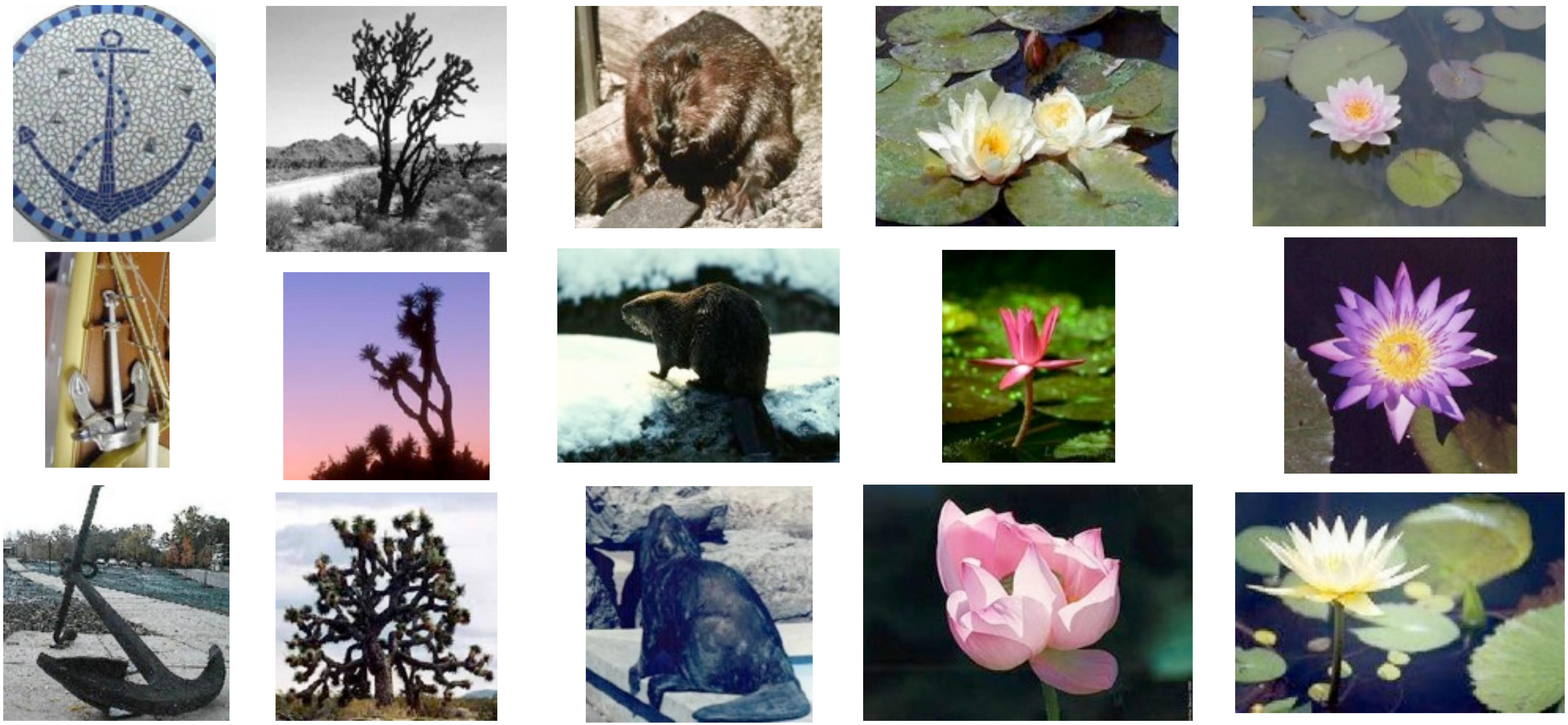
# Scattering Invariant Deep Networks for Classification


*Stéphane Mallat*
**IHES**
**Ecole Polytechnique**

# Image Classification

## CalTech 101:

Anchor | Joshua Tree | Beaver | Lotus | Water Lily



- Considerable variability in each class.
- A Euclidean norm does not measure signal «similarities».

# Metric for Classification

- Classification requires finding a metric to compare signals, with:
  - small distances $d(f,g)$ within a class
  - large distances $d(f,g)$ across classes.

- If one finds a representation $\Phi(f)$ such that

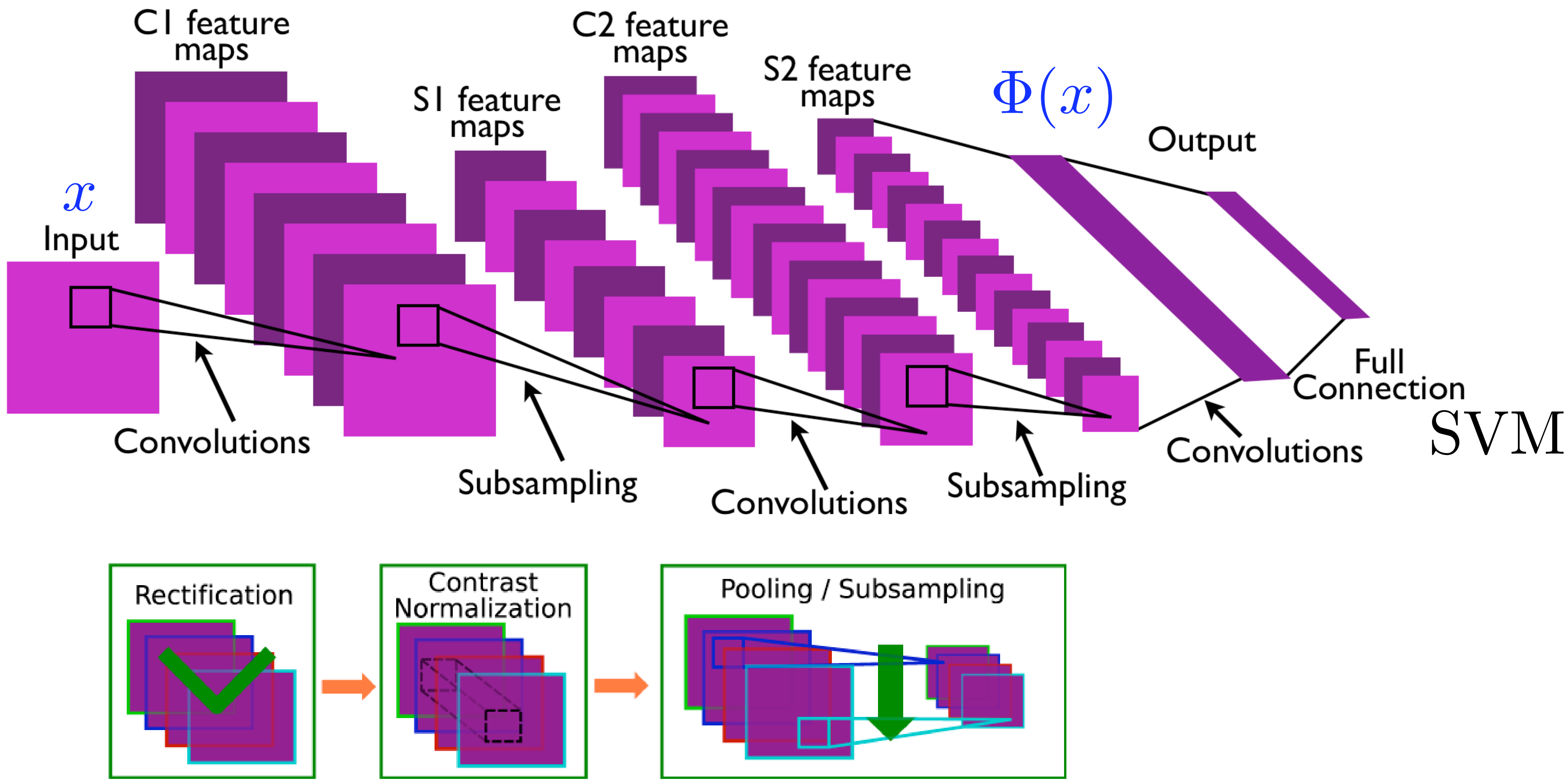
$$d(f,g) = \|\Phi(f) - \Phi(g)\| \quad \text{(kernel metric)}$$

  then the classification may be linearized (SVM, PCA,...).

- Is there an appropriate kernel metric, which $\Phi$ ?

# A View of Convolution Networks

*Y. LeCun et. al.*




- Deep convolution networks are very efficient image and audio classifiers: WHY ?

# Representation for Classification

- What principles to construct such representations ?

- Deep convolution networks:
  - Why convolutions ?
  - Which filters ?
  - Why multistage and how deep ?
  - Why pooling ? How to pool ?
  - Why non-linear, which non-linearities ?
  - Why normalizing ?
  - What is the role of sparsity ?

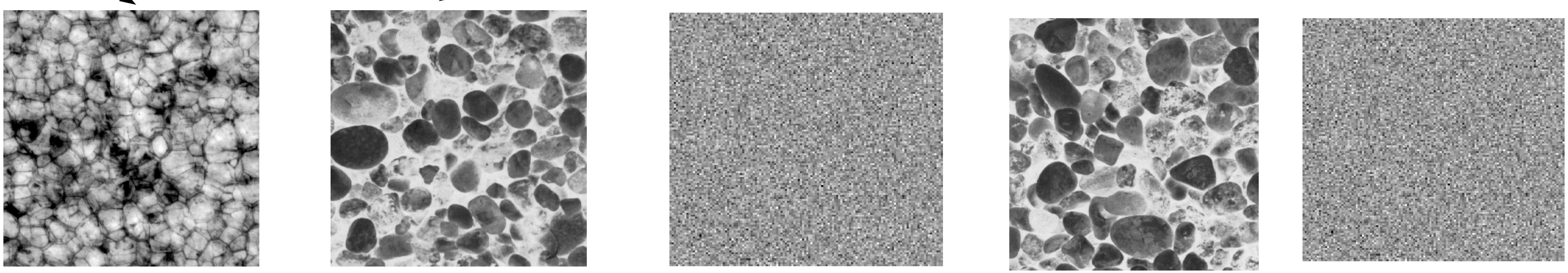
- What are the underlying useful mathematics ?

# Texture Discrimination

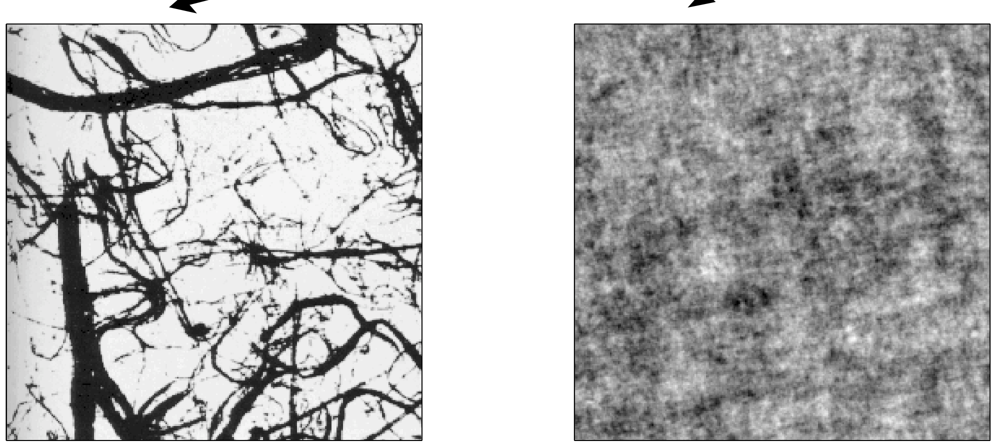
- Textures define high-dimensional image classes.

- Realizations of stationary processes $X$ but typically not Gaussian, not Markovian and not characterized by second order moments.

same power spectrum



same power spectrum

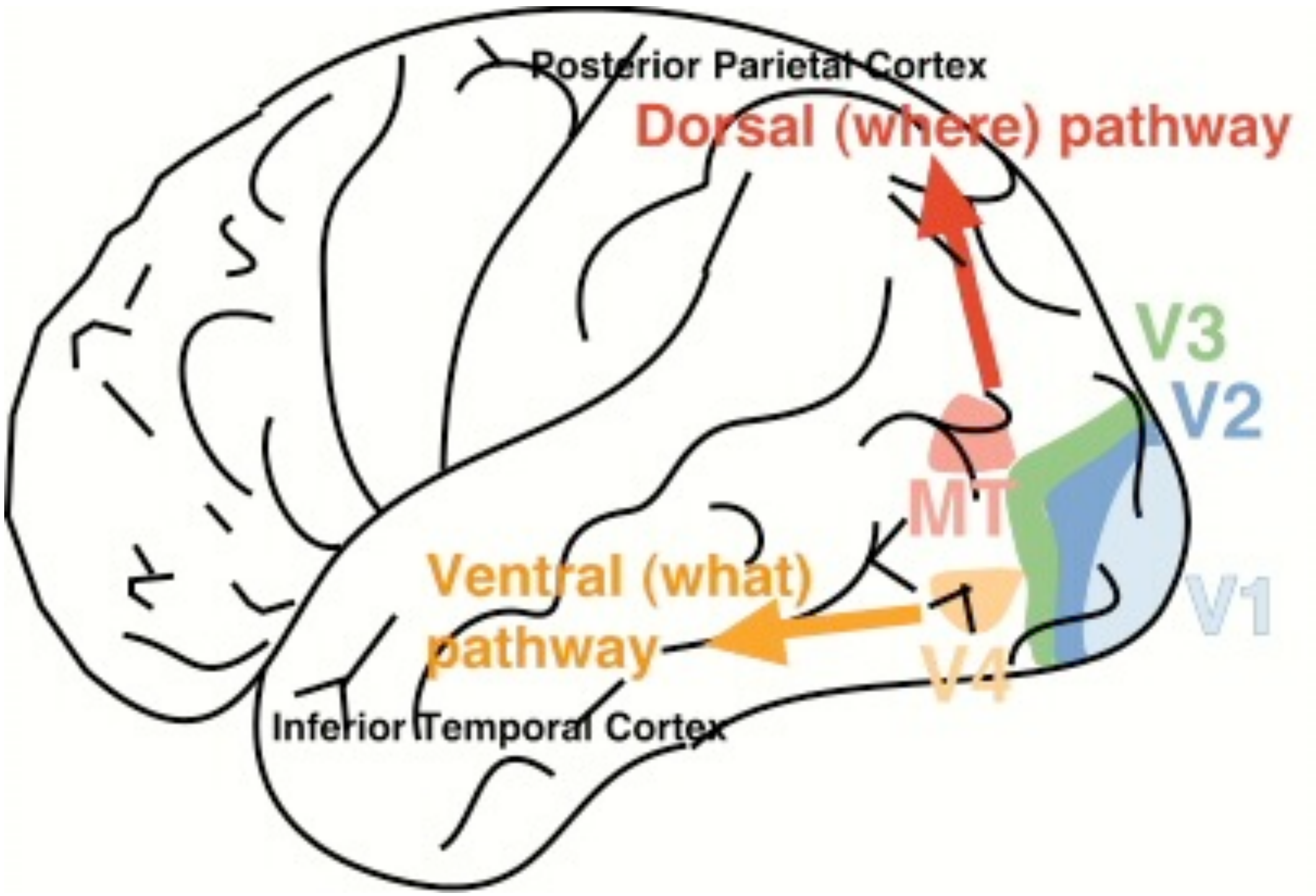


# Audio Textures

J. McDermott textures

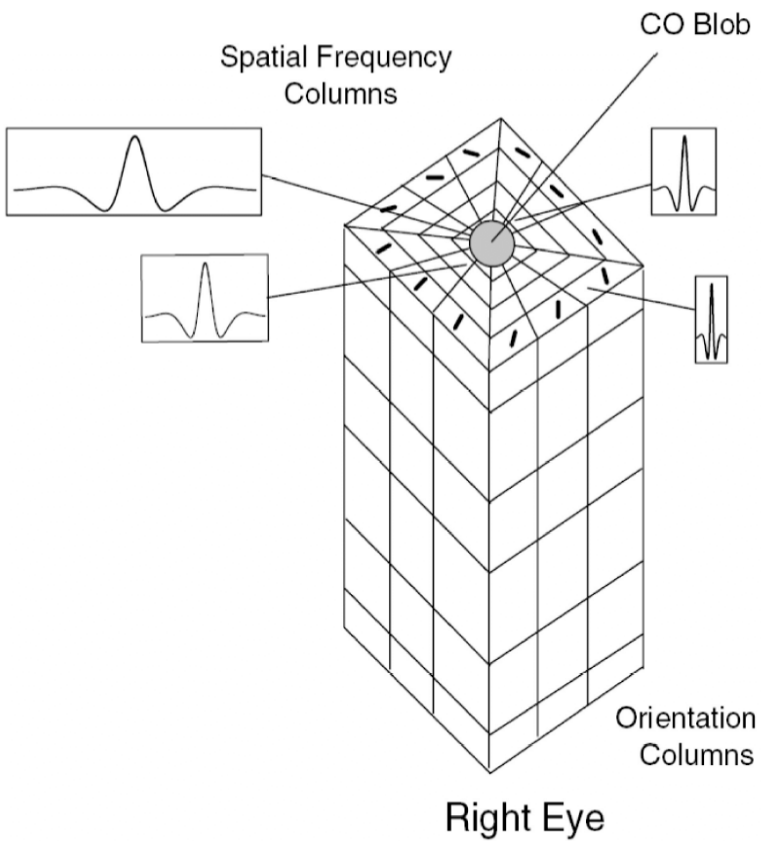
- Natural Sounds (1s)   Original   Gaussian model
  - Hammer
  - Insect
  - Water
  - Applause

# The Best Image Classifier

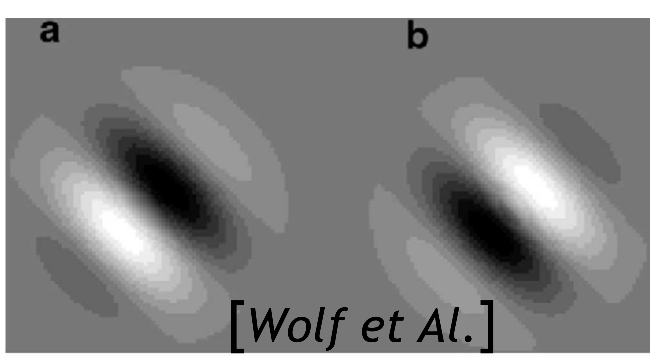


# Psychophysics of Vision

## **Hypercolumns in V1:**
directional wavelets



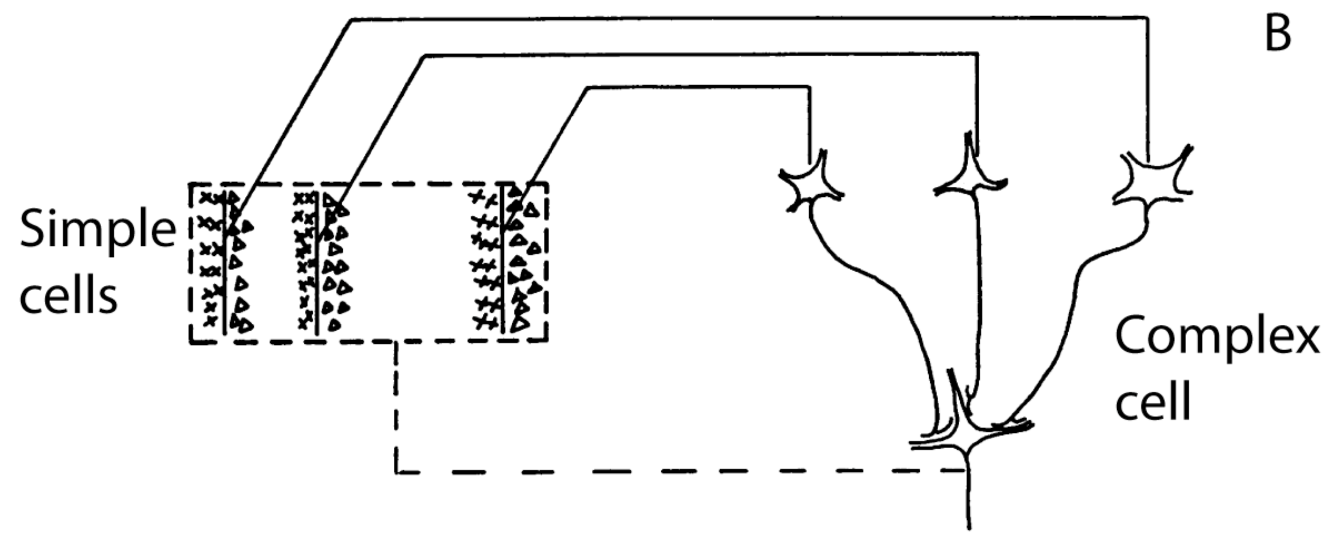

Simple cells Gabor linear models




$$\psi(x) = \theta(x) e^{i\xi x}$$

## **Complex Cells**

- Non-linear
- Large receptive fields
- Some forms of invariance

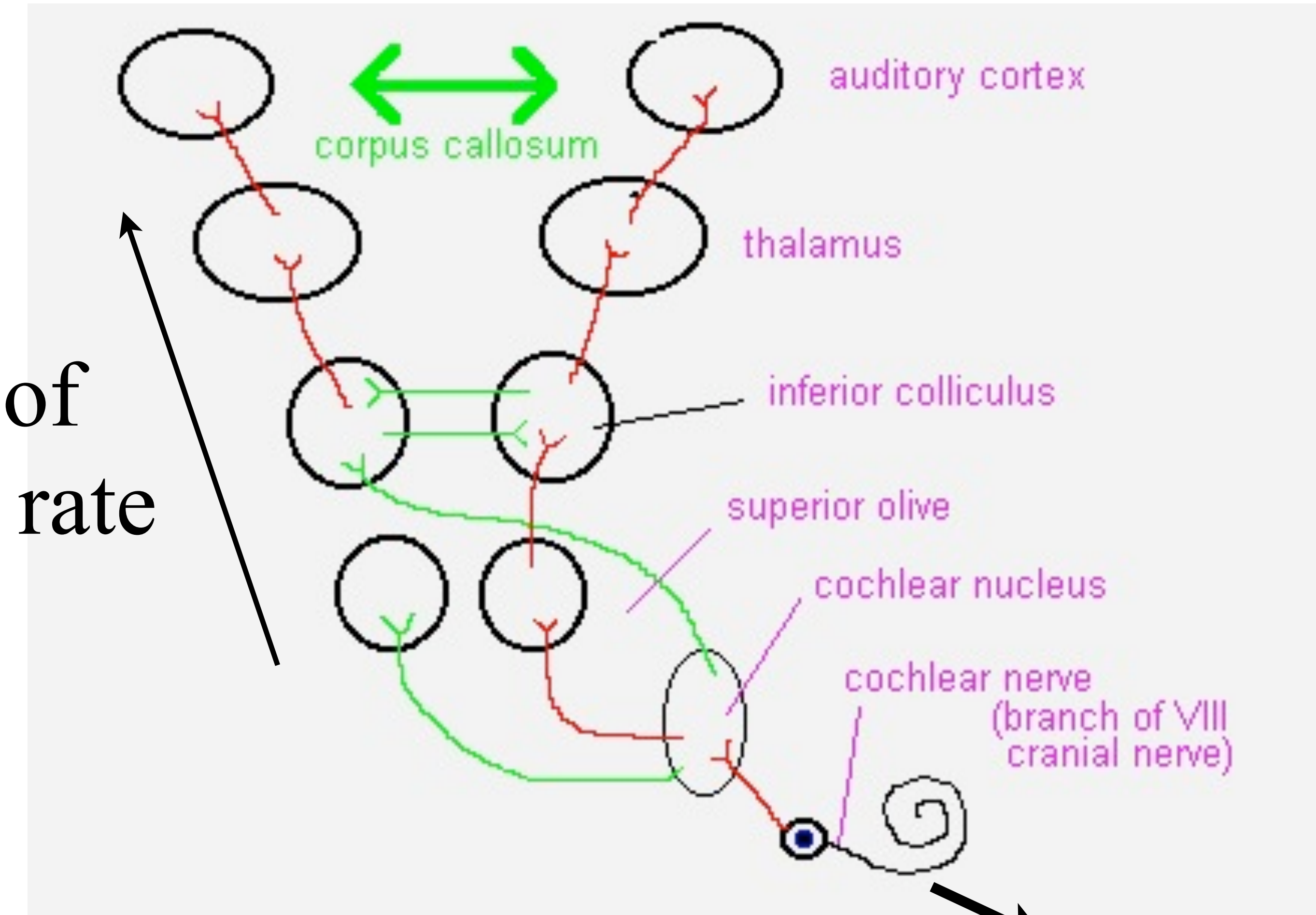



## **«What» Pathway towards V4:**

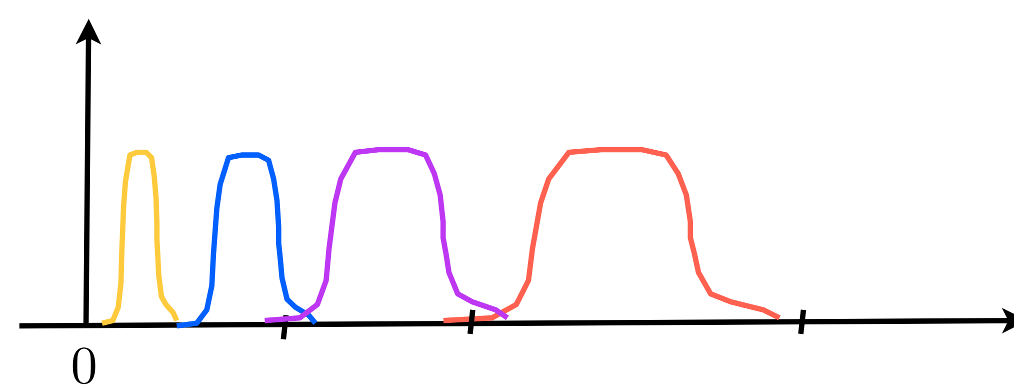
- More specialized invariance
- «Grand mother cells»

# Audio Psychophysics

Reduction of processing rate




**Cochlea:**
dilated wavelet filters




- Wavelets appear at early stages of vision and audition. WHY ?

# Low-Level Signal Representation

- Low-level signal processing:
  - compression/information theory for storage and transmission
  - inverse problems from partial and degraded measurements

- A key idea: find **sparse** accurate representations with few parameters.

- Mathematical tools: Fourier transform, **wavelet bases**, adaptive dictionary representations, variational formulations...
  A relatively well understood framework.

- Classification problems: discriminate not reconstruct.
- Different problems where sparsity yields ***instabilities***.

# Analysis versus Synthesis

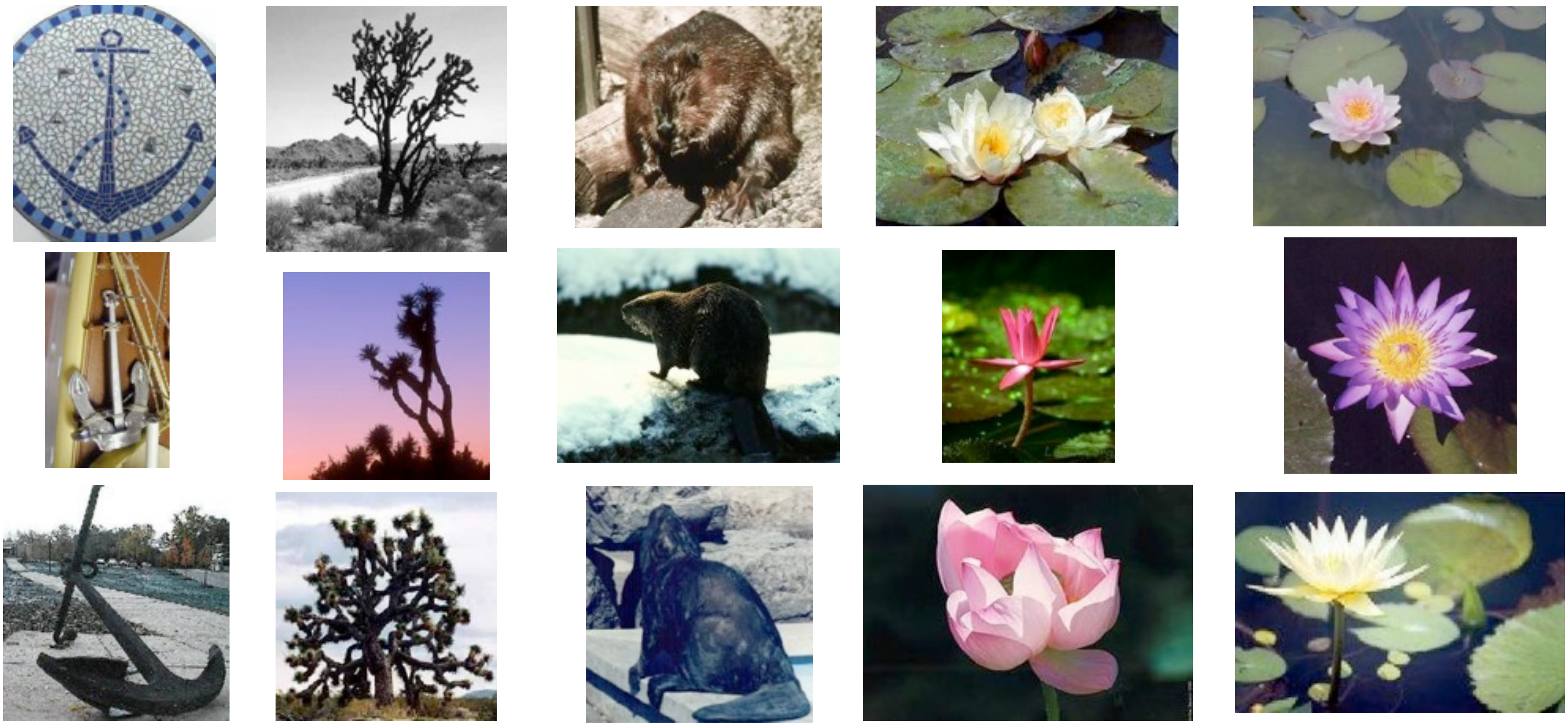
- How to construct a sparse representation ?

- What about stability ?

# Image Classification

## CalTech 101:

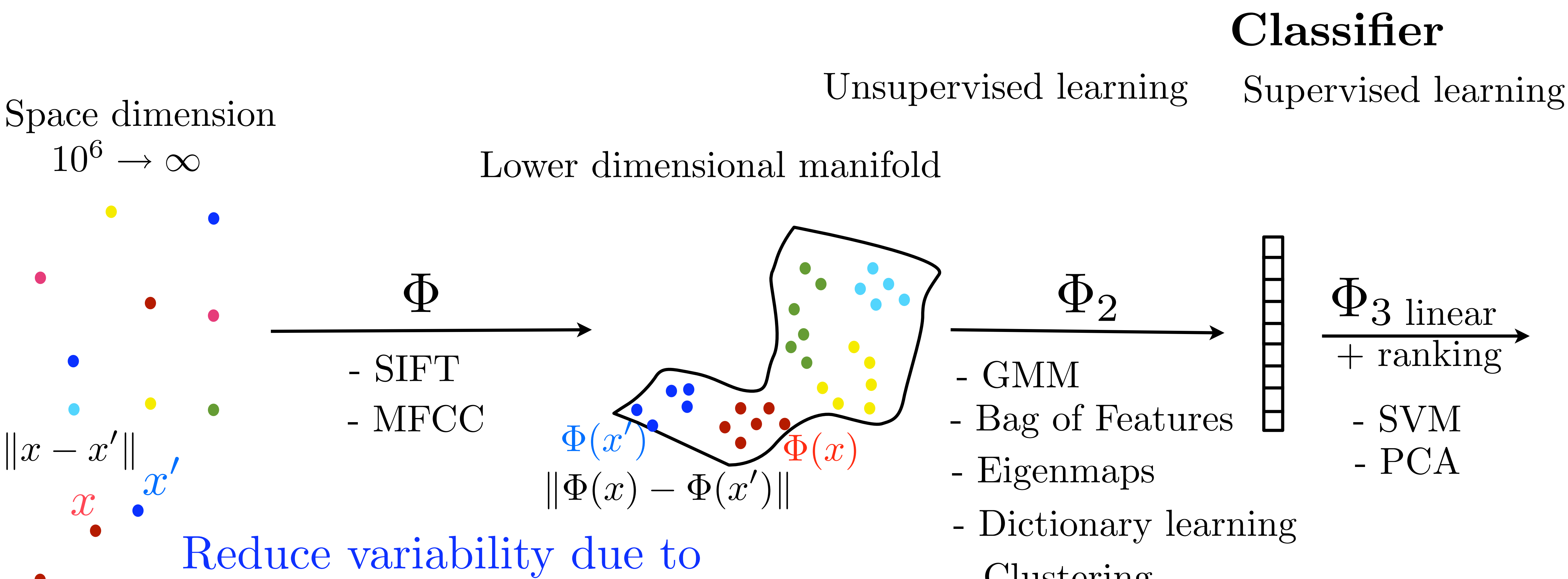
Anchor | Joshua Tree | Beaver | Lotus | Water Lily



- Considerable variability in each class.
- Reduce variability means constructing invariants.

# Signal Classification

- Very high dimensional space $N \geq 10^6$.
- Few training samples per class $P \ll N$.
- Signals do not belong to a low-dimensional manifold.




Reduce variability due to translations, transposition (audio) rotations, scaling (images)
*action of groups*

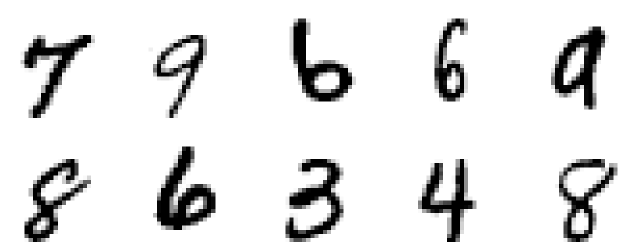
Reduce structural variability

# Stable Translation Invariants

- **Invariance** to translations $x_c(t) = x(t-c)$

$$\forall c \in \mathbf{R} \quad , \quad \Phi(x_c) = \Phi(x) \ .$$

- **Metric stability** with deformations $x_\tau(t) = x(t - \tau(t))$



small deformations of $x \implies$ small modifications of $\Phi(x)$

$$\forall \tau \quad , \quad \|\Phi(x_\tau) - \Phi(x)\| \leq C \ \underbrace{\sup_t |\nabla \tau(t)|}_{\text{deformation size}} \ \|x\| \ .$$

- **Preserve information**

# Overview

- **Part 1:** *Invariance and deformation stability*

- Fourier failure
- Wavelet stability to deformations
- Scattering invariants and deep convolution networks
- Mathematical properties of deep scattering networks
- Classification of images

- **Part 2:** *Inverse, Textures and Multiple Invariants*

- Inverse scattering by phase retrieval and sparsity
- Scattering models of stationary processes
- Texture classification
- Invariants over multiple groups: transposition, rotation, scaling

# Fourier & Correlation Invariance

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

- Translation Invariance: if $x_c(t) = x(t-c)$ then

$$|\widehat{x_c}(\omega)| = |\hat{x}(\omega)|$$

- For the auto-correlation $Cx(u) = \int x(t)\, x(t-u)\, dt$

$$Cx(u) = Cx_c(u)\ .$$

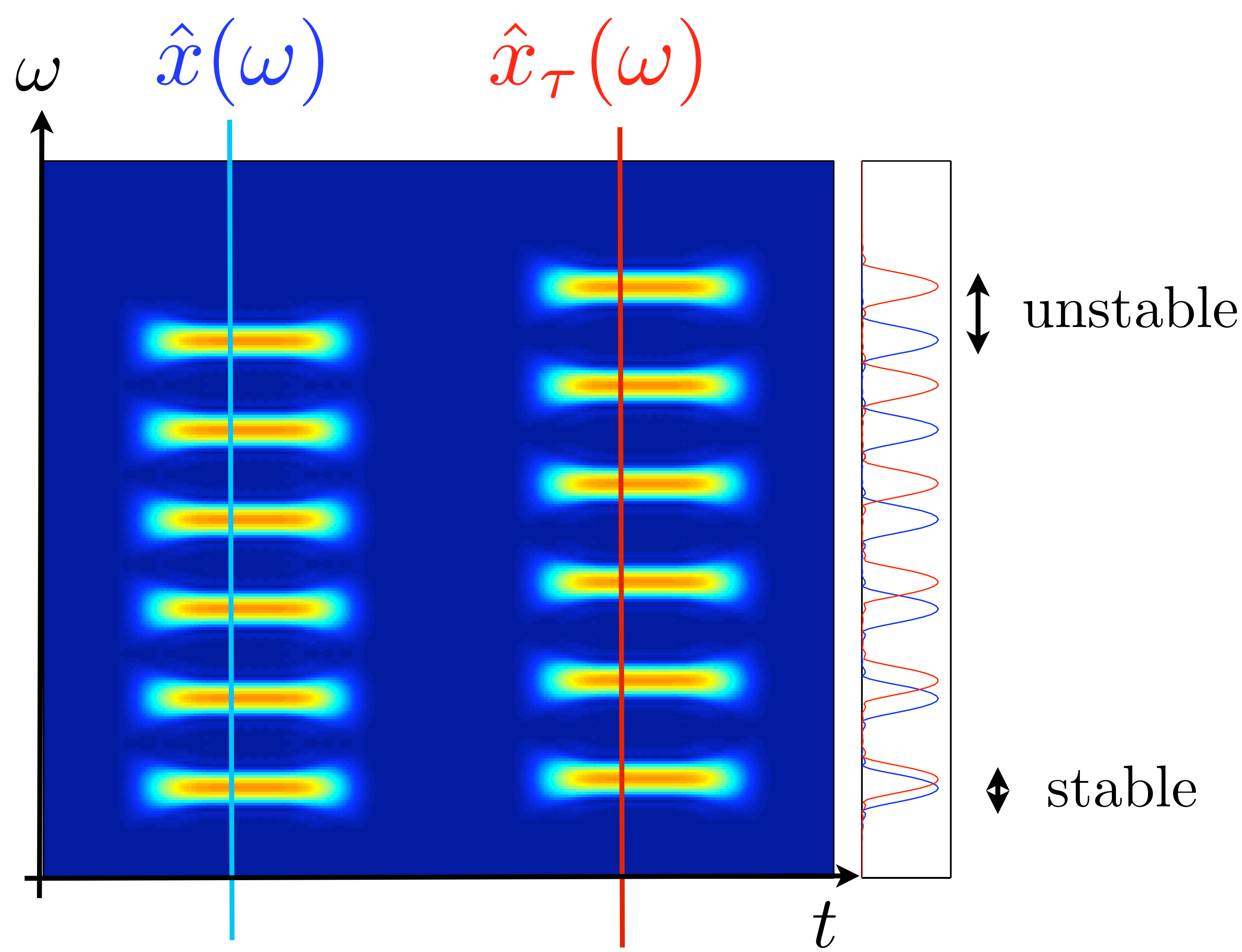
# Fourier & Correlation Instabilities

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

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

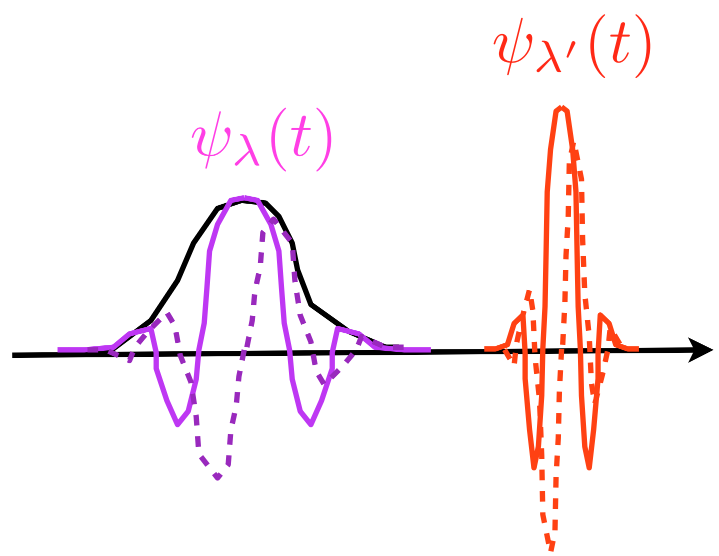
$$\Rightarrow \; \| |\hat{x}_\tau|^2 - |\hat{x}|^2 \| = \| C x_\tau - C x \| \text{ is big .}$$

$$\tau(t) = \epsilon\, t$$

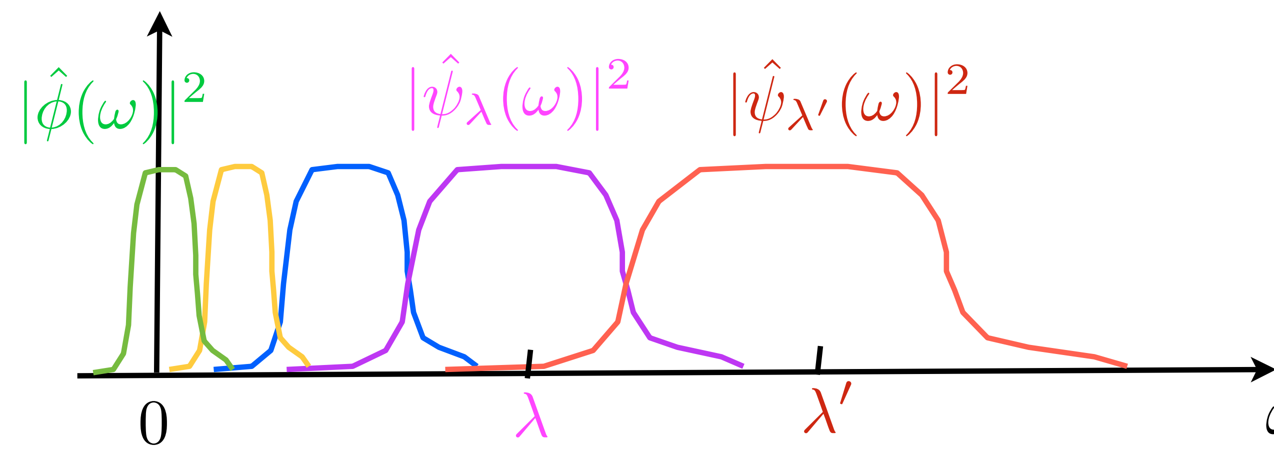


# Wavelet Transform

- Dilated wavelets: $\psi_\lambda(t) = 2^{-jQ}\,\psi(2^{-jQ} t)$ with $\lambda = 2^{-jQ}$ .




quadrature pairs




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

- Wavelet transform: $Wx(t) = \Big\{ x \star \phi\,(t)\ ,\ x \star \psi_\lambda(t) \Big\}_\lambda$

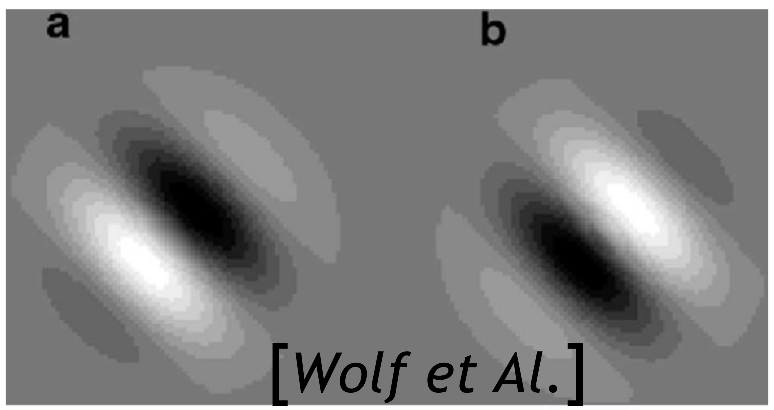
- If $|\hat\phi(\omega)|^2 + \sum_\lambda |\hat\psi_\lambda(\omega)|^2 = 1$ then $W$ is unitary :

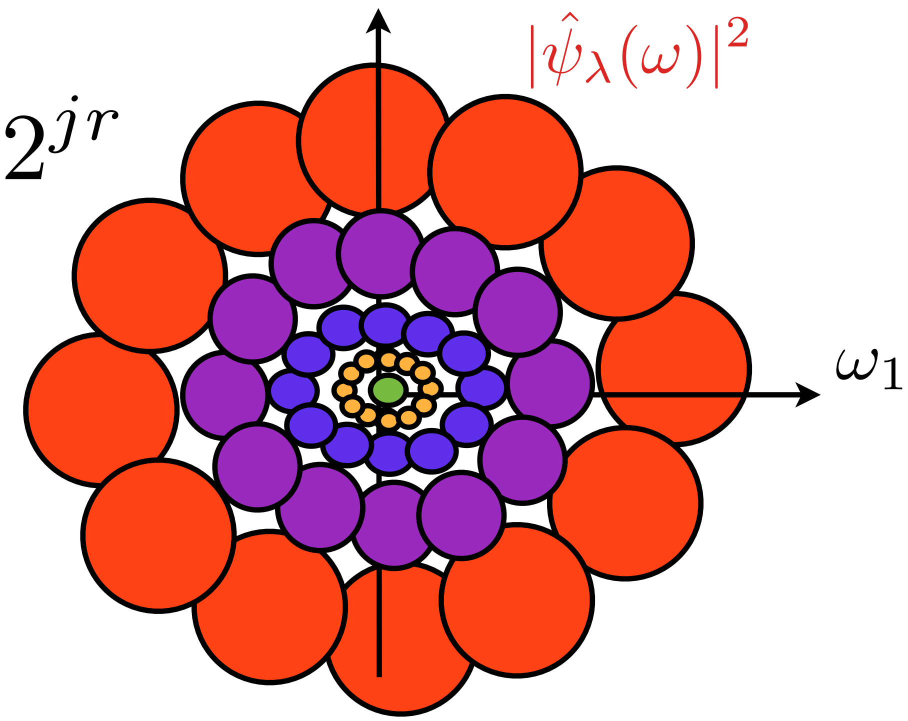
$$\|Wx\|^2 = \|x \star \phi\,\|^2 + \sum_\lambda \|x \star \psi_\lambda\|^2 = \|x\|^2 .$$

# Wavelet Transform

- For images, dilated and rotated wavelets:

$$\psi_\lambda(t) = 2^j\, \psi(2^j r t) \quad \text{with} \quad \lambda = 2^j r$$



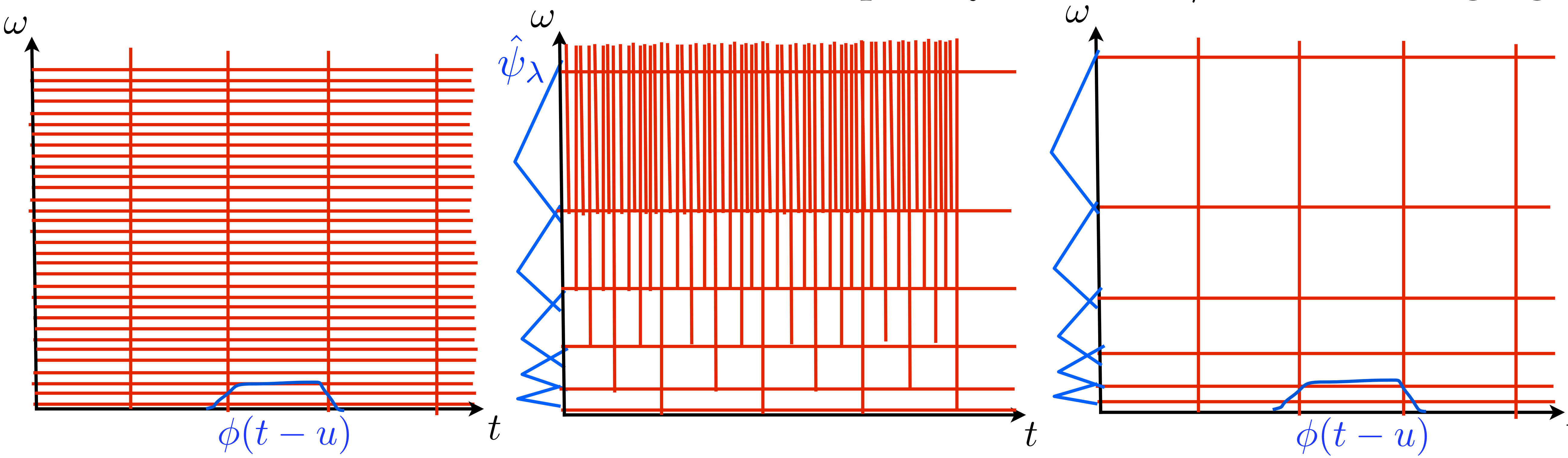

- Wavelet transform: $W x(t) = \Big\{ x \star \phi\,(t)\ ,\ x \star \psi_\lambda(t) \Big\}_\lambda$

- If $|\hat{\phi}(\omega)|^2 + \sum_\lambda |\hat{\psi}_\lambda(\omega)|^2 = 1$ then $W$ is unitary :

# Wavelet Stabilization

$$\left\{ |x \star \psi_\lambda(t)| \right\}_\lambda \qquad \left\{ |x \star \psi_\lambda| \star \phi(t) \right\}_\lambda$$

Window Fourier — Wavelet time-frequency — Time/Space averaging




Locally invariant to translations

and stable to deformations

MFSC (audio)

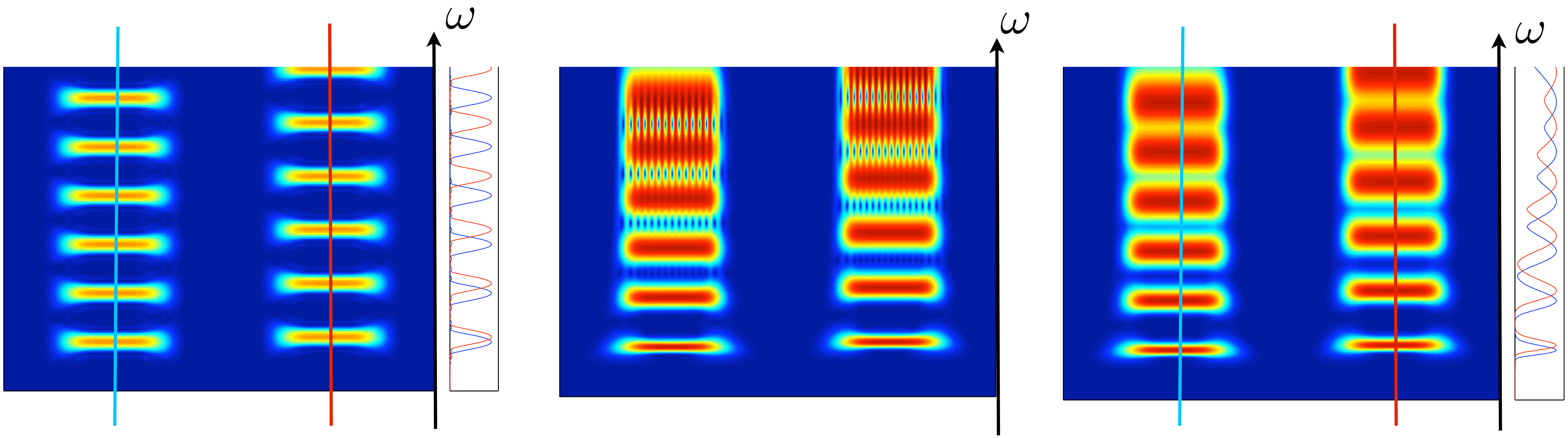
SIFT (images)

**But loss of information.**

# Wavelet Stabilization

$$\left\{ |x \star \psi_\lambda(t)| \right\}_\lambda \qquad \left\{ |x \star \psi_\lambda| \star \phi(t) \right\}_\lambda$$

Window Fourier | Wavelet time-frequency | Time/Space averaging




Locally invariant to translations

and stable to deformations

MFSC (audio)

SIFT (images)

**But loss of information.**

# Wavelet Stabilization

$$\left\{ |x \star \psi_\lambda(t)| \right\}_\lambda$$

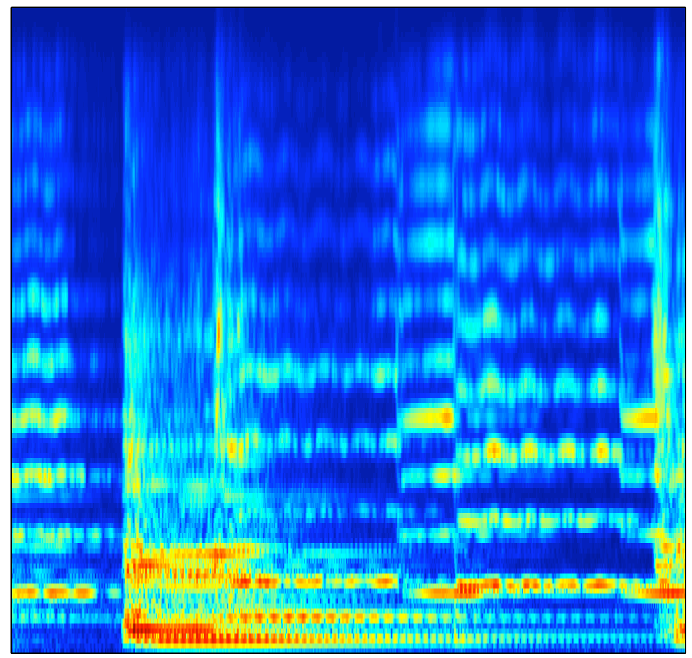
Wavelet time-frequency

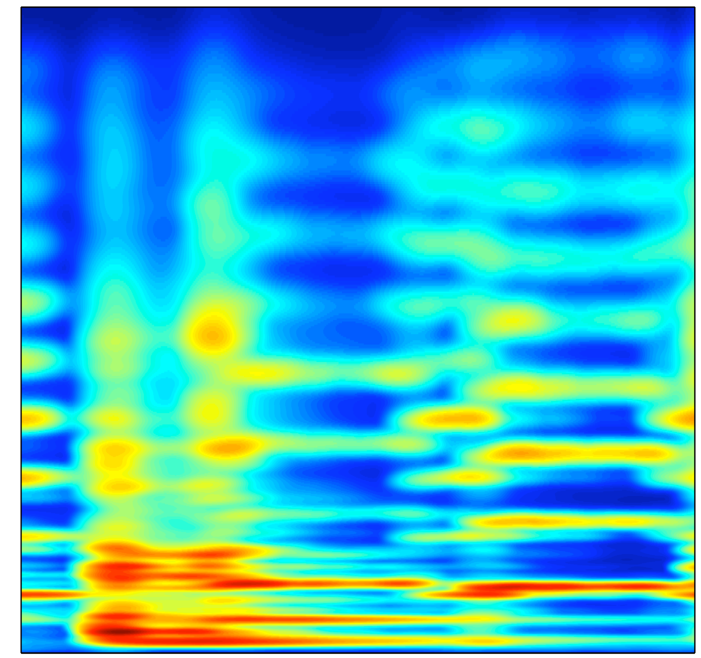
$$\left\{ |x \star \psi_\lambda| \star \phi(t) \right\}_\lambda$$

Time/Space averaging
370ms window

Non-linearity is needed to have a non-zero invariant

A modulus is "optimal"





Locally invariant to translations and stable to deformations

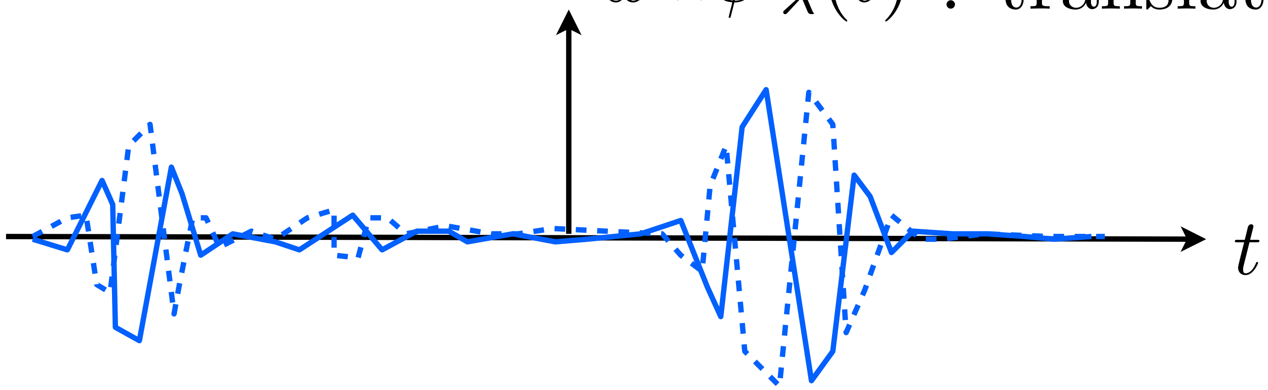
MFSC (audio)

SIFT (images)

**But loss of information.**

# Stable Translation Invariance

$x \star \psi_\lambda(t)$ : translation covariant, not invariant, and




$$\int x \star \psi_\lambda(t)\, dt = 0$$

- Translation invariant representation: $\displaystyle\int M(x \star \psi_\lambda)(t)\, dt$

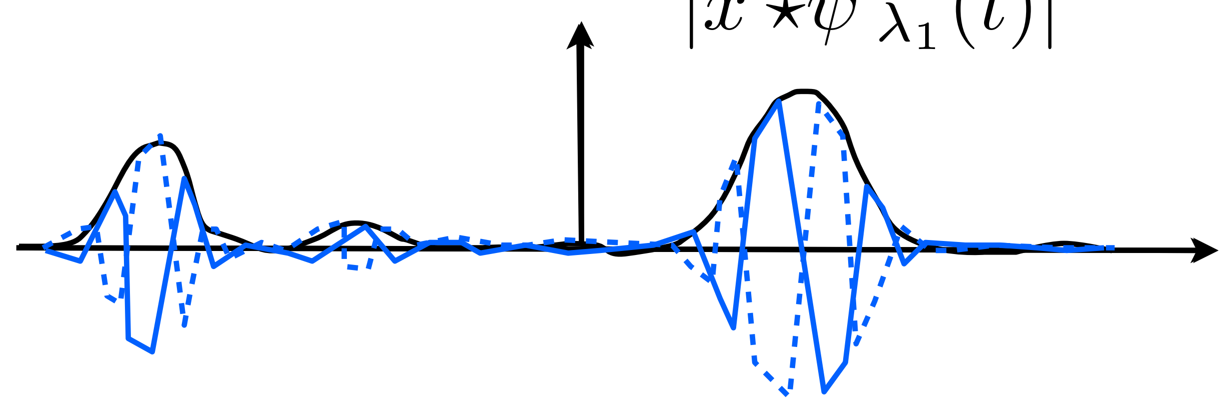
- Diffeomorphism stability: $M$ commutes with diffeomorphisms.

- $\mathbf{L^2}$ stability: $\|Mh\| = \|h\|$ and $\|Mg - Mh\| \leq \|g - h\|$

$$\Rightarrow \quad M(h)(t) = |h(t)| = \sqrt{|h_r(t)|^2 + |h_i(t)|^2}$$

- A modulus computes a lower frequency envelop

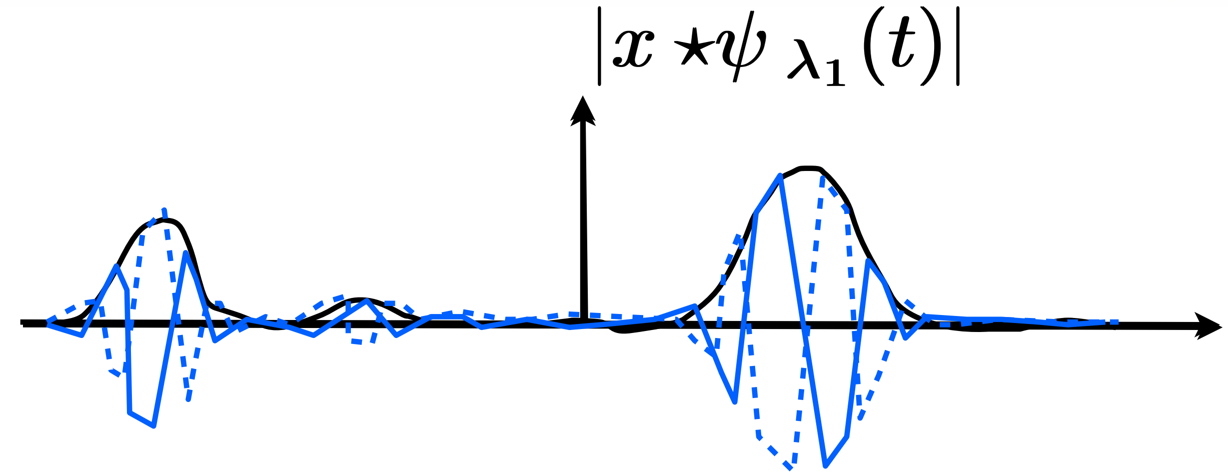
$|x \star \psi_{\lambda_1}(t)|$



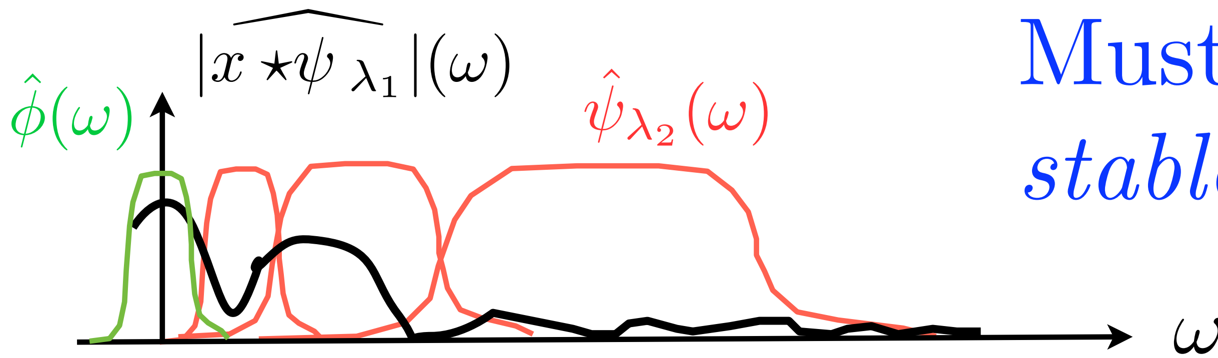
- Stable invariant: $\displaystyle\int |x \star \psi_\lambda(t)|\, dt = \|x \star \psi_\lambda\|_1$ .

# Recovering Lost Information

- A modulus computes a lower frequency envelop




- The averaging $|x \star \psi_{\lambda_1}| \star \phi$ removes high frequencies:




Must recover high frequencies: *stable modulation spectrum*

- Wavelet transform: $\{|x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\}_{\lambda_2}$

- Translation invariance by time averaging the amplitude:

$$\forall \lambda_1, \lambda_2, \quad ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi \quad : \text{ stable to deformations}$$

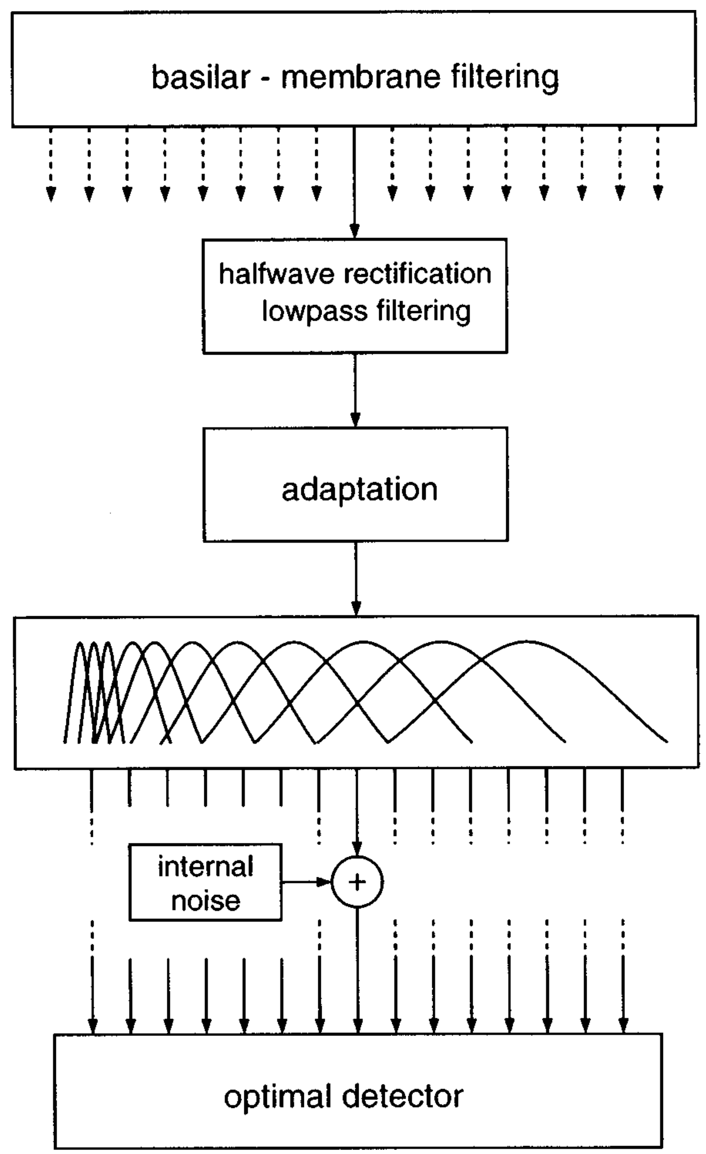
# Windowed Scattering

For any path $p = (\lambda_1, \lambda_2, ..., \lambda_m)$ of order $m$

$$S[p]x(t) = | \, |x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \dots | \star \psi_{\lambda_m}| \star \phi(t)$$

A window of size $N$ yields $O(Q^m \log^m N)$ coefficients of order $m$

First two orders:
Törsten Dau model




1 channel
$x \star \psi_{\lambda_1}$

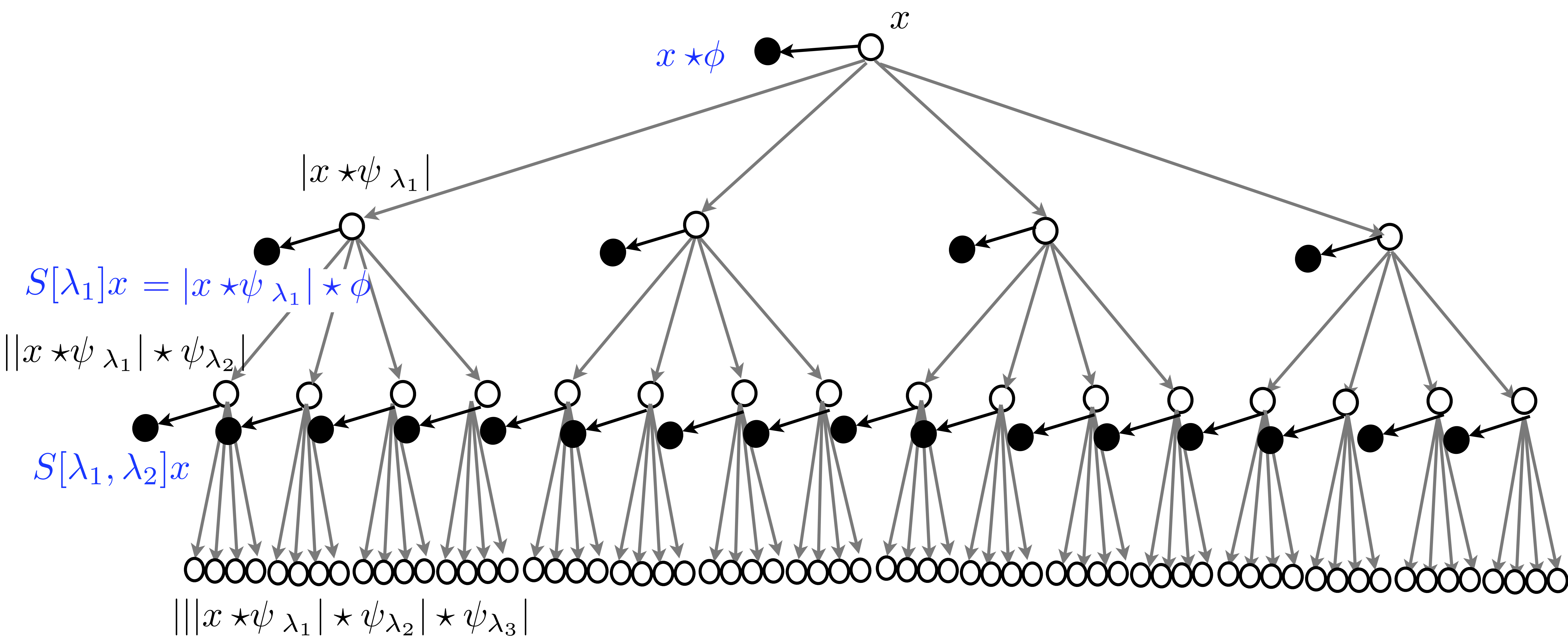
$|x \star \psi_{\lambda_1}|$
$Q \log N$ channels

$||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|$
$(Q \log N)^2$ channels

# Deep Convolution Network

*Y. LeCun et. al.*

- Iteration on $Ux = \{x \star \phi\,,\ |x \star \psi_{\lambda}|\}_{\lambda}$ , contracting.




- Output at **all layers**: $\{S[p]x\}_{p \in \mathcal{P}}$ .

  MFSC and SIFT are 1st layer outputs: $S[\lambda_1]x$

# Amplitude Modulations

- Amplitude modulations such as tremolos or attacks:

$$x(t) = h \star e\,(t)\,.\,a(t) \quad \text{with} \quad e(t) = \sum_n \delta(t - n/\xi_1)$$

$\hat{h}(\omega)$: formant, $\xi_1$ : pitch, $a(t)$: amplitude modulation

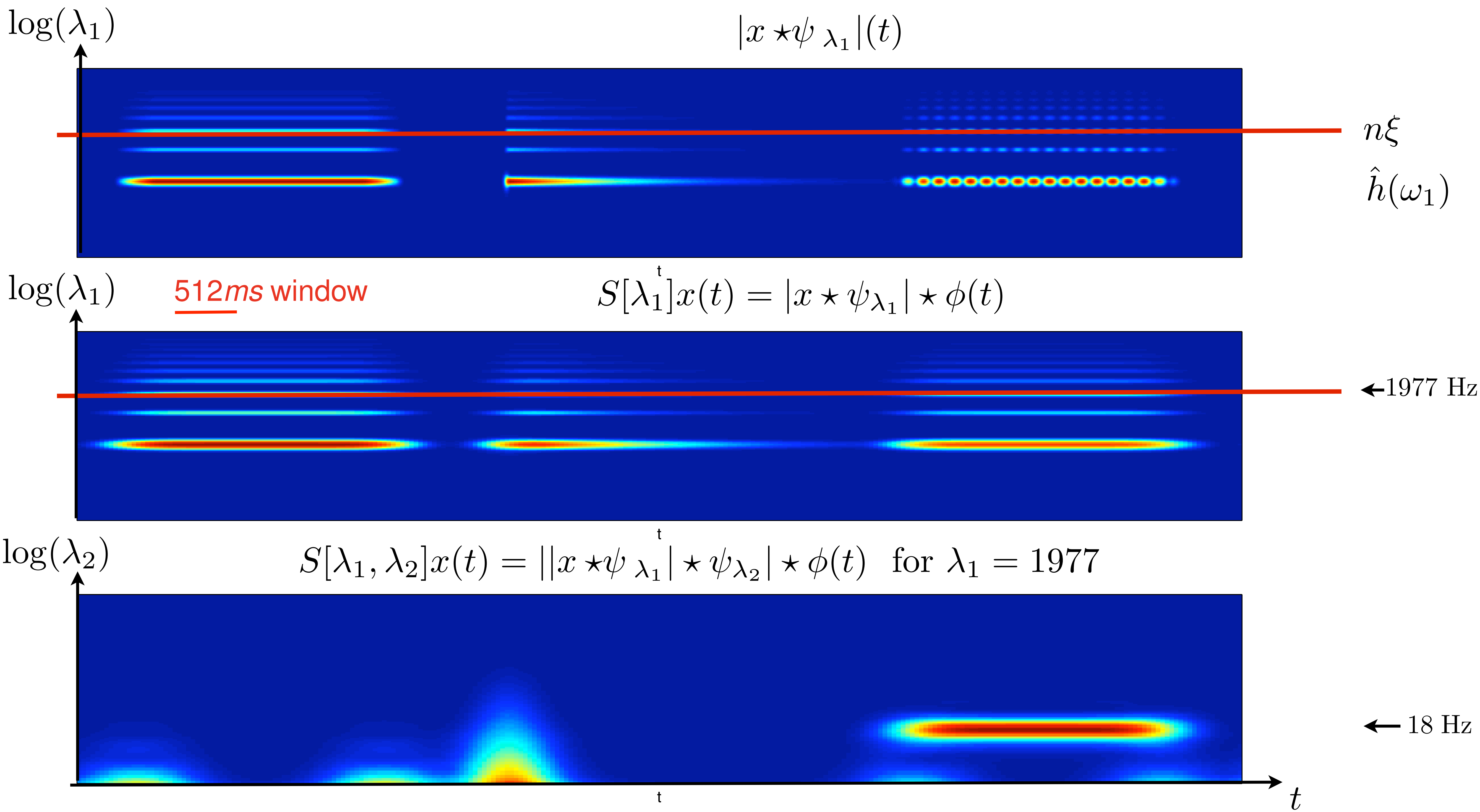
- Pitch harmonics: if $\lambda_1 = k\,\xi_1$ then

$$S[\lambda_1]x(t) = |x \star \psi_{\lambda_1}| \star \phi(t) = |\hat{h}(\lambda_1)|\; a \star \phi\,(t)$$

- Amplitude modulation spectrum:

$$S[\lambda_1, \lambda_2]x(t) = |\,|f \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi(t) = |\hat{h}(\lambda_1)|\; |\hat{a}(\lambda_2)|$$

# Amplitude Modulation



# Frequency Modulated Sounds

- Frequence modulations such as vibratos:

$$x(t) = h \star \widetilde{e}(t) \quad \text{with} \quad \widetilde{e}(t) = \sum_n \delta(t - \epsilon \cos \xi_2 t - n/\xi_1) \ .$$

$\hat{h}(\omega)$: formant, $\quad \xi_1$: pitch, $\quad \xi_2$: vibrato frequency.

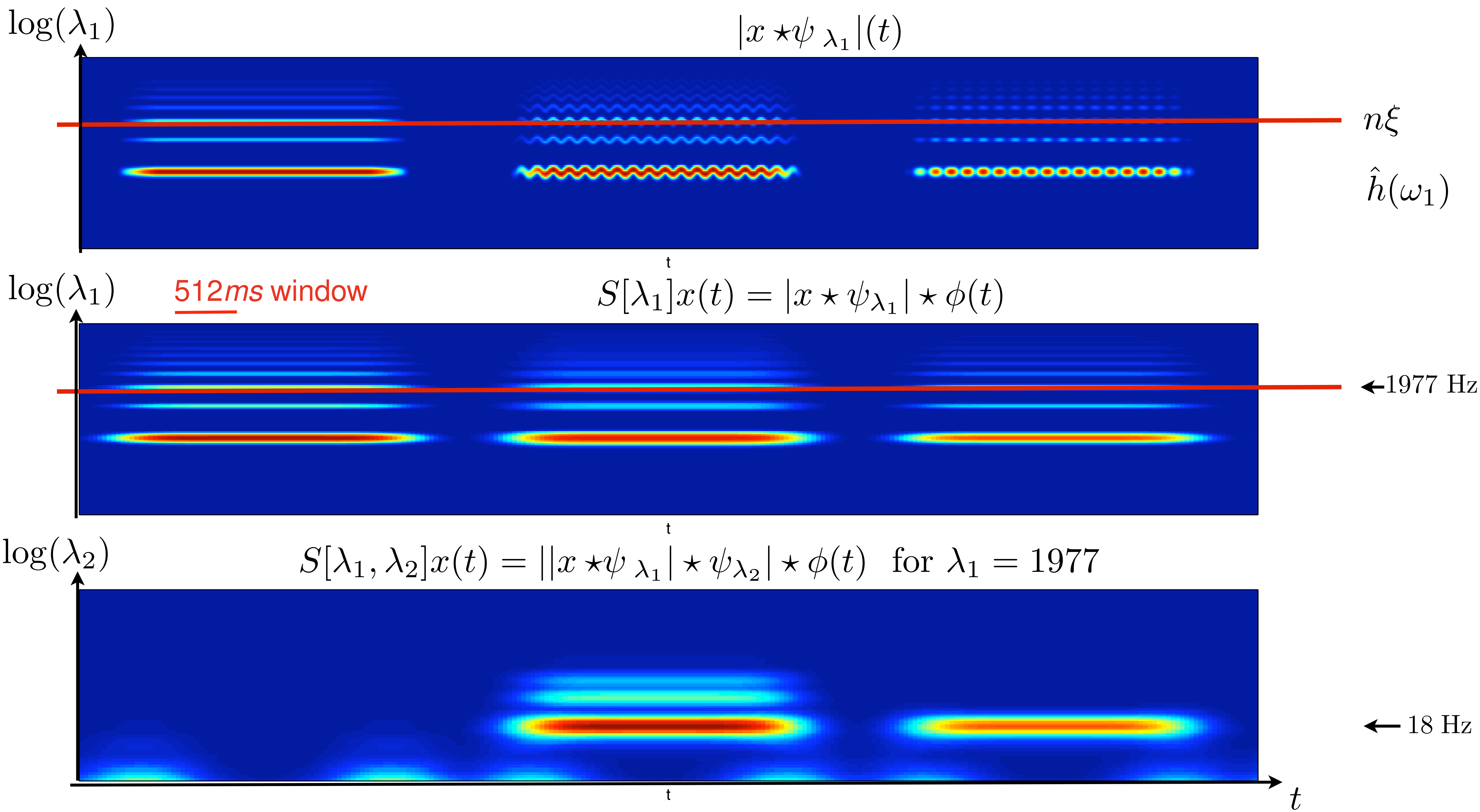
- Pitch harmonics: if $\lambda_1 = k\, \xi_1$ then

$$S[\lambda_1] x(t) = |\hat{h}(\lambda_1)|$$

- Vibrato harmonics: if $\lambda_2 = l\, \xi_2$ then

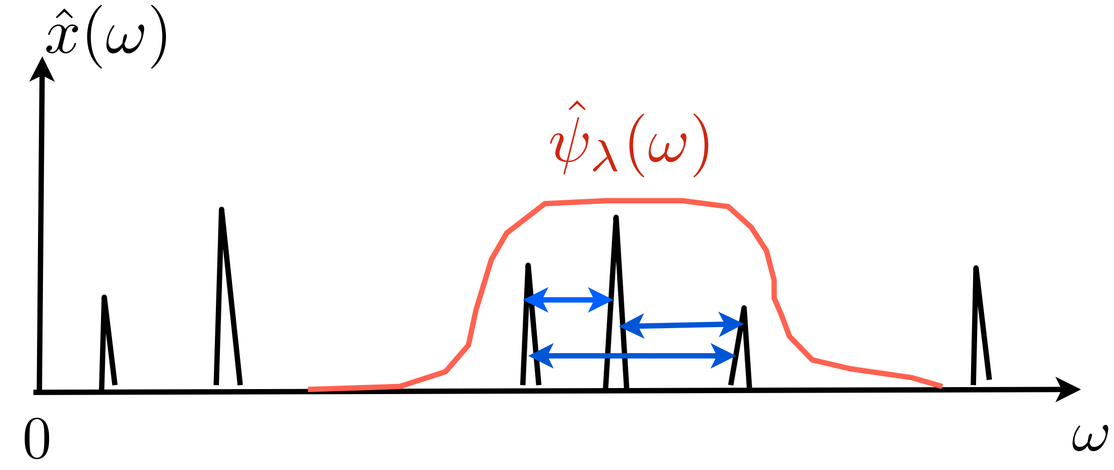
$$S[\lambda_1, \lambda_2] x(t) = C_l \, |\hat{h}(\lambda_1)| \, \epsilon^{2l} \, \xi_2^{2l}$$

# Frequency Modulation



# Interferences

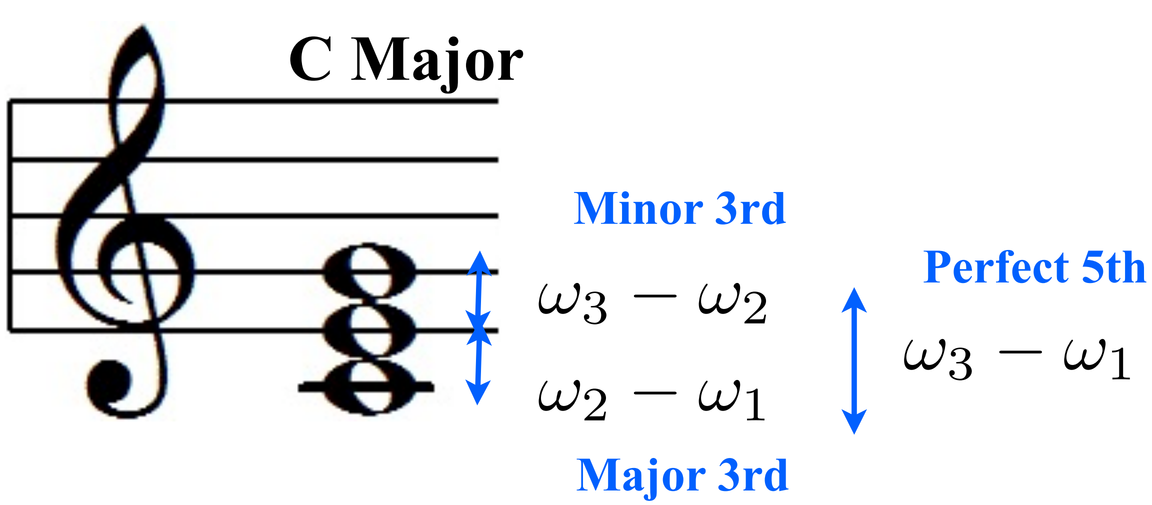
$$x(t) = \sum_m a_m \cos(\omega_m t)$$

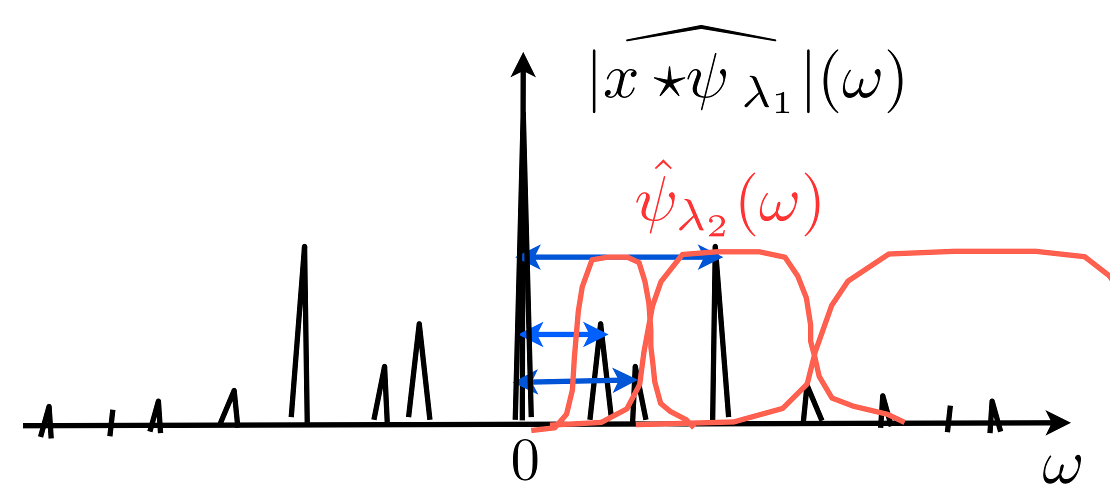



***Interferences :***

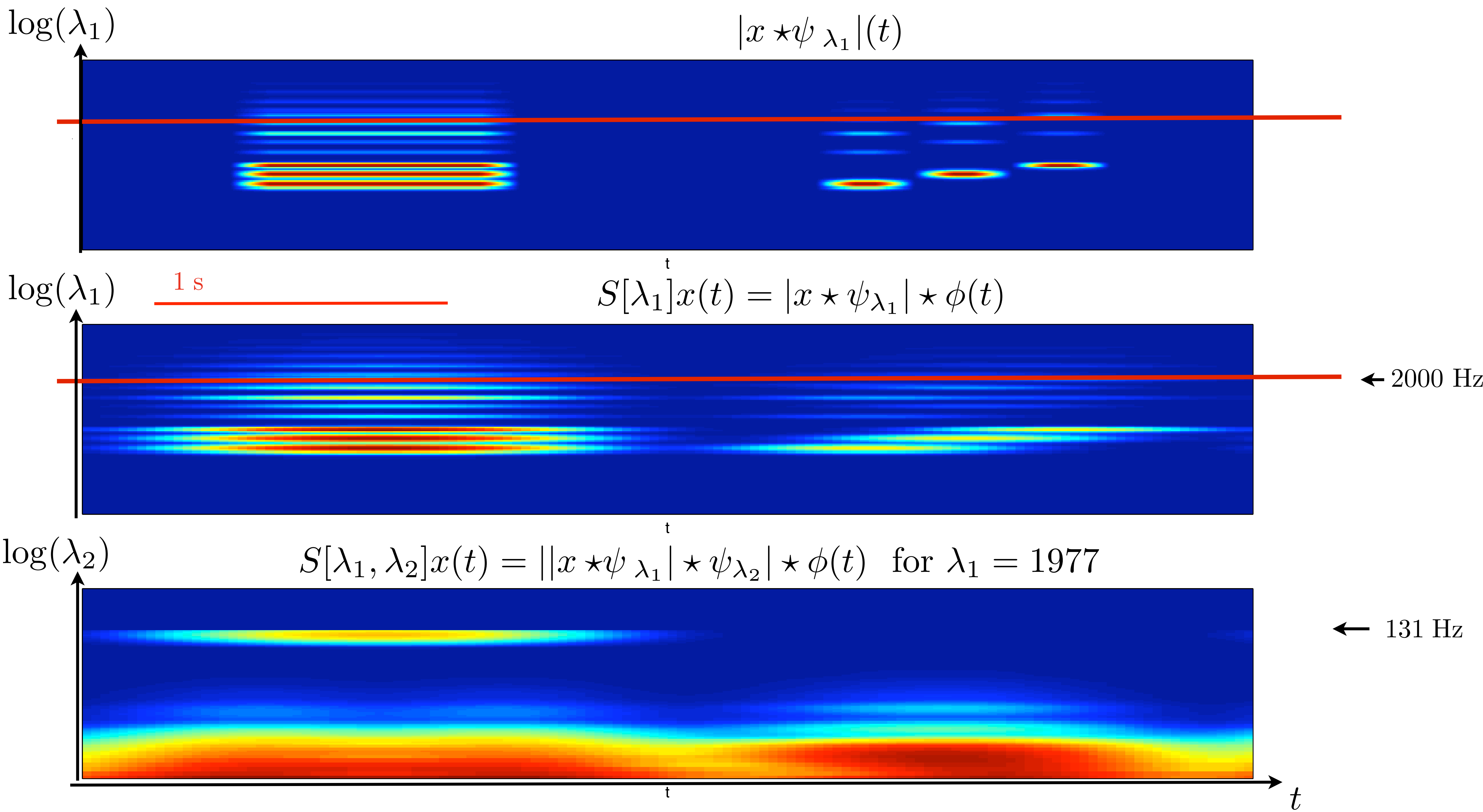
$$|x \star \psi_{\lambda}(t)|^2 = e_\lambda^2 + \sum_{m' \neq m} c_{m,m'} \cos(\omega_m - \omega_{m'})t$$

Music chord :



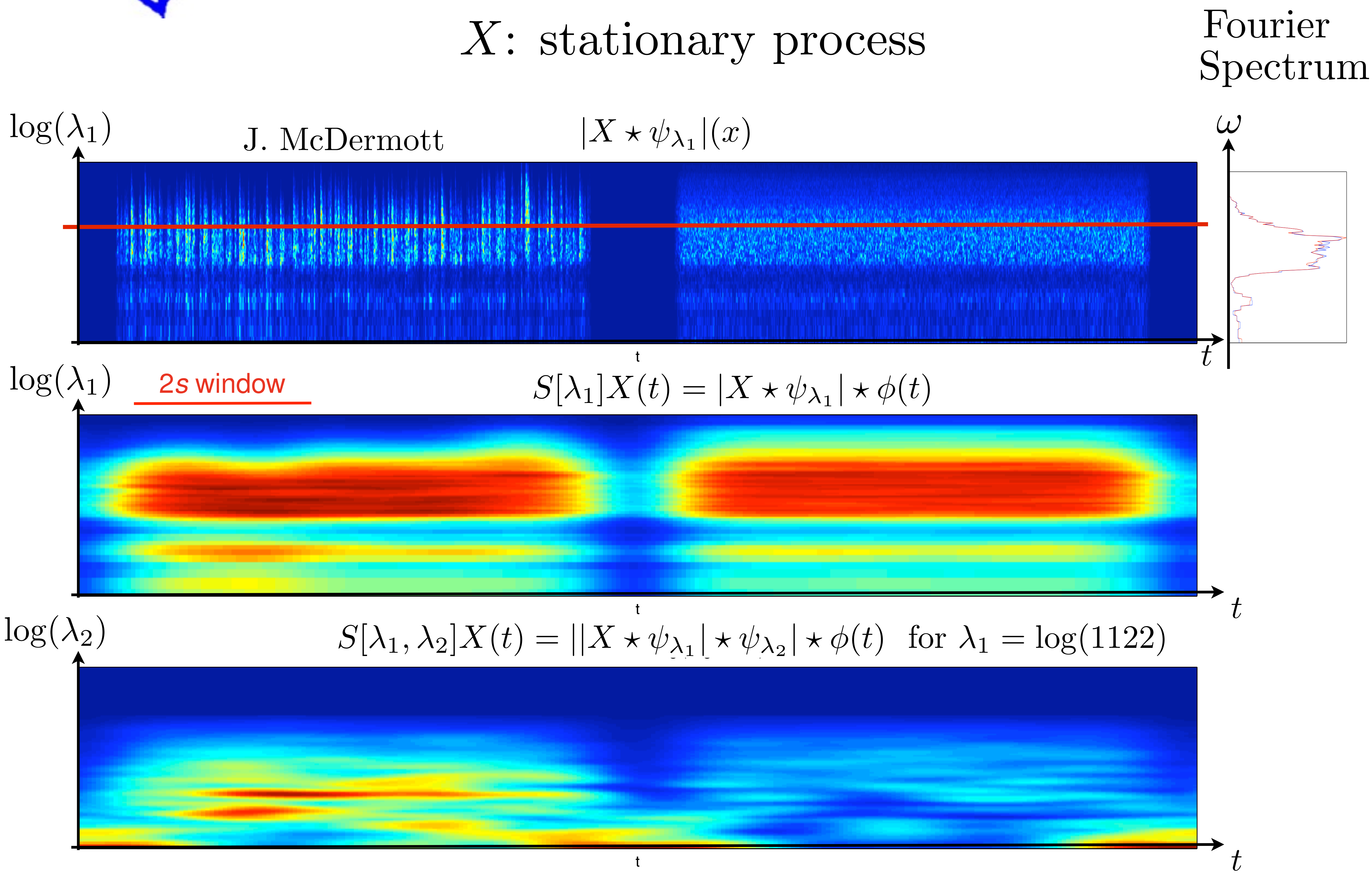



# Arpeggio


$\log(\lambda_1)$
$|x \star \psi_{\lambda_1}|(t)$
t
$\log(\lambda_1)$
1 s
$S[\lambda_1]x(t) = |x \star \psi_{\lambda_1}| \star \phi(t)$
2000 Hz
t
$\log(\lambda_2)$
$S[\lambda_1, \lambda_2]x(t) = ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi(t)$ for $\lambda_1 = 1977$
131 Hz
t
$t$

# Sounds with Same Spectrum

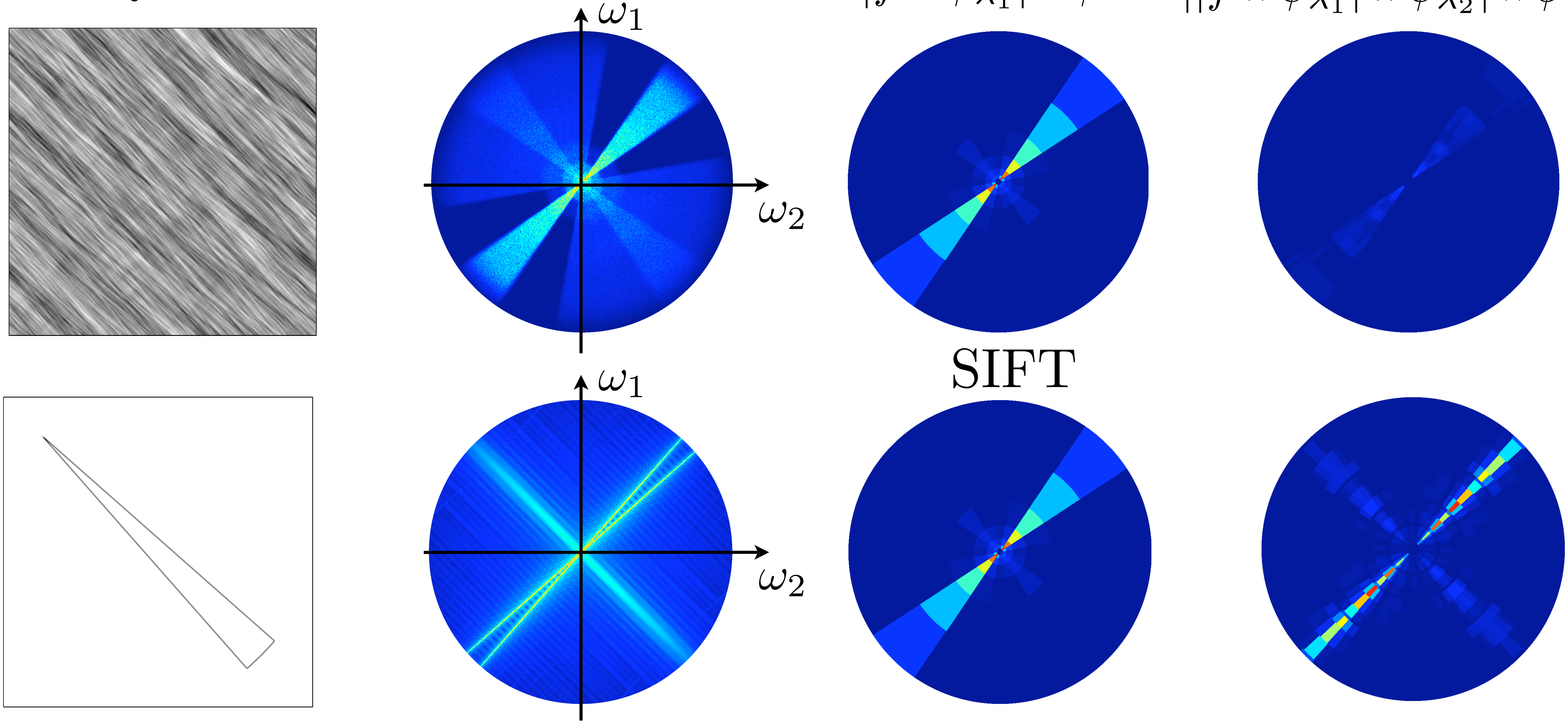
$X$: stationary process



# Image Wavelet Scattering

Images — Fourier — Wavelet Scattering

$f$ — $\hat{f}$ — $|f \star \psi_{\lambda_1}| \star \phi$ — $||f \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi$

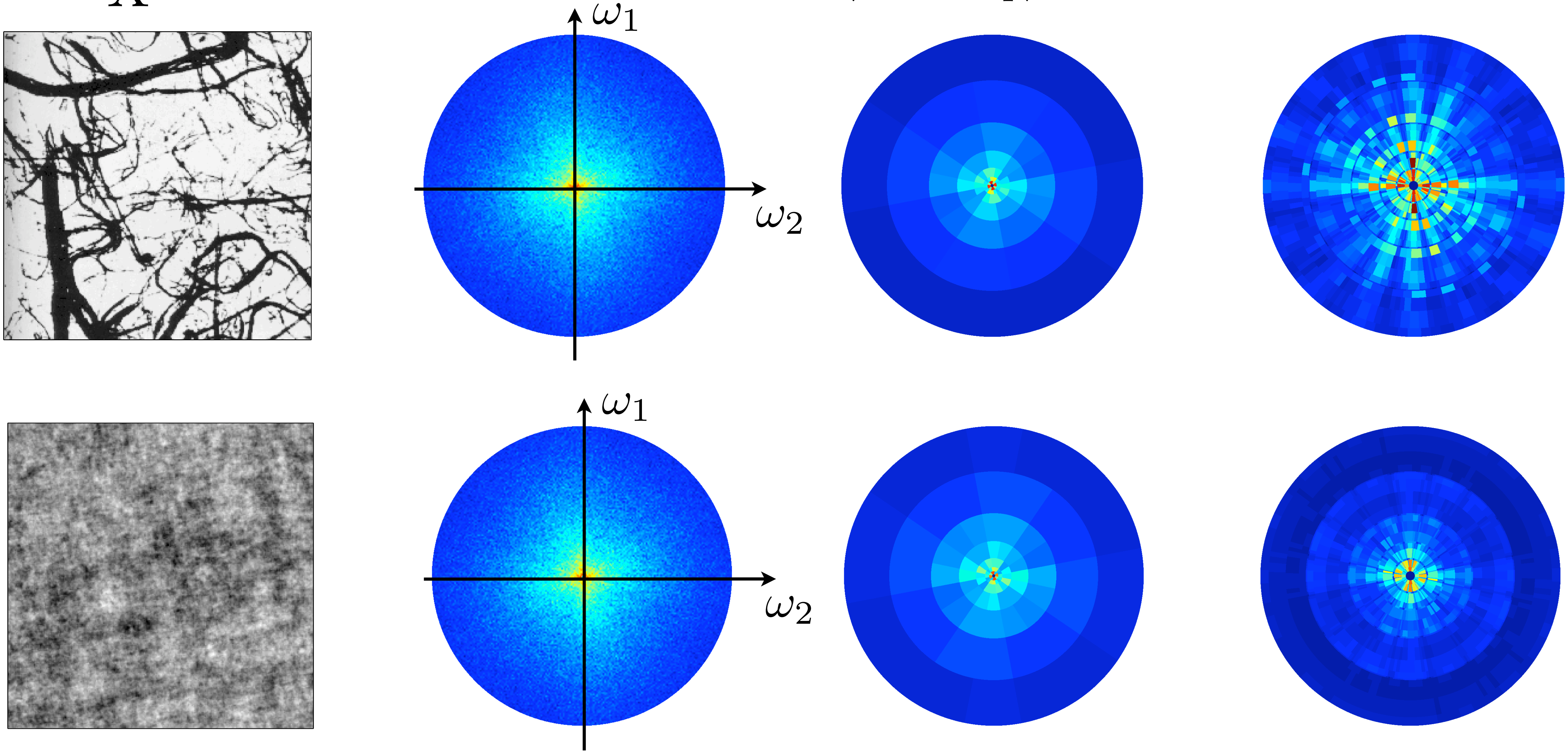



$$\text{window size} = \text{image size}$$

# Textures with Same Spectrum

$X$: stationary process

Textures X | Fourier Power Spectrum | Wavelet Scattering $|X \star \psi_{\lambda_1}| \star \phi$ $\quad ||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi$

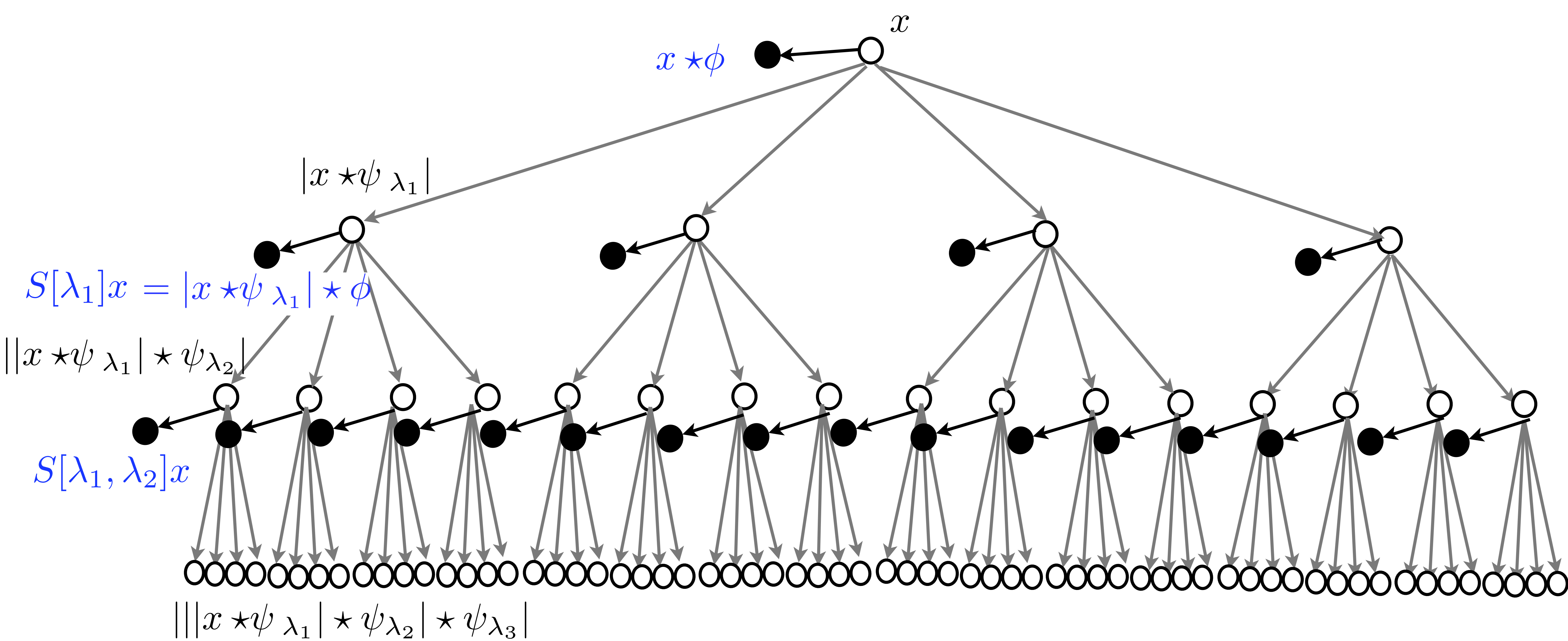



window size = image size

# Deep Convolution Network

*Y. LeCun et. al.*

- Iteration on $Ux = \{x \star \phi\,,\ |x \star \psi_\lambda|\}_\lambda$ , contracting.




- Output at **all layers**: $\{S[p]x\}_{p \in \mathcal{P}}$ .

# Scattering Properties

For any path $p = (\lambda_1, \lambda_2, ..., \lambda_m)$ of order $m$

$$S[p]x(t) = |\,|x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \ldots | \star \psi_{\lambda_m}| \star \phi(t)$$

$$\|Sx\|^2 = \sum_{p \in \mathcal{P}} \|S[p]x\|^2$$

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

*contracting* $\|Sx - Sy\| \leq \|x - y\|$

*preserves energy* $\|Sx\|^2 = \|x\|^2$

*stable to deformations* $\|Sx - Sx_\tau\| \leq C \sup_t |\nabla\tau(t)|\, \|x\|$

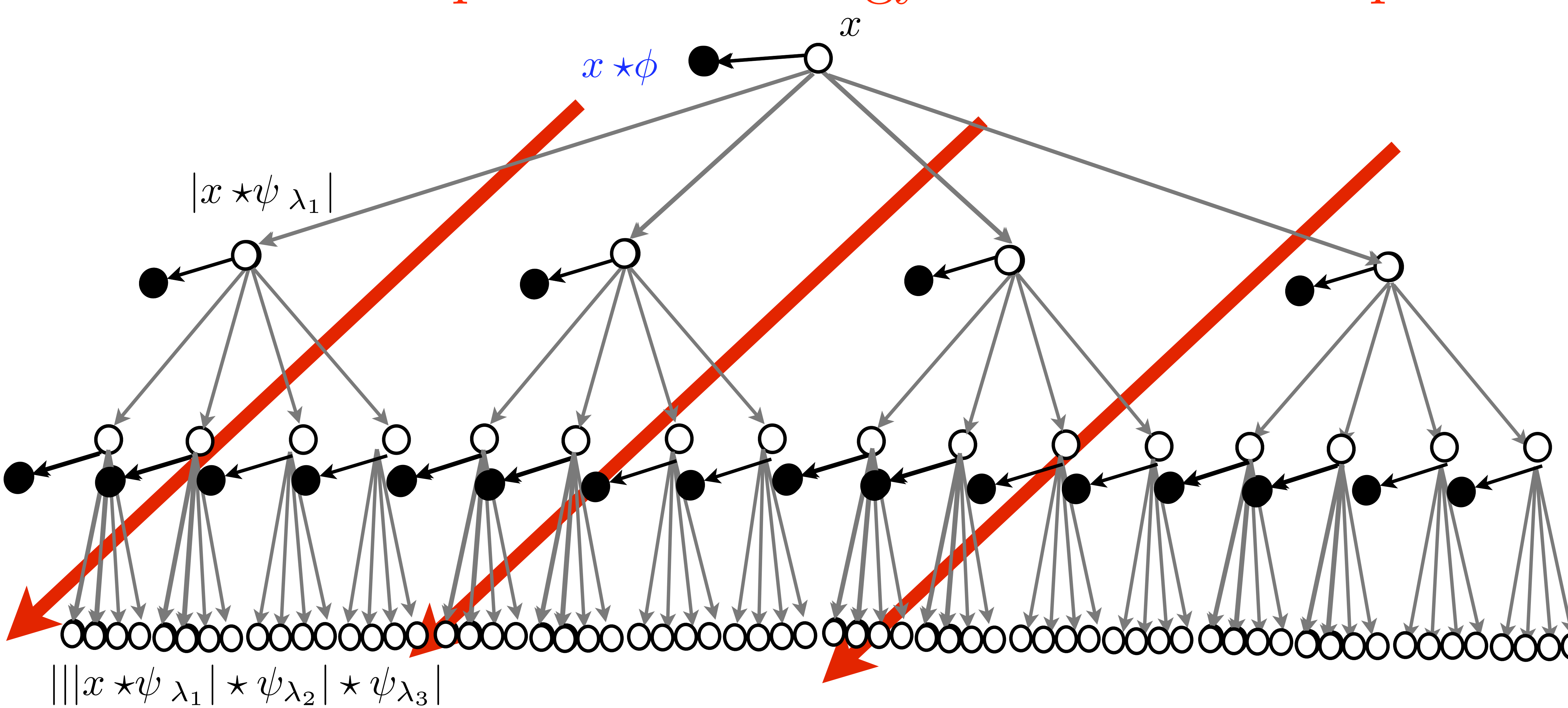
*when* $\phi$ *goes to* $1$, $Sx$ *converges to* $\overline{S}x(p) \in \mathbf{L^2}(\mathcal{P}_\infty)$

*which is translation invariant.*

# Energy Conservation

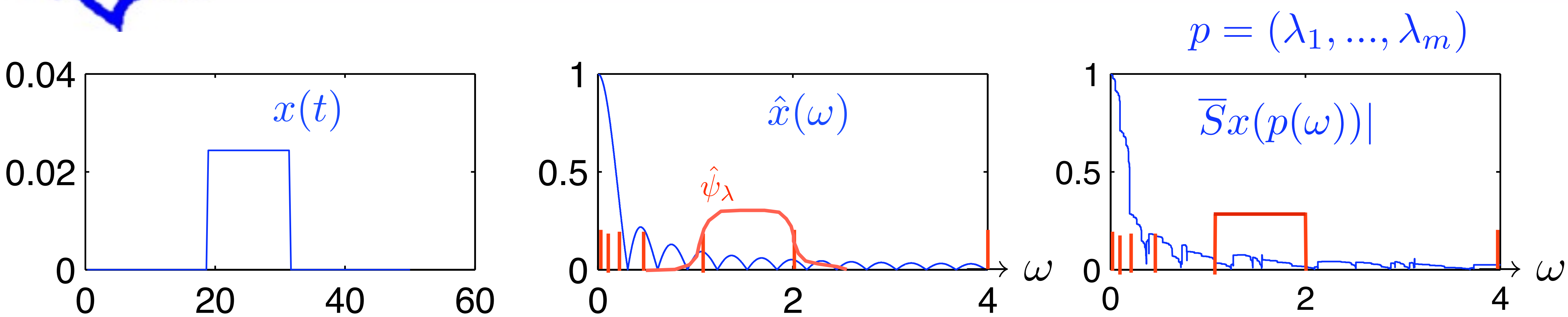
$$\|Ux\| = \|Wx\| = \|x\|$$

Proof: The modulus pushes the energy towards low frequencies

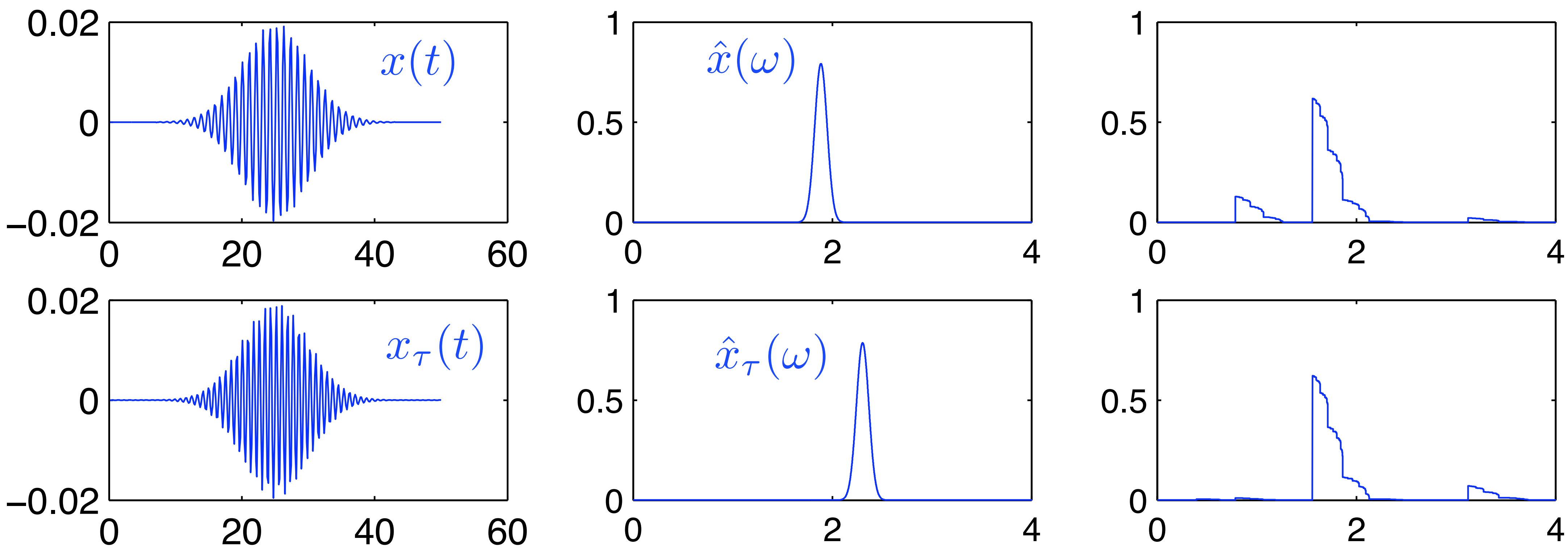



- Fast decay across layers of $\|U[p]x\| \Rightarrow \|Sx\| = \|x\|$
- Reduced number of paths with non-negligible output.
- Computational complexity: $O(N \log N)$.

# Frequency to Paths Mapping




$$x_\tau(t) = x(t - \tau(t)) \quad \text{with} \quad \tau(t) = \epsilon\, t\ .$$



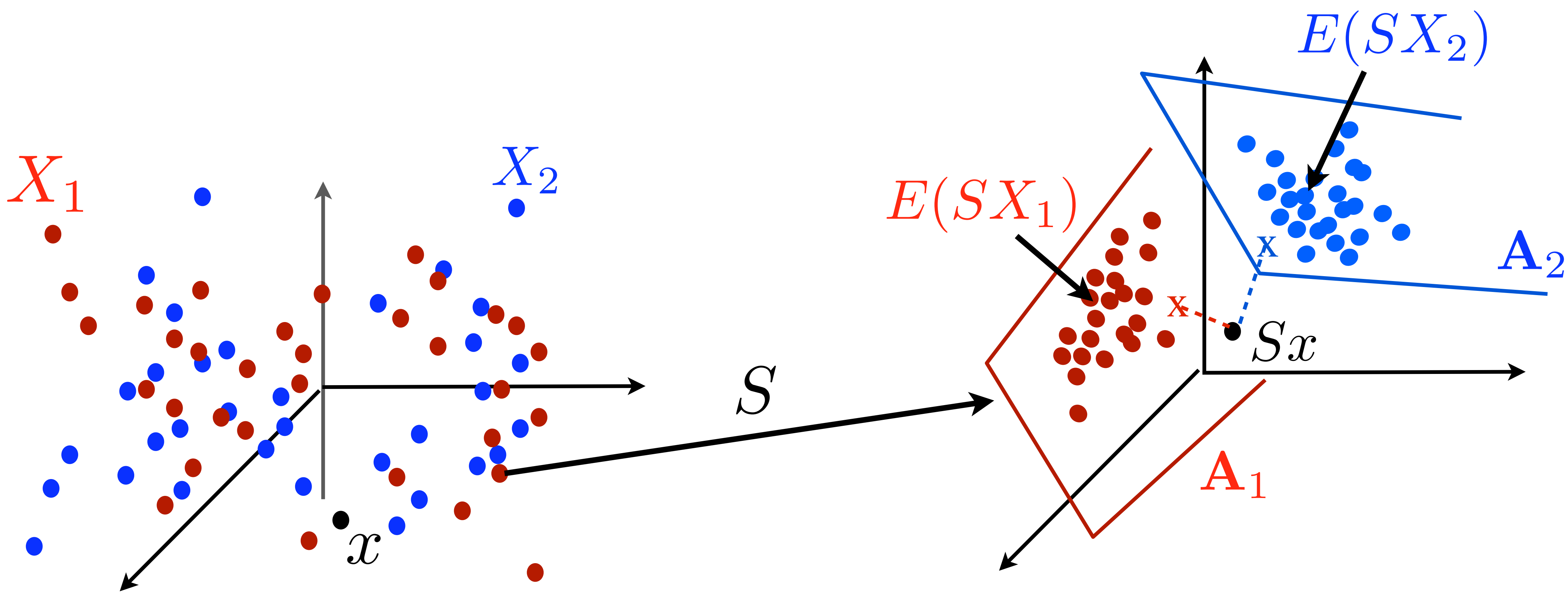

$$\frac{\|\,|\hat{x}| - |\,\hat{x}_\tau|\,\|}{\|x\|\,\|\tau'\|_\infty} = 13 \qquad \frac{\|\overline{S}x - \overline{S}x_\tau\|_{\overline{\mathcal{P}}_\infty}}{\|x\|\,\|\tau'\|_\infty} = 1.4$$

# Affine Space Classification

*Joan Bruna*

- Each class $X_k$ is represented by a scattering centroid $E(SX_k)$ and a space $\mathbf{V}_k$ of principal variance directions (PCA).

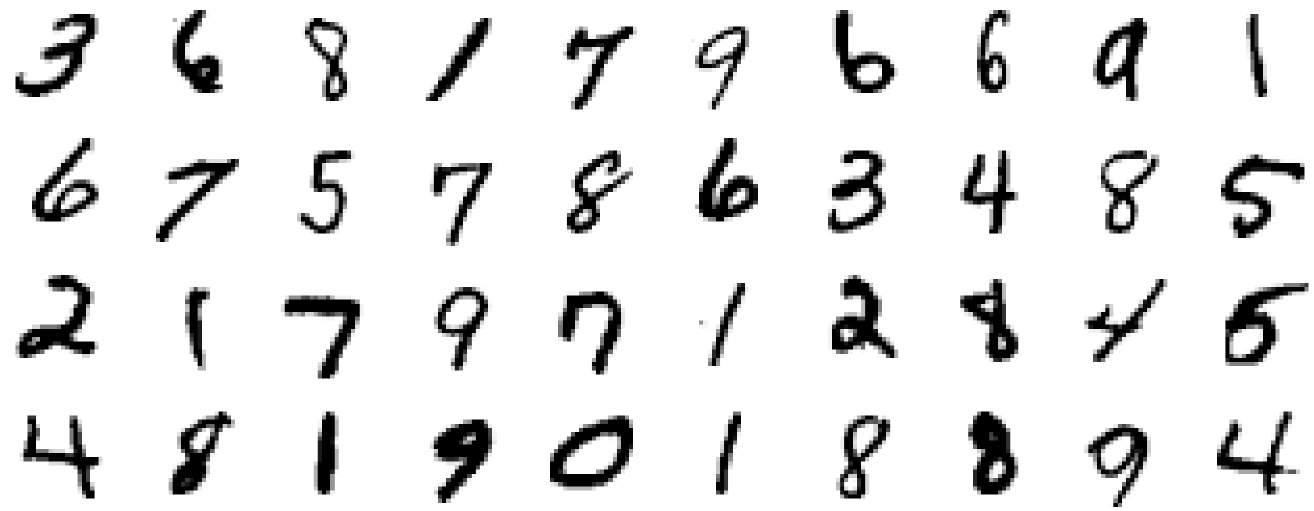
  Affine space model $\mathbf{A}_k = E(SX_k) + \mathbf{V}_k$.



# Affine Space Learning

- **Estimation** of affine approximation spaces with PCA
  - Estimation of the mean $E(SX_k)$ and the covariance $\Sigma_k$ from transformed labeled examples $Sx_n$ in each class
  - The best approximation space $\mathbf{V}_k$ of dimension *d* is generated by the *d* eigenvectors of $\Sigma_k$ of largest eigenvalues. It carries the principal deformation directions of each class.
  - The dimension *d* is optimized by cross-validation.

# Digit Classification: MNIST

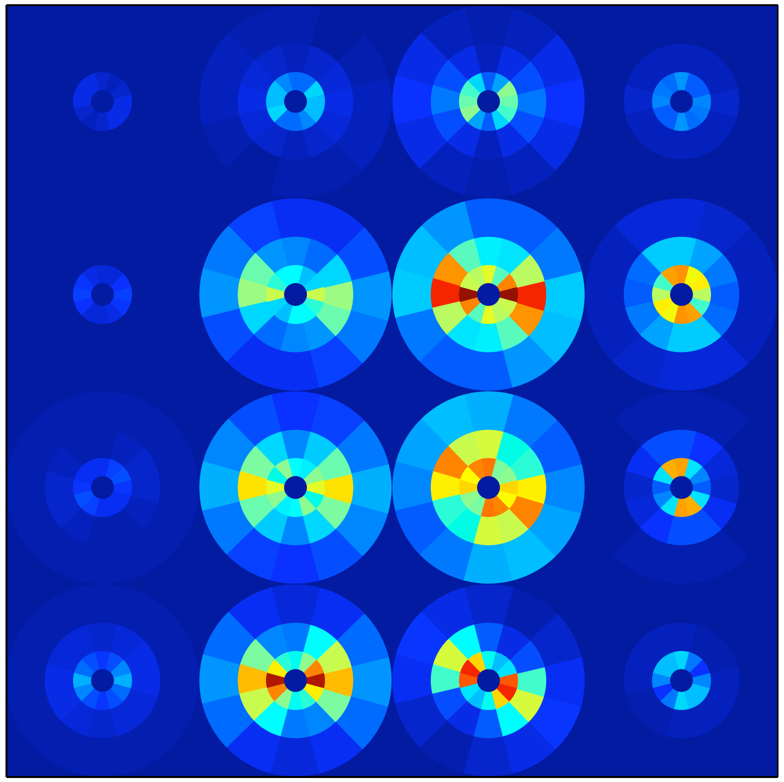


# Digit Classification: MNIST

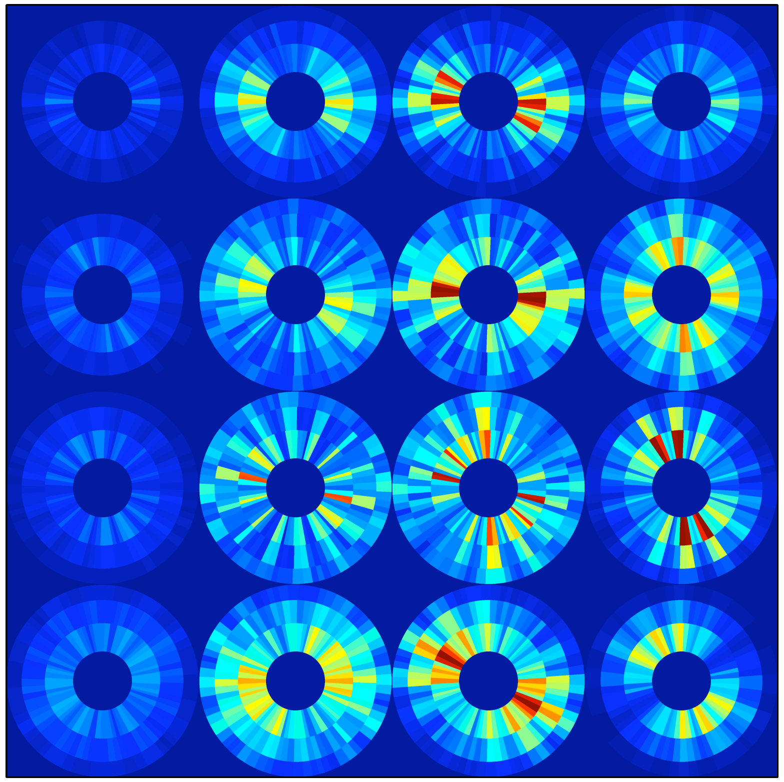
Wavelet Scattering

$x$

$|x \star \psi_{\lambda_1}| \star \phi(2^J n)$

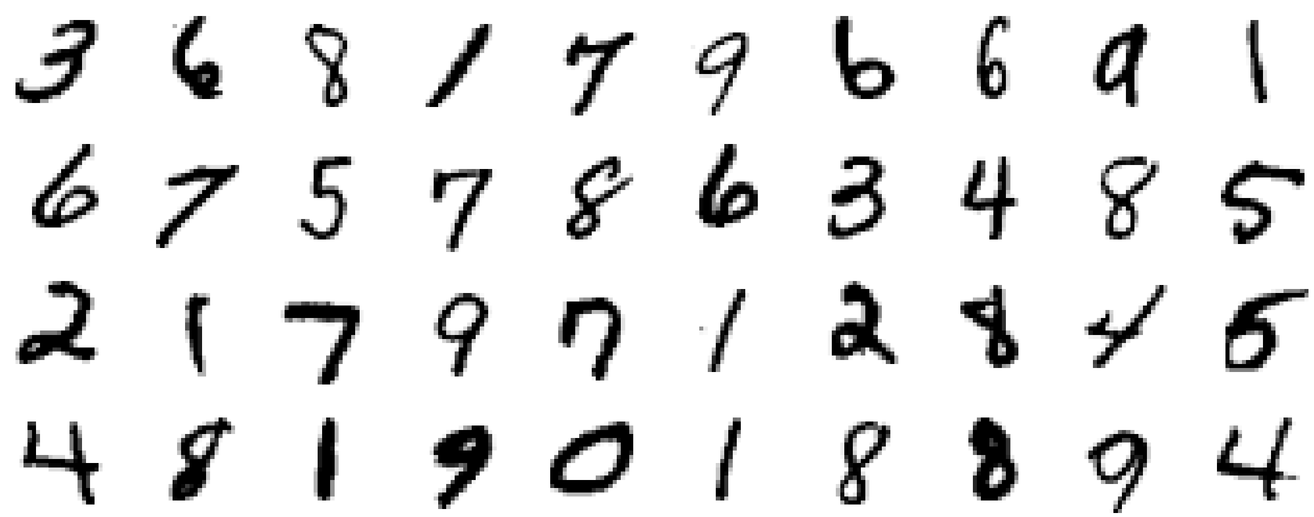


$||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi(2^J n)$



$2^J = 8$ : window size
cross-validated

# Digit Classification: MNIST



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.

# Overview

- **Part 1:** *Invariance and deformation stability*

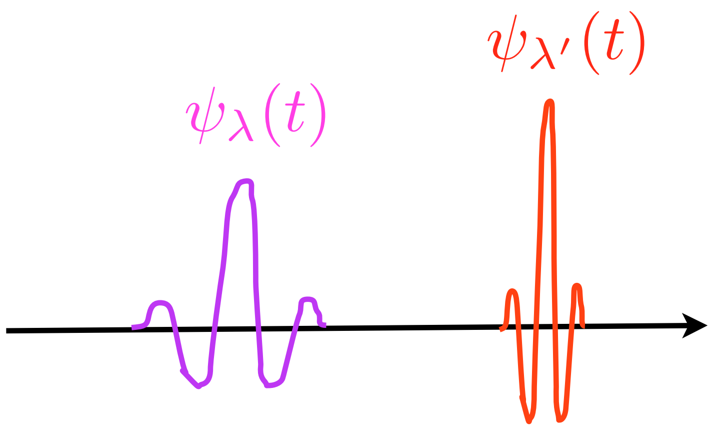
- Fourier failure
- Wavelet stability
- Scattering transform invariants and deep convolution networks
- Mathematical properties of deep networks
- Classification of images

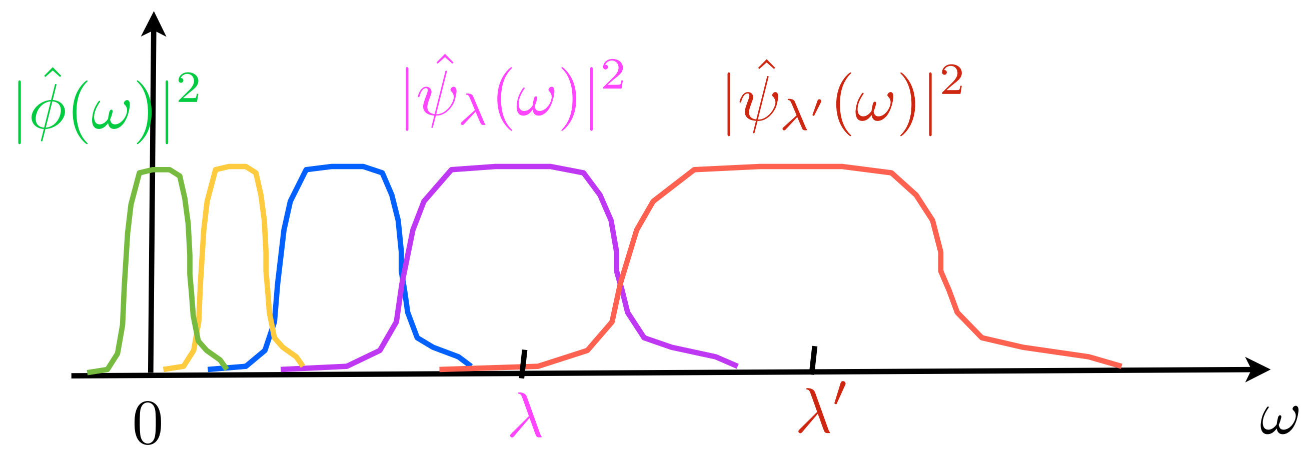
- **Part 2:** *Inverse, Textures and Multiple Invariants*

- Inverse scattering by phase retrieval and sparsity
- Scattering models of stationary processes
- Texture classification
- Invariants over multiple groups: transposition, rotation, scaling

# Wavelet Transform

- Dilated wavelets: $\psi_\lambda(t) = 2^{-jQ}\, \psi(2^{-jQ} t)$ with $\lambda = 2^{-jQ}$ .




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

- Wavelet transform: $Wx(t) = \left\{ x \star \phi\,(t)\ ,\ x \star \psi_\lambda(t) \right\}_\lambda$

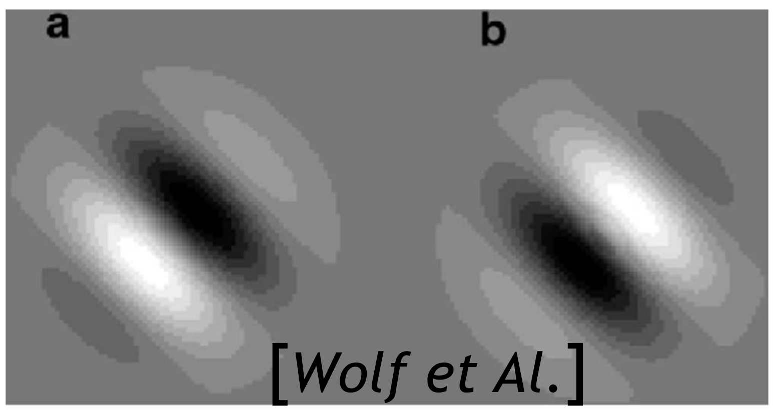
- If $|\phi|^2 + \sum_\lambda |\hat\psi_\lambda|^2 = 1$ then $W$ is unitary.

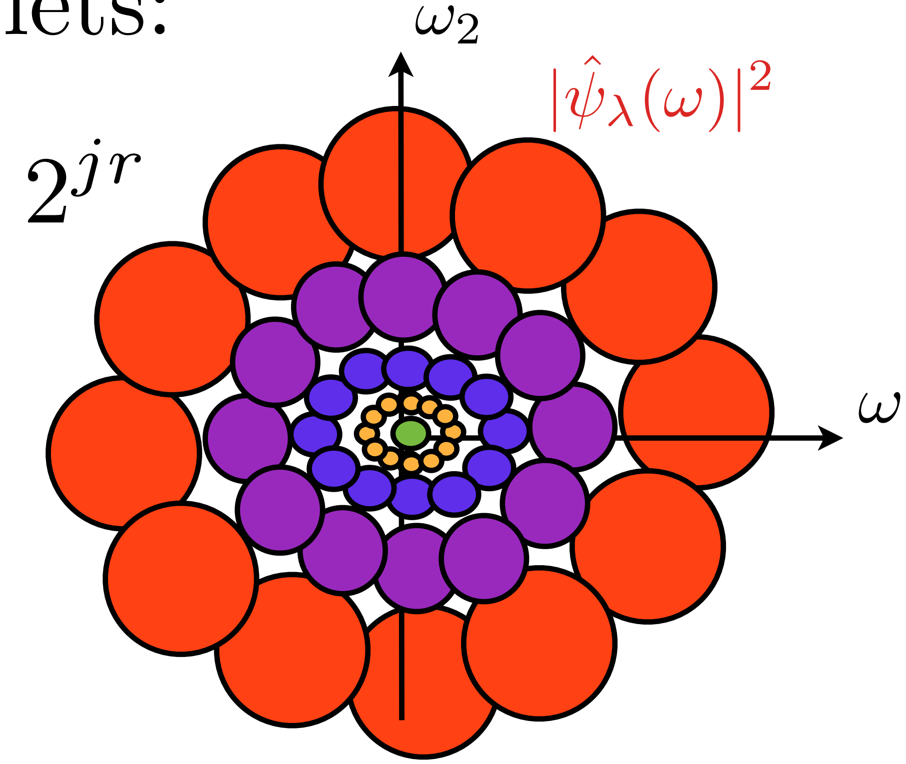
$$\|Wx\|^2 = \|x \star \phi\,\|^2 + \sum_\lambda \|x \star \psi_\lambda\|^2 = \|x\|^2 \ .$$

# Wavelet Transform

- For images, dilated and rotated wavelets:

$$\psi_\lambda(t) = 2^j \, \psi(2^j r t) \quad \text{with} \quad \lambda = 2^j r$$



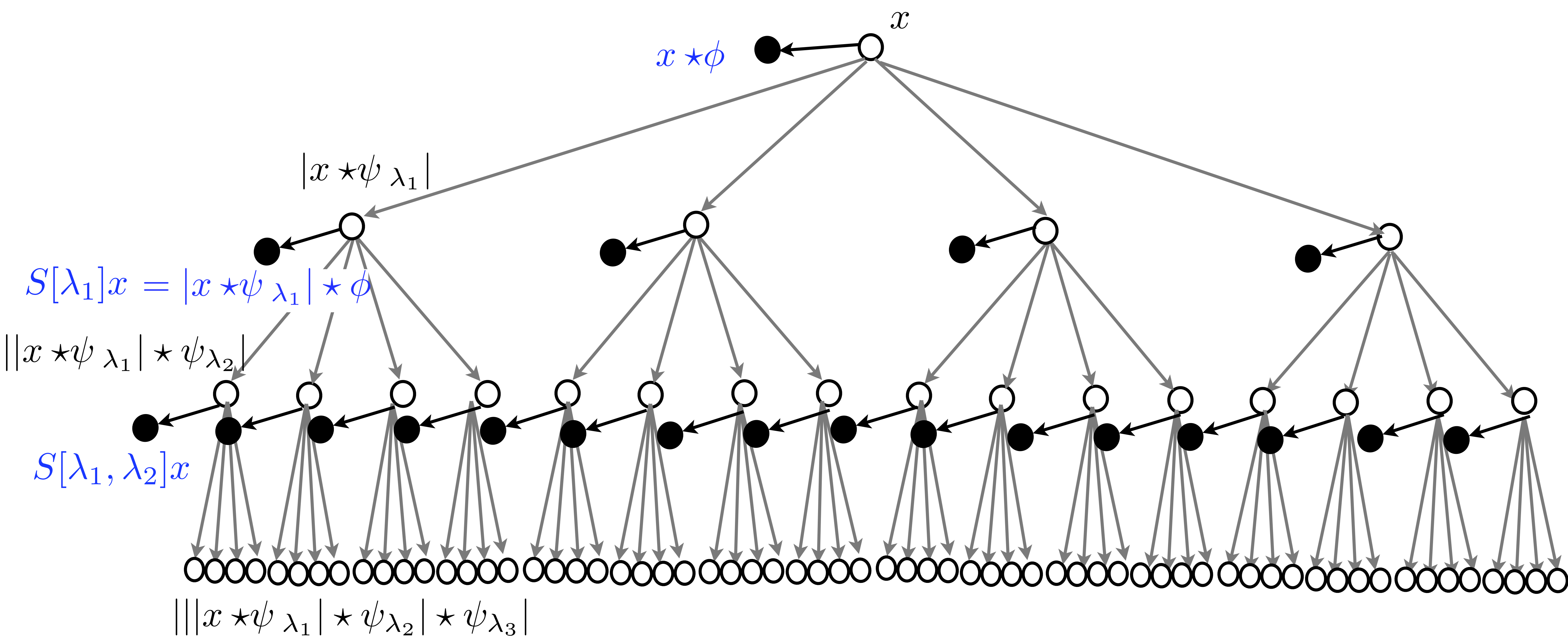

- Wavelet transform: $Wx(t) = \left\{ x \star \phi\,(t)\ ,\ x \star \psi_\lambda(t) \right\}_\lambda$

- If $|\phi|^2 + \sum_\lambda |\hat{\psi}_\lambda|^2 = 1$ then $W$ is unitary.

# Deep Convolution Network

*Y. LeCun et. al.*

- Iteration on $Ux = \{x \star \phi \,, \, |x \star \psi_{\lambda}|\}_{\lambda}$ , contracting.




- Output at **all layers**: $\{S[p]x\}_{p \in \mathcal{P}}$ .

  MFSC and SIFT are 1st layer outputs: $S[\lambda_1]x$

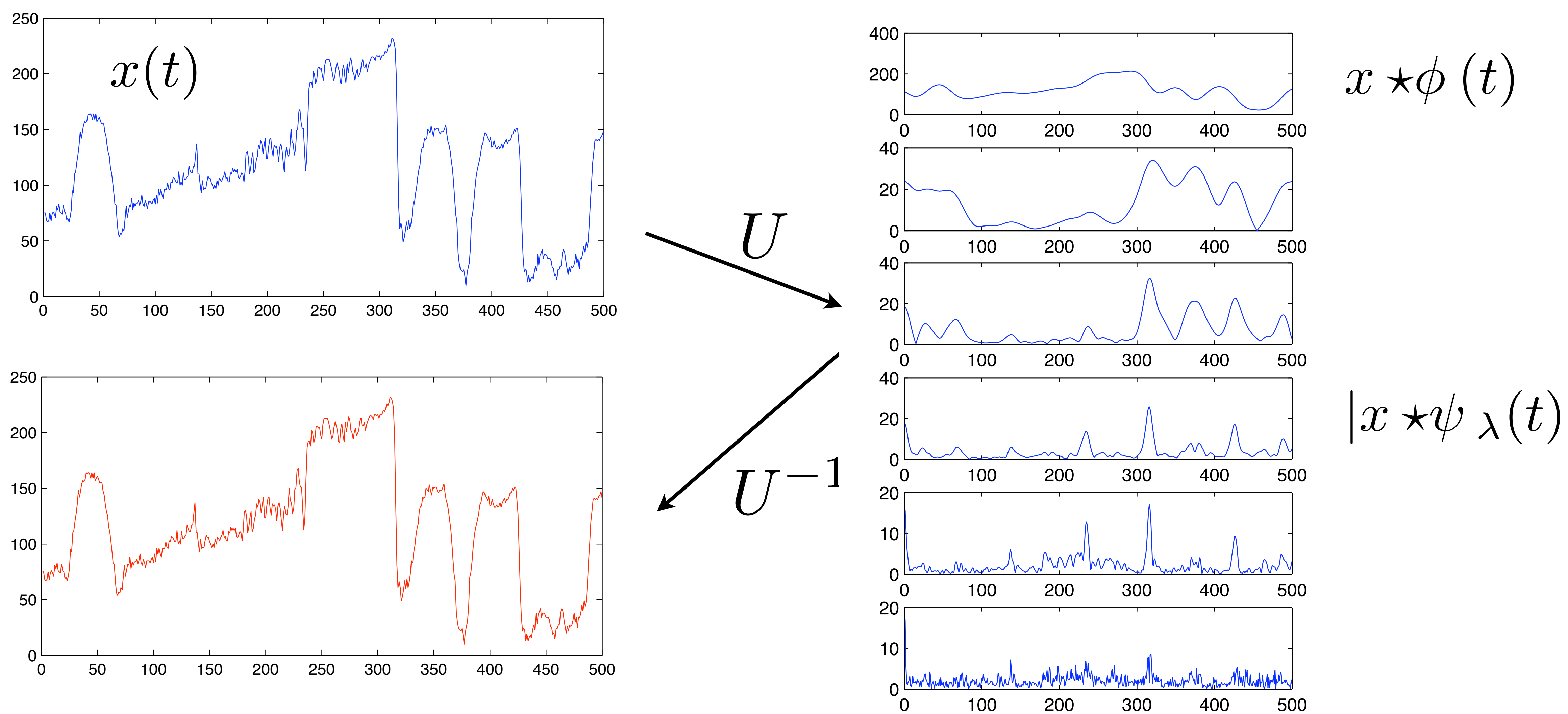
# Reconstruction, Phase Retrieval

*Irène Waldspurger*

**Theorem** For appropriate wavelets

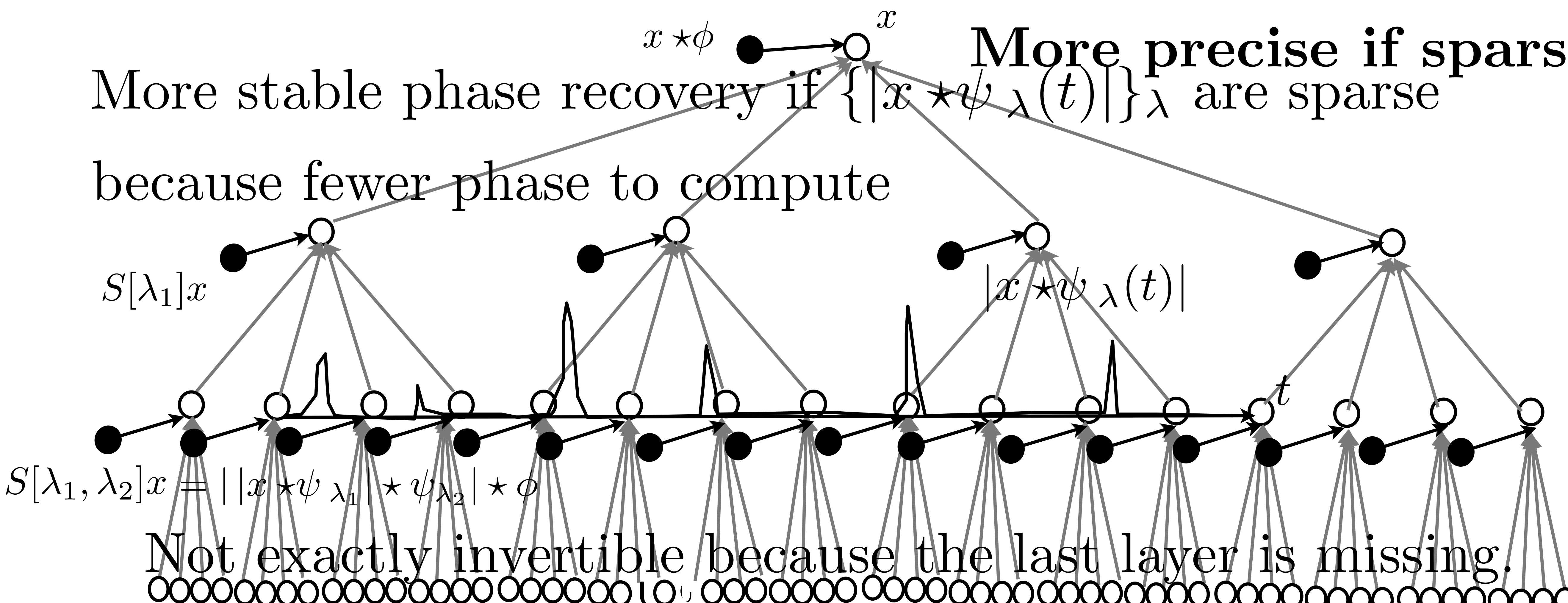
$$Ux = \left\{ x \star \phi\,(t)\,,\ |x \star \psi_{\lambda}(t)| \right\}_{\lambda}$$

is invertible and the inverse is continuous.



# Scattering Inversion: Sparsity

Inverse scattering: Progressive inversions of $U$




**More precise if sparse**

More stable phase recovery if $\{|x \star \psi_\lambda(t)|\}_\lambda$ are sparse

because fewer phase to compute

Not exactly invertible because the last layer is missing.

Smaller information loss if sparse: sparse deconvolution.

- Scattering invariants discriminate signals that are sparse

# Audio Reconstruction

*Joakim Anden*

Original audio signal $x$

Reconstruction from $Sx$ for a window of 3 s with $N$ samples

$Q = 8$

From order 1 $S[\lambda_1]x$ : $Q \log N$ coefficients

From order 2 $S[\lambda_1, \lambda_2]x$ : $(Q \log N)^2/2$ coefficients

# Sparsity for Learning

- Need a sparse analysis representation:

$$\Big\{ \langle x(t), \psi_\lambda(t-u) \rangle = x \star \psi_\lambda(u) \Big\}_{\lambda,u}$$

But we do not know how to learn them...

- We know how to learn sparse analysis representations:

$$x \approx \sum_\gamma \alpha_\gamma \, \psi_\gamma \quad \text{(unstable)}$$

by finding $\mathcal{D} = \{\psi_\gamma\}_\gamma$ which minimizes:

$$\|x - \sum_\gamma \alpha_\gamma \, \psi_\gamma\| + \mu \sum_\gamma |\alpha_\gamma|$$

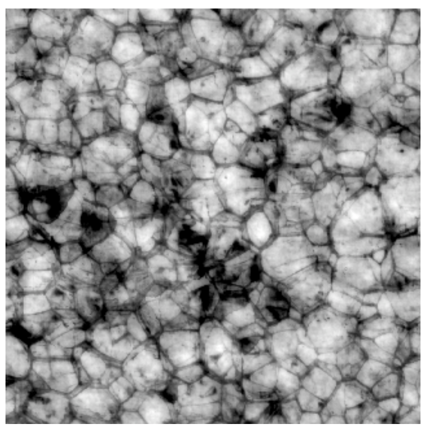
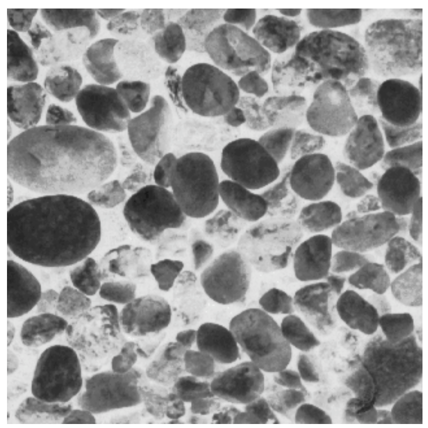
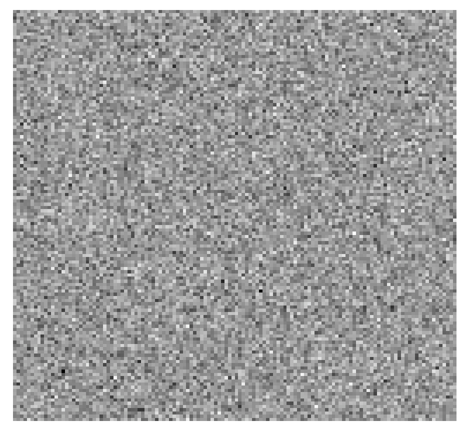
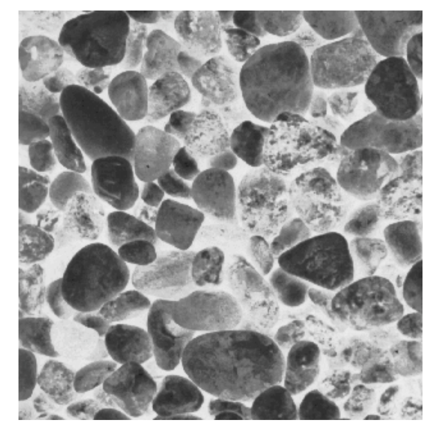
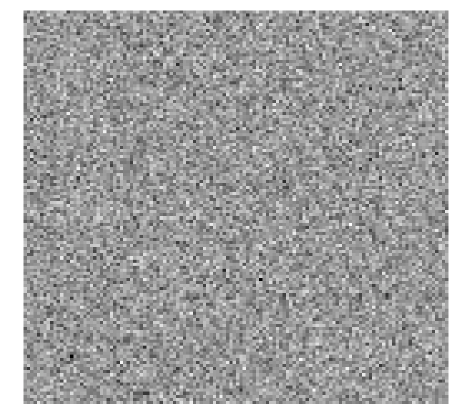
$\Rightarrow$ learn by synthesis and classify with analysis operators:

$$\{\langle x, \psi_\gamma \rangle\}_\gamma \;\; : \text{stable} \;\; \text{(autoencoders)}$$

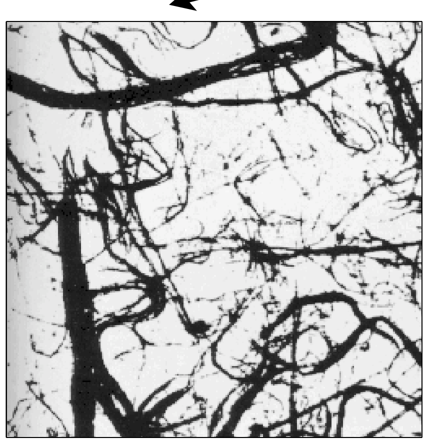
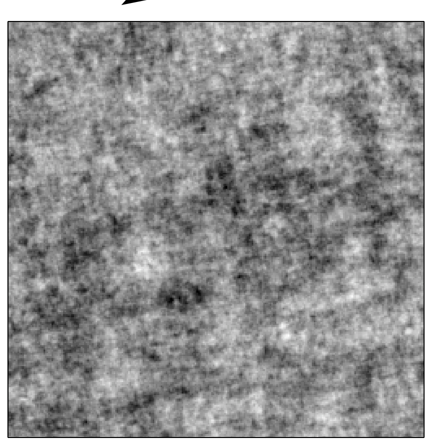
# Texture Discrimination

- Textures define high-dimensional image classes.
  - Realizations of stationary processes $X$ but typically not Gaussian, not Markovian and not characterized by second order moments.

same power spectrum

same power spectrum

# Scattering Stationary Processes

- If $X(t)$ is stationary then

$$U[p]X = |\cdots|X \star \psi_{\lambda_1}| \star \cdots | \star \psi_{\lambda_m}| \quad \text{is stationary}$$

- Expected scattering: $\overline{S}X(p) = E(U[p]X)$

depends on normalized moments of order $2^m$ of $X$.

- A windowed scattering

$$S[p]X(t) = U[p]X \star \phi(t)$$

is an unbiased estimator of $\overline{S}X(p) = E(U[p]x)$.

# Maximum Entropy Distribution

*Joan Bruna*

- Given $\overline{S}X(p) = E(U[p]X)$ for $p \in \mathcal{P}$

the maximum entropy distribution is (Boltzman theorem):

$$p(x) = \frac{1}{Z} \exp\Big( \sum_{p \in \mathcal{P}} \alpha_p \, U[p] x \Big)$$

where $\alpha_p$ are Lagrange multipliers and $Z$ is defined by

$$\int p(x)\, dx = 1 \ .$$

- Metropolis-Hasting algorithm samples the distribution, but computationaly very expansive.

- Faster iterative algorithm with sparsity condition on $\mathbf{l^0}$ norm.

# Synthesis from Second Order

J. McDermott textures

*Joan Bruna*
*Joakim Anden*

- Estimation of $X(x)$ from $\log^2 N$ second order coefficients:

$Q = 1$

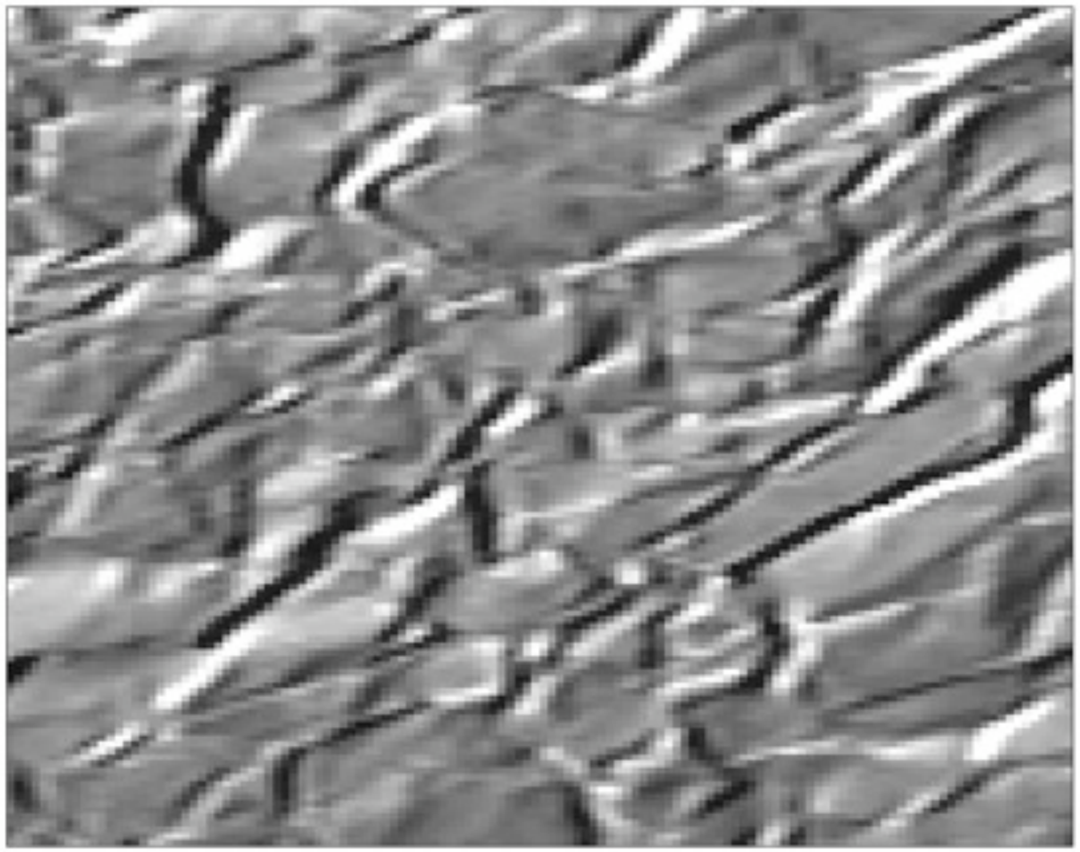
- Original jackhammer
- Synthesized

- Original water
- Synthesized

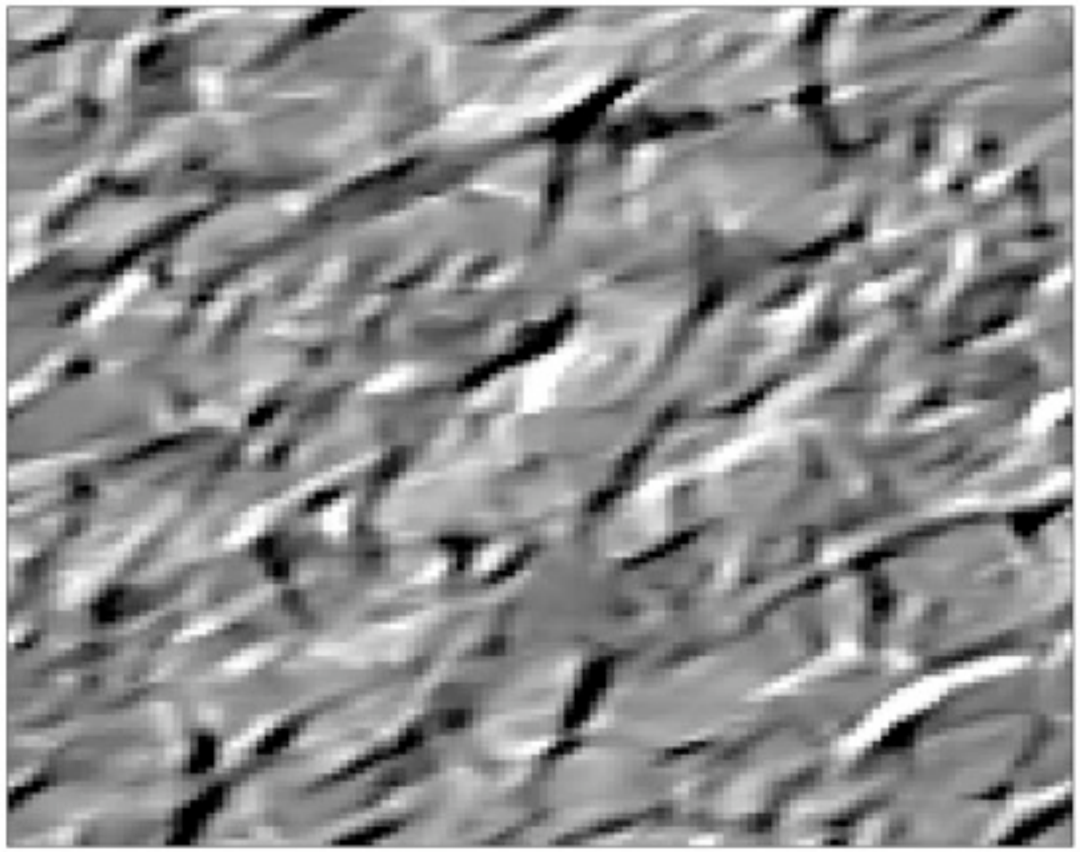
- Original applause
- Synthesized

# Image Reconstruction

Original



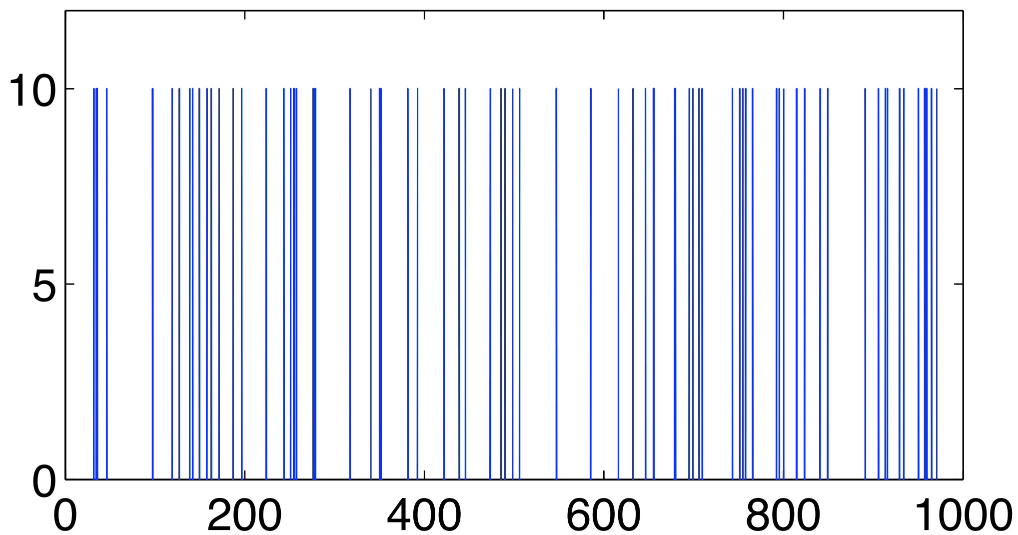
Reconstructed



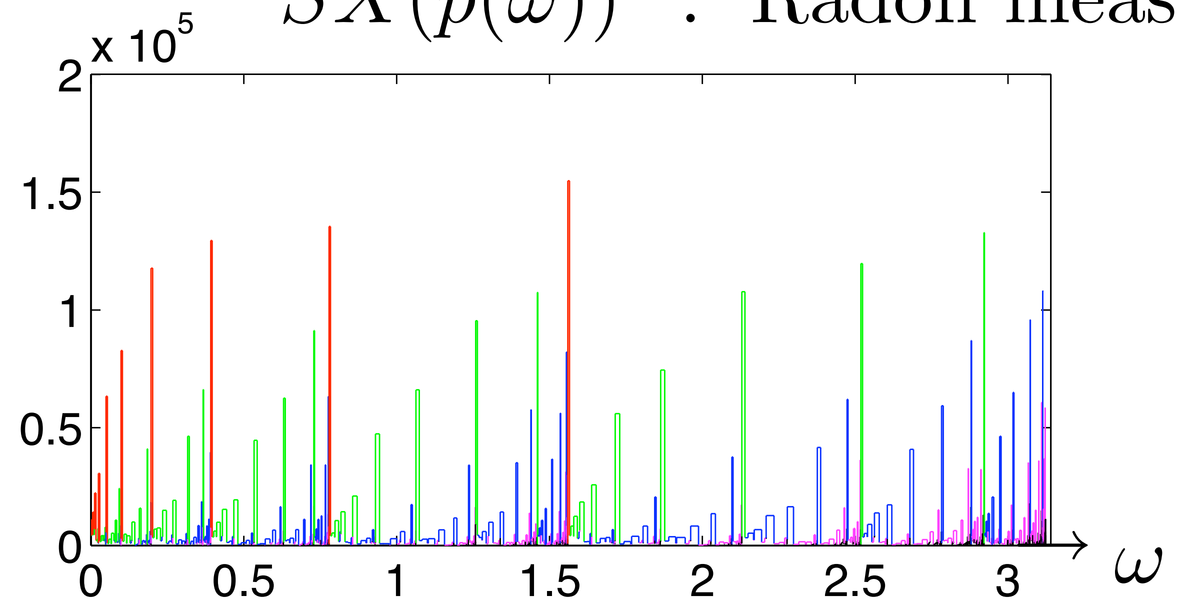
# Scattering White Noises

Constant Fourier power spectrum: $\hat{R}_X(\omega) = \sigma^2$.

Bernoulli $X(x)$



$\overline{S}X(p(\omega))^2$ : Radon measure

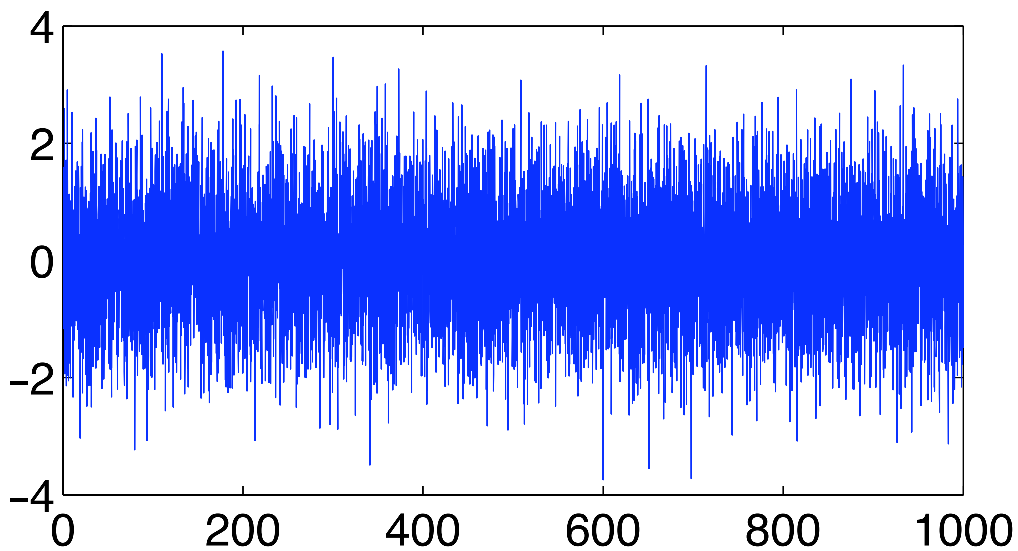


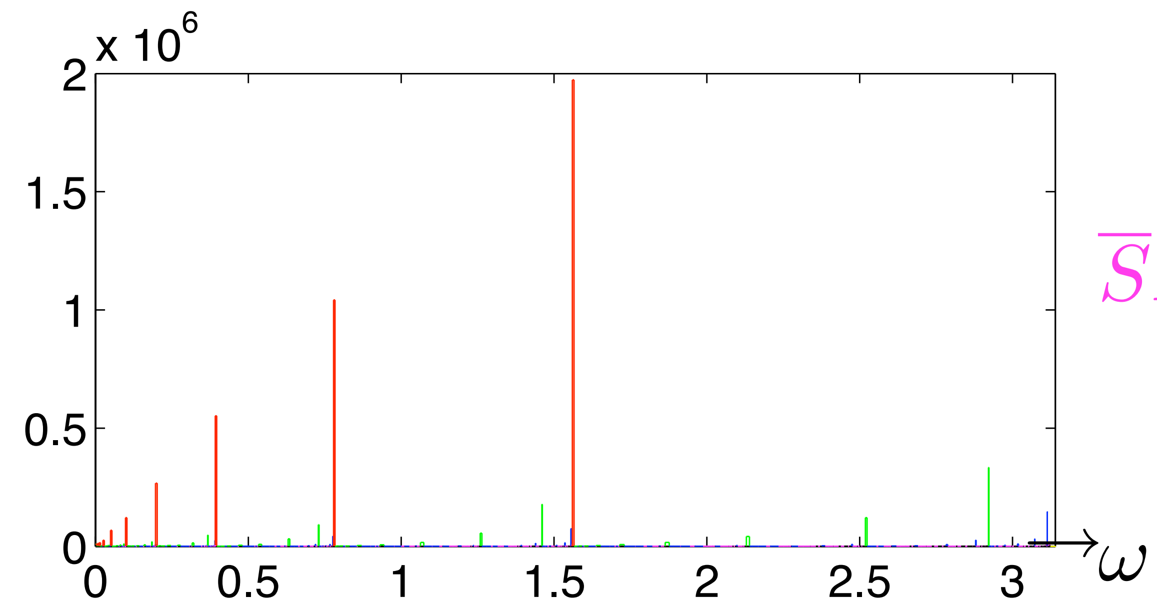

$\overline{S}X(\lambda_1)$

$\overline{S}X(\lambda_1, \lambda_2)$

$\overline{S}X(\lambda_1, \lambda_2, \lambda_3)$

$\overline{S}X(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$






Gaussian White

$$\int \hat{R}_X(\omega)\, |\hat{\psi}_{2^j}(\omega)|^2\, d\omega = \int_{2^j}^{2^{j+1}} \overline{S}X(p(\omega))^2\, d\omega \ .$$

# Gain Control and Normalization

- Invariant information is in transfer functions:

$$\frac{S[\lambda_1, ..., \lambda_{m-1}, \lambda_m] x(t)}{S[\lambda_1, ..., \lambda_m] x(t)} \quad : \text{ tuned gain control}$$

computed by cascading a normalized propagator

$$\overline{U} x = \left\{ x \star \phi \ , \ \frac{|x \star \psi_\lambda|}{x \star \phi} \right\} \quad : \quad \text{surround suppression}$$

# Multifractal Scattering

- Multifractal scaling:

$$\overline{S}X(\lambda_1) \sim \lambda_1^{-\gamma_1} \qquad \frac{\overline{S}X(\lambda_1,\lambda_2)}{\overline{S}X(\lambda_1)} \sim (\lambda_2\,\lambda_1^{-1})^{-\gamma_2}$$

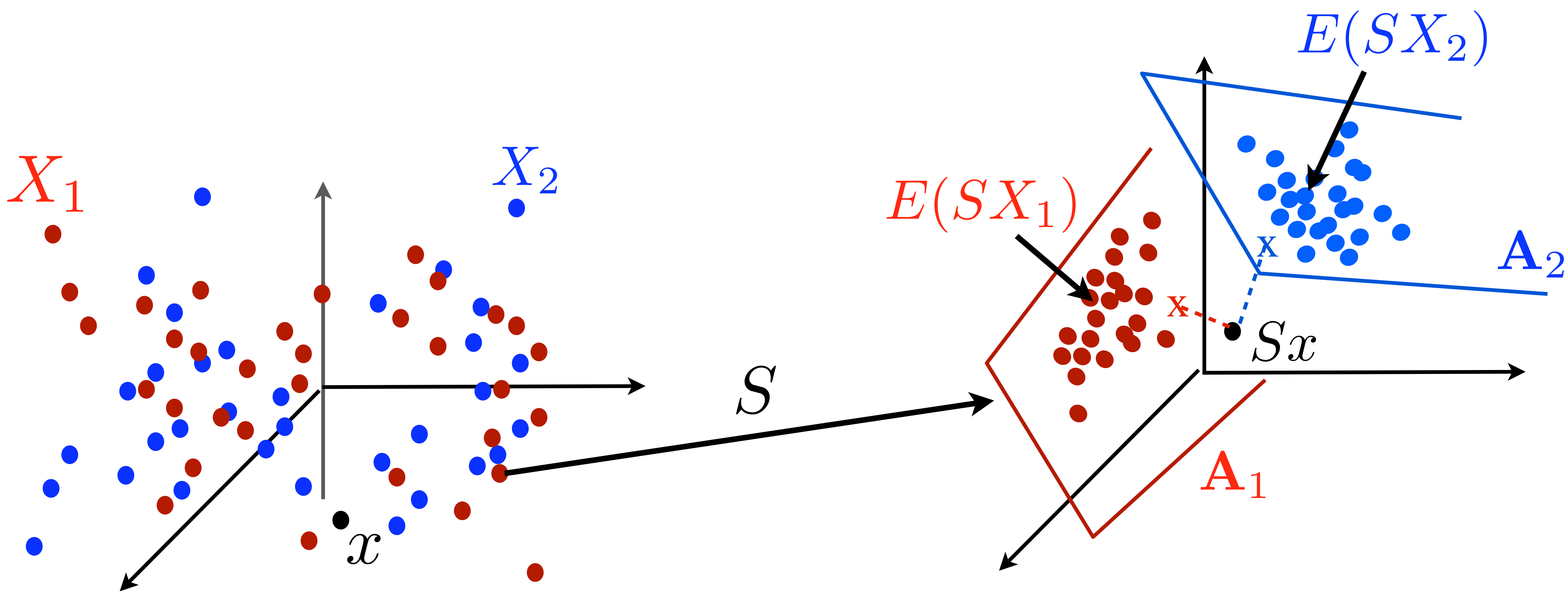
| **Process** | $\gamma_1$ | $\gamma_2$ |
|---|---|---|
| White Gaussian | $-1/2$ | $-1/2$ |
| Fractional Brownian Noise $B_H(t)$ | $H$ | $-1/2$ |
| Mandelbrot cascade | $\gamma_1$ | 0 |
| NASDAQ:AAPL | $2/3$ | $-0.15$ |
| Dirac measure | 0 | 0 |
| Poisson pp density $\alpha$ | $0$ if $\lambda < \alpha$<br>$-1/2$ if $\lambda \geq \alpha$ | $0$ if $\lambda_1+\lambda_2 < \alpha$<br>$-1/2$ if $\lambda_1+\lambda_2 \geq \alpha$ |

# Affine Space Classification

*Joan Bruna*

- Each class $X_k$ is represented by a scattering centroid $E(SX_k)$ and a space $\mathbf{V}_k$ of principal variance directions (PCA).

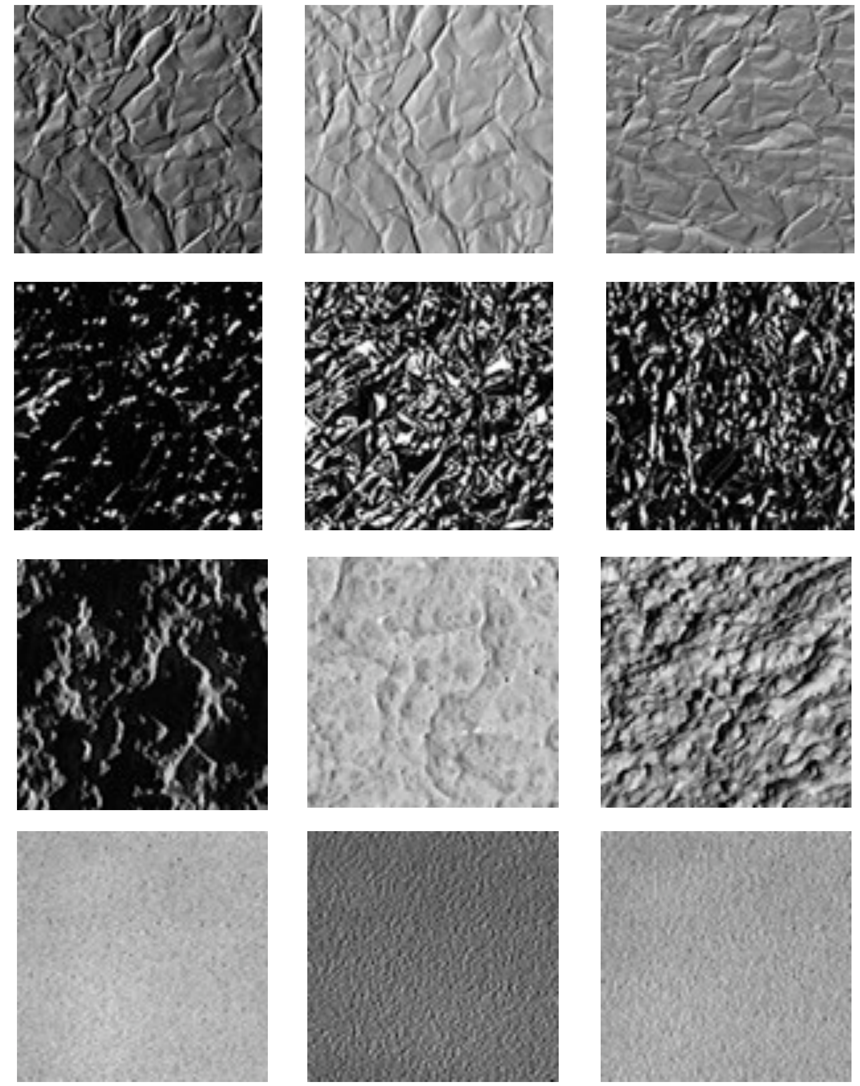
  Affine space model $\mathbf{A}_k = E(SX_k) + \mathbf{V}_k$.



# Classification of Textures

CUREt database

61 classes

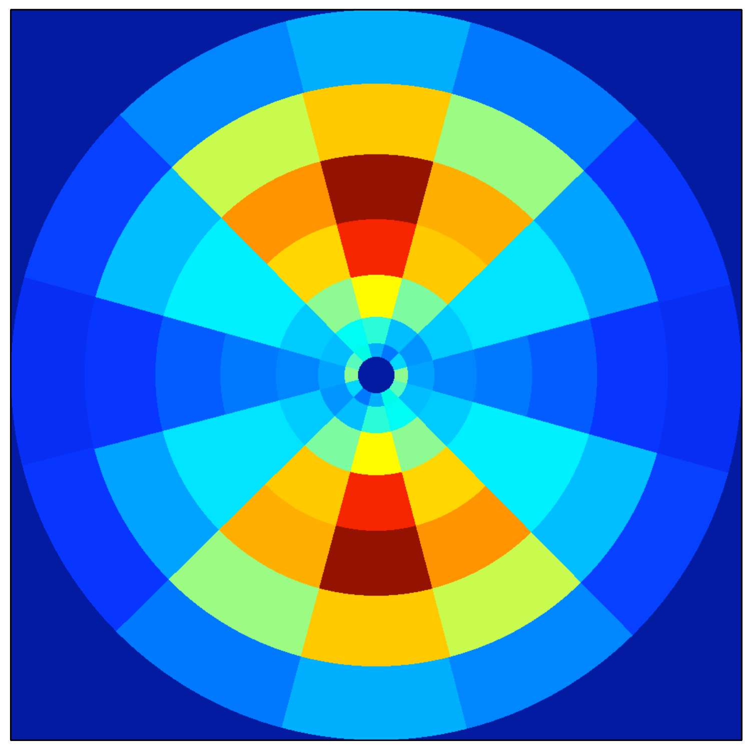


Rotations and illumination variations.

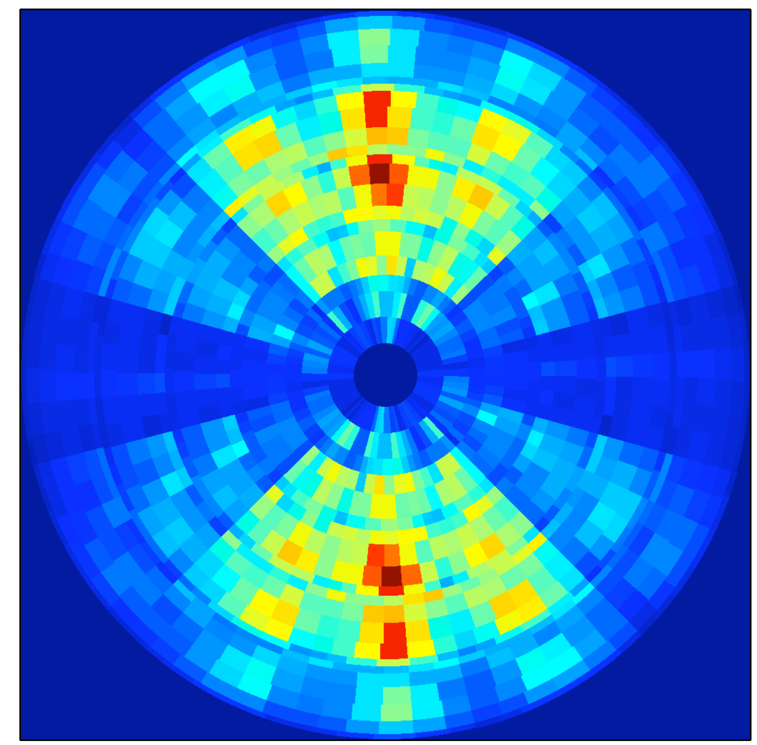
# Classification of Textures

$X$

$||X \star \psi_{\lambda_1}|| \star \phi$



$||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi$



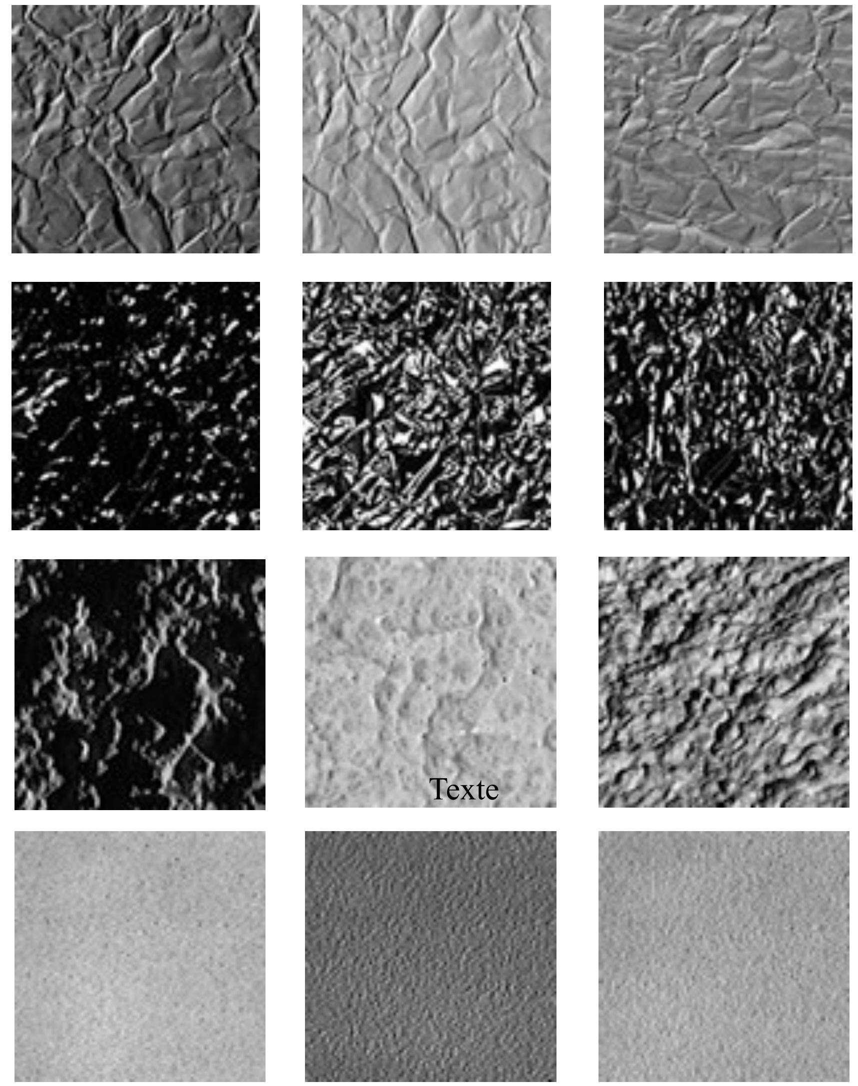
window size = image size

cross-validated

# Classification of Textures

CUREt database

61 classes




Rotations and illumination variations.

Classification Errors

| Training per class | Fourier Spectr. | Markov Field | Scattering |
|---|---|---|---|
| 46 | 2.15% | 2.46% | **0.2** % |

Varma & Zisserman

# Audio Genre Classification

*Joakim Anden*

GTZAN: music genre classification (jazz, rock, classic,...)

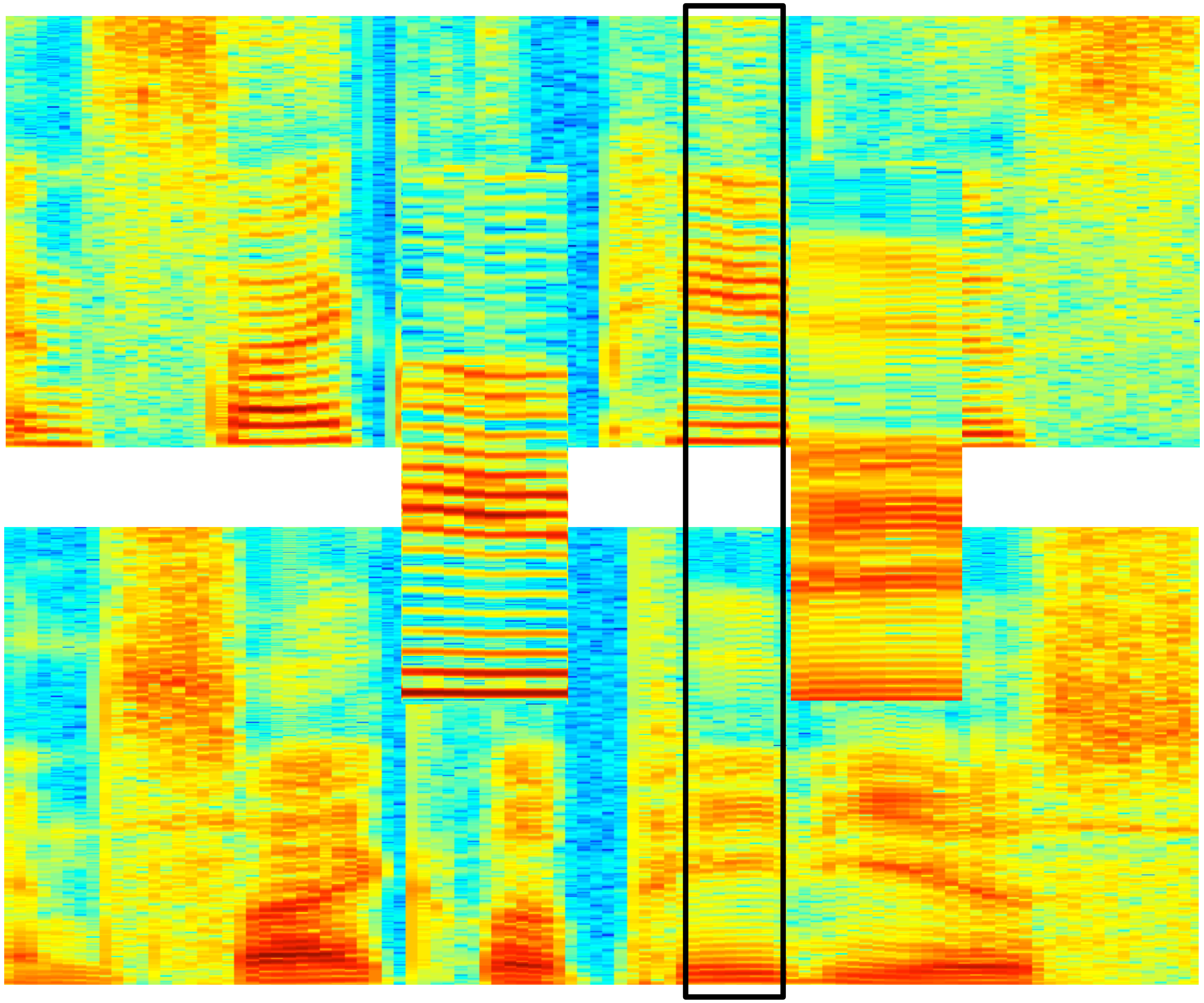
10 classes and 30 seconds tracks.

Classification errors

| Feature Set | Error (%) |
|---|---|
| MFCC | 32 |
| Delta-MFCC | 23 |
| Scattering, m=1 | 28 |
| Scattering, m=2 | 16 |

# Same or Different ?

*Aren Jensen*



encyclopaedias

# Frequency Transposition Invariance

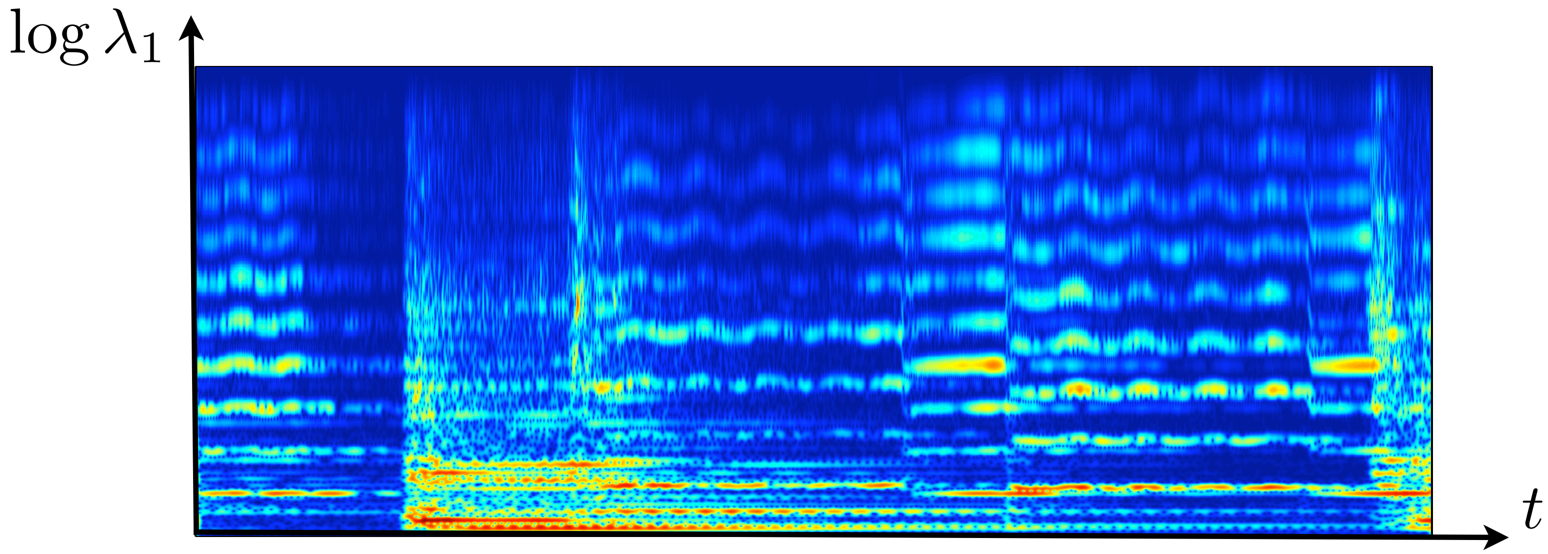
Same words by different people

Change of pich $\Rightarrow$ frequency scaling: $\omega \to \alpha\,\omega$

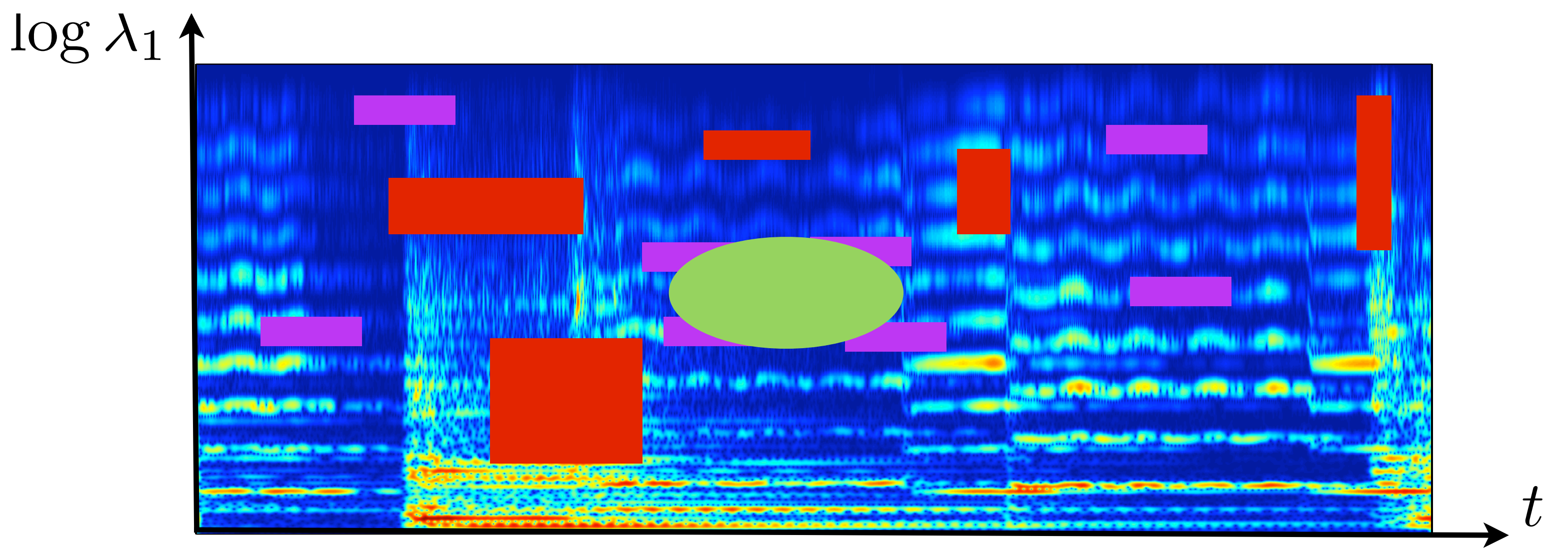
$\Rightarrow$ log frequency translation: $\log\omega \to \log\alpha + \log\omega$

$\Rightarrow$ translation invariance in $(t, \log\omega)$ with deformation stability

Wavelet modulus: $|x \star \psi_{\lambda_1}(t)|$



# Transposition Invariant Scattering




Separable wavelets in $t$ and $\log \lambda_1$

$$\Psi_{\lambda_2,\tilde{\lambda}_2}(t, \log \lambda_1) = \psi_{\lambda_2}(t)\ \tilde{\psi}_{\tilde{\lambda}_2}(\log \lambda_1)$$

Separable 2D wavelet transform of: $y(t, \log \lambda_1) = |x \star \psi_{\lambda_1}(t)|$

$$y \star \Psi_{\lambda_2,\tilde{\lambda}_2}(t, \log \lambda_1)$$

Invariance by wavelet amplitude averaging in $(t, \log \lambda_1)$:

$$|y \star \Psi_{\lambda_2,\tilde{\lambda}_2}| \star \Phi(t, \log \lambda_1)$$

Invariant scattering:

$$|\ |y \star \Psi_{\lambda_2,\tilde{\lambda}_2}| ... \star \Psi_{\lambda_m,\tilde{\lambda}_m}| \star \Phi(t, \log \lambda_1)$$

# Classification with Transp. Invariants

*Joakim Anden*

GTZAN: music genre classification (jazz, rock, classic,...)

10 classes and 30 seconds tracks.

Classification errors

| Feature Set | Error (%) |
|---|---|
| Scattering, m=2 | 16 |
| Scat.+ Transp. Inv., m=2 | 13 |

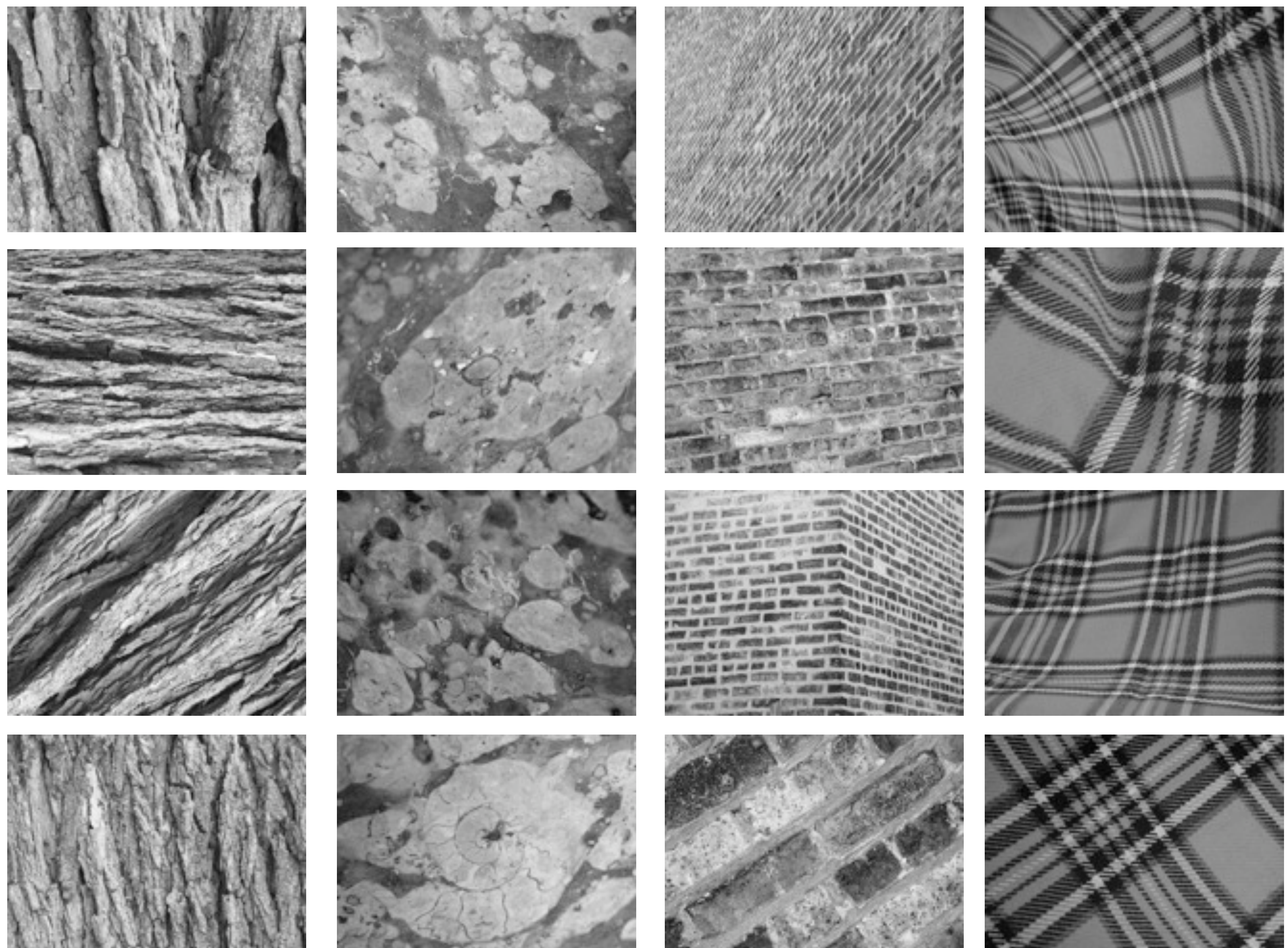
# Rotation and Affine Invariance

*Laurent Sifre*

- Scatterings along translation, rotation and affine groups:

$$x \longrightarrow \boxed{\text{Translat. Invar.}} \longrightarrow \boxed{\text{Rotation Invar.}} \longrightarrow \boxed{\text{Affine Invar.}} \longrightarrow Sx$$

UIUC database:



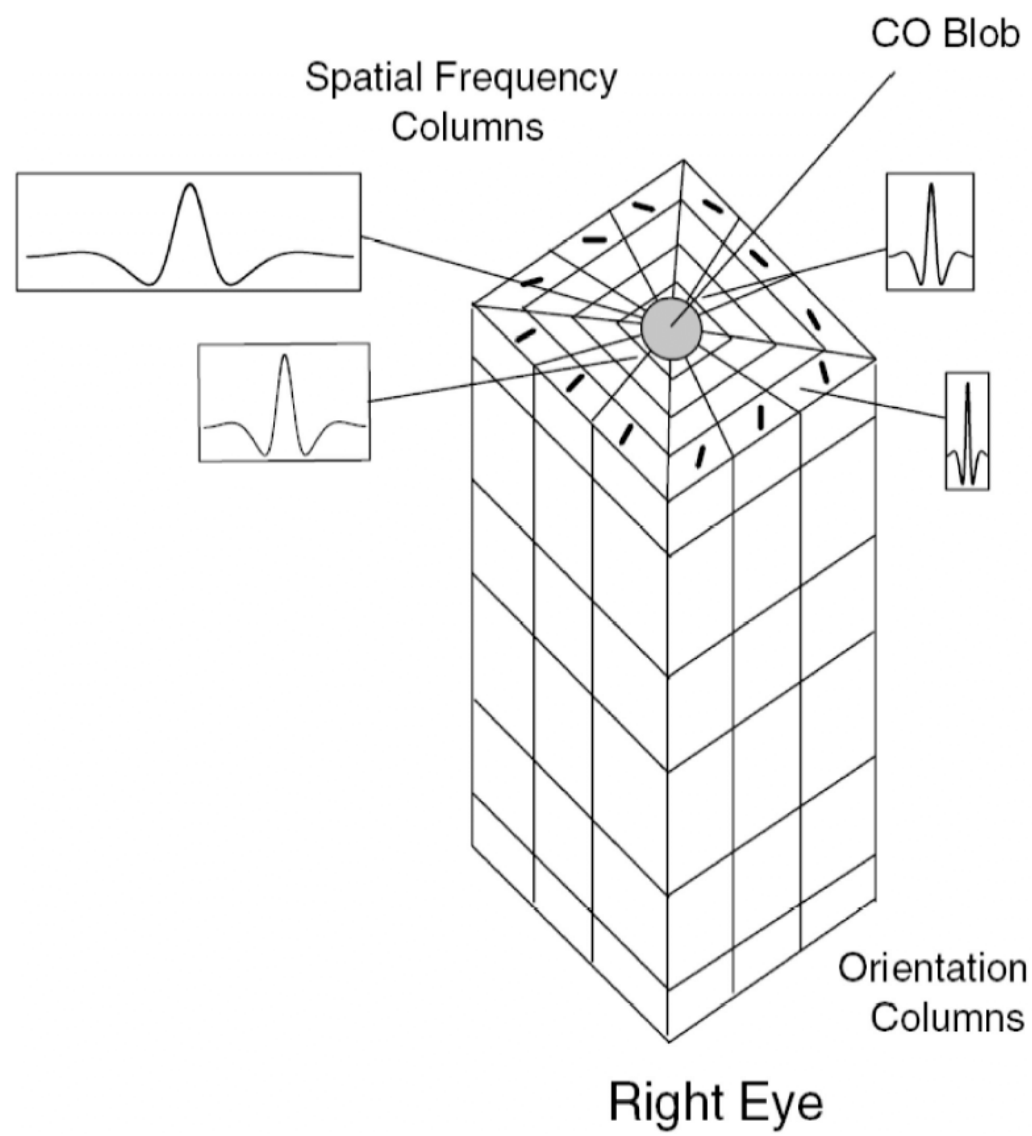
# Rotation and Affine Invariance

*Laurent Sifre*

- Scatterings along translation, rotation and affine groups:

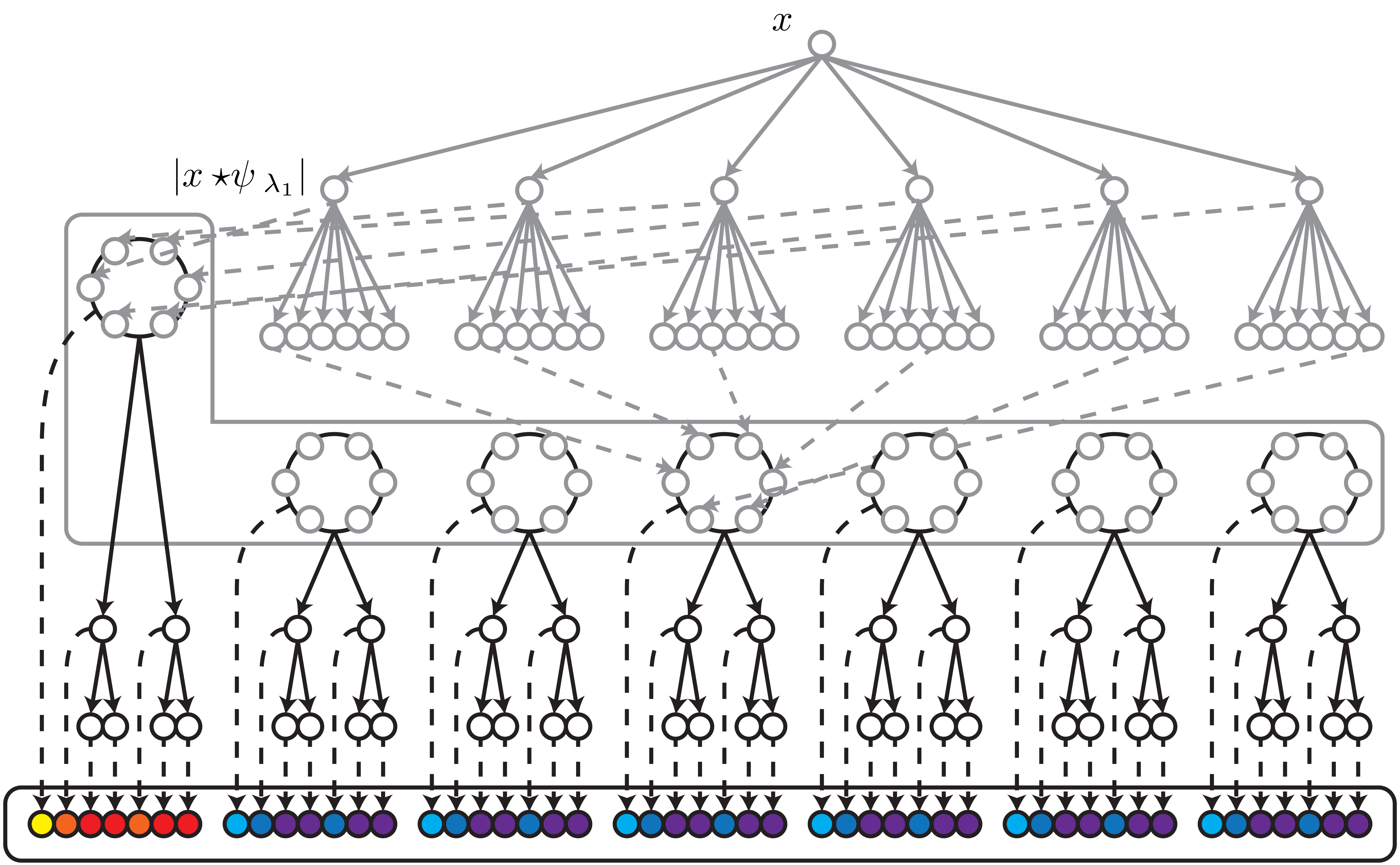
$x \longrightarrow$ [Translat. Invar.] $\longrightarrow$ [Rotation Invar.] $\longrightarrow$ [Affine Invar.] $\longrightarrow Sx$

Wavelet transform along positions, rotations and scales in "V1 hypercolumns"



# Translation and Rotation Invariance

*Laurent Sifre*



# Multiple Scattering Invariants

$x$

↓

Wavelet convolutions along space

↓

Modulus Pooling

↓

Wavelet convolutions along space, rotations, scales

↓

Modulus Pooling

↓

Linear averaging along space, rotations, scales

↓

$S[p]x$

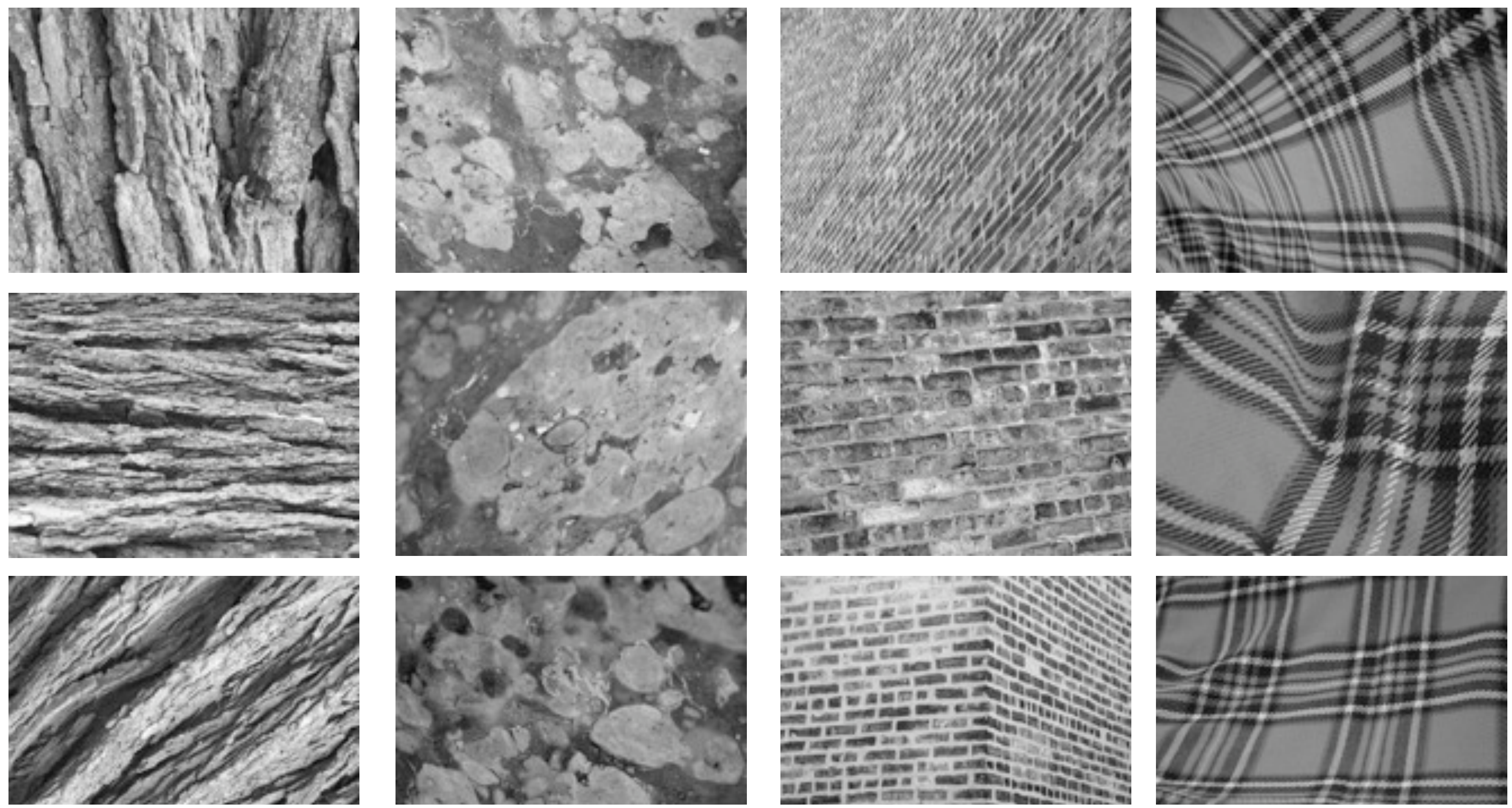
# Rotation and Affine Invariance

*Laurent Sifre*

- Scatterings along translation, rotation and affine groups:

$$x \longrightarrow \boxed{\text{Translat. Invar.}} \longrightarrow \boxed{\text{Rotation Invar.}} \longrightarrow \boxed{\text{Affine Invar.}} \longrightarrow Sx$$

UIUC database:



Classification Errors

| Training | Translation | Transl + Rotation | Affine |
|---|---|---|---|
| 20 | 15 % | 3% | **1**% |

# Unsupervised Learning

- Need to learn informative, stable invariants.
- Over general manifolds as opposed to groups
- The final linear averaging providing adapted invariants can be learned by supervised classifiers (SVM).

- Problem: unsupervised learning of the dictionary and of the non-linear pooling.
- Sparsity is important to build informative invariants: auto-encoders with group sparsity.
- Why does it work ? still a mathematical mystery.

# Conclusion

- An interpretation of convolution networks for groups:
  - Filters must be wavelets
  - Stable pooling: complex modulus + averaging
  - Multilayers: recover lost information and refine invariants
  - Sparsity is needed to preserve information in invariants
  - Normalisation: to «decorrelate» outputs
  - Learning: needed but not for first layers.

- Unsupervised deep learning: still not understood mathematically

- Papers and softwares: www.cmap.polytechnique.fr/scattering