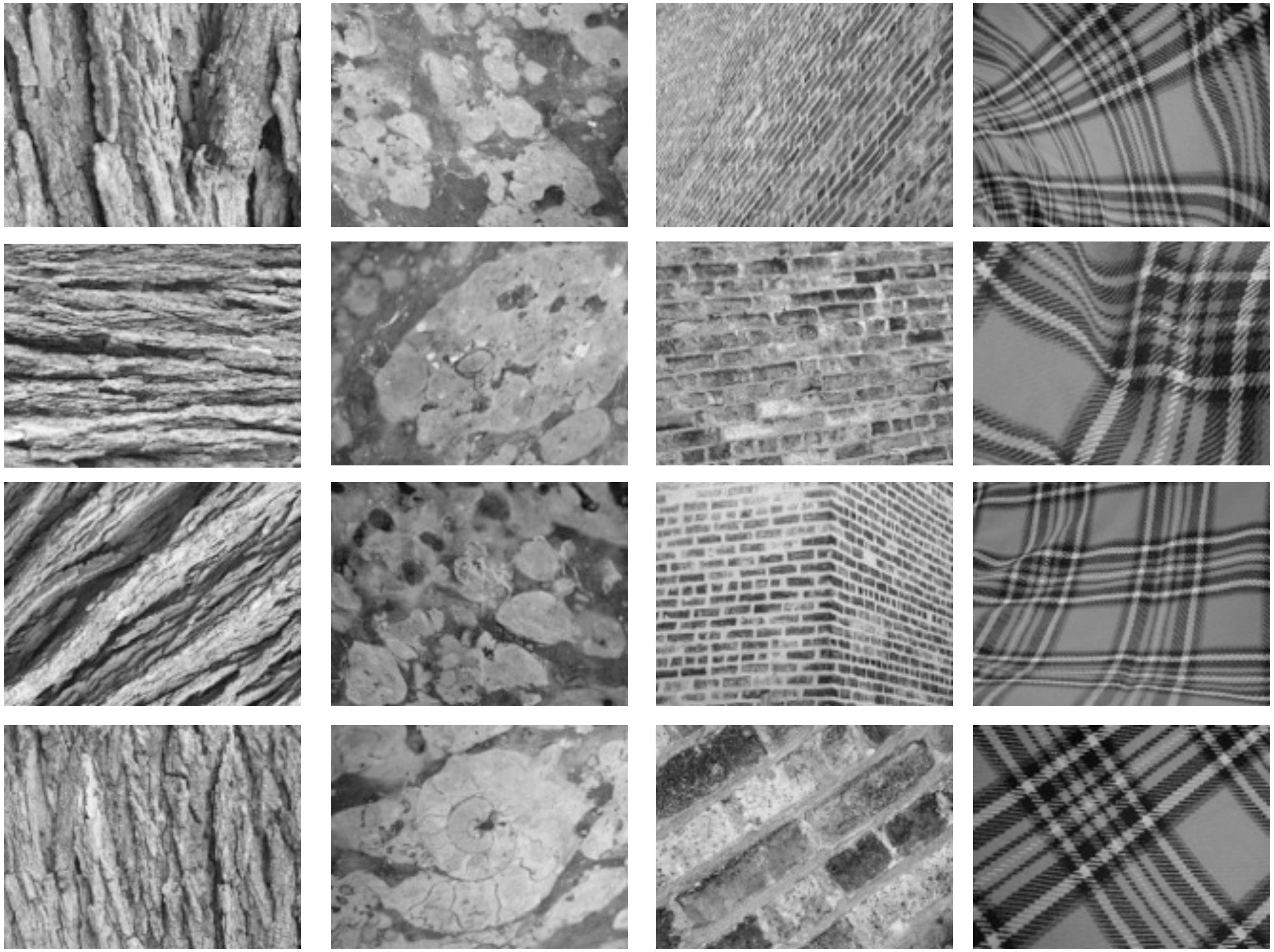
Training per class	Fourier Spectr.	Histogr. Features	Scattering
46	1%	1%	**0.2** %

- Can characterise non-Gaussian processes

Rotation and Scaling Invariance

Laurent Sifre

UIUC database:
25 classes



Scattering classification errors

Training	Scat. Translation
20	20 %

Extension to Rigid Mouvements

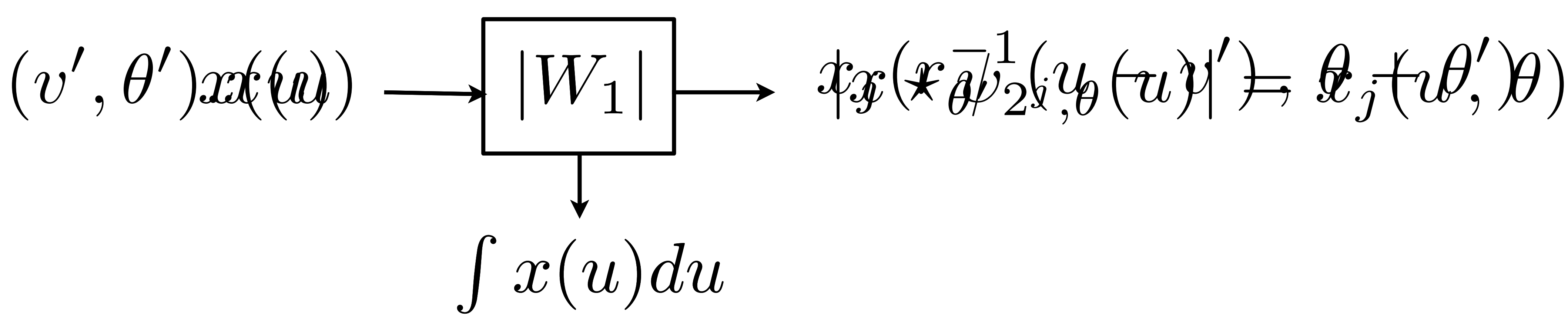
Laurent Sifre

- Special Euclidean group $G = \{(v, \theta) \in \mathbb{R}^2 \times [0, 2\pi)\}$

  action on an image: $(v, \theta) \,.\, x(u) = x(r_\theta^{-1}(u - v))$

$$(v, \theta)^{-1} = (-r_{-\theta} v, -\theta)$$

- Action on wavelet coefficients:



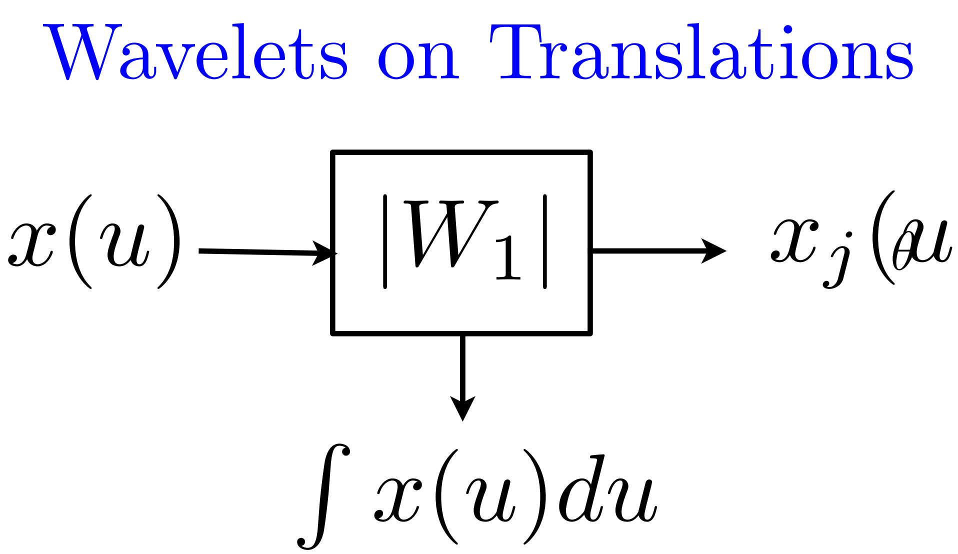
Extension to Rigid Mouvements

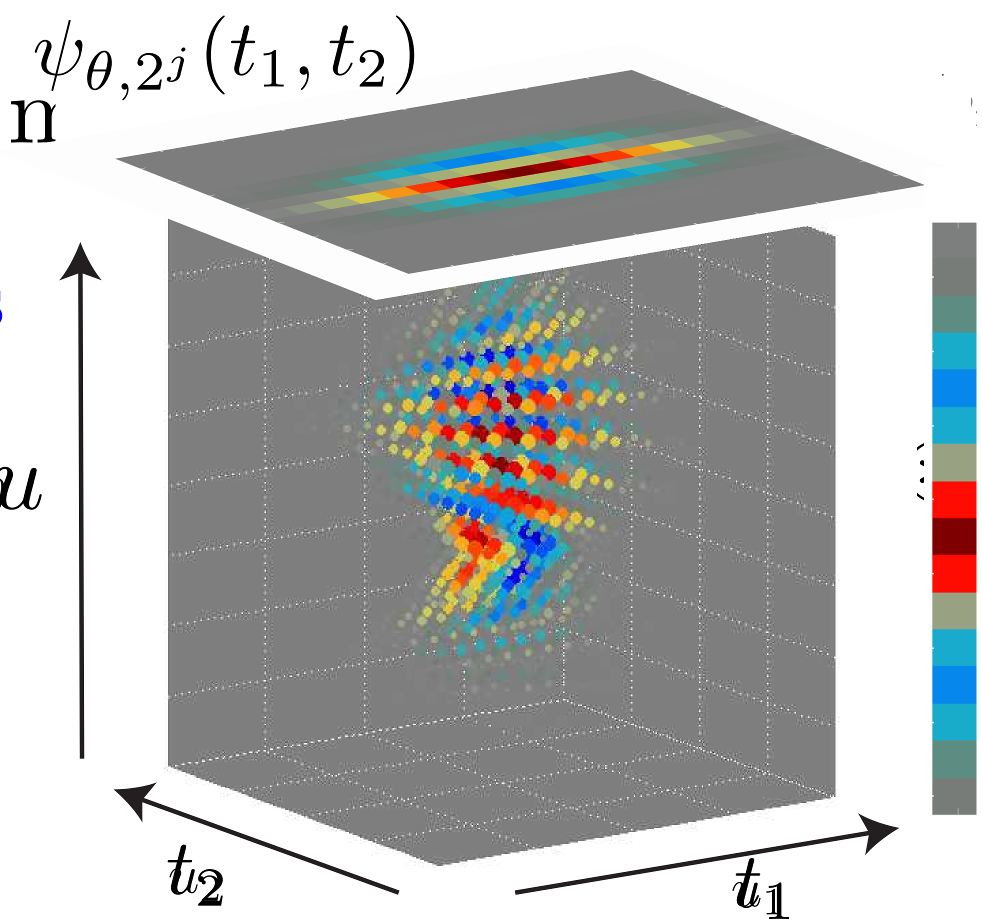
Laurent Sifre

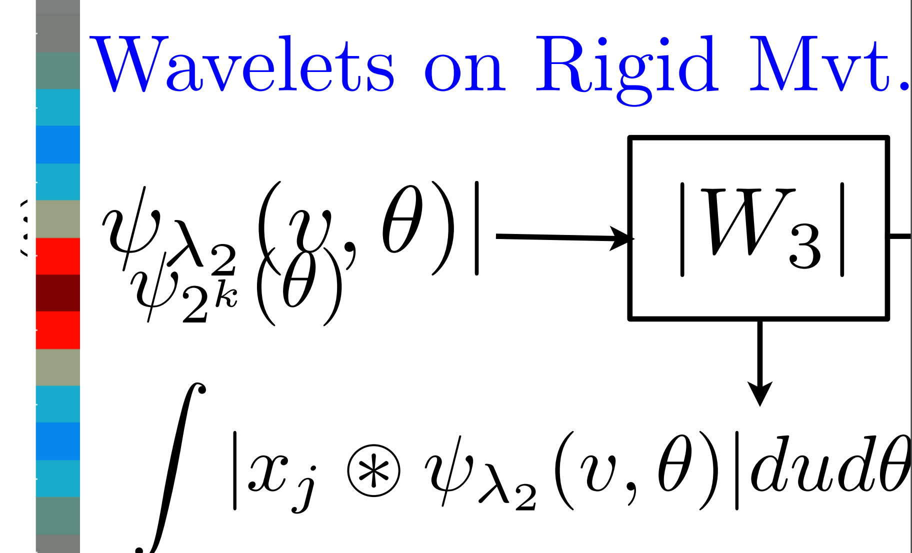
- To build invariants: second wavelet transform on $\mathbf{L^2}(G)$: group convolutions of $x_j(u,\theta)$ with wavelets $\psi_{\lambda_2}(u,\theta)$

$$x_j \circledast \psi_{\lambda_2}(u,\theta) = \int_{\mathbb{R}^2}\int_0^{2\pi} x_j(v',\theta')\, \psi_{\lambda_2}\Big((v',\theta')^{-1}(u,\theta)\Big)\, dv'\, d\theta'$$

- Scattering on rigid m



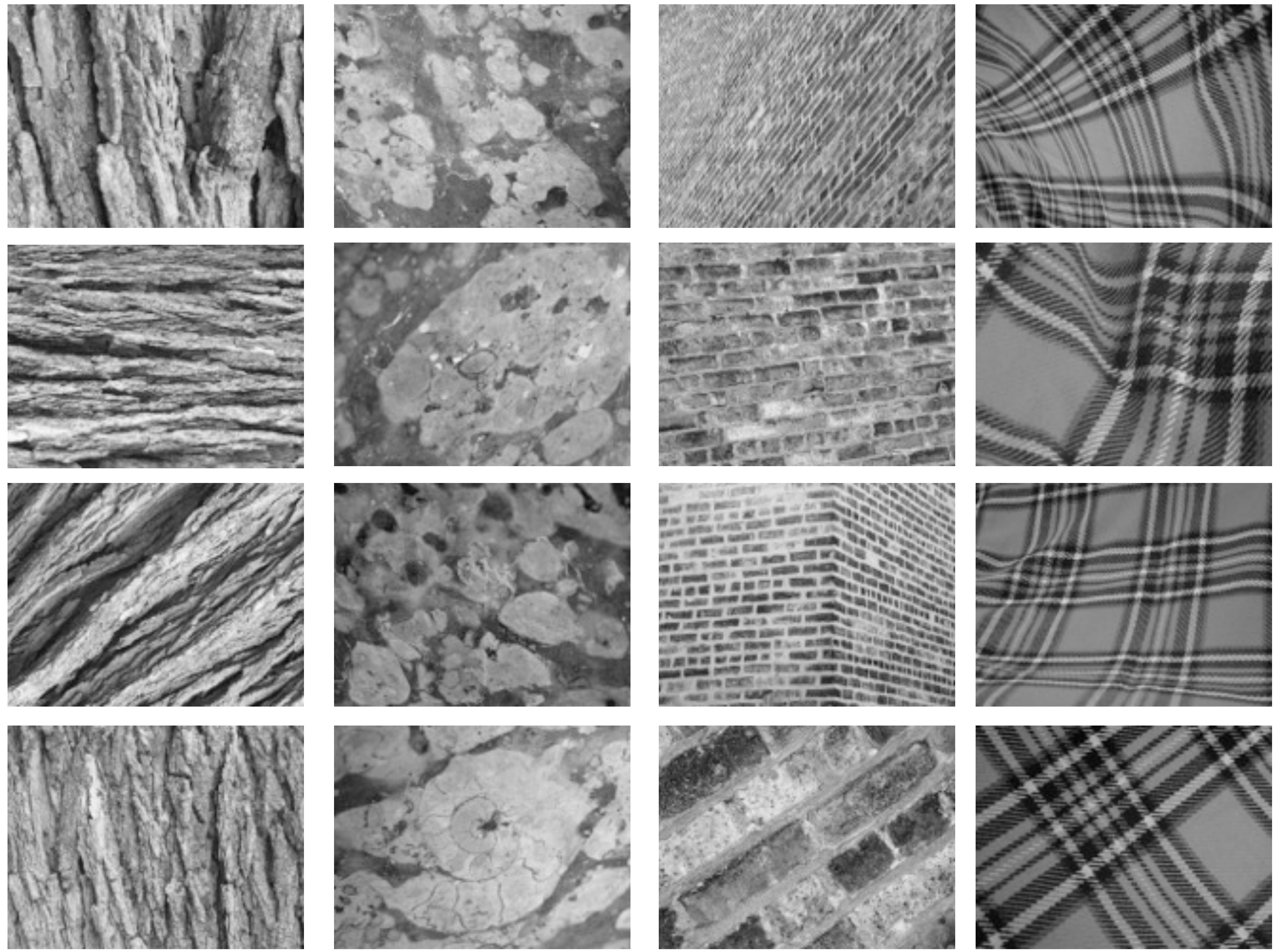




Rotation and Scaling Invariance

Laurent Sifre

UIUC database:
25 classes



Scattering classification errors

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

Scattering Inversion: Phase Recovery

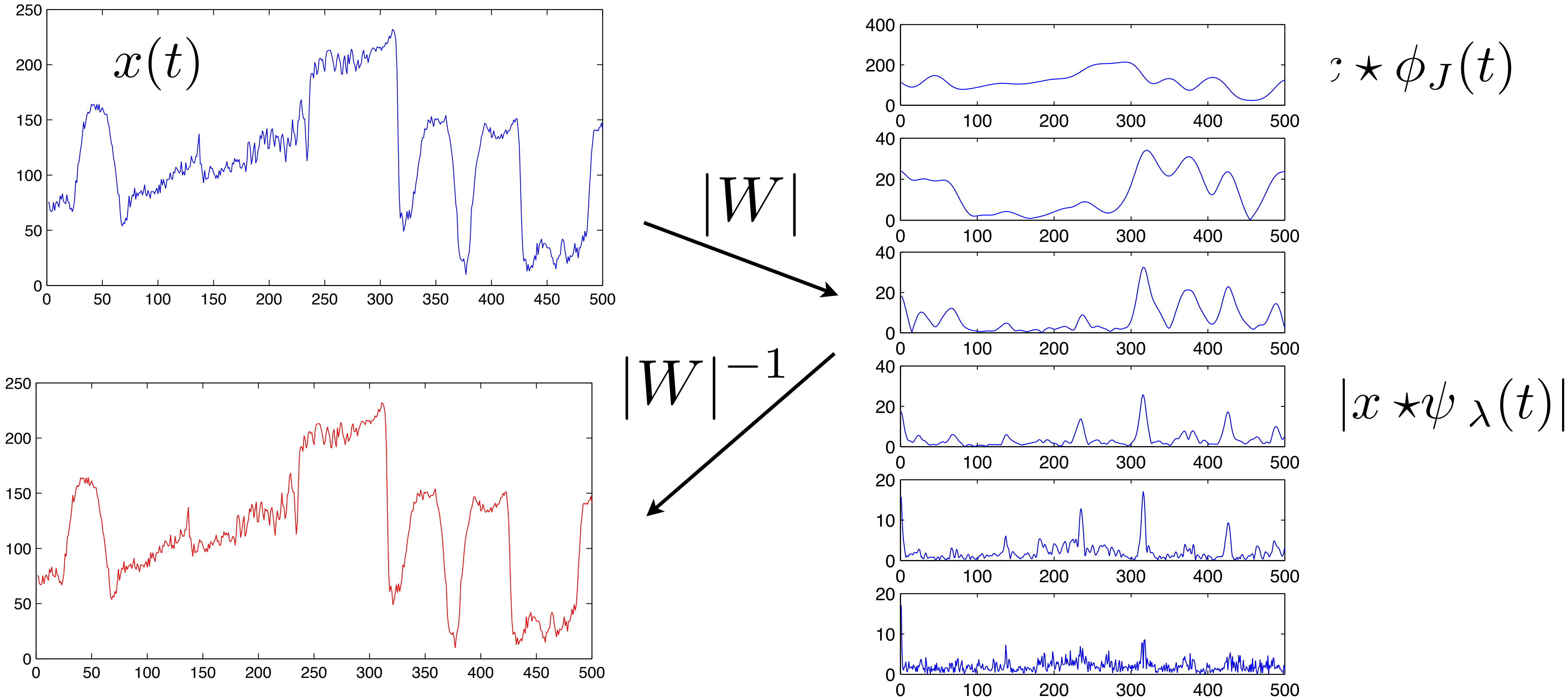
I. Waldspurger

Theorem For appropriate wavelets and any $J \leq \infty$

$$|W|x = \Big\{ x \star \phi_J \, , \, |x \star \psi_\lambda| \Big\}_\lambda$$

is invertible and the inverse is weakly continuous.

Proof: Complex analysis.



Inverse Scattering Transform

Joan Bruna

- Compute $\tilde{x}$ such that:

$$\forall k,\ \forall \lambda_1, ..., \lambda_k\ ,\ S_J\tilde{x}(\lambda_1, ..., \lambda_k) = S_J x(\lambda_1, ..., \lambda_k)$$

- If x is of period $N = 2^J$, at orders $k \leq m$ there are $O(\log_2^m N)$ scattering coefficients:

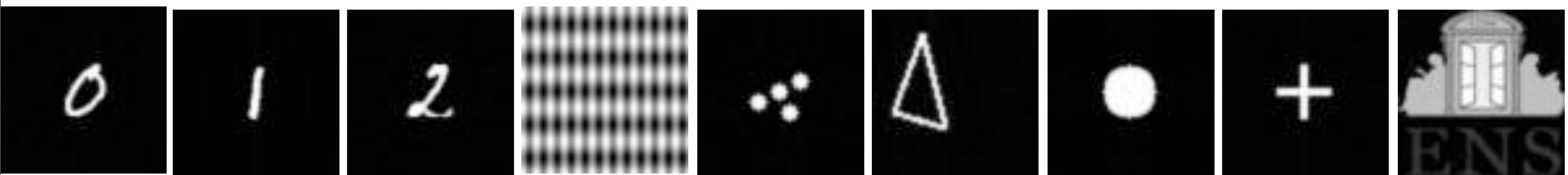
$$S_J x(\lambda_1, ..., \lambda_k) = \int_{[0,2^J]^2} ||x \star \psi_{\lambda_1}| \star ...| \star \psi_{\lambda_k}(u)|\, du$$

Compressed Shape Sensing

Joan Bruna

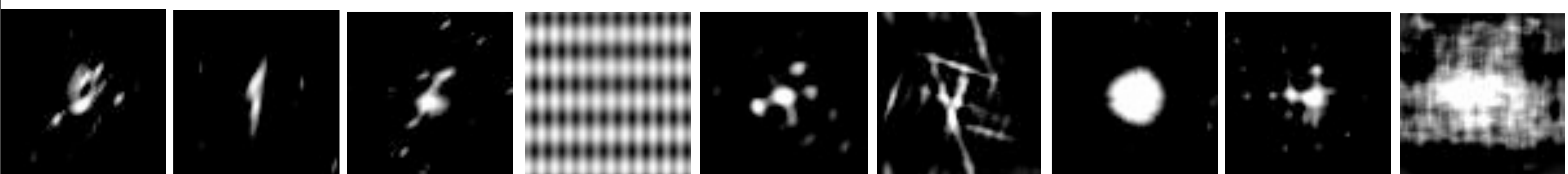
- Numerical recovery from 1st and 2nd order coefficients:

Original images of N^2 pixels:



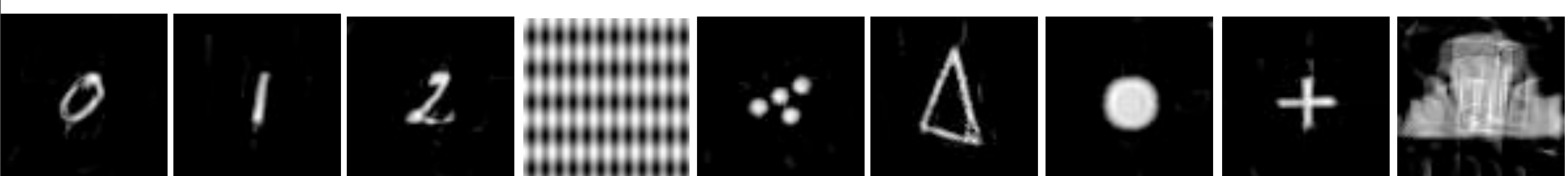
For $2^J = N$, $m = 1$

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



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.

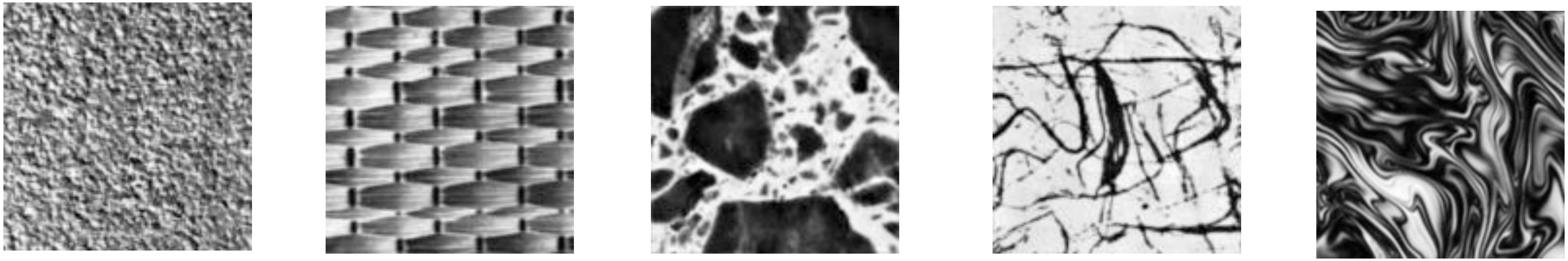


Ergodic Texture Reconstructions

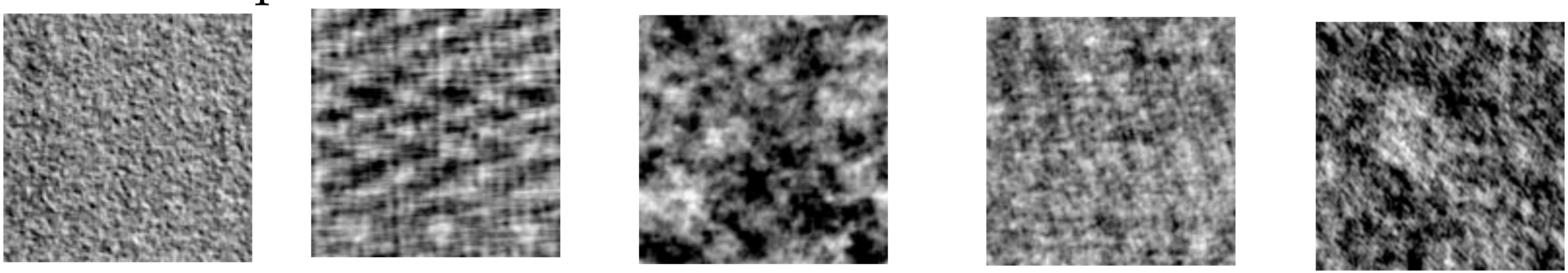
Joan Bruna

Original Textures

2D Turbulence

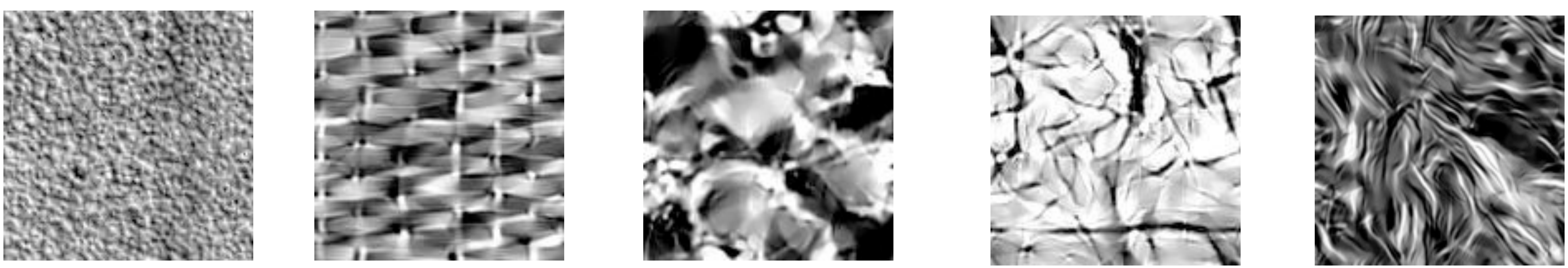


Gaussian process model with same second order moments



For $2^J = N$: $O(\log N^2)$ scattering moments:

$$\|x \star \psi_{\lambda_1}\|_1 \approx \mathbb{E}(|x \star \psi_{\lambda_1}|) \quad , \quad \||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1 \approx \mathbb{E}(||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|)$$



Representation of Audio Textures

Joan Bruna

- $x \in \mathbb{R}^d$ realization of a stationary process

	Original	Gaussian model	Scattering
Water			
Paper			
Cocktail Party			

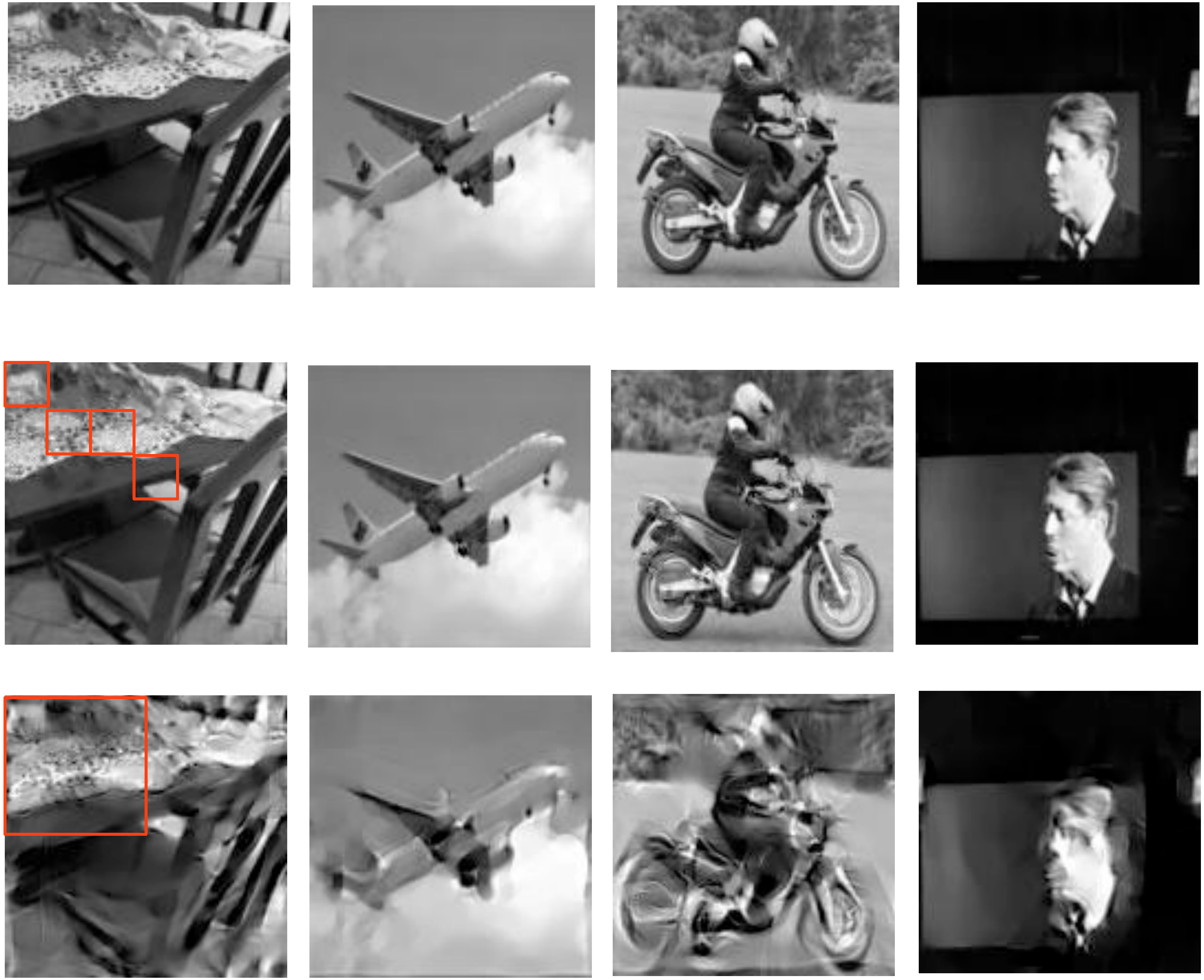
Multiscale Scattering Reconstructions

Original Images
N^2 pixels

Scattering Reconstruction

$2^J = 16$
$1.4\, N^2$ coeff.

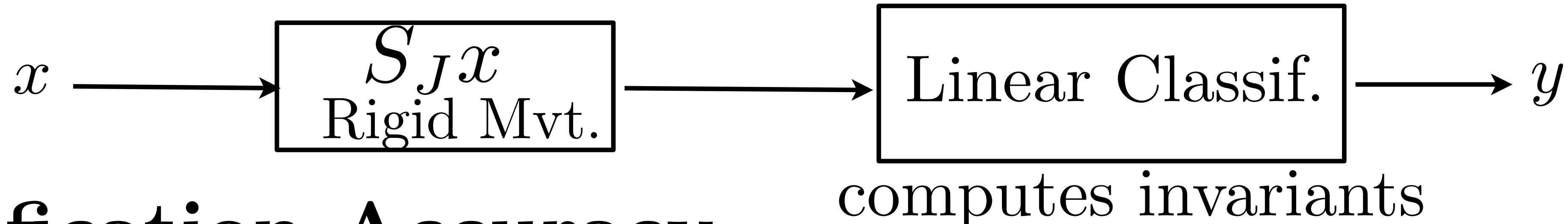
$2^J = 64$
$N^2/8$ coeff.



Complex Image Classification

CalTech 101 data-basis:

Edouard Oyallon



Classification Accuracy

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

Scattering almost linearises these classification problems.

Learning Physics: N-Body Problem

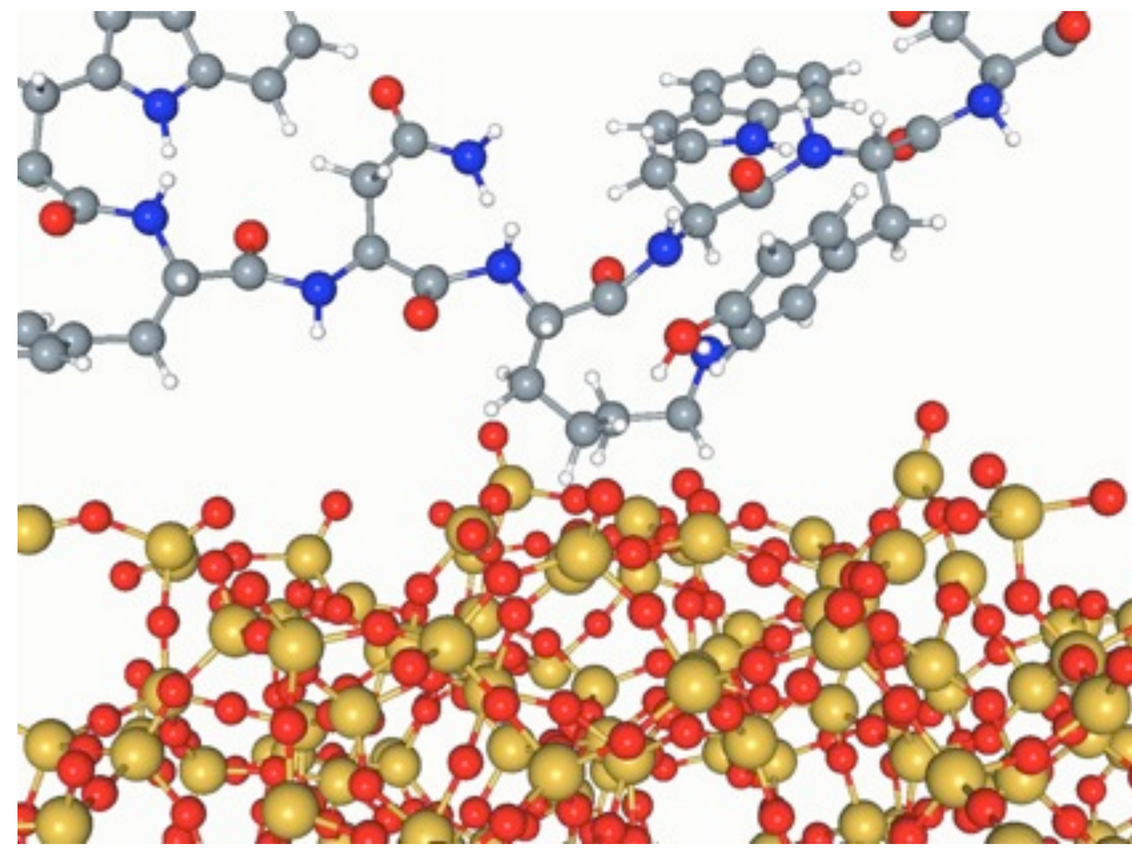
N. Poilvert
Matthew Hirn

- Energy of d interacting bodies:

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

Astronomy

Quantum Chemistry



Second Order Interactions

- Energy of d interacting bodies (Coulomb):

for point charges $x(u) = \sum_{k=1}^{d} q_k \, \delta(u - p_k)$ then

potential $V(r) = |r|^{-\beta}$: $f(x) = \sum_{k=1}^{d} \sum_{k'=1}^{d} \frac{q_k \, q_{k'}}{|p_k - p_{k'}|^{\beta}}$

diagonalized in Fourier : $f(x) = (2\pi)^{-2} \int |\hat{x}(\omega)|^2 \, \hat{V}(\omega) \, d\omega$

can be approximated at best by summing $\sim d$ terms.

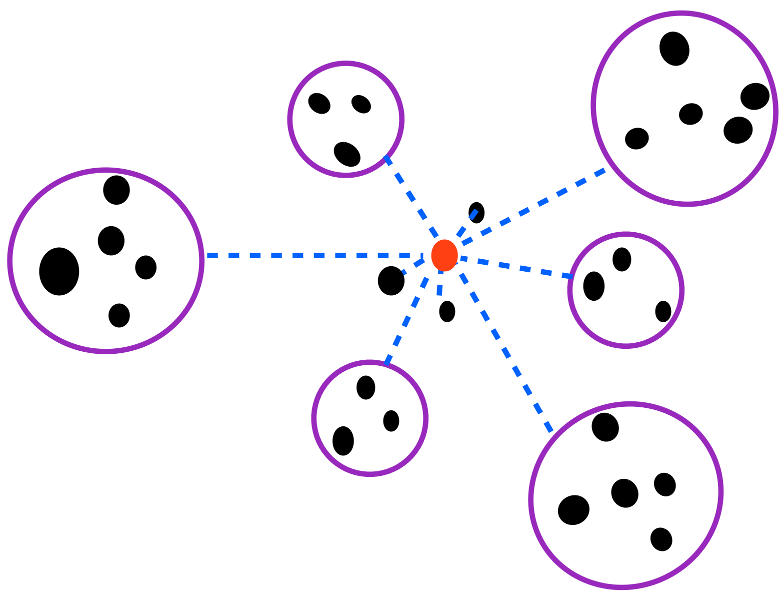
Many Body Interactions

N. Poilvert
Matthew Hirn

- Energy of d interacting bodies (Coulomb):

Fast multipoles: each particle interacts with $O(\log d)$ groups
(Rocklin, Greengard)

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



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

$$f(x) = \sum_{\lambda} v_\lambda \, \|x \star \psi_\lambda\|^2 \, (1 + \epsilon)$$

$O(\log d)$ terms

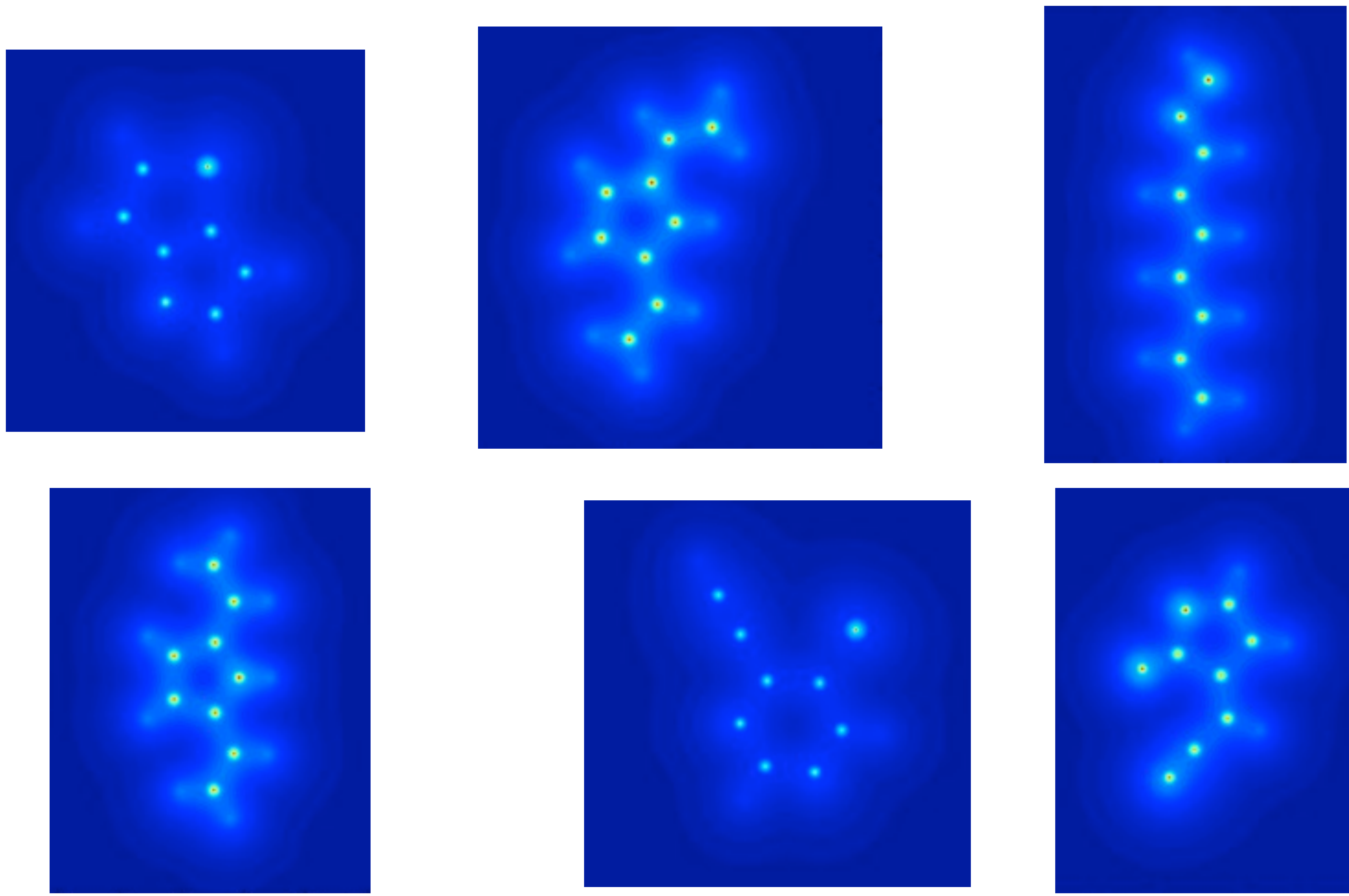
Quantum Chemistry

Protonic charges of a molecule: $x(u) = \sum_{k=1}^{d} q_k \, \delta(u - p_k)$

Atomic energy $f(x)$ = molecule energy - isolated atoms energy

Density Functional Theory: computes the electronic density $\rho(u)$

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



Quantum Chemistry

Atomic energy f is computed from each electronic orbital $\phi_k(u)$

$$\rho(u) = \sum_{k=1}^{K} |\phi_k(u)|^2$$

Kohn-Sham model:

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

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

where ρ minimises the energy $E(\rho)$

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

Quantum Chemistry

N. Poilvert
Matthew Hirn

- Data bases $\{x_i\,,\, f(x_i)\}_i$ of 2D molecules with up to 20 atoms

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

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

Partial Least Square regression on the training set:

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

Quantum Chemistry

N. Poilvert
Matthew Hirn

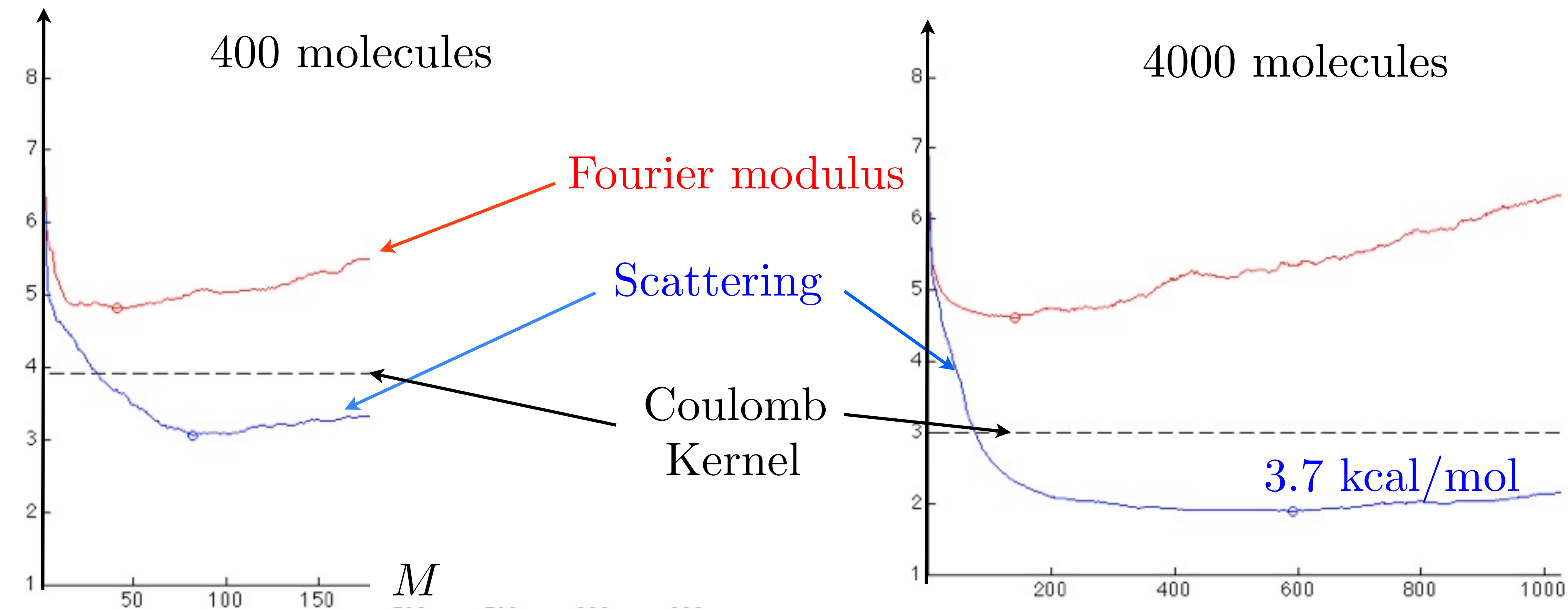
- Data bases $\{x_i \, , \, f(x_i)\}_i$ of 2D molecules with up to 20 atoms

$$f_M(x) = \sum_{k=1}^{M} w_k \, \phi_{n_k}(x) \quad \Rightarrow \text{forces: } \nabla f(x) = \sum_k w_k \, \nabla \phi_{n_k}(x)$$

$$\phi_{n_1}(x) = \int x(u)\,du: \text{ total charge}$$

$$\phi_{n_2}(x) = \|x \star \psi_{\lambda_1}\|_1$$

$\log_2 \mathbb{E}|f(x) - f_M(x)|^2$: testing error

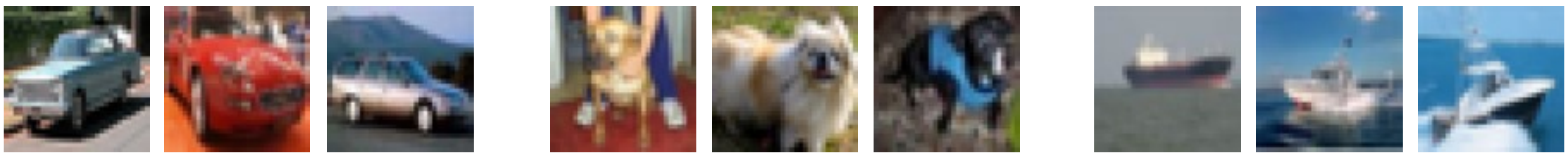


Learning with Unknown Geometry

Xu Chen, Xiu Cheng

CIFAR-10: 10 classes with 500 training images per class

Cars Dogs Ships



If the geometry is unknown (permutation of pixels):

Do not learn the geometry (NP complete)

Learn the support of multiscale wavelets (polynomial algo.)

Learned Haar Scattering : **27%** errors (state of the art)

Conclusion

- A major challenge of data analysis is to find Euclidean embeddings of metrics.

- 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.

- Multitude of open mathematical problems at interface of: geometry, harmonic analysis, probability, statistics, PDE.