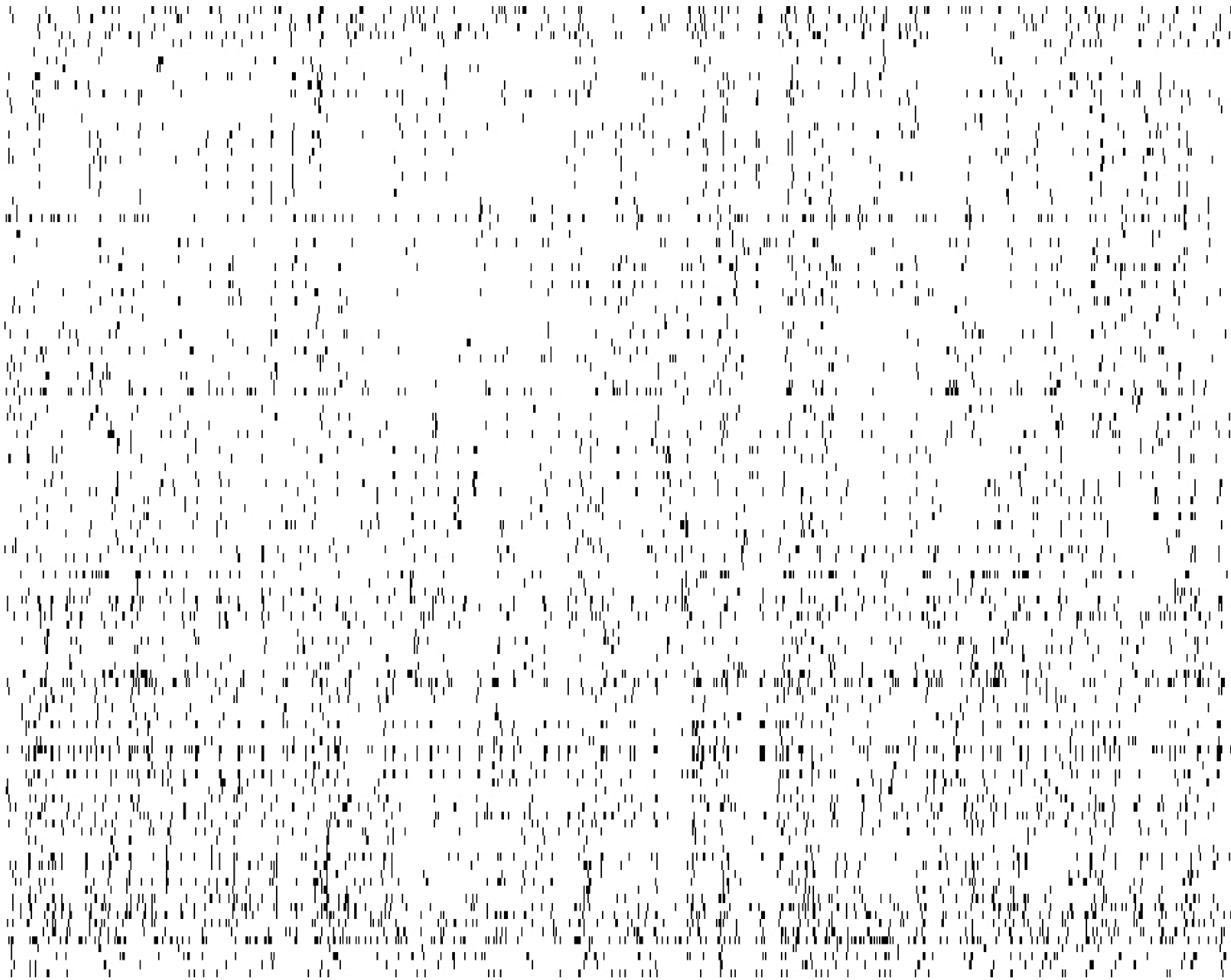
Lecture 11: *Models of neural population activity*



with thanks to Elad Schneidman, Weizmann Institute

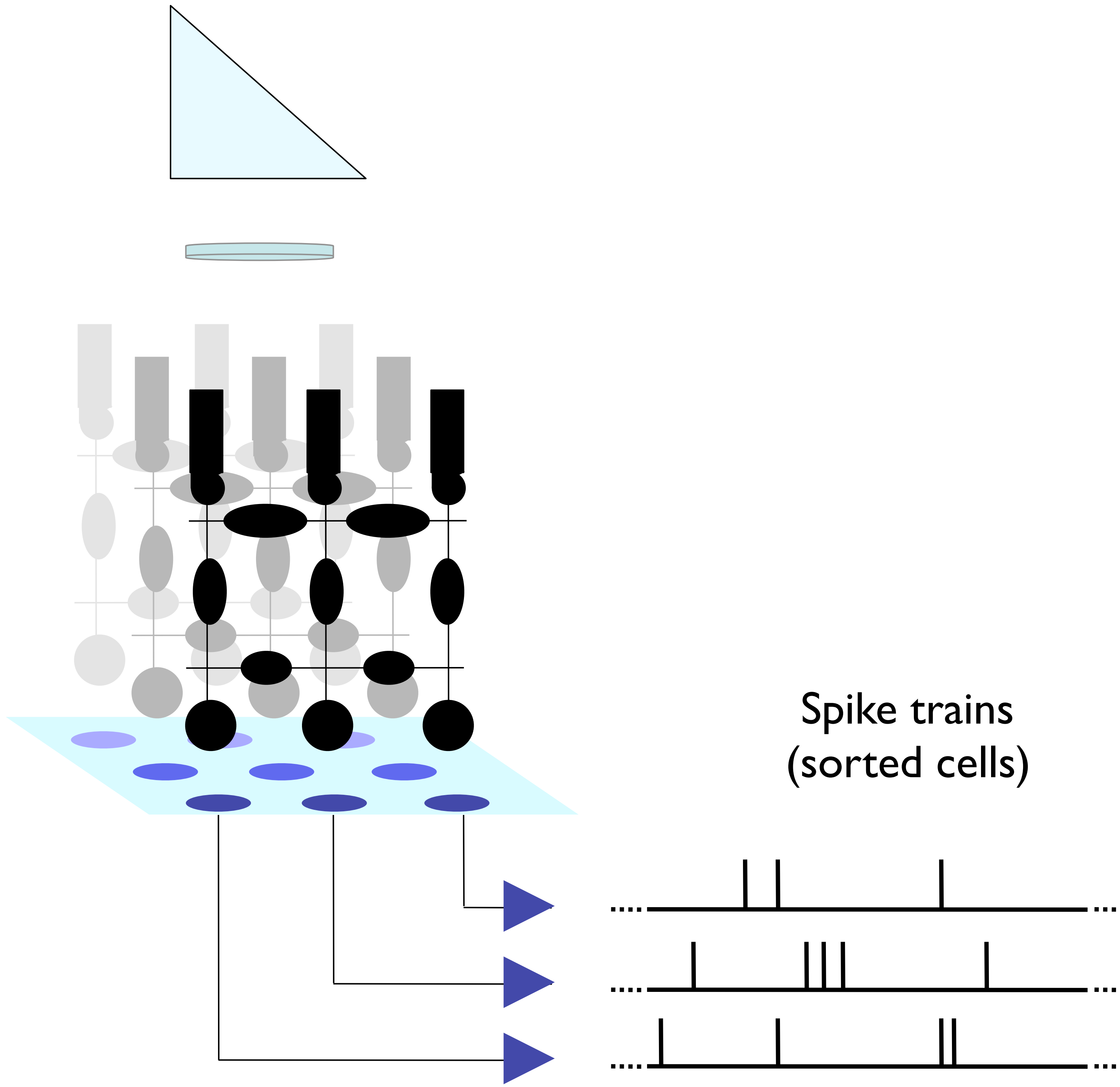
Understanding the neural population code:

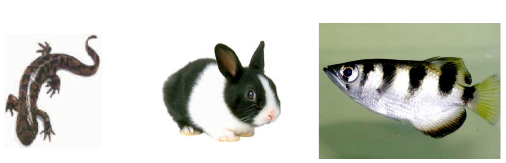


\# of patterns >1,267,650,600,228,229,000,000,000,000,000

Recording from a large neural population:

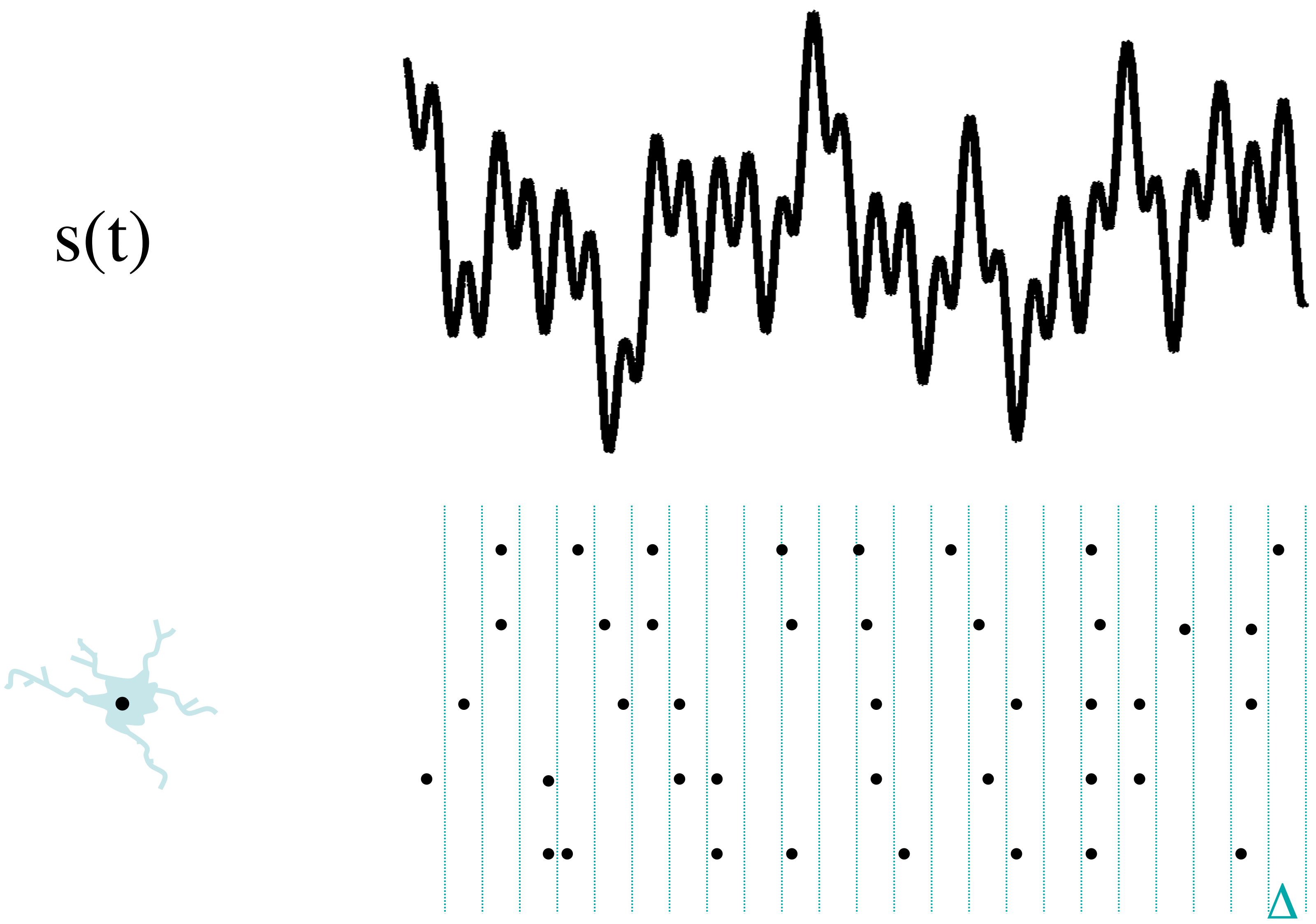
s(t)



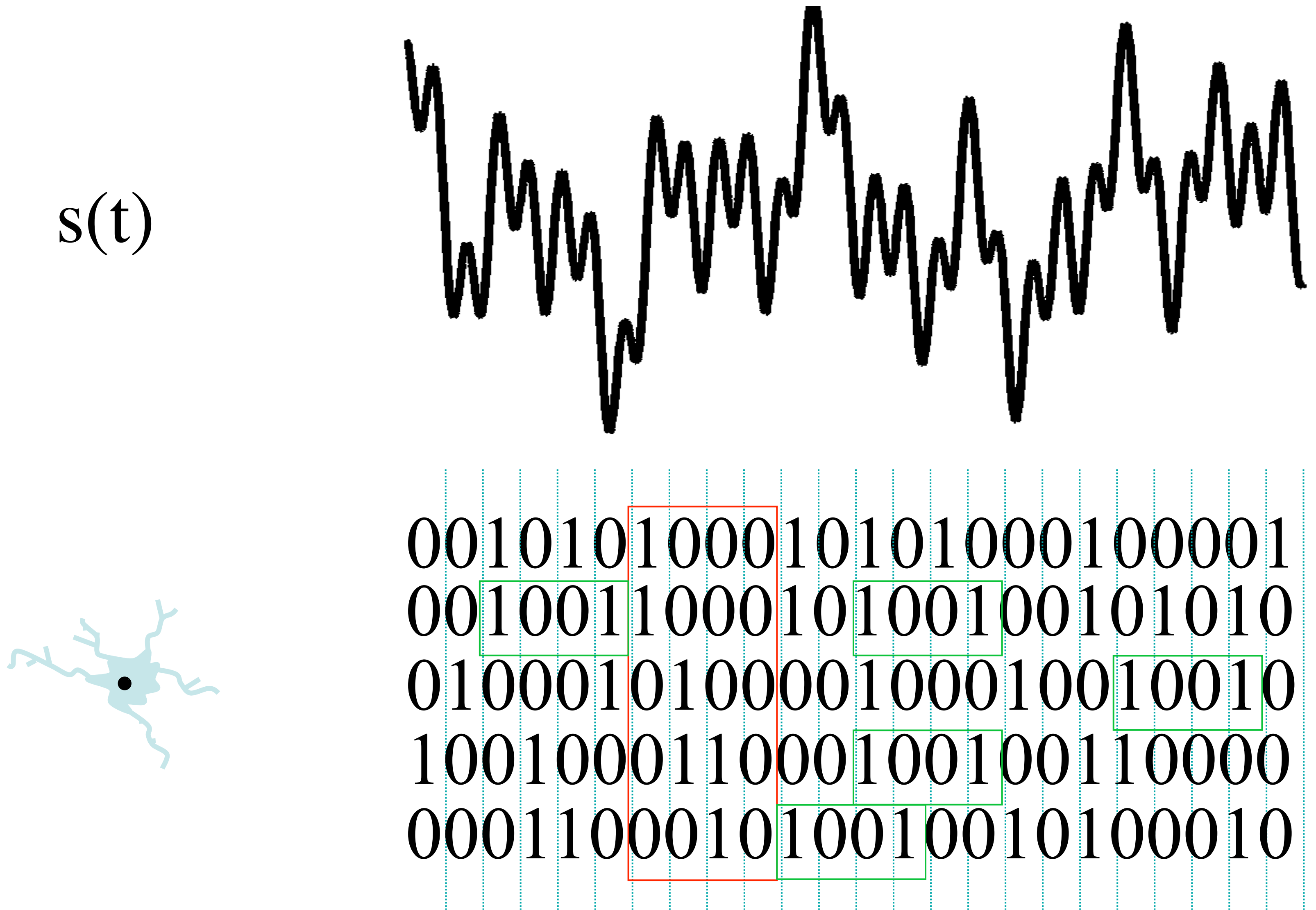


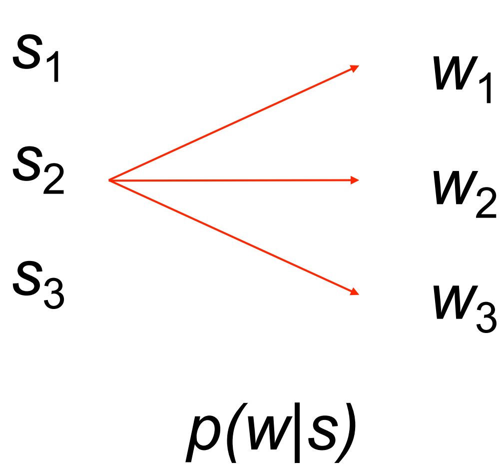
Berry Lab, Princeton & Segev Lab, Ben-Gurion University

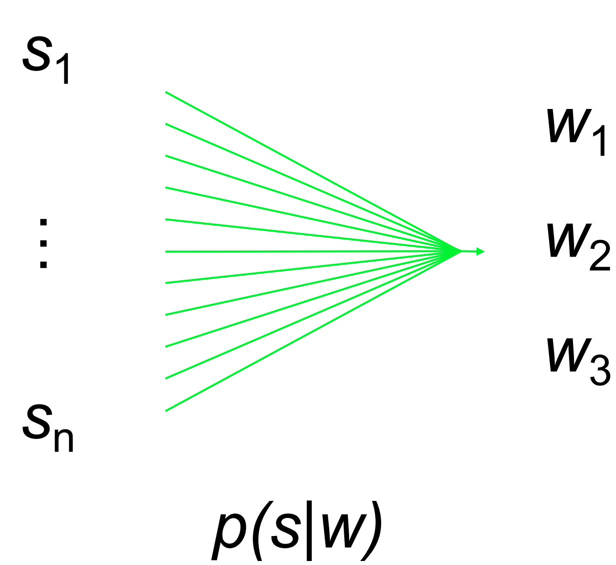
Defining spiking patterns across many cells:



The concept of a neural ‘dictionary’:



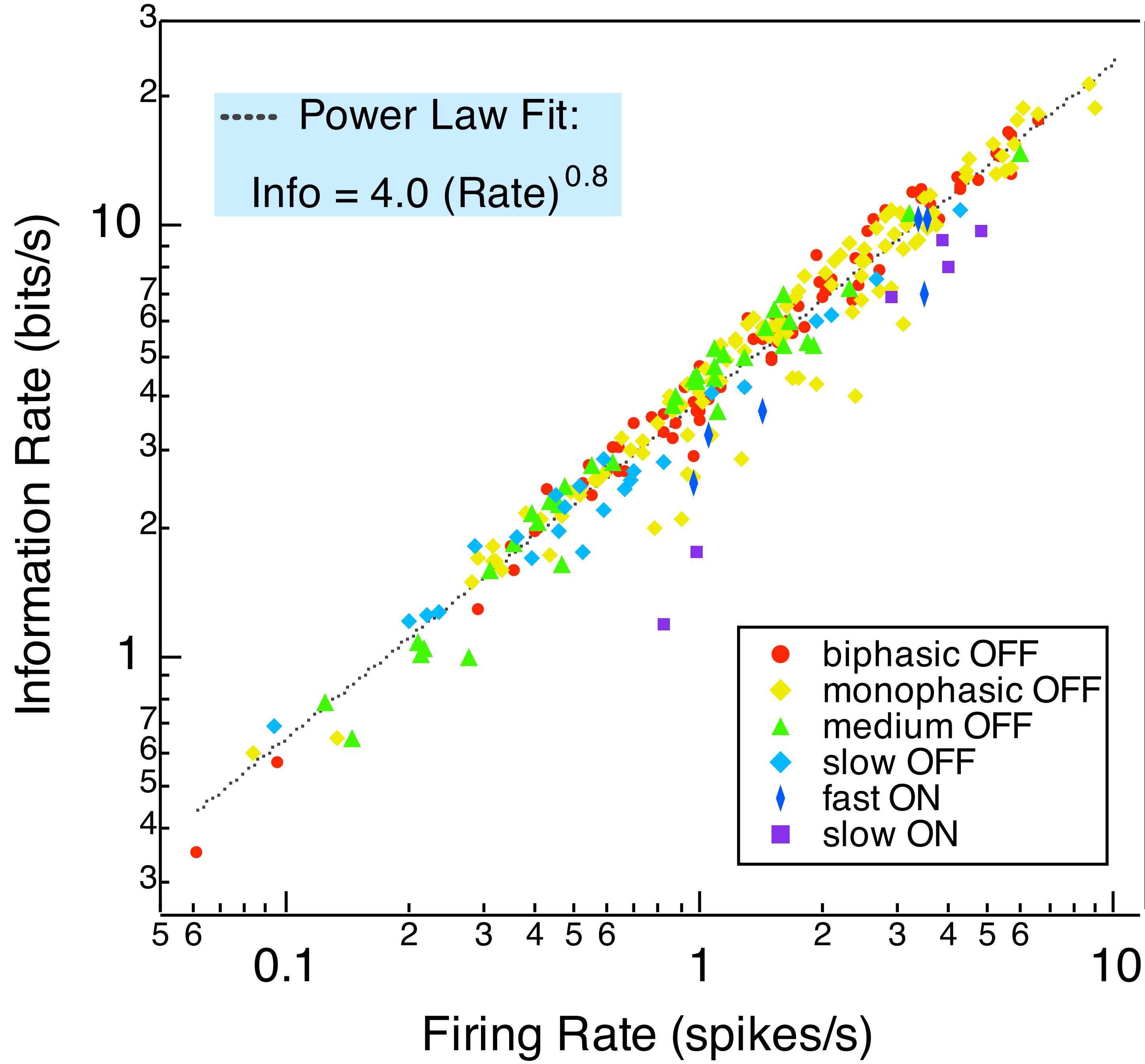




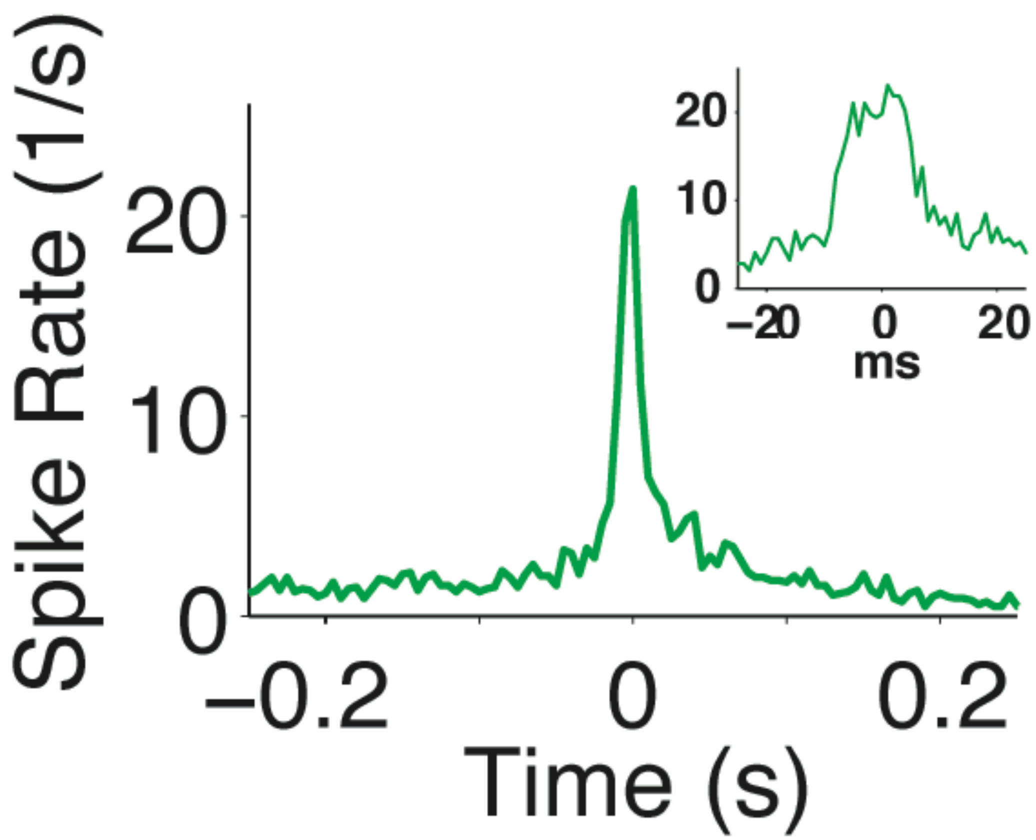
Information in single cells:

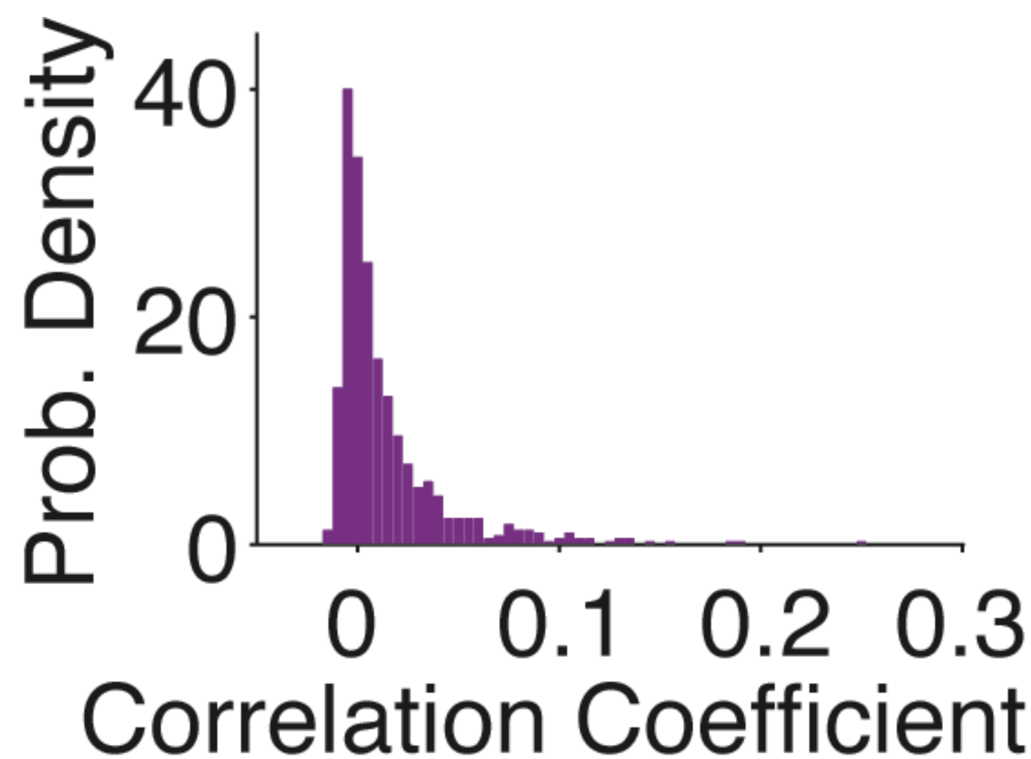
$$I(W;S) = H\left[P(w)\right] - \left\langle H\left[P(w|s)\right]\right\rangle_s$$

$$= -\sum_W P(w)\log_2 P(w) + \left\langle \sum_W P(w|s)\log_2 P(w|s)\right\rangle_s$$

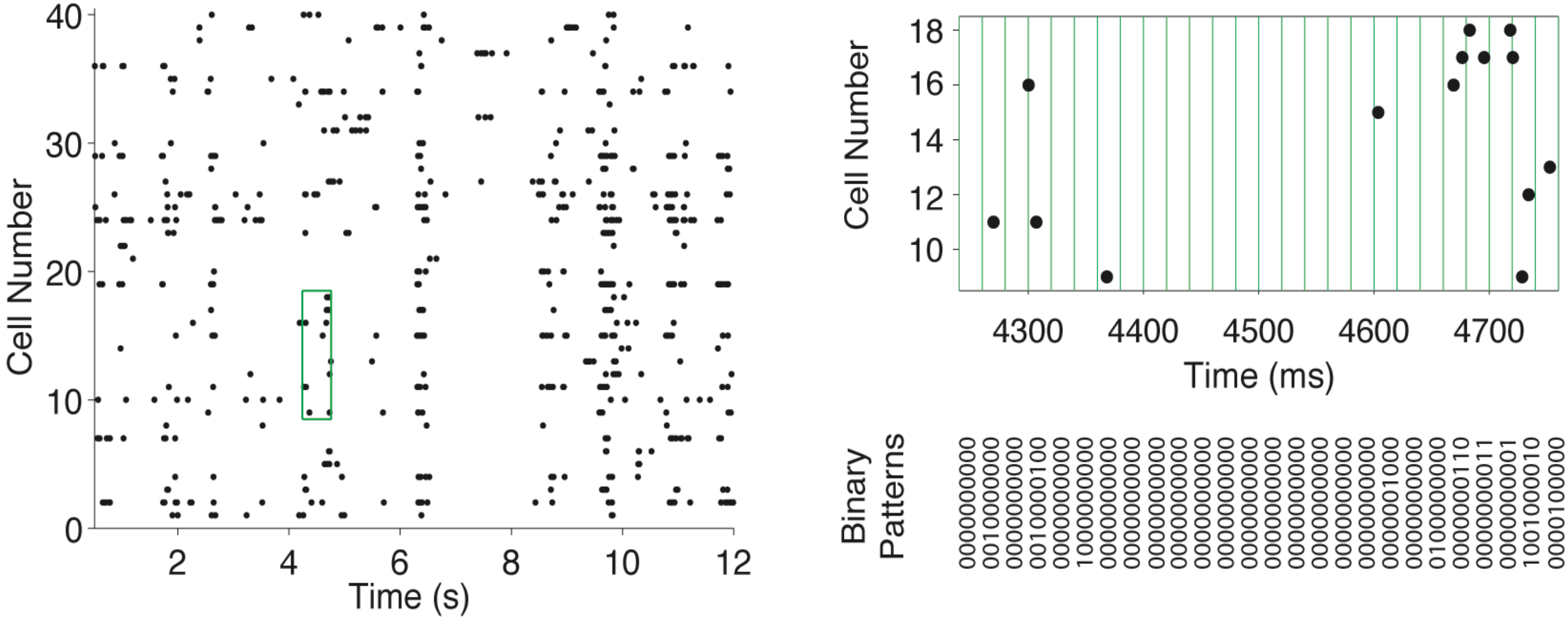


Pairwise correlations are weak:





Do we need correlations to describe population



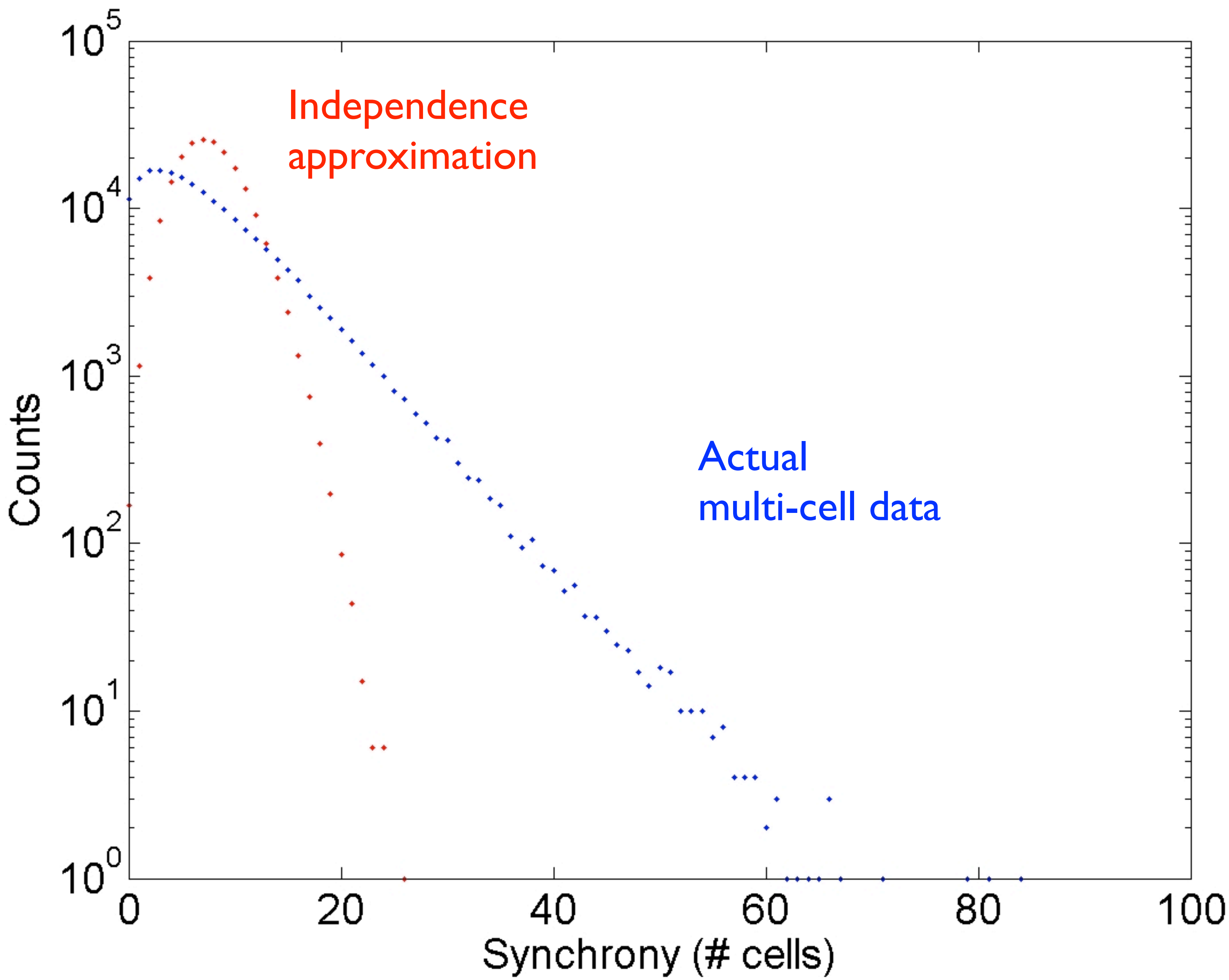
Can we replace $P(x_1, x_2, \ldots, x_n)$, with $P(x_1)P(x_2)\ldots P(x_n)$?

$o(n)$ parameters instead of $o(2^n)$

Redundancy reduction and efficient coding (Barlow 1961 | 2001, Attneave, 1954)

Large networks are highly correlated:

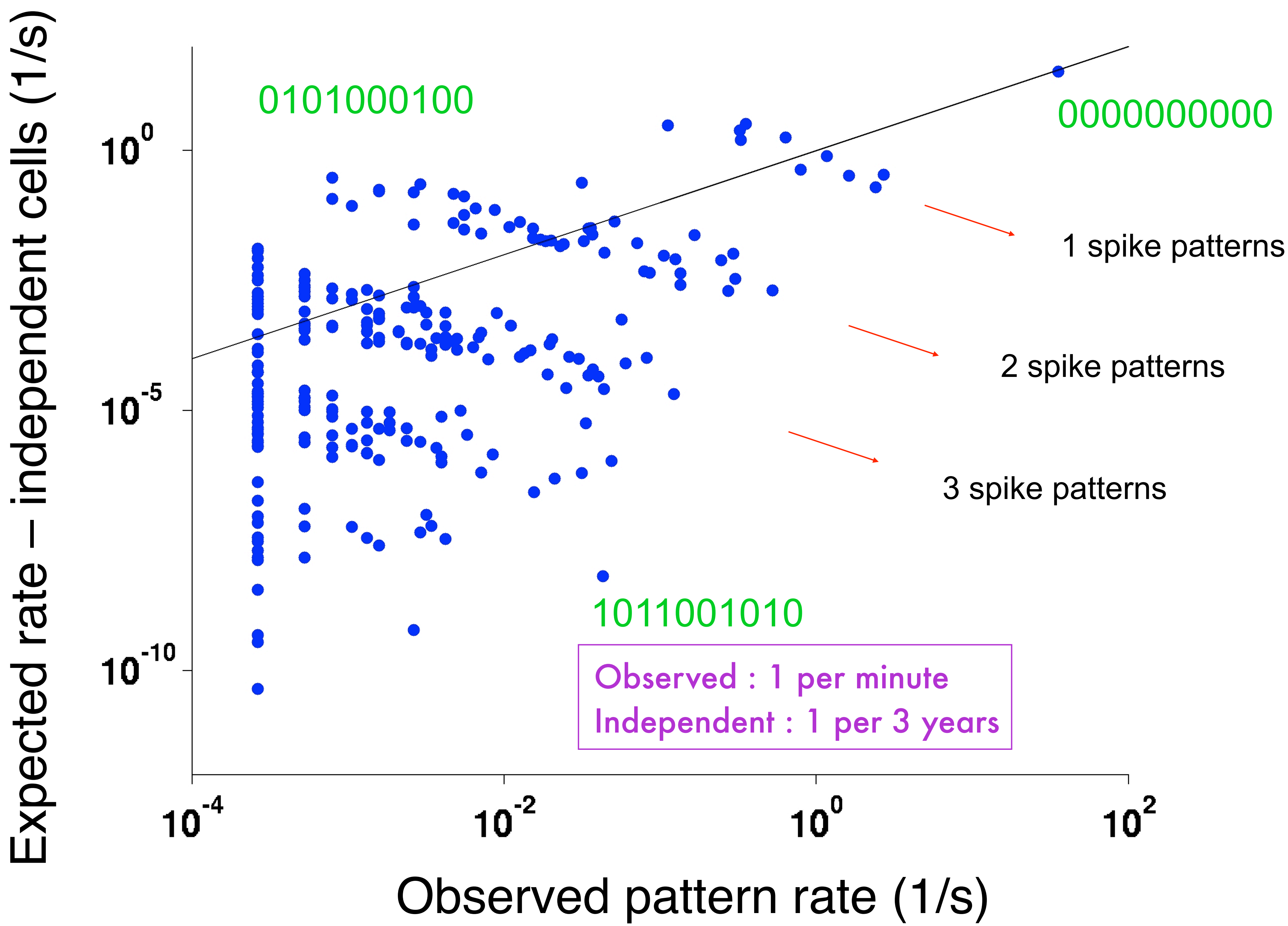
Synchrony of >100 retinal ganglion cells responding to a natural movie



Ganmor *et al.* (J Neurosci, 2011)

Even small networks are strongly correlated:

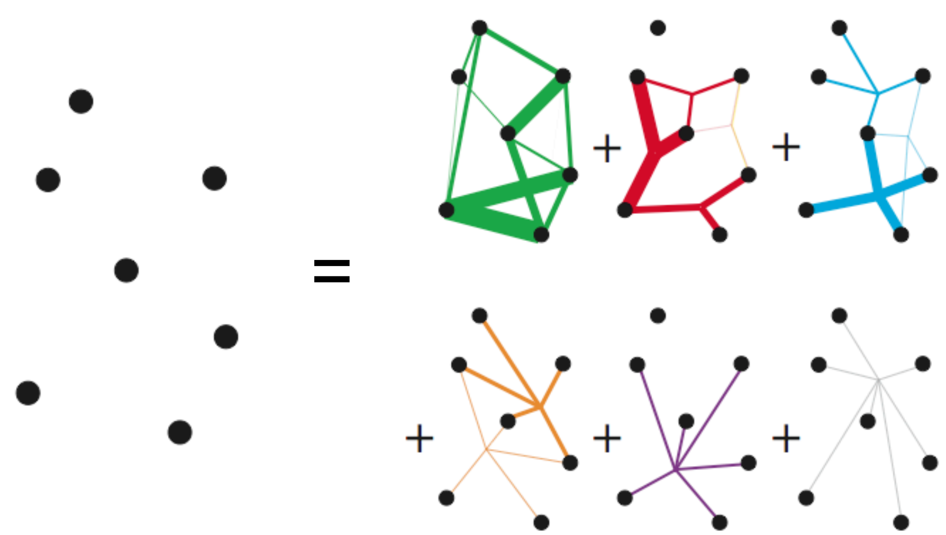
Example: 10 neurons responding to a natural movie



Schneidman *et al. Nature* (2006)

Is the network dominated by higher order corr's?

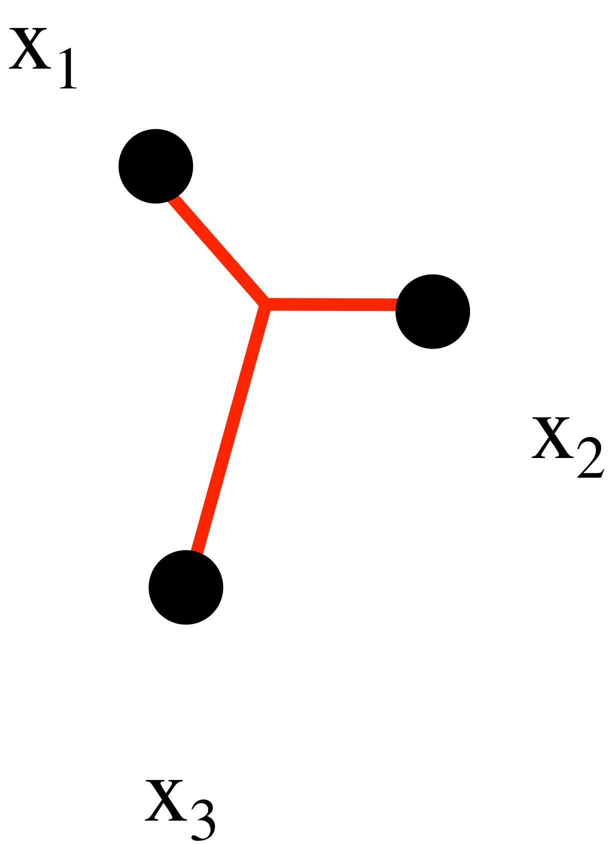
- $P(x_1)P(x_2)\ldots P(x_n)$ is a bad approximation
- Pairwise correlations seem to be weak
- How should we quantify the (high order) network correlations?



- Multi-Information

$$I\left(\{x_i\}\right) = H\left[p(x_1)p(x_2)...p(x_N)\right] - H\left[p(x_1, x_2, ..., x_N)\right]$$

Example of a pure 3-body correlation:



x_1	x_2	x_3	$P(x_1,x_2,x_3)$
0	0	0	0.25
0	1	1	0.25
1	0	1	0.25
1	1	0	0.25
0	0	1	
0	1	0	
1	0	0	
1	1	1	

The maximum entropy principle:

- Recall that the entropy of a distribution measures how much we don't know about the distribution
- Key Idea: of all the distributions that obey a given set of requirements, the maximum entropy distribution is the most parsimonious description we can look for
- (Intuitively, this is the most random/flat model that is consistent with the constraints)

Jaynes, *Phys Rev* (1957)

Defining the maximum entropy model:

$$\max H\left[p(x)\right] = -\sum_x p(x)\log p(x)$$

While obeying:

$$\sum_x p(x) O_i(x) = \langle O_i \rangle_{empr}$$

$$\sum_x p(x) = 1$$

The maximum entropy distribution:

$$L = -\sum_x p(x)\log p(x) - \sum_i g_i \left(\sum_x p(x) O_i(x) - \langle \hat{O}_i \rangle \right) - \mu \left(\sum_x p(x) - 1 \right)$$

$$\frac{\partial L}{\partial g_i} = \sum_x P(x) O_i(x) - \langle O_i(x) \rangle_{empr}$$

$$\frac{\partial L}{\partial \mu} = \sum_x p(x) - 1$$

$$p(x|\{g_i\}) = \frac{1}{Z(\{g_i\})} exp\{\sum_i g_i O_i(x)\}$$

$$Z(g) = \sum_x \exp\left\{ \sum_i g_i O_i(x) \right\}$$

Maximum entropy + constraints give an exponential distribution

Fitting a maximum entropy model:

- Convexity
- Generalized Iterative Scaling (Darroch+Ratcliff '72)
  update model using the ratios of empirical marginals to model marginals:

$$p^{(t+1)}(x) = p^{(t)}(x)\prod_i \left(\frac{\sum_y p_{empr}(y) O_i(y)}{\sum_y p^{(t)}(y) O_i(y)} \right)^{O_i(x)}$$

- Max Likelihood and MaxEnt:

$$argmax_{\{g_i\}} \log\{\prod_{k=1}^{M} p(x^k|\{g_i\})\}$$

use Gradient ascent:

$$\Delta g_i^{(t+1)} = -\eta \left[\left\langle O_i(x) \right\rangle_{p^{(t)}(x)} - \left\langle O_i(x) \right\rangle_{p_{\text{empr}}} \right]$$

Maximum entropy distribution based on pairwise

What is the minimal model which is consistent with the firing rates $\langle x_i \rangle$ and the pairwise correlations $\langle x_i x_j \rangle$, but makes no other assumptions?

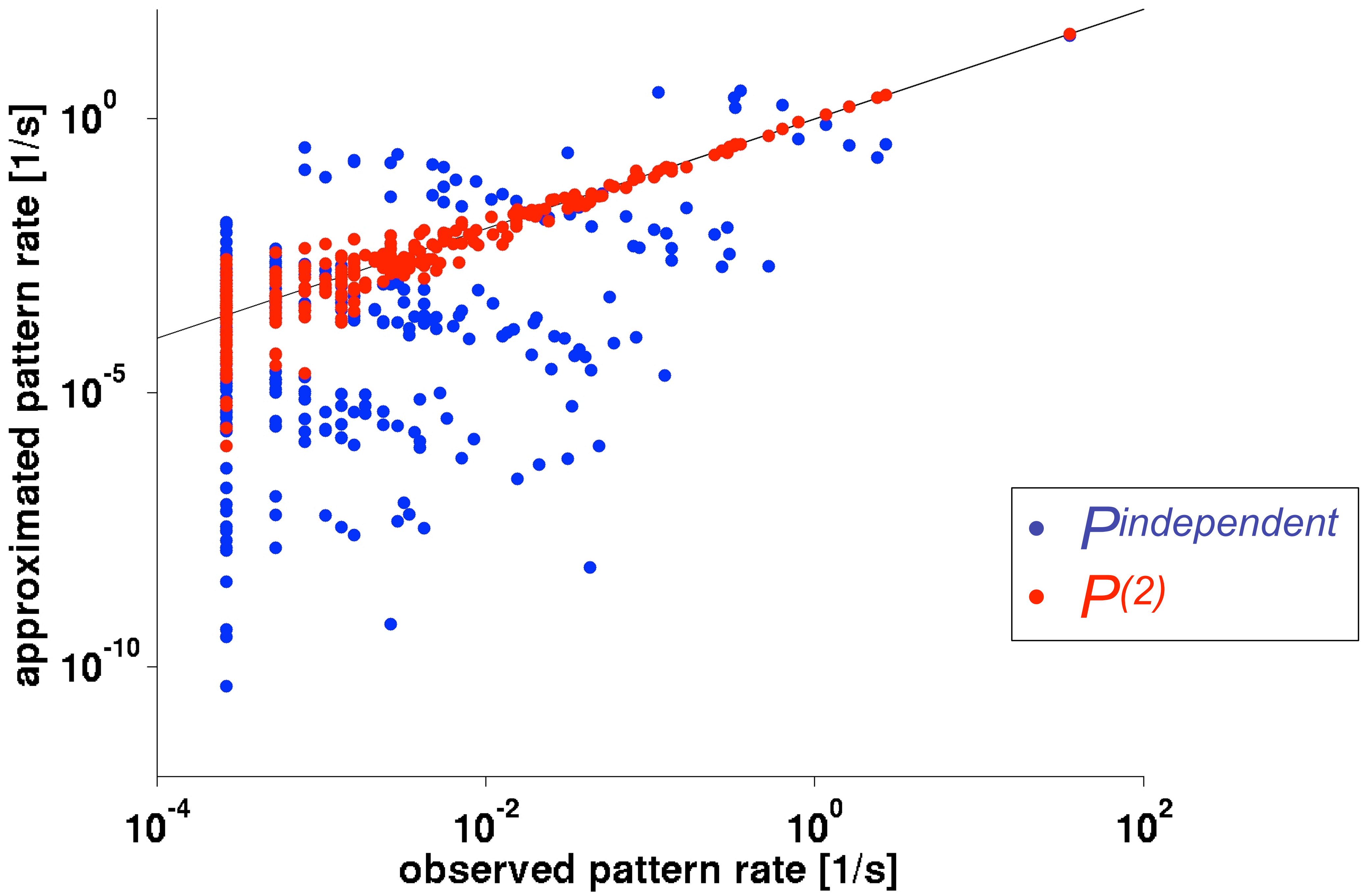
$$\begin{aligned} L\left(p(x_1,x_2,...,x_N),\{\alpha_i\},\{\beta_i\}\right) = & - \sum_{x_1,x_2,\ldots,x_N} p(x_1,x_2,...,x_N)\log p(x_1,x_2,...,x_N) \\ & - \sum_i \alpha_i \left(\langle x_i \rangle_p - \langle x_i \rangle_{p_{empr}}\right) - \sum_{i<j} \beta_{ij}\left(\langle x_i x_j \rangle_p - \langle x_i x_j \rangle_{p_{empr}}\right) \\ & - \mu \left(\sum_{x_1,x_2,\ldots,x_N} p(x_1,x_2,...,x_N) - 1 \right) \end{aligned}$$

Unique solution (numerical convex optimization):

$$P^{(2)}\left(x_1,x_2,\ldots,x_n\right) = \frac{1}{Z}\exp\left[-\sum_i \alpha_i x_i - \sum_{i,j}\beta_{ij}x_i x_j\right]$$

N(N+1)/2 parameters instead of 2^N

This pairwise model is an excellent fit to data:



Schneidman *et al.*, *Nature* (2006)

Quantifying contributions of k-th order interactions:

The total network correlation is given by

$$I\left(\{x_i\}\right) = H\left[p(x_1)p(x_2)...p(x_N)\right] - H\left[p(x_1, x_2, ..., x_N)\right]$$

Build a hierarchy of correlation based models:

$$p^{(k)}(x_1, x_2, ..., x_n) = argmax_p H\left[\tilde{p}(x_1, x_2, ..., x_N)\right]$$

with constraints for all correlations up to order k

Schneidman *et al.* *Phys Rev Lett* (2003); see a geometric derivation in Amari IEEE-IT, 2001

Quantifying the contributions of k-th order

$$p(x_1)p(x_2)...p(x_N) \qquad \langle x_i \rangle$$

$$p^{(2)}(x_1, x_2, ..., x_N) \qquad \langle x_i \rangle \; \langle x_i x_j \rangle$$

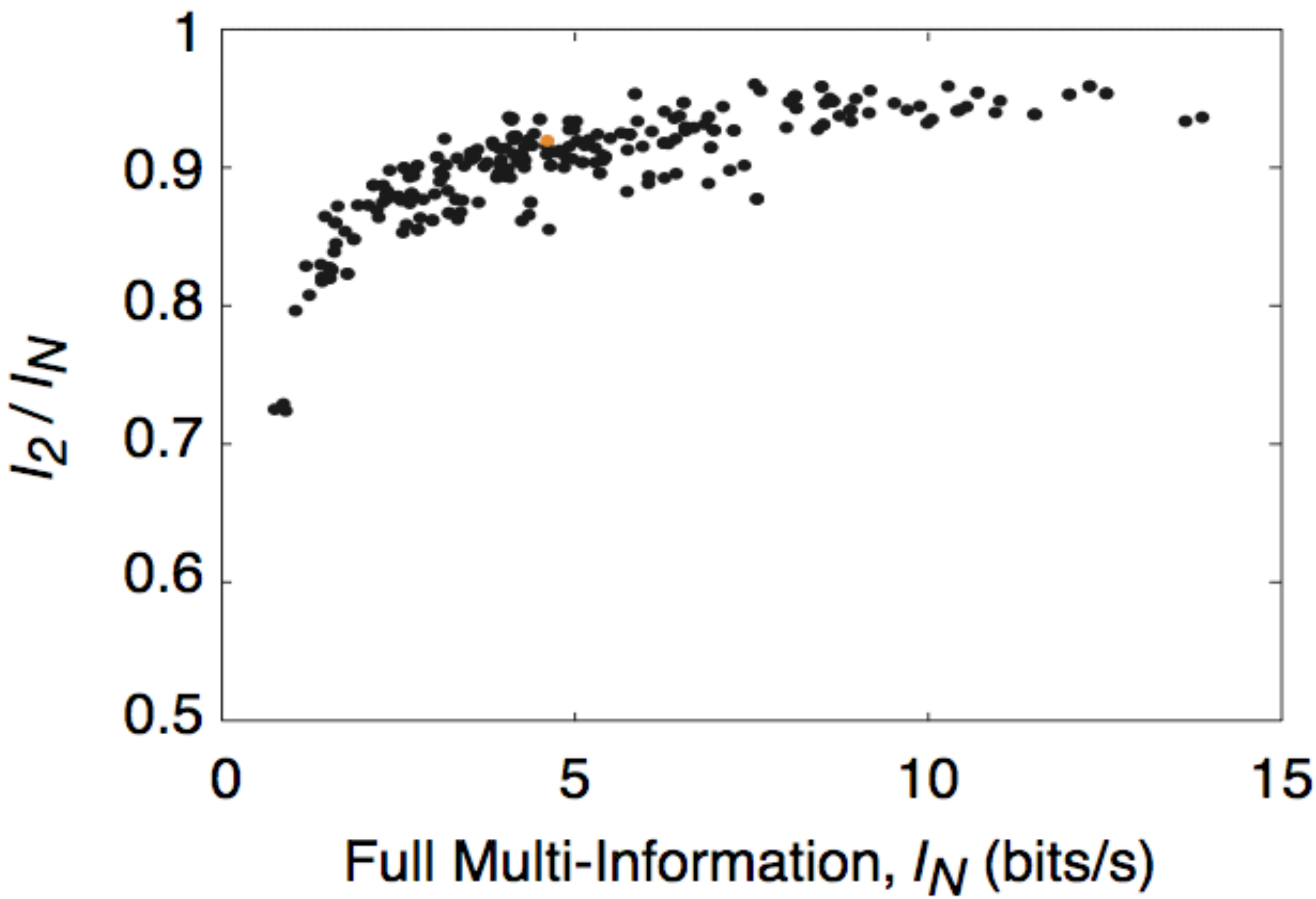
$$p^{(3)}(x_1, x_2, ..., x_N) \qquad \langle x_i \rangle \; \langle x_i x_j \rangle \; \langle x_i x_j x_k \rangle$$

$$p(x_1, x_2, ..., x_N)$$

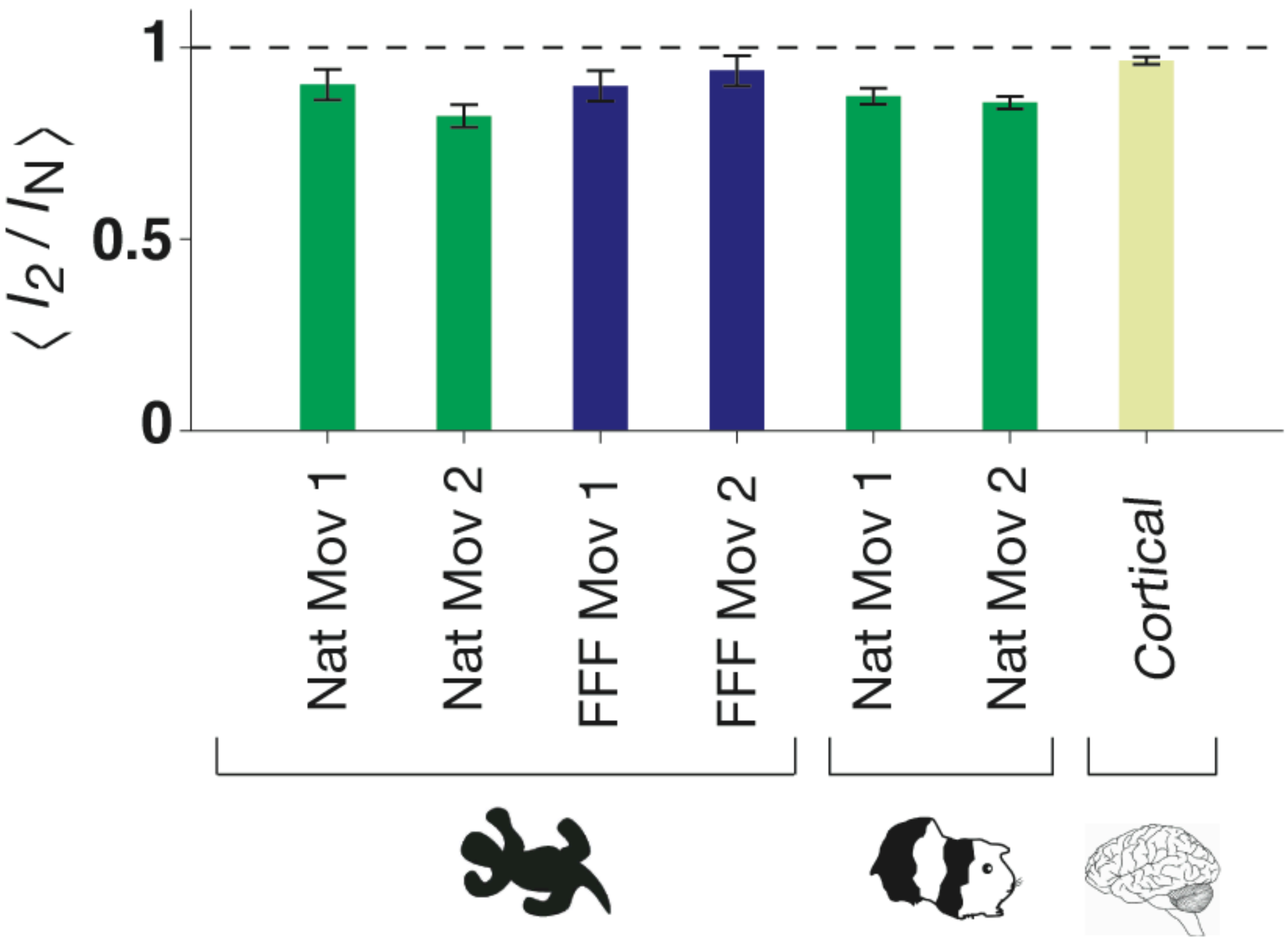
The contribution of correlations of order k

$$I^{(k)}\left(\{x_i\}\right) = H\left[p^{(k)}(x_1, x_2, ..., x_N)\right] - H\left[p^{(k-1)}(x_1, x_2, ..., x_N)\right]$$

The pairwise information captures nearly all of the



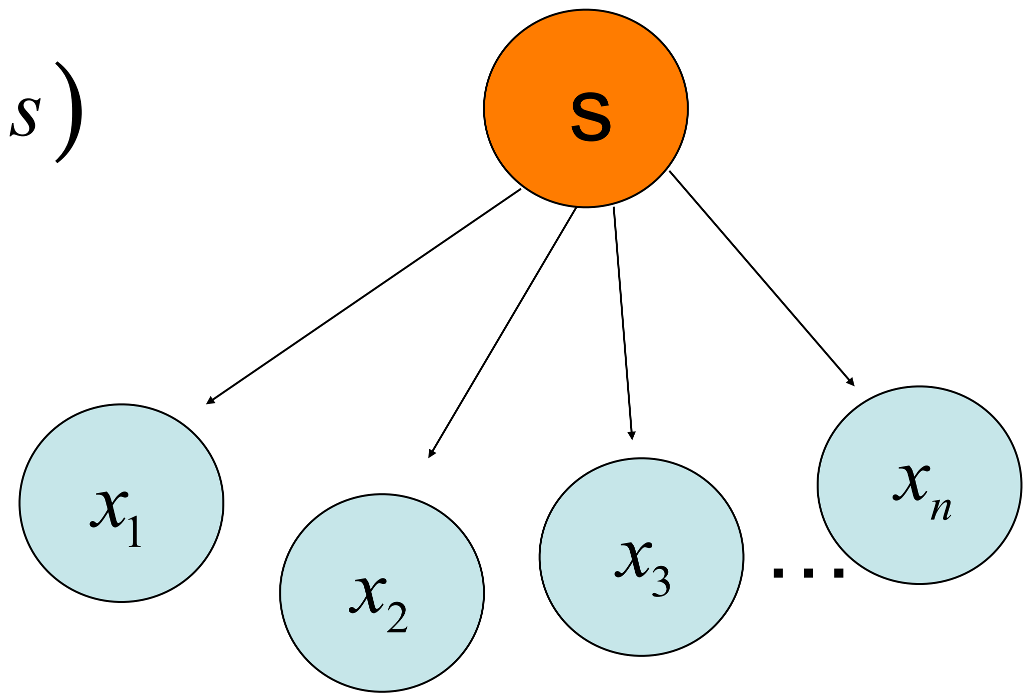
This works for a variety of preps, different stims:



Perhaps the cells are conditionally independent?

Maybe the cells are independent given the stimulus?

$$P_{cond-ind}\left(\{x_i\} \mid s\right) = P\left(x_1 \mid s\right) P\left(x_2 \mid s\right) \ldots P\left(x_n \mid s\right)$$



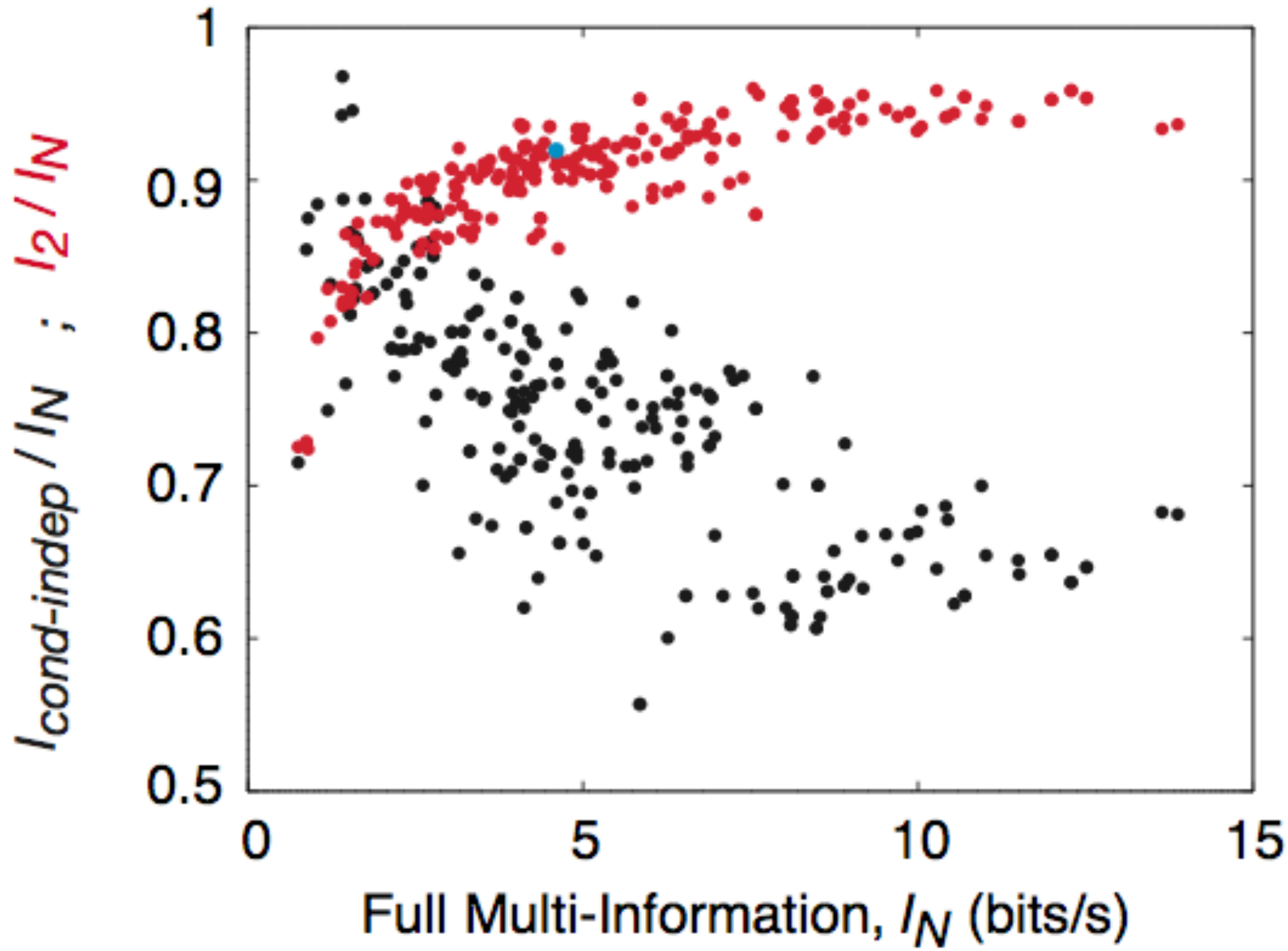
Benefits:

- Easy to construct, using repeated presentations of the same stimulus
- Can be used to combine cells from different experiments

Drawbacks:

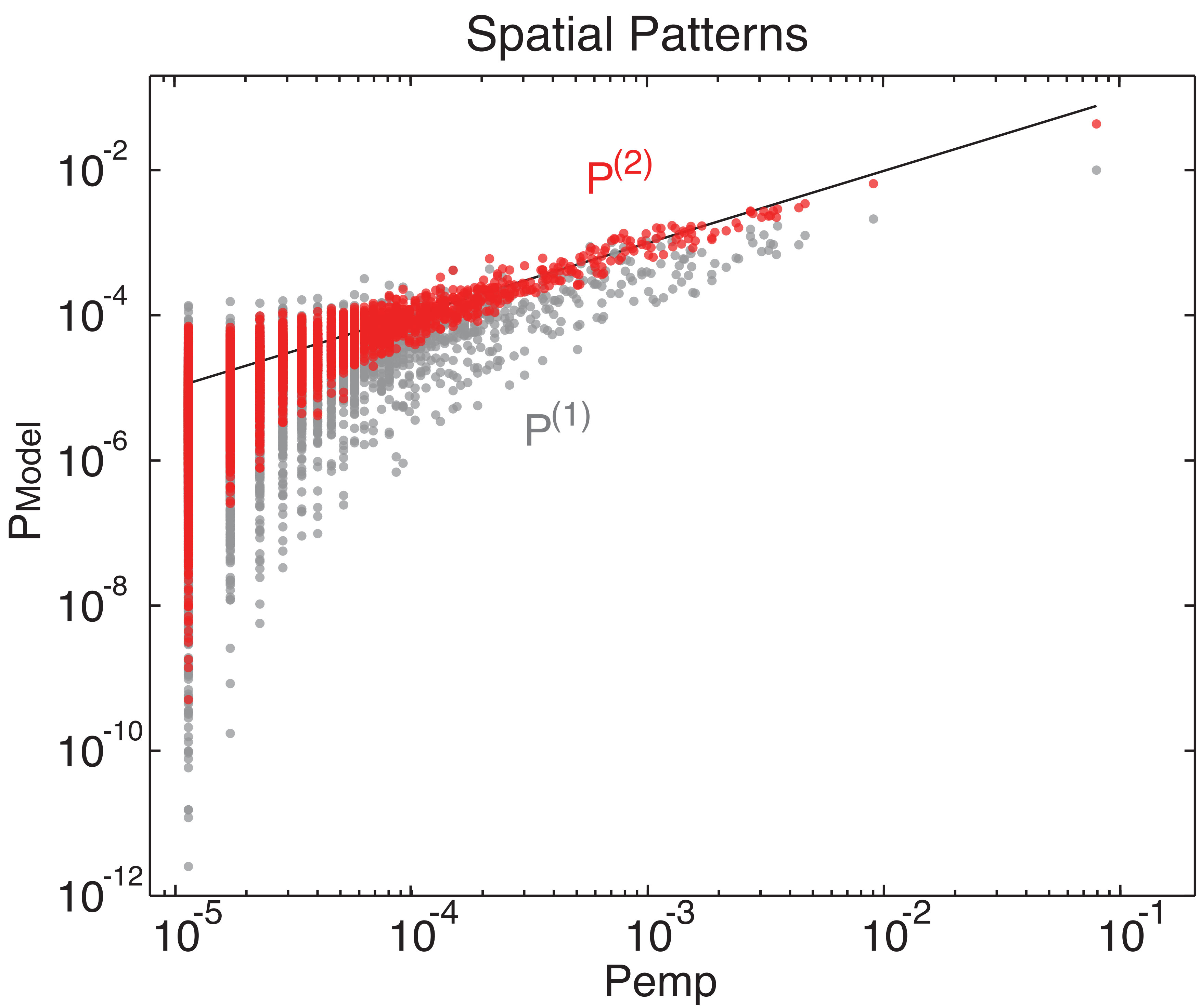
- Ignoring stimulus induced correlations
- The brain has no access to repeated presentations…

...but this isn't a good model:

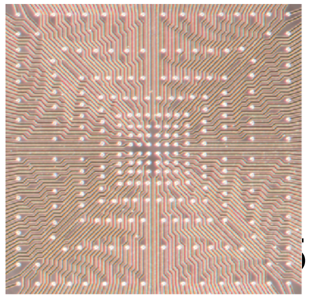


In this case $P^{(2)}$ has 55 parameters, $P^{\text{cond-indep}}$ has ~10,000

A pairwise maxent model for 100 retinal cells:



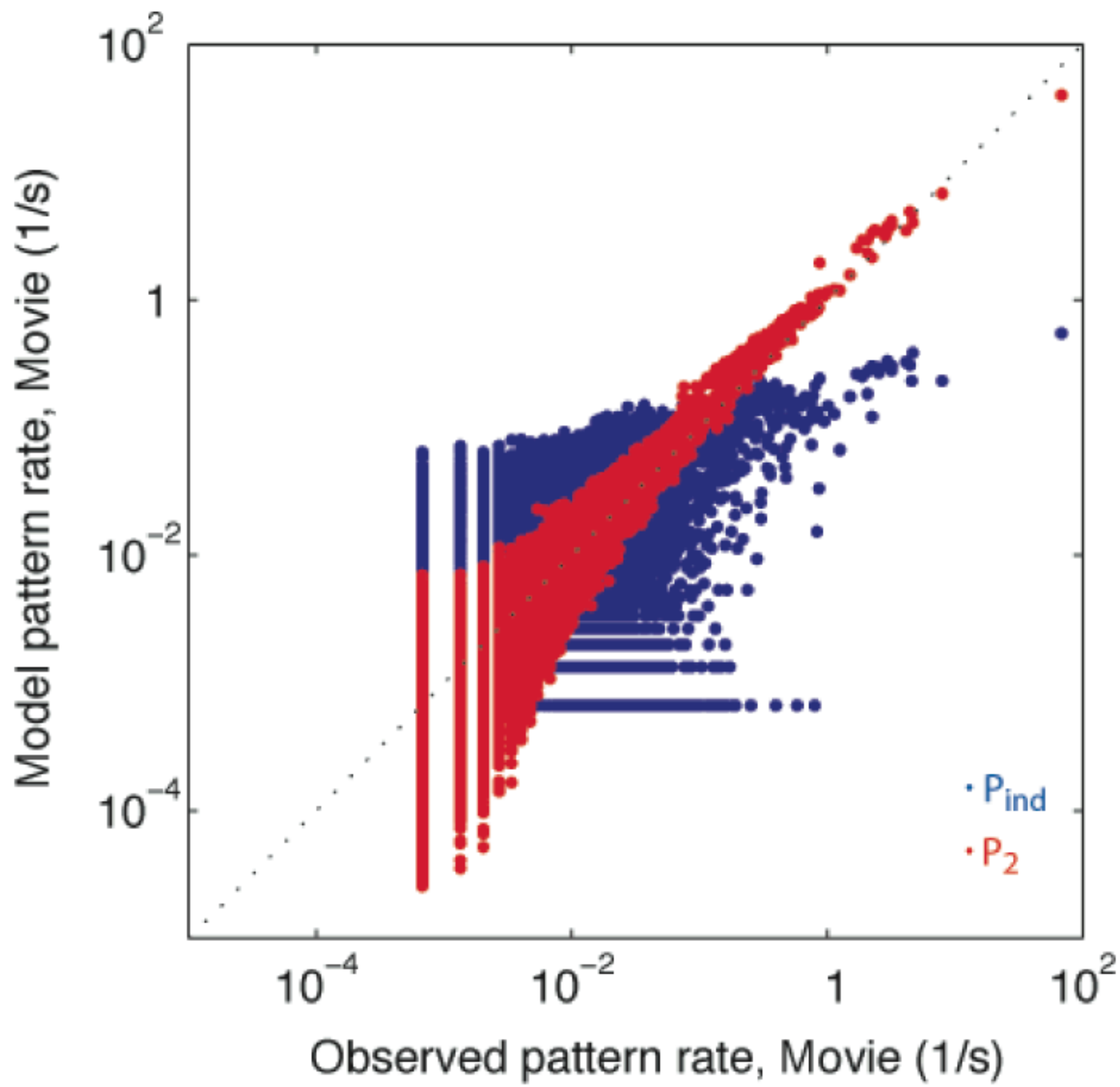
Wide MEA



Ganmor et al. (J Neurosci 2011)

A pairwise model works for the cortex, too:

V1 of awake ferrets watching a movie; 16 multi-unit, 2ms bins



Schneidman + Weliky (unpublished)

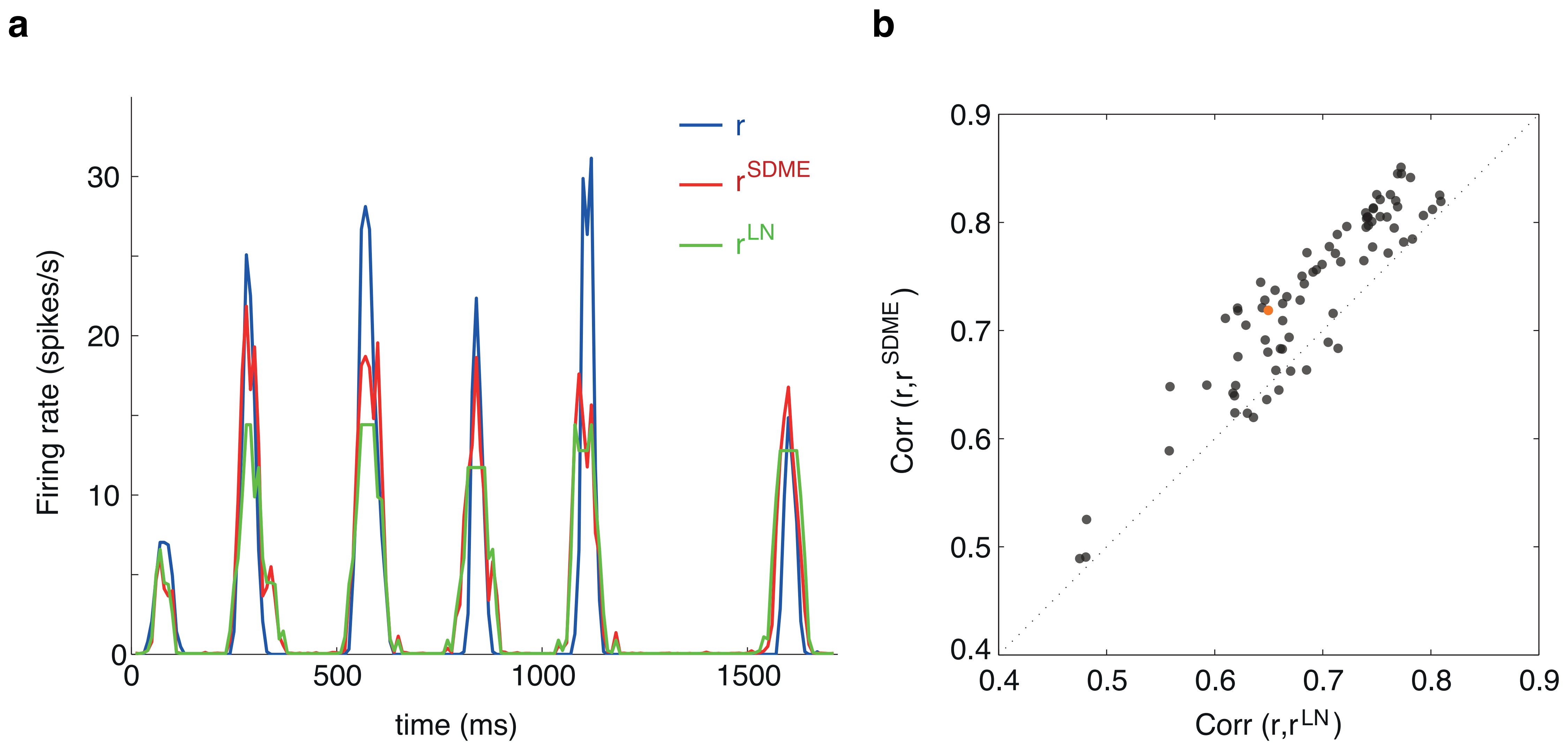
Adding in stimulus dependencies in the code:

- Combine the LN model with MaxEnt:
  Fit LN model for each of the cells,
  then solve the maxent model given the
  firing rates and correlations

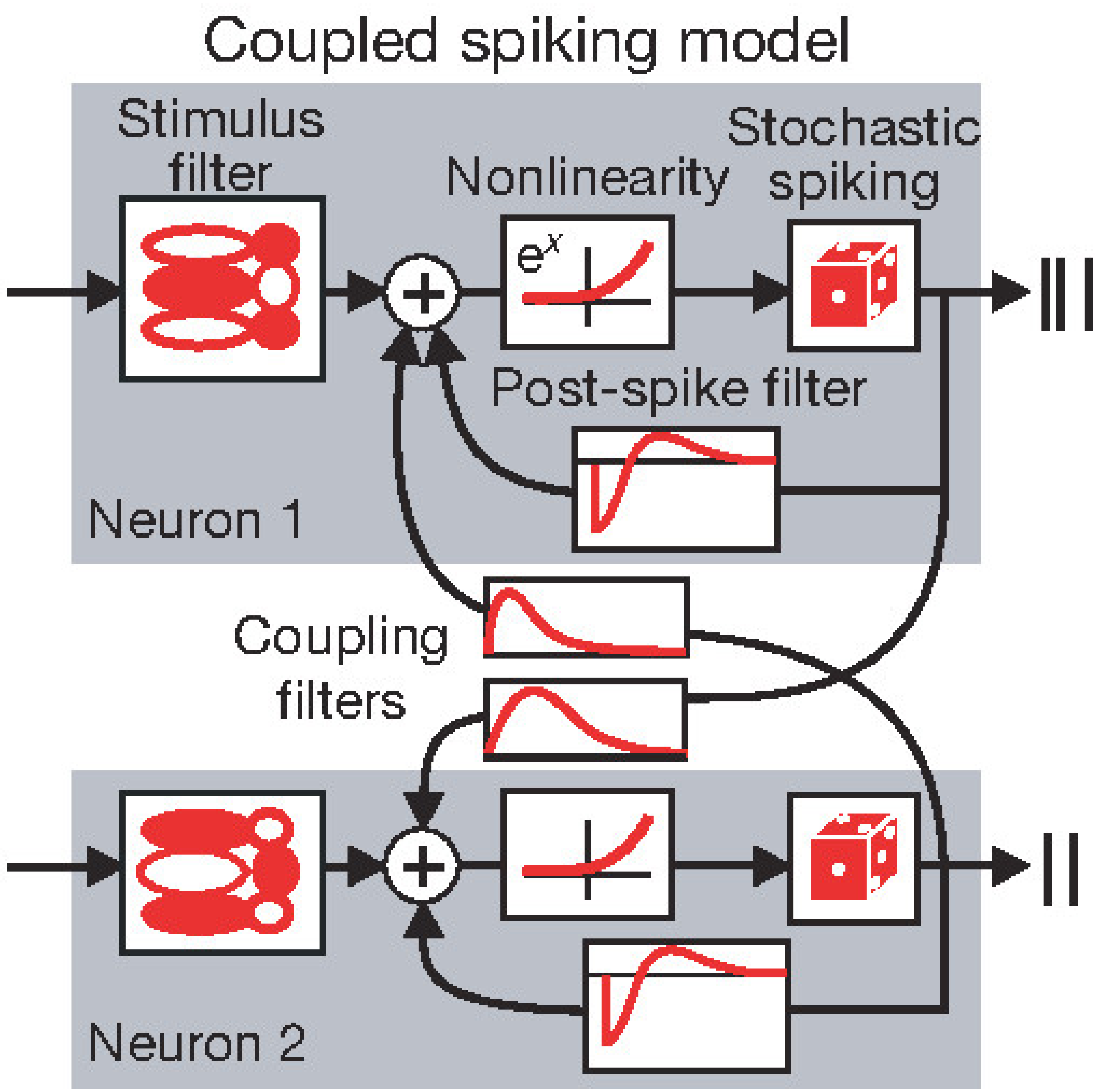
$$p(x_1, x_2, ..., x_n|s) = \frac{1}{Z} exp \left(\sum_i \alpha_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_i \theta_i(s) x_i \right)$$

Granot et al.(Plos Comp Bio 2013)

This model is better than LN alone:



Fitting correlated spike trains: GLMs



Pillow et al., Nature (2008)

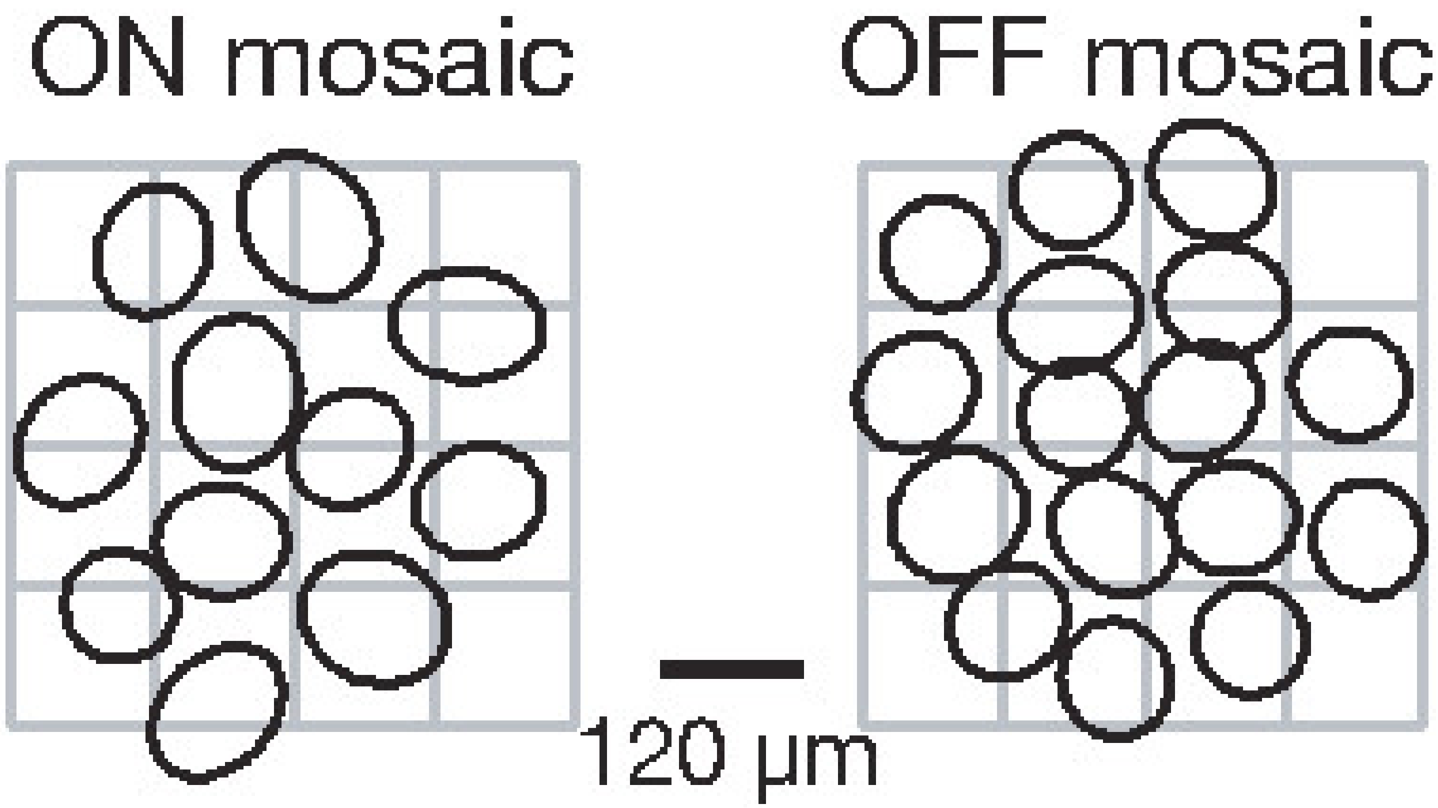
GLMs under the hood:

- Spike trains: inhomogeneous Poisson processes
- Instantaneous firing rate:

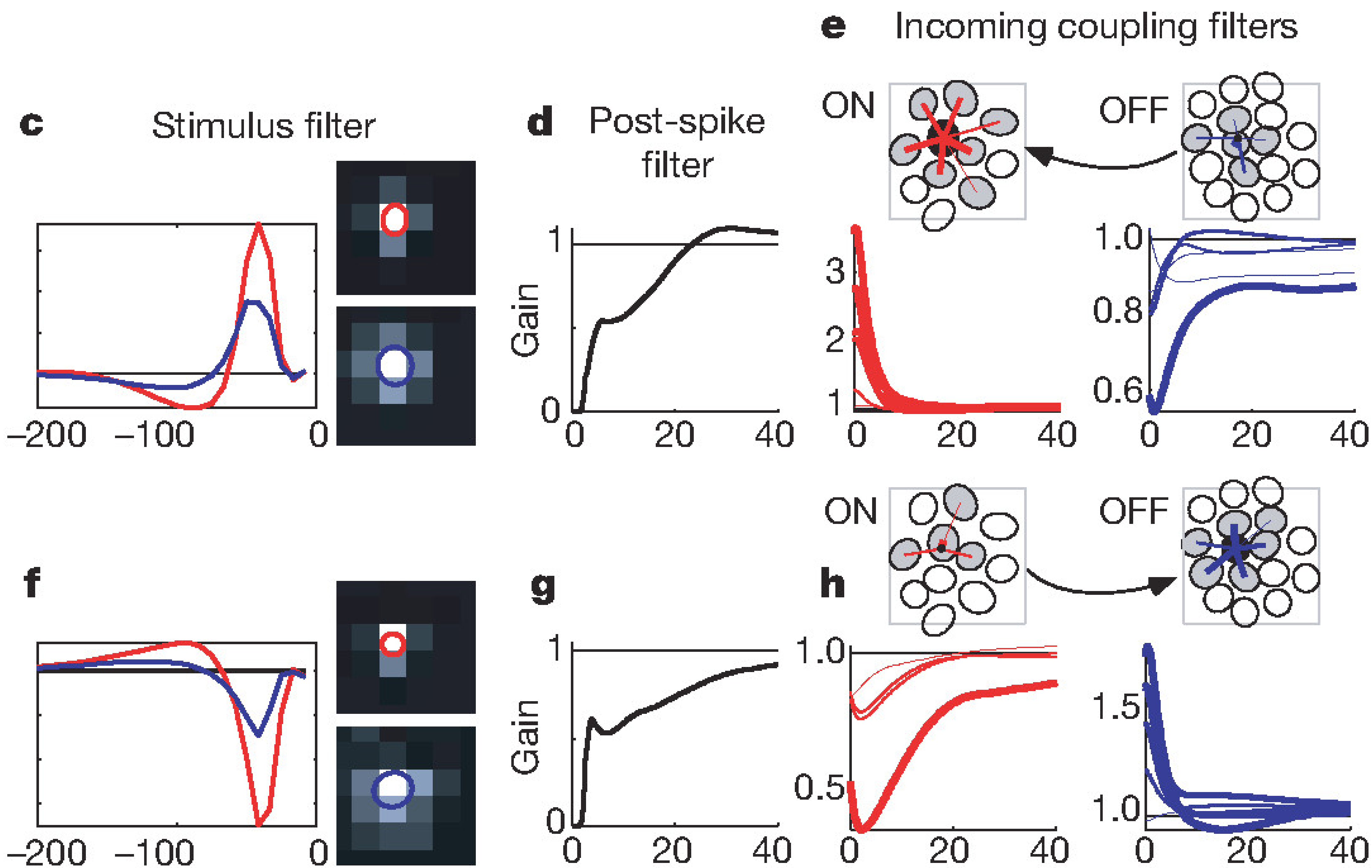
$$\lambda_i(t) = \Phi\left(k_i * s_i(t) + h_i * y_i(t) + \sum_j h_{ij} * y_j(t)\right)$$

- Can be used to fit data using maximum likelihood (find filters k and h that maximize likelihood of observed spike trains)

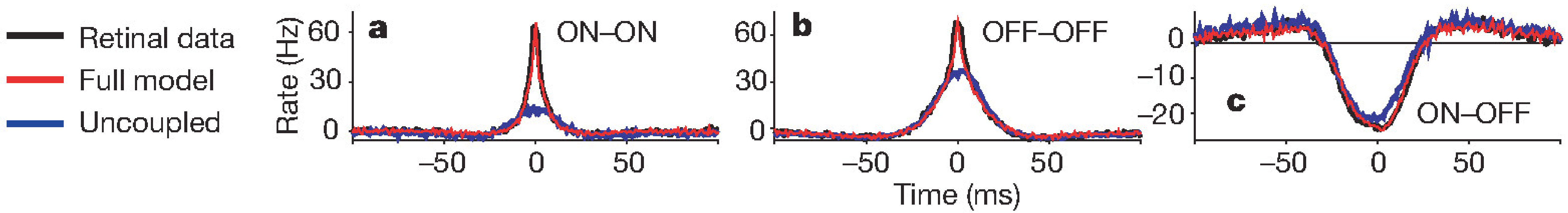
Fitting retina data, macaque retina:



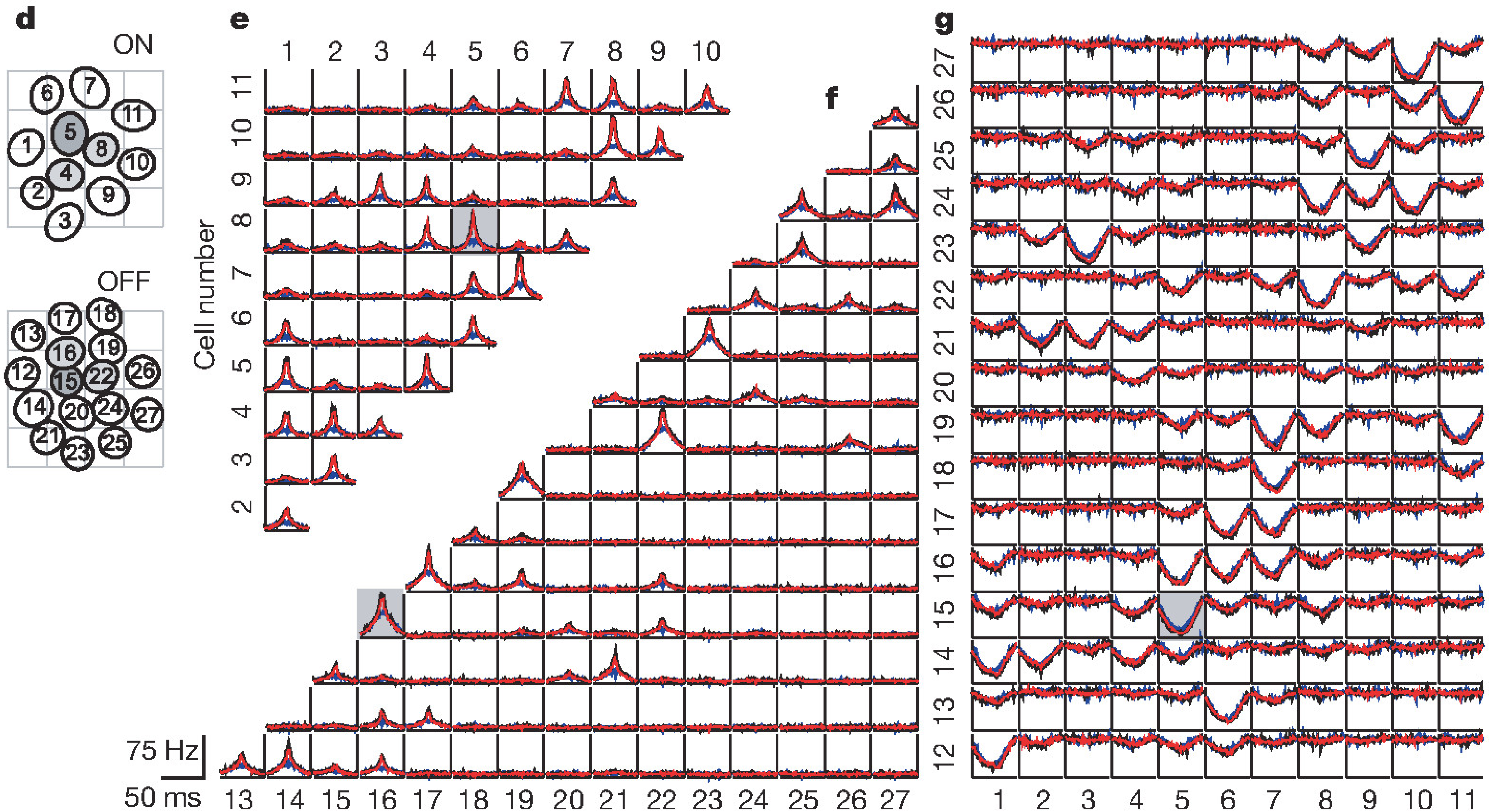
What sorts of filters do you get?



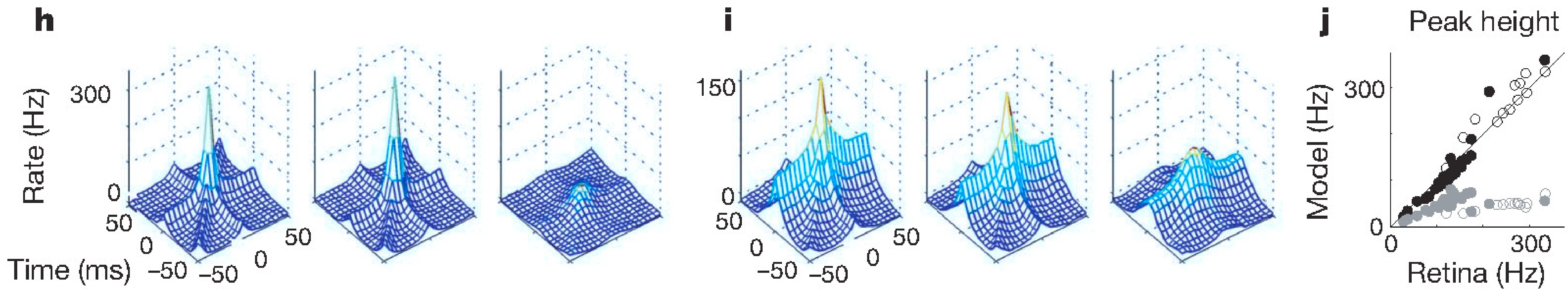
Fitting the cross-correlations between cell types:



Fitting the temporal cross-correlations:



Cross-correlations:



PSTHs from the GLM match fairly nicely: