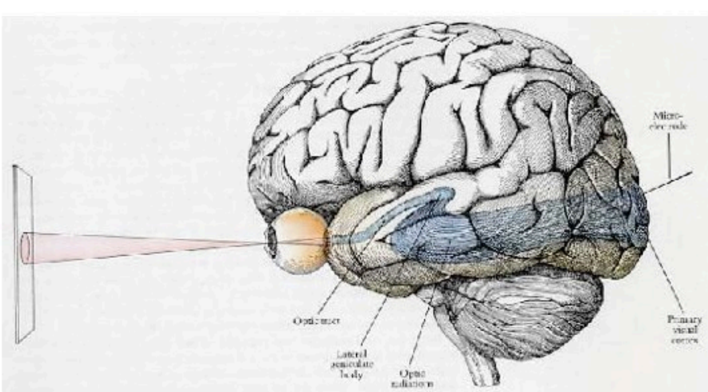
Image Statistics and Efficient Sensory Coding

NEURL-GA 3042: Visual Neuroscience

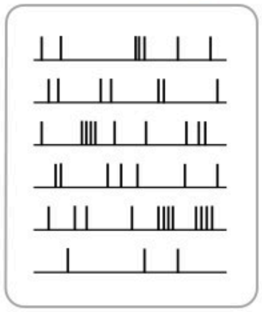
Eero Simoncelli

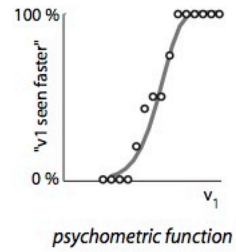
Why image statistics?

- Engineering: compression, denoising, restoration, enhancement/modification, synthesis, manipulation
- Science: optimality principles for neurobiology (evolution, development, learning, adaptation)



Stimulus — Neural response — Behavior





"Encoding" "Decoding"

Efficient coding theory
[Barlow '61]

Optimal estimation/decision
[Al Hazan, 1040; Helmholtz, 1866]

Both theories rely on statistical models of environment

13

H. B. BARLOW
Physiological Laboratory, Cambridge University

Possible Principles Underlying the Transformations of Sensory Messages

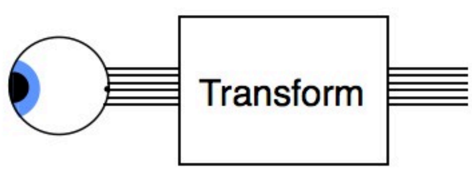
A wing would be a most mystifying structure if one did not know that birds flew. One might observe that it could be extended a considerable distance, that it had a smooth covering of feathers with conspicuous markings, that it was operated by powerful muscles, and that strength and lightness were prominent features of its construction. These are important facts, but by themselves they do not tell us that birds fly. Yet without knowing this, and without understanding something of the principles of flight, a more detailed examination of the wing itself would probably be unrewarding. I think that we may be at an analogous point in our understanding of the sensory side of the central nervous system. We have got our first batch of facts from the anatomical, neurophysiological, and psychophysical study of sensation and perception, and now we need ideas about what operations are performed by the various structures we have examined. For the bird's wing we can say that it accelerates downwards the air flowing past it and so derives an upward force which supports the weight of the bird; what would be a similar summary of the most important operation performed at a sensory relay?

... to have in mind possible answers

A mathematical theory of communication
CE **Shannon** - Bell System Technical Journal, v 27, 1948
... CE **Shannon** ... The message may be of various types: (a) A sequence of letters as in a telegraph of teletype system; (b) A single function of time f(t) as in radio or telephony; (**c**) A function of time and other variables as in black and white television -- here the message may be ...
Cited by 62650 Related articles All 782 versions Import into BibTeX More

Efficient Coding

[Attneave '54; Barlow '61; Laughlin '81; Atick '90; Bialek etal '91]

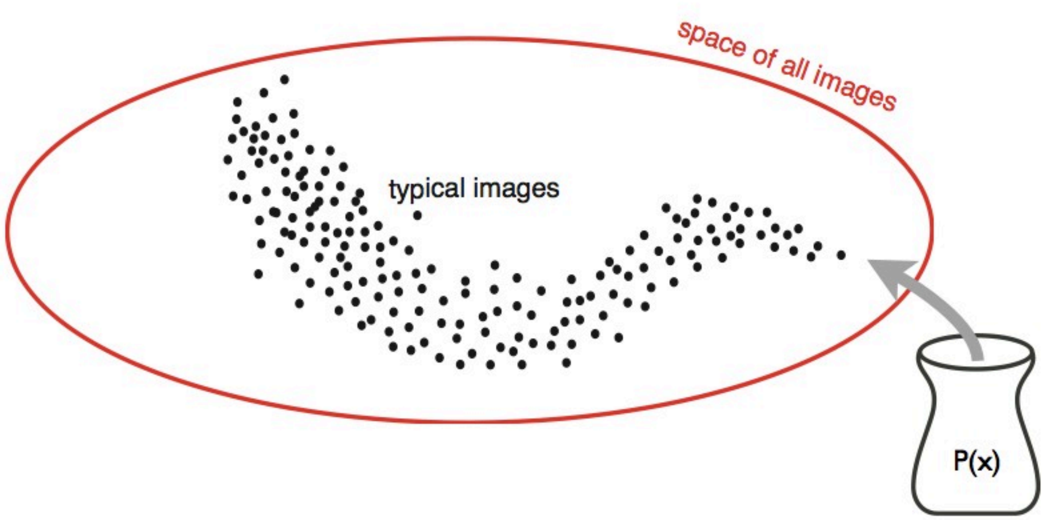


Maximize information about stimulus in response, subject to constraints (e.g., number of neurons, spike rate, metabolic costs, wiring length)

Why should efficient coding of sensory information matter to the brain?

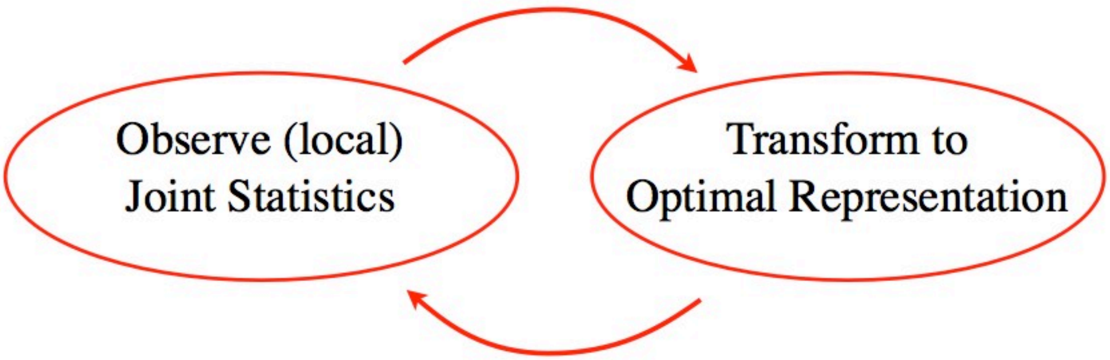
- Incoming stimuli are highly *redundant*, and do not span the full space of possible input signals
- Resources (for communicating / processing / storing) are *limited*
- Not all incoming sensory information is behaviorally *relevant*

Image statistics



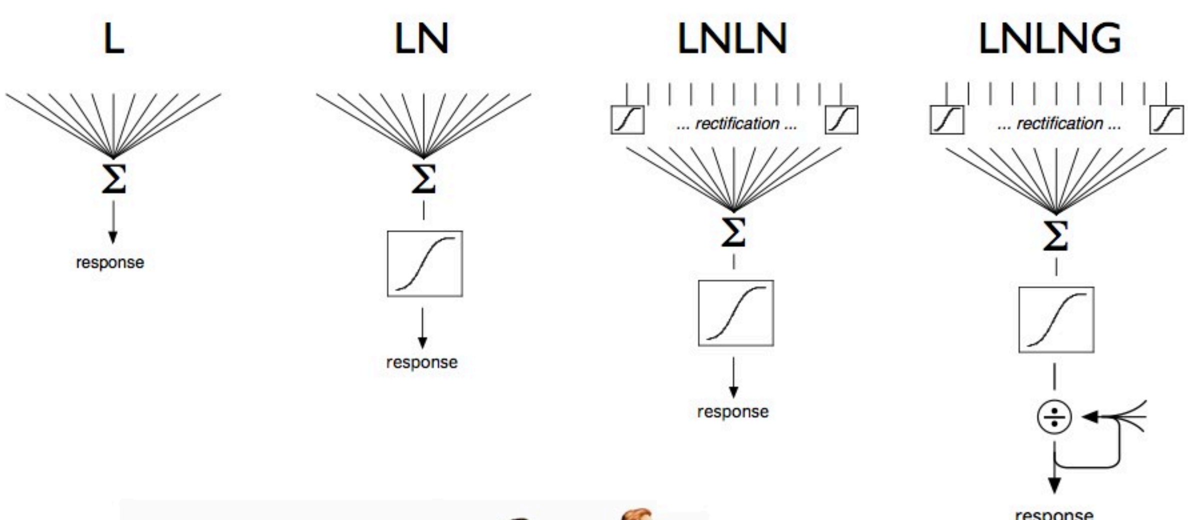
Can't characterize this directly, because dimensionality is too high!

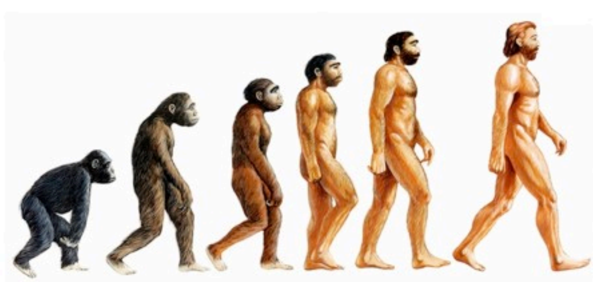
General methodology



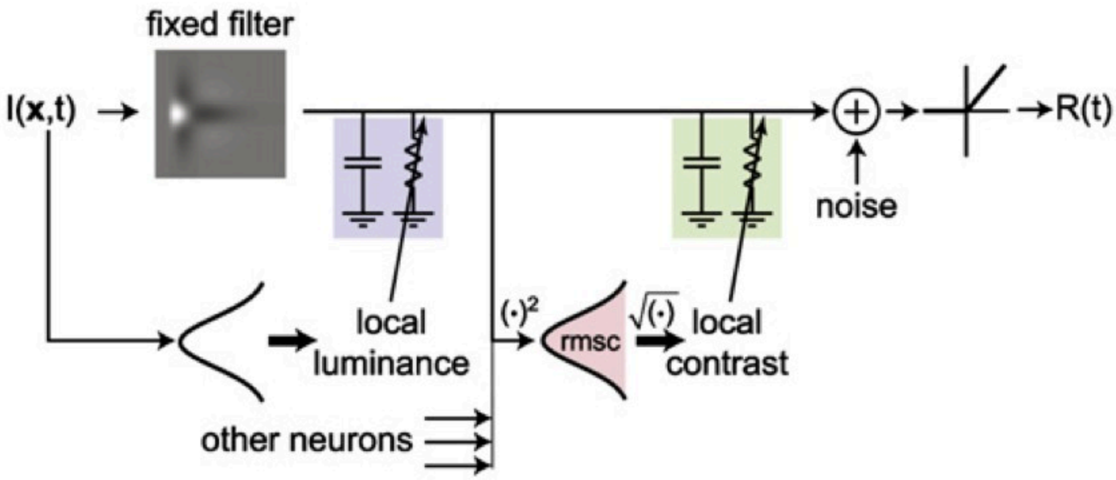
"Onion peeling"

Canonical neural models - retina, lgn, V1, MT, V2 ...



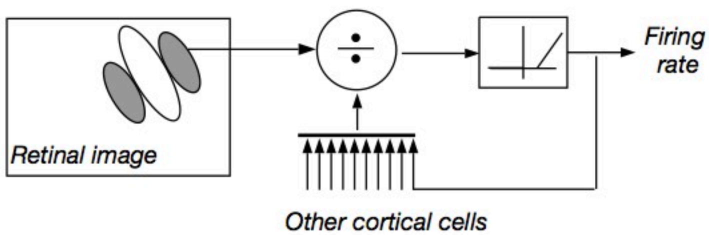


LGN

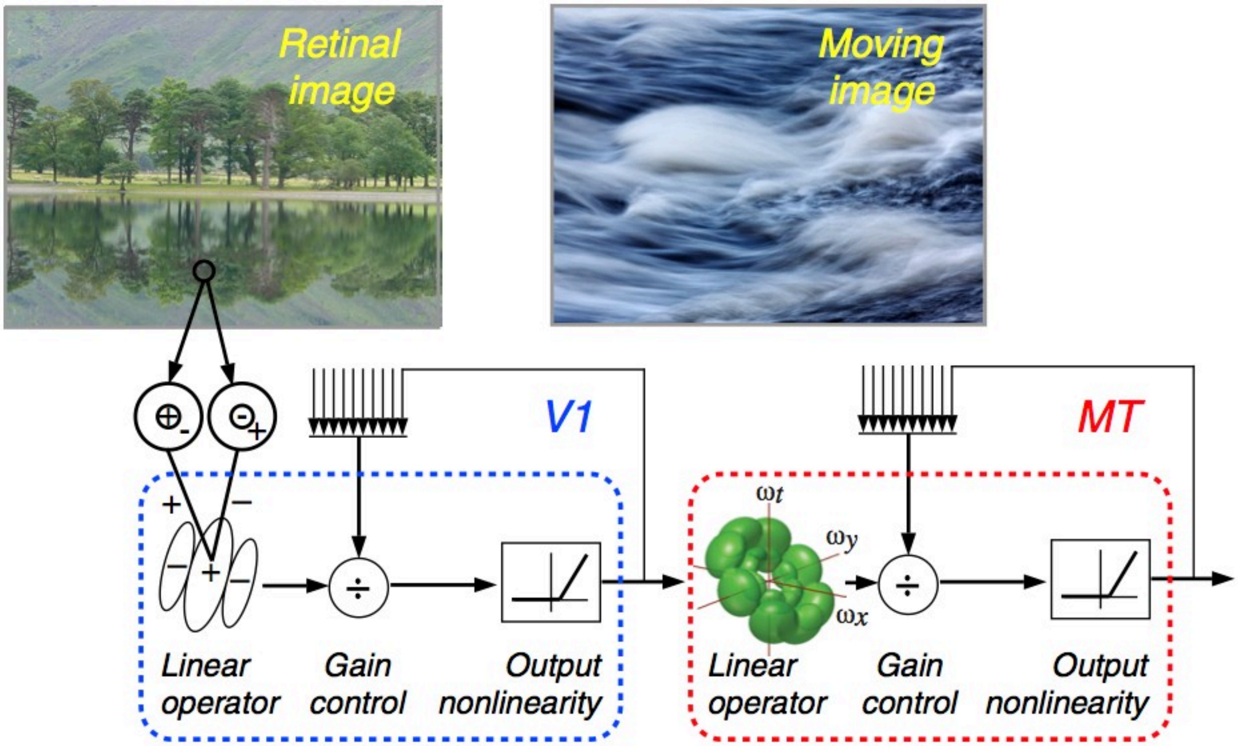


[Mante, Bonin, Carandini 2008]

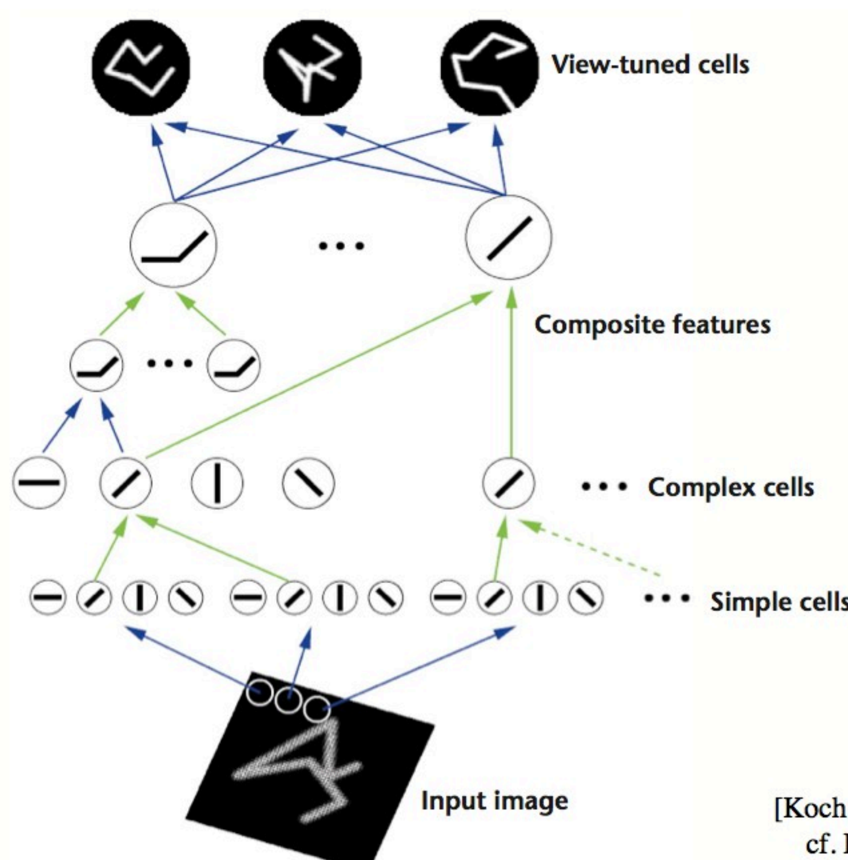
The normalization model of simple cells



[Carandini, Heeger, and Movshon, 1996]

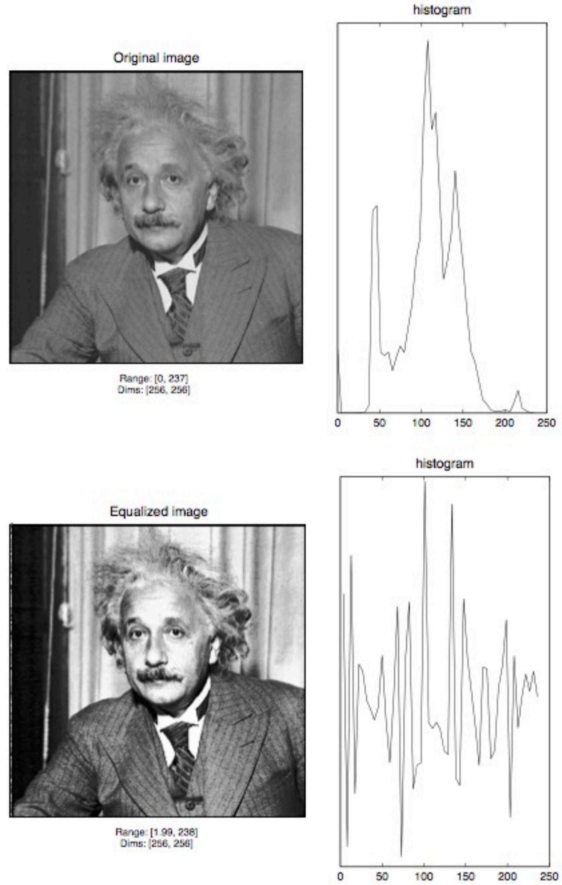


Simoncelli & Heeger, 1998; Rust, Mante, Simoncelli & Movshon, 2006; Nishimoto & Gallant, 2011



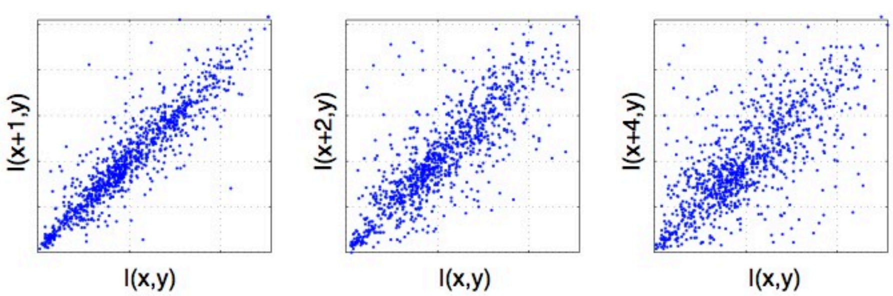
[Koch & Poggio, 1999;
cf. Fukishima, 1980;
Serre, Oliva, Poggio 2007; etc]

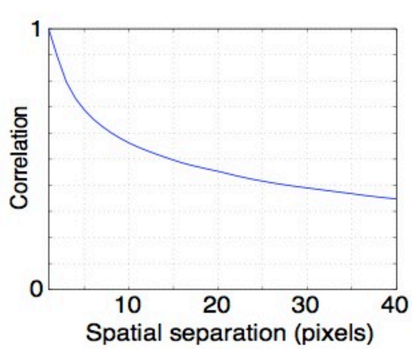
Pixel histograms carry very little information about image content...



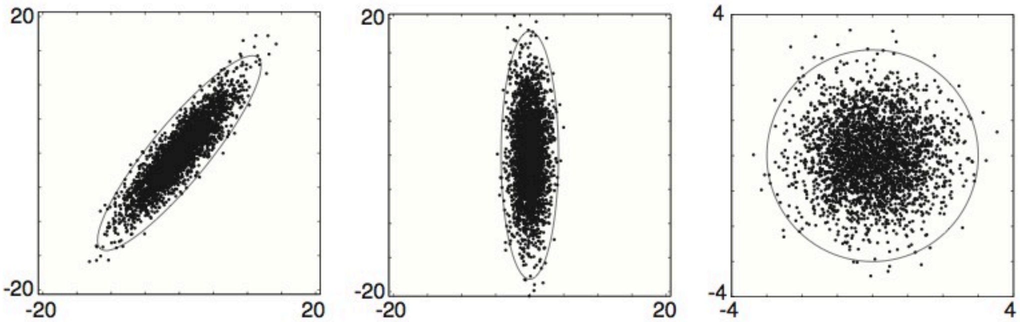
different histogram

Pixel correlation





Principal Components Analysis (PCA)



For Gaussian sources: guaranteed independence

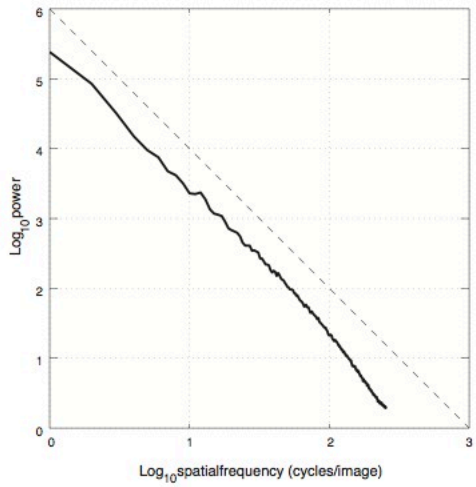
PCA on images

- Translation-invariance => Fourier PCA
- Scale-invariance => power law spectrum

$$\tilde{I}(\omega) = A\omega^p \quad \Longrightarrow \quad \tilde{I}(s\omega) = As^p\omega^p$$

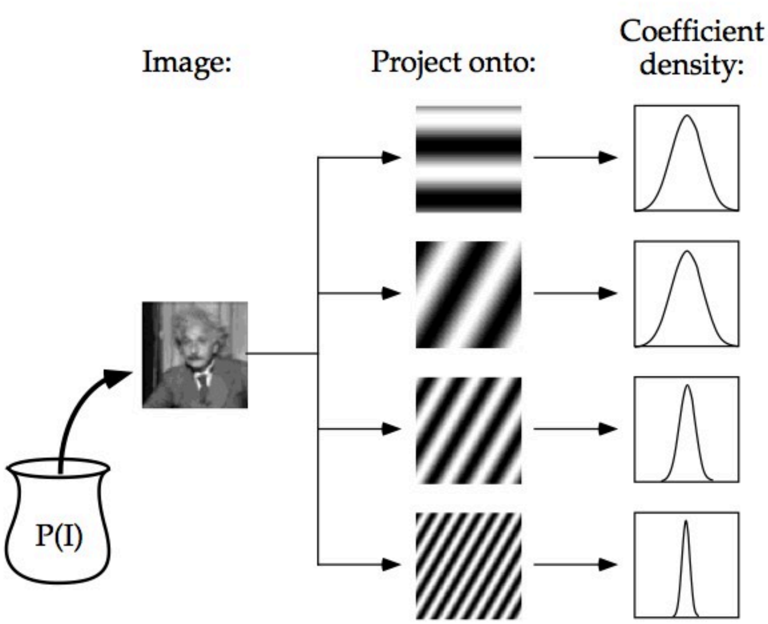
That is, the shape of the Fourier spectrum does not change!

Spectral power



Power law: fourier power falls as $1/f^{\alpha}, \alpha \approx 2$

[Ritterman '52; DeRiugin '56; Field '87; Tohurst '92; Ruderman/Bialek '94]

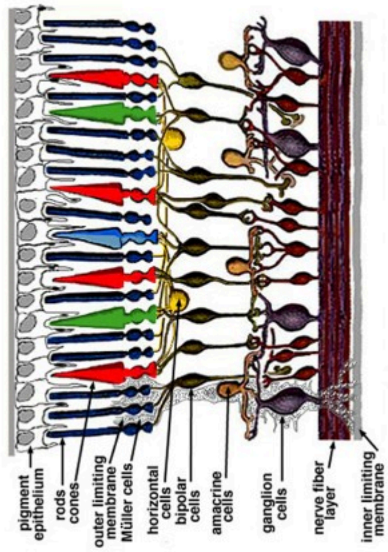


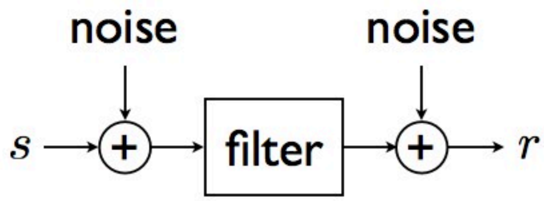
Most image processing engineering is based on this "classic" model

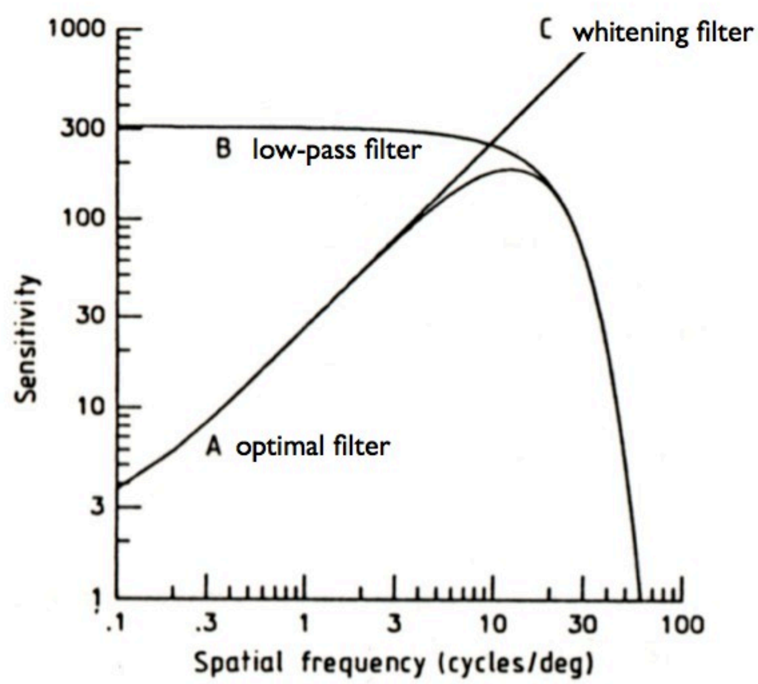
Efficient coding in the retina?

[Srinivasan et. al. 82;
Atick & Redlich 90;
van Hateren 92]

Optimize RF to maximize encoded information information
... over Gaussian image model
... constrained by response variance

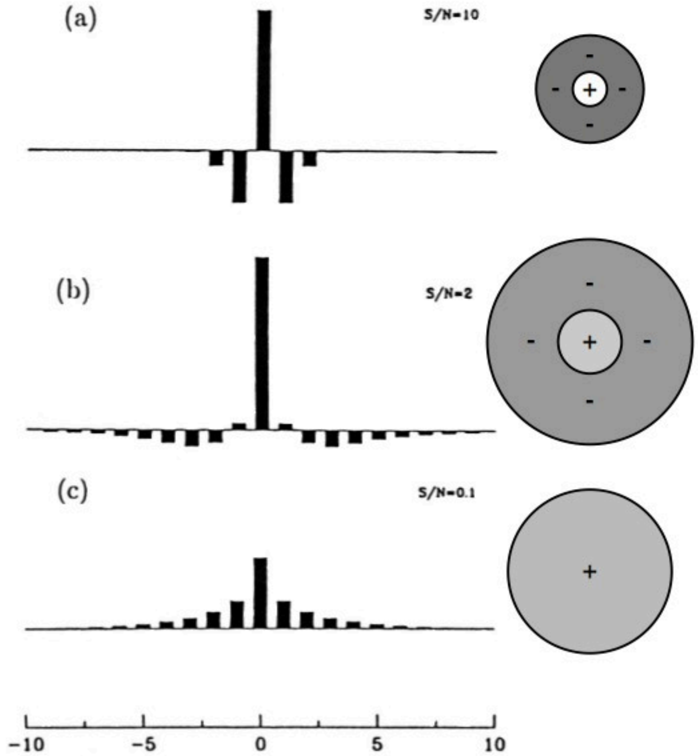


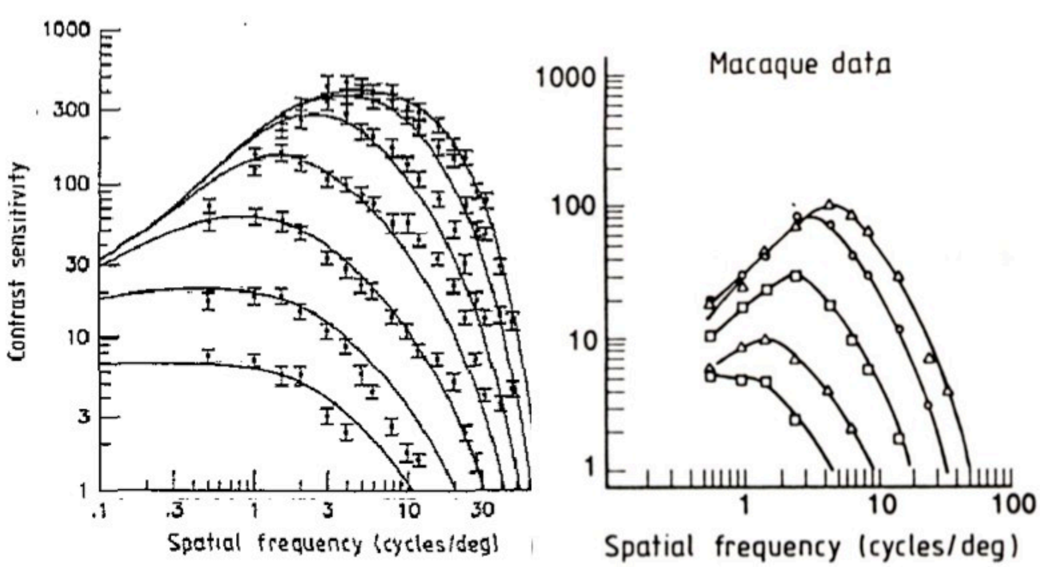




[Atick & Redlich '92]

Optimal filters depend on SNR





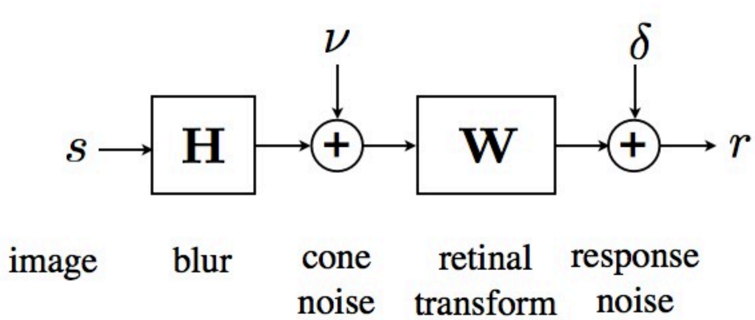
compared with contrast sensitivity functions (human)

[Atick '92]

Limitations

- Signal non-Gaussian
- Noise non-Gaussian, possibly correlated
- Transformations nonlinear
- Cone:RGC ratio is not 1:1
- Cone lattice irregular
- Ganglion cells irregular (shape and sampling)

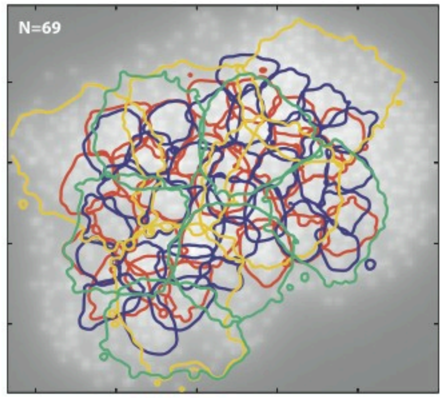
Efficient coding in retina, redux



[Doi et. al., J. Neuro 2012]

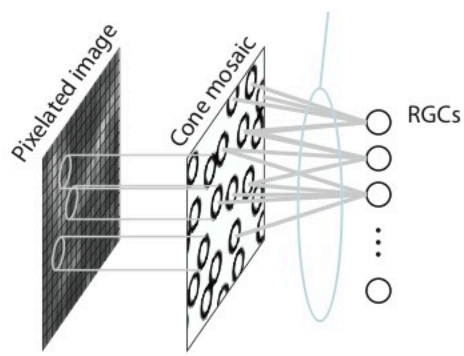
Measured W

- Macaque retina, 27 deg eccentricity
- full mosaics: ON/OFF Midget/ Parasol
- 145 RGCs, 665 cones [ratio = 4.6]
- receptive fields, as weights on cones, determined by STA
- W contains receptive fields



[Gauthier et. al., 09]

Theoretical W



Solution based on:

- 1/f Gaussian model for images
- Human optical blur at 30 deg ecc [Navarro etal 93]
- 10 dB photoreceptor SNR
- 10 dB ganglion cell SNR [Borst & Theunissen 99]

[Doi et. al., J. Neuro 2012]

Information Maximization

$$I(s;r) = \frac{1}{2}\log\frac{|\mathbf{W}\mathbf{H}\mathbf{\Sigma_s}\mathbf{H}^T\mathbf{W}^T + \sigma_\nu^2\mathbf{W}\mathbf{W}^T + \sigma_\delta^2\mathbf{I}_M|}{|\sigma_\nu^2\mathbf{W}\mathbf{W}^T + \sigma_\delta^2\mathbf{I}_M|}$$

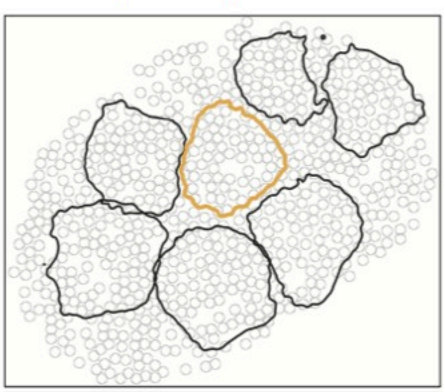
(total power of responses is constrained)

Optimal solution is not unique - there is a whole family of W's that are equally good!

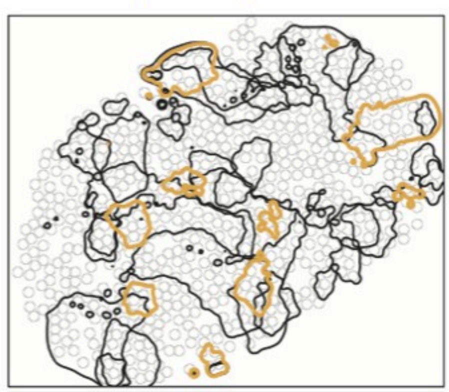
[Doi et. al., J. Neuro 2012]

Optimal RF solution is not unique...

optimally efficient

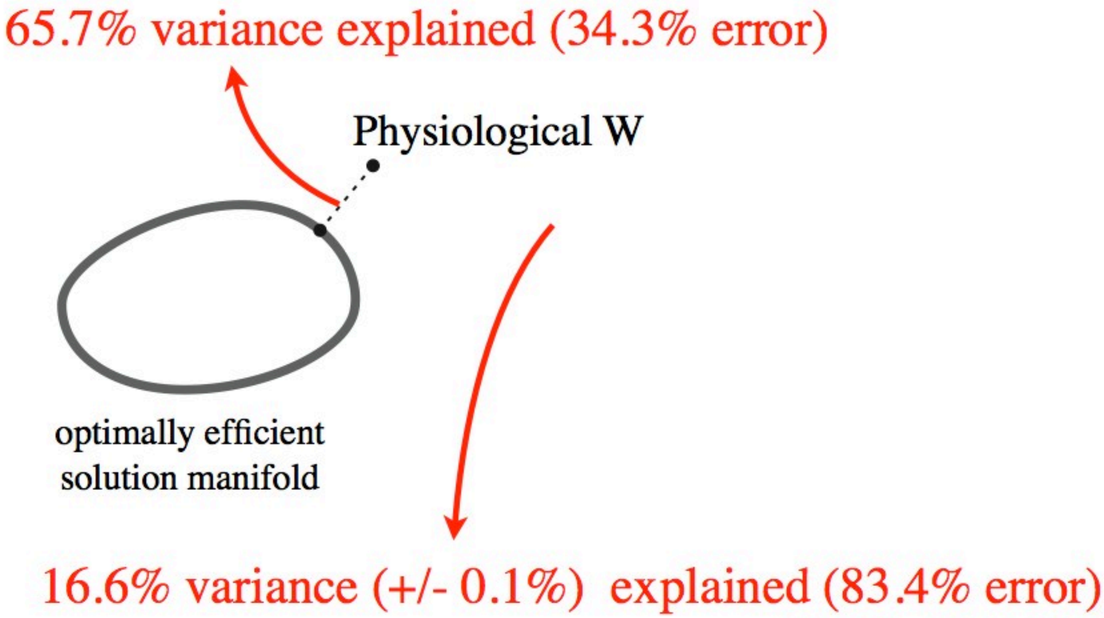


also optimally efficient

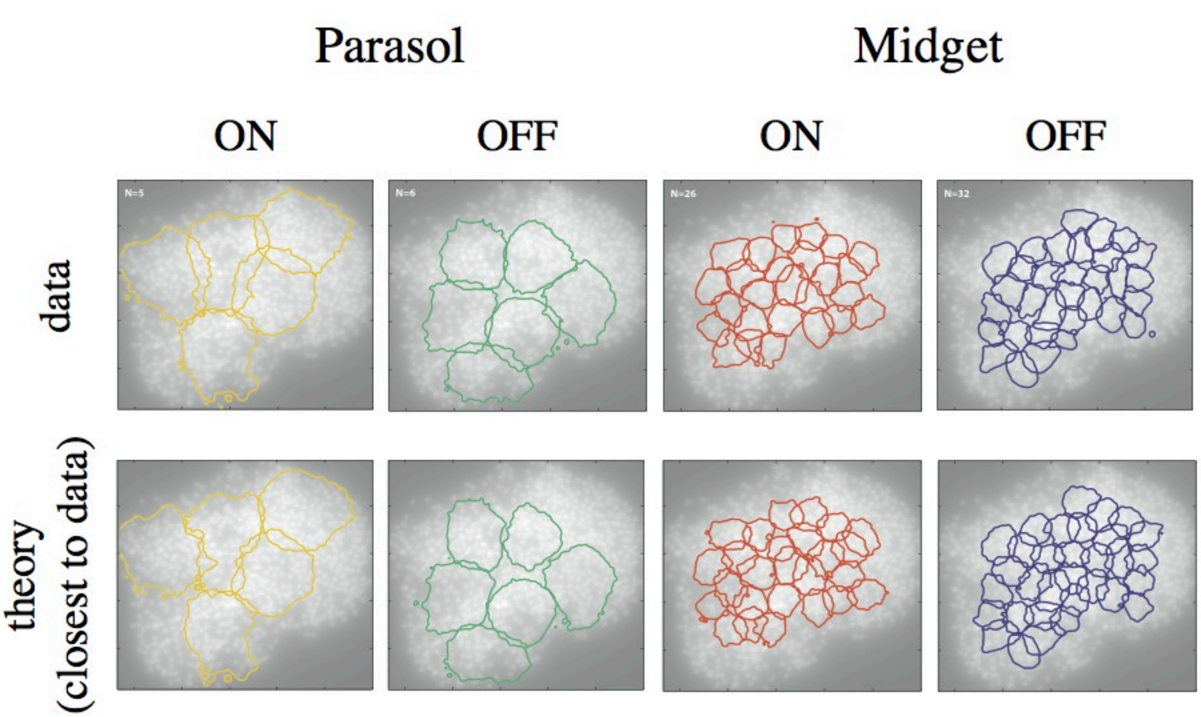


[Doi et. al., 2012]

Comparison to data

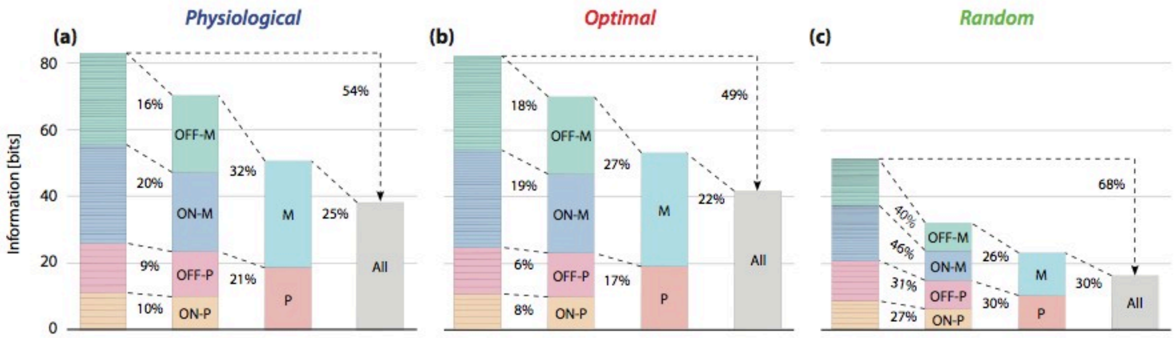


[Doi et. al., J. Neuro 2012]



[Doi et. al., J. Neuro 2012]

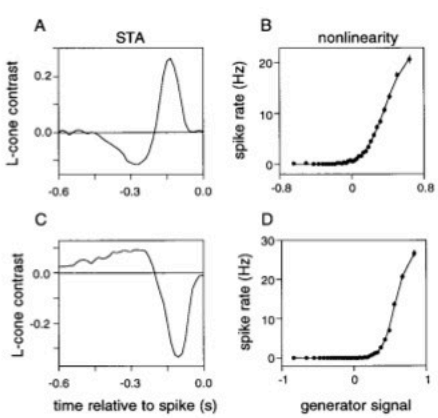
Efficiency & Redundancy



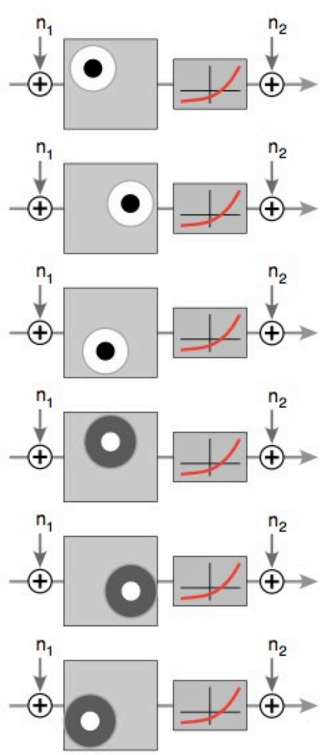
[Doi, et. al., unpublished]

Retinal nonlinearities?

1) Assume a noisy L-N model



[Chander & Chichilnisky, 2001]



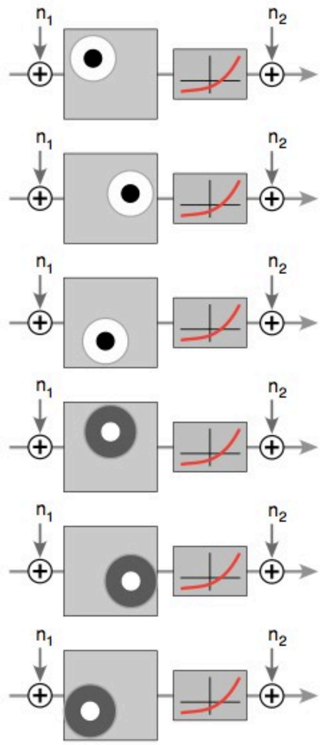
[Karklin & Simoncelli, 2011]

Retinal nonlinearities?

1) Assume a *noisy* LN model

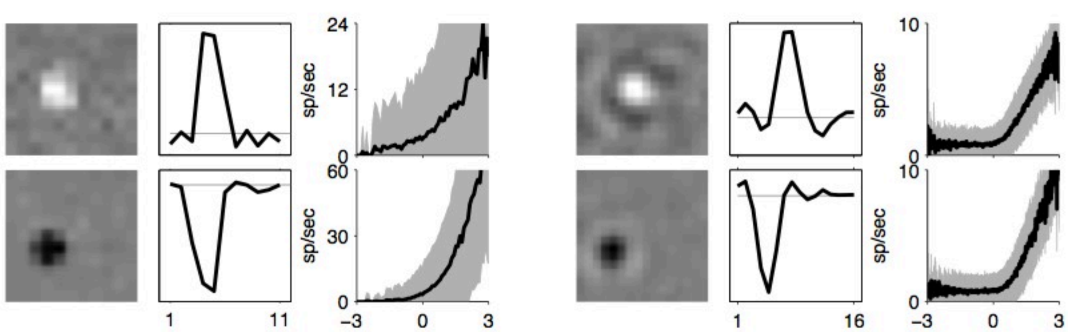
2) Optimize both L and N stages ...

- for information transmission
- subject to constraint on mean response
- over a set of photographic images



[Karklin & Simoncelli, 2011]

Optimal nonlinearities are rectifying, and population naturally separates into ON and OFF

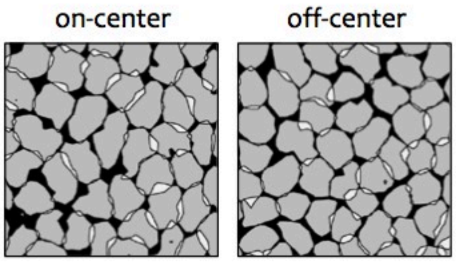


Example On/Off RGCs [Chichilnisky & Kalmar, 02]

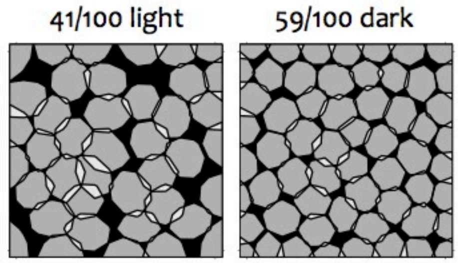
Example optimally efficient On/Off RGCs

[Karklin & Simoncelli, NIPS 2011]

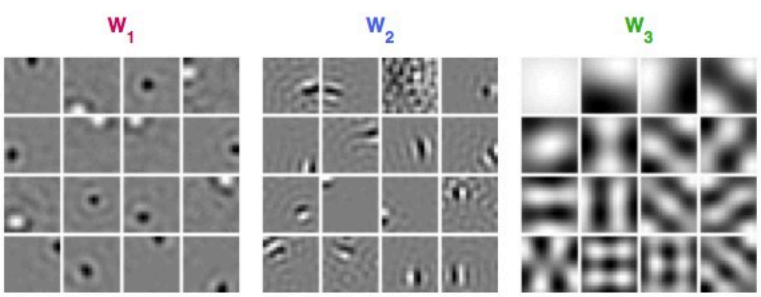
Receptive field populations

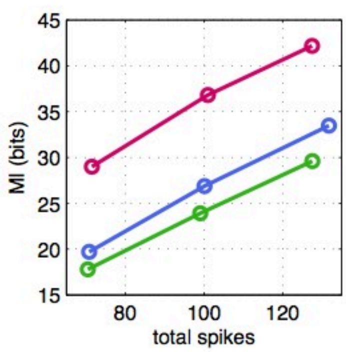


[Gauthier et. al., 2009]



[Karklin & Simoncelli, 2011]





Optimal noisy

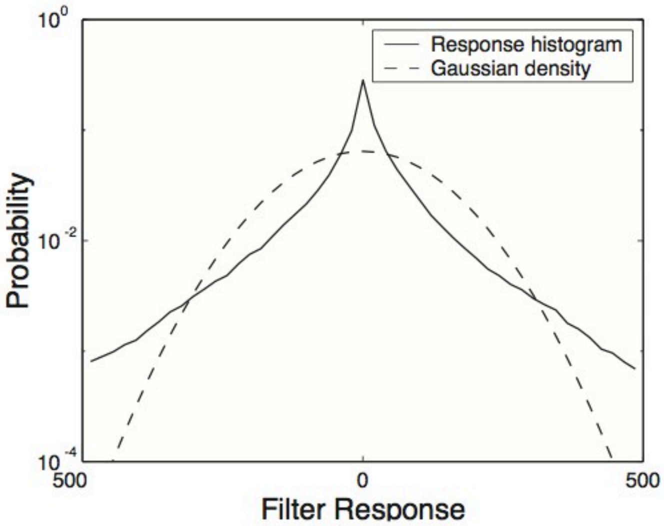
Optimal low-noise

PCA

Solution depends on noise level
Optimal solution transmits significantly more information than optimal low-noise or whitening (PCA) solutions

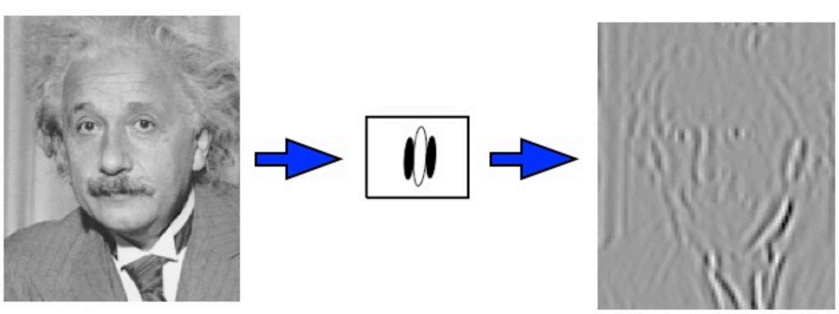
[Karklin & Simoncelli, NIPS 2011]

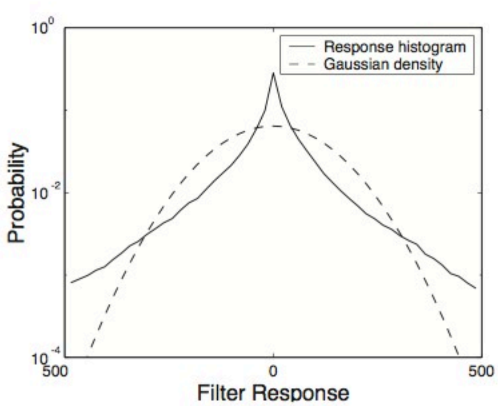
Marginal densities



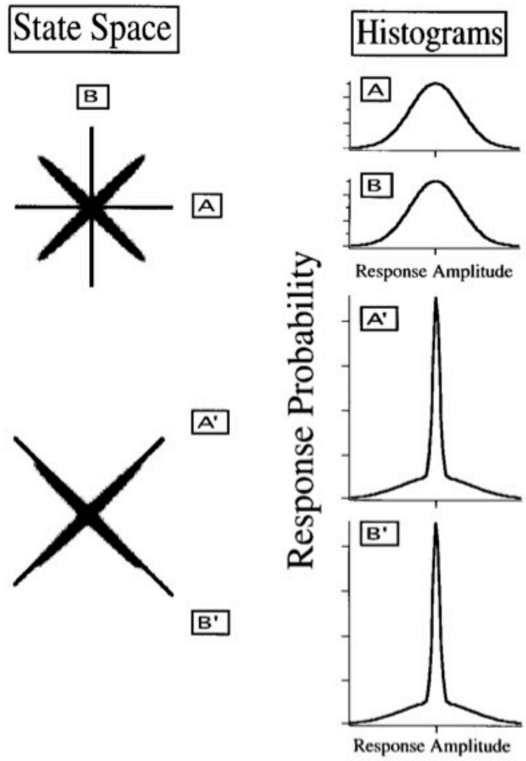
[Burt&Adelson 82; Field 87]

Marginal statistics



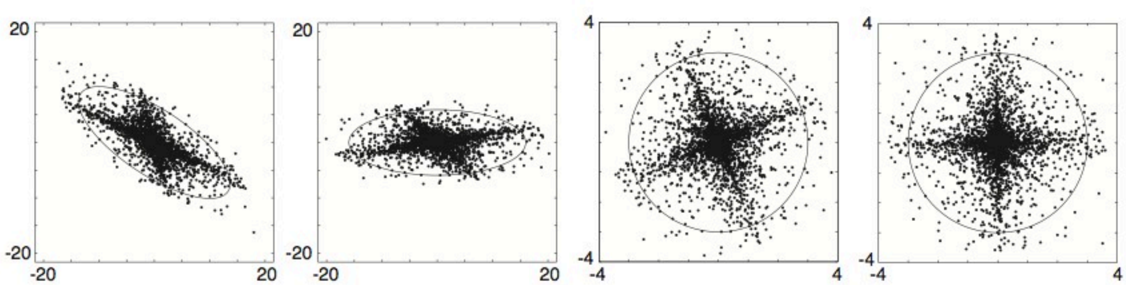


[Burt&Adelson 82; Field 87; Mallat 89]



[Field, “what is the goal of sensory coding”, 1994]

“Independent” Components Analysis (ICA)



For Linearly Transformed Factorial sources:
guaranteed independence
(with some minor caveats)

Independent Component Analysis

Solve for a set of axes (not necessarily orthogonal) along which the data are least Gaussian.

Examples:

- FOBI - simplest algorithm (Cardoso, 1989)
- Fast ICA - fixed-point algorithm with fast convergence (Hyvarinen, 1997)

Closely related: **Projection pursuit**. Seek projections of data that are non-Gaussian (Friedman & Tukey, 1974).

Icassp'89. pp. 2109-2112

SOURCE SEPARATION USING HIGHER ORDER MOMENTS

Jean-Francois Cardoso

Ecole Nat. Sup. des Telecommunications - Dept SIGNAL
46 rue Barrault, 75634 PARIS CEDEX 13, FRANCE.
and
CNRS-URA 820, GRECO-TDSI.

ABSTRACT

This communication presents a simple algebraic method for the extraction of independent components in multidimensional data. Since statistical independence is a much stronger property than uncorrelation, it is possible, using higher-order moments, to identify source signatures in array data without any a-priori model for propagation or reception, that is, without directional vector parametrization, provided that the emitting sources be independent with different probability distributions. We propose such a "blind" identification procedure. Source signatures are directly identified as covariance eigenvectors after data have been orthonormalized and non linearly weighted. Potential applications to Array Processing are illustrated by a simulation consisting in a simultaneous range-bearing estimation with a passive array.

INTRODUCTION

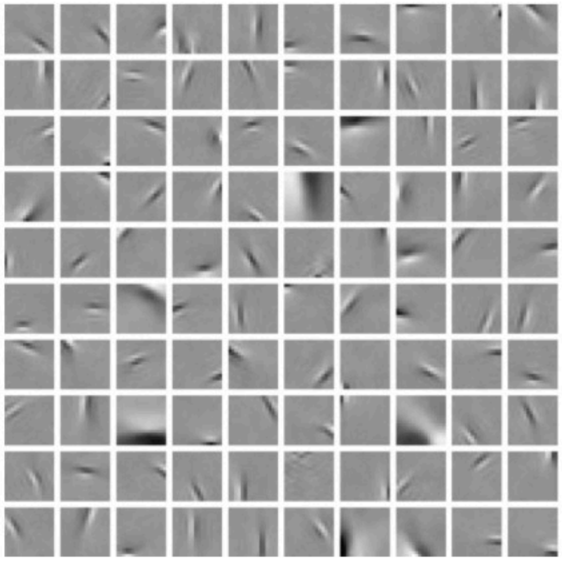
For a lot of reasons (of various kinds), the most common Signal Processing methods deal with second-order statistics, expressed in terms of covariance matrices. It is well known that Gaussian stochastic processes are exhaustively described by their second-order statistics. Nonetheless, when the Gaussian assumption is not valid, some information is lost by retaining only second-order statistics.

in Array Processing has been done within this framework [6,7,8]. However, actual physical settings are often such that source signatures (directional vectors) depart from the assumed model. As expected, model-based methods are very sensitive to such discrepancies. Multipath, unknown antenna deformation are among the common causes of severe performance degradation.

It is the purpose of this communication to present a simple algebraic method allowing source identification when NO a priori information about the propagation and the reception is available. The key requirement is that the observed data consist in a linear superimposition of statistically independent components. It may seem strange that such a blind identification procedure be possible, but it should be recalled that statistical independence between sources is a much stronger requirement than mere uncorrelation. The question of blind separation of multidimensional components by taking advantage of statistical independence has already been adressed in recent litterature. A non-linear adaptive procedure has been proposed in [9,10] while a direct solution using explicitely cumulants was given for the case of two sources and two sensors in [11]. In contrast, we propose here a simple algebraic method to separate an arbitrary number of sources, given measurements from a larger number of sensors.

THE SOURCE SEPARATION PROBLEM

ICA on image blocks



[Bell/Sejnowski '97]
[example obtained with FastICA, Hyvarinen]

Alt: Sparse representation

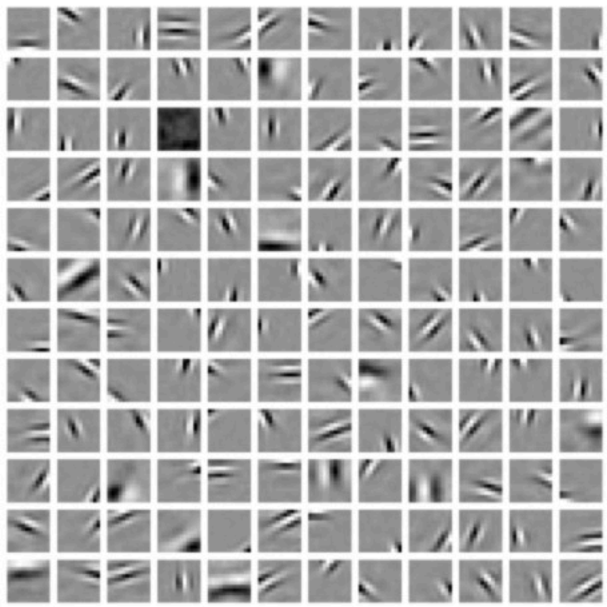
$$E(\vec{c}) = ||\vec{x} - B\vec{c}||^2 + \lambda S_p(\vec{c})$$ [Olshausen & Field '95]

$$S_p(\vec{c}) = \sum |c_k|^p$$

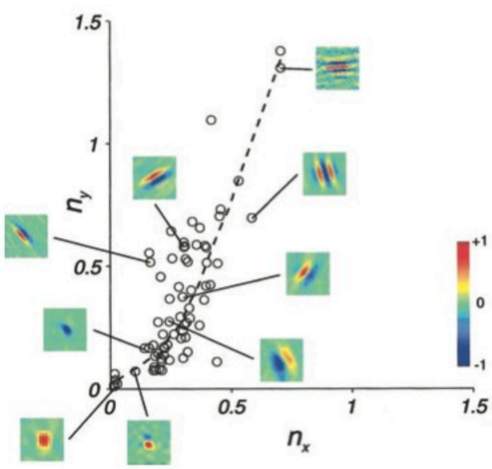
- If p >= 1, the objective function is convex (and thus can solve with descent algorithms)
- The p=1 case is widely used [LASSO - Tibshirani, 1996] [Basis Pursuit - Chen, Donoho, Sanders, 1998]
- Finding efficient solutions, and/or solutions for p<1, has become a major research area

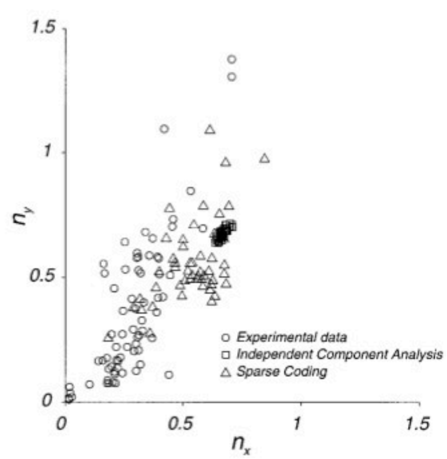
  [e.g., Figueiredo&Nowak 01; Daubechies etal 03; Starck etal 03; Bect etal 04; Elad etal 06; Chartrand 08]

Sparse basis for images

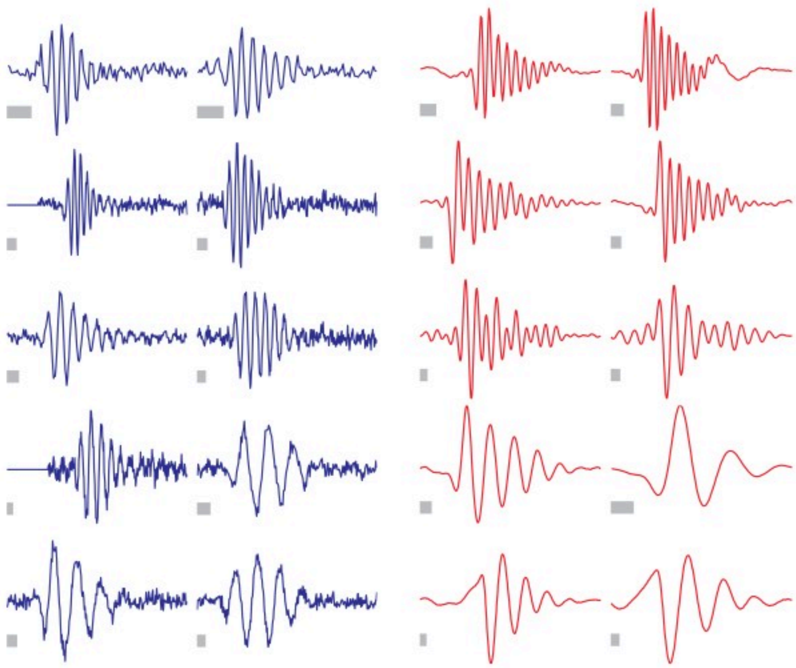


[Olshausen/Field ’96]





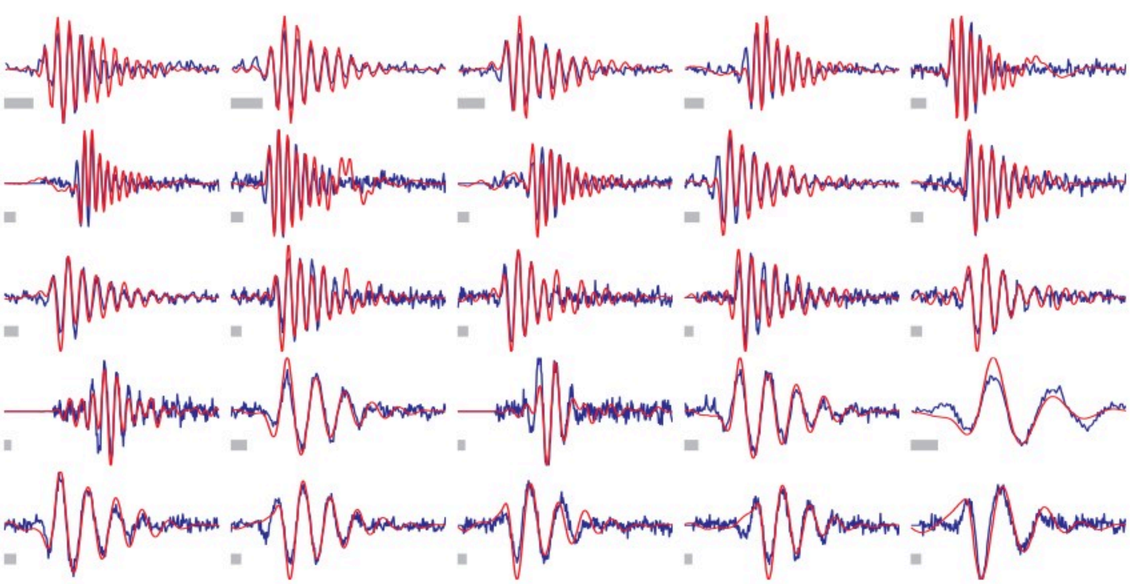
[Ringach 2002]



Auditory nerve filters
from Carney, McDuffy, and Shekhter, 1999

Optimized kernels
scale bar = 1 msec

- Smith & Lewicki, 2006

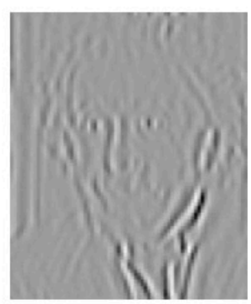


For most filters, there's a matching auditory nerve fiber!

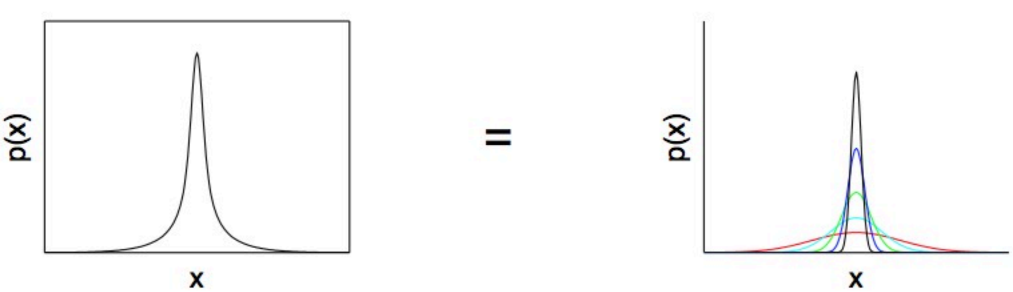
Refinements

- Overcompleteness
  [Olshausen/Field '97; Lewicki&Sejnowski '00]
- Complex cells
  [Berkes&Wiskott '02; Hyvarinen&Hoyer '01, etc.]
- Nonlinearities
  [Rao&Ballard '99; Schwartz&Simoncelli '01]

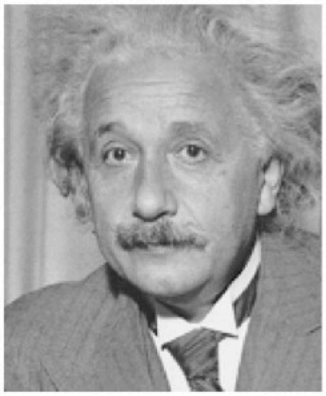
Instead of focus on marginal distribution, note that subbands are *heteroskedastic* (they have variable local variance):

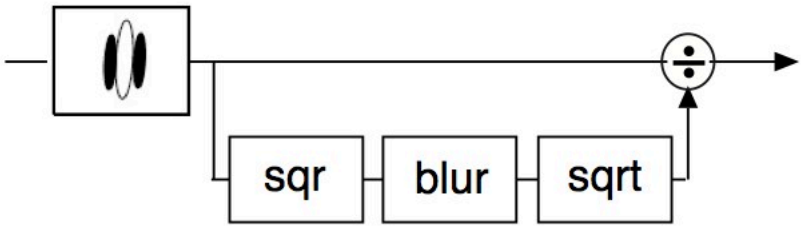


We can model this behavior using a Gaussian scale mixture (GSM):
[Wainwright & Simoncelli 2000]



Estimate local stdev, and then *divide*

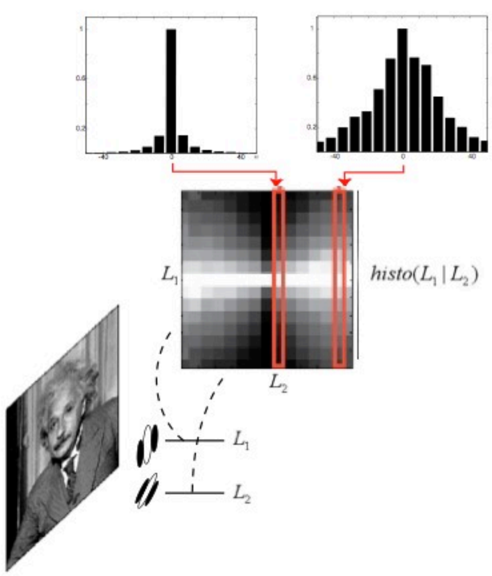




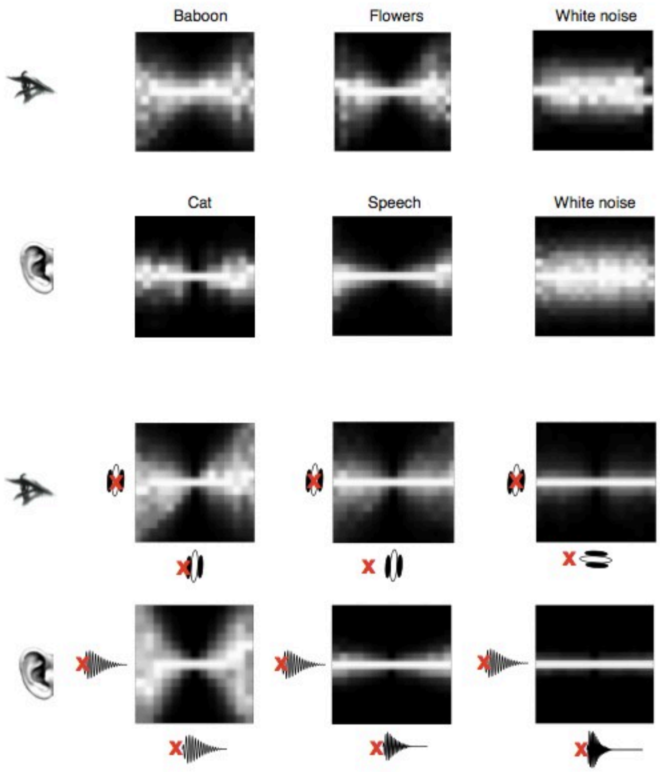
Output distribution is approximately Gaussian!

[eg, Ruderman & Bialek ’94; Wainwright & Simoncelli ’00; Fairhall et al ‘01]

Conditional densities



[Simoncelli ’97; Schwartz&Simoncelli ‘01]

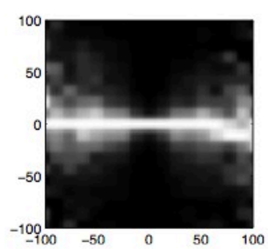


[Schwartz & Simoncelli, 2001]

Modeling the Dependency

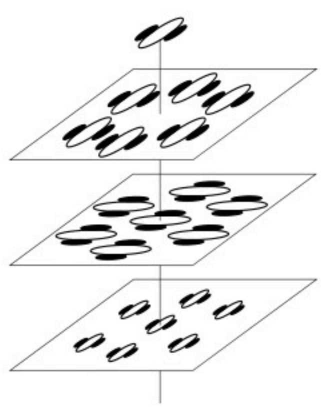
One filter:

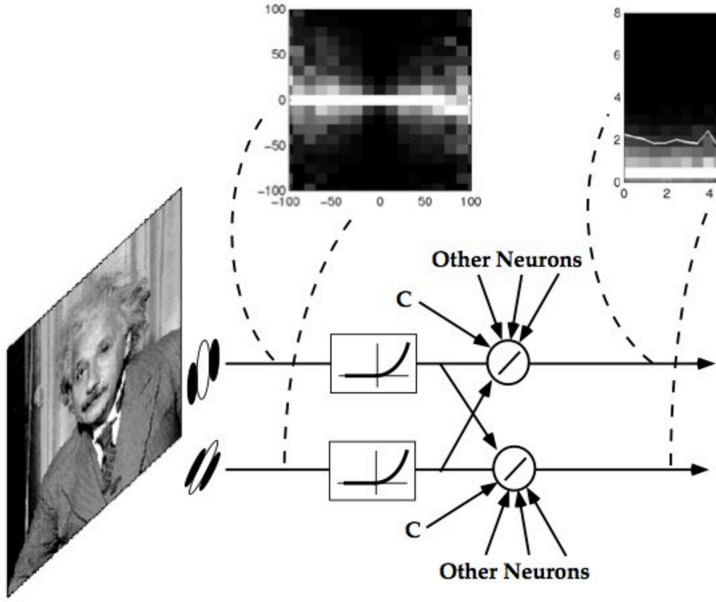
$$\mathrm{var}(L_1|L_2) = w\ L_2^2 + \sigma^2$$



Generalized neighborhood:

$$\mathrm{var}(L_1|\{L_n\}) = \Sigma_n w_n\ {L_n}^2 + \sigma^2$$





Divisive normalization reduces dependency

[Schwartz & Simoncelli, '01]

Divisive Normalization: Physiological Evidence

Steady-state neural responses = linear projection, rectification, and division by the summed responses of other neurons [Heeger '92; Carandini/Heeger/Movshon '97]

Such models can account for some nonlinear striate cortical behaviors. Examples [Carandini et al. 1997]:

- Tuning curves independent of contrast
- Contrast saturation level depends on stimulus parameters
- Cross-orientation suppression
- Increasing phase lag at lower contrast

Methods

- Statistically-determined model:
  1. Linear basis: multi-scale, oriented 3rd derivative operators
  2. "Neuron": vertical, optimal spatial frequency 0.125 cycle/pixel
  3. Neighborhood: 2 scales, 4 orientations, 3×3 array
  4. Weights: optimized (ML) for statistics of 10 images (faces, landscapes, and animals).
- Physiological simulations:
  1. Compute linear responses of full neighborhood
  2. Square
  3. Divide chosen neuron response by weighted combination of squared neighbor responses.

Parameter Optimization

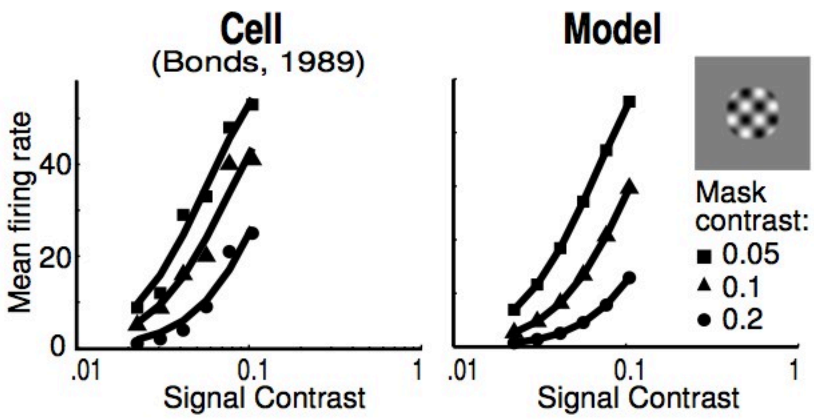
Assume a Gaussian form for the conditional distribution:

$$\mathcal{P}(L_n \mid \{L_k\}) \;\sim\; \mathcal{N}\left(0; \sum_k w_{nk}\,|L_k|^2 + \sigma^2\right)$$

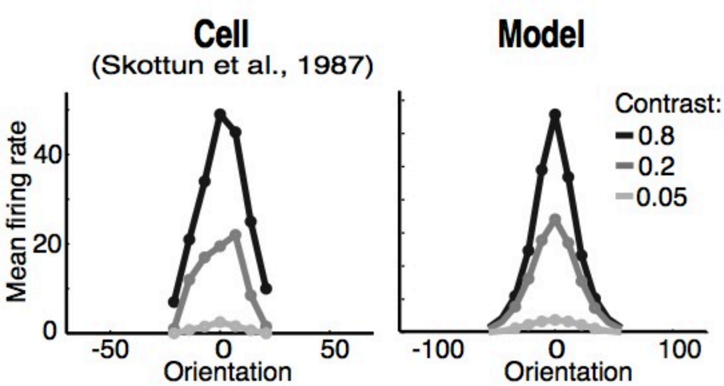
Maximize the likelihood over the image data:

$$\hat{w}_{nk}, \hat{\sigma} = \arg\max_{w_{nk},\sigma} \prod_i \frac{1}{\sqrt{2\pi \sum_k \, w_{nk}\,|L_k|^2 + \sigma^2}} \exp\left[\frac{-{L^2}_n}{2\sum_k \, w_{nk}\,|L_k|^2 + \sigma^2}\right]$$

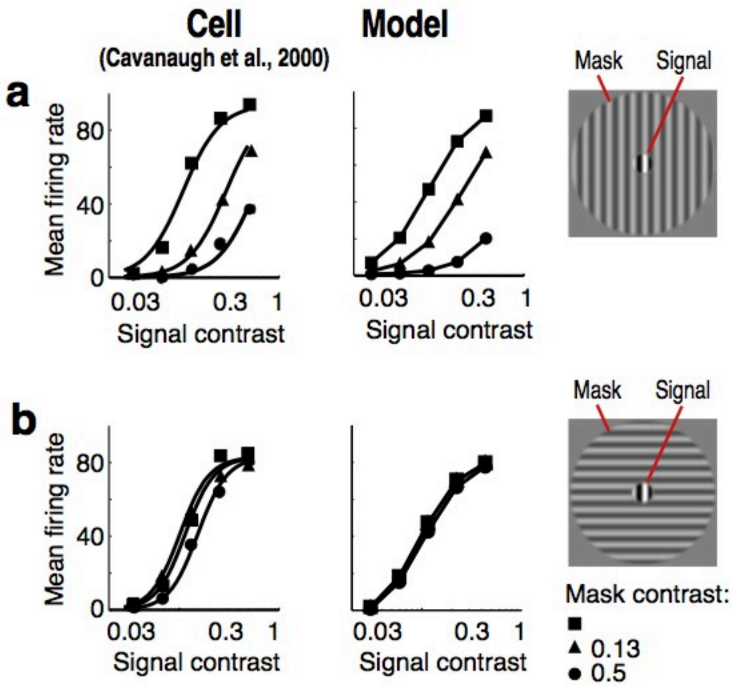
Cross-orientation Suppression



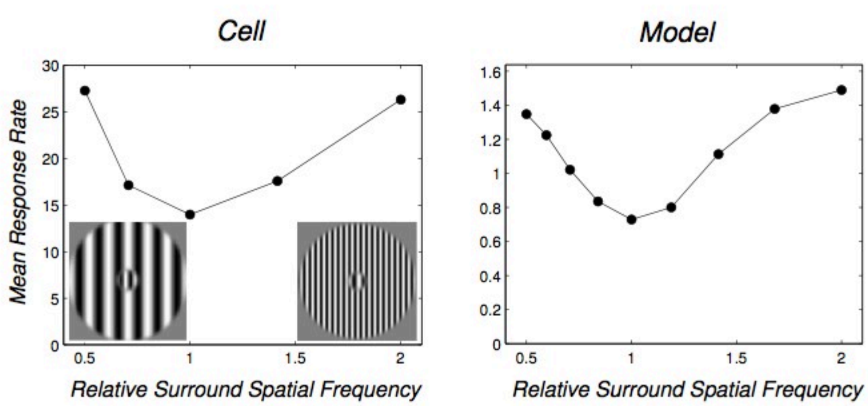
Tuning Curves Independent of Contrast



Surround Suppression



Surround Spatial Frequency



[Data: Müller, Krauskopf, & Lennie.]

Stimulus Diameter