Challenges and opportunities in statistical neuroscience

Liam Paninski

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

November 1, 2012

Support: NIH/NSF CRCNS, Sloan, NSF CAREER, DARPA, McKnight.

A golden age of statistical neuroscience

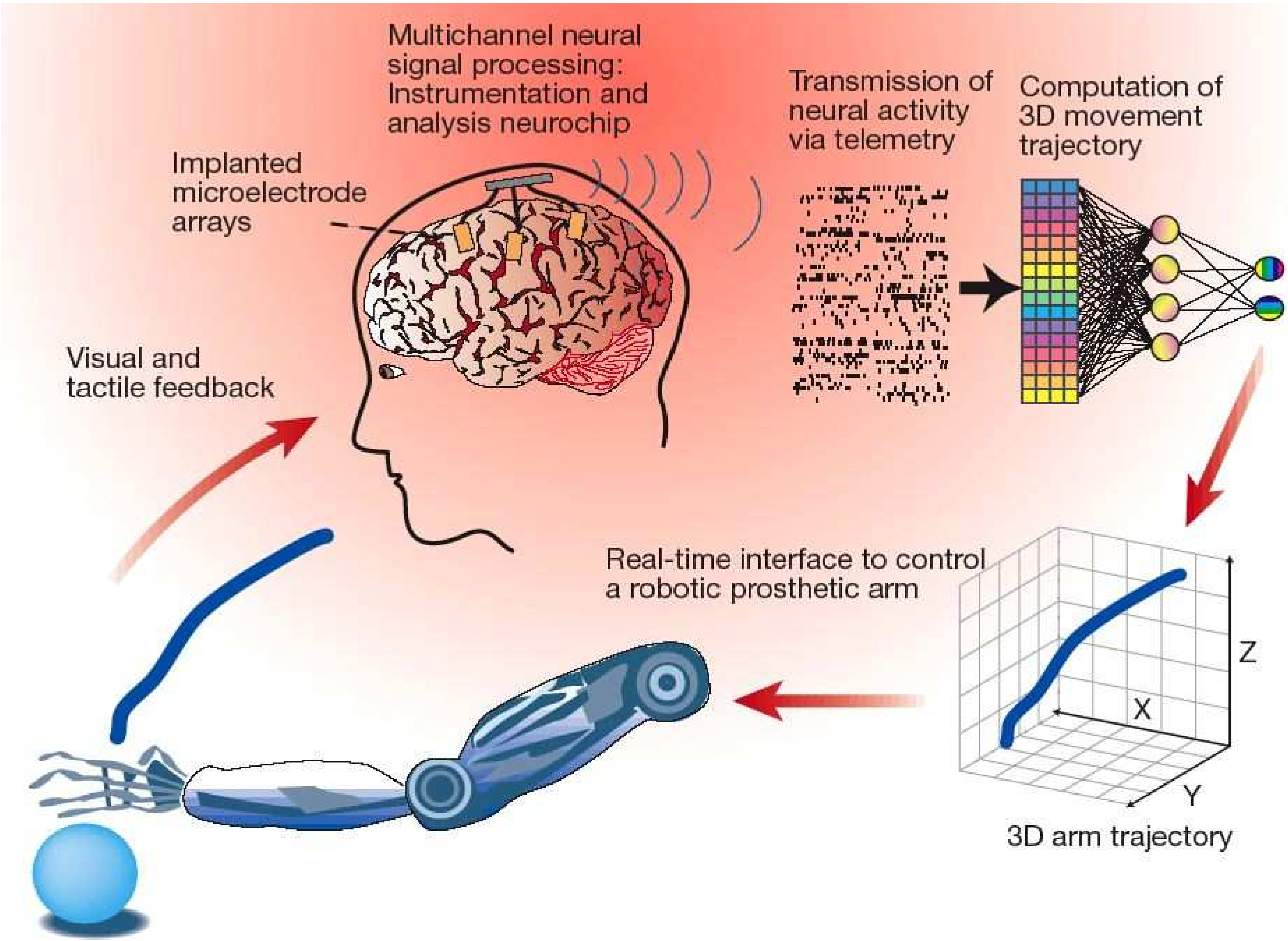
Some notable recent developments:

- machine learning / statistics / optimization methods for extracting information from high-dimensional data in a computationally-tractable, systematic fashion
- computing (Moore's law, massive parallel computing)
- optical and optogenetic methods for recording from and perturbing neuronal populations, at multiple scales
- large-scale, high-density multielectrode recordings

A few grand challenges

- Optimal decoding and dimensionality reduction of large-scale multineuronal spike train data
- Circuit inference from multineuronal spike train data
- Optimal control of spike timing in large neuronal populations
- Hierarchical nonlinear models for encoding information in neuronal populations
- Robust, expressive neural prosthetic design
- Understanding dendritic computation and location-dependent synaptic plasticity via optical imaging (statistical spatiotemporal signal processing on trees)

Example: neural prosthetics



Example: neural prosthetics

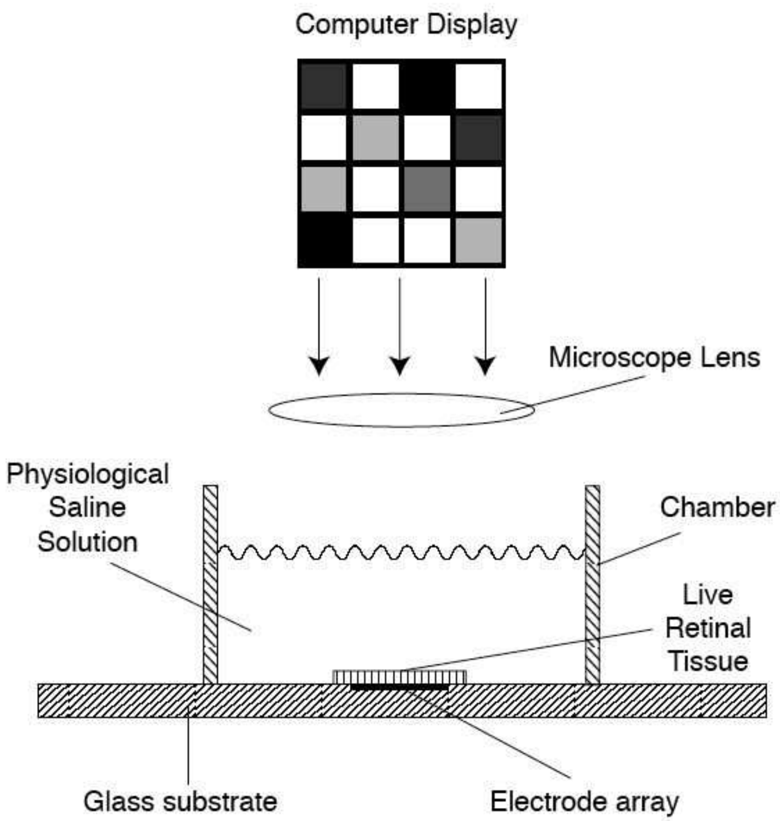
(Loading monkey-zombies.mp4)

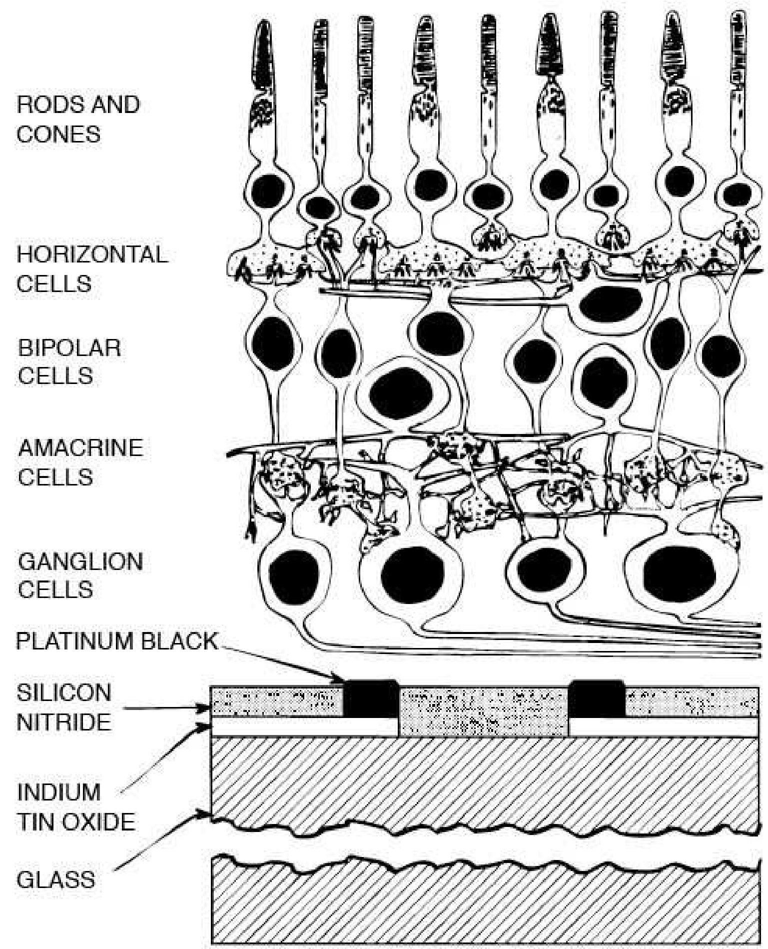
w/ B. Pesaran (NYU), D. Pfau, J. Merel

Example: modeling the output of the retina

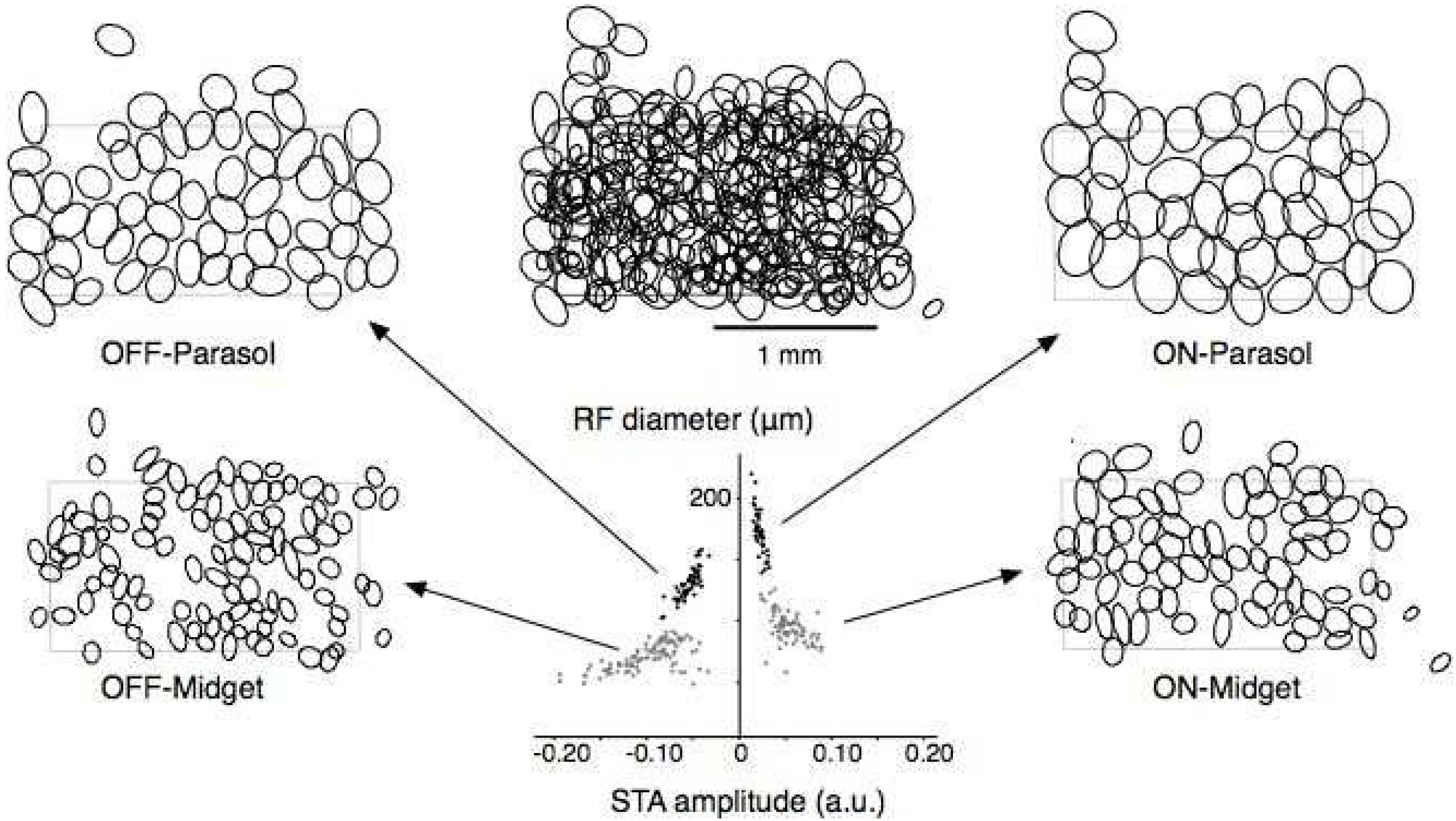
Preparation: dissociated macaque retina (Chichilnisky lab, Salk)

— extracellularly-recorded responses of populations of retinal ganglion neurons



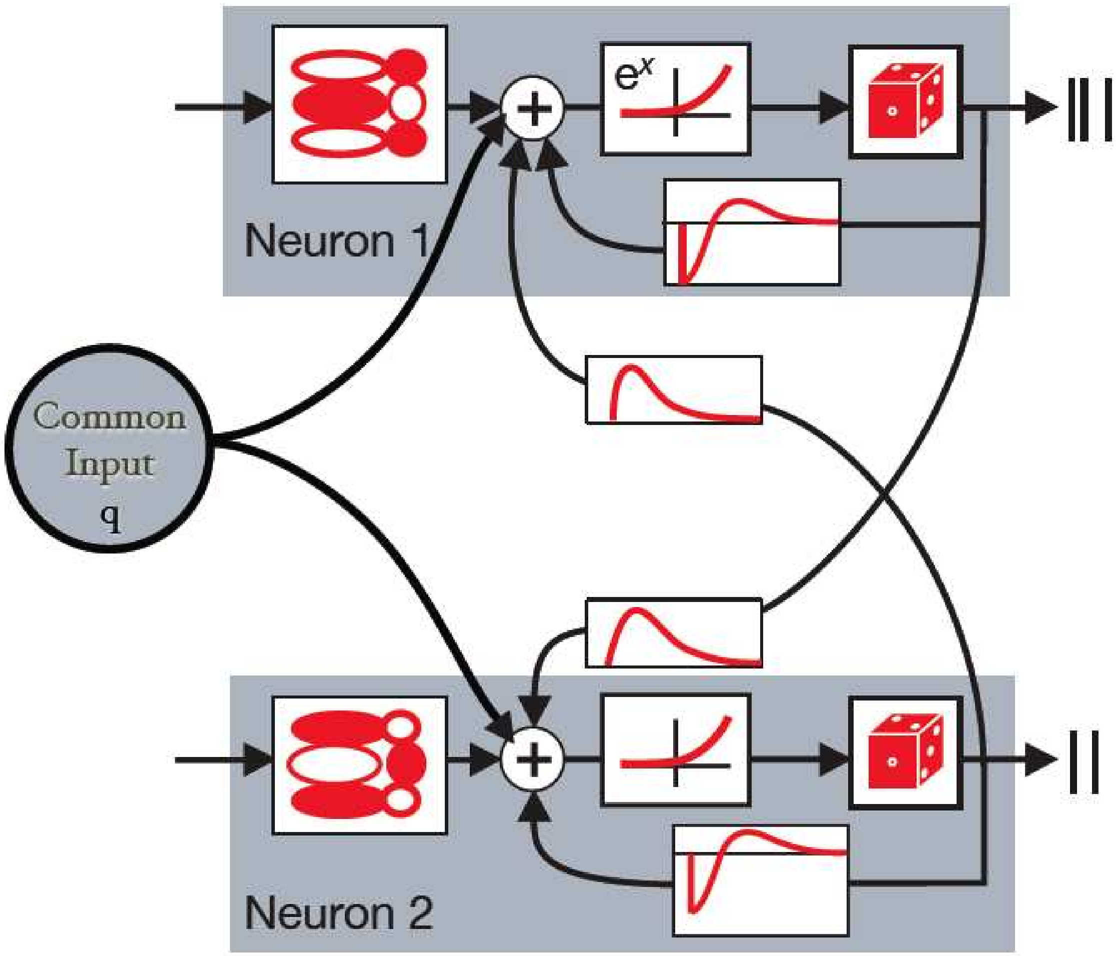


Sampling the complete receptive field mosaic



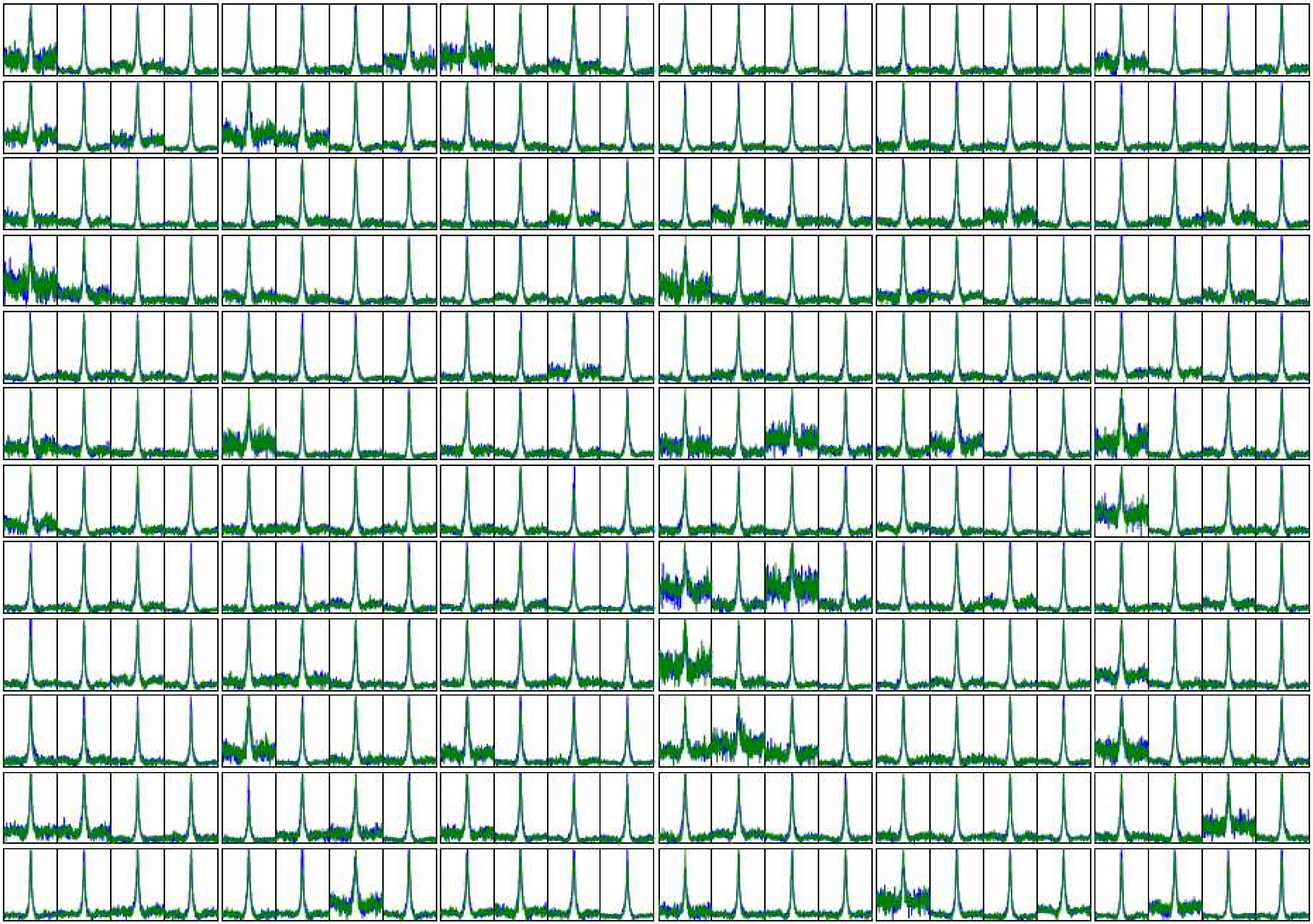
Multineuronal point-process model

$$\lambda_i(t) = \exp\left(k_i \cdot x(t) + h_i \cdot y_i(t) + \sum_{i \neq j} l_{i,j} \cdot y_j(t) + Lq(t)\right)$$



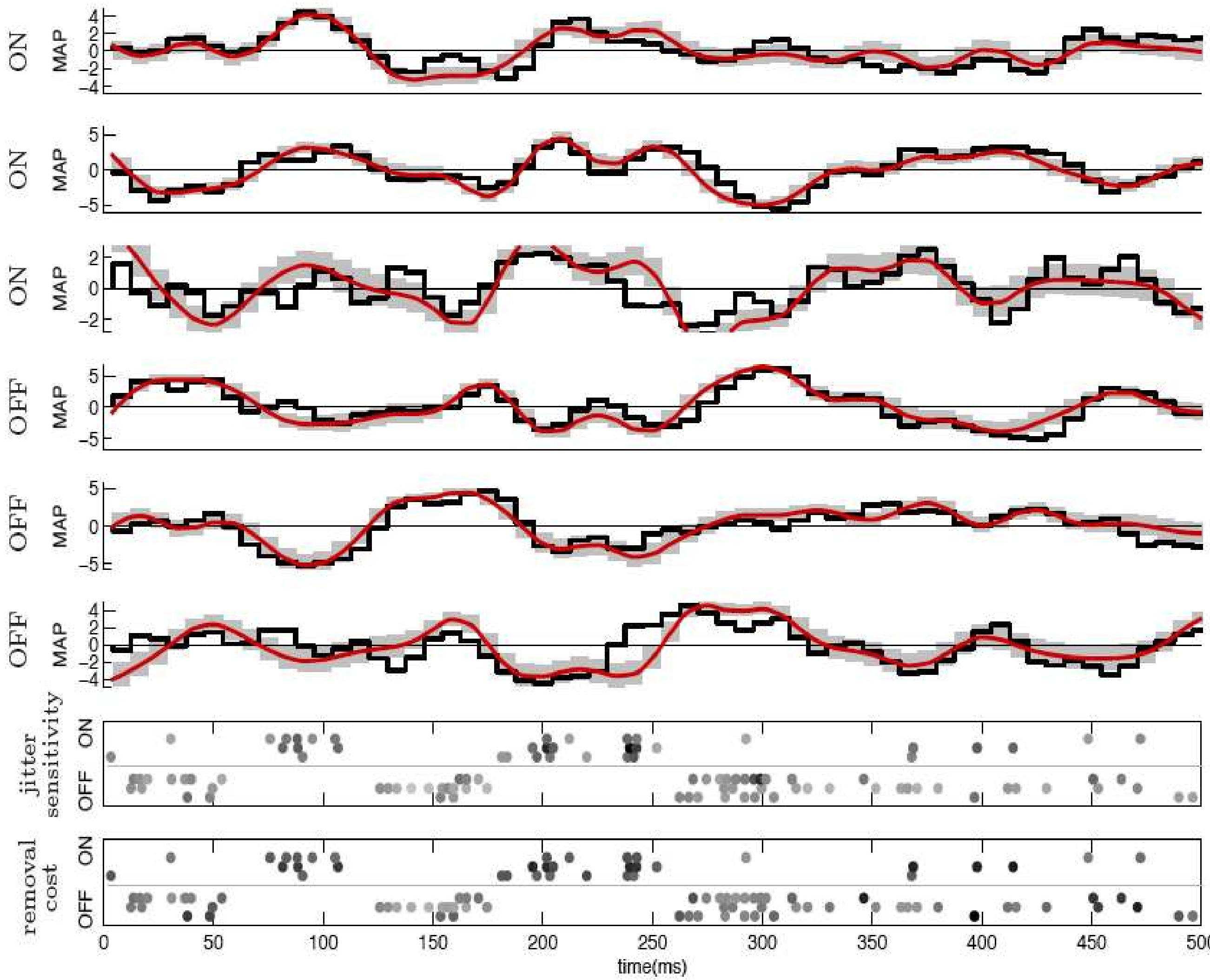
— likelihood is tractable to compute and to maximize (concave optimization)
(Paninski, 2004; Paninski et al., 2007; Pillow et al., 2008; Paninski et al., 2010)

Network model predicts correlations correctly



— single and triple-cell activities captured as well (Vidne et al., 2009)

Optimal Bayesian decoding

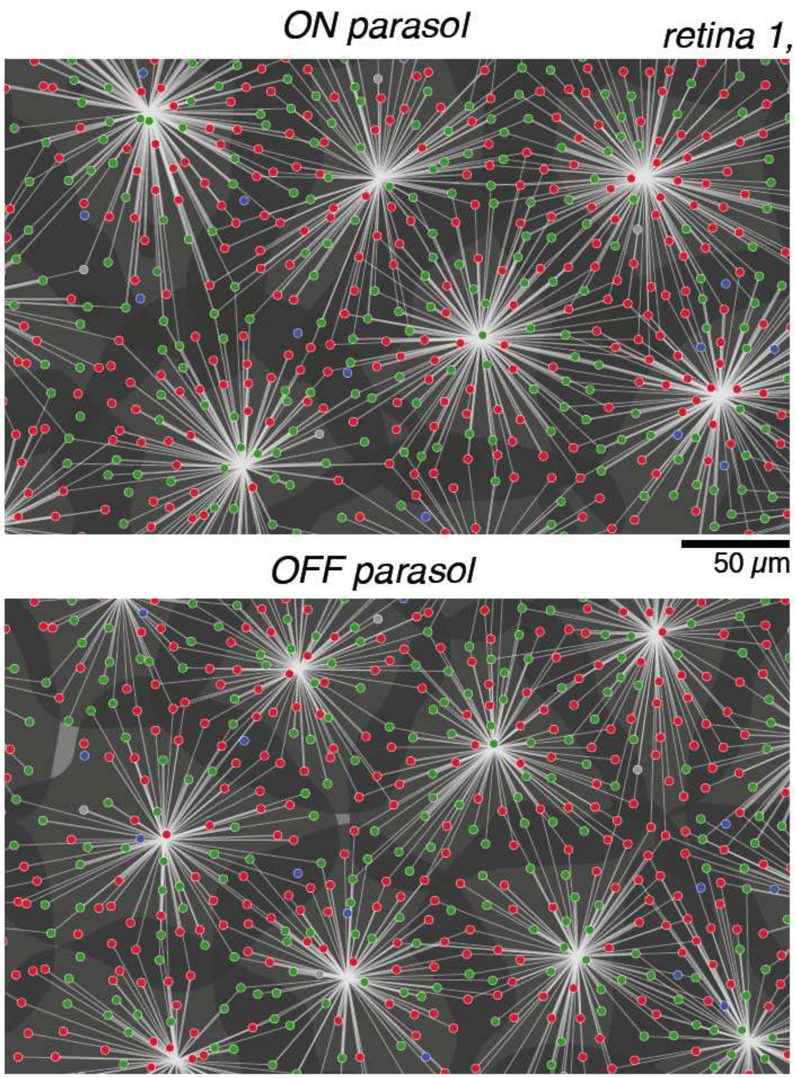


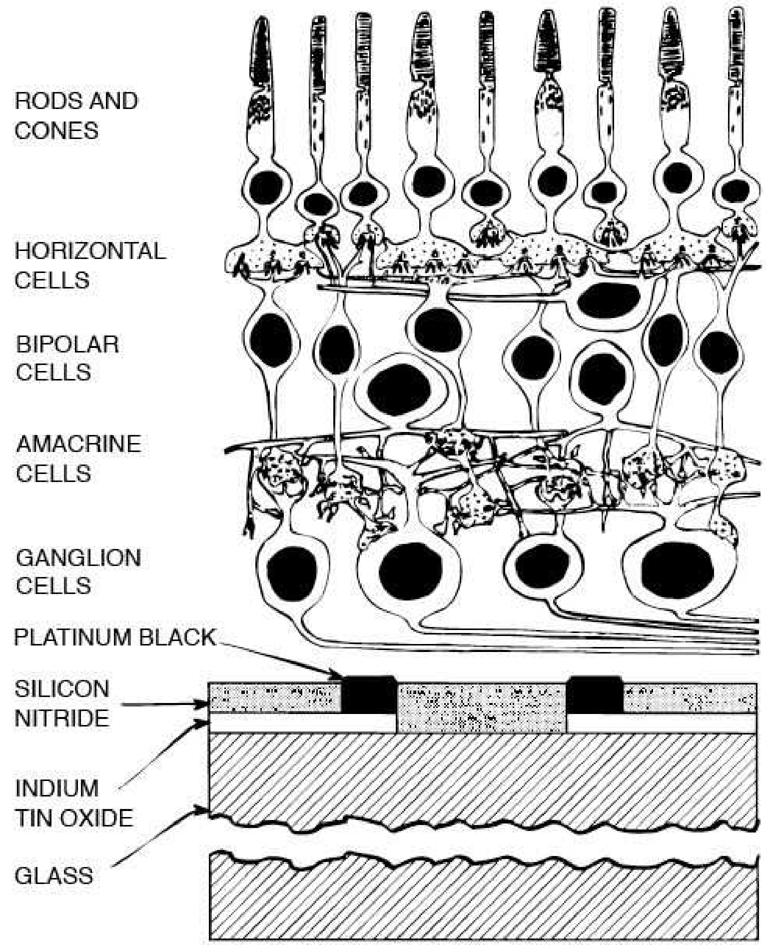
— properly modeling correlations improves decoding accuracy (Pillow et al., 2008).

— further applications: decoding velocity signals (Lalor et al., 2009); tracking images perturbed by eye jitter (Pfau et al., 2009); retinal prosthetics (Ahmadian et al., 2011)

— convex optimization approach requires just $O(T)$ time. Open challenge: real-time decoding / optimal control of large populations

Inferring cone maps

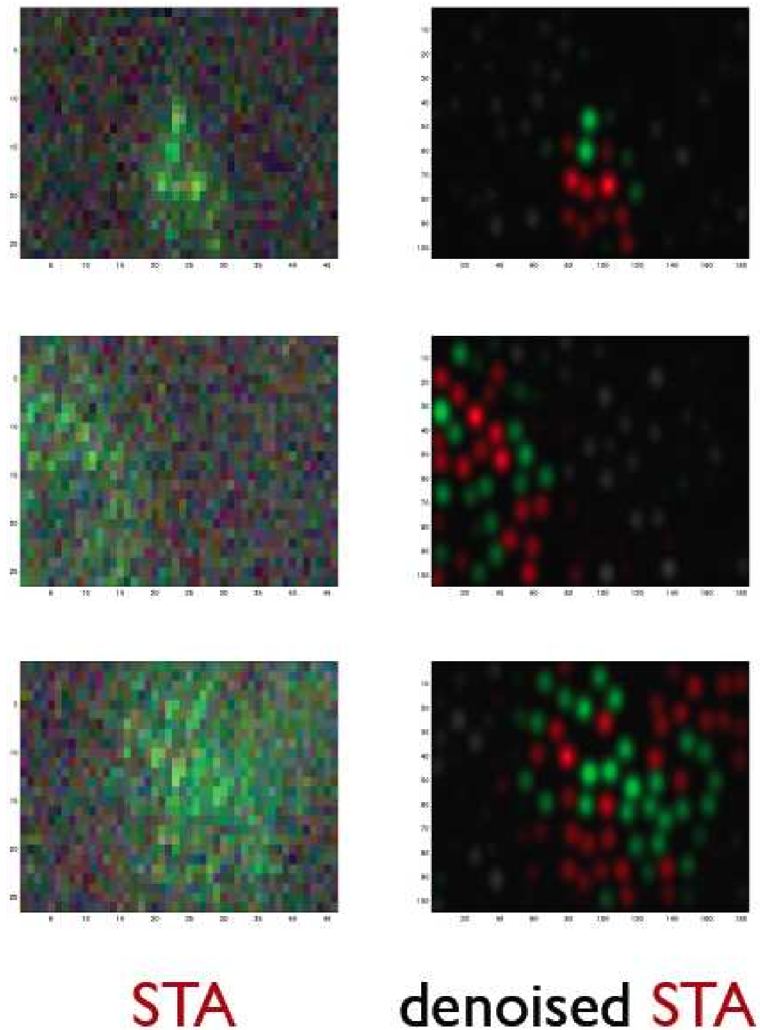




— cone locations and color identity inferred accurately with high-resolution stimuli; Bayesian hierarchical approach integrates information over multiple simultaneously recorded neurons (Field et al., 2010).

Opportunity: hierarchical models

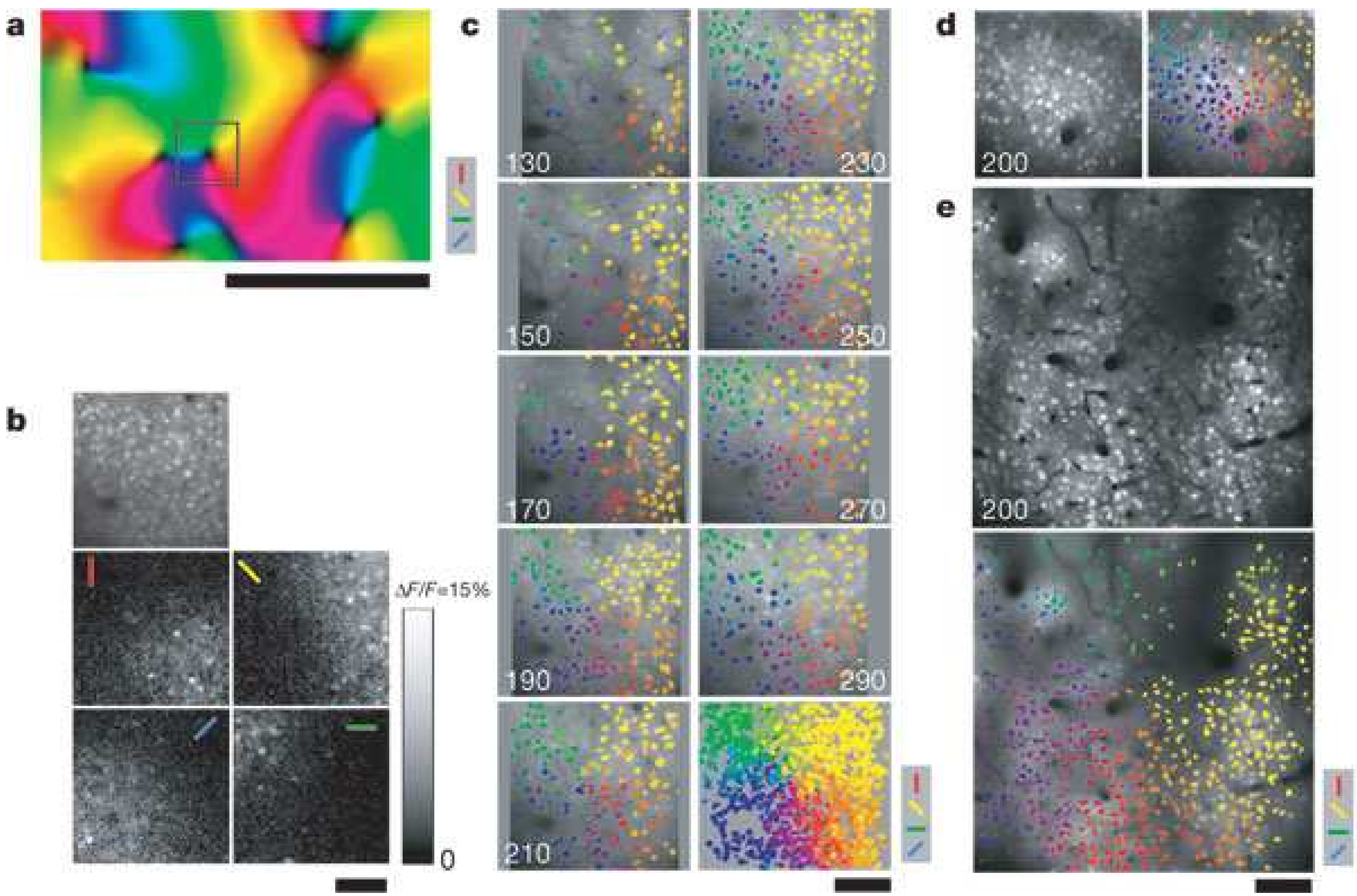
More general idea: sharing information across multiple simultaneously-recorded cells can be very useful (Sadeghi et al, 2012).



Open challenge: extension to richer nonlinear models (J. Merel, E. Pnevmatikakis, J. Freeman, E. Simoncelli, A. Ramirez, ongoing)

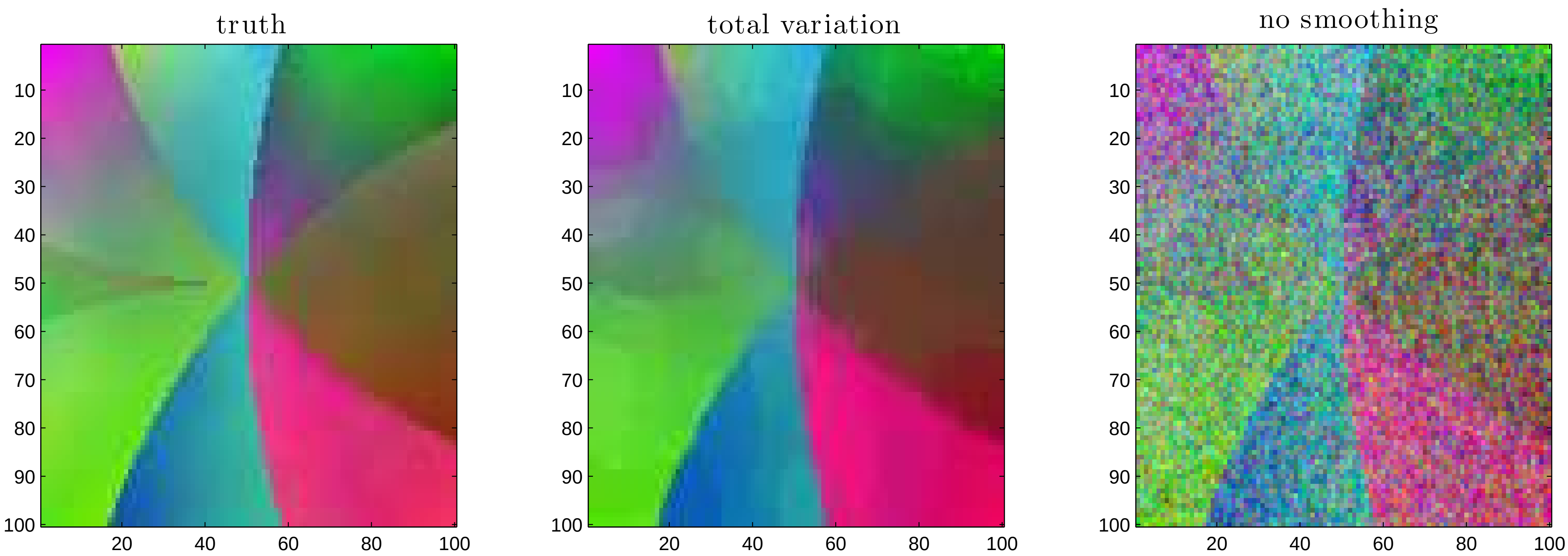
Opportunity: hierarchical models

More general idea: sharing information across multiple simultaneously-recorded cells can be very useful. Exploit location, genetic markers, other information to extract more information from noisy data.



Ohki '06

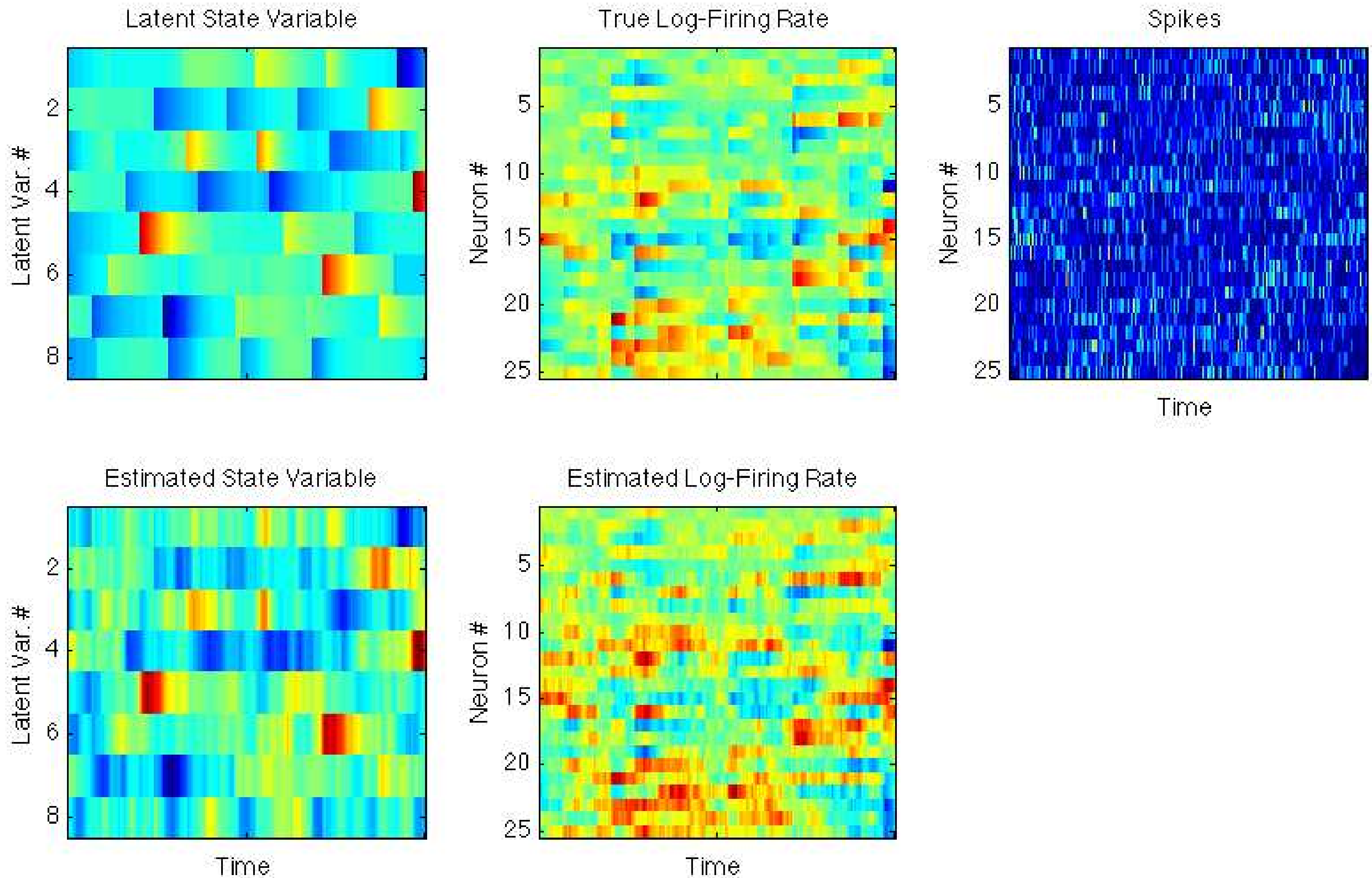
Opportunity: hierarchical models



Scalable convex edge-preserving neighbor-penalized likelihood methods; K. Rahnama Rad, C. Smith, G. Lacerda, ongoing

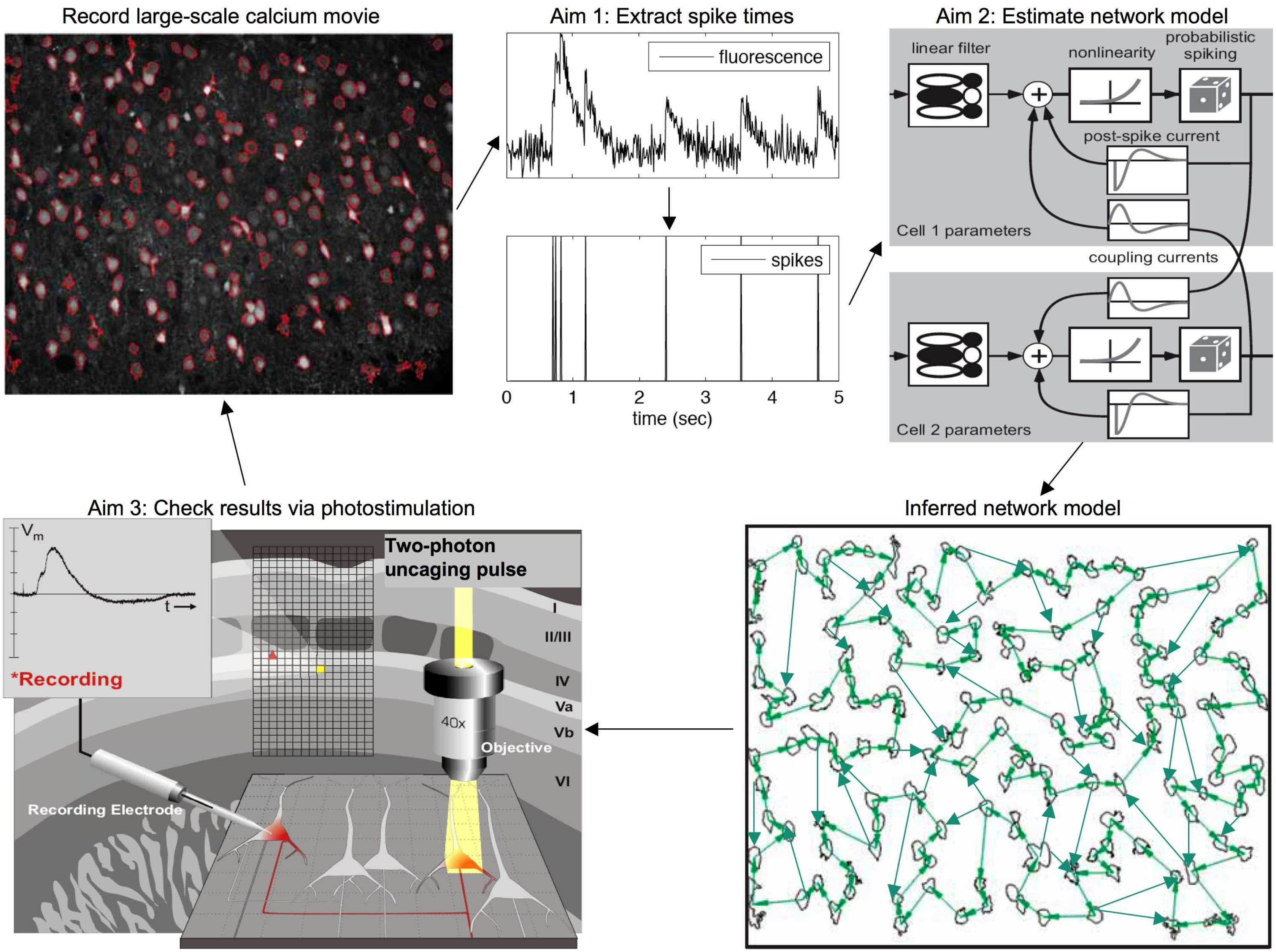
Dimensionality reduction; inferring hidden dynamics

Dynamic generalized factor analysis model: q_t evolves according to a simple linear dynamical system, with "kicks." Log-firing rates modeled as linear functions of q_t. Convex rank-penalized optimization methods to infer q_t given spike train.



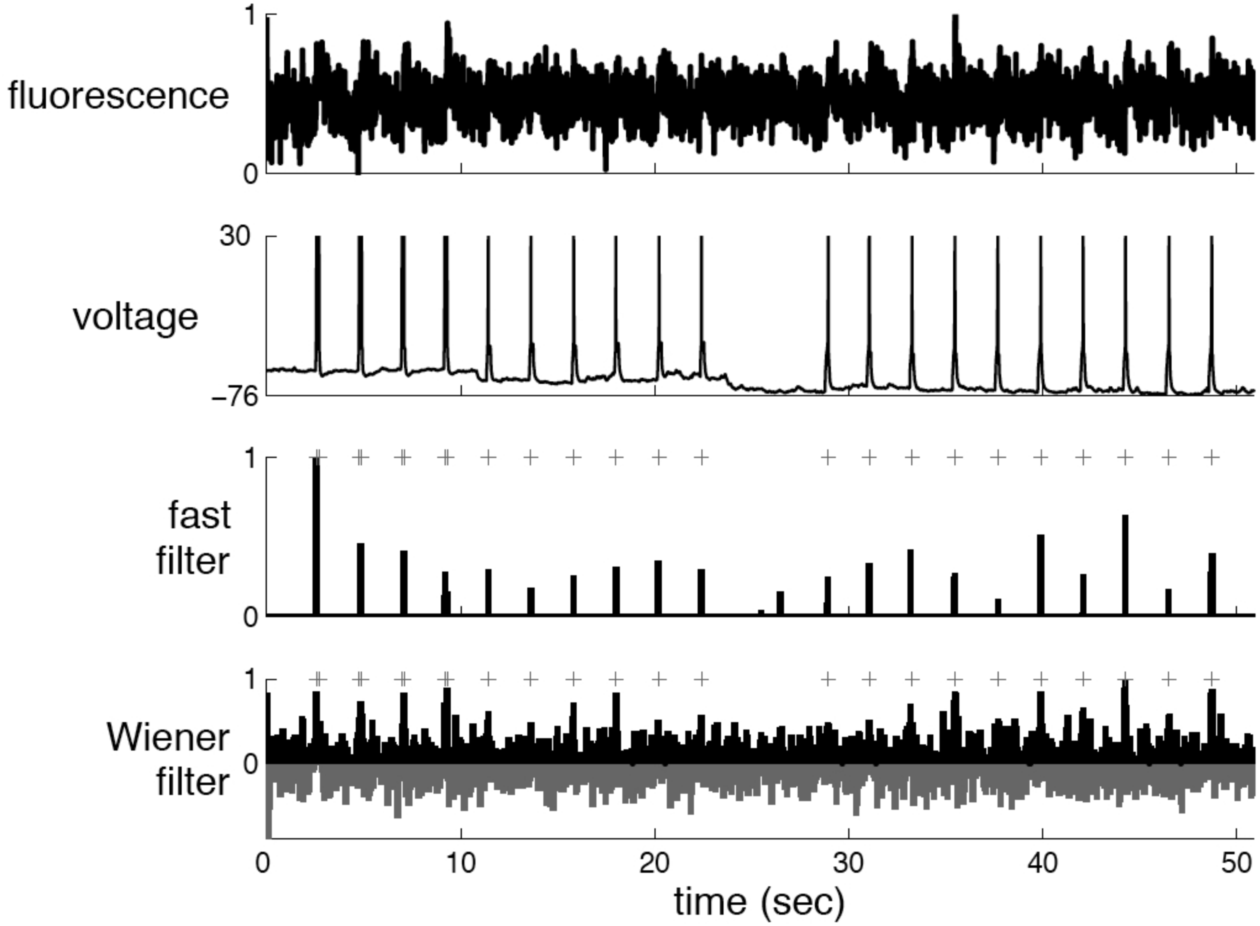
Open challenge: richer nonlinear models. E. Pnevmatikakis and D. Pfau, ongoing

Circuit inference from large-scale Ca^{2+} imaging



w/ R. Yuste, K. Shepard, Y. Ahmadian, J. Vogelstein, Y. Mishchenko, B. Watson, A. Murphy

Challenge: slow, noisy calcium data



First-order model:

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

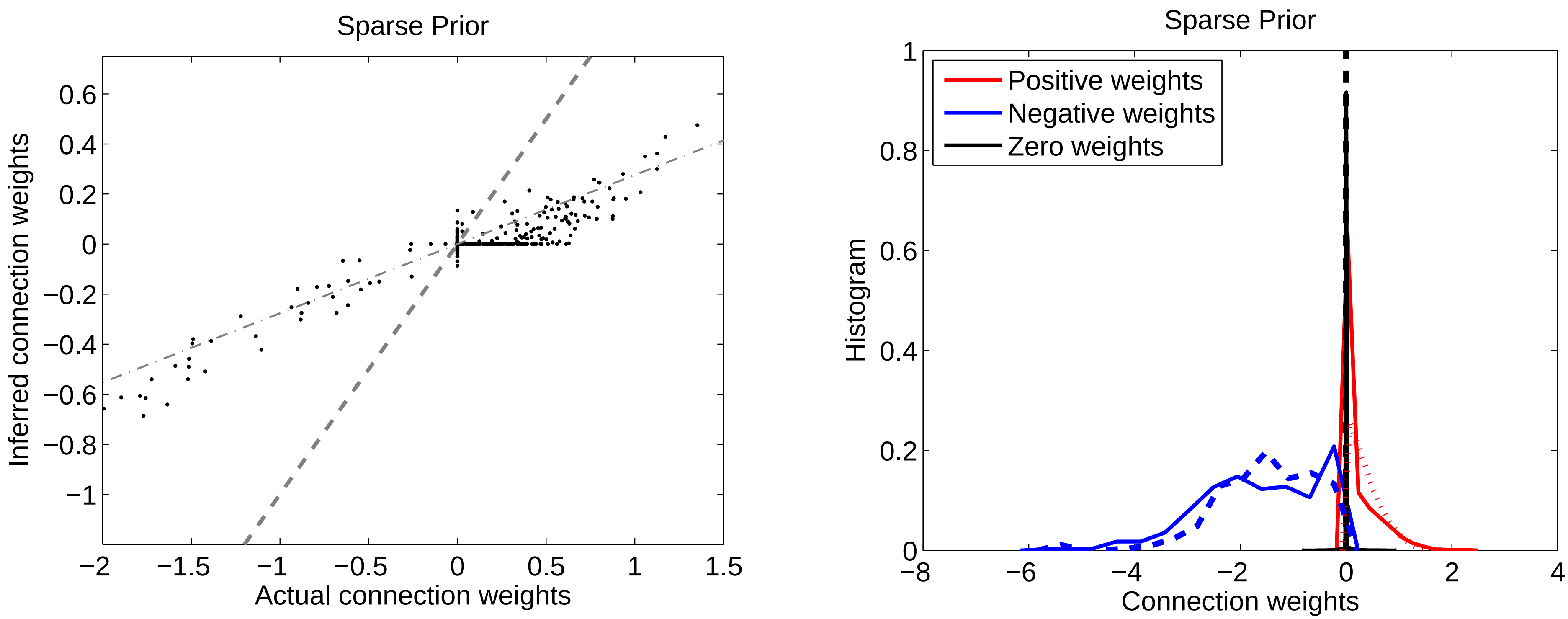
— $\tau \approx 100$ ms; nonnegative deconvolution problem. Interior-point approach leads to $O(T)$ solution (Vogelstein et al., 2009; Vogelstein et al., 2010; Mishchenko et al., 2010).

Spatiotemporal Bayesian spike estimation

(Loading Tim-data0b2.mp4)

Rank-penalized convex optimization with nonnegativity constraints. E. Pnevmatikakis and T. Machado, ongoing

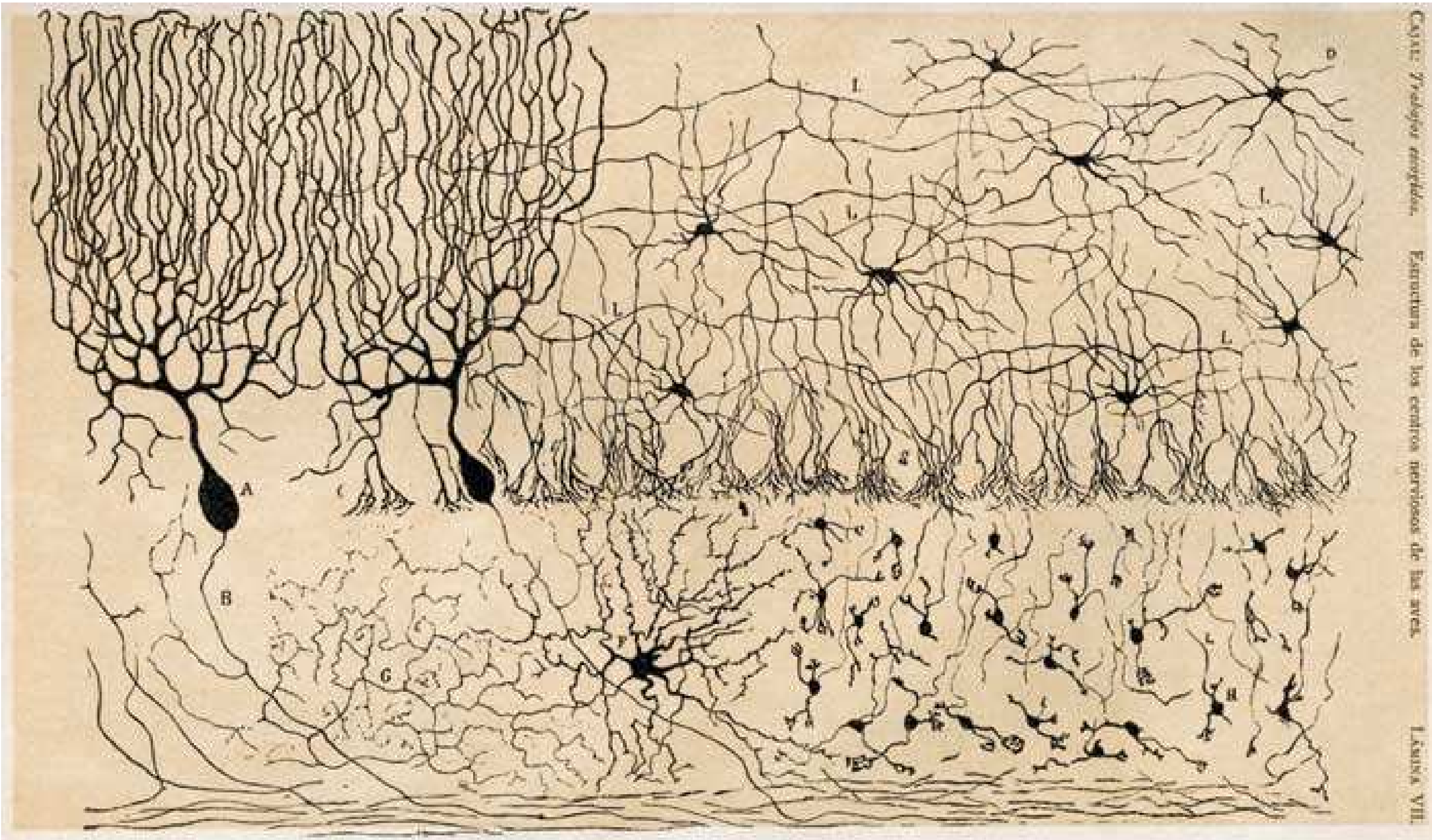
Simulated circuit inference



Good news: connections are inferred well in biologically-plausible simulations (Mishchencko et al., 2009), if most neurons in circuit are observable. Fast enough to estimate connectivity in real time (T. Machado). Preliminary experimental results are encouraging (correct identification checked w/ intracellular recordings).

Open challenge: method is non-robust when smaller fractions of the network are observable. Massive hidden data problem. Some progress in (Vidne et al., 2009), but remains open for new ideas.

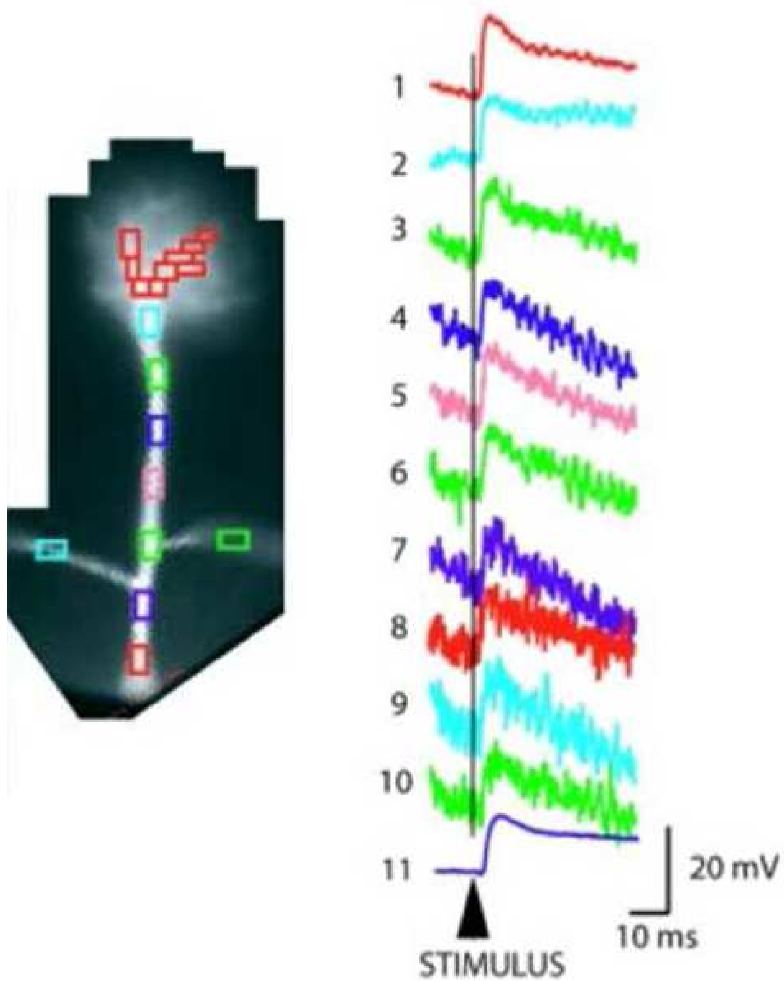
A final challenge: understanding dendrites



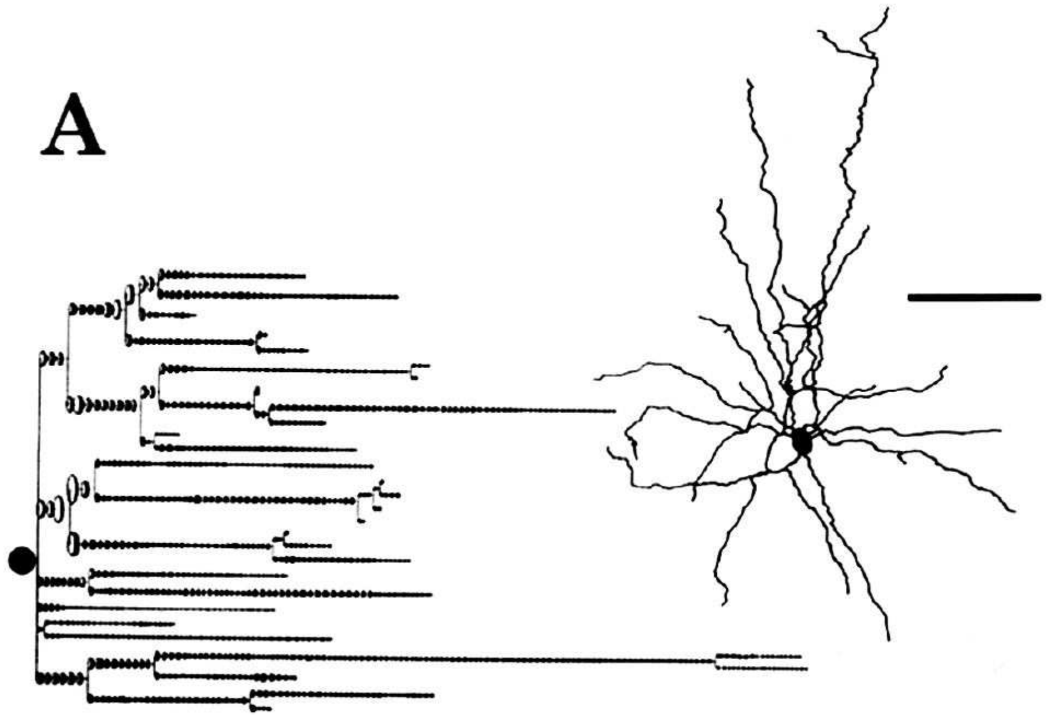
Ramon y Cajal, 1888.

A spatiotemporal 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.



Basic paradigm: compartmental models



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

Inference of spatiotemporal neuronal state given noisy observations

Variable of interest, V_t, evolves according to a noisy differential equation (e.g., cable equation):

$$dV/dt = f(V) + \epsilon_t.$$

Make noisy observations:

$$y(t) = g(V_t) + \eta_t.$$

We want to infer $E(V_t|Y)$: optimal estimate given observations. We also want errorbars: quantify how much we actually know about V_t.

If $f(.)$ and $g(.)$ are linear, and ϵ_t and η_t are Gaussian, then solution is classical: Kalman filter. (Many generalizations available; e.g., (Huys and Paninski, 2009).)

Even Kalman case is challenging, since $d = \dim(\vec{V})$ is very large: computation of Kalman filter requires $O(d^3)$ computation per timestep

(Paninski, 2010): methods for Kalman filtering in just $O(d)$ time: take advantage of sparse tree structure.

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 $C_t = cov(q_t|Y_{1:t})$ as a perturbation of the marginal covariance $C_0 + U_t D_t U_t^T$, with $C_0 = \lim_{t\to\infty} cov(q_t)$.

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

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

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

and can be done in $O(\dim(q))$ time (Paninski, 2010). Generalizable to many other state-space models (Pnevmatikakis and Paninski, 2011).

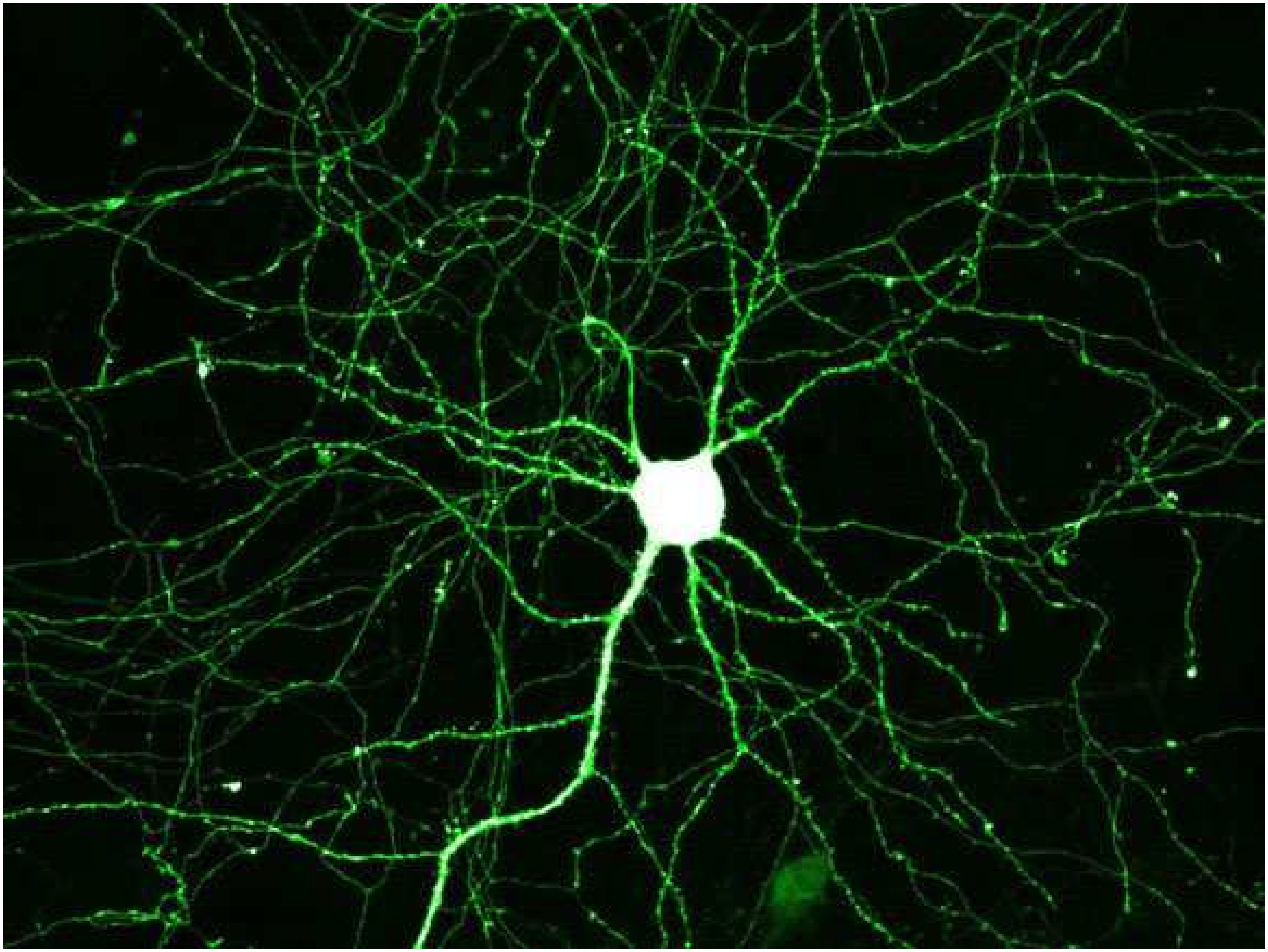
Example: inferring voltage from subsampled observations

(Loading low-rank-speckle.mp4)

Applications

- Optimal experimental design: which parts of the neuron should we image? Submodular optimization (Huggins and Paninski, 2011)
- Estimation of biophysical parameters (e.g., membrane channel densities, axial resistance, etc.): reduces to a simple nonnegative regression problem once $V(x,t)$ is known (Huys et al., 2006)
- Detecting location and weights of synaptic input

Application: synaptic locations/weights



Application: synaptic locations/weights

Including known terms:

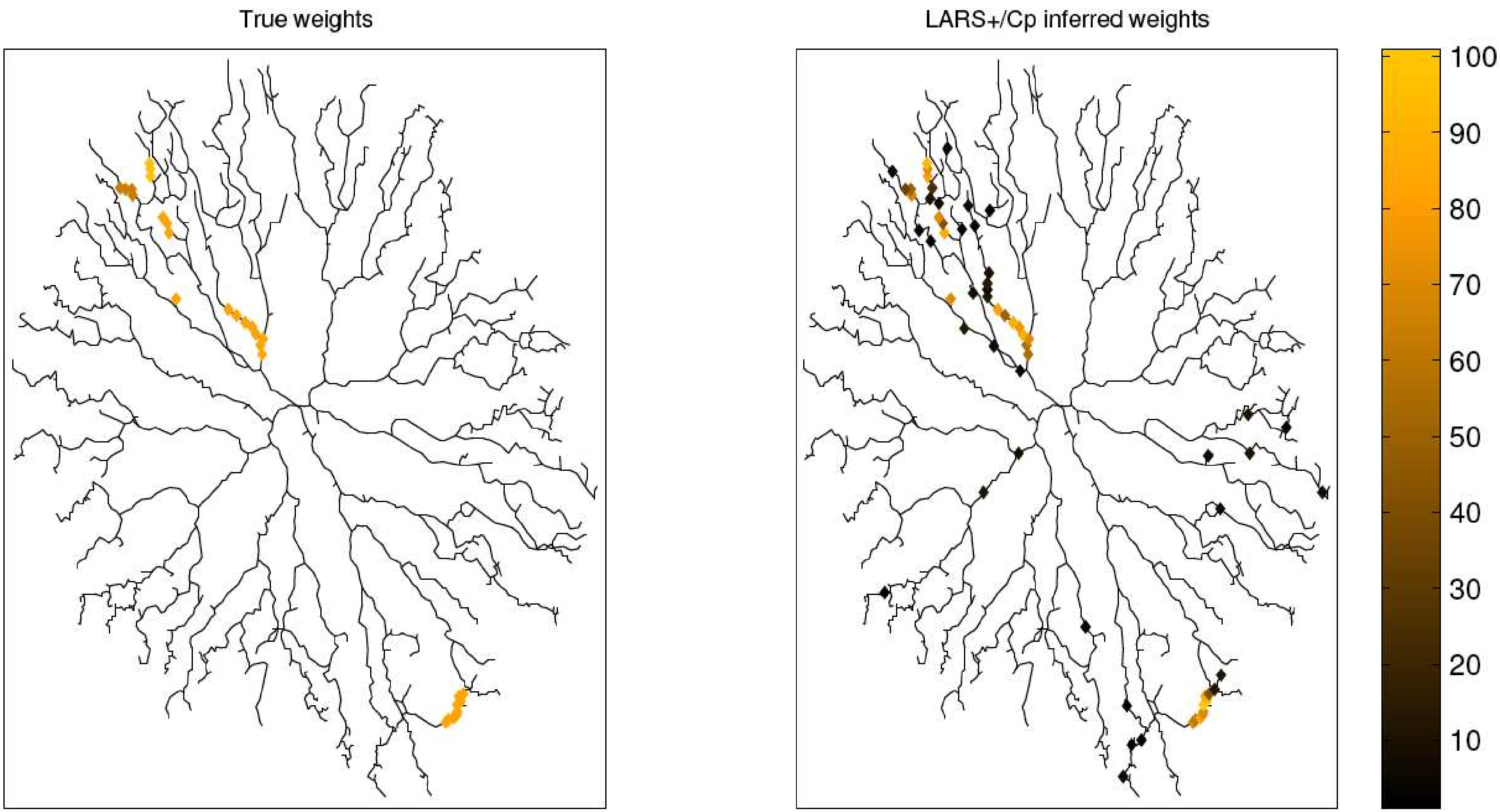
$$d\vec{V}/dt = A\vec{V}(t) + W\vec{U}(t) + \vec{\epsilon}(t);$$

$U(t)$ are known presynaptic spike times, and we want to detect which compartments are connected (i.e., infer the weight matrix W).

Loglikelihood is quadratic; W is a sparse vector. Adapt standard LARS-like (homotopy) approach (Pakman et al., 2012).

Total computation time: $O(dTk)$; $d = \#$ compartments, $T = \#$ timesteps, $k = \#$ nonzero weights.

Example: inferring dendritic synaptic maps



700 timesteps observed; 40 compartments (of $>$ 2000) observed per timestep

Note: random access scanning essential here: results are poor if we observe the same compartments at each timestep. "Compressed sensing" observations improve results further.

Conclusions

- Modern statistical approaches provide flexible, powerful methods for answering key questions in neuroscience. Many neuroscience problems are actually statistics problems, thinly disguised.
- Close relationships between biophysics and statistical modeling
- Modern optimization methods make computations very tractable; suitable for closed-loop experiments
- Experimental methods progressing rapidly; many new challenges and opportunities for breakthroughs based on statistical ideas. Rich open ground for collaboration between neuroscience, statistics, CS, optimization theory, ...

References

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.

Field et al. (2010). Mapping a neural circuit: A complete input-output diagram in the primate retina. *Under review.*

Huggins, J. and Paninski, L. (2011). Optimal experimental design for sampling voltage on dendritic trees. *J. Comput. Neuro.*, In press.

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.

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.

Lalor, E., Ahmadian, Y., and Paninski, L. (2009). The relationship between optimal and biologically plausible decoding of stimulus velocity in the retina. *Journal of the Optical Society of America A*, 26:25–42.

Mishchenko, Y., Vogelstein, J., and Paninski, L. (2010). A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. *Annals of Applied Statistics*, In press.

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

Paninski, L. (2010). Fast Kalman filtering on quasilinear dendritic trees. *Journal of Computational Neuroscience*, 28:211–28.

Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29:107–126.

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.

Pfau, D., Pitkow, X., and Paninski, L. (2009). A Bayesian method to predict the optimal diffusion coefficient in random fixational eye movements. *Conference abstract: Computational and systems neuroscience.*

Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A., Chichilnisky, E., and Simoncelli, E. (2008). Spatiotemporal correlations and visual signaling in a complete neuronal population. *Nature*, 454:995–999.