Wavelet Translation Invariance

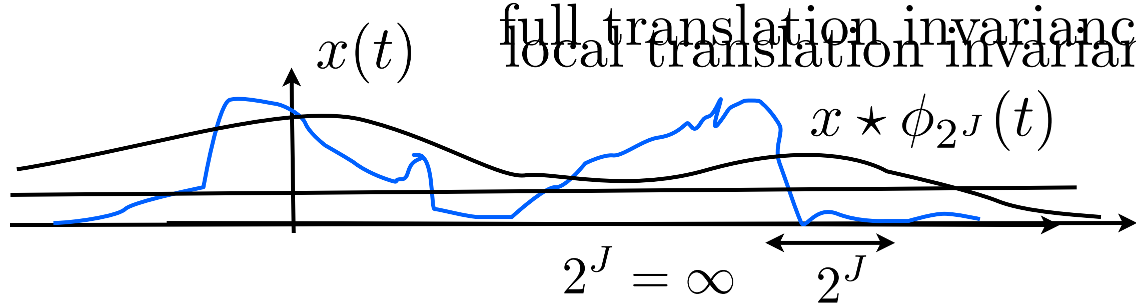
First wavelet transform

full translation invariance

local translation invariance

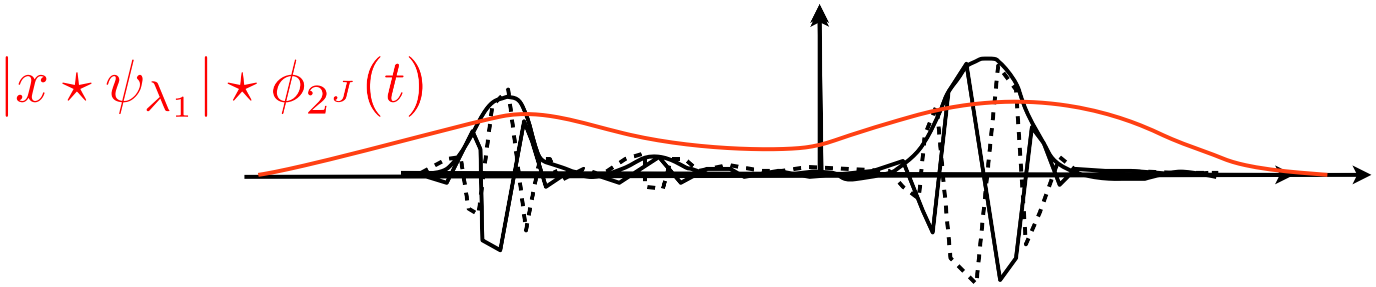
$$W_1 x = \begin{pmatrix} x \star \phi_{2^J} \\ x \star \psi_{\lambda_1} \end{pmatrix}_{\lambda_1}$$

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



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

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

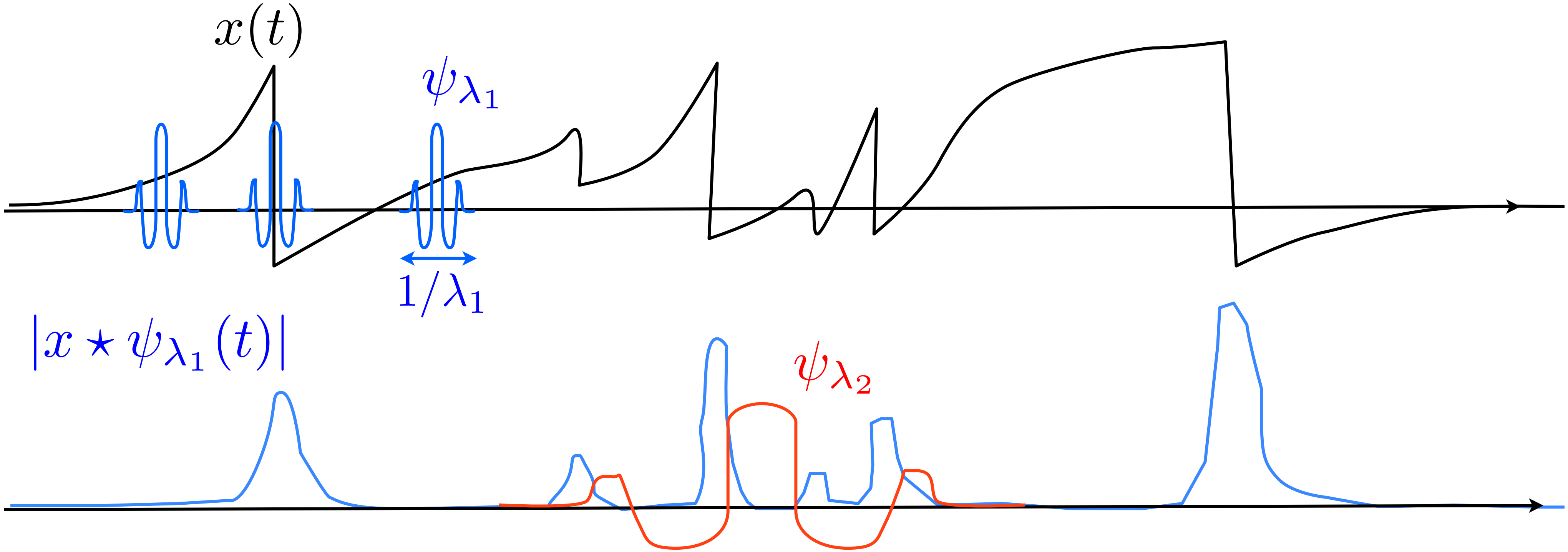


Second wavelet transform modulus

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

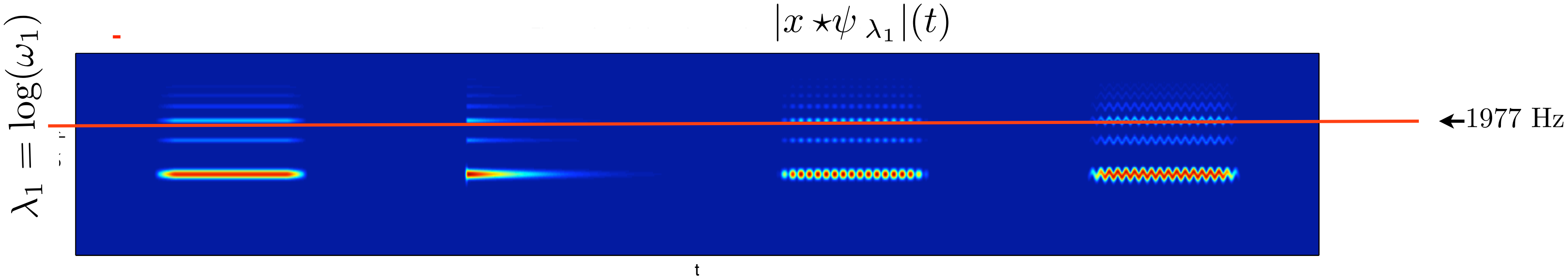
Singular Functions

$$|x \star \psi_{\lambda_1}(t)| = \left| \int x(u)\psi_{\lambda_1}(t-u)\,du \right|$$

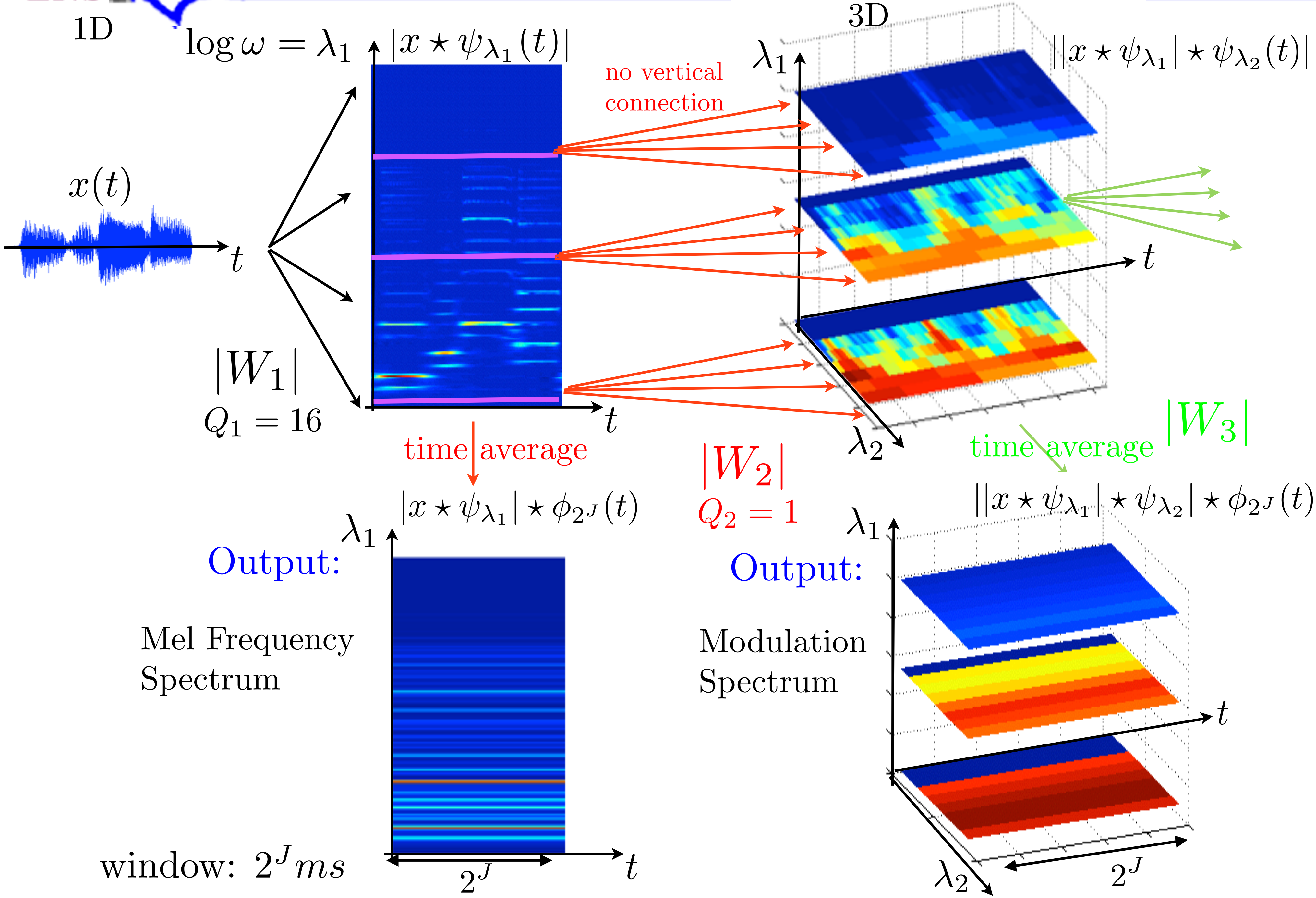


Amplitude Modulation

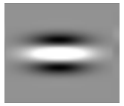
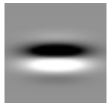
Harmonic sound: $x(t) = a(t)\, e \star h(t)$ with varying $a(t)$



ScatteringConvolution Network

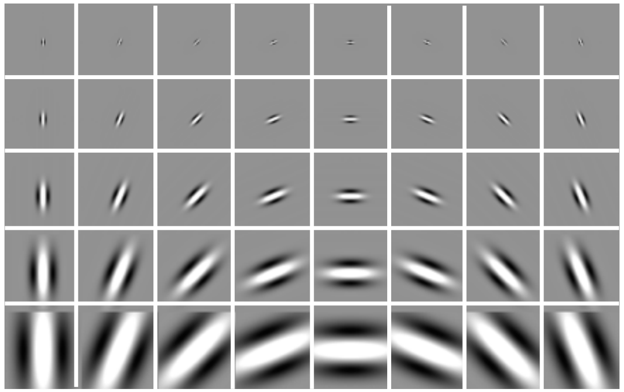


Scale separation with Wavelets

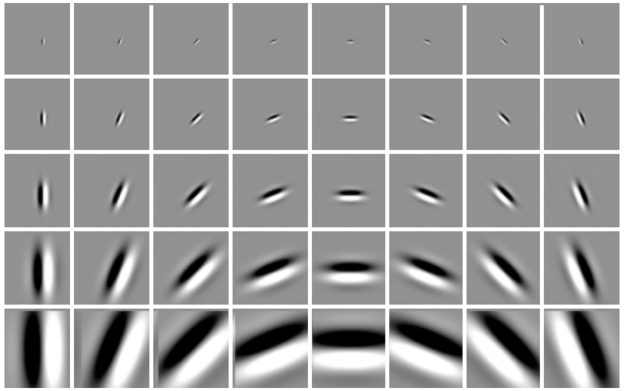
- Wavelet filter $\psi(u)$:  $+\, i$ 

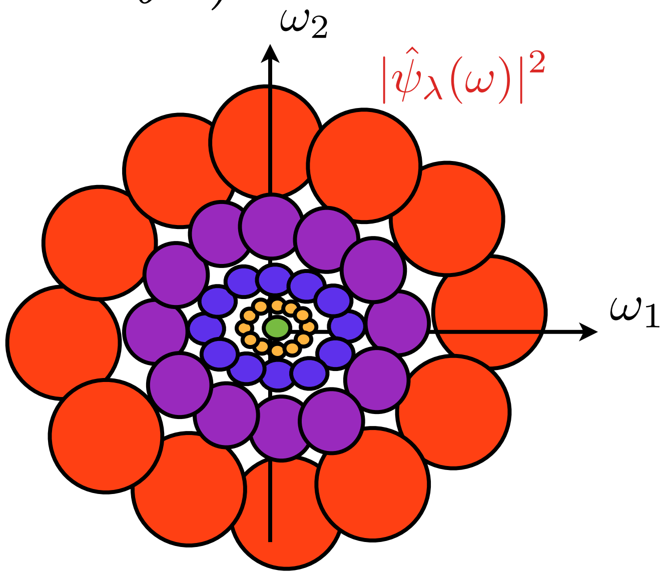
rotated and dilated: $\psi_{2^j,\theta}(u) = 2^{-j}\,\psi(2^{-j} r_\theta u)$

real parts



imaginary parts





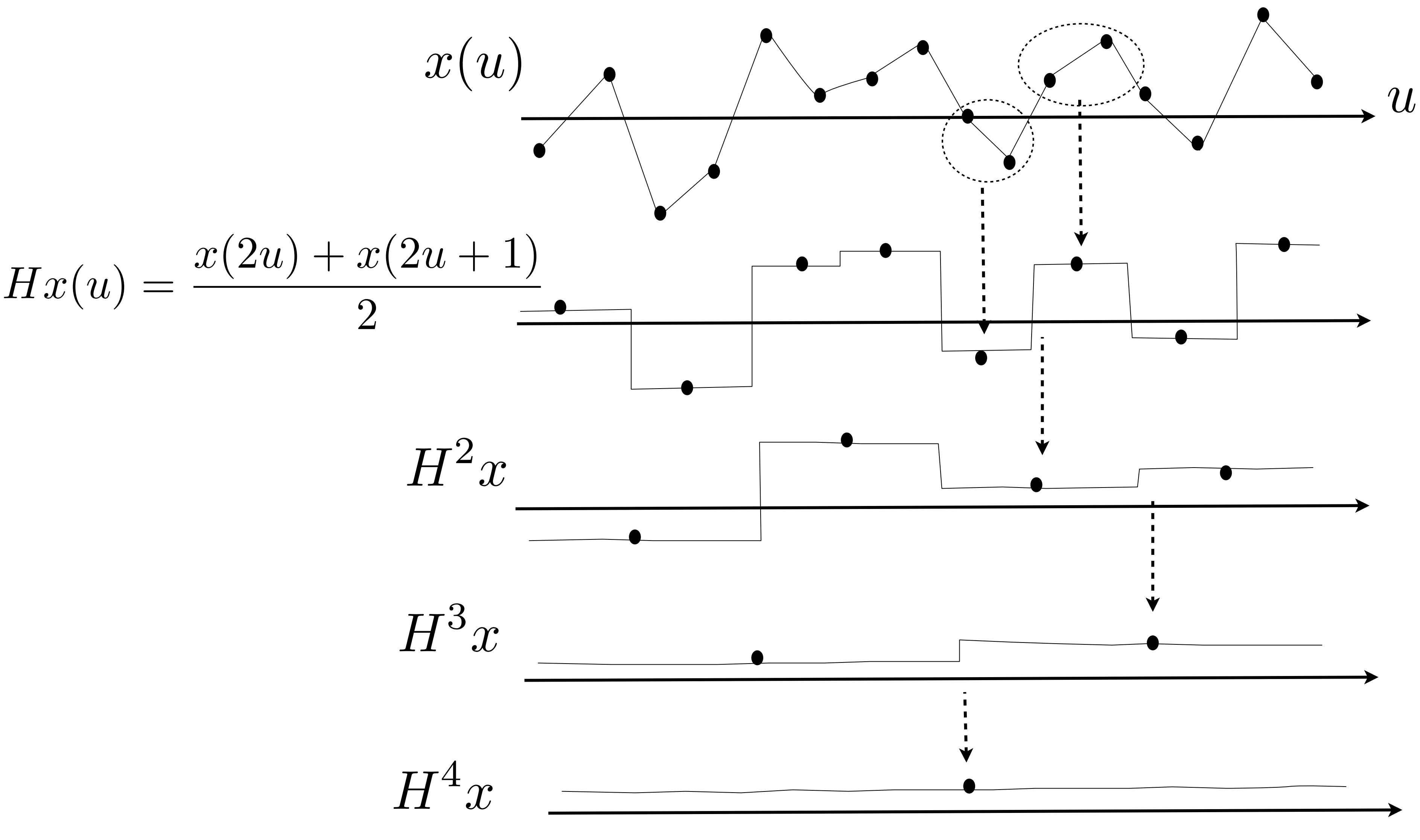
$$x \star \psi_{2^j,\theta}(u) = \int x(v)\, \psi_{2^j,\theta}(u-v)\, dv$$

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

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

Averaging Pyramid

- Multiscale averaging by cascade of pair averaging:



Haar Filtering

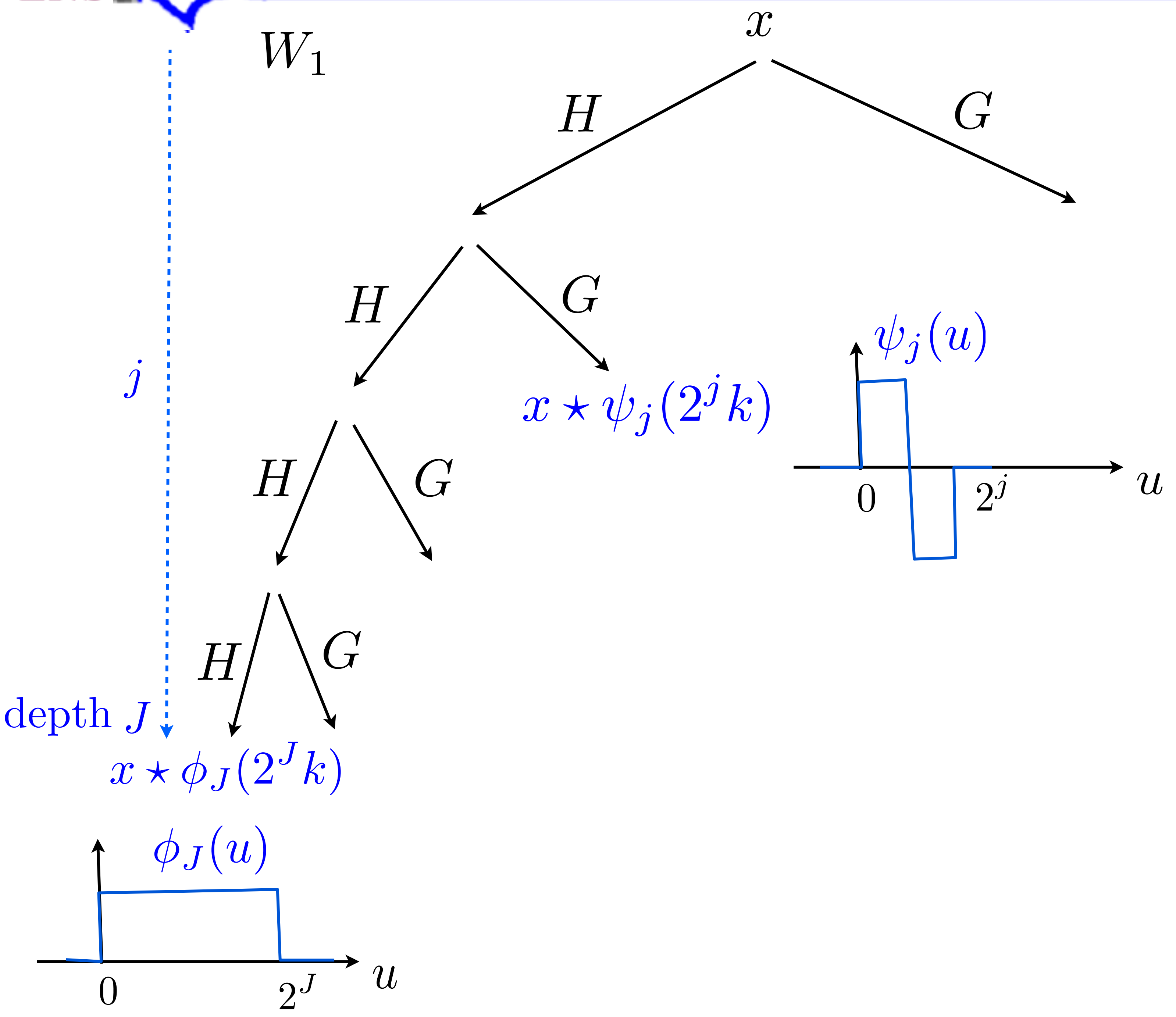
$$\{x(u)\}_{u \le d}$$

$$\left\{ \frac{x(2u) + x(2u+1)}{\sqrt{2}} \right\}_{u \le d/2} \xleftarrow{\;H\;} \{x(u)\}_{u \le d} \xrightarrow{\;G\;} \left\{ \frac{x(2u) - x(2u+1)}{\sqrt{2}} \right\}_{u \le d/2}$$

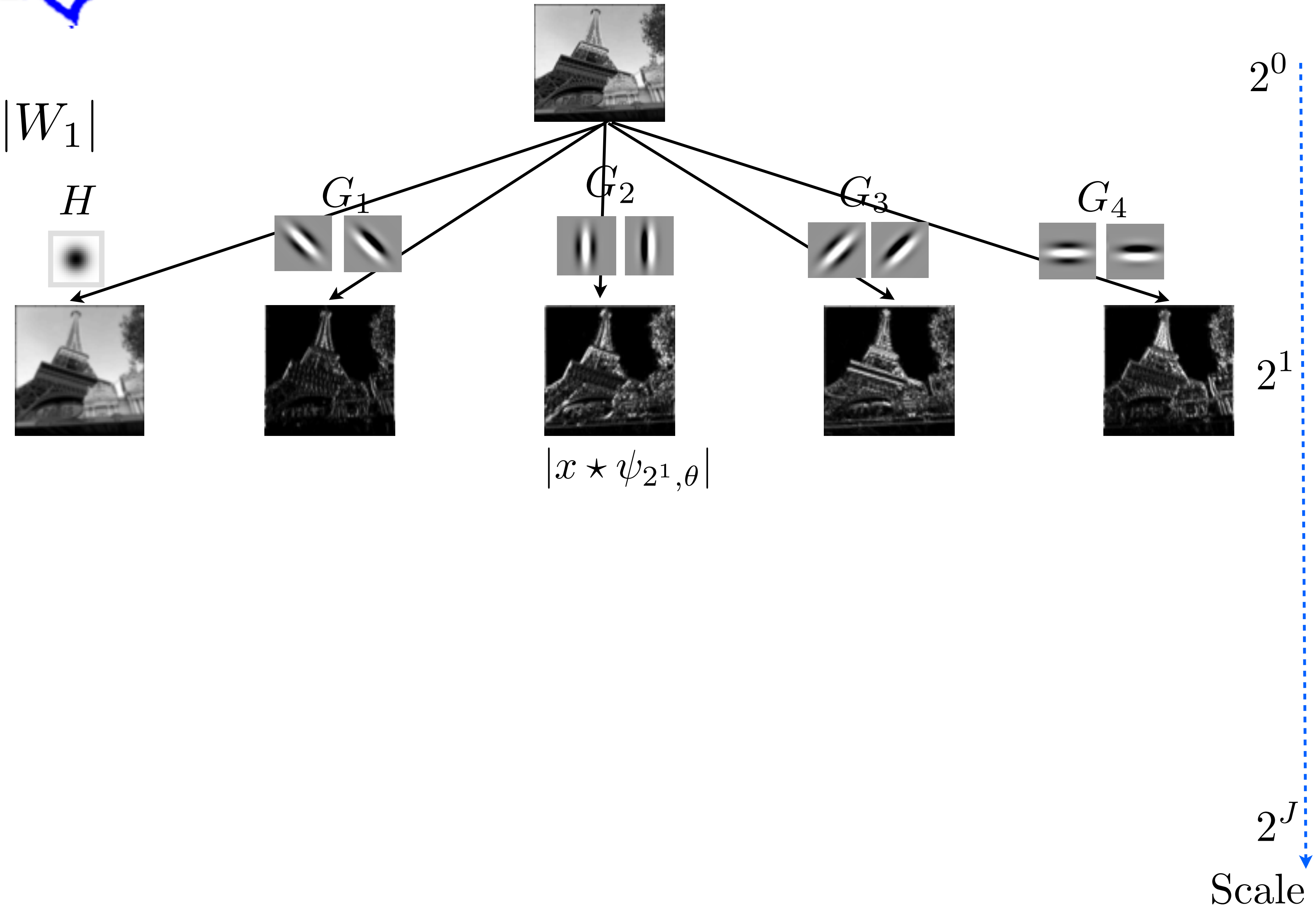
$$Hx(u) = x \star h(2u) \quad \text{and} \quad Gx(u) = x \star g(2u)$$

where h is a low frequency and g is a high frequency filter.

Haar Wavelet Transform



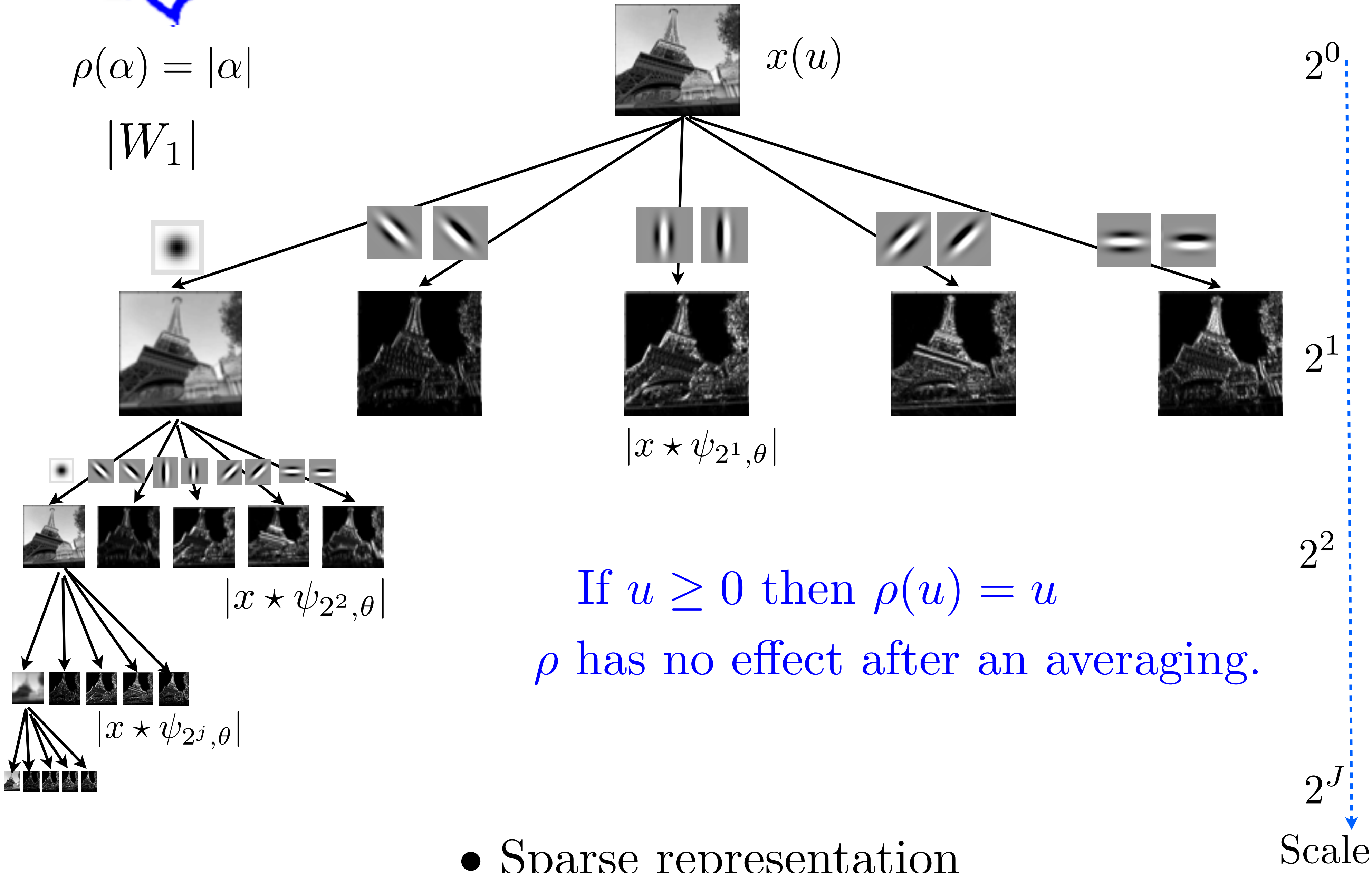
Fast Wavelet Filter Bank



Wavelet Filter Bank

$$\rho(\alpha) = |\alpha|$$

$$|W_1|$$

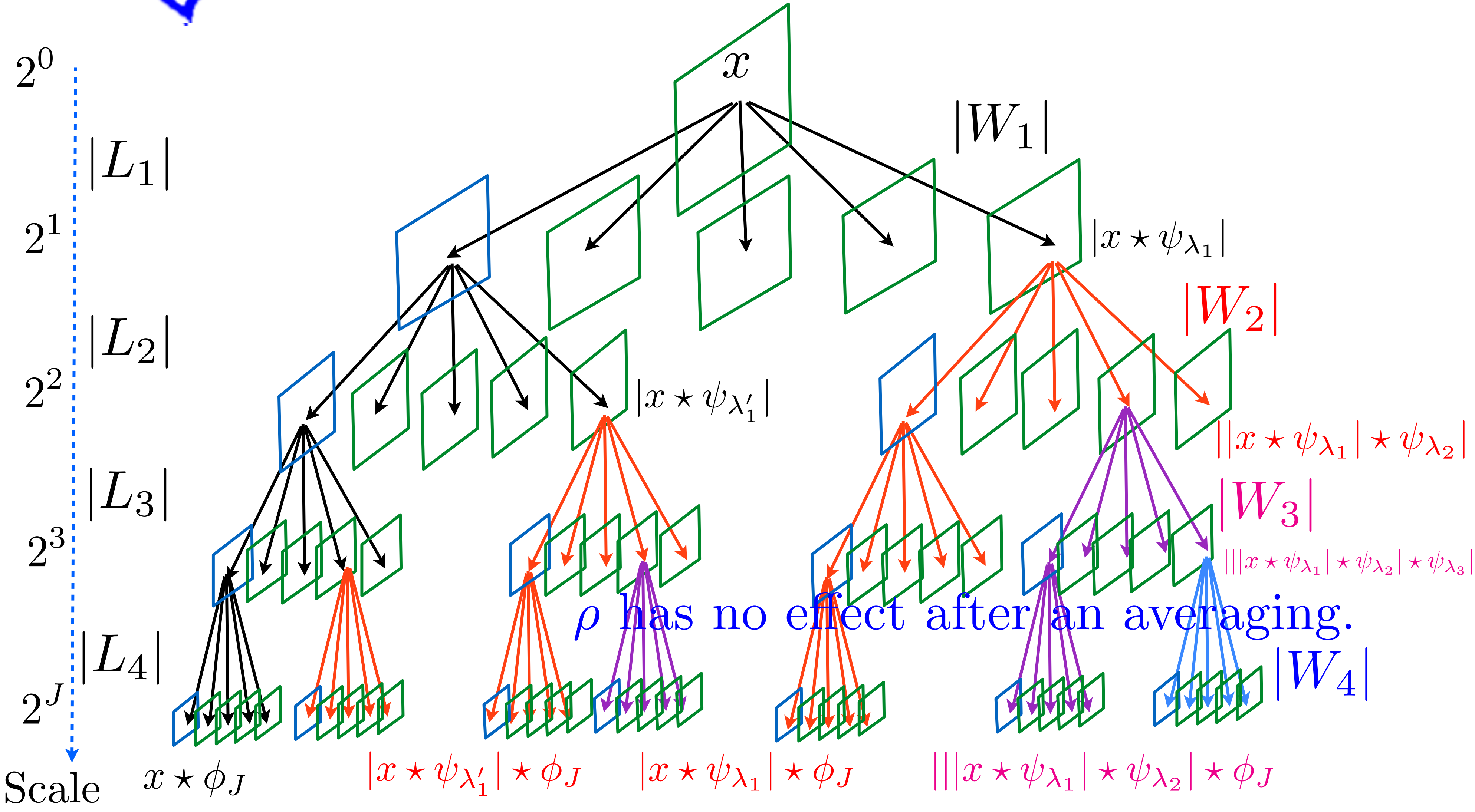


If $u \geq 0$ then $\rho(u) = u$

ρ has no effect after an averaging.

- Sparse representation

Wavelet Convolution Network Tree



$$S_4 x = |L_4|\,|L_3|\,|L_2|\,|L_1| x = |W_4|\,|W_3|\,|W_2|\,|W_1|\,x$$

Contraction

$$Wx = \begin{pmatrix} x \star \phi(t) \\ x \star \psi_\lambda(t) \end{pmatrix}_{t,\lambda} \quad \text{is linear and } \|Wx\| = \|x\|$$

$$\rho(u) = |u|$$

$$|W|x = \begin{pmatrix} x \star \phi(t) \\ |x \star \psi_\lambda(t)| \end{pmatrix}_{t,\lambda} \quad \text{is non-linear}$$

- it is contractive $\| |W|x - |W|y \| \leq \|x - y\|$

  because for $(a,b) \in \mathbb{C}^2$ $||a| - |b|| \leq |a - b|$

- it preserves the norm $\| |W|x \| = \|x\|$

Scattering Properties

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

$$\|W_k x\| = \|x\|$$

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

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 $\|S_J x\| = \|x\|$

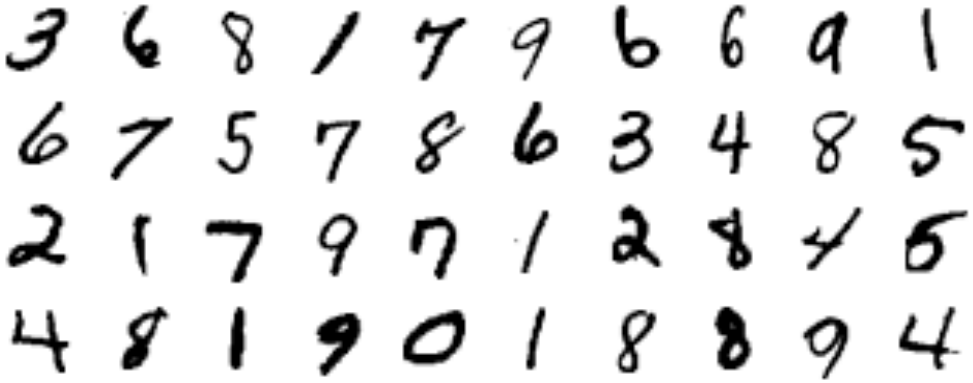
translations invariance and deformation stability:

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



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

Invariants to translations
Linearises small deformations
No learning

Invariants to specific deformations
Separates different patterns

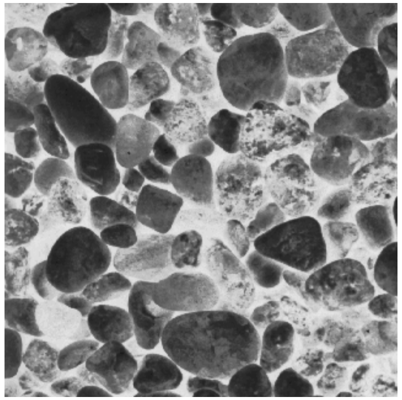
Classification Errors

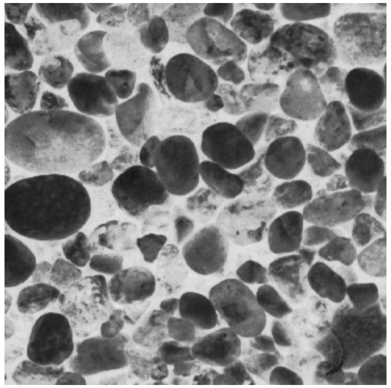
Training size	Conv. Net.	Scattering
50000	0.4%	0.4%

LeCun et. al.

Classification of Stationary Textures

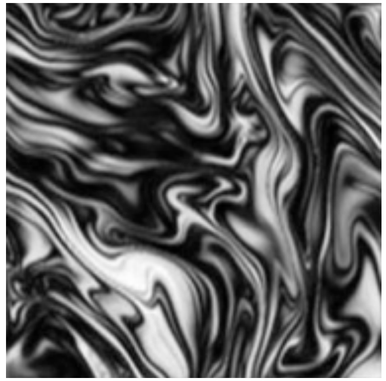
Ω_1





2D Turbulence

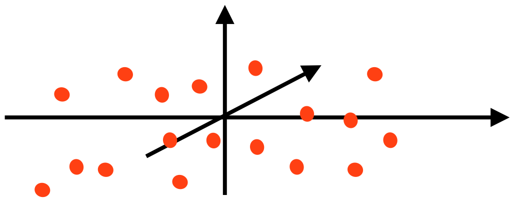
Ω_2



- What stochastic models ?

  Non Gaussian with long-range dependance.

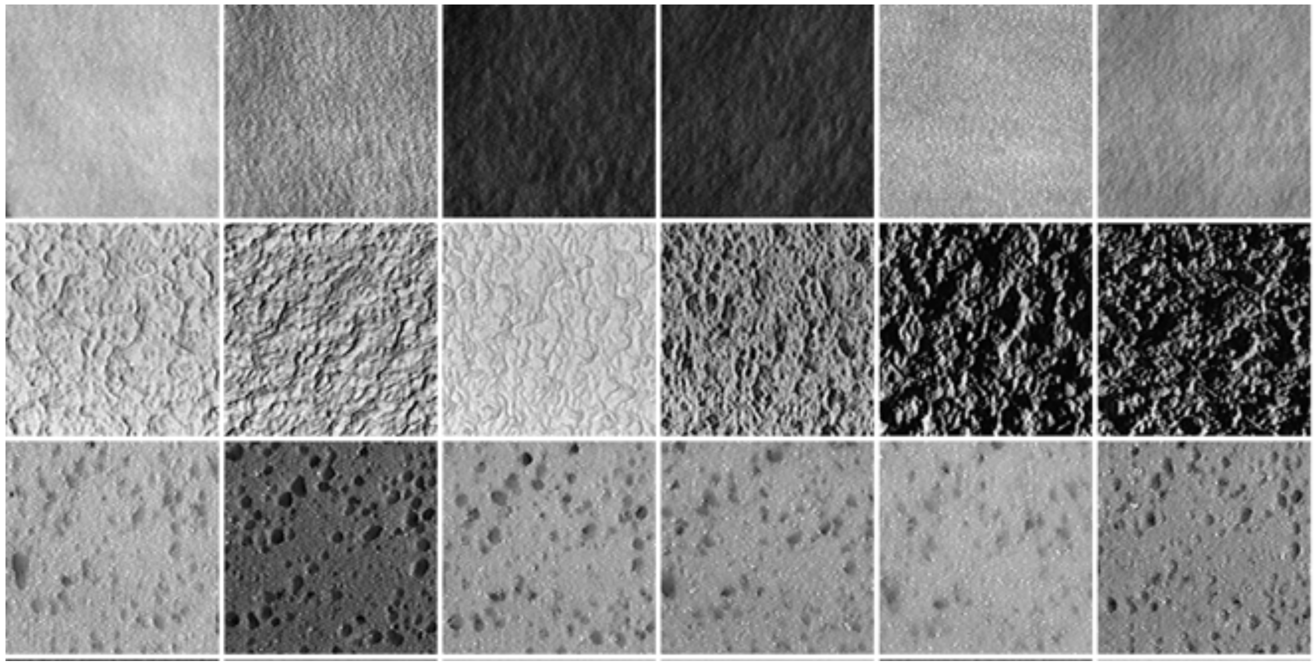
- Can we "Gaussianize" (linearize) such distributions in a reduced dimensional space ?



Classification of Textures

J. Bruna

CURET database



$$x \longrightarrow \boxed{S_J x} \longrightarrow \boxed{\begin{array}{c}\text{Supervised}\\ \text{Linear classifier}\end{array}} \longrightarrow y = f(x)$$

Classification Errors

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

Scattering Moments of Processes

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

$$S_J X = \begin{pmatrix} X \star \phi_{2^J}(t) \\ |X \star \psi_{\lambda_1}| \star \phi_{2^J}(t) \\ ||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_{2^J}(t) \\ |||X \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \star \phi_{2^J}(t) \\ \dots \end{pmatrix}_{\lambda_1,\lambda_2,\lambda_3,\dots} \quad : \text{stationary vector}$$

$J \to \infty$

Central limit theorem
with "weak" ergodicity conditions

J. Bruna

Gaussian distribution: $\mathcal{N}\Big(\mathbb{E}(SX), \Sigma_J \to 0\Big)$

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

Scattering Moments of Processes

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

$$S_J X = \begin{pmatrix} X \star \phi_{2^J}(t) \\ |X \star \psi_{\lambda_1}| \star \phi_{2^J}(t) \\ ||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_{2^J}(t) \\ |||X \star \psi_{\lambda_2}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \star \phi_{2^J}(t) \\ \cdots \end{pmatrix}_{\lambda_1, \lambda_2, \lambda_3, \ldots} \quad : \text{ stationary vector}$$

$J \to \infty$

Central limit theorem
with "weak" ergodicity conditions

Gaussian distribution: $\mathcal{N}\Big(\mathbb{E}(SX), \Sigma_J \to 0\Big)$

- Reconstruction: compute $\tilde{X}$ which minimises

$$\|S_J \tilde{X} - S_J X\|^2$$

- Gradient descent

Representation of Audio Textures

Joan Bruna

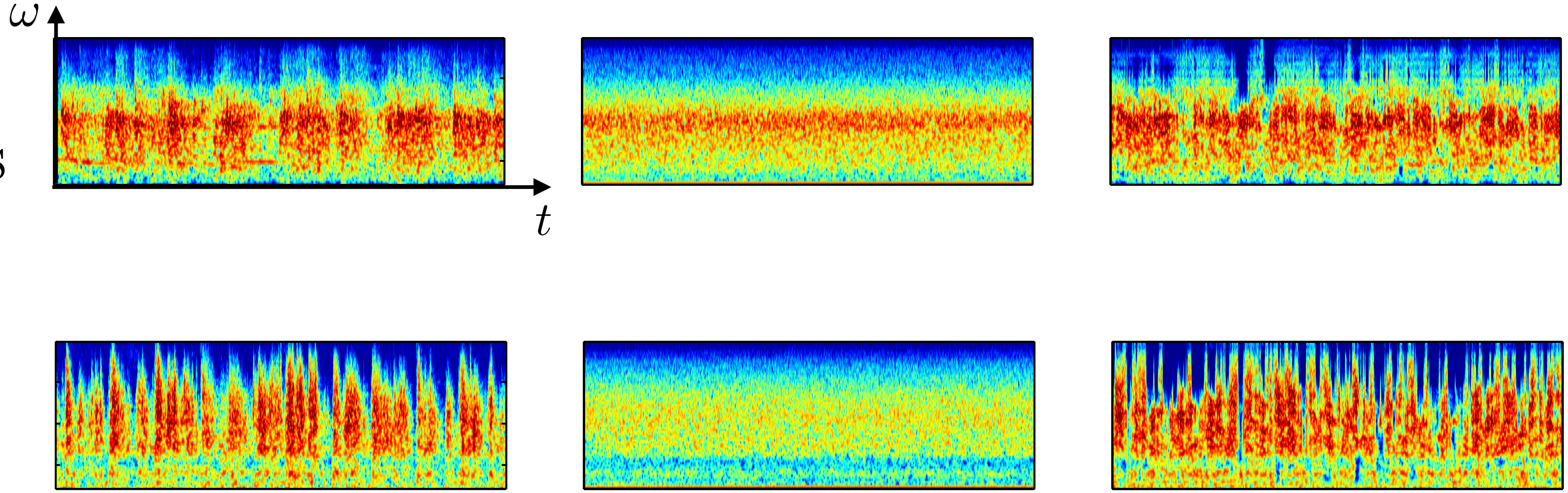
Original

Gaussian in time

Gaussian in scattering

Applauds

Paper



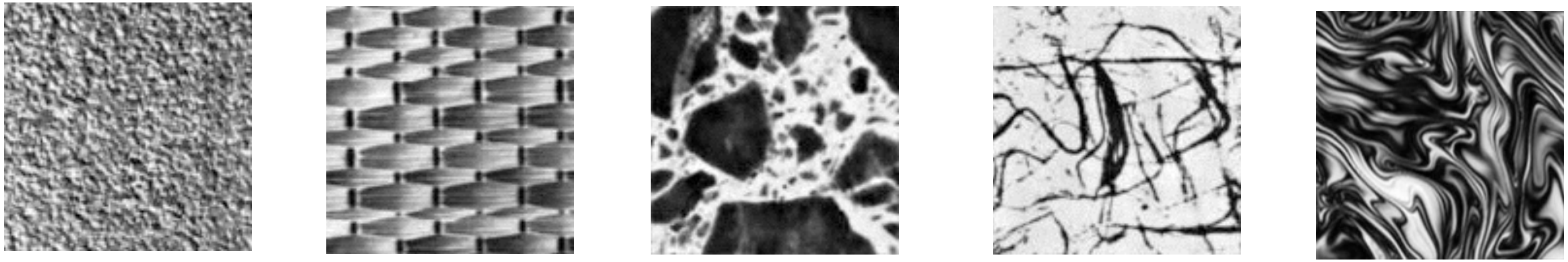
Cocktail Party

Ergodic Texture Reconstructions

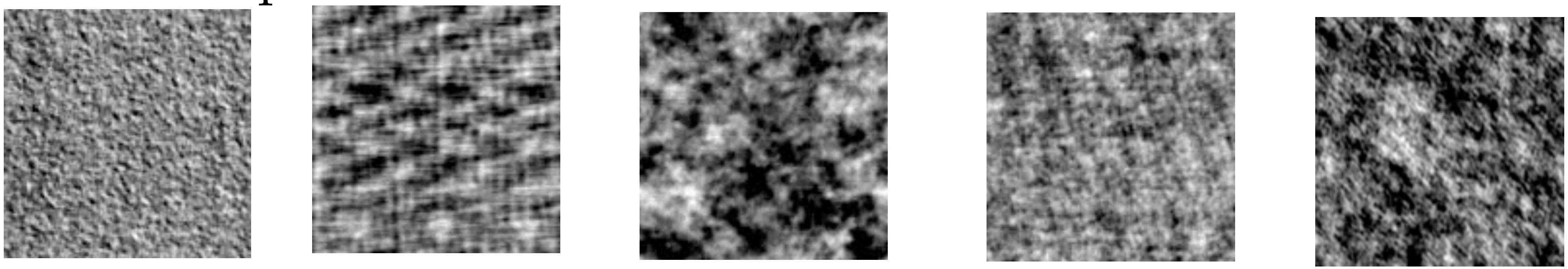
Joan Bruna

Textures of N pixels

2D Turbulence

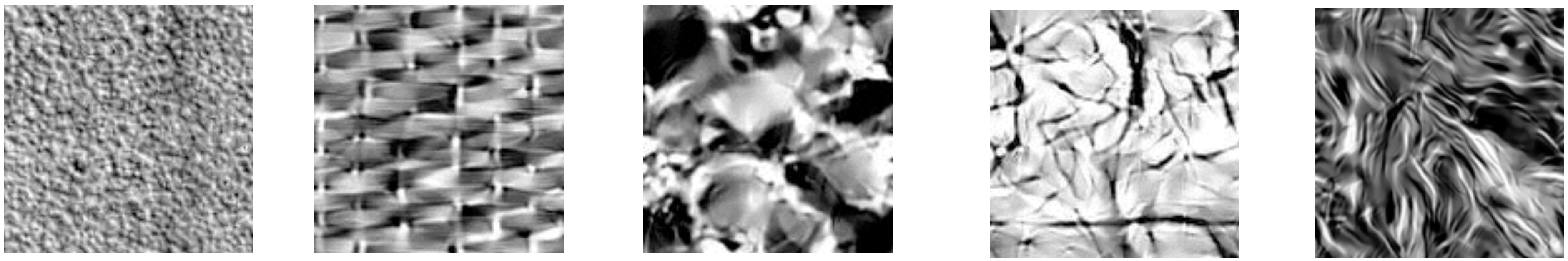


Gaussian process model with N second order moments



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

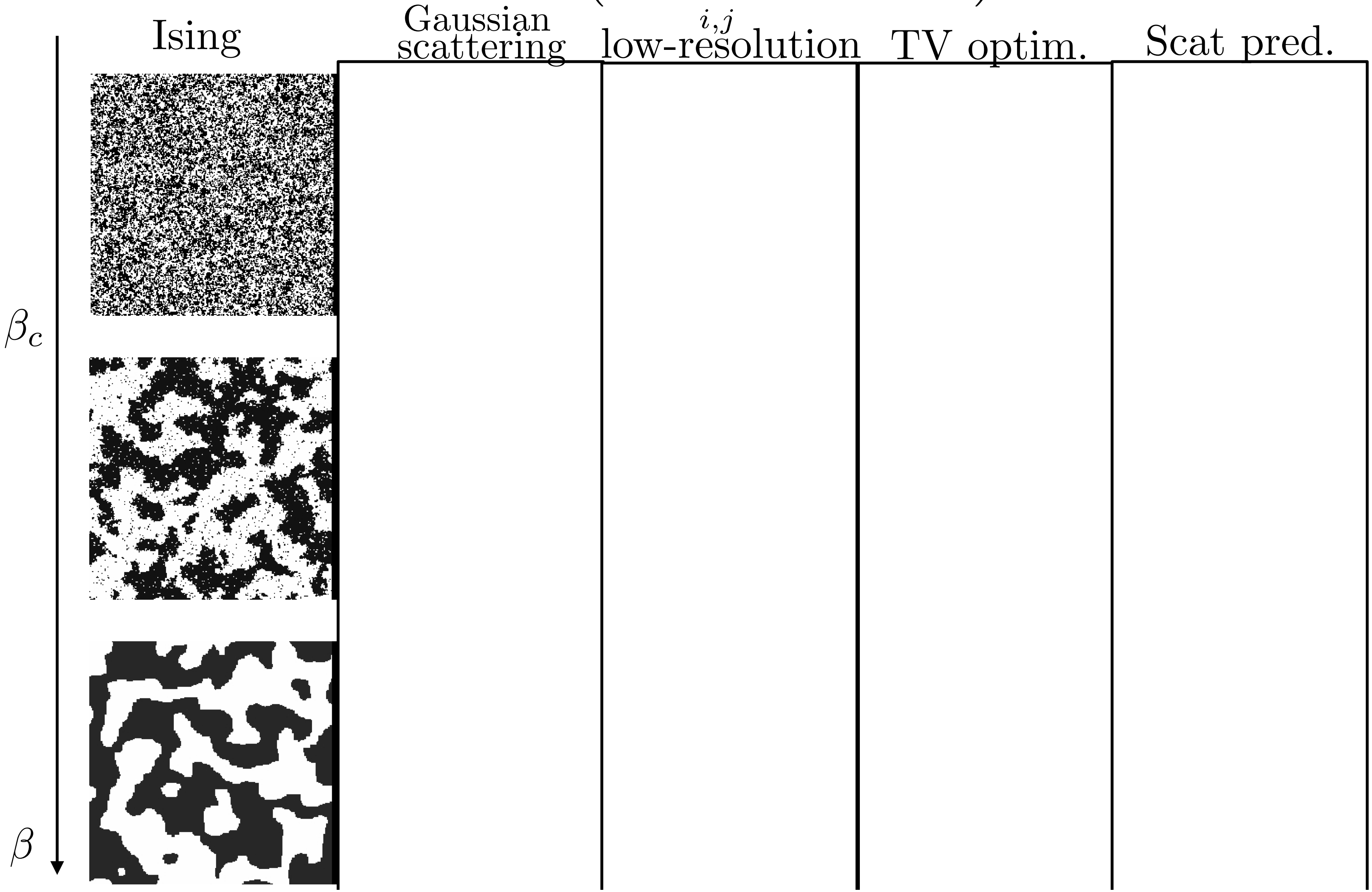
$\mathbb{E}(|x \star \psi_{\lambda_1}|)$, $\mathbb{E}(||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|)$



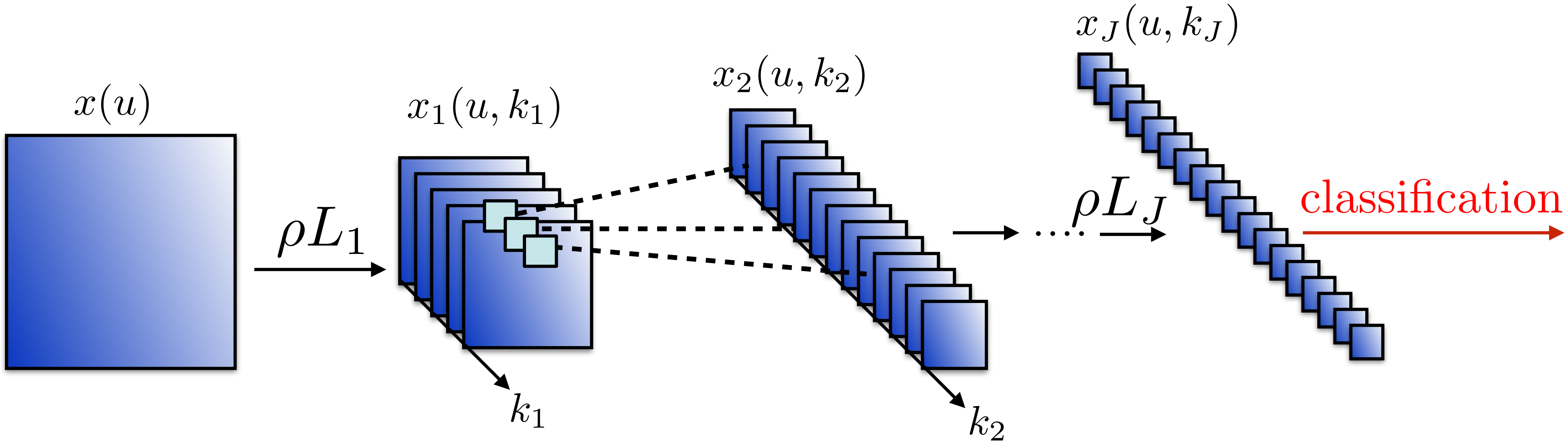
Ising Model and Inverse Problem

Bruna, Dokmanic, Maarten de Hoop

$$p(x) = Z_\beta^{-1} \exp\Big(-\beta \sum_{i,j} J_{i,j}\, x(i)\, x(j) \Big) \quad \text{with} \quad x(i) = \pm 1$$



Deep Convolutional Trees



$$x_j = \rho\, L_j\, x_{j-1}$$

L_j is composed of convolutions and subs samplings:

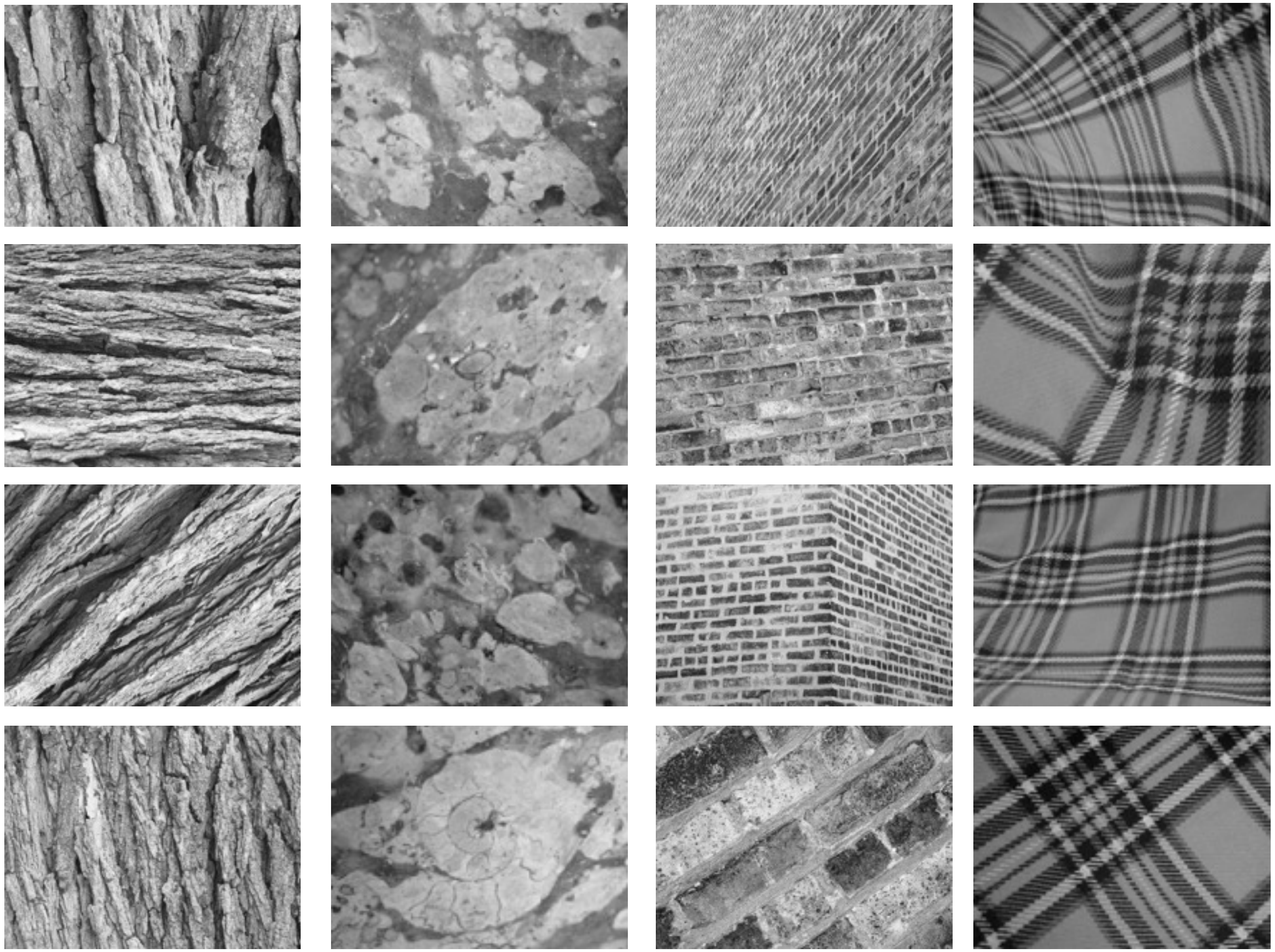
$$x_j(u, k_j) = \rho\Big(x_{j-1}(\cdot, k) \star h_{k_j,k}(u)\Big)$$

No channel communication: what limitations ?

Rotation and Scaling Invariance

Laurent Sifre

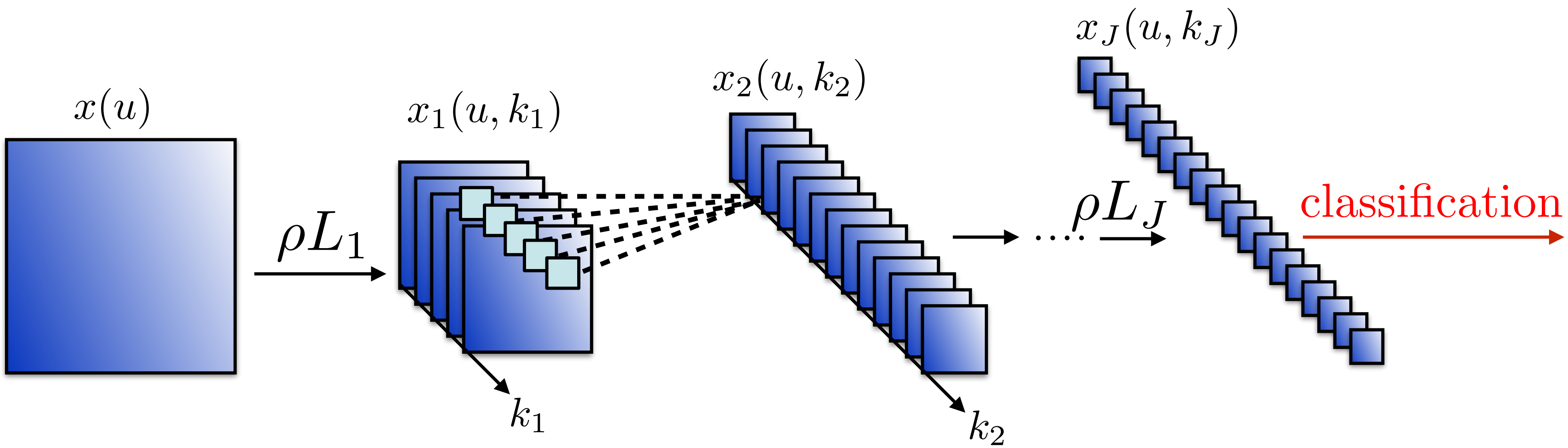
UIUC database:
25 classes



Scattering classification errors

Training	Scat. Translation
20	20 %

Deep Convolutional Networks



$$x_j = \rho \, L_j \, x_{j-1}$$

- L_j is a linear combination of convolutions and subsampling:

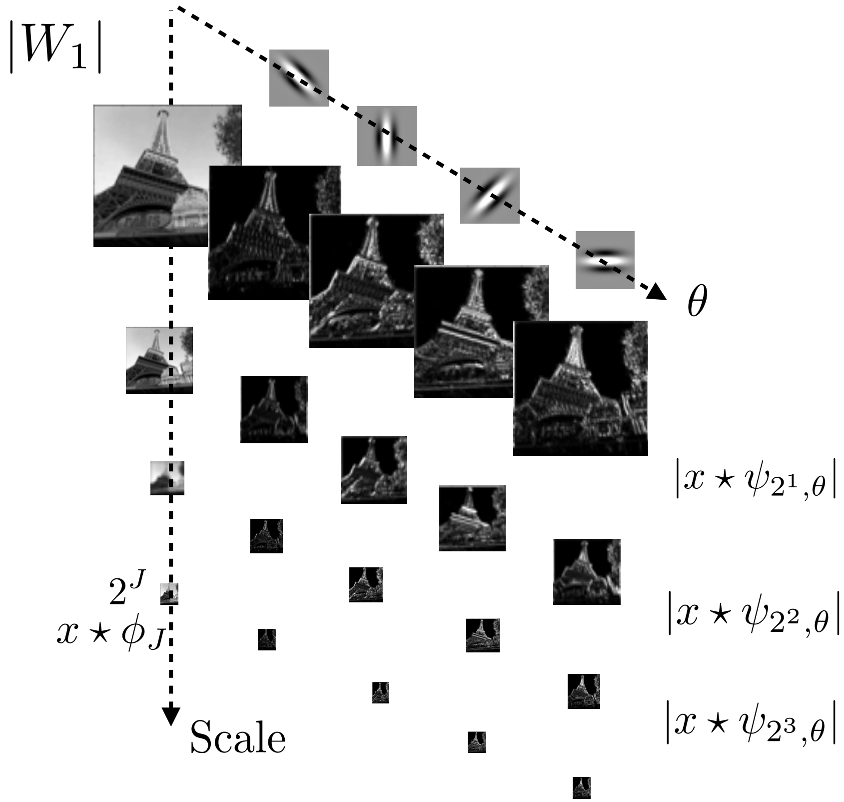
$$x_j(u, k_j) = \rho \Big(\underbrace{\sum_k}_{\text{sum across channels}} x_{j-1}(\cdot, k) \star h_{k_j,k}(u) \Big)$$

What is the role of channel connections ?

Linearize other symmetries beyond translations.

Rotation Invariance

- Channel connections linearize other symmetries.



- Invariance to rotations are computed by convolutions along the rotation variable θ with wavelet filters.

  $\Rightarrow$ invariance to rigid mouvements.

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

$$x(r_\alpha(u)) \quad x(u)$$

rotation & translation , angle translation

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

$$\downarrow$$

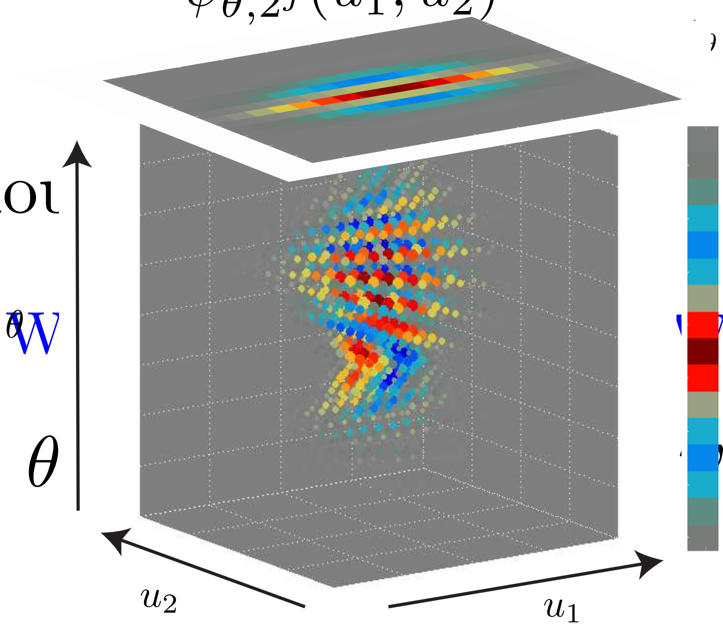
$$\int x(u)du$$

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 $\psi_{\lambda_2}(u,\theta)$

$$x \circledast \psi_\lambda(u,\theta) = \int_0^{2\pi} \left(\int_{\mathbb{R}^2} x(u',\theta')\, \psi_{\theta,2^j}(r_{-\theta'}(u-u')) \right) \psi_{2^k}(\theta-\theta') d\theta' dt'$$



- Scattering on rigid mou

Wavelets on Translations W ... avelets on Rigid Mvt.

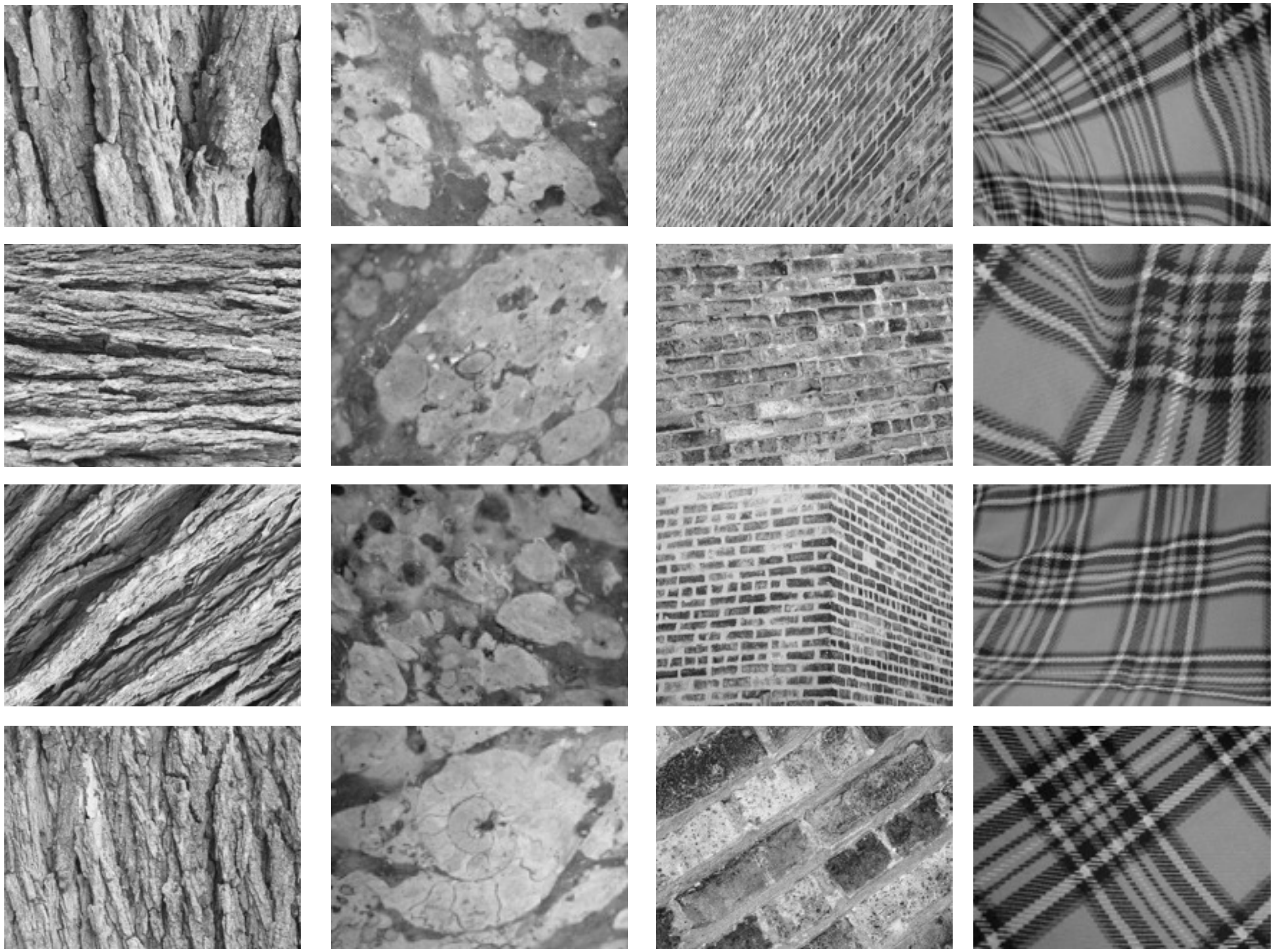
$x(u) \longrightarrow \boxed{|W_1|} \longrightarrow x_j(u,\theta$ ... $\lambda_2(v,\theta)| \longrightarrow \boxed{|W_3|}$

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

Rotation and Scaling Invariance

Laurent Sifre

UIUC database:
25 classes



Scattering classification errors

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

Learning Physics: N-Body Problem

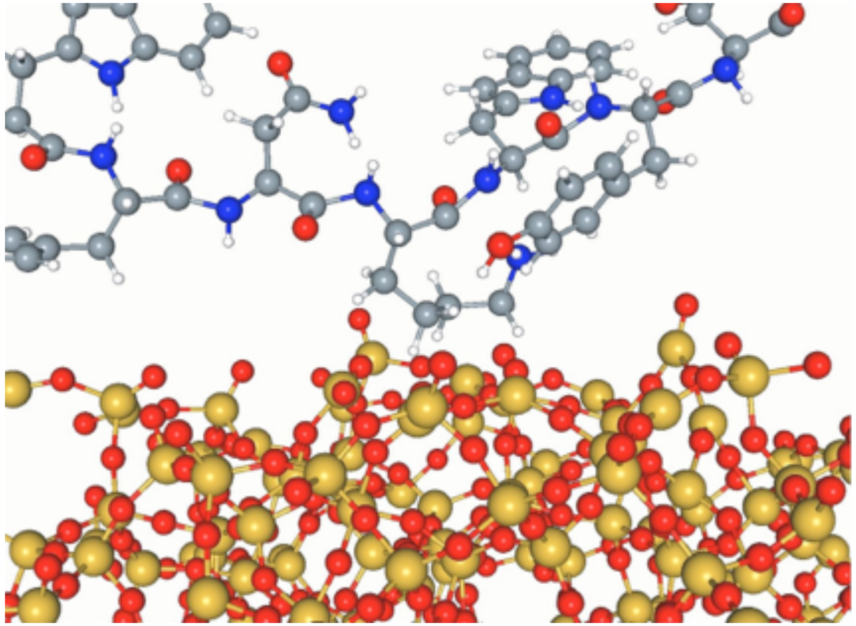
N. Poilvert
Matthew Hirn

- Energy of d interacting bodies:

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

Astronomy

Quantum Chemistry



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:

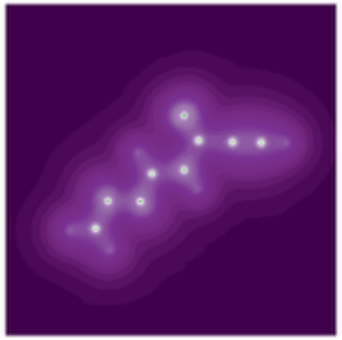
$$f(x) = E(\rho_x) = \min_{\rho} E(\rho)$$

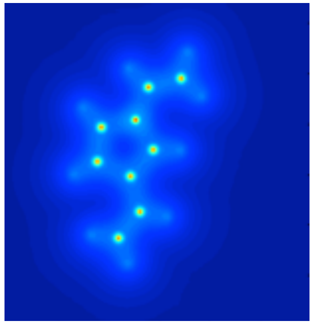
Quantum Chemistry Invariants

Quantum chemistry: $f(x)$ is invariant to rigid mouvements, stable to deformations.

Depends on the true electronic density (Kohn-Sham)

Ground state electronic density computed with Schroedinger





- Can we estimate $f(x)$ from a naive electronic density ?

Density $\tilde{\rho}_x$ computed as a sum of blobs

Quantum Regression

N. Poilvert
Matthew Hirn

- Linear regressions computed with invariant change of variables:

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

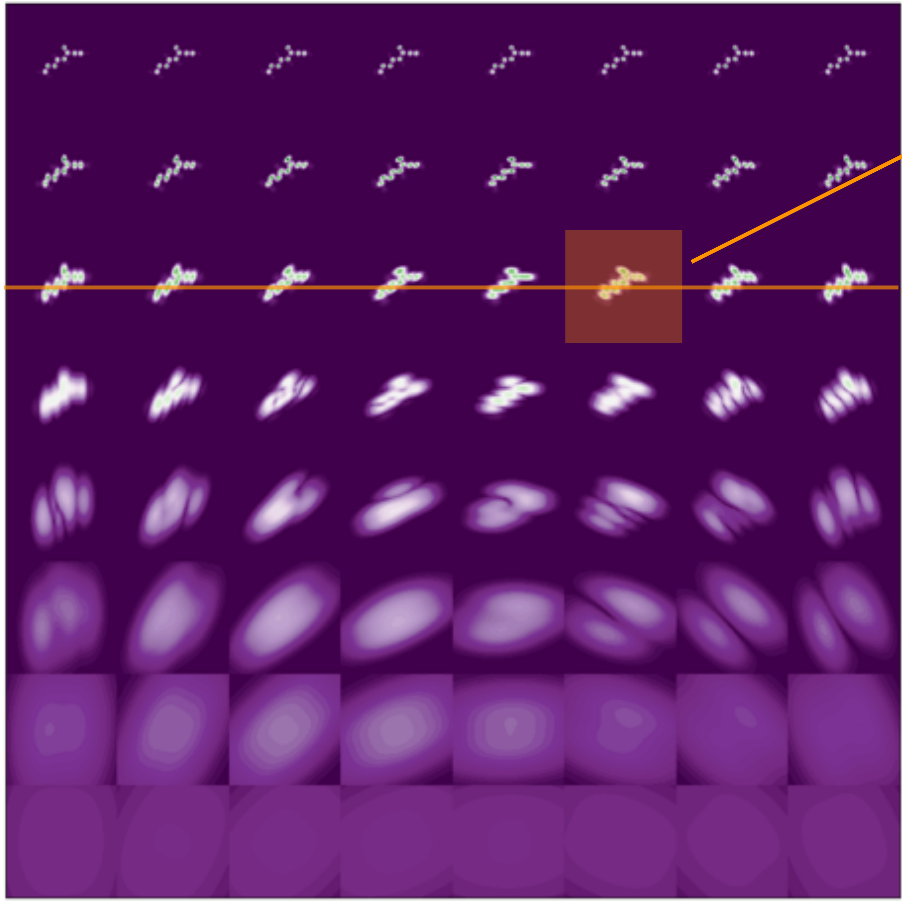
$$f_M(x) = \sum_{k=1}^{M} w_k \, \phi_{n_k}(\tilde{\rho}_x)$$

Regression coefficients w_k: equivalent potential.

Scattering Dictionary

$\rho(u$

$|\rho * \psi_{j_1,\theta_1}(u)|$



Scales j_1

Rotations θ_1

2nd Order Interferences

Recover translation variability:

$$|\rho * \psi_{j_1,\theta_1}| * \psi_{j_2,\theta_2}(u)$$

Recover rotation variability:

$$|\rho * \psi_{j_1,\cdot}(u)| \circledast \overline{\psi}_{l_2}(\theta_1)$$

Combine to recover
roto-translation variabiltiy:

$$||\rho * \psi_{j_1,\cdot}| * \psi_{j_2,\theta_2}(u) \circledast \overline{\psi}_{l_2}(\theta_1)|$$

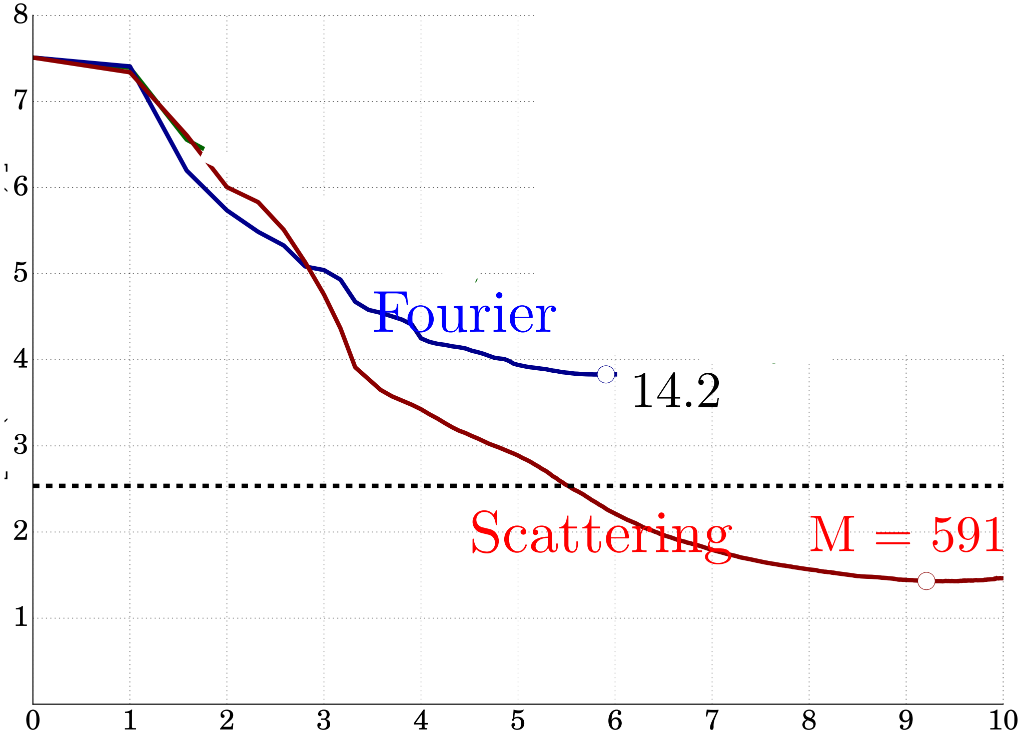
Scattering Regression

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

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

Interaction terms
across scales

Testing error
$2^{-1} \log_2 \mathbb{E}|f_M(x) - y(x)|^2$



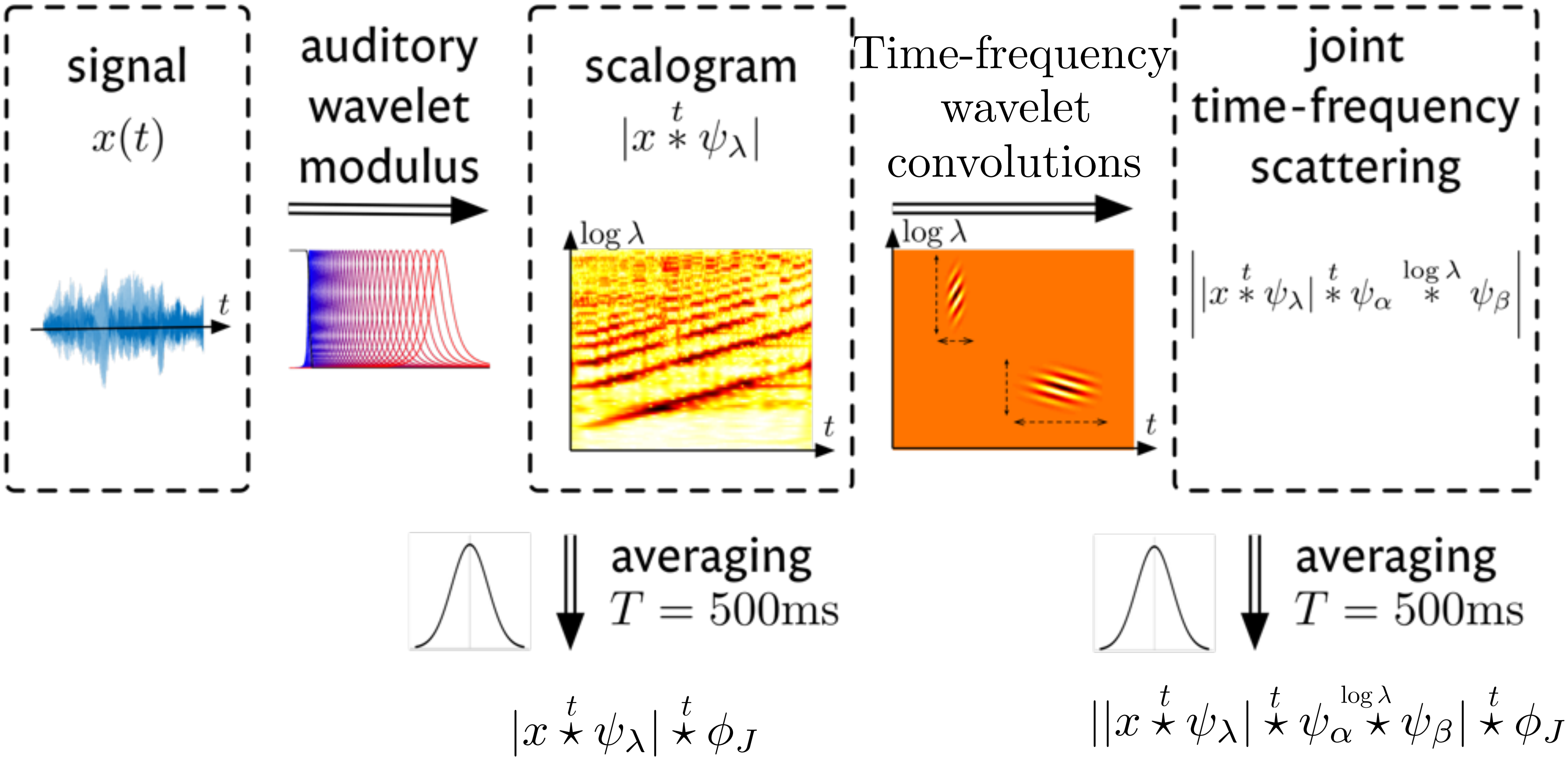
5.8 : State of the art

1.8 kcal/mol

$\log_2 M$

Time-Frequency Translation Group

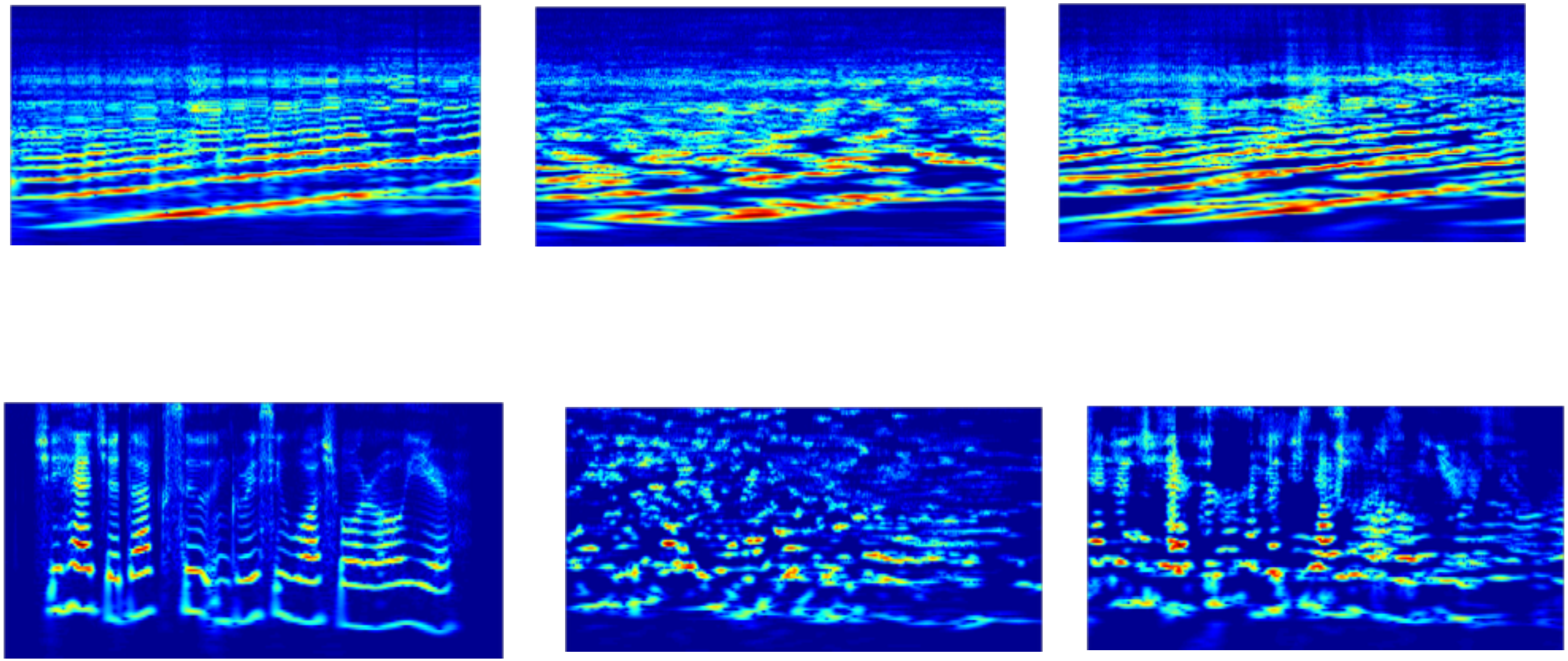
J. Anden and V. Lostanlen



Joint Time-Frequency Scattering

J. Anden and V. Lostanl

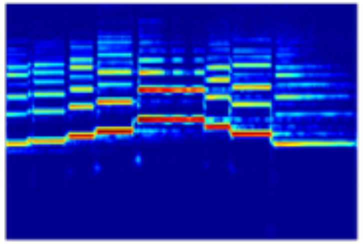
Original | Time Scattering | Time/Freq Scattering



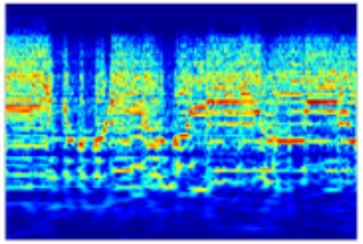
Musical Instrument Classificaiton

J. Anden and V. Lostanlen

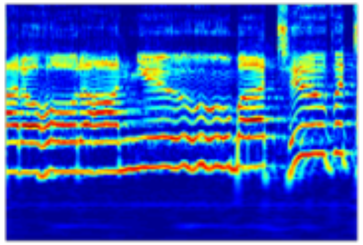
clarinet



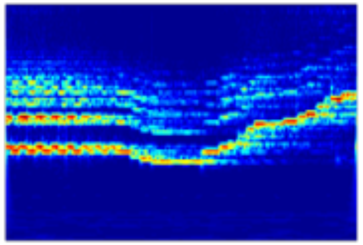
electric guitar



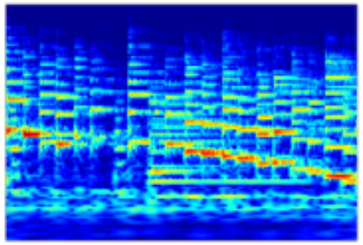
female singer



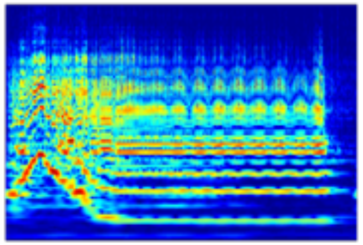
flute



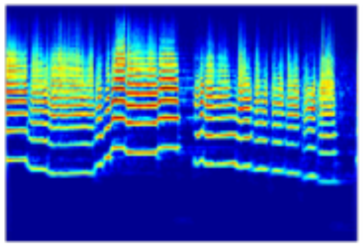
piano



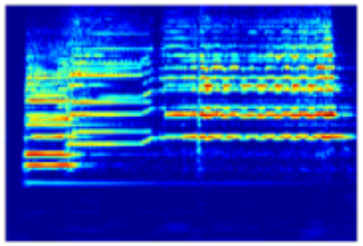
tenor saxophone



trumpet



violin

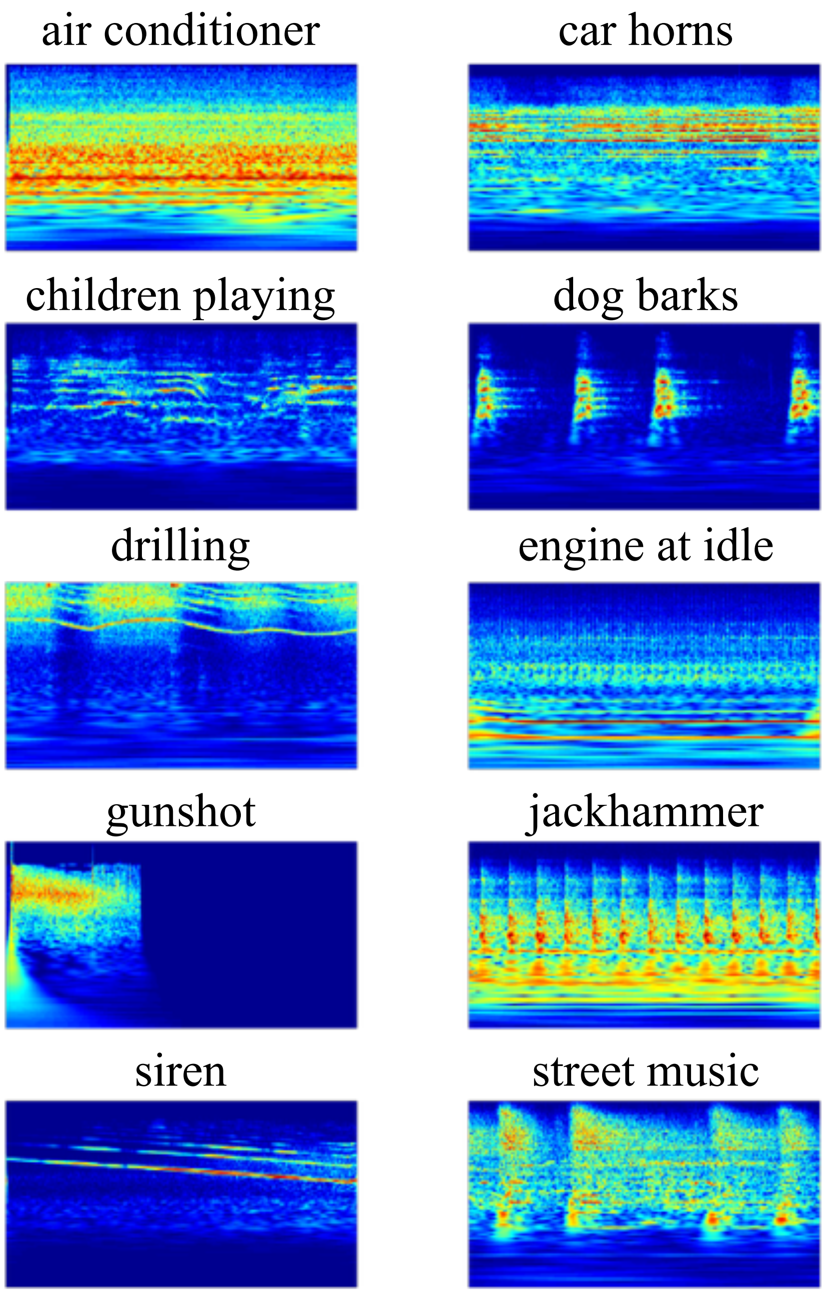


MedleyDB: 8 classes
10k training examples

class-wise average error

MFCC audio descriptors	0,39
time scattering	0,31
ConvNet	0,31
time-frequency scattering	0,18

Environmental Sound Classification



J. Anden and V. Lostanlen

UrbanSound8k: 10 classes
8k training examples

class-wise average error

MFCC audio descriptors	0,39
time scattering	0,27
ConvNet (Piczak, MLSP 2015)	0,26
time-frequency scattering	0,2

Complex Image Classification

Edouard Oyallon

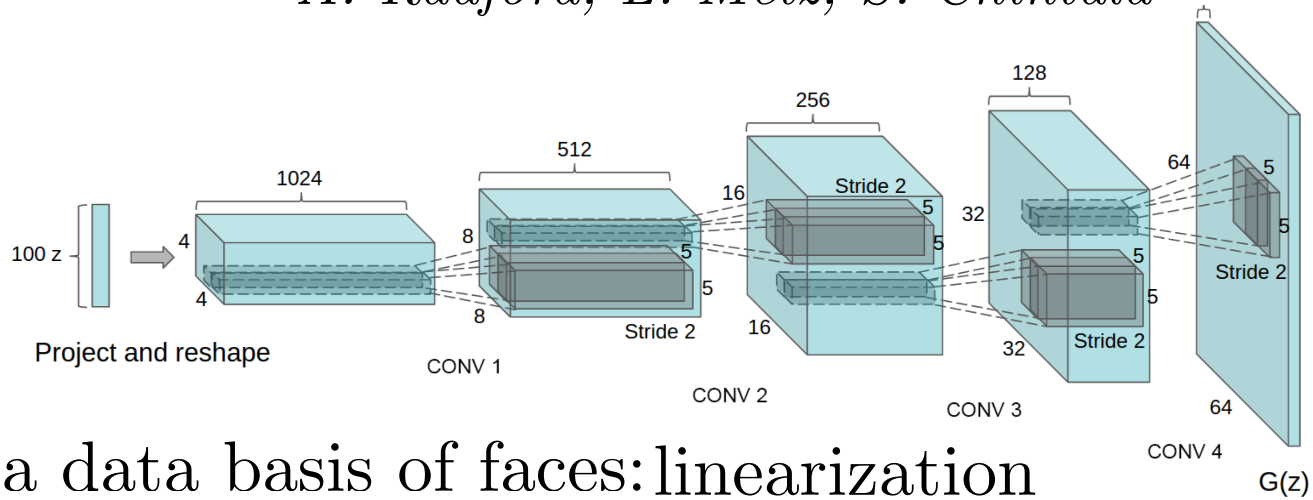
$$x \longrightarrow \boxed{S_J x} \longrightarrow \boxed{\begin{array}{c}\text{Supervised}\\ \text{Linear classifier}\end{array}} \longrightarrow y = f(x)$$

No learning

Data Basis	Deep-Net	Scat/Unsupervised
CIFAR-10	**7%**	20%

Linearisation in Deep Networks

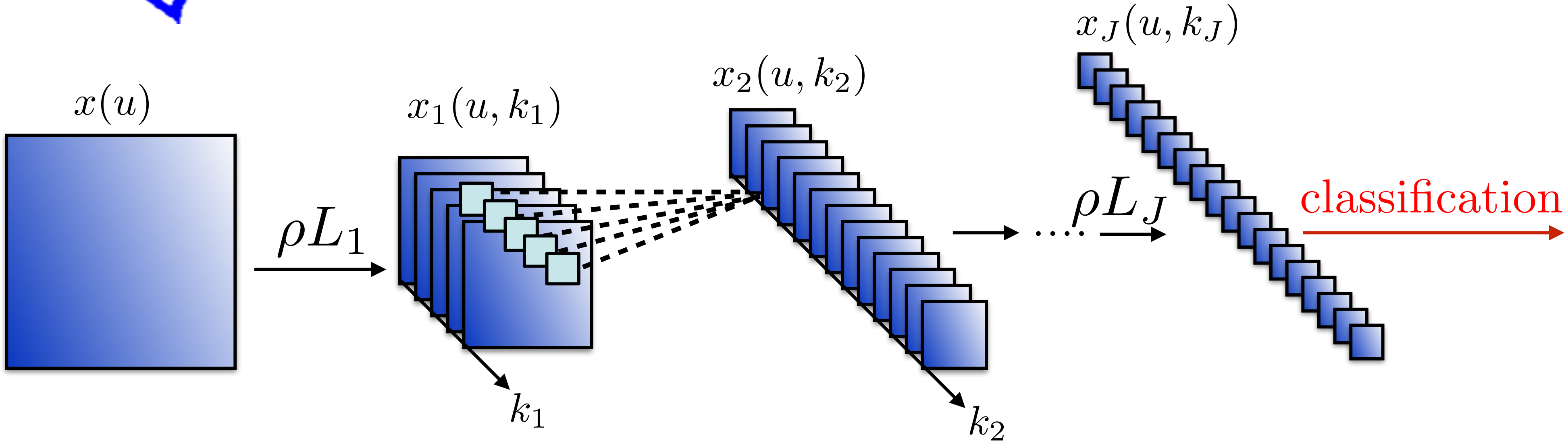
A. Radford, L. Metz, S. Chintala



- Trained on a data basis of faces: linearization

- On a data basis including bedrooms: interpolaitons

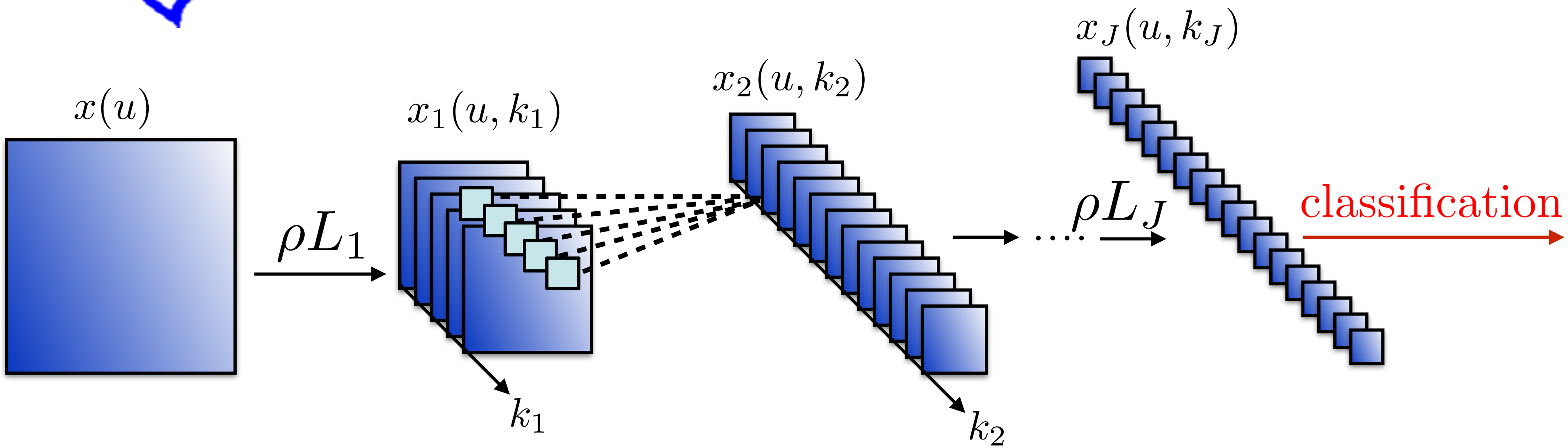
Deep Convolutional Networks



- The convolution network operators L_j have many roles:
  - Linearize non-linear transformations (symmetries)
  - Reduce dimension with projections
  - Memory storage of « characteristic » structures

- Difficult to separate these roles when analyzing learned networks

Open Problems



- Can we recover symmetry groups from the matrices *Lj* ?
- What kind of groups ?
- Can we characterise the regularity of *f(x)* from these groups ?
- Can we define classes of high-dimensional « regular » functions that are well approximated by deep neural networks ?
- Can we get approximation theorems giving errors depending on number of training exemples, with a fast decay ?

Conclusions

- Deep convolutional networks have spectacular high-dimensional approximation capabilities.

- Seem to compute hierarchical invariants of complex symmetries

- Used as models in physiological vision and audition

- Close link with particle and statistical physics

- Outstanding mathematical problem to understand them:
  notions of complexity, regularity, approximation theorems…

Understanding Deep Convolutional Networks, arXiv 2016.