Scene Statistics
Part 2

Odelia Schwartz
University of Miami
Summer 2016

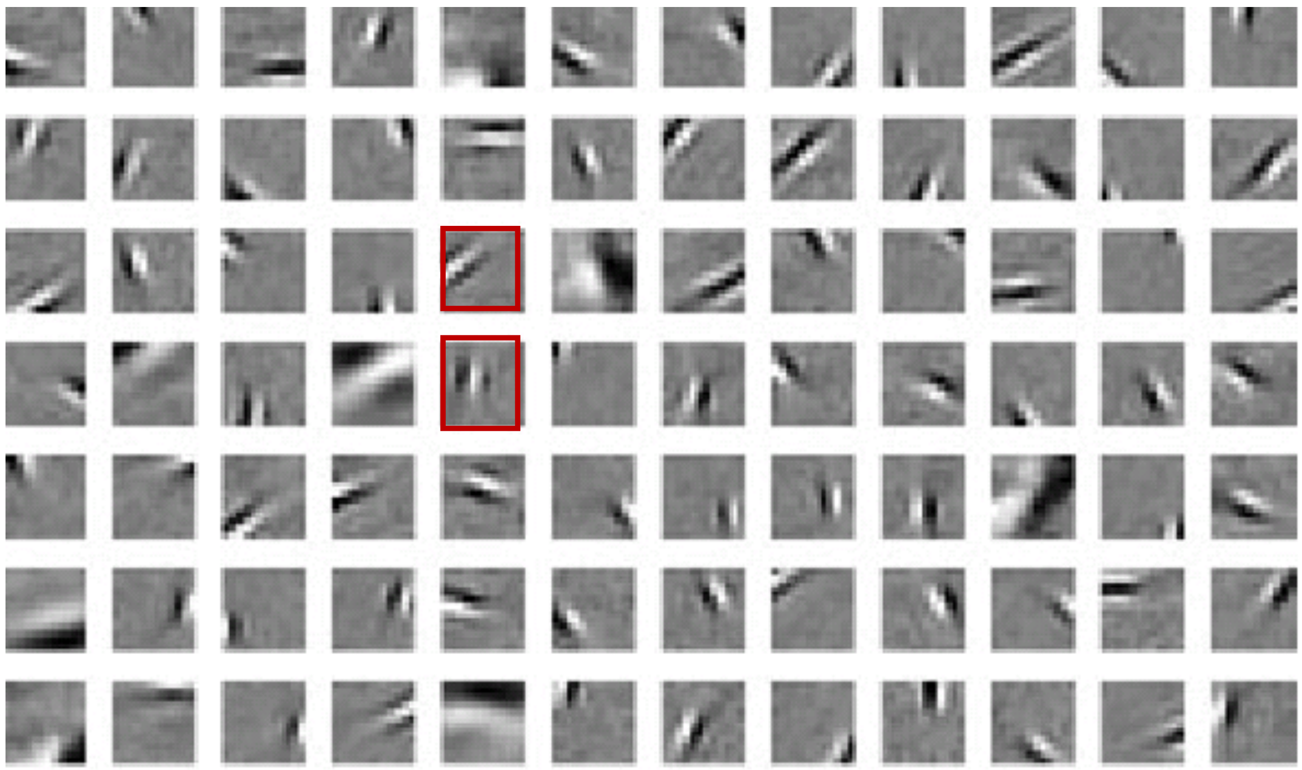
Summary

- We've considered bottom-up scene statistics, efficient coding, and relation of linear transforms to visual filters

- This class: nonlinearities

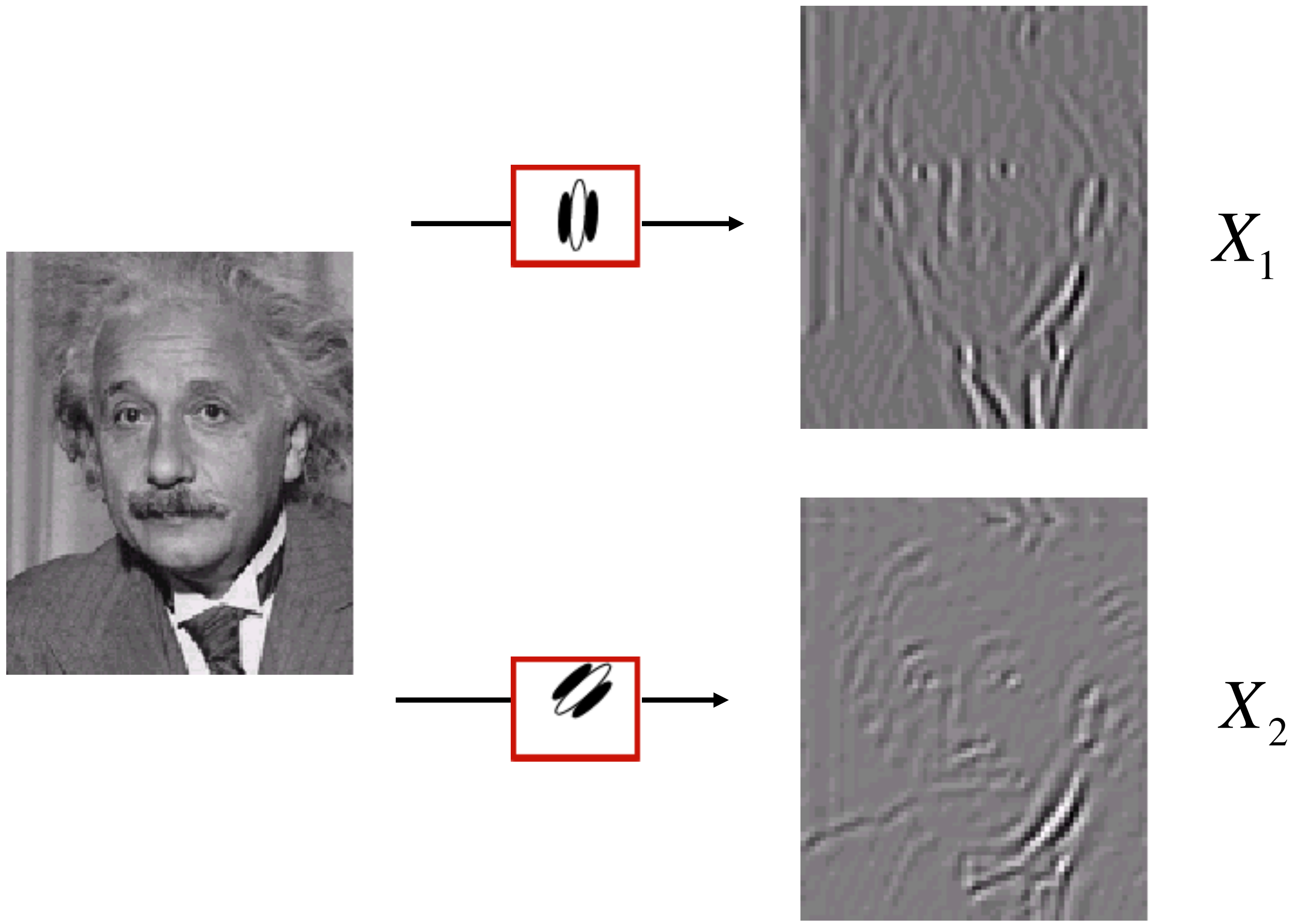
- This class: generative, top-down, perspective

Beyond linear

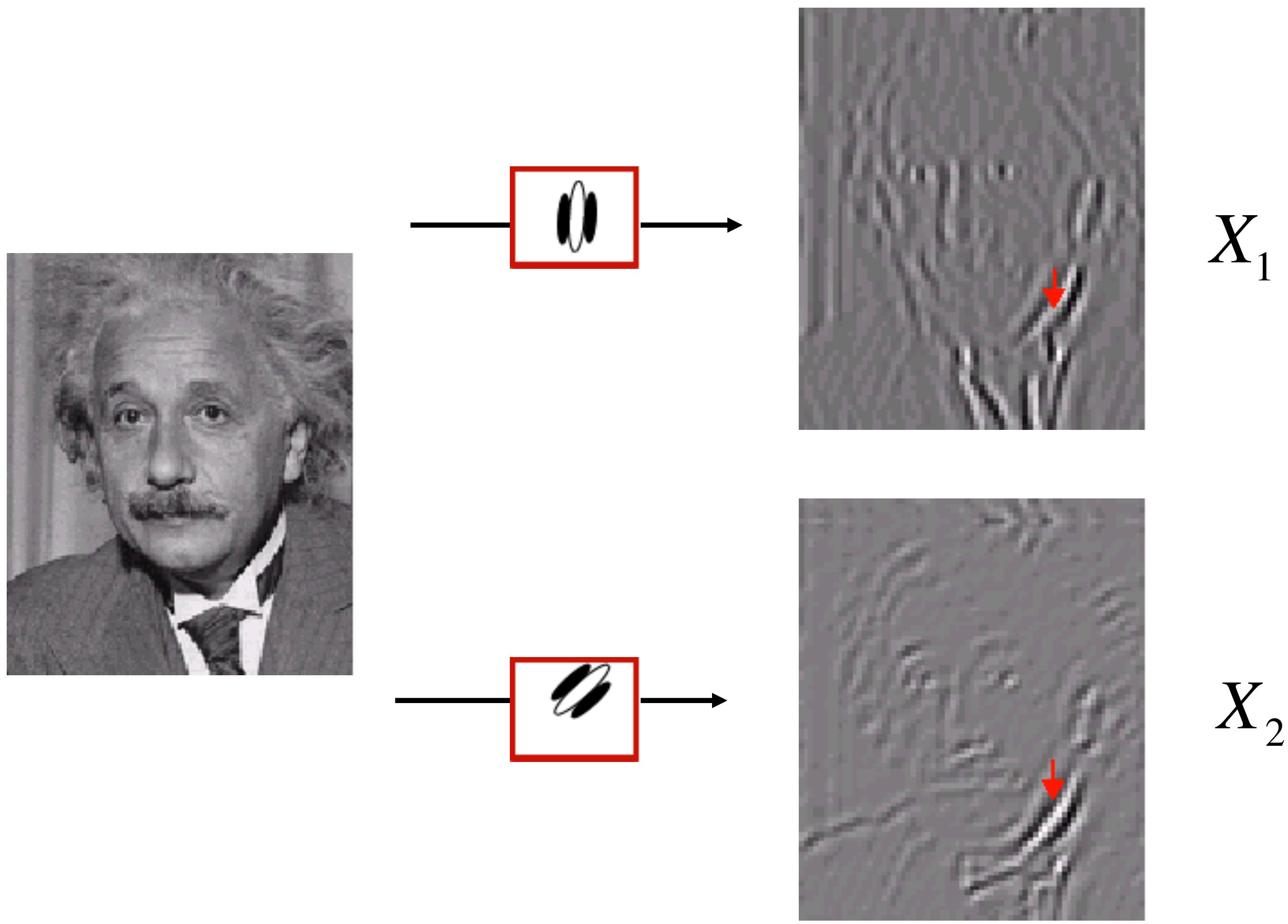


- Filter responses as independent as possible assuming a linear transform
- But are they independent?

Bottom-up Joint Statistics

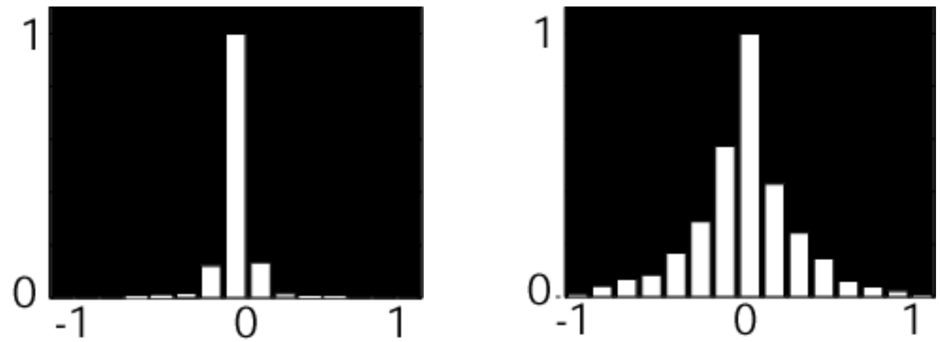


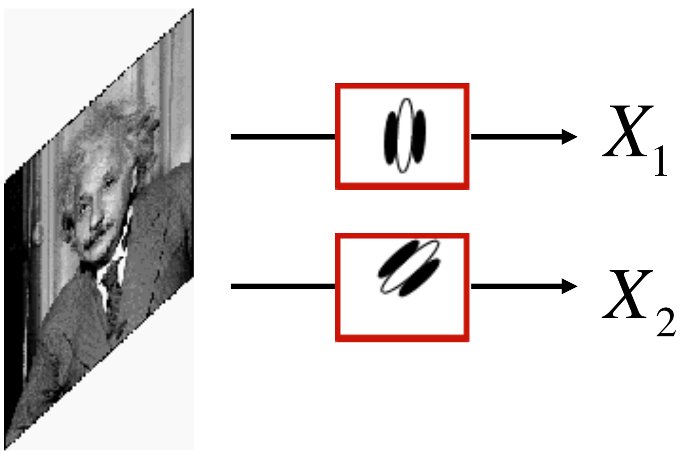
Bottom-up Joint Statistics



Bottom-up Joint Statistics

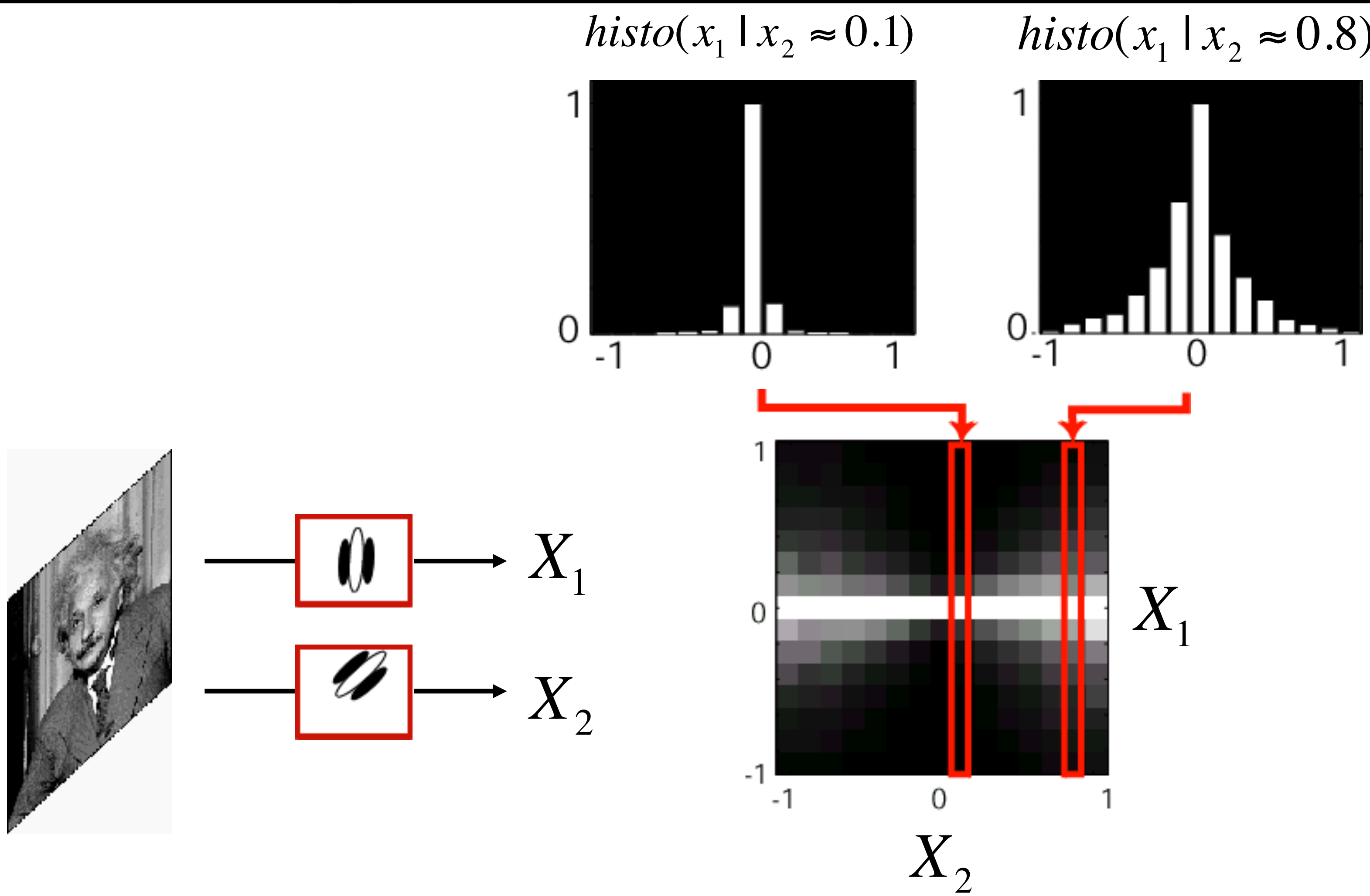
$histo(X_1 \mid X_2 \approx 0.1)$ $histo(X_1 \mid X_2 \approx 0.8)$





Are X_1 and X_2 statistically independent?

Bottom-up Joint Statistics

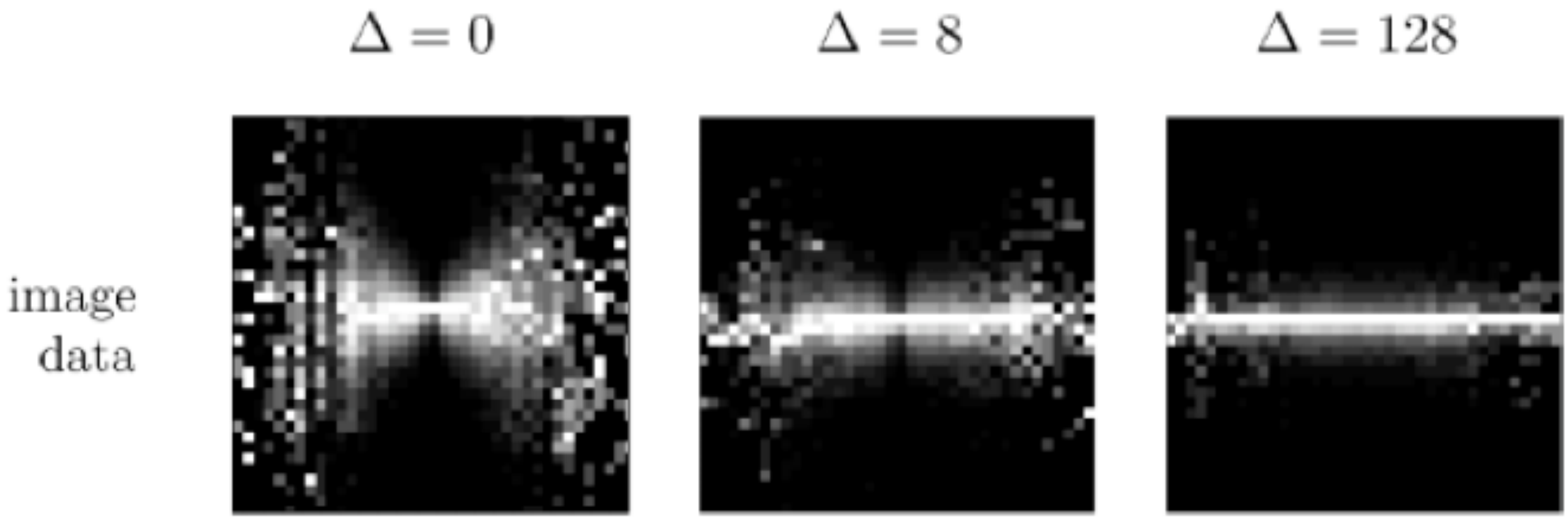


X_1 and X_2 are **not** statistically independent

Schwartz and Simoncelli, 2001

Bottom-up (contour plots)

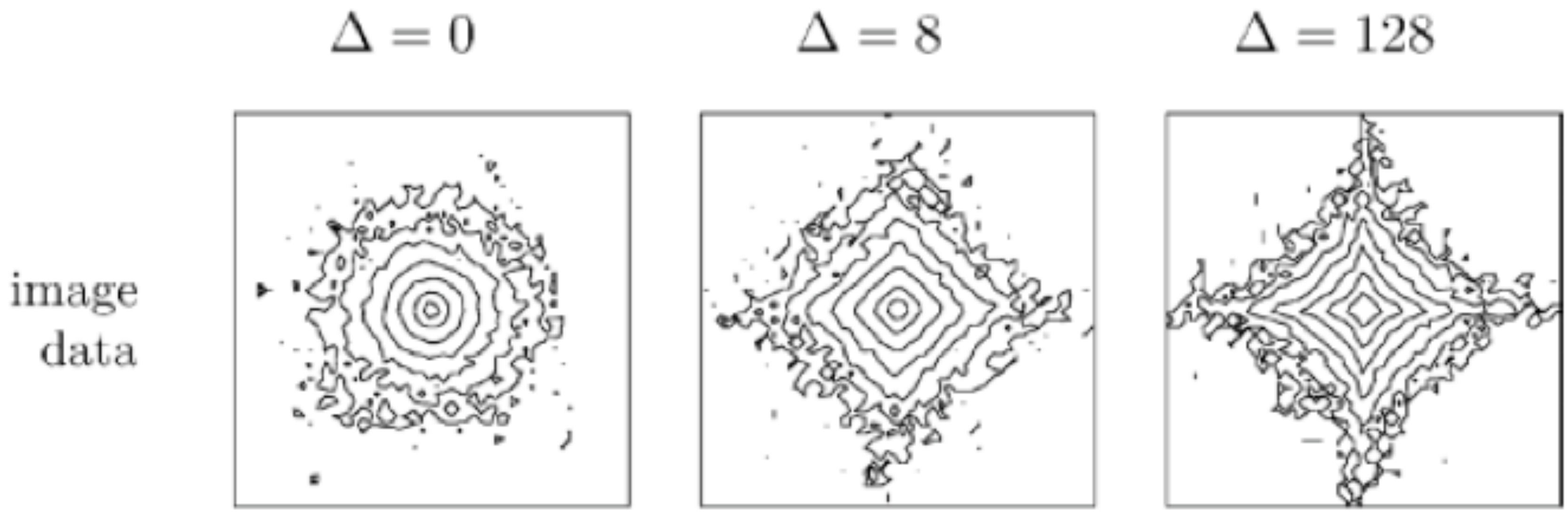
Image patch and different filter pairs (even and odd Phase at different distances) …



Wainwright et al. 2001

Bottom-up (contour plots)

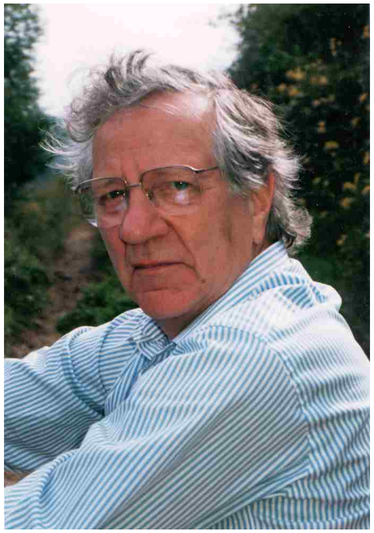
Image patch and different filter pairs (even and odd Phase at different distances) …



Wainwright et al. 2001

Theoretical Approaches

• Sensory systems aim to form an *efficient code* by reducing the redundancies and statistical dependencies of the input; influenced by information theory in the 1950s



Barlow (also Attneave)

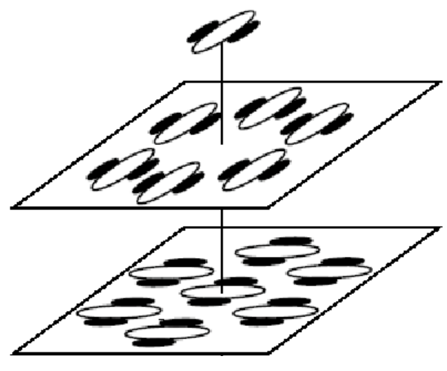
Theoretical Approaches

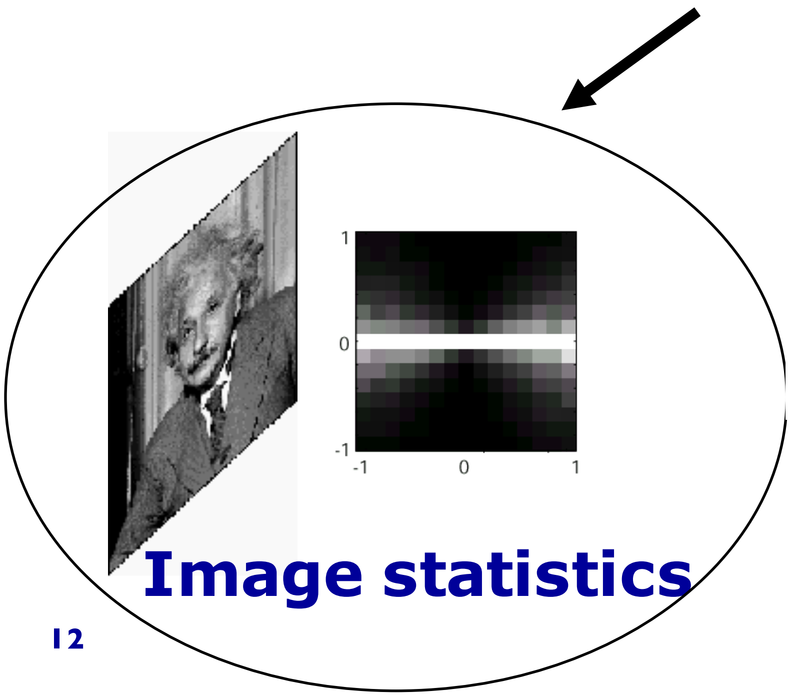
• Sensory processing as inference of properties of the input (can be formalized via probabilistic *Bayesian inference*)

Helmholtz

Bayes

Generative Model (nonlinear)

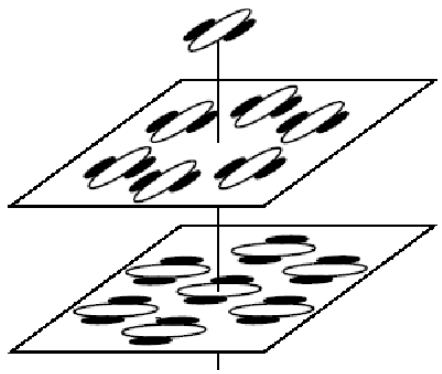




If we have a front end of filters, how do we statistically model their coordination?

Generative Model (nonlinear)

Modeling filter coordination in images

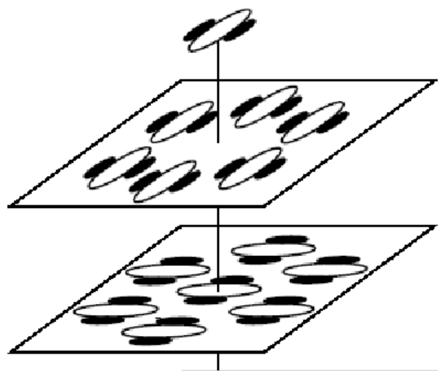


- Learning how more complex representations build up from the structure of images

- Understanding spatial contextual effects in visual processing and how they relate to the coordinated filter structure in images

Generative Model (nonlinear)

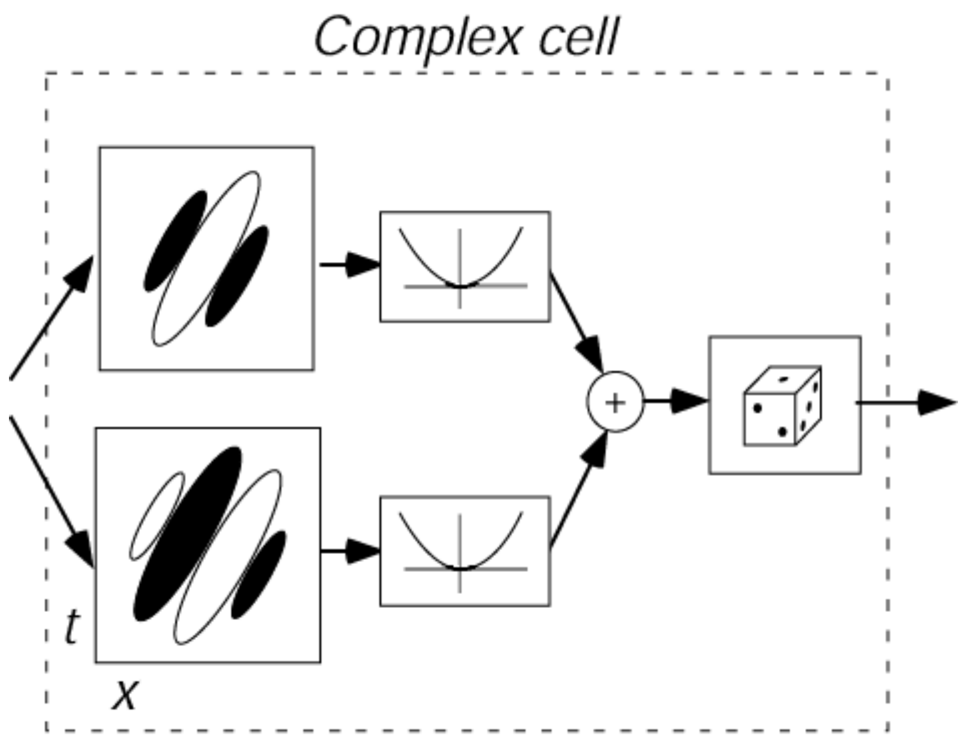
Modeling filter coordination in images



What kind of complex representations?

More complex representations

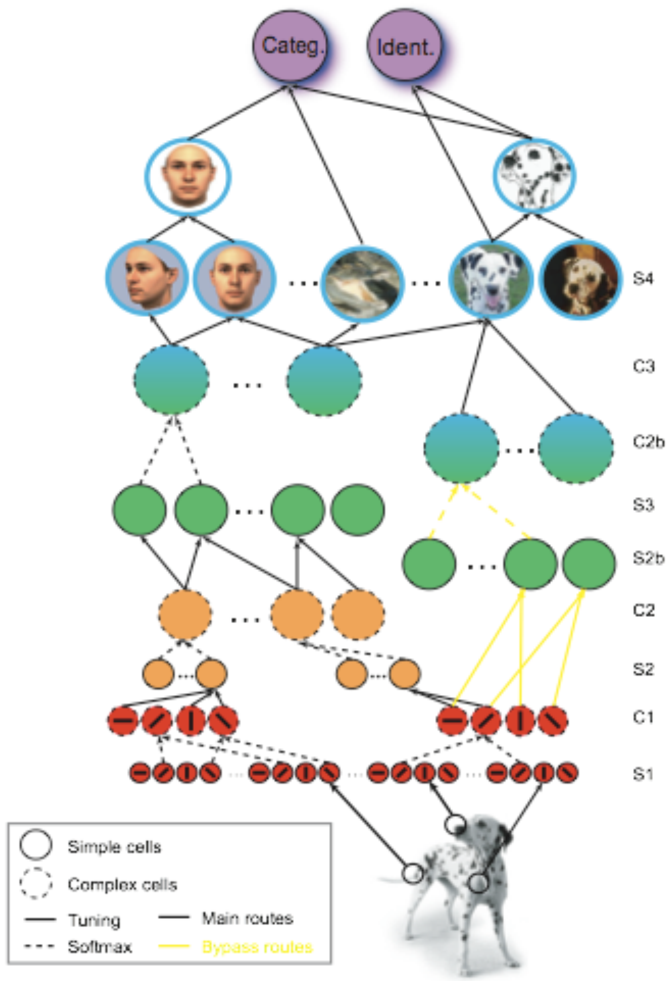
Descriptive models (capturing an invariance)



Adelson & Bergen (1985)

More complex representations

Descriptive models (selectivity and invariance)

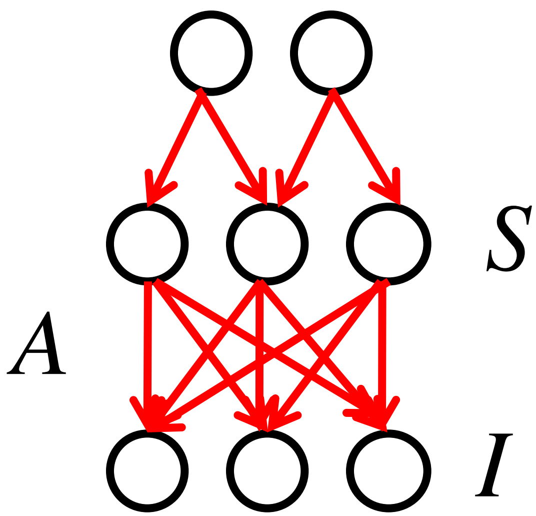


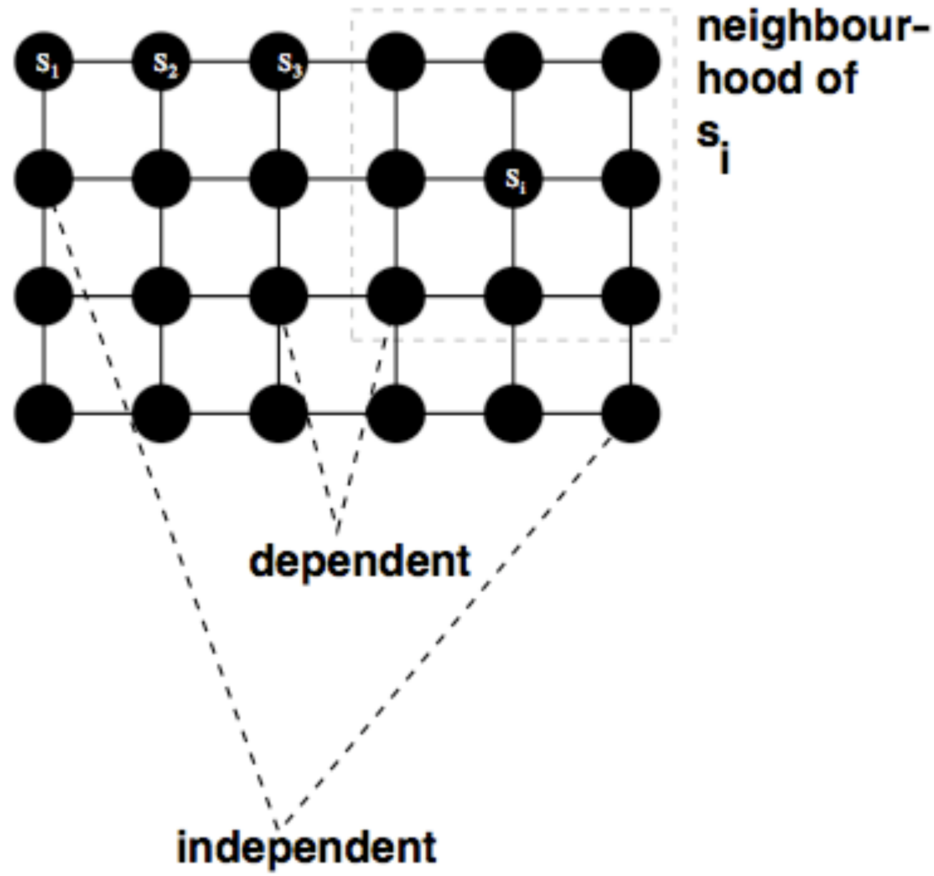
Reisenhuber and Poggio

More complex representations

What about learning from natural scenes?

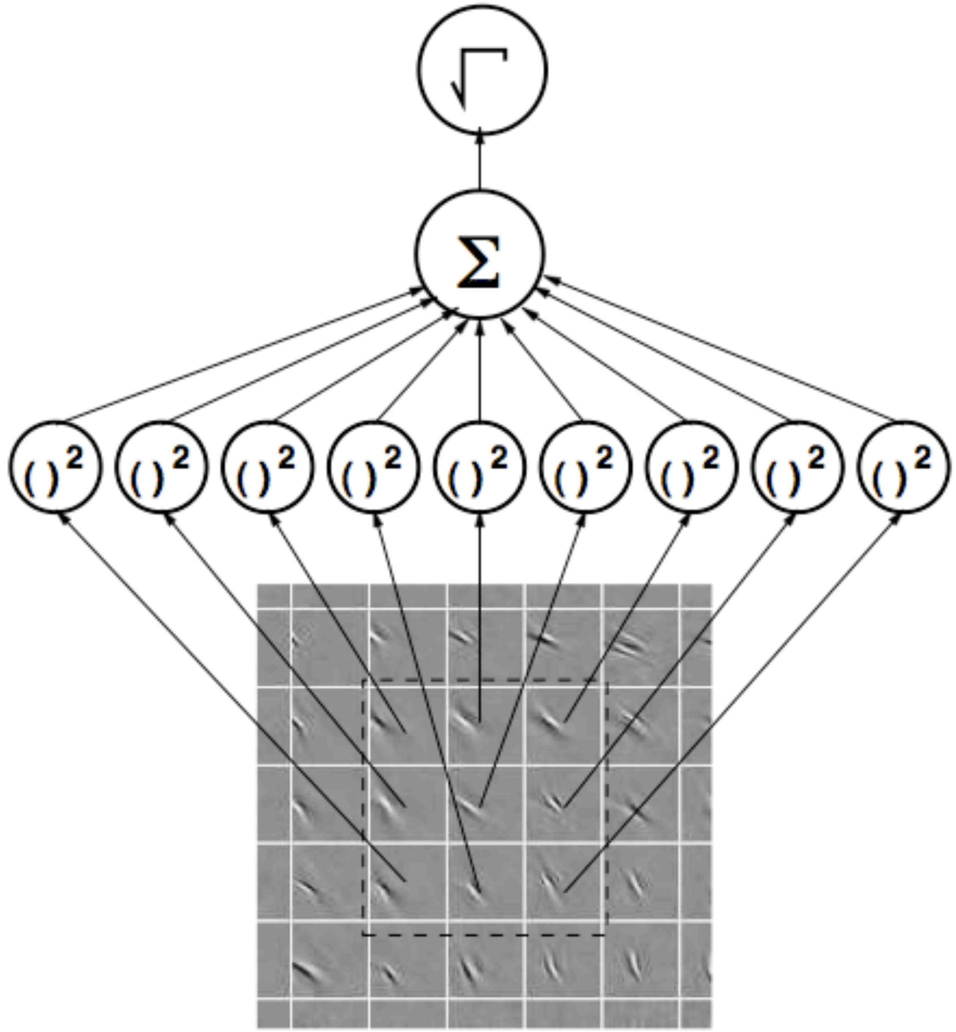
Extensions to ICA





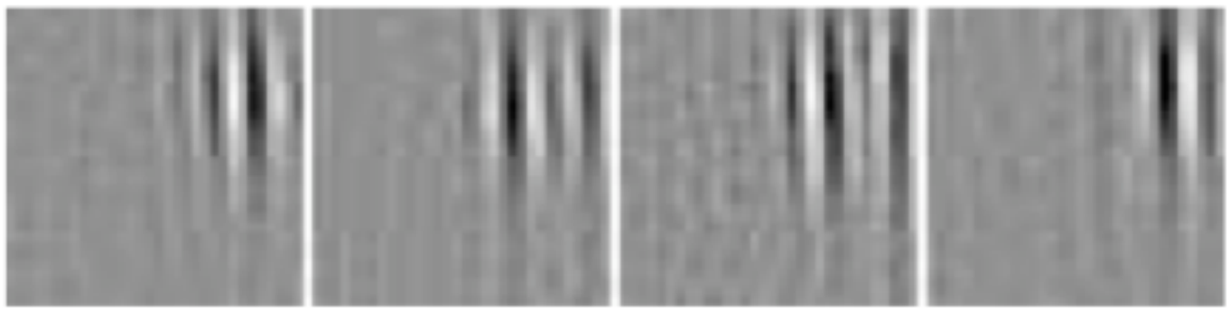
- from Hyvarinen and Hoyer: relax independence assumption; nearby hidden variables S no longer independent; but different neighborhoods independent of one another…

Extensions to ICA



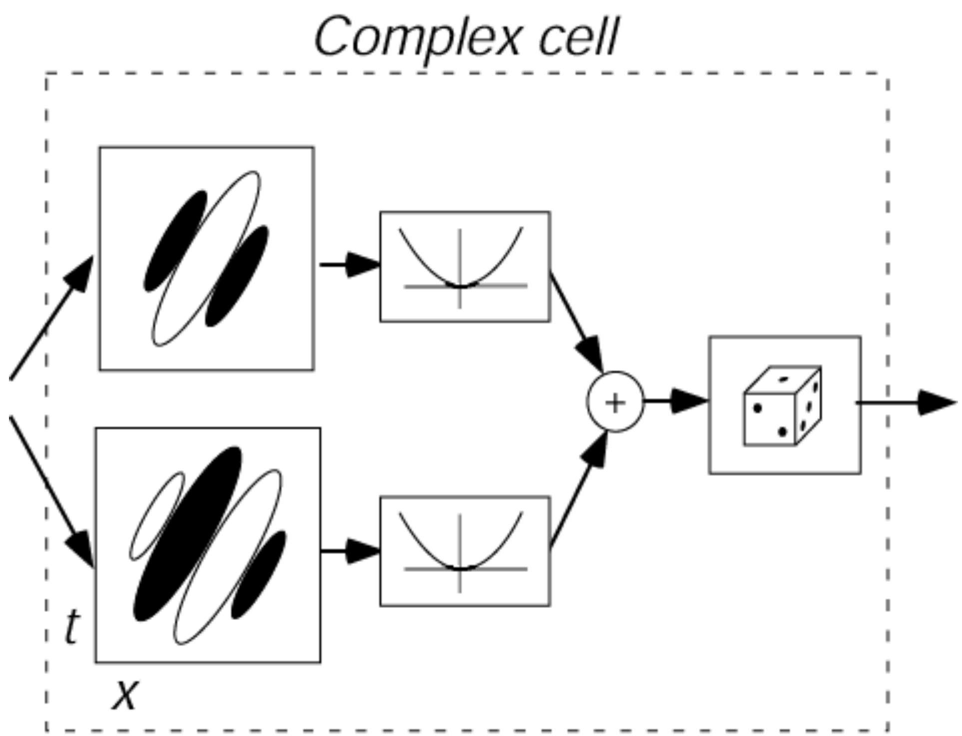
- Hyvarinen and Hoyer, 2001: Invariant features (complex cell outputs) from summing squares of linear (simple cell outputs)

Extensions to ICA



- Hyvarinen book: shown smaller group of dependent filters

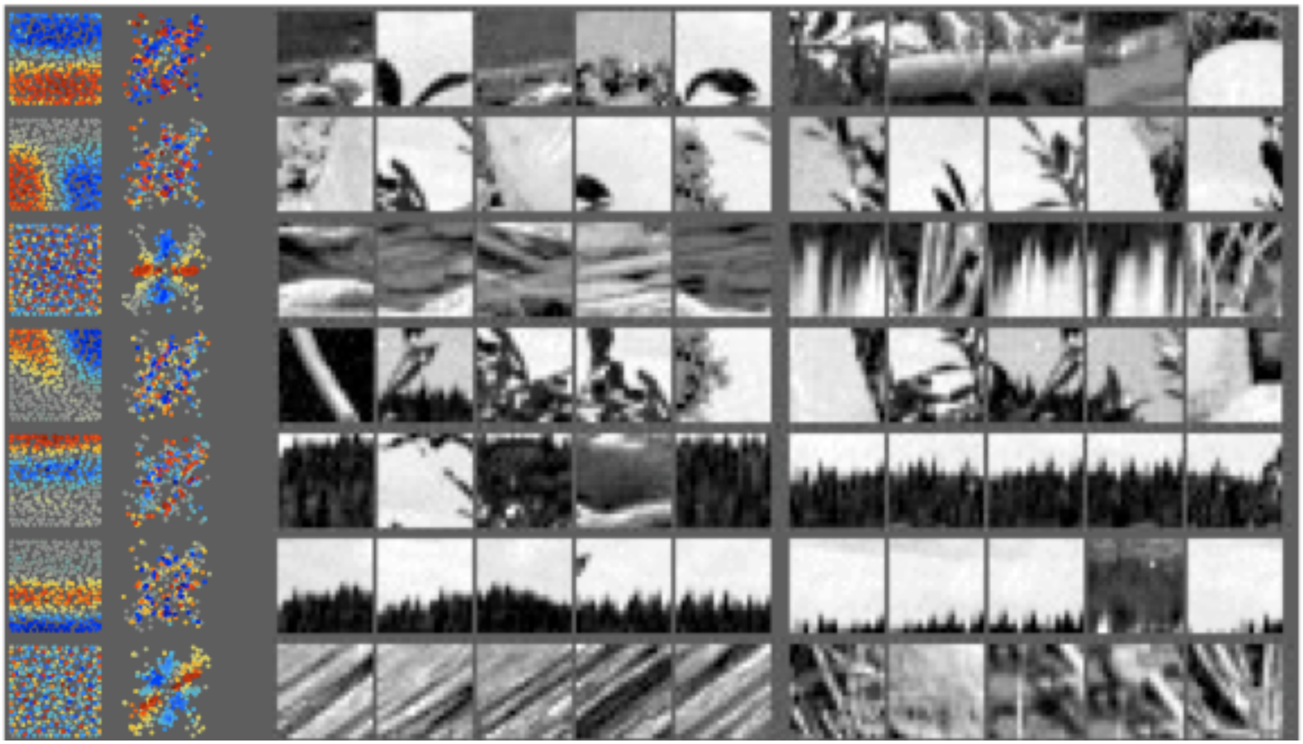
Complex cell



Adelson & Bergen (1985)

Relates to complex cells and invariances…

Hierarchical ICA



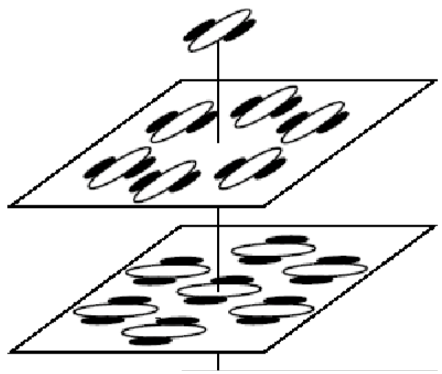
Karklin & Lewicki, 2003; 2005: variance of each unit in the first (ICA) layer arises from an additive combination of a set of basis functions in a higher layer (more complex patterns of dependencies)

Hierarchical representations

We will look at recent advances in learning hierarchical structure more later in the context of deep learning

Generative Model (nonlinear)

Modeling filter coordination in images



- Learning how more complex representations build up from the structure of images

- Understanding spatial contextual effects in visual processing and how they relate to the structure of images – will connect to neural nonlinearities and specifically divisive normalization

Motivation

- Spatial context plays critical role in object *grouping* and recognition, and in *segmentation*. It is key to everyday behavior; deficits have been implicated in neurological and developmental disorders and aging

- Range of existing experimental data on spatial context (neural; perceptual). Lacking principled explanation

- Poor understanding for how we (and our cortical neurons) process complex, natural images

- Spatial context in scenes
- Temporal context in image sequences
- Context given by attention and task

Contextual influences

Perceptual illusions:

Review paper on context:
Schwartz, Hsu, Dayan, Nature Reviews Neuroscience 2007

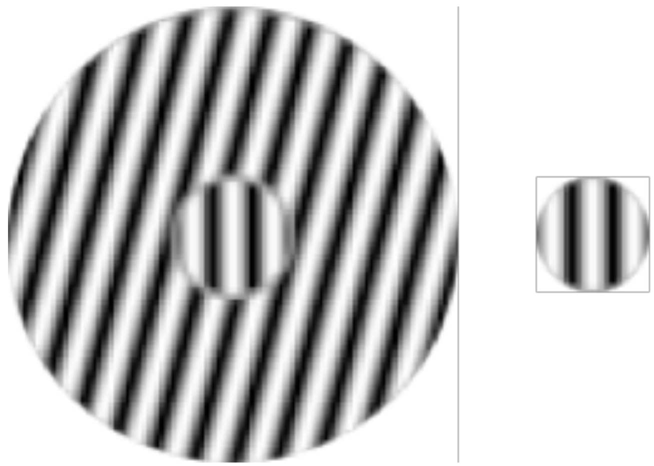
Contextual influences

Perceptual illusions:

Review paper on context:
Schwartz, Hsu, Dayan, Nature Reviews Neuroscience 2007

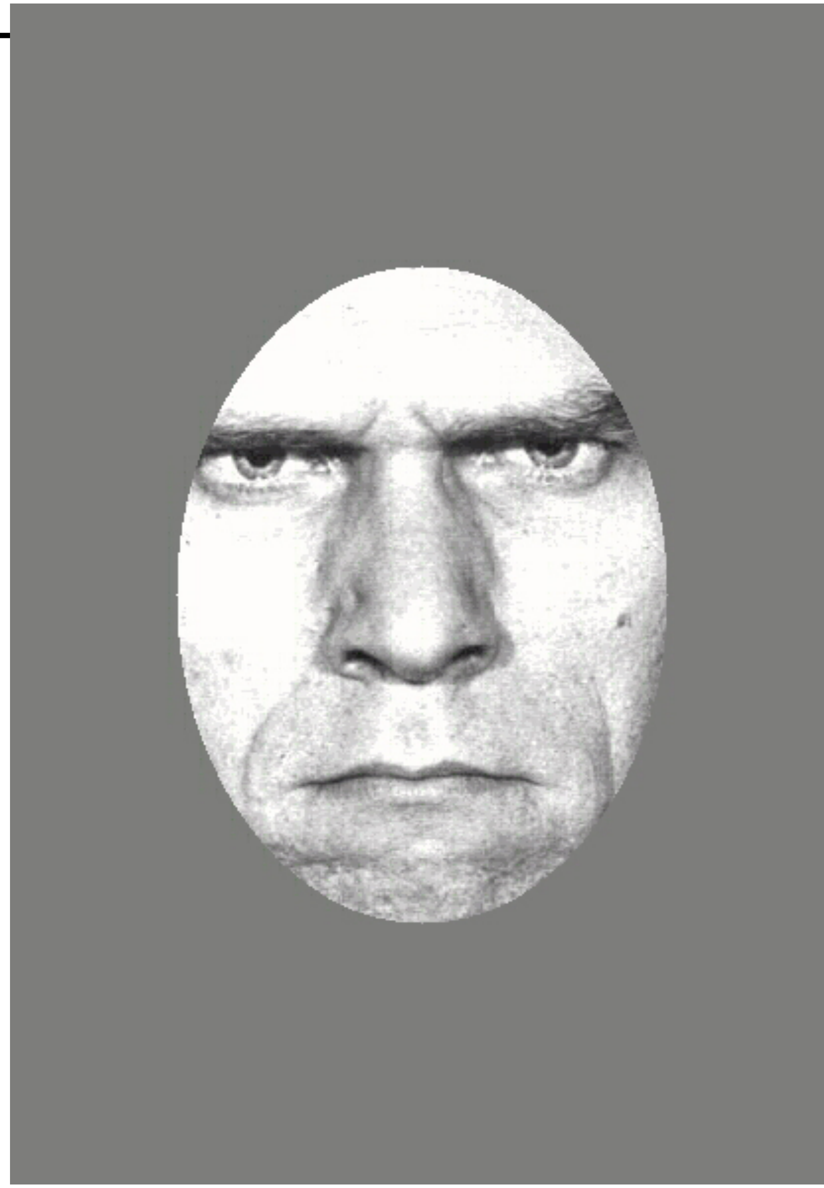
Spatial context

Perceptual illusions



Contextual effects in time...

Adaptation to expression: pre-adapt (from Michael Webster)



adapt

post-adapt

Spatial context

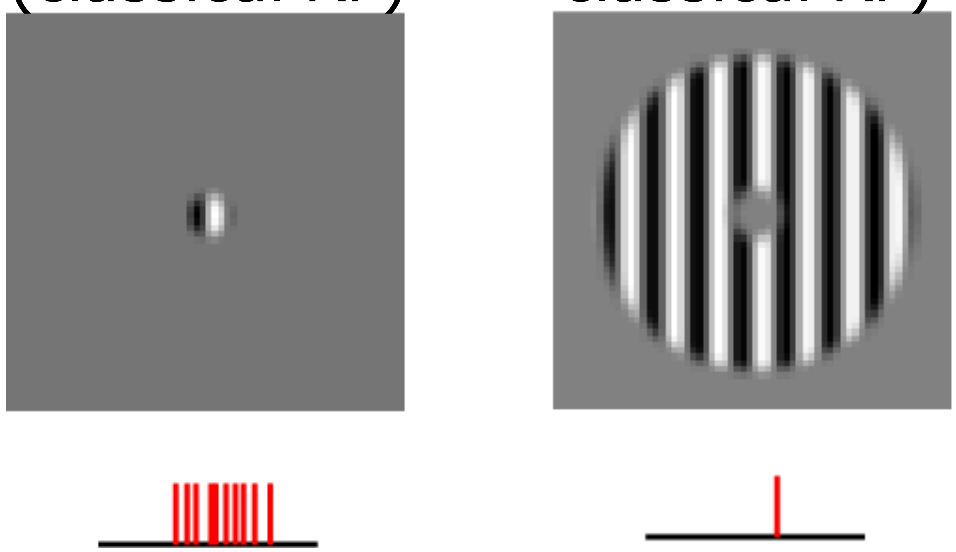
Cortical neurons

Center (classical RF)

Surround (non classical RF)

Stimulus

Spikes



Surround stimulus is defined such that by itself elicits no response

Spatial context

Cortical neurons

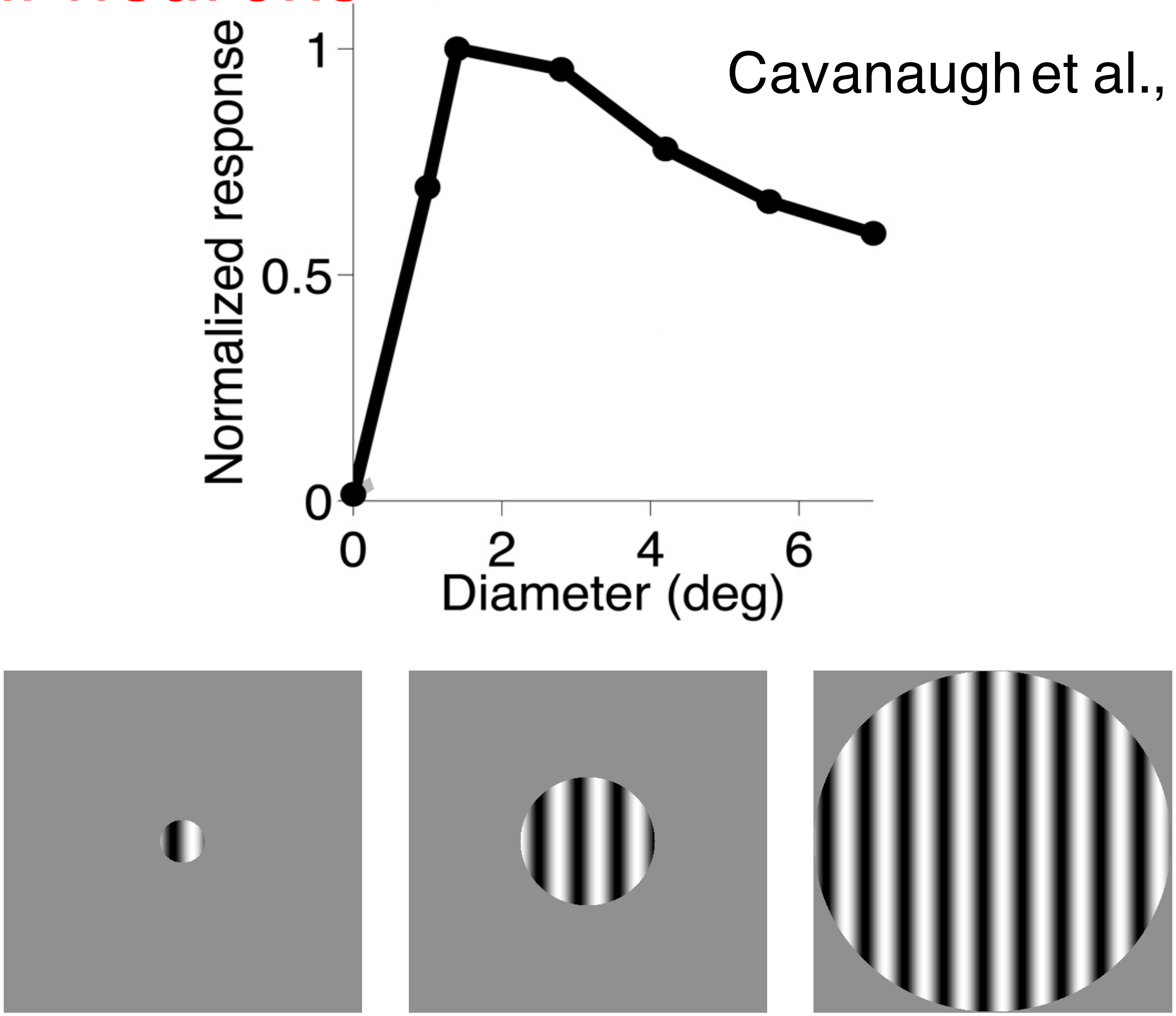
Stimulus

Spikes

But surround stimulus can modulate response to center. Cortical neurons are affected by spatial context.

Spatial context

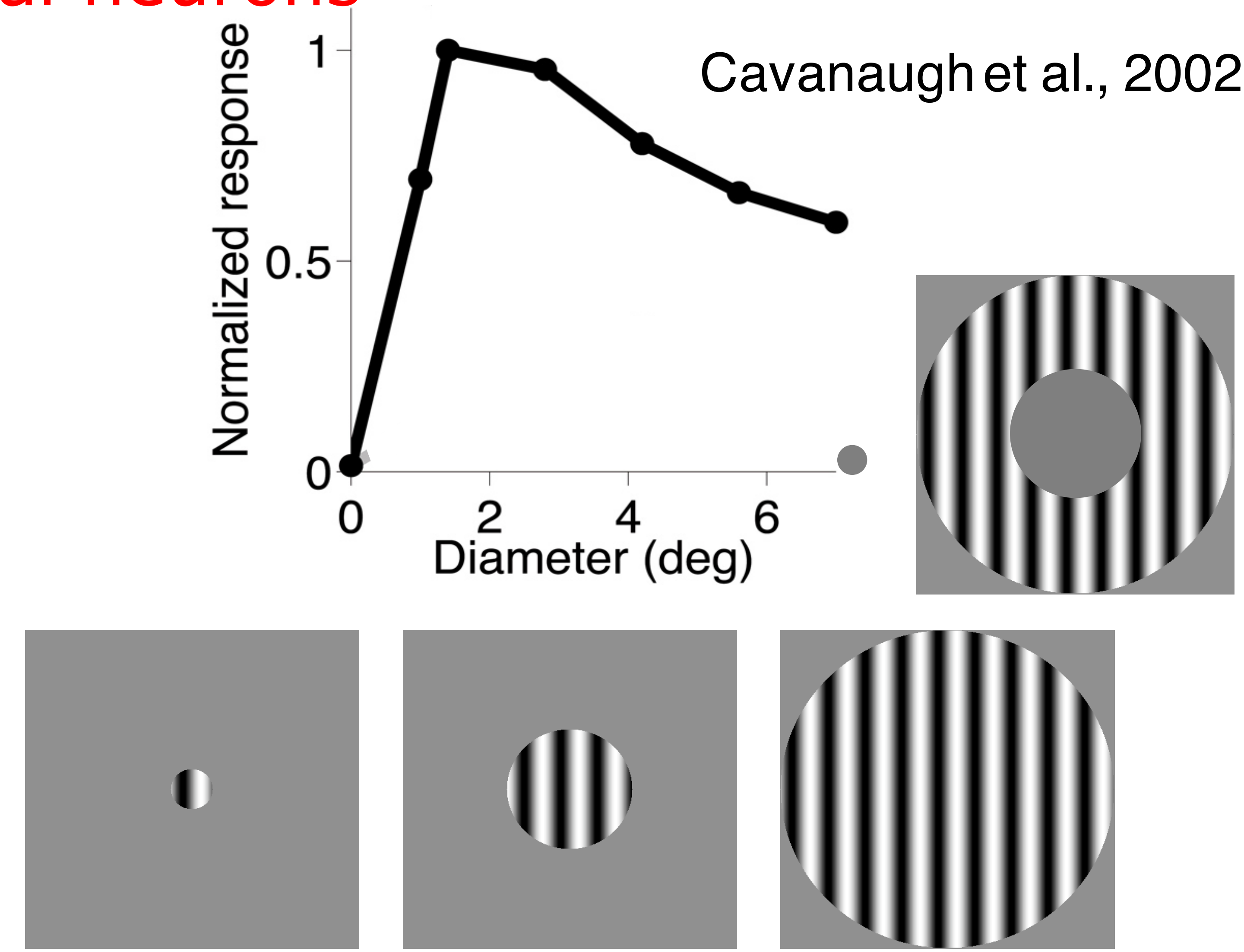
Cortical neurons



Cavanaugh et al., 2002

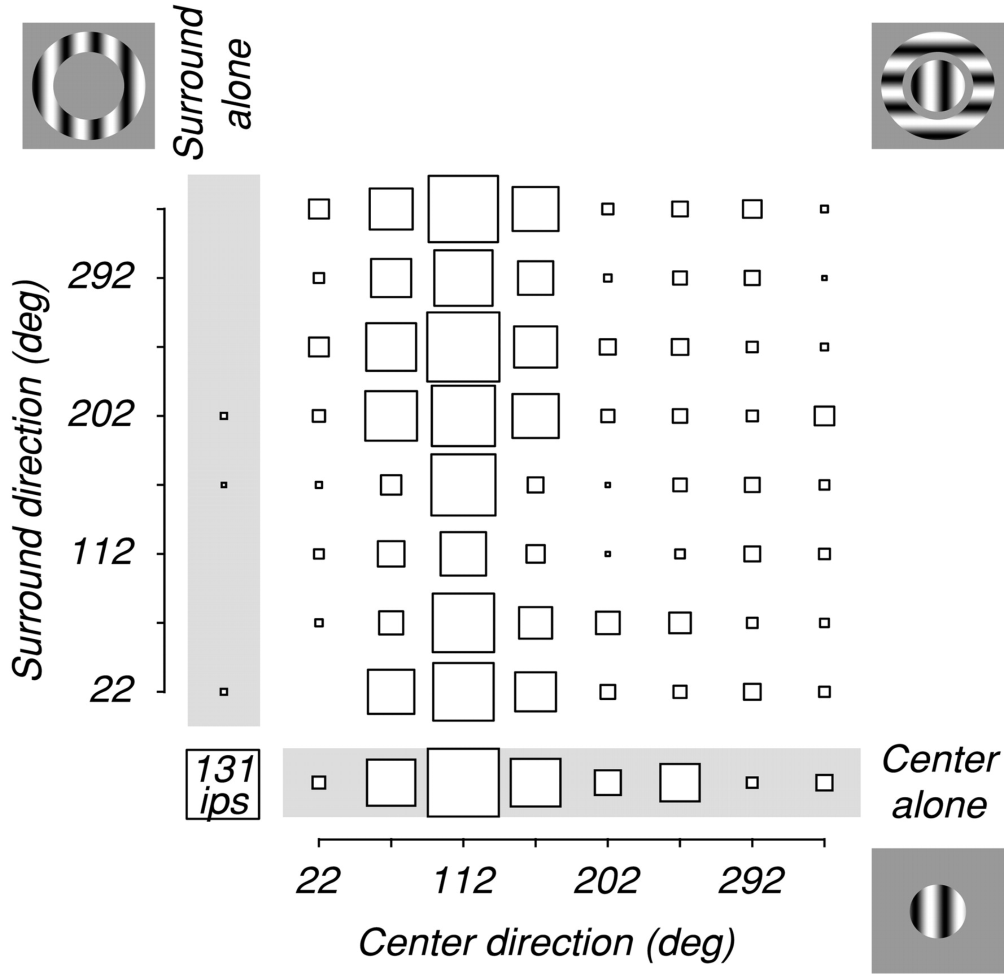
Spatial context

Cortical neurons



Spatial context

Cortical neurons

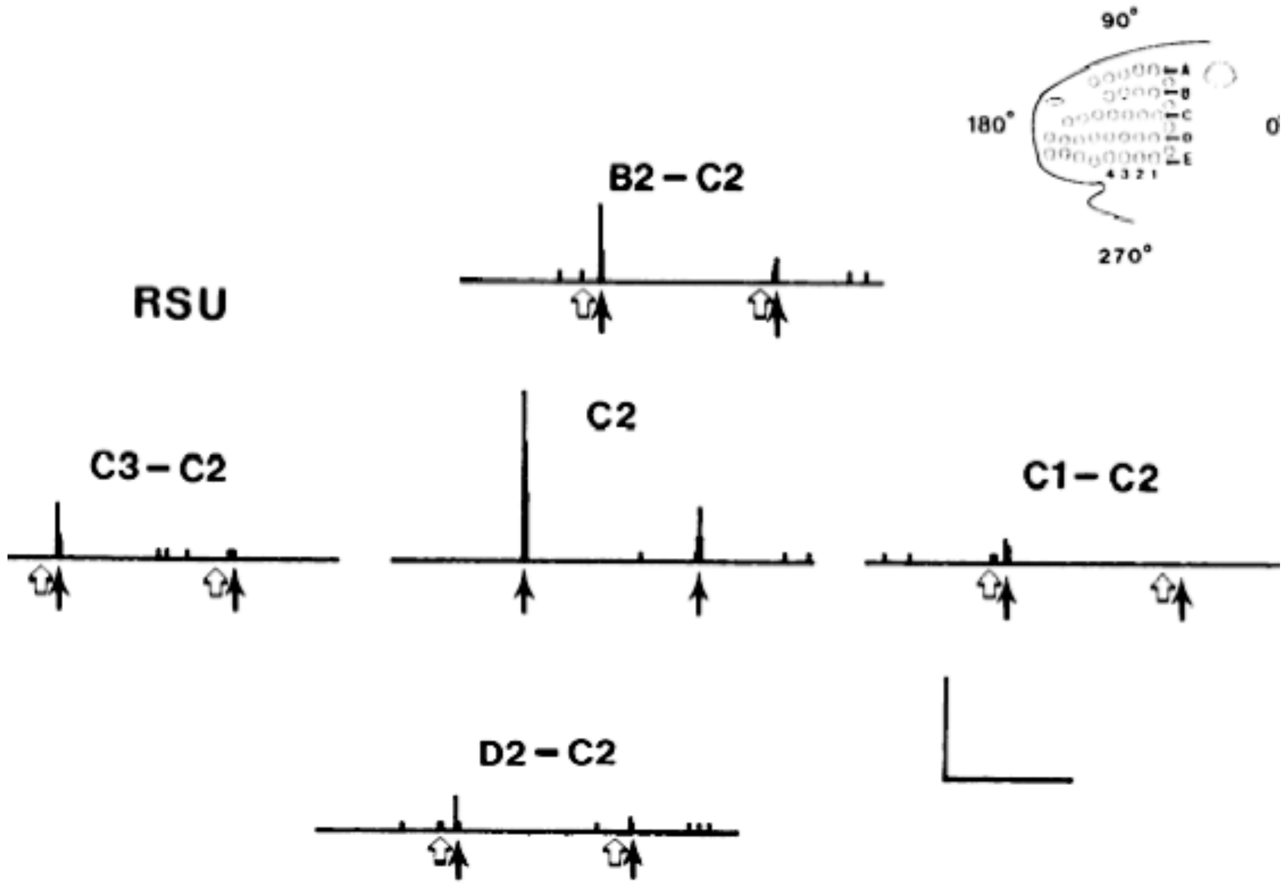


Cavanaugh et al. 2002

In what other sensory systems might we expect contextual effects?

Spatial context

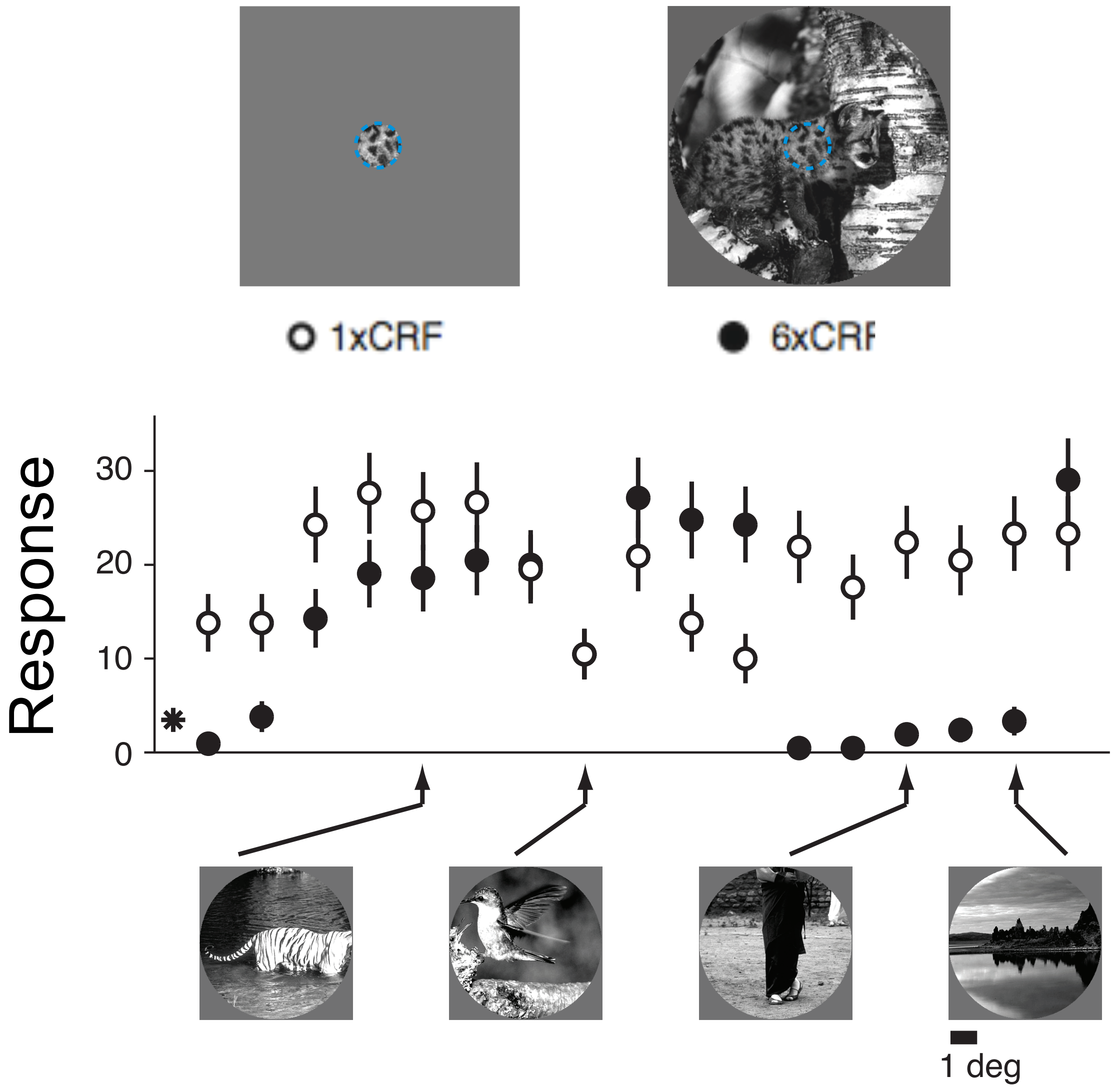
Cortical neurons: other modalities…



Simons and Carvell, 1989

Cortical Neurons

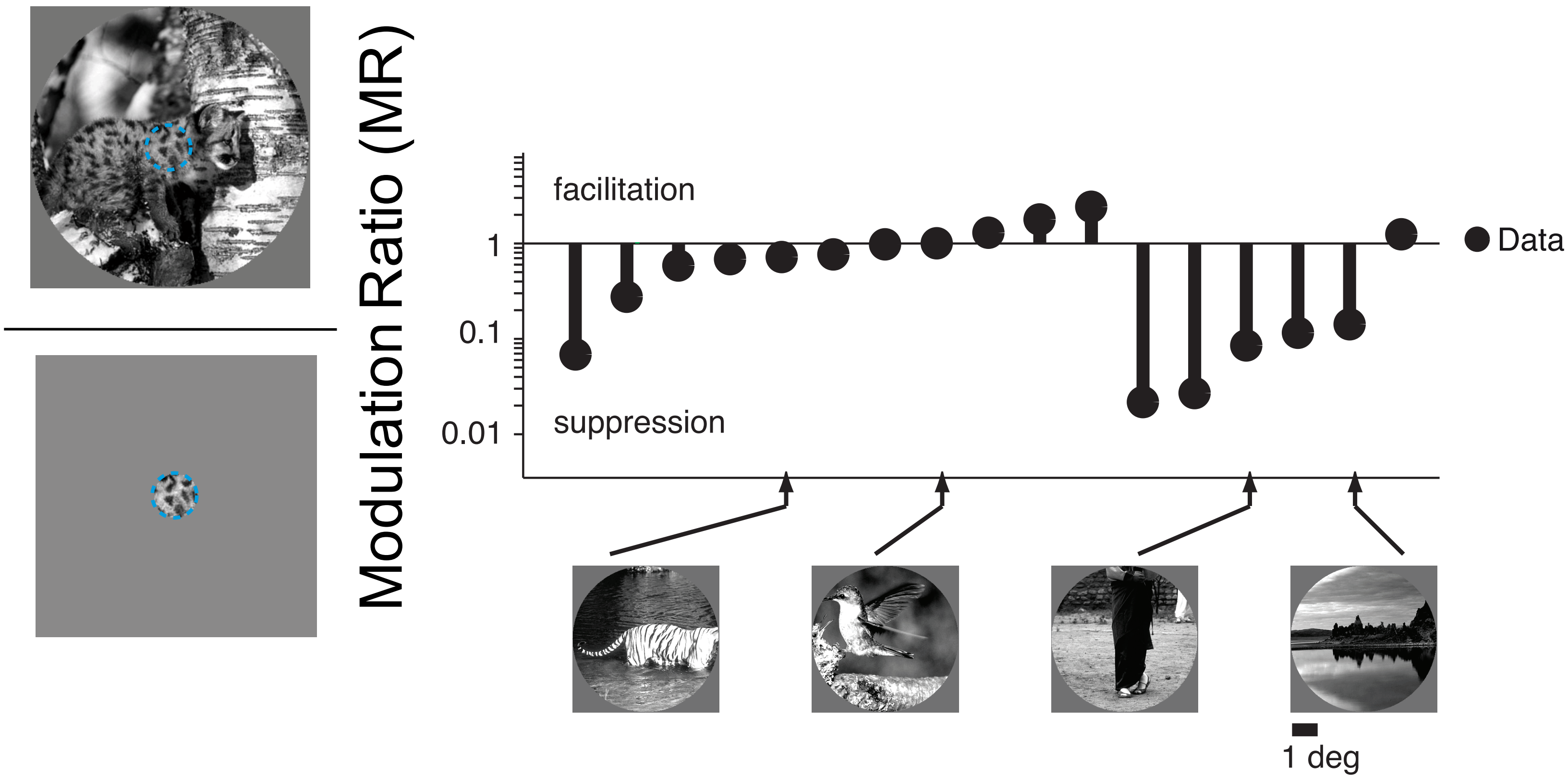
- Spatial context and natural scenes



Data: Adam Kohn lab (Coen-Cagli, Kohn, Schwartz, 2015)

Cortical Neurons

- Spatial context and natural scenes



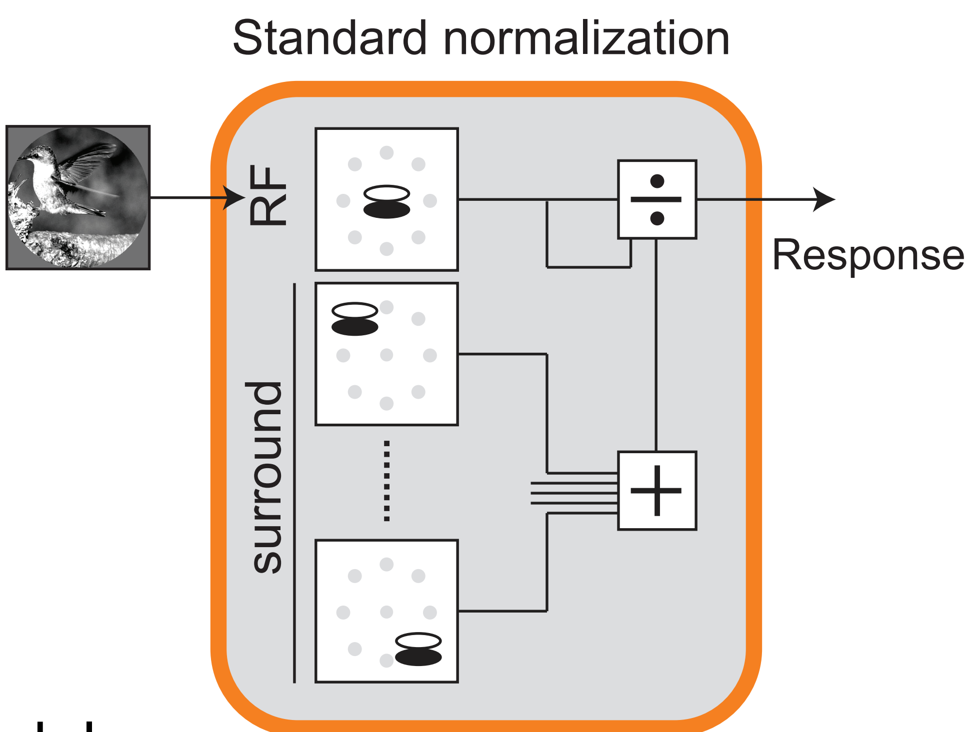
Data: Adam Kohn lab (Coen-Cagli, Kohn, Schwartz, 2015)

Cortical Neurons

- Spatial context and natural scenes

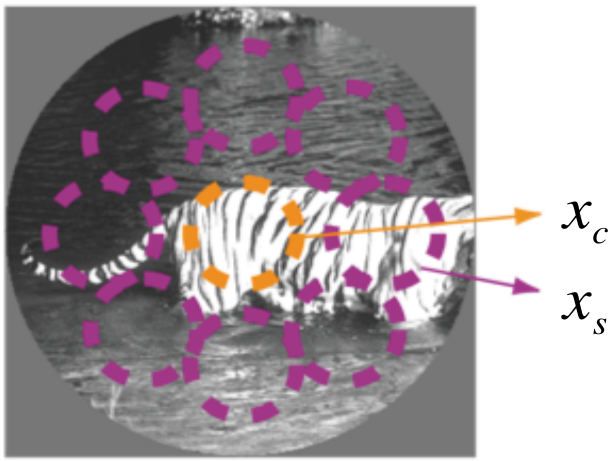
Can we capture data with canonical divisive normalization? (descriptive model)

Divisive normalization



- Descriptive model
- Canonical computation (Carandini, Heeger, Nature Reviews Neuro, 2012)
- Has been applied to visual cortex, as well as other systems and modalities, multimodal processing, value encoding, etc

Cortical Neurons

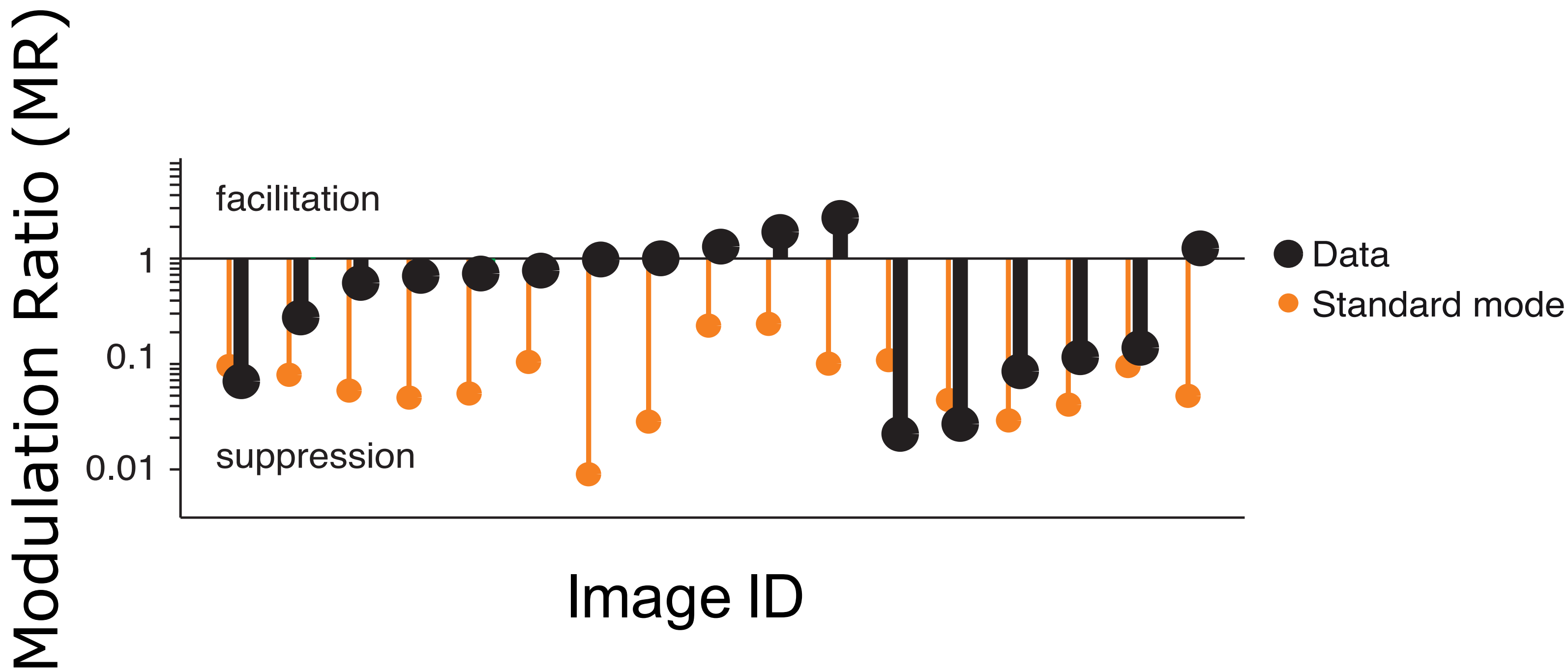


Canonical divisive normalization:

$$R_c \prec \frac{x_c}{\sqrt{x_c^2 + x_s^2}}$$

V1 Data: Kohn lab

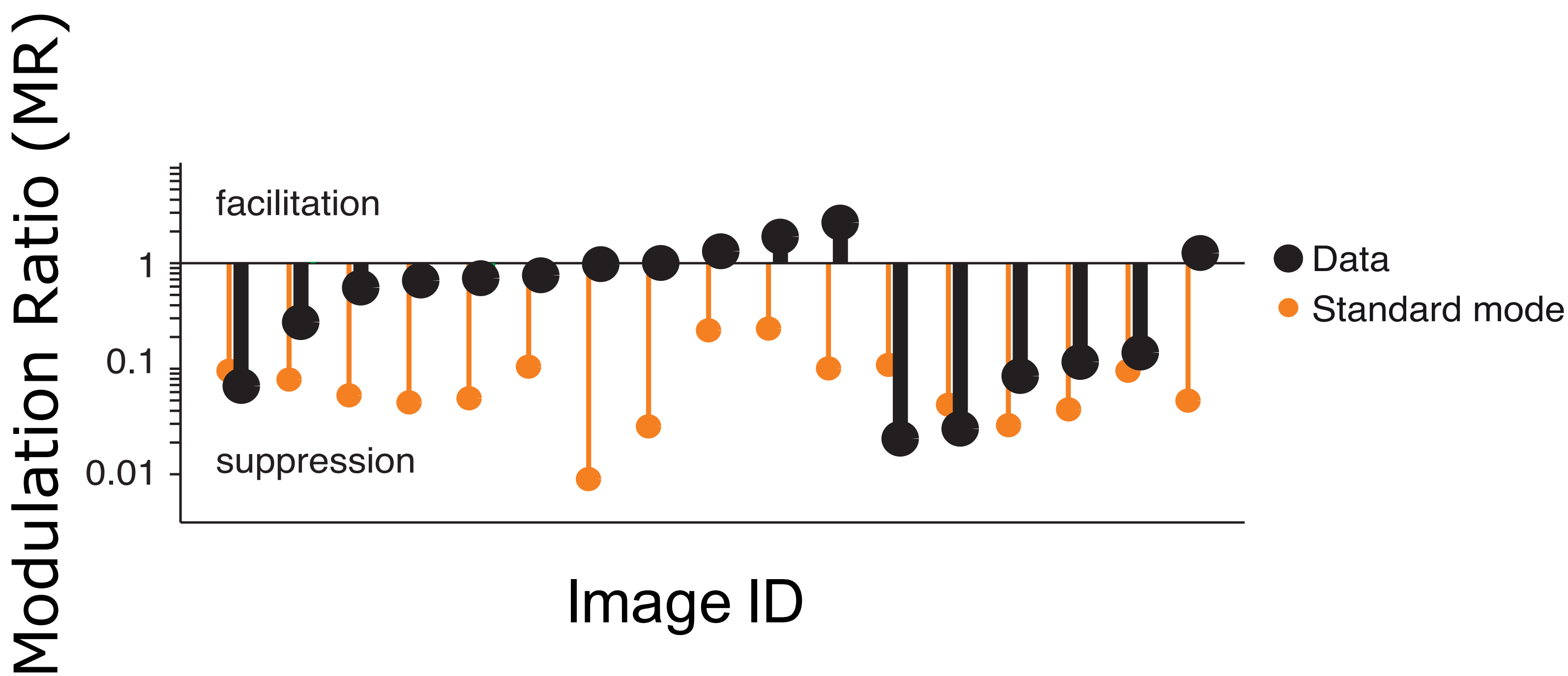
Cortical responses to natural images



- We fit the standard normalization model to neural data
- Poor prediction quality

Data: Adam Kohn lab
Coen-Cagli, Kohn, Schwartz, 2015

Cortical responses to natural images



- Can we explain as strategy to encode natural images optimally based on expected contextual regularities?

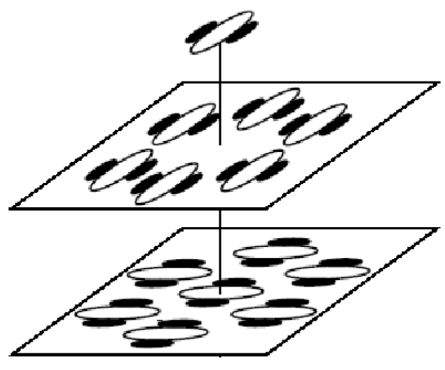
Data: Adam Kohn lab
Coen-Cagli, Kohn, Schwartz, 2015

Outline

- Generative model that captures the joint statistics and its relation to cortical neural processing

- Application and testing on biological data

Generative model



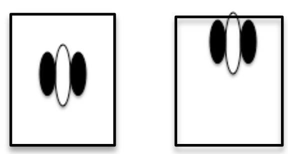
Generative model
of the coordination of
filter activations

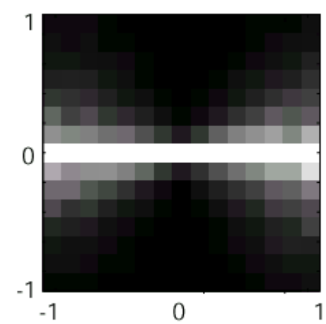
Generate filter coordination

Hidden variables

generative

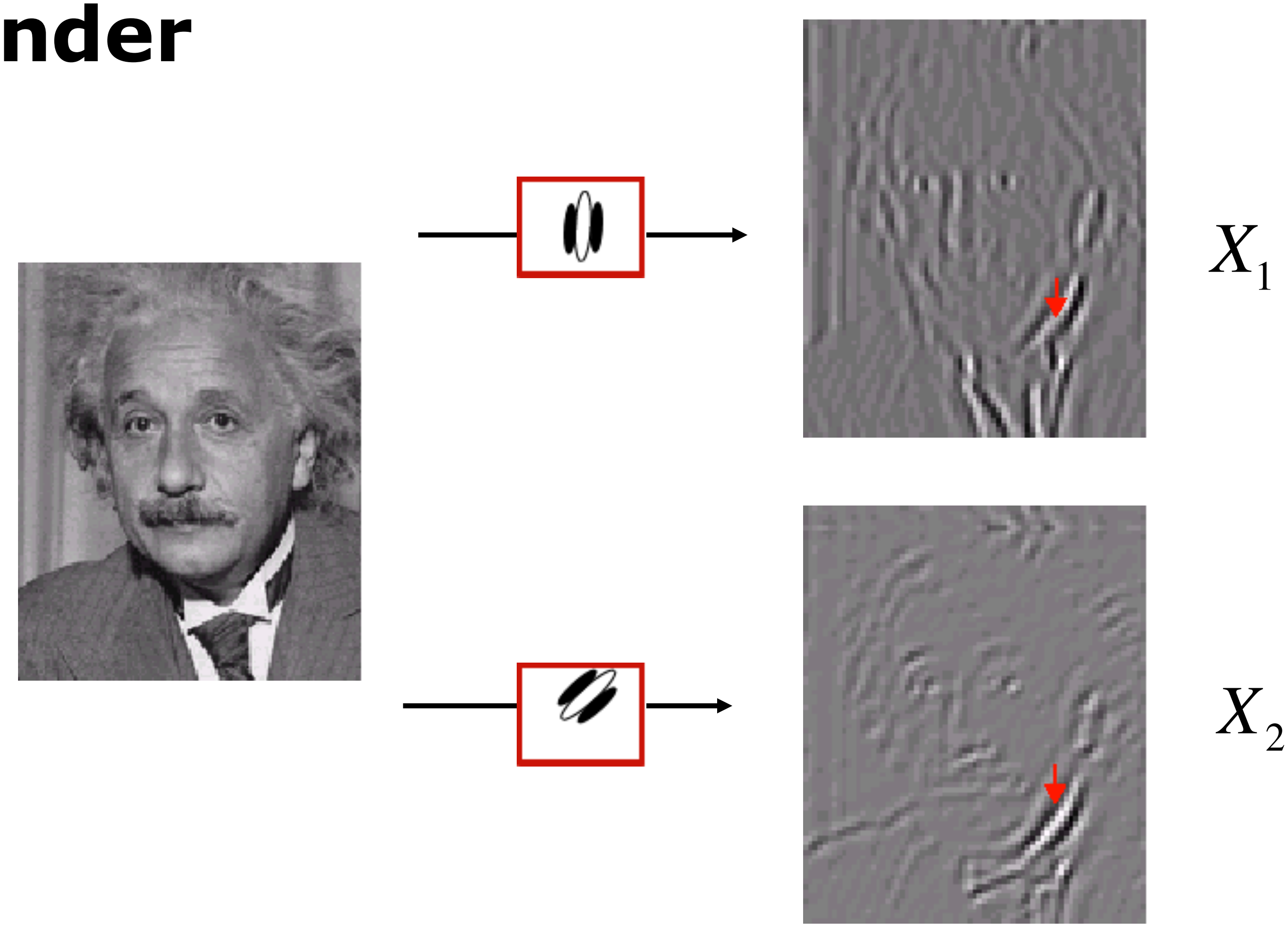
Filter activations





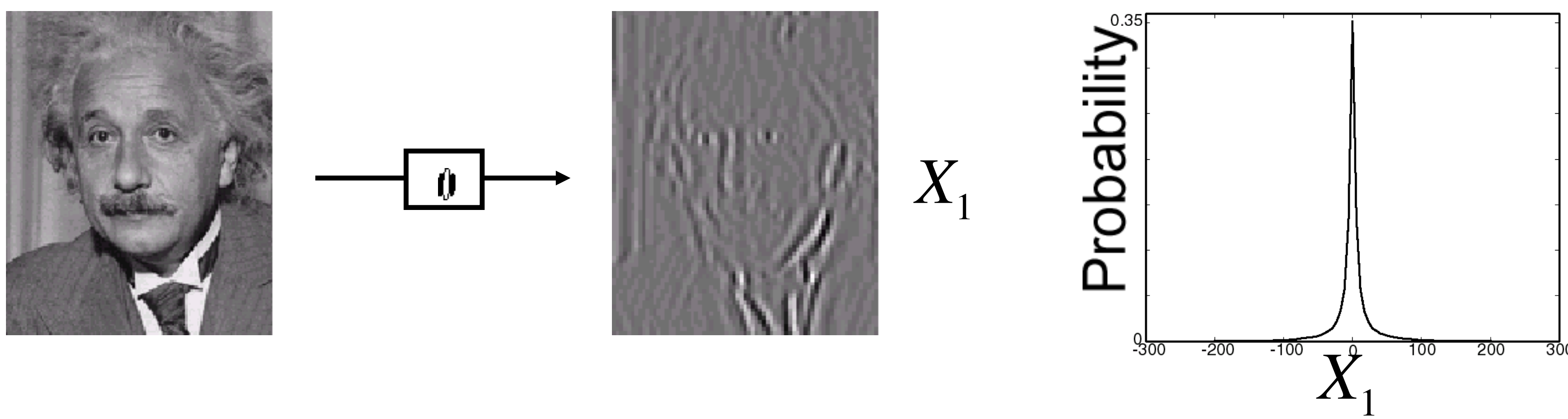
Bottom-up Joint Statistics

Reminder



Bottom-up Statistics

Reminder

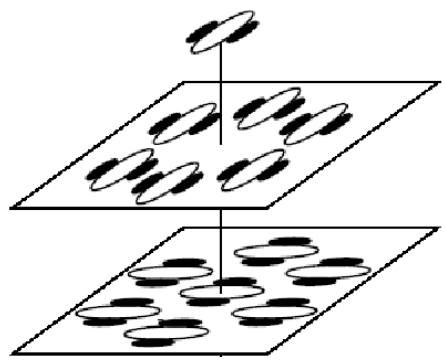


After Field, 1987

GSM Generative model

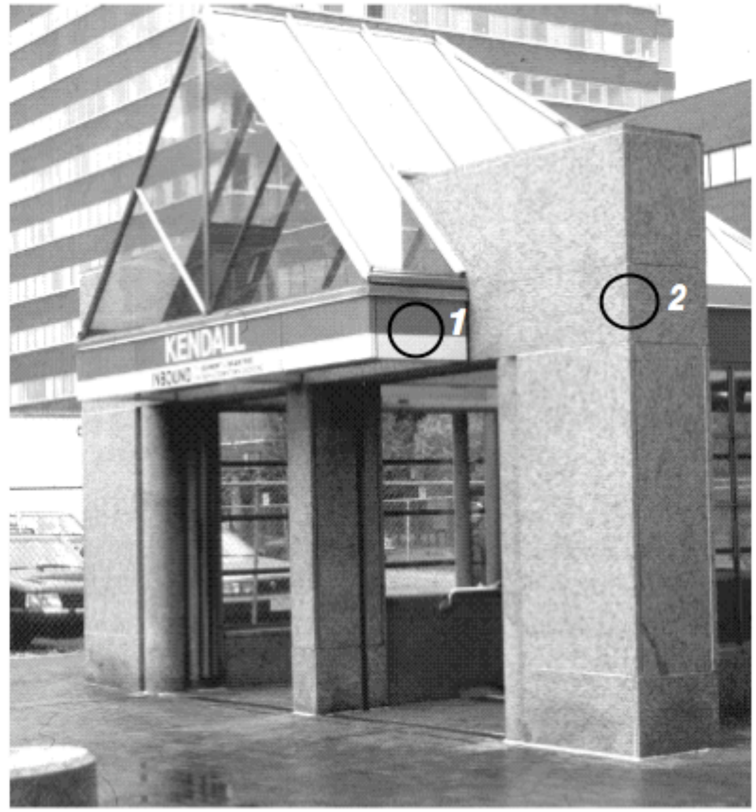
Gaussian Scale Mixture (GSM) – multiplicative model
(Andrews & Mallows, 1974. Wainwright & Simoncelli, 2000 **)**

$$x_1 = v g_1$$
$$x_2 = v g_2$$



Coordinating filter activations via a *common multiplicative hidden variable*

Analogy…Brightness percept as multiplicative

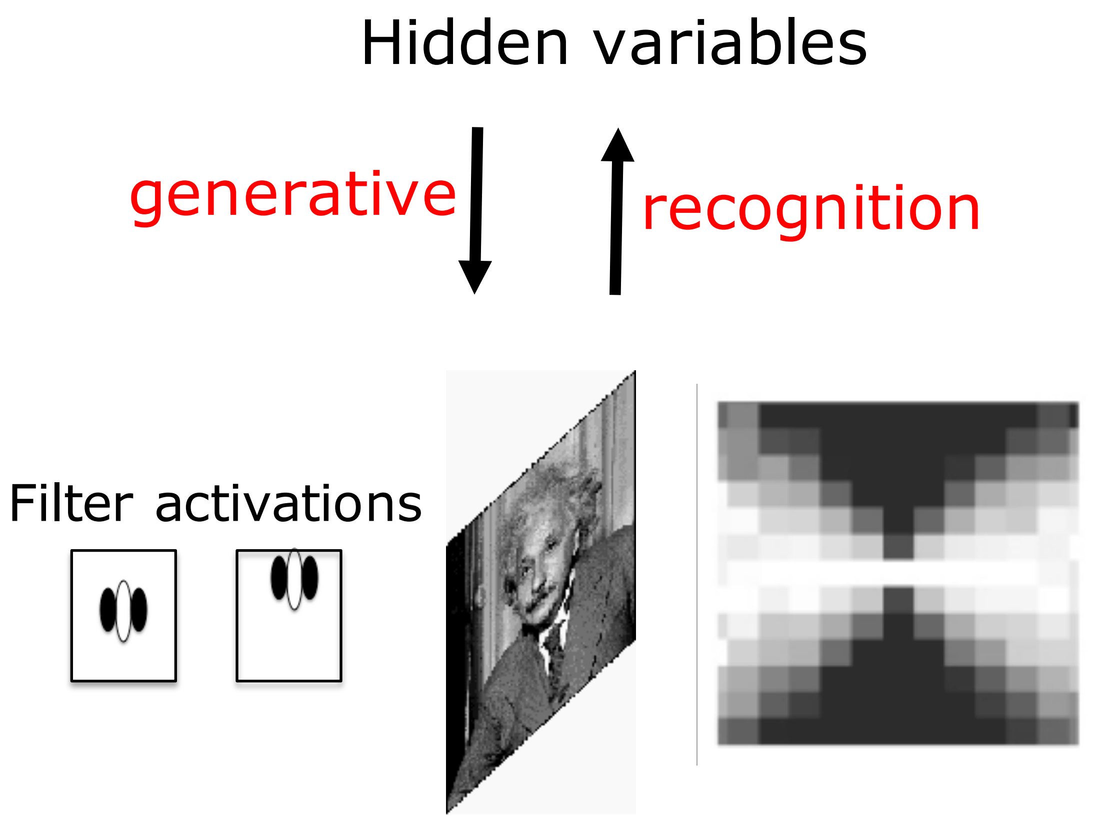


Observed brightness = illuminant times reflectance

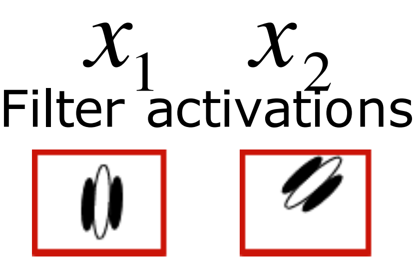
global local

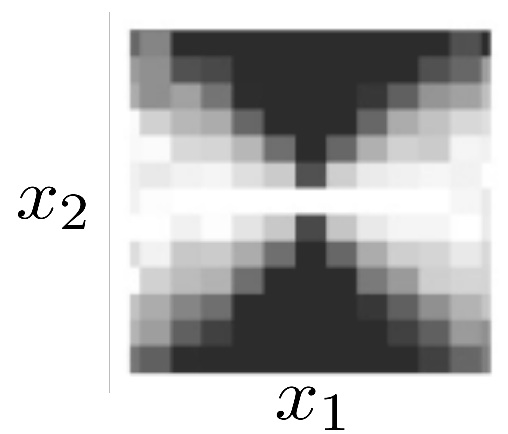
From Adelson 1993

GSM Generative model



Modeling Statistical dependencies: Gaussian Scale Mixture (GSM)

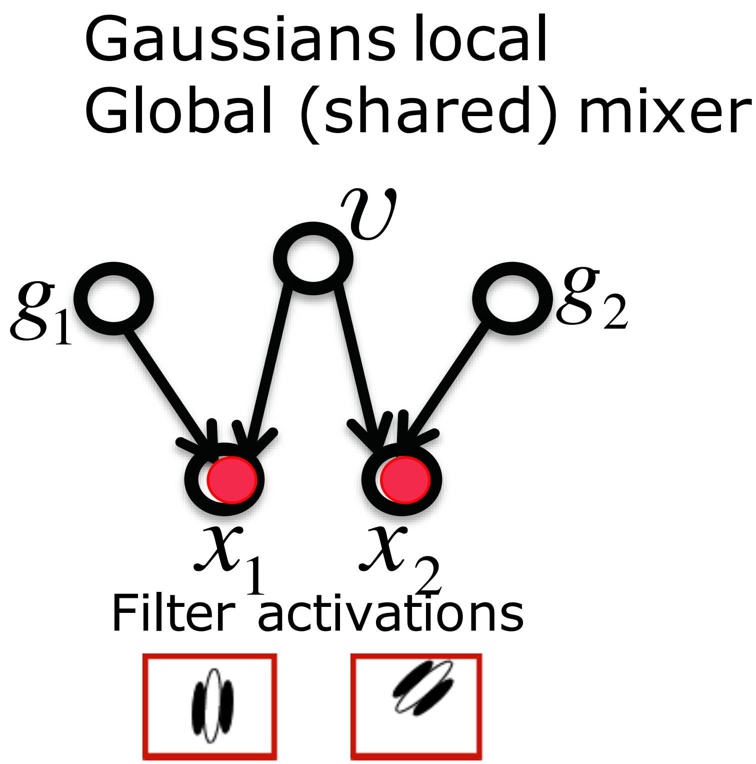


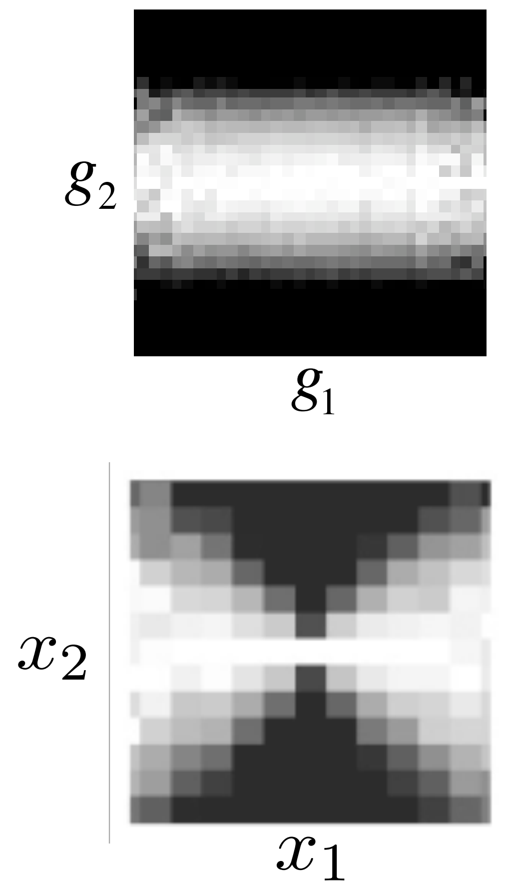


Andrews & Mallows, 1974; Wainwright & Simoncelli, 2000

Modeling Statistical dependencies: Gaussian Scale Mixture (GSM)

$$x_1 = v g_1$$
$$x_2 = v g_2$$

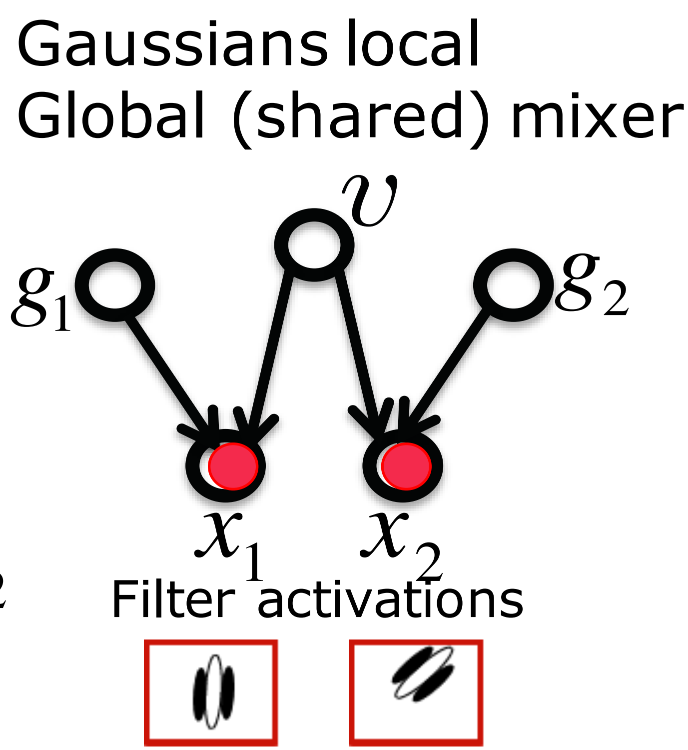




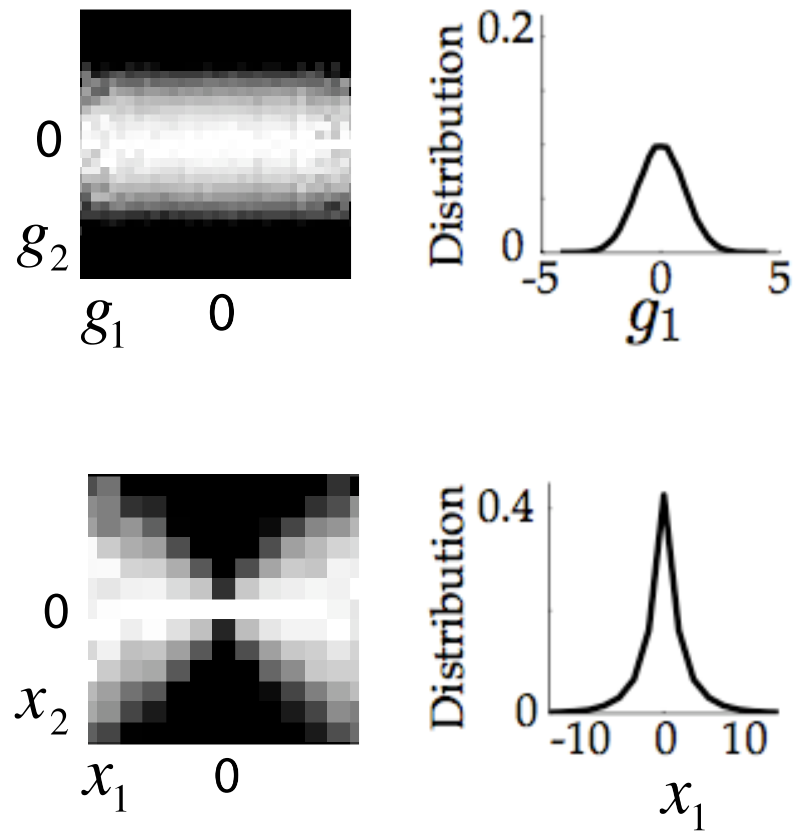
Andrews & Mallows, 1974; Wainwright & Simoncelli, 2000

GSM Generative model

Gaussian Scale Mixture (GSM) – multiplicative model
Also generates sparse marginal…



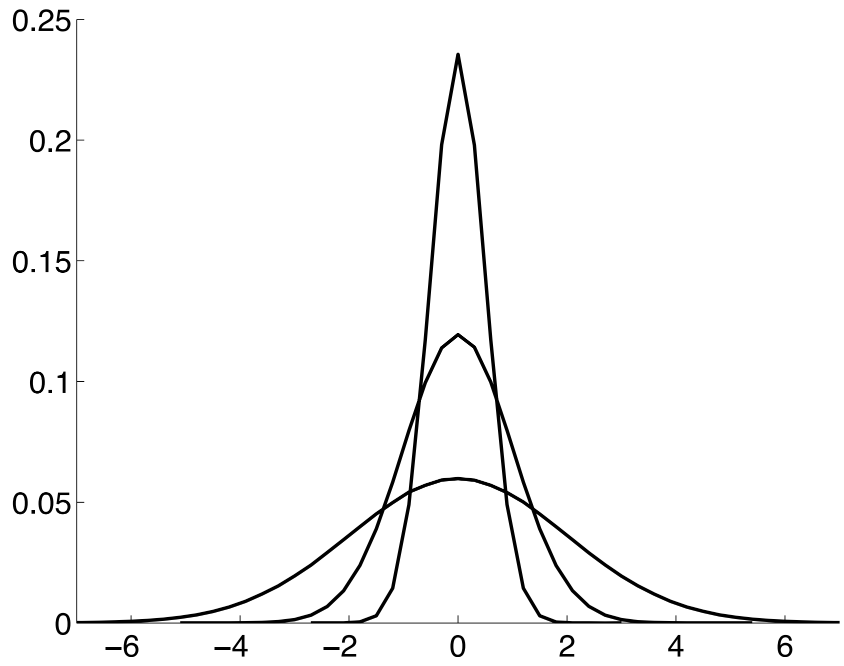
$$x_1 = vg_1$$
$$x_2 = vg_2$$



GSM Generative model

Gaussian Scale Mixture (GSM) – multiplicative model
Also generates sparse marginal…

Example: mixing few zero mean Gaussians multiplied
By different scale factors or mixer values:

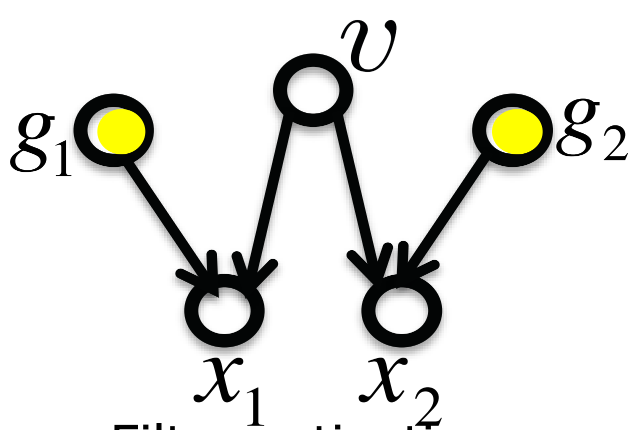


Modeling Statistical dependencies: Gaussian Scale Mixture (GSM)

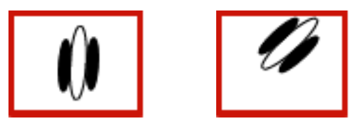
Gaussians local
Global (shared) mixer

Infer local Gaussian

$$x_1 = v g_1$$
$$x_2 = v g_2$$



Filter activations

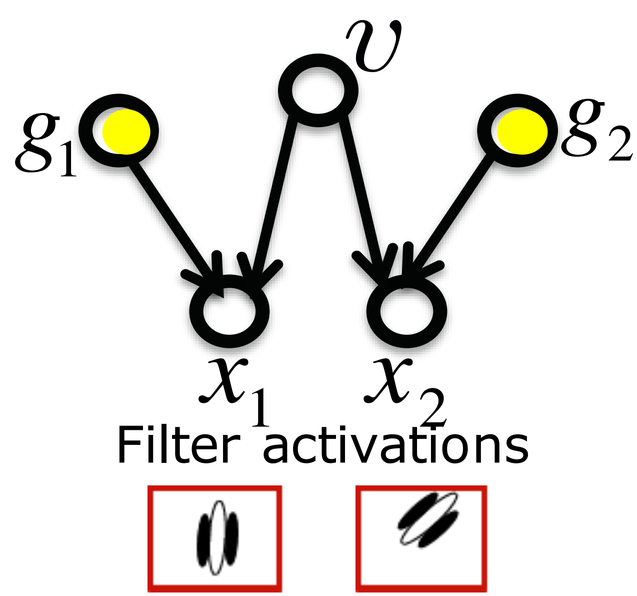


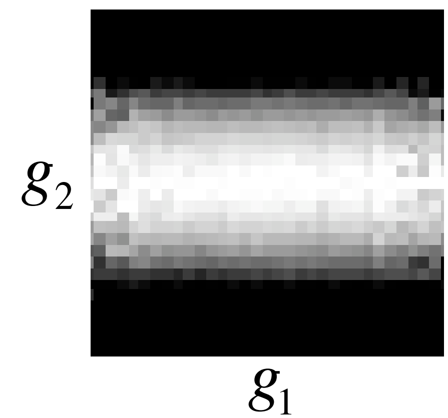
$E(g_1 \mid x_1, x_2) =$ Model neuron activity

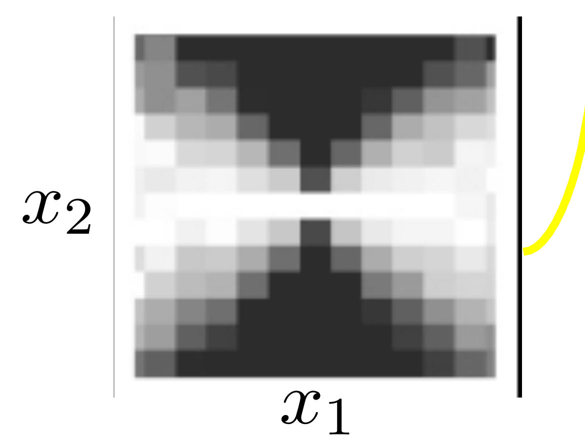
Modeling Statistical dependencies: Gaussian Scale Mixture (GSM)

Gaussians local
Global (shared) mixer

$$x_1 = v g_1$$
$$x_2 = v g_2$$



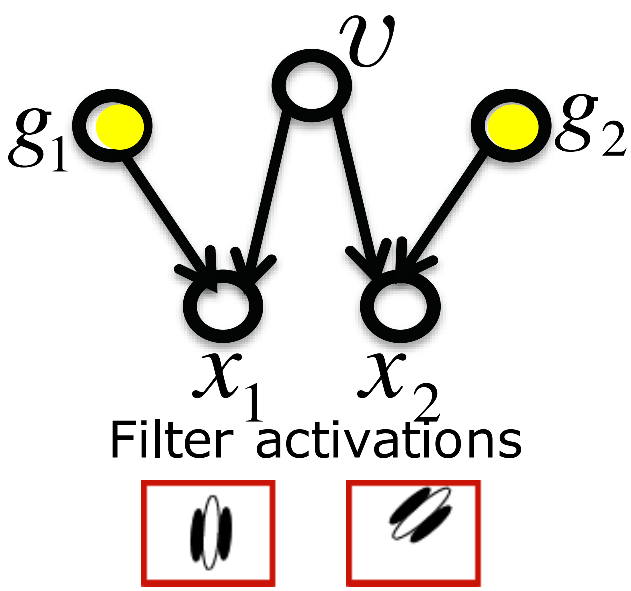




EFFICIENT CODING

Modeling Statistical dependencies: Gaussian Scale Mixture (GSM)

Gaussians local
Global (shared) mixer



Computed via Bayes rule

$$E(g_1 \mid x_1, x_2) \prec \boxed{\frac{x_1}{\sqrt{l}}};$$

$$l = \sqrt{x_1^2 + x_2^2}$$

DIVISIVE NORMALIZATION

Divisive Normalization Canonical Model

Divisive normalization *descriptive* models have been applied in many neural systems. Here we provide a *principled explanation.* We will next show that it also leads to a richer model based on image statistics and makes predictions

GSM Generative model

- We'll pause for a moment and take a more detailed look at the math…

GSM Generative model

- Mixer and Gaussian component estimation is formally ill-posed, but regularized through prior distributions.

Mixer prior:

$$P(v) \propto [v \exp(-v^2/2)]$$

Gaussian prior:

$$P(g_i) \propto Gauss(0,1)$$

- Posterior statistics have closed form expressions!

Generative model (GSM)

- Mixer and Gaussian component estimation is formally ill-posed, but regularized through prior distributions.

Mixer prior:

$$P(v) \propto [v \exp(-v^2/2)]$$

Gaussian prior:

$$P(g) \propto Gauss(0,1)$$

- Posterior statistics have closed form expressions!

Let $l = \sqrt{\sum_i x_i^2}$ sum of squared filter responses

$$E(v \mid x) = \sqrt{l}\,\frac{Besselk(\frac{3}{2}-\frac{n}{2},l)}{Besselk(1-\frac{n}{2},l)} \qquad E(g_1 \mid x) = sign(x_1)\sqrt{|x_1|}\sqrt{\frac{|x_1|}{l}}\,\frac{Besselk(\frac{n}{2}-\frac{1}{2},l)}{Besselk(\frac{n}{2}-1,l)}$$

GSM Generative model

Gaussian Scale Mixture (GSM) – multiplicative model
(Andrews & Mallows, 1974. Wainwright & Simoncelli, 2000 **)**

$$p(v) = v \exp(-v^2/2);$$

$$p(g_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-g_1^2/2);$$

$$x_1 = vg_1; x_2 = vg_2$$

Use Bayesian inference to get out the Gaussian estimate:

$$p(g_1 \mid x_1) \prec p(g_1)\, p(x_1 \mid g_1);$$

$$p(x_1 \mid g_1) \prec p(v = x_1 / g_1) = \frac{|x_1| \exp(-x_1^2 / 2g_1^2)}{g_1^2};$$

GSM Generative model

- Mixer and Gaussian component estimation is formally ill-posed, but regularized through prior distributions.

Mixer prior:

$$P(v) \propto [v \exp(-v^2/2)]$$

Gaussian prior:

$$P(g_i) \propto Gauss(0,1)$$

- Posterior statistics have closed form expressions!

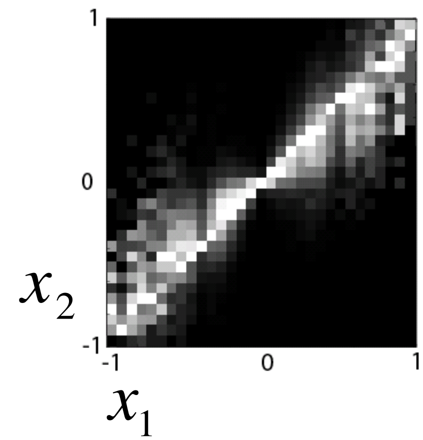
$$p(g_1 \mid x_1, x_2) \prec p(g_1)\, p(x_1, x_2 \mid g_1) = p(g_1)\, p(x_1 \mid g_1)\, p(x_2 \mid v = x_1 / g_1)$$

$$E(g_1 \mid x_1, x_2) = sign(x_1)\sqrt{|x_1|}\sqrt{\frac{|x_1|}{l}}\,\frac{B(\frac{n}{2}-\frac{1}{2}, l)}{B(\frac{n}{2}-1, l)}$$

$$l = \sqrt{\sum_i x_i^2}$$

Extension 1: include linear correlations

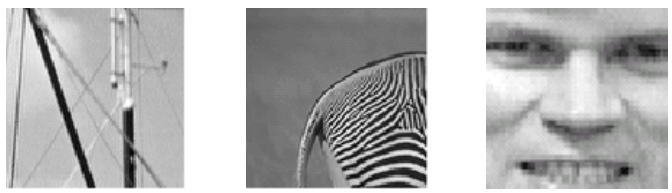
$$x_1 = v g_1$$
$$x_2 = v g_2$$

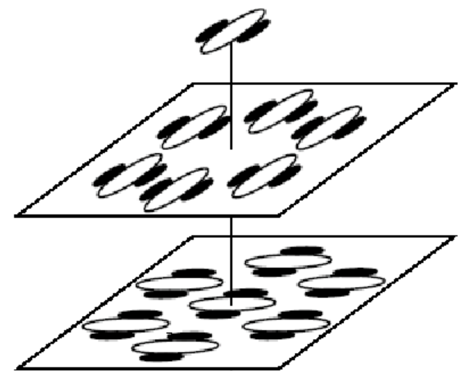


$$l = \sqrt{\sum_i x_i^2} \longrightarrow l = \sqrt{\sum_{i,j} \Sigma_{i,i}^{-1} x_i^2 + \Sigma_{i,j}^{-1} x_i x_j}$$

$$E(g_1 \mid x_1, x_2) = \operatorname{sign}(x_1)\sqrt{x_1}\sqrt{\frac{|x_1|}{l}}\,\frac{B\left(\frac{n}{2}-\frac{1}{2}, l\right)}{B\left(\frac{n}{2}-1, l\right)}$$

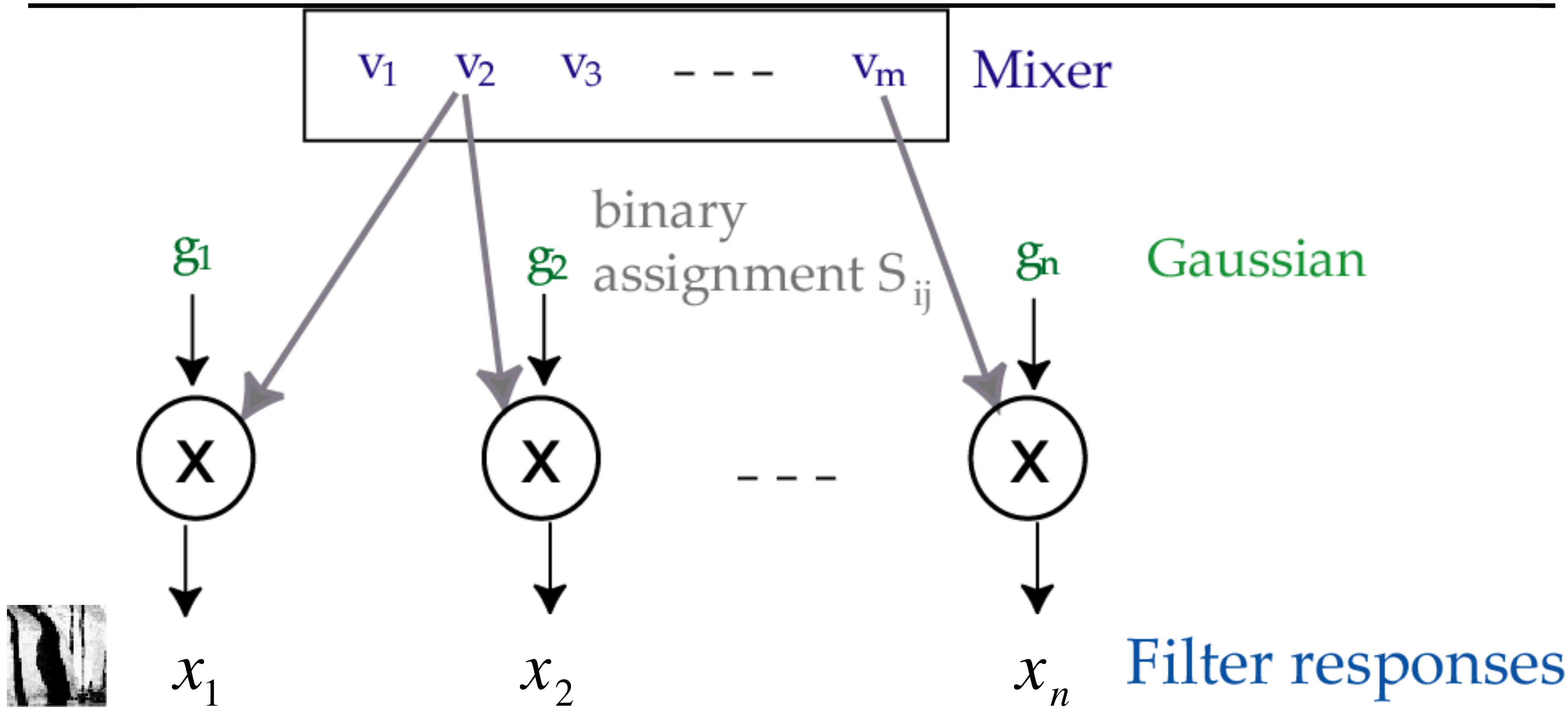
Extension 2: Non homogeneity of scenes





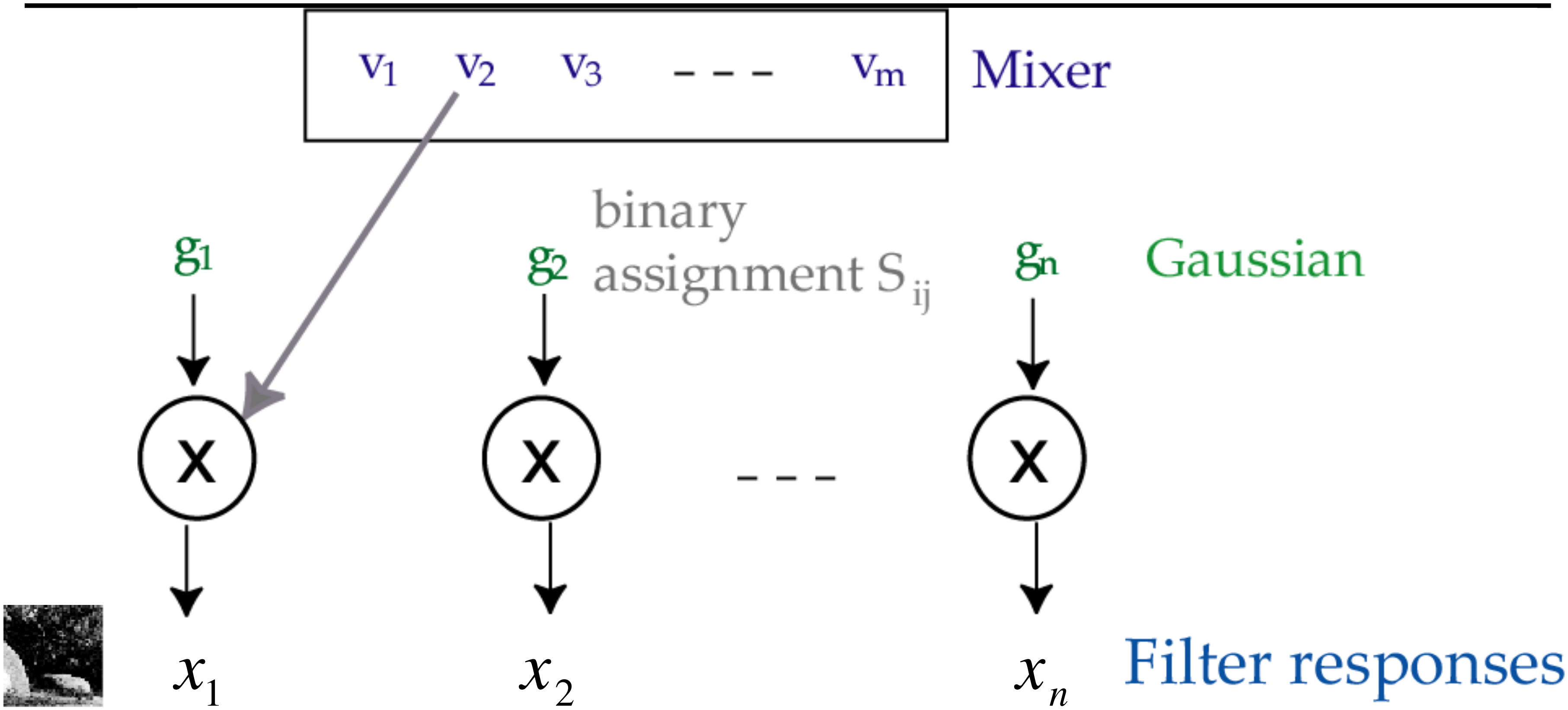
Just one common mixer for all filters
and image patches ??

Extension 2: Non homogeneity of scenes



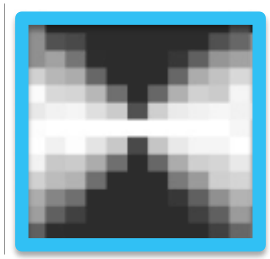
- From Schwartz, Sejnowski, Dayan, 2006
- Wainwright, Simoncelli & Willsky, 2001: dependence arising from tree.
- Karklin & Lewicki, 2003: generating mixer value for each filter as linear combination of the values of a small number of underlying components.
- Williams & Adams, 1999; 2003: similar assignment problem in dynamical tree.
- Lyu & Simoncelli, NIPS 2006: Mixers in a MRF.

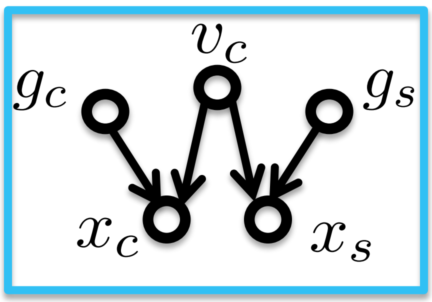
Extension 2: Non homogeneity of scenes

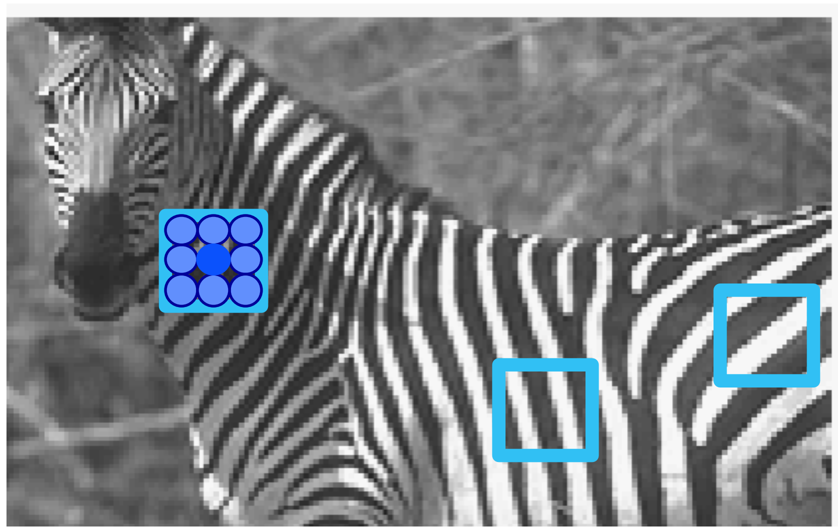


- From Schwartz, Sejnowski, Dayan, 2006
- Wainwright, Simoncelli & Willsky, 2001: dependence arising from tree.
- Karklin & Lewicki, 2003: generating mixer value for each filter as linear combination of the values of a small number of underlying components.
- Williams & Adams, 1999; 2003: similar assignment problem in dynamical tree.
- Lyu & Simoncelli, NIPS 2006: Mixers in a MRF.

Non-homogeneity of Scenes



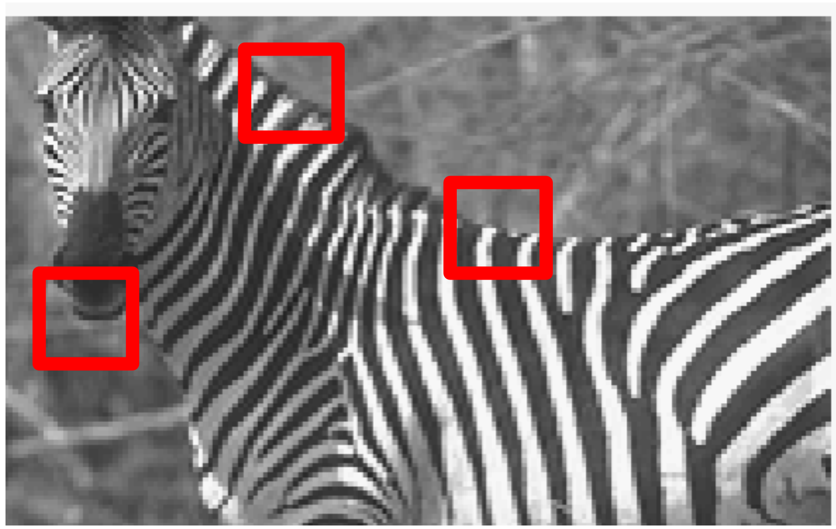


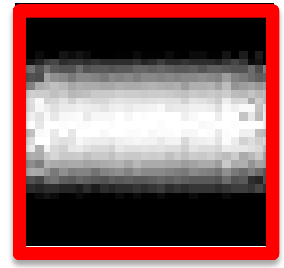


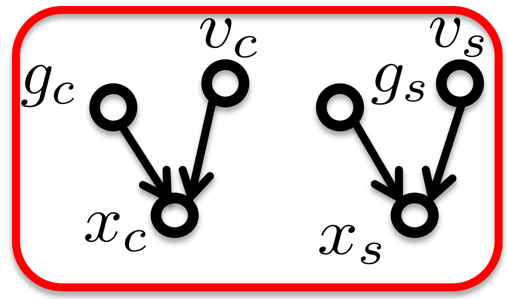
homogenous image patches

Schwartz, Sejnowski, Dayan, 2006; Coen-Cagli, Dayan, Schwartz, 2012
 (See also: Karklin, Lewicki 2005; Guerrero-Colon, Simoncelli, Portilla 2008)

Non-homogeneity of Scenes



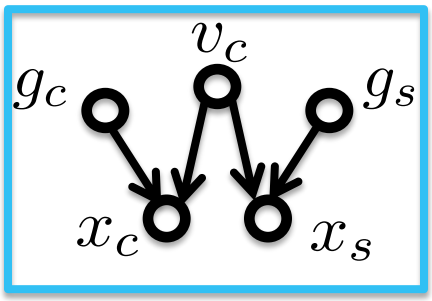


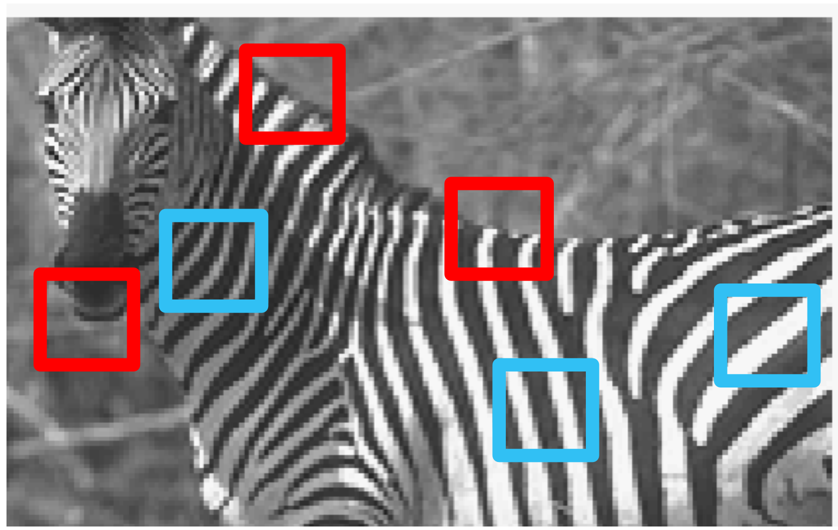


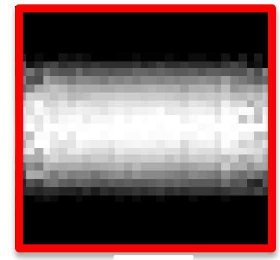
non-homogenous image patches

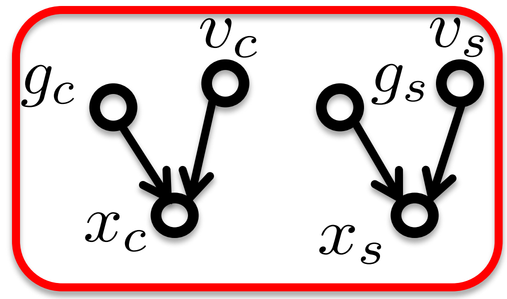
Schwartz, Sejnowski, Dayan, 2006; Coen-Cagli, Dayan, Schwartz, 2012
(See also: Karklin, Lewicki 2005; Guerrero-Colon, Simoncelli, Portilla 2008)

Non-homogeneity of Scenes









homogenous

non-homogenous

Coen-Cagli, Dayan, Schwartz, 2012

Flexible Divisive Normalization

divisive normalization ON

divisive normalization OFF

$$E[g_c \mid x_c, x_s] = p(\xi_1 \mid x_c, x_s)\, E[g_c \mid x_c, x_s, \xi_1] + p(\xi_2 \mid x_c, x_s)\, E[g_c \mid x_c, \xi_2]$$

Coen-Cagli, Dayan, Schwartz, 2012

Non-homogeneity of images

$$E[g_c \mid x_c, x_s] = p(\xi_1 \mid x_c, x_s) E[g_c \mid x_c, x_s, \xi_1] + p(\xi_2 \mid x_c, x_s) E[g_c \mid x_c, \xi_2]$$

$$p(\xi_1 \mid x) \prec p(\xi_1) p(x \mid \xi_1) = p(\xi_1) \int dv_c\, p(v_c) p(x \mid v_c, \xi_1);$$

$$p(\xi_2 \mid x) \prec p(\xi_2) p(x \mid \xi_2) = p(\xi_2) \int dv_c\, p(v_c) p(x_c \mid v_c, \xi_2) \int dv_s\, p(v_s) p(x_s \mid v_c, \xi_2)$$

Schwartz, Sejnowski, Dayan, 2009; Coen-Cagli, Dayan, Schwartz, PLoS Comp Biology 2012

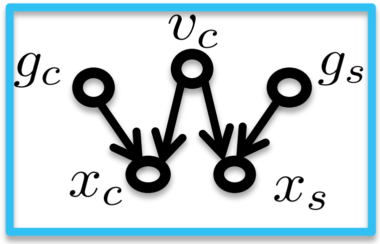
Related to...

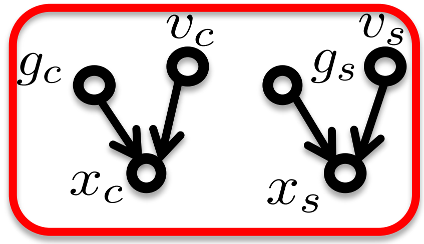
• Related to divisive normalization bottom-up approach (via estimating the Gaussian)

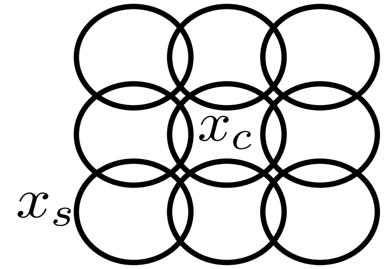
• But allows richer characterization (linear dependencies and non homogeneity of scenes)

• Related to hierarchical ICA (eg, in Karklin and Lewicki mixer is given by a linear combination of small number of underlying components).

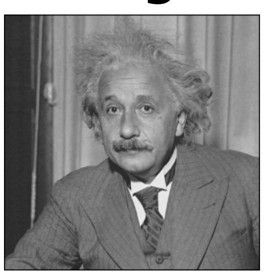
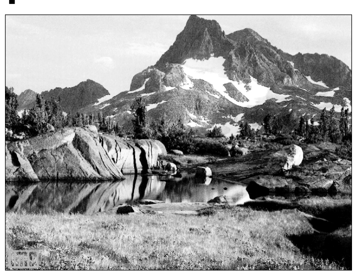
Optimizing Scene Ensemble





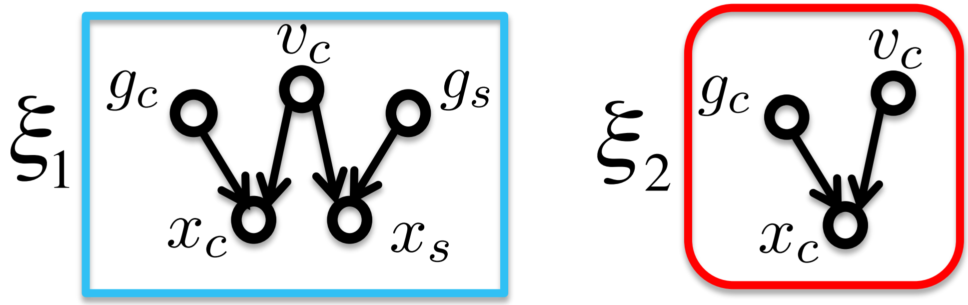


- 3x3 spatial positions, 6px separation
- 4 orientations in the center
- 4 orientations in the surround
- 2 phases (quadrature)
- model parameters (prior probability ξ_1, ξ_2 and also linear covariance matrices) optimized to maximize the likelihood of a database of natural images using Expectation Maximization

Coen-Cagli, Dayan, Schwartz, 2012

Optimizing scene ensemble



$$p(x) = p(\xi_1)\,p(x \mid \xi_1) + p(\xi_2)\,p(x \mid \xi_2)$$

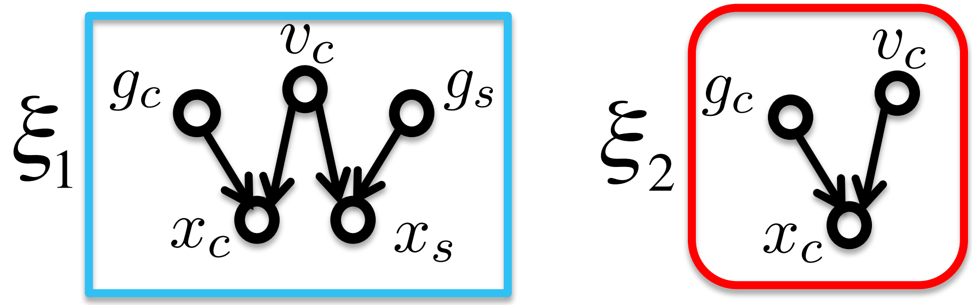
We would like to find the model parameters from an ensemble of images:

(1) The prior probablilities $p(\xi_1)$ and $p(\xi_2) = 1 - p(\xi_1)$

(2) The linear covariance matrices

Coen-Cagli, Dayan, Schwartz, 2012

Optimizing scene ensemble



$$p(x) = p(\xi_1)\,p(x \mid \xi_1) + p(\xi_2)\,p(x \mid \xi_2)$$

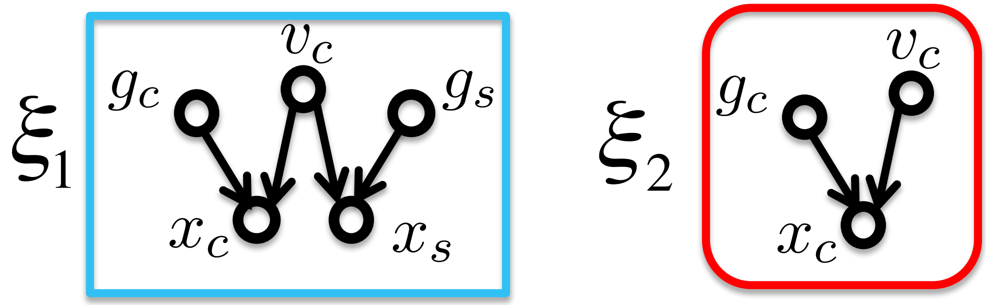
We maximize the average log likelihood of the observed filter responses with respect to the parameters of the model:

(1) The prior probablilities $p(\xi_1)$ and $p(\xi_2) = 1 - p(\xi_1)$

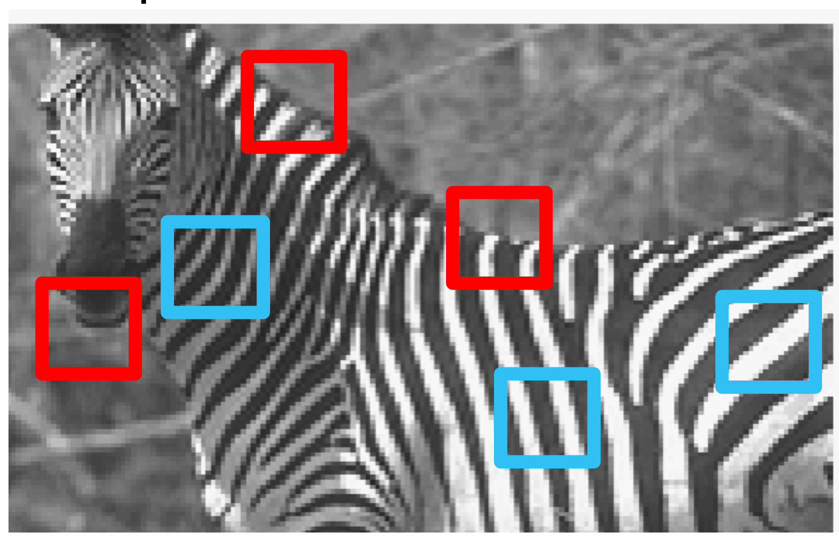
(2) The linear covariance matrices

Coen-Cagli, Dayan, Schwartz, 2012

Optimizing scene ensemble

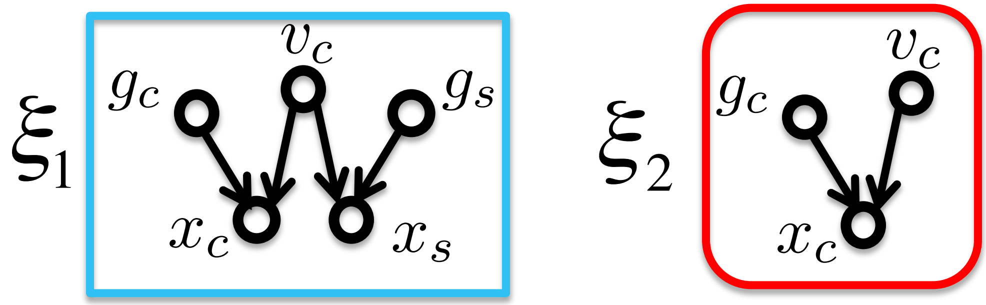


If we knew the posterior assignment probabilities $p(\xi_1 \mid x); p(\xi_2 \mid x)$
(which assignment group each patch belonged to)
we could find the model parameters…

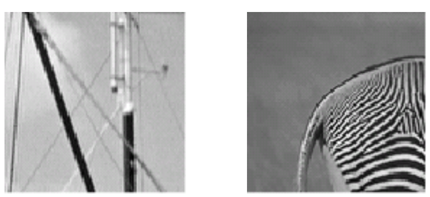


Coen-Cagli, Dayan, Schwartz, 2012

Optimizing scene ensemble



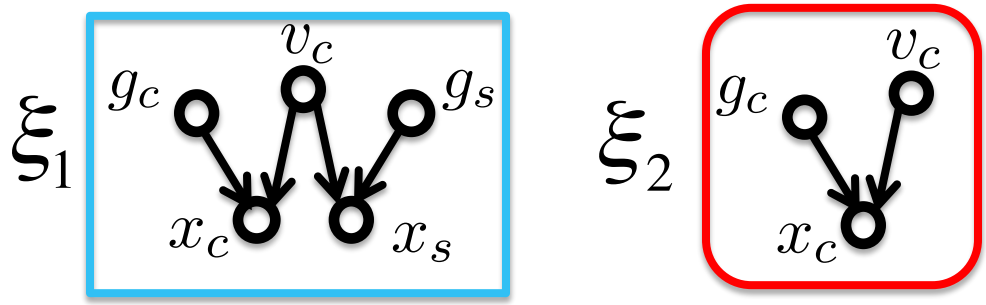
If we knew the the parameters of the model, we could calculate the posterior assignments $p(\xi_1 \mid x); p(\xi_2 \mid x)$ for each patch and its filter activations



...

Coen-Cagli, Dayan, Schwartz, 2012

Optimizing scene ensemble



Expectation Maximization:

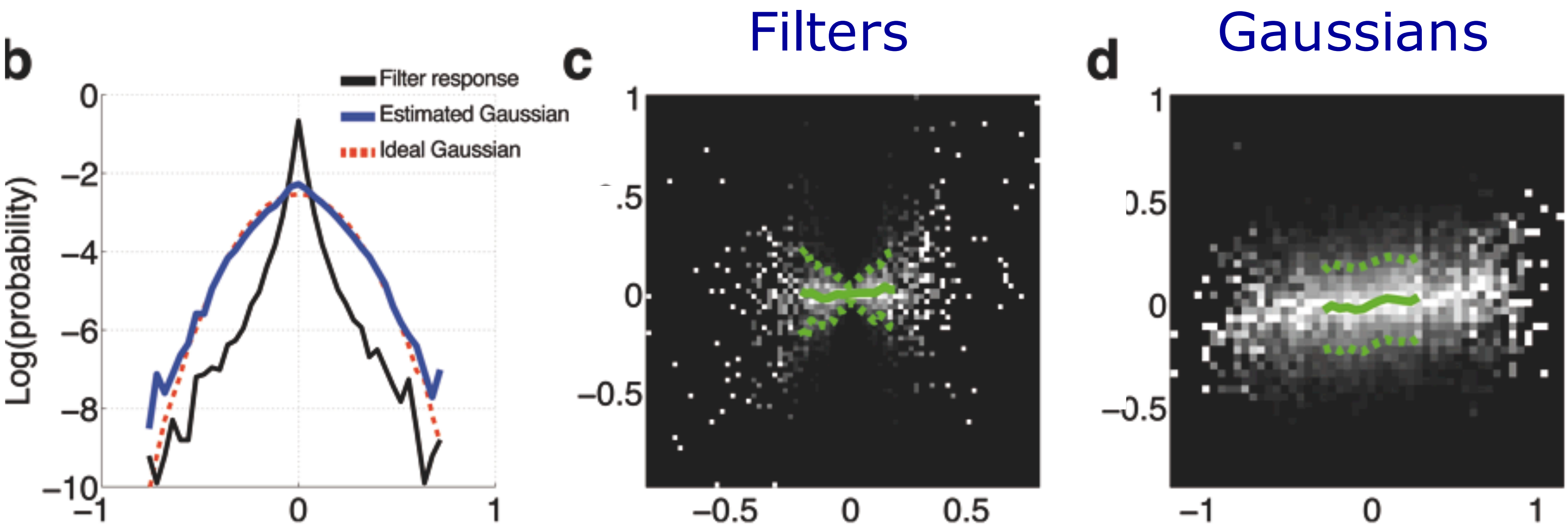
E-step: calculate $p(\xi_1 | x); p(\xi_2 | x)$ given current estimate of model parameters

M-step: update the model parameters by maximizing the log likelihood

Optimizing scene ensemble

When estimating a statistical model, need to go back and check that model estimates match statistical assumptions…

Optimizing scene ensemble



Coen-Cagli, Dayan, Schwartz, 2012

Optimizing scene ensemble

Learned linear correlations (covariance)

 Coen-Cagli, Dayan, Schwartz, 2012

Outline

- Generative that captures the joint statistics and its relation to cortical neural processing

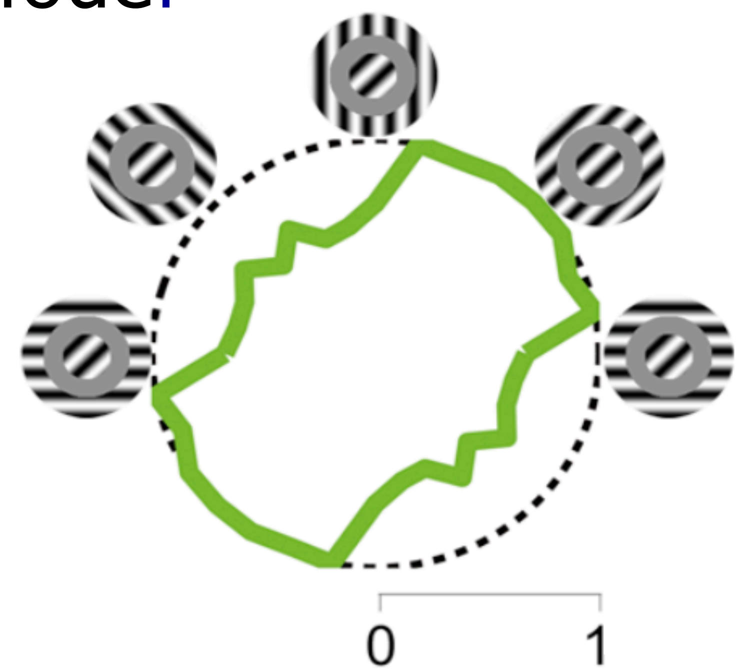
- Application and testing on biological data

Application to biological data

Neurophysiology

Gratings: Examples

Model

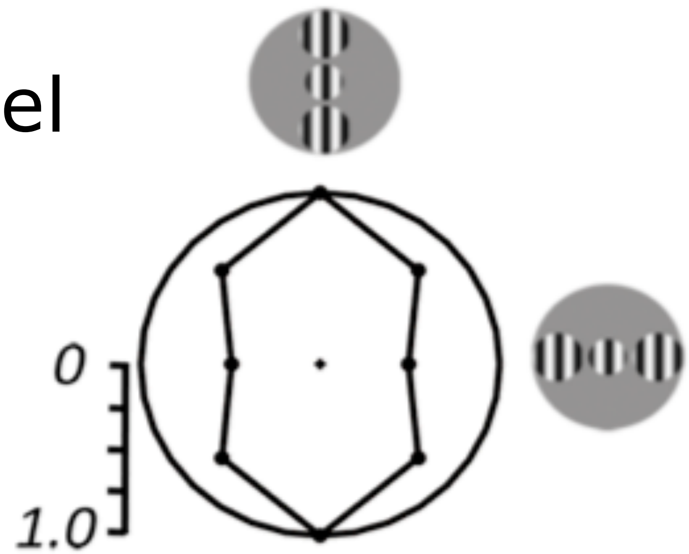


Probability center and surround dependent

Coen-Cagli, Dayan, Schwartz, 2012

Gratings: Examples

Model

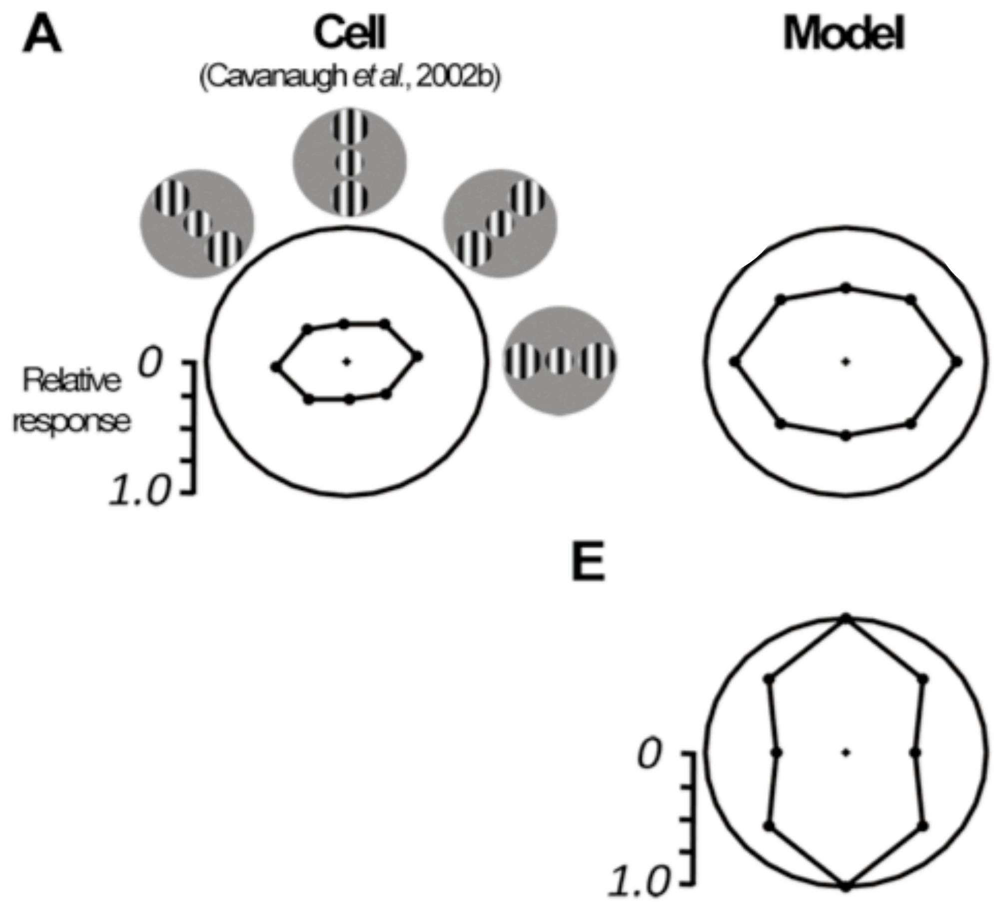


Probability center and surround dependent

Coen-Cagli, Dayan, Schwartz, 2012

Gratings: Examples

Cortical neurons



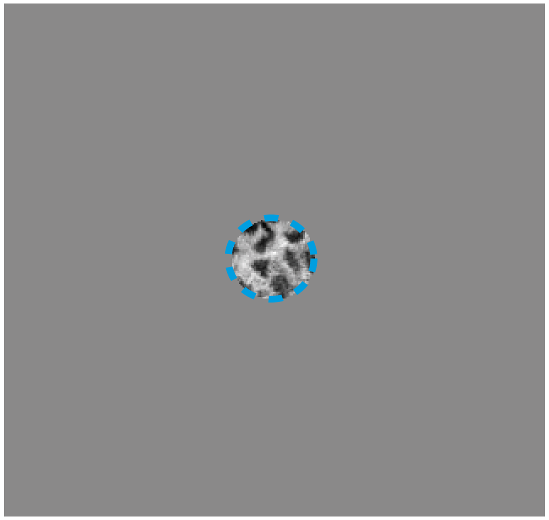
Coen-Cagli, Dayan, Schwartz, 2012

Application to data

Simple stimuli: spatial context effects (suppression and facilitation) as a function of orientation, size, spatial position, and contrast.

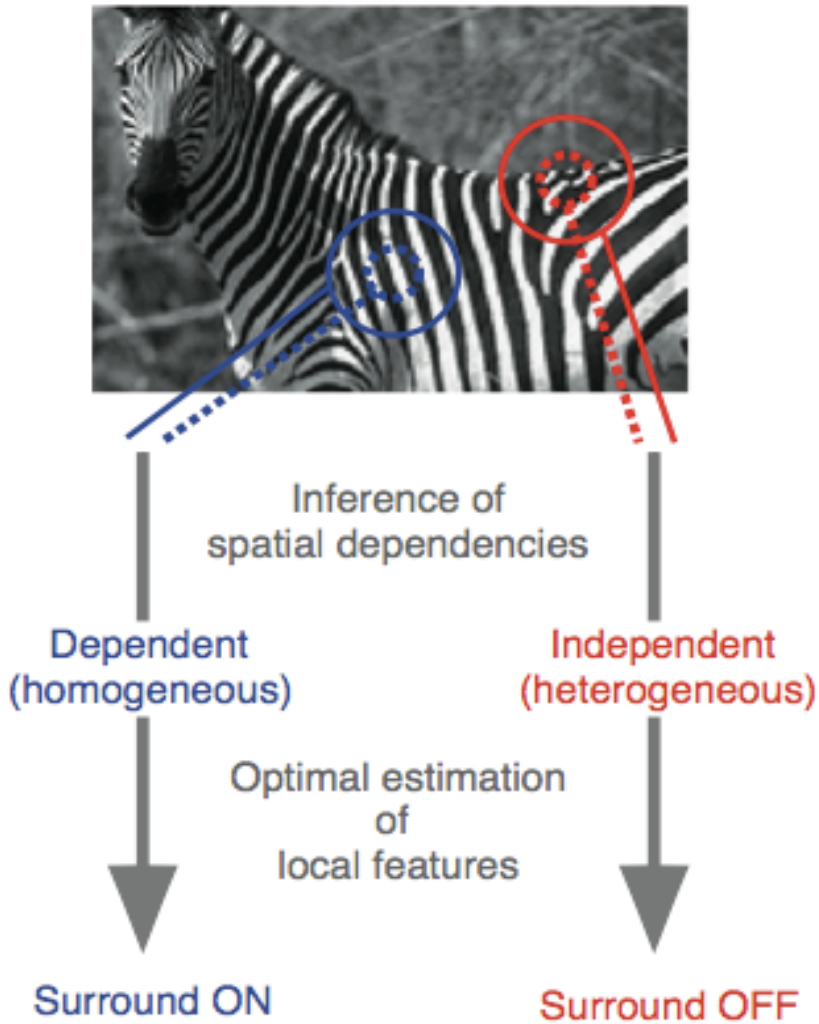
Model Makes Predictions

Natural images:

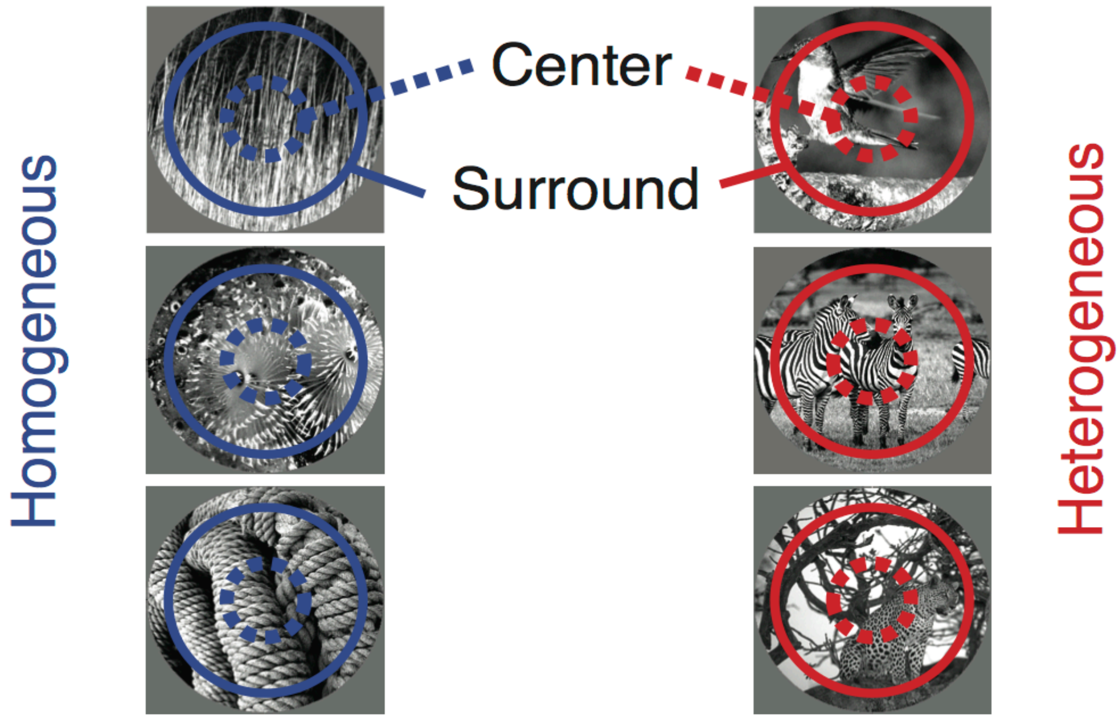


Coen-Cagli, Kohn, Schwartz, 2015

Flexible Divisive Normalization



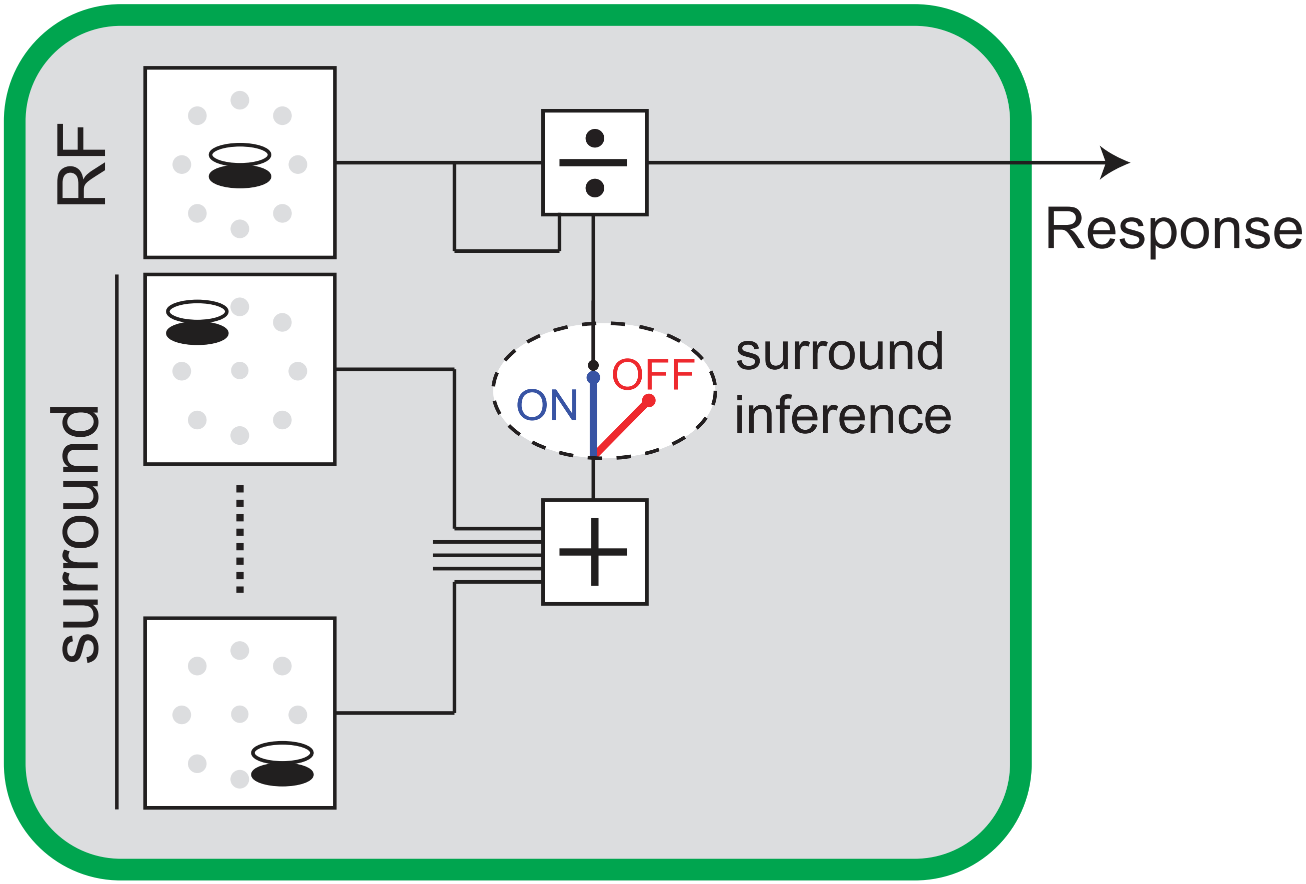
Model predictions for natural images



- **Homogeneous and heterogeneous determined by model!**
- Expect more suppression in neurons for homogeneous
- Related to salience (eg, Zhaoping)

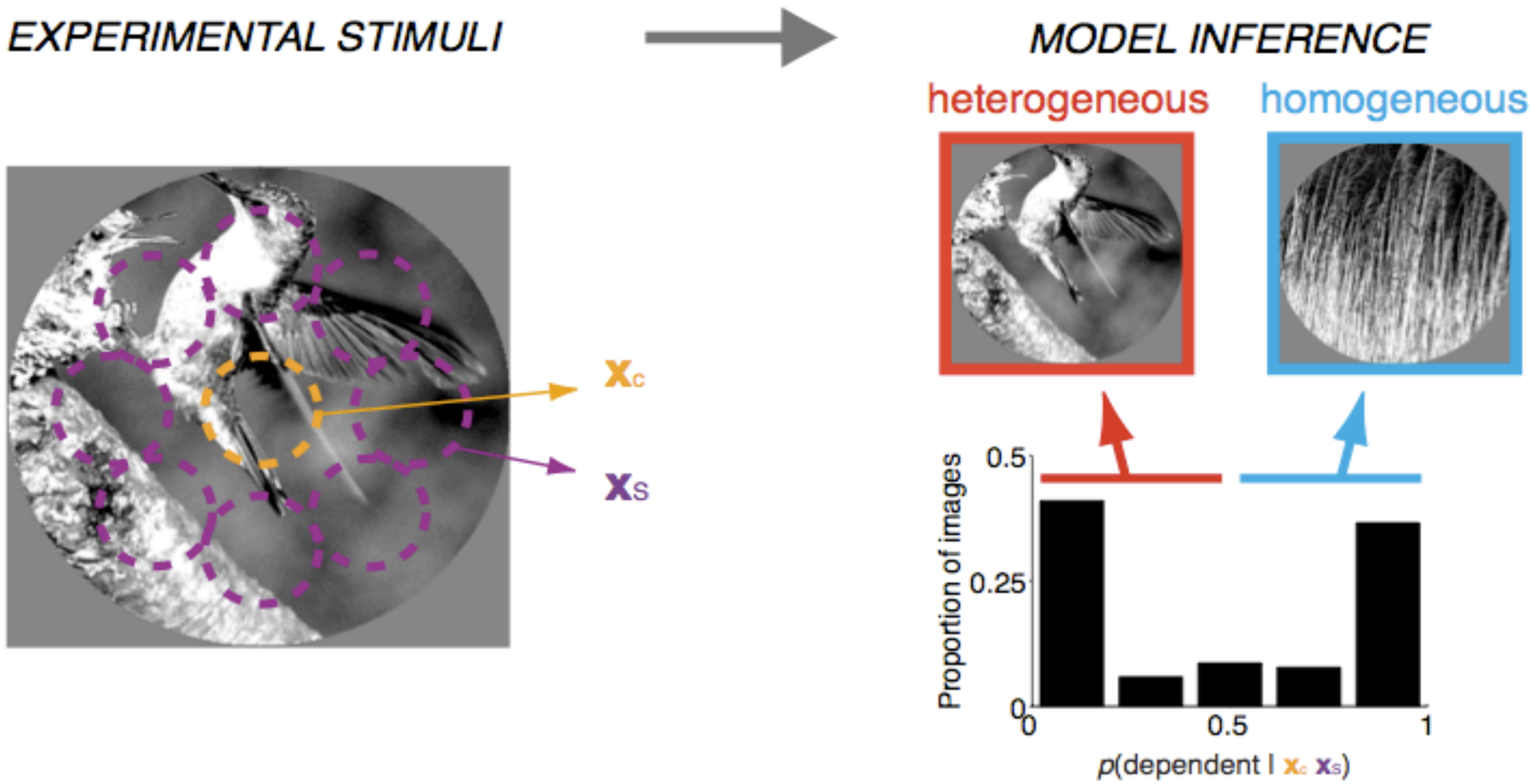
 Coen-Cagli, Kohn, Schwartz, 2015; in review

Model summary



Inference determined by model

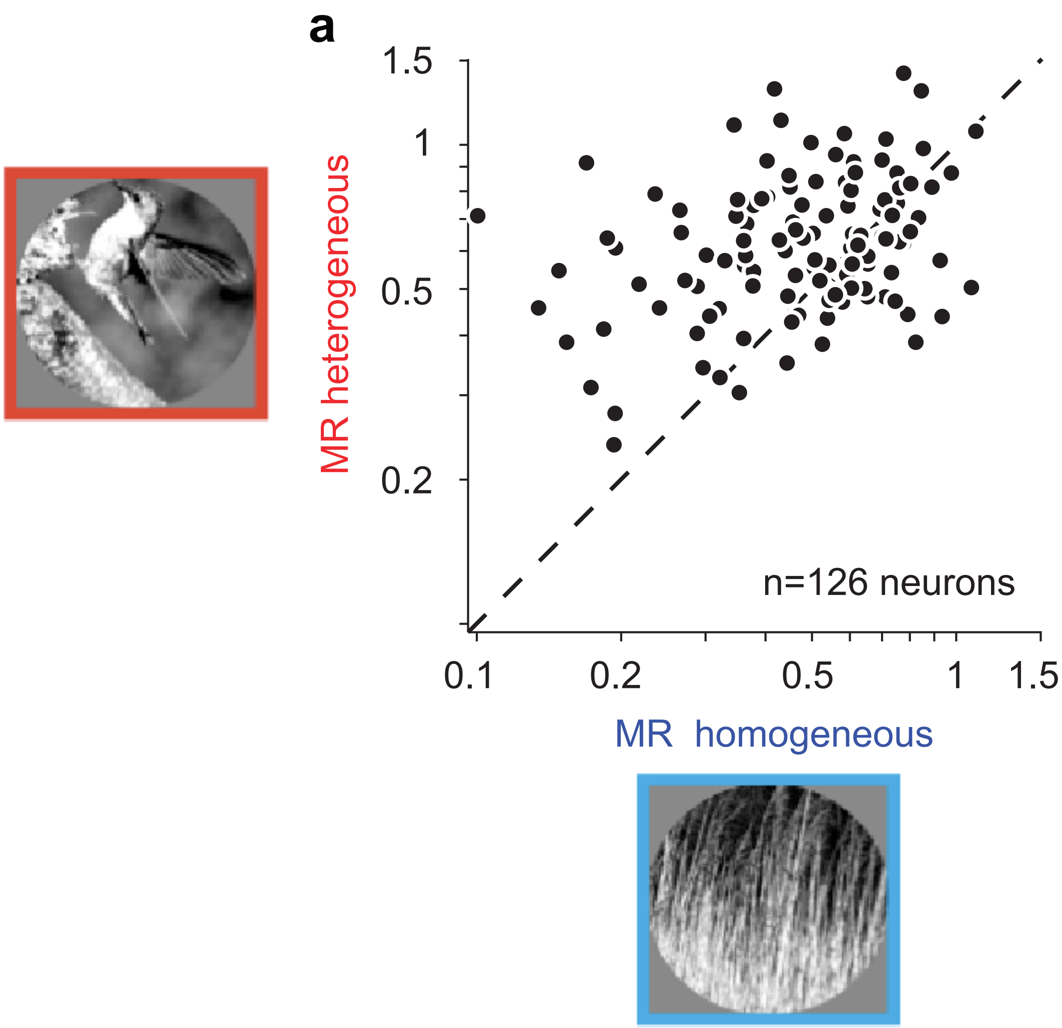
Model Predictions for Natural Scenes



homogeneous versus heterogeneous determined by the model

Model Predictions for Natural Scenes

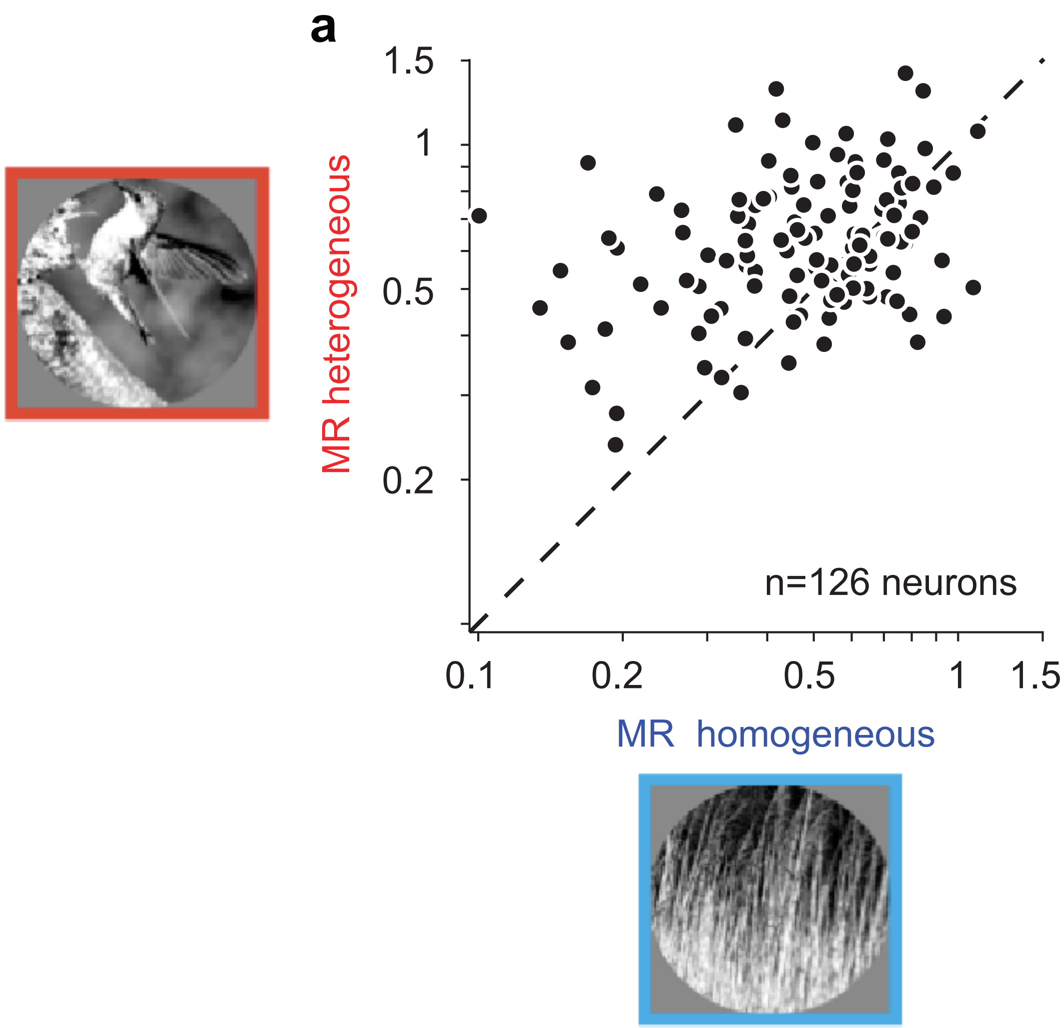
Cortical V1 data:



Coen-Cagli, Kohn, Schwartz, 2015

Model Predictions for Natural Scenes

Cortical V1 data:

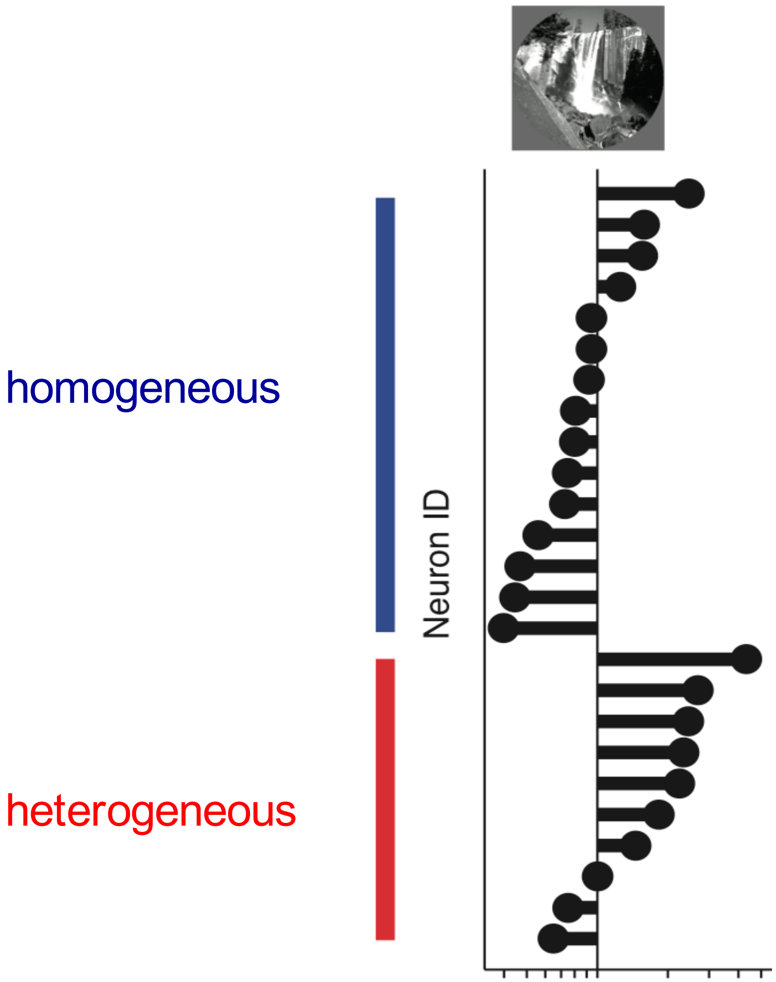


Not explained by:
- firing rate with small frames
- surround energy

Coen-Cagli, Kohn, Schwartz, 2015

Model predictions for natural images

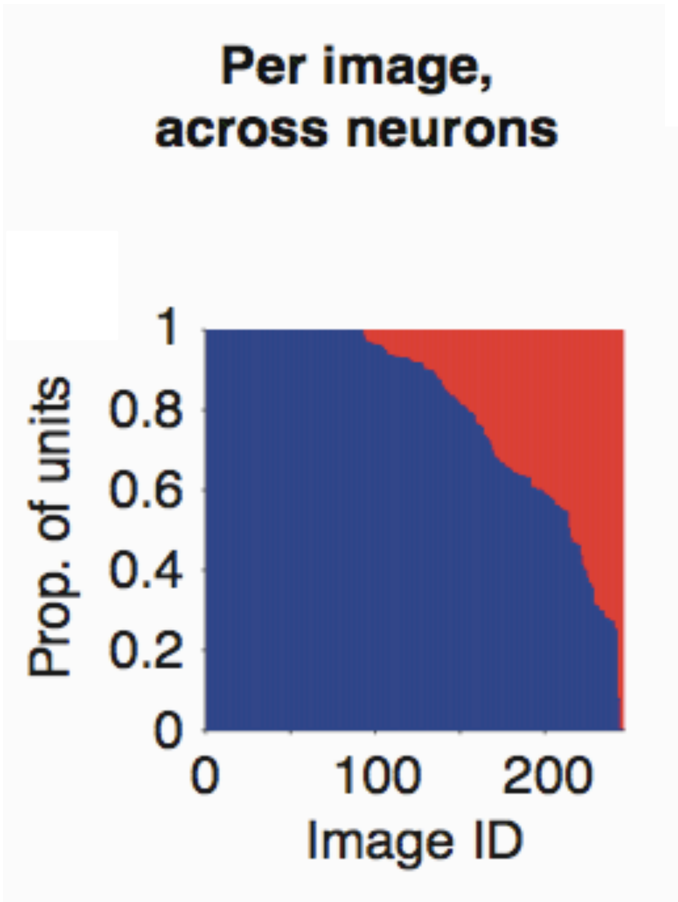
- Per image, across neurons

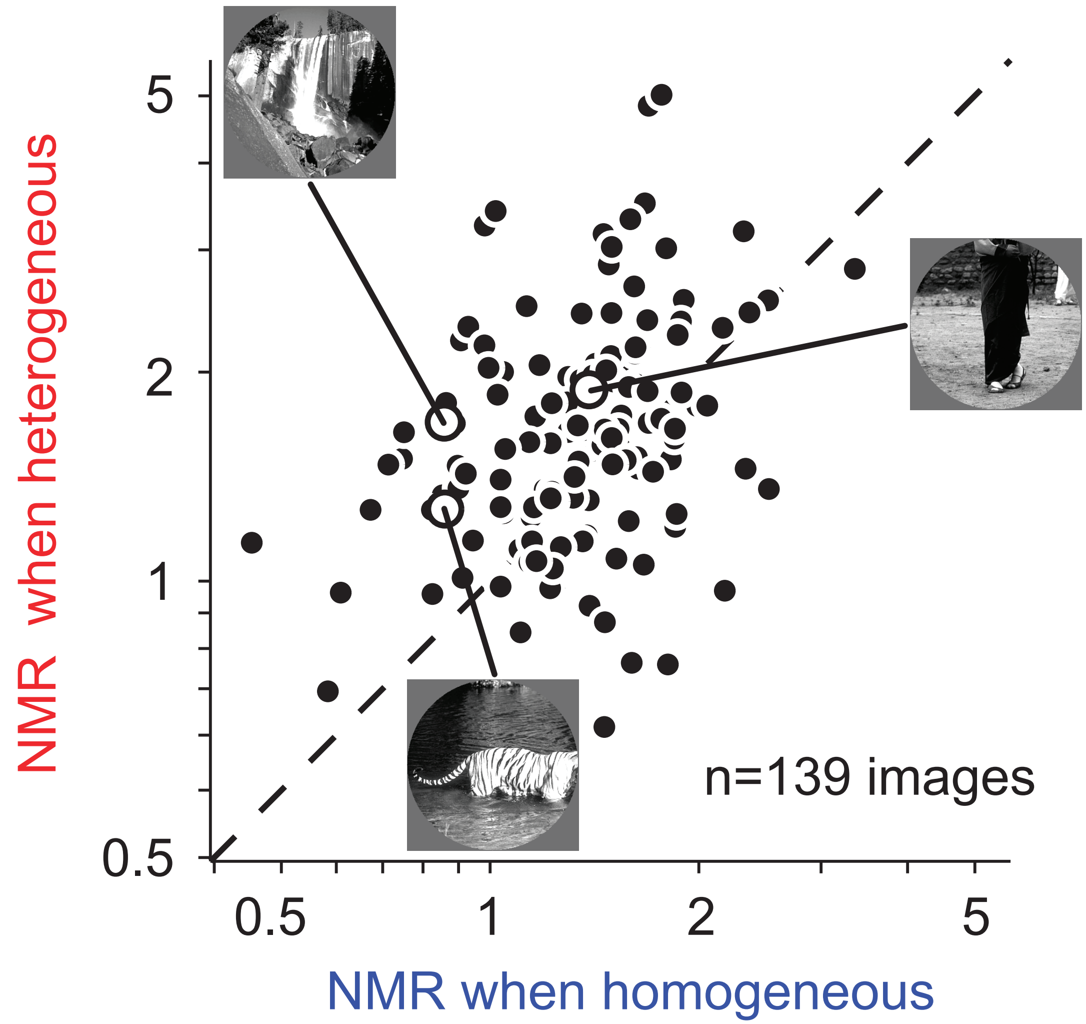


99 Coen-Cagli, Kohn, Schwartz, 2015

Model predictions for natural images

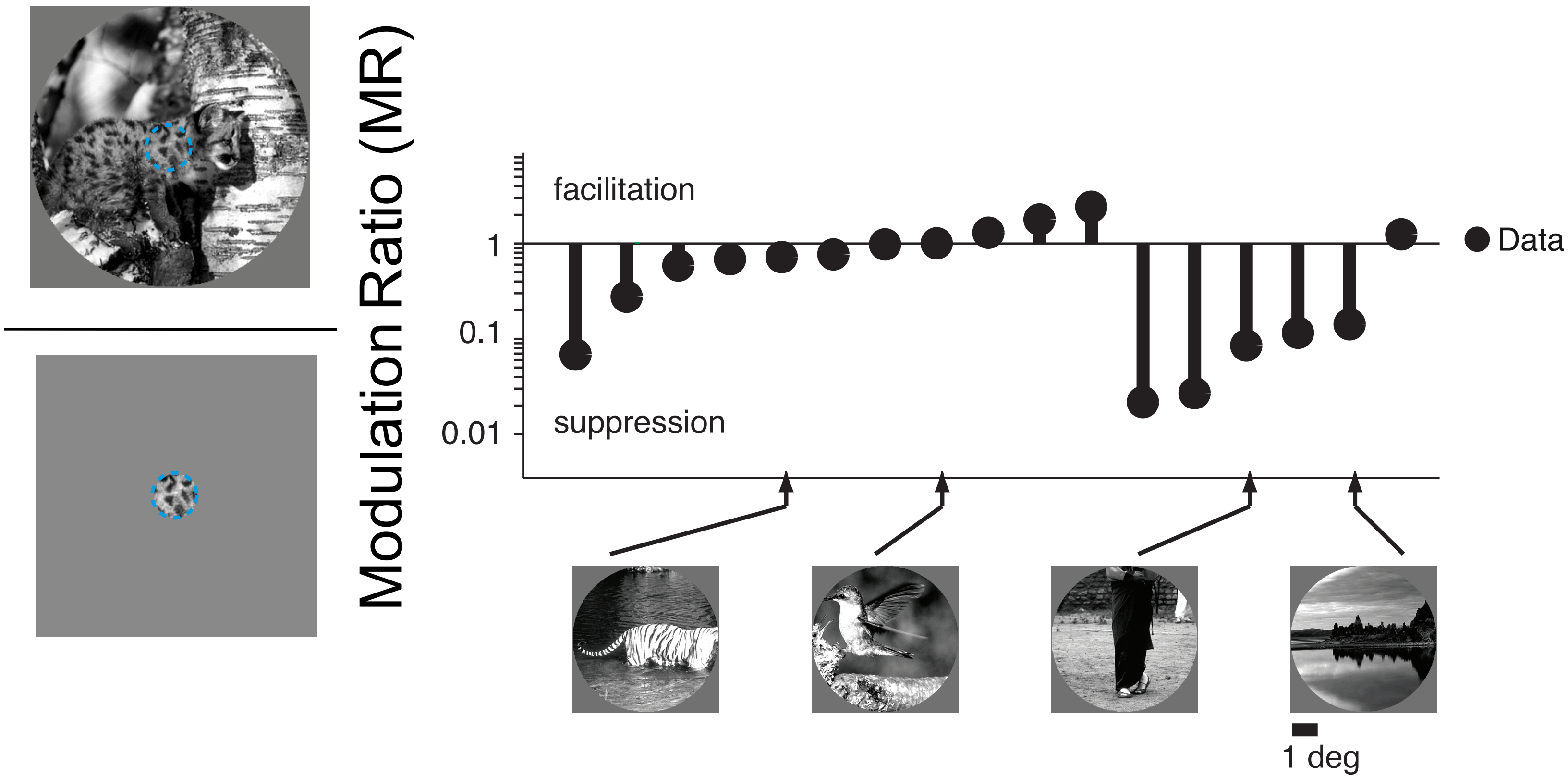
- Testing predictions with cortical data





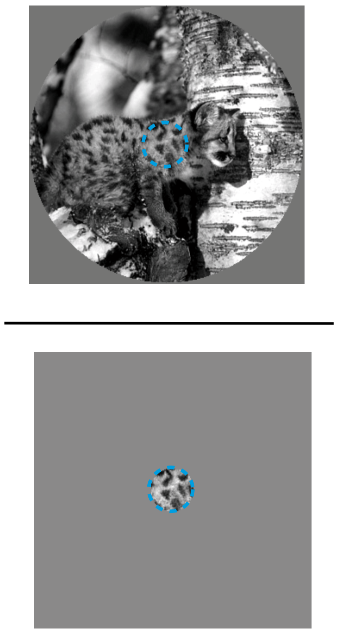
Coen-Cagli, Kohn, Schwartz, 2015

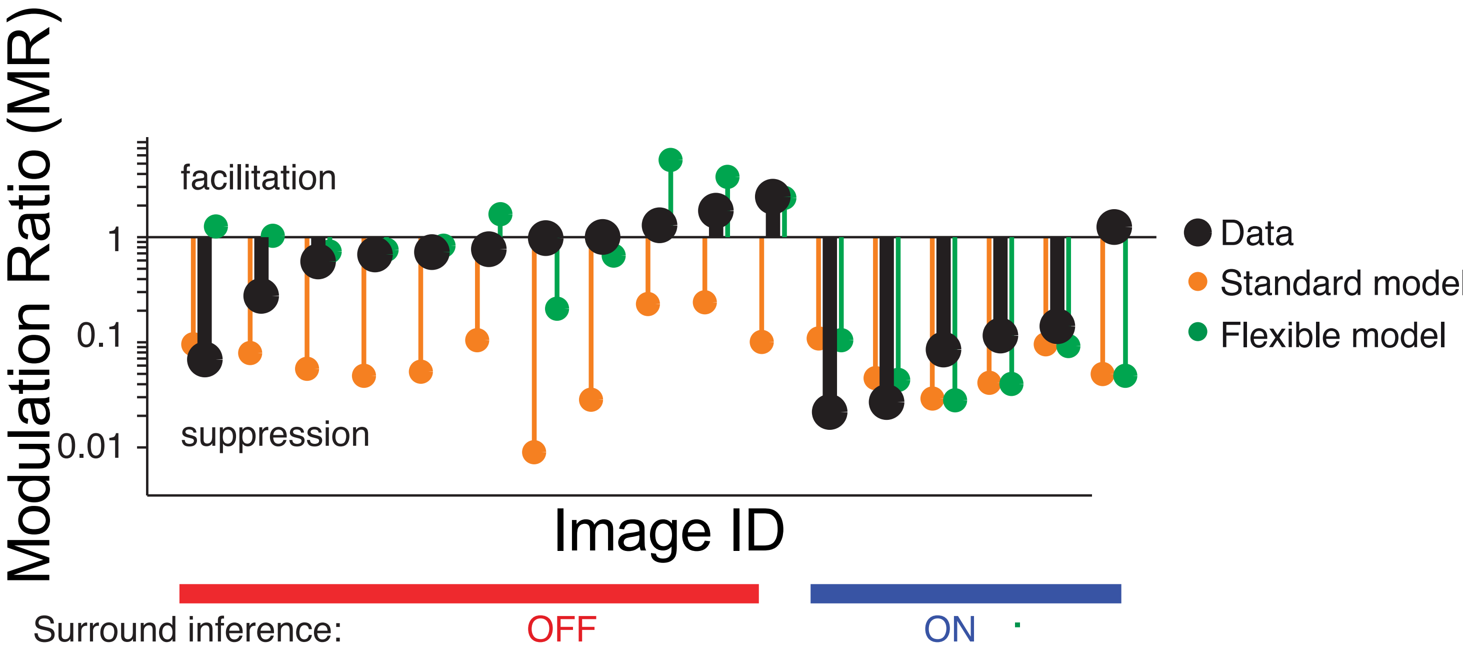
Natural scenes data



Coen-Cagli, Kohn, Schwartz, 2015

Natural scenes data





Coen-Cagli, Kohn, Schwartz, 2015

Model fit for Natural Scenes

$$R_i = \alpha \left(\frac{E_{c,\phi_{pref}}}{\varepsilon + \beta E_c + \gamma E_s} \right)^n$$

$R_i - O_i$ Gaussian distribution

$$X^2 = \sum_i \frac{(R_i - O_i)^2}{\sigma_i^2}$$

O_i Observed mean firing rate

$$\sigma_i^2 = a O_i^b$$

Fit to the data

Minimize chi-squared error

Model predictions for natural images

- Comparing model performance for cortical data

Standard divisive normalization

$$R_i = \alpha \left(\frac{E_{c,\phi_{pref}}}{\varepsilon + \beta E_c + \gamma E_s} \right)^n$$

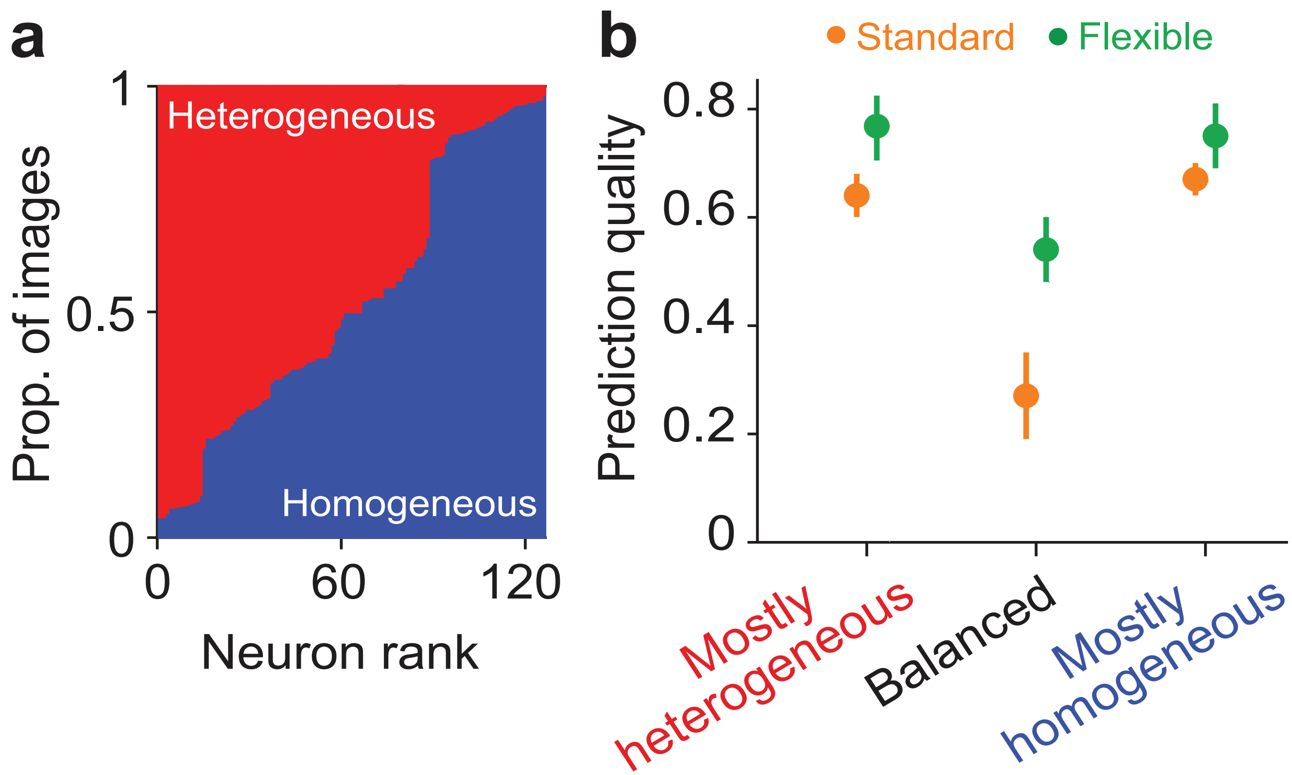
Flexible divisive normalization:

$$R_i = \alpha \left(\frac{E_{c,\phi_{pref}}}{\varepsilon + \beta E_c + \boxed{q(c,s)} \gamma E_s} \right)^n$$

Determined by the model (not fit!)
1 if $p(\xi_1 | c,s) \geq 0.5$
0 otherwise

(similar results if non binary)

Natural scenes data



Coen-Cagli, Kohn, Schwartz, 2015

Model Mechanisms

Divisive normalization:

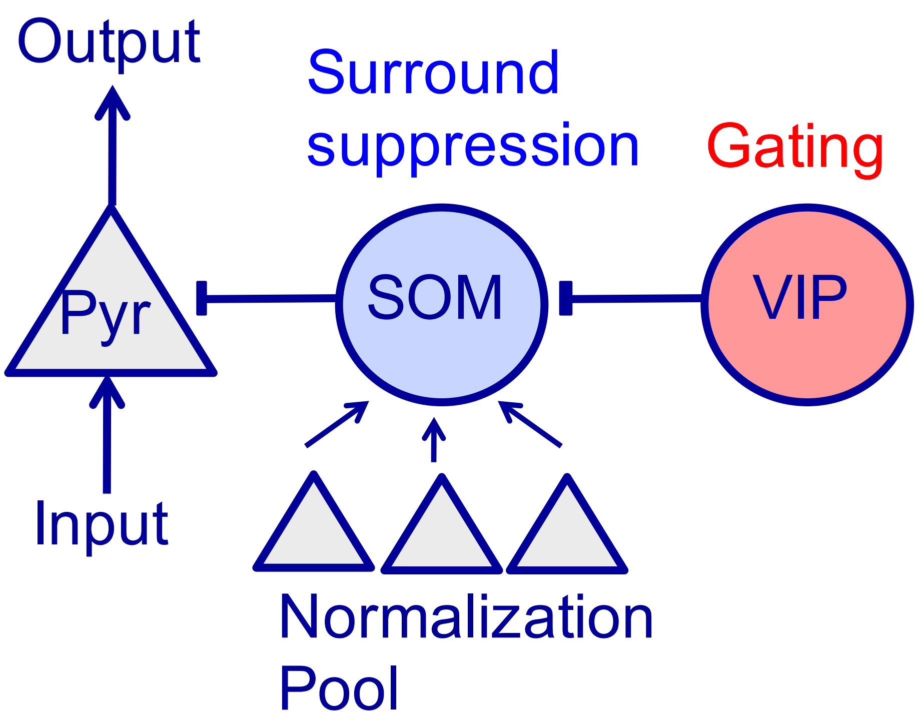
- Feedback inhibition
- Distal dendrite inhibition
- Depressing synapses
- Internal biochemical adjustments
- Non-Poisson spike generation

Flexible normalization pools, possibly:

- Inhibitory interneurons pooling outputs of subpopulations
- Stimulus-dependent switching between cortical network states

Flexible Normalization Mechanism?

- Adjusting gain by circuit or postsynaptic mechanisms?
- Distinct classes of inhibitory interneurons? (eg, Adesnik, Scanziani et al. 2012; Pfeffer, Scanziani et al. 2013; Pi, Kepecs et al. 2013; Lee, Rudy et al. 2013)



Key take-home points

•New approach to understanding cortical processing of natural images. Rather than fitting more complicated models, use insights from image statistics

•Connects to neural computations that are ubiquitous, but enriches the “standard” model

•Our results suggest flexibility of contextual influences in natural vision, depending on whether center and surround are deemed statistically homogeneous—a kind of gating

Generative models summary

- Progress in building generative probabilistic models of scenes (e.g., based on joint statistics)

- Generative and bottom-up can be closely tied

- Applied largely to building more complex visual representations and to contextual effects, and addressing nonlinearities such as complex cells and divisive normalization in a richer way

- Testing statistical models in experiments with naturalistic stimuli (using scene statistics to enrich models; going beyond descriptive models)

- Next: Deep (discriminative, supervised) learning