Coding and computation by neural ensembles in the primate retina

Liam Paninski

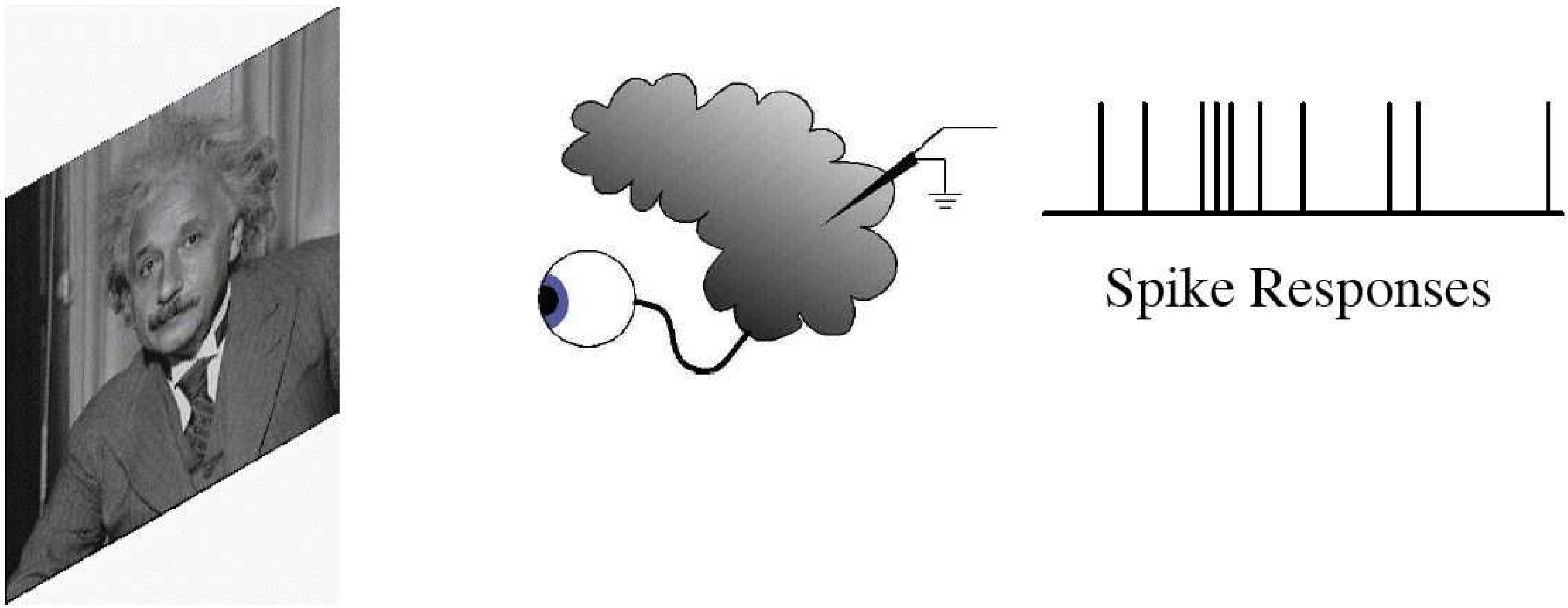
Department of Statistics and Center for Theoretical Neuroscience
Columbia University
http://www.stat.columbia.edu/∼liam
liam@stat.columbia.edu

October 13, 2008

— with J. Pillow (Gatsby), E. Simoncelli (NYU), E.J. Chichilnisky, J. Shlens (Salk), E. Lalor (TC Dublin), S. Koyama (CMU), Y. Ahmadian, J. Kulkarni, D. Pfau, X. Pitkow, K. Rahnama Rad, T. Toyoizumi, M. Vidne (Columbia).

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

The neural code



Input-output relationship between

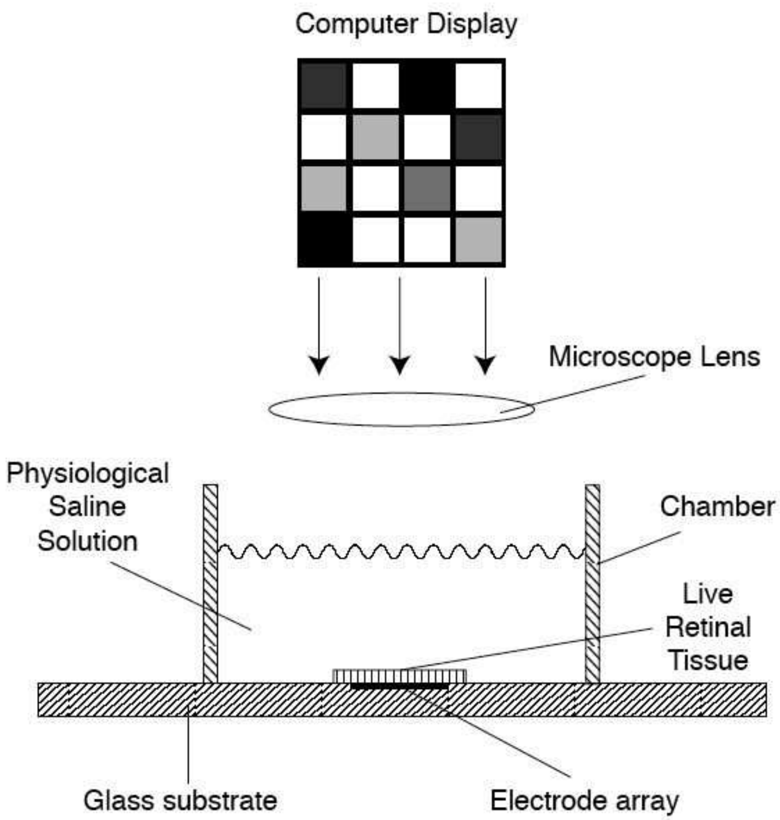
- External observables x (sensory stimuli, motor responses...)
- Neural variables y (spike trains, population activity...)

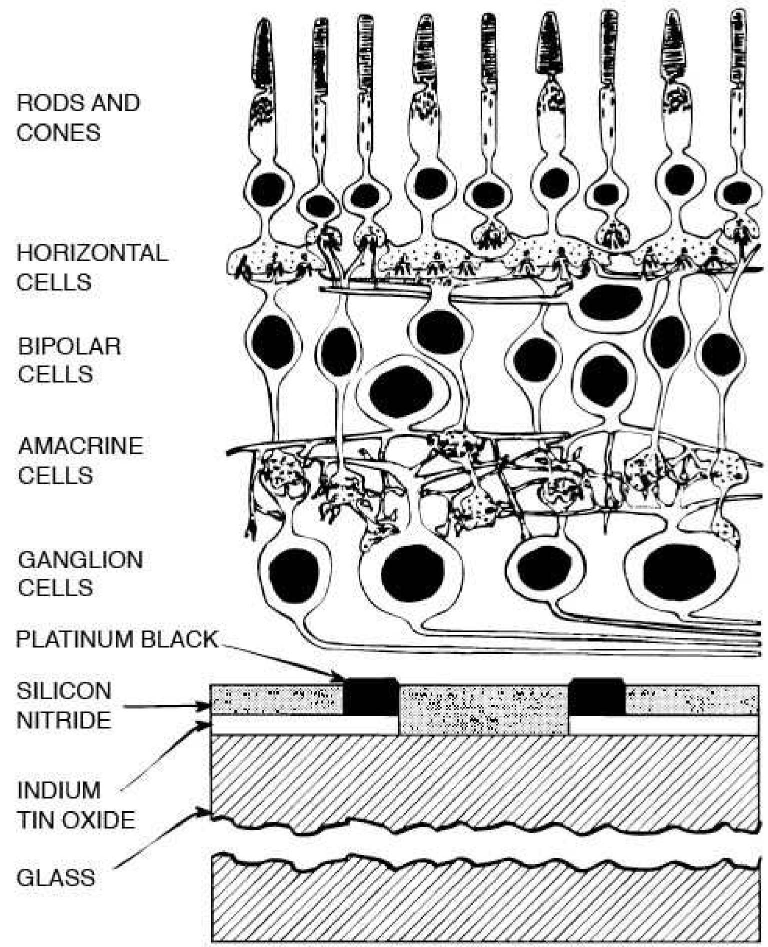
Encoding problem: $p(y|x)$; decoding problem: $p(x|y)$

Retinal ganglion neuronal data

Preparation: dissociated macaque retina

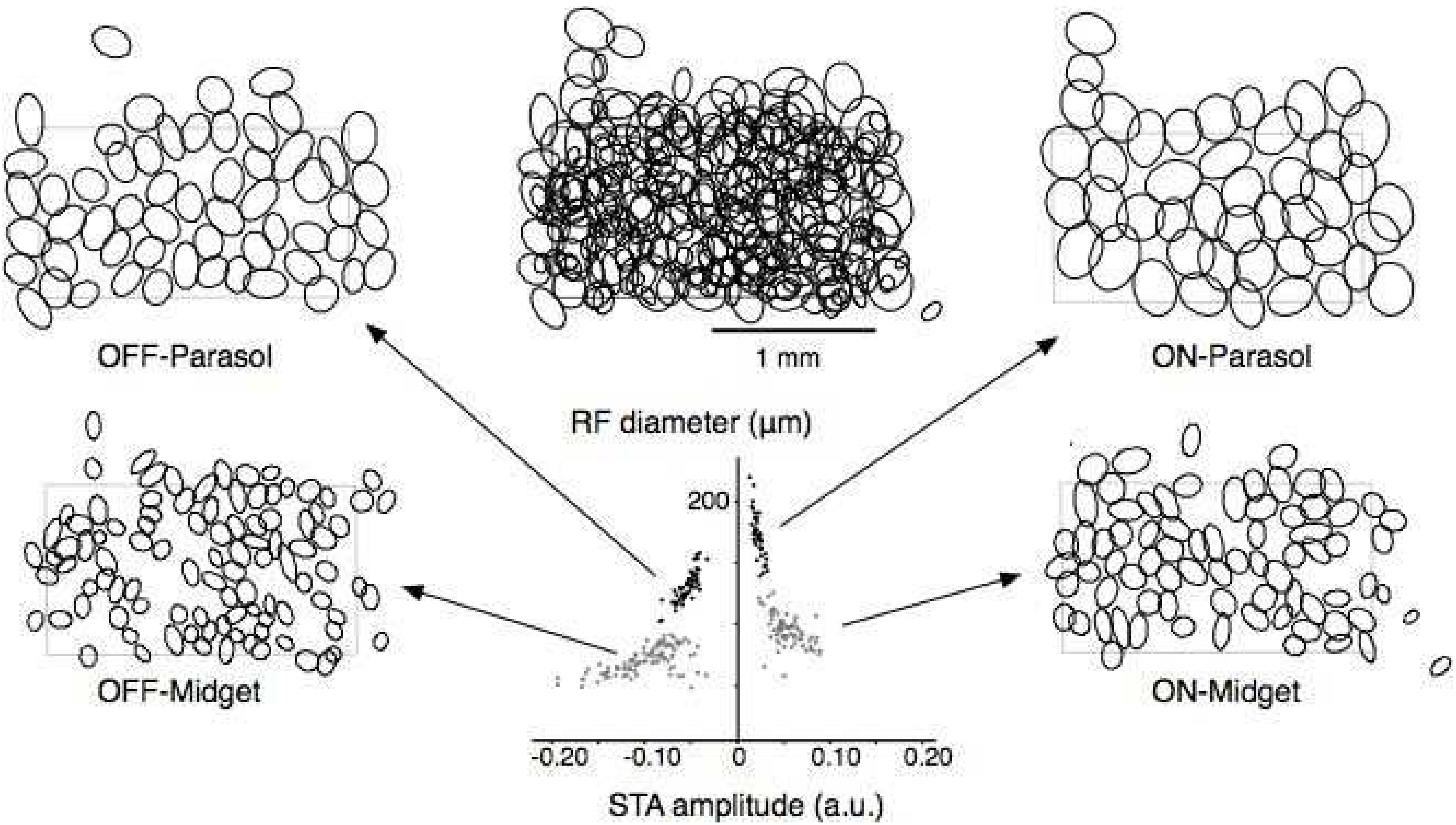
— extracellularly-recorded responses of populations of RGCs



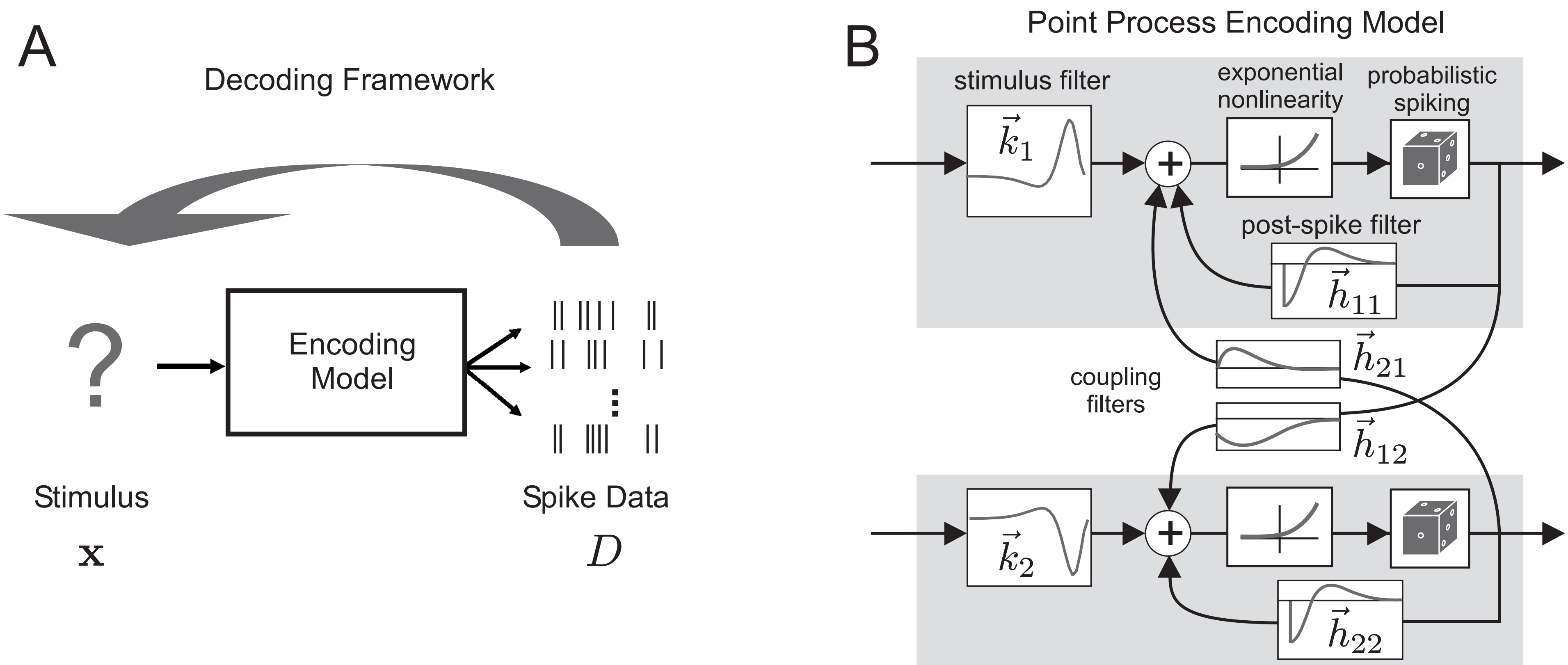


Stimulus: random spatiotemporal visual stimuli (Pillow et al., 2008b)

Receptive fields tile visual space



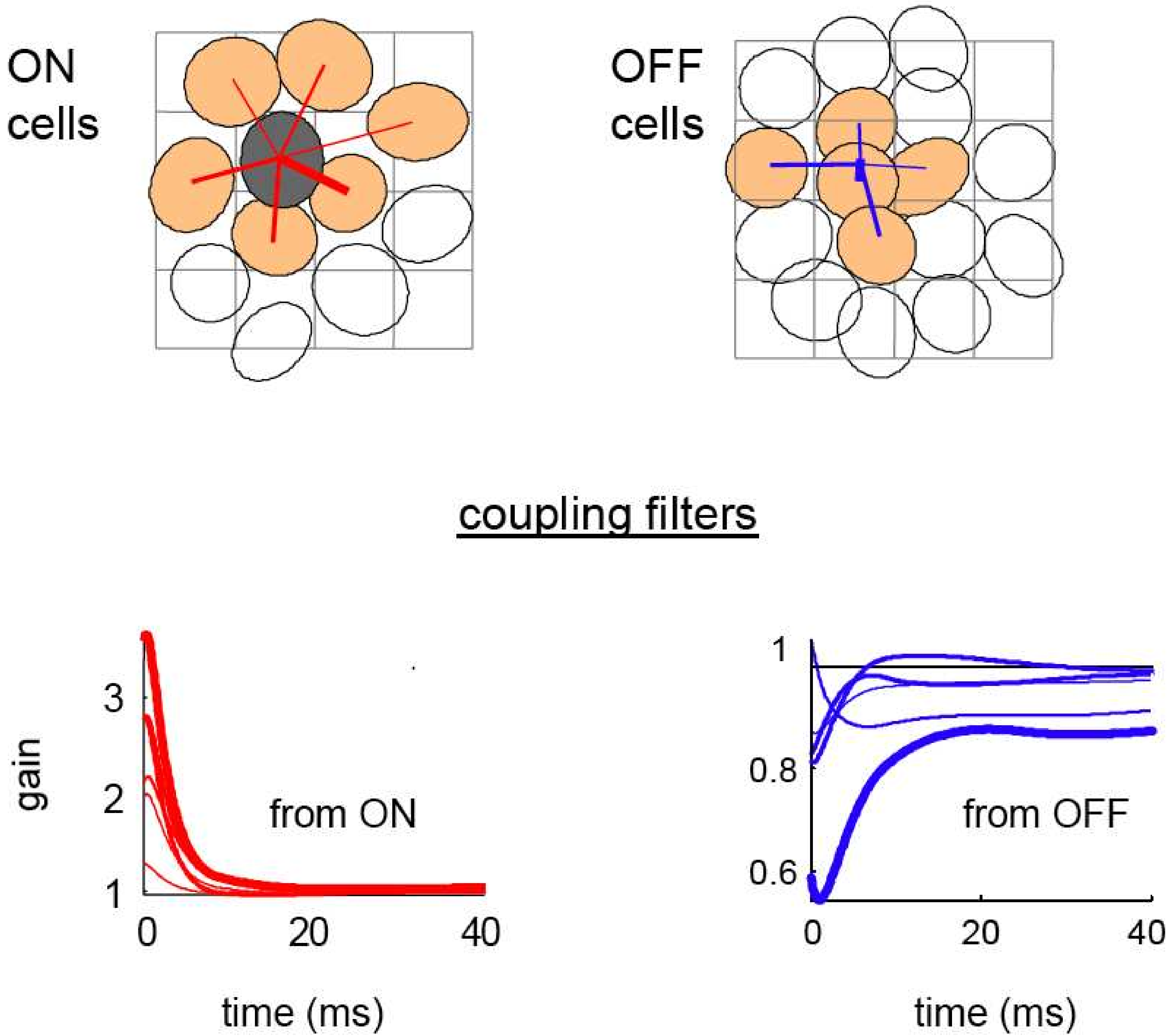
Multineuronal point-process model



$$\lambda_i(t) = f\left(b_i + \vec{k}_i \cdot \vec{x}(t) + \sum_{i',j} h_{i',j} n_{i'}(t-j)\right),$$

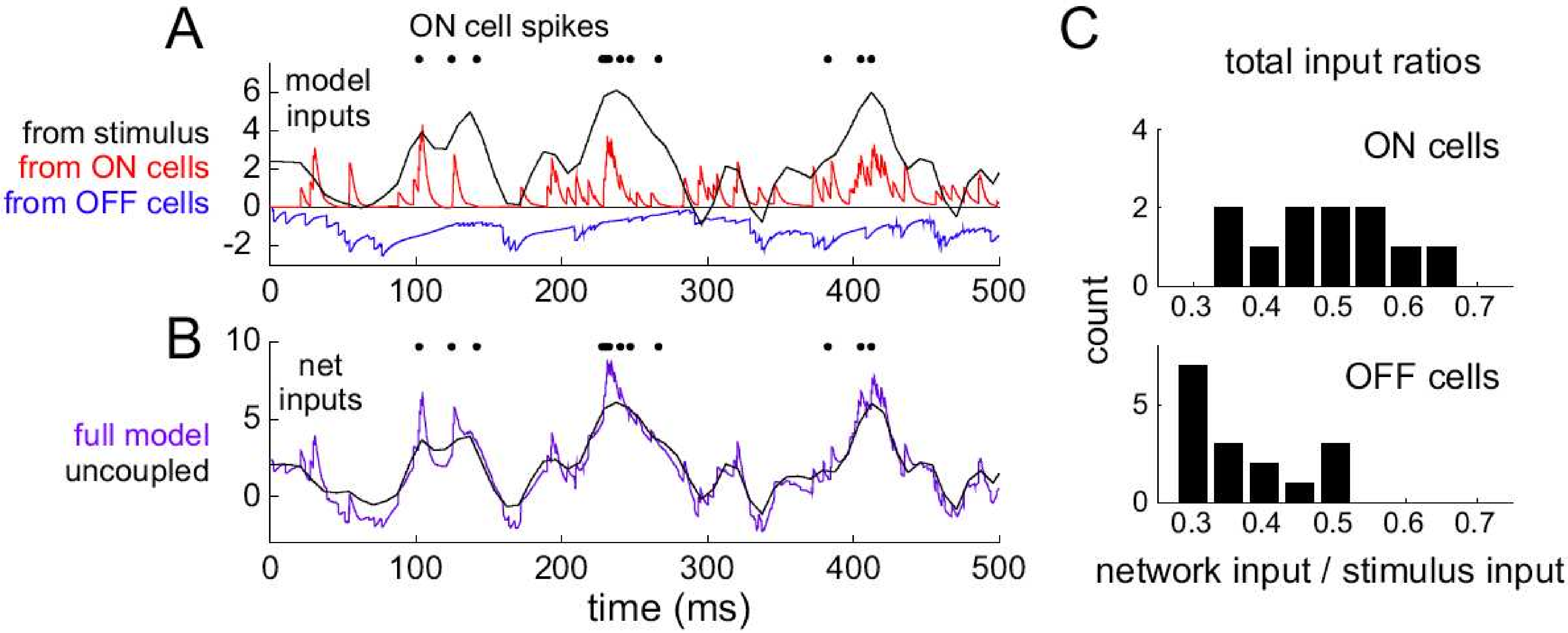
— GLM; fit by L_1-penalized maximum likelihood (concave optimization)

(Paninski, 2004; Truccolo et al., 2005)



ON cells
OFF cells
coupling filters
gain
3
2
1
0
20
40
from ON
time (ms)
1
0.8
0.6
0
20
40
from OFF
time (ms)

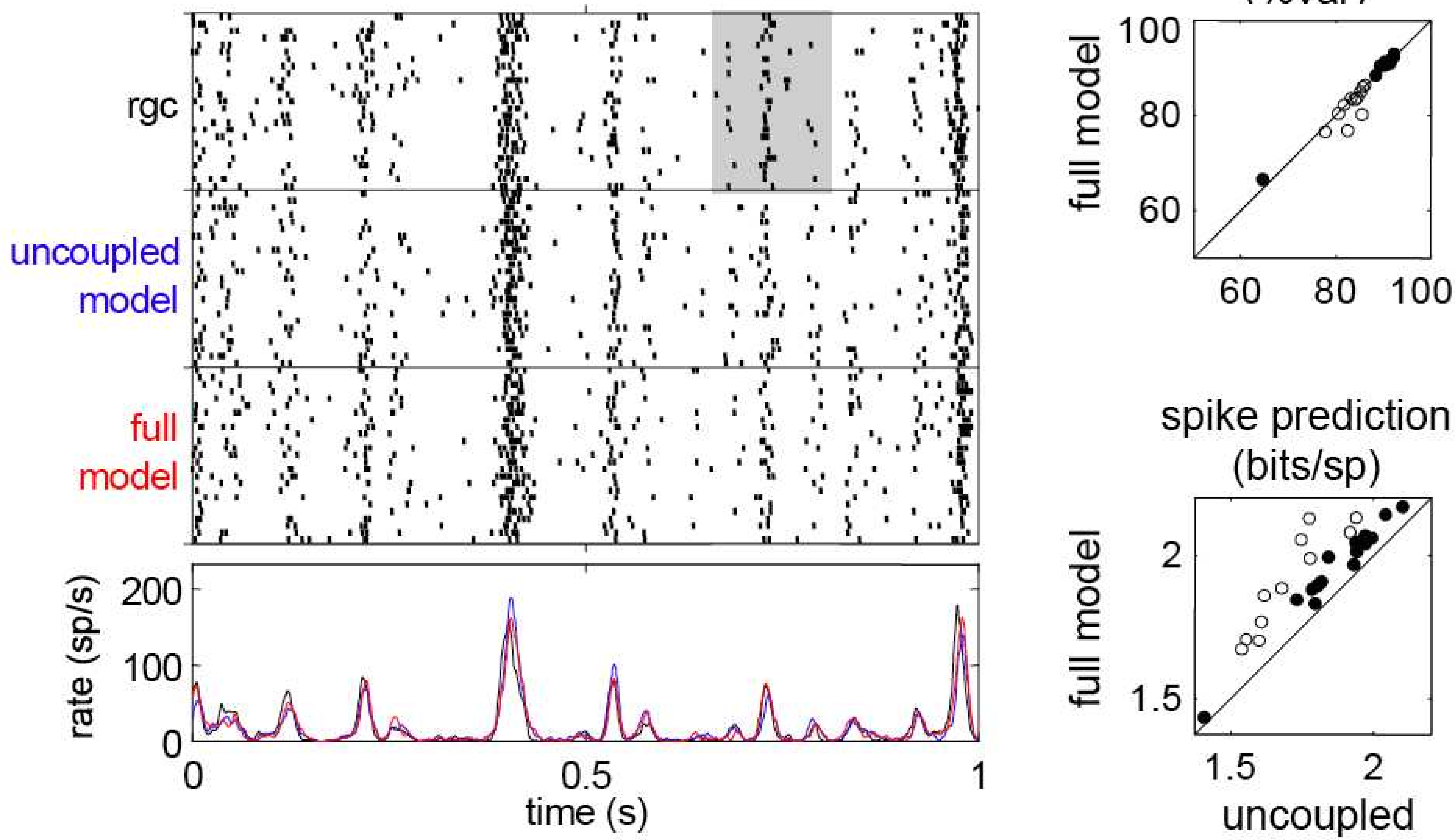
Network vs. stimulus drive



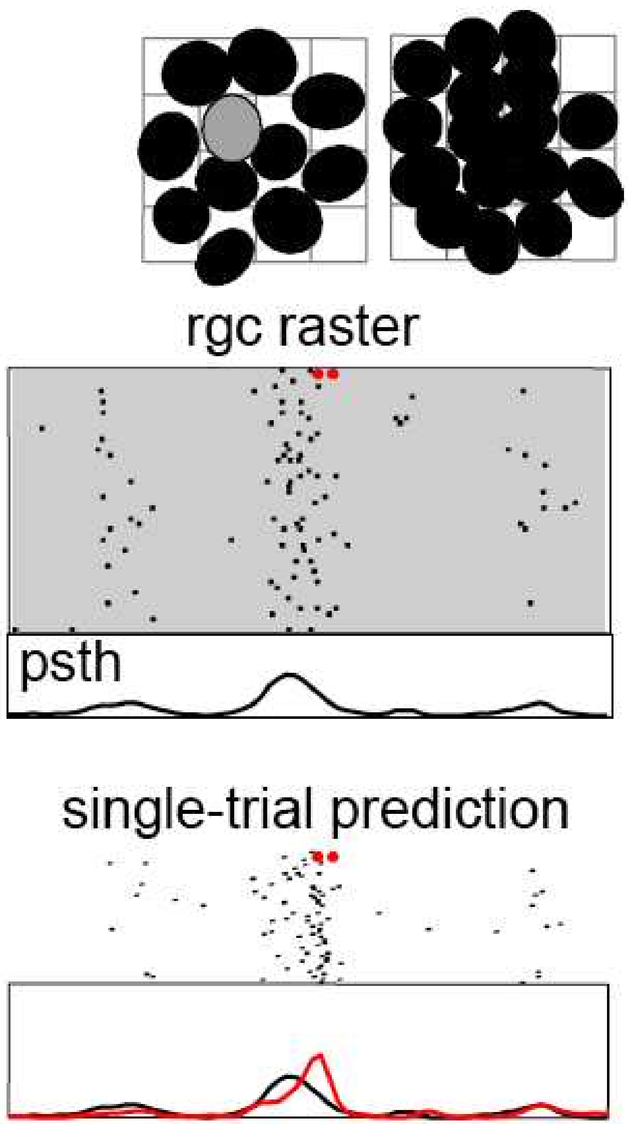
— Network effects are ≈ 50% as strong as stimulus effects

Spike Train Prediction

- improved prediction, but not of mean rate!

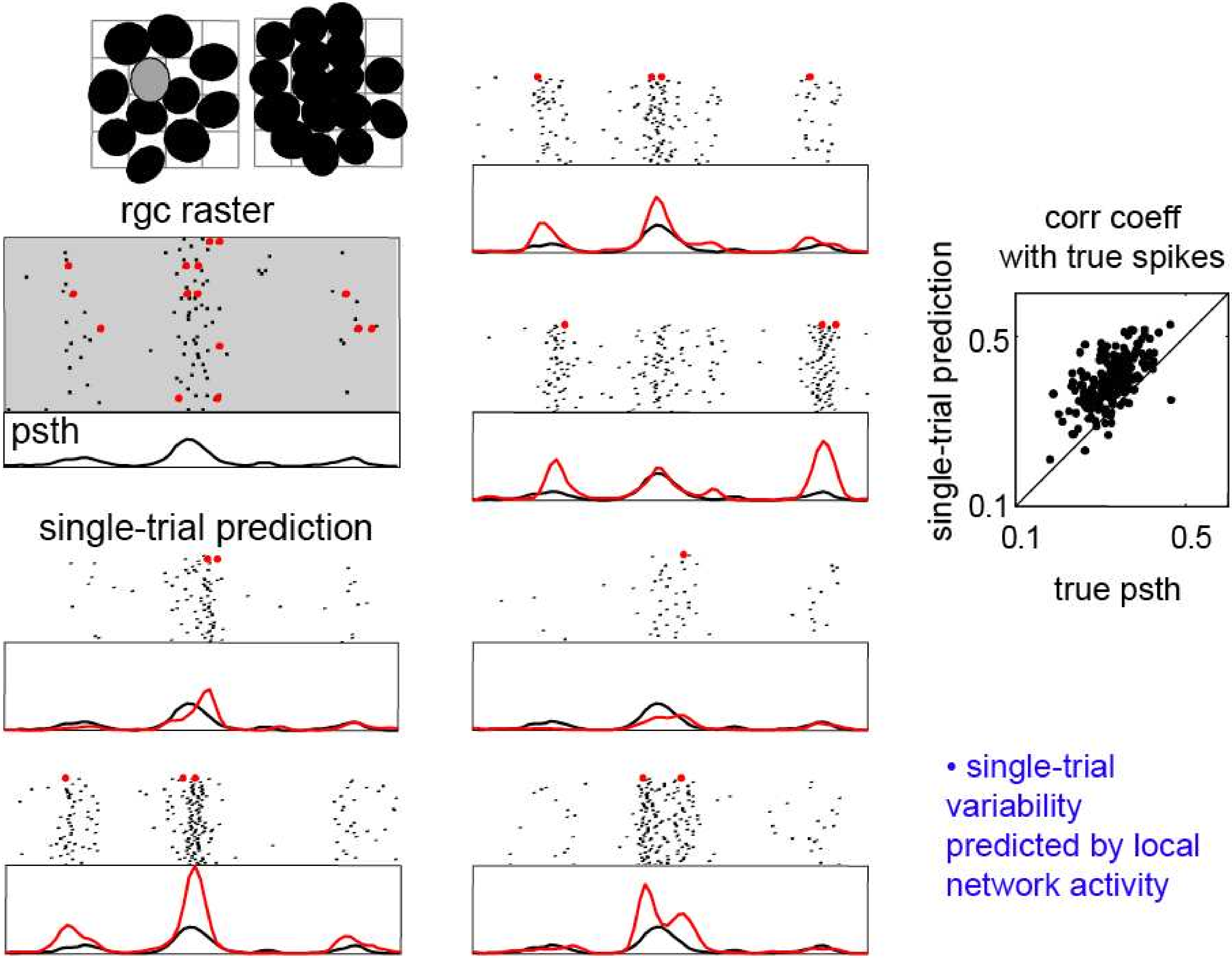


Network predictability analysis



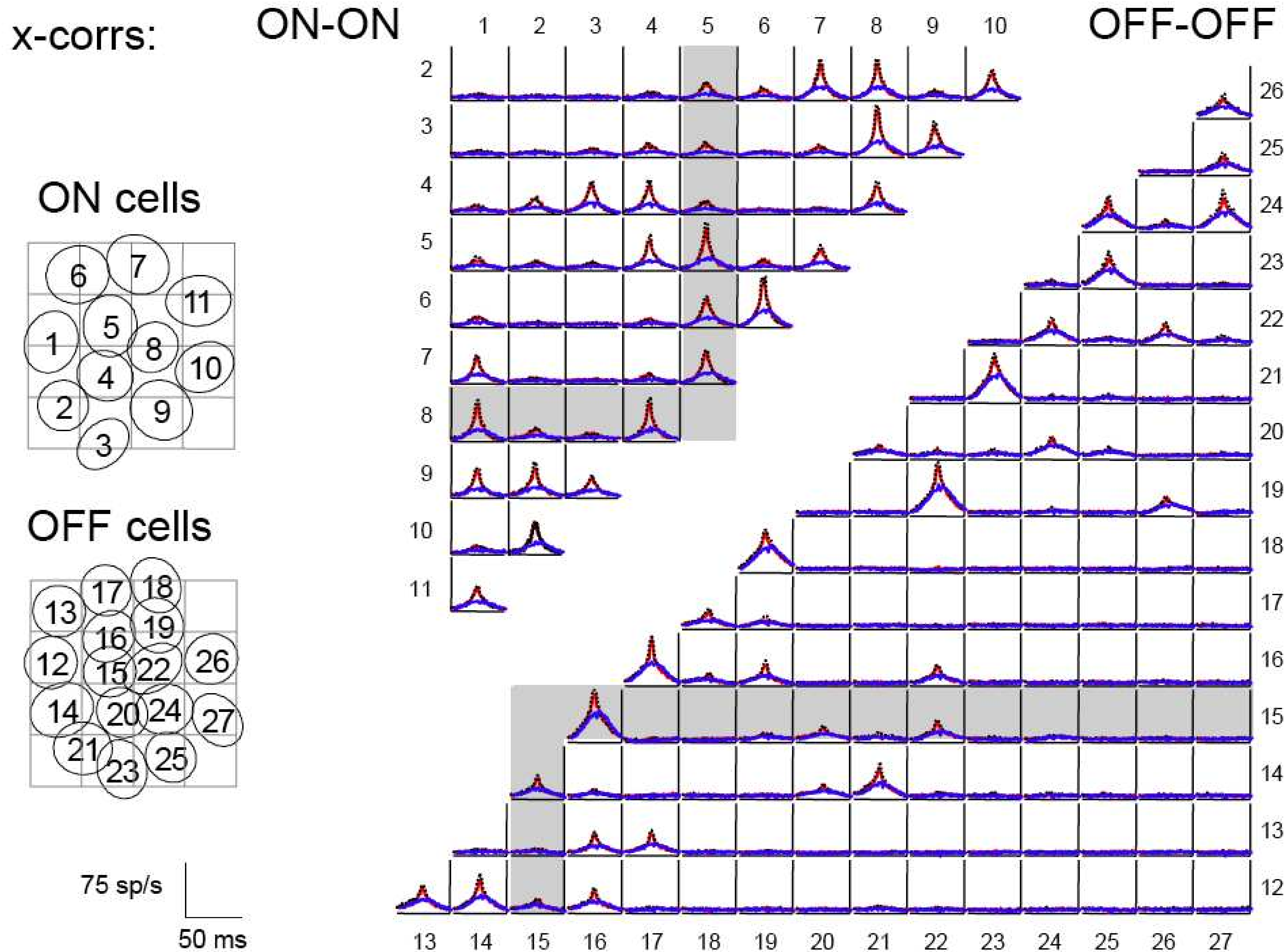
• fix all other neurons for a single trial

• draw single-trial predictions of this cell's spike train

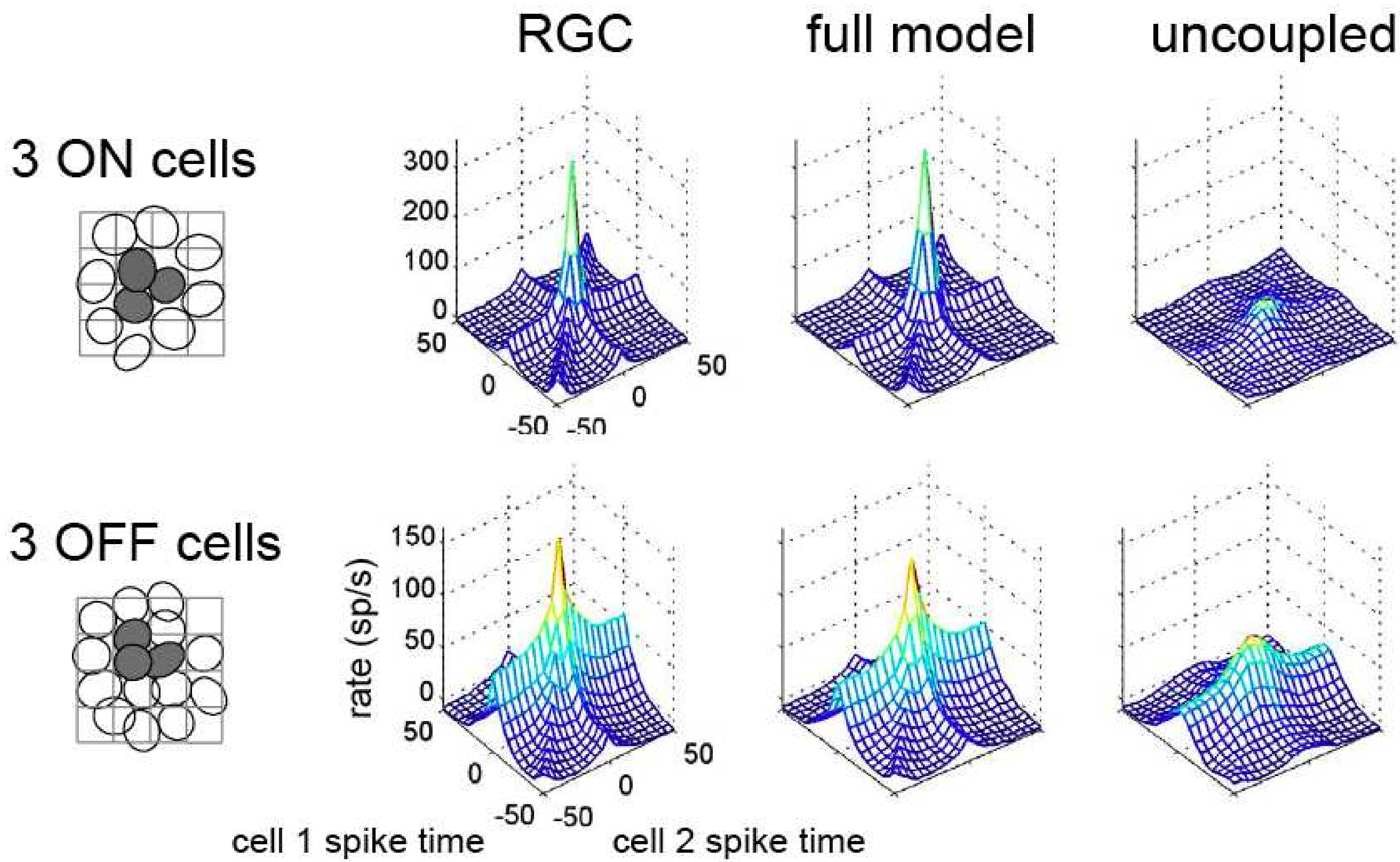


• single-trial variability predicted by local network activity

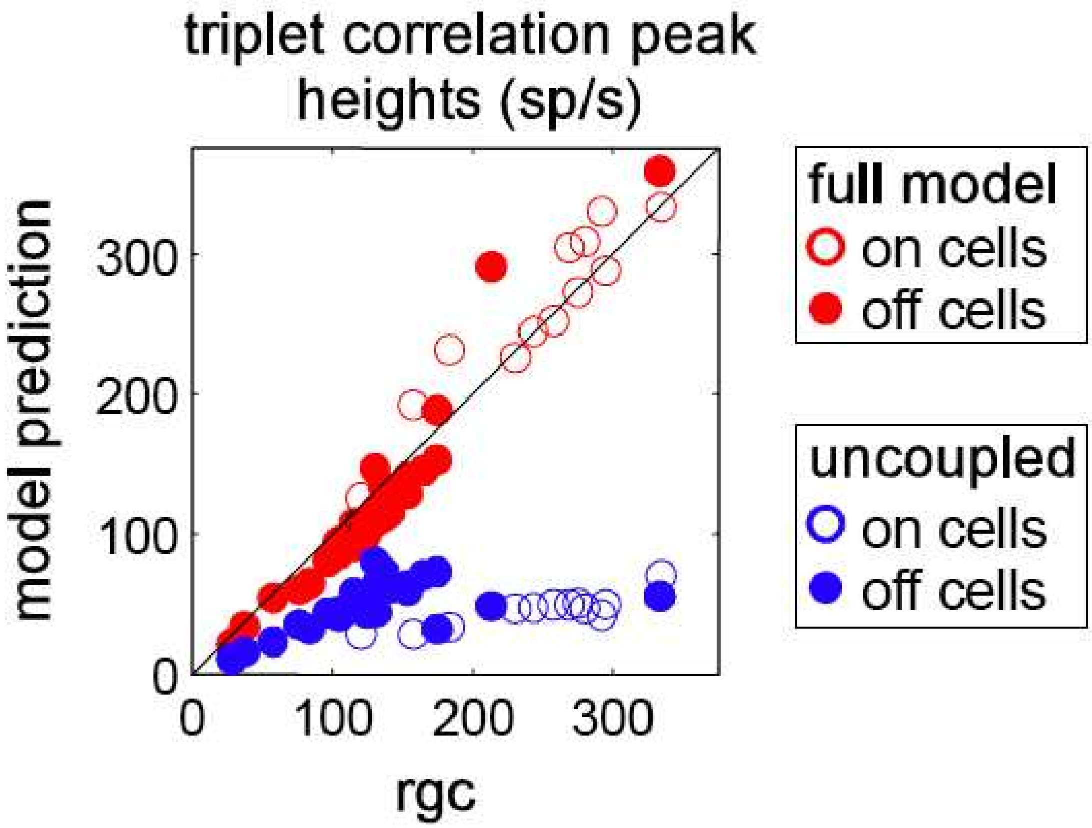
Model captures spatiotemporal cross-corrs



Triplet correlations

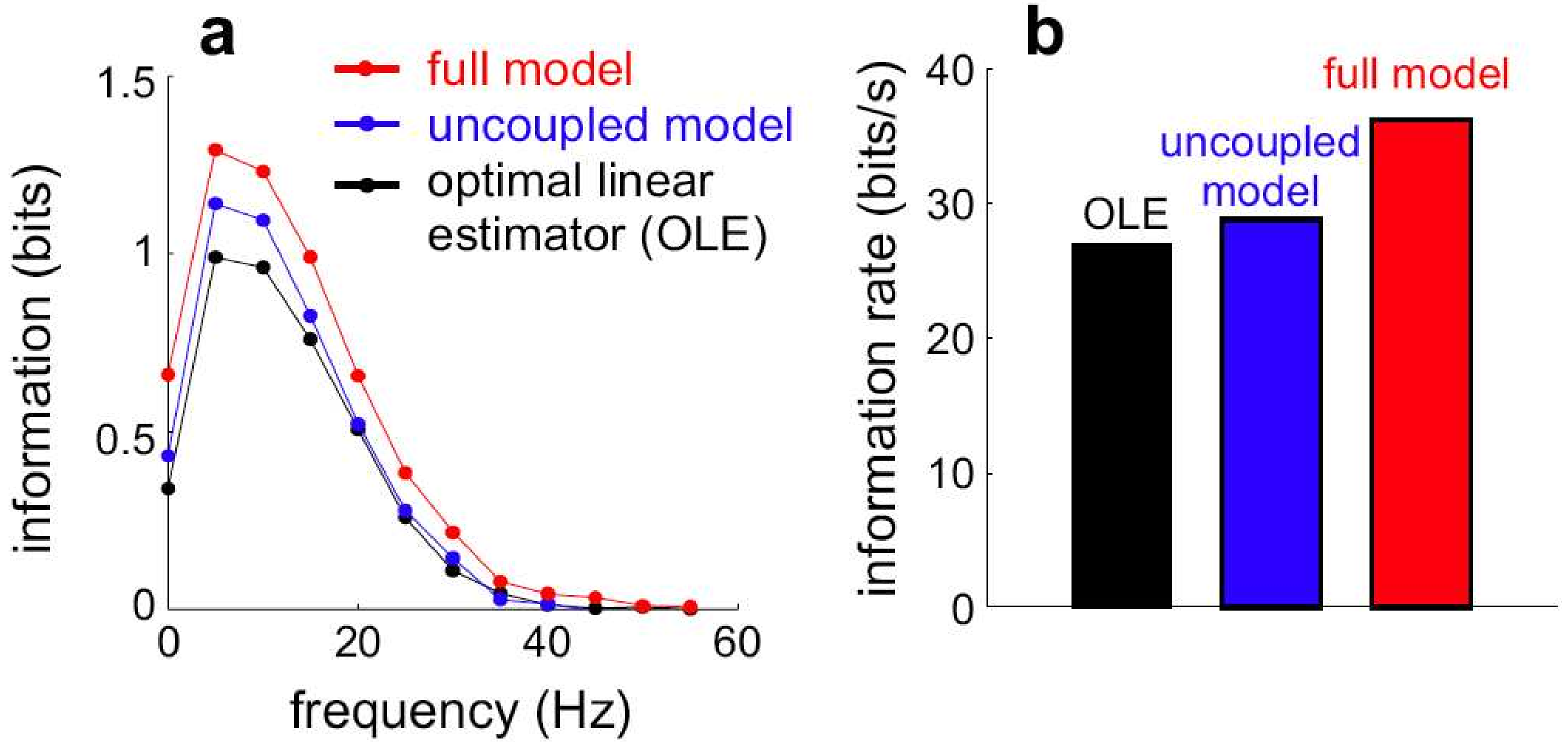


Triplet correlations



Optimal Bayesian decoding

$$E(\vec{x}|spikes) \approx \arg\max_{\vec{x}} \log P(\vec{x}|spikes) = \arg\max_{\vec{x}} [\log P(spikes|\vec{x}) + \log P(\vec{x})]$$

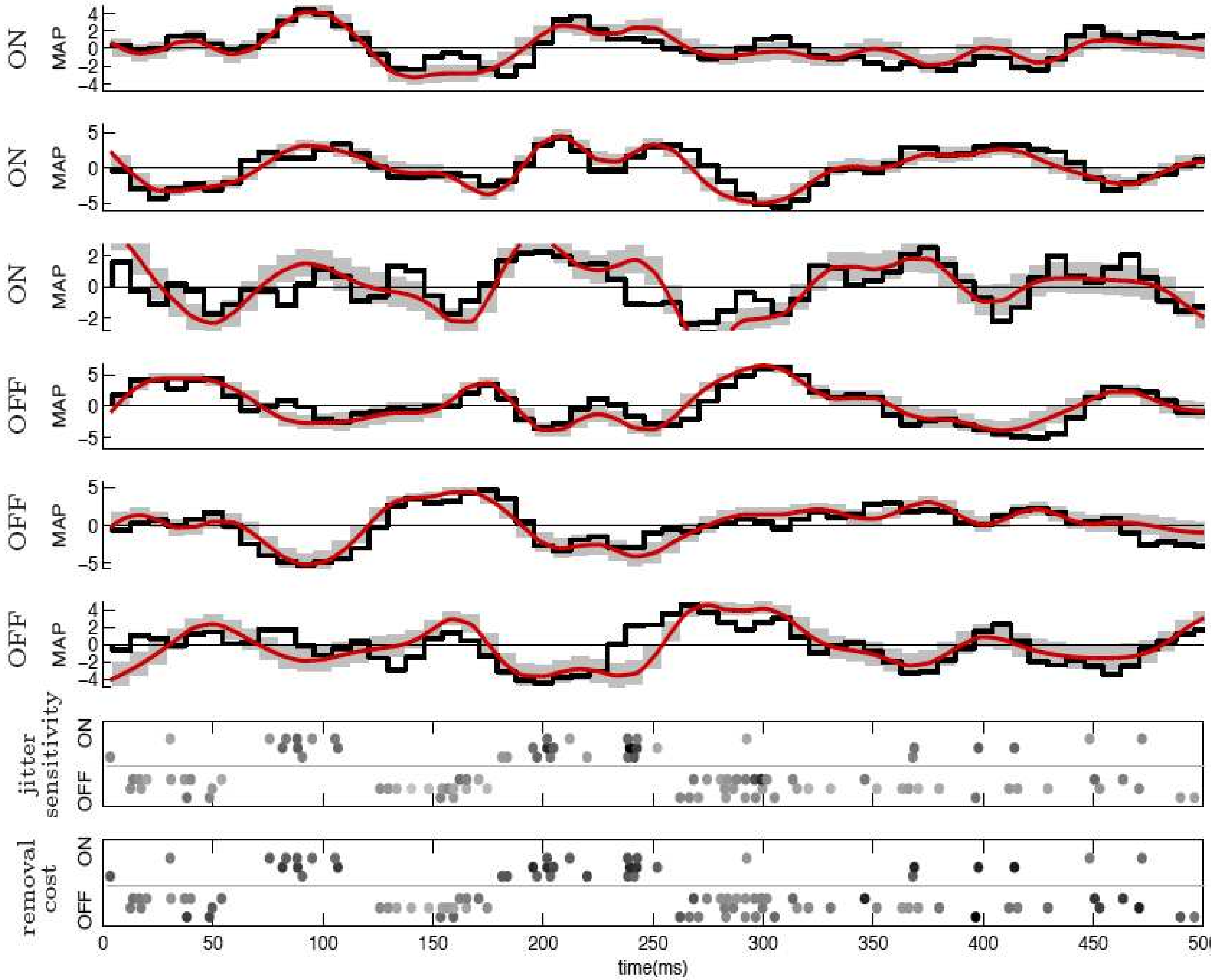


— Computational points:

- $\log P(spikes|\vec{x})$ is concave in $\vec{x}$: concave optimization again.
- Decoding can be done in linear time via standard Newton-Raphson methods, since Hessian of $\log P(\vec{x}|spikes)$ w.r.t. $\vec{x}$ is banded (Pillow et al., 2008a).

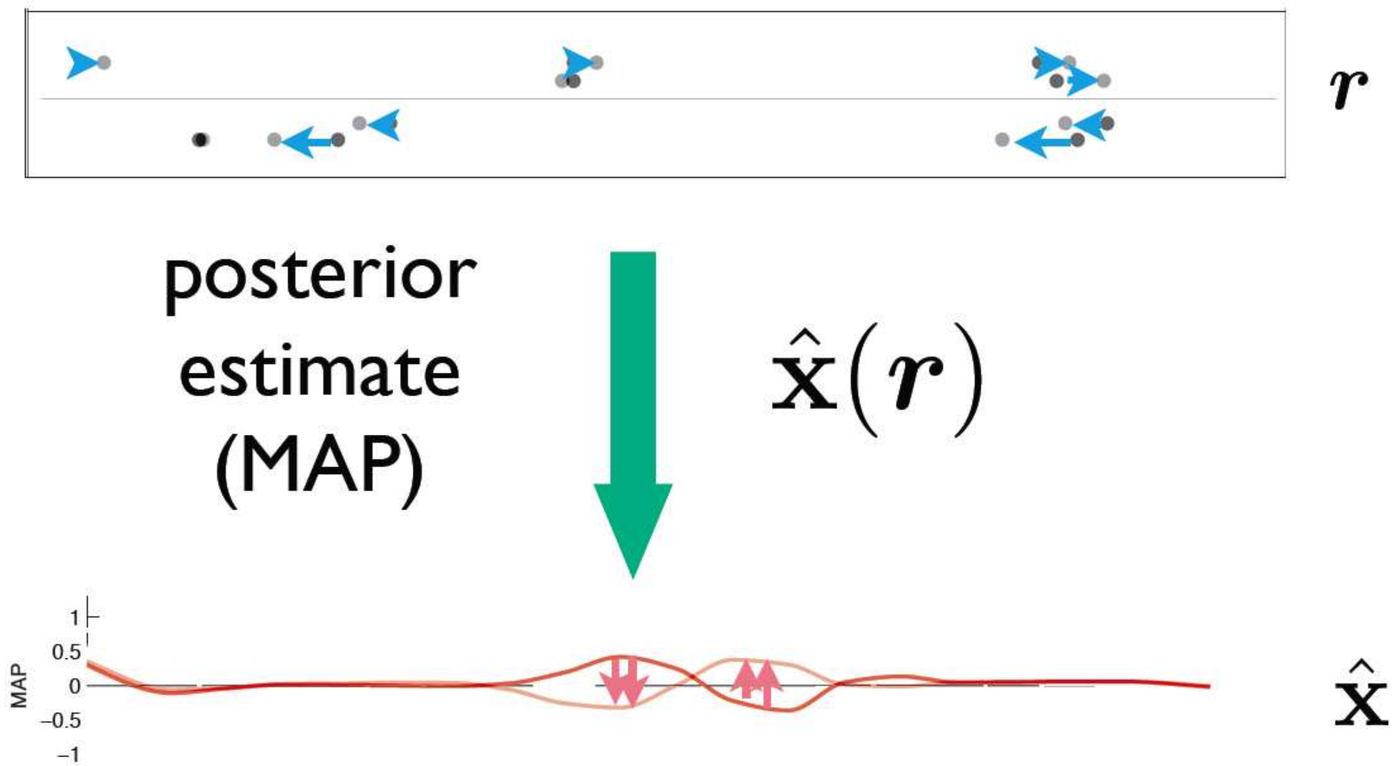
— Biological point: paying attention to correlations improves decoding accuracy.

Application: how important is timing?



— Fast decoding methods let us look more closely (Ahmadian et al., 2008)

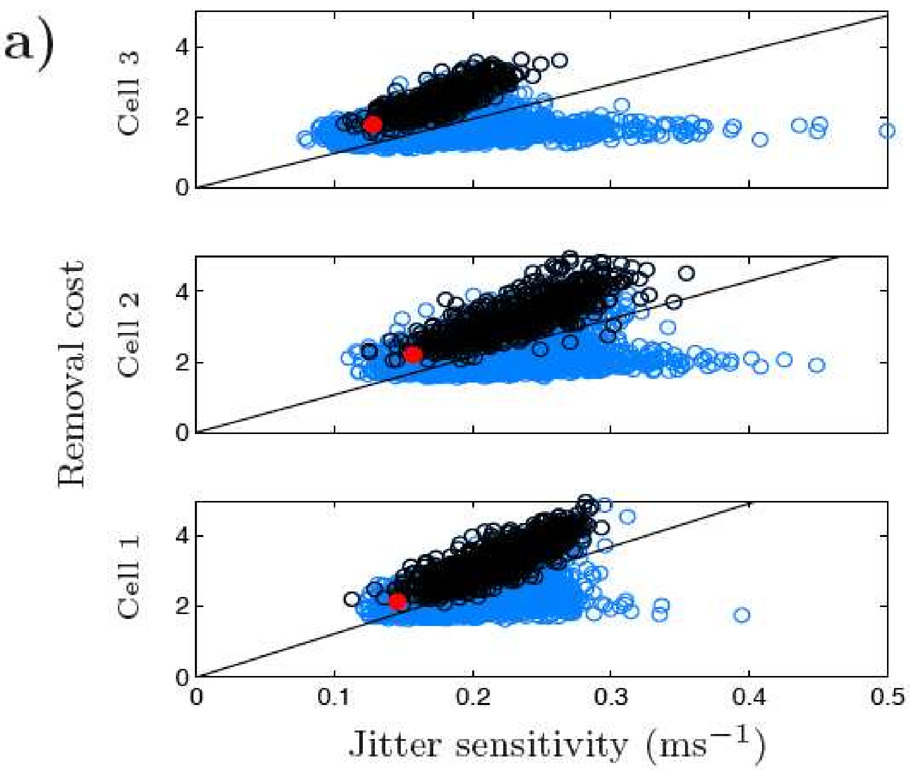
Constructing a metric between spike trains

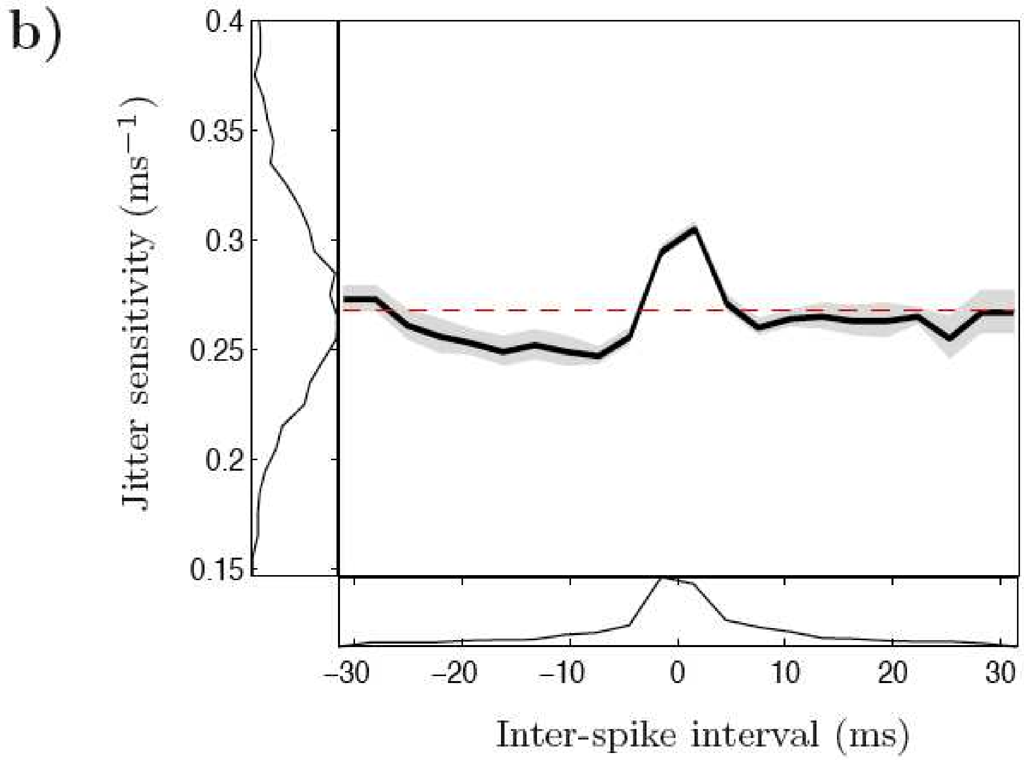


$$d(r_1, r_2) \equiv d_x\left(\hat{x}(r_1), \hat{x}(r_2)\right)$$

Locally, $d(r, r + \delta r) = \delta r^T G_r \delta r$: interesting information in G_r.

Spike sensitivity is strongly context-dependent

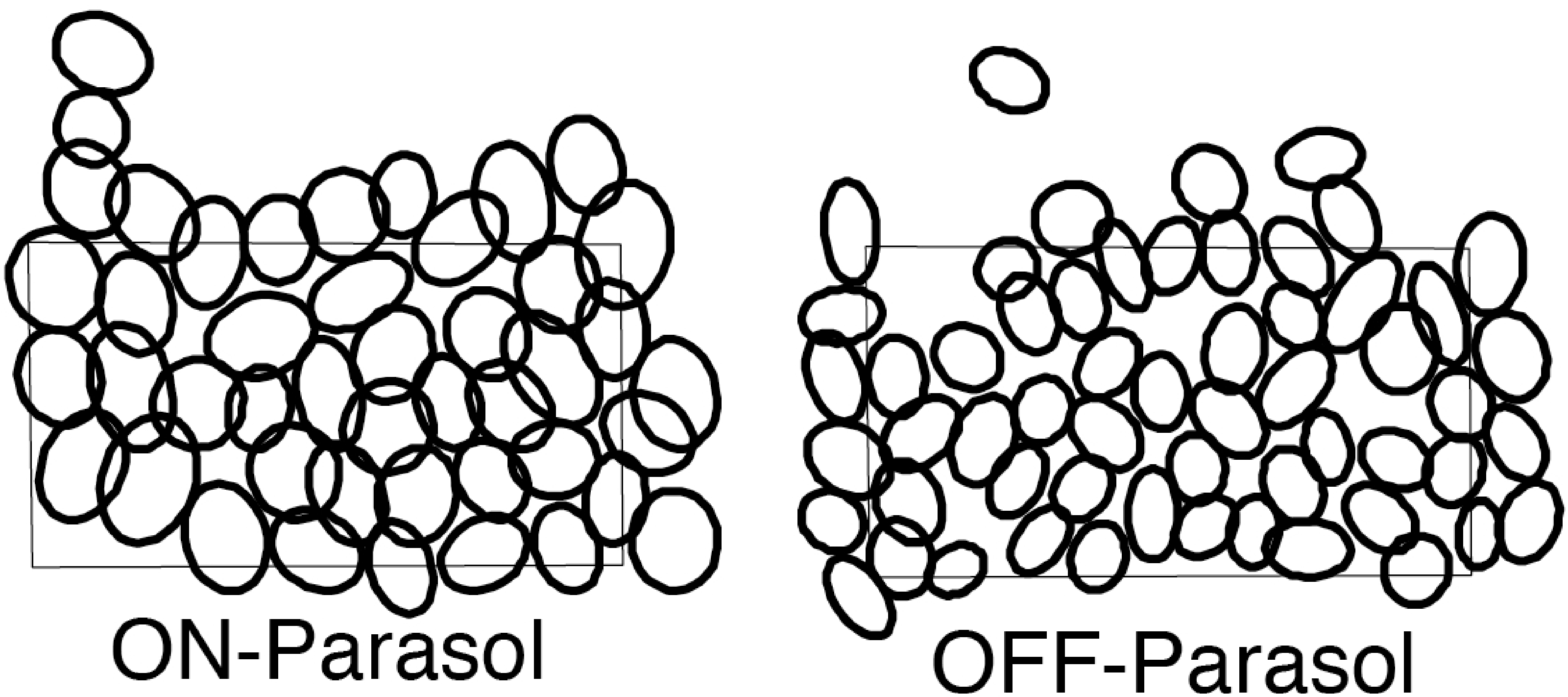




— Reflects nonlinearity of decoder $\hat{x}(r)$: linear decoder is context-independent

— Cost of spike addition/deletion $\approx$ cost of jittering by 10 ms (Victor, 2000): natural time scale of spike train.

Application: recurrent network modeling



— Do observed local connectivity rules lead to interesting network dynamics? What are the implications for retinal information processing? Can we capture these effects with a reduced dynamical model?

— Mean-field analysis (Toyoizumi et al., 2008)

Application: optimal velocity decoding

How to decode behaviorally-relevant signals, e.g. image velocity?

If image I is known, use Bayesian estimate (Weiss et al., 2002):

$$p(v|spikes, I) \propto p(v)p(spikes|v, I)$$

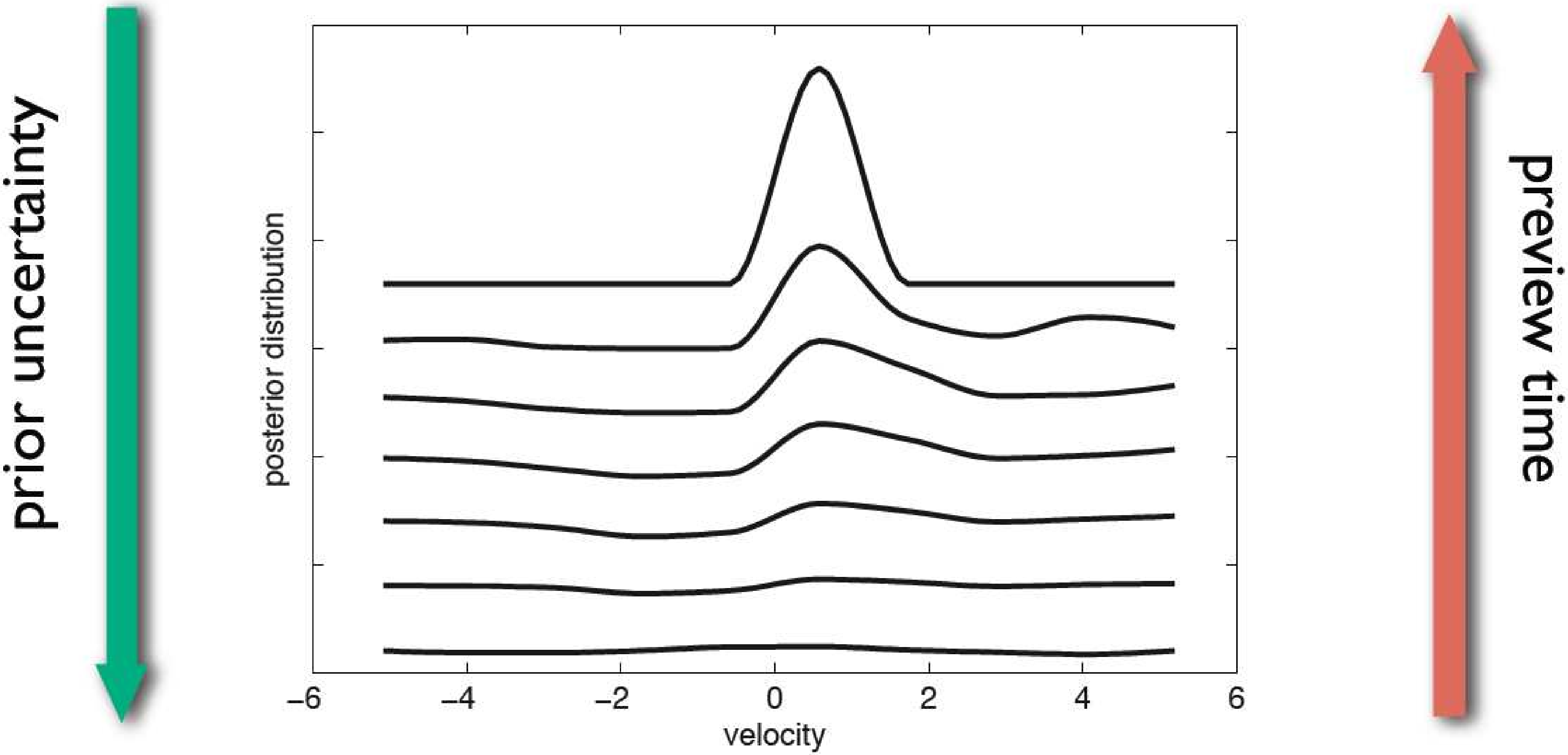
If image is unknown, we have to integrate out:

$$p(v|spikes) \propto p(v)p(spikes|v) = p(v) \int p(I)p(spikes|v, I)dI;$$

$p(I)$ denotes *a priori* image distribution.

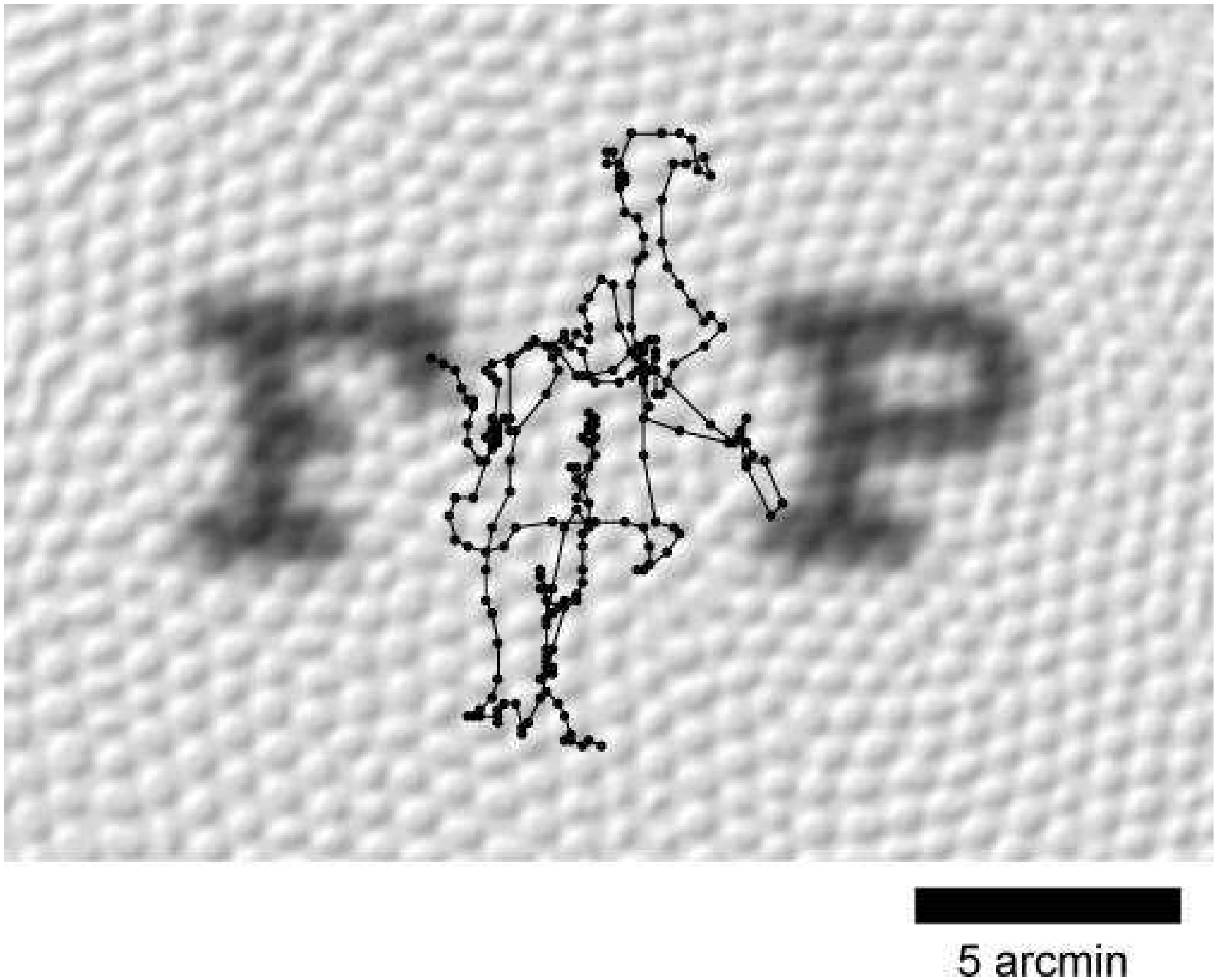
— connections to standard energy models
(Frechette et al., 2005; Lalor et al., 2008)

Optimal velocity decoding



— estimation improves with knowledge of image; can compare directly to human psychophysics (Frechette et al., 2004)

Application: image stabilization



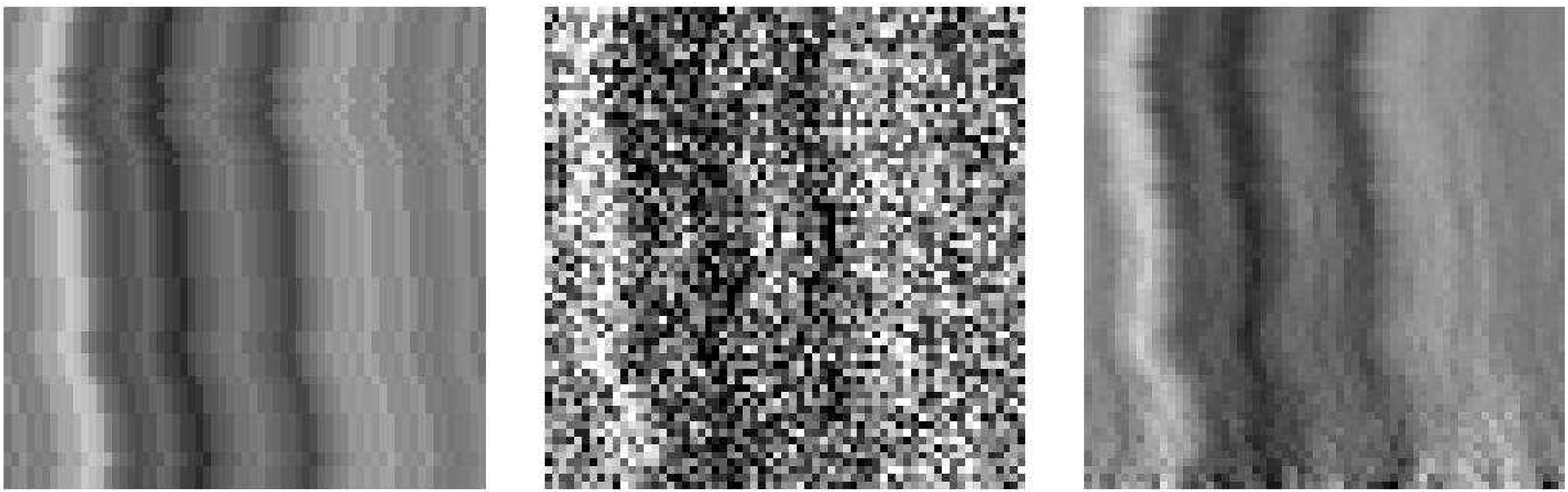
From (Pitkow et al., 2007): neighboring letters on the 20/20 line of the Snellen eye chart. Trace shows 500 ms of eye movement.

Bayesian methods for image stabilization

Similar marginalization idea as in velocity estimation:

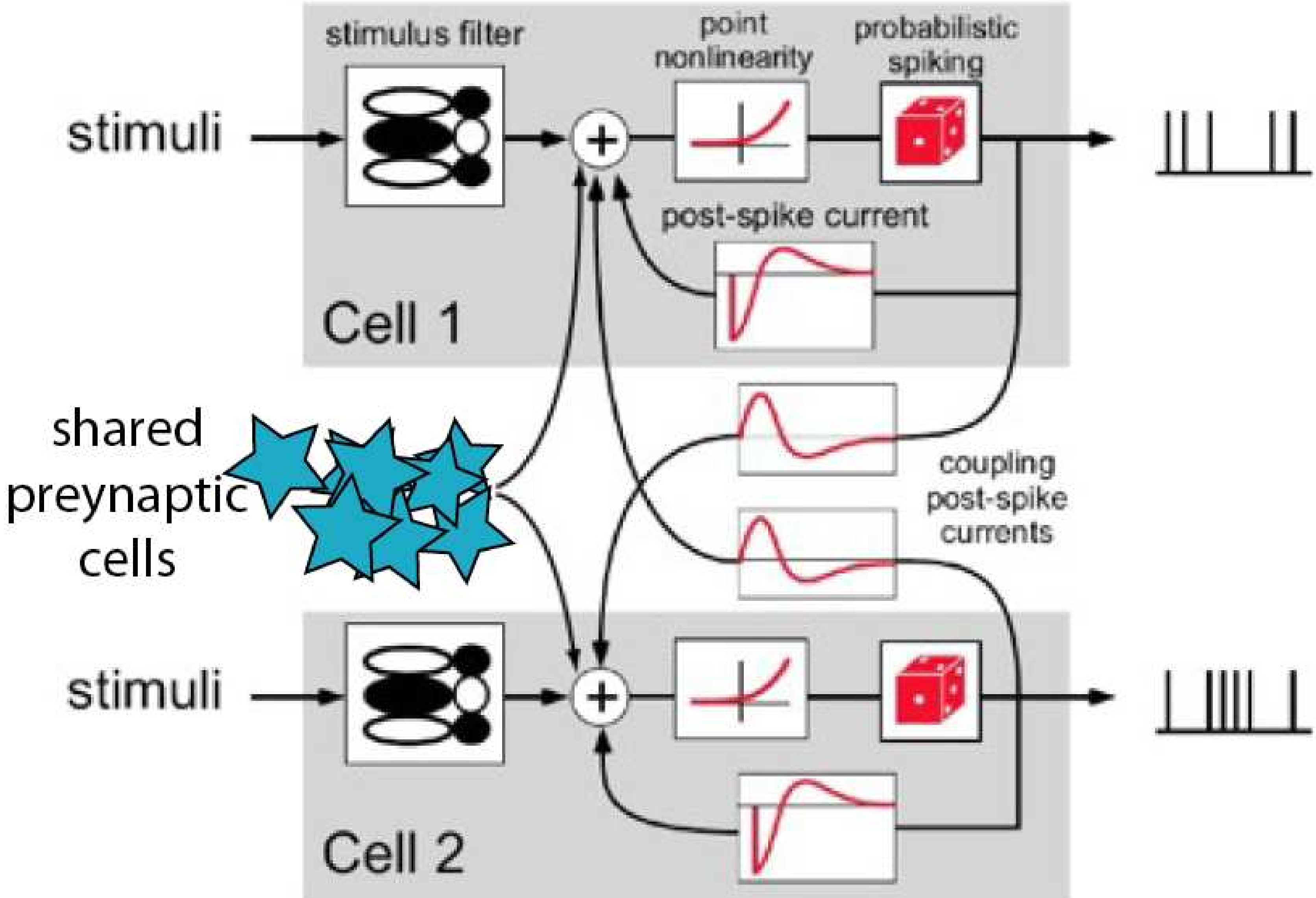
$$p(I|spikes) \propto p(I)p(spikes|I) = p(I) \int p(spikes|e, I)p(e)de;$$

e denotes eye jitter path; integration by particle-filter methods.



true image w/ translations; observed noisy retinal responses; estimated image.

Extension: including common input effects



State-space setting (Kulkarni and Paninski, 2007; Khuc-Trong and Rieke, 2008; Wu et al., 2008)

Direct state-space optimization methods

$$\begin{aligned}\lambda_i(t) &= f\left[b_i + \vec{k}_i \cdot \vec{x}(t) + \sum_{i',j} h_{i',j} n_{i'}(t-j) + q_i(t)\right] \\ &= f\left[X_t \theta + q_i(t)\right]\end{aligned}$$

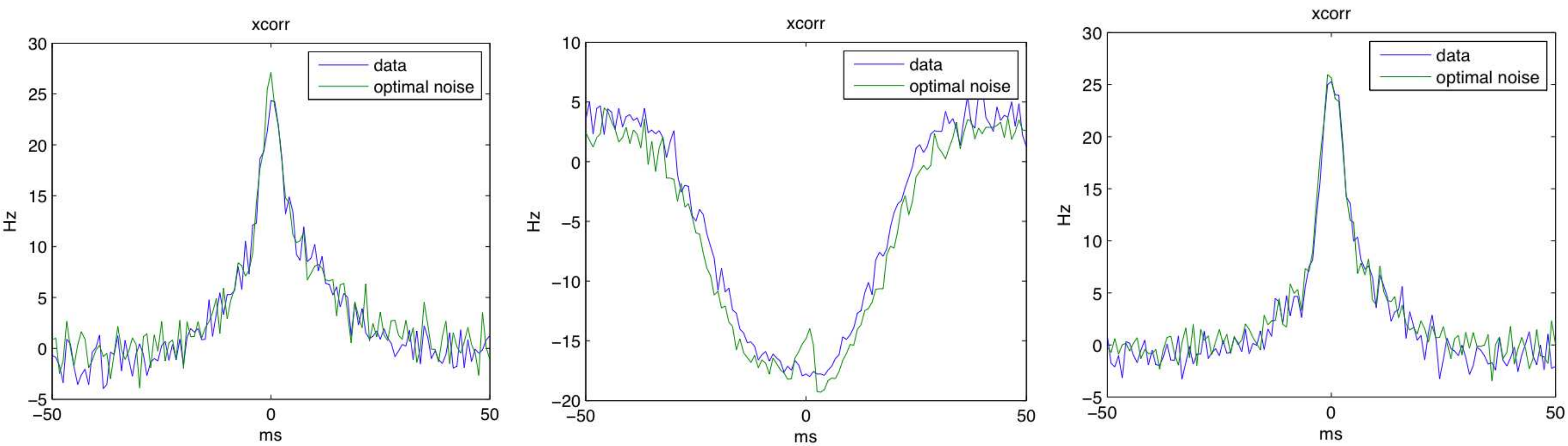
— Q is a very high-dimensional latent (unobserved) "common input" term. Taken to be a Gaussian process here with autocorrelation time ≈ 5 ms

— Parameter θ is high-d; standard Expectation-Maximization approach is very slow. Instead, optimize Laplace-approximated marginal likelihood directly:

$$\begin{aligned}\log p(spikes|\theta) &= \log \int p(Q|\theta) p(spikes|\theta, Q) dQ \\ &\approx \log p(\hat{Q}_\theta|\theta) + \log p(spikes|\hat{Q}_\theta) - \frac{1}{2}\log|J_{\hat{Q}_\theta}| \\ \hat{Q}_\theta &= \arg\max_Q \{\log p(Q|\theta) + \log p(spikes|Q)\}\end{aligned}$$

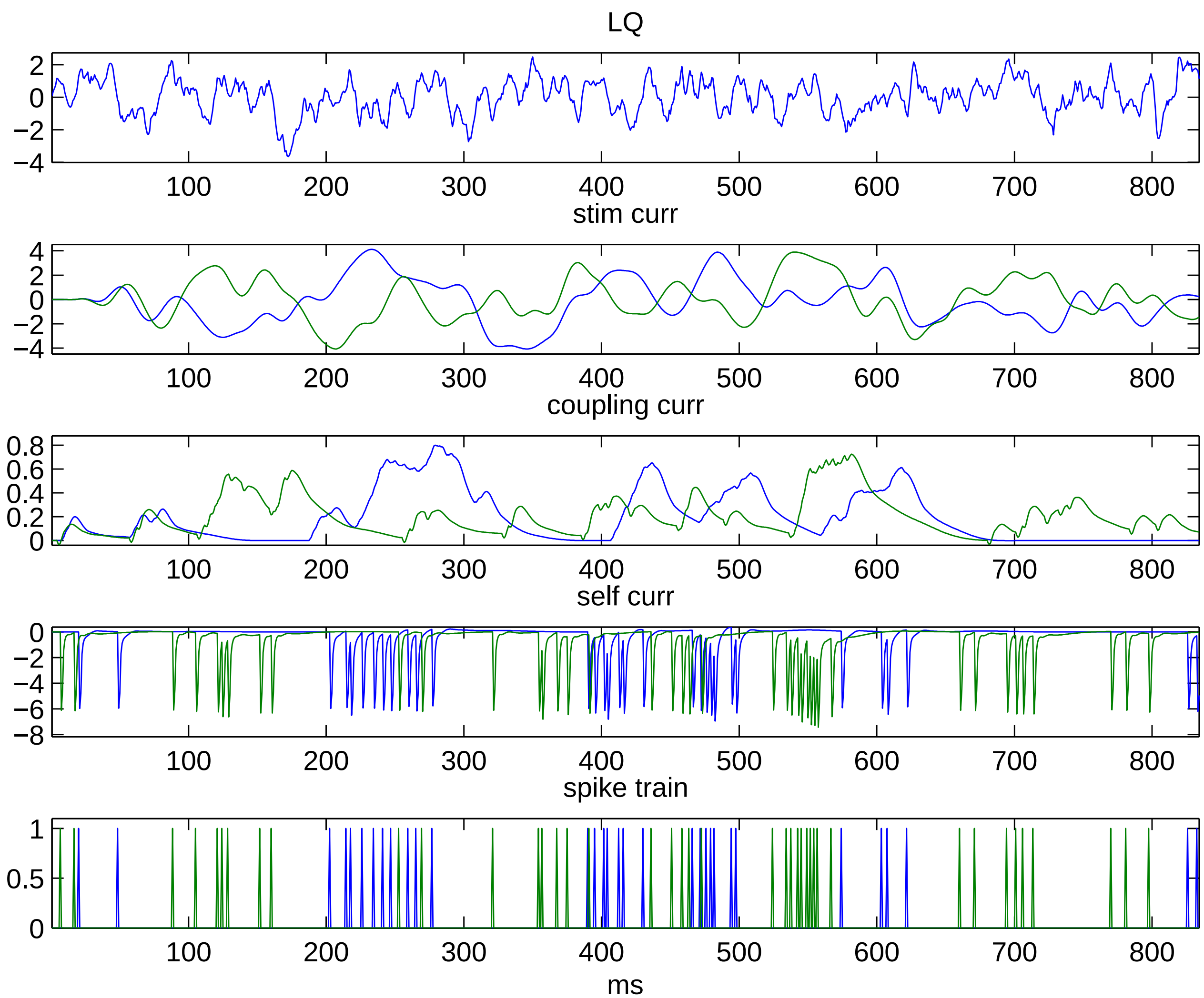
— all terms can be computed in linear time via block-tridiagonal matrix methods (Koyama et al., 2008). Number of applications (Paninski et al., 2008).

Common input model predicts x-corrs well



(analysis of full population is in progress...)

Inferred common input effects are strong



— Much more consistent with biophysical data (Khuc-Trong and Rieke, 2008).

— Next steps: what is impact on statistical properties of the model? Can inferred common inputs be mapped directly onto biophysical currents?

Conclusions

- Standard statistical models (GLM) provide flexible, powerful tools for answering key questions in neuroscience
- Close relationships between encoding, decoding, and experimental design (Paninski et al., 2007)
- Log-concavity and suitable matrix structure makes computations very tractable
- Many opportunities for machine learning / fast computational techniques in neuroscience

References

Ahmadian, Y., Pillow, J., and Paninski, L. (2008). Efficient Markov Chain Monte Carlo methods for decoding population spike trains. *Under review, Neural Computation*.

Frechette, E., Sher, A., Grivich, M., Petrusca, D., Litke, A., and Chichilnisky, E. (2005). Fidelity of the ensemble code for visual motion in the primate retina. *J Neurophysiol*, 94(1):119–135.

Frechette, E. S., Grivich, M. I., Kalmar, R. S., Litke, A. M., Petrusca, D., Sher, A., and Chichilnisky, E. J. (2004). Retinal motion signals and limits on speed discrimination. *J. Vis.*, 4(8):570.

Khuc-Trong, P. and Rieke, F. (2008). Origin of correlated activity between parasol retinal ganglion cells. *In press*.

Koyama, S., Kass, R., and Paninski, L. (2008). Efficient computation of the most likely path in integrate-andfire and more general state-space models. *COSYNE*.

Kulkarni, J. and Paninski, L. (2007). Common-input models for multiple neural spike-train data. *Network: Computation in Neural Systems*, 18:375–407.

Lalor, E., Ahmadian, Y., and Paninski, L. (2008). Optimal decoding of stimulus velocity using a probabilistic model of ganglion cell populations in primate retina. *Journal of Vision*, Under review.

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

Paninski, L., Kass, R., Eden, U., and Brown, E. (2008). *Statistical analysis of neurophysiological data*. Book under review.

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.

Pillow, J., Ahmadian, Y., and Paninski, L. (2008a). Model-based decoding, information estimation, and change-point detection in multi-neuron spike trains. *Under review, Neural Computation*.

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

Pitkow, X., Sompolinsky, H., and Meister, M. (2007). A neural computation for visual acuity in the presence of eye movements. *PLOS Biology*, 5.

Toyoizumi, T., Rahnama Rad, K., and Paninski, L. (2008). Mean-field approximations for coupled populations of generalized linear model spiking neurons with Markov refractoriness. *Neural Computation*, In press.

Truccolo, W., Eden, U., Fellows, M., Donoghue, J., and Brown, E. (2005). A point process framework for