Continuous-time models for neural encoding

Liam Paninski

Gatsby Computational Neuroscience Unit
University College London
http://www.gatsby.ucl.ac.uk/∼liam
liam@gatsby.ucl.ac.uk

November 5, 2004

Review...

Main idea: build flexible, fittable models for input-output relationship $p(\{t_i\}|\vec{x})$

Introduced cascade idea, geometric intuition

Discussed STA, STC fitting procedures

Extended cascade idea to include interneuronal effects

Illustrated importance of "complexity control," reducing overfitting

Continuous-time neural encoding models

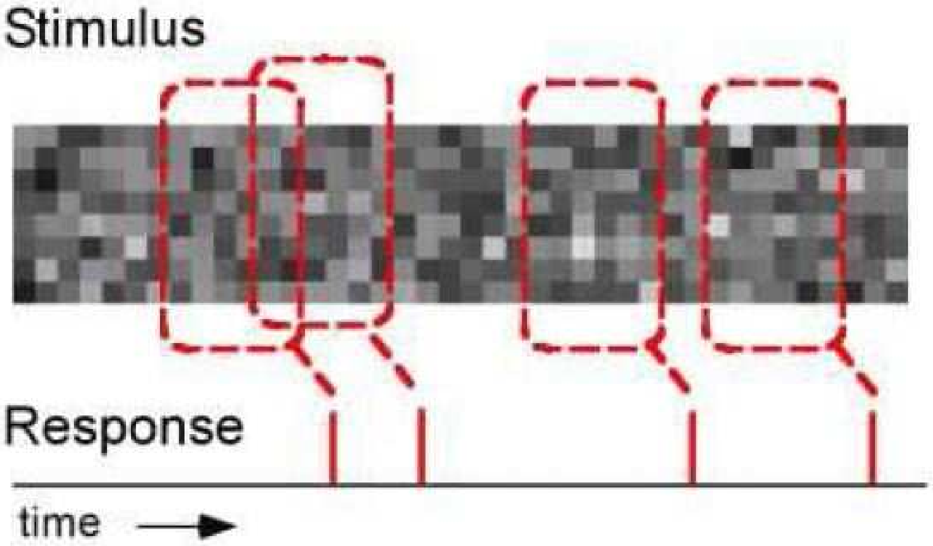
Sounds harder than discrete time. Why would this be useful?

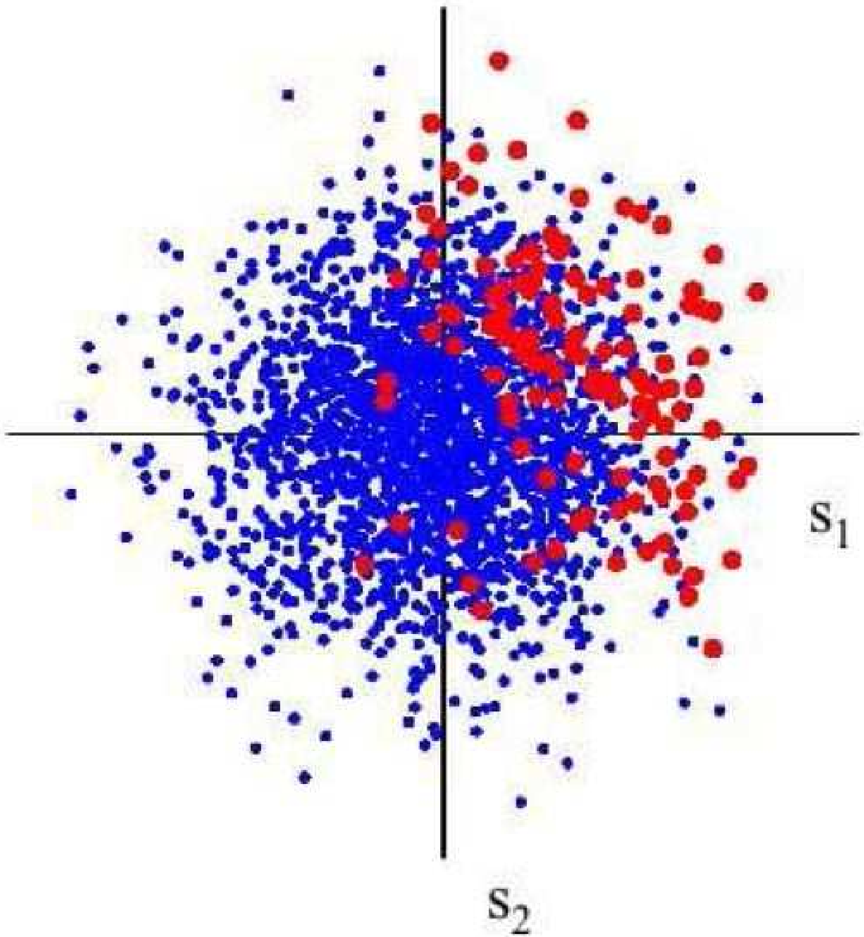
— Makes it much easier to model spike time-dependent effects: refractory periods, burstiness, adaptation, etc.

— LNP (linear - nonlinear - Poisson) model ignores spike history

(Kass et al., 2003; Brown et al., 2002)

Spike-triggered ensemble idea ignores response history





Starting point: point process likelihood

Models give "intensity function" $\lambda(t) = \lambda(t|\vec{x}, spike\ history)$ — "instantaneous firing rate."

$$
\begin{aligned}
L(spikes = \{t_i\}|\{\lambda(t)\}) &= \log p(spikes = \{t_i\}|\{\lambda(t)\}) \\
&\sim \log\left(e^{-\int_0^T \lambda(t)dt} \prod_i \lambda(t_i)\right) \\
&= \sum_i \log \lambda(t_i) - \int_0^T \lambda(t)dt
\end{aligned}
$$

Reminder: where does this come from?

Discrete-time likelihood

Prob of spiking in small bin dt:

$$p(spike \in [t, t + dt] | \lambda(t)) \approx \lambda(t) dt.$$

Discrete-time likelihood:

$$\begin{aligned} L_{discrete} &\sim \log \left((\prod_i \lambda(t_i))(\prod_j (1 - \lambda(t_j) dt)) \right) \\ &= \sum_i \log \lambda(t_i) + \sum_j \log(1 - \lambda(t_j) dt) \end{aligned}$$

i = spikes; j = no spikes

Continuous limit

Use $\log(1 - \lambda(t)dt) \approx -\lambda(t)dt$:

$$\begin{aligned}
\sum_i \log \lambda(t_i) + \sum_j \log(1 - \lambda(t_j)dt) &\approx \sum_i \log \lambda(t_i) - \sum_j \lambda(t_j)dt \\
&\approx \sum_i \log \lambda(t_i) - \int_0^T \lambda(t)dt
\end{aligned}$$

Maximum likelihood

Choose model parameters θ to maximize probability of data $\{t_i\}$, under model θ:

$$\begin{aligned} L(\theta) &= L(spikes = \{t_i\} | \{\lambda_\theta(t)\}) \\ &= \sum_i \log \lambda_\theta(t_i) - \int_0^T \lambda_\theta(t) dt. \end{aligned}$$

Maximum likelihood (ML) estimators have good theoretical properties assuming model is flexible enough (model space Θ contains some θ_0 that gives a good model for data).

In particular, ML provides asymptotically optimal θ_0, given enough data ("consistency") and gets close to this optimal θ_0 asymptotically as fast as any other estimator ("efficiency").

Concavity

Idea: 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.

— Find flexible models whose loglikelihoods are guaranteed to be concave.

Cascade models, again

$$\lambda(t) = f(\vec{k} \cdot \vec{x}(t))$$

$$L(\vec{k}) = \sum_i \log f(\vec{k} \cdot \vec{x}(t_i)) - \int_0^T f(\vec{k} \cdot \vec{x}(t))dt$$

Sums, integrals of concave functions are concave $\Longrightarrow$

If $f(u)$ convex, $\log f(u)$ concave, $L(\vec{k})$ is always concave.

— models with $f(u)$ convex, $\log f(u)$ concave are easy to fit.

Examples

$f(u)$ convex, $\log f(u)$ concave:

$f(u) = e^u$.

$f(u) =$ linear rectifier

$f(u) = u^\alpha$, $\alpha \geq 1$

$$f(u) = \begin{cases} e^u & u < 0 \\ 1 + u & u \geq 0 \end{cases} .$$

Connection to STA

Let's examine likelihood for this model.

$$L(\vec{k}) = \sum_i \log f(\vec{k} \cdot \vec{x}(t_i)) - \int_0^T f(\vec{k} \cdot \vec{x}(t))dt$$

At maximum (MLE), gradient of likelihood must be 0.

$$\begin{aligned} \nabla L(\vec{k}_{MLE}) &= \sum_i \frac{f'(\vec{k}_{MLE} \cdot \vec{x}(t_i))}{f(\vec{k}_{MLE} \cdot \vec{x}(t_i))} \vec{x}(t_i) - \int_0^T \vec{x}(t) f'(\vec{k}_{MLE} \cdot \vec{x}(t))dt \\ &= 0 \end{aligned}$$

Connection to STA

Gradient equation for MLE:

$$\sum_i \frac{f'(\vec{k}_{MLE} \cdot \vec{x}(t_i))}{f(\vec{k}_{MLE} \cdot \vec{x}(t_i))} \vec{x}(t_i) = \int_0^T \vec{x}(t) f'(\vec{k}_{MLE} \cdot \vec{x}(t)) dt$$

Two weighted averages:

$\sum_i \frac{f'}{f}(\vec{k}_{MLE} \cdot \vec{x}(t_i))\vec{x}(t_i)$ = a weighted STA

$\int_0^T \vec{x}(t) f'(\vec{k}_{MLE} \cdot \vec{x}(t)) dt \approx E_{p(\vec{x})}\left(\vec{x} f'(\vec{k}_{MLE} \cdot \vec{x})\right)$

Connection to STA

Geometric argument: under elliptically symmetric $p(\vec{x})$,

$$E_{p(\vec{x})}\left(\vec{x}f'(\vec{k}_{MLE}\cdot\vec{x})\right)\sim C\vec{k}_{MLE}$$

$C =$ stimulus correlation matrix, $E(\vec{x}^t\vec{x})$.

So

$$\vec{k}_{MLE}\approx C^{-1}\sum_i \frac{f'}{f}(\vec{k}_{MLE}\cdot\vec{x}(t_i))\vec{x}(t_i)$$

i.e., MLE is like a weighted correlation-corrected STA.

Making the MLE more robust to bias

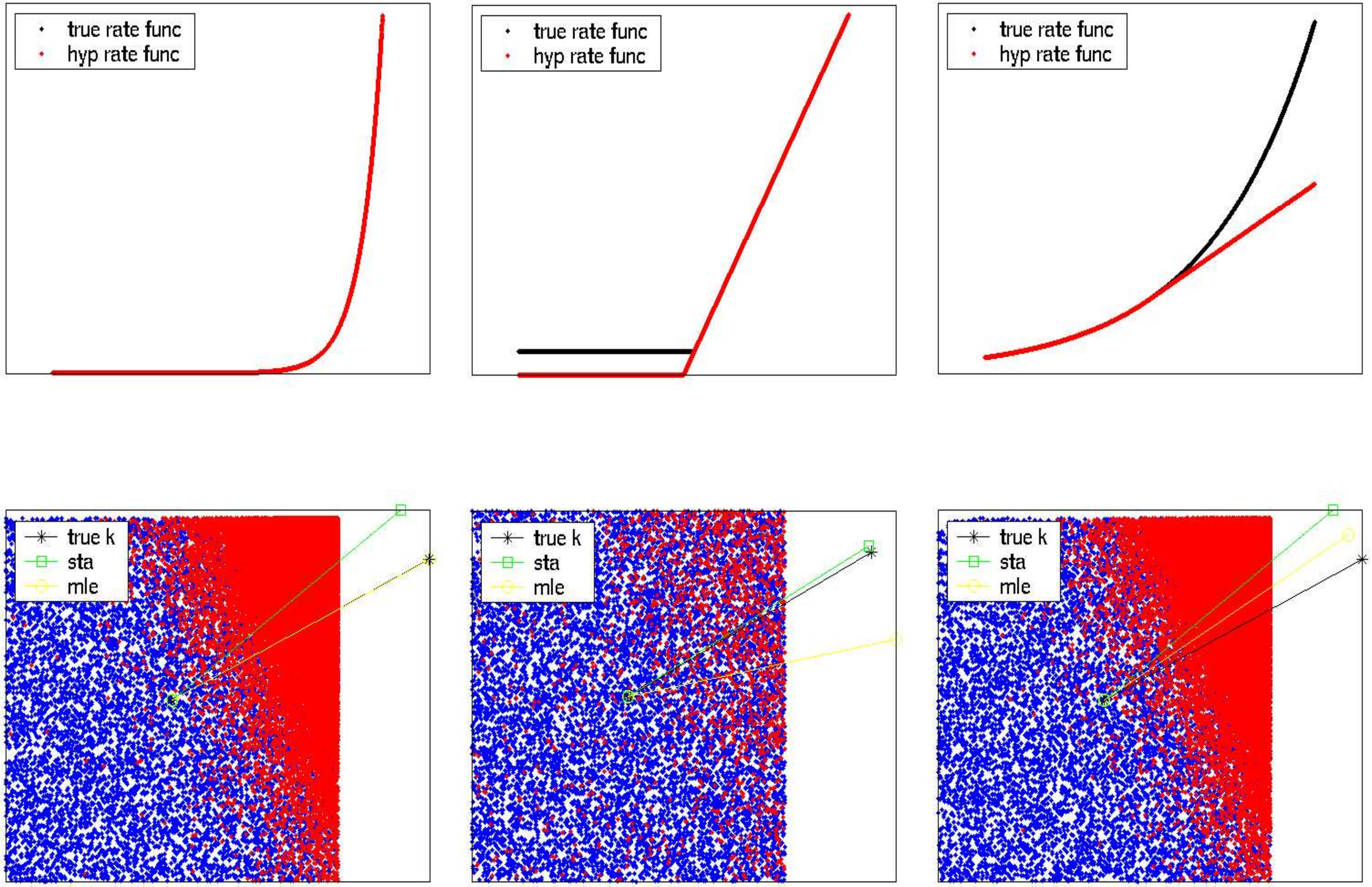
If $f(u) = e^u$, $\frac{f'}{f} = 1$, so

$$\sum_i \frac{f'}{f}(\vec{k} \cdot \vec{x}(t_i))\vec{x}(t_i) = \vec{k}_{STA}$$

(assuming elliptic symmetry).

If f increases more slowly for large u, decreases more slowly for small u: more robust weighted STA.

Making the MLE more robust to bias



(Paninski, 2004)

Incorporating spike-history dependence: multiplicative models

First idea: take your model, whatever it is, and tack on a mulitplicative term.

Start with model: $\lambda_0(t|\vec{x})$

New model:

$$\lambda_1(t|\vec{x}, spike\ history) = \lambda_0(t|\vec{x})r(t - t_{i-1})$$

$t - t_{i-1}$ = time since last spike.

$r(t - t_{i-1})$ = absolute and/or relative refractory period, burstiness, etc.

Multiplicative models

How to estimate $r(t - t_{i-1})$?

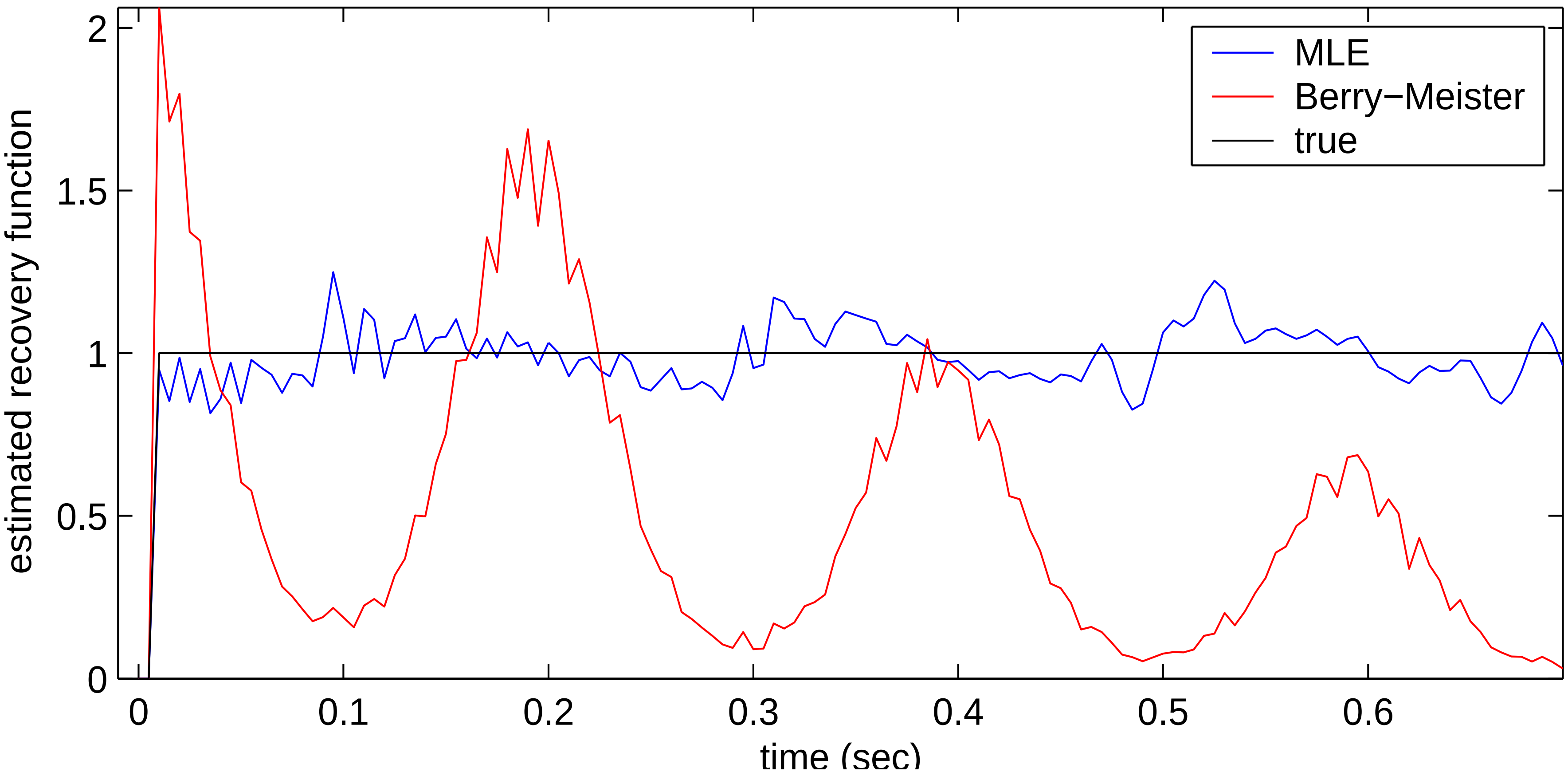
Loglikelihood:

$$
\begin{aligned}
L(\vec{k}) &= \sum_i \log \lambda_1(t_i) - \int_0^T \lambda_1(t)dt \\
&= \sum_i \log(\lambda_0(t_i) r(t_i - t_{i-1})) - \int_0^T (\lambda_0(t) r(t - t_{i-1}))dt \\
&= K + \sum_i \left(\log r(t_i - t_{i-1}) - \int_{t_{i-1}}^{t_i} \lambda_0(t) r(t - t_i) dt \right)
\end{aligned}
$$

Key point: this is concave in $r(.)$. Unique maximum.

Can find MLE for $r(.)$ analytically (take gradient, set to zero; exercise).

Estimating multiplicative refractory term



(Berry and Meister, 1998; Paninski, 2004)

Incorporating spike-history dependence: subthreshold models

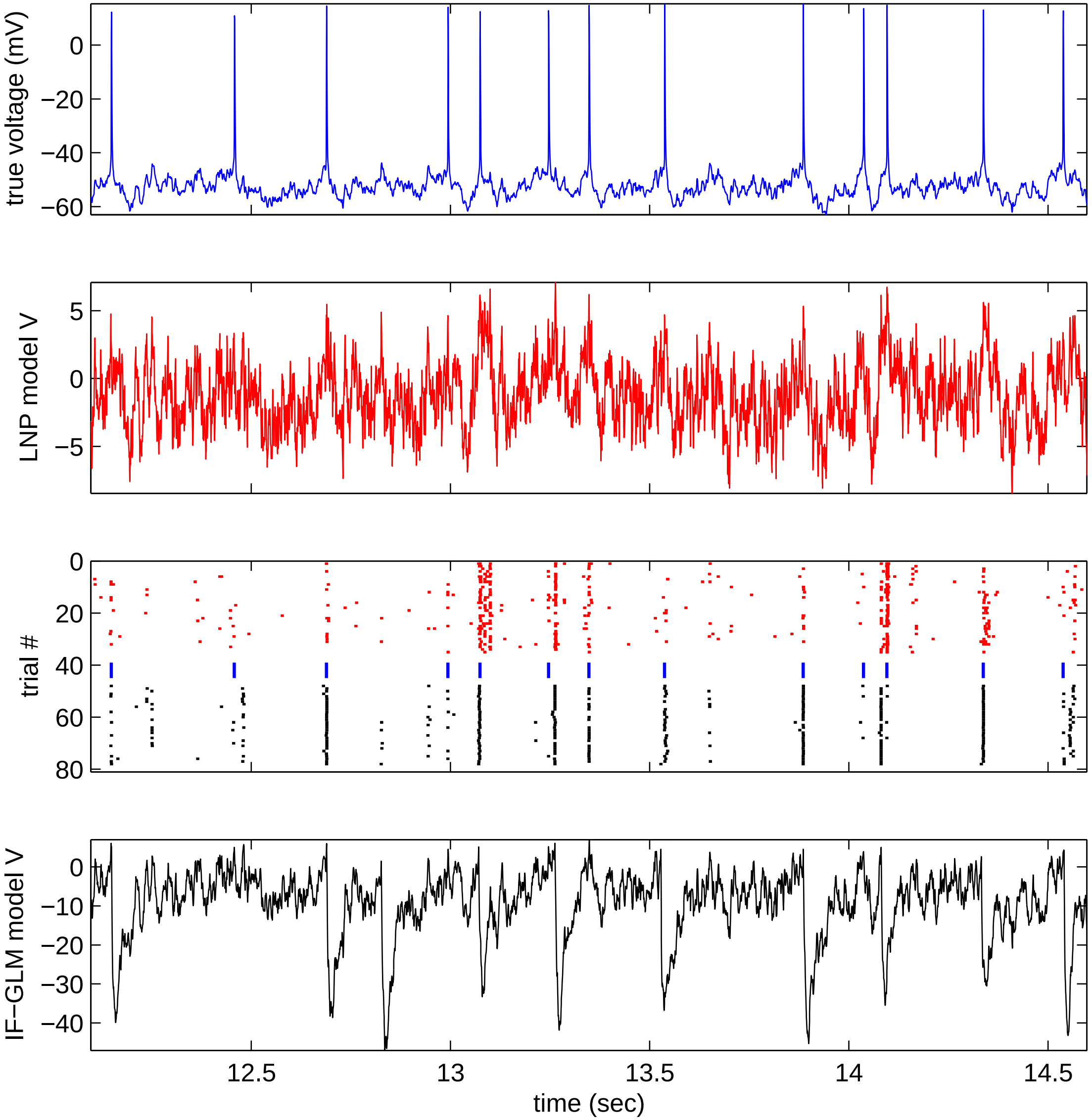
Multiplicative model only looks back to last spike. How to incorporate multi-spike effects?

$$\lambda(t) = f\left(\vec{k} \cdot \vec{x}(t) + \sum_{j=-\infty}^{0} h(t - t_j)\right)$$

$h(t)$ is "current" injected after every spike; can model refractoriness, burstiness, adaptation, etc.

$f(u)$ convex, $\log f(u)$ concave $\implies$
loglikelihood concave in $(\vec{k}, h(t)) \implies$
easy to solve for optimal $\vec{k}$ and $h(t)$ together

Modeling spike-history dependence greatly improves accuracy



(Paninski et al., 2003; Paninski, 2004)

Regularization, likelihood version

Instead of maximizing $L(\vec{k})$, maximize

$$L(\vec{k}) - \lambda Q(\vec{k}),$$

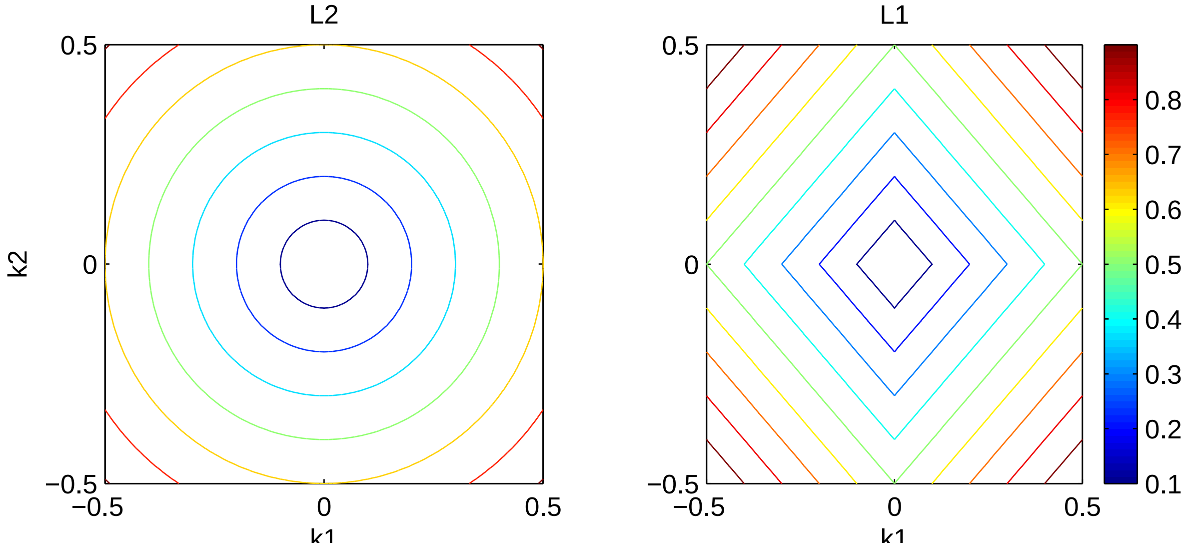
$Q(\vec{k})$ some convex "penalty" function. Doesn't have to be quadratic!

Bayesian interpretation: $-Q(\vec{k}) = \log(prior(\vec{k}))$. Maximizing penalized likelihood = maximum *a posteriori* estimate

Regularization, likelihood version

$Q(\vec{k}) = ||\vec{k}||_2 = \sum_i k_i^2$: radially-symmetric $\vec{k}$ penalty

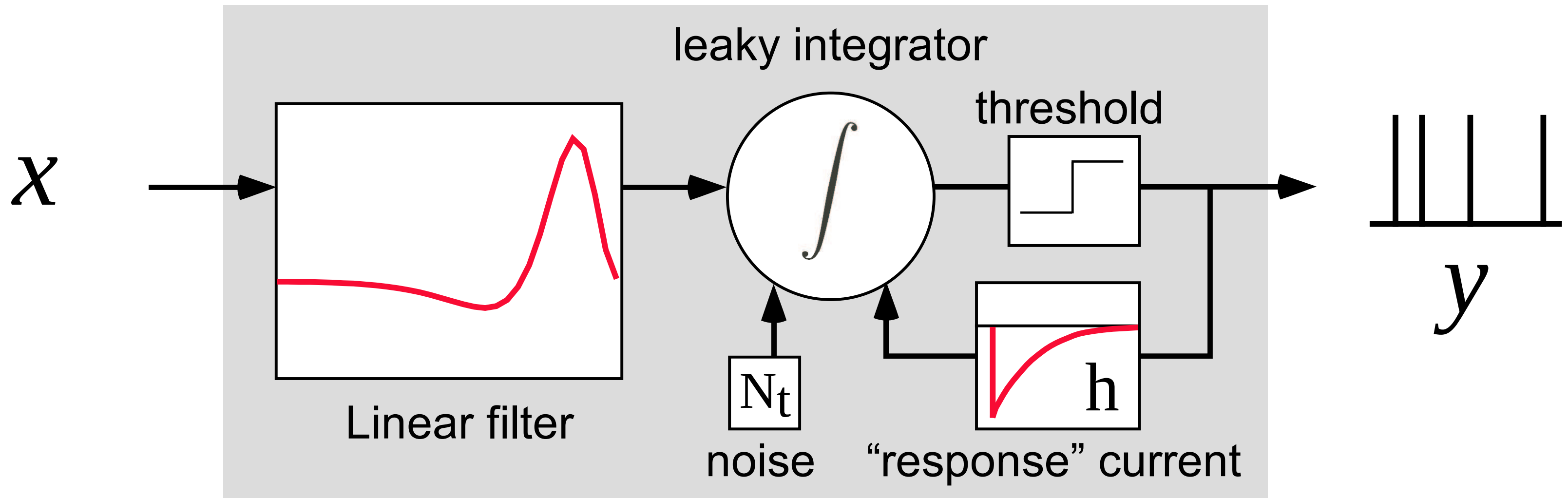
$Q(\vec{k}) = ||\vec{k}||_1 = \sum_i |k_i|$: $\vec{k}$ with many positive elements penalized more $\implies$ sparser $\vec{k}_{reg}$



These generalized cascade models are flexible, powerful, and easy to estimate (due to concavity).

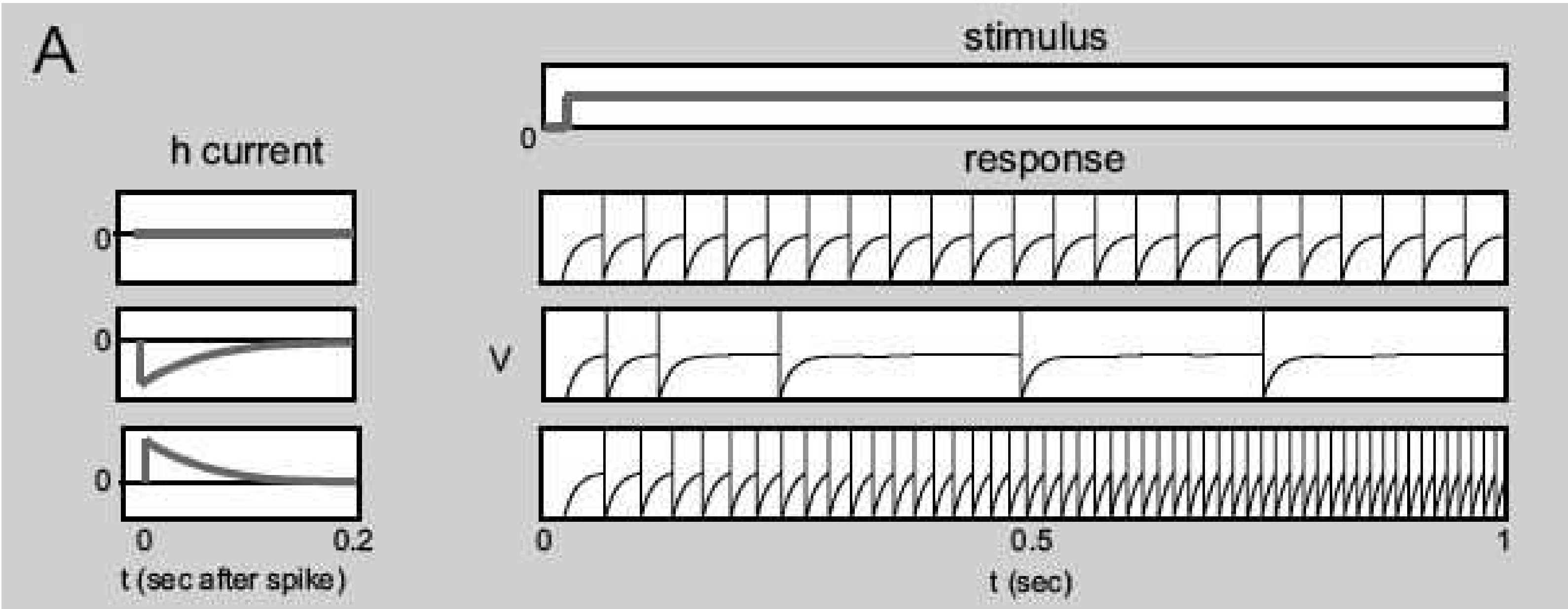
Can we make a connection to the underlying physiology?

Filtered integrate-and-fire model

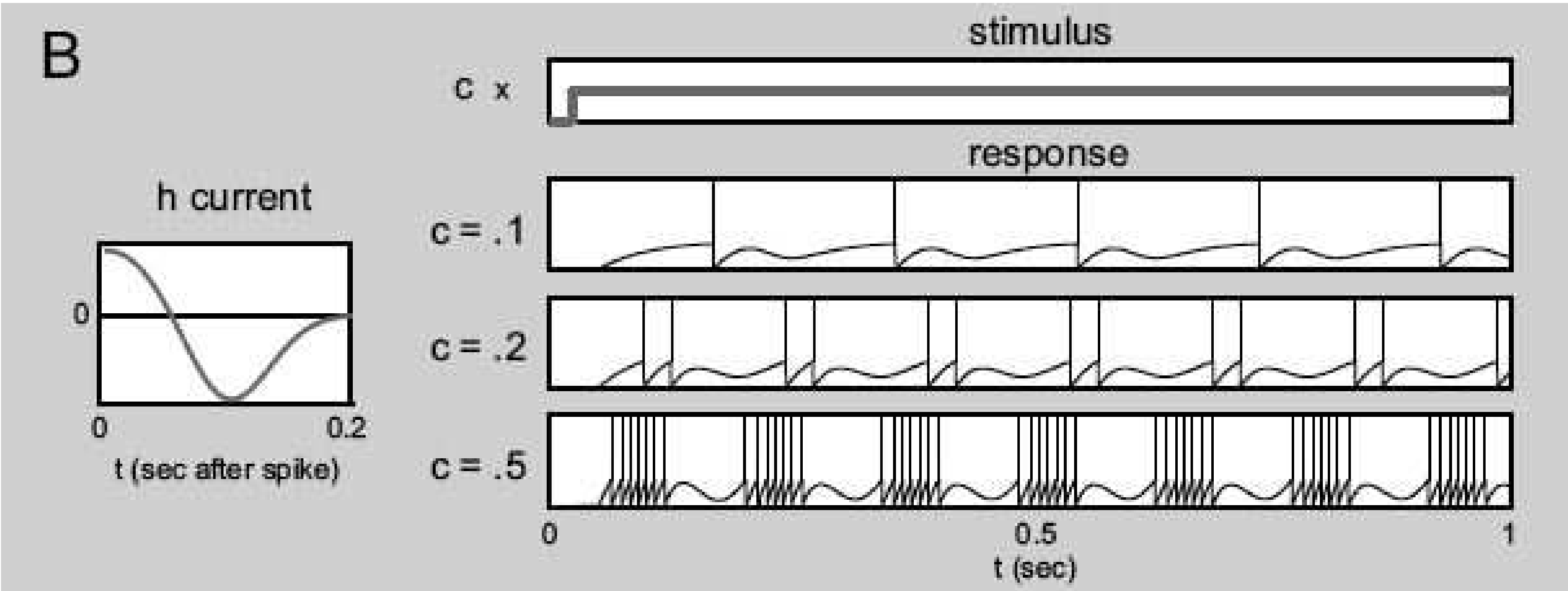


(Gerstner and Kistler, 2002; Paninski et al., 2004b)

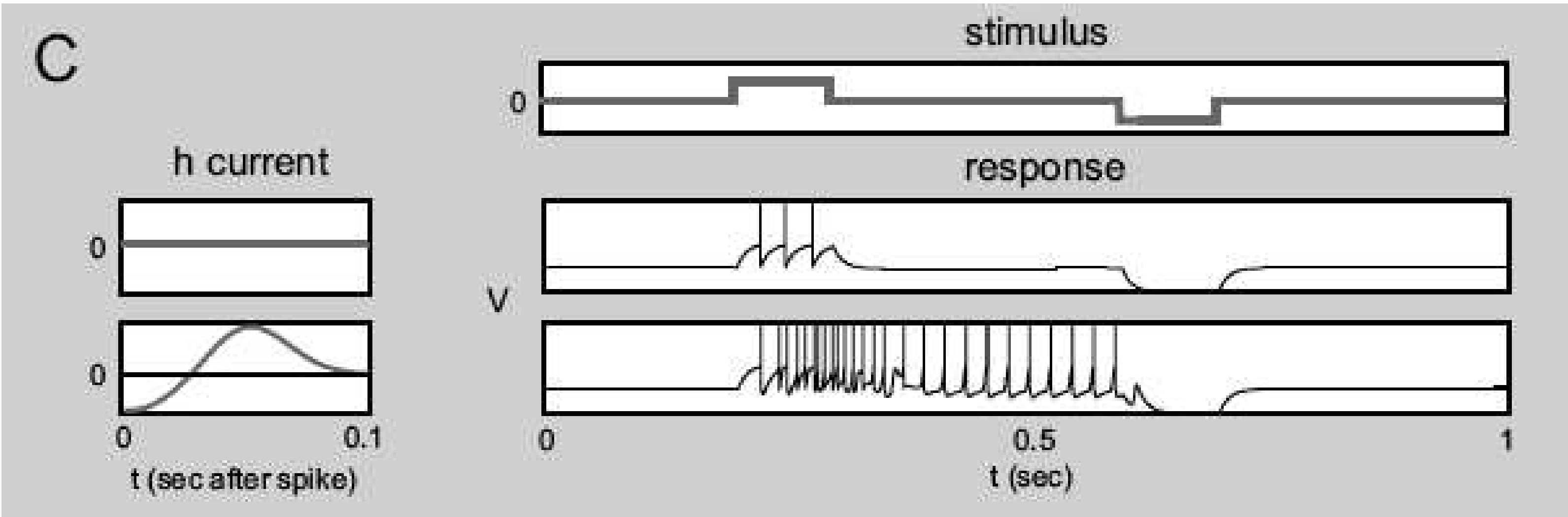
Model flexibility: Adaptation



Model flexibility: Burstiness



Model flexibility: Bistability



The Estimation problem

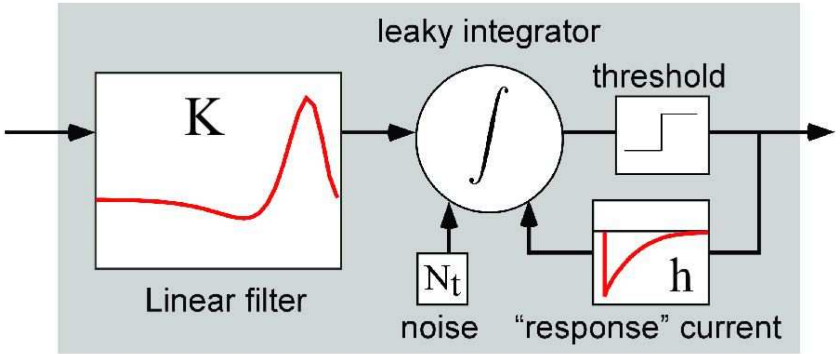
Learn the model parameters:

$\vec{K} =$ stimulus filter

$g =$ leak conductance

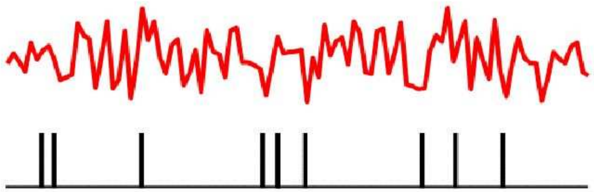
$\sigma^2 =$ noise variance

$\vec{h} =$ response current



From: stimulus train $x(t)$

spike times t_i

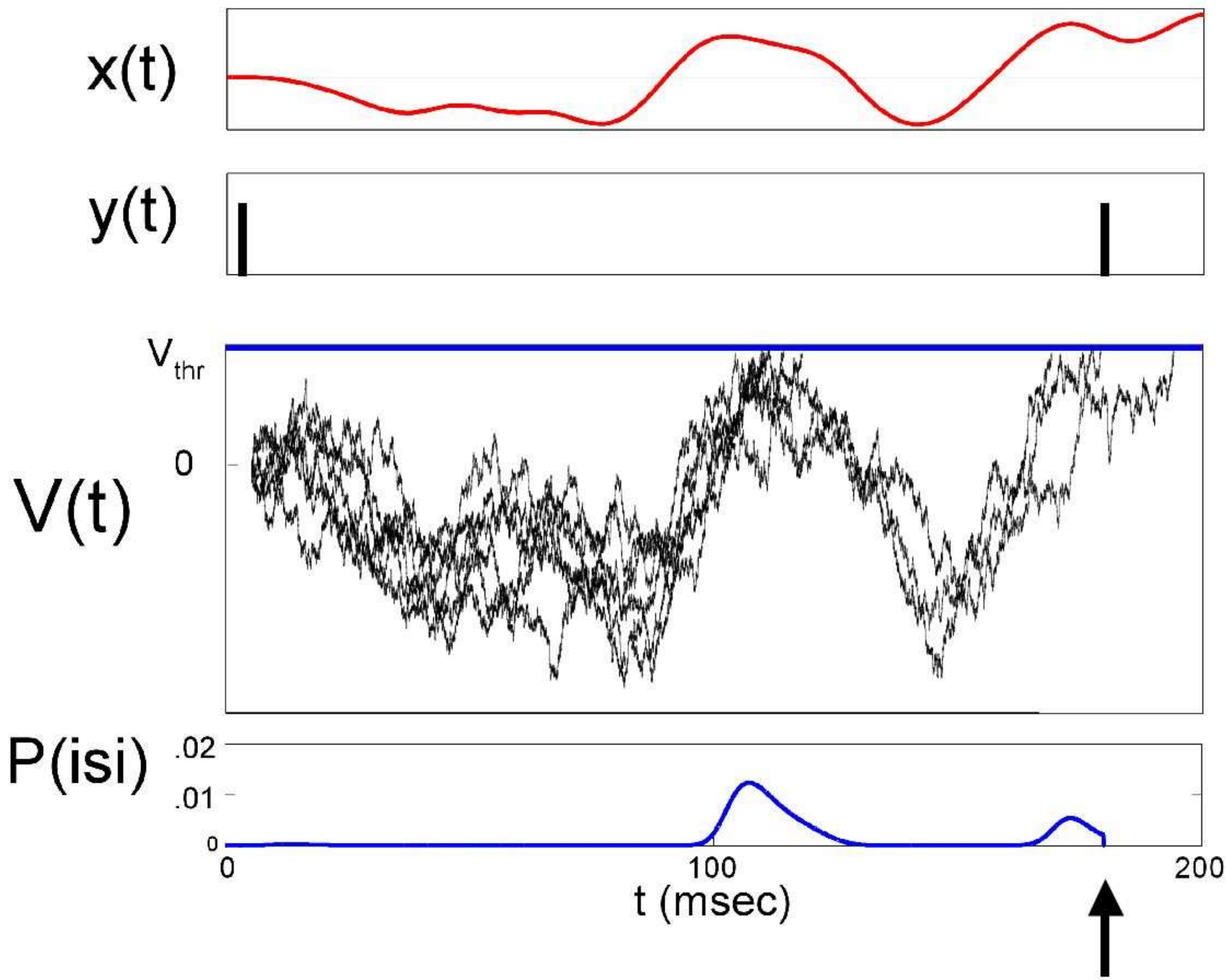


Solution: use Maximum Likelihood

(Paninski et al., 2004b)

Likelihood function

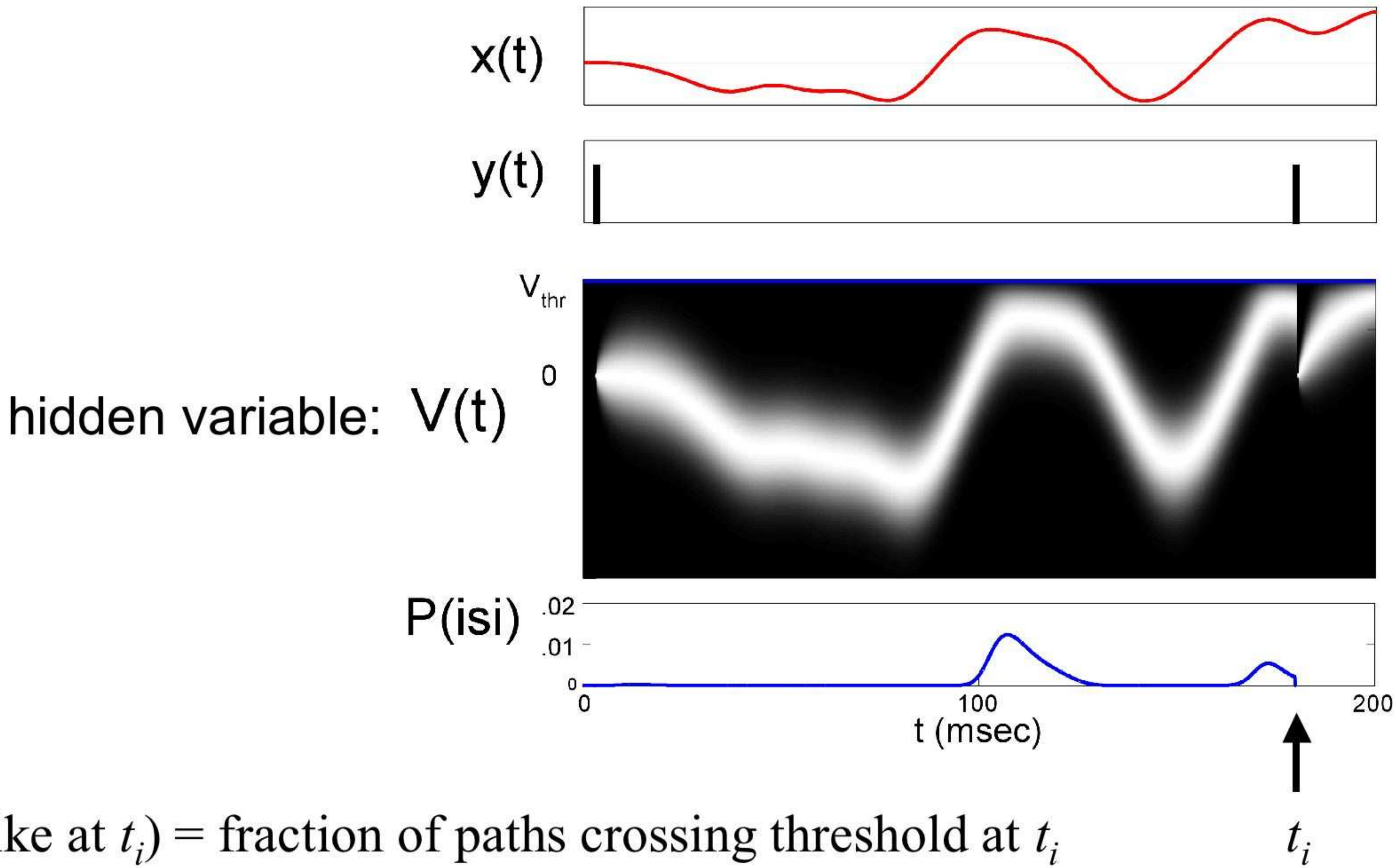
hidden variable:



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

t_i

Likelihood function



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

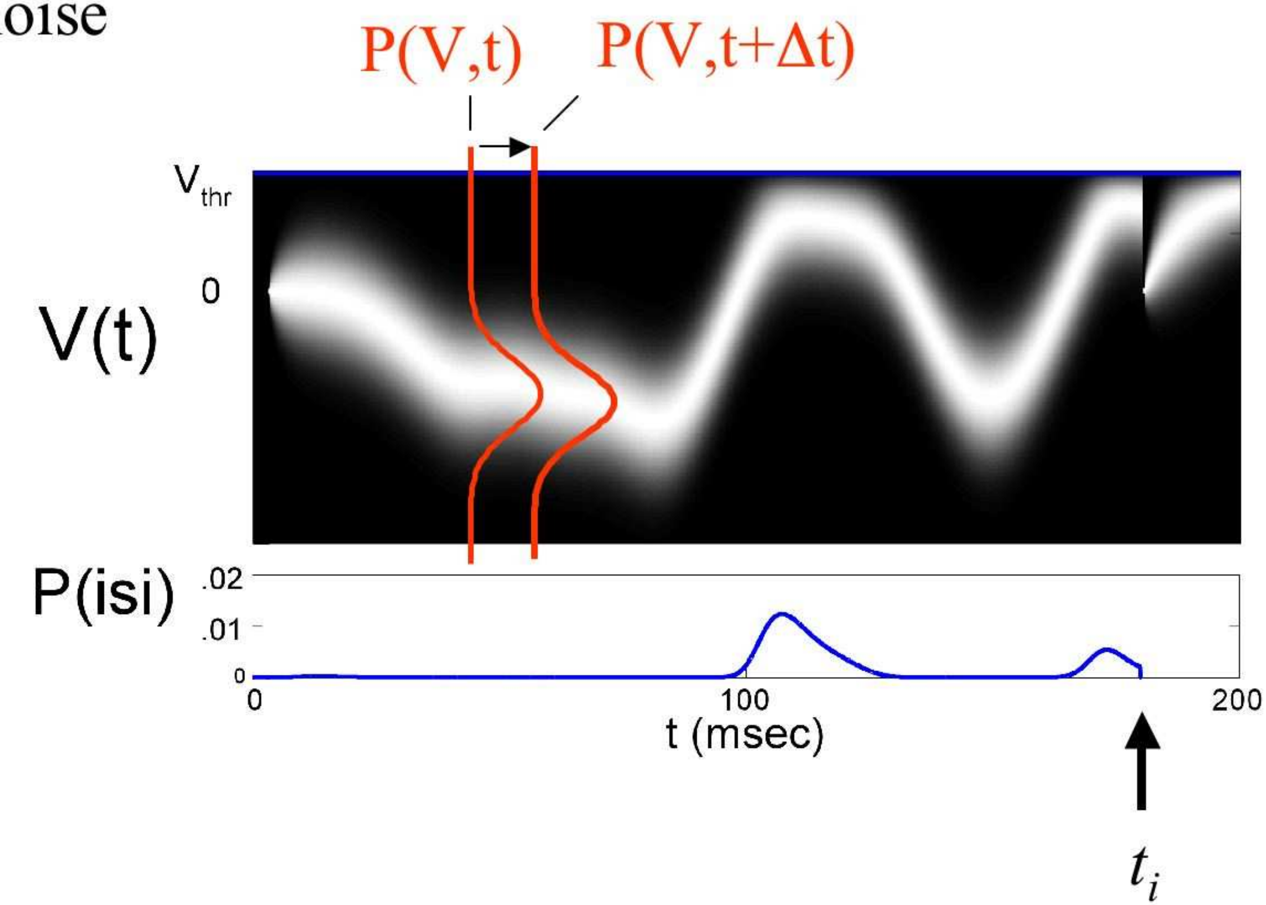
Computing Likelihood

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

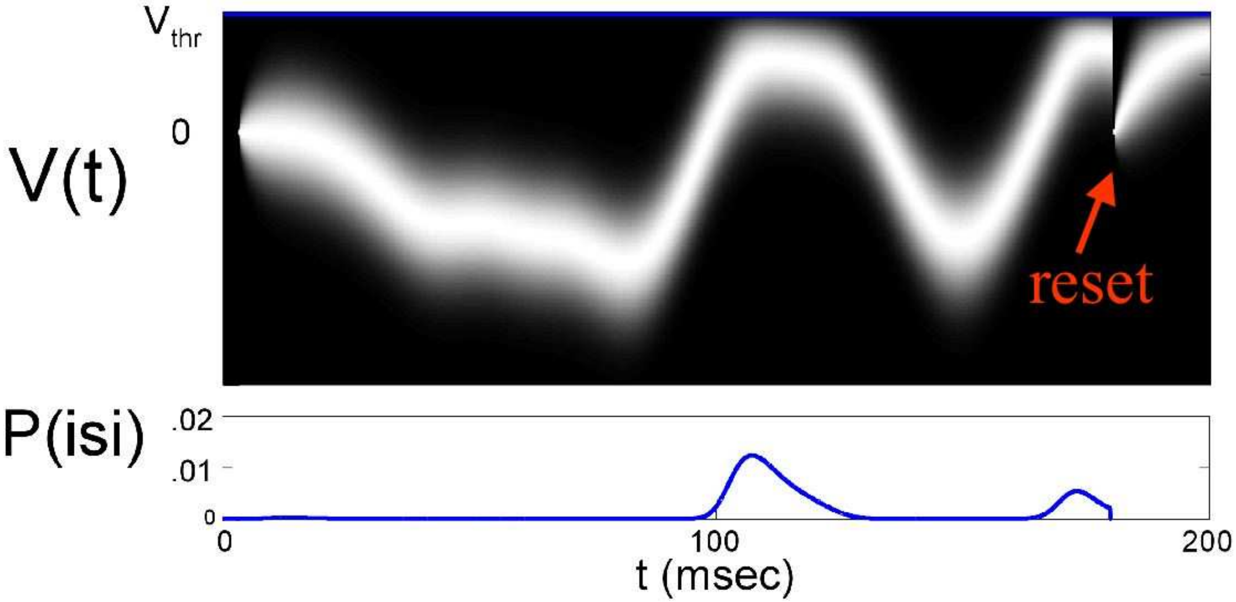
Diffusion Equation: $$\frac{\partial P(V,t)}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial V^2} + g\frac{\partial[(V-V_0)P]}{\partial V},$$

- linear dynamics
- additive Gaussian noise

fast methods for solving linear PDE

$\Rightarrow$

efficient procedure for computing likelihood



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

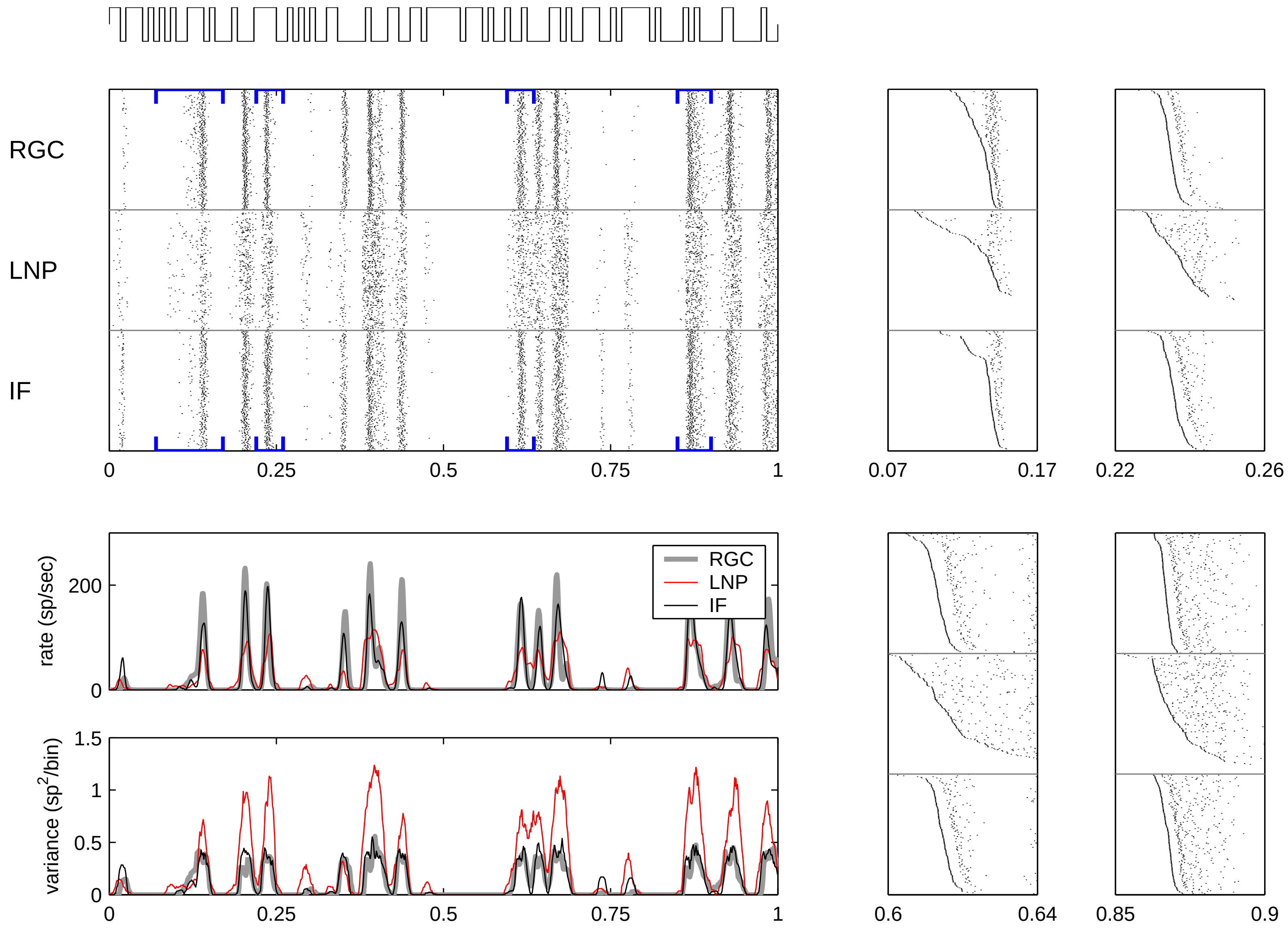
Maximizing the likelihood

- parameter space is large ($\approx$ 10 to 100 dimensions)
- parameters interact nonlinearly

Main Theorem: The log likelihood is concave in the parameters $\{\vec{K}, g, \sigma, \vec{h}\}$, for any data $\{x(t), t_i\}$

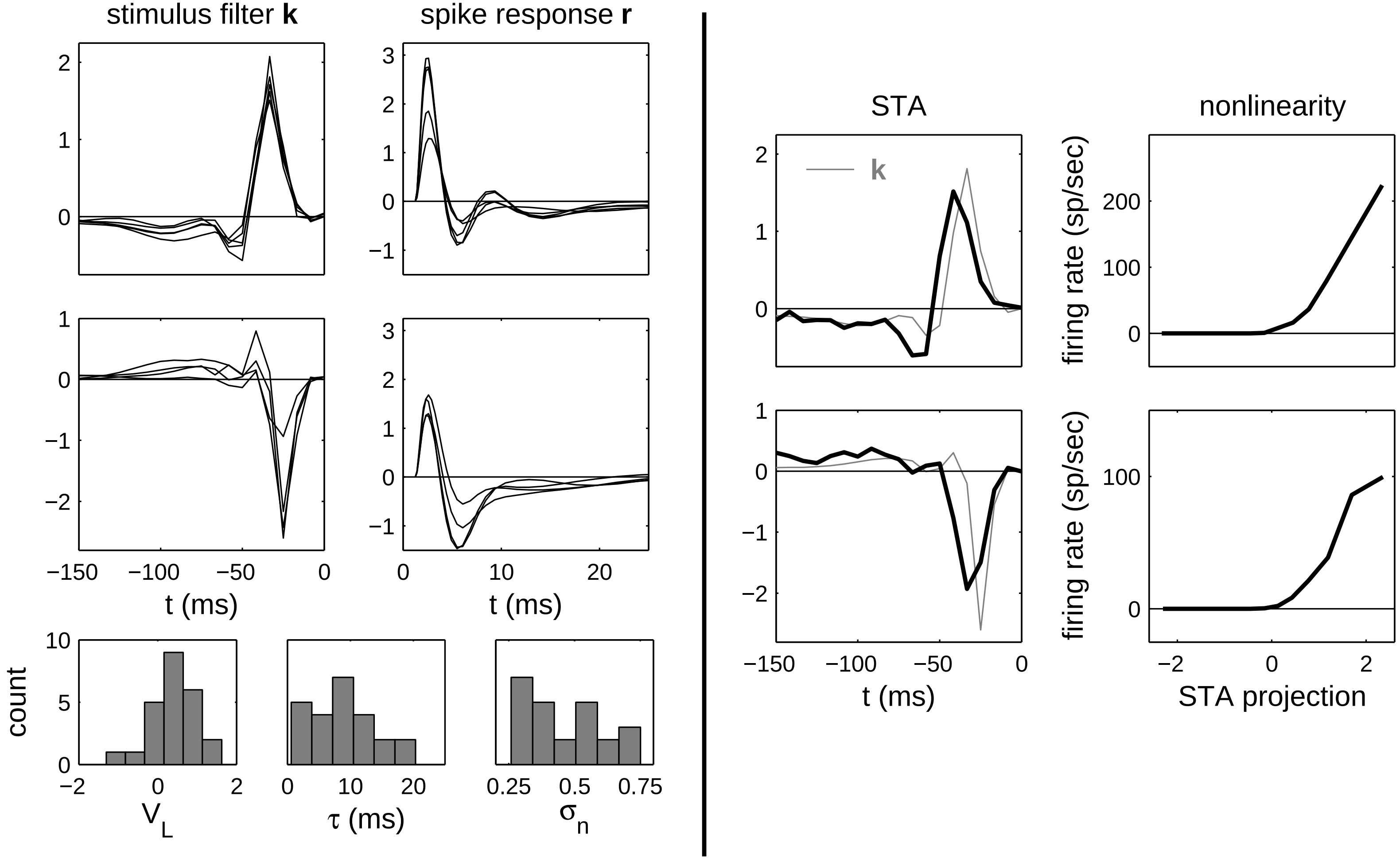
$\Rightarrow$ gradient ascent guaranteed to converge to global maximum!

Spike timing precision in retina



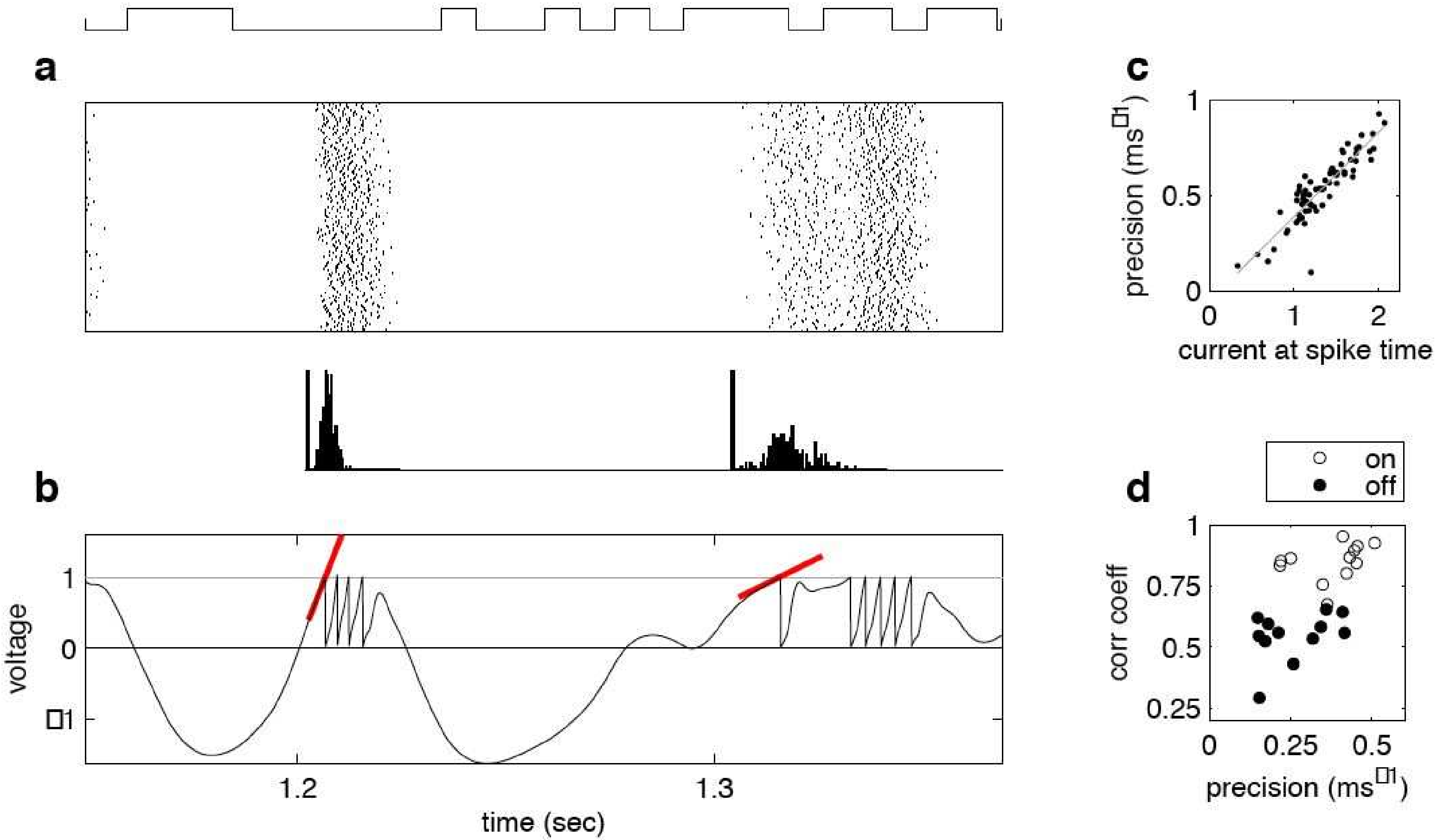
(Pillow et al., 2004)

Spike timing precision in retina



(Pillow et al., 2004)

Linking spike reliability and subthreshold noise



(Pillow et al., 2004)

"Escape-rate" models

Can we approximate likelihood in IF model to speed up computations?

Idea: firing rate in noisy IF model increases as $V(t)$ gets closer to threshold.

Keep subthreshold mean idea, but use "escape rate" approximation for firing rate.

$$\lambda(t) = f(V(t))$$

$$\dot{V}(t) = -g(V(t) - V_{rest}) + \vec{k} \cdot \vec{x} + \sum_j h(t - t_j)$$

$f(u)$ convex, $\log f(u)$ concave $\implies$
loglikelihood concave in $\{gV_{rest}, \vec{k}, h(.)\}$

Intracellular models

We've been discussing fitting models to superthreshold observations.

What if we are recording intracellularly and have subthreshold voltage as well?

Instead of $p(spike(t)|\vec{x})$, model $p(V(t)|\vec{x})$

Intracellular models

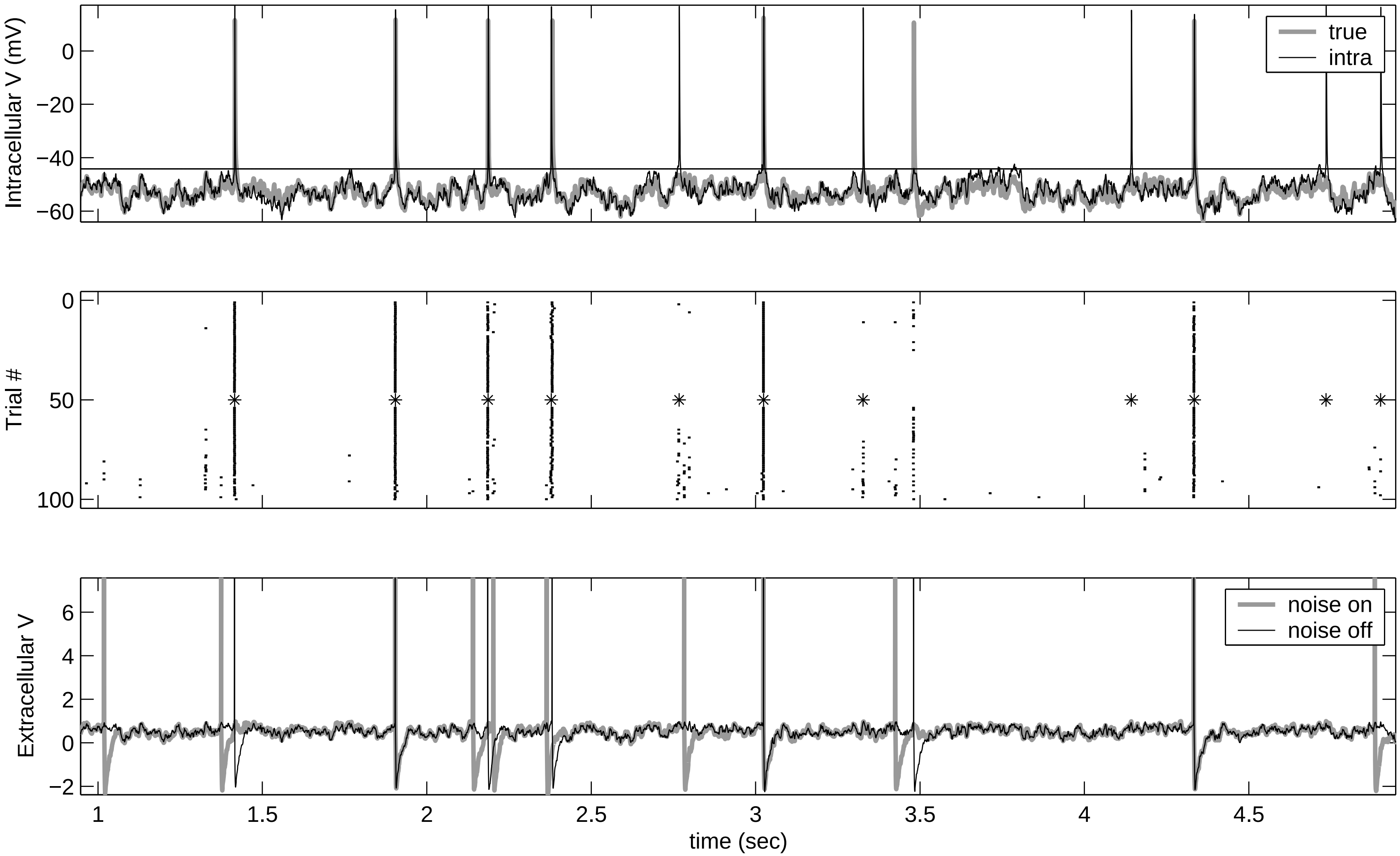
How to fit intracellular models? E.g., IF model:

$$\dot{V}(t) = -g(V(t) - V_{rest}) + \vec{k} \cdot \vec{x} + \sum_j h(t - t_j)$$
$$= -gV(t) + I_{DC} + \vec{k} \cdot \vec{x} + \sum_j h(t - t_j)$$

Notice: given data $V, \vec{x}$, and $\{t_j\}$, $\dot{V}$ is linear in parameters $\{g, I_{DC}, \vec{k}, h(.)\}$

$\Longrightarrow$ parameters may be fit by linear regression.

Fit using white noise stimuli *in vitro*



(Paninski et al., 2003; Paninski et al., 2004a)

Summary of continuous-time models

Cascade methodology gives flexible framework

Continuous time models allows accurate modeling of spike-history effects

Concavity idea leads to easily-fittable class of models

Integrate-and-fire model provides connection to biophysical parameters

Summary of encoding

Toolbox of models: additive models, cascade models, integrate-and-fire models, escape-rate models

Well-understood, straightforward estimation methods

Good grasp of bias/variance properties of estimators: when methods will work, when they will fail

Next: how to decode spike trains

References

Berry, M. and Meister, M. (1998). Refractoriness and neural precision. *Journal of Neuroscience*, 18:2200–2211.

Brown, E., Barbieri, R., Ventura, V., Kass, R., and Frank, L. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14:325–346.

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

Kass, R., Ventura, V., and Cai, C. (2003). Statistical smoothing of neuronal data. *Network: Computation in Neural Systems*, 14:5–15.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15:243–262.

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*, In press.

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

Pillow, J., Paninski, L., Uzzell, V., Simoncelli, E., and Chichilnisky, E. (2004). Accounting for timing and variability of retinal ganglion cell light responses with a stochastic integrate-and-fire model. *SFN Abstracts*.