NEURAL DATA ANALYSIS

ALEXANDER ECKER, PHILIPP BERENS, MATTHIAS BETHGE

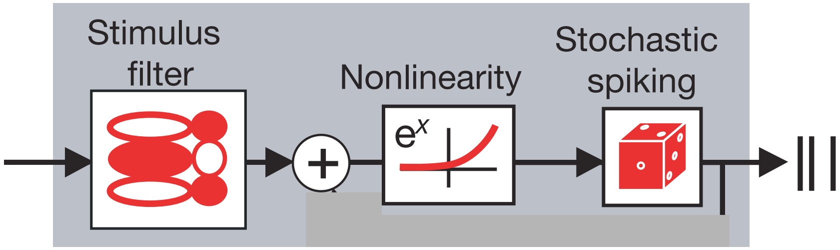
COMPUTATIONAL VISION AND NEUROSCIENCE GROUP

GENERALIZED LINEAR MODELS

SHORTCOMINGS OF MODELS SO FAR

What we have:

- **Stimulus (in space and time)** (**L**inear)
- **(Static) non-linearity** (**N**on-linear)
- **Probabilistic spiking** (**P**oisson)



Pillow et al. 2008

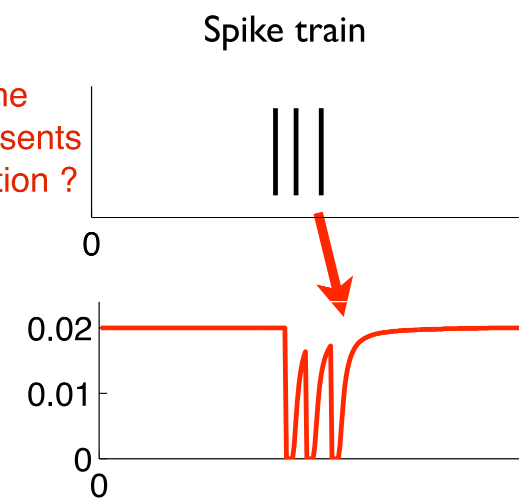
Not included so far:

- **Spiking dynamics**
  - Refractory period
  - Bursting
- **Interactions between neurons**

REFRACTORY PERIODS AND BURSTING

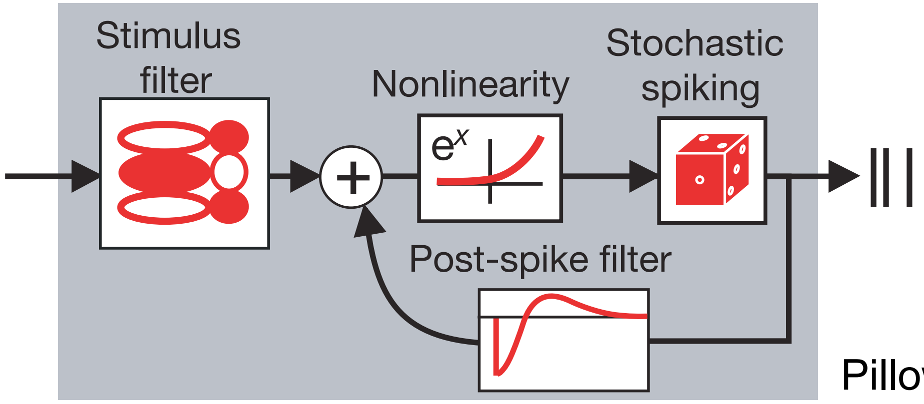
Include ***self-feedback*** **term which excites/inhibits the neuron after spikes**

I suppose the y axis represents autocorrelation ?



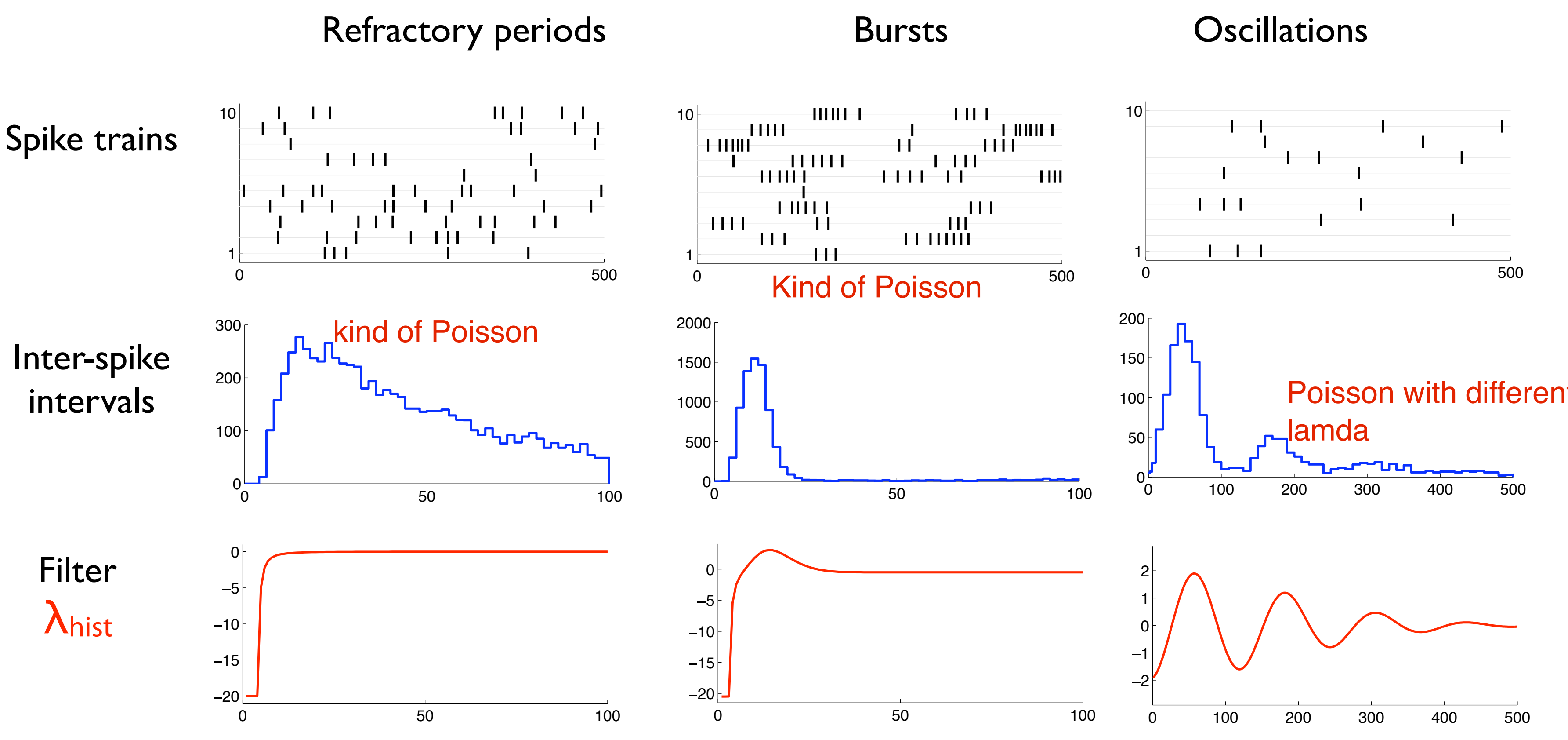
$$r(t) = \exp(\lambda_{\text{stimulus}} + \lambda_{\text{history}})$$

$$\lambda_{\text{history}}(t) = \sum_{\tau} h_{\tau} y_{t-\tau}$$



Pillow et al. 2008

EXAMPLES OF HISTORY EFFECTS

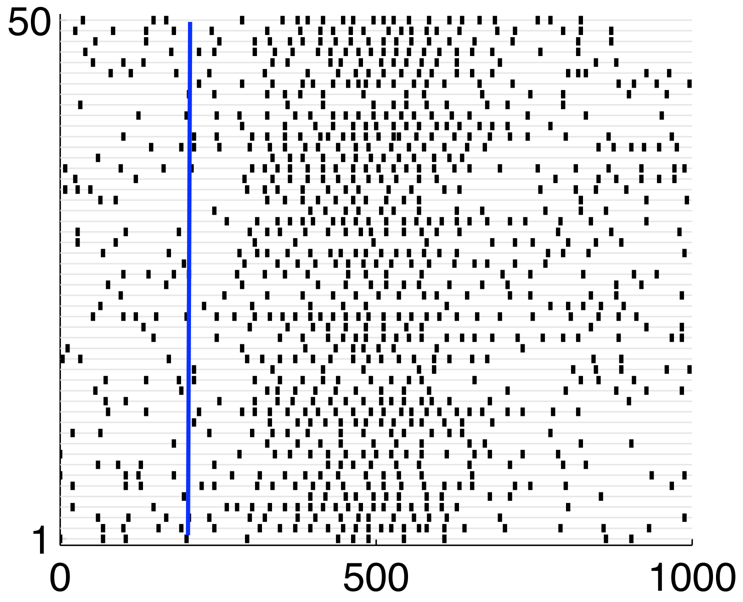


Infer history terms by optimizing likelihood as before

FITTING STIMULUS AND HISTORY SIMULTANEOUSLY

If a spike occurs at time *t*

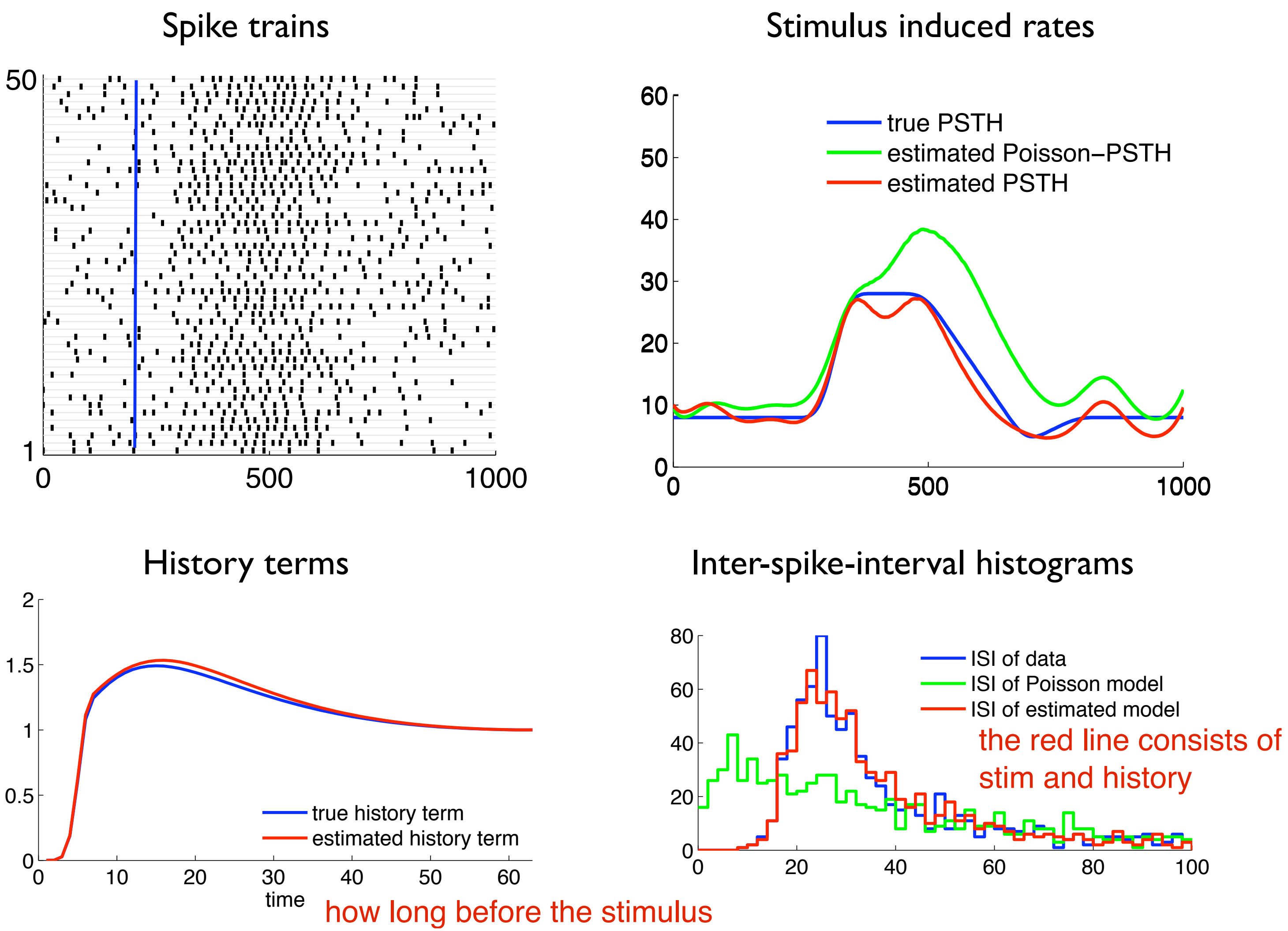
- **Stimulus-induced rate is high (e.g. stimulus onset)**
- **Spike is consequence of recent spiking history (e.g. burst)**



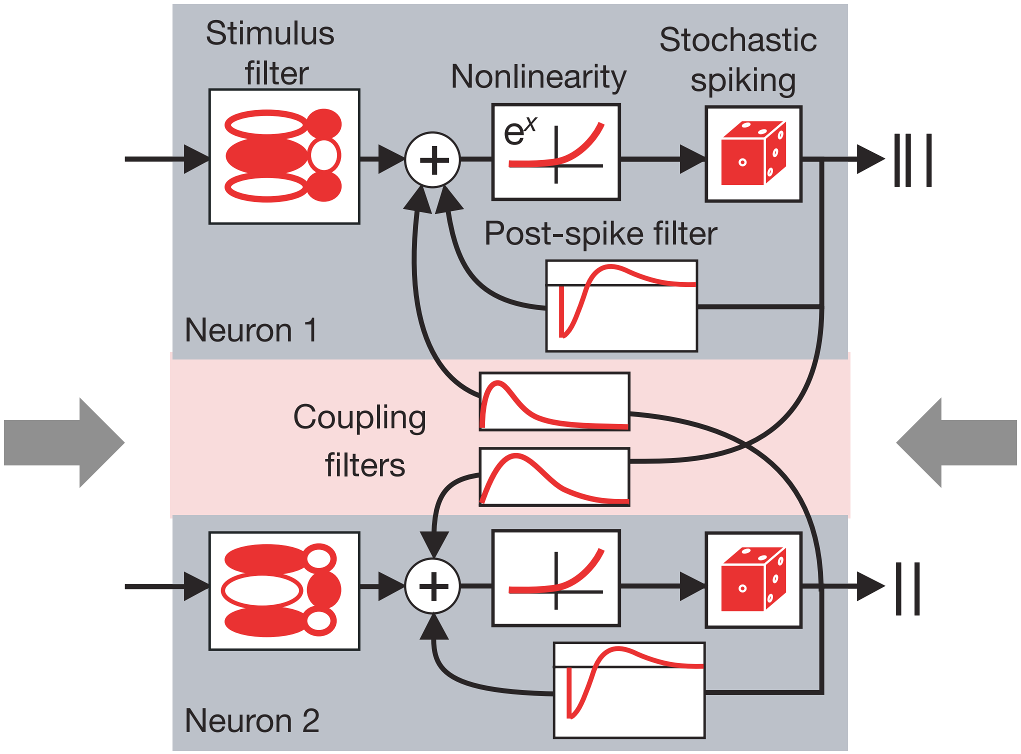
Averaging spikes across trials confounds the two!

→ Infer stimulus and history terms simultaneously

FITTING STIMULUS AND HISTORY SIMULTANEOUSLY

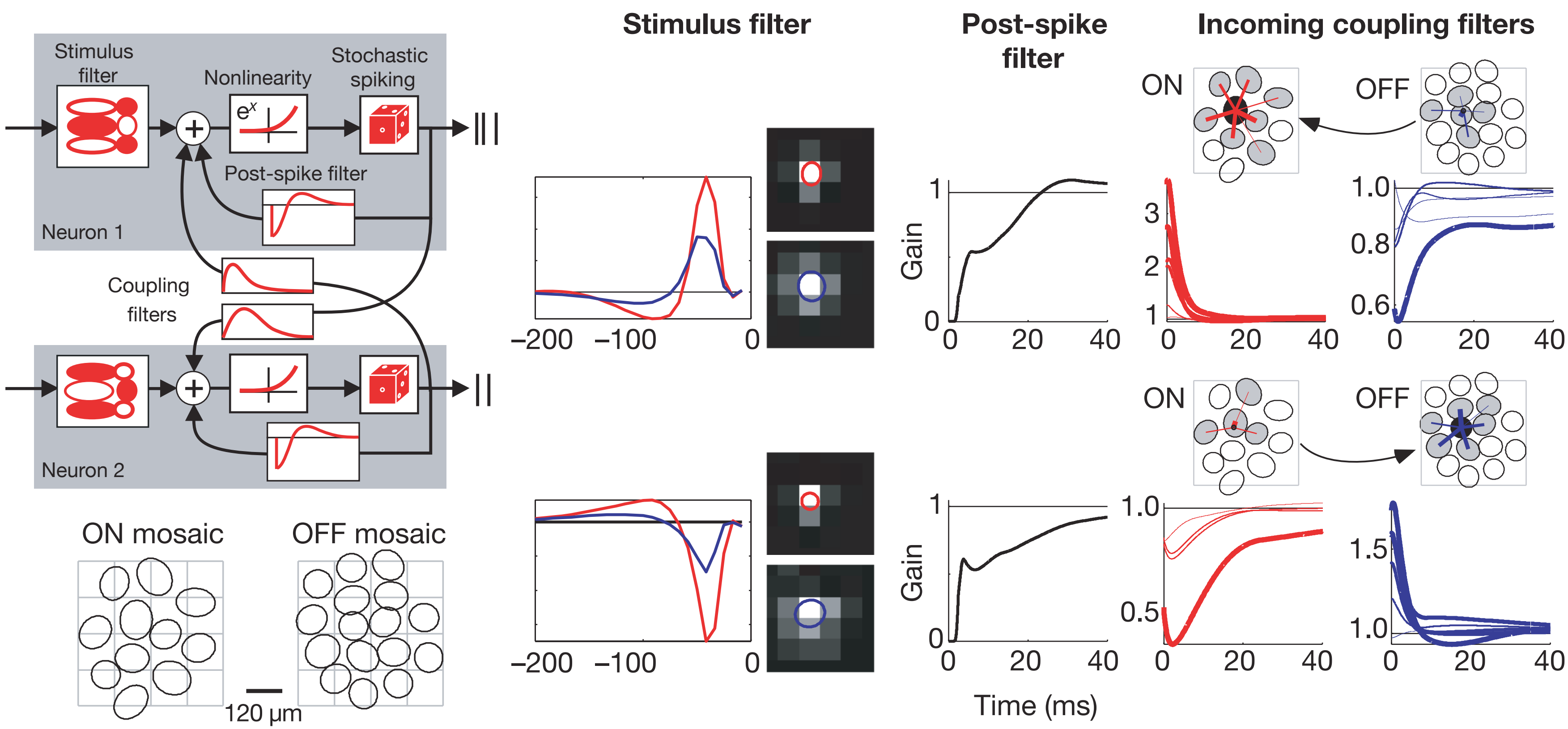


BEYOND SINGLE CELLS: PAIRWISE COUPLINGS



Pillow et al. 2008

GLM FOR RETINAL POPULATION



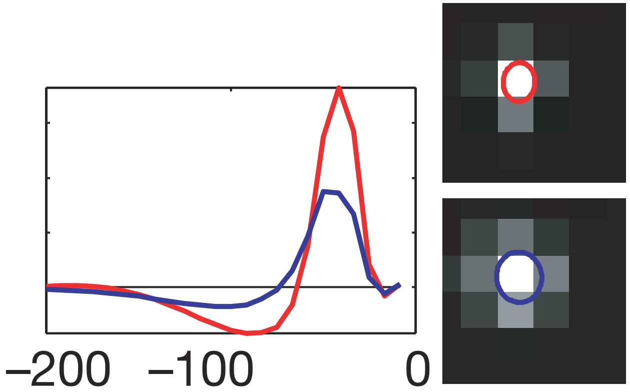
Pillow et al. 2008

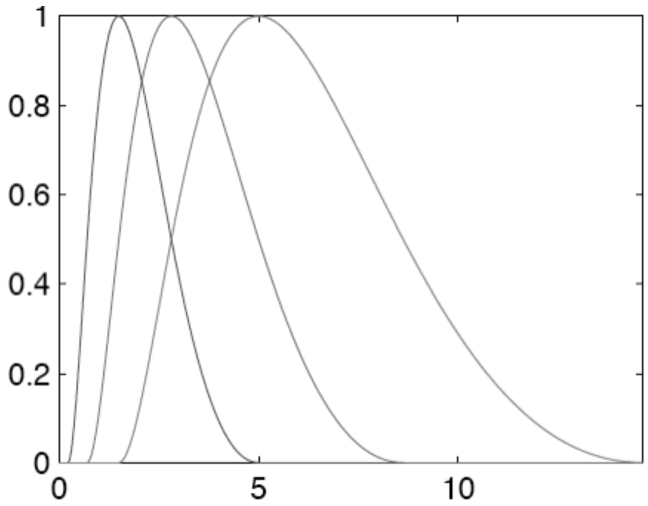
MODEL COMPLEXITY

Main challenge: too many parameters for too few data points

Solutions:

- **Assume space-time separability**
- **Use basis functions**



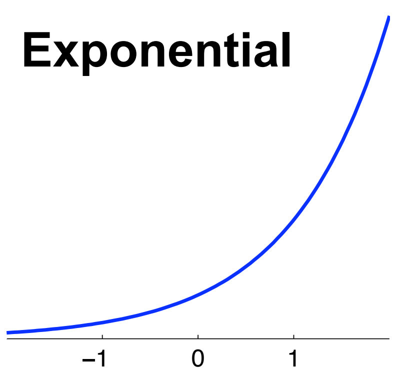


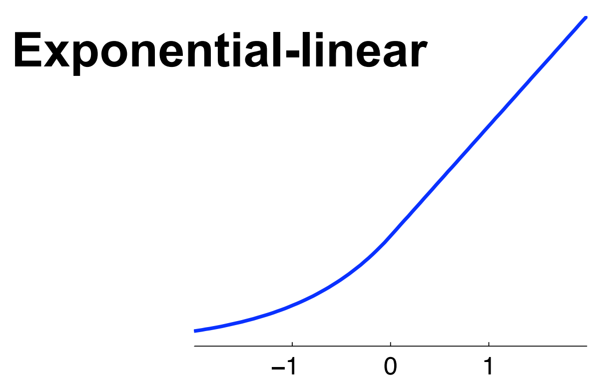
DIFFERENT NON-LINEARITIES?

$$r(t) = \exp(\lambda) \quad \overset{?}{\longrightarrow} \quad r(t) = f(\lambda)$$

Log-likelihood is convex if *f* is convex and log-concave
(i.e. grows at least linearly and at most exponentially)

→ Unique maximum





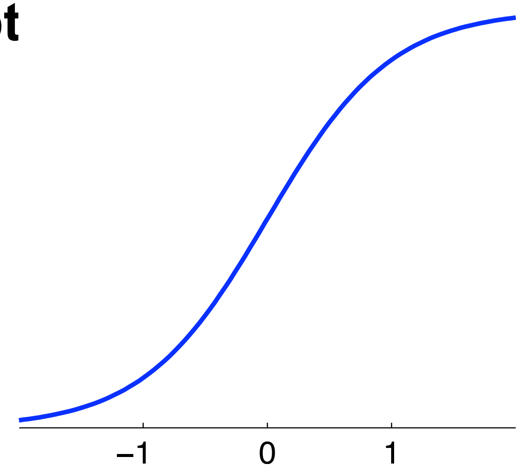
Paninski 2004

SIGMOIDAL NON-LINEARITY?

Problem is non-convex (local maxima) except in case of binary outputs

For each bin assume either 0 or 1 spike

→ Equivalent to logistic regression



Not a strong limitation: spike history can take care of saturation

LIMITATIONS OF GLMS AND SOME REMEDIES

Stimulus fiters are linear: can model only simple cells

Work in non-linear feature space (e.g. Gerwinn et al. 2007)

Linear history filters and temporal precision

Estimate non-linear interactions/ hidden neurons (e.g. Pillow & Latham 2008)

Correlations between cells are modeled as direct interactions

Model common input (e.g. Kulkarni & Paninski 2007, Vidne et al. 2009)

No state dependence

Model state dependence (e.g. Hidden Markov Models: Escola et al. 2011, Latent Linear Dynamical Systems: Macke et al. 2011)

Most extensions come at the cost of added computational cost and non-convexity!

SUMMARY: GLMS

Generalized Linear Models are

- flexible
- straightforward to use (convex, glmfit)
- well understood

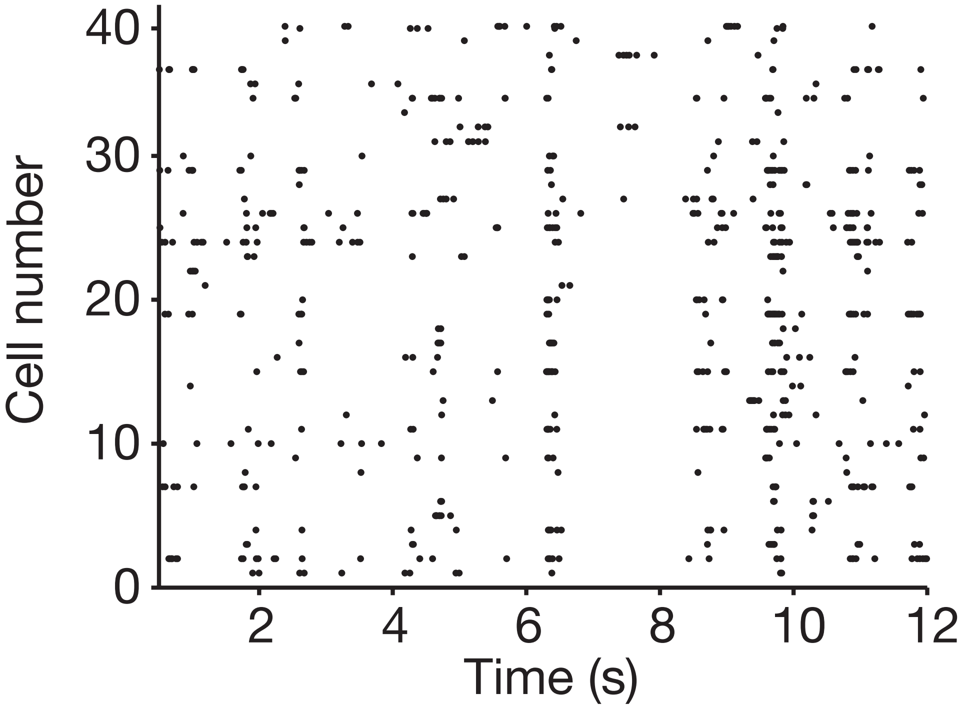
There are numerous extensions within the same framework

MAXIMUM ENTROPY MODELS

GOAL

Describe probability distribution over neural responses

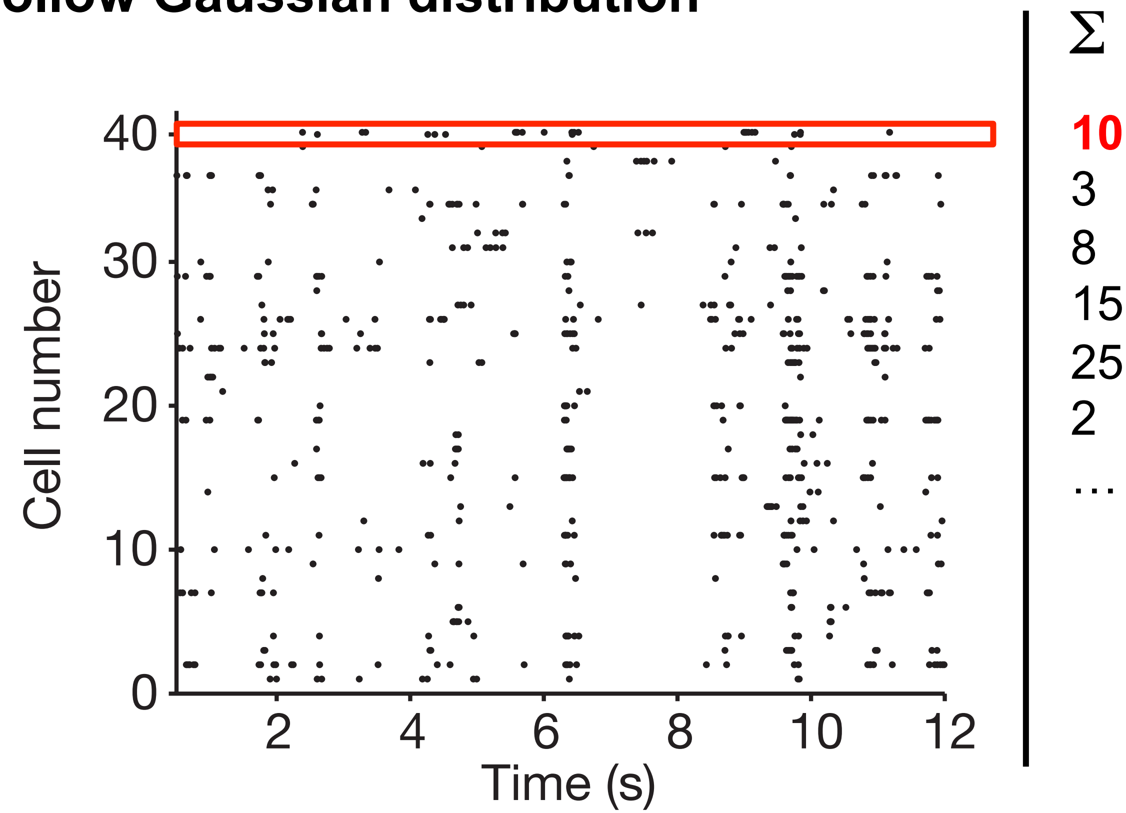
$$P(x_1, x_2, \cdots, x_n) = ?$$



A SIMPLE APPROACH

Count spikes over large window

Assume spike counts follow Gaussian distribution

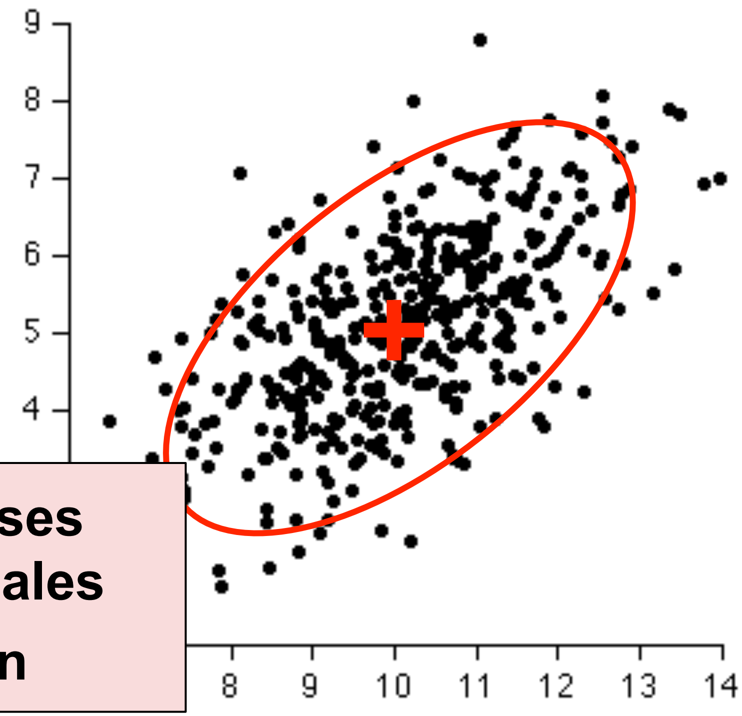


A SIMPLE APPROACH

Count spikes over large window

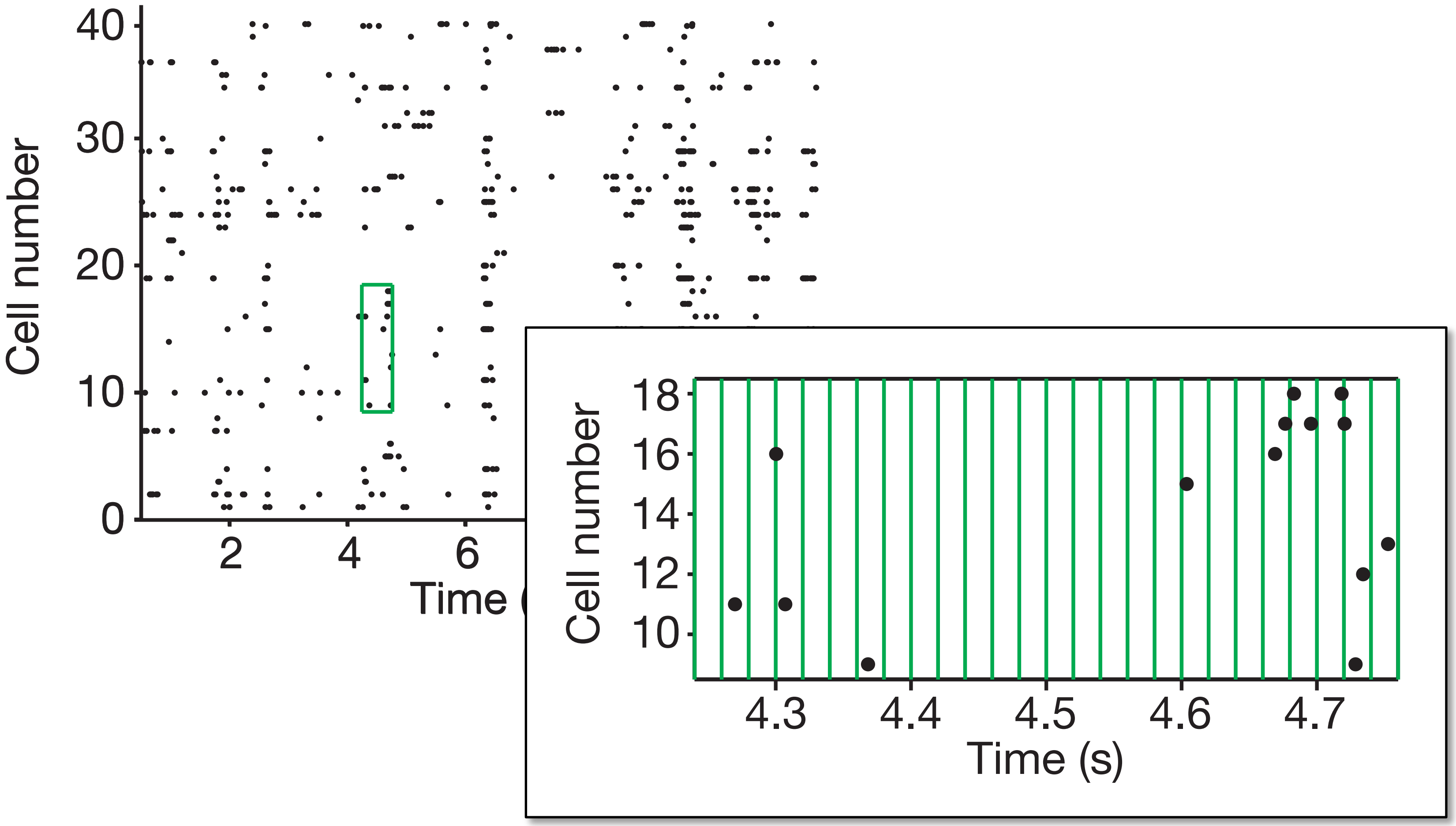
Assume spike counts follow Gaussian distribution

→ **Mean and Covariance fully describe the distribution**

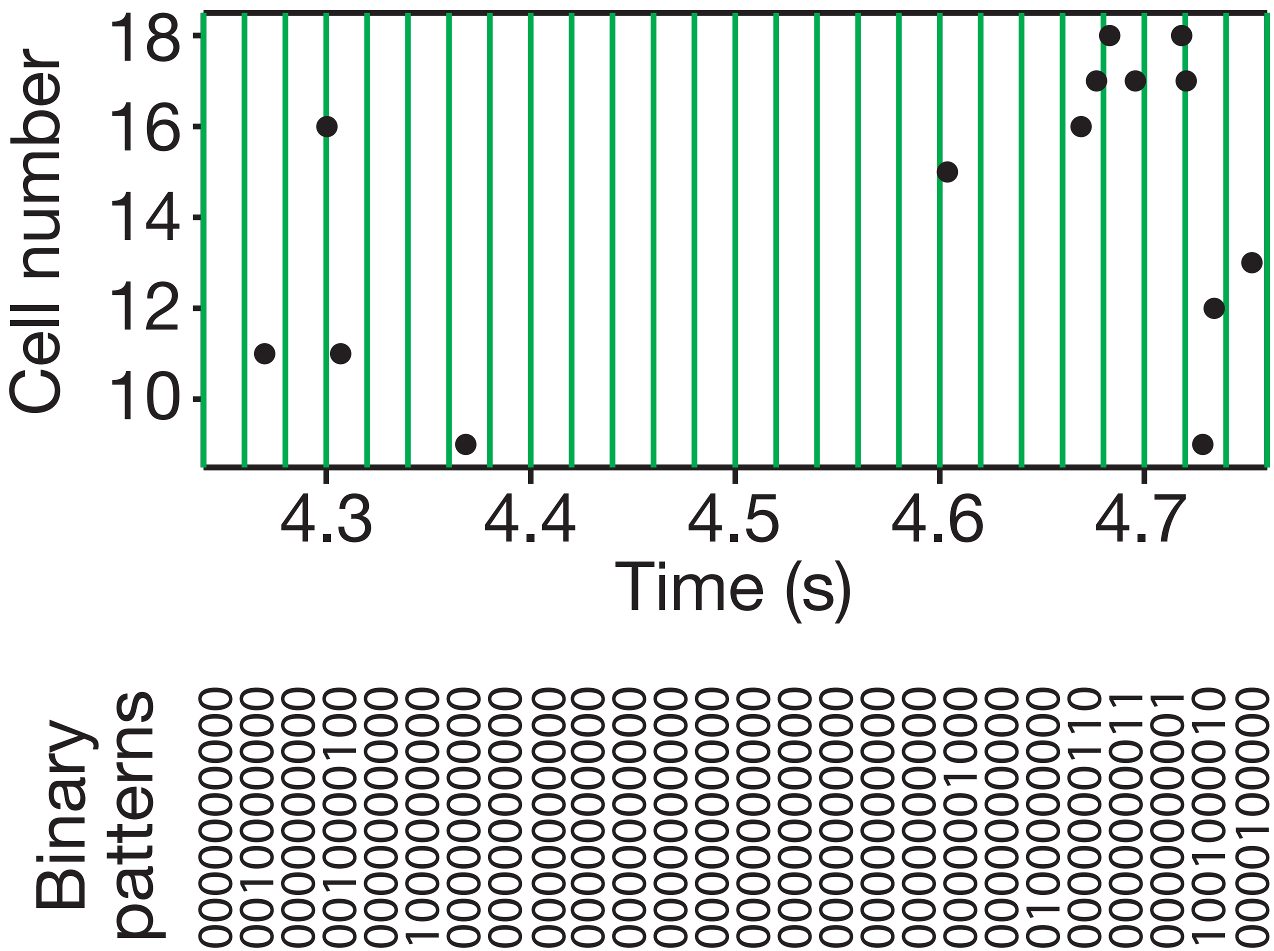


- **Most interesting brain processes happen at much faster timescales**
- **Spike counts are not Gaussian**

MODELING BINARY PATTERNS

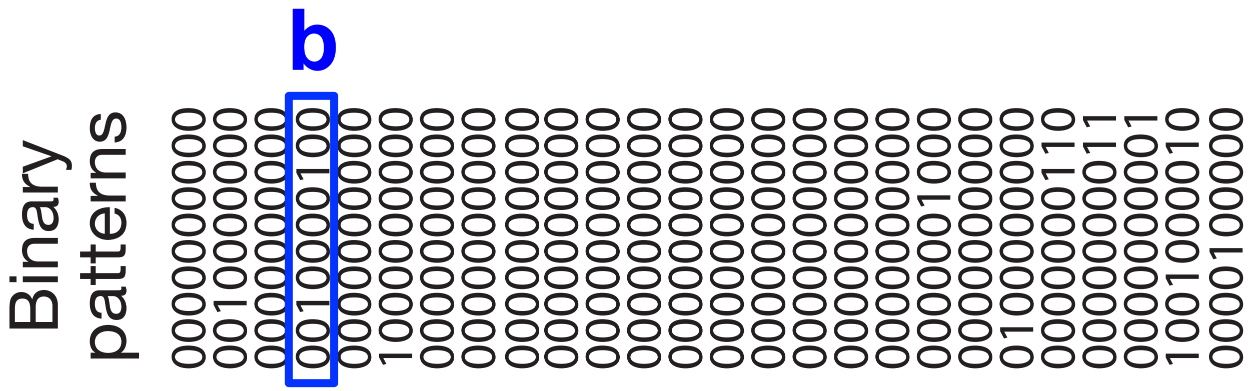


MODELING BINARY PATTERNS



MODELING BINARY PATTERNS

$$P(\mathbf{b}) = ?$$



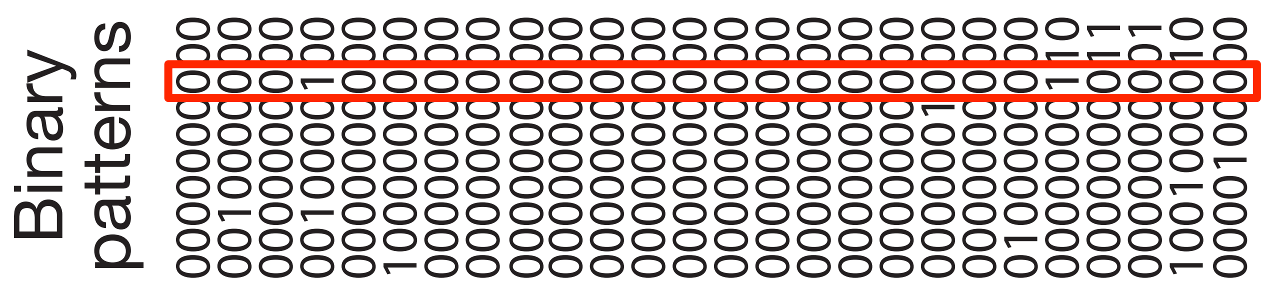
INDEPENDENT MODEL

$$P_1(\mathbf{b}) = \prod_{k=1}^{N} P_k(b_k)$$

$$P_k(b_k) = \lambda_k^{b_k}(1-\lambda_k)^{1-b_k}$$

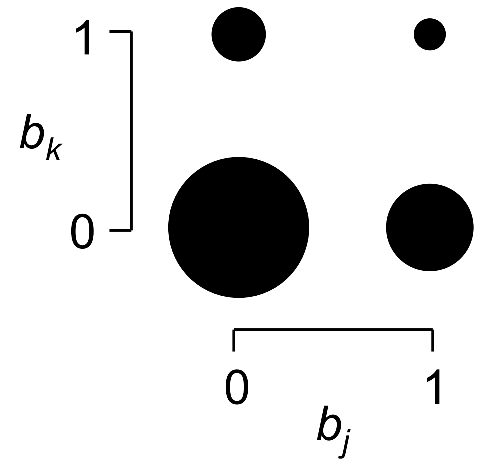
Bernoulli Dist. bk = 1 fire; bk = 0 not fire.

Simple, but does not capture dependencies between neurons



$$\hat{\lambda}_k = \frac{2}{26}$$

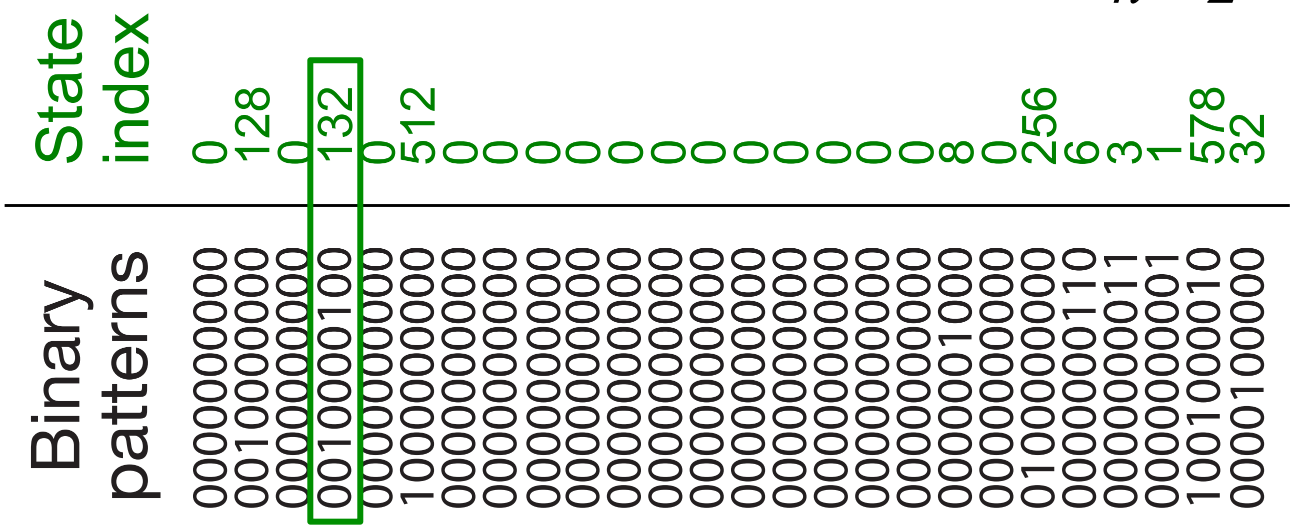
FULL MODEL



$$P_{\text{full}}(\mathbf{b}) \approx \frac{h_{\mathbf{b}}}{\sum_{\mathbf{b}} h_{\mathbf{b}}}$$

Captures all dependencies between neurons, but is too complex

$$n = 2^{10} = 1024 \text{ states}$$



MAXIMUM ENTROPY PRINCIPLE

Assume we know some statistics of the activity patterns

(e.g. mean, correlations, population spike count, etc.)

What is the distribution that satisfies these constraints but imposes no additional structure?

$$P(\mathbf{b}) = \frac{1}{Z}\exp\left(-\sum_k \lambda_k f_k(\mathbf{b})\right)$$

$f_k(\mathbf{b})$**:** *statistic* **(function of b whose expected value we know)**

MAXIMUM ENTROPY DISTRIBUTION

Some example statistics:

some limitation (statistics) we know about b that can be used as constraint of choice of model.

$$f_k(\mathbf{b}) = \mathbf{b}$$ **Mean**

$$f_k(\mathbf{b}) = (\mathbf{b} - \langle \mathbf{b} \rangle)^2$$ **Covariance**

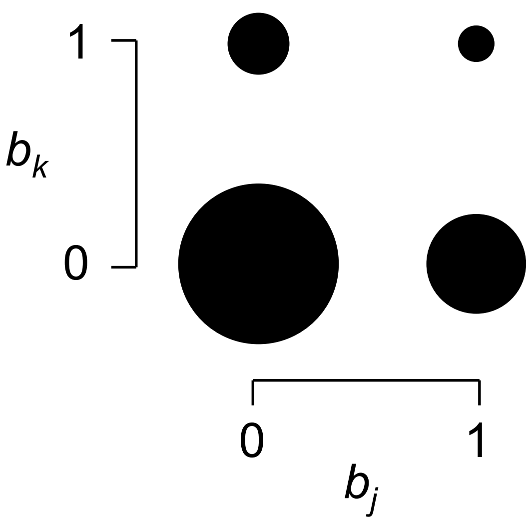
$$f_k(\mathbf{b}) = \Sigma_k b_k$$ **Population spike count**

PAIRWISE MODEL

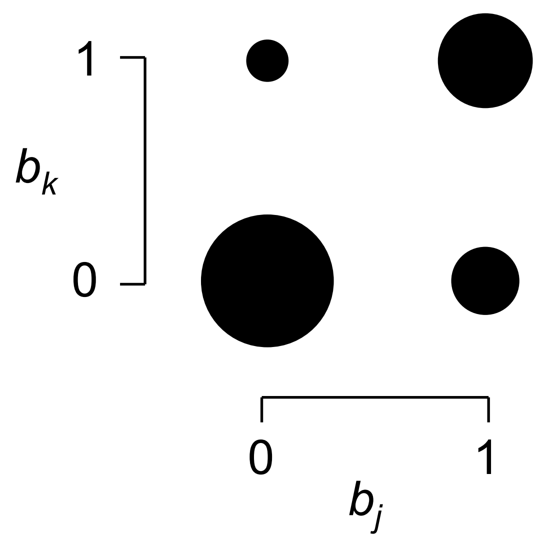
Pairwise correlations:

$$\rho_{jk} = \frac{\mathrm{E}[b_j b_k] - \mathrm{E}[b_j]\mathrm{E}[b_k]}{\sqrt{\mathrm{Var}[b_j]\mathrm{Var}[b_k]}}$$

$\rho_{jk} = 0$



$\rho_{jk} > 0$



PAIRWISE MODEL

Second-order MaxEnt model (a.k.a. *Ising Model*):

$$P_2(\mathbf{b}) = \frac{1}{Z}\exp(-\mathbf{h}^{\mathrm{T}}\mathbf{b} - \mathbf{b}^{\mathrm{T}}J\mathbf{b})$$
$$= \frac{1}{Z}\exp\left(-\sum h_k b_k - \sum J_{jk} b_j b_k\right)$$

Number of parameters: $n = \dfrac{N(N+1)}{2} = 65$

LEARNING MAXENT MODELS

Ising model: Boltzmann-Machine learning

No closed form for maximum likelihood solution

$$\frac{\partial L}{\partial J_{jk}} = \langle b_j b_k \rangle_{\mathrm{Data}} - \langle b_j b_k \rangle_{\mathrm{Model}}$$

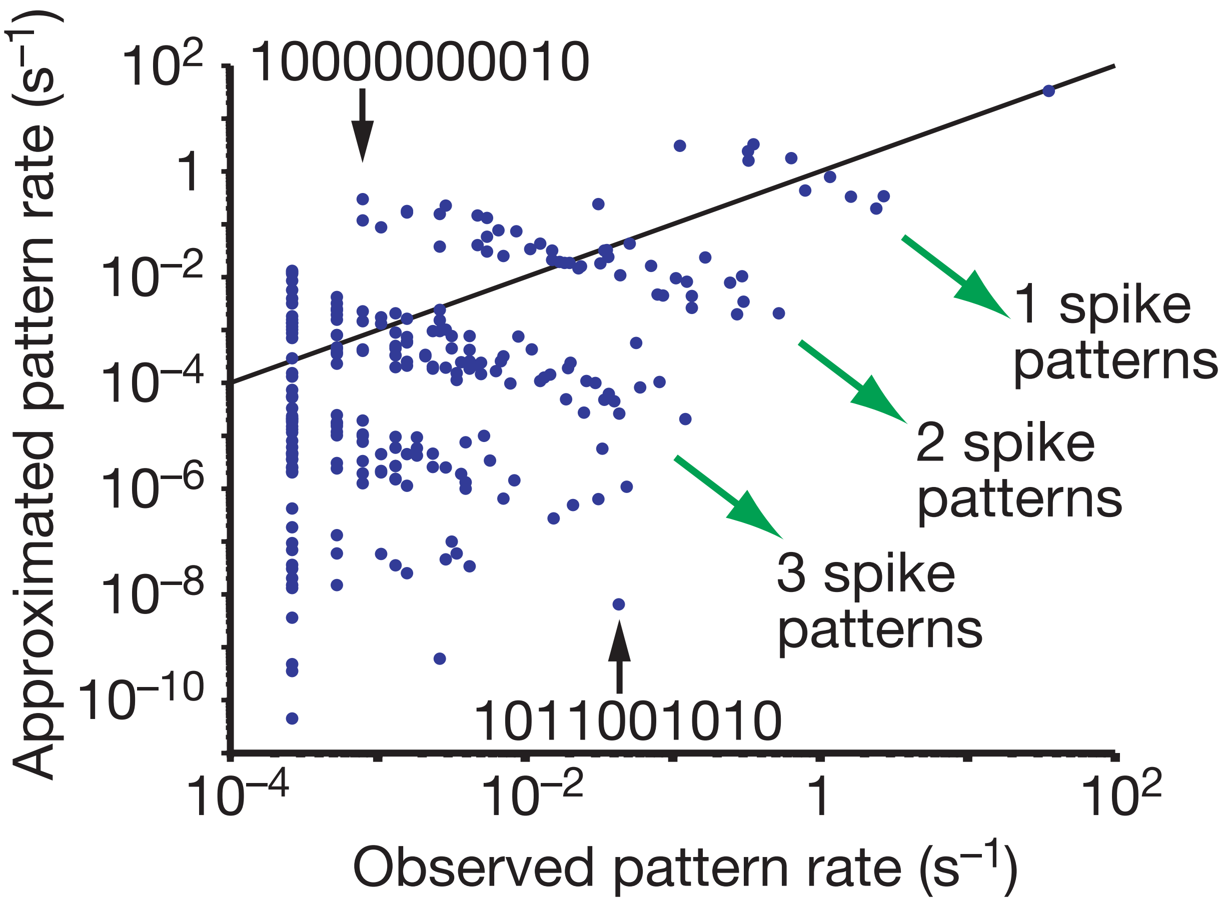
Need to sample from model
→ Gibbs sampling

$$P_2(\mathbf{b}) = \boxed{\frac{1}{Z}} \exp(-\mathbf{h}^{\mathrm{T}}\mathbf{b} - \mathbf{b}^{\mathrm{T}} J \mathbf{b})$$

e.g. Bethge & Berens 2008

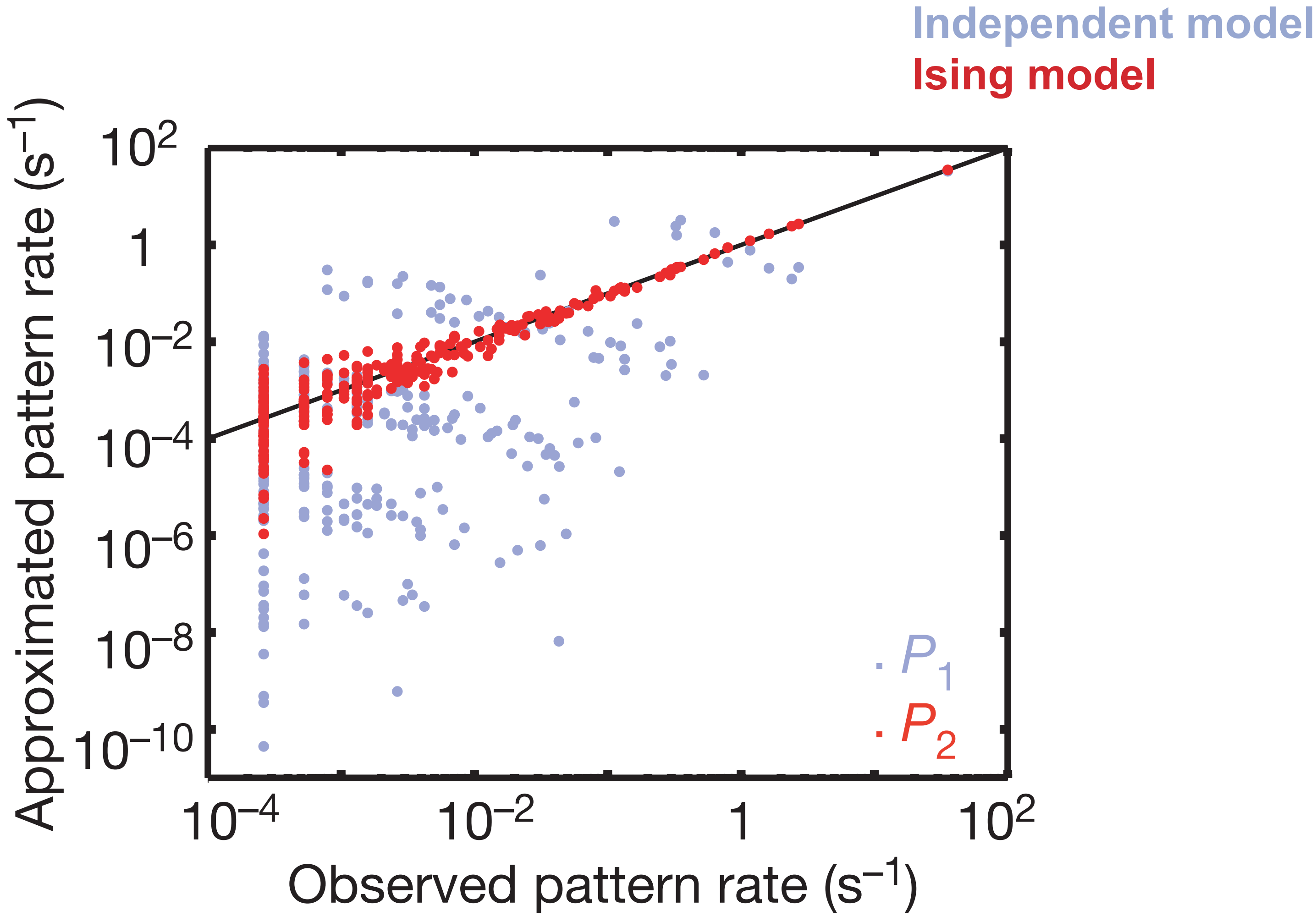
EXAMPLE: ISING MODEL IN THE RETINA

Independent model



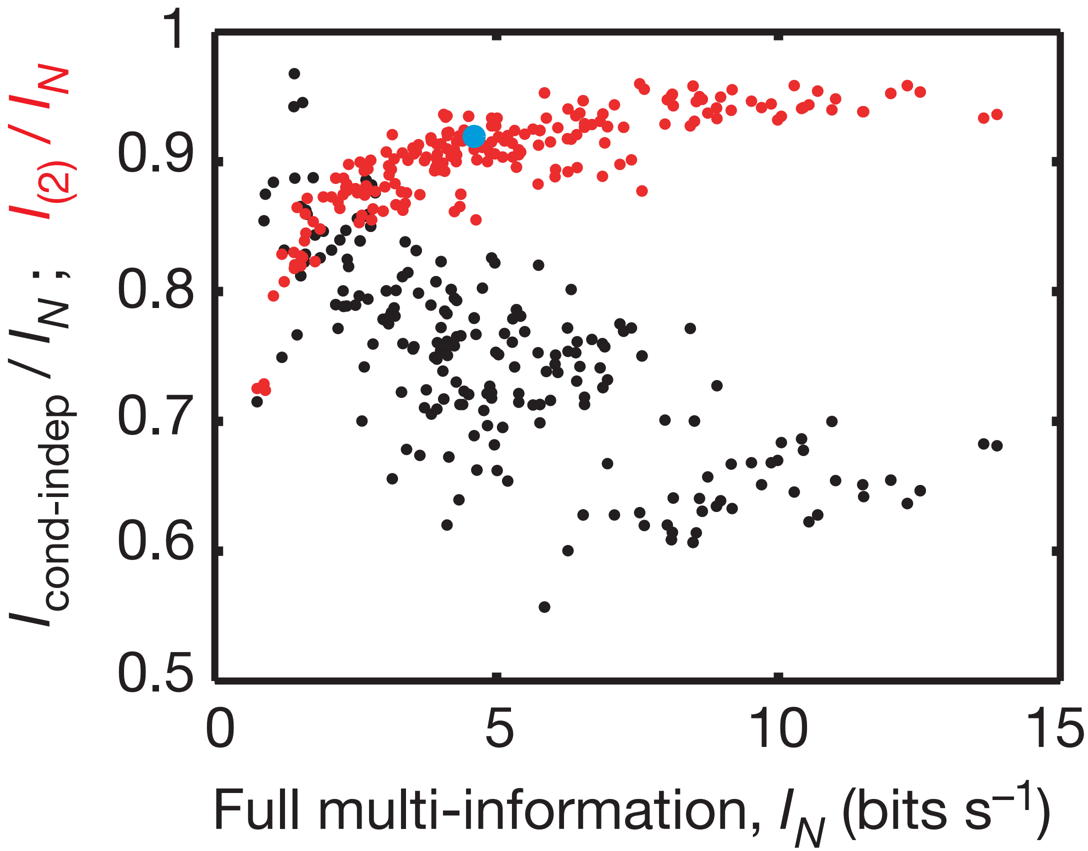
Schneidman et al. 2006

EXAMPLE: ISING MODEL IN THE RETINA



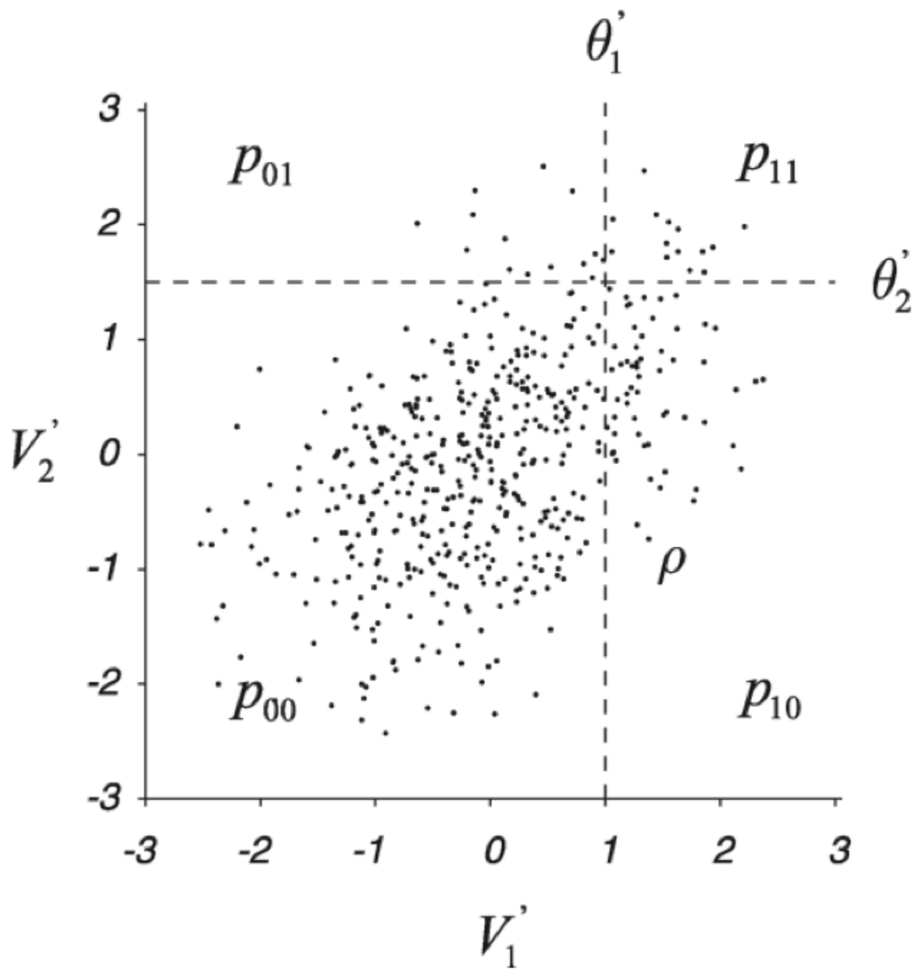
Schneidman et al. 2006

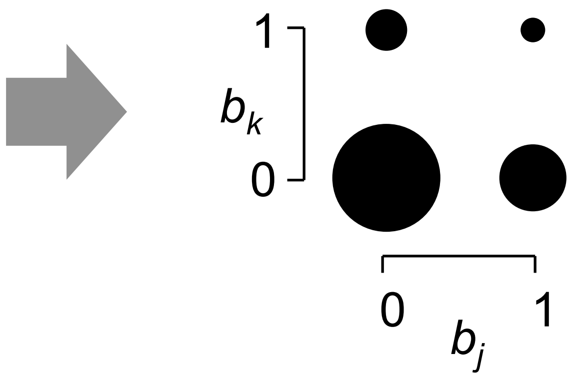
EXAMPLE: ISING MODEL IN THE RETINA



Schneidman et al. 2006

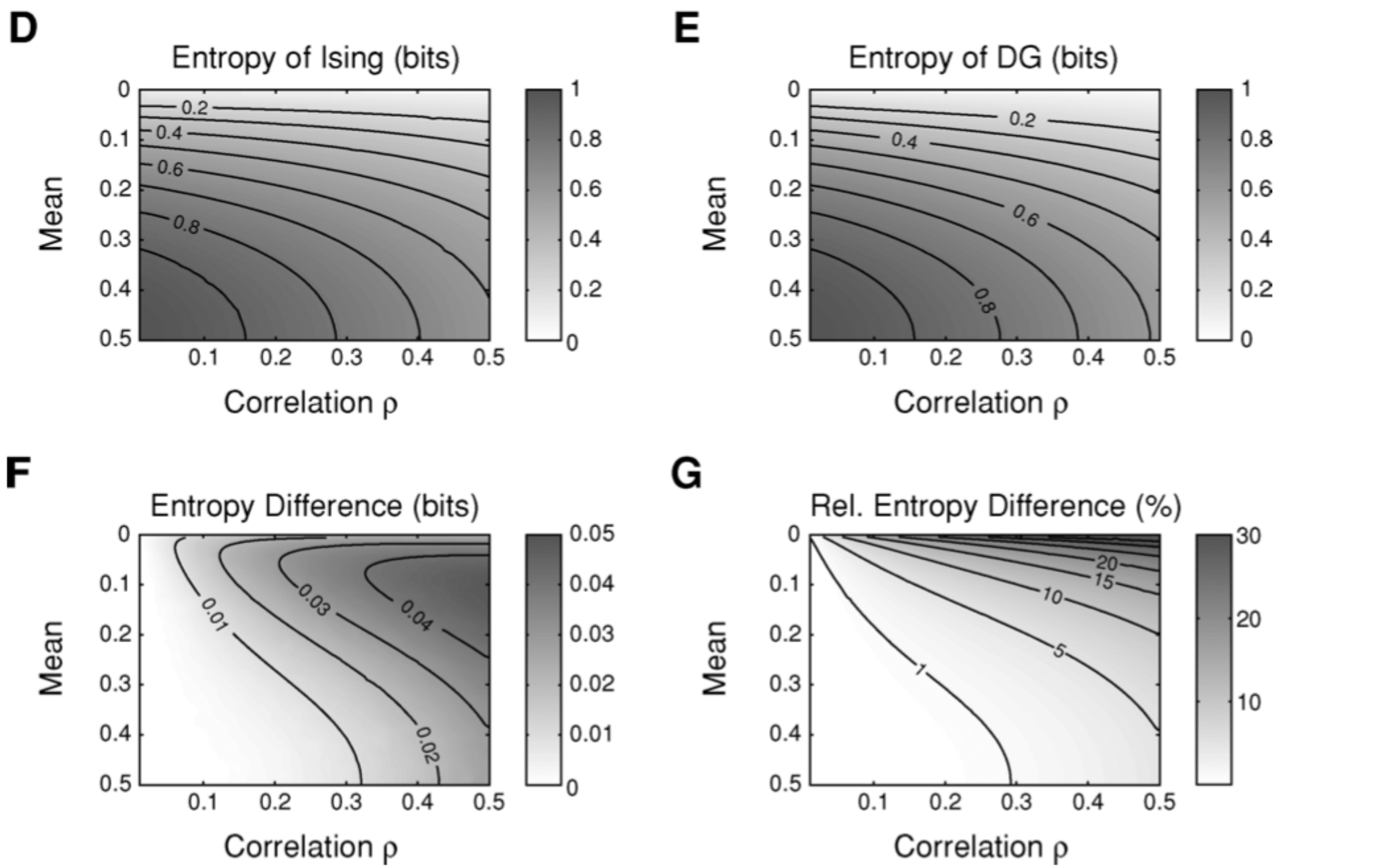
ALTERNATIVE: DICHOTOMIZED GAUSSIAN (DG)





From: Dorn & Ringach 2003

ENTROPY: ISING VS. DG



Macke, Berens, Ecker, Tolias, Bethge 2009

ADVANTAGES OF DG

Much easier to fit than MaxEnt/Ising model

Close to maximum entropy
(not much more structure than necessary)

Can also be interpreted as thresholded membrane potential
(see Dorn & Ringach 2003)

Can be generalized to spike counts > 0/1
(see Macke et al. 2009)

Disadvantages

Binary distributions exist for which no DG model exists

No simple way of computing likelihoods