Signal processing by stochastic biochemical networks

Ilya Nemenman
CCS-3/CNLS LANL

with

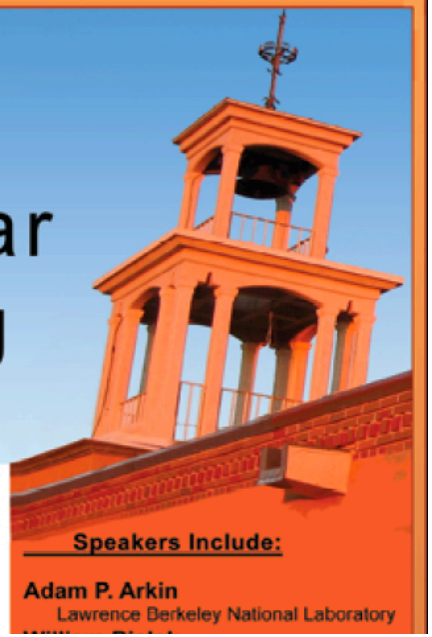
Etay Ziv, Chris Wiggins (Columbia)
Nikolai Sinitsyn (CNLS/CCS-3 LANL)

Center for Nonlinear Studies

The First q-bio Conference on Cellular Information Processing

August 8-11, 2007 | Santa Fe, New Mexico, USA

http://cnls.lanl.gov/q-bio
q-bio@cnls.lanl.gov



Speakers Include:

Adam P. Arkin
Lawrence Berkeley National Laboratory

Mark Goulian
University of Pennsylvania
Jeff Hasty
University of California, San Diego
Jason M. Haugh
North Carolina State University
Alexander Hoffmann
University of California, San Diego
Ravi Iyengar
Mount Sinai School of Medicine
Boris N. Kholodenko
Thomas Jefferson University
Andre Levchenko
Johns Hopkins University
Wendell A. Lim
University of California, San Francisco
Rune Linding
Samuel Lunenfeld Research Institute
John W. Little
University of Arizona
Leslie M. Loew
University of Connecticut Health Center
Bernhard Ø. Palsson
University of California, San Diego
Johan Paulsson
Harvard Medical School
John Reinitz
Stony Brook University
Micha
Uni
Anirv
Rut
Yuhai
IBM
Alexa
Ma
Jose
Slo
H. Ste
Pac
Bridg
Uni
X. Su
Har

Deadlines:

Abstract submission	April 15, 2007
Summer School registration	April 15, 2007
Early registration	June 1, 2007

**Travel awards for graduate students and postdocs may be available.*
More information and applications are available on the conference website.

Center for Nonl

The First q-bio Conference Information P

August 8-11, 2007 | Santa Fe

http://cnls.lanl.gov
q-bio@cnls.lanl

First q-bio Conference on Cellular Information Processi
This conference is intended to advance predictive mod
genetic regulatory systems. The emphasis is
experimentation for the purposes of understanding
particular regulatory systems and of elucidating gene
information processing.

The single-track program will include invited talks
theoretical researchers as well as shorter talks, pos
demonstrations selected from contributed submissio
banquets, six sessions covering a range of topics, ar
sessions.

There will be an opportunity for selected participants
presentations made at the conference to a special is
journal indexed by ISI and PubMed.

Lodging is available for participants on the campus of St. John's College, which should facilitate interactions and stimulating informal discussions about quantitative biology. Space is available for 200 participants. In the event that registration demand exceeds capacity, preference will be given to individuals selected to present contributed talks or posters. Abstracts should be submitted for review via the conference web site.

First q-bio Summer School on Cellular Information Processing
This school is designed for researchers new to modeling cellular regulatory systems. It will take place in Los Alamos from July 23 to August 7. Participants will attend daily lectures about signal transduction, gene regulation and stochastic effects in biochemical networks and work in small teams on selected research projects. Tuition includes conference registration.

Deadlines:

Abstract submission	April 15, 2007
Summer School registration	April 15, 2007
Early registration	June 1, 2007

**Travel awards for graduate students and postdocs may be available.*
More information and applications are available on the conference website.

Organizing Committee: Jeremy S. Edwards (University of New Mexico); James R. Faeder, William S. Hlavacek, Yi Jiang, Ilya Nemenman, and Michael E. Wall (Los Alamos National Laboratory).

Advisory Committee: William Bialek (Princeton University); Byron Goldstein, John E. Pearson, William H. Press, David H. Sharp, and Pat J. Unkefer (Los Alamos National Lab); Michael A. Savageau (University of California, Davis)

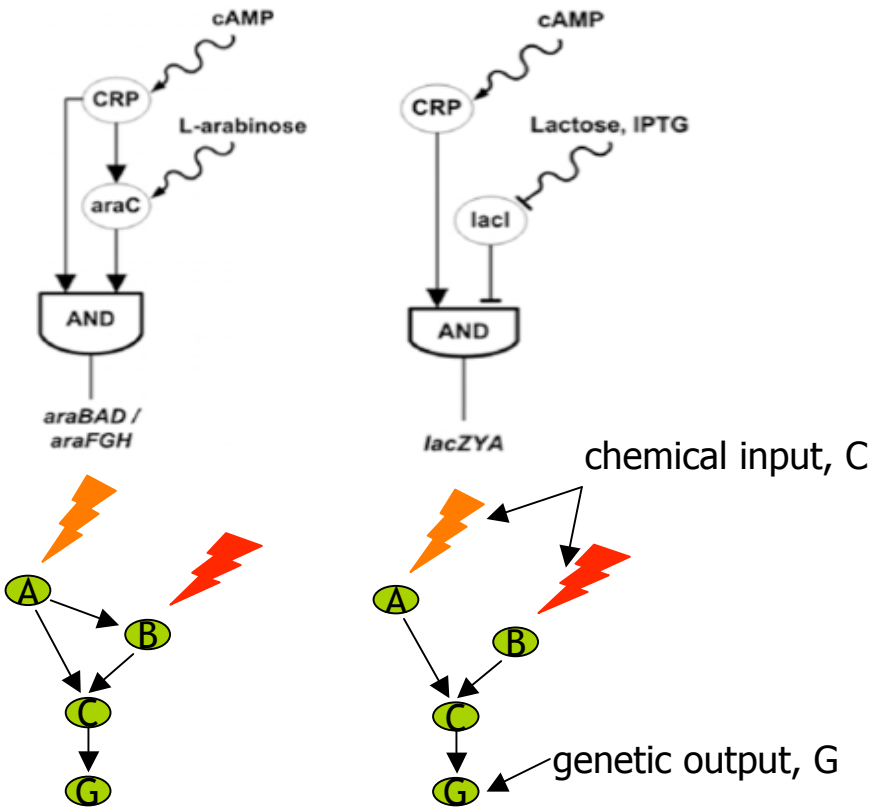
Los Alamos NATIONAL LABORATORY

New Mexico Consortium's LAS Institute for Advanced Studies at Los Alamos National Laboratory

Center for Nonlinear Studies

Images are of St. John's College, the conference location

Signal processing by small networks: Does topology have a function?

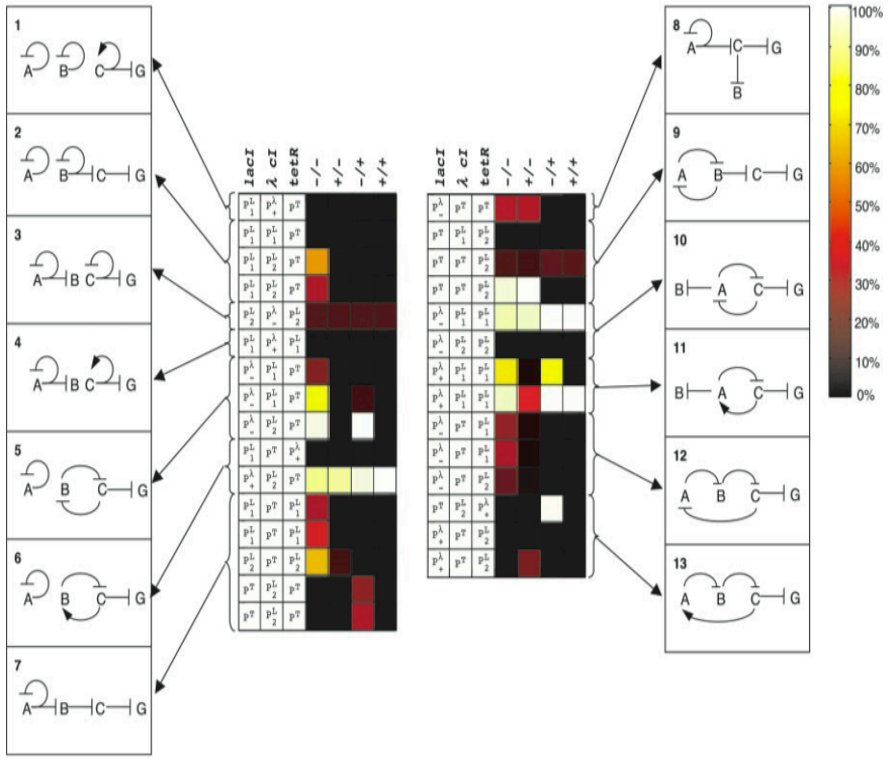


Multiple functions (Wall et al.)
Stochasticity?

A Sign-Sensitive Delay

From Mangan et al., 2003

Signal processing by small networks: Are some networks better than others?



What if wrong parameters were explored?

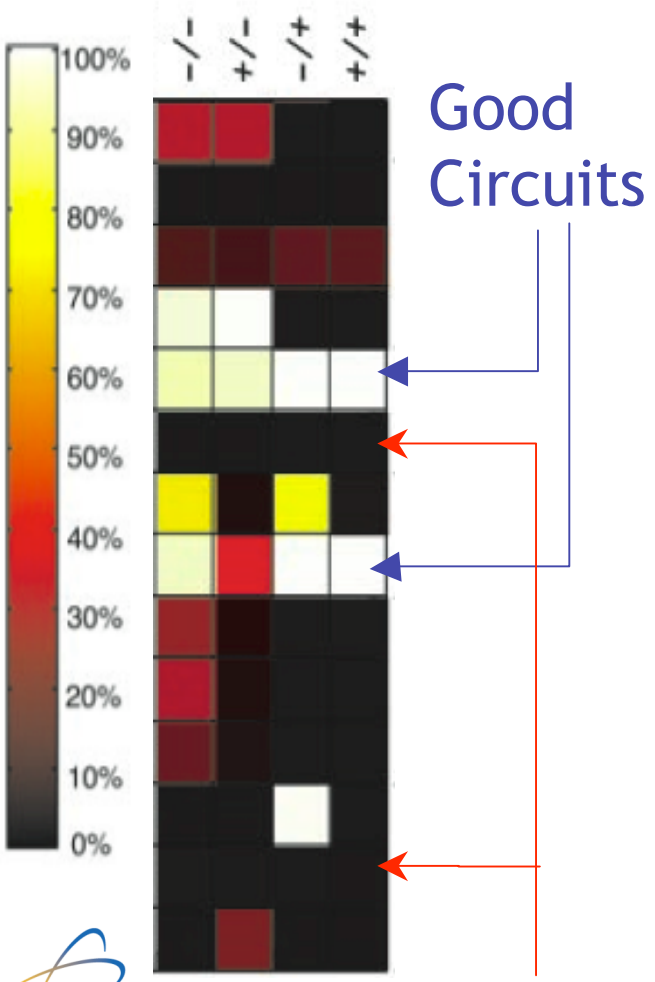
Logic Gates

From Guet et al., 2002

Signal processing by small networks: How much info can be transduced?

- Cross-talk "paradox"
  - Single 2-state MAPKKK (channel capacity of 1 bit)
  - Multiple *on/off* signals (>1bit) passing through
- How?
  - Compartmentalization, extra signals, timing…
  - Concentration of MAPKKK is real-valued! (multi-bit)
- Only ~100 molecules to make a decision
  - Noisy
  - How many bits can be sent with a few molecules?

How to measure circuit quality without knowing its function?



$$g = g(c,t)\big|_{t\to\infty} + \text{noise} \quad \Rightarrow \quad P(g \mid c)$$

$$I(C,G) = \int dc\, dg \;\; p(c,g) \log \frac{p(c,g)}{p(c)p(g)}$$

$$0 \le I(C,G) \le \min\{S(C), S(G)\}$$

Guet example:
C={(0,0),(1,0),(0,1),(1,1)}
G={+1, -1}
Broken circuit: $I(C,G)=S(G)=0$

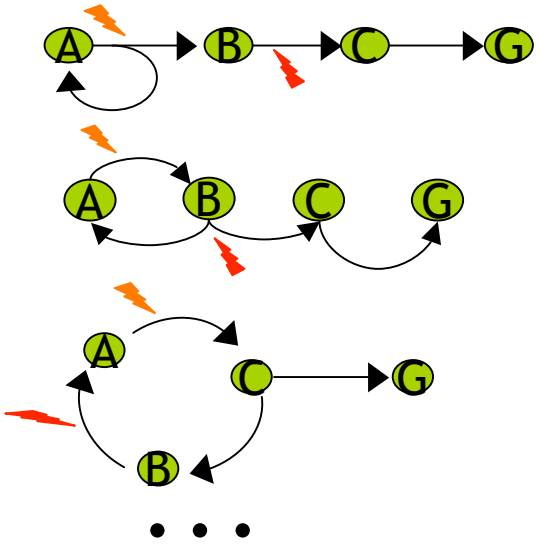
What hides beneath?

- Circuits may not have oscillations
- Circuits may have multiple fixed points
- Fixed points may have different basins of attraction
- What defines *P(c)*?

How good are circuits?

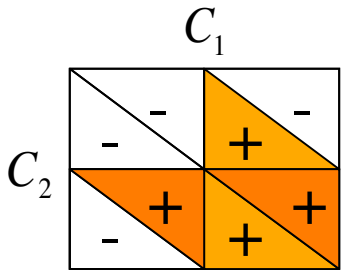
1. For a given topology,

exactly one promoter per gene,
each TF binds to one promoter type



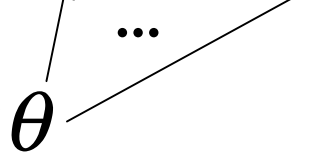
2. For a given $p(C)$,

each input is binary



3. Calculate $g=g(c)$ for all $c \subset C$

$$\frac{dg}{dt} = -R_g g + a_0 + \alpha\left(\{g, c\}\right)$$



(actually do for 3 inputs)

How good are circuits?

4. And maximize information.

$$\hat{\theta} = \arg\max_{\theta \subset \begin{pmatrix}\text{biologically}\\ \text{realistic}\end{pmatrix}} I(C,G)$$

$$\hat{\theta} = \arg\max_{\theta} I(G,C) - \lambda\langle N\rangle$$
$$\hat{\theta} = \arg\max_{\theta} I(G,C) - \gamma\,\tau_{\max}/\tau_{\min}$$
$$\hat{\theta} = \arg\max_{\theta} I(G,C) - \lambda\langle N\rangle - \gamma\,\tau_{\max}/\tau_{\min}$$

Constraints on time to and the copy # at the steady state.

max= high fidelity differentiation in development
high capacity signal transduction (*lac*, photoreceptor)

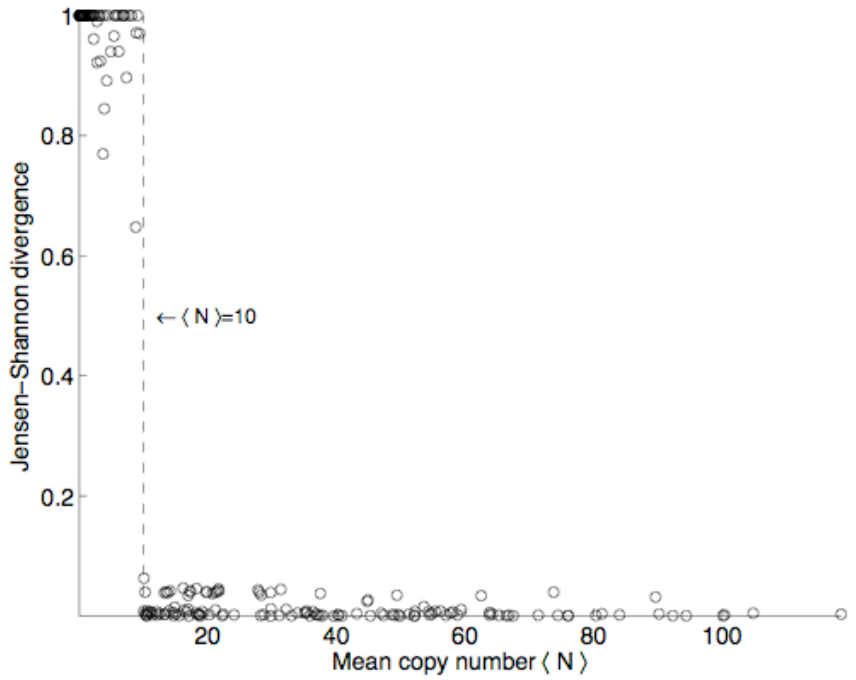
5. How does max(I) depend on constraints?
On the topology?

Linear noise: How good is it?

- Poisson reactions

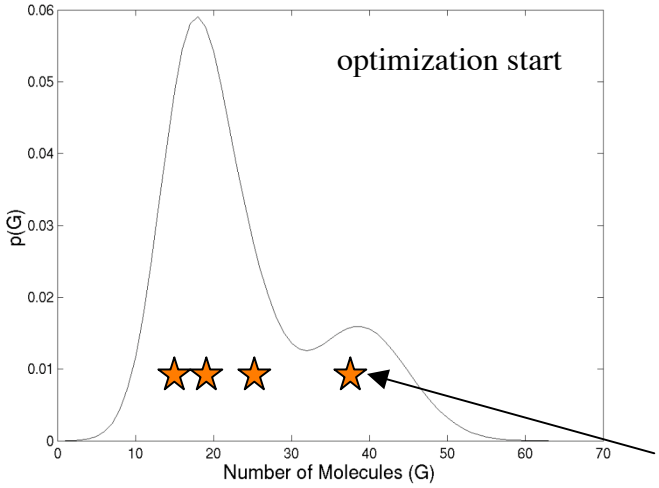
$$n_i - 1 \xleftarrow{r} n_i \xrightarrow{\text{Hill}} n_i + 1$$

- Master equation with large N
- Fokker-Planck equation
- Steady state
- Steady state $P(g|c)$: multivariate normals

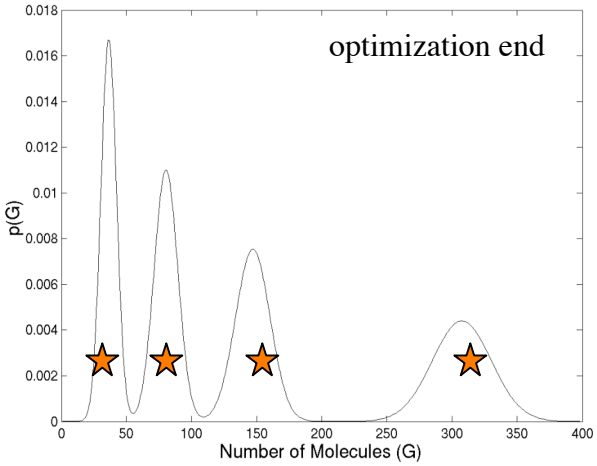


Van Kampen, 1997
Elf and Ehrenberg, 2003
Paulsson et al., 2004

Numerics: increasing MI

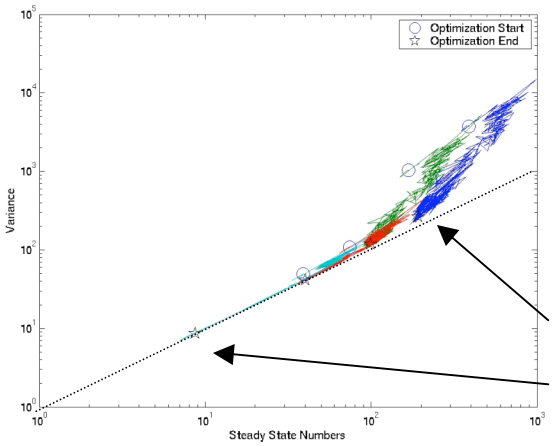


separating peaks



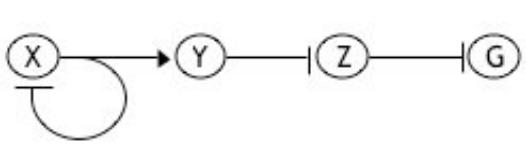
Mean outputs in response to different inputs

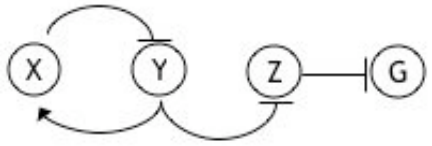
decreasing the reporter variance to the Poisson limit (low pass filtering upstream noise by slow reporter); variance of the other species may be sub-Poisson (negative feedback)

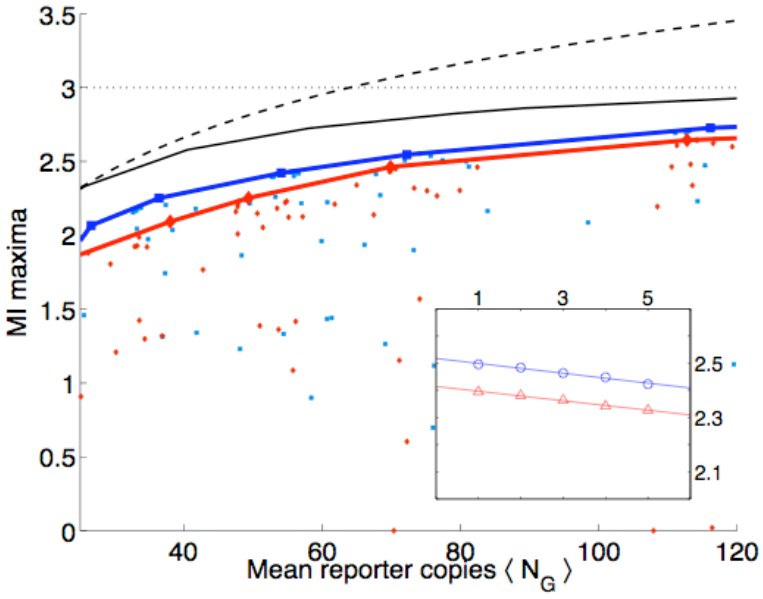


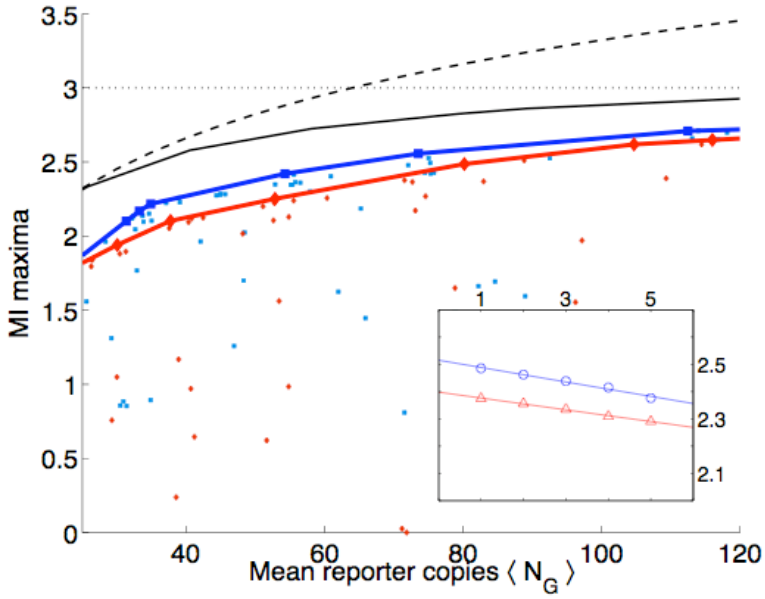
other species reporter

Specific circuits: more than 1 bit, almost optimal



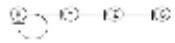
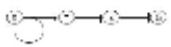
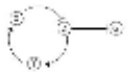
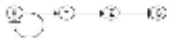


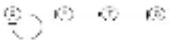
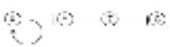
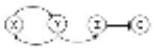
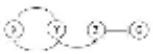
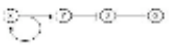
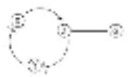
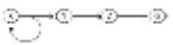




Maxima: analytics and numerics

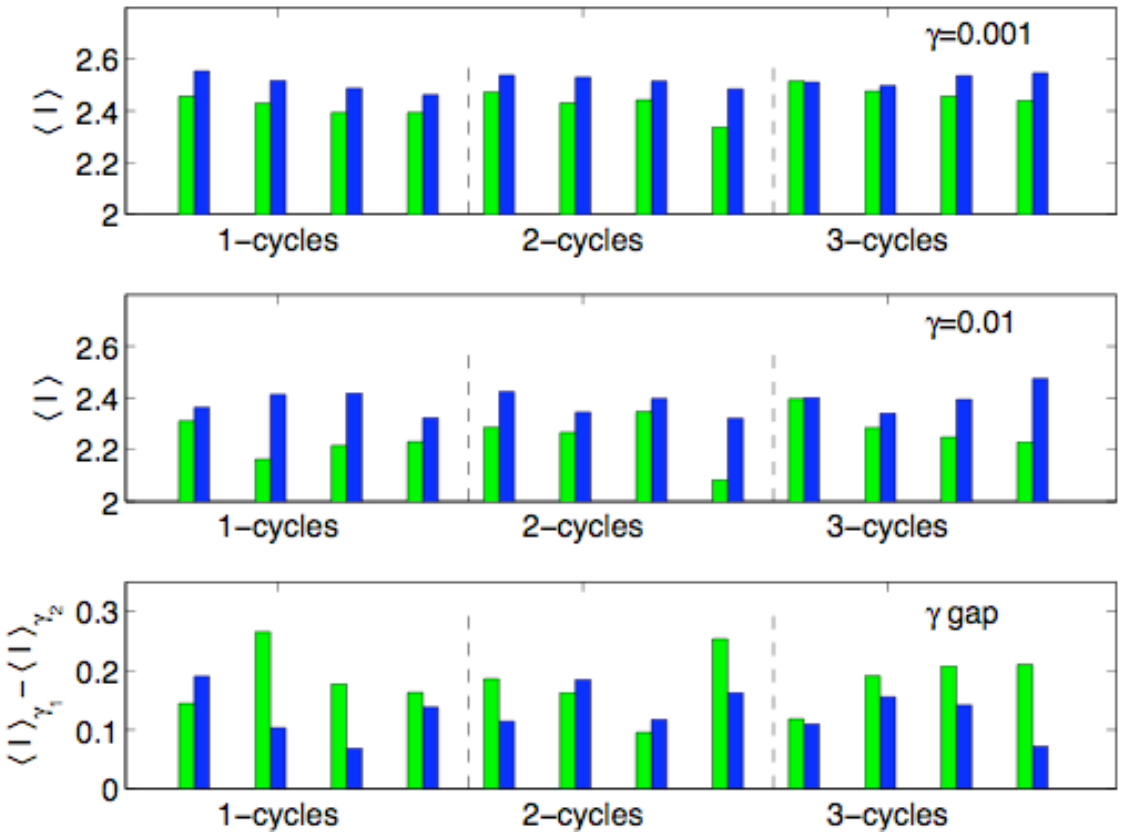
Is topology important?

Number	Topology	$\gamma = 0.001$	$\gamma = 0.01$
1		2.5570	2.3638
20		2.5524	2.3970
6		2.5451	2.4818
2		2.5357	2.3549
22		2.5354	2.3909
19		2.5218	2.3718
10		2.5172	2.3925
13		2.5055	2.4058
8		2.5002	2.3463
23		2.4976	2.3831
14		2.4874	2.4251
12		2.4809	2.3219

Number	Circuit	$\gamma = 0.001$	$\gamma = 0.01$
17		2.4695	2.2876
5		2.4659	2.2806
4		2.4624	2.2930
21		2.4605	2.23121
9		2.4497	2.3491
7		2.4420	2.2773
15		2.4244	2.1587
24		2.4234	2.2123
11		2.3958	2.2143
16		2.3943	2.2281
18		2.3603	2.0751
3		2.3099	2.2471

All are great!
Some are better than others

Positive vs. negative feedback



negative
positive

$p = 0.0002$

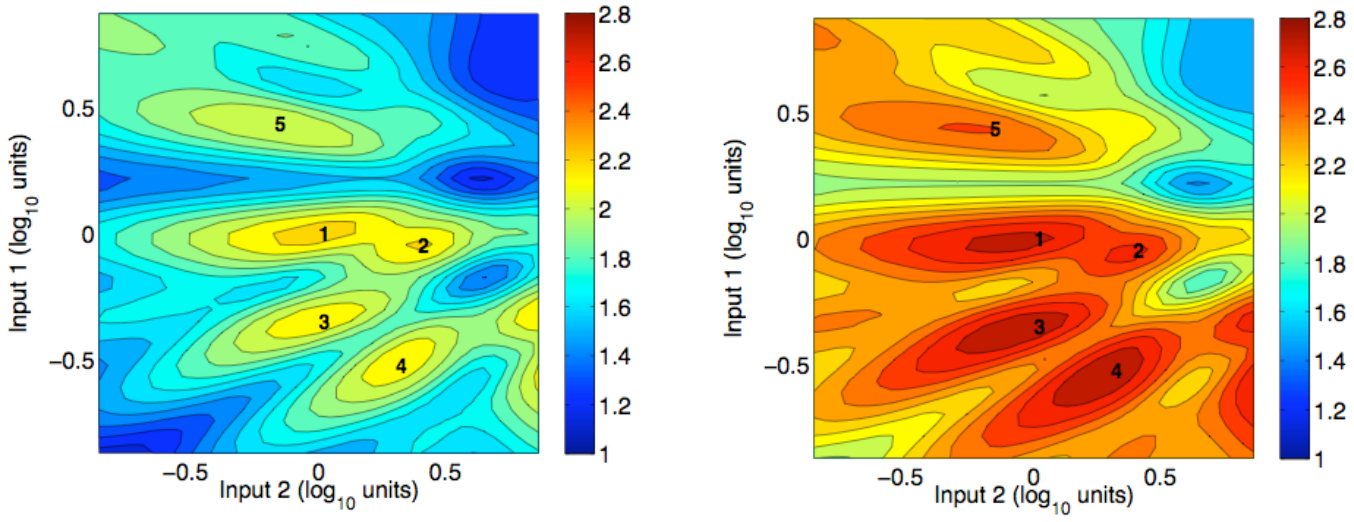
$p = 0.0003$

$p = 0.01$

NF circuits have higher capacity and reach it easier

Explanation available in terms of decreased state variance

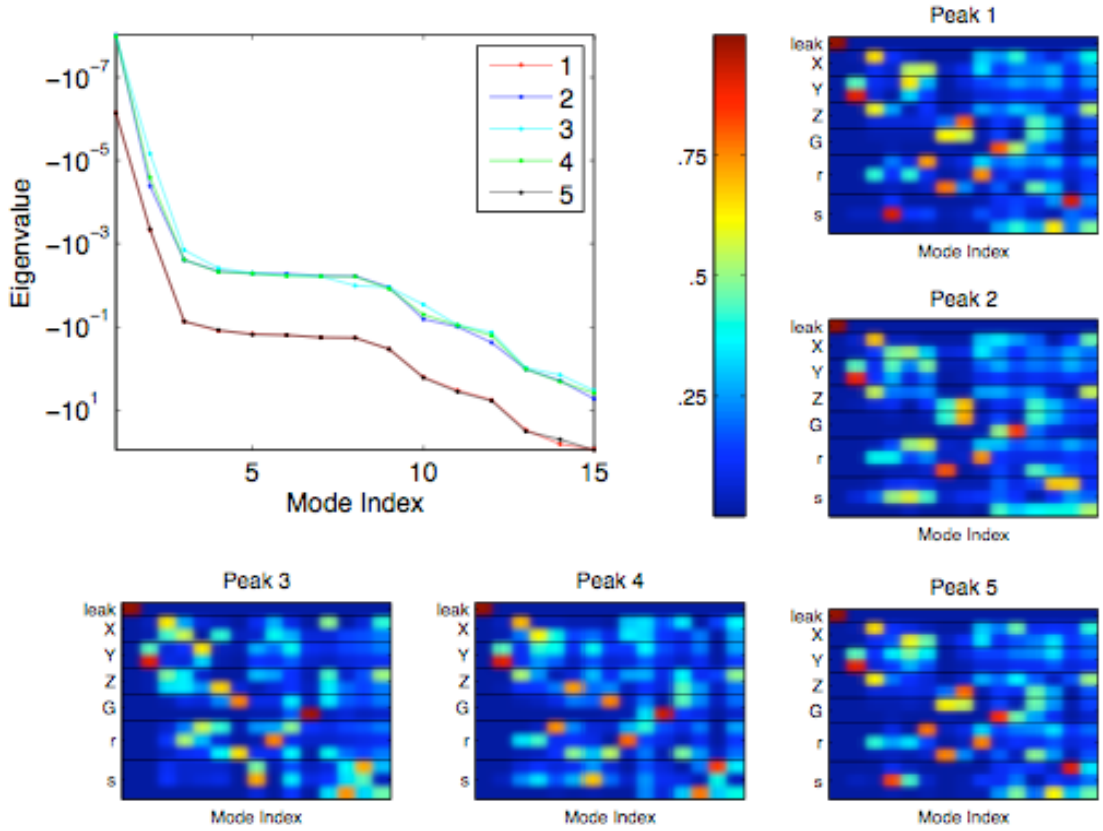
Multiple functions?



Topology 2

Chemical State	000	001	010	011	100	101	110	111
Peak 1	2	6	1	5	4	8	3	7
Peak 2	2	6	4	1	5	8	3	7
Peak 3	2	1	4	6	3	5	8	7
Peak 4	2	1	6	4	5	3	8	7
Peak 5	6	2	5	1	8	4	7	3

Robust maxima?



Predictions

Fast response and autorepression - correlated
Rosenfeld et al. (2002) - autorepression causes fast response
Alternative: Fast response *requires* negative feedback (cannot average)

	Negative Feedback	No Negative Feedback
Proteolysis	9	4
No Proteolysis	44	88

$p=0.013$

Conclusions 1

- Small, noisy, generic biochemical networks easily achieve >1 bit of information throughput over short times with a handful of molecules.
- The circuits come very close to transmitting the maximum information.
- No fine tuning is required.
- While all circuits are good, negative feedback circuits are marginally better (skipped in this talk).
- Multiple functions per circuit (more exploration is needed).

How good is analysis?

- A multi-step transcription/translation/binding reaction modeled as a single-step elementary one

$$n_i - 1 \xleftarrow{r} n_i \xrightarrow{\text{Hill}} n_i + 1$$

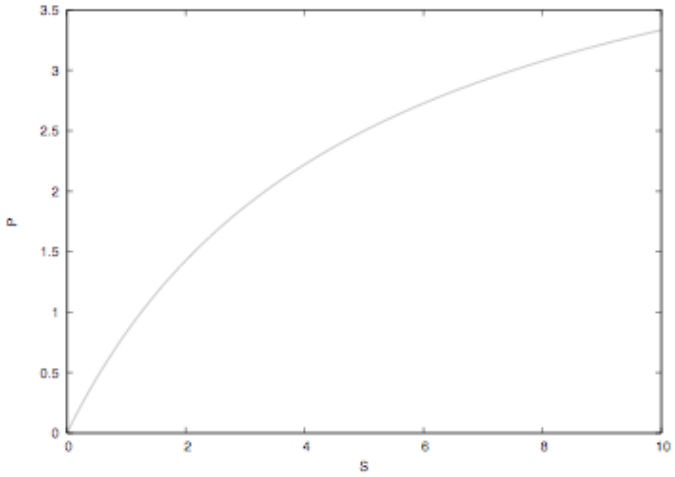
- Is this valid?
- In general, how do we coarse-grain biochemical reactions? (modeling each one is infeasible)
- What is the right way to simulate a biochemical network?

Michaelis-Menten reaction: Deterministic coarse-graining

$$S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \xrightarrow{k_2} P$$

$$\frac{d[SE]}{dt} = k_1[S][E] - (k_{-1} + k_2)[SE] = 0$$

$$\frac{dP}{dt} = \frac{k_1 k_2 [E][S]}{k_2 + k_{-1} + k_1[S]} = J_{cl}$$

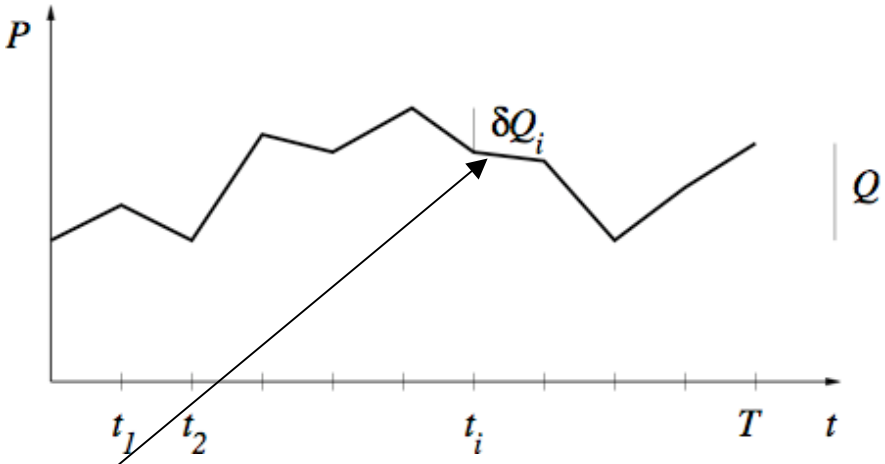


- Adiabatic approximation
  - Many enzyme turnovers for small fractional change in [*P*], [*S*]
- How to do coarse-graining with fluctuations?

Michaelis-Menten reaction (or a pore): Stochastic coarse-graining

$$S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} P$$

$$Q = P(T) - P(0)$$



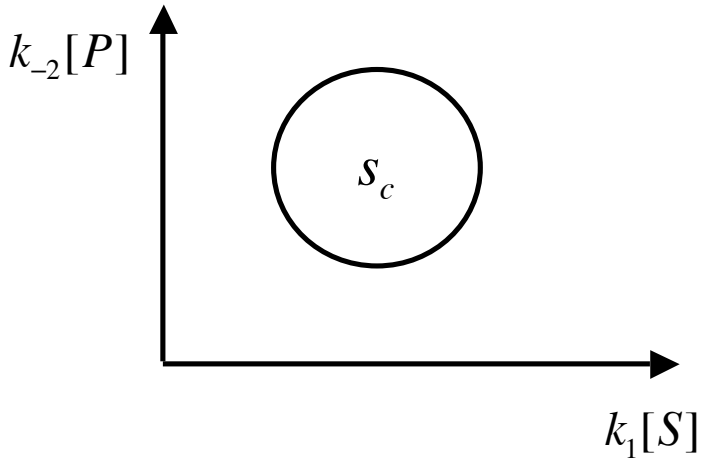
4 Poisson processes
with (almost) constant rates k_i

$$P(Q) = \int d\delta Q_1 \int d\delta Q_2 \ldots \int d\delta Q_{T/\delta t} \prod P(\delta Q_i)\,\delta\left(Q - \sum \delta Q_i\right)$$

Functional integral over all paths - can get full MGF

(Simper version of Sinitsyn and Nemenman, 2007)

Michaelis-Menten reaction: Periodic modulation of two rates

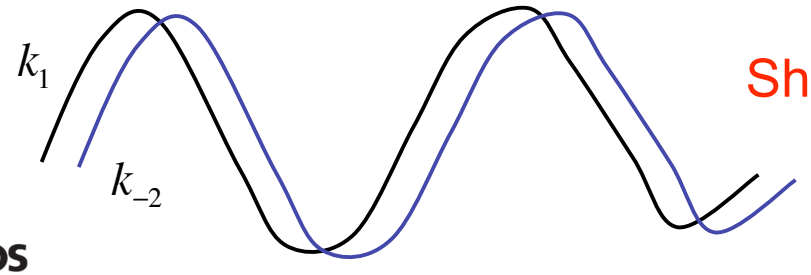


$$\mathrm{MGF} = \mathrm{MGF}_{\mathrm{cl}} + \iint_{s_c} dSdP\ F$$

Berry curvature
As in adiabatic QM

$$J = J_{cl} + J_{pump} = J_{cl} + \iint_{s_c} d^2k \frac{k_2 + k_{-1}}{T_0\left(\sum k_i\right)^3}$$

Pump, ratchet

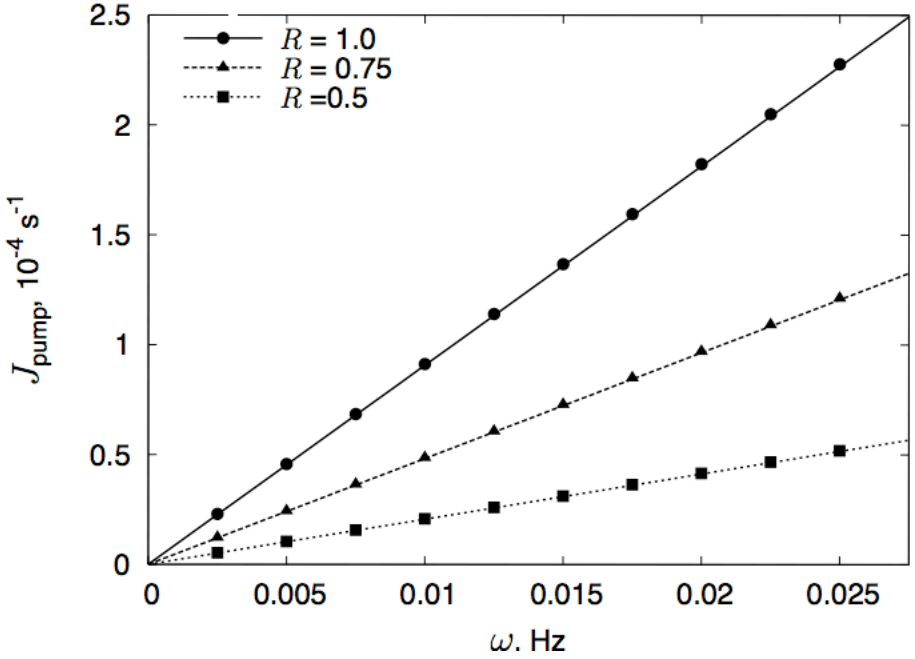


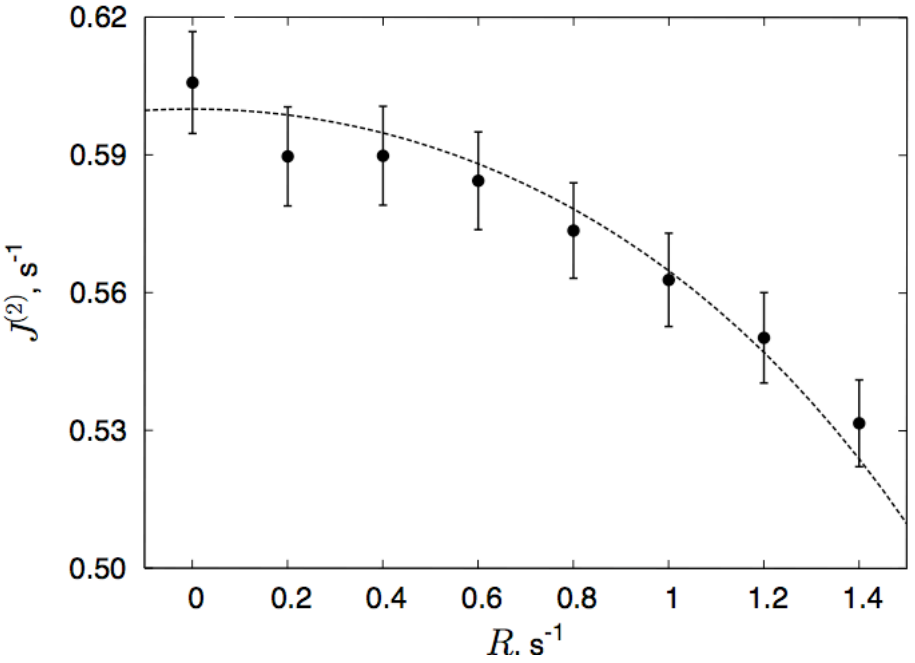
Shielding

Example 1: Bulk

$$k_1 = 1.5 + R\cos\omega t; \qquad k_{-2} = 1.5 + R\sin\omega t; \qquad k_{-1} = k_2 = 1$$

equilibrium, on average: $J_{cl} = 0$



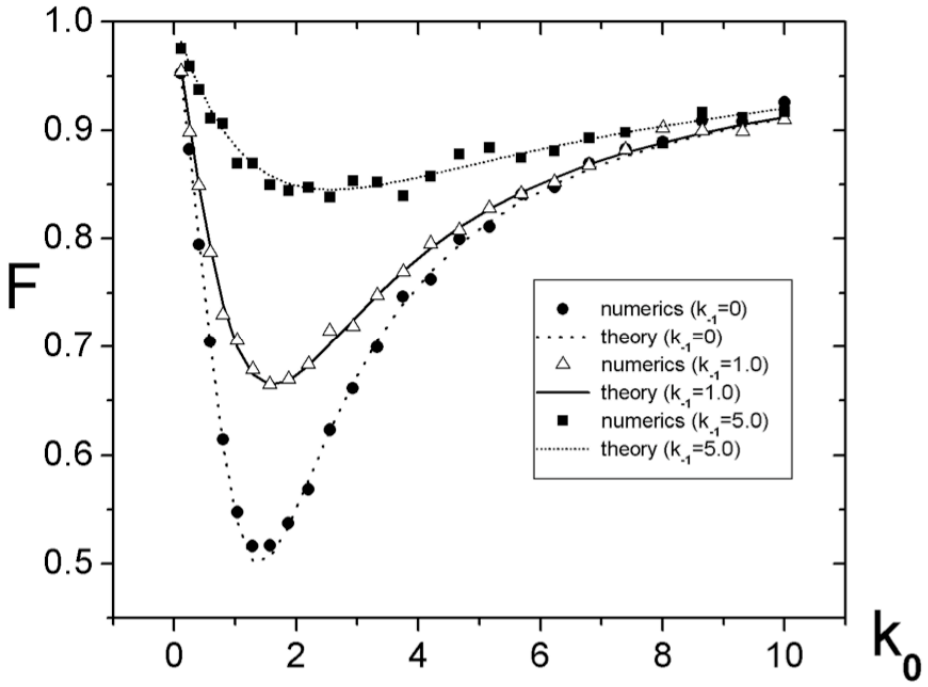


Pump current up to 10% for realistic enzymes

Example 2:
Single molecule experiments

$$\varnothing \underset{q}{\overset{k_0}{\rightleftharpoons}} S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \xrightarrow{k_2} P$$

Xie et al.
Bezrukov et al.



Conclusions 2

- Can coarse-grain biochemical reactions
- Pump effect (nonzero mean noise)
- Fano factor non-unity: non-Poisson statistics