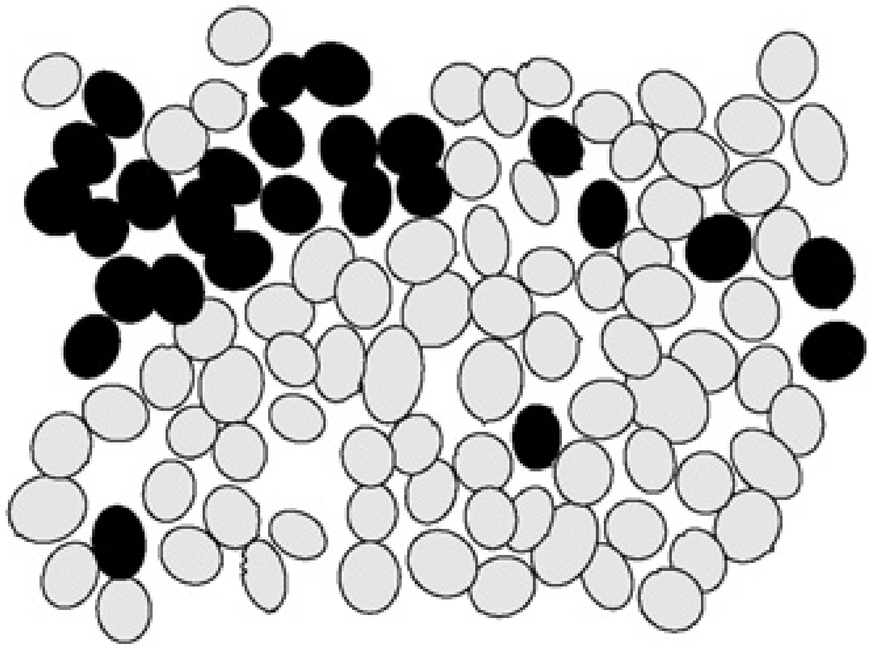
When are microcircuits well-modeled by pairwise maximum entropy methods?

Andrea K. Barreiro

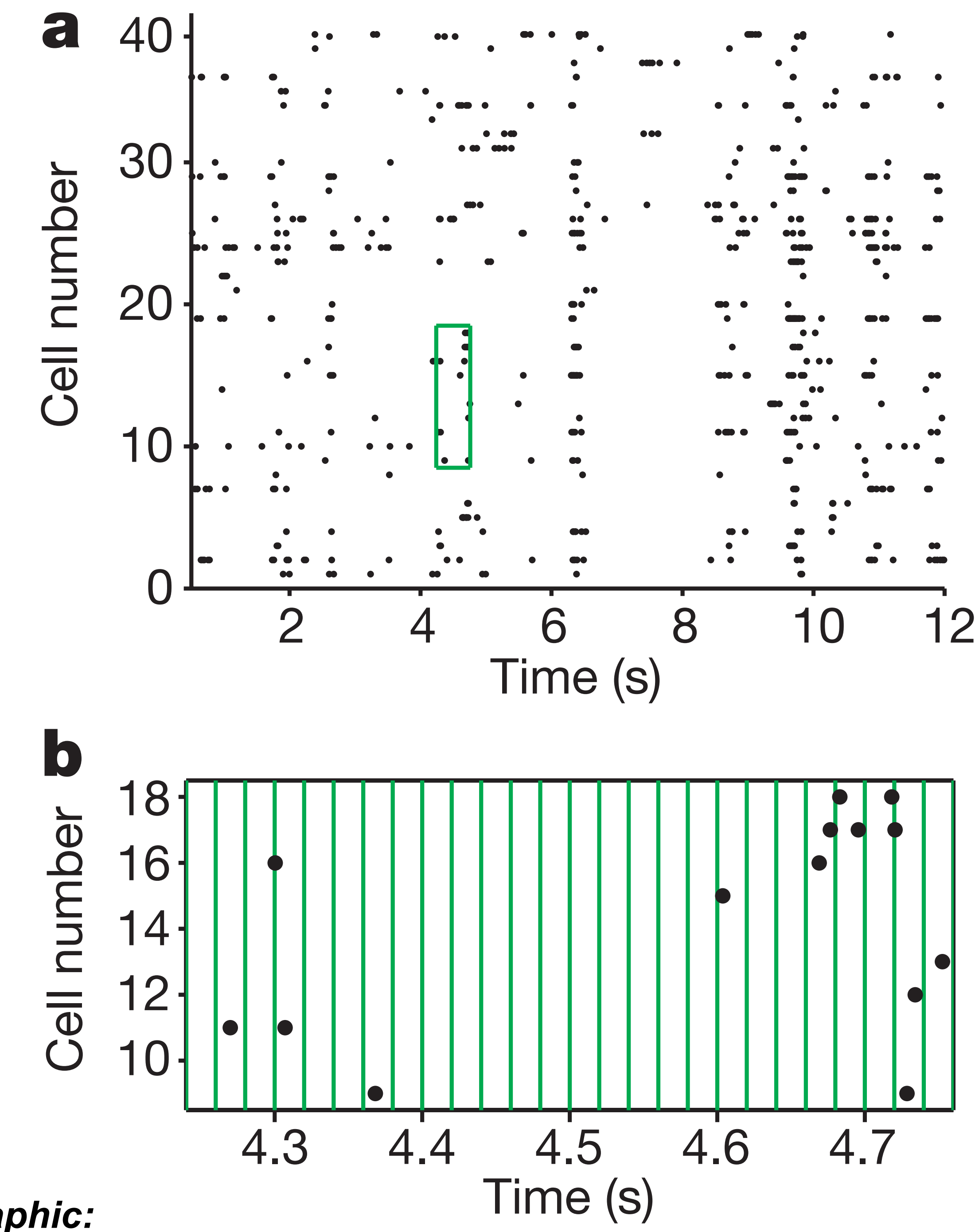
Department of Applied Mathematics
University of Washington

BIRS, October 2010

How we describe spiking activity

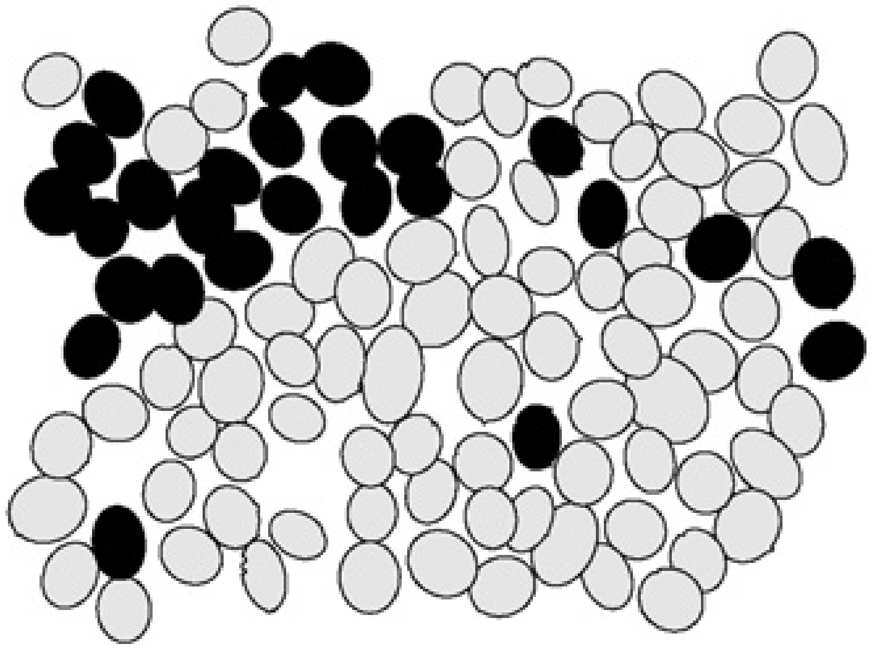


Graphic:
Shlens, Rieke and Chichilnisky, 2008

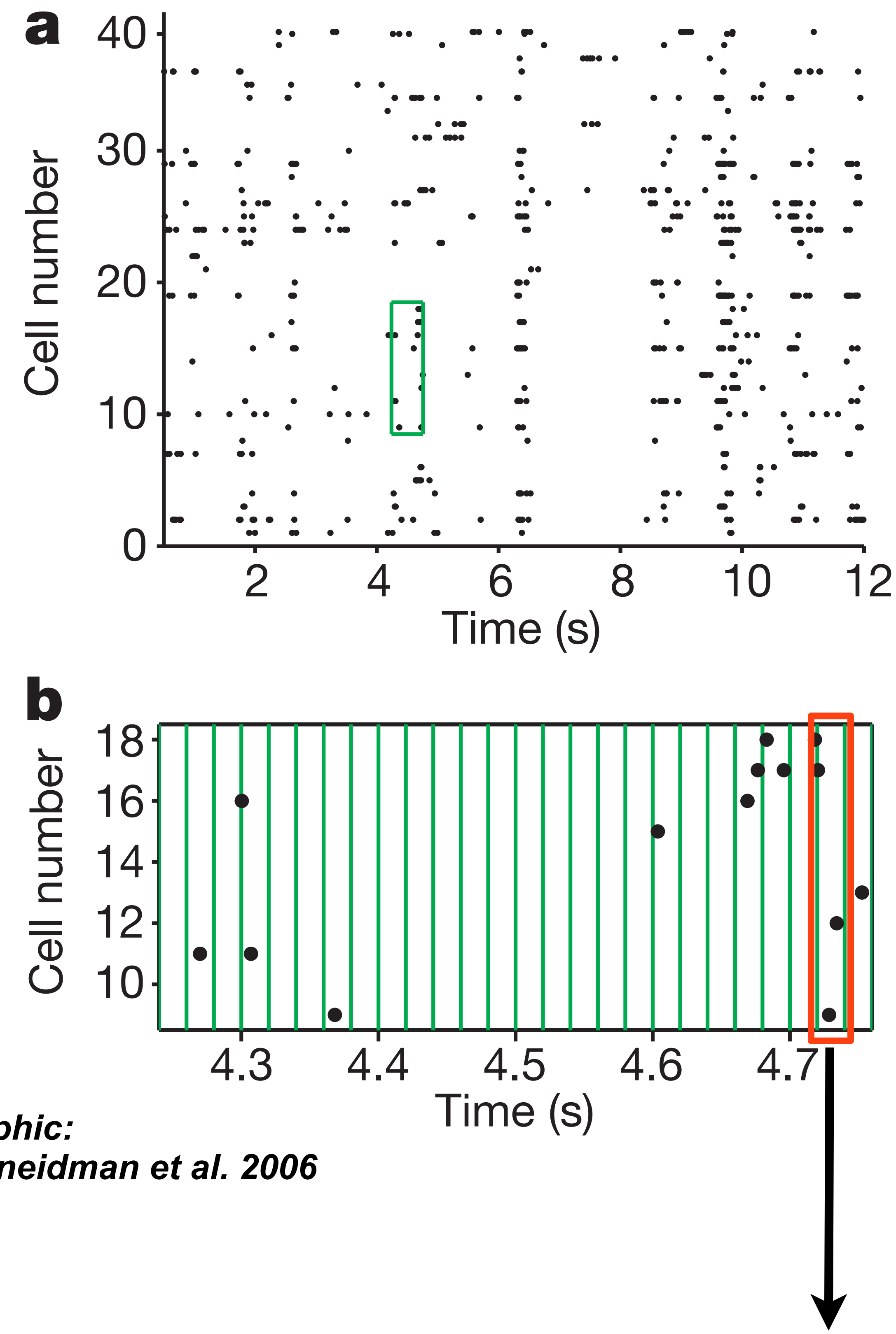


Graphic:
Schneidman et al. 2006

How we describe spiking activity



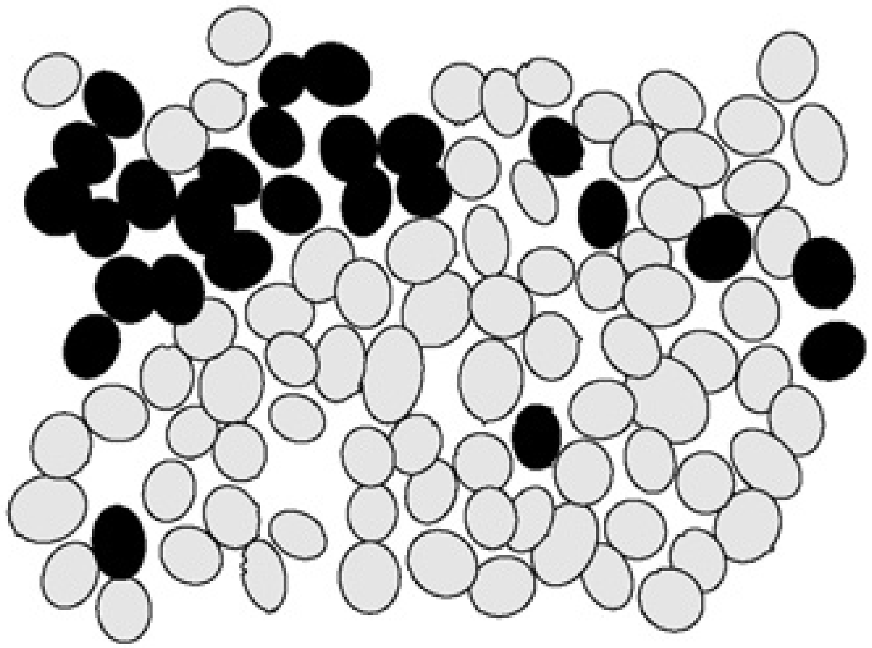
Graphic:
Shlens, Rieke and Chichilnisky, 2008



Graphic:
Schneidman et al. 2006

1001000010

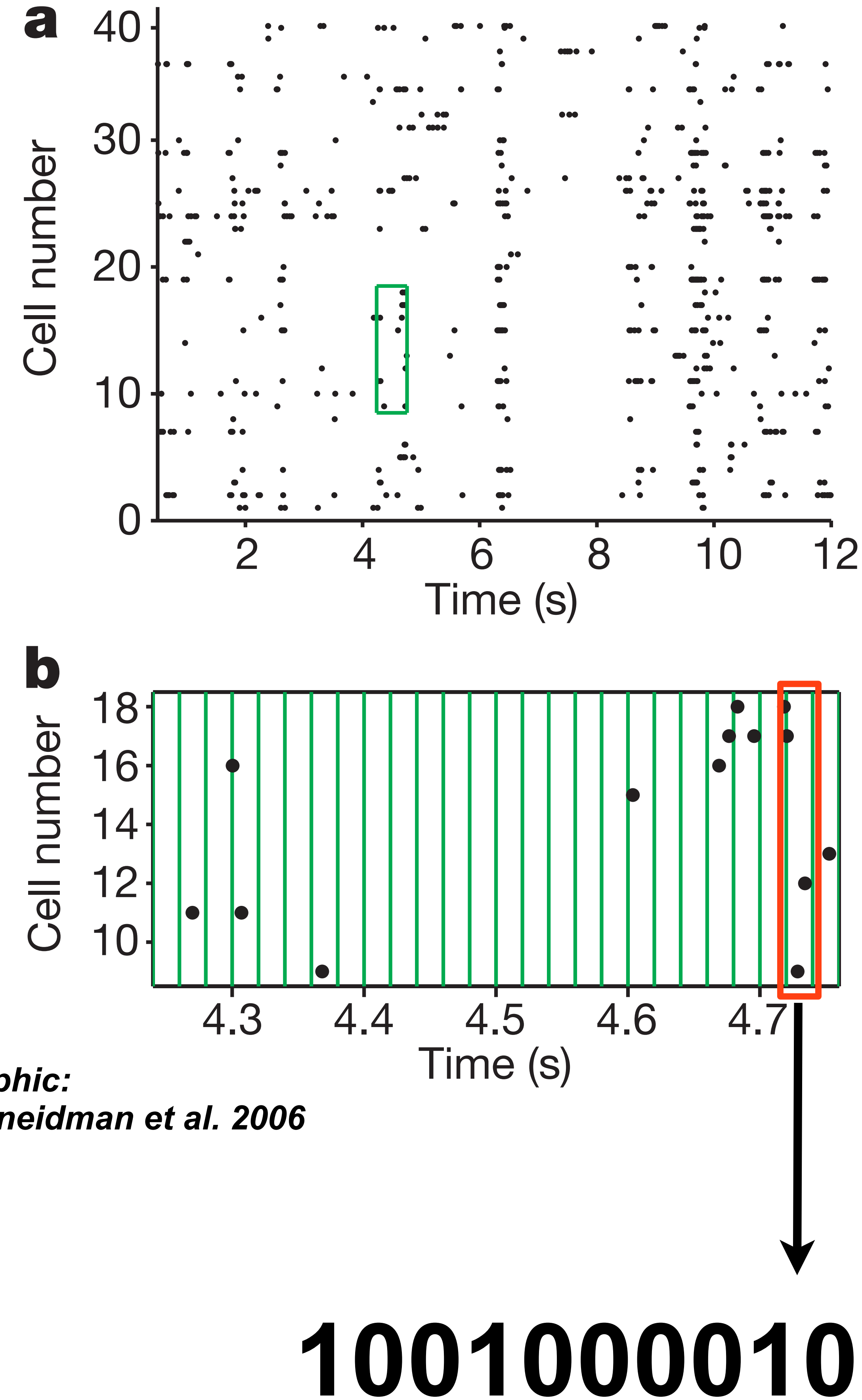
How we describe spiking activity



Graphic:
Shlens, Rieke and Chichilnisky, 2008

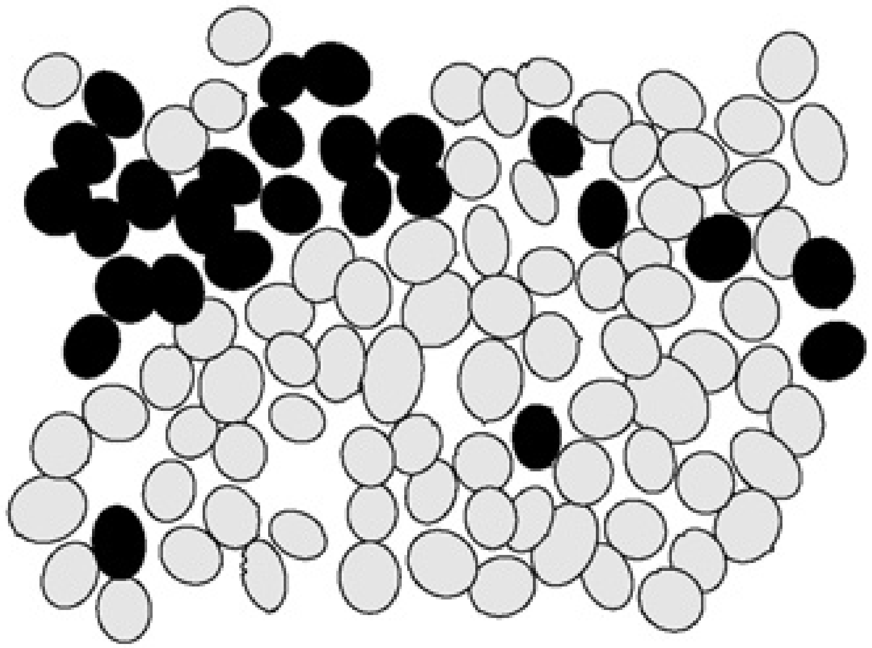
$$x_j = \{0,1\}$$

$$P(x_1, x_2, \cdots, x_N)$$



Graphic:
Schneidman et al. 2006

Retinal ganglion cells (RGCs) do not fire independently

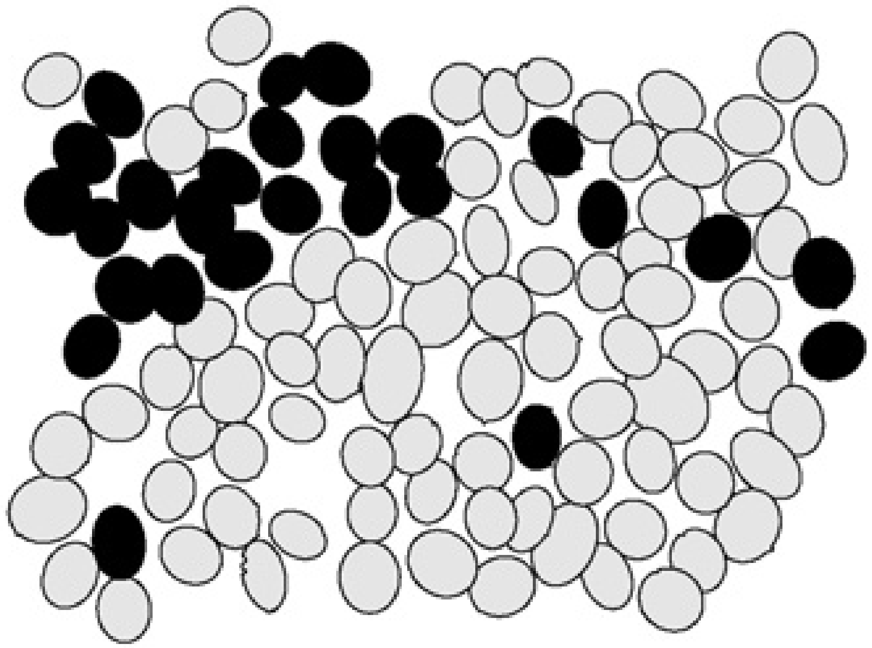


Graphic:
Shlens, Rieke and Chichilnisky, 2008

$$x_j = \{0,1\}$$

$$P(x_1, x_2, \cdots, x_N) \neq P(x_1)P(x_2)\cdots P(x_N)$$

Retinal ganglion cells (RGCs) do not fire independently



Graphic:
Shlens, Rieke and Chichilnisky, 2008

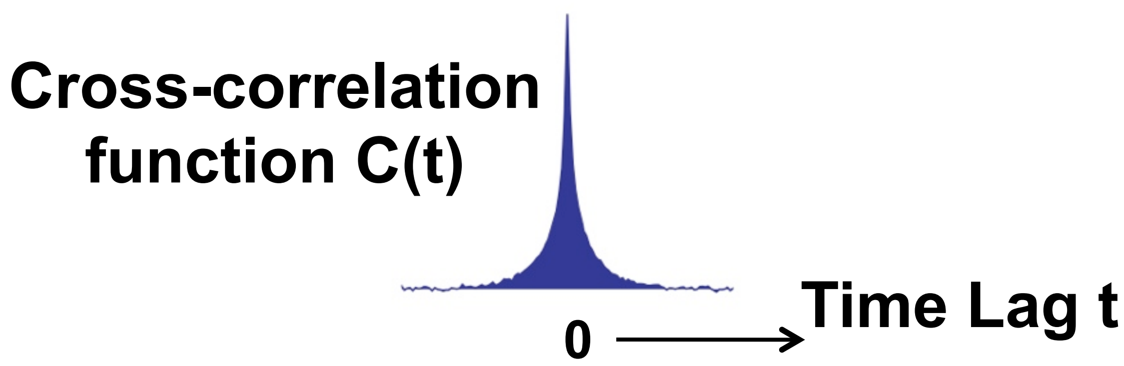
$$x_j = \{0,1\}$$

$$P(x_1, x_2, \cdots, x_N) \neq P(x_1)P(x_2)\cdots P(x_N)$$

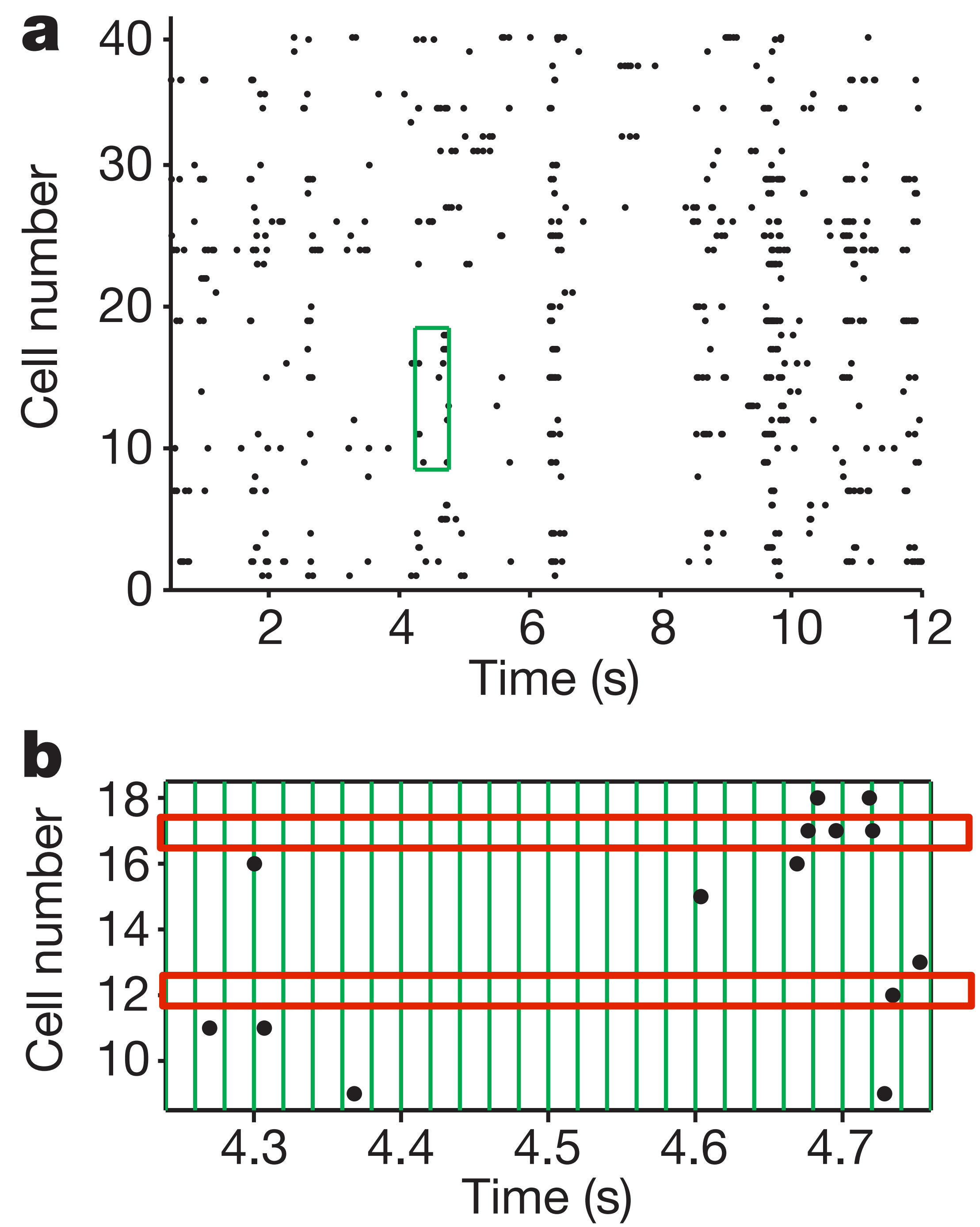
Time to observe event under independent assumption: 100 days

Actual time to observe event: 1 minute

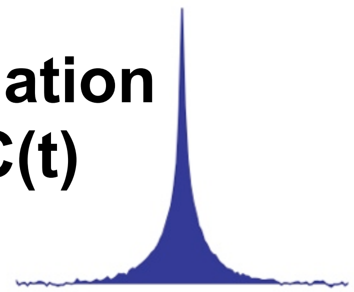
How we describe spiking activity... with a *pairwise* model

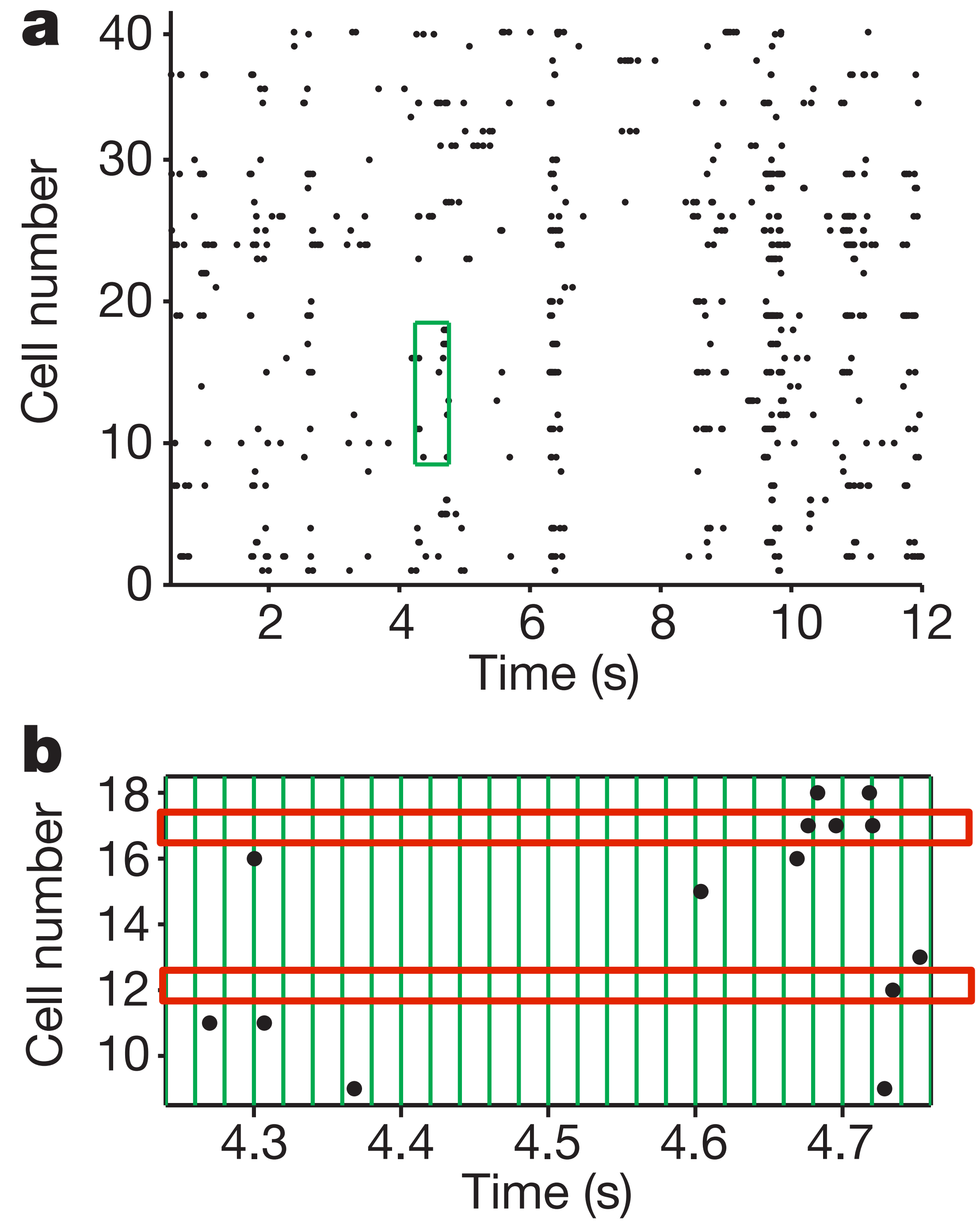


$$C_{ij}(0) = P(x_i x_j) - P(x_i)P(x_j)$$



How we describe spiking activity... with a *pairwise* model





$$C_{ij}(0) = P(x_i x_j) - P(x_i)P(x_j)$$

- # of parameters needed to specify *pairwise* information: i.e. firing rates and cross-correlations: ~N^2 ($100^2 \sim 10^4$) (*vs.* 2^N (10^{30}))

How to specify pairwise model: Maximum entropy

Suppose we have a distribution, $P(x_1,\ldots,x_n)$**, with moments**

$$E[x_i] = \mu_i$$ **(firing rate)**

$$E[x_i x_j] = \sigma_{ij}$$ **(covariance)**

Find, among distributions consistent with these moments, the one with maximal entropy $H(P) = -\sum_{\{\vec{x}\in S\}} P(\vec{x}) \log P(\vec{x})$

Then we know

$$P_2 = \frac{1}{Z}\exp\left(\sum_i \lambda_i x_i + \sum_{i,j} \lambda_{ij} x_i x_j\right)$$

(equivalent to Ising model)

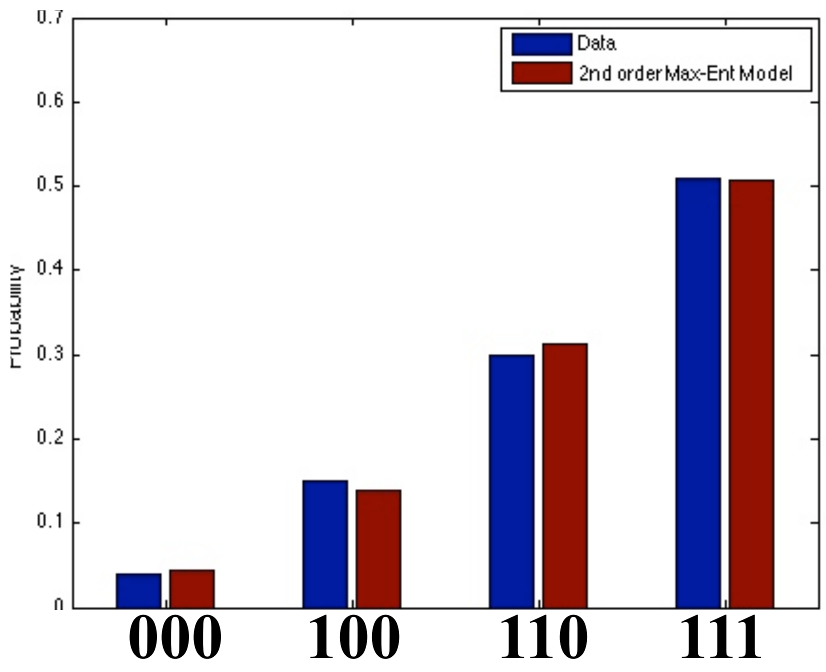
How to quantify higher order correlations?

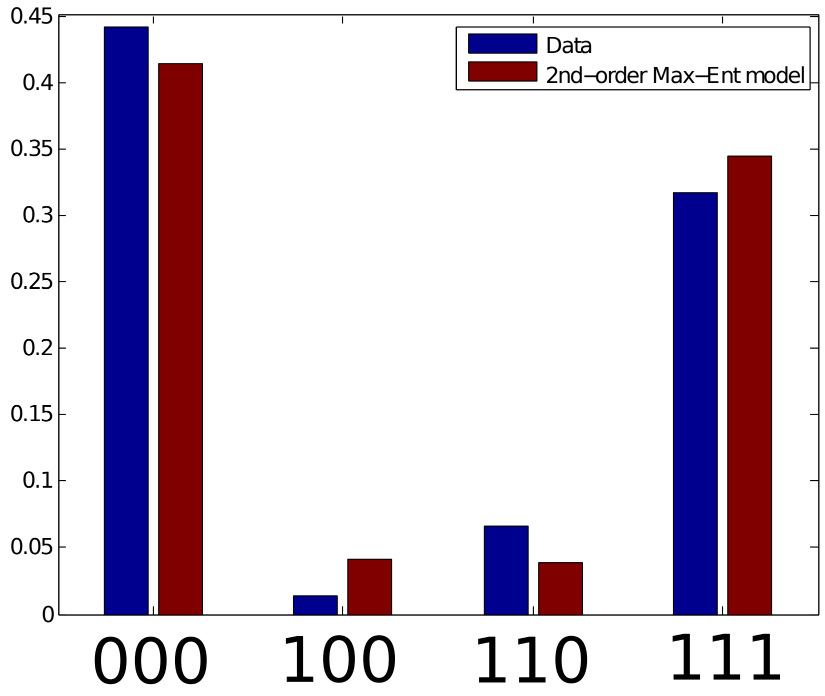
1) Given *P*, find pairwise maximum entropy fit P_2

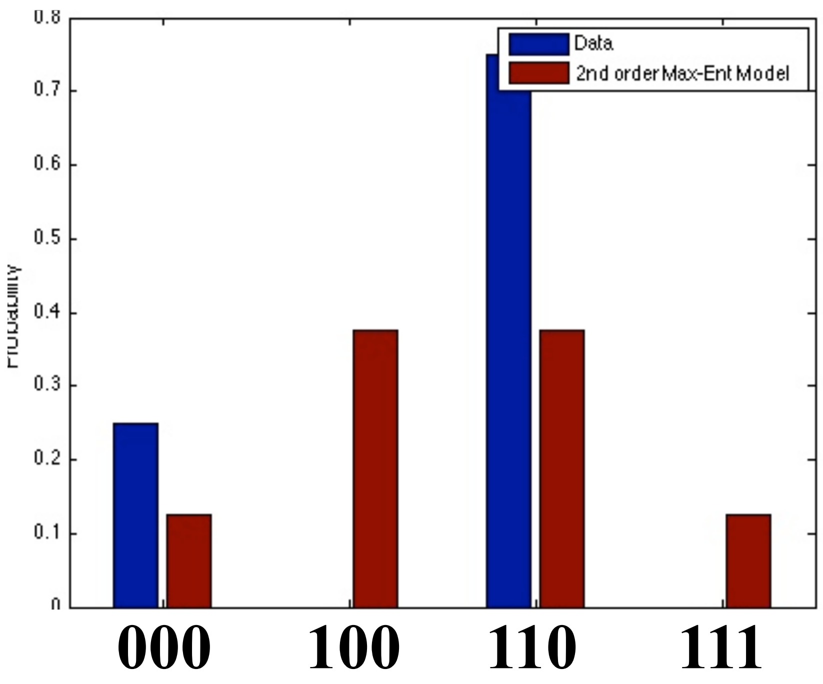
2) Compute distance between *P*, P_2 using Kullback-Leibler divergence $D_{KL}(P, P_2)$

H_2 = entropy of P_2
H_N = entropy of P

$$D_{KL}(P, P_2) = H_2 - H_N$$





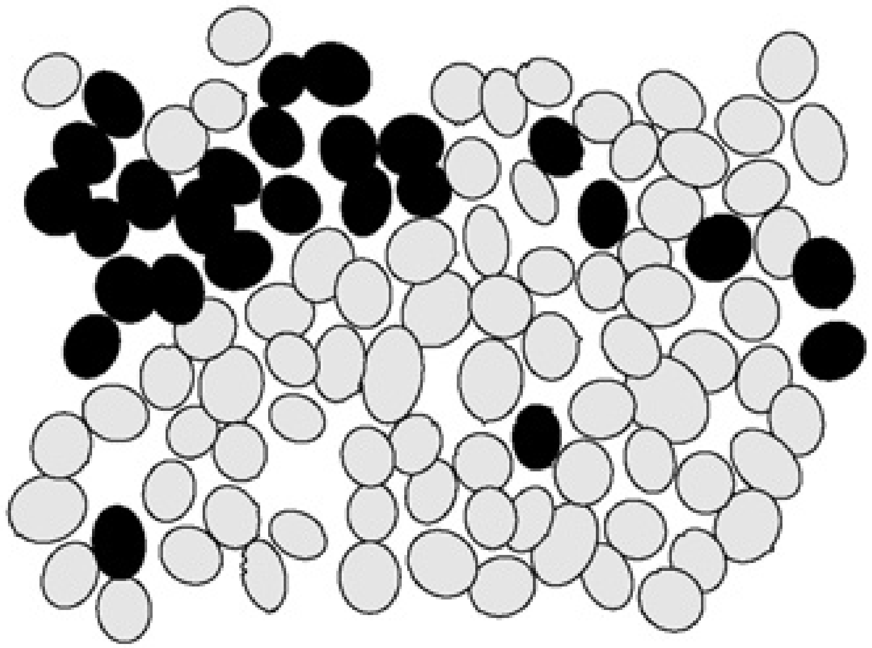


H_2-H_N=0.0013

H_2-H_N=0.0908

H_2-H_N=1

Retinal ganglion cells (RGCs) are well modeled with pairwise maximum entropy model (PME)



$$x_j = \{0,1\}$$

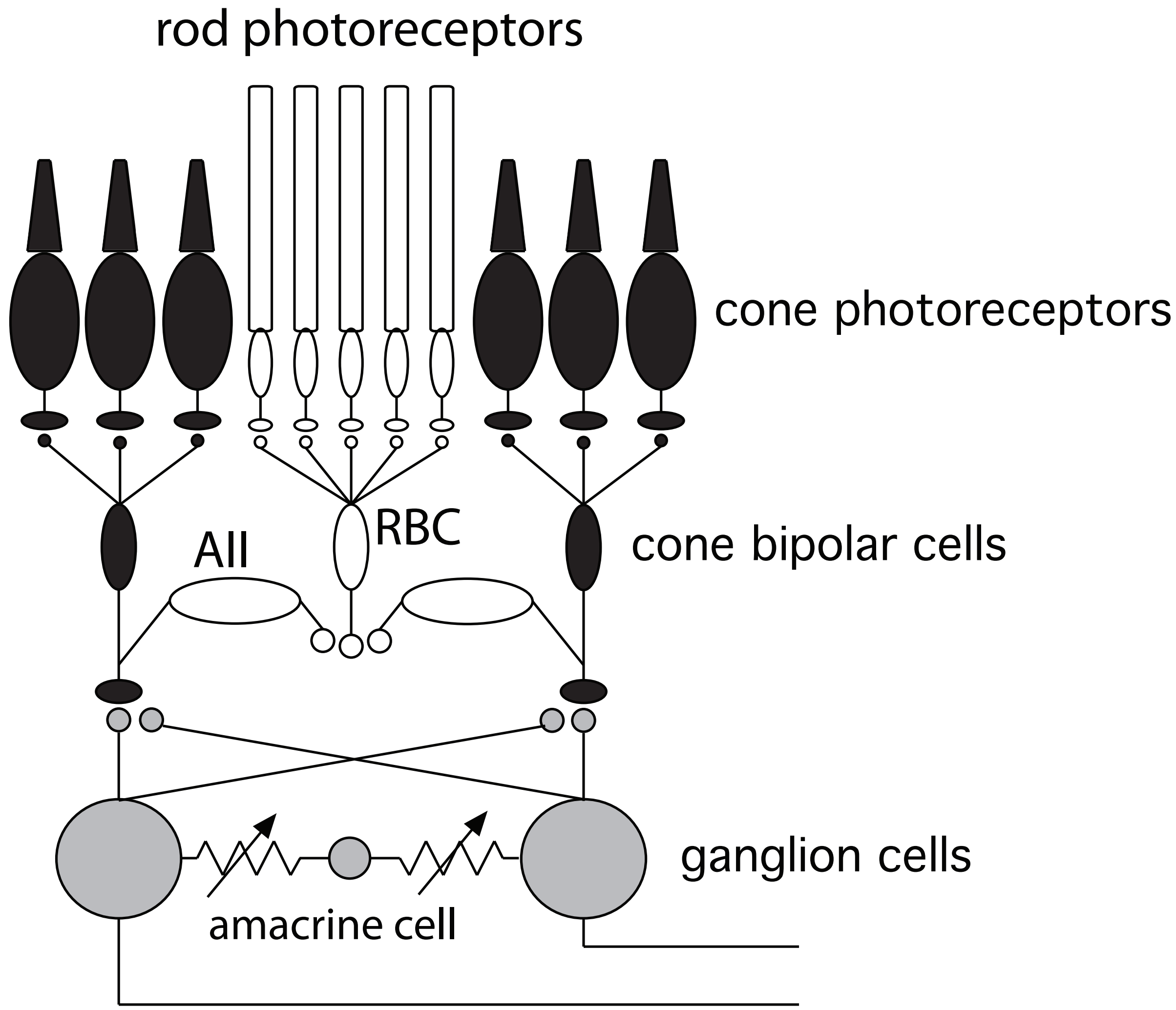
$$P(x_1, x_2, \cdots, x_N) \approx \frac{1}{Z} \exp\left(\sum_i \lambda_i x_i + \sum_{i,j} \lambda_{ij} x_i x_j \right)$$

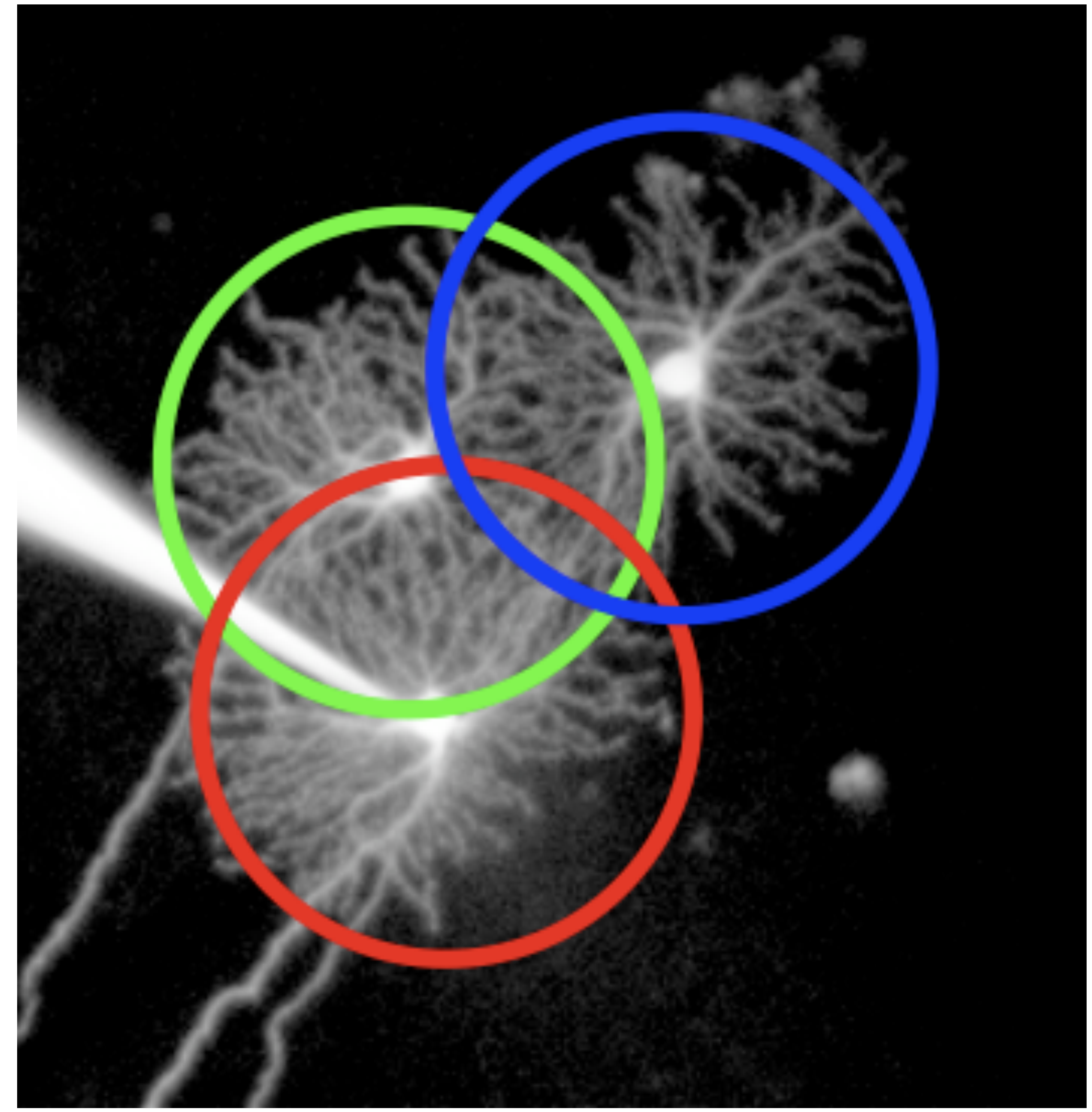
Graphic:
Shlens, Rieke and Chichilnisky, 2008

	D_{KL} (bits per neuron)
Shlens et al. 2006,	***1.62×10^{-4}***
Shlens et al. 2009,	***$1.30\text{-}1.74 \times 10^{-4}$***
Schneidman et al. 2006;	***$0.3\text{-}3 \times 10^{-4}$***

(contrast cortex (Montani et al. 2009, Martingnon 2000, Oizumi et al. 2010, Ohiorhenuan et al. 2010, Tang et al. 2008, Spacek and Swindale (unpublished)))

Retinal ganglion cells share common input

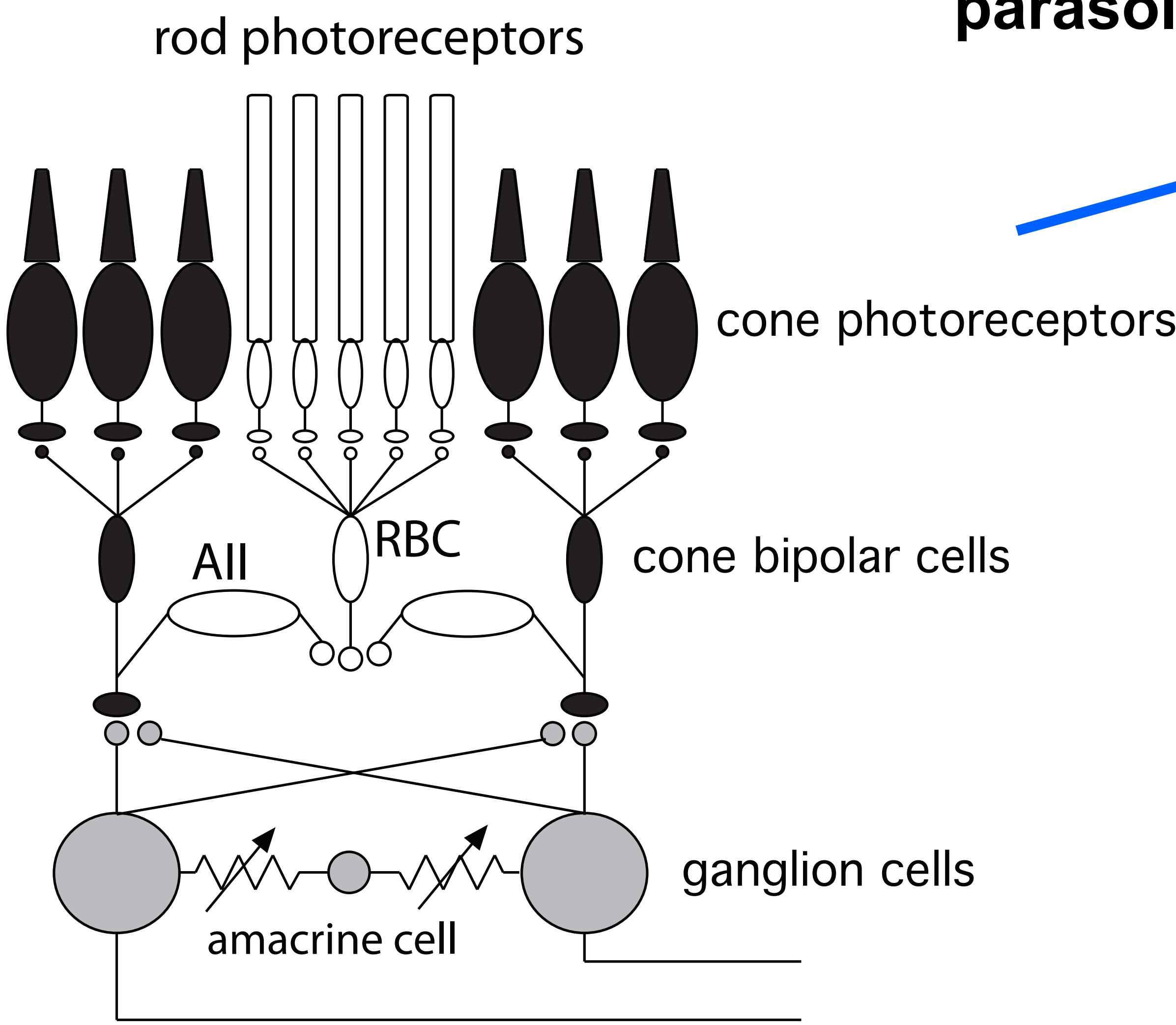




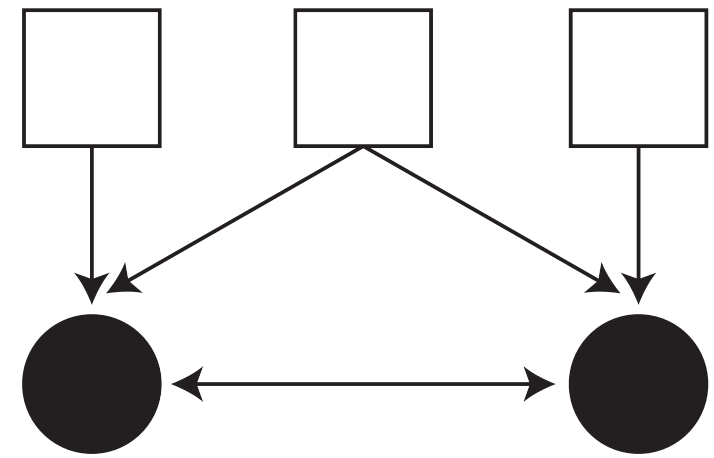
Graphic: Rieke lab

…input is shared among > 2 cells,
so where are the higher order correlations?

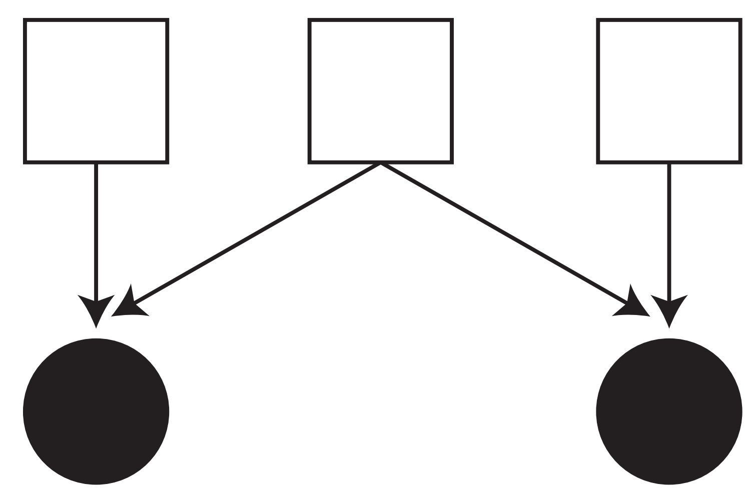
Which features of RGC pathway to keep?



ON and OFF parasol cells

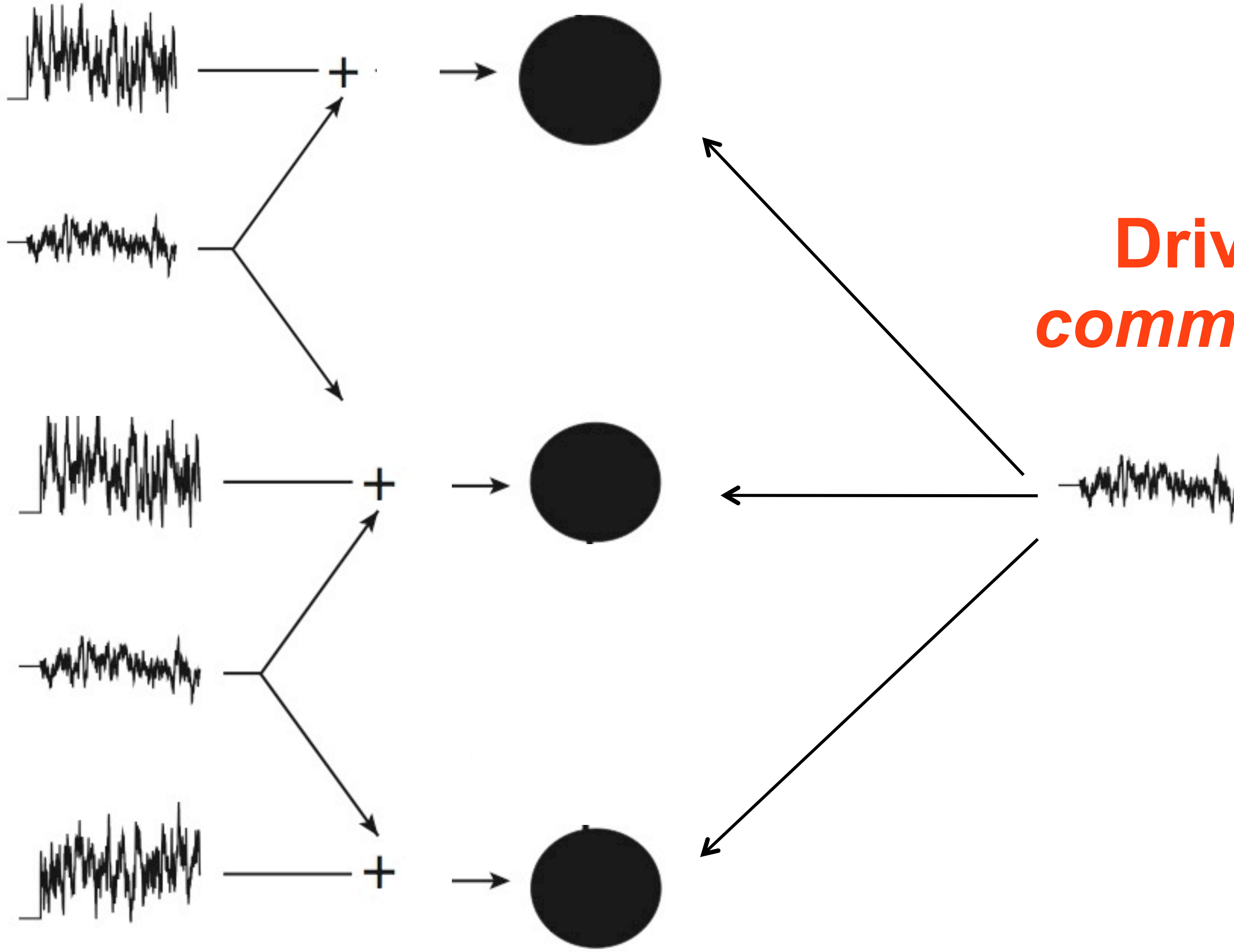


Trong and Rieke 2008



Idea: quantify higher-order correlations systematically in RGC-like circuit

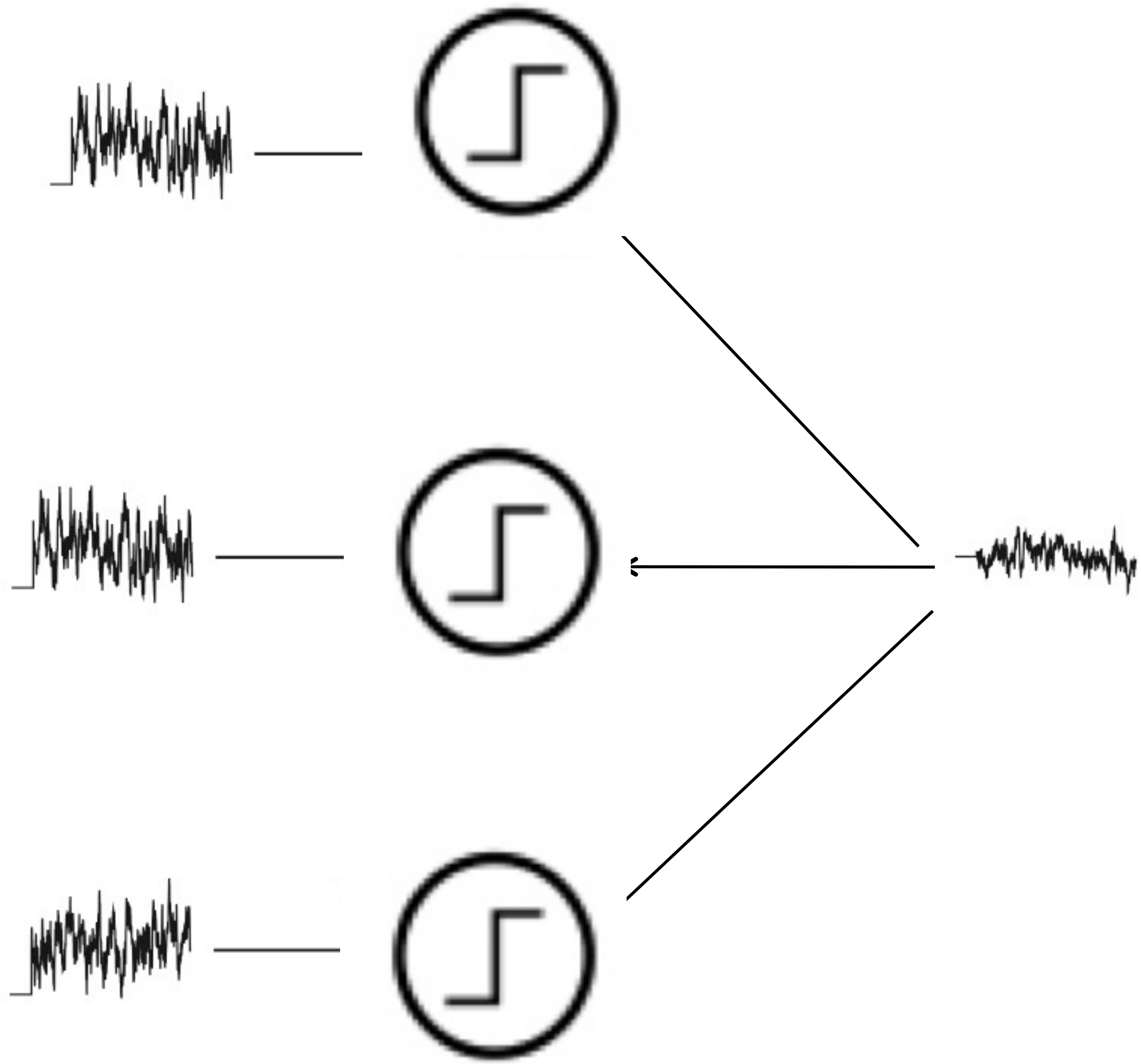
Feedforward structure



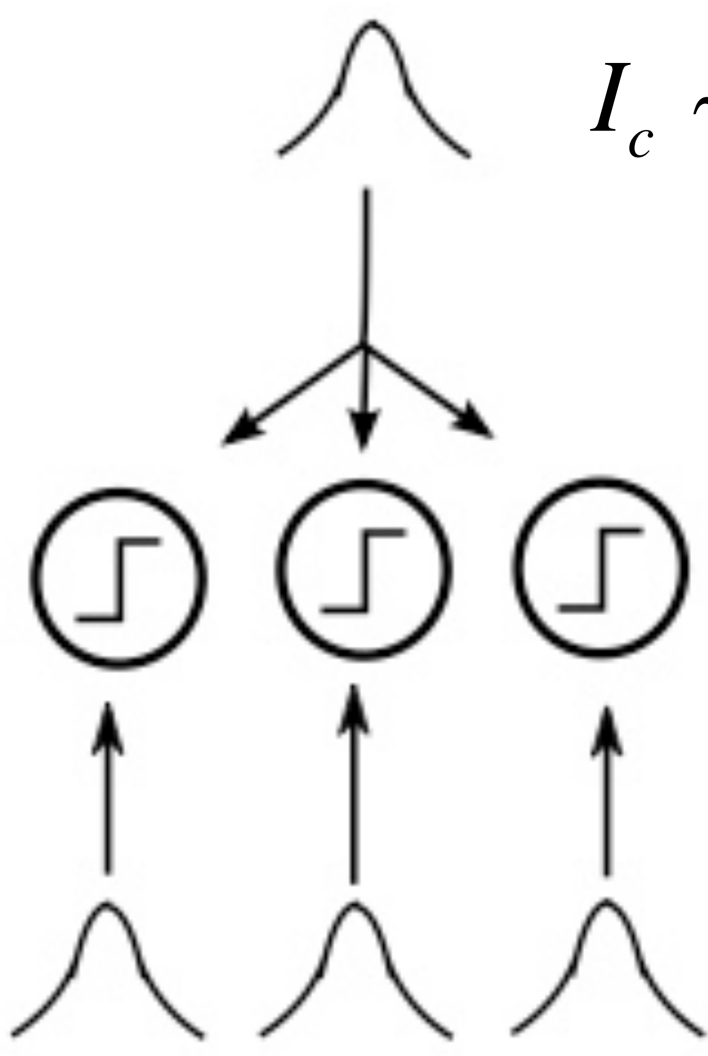
Driven by *common input*

Simplification 1 – triplet input only

Simplification 2 – “threshold” neuron, 0 / 1 spikes



Testing pairwise methods in feed-forward circuits



$I_c \sim N(0, c\sigma^2)$

$I_{1,2,3} \sim N(0, (1-c)\sigma^2)$

p **- observed distribution**

p_2 **- pairwise fit**

There is a triplet common input: so there should be third order correlations, right?

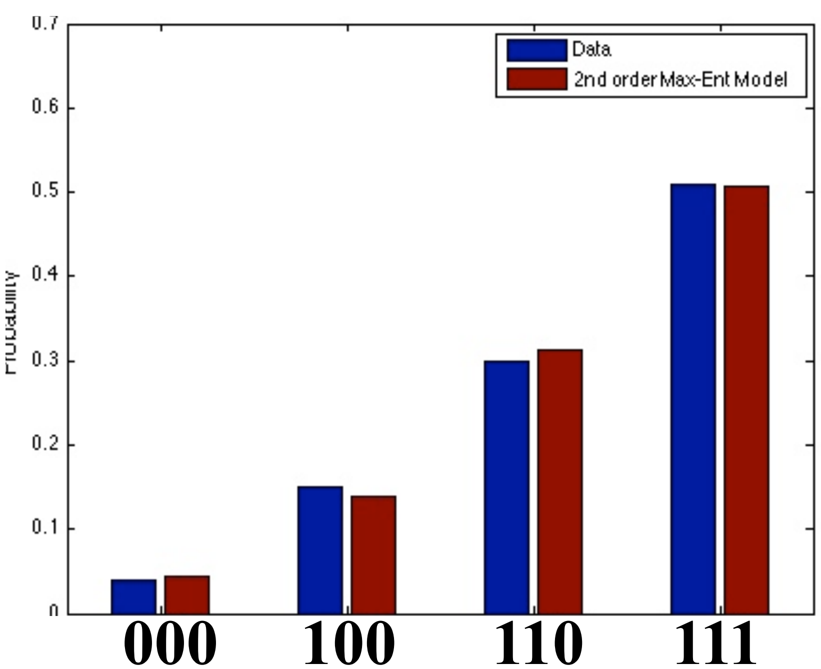
How to quantify higher order correlations?

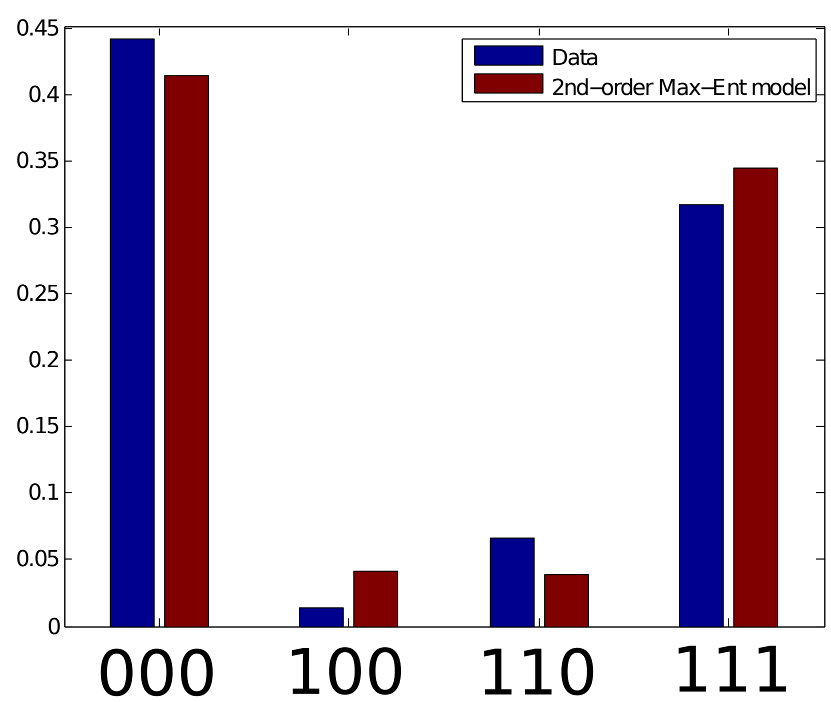
1) Given *P*, find pairwise maximum entropy fit P_2

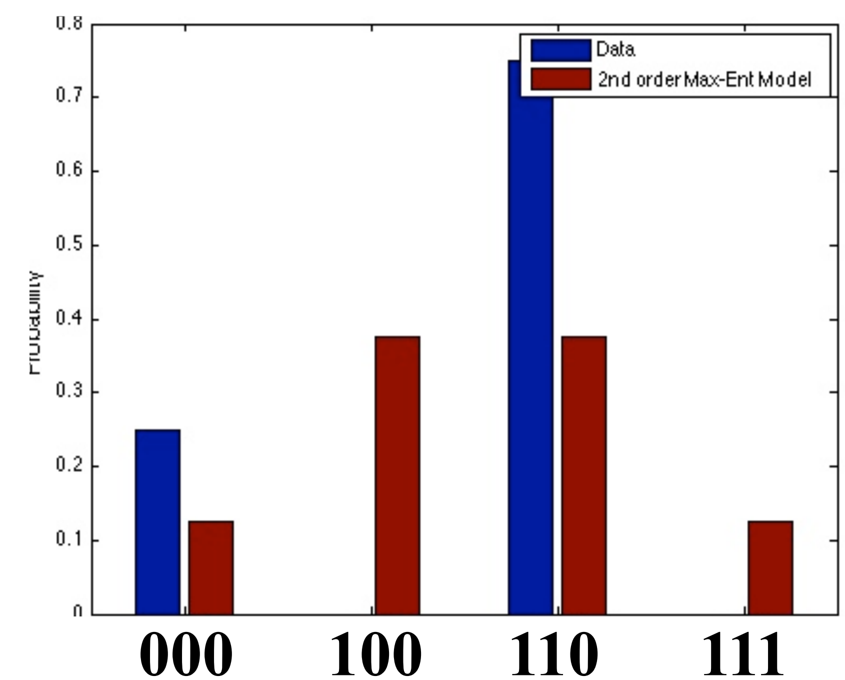
2) Compute distance between *P*, P_2 using Kullback-Leibler divergence $D_{KL}(P, P_2)$

H_2 = entropy of P_2
H_N = entropy of P

$$D_{KL}(P, P_2) = H_2 - H_N$$



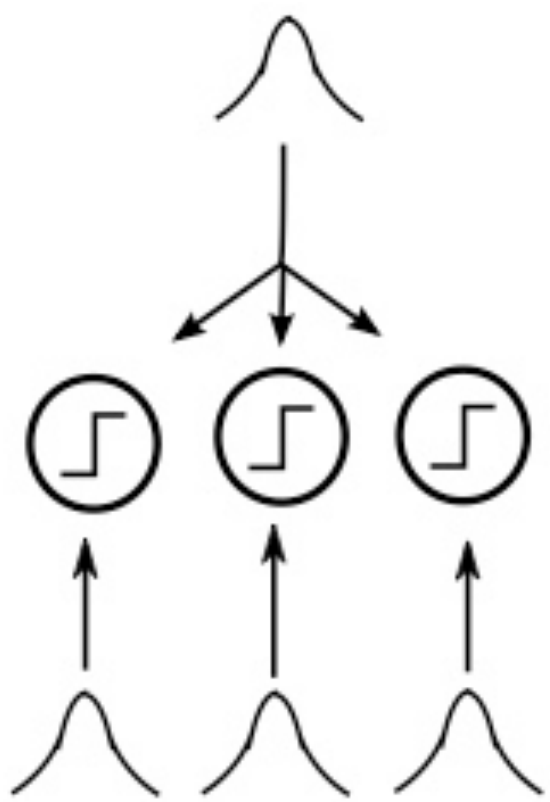


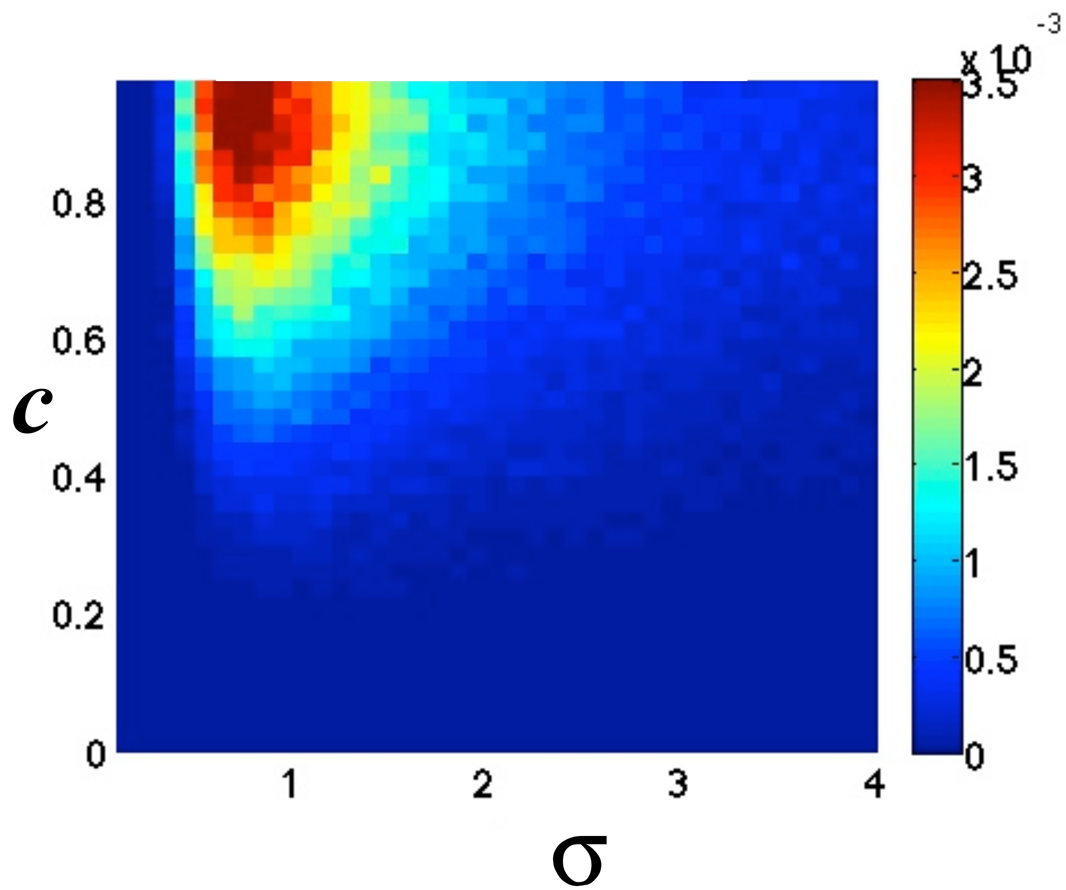


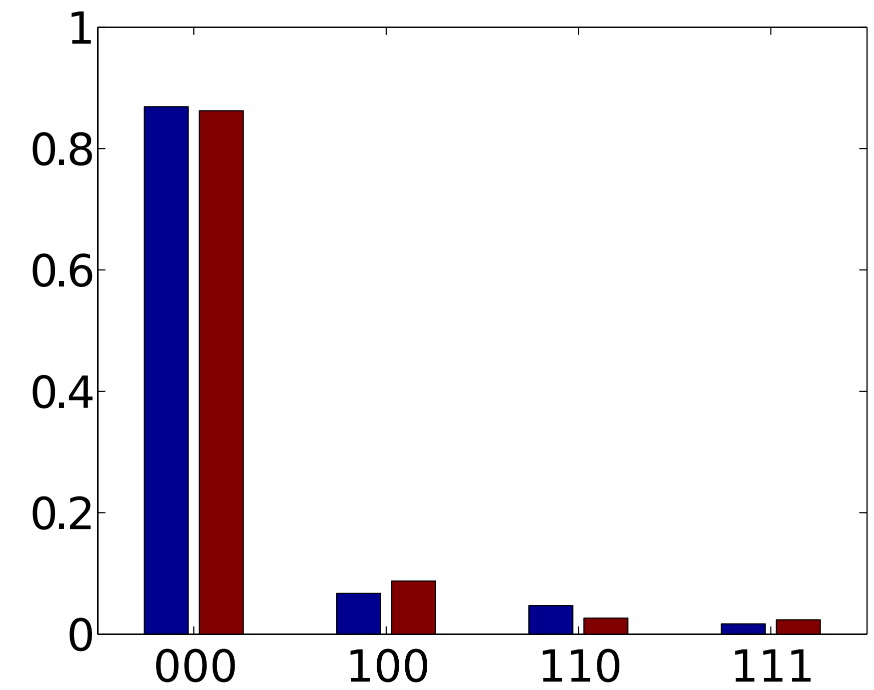
H_2-H_N=0.0013 **H_2-H_N=0.0908** **H_2-H_N=1**

Our setup with gaussian inputs is well-approximated by pairwise fit

H_2-H_N < .0038



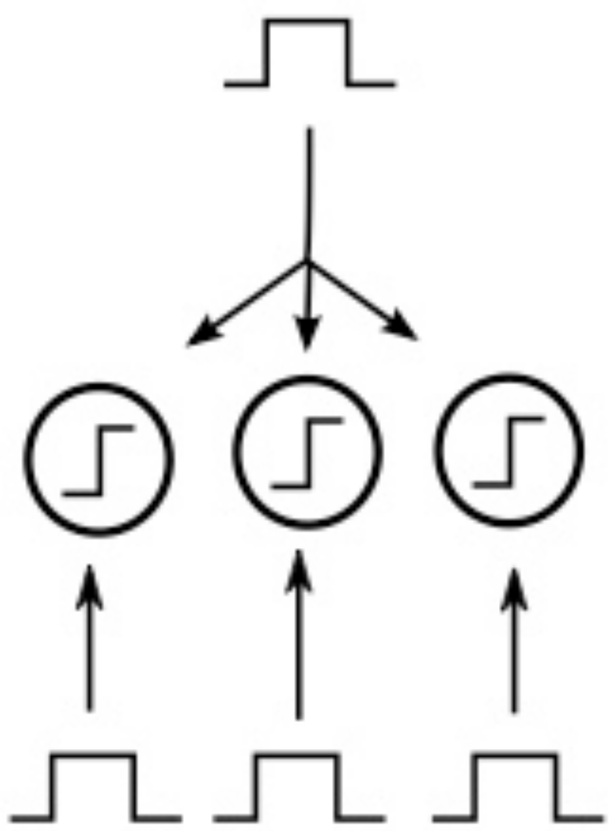


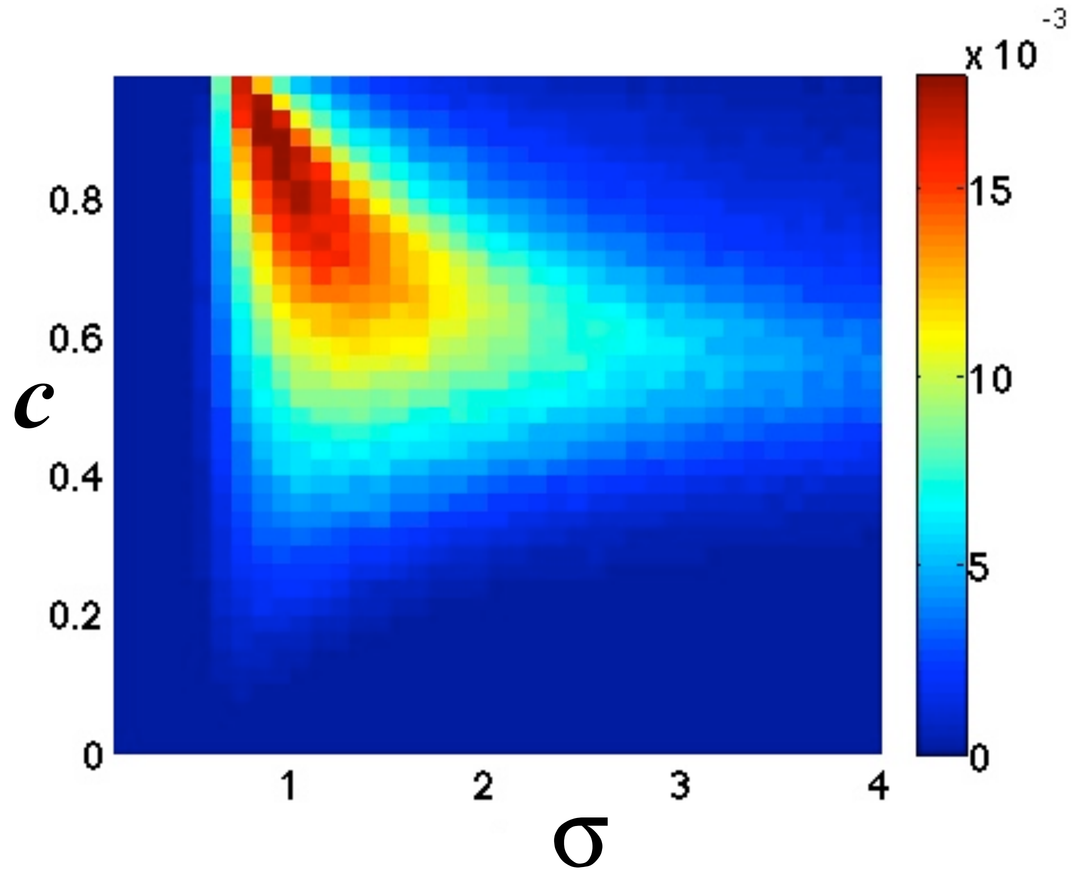


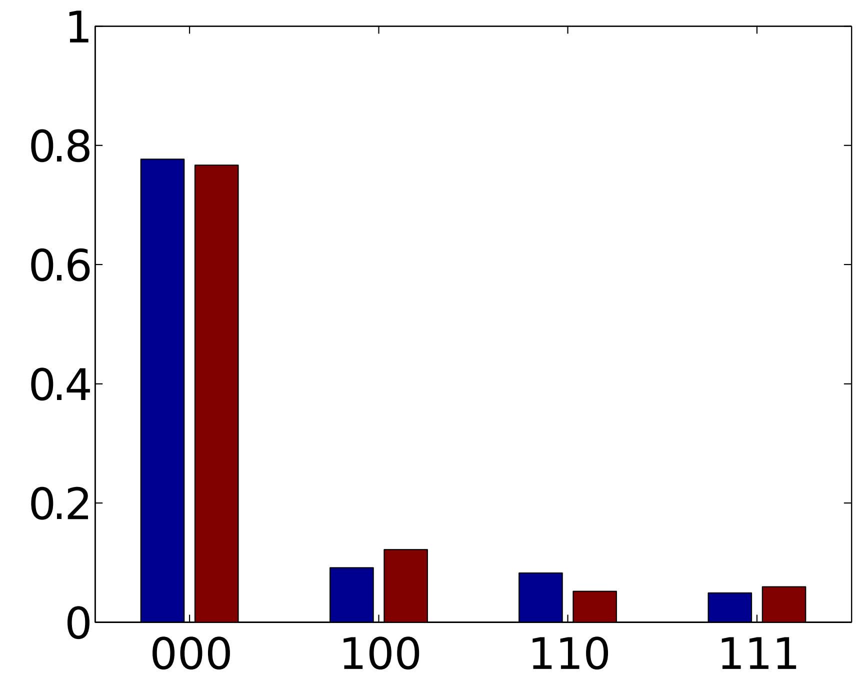
(this is the dichotomized gaussian: e.g. Macke et al. 2009)

We see this with uniform inputs as well…

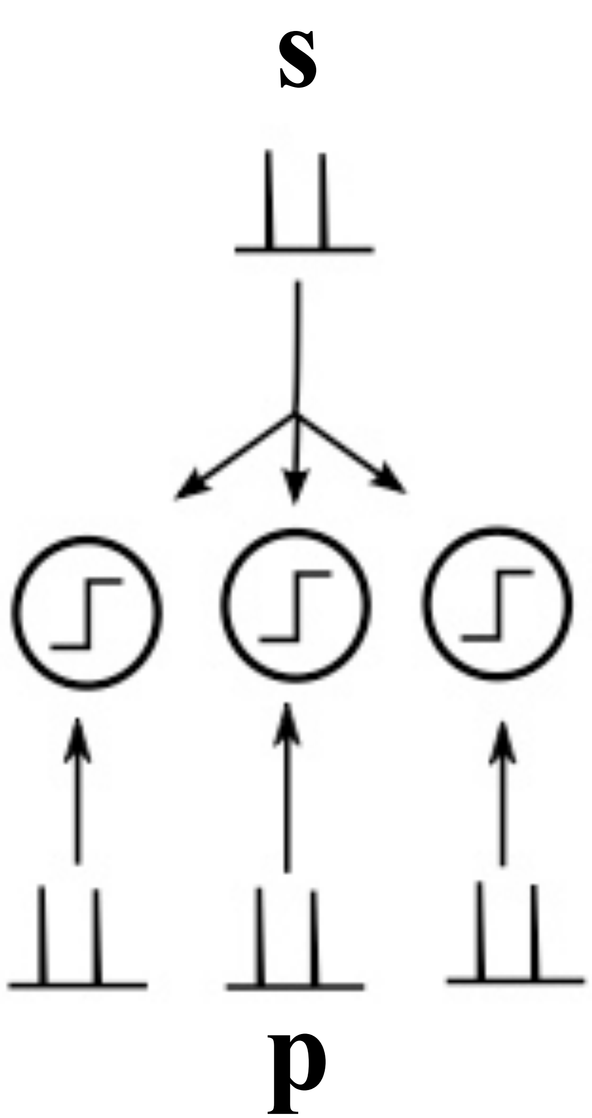
H_2-H_N < .018

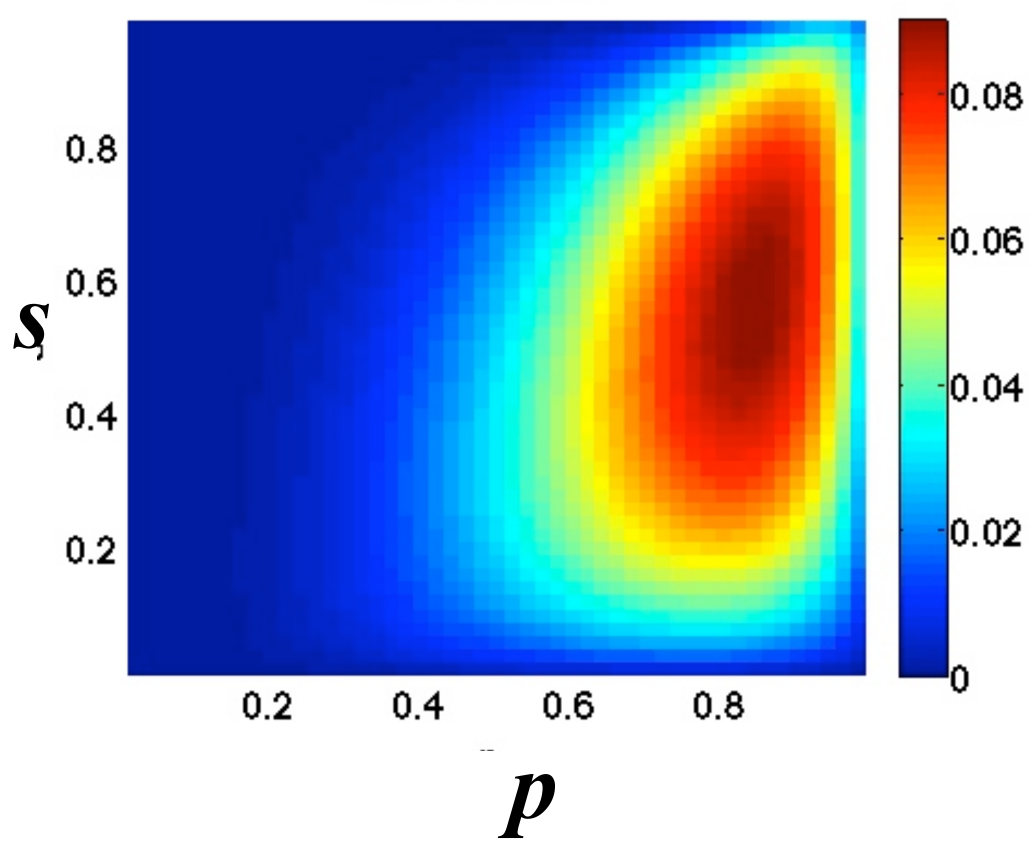




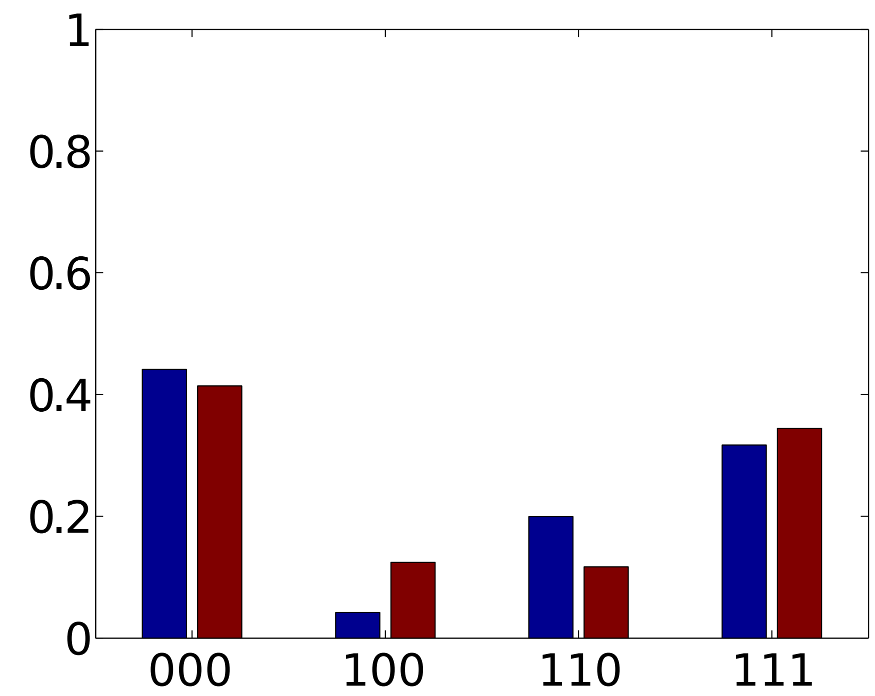


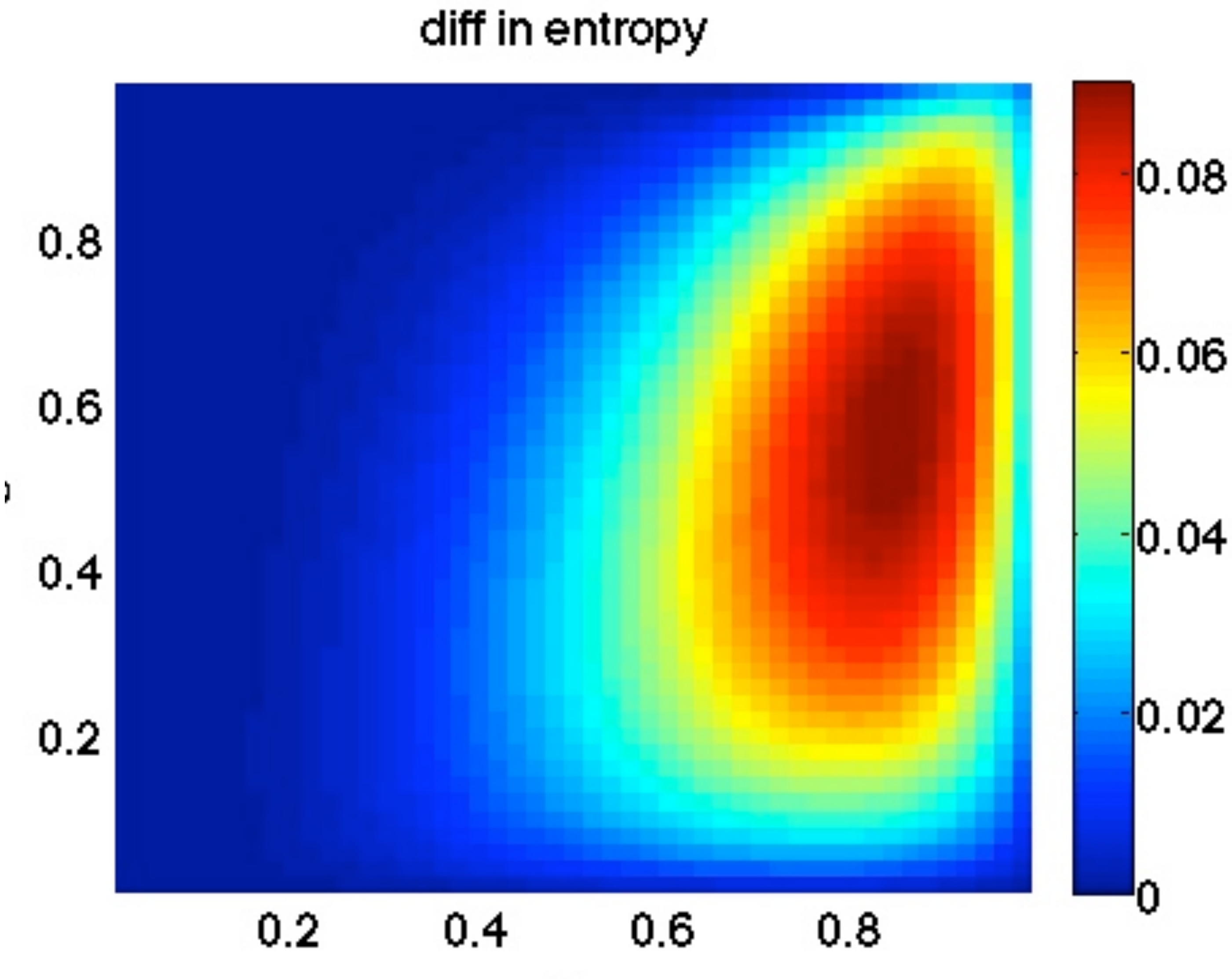
“Binary” model: moderate departure from max-ent





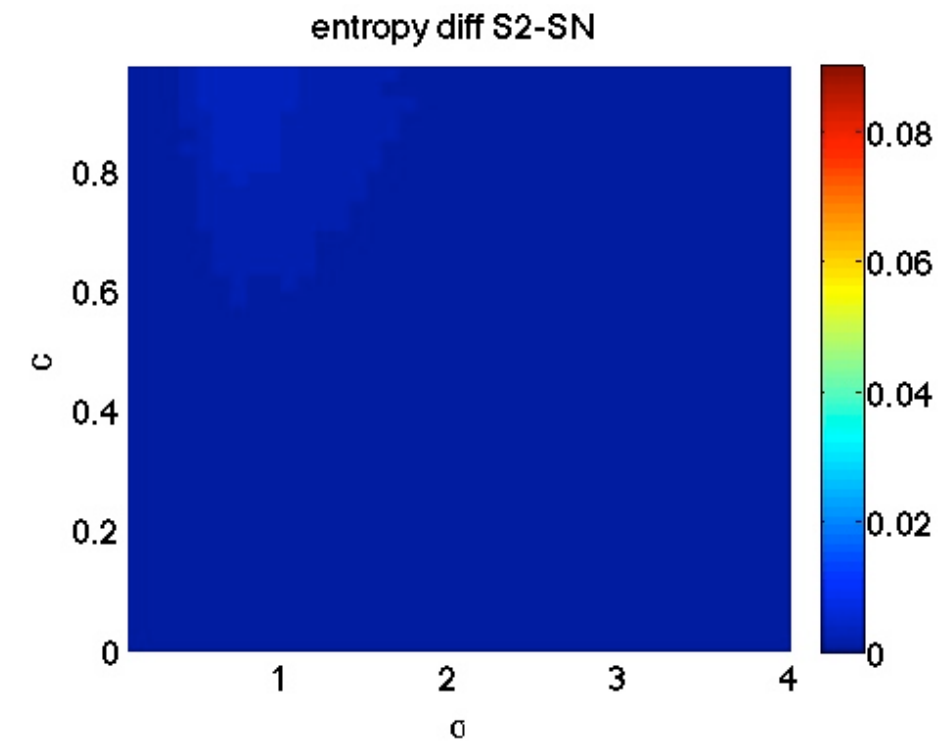
$H_2\text{-}H_N < 0.1$

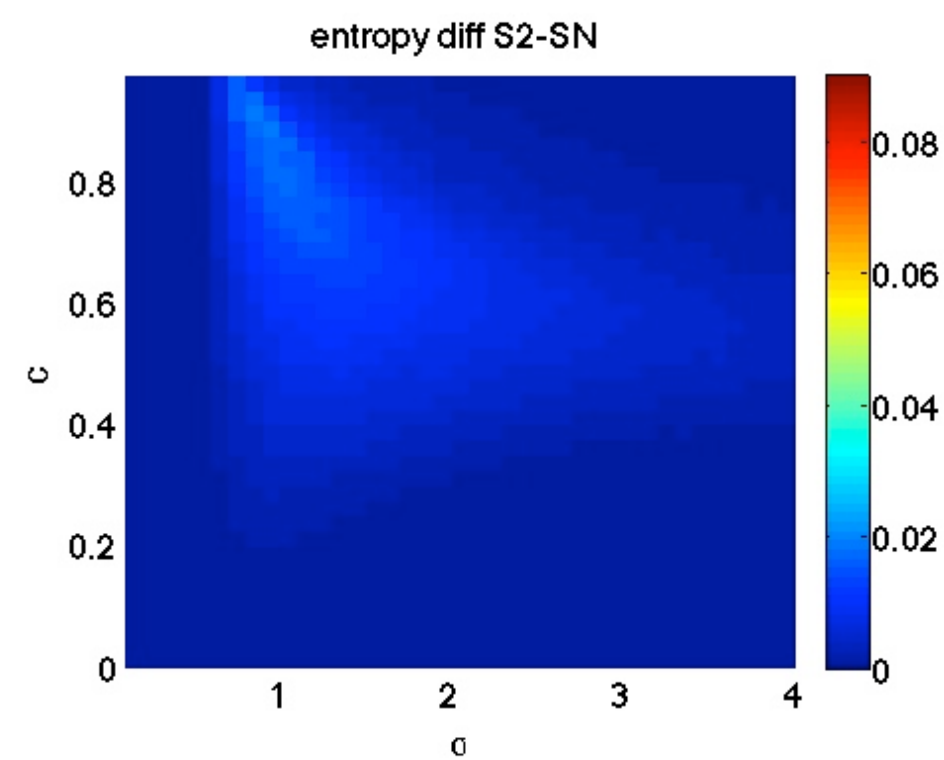




$0 < S_2 - S_N < 0.1$

(25 times larger than in unimodal case)





How can we interpret these results?

- **Consider symmetric distributions on $[0,1]^3$ - That is, stats of cell 1 = stats of cell 2**

- **Max-ent**

$$\Rightarrow p(x_1, x_2, x_3) = \frac{1}{Z}\exp\left(\lambda_1(x_1 + x_2 + x_3) + \lambda_2(x_1x_2 + x_2x_3 + x_1x_3)\right)$$

How can we interpret these results?

- **Consider symmetric distributions on $[0,1]^3$ - That is, stats of cell 1 = stats of cell 2**

- **Max-ent**

$$\Rightarrow p(x_1,x_2,x_3) = \frac{1}{Z}\exp\left(\lambda_1(x_1+x_2+x_3)+\lambda_2(x_1x_2+x_2x_3+x_1x_3)\right)$$

$$p_3 = p(1,1,1)$$
$$p_2 = p(1,1,0)$$
$$p_1 = p(1,0,0)$$
$$p_0 = p(0,0,0)$$

$$\Rightarrow \left(\frac{p_3}{p_0}\right) = \left(\frac{p_2}{p_1}\right)^3$$

How can we interpret these results?

- **Consider symmetric distributions on $[0,1]^3$ -**
  That is, stats of cell 1 = stats of cell 2

- **Max-ent**

$$\Rightarrow p(x_1,x_2,x_3) = \frac{1}{Z}\exp\left(\lambda_1(x_1+x_2+x_3)+\lambda_2(x_1x_2+x_2x_3+x_1x_3)\right)$$

$$p_3 = p(1,1,1)$$
$$p_2 = p(1,1,0)$$
$$p_1 = p(1,0,0)$$
$$p_0 = p(0,0,0)$$

$$\Rightarrow \left(\frac{p_3}{p_0}\right) = \left(\frac{p_2}{p_1}\right)^3$$

a coordinate change

$$(p_1, p_2, p_3) \rightarrow (f_p, f_{1m}, f_{1p})$$

simplifies our constraint...

How can we interpret these results?

- **Consider symmetric distributions on $[0,1]^3$ -**
  That is, stats of cell 1 = stats of cell 2

- **Max-ent**

$$\Rightarrow p(x_1,x_2,x_3) = \frac{1}{Z}\exp\left(\lambda_1(x_1+x_2+x_3)+\lambda_2(x_1x_2+x_2x_3+x_1x_3)\right)$$

$$p_3 = p(1,1,1)$$
$$p_2 = p(1,1,0)$$
$$p_1 = p(1,0,0)$$
$$p_0 = p(0,0,0)$$

$$\Rightarrow \left(\frac{p_3}{p_0}\right) = \left(\frac{p_2}{p_1}\right)^3$$

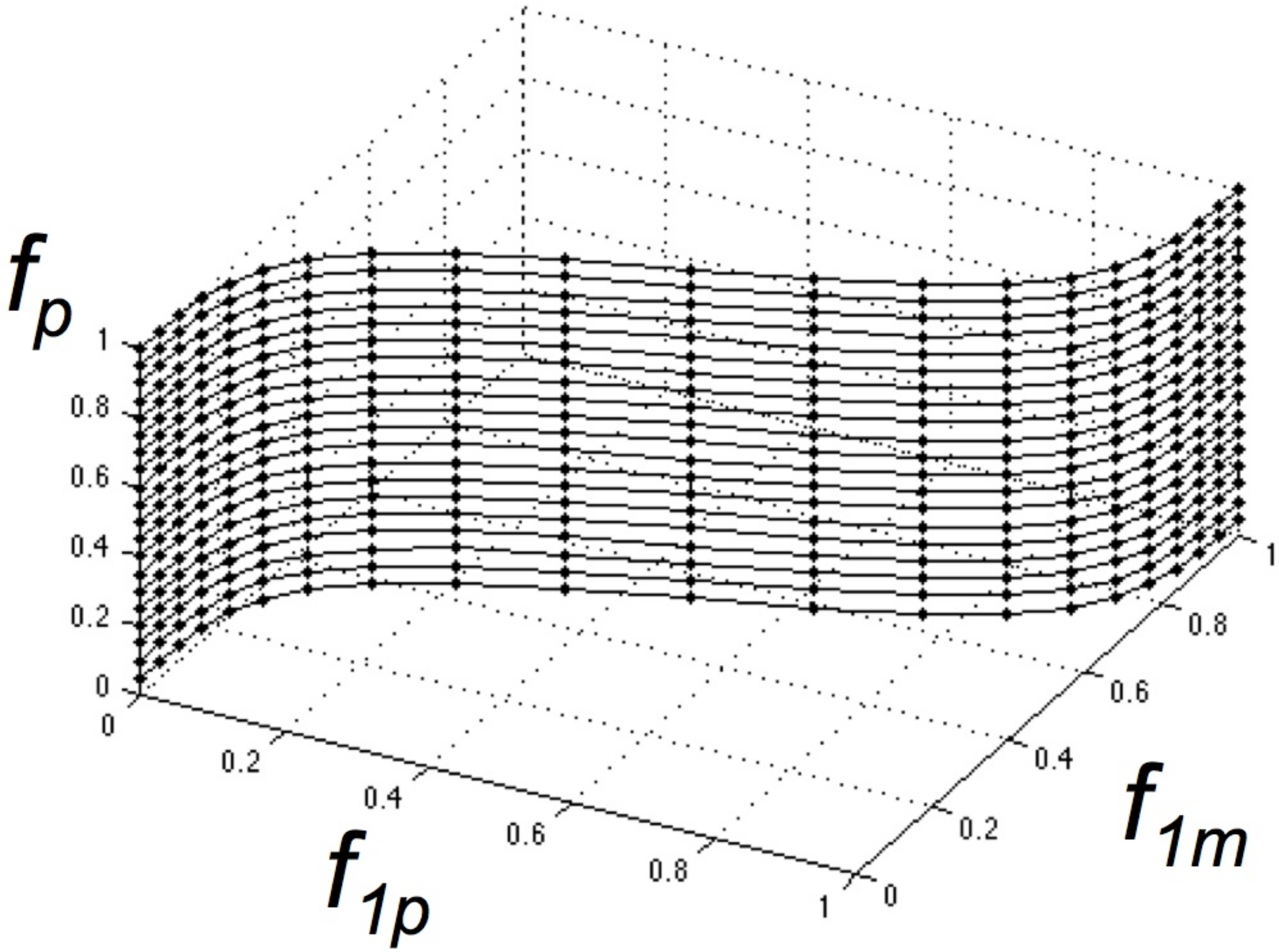
a coordinate change

$$(p_1, p_2, p_3) \rightarrow (f_p, f_{1m}, f_{1p})$$

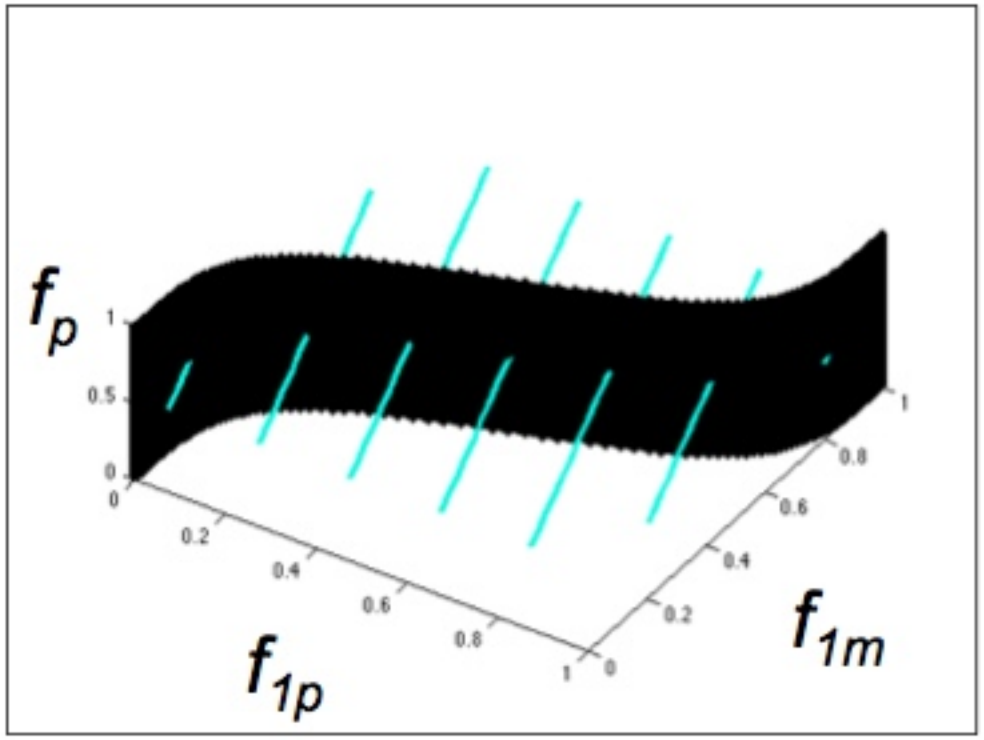
simplifies our constraint...

$$\Rightarrow f_{1p} = \frac{f_{1m}^3}{1-3f_{1m}+3f_{1m}^2}$$

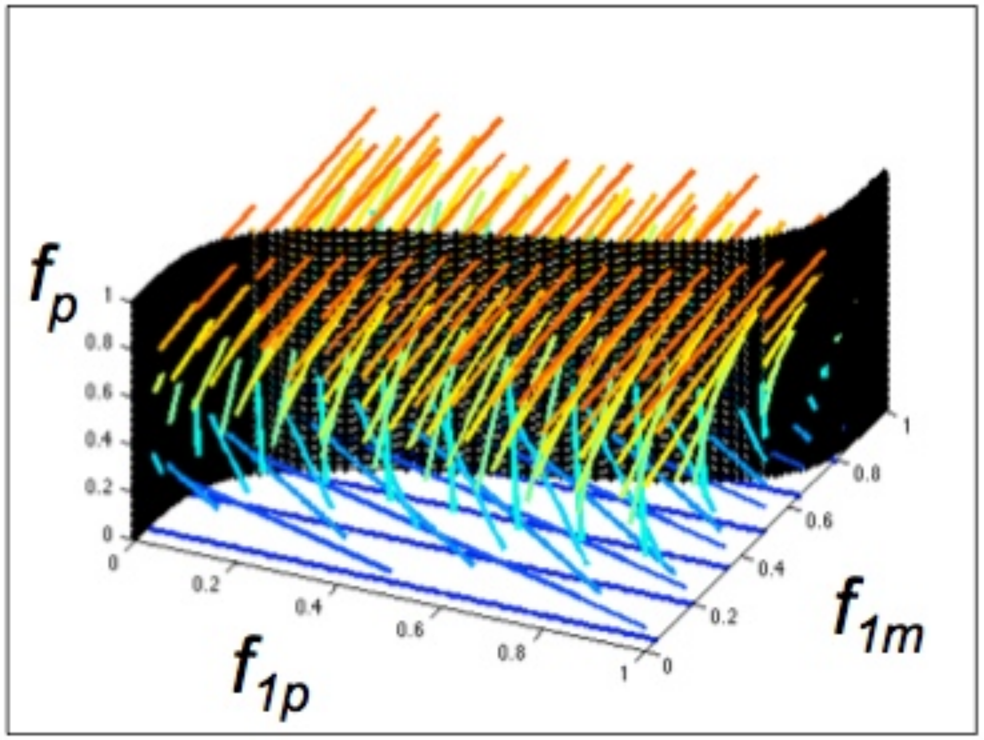
Constraint surface in new coordinates



Distance from surface gives info about $D_{KL}(p, p_2)$



Family of lines for one value of f_p

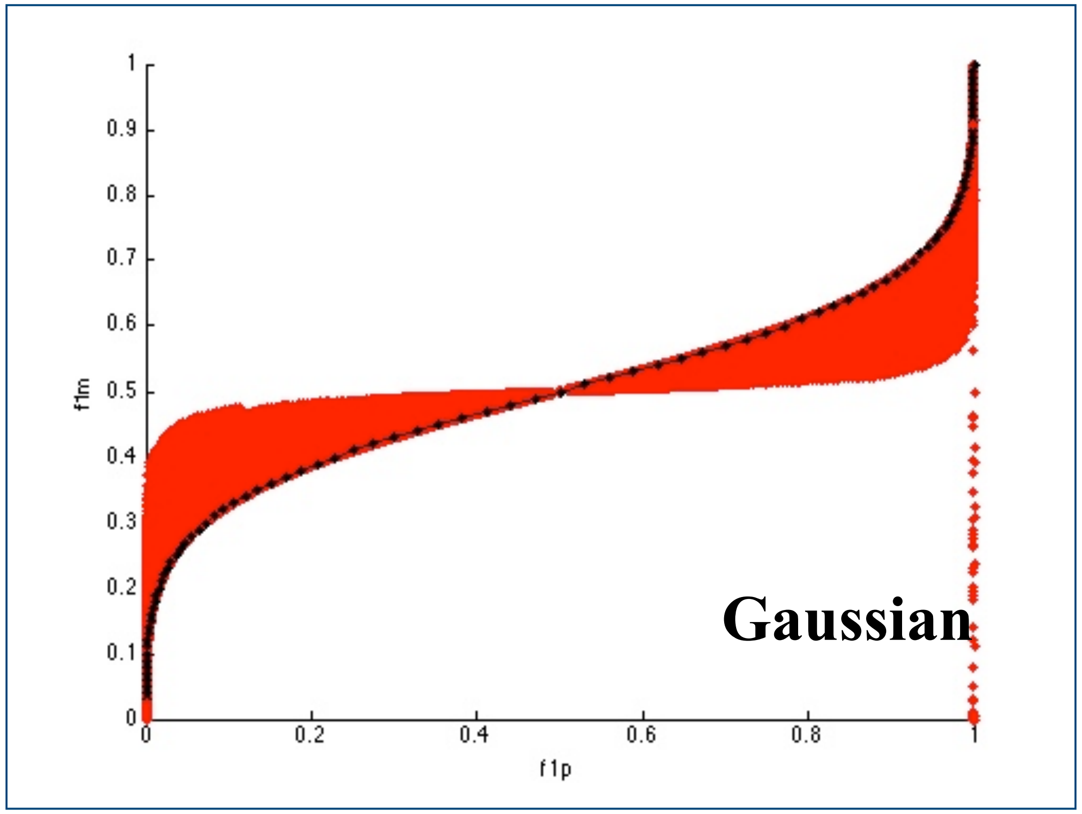


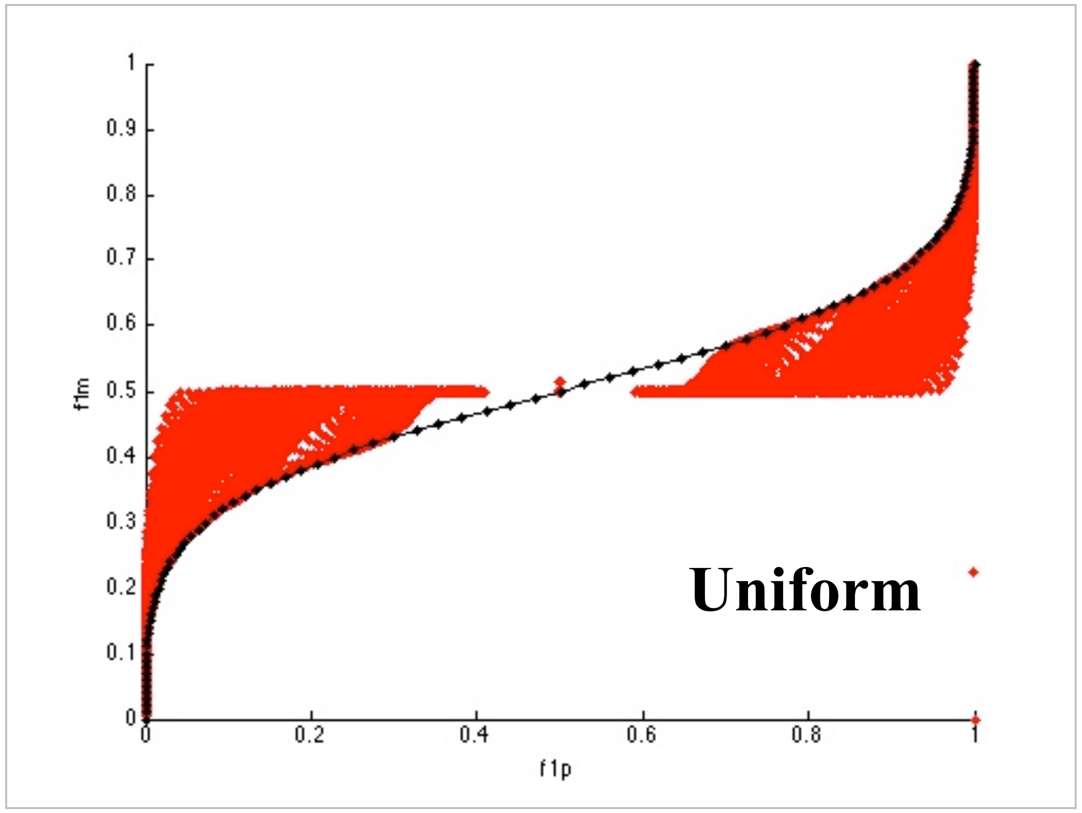
Families of lines for several values of f_p

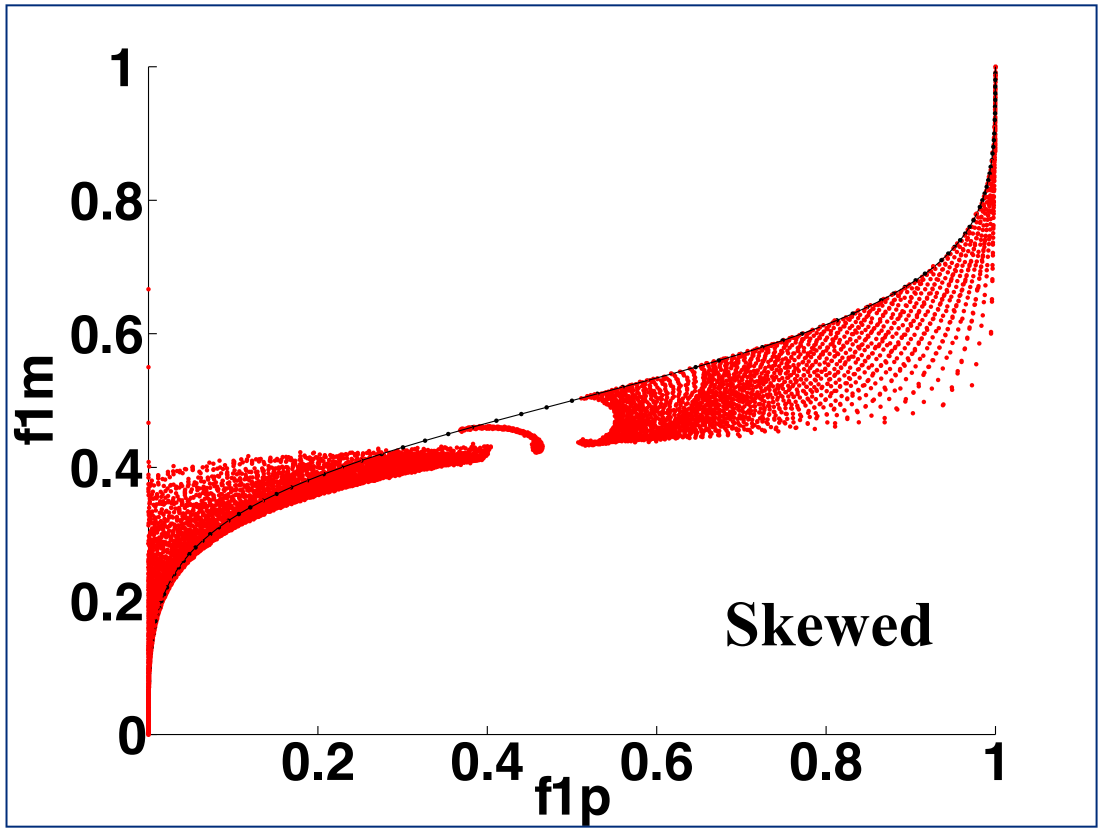
As one moves away from Σ, $D_{KL}(P,P_2)$ increases *quadratically:*

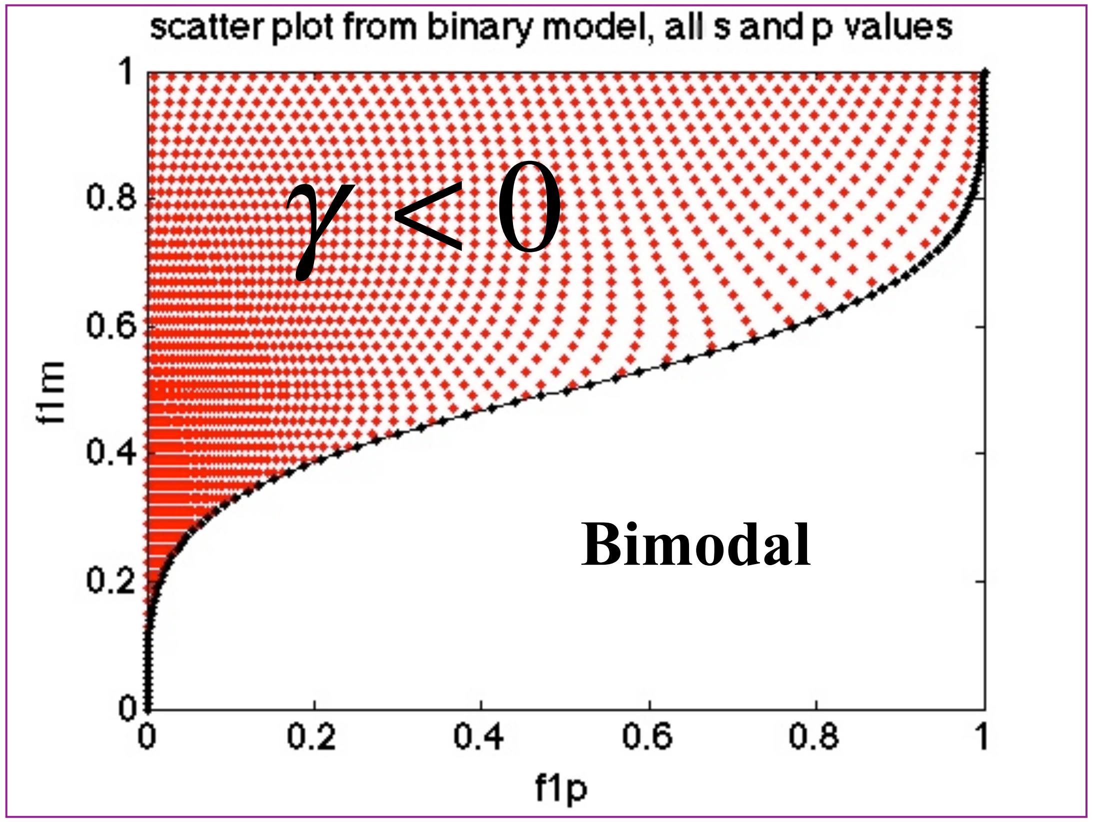
$$D_{KL}(P,P_2) \approx (f_{1m} - f_{1m,0})^2 H(\nu, f_{p,0})$$

$f_{1m,0}, f_{p,0}$ *denote quantities evaluated on Σ*


Gaussian
f1m
f1p


Uniform
f1m
f1p


Skewed
f1m
f1p


scatter plot from binary model, all s and p values
$\gamma < 0$
Bimodal
f1m
f1p

For small common inputs, bimodal > unimodal

Start with cells firing independently...

Perturb with *unimodal* common input (variance *c*):

$$p(x) = \frac{1}{\sqrt{c}} f\left(\frac{x-\mu}{\sqrt{c}}\right) \quad \rightarrow \quad D_{KL}\left(P, P_2\right) \approx c^3 C_f^U$$

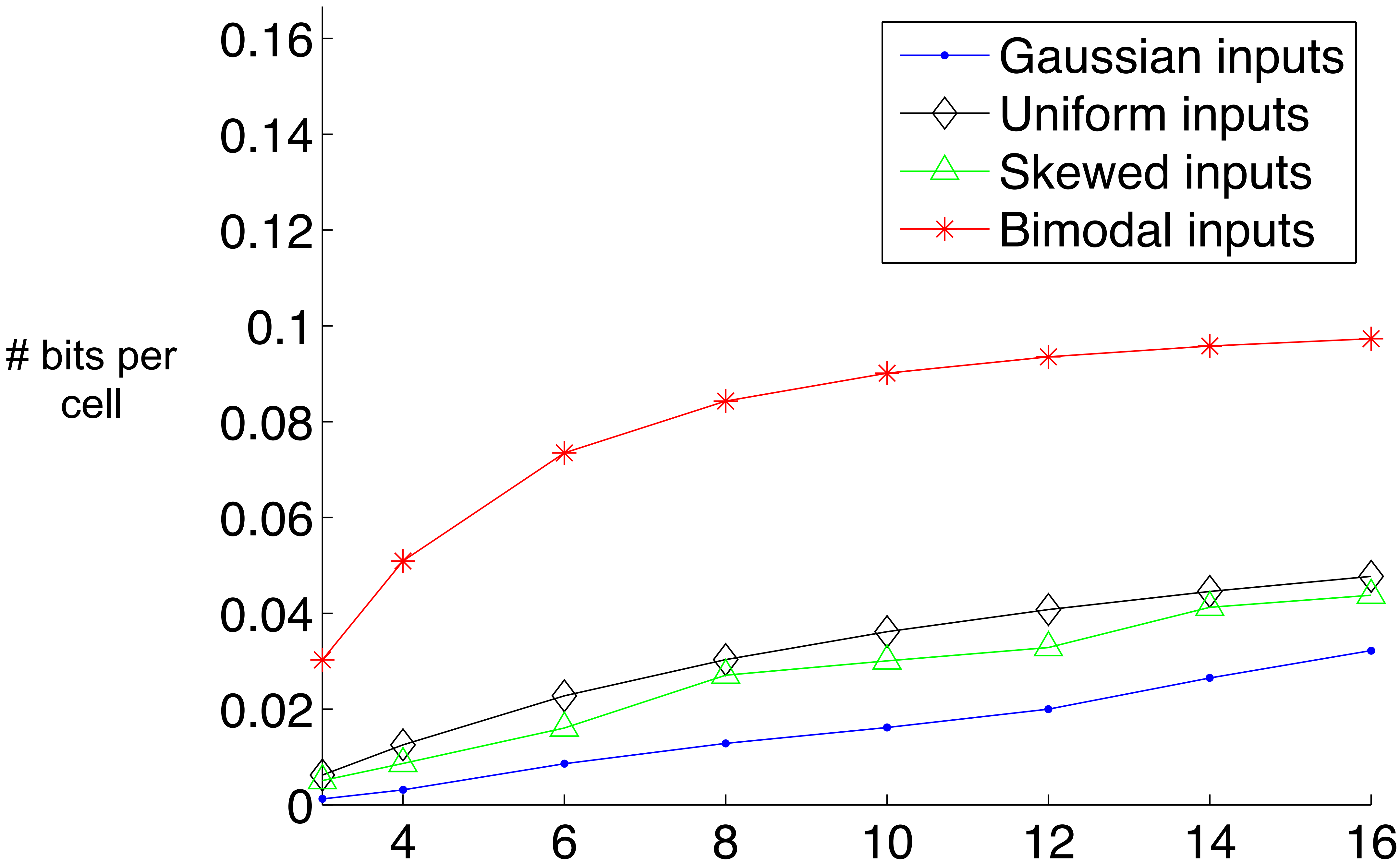
in special cases,
***c*³ ⟶ *c*⁴**

Perturb with *bimodal* common input (variance *c*):

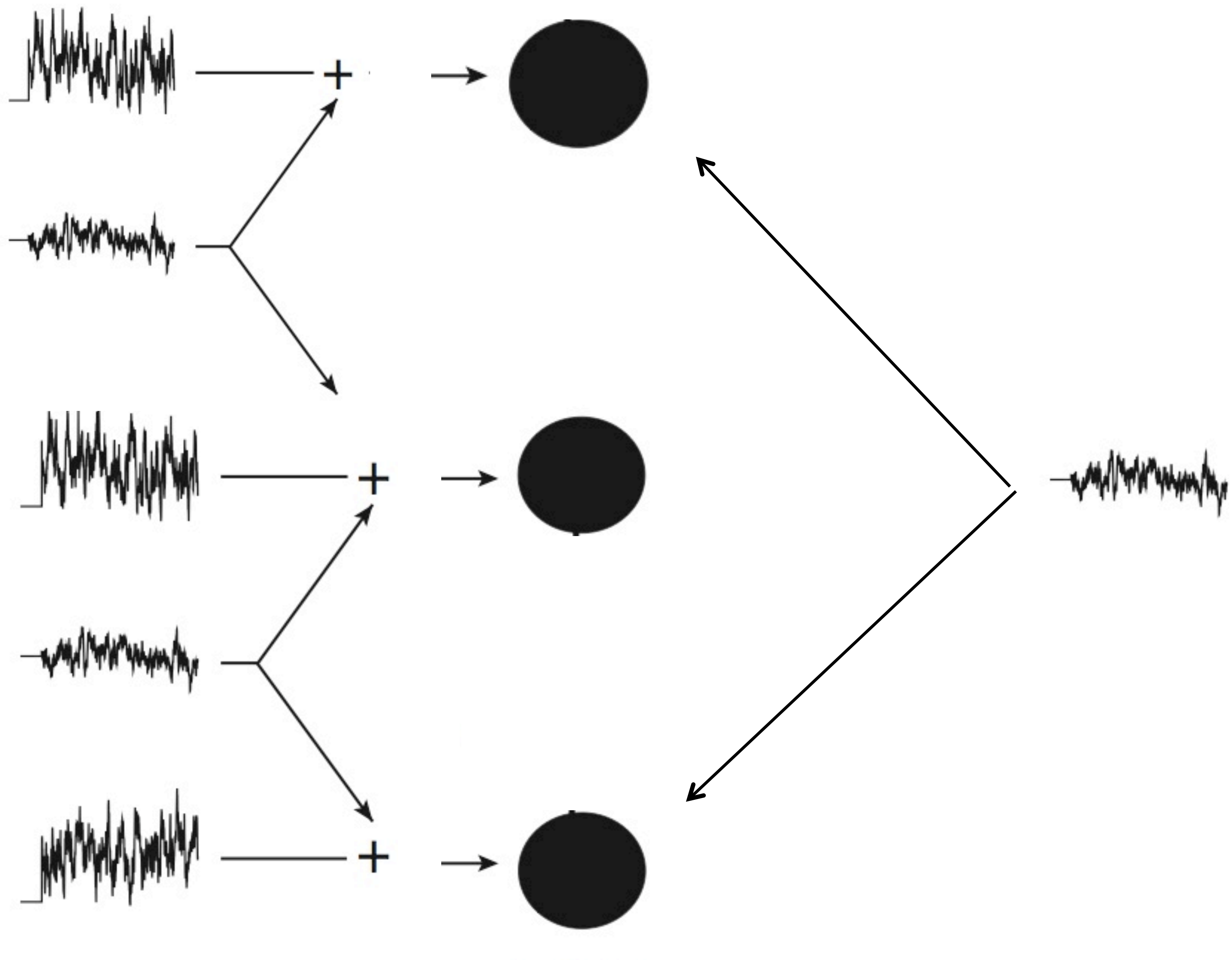
$$p(x) = \frac{1-c}{\sqrt{c}} f\left(\frac{x}{\sqrt{c}}\right) + c f\left(\frac{x-\mu}{\sqrt{c}}\right) \quad \rightarrow \quad D_{KL}\left(P, P_2\right) \approx c^2 C_f^B$$

Patterns persist for larger N …

$$\max\left(D_{KL}(P, P_2)\right) / N$$

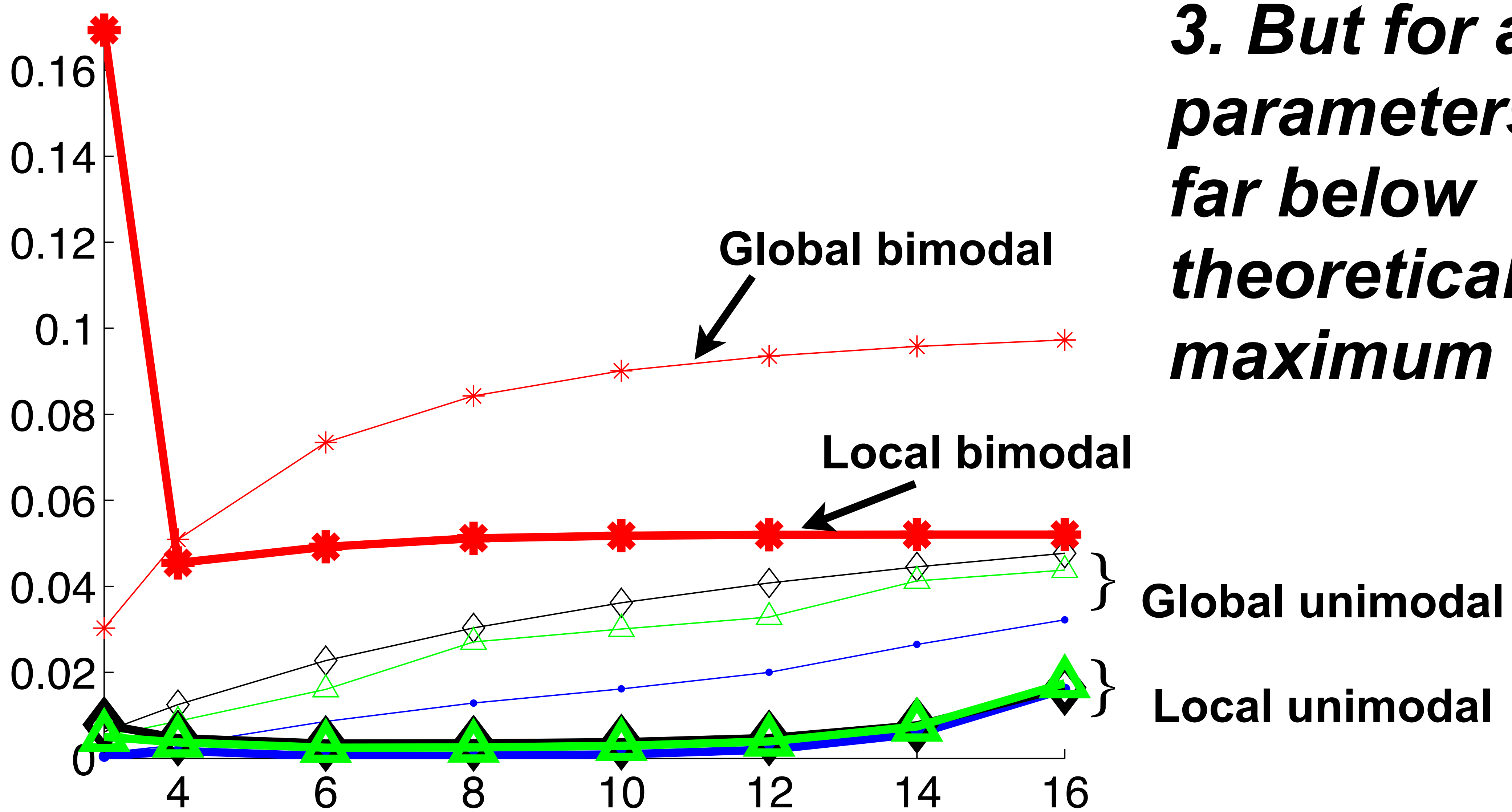


Pairwise inputs



Global vs. pairwise for moderate N, all input types

$$\max\left(D_{KL}(P,P_2)\right)/N$$

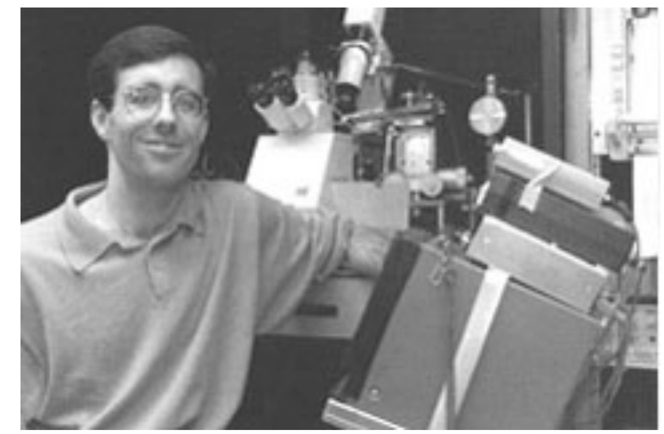


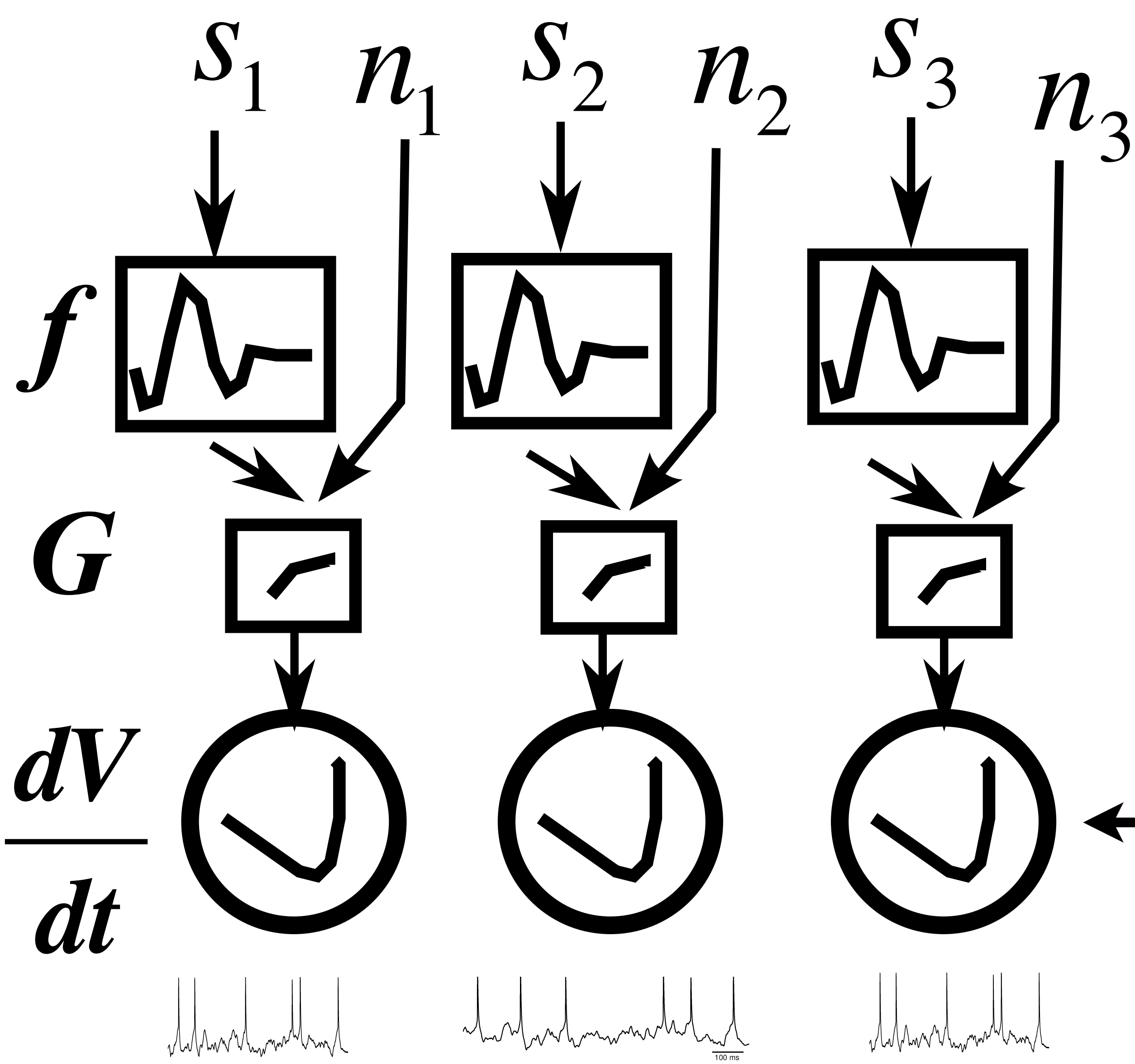
1. Global generates more than local
2. Bimodal generates more than unimodal...

3. But for all parameters, level is far below theoretical maximum (1/3)

Realistic "RGC-like" network

- Construct a detailed model of the response of a primate ON parasol cell: constrain with intracellular recordings





processing of stimuli by upstream circuitry

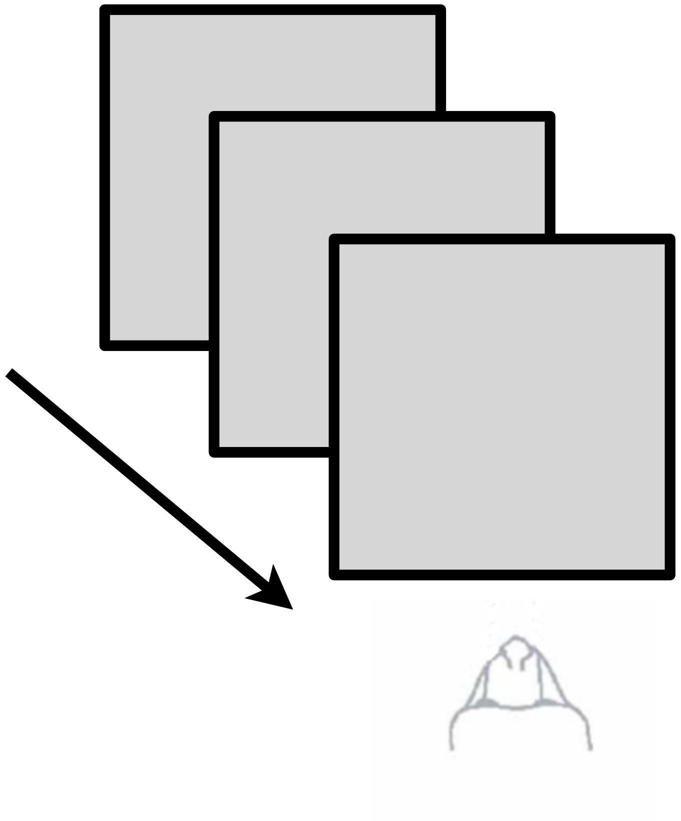
spike generator

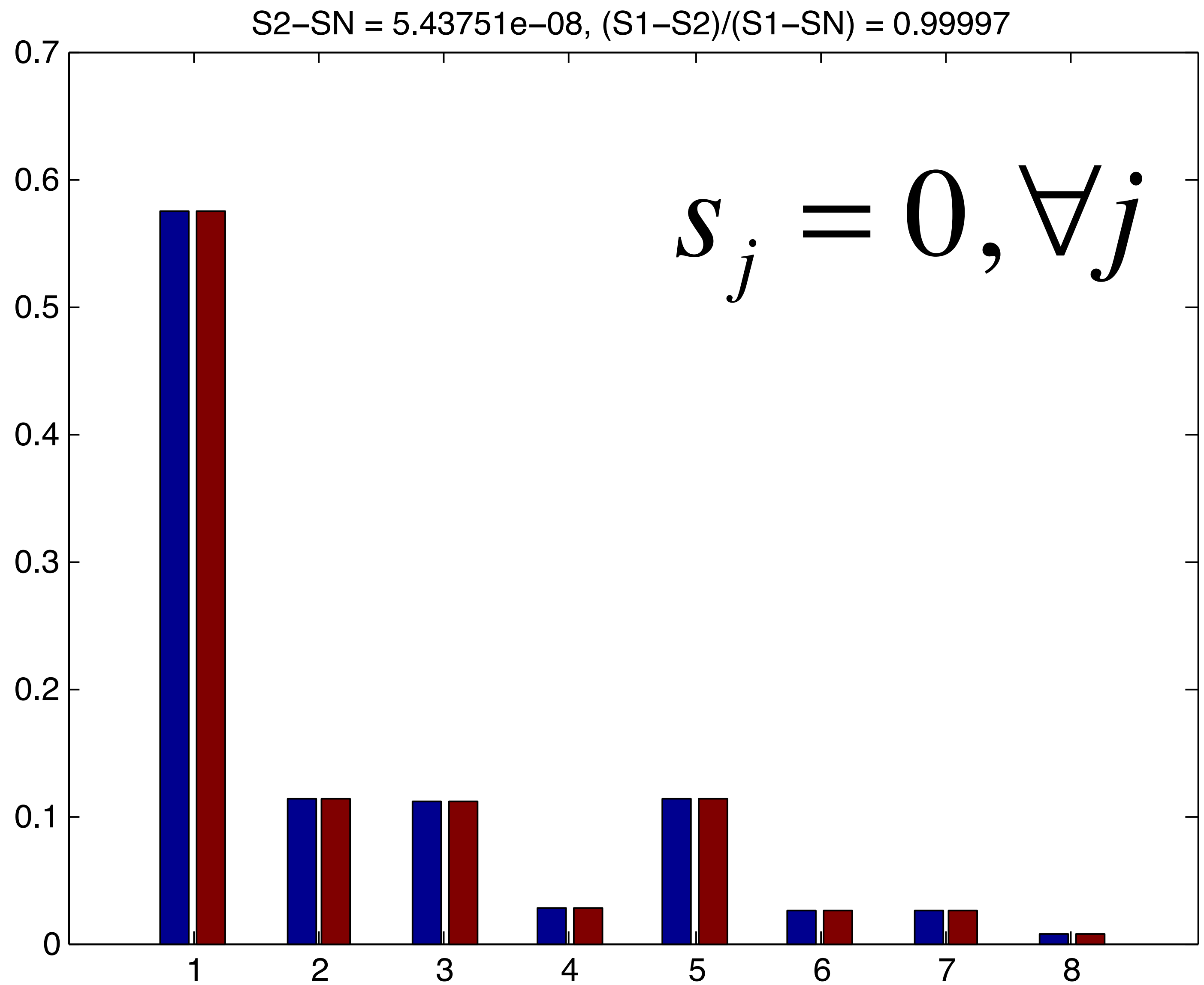
Can any light stimuli bring out higher order statistics in this circuit, and if so, what are the required spatiotemporal statistics?

- **With correlated noise, and constant light stimuli, responses very well fit by pairwise maximum entropy model**

$$I_j = G(f * s_j + n_j)$$

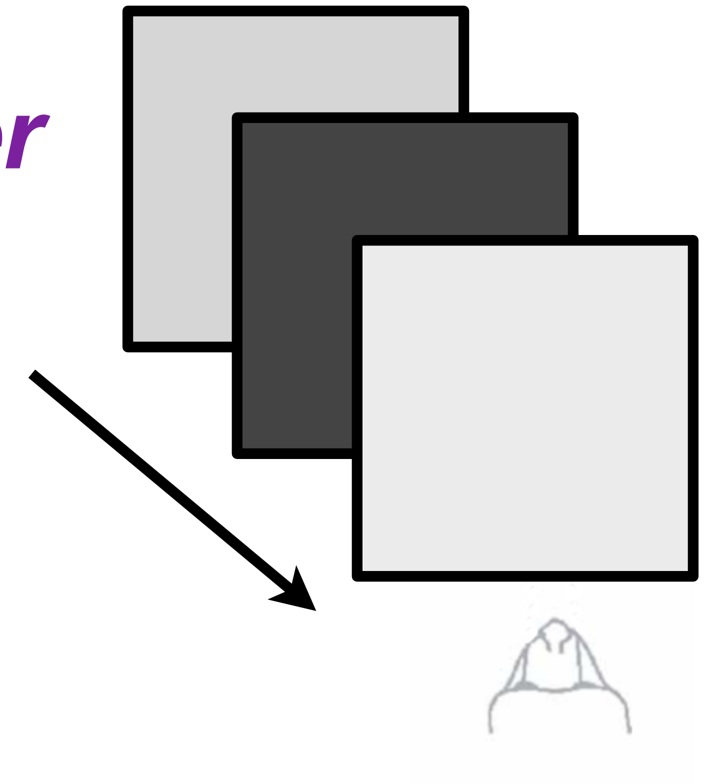
$$Cov\left(n_i, n_j\right) \approx 0.3$$





Can any light stimuli bring out higher order statistics in this circuit, and if so, what are the required spatiotemporal statistics?

ex: full-field flicker

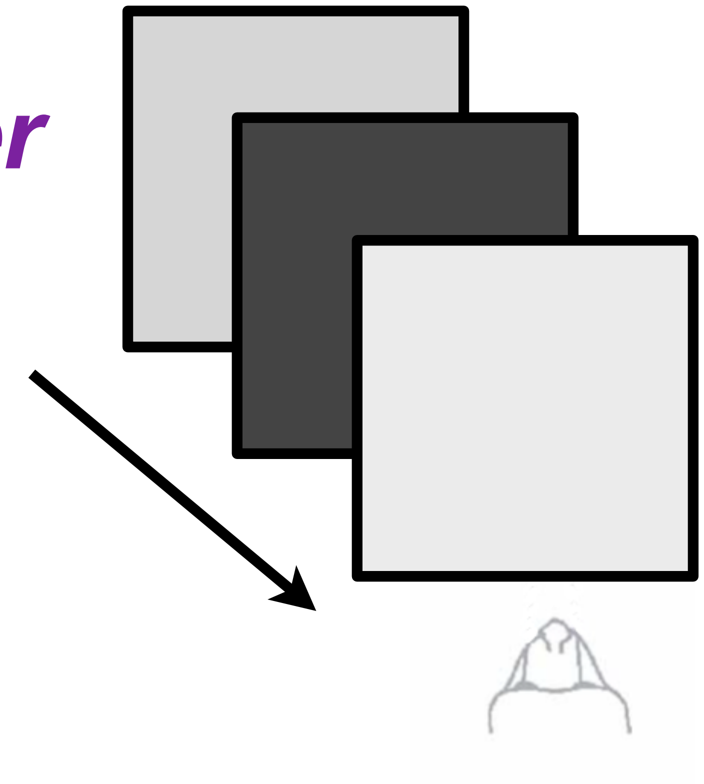


$$I_j = G(f * s_j + n_j)$$

$$s_j = c(t), \forall j$$

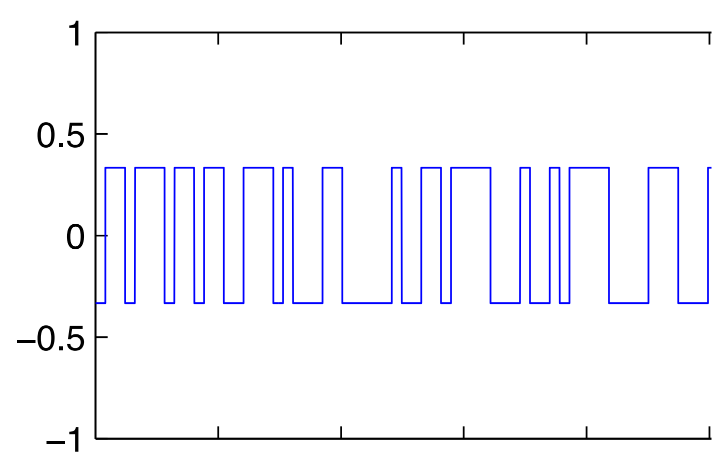
Can any light stimuli bring out higher order statistics in this circuit, and if so, what are the required spatiotemporal statistics?

ex: full-field flicker



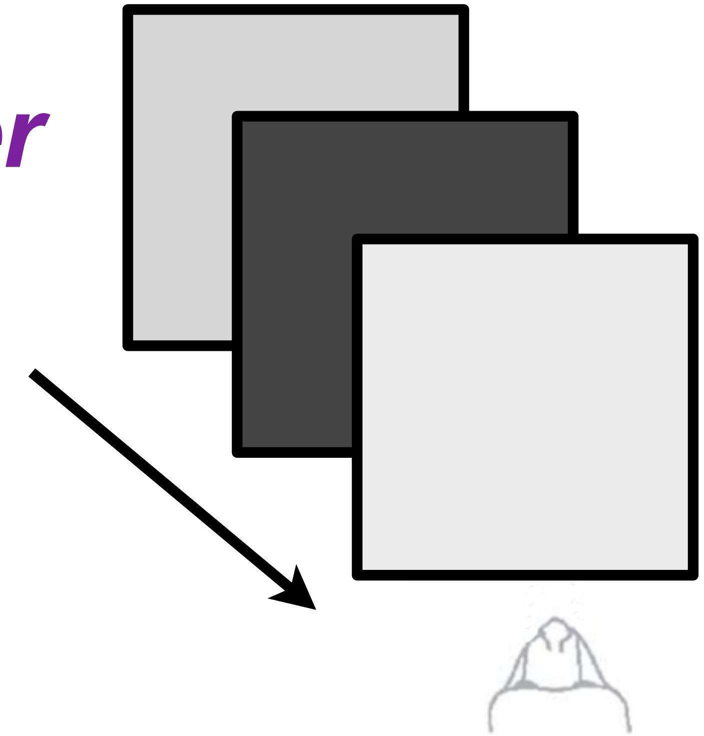
$$I_j = G(f * s_j + n_j)$$

$$s_j = c(t), \forall j$$



Can any light stimuli bring out higher order statistics in this circuit, and if so, what are the required spatiotemporal statistics?

ex: full-field flicker

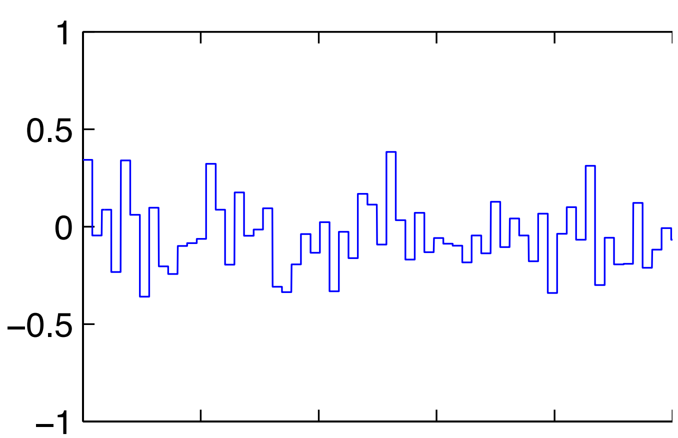


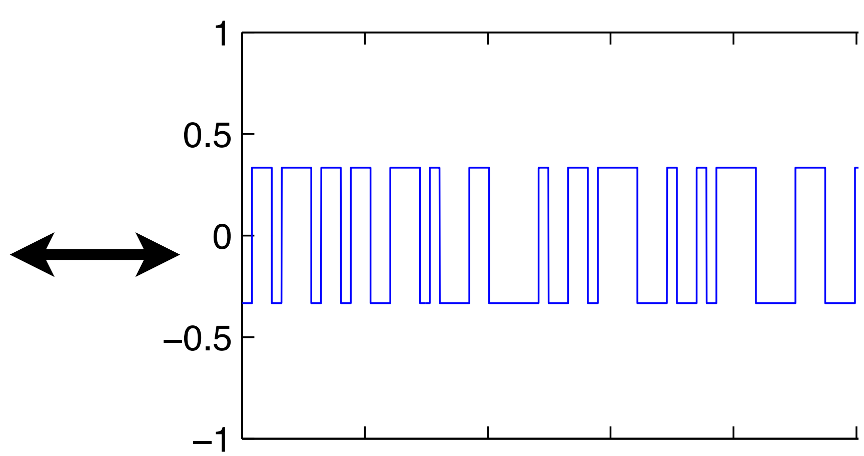
$$I_j = G(f * s_j + n_j)$$

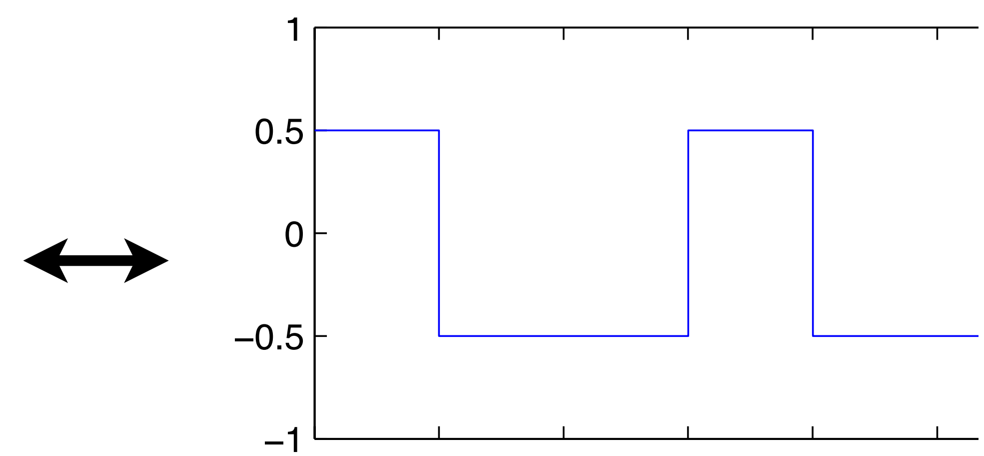
$$s_j = c(t), \forall j$$

vary marginal stats

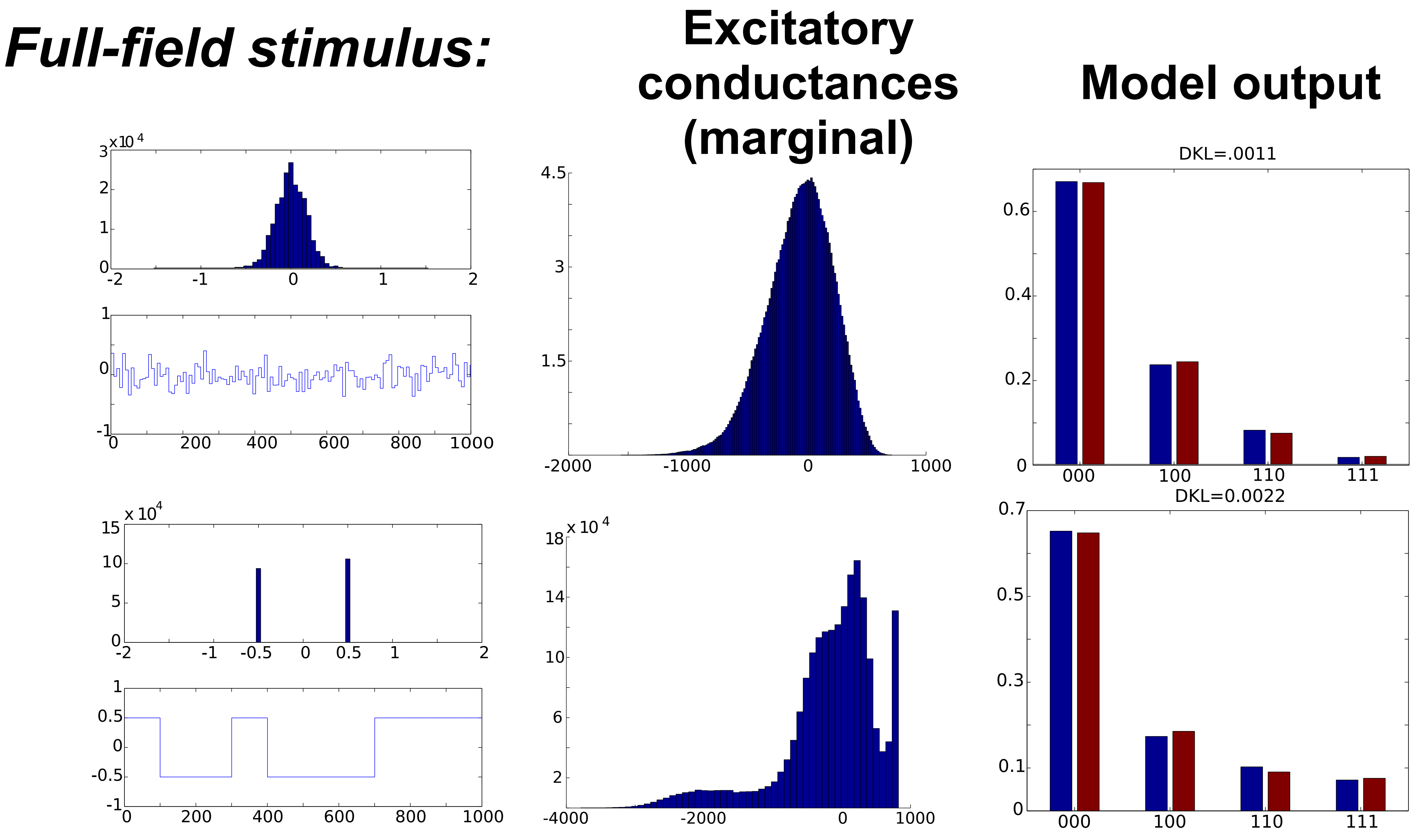
vary refresh rate or variance of input





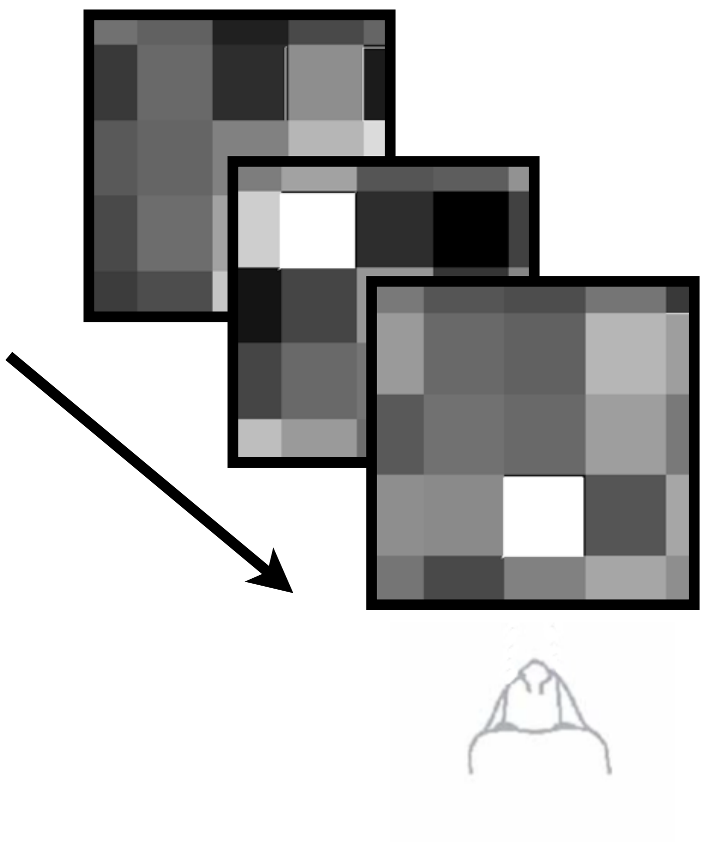


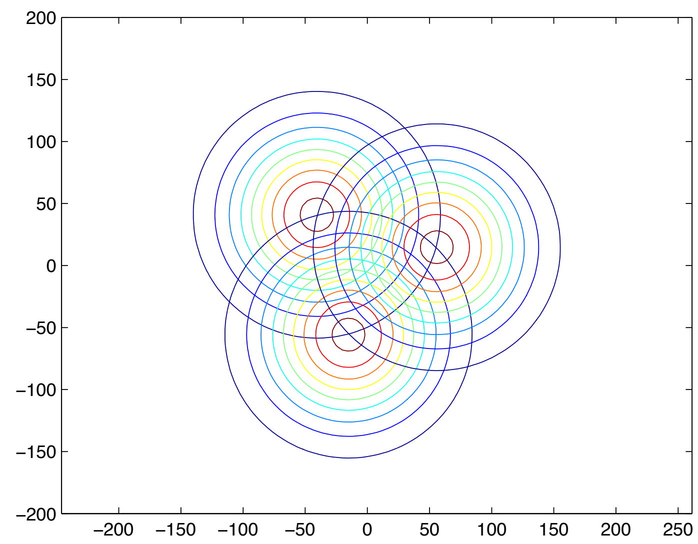
Answer for full-field flicker: no!



$D_{KL}(P, P_2)$ ***under 0.007 (0.002333 per cell) for all conditions***

What about a spatially variable stimulus?





$$D_{KL}(P, P_2)$$

remains under 0.0045 (0.0015 per cell) for all conditions

$$I_j = G(f * s_j + n_j)$$

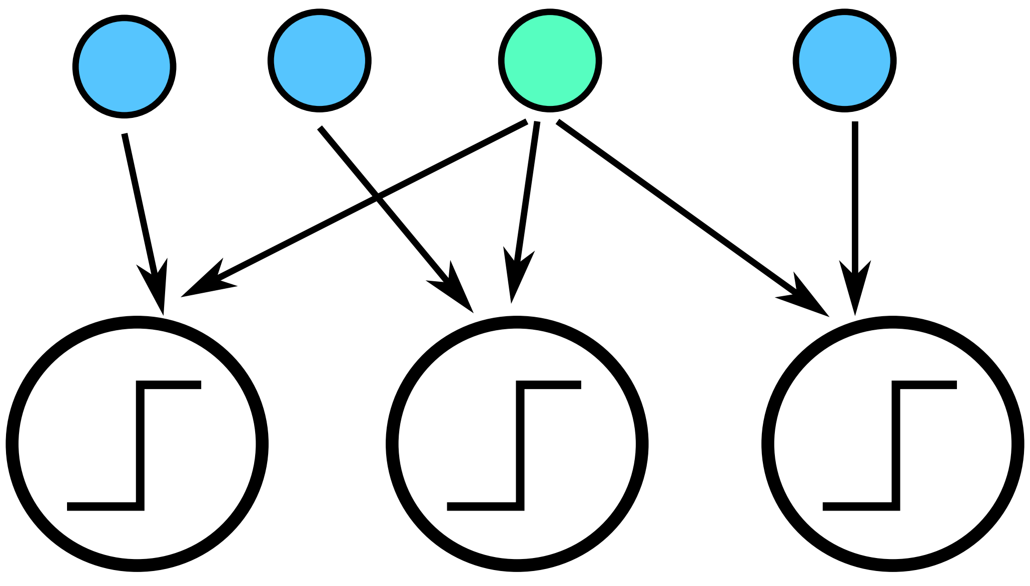
$$s_j = c_j(t), \forall j$$

$$c_j(t) = \int V(\mathbf{x}, t) R_j(\mathbf{x}) d\mathbf{x}$$

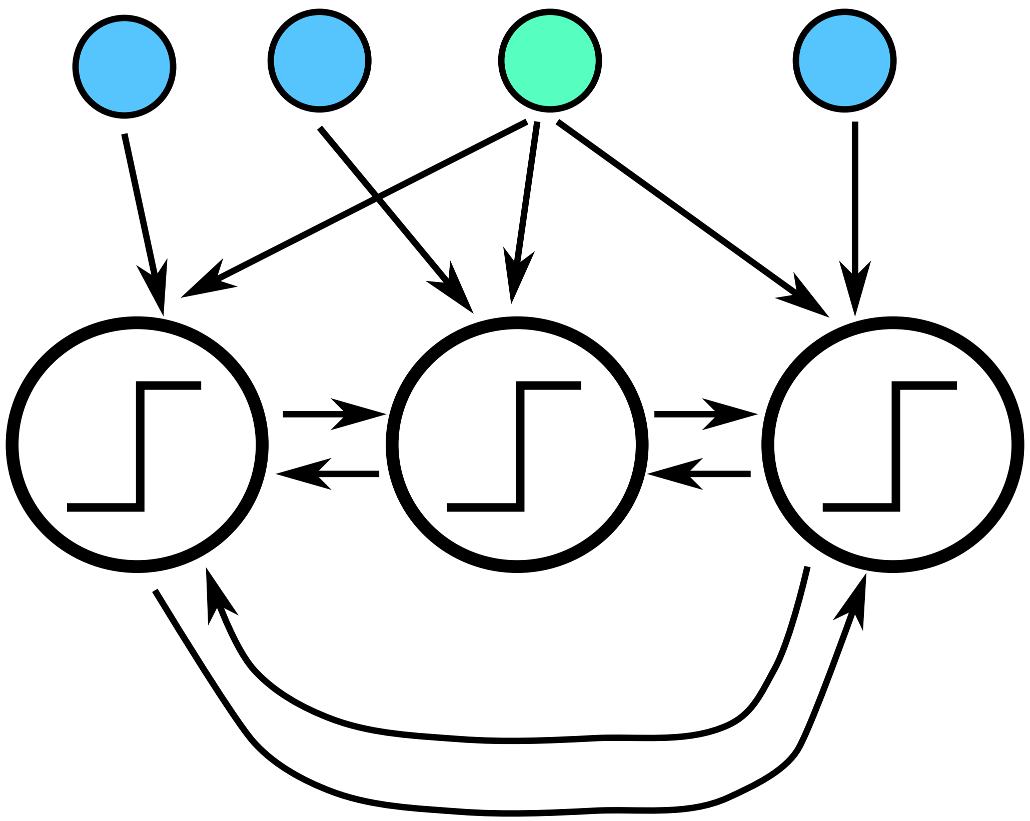
→ ***Retinal circuitry produces nearly pairwise interactions***

Feedforward circuits generate limited higher-order interactions

What if I add recurrence to this circuit?



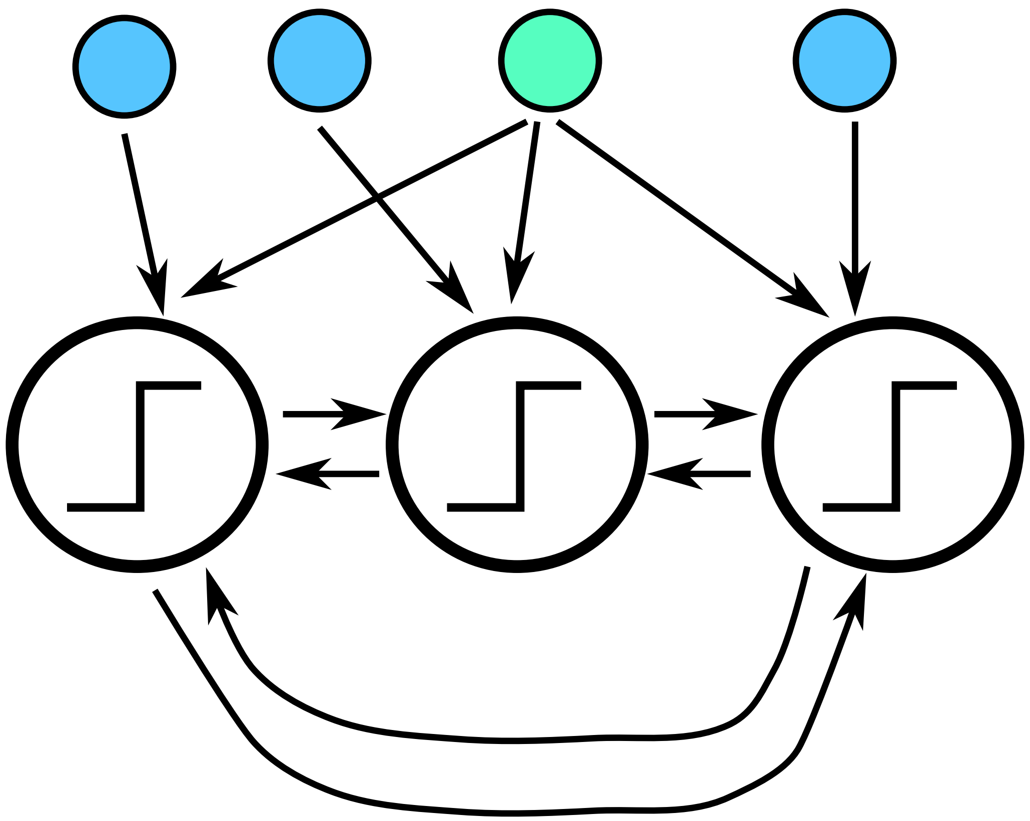
What if I add recurrence to this circuit?



Deterministic, strong, excitatory synapses:
000 -> 000
100 -> 100
110 -> 111
111 -> 111

"Strong synapses":
110 -> 111

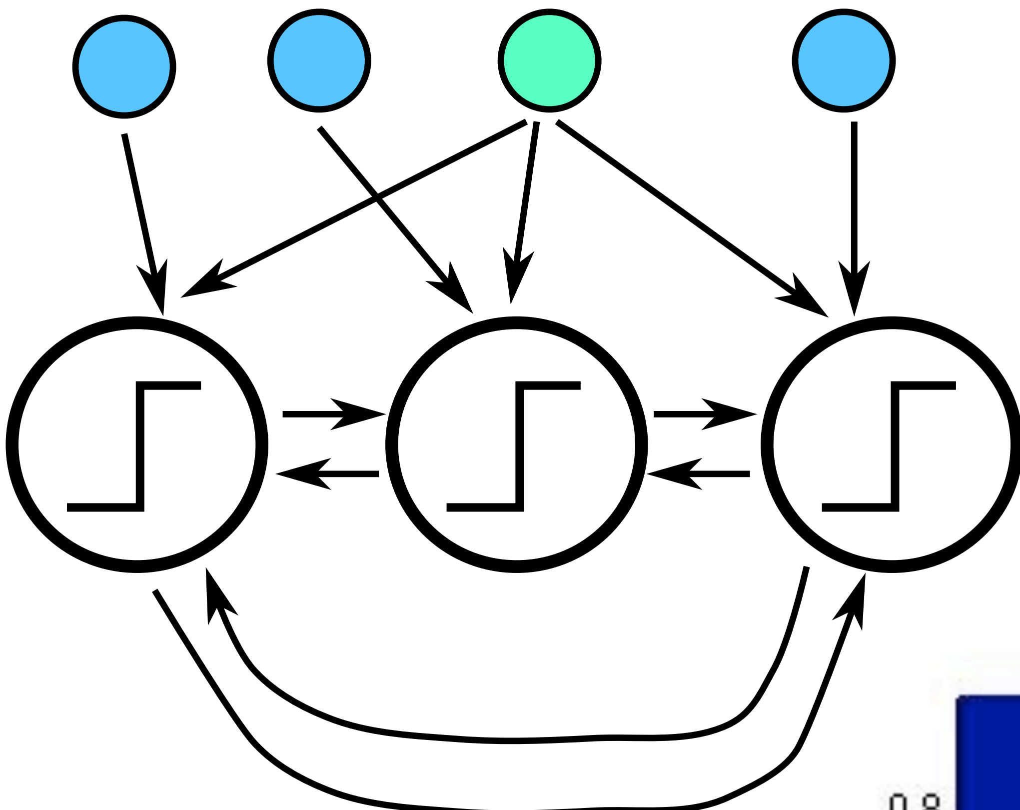
What if I add recurrence to this circuit?



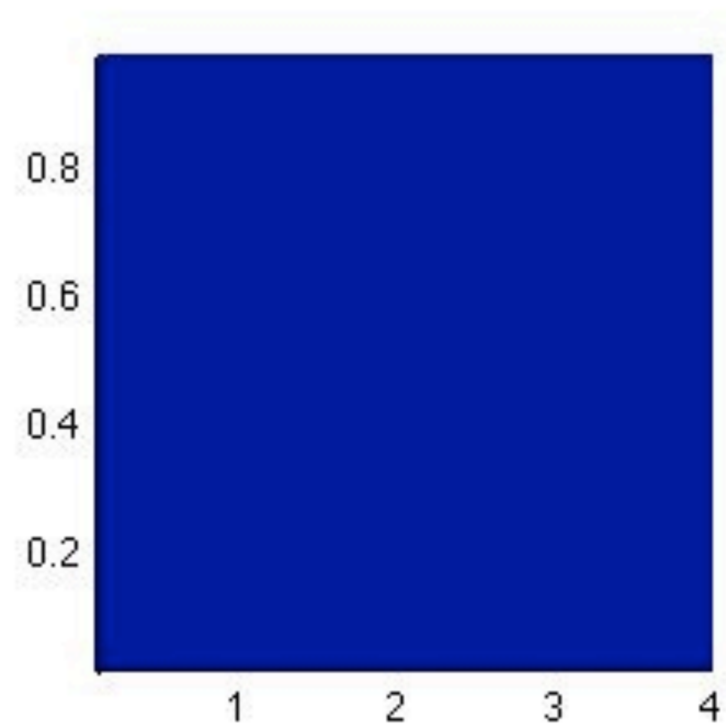
Idealized excitatory synapses:
000 -> 000
100 -> 100
110 -> 111
111 -> 111

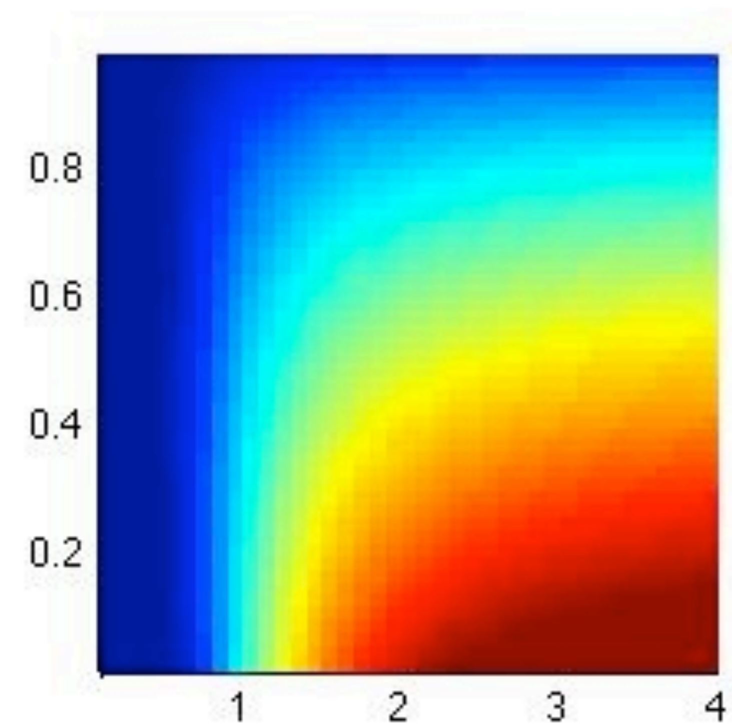
$$\frac{p_3}{p_0} = \left(\frac{p_2}{p_1}\right)^3$$

What if I add recurrence to this circuit?



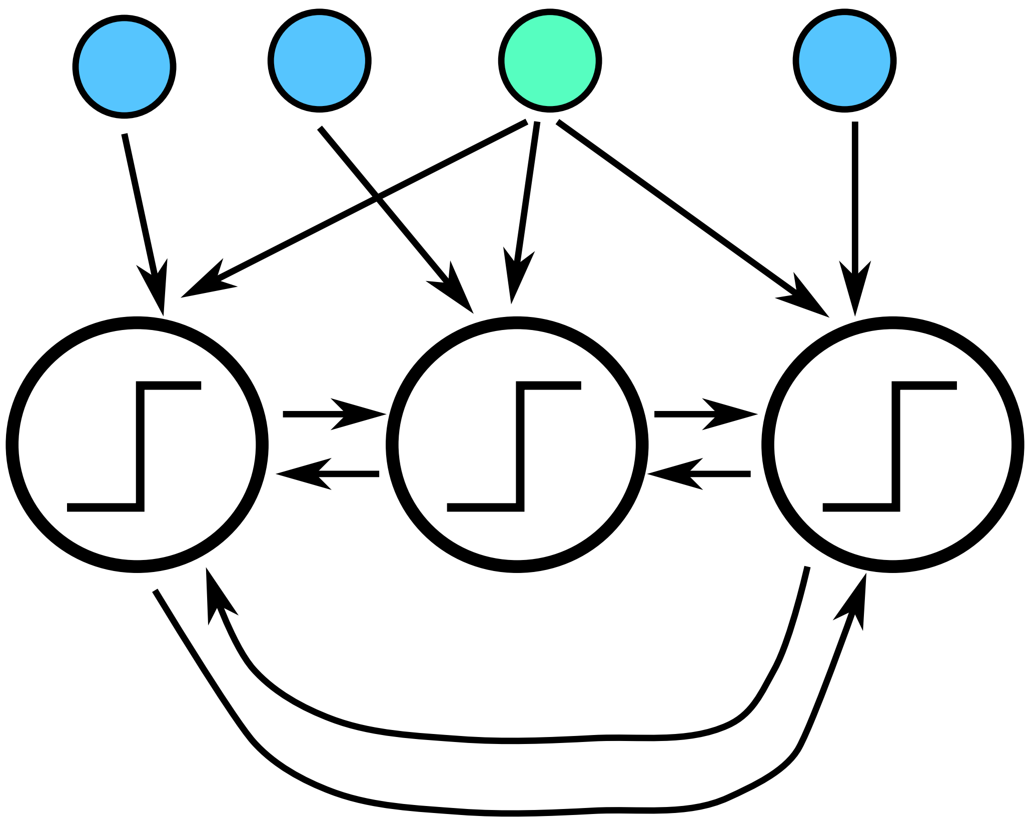
Deterministic, strong, excitatory synapses:
000 -> 000
100 -> 100
110 -> 111
111 -> 111





max(D_{KL}) **= 0.0037** ***= 0.45***

Excitatory synapses that interact with membrane potential?



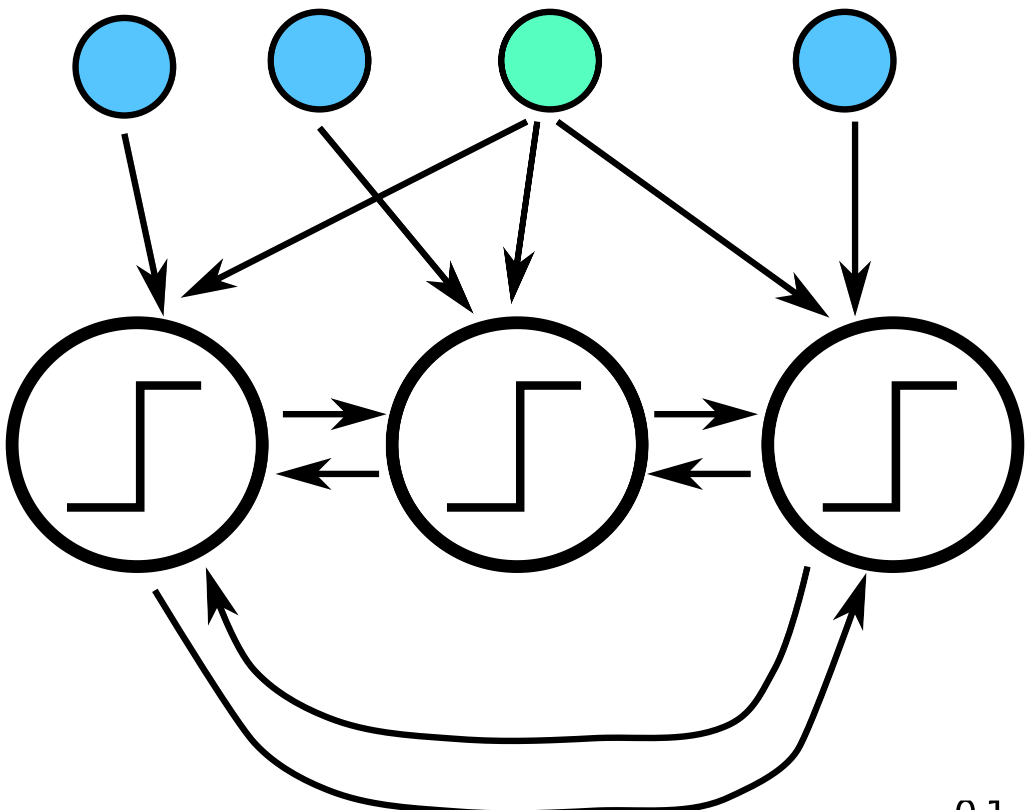
Prediction step...

$$y_j = H\left(I_j + I_c - \theta\right)$$

Decision step:

$$x_j = H\left(I_j + I_c - \theta + q\left(\sum_{k \neq j} y_k \right) \right)$$

Excitatory synapses that interact with membrane potential?

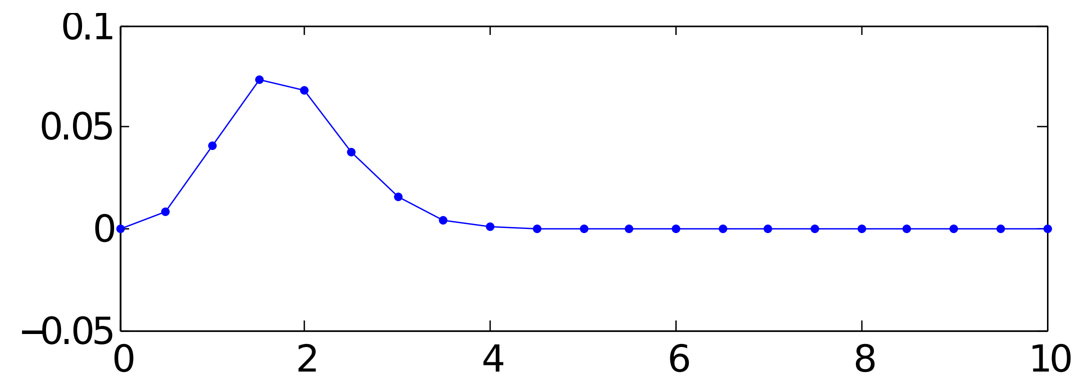


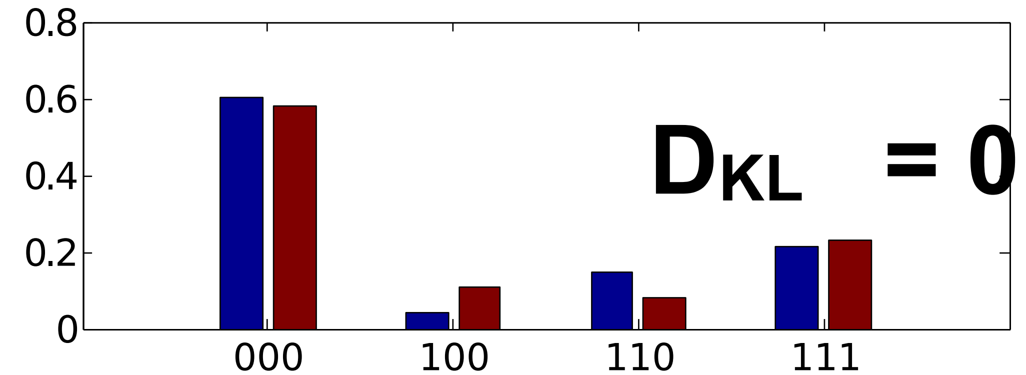
Prediction step...

$$y_j = H\left(I_j + I_c - \theta\right)$$

Decision step:

$$x_j = H\left(I_j + I_c - \theta + q\left(\sum_{k \neq j} y_k\right)\right)$$

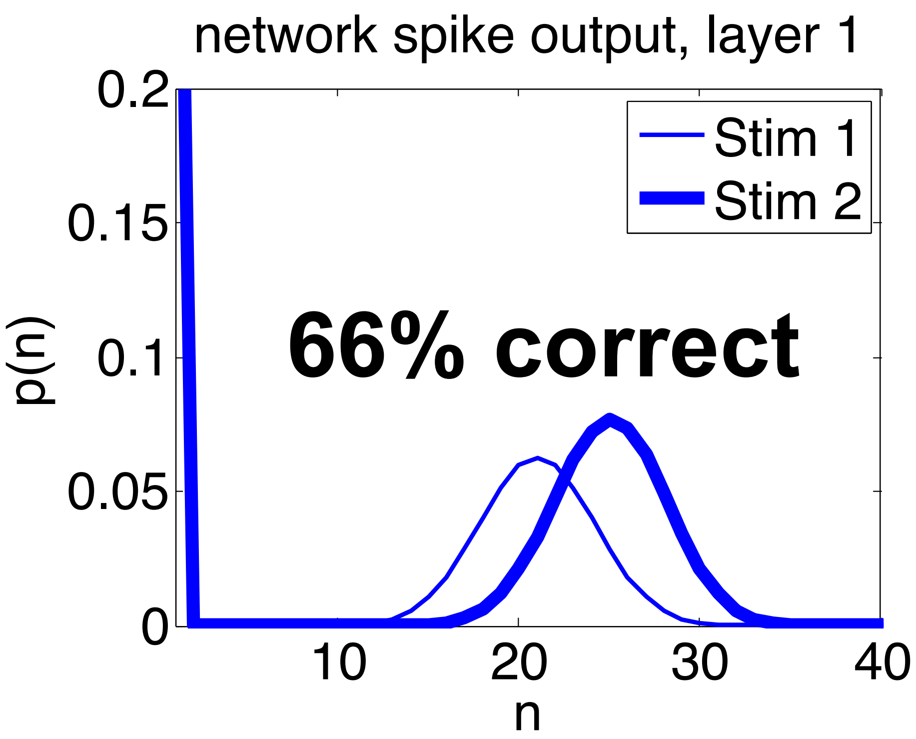


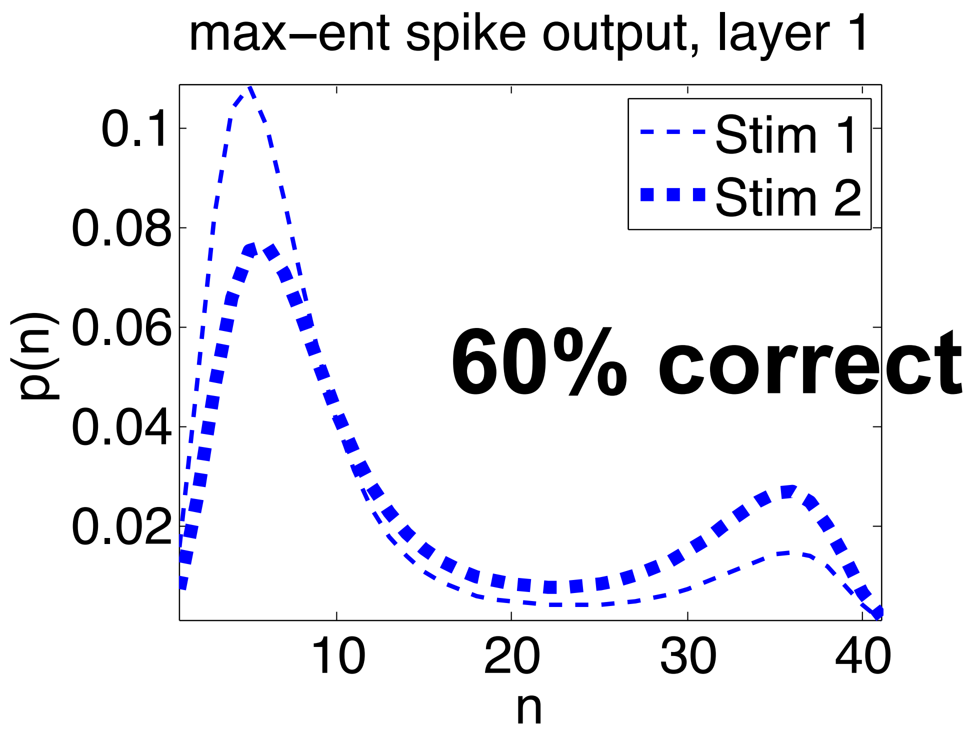


D_{KL} = 0.073

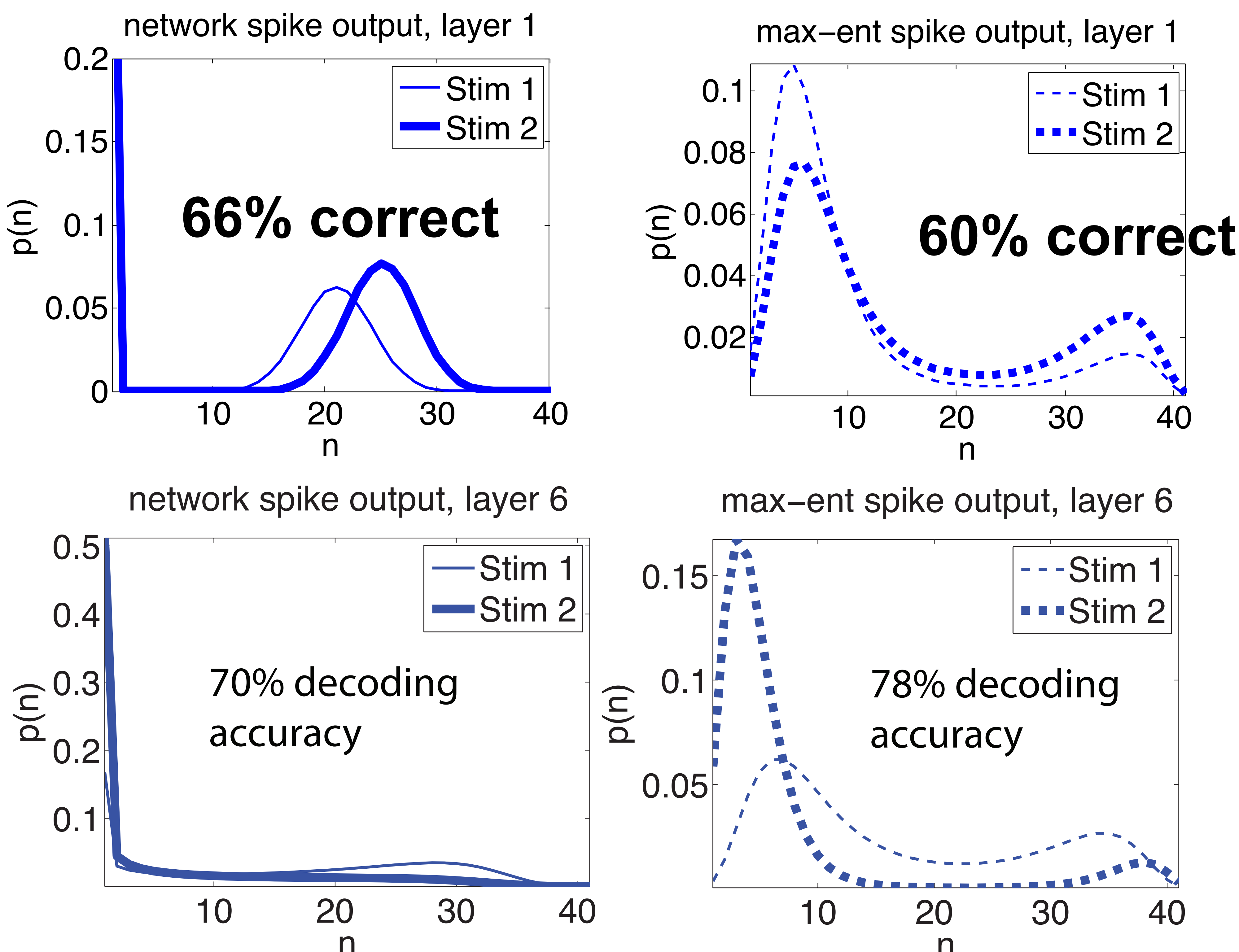
(compare with 0.0038)

Are higher order correlations good for coding??





Are higher order correlations good for coding??



Thank you!

Collaborators:
Eric Shea-Brown (UW)
Fred Rieke (UW)
Julijana Gjorgjieva (Cambridge)
Evan Thilo (UW)

Funding:
Burroughs-Wellcome (ESB)
NSF DMS (ESB)
Howard Hughes Medical Institute (FMR)