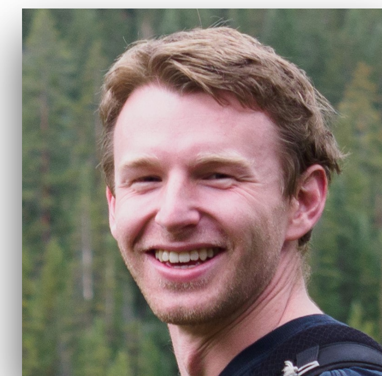
Coarse graining to capture "relevant" information in biological systems

Adam Kline
William Bialek
Thierry Mora
Vudtiwat Ngampruetikorn
Vedant Sachdeva
David Schwab
Aleksandra Walczak



S. E. Palmer
University of Chicago
Department of Organismal Biology and Anatomy
Department of Physics
Princeton Center for the Physics of Biological Function
sepalmer@uchicago.edu

Outline:

- *your eye is a computer*
- *neurons are a bit like spins*
- *connections to NPRG*
- *progress on linking biological to RG notions of “relevance”*

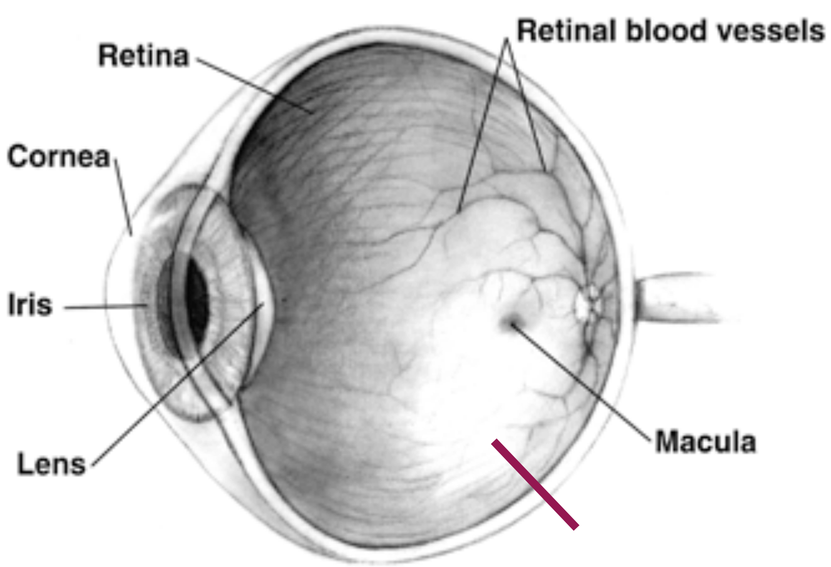
Is anything in biology “optimal”?

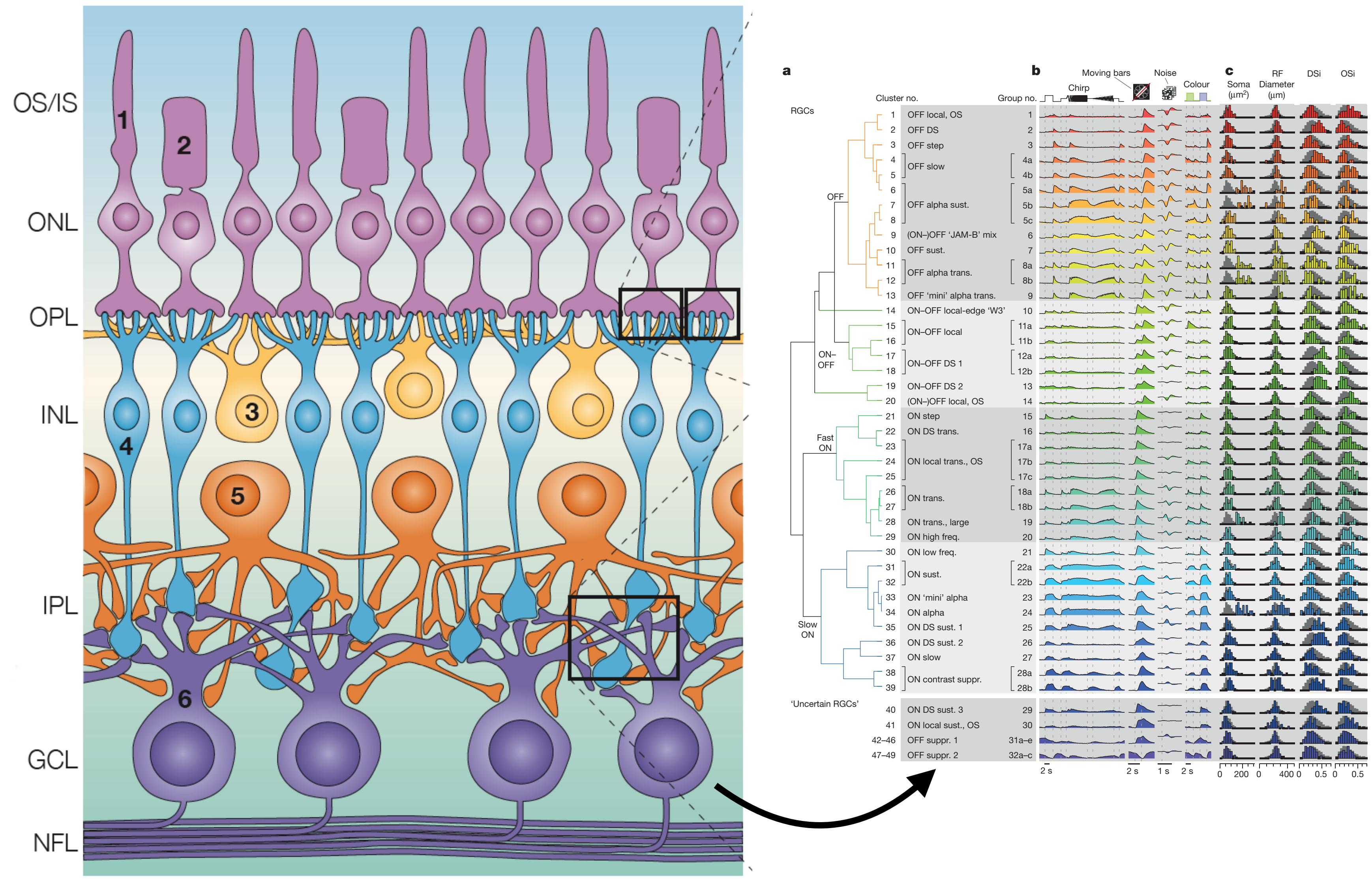
optimal coding:

max(info) + constraints

behavioral goals

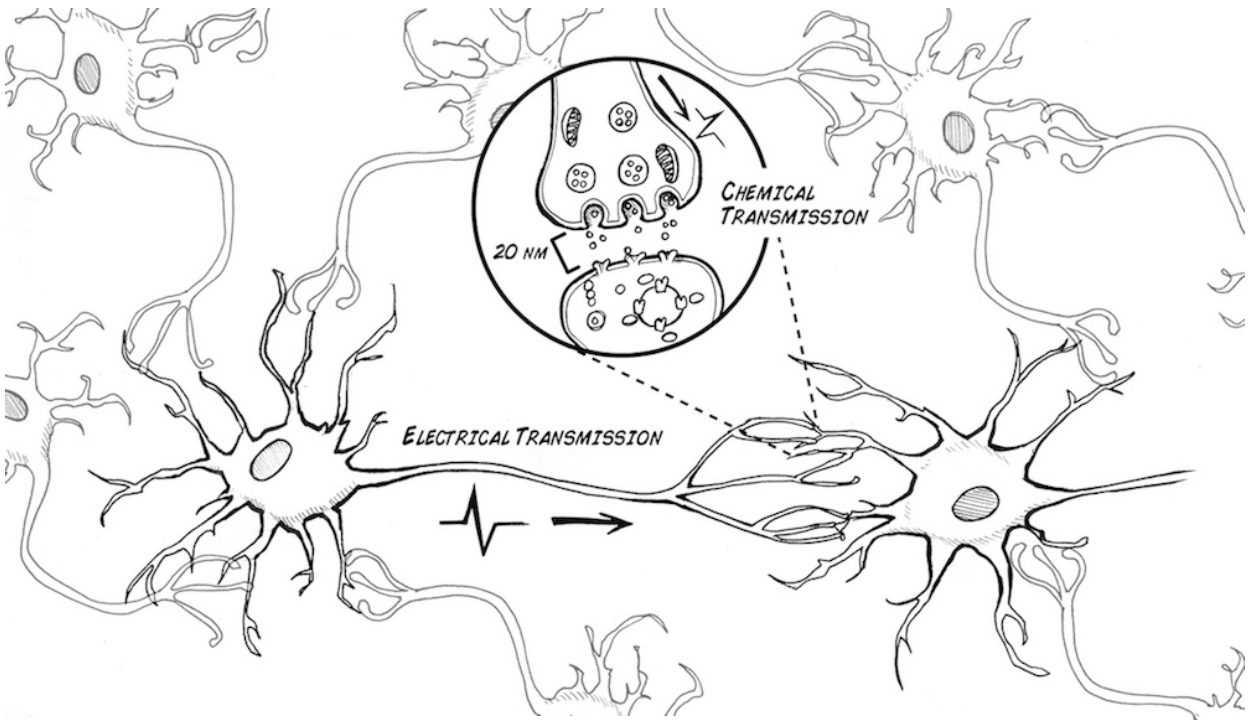
The retina is a piece of the central brain out in the eye:

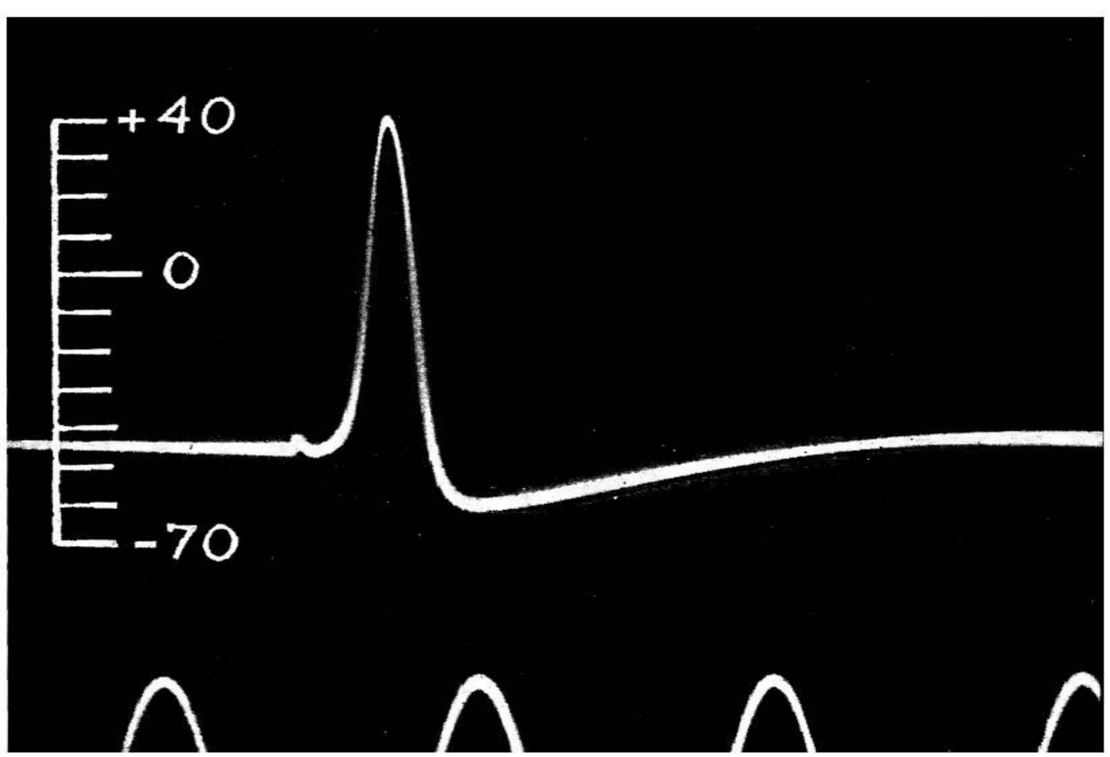




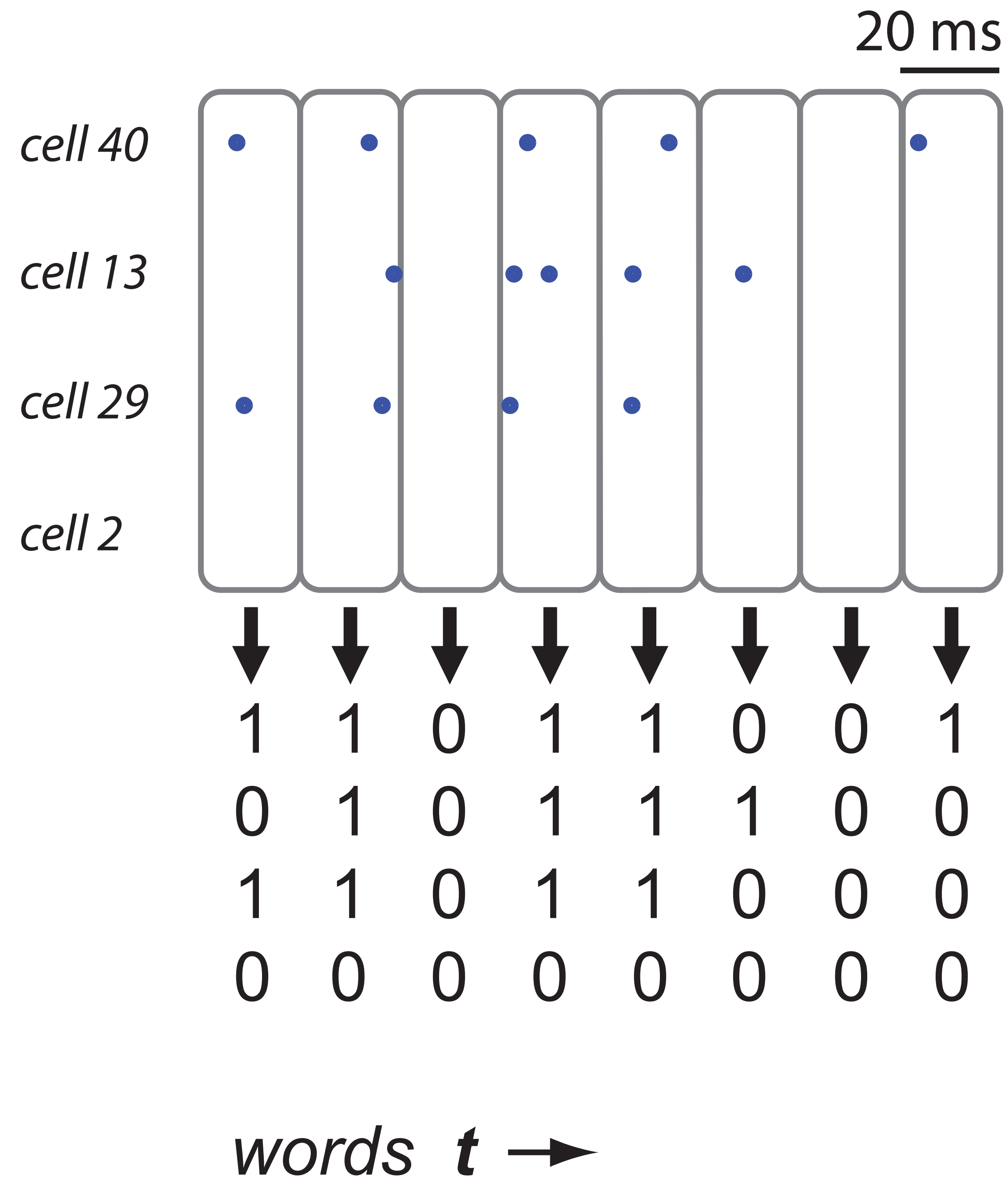
H Wässle, *Nature Reviews Neurosci* (2004)
Baden, Berens, Franke, Roson, Bethge, Euler (2016)

Neurons "speak" in binary words:

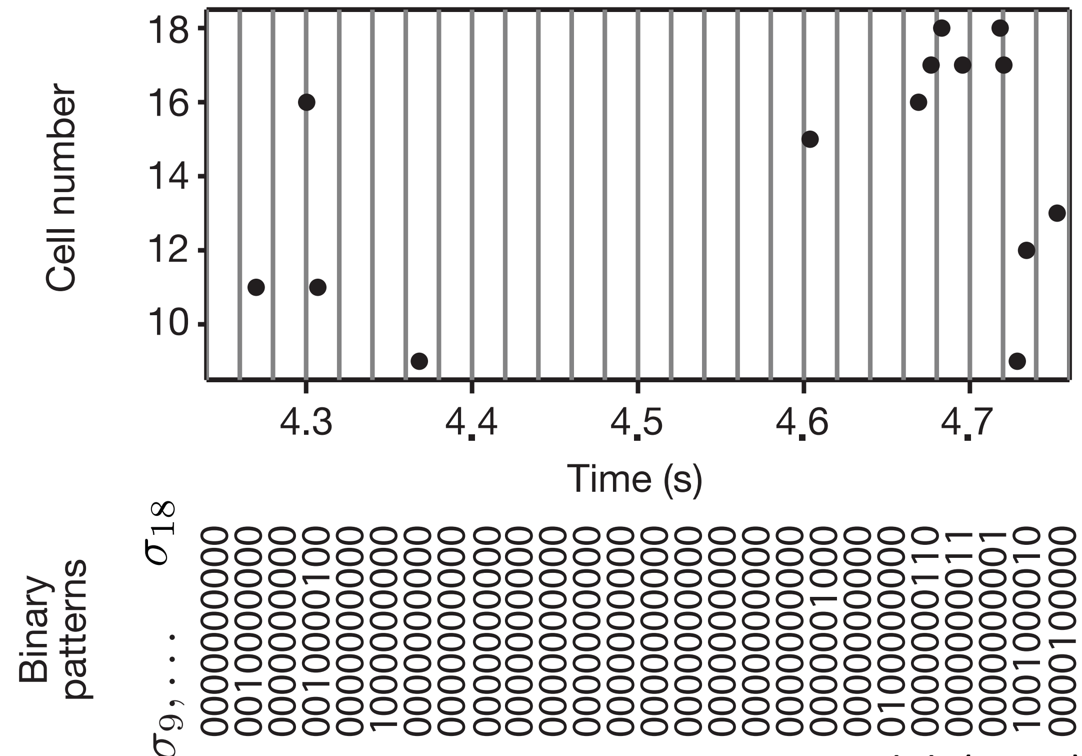




Hodgkin & Huxley (1939) Nature 144:710-711

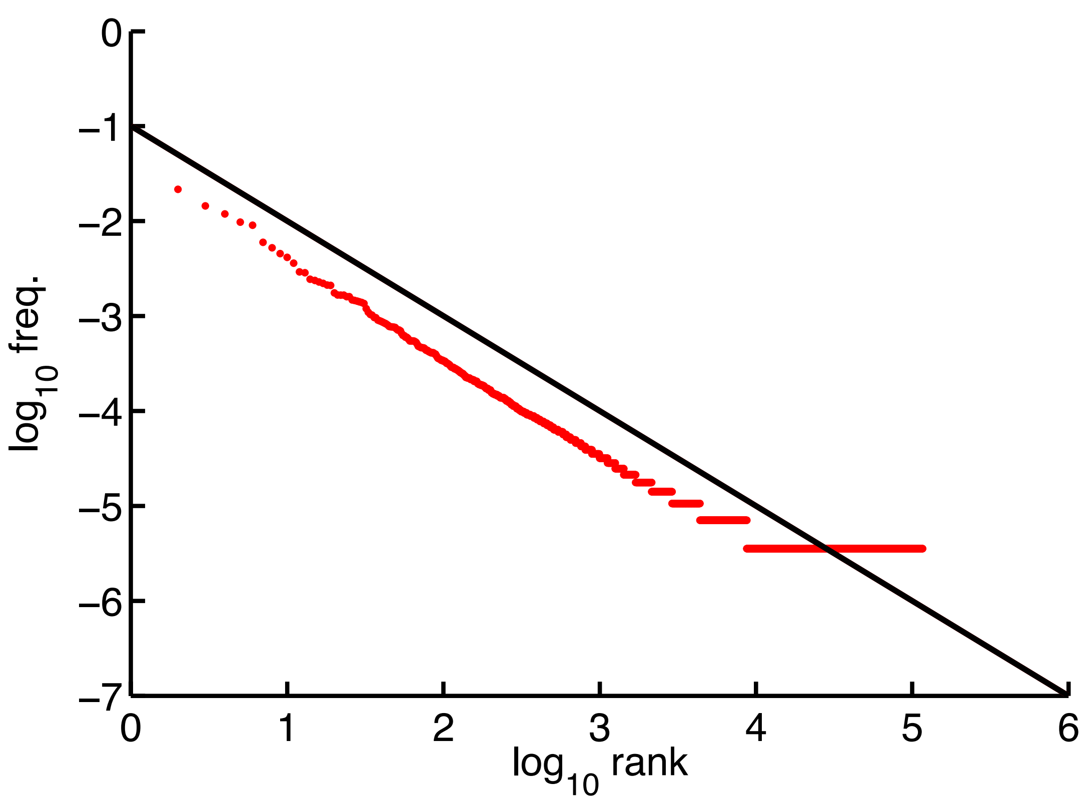


Neural word distributions are ~ power law:

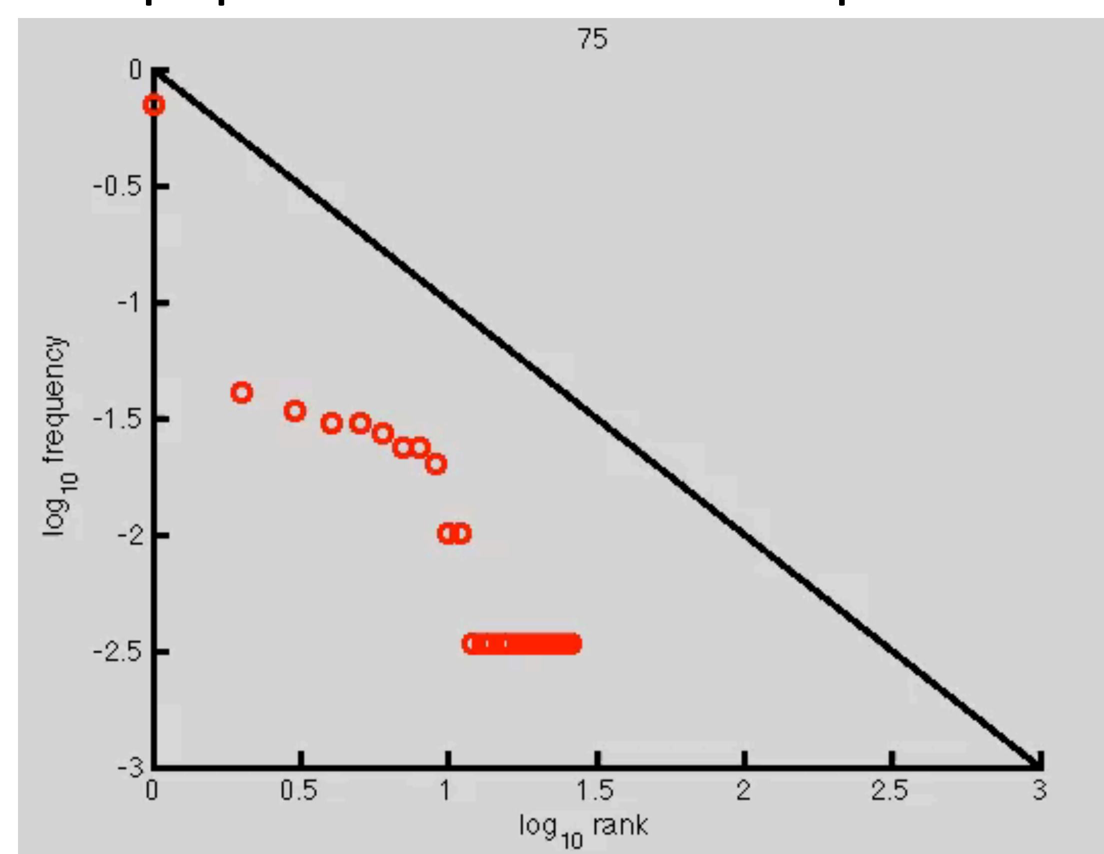


Mora, Bialek (2011)

Zipf plot of all data:



Zipf plot at fixed time in repeat:

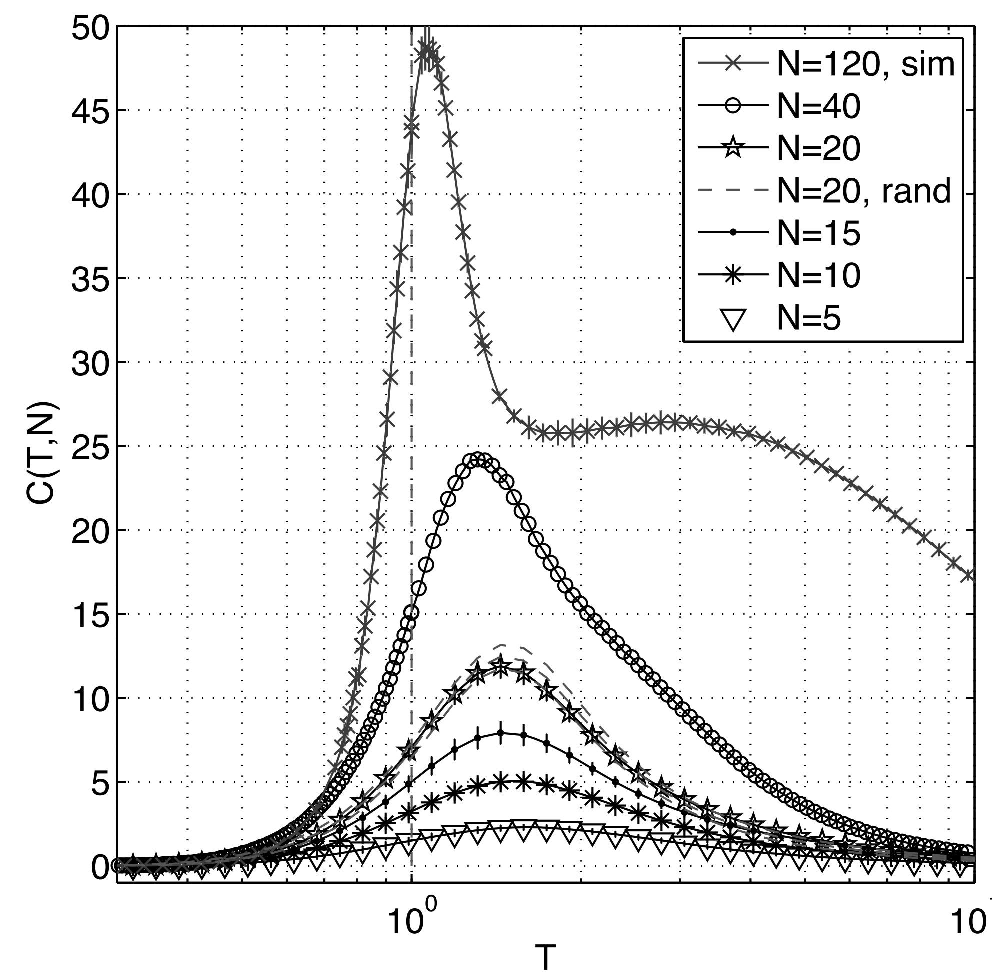


SEP + Schwab + Berry + Marre, unpublished

Standard approach to modeling neural activity: (pairwise) maximum entropy

$$P_2(\boldsymbol{\sigma}) = \frac{1}{Z} e^{-E(\boldsymbol{\sigma})}$$

$$E(\boldsymbol{\sigma}) = -\sum_{i=1}^{N} h_i \sigma_i - \sum_{i<j} J_{ij} \sigma_i \sigma_j$$



Find J_{ij} that reproduces observed pairwise correlations $\langle \sigma_i \sigma_j \rangle$

Introduce fictitious temperature to calculate specific heat - "real" system sits at T=1

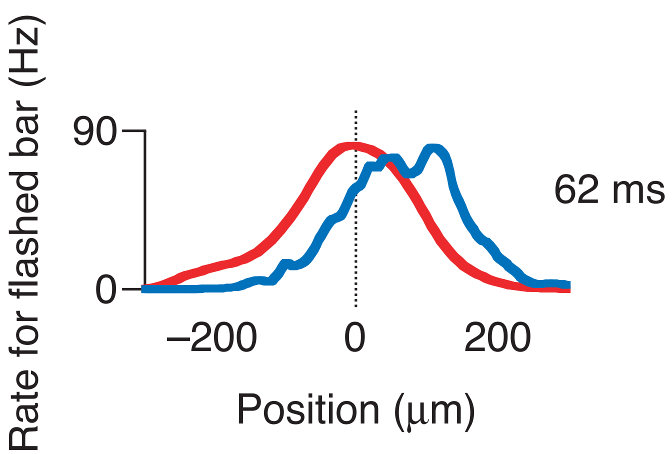
Mora + Bialek PRL (2011)

Tkačik, Mora, Marre, Amodei, SEP, Berry, Bialek PNAS (2015)

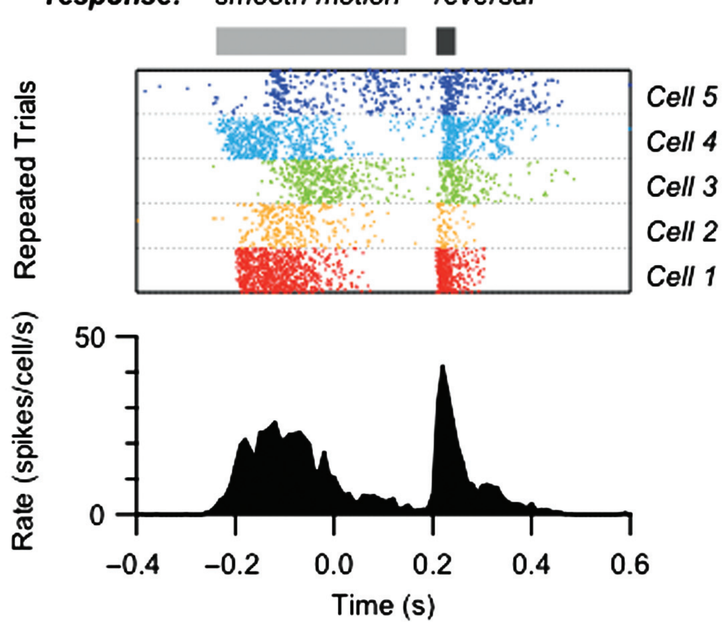
Processing delays mean the brain has to make predictions:

The retina performs a lot of complex computations:

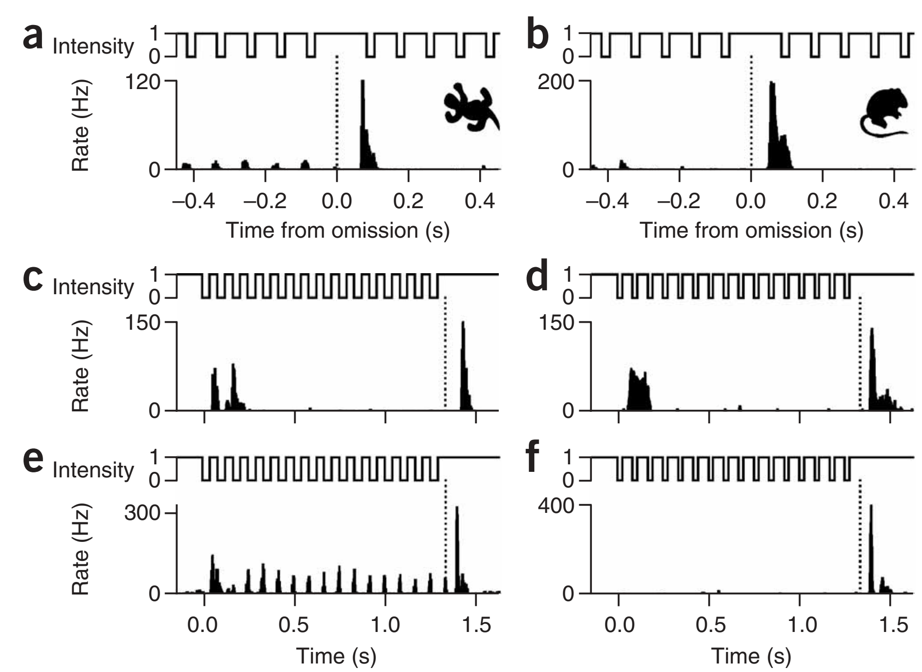
Motion anticipation



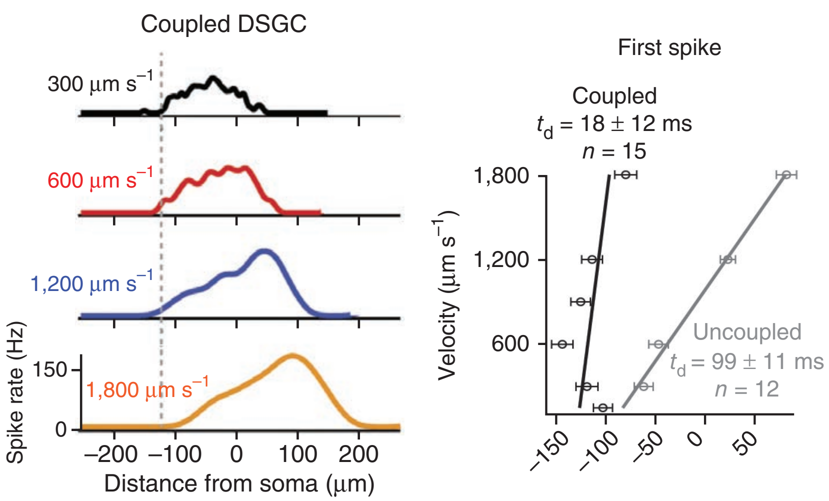
Reversal response

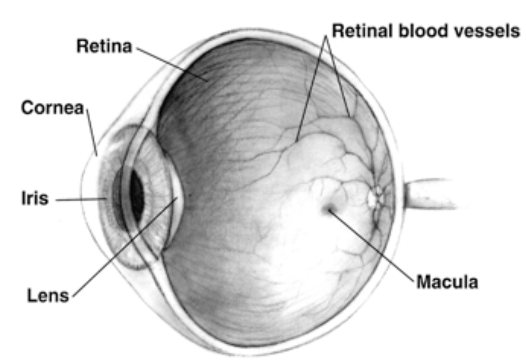


Omitted stimulus response

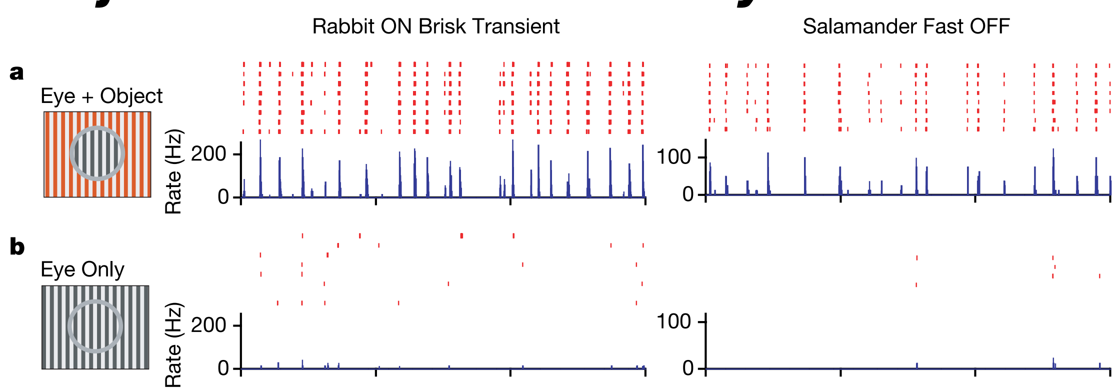


Lag normalization

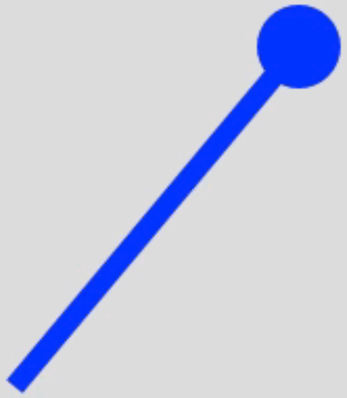




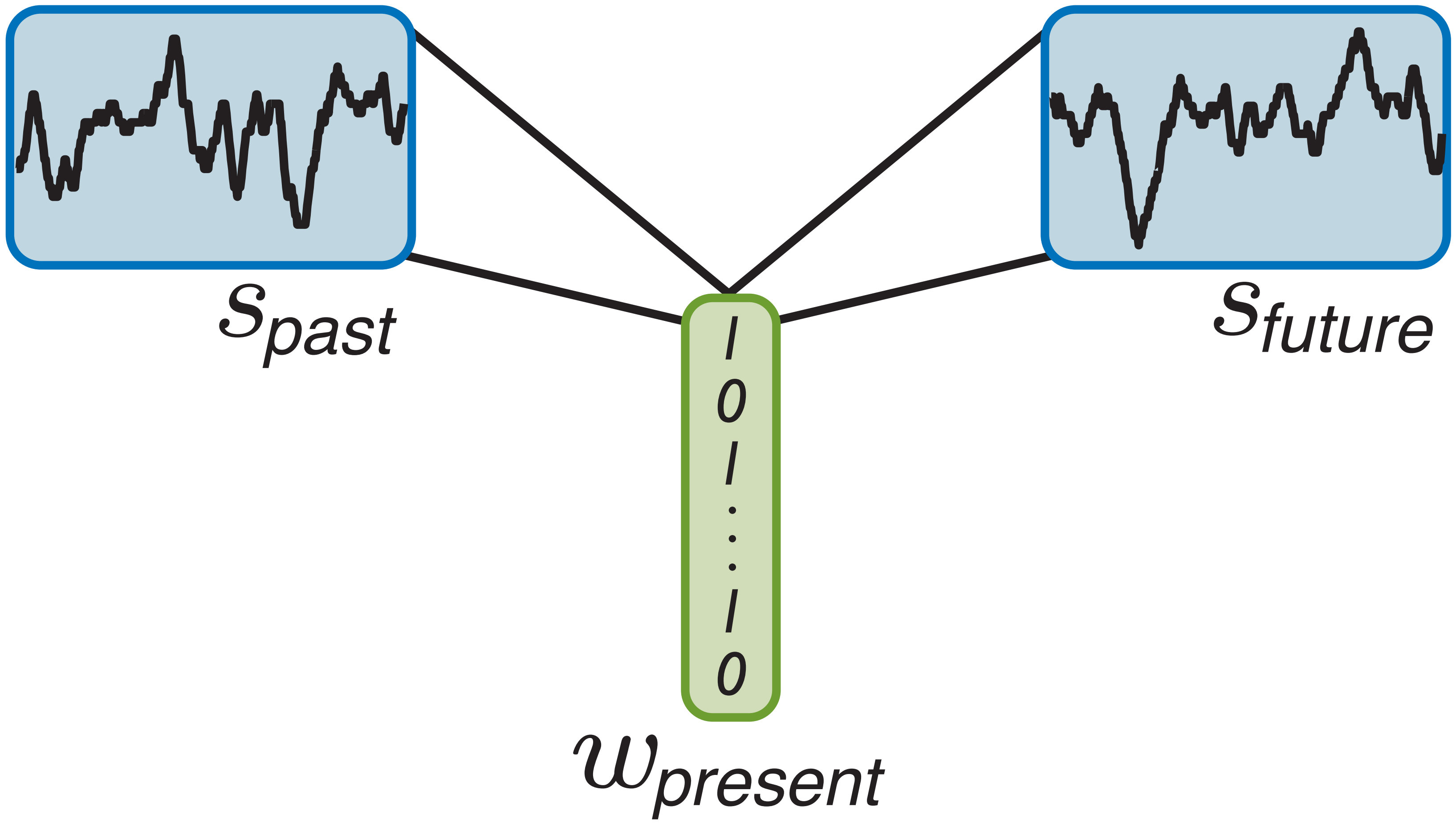
Object motion sensitivity



Berry, Brivanlou, Jordan, and Meister (1999)
Olveczky, Baccus, Meister (2003)
Schwartz, Taylor, Fisher, Harris, Berry (2007)
Schwartz, Harris, Shrom, Berry (2007)
Trenholm, Schwab, Balasubramanian, Awatramani (2013)

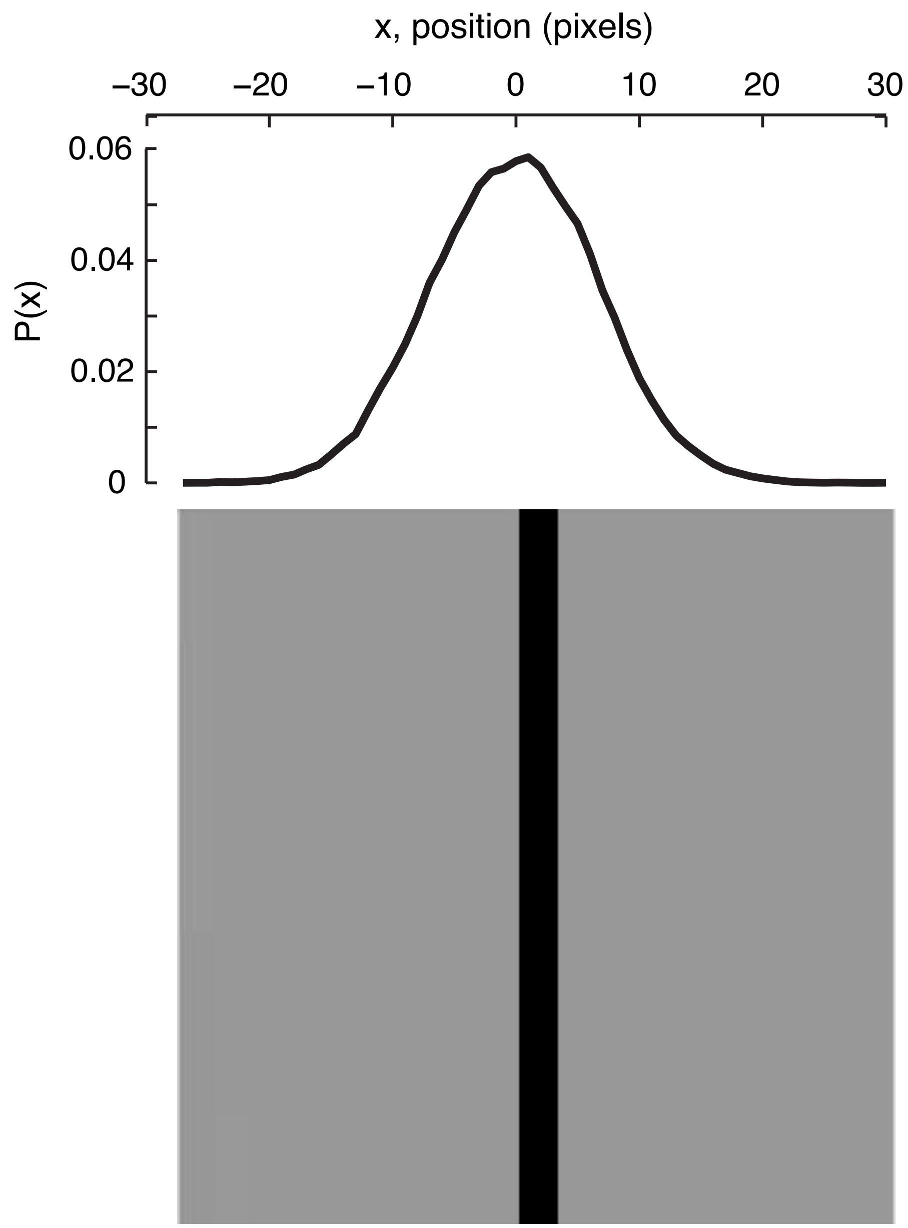


The Information Bottleneck tradeoff:

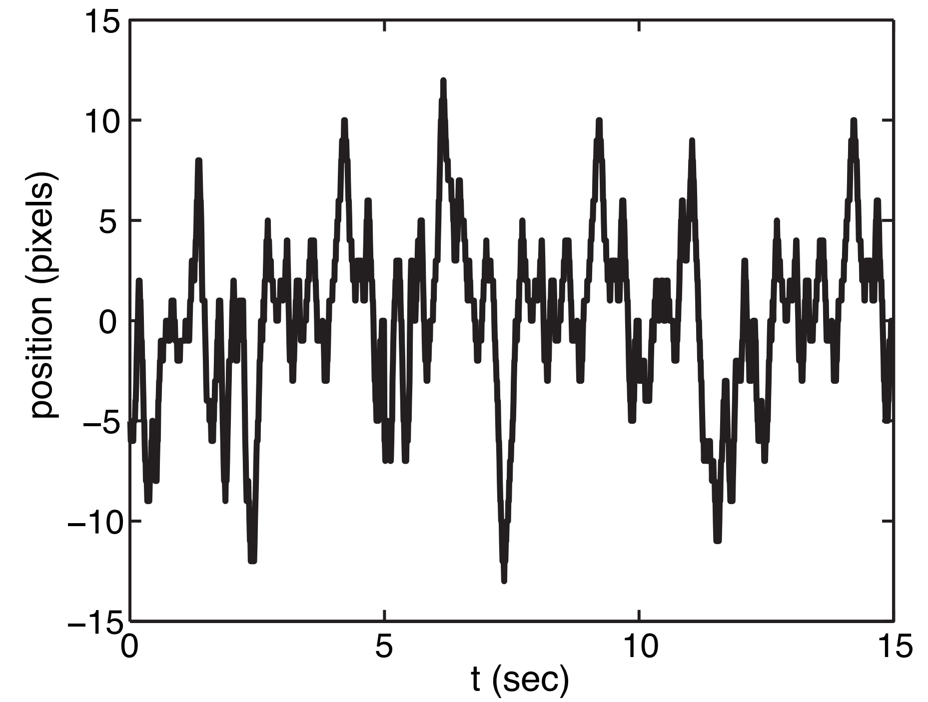


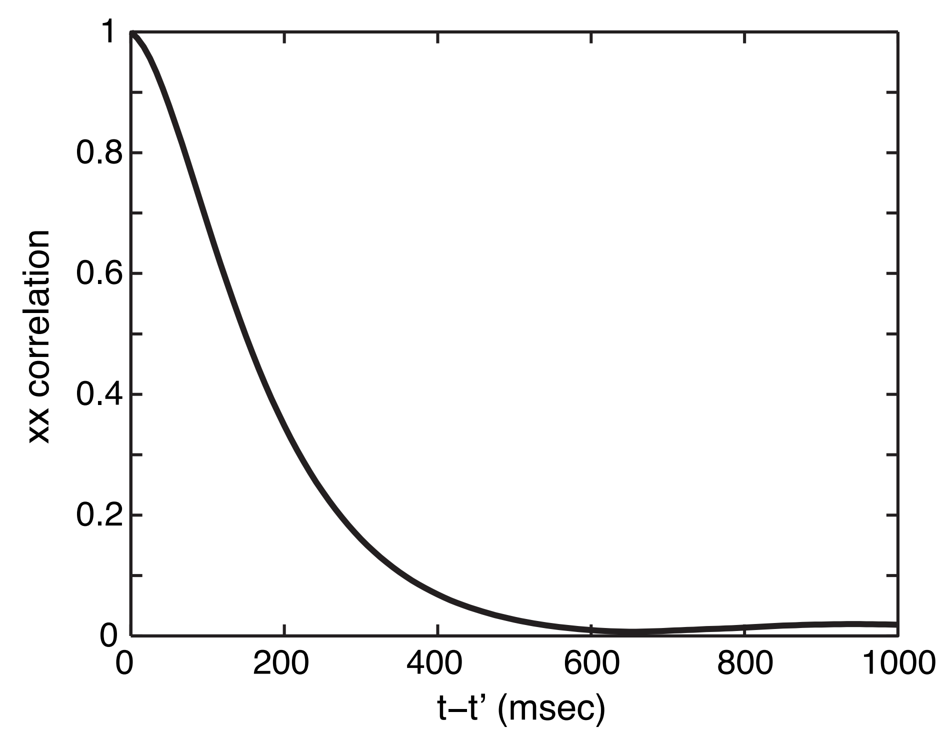
$$L = I_{past}(W_t; \vec{S}_{t-\Delta t}) - \beta I_{future}(W_t; \vec{S}_{t+\Delta t})$$

A bar stimulus with both predictable and non-predictable motion components:

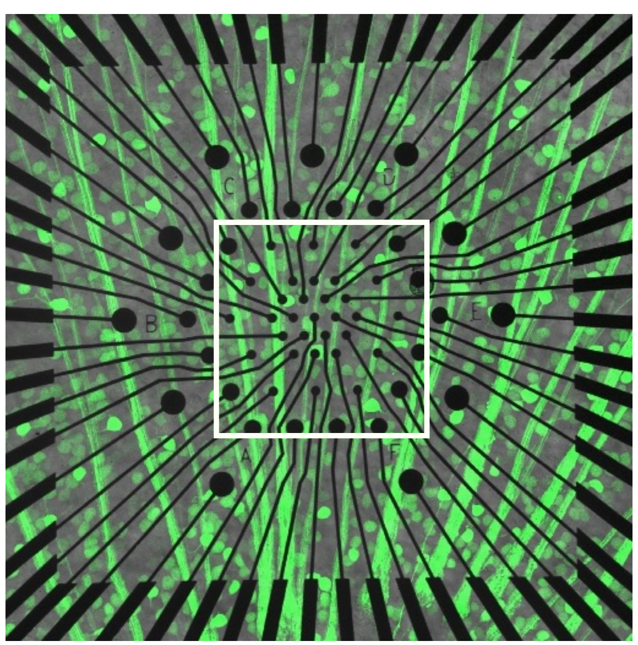


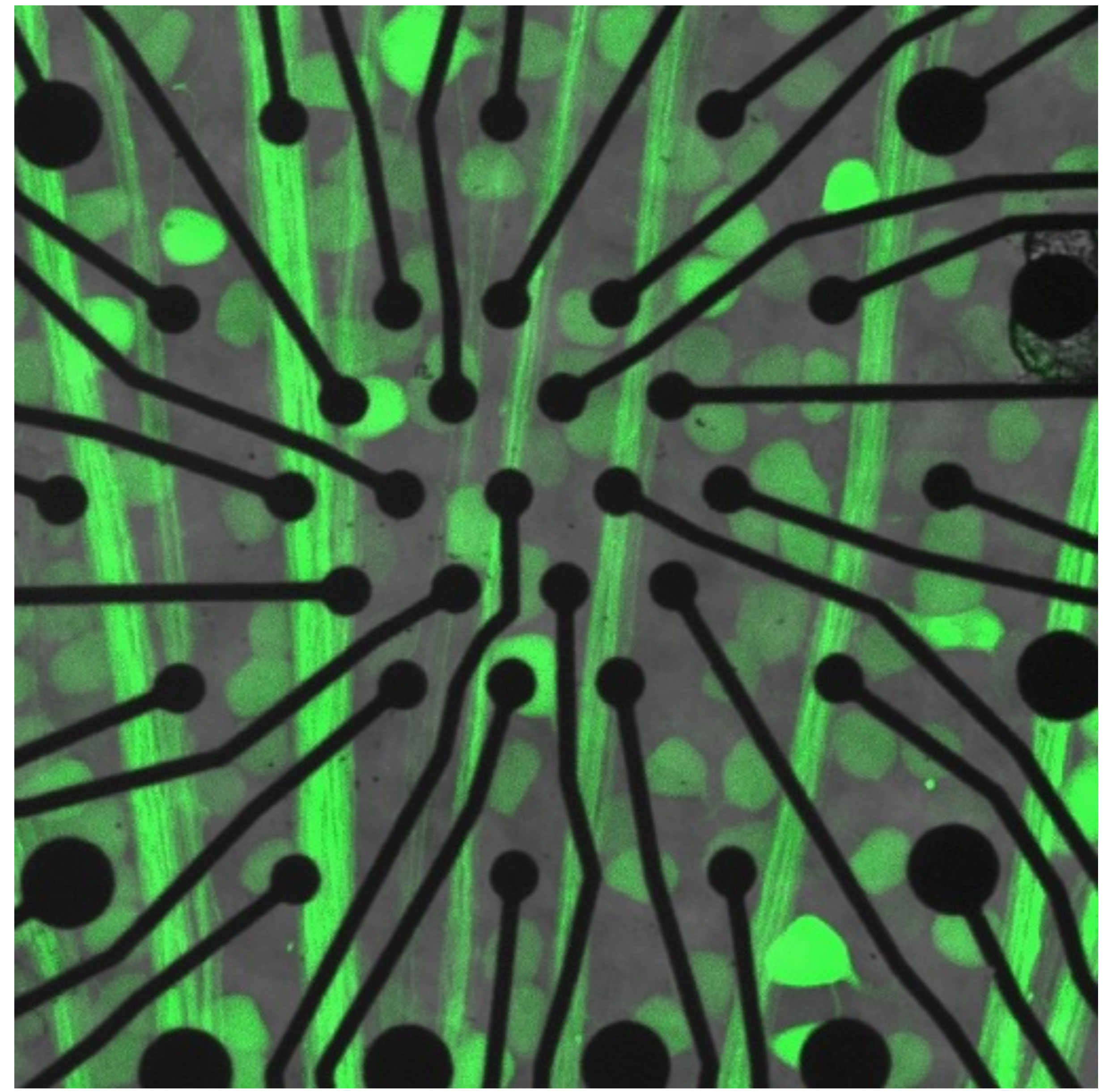
$$\frac{dv}{dt} = -\frac{v}{\tau} + D^{1/2}\Gamma(t) - \omega_0^2 x$$





Recording from the salamander retina:



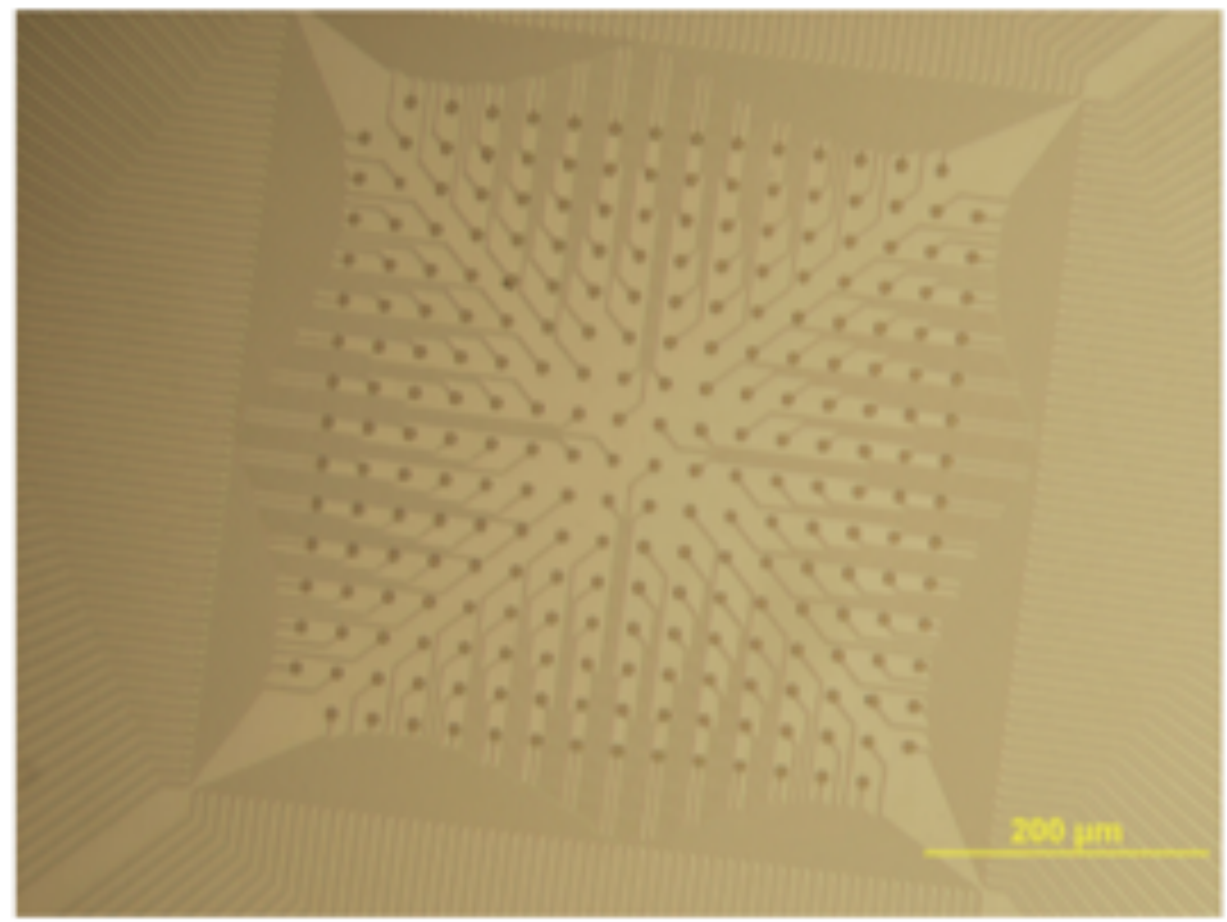


$30\mu m$

images courtesy of Ronen Segev

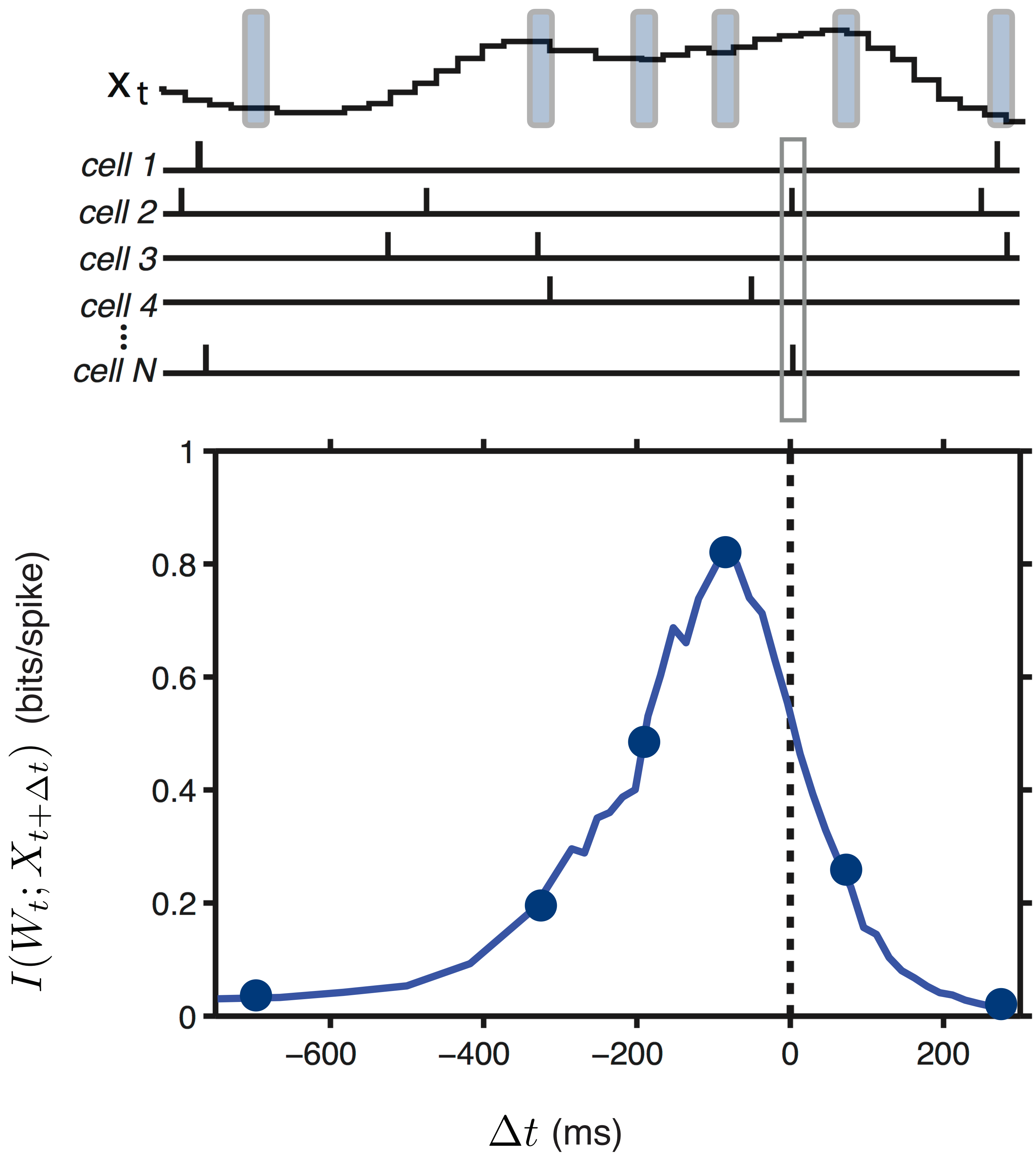
Recording from the rat retina:

wired.co.uk

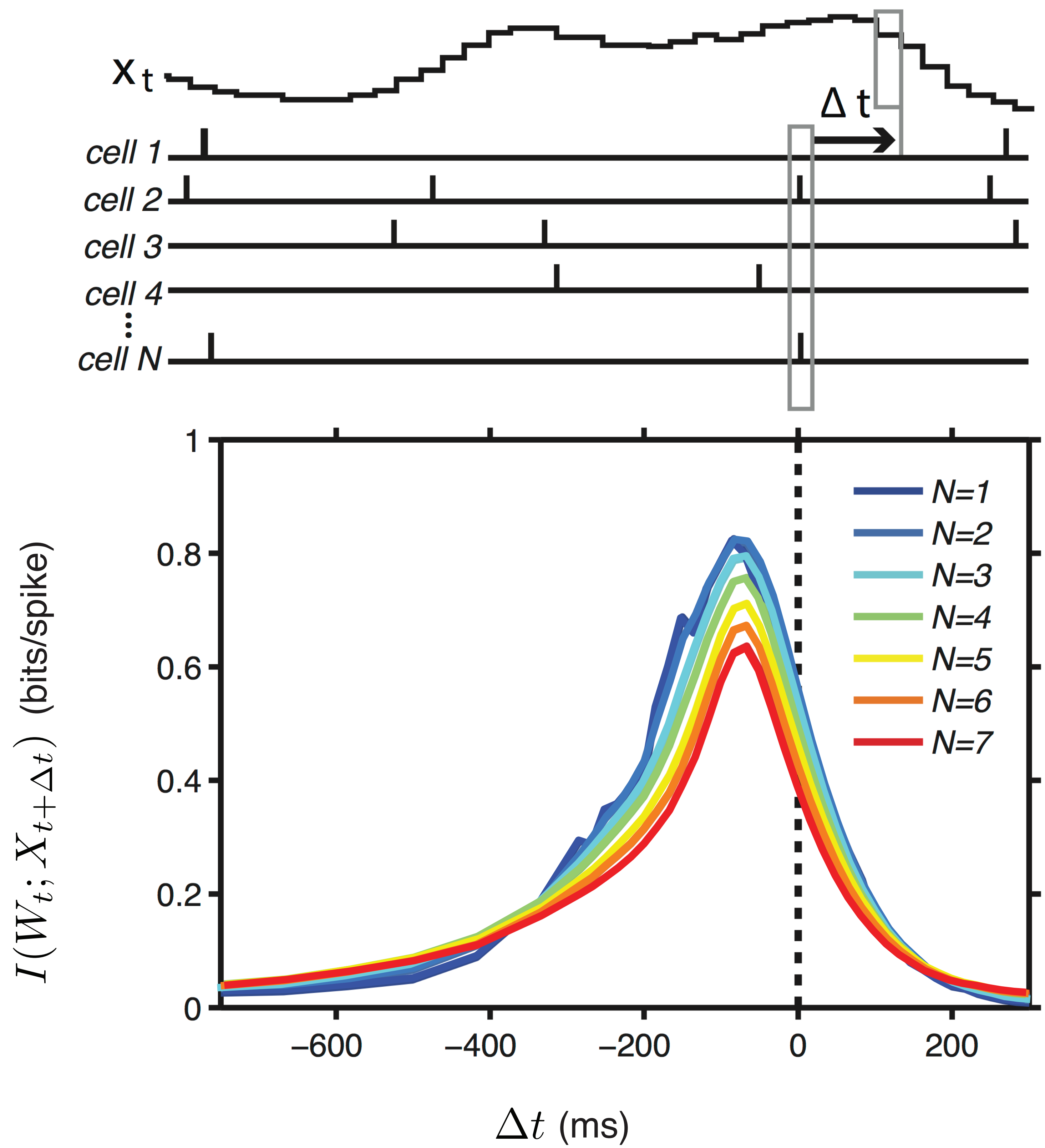


data collected with the Marre Lab, Paris, France

Retina populations carry info about the future:

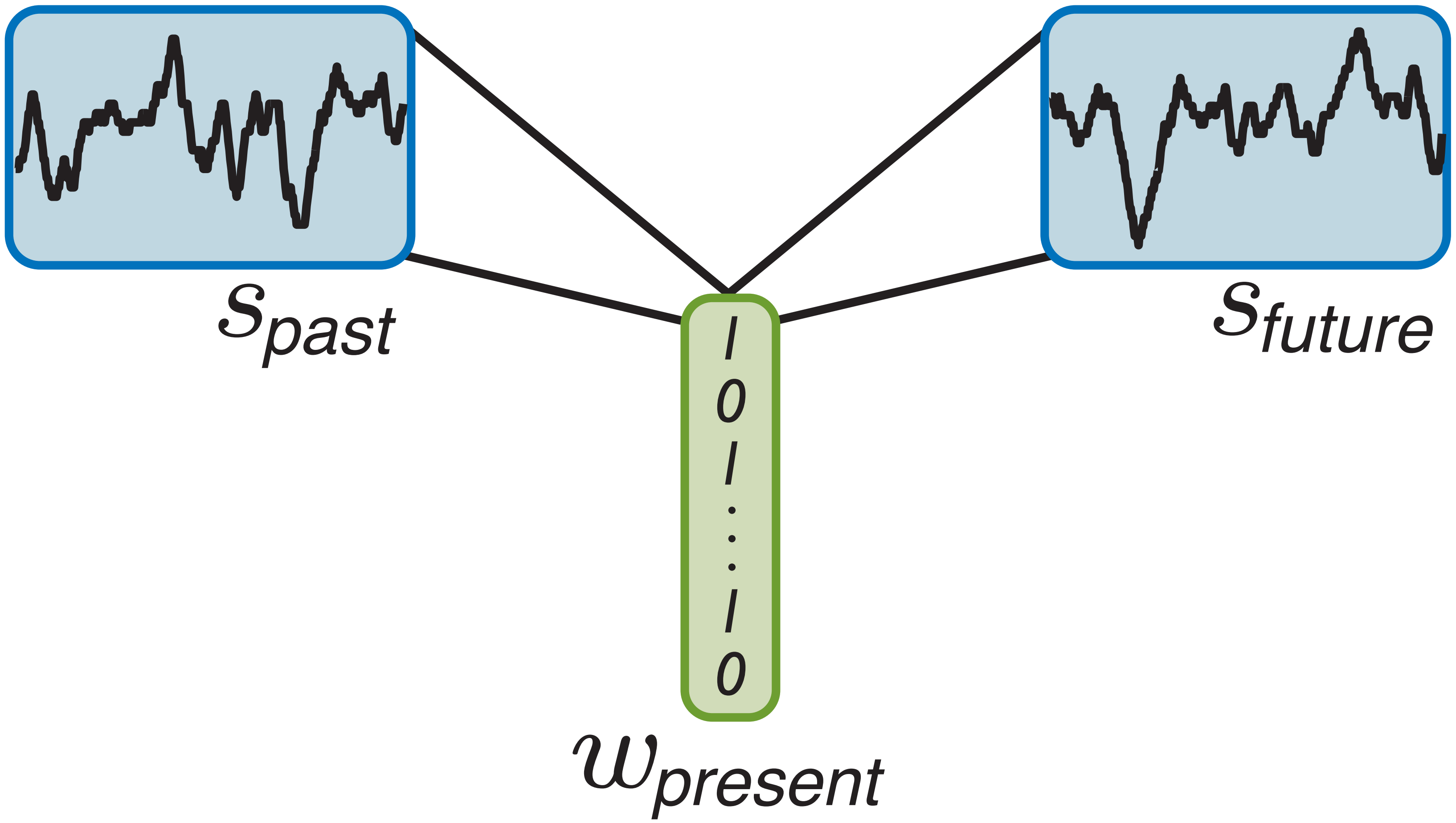


Retina populations carry info about the future:



SEP, Marre, Berry, Bialek PNAS (2015)
Salisbury + SEP J Stat. Phys. (2016)

Adding computational goals to optimal coding:



$$L = I_{past}(W_t; \vec{S}_{t-\Delta t}) - \beta I_{future}(W_t; \vec{S}_{t+\Delta t})$$

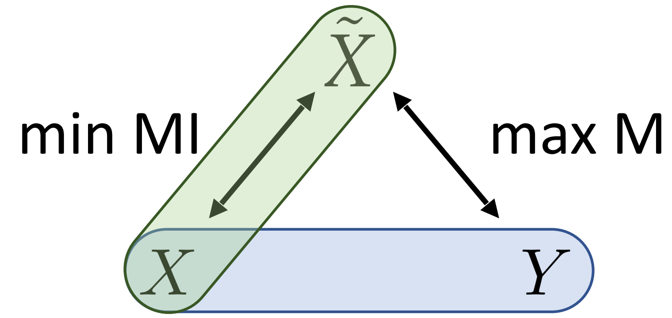
behavioral goals

Gaussian IB is analytically tractable:

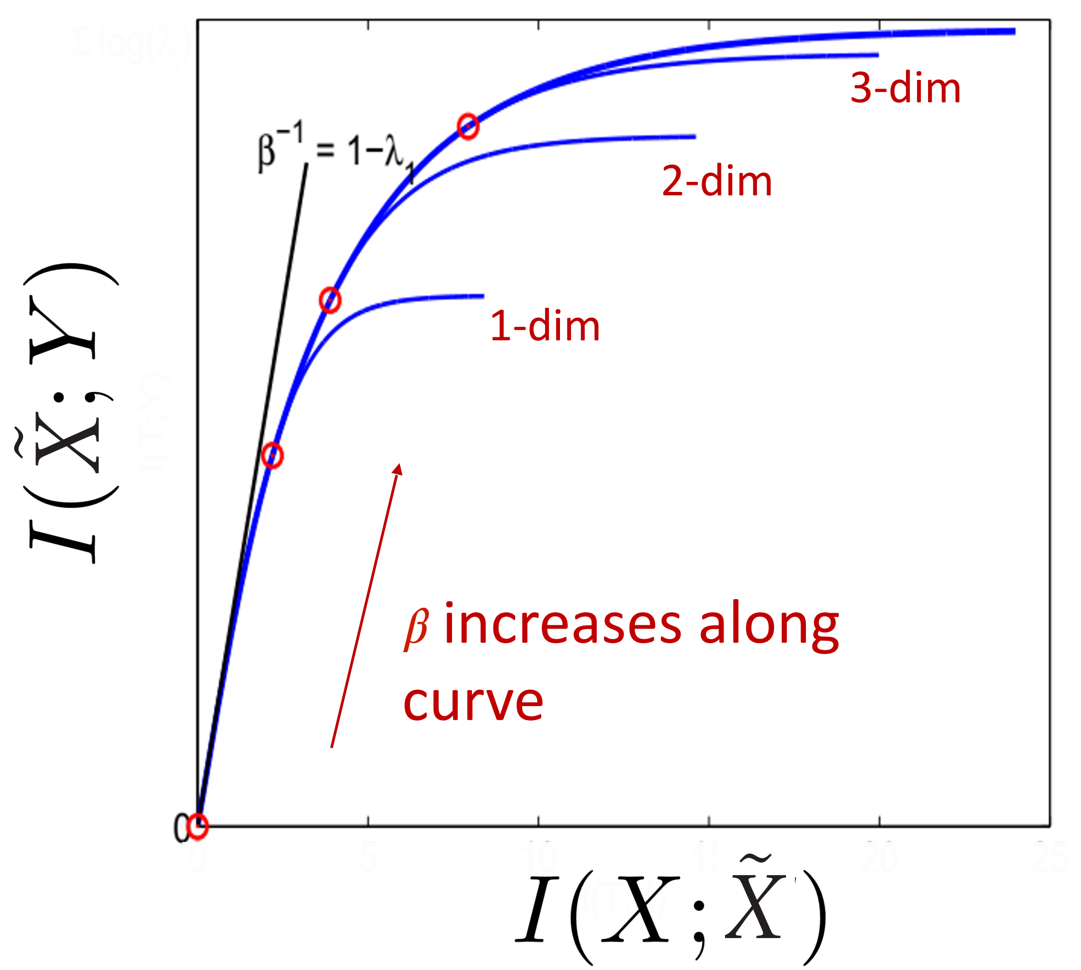
$$\min_{P(\tilde{X}|X)} I(X;\tilde{X}) - \beta I(\tilde{X};Y)$$

Start with $P(X,Y)$

Solve for $P_\beta(X,\tilde{X})$



How you coarse-grain a signal X depends on what you wish to recover from the coarse-graining



Gaussian $P(X,Y)$ is exactly solvable!

$$\tilde{X} = A\vec{X} + \vec{\xi}$$

Linear projection + noise

$$= \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_M \end{bmatrix} \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \\ \vec{v}_M \end{bmatrix} \vec{X} + \vec{\xi}$$

Weights depend on IB scale β

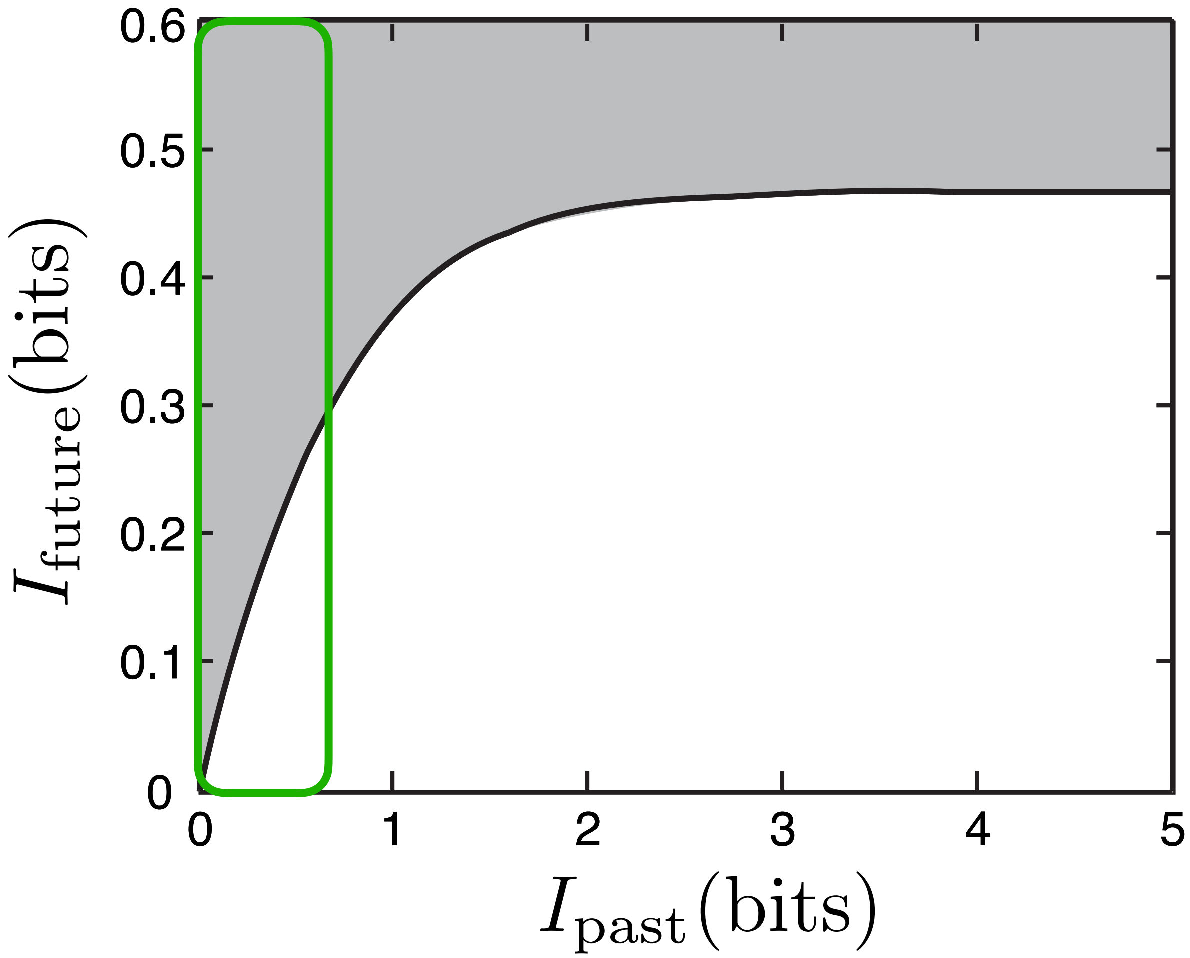
“Collective modes”

Unit variance, zero mean, Gaussian

Tishby, Pereira, Bialek (1999)
Chechik, Globerson, Tishby, Weiss (2005)

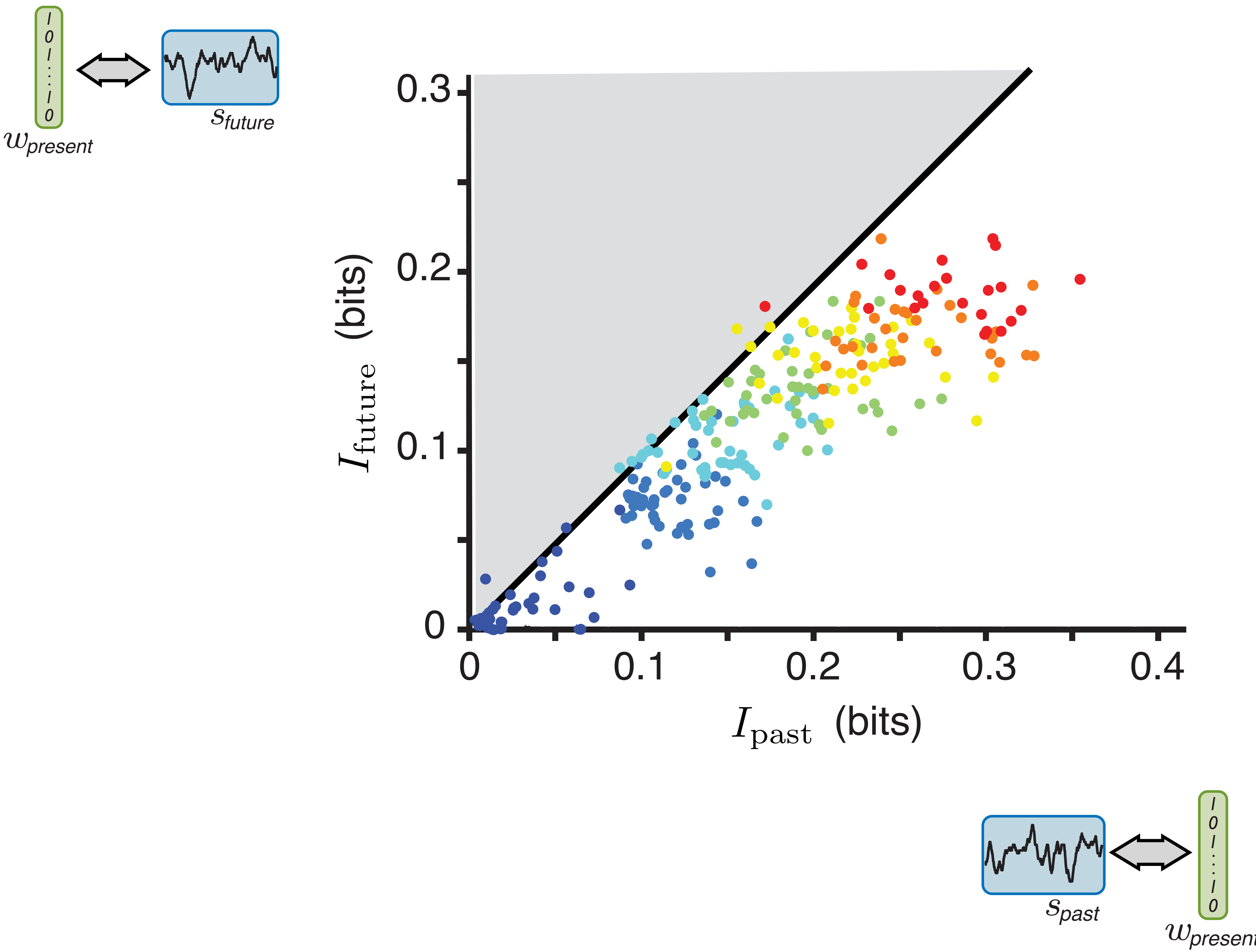
Measuring optimal prediction in the retina:

$$\mathcal{L} = I_{\text{past}} - \beta I_{\text{future}}$$

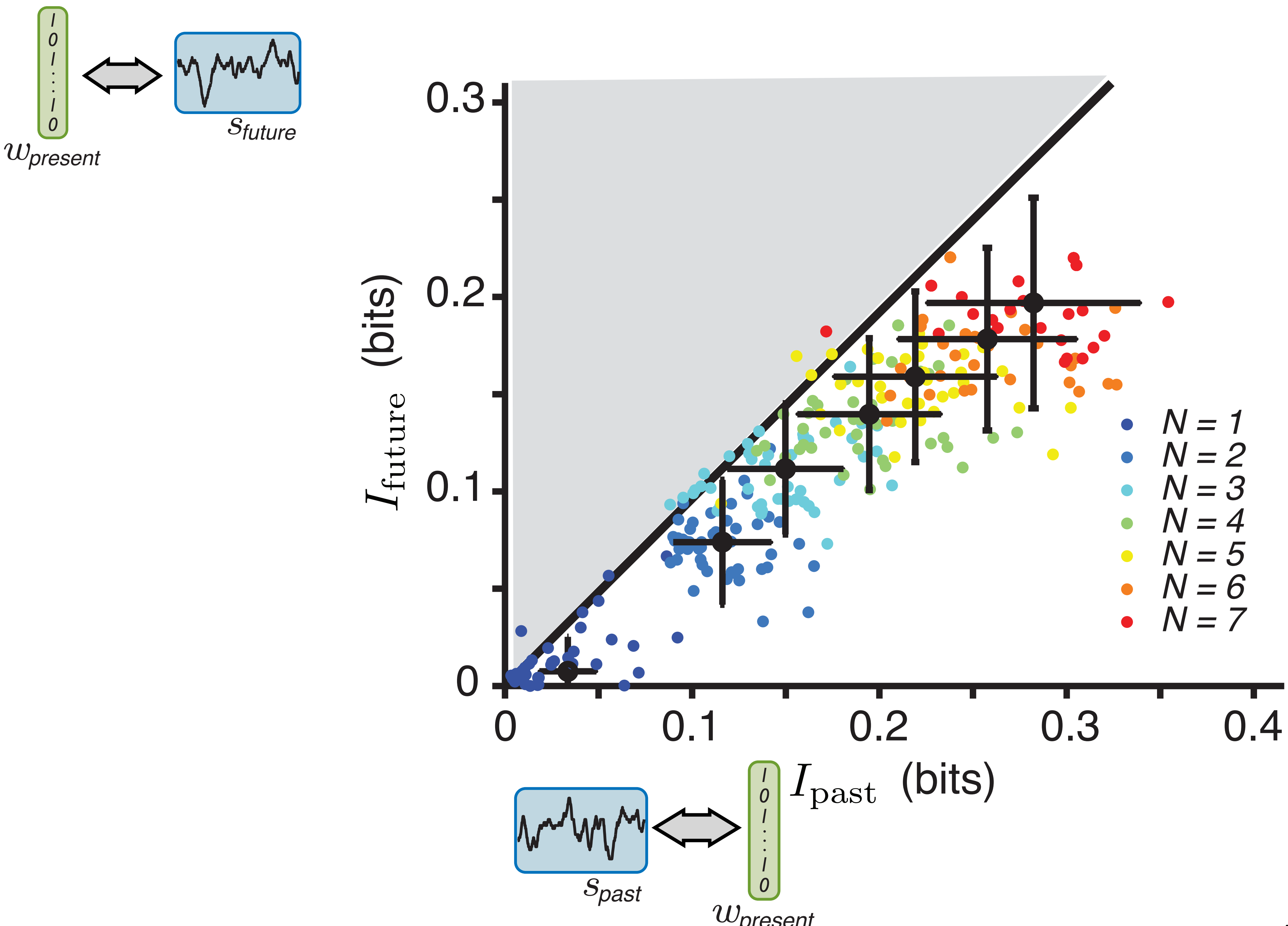


Tishby, Pereira, Bialek (1999)
Bialek, Nemenman, Tishby (2001)
Chechik, Globerson, Tishby, Weiss (2005)
Palmer et al. PNAS (2015)
Salisbury & Palmer J Stat Phys (2016)
Sachdeva, Mora, Walczak, Palmer PLoS CB (2021)
Palmer and Kline, New J. Phys (2022)

Spiking patterns sit close to the bound:



Retinal populations saturate the bound:

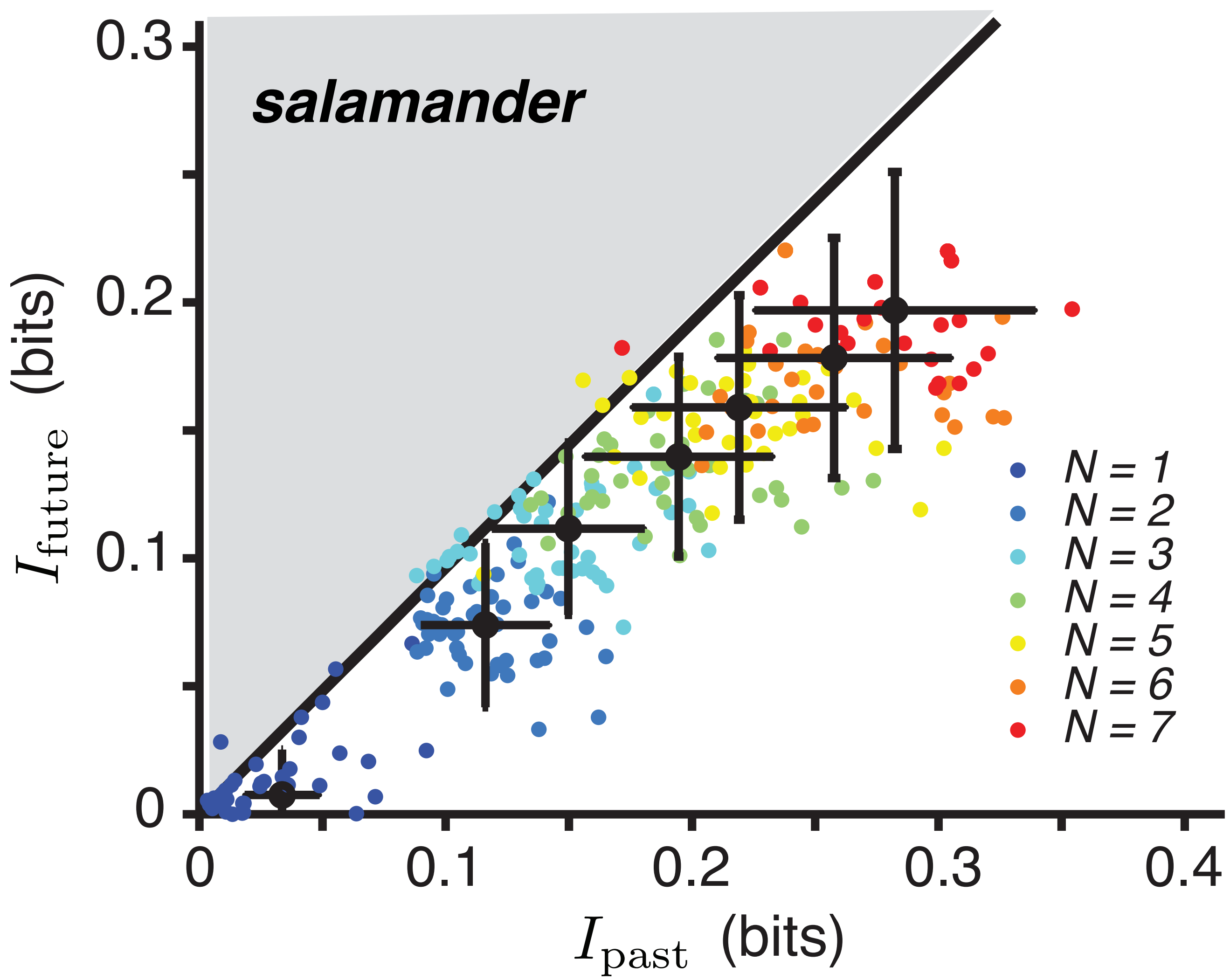


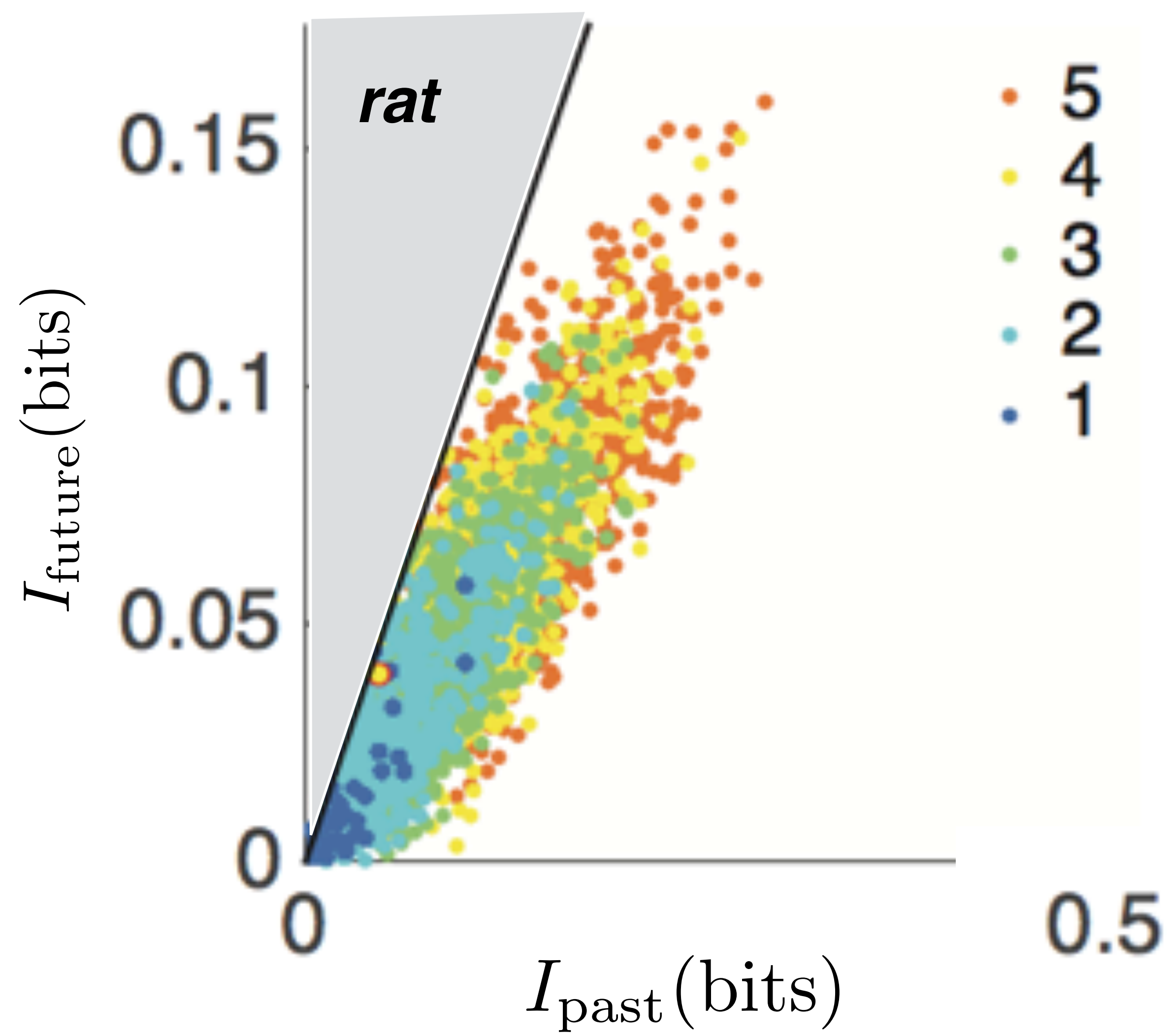
Palmer et al. PNAS (2015)

Salisbury and Palmer JSP (2016)

Sederberg, MacLean, Palmer PNAS (2018)

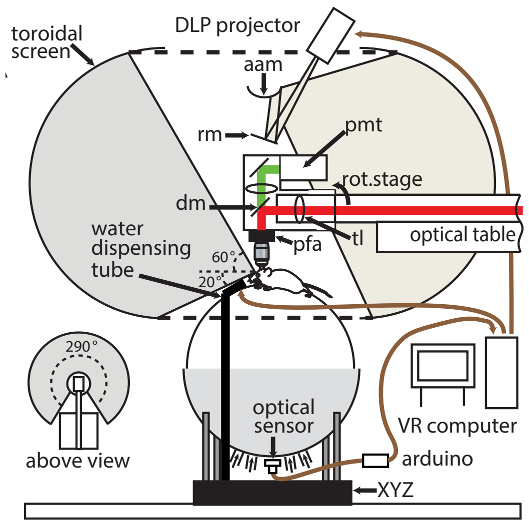
Retinal populations have as much predictive info as they possibly could:

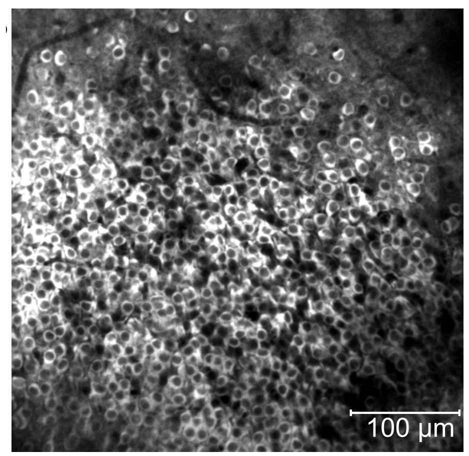


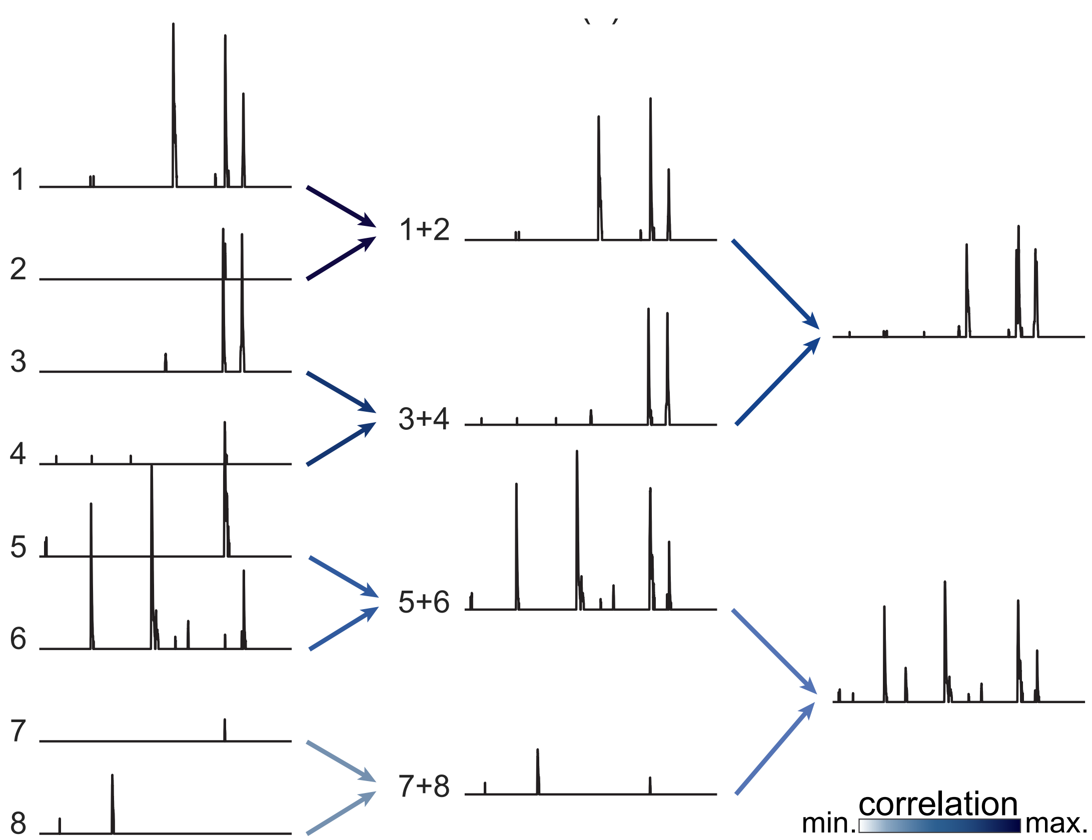


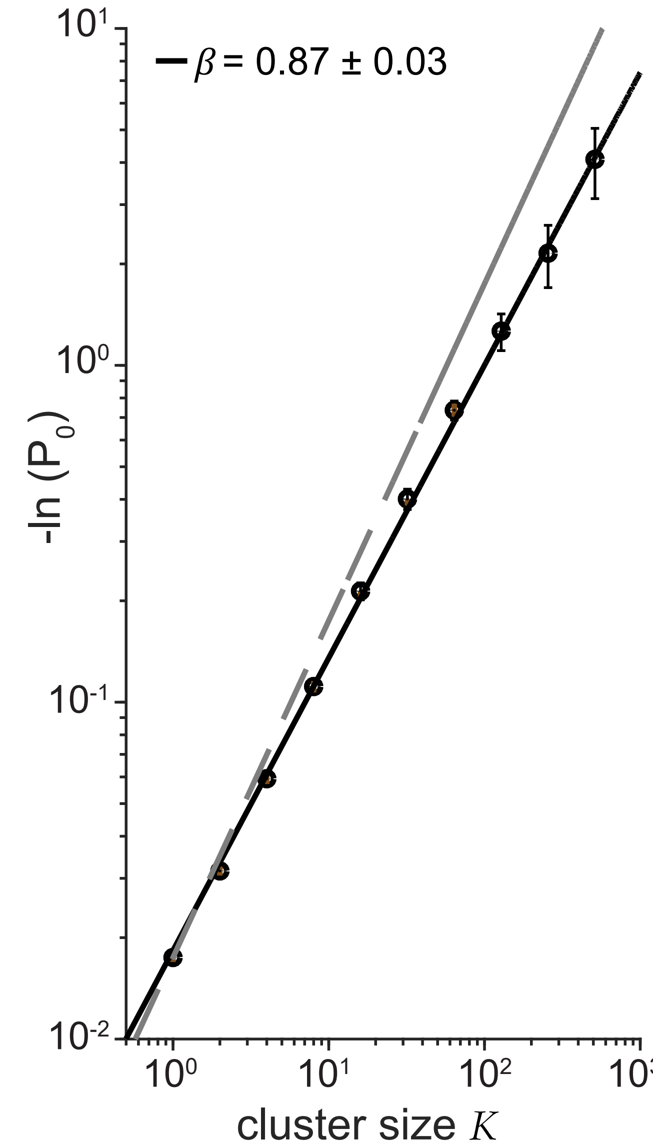
Palmer et al. PNAS (2015)
Salisbury and Palmer JSP (2016)
Sederberg, MacLean, Palmer PNAS (2018)
Salisbury, Marre, Palmer (in prep.)

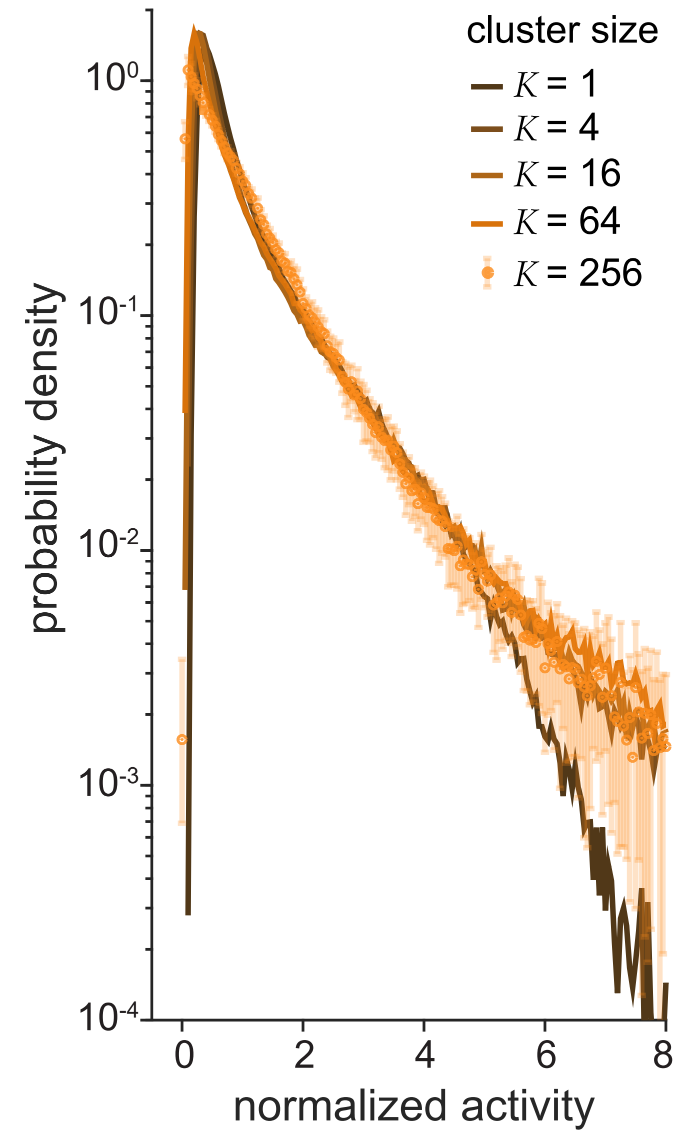
Coarse-graining in a neural population:











Meshulam, Gauthier, Brody, Tank, Bialek arXiv:1812.11904v1 PRL (2019)

Connecting IB to RG via NPRG:

$$P_\beta(\tilde{x}|x) = \mathrm{argmin}_{P(\tilde{x}|x)}\, I(X;\tilde{X}) - \beta I(\tilde{X};Y)$$

Gaussian coarsening

$$\tilde{X} = AX + \xi \quad \begin{cases} \Sigma_\xi = I \\ A(\beta) = \mathrm{diag}\left\{\alpha_i(\beta)\right\} V^T \\ \alpha_i(\beta) = \left[\dfrac{\beta(1-\lambda_i)-1}{\lambda_i s_i}\right]^{1/2} \Theta\left(\beta - \dfrac{1}{1-\lambda_i}\right) \end{cases}$$

$$s_i = [V^T \Sigma_X V]_{ii}$$

$$\Sigma_X^{-1}\Sigma_{X|Y} V = V\ \mathrm{diag}\left\{\lambda_i\right\}$$

$$W_k[J] = \log \int \mathcal{D}\chi\, \exp\left[-S[\chi] - \Delta S_k[\chi] + \sum_a \int \mathrm{d}^d x\, J_a(x)\chi_a(x)\right]$$

$$\Delta S_k[\chi] = \frac{1}{2}\chi^\dagger R_k \chi$$

Gaussian coarsening

Tishby, Pereira, Bialek (1999)
Chechik, Globerson, Tishby, Weiss (2005)
Kline + SEP, *New J. Phys.* 24.3 (2022)

A soft cutoff, and computing the regulator:

$$\Delta S_k[\chi] = \frac{1}{2}\chi^\dagger A_k^\dagger \Delta_k^{-1} A_k \chi$$

$$P_k[\tilde{\chi}|\chi] = \exp\left[-\frac{1}{2}(\tilde{\chi} - A_k\chi)^\dagger \Delta_k^{-1}(\tilde{\chi} - A_k\chi) - C_k\right]$$

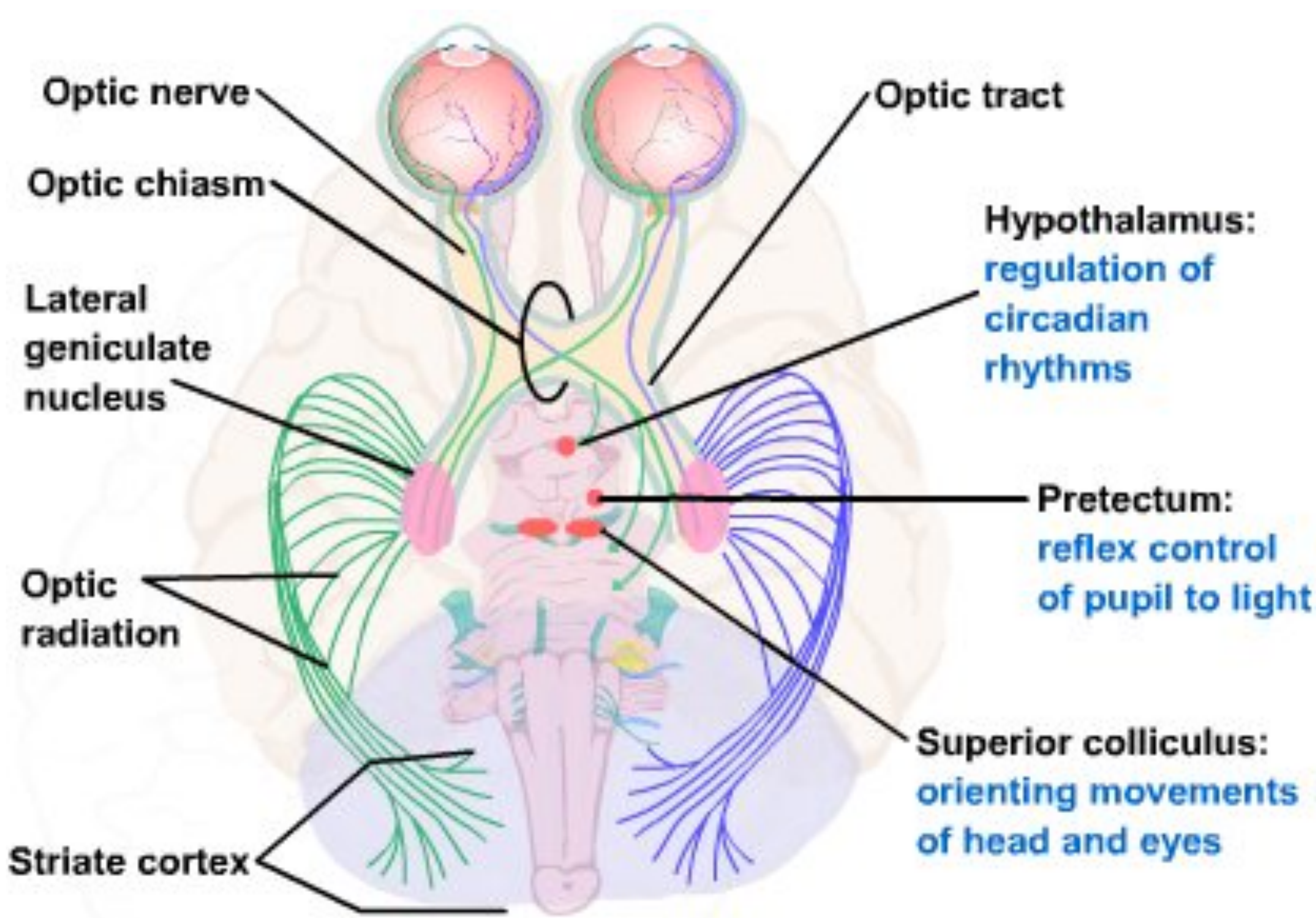
$$\min_{P(\tilde{X}|X)} I(X;\tilde{X}) - \beta I(\tilde{X};Y)$$

$$\Delta S_\beta(x) = \frac{1}{2}x^\dagger R_\beta x$$

$$\left[R_\beta^{(\mathrm{IB})}\right]_{ij} = \frac{\beta - \beta_i}{s_i(\beta_i - 1)}\Theta(\beta - \beta_i)\delta_{ij}$$

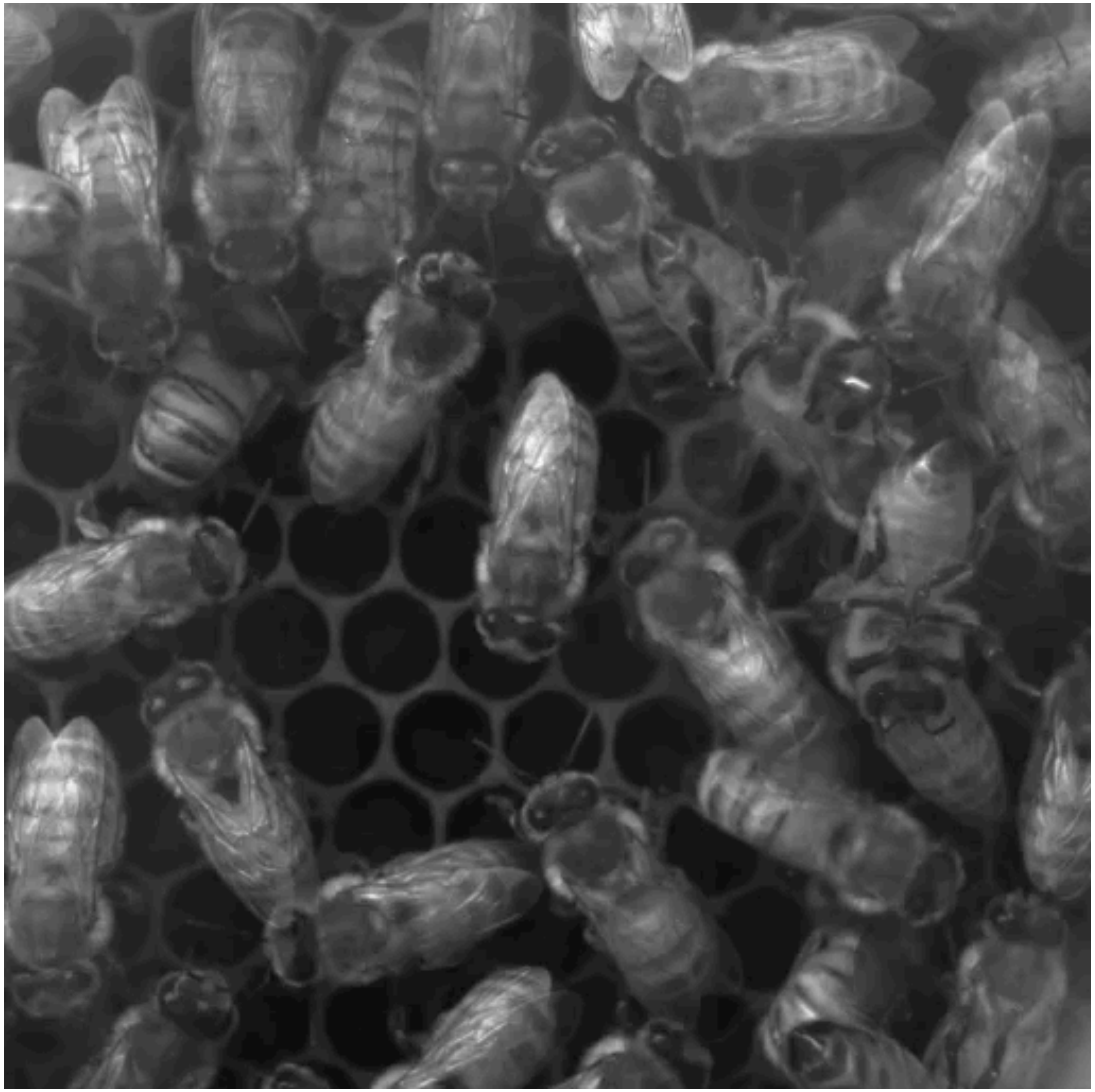
coarsening	IB	NPRG
Gaussian	only when P(X,Y) Gaussian	almost always
non-Gaussian	almost always the case	probably intractable

Retinal neurons contribute to many behaviors:



Neuroscience *online*, Ch. 15 Valentin Dragoi (2020)

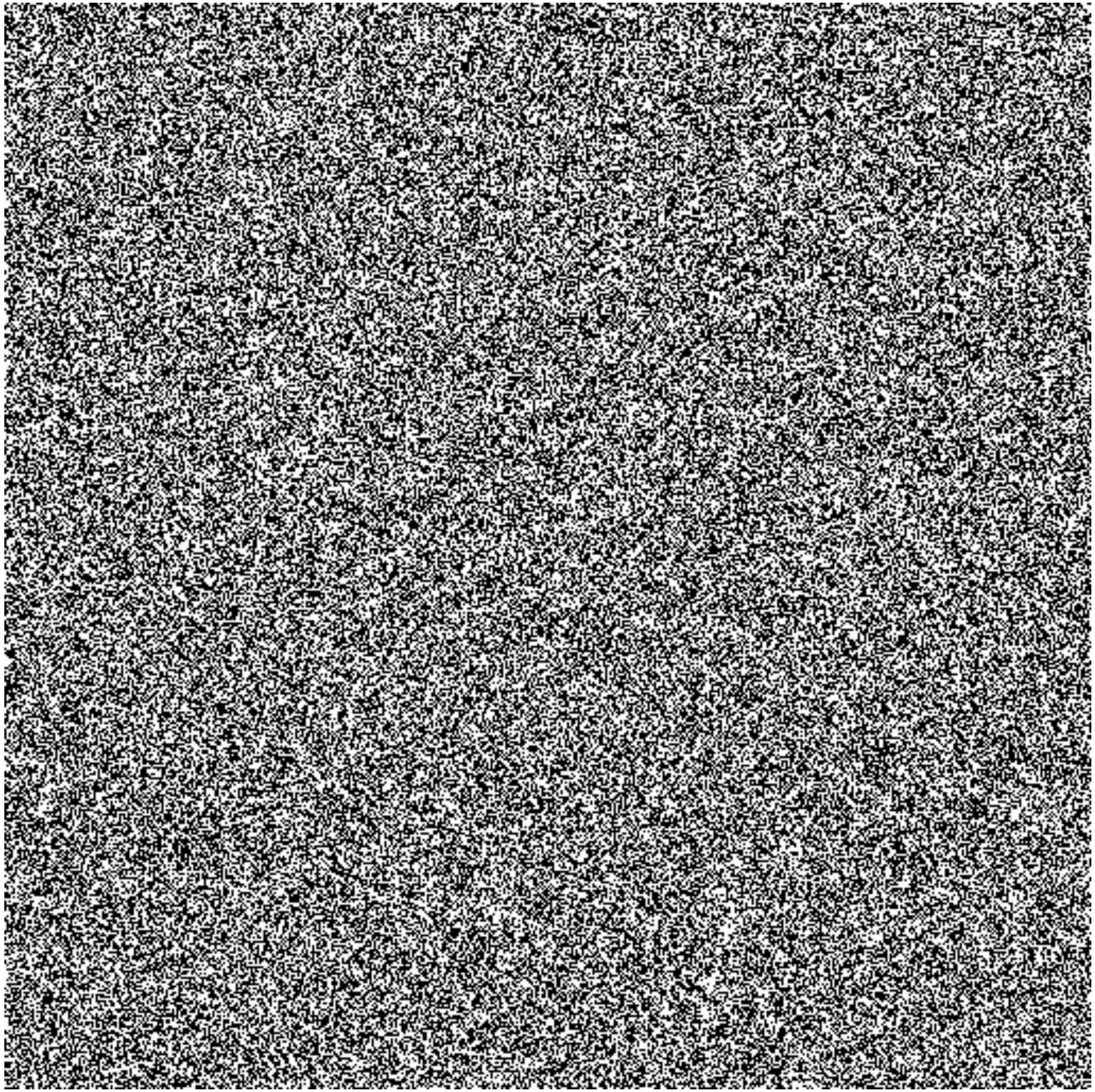
Natural scenes are complex, full of “relevance” features:



Analyzing local flow:

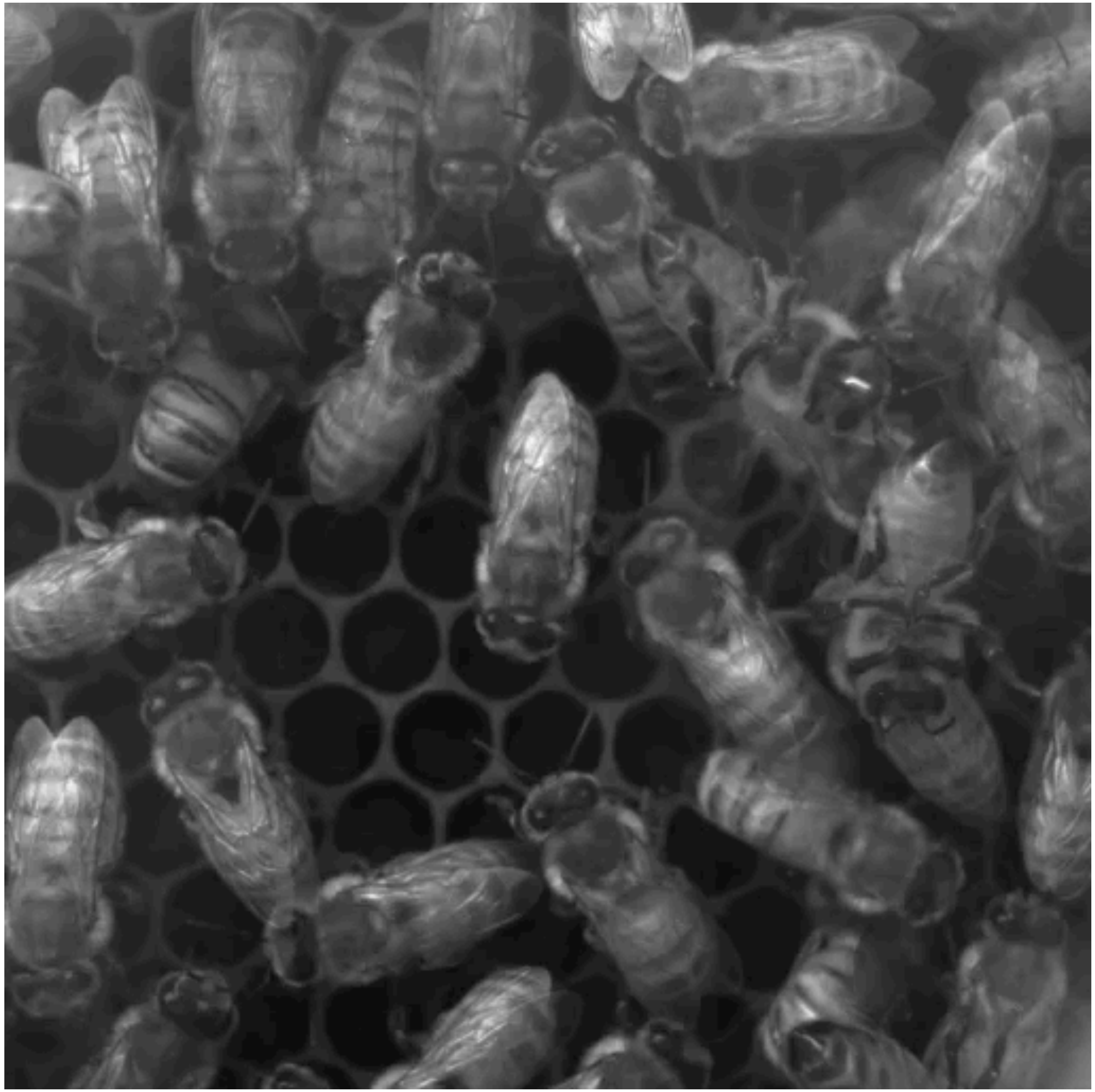
Jared Salisbury

Removing static spatial correlations:

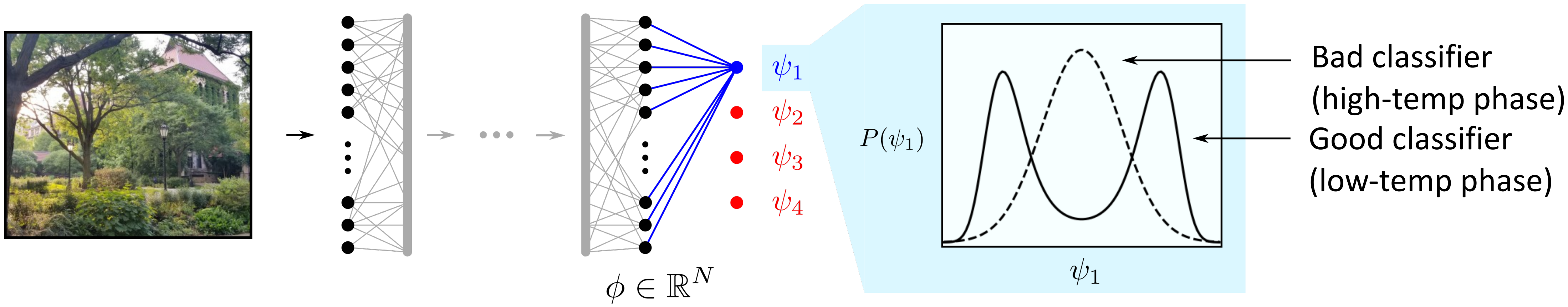


Jared Salisbury

Multiple notions of scale in natural scenes:



Multiple notions of scale in one system:



$$\mathcal{H}(\phi) = \frac{1}{2} u^{(2)}_{ab}\, \phi_a \phi_b + \frac{1}{4!} u^{(4)}_{abcd} \phi_a \phi_b \phi_c \phi_d$$

$$R_{k\,\mu,-\mu} = \min\left\{ k\left(\Gamma^{(4)}_{k\,\psi,\psi,\mu,-\mu} - 2\,\Gamma^{(3)}_{k\,\psi,\mu} \cdot G_k \cdot \Gamma^{(3)}_{k\,\psi,-\mu}\right) - \Gamma^{(2)}_{l\,\mu,-\mu},\, 0 \right\}$$

One can change the scaling properties and collective mode ordering by changing how third and fourth-order vertices relate to the second-order vertex.

Kline + SEP, *New J. Phys.* 24.3 (2022)
C Wetterich, *Phys. Lett. B* 301.1 (1993)
DF Litim, *Phys. Rev. D* 64.10 (2001)

Summary:

- *brains "see" different things*
- *working on an RG framework*
- *compute your regulator!*

questions?

selected Palmer group paper refs:

SEP, Marre, Berry, Bialek PNAS (2015)
Salisbury + SEP J Stat Phys (2016)
Sederberg, MacLean, SEP PNAS (2018)
Buerkle + SEP bioRxiv 891382 (2019)
Ding + Chen ++ SEP + Wei bioRxiv eLife (2021)
Sachdeva, Mora, Walczak, SEP bioRxiv PLoS CB (2021)
Wang, Borst, Segev, SEP bioRxiv 814319 PLoS CB (2021)
Kline + SEP arXiv 2107.13700 NJP (2022)