High Dimensional Classification with Deep Networks

Joan Bruna, Stéphane Mallat,
Edouard Oyallon, Laurent Sifre

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

High-Dimensional Classification

- Estimate a label $y(x)$ of $x \in \mathbb{R}^d$ given examples $\{x_i, y_i\}_i$

- In high dimension $\|x - x'\|$ is not a good similarity measure

- Compute $\Phi x \in \mathbb{R}^D$ so that $\|\Phi x - \Phi x'\|$ measures similarity

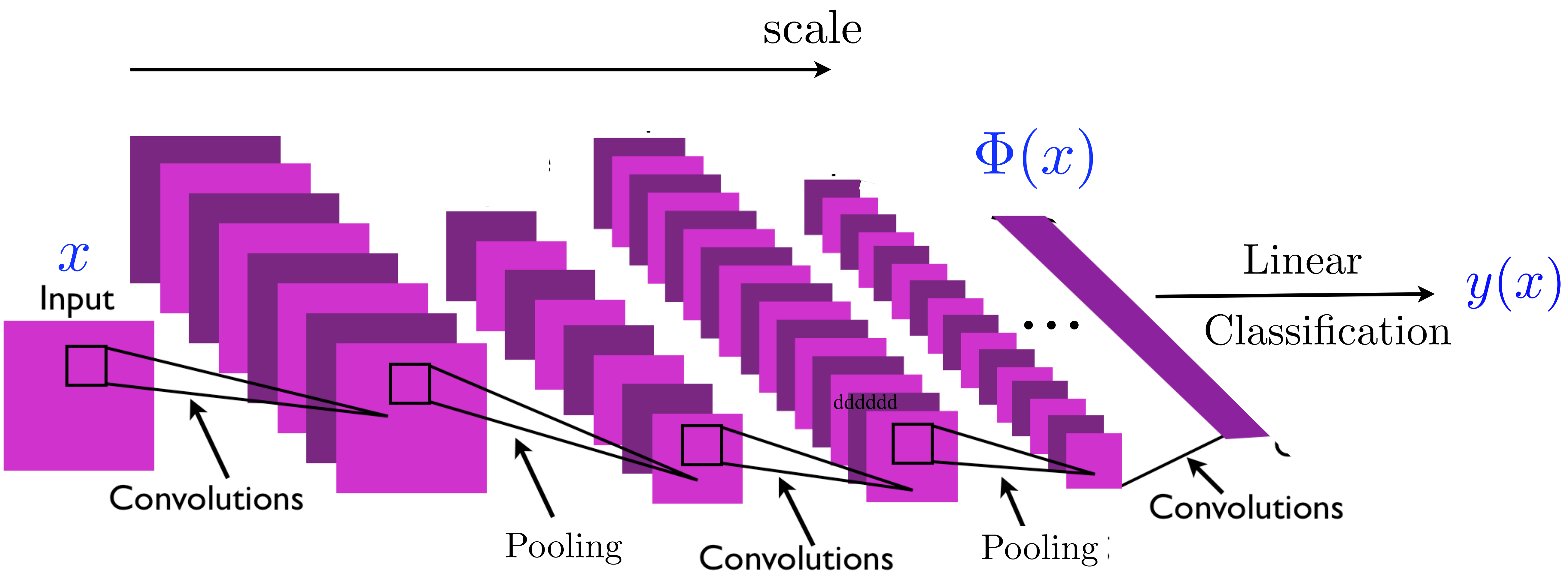
  then a linear classifier applied to Φx is highly effective.

$$x \longrightarrow \boxed{\Phi} \longrightarrow \Phi x \longrightarrow \boxed{\begin{array}{c}\text{Supervised}\\ \text{Linear Classifier}\end{array}} \longrightarrow \tilde{y}(x)$$

- How to define Φ ? Should we learn it ?

Deep Convolution Networks

Y. LeCun 1989



Millions to Billions of parameters

Supervised learning of filter coefficients

Genericity: one network (Alex net) yields state of the art on very different image classification problems.

Overview of Outrageous Claims

- No need to learn deep net for structured signals (images) just wavelet filters derived from geometry.

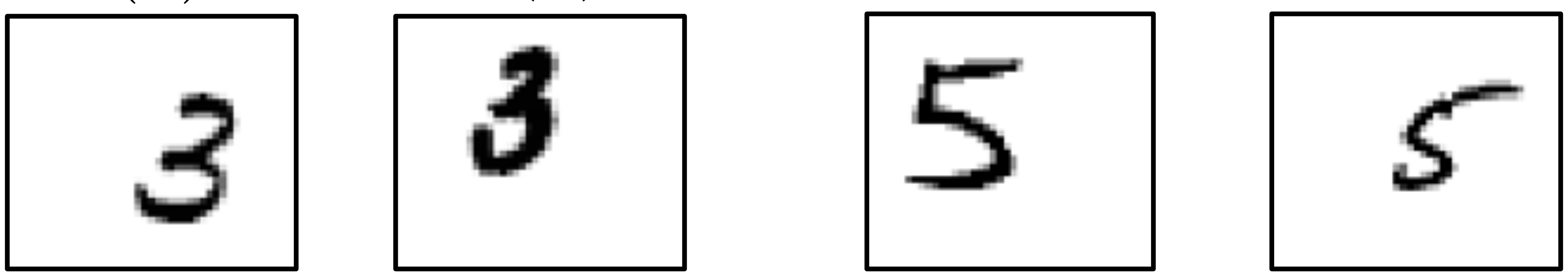
- Deep wavelet networks are signal coders.

- One can learn physical interactions: quantum chemistry.

Image Metrics

- Low-dimensional "geometric shapes"

$x(u)$ $\quad$ $x'(u)$



Deformation metric: *Grenander, Trouvé, Younes*

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

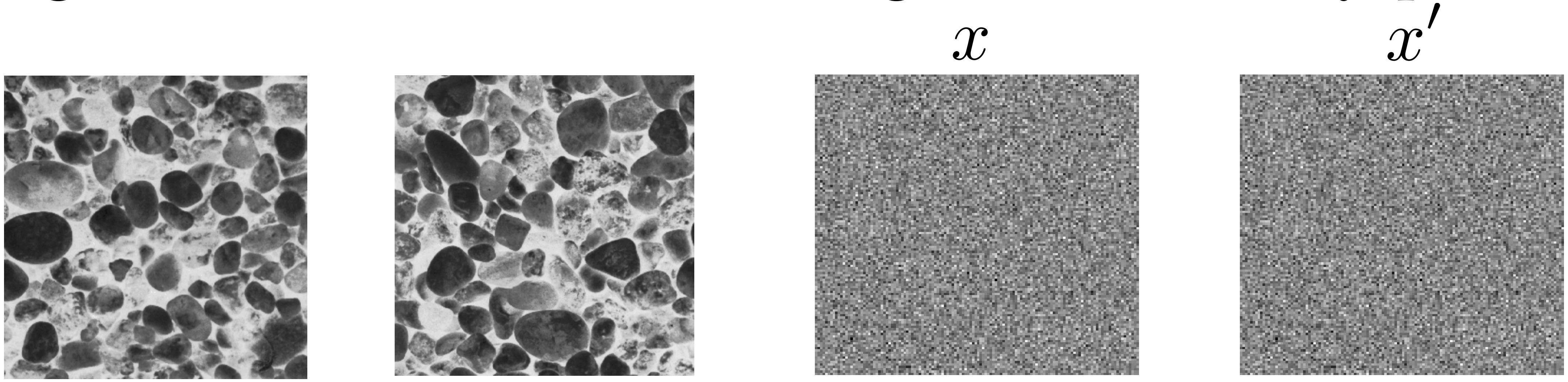
$$\Delta(x, x') \sim \min_\tau \|D_\tau x - x'\| + \|\nabla \tau\|_\infty \|x\|$$

Invariant to translations

diffeomorphism amplitude

Image Metrics

- High dimensional textures: ergodic stationary processes

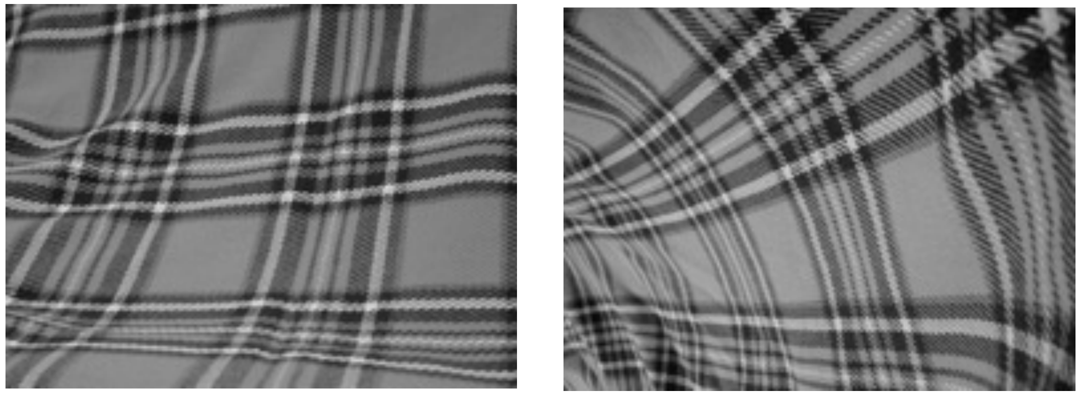


What metric on stationary processes ?

- Invariant to translations and stable to deformations

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

Reverse inequality is wrong



- $\Delta(x', x) = 0$ for realisations of a "same stationary process"

Image Geometry and Metric

- High dimensional ”structured” images




What metric on images ?

- Invariant to translations and stable to deformations

- What else ?

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

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

$\Rightarrow$ Invariance to translation

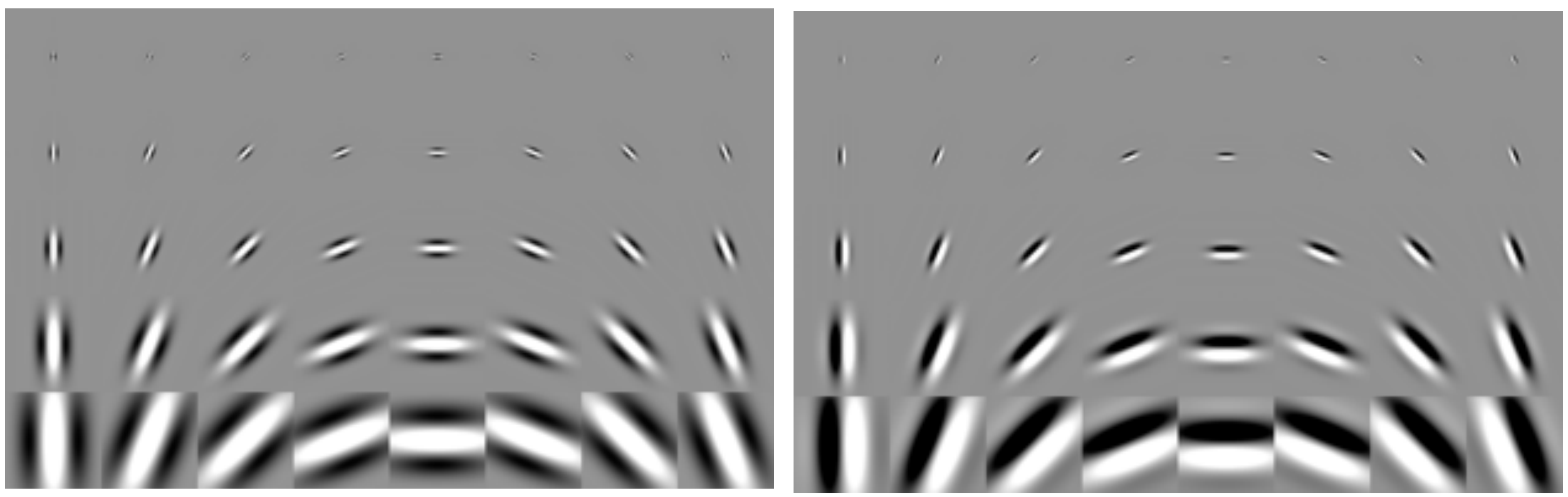
Failure of classical math invariants: Fourier, canonical...

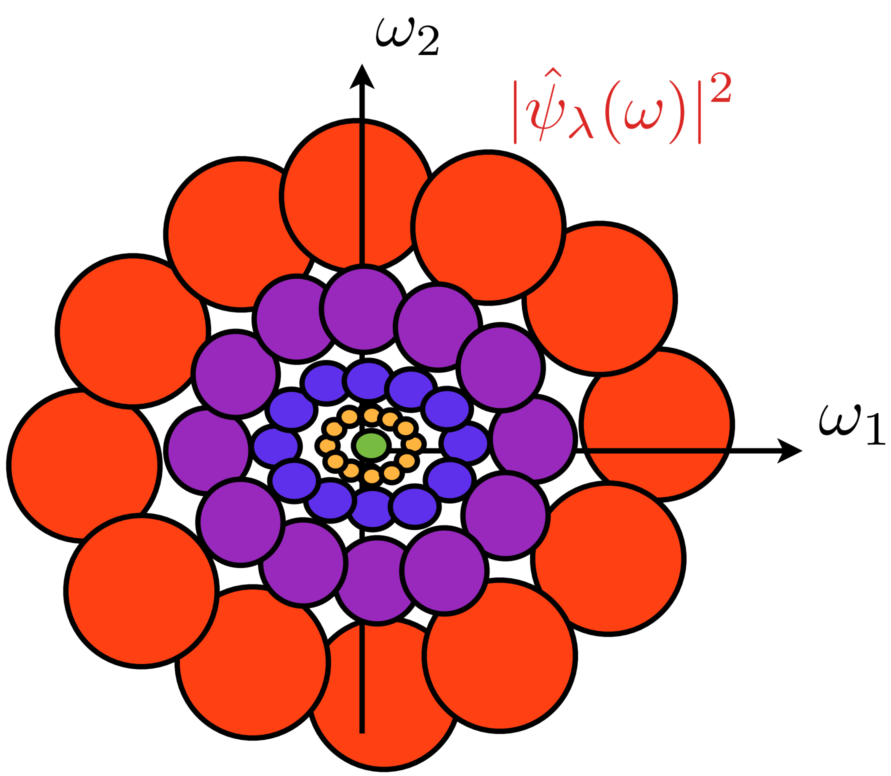
Wavelet Transform of Images

- 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

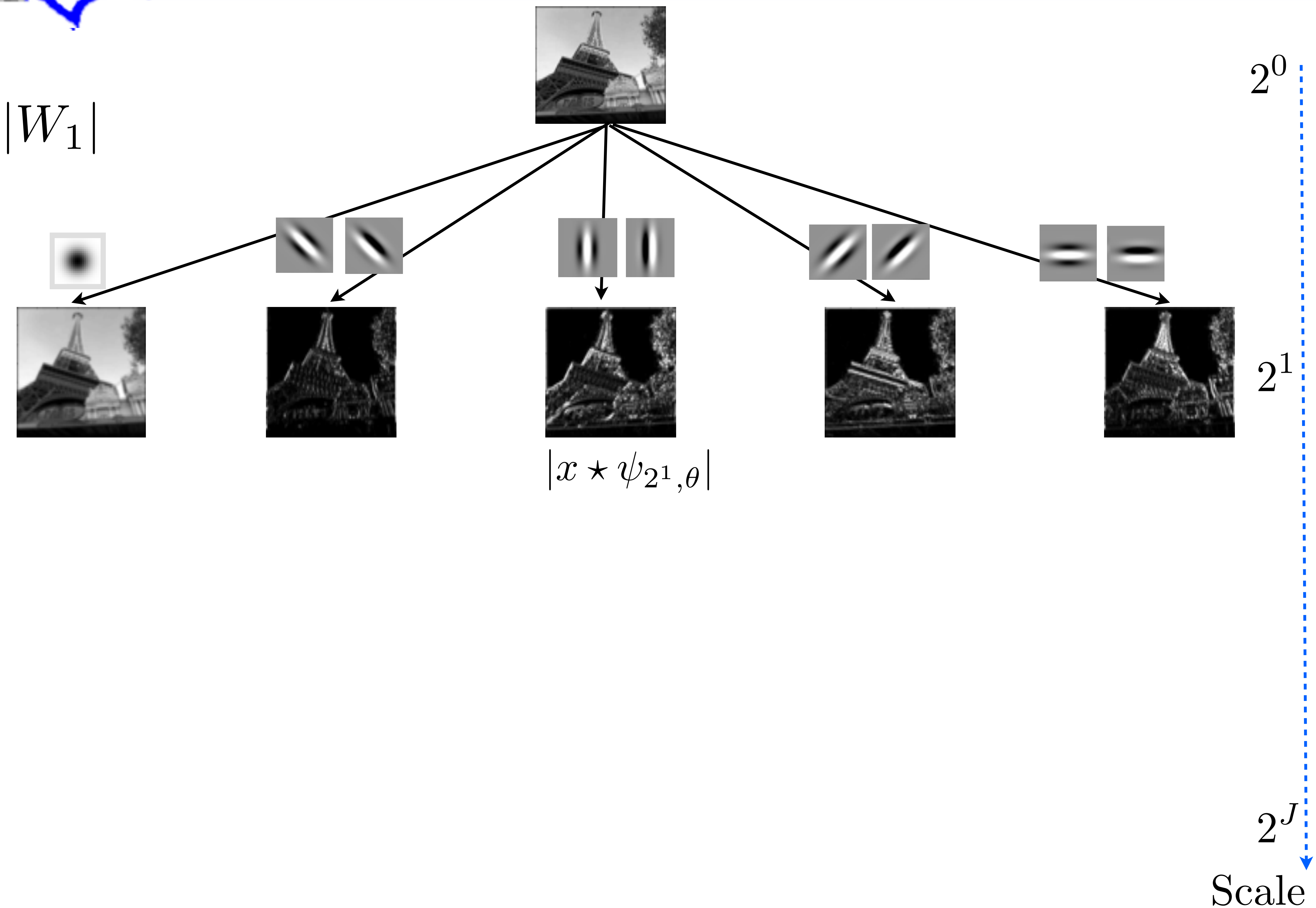




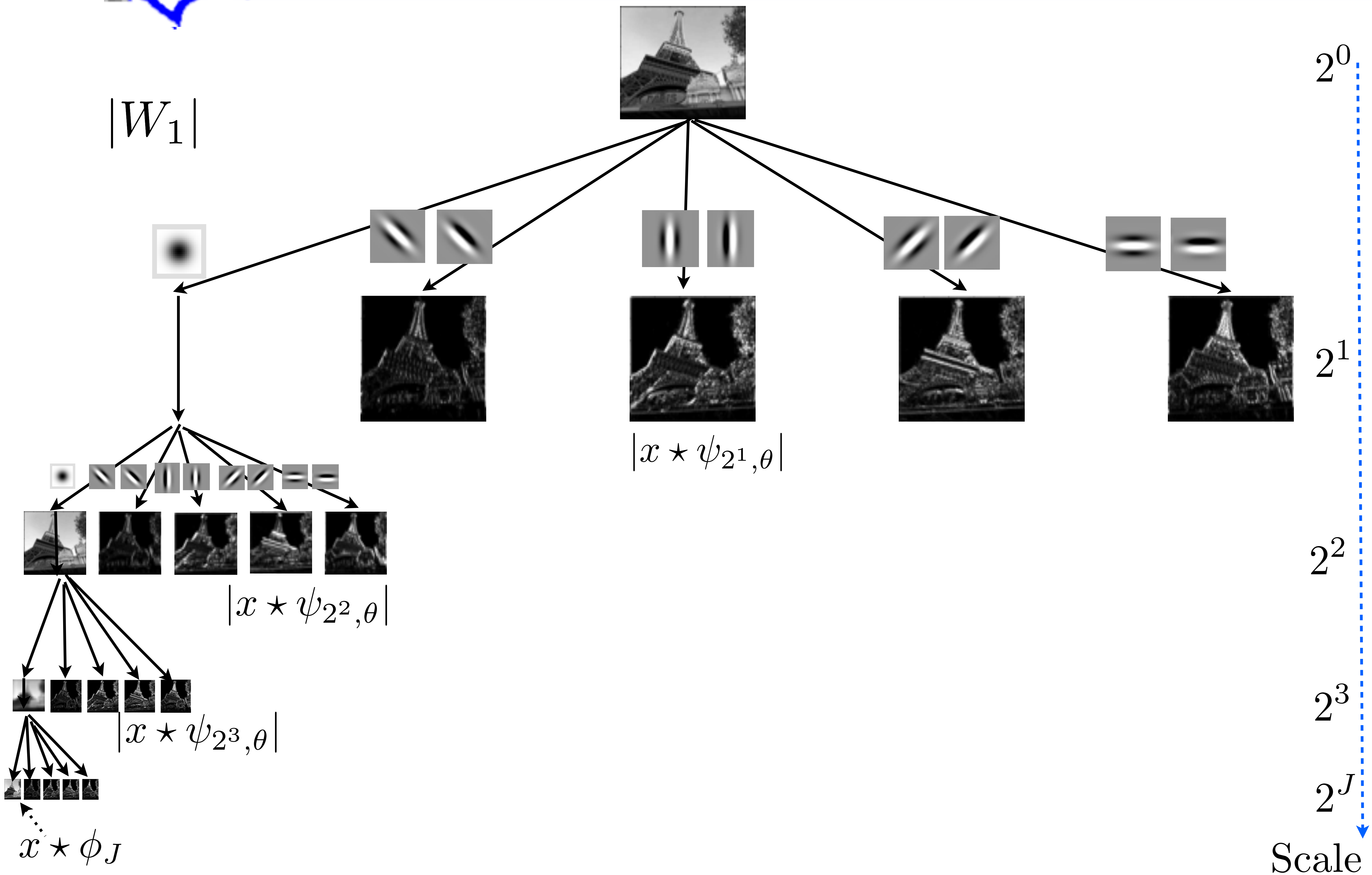
- Wavelet transform: $Wx = \left(\begin{array}{c} x \star \phi_{2^J}(t) \\ x \star \psi_\lambda(t) \end{array} \right)_{\lambda \leq 2^J}$

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

Fast Wavelet Transform



Fast Wavelet Transform



Wavelet Translation Invariance

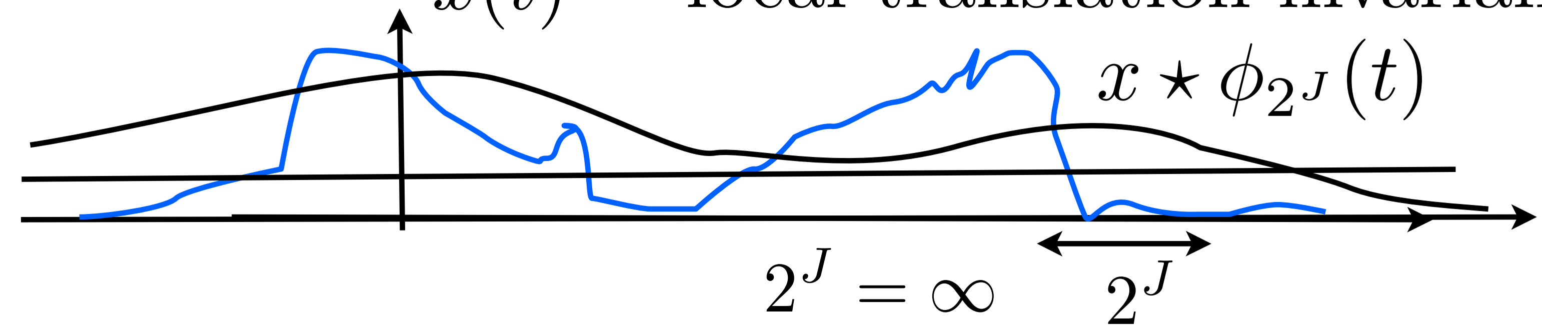
First wavelet transform

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

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

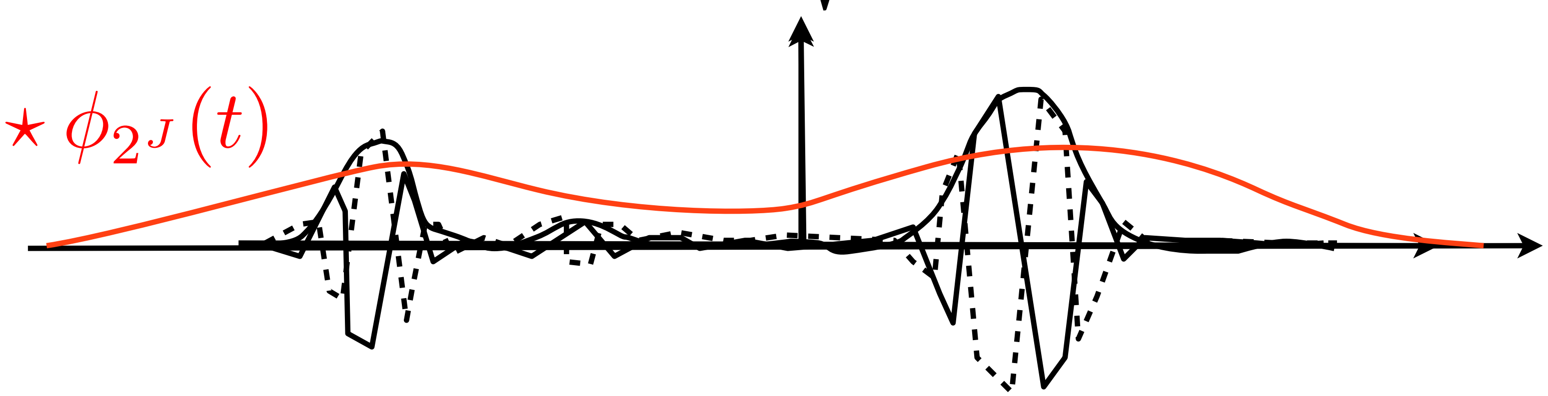
full translation invariance

local translation invariance



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

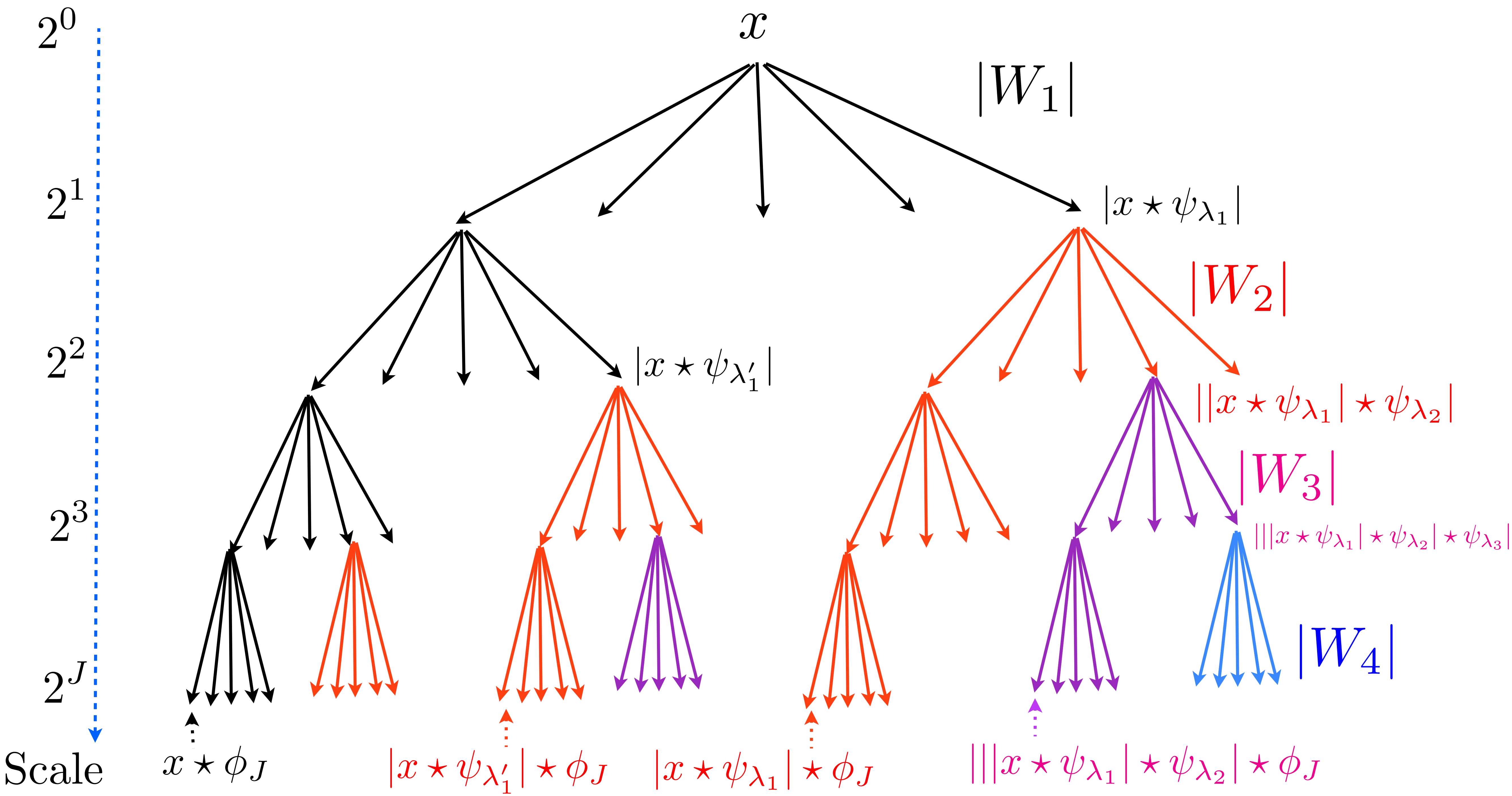
Modulus improves invariance: $|x \star \psi_{\lambda_1}(t)| = \sqrt{|x \star \psi^a_{\lambda_1}(t)|^2 + |x \star \psi^b_{\lambda_1}(t)|^2}$



Second wavelet transform modulus

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

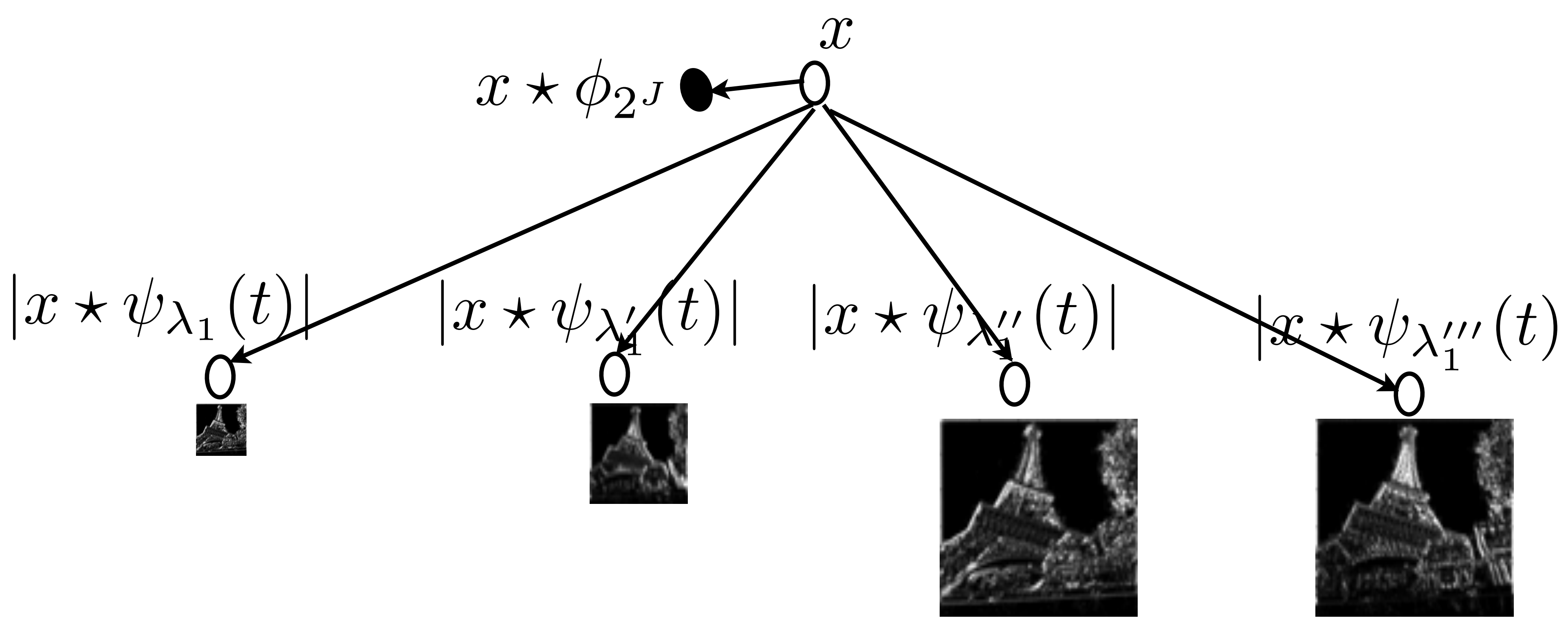
Wavelet Scattering Network



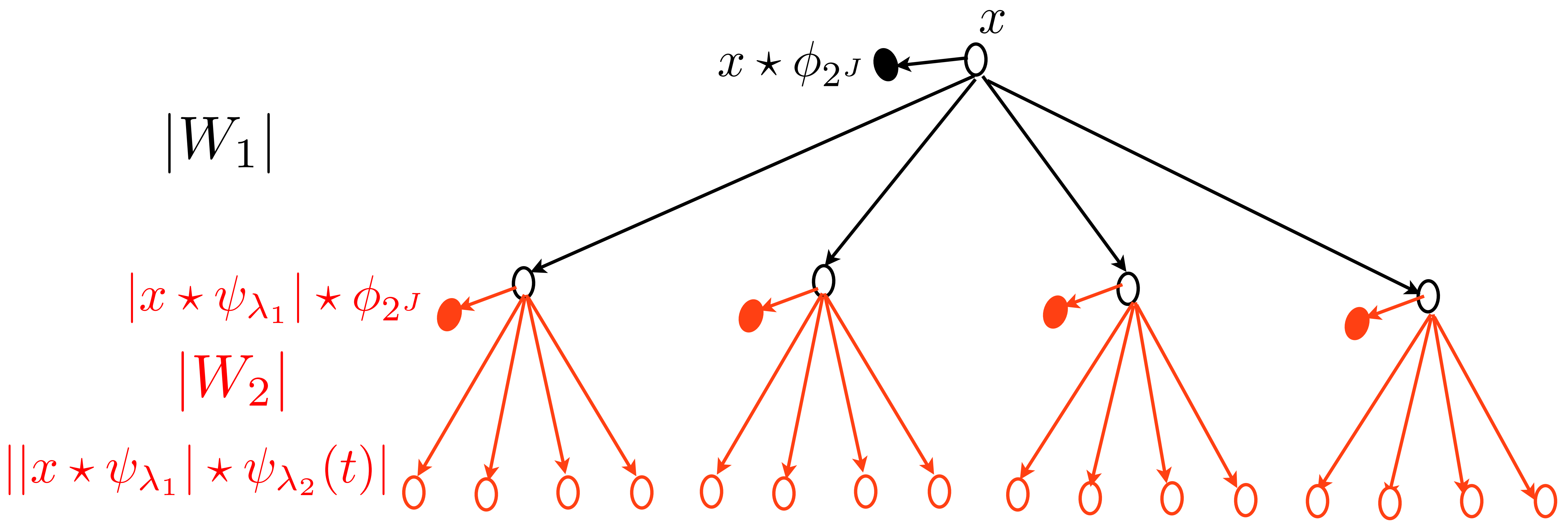
$$S_J x = |W_J| \ldots |W_4|\, |W_3|\, |W_2|\, |W_1|\; x$$

Scattering Neuronal Network

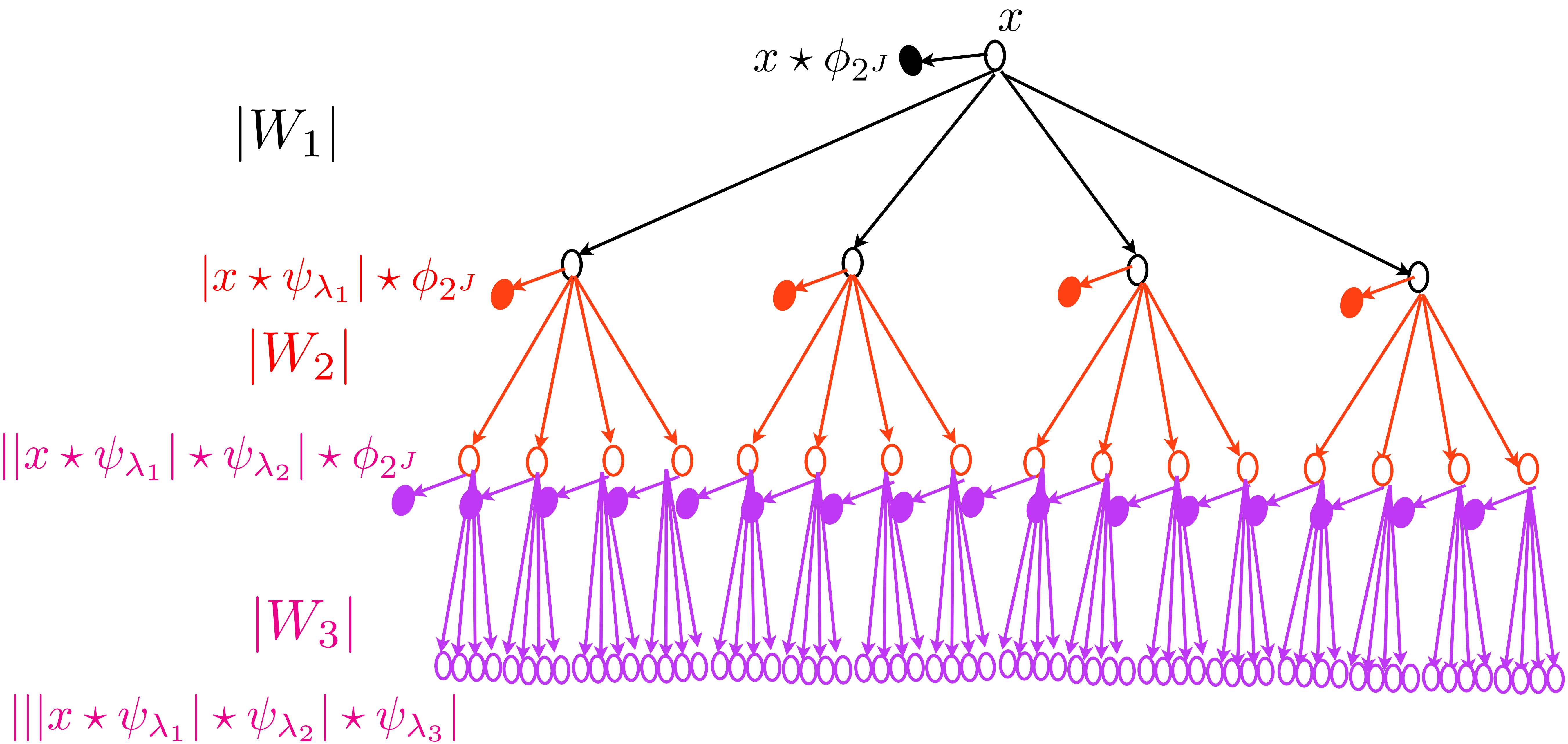
$|W_1|$



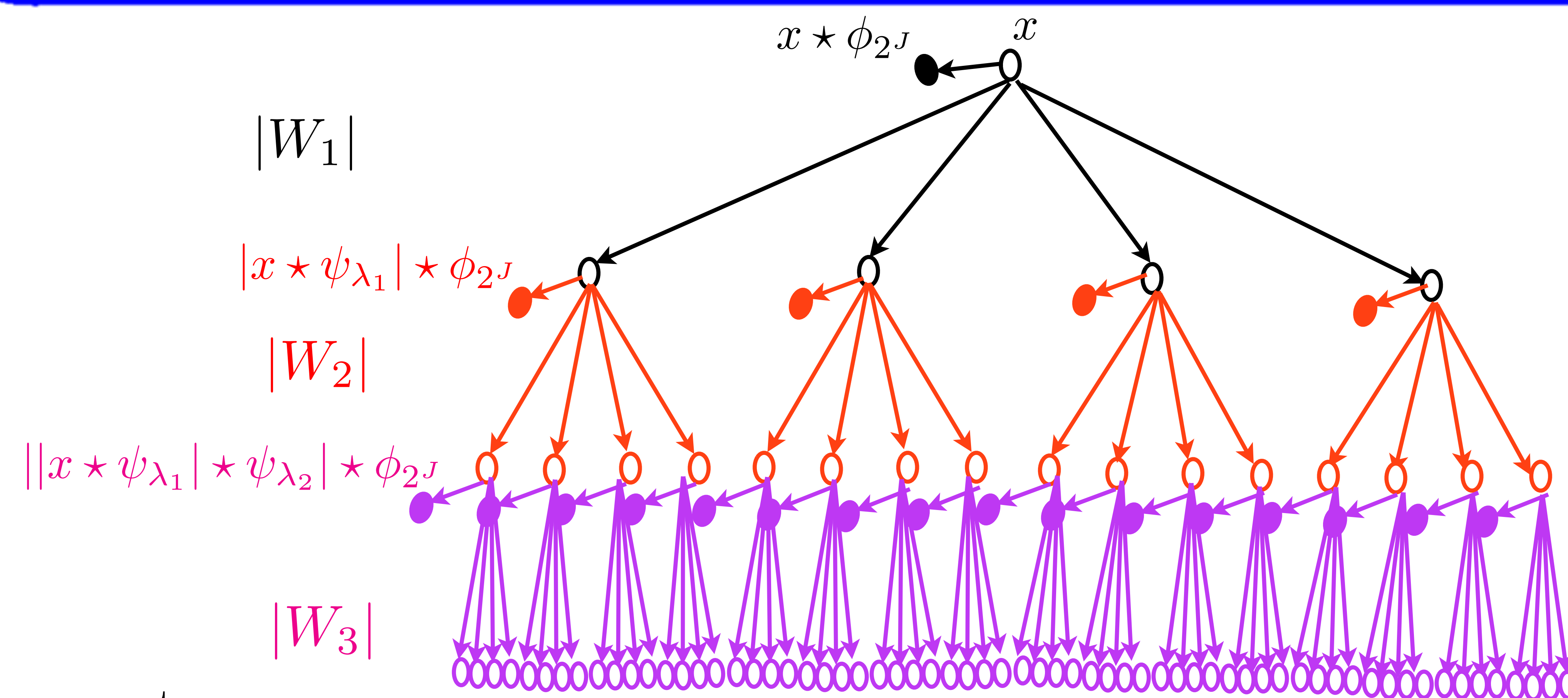
Scattering Neuronal Network



Scattering Neuronal Network



Wavelet Scattering



Scattering operator:

$$S_J x = \begin{pmatrix} x \star \phi_{2^J} \\ |x \star \psi_{\lambda_1}| \star \phi_{2^J} \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_{2^J} \\ |||x \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \star \phi_{2^J} \\ \cdots \end{pmatrix}_{\lambda_1,\lambda_2,\lambda_3} \xrightarrow[J \to \infty]{} \begin{pmatrix} \int x(u)du \\ \|x \star \psi_{\lambda_1}\|_1 \\ \||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1 \\ \|||x \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}\|_1 \\ \cdots \end{pmatrix}_{\lambda_1,\lambda_2,\lambda_3}$$

Theorem: The total energy of coefficients converge to 0 as the depth (number of modulus) increases.

Scattering Properties

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

Theorem: *For appropriate wavelets, a scattering is*

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

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

stable to deformations $D_\tau x(u) = x(u - \tau(u))$

$$\|S_J D_\tau x - S_J x\| \leq C \left(\|\nabla \tau\|_\infty \, \|x\| + 2^{-J} \|\tau\|_\infty \right)$$

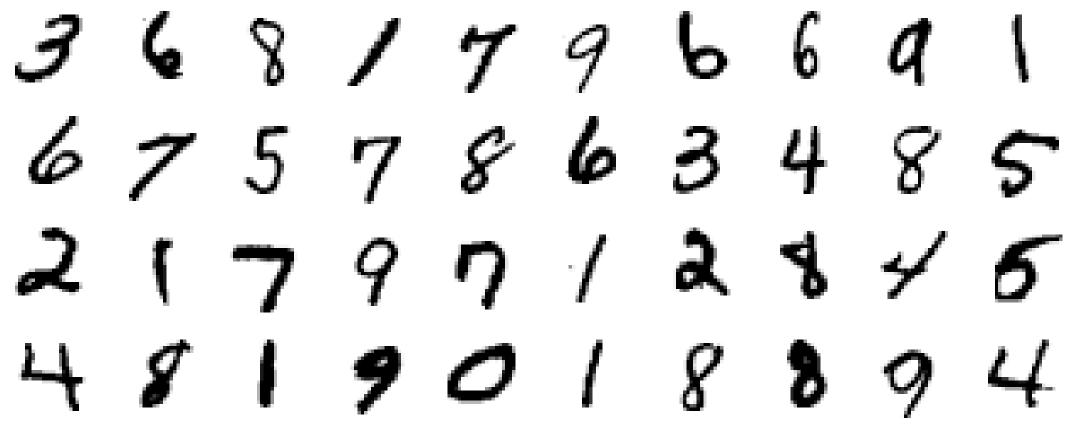
$$\underset{J \to \infty}{\Rightarrow} \|Sx - Sx'\| \leq C \left(\min_\tau \|x - D_\tau x'\| + \|\nabla \tau\|_\infty \, \|x\| \right)$$

Image Classification

Joan Bruna

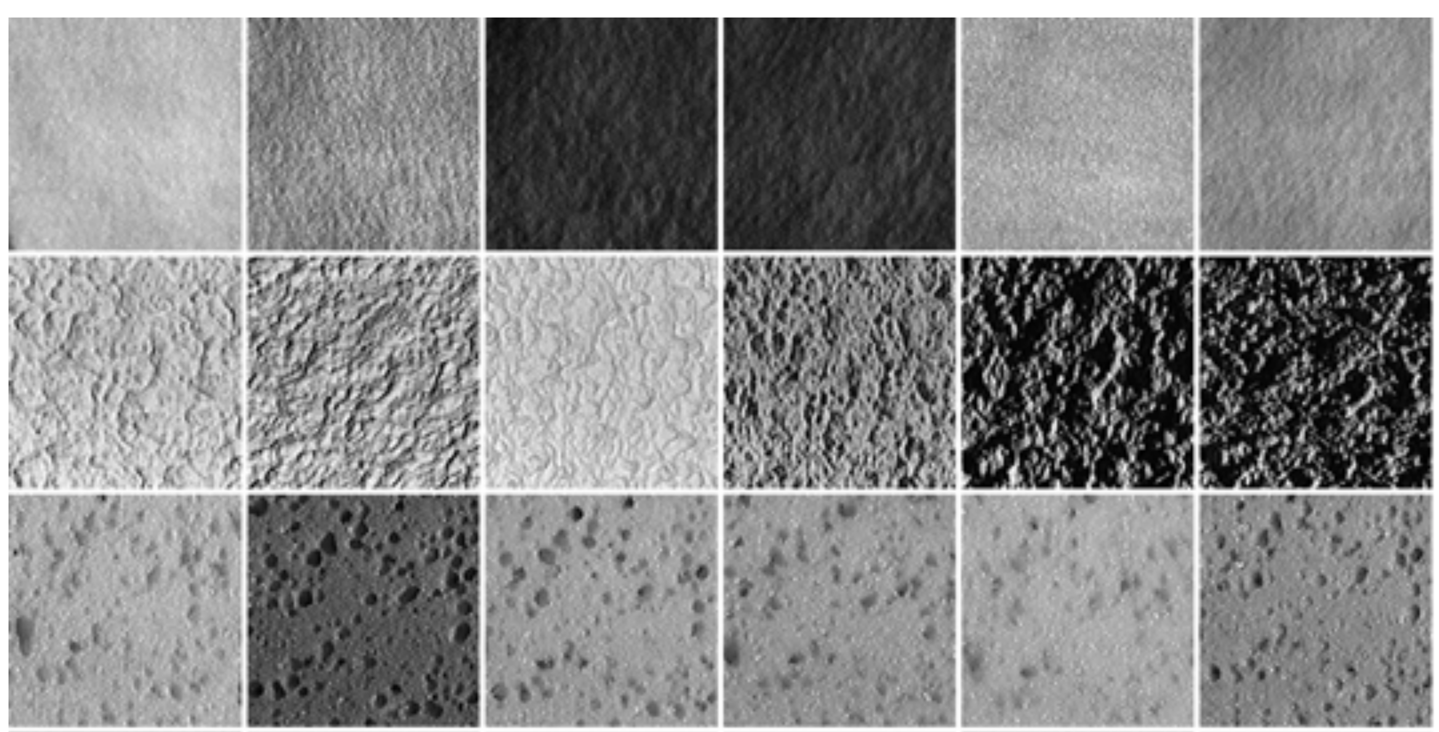
$$x \longrightarrow \boxed{S_J x} \longrightarrow \boxed{\text{Linear Classif. SVM}} \longrightarrow y$$

MNIST:
$6\,10^4$ chiffres



$\mathbf{0.4\%}$ *errors*
$2^J = 2^3$

CUREt
61 classes



$\mathbf{0.2\%}$ *errors*
$2^J = \text{image size}$

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 $X(t)$

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

and $SX \underset{J \to \infty}{\longrightarrow} \overline{S}X$ if X is ergodic.

- But does $\overline{S}X$ "characterize" X ?

Adapt Convolutions to Invariants

Laurent Sifre

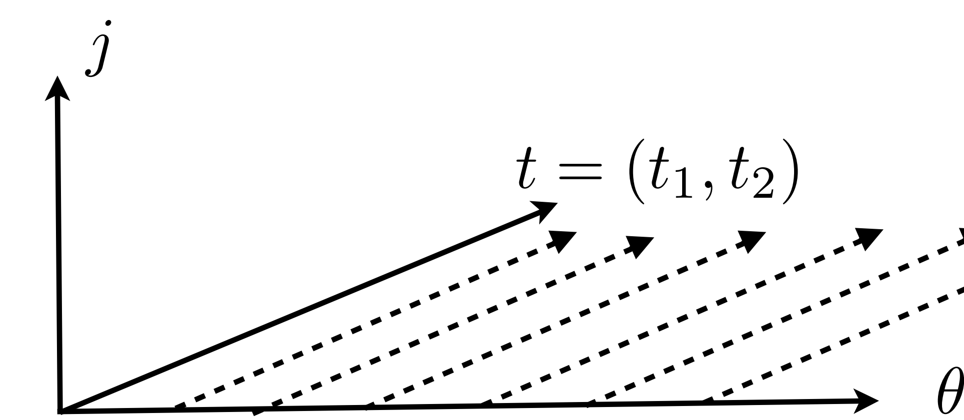
translation — translation

1st order

$$x(t) \longrightarrow \boxed{|W_1|} \longrightarrow |x \star \psi_{j,\theta}(t)| = x_1(j,\theta,t) \longrightarrow \boxed{|W_2|} \longrightarrow |x_1(j,\theta,.) \star \psi_{j'}(t)|$$

W_2 computes wavelet convolutions along (t_1, t_2)

4D space



Rotation-Translation Invariance

Laurent Sifre

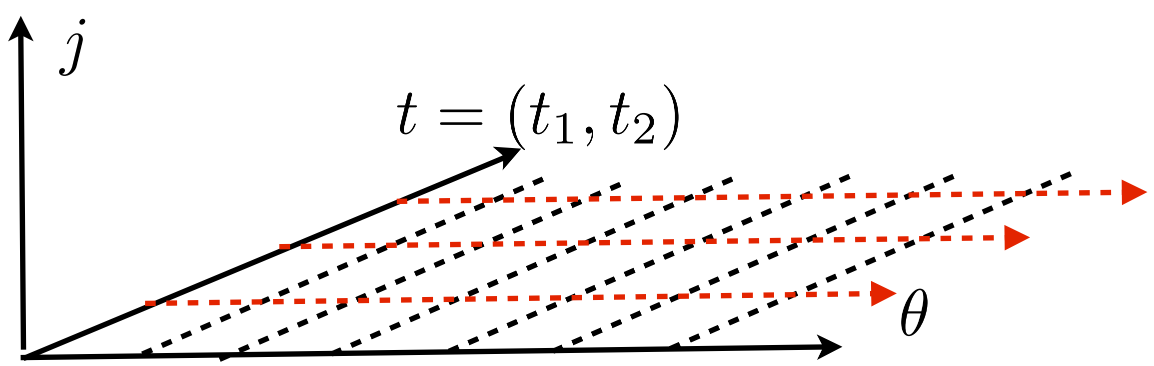
translation — roto-translation

1st order

$$x(t) \rightarrow \boxed{|W_1|} \rightarrow |x \star \psi_{j,\theta}(t)| = x_1(j,\theta,t) \rightarrow \boxed{|W_2|} \rightarrow |x_1 \star \overline{\psi}_{j',l'}(j,\theta,t)|$$

W_2 computes wavelet convolutions along (t_1, t_2, θ)

4D space



Scalo-Roto-Translation Invariance

Laurent Sifre

translation

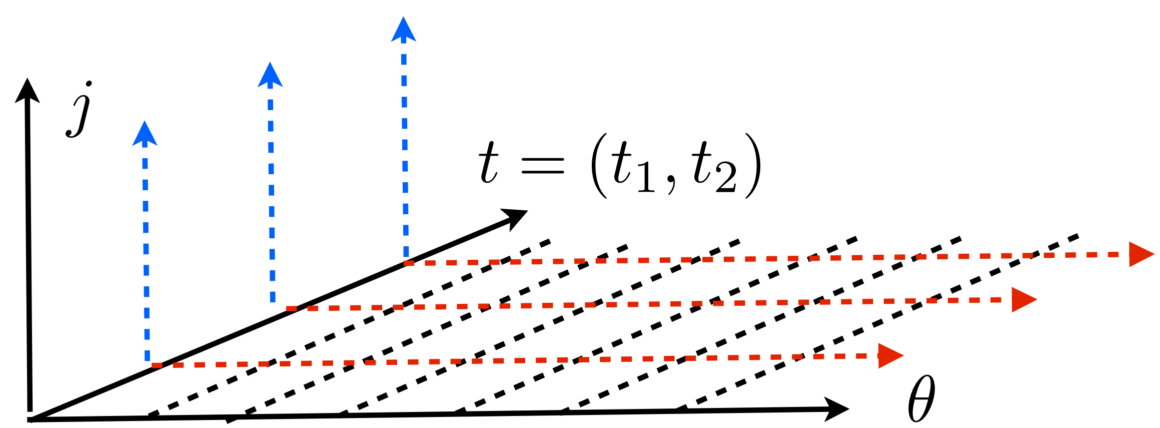
1st order

scalo-roto-translation

$$x(t) \longrightarrow \boxed{|W_1|} \longrightarrow |x \star \psi_{j,\theta}(t)| = x_1(j,\theta,t) \longrightarrow \boxed{|W_2|} \longrightarrow |x_1 \star \overline{\psi}_{j',l'}(j,\theta,t)|$$

W_2 computes wavelet convolutions along (t_1, t_2, θ, j)

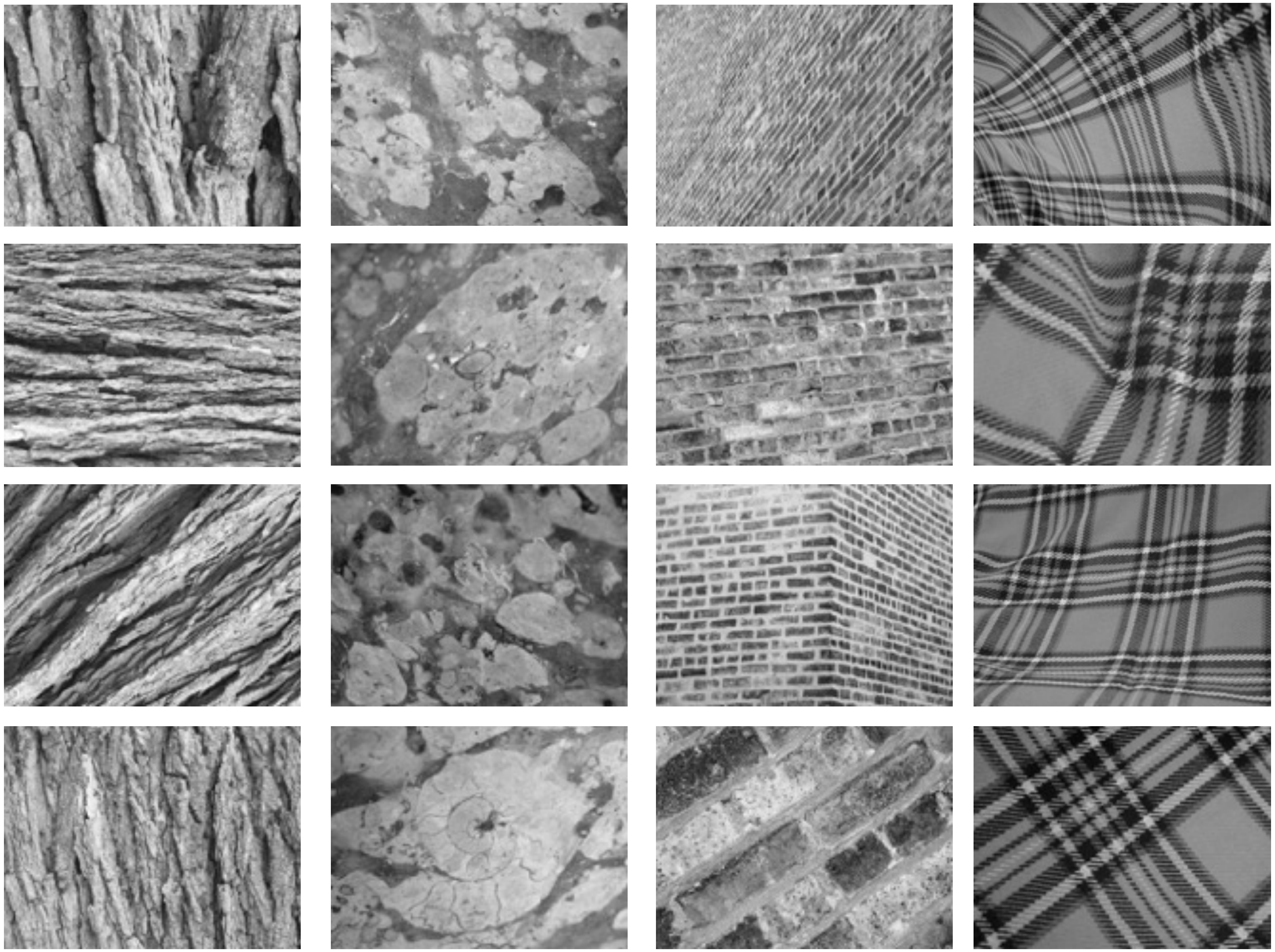
4D space



Rotation and Scaling Invariance

Laurent Sifre

UIUC database:
25 classes



Scattering classification errors

Training	Translation	Transl + Rotation	+ Scaling
20	20 %	2%	**0.6%**

Complex Image Classification

CalTech 101 data-basis: *Edouard Oyallon*

$$x \longrightarrow \boxed{\begin{array}{c} S_J x \\ \text{Roto-Trans.} \end{array}} \longrightarrow \boxed{\text{Linear Classif.}} \longrightarrow y$$

Classification Accuracy $2^J = 2^5$

Data Basis	2012	Deep-Net	Scat.-1	Scat.-2
CalTech-101	80%	**85%**	50%	80%
CalTech-256	50%	**70%**	30%	50%
CIFAR-10	80%	**90%**	55%	80%

Complex Image Classification

CalTech 101 data-basis:

Edouard Oyallon

$$x \longrightarrow \boxed{\begin{array}{c} S_J x \\ \text{Roto-Trans.} \end{array}} \longrightarrow \boxed{\text{Linear Classif.}} \longrightarrow y$$

Classification Accuracy $2^J = 2^5$

Data Basis	2012	Deep-Net	Scat.-2
CalTech-101	80%	**85%**	80%
CalTech-256	50%	**70%**	50%
CIFAR-10	80%	**90%**	80%

Inverse Scattering Transform

Joan Bruna

- Given $S_J x$ we want to compute $\tilde{x}$ such that:

$$S_J \tilde{x} = \begin{pmatrix} \tilde{x} \star \phi_{2^J} \\ |\tilde{x} \star \psi_{\lambda_1}| \star \phi_{2^J} \\ \dots \\ |||\tilde{x} \star \psi_{\lambda_1}| \star ..| \star \psi_{\lambda_m}| \star \phi_{2^J} \end{pmatrix}_{\lambda_1,\dots,\lambda_m} = \begin{pmatrix} x \star \phi_{2^J} \\ |x \star \psi_{\lambda_1}| \star \phi_{2^J} \\ \dots \\ |||x \star \psi_{\lambda_1}| \star ..| \star \psi_{\lambda_m}| \star \phi_{2^J} \end{pmatrix}_{\lambda_1,\dots,\lambda_m} = S_J x$$

with $\|\tilde{x}\|$ minimum. Non convex optimisation problem.

- For $m = 1$ and $2^J = \infty$, minimize $\|\tilde{x}\|$ subject to:

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

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

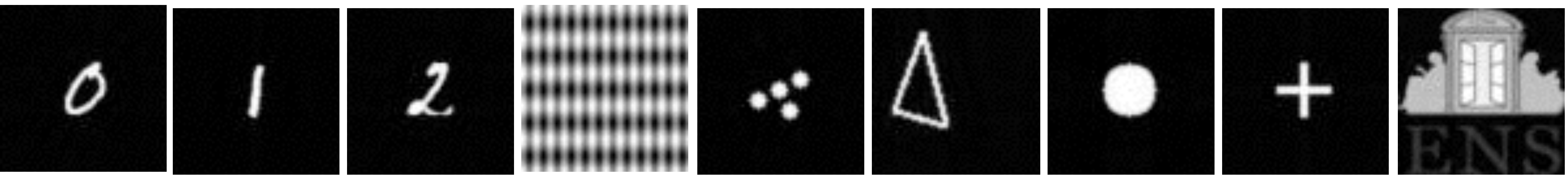
- If $x(u)$ is a Dirac, or a straight edge or a sinusoid then $\tilde{x}$ is equal to x up to a translation.

Sparse Shape Reconstruction

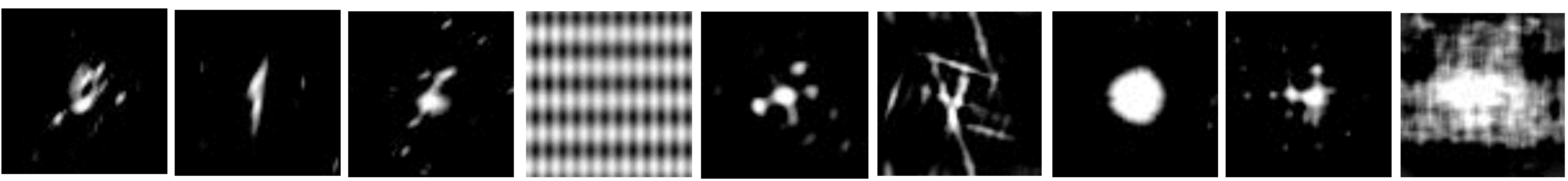
Joan Bruna

With a gradient descent algorithm:

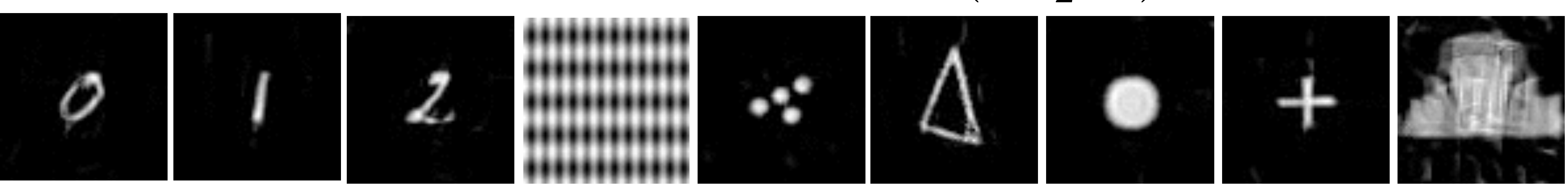
Original images of N^2 pixels:



$m = 1,\ 2^J = N$: reconstruction from $O(\log_2 N)$ scattering coeff.



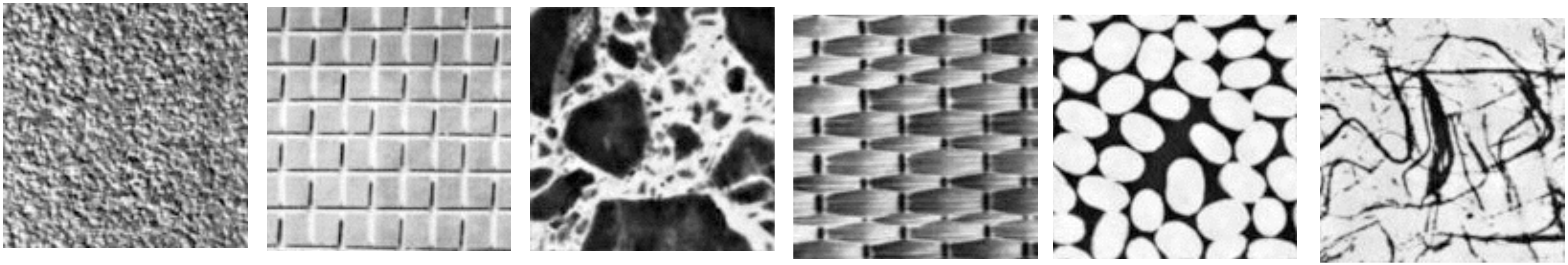
$m = 2,\ 2^J = N$: reconstruction from $O(\log_2^2 N)$ scattering coeff.



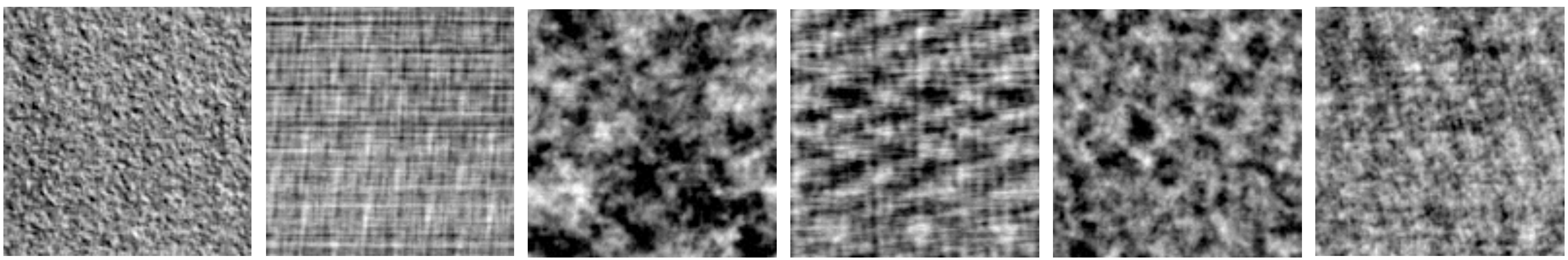
Ergodic Texture Reconstructions

Joan Bruna

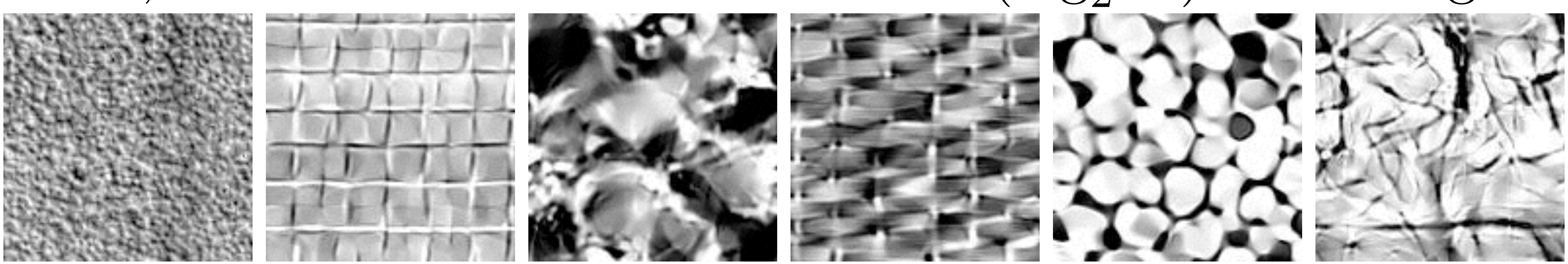
Original Textures



Gaussian process model with same second order moments



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



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.

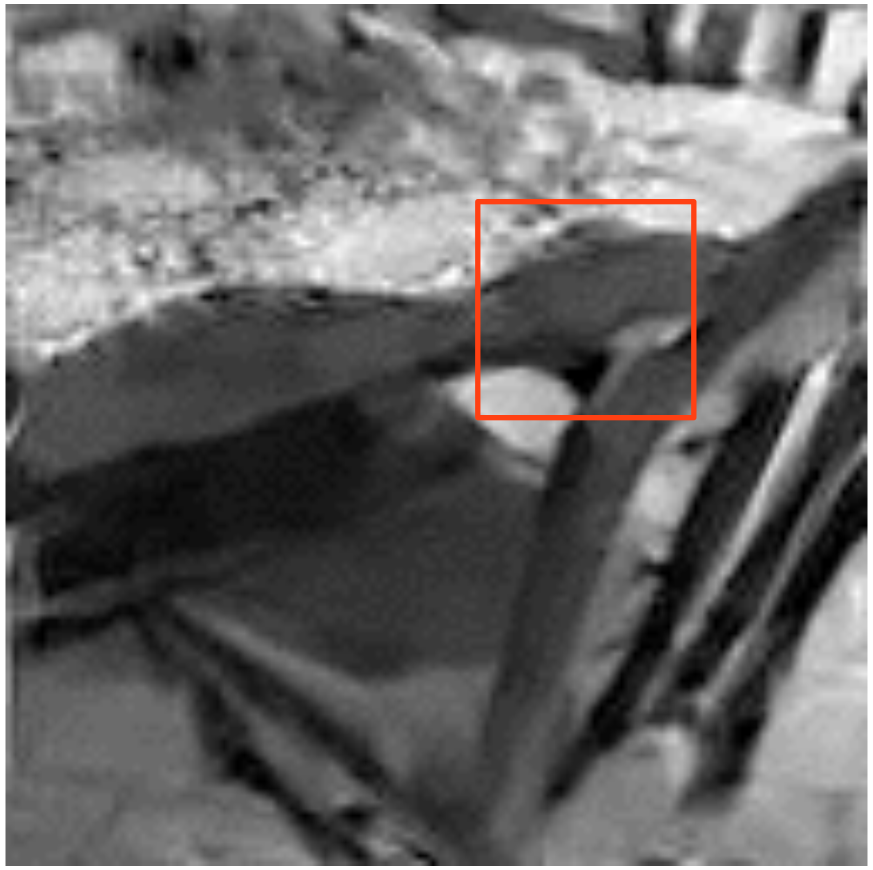
$2^J = 64$

$2^J = 128 = N$

Scattering Reconstructions

Original Images

Scat-2. Reconstr. $2^J = 32$



Learning Physics: N-Body Problem

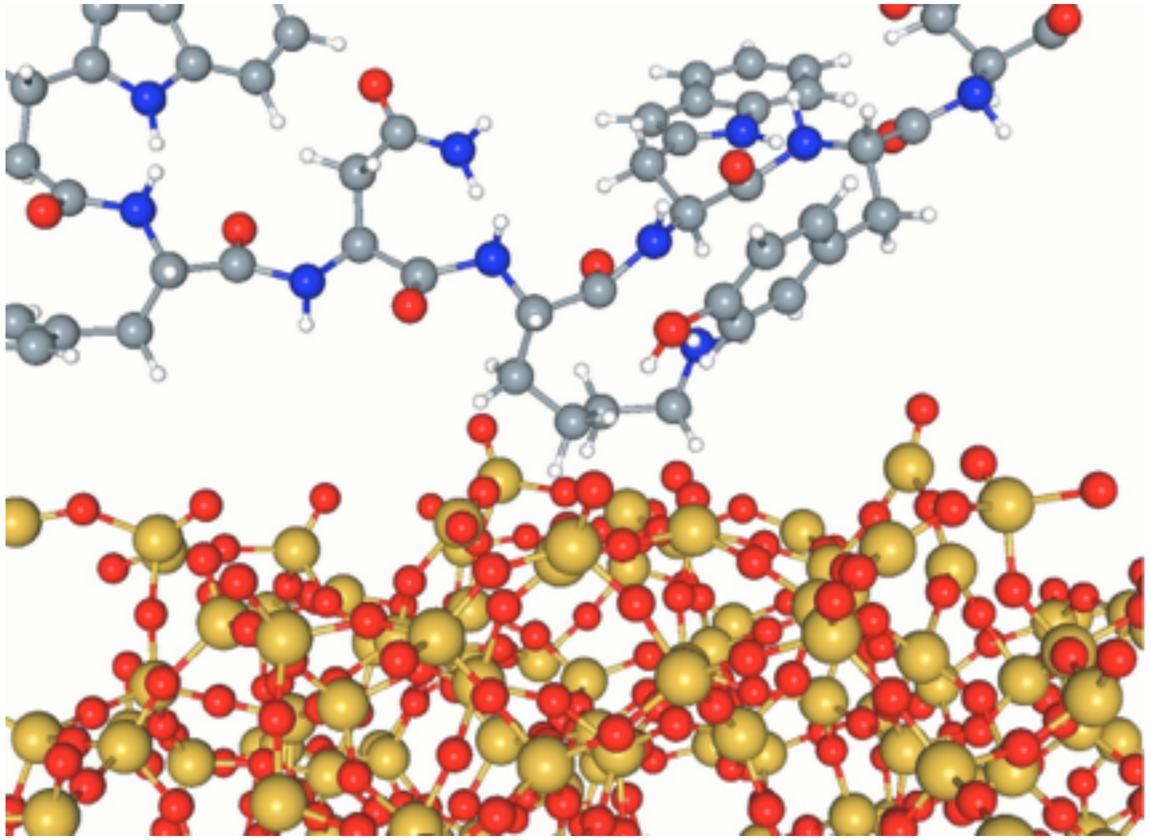
N. Poilvert
Matthew Hirn

- 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



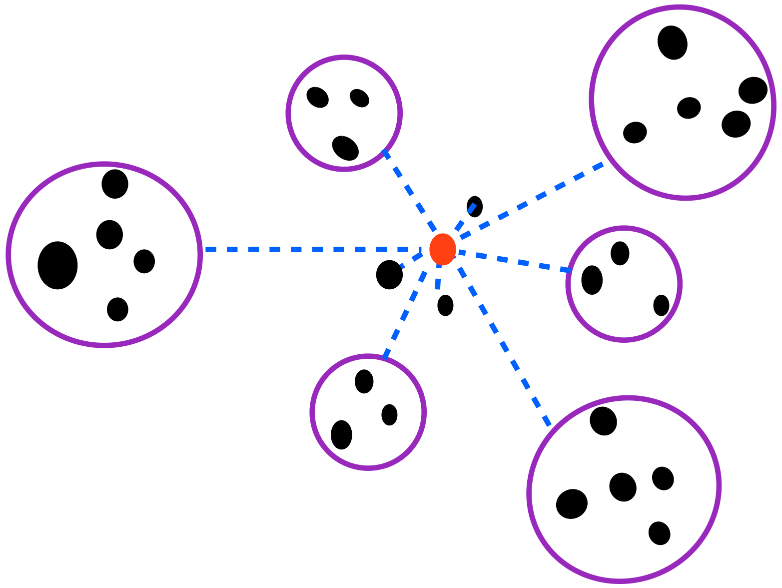
Learning Physics: N-Body Problem

N. Poilvert
Matthew Hirn

- Classic energy of d interacting bodies:

$$\text{If} \quad x(u) = \sum_{k=1}^{d} q_k \, \delta(u - p_k) \quad \text{then} \quad f(x) = \sum_{k=1}^{d} \sum_{k'=1}^{d} \frac{q_k \, q_{k'}}{|p_k - p_{k'}|^\beta}$$

Each particle interacts with $O(\log d)$ groups



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

$$f(x) = \sum_{m=0}^{M} \sum_{\lambda_1, \lambda_m} \alpha(\lambda_1, .., \lambda_m) \, S^2 x(\lambda_1, ..., \lambda_m)(1 + \epsilon)$$

N. Poilvert
Matthew Hirn

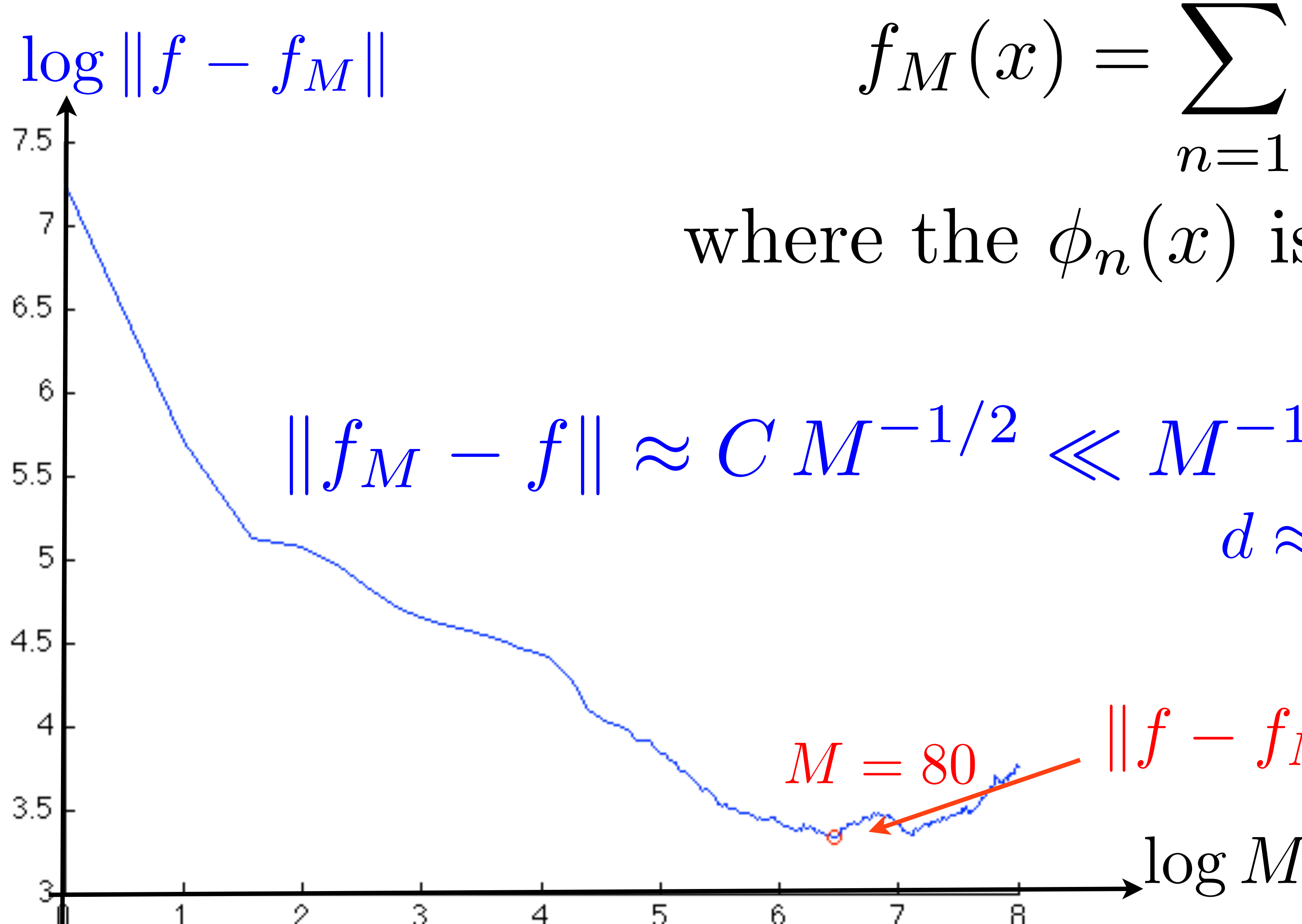
- Complex orbital interactions: no analytical energy $f(x)$.
  Invariant to translations, rotations, stable to deformations.
- Data basis $\{x_i \, , \, f(x_i)\}_i$ of 700 2D molecules (about 20 atoms)
- Best M-term scattering approximation f_M of f:

$$f_M(x) = \sum_{n=1}^{M} \alpha_n \, \phi_n(x)$$

where the $\phi_n(x)$ is a 1st or 2nd order term

$$\|f_M - f\| \approx C \, M^{-1/2} \ll M^{-1/d}$$

$d \approx 60$ degrees of freedom



Conclusion

- Do we need to learn deep net filters ?
- Can we analyse geometry in Euclidean spaces ?
- How much physics can we learn and why ?

Looking for Post-Doc!

www.di.ens.fr/data/scattering