Statistical methods for understanding neural codes:

Multineuronal spike coding in primate retina

Liam Paninski

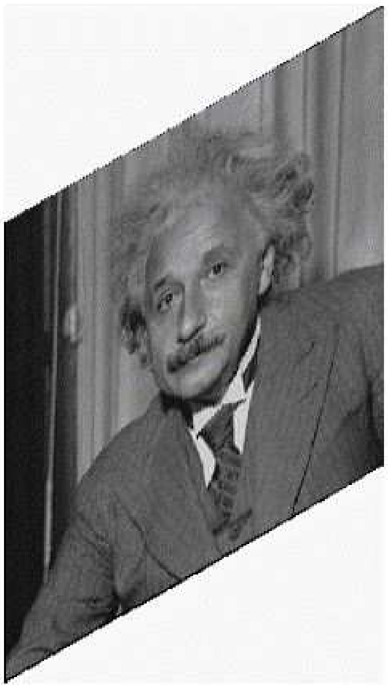
Department of Statistics and Center for Theoretical Neuroscience
Columbia University
http://www.stat.columbia.edu/∼liam
liam@stat.columbia.edu

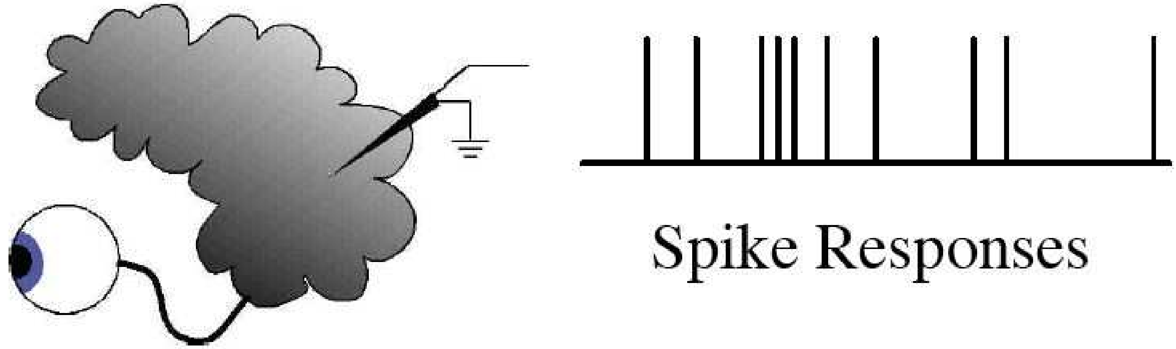
April 19, 2007

— with J. Pillow (Gatsby), E. Simoncelli (NYU), E.J. Chichilnisky, V. Uzzell, J. Shlens (Salk), T. Toyoizumi (Columbia).

Support: NIH CRCNS award, Sloan Research Fellowship.

The neural code





Input-output relationship between

- External observables x (sensory stimuli, motor responses...)
- Neural variables y (spike trains, population activity...)

Probabilistic formulation: $p(y|x)$

Basic goal

...learning the neural code.

Fundamental question: how to estimate $p(y|x)$ from experimental data?

General problem is too hard — not enough data, too many inputs x and spike trains y

Avoiding the curse of insufficient data

Many approaches to make problem tractable:

1: Estimate some functional $f(p)$ instead

e.g., information-theoretic quantities (Nemenman et al., 2002; Paninski, 2003)

2: Select stimuli as efficiently as possible (Foldiak, 2001; Machens, 2002; Paninski, 2005; Lewi et al., 2006)

3: Fit a model with small number of parameters

Neural encoding models

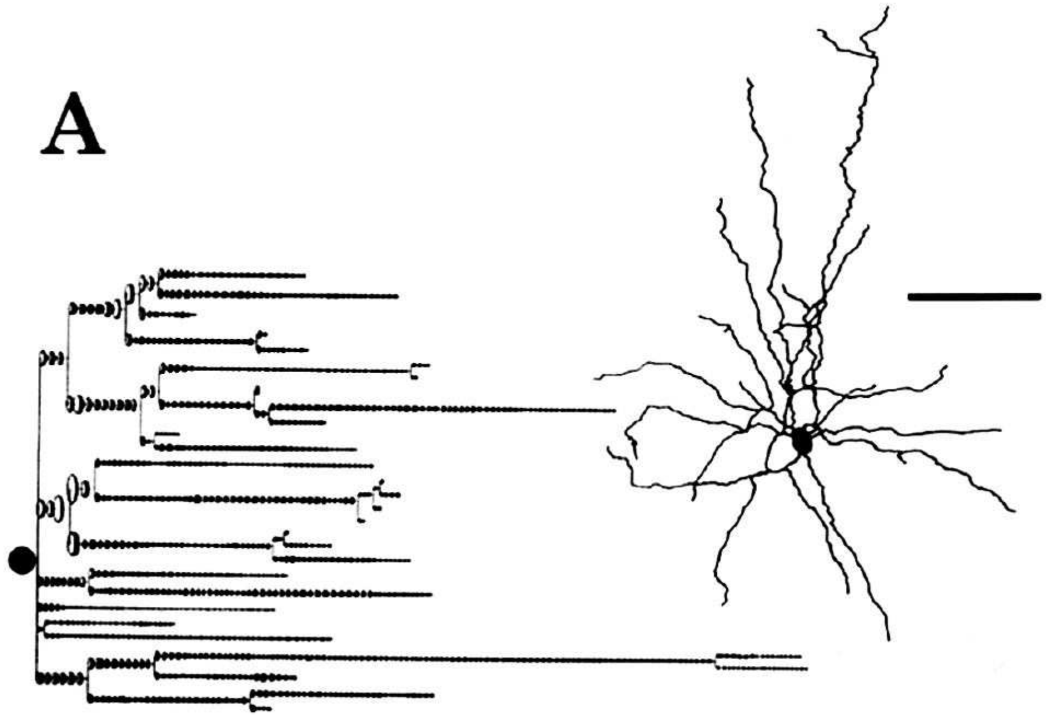
"Encoding model": $p_\theta(y|x)$.

— Fit parameter θ instead of full $p(y|x)$

Main theme: want model to be flexible but not overly so

Flexibility vs. "fittability"

Multiparameter HH-type model



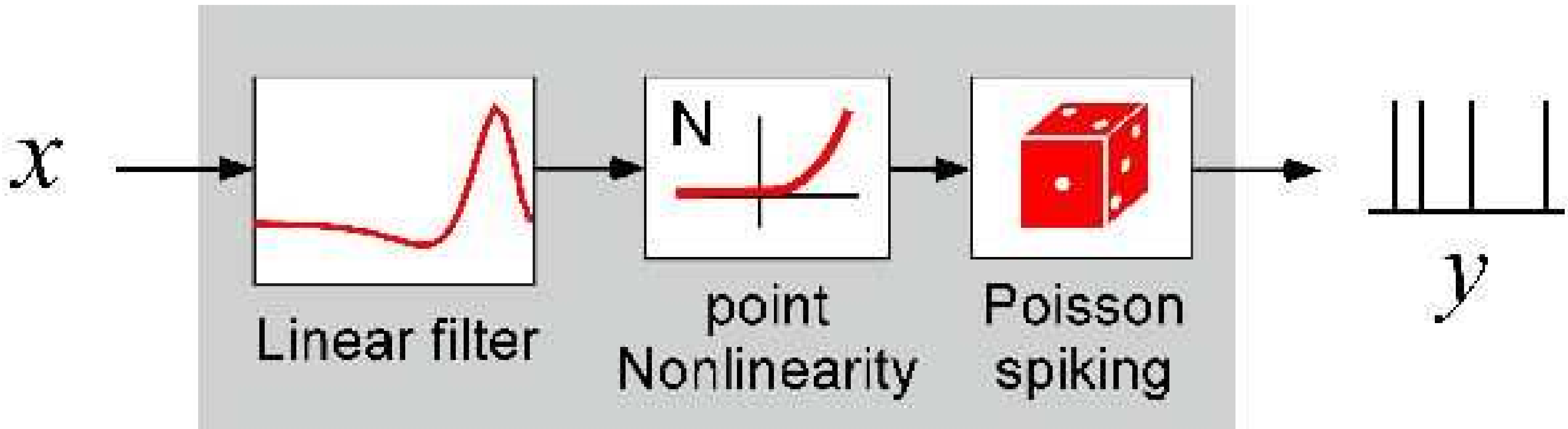
Model	Current	Regional Conductances (mS/cm²) Dendrites	Soma	AH	NR	Axon
EC2.5 REAL	I_{Ca}	2.0	1.5	1.5	—	—
$j = 1$	$I_{K,Ca}$	0.001	0.065	0.065	0.065	0.065
SD* (real) = 21.9 μm	I_{Na}	25	80	100–150†	100	40–70‡
SD (EC2.5) = 20 μm	I_K	12	18	18	18	12–18‡
$\tau_{Ca} = 1.5$	I_A	36	54	54	54	—
$E_L = -60$ mV	Leak (Real)	0.008	0.008	0.008	0.008	0.008
$E_{Na} = 35$ mV	(EC2.5)	0.005	0.005	0.005	0.005	0.005

— highly biophysically plausible, flexible

— **but** very difficult to estimate parameters given spike times alone

(figure adapted from (Fohlmeister and Miller, 1997))

Cascade ("LNP") model



— easy to estimate via correlation-based methods (Simoncelli et al., 2004)

— **but** not biophysically plausible (fails to capture spike timing details: refractoriness, burstiness, adaptation, etc.)

Two key ideas

1. Use likelihood-based methods for fitting.

   — well-justified statistically

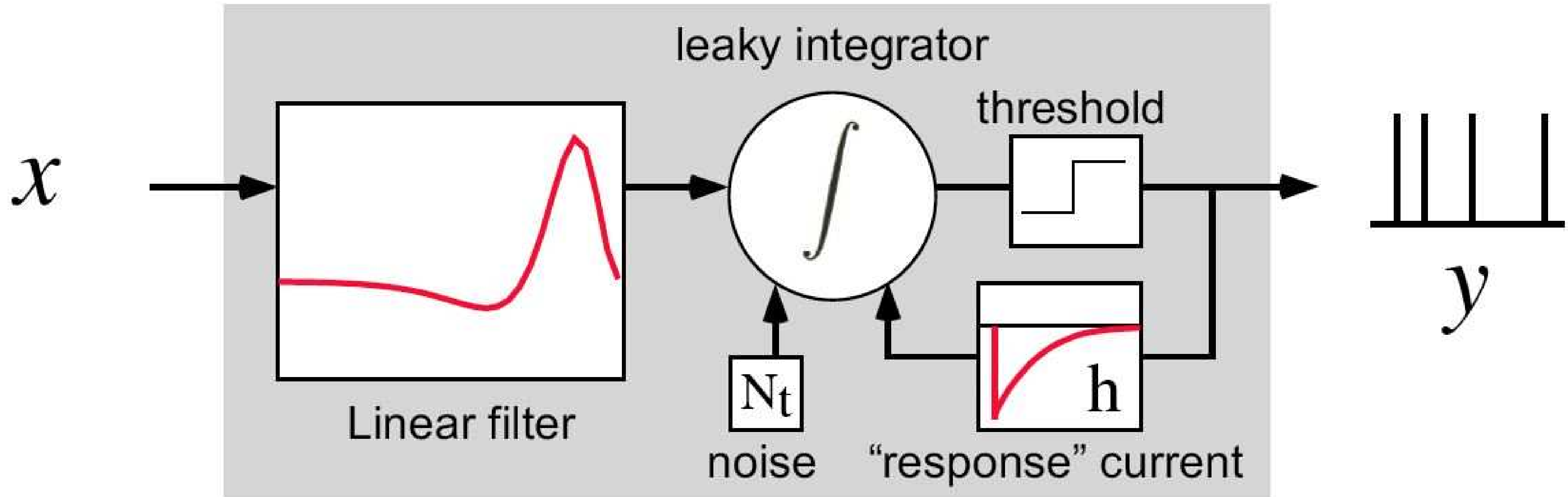
   — easy to incorporate prior knowledge, explicit noise models, etc.

2. Use models that are easy to fit via maximum likelihood

   — **concave** (downward-curving) functions have no non-global local maxima $\Longrightarrow$ concave functions are easy to maximize by gradient ascent.

Recurring theme: find flexible models whose loglikelihoods are guaranteed to be concave.

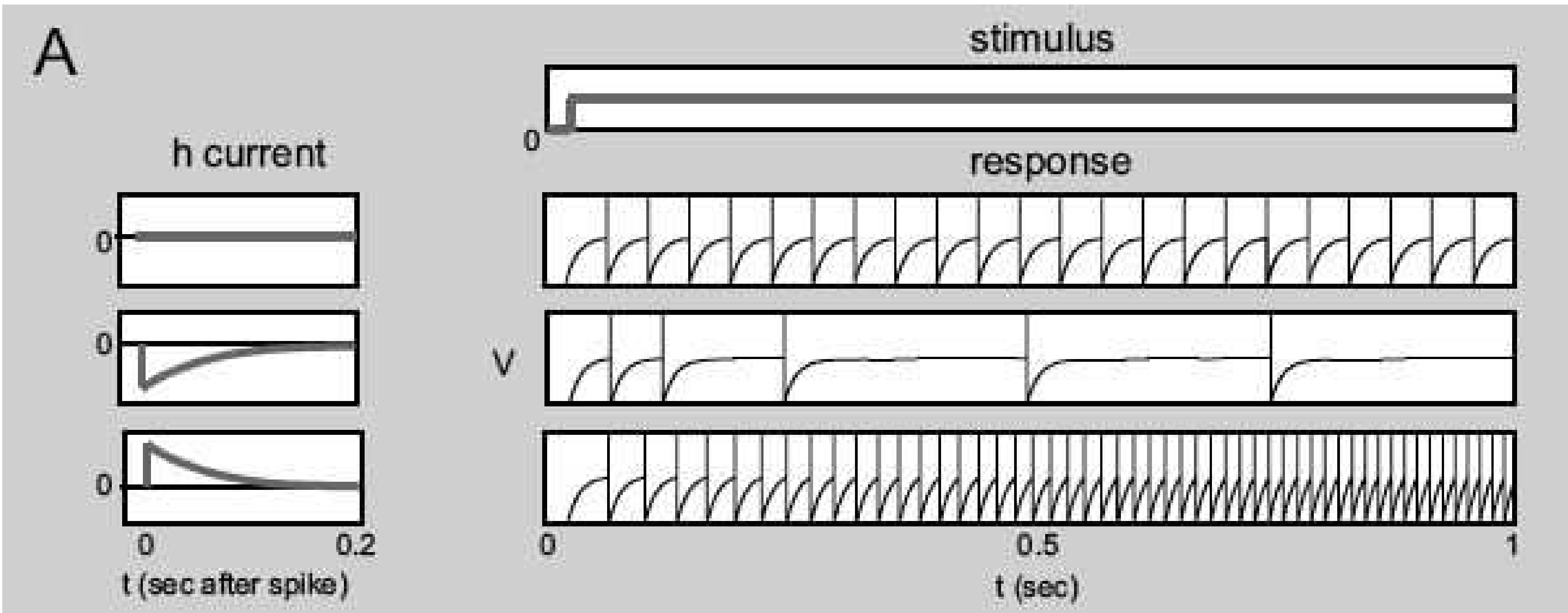
Filtered integrate-and-fire model



$$dV(t) = \left(-g(t)V(t) + I_{DC} + \vec{k} \cdot \vec{x}(t) + \sum_{j=-\infty}^{0} h(t - t_j) \right) dt + \sigma dN_t;$$

(Gerstner and Kistler, 2002; Paninski et al., 2004)

Model flexibility: Adaptation



The estimation problem

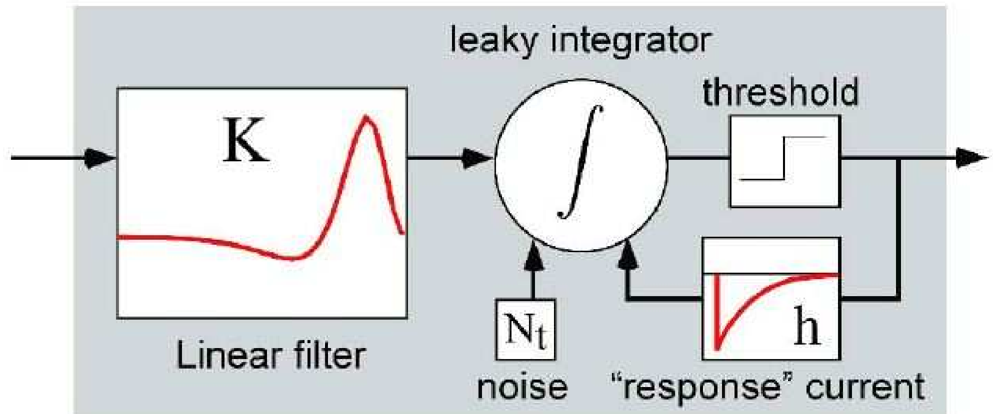
Learn the model parameters:

$\vec{K} =$ stimulus filter

$g =$ leak conductance

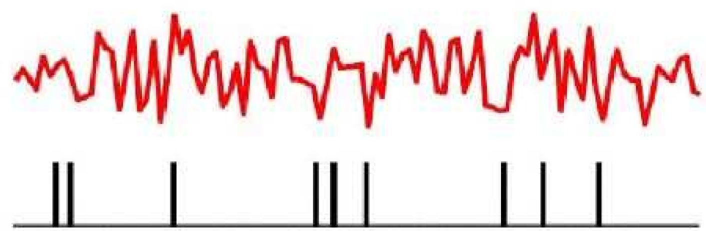
$\sigma^2 =$ noise variance

$\vec{h} =$ response current



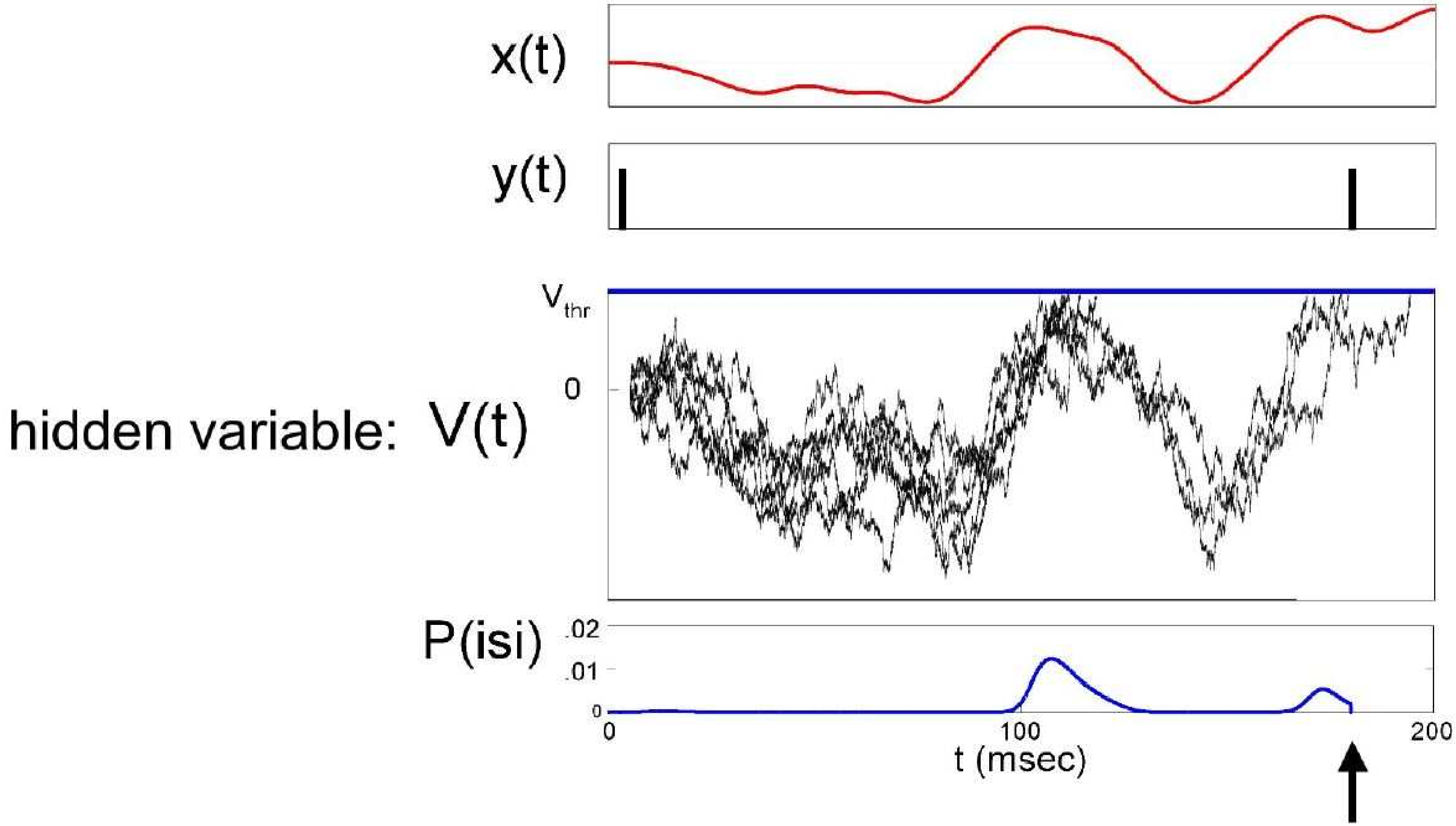
From: stimulus train $x(t)$

spike times t_i



(Paninski et al., 2004)

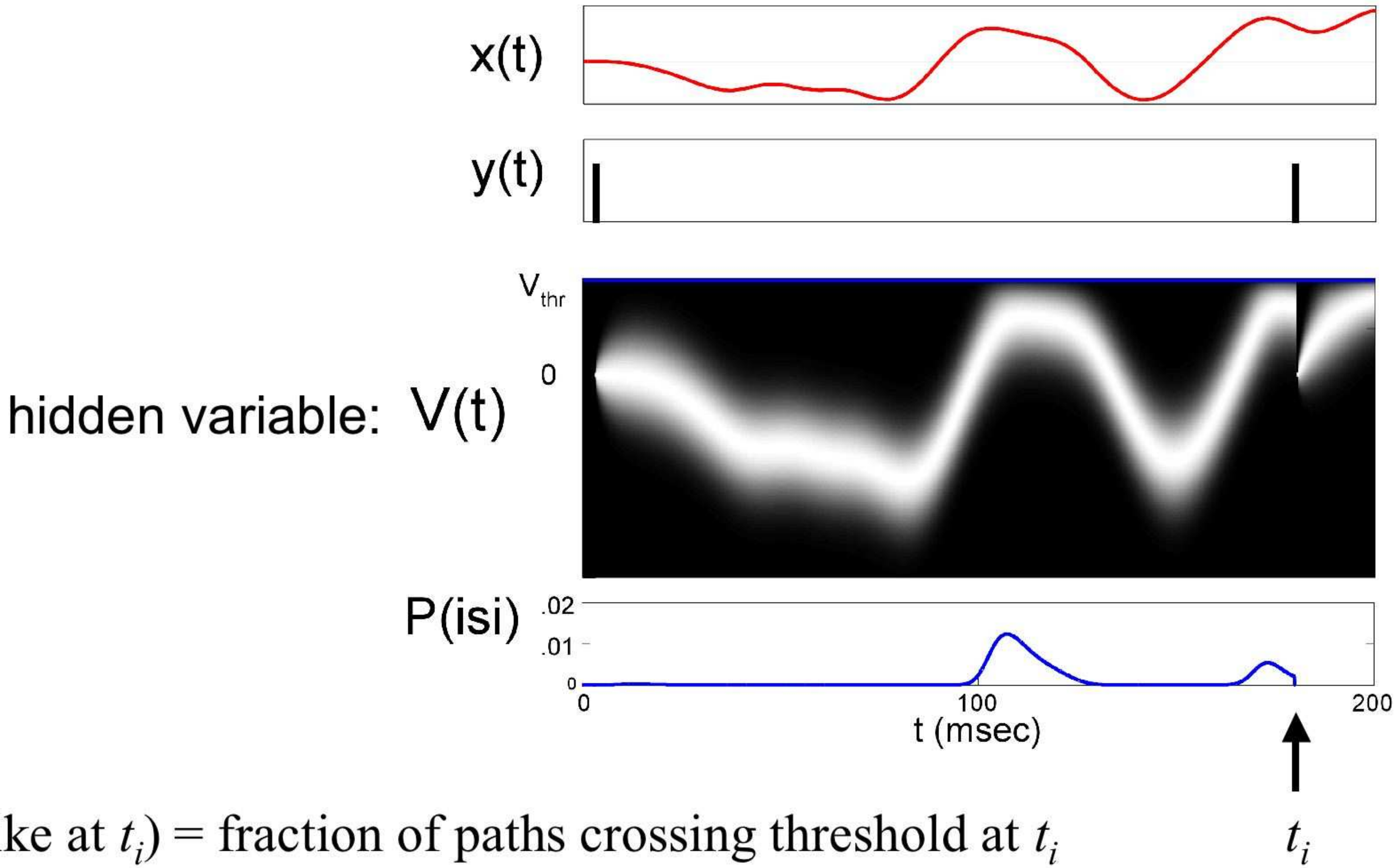
First passage time likelihood



$P(\text{spike at } t_i) = \text{fraction of paths crossing threshold for first time at } t_i$

(computed numerically via Fokker-Planck or integral equation methods)

Likelihood function



P(spike at t_i) = fraction of paths crossing threshold at t_i

Computing Likelihood

Diffusion Equation:

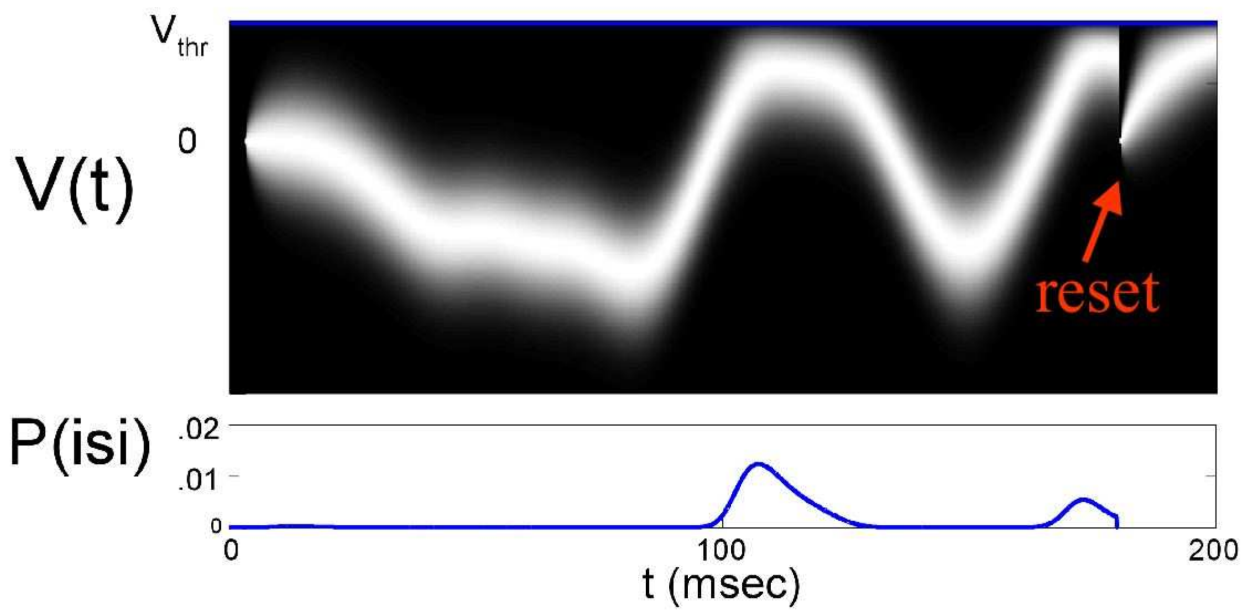
$$\frac{\partial P(V,t)}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial V^2} + g\frac{\partial[(V-V_0)P]}{\partial V},$$

- linear dynamics
- additive Gaussian noise

fast methods for solving linear PDE

$\Rightarrow$

efficient procedure for computing likelihood



ISIs are conditionally independent $\Rightarrow$ likelihood is product over ISIs

Maximizing likelihood

Maximization seems difficult, even intractable:

— high-dimensional parameter space

— likelihood is a complex nonlinear function of parameters

Main result: The loglikelihood is concave in the parameters, no matter what data $\{\vec{x}(t), t_i\}$ are observed.

$\Longrightarrow$ no non-global local maxima

$\Longrightarrow$ maximization easy by ascent techniques.

Proof of log-concavity theorem

Based on probability integral representation of likelihood:

$$L(\theta) = \int 1(\mathbf{V} \in C) dG_{\vec{x},\theta}(\mathbf{V})$$

$G_{\vec{x},\theta}(\mathbf{V})$ = OU-measure on voltage paths $\mathbf{V}$

C = set of voltage paths $V(t)$ consistent with spike data:

$$V(t) \leq V_{th};\ V(t_i) = V_{th};\ V(t_i^+) = V_{reset}$$

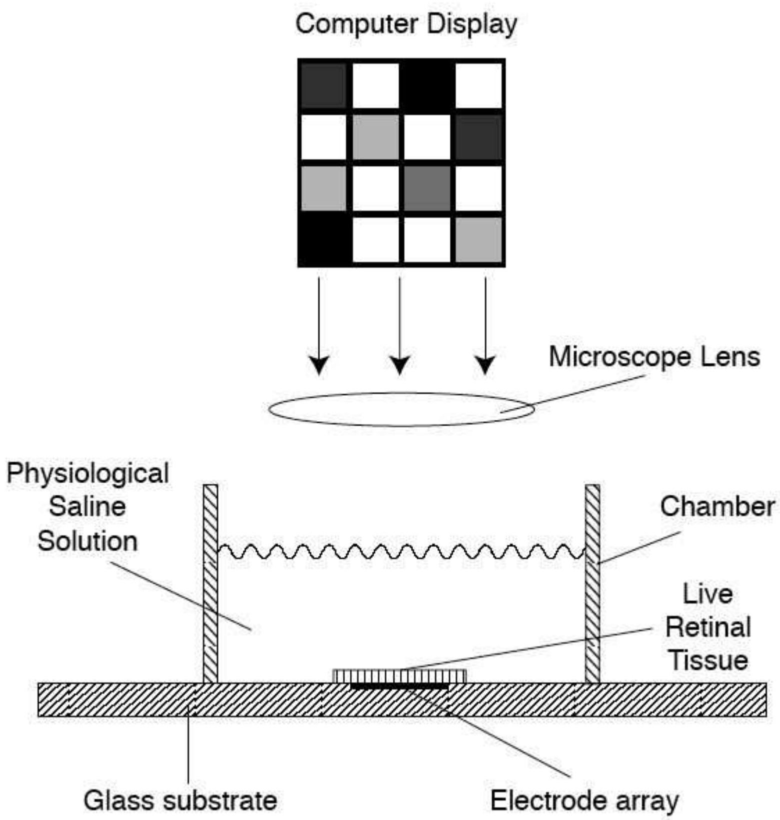
Now use fact that marginalizing preserves log-concavity (Prekopa, 1973): if $f(\vec{x}, \vec{y})$ is jointly l.c., then so is

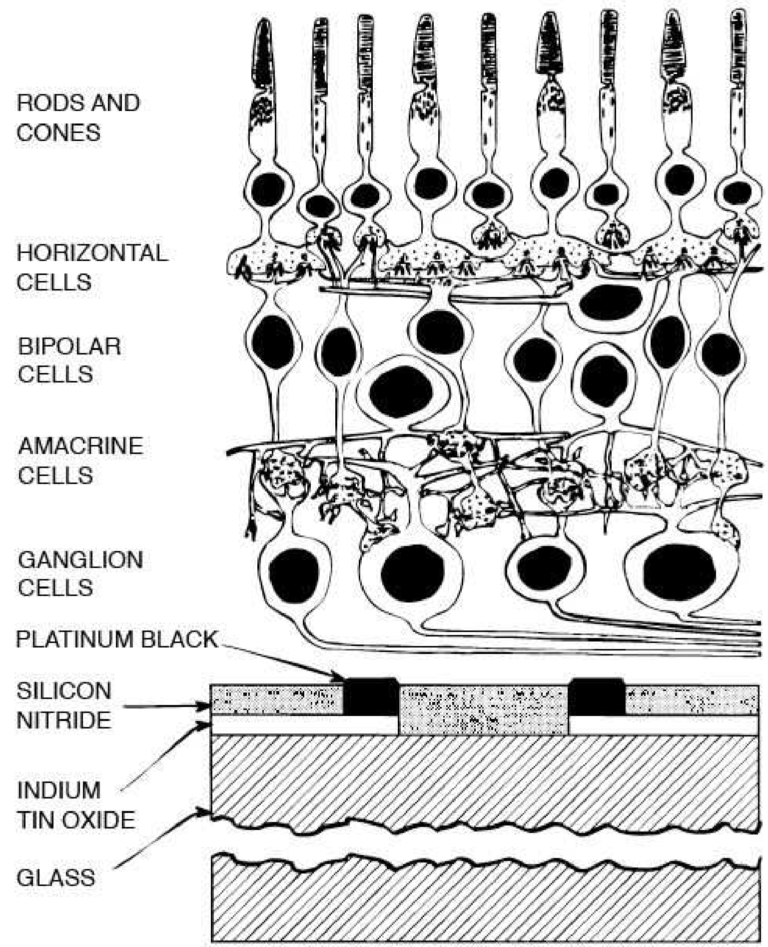
$$f_0(\vec{x}) \equiv \int f(\vec{x}, \vec{y}) d\vec{y}.$$

Application: retinal ganglion cells

Preparation: dissociated salamander and macaque retina

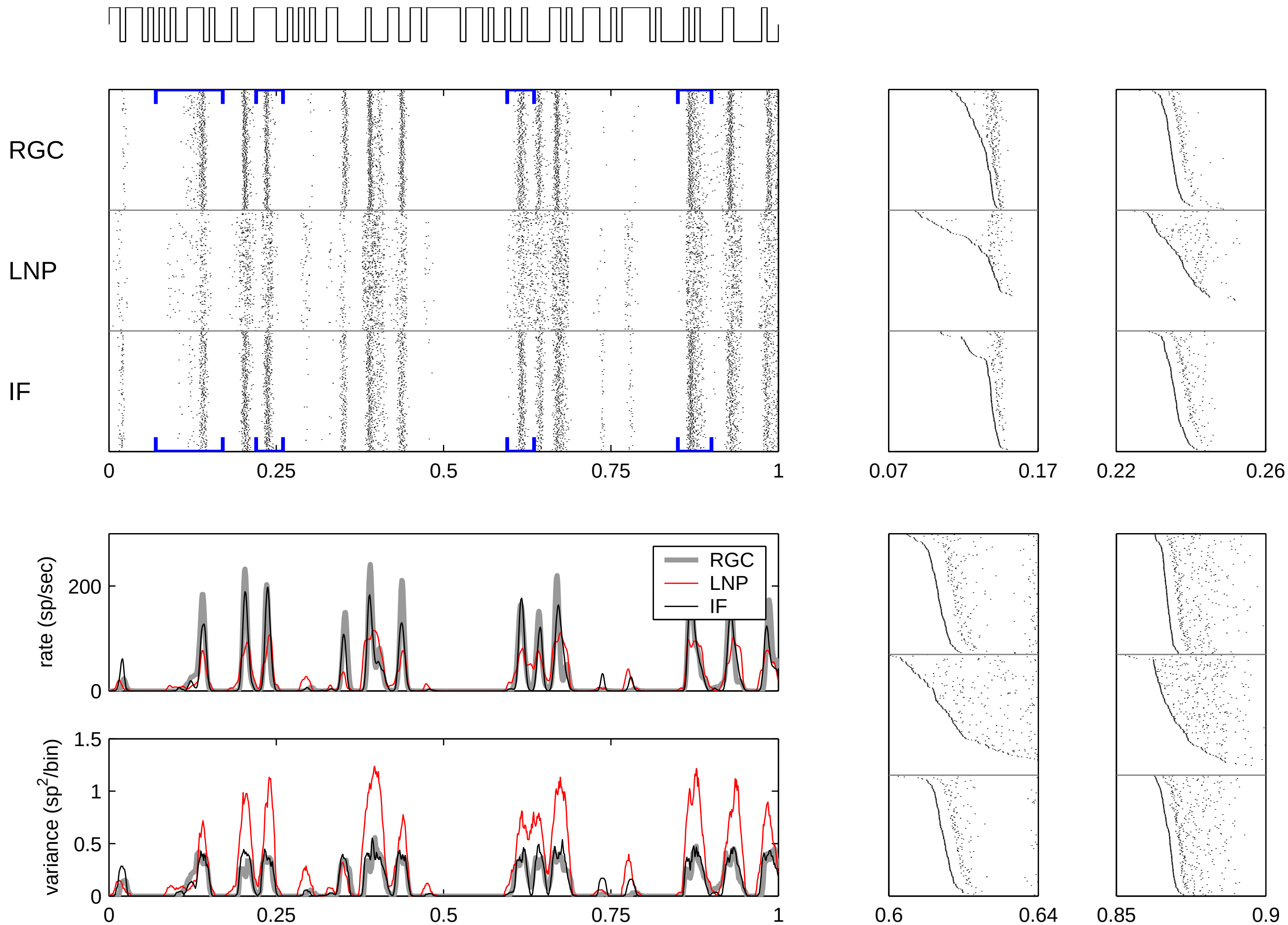
— extracellularly-recorded responses of populations of RGCs





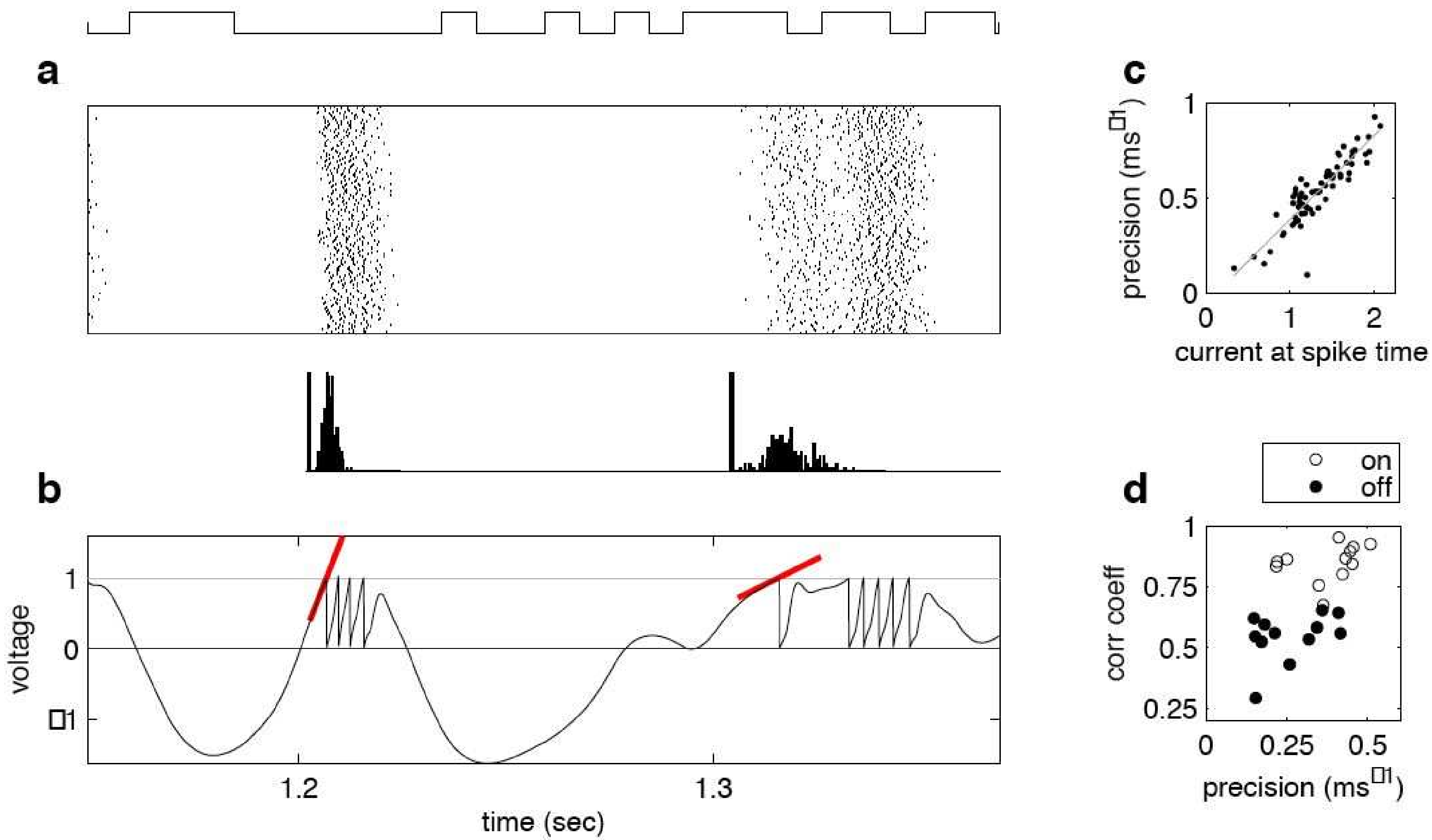
Stimulus: random "flicker" visual stimuli (Chander and Chichilnisky, 2001)

Spike timing precision in retina



(Pillow et al., 2005b)

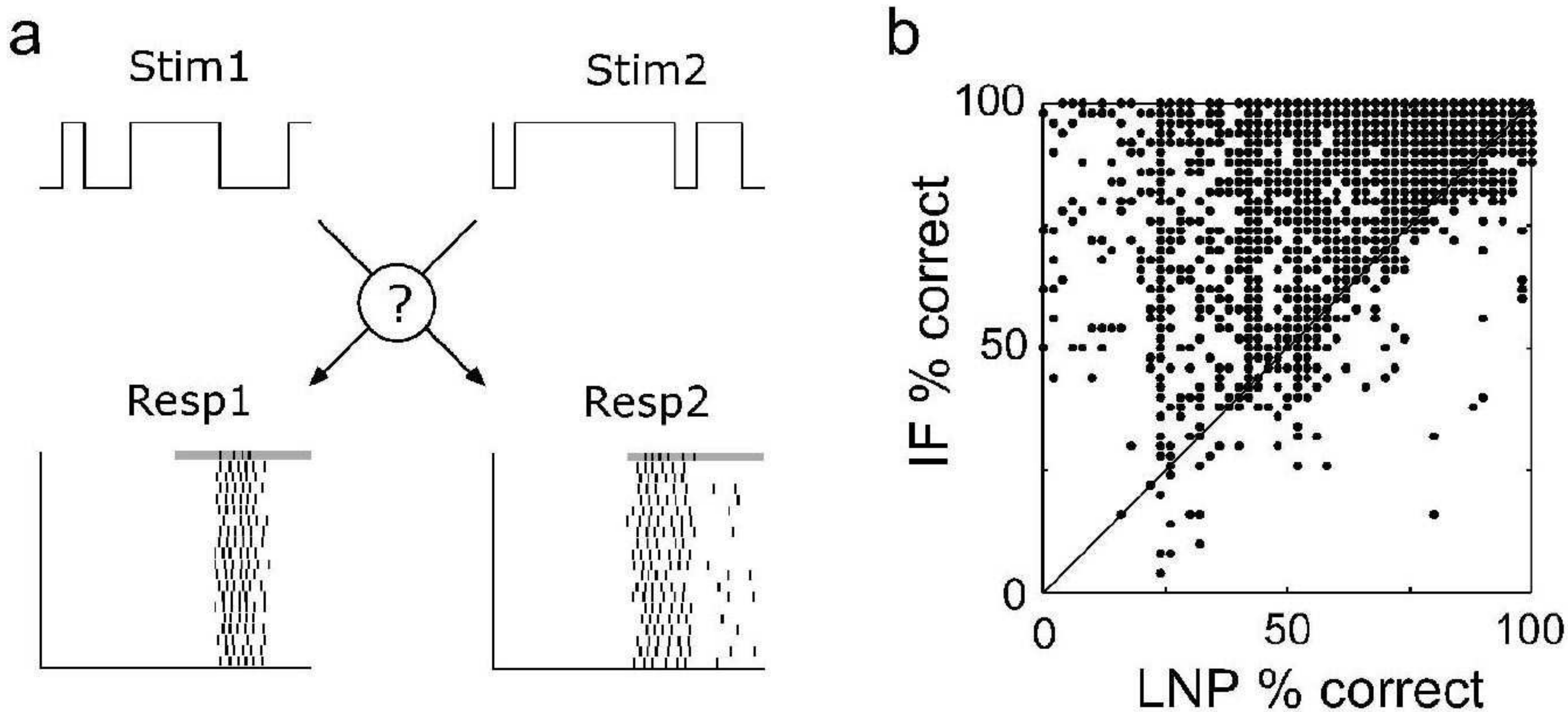
Linking spike reliability and subthreshold noise



(Pillow et al., 2005b)

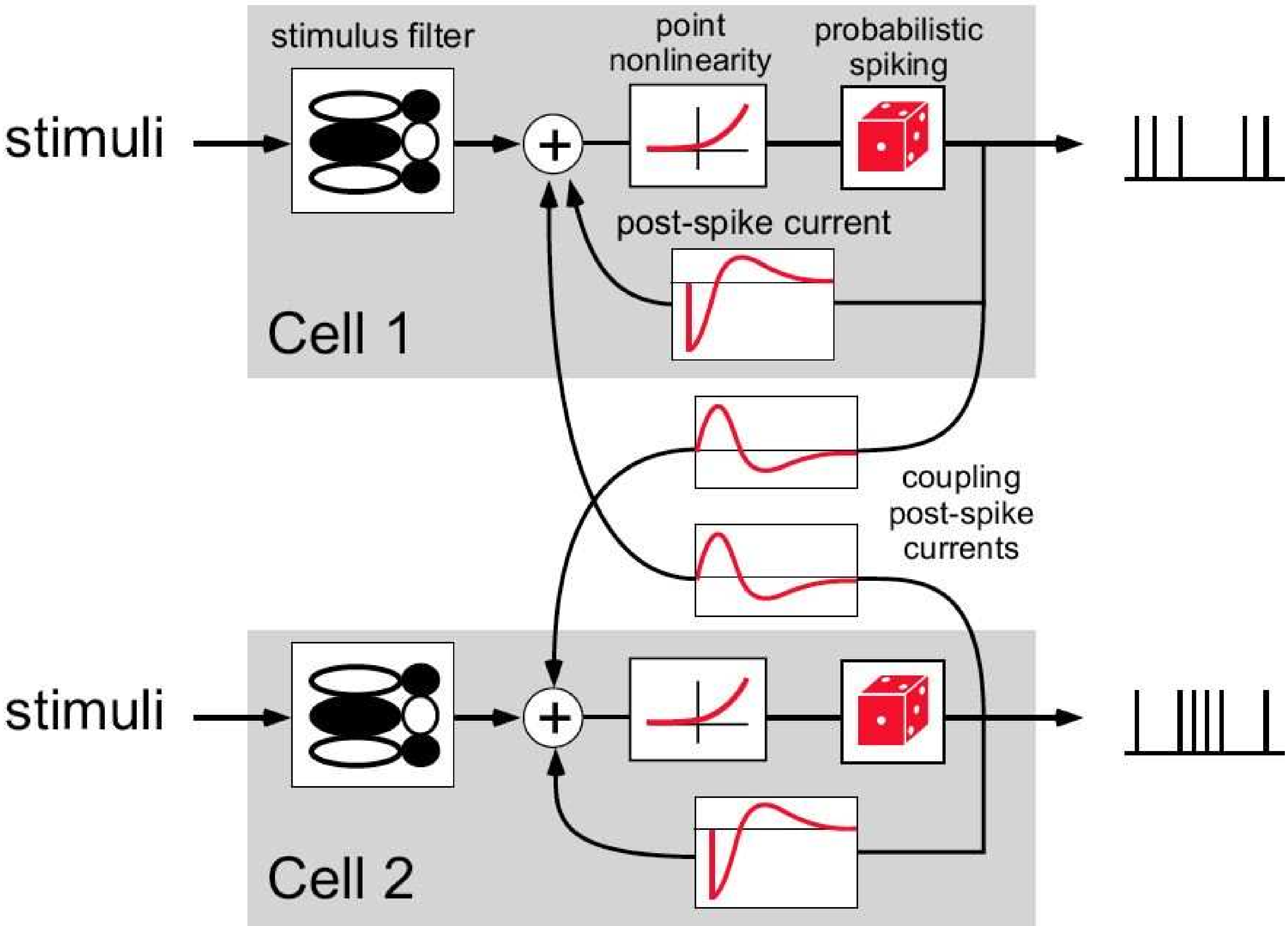
Likelihood-based discrimination

Given spike data, optimal decoder chooses stimulus $\vec{x}$ according to likelihood: $p(spikes|\vec{x}_1)$ vs. $p(spikes|\vec{x}_2)$.

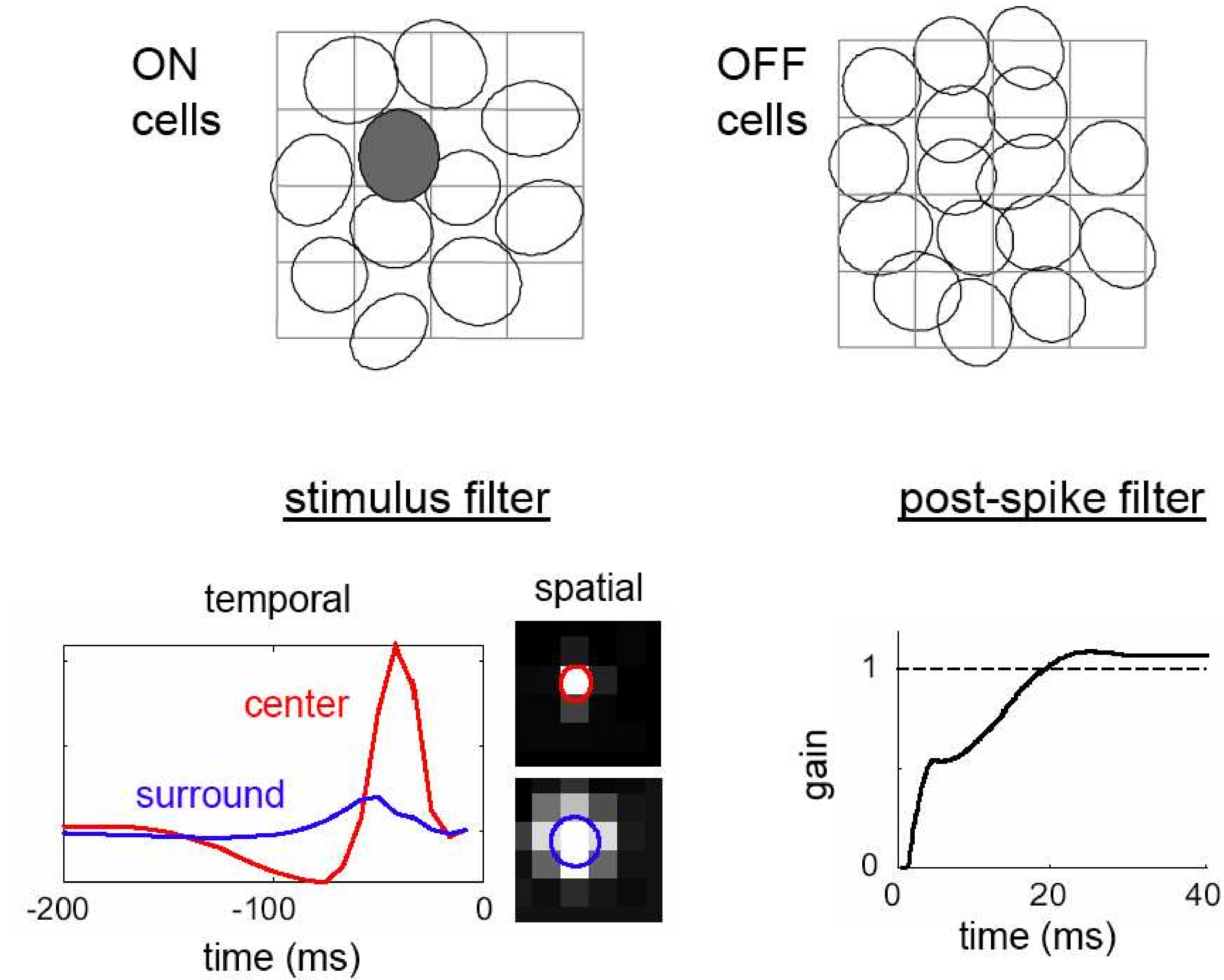


Using accurate model is essential (Pillow et al., 2005b)

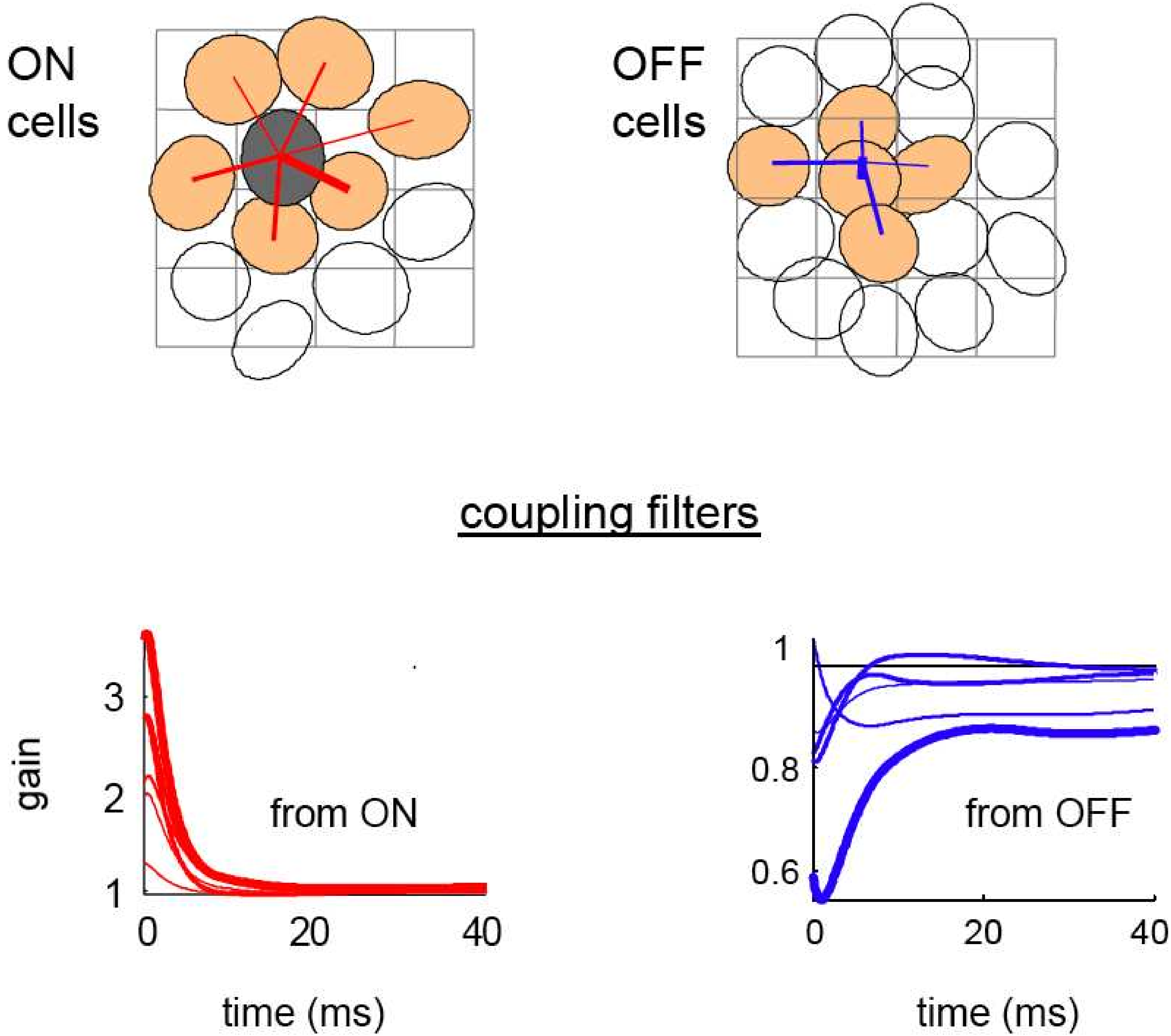
Generalization: population responses



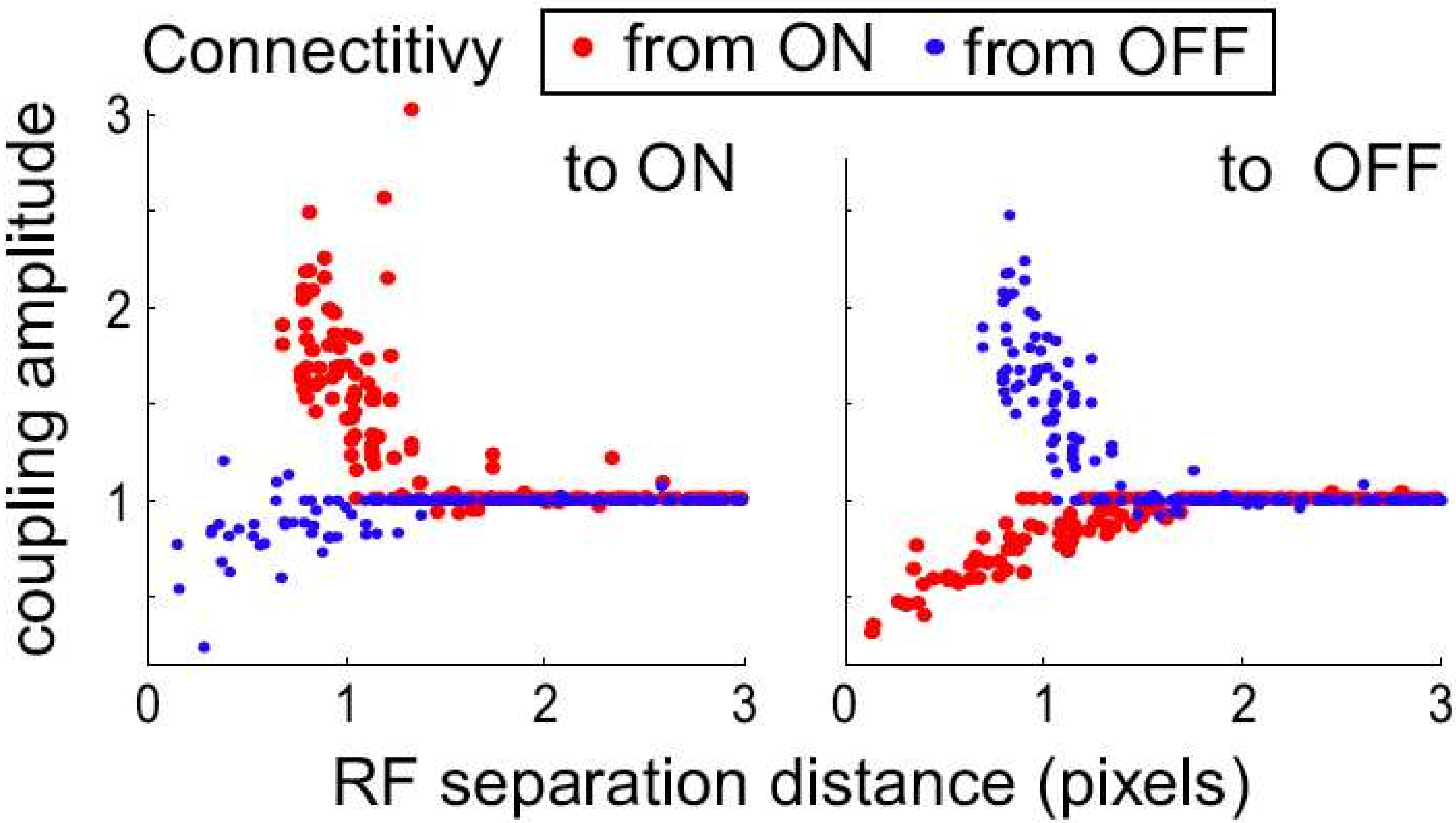
(Pillow et al., 2005a)



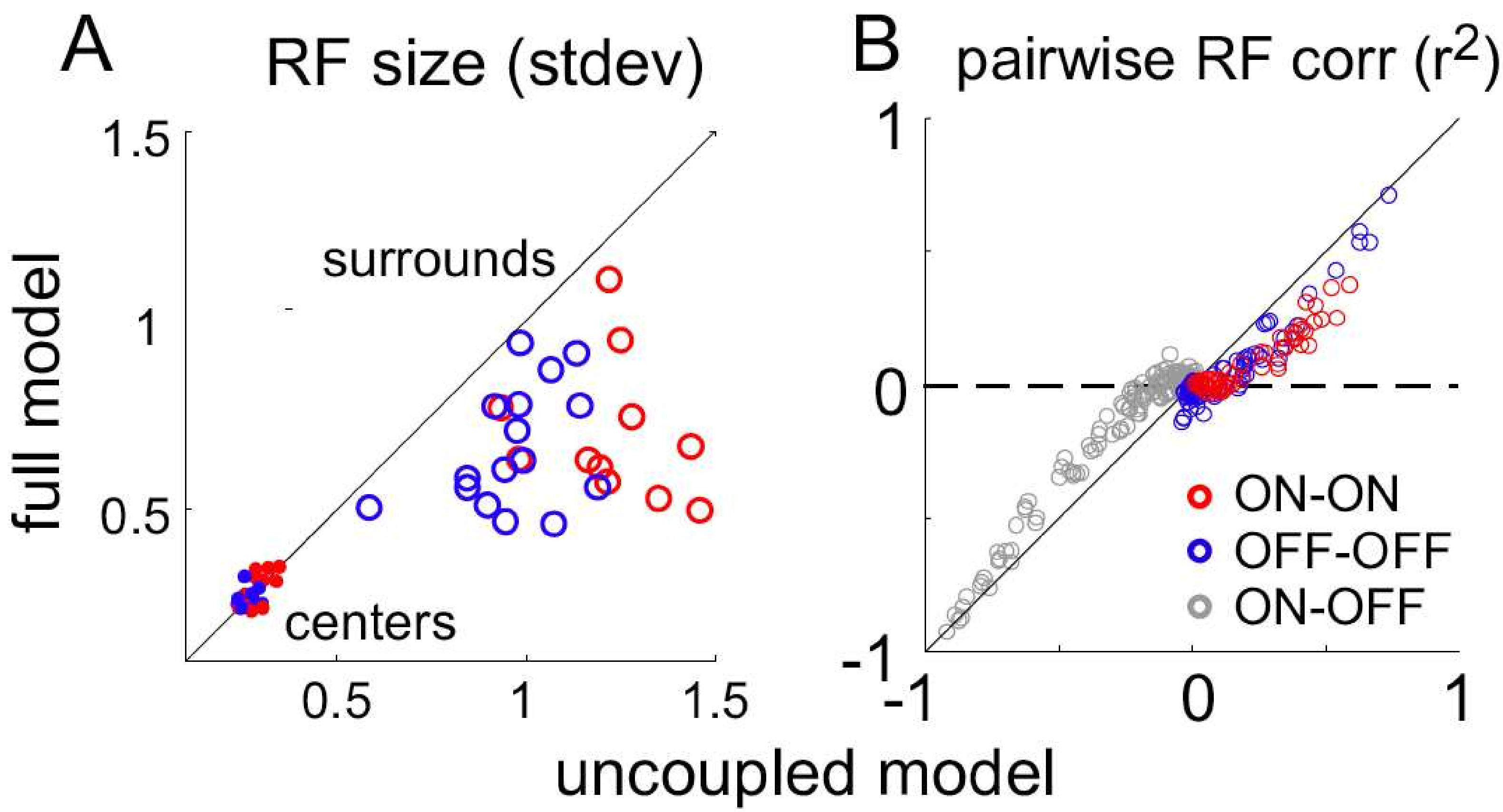
ON
cells
OFF
cells
stimulus filter
temporal
spatial
center
surround
-200
-100
0
time (ms)
post-spike filter
gain
1
0
0
20
40
time (ms)


ON
cells
OFF
cells
coupling filters
gain
3
2
1
0
20
40
from ON
time (ms)
1
0.8
0.6
0
20
40
from OFF
time (ms)

Nearest-neighbor connectivity

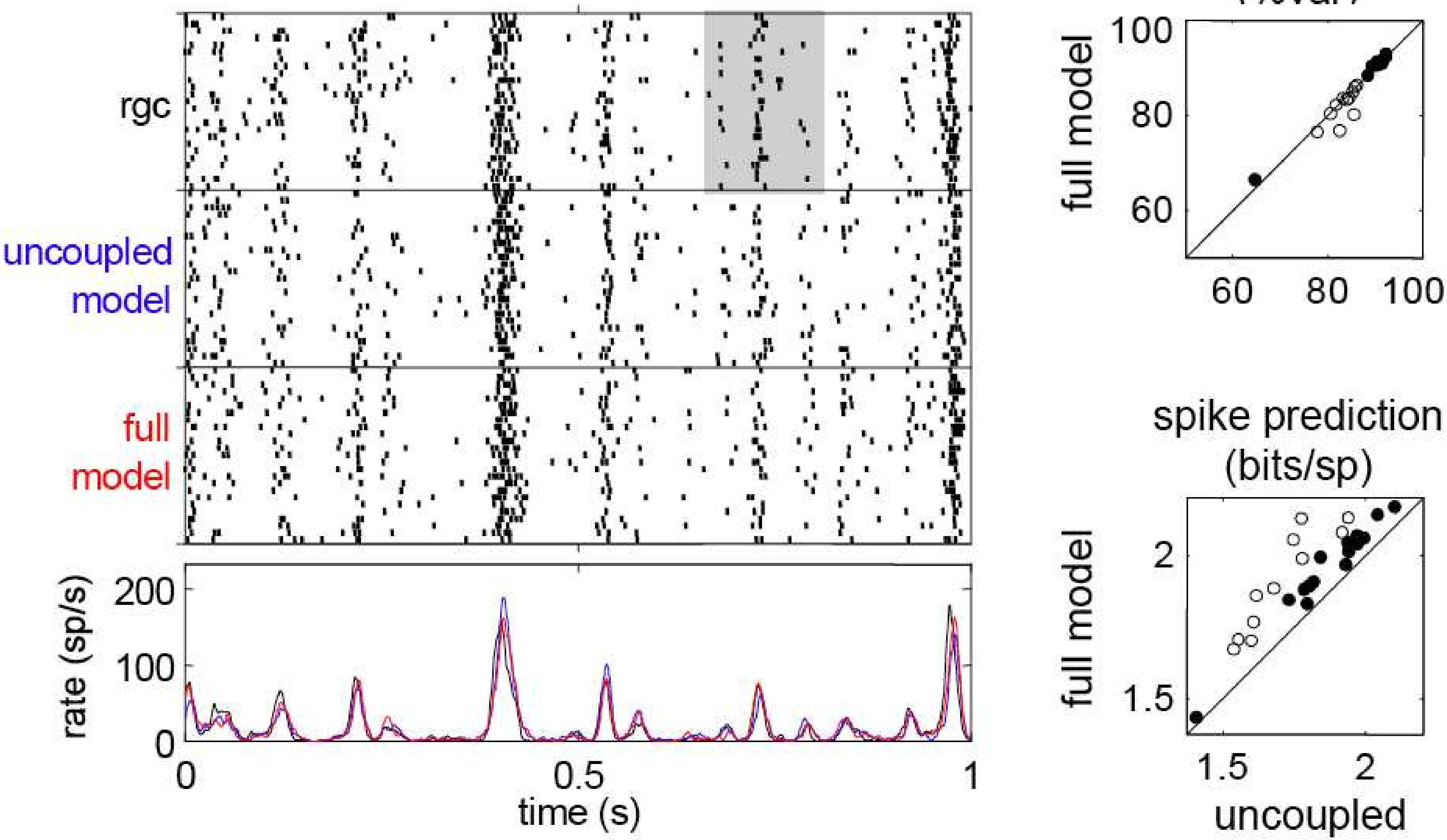


Fitting coupling terms exposes smaller receptive fields

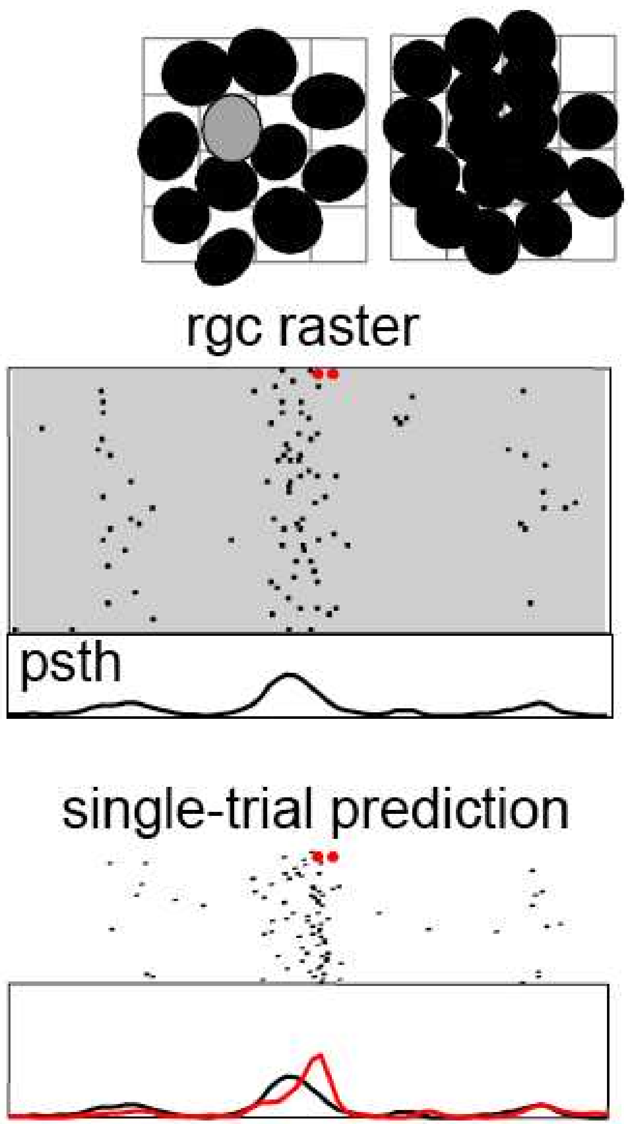


Spike Train Prediction

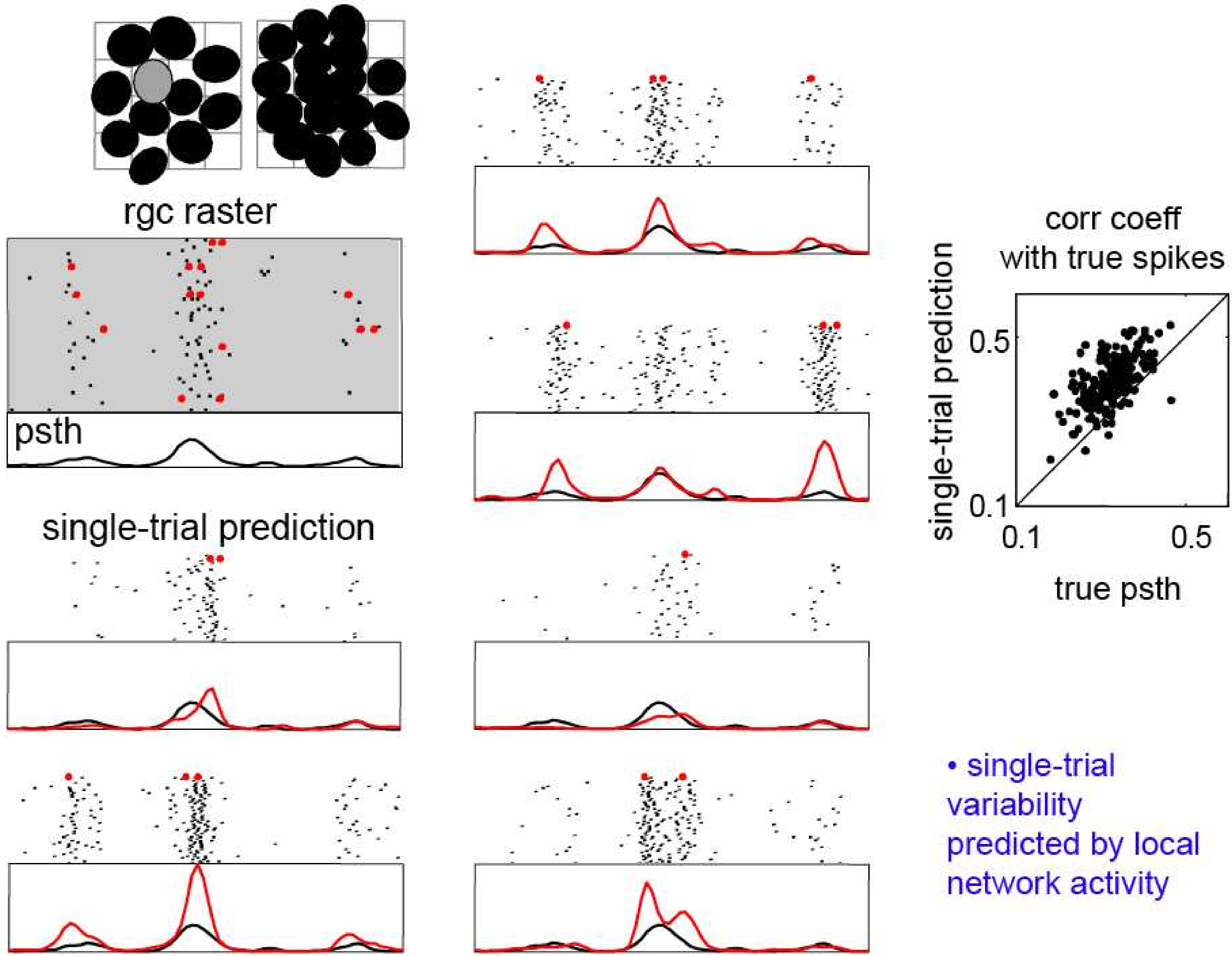
- improved prediction, but not of mean rate!



Network predictability analysis

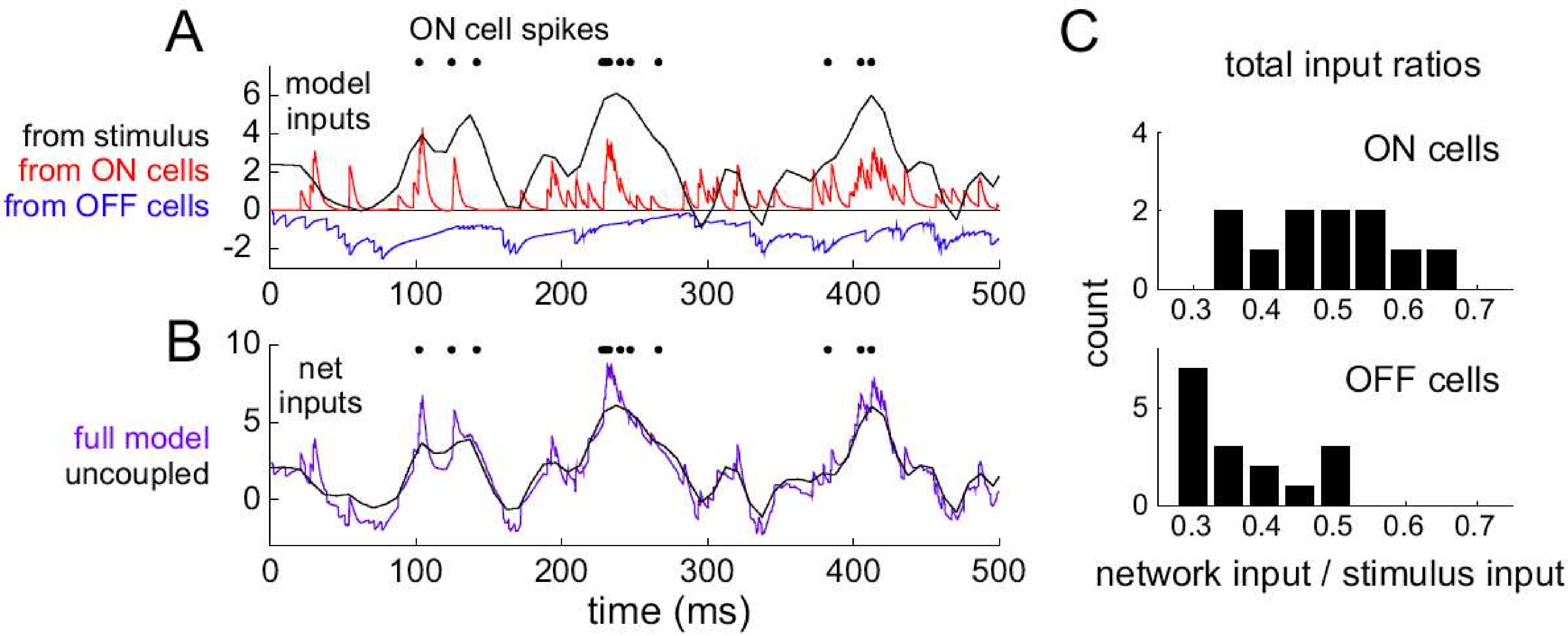


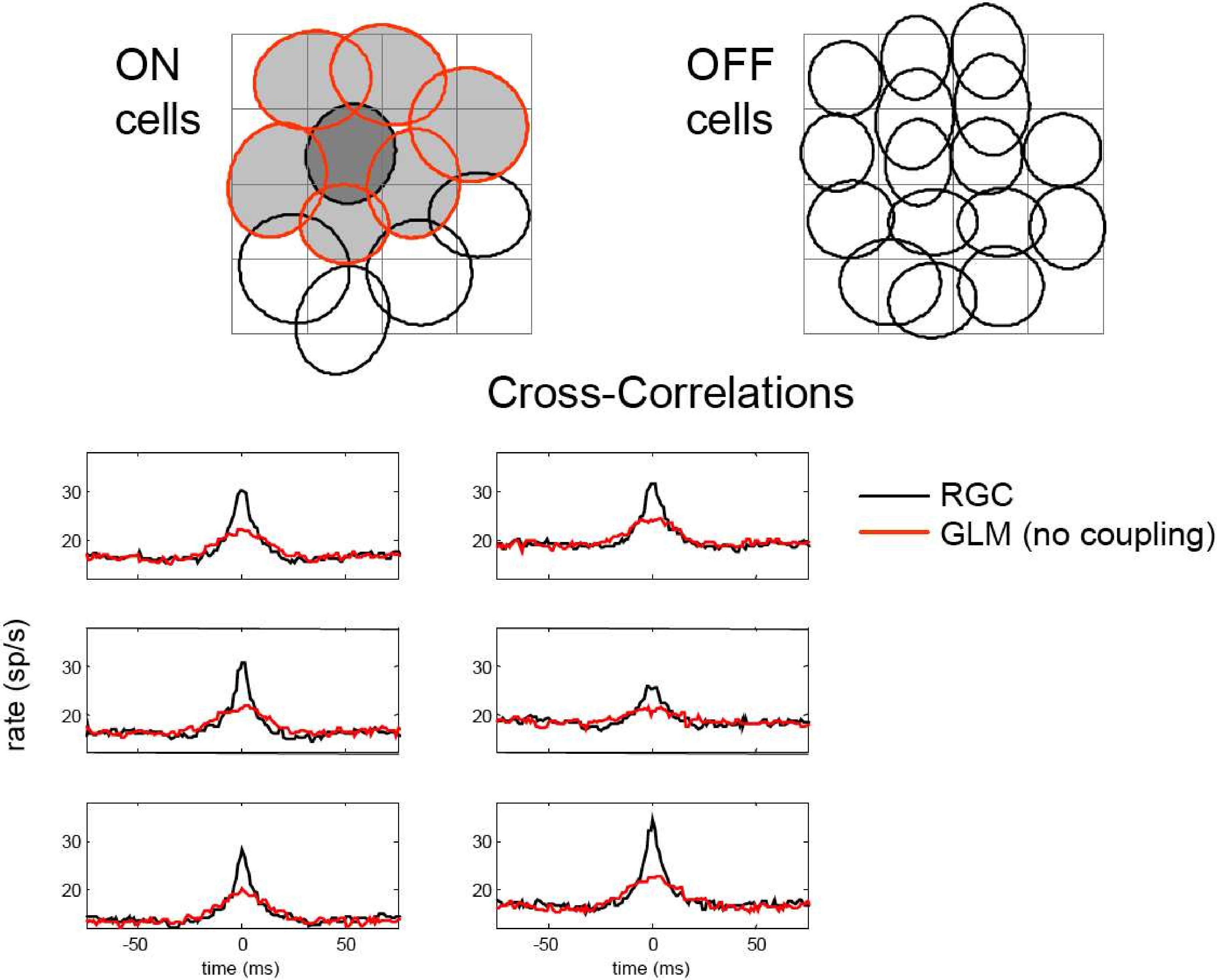
- fix all other neurons for a single trial
- draw single-trial predictions of this cell's spike train

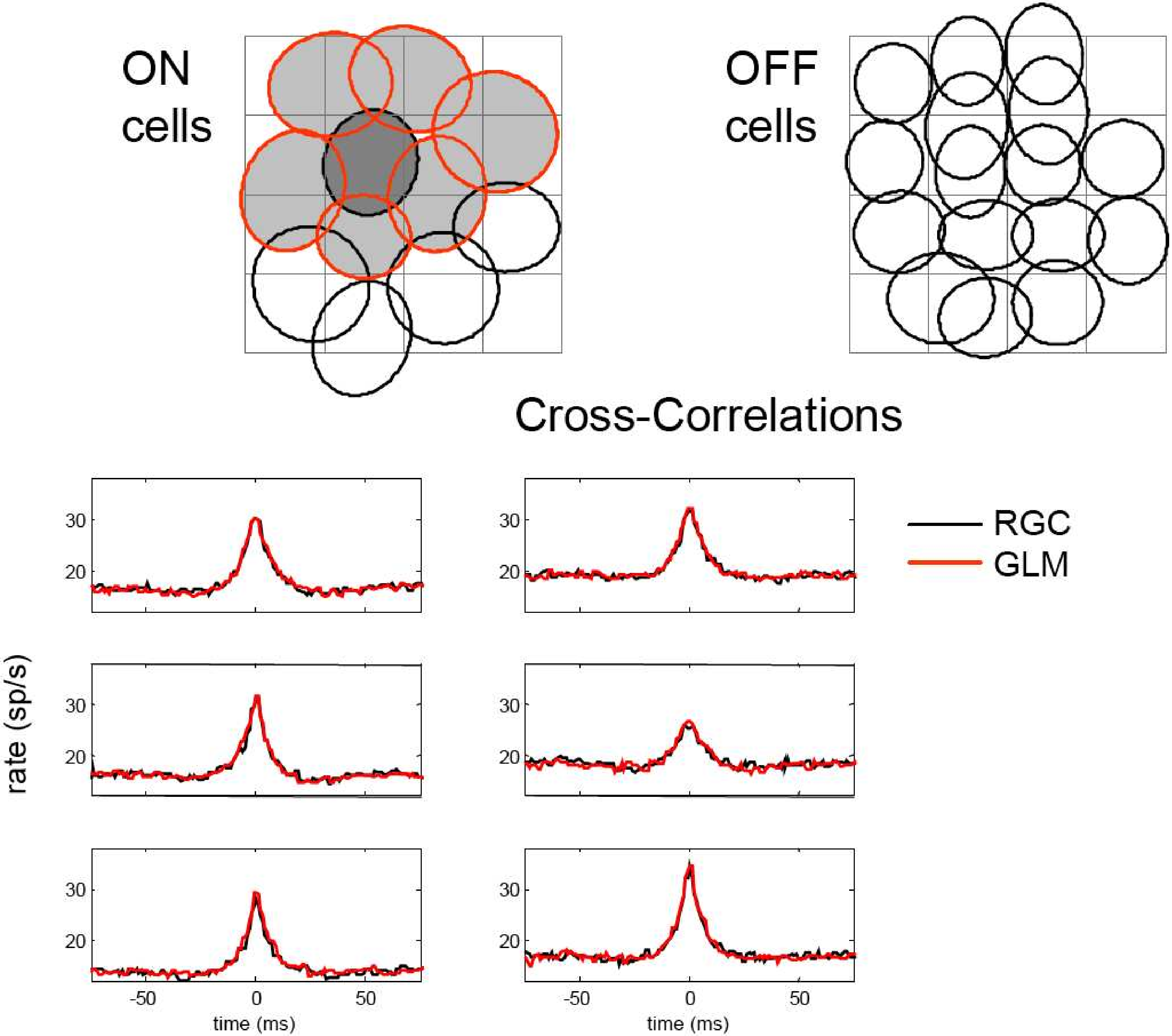


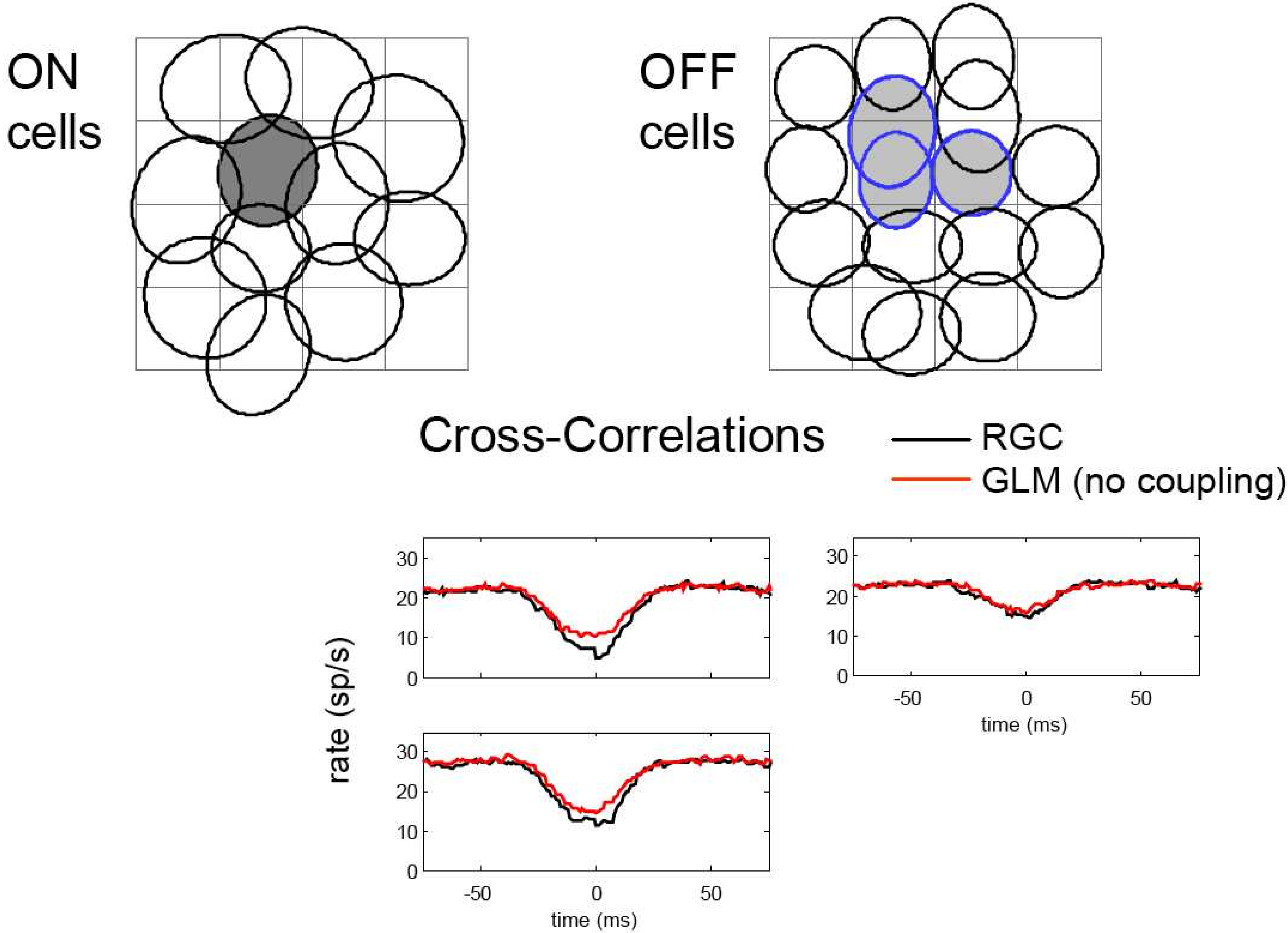
rgc raster
psth
single-trial prediction
corr coeff
with true spikes
single-trial prediction
0.5
0.1
0.1
0.5
true psth
• single-trial
variability
predicted by local
network activity

Network vs. stimulus drive

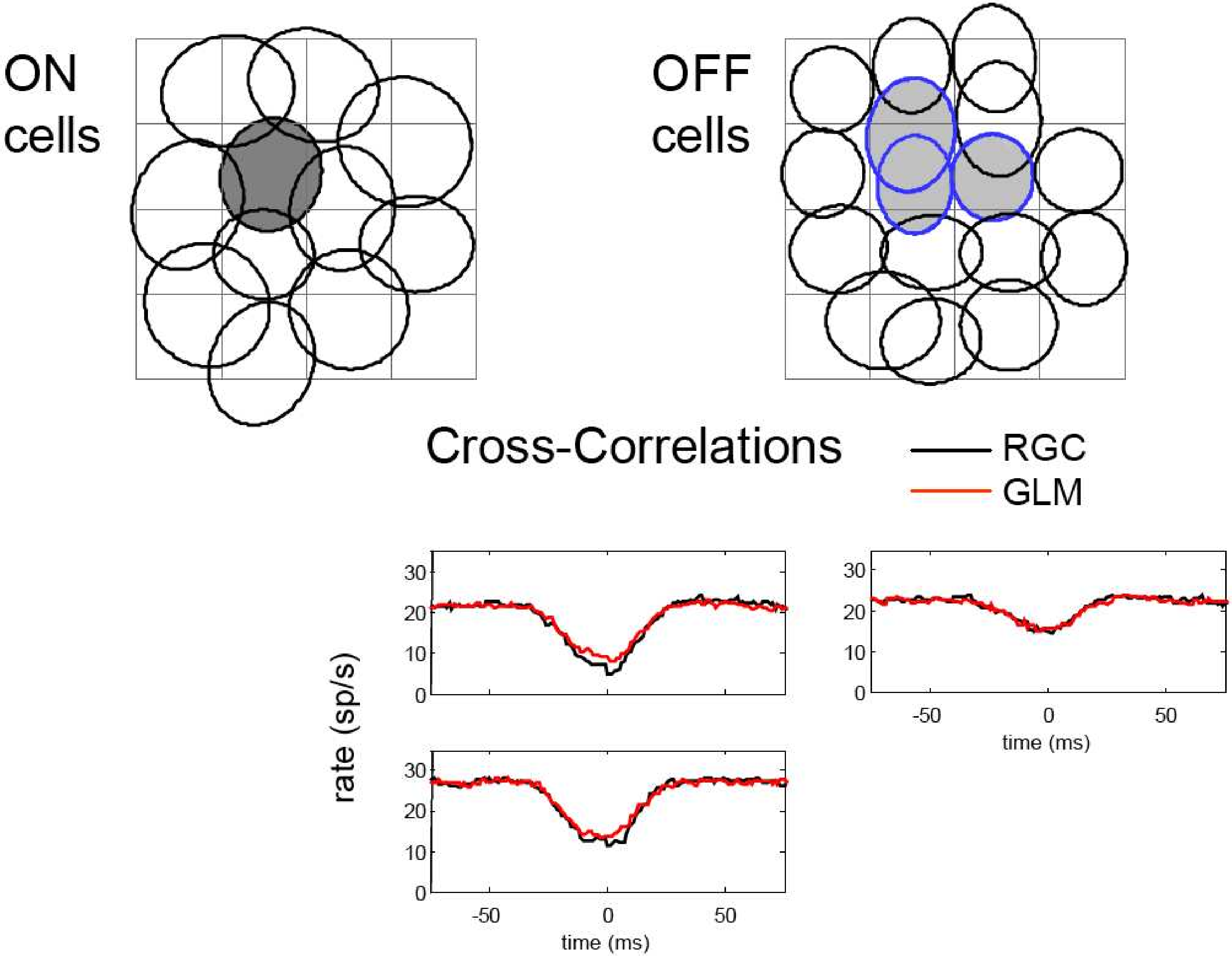



ON
cells
OFF
cells
Cross-Correlations
RGC
GLM (no coupling)
rate (sp/s)
30
20
-50
0
50
time (ms)

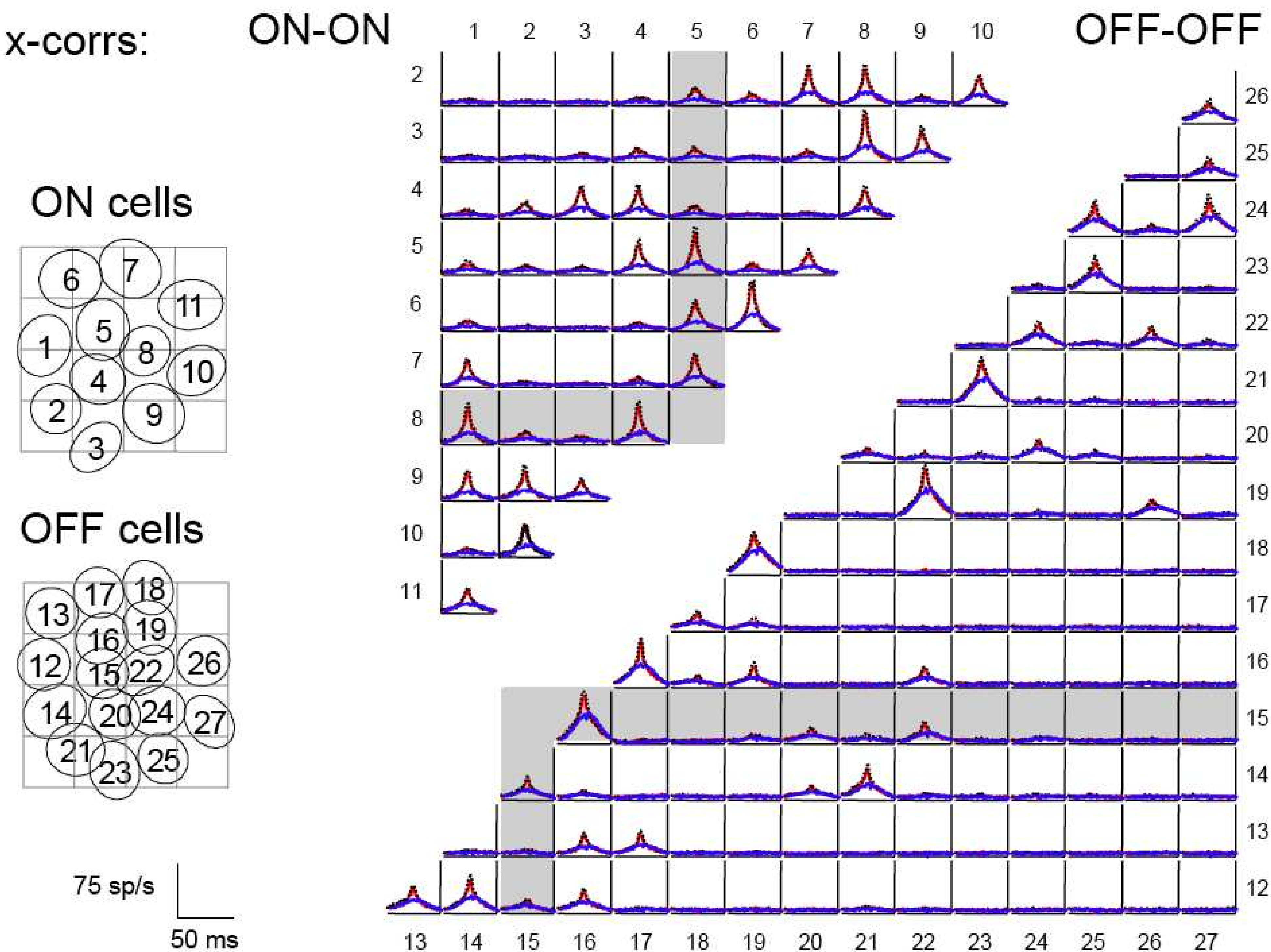

ON cells
OFF cells
Cross-Correlations
RGC
GLM
rate (sp/s)
30
20
-50
0
50
time (ms)



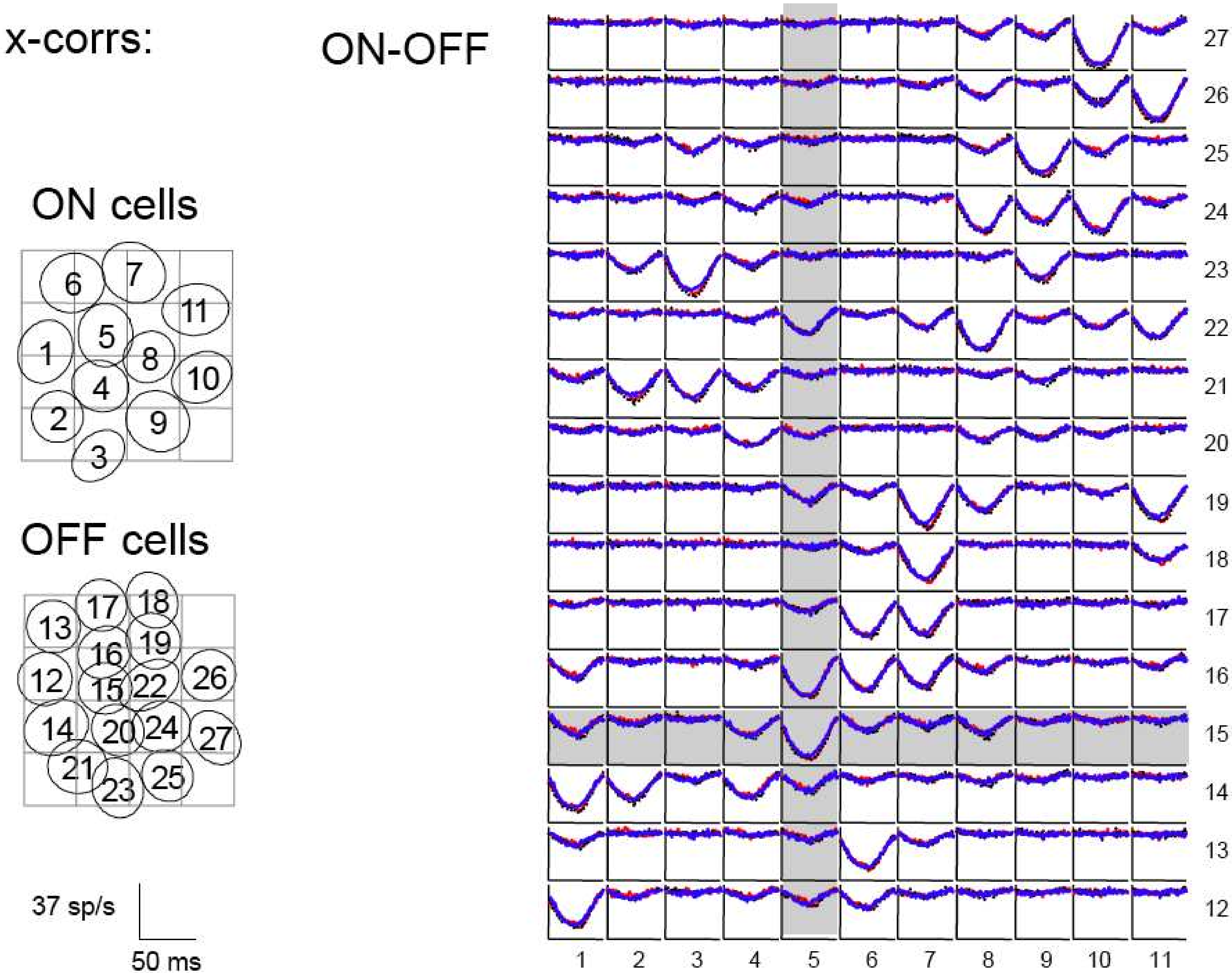
ON
cells
OFF
cells
Cross-Correlations
RGC
GLM (no coupling)
rate (sp/s)
30
20
10
0
-50
0
50
time (ms)



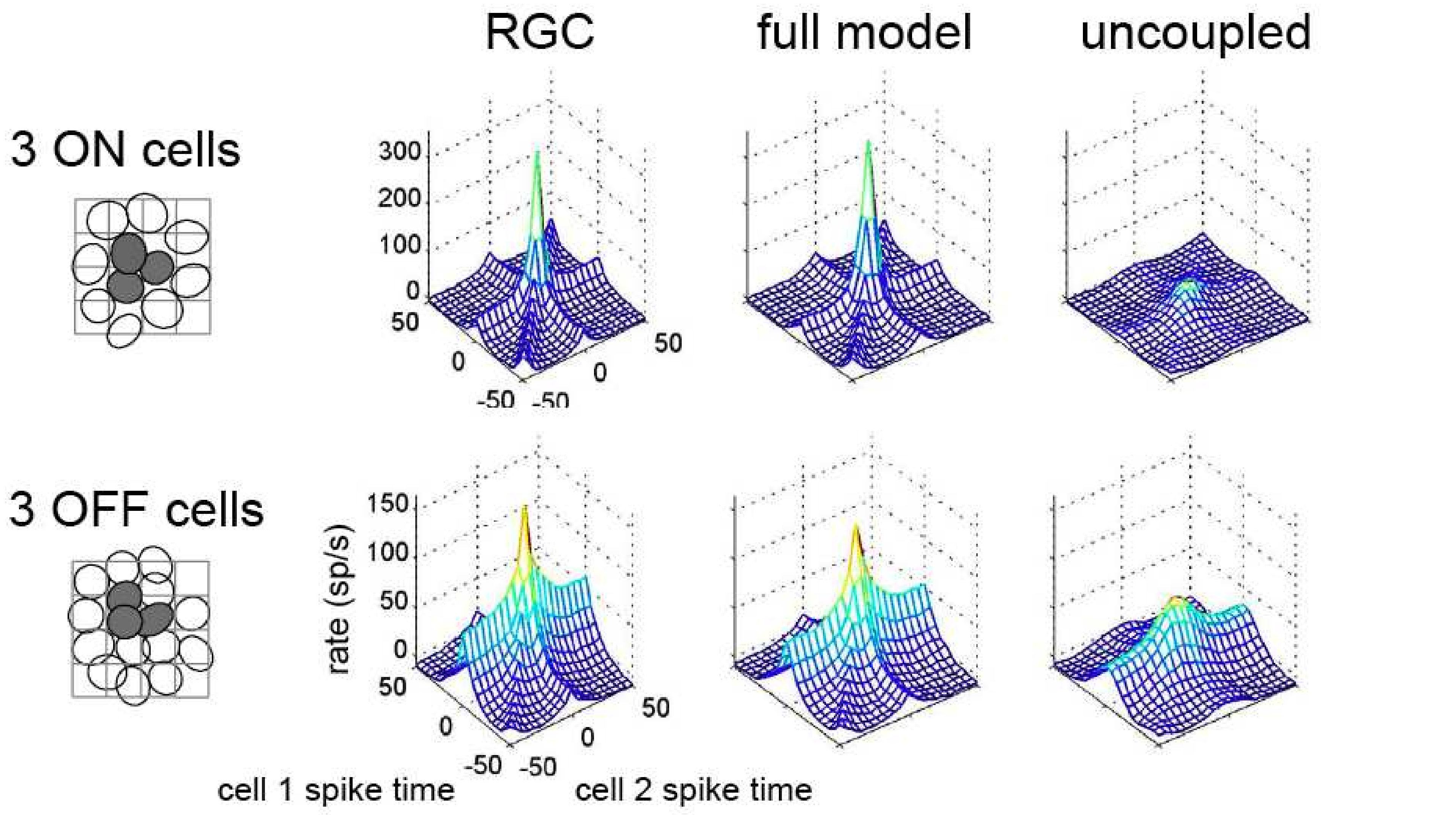
ON
cells
OFF
cells
Cross-Correlations
RGC
GLM
rate (sp/s)
30
20
10
0
-50
0
50
time (ms)



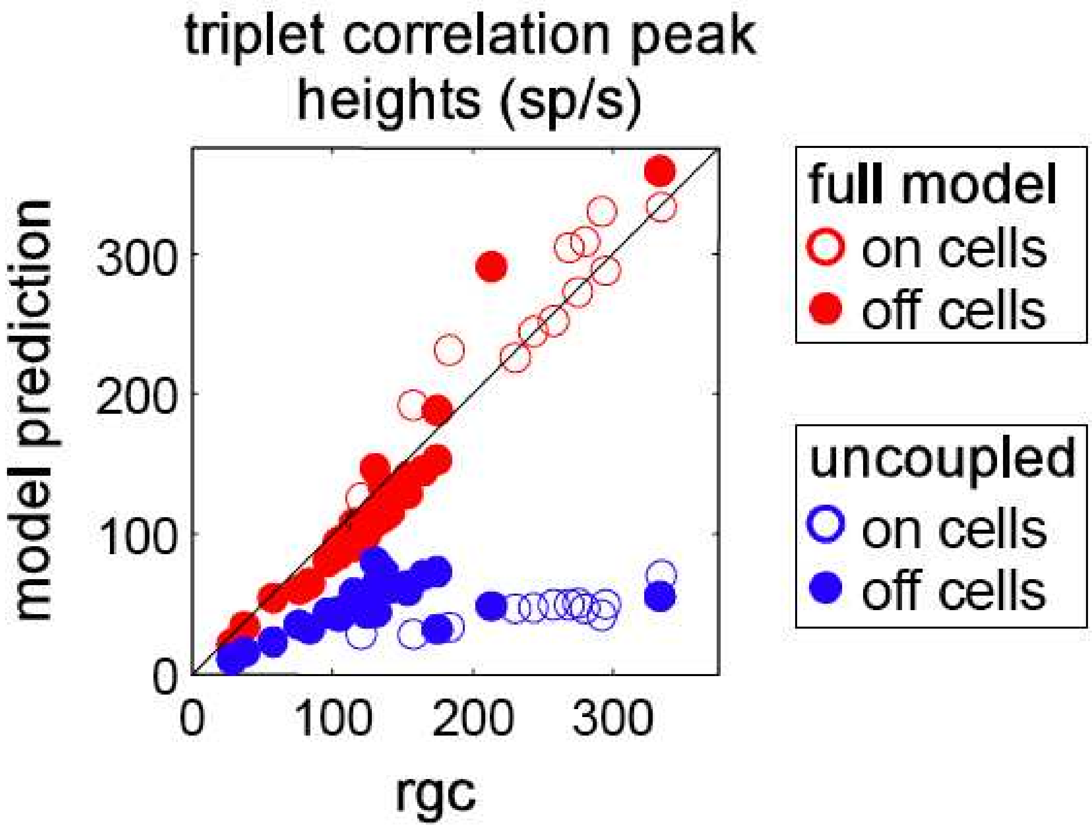
x-corrs:
ON-ON
OFF-OFF
ON cells
OFF cells
75 sp/s
50 ms


x-corrs:
ON-OFF
ON cells
OFF cells
37 sp/s
50 ms
1 2 3 4 5 6 7 8 9 10 11
27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12

Triplet correlations

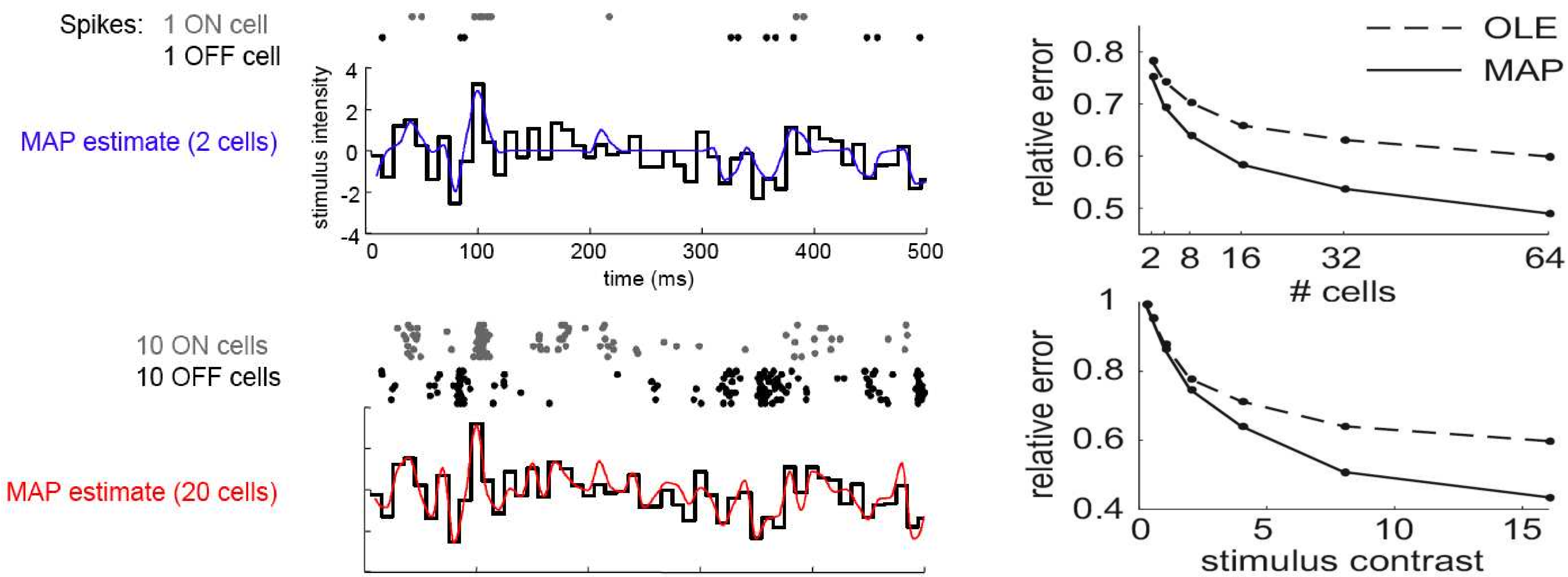


Triplet correlations



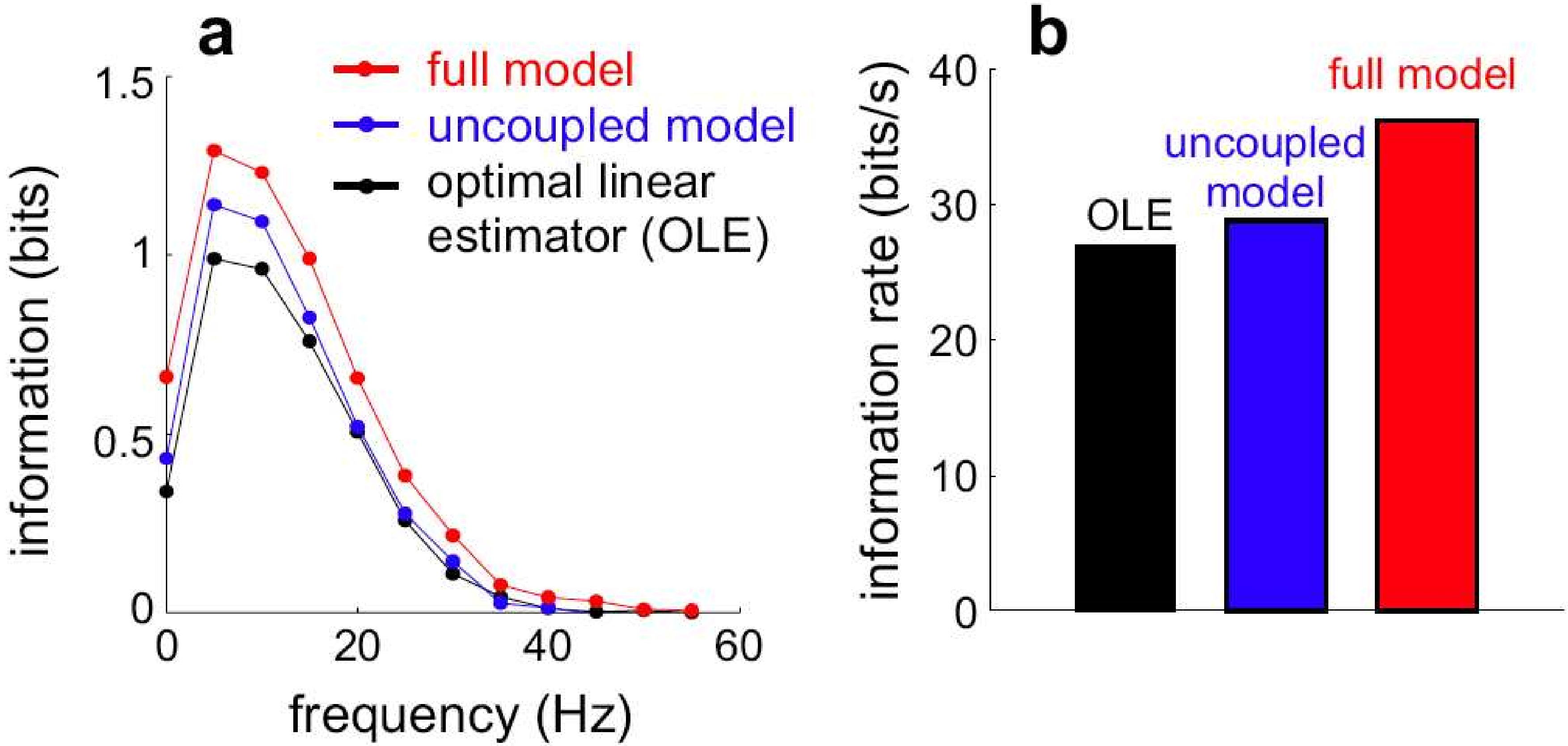
Model-based optimal decoding

— Use Bayes' rule to compute $\vec{x}_{MAP} =$
$\arg\max_{\vec{x}} \log P(\vec{x}|spikes) = \arg\max_{\vec{x}} \log P(spikes|\vec{x}) + \log P(\vec{x})$.

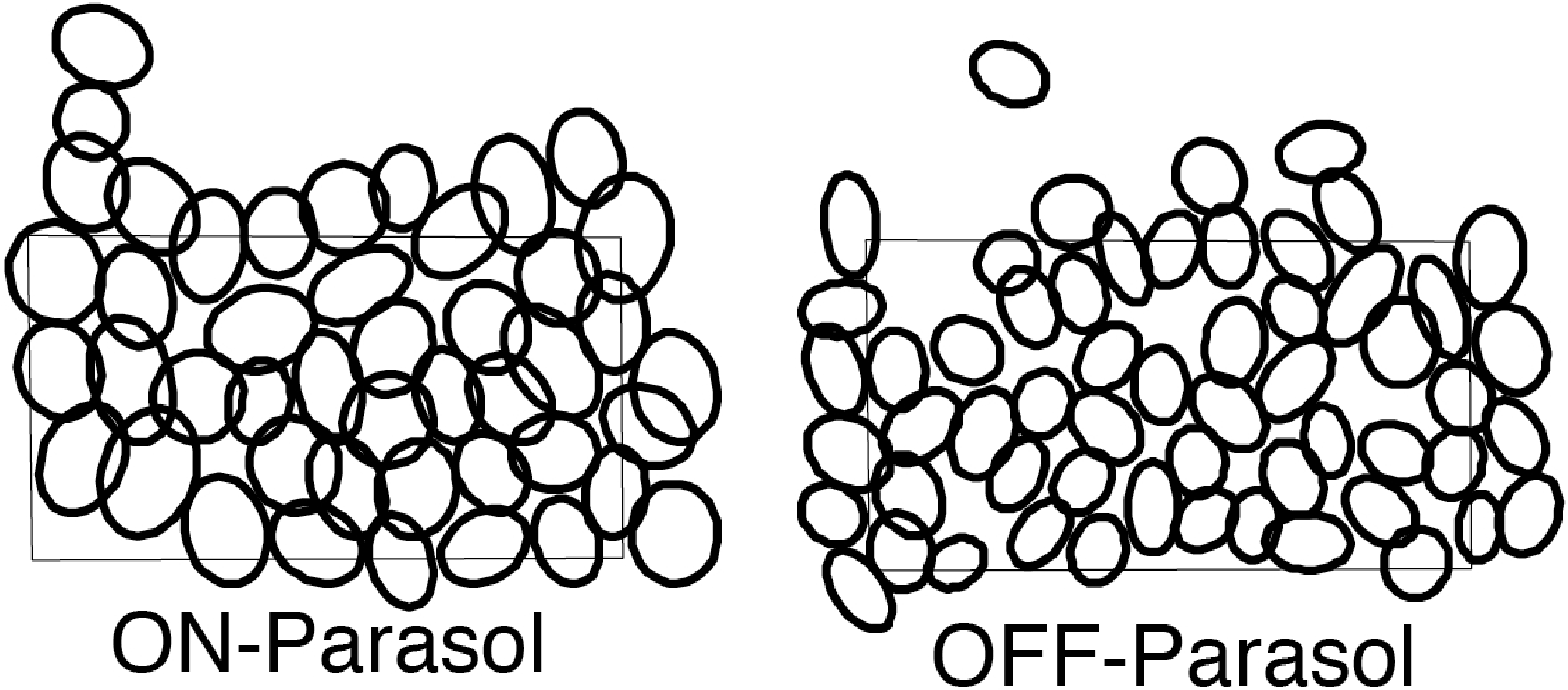


Again, correct model $P(spikes|\vec{x})$ is essential (Pillow and Paninski, 2007).

Coupled model decodes more accurately



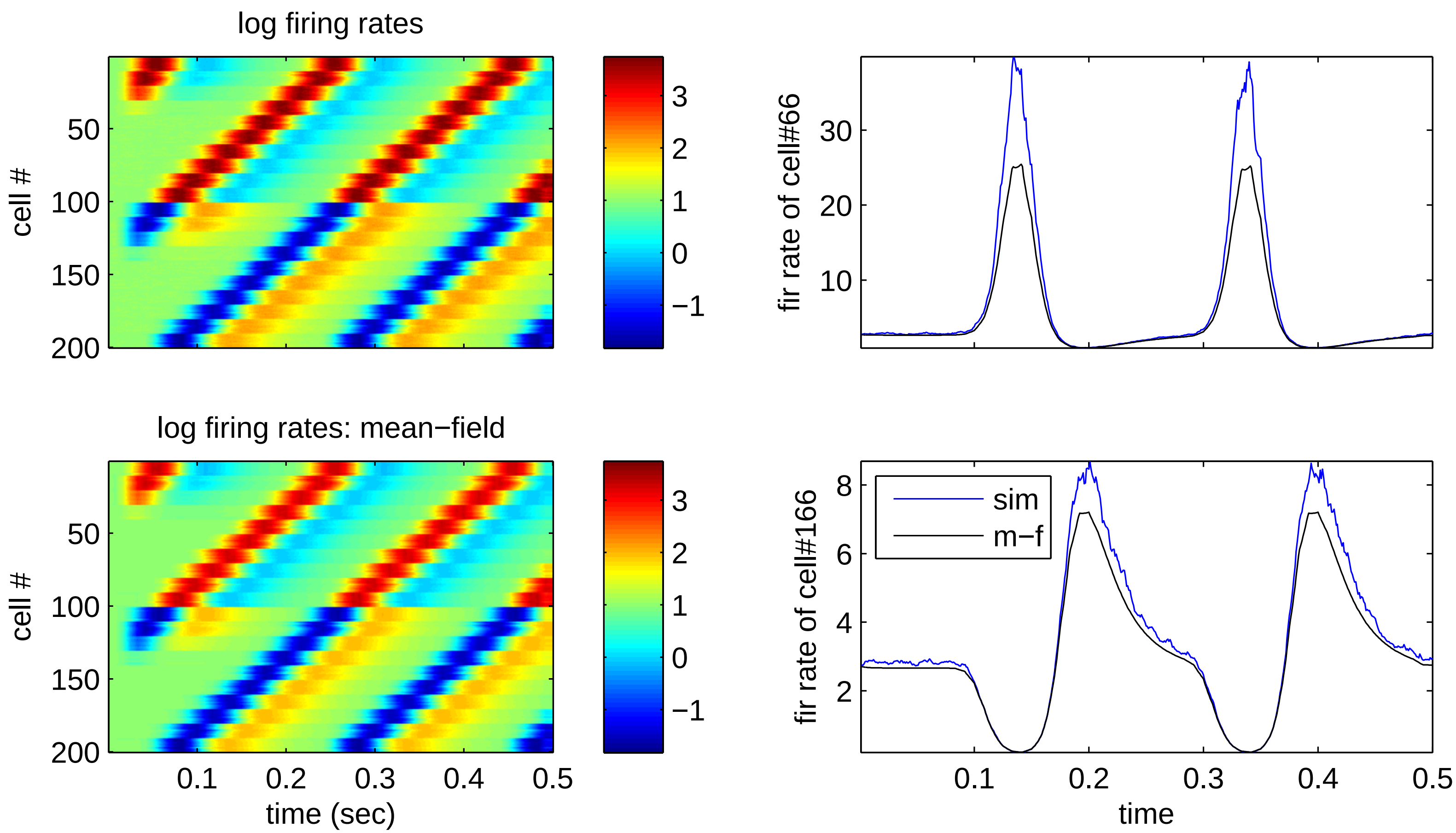
Next: Large-scale network modeling



— Do observed local connectivity rules lead to interesting network dynamics? What are the implications for retinal information processing?

Mean-field model

$$\begin{aligned}\lambda_i(t) &= f\left[\vec{k}_i^T \vec{x}(t) + b_i + \sum_{i',j} h_{i',i}(t - t_{i',j})\right] \\ &\approx f\left[\vec{k}_i^T \vec{x}(t) + b_i + \sum_{i'} h_{i',i}(t) * \lambda_{i'}(t)\right]\end{aligned}$$



References

Chander, D. and Chichilnisky, E. (2001). Adaptation to temporal contrast in primate and salamander retina. *Journal of Neuroscience*, 21:9904–16.

Fohlmeister, J. and Miller, R. (1997). Mechanisms by which cell geometry controls repetitive impulse firing in retinal ganglion cells. *Journal of Neurophysiology*, 78:1948–1964.

Foldiak, P. (2001). Stimulus optimisation in primary visual cortex. *Neurocomputing*, 38–40:1217–1222.

Gerstner, W. and Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.

Lewi, J., Butera, R., and Paninski, L. (2006). Real-time adaptive information-theoretic optimization of neurophysiological experiments. *NIPS*.

Machens, C. (2002). Adaptive sampling by information maximization. *Physical Review Letters*, 88:228104–228107.

Nemenman, I., Shafee, F., and Bialek, W. (2002). Entropy and inference, revisited. *NIPS*, 14.

Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Computation*, 15:1191–1253.

Paninski, L. (2005). Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17:1480–1507.

Paninski, L., Pillow, J., and Simoncelli, E. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural model. *Neural Computation*, 16:2533–2561.

Pillow, J. and Paninski, L. (2007). Model-based decoding, information estimation, and change-point detection in multi-neuron spike trains. *Submitted*.

Pillow, J., Paninski, L., Shlens, J., Simoncelli, E., and Chichilnisky, E. (2005a). Modeling multi-neuronal responses in primate retinal ganglion cells. *Comp. Sys. Neur. '05*.

Pillow, J., Paninski, L., Uzzell, V., Simoncelli, E., and Chichilnisky, E. (2005b). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25:11003–11013.

Prekopa, A. (1973). On logarithmic concave measures and functions. *Acad Sci. Math.*, 34:335–343.

Simoncelli, E., Paninski, L., Pillow, J., and Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. In *The Cognitive Neurosciences*. MIT Press, 3rd edition.