Statistical methods for understanding neural computation

Liam Paninski

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

October 7, 2009

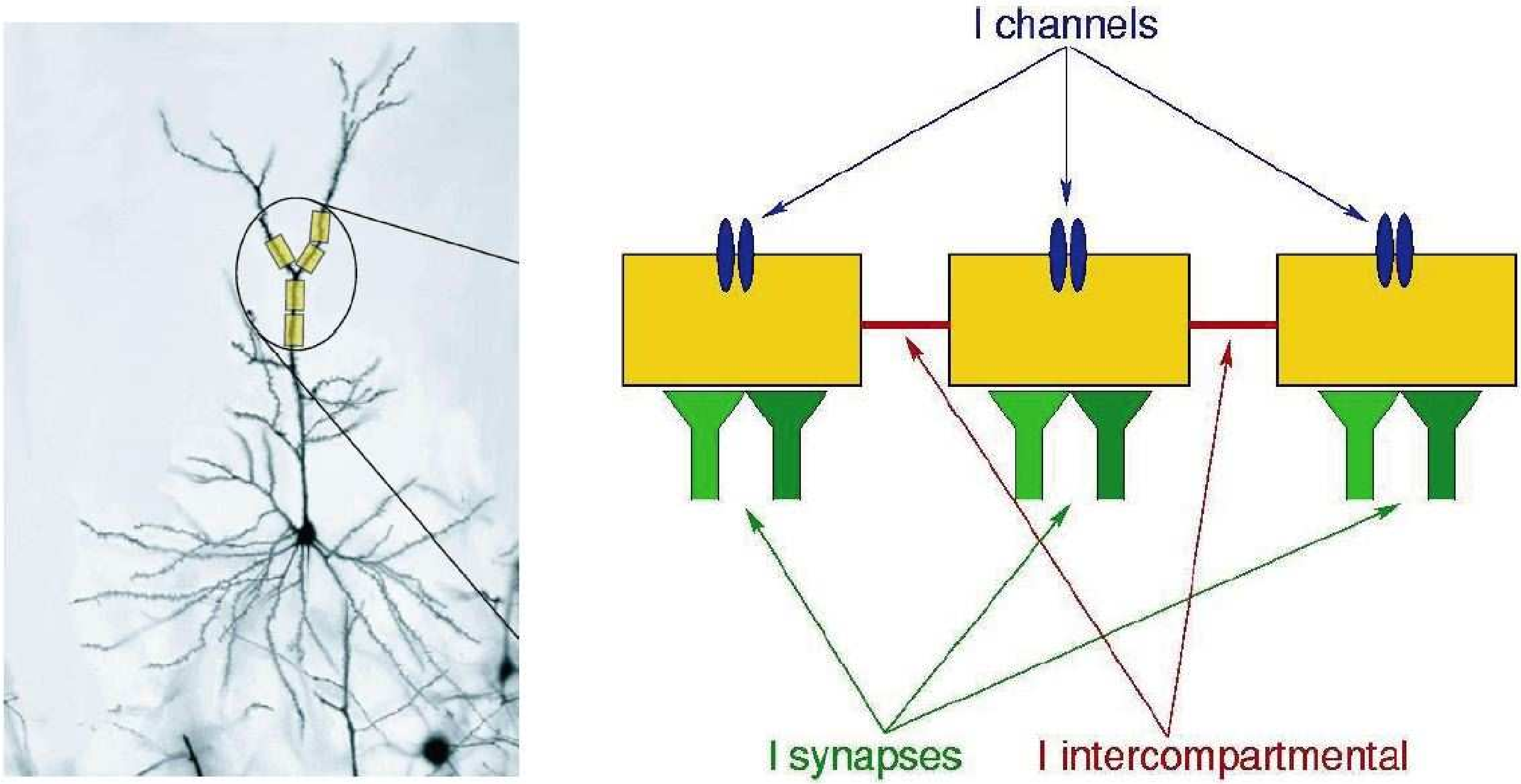
Support: Sloan Fellowship, NIH/NSF CRCNS, NSF CAREER, McKnight Scholar award.

Some exciting open challenges for statistical neuroscience

- 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

An inverse problem: inferring cable equation parameters



Can we recover detailed biophysical properties?

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

Estimating biophysical parameters from $V(x,t)$

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

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

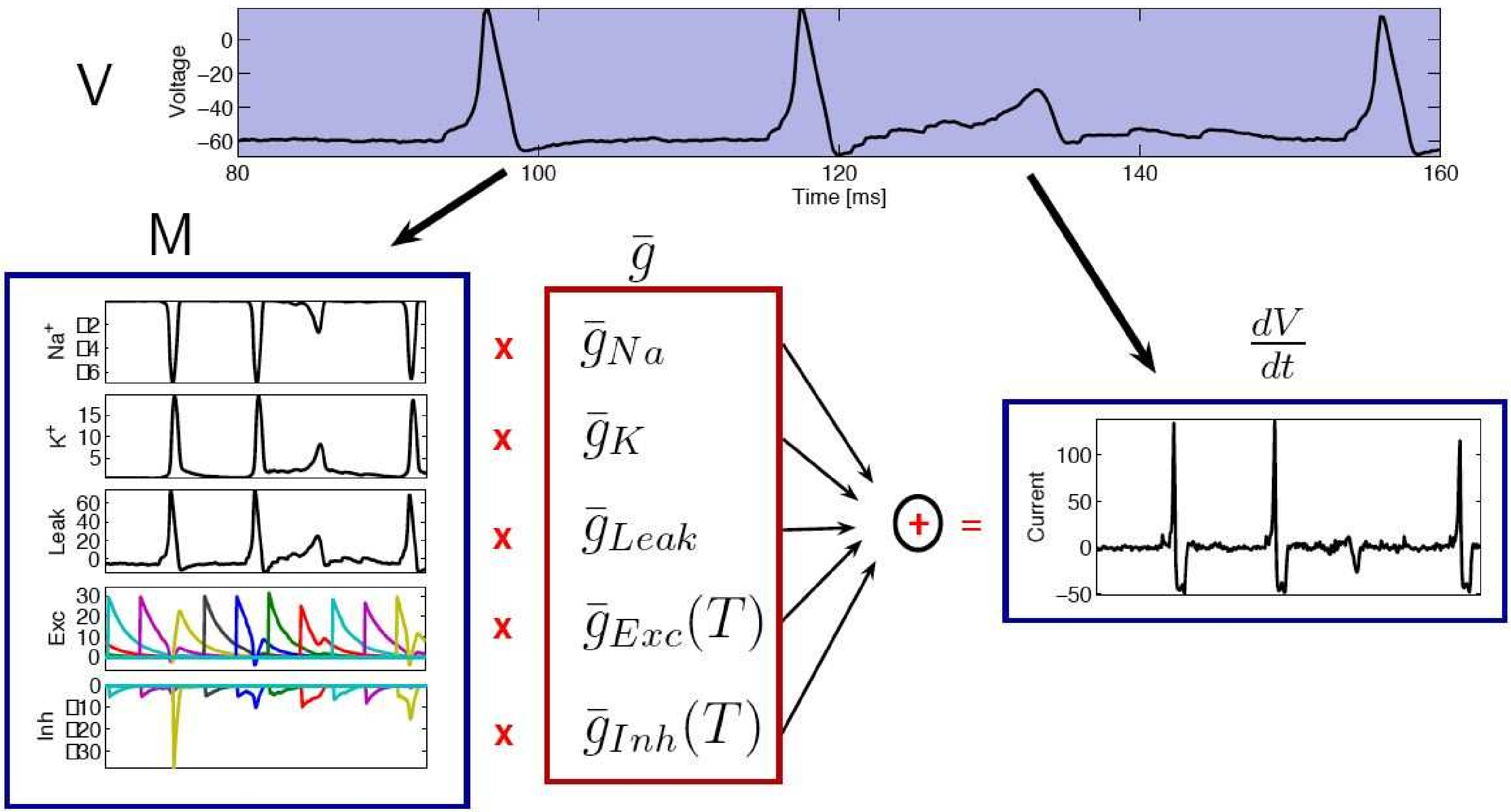
$$I_i^{\text{synapses}} = \sum_s (\xi_s * k_s)(t)(E_s - V_i(t))$$

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

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

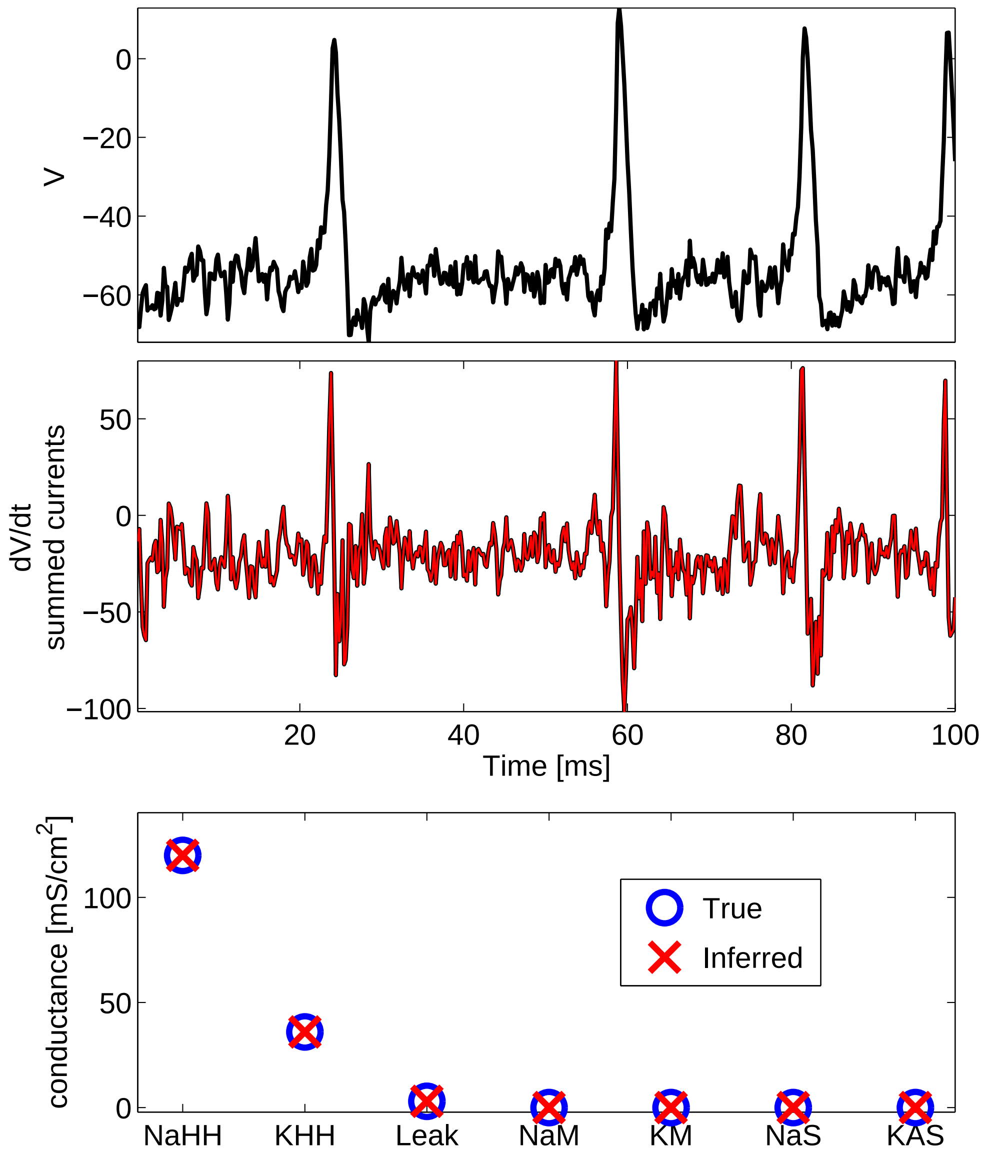
then estimating unknown parameters is standard convex nonnegative regression problem (albeit high-d): $\min_{\theta \geq 0} ||Y - X\theta||^2$.

Estimating channel densities from $V(t)$



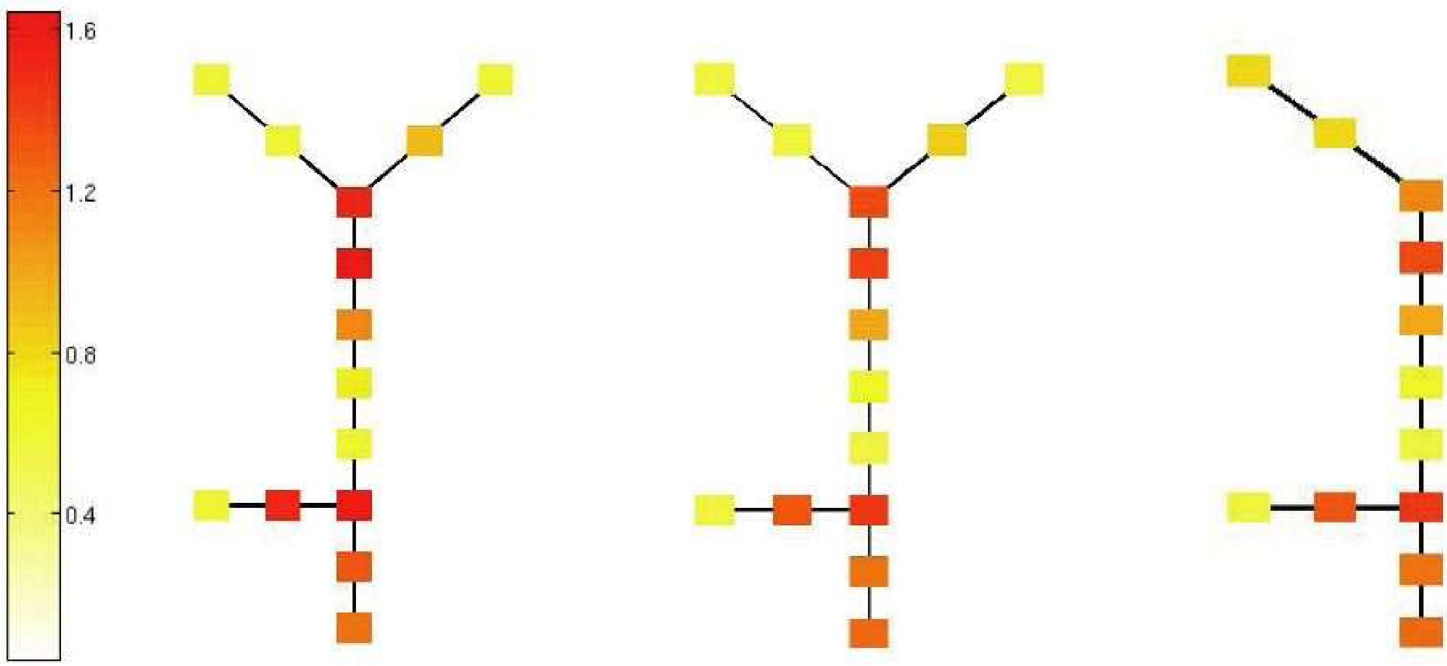
(Huys et al., 2006)

Estimating channel densities from $V(t)$



Estimating non-homogeneous channel densities

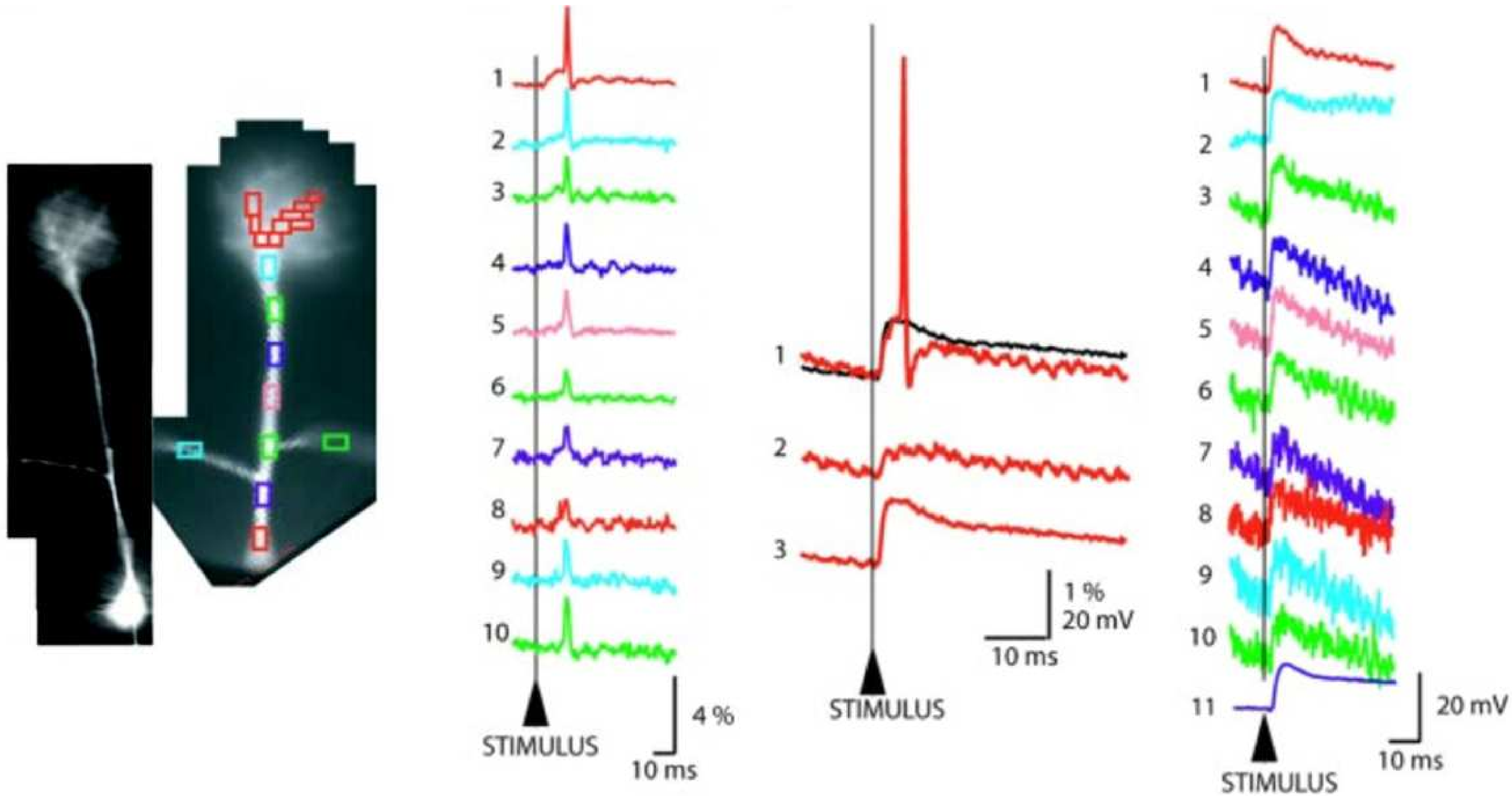
$$I_i^{\text{channels}} = \sum_c \bar{g}_c g_c(t)(E_c - V_i(t))$$



True g_{Na} Estimated g_{Na}

The filtering problem

Spatiotemporal imaging data is very exciting, but we have to deal with noise and intermittent observations.



(Djurisic et al., 2004)

Basic paradigm: the Kalman filter

Variable of interest, q_t, evolves according to a noisy differential equation (Markov process):

$$dq/dt = f(q_t) + \epsilon_t.$$

Make noisy observations:

$$y_t = g(q_t) + \eta_t.$$

We want to infer $E(q_t|Y)$: optimal estimate given observations.

If $f(.)$ and $g(.)$ are linear, and ϵ_t and η_t are Gaussian, then solution is classical: Kalman filter. More general problems: particle filter (Huys and Paninski, 2009).

Basic Kalman filter requires $O(\dim(q)^3 T)$ time. Reduction to $O(qT)$ by exploiting tree structure of dendrite (Paninski, 2009).

Example: inferring voltage from subsampled observations

(Loading low-rank-speckle.mp4)

Example: summed observations

(Loading low-rank-horiz.mp4)

Part 2: Reinterpreting the STRF

Classic method for estimating spectrotemporal receptive field: fit the linear-Gaussian regression model

$$n_t = \vec{k} \cdot \vec{x}_t + \epsilon_t, \;\; \epsilon_t \sim \mathcal{N}(0, \sigma^2).$$

The STRF $\vec{k}$ weights the stimulus $\vec{x}_t$; ϵ_t models variability of response n_t.

Pros:

- analytical solution for optimal $\hat{k}$.
- easy to incorporate prior assumptions on $\vec{k}$ (e.g., smoothness); Bayesian smoothing methods built in to STRFPak.

Part 2: Reinterpreting the STRF

Classic method for estimating spectrotemporal receptive field: fit the linear-Gaussian regression model

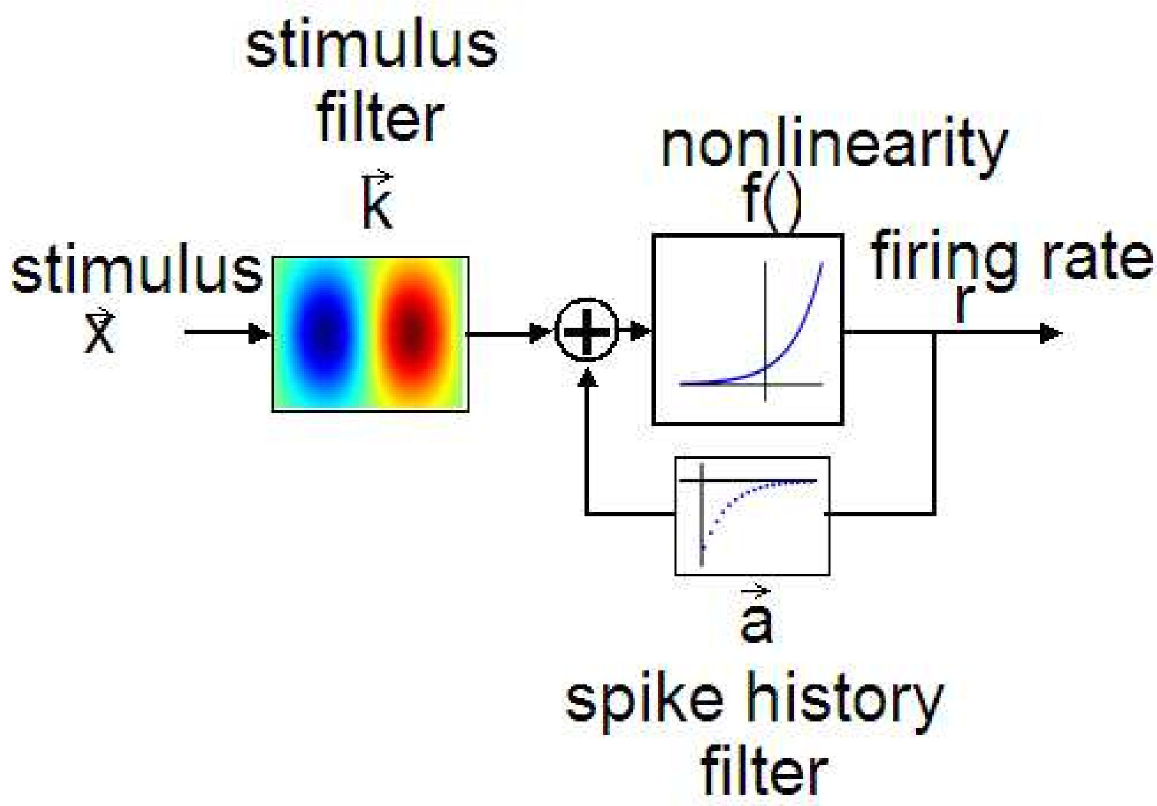
$$n_t = \vec{k} \cdot \vec{x}_t + \epsilon_t, \;\; \epsilon_t \sim \mathcal{N}(0, \sigma^2).$$

The STRF $\vec{k}$ weights the stimulus $\vec{x}_t$; ϵ_t models variability of response n_t.

Cons:

- Gaussian model is not really accurate for spike trains.
- responses n_t can be negative.
- given stimulus $\vec{x}_t$, responses n_t are independent: no refractoriness, burstiness, firing-rate adaptation, etc.

Generalized linear model



$$
\begin{aligned}
p(n_t = 1) &= \lambda_t dt \\
\lambda_t &= f(\vec{k} \cdot \vec{x}_t + \sum_j a_j n_{t-j})
\end{aligned}
$$

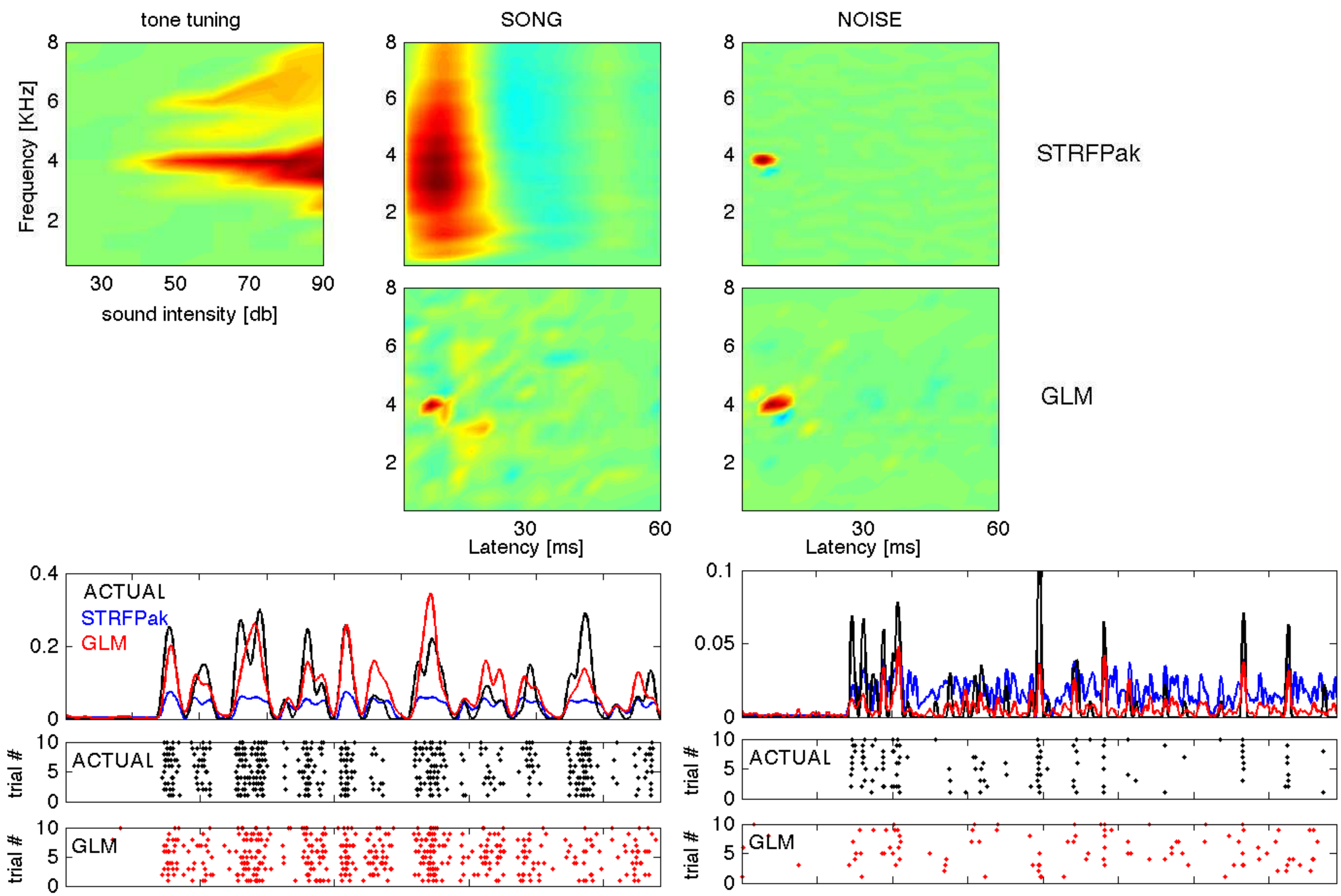
GLM likelihood

$$\lambda_t = f(\vec{k} \cdot \vec{x}_t + \sum_j a_j n_{t-j})$$

Key points:

- f convex and log-concave $\implies$ log-likelihood concave in $\vec{\theta}$. Easy to optimize, so estimating $\hat{\theta}$ is very tractable.
- Easy to include smoothing (as in STRFPak) or sparsening priors.
- Can also include nonlinear terms easily (Gill et al., 2006; Ahrens et al., 2008)

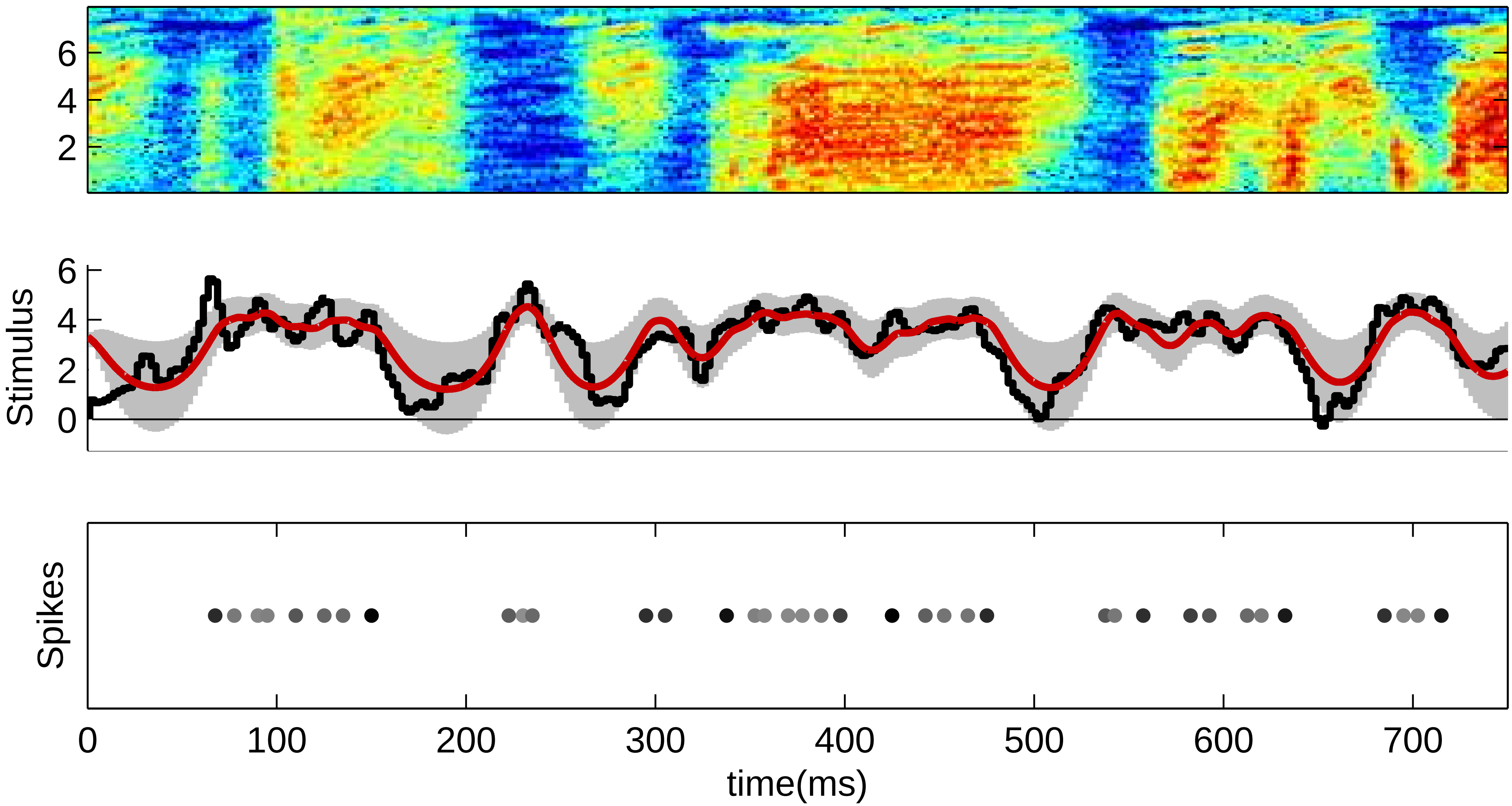
Model performance: zebra finch MLd



(Calabrese, Schneider, Woolley et al. 2009)

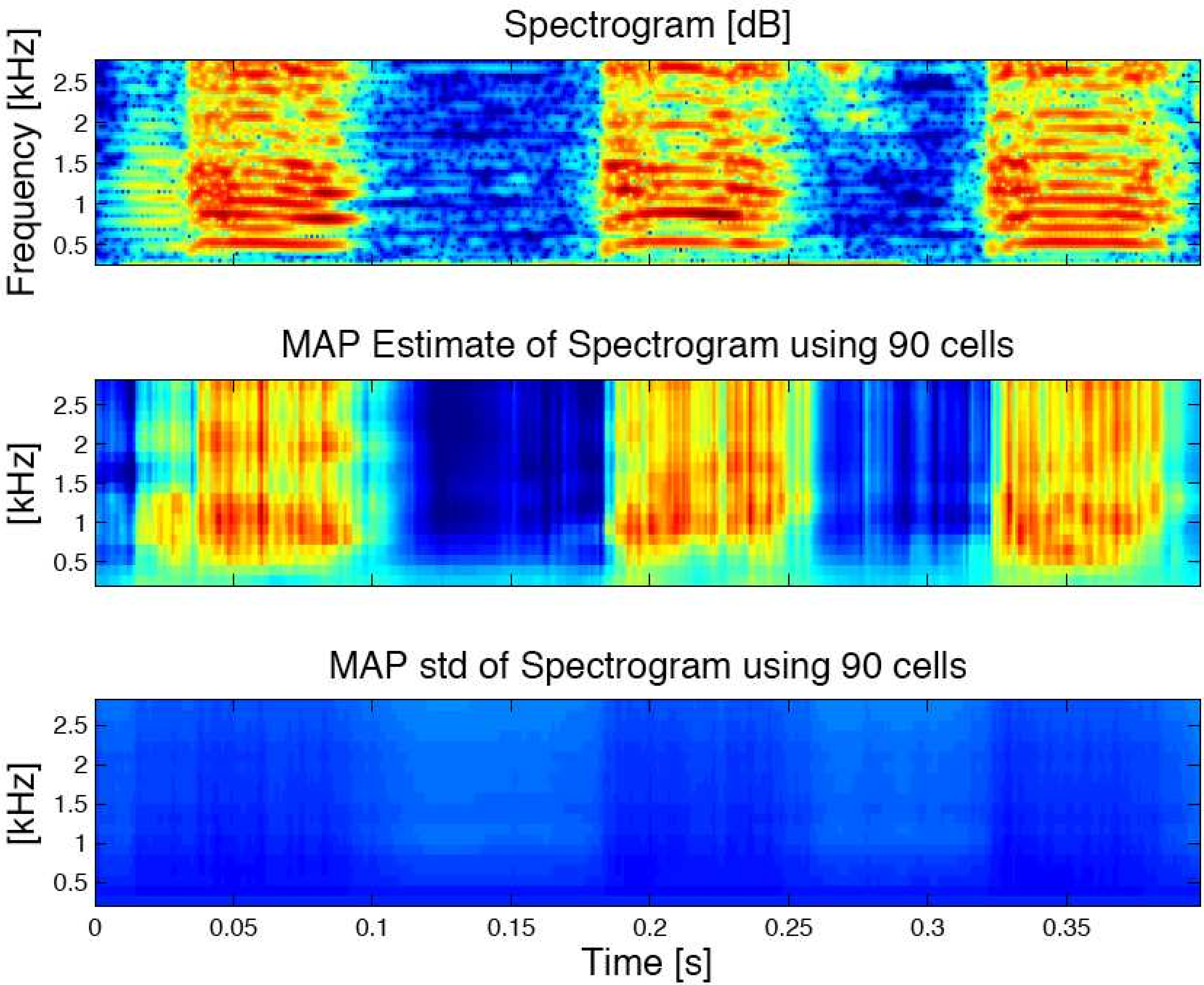
Application: fast optimal decoding

Maximize $\log p(\vec{x}|n, \vec{\theta})$ with respect to $\vec{x}$. Concave optimization.



Can be computed quickly: $O(T)$ time (Ahmadian et al., 2009). Fast decoding enables perturbation analysis: how important is each spike? Leads to decoding-based spike-train metric (Ahmadian et al., 2008).

Decoding a full song



Application: optimal stimulus design

Idea: we have full control over the stimuli we present. Can we choose stimuli $\vec{x}_t$ to maximize the informativeness of each trial?

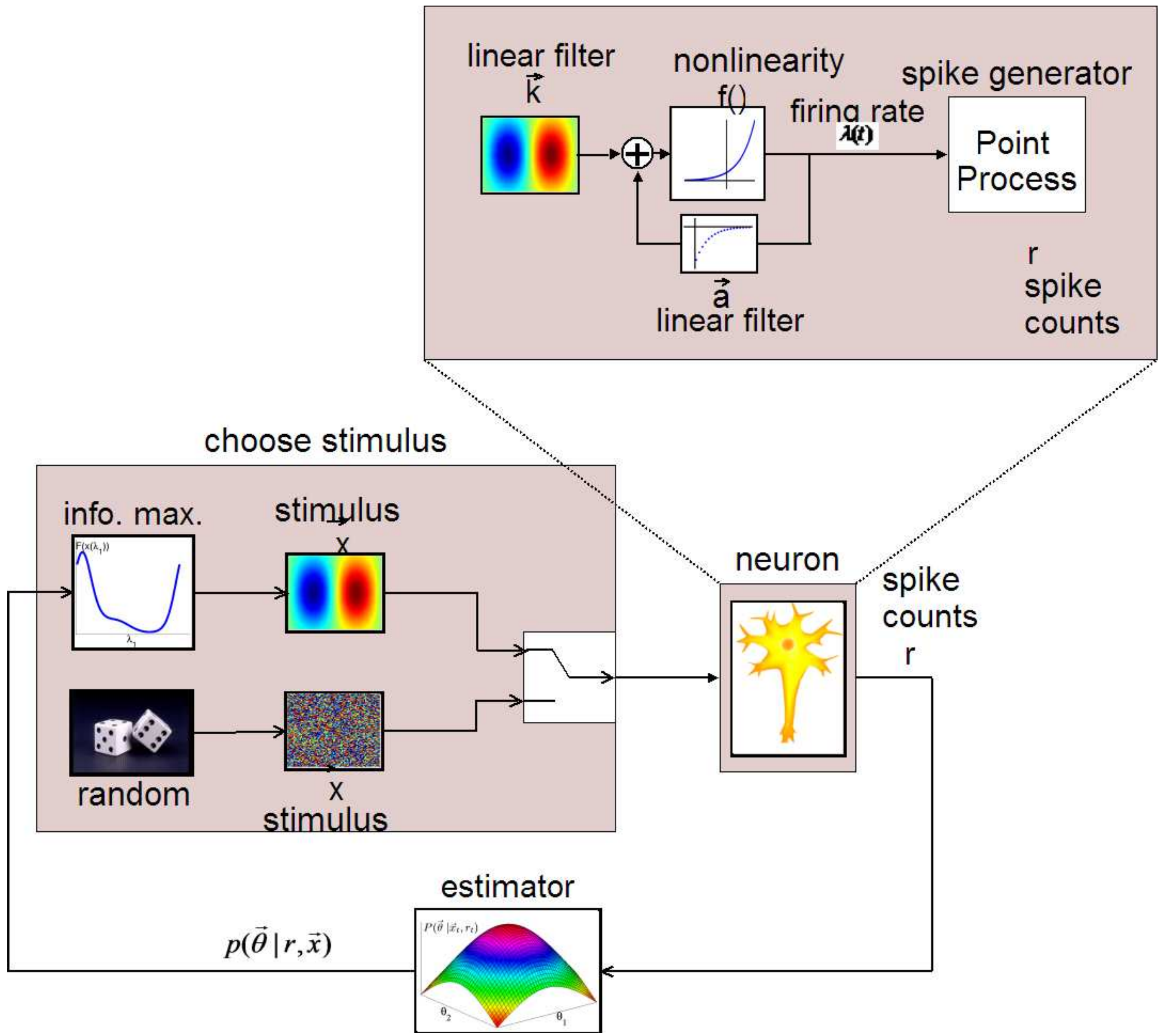
— More quantitatively, optimize $I(n_t; \theta | \vec{x}_t)$ with respect to $\vec{x}_t$.

Maximizing $I(n_t; \theta; \vec{x}_t) \implies$ minimizing uncertainty about θ.

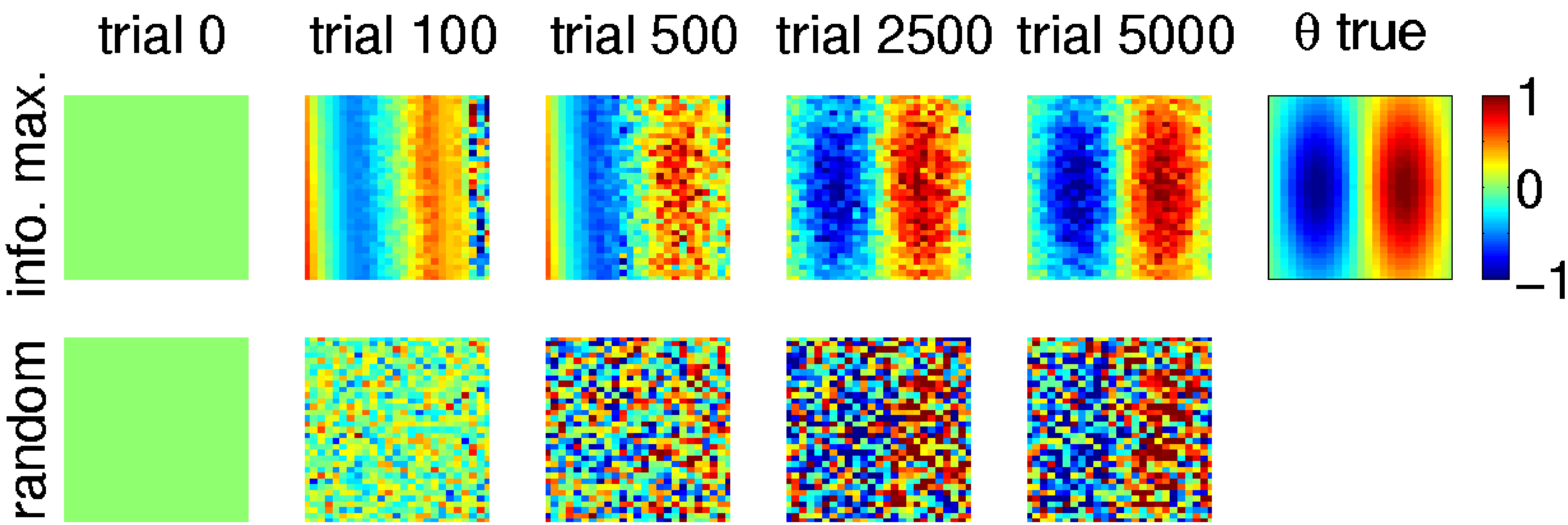
In general, very hard to do: high-d integration over θ to compute $I(n_t; \theta | \vec{x}_t)$, high-d optimization to select best $\vec{x}_t$.

GLM setting makes this surprisingly tractable
(Lewi et al., 2009).

Infomax vs. randomly-chosen stimuli

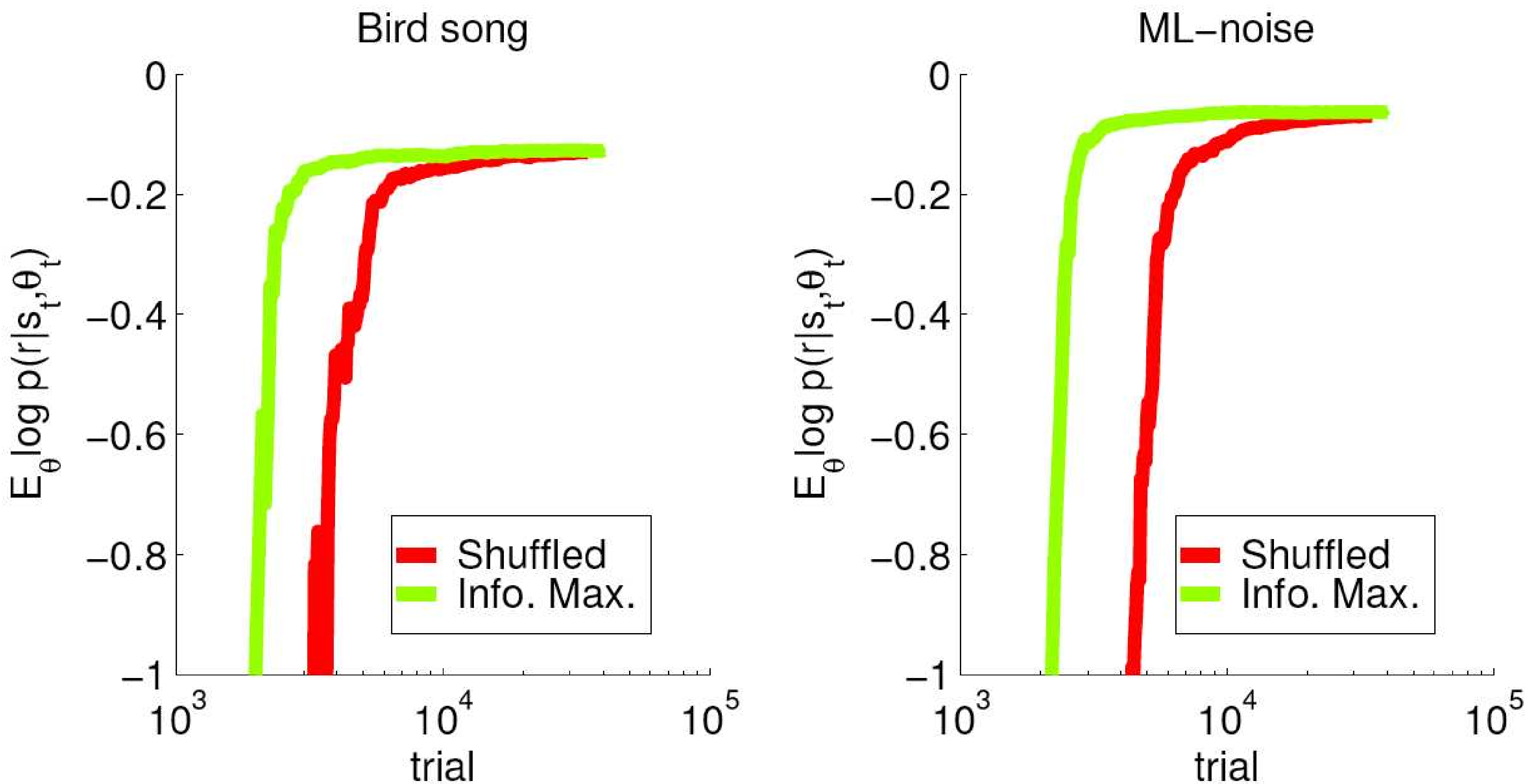


Simulated example



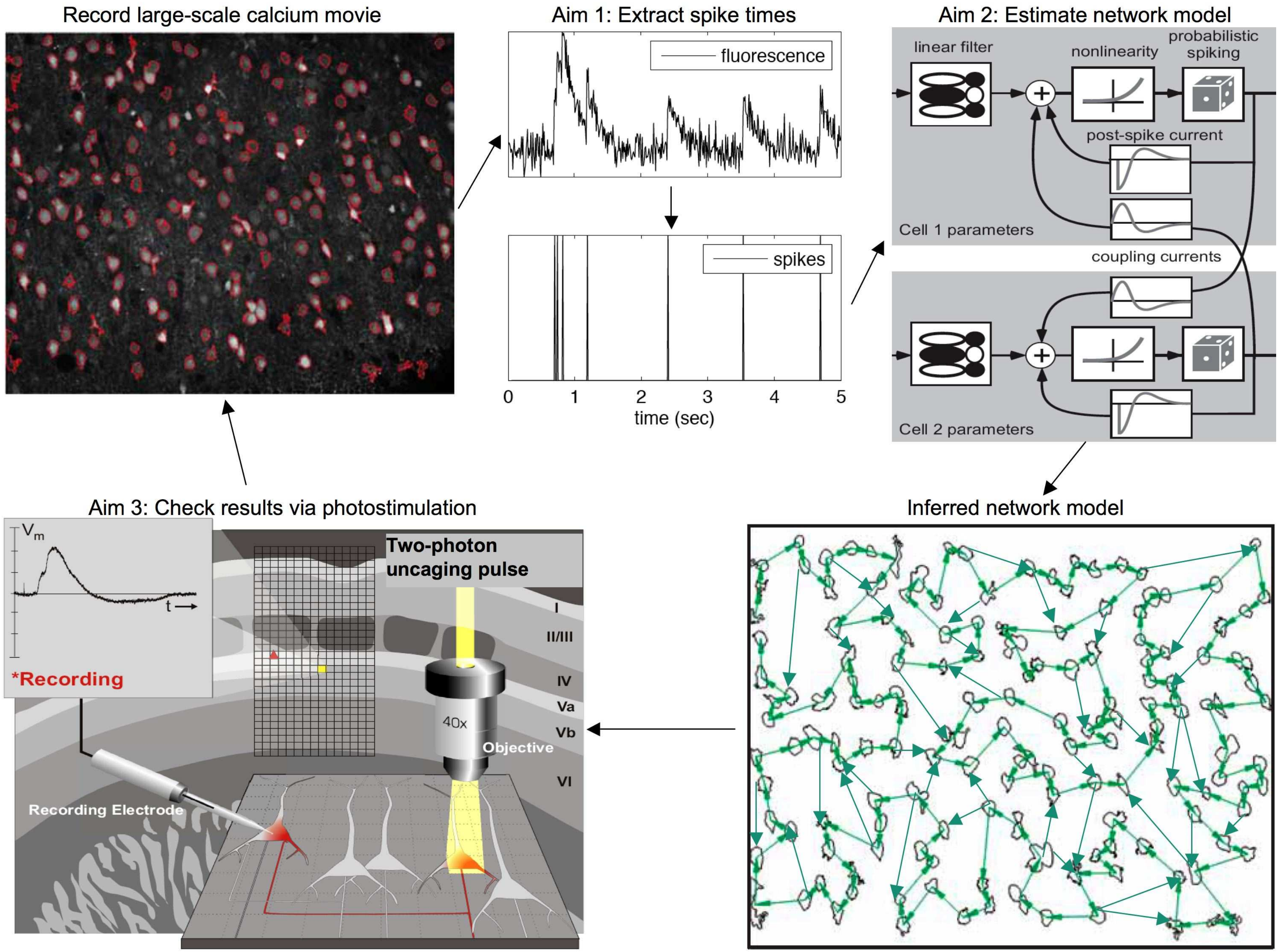
— infomax can be an order of magnitude more efficient.

Application to real data: choosing an optimal stimulus sequence

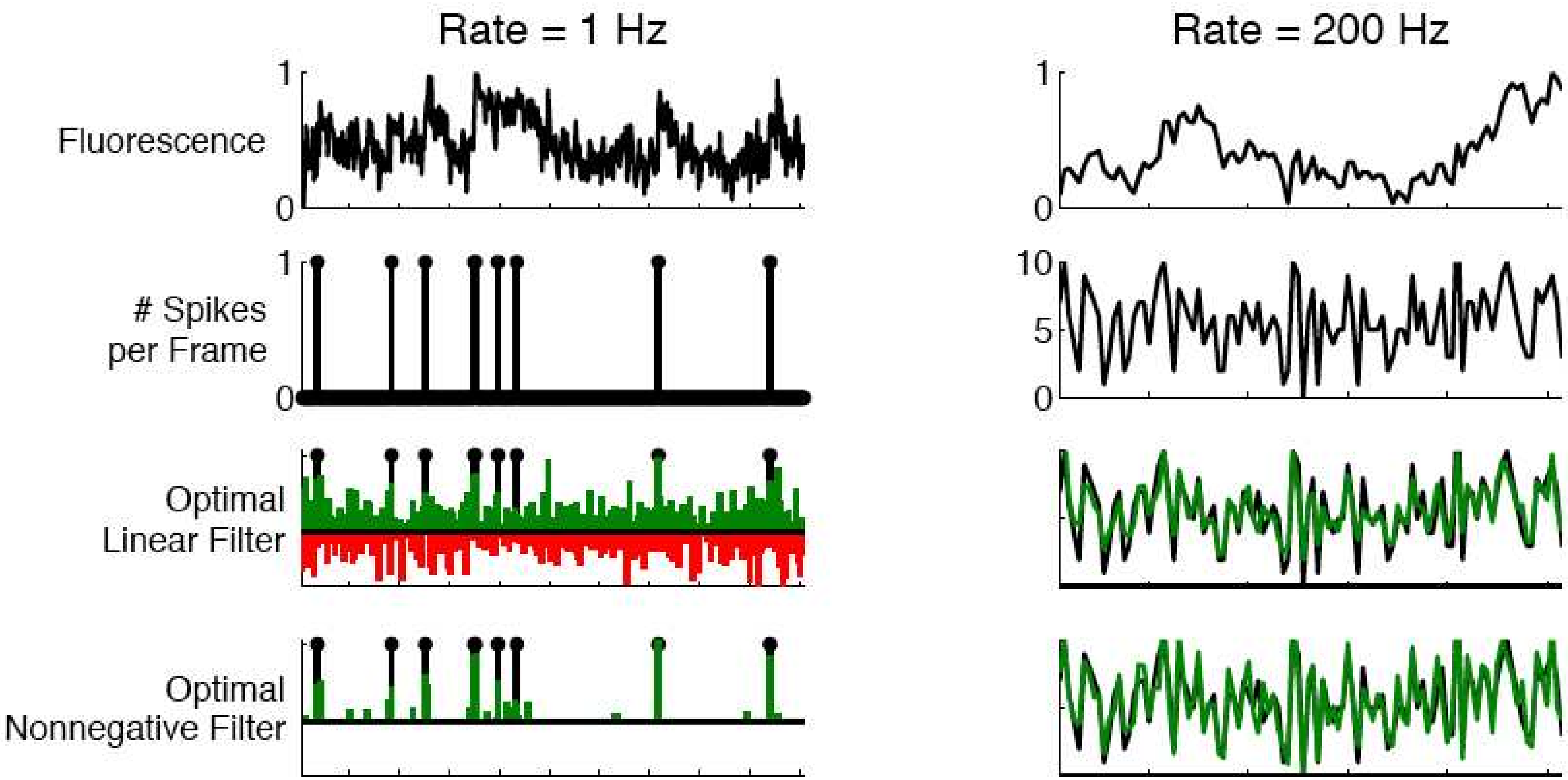


— stimuli chosen from a fixed pool; greater improvements expected if we can choose arbitrary stimuli on each trial.

Part 3: circuit inference



Challenge: slow, noisy calcium data

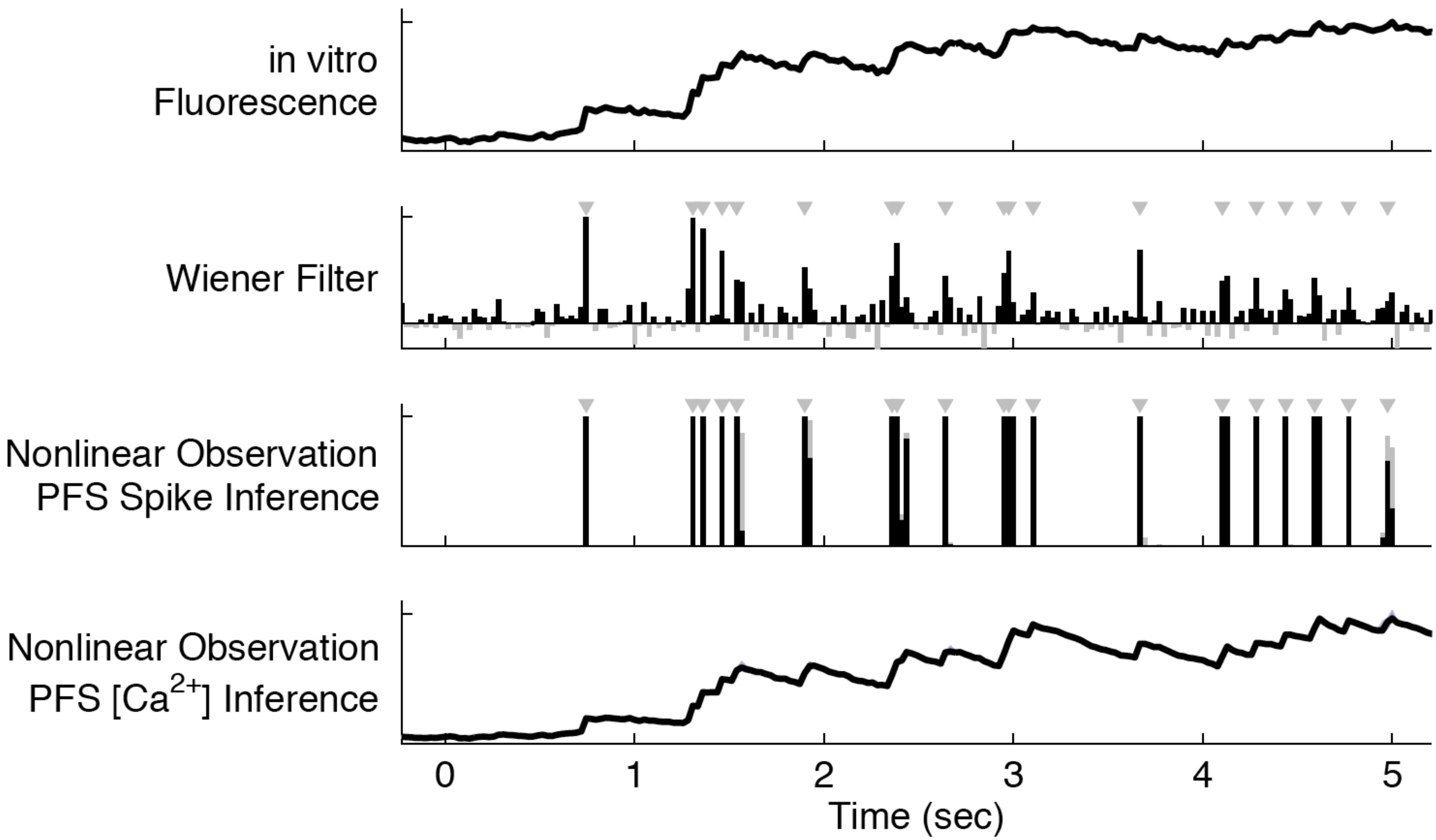


First-order model:

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

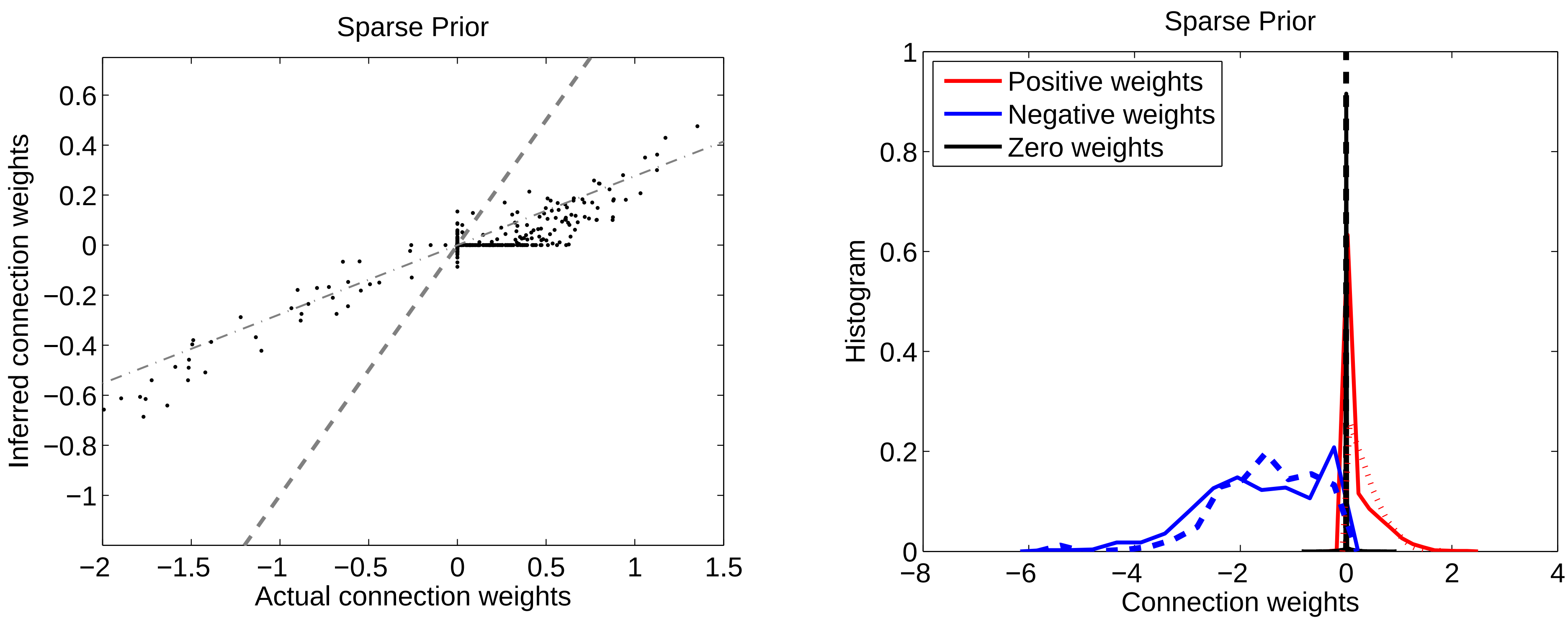
— $\tau \approx 100$ ms; nonnegative deconvolution problem. Can be solved by $O(T)$ relaxed constrained optimization methods (Vogelstein et al., 2008) or sequential Monte Carlo (Vogelstein et al., 2009).

Particle filter can extract spikes from saturated recordings



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

Simulated circuit inference



— Connections are inferred with the correct sign in conductance-based integrate-and-fire networks with biologically plausible connectivity matrices (Mishchencko et al., 2009).

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.

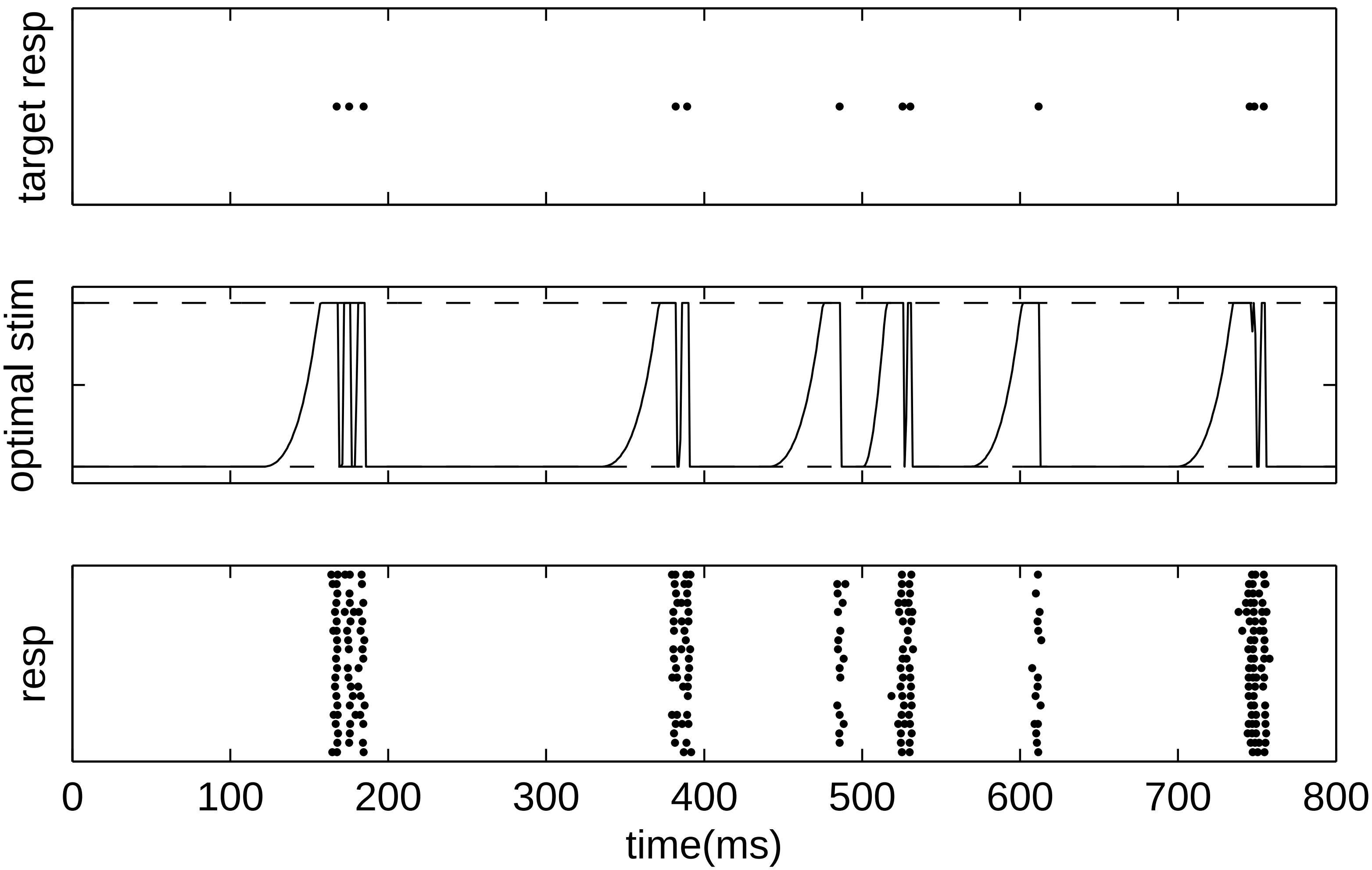
Start with a simple model:

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

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.

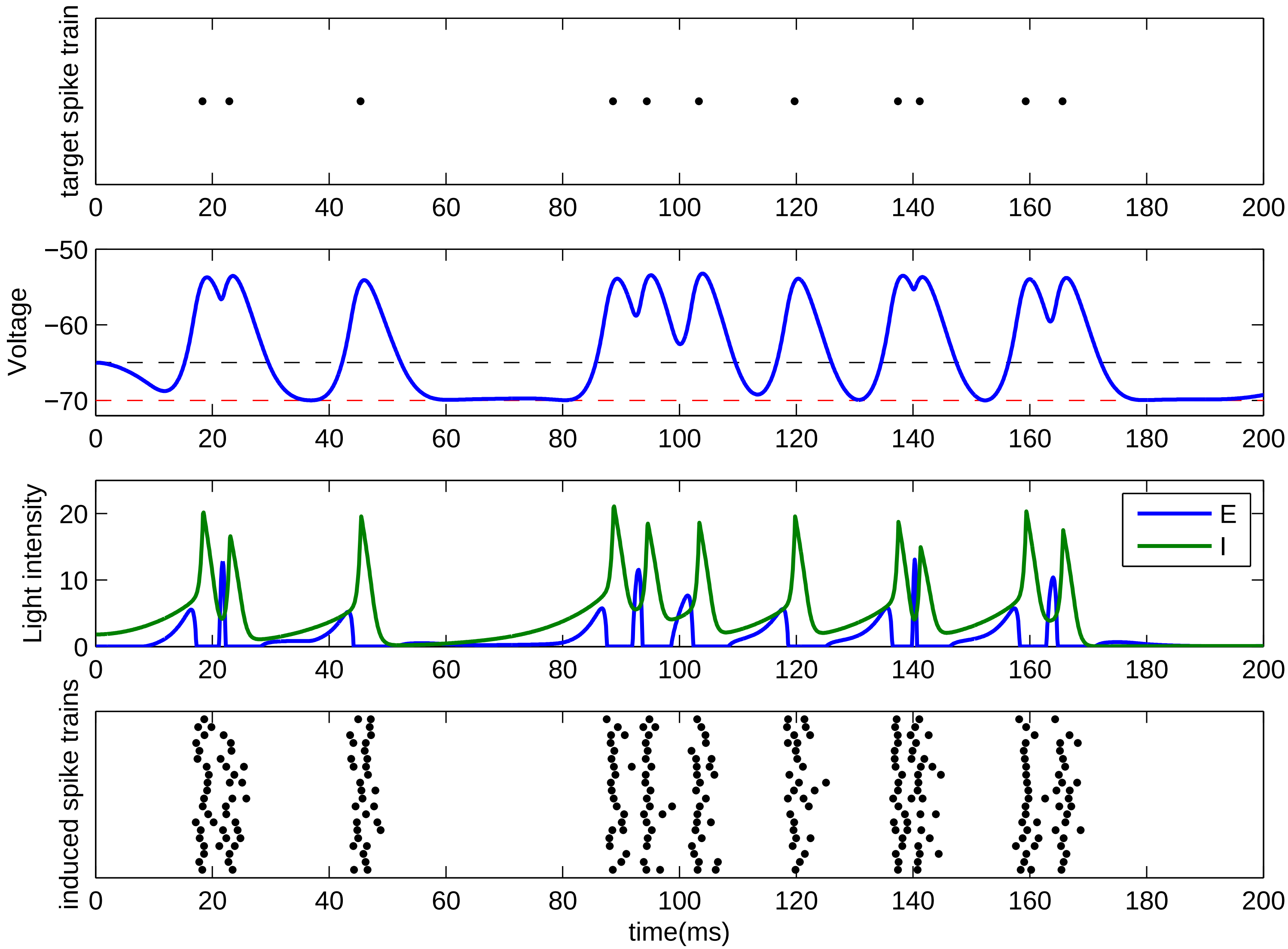
Optimal electrical control of spike timing



Optical conductance-based control of spiking

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

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



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., Kulkarni, J., Shlens, J., Simoncelli, E., Chichilnisky, E., and Paninski, L. (2008). Analyzing the neural code in the primate retina using efficient model-based decoding techniques. *SFN Abstract*.

Ahmadian, Y., Pillow, J., and Paninski, L. (2009). Efficient Markov Chain Monte Carlo methods for decoding population spike trains. *Under review, Neural Computation*.

Ahrens, M., Paninski, L., and Sahani, M. (2008). Inferring input nonlinearities in neural encoding models. *Network: Computation in Neural Systems*, 19:35–67.

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.

Gill, P., Zhang, J., Woolley, S., Fremouw, T., and Theunissen, F. (2006). Sound representation methods for spectro-temporal receptive field estimation. *Journal of Computational Neuroscience*, 21:5–20.

Huys, Q., Ahrens, M., and Paninski, L. (2006). Efficient estimation of detailed single-neuron models. *Journal of Neurophysiology*, 96:872–890.

Huys, Q. and Paninski, L. (2009). Model-based smoothing of, and parameter estimation from, noisy biophysical recordings. *PLOS Computational Biology*, 5:e1000379.

Lewi, J., Butera, R., and Paninski, L. (2009). Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21:619–687.

Paninski, L. (2009). Fast Kalman filtering on dendritic trees. *In progress*.

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.

Vogelstein, J., Babadi, B., Watson, B., Yuste, R., and Paninski, L. (2008). Fast nonnegative deconvolution via tridiagonal interior-point methods, applied to calcium fluorescence data. *Statistical analysis of neural data (SAND) conference*.

Vogelstein, J., Watson, B., Packer, A., Jedynak, B., Yuste, R., and Paninski, L. (2009). Model-based optimal inference of spike times and calcium dynamics given noisy and intermittent calcium-fluorescence imaging. *Biophysical Journal*, In press; http://www.stat.columbia.edu/∼liam/research/abstracts/vogelstein-bj08-abs.html.