Combining biophysical and statistical methods for understanding neural codes

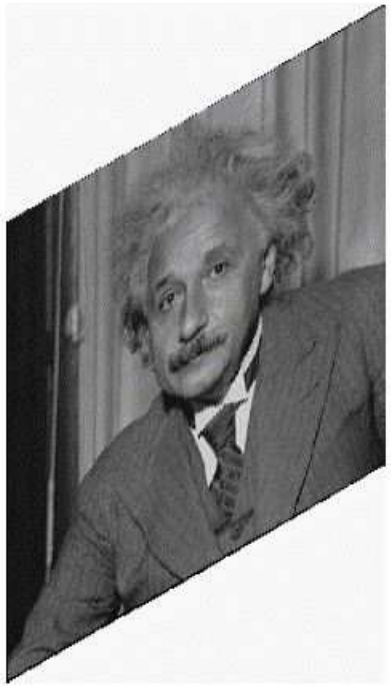
Liam Paninski

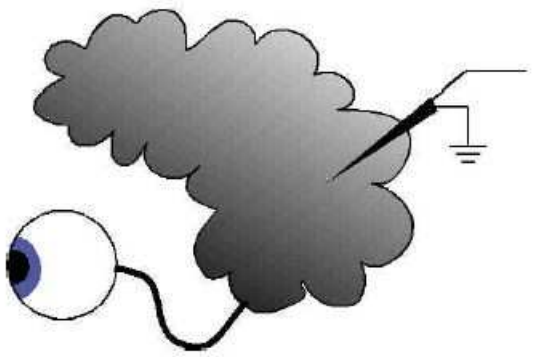
Department of Statistics and Center for Theoretical Neuroscience
Columbia University
http://www.stat.columbia.edu/∼liam
liam@stat.columbia.edu

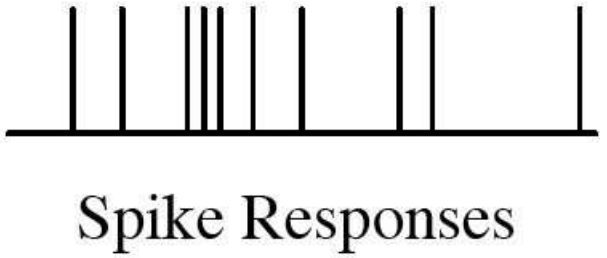
October 22, 2007

Support: NIH CRCNS award, Sloan Research Fellowship, NSF CAREER award.

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 (Paninski, 2003)

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

3: Fit a model with small number of parameters

Neural encoding models

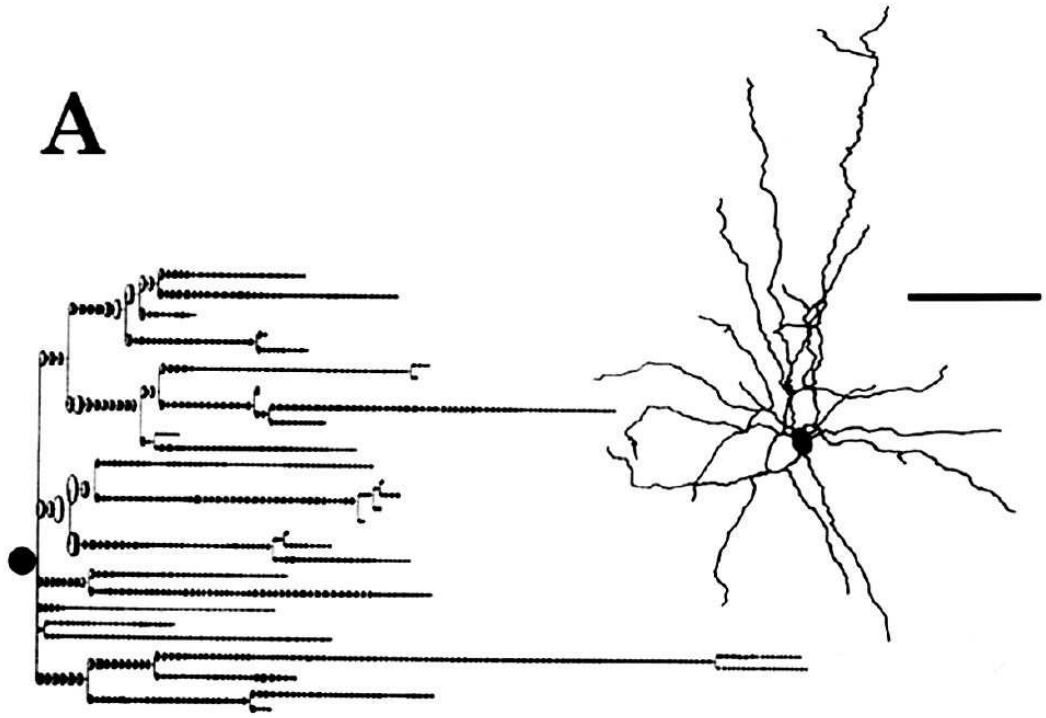
"Encoding model": $p_\theta(y|x)$.

— 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



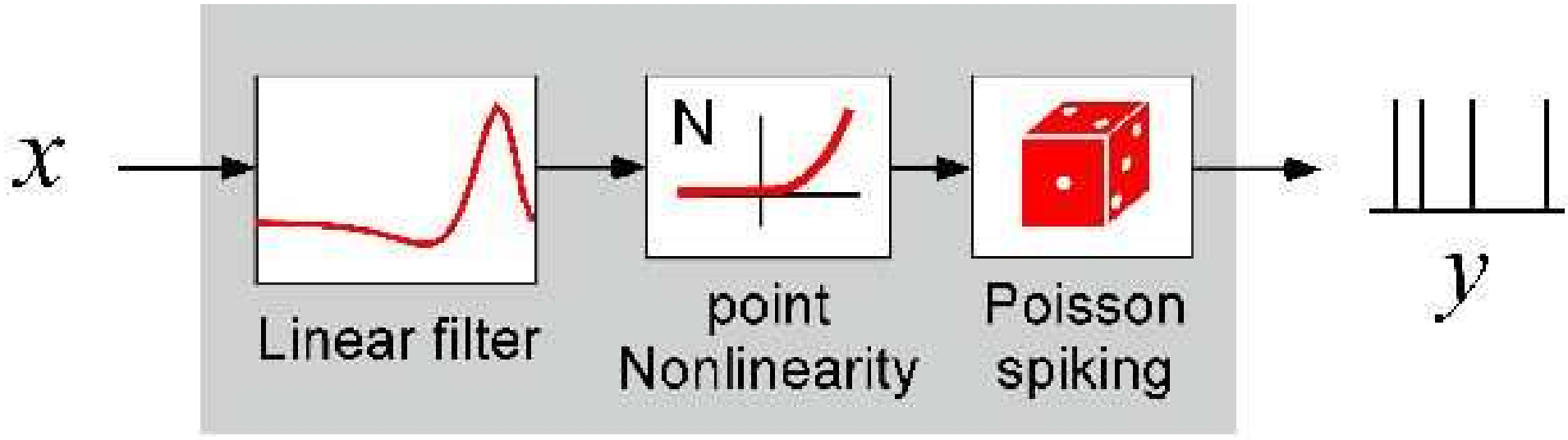
		Regional Conductances (mS/cm²)				
Model	Current	Dendrites	Soma	AH	NR	Axon
EC2.5 REAL	I_{Ca}	2.0	1.5	1.5	—	—
$j = 1$	$I_{K,Ca}$	0.001	0.065	0.065	0.065	0.065
SD* (real) = 21.9 μm	I_{Na}	25	80	100–150†	100	40–70‡
SD (EC2.5) = 20 μm	I_K	12	18	18	18	12–18‡
$\tau_{Ca} = 1.5$	I_A	36	54	54	54	—
$E_L = -60$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

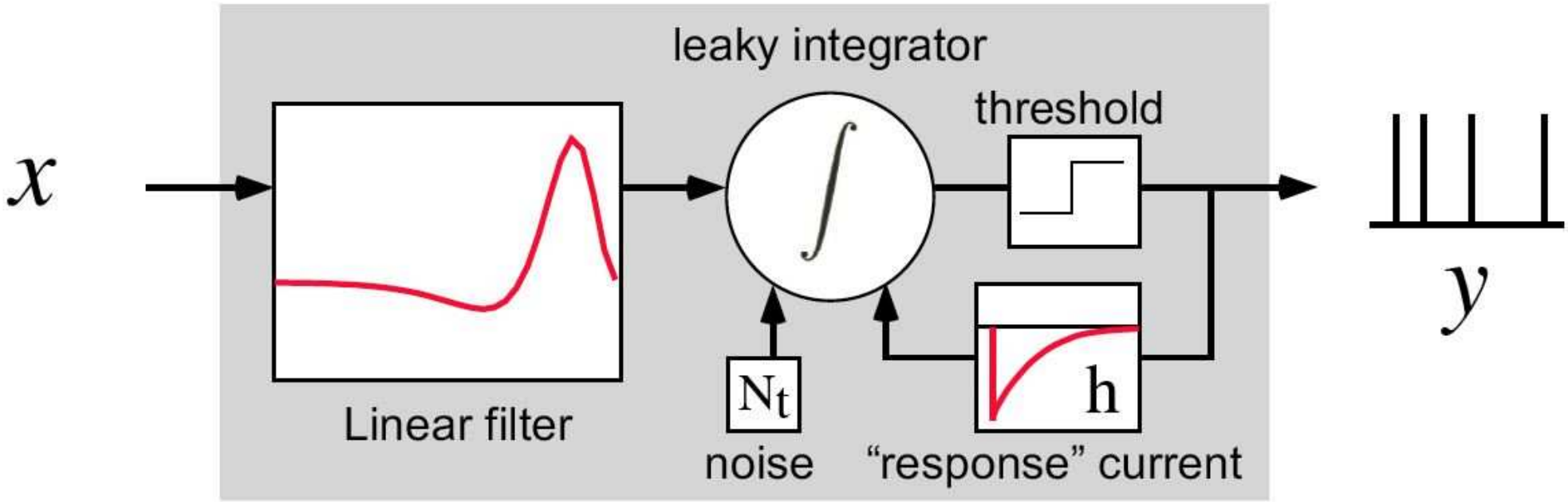
   — easy to incorporate prior knowledge, explicit noise models, etc.

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.

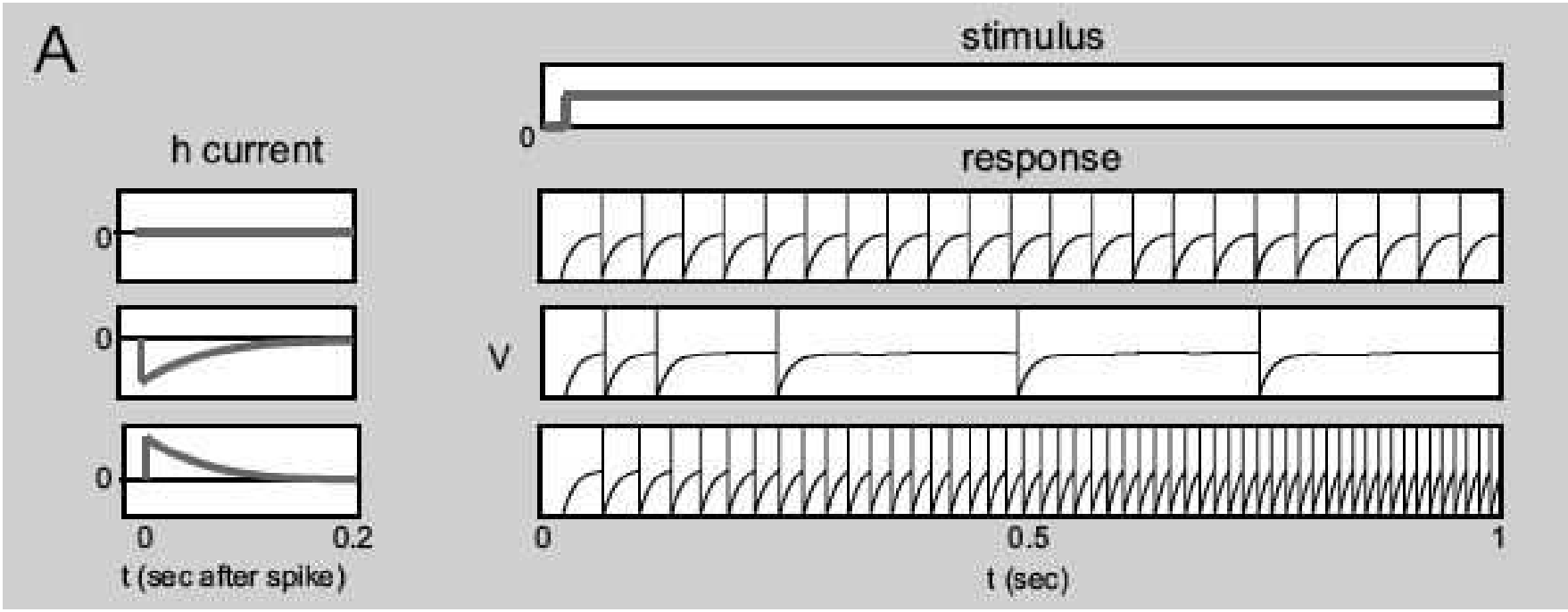
Filtered integrate-and-fire model



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

(Paninski et al., 2004b)

Model flexibility: Adaptation



The estimation problem

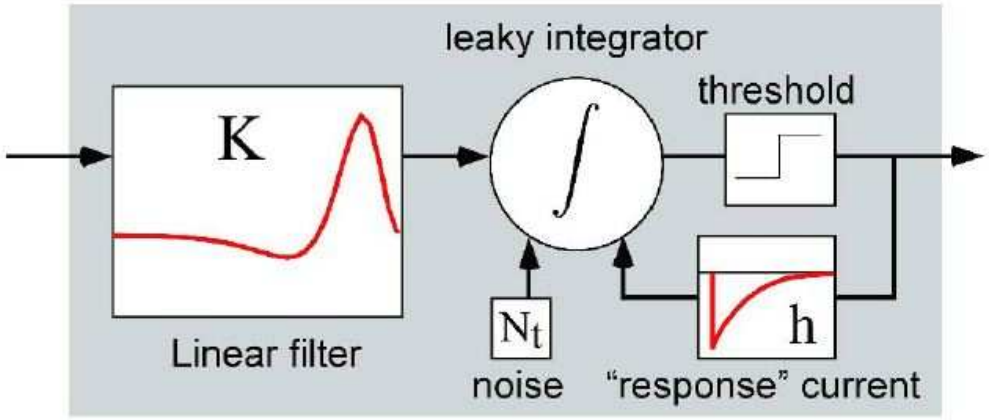
Learn the model parameters:

$\vec{K}$ = stimulus filter

g = leak conductance

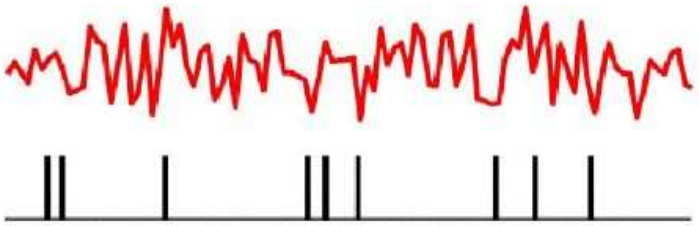
σ^2 = noise variance

$\vec{h}$ = response current



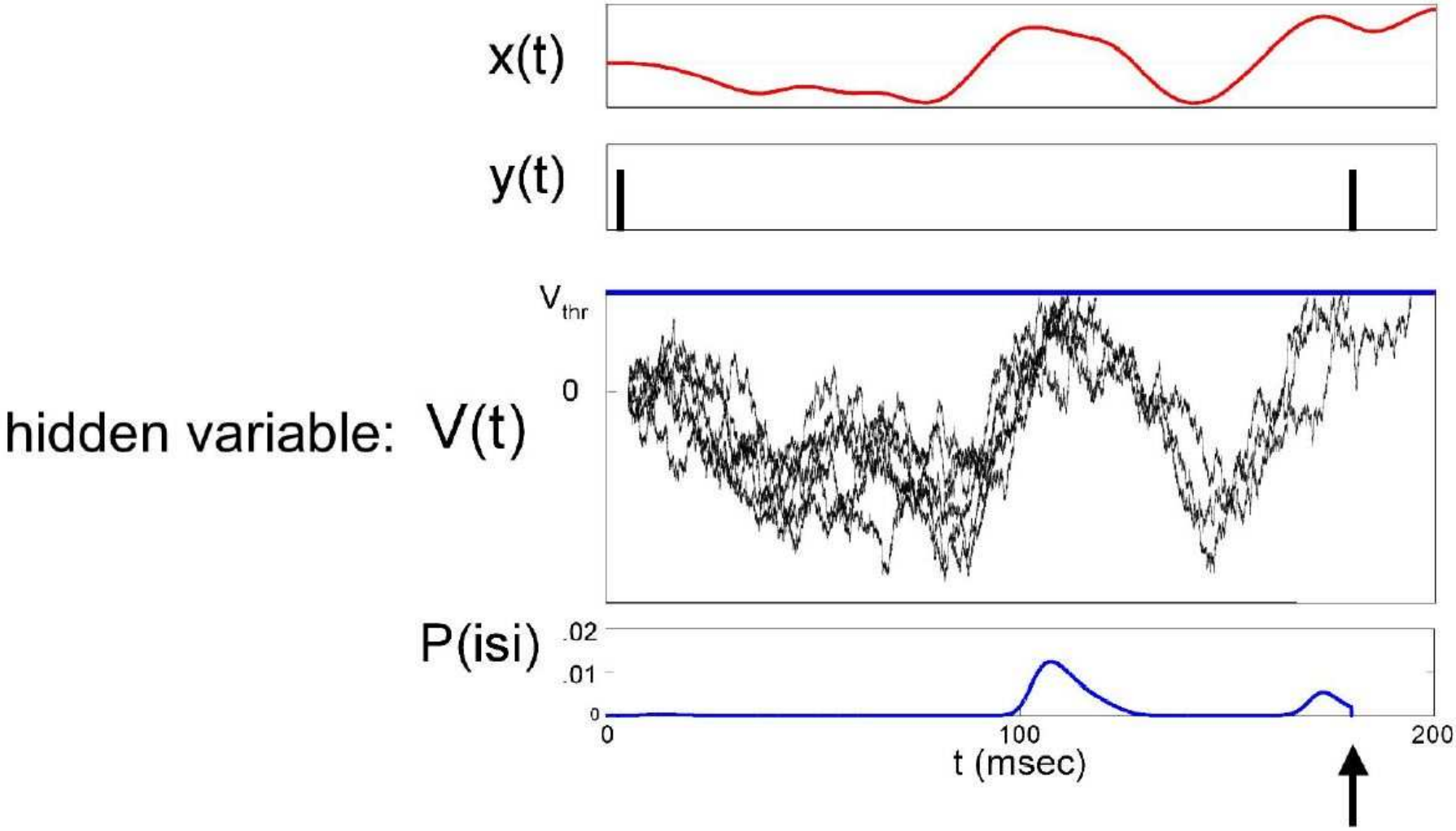
From: stimulus train $x(t)$

spike times t_i



(Paninski et al., 2004b)

First passage time likelihood



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

(via Fokker-Planck, integral equation, or EM; (Paninski et al., 2004b; Paninski et al., 2007; Nikitchenko and Paninski, 2007))

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.

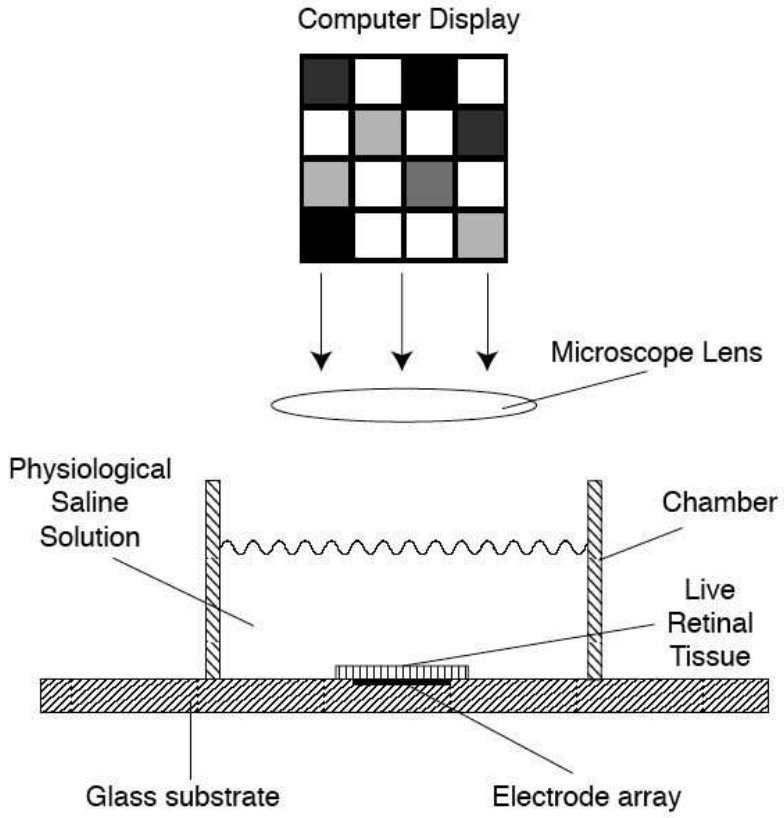
$\Longrightarrow$ no non-global local maxima

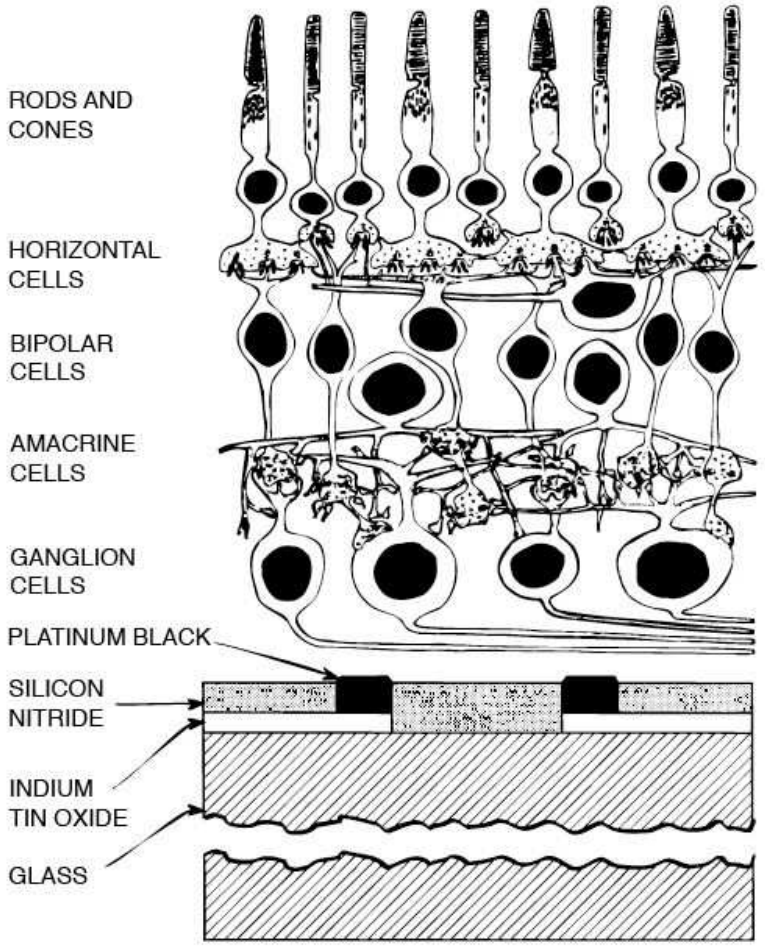
$\Longrightarrow$ maximization easy by ascent techniques.

Application: retinal ganglion cells

Preparation: dissociated macaque retina

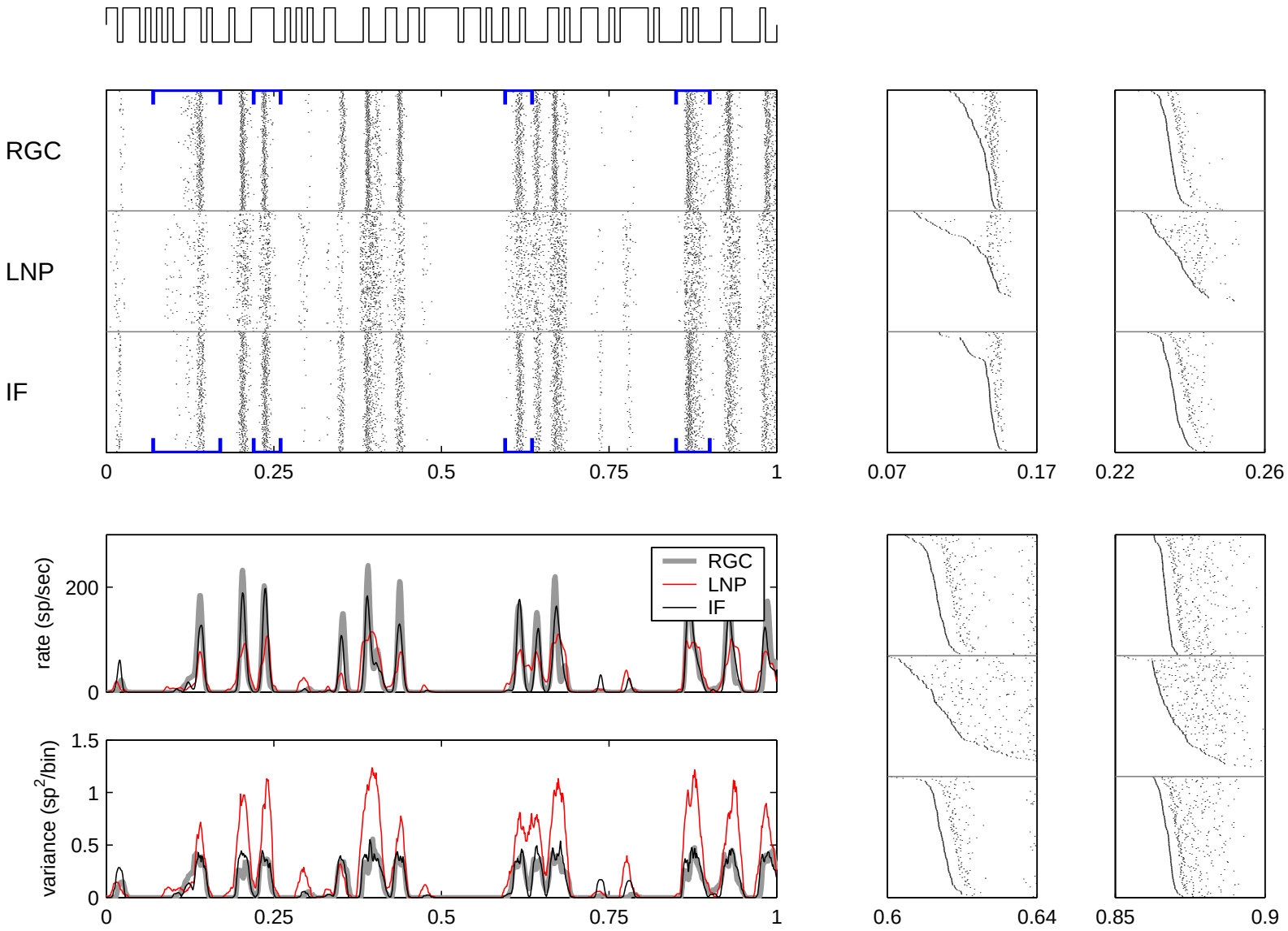
— extracellularly-recorded responses of populations of RGCs





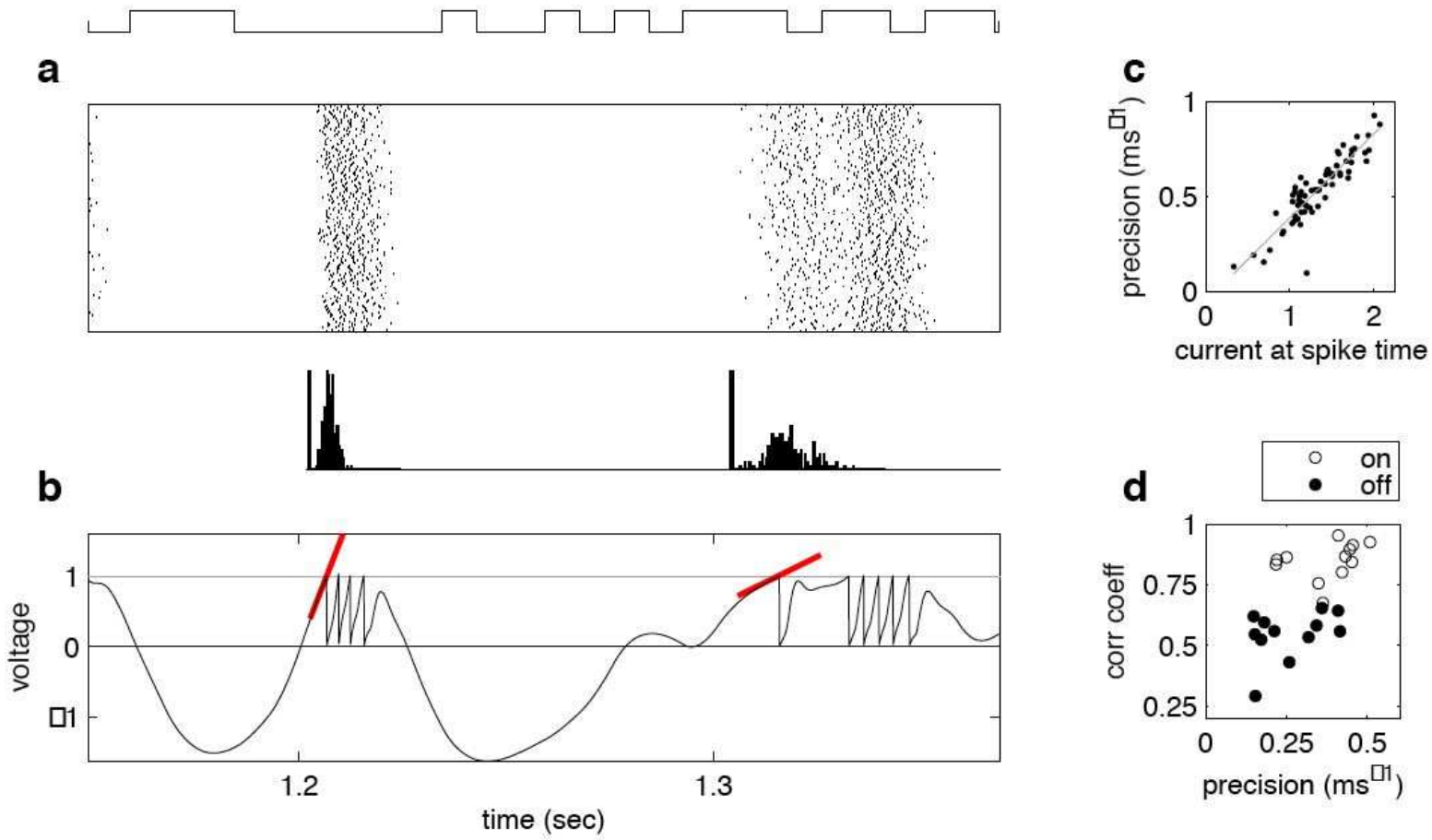
Stimulus: random "flicker" visual stimuli

Spike timing precision in retina



(Pillow et al., 2005)

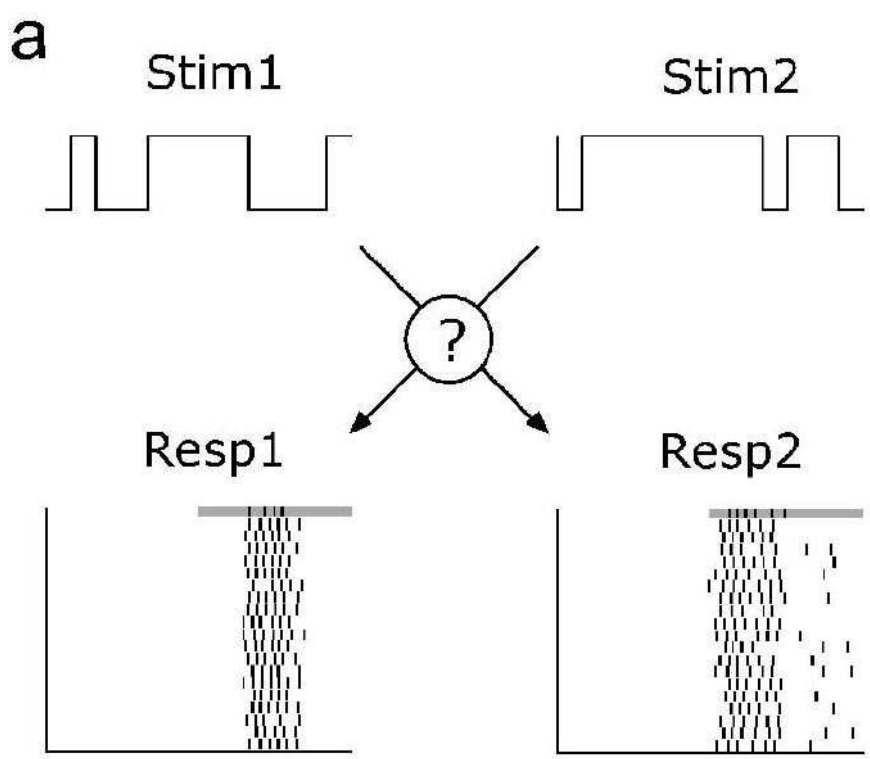
Linking spike reliability and subthreshold noise

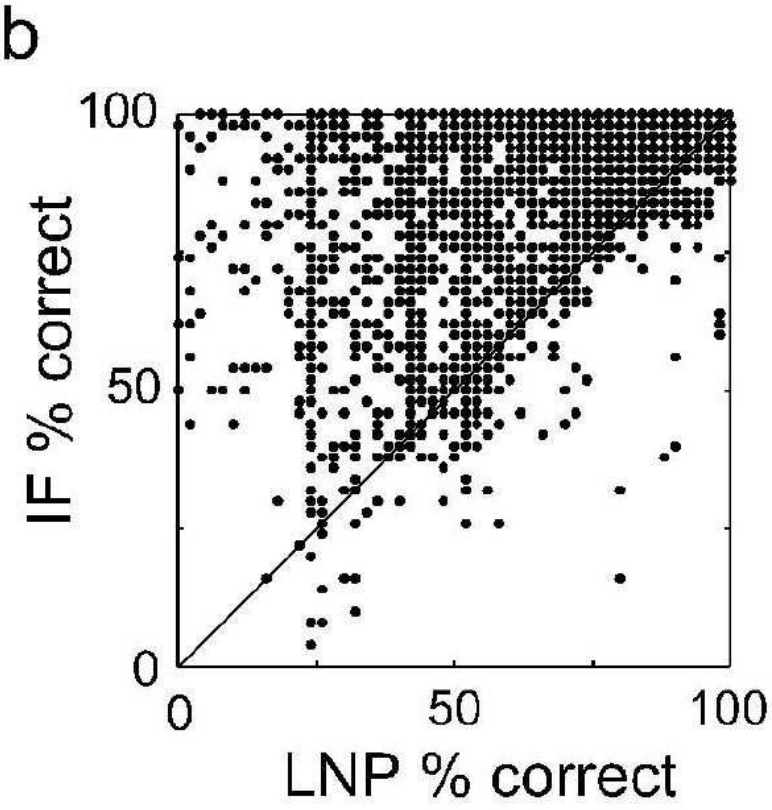


(Pillow et al., 2005)

Likelihood-based discrimination

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

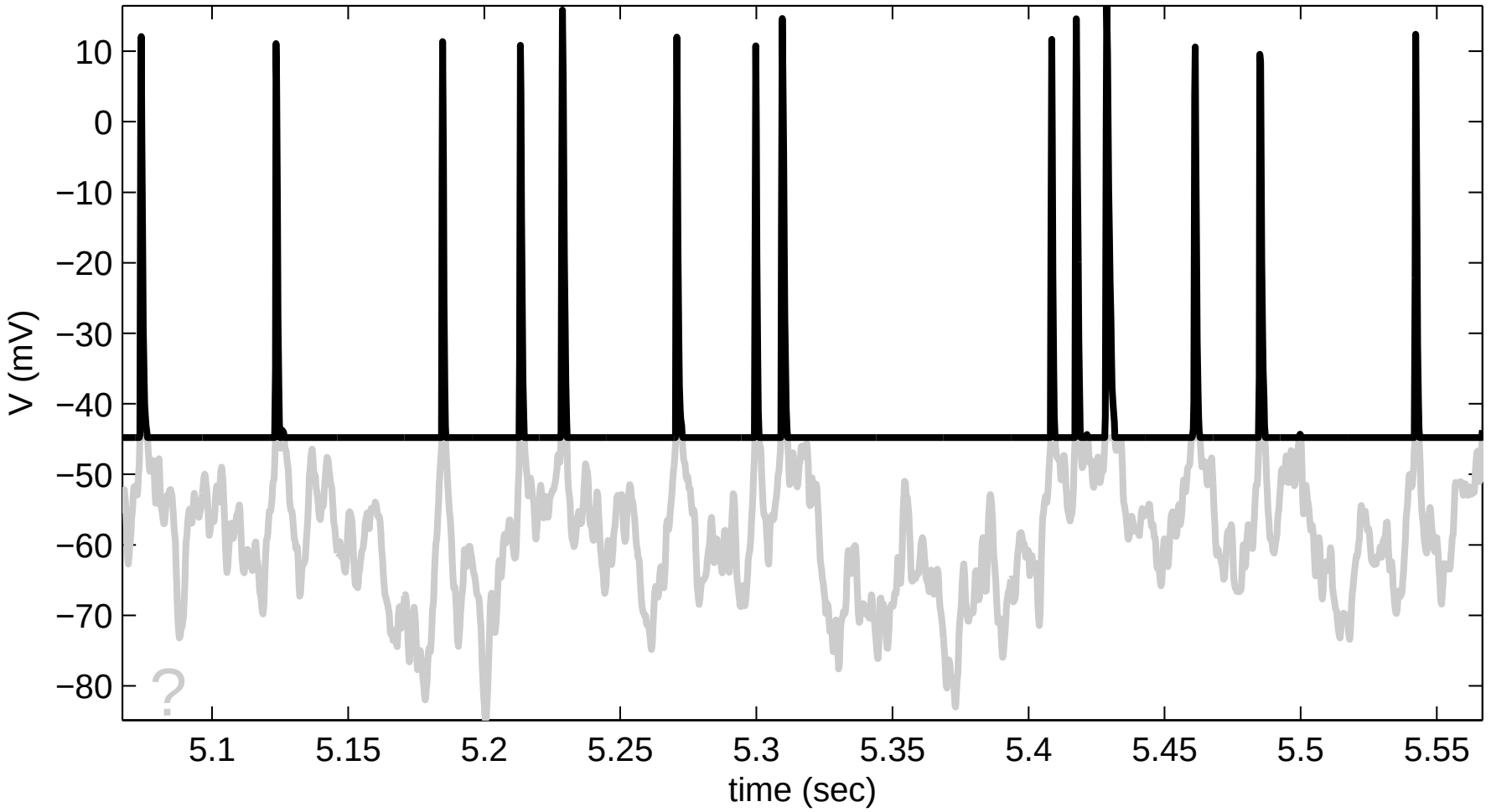




Using accurate model is essential (Pillow et al., 2005)

Example 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):

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

Constraints:

- Reset at $t = 0$:

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

- Spike at $t = T$:

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

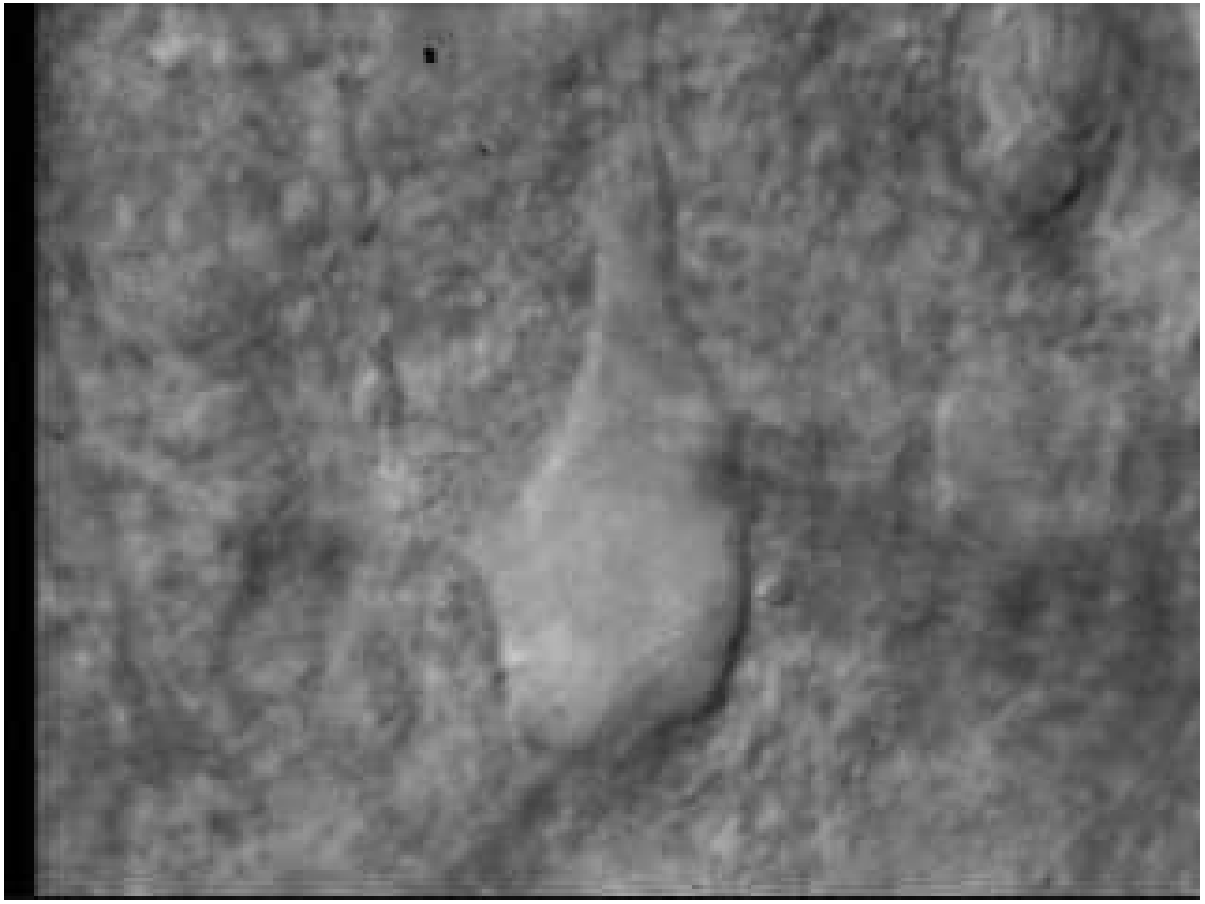
- No spike for $0 < t < T$:

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

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

Application: *in vitro* data

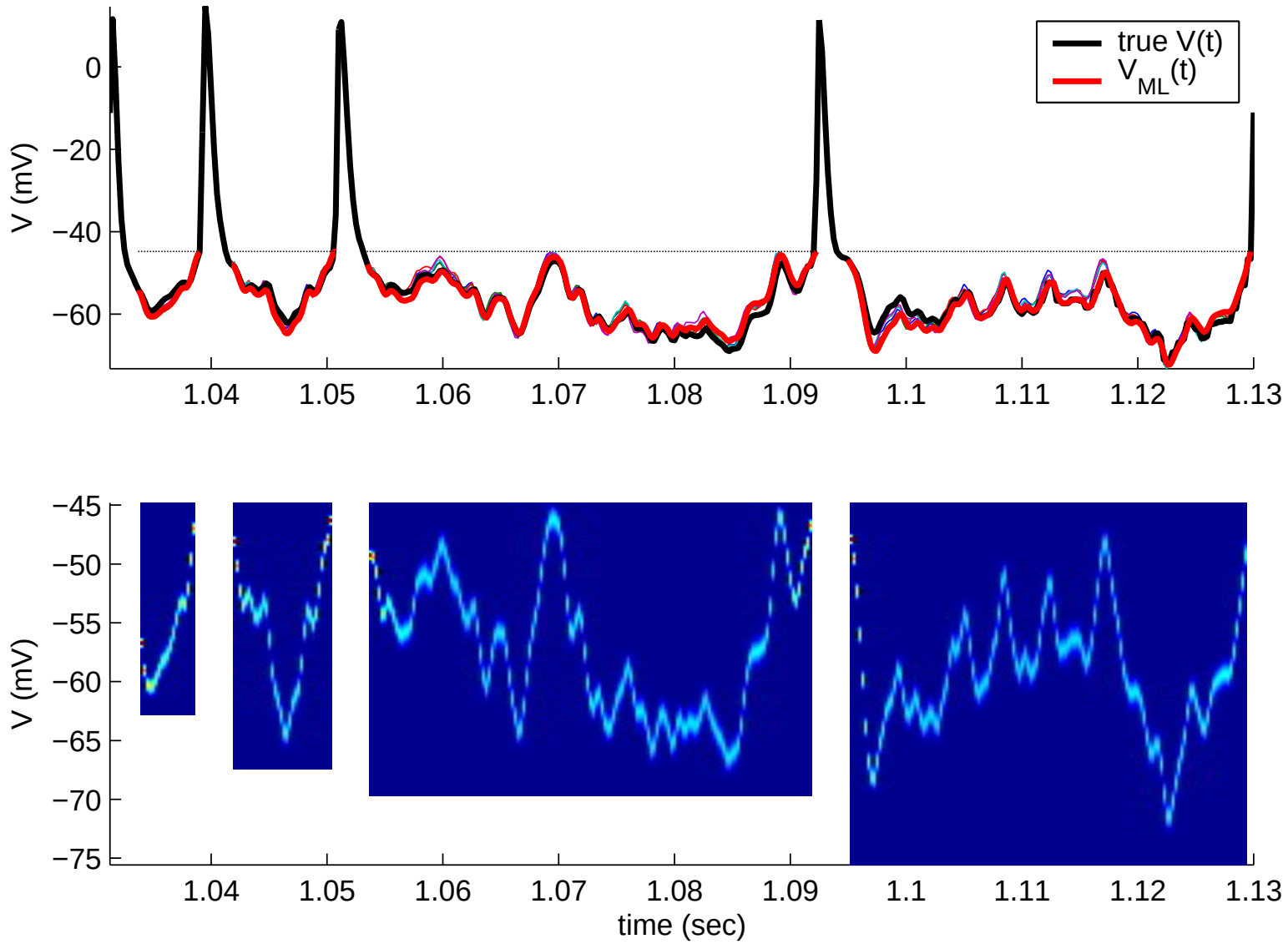
Recordings: rat sensorimotor cortical slice; dual-electrode whole-cell



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

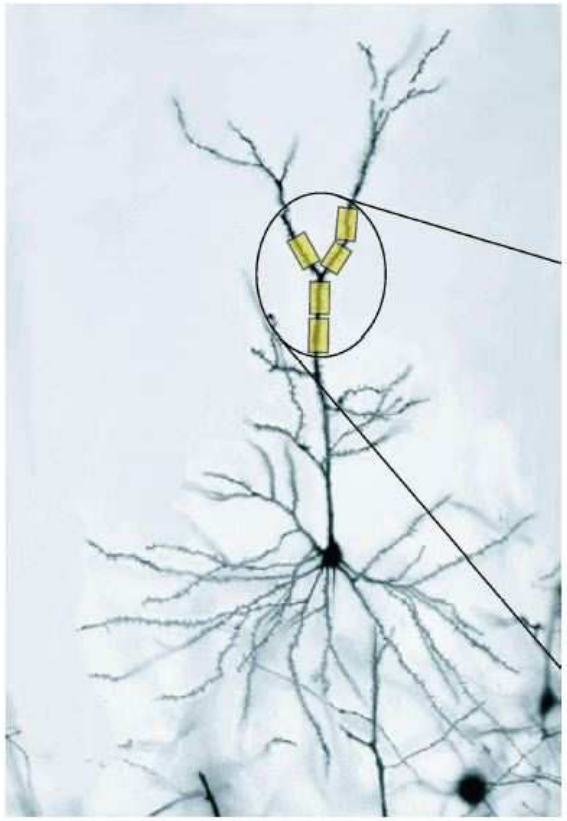
Analysis: fit IF model parameters $\{g, \vec{k}, h(.), V_{th}, \sigma\}$ by maximum likelihood (Paninski et al., 2003; Paninski et al., 2004a), then compute $V_{ML}(t)$

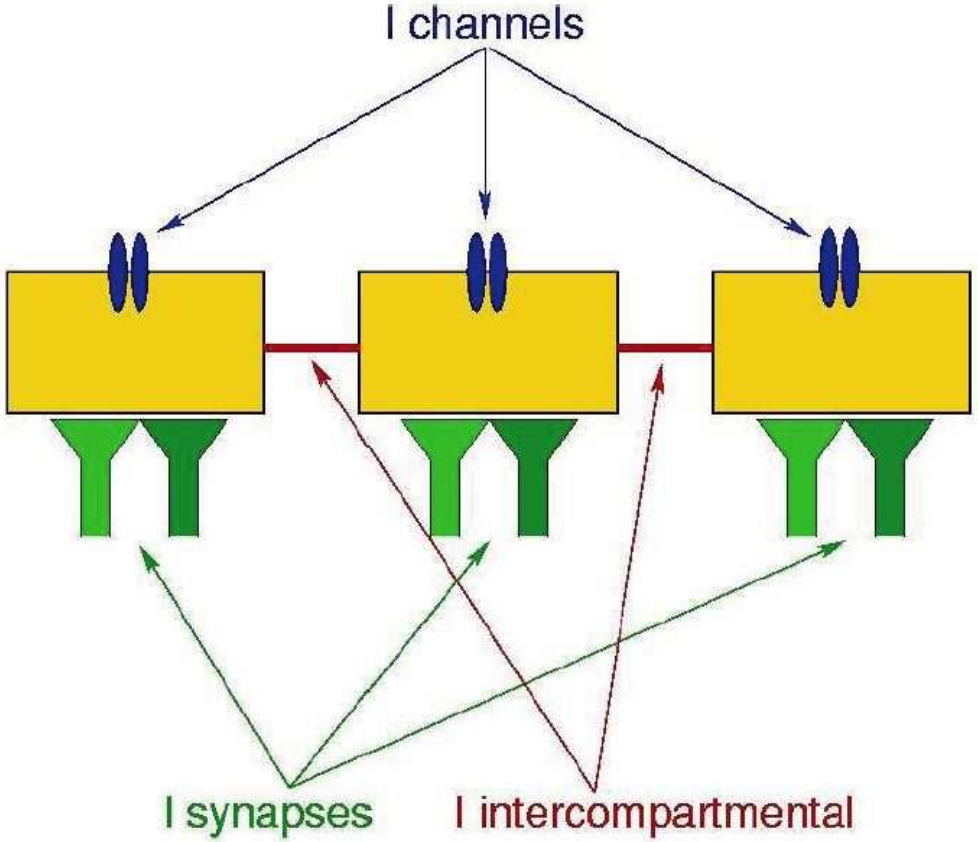
Application: *in vitro* data



(Applications to spike-triggered average (Paninski, 2006a; Paninski, 2006b).)

Part 3: Back to detailed models

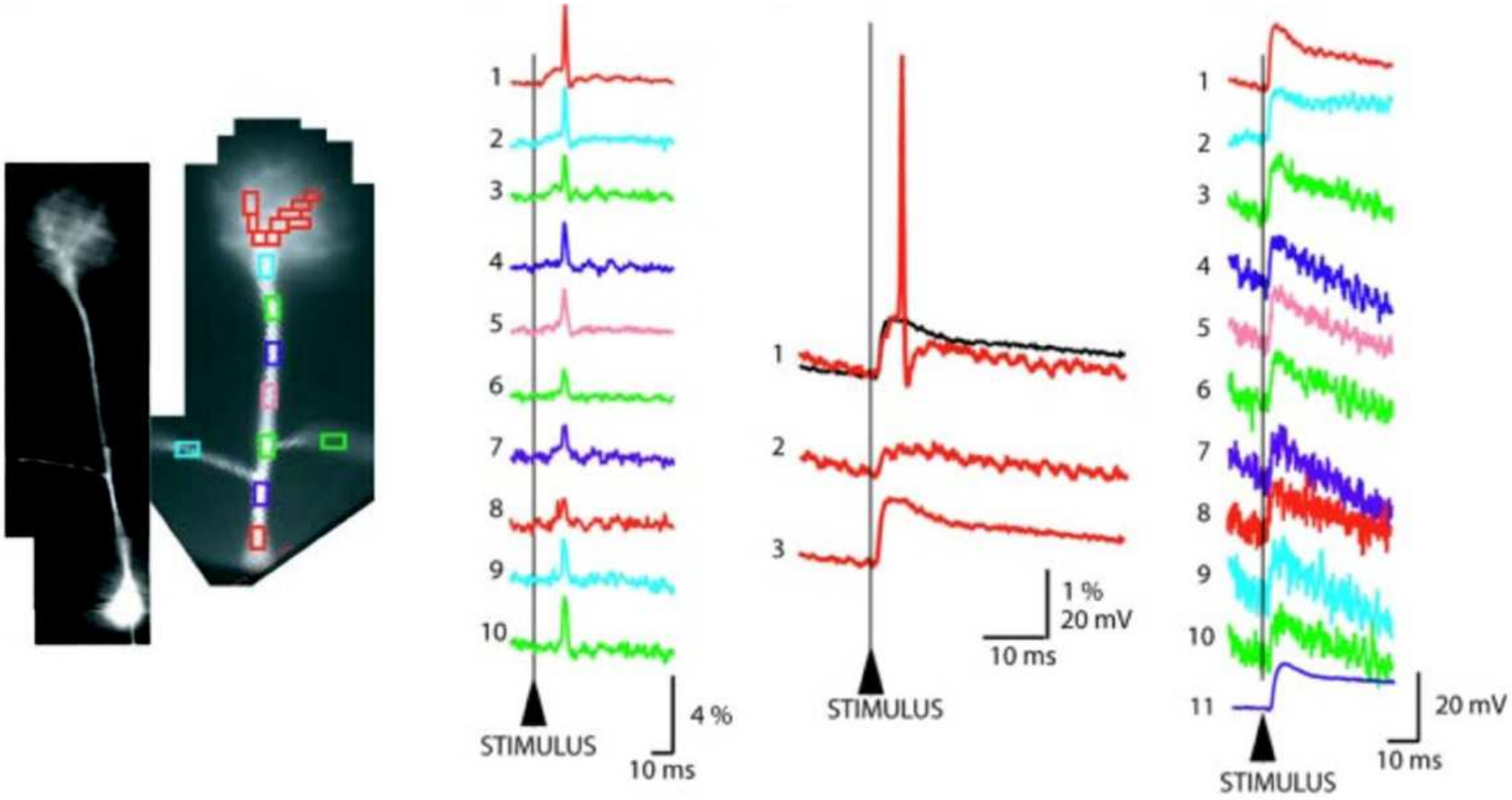




Can we recover detailed biophysical properties?

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

Spatiotemporal voltage recordings



Djurisic et al, 2004

Conductance-based models

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

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

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

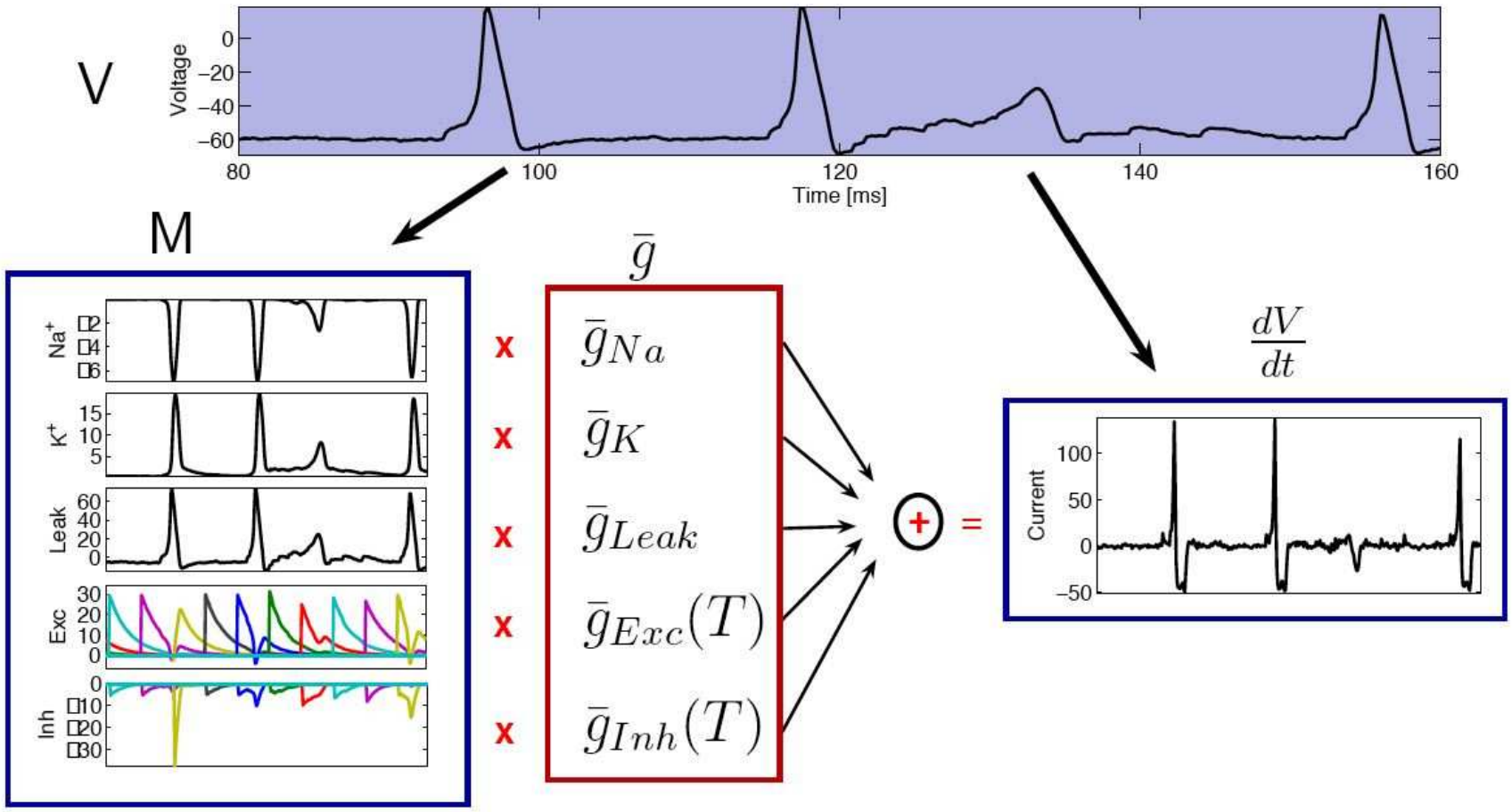
$$I_i^{\text{intercompartmental}} = \sum_a g_a \Delta V_a(t)$$

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.

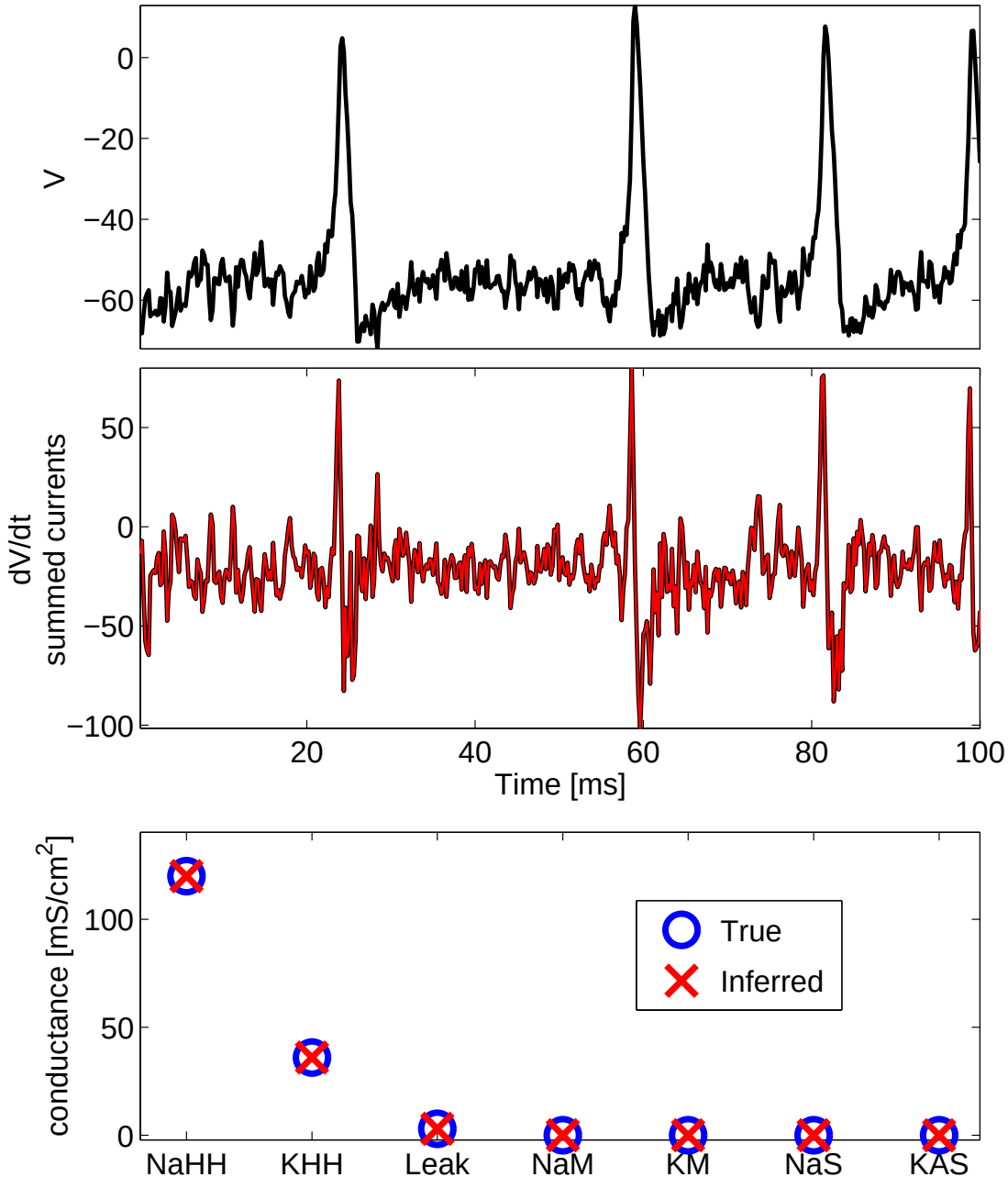
Gaussian noise $\implies$ standard nonnegative regression (albeit high-d).

Estimating channel densities from $V(t)$



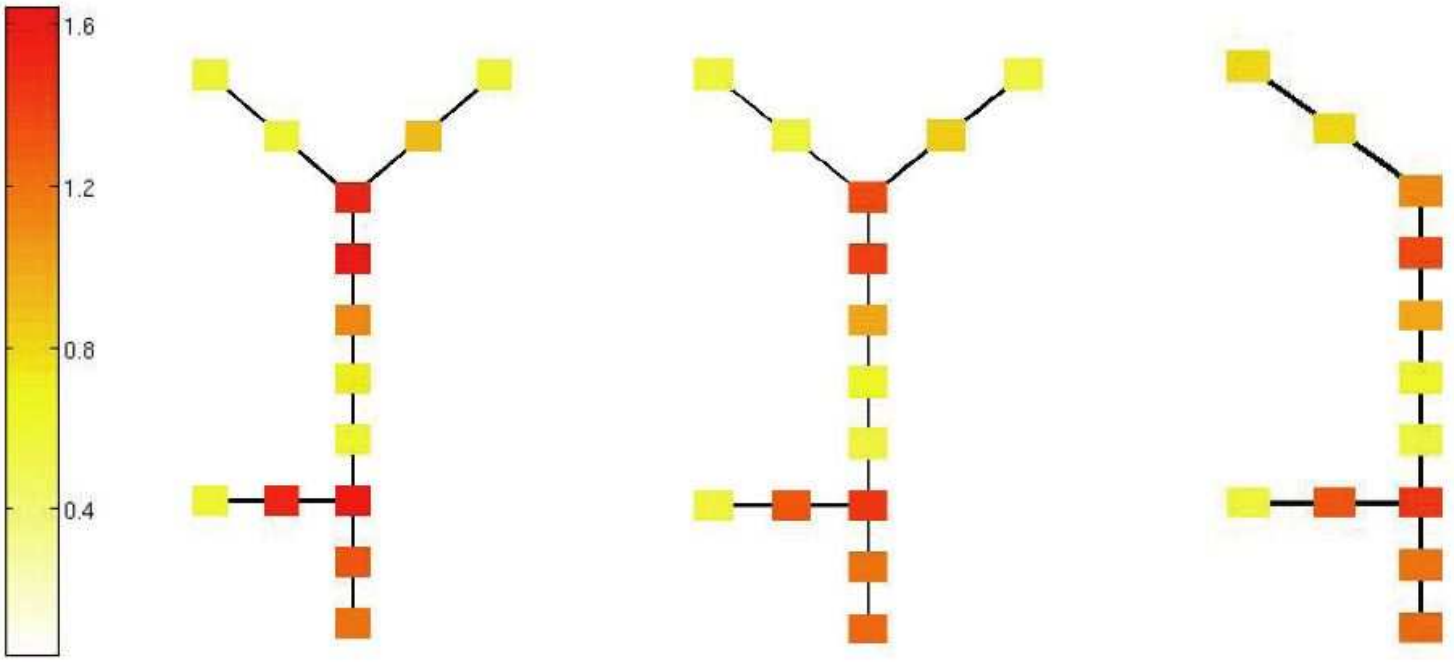
(Huys et al., 2006)

Estimating channel densities from $V(t)$



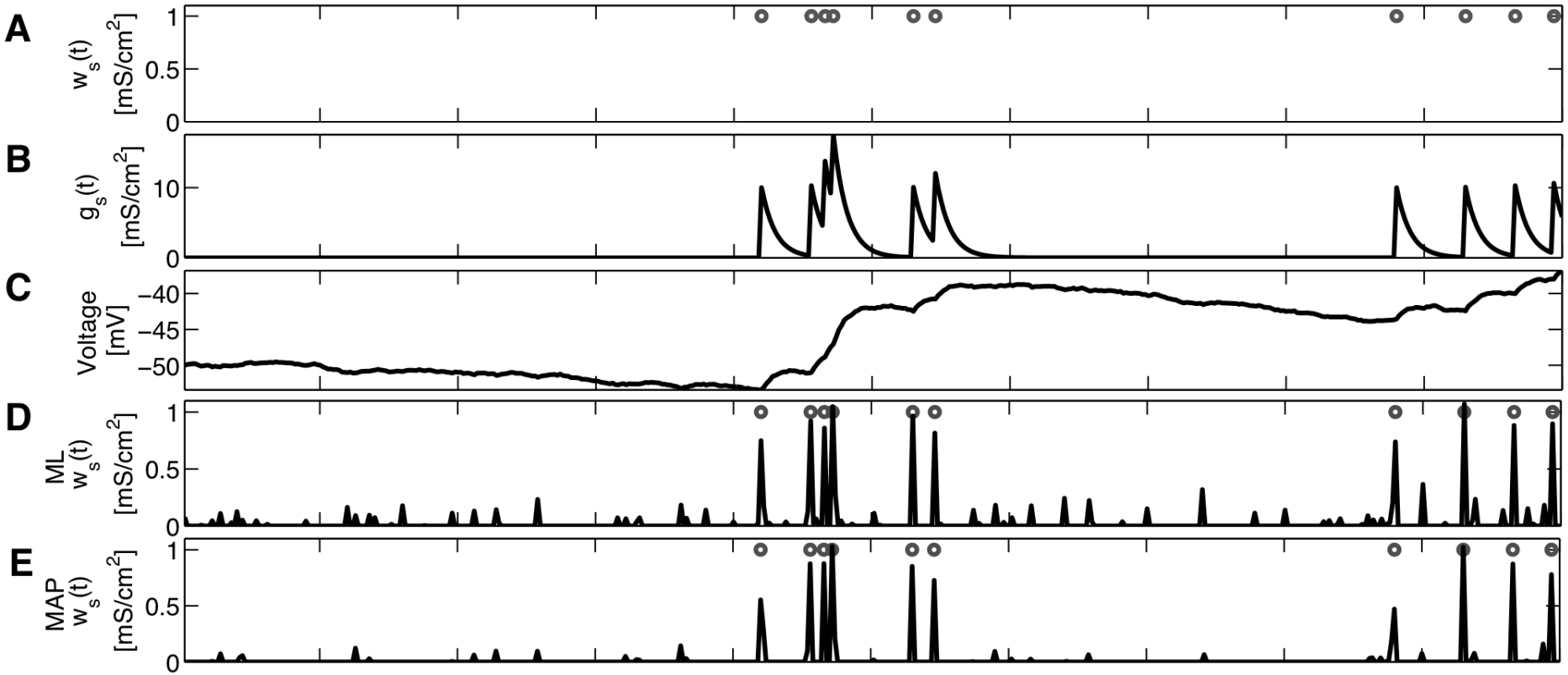
Estimating non-homogeneous channel densities and axial resistances from spatiotemporal voltage recordings

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

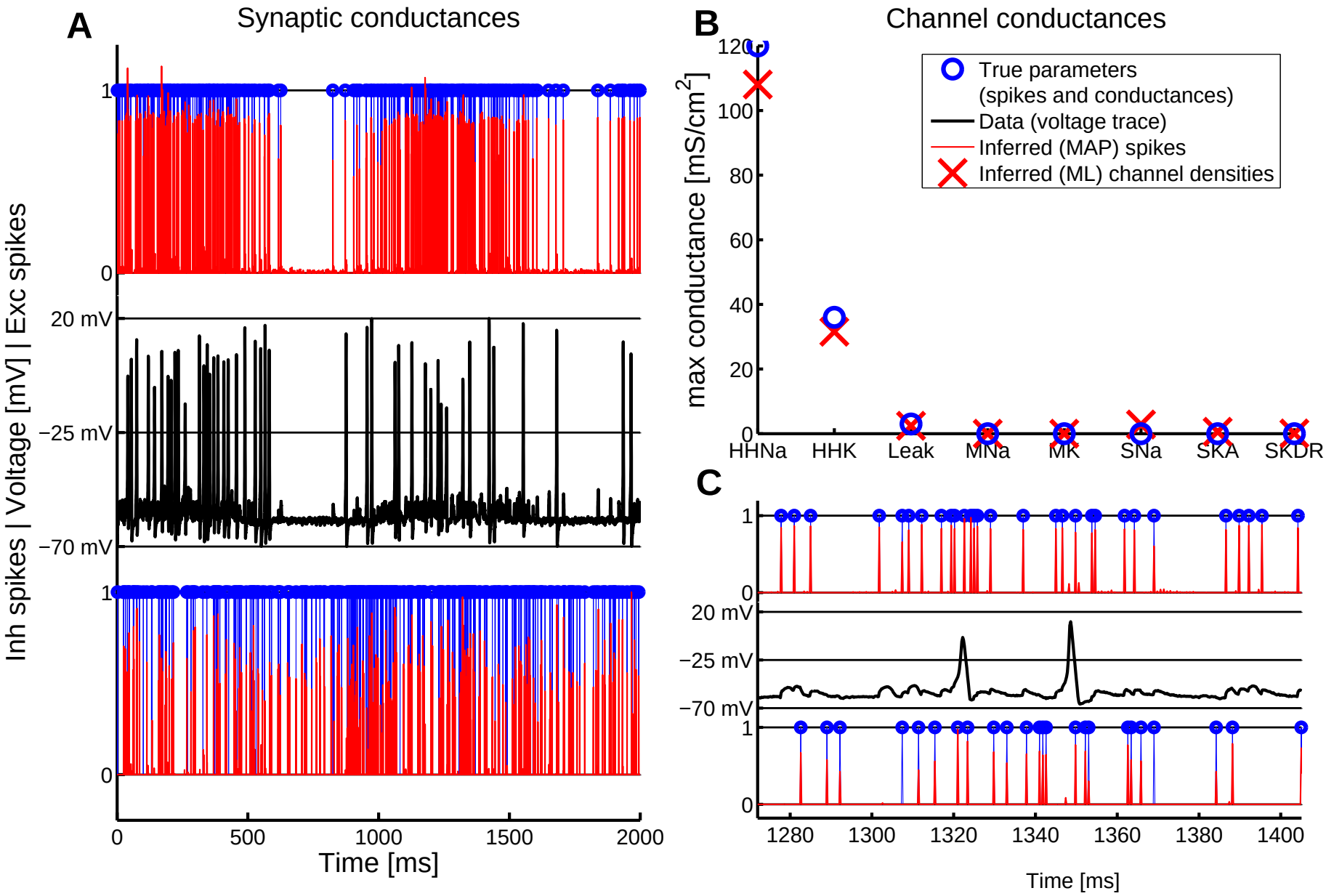


True g_{Na} Estimated g_{Na}

Estimating synaptic inputs given $V(t)$

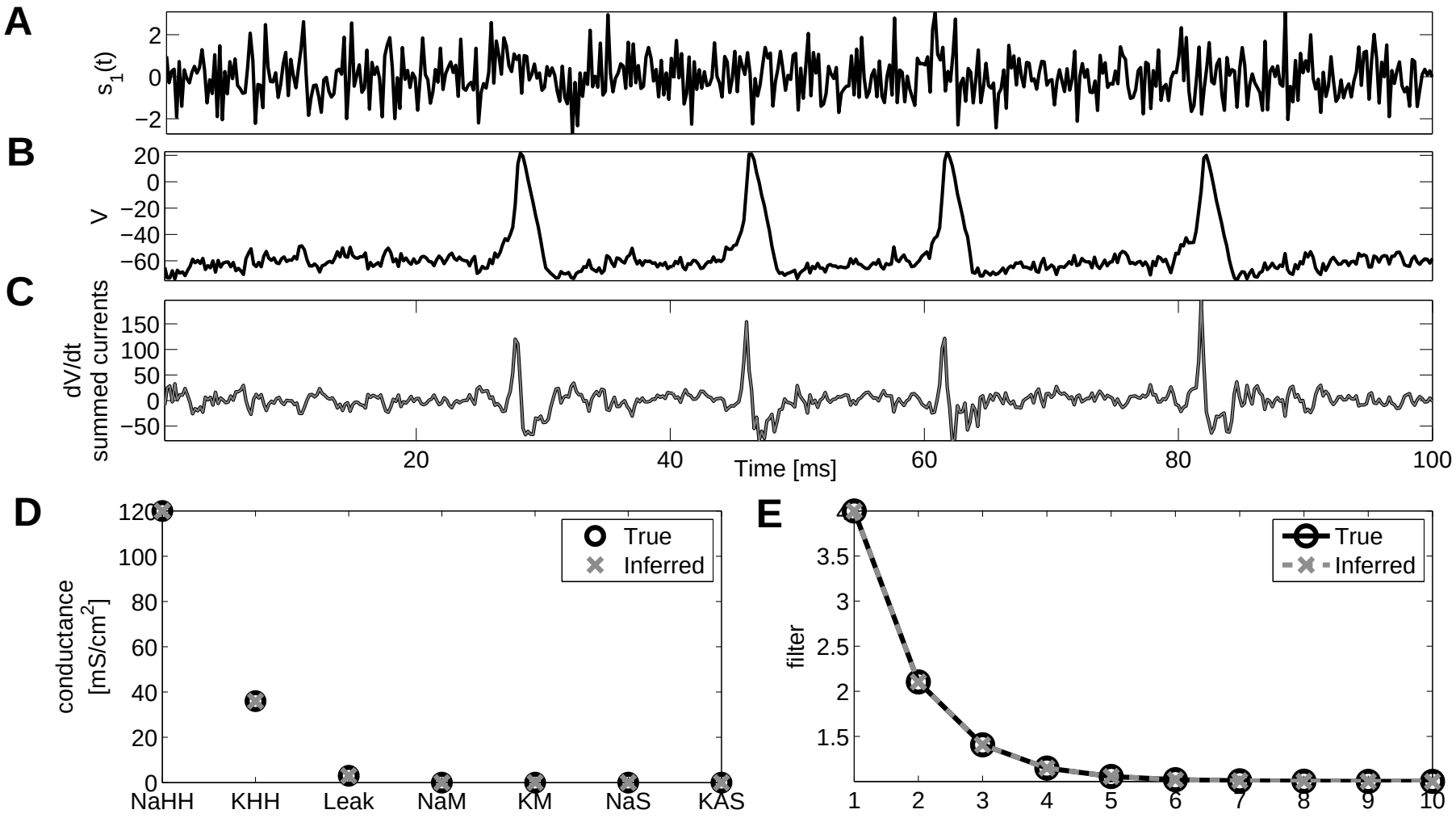


Estimating synaptic inputs given $V(t)$

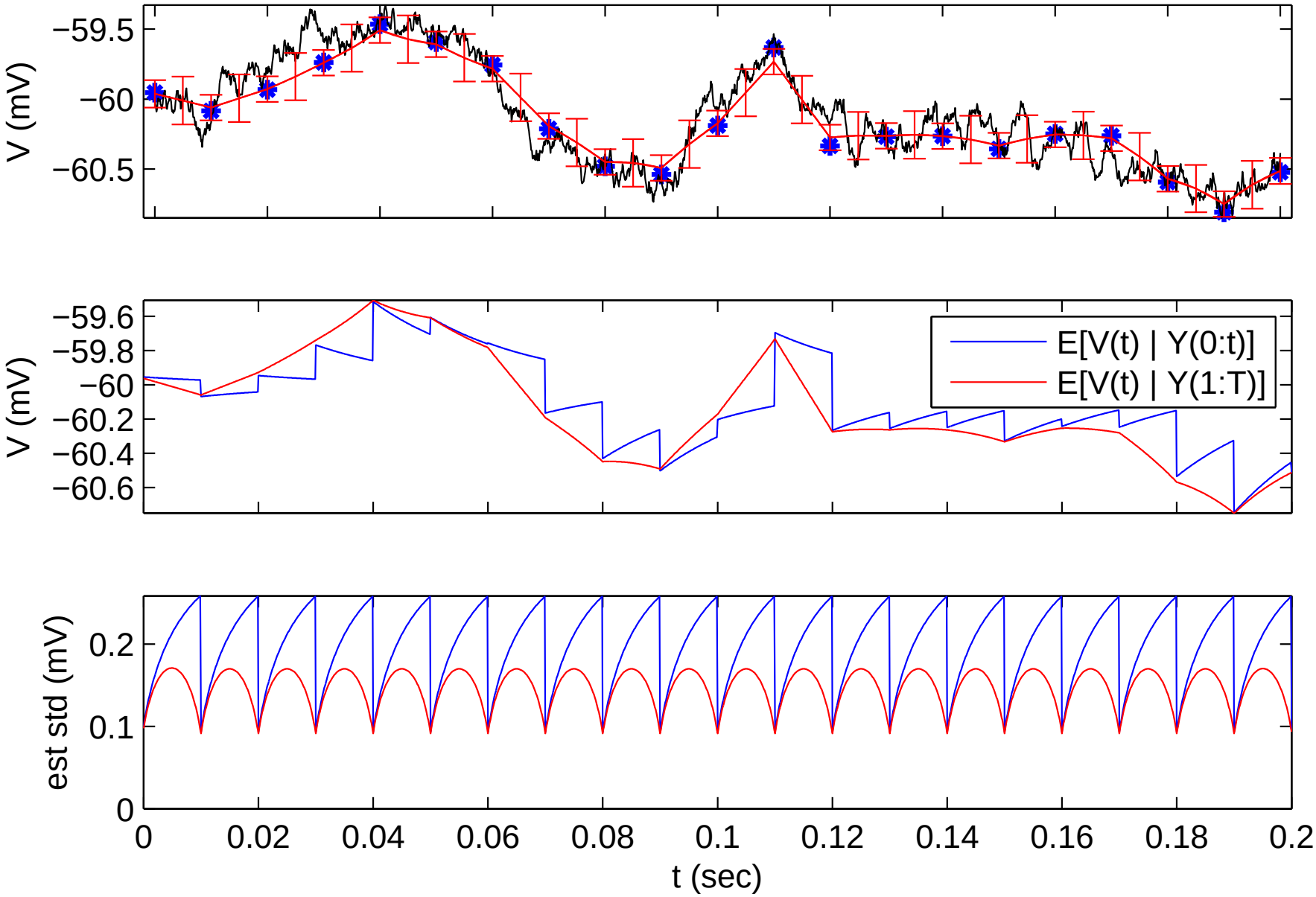


Estimating stimulus effects

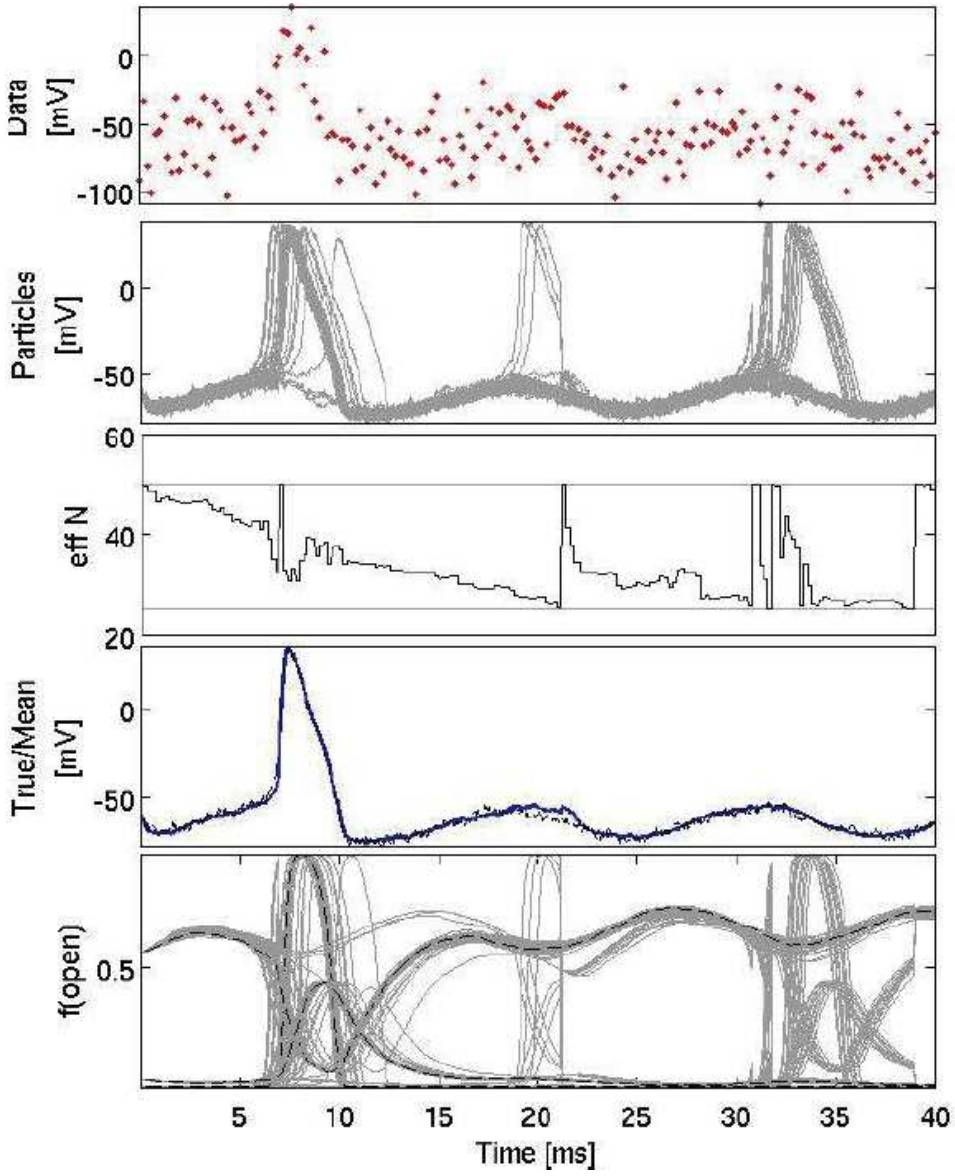
$$dV/dt = I_{channel} + \vec{k} \cdot \vec{x}(t) + \sigma N_t$$



Dealing with incomplete observations: Kalman filter

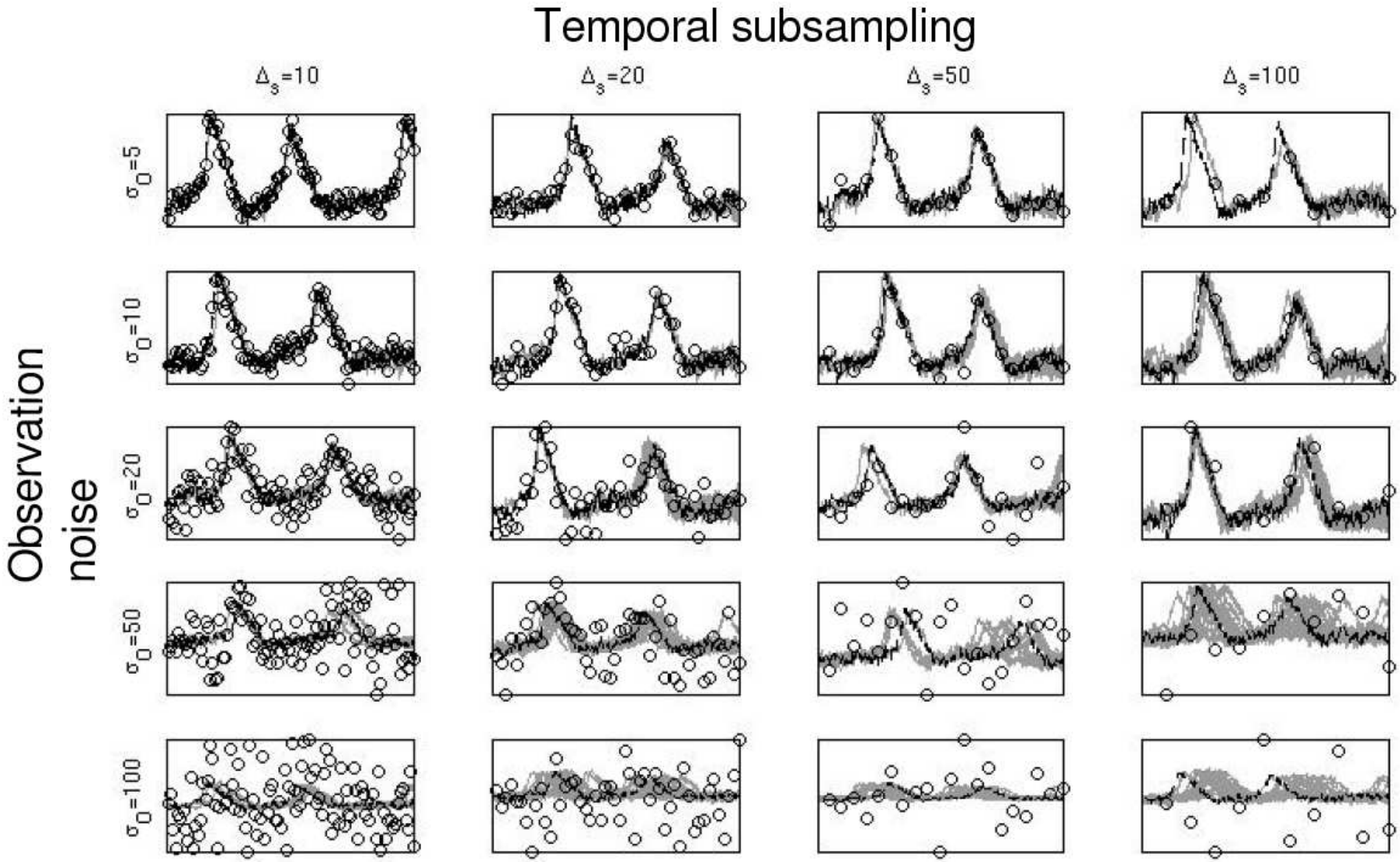


Smoothing given nonlinear dynamics

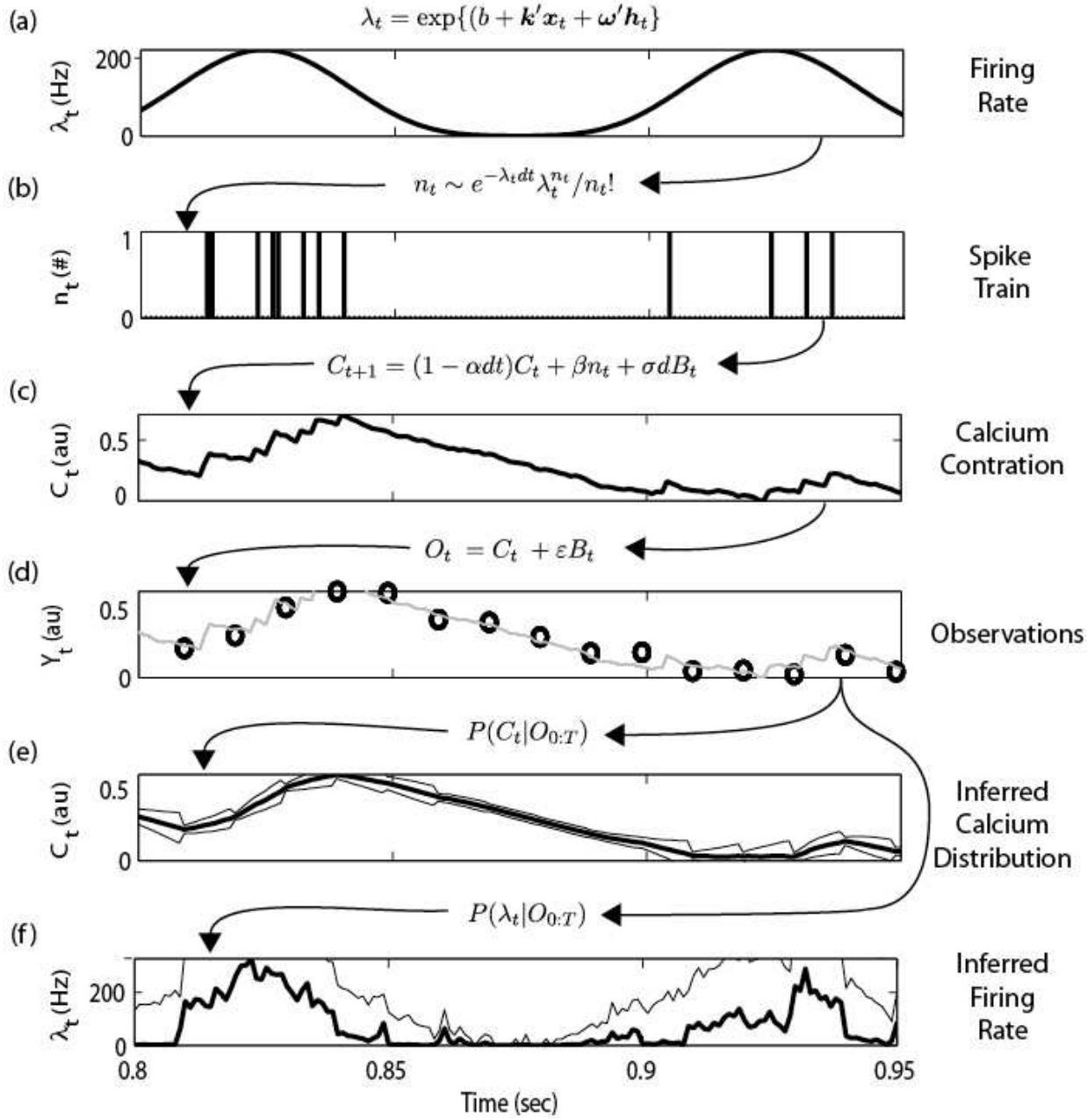


— via particle filtering (Huys and Paninski, 2006)

Subsampling and noise

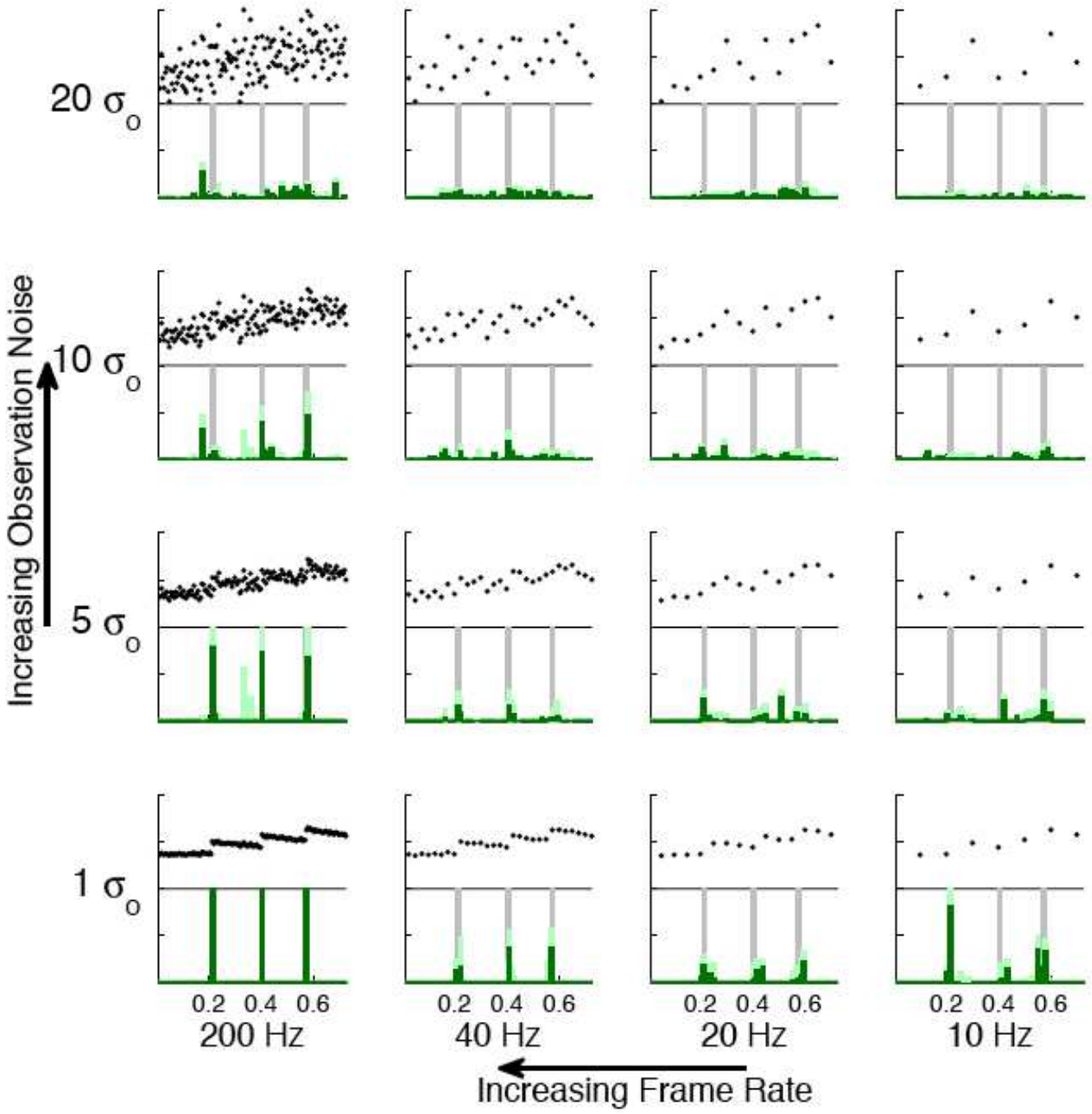


Inferring spike rates from calcium observations



(Vogelstein et al., 2007)

Inferring spike rates from calcium observations



Conclusions

Advantages of model-based approach:

- Flexibility of generative probabilistic framework
- Direct biophysical interpretability of estimated parameters
- Connections to statistical decoding methods, optimal experimental design (Paninski et al., 2008)
- Direct quantification of uncertainty

Next steps:

- Further applications to data
- Further relaxation of assumptions

Collaborators

Theory and numerical methods

- Y. Ahmadian, S. Escola, G. Fudenberg, Q. Huys, J. Kulkarni, M. Nikitchenko, K. Rahnama, G. Szirtes, T. Toyoizumi, Columbia
- E. Simoncelli, NYU
- A. Haith, C. Williams, Edinburgh
- M. Ahrens, J. Pillow, M. Sahani, Gatsby
- J. Lewi, Georgia Tech
- J. Vogelstein, Johns Hopkins

Retinal physiology

- E.J. Chichilnisky, J. Shlens, V. Uzzell, Salk

Cortical *in vitro* physiology

- B. Lau and A. Reyes, NYU

References

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.

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. (2006). Model-based optimal interpolation and filtering for noisy, intermittent biophysical recordings. *CNS*06 Meeting, Edinburgh.*

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.

Nikitchenko, M. and Paninski, L. (2007). An expectation-maximization Fokker-Planck algorithm for the noisy integrate-and-fire model. *COSYNE.*

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. (2006a). The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *Journal of Computational Neuroscience*, 21:71–87.

Paninski, L. (2006b). The spike-triggered average of the integrate-and-fire cell driven by Gaussian white noise. *Neural Computation*, 18:2592–2616.

Paninski, L., Haith, A., Pillow, J., and Williams, C. (2007). Integral equation methods for computing likelihoods and their derivatives in the stochastic integrate-and-fire model. *Journal of Computational Neuroscience*, In press.

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 Lewi, J. (2008). 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.

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.

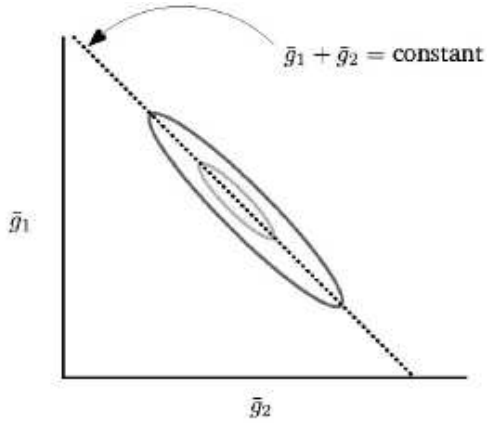
Pillow, J., Paninski, L., Uzzell, V., Simoncelli, E., and Chichilnisky, E. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25:11003–11013.

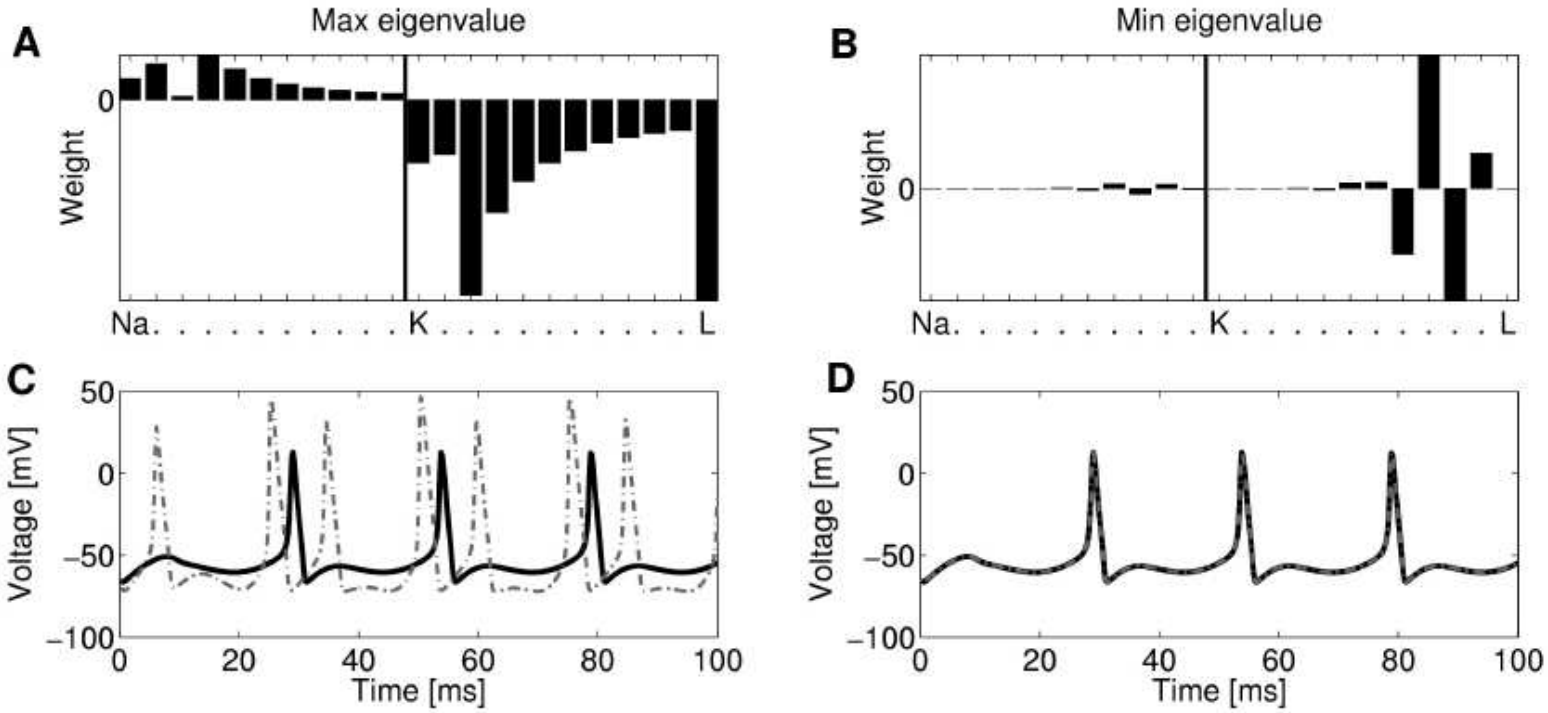
Simoncelli, E., Paninski, L., Pillow, J., and Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. In *The Cognitive Neurosciences*. MIT Press, 3rd edition.

Vogelstein, J., Zhang, K., Jedynak, B., and Paninski, L. (2007). Inferring the structure of populations of neurons using a sequential Monte Carlo EM algorithm. *COSYNE*.

Measuring uncertainty in channel densities

$$\begin{aligned}\hat{\mathbf{a}} &= \arg\min_{\mathbf{a}} ||\dot{\mathbf{V}} - \mathbf{J}\mathbf{a}||^2 \\ &= \arg\min_{\mathbf{a}} \mathbf{a}^T\mathbf{H}\mathbf{a} - 2\mathbf{a}^T\mathbf{f} \quad s.t. \quad a_i \geq 0 \forall i\end{aligned}$$





Forward equation method

Let $P(V,t) =$ density of $V(t)$.

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

under boundary conditions

$$P(V, t_{i-1}) = \delta(V - V_{reset}),$$

$$P(V_{th}, t) = 0;$$

$$V_{rest}(t) \equiv \frac{1}{g}\left(I_{DC} + \vec{k}\cdot\vec{x}(t) + \sum_{j=0}^{i-1} h(t - t_j)\right).$$

Linear PDE may be solved efficiently via, e.g., Crank-Nicholson.

$$p_{t_{i-1},V_{reset}\to V_{th}}(t_i|\{t_j\}_{j<i},\theta,\vec{x}) = -\frac{\partial}{\partial t}\int P(V,t)dV.$$

Integral equation method

$p(t)$ solves several Volterra integral equations (via "method of images"; goes back to Schrodinger), e.g.:

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

$G_\theta(y, t|x, s) =$ probability of $V(t) = y$, given $V(s) = x$
(Gaussian; analytically computable for OU process)

Discretized linear system is lower-triangular: $O((\frac{T}{dt})^2)$ solution

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