

# Statistical challenges in neural data analysis

Liam Paninski

Department of Statistics and Center for Theoretical Neuroscience  
Columbia University

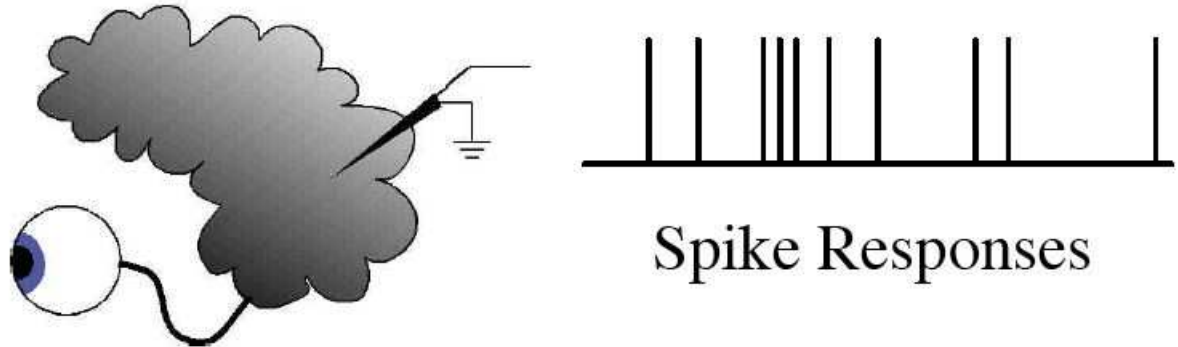
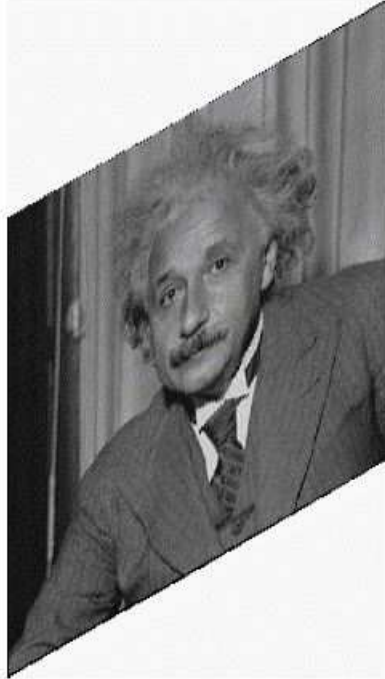
<http://www.stat.columbia.edu/~liam>

*liam@stat.columbia.edu*

August 24, 2009

Support: Sloan Fellowship, NIH collaborative research in computational neuroscience, NSF CAREER, McKnight Scholar award.

# The neural code



Basic goal: infer input-output relationship between

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

# Several levels of neural data analysis

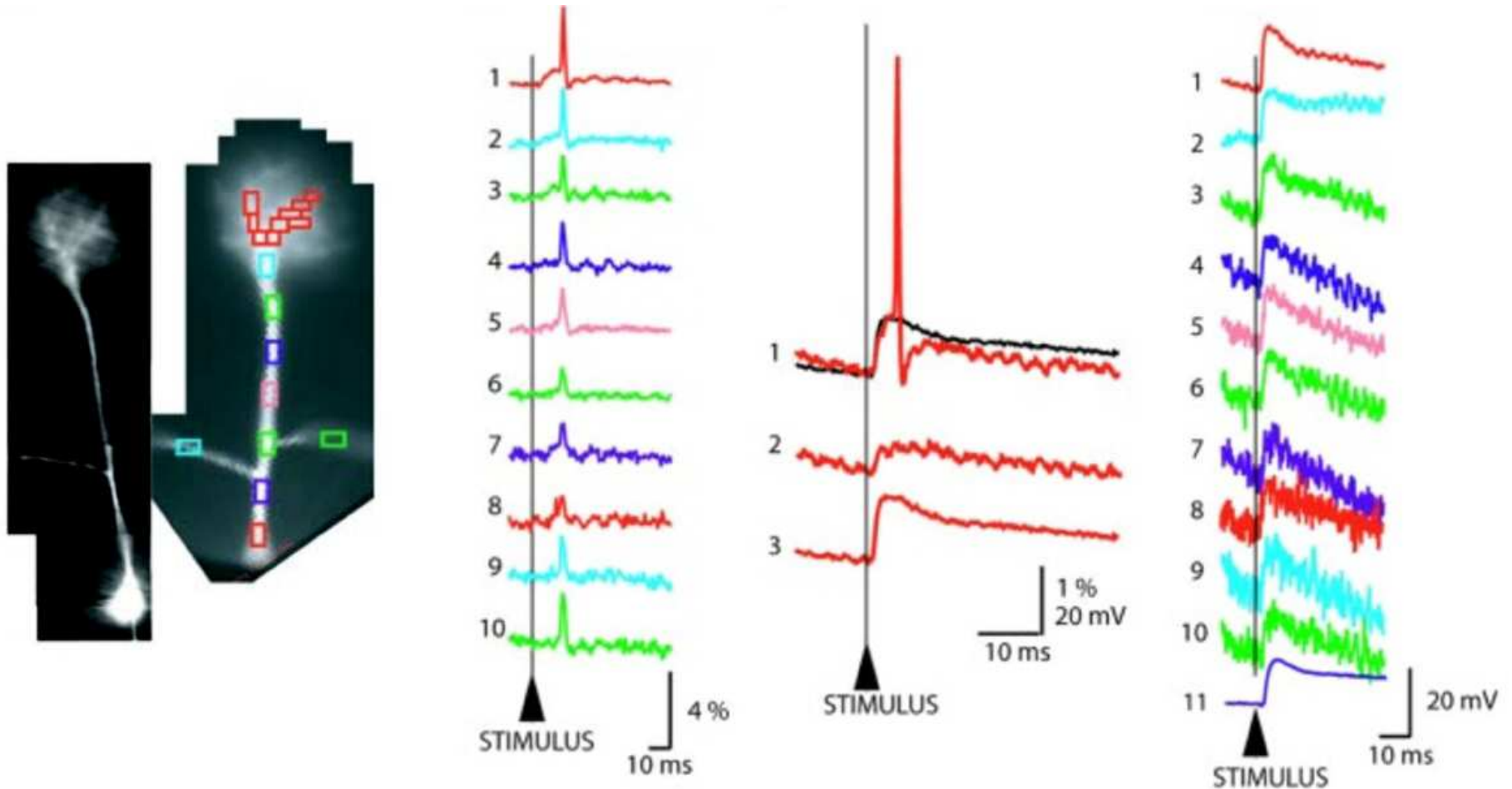
- “Subcellular” level: measurements of intracellular voltage or ionic concentrations (intracellular “patch” electrodes, two-photon imaging)
- “Circuit” level: electrical activity of single neurons or small groups of isolated neurons (multi-electrode recordings, calcium-sensitive microscopy)
- “Systems” level: blood flow or other indirect measurements of electrical activity in coarsely-defined brain areas (fMRI, EEG, MEG...)

# Three challenges

1. Reconstructing the full spatiotemporal voltage on a dendritic tree given noisy, intermittently-sampled subcellular measurements
2. Decoding behaviorally-relevant information from multiple spike trains
3. Inferring connectivity from large populations of noisily-observed spike trains

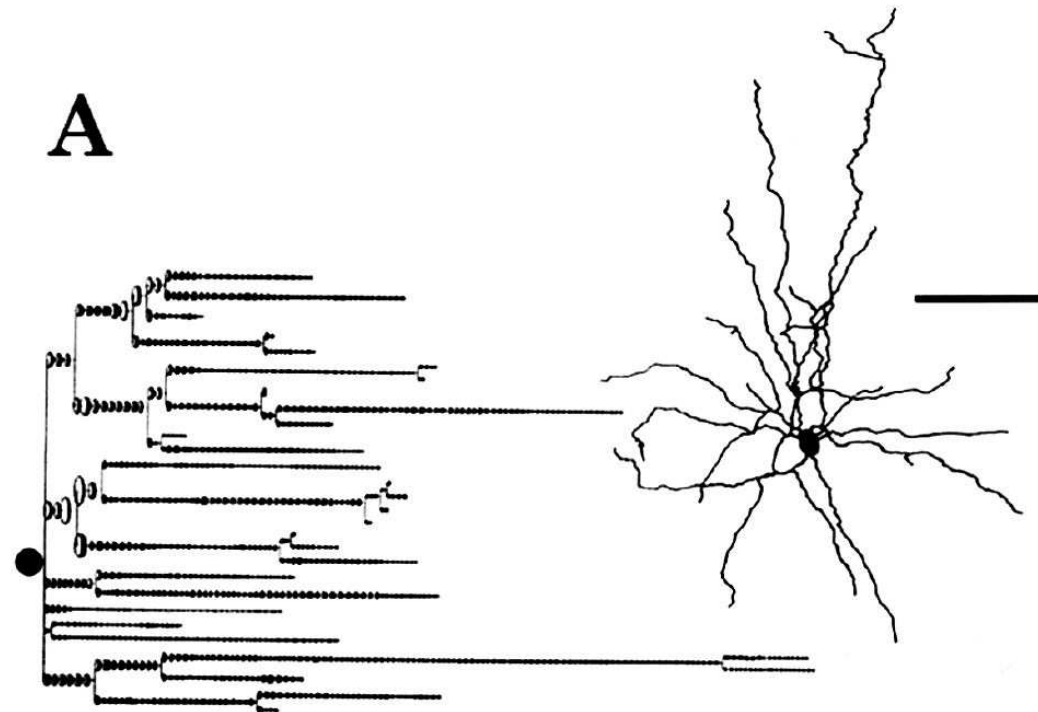
# The filtering problem

Spatiotemporal imaging data opens an exciting window on the computations performed by single neurons, but we have to deal with noise and intermittent observations.



(Djurisic et al., 2004; Knopfel et al., 2006)

# Basic paradigm: compartmental models



- write neuronal dynamics in terms of equivalent nonlinear, time-varying RC circuits (Koch, 1999)
- leads to a coupled system of stochastic differential equations

# Inference of spatiotemporal neuronal state given noisy observations

State-space approach:  $q_t$  = state of neuron at time  $t$ .

We want  $p(q_t|Y_{1:t}) \propto p(q_t, Y_{1:t})$ . Markov assumption:

$$p(Q, Y) = p(Q)p(Y|Q) = p(q_1) \left( \prod_{t=2}^T p(q_t|q_{t-1}) \right) \left( \prod_{t=1}^T p(y_t|q_t) \right)$$

To compute  $p(q_t, Y_{1:t})$ , just recurse

$$p(q_t, Y_{1:t}) = p(y_t|q_t) \int_{q_{t-1}} p(q_t|q_{t-1})p(q_{t-1}, Y_{1:t-1})dq_{t-1}.$$

Linear-Gaussian case: requires  $O(\dim(q)^3T)$  time; in principle, just matrix algebra (Kalman filter). Approximate solutions in more general case via sequential Monte Carlo (Huys and Paninski, 2009).

Major challenge:  $\dim(q)$  can be  $\approx 10^4$  or greater.

# Low-rank approximations

Key fact: current experimental methods provide just a few low-SNR observations per time step.

Basic idea: if dynamics are approximately linear and time-invariant, we can approximate Kalman covariance  $C_t = \text{cov}(q_t|Y_{1:t})$  as  $C_0 + U_t D_t U_t^T$ , with  $C_0 = \lim_{t \rightarrow \infty} \text{cov}(q_t)$ .

$C_0$  is the solution to a Lyapunov equation. It turns out that we can solve linear equations involving  $C_0$  in  $O(\text{dim}(q))$  time via Gaussian belief propagation (Shental et al., 2008), using the fact that the dendrite is a tree.

The necessary recursions here — i.e., updating  $U_t, D_t$  and the Kalman mean  $E(q_t|Y_{1:t})$  — involve linear manipulations of  $C_0$ , using

$$\begin{aligned} C_t &= [(AC_{t-1}A^T + Q)^{-1} + B_t]^{-1} \\ C_0 + U_t D_t U_t^T &= ([A(C_0 + U_{t-1} D_{t-1} U_{t-1}^T)A^T + Q]^{-1} + B_t)^{-1}, \end{aligned}$$

and can be done in  $O(\text{dim}(q))$  time (Paninski, 2009).



# Example: inferring voltage from subsampled observations

(Loading low-rank-speckle.mp4)

# Example: summed observations

(Loading low-rank-horiz.mp4)

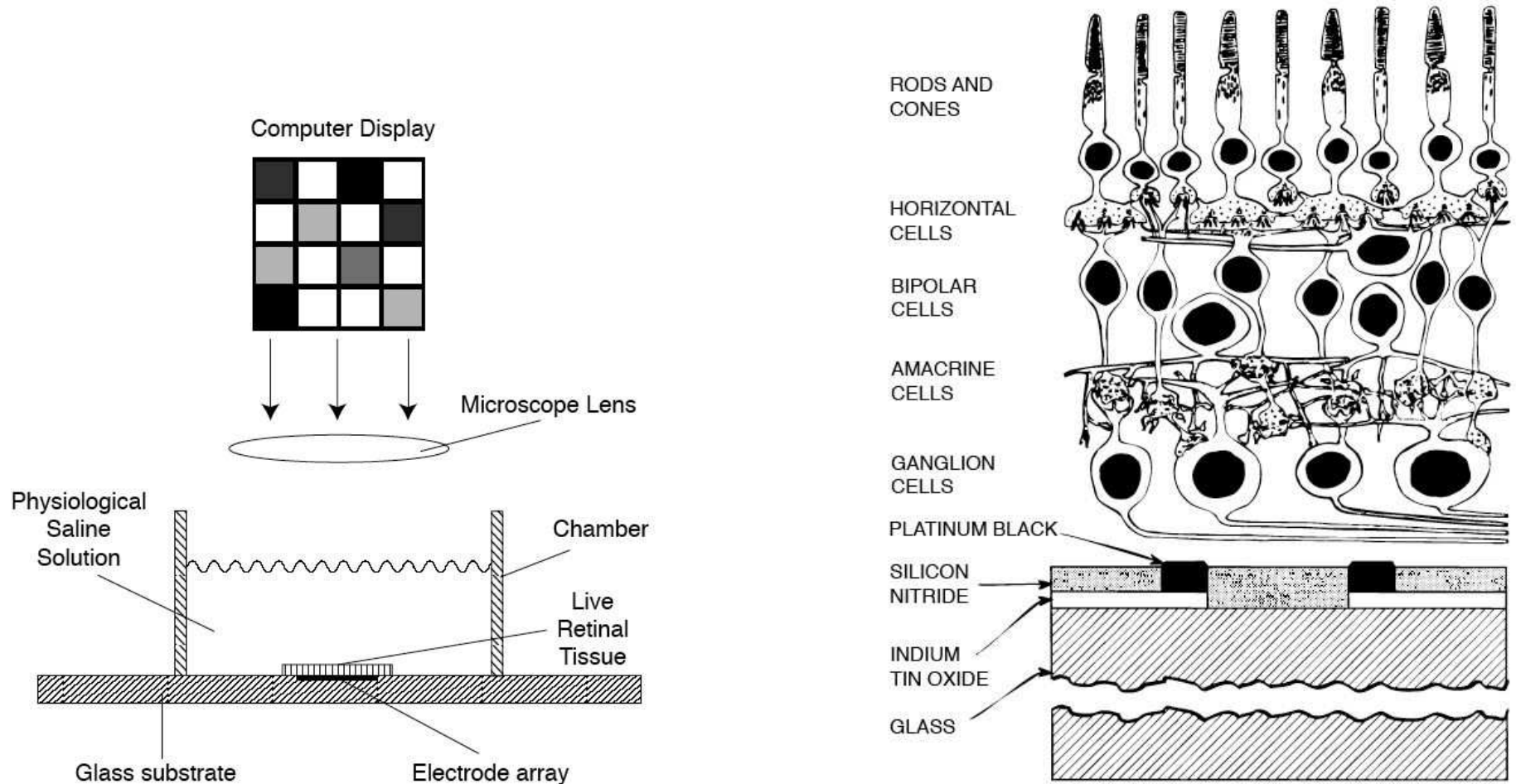
# Open challenges

- Application to real data
- Extension to strongly nonlinear case
- Other (non-neural) applications?

# Part 2: optimal decoding of spike train data

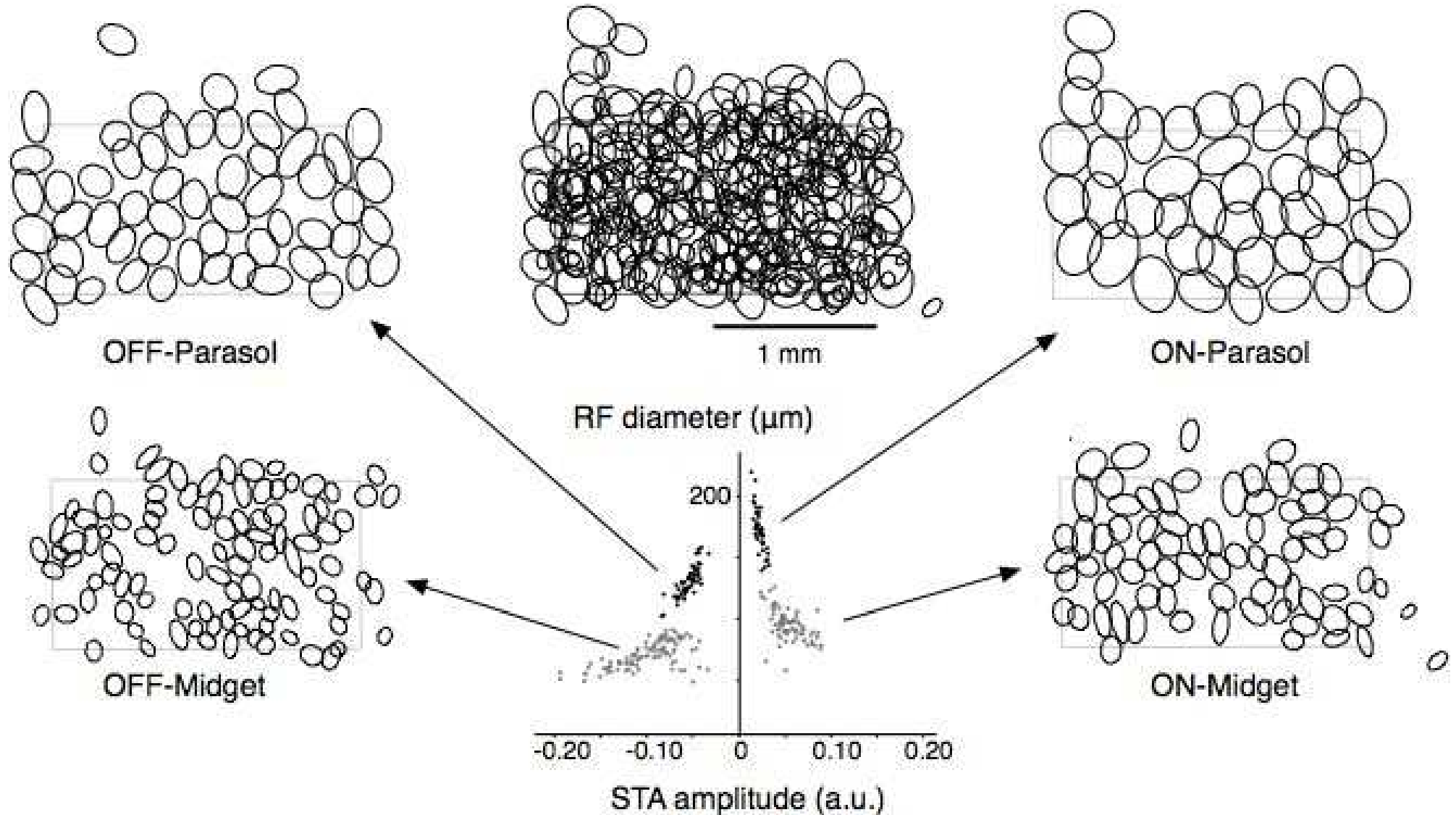
Preparation: macaque retina *in vitro* (Litke et al., 2004)

— extracellularly-recorded responses of populations of ganglion cells

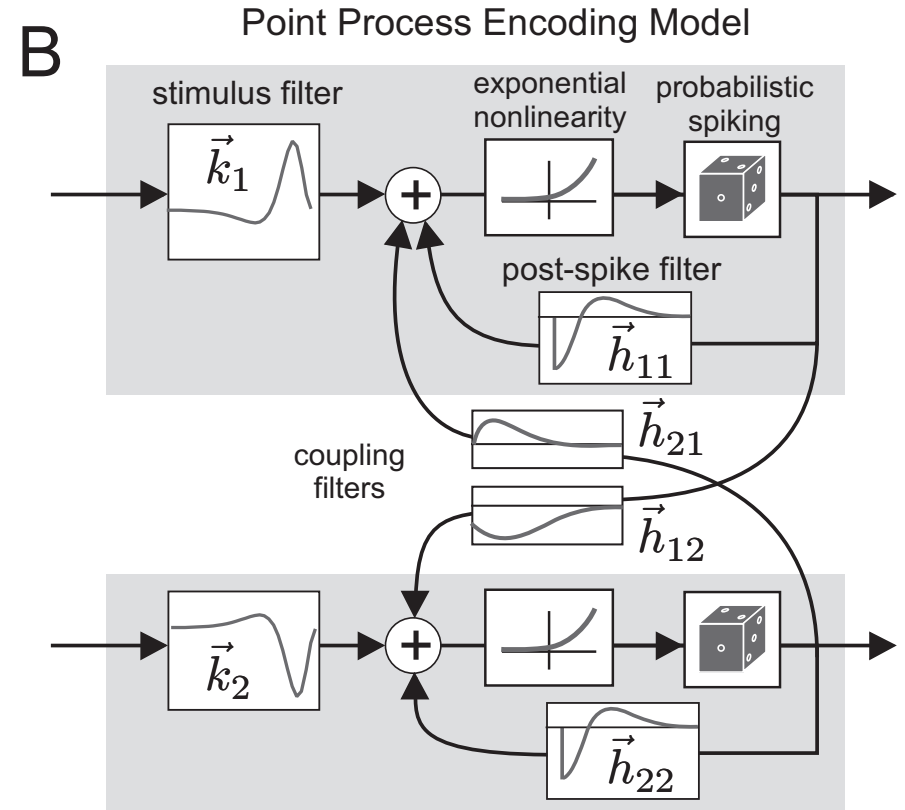
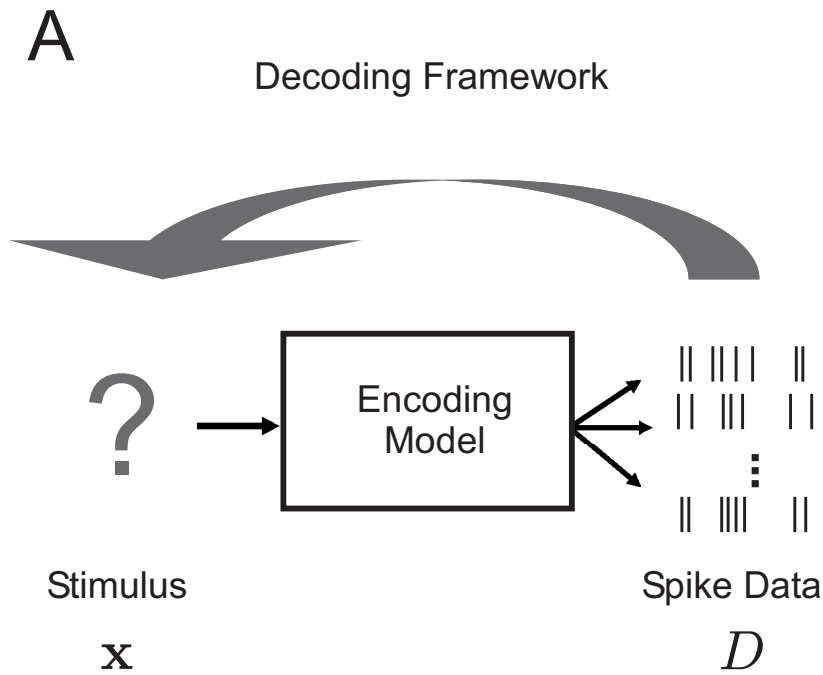


Full control over input; complete observation of output

# Receptive fields tile visual space



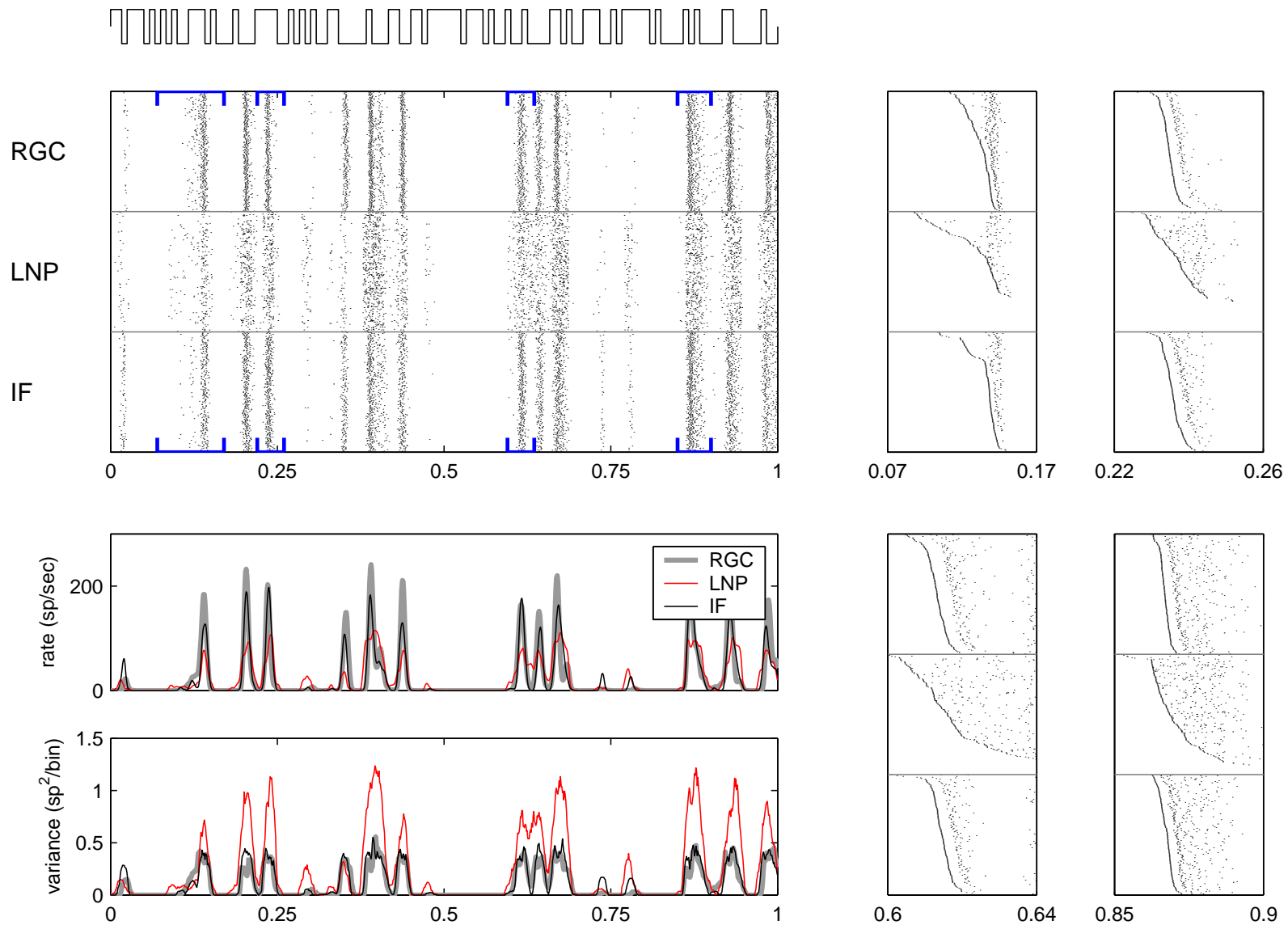
# Multineuronal point-process GLM



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

Fit by  $L_1$ -penalized maximum likelihood (parallel, concave optimization)  
 (Brillinger, 1988; Paninski et al., 2007; Pillow et al., 2008)

# Model captures spike timing precision



(Pillow et al., 2005)

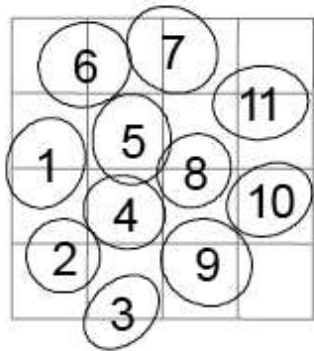
# Model captures cross-correlations

x-corrs:

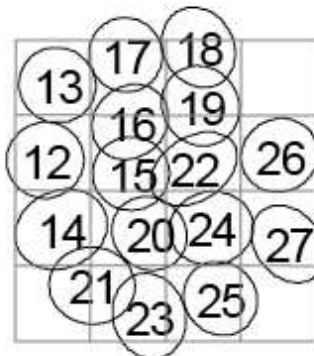
ON-ON

OFF-OFF

ON cells

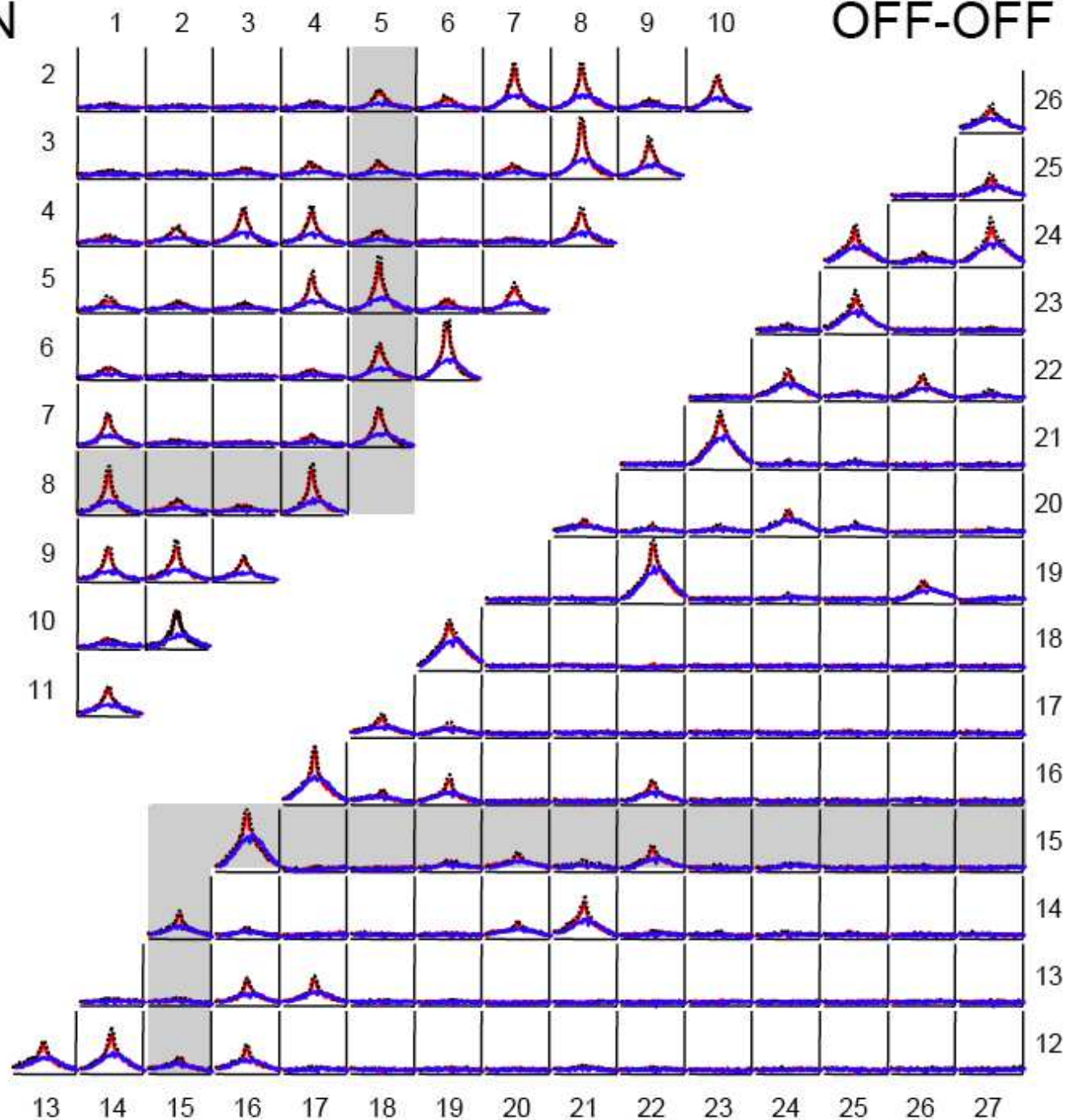


OFF cells



75 sp/s

50 ms





# MAP stimulus decoding

It is reasonable to estimate the movie  $X$  that led to a response  $Y$  via the MAP

$$\hat{X} = \arg \max_X p(X|Y).$$

(Note that  $X$  is very high-dimensional!) For this model, we have:

$$\begin{aligned} \log p(X|Y) &= \log p(X) + \log p(Y|X) + \text{const.} \\ &= \log p(X) + \sum_t \log p(y_t|X, Y_{\dots, t-1}) + \text{const.} \end{aligned}$$

Two basic observations:

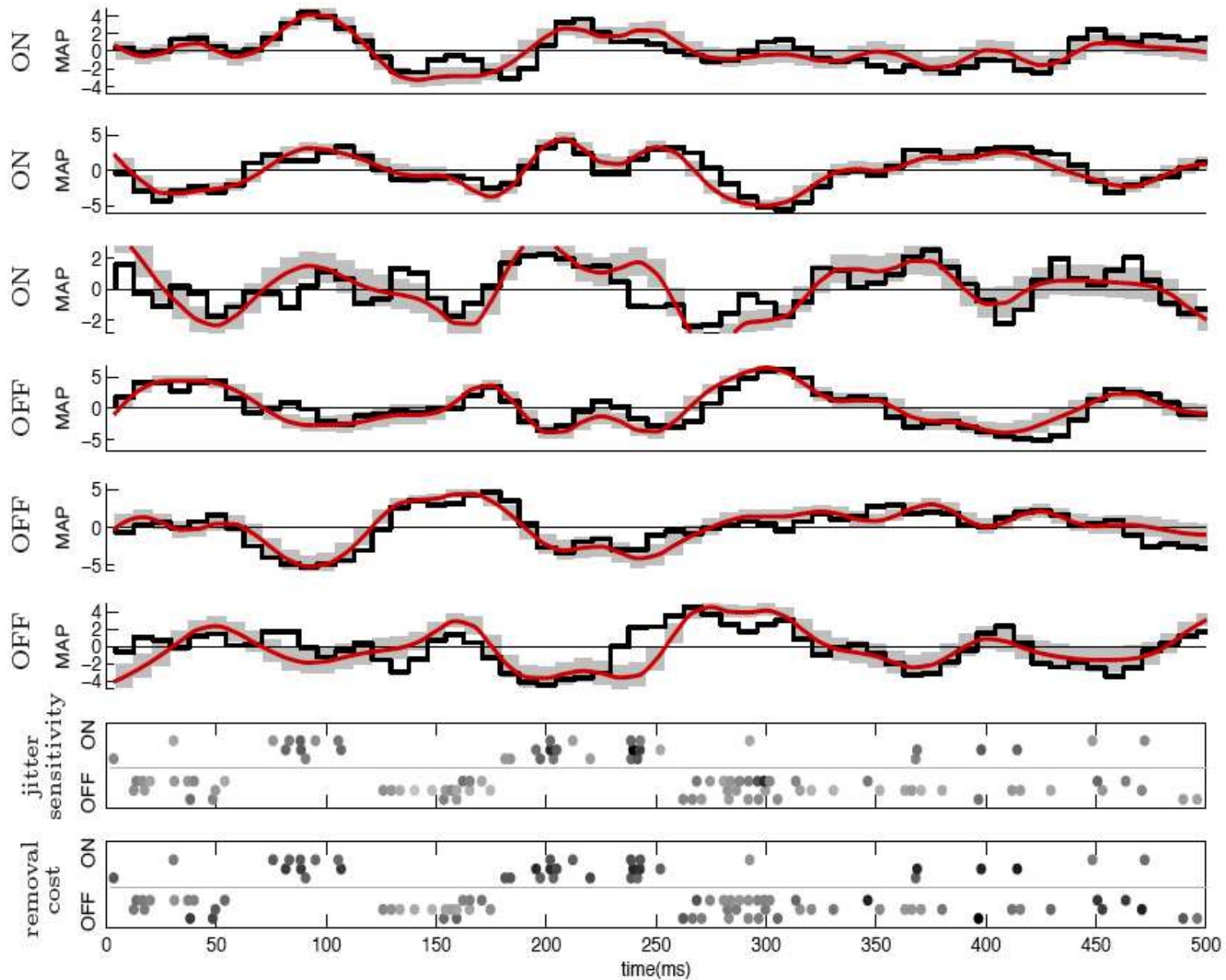
- If  $\log p(X)$  is concave, then so is  $\log p(X|Y)$ , since each  $\log p(y_t|X, Y_{\dots, t-1})$  is.
- Hessian  $H$  of  $\log p(Y|X)$  w.r.t.  $X$  is banded: each  $p(y_t|X, Y_{\dots, t-1})$  depends on  $X$  locally in time, and therefore contributes a banded term.

Newton's method iteratively solves  $HQ_{dir} = \nabla$ . Solving banded systems requires  $O(T)$  time, so computing MAP requires  $O(T)$  time if log-prior is concave with a banded Hessian (Fahrmeir and Kaufmann, 1991;

Davis and Rodriguez-Yam, 2005; Jungbacker and Koopman, 2007).

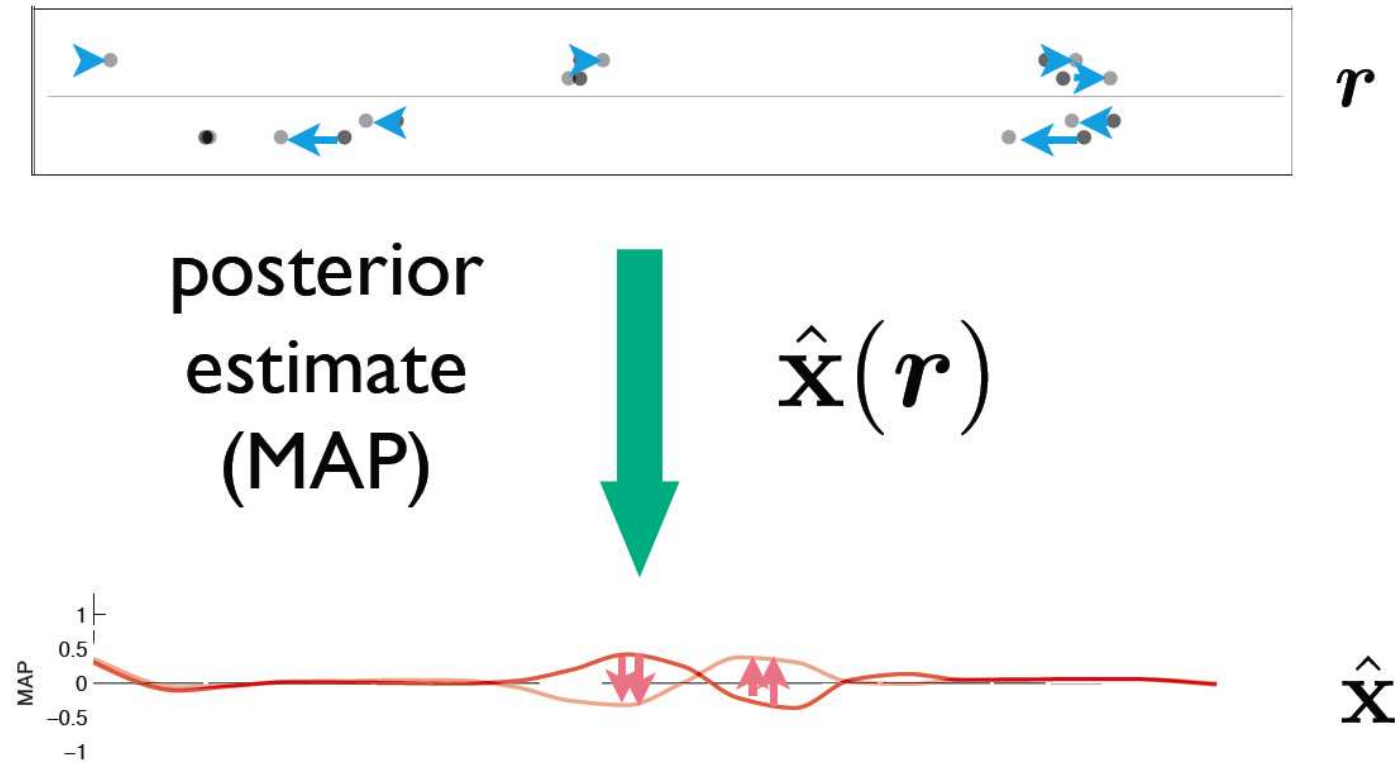
— similar speedups available in constrained cases (Paninski et al., 2009), and in MCMC setting (e.g., fast hybrid Monte Carlo methods (Ahmadian et al., 2009)).

# Application: how important is timing?



— Fast decoding methods let us look more closely (Ahmadian et al., 2009)

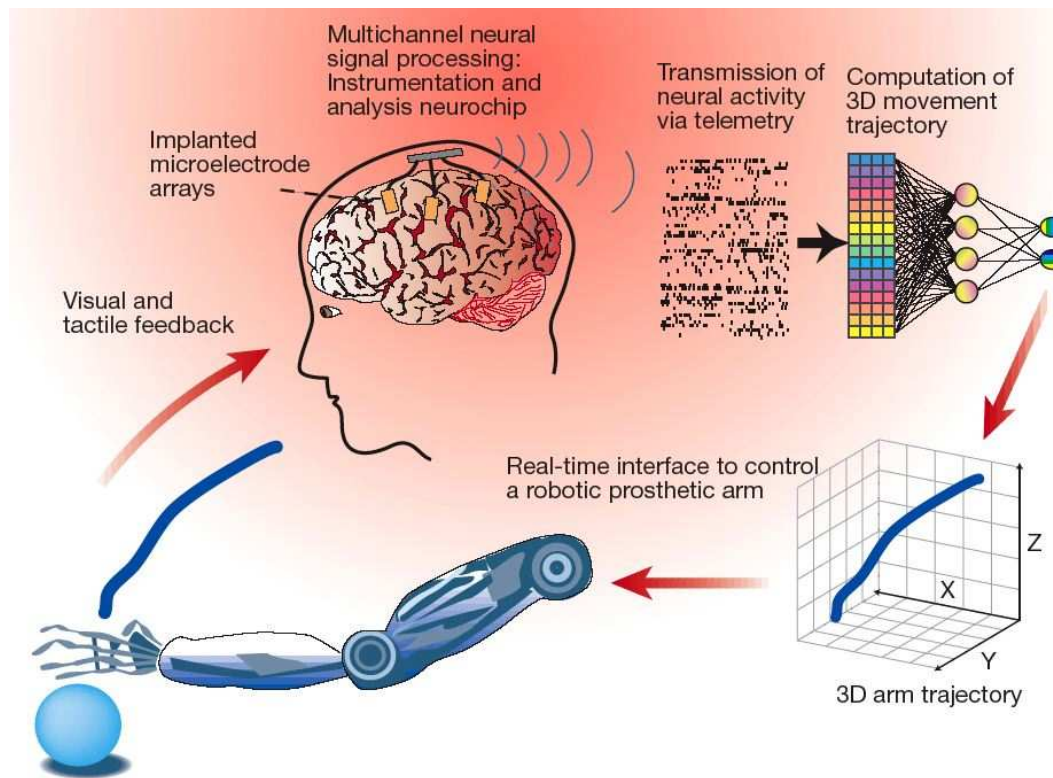
# Constructing a metric between spike trains



$$d(r_1, r_2) \equiv d_x(\hat{x}(r_1), \hat{x}(r_2))$$

Locally,  $d(r, r + \delta r) = \delta r^T G_r \delta r$ : interesting information in  $G_r$ .

# Real-time application: neural prosthetics

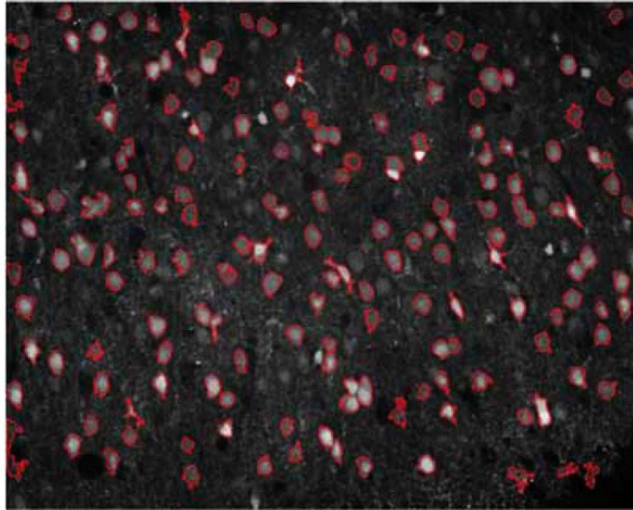


(Loading decoding.mp4)

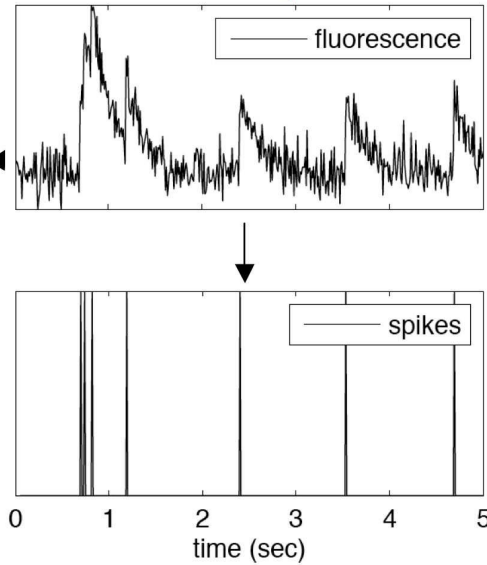
(Donoghue, 2002; Wu et al., 2006; Wu et al., 2009)

# Part 3: circuit inference

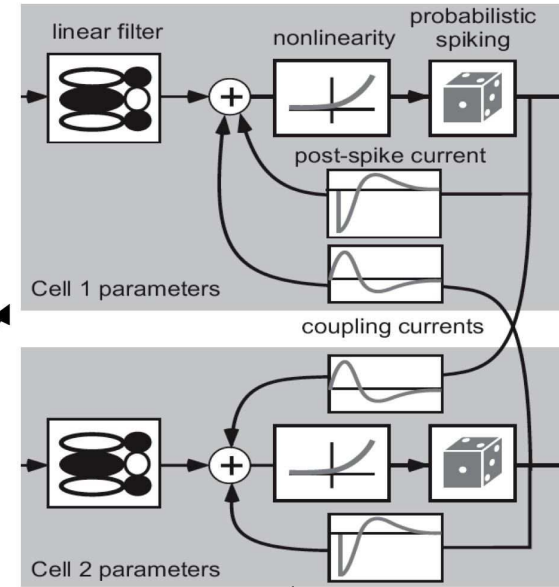
Record large-scale calcium movie



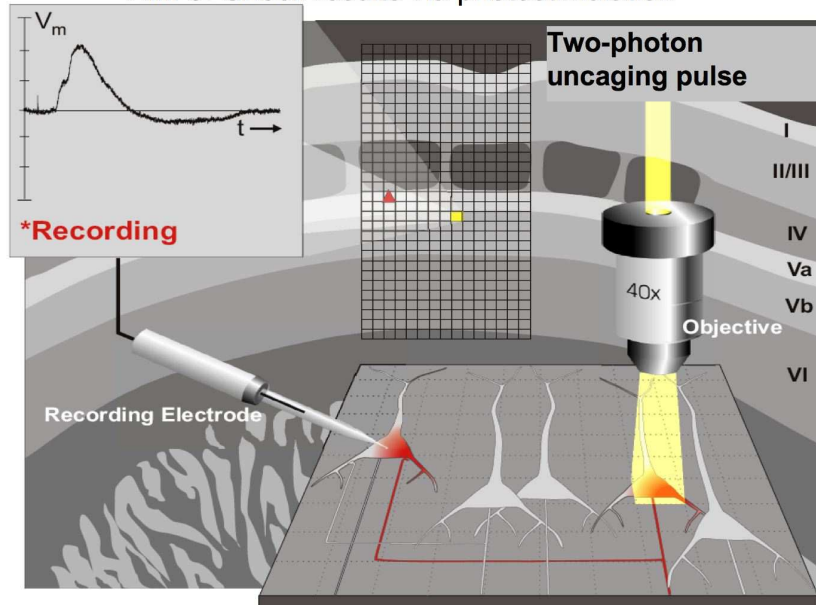
Aim 1: Extract spike times



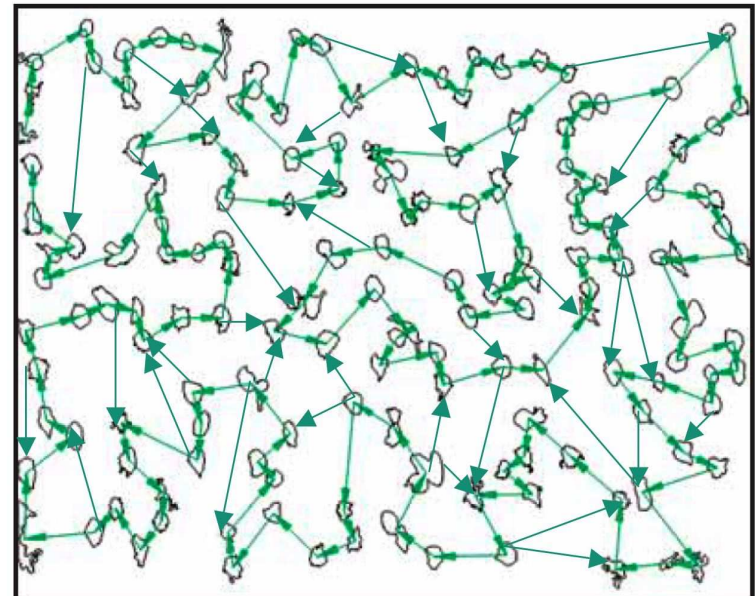
Aim 2: Estimate network model



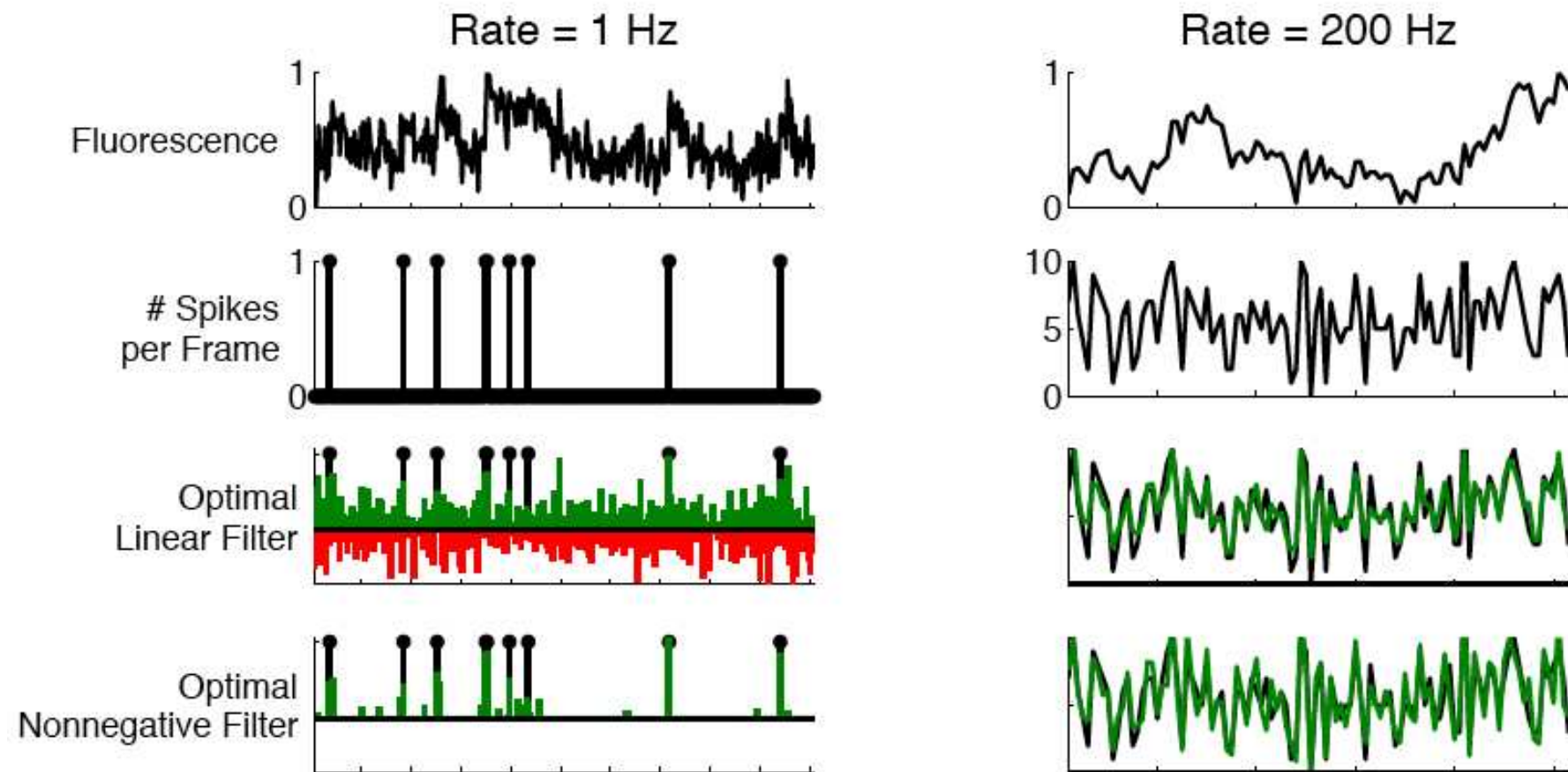
Aim 3: Check results via photostimulation



Inferred network model



# Challenge: slow, noisy calcium data

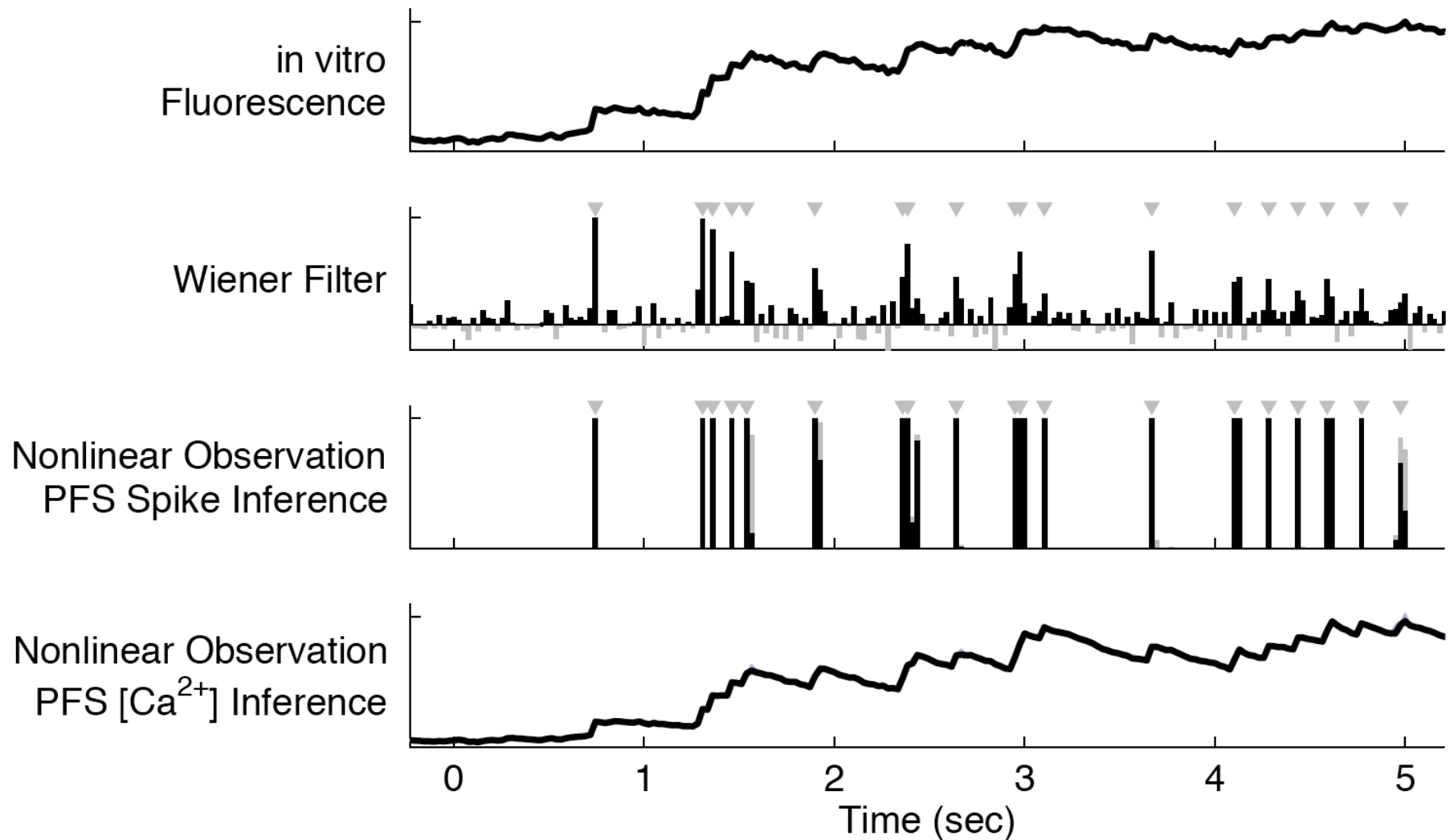


First-order model:

$$C_{t+dt} = C_t - dtC_t/\tau + N_t; \quad N_t > 0; \quad y_t = C_t + \epsilon_t$$

—  $\tau \approx 100$  ms; nonnegative deconvolution problem. Can be solved by  $O(T)$  relaxed constrained optimization methods (Vogelstein et al., 2008) or sequential Monte Carlo (Vogelstein et al., 2009).

# Particle filter can extract spikes from saturated recordings



— saturation model:  $y_t = g(C_t) + \epsilon_t$  (Vogelstein et al., 2009)

# Monte Carlo EM approach

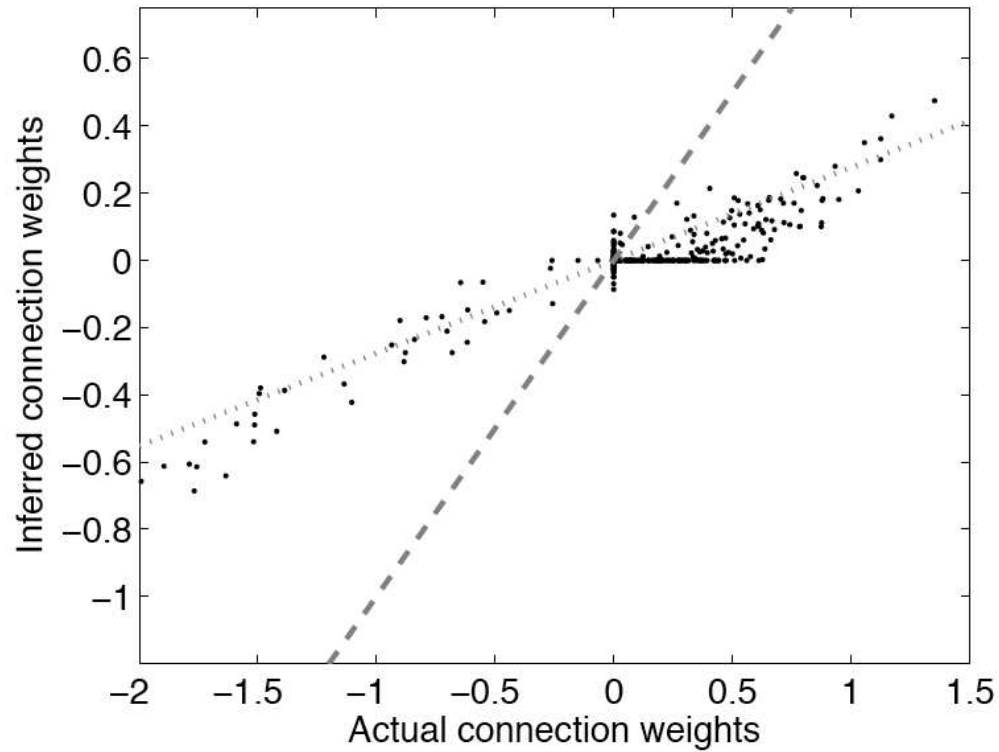
Given the spike times in the network,  $L_1$ -penalized likelihood optimization is easy. But we only have noisy calcium observations  $Y$ ; true spike times are hidden variables. Thus an EM approach is natural.

- E step: sample spike trains  $Z$  from  $p(Z|Y, \theta)$
- M step: given sampled spike trains, perform  $L_1$ -penalized likelihood optimization to update parameters  $\theta$ .

E step is hard part here. Use the fact that neurons interact fairly weakly; thus we need to sample from a collection of weakly-interacting Markov chains. Efficient Metropolis-within-blockwise-Gibbs forward-backward methods (Neal et al., 2003).



# Simulated circuit inference



Good news: connections are inferred with the correct sign. But process is slow; improved sampling methods would have a major impact.

# Optimal control of spike timing

Optimal experimental design and neural prosthetics applications require us to perturb the network at will. How can we make a neuron fire exactly when we want it to?

Assume bounded inputs; otherwise problem is trivial.

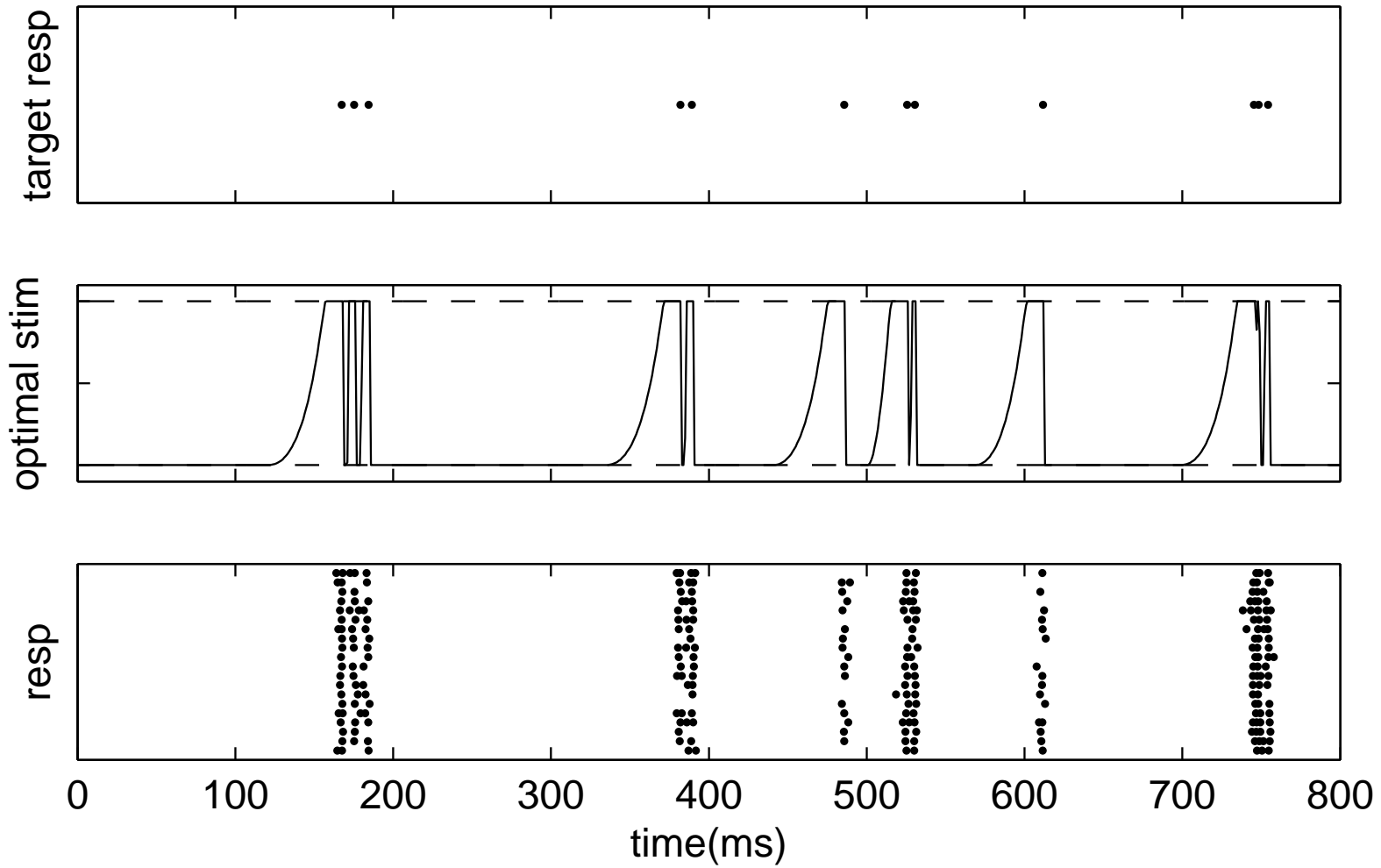
Start with a simple model:

$$\lambda_t = f(\vec{k} * I_t + h_t).$$

Now we can just optimize the likelihood of the desired spike train, as a function of the input  $I_t$ , with  $I_t$  bounded.

Concave objective function over convex set of possible inputs  $I_t$   
+ Hessian is banded  $\implies O(T)$  optimization.

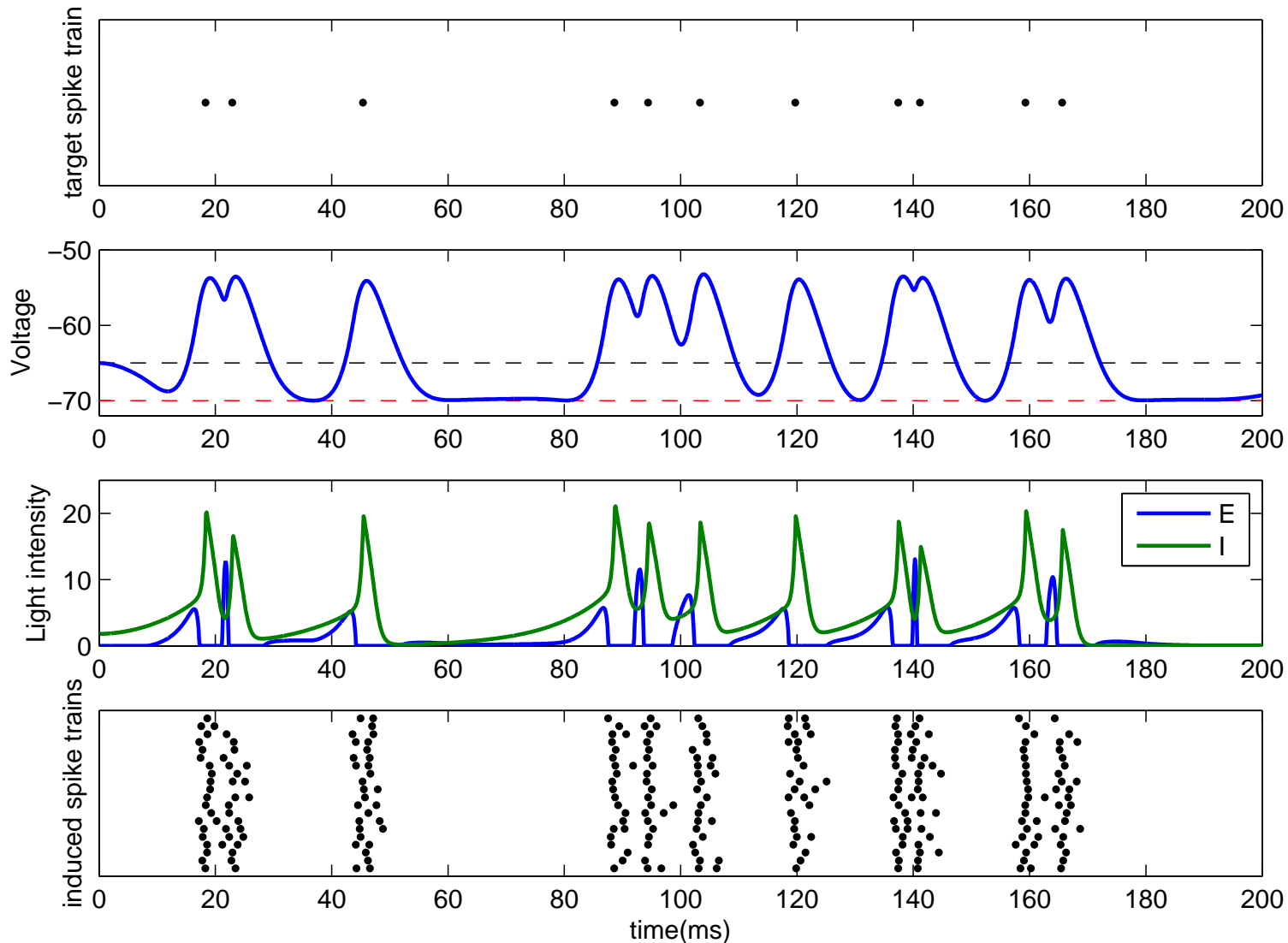
# Optimal electrical control of spike timing



# Optical conductance-based control of spiking

$$V_{t+dt} = V_t + dt \left( -gV_t + g_t^i(V^i - V_t) + g_t^e(V^e - V_t) \right) + \sqrt{dt}\sigma\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1)$$

$$g_{t+dt}^i = g_t^i + dt \left( -\frac{g_t^i}{\tau_i} + a_{ii}L_t^i + a_{ie}L_t^e \right); \quad g_{t+dt}^e = g_t^e + dt \left( -\frac{g_t^e}{\tau_i} + a_{ee}L_t^e + a_{ei}L_t^i \right)$$



# Conclusions

- GLM and state-space approaches provide flexible, powerful methods for answering key questions in neuroscience
- Close relationships between encoding, decoding, and experimental design (Paninski et al., 2007)
- Log-concavity, banded matrix methods make computations very tractable
- Experimental methods progressing rapidly; many new challenges and opportunities for applications of statistical ideas

# References

- Ahmadian, Y., Pillow, J., and Paninski, L. (2009). Efficient Markov Chain Monte Carlo methods for decoding population spike trains. *Under review, Neural Computation*.
- Brillinger, D. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cyberkinetics*, 59:189–200.
- Davis, R. and Rodriguez-Yam, G. (2005). Estimation for state-space models: an approximate likelihood approach. *Statistica Sinica*, 15:381–406.
- Djurisic, M., Antic, S., Chen, W. R., and Zecevic, D. (2004). Voltage imaging from dendrites of mitral cells: EPSP attenuation and spike trigger zones. *J. Neurosci.*, 24(30):6703–6714.
- Donoghue, J. (2002). Connecting cortex to machines: recent advances in brain interfaces. *Nature Neuroscience*, 5:1085–1088.
- Fahrmeir, L. and Kaufmann, H. (1991). On Kalman filtering, posterior mode estimation and fisher scoring in dynamic exponential family regression. *Metrika*, 38:37–60.
- Huys, Q. and Paninski, L. (2009). Model-based smoothing of, and parameter estimation from, noisy biophysical recordings. *PLOS Computational Biology*, 5:e1000379.
- Jungbacker, B. and Koopman, S. (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*, 94:827–839.
- Knopfel, T., Diez-Garcia, J., and Akemann, W. (2006). Optical probing of neuronal circuit dynamics: genetically encoded versus classical fluorescent sensors. *Trends in Neurosciences*, 29:160–166.
- Koch, C. (1999). *Biophysics of Computation*. Oxford University Press.
- Litke, A., Bezayiff, N., Chichilnisky, E., Cunningham, W., Dabrowski, W., Grillo, A., Grivich, M., Grybos, P., Hottowy, P., Kachiguine, S., Kalmar, R., Mathieson, K., Petrusca, D., Rahman, M., and Sher, A. (2004). What does the eye tell the brain? development of a system for the large scale recording of retinal output activity. *IEEE Trans Nucl Sci*, pages 1434–1440.
- Neal, R., Beal, M., and Roweis, S. (2003). Inferring state sequences for non-linear systems with embedded hidden Markov models. *NIPS*, 16.
- Paninski, L. (2009). Fast Kalman filtering on dendritic trees. *In progress*.
- Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama, K., Vidne, M., Vogelstein, J., and Wu, W. (2009). 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.