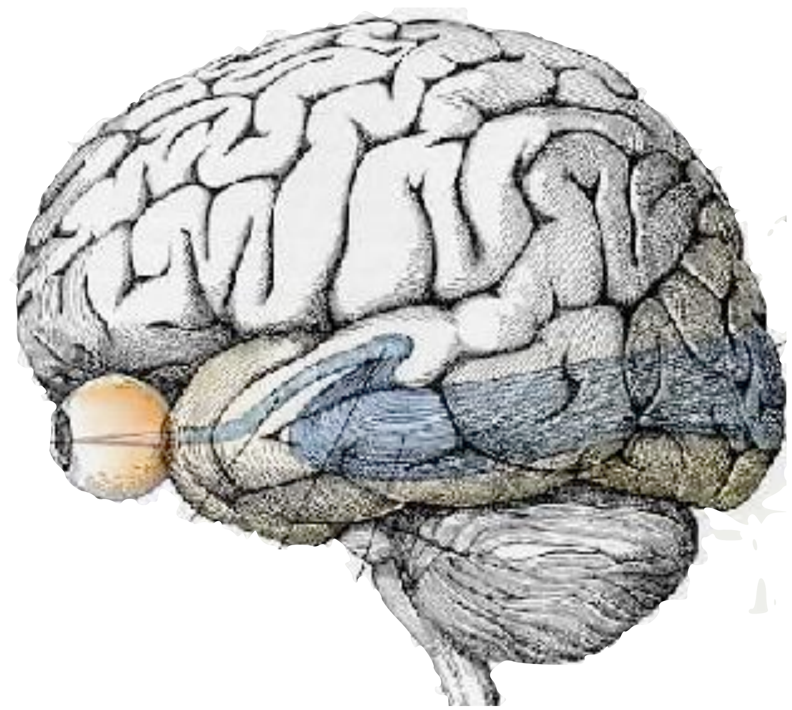
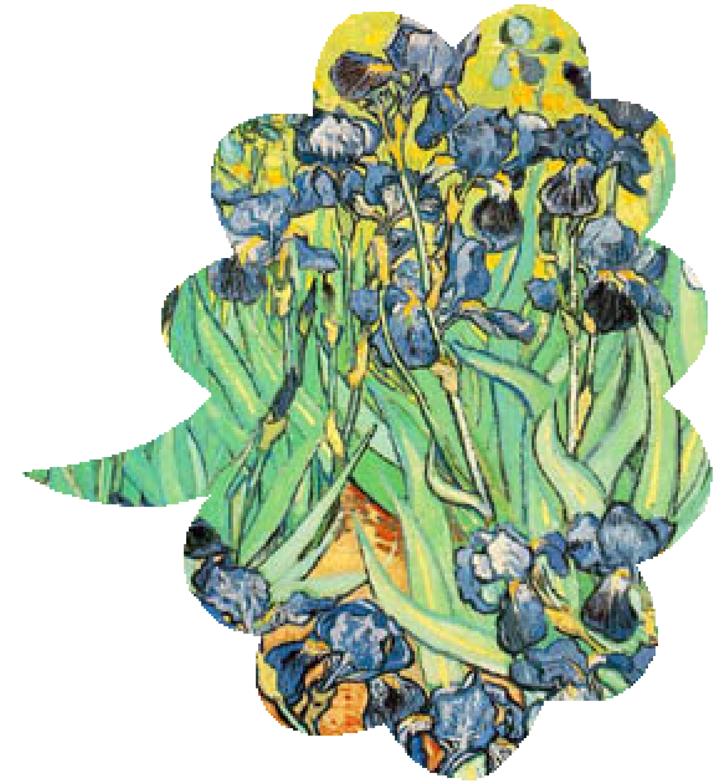
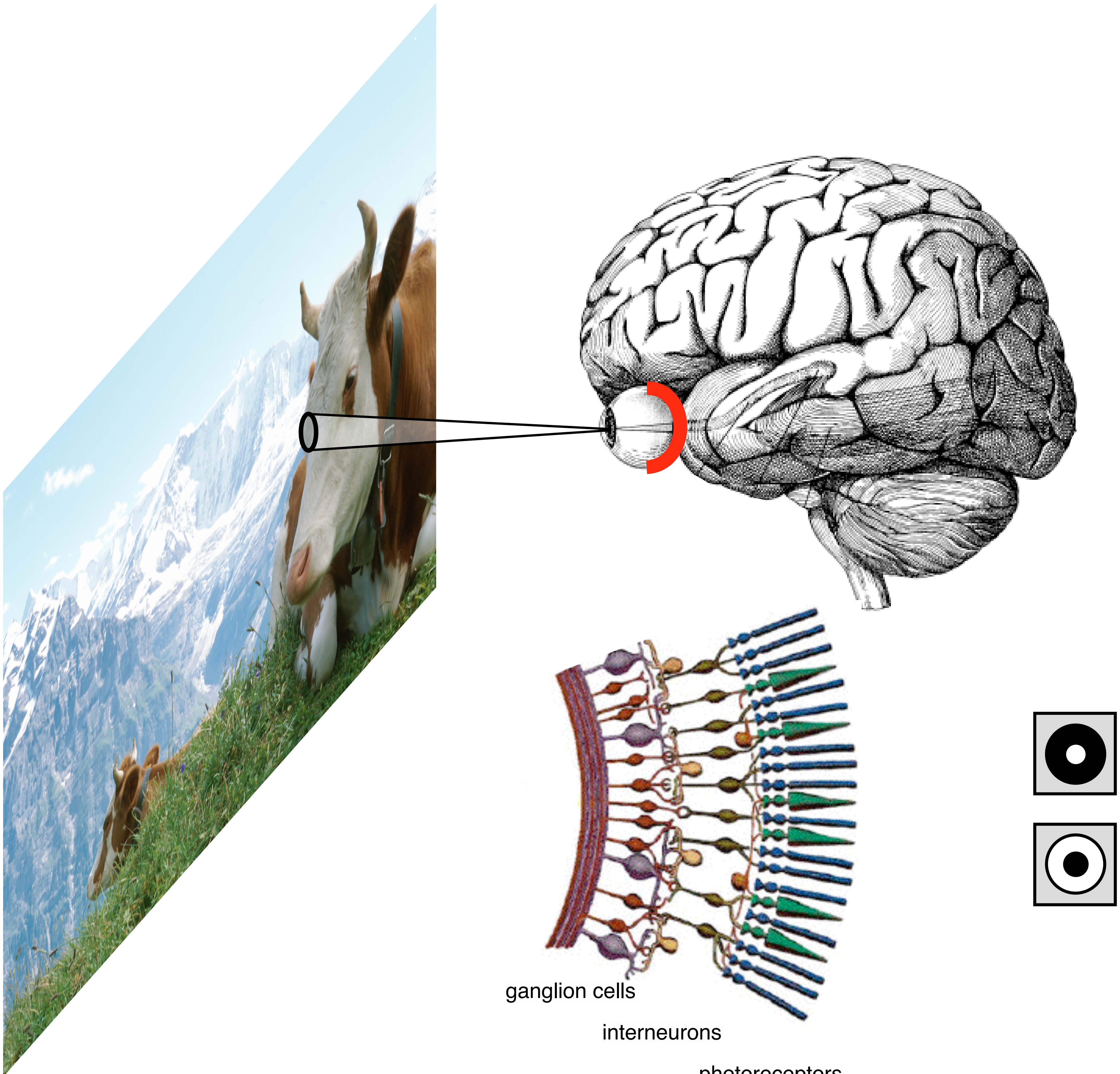
Nonlinear Image Representation with Cascaded Local Gain Control

Eero Simoncelli
Center for Neural Science, and
Courant Inst. of Mathematical Sciences,
New York University

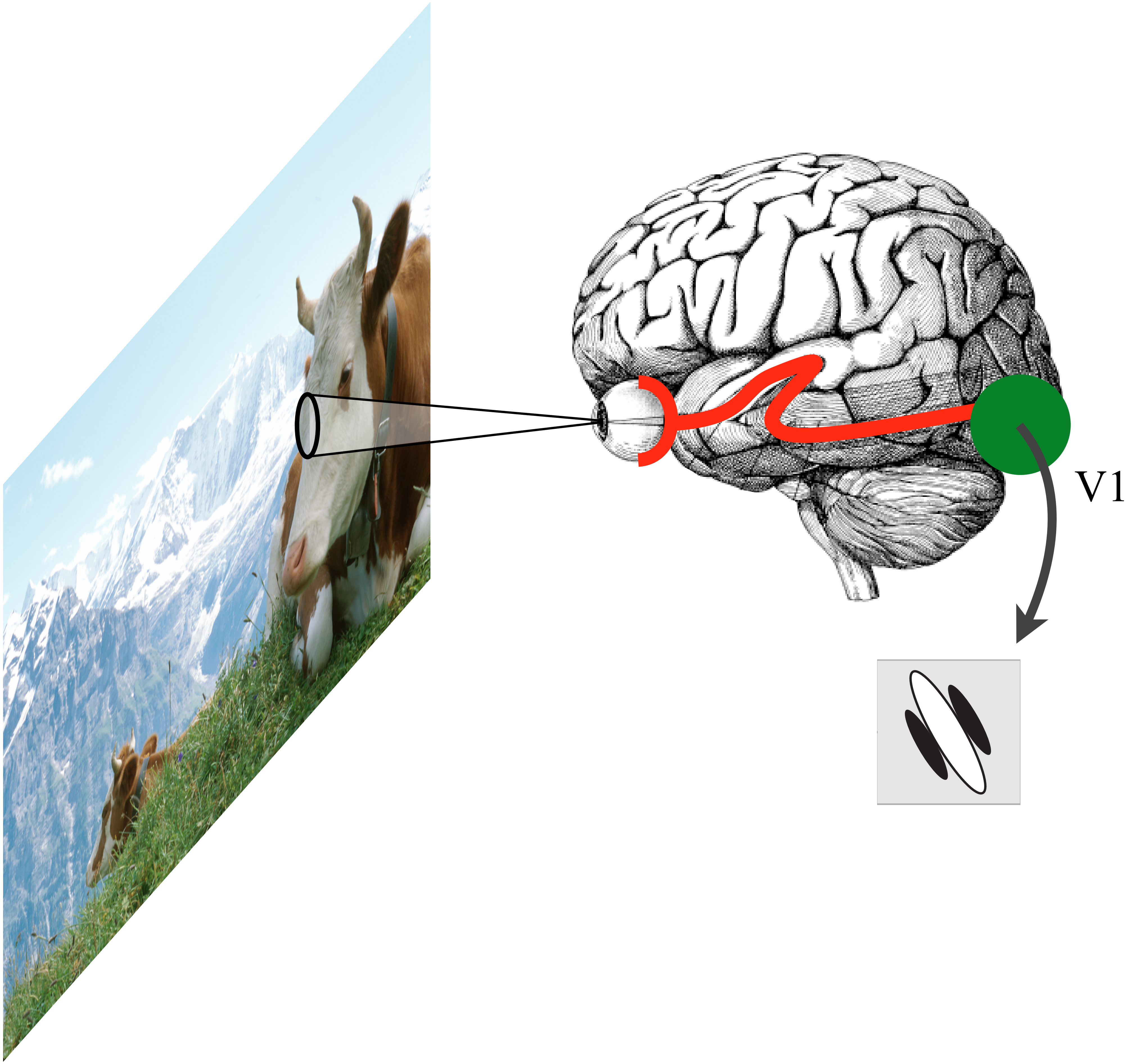
  

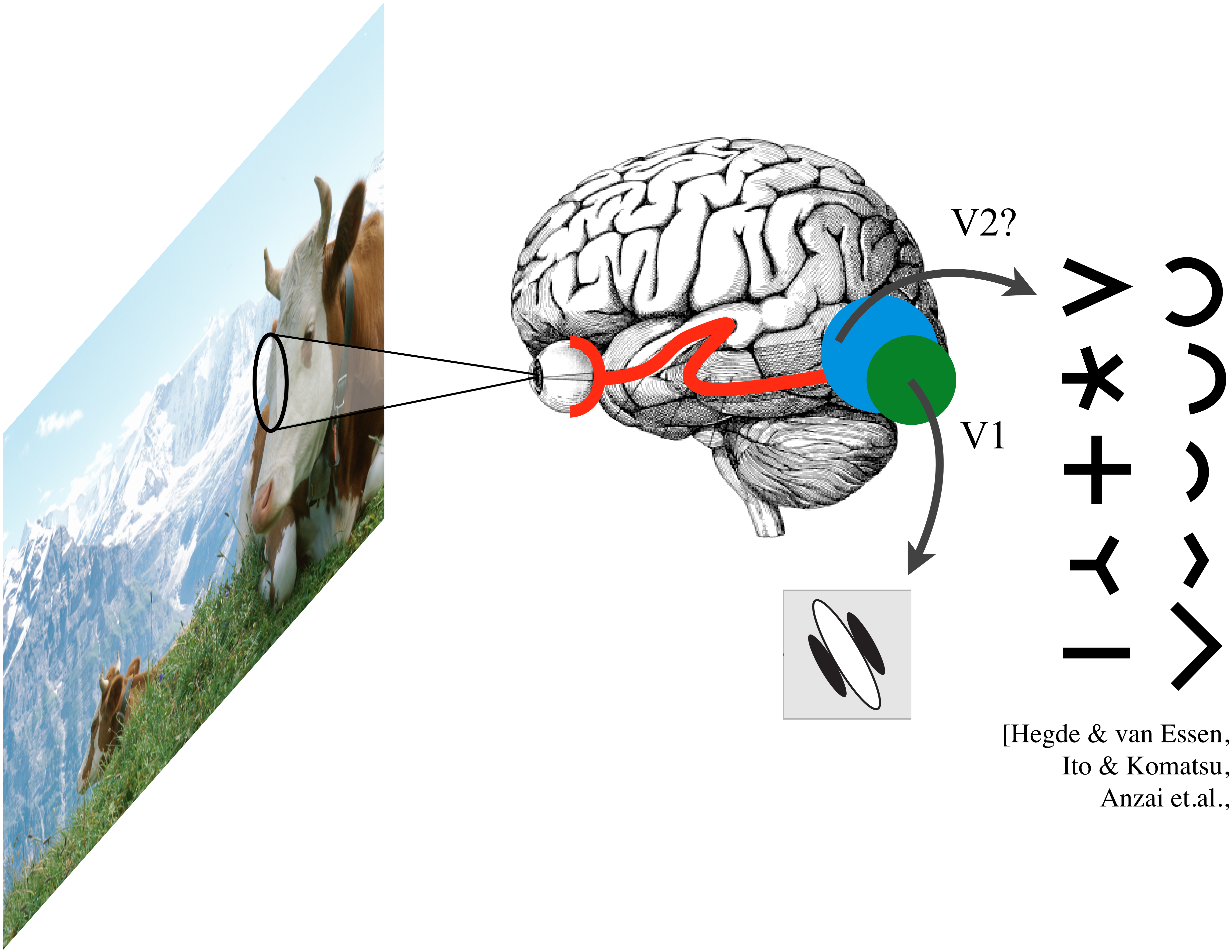
Environment → Physiology → Perception

- How do populations of neurons extract/represent visual information?
- In what ways is this matched to, or optimized for, our visual environment?
- How do these representations enable/limit perception?
- What new principles may be gleaned from these representations, and applied to engineered imaging or vision systems?



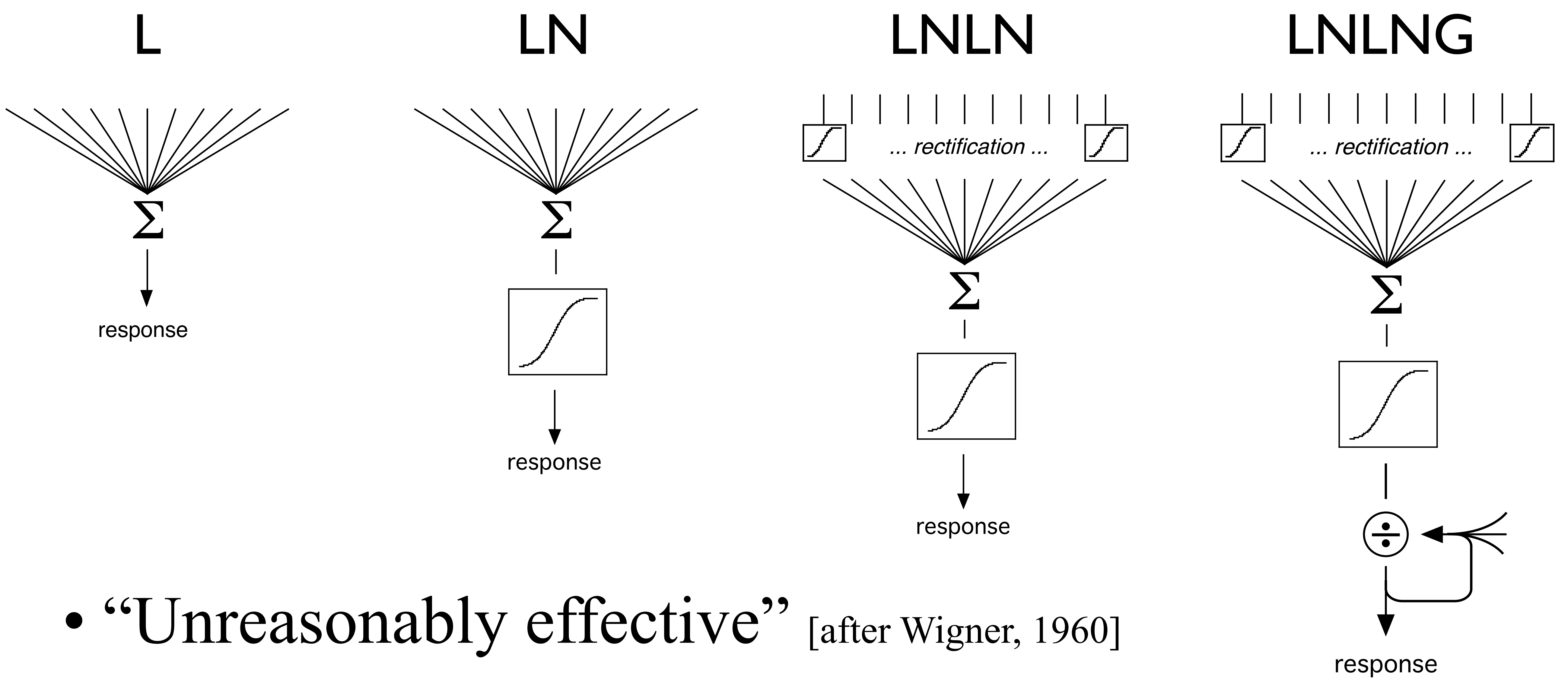
[figure: Hubel '95]


V1

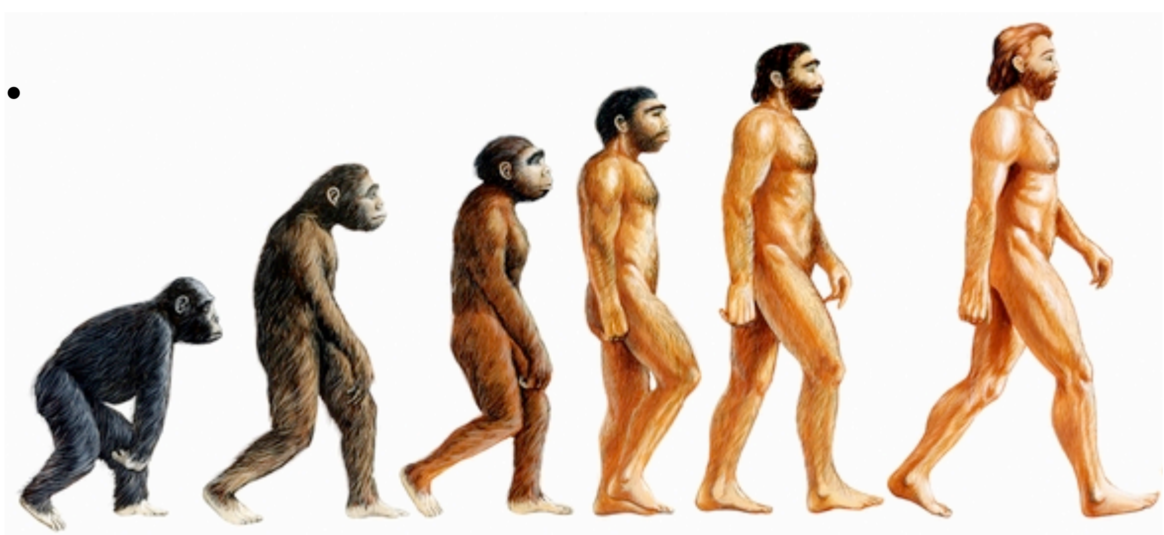


[Hegde & van Essen, 2000
Ito & Komatsu, 2004
Anzai et.al., 2007
etc]

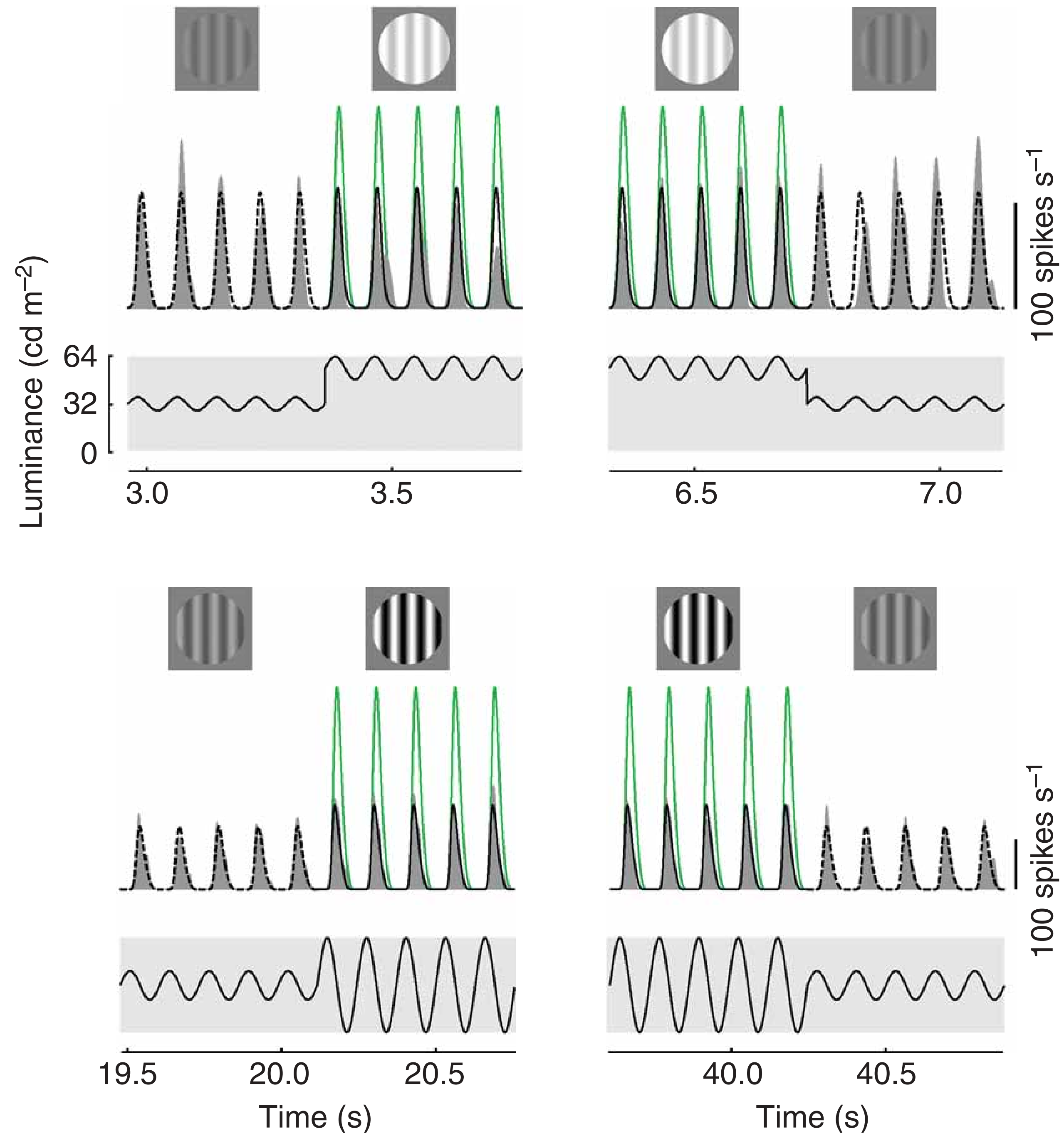
Canonical functional models for sensory neurons



- "Unreasonably effective" [after Wigner, 1960]
- Evolving...

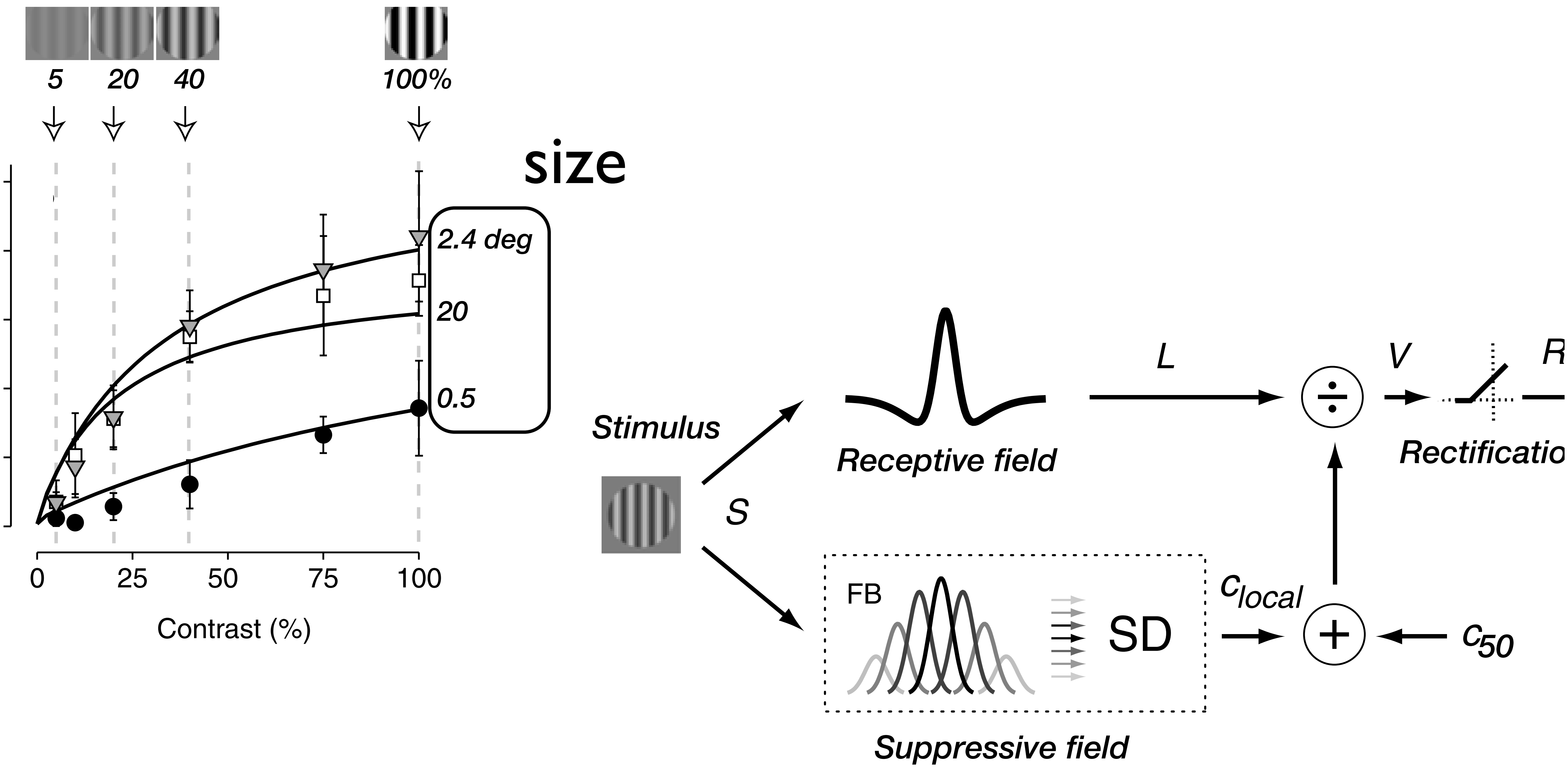


Temporally adaptive gain control



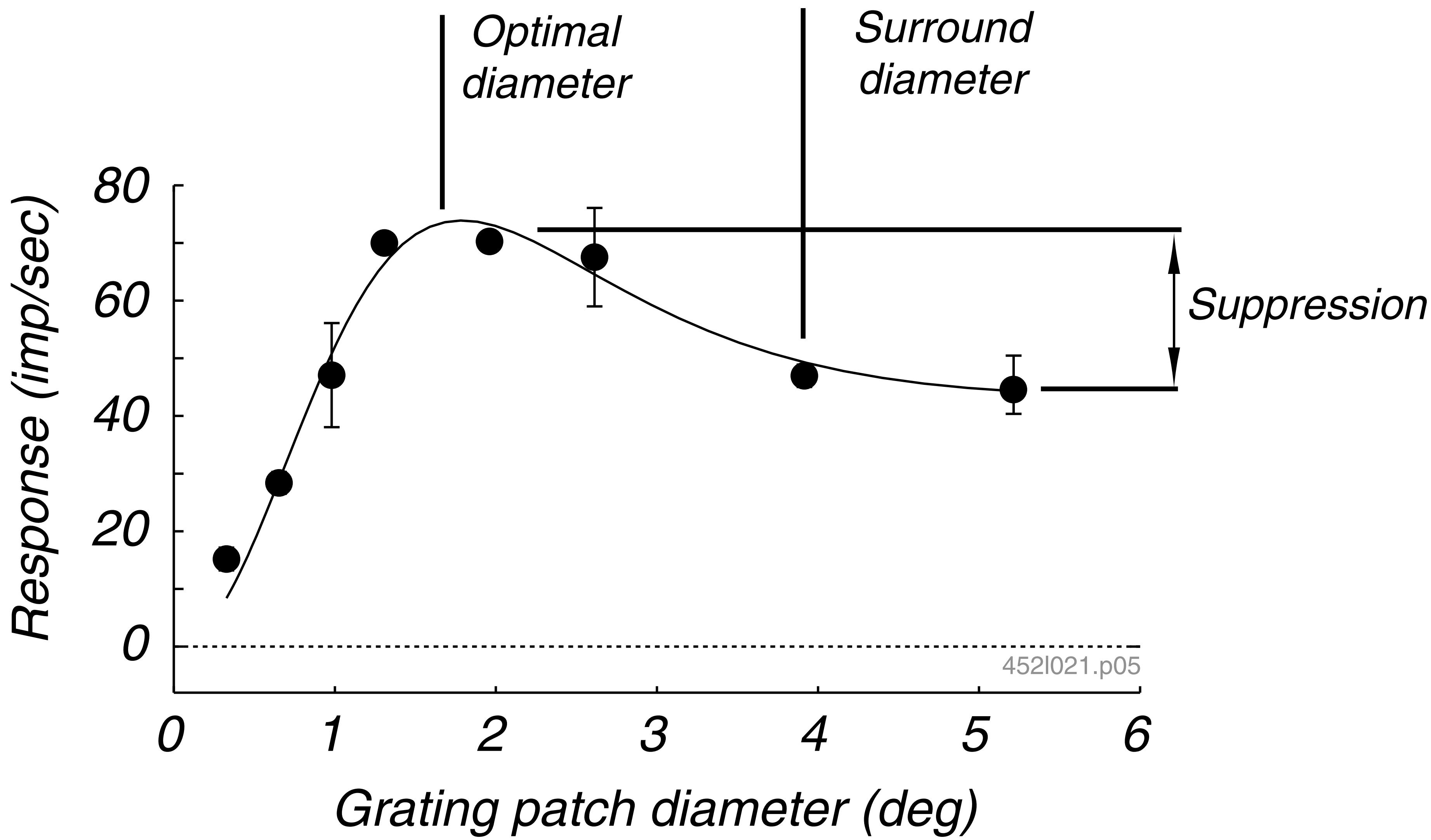
[Mante et. al. 2005]

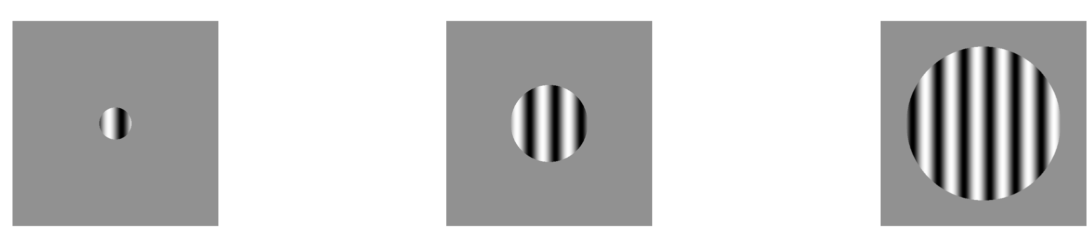
Local **spatial** gain control



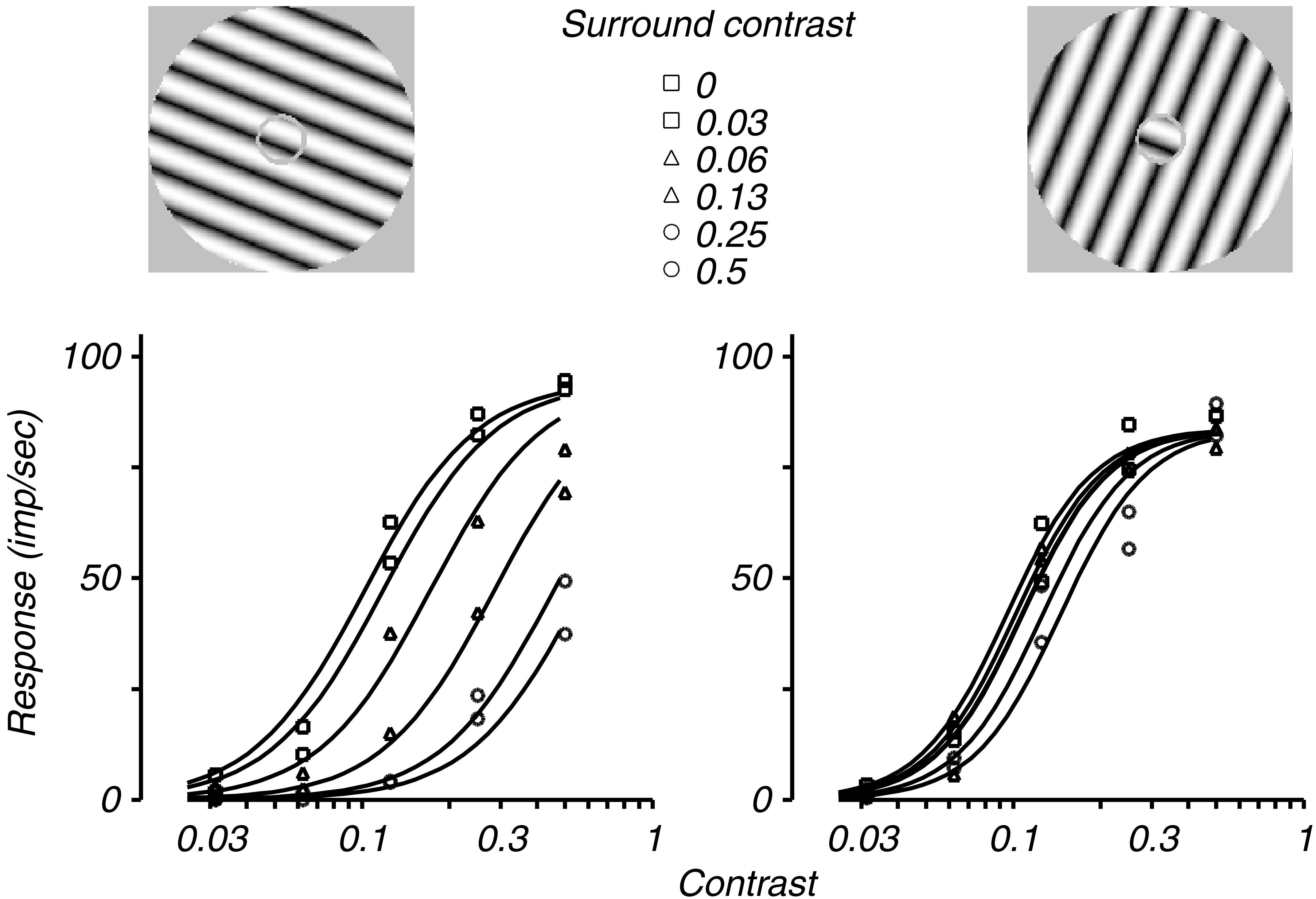
[Bonin, et al 2005]

V1: Surround suppression



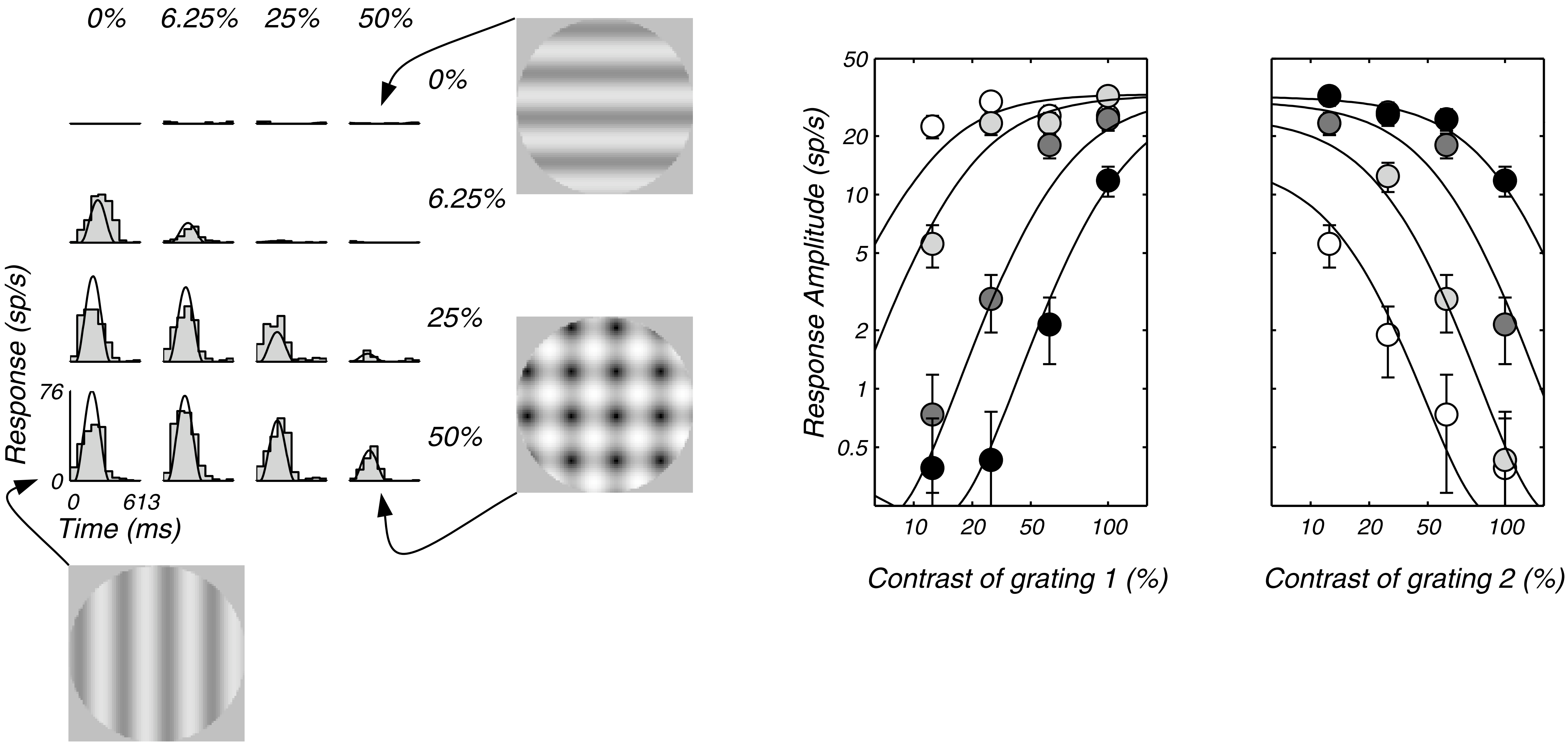


[Cavanaugh etal 02]



[Cavanaugh etal 02]

V1: Cross-orientation suppression

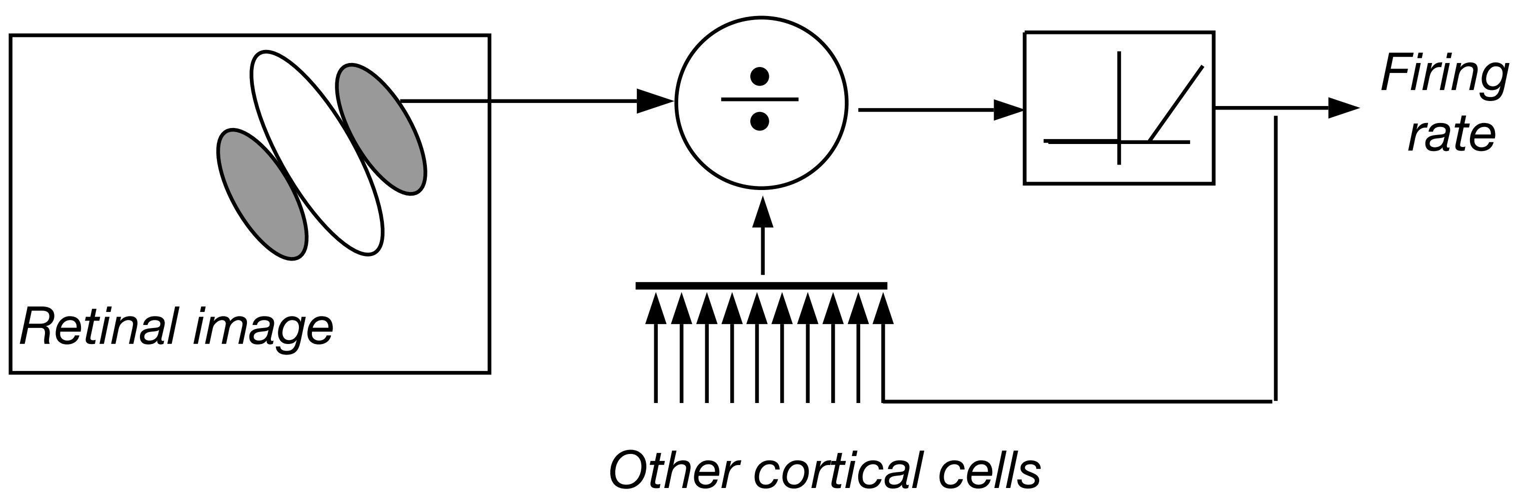


[Carandini etal 1997]

V1 Normalization Model

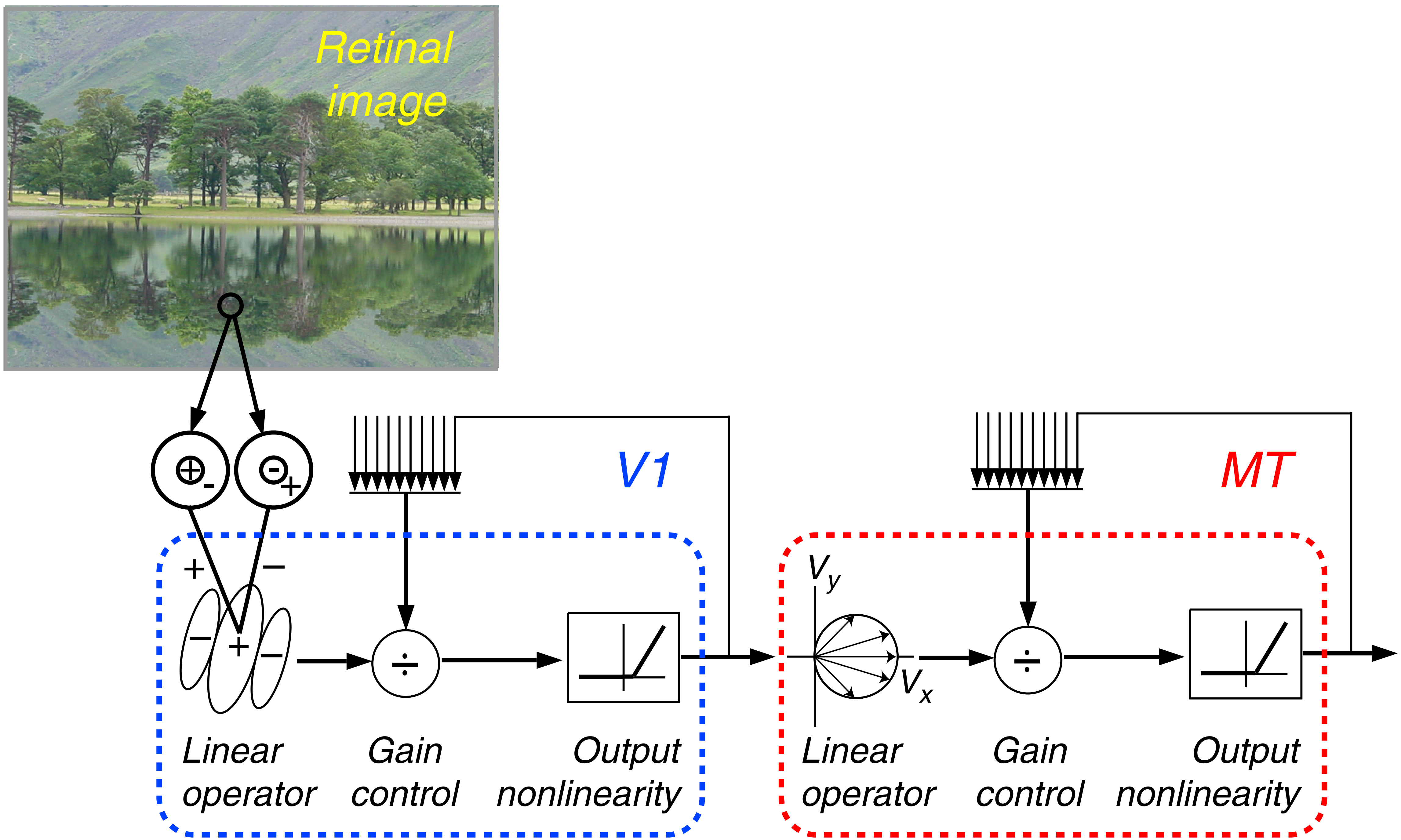
The linear model of simple cells

The normalization model of simple cells



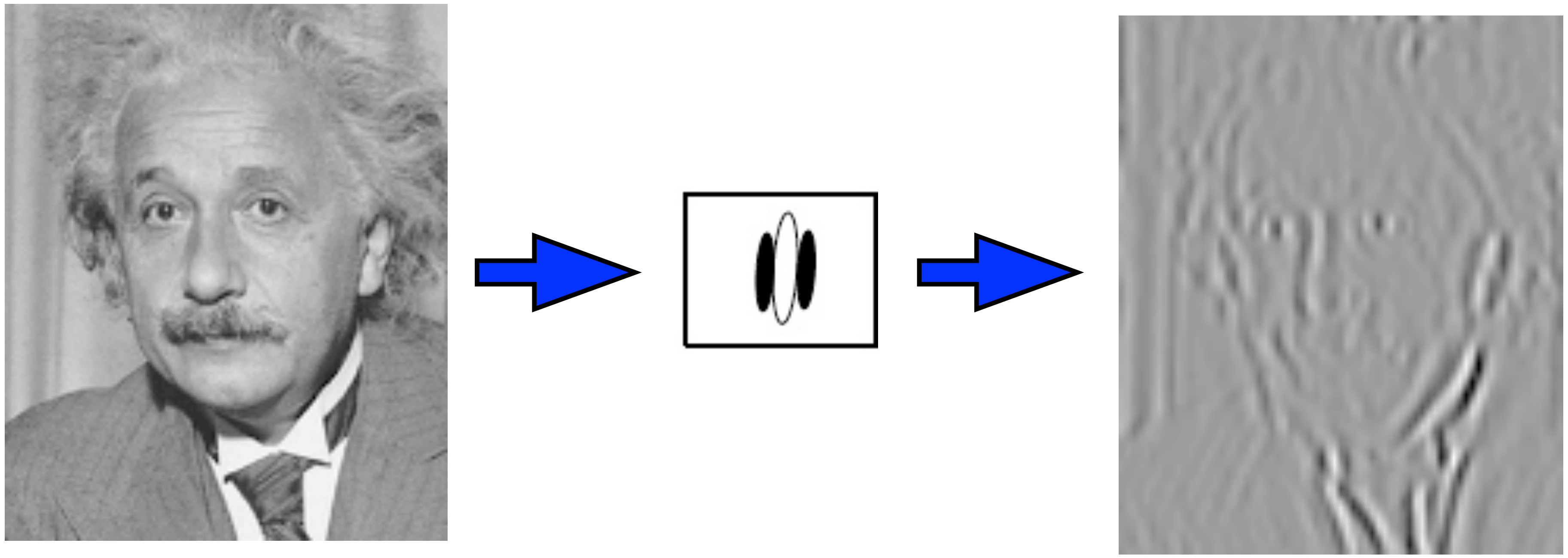
[Carandini, Heeger, and Movshon, 1996;
Carandini & Heeger, 2012]

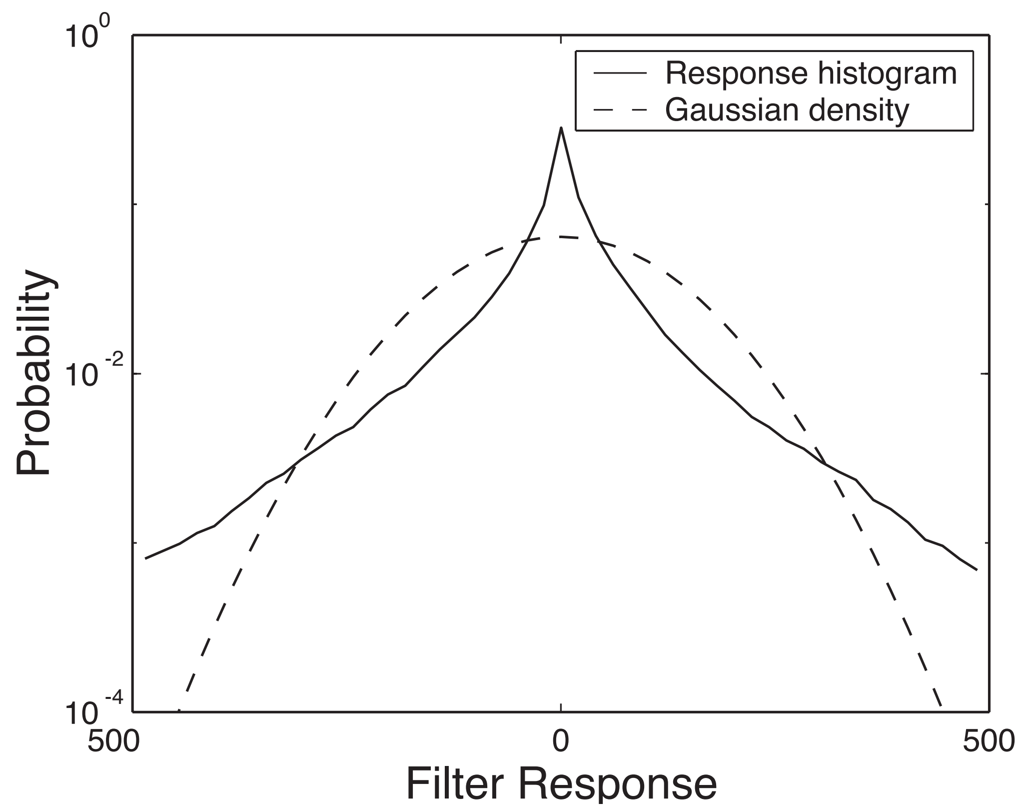
Example: Area MT



[Simoncelli & Heeger, 1998]

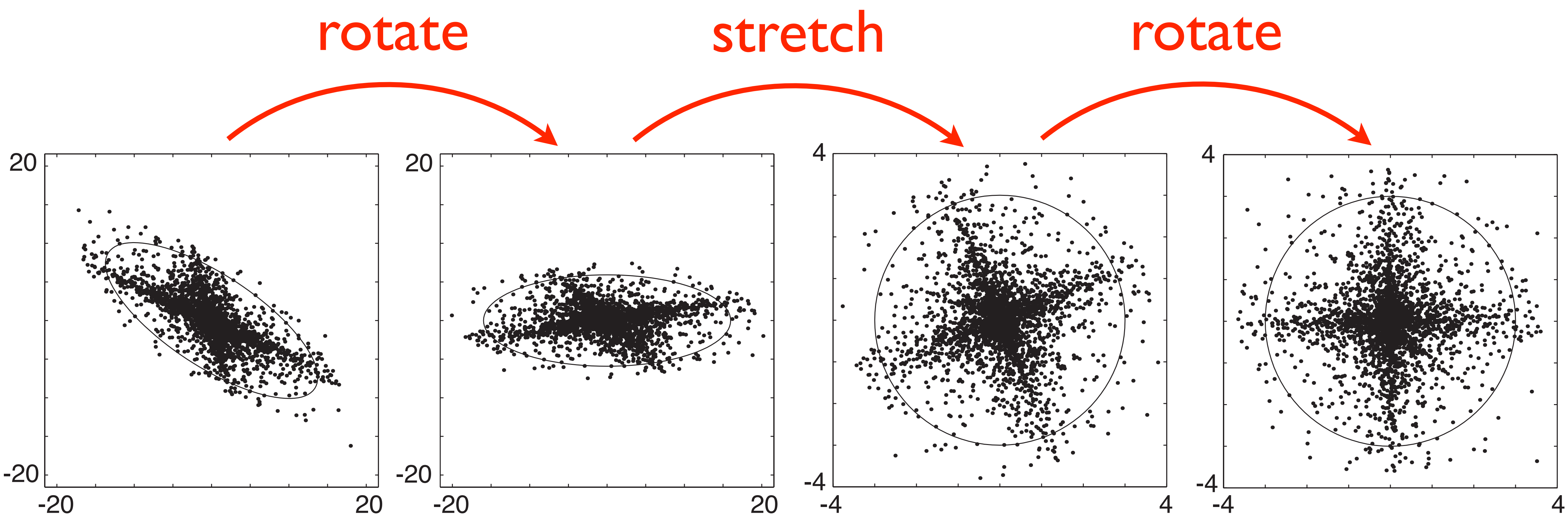
Sparse marginal statistics





[Burt&Adelson 82; Field 87; Mallat 89; etc]

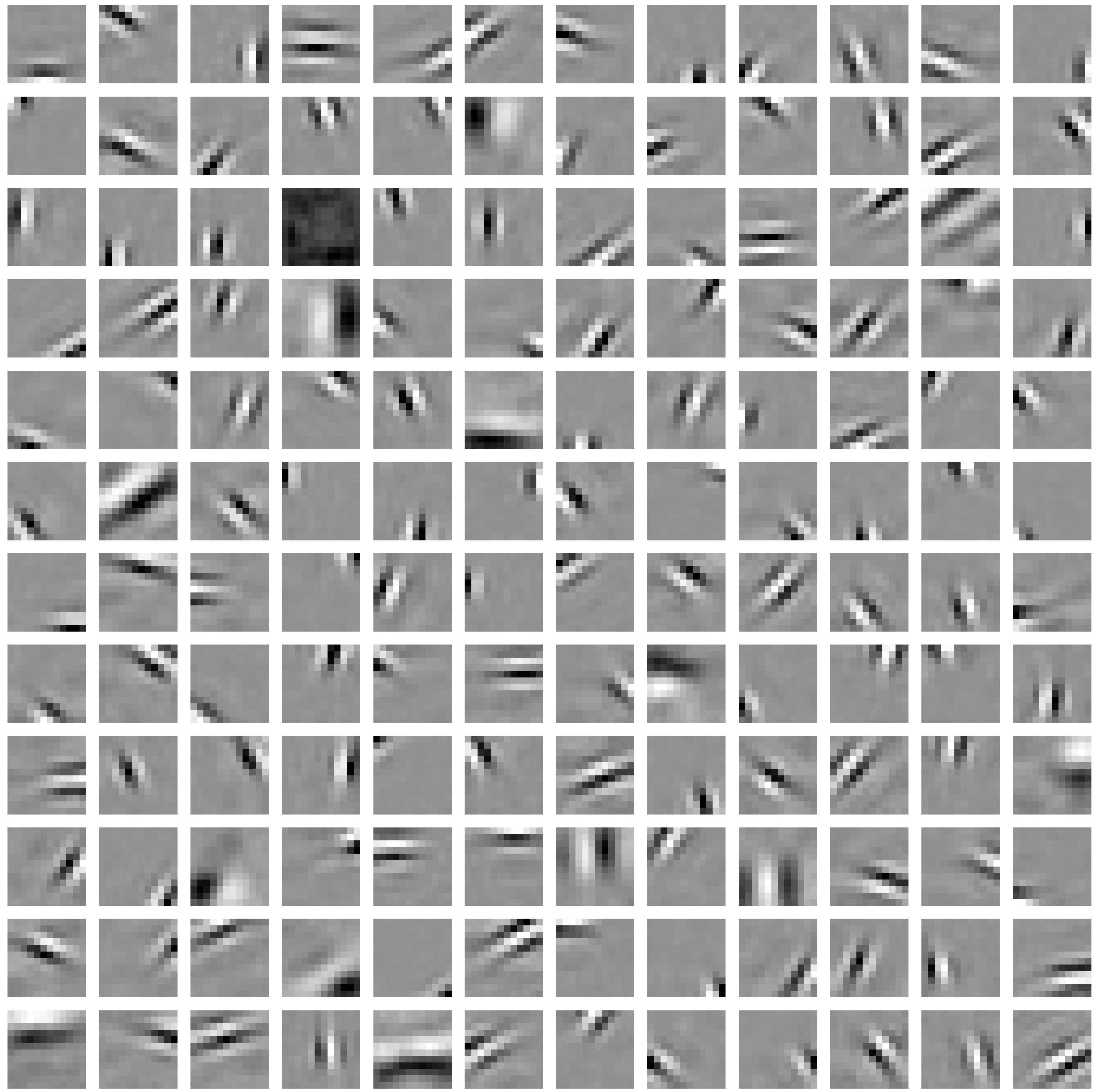
Independent Components Analysis (ICA)



For linearly-transformed-factorial sources:
guaranteed independence
(with some minor caveats)

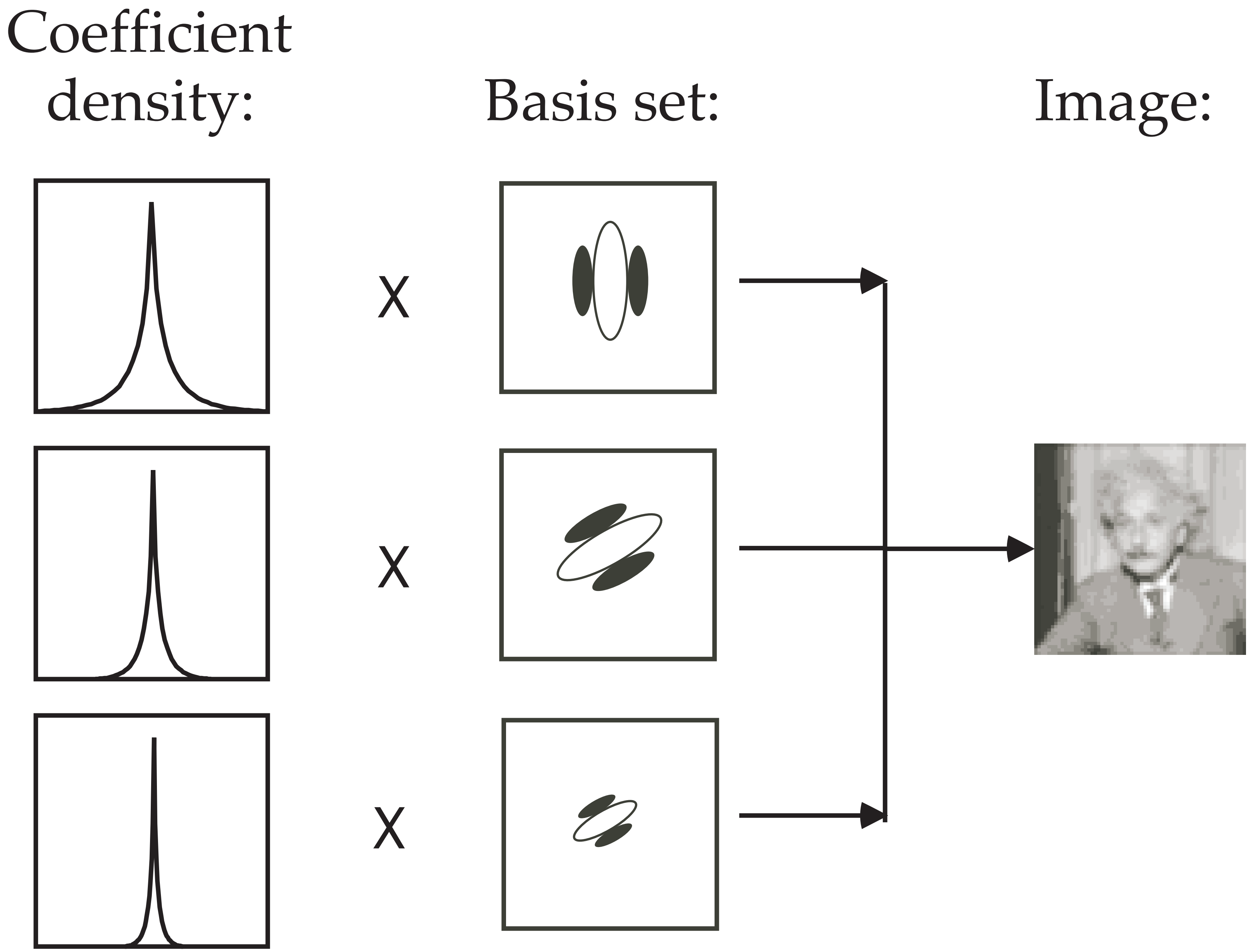
[Cardoso 89; Jutten & Herault 91; Comon 94; Bell & Sejnowski 96; etc]

ICA on image blocks

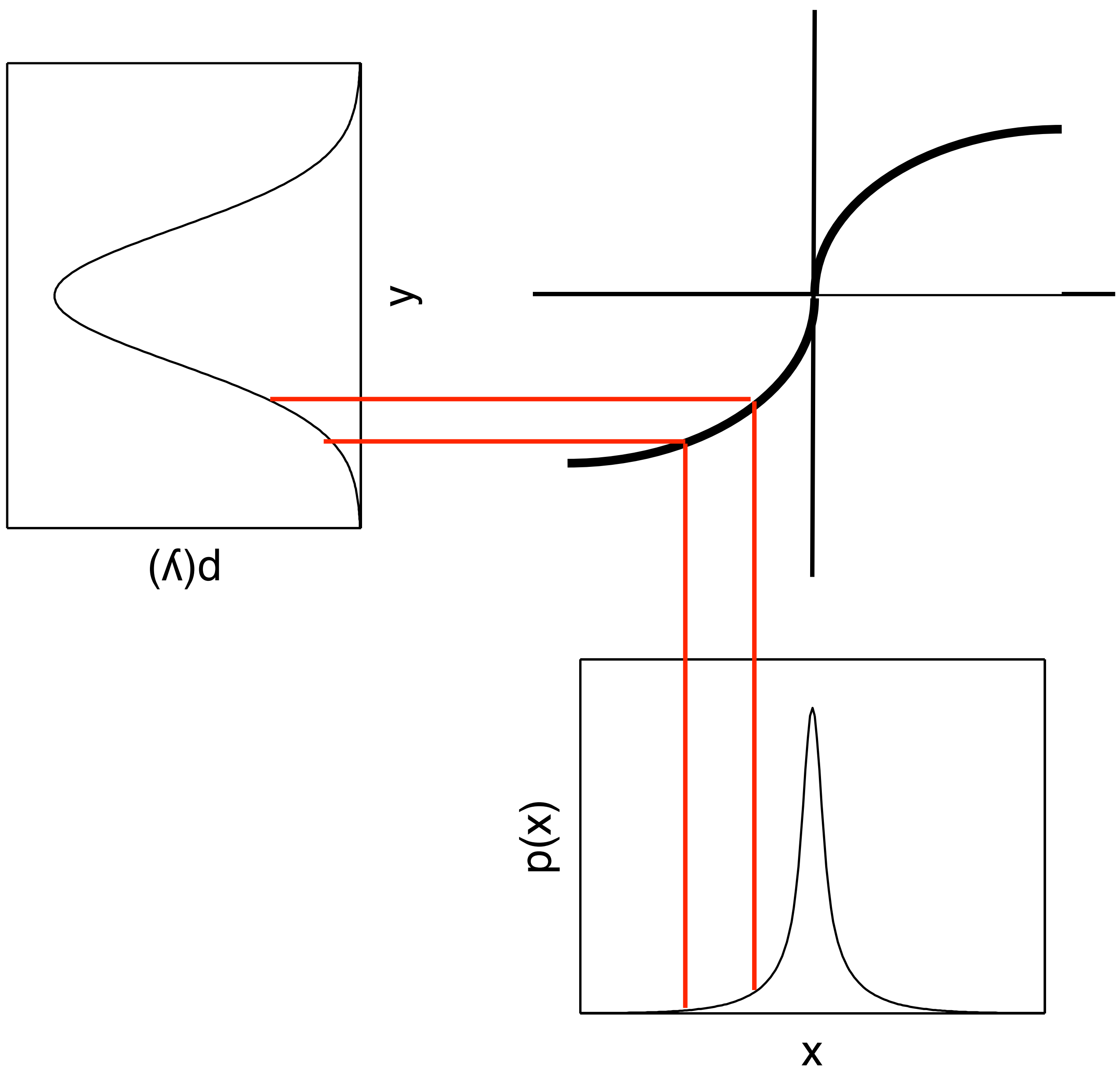


[Bell/Sejnowski '97; see also Olshausen/Field '96]

Linearly-transformed factorial model

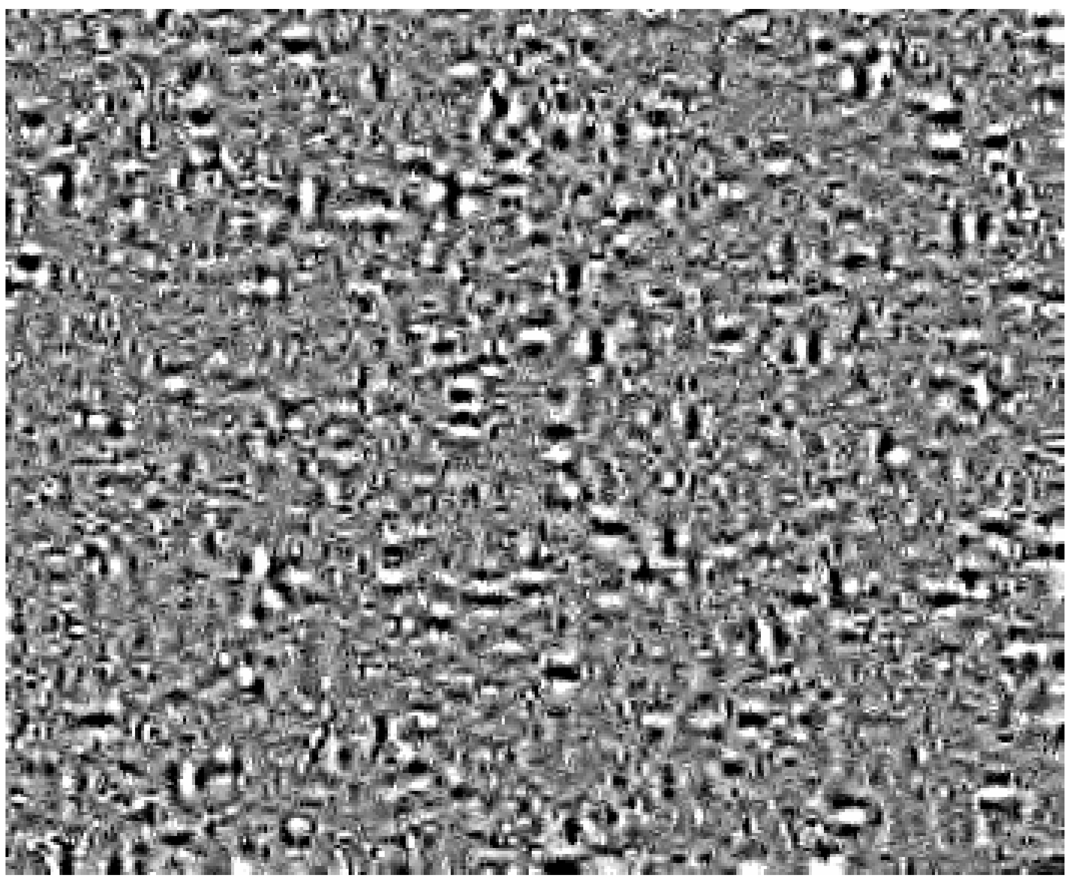


Marginal Gaussianization

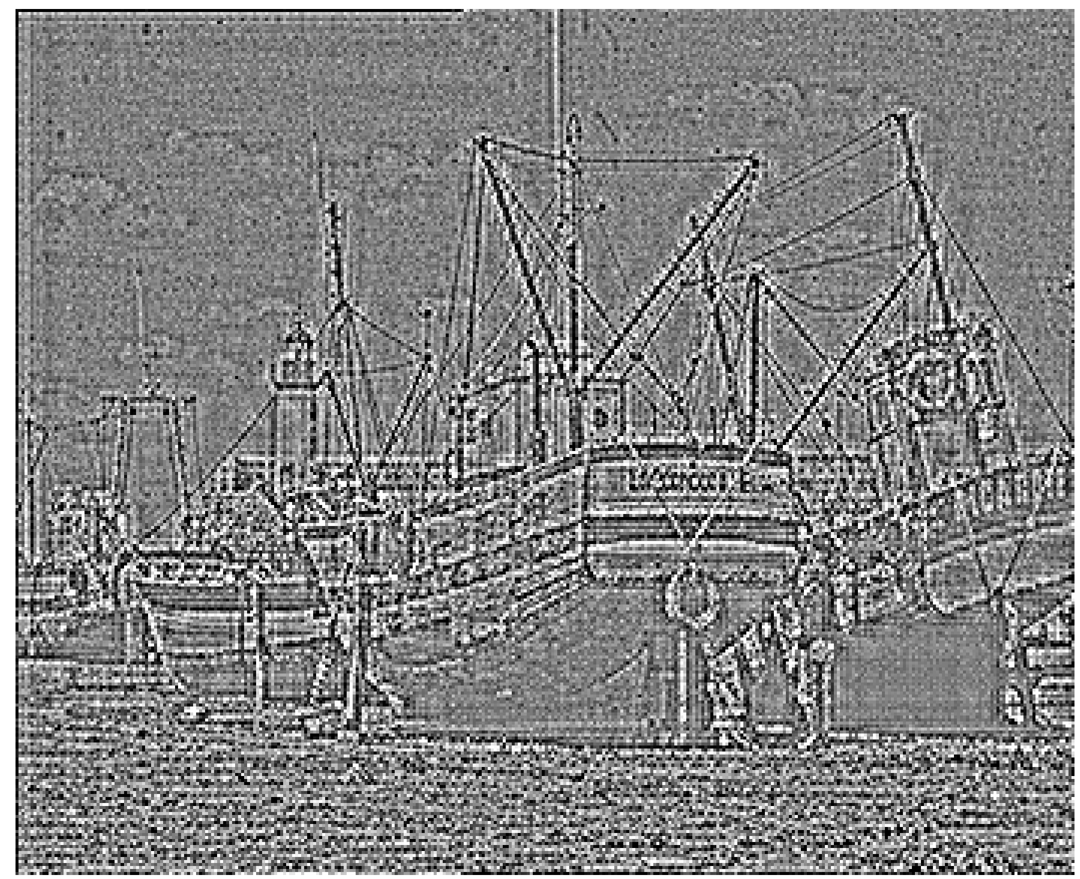


[Chen & Gopinath 01]

Indications that the model is weak...

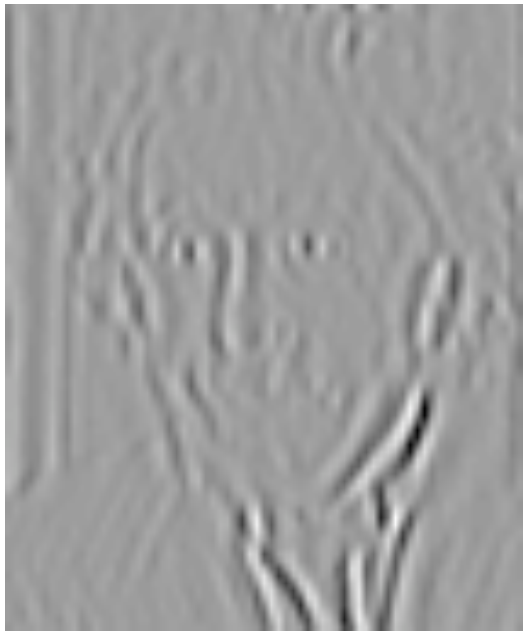


Sample from model

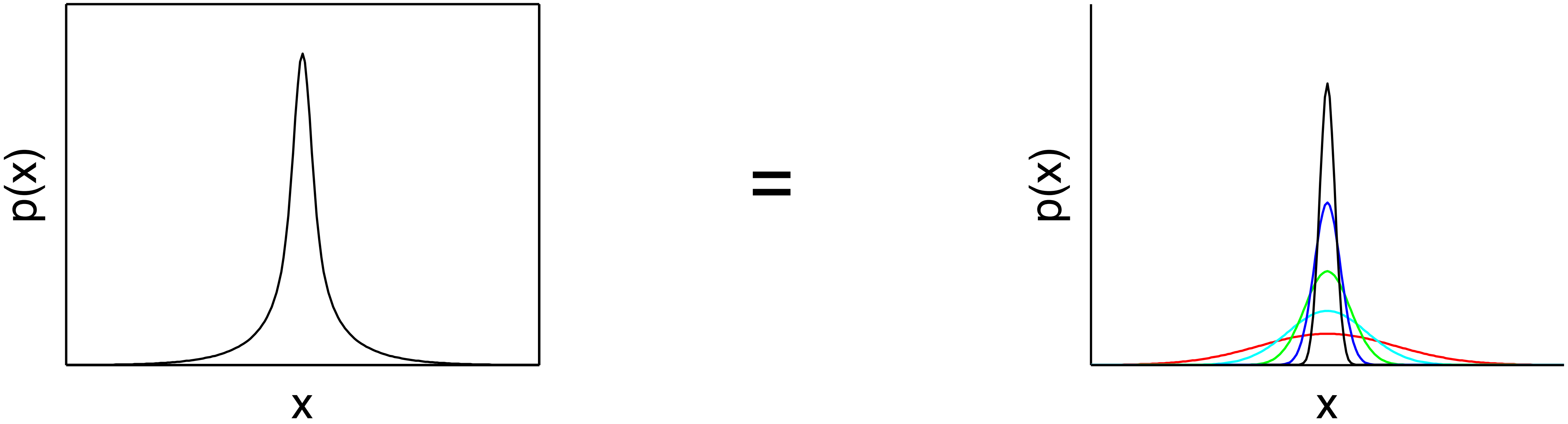


Image, ICA-transformed
and marginally Gaussianized

Subbands are *heteroskedastic*
(they have variable variance):



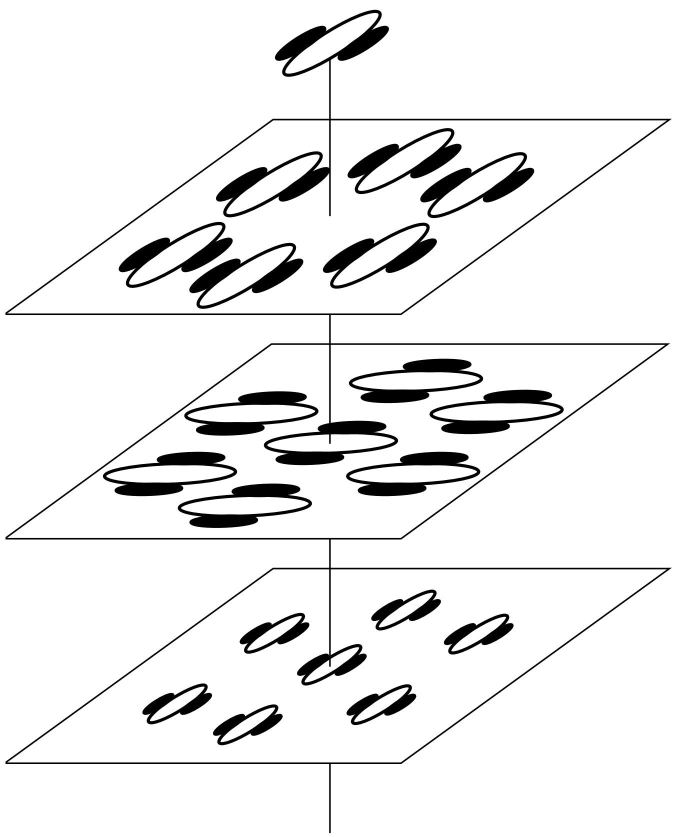
We can model this behavior using a Gaussian scale mixture (GSM):
[Wainwright & Simoncelli 2000]



GSM

Model generalized coefficient neighborhood as a Gaussian scale mixture (GSM) [Andrews&Mallows '74]:

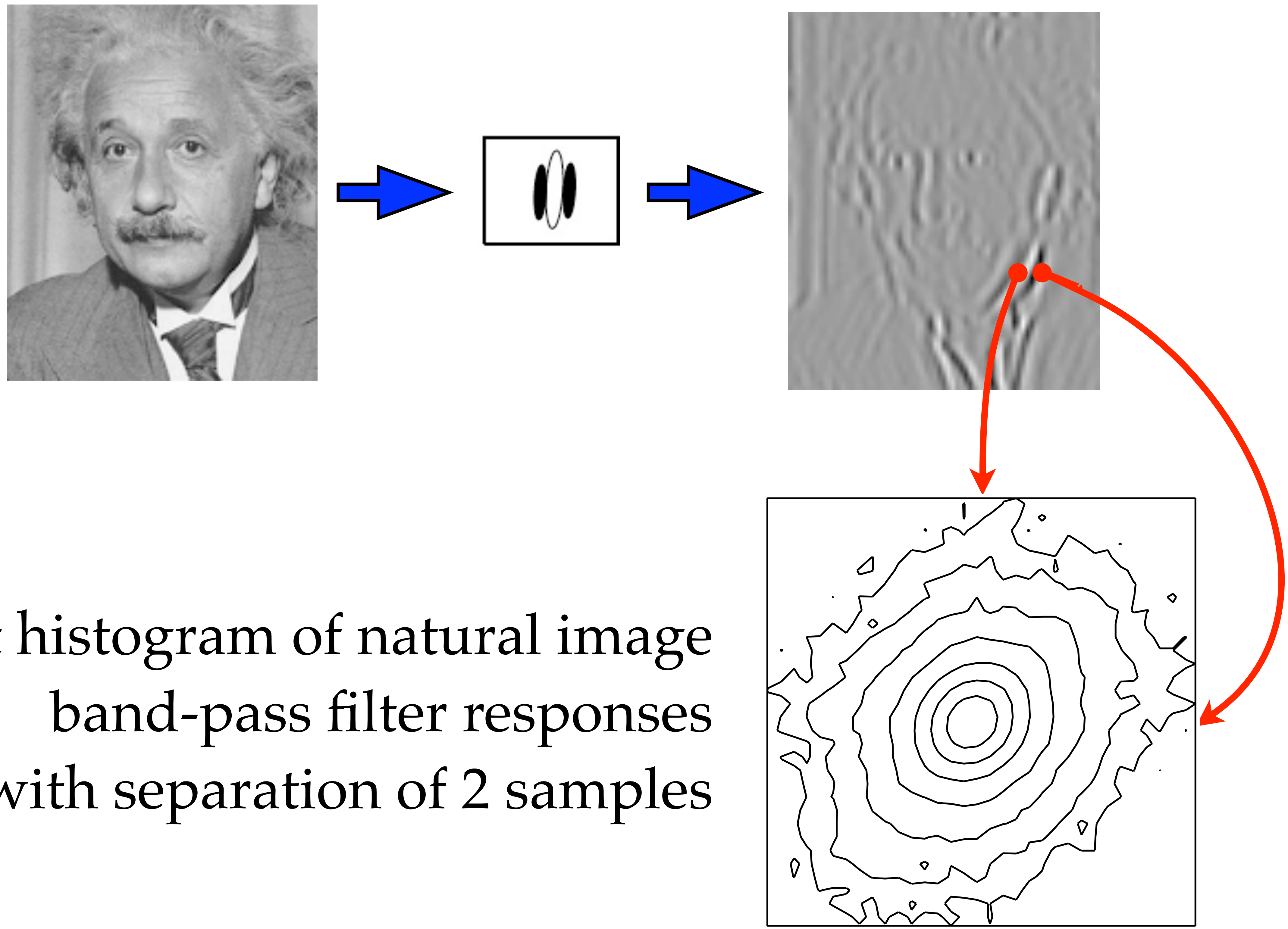
$$\vec{x} = \sqrt{z}\vec{u}$$



- $\vec{u}$ is Gaussian, $z > 0$
- z and $\vec{u}$ are independent
- $\vec{x}$ is elliptically symmetric, with covariance zC_u
- marginals of $\vec{x}$ are leptokurtotic

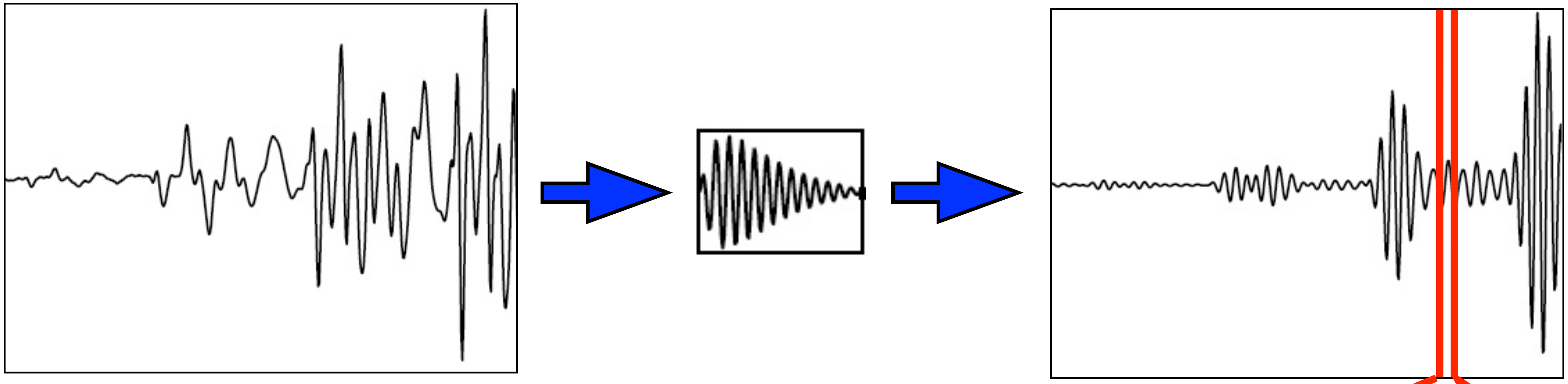
[Wainwright&Simoncelli, '99]

Joint statistics

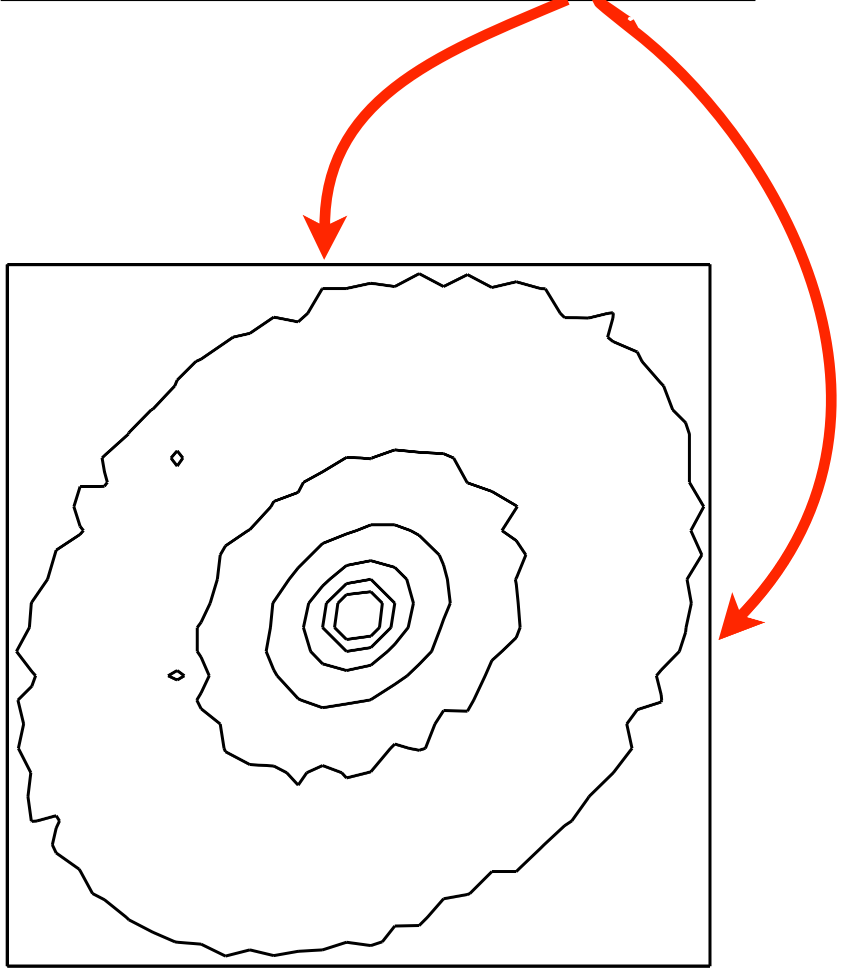


joint histogram of natural image
band-pass filter responses
with separation of 2 samples

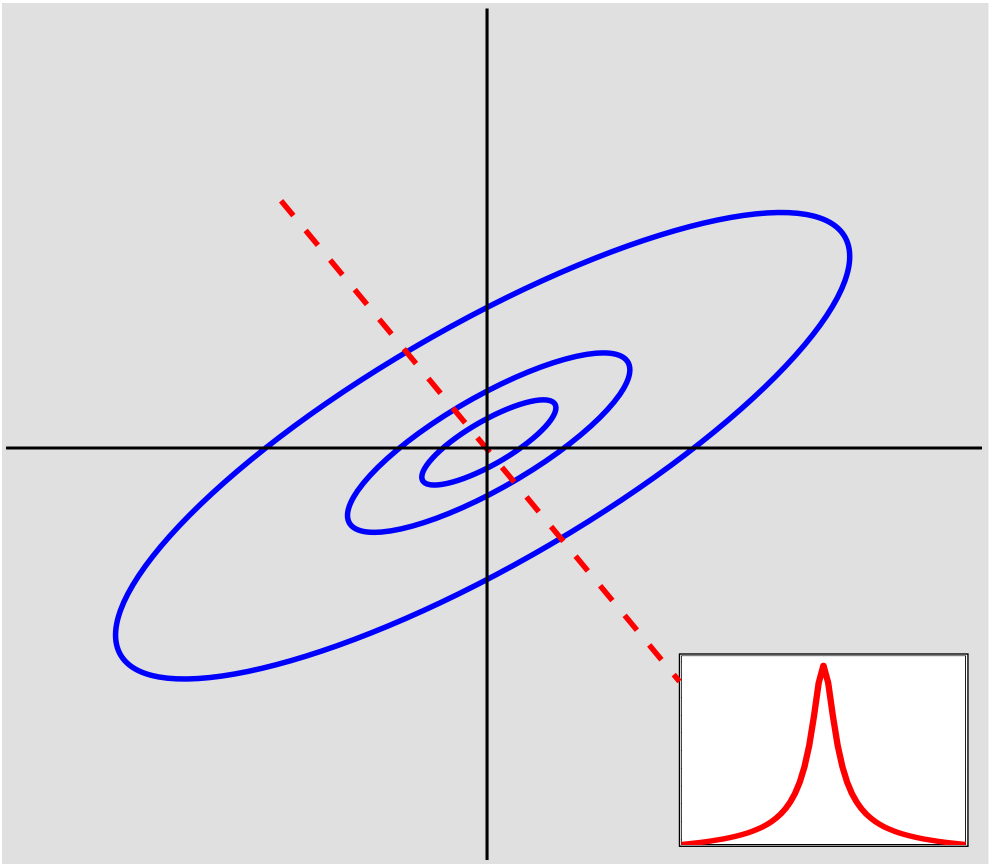
Joint statistics - sound



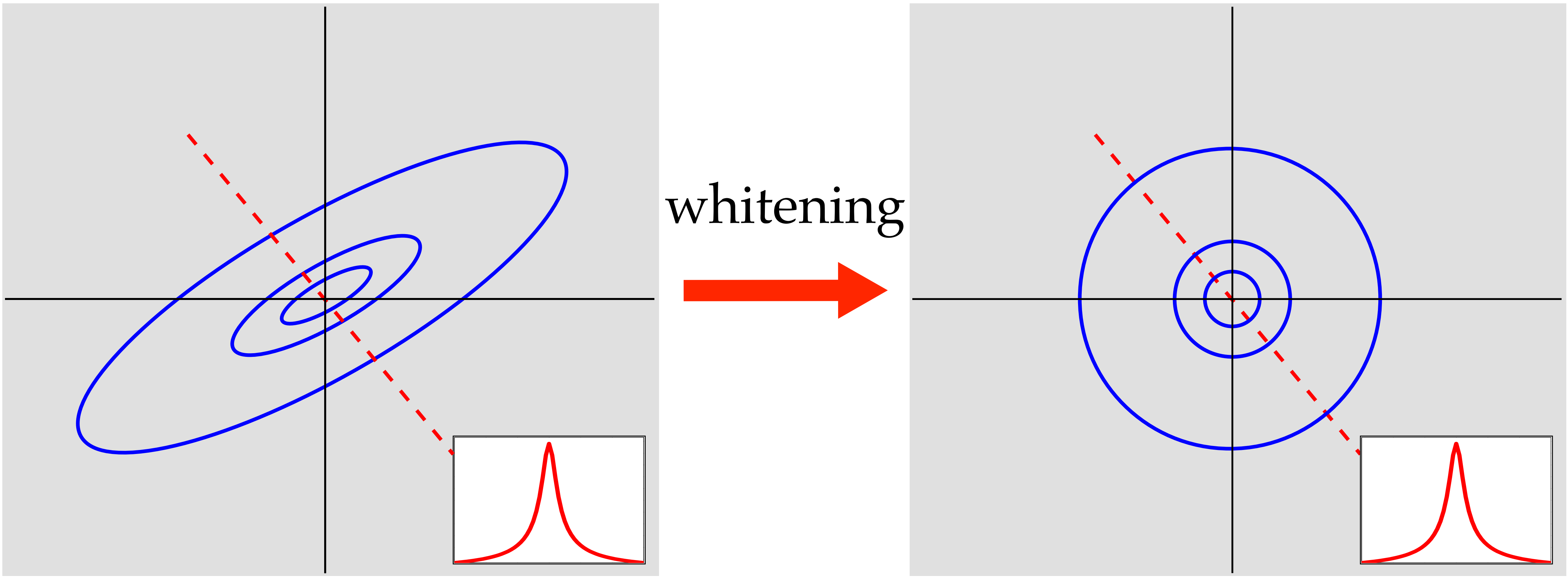
joint histogram of natural audio signal gammatone filter responses with separation of 0.1 msec



non-Gaussian elliptical models of natural images:

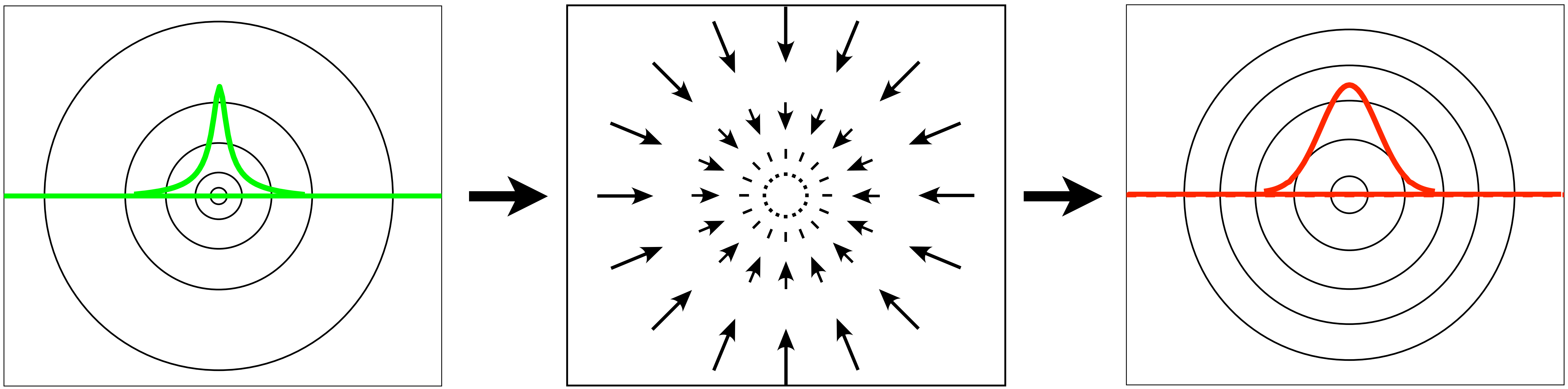


- Simoncelli, 1997;
- Zetzsche & Krieger, 1999;
- Huang & Mumford, 1999;
- Wainwright & Simoncelli, 2000;
- Hyvärinen and Hoyer, 2000;
- Parra et al., 2001;
- Srivastava et al., 2002;
- Sendur & Selesnick, 2002;
- Teh et al., 2003;
- Gehler and Welling, 2006
- Lyu & Simoncelli, 2008
- etc.



- Density is elliptical, but *not* Gaussian
- Whitening makes spherical, but not independent!

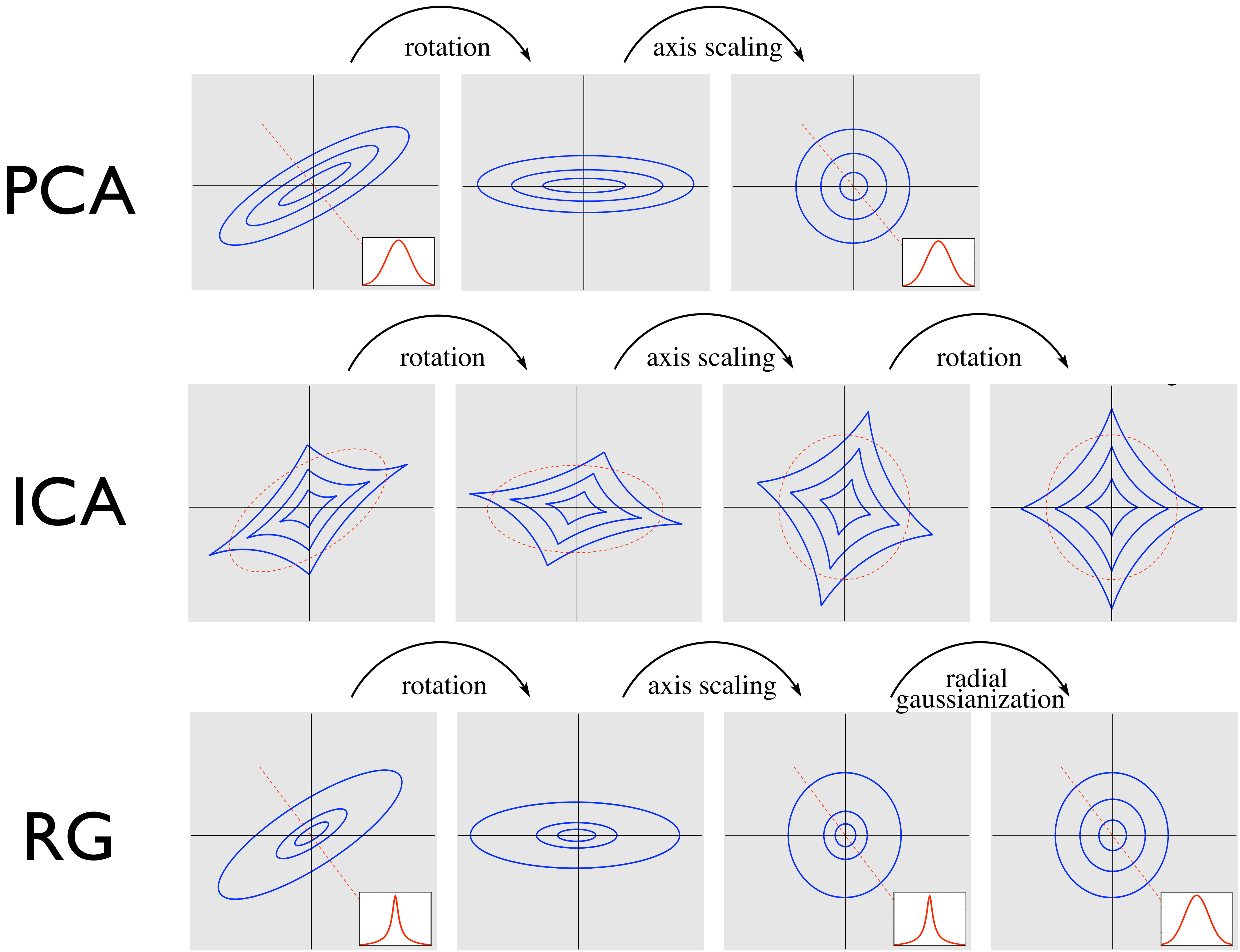
radial Gaussianization (RG)



Gaussianize the *radial component* of the density

Approximate version: estimate local L2 norm, and divide (i.e., local gain control)

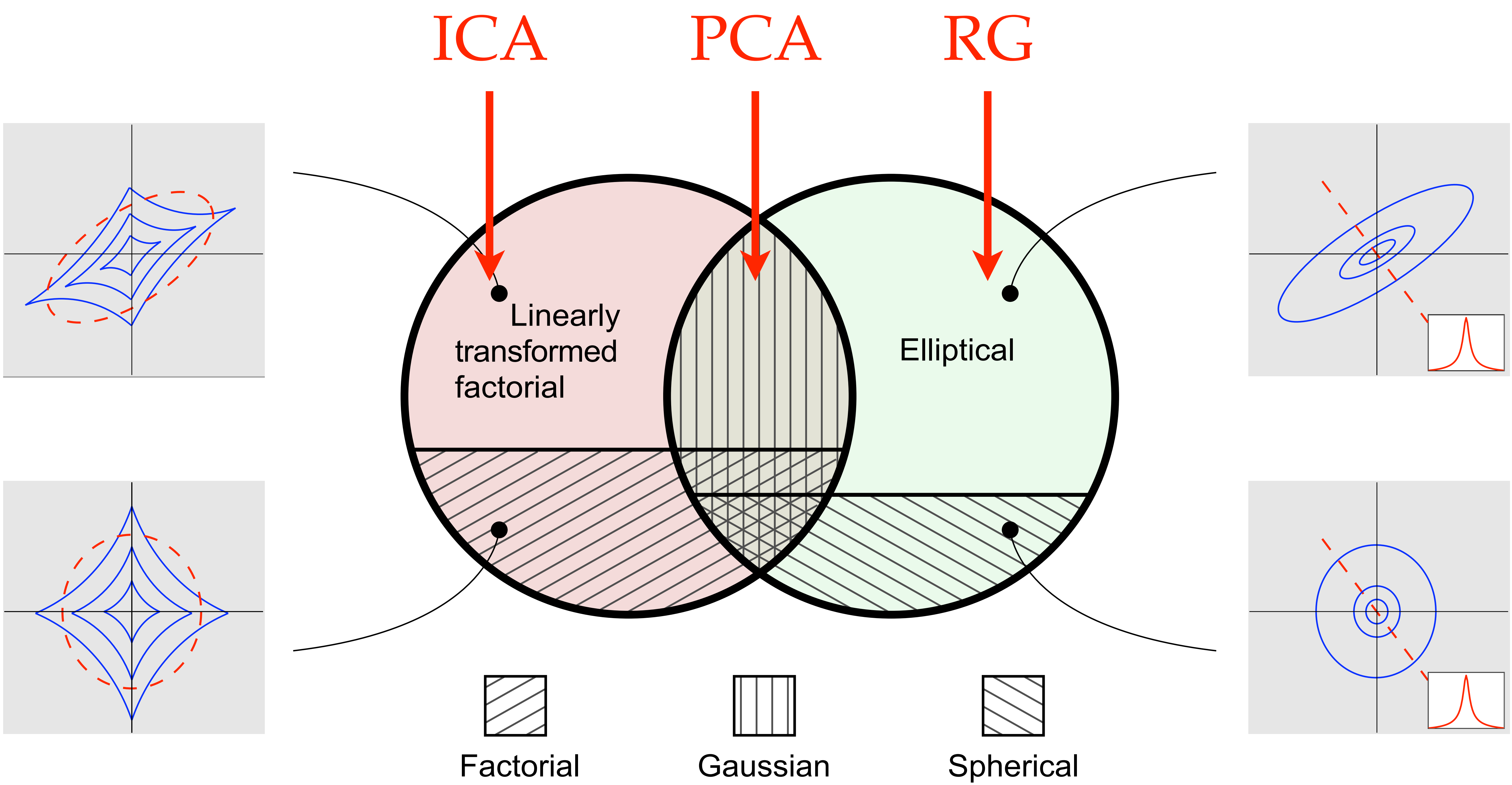
[Lyu & Simoncelli, 2008]



... all paths lead to a spherical/factorial Gaussian

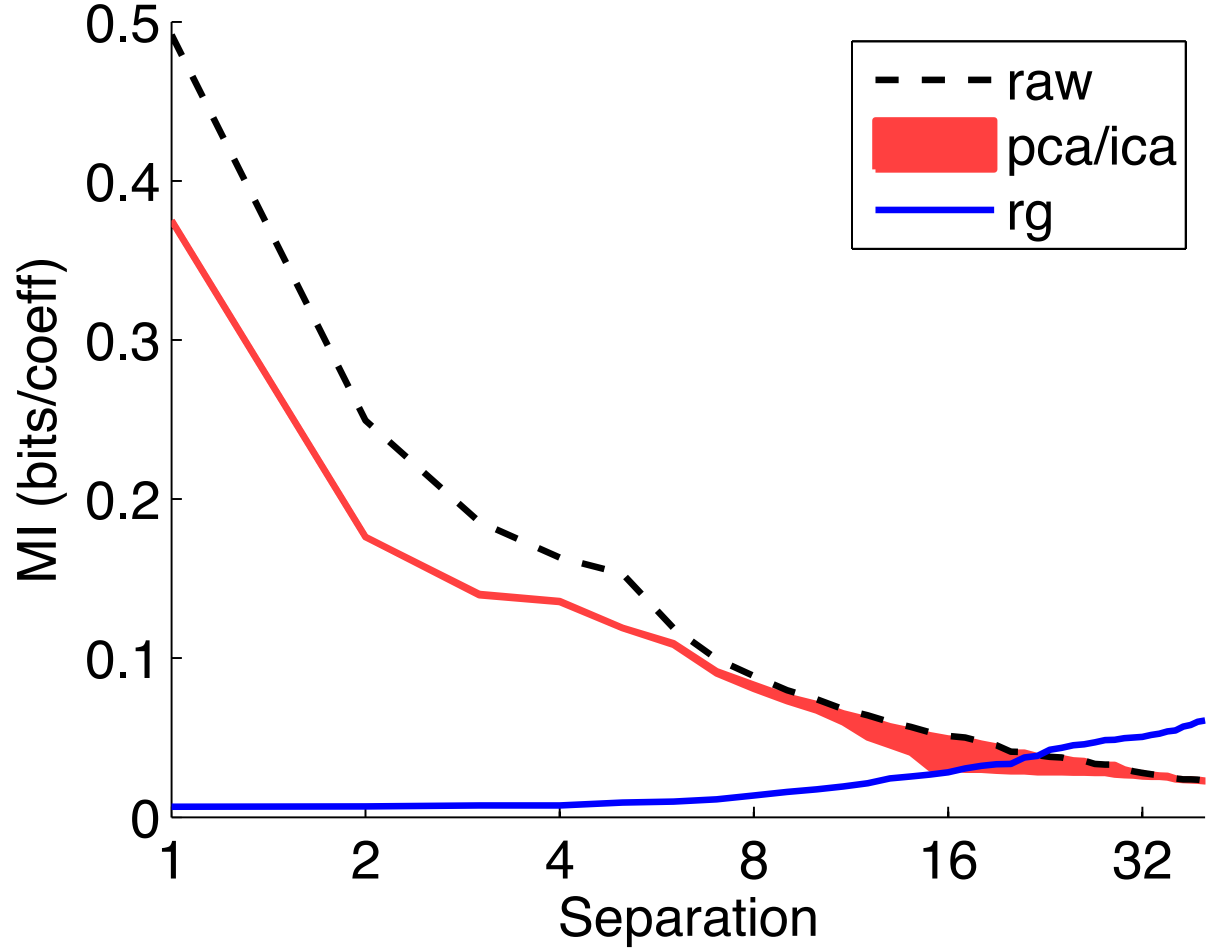
[Lyu & Simoncelli, 2008]

Densities and their factorizations



[Lyu & Simoncelli, 2008]

RG vs. ICA on coefficient pairs



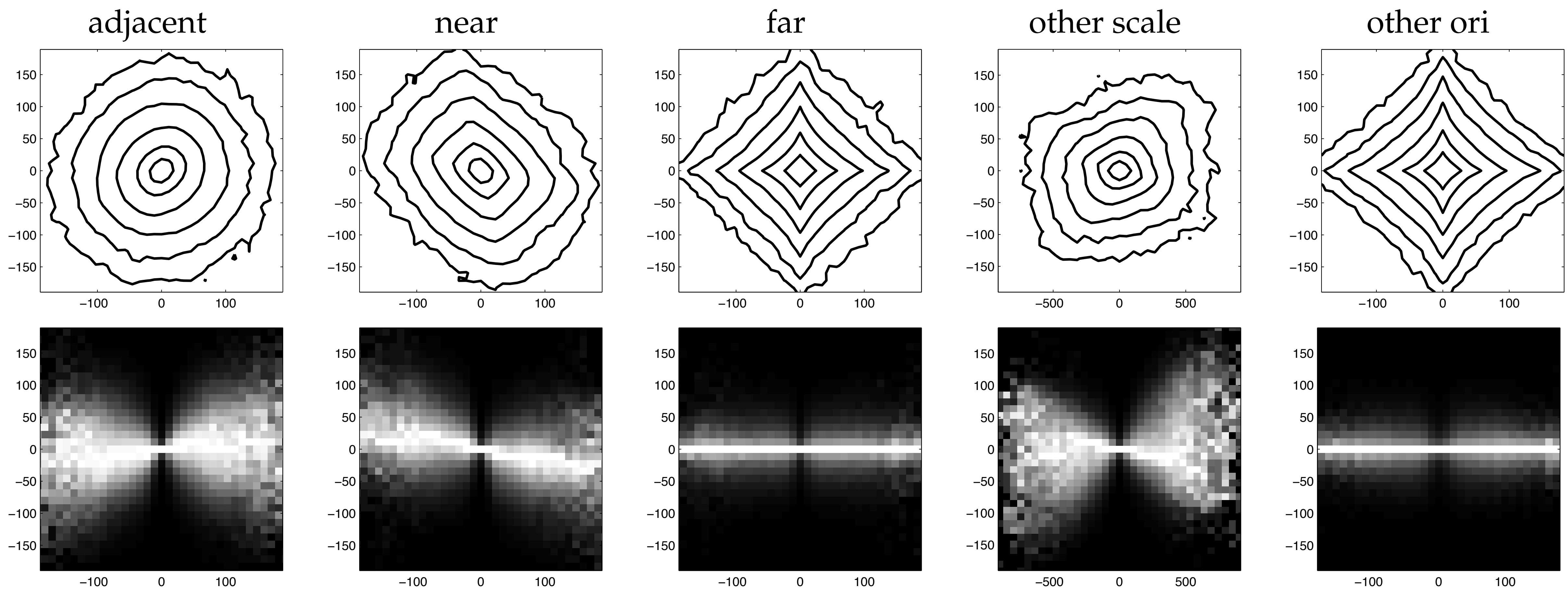
RG eliminates most dependency for *nearby* coeffs

ICA offers minimal advantage over PCA

Similar behaviors for coefficient blocks

[Lyu & Simoncelli, 2008]

Joint densities



- Nearby: densities are approximately circular/elliptical
- Distant: densities are approximately factorial

[Simoncelli, '97; Wainwright&Simoncelli, '99]

How do we build a global model that captures the full range of observed statistical behaviors?

1) Random Field of Gaussian Scale Mixtures

[Lyu & Simoncelli, 2008]

2) Build an implicit model, using local gain control. I'll show three recent examples...

Example 1: Density estimation

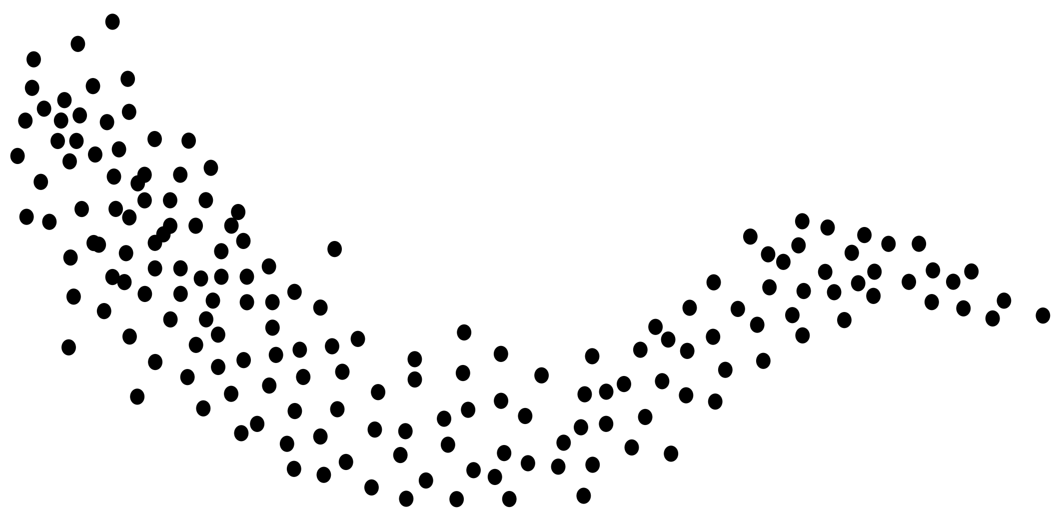
[Ballé, Laparra, Simoncelli, ICLR-16]

Density estimation (parametric density)

$$p_x(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\big(-f(\boldsymbol{x}; \boldsymbol{\theta})\big)$$

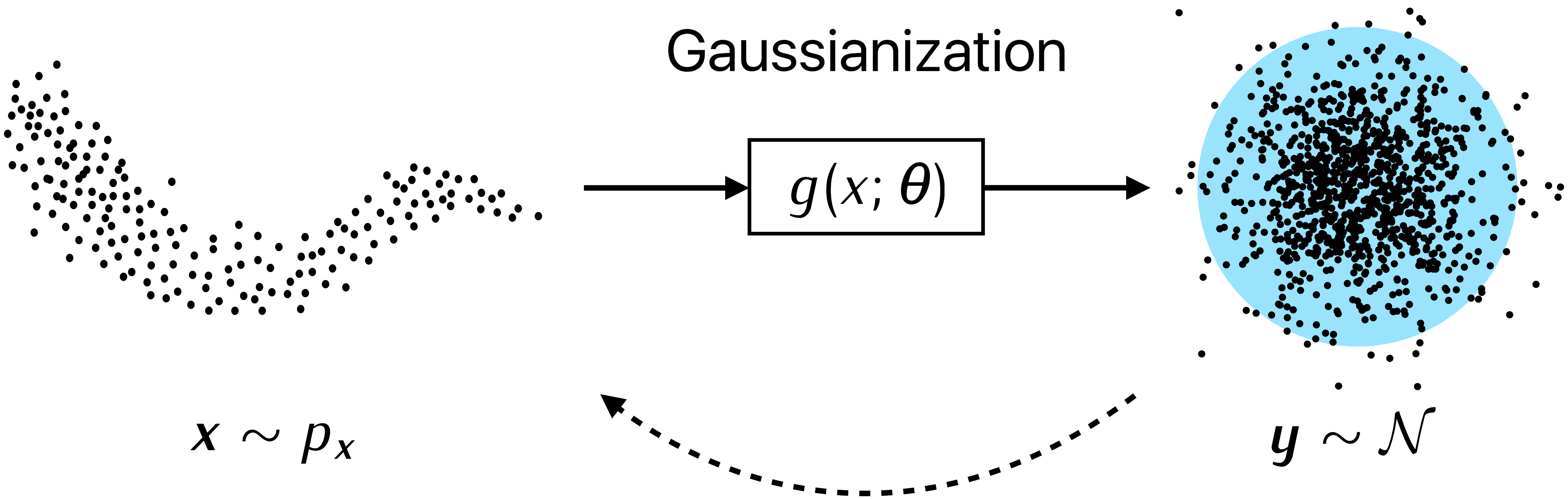
$$Z(\boldsymbol{\theta}) = \int \exp\big(-f(\boldsymbol{x}; \boldsymbol{\theta})\big) \, \mathrm{d}\boldsymbol{x}$$

tractable?



[Balle, Laparra, Simoncelli, ICLR-16]

Density estimation (parametric transformation)



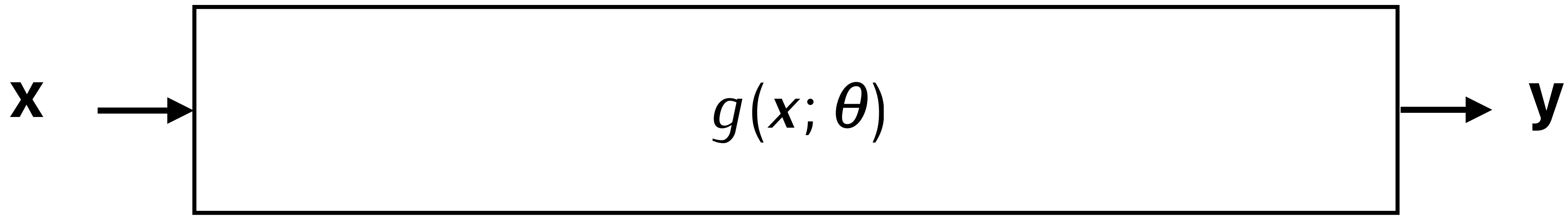
"inferred" density:

$$p_x(\boldsymbol{x}) = \left| \frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}} \right| \mathcal{N}\big(g(\boldsymbol{x};\boldsymbol{\theta})\big)$$

Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010
Dinh et al., 2015

[Balle, Laparra, Simoncelli, ICLR-16]

Parameter estimation

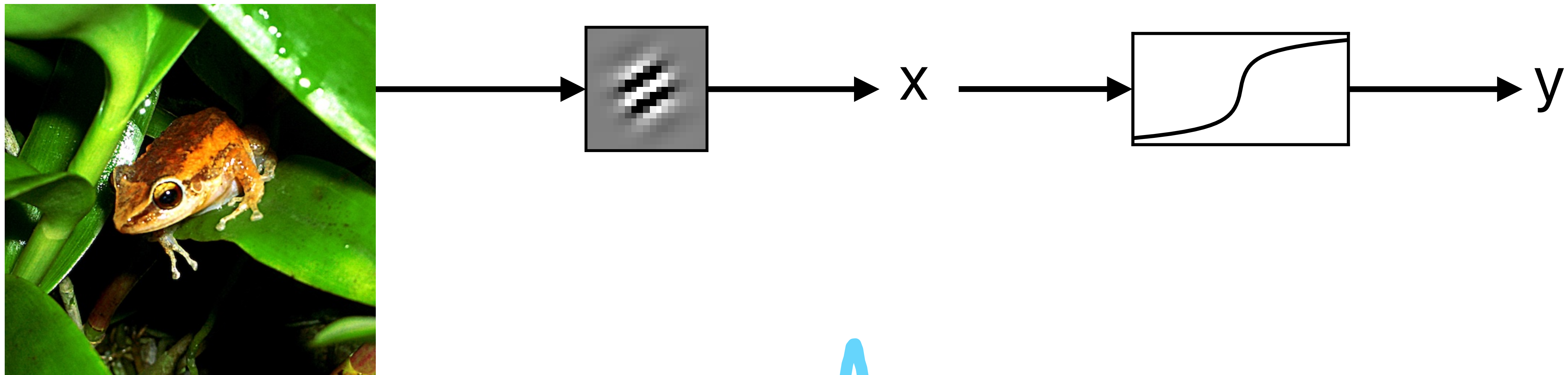


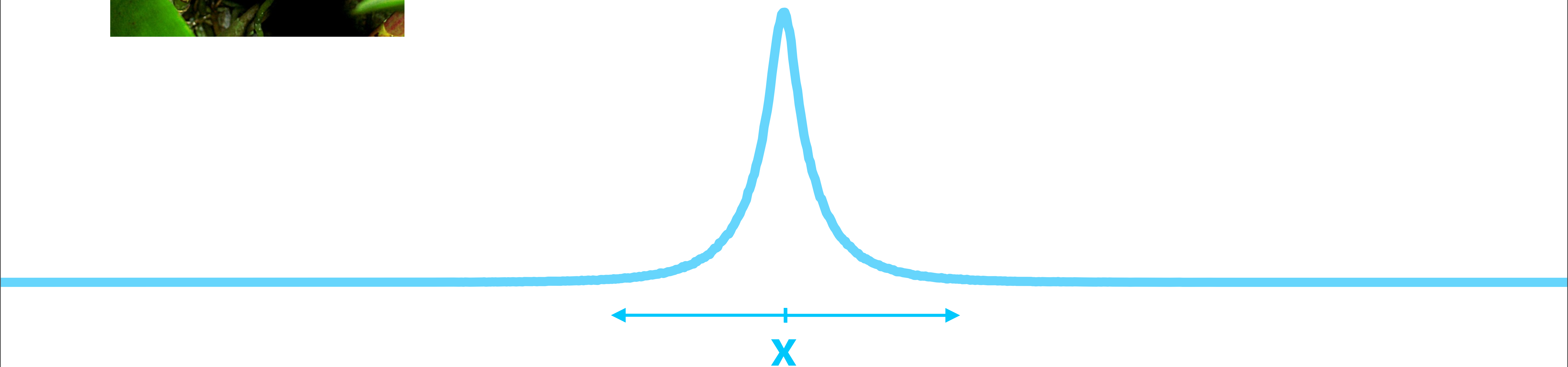
$$-\log p_{\boldsymbol{x}}(\boldsymbol{x}) = -\log\left|\frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}}\right| + \frac{1}{2}\|g(\boldsymbol{x};\boldsymbol{\theta})\|_2^2 + C$$

minimize wrt. θ using stochastic gradient descent

[Balle, Laparra, Simoncelli, ICLR-16]

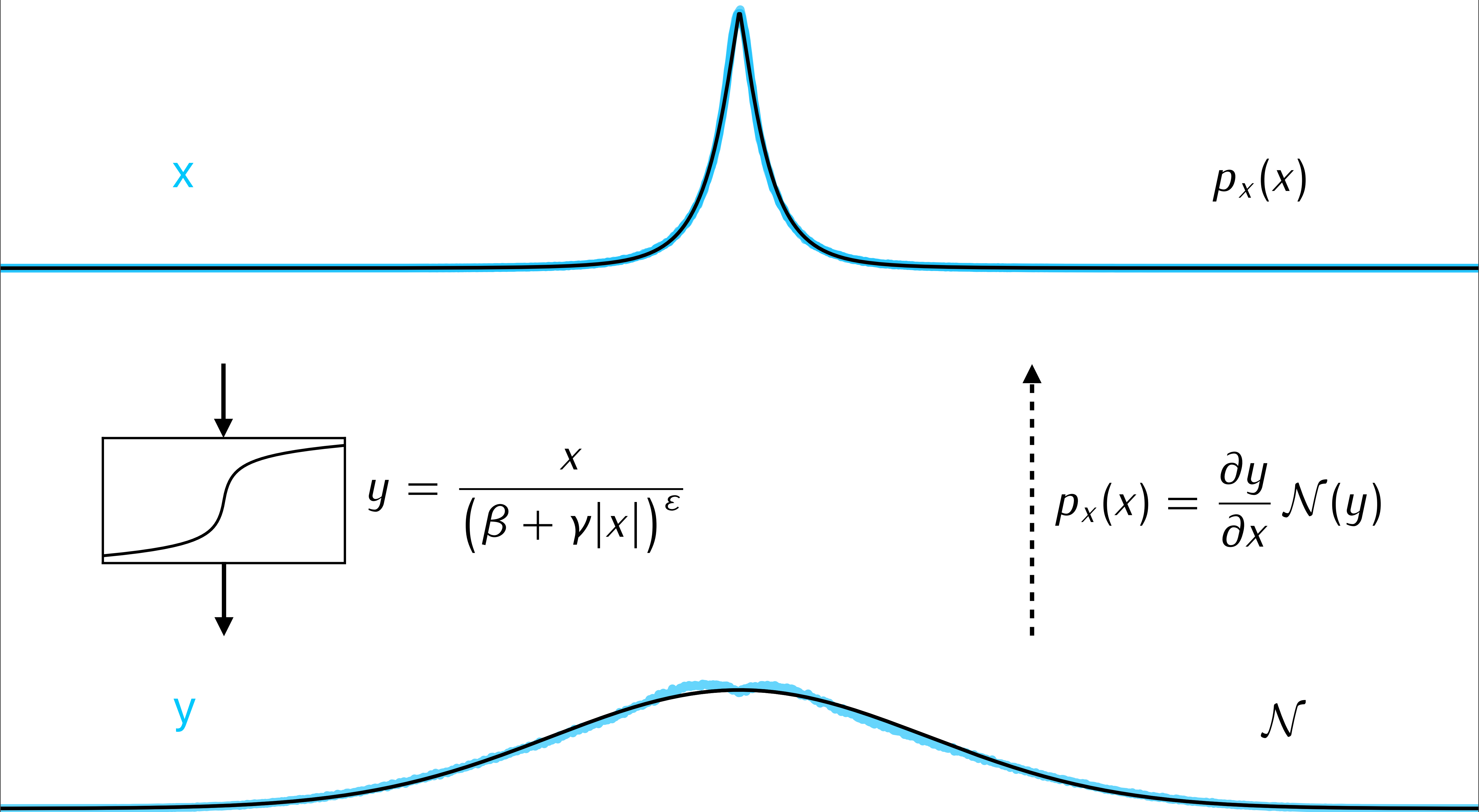
Marginal distribution of linear filter responses



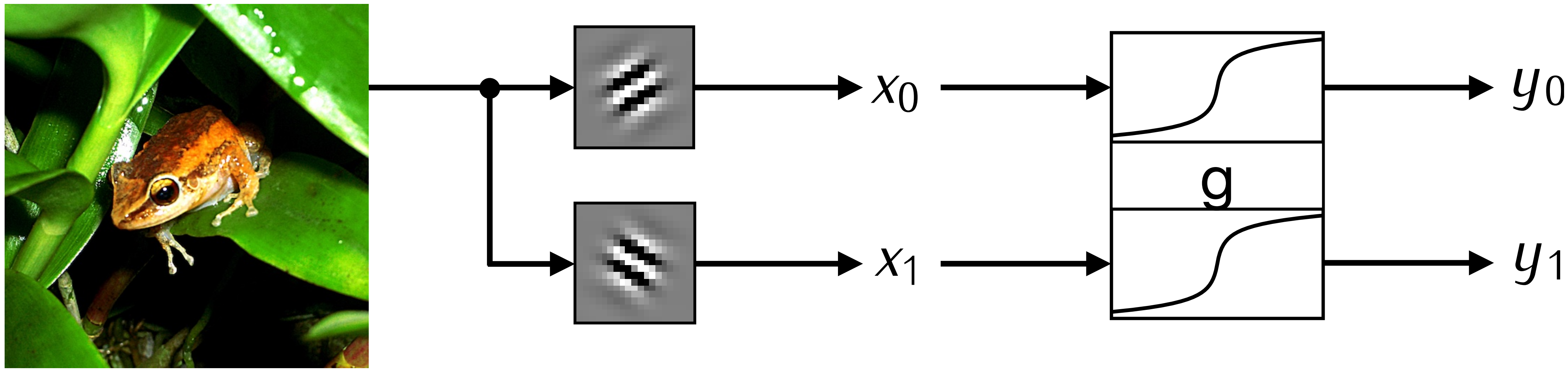


Burt & Adelson, 1981
Field, 1987
Mallat, 1989

[Balle, Laparra, Simoncelli, ICLR-16]



[Balle, Laparra, Simoncelli, ICLR-16]

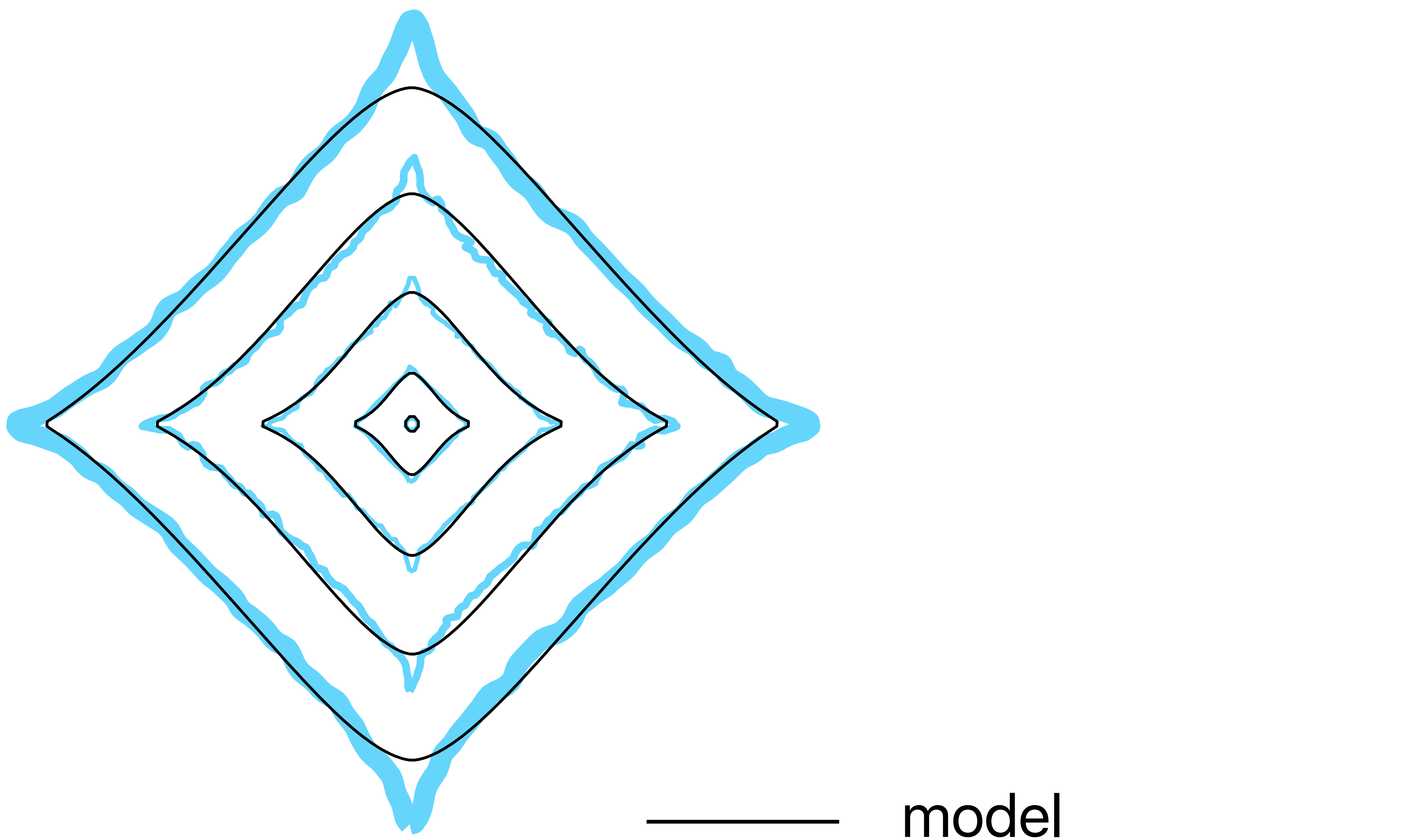


$$y_0 = \frac{x_0}{\left(\beta_0 + \gamma_{01}|x_1|^{\alpha_{01}} + \gamma_{00}|x_0|^{\alpha_{00}}\right)^{\varepsilon_0}}$$

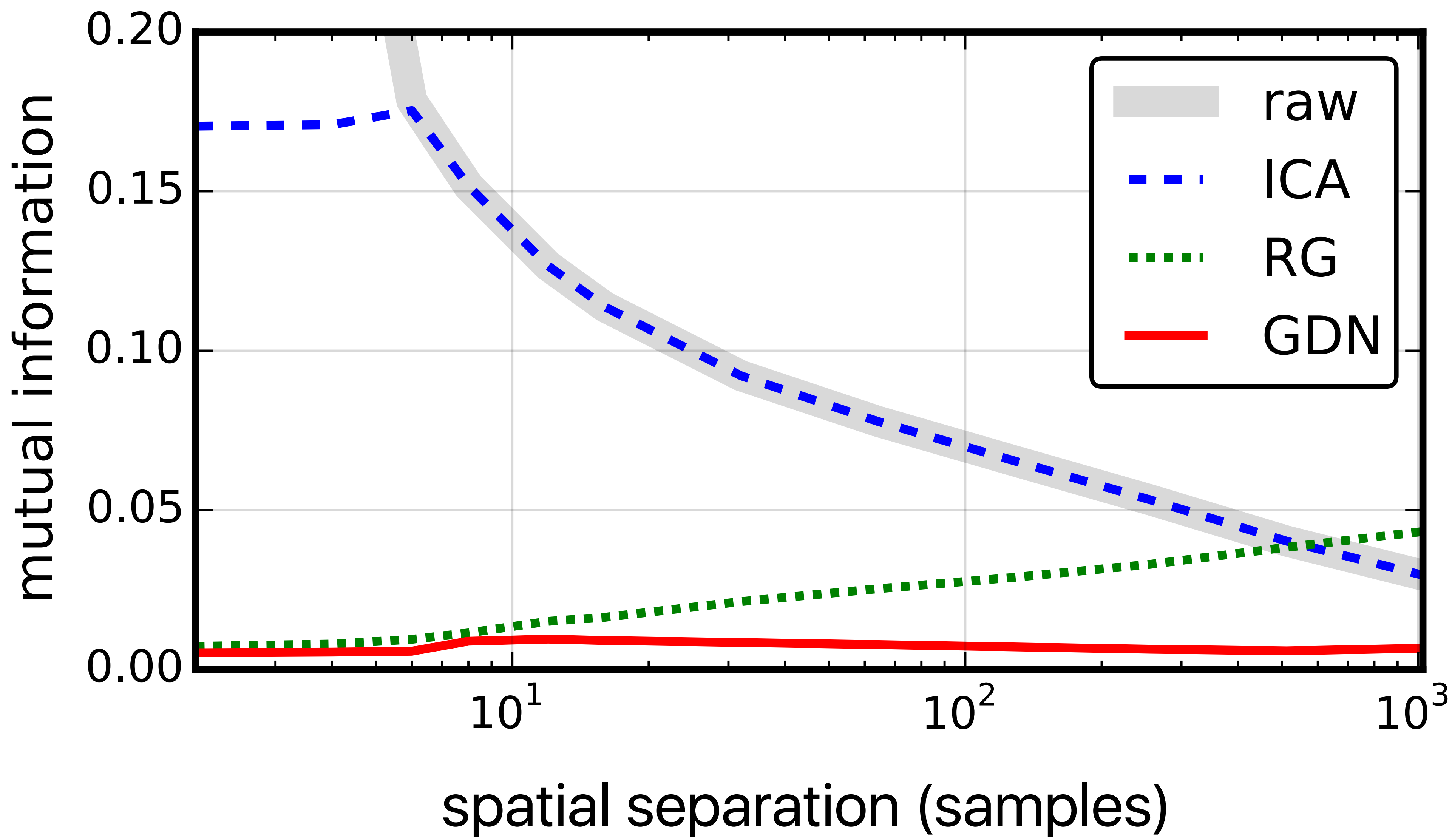
$$y_1 = \frac{x_1}{\left(\beta_1 + \gamma_{10}|x_0|^{\alpha_{10}} + \gamma_{11}|x_1|^{\alpha_{11}}\right)^{\varepsilon_1}}$$

[Balle, Laparra, Simoncelli, ICLR-16]

Contour lines, linear filter responses

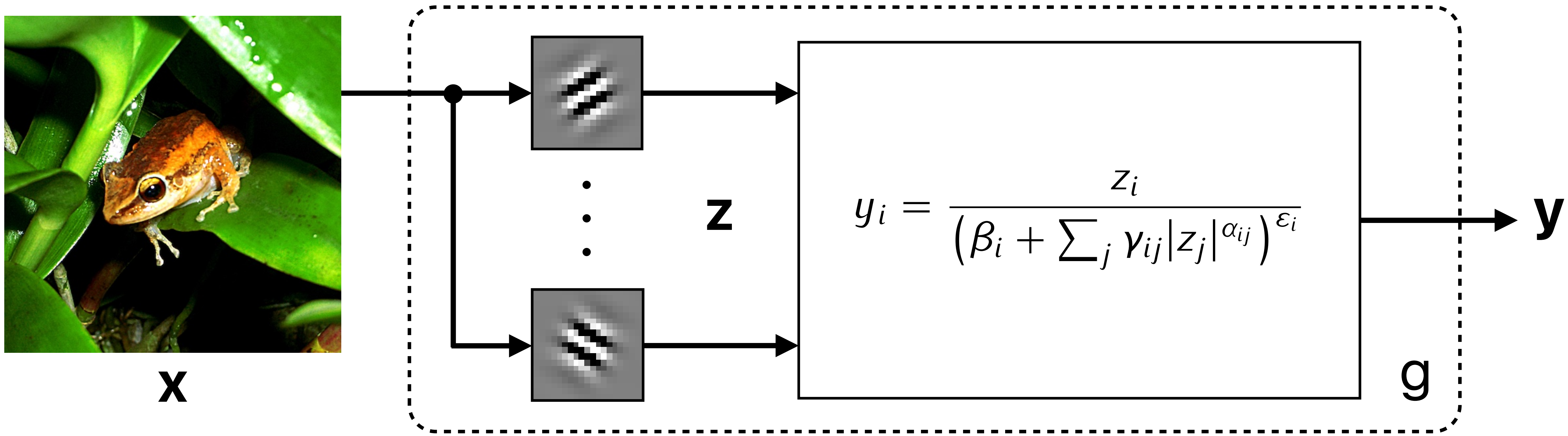


[Balle, Laparra, Simoncelli, ICLR-16]



[Balle, Laparra, Simoncelli, ICLR-16]

Generalized divisive normalization (GDN)

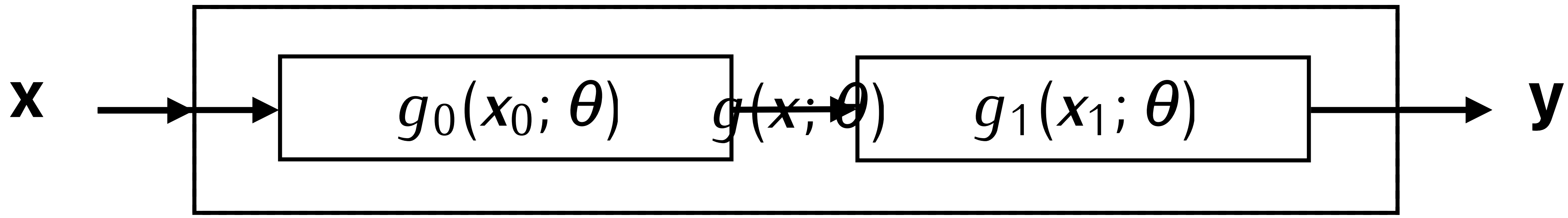


Special cases/related models:

- Independent Component Analysis, Cardoso, 2003
- Independent Subspace Analysis, Hyvärinen & Hoyer, 2000
- Weighted normalization model, Schwartz & Simoncelli, 2001
- Topographic ICA, Hyvärinen et al., 2001
- Radial Gaussianization, Lyu & Simoncelli, 2009
- L_p-nested symmetric distributions, Sinz & Bethge, 2010
- "Two-layer model", Köster & Hyvärinen, 2010

[Balle, Laparra, Simoncelli, ICLR-16]

Parameter estimation (multiple layers)

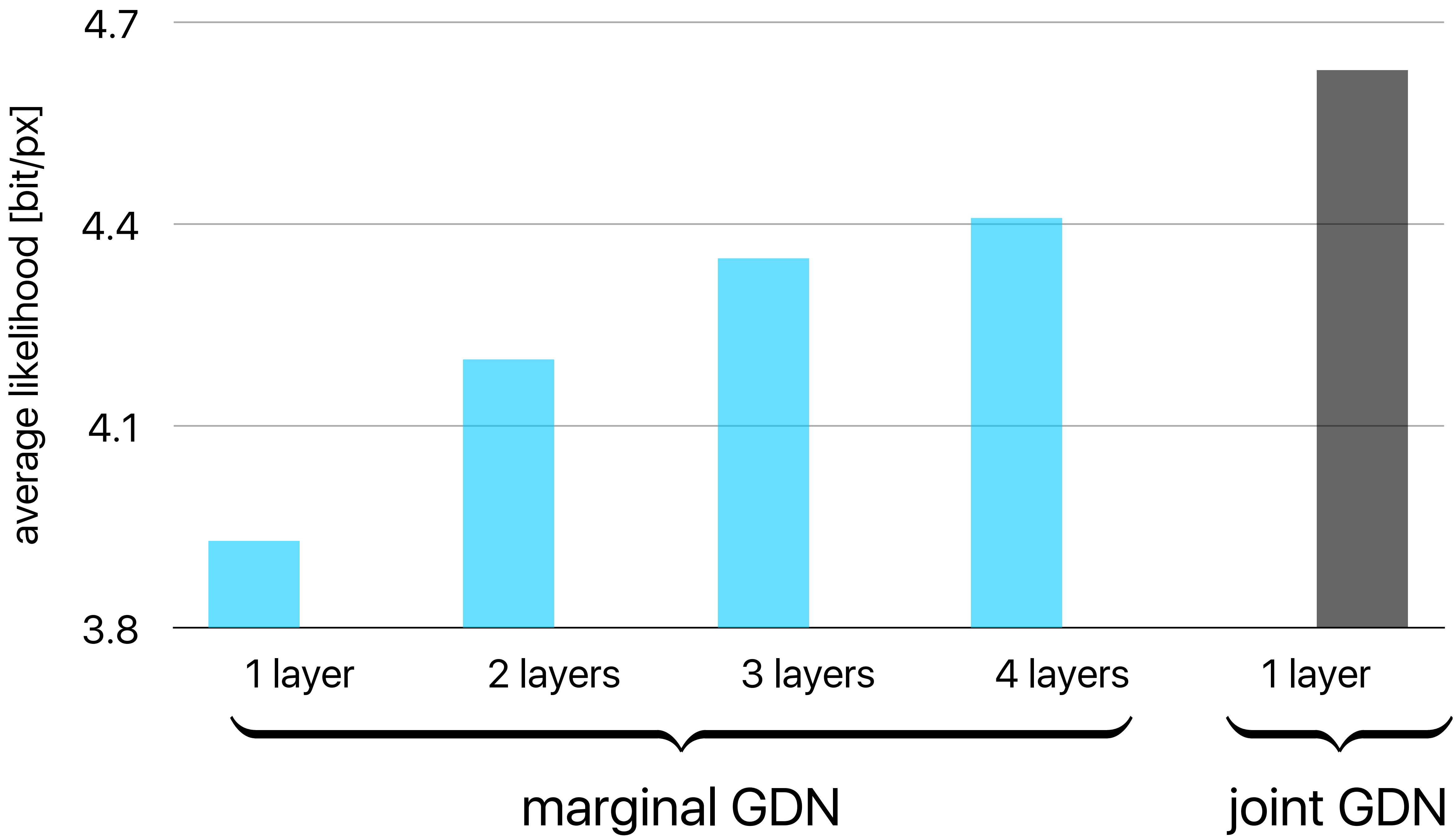


$$-\log p_{\boldsymbol{x}}(\boldsymbol{x}) = \underbrace{-\log\left|\frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}}\right|}_{-\log\left|\frac{\partial g_0(\boldsymbol{x}_0;\boldsymbol{\theta})}{\partial \boldsymbol{x}_0}\right| - \log\left|\frac{\partial g_1(\boldsymbol{x}_1;\boldsymbol{\theta})}{\partial \boldsymbol{x}_1}\right| - \dots} - \frac{1}{2}\left\|g(\boldsymbol{x};\boldsymbol{\theta})\right\|_2^2 + C$$

minimize wrt. $\boldsymbol{\theta}$ using stochastic gradient descent

[Balle, Laparra, Simoncelli, ICLR-16]

One layer of joint GDN > many layers of marginal GDN



[Balle, Laparra, Simoncelli, ICLR-16]

Example 2: Perceptually-optimized rendering

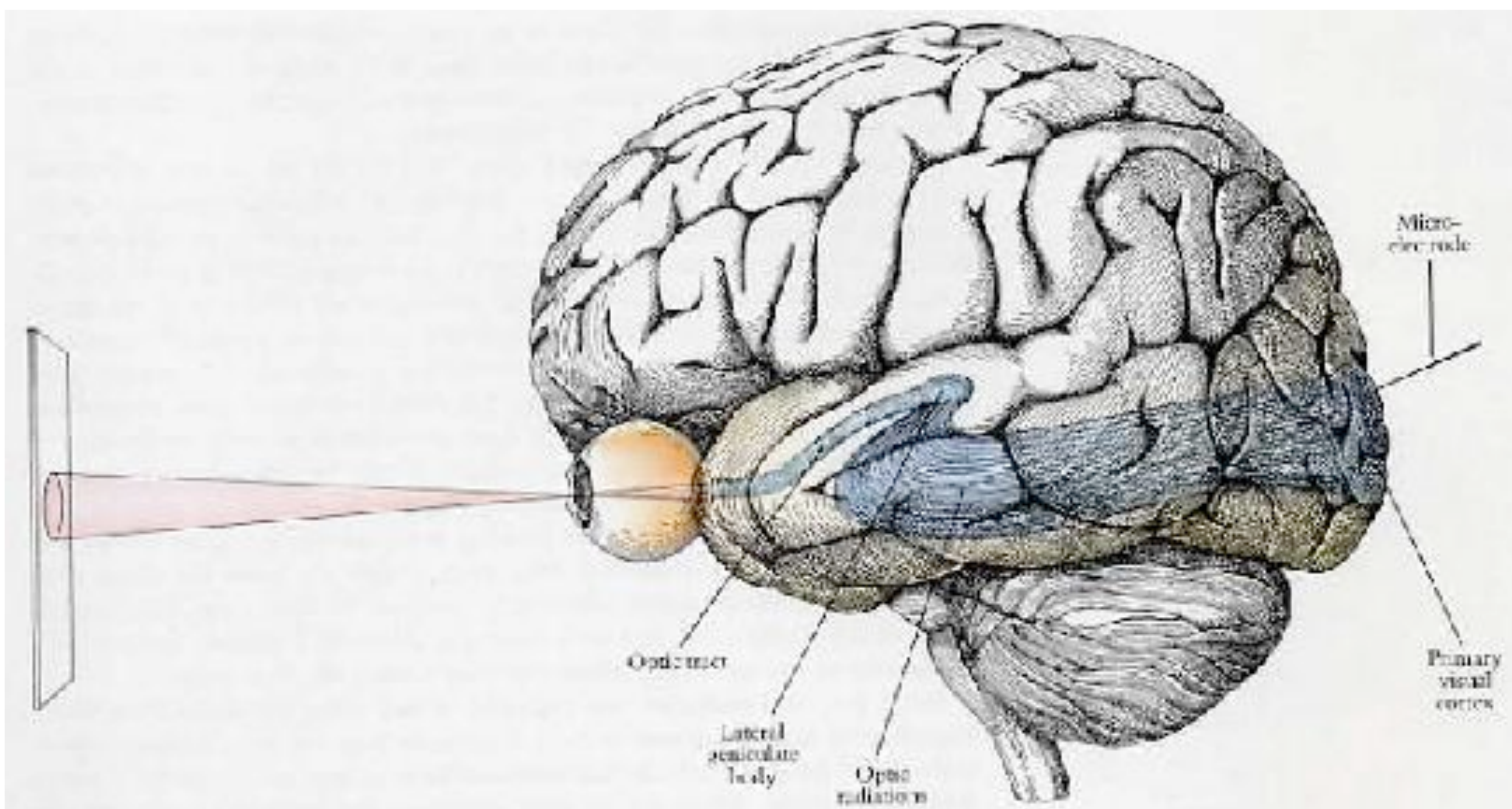
[Laparra, Ballé, Berardino & Simoncelli, in preparation]

Perceptual error is not consistent with Mean Squared Error

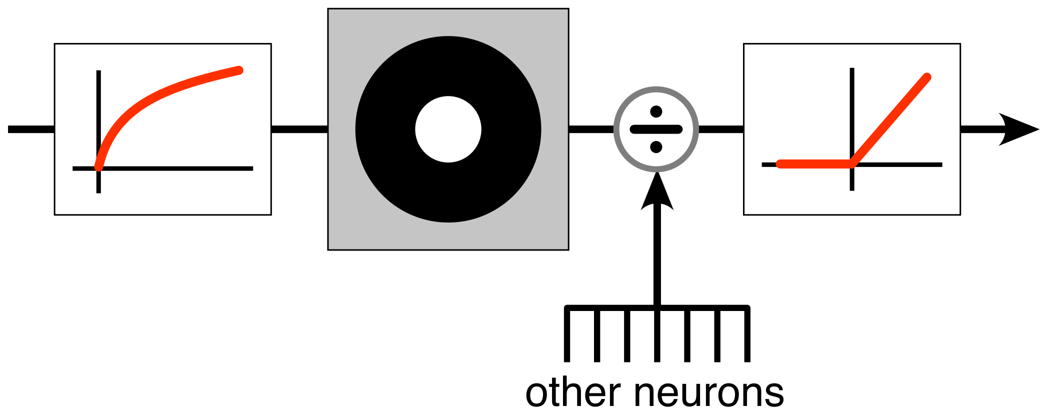
Equal MSE:

Why does MSE fail?

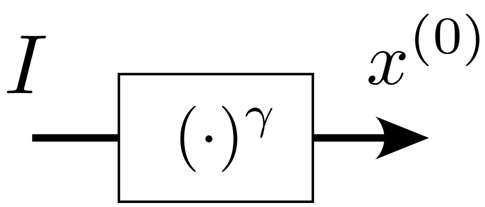
Human visual system constructs a nonlinear visual representation, and is sensitive to distortions in this space

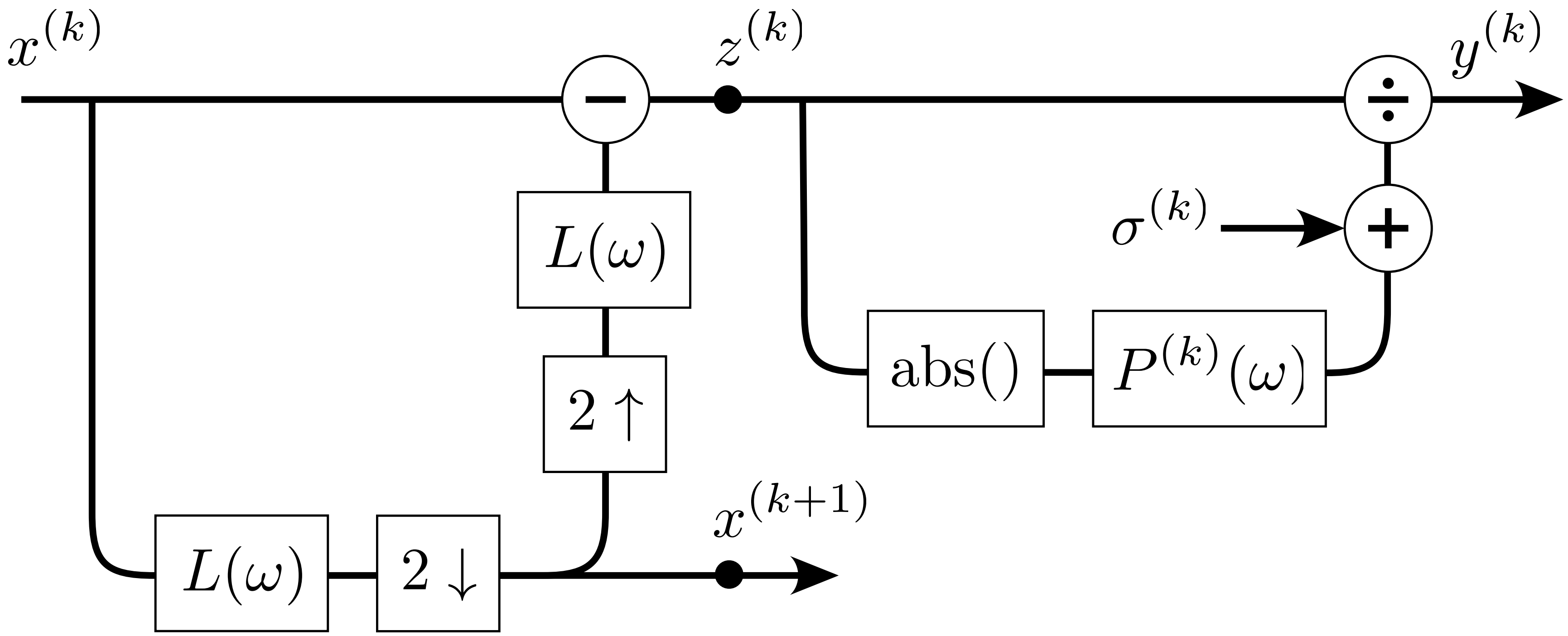


Simple retinal gain-control model:



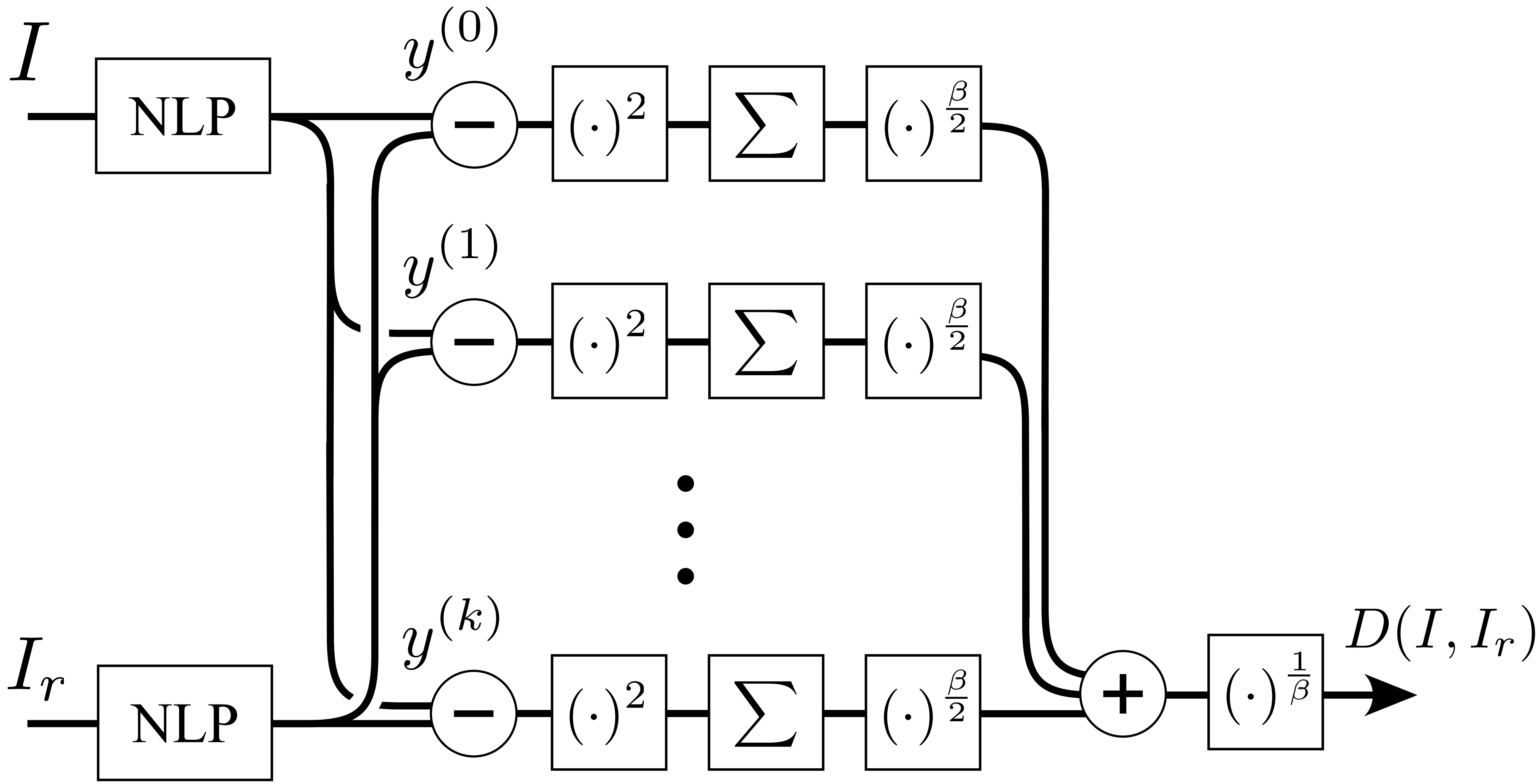
Multiscale version: Normalized Laplacian Pyramid (NLP)





[Laparra, et. al. 2016]

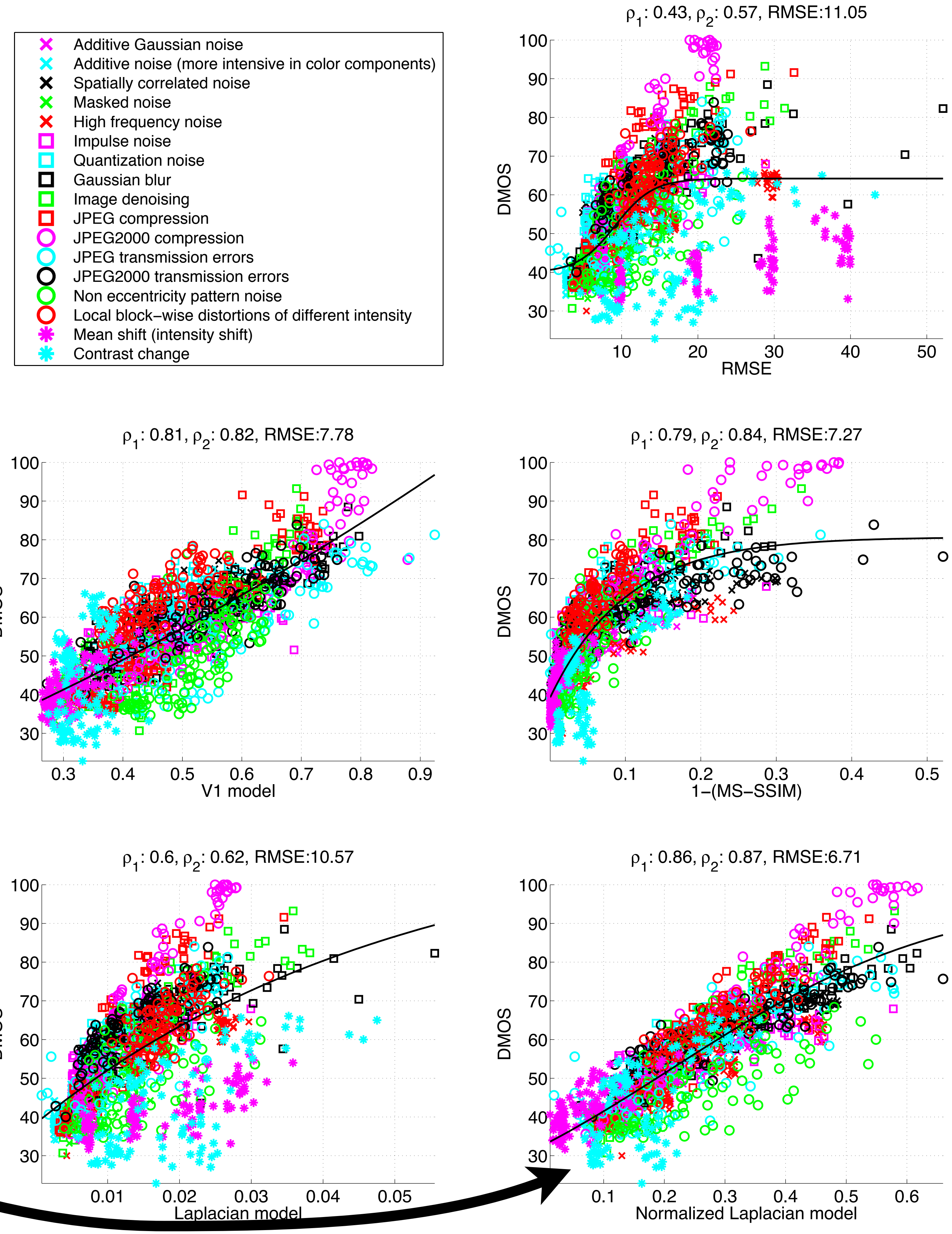
NLP Distortion metric



[Laparra, et. al. 2016]

TID2008 database

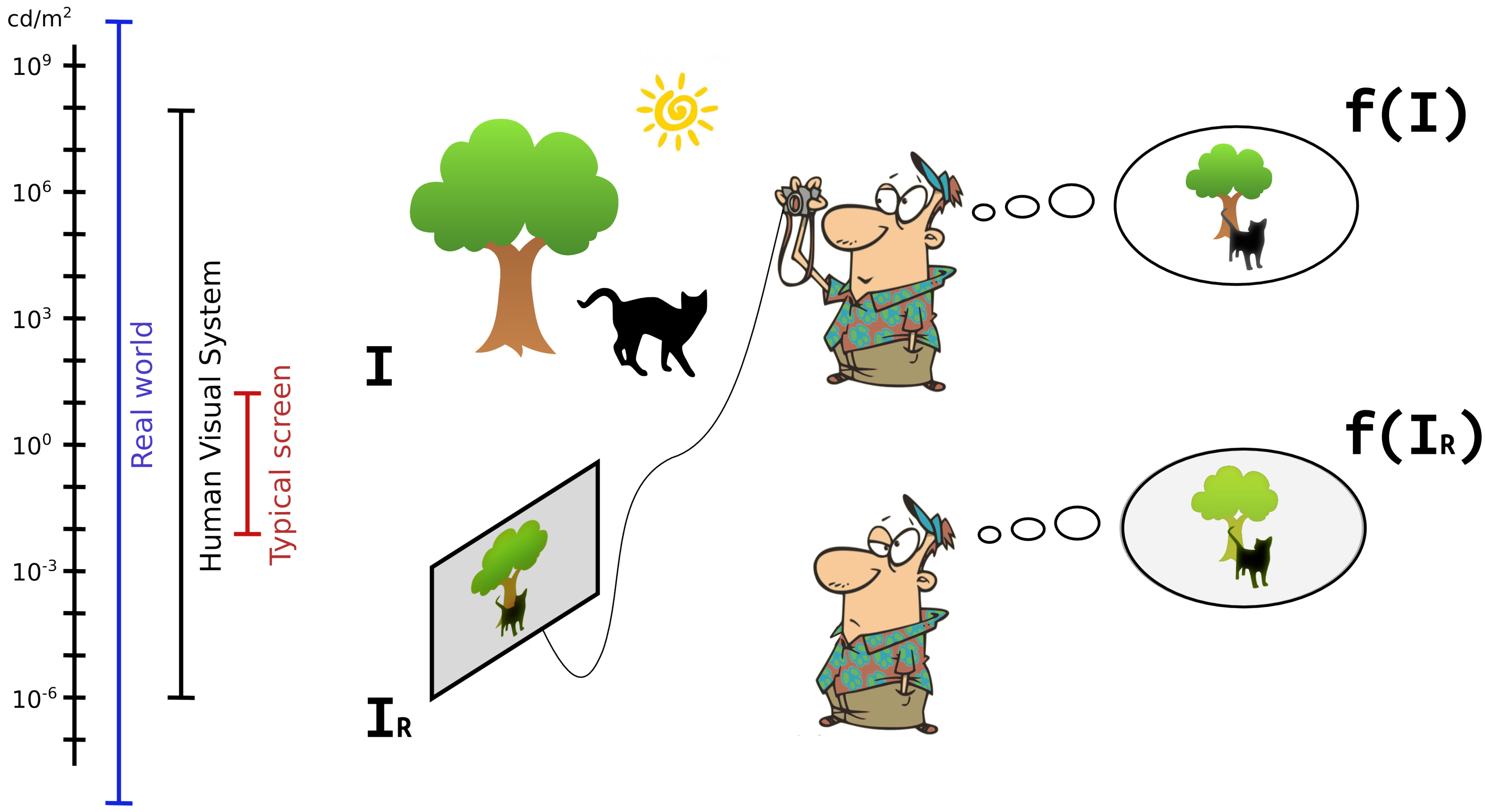
[Ponomarenko et al., 2009]



MSE in retinal model response space explains perceptual data

[Laparra et al, 2016]

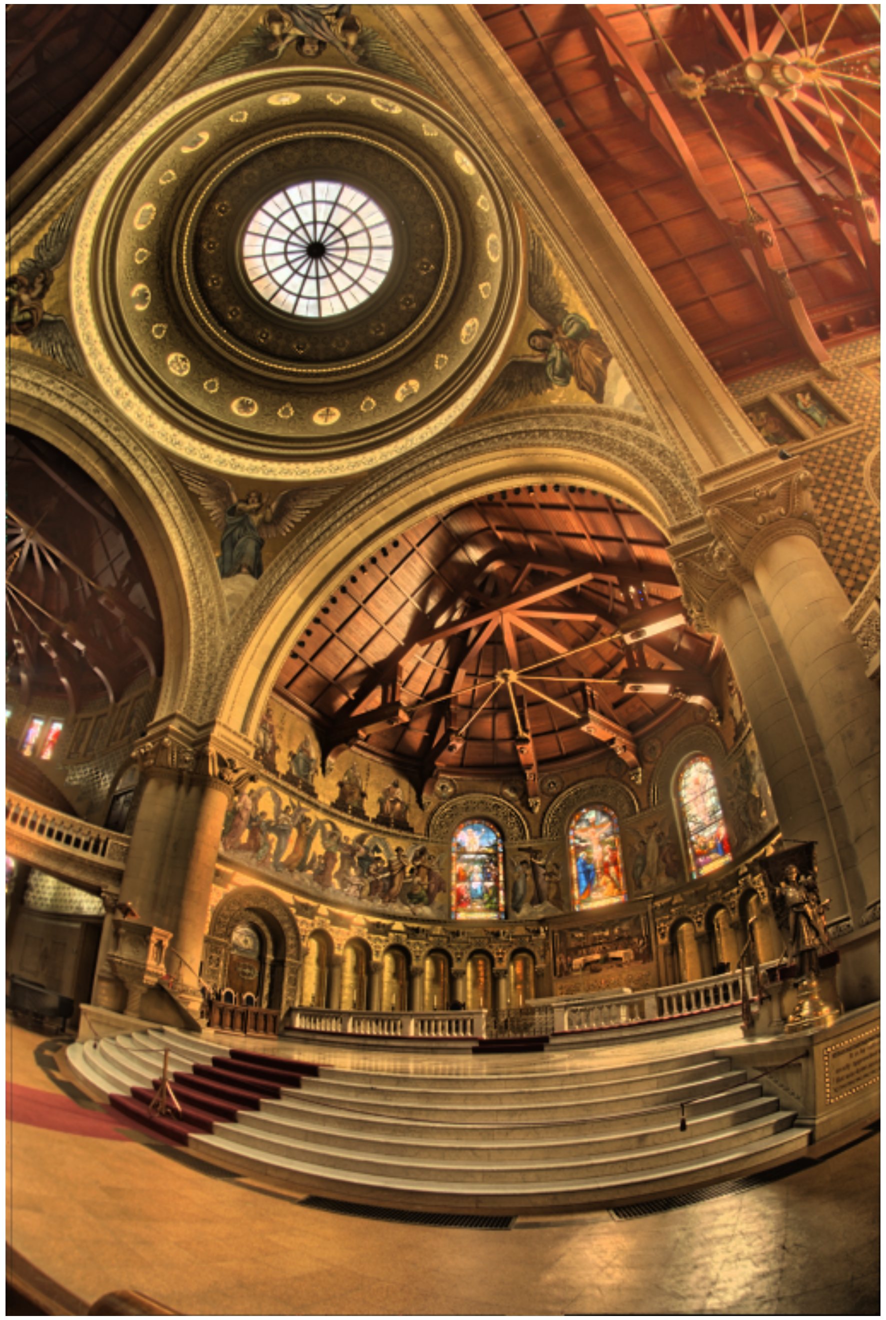
The image rendering problem



Automatic high dynamic range image rendering

original

Paris et. al., 2015



Laparra et. al. (in preparation)

Perceptually-optimized image rendering

original

Paris et. al., 2015

Lmax = 10^3

Lmax = 10^4

Lmax = 10^5

Dithering (binary rendering)

original
(grayscale)

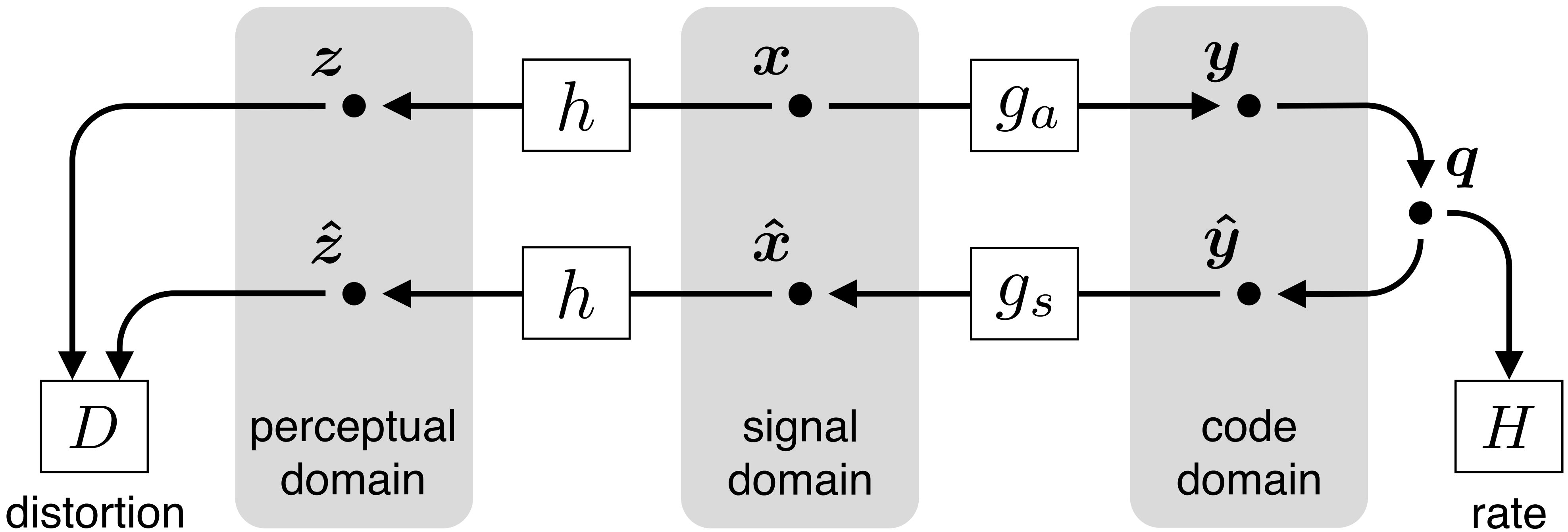
standard
(Floyd-Steinberg 1976)

Laparra et. al.
(in preparation)

Example 3: Compression

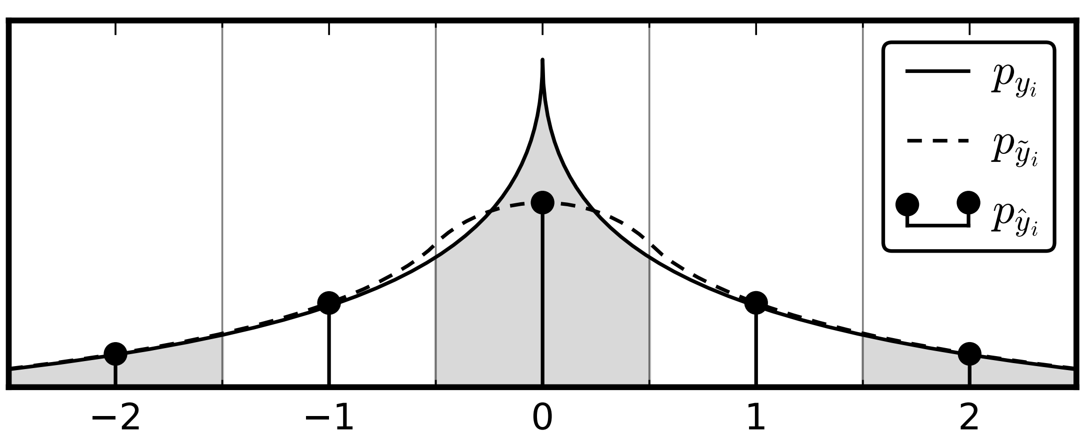
[Ballé, Laparra & Simoncelli, PCS-16+]

End-to-end rate-distortion optimization



objective: $L[g_a, g_s] = H[P_{\boldsymbol{q}}] + \lambda \, \mathbb{E} \, \|\boldsymbol{z} - \hat{\boldsymbol{z}}\|$

relaxation to differential entropy:

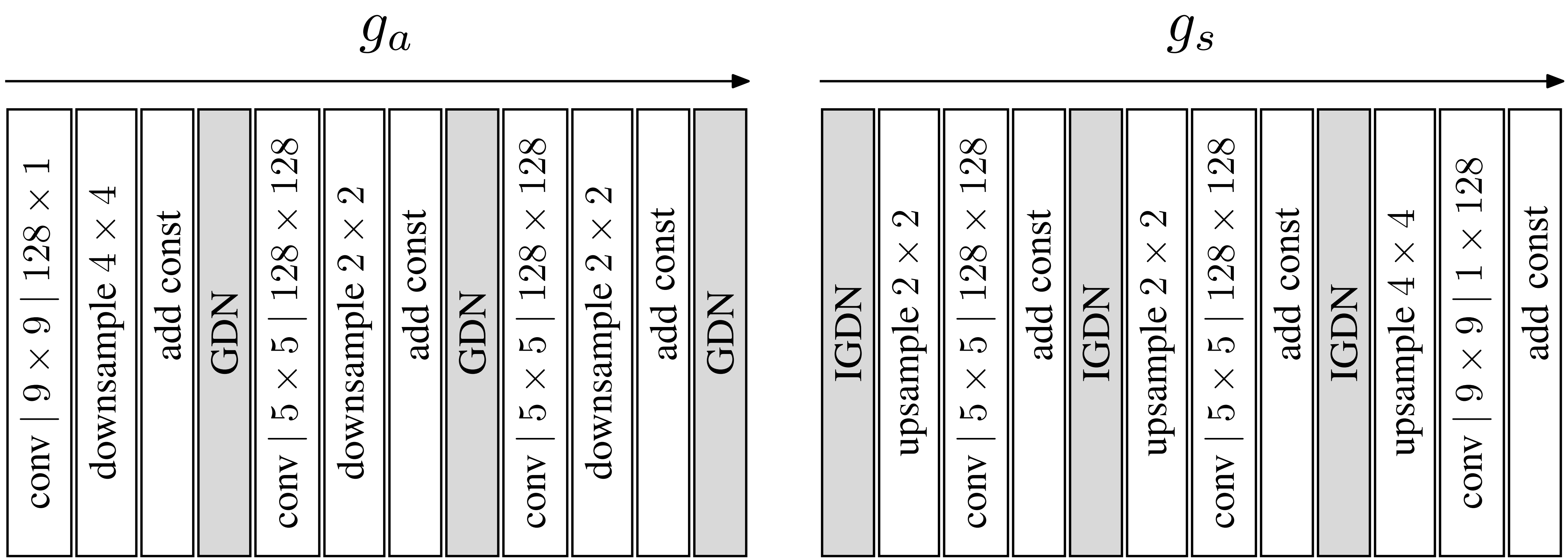


$$p_{\hat{y}_i}(t) = \sum_{n=-\infty}^{\infty} P_{q_i}(n) \, \delta(t - n),$$

$$P_{q_i}(n) = \big(p_{y_i} * \mathcal{U}(0,1)\big)(n), \text{ for all } n \in \mathbb{Z},$$

[Balle et al., PCS-16]

3-stage GDN cascade



GDN:

$$y_i = \frac{x_i}{(\beta_i + \sum_j \gamma_{ij} x_j^2)^{\frac{1}{2}}}$$

γ_{ij} symmetric

IGDN (approximate):

$$x_i = y_i \cdot (\beta_i + \sum_j \gamma_{ij} y_j^2)^{\frac{1}{2}}$$

(one step of fixed-point inverse)

[Balle et al., in preparation]

original (cropped)

jpeg: 9094 bytes, RMSE: 12.032

jpeg2000: 8362 bytes, RMSE: 11.264

3-stage GDN: 8360 bytes, RMSE: 8.16

[Balle et al., in preparation]

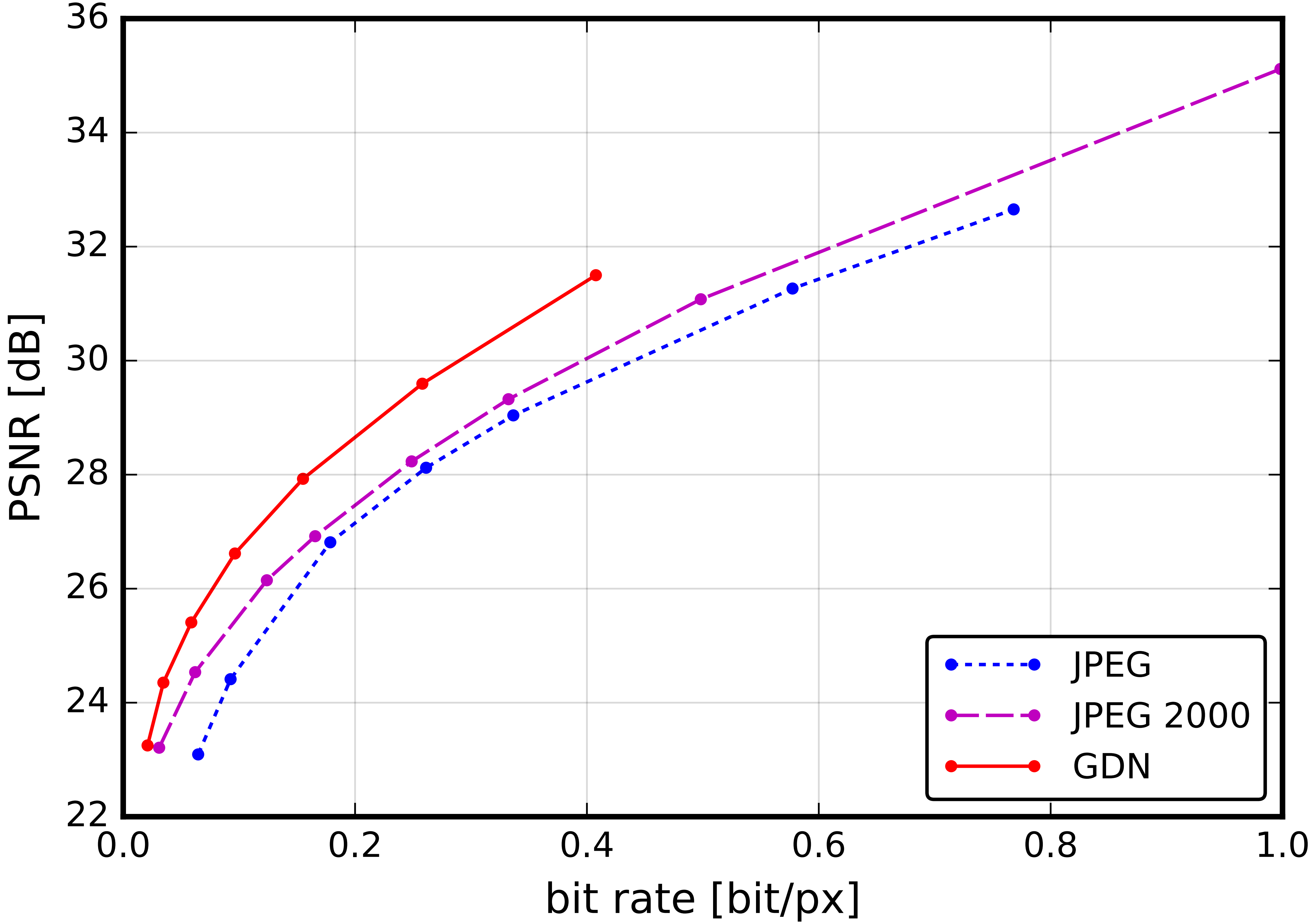
original (cropped)

jpeg: 9851 bytes, RMSE: 18.84

jpeg2000: 8127 bytes, RMSE: 18.37

3-stage GDN: 8115 bytes, RMSE: 15.95

[Balle et al., in preparation]



[Balle et al., in preparation]

Local gain control...

- is found throughout biological sensory systems
- can be implemented as an invertible nonlinear transform
- can Gaussianize natural signals, eliminating dependencies
- can mimic human perception of visual distortions
- can be used, cascaded, for image compression
- but we need a more complete characterization/ design toolbox!

Thanks

Johannes Ballé

Valero Laparra

Alex Berardino

Siwei Lyu

hhmi
Howard Hughes Medical Institute

National Eye Institute
NATIONAL INSTITUTES OF HEALTH

NSF