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

June 1, 2014

— with Y. Ahmadian, L. Buesing, Y. Gao, J. Huggins, S. Keshri, T. Machado, G. Mena, T. Paige, A. Pakman, D. Pfau, E. Pnevmatikakis, B. Shababo, C. Smith, D. Soudry, J. Vogelstein

Support: NIH/NSF CRCNS, NSF CAREER, DARPA, ARO, ONR, 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
- growing acceptance that many fundamental neuroscience questions are in fact statistics questions in disguise

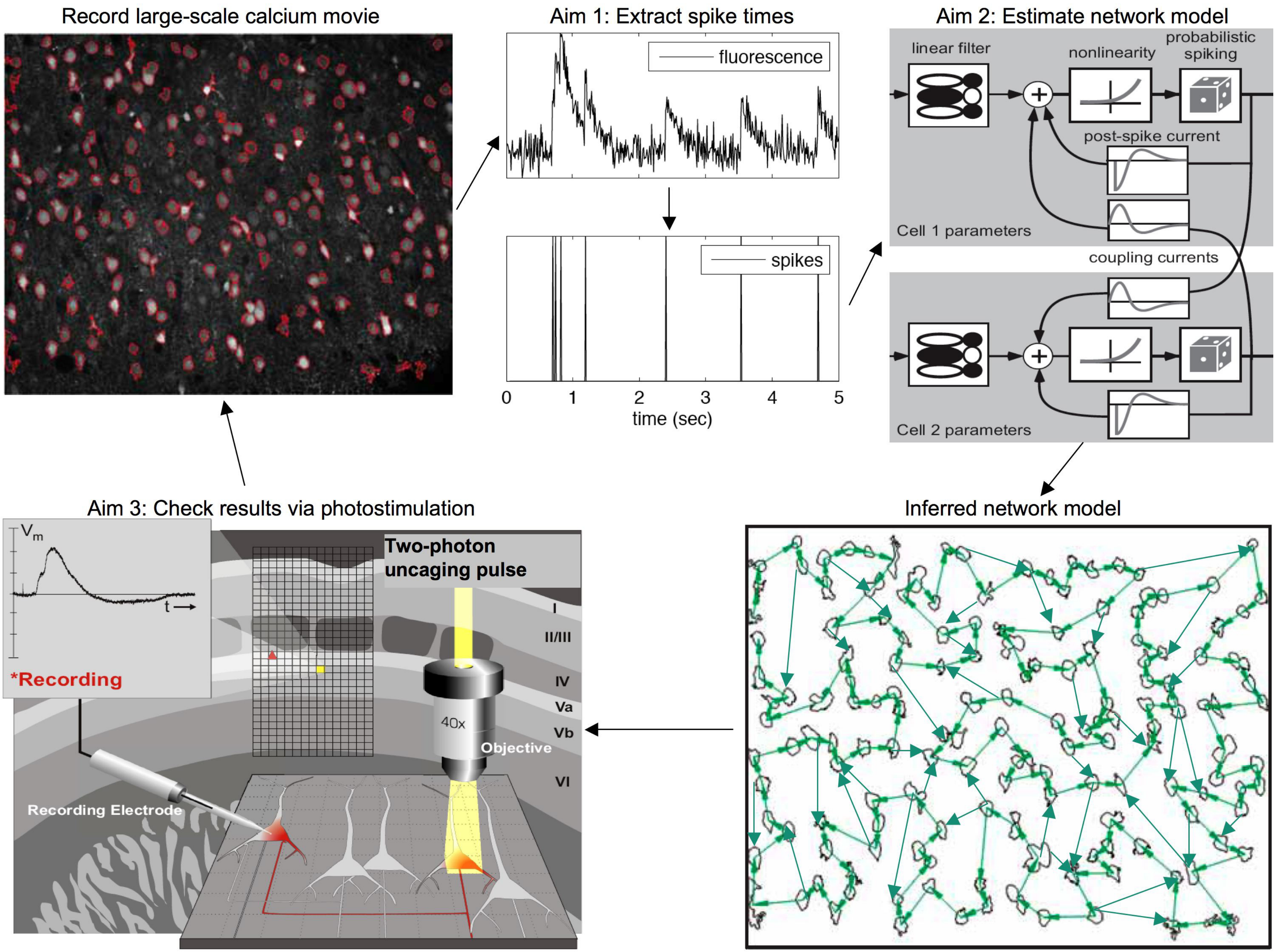
A few grand challenges

- Optimal decoding and dimensionality reduction of large-scale multineuronal point process / count data
- Network inference from multineuronal spike train data
- Optimal control of large networks
- Hierarchical nonlinear models for input-output relationships in neuronal networks
- Robust, expressive brain-machine interfaces; brain reading and writing
- Understanding dendritic computation and location-dependent synaptic plasticity via optical imaging (statistical spatiotemporal signal processing on trees)

A few grand challenges

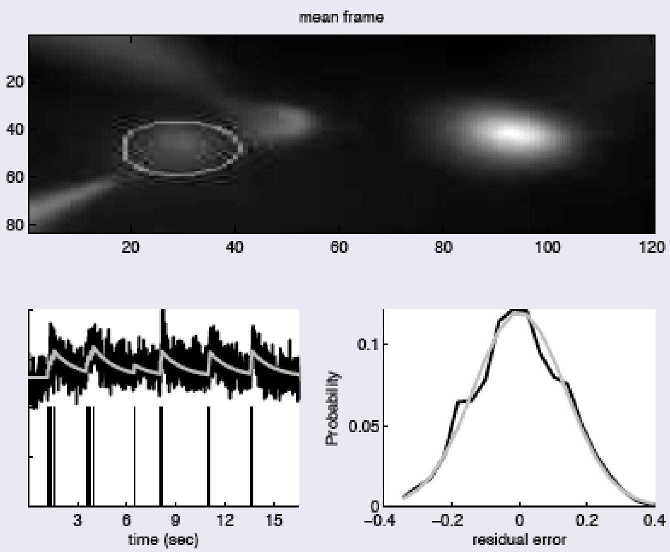
- Optimal decoding and dimensionality reduction of large-scale multineuronal point process / count data
- Network inference from multineuronal spike train data
- Optimal control of large networks
- Hierarchical nonlinear models for input-output relationships in neuronal networks
- Robust, expressive brain-machine interfaces; brain reading and writing
- Understanding dendritic computation and location-dependent synaptic plasticity via optical imaging (statistical spatiotemporal signal processing on trees)

Circuit inference via optical methods

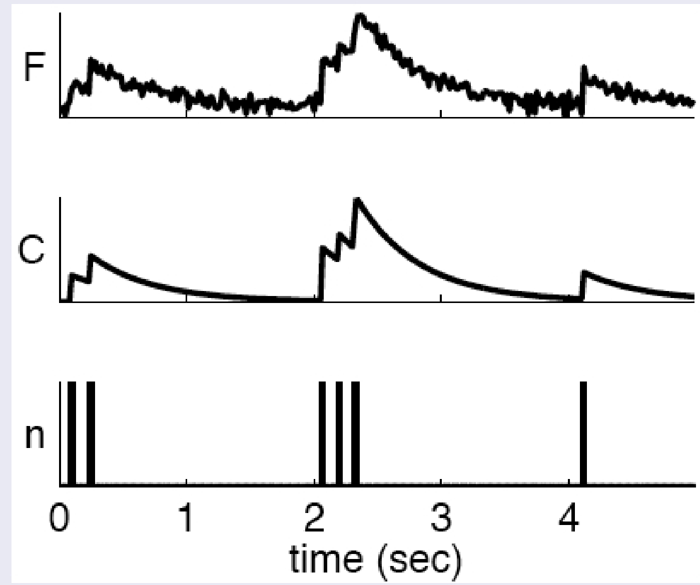


Aim 1: Model-based estimation of spike rates

data



schematic



equations

$$F_t = \alpha C_t + \beta + \sigma \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$$

$$C_t = -(1 - \Delta/\tau) C_{t-1} + n_t$$

$$n_t \sim \text{poisson}(\lambda \Delta)$$

Note: each component here can be generalized easily.

Fast maximum a posteriori (MAP) estimation

Recipe: biophysical model, then likelihood, then computation.

Start by writing out the posterior:

$$\begin{aligned}\log p(C|F) &= \log p(C) + \log p(F|C) + const. \\ &= \sum_t \log p(C_{t+1}|C_t) + \sum_t \log p(F_t|C_t) + const.\end{aligned}$$

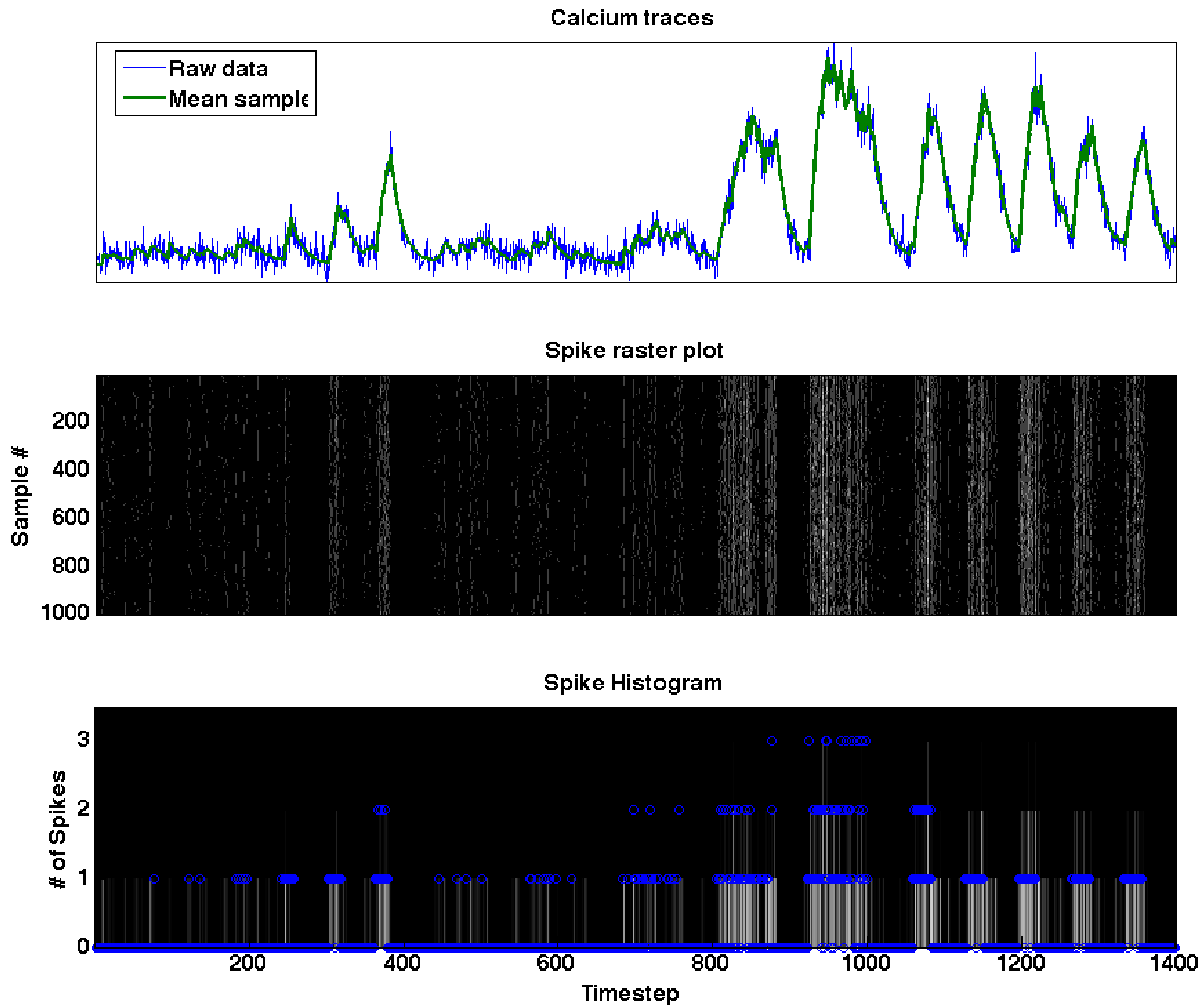
Three basic observations:

- If $\log p(C_{t+1}|C_t)$ and $\log p(F_t|C_t)$ are concave, then so is $\log p(C|F)$.
- Hessian H of $\log p(C|F)$ is tridiagonal: $\log p(F_t|C_t)$ contributes a diag term, and $\log p(C_{t+1}|C_t)$ contributes a tridiag term (Paninski et al., 2010).
- C is a linear function of n.

Newton's method: iteratively solve $HC_{dir} = \nabla$. Tridiagonal solver requires $O(T)$ time. Can include nonneg constraint $n_t \geq 0$ via log-barrier (Koyama and Paninski, 2010) — real-time deconvolution (Vogelstein et al., 2010).

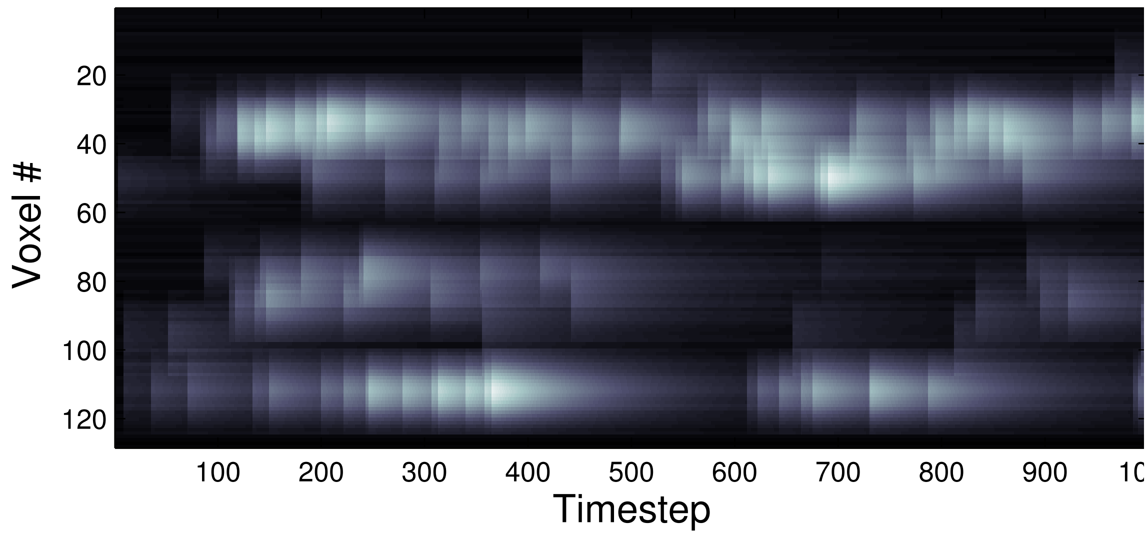
More recently: constrained formulation that eliminates the need to estimate the firing rate hyperparameter (Pnevmatikakis et al 2013).

Markov chain Monte Carlo sampling



— Pnevmatikakis et al (2014)

Multineuronal case: spatiotemporal demixing



Model:

$$
\begin{aligned}
Y &= C + \epsilon \\
C(x,t) &= \sum_{i=1}^{r} s_i(x) f_i(t) \\
f_i(t+dt) &= \left(1 - \frac{dt}{\tau_i}\right) f_i(t) + n_i(t)
\end{aligned}
$$

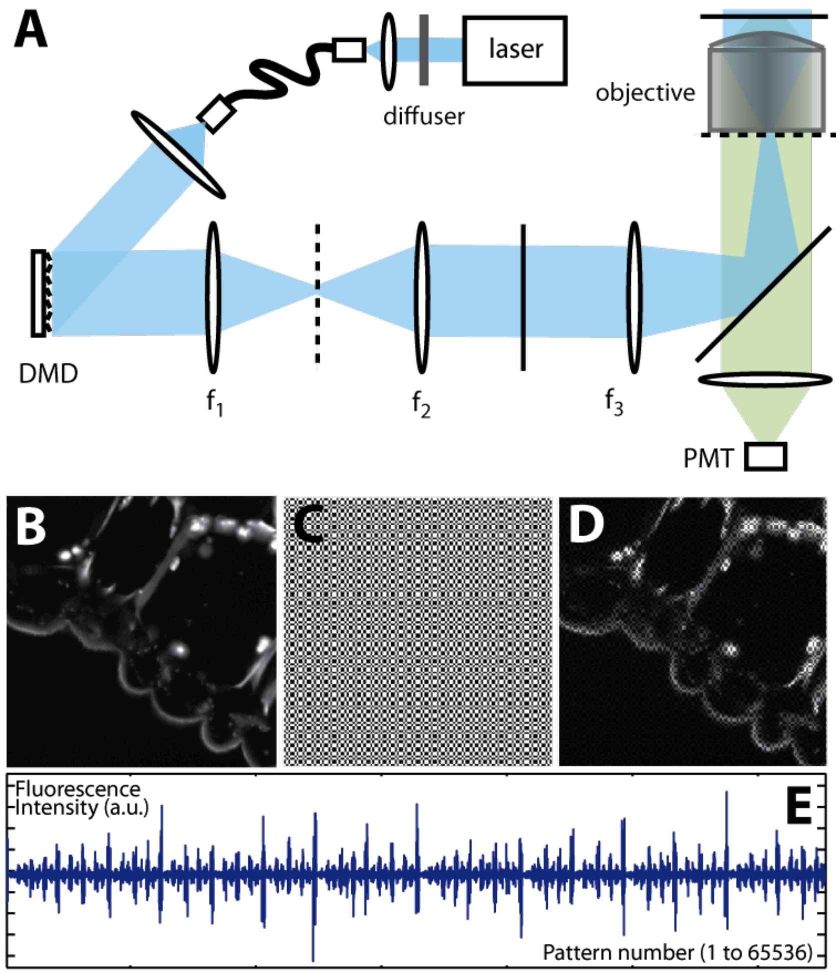
Goal: infer low-rank matrix C from noisy Y. Rank r = number of visible neurons

Additional structure: jumps in $f_i(t)$ are non-negative

Locally rank-penalized convex optimization with nonnegativity constraints to infer C, followed by iterative matrix factorization under nonnegativity constraints to infer $s_i(x), f_i(t)$ (Pnevmatikakis et al, 2013). Examples: Machado, Lacefield, Kira, Yuanjun

Compressed sensing imaging

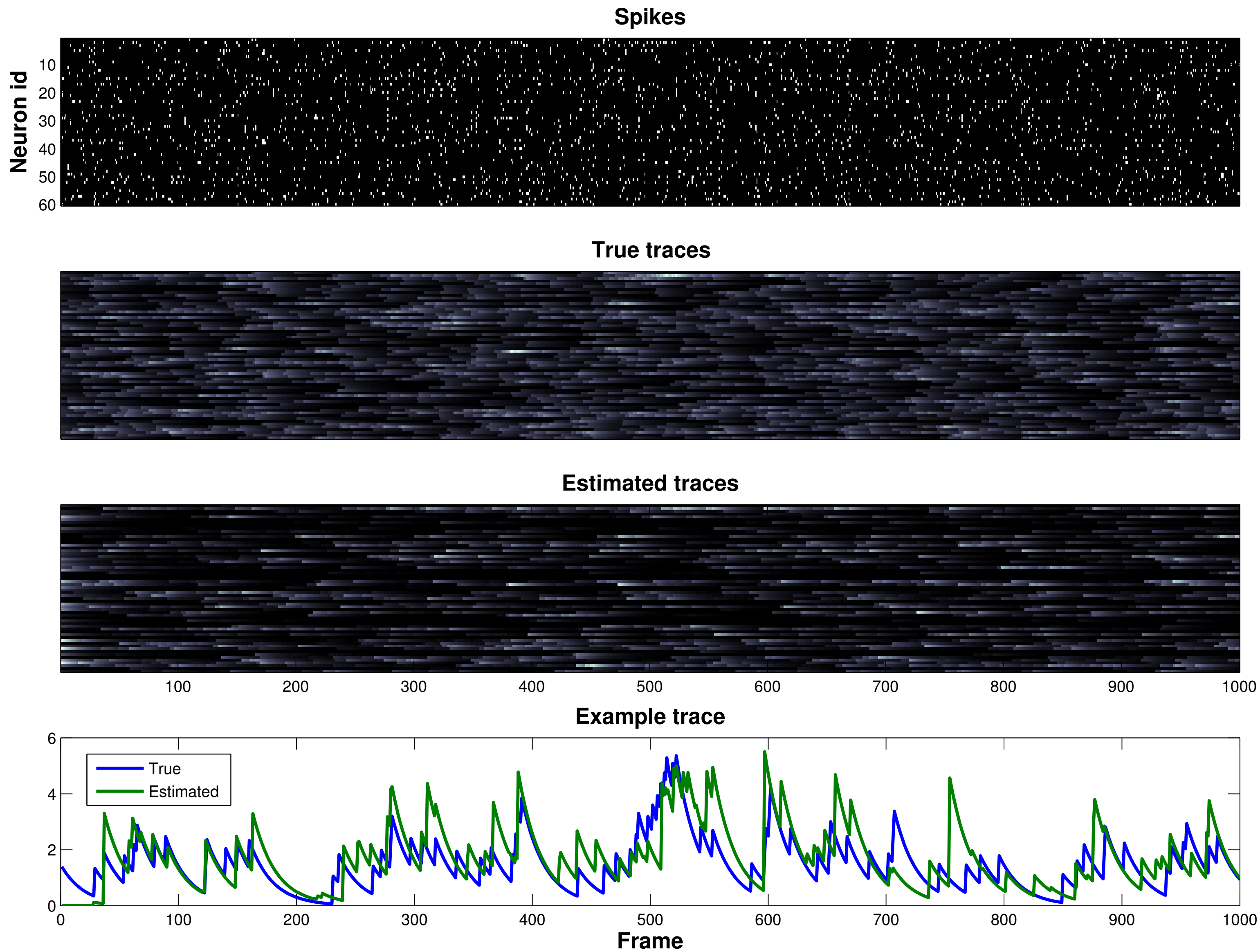
Idea: instead of raster scans, take randomized projections in each frame.



(from Studer et al, 2011)

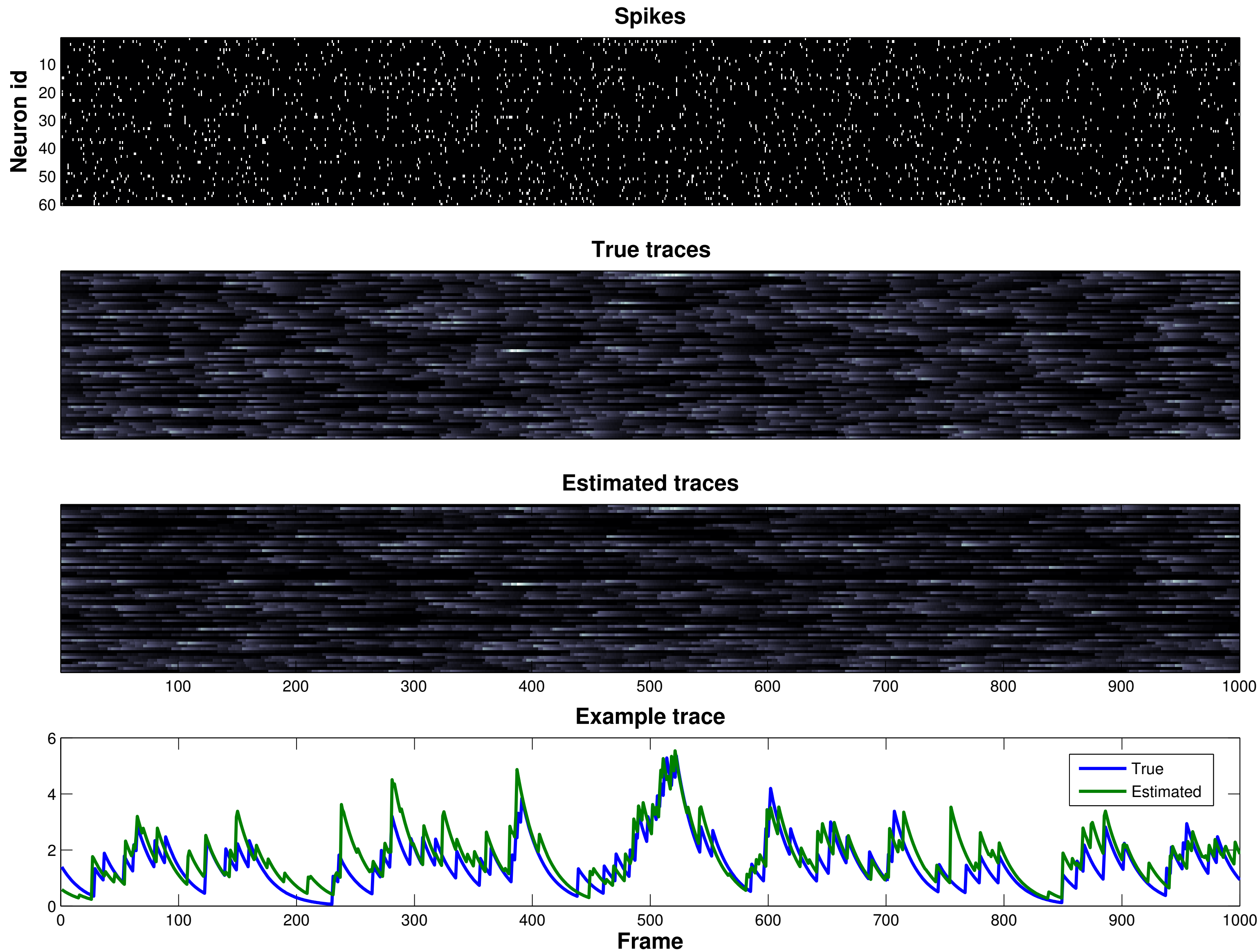
Estimating C given randomized projections Y can be cast as a similar convex optimization.

Compressed sensing imaging



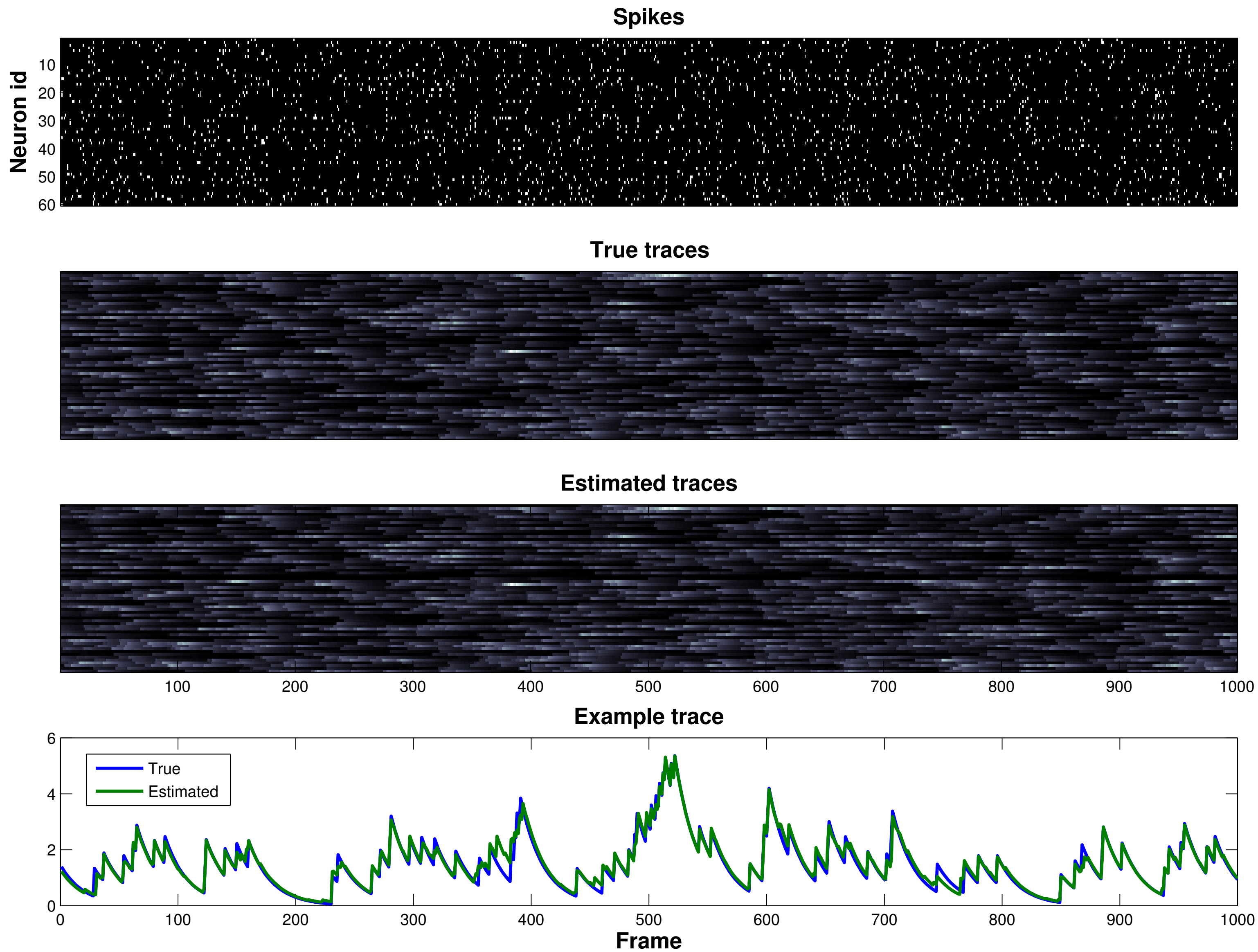
2 measurements per timestep (30x undersampling); Pnevmatikakis et al (2013)

Compressed sensing imaging



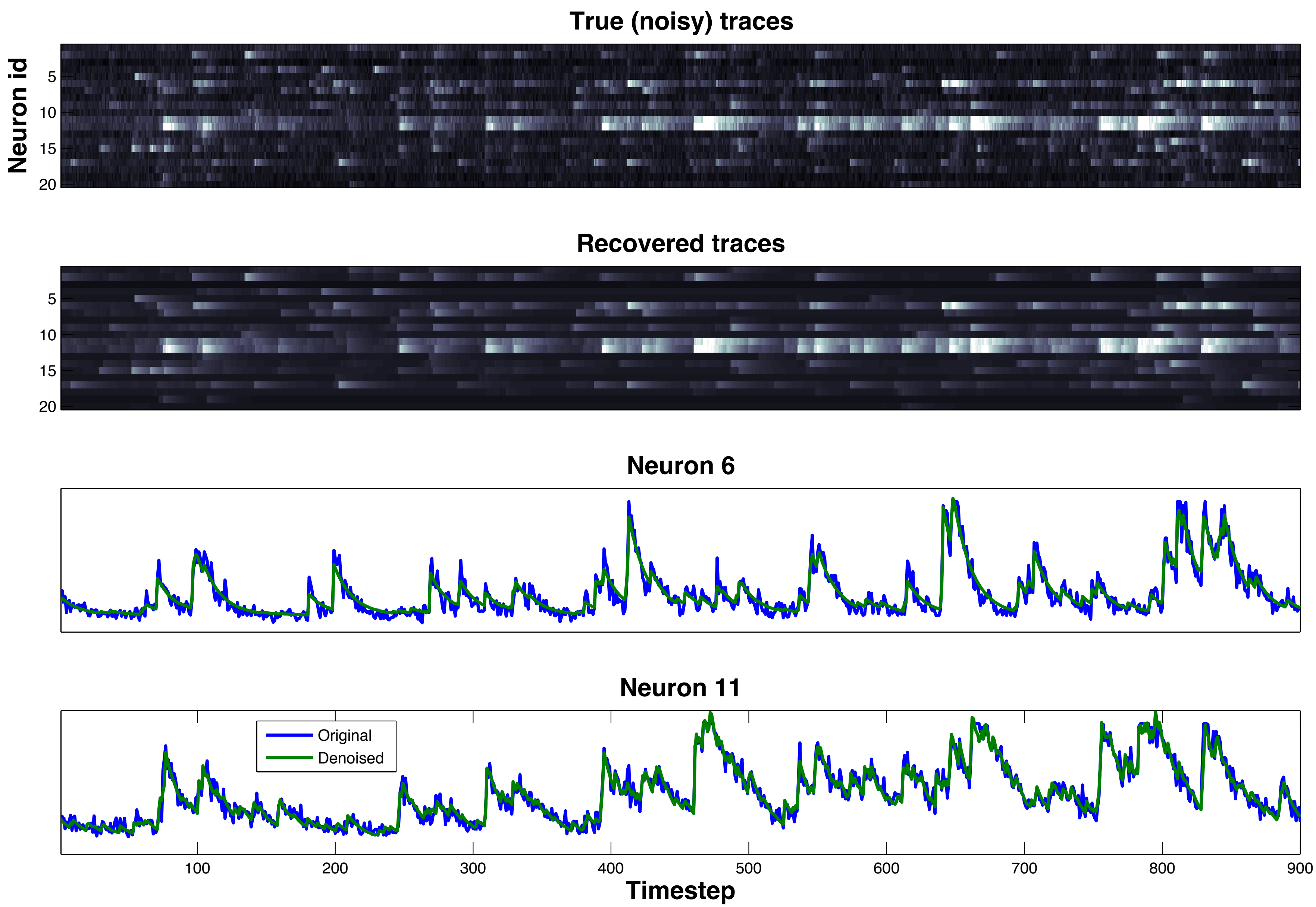
4 measurements per timestep (15x undersampling); Pnevmatikakis et al (2013)

Compressed sensing imaging



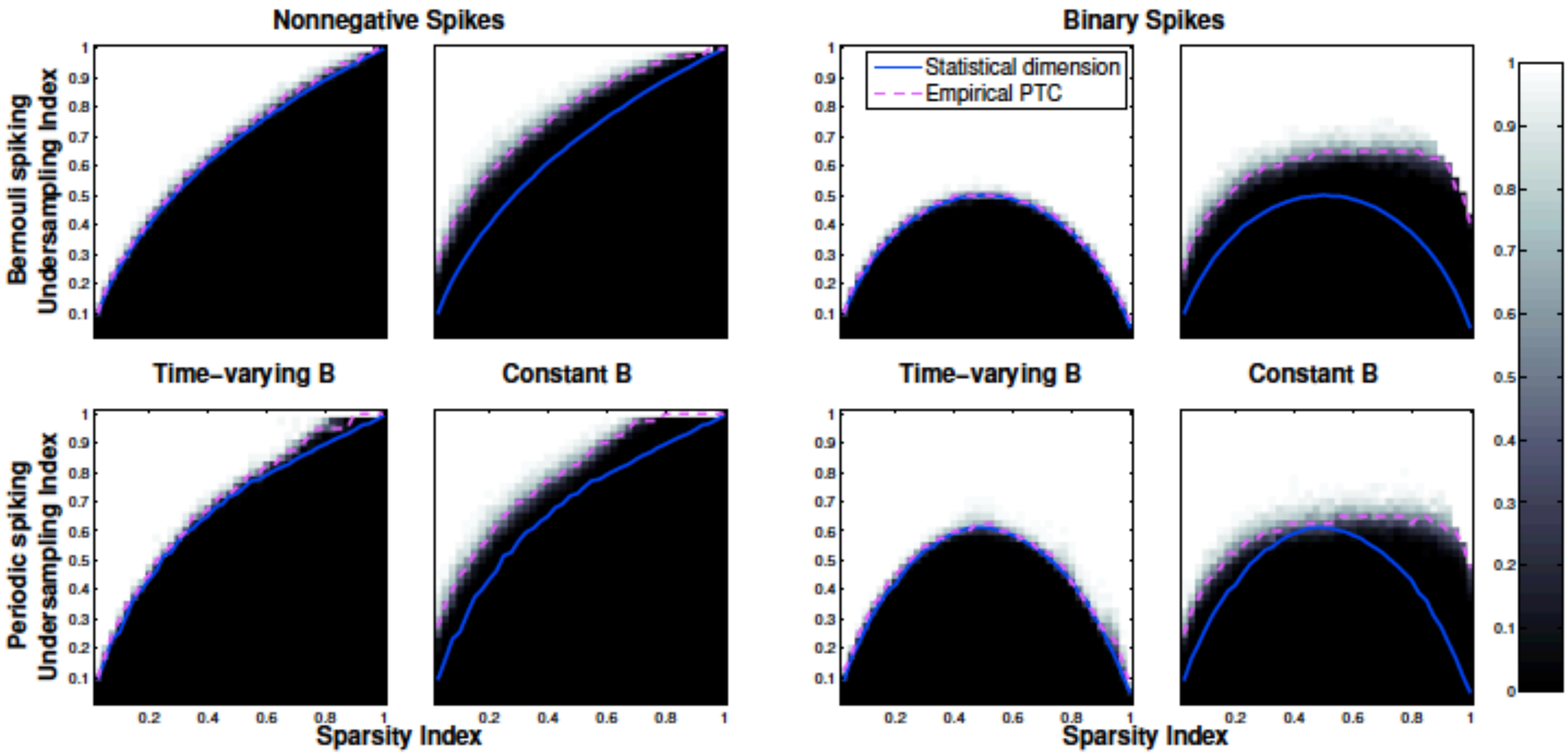
8 measurements per timestep (7.5x undersampling); Pnevmatikakis et al (2013)

Compressed real data



∼ 2x undersampling; Pnevmatikakis et al (2013). superposition movie

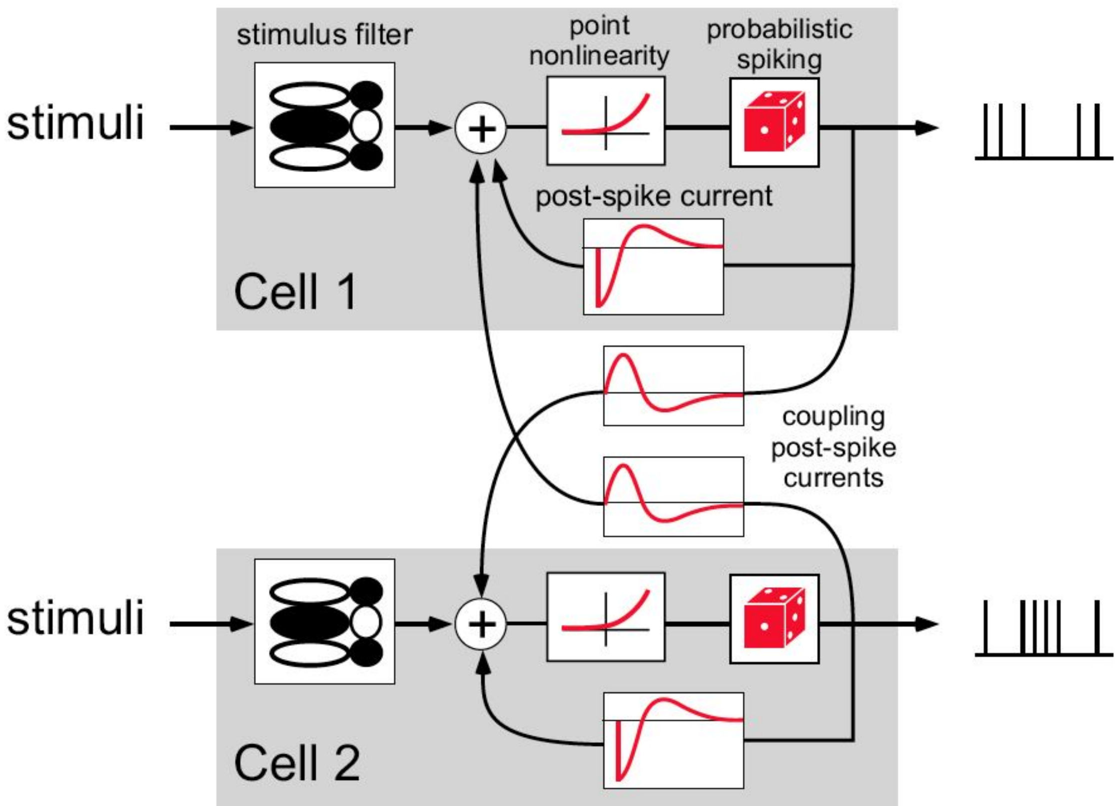
Phase transitions in decoding accuracy



New tool in compressed sensing theory: "statistical dimension" (Amelunxen, Lotz, McCoy, Tropp '13).

Interesting feature of this problem: phase transition depends on pattern of spikes, not just sparsity (as in standard LASSO problem).

Aim 2: estimating network connectivity



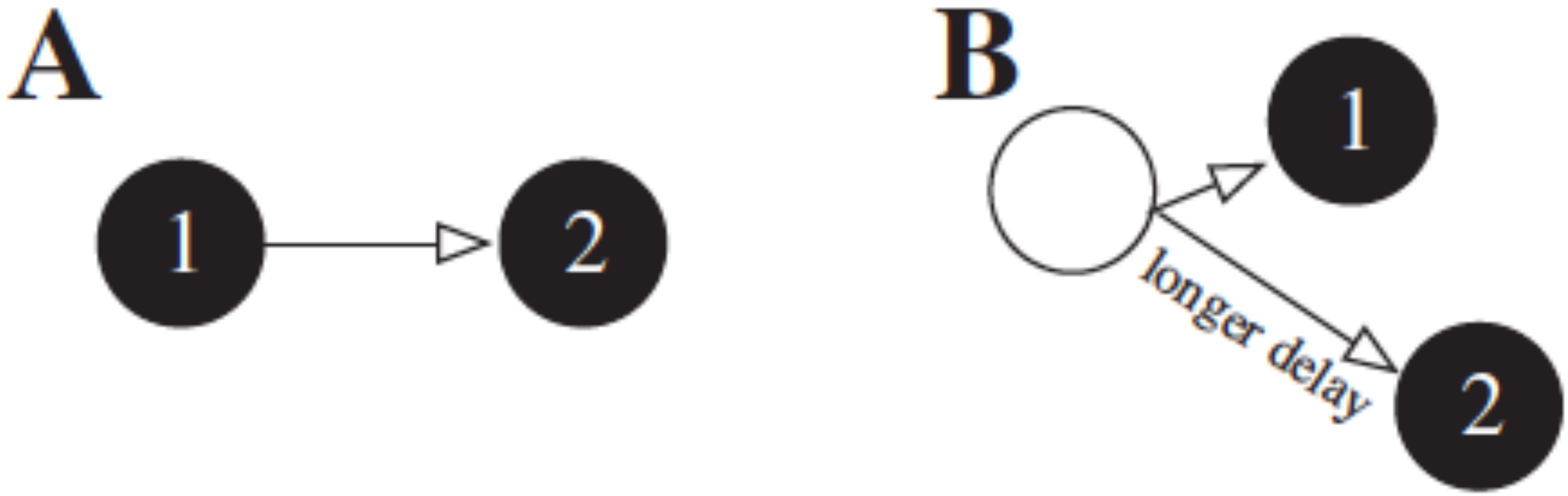
Model:

$$n_{i,t} \sim Poiss(\lambda_{i,t}), \quad \lambda_{i,t} = \exp(b_i + W_i n_{t-1} + stim.)$$

Coupled generalized linear model structure; concave loglikelihoods, optimization is straightforward (Paninski, 2004; Pillow et al., 2008). Easy to incorporate prior information about sparsity of connections, cell types, etc.

The dreaded common input problem

How to distinguish direct connectivity from common input?



(from Nykamp '07)

Previous work (e.g., Vidne et al, 2012) modeled common input terms explicitly as latent variables; works well given enough a priori information, but not a general solution.

A “shotgun sampling” approach

We can only observe K cells at a time.

Idea: don’t observe the same subset of K cells throughout the experiment.

Instead, observe as many different K-subsets as possible.

Hard with multi-electrode arrays; easy with imaging approaches.

Statistics problem: how to patch together all of the estimated subnetworks?

Want to integrate over $\{n_i(t)\}$, but scaling to large networks is a big challenge.

Approximate sufficient statistics in large Poisson regression network models

Model:

$$n_{i,t} \sim Poiss(\lambda_{i,t}), \quad \lambda_{i,t} = \exp(b_i + W_i n_{t-1})$$

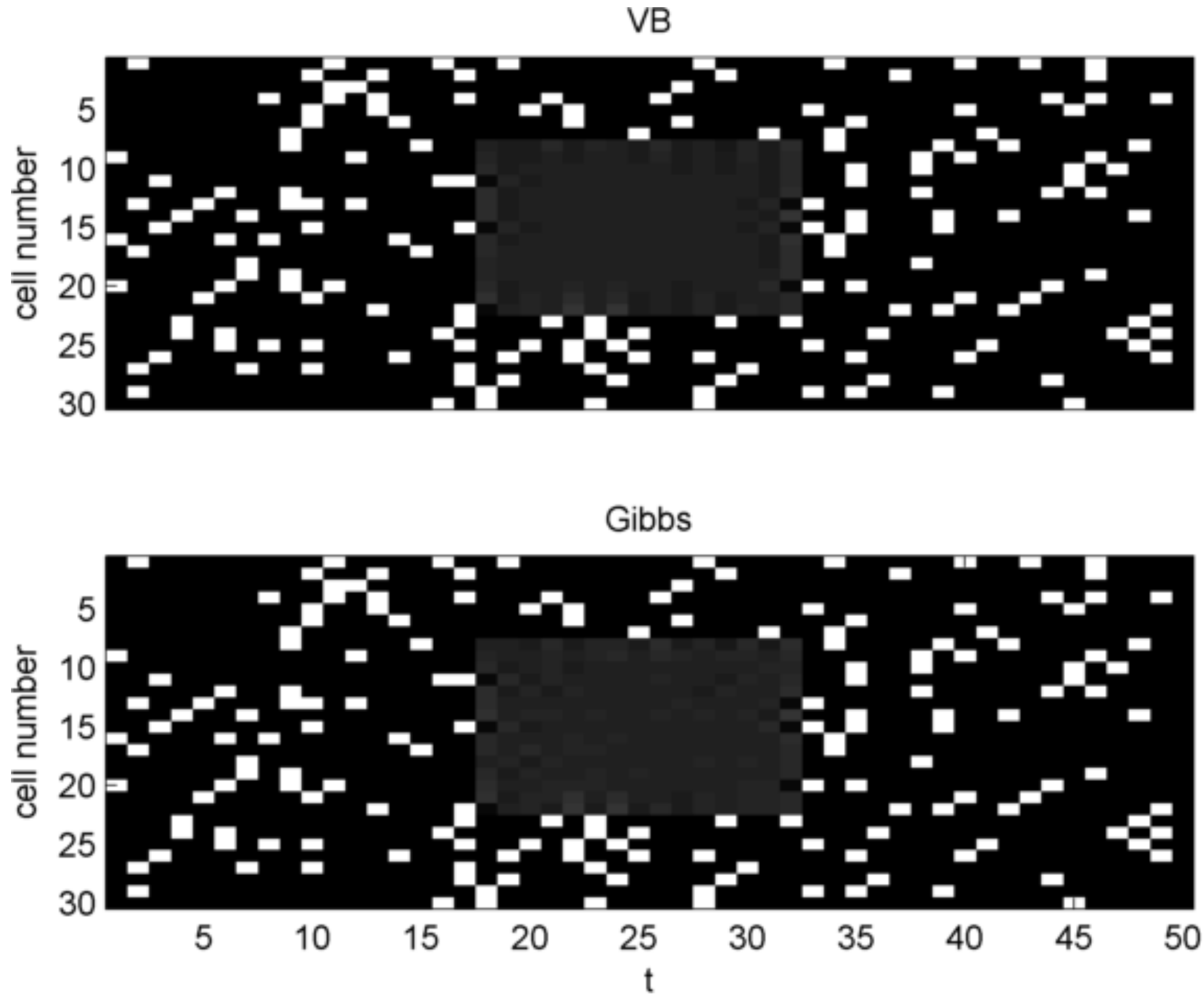
$$LL_i = \sum_t n_{i,t}(b_i + W_i n_{t-1}) - \sum_t \exp(b_i + W_i n_{t-1})$$

Idea: central limit theorem approximation for second term: $W_i n_{t-1}$ is a big sum.

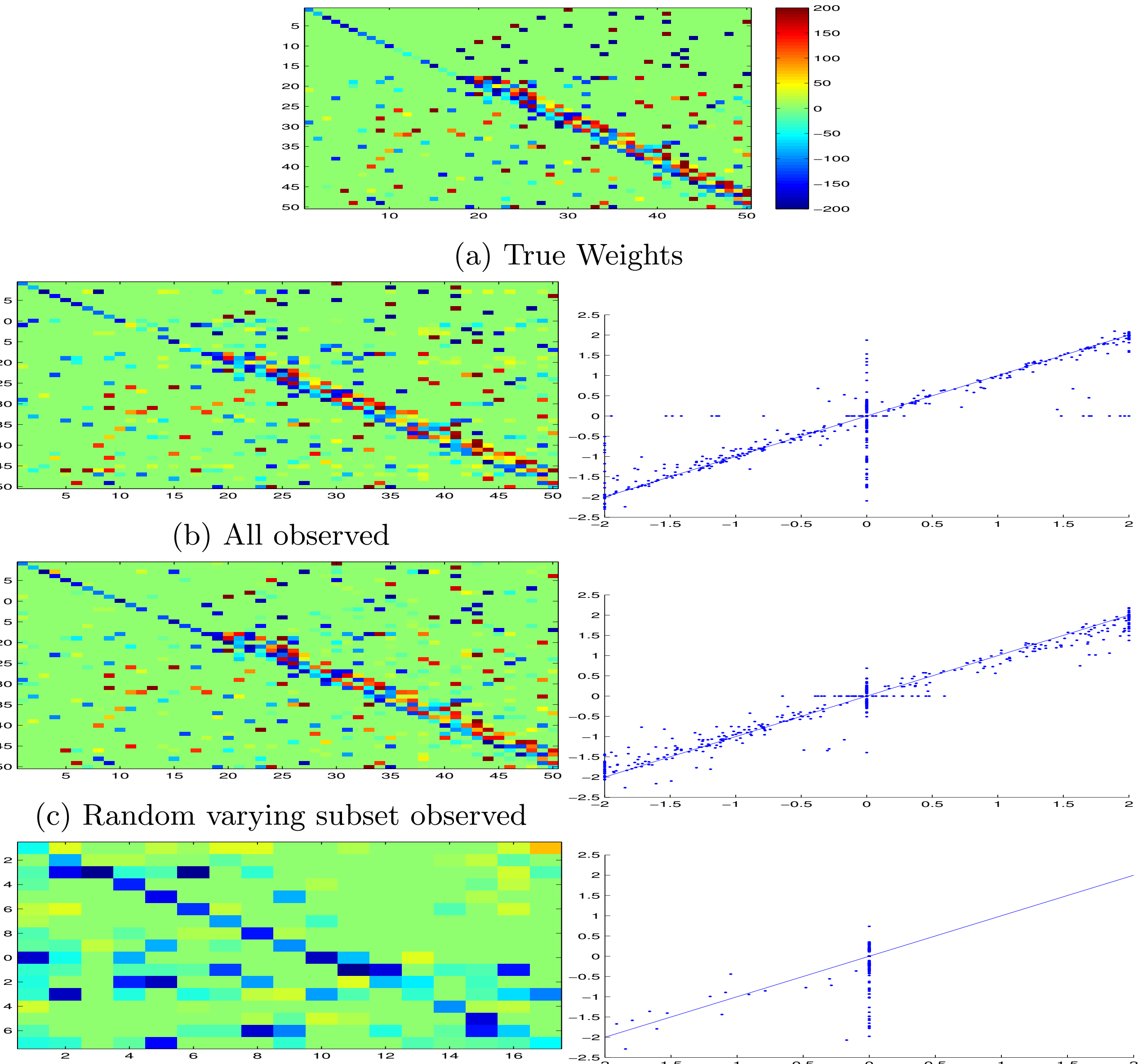
Dramatic simplification: approximate log-likelihood is quadratic! (Ramirez and Paninski '13)

Approximate sufficient statistics: $E(n_t), E(n_t n_t^T), E(n_t n_{t-1}^T)$. Can be estimated from just the observed data, or can be augmented with imputed unobserved $\{n_{i,t}\}$. Can further regularize $E(n_t n_t^T)$ via sparse-inverse penalty.

Filling in missing spikes



Simulated "shotgun" results



(a) True Weights

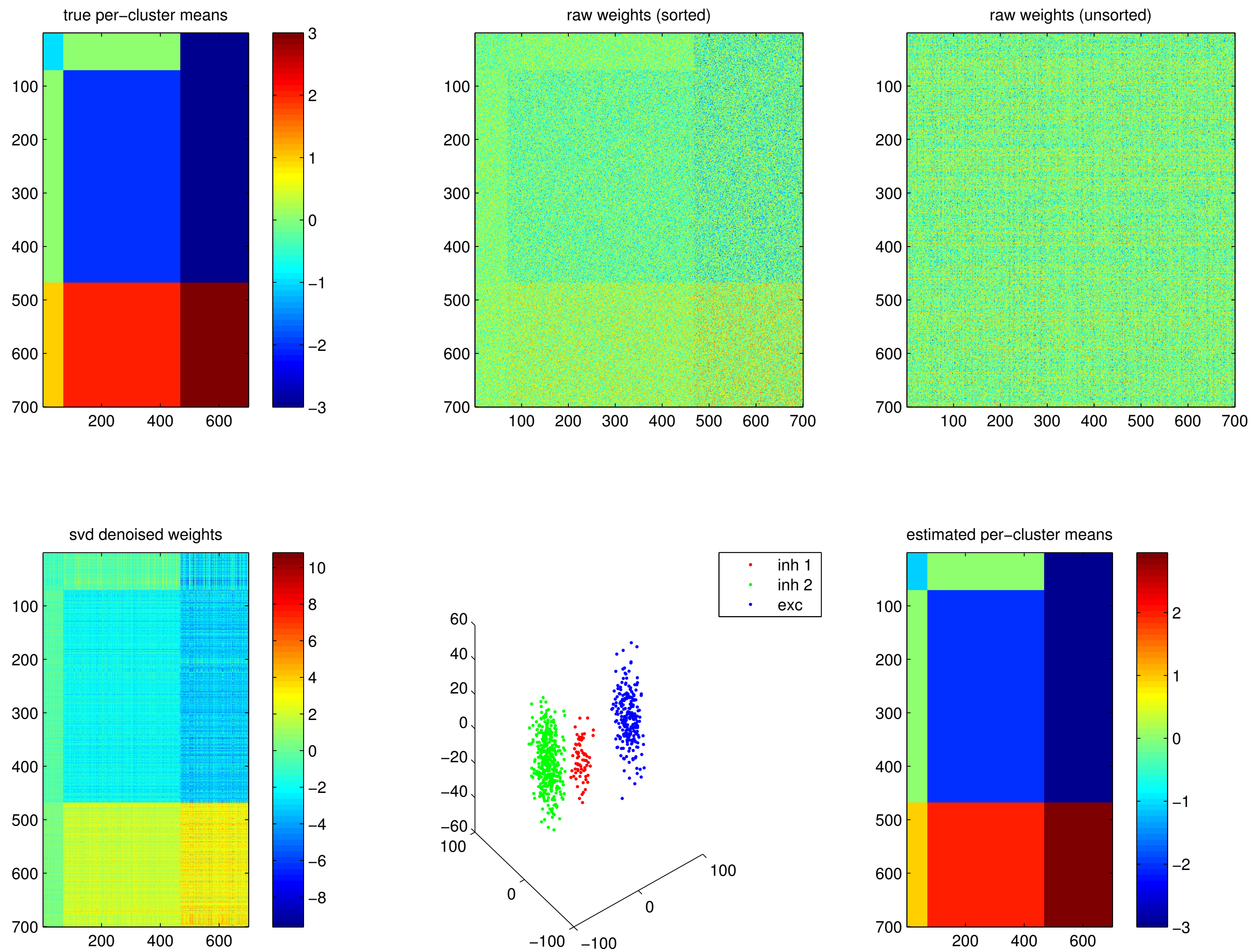
(b) All observed

(c) Random varying subset observed

(d) Fixed subset observed

$K = 20\%$ of network size; spike-and-slab priors (Keshri et al, 2013)

Incorporating cell type structure



— simple spectral clustering methods work well to infer stochastic block model structure

Aim 3: Optimal control of spike timing

To test our results, we want 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 integrate-and-soft-threshold model:

$$\begin{aligned} \lambda_t &= f(V_t + h_t) \\ V_{t+dt} &= V_t + dt\,(-gV_t + aI_t) + \sqrt{dt}\sigma\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1). \end{aligned}$$

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 tridiagonal $\implies O(T)$ optimization.

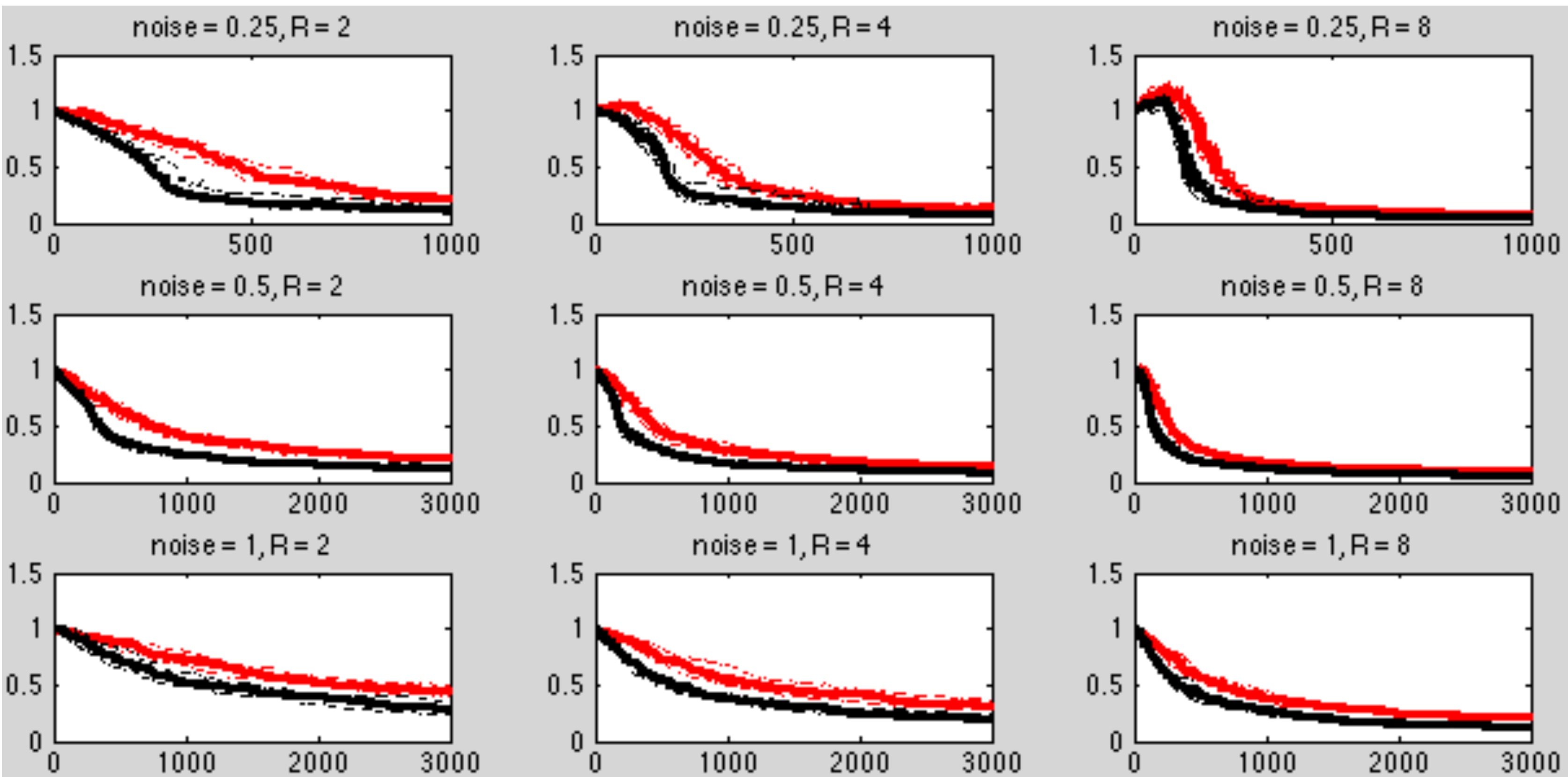
— again, can be done in real time (Ahmadian et al., 2011)... though some open challenges when I_t is high-d, spatiotemporal

Applications

- sensory prosthetics, e.g. retinal prosthetics

- fine-grained behavioral control

- online adaptive experimental design: choose stimuli which provide as much information about network as possible. Major problem here: updating sparse posteriors. Can speed inference significantly (Shababo, Paige et al, '13)



Robust point-process dimensionality reduction

Low-dimensional latent variable z_t

Fixed matrix B mapping z_t up to the higher-dimensional neural rate space

Simplest firing rate model:

$$n_i(t) \sim Poiss(\lambda_i(t))$$

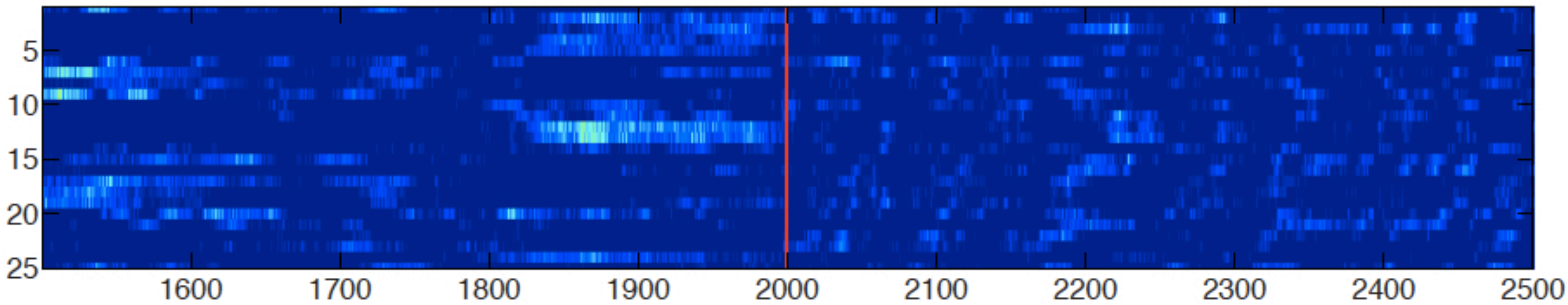
$$\log \lambda_i(t) = B_i z_t = r_i(t)$$

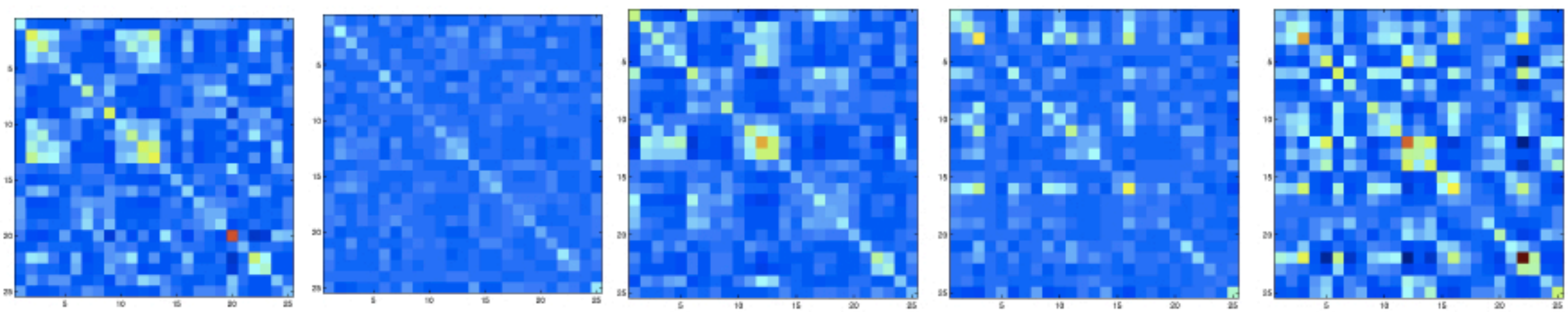
(can be generalized easily to include stimulus terms, spike history effects, etc.)

How to estimate low-d structure B without making a lot of assumptions about the dynamics (e.g., linear, Gaussian) of z_t?

Generalization of PCA: max. likelihood of spike data $n = \{n_i(t)\}$ as function of $R = \{r_i(t)\}$, while minimizing rank of R: $rank(R) \leq dim(z)$. Nuclear norm $\implies$ convex problem.

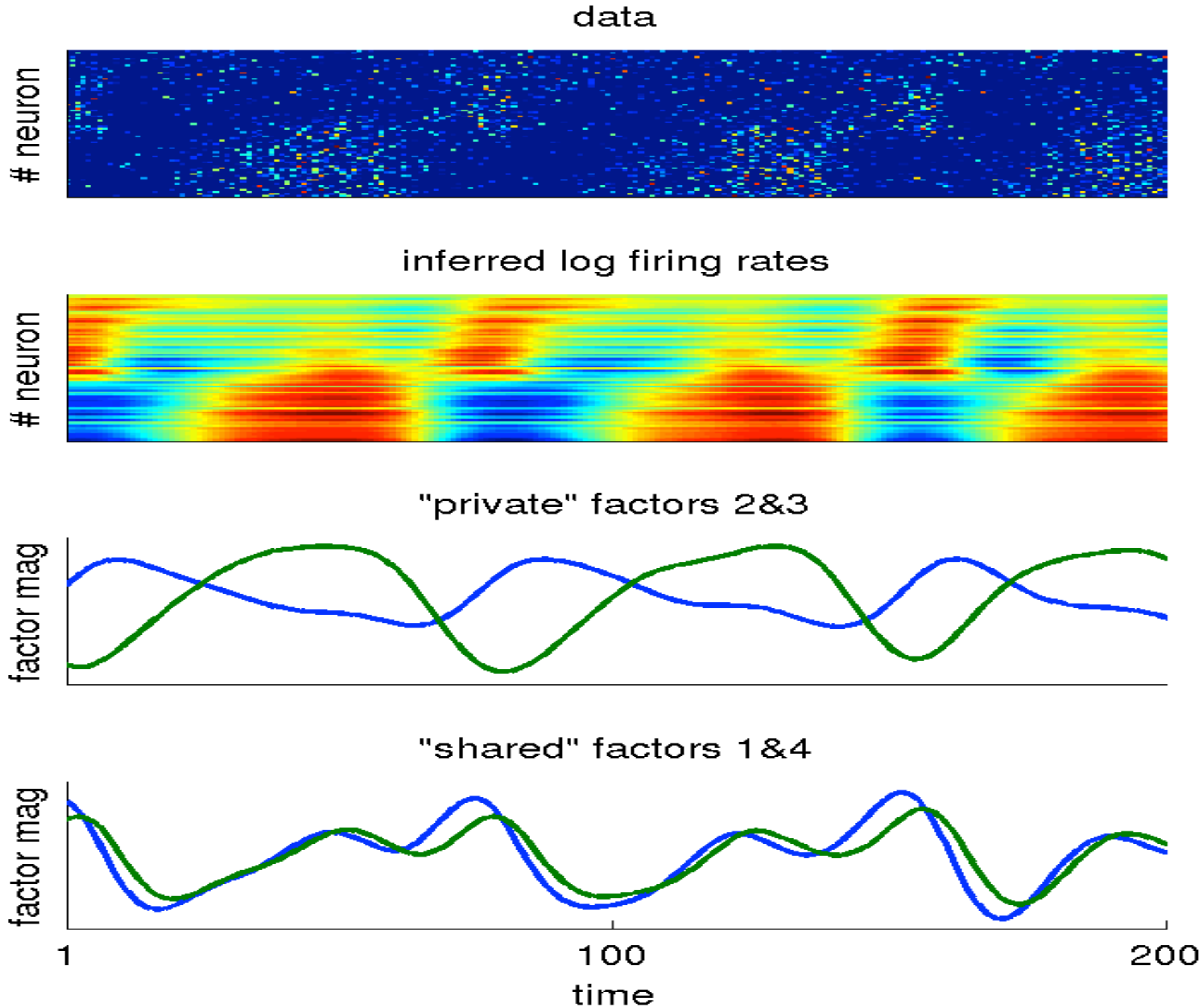
Dimensionality reduction of non stationary spiking data





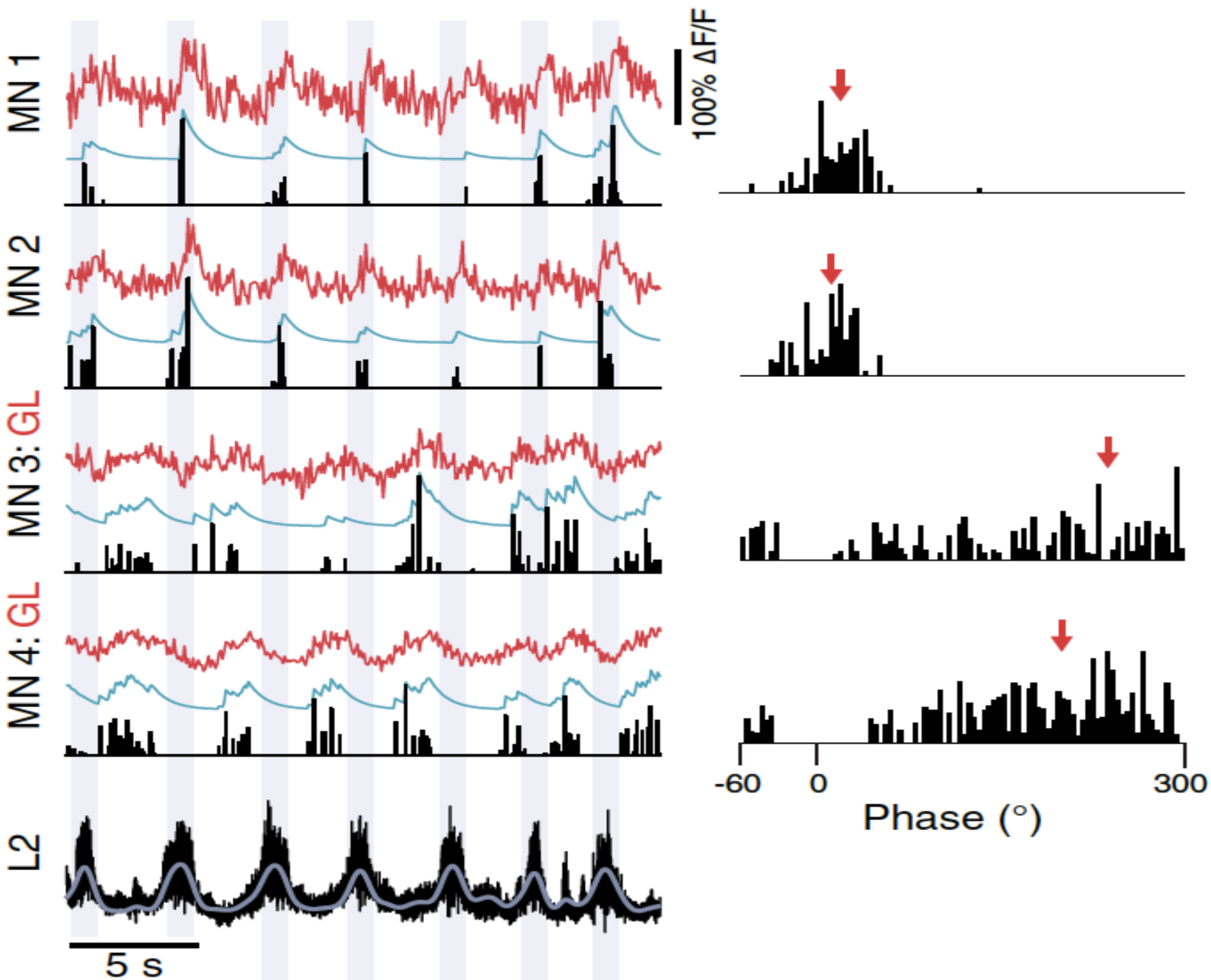
Method can infer B well even in highly non stationary settings — see Pfau, Pnevmatikakis et al (2013) for details.

Smooth, clustered activity in spinal cord

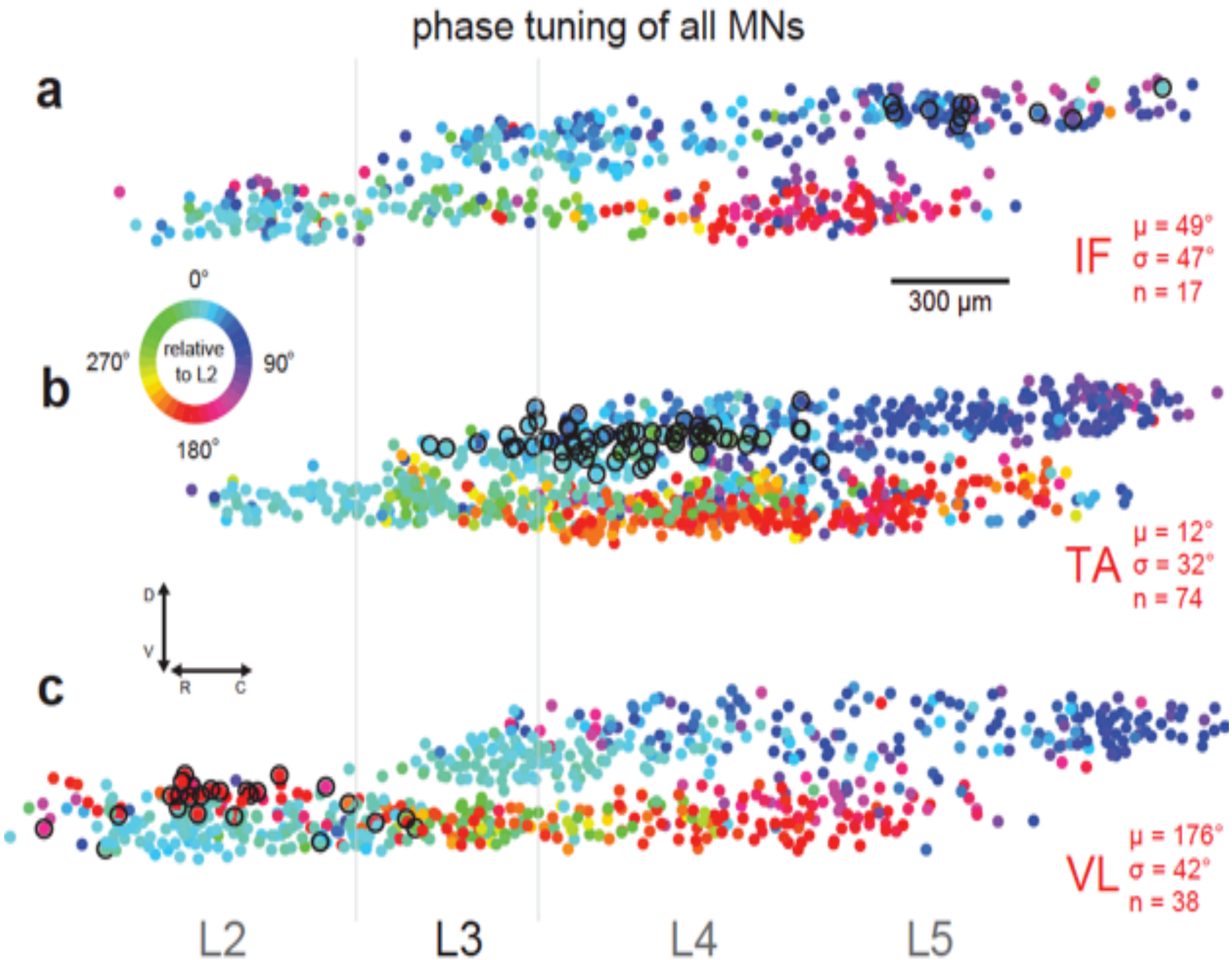


— Machado, Buesing et al. 2014. data; spikes and rates

Measuring phase tuning in single neurons

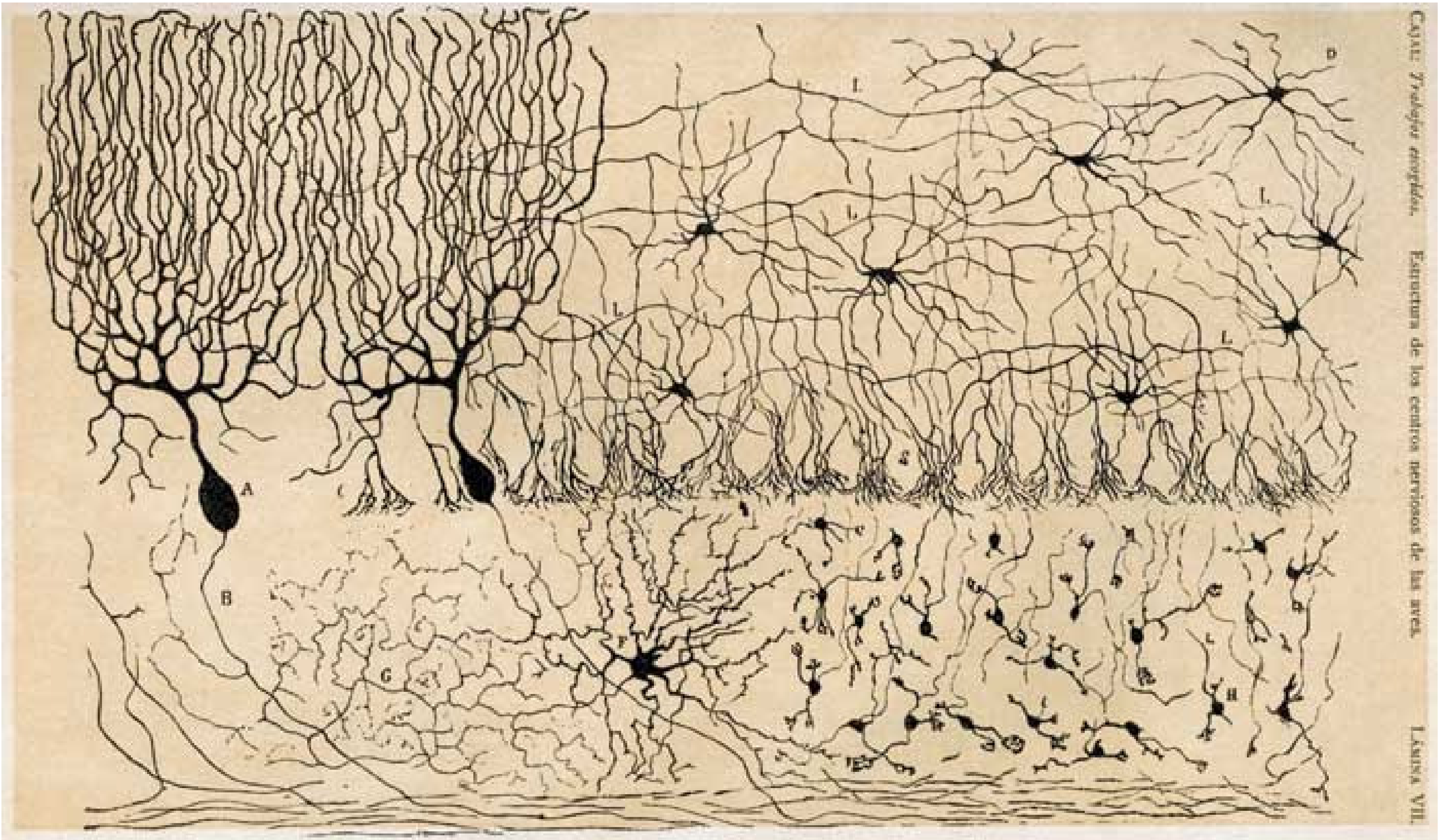


Mapping phase tuning across the population



— Machado, Buesing et al. 2014

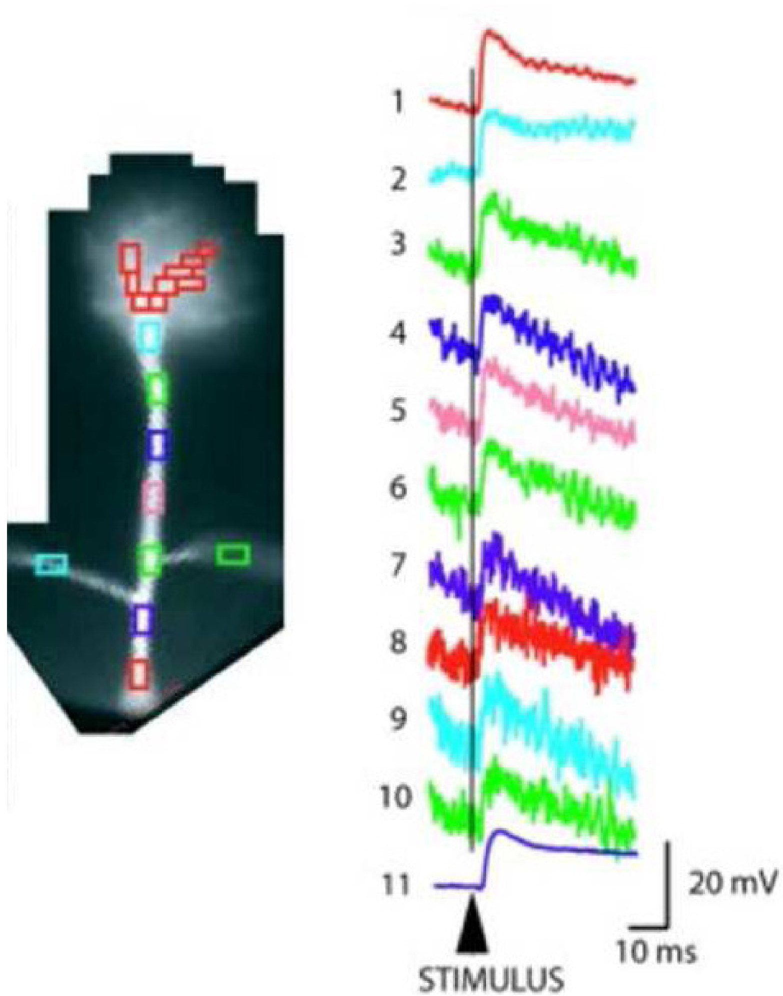
Extension: Connectivity at the dendritic scale



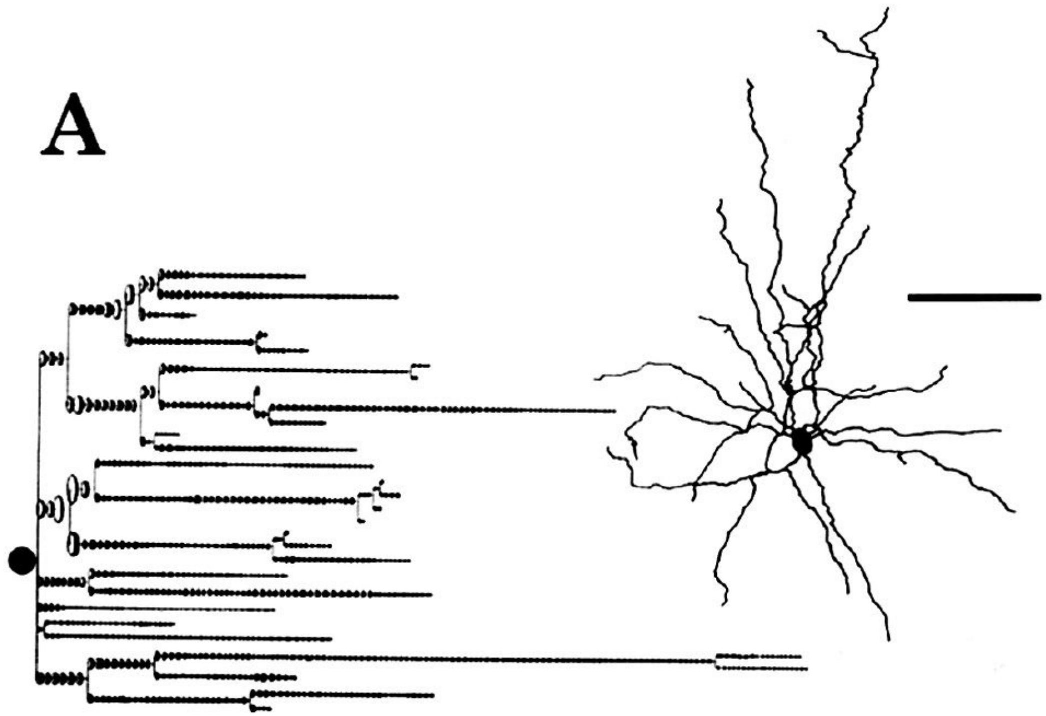
Ramon y Cajal, 1888.

Signal processing on trees

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

Simplest case: Kalman filter

Dynamics and observation equations:

$$d\vec{V}/dt = A\vec{V} + \vec{\epsilon}_t$$

$$\vec{y}_t = B_t\vec{V} + \vec{\eta}_t$$

$V_i(t)$ = voltage at compartment i

A = cable dynamics matrix: includes leak terms ($A_{ii} = -g_l$) and intercompartmental terms ($A_{ij} = 0$ unless compartments are adjacent)

B_t = observation matrix: point-spread function of microscope

Even this case is challenging, since $d = \dim(\vec{V})$ is very large

Standard Kalman filter: $O(d^3)$ computation per timestep (matrix inversion)

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.

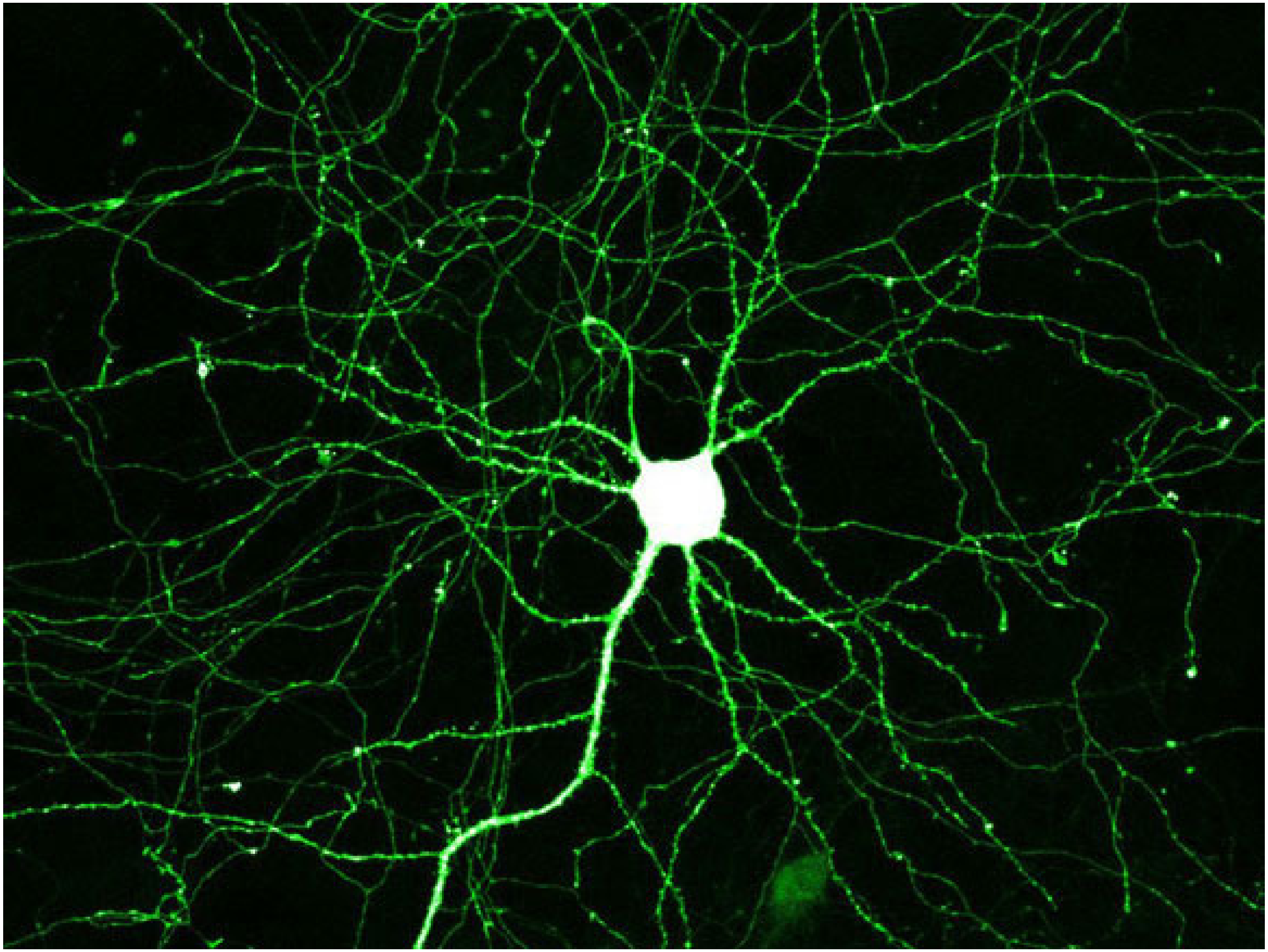
The necessary recursions — i.e., updating U_t, D_t and the Kalman mean $E(q_t|Y_{1:t})$ — involve linear manipulations of C_0, using

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

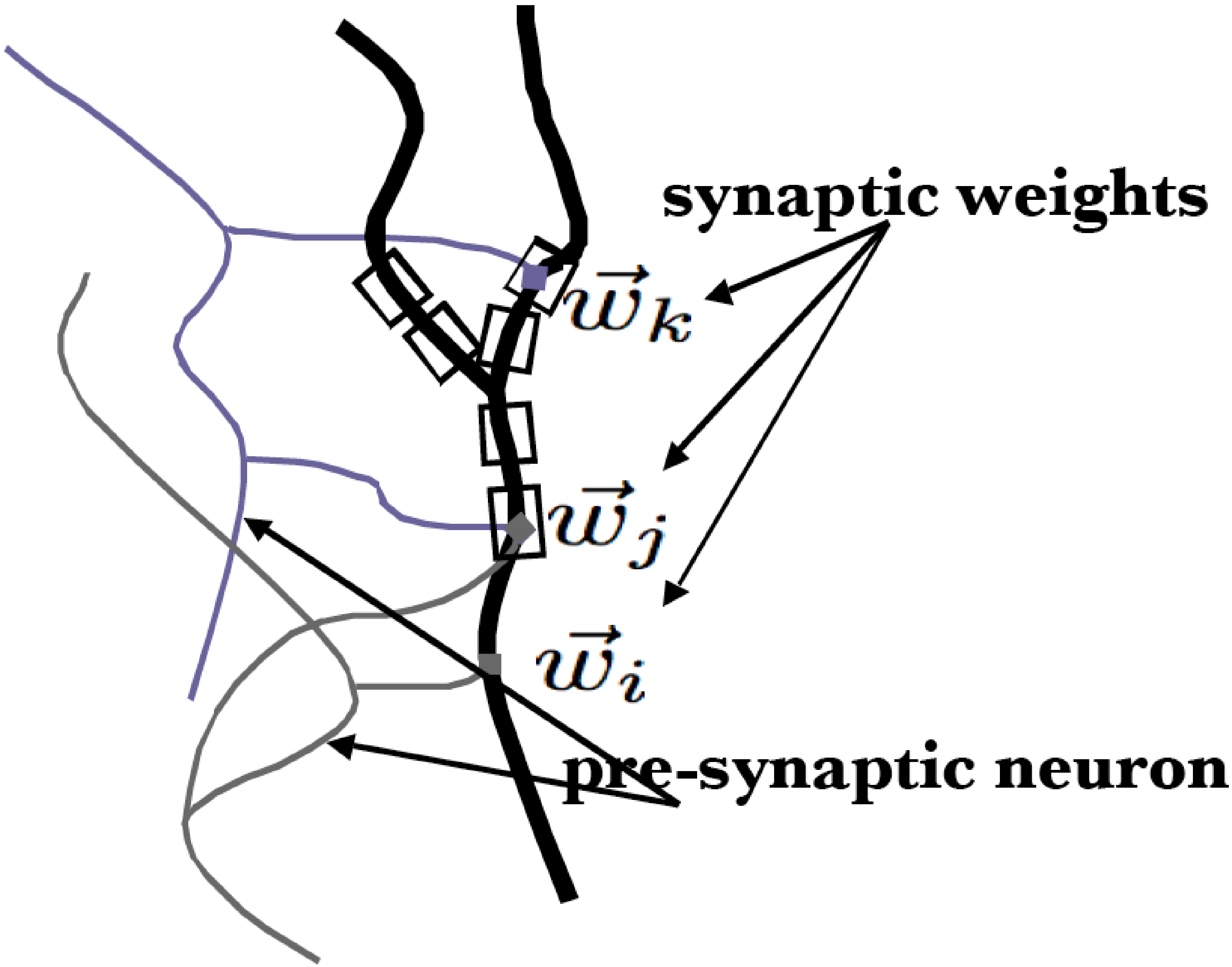
and can be done in $O(\dim(q))$ time (Paninski, 2010). Generalizable to many other state-space models (Pnevmatikakis and Paninski, 2011).

Examples: speckle, vertical

Application: synaptic locations/weights



Application: synaptic locations/weights



Application: synaptic locations/weights

Including known terms:

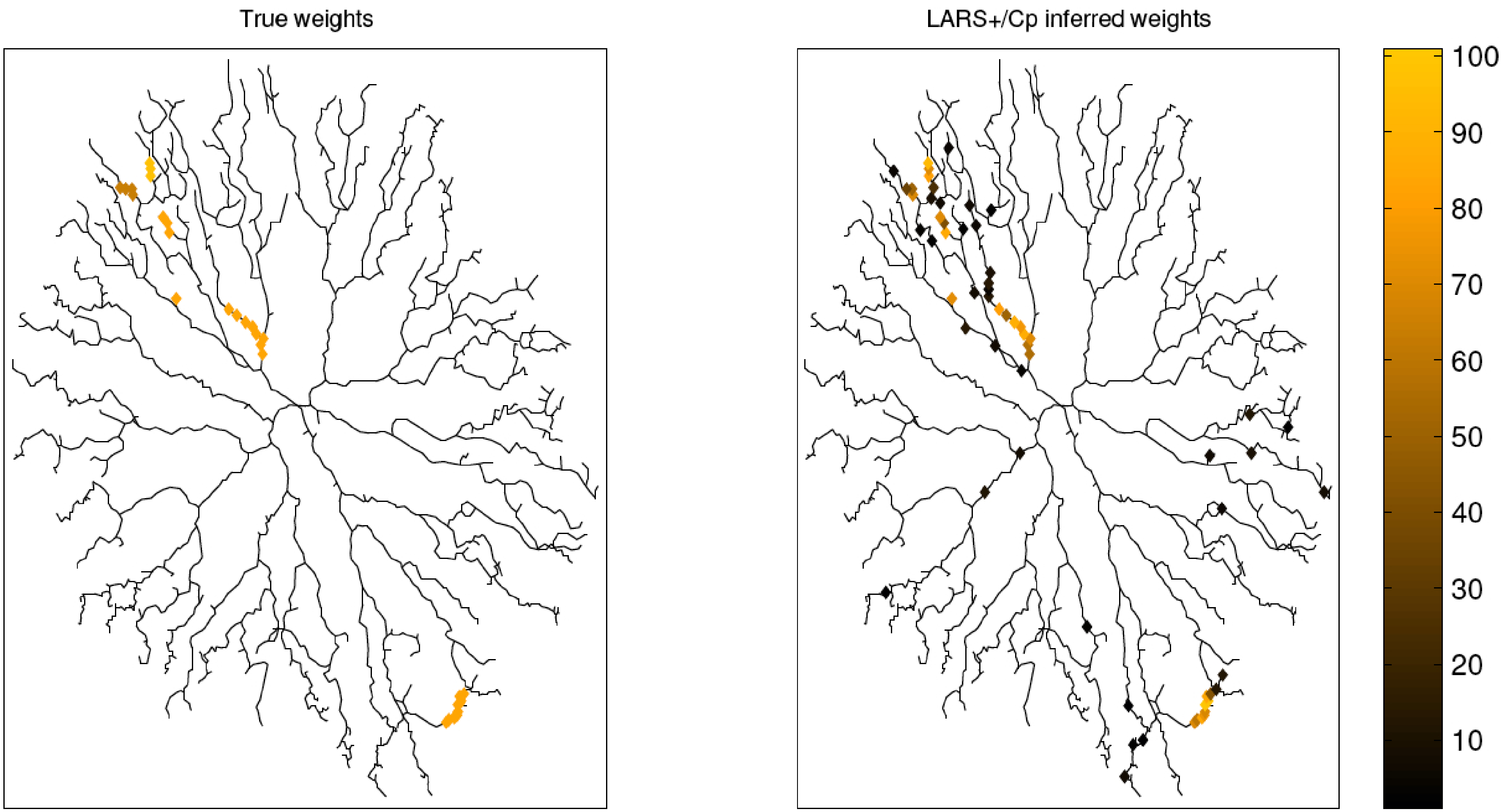
$$d\vec{V}/dt = A\vec{V}(t) + W\vec{U}(t) + \vec{\epsilon}(t);$$

$U(t)$ 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. L_1-penalized loglikelihood can be optimized efficiently with homotopy (LARS) approach.

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

Example: real neural geometry



700 timesteps observed; 40 random compartments (of > 2000) observed per timestep. Zecevic data

Compressed sensing measurements improve accuracy further (Pakman et al 2013).

Conclusions

- Modern statistical approaches provide flexible, powerful methods for answering key questions in neuroscience — many of these problems are statistics problems in disguise
- Close relationships between biophysics, statistical modeling, and experimental design
- 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

References

Ahmadian, Y., Packer, A. M., Yuste, R., and Paninski, L. (2011). Designing optimal stimuli to control neuronal spike timing. *Journal of Neurophysiology*, 106(2):1038–1053.

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.

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.

Koyama, S. and Paninski, L. (2010). Efficient computation of the map path and parameter estimation in integrate-and-fire and more general state-space models. *Journal of Computational Neuroscience*, 29:89–105.

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.

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.

Pnevmatikakis, E. and Paninski, L. (2011). Fast interior-point inference in high-dimensional sparse, penalized state-space models. *Submitted.*

Vogelstein, J., Packer, A., Machado, T., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast non-negative deconvolution for spike train inference from population calcium imaging. *J. Neurophys.*, 104:3691–704.