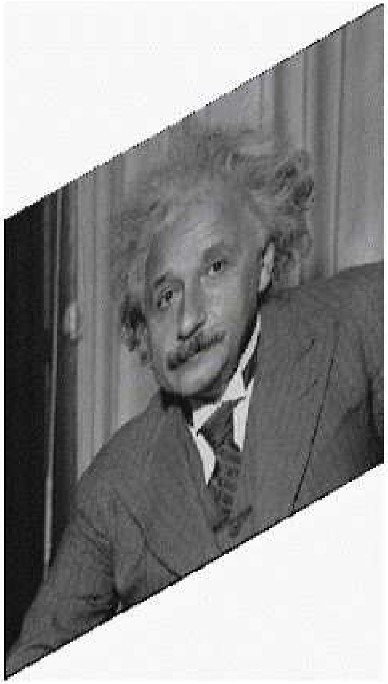
Statistical methods for understanding neural codes

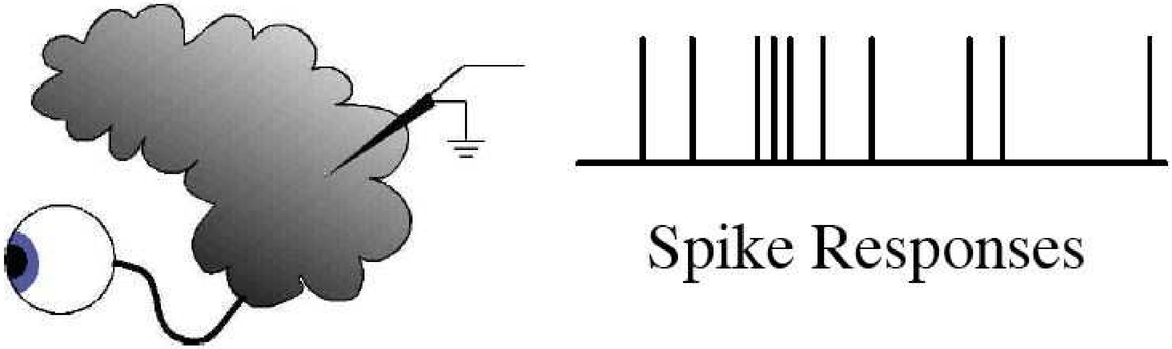
Liam Paninski

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

December 9, 2005

The neural code



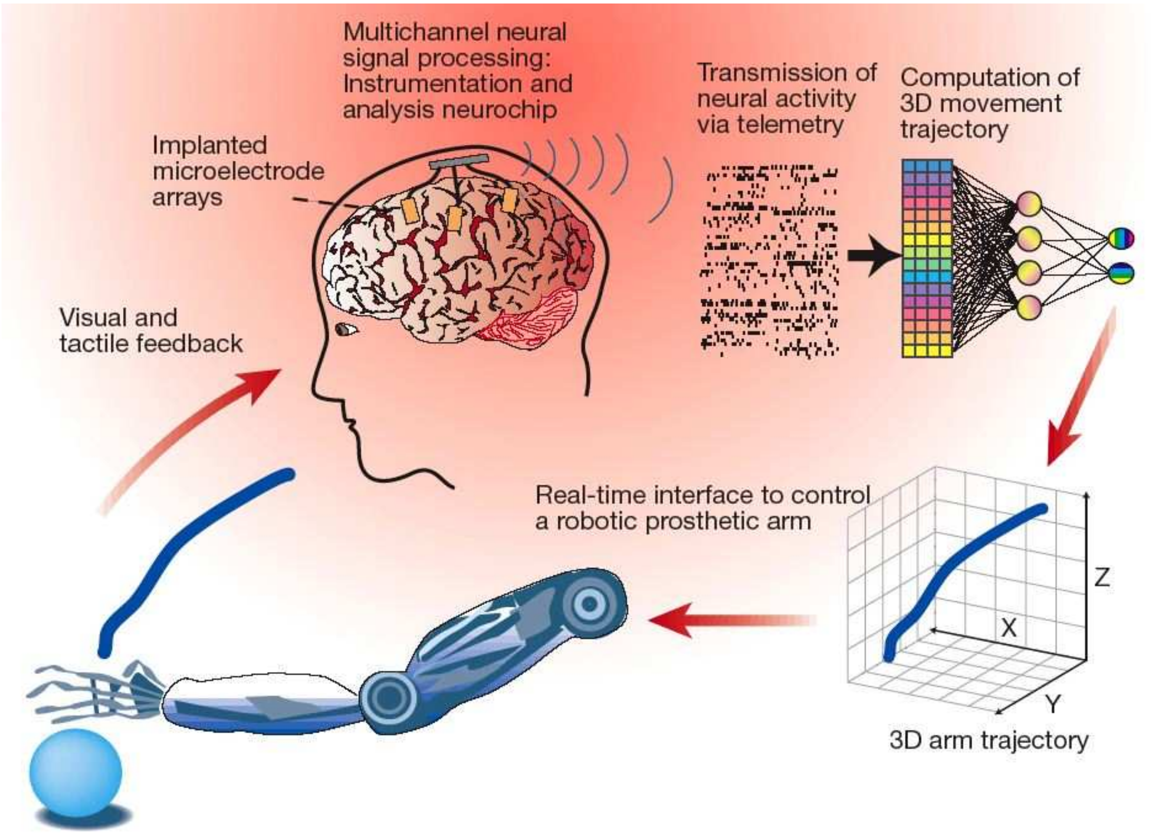


Input-output relationship between

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

Probabilistic formulation: $p(y|x)$

Example: neural prosthetic design



Nicolelis, Nature '01

Donoghue; Cyberkinetics, Inc. '04

(Paninski et al., 1999; Serruya et al., 2002; Shoham et al., 2005)

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, 2003b)

2: Select stimuli as efficiently as possible

e.g., (Foldiak, 2001; Machens, 2002; Paninski, 2003a)

3: Fit a model with small number of parameters

Part 1: Neural encoding models

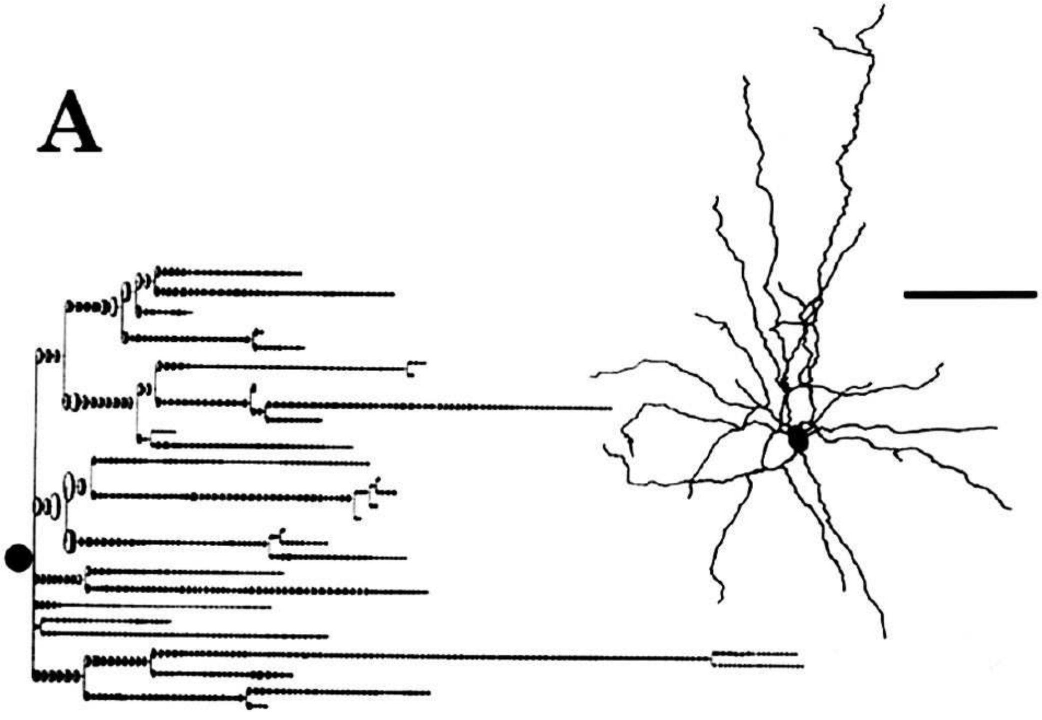
“Encoding model”: $p_\theta(y|x)$.

— 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



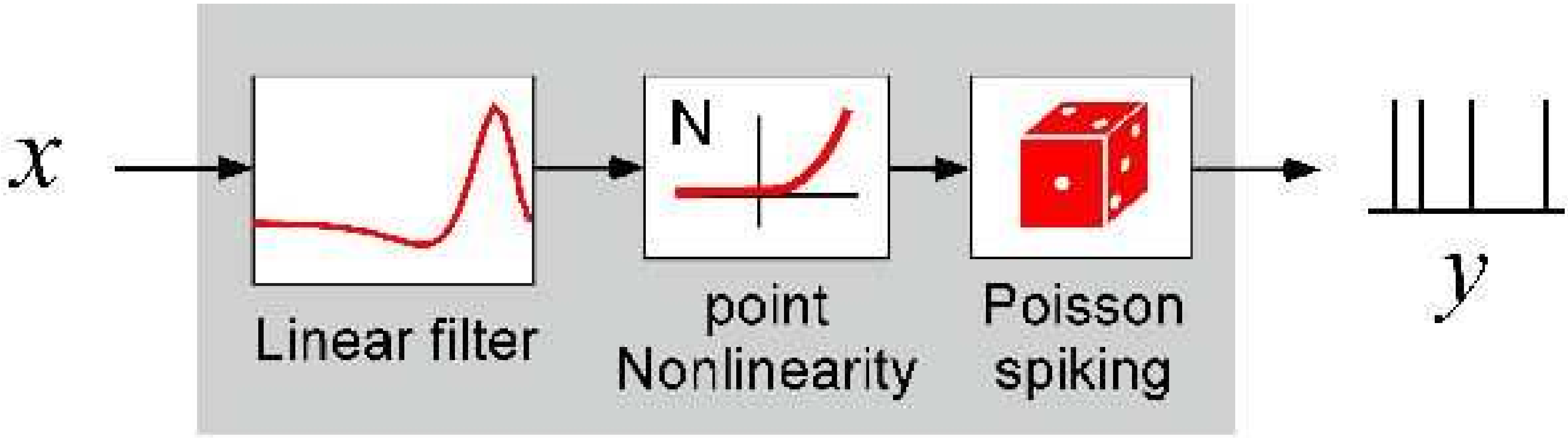
		Regional Conductances (mS/cm^2)				
Model	Current	Dendrites	Soma	AH	NR	Axon
EC2.5 REAL	I_{Ca}	2.0	1.5	1.5	—	—
$j = 1$	$I_{K,Ca}$	0.001	0.065	0.065	0.065	0.065
SD* (real) = 21.9 μm	I_{Na}	25	80	100–150†	100	40–70‡
SD (EC2.5) = 20 μm	I_K	12	18	18	18	12–18‡
$\tau_{Ca} = 1.5$	I_A	36	54	54	54	—
$E_L = -60$ 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: STA, STC (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

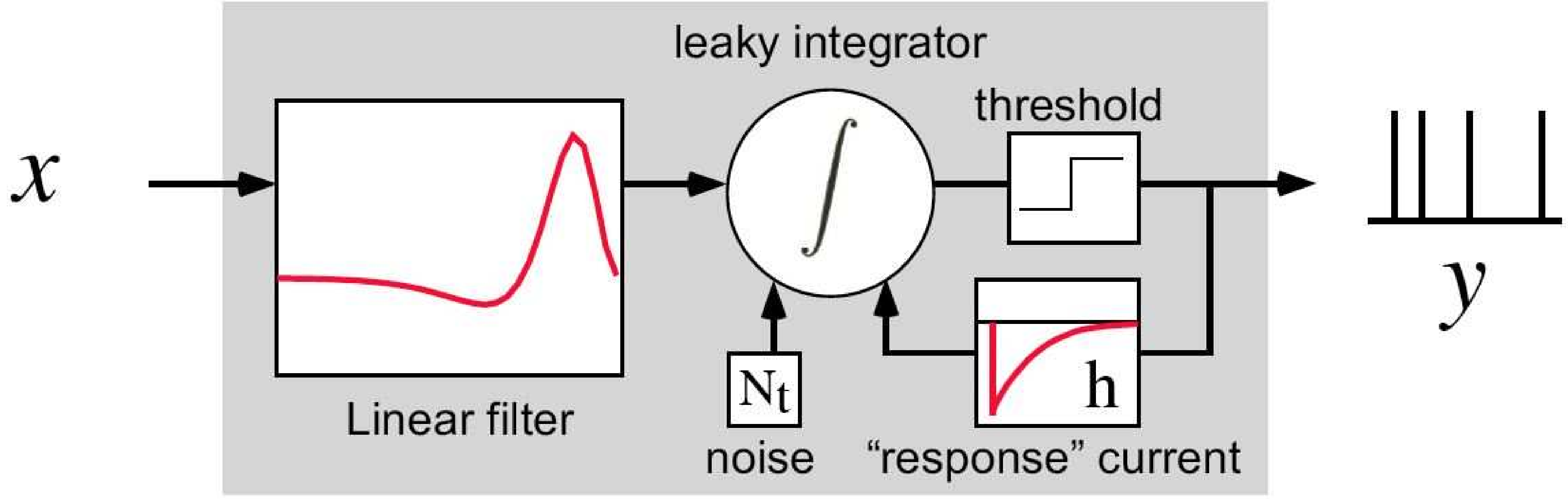
   — easy to incorporate prior knowledge, explicit noise models, etc.

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.

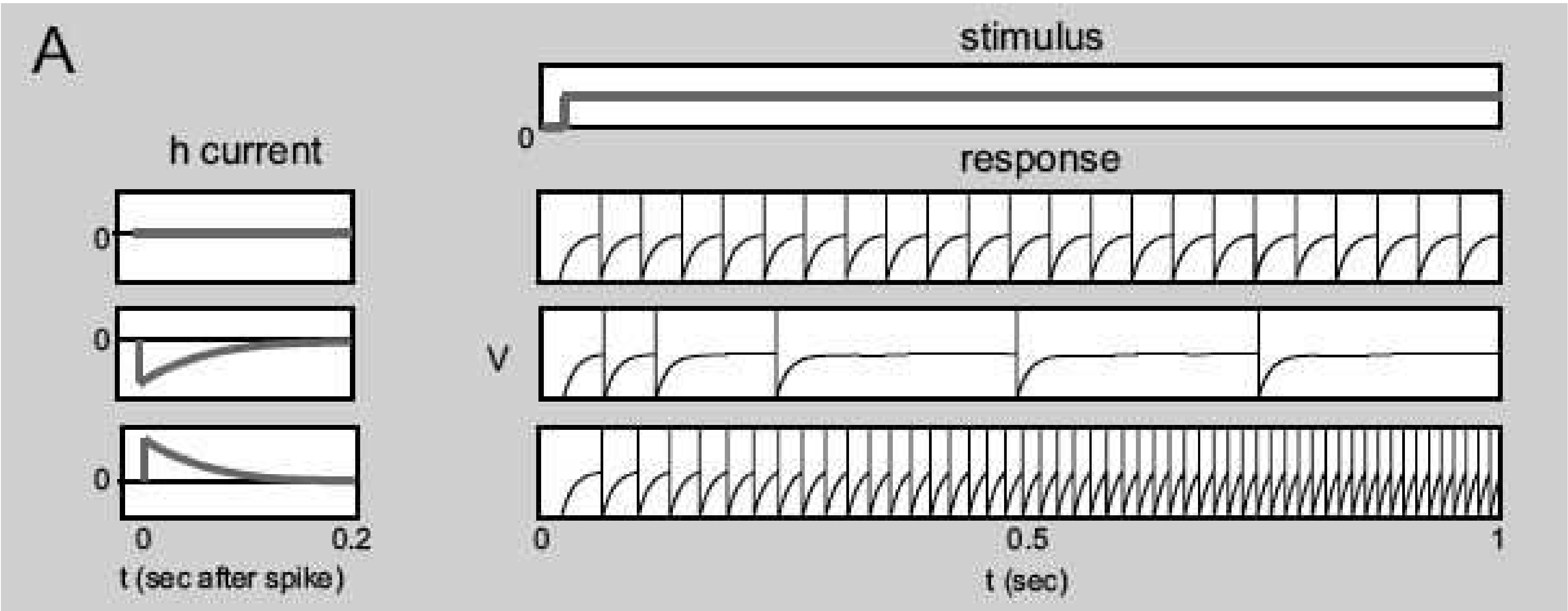
Filtered integrate-and-fire model



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

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

Model flexibility: Adaptation



The estimation problem

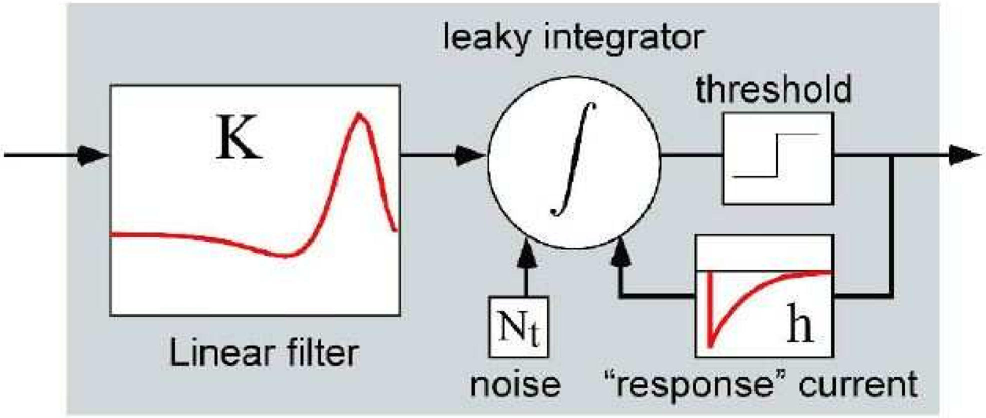
Learn the model parameters:

$\vec{K}$ = stimulus filter

g = leak conductance

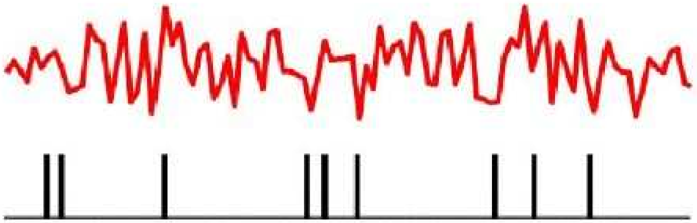
σ^2 = noise variance

$\vec{h}$ = response current



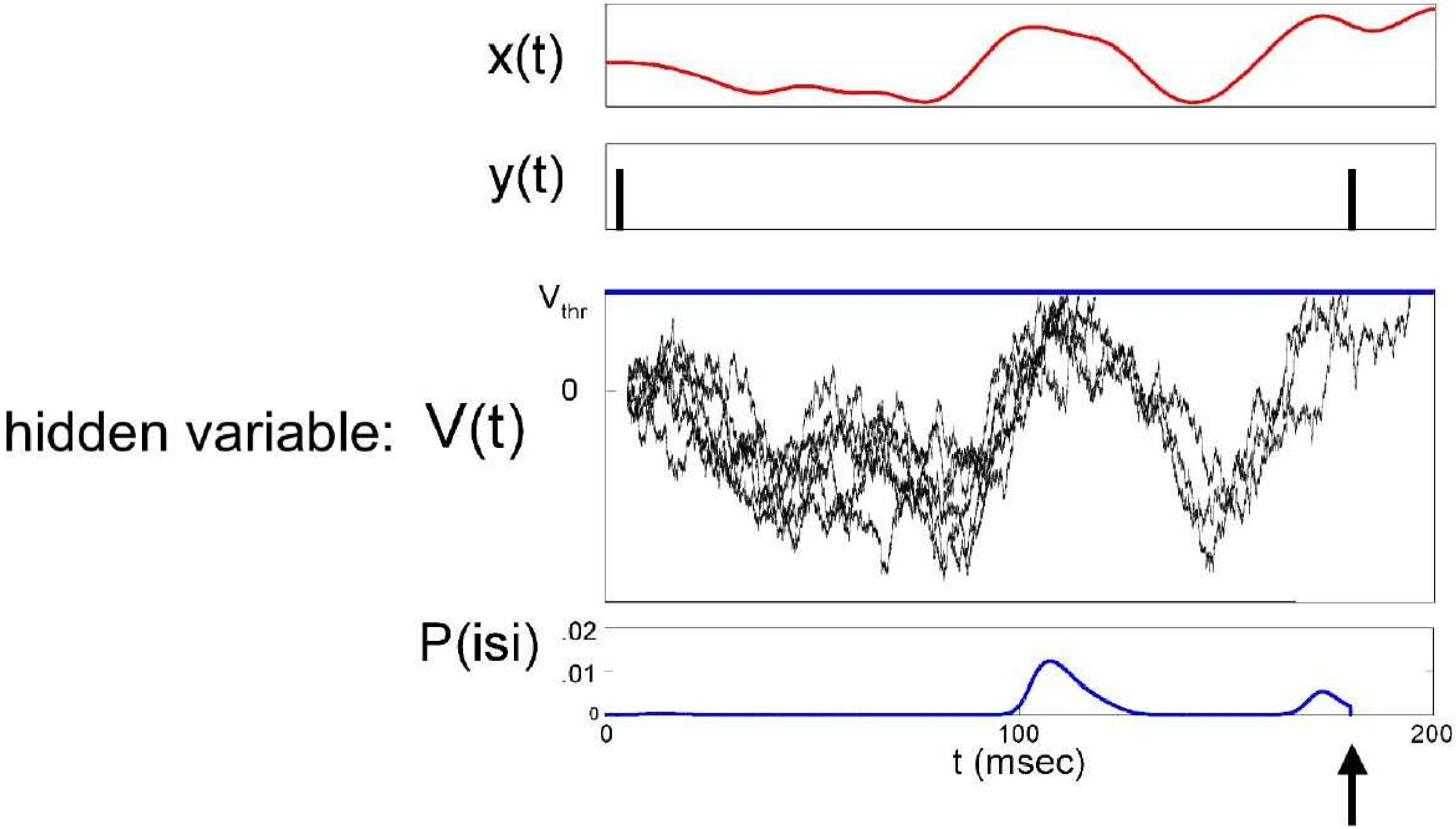
From: stimulus train $x(t)$

spike times t_i



(Paninski et al., 2004b)

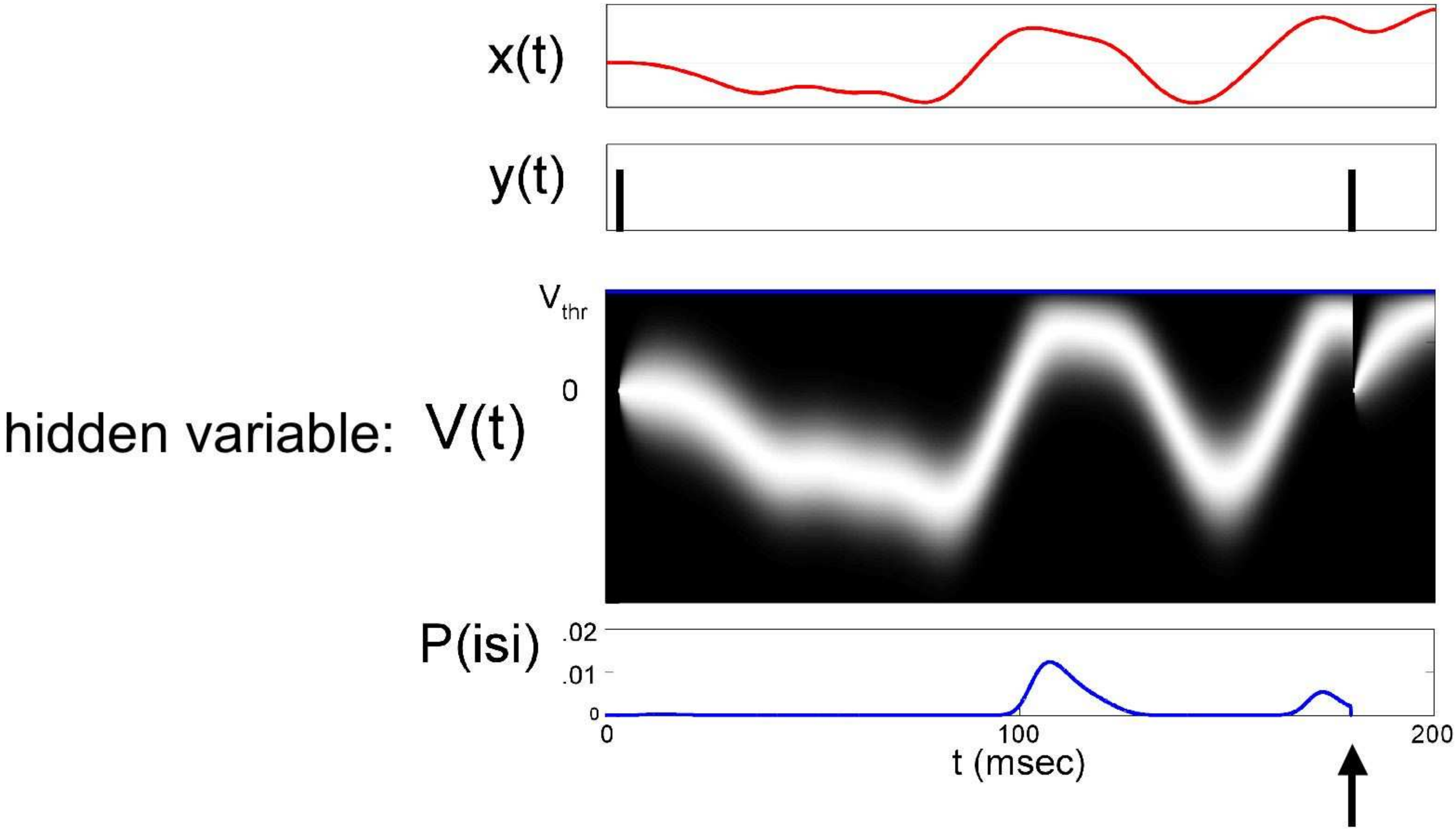
First passage time likelihood



$P(\text{spike at } t_i)$ = 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:

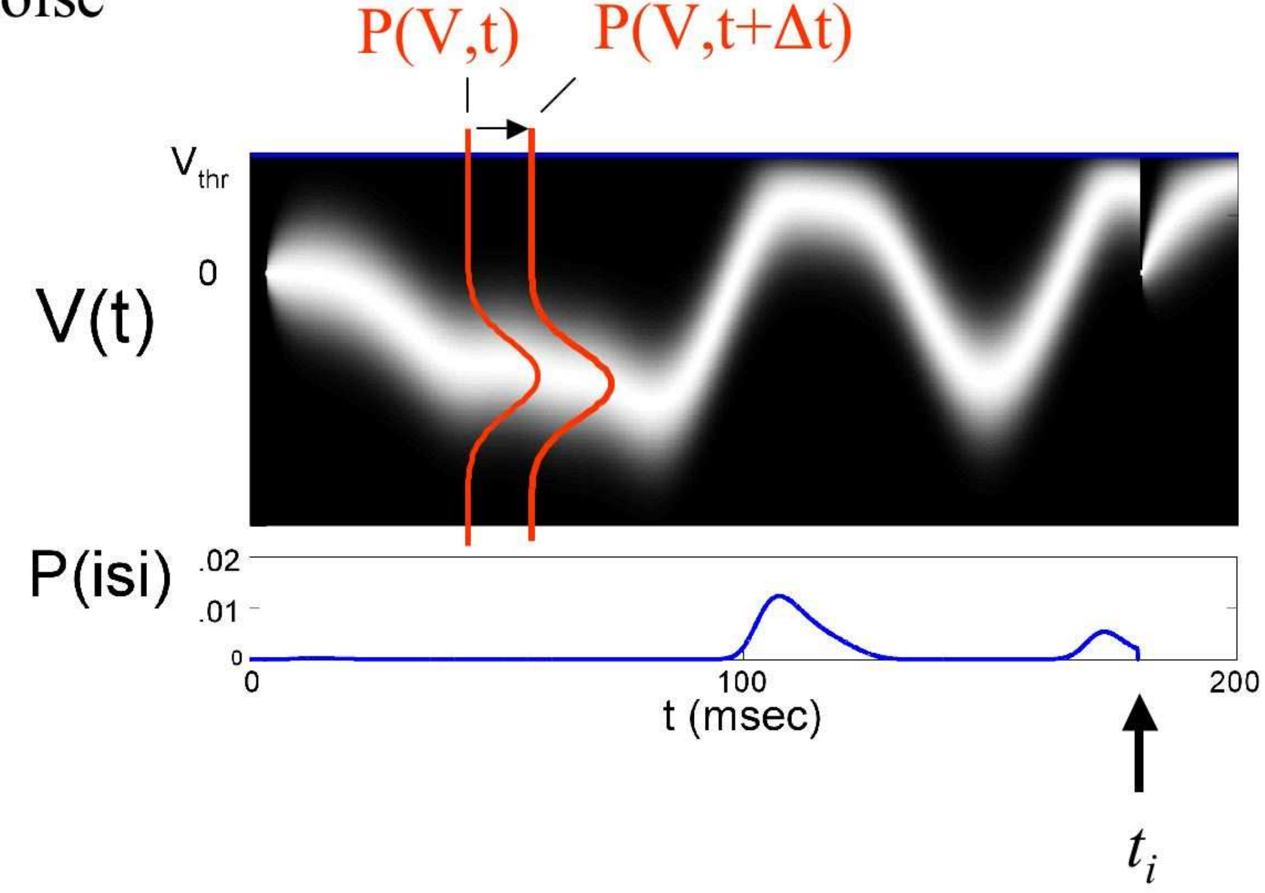
$$\frac{\partial P(V,t)}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial V^2} + g\frac{\partial[(V - V_0)P]}{\partial V},$$

- linear dynamics
- additive Gaussian noise

fast methods for solving linear PDE

$\Rightarrow$

efficient procedure for computing likelihood



Computing Likelihood

Diffusion Equation:

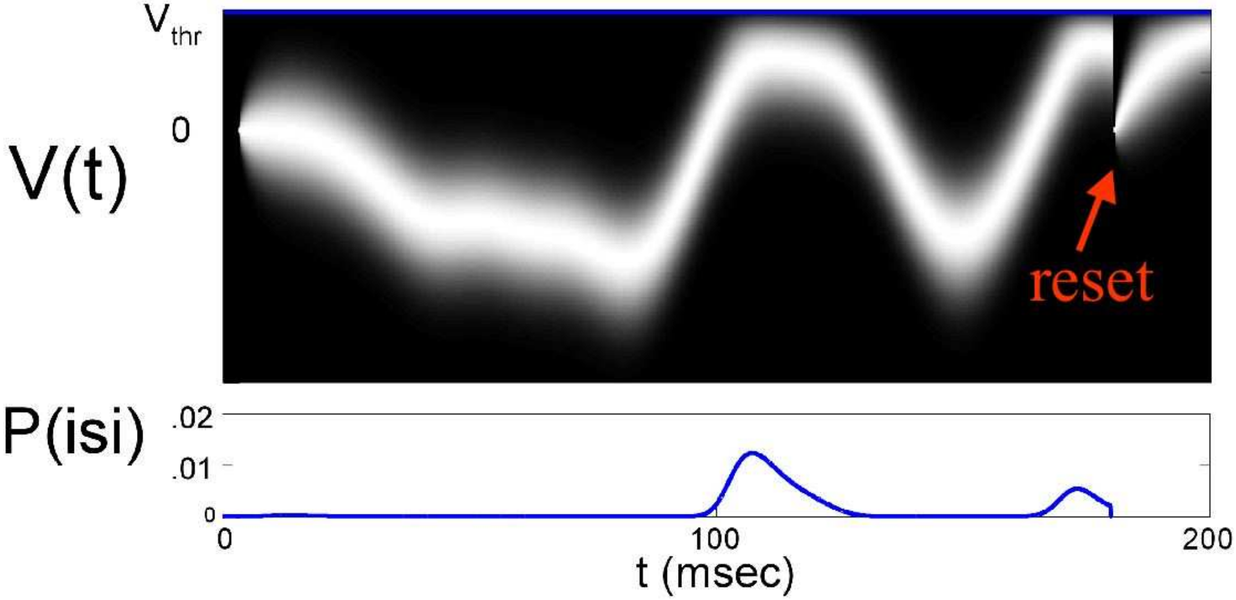
$$\frac{\partial P(V,t)}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial V^2} + g\frac{\partial[(V - V_0)P]}{\partial V},$$

- linear dynamics
- additive Gaussian noise

fast methods for solving linear PDE

$\Longrightarrow$

efficient procedure for computing likelihood



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

Integral equation method

Let $p(t) = P(\text{spike at time t})$.

Then $p(t)$ solves Volterra integral equation (Plesser and Tanaka, 1997) (via "method of images"; goes back to Schrodinger):

$$G_\theta(V_{th}, t|V_{th}, 0) = \int_0^t G_\theta(V_{th}, t|V_{th}, s)p(s)ds$$

$G_\theta(y, t|x, s)$ = probability of $V(t) = y$, given $V(s) = x$ (analytically computable)

Lower-triangular linear system: $O(d^2)$ (Paninski et al., 2005)

Can compute gradient $p(t)$ w.r.t. θ via matrix perturbation: efficient maximization

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.

$\implies$ no non-global local maxima

$\implies$ 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:

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

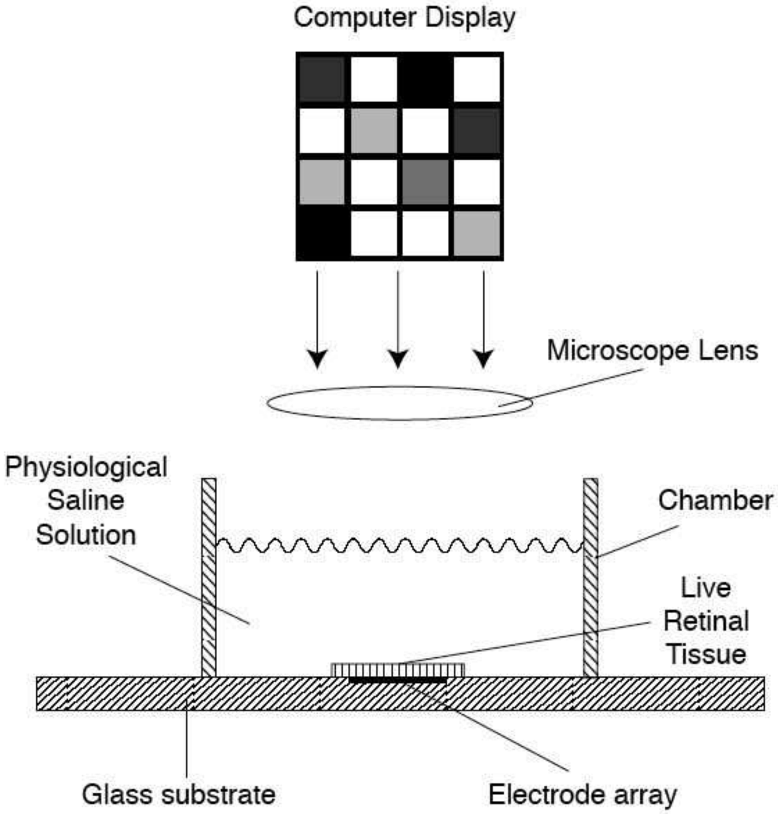
Now use fact that marginalizing preserves log-concavity (Prekopa, 1973): if $f(\vec{x}, \vec{y})$ is jointly l.c., then so is

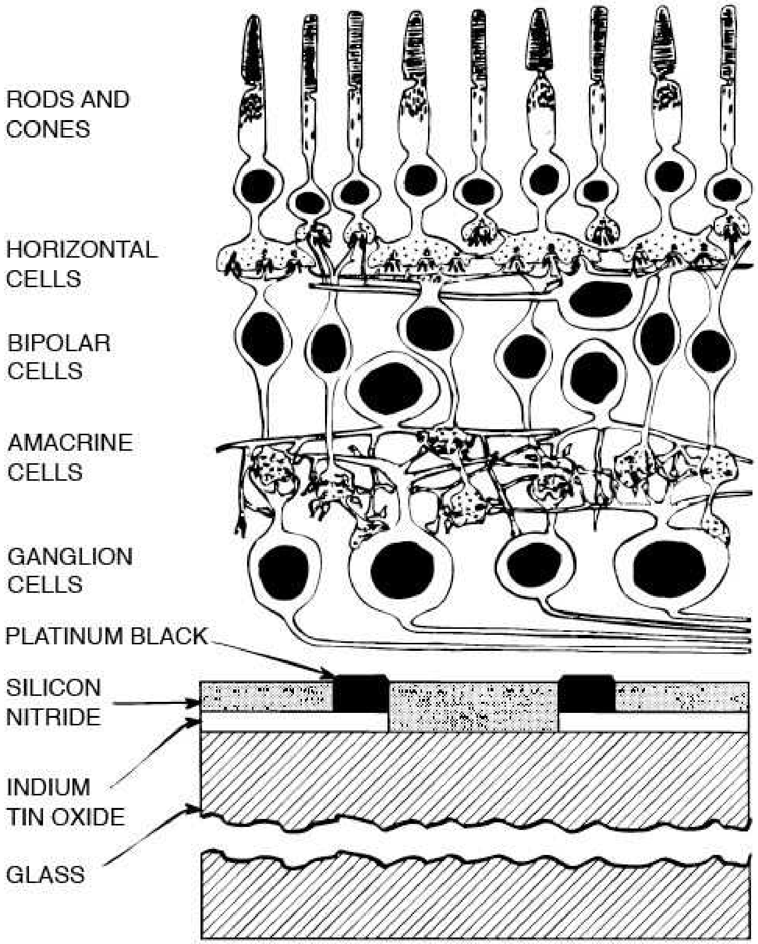
$$f_0(\vec{x}) \equiv \int f(\vec{x}, \vec{y}) d\vec{y}.$$

Application: retinal ganglion cells

Preparation: dissociated salamander and macaque retina

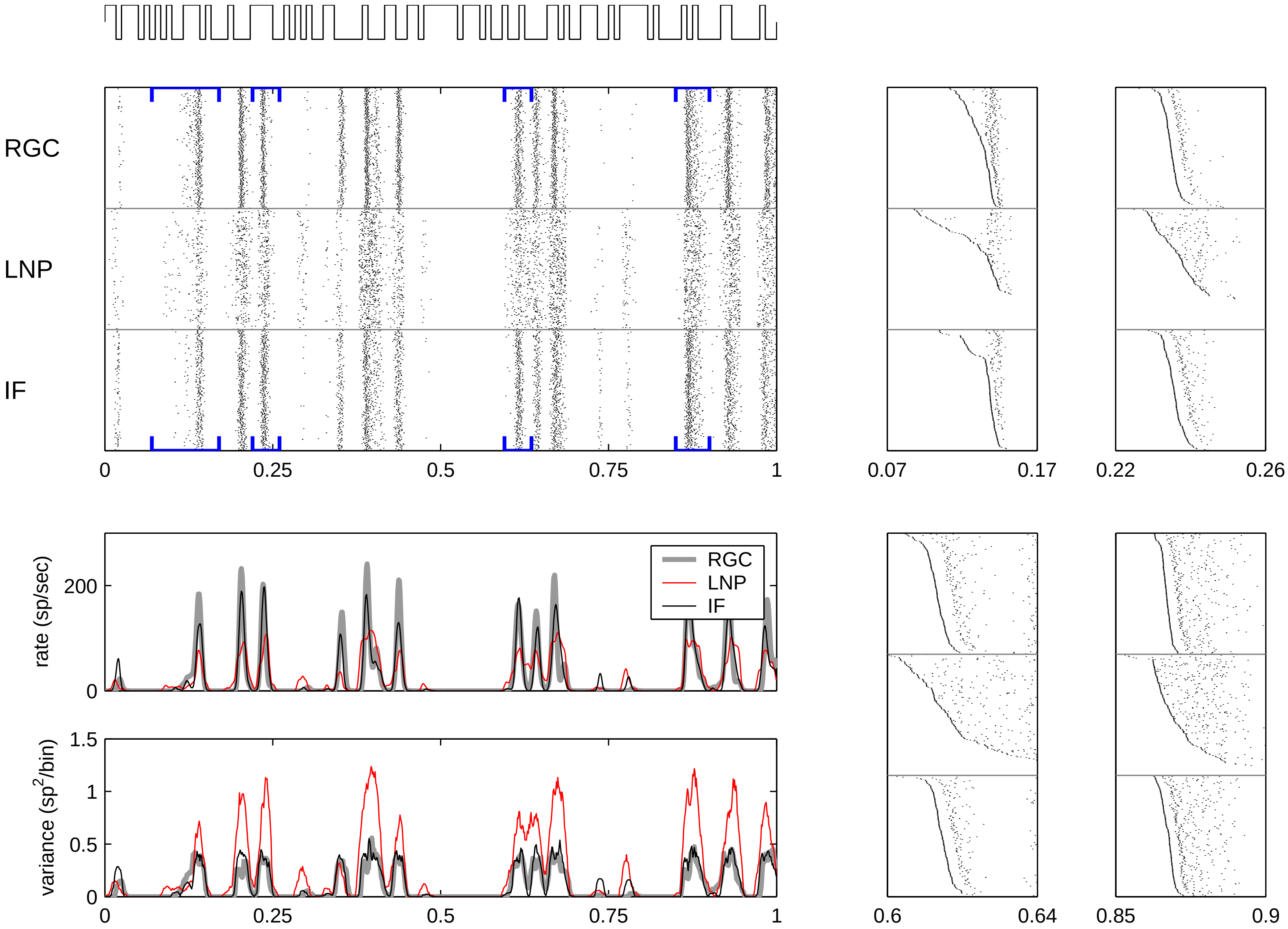
— extracellularly-recorded responses of populations of RGCs





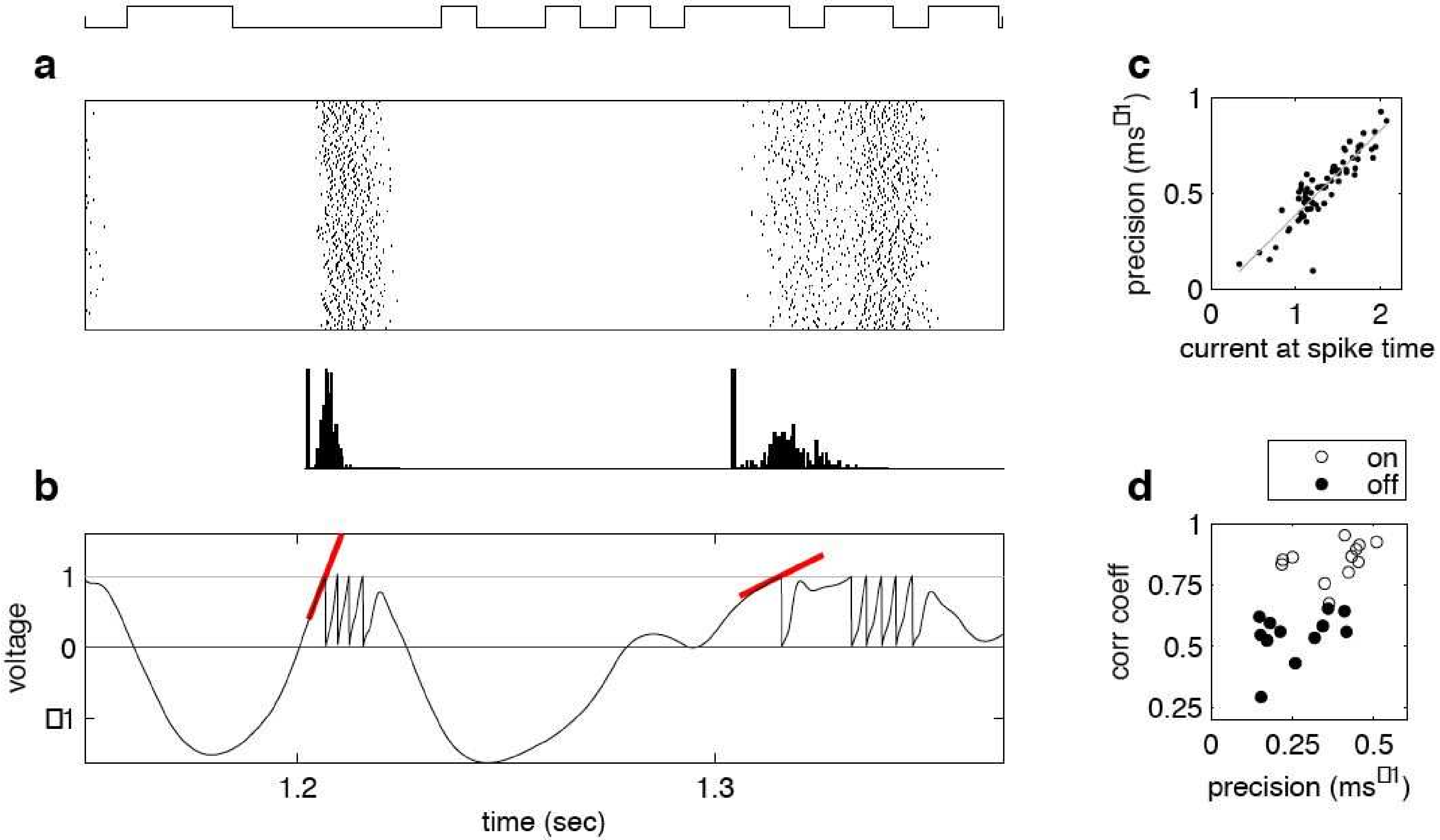
Stimulus: random “flicker” visual stimuli (Chander and Chichilnisky, 2001)

Spike timing precision in retina



(Pillow et al., 2005)

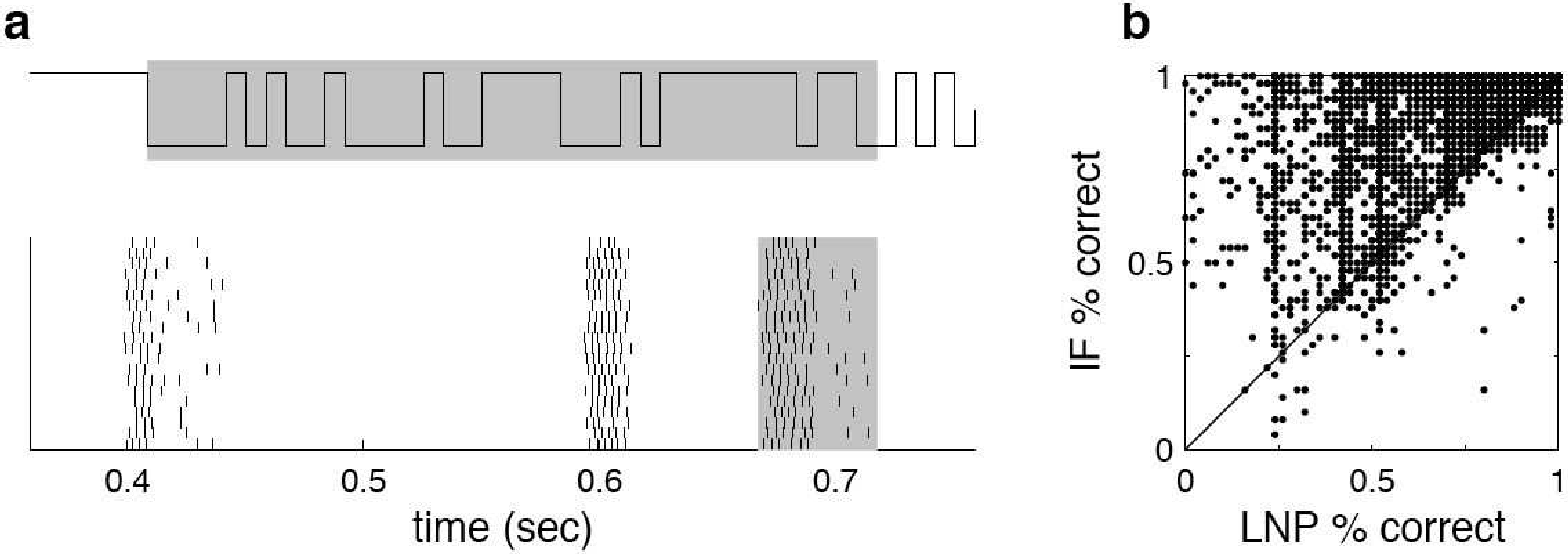
Linking spike reliability and subthreshold noise



(Pillow et al., 2005)

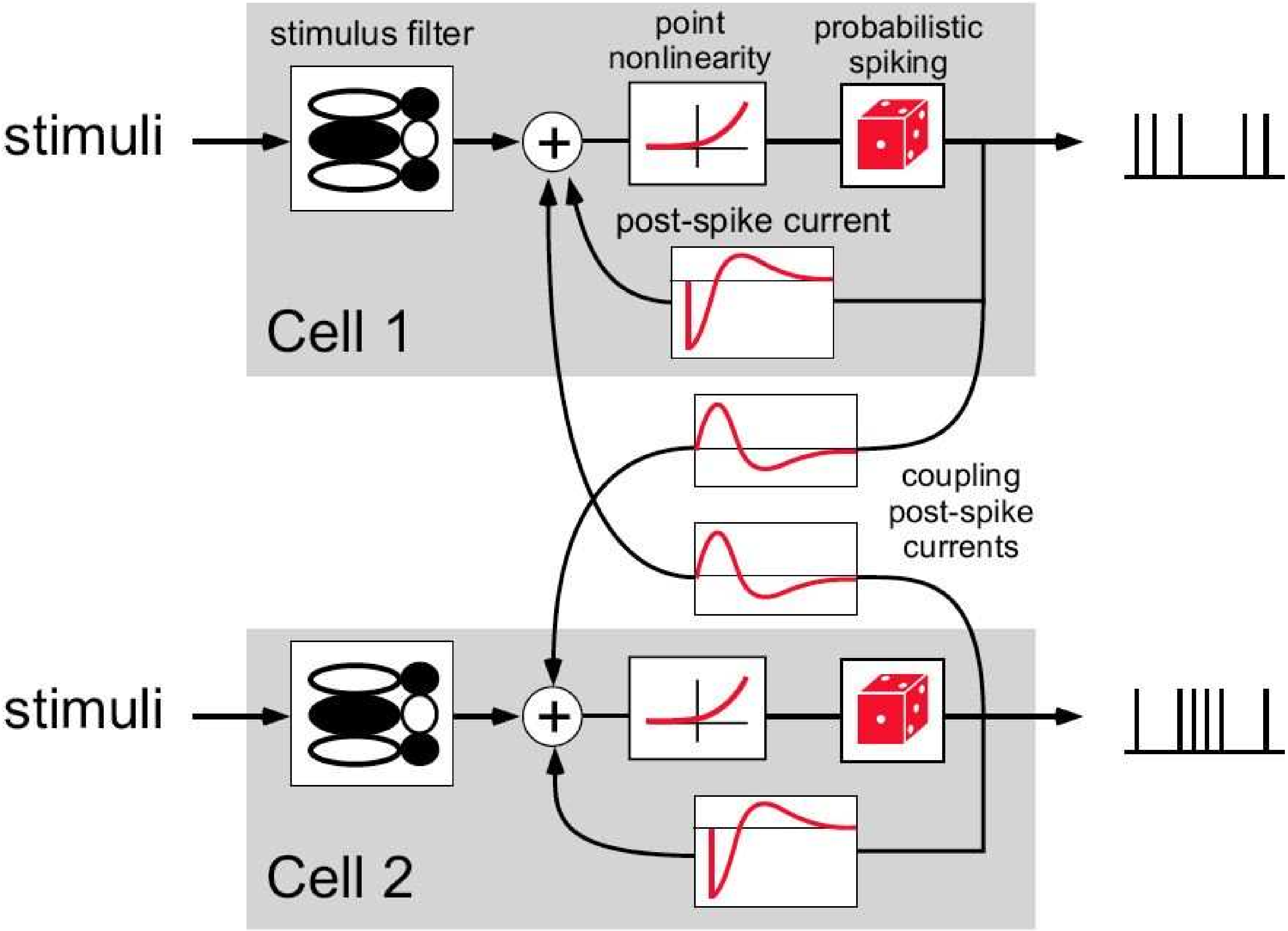
Likelihood-based discrimination

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



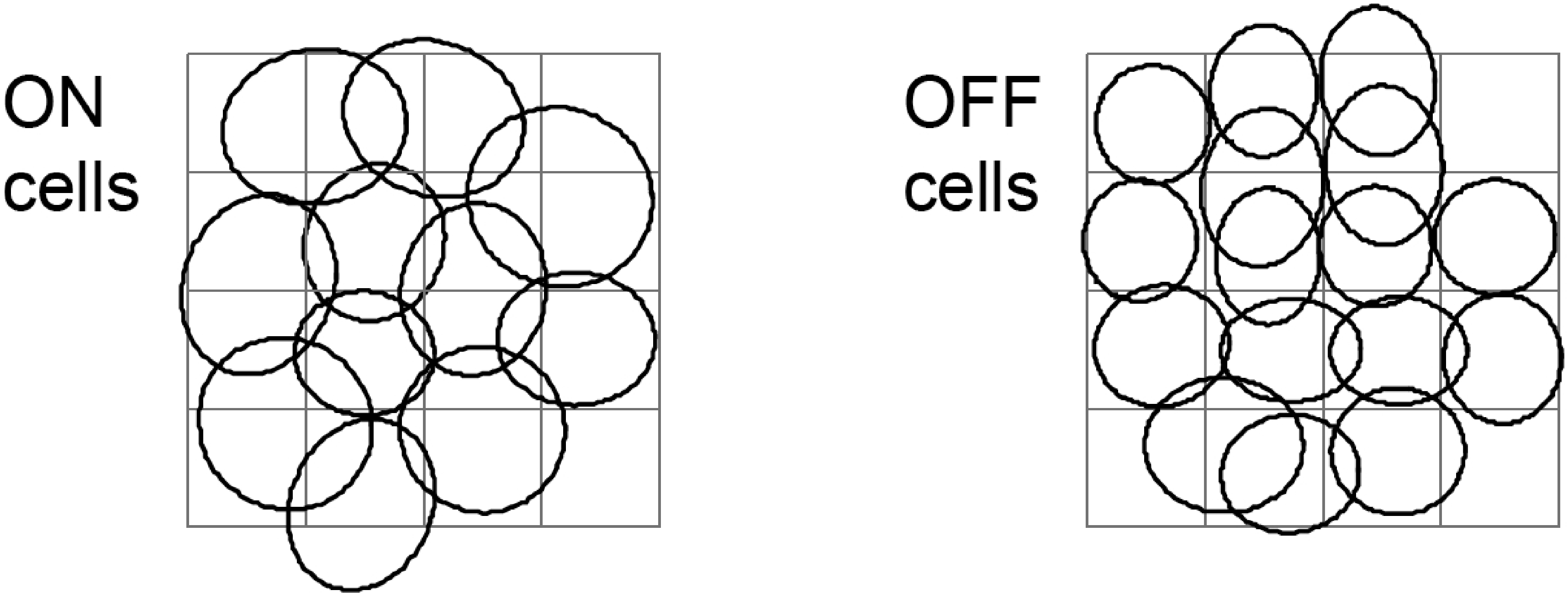
Using correct model is essential (Pillow et al., 2005)

Generalization: population responses

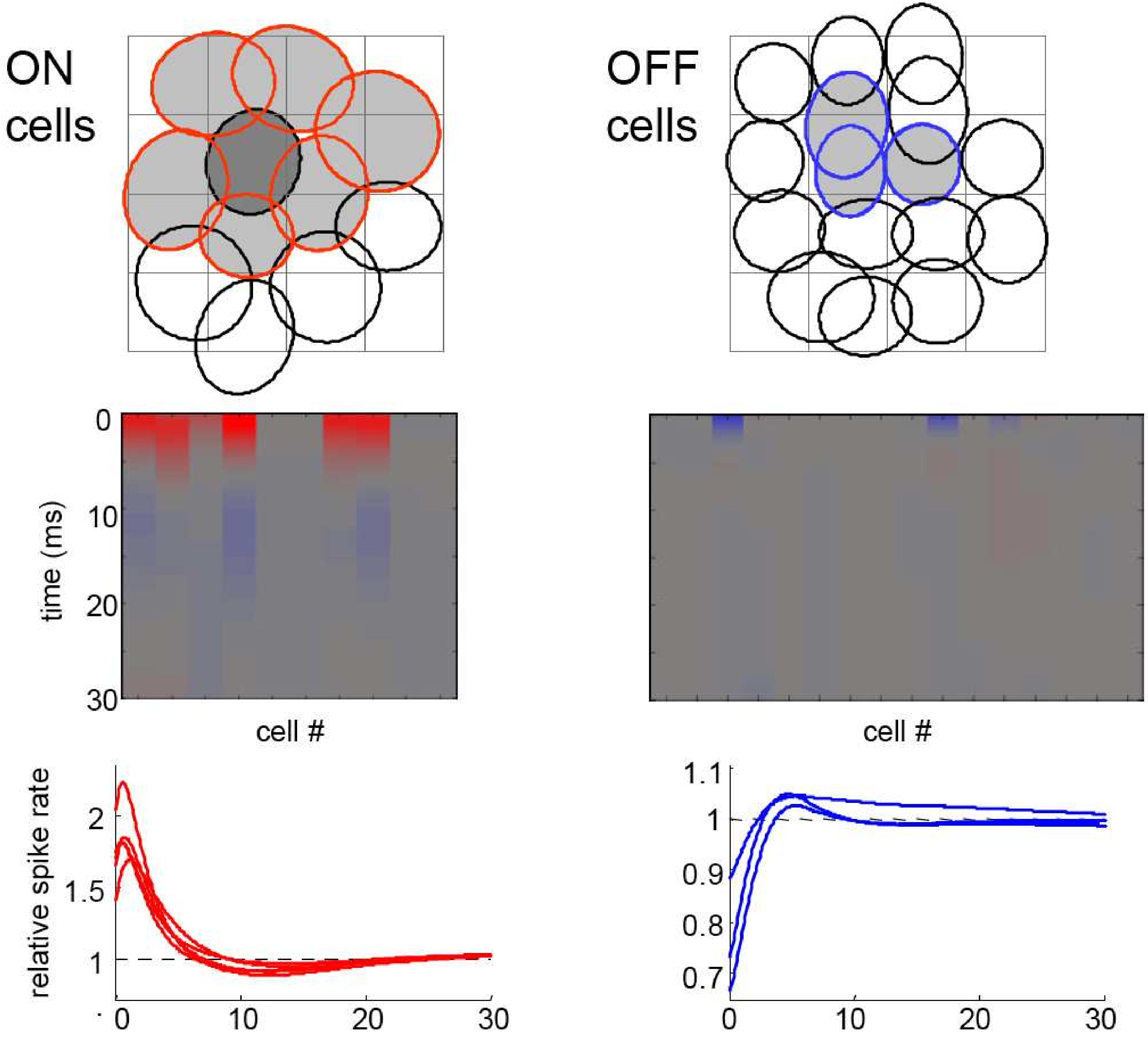


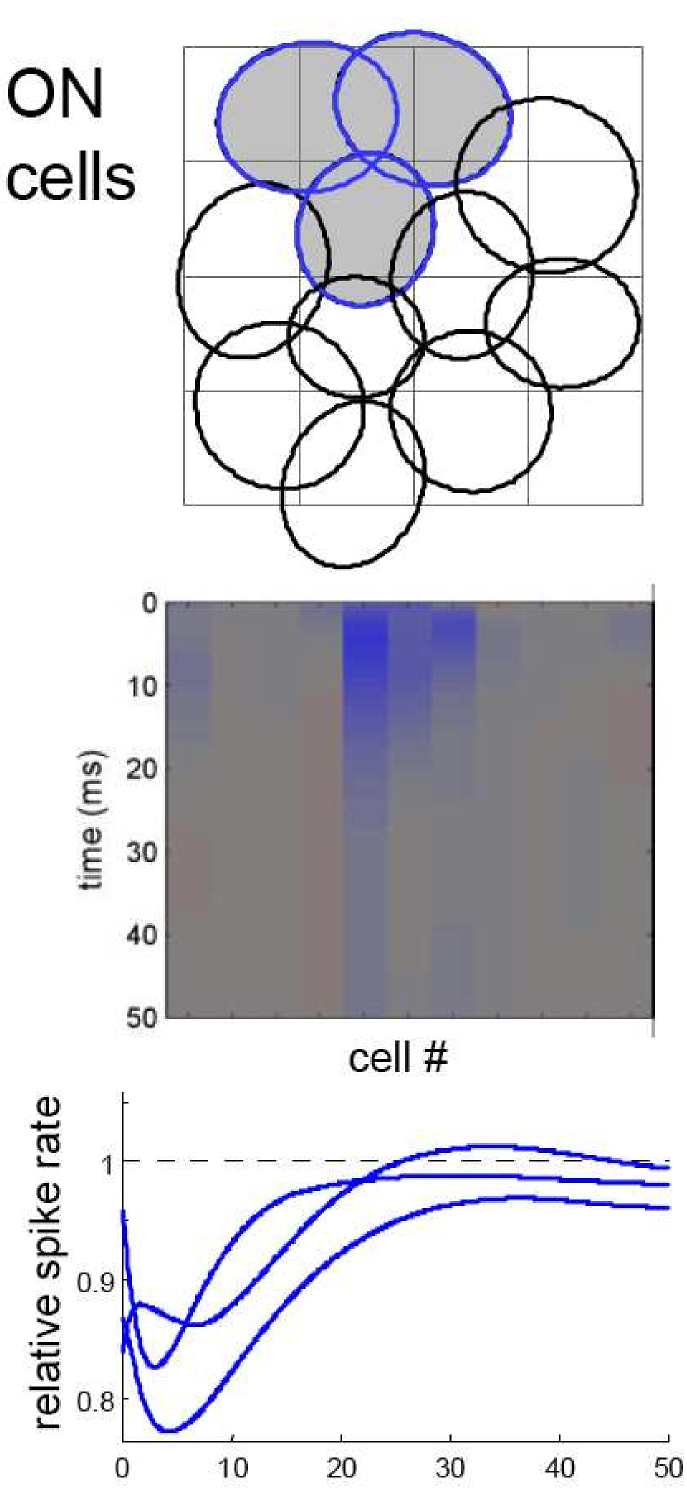
Pillow et al., SFN ’05

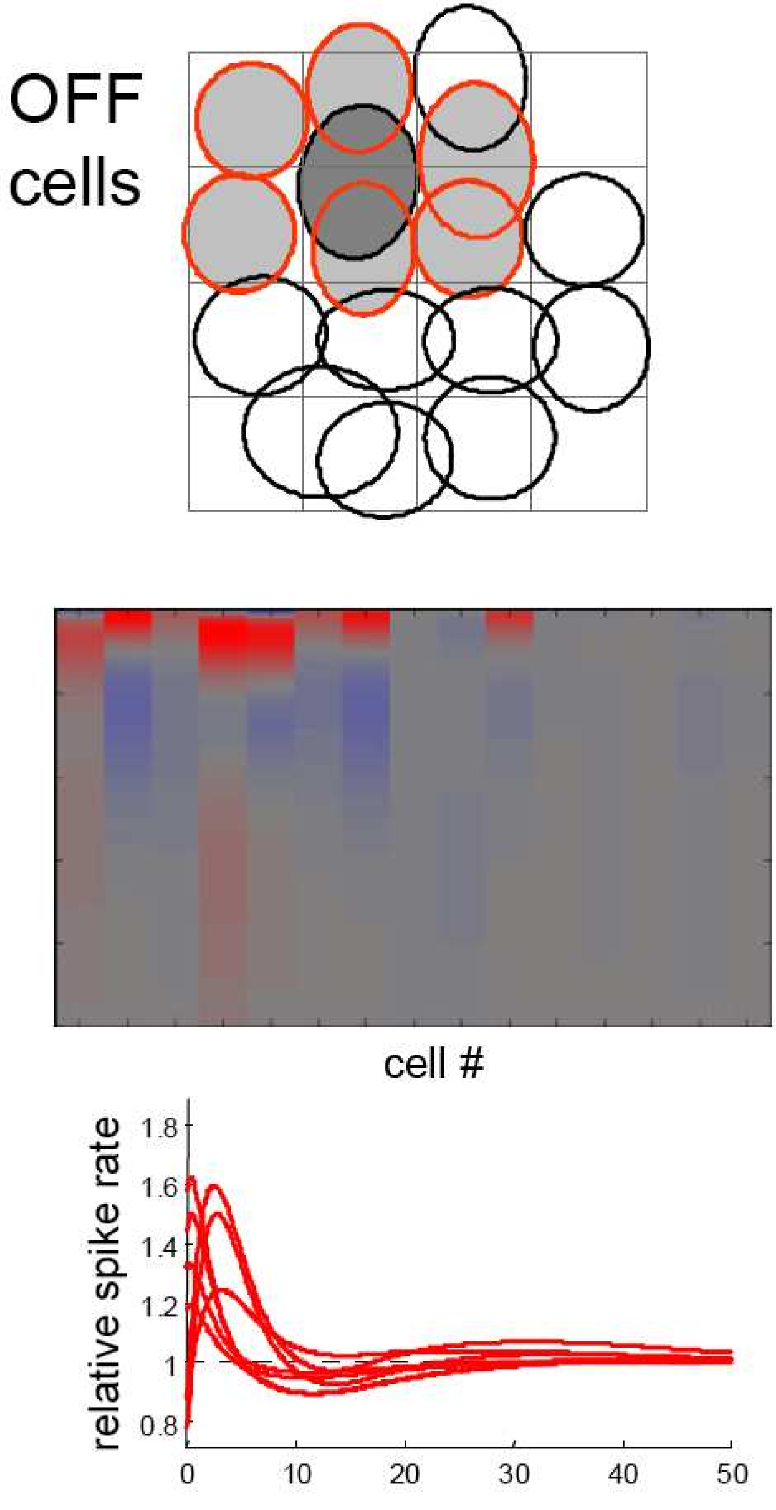
Population retinal recordings

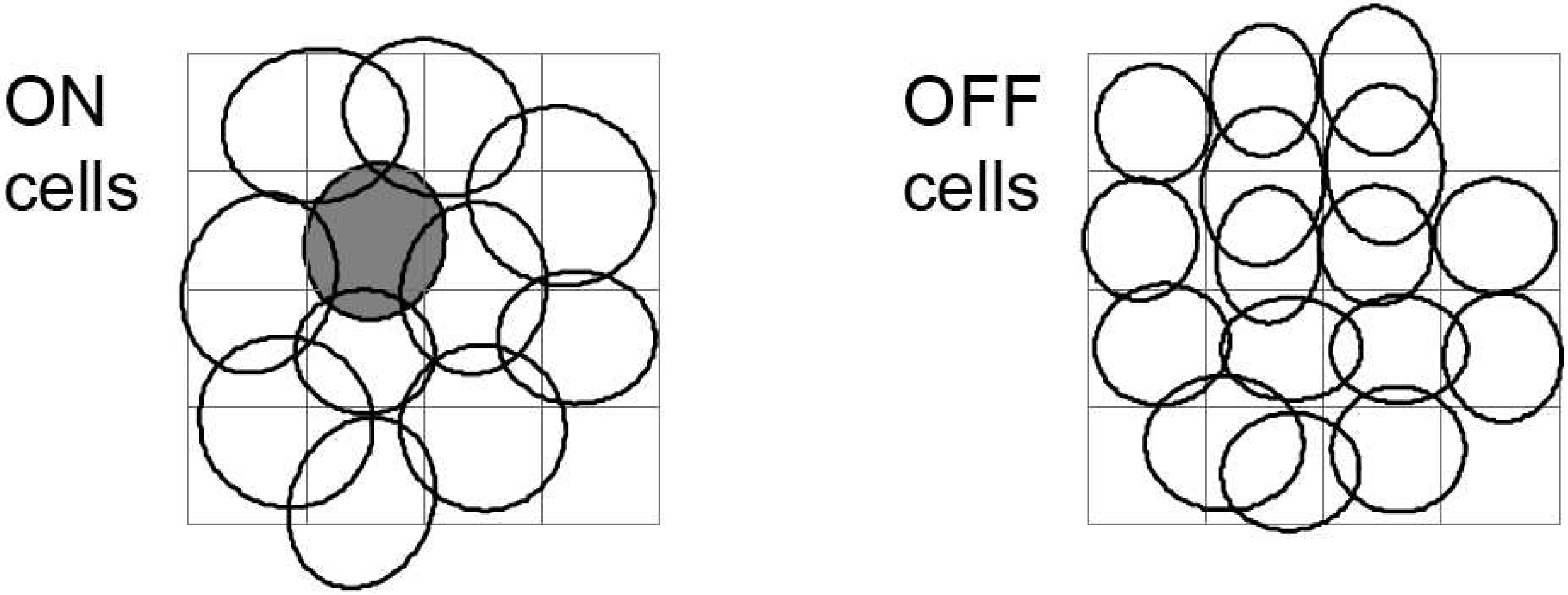


Pillow et al., SFN '05

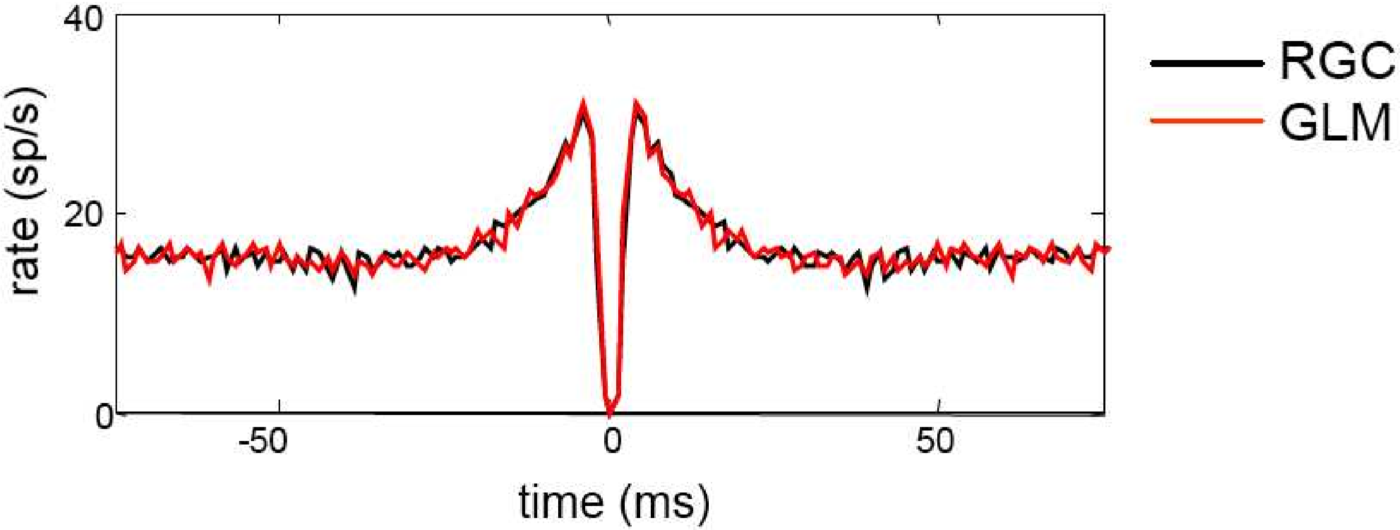

ON
cells
OFF
cells
0
10
20
30
time (ms)
cell #
cell #
relative spike rate
2
1.5
1
0
10
20
30
1.1
1
0.9
0.8
0.7
0
10
20
30

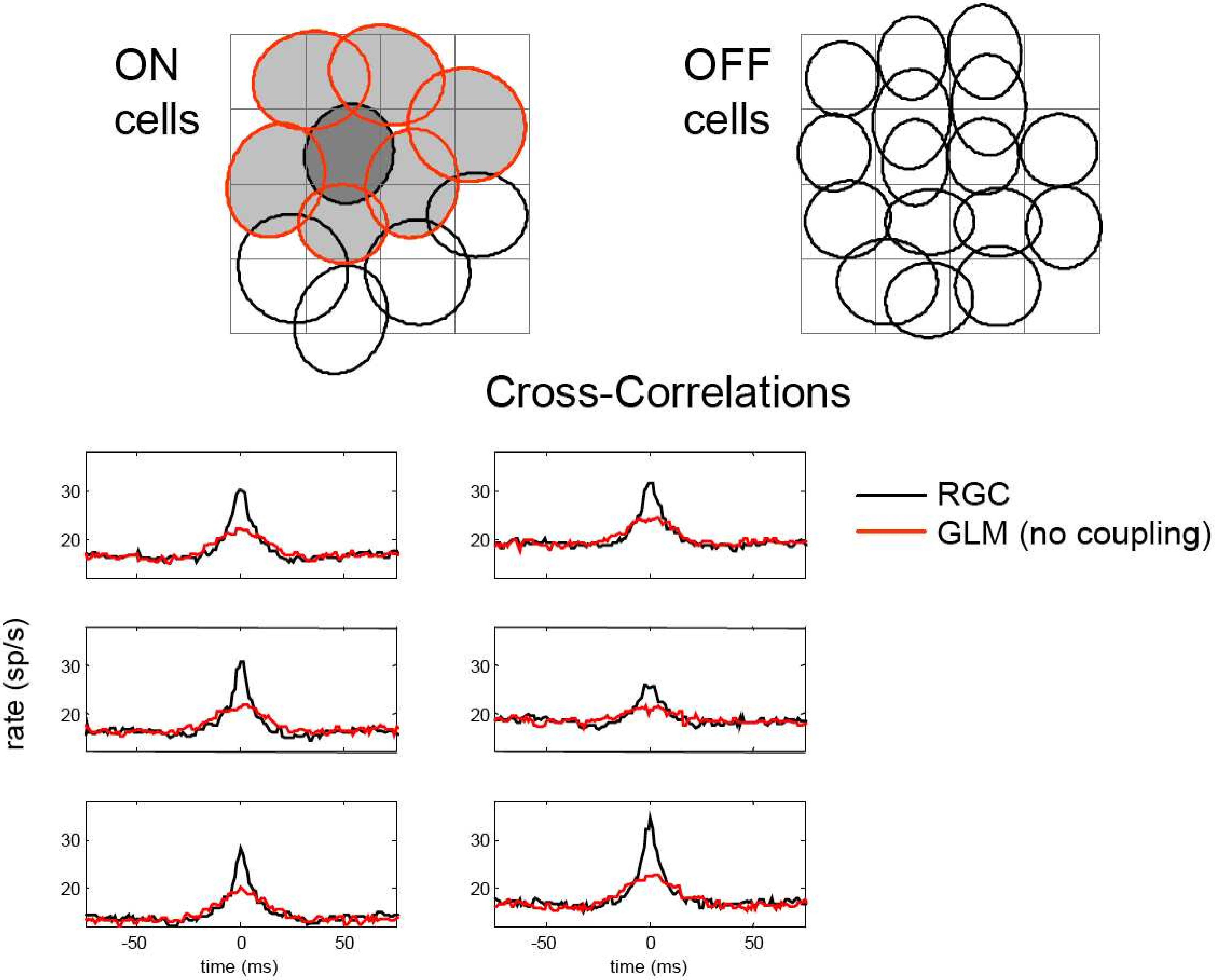

ON
cells
0
10
20
30
40
50
time (ms)
cell #
relative spike rate
1
0.9
0.8
0
10
20
30
40
50

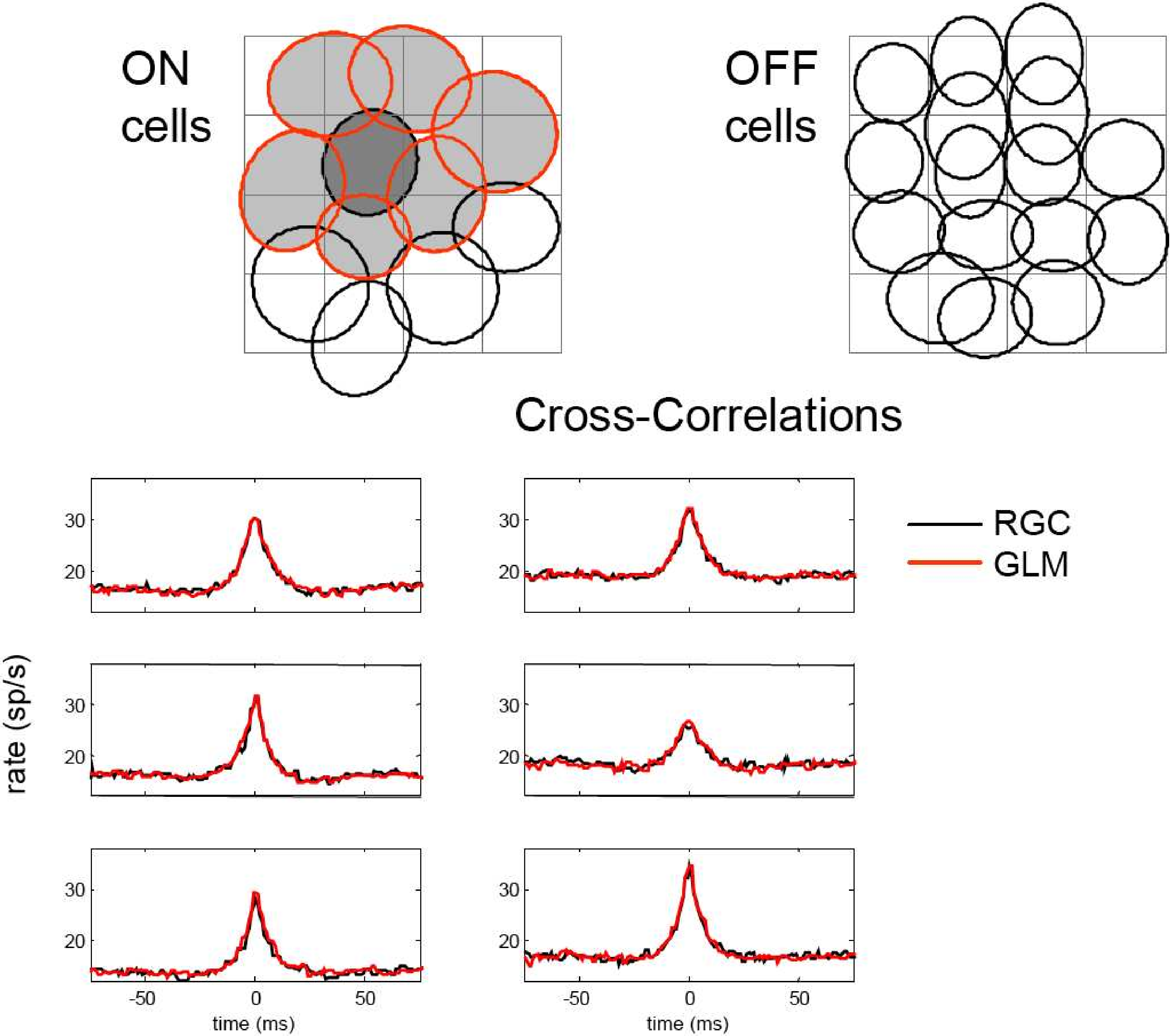

OFF
cells
cell #
relative spike rate
1.8
1.6
1.4
1.2
1
0.8
0
10
20
30
40
50



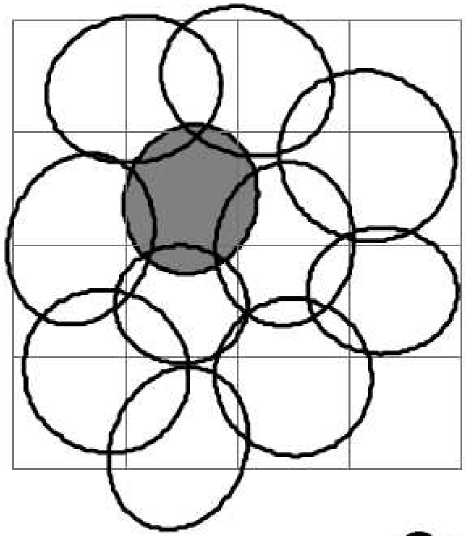
auto-correlation



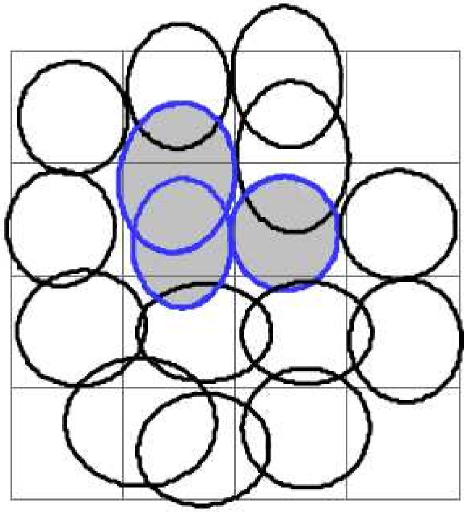

ON cells
OFF cells
Cross-Correlations
rate (sp/s)
30
20
-50
0
50
time (ms)
RGC
GLM (no coupling)


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

ON
cells

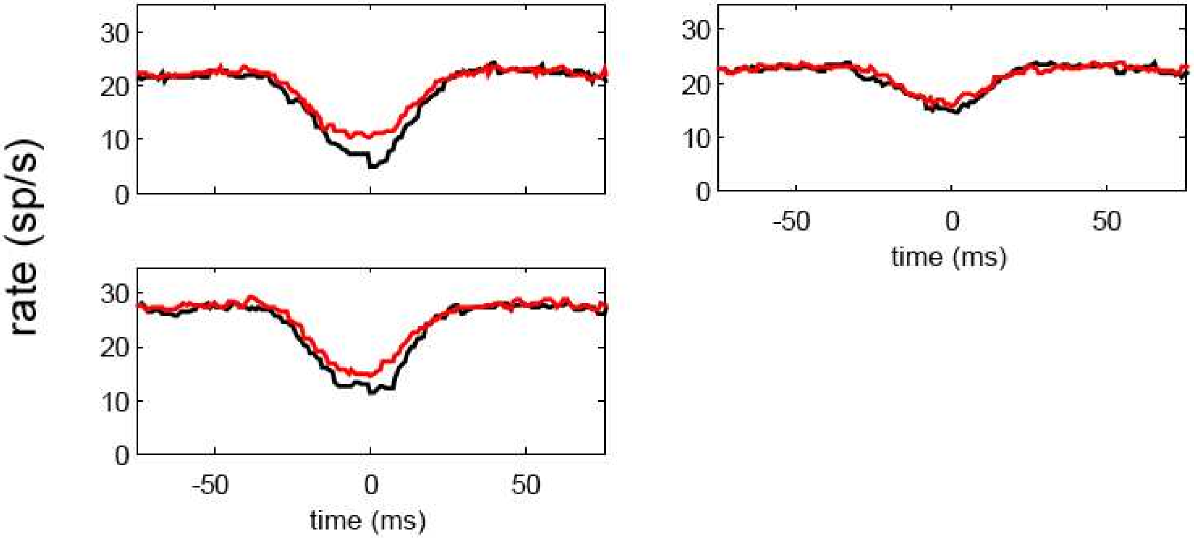


OFF
cells

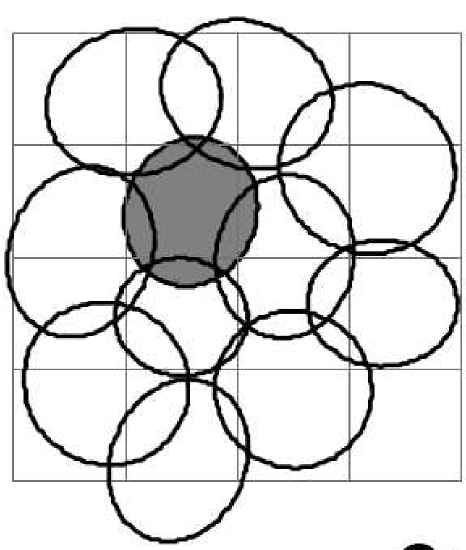


Cross-Correlations

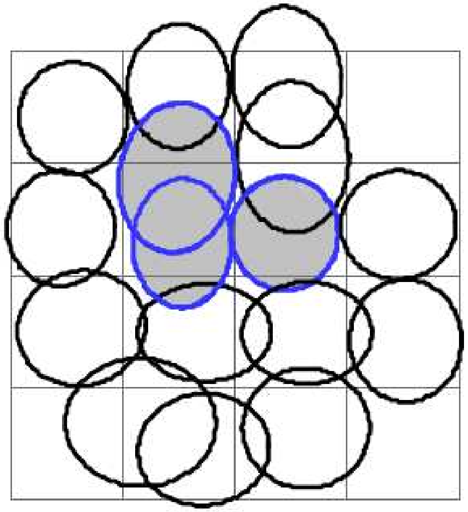
— RGC
— GLM (no coupling)



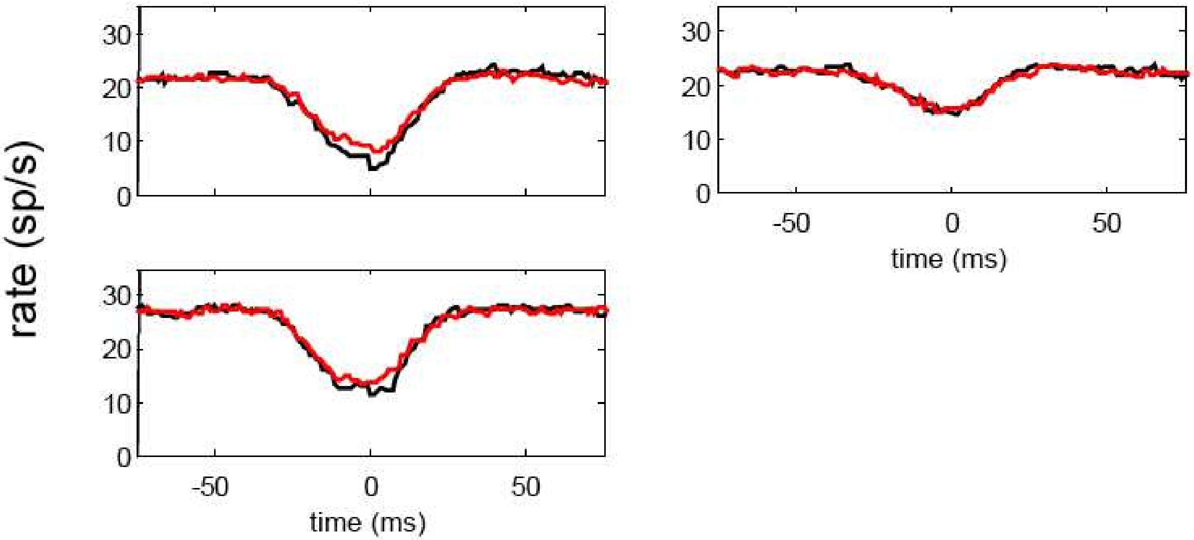
ON cells



OFF cells

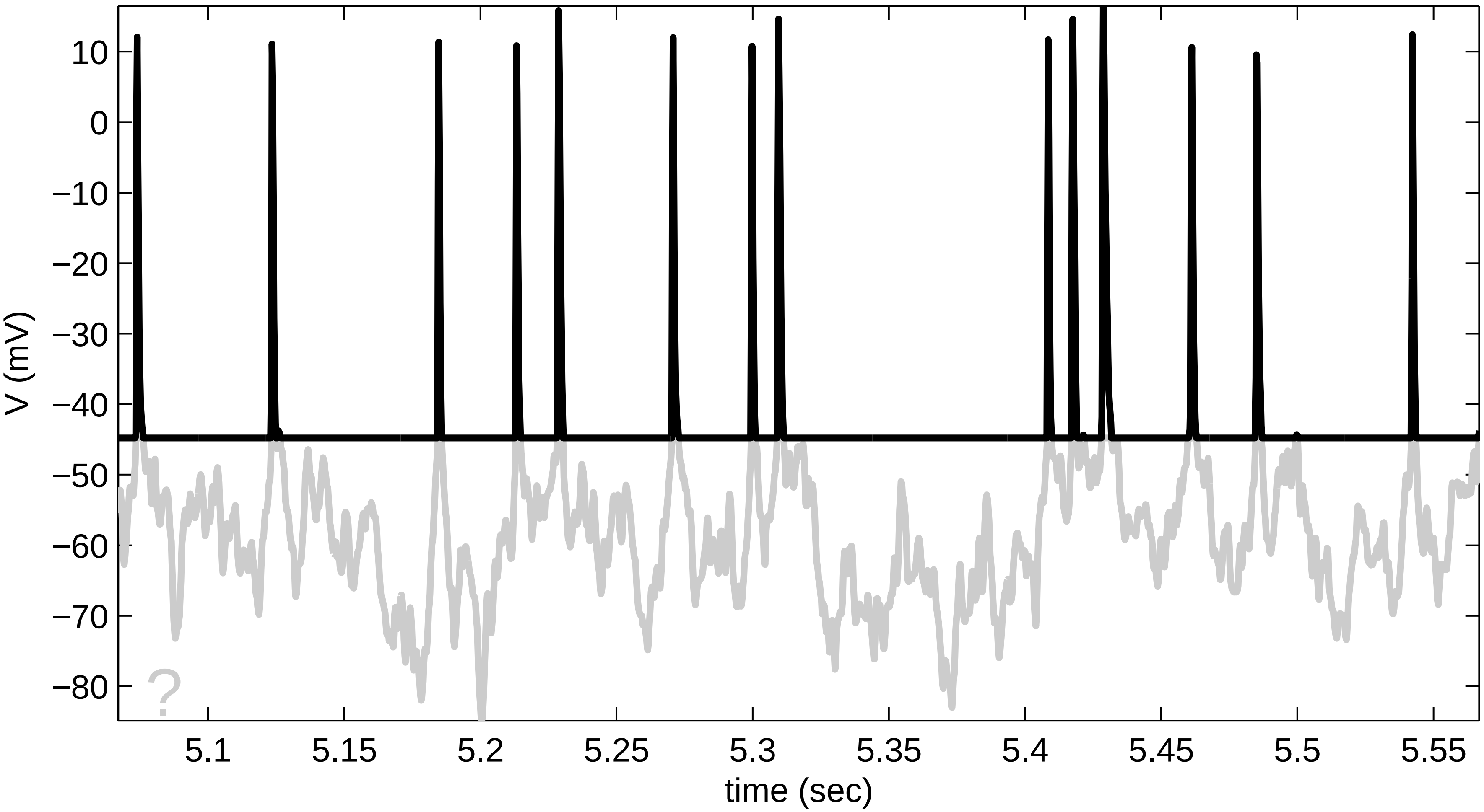


Cross-Correlations



Part 2: Decoding subthreshold activity

Given extracellular spikes, what is most likely intracellular $V(t)$?



Computing $V_{ML}(t)$

Loglikelihood of $V(t)$ (given LIF parameters, white noise N_t):

$$L(\{V(t)\}_{0\leq t\leq T}) = -\frac{1}{2\sigma^2}\int_0^T \left[\dot{V}(t) - \Big(- gV(t) + I(t)\Big)\right]^2 dt$$

Constraints:

- Reset at $t = 0$:

$$V(0) = V_{reset}$$

- Spike at $t = T$:

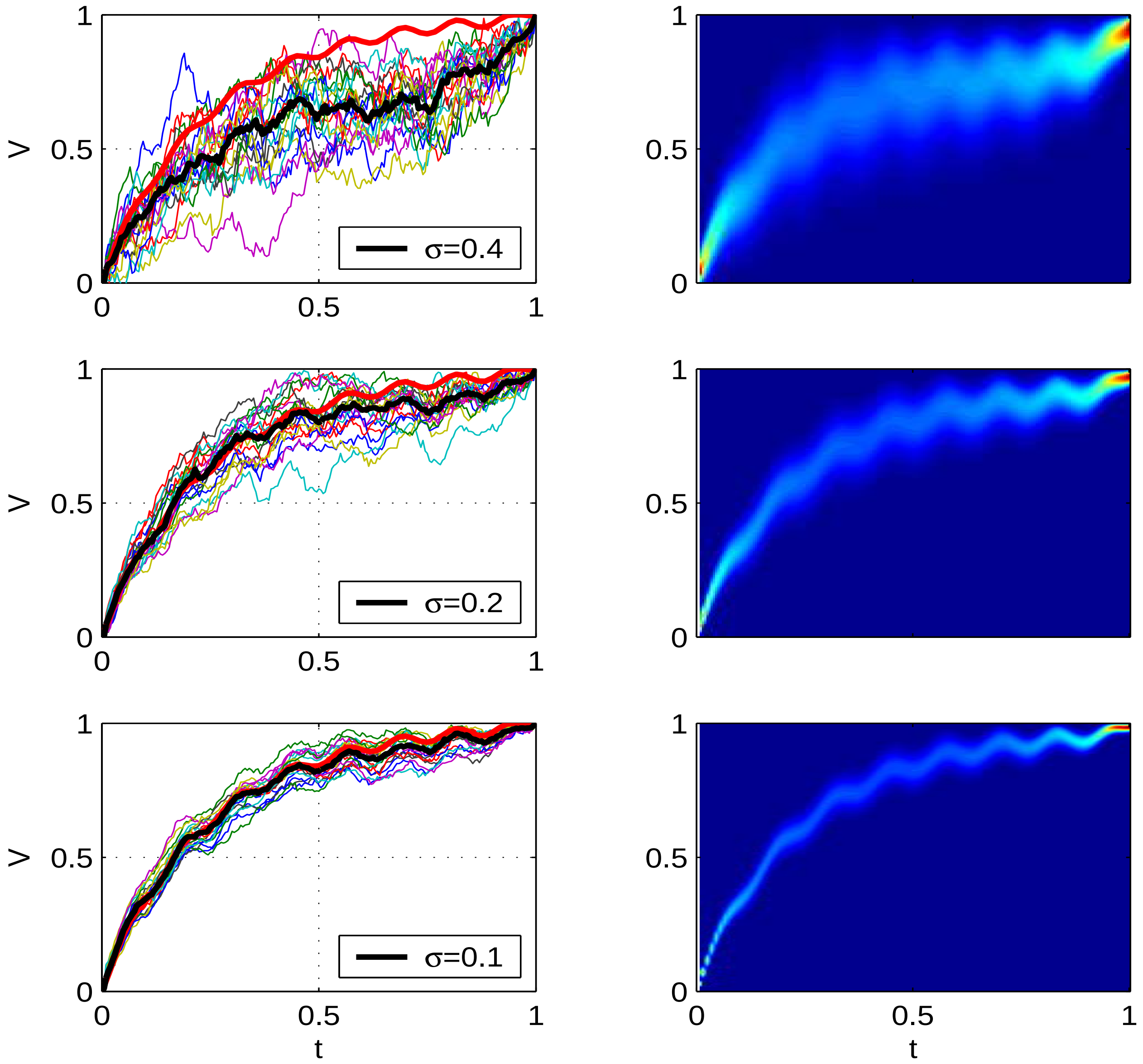
$$V(T) = V_{th}$$

- No spike for $0 < t < T$:

$$V(t) < V_{th}$$

Quadratic programming problem: optimize quadratic function under linear constraints. **Concave**: unique global optimum.

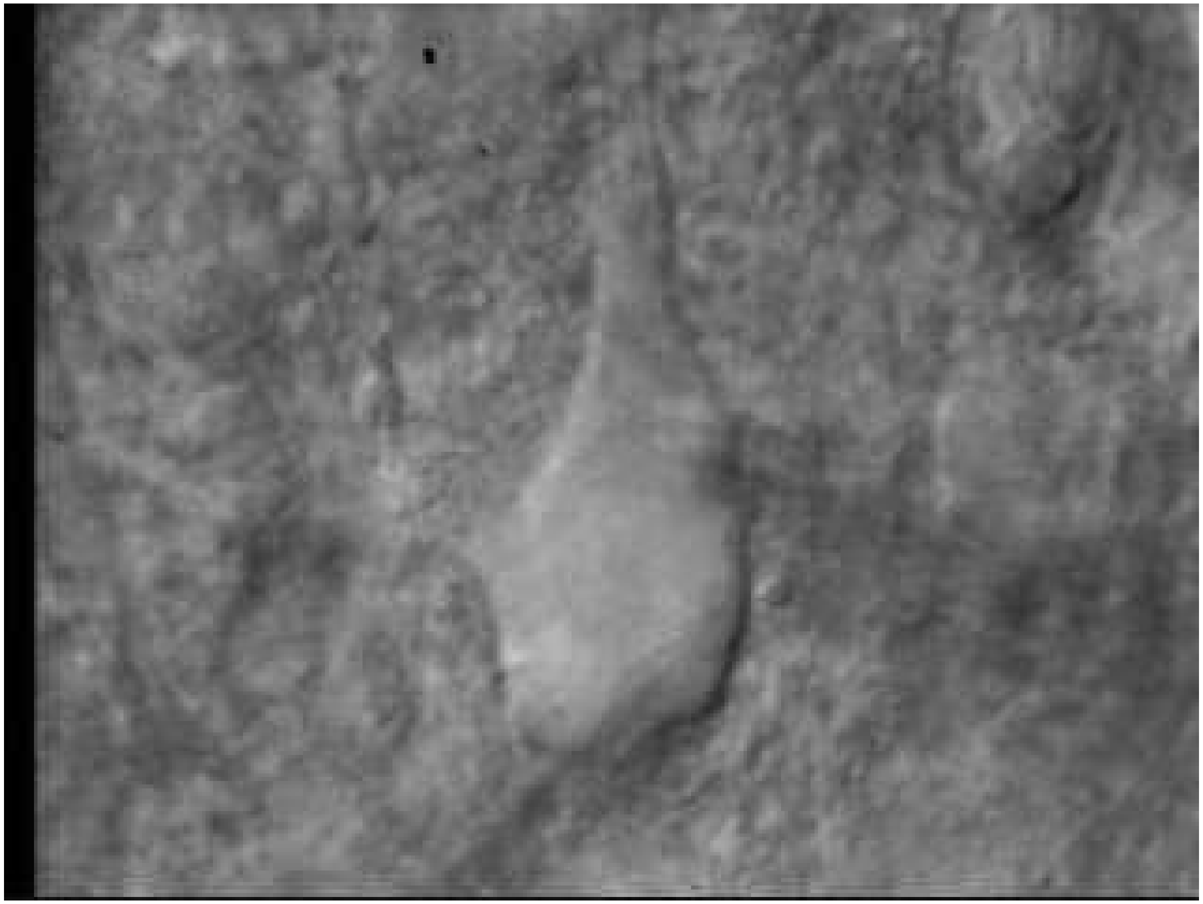
Most likely vs. average $V(t)$



(Applications to spike-triggered average (Paninski, 2005a; Paninski, 2005b))

Application: *in vitro* data

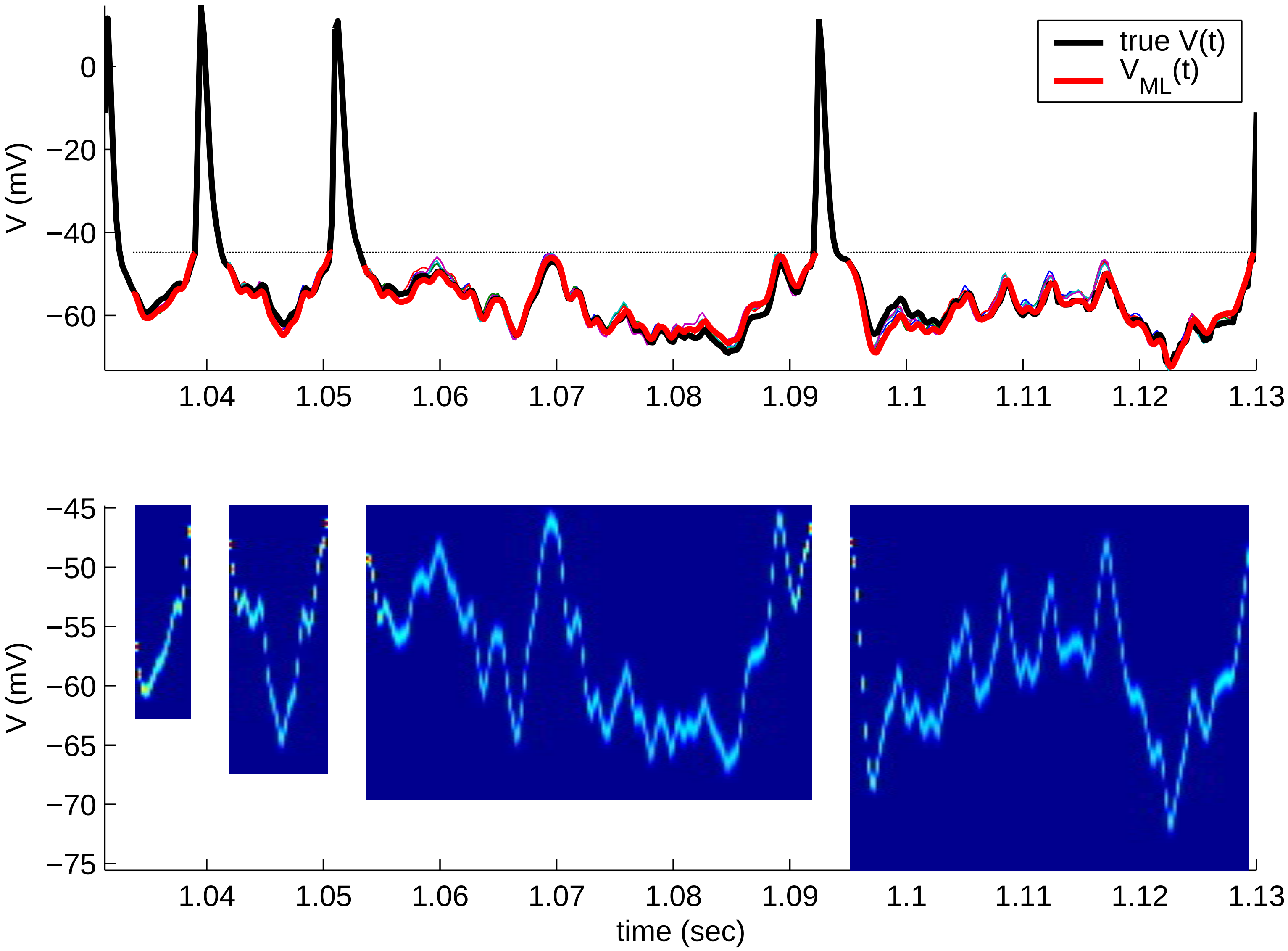
Recordings: rat sensorimotor cortical slice; dual-electrode whole-cell



Stimulus: Gaussian white noise current $I(t)$

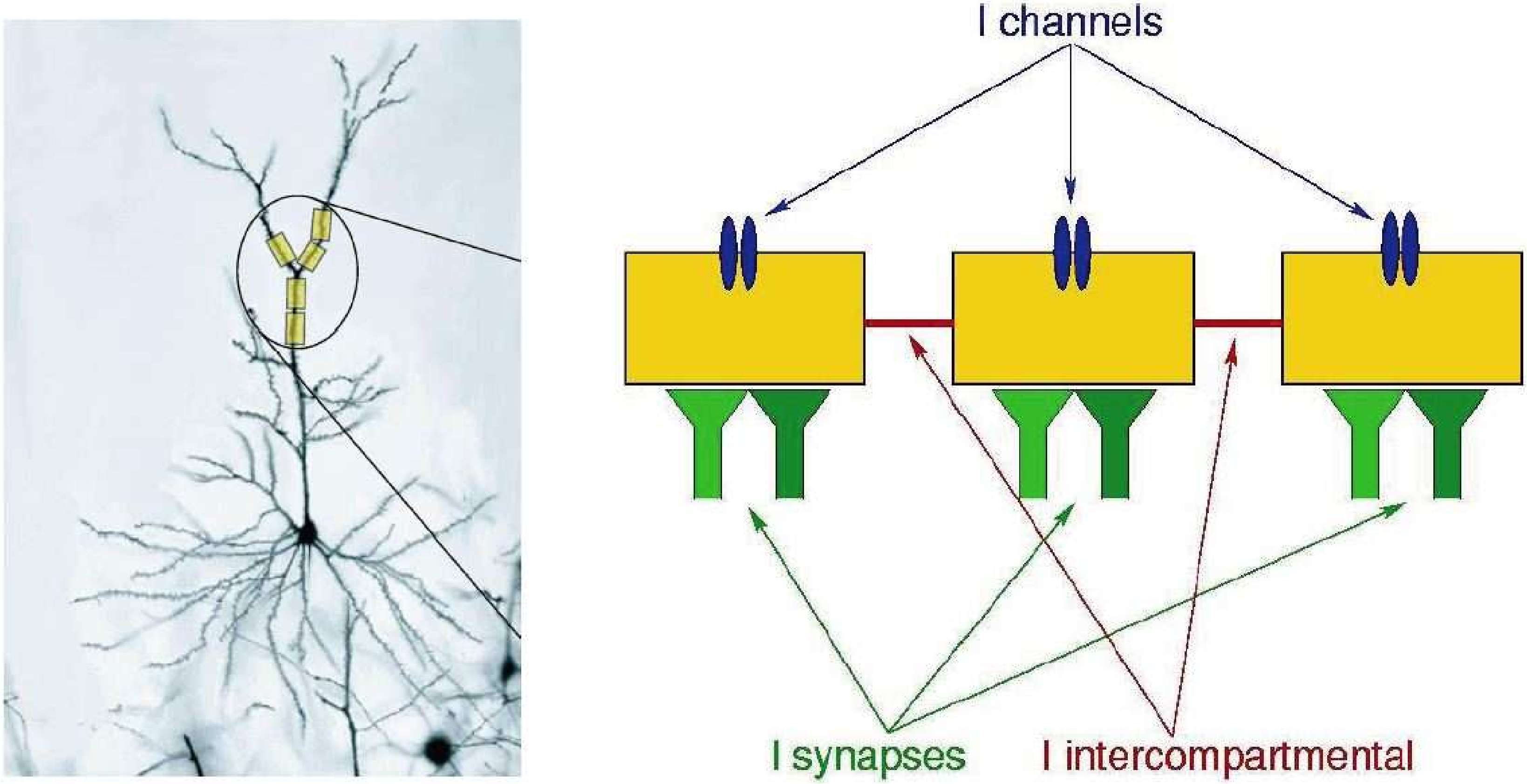
Analysis: fit IF model parameters $\{g, \vec{k}, h(.), V_{th}, \sigma\}$ by maximum likelihood (Paninski et al., 2003; Paninski et al., 2004a), then compute $V_{ML}(t)$

Application: *in vitro* data



$P(V(t)|\{t_i\}, \hat{\theta}_{ML}, \vec{x})$ computed via forward-backward hidden Markov model method (Paninski, 2005a).

Part 3: Back to detailed models



Can we recover detailed biophysical properties?

- Active: membrane channel densities
- Passive: axial resistances, “leakiness” of membranes
- Dynamic: spatiotemporal synaptic input

Conductance-based models

$$C\frac{dV_i}{dt} = I_i^{\text{channels}} + I_i^{\text{synapses}} + I_i^{\text{intercompartmental}}$$

$$I_i^{\text{channels}} = \sum_c \bar{g}_c g_c(t)(E_c - V_i(t))$$

$$I_i^{\text{synapses}} = \sum_s (\xi_s * k_s)(t)(E_s - V_i(t))$$

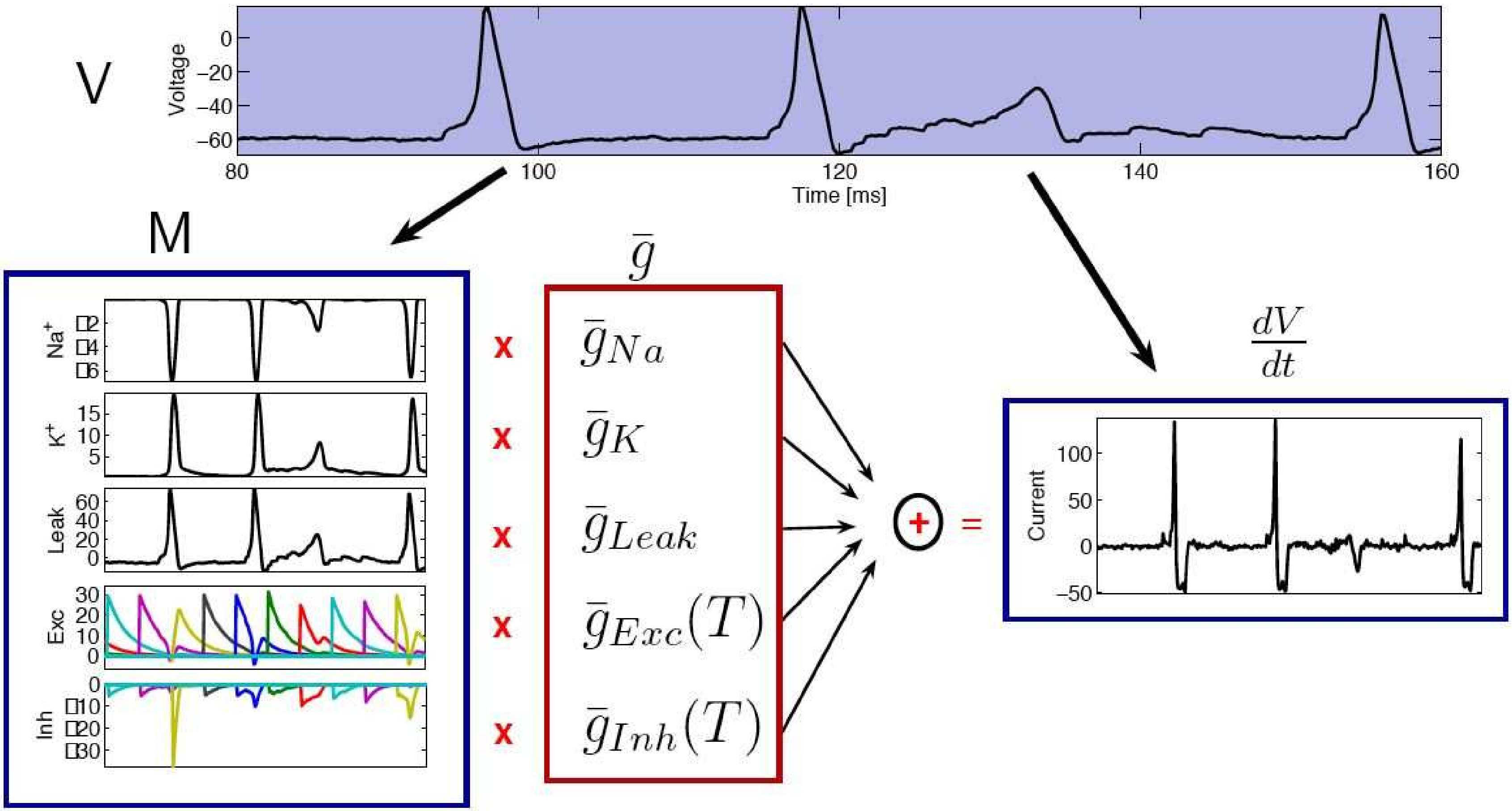
$$I_i^{\text{intercompartmental}} = \sum_a g_a \Delta V_a(t)$$

Key point: **if** we observe full $V_i(t)$ + cell geometry, channel kinetics known + current noise is log-concave,

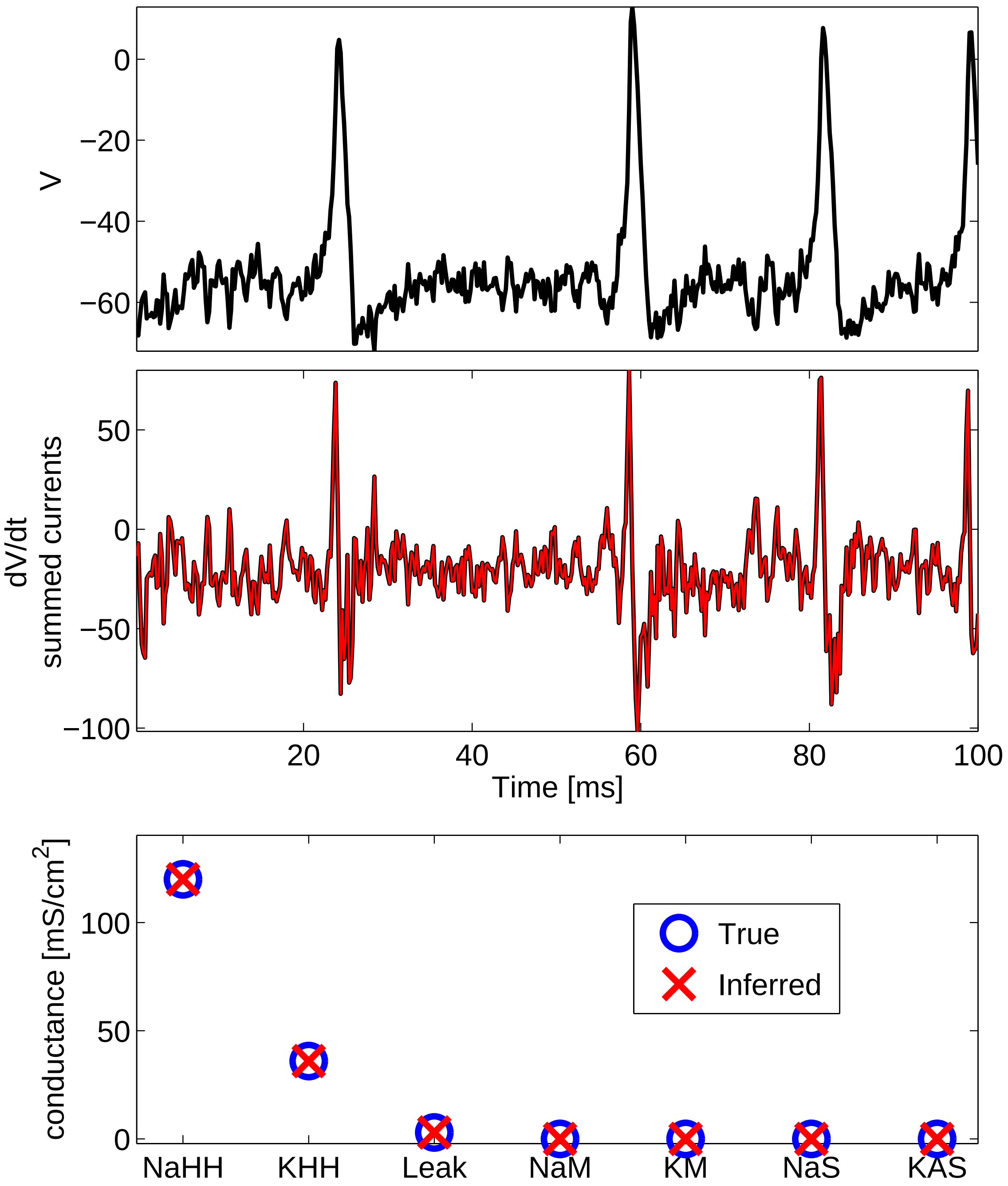
then loglikelihood of unknown parameters is concave.

Gaussian noise $\Longrightarrow$ standard nonnegative regression (albeit high-d).

Estimating channel densities from $V(t)$



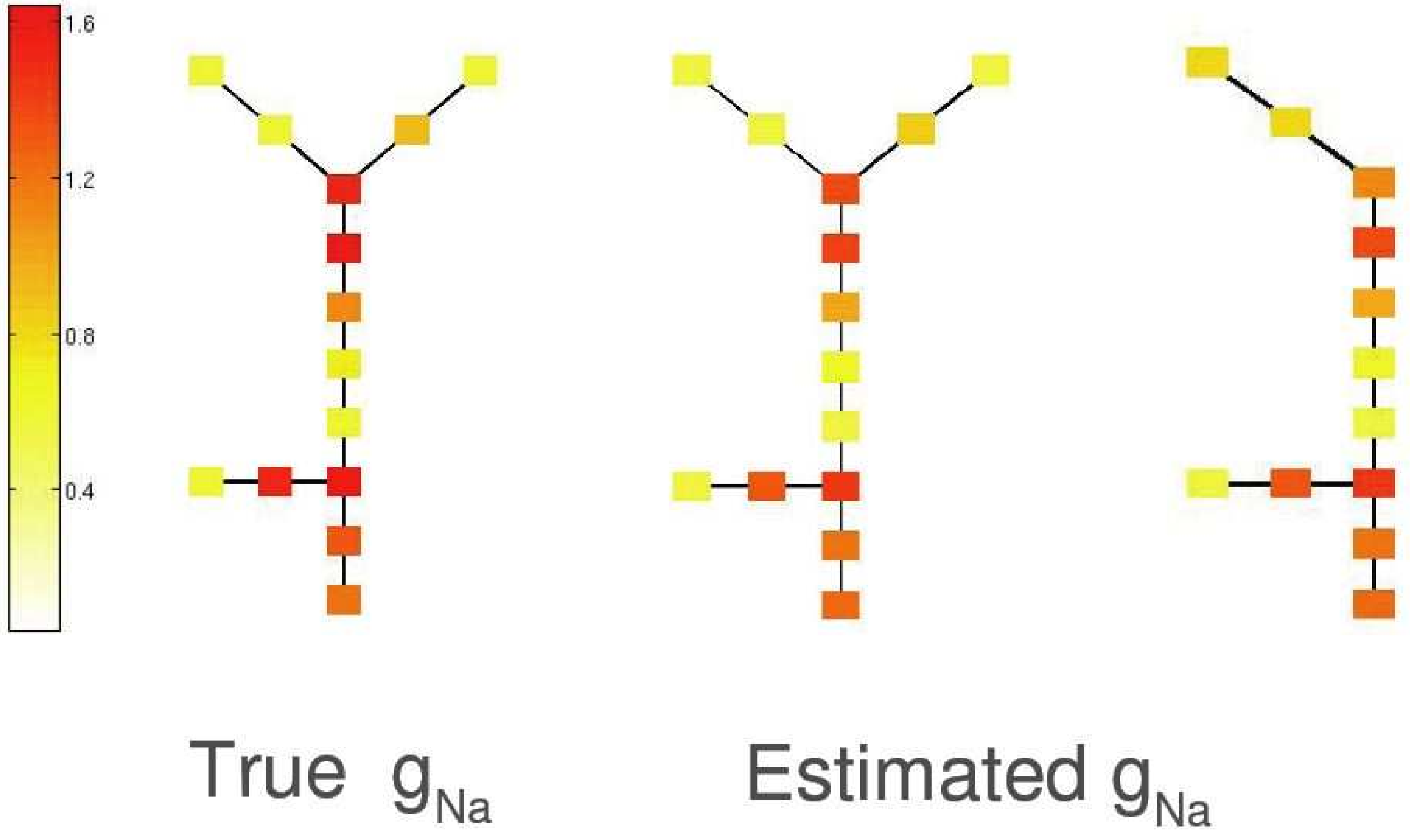
Ahrens, Huys, Paninski, NIPS ’05



Ahrens, Huys, Paninski, NIPS ’05

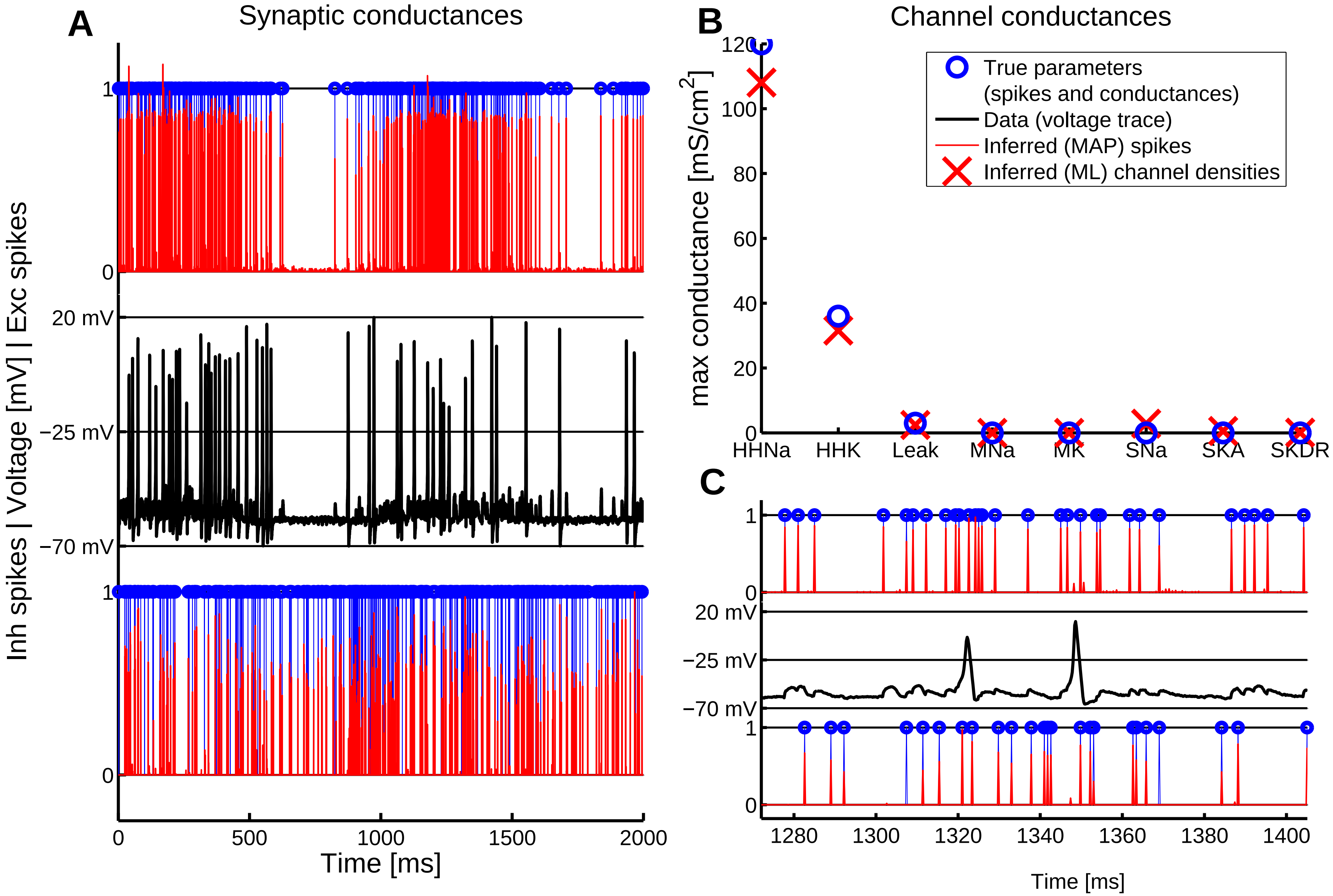
Estimating non-homogeneous channel densities and axial resistances from spatiotemporal voltage recordings

$$I_i^{\text{channels}} = \sum_c \bar{g}_c g_c(t)(E_c - V_i(t))$$



Ahrens, Huys, Paninski, COSYNE ’05

Estimating synaptic inputs given $V(t)$



Ahrens, Huys, Paninski, NIPS '05

Collaborators

Theory and numerical methods

— J. Pillow, E. Simoncelli, NYU

— S. Shoham, Princeton

— A. Haith, C. Williams, Edinburgh

— M. Ahrens, Q. Huys, Gatsby

Motor cortex physiology

— M. Fellows, J. Donoghue, Brown

— N. Hatsopoulos, U. Chicago

— B. Townsend, R. Lemon, U.C. London

Retinal physiology

— V. Uzzell, J. Shlens, E.J. Chichilnisky, UCSD

Cortical *in vitro* physiology

— B. Lau and A. Reyes, NYU

References

Paninski, L. (2003a). Design of experiments via information theory. *Advances in Neural Information Processing Systems*, 16.

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

Paninski, L. (2005a). The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *Journal of Computational Neuroscience*, under review.

Paninski, L. (2005b). The spike-triggered average of the integrate-and-fire cell driven by Gaussian white noise. *Submitted.*

Paninski, L., Fellows, M., Hatsopoulos, N., and Donoghue, J. (1999). Coding dynamic variables in populations of motor cortex neurons. *Society for Neuroscience Abstracts*, 25:665.9.

Paninski, L., Haith, A., Pillow, J., and Williams, C. (2005). Improved numerical methods for computing likelihoods in the stochastic integrate-and-fire model. *Comp. Sys. Neur. '05.*

Paninski, L., Lau, B., and Reyes, A. (2003). Noise-driven adaptation: in vitro and mathematical analysis. *Neurocomputing*, 52:877–883.

Paninski, L., Pillow, J., and Simoncelli, E. (2004a). Comparing integrate-and-fire-like models estimated using intracellular and extracellular data. *Neurocomputing*, 65:379–385.

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

Pillow, J., Paninski, L., Uzzell, V., Simoncelli, E., and Chichilnisky, E. (2005). Accounting for timing and variability of retinal ganglion cell light responses with a stochastic integrate-and-fire model. *Journal of Neuroscience*, 25:11003–11013.

Serruya, M., Hatsopoulos, N., Paninski, L., Fellows, M., and Donoghue, J. (2002). Instant neural control of a movement signal. *Nature*, 416:141–142.

Shoham, S., Paninski, L., Fellows, M., Hatsopoulos, N., Donoghue, J., and Normann, R. (2005). Optimal decoding for a primary motor cortical brain-computer interface. *IEEE Transactions on Biomedical Engineering*, 52:1312–1322.

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