Coding and computation by neural ensembles in the primate retina

Liam Paninski

Department of Statistics and Center for Theoretical Neuroscience
Columbia University
http://www.stat.columbia.edu/∼liam
liam@stat.columbia.edu
February 14, 2010

The coming statistical neuroscience decade

Some notable recent developments:

- machine learning / statistics methods for extracting information from high-dimensional data in a computationally-tractable, systematic fashion

- computing (Moore's law, massive parallel computing)

- optical methods (eg two-photon, FLIM) and optogenetics (channelrhodopsin, viral tracers, "brainbow")

- high-density multielectrode recordings (Litke's 512-electrode retinal readout system; Shepard's 65,536-electrode active array)

Some exciting open challenges

- inferring biophysical neuronal properties from noisy recordings

- reconstructing the full dendritic spatiotemporal voltage from noisy, subsampled observations

- estimating subthreshold voltage given superthreshold spike trains

- extracting spike timing from slow, noisy calcium imaging data

- reconstructing presynaptic conductance from postsynaptic voltage recordings

- inferring connectivity from large populations of spike trains

- decoding behaviorally-relevant information from spike trains

- optimal control of neural spike timing

— to solve these, we need to combine the two classical branches of computational neuroscience: dynamical systems and neural coding
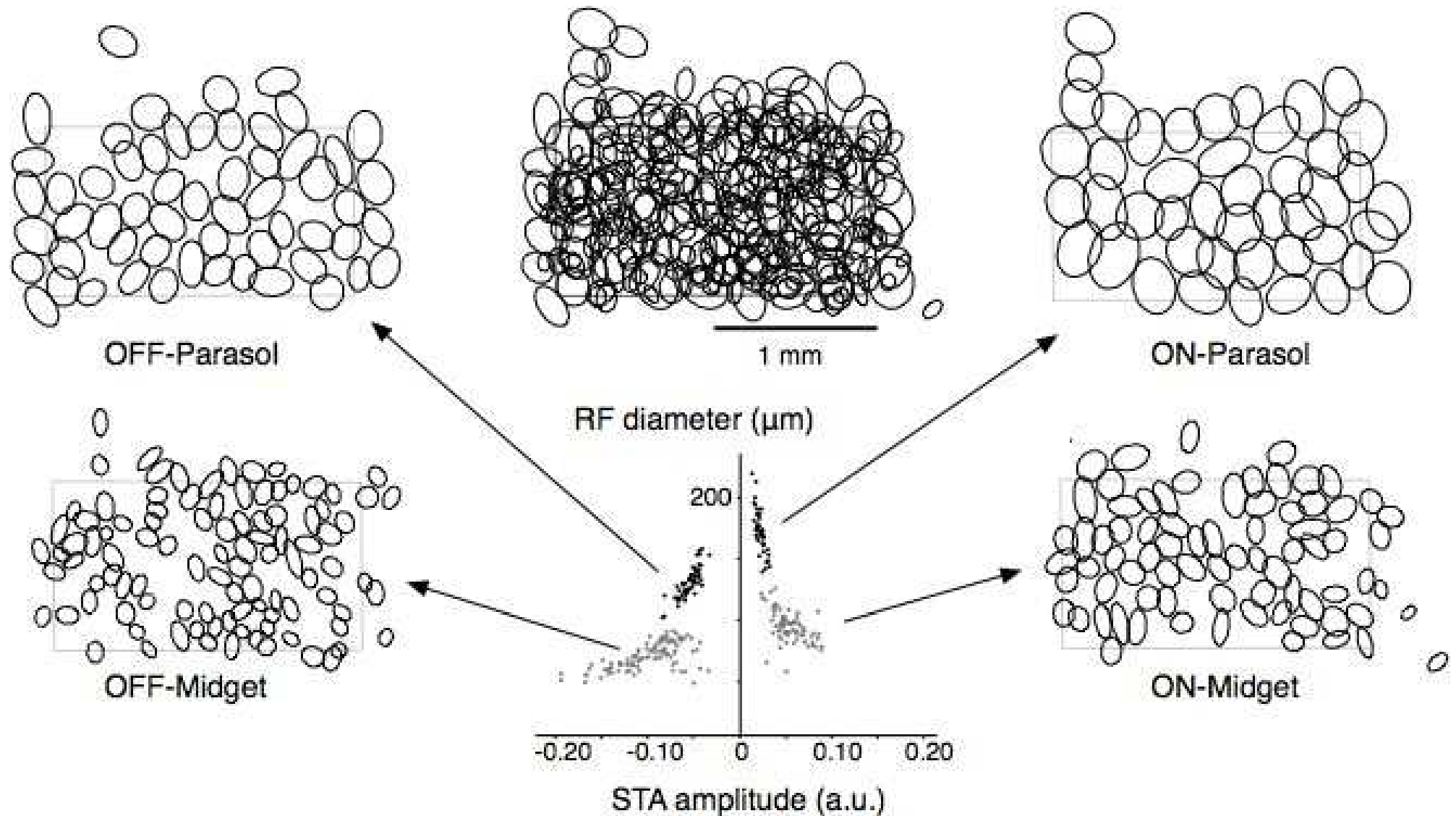
Retinal ganglion neuronal data

Preparation: dissociated macaque retina

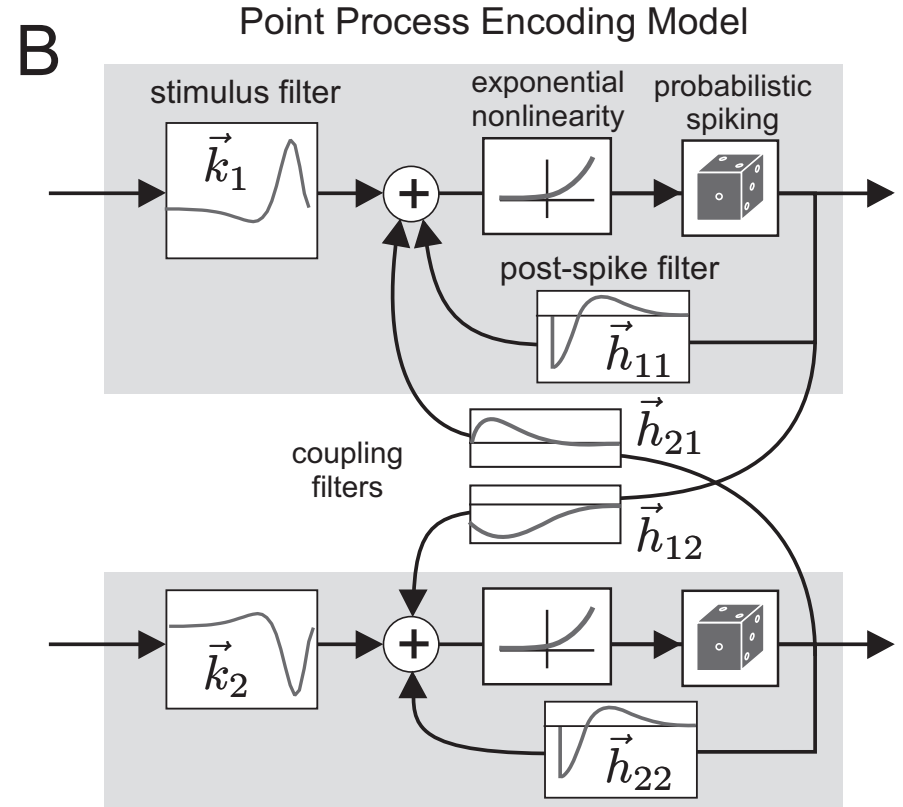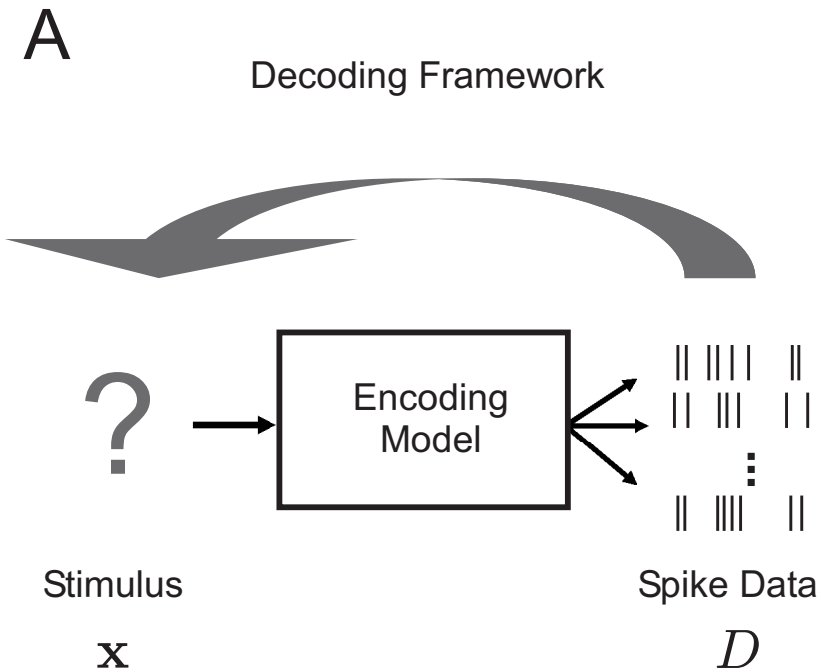— extracellularly-recorded responses of populations of RGCs
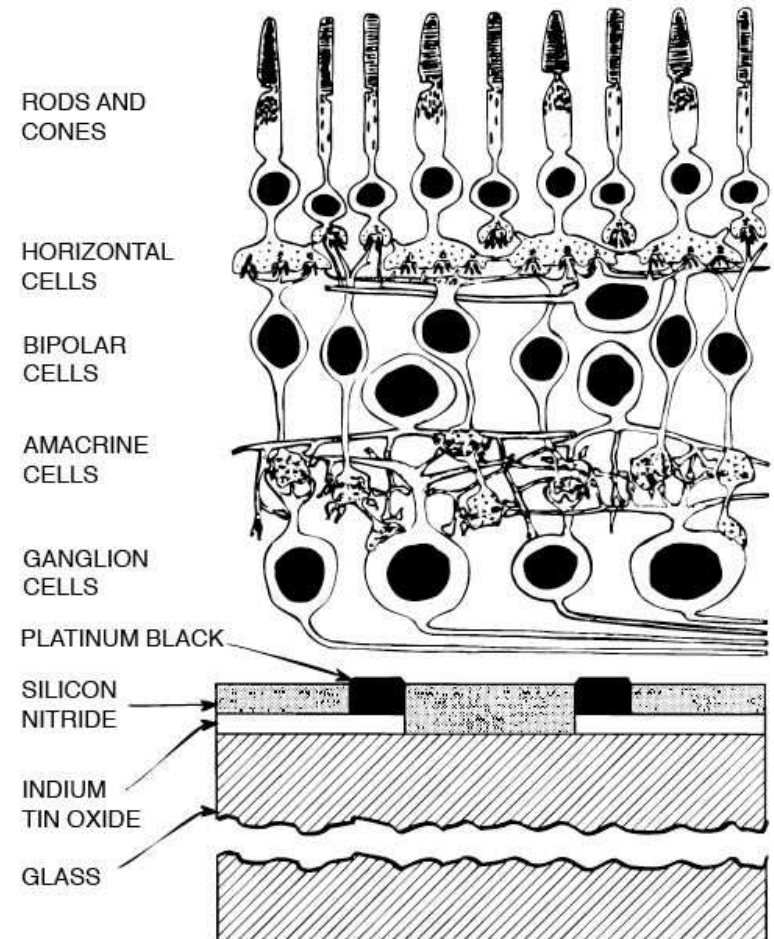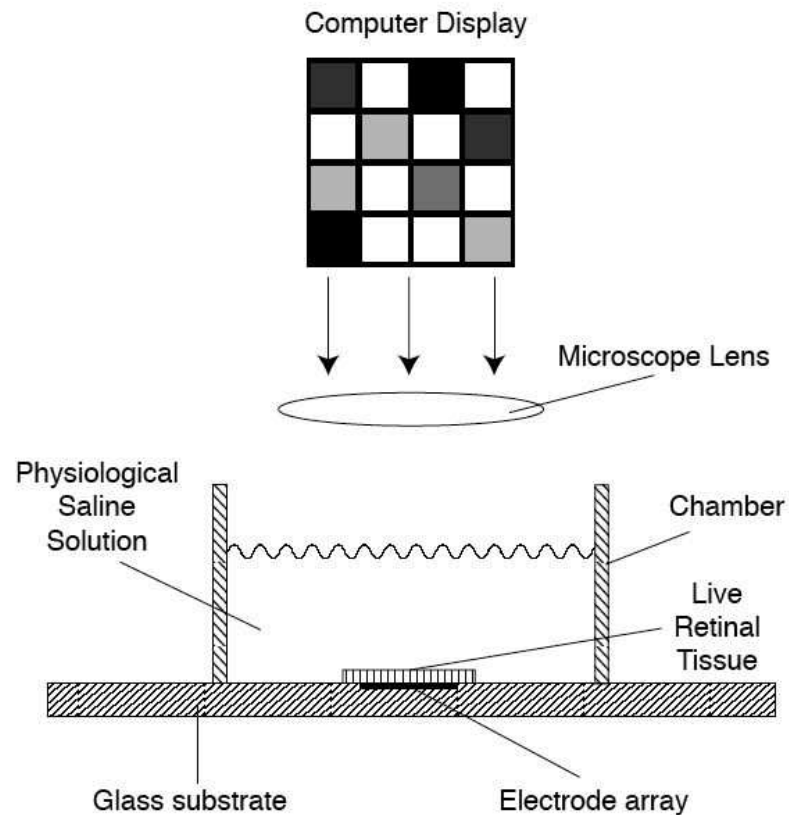
Stimulus: random spatiotemporal visual stimuli (Pillow et al., 2008)

# Receptive fields tile visual space

OFF-Parasol

ON-Parasol

1 mm

RF diameter (µm)

OFF-Midget

ON-Midget

200

-0.20   -0.10   0   0.10   0.20

STA amplitude (a.u.)

Multineuronal point-process model

A

Decoding Framework

? → Encoding Model →

Stimulus
x

Spike Data
D

B

Point Process Encoding Model

stimulus filter $\vec{k}_1$ — exponential nonlinearity — probabilistic spiking

post-spike filter $\vec{h}_{11}$

$\vec{h}_{21}$

coupling filters

$\vec{h}_{12}$

$\vec{k}_2$

$\vec{h}_{22}$

$$\lambda_i(t) = f\left(b_i + \vec{k}_i \cdot \vec{x}(t) + \sum_{i',j} h_{i',j}\, n_{i'}(t-j) \right),$$

— likelihood is easy to compute and to maximize (concave optimization)
(Paninski, 2004; Truccolo et al., 2005; Pillow et al., 2008)

— close connections to noisy integrate-and-fire model
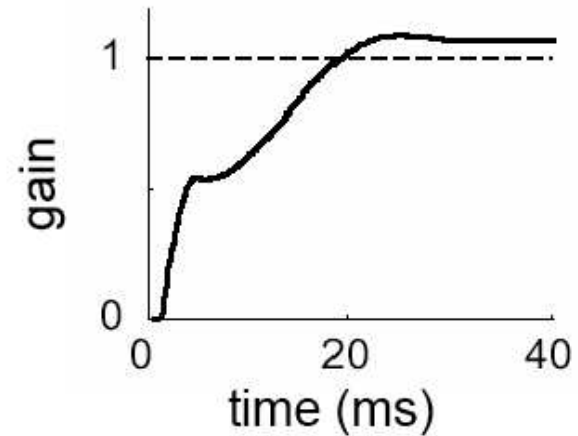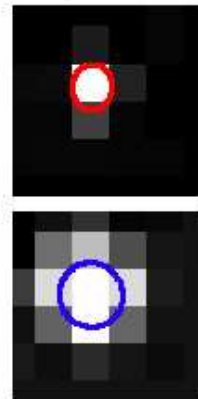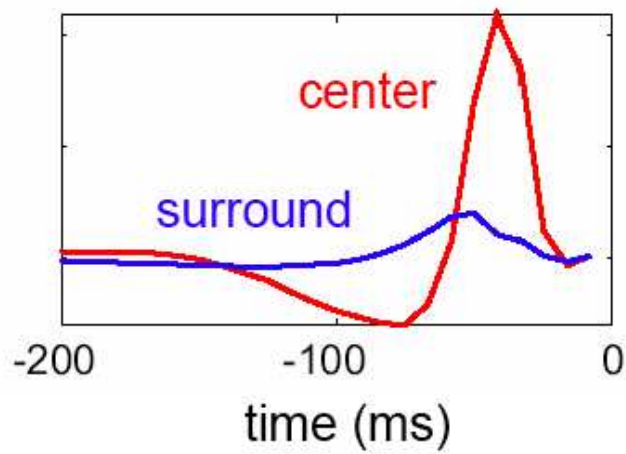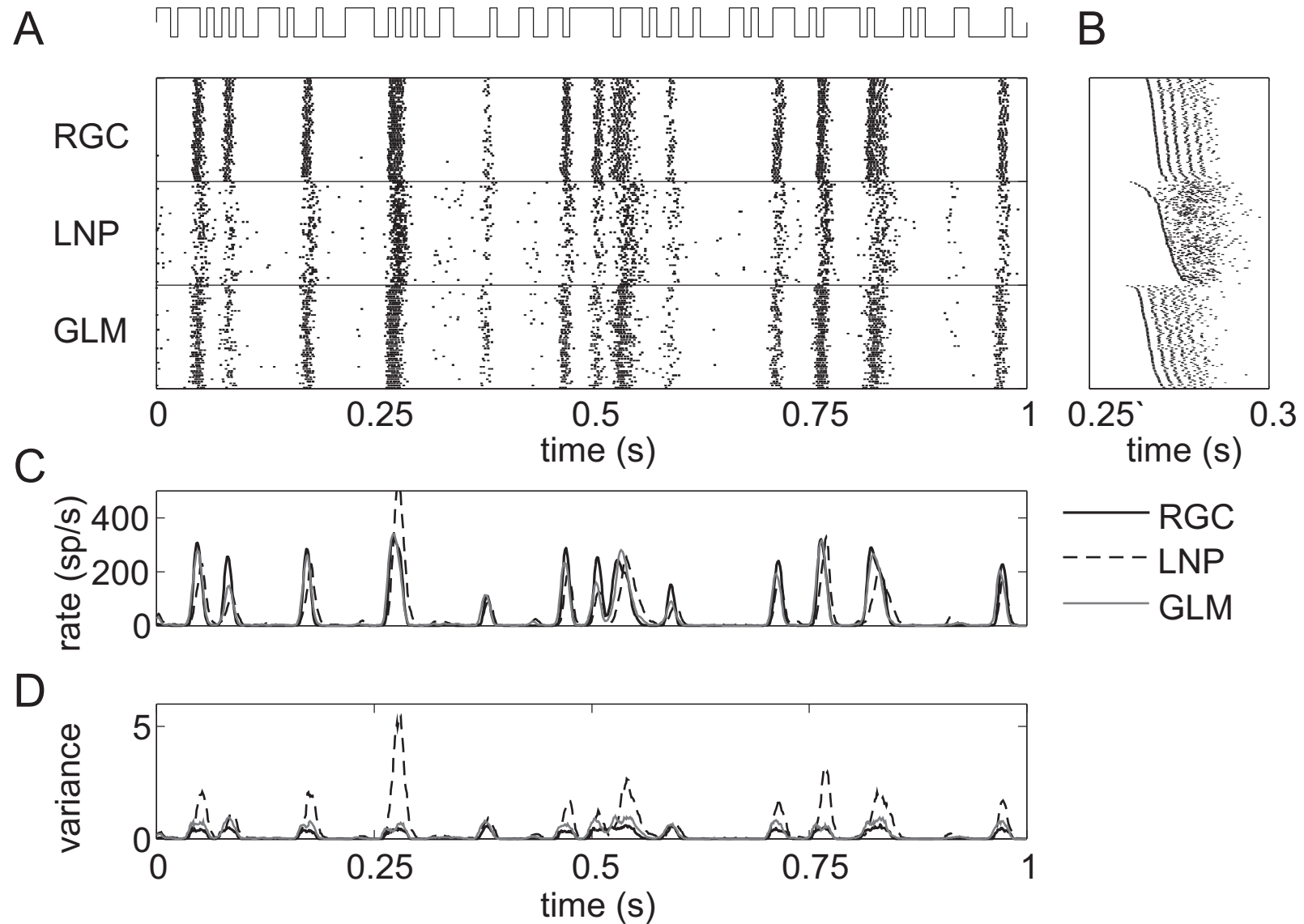
ON cells
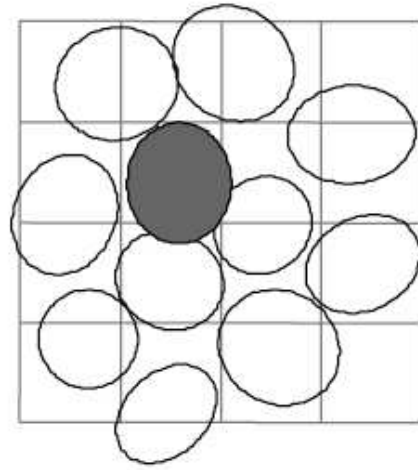
OFF cells

stimulus filter

temporal

center

surround

spatial

-200   -100   0

time (ms)

post-spike filter

gain

1

0

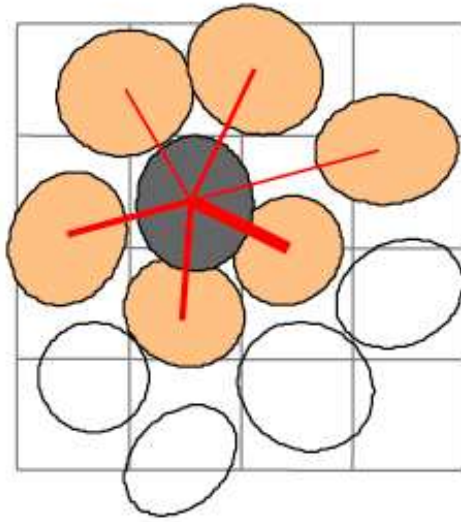0   20   40

time (ms)
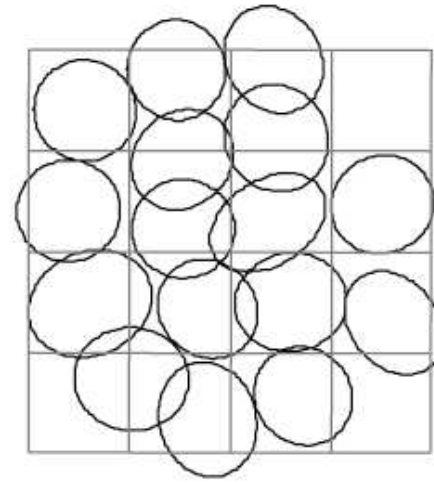
Predicting single-neuron responses

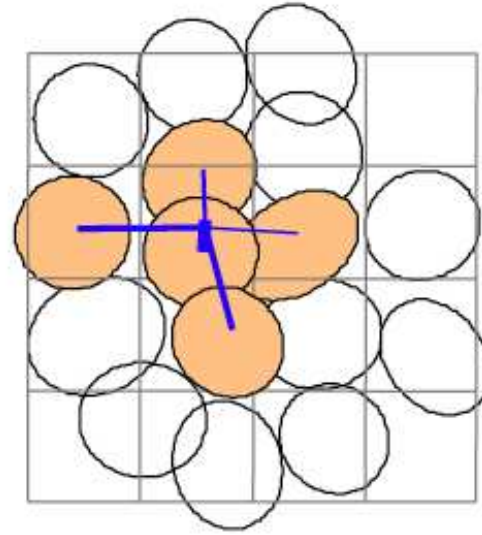
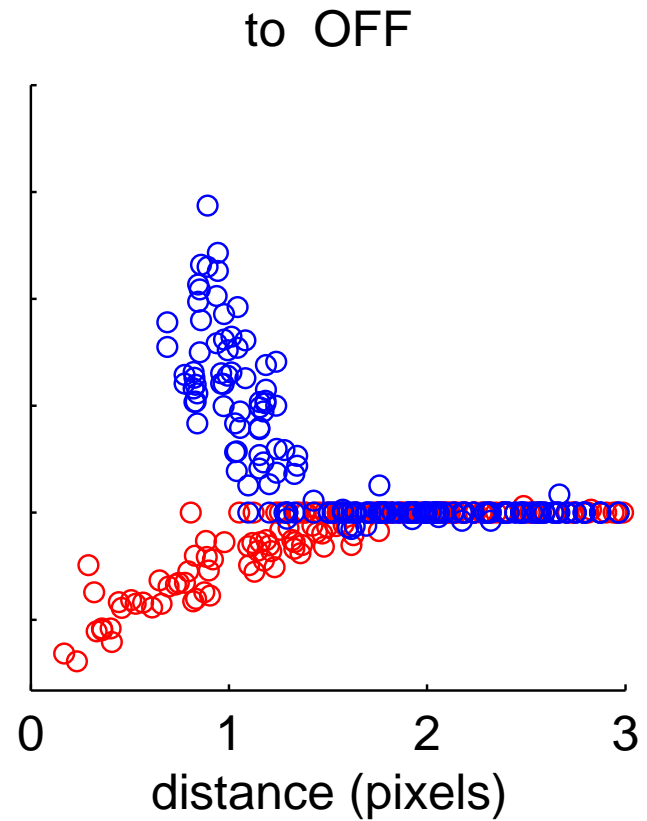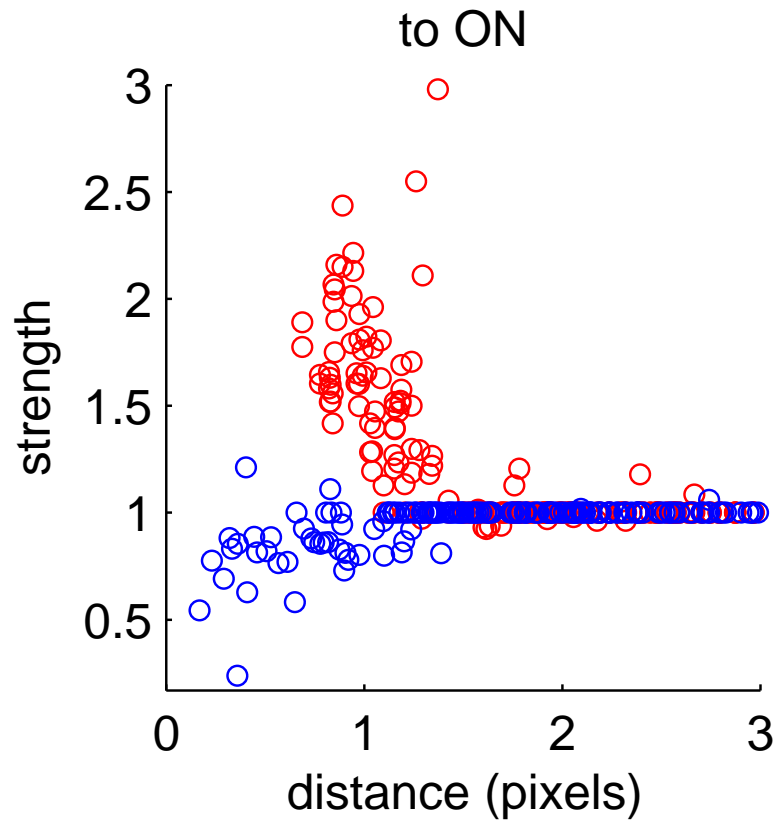— model captures high precision of retinal responses.

ON
cells

OFF
cells

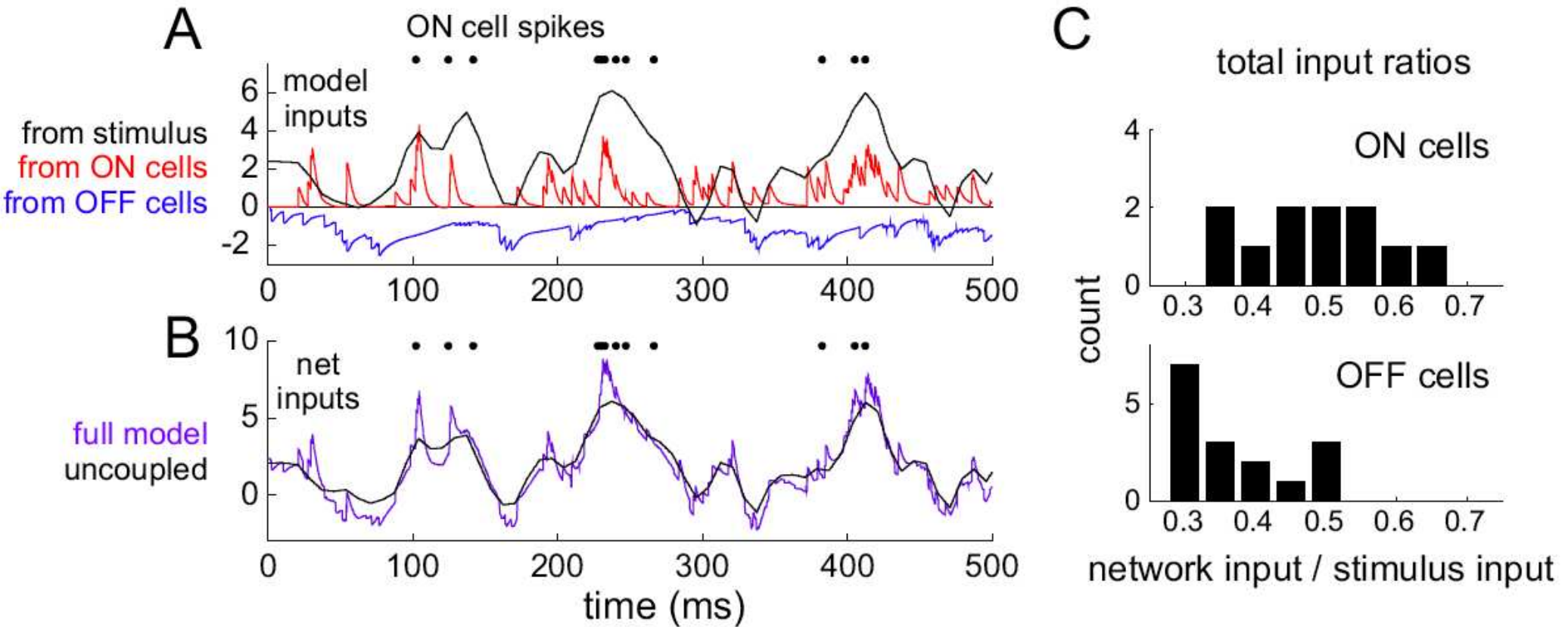coupling filters

from ON

gain

time (ms)

from OFF

time (ms)
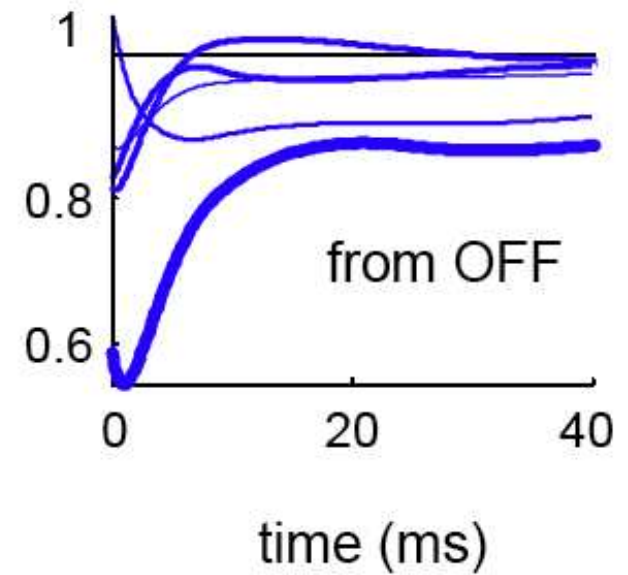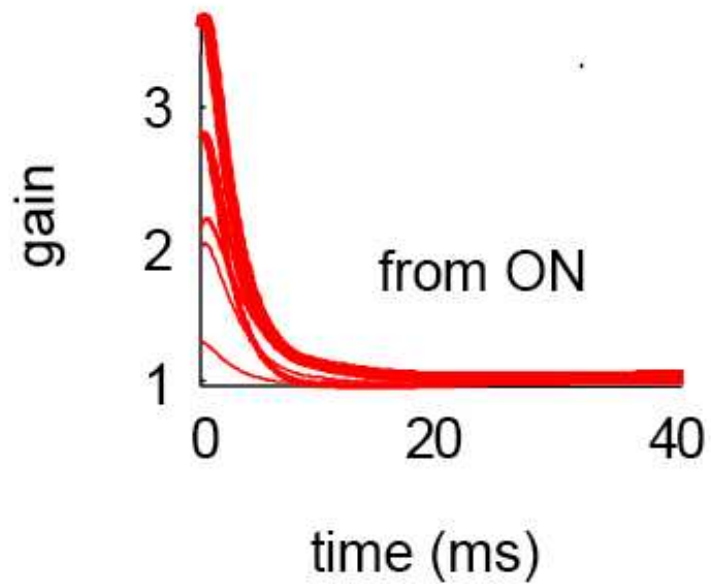
Nearest-neighbor effective connectivity

to ON

to OFF

Network vs. stimulus drive
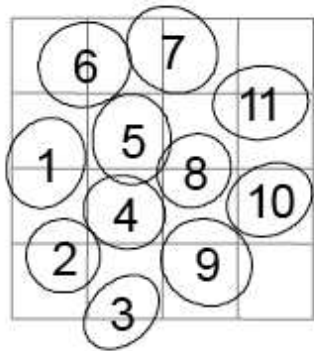

— Network effects are ≈ 50% as strong as stimulus effects

Model captures spatiotemporal cross-corrs

x-corrs:

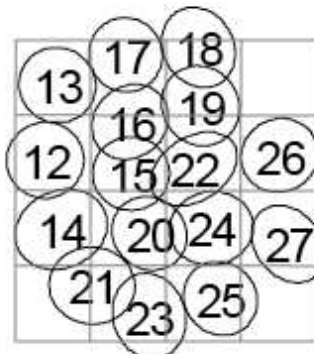ON cells

OFF cells

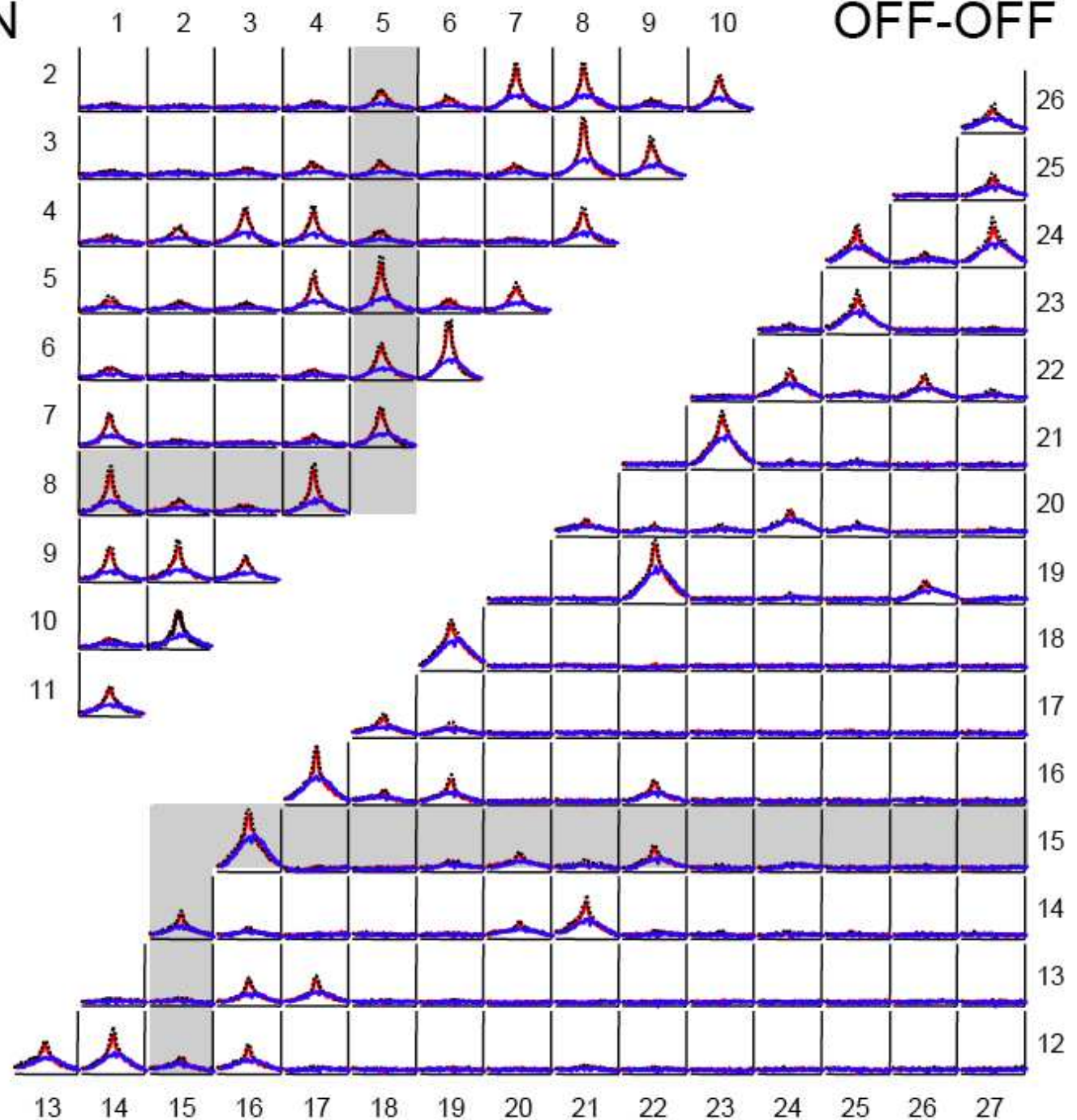75 sp/s

50 ms

ON-ON   OFF-OFF

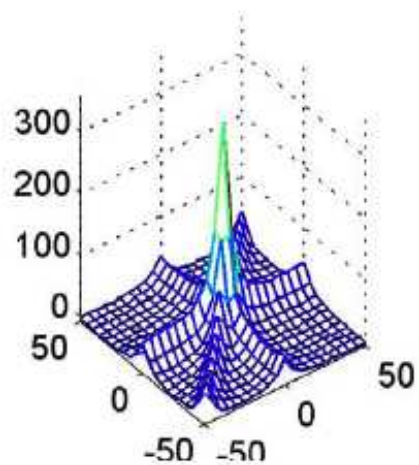# Triplet correlations

3 ON cells
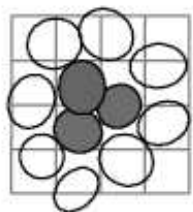
3 OFF cells

RGC   full model   uncoupled
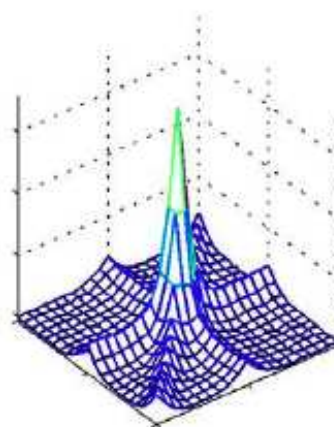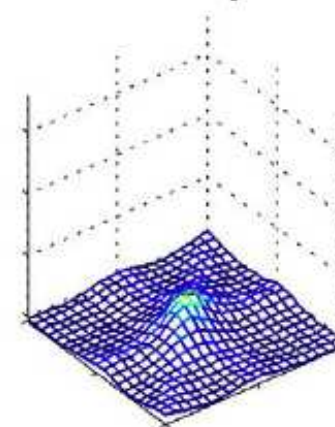
rate (sp/s)
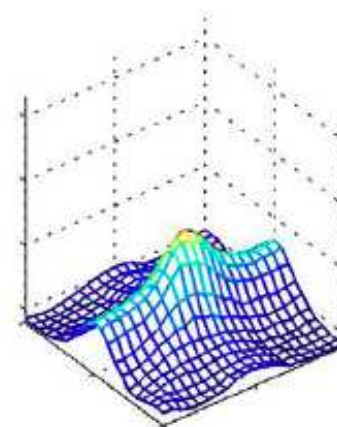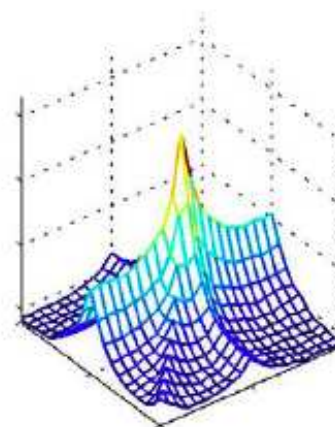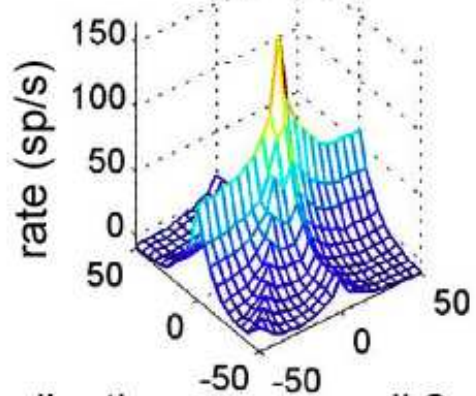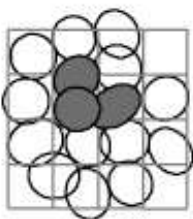
cell 1 spike time   cell 2 spike time

Optimal Bayesian decoding

$$E(\vec{x}|spikes) \approx \arg\max_{\vec{x}} \log P(\vec{x}|spikes) = \arg\max_{\vec{x}} \left[\log P(spikes|\vec{x}) + \log P(\vec{x})\right]$$

(Loading yashar-decode.mp4)

— Computational points:

- $\log P(spikes|\vec{x})$ is concave in $\vec{x}$: concave optimization again.

- Decoding can be done in linear time via standard Newton-Raphson methods, since Hessian of $\log P(\vec{x}|spikes)$ w.r.t. $\vec{x}$ is banded (Pillow et al., 2009).

Optimal Bayesian decoding
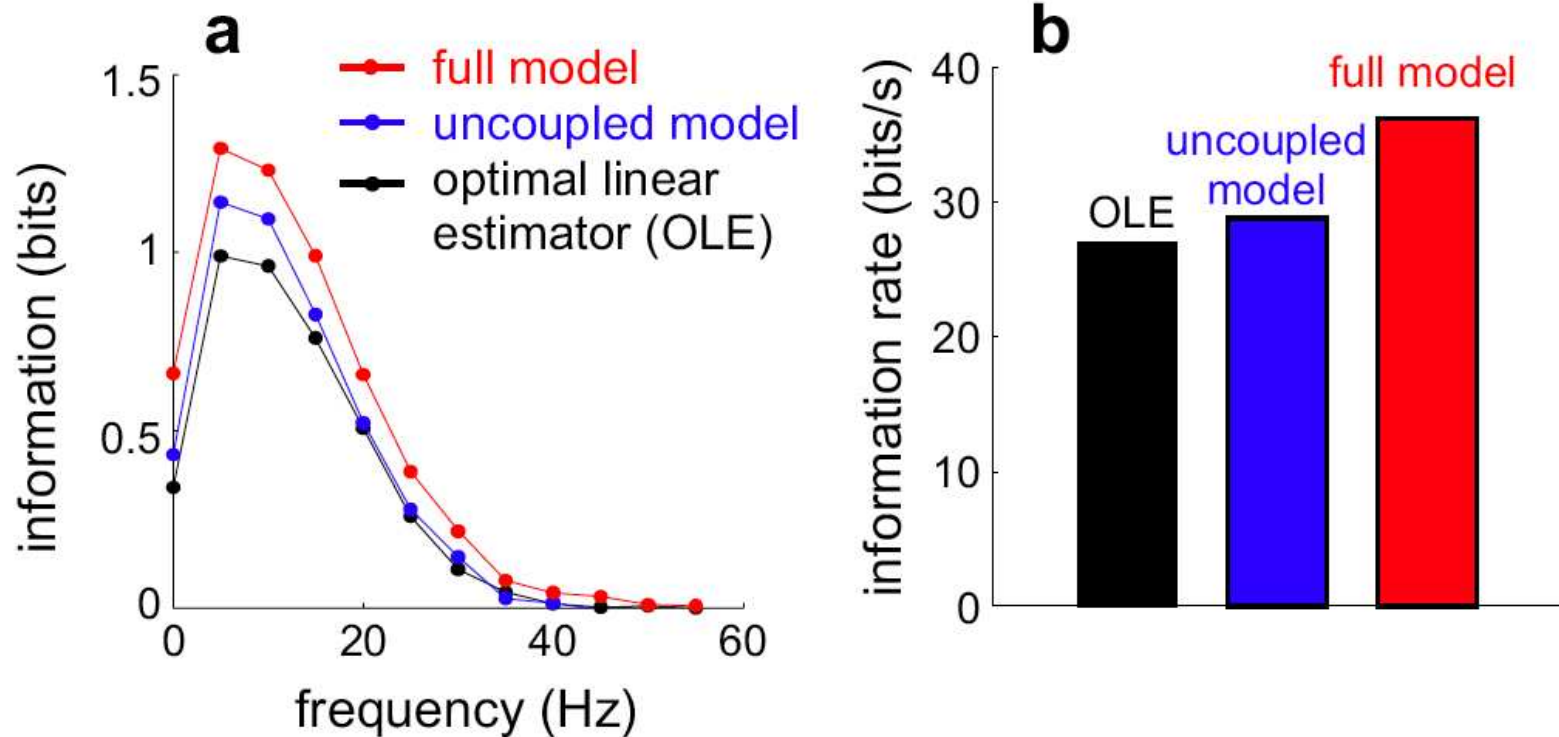
$$E(\vec{x}|spikes) \approx \arg\max_{\vec{x}} \log P(\vec{x}|spikes) = \arg\max_{\vec{x}} \left[\log P(spikes|\vec{x}) + \log P(\vec{x})\right]$$

— Computational points:

- $\log P(spikes|\vec{x})$ is concave in $\vec{x}$: concave optimization again.

- Decoding can be done in linear time via standard Newton-Raphson methods, since Hessian of $\log P(\vec{x}|spikes)$ w.r.t. $\vec{x}$ is banded (Pillow et al., 2009).

— Biological point: paying attention to correlations improves decoding accuracy.

Application: how important is timing?
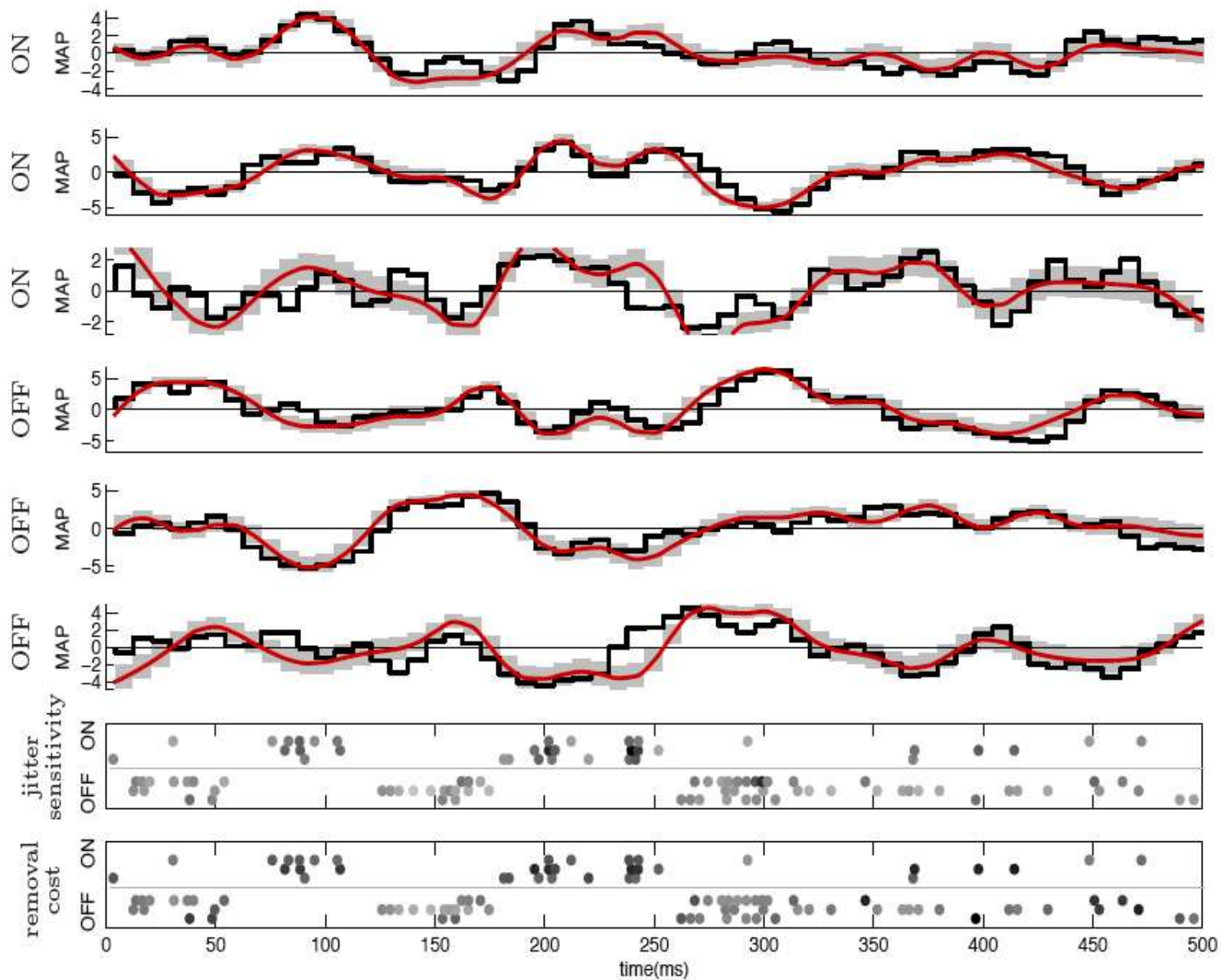
— Fast decoding methods let us look more closely (Ahmadian et al., 2010)

Application: optimal velocity decoding

A

B

raw responses

cell: 1 to 56
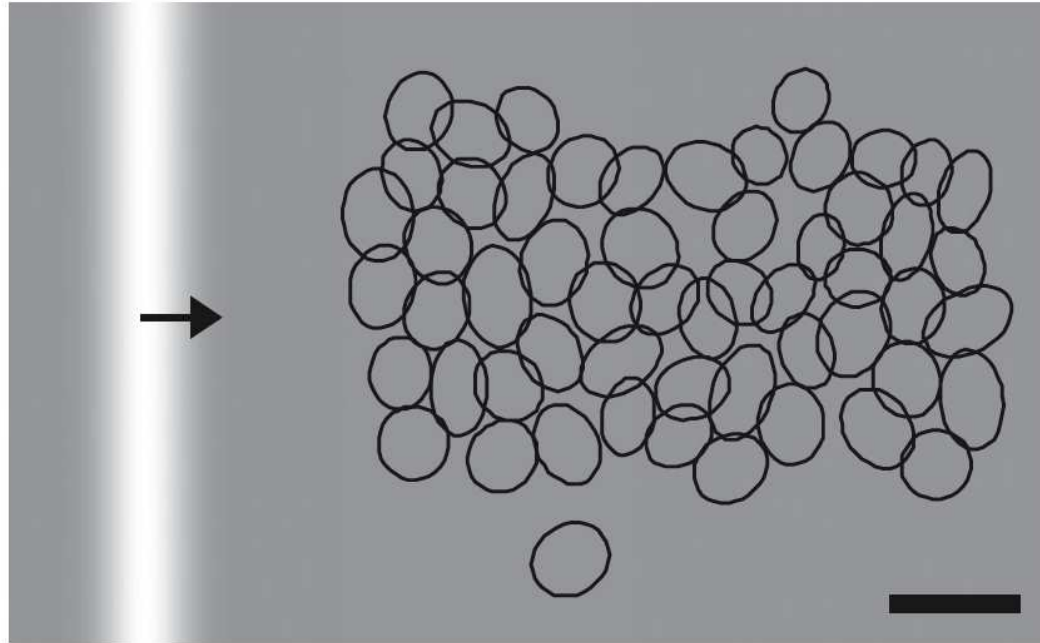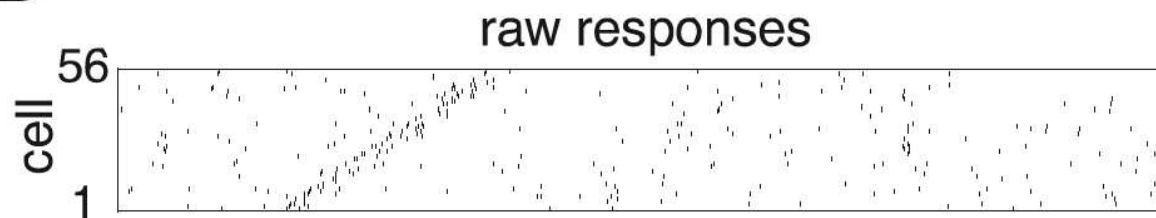
Bayesian estimate requires us to integrate out unknown image I:

$$p(v|spikes) \propto p(v)p(spikes|v) = p(v) \int p(I)p(spikes|v, I)dI;$$

(Frechette et al., 2005; Lalor et al., 2009)

Application: image stabilization

(Loading rossi-roorda.avi)

— from (Rossi and Roorda '09): quite a bit of motion in 1 sec.

Bayesian methods for image stabilization

Have to integrate out random eye movements:

$$p(I|spikes) \propto p(I)p(spikes|I) = p(I) \int p(spikes|e, I)p(e)de;$$

e denotes eye path; integration by particle-filter methods.
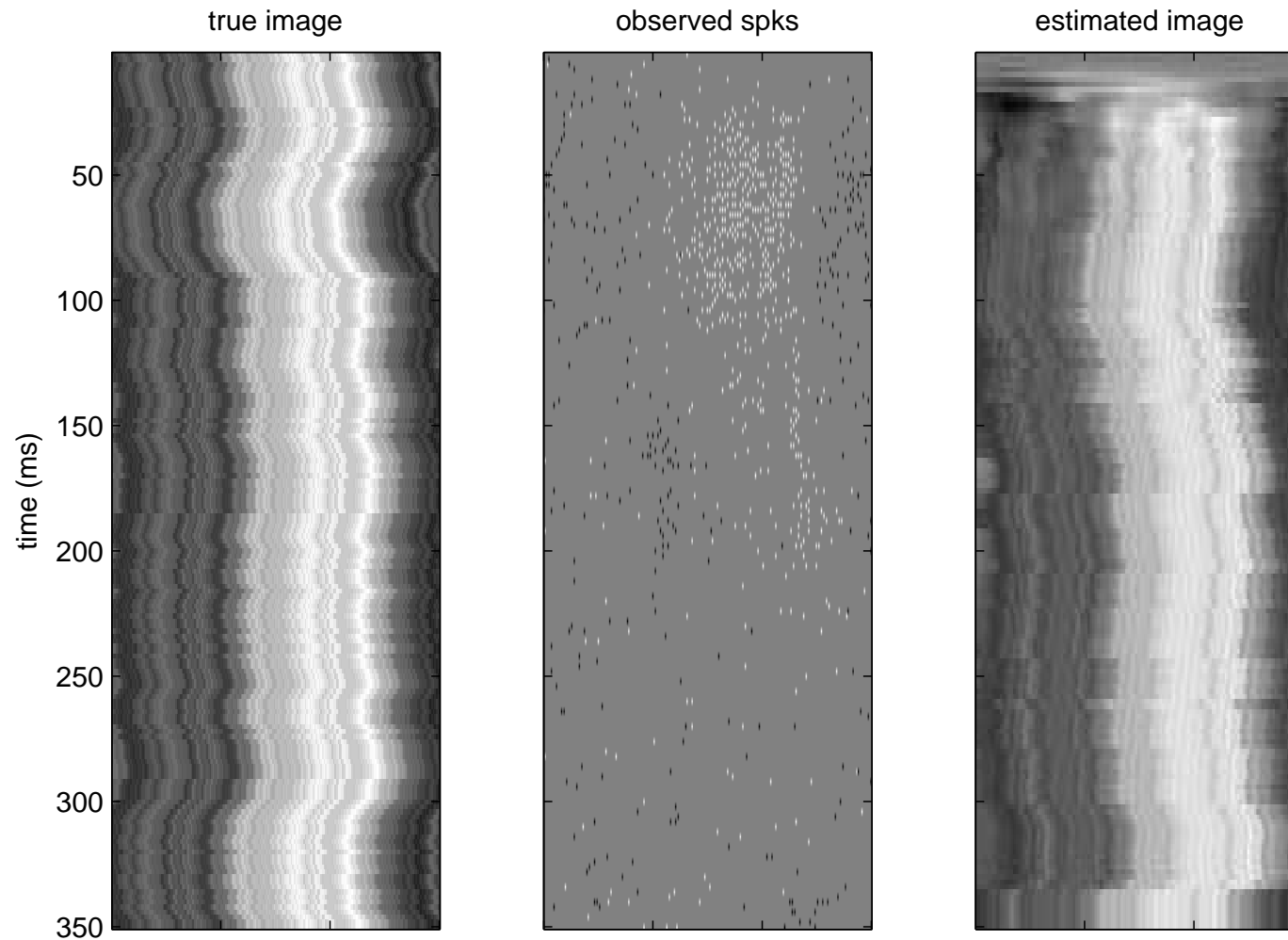
Next steps: reconsidering the model

$$\lambda_i(t) = \exp\left(k_i \cdot x(t) + h_i \cdot y_i(t) + \sum_{i \neq j} l_{i,j} \cdot y_j(t)\right)$$
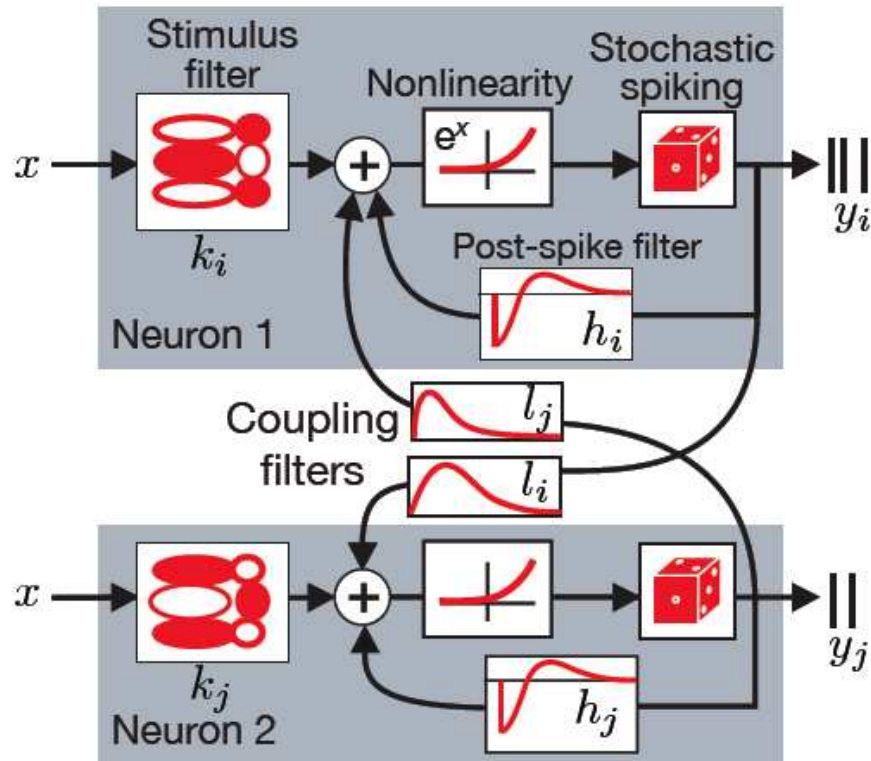
Pros:

- Tractable model-fitting and optimal decoding

- Captures response statistics

Cons:

- Instantaneous coupling filters

- No explicit Common Input

Considering common input effects

RODS AND CONES

HORIZONTAL CELLS

BIPOLAR CELLS

AMACRINE CELLS

GANGLION CELLS

PLATINUM BLACK

SILICON NITRIDE

INDIUM TIN OXIDE

GLASS

Intracellular findings:

- RGCs receive strongly correlated synaptic input in the absence of modulated light stimuli

ON   OFF

1 nA   100 pA

50 ms   50 ms

correlation 0.3 — cross   correlation 0.3
             0.2   0.2
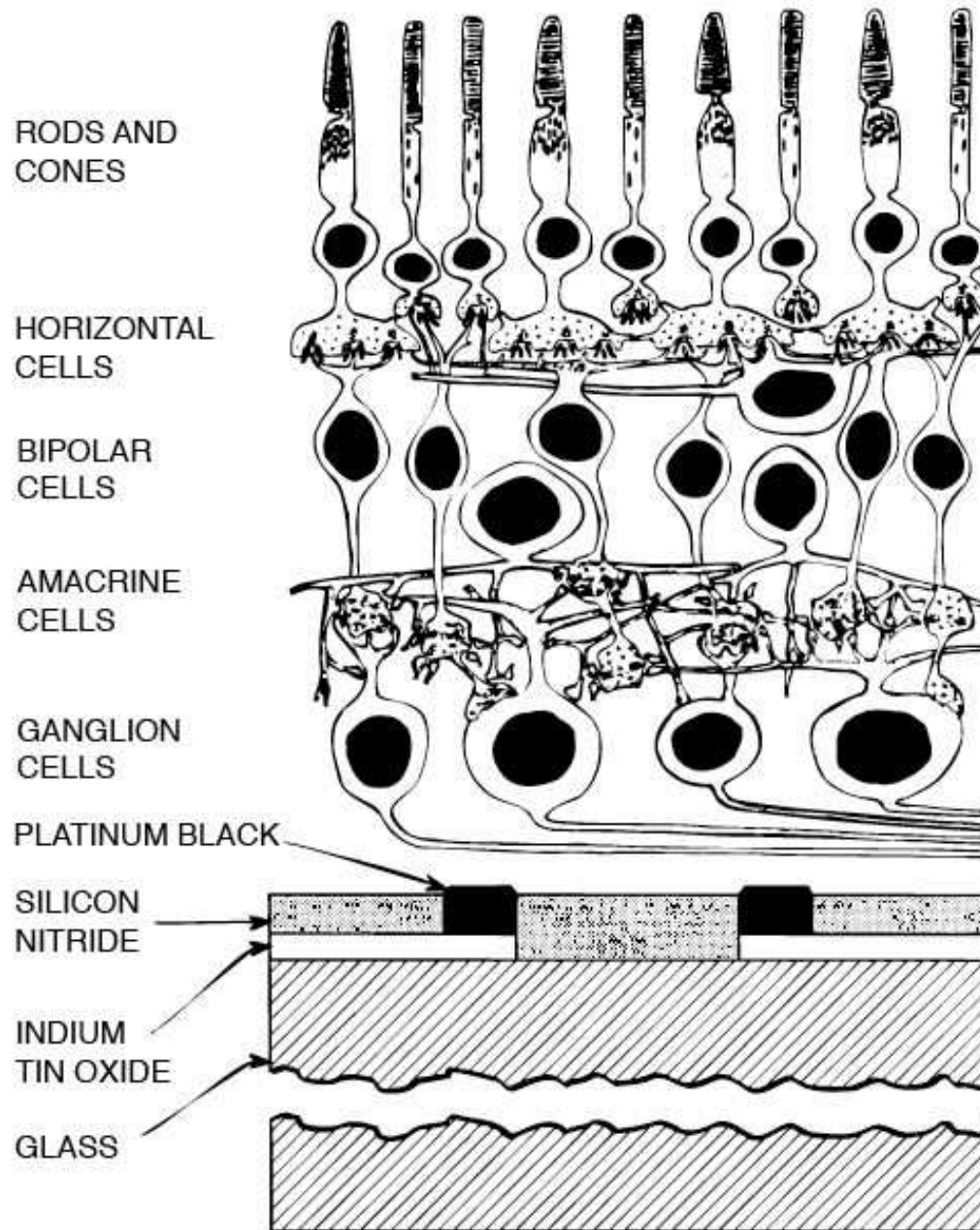   0.1   0.1
             0.0   0.0
        -50   50 ms   -50   50 ms
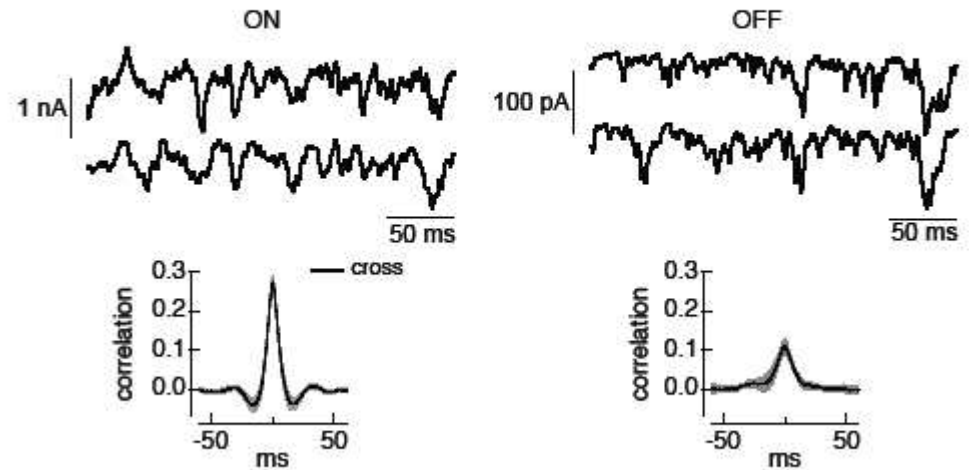
Khuc Trong & Rieke Nature Neuro 2008

- ON RGCs are weakly electrically coupled

- ON RGCs
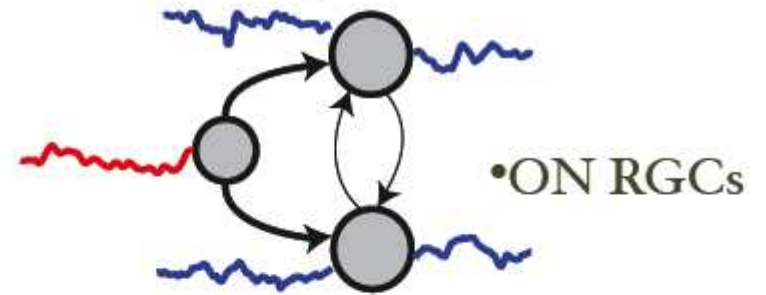
- No electrical coupling seen between OFF RGCs

- OFF RGCs

Extension: including common input effects

$$\lambda_i(t) = \exp\left(k_i \cdot x(t) + h_i \cdot y_i(t) + \sum_{i \neq j} l_{i,j} \cdot y_j(t) + Lq(t)\right)$$

Direct state-space optimization methods

To fit parameters, optimize approximate marginal likelihood:

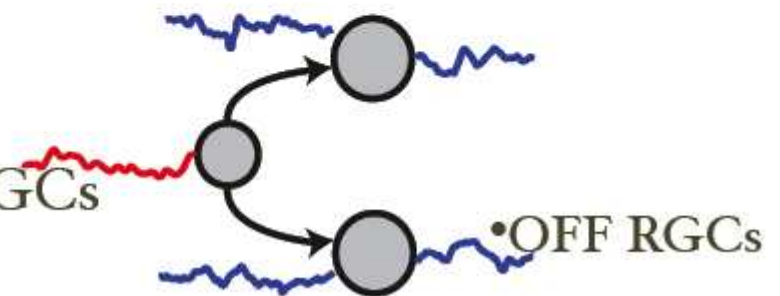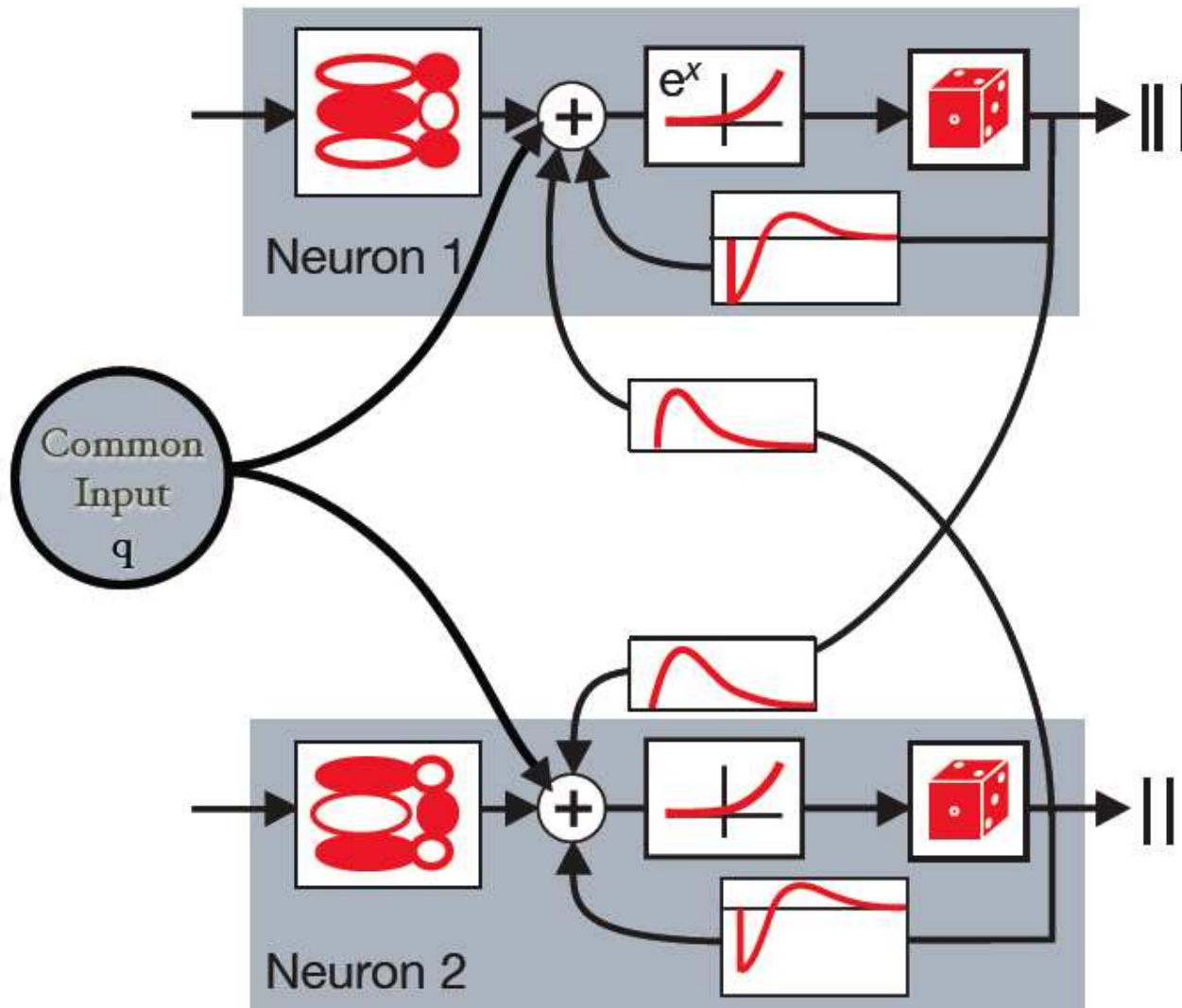$$\log p(spikes|\theta) = \log \int p(Q|\theta)p(spikes|\theta, Q)dQ$$

$$\approx \log p(\hat{Q}_\theta|\theta) + \log p(spikes|\hat{Q}_\theta) - \frac{1}{2}\log|J_{\hat{Q}_\theta}|$$

$$\hat{Q}_\theta = \arg\max_Q \{\log p(Q|\theta) + \log p(spikes|Q)\}$$

— Q is a very high-dimensional latent (unobserved) "common input" term. Taken to be a Gaussian process here with autocorrelation time ≈ 5 ms (Khuc-Trong and Rieke, 2008).

— correlation strength specified by one parameter per cell pair.

— all terms can be computed in $O(T)$ via banded matrix methods (Paninski et al., 2010).

Inferred common input effects are strong
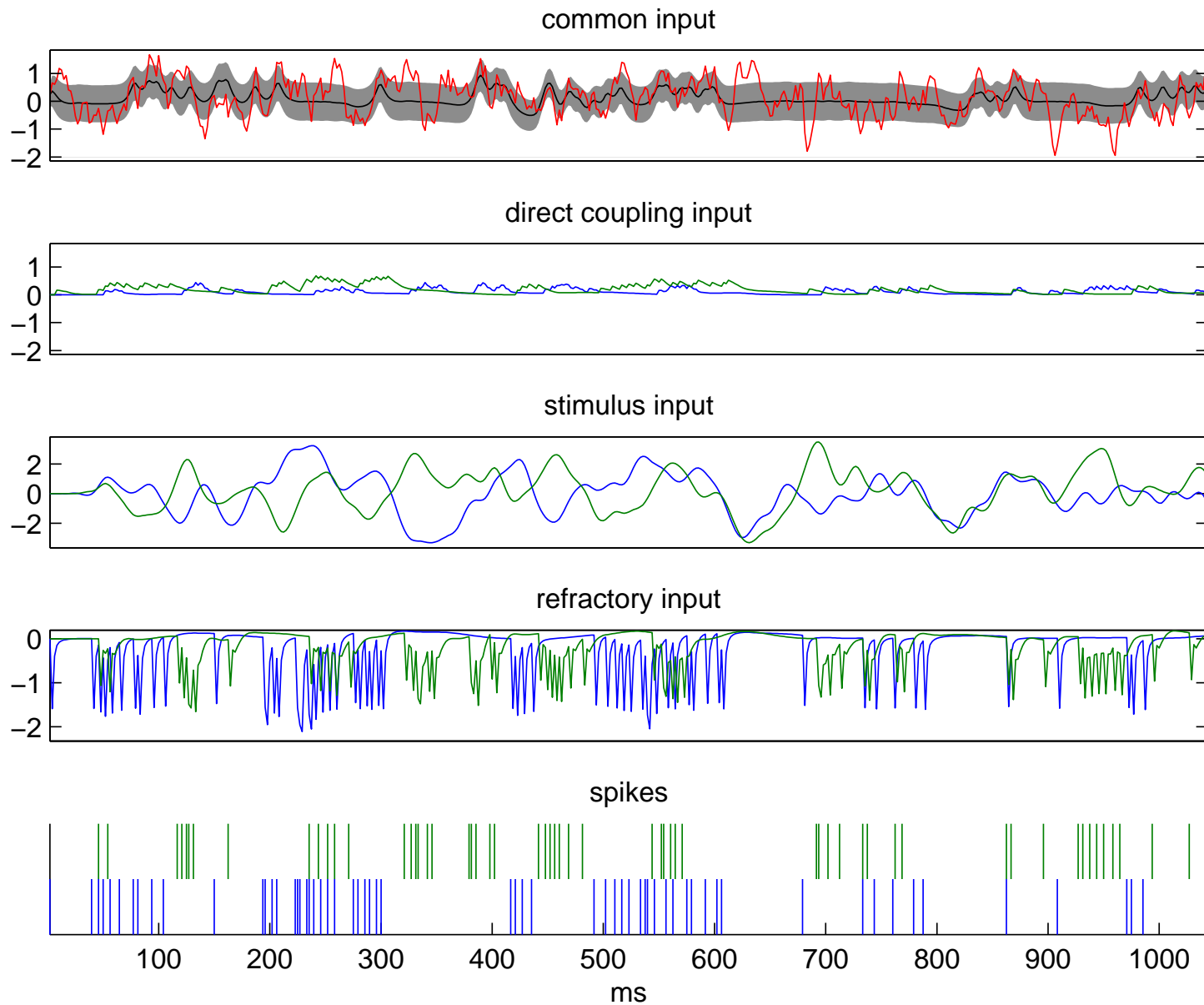
— note that inferred direct coupling effects are now relatively small.

Common-input-only model captures x-corrs
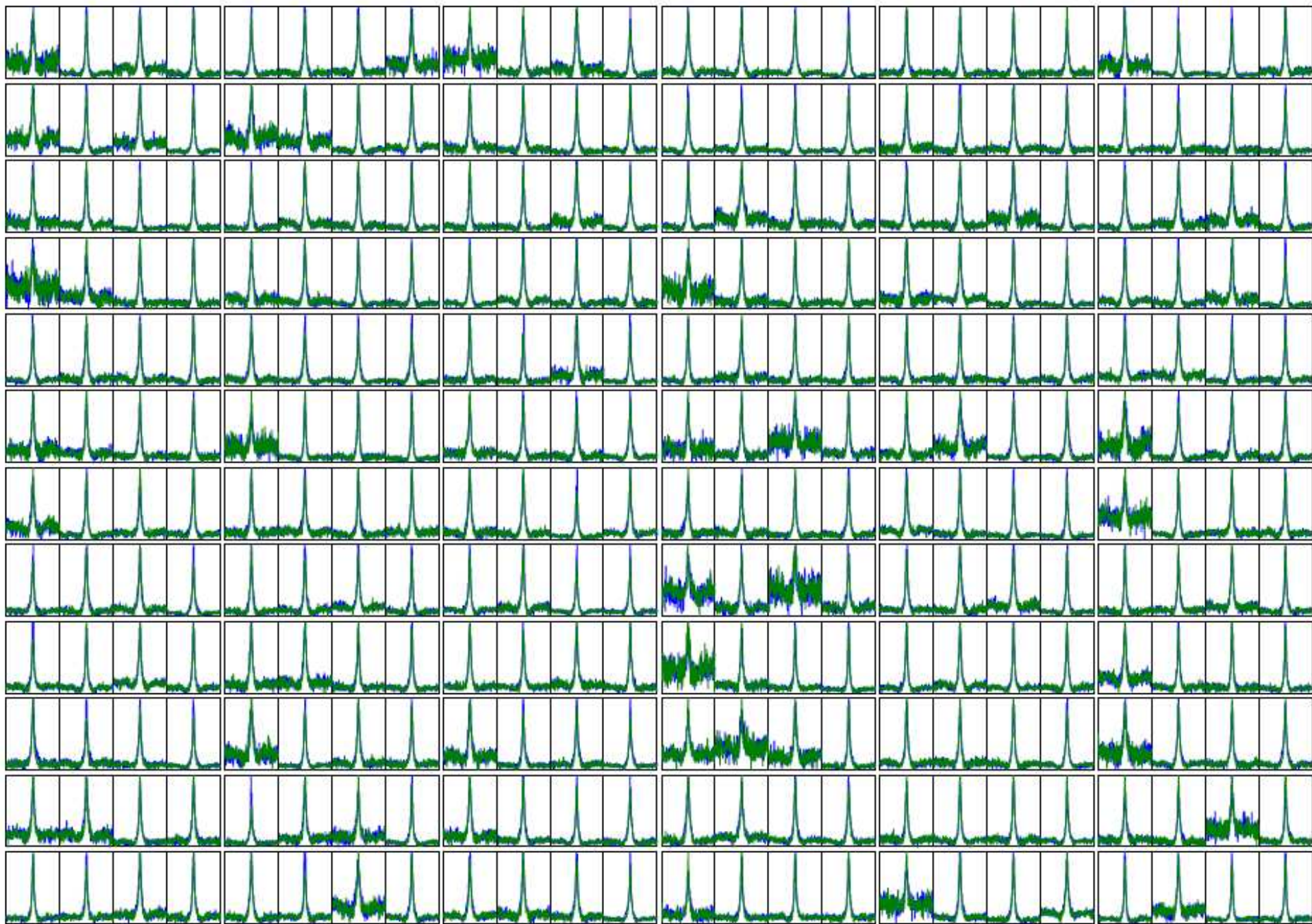
Decoding the stimulus and hidden input
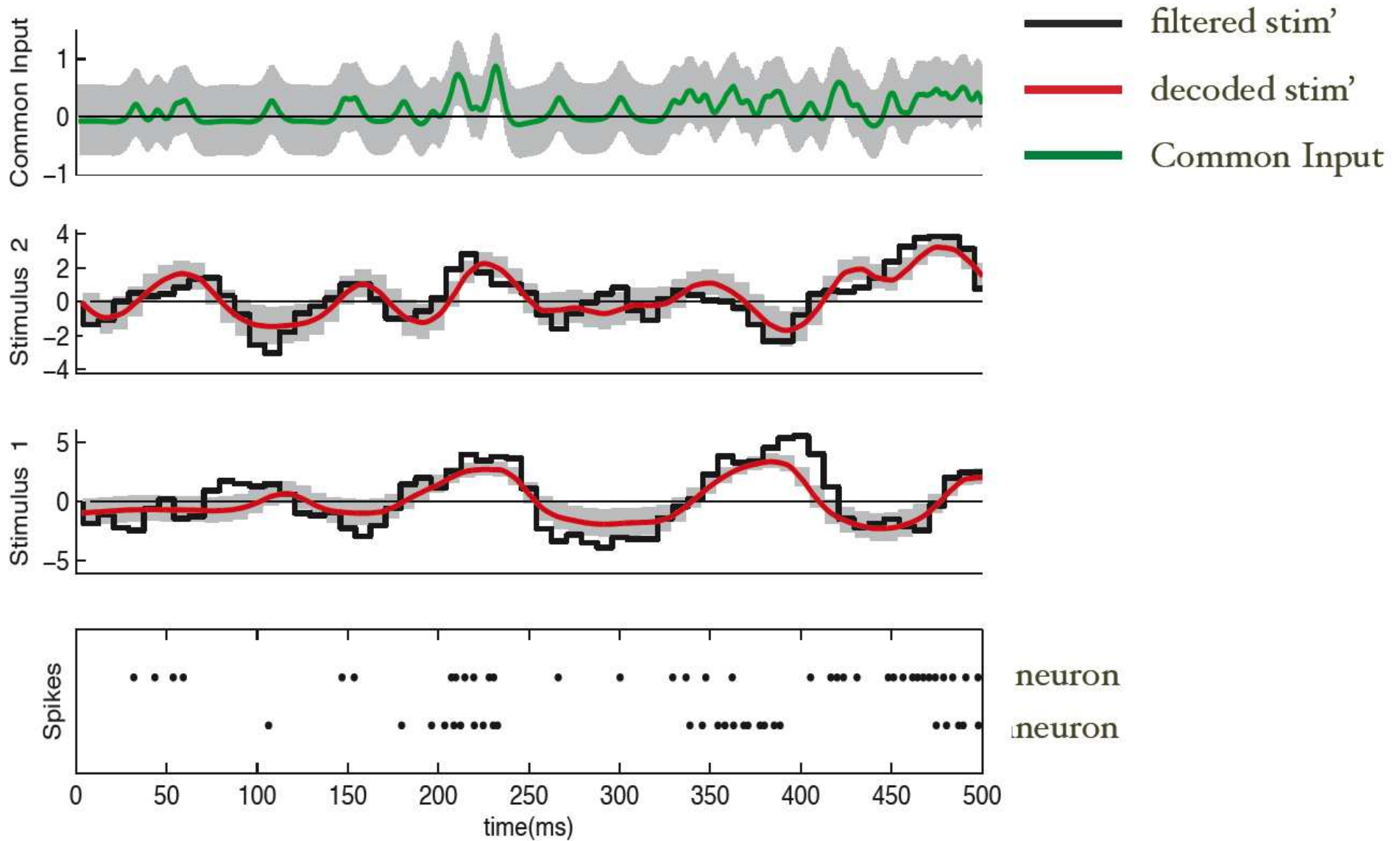
$$\arg\max_{\vec{x}} p(\vec{x}|y,\theta) = \arg\max_{\vec{x}} \int p(\vec{x},Q|y,\theta)dQ \approx \arg\max_{\vec{x},Q} p(\vec{x},Q|y,\theta)$$

Models lead to similar decoding performance
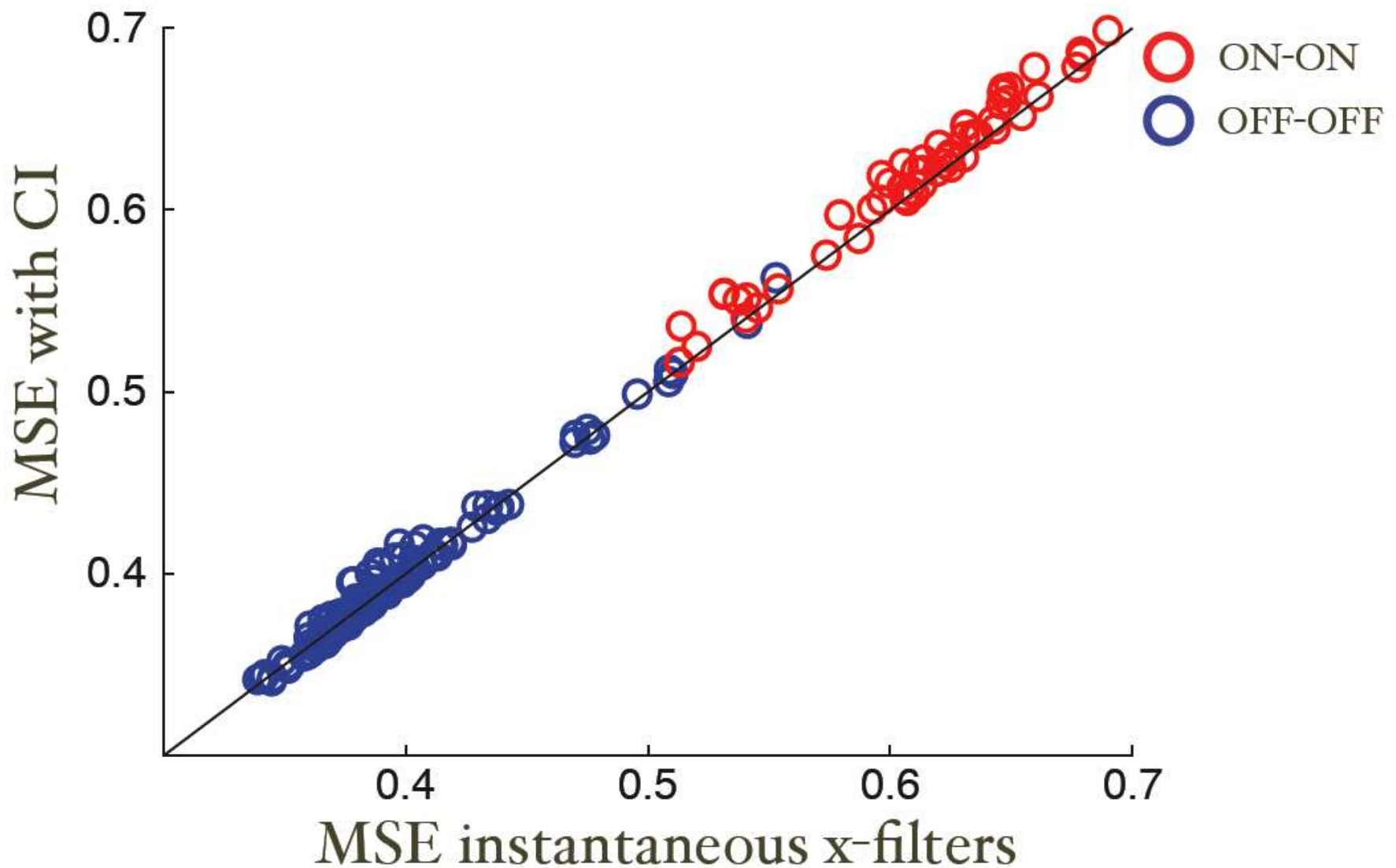
...but CI model is more robust to spike jitter and deletions (Vidne et al. 2010).

Next steps: inferring cones

midget cell   parasol cell   all cells   cone mosaic
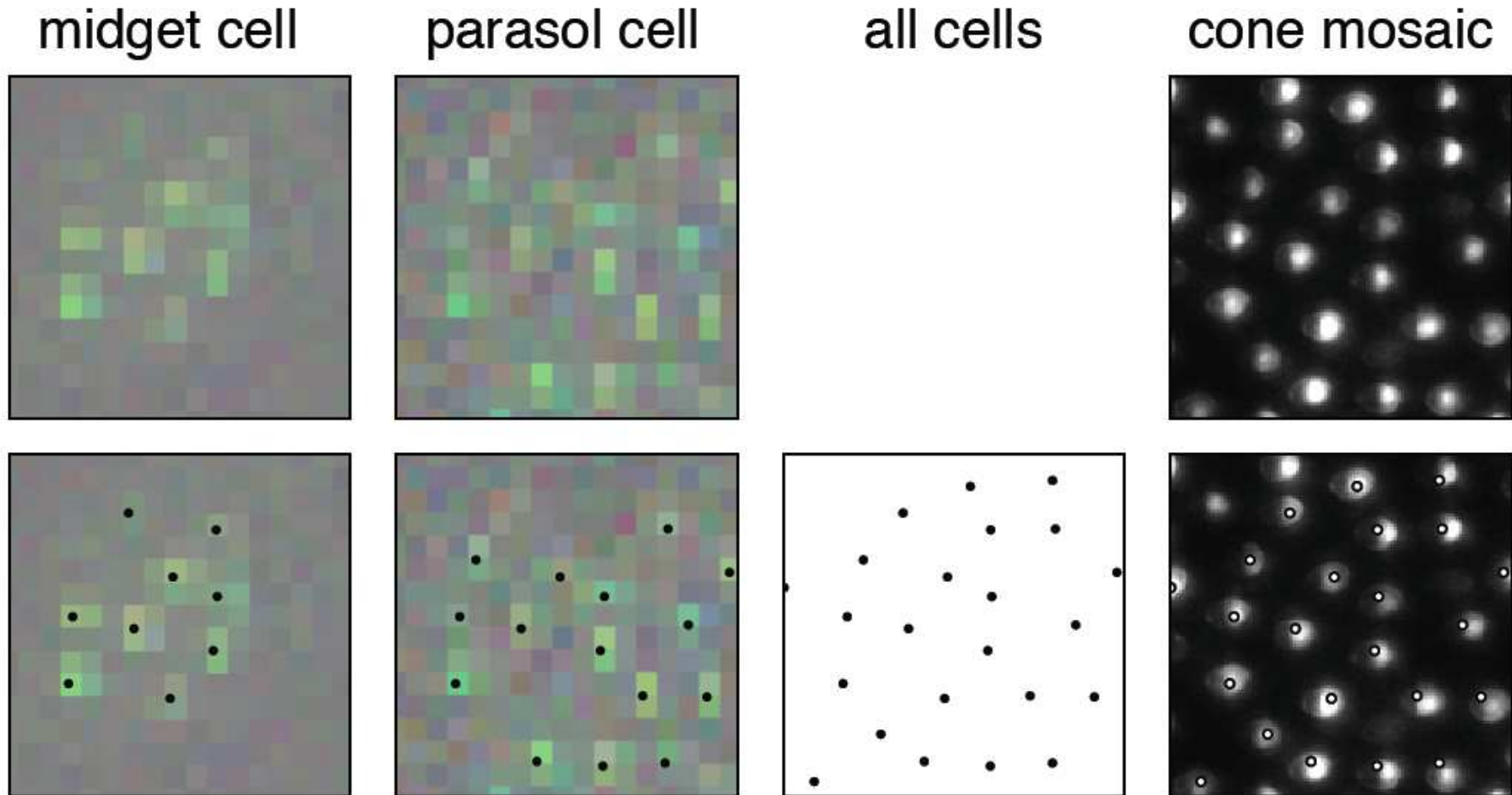
— cone locations and color identity can be inferred accurately with high spatial-resolution stimuli via maximum a posteriori estimates.

ON midget
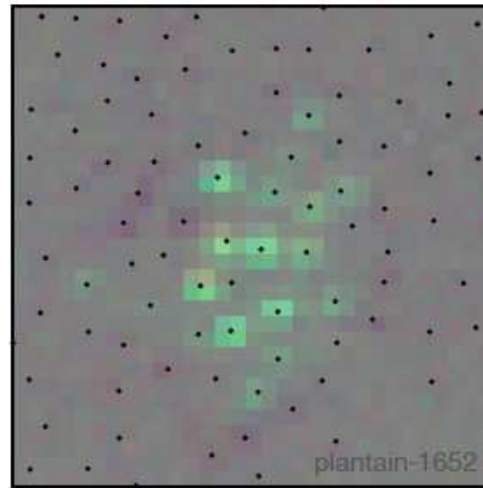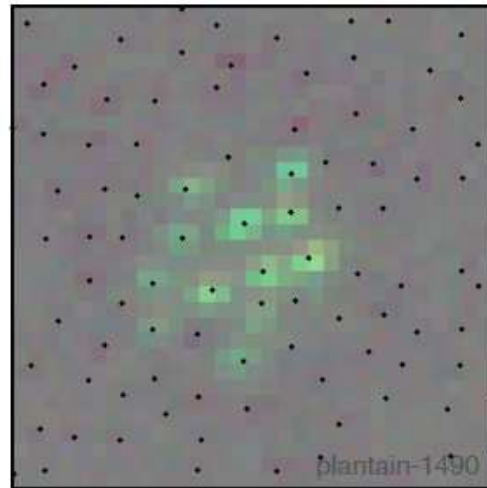
OFF midget
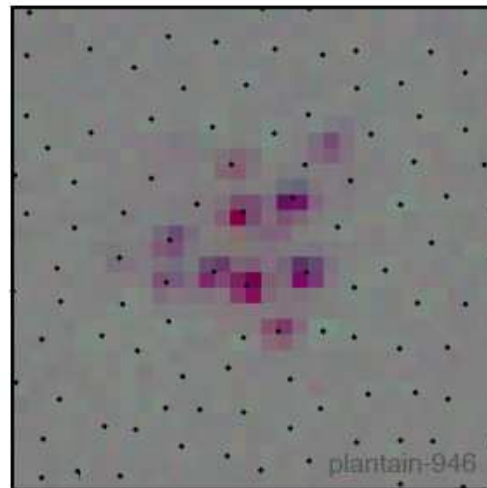
Next steps: inferring circuitry?

ON parasol   retina 1,
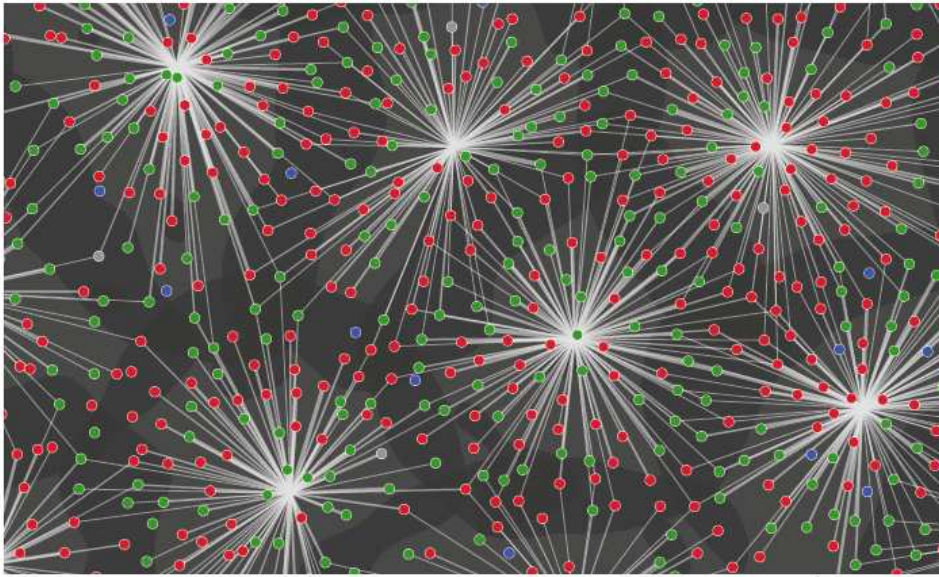
OFF parasol   50 μm
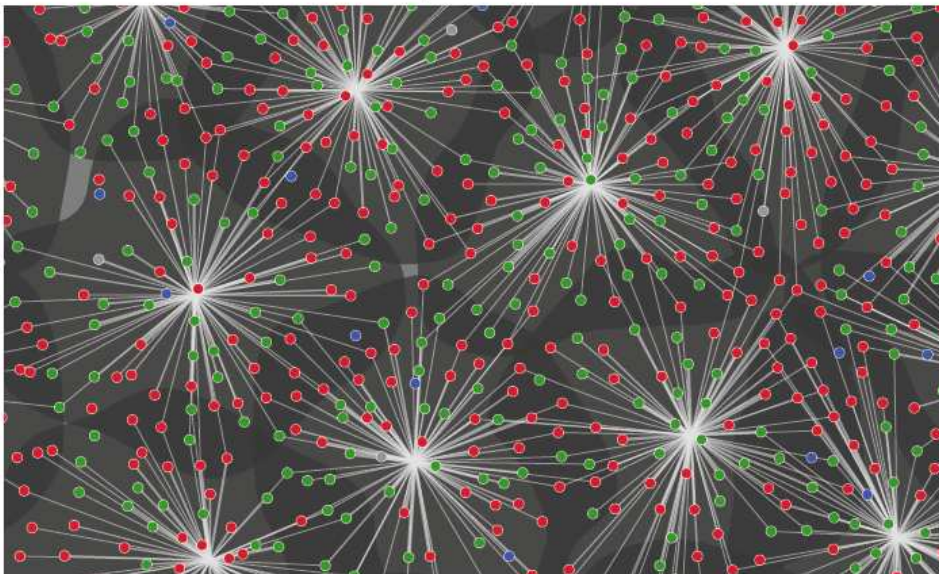
RODS AND CONES

HORIZONTAL CELLS

BIPOLAR CELLS

AMACRINE CELLS

GANGLION CELLS

PLATINUM BLACK

SILICON NITRIDE

INDIUM TIN OXIDE

GLASS
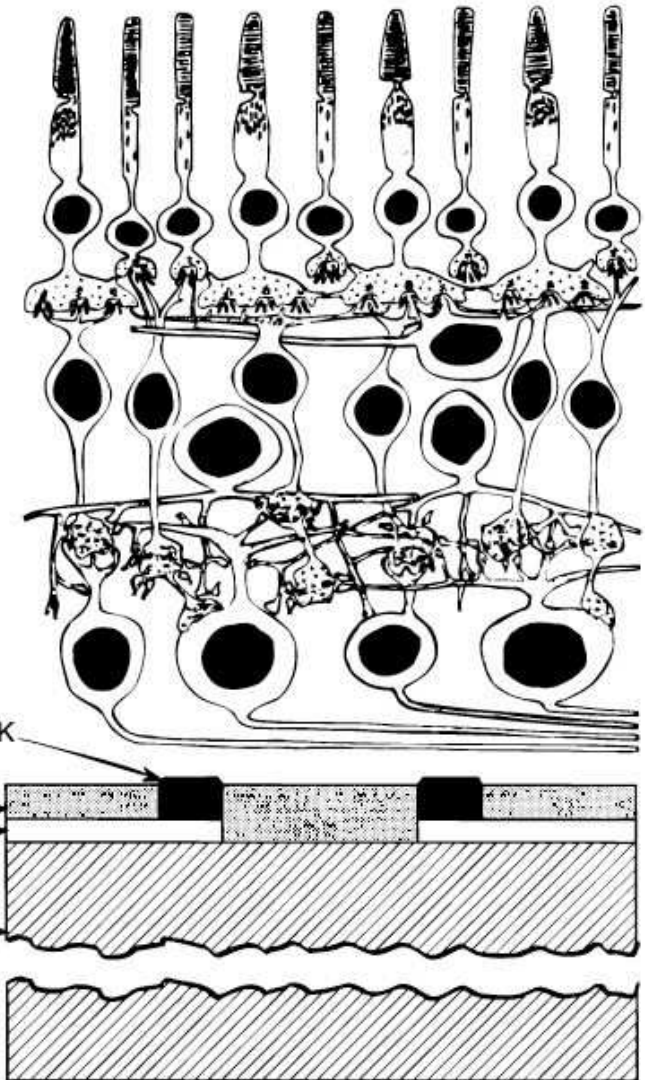
Conclusions

- GLM and state-space approaches provide flexible, powerful methods for answering key questions in neuroscience

- Close relationships between encoding and decoding (Paninski et al., 2007)
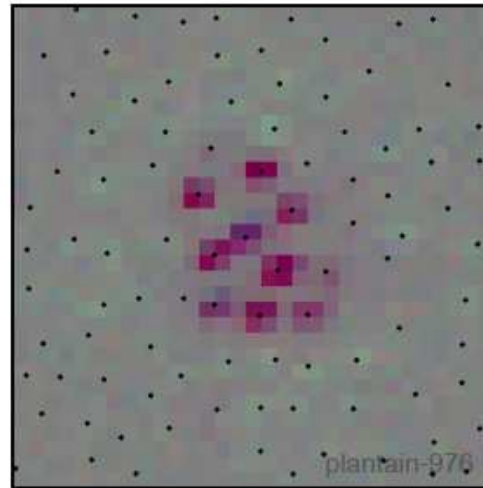
- Log-concavity, banded matrix methods make computations very tractable

- Experimental methods progressing rapidly; many new challenges and opportunities for breakthroughs based on statistical ideas

References

Ahmadian, Y., Pillow, J., and Paninski, L. (2010). Efficient Markov Chain Monte Carlo methods for decoding population spike trains. *Under review, Neural Computation*.

Frechette, E., Sher, A., Grivich, M., Petrusca, D., Litke, A., and Chichilnisky, E. (2005). Fidelity of the ensemble code for visual motion in the primate retina. *J Neurophysiol*, 94(1):119–135.

Khuc-Trong, P. and Rieke, F. (2008). Origin of correlated activity between parasol retinal ganglion cells. *Nature Neuroscience*, 11:1343–1351.

Lalor, E., Ahmadian, Y., and Paninski, L. (2009). The relationship between optimal and biologically plausible decoding of stimulus velocity in the retina. *Journal of the Optical Society of America A*, 26:25–42.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15:243–262.

Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, In press.

Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. In Cisek, P., Drew, T., and Kalaska, J., editors, *Computational Neuroscience: Progress in Brain Research*. Elsevier.

Pillow, J., Ahmadian, Y., and Paninski, L. (2009). Model-based decoding, information estimation, and change-point detection in multi-neuron spike trains. *Under review, Neural Computation*.

Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A., Chichilnisky, E., and Simoncelli, E. (2008). Spatiotemporal correlations and visual signaling in a complete neuronal population. *Nature*, 454:995–999.

Truccolo, W., Eden, U., Fellows, M., Donoghue, J., and Brown, E. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *Journal of Neurophysiology*, 93:1074–1089.