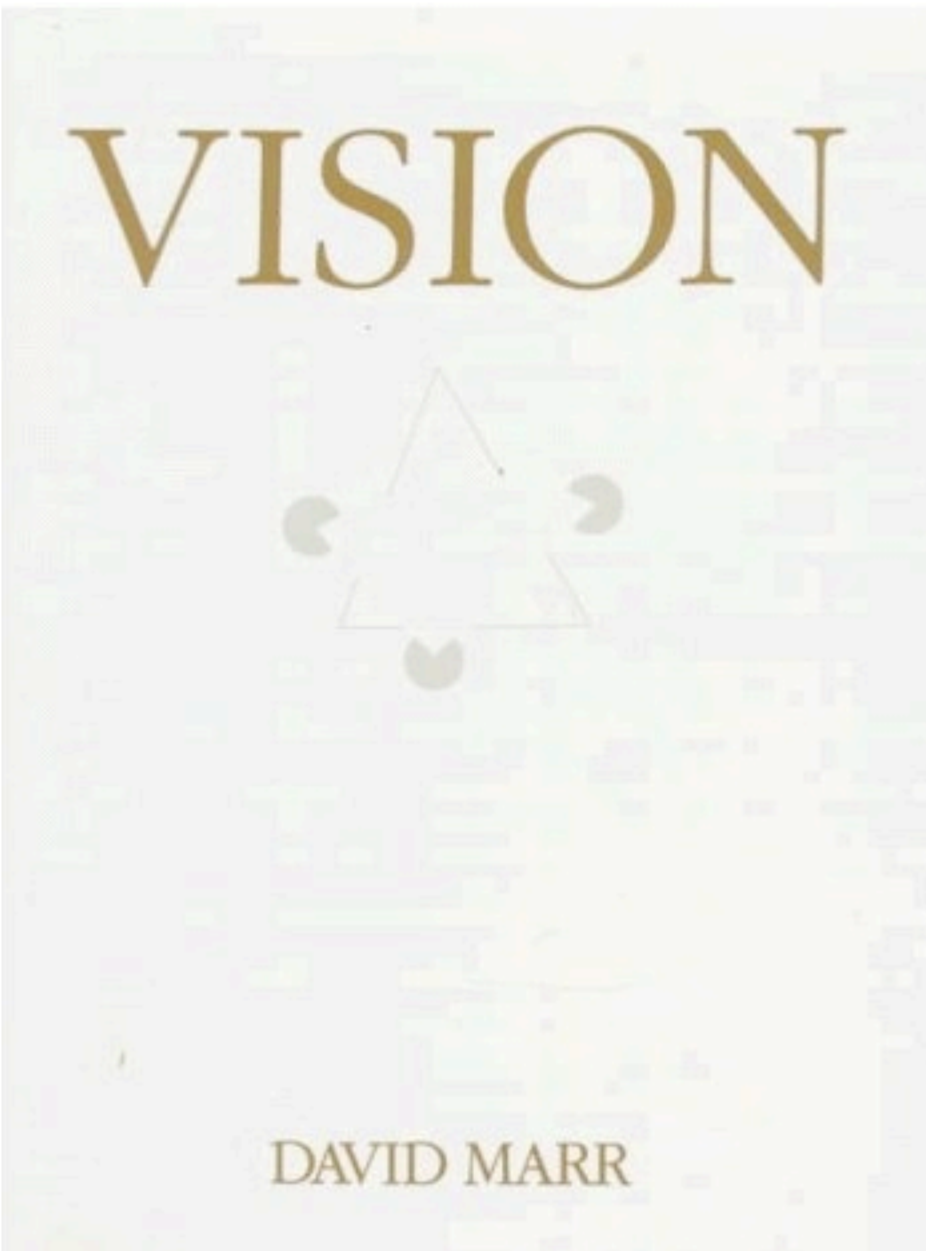
Natural Image Statistics

Siwei Lyu

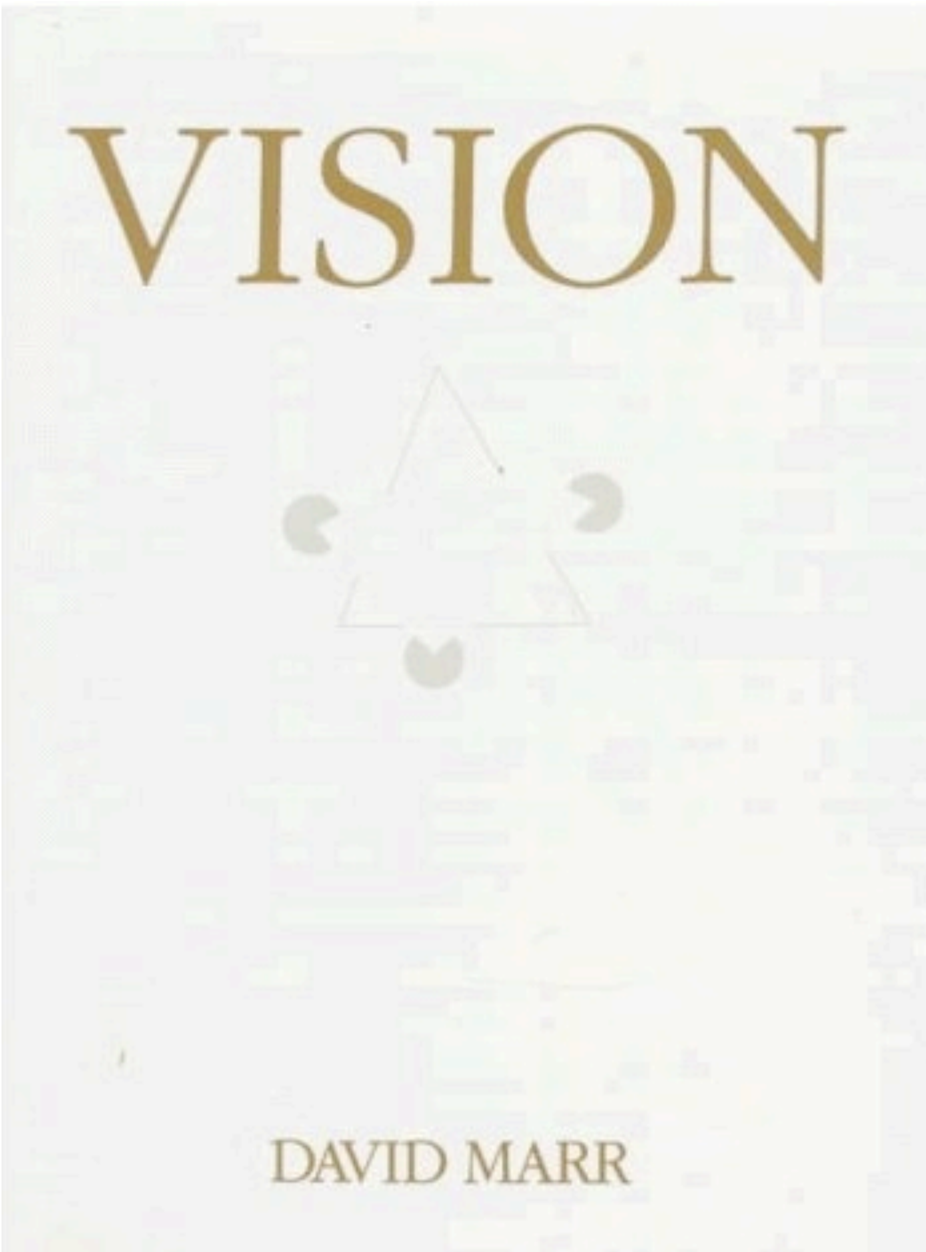
vision and image



Vision is a process that produces from images of the external world a description that is useful to the viewer.

[Marr, 1982]

vision and image



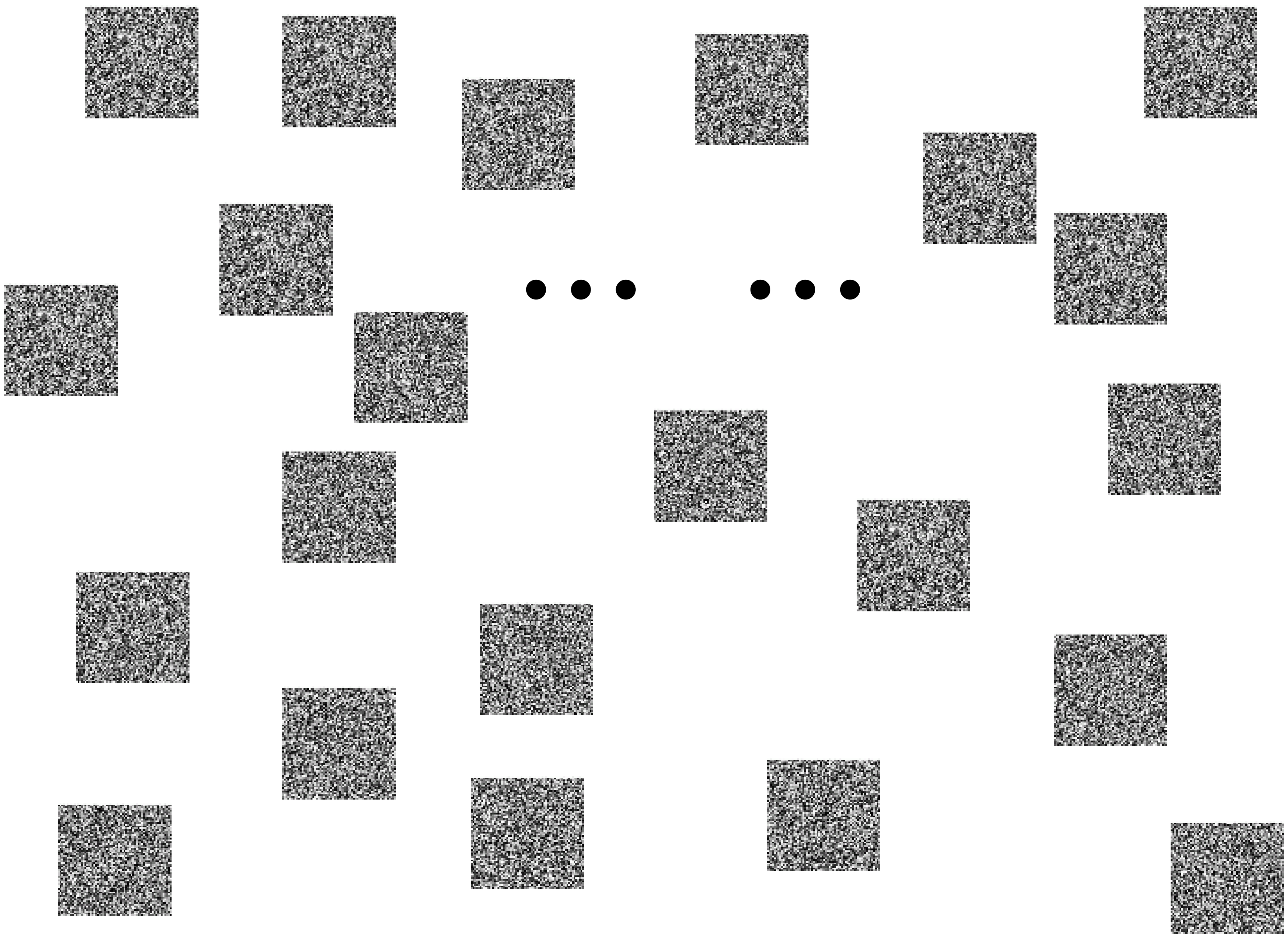
Vision is a process that produces from images of the external world a description that is useful to the viewer.

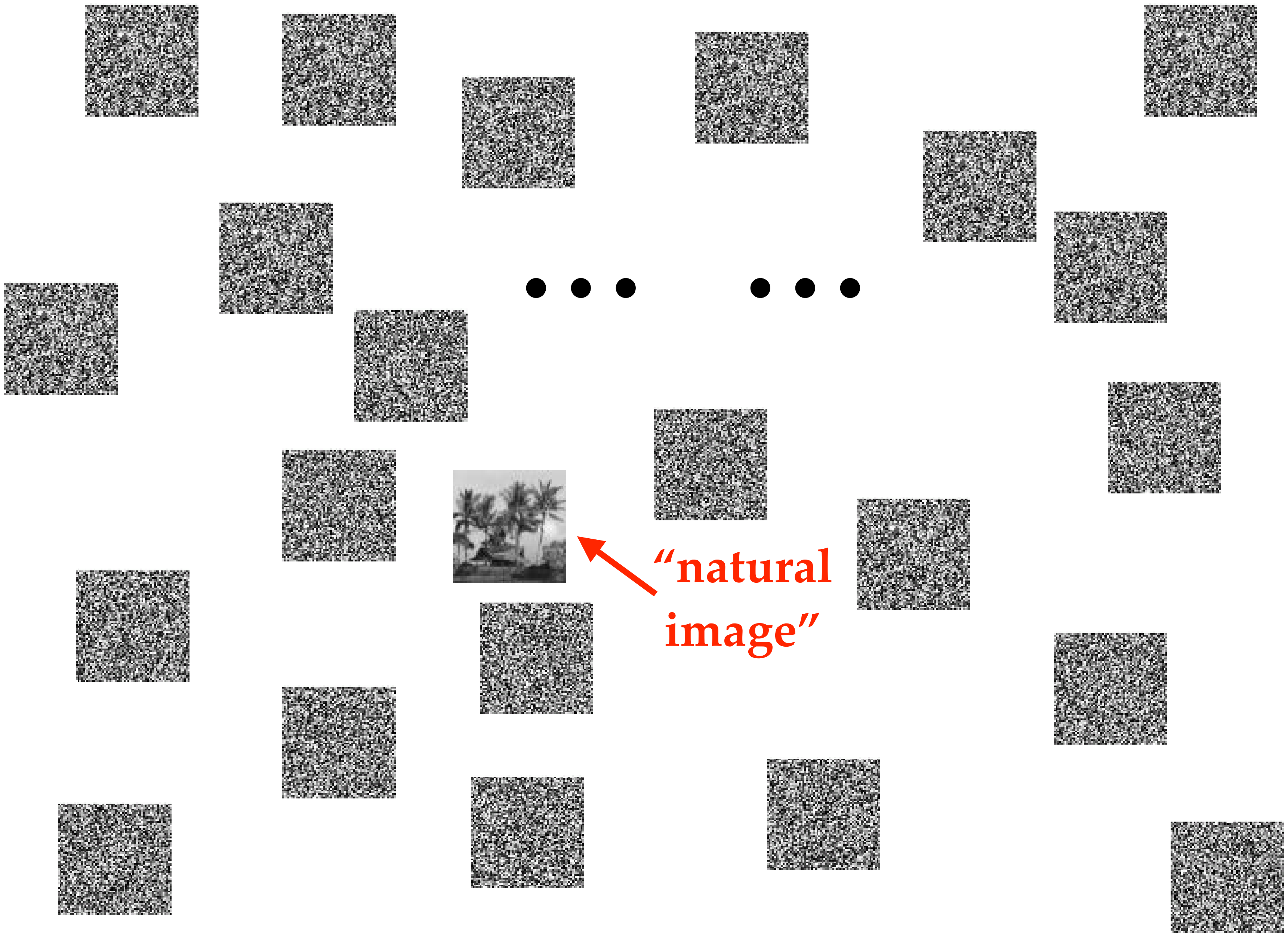
[Marr, 1982]

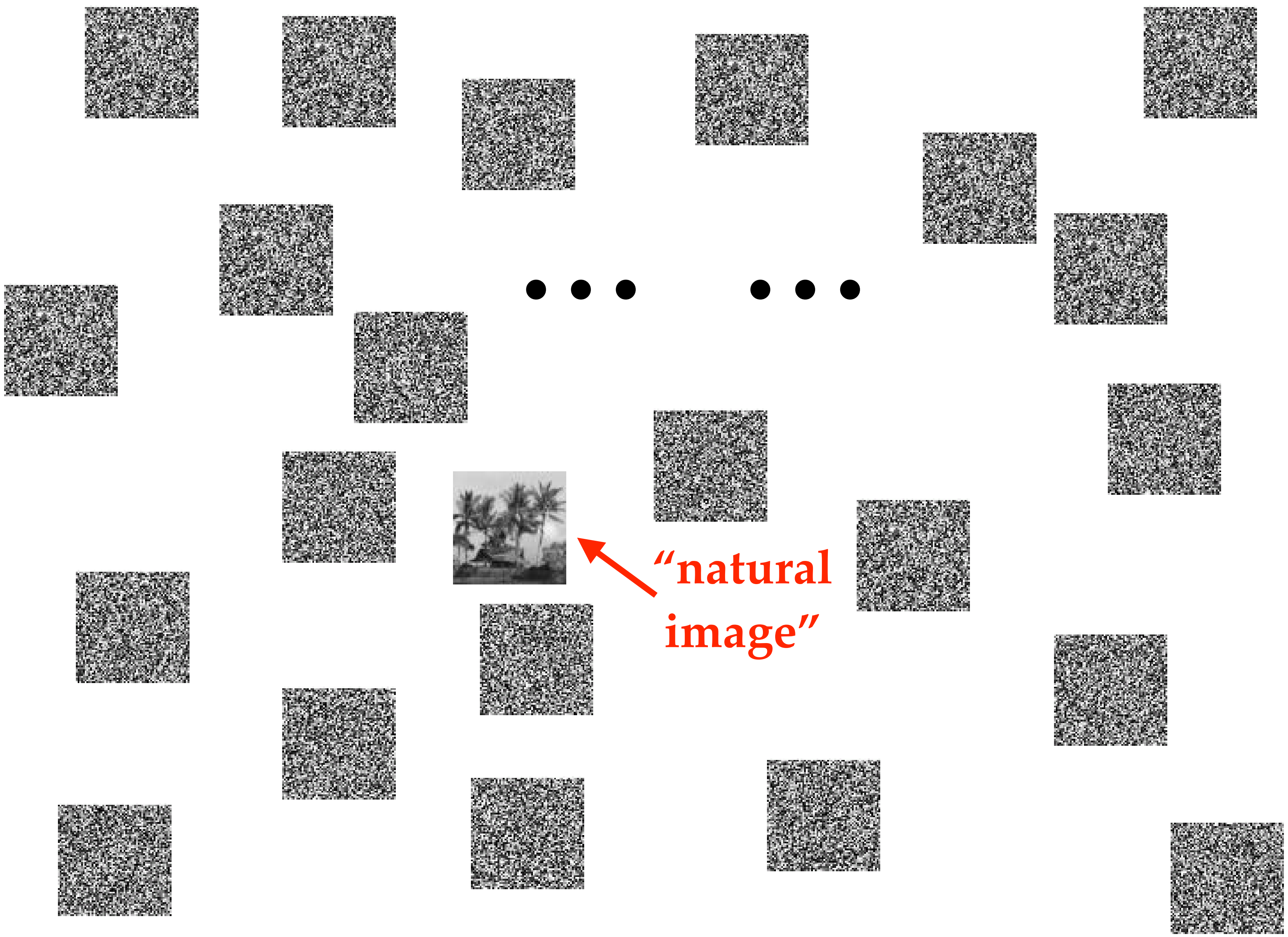
65×65 8-bit gray-scale images: $256^{65\times 65} \sim \mathbf{10^{10^5}}$

seconds since big bang: $\sim \mathbf{10^{17}}$

atoms in the universe: $\sim\mathbf{10^{80}}$




“natural image”

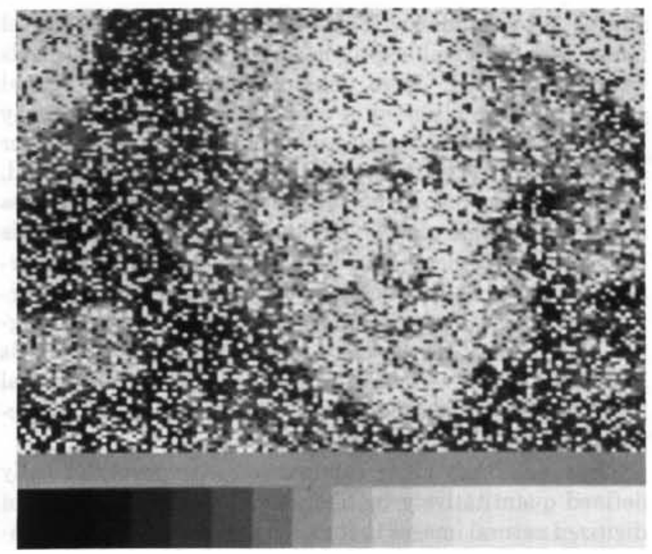


The distribution of natural images is complicated. Perhaps it is something like ***beer foam***, which is mostly empty but contains a thin mesh-work of fluid which fills the space and occupies almost no volume. The fluid region represents those images which are natural in character.

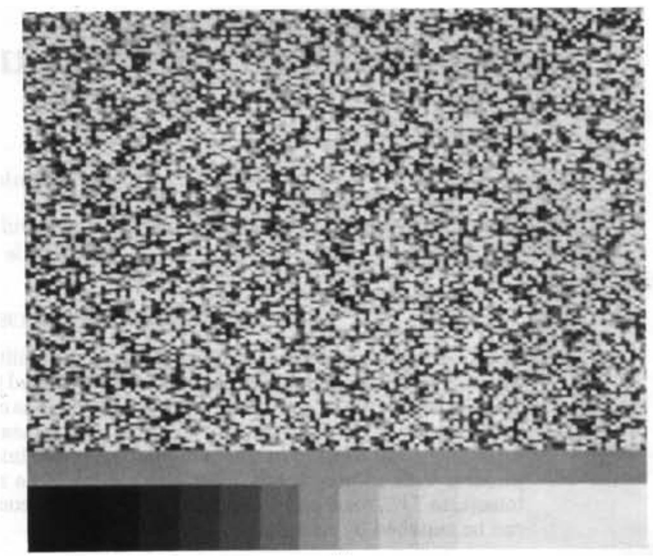
[Ruderman, 1996]

dependency in natural images

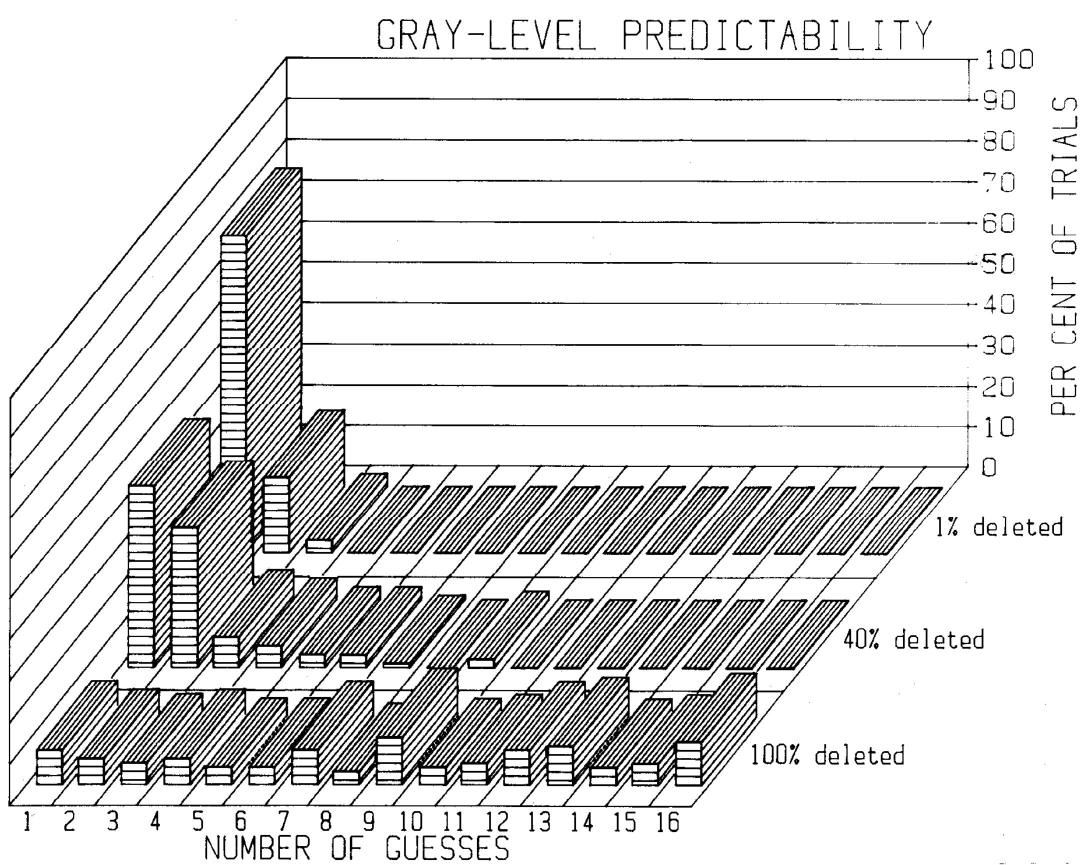
1% deleted



40% deleted



100% deleted



structure
= predictivity
= redundancy
= statistical dependency

[Kersten, 1987]

natural image statistics

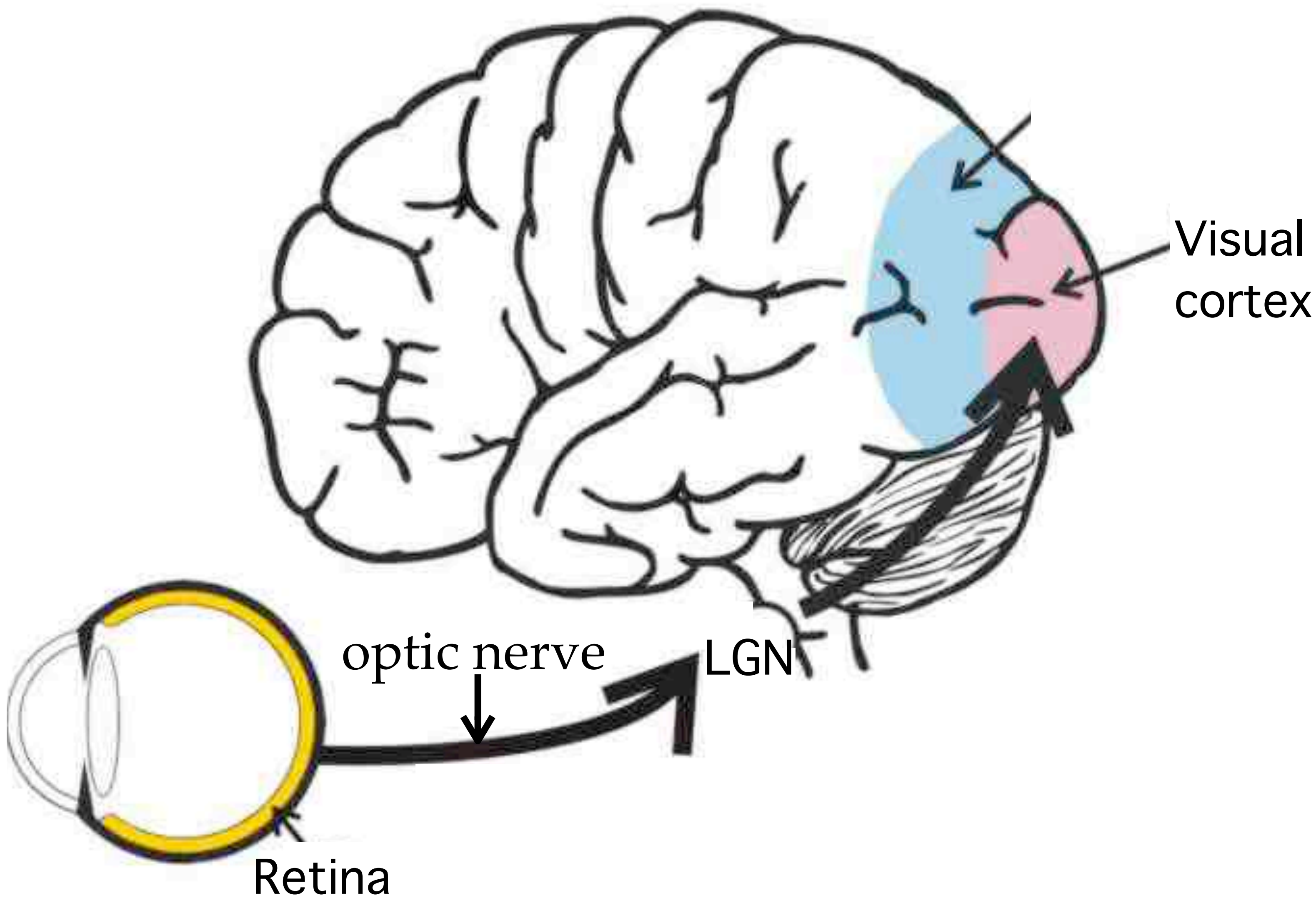
- natural images are a small subset in the image space
- natural images have non-random structures that reflect regularities in the physical world
- natural images as an ensemble can be studied by their common regular statistical properties

biological vision

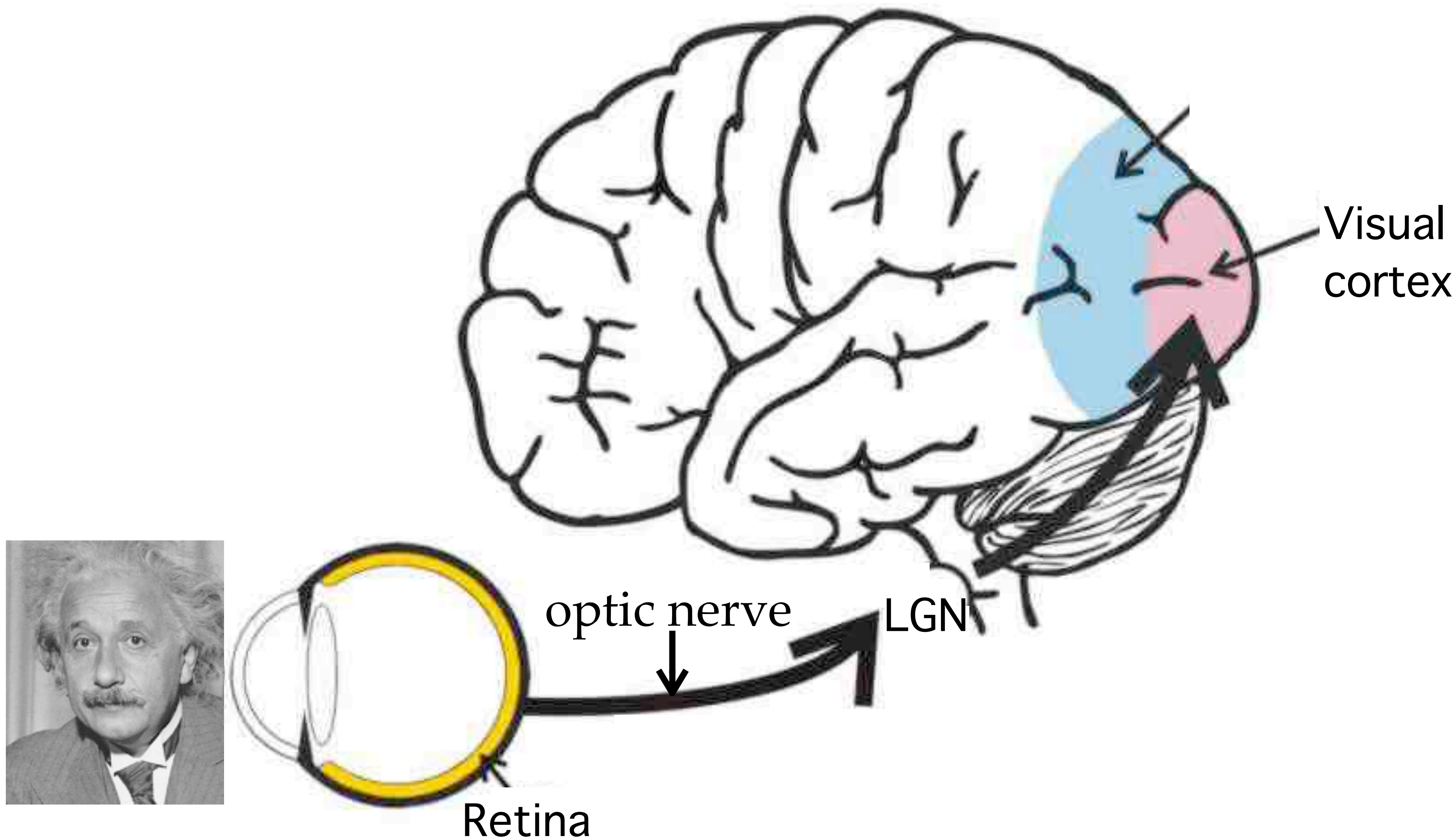
“the [neurally] encoded image is a very partial representation of the light that arrives at the eye: there is only a narrow region of high visual acuity in the fovea; the dynamic range of the sensors is very small; and the representation of wavelet is very coarse. **You would never buy a camera with such poor optics and coarse spatial encoding.** Yet, the visual algorithms can interpret the properties of objects from this poor encoding”

- Brian Wandell, *Foundation of Vision*, 1995

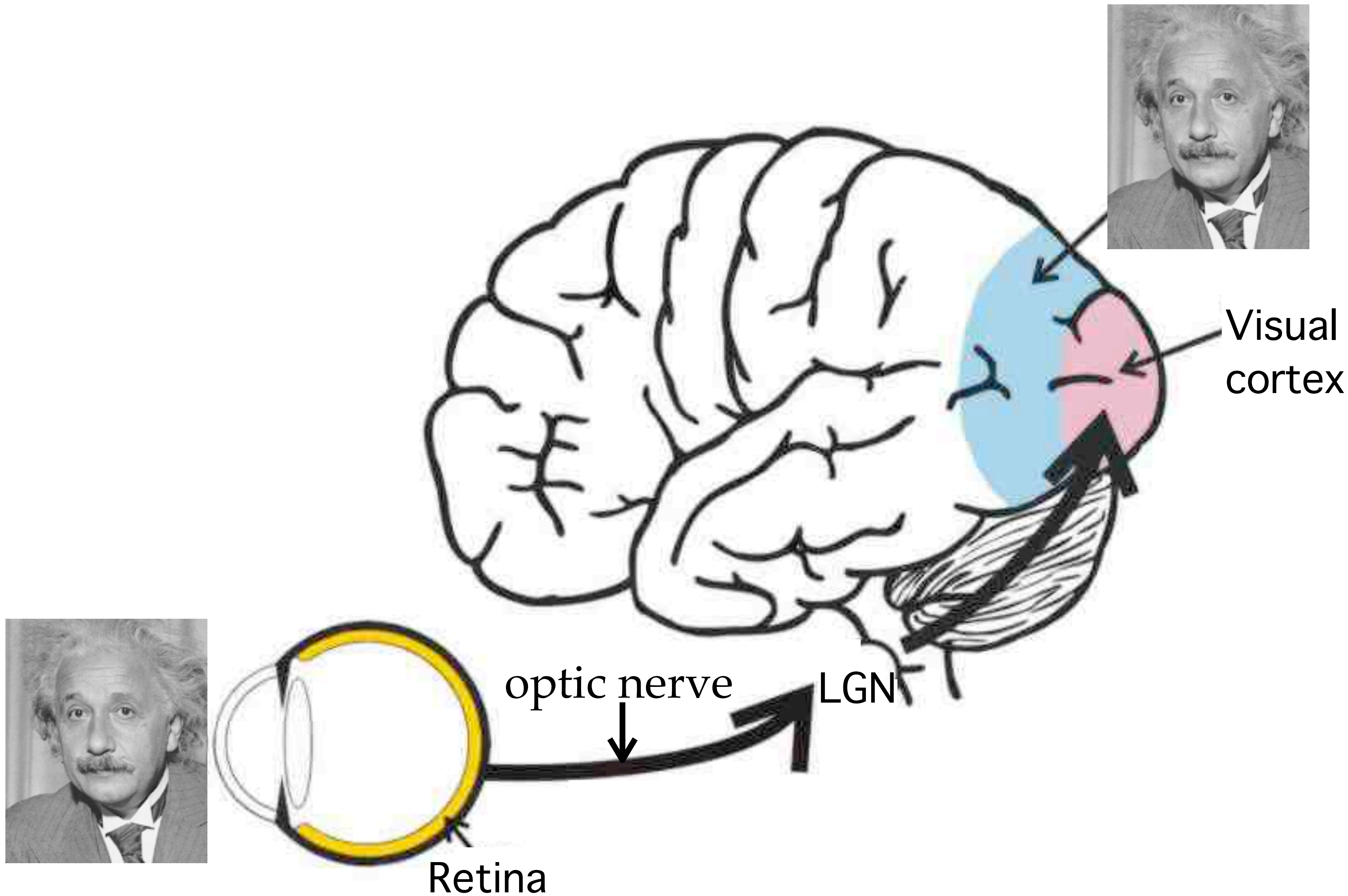
biological vision



biological vision



biological vision



biological vision

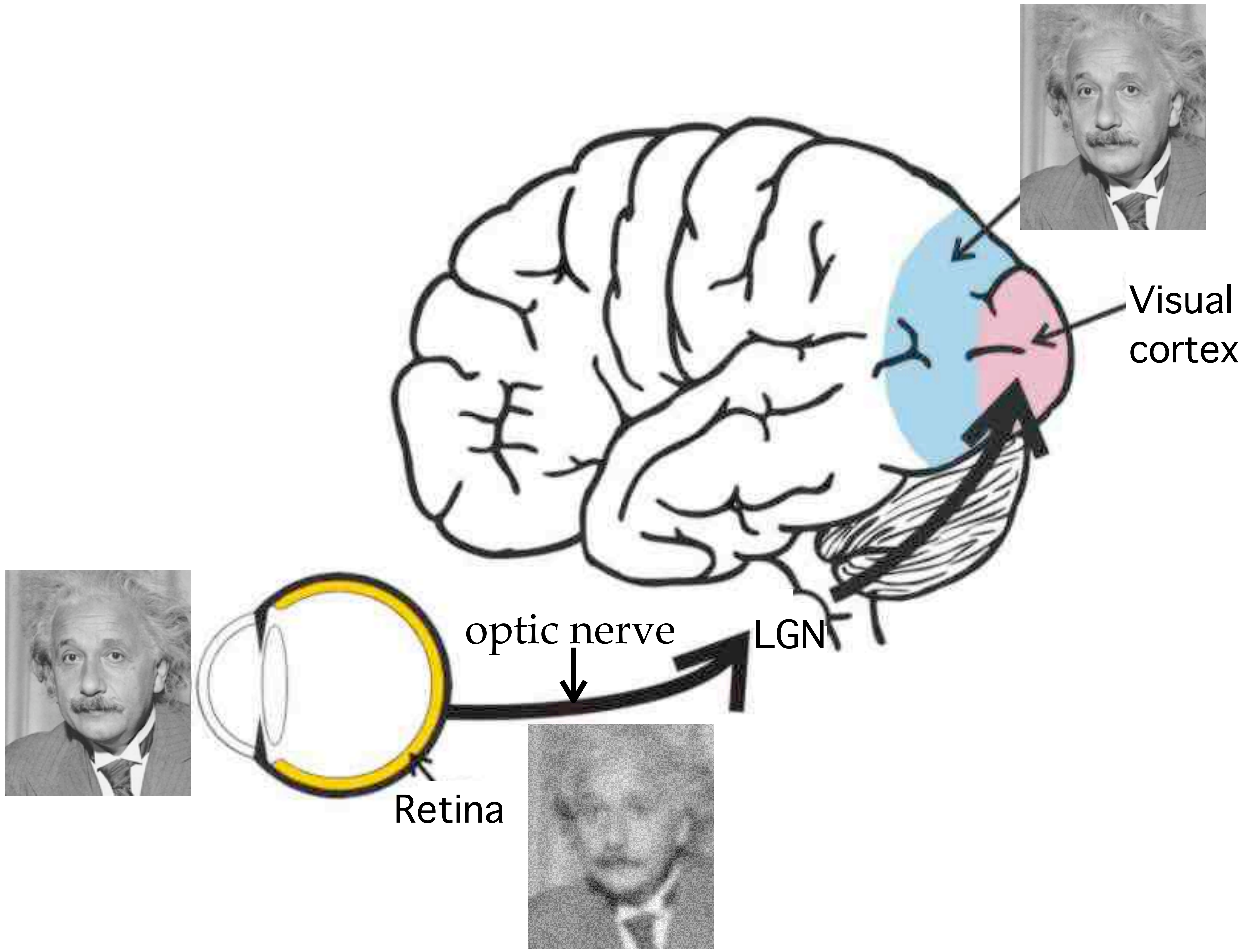
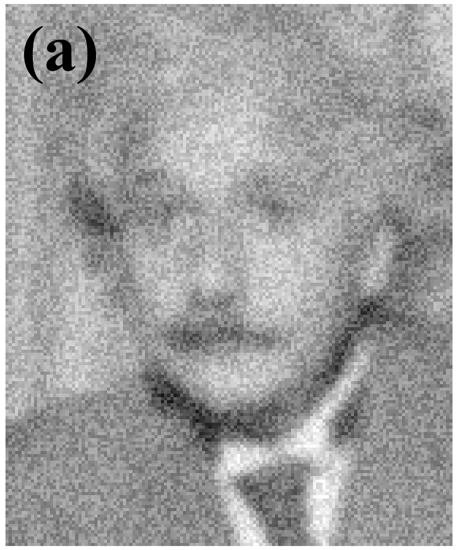


image restoration



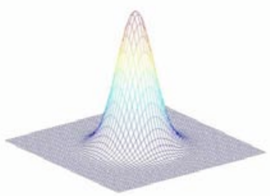
$= \boxed{\;?\;} \otimes$  $+$ noise

image restoration

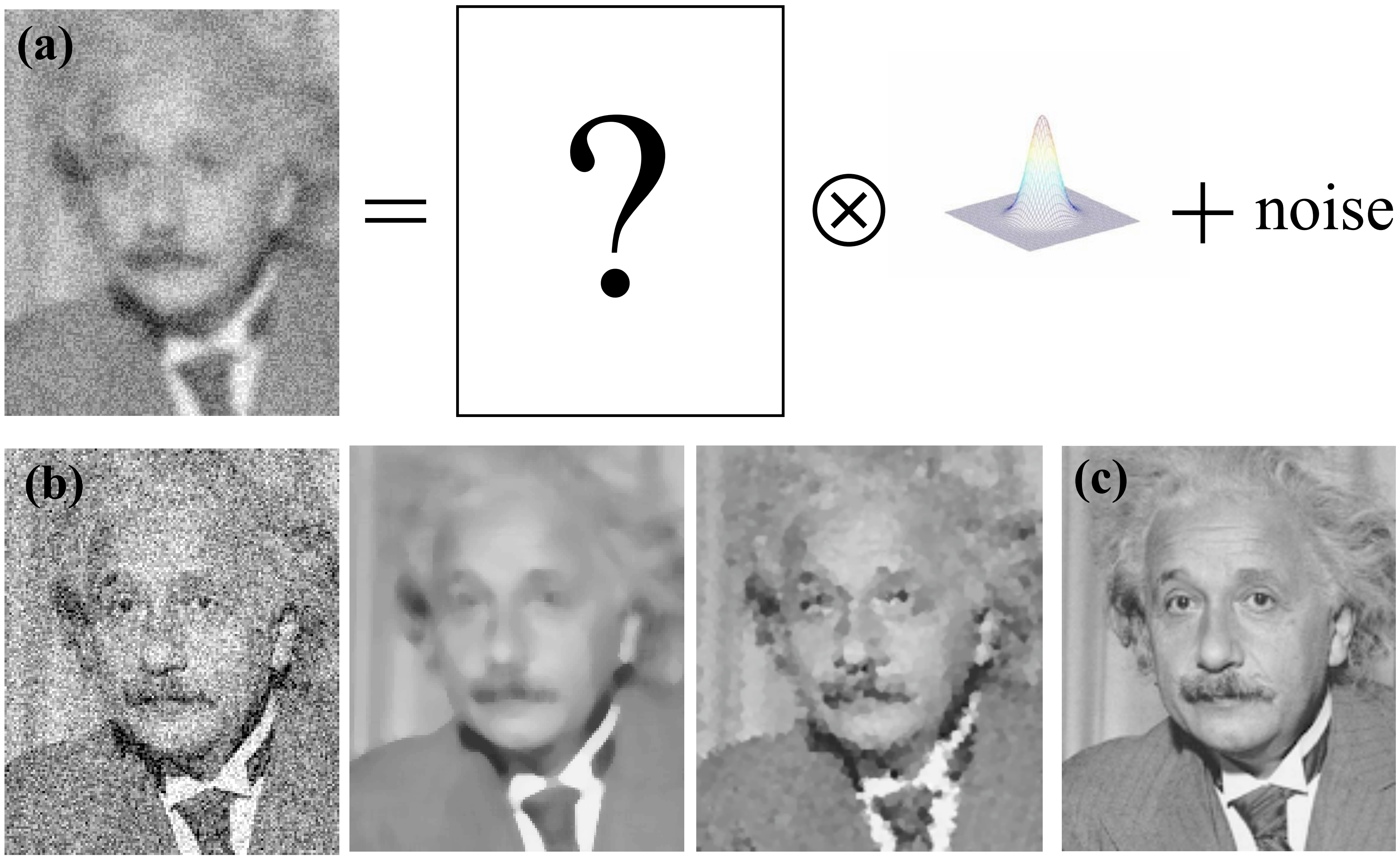
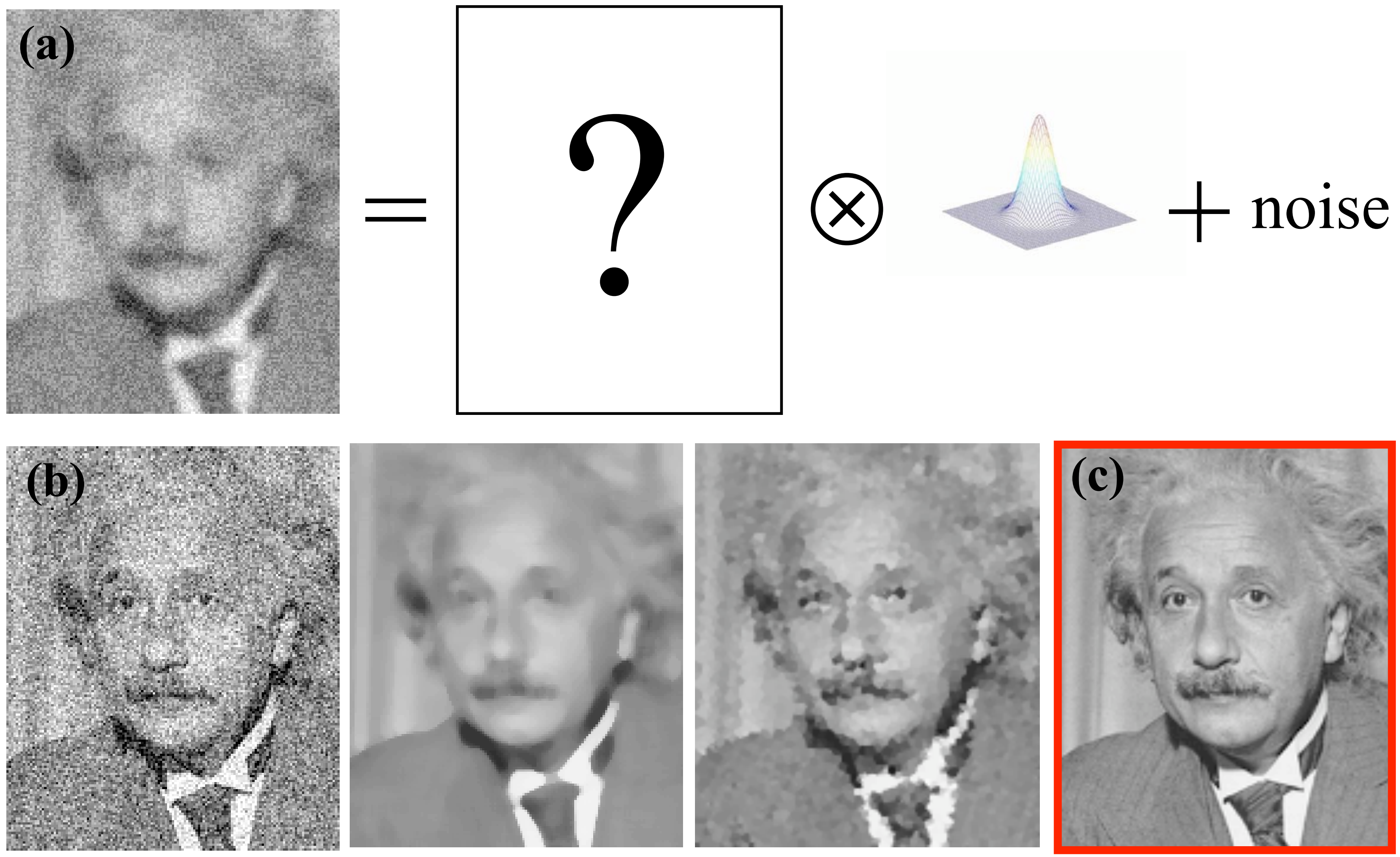
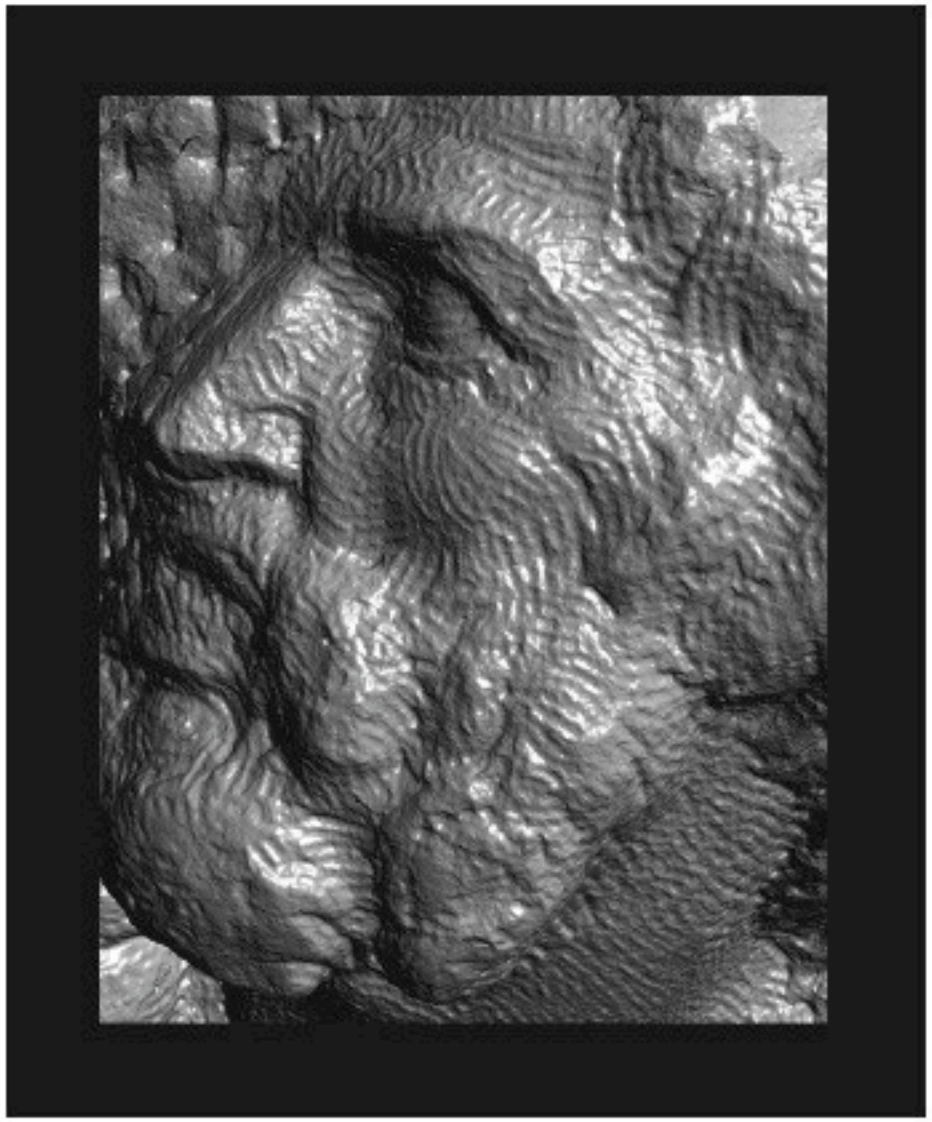


image restoration



surface perception



high skewness

low skewness

[Motoyoshi *etal.*, 2007]

engineering applications

- image compression
  - e.g., JPEG, JPEG 2000
- noise and blur removal, inpainting, super-resolution
  - e.g., [Freeman etal. 2000; Roth & Black, 2005; Levin etal, 2009]
- texture synthesis
  - e.g., [Heeger & Bergen, 1995; Zhu, Wu & Mumford, 2001; Portilla & Simoncelli, 2003]
- visual saliency
  - e.g., [Itti etal, 2003; Gao & Vasconcelos, 2009]
- low level features for object/scene recognition
  - e.g., [Oliva & Torrolba, 2001; Kouh & Poggio, 2009]
- and many more … …

scope

- statistical approach to the study of natural images
  - gray-scale static images
- focus on concepts and their relations, but not on
  - specific mathematical/computational details
  - specific applications in biology/engineering
- follow one particular theme of developments
  - statistical properties observed on ensembles of natural images
  - probabilistic models that capture such properties
  - image representations that simplify such properties

scope

- statistical approach to the study of natural images
  - gray-scale static images
- focus on concepts and their relations, but not on
  - specific mathematical/computational details
  - specific applications in biology/engineering
- follow one particular theme of developments
  - **statistical properties** observed on ensembles of natural images
  - **probabilistic models** that capture such properties
  - **image representations** that simplify such properties

how natural images can be studied

- step 1: collect an image database
  - find a lot of nice-looking images

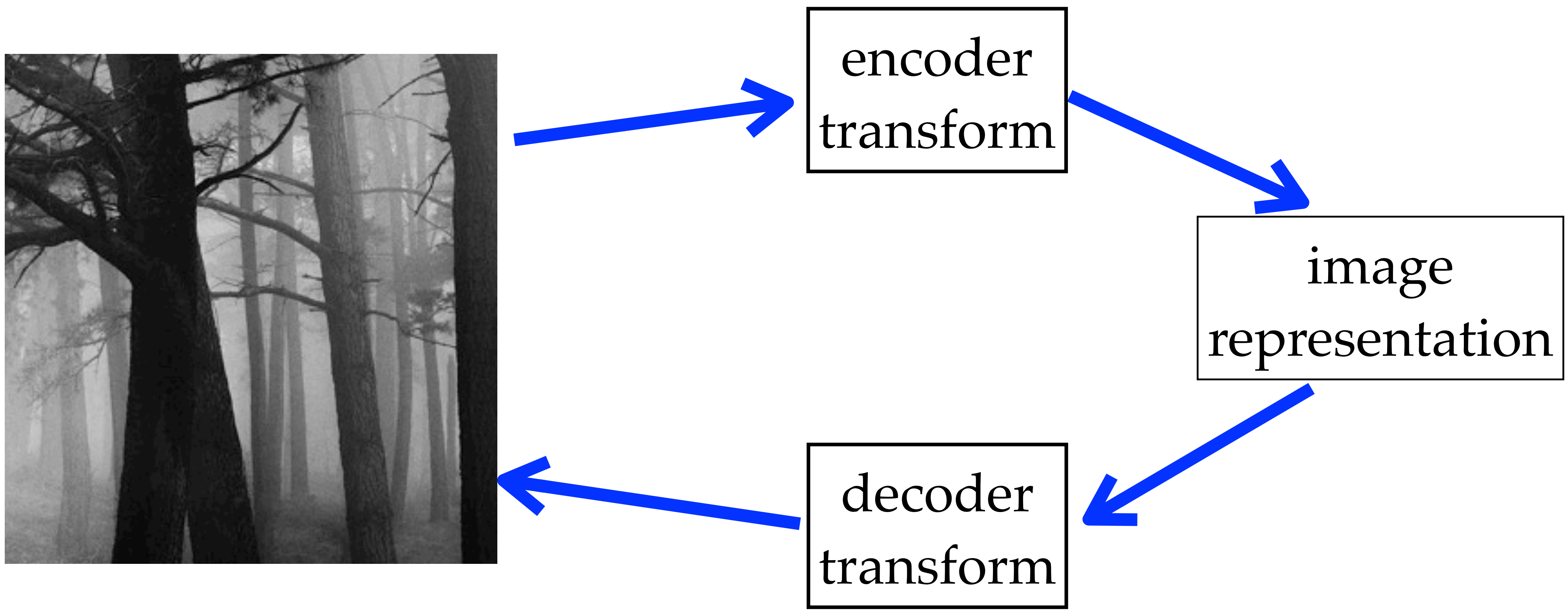
[van Hateren & van der Schaaf, 1998]

how natural images can be studied

- step 1: collect an image database
  - find a lot of nice-looking images
- step 2: choose an image representation
  - a language describe and a tool to probe these images

image representations

- encoder / decoder: information bottleneck
  - preservation of essential and relevant structures
  - special case: perfect reconstruction



why representation matters?

why representation matters?

- example: numbers
  - Arabic: **123**
  - Roman: **MCXXIII**
  - binary: **1111011**
  - English: **one hundred and twenty three**
  - Japanese: 百二十三

why representation matters?

- example: numbers
  - Arabic: **123**
  - Roman: **MCXXIII**
  - binary: **1111011**
  - English: **one hundred and twenty three**
  - Japanese: 百二十三
- operations
  - multiply by **10**
  - multiply by **4**

pixel representation

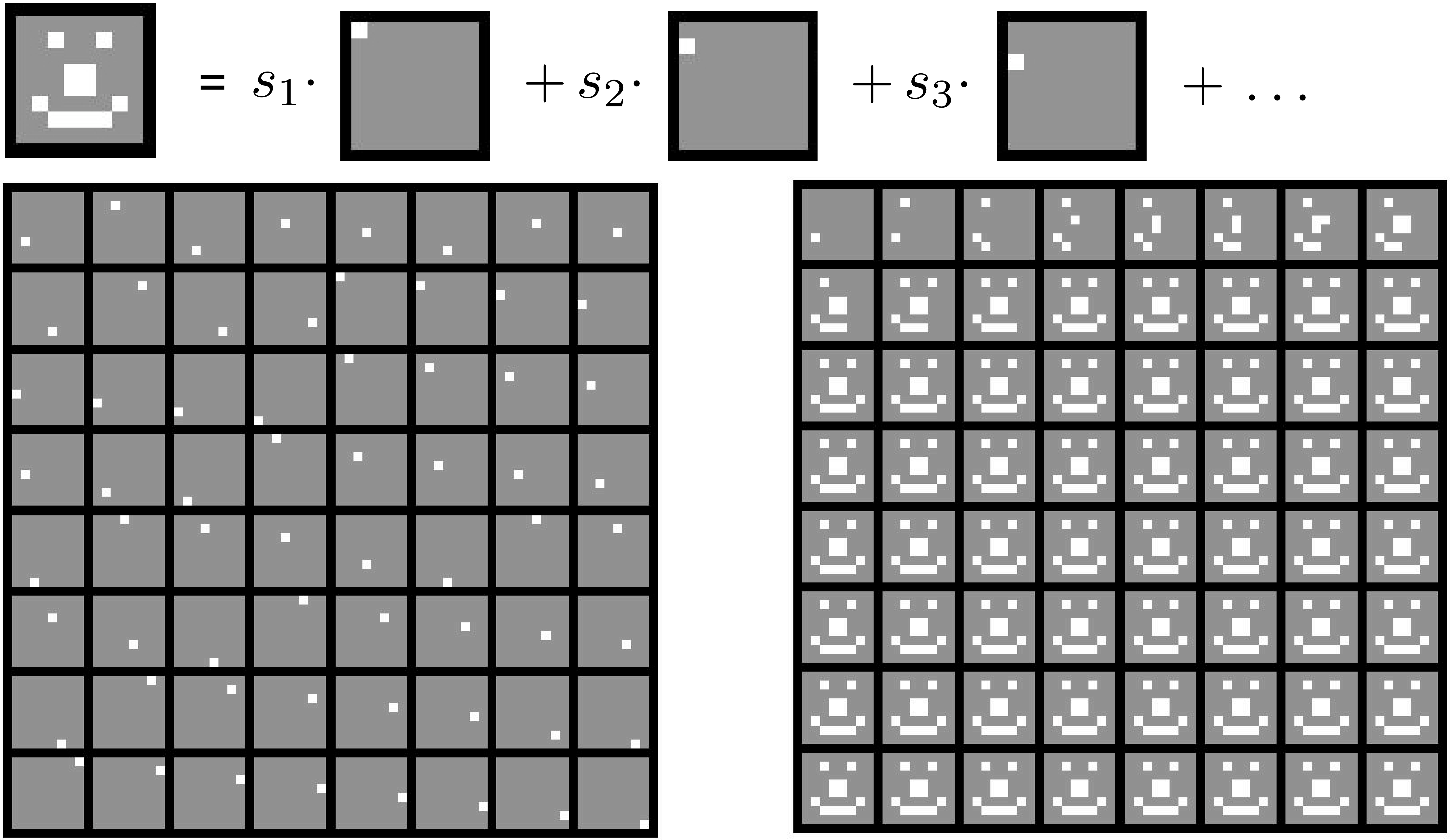
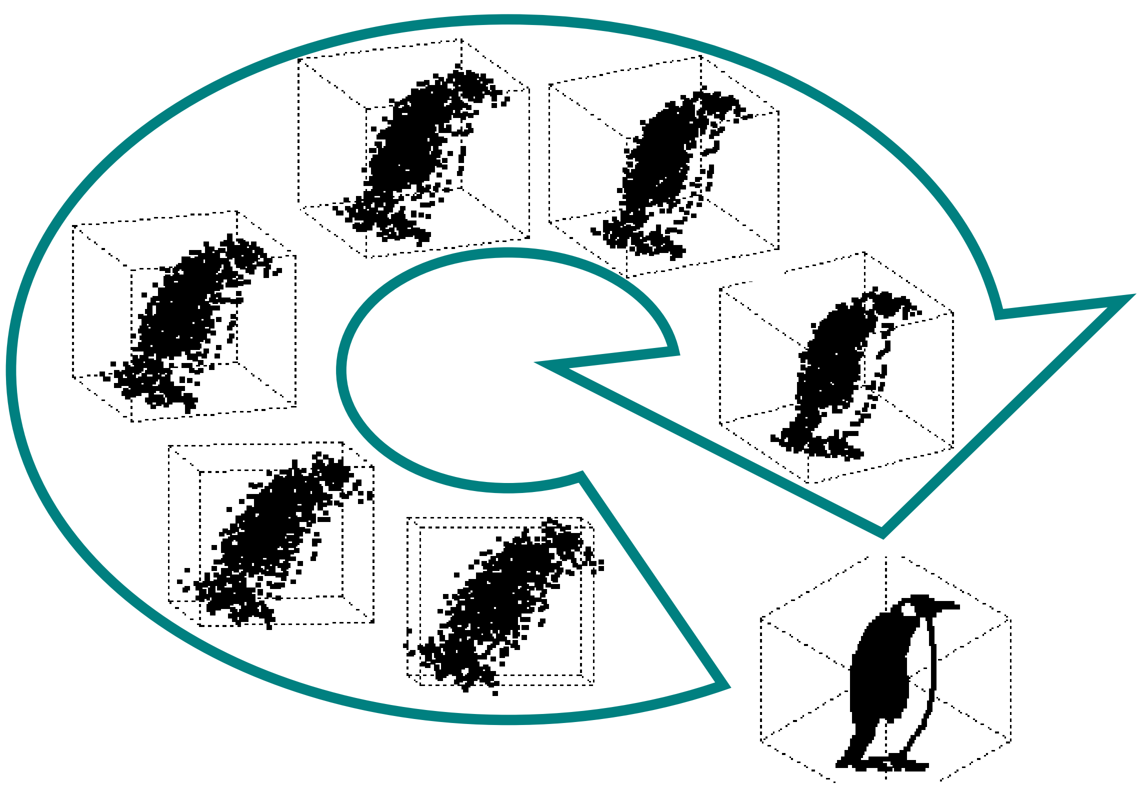
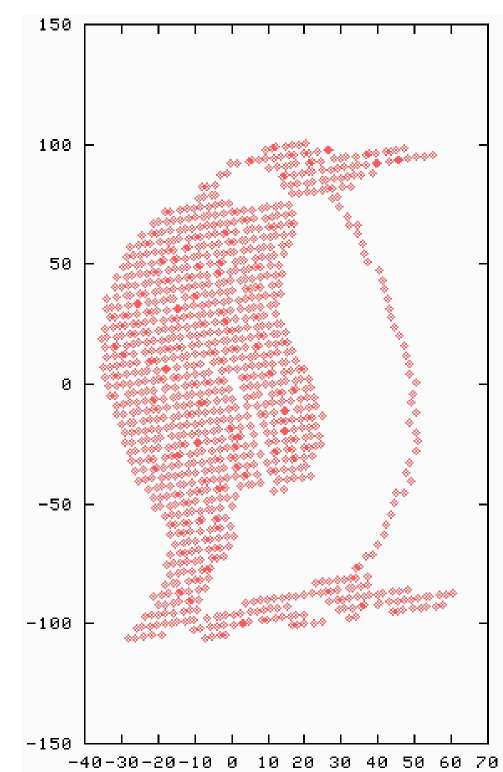


figure courtesy of M. Bethge

desiderata

- simplicity of the encoder/decoder transforms
  - linear transform is preferred
- simplicity of the representation
  - e.g., reveal lower intrinsic dimension





how natural images can be studied

- step 1: collect an image database
  - find a lot of nice-looking images
- step 2: choose an image representation
  - a language describe and a tool to probe these images
- step 3: make observations of statistical properties
  - find something interesting and unexpected

statistical observations

- pixel representation
  - second-order pixel correlations
  - scale invariance
- frequency representation
  - power law distribution of power
- band-pass filtered representation
  - heavy-tail non-Gaussian marginals
  - sparsity of representationss
  - strong higher-order dependency of nearby representationss
  - decay of dependency with distance
- many more ....

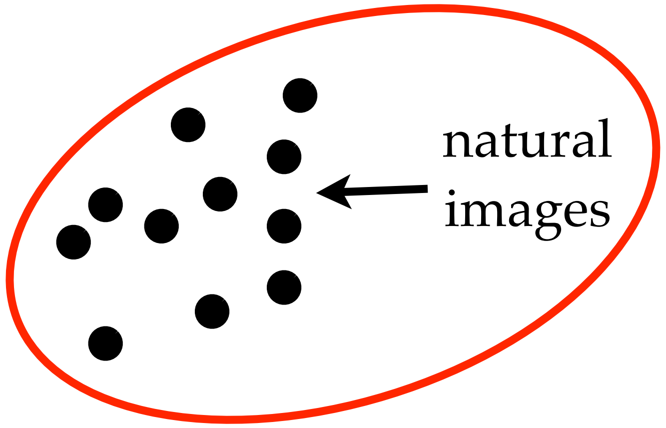
how natural images can be studied

- step 1: collect an image database
  - find a lot of nice-looking images
- step 2: choose an image representation
  - a language describe and a tool to probe these images
- step 3: make observations of statistical properties
  - find something interesting and unexpected
- step 4: devise a mathematical model for these observations
  - give a concise description and/or an (formal) explanation why natural images have such properties

how to construct model

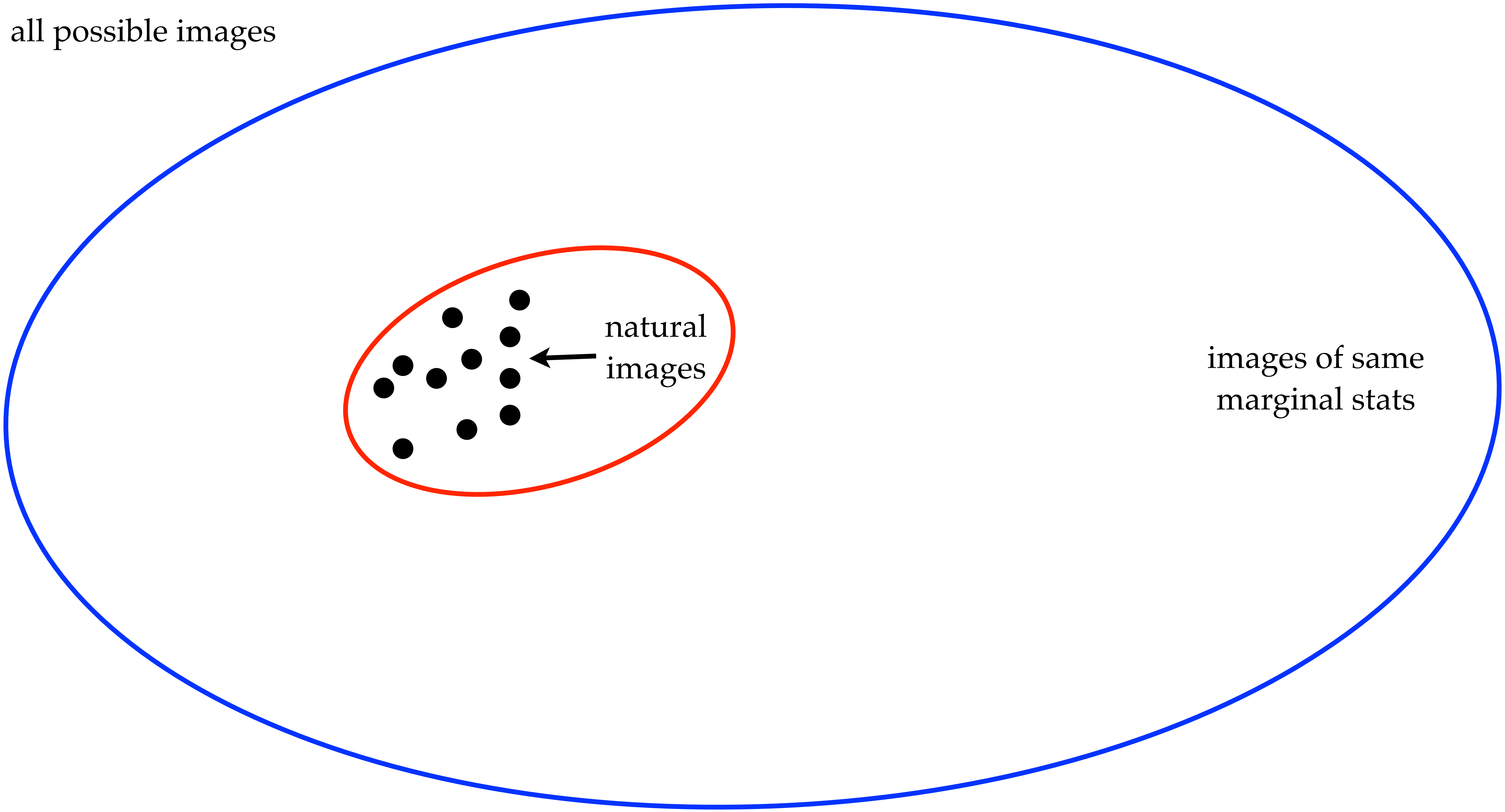
onion peeling

all possible images



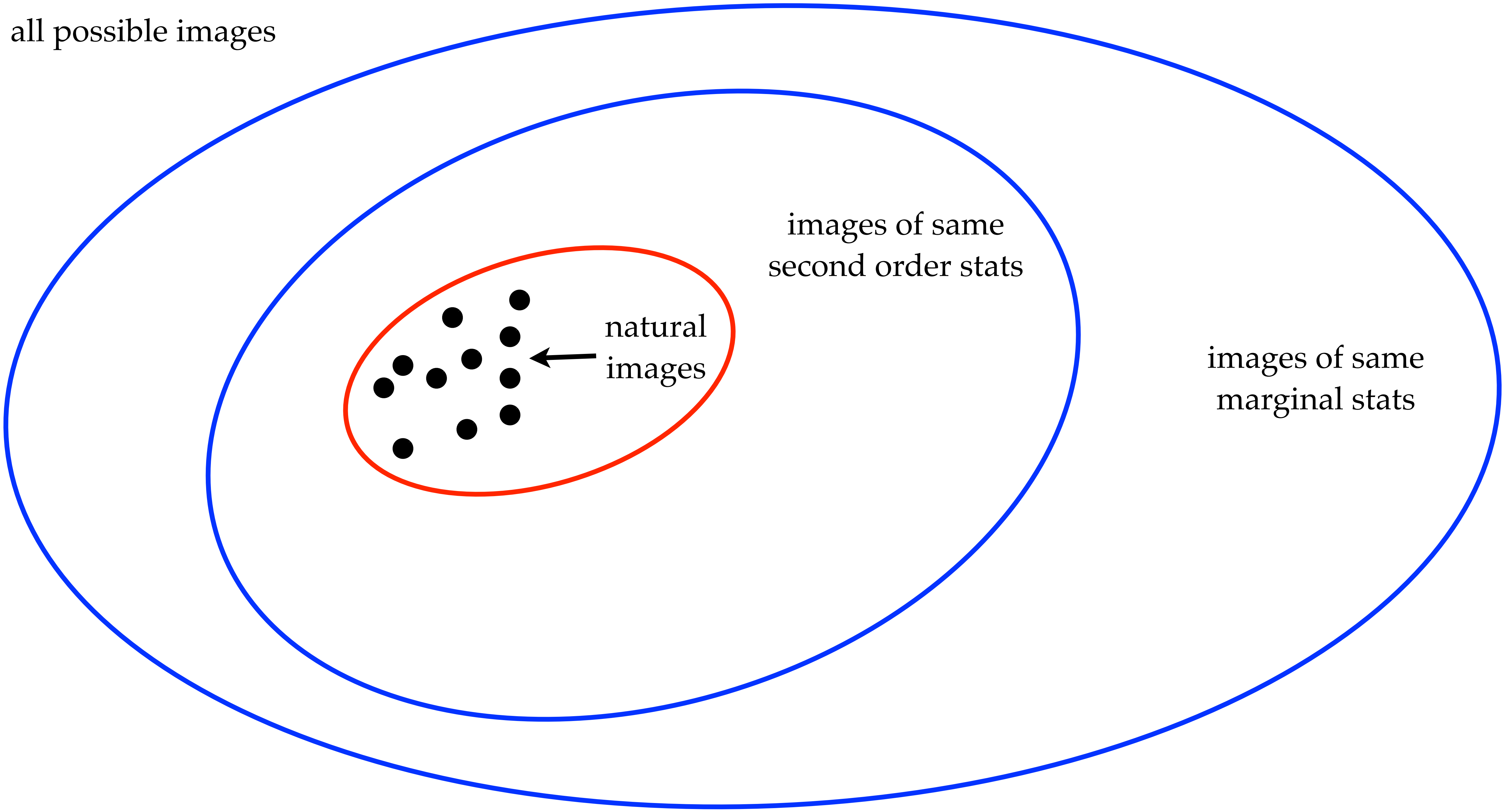
how to construct model

onion peeling

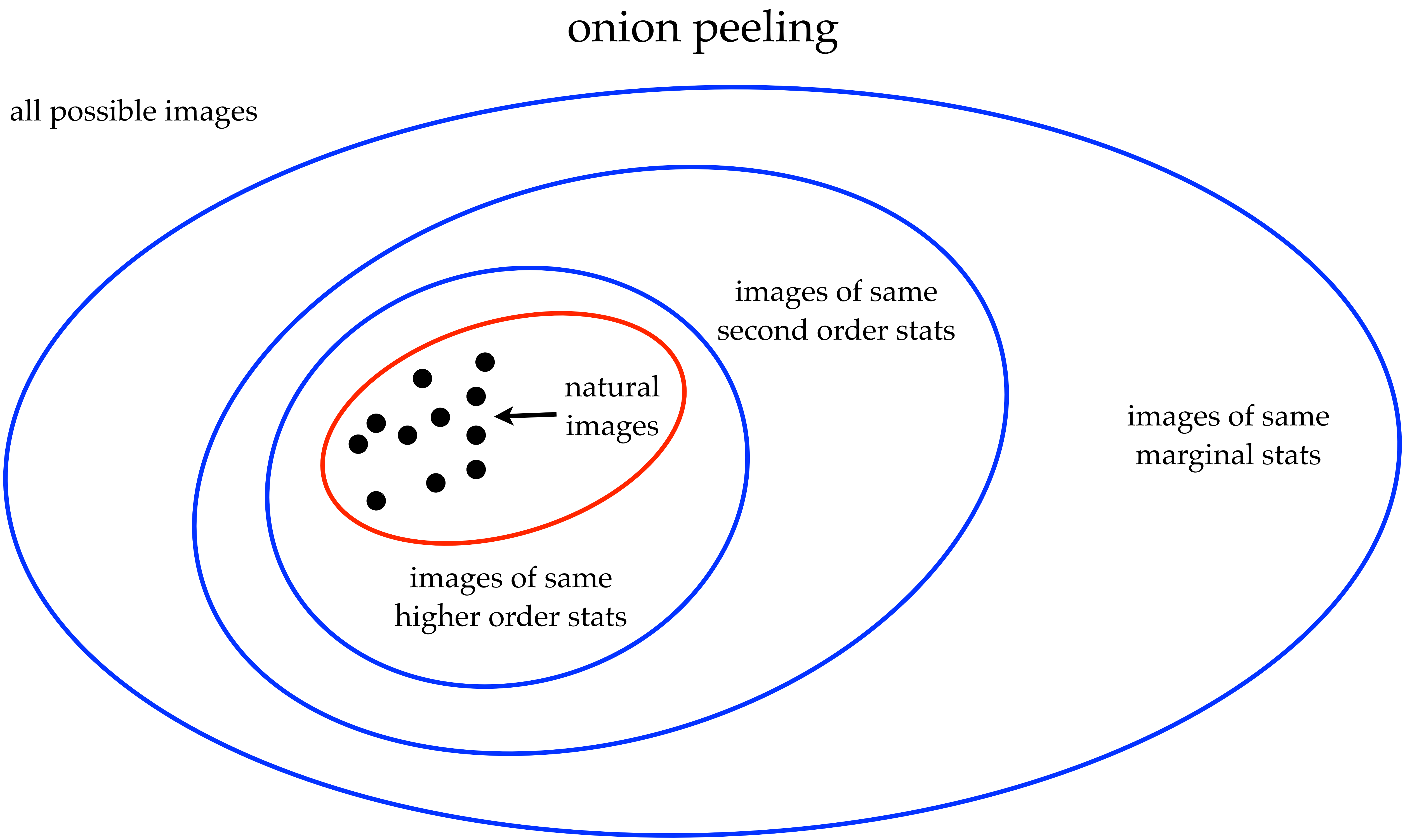


how to construct model

onion peeling



how to construct model



from statistics to model

- principle of maximum entropy [Jaynes, 1954]
  - given a set of statistical constraints on data

$$E(f(x)) = c$$

  - choose a probabilistic model with maximum entropy

$$p^\star = \operatorname*{argmax}_p \mathcal{H}(p)$$

  - solution

$$p^\star(x) \propto \exp(-\lambda f(x))$$

λ is determined by c

maxEnt examples

- constraint on range -> uniform
- matching mean -> exponential
- matching covariance -> Gaussian
- matching all singleton marginals -> factorial model

$$\forall i, p_i(x_i) = q_i(x_i) \Rightarrow p^\star(\vec{x}) = \prod_i q_i(x_i)$$

- matching all clique marginals -> Markov random field

$$\forall \text{clique } c, p_c(\vec{x}_c) = q_c(\vec{x}_c) \Rightarrow p^\star(\vec{x}) \propto \exp(-\sum_c \lambda_c(\vec{x}_c))$$

[Schneidman etal., 2003]

Bayesian inference

- maximum a posterior (MAP)

$$\mathbf{x}_{\mathrm{MAP}} = \operatorname*{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \operatorname*{argmax}_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

- minimum mean squares error (MMSE)

$$\begin{aligned} \mathbf{x}_{\mathrm{MMSE}} &= \operatorname*{argmin}_{\mathbf{x}'} \int_{\mathbf{x}} \|\mathbf{x} - \mathbf{x}'\|^2 p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \\ &= \frac{\int_{\mathbf{x}} \mathbf{x} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}}{\int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}} = E(\mathbf{x}|\mathbf{y}) \end{aligned}$$

how natural images can be studied

- step 1: collect an image database
  - find a lot of nice-looking images
- step 2: choose an image representation
  - a language describe and a tool to probe these images
- step 3: make observations of statistical properties
  - find something interesting and unexpected
- step 4: devise a mathematical model for these observations
  - give a concise description and/or an (formal) explanation why natural images have such properties
- step 5: improve the representation, go back to step 3

desiderata

- simplicity of the encoder/decoder transforms
  - linear transform is preferred
- simplicity of the representation
  - lower intrinsic dimension
  - simplified statistical structure
    - reduce statistical dependency

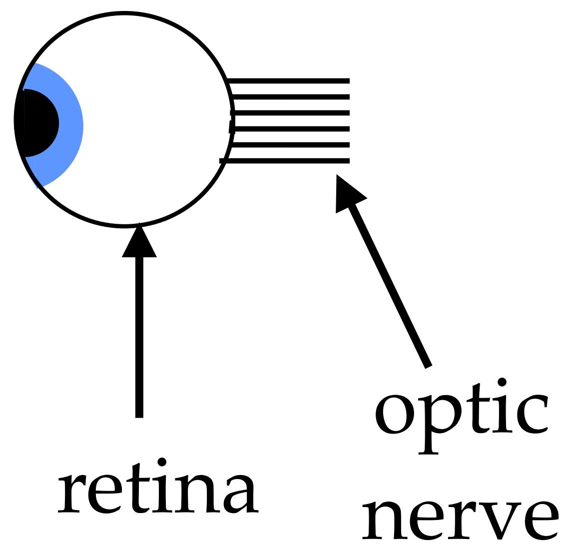
measure statistical dependency

- multi-information
  - [Studeny and Vejnarova, 1998]

$$
\begin{aligned}
I(\vec{x}) &= D_{\mathrm{KL}}\left(p(\vec{x}) \,\middle\|\, \prod_k p(x_k)\right) \\
&= \sum_k H(x_k) - H(\vec{x})
\end{aligned}
$$

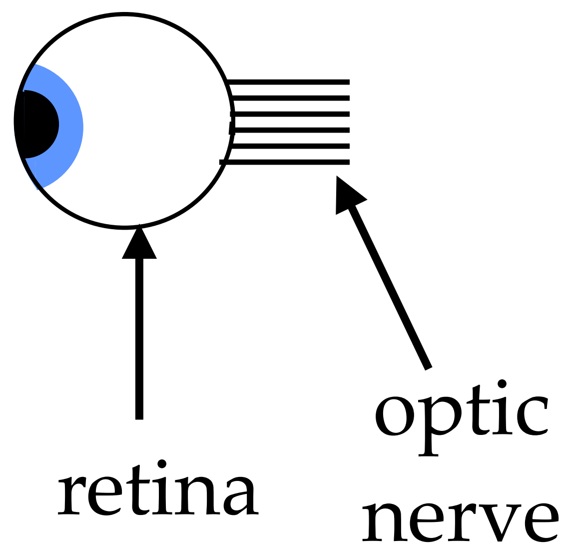
- non-negative with any density over x
- zero when p(x) is factorial
  - elements of x are mutually independent
  - justifies factorial models have maximum entropy with constraints on singleton marginal densities

biology: efficient coding [Attneave, 1954; Barlow, 1961]



optic nerve has a channel capacity C

biology: efficient coding [Attneave, 1954; Barlow, 1961]

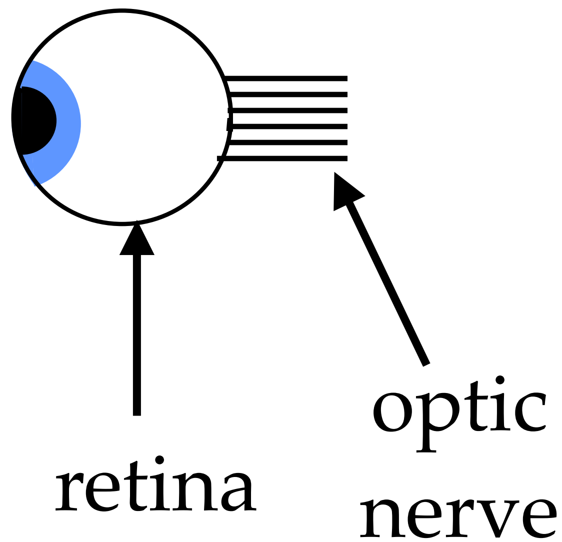


coding efficiency [Attick, 1991]

$$
\begin{aligned}
E &= \frac{H(\vec{x})}{C} = \frac{\sum_i H(x_i)}{C} \frac{H(\vec{x})}{\sum_i H(x_i)} \\
&= \frac{\sum_i H(x_i)}{C} \frac{\sum_i H(x_i) - I(\vec{x})}{\sum_i H(x_i)}
\end{aligned}
$$

optic nerve has a channel capacity C

biology: efficient coding [Attneave, 1954; Barlow, 1961]



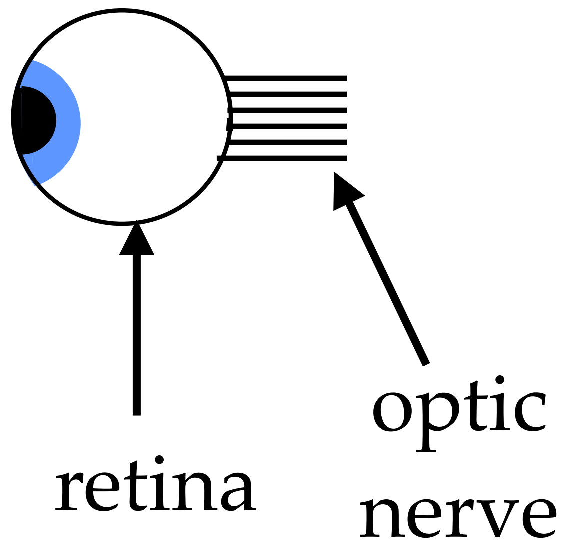
coding efficiency [Attick, 1991]

$$
\begin{aligned}
E &= \frac{H(\vec{x})}{C} = \frac{\sum_i H(x_i)}{C} \frac{H(\vec{x})}{\sum_i H(x_i)} \\
&= \boxed{\frac{\sum_i H(x_i)}{C}} \frac{\sum_i H(x_i) - I(\vec{x})}{\sum_i H(x_i)}
\end{aligned}
$$

channel usage efficiency

optic nerve has a channel capacity C

biology: efficient coding [Attneave, 1954; Barlow, 1961]



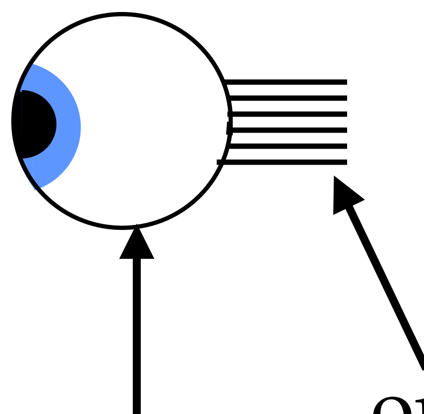
coding efficiency [Attick, 1991]

$$E = \frac{H(\vec{x})}{C} = \frac{\sum_i H(x_i)}{C} \frac{H(\vec{x})}{\sum_i H(x_i)}$$

$$= \underbrace{\boxed{\frac{\sum_i H(x_i)}{C}}}_{\text{channel usage efficiency}} \underbrace{\boxed{\frac{\sum_i H(x_i) - I(\vec{x})}{\sum_i H(x_i)}}}_{\text{code efficiency}}$$

optic nerve has a channel capacity C

biology: efficient coding [Attneave, 1954; Barlow, 1961]



optic nerve has a channel capacity C

coding efficiency [Attick, 1991]

$$E = \frac{H(\vec{x})}{C} = \frac{\sum_i H(x_i)}{C} \frac{H(\vec{x})}{\sum_i H(x_i)}$$

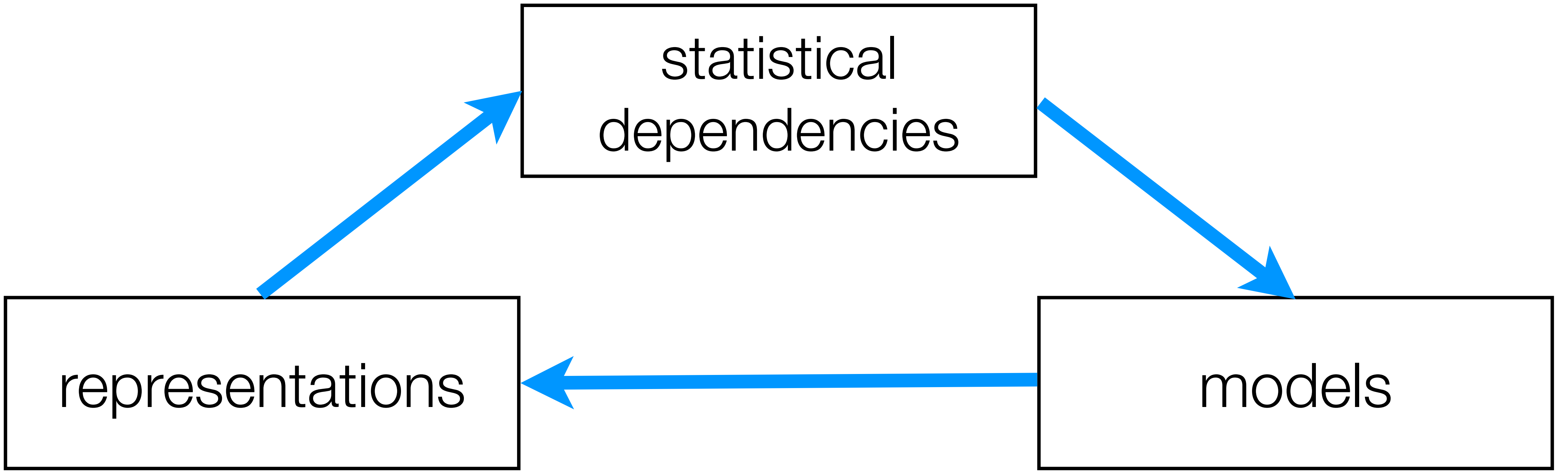
$$= \underbrace{\boxed{\frac{\sum_i H(x_i)}{C}}}_{\text{channel usage efficiency}} \underbrace{\boxed{\frac{\sum_i H(x_i) - I(\vec{x})}{\sum_i H(x_i)}}}_{\text{code efficiency}}$$

efficient code
- match channel marginals
- independent

dependency reduction

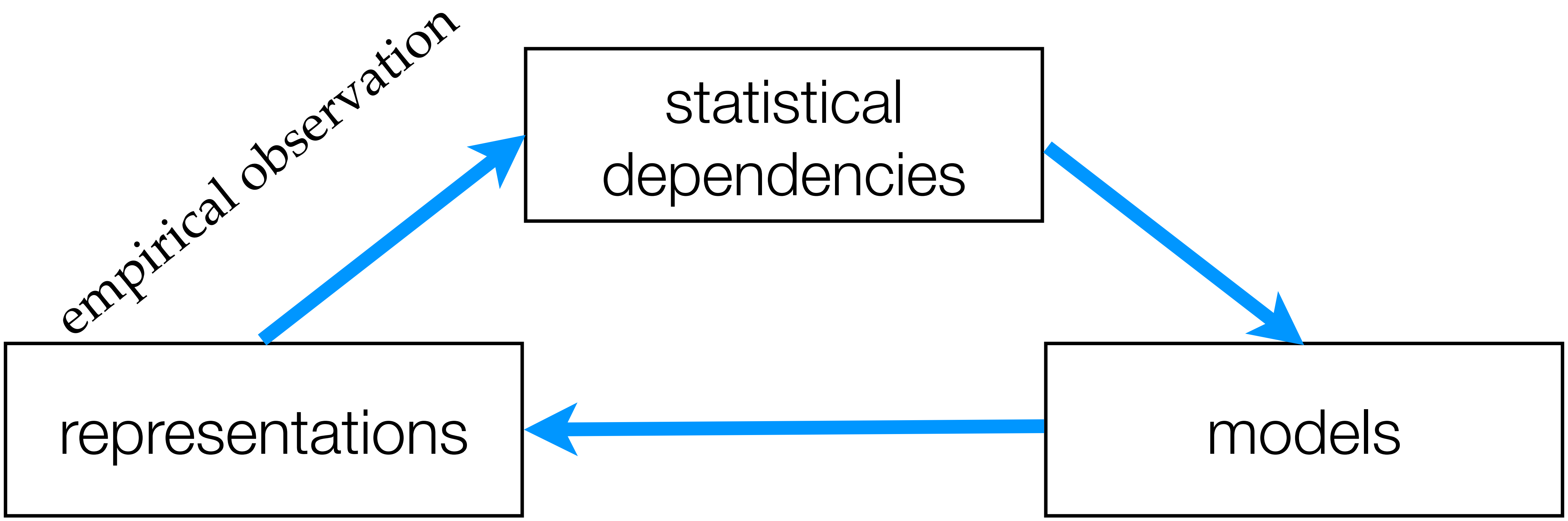
- simplify modeling
  - if components of x are independent, the joint density of x can be expressed as the product of marginals on each component
  - *dimensionality reduction in the parameter space*
- parallel manipulation
  - if components of x are independent, each component can be processed independently
- parallel sampling

closed loop



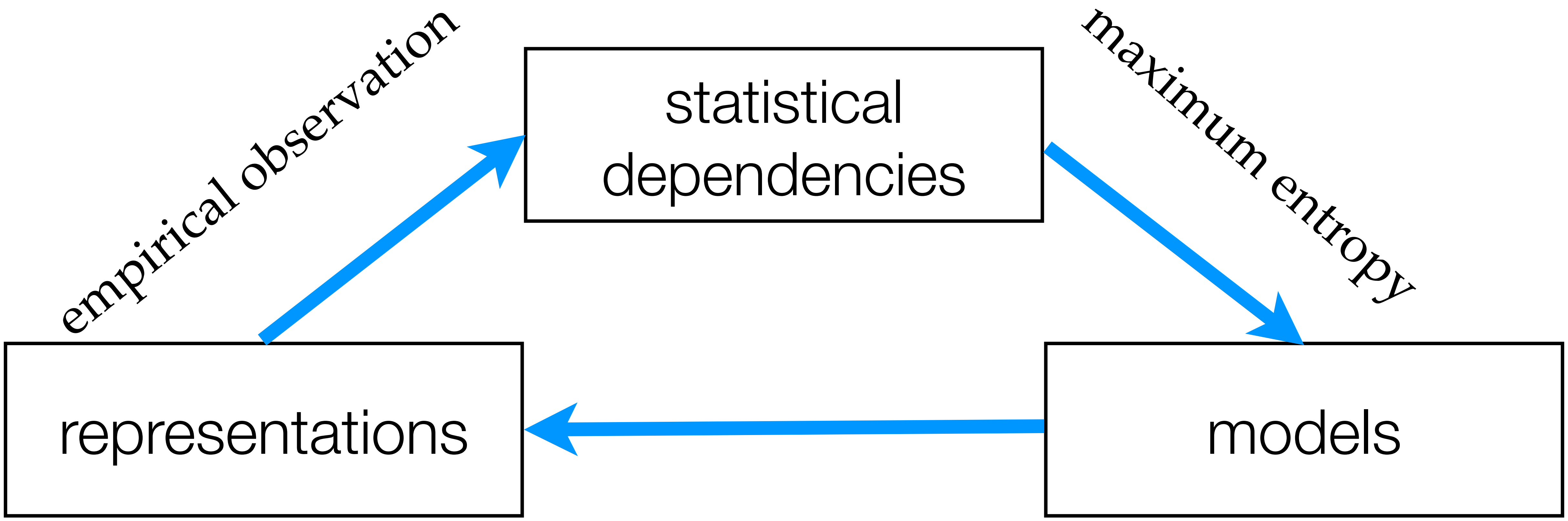
key question: where to put the complexity?

closed loop



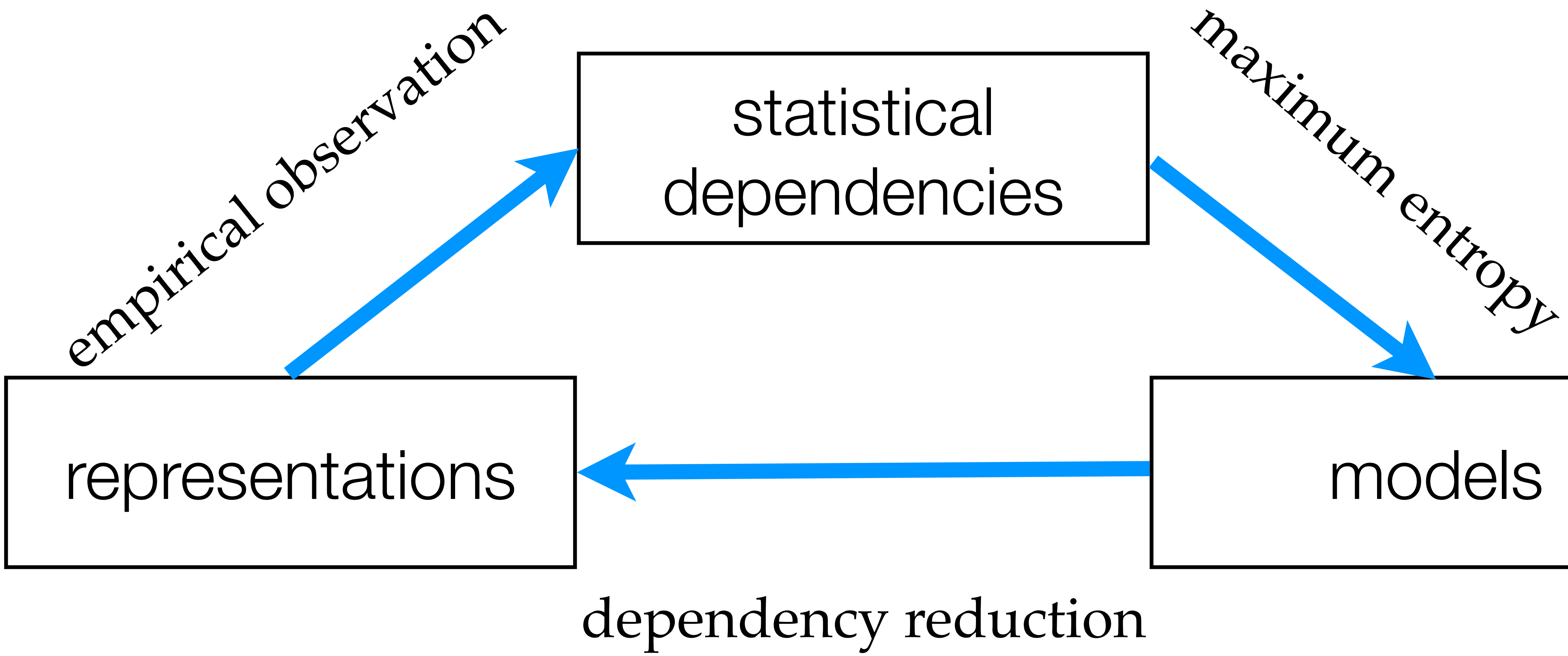
key question: where to put the complexity?

closed loop



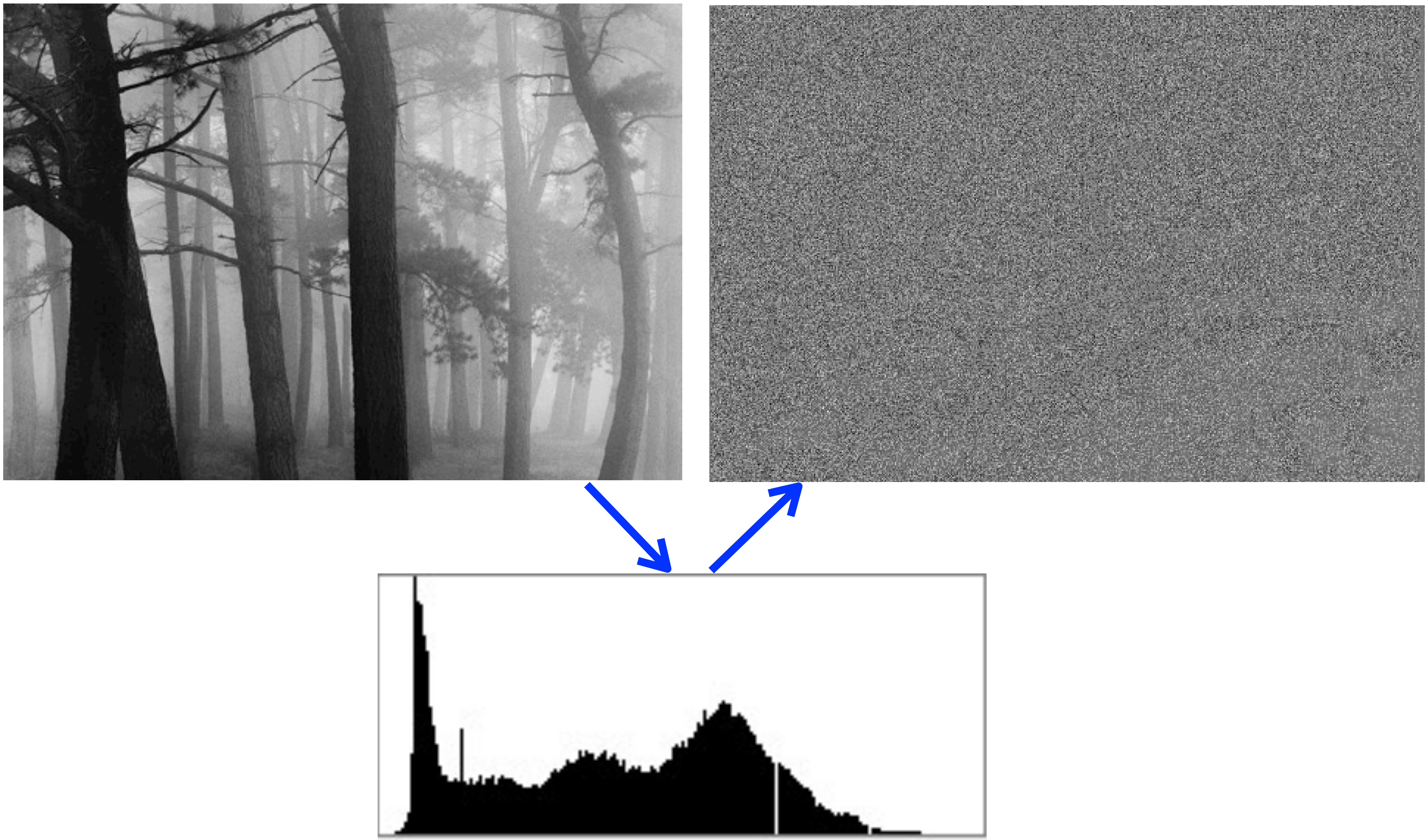
key question: where to put the complexity?

closed loop

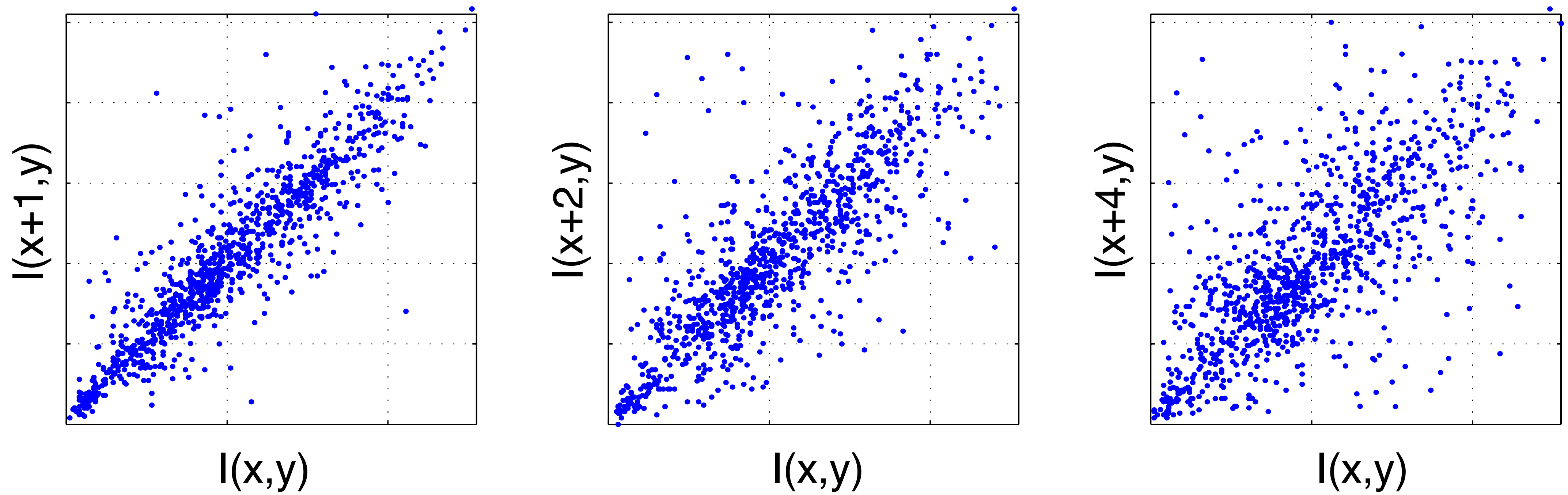


key question: where to put the complexity?

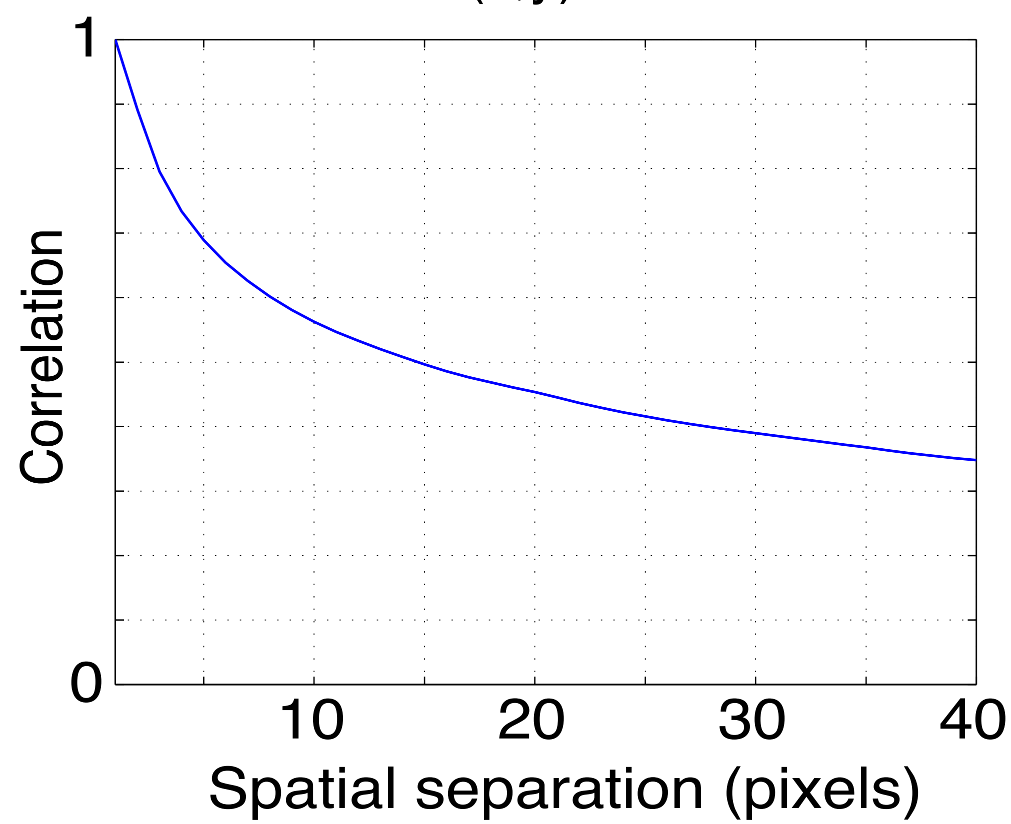
pixel - marginal distributions



pixel - second order correlation



second-order correlation

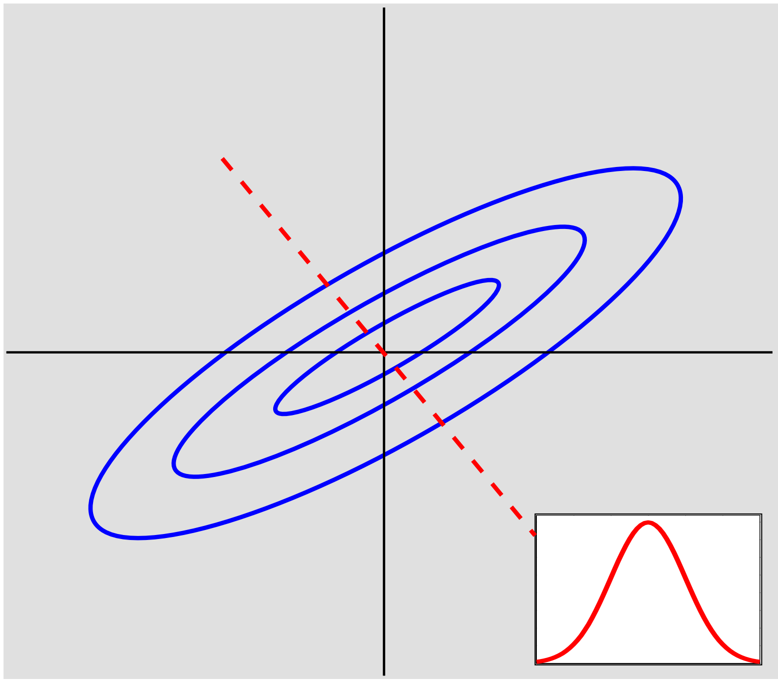


Gaussian model

- assume zero mean and match second order statistics
  - covariance matrix $\Sigma = E(\vec{x}\vec{x}^T)$
- maximum entropic model is Gaussian

$$p(\vec{x}) \propto \exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x}\right)$$

- extension: Gaussian Markov random field for large images
  - specified by the inverse covariance (precision/structure) matrix



Bayesian denoising

- additive white Gaussian noise $\vec{y} \quad = \quad \vec{x} + \vec{w}$
  - likelihood $$p(\vec{y}|\vec{x}) \quad \propto \quad \exp[-\|\vec{y} - \vec{x}\|^2/2\sigma_w^2]$$
- prior model

$$p(\vec{x}) \propto \exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x}\right)$$

- posterior density (another Gaussian)

$$p(\vec{x}|\vec{x}) \propto \exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x} - \frac{\|\vec{x} - \vec{y}\|^2}{2\sigma_w^2}\right)$$

- solution: Wiener filter

$$\vec{x}_{\mathrm{MAP}} = \vec{x}_{\mathrm{MMSE}} = \Sigma(\Sigma + \sigma_w^2 I)^{-1}\vec{y}$$

PCA representation

- Gaussians only have second-order dependency

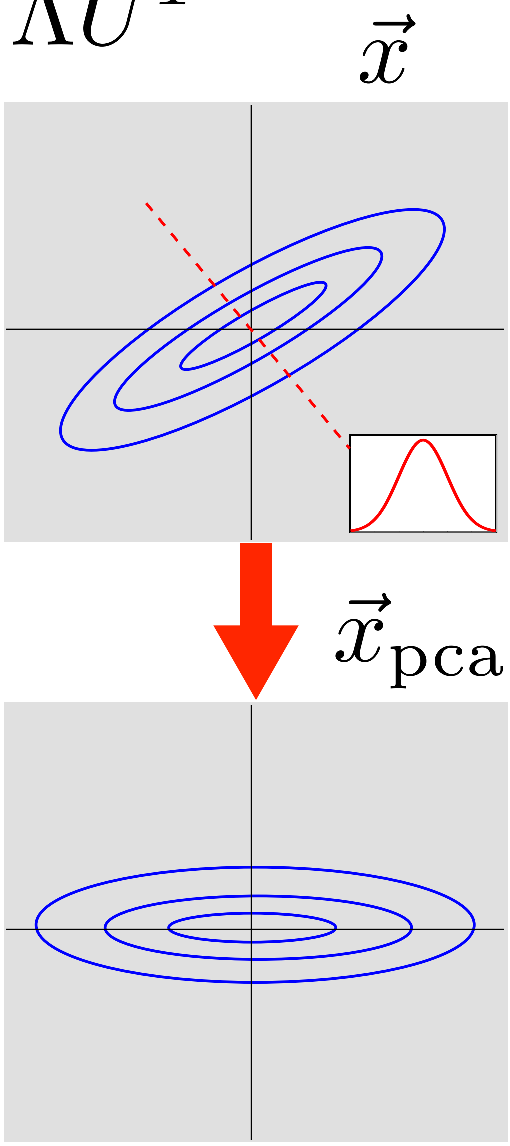
$$I(\vec{x}) \propto \sum_{i=1}^{d} \log(\Sigma)_{ii} - \log\det(\Sigma)$$

- minimum (independent) when Σ is diagonal
    - Hadamard's inequality
- a transform that ***diagonalizes*** Σ can eliminate all dependencies (second-order)
- result: principal component analysis (PCA)

PCA

- eigen-decomposition of covariance $\Sigma = U \Lambda U^T$
  - U: orthonormal matrix (rotation)
  - Λ: diagonal matrix of eigenvalues

$$\vec{x}_{\text{pca}} = U^T \vec{x}$$

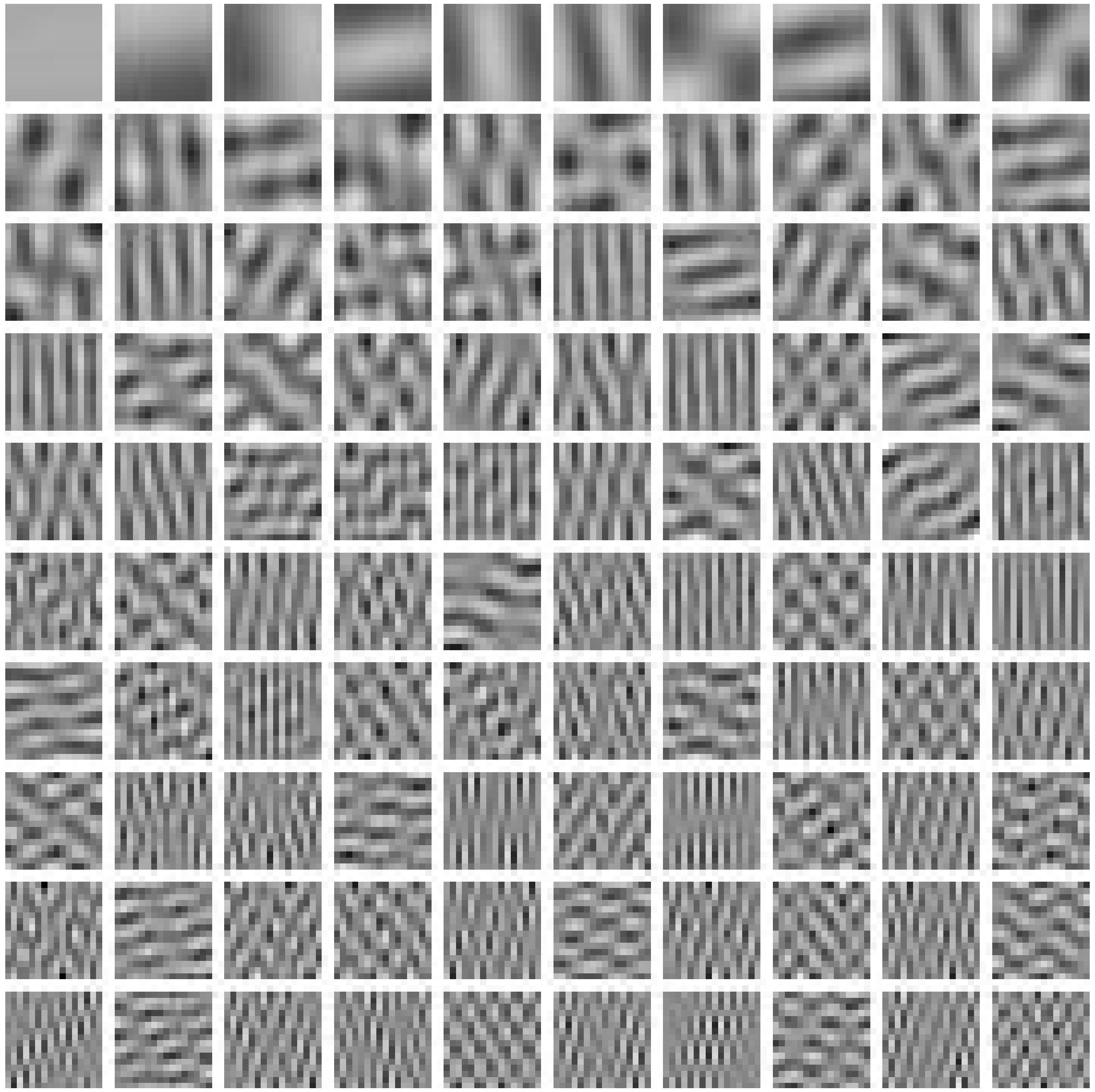


  - covariance becomes diagonal

$$\begin{aligned} & E\{\vec{x}_{\text{pca}} \vec{x}_{\text{pca}}^T\} \\ = \quad & U^T E\{\vec{x}\vec{x}^T\} U \\ = \quad & U^T U \Lambda U^T U = \Lambda \end{aligned}$$

  - independent Gaussian, if x is Gaussian
  - no correlation, if x is from arbitrary source

PCA basis from image patches



U

whitening

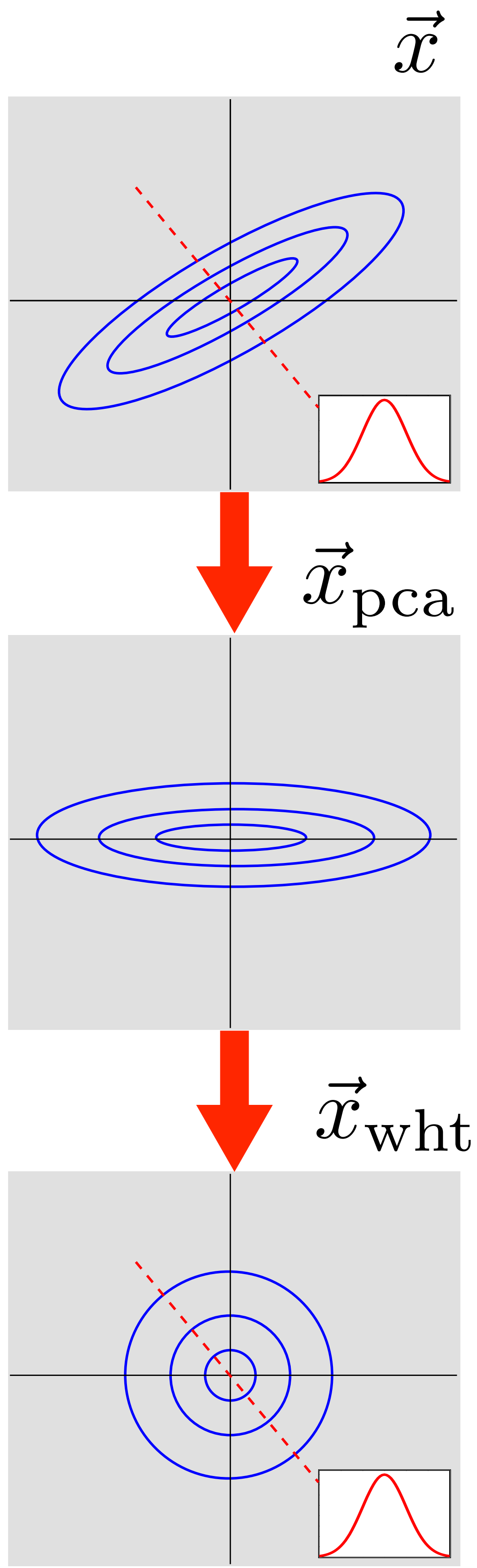
- making the PCA representation isotropic in variances

$$\vec{x}_{\text{wht}} = V\Lambda^{-\frac{1}{2}}\vec{x}_{\text{pca}} = V\Lambda^{-\frac{1}{2}}U^T\vec{x}$$

  - V is an orthonormal matrix (rotation)

$$\begin{aligned} & E\{\vec{x}_{\text{wht}}\vec{x}^T_{\text{wht}}\} \\ = \quad & V\Lambda^{-1/2}U^T E\{\vec{x}\vec{x}^T\}U\Lambda^{-1/2}V^T \\ = \quad & V\Lambda^{-1/2}U^T U\Lambda U^T U\Lambda^{-1/2}V^T = I \end{aligned}$$

  - isotropic Gaussian, if x is Gaussian
  - whitened, if x is from arbitrary source
  - whitening transform is not unique



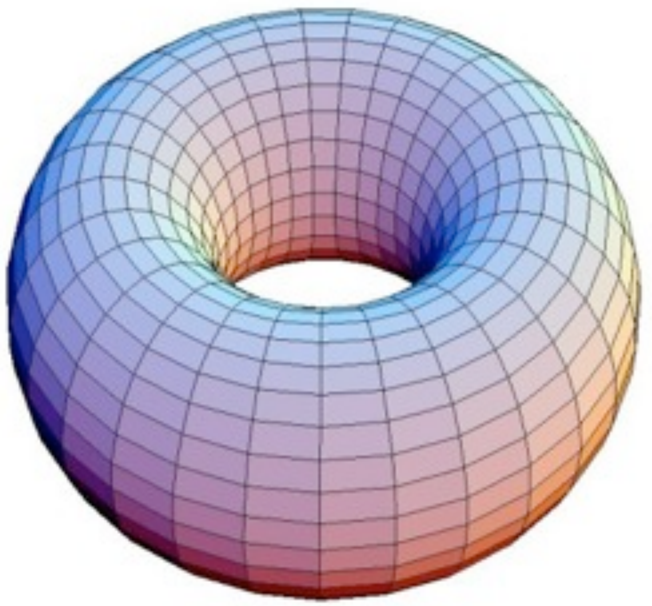
ZCA whitening

- zero-phase component analysis [Bell & Sejnowski, 1996]

$$\vec{x}_{\mathrm{zca}} = U\Lambda^{-\frac{1}{2}}U^T\vec{x}$$

  - choose V = U, the result is a symmetric linear transform
  - minimizing squared distortion between data and representation
    - minimum wiring length principle [Vincent & Baddeley, 2003]
  - similar to the center-surround receptive fields for retina gangalion cells

fixed transform



- translation invariance

$$\mathrm{cov}(I(x,y), I(x+\triangle x, y+\triangle y)) = \mathrm{cov}(I(0,0), I(\triangle x, \triangle y))$$

- circular boundary handling

- covariance matrix Σ is a ***circulant matrix***
- example:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \\ 4 & 5 & 1 & 2 & 3 \\ 3 & 4 & 5 & 1 & 2 \\ 2 & 3 & 4 & 5 & 0 \end{pmatrix}$$

Fourier representation

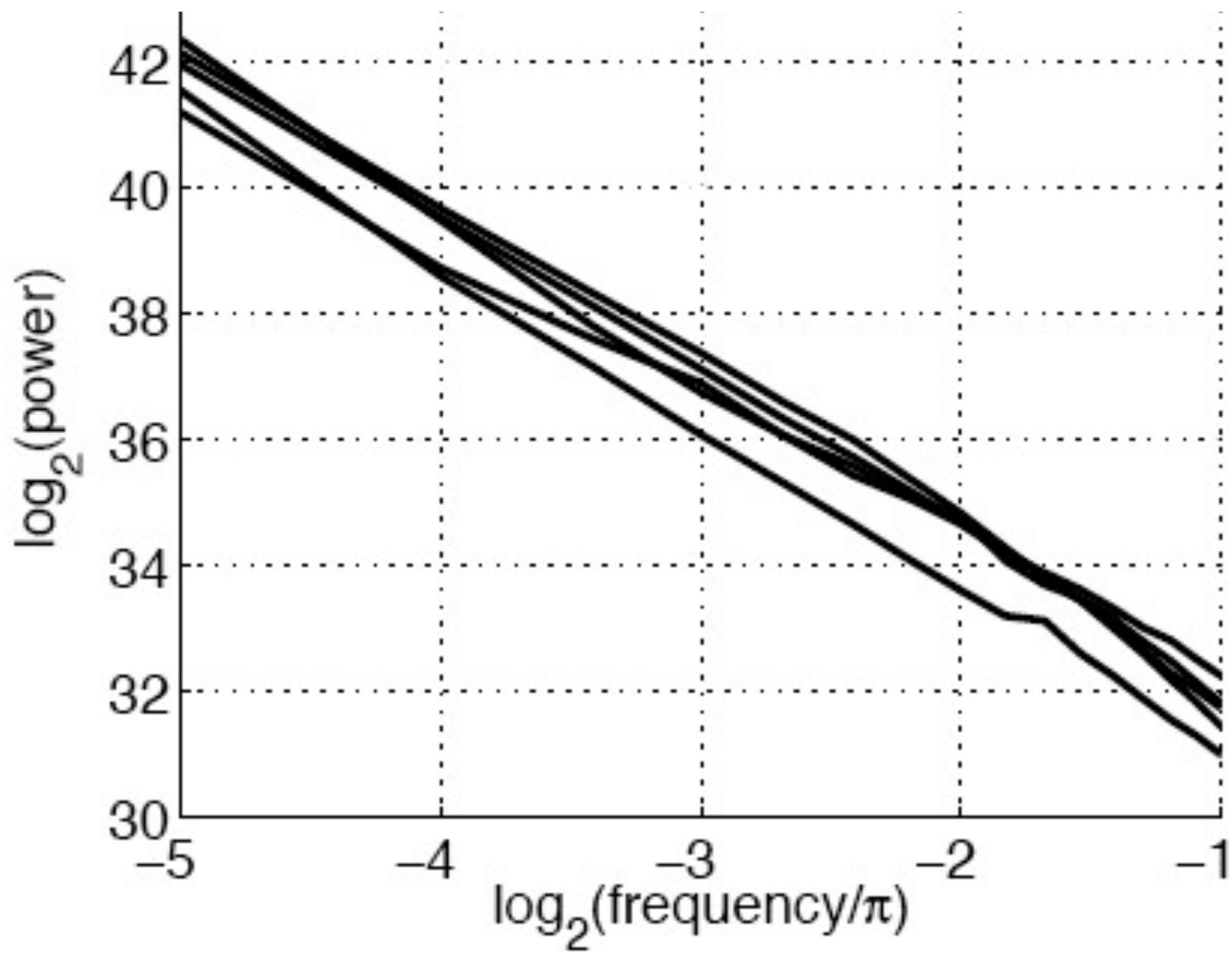
- Fourier transform diagonalizes the circulant covariance matrix
  - discrete Fourier transform basis are eigenvectors
  - Fourier transform of the circulant kernel are the eigenvalues

$$\Sigma = \mathrm{circ}(\vec{v}) = \mathcal{F}\ \mathrm{diag}(\mathcal{F}^* \vec{v})\mathcal{F}^*$$

- DFT is the eigen-system for translational invariant ensembles of images with circular boundary condition
  - question: why complex-valued?

Fourier - marginal

- spectral power



$$F(\omega) = \frac{A}{\omega^\gamma}$$

$$F(s\omega) = s^p F(\omega)$$

figure from [Simoncelli, 2005]

scale invariance of image variance

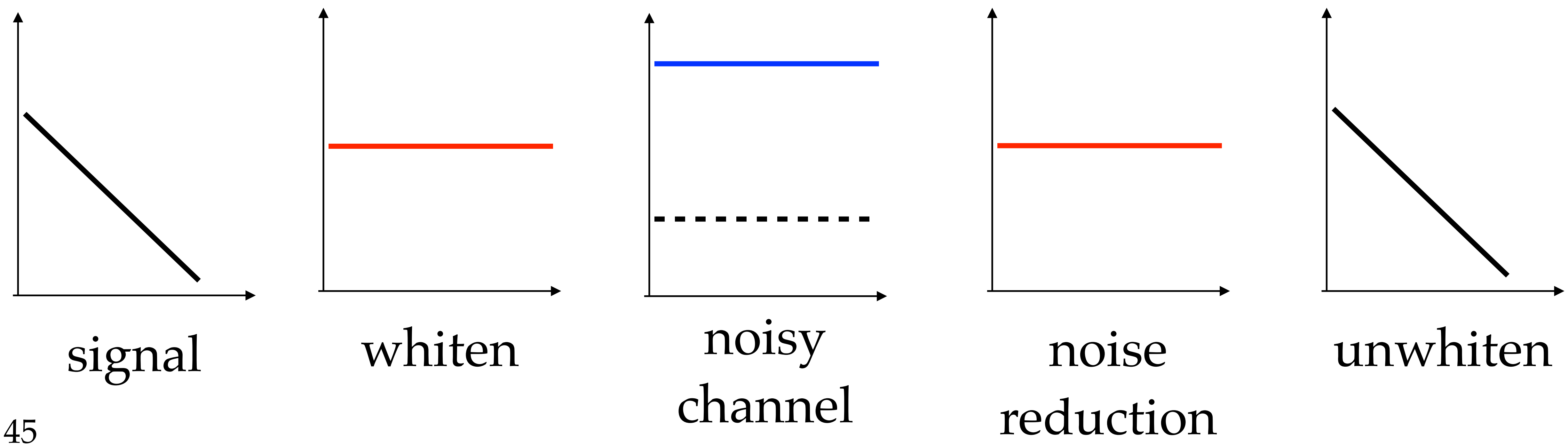
E()=1/4 E()

[Ritterman 52; DeRiugin 56; Field 87; Tolhurst 92; Ruderman/Bialek 94; ...]

applications

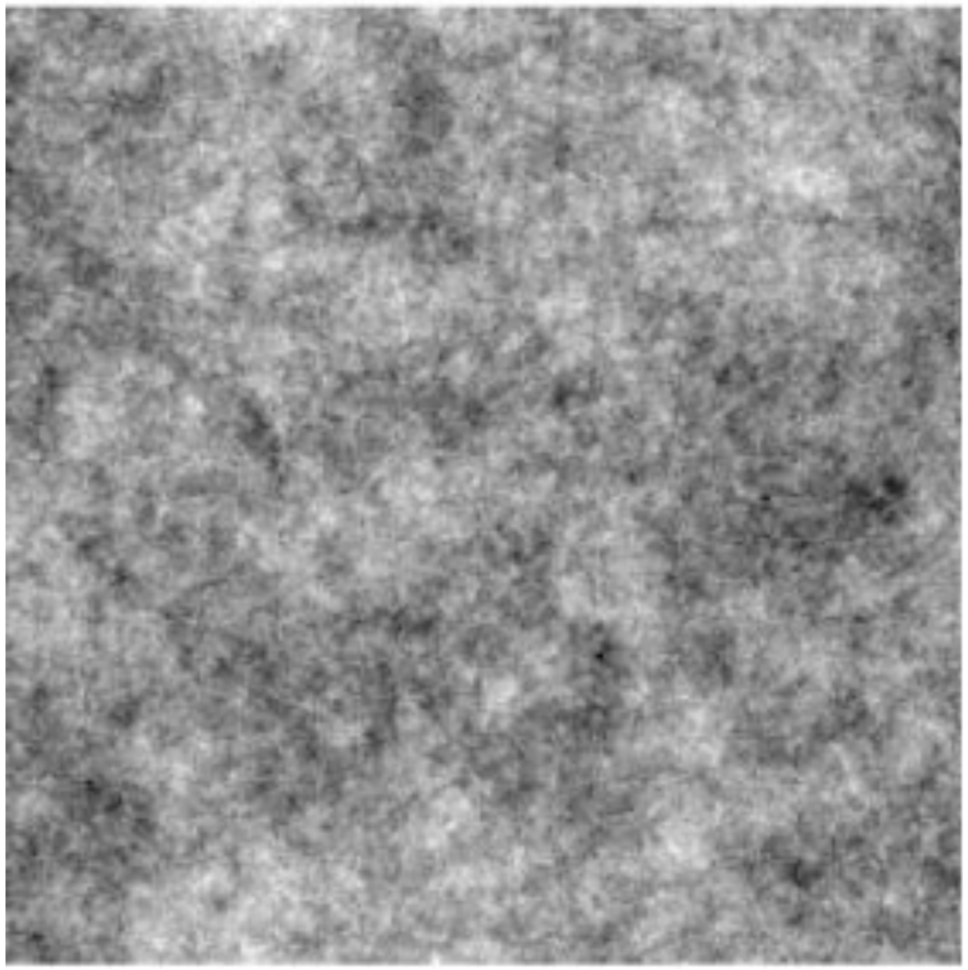
$$F(\omega) = \frac{A}{\omega^\gamma}$$

- denoising (Wiener filter in frequency domain)
- JPEG compression
- Dolby noise reduction



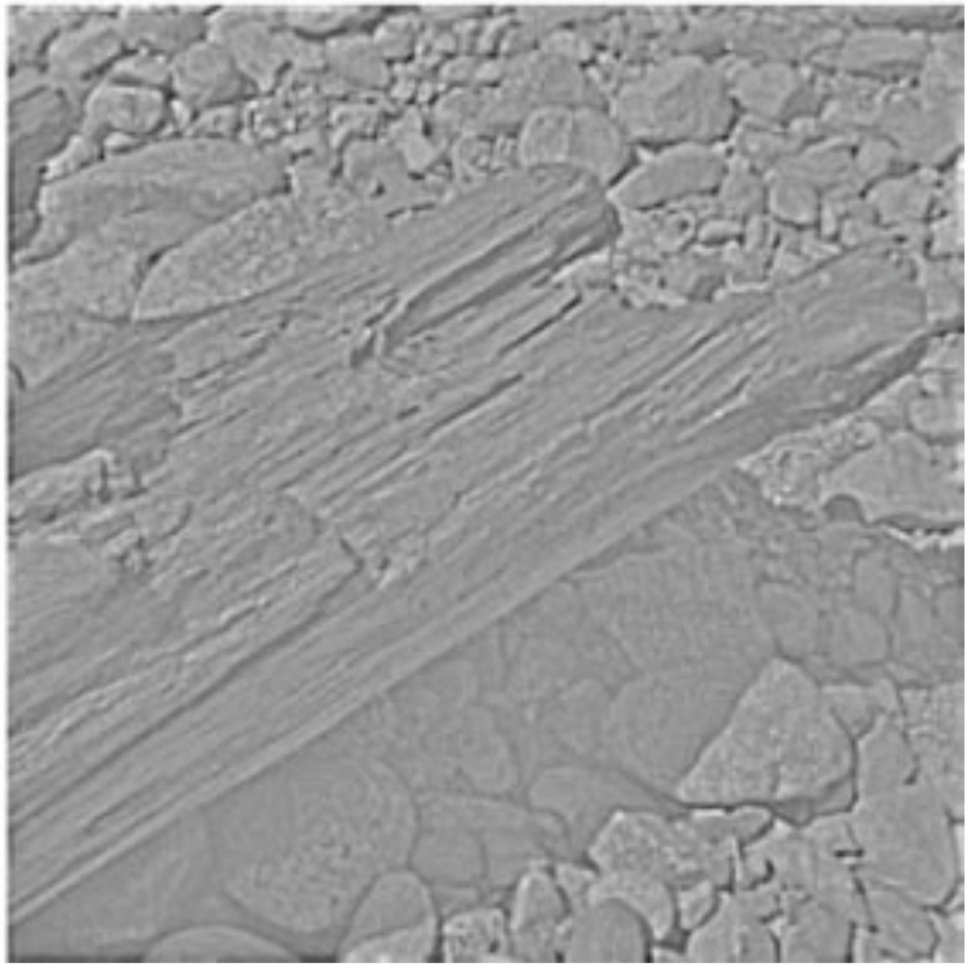
not sufficient

sample from power law
Gaussian sample



not natural image

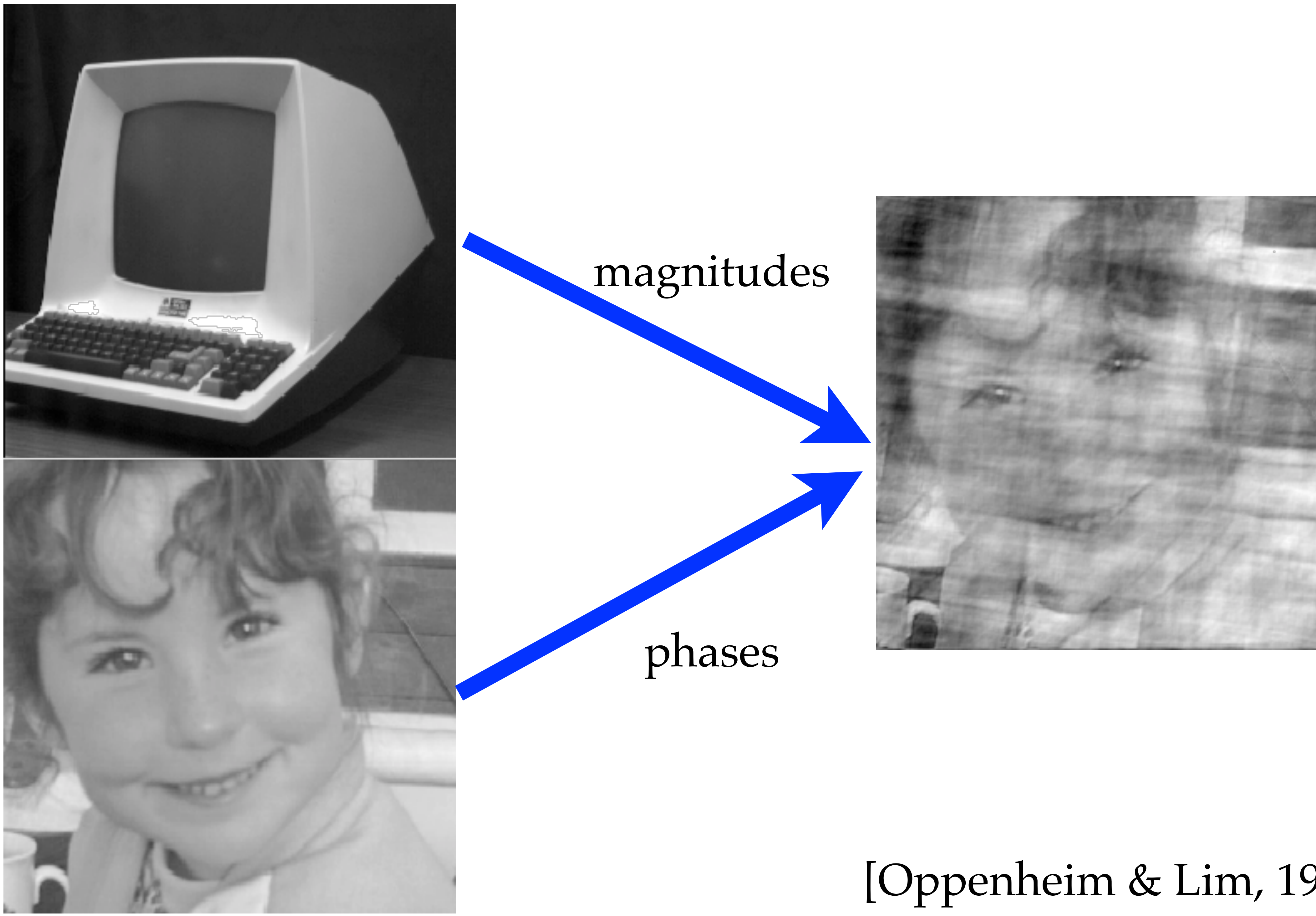
natural image
after whitening



not independent noise

[Simoncelli and Olshausen, 2001]

structures in phases



[Oppenheim & Lim, 1981]

dependency

$$I(\vec{x}) \quad = \quad \boxed{\sum_{k=1}^{d} \log(\Sigma_{kk}) - \log|\Sigma|}$$

second-order dependency

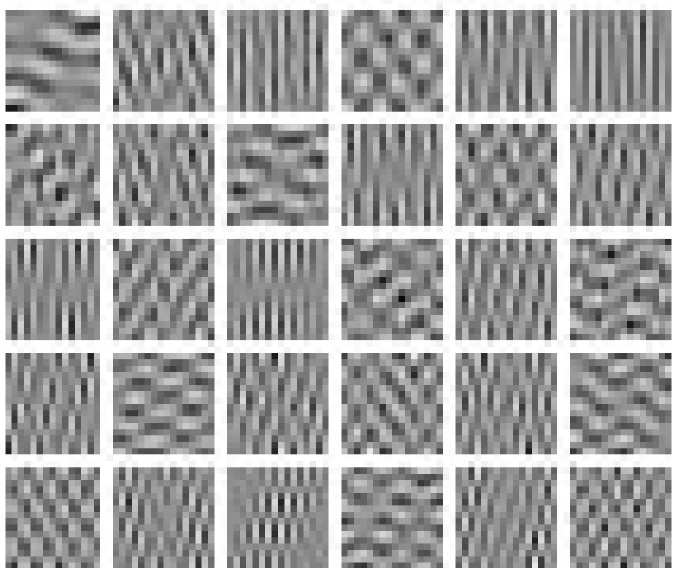
$$+ \quad \boxed{D_{\mathrm{KL}}\left(p(\vec{x}) \,\|\, \mathcal{G}(\vec{x})\right) - \sum_{k=1}^{d} D_{\mathrm{KL}}\left(p(x_k) \,\|\, \mathcal{G}(x_k)\right)}$$

higher-order dependency

summary

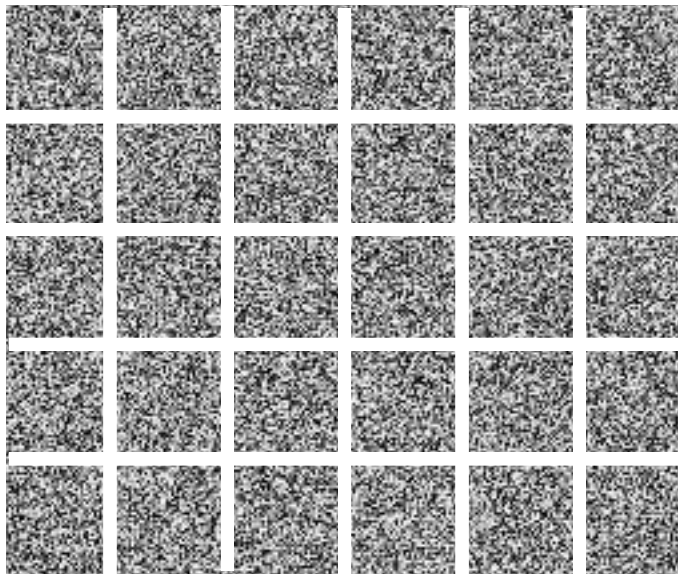
- pixel domain matching second-order statistics leads to Gaussian image models
- eliminating dependencies in Gaussian models leads to PCA/whitening based representations
- extending PCA to global image domain leads to frequency domain representations
- Gaussian model + PCA representations are not sufficient to model natural images
  - higher-order statistical dependencies not being captured

band pass filters



PCA

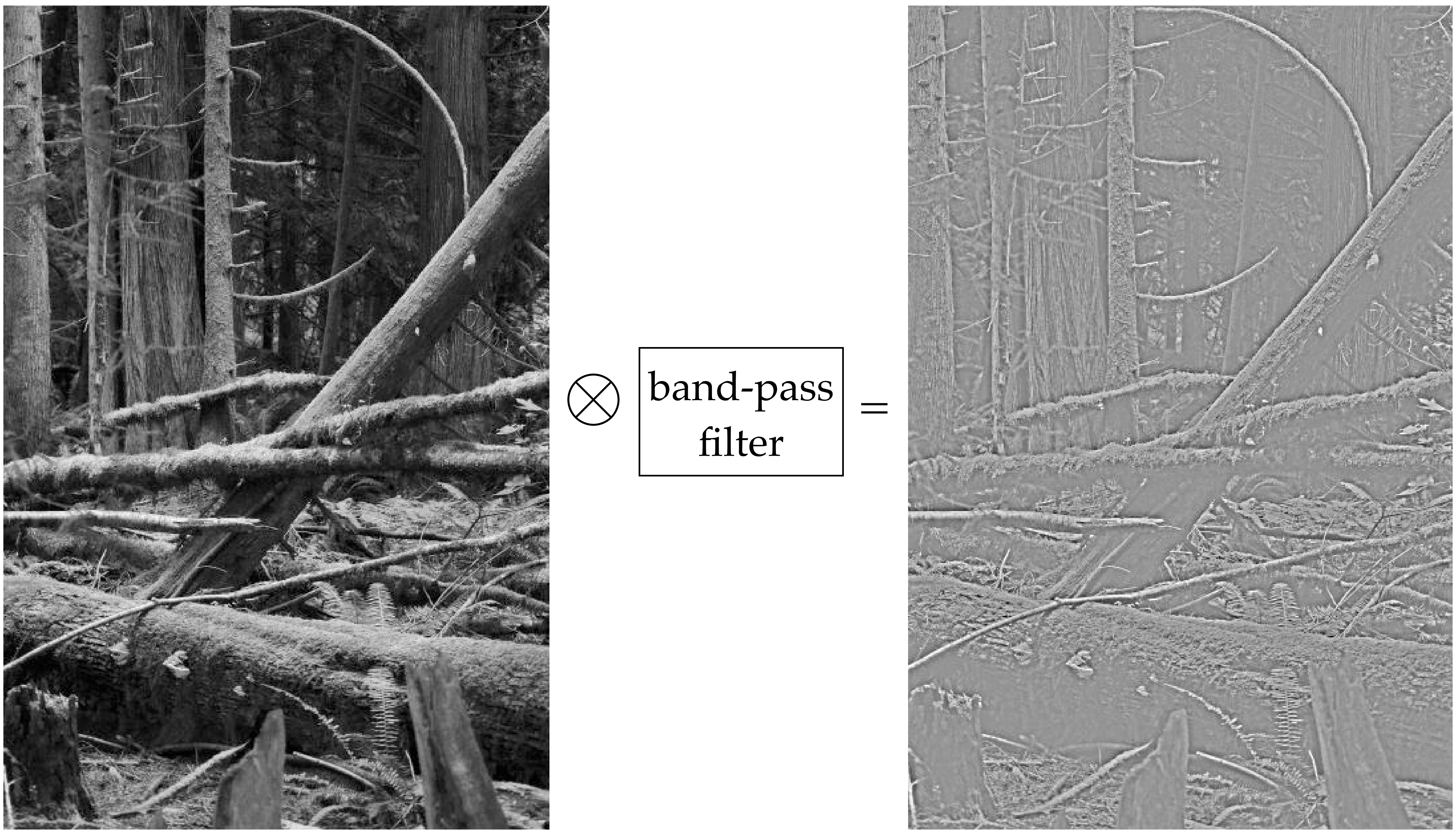
ZCA



random

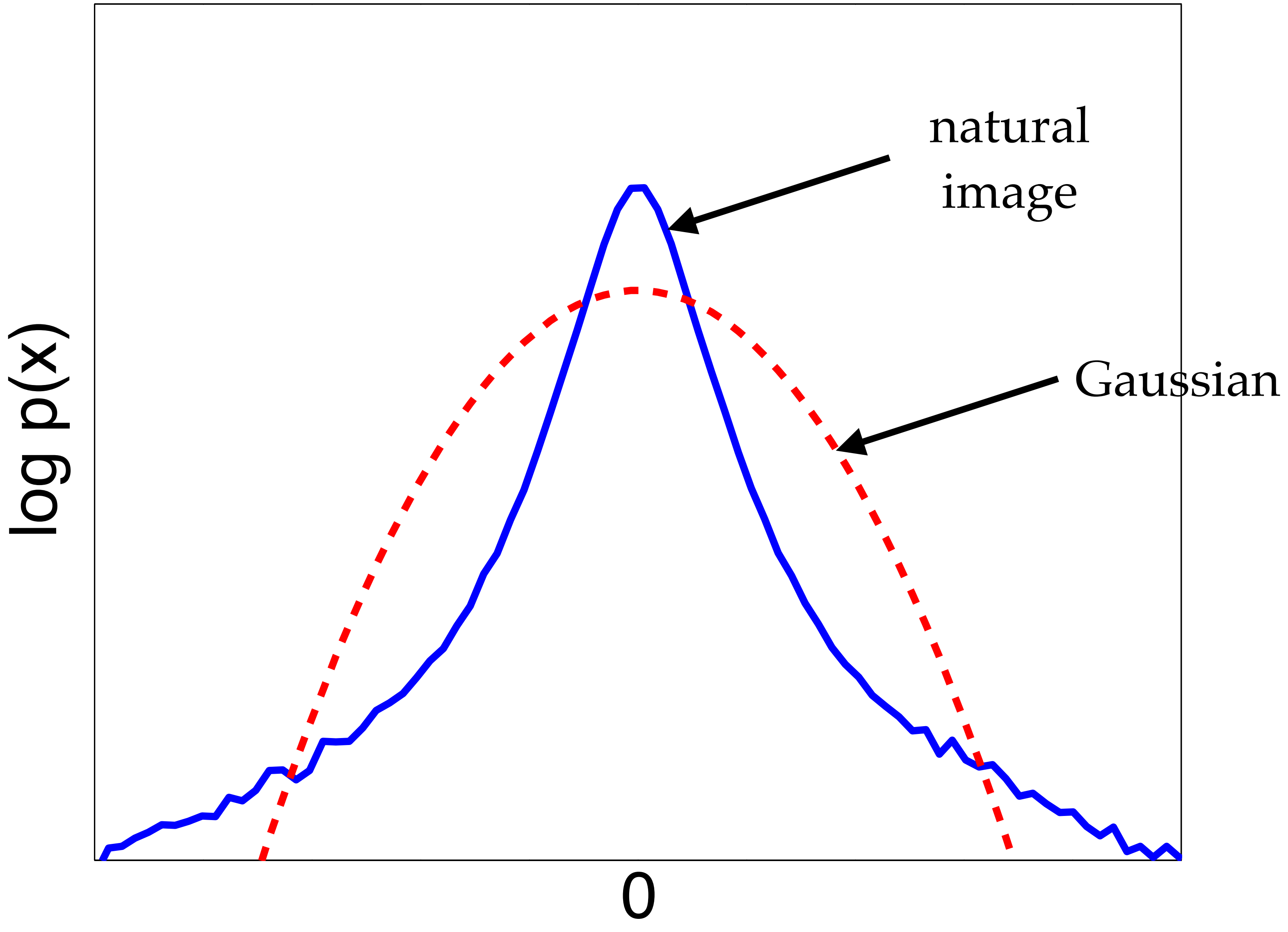
- localize in space and frequency
- reduce low-frequency components

bandpass filter domain



band-pass filter domain

- marginal density



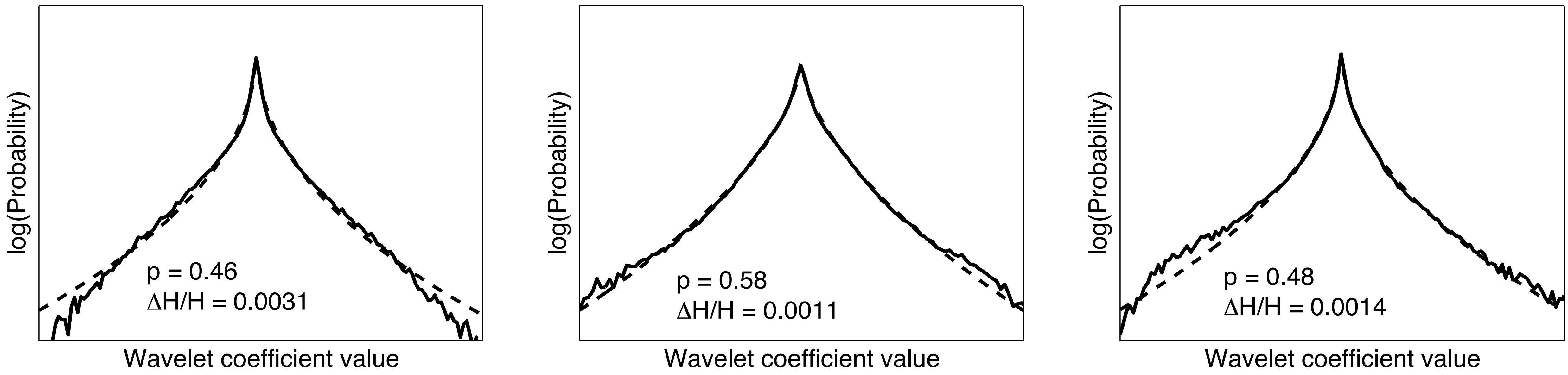
[Burt&Adelson 82; Field 87; Mallat 89; Daugman 89, ...]

marginal model

- well fit with generalized Gaussian

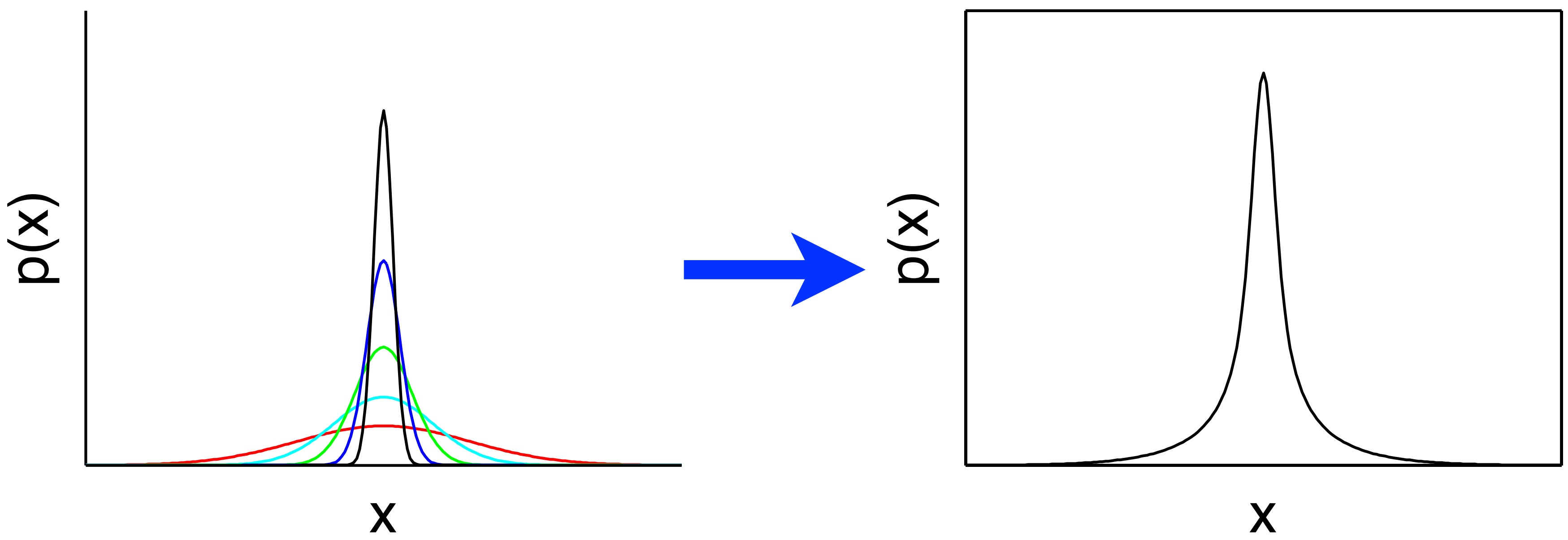
$$p(s) \propto \exp\left(-\frac{|s|^p}{\sigma}\right)$$

[Mallat 89; Simoncelli&Adelson 96; Moulin&Liu 99; …]



Gaussian scale mixtures

[Andrews & Mallows 74, Wainwright & Simoncelli, 99]



$$x = u\sqrt{z}$$

- u: zero mean Gaussian with unit variance
- z: positive random variable
- different p(z)

  generalized Gaussian, Student's t, Bessel's K, Cauchy, α-stable, etc

factorial model

enforce consistency on singleton marginal densities, i.e., $p(x_i) = q_i(x_i)$, maximum entropic density is the factorial density

$$H(\vec{x}) = \sum_i H(x_i) - I(\vec{x})$$

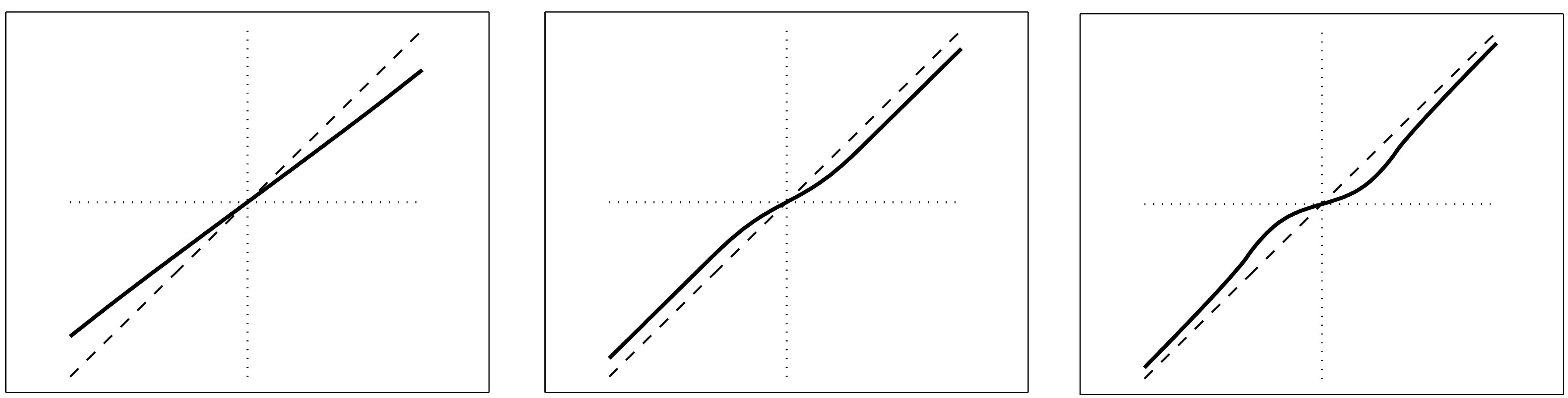
$$\Big\downarrow \text{maximum entropy}$$

$$p(\vec{x}) = \prod_{i=1}^{d} q_i(x_i)$$

Bayesian denoising - coring

$$\hat{x}(y) = \int dx\, \mathcal{P}_{x|y}(x|y)\, x = \frac{\int dx\, \mathcal{P}_{y|x}(y|x)\, \mathcal{P}_x(x)\, x}{\int dx\, \mathcal{P}_{y|x}(y|x)\, \mathcal{P}_x(x)}$$

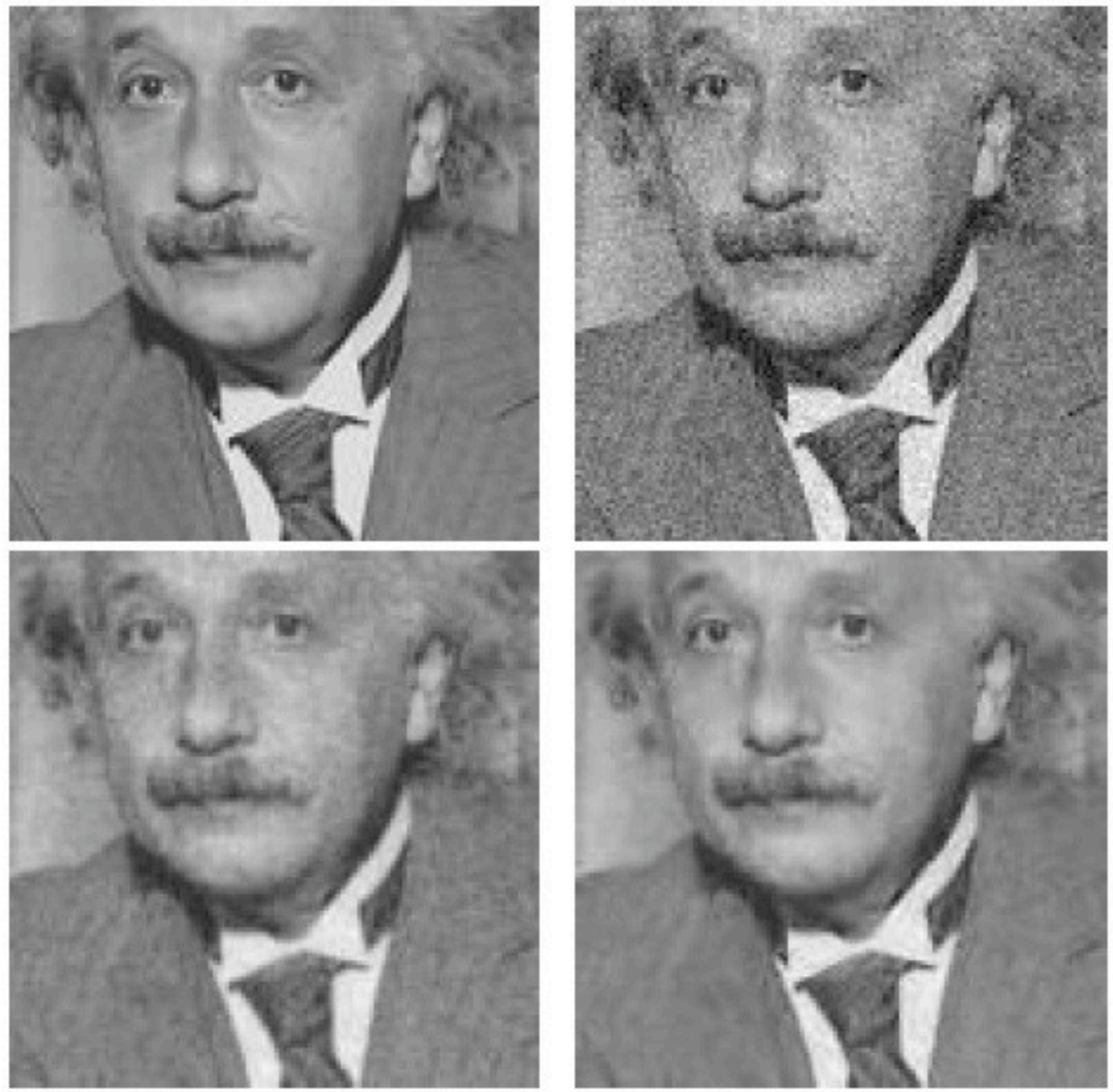
$$= \frac{\int dx\, \mathcal{P}_n(y-x)\, \mathcal{P}_x(x)\, x}{\int dx\, \mathcal{P}_n(y-x)\, \mathcal{P}_x(x)},$$



p = 2.0 *p = 1.0* *p = 0.5*

[Simoncelli & Adelson, '96]

original image noise (SNR = 9dB)

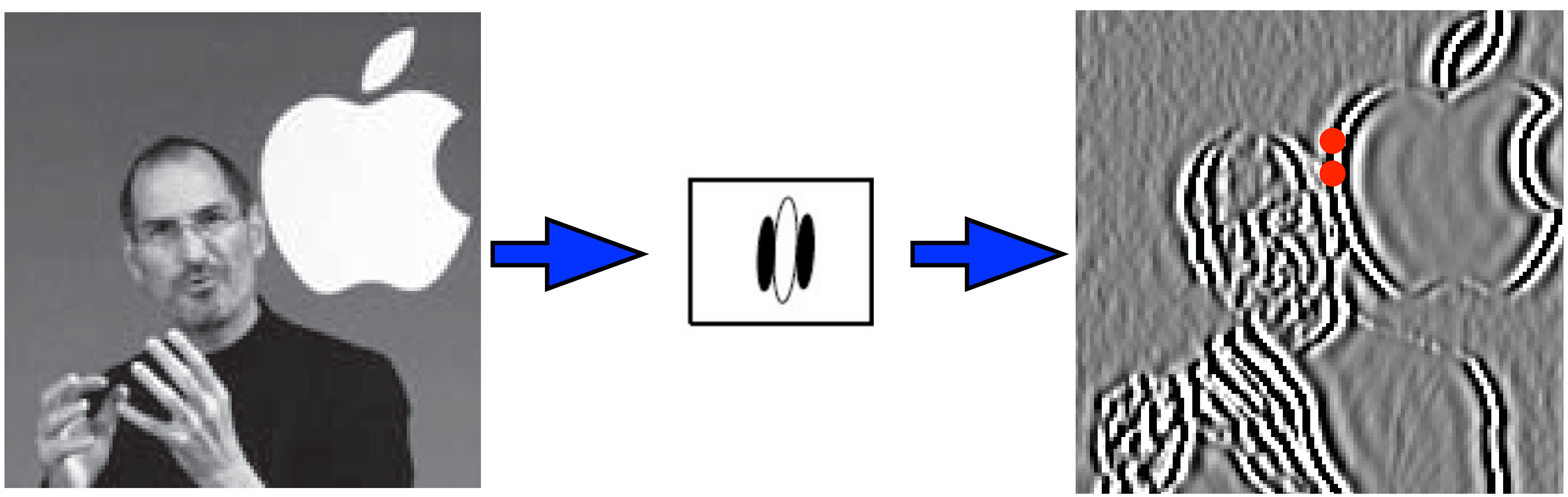


Wiener filter (11.88dB) coring (13.82dB)

[Simoncelli & Adelson, 1996]

dependencies

- band-pass filtered representationss of natural images are not independent



LTF model

- linearly transformed factorial (LTF)
- each component in x is a linear mixing of independent super-Gaussian sources, so they are not independent

$$\begin{aligned} \vec{x} &= A\vec{s} = \begin{pmatrix} | & \cdots & | \\ \vec{a}_1 & \cdots & \vec{a}_d \\ | & \cdots & | \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_d \end{pmatrix} \\ &= s_1\vec{a}_1 + \cdots + s_d\vec{a}_d \end{aligned}$$

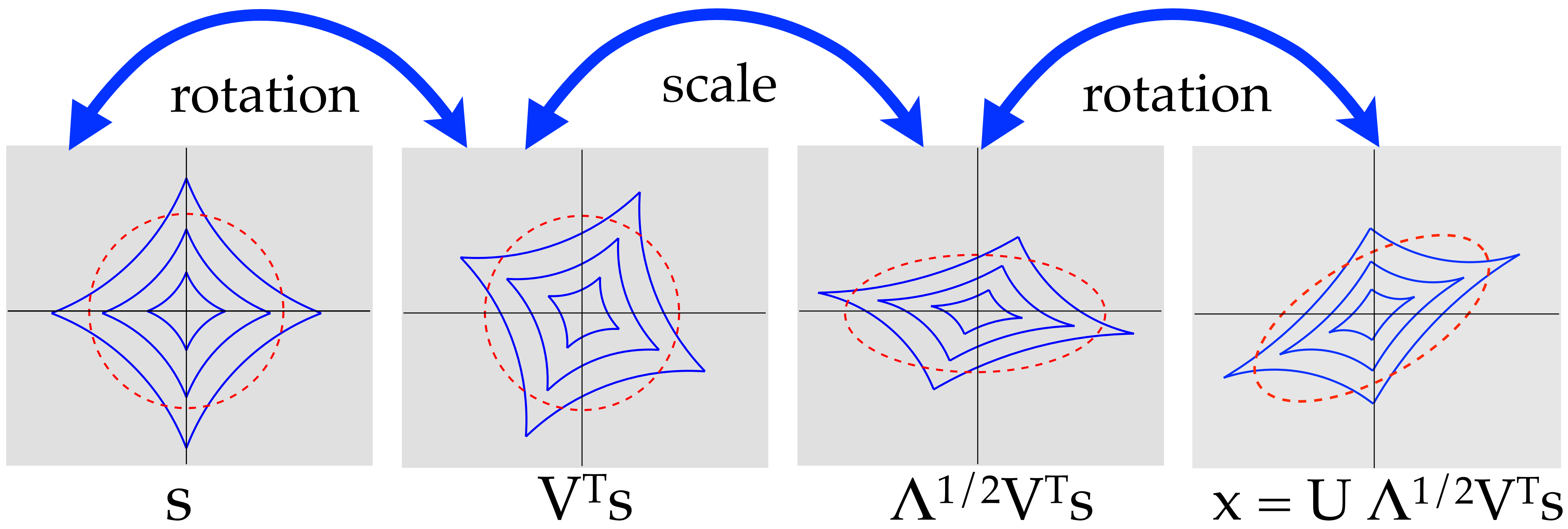
$$p(\vec{s}) = \prod_{i=1}^{d} p(s_i)$$

- A is an invertible linear transform (basis), A^{-1} are the encoding transform

LTF model - generative view

- SVD of matrix A: $A = U\Lambda^{1/2}V^T$
  - U,V: orthonormal matrices (rotation)

  - Λ: diagonal matrix
    $(\Lambda_{ii})^{1/2} \geq 0$ -- singular value



representation

- independent component analysis (ICA)
  [Comon 94; Cardoso 96; Bell/Sejnowski 97; …]
  - many different implementations
    - JADE, InfoMax, FastICA, etc.
- interpretation using SVD

$$\vec{x}_{\mathrm{ica}} = A^{-1}\vec{x} = V\Lambda^{-1/2}U^T\vec{x}$$

  - U and Λ obtained from PCA

independent
components
are decorrelated

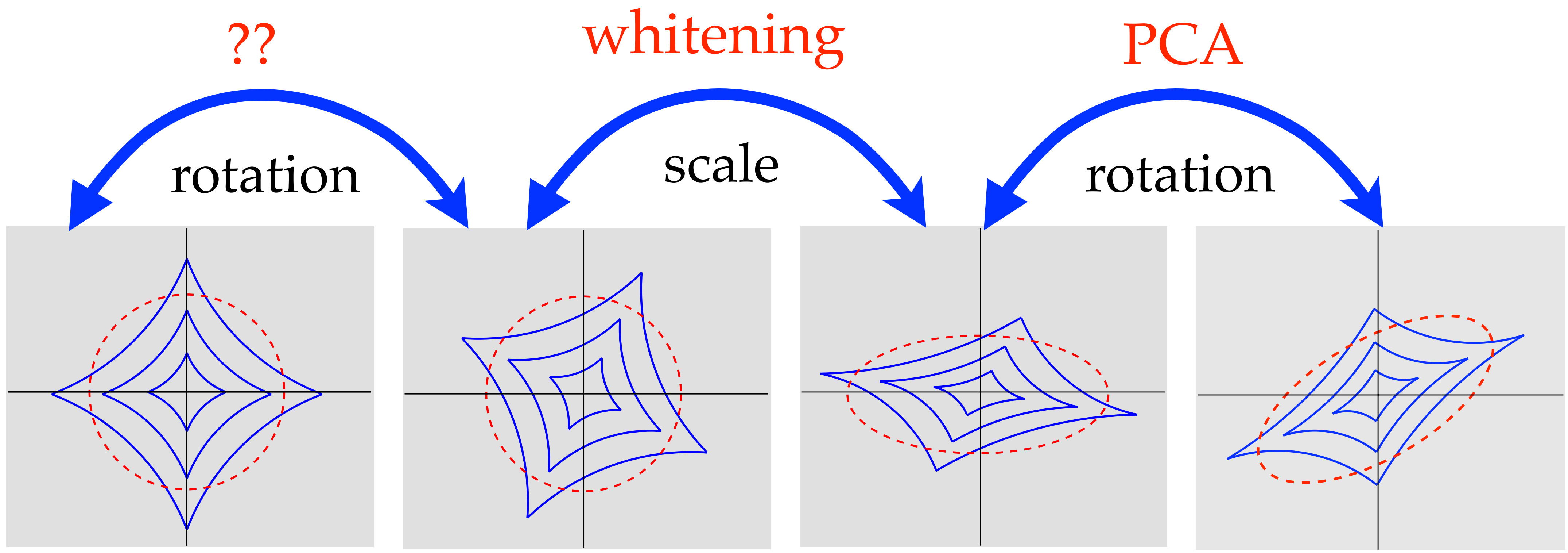
$$\begin{aligned} E\{\vec{x}\vec{x}^T\} &= A\boxed{E\{\vec{x}_{\mathrm{ica}}\vec{x}_{\mathrm{ica}}^T\}}A^T \\ &= U\Lambda^{1/2}V^T I V\Lambda^{1/2}U^T \\ &= U\Lambda U^T \end{aligned}$$

ICA

- ICA can be seen as a whitening operation

$$\vec{x}_{\mathrm{ica}} = A^{-1}\vec{x} = V\Lambda^{-1/2}U^T\vec{x}$$

- how to find the last rotation V



search for the last rotation in ICA

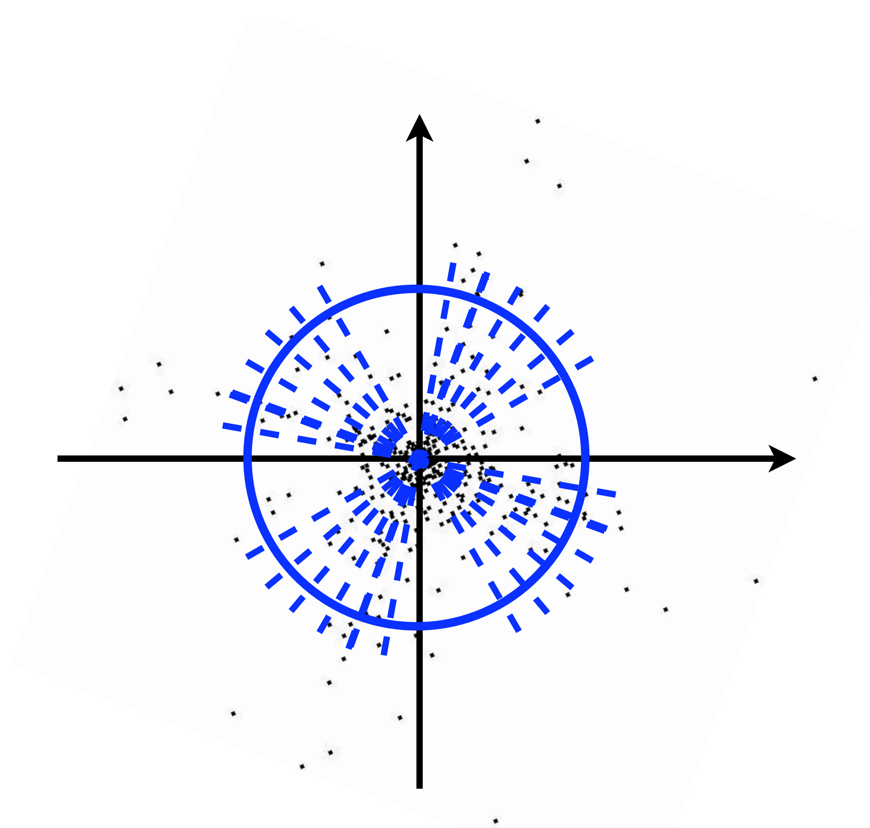
- minimizing multi-information

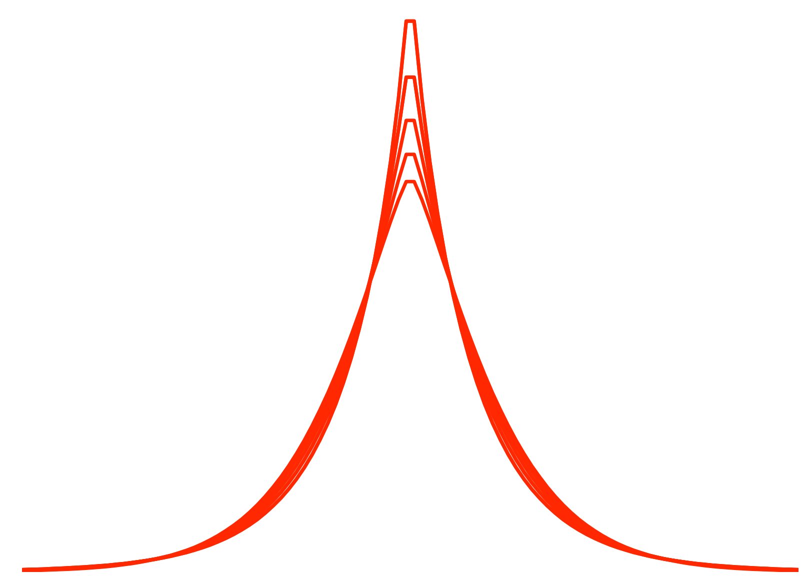
$$I(\vec{x}) = \boxed{\sum_k H(x_k)} - \boxed{H(\vec{x})}$$

minimize singleton entropy

not changed by rotation

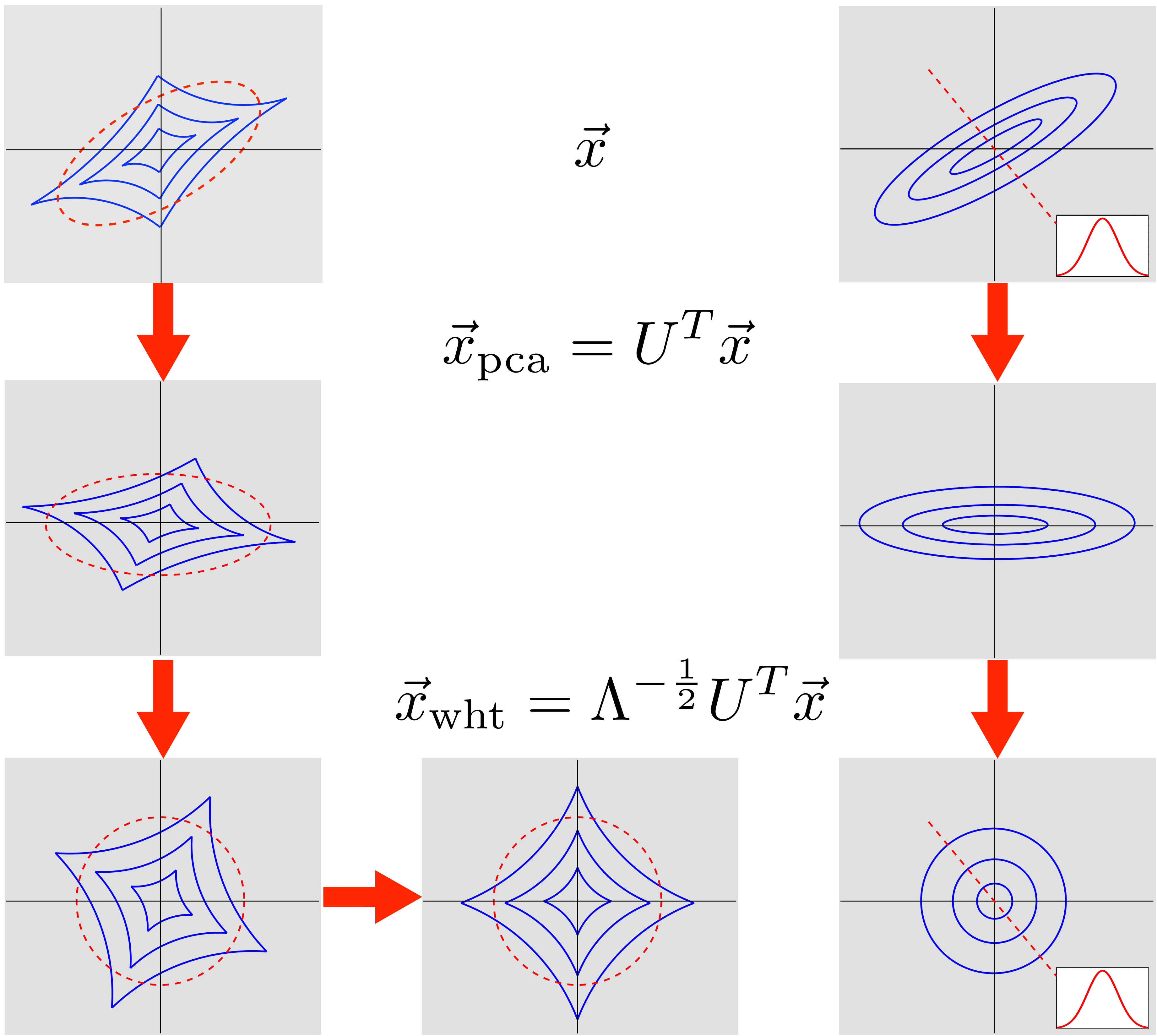
- for super-Gaussian densities, lower kurtosis suggests lower entropy





ICA / whitening

PCA / whitening



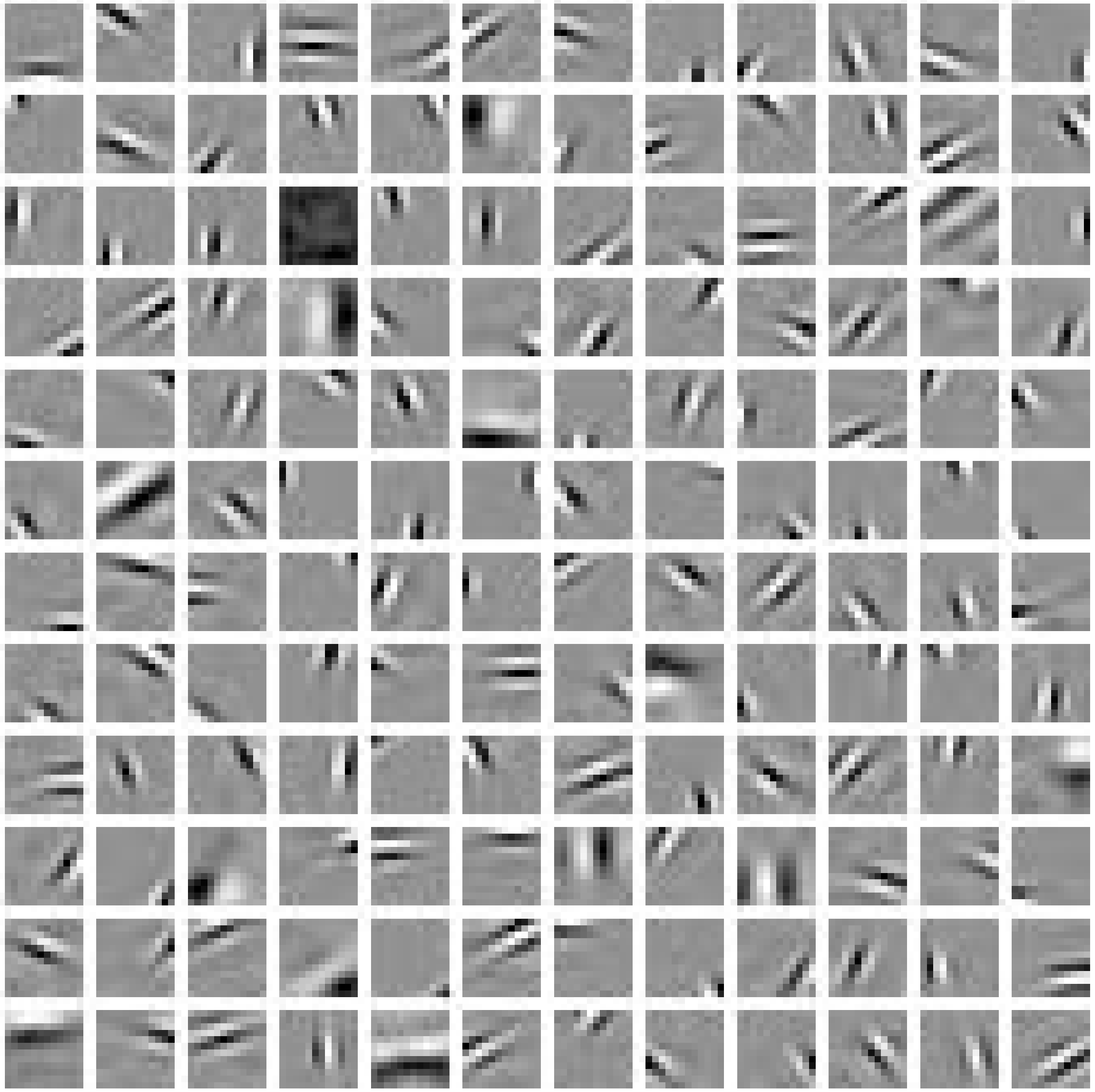
$$\vec{x}$$

$$\vec{x}_{\text{pca}} = U^T \vec{x}$$

$$\vec{x}_{\text{wht}} = \Lambda^{-\frac{1}{2}} U^T \vec{x}$$

$$\vec{x}_{\text{ica}} = V \Lambda^{-\frac{1}{2}} U^T \vec{x}$$

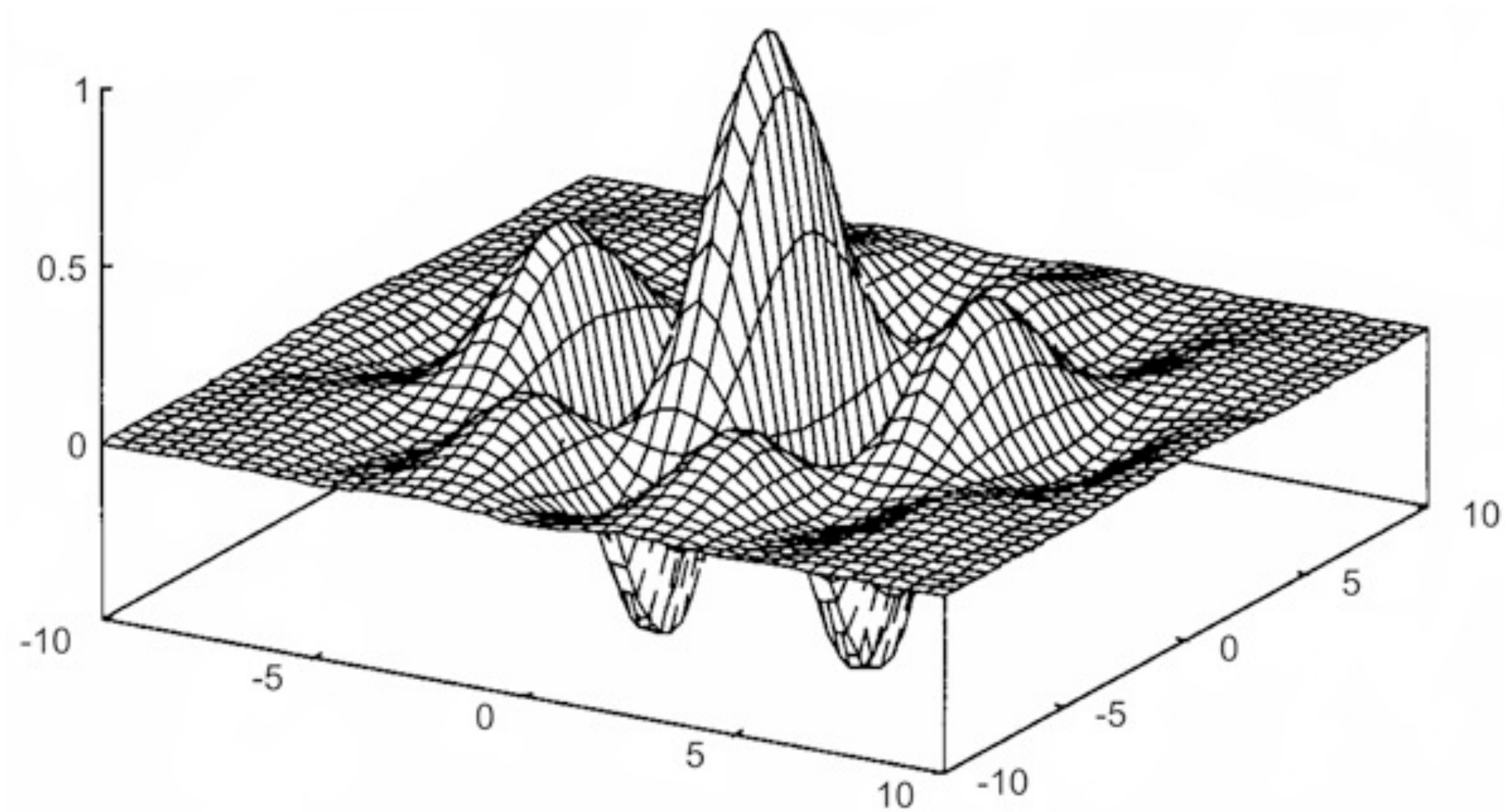
ICA basis from image patches



similar to the receptive field of V1 simple cells
[Olshausen & Field 1996, Bell & Sejnowski 1997]

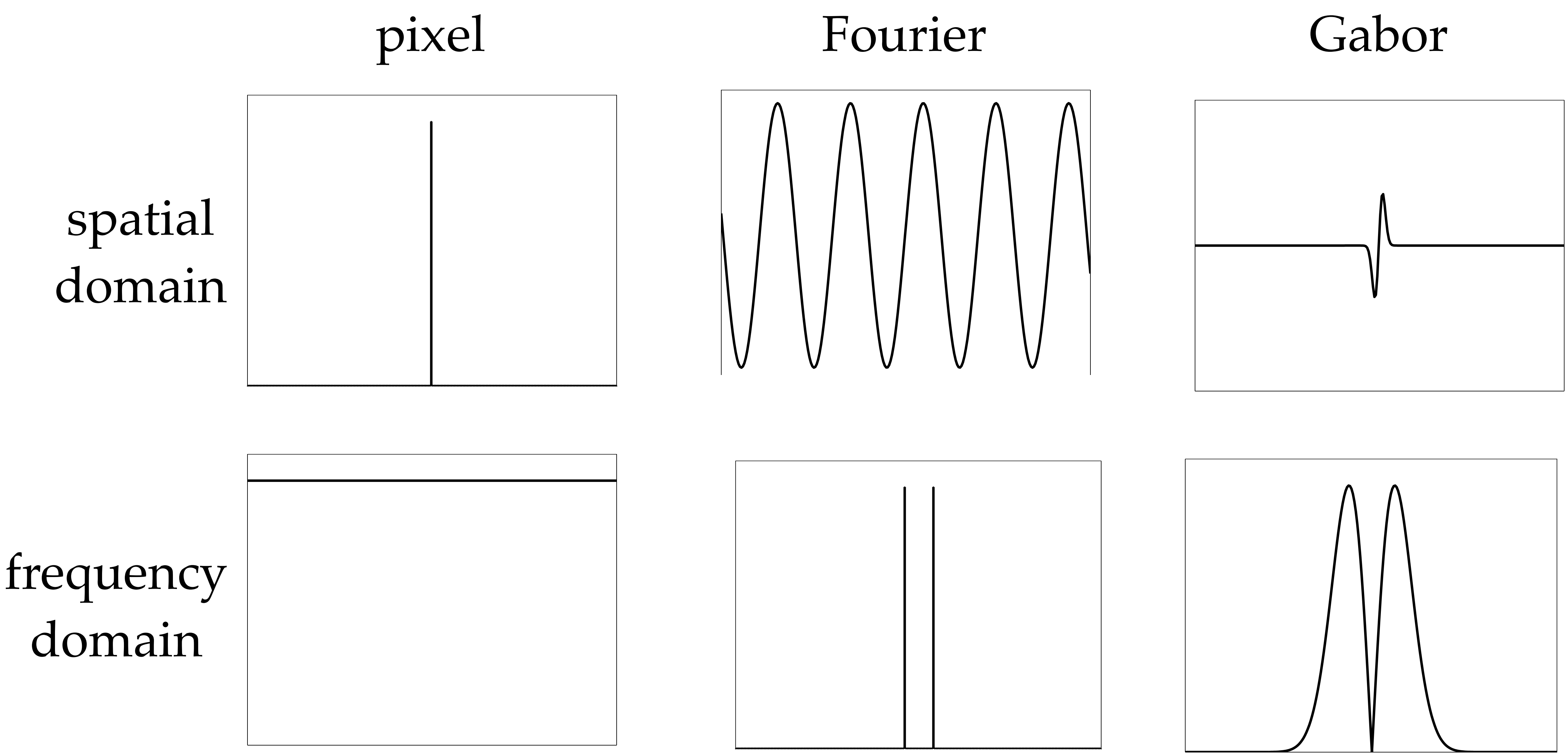
ICA basis

- approximated by Gabor functions
  - localized in space/frequency
  - orientation preference



- connection with wavelet

linear representations



wavelet

- developed in parallel with the ICA methodology
  - [Burt & Adelson, 1981; Mallat, 1989]
- data independent
  - implemented as filter banks
- wavelet filters are similar to those found by ICA
  - localized in space/time
  - orientation selective
- applicable to whole image
  - incorporate scale invariance with multi-scale pyramid structure

pyramid

figure courtesy of Jeremy Freeman

pyramid

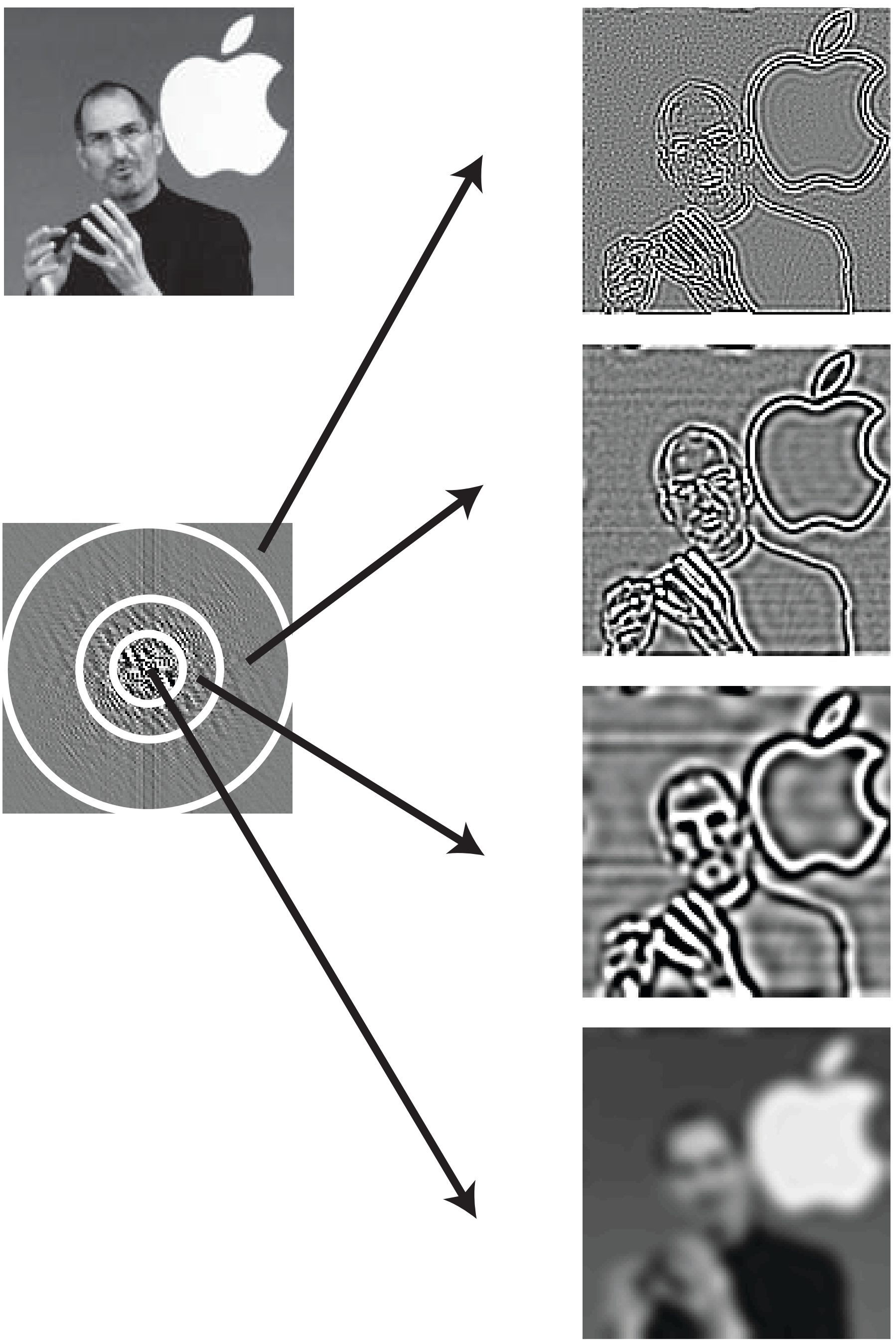


figure courtesy of Jeremy Freeman

pyramid

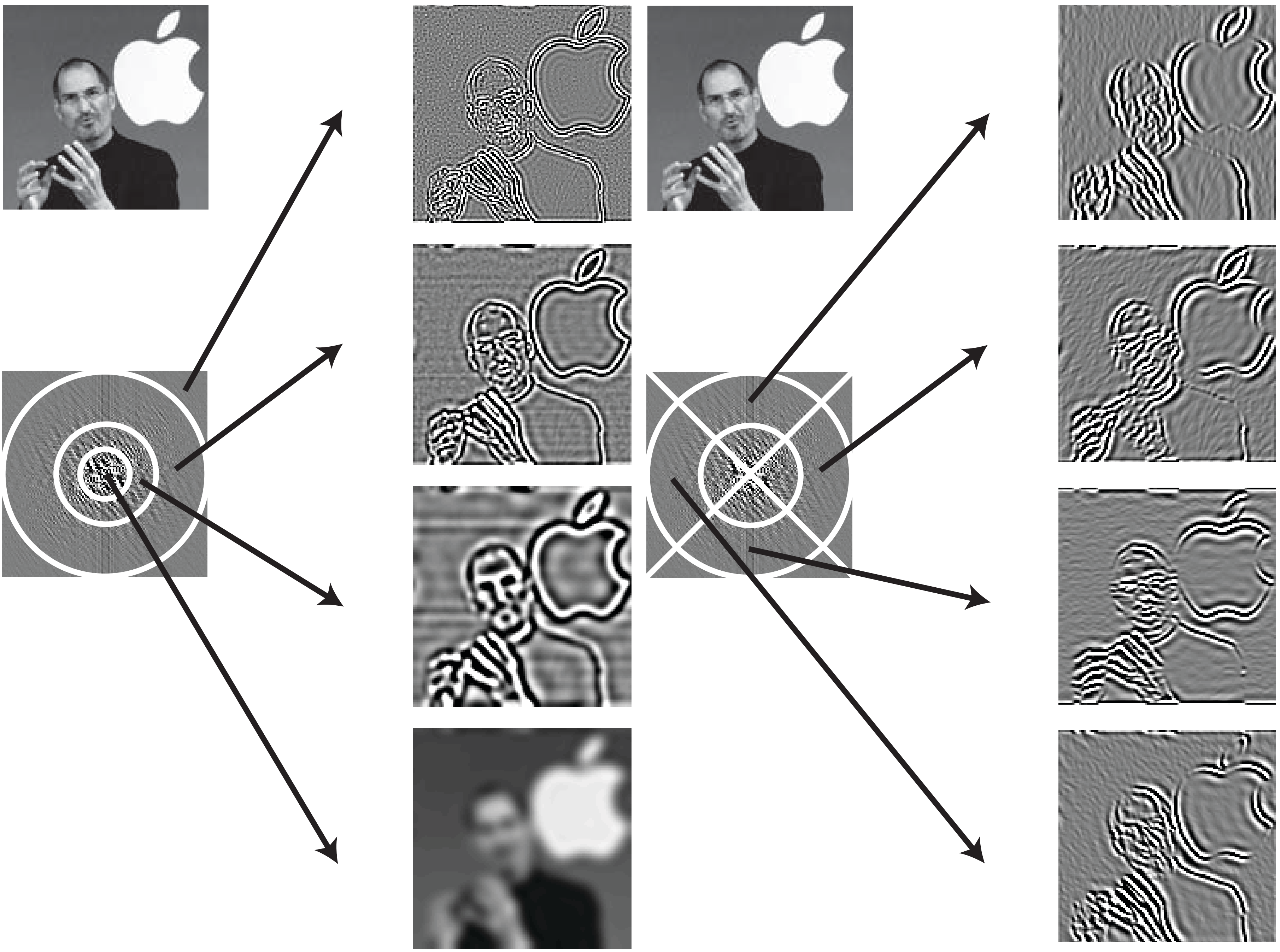


figure courtesy of Jeremy Freeman

pyramid

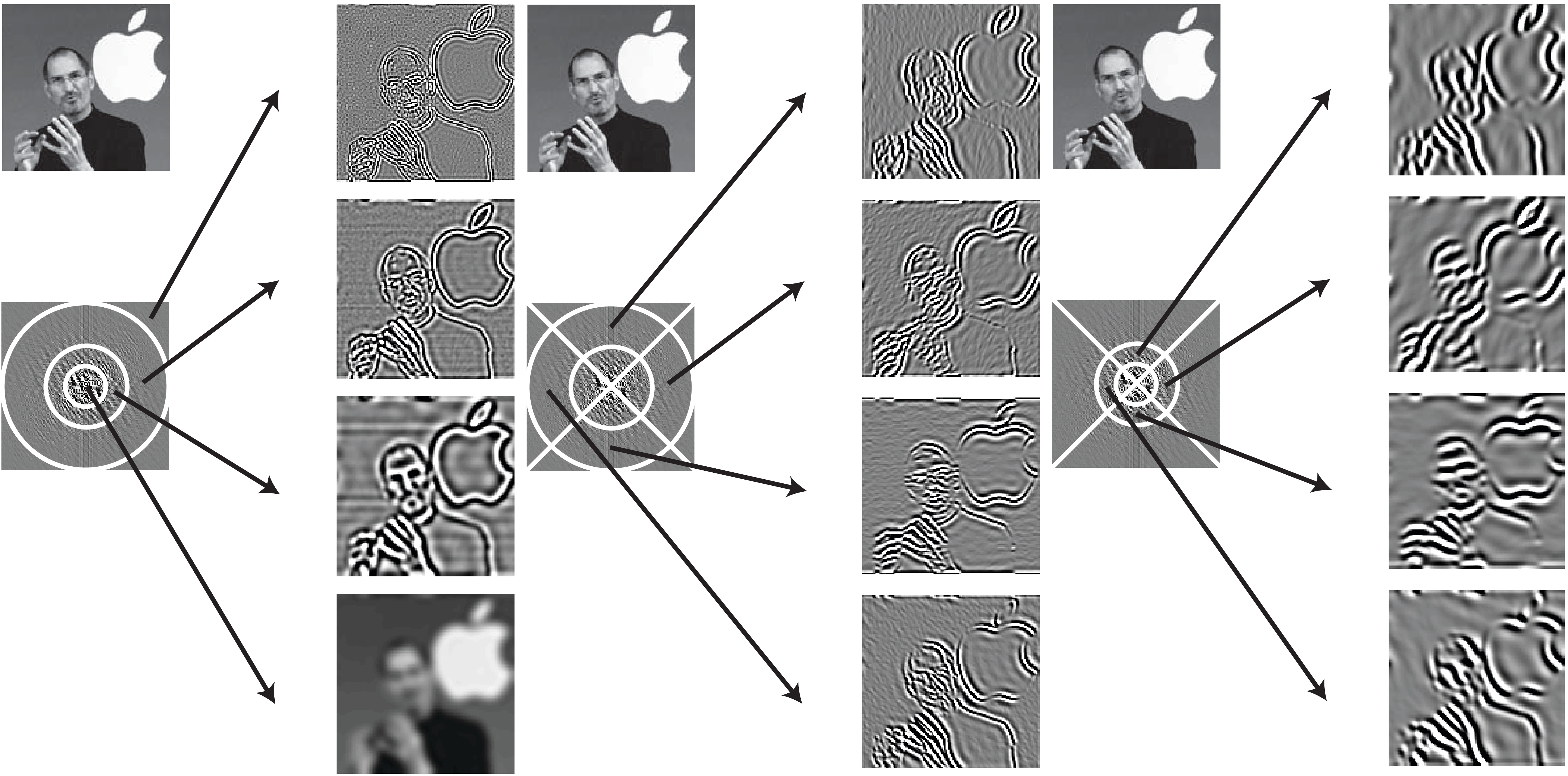


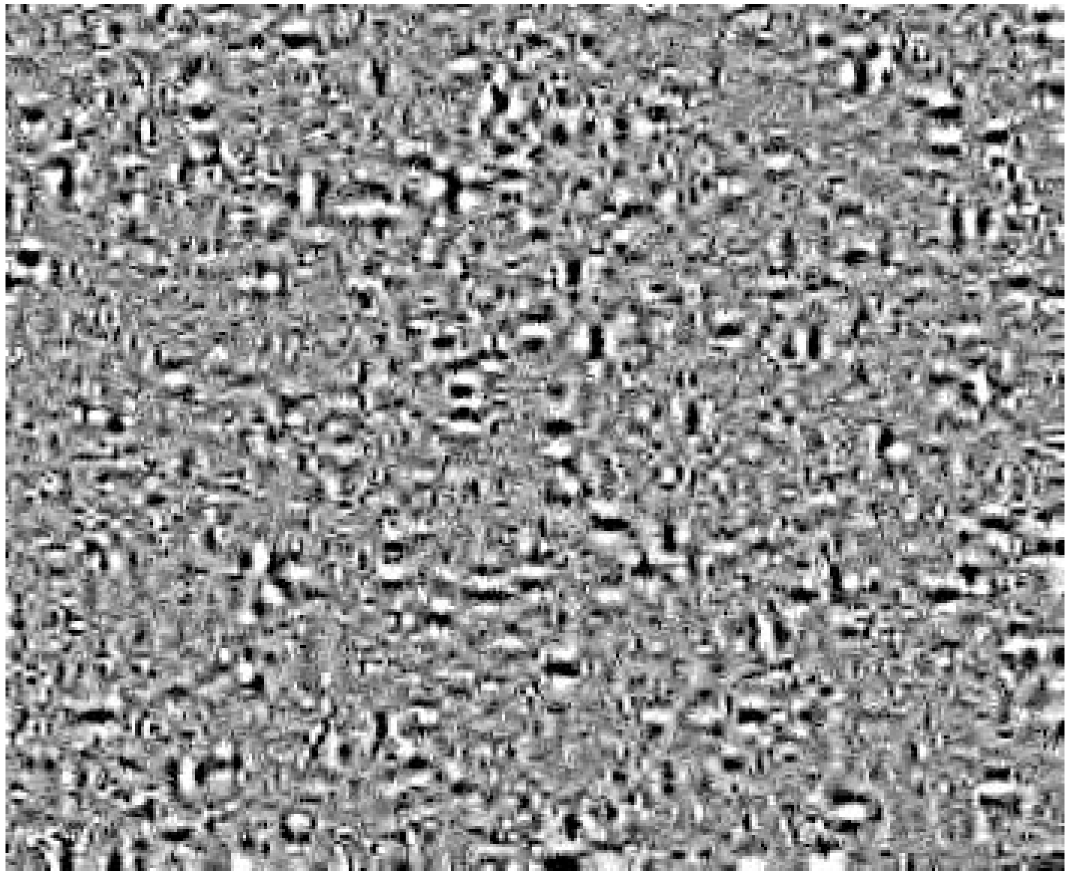
figure courtesy of Jeremy Freeman

application

- ICA and wavelet methodology brings forth a revolutionary breakthrough for image processing and computer vision, for every application there is a significant improvement in performance
  - compression
  - denoising
  - image features
  - texture synthesis
  - … ...

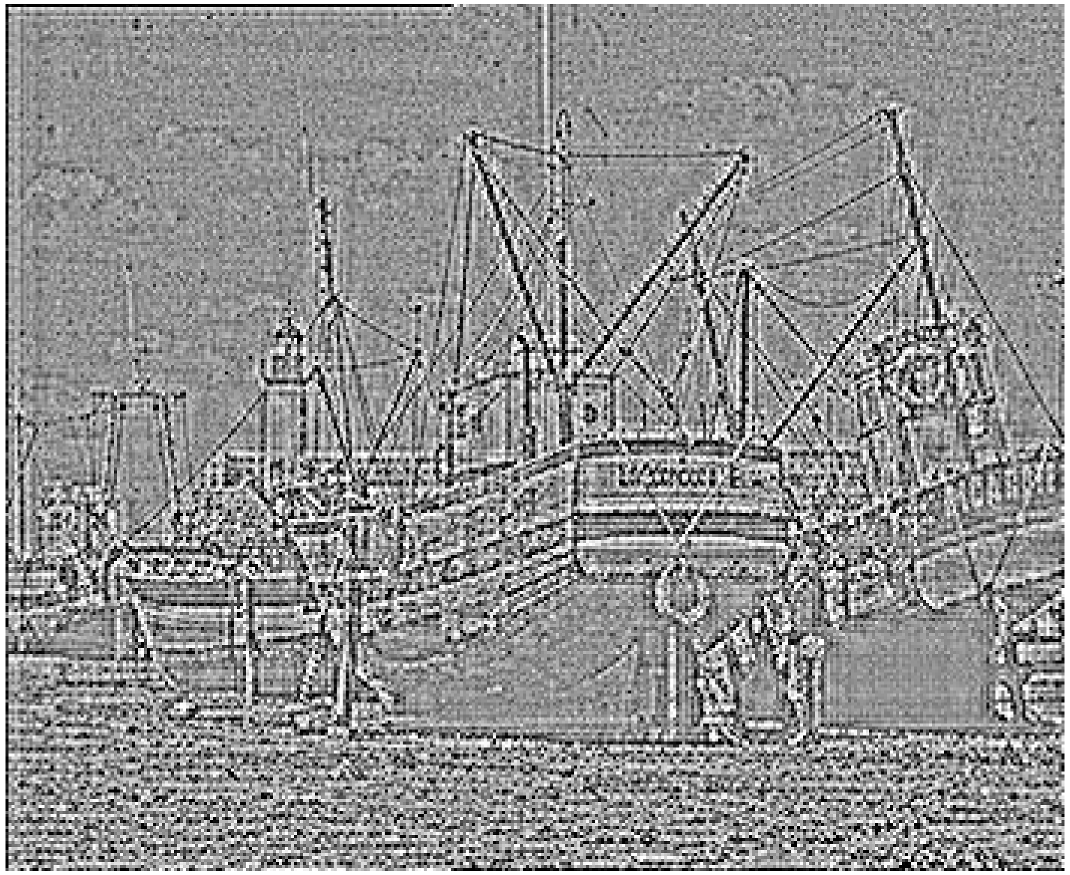
not sufficient

sample from LTF model
+ wavelet representation



not natural image

natural images after
filtered with ICA basis



not independent noise

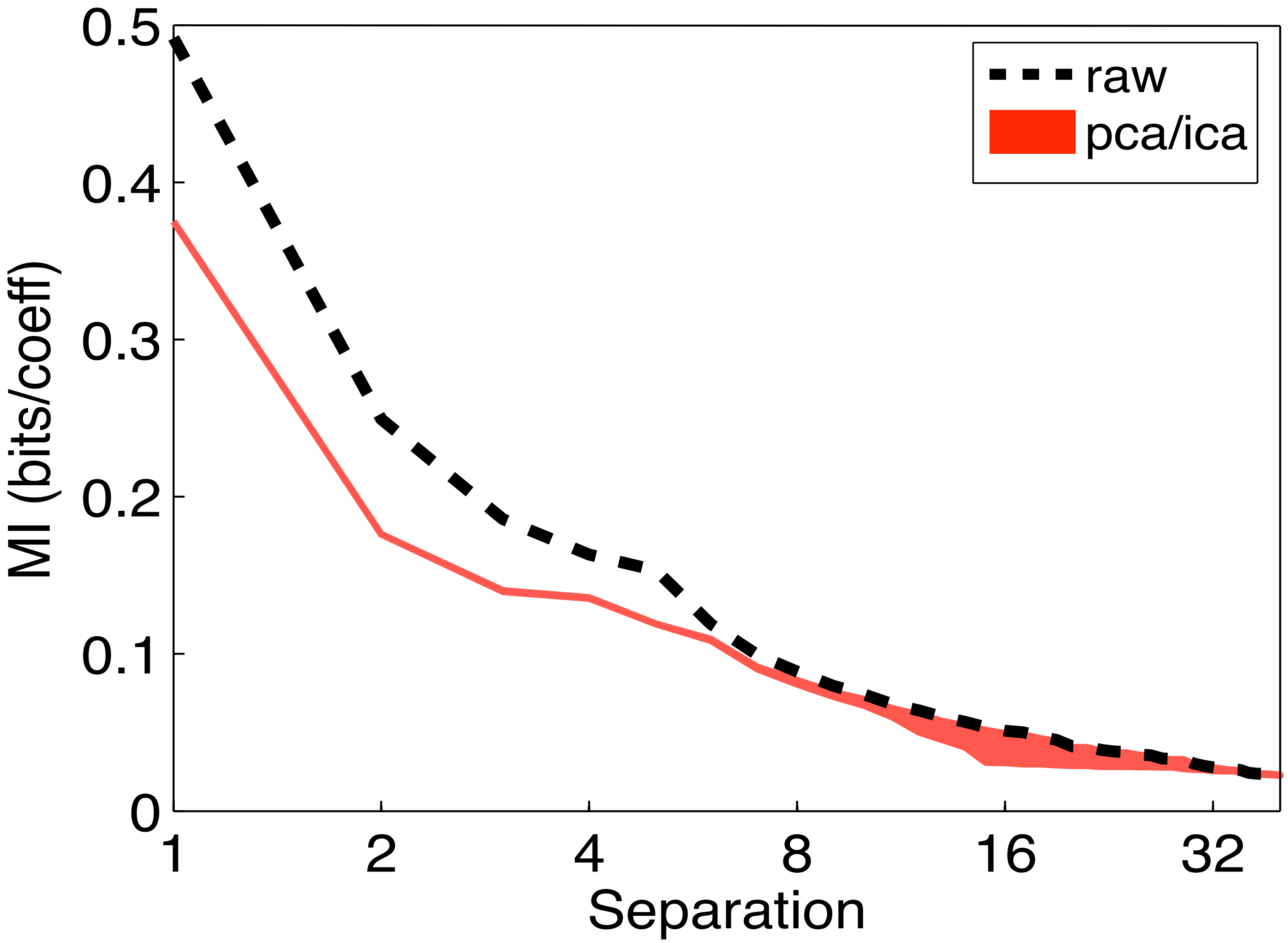
figure courtesy of Eero Simoncelli

problems with LTF/ICA

- any band-pass filter will lead to heavy tail marginals
  - even random ones
- according to LTF model, random projection (filtering) should look like Gaussian
  - central limit theorem

dependency reduction of ICA

ICA reduces less than 5% of statistical dependency
compared to PCA on natural images

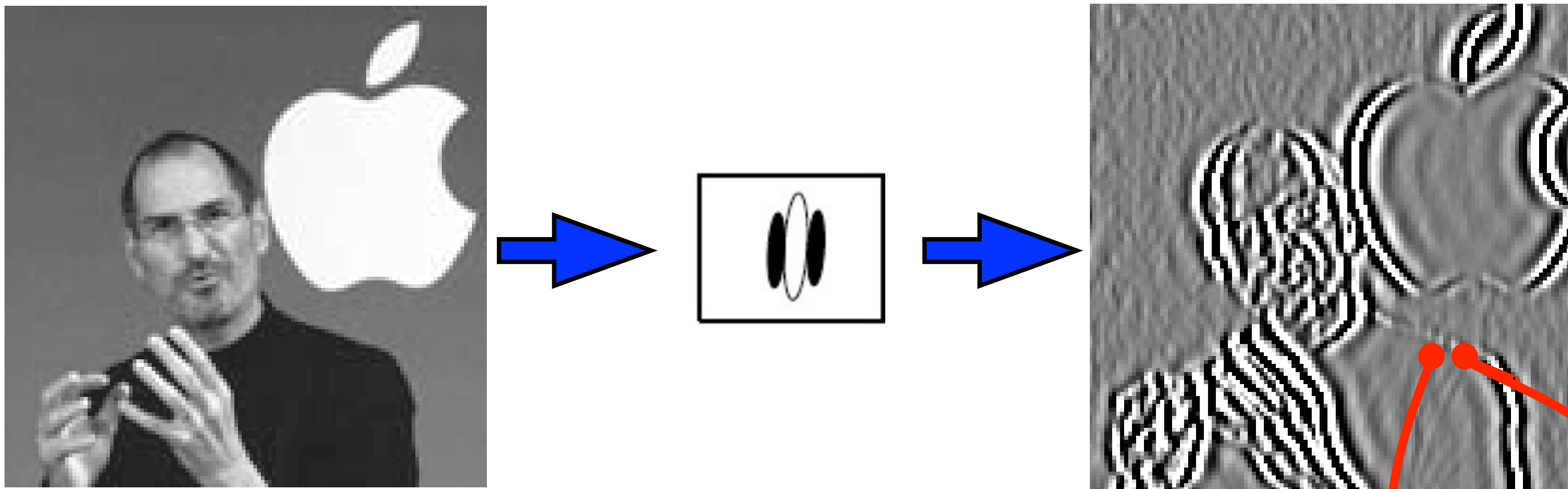


[Bethge, 06; Lyu & Simoncelli, 09]

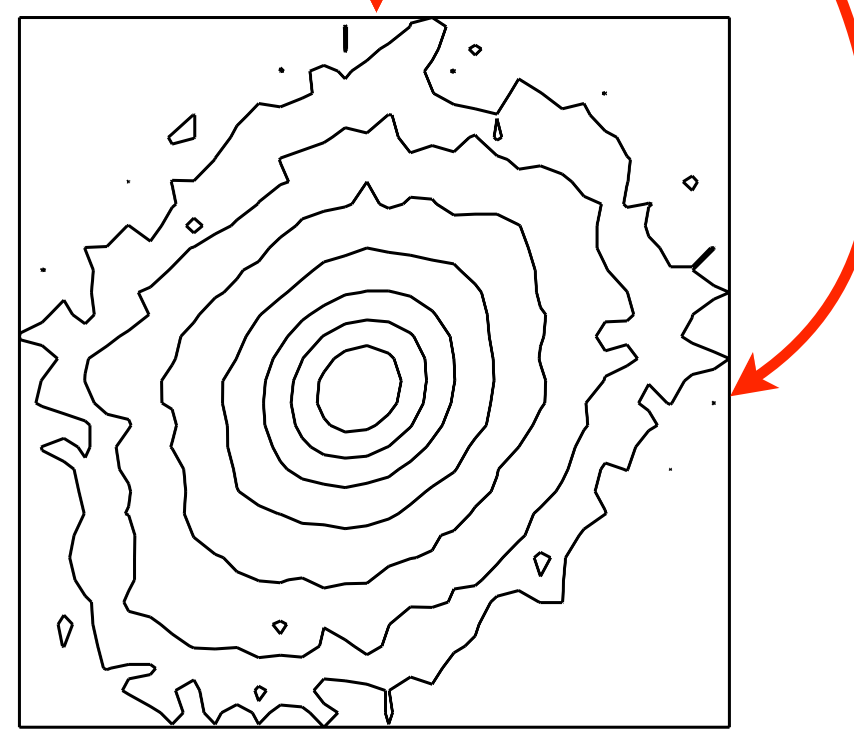
summary

- in band-pass filter domain, natural images have
  - non-Gaussian marginal distributions
  - higher-order dependency
- statistical properties lead to LTF model
- LTF model leads to ICA/wavelet representations
- not sufficient to describe natural images

problem - joint density

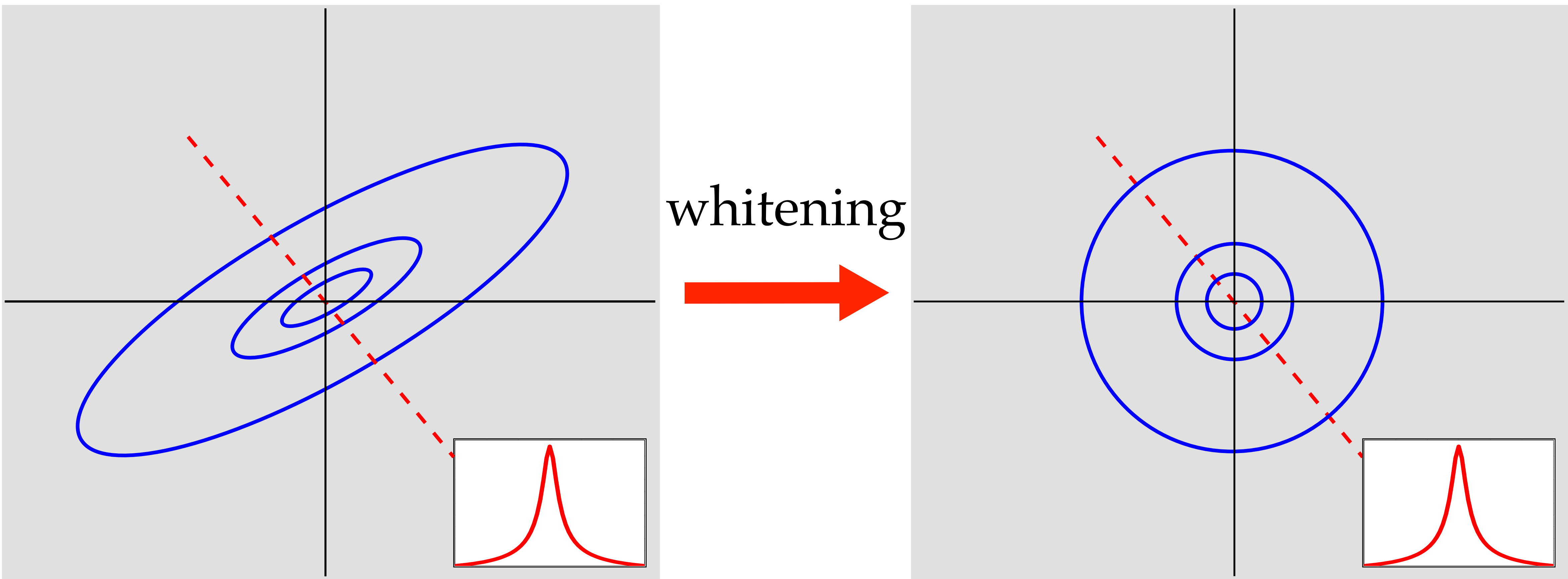


joint density of natural image
band-pass filter representationss
with separation of 2 pixels



[Wegmann & Zetzsche, 1990; Baddeley, 1996; Simoncelli, 1997] elliptically symmetric density

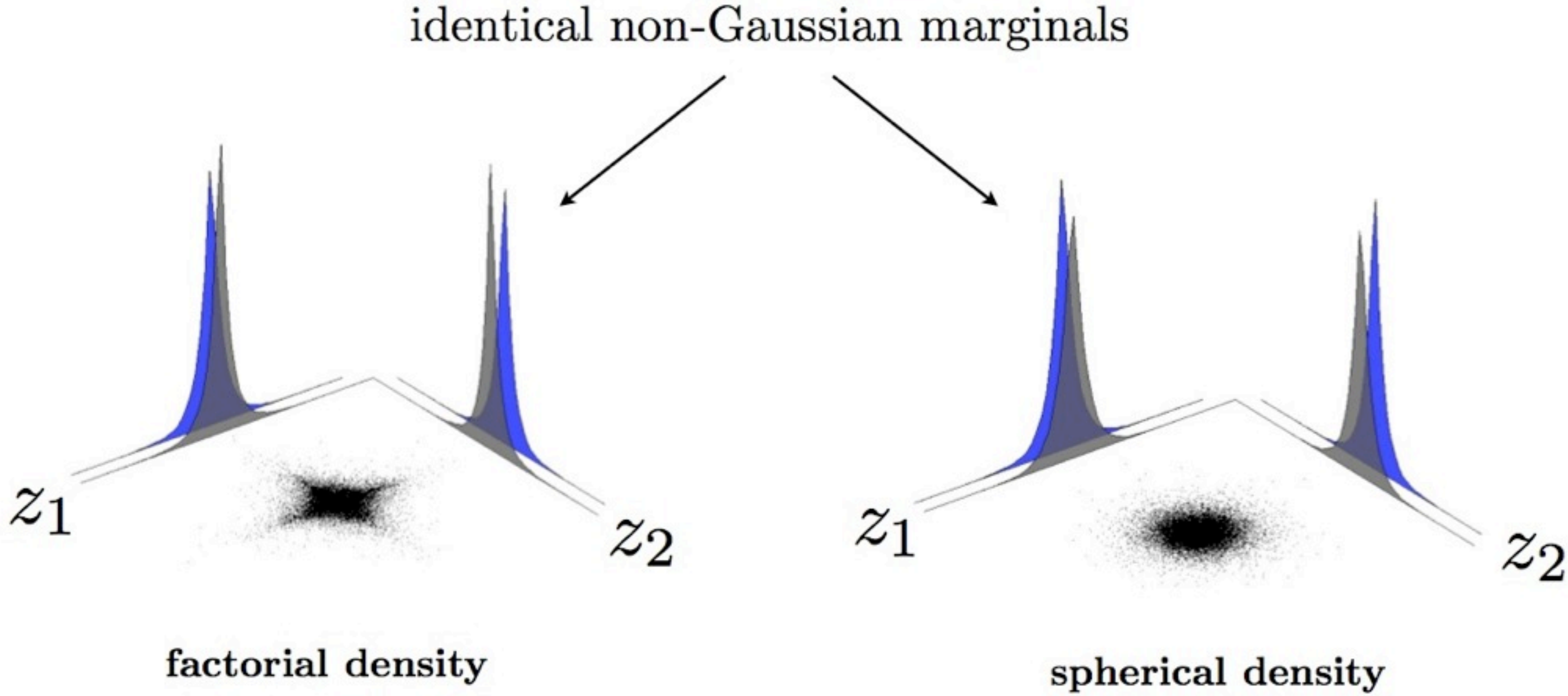
spherically symmetric density

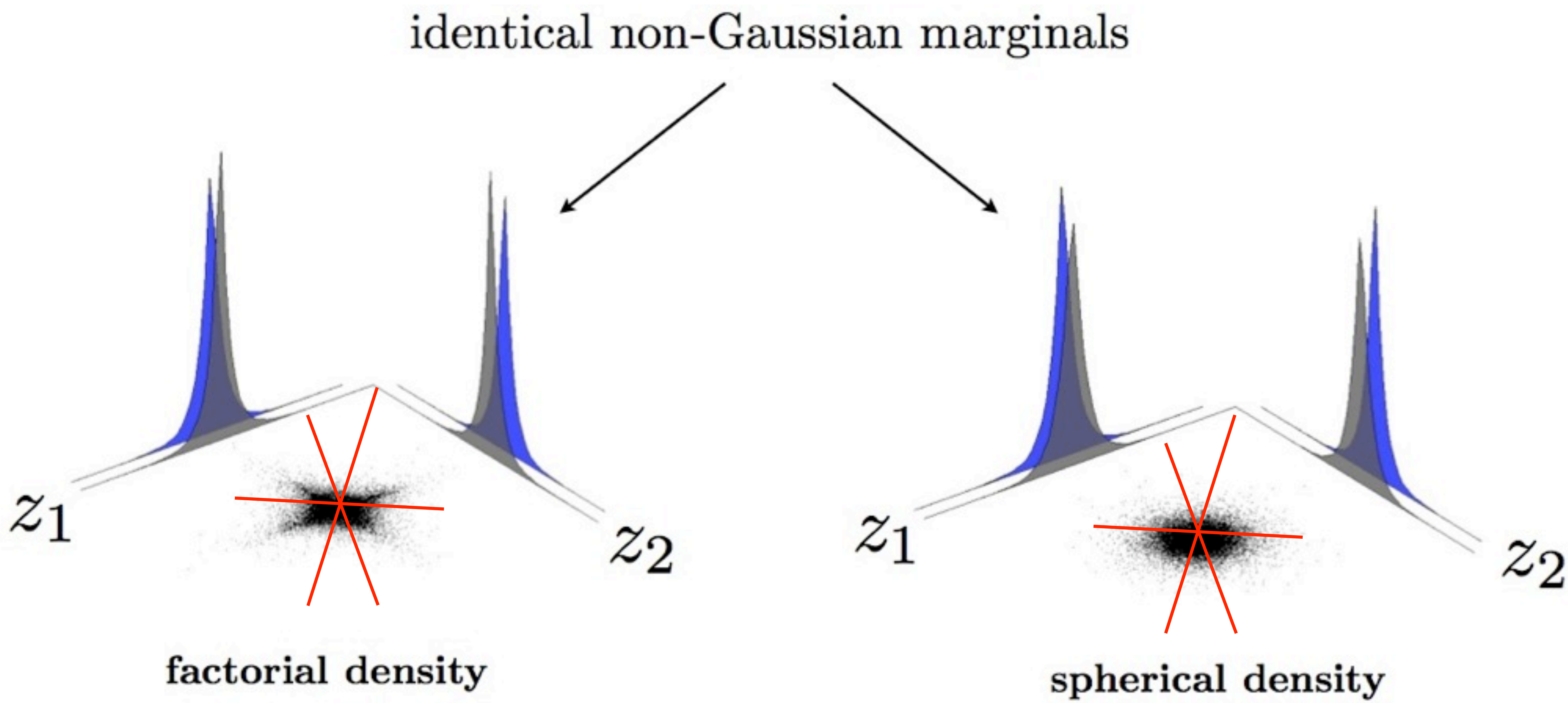


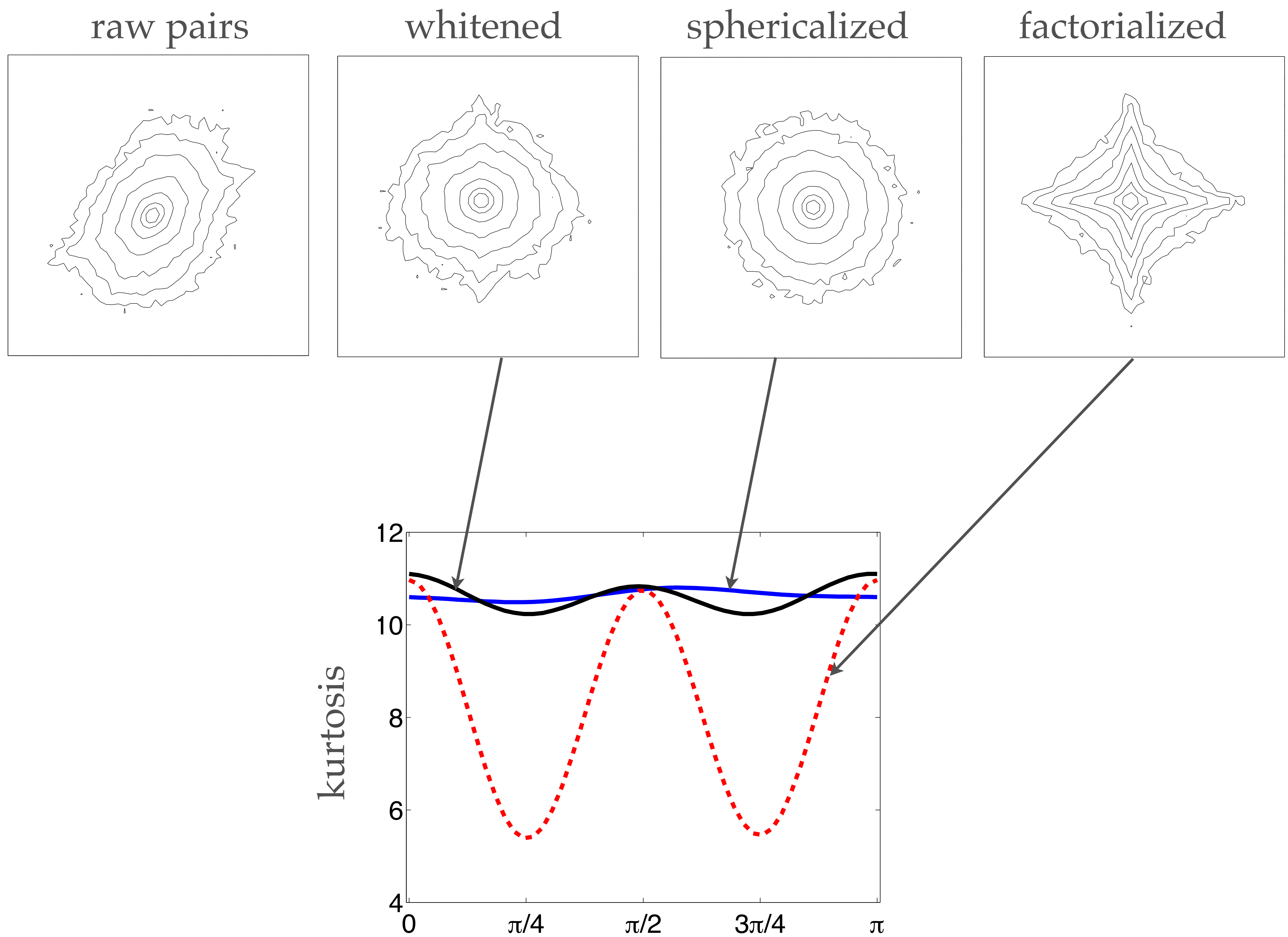
$$p_{\mathrm{esd}}(\vec{x}) = \frac{1}{\alpha|\Sigma|^{\frac{1}{2}}} f\left(-\frac{1}{2}\vec{x}^T \Sigma^{-1} \vec{x}\right)$$

$$p_{\mathrm{ssd}}(\vec{x}) = \frac{1}{\alpha} f\left(-\frac{1}{2}\vec{x}^T \vec{x}\right)$$

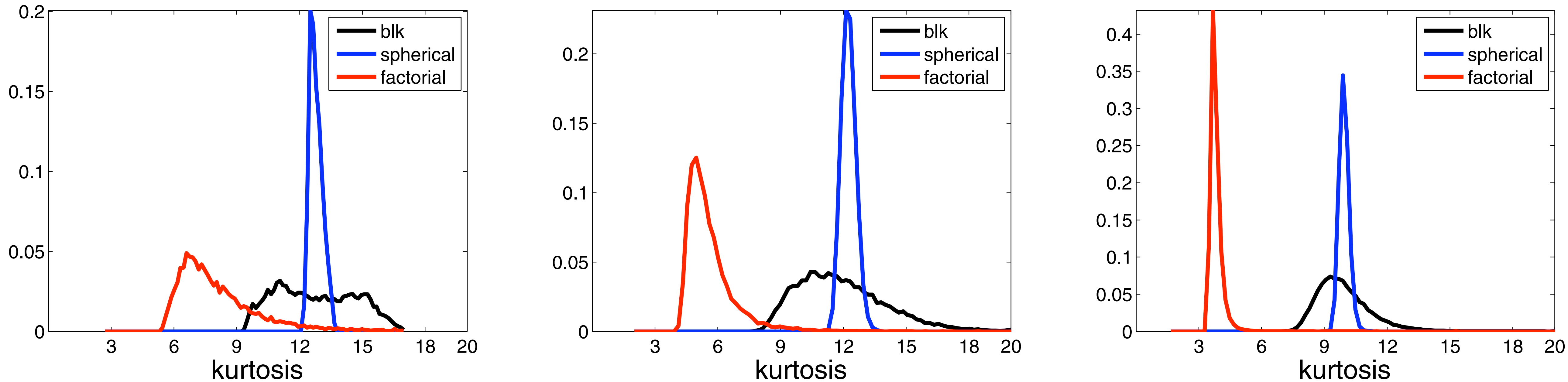
(Fang et.al. 1990)


identical non-Gaussian marginals
z_1
z_2
factorial density
z_1
z_2
spherical density


identical non-Gaussian marginals
z_1
z_2
factorial density
z_1
z_2
spherical density



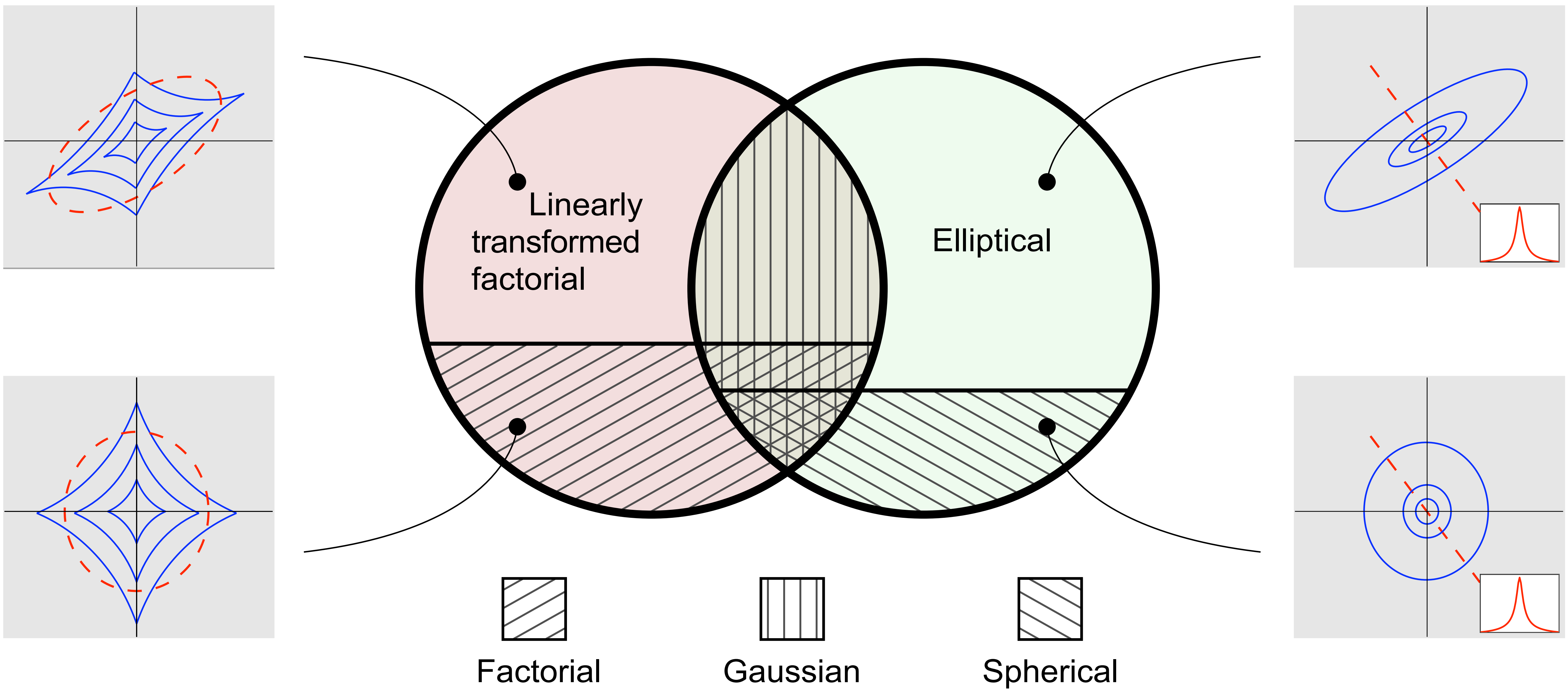
raw pairs
whitened
sphericalized
factorialized
kurtosis
12
10
8
6
4
0
π/4
π/2
3π/4
π

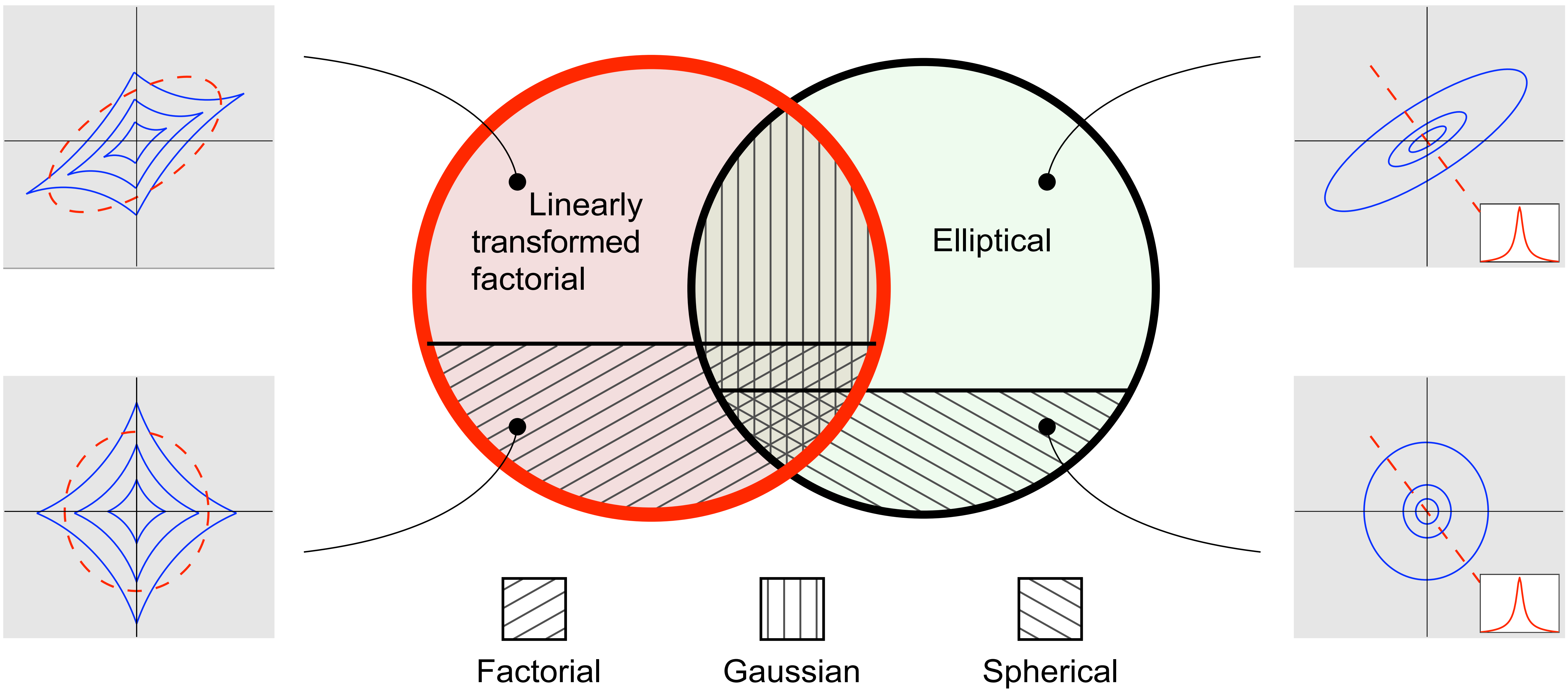


3x3 7x7 15x15

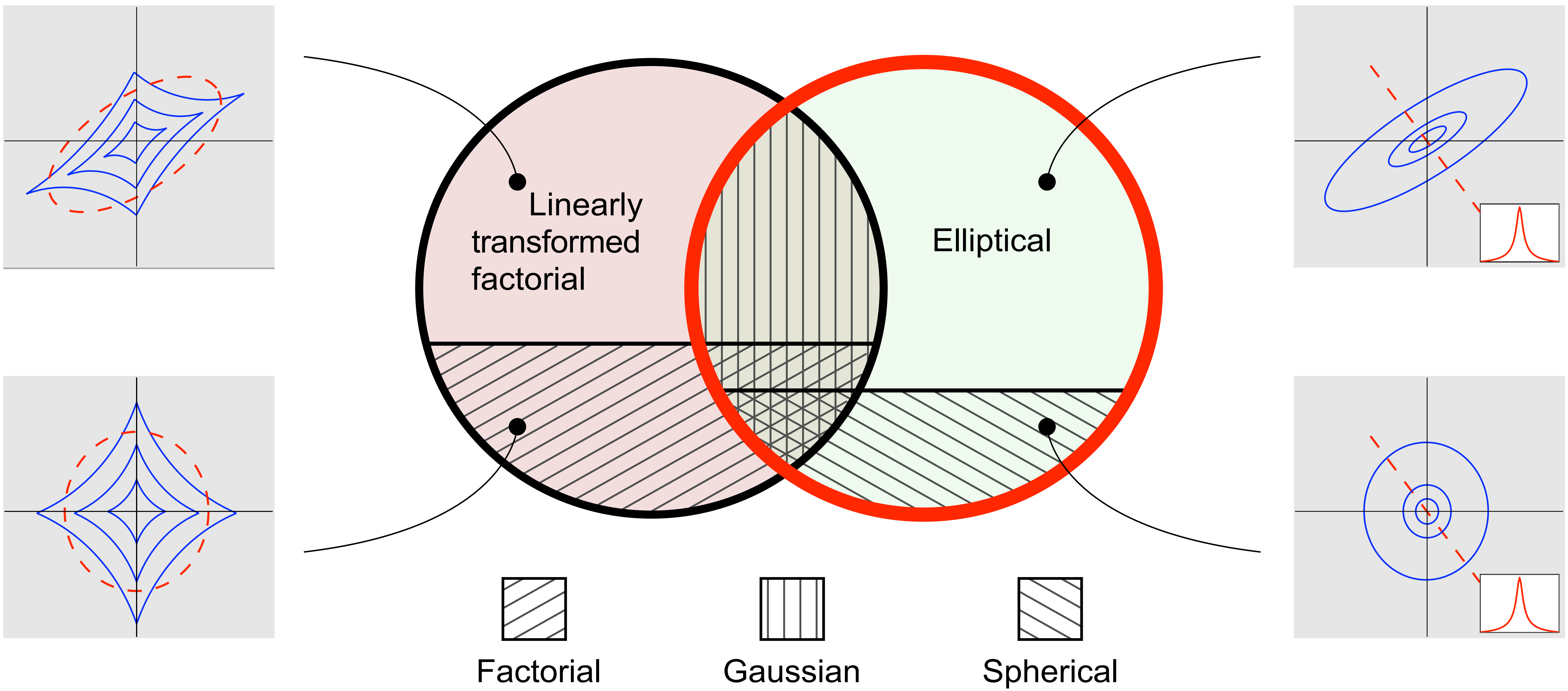
data (ICA'd): sphericalized: factorialized:

- Histograms, kurtosis of projections of image blocks onto random unit-norm basis functions.
- These imply data are closer to spherical than factorial

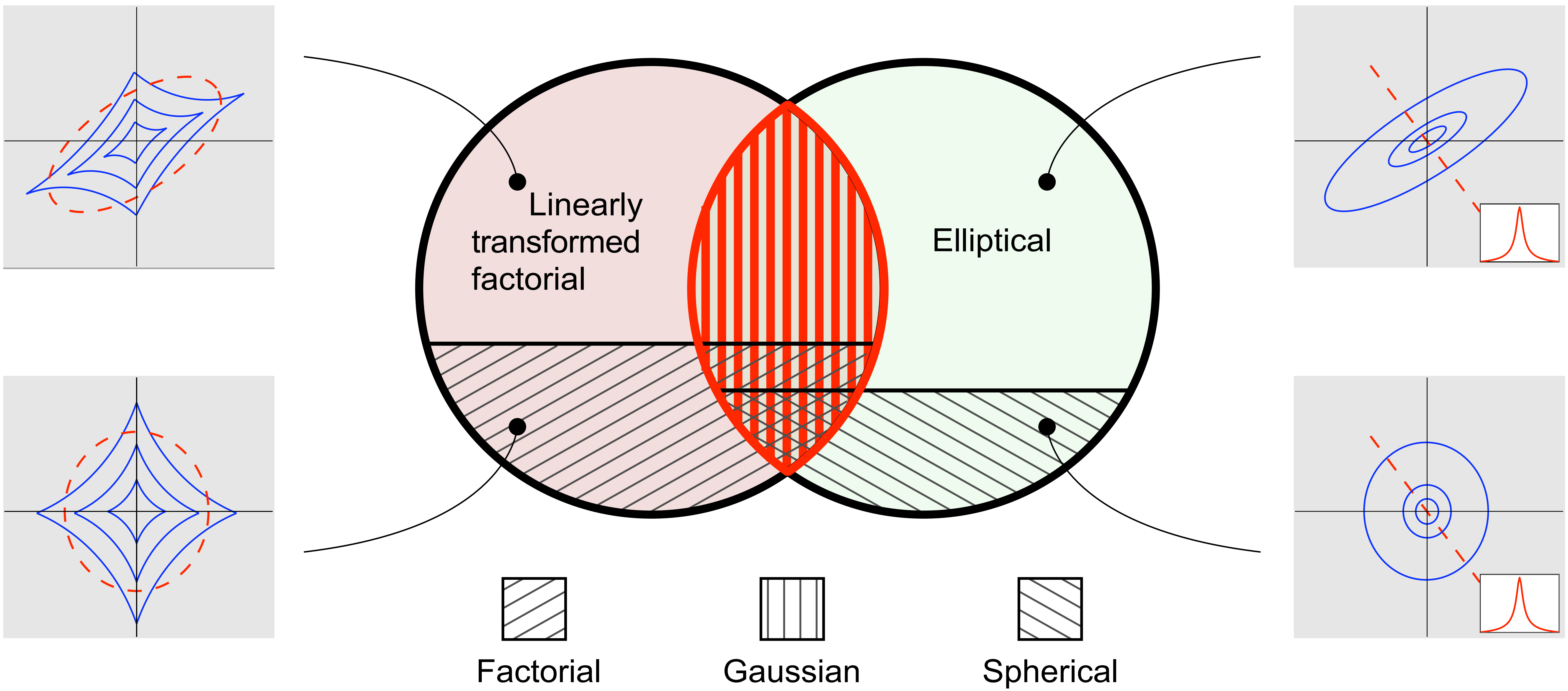

Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical



Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical



Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical

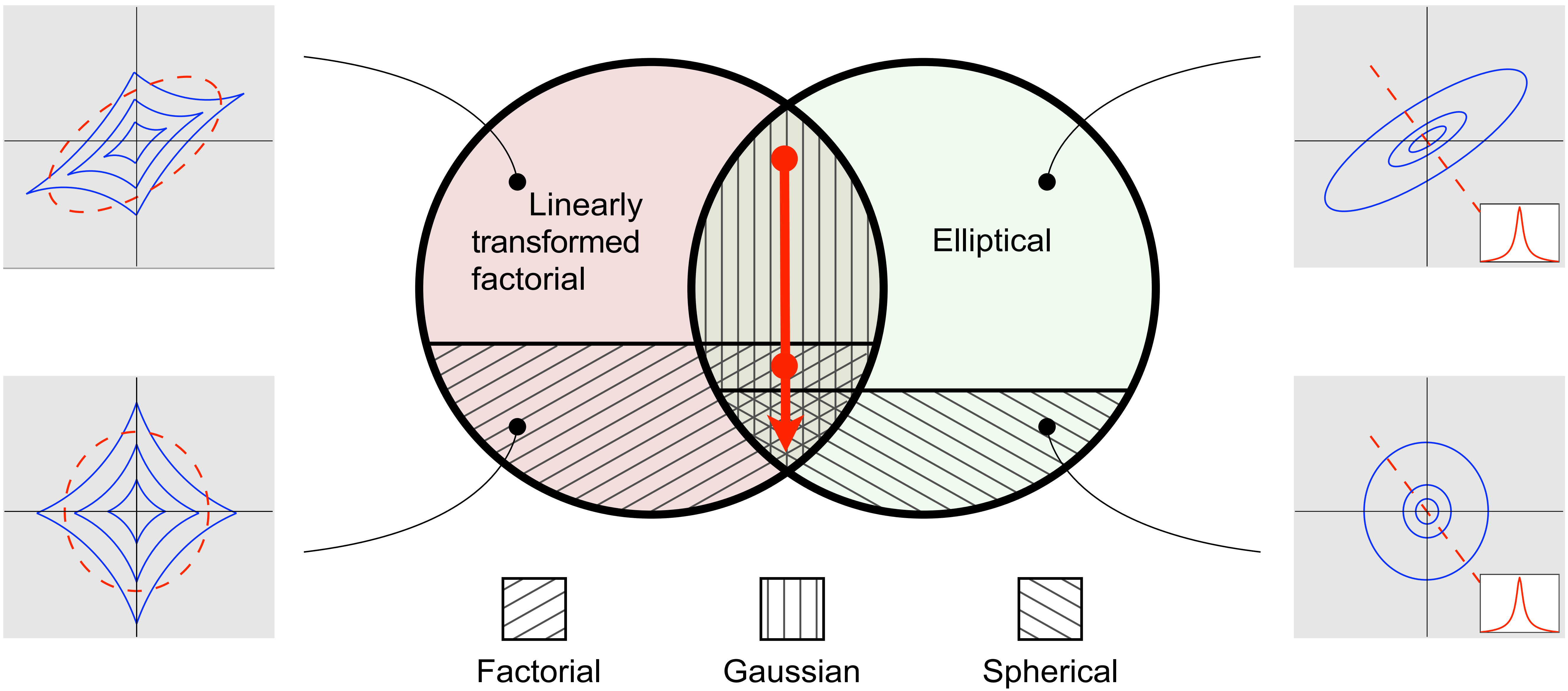


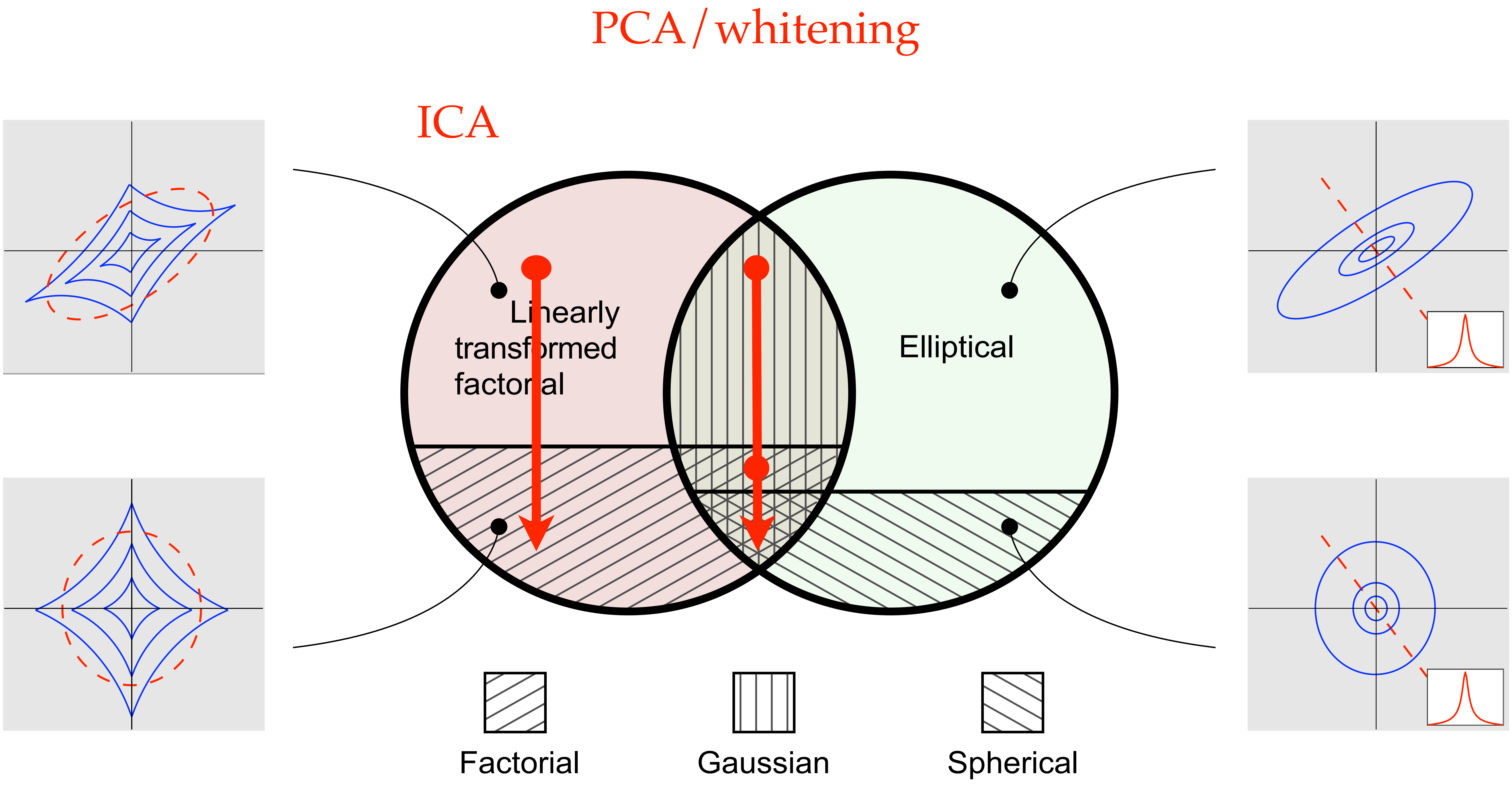
Linearly
transformed
factorial
Elliptical
Factorial
Gaussian
Spherical

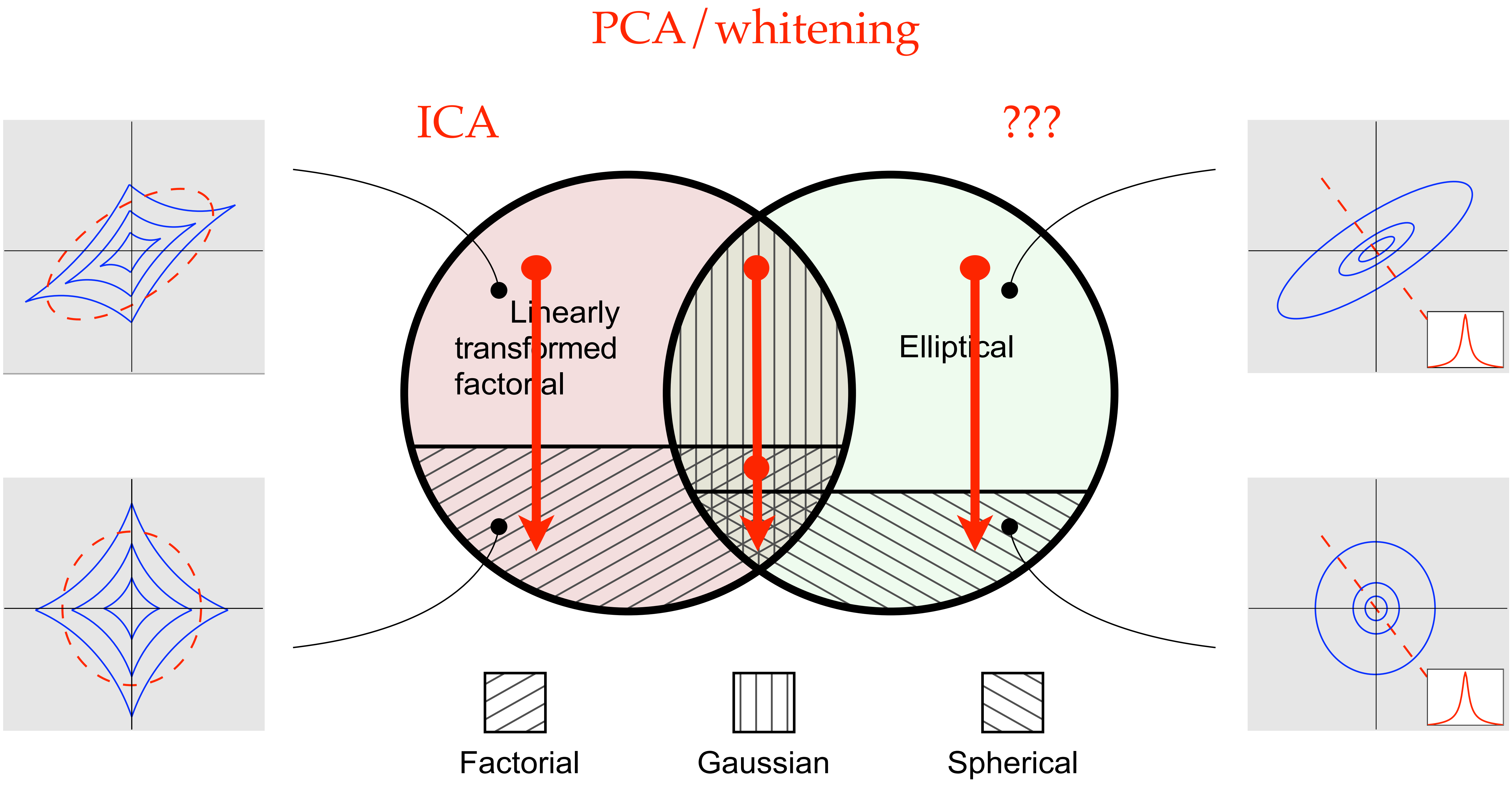
$$
\begin{aligned}
p(\vec{x}) &= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{\vec{x}^T\vec{x}}{2}\right) \\
&= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\sum_{i=1}^{d} x_i^2\right) \\
&= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \\
&= \prod_{i=1}^{d} p(x_i)
\end{aligned}
$$

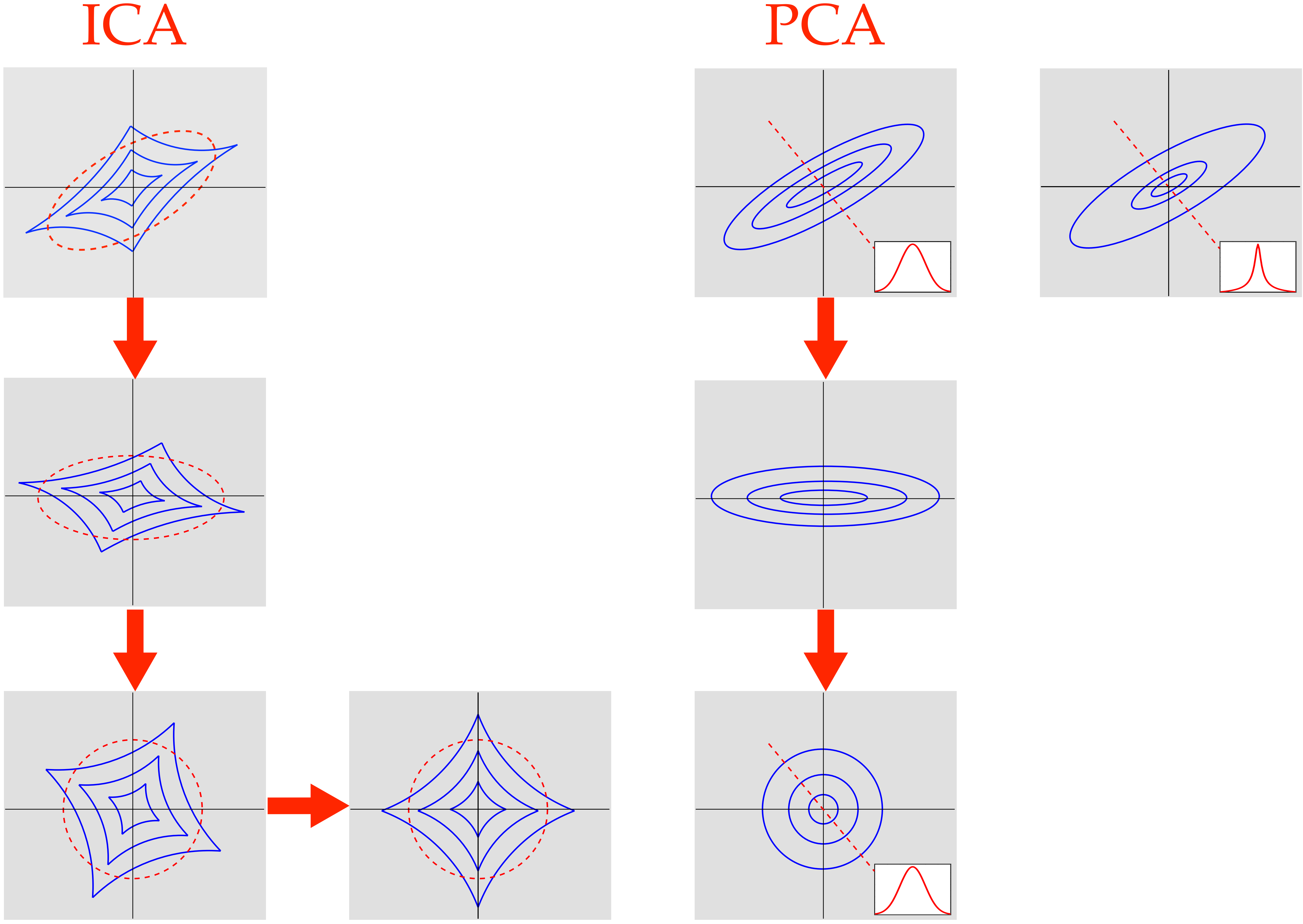
Gaussian is the **only** density that can be both factorial and spherically symmetric [Nash and Klamkin 1976]

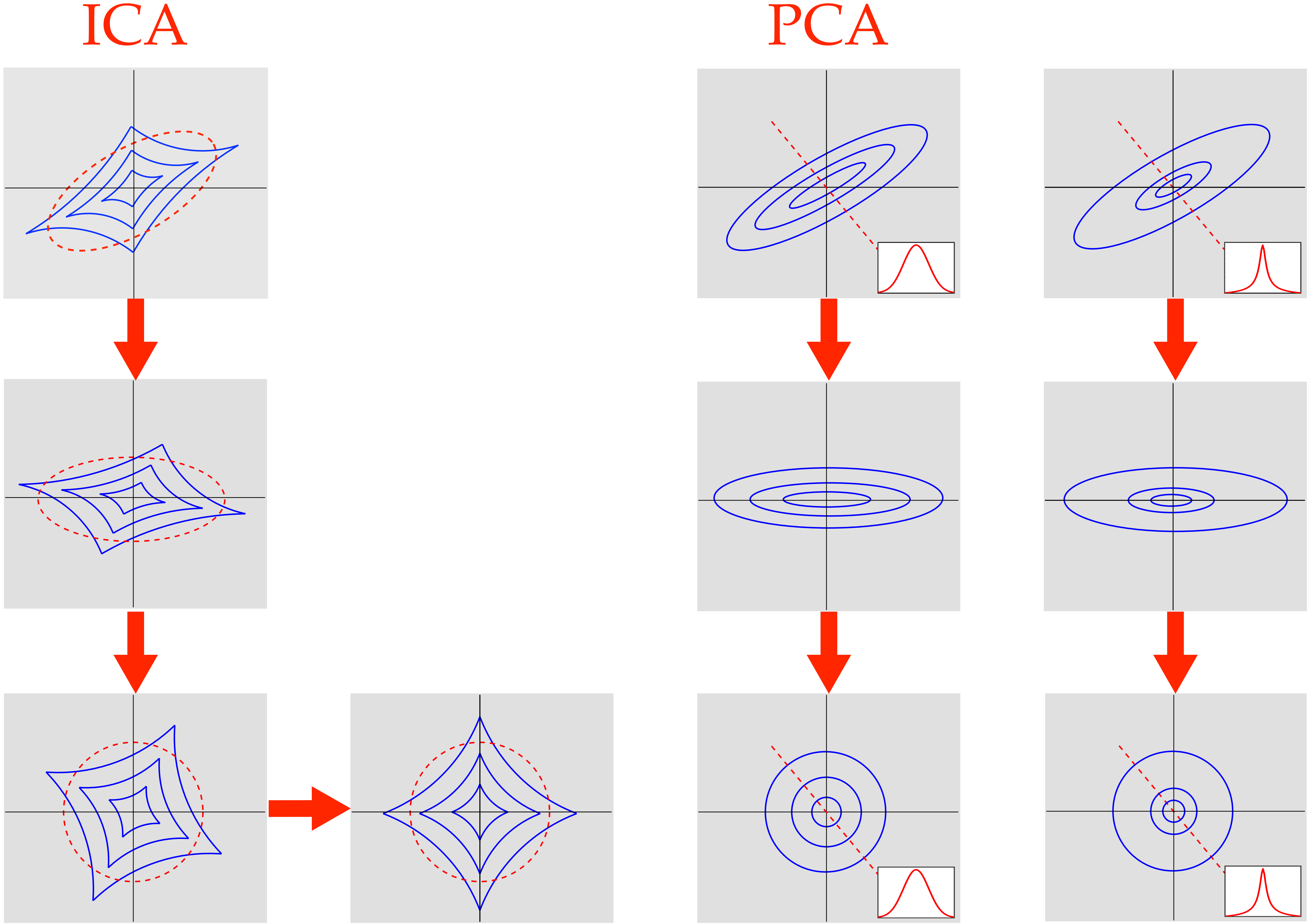
PCA / whitening




PCA / whitening
ICA
Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical


PCA / whitening
ICA
???
Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical


ICA
PCA


ICA
PCA

assumptions of LTF/ICA

$$\vec{x} \quad = \quad A\vec{s}$$

- linear transform between signal and representation
- one signal corresponds to one representation
- one representation corresponds to one signal
- representationss are mutually independent

assumptions of LTF / ICA

$$\vec{x} \quad = \quad A\vec{s}$$

- linear transform between signal and representation

- one signal corresponds to one representation

- one representation corresponds to one signal

- representationss are mutually independent

complete
linear sys

assumptions of LTF / ICA

- linear transform between signal and representation
- one signal corresponds to one representation
- one representation corresponds to one signal
- ~~representationss are mutually independent~~
  - explicitly modeling dependencies in representation

complete representation

- independent subspace analysis and topographic ICA
  - [Hyvarinen & Hoyer, 2000; Hyvarinen, Hoyer & Inki, 2000]
- hierarchical models
  - e.g., [Karklin & Lewicki, 2003, 2009; Ranzato & Hinton, 2010; … …]
- joint GSM model for wavelet coefficients
  - [Wainwright & Simoncelli, 1999; Portilla etal., 2003]
- MRF models for wavelet coefficients
  - e.g., [Crouse etal, 1999; Lyu & Simoncelli, 2008; Lyu, 2009; … …]
- tree dependent component analysis
  - [Zoran & Weiss, 2009]

assumptions of LTF/ICA

- linear transform between signal and representation
- one signal corresponds to one representation
- ~~one representation corresponds to one signal~~
  - multiple representation can lead to one signal
- representation are mutually independent

- achieving sparsity can be a driving principle itself
  - over-complete sparse coding (nonlin encoding, lin decoding)
    - [Olshausen & Field, 1996]
  - compressed sensing (lin encoding, nonlin decoding)
    - [Candes & Donoho, 2003]
  - PCA/whitening/ICA (lin encoding, lin decoding)

assumptions of LTF / ICA

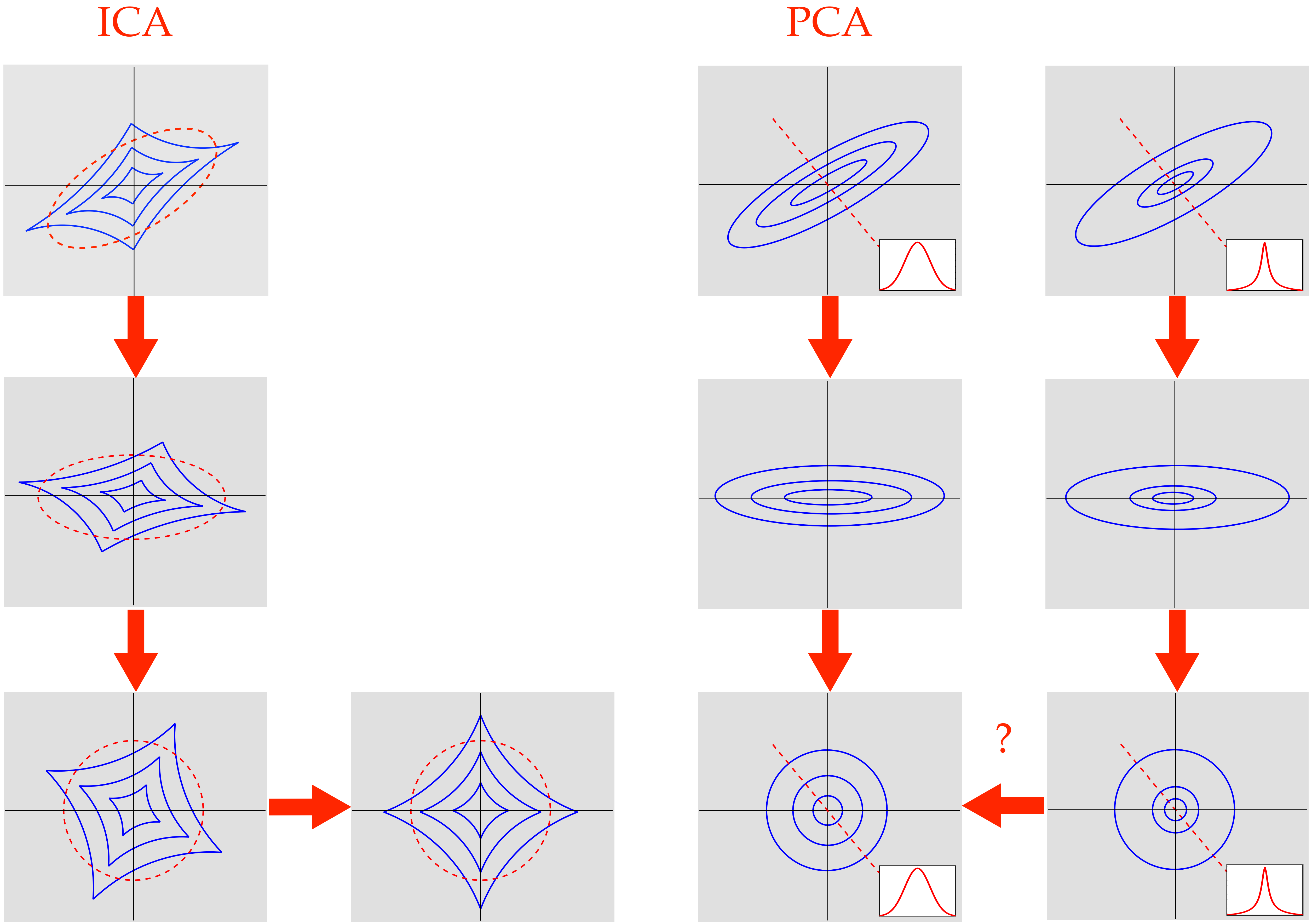
- linear transform between signal and representation
- ~~one signal corresponds to one representation~~
  - focusing on the analysis $\vec{s} = B\vec{x}$
- one representation corresponds to one signal
- representations are mutually independent

maximum entropy models

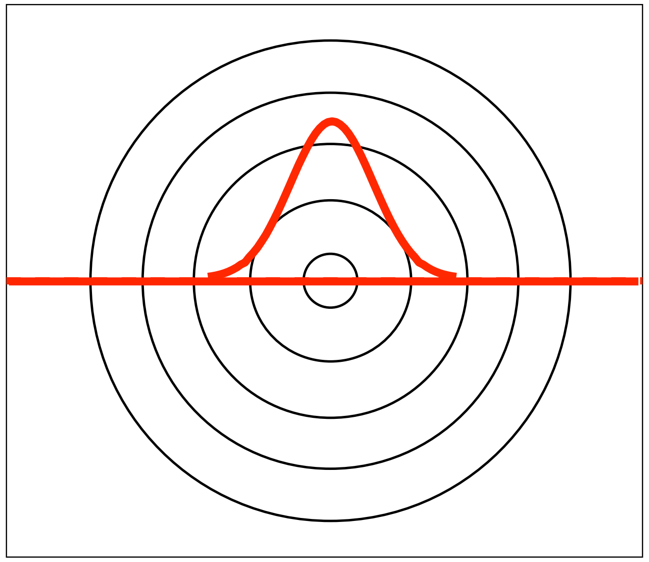
- use representationss as constraints to build statistical models
  - patch models
    - product of experts [Teh etal., 2003]
    - product of edgeperts [Gehler and Welling, 2006]
  - image models
    - FRAME [Zhu, Wu & Mumford, 2001]
    - field of experts [Roth & Black, 2005]

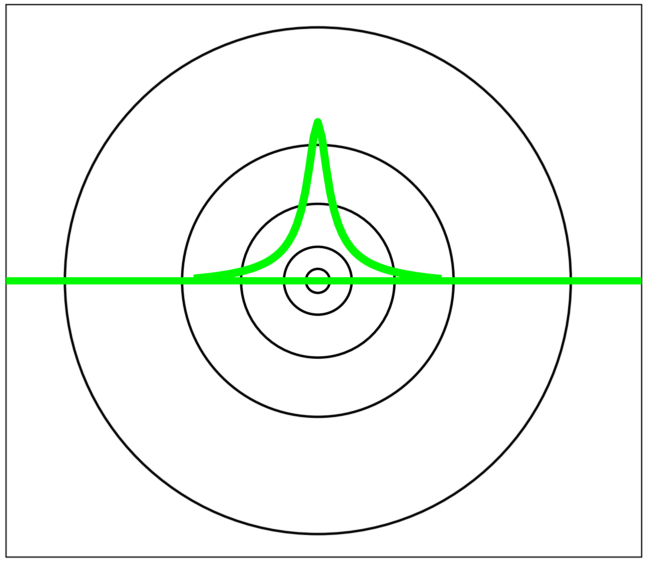
assumptions of LTF/ICA

- ~~linear transform between signal and representations~~
  - find nonlinear encoding/decoding transforms
- one signal corresponds to one representations
- one representations corresponds to one signal
- representationss are mutually independent


ICA
PCA
?

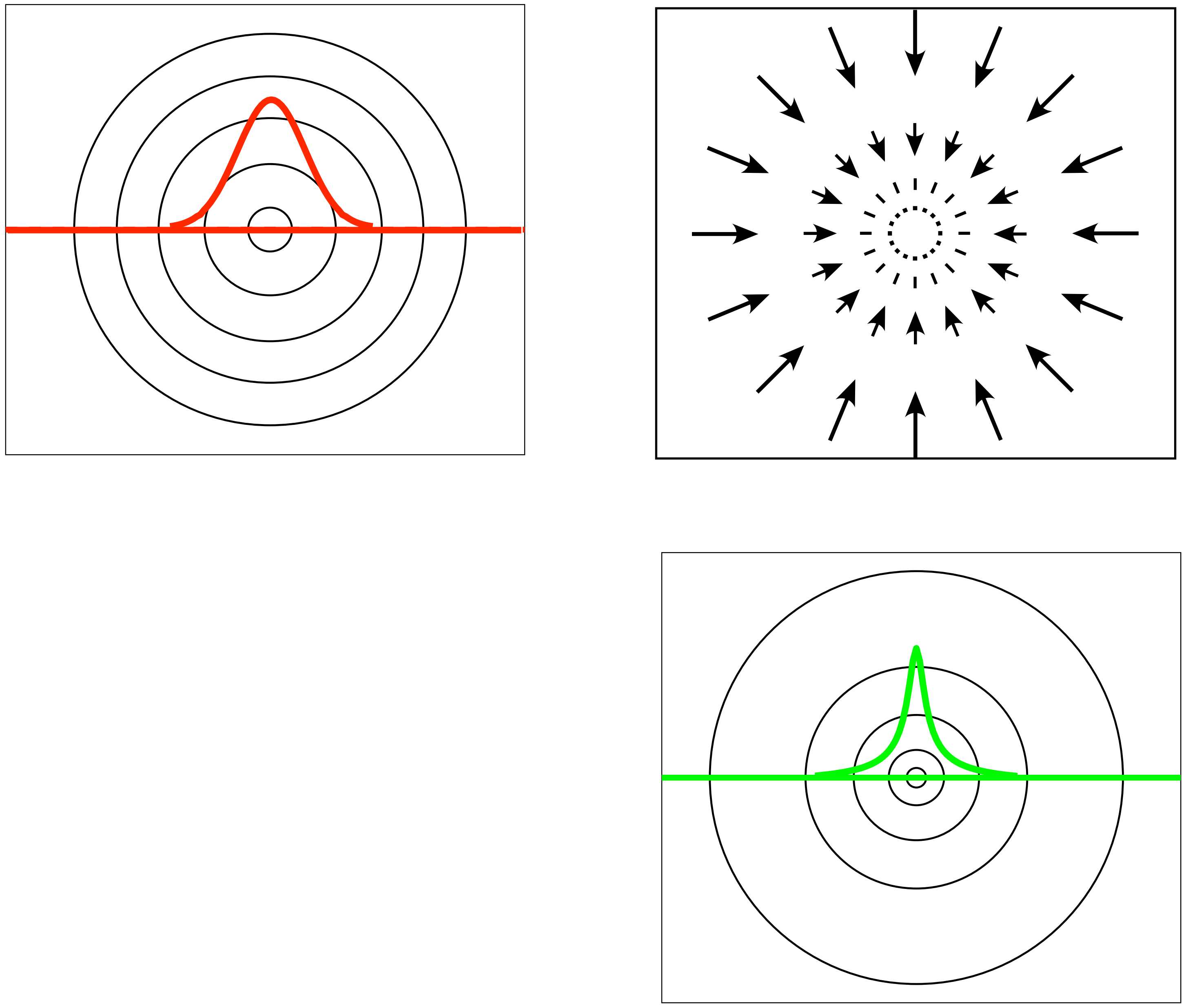
radial Gaussianization (RG)





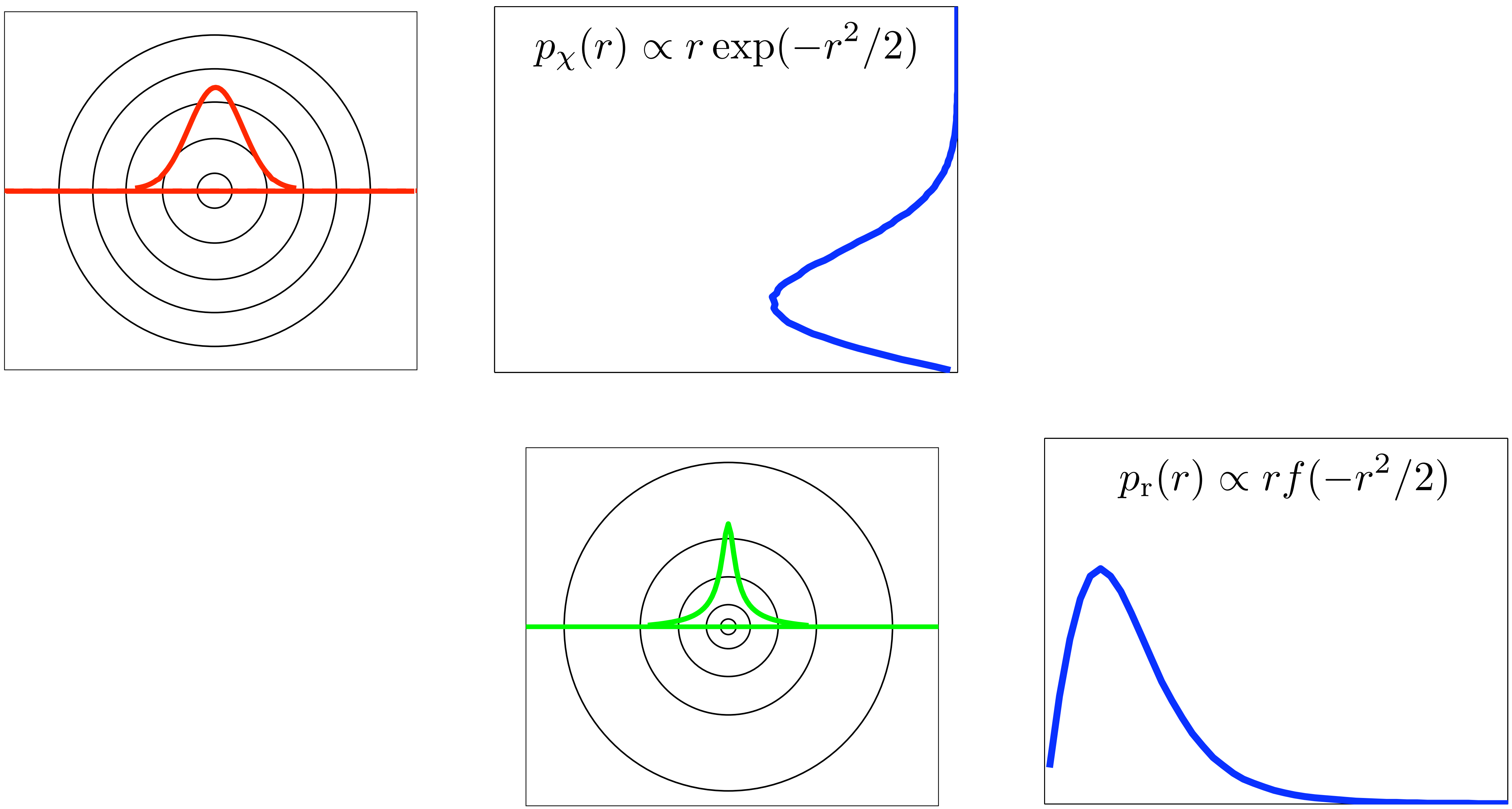
[Lyu & Simoncelli, 2009]

radial Gaussianization (RG)



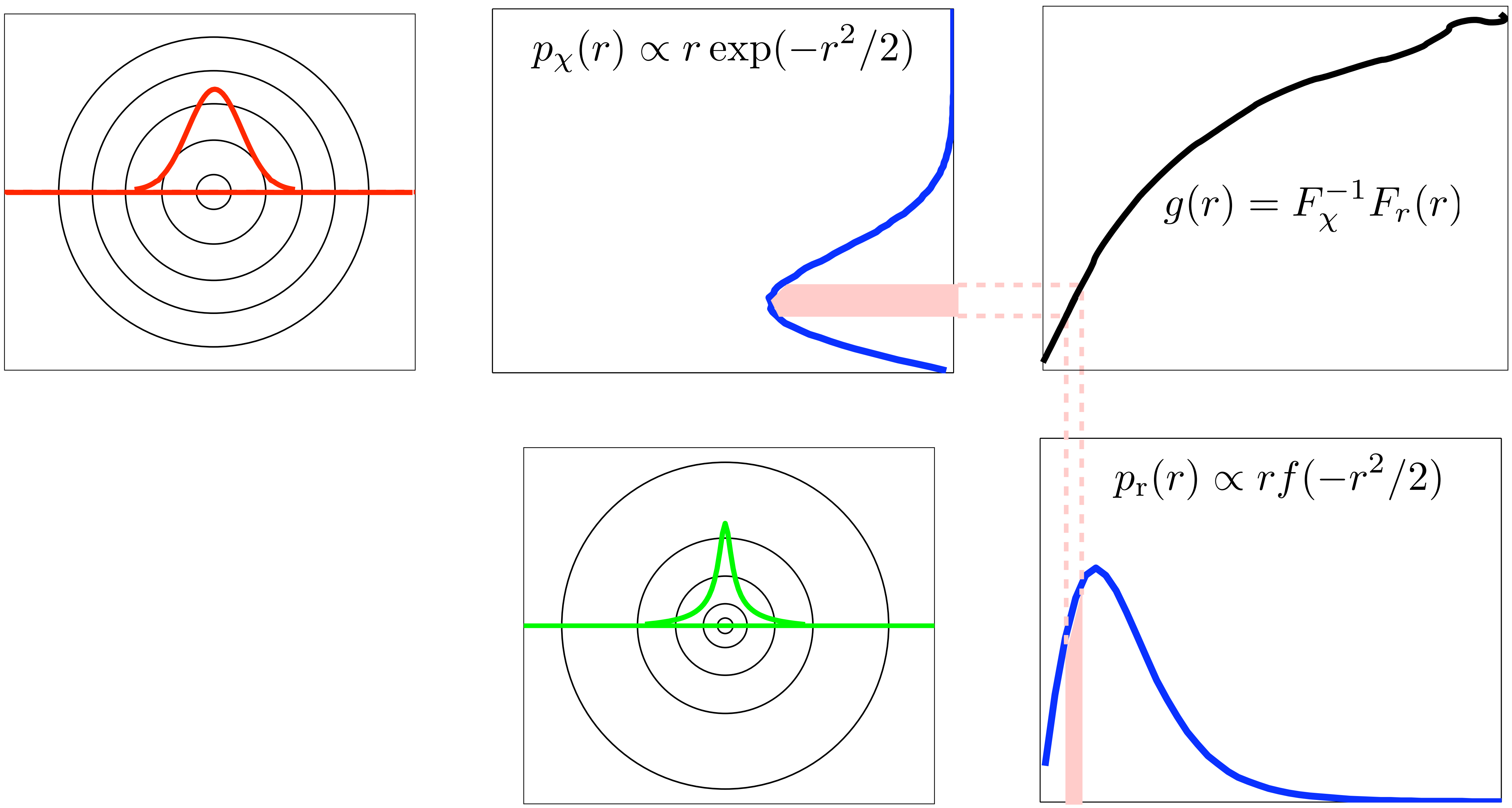
[Lyu & Simoncelli, 2009]

radial Gaussianization (RG)



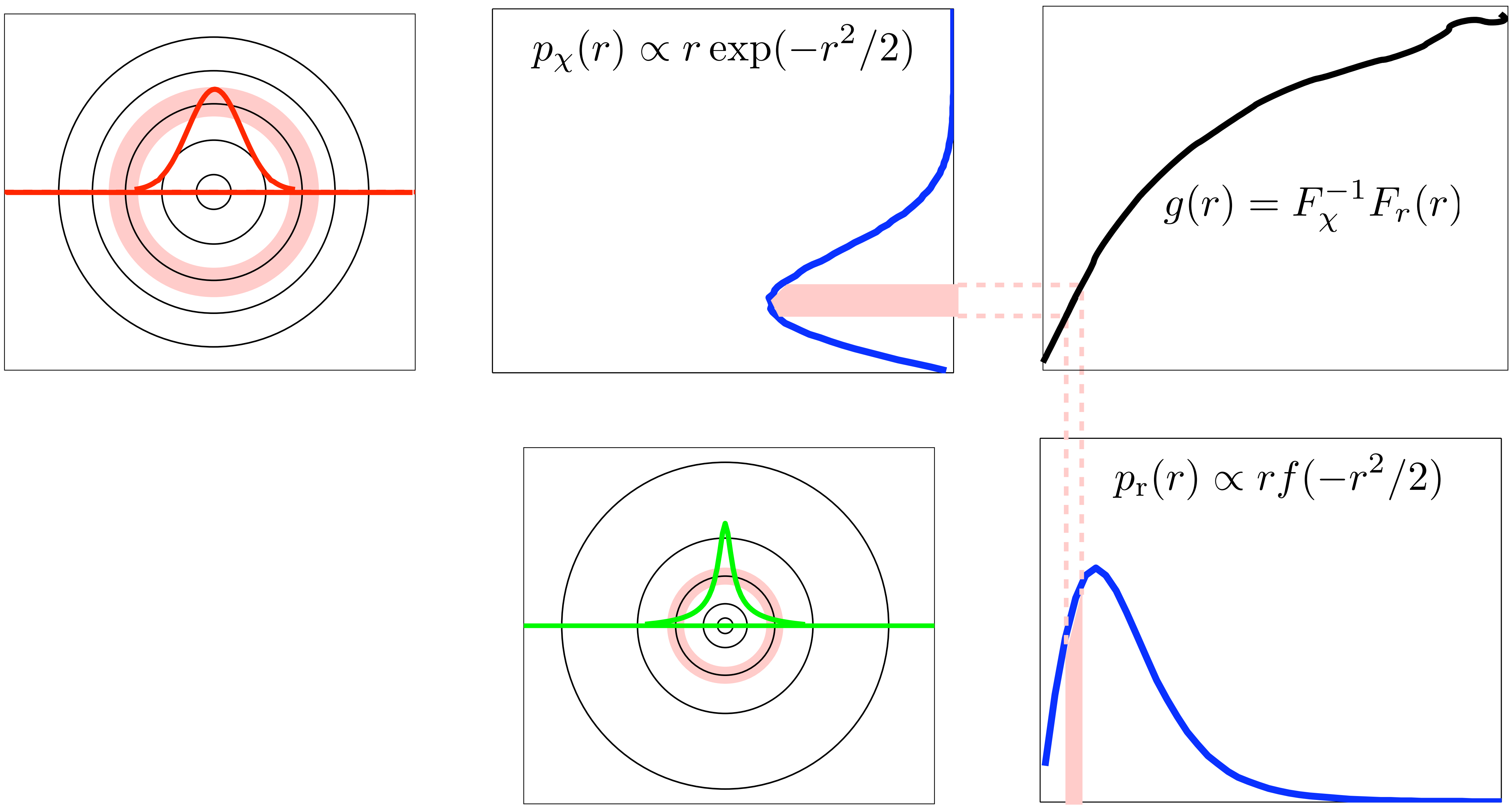
[Lyu & Simoncelli, 2009]

radial Gaussianization (RG)



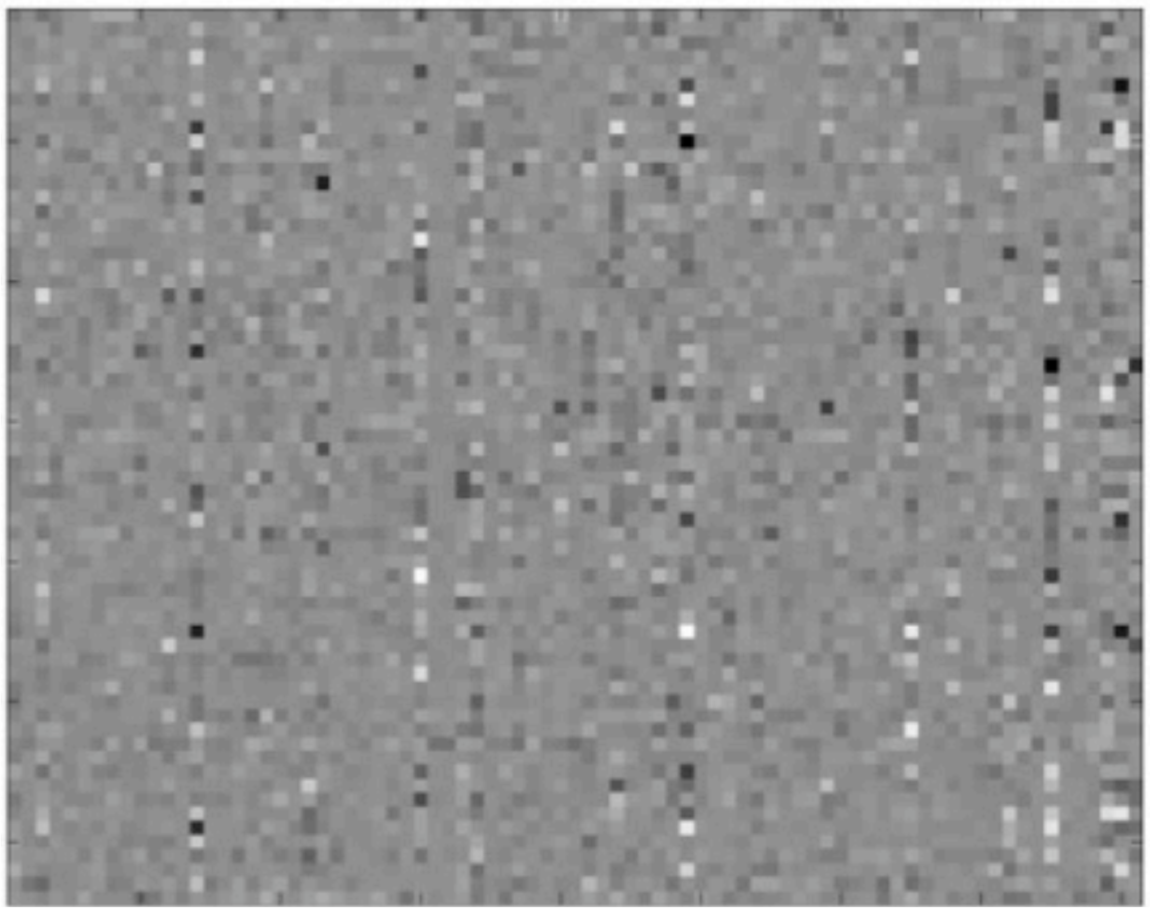
[Lyu & Simoncelli, 2009]

radial Gaussianization (RG)

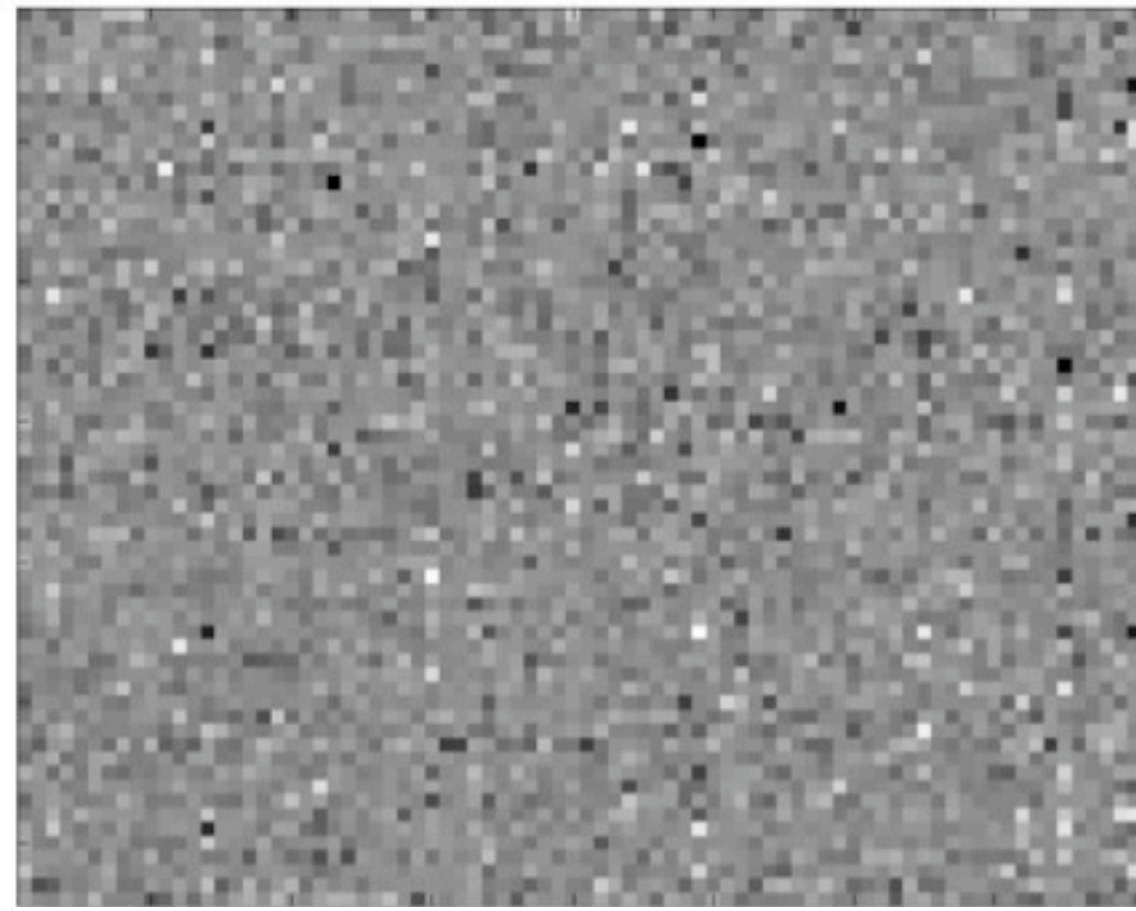


$$\vec{x}_{\mathrm{rg}} = \frac{g(\|\vec{x}_{\mathrm{wht}}\|)}{\|\vec{x}_{\mathrm{wht}}\|}\vec{x}_{\mathrm{wht}}$$

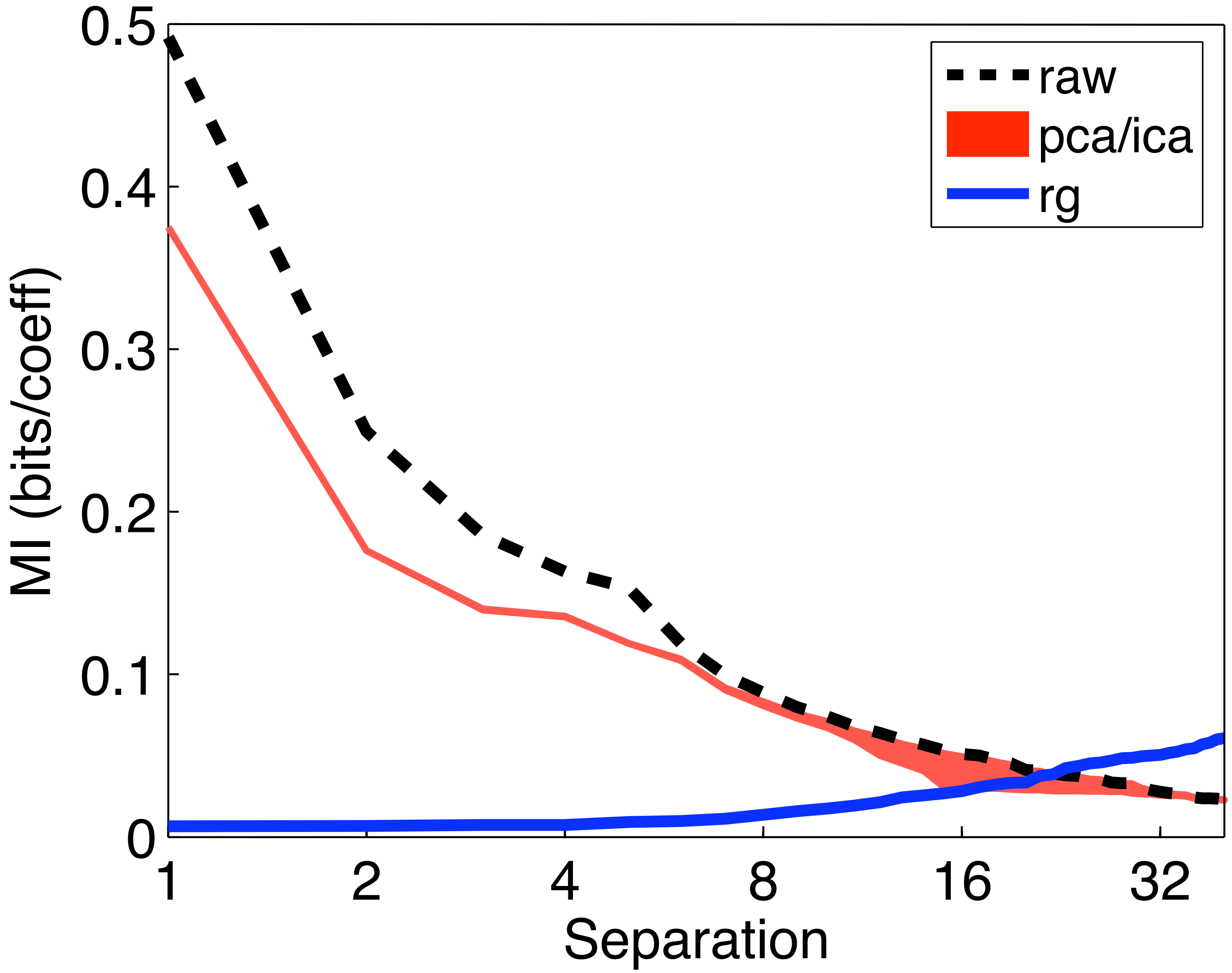
[Lyu & Simoncelli, 2009]



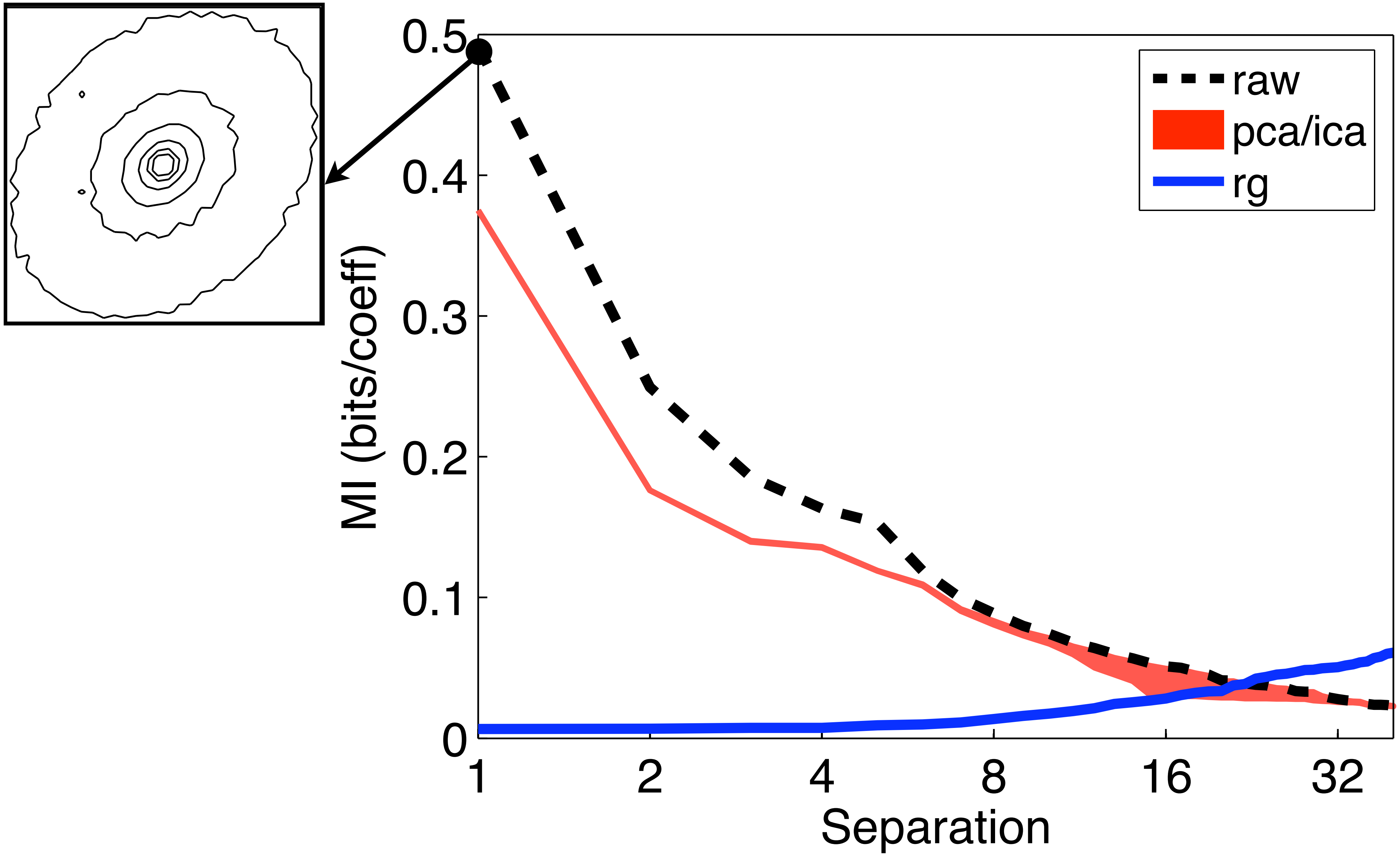
ICA coefficients



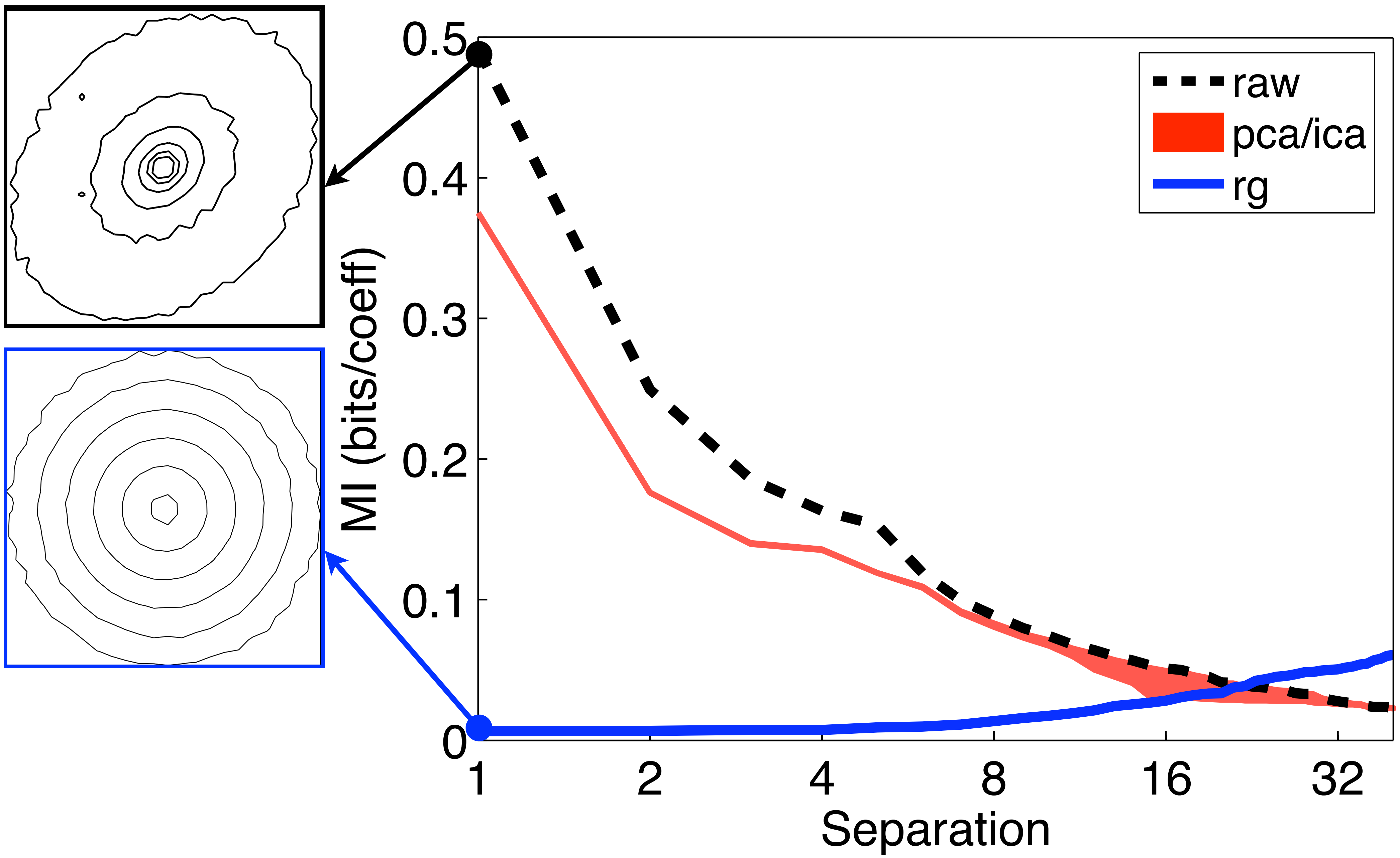
Radially factorized coefficients



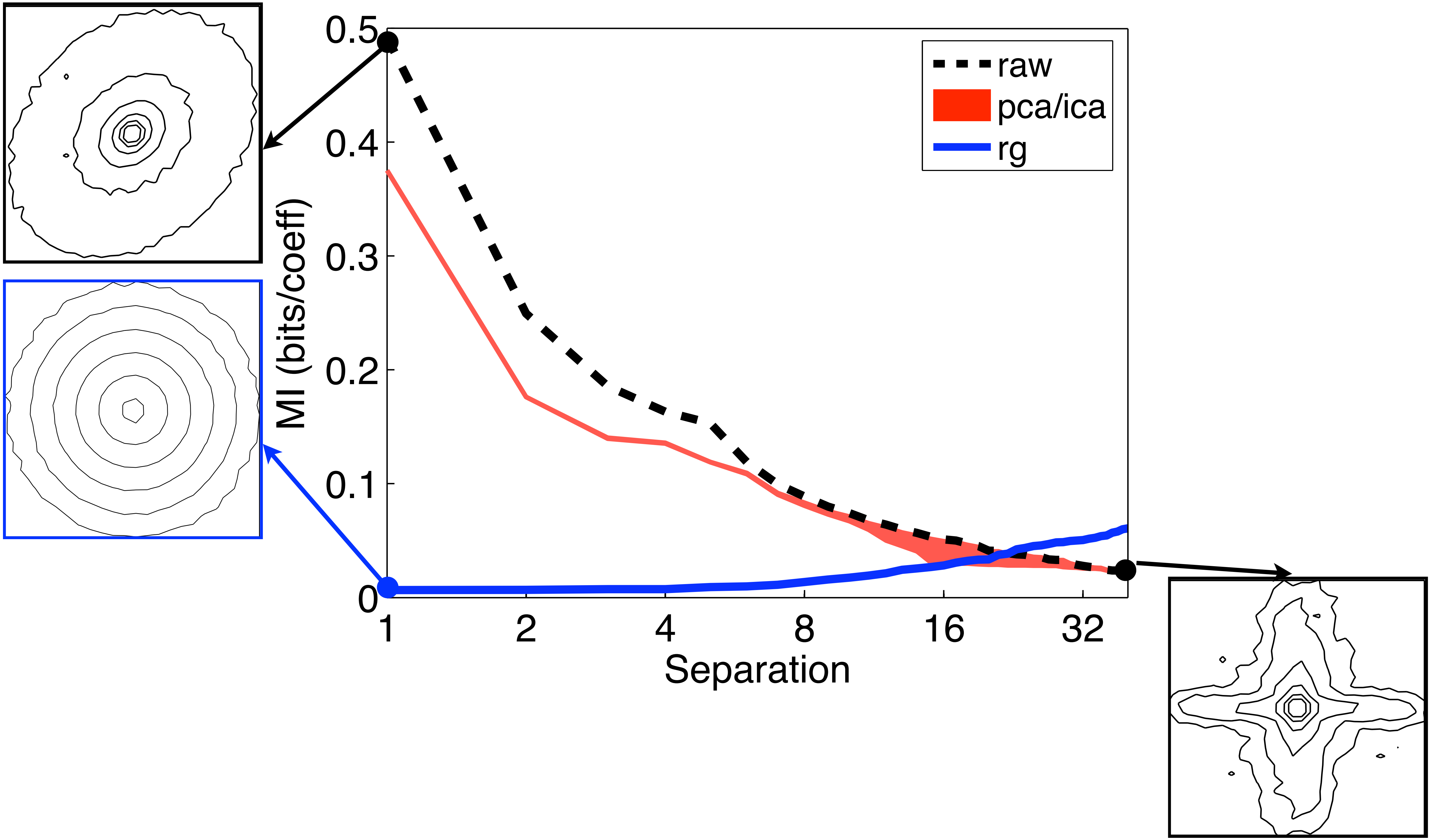
raw
pca/ica
rg
0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation

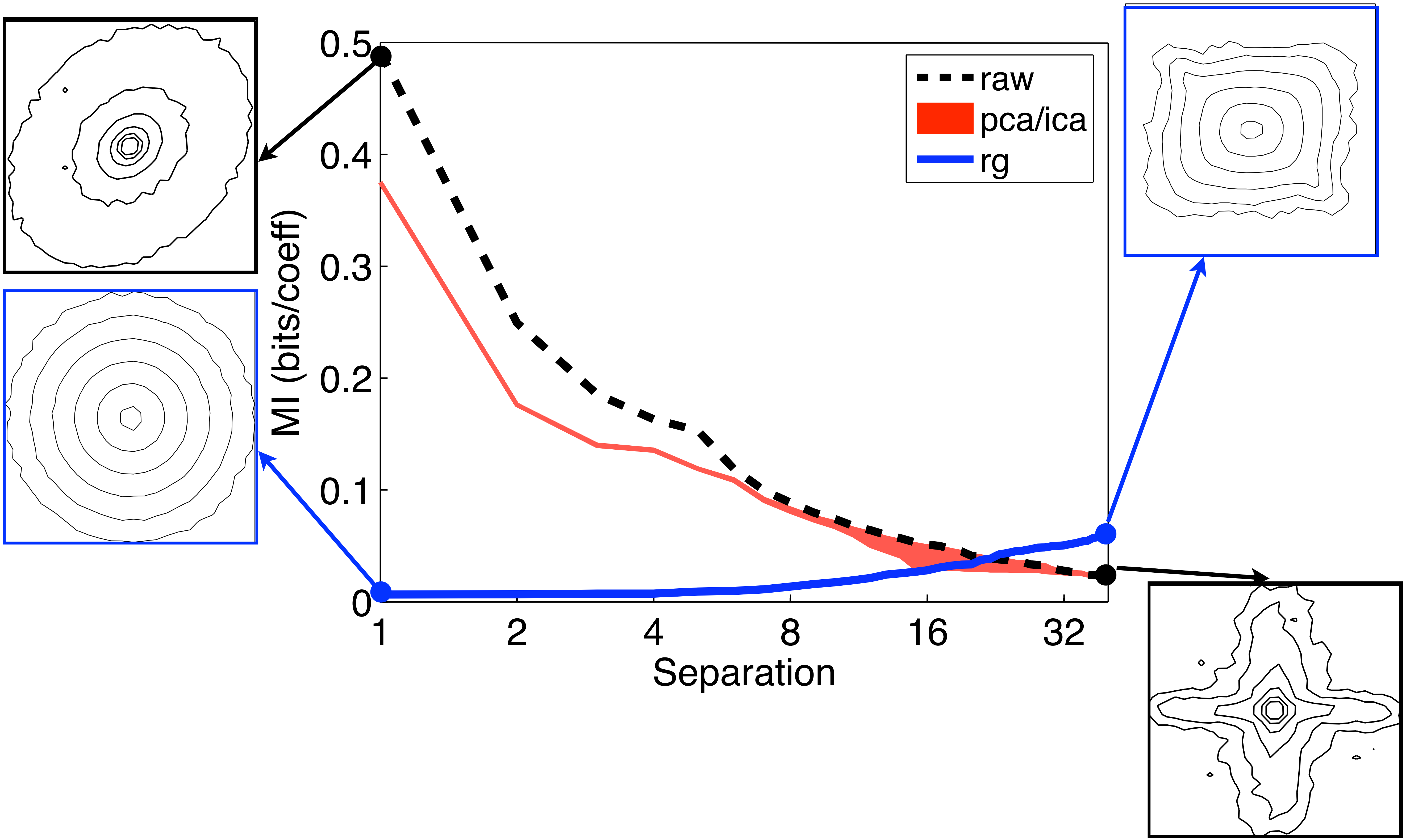


raw
pca/ica
rg
0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation

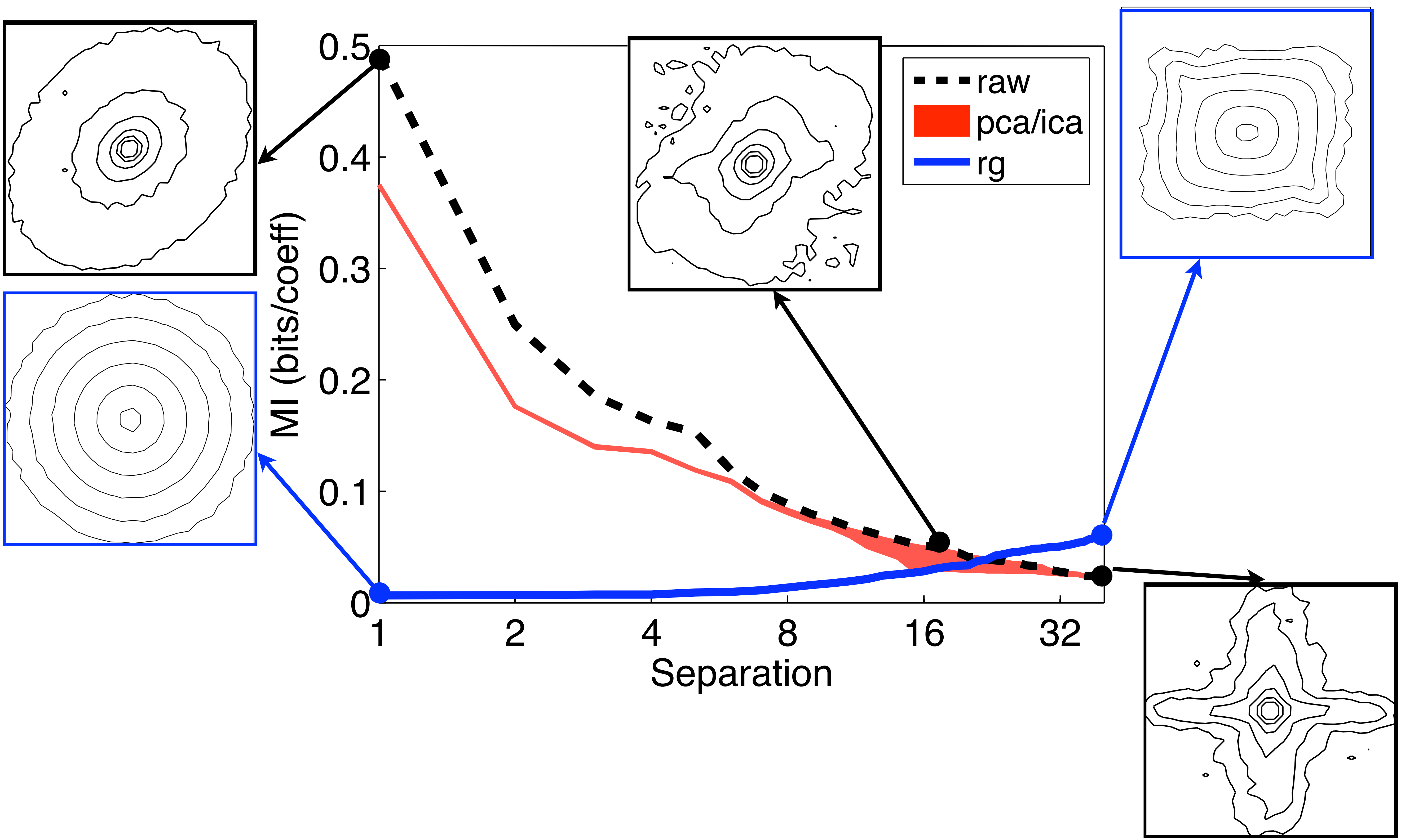


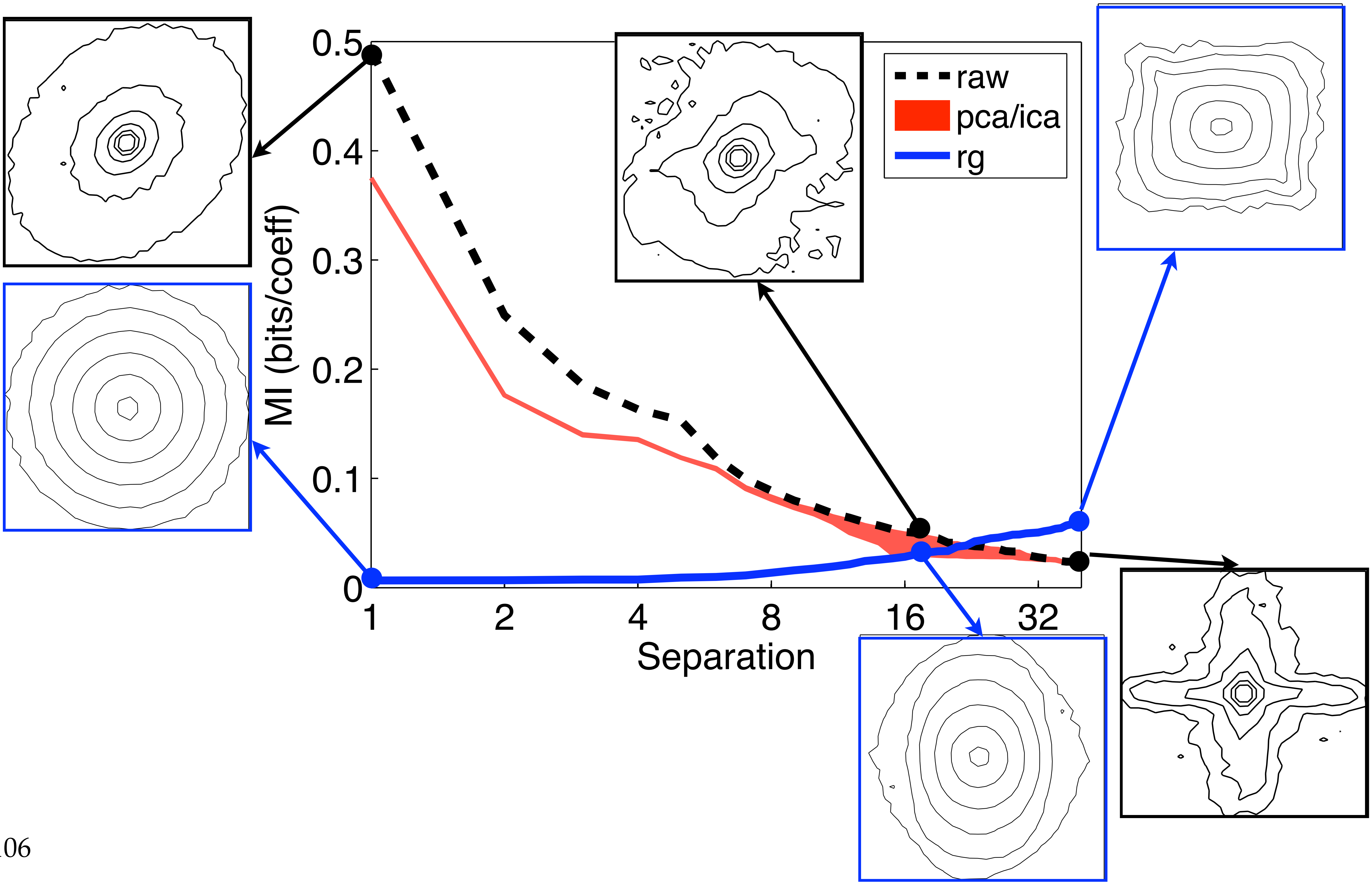
0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg

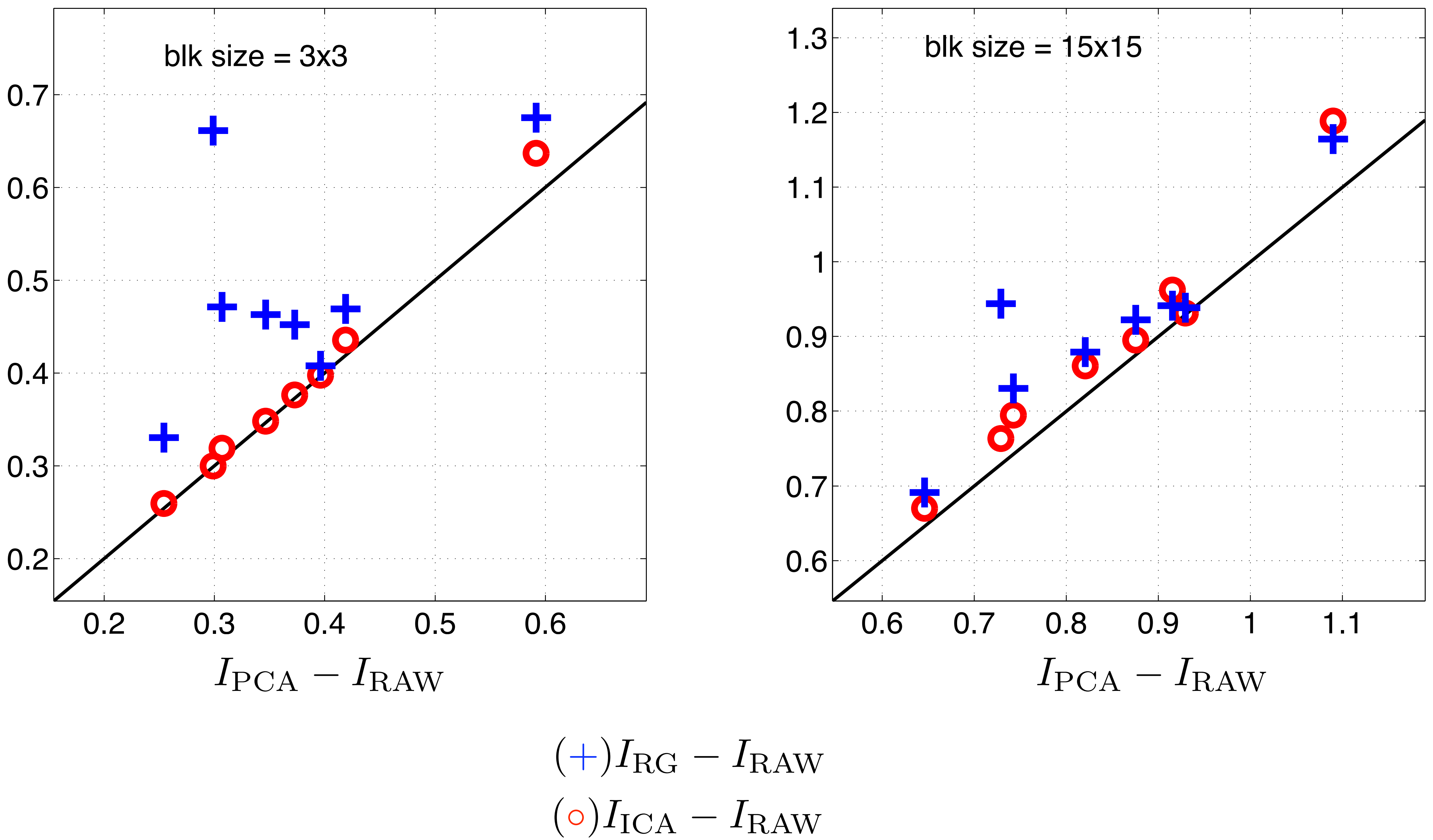

0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg



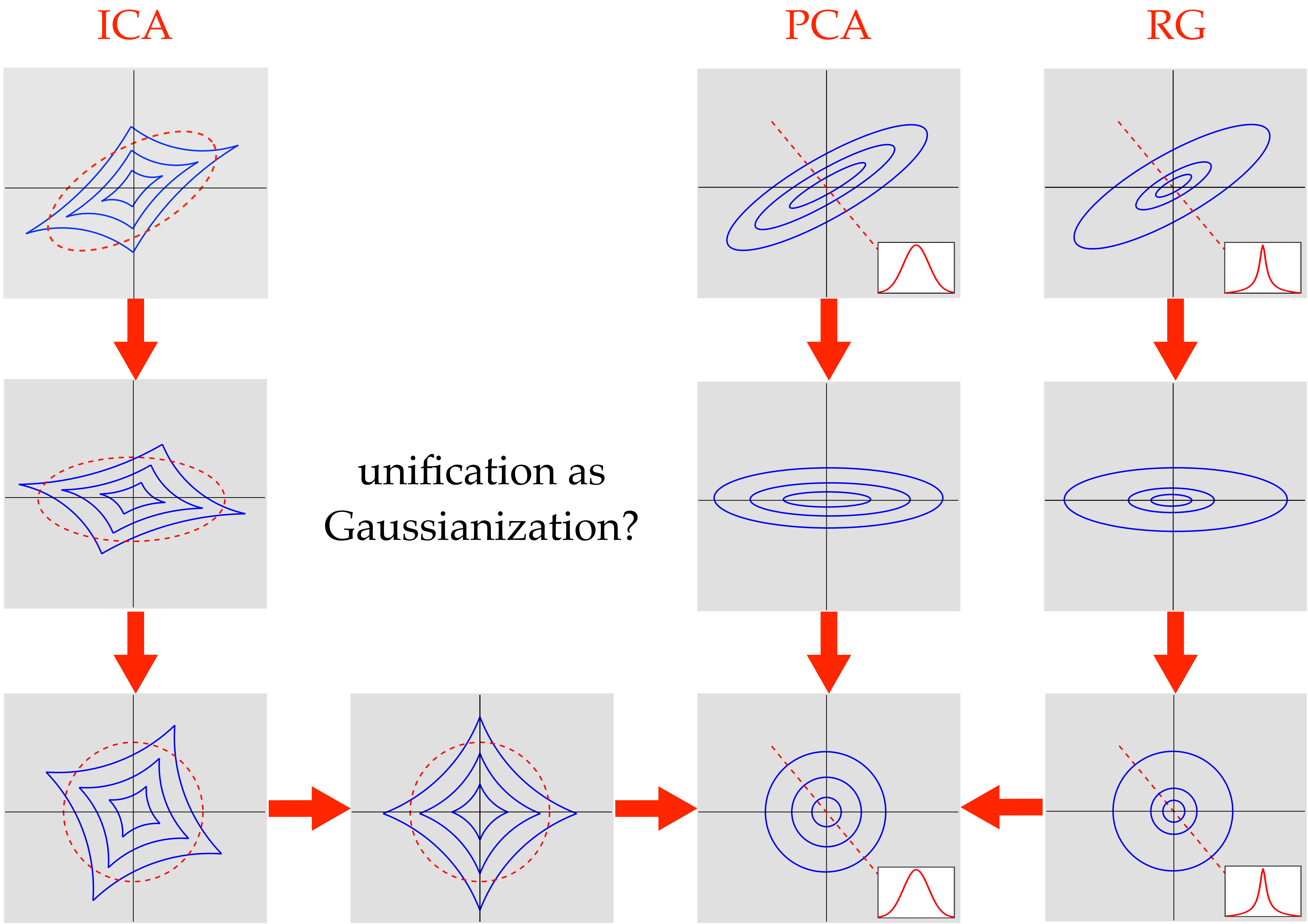
0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg


0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg

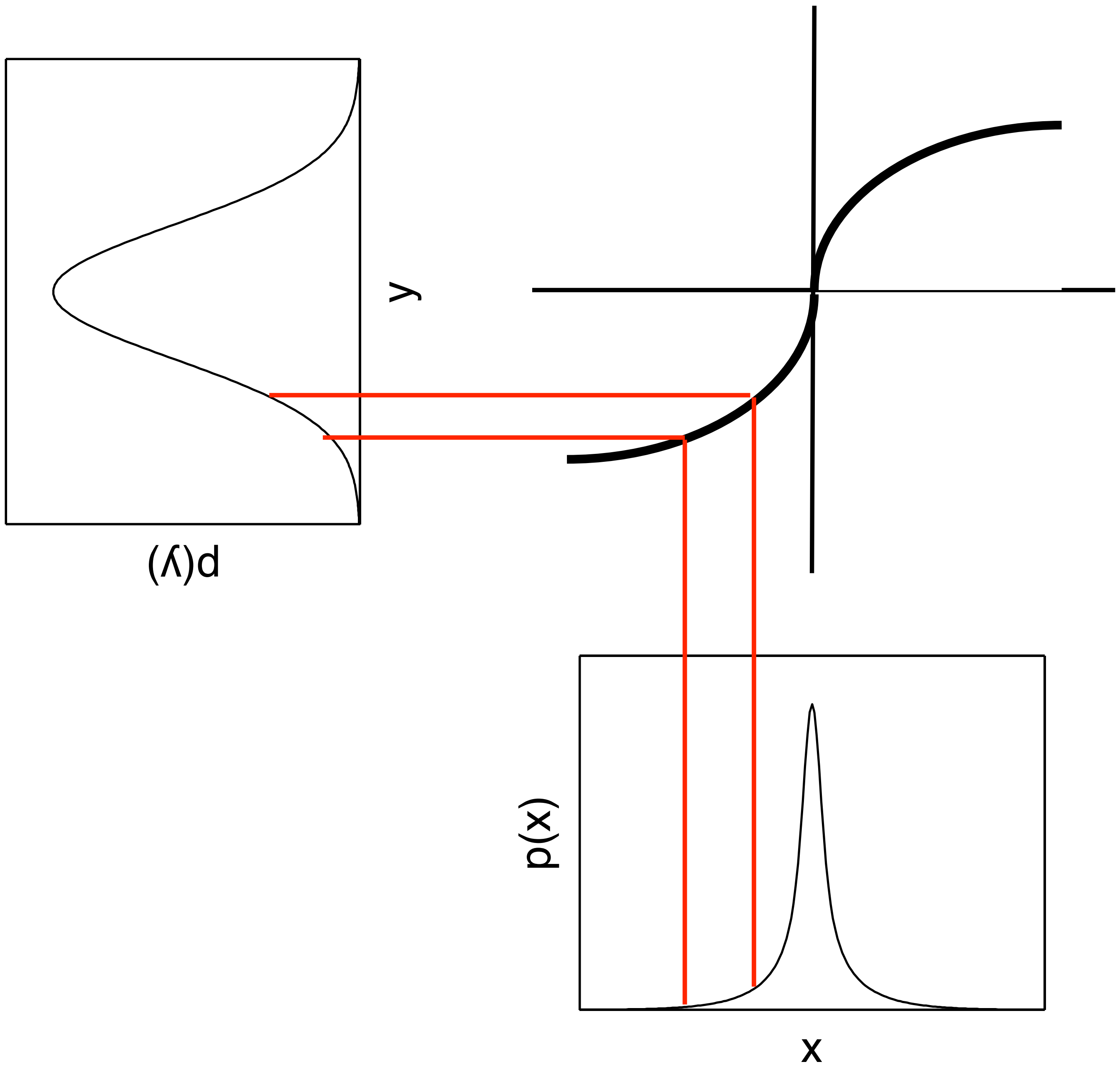

0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg



blocks of local mean removed pixel blocks of natural images

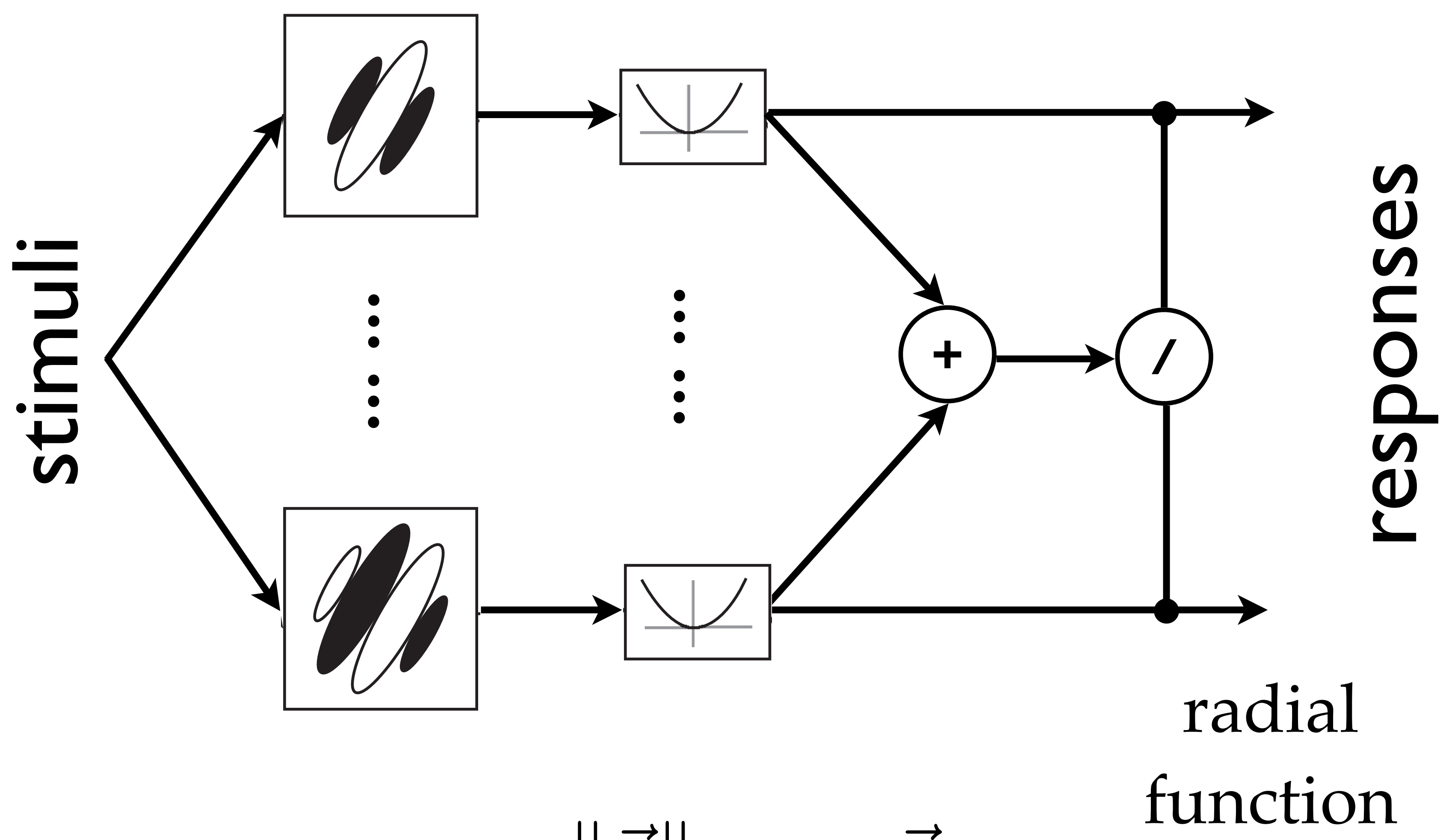

ICA
PCA
RG
unification as
Gaussianization?

marginal Gaussianization



divisive normalization

- nonlinear transform



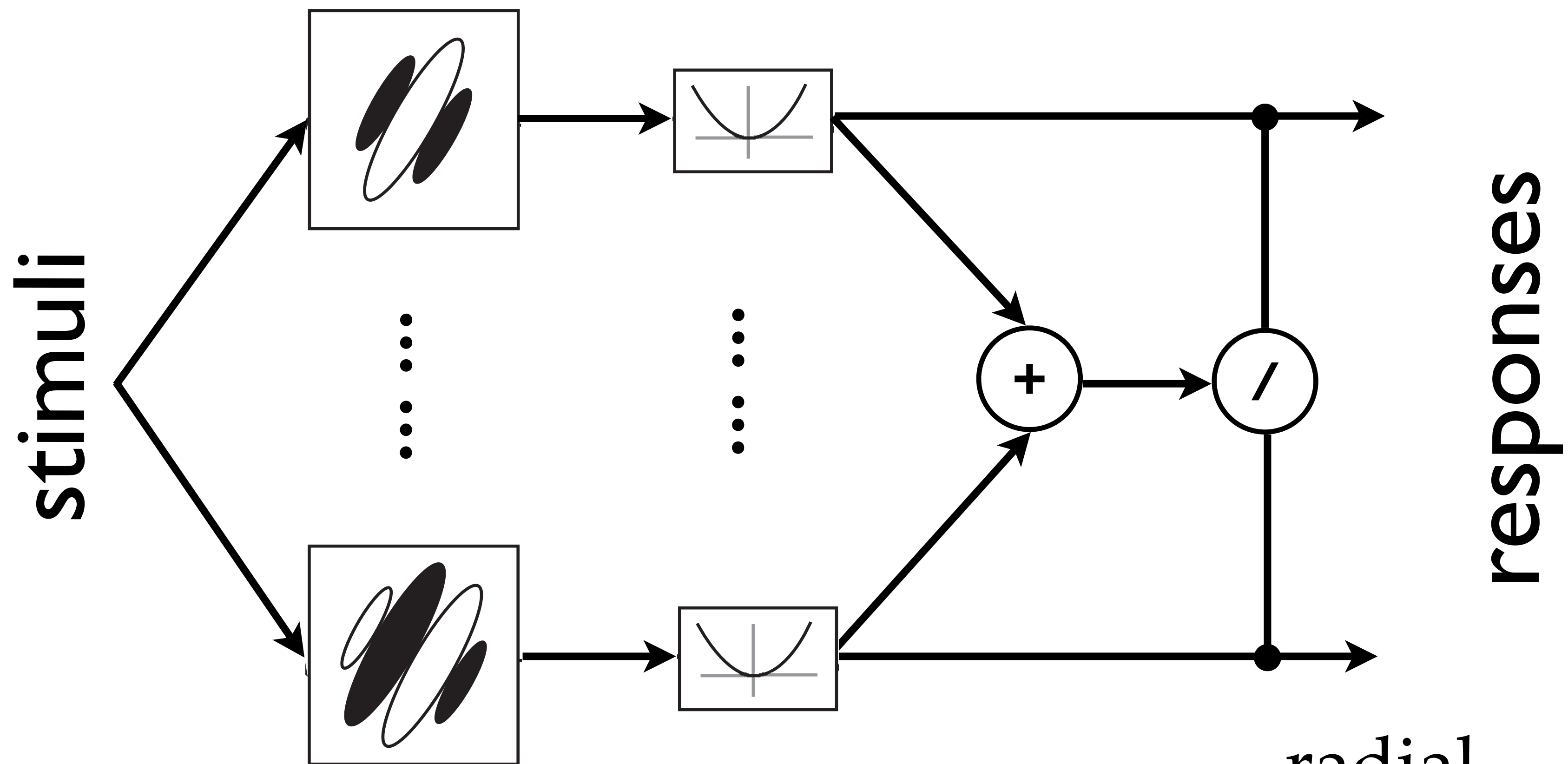
$$\vec{s} = \frac{\|\vec{x}\|}{\sqrt{\alpha + \|\vec{x}\|^2}} \frac{\vec{x}}{\|\vec{x}\|}$$

divisive normalization

- observation
  - visual cortex [Heeger, 1991]
  - retina/LGN [Caradini etal. 2008]
  - auditory [Schwartz & Simoncelli, 1999]
  - olfactory [Wilson etal, 2010]
- underlying principle
  - dynamic gain control
  - dependency reduction [Schwartz & Simoncelli, 2001]

divisive normalization

- nonlinear transform

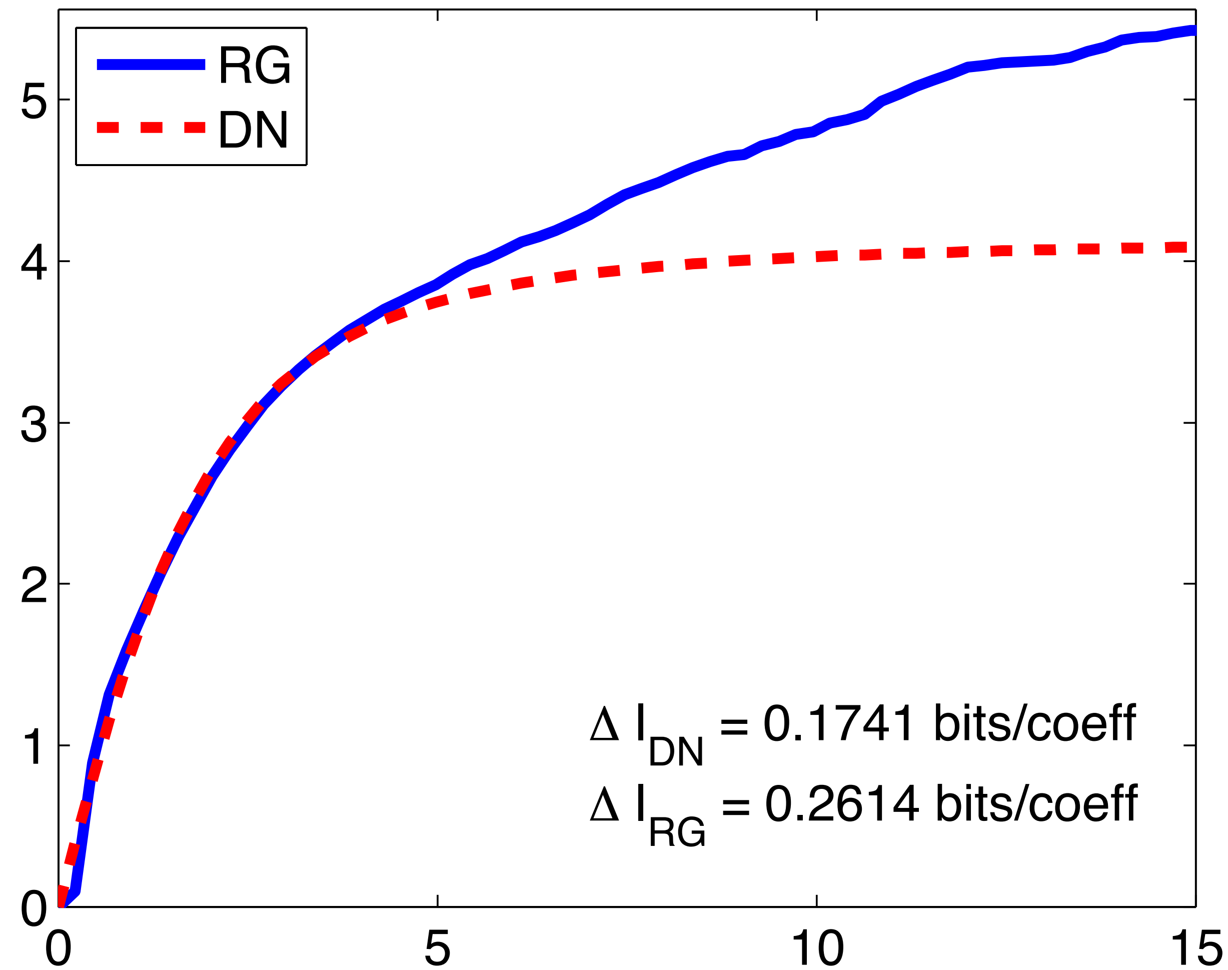


radial function

$$\vec{s} = \boxed{\frac{||\vec{x}||}{\sqrt{\alpha + ||\vec{x}||^2}}} \frac{\vec{x}}{||\vec{x}||}$$

divisive normalization

- comparing the two radial transforms

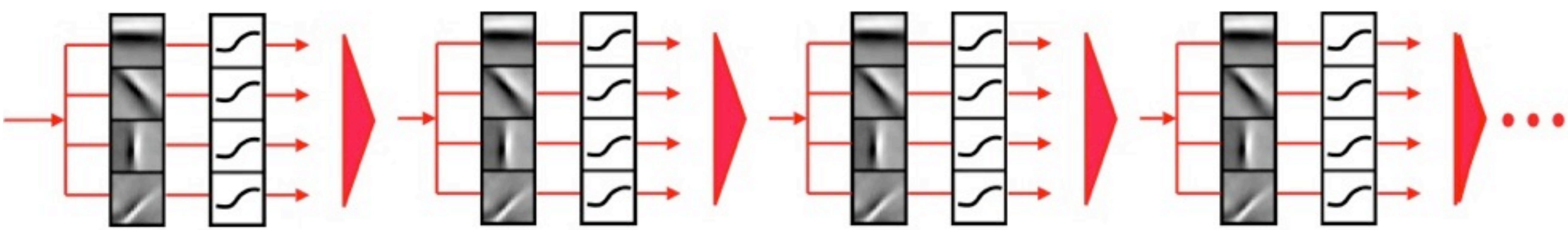


summary

- in band-pass filter domain, we observe
  - non-Gaussian marginal densities
  - elliptically symmetric joint densities
- observations lead to elliptically symmetric models
- ESD models lead to nonlinear radial Gaussianization
- extended to L_p elliptically symmetric models
  - [Sinz & Bethge, 2009]
- not sufficient
  - not effective for longer-range dependencies

next step: building hierarchies

- hierarchical representations
  - iterative Gaussianization/hierarchical ICA/bio-inspired
    - [Chen & Gopinath, 2000; Shan etal, 2007; Karklin & Lewicki, 2003; Serre & Poggio, 2006]



- hierarchical model
  - DBN type models, convolutional net

summary

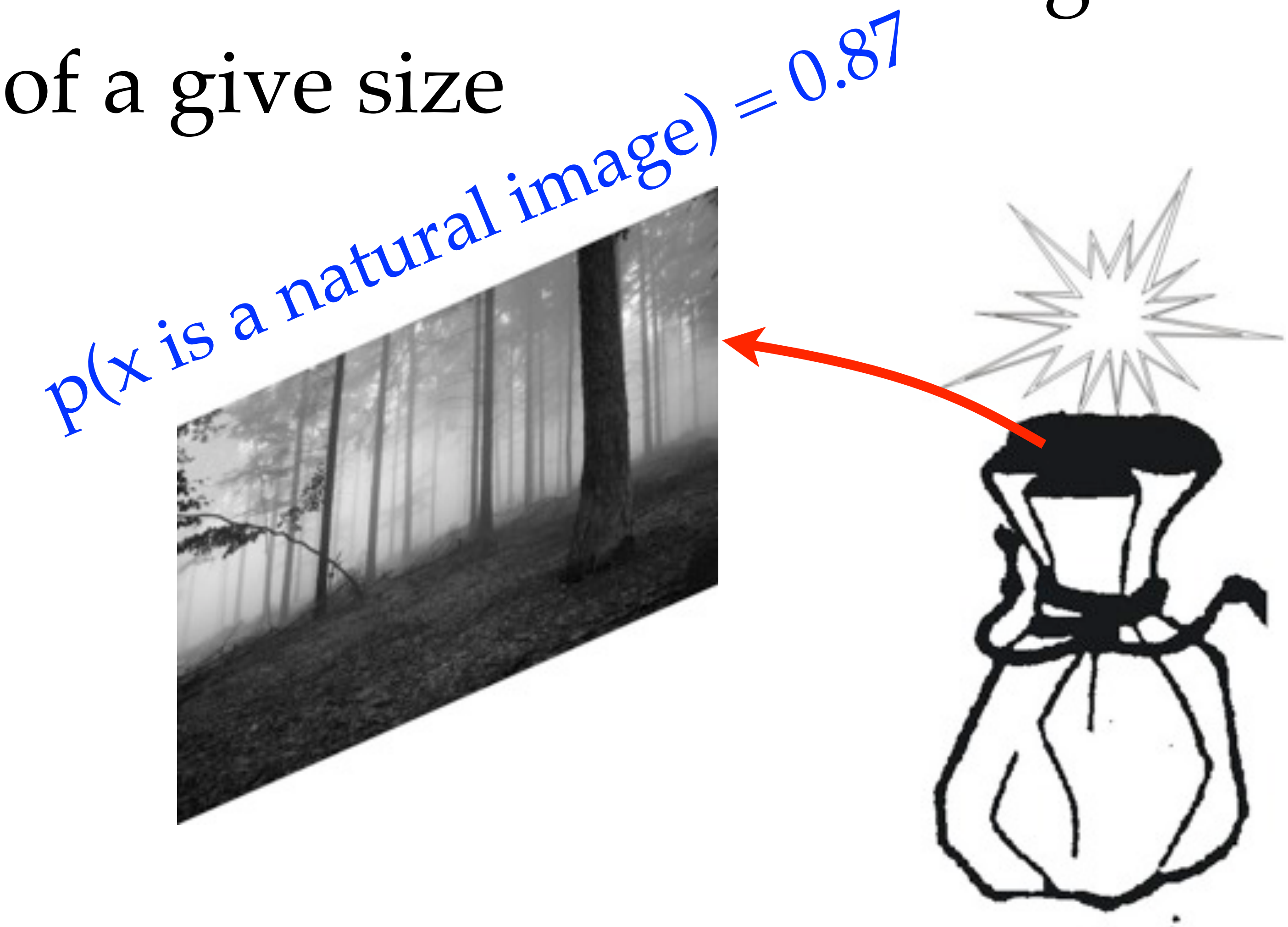
- natural images are special in the space of all possible images and have regular statistical properties
- these properties can be captured using representation and statistical models
  - dependency reduction
  - maximum entropy with constraints
- key question: where to put the complexity

afterthoughts

- are the observed properties real or results of "*artifacts of the lens through which we view the data*"

we believe but cannot prove ...

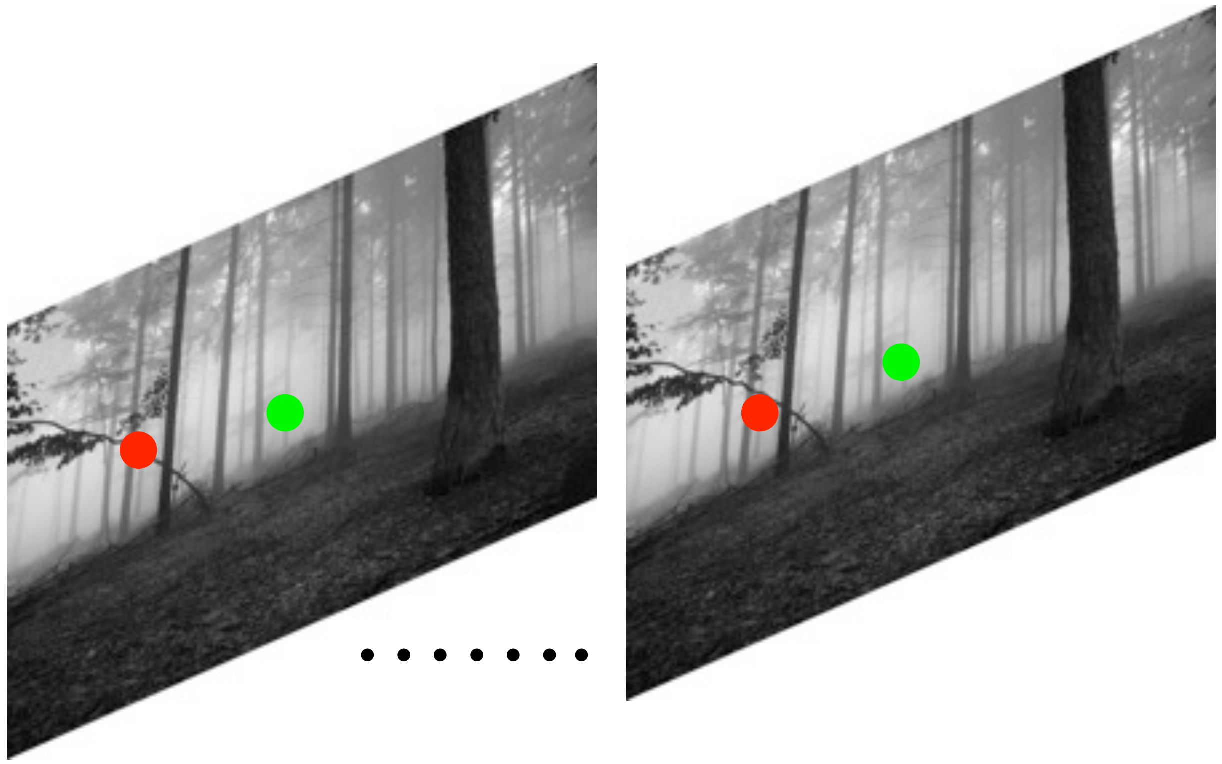
- there is a probabilistic model over natural images in the space of all images of a give size



p(x)

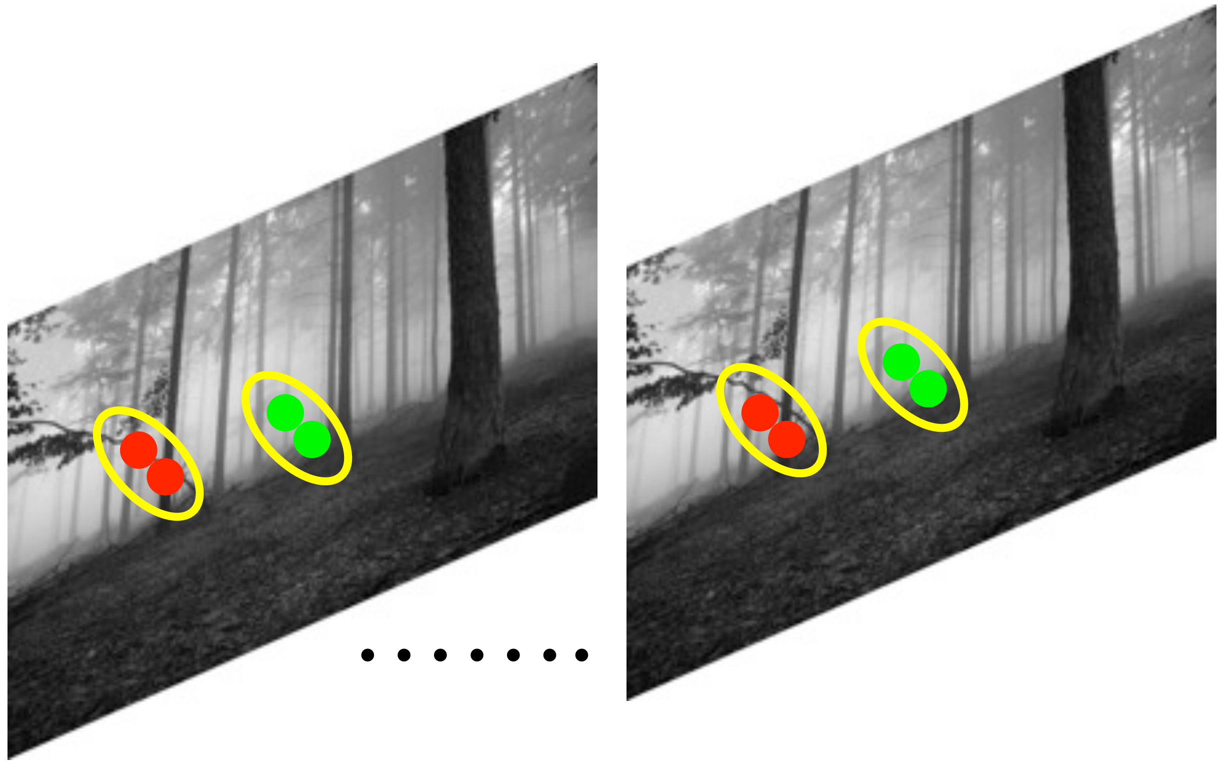
we believe but cannot prove ...

- there is a probability measure over natural images in the space of all images of a give size
- this probability measure has invariance
  - translation invariance (a.k.a., stationary, homogeneous)
    - marginal densities have no dependency with spatial locations



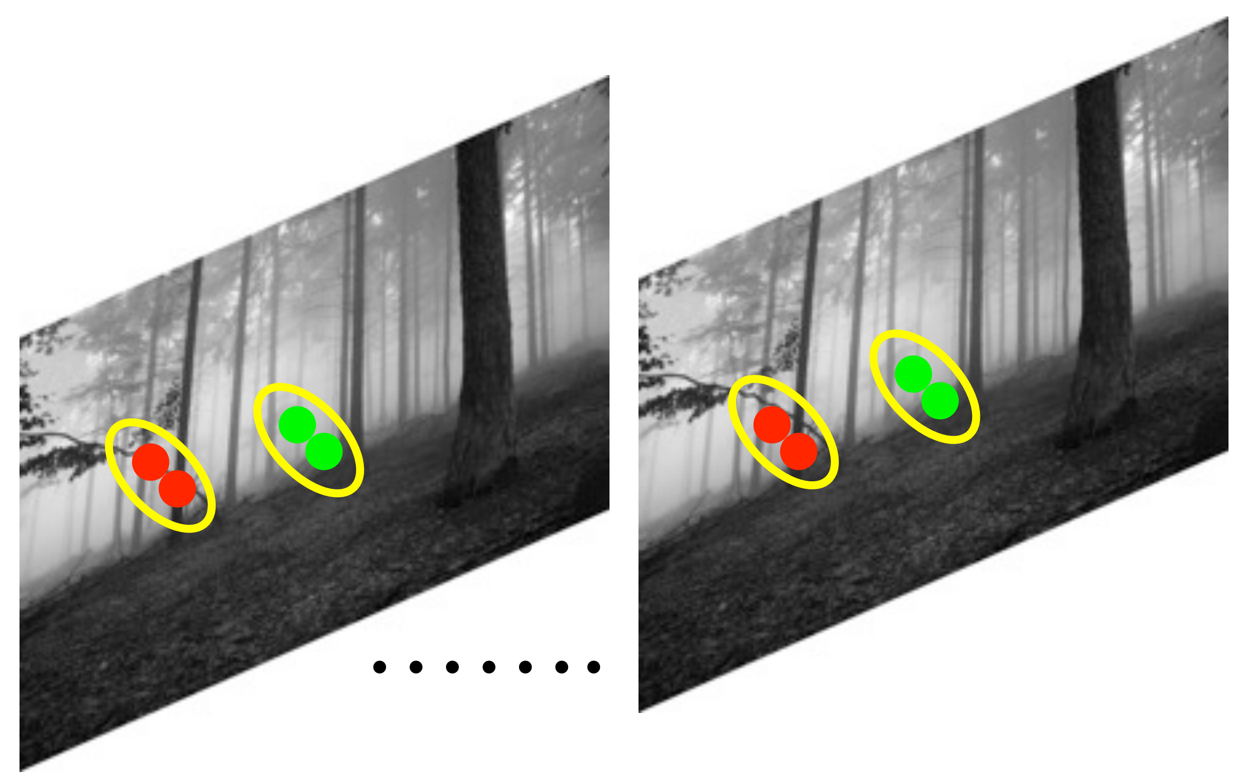
we believe but cannot prove ...

- there is a probability measure over natural images in the space of all images of a give size
- this probability measure has invariance
  - translation invariance (a.k.a., stationary, homogeneous)
    - marginal densities have no dependency with spatial locations
    - joint densities have no dependency with spatial locations



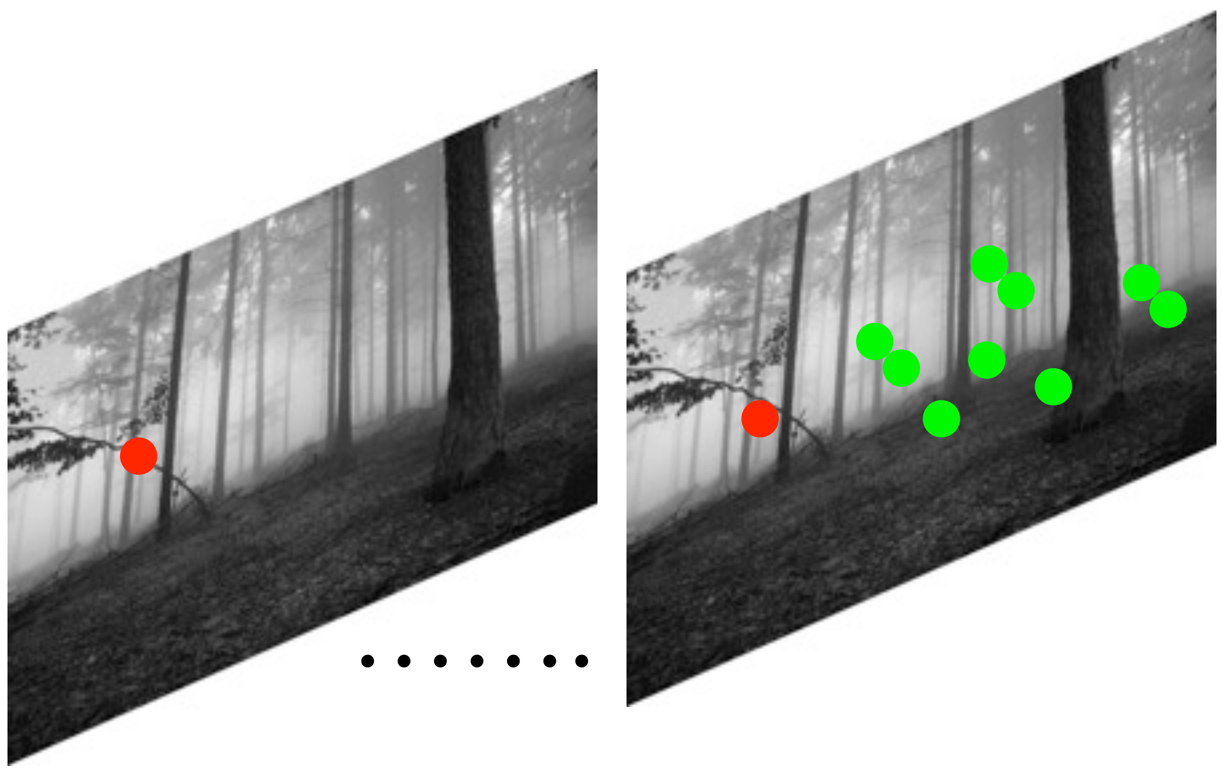
we believe but cannot prove ...

- there is a probability measure over natural images in the space of all images of a give size
- this probability measure has invariance
  - translation invariance (a.k.a., stationary, homogeneous)
    - marginal densities have no dependency with spatial locations
    - joint densities have no dependency with spatial locations
    - practical issue: proper boundary handling



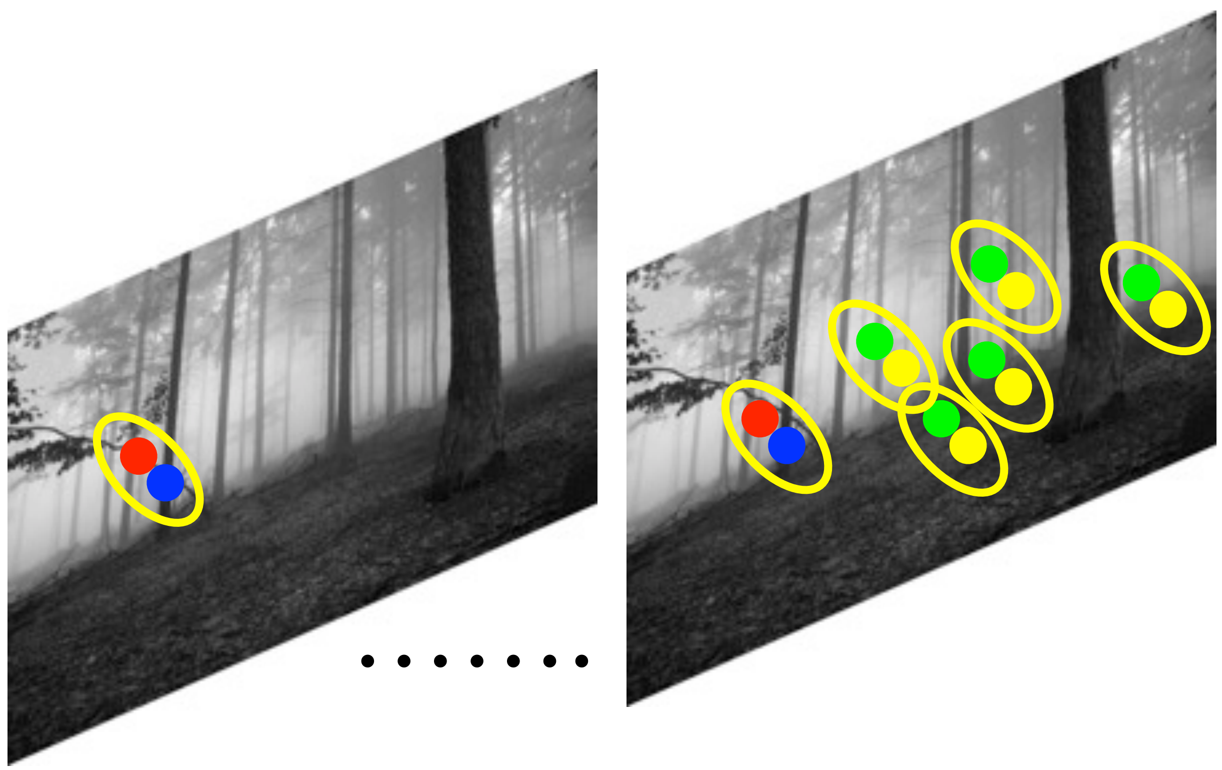
we believe but cannot prove ...

- there is a probability measure over natural images in the space of all images of a give size
- this probability measure has invariance
  - translation invariance (a.k.a., stationary, homogeneous)
  - (empirical) ergodic
    - ensemble average = spatial average
    - ensemble marginal = spatial marginal

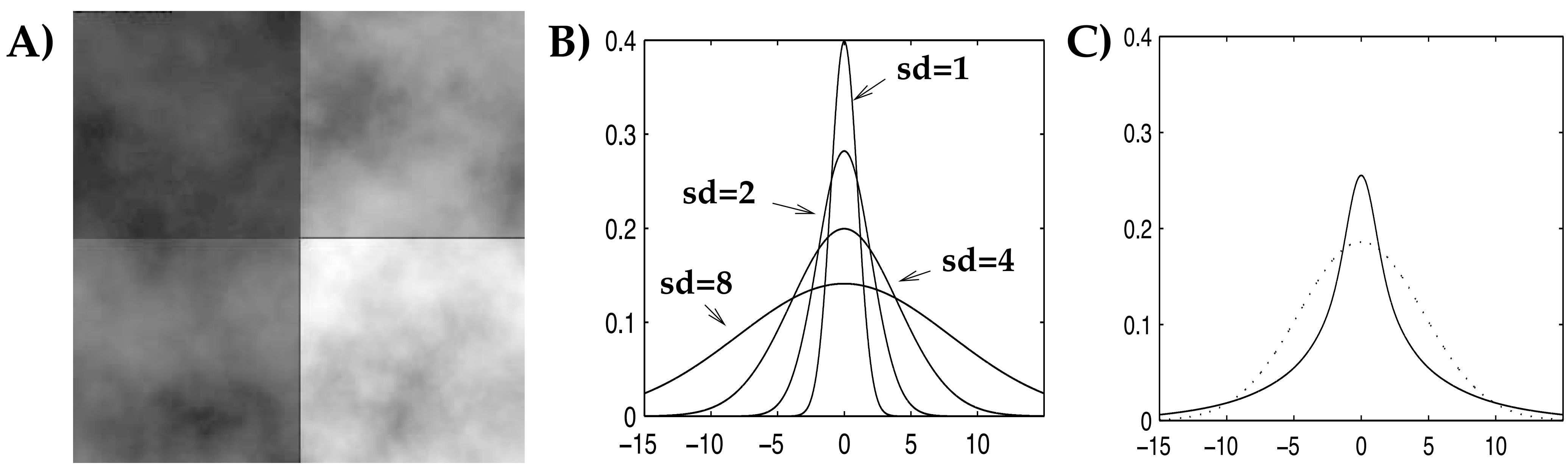


we believe but cannot prove ...

- there is a probability measure over natural images in the space of all images of a give size
- this probability measure has invariance
  - translation invariance (a.k.a., stationary, homogeneous)
  - (empirical) ergodic
    - ensemble average = spatial average
    - ensemble joint = spatial joint



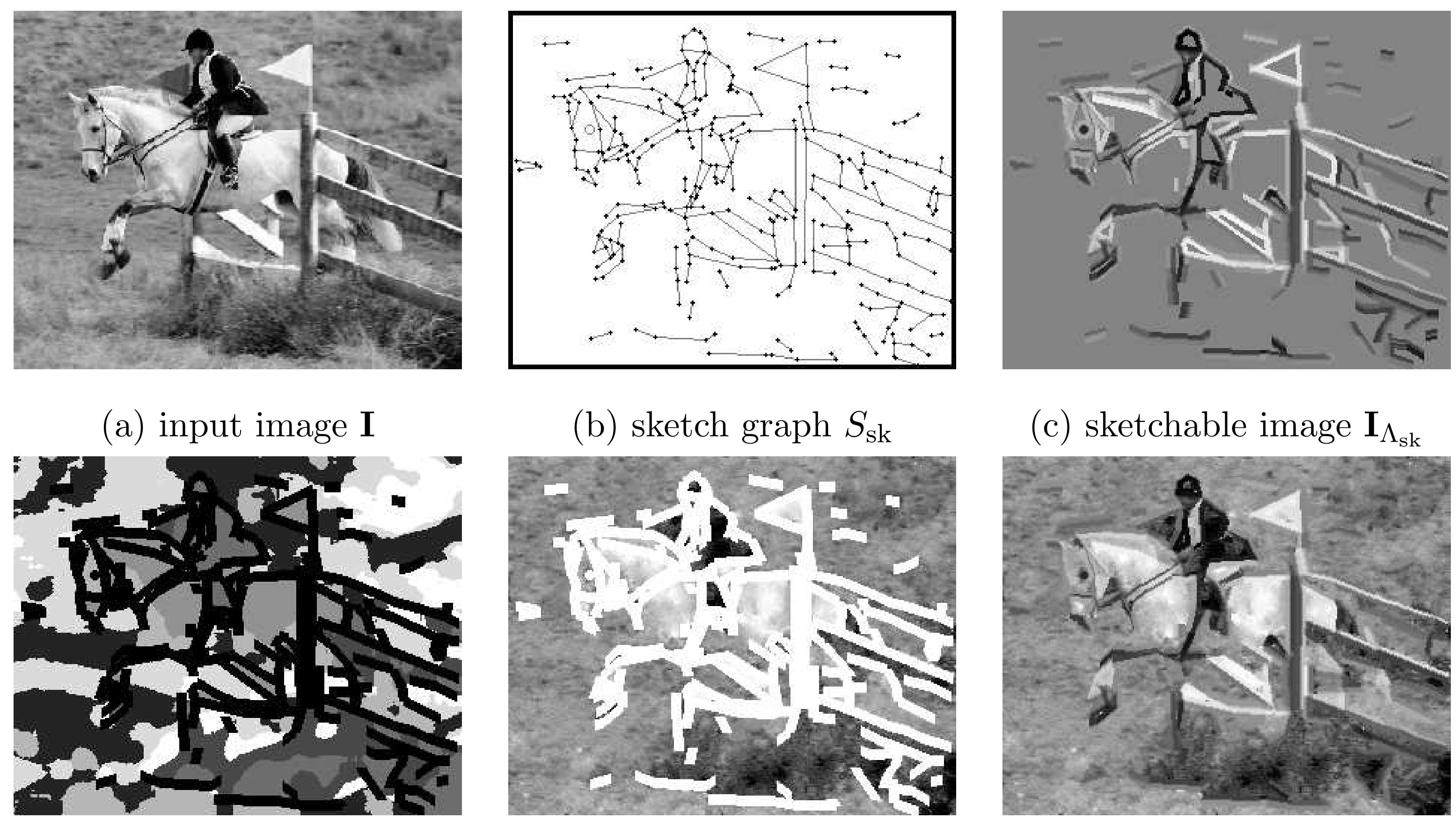
afterthoughts



[Baddeley 1996]

afterthoughts

- image specific model
  - CRF image models for denoising, directly model p(x | y)
    - [Tappen etal., 2007; 2009]
  - primary sketch model
    - [Guo, Zhu and Wu, 2007]



(a) input image $\mathbf{I}$ (b) sketch graph S_{sk} (c) sketchable image $\mathbf{I}_{\Lambda_{\mathrm{sk}}}$

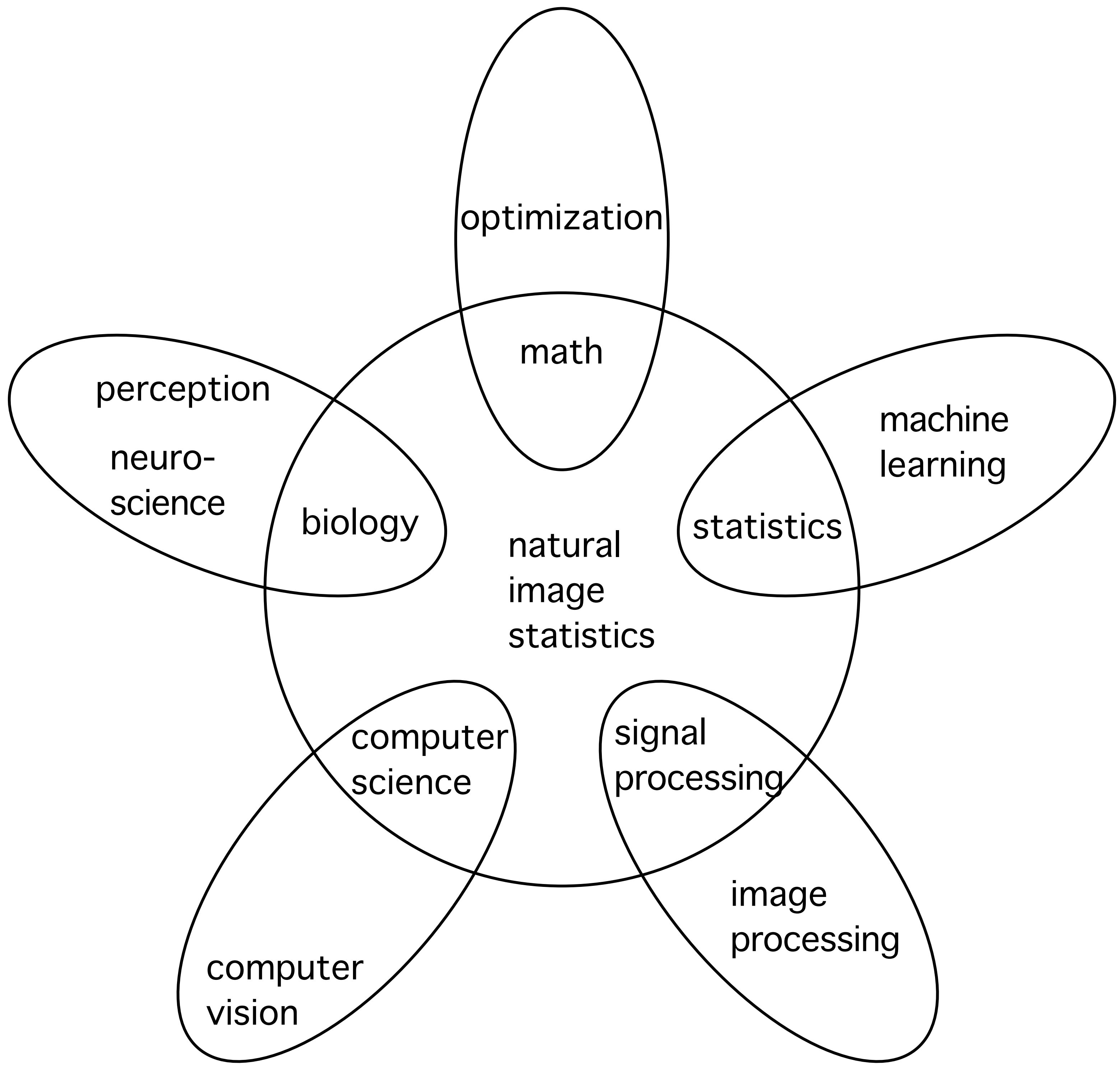
(d) texture regions $S_{\Lambda_{\mathrm{nsk}}}$ (e) synthesized textures $\mathbf{I}_{\Lambda_{\mathrm{nsk}}}$ (f) synthesized image $\mathbf{I}^{\mathrm{syn}}$

big question marks

- what are natural images, anyway?

ironically, white noises are "natural" as they are the result of cosmic radiations

- naturalness is subjective


optimization
math
perception
neuro-
science
biology
machine
learning
statistics
natural
image
statistics
computer
science
signal
processing
image
processing
computer
vision

want to know more?

- D. L. Ruderman. *The statistics of natural images*. Network: Computation in Neural Systems, 5:517–548, 1996.
- E. P. Simoncelli and B. Olshausen. *Natural image statistics and neural representation*. Annual Review of Neuroscience, 24:1193–1216, 2001.
- S.-C. Zhu. *Statistical modeling and conceptualization of visual patterns*. IEEE Trans. PAMI, 25(6), 2003.
- A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu. *On advances in statistical modeling of natural images*. J. Math. Imaging and Vision, 18(1): 17–33, 2003.
- E. P. Simoncelli. *Statistical modeling of photographic images*. Handbook of Image and Video Processing. Academic Press, 2005.
- A. Hyvärinen, J. Hurri, and P. O. Hoyer. *Natural Image Statistics: A probabilistic approach to early computational vision*. Springer, 2009.


thank you