The Abdus Salam International Centre for Theoretical Physics

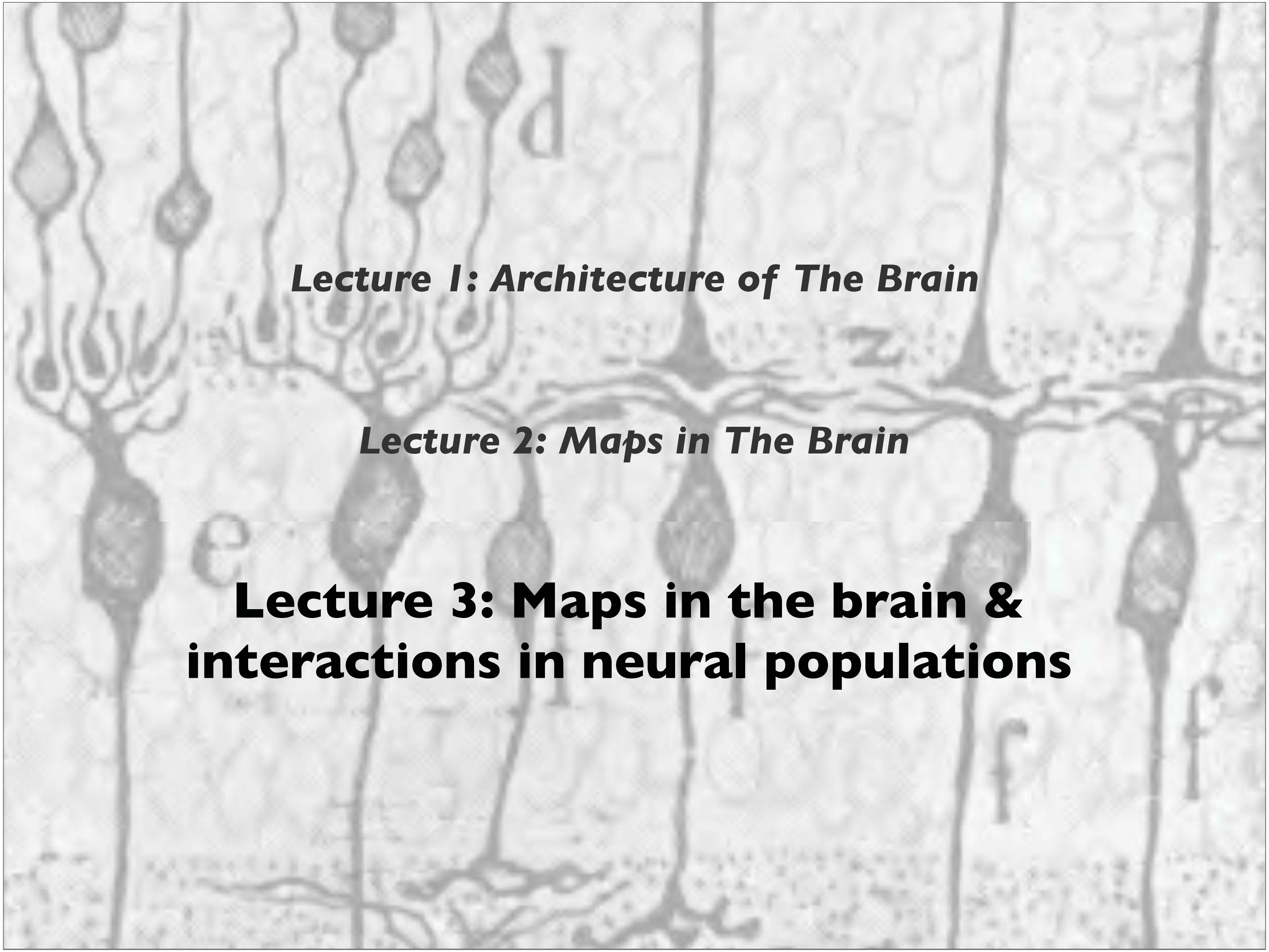
2415-11

Winter School on Quantitative Systems Biology

26 November - 7 December, 2011

Topics in Theoretical Neuroscience Pt.3

V. Balasubramanian
Univ. of Pennsylvania
Philadelphia
USA

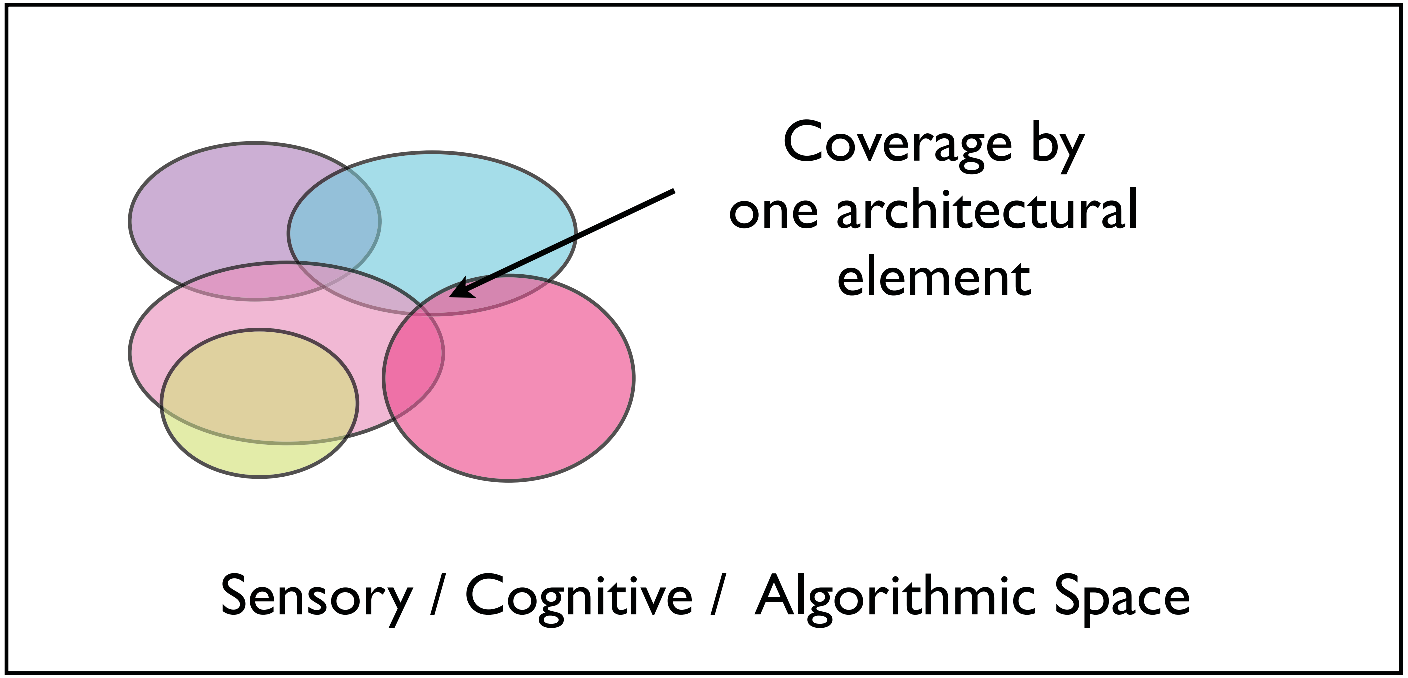


Lecture 1: Architecture of The Brain

Lecture 2: Maps in The Brain

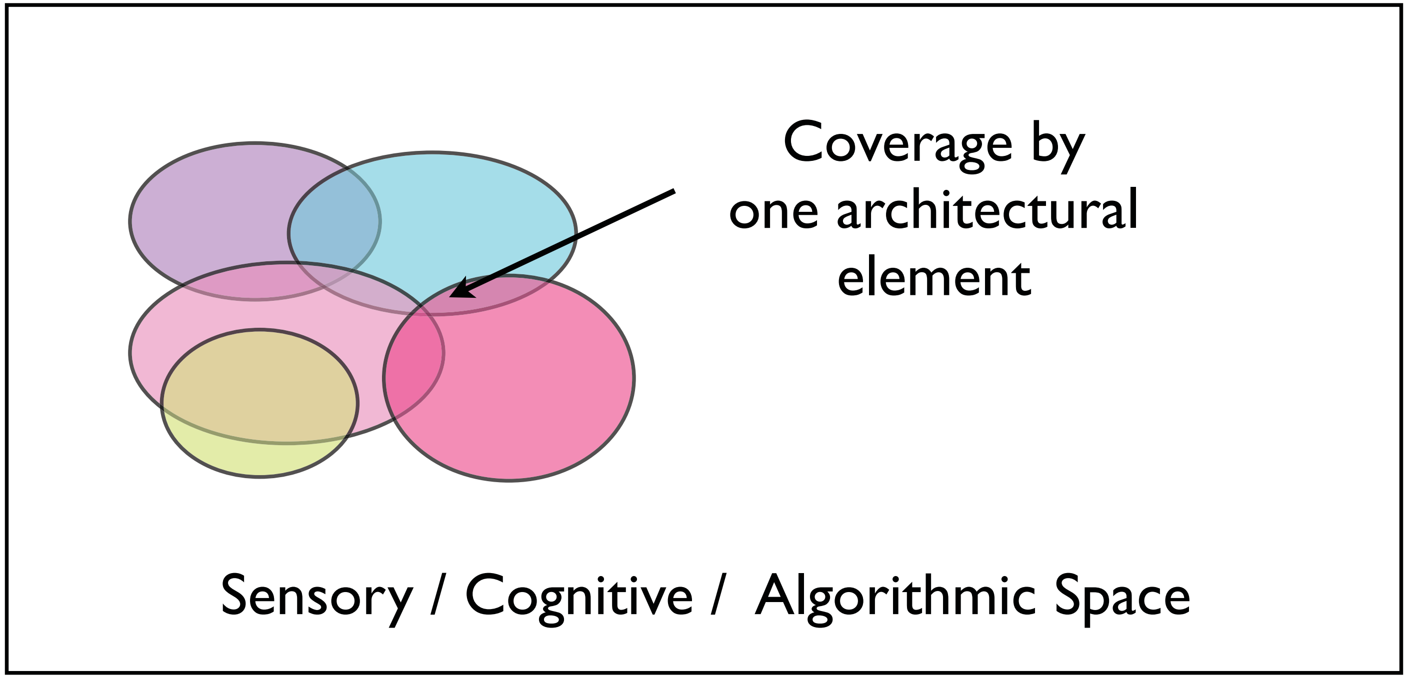
Lecture 3: Maps in the brain & interactions in neural populations

Resource allocation in the brain



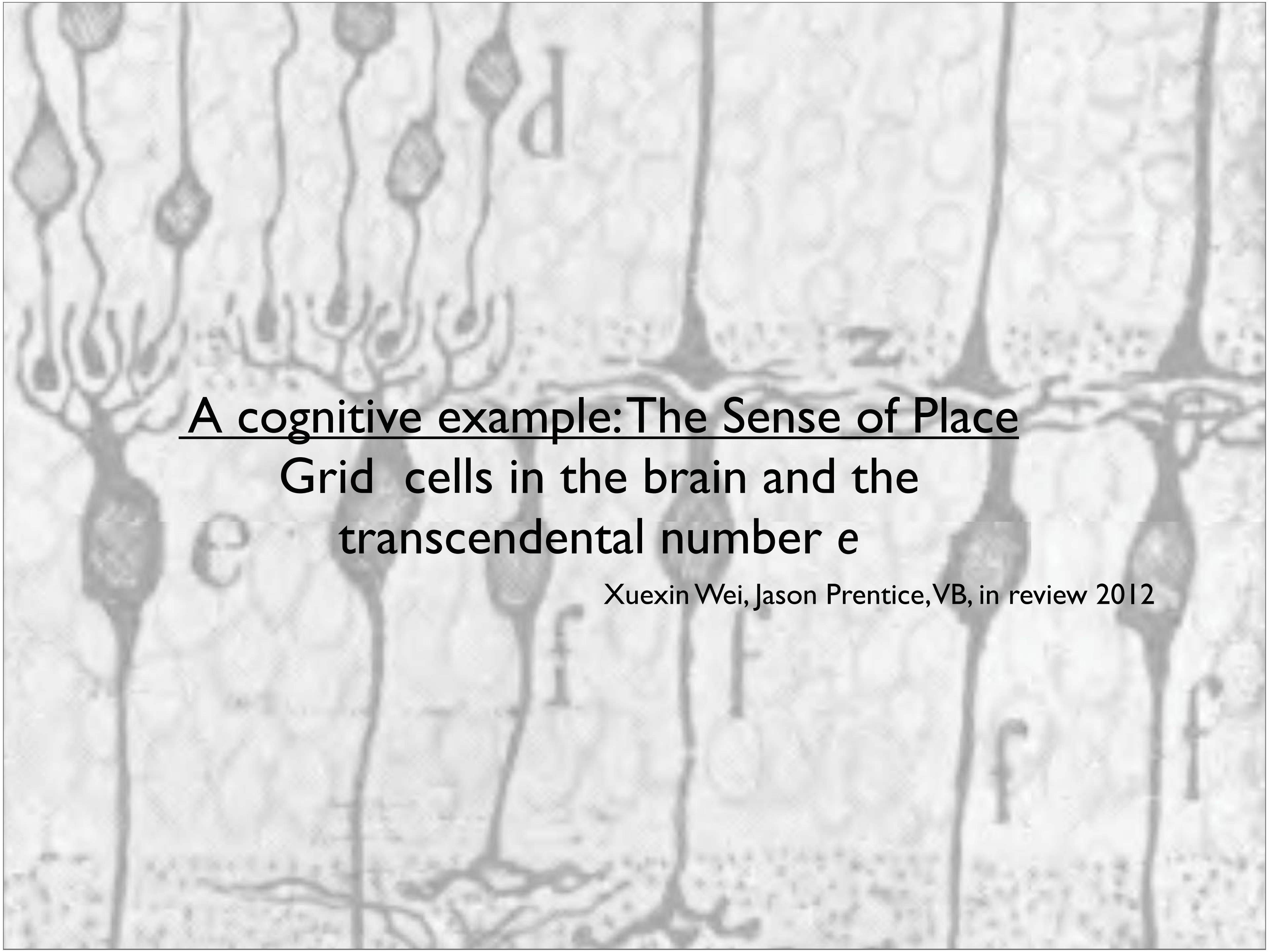
IDEA: The huge complexity of living systems, arising from the division into diverse functional units, can be understood in terms of a principle -- minimize resources while maximizing function.

Resource allocation in the brain



Previous Lecture:

(1) Retina has multiple output cell types.
(2) We explained the structure of receptive field mosaics of each type.
(3) We then explained the relative proportions of two types: OFF cells and ON cells



A cognitive example: The Sense of Place

Grid cells in the brain and the transcendental number e

Xuexin Wei, Jason Prentice, VB, in review 2012

A simple model for a representation of space

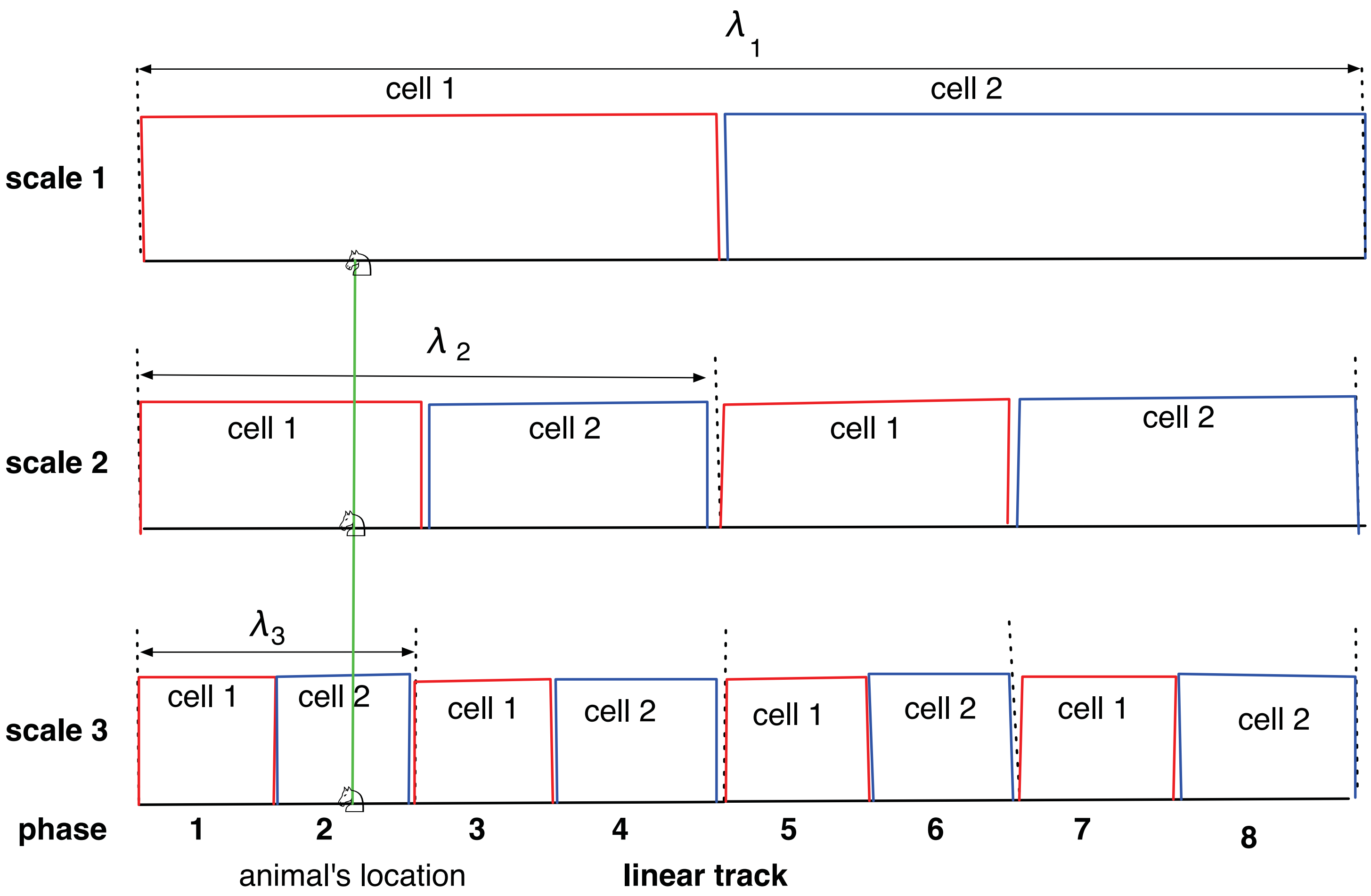
cell 1	cell 2	cell 3	cell 4	cell 5	cell 6	cell 7	cell 8
1m	1m	1m	1m	1m	1m	1m	1m

8m linear track

To achieve 1m resolution on an 8m track have 8 place cells, each of which fire when you are in a particular 1m wide location. **This requires 8 cells.**

Something like this may happen in a brain region called the hippocampus (but it’s more complicated, and the map is unstable to perturbations of the environment).

A better model for a representation of space



This is a "binary" representation of space and requires only 6 cells.

Grid Cells (Medial Entorhinal Cortex): a cortical representation of space

Hafting et al Nature 2005

Scale—increases along dorsal ventral axis

Orientation—local ensemble share same orientation

Phase—randomly distributed

Angle— Hexagonal grids in 2d

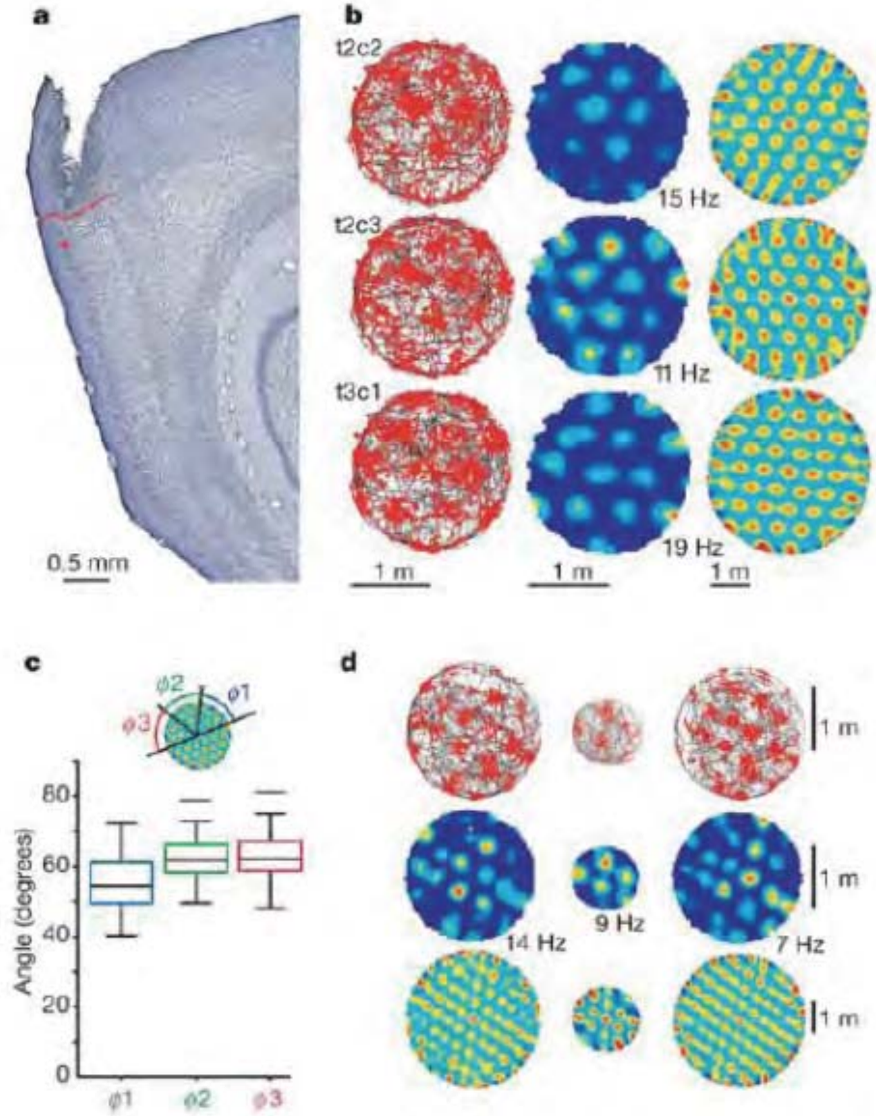
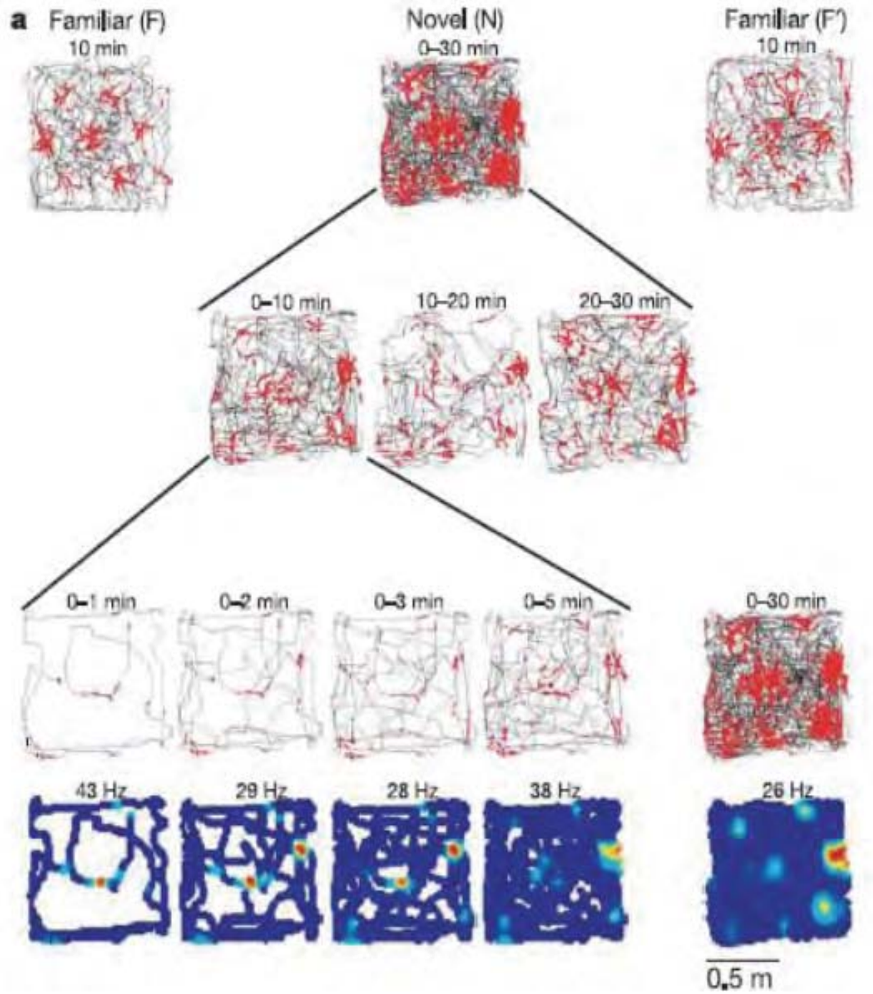


Figure 1 | Firing fields of grid cells have a repetitive triangular structure.

Grid pattern forms at the first encounter of an environment



Grids persist in darkness

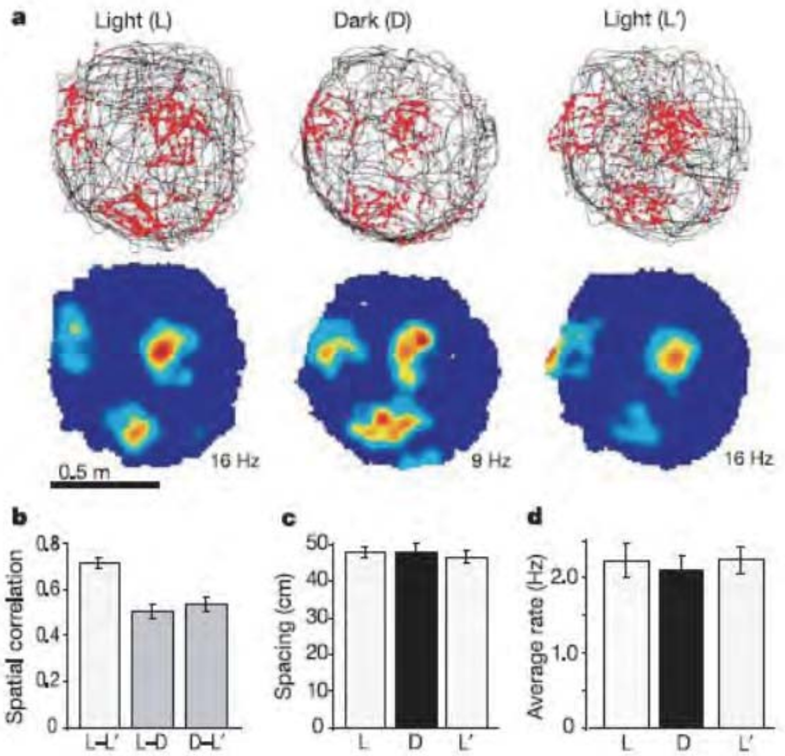
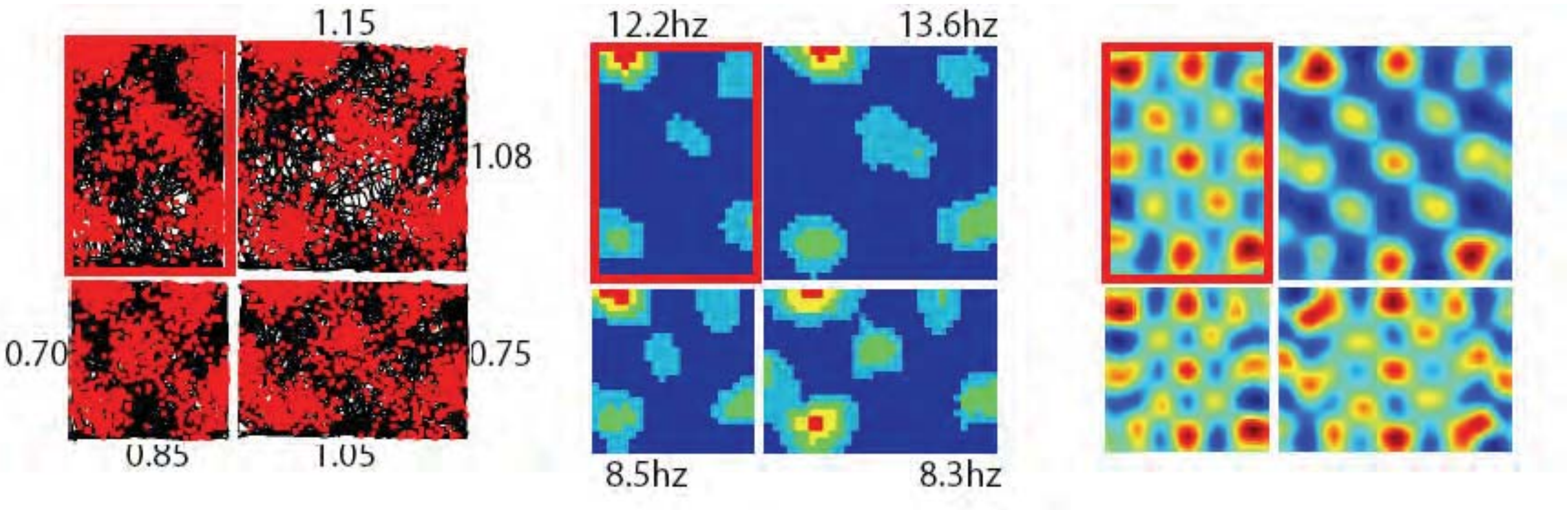


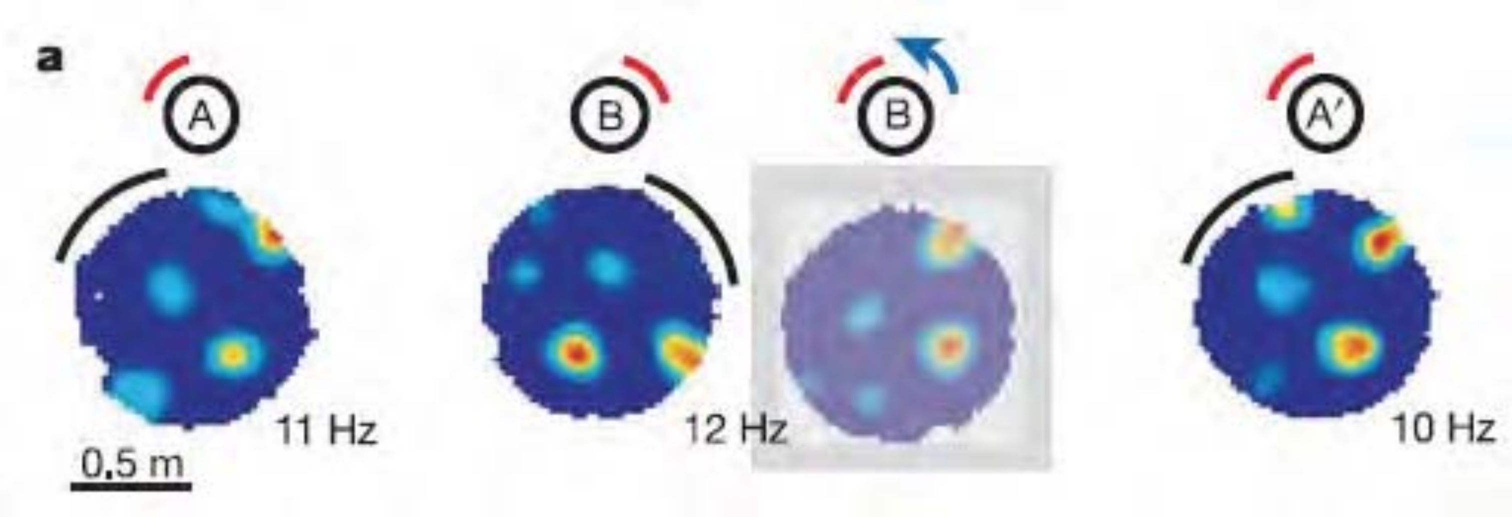
Figure 5 | Grids persist in darkness. **a**, Trajectory and rate maps for a representative dMEC cell after onset of darkness. Room lights were on for 10 min (L), off for 10 min (D), and on for another 10 min (L′). **b–d**, Firing properties of grid cells in L, D and L′ (all rats; means ± s.e.m.).

d

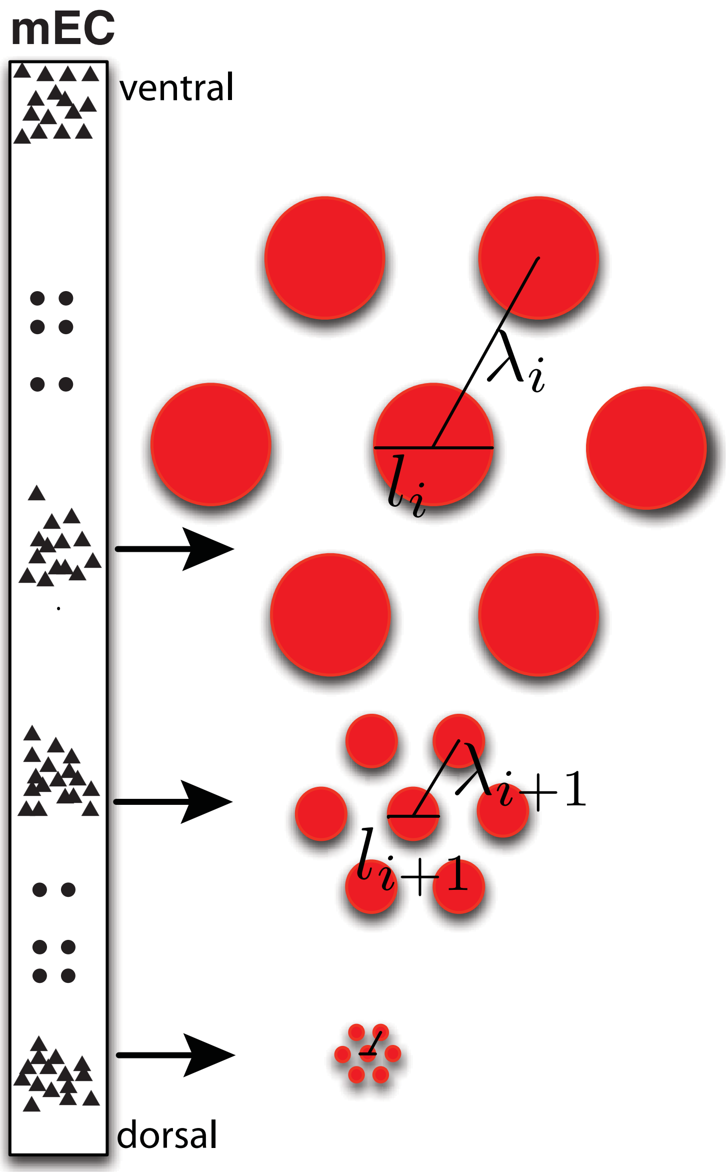
Grid rescale
(C. Barry et al 2007 Nat. Neuro)



Grids rotate according to the visual cue



Spatial characteristics of the grid system



λ_i = Scale of grid

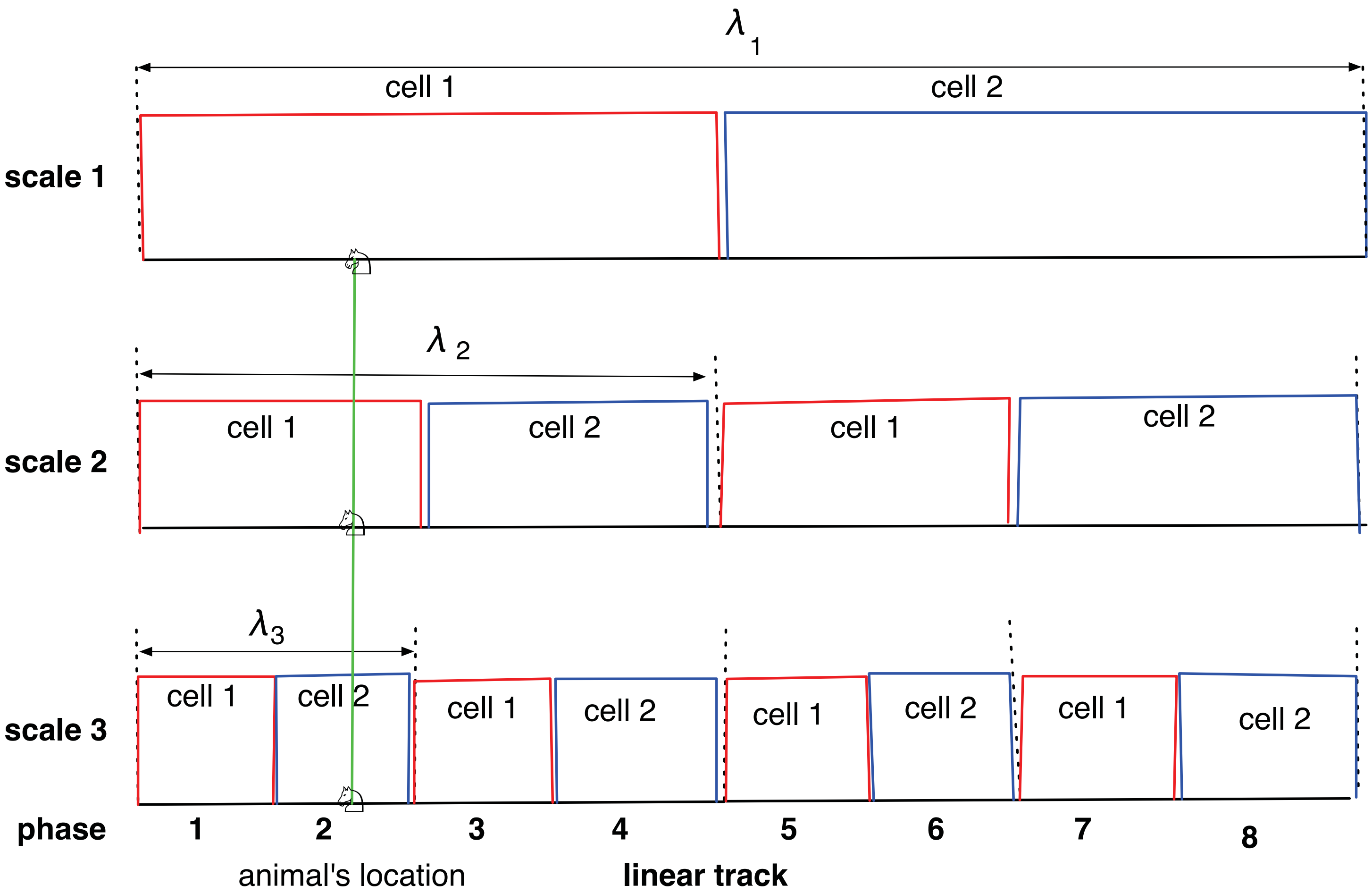
l_i = Diameter of single grid field

$$r_i = \frac{\lambda_i}{\lambda_{i+1}}$$ = Ratio of grid scales

$$R = \frac{\lambda_1}{\lambda_n} = \prod_i r_i$$ = Resolution

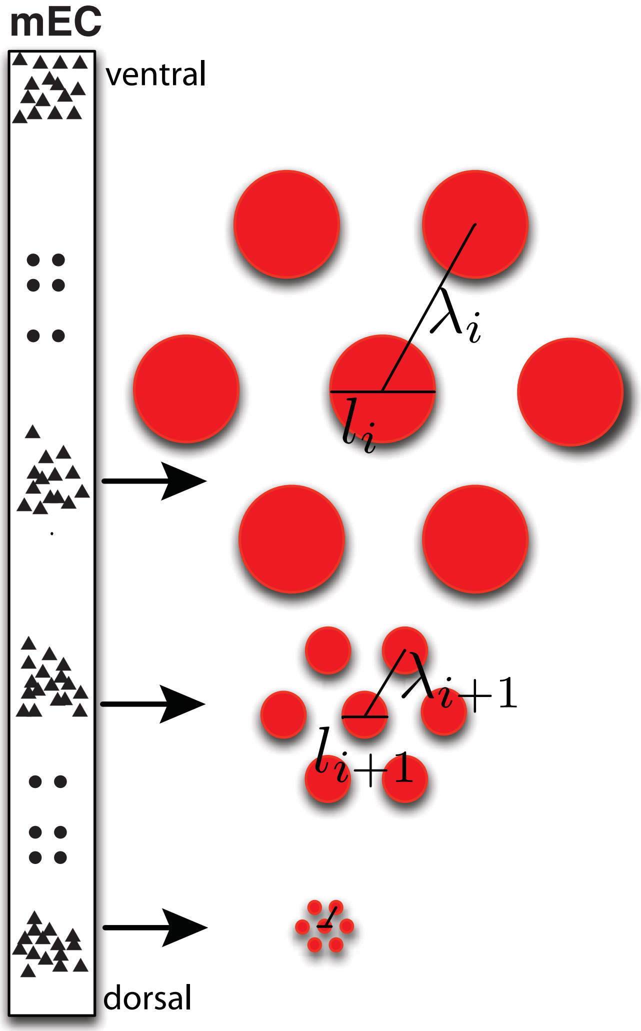
$$N \propto \sum_i r_i$$ = Number of grid cells

A simple modular representation of space



What ratio between scales minimizes the number of cells required achieve a given spatial resolution?

Optimizing the grid (simple model)



$$r_i = \frac{\lambda_i}{\lambda_{i+1}}$$ = Ratio of grid scales

$$R = \frac{\lambda_1}{\lambda_n} = \prod_i r_i$$ = Resolution

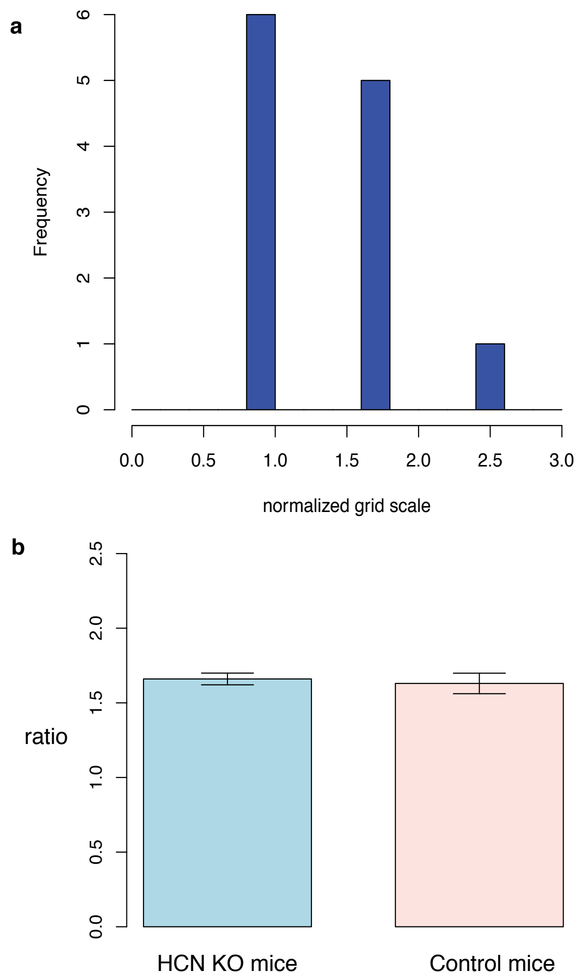
$$N \propto \sum_i r_i$$ = Number of grid cells

Minimize number of cells (N) for fixed resolution (R). This requires all r_i equal.

Result: (a) 1 dimension: optimal grid scale ratio = e
(b) 2 dimensions: optimal grid scale ratio = $\sqrt{e}$

Empirical data

The ratios of grid scales in rats very close to $\sqrt{e}$ ~ 1.6 as predicted by model.



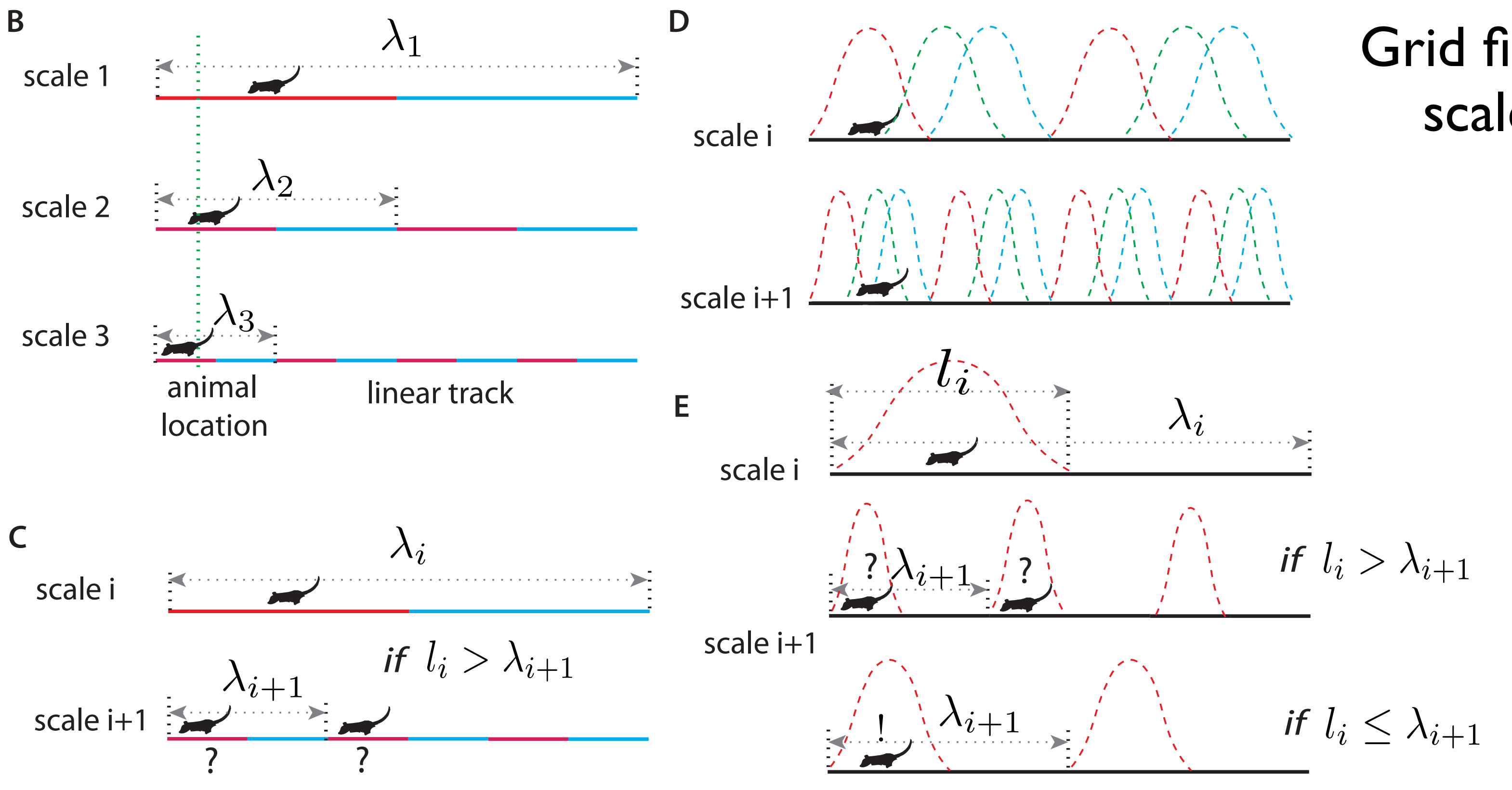
Barry et al., 2007

The ratios of grid scales and grid field diameter in mice is close to $\sqrt{e}$. Also predicted by model.

Giocomo et al., 2011

Good match with our simple theory!

More realistic model



Grid fields at each scale overlap

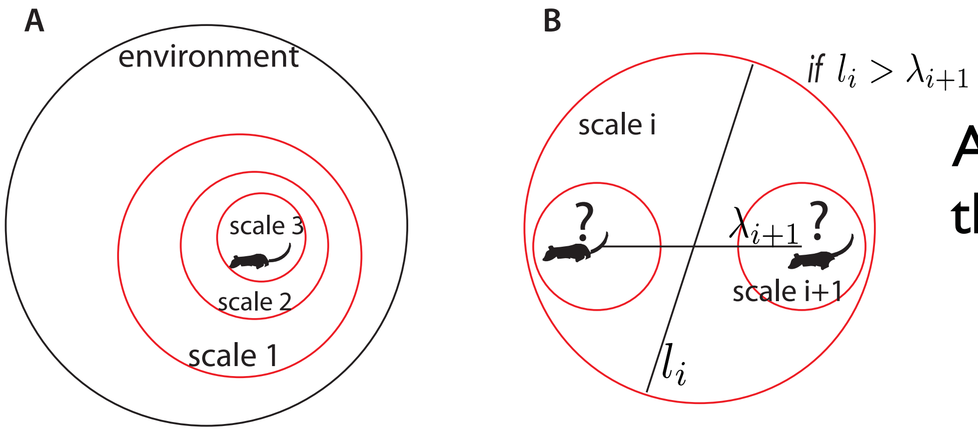
Ambiguities arise if the grid field width is too large compared to the next scale

Alternative decoding strategies

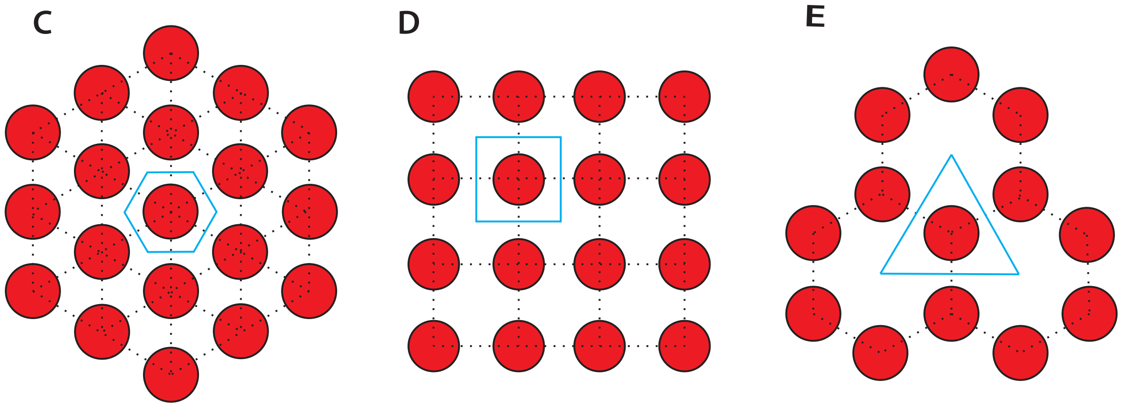
- "Winner take all"
- Optimal probabilistic decoding ("Bayesian")

Two dimensional grids

Circular, overlapping firing fields



Ambiguities arise if the grid field width is too large compared to the next scale

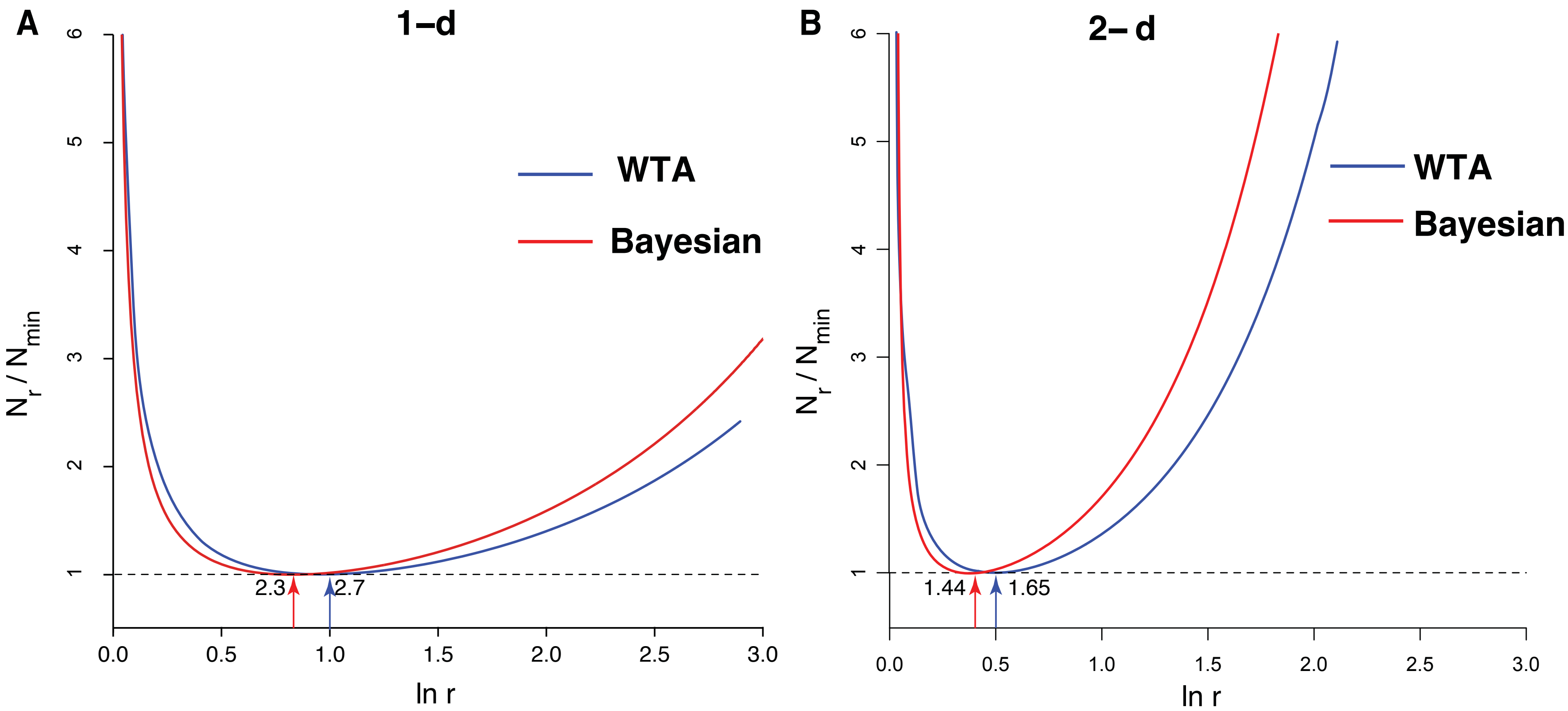


Invariances dictate three possible grids

Alternative decoding strategies

- "Winner take all"
- Optimal probabilistic decoding ("Bayesian")

Number of neurons as a function of scaling ratio

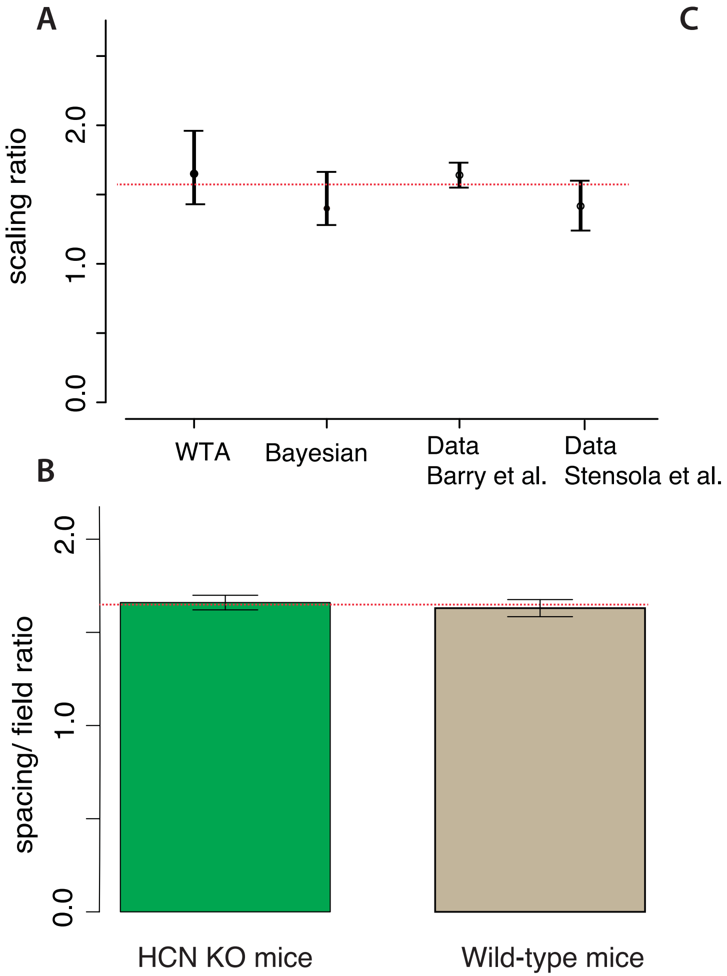


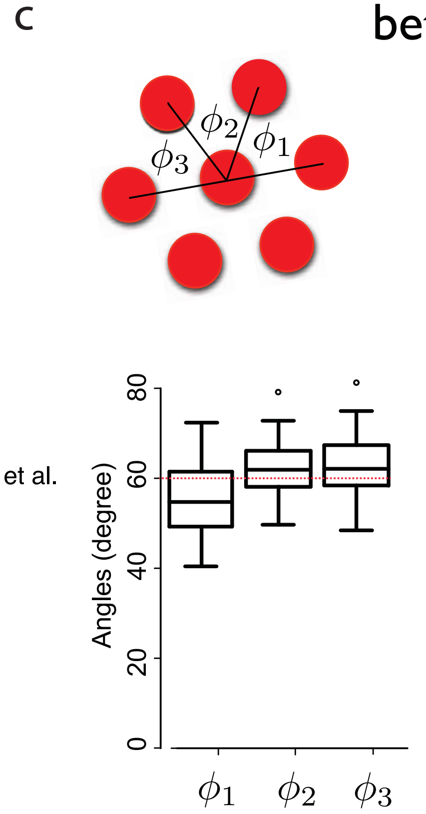
- “Winner take all” = WTA
- Optimal probabilistic decoding = Bayesian

Theory matches experiment

(A) Ratio between neighboring grid scales (Theory error bars = 95% of minimal neuron number)

(B) Ratio between grid field width and grid scale (Red line = prediction)

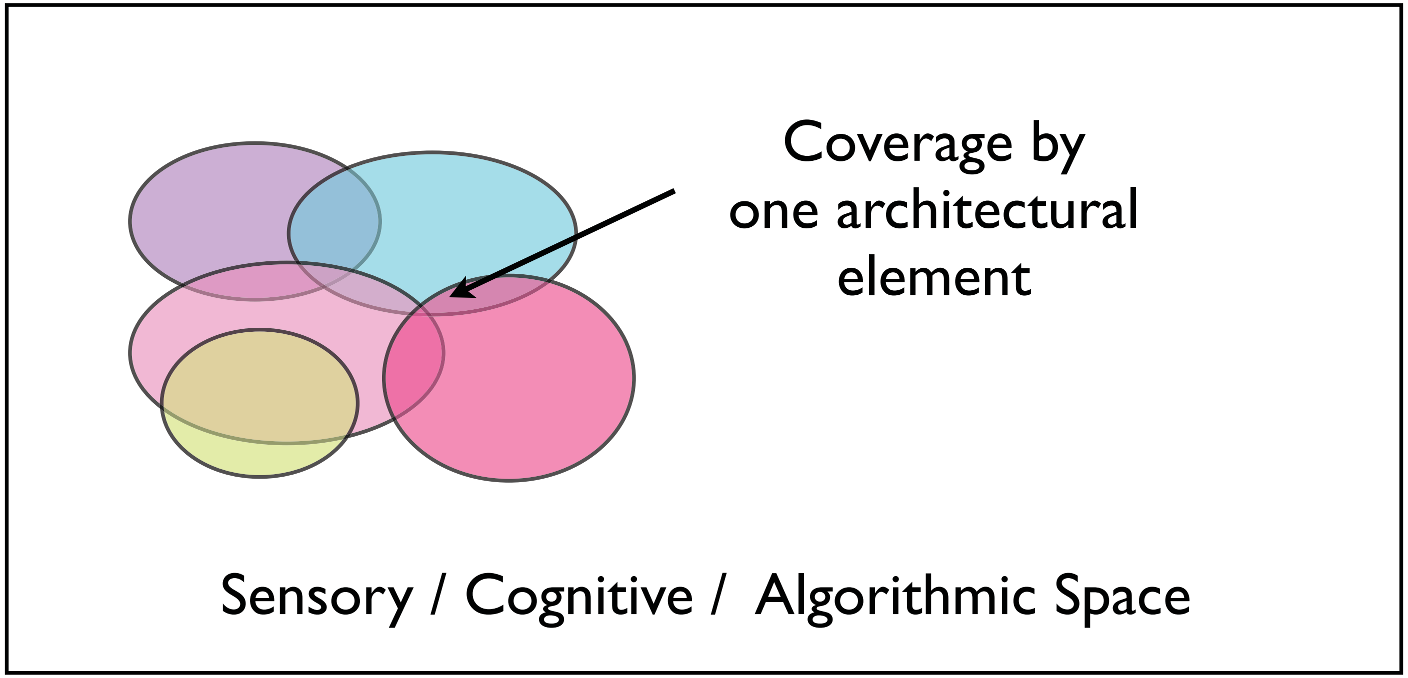




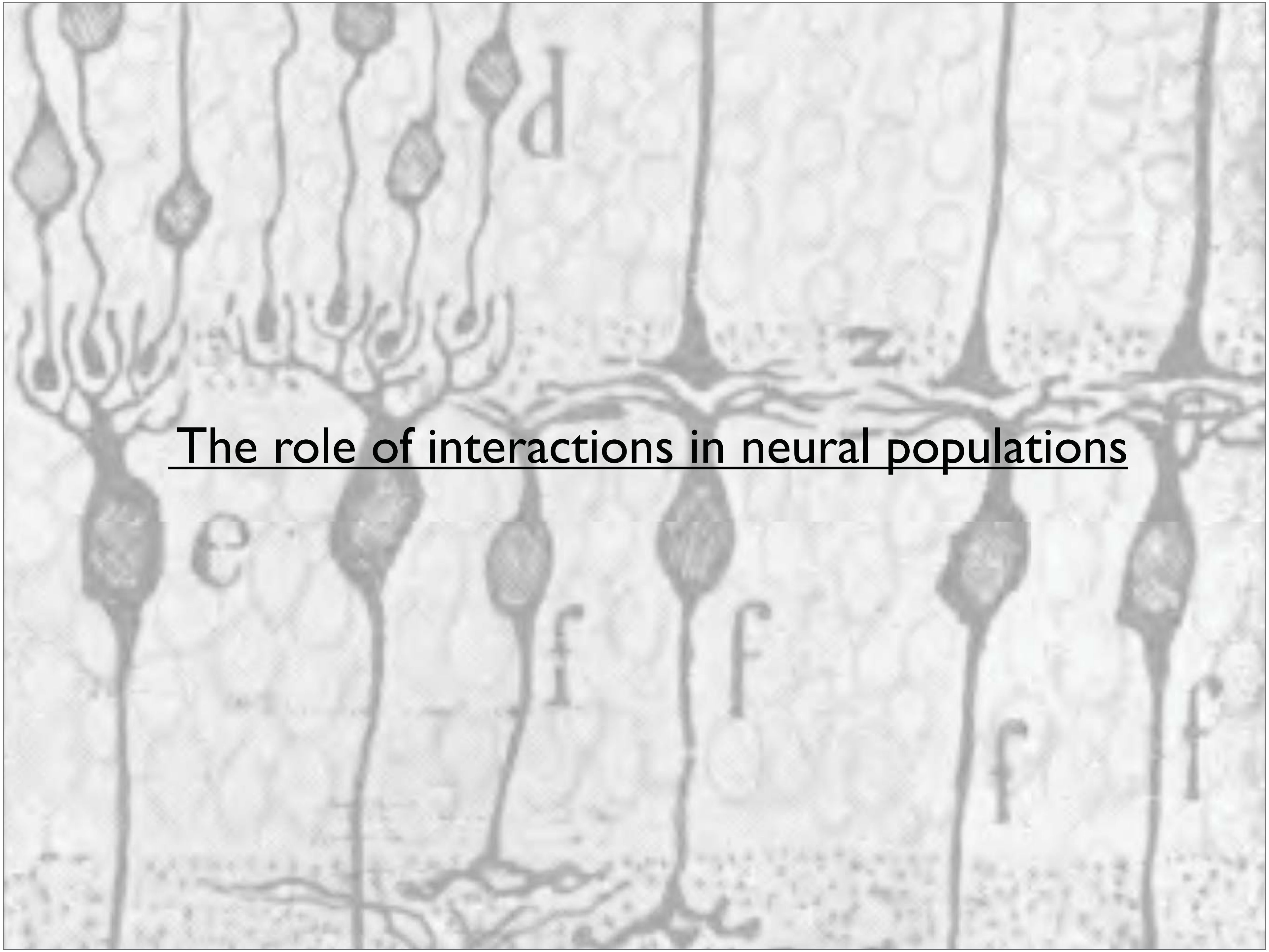
(C) Angles between grid fields (Red line = prediction)

- “Winner take all” = WTA
- Optimal probabilistic decoding = Bayesian

Resource allocation in the brain

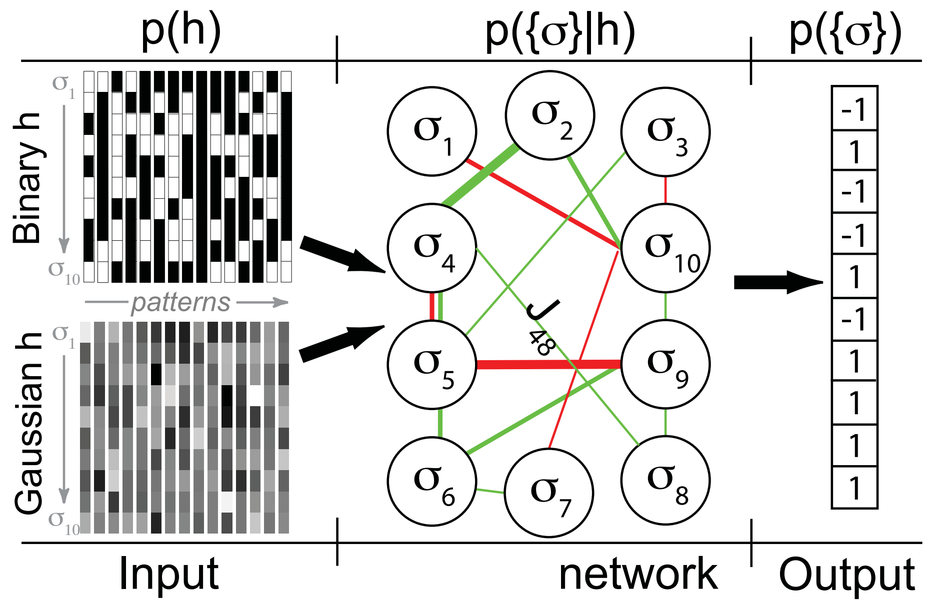


But what about interactions between architectural elements?



The role of interactions in neural populations

Interacting networks responding to correlated inputs



Networks can store memories, learn things, perform complex computations.

Pairwise interacting models of response

- The retinal ganglion cell network (and cortical slices) have response distributions with correlations that are well-summarized by a distribution with only pairwise interactions (Schneidman et al., Shlens et al., 2006):

$$\hat{P}(\{\sigma_i\}) = \frac{1}{Z(h,J)} \exp\left[\sum_i h_i \sigma_i + \frac{1}{2}\sum_{i,j} J_{ij}\sigma_i\sigma_j\right]$$

- Generalize to have stimulus dependence (β measures *neural reliability*):

$$P(\{\sigma_i\}\,|s) = \frac{\exp\left\{\beta\left(\sum_i (h_i^0 + h_i(s))\sigma_i + \frac{1}{2}\sum_{i,j} J_{ij}\sigma_i\sigma_j\right)\right\}}{Z(h,J)}$$

- When J_{ij}=0, these are independent Linear-Nonlinear-Poisson neurons:

$$P(\{\sigma_i\}|s) = Z^{-1}\prod_i \exp\left[\beta\left(h_i^0 + h_i(s)\right)\sigma_i\right] \Longrightarrow \langle\sigma_i(s)\rangle = \tanh\left[\beta\left(h_i^0 + h_i(s)\right)\right]$$

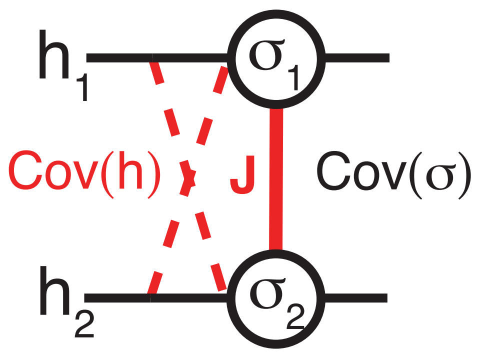
What can we do with such representations of population activity?

Two neurons: optimal strategy varies with noise

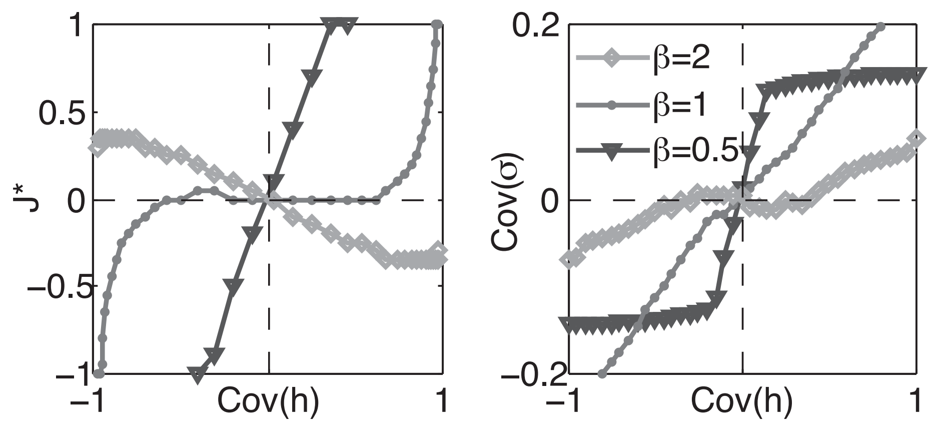
Population coding strategies

- Independence: Each neuron responds independently to input
- Decorrelation: The network decorrelates its inputs
- Error correction: The network fights noise through redundancy
- Synergy: Neural responses synergistically encode stimuli

What strategy should an optimal network follow?

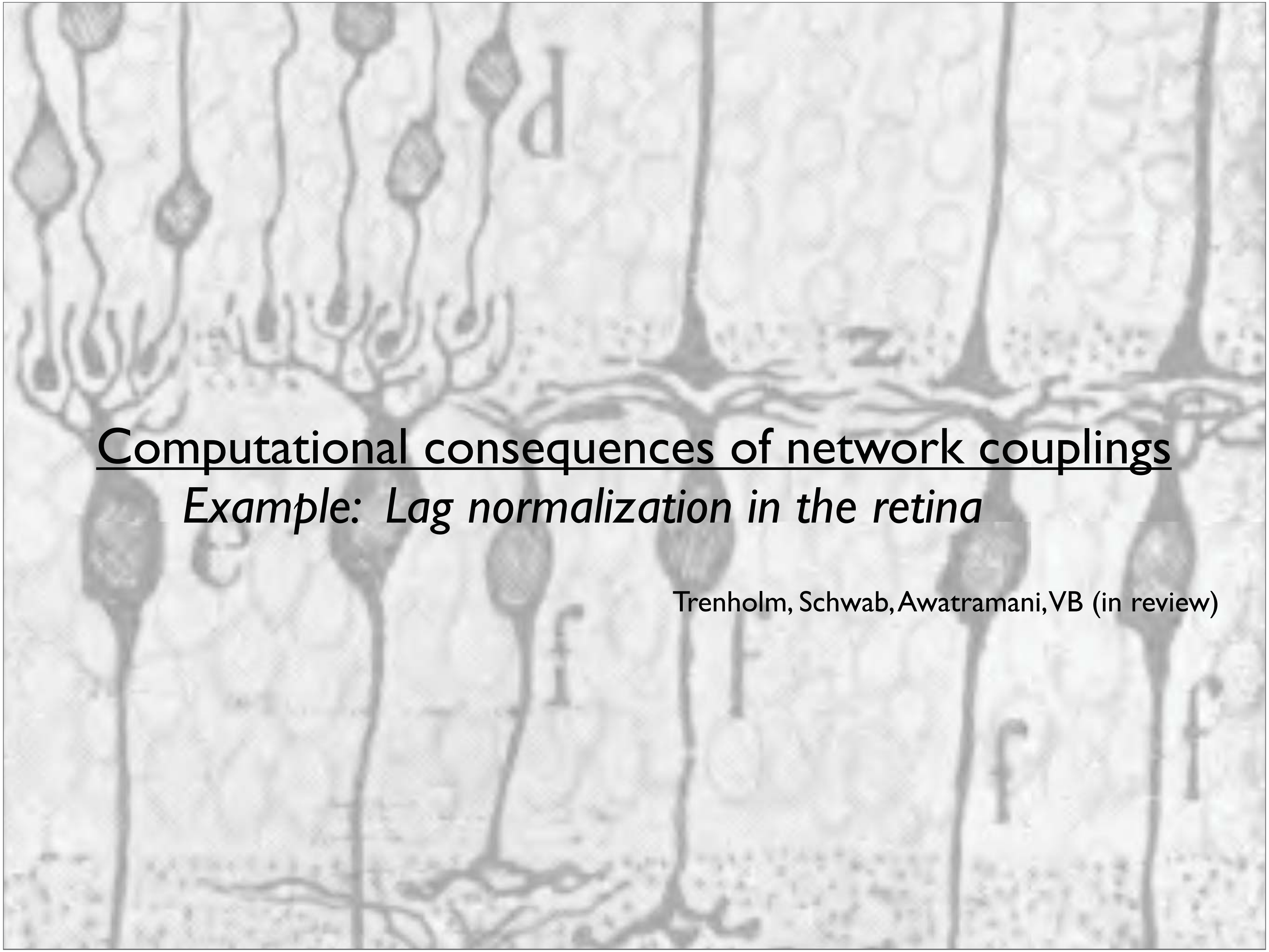


Two interacting neurons encoding correlated inputs



At high reliability (β), the network decorrelates the inputs. At low reliability (β), the network enhances input correlations to achieve error correction via redundancy.

Tkacik, Prentice, Schneidman, VB (PNAS 2010)

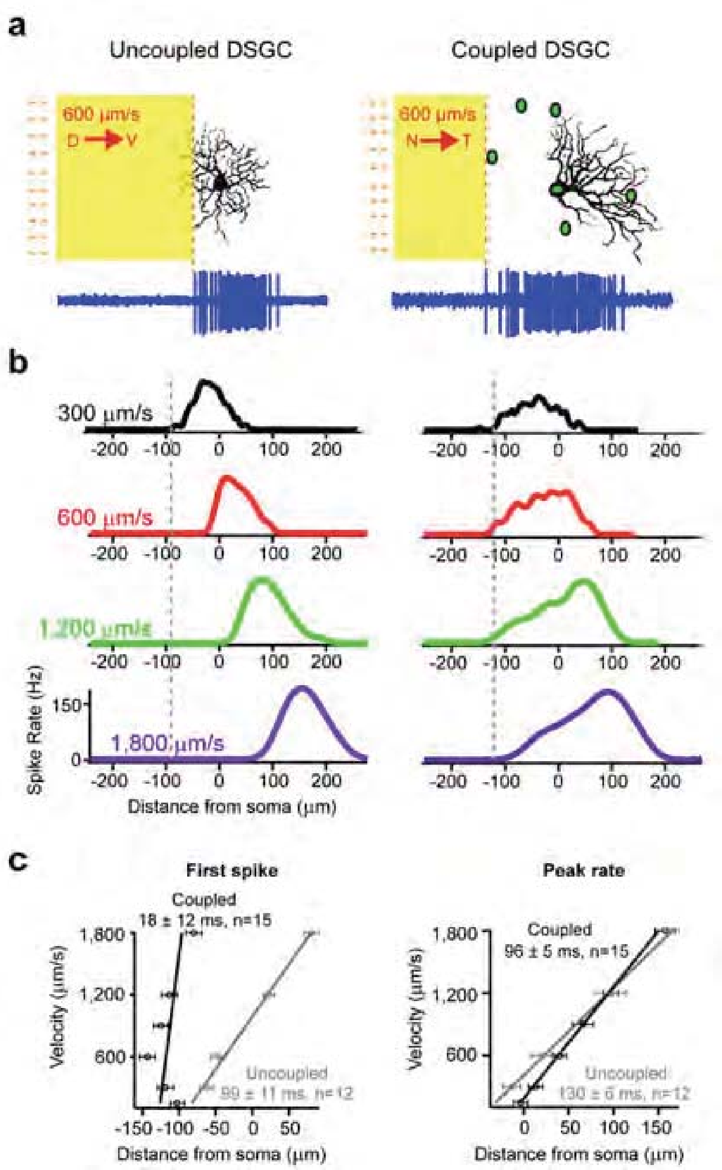


Computational consequences of network couplings

Example: Lag normalization in the retina

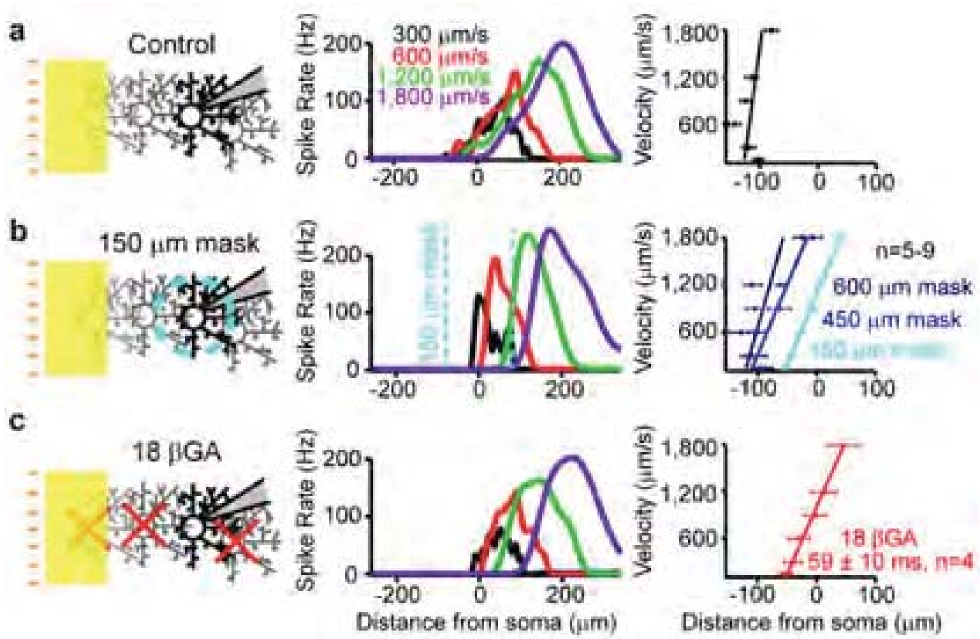
Trenholm, Schwab, Awatramani, VB (in review)

“Lag normalization” in the retina



A class of Direction Selective Ganglion Cells (DSGCs) removes the effects of retinal circuit delays

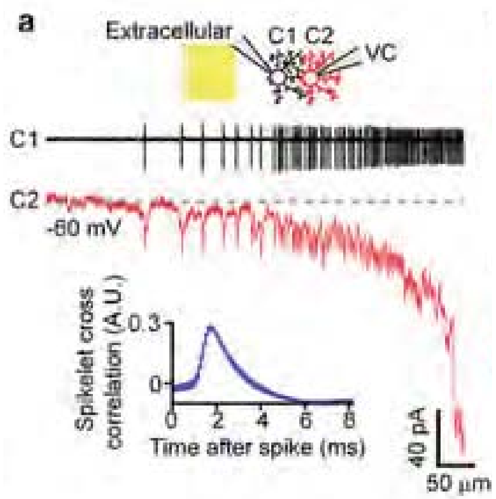
Lag normalization is a population effect of gap junctions



Control cells

Masking destroys lag normalization

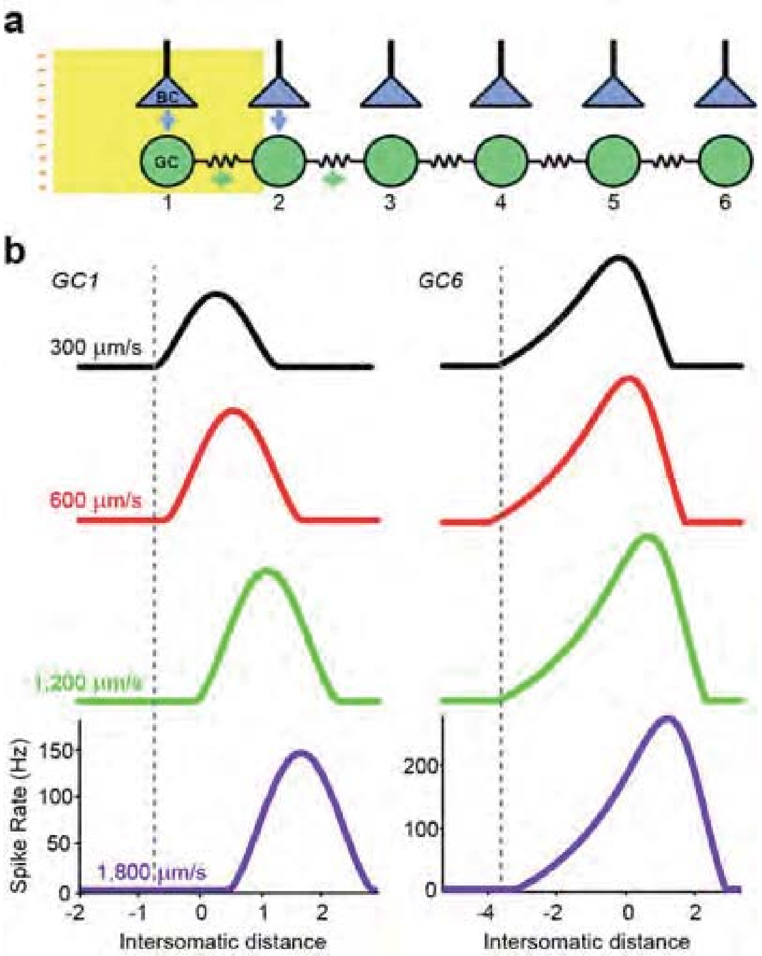
Blocking gap junctions destroys lag normalization



Paired recordings show that spikes in one DSGC reliably produce spikelets in its neighbor

Lateral coupling explains lag normalization

Model schematic



Lag normalization develops by the sixth neuron in the model array

$$I_n(t) = J_n(t) + \alpha I_{n-1}(t)$$

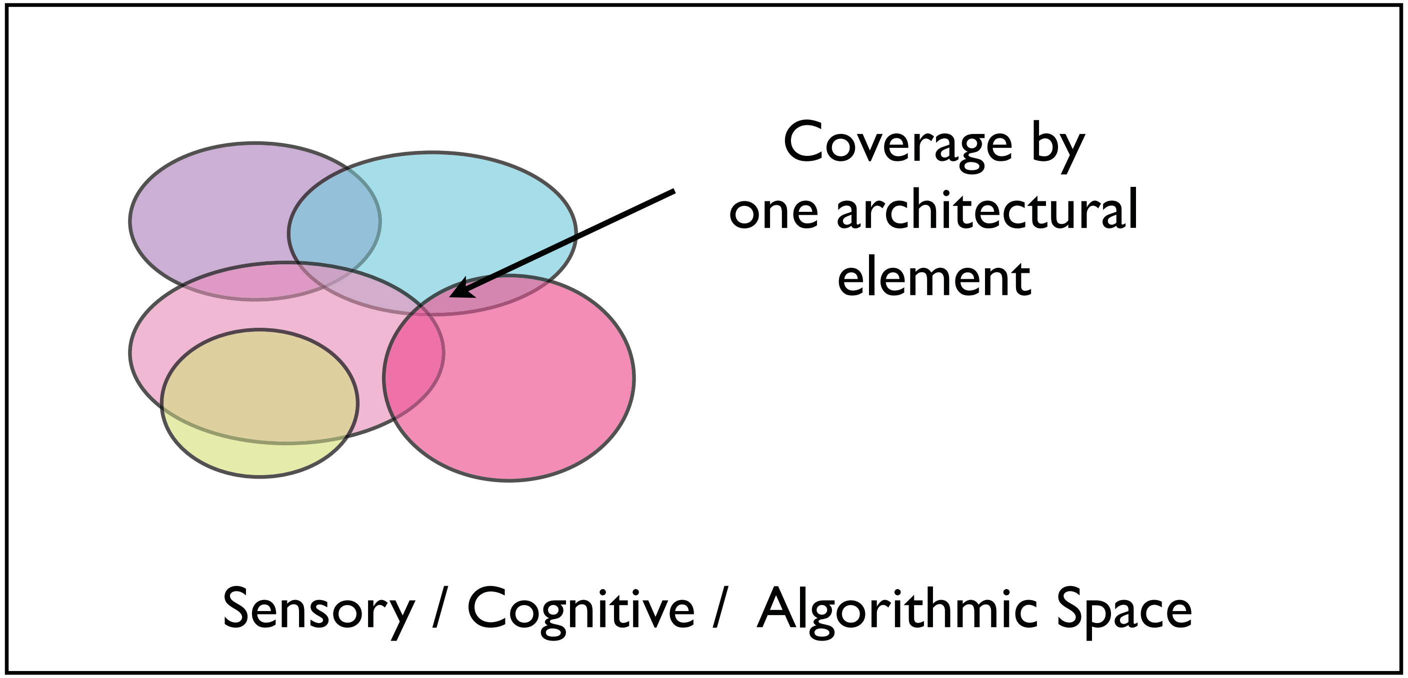
$J_n(t)$ = Gaussian spatial tuning curve with velocity dependent amplitude g(v)

$\alpha =$ gap junction coupling

$$I_n(t) = \sum_{m=1}^{n} \alpha^{n-m} J_m(t)$$

Spiking occurs when the current I_n reaches some threshold C.

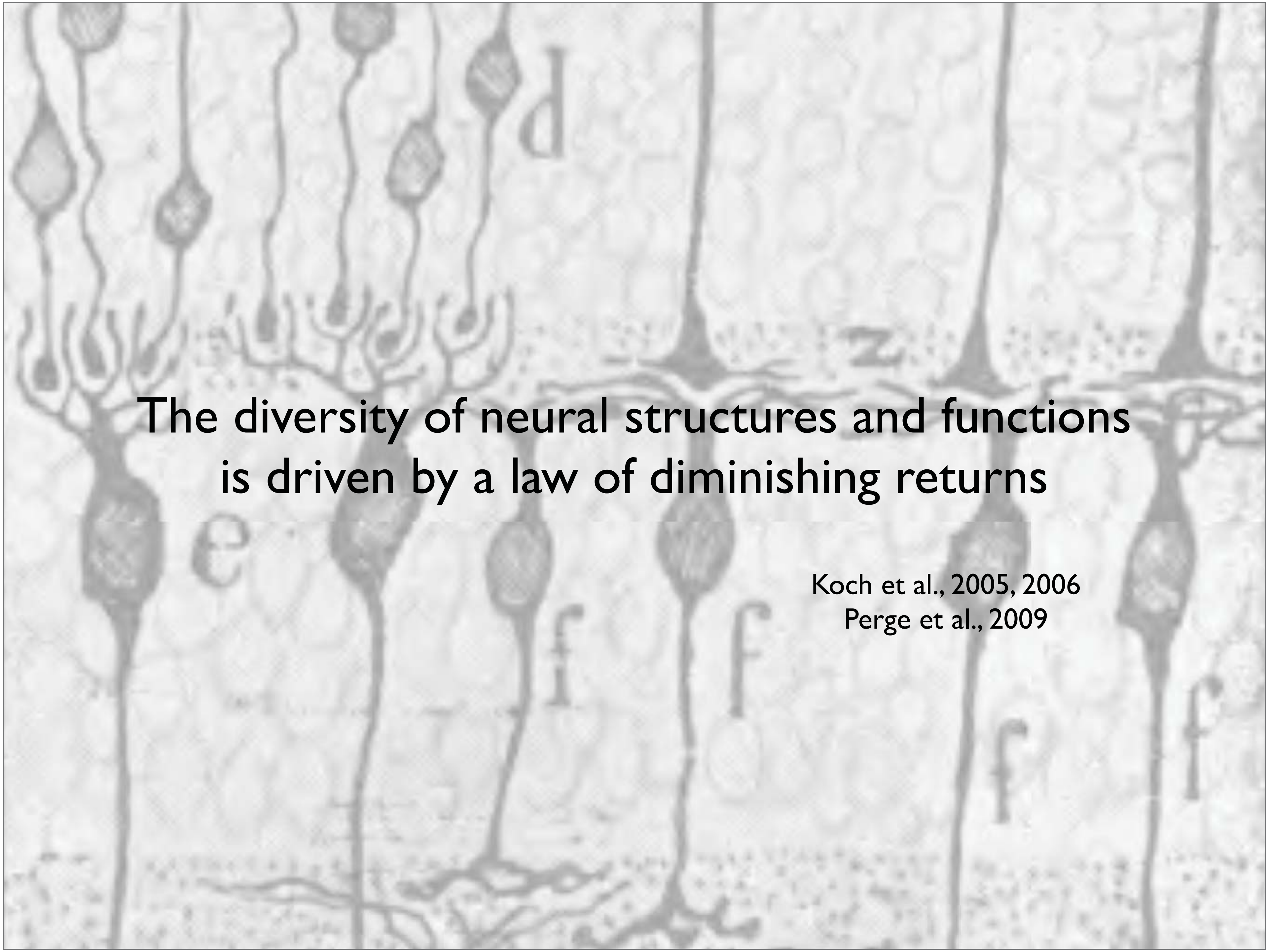
Resource allocation in the brain



Ongoing work:

- Representation of textures in visual cortex
- Planning of paths based on hippocampal place cells
- Representation of odors by the olfactory cortex
- Representation of structured stimuli by retina

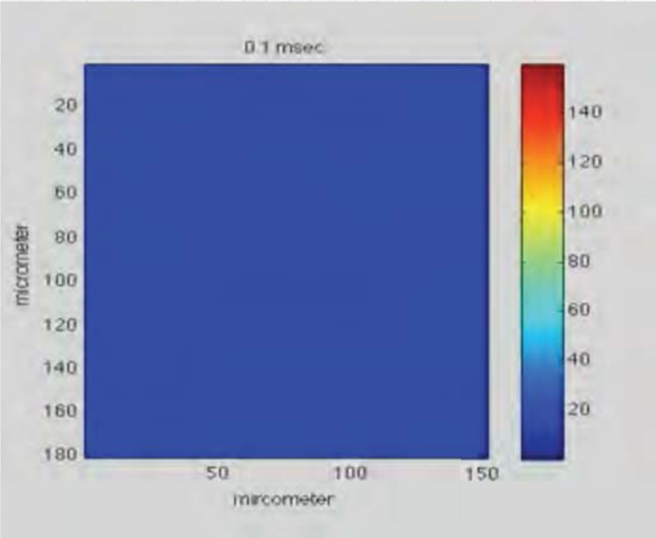
CHALLENGE: Understanding the interactions



The diversity of neural structures and functions is driven by a law of diminishing returns

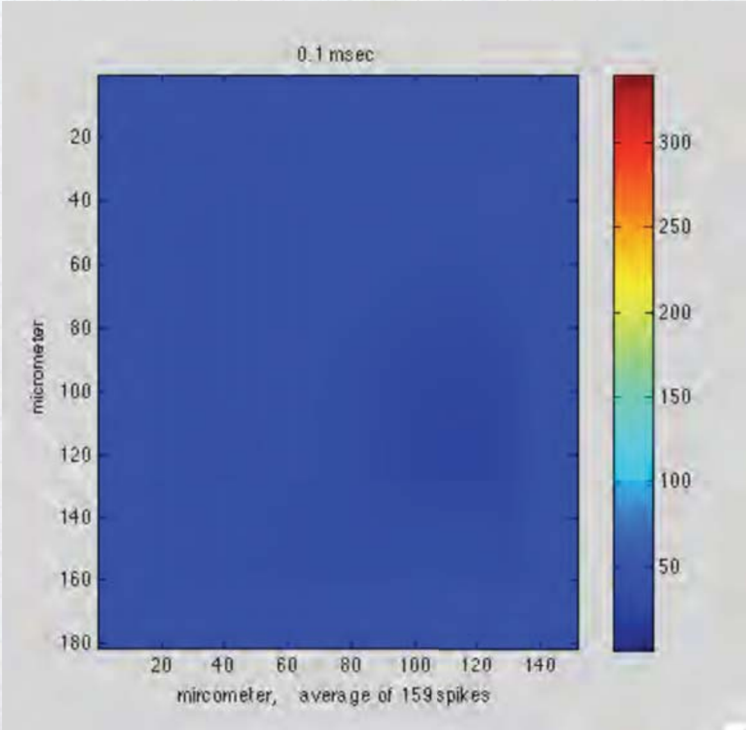
Koch et al., 2005, 2006
Perge et al., 2009

Firing events

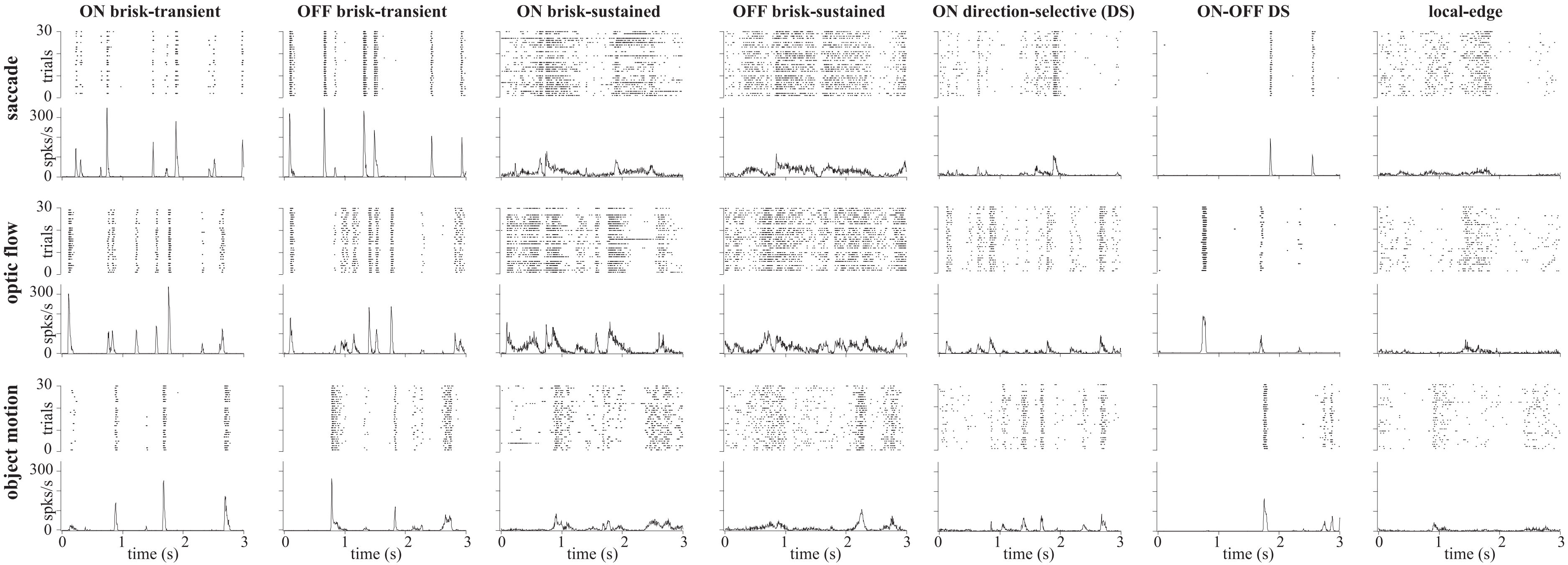


67 ms of data,
viewed as a movie.
[data have been smoothed]

Some spikes move across the array:



Ganglion cell responses to natural stimuli

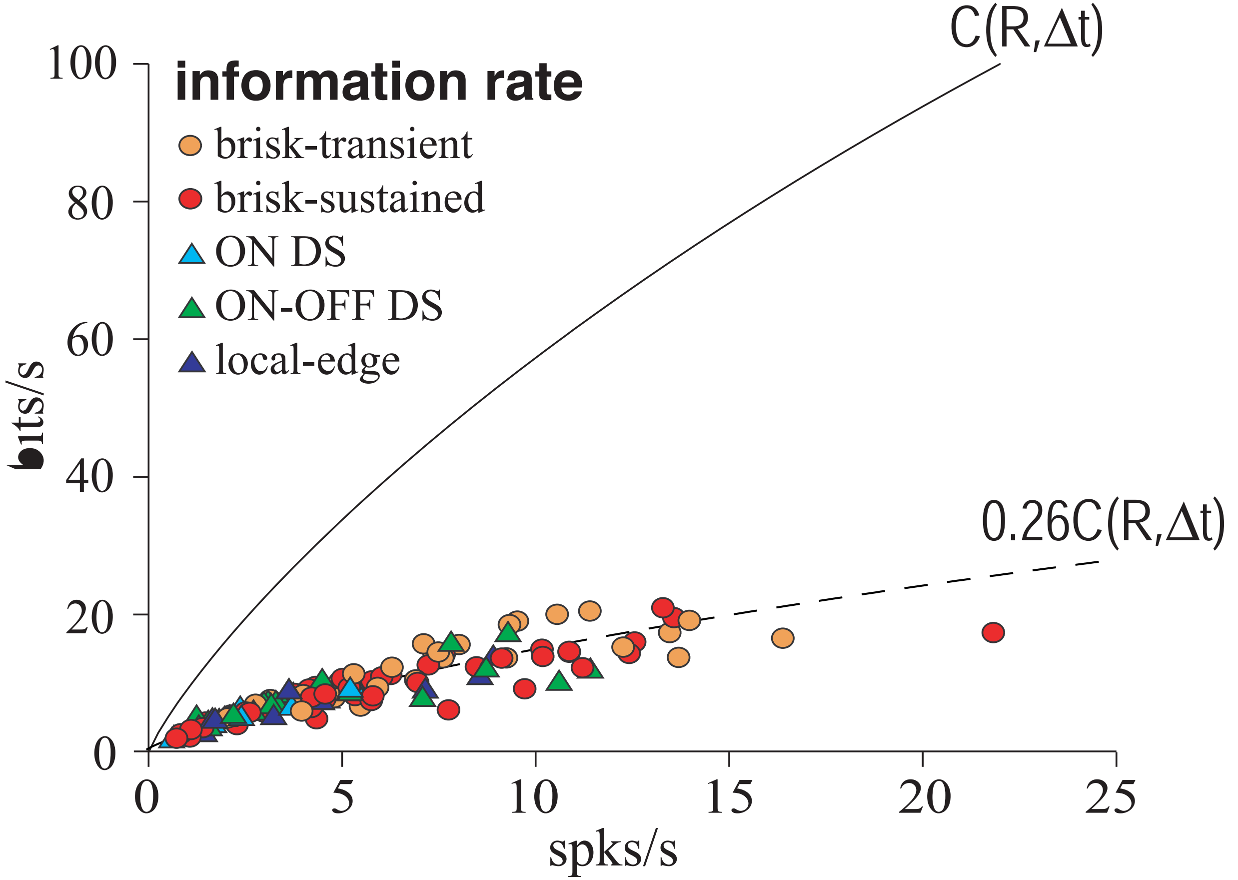


Guinea Pig (Koch et al., 2005; Koch et al., 2006)

How and why is the total information traffic divided up between so many different types of channels?

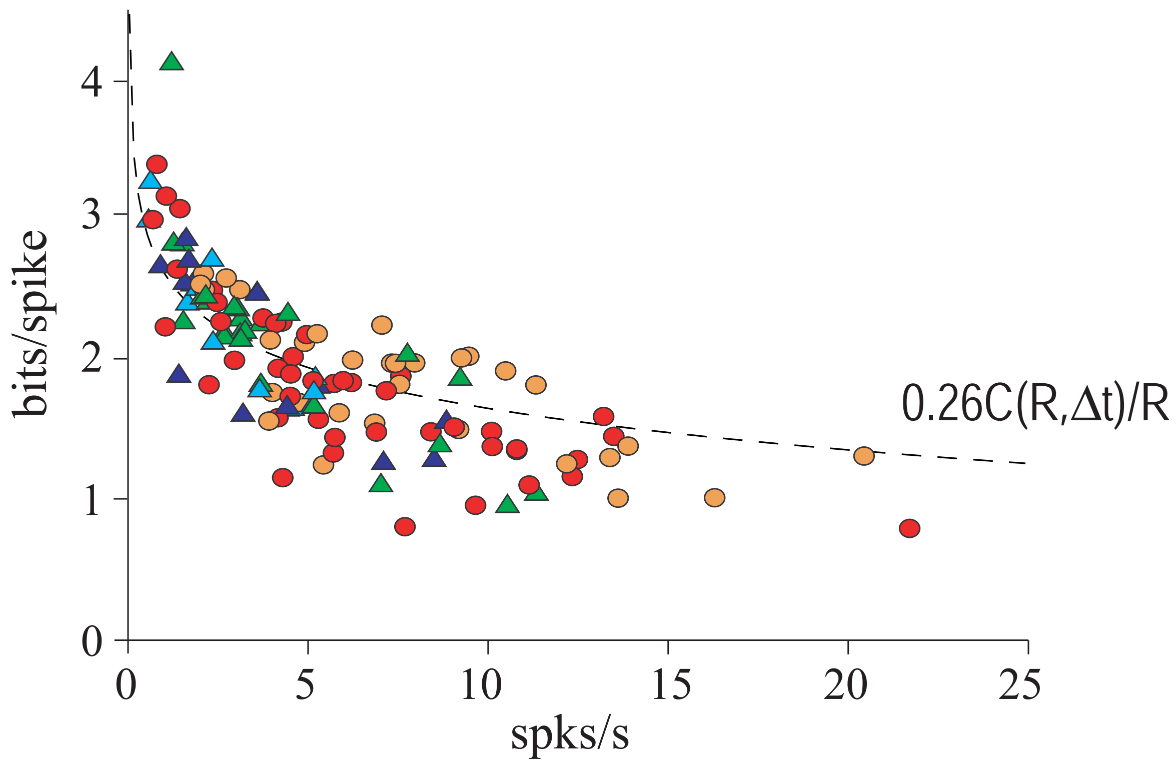
All cells are equally efficient for all stimuli

Capacity = $C(R, \Delta t) = -R\,\Delta t\,\log(R\Delta t) - (1 - R\,\Delta t)\log(1 - R\,\Delta t)$



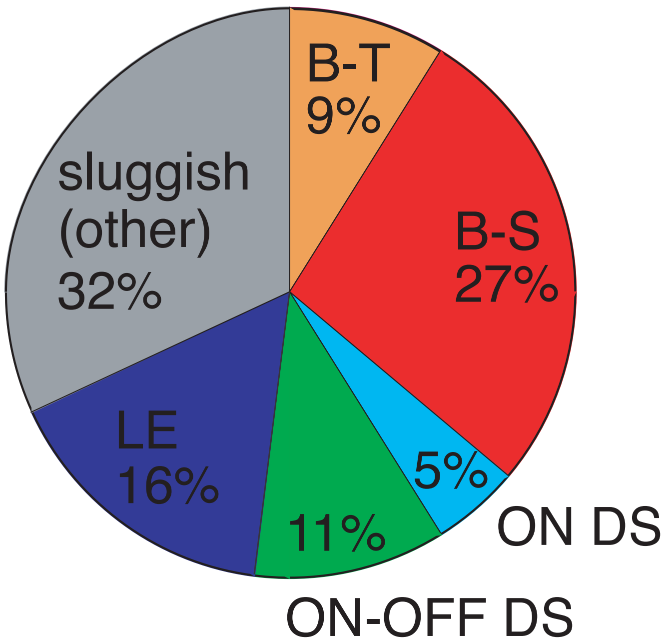
All cell types are ~26% efficient for all naturalistic stimuli.

Efficiency of a single spike on depends on firing rate



Information traffic along the optic nerve

cell type	bits/s	bits/spike	cells	bits	spikes
Brisk-transient	13	1.9	6,000	78,000	41,000
Brisk-sustained	10	1.8	24,000	240,000	133,000
ON DS	6	2.2	7,000	42,000	19,000
ON-OFF DS	8	2.2	12,000	96,000	44,000
Local-edge	7	2.1	20,000	140,000	67,000
Sluggish (other)	9	2.2	31,000	279,000	127,000
Total			**100,000**	**875,000**	**431,000**



- Messages are transmitted asymmetrically - less studied sluggish cells account for most of the traffic

- Scaling up to human optic nerve (million axons) gives a traffic of order an Ethernet cable. This is the same order of magnitude as the amount of information in natural scenes according to Ruderman, 1994.

Why this organization?

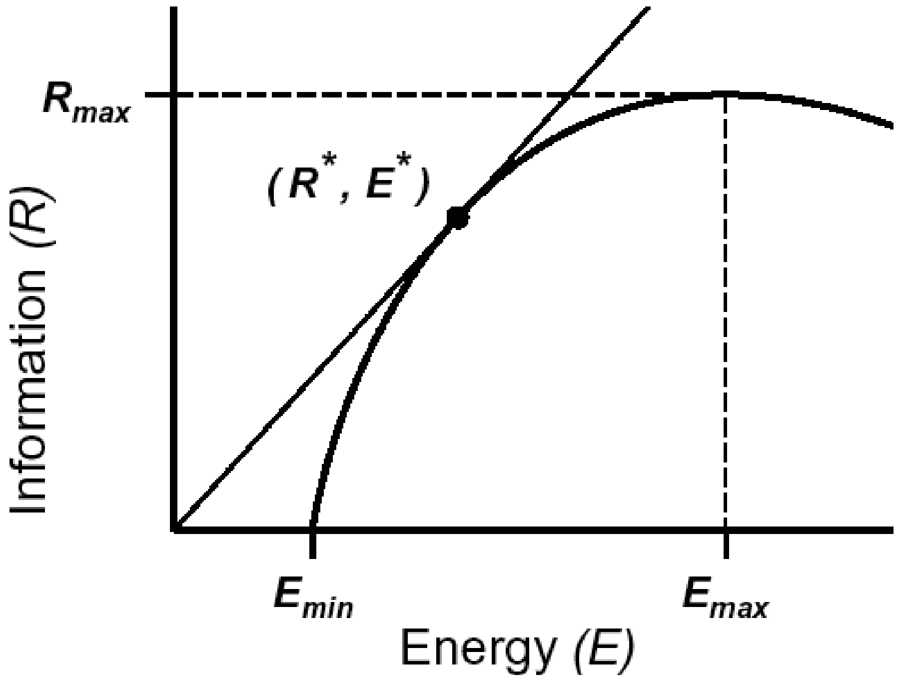
- Given equal coding efficiency for all types, why not use a single type to send all information over a high firing rate axon?
- Perhaps because the cost of signaling increases non-linearly with firing rate.

Comparison of three kinds of neural cables			
Type	LED	Brisk-Transient	Super
Mean spike rate for cell	4 Hz	8 Hz	40 Hz
bits/spike	2.1	1.8	1.1
spikes/s for 300 bits/s traffic over cable	~140	~170	~270

using 26% efficiency)

The dominant metabolic cost in neural signaling is associated with spiking (Attwell & Laughlin; Lennie).

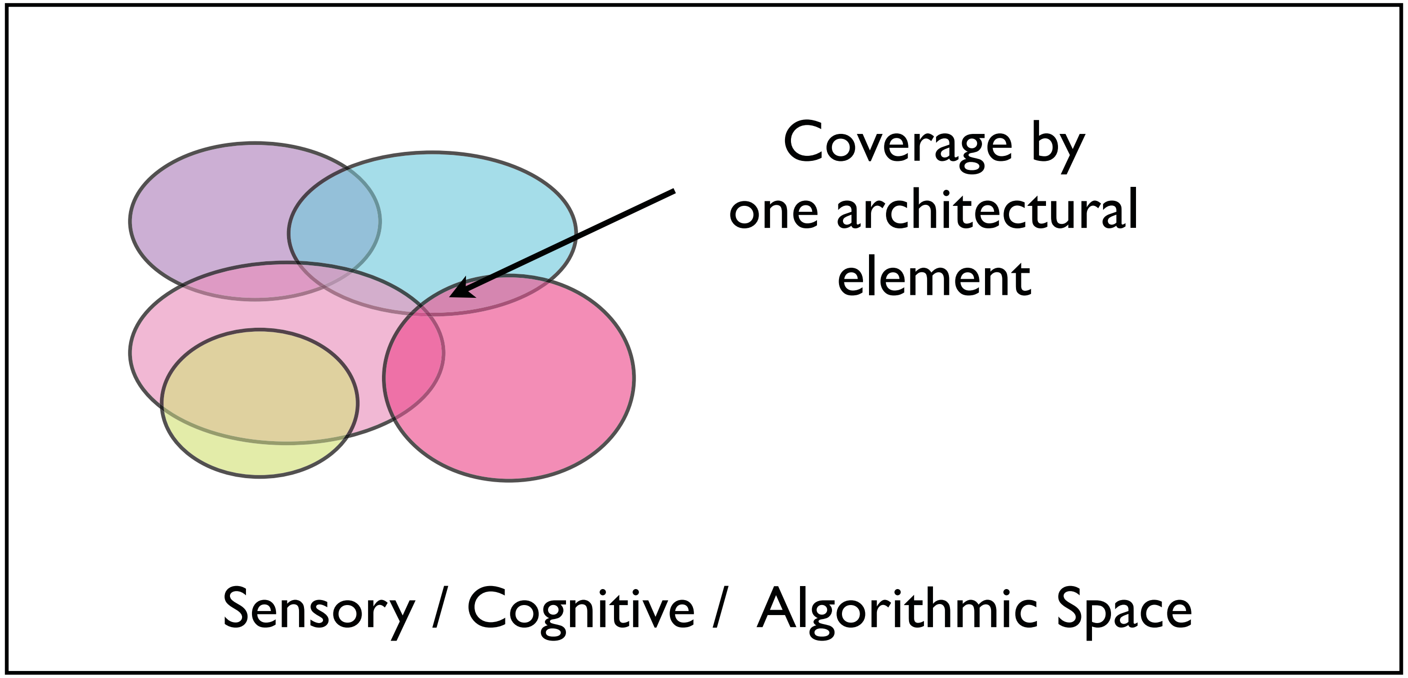
A law of diminishing returns



In any given information channel bits/energy is optimized at a particular firing rate (Balasubramanian, Berry & Kimber, 2001).

Perhaps this is related to the surprising similarity of information rates of all retinal ganglion cells (6-13 bits/s) responding the to natural stimuli.

Resource allocation in the brain



Ongoing work:

- Representation of textures in visual cortex
- Planning of paths based on hippocampal place cells
- Representation of odors by the olfactory cortex
- Representation of structured stimuli by retina

CHALLENGE: Understanding the interactions

The End

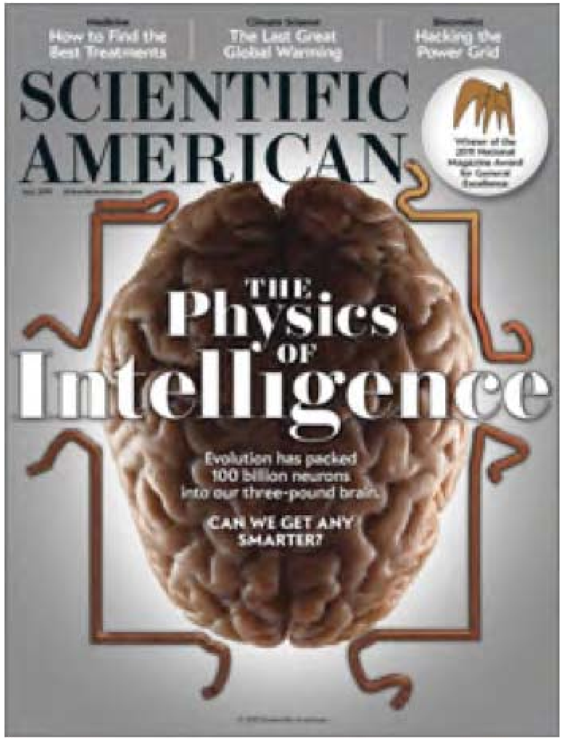
Conclusion: How smart can you get?

- *Faster neurons*: Requires disproportionate investment of resources because of the law of diminishing returns

- *More neurons:* Requires larger brains, and therefore slower communications, more power. Reducing neuron size to compensate leads to more noise and uncertainty

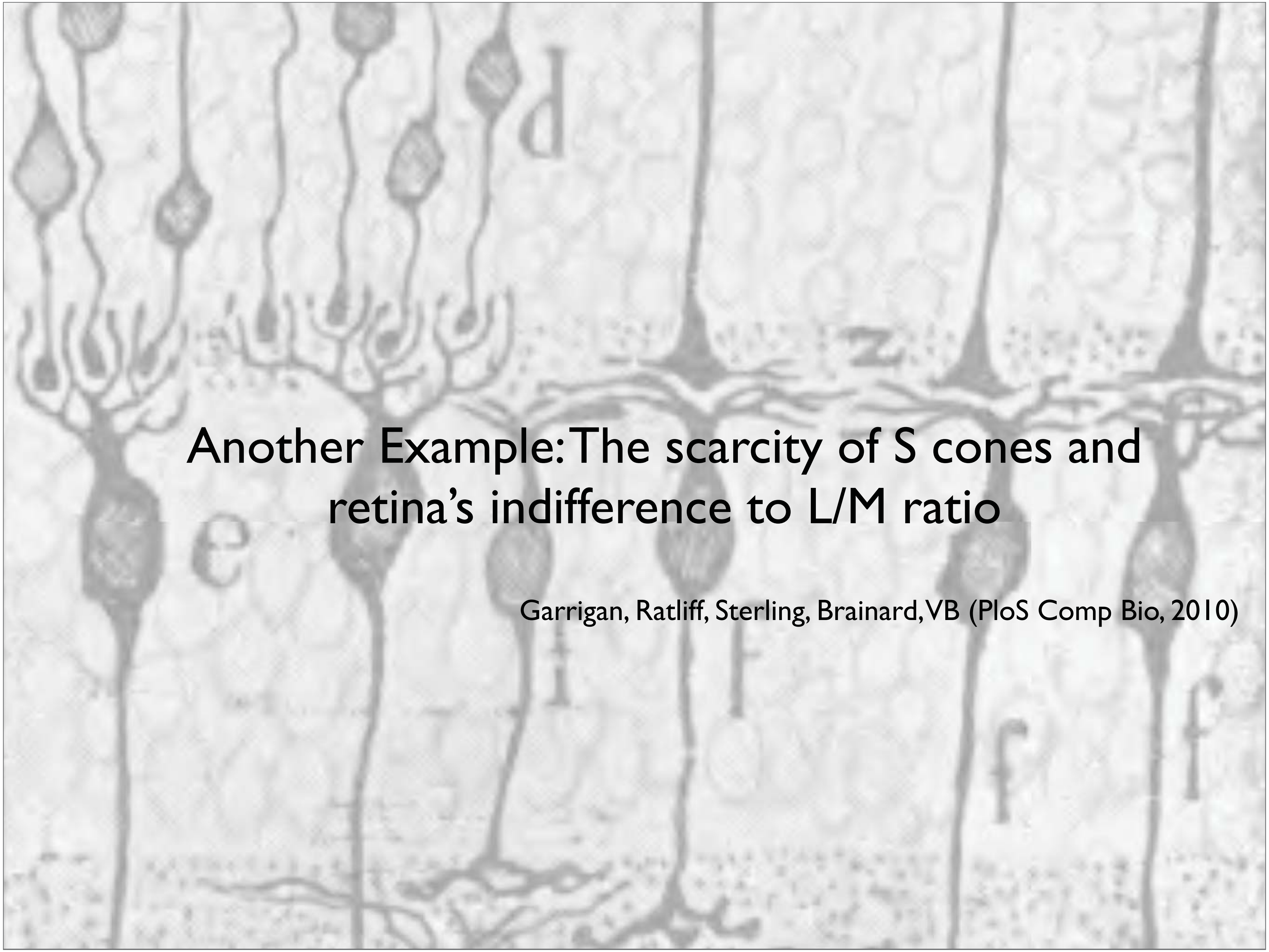
- *Communities:* Communicate with other humans specialized to other tasks and think collectively?

- *Technology:* The book, the computer, plug-in cognitive modules?

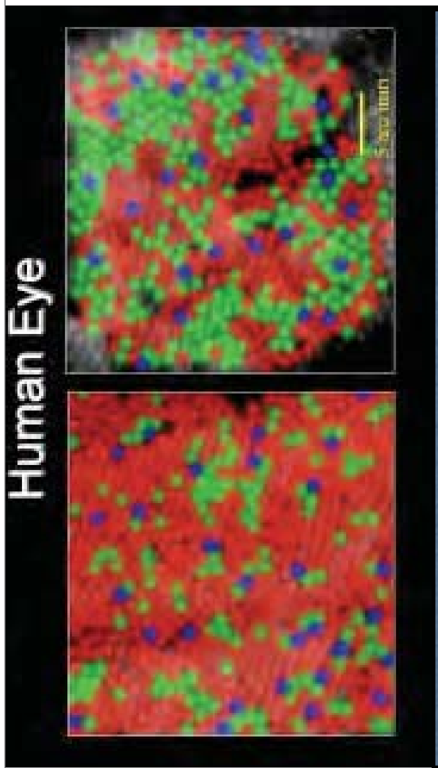


Another Example: The scarcity of S cones and retina's indifference to L/M ratio

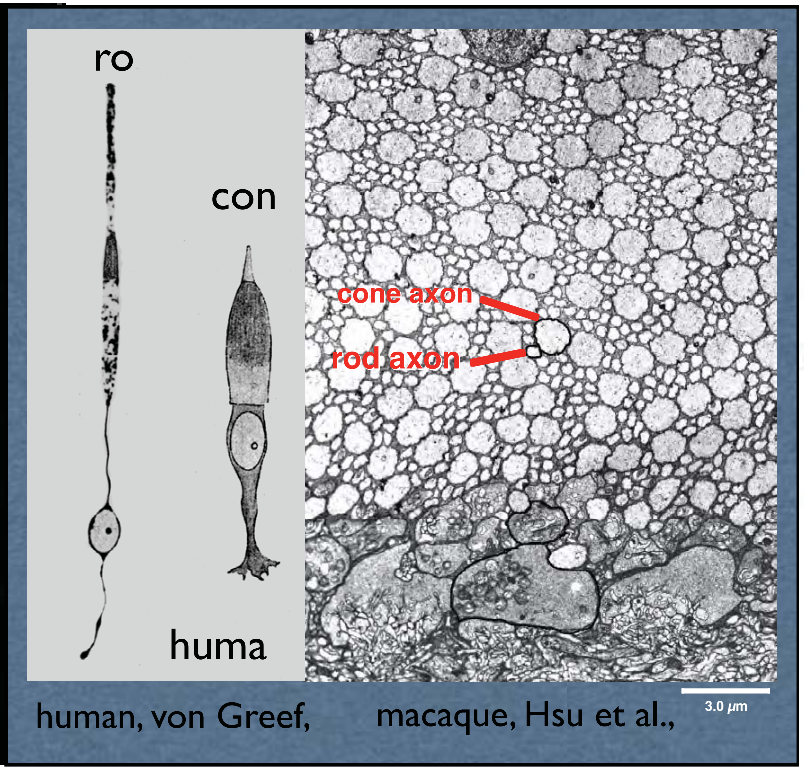
Garrigan, Ratliff, Sterling, Brainard, VB (PloS Comp Bio, 2010)

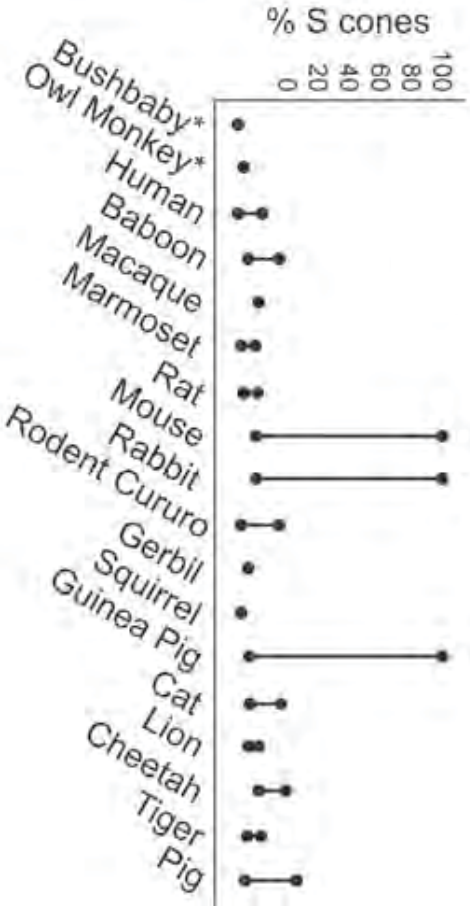


The photoreceptor distribution



Roorda, Williams

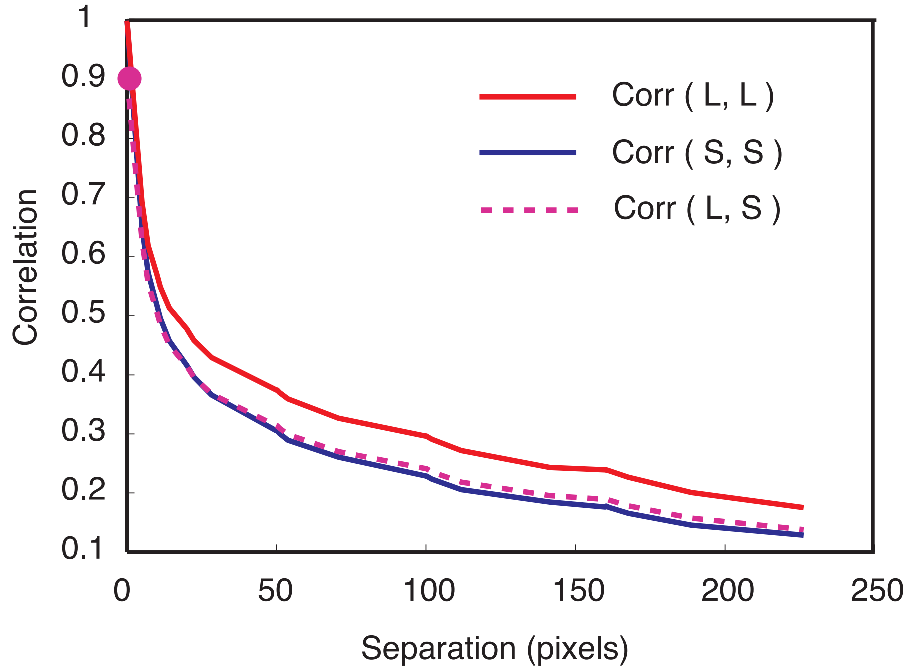


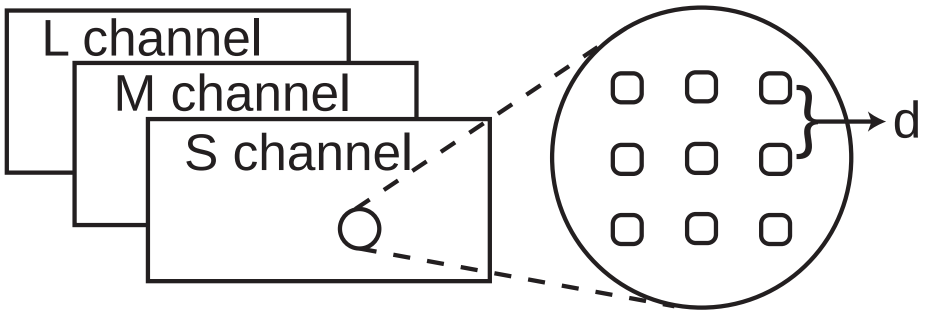


- ~10-fold variability in human L/M ratios (e.g., Hagstrom et al., 1997)
- S cones are scarce (<10%) (e.g., Curcio et al., 1991)

HYPOTHESIS: This is a consequence of maximizing chromatic information transfer from natural scenes.

Cone responses are highly correlated

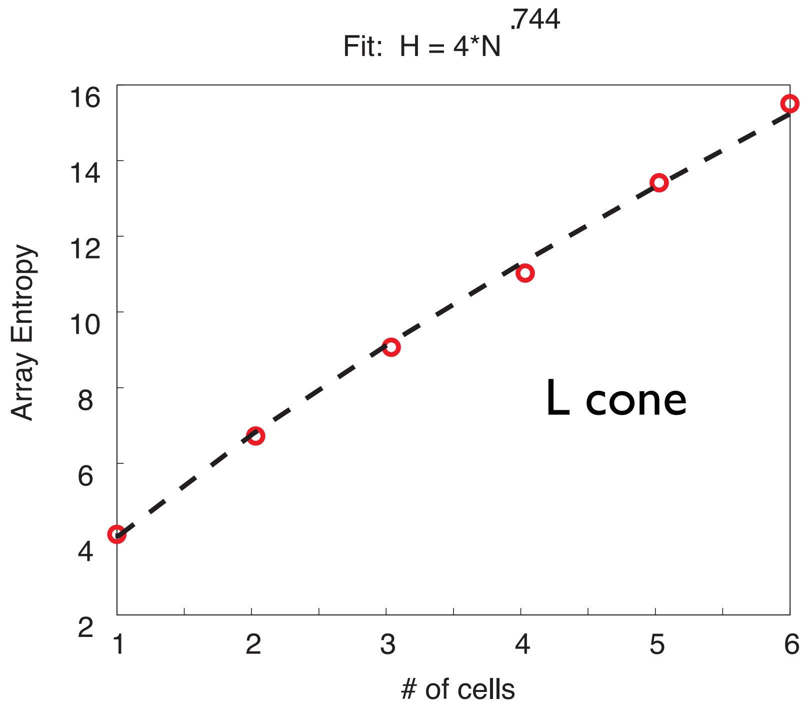


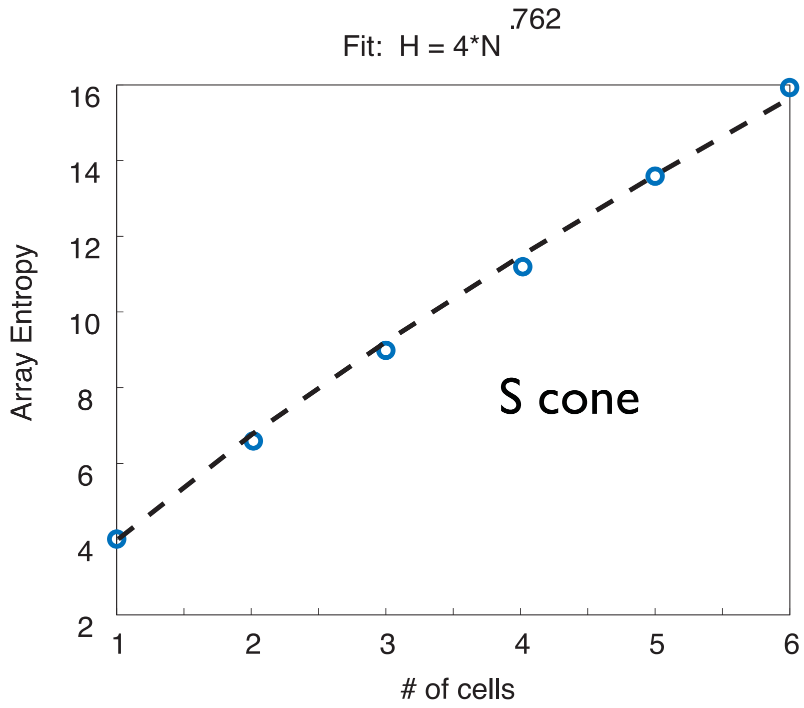


Measure the joint distribution of intensities from an *n* pixel lattice, separated by spacing *d*

$$H(x_1, x_2, ...x_n) = -\sum_{(x_1, x_2, ...x_n)} p(x_1, x_2, ...x_n)\, log\,(p(x_1, x_2, ...x_n))$$

Scaling of information in cone arrays





$$H_L = N_L^{0.74} \, H_L^1$$

$$H_S = N_S^{0.76} \, H_S^1$$

$$H_{\text{tot}} = H_L + H_S - \text{Mutual Information}$$

Let $N_S = N - N_R$ and optimize

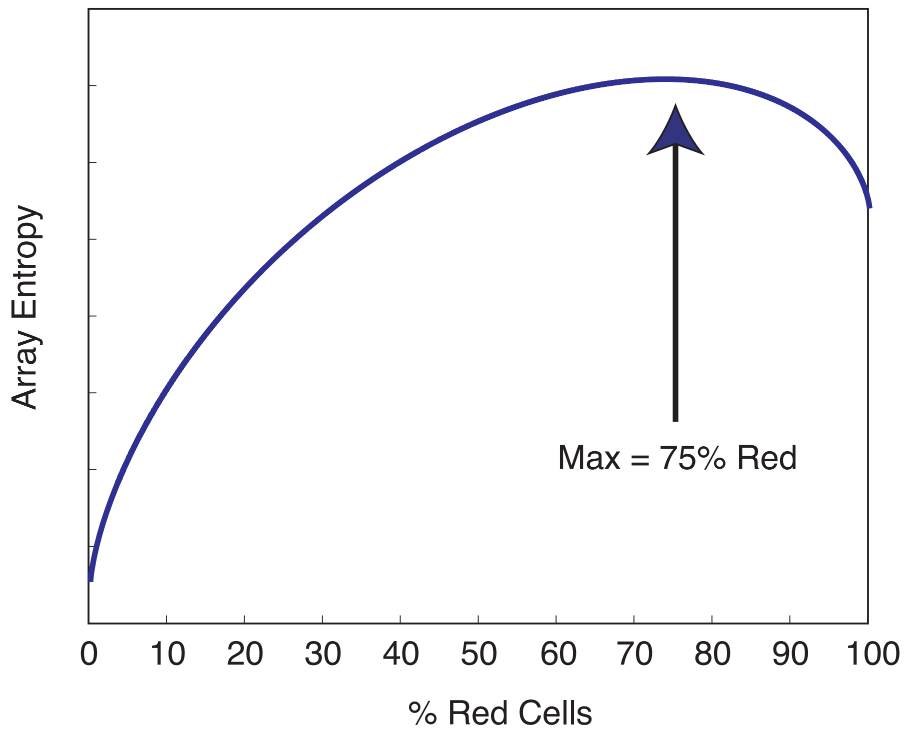
Back of the envelope estimate

- Neglect the shared information between L & S
- Set the scaling exponents of both L and S arrays to be ~ 0.75

•Then: $$H_{\text{tot}} = N_L^{0.75}\, H_L^1 + (N - N_L)^{0.75}\, H_S^1$$

- Optimize: $$\frac{\partial H_{\text{tot}}}{\partial N_L} = 0 \quad \Longrightarrow \quad \frac{N_L}{N} = \frac{1}{1 + \left(\frac{H_L^1}{H_S^1}\right)^{\frac{1}{0.75-1}}}$$

- Putting in the daylight values for the single pixel entropy:



The optimal mosaic including optical factors

- Attenuation of high frequencies by macular pigment
- Chromatic aberration by the lens
- Shared information between cone types