Statistical models of the early visual system

Yan Karklin

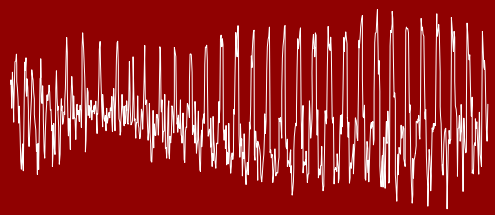
Center for the Neural Basis of Cognition / Computer Science Dept.
Carnegie Mellon University

joint work with Mike Lewicki

DOE Conf. june 22 2006.

computation in the visual system

- complex tasks
- computation is fast, reliable, robust to noise

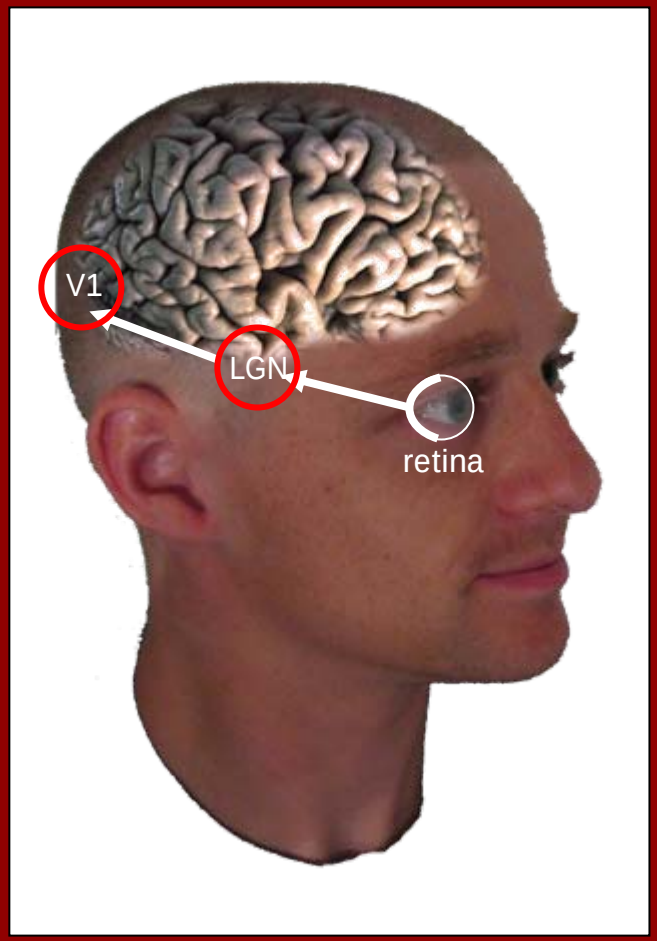


pattern of light on the retina

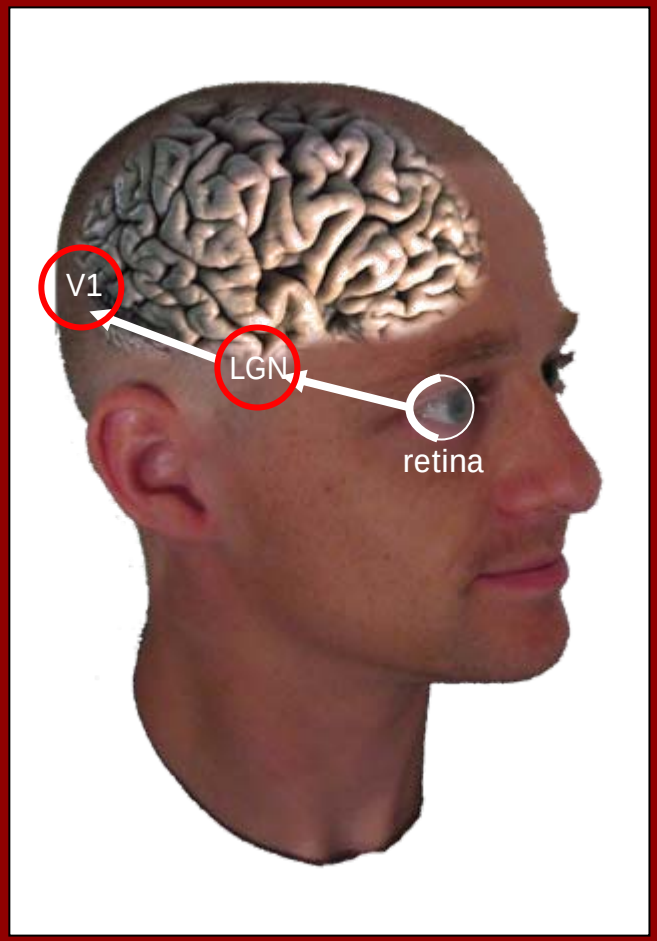
machines don’t even come close!

how does it work?

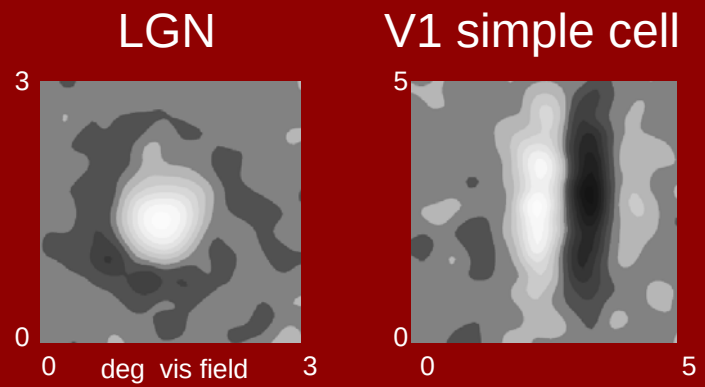
- early visual system (retina, LGN)



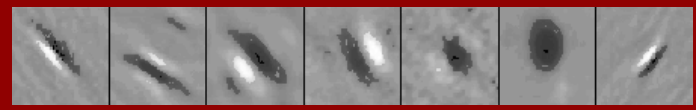
how does it work?



- early visual system (retina, LGN)



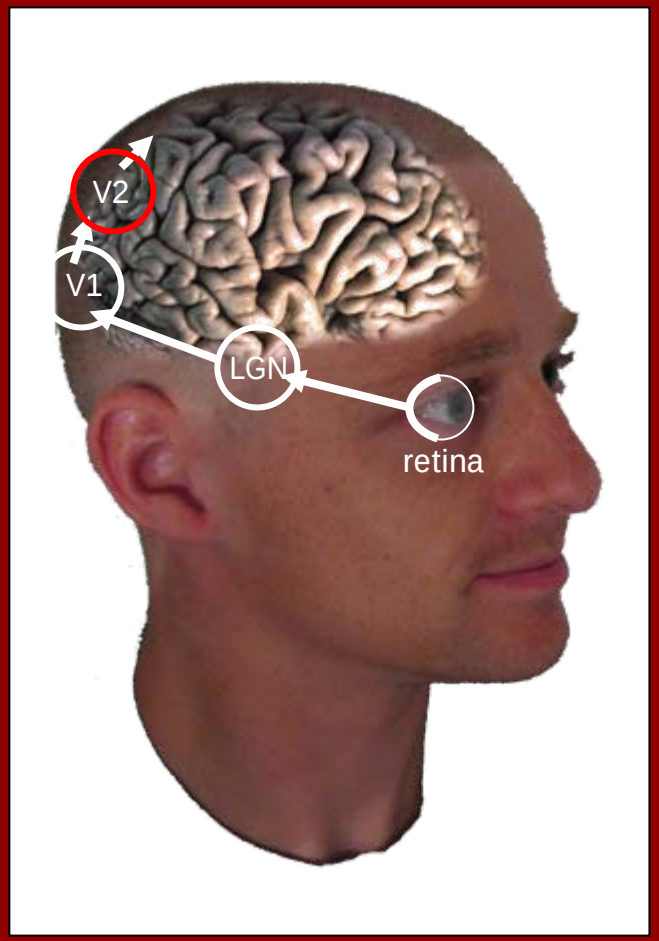
(DeAngelis, Ohzawa, Freeman 95)



(Ringach 2002)

- linear responses
- it's possible to fully characterize neural behavior

how does it work?



- early visual system (retina, LGN)
  - linear responses
  - it's possible to fully characterize neural behavior
- further along the pathway

V2

(Hegde and Van Essen 00)

  - complex, non-linear responses
  - increasingly invariant to transformations
  - difficult to map out the response
  - **need theory!**

a theoretical approach

computational goal:

assume the brain has evolved to process input as efficiently as possible

"efficient coding hypothesis" (Attneave 1954, Barlow 1961)

- the brain is adapted specifically to process natural stimuli
- natural images are full of redundancies
- first stage of visual processing: recode the image **efficiently**

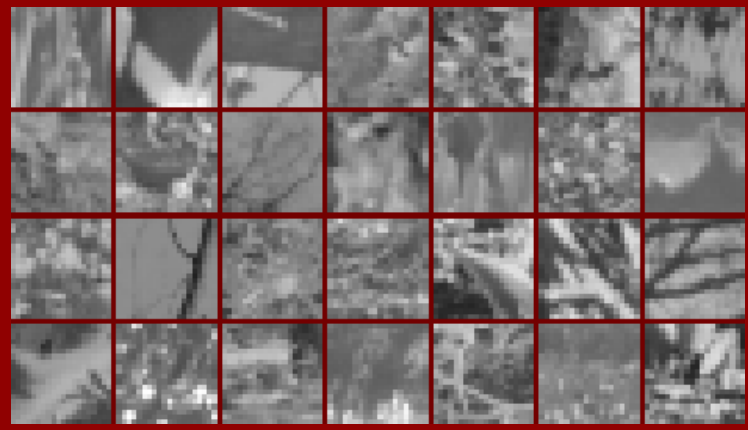
what kind of code is efficient?

*optimal codes depend on good underlying **statistical models***

(Shannon 1948)

- reduce redundancy by encoding the **independent** elements of the scene

statistics of natural scenes

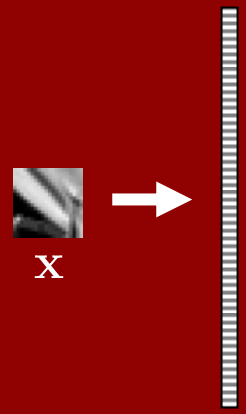


data

- small image patches
- sampled from natural images

notation

- image patches as vectors of pixel intensities (**x**)

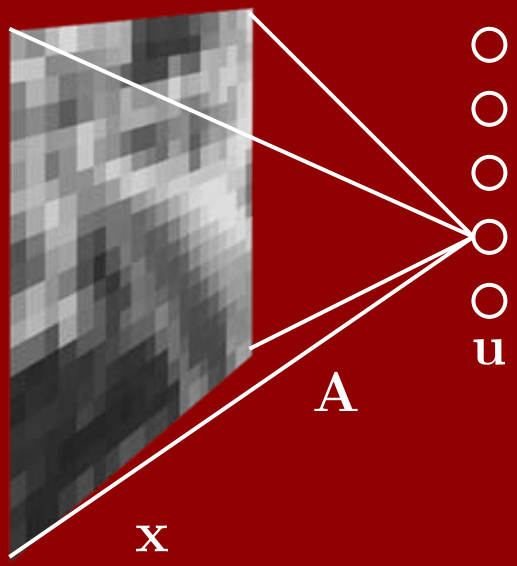


goal

- model the distribution of the data

$$P(\mathbf{x}) = ?$$

a linear model



a simple, scalable model

$$x_i = \sum_j A_{ij} u_j$$

independent coefficients

$$P(\mathbf{u}) = \prod_j P(u_j)$$

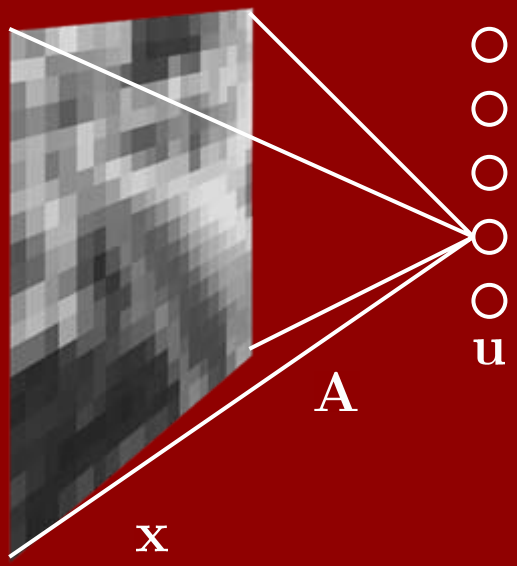
simple, but non-Gaussian probability densities $P(u_j)$

estimate parameters that maximize data likelihood

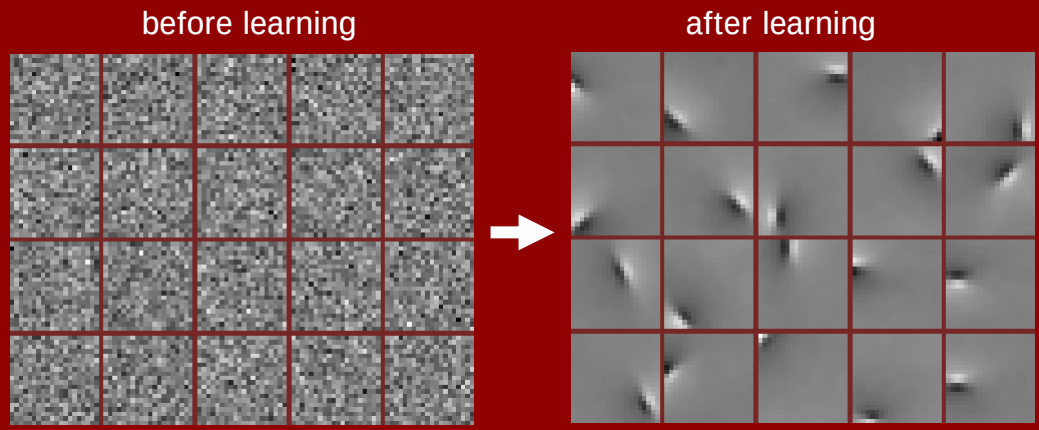
$$\mathbf{A}^* = \arg\max_{\mathbf{A}} P(\mathbf{x}|\mathbf{A})$$

Independent Component Analysis
(Bell and Sejnowski 1995)
Sparse Coding
(Olshausen and Field 1996)

a linear model: results



parameters **A** trained on a large collection of image patches:



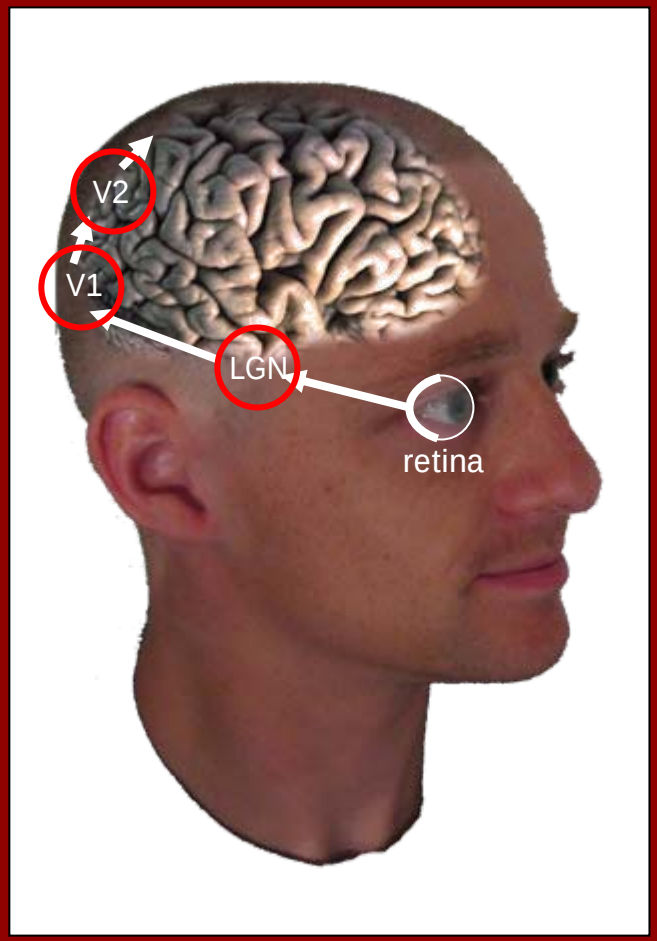
(Olshausen and Field 1996, Bell and Sejnowski 1997)

receptive fields of V1 simple cells

(Ringach 2002)

- theoretical predictions match experimental results
- provide a *functional* explanation for observed properties of visual neurons
- allow analysis of population properties

a linear model: limitations



explained V1 simple cell properties...

what about later stages?

linear model, assumed independent coefficients **u**

$$x_i = \sum_j A_{ij} u_j$$

$$P(\mathbf{u}) = \prod_j P(u_j)$$

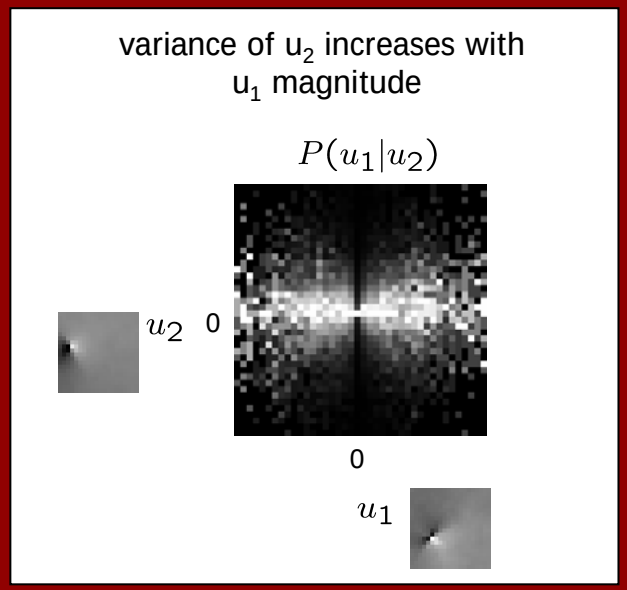
is this a good enough model of the data?

a linear model: limitations

$$x_i = \sum_j A_{ij} u_j$$

$$P(\mathbf{u}) = \prod_j P(u_j)$$

higher-order structure (dependence not captured by model)

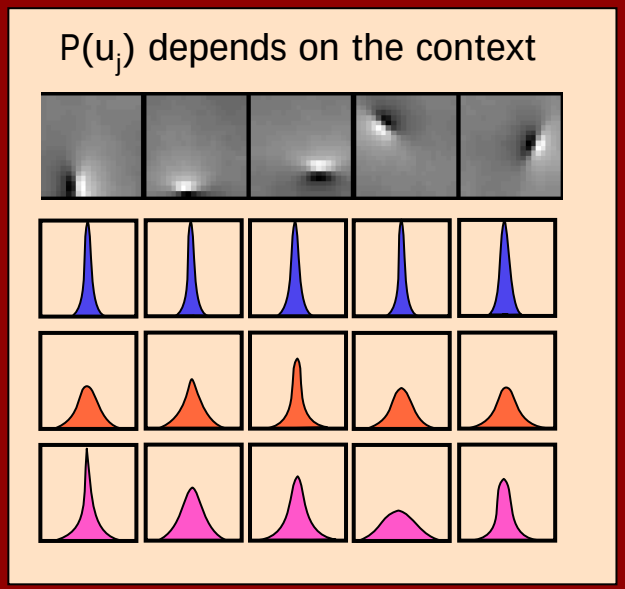


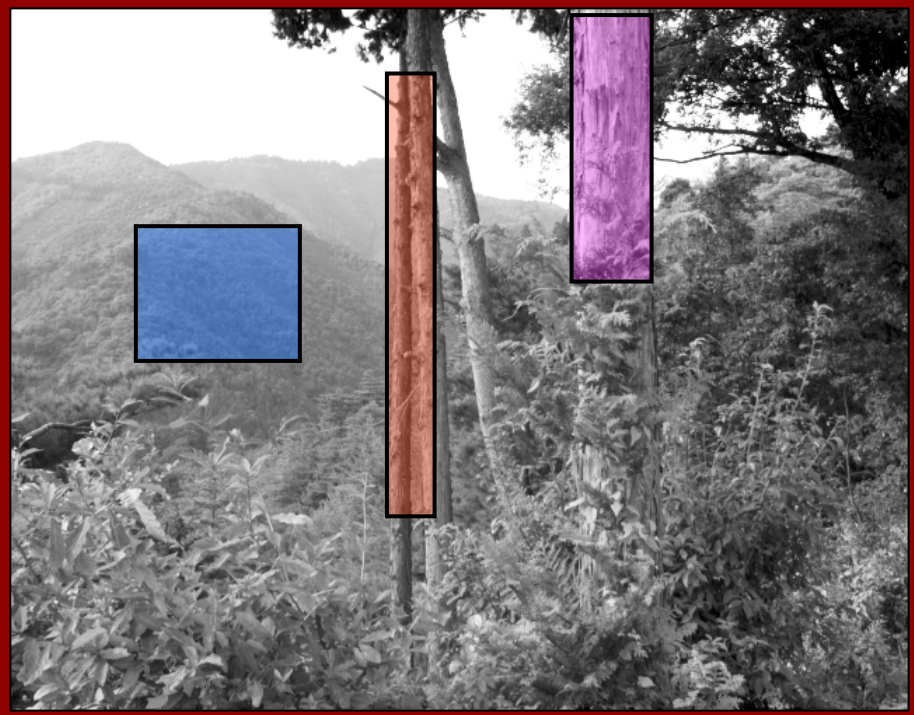
a linear model: limitations

$$x_i = \sum_j A_{ij} u_j$$

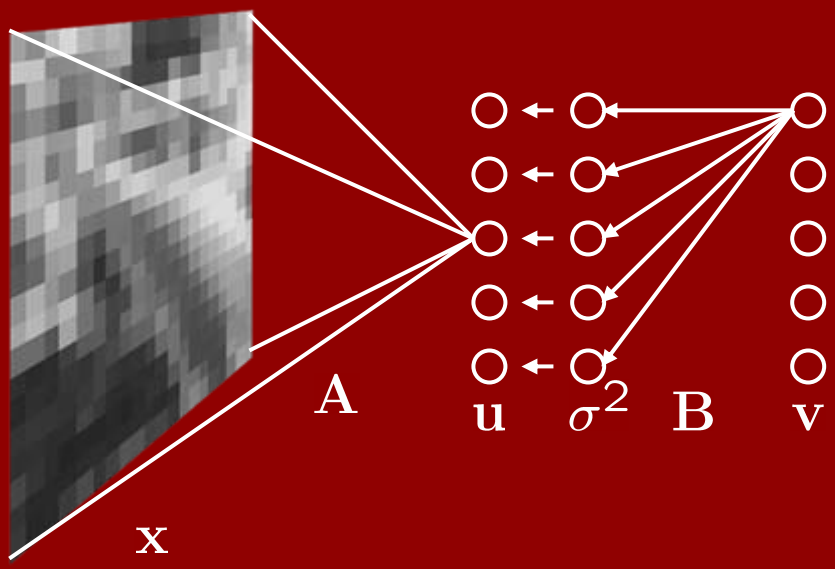
$$P(\mathbf{u}) = \prod_j P(u_j)$$

non-stationarity





solution: a hierarchical statistical model



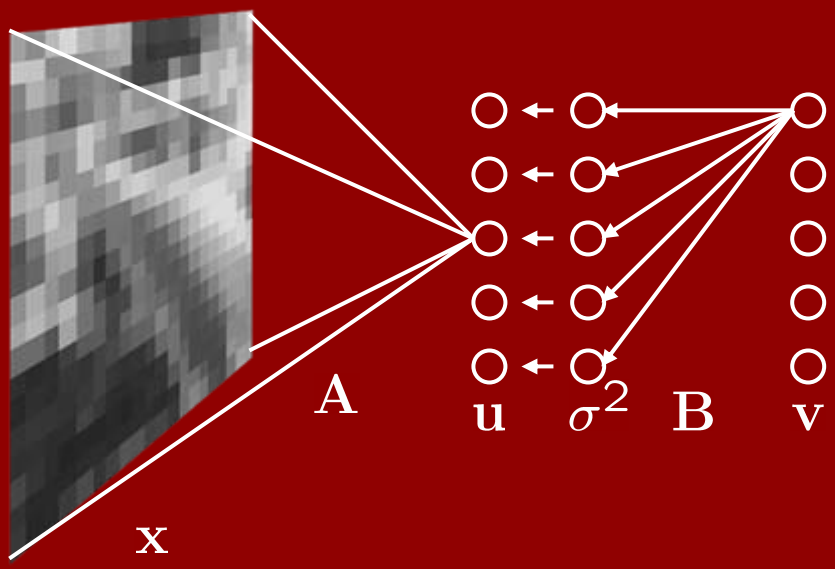
$$P(\mathbf{u}|\mathbf{v}) = \prod_i P(u_i|\mathbf{v})$$

$$P(u_i|\mathbf{v}) = \mathcal{N}(0, \sigma_i^2)$$

$$\log \sigma_i^2 = [\mathbf{B}\mathbf{v}]_i$$

- different distributions P(**u**) depending on **v**
- columns in the matrix **B** represent common variance patterns
- **B** is adapted to the data, captures higher-order image structure

solution: a hierarchical statistical model



$$P(\mathbf{u}|\mathbf{v}) = \prod_i P(u_i|\mathbf{v})$$

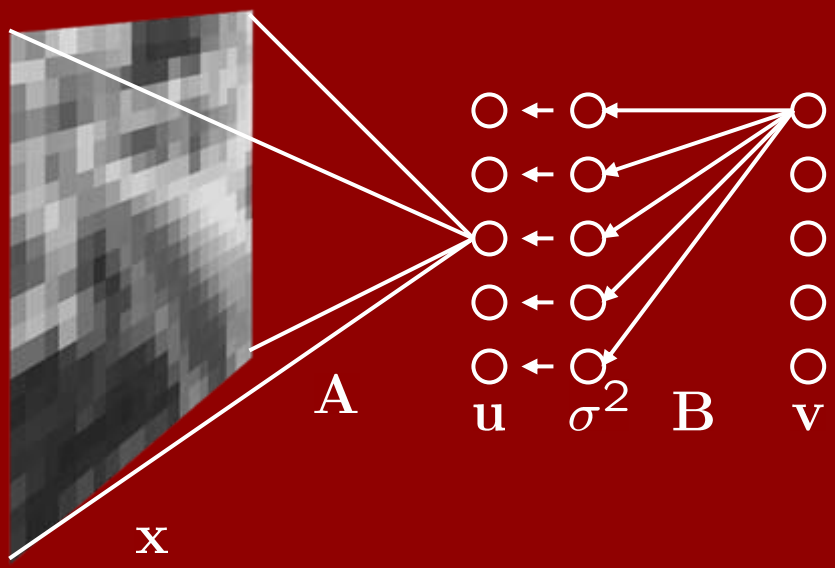
$$P(u_i|\mathbf{v}) = \mathcal{N}(0, \sigma_i^2)$$

$$\log \sigma_i^2 = [\mathbf{B}\mathbf{v}]_i$$

- latent variables: **v**
- model parameters: **A**, **B**
- adapt parameters to maximize data likelihood

$$\mathbf{A}^*, \mathbf{B}^* = \arg\max_{\mathbf{A},\mathbf{B}} P(\mathbf{x}|\mathbf{A}, \mathbf{B})$$

solution: a hierarchical statistical model



model accounts for observed dependencies

- variance of u_i depends on u_j
- $P(u_i)$ depends on context

higher-order representation of image structure

- distributed higher-order code **B**

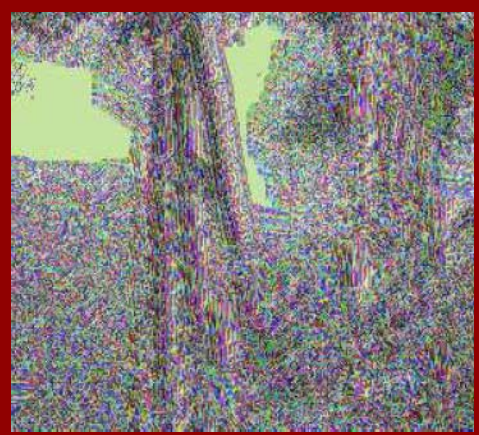
more invariant responses

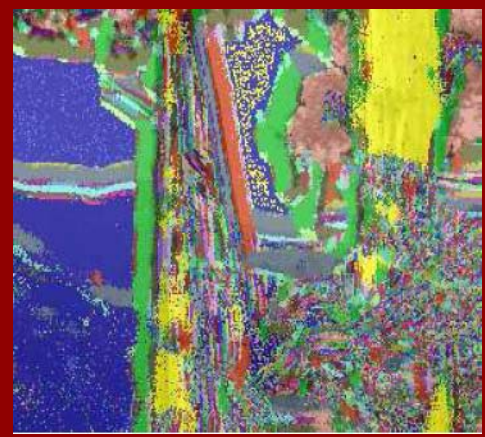
parallels with physiology

- normalization
- invariance

maximally active u_i

maximally active v_j





thank you to DOE and Krell Inst. for CSGF