High Dimensional Learning

From Images to Quantum Chemistry

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

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

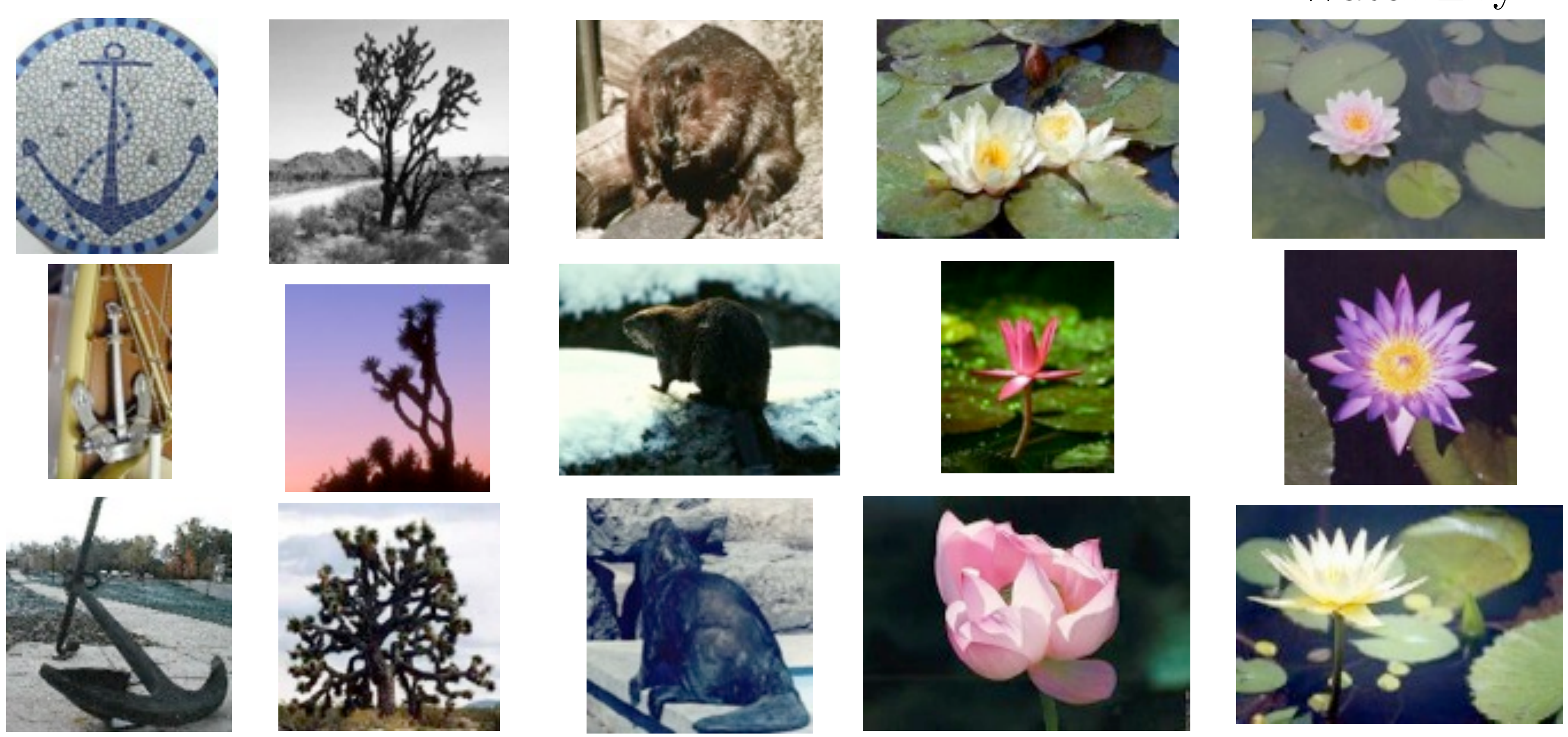
High Dimensional Learning

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

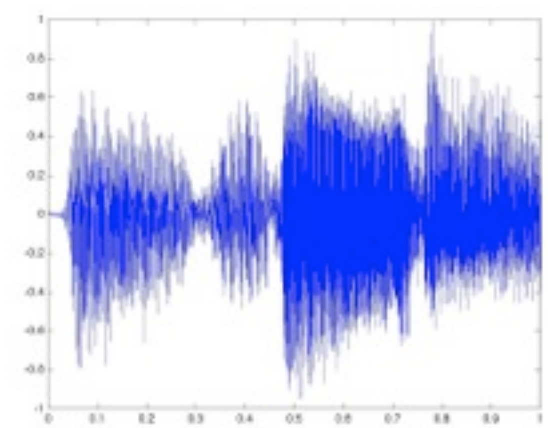
- **Classification:** estimate a class label $f(x)$ given n sample values $\{x_i \, , \, y_i = f(x_i)\}_{i \leq n}$

Image Classification $d = 10^6$

Anchor | Joshua Tree | Beaver | Lotus | Water Lily



Sons
$d = 10^4 / s$



High Dimensional Learning

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

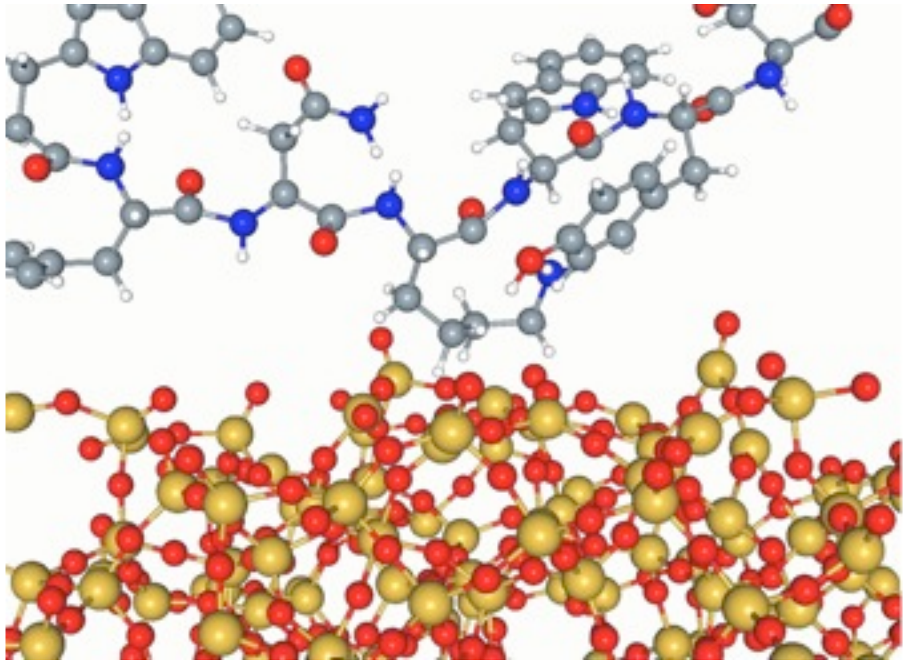
- **Regression:** approximate a *functional* $f(x)$ given n sample values $\{x_i \, , \, y_i = f(x_i)\}_{i \leq n}$

Physics: Many Body Problem

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

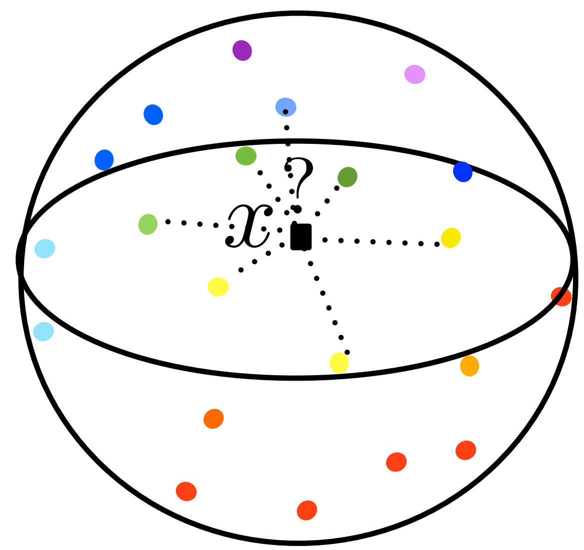
Astronomy

Quantum Chemistry



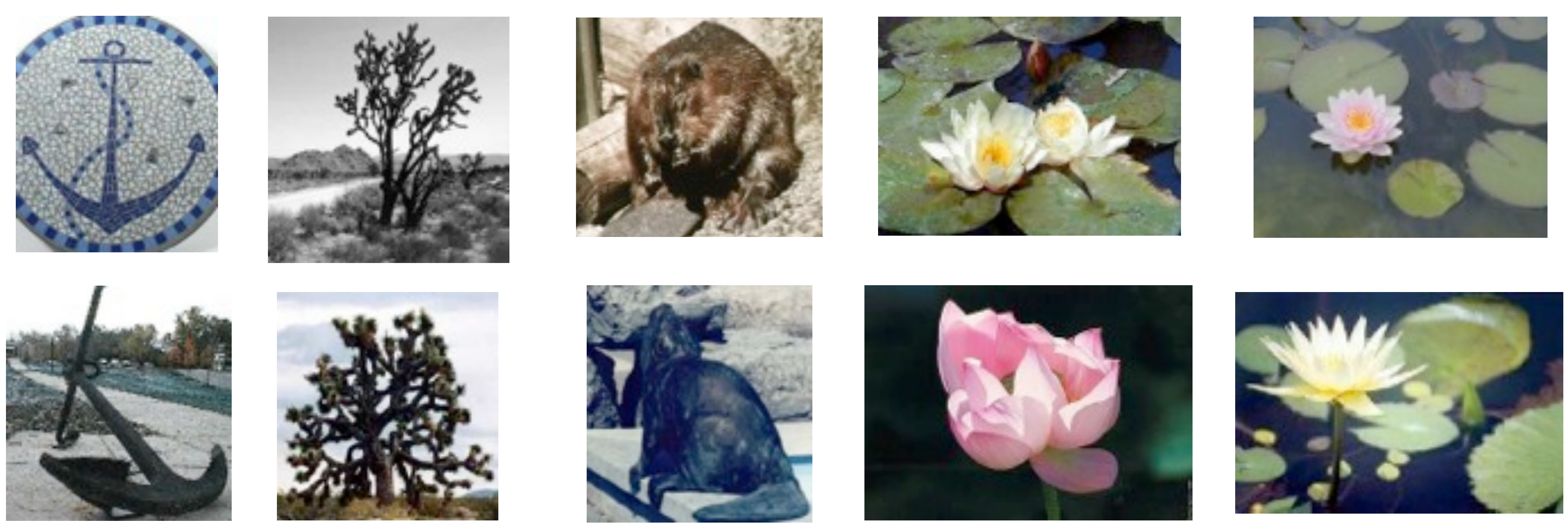
Curse of Dimensionality

- $f(x)$ 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

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



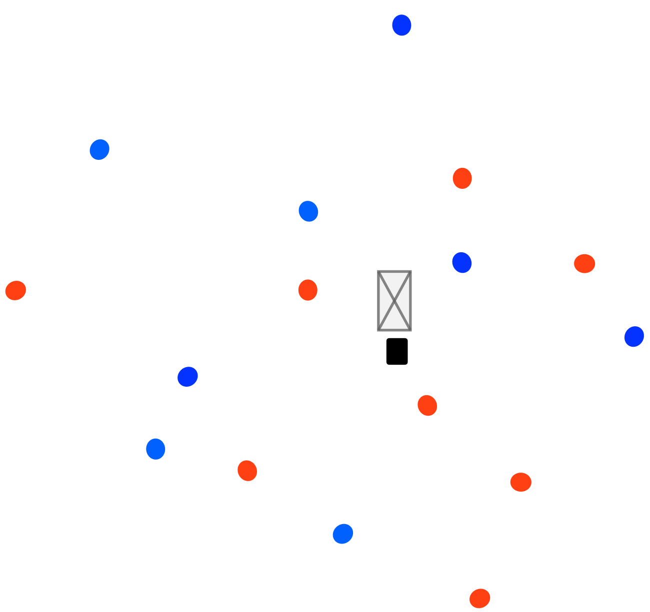
Euclidean Embedding

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

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

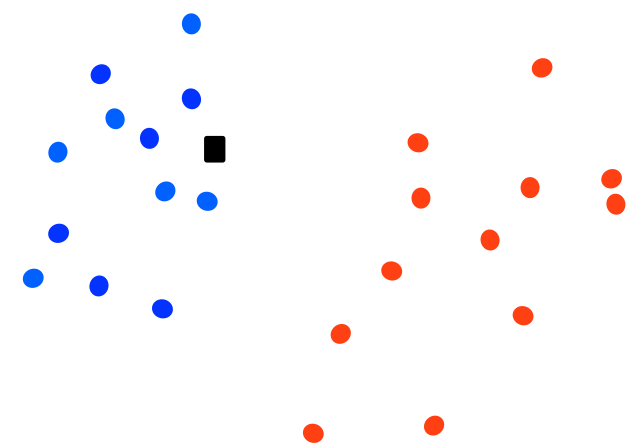
Representation

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



Intelligence

$$\xrightarrow{\Phi}$$



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

Bi-Lipschitz Euclidean metric embedding:

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

How to define Φ ?

Linear Regression

- Representation of x: $\Phi(x) = \{\phi_n(x)\}_n$
- Regression $\tilde{f}(x)$ of $f(x)$ linear in $\Phi(x)$:

$$\tilde{f}(x) = \langle w, \Phi(x) \rangle = \sum_n w_n \, \phi_n(x)$$

interpolates: $\forall i \; , \; f(x_i) = \tilde{f}(x_i) = \langle w, \Phi(x_i) \rangle$

regular: $\min \|w\|^2 = \sum_n w_n^2 \quad \Rightarrow \quad w = \sum_i \alpha_i \, \Phi(x_i)$

$$\Rightarrow \quad \tilde{f}(x) = \sum_i \alpha_i \, \underbrace{\langle \Phi(x) \, , \, \Phi(x_i) \rangle}_{\text{kernel similarity of } x \text{ and } x_i}$$

Linear Classifiers

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

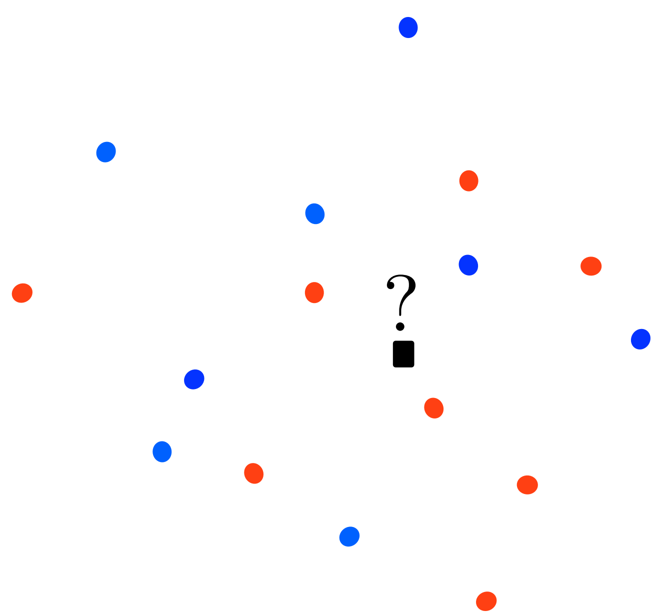
$\|x - x'\|$: non-informative

Representation

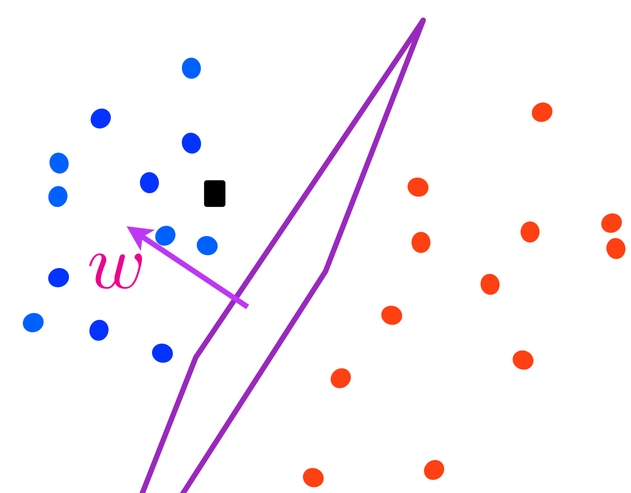
$\Phi x \in \mathcal{H}$

Linear Classifier

$\mathrm{sign}(\langle w, \Phi x \rangle + b)$



$\xrightarrow{\Phi}$



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

How to define Φ ?

Known Euclidean Embeddings

- If the data is in a low-dimensional manifold, embedding of manifold metrics with a heat kernel:

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

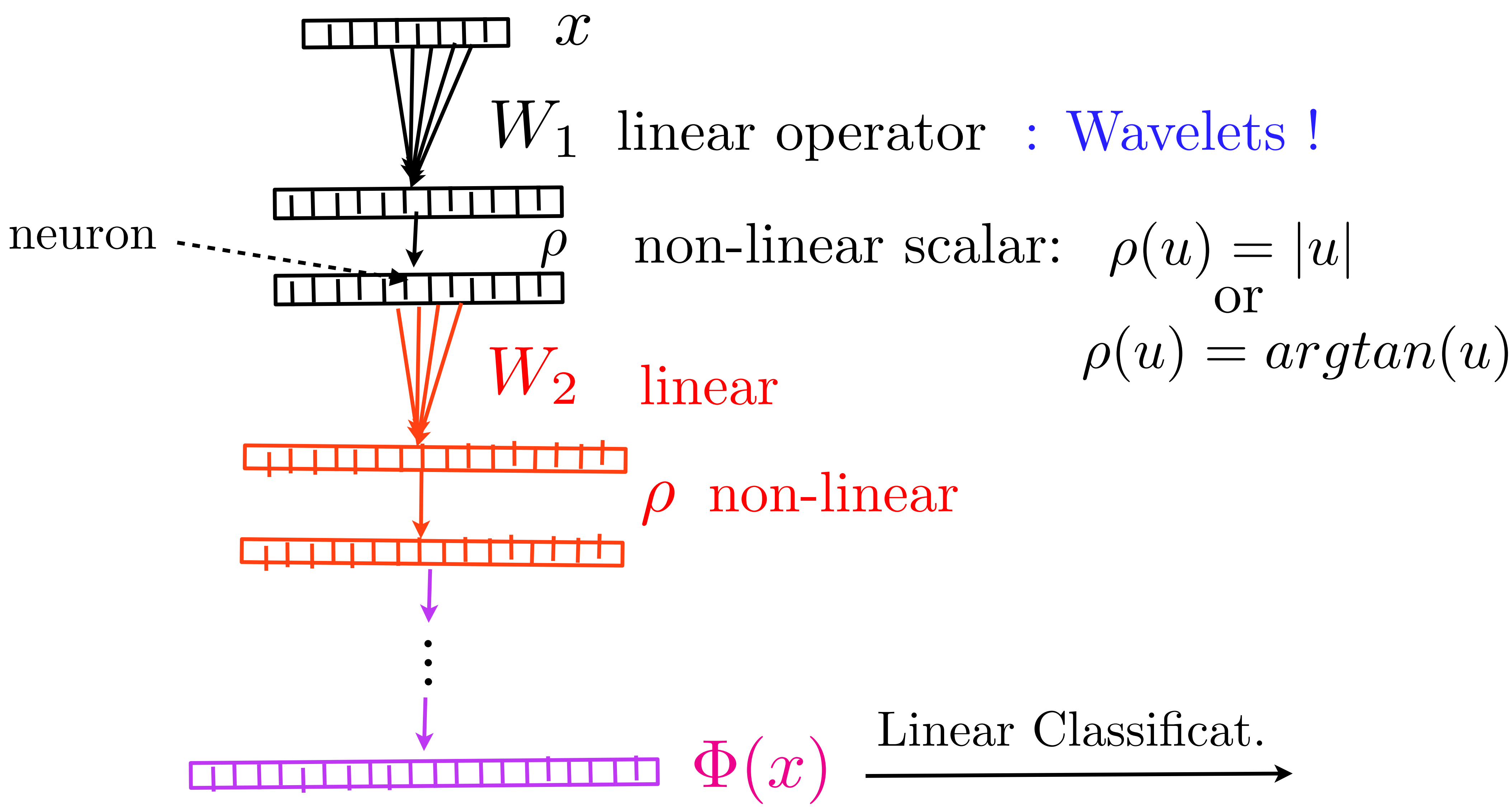
- Embedding of Banach metrics over finite set of points $\{x_i\}_i$ but problem of generalisation for all x. *(Bourgain)*

  Need to embed the full space.

- Can we learn $\Phi(x)$ from data ?

Deep Neural Neworks

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



Optimize the W_k: 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 ?

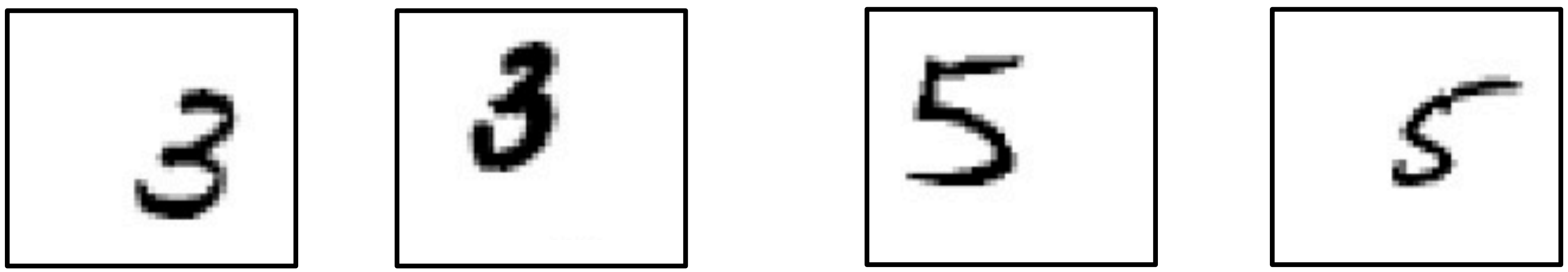
Overview

- Embedding geometry: invariance and stability to deformations
- Image classification
- Learning physics: quantum chemistry energy regression

Image Metrics

- Low-dimensional "geometric shapes"

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



Deformation metric: (classic mechanics)

Deformation: $D_\tau x(u) = x(u - \tau(u))$

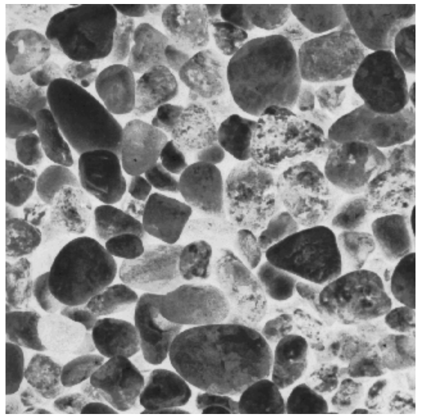
$$\Delta(x, x') \sim \min_\tau \|D_\tau x - x'\| + \|\nabla \tau\|_\infty \|x\|$$

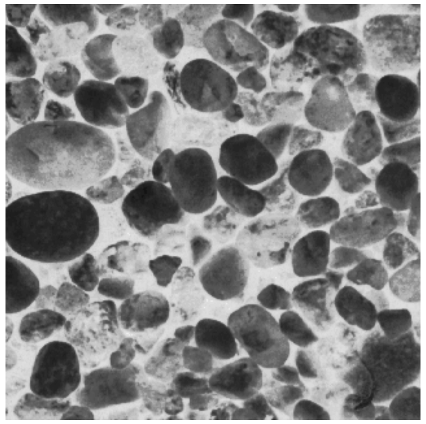
Invariant to translations

diffeomorphism amplitude

Image Metrics

- High dimensional textures:
  ergodic stationary processes





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

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

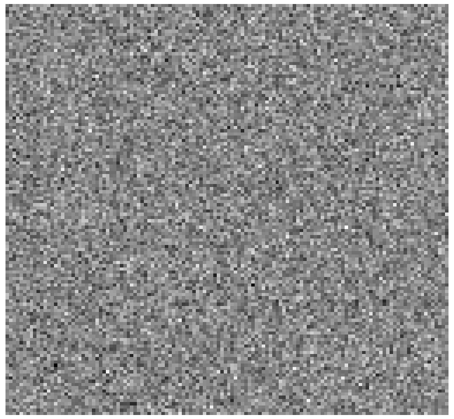
x 

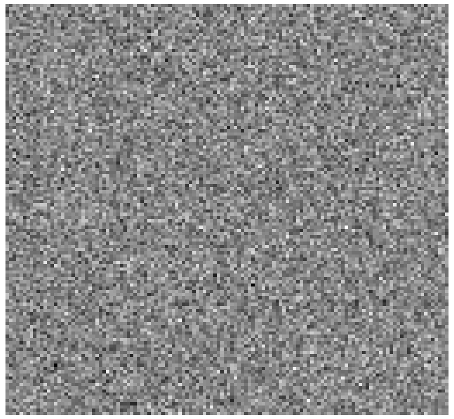
x' 

But not equivalent:

$\Delta(x', x) = 0$ 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:

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

- **Lipschitz stable** to diffeomorphisms

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

$\Rightarrow$ Invariance to translation

Fourier Deformation Instability

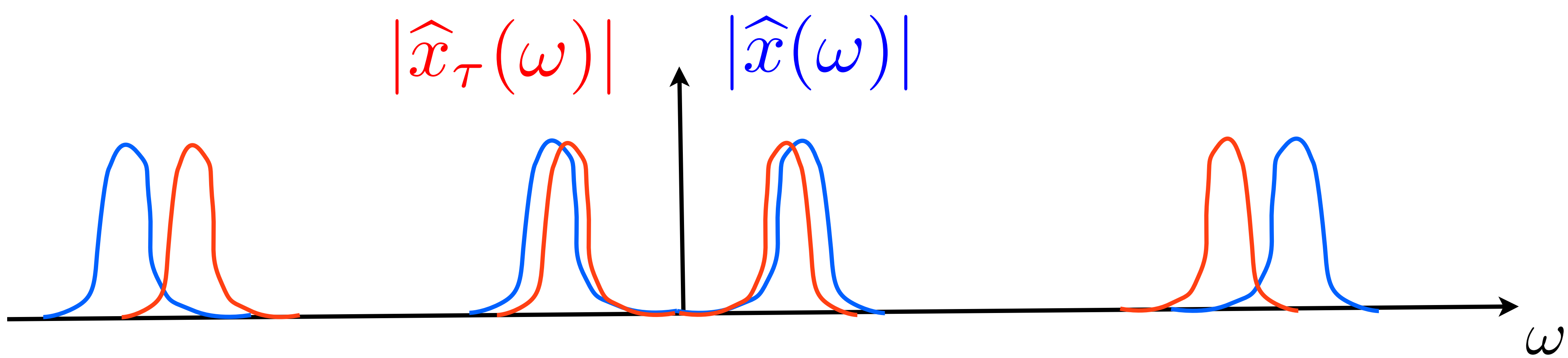
- Fourier transform $\hat{x}(\omega) = \int x(t)\, e^{-i\omega t}\, dt$

  The modulus is invariant to translations:

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

- 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



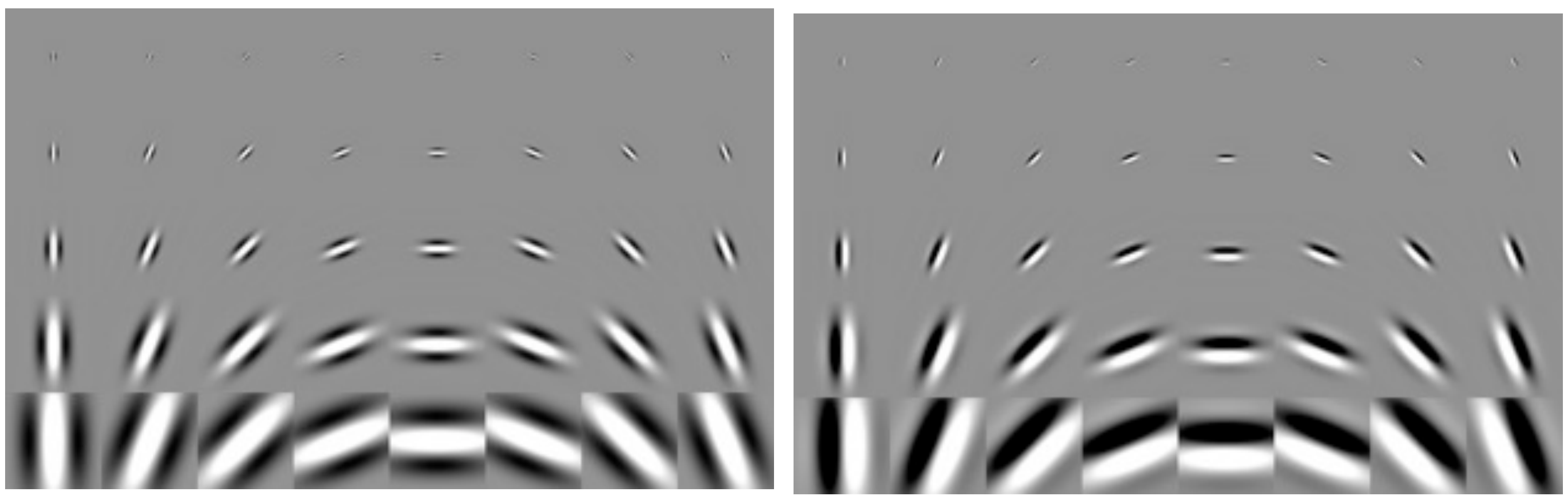
$$\Rightarrow \quad \|\Phi(x) - \Phi(x_\tau)\| \gg \|\nabla\tau\|_\infty\, \|x\|$$

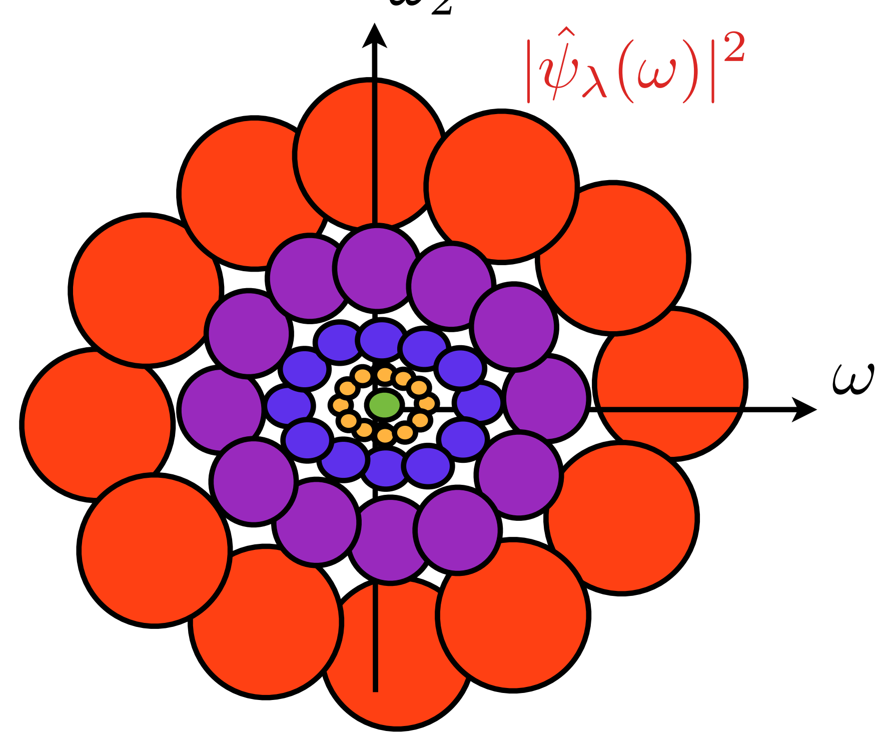
Scale separation with Wavelets

- 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

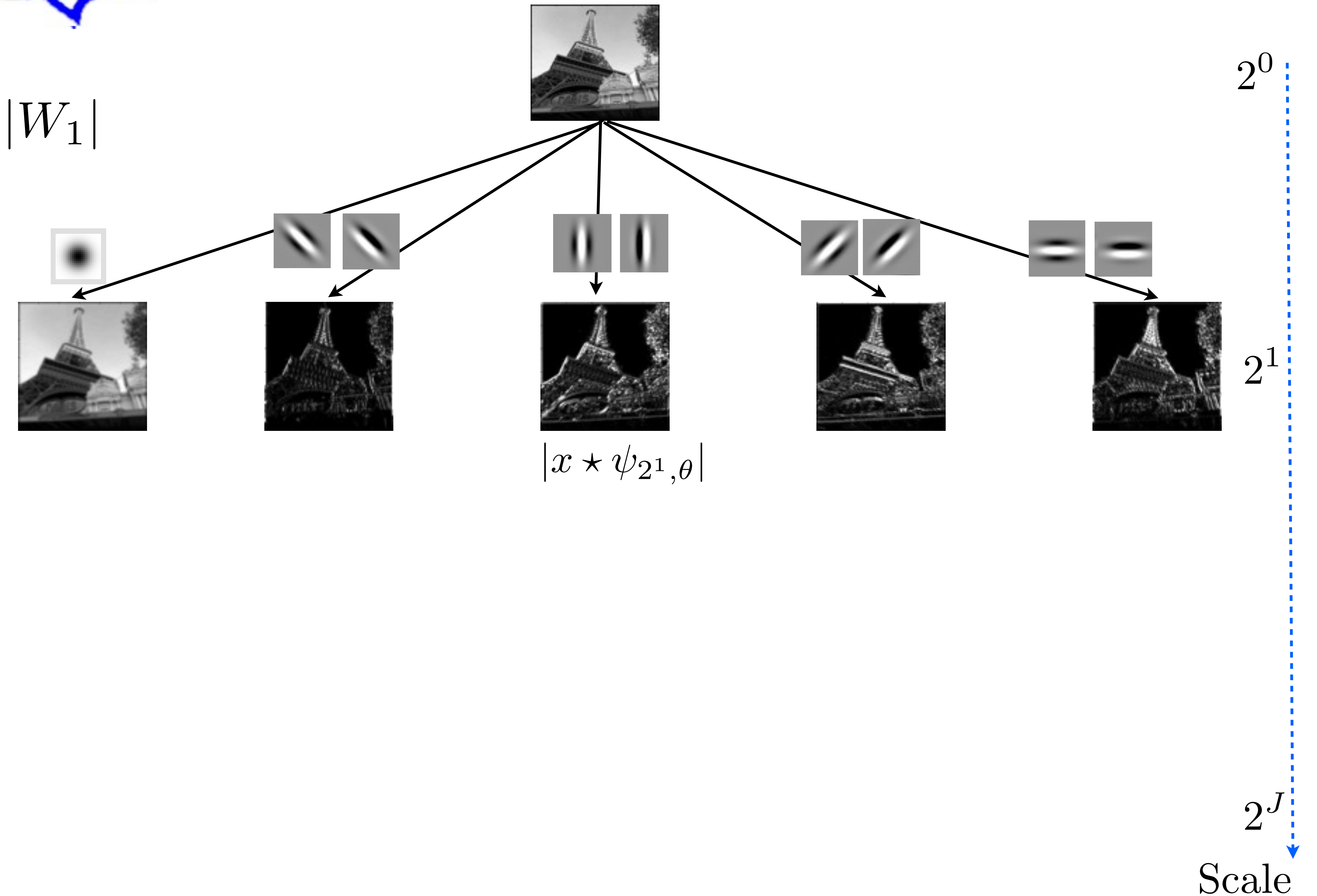




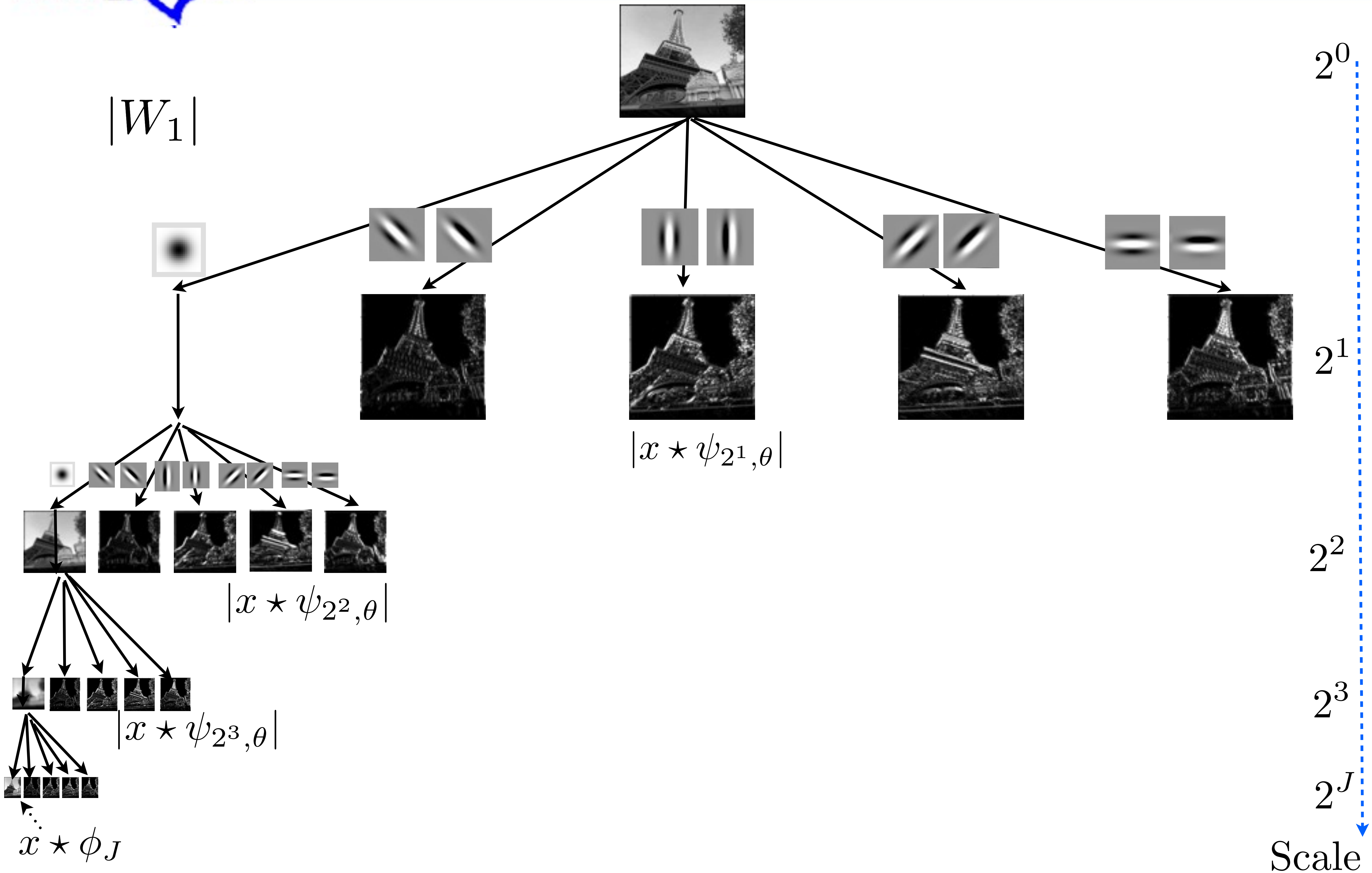
- Wavelet transform: $Wx = \begin{pmatrix} x \star \phi_{2^J}(t) \\ x \star \psi_\lambda(t) \end{pmatrix}_{\lambda \leq 2^J}$

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

Fast Wavelet Transform



Wavelet Transform

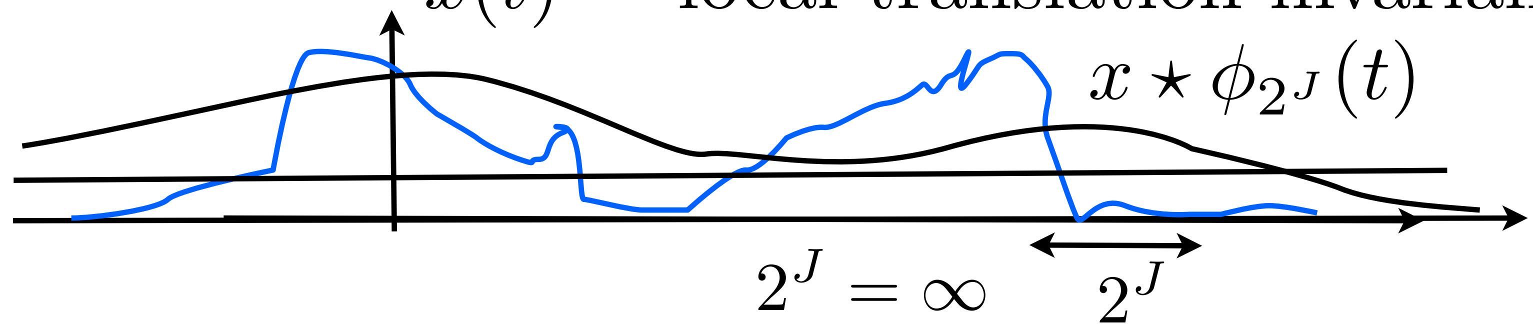


Wavelet Translation Invariance

First wavelet transform

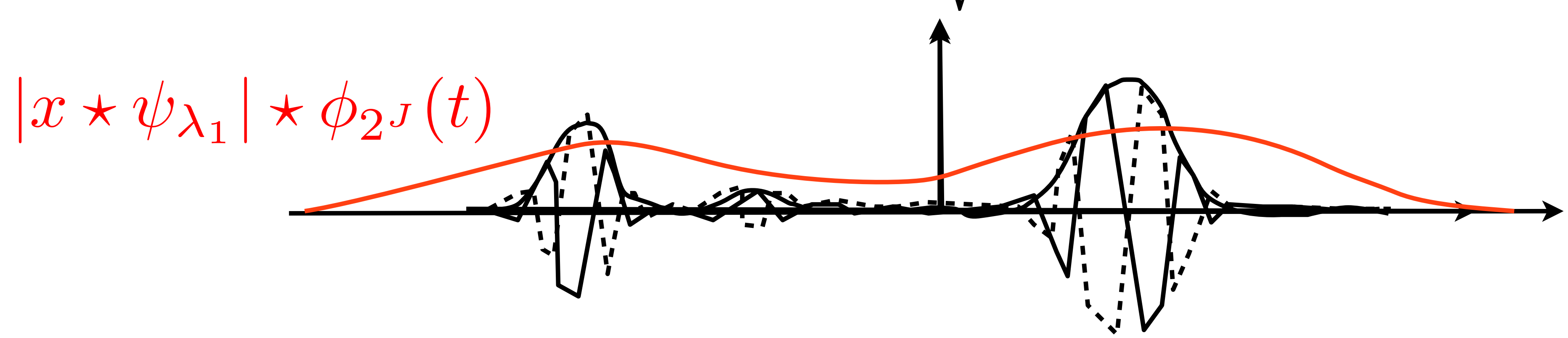
$$W_1 x = \begin{pmatrix} x \star \phi_{2^J} \\ x \star \psi_{\lambda_1} \end{pmatrix}_{\lambda_1}$$

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



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

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

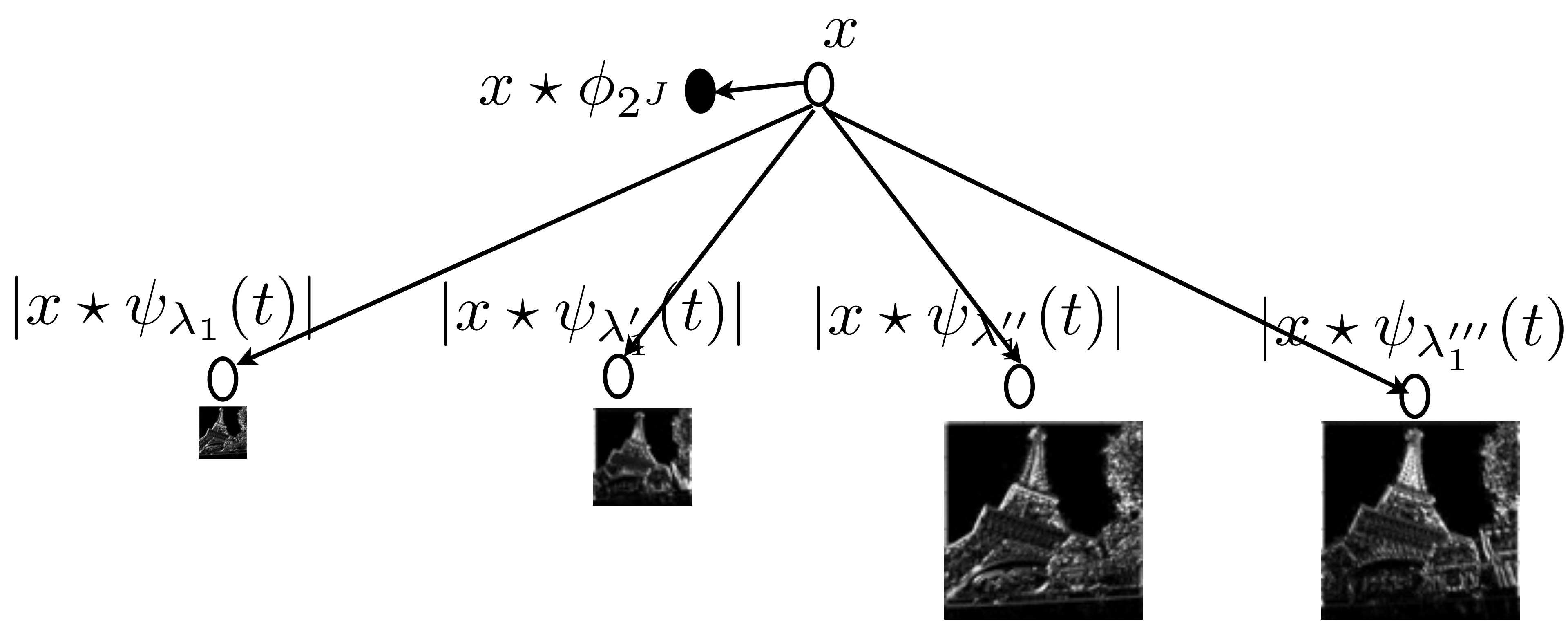


Second wavelet transform modulus

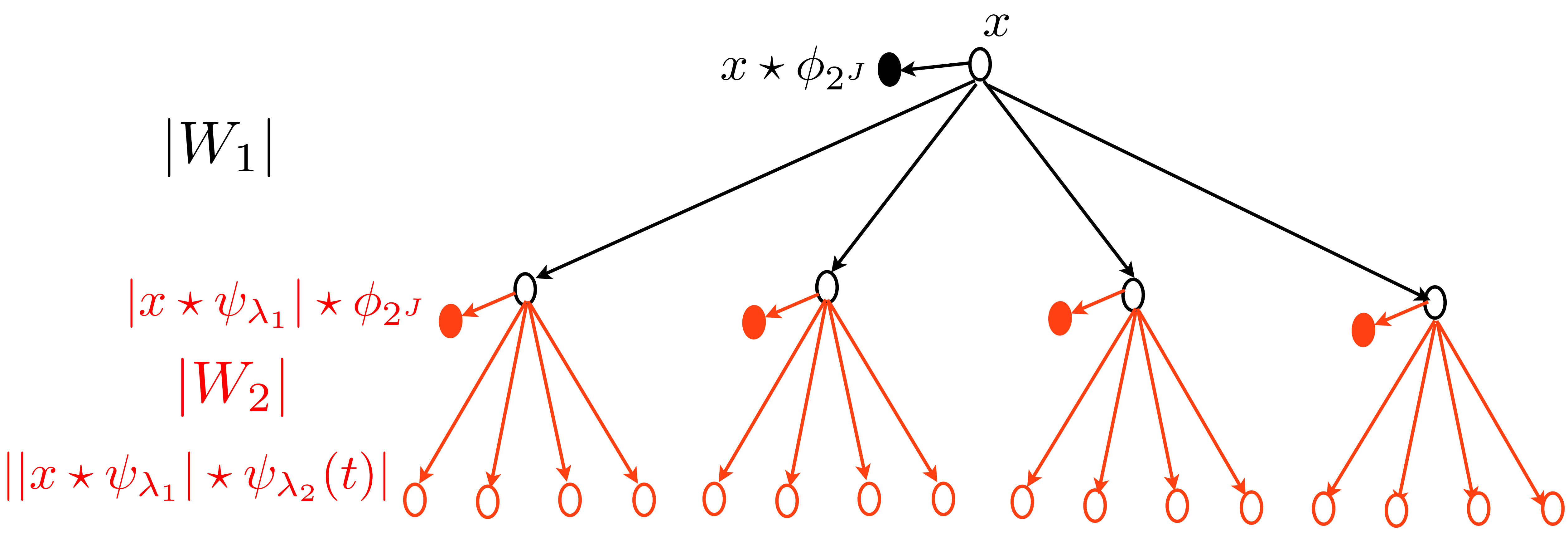
$$|W_2|\ |x \star \psi_{\lambda_1}| = \begin{pmatrix} |x \star \psi_{\lambda_1}| \star \phi_{2^J}(t) \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}(t)| \end{pmatrix}_{\lambda_2}$$

Scattering Transform

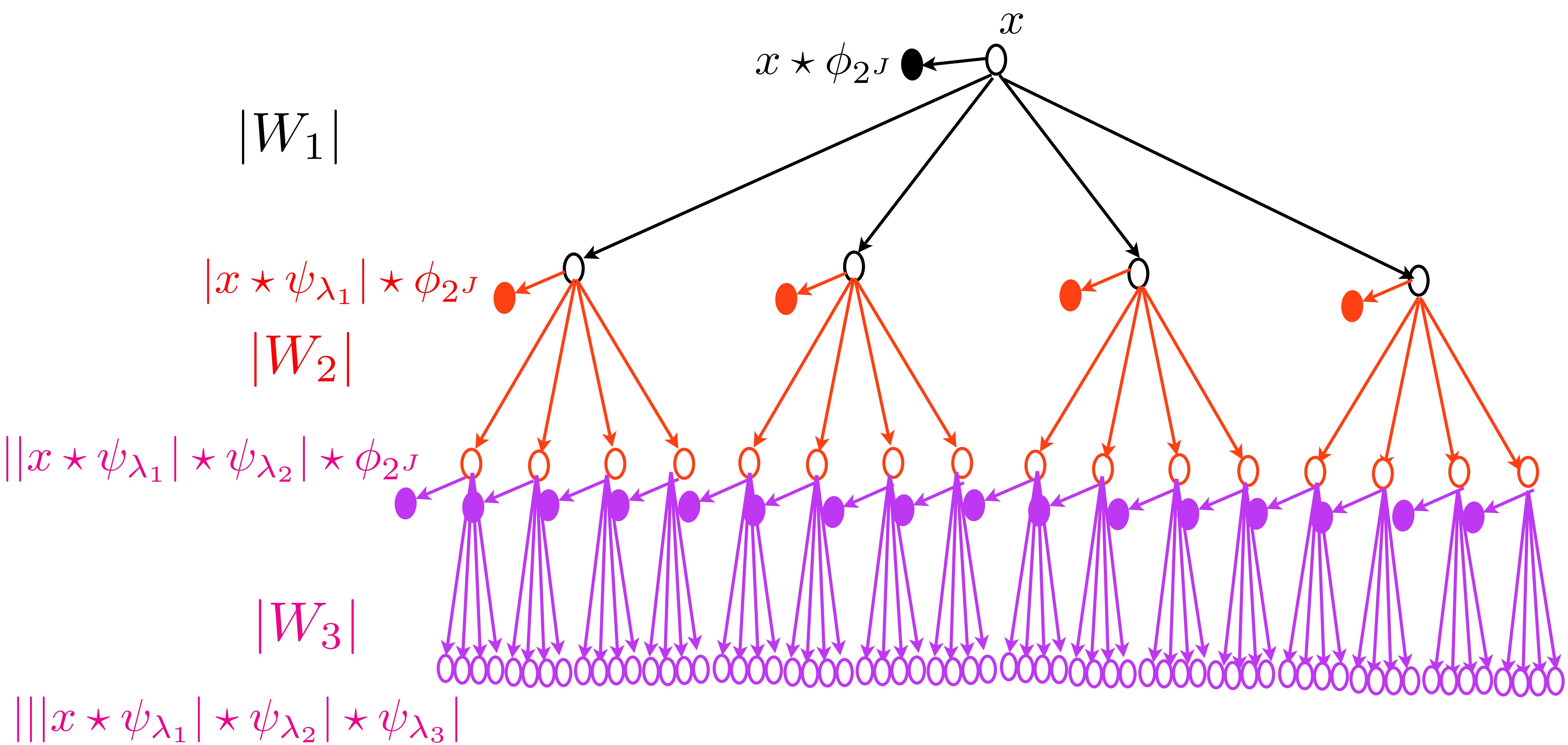
$|W_1|$



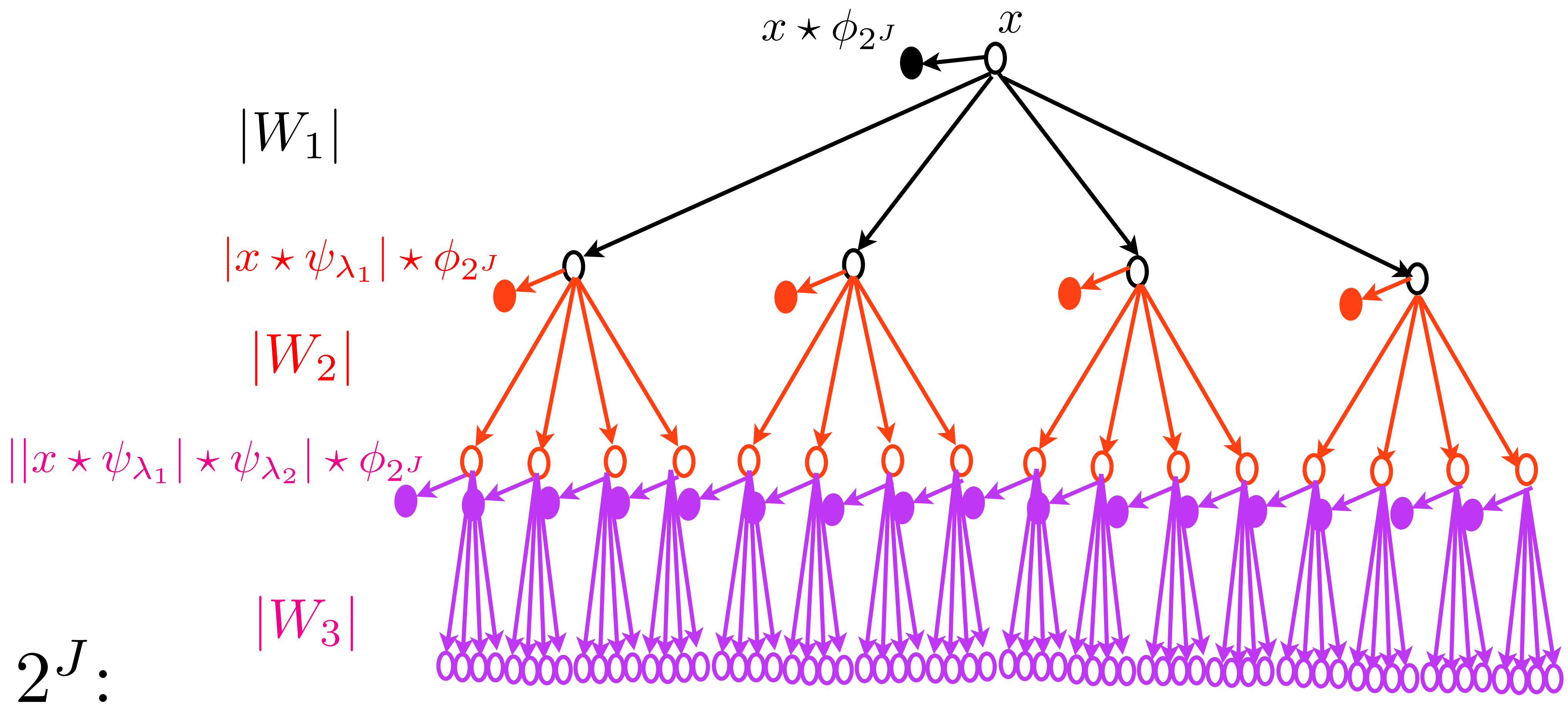
Scattering Transform



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*)

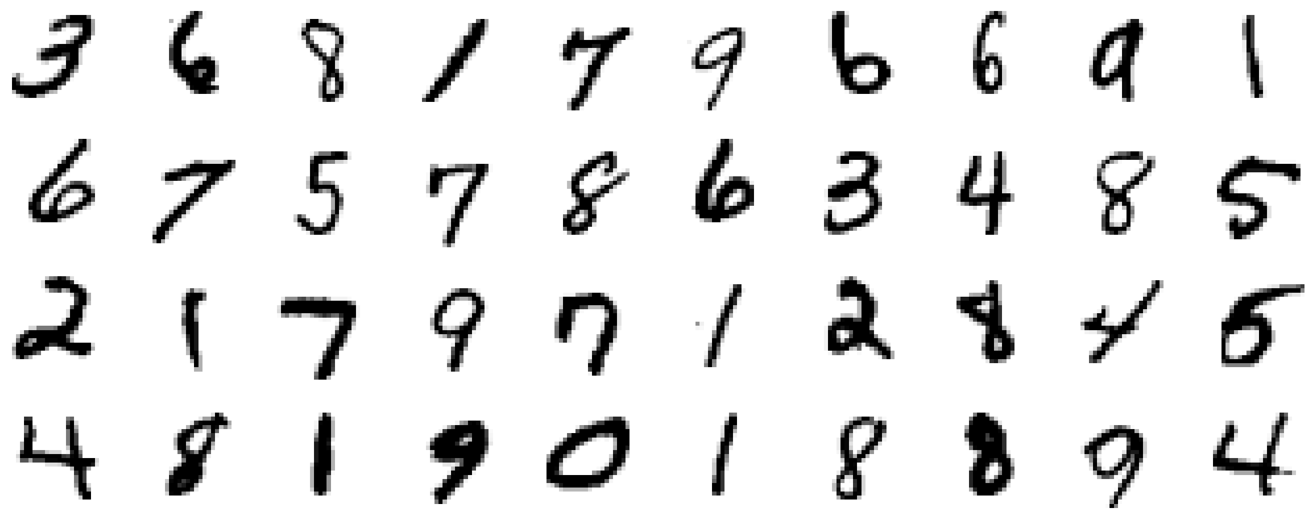
preserves norms $\|S_J x\| = \|x\|$

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

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

Classification Errors

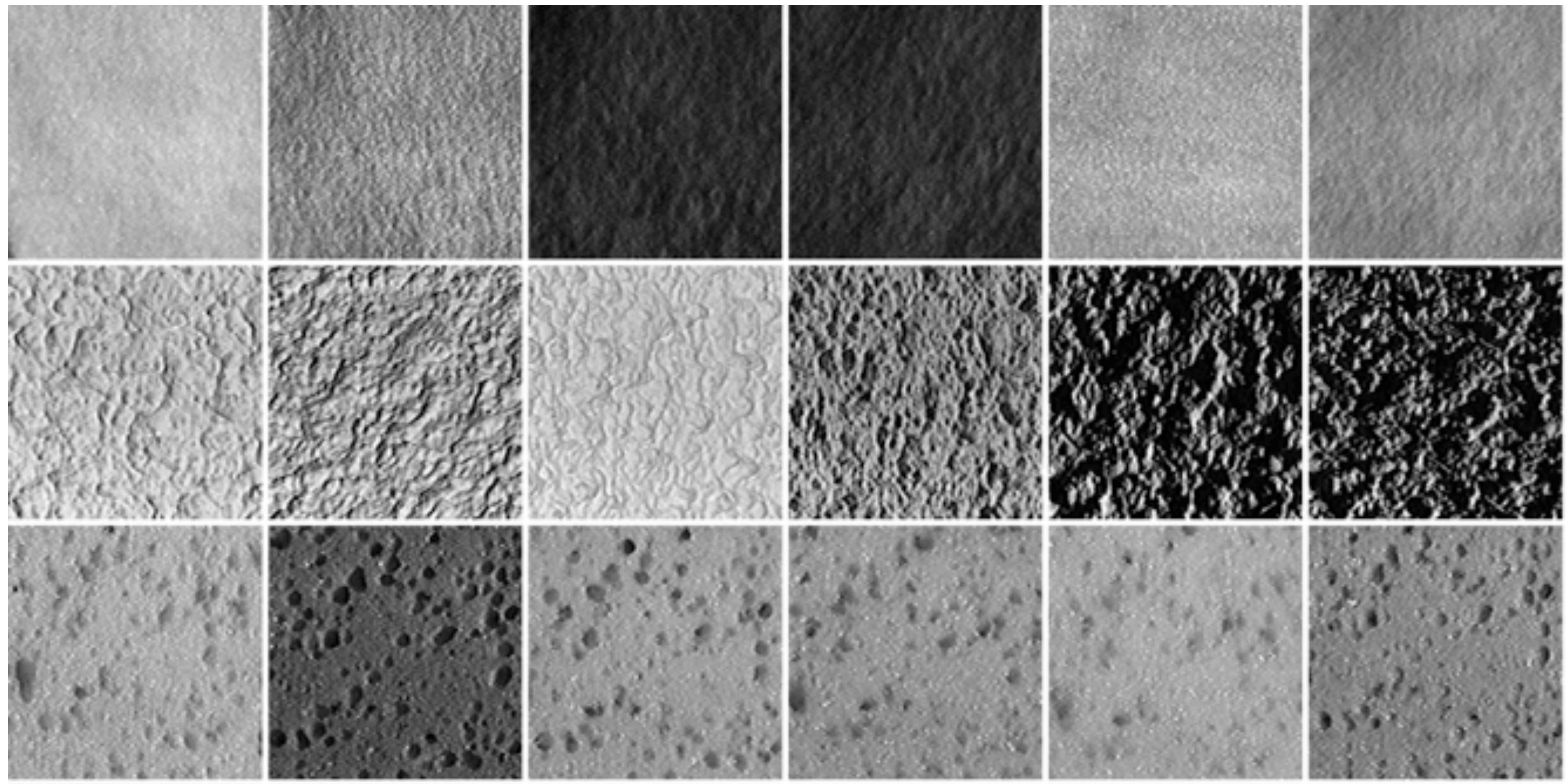
Training size	Conv. Net.	Scattering
300	7.2%	**4.4**%
5000	1.5%	**1.0**%
20000	0.8%	**0.6**%
60000	0.5%	**0.4**%

LeCun et. al.

Classification of Textures

J. Bruna

CUREt database
61 classes



Scattering Moments of Processes

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

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

is a low-variance estimator of the scattering moments of $X(t)$

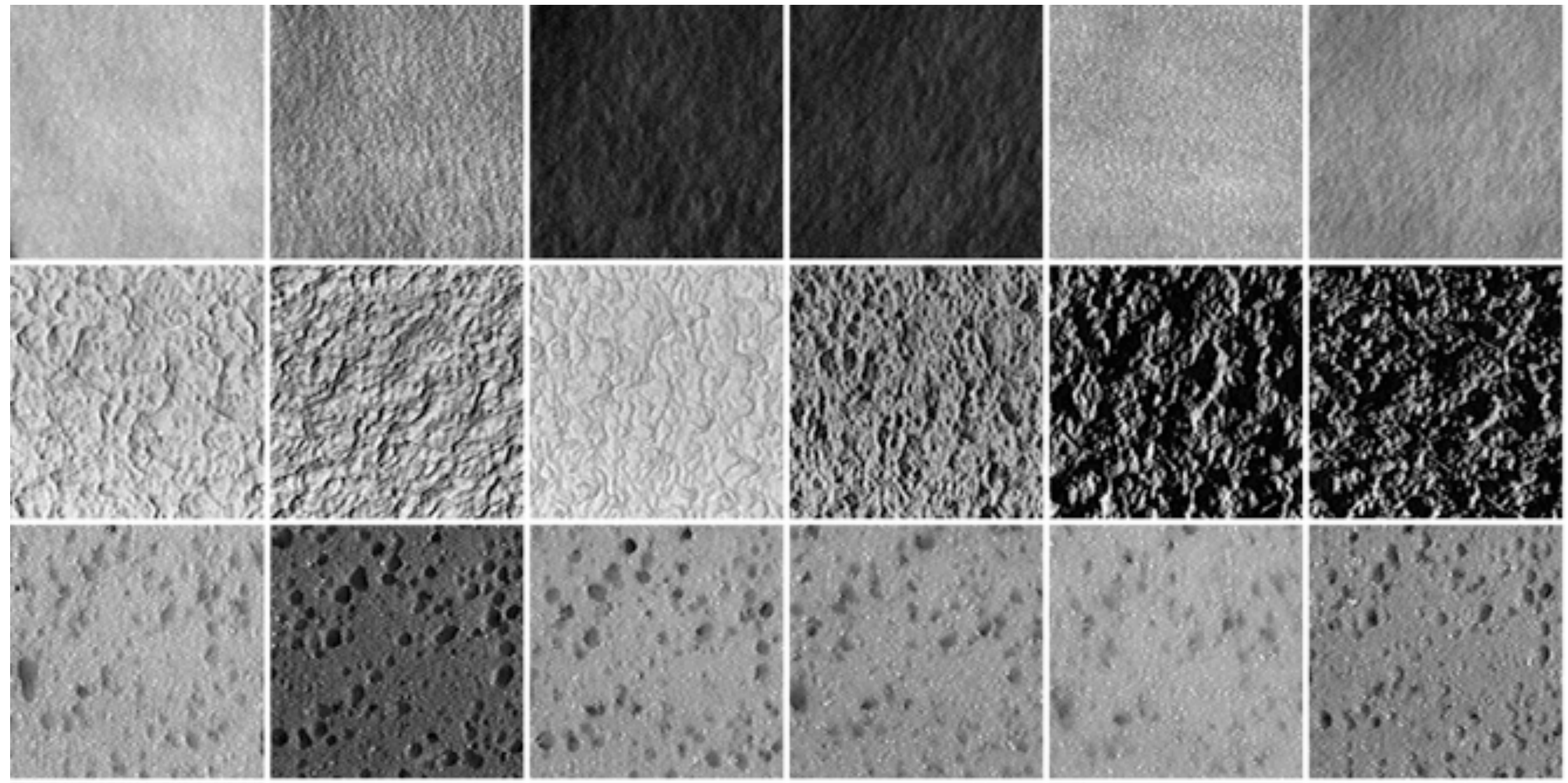
$$\overline{S} X = \begin{pmatrix} E(X) \\ E(|X \star \psi_{\lambda_1}|) \\ E(||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|) \\ E(|||X \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}|) \\ ... \end{pmatrix}_{\lambda_1,\lambda_2,\lambda_3,...}$$

$$\lim_{J \to \infty} S_J X = \overline{S} X \quad \text{in mean-square, if } X \text{ is "ergodic" .}$$

Classification of Textures

J. Bruna

CUREt database
61 classes



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

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

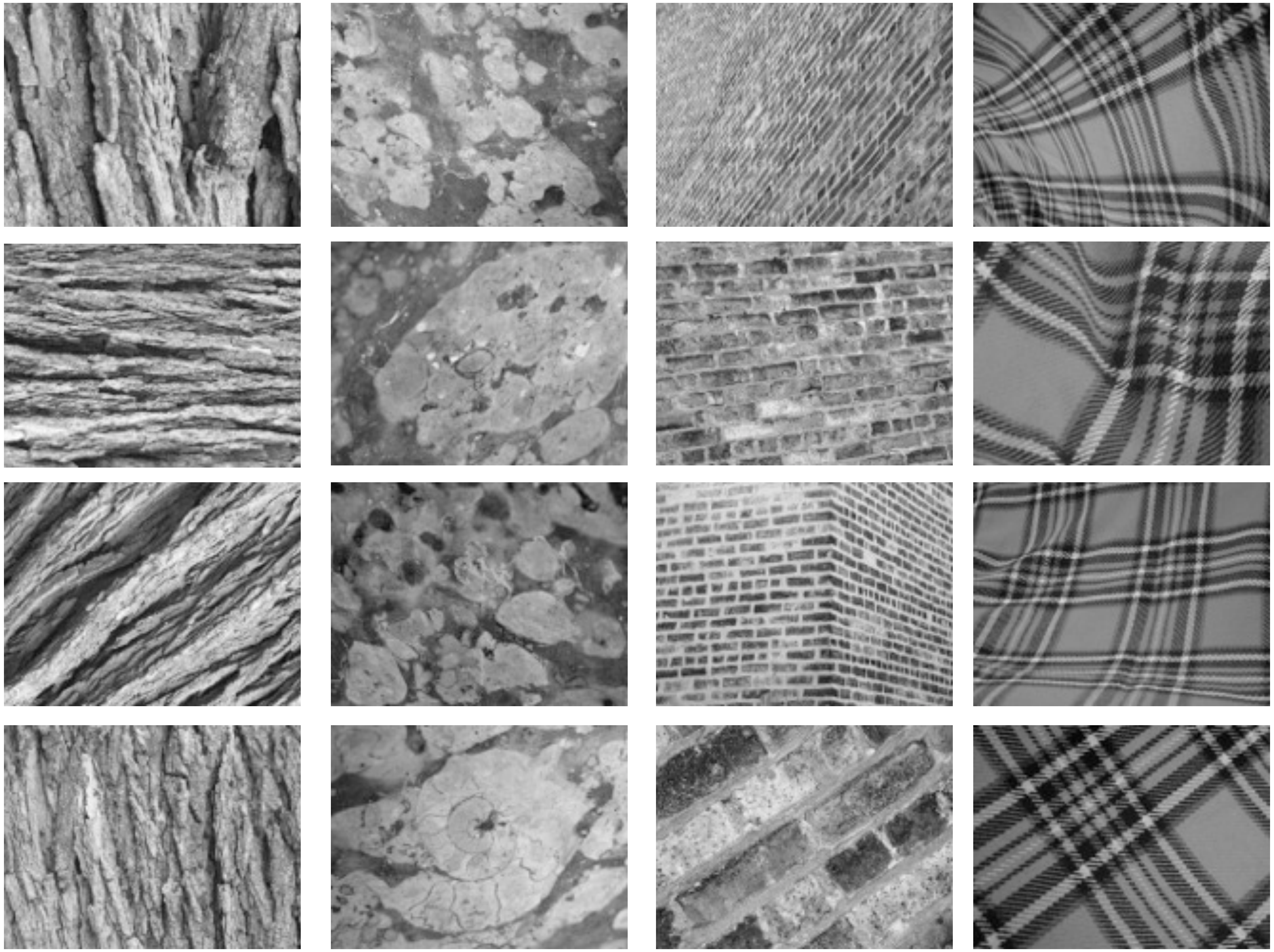
Training per class	Fourier Spectr.	Histogr. Features	Scattering
46	1%	1%	**0.2** %

- Can characterise non-Gaussian properties of 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

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

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

  $$(v', \theta')\,(v, \theta) = (v' + r_{\theta'} v \,,\, \theta + \theta') \ \ : \text{non-commutative}$$

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

- Action on wavelet coefficients:

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

$$(v', \theta') \,.\, x(u) \longrightarrow \boxed{|W_1|} \longrightarrow x_j(r_{\theta'}^{-1}(u - v'), \theta - \theta')$$

$$\big\downarrow$$

$$\int x(u) du$$

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 Isometries:

Wavelets on Translations — Wavelets on Isometries — Wavelets on Isometries

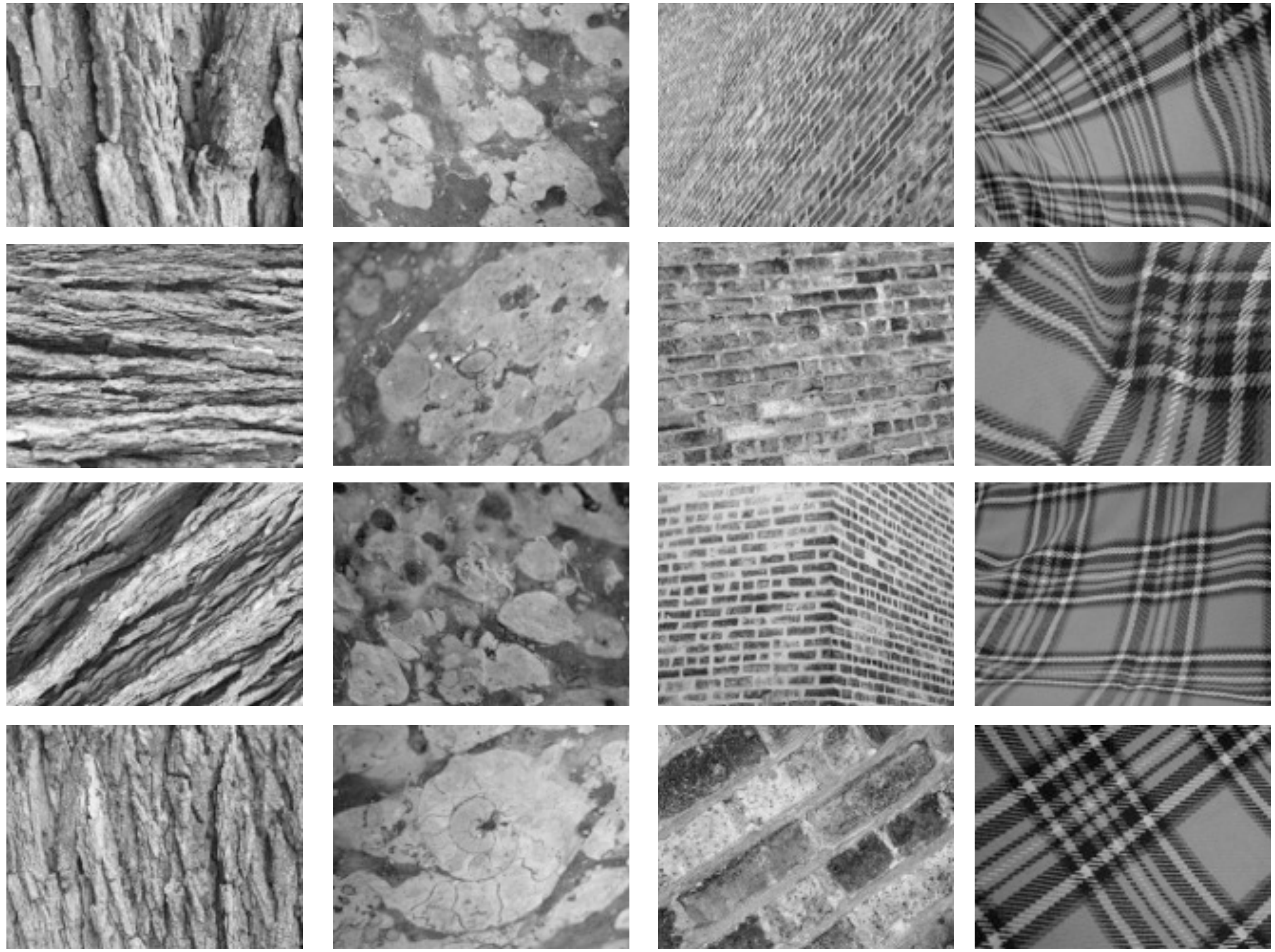
$$x(u) \longrightarrow \boxed{|W_1|} \longrightarrow x_j(u\,,\,\theta) \longrightarrow \boxed{|W_2|} \longrightarrow |x_j \circledast \psi_{\lambda_2}(v,\theta)| \longrightarrow \boxed{|W_3|} \longrightarrow$$

$$\boxed{|W_1|} \downarrow \int x(u)du \qquad \boxed{|W_2|} \downarrow \int x_j(u,\theta)\, dud\theta \qquad \boxed{|W_3|} \downarrow \int |x_j \circledast \psi_{\lambda_2}(v,\theta)| dud\theta$$

Rotation and Scaling Invariance

Laurent Sifre

UIUC database:
25 classes



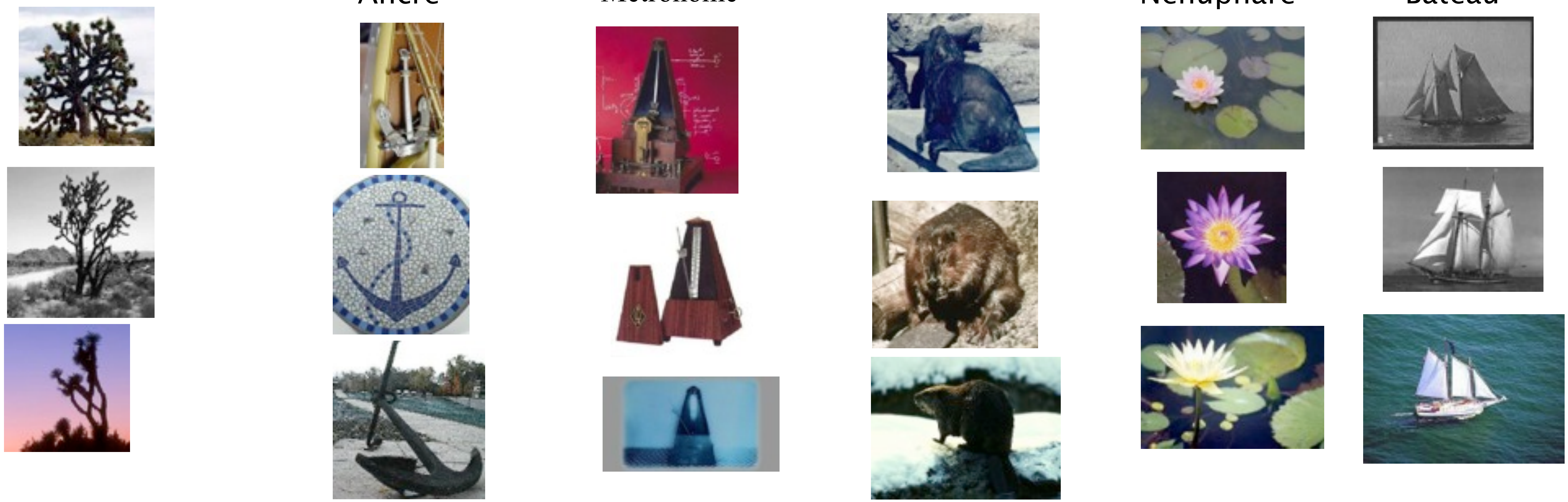
Scattering classification errors

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

Complex Image Classification

CalTech 101 data-basis: *Edouard Oyallon*

Arbre de Joshua | Ancre | Metronome | Castore | Nénuphare | Bateau



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

Classification Accuracy $2^J = 2^5$

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

Scattering almost linearises these classification problems.

Inverse Scattering Transform

Joan Bruna

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

$$\forall m \ , \ \forall \lambda_1, ..., \lambda_m \ , \ S_J \tilde{x}(\lambda_1, ..., \lambda_m) = S_J x(\lambda_1, ..., \lambda_m)$$

- At the second order for $J = \infty$:

$$\min \|\tilde{x}\|$$

such that: $$\int x(u)\, du = \int \tilde{x}(u)\, du$$

$$\forall \lambda_1 \ , \ \|\tilde{x} \star \psi_{\lambda_1}\|_1 = \|x \star \psi_{\lambda_1}\|_1$$

$$\forall \lambda_1, \lambda_2 \ , \ \| |\tilde{x} \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1 = \| |x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1$$

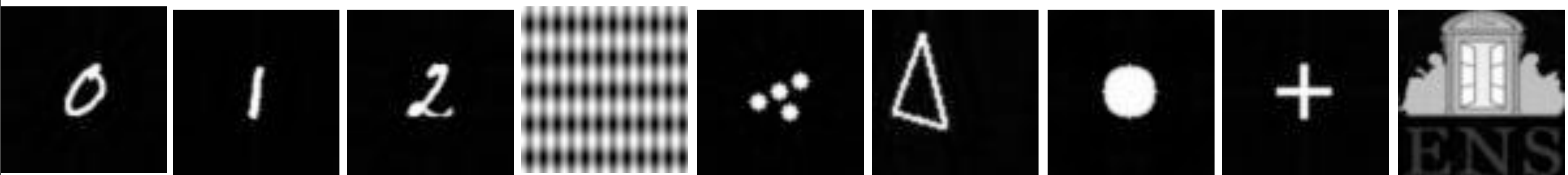
Non convex optimization.

Sparse Shape Reconstruction

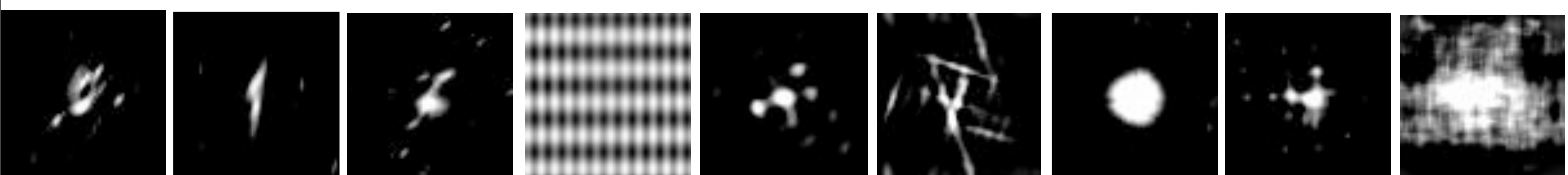
Joan Bruna

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

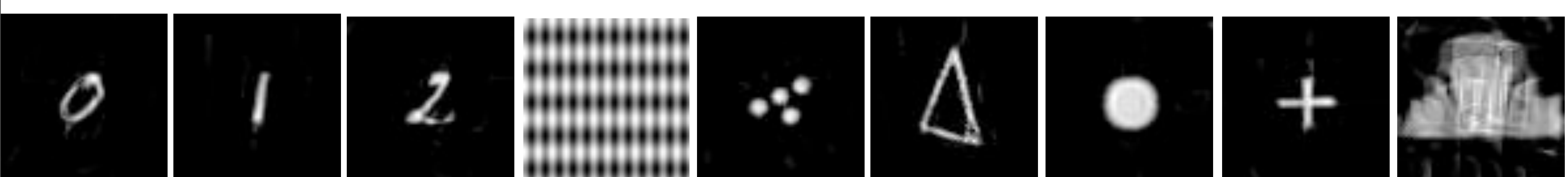
Original images of N^2 pixels:



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



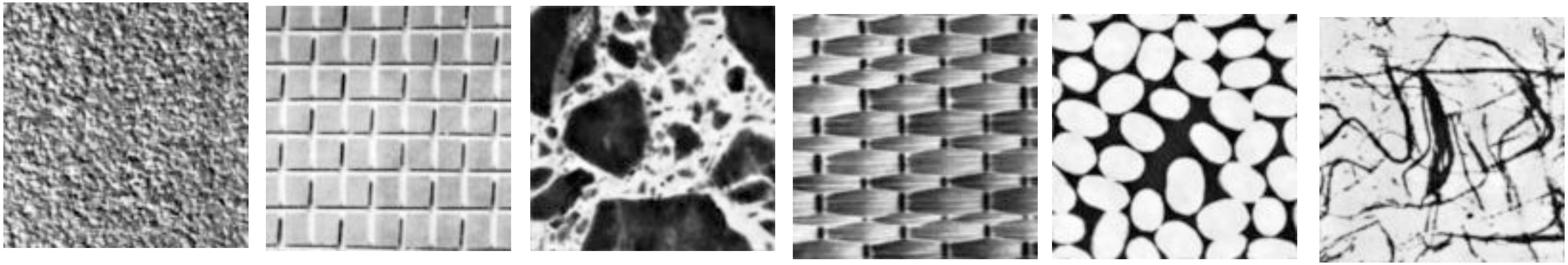
Reconstruction from $\{\|x\|_1 \,, \, \|x \star \psi_{\lambda_1}\|_1 \,, \, \||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1\} : \; O(\log_2^2 N)$ coeff.



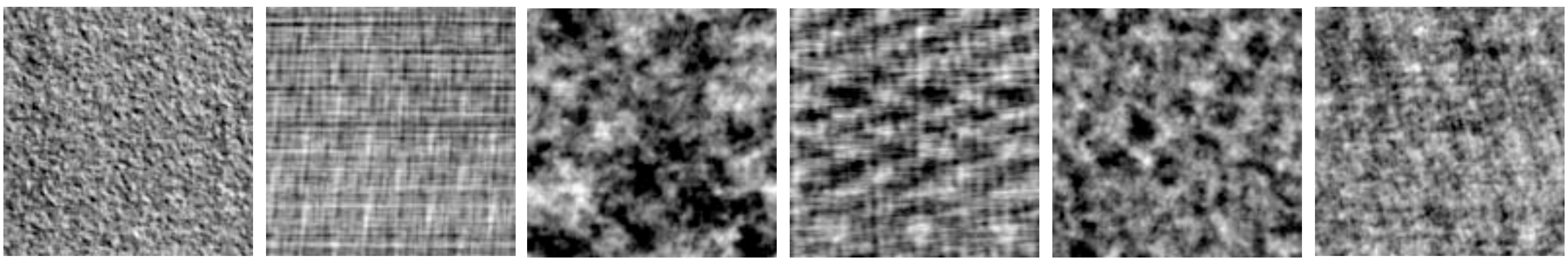
Ergodic Texture Reconstructions

Original Textures

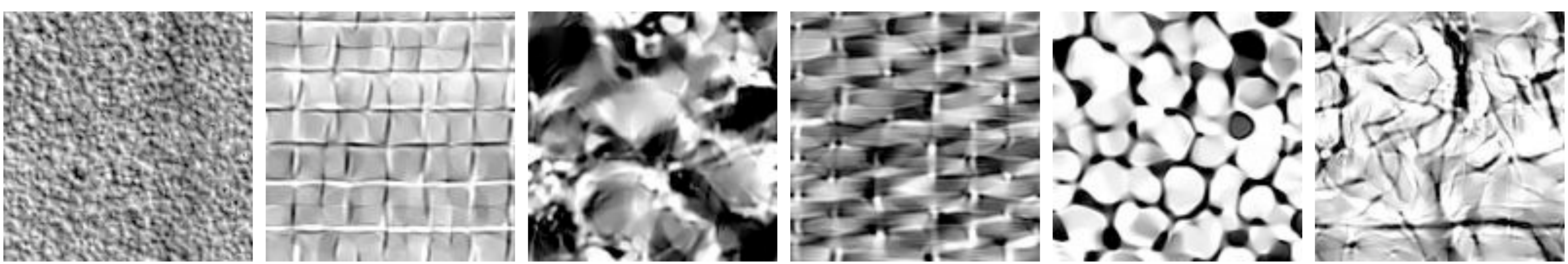
Joan Bruna



Gaussian process model with same second order moments



Reconstruction from $\{\|x\|_1\,,\ \|x \star \psi_{\lambda_1}\|_1\,,\ \||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1\}_{\lambda_1,\lambda_2}$



Multiscale Scattering Reconstructions

Original Images

N^2 pixels

Scattering Reconstruction

$2^J = 16$

$1.4\, N^2$ coeff.

$2^J = 32$

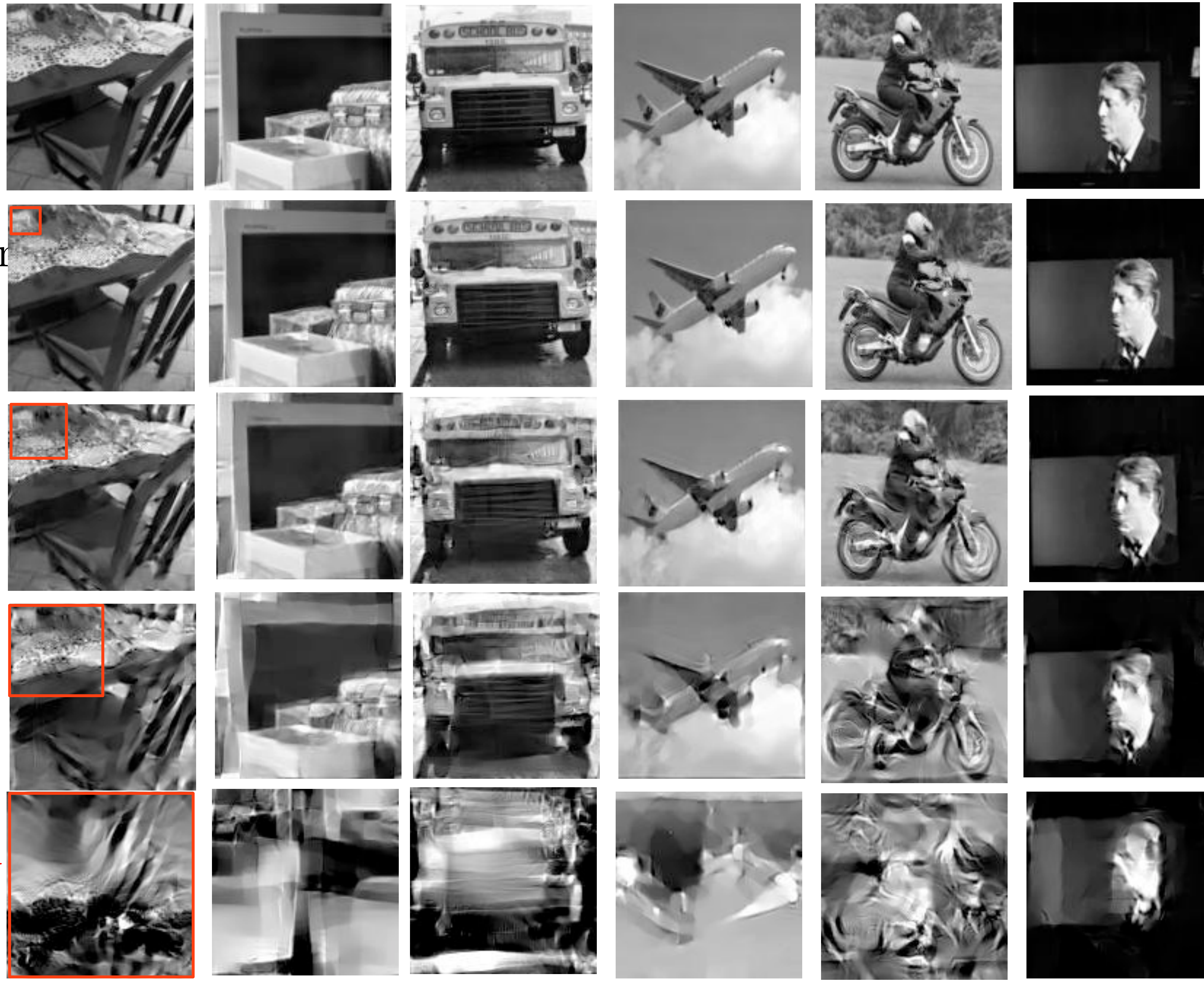
$0.5\, N^2$ coeff.

$2^J = 64$

$N^2/8$ coeff.

$2^J = 128 = N$

$N^2/32$ coeff.



Representation of Audio Textures

Joan Bruna

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

	Original	Gaussian model	Scattering
Water			
Paper			
Cocktail Party			

Learning Physics: N-Body Problem

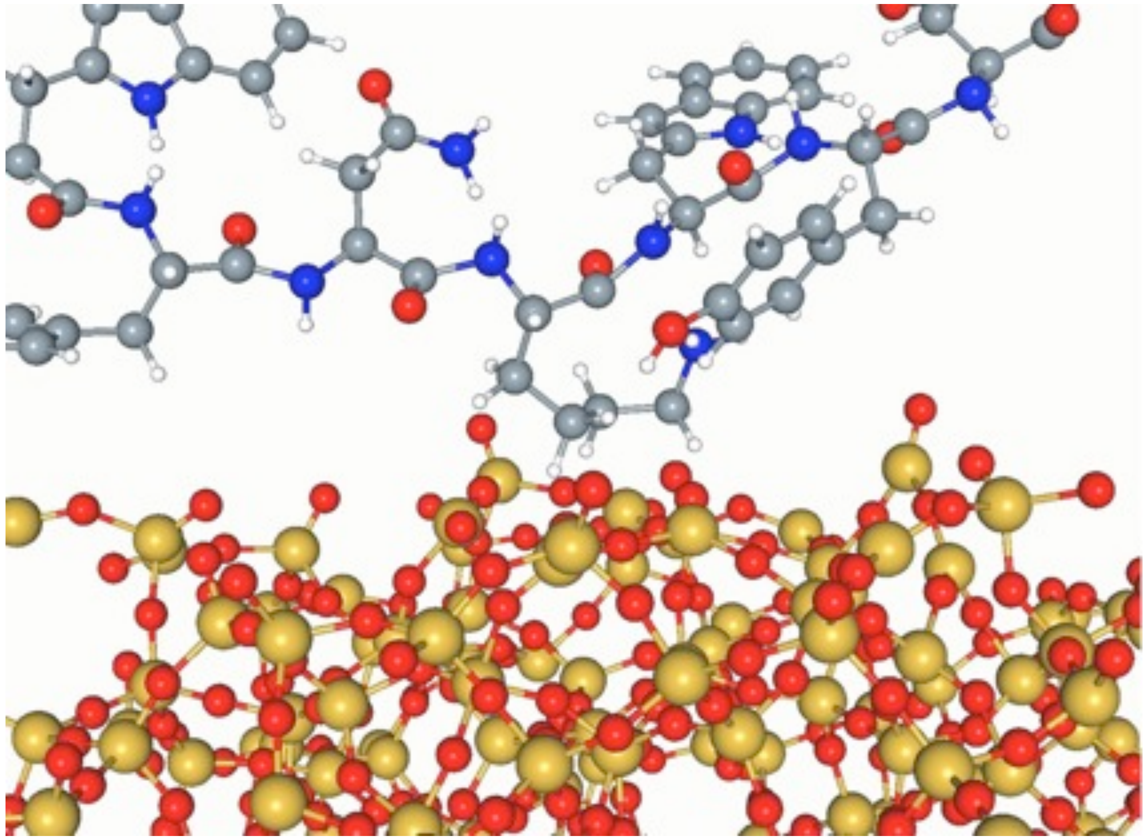
N. Poilvert
Matthew Hirn

- 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



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

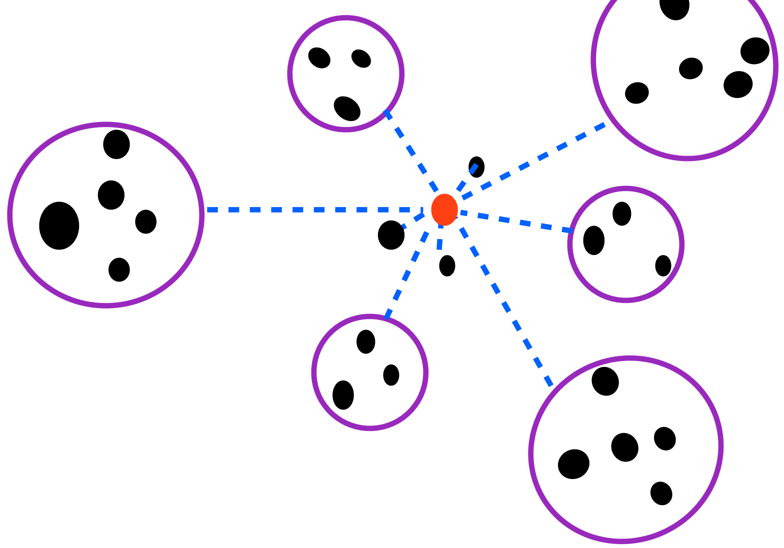
Many Body Interactions

N. Poilvert
Matthew Hirn

- Energy of d interacting bodies (Coulomb):

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

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



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

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

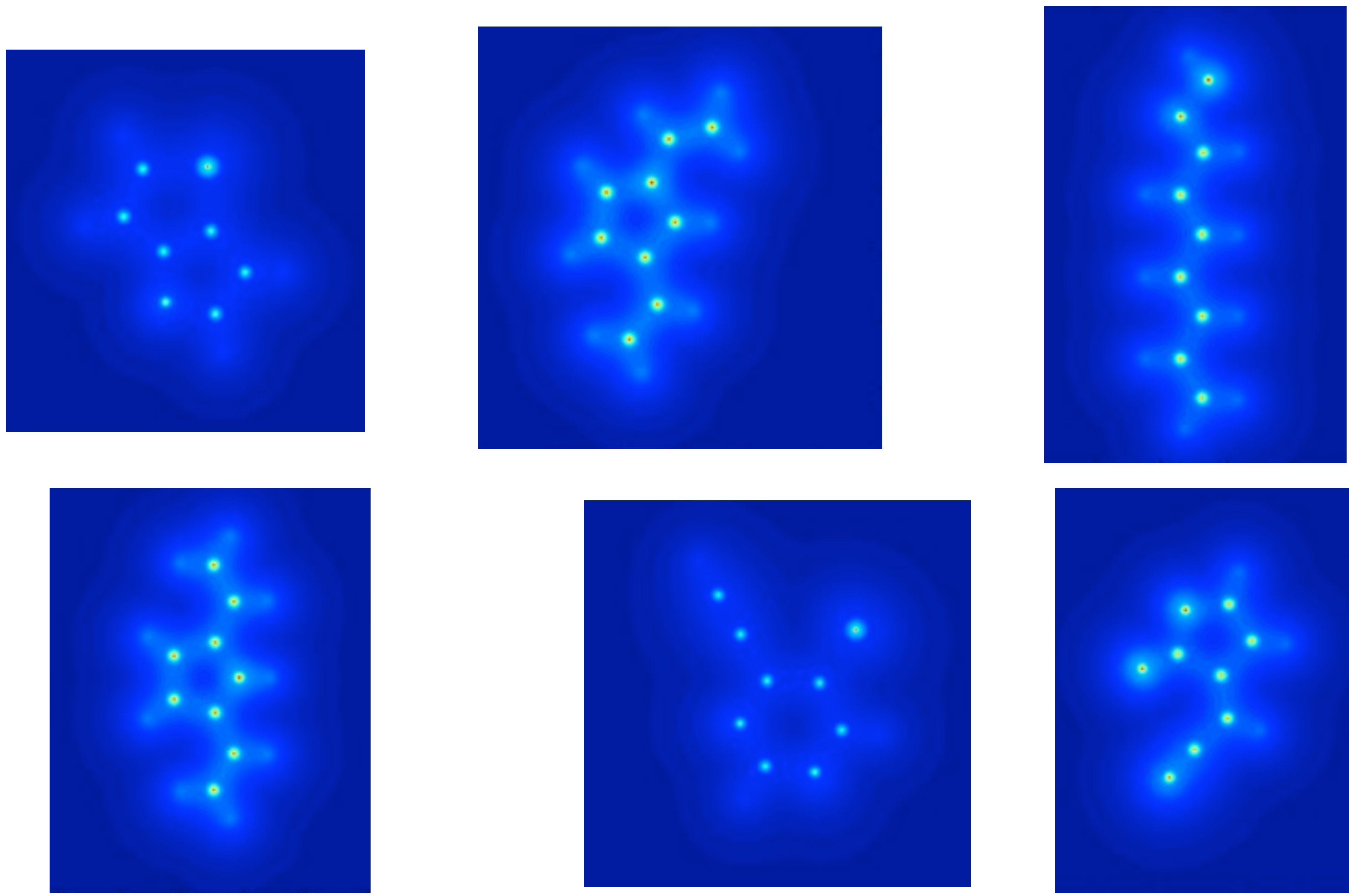
Quantum Chemistry

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 $\phi_k(u)$

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

$$f(\rho(x)) = \underbrace{T(\rho)}_{\text{Kinetic energy}} + \underbrace{\int \rho(u)\, V(u)}_{\text{electron-nuclei attraction}} + \underbrace{\frac{1}{2}\int \frac{\rho(u)\rho(v)}{|u-v|} du dv}_{\text{electron-electron Coulomb repulsion}} + \underbrace{E_{xc}(\rho)}_{\text{Exchange correlat. energy}}$$

- ρ is computed with a variational problem in $O(K^3)$
- Orbitals have "sparse" multiscale wavelet decompositions.
- $f(x)$ is invariant by rigid movements and deformation stable

Quantum Chemistry

N. Poilvert
Matthew Hirn

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

- Regression on scattering coefficients:

$$\Phi x = \{\phi_n(x)\}_n : \left| \begin{array}{c} \text{Fourier modulus coefficients and squared} \\ \text{or} \\ \text{order 2 scattering coefficients and squared} \end{array} \right.$$

M-term sparse regression with a greedy Partial Least Square:
computed on training set:

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

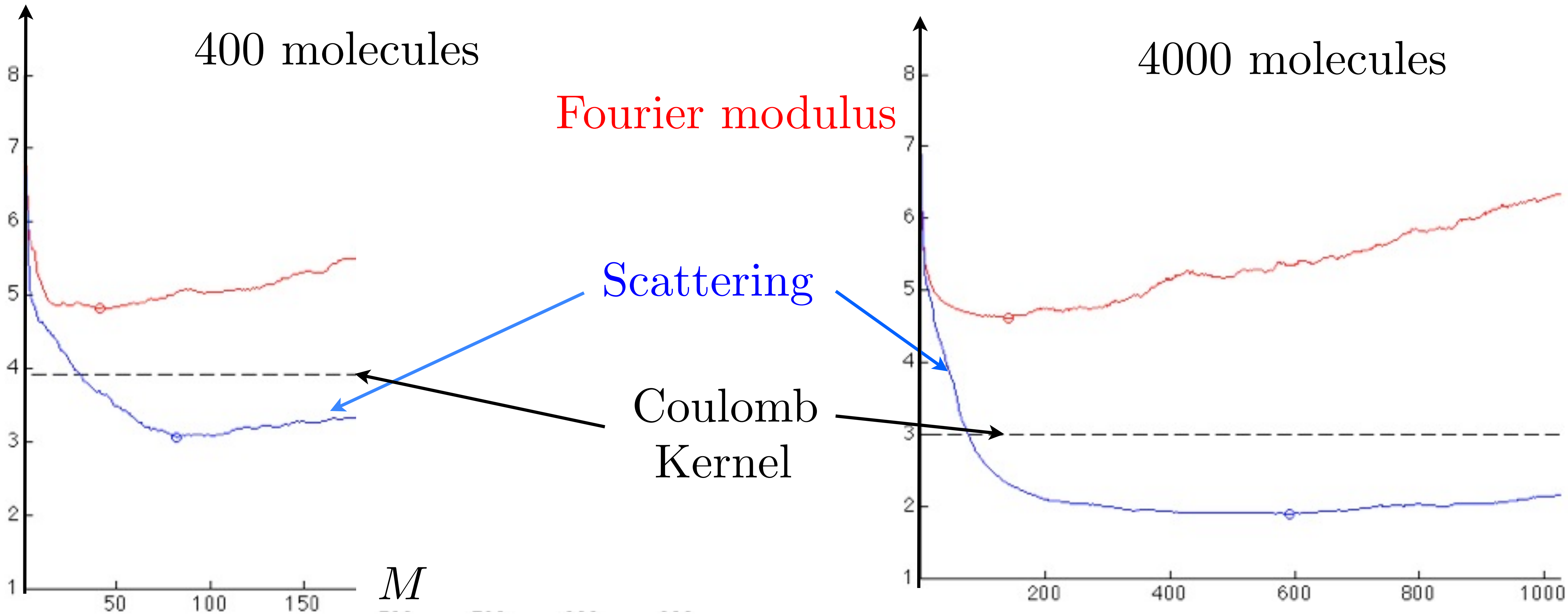
Quantum Chemistry

N. Poilvert
Matthew Hirn

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

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

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



Quantum Chemistry

N. Poilvert

Matthew Hirn

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

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

Mean-square error $\mathbb{E}(|f(X) - f_M(X)|^2)^{1/2}$ in kcal/mol

	Fourier	Coulomb	Scattering
400 atoms	30	15	**8**
4000 atoms	24	8	**3.7**

WHY ?

First terms of scattering expansions:

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

$\phi_{n_2}(x) = \|x \star \psi_{\lambda_1}\|_1$: where λ_1 is the main geometric scale

Conclusion

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

- One can learn physics through data and compute fast

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

www.di.ens.fr/data/scattering