Signal processing by stochastic biochemical networks

Ilya Nemenman
CCS-3/CNLS LANL

with

Etay Ziv, Chris Wiggins (Columbia)
Nikolai Sinitsyn (CNLS/CCS-3 LANL)
First q-bio Conference on Cellular Information Processing

This conference is intended to advance predictive molecular genetic regulatory systems. The emphasis is experimentation for the purposes of understanding particular regulatory systems and of elucidating general information processing.

The single-track program will include invited talks by theoretical researchers as well as shorter talks, poster demonstrations selected from contributed submissions, and a banquet. Six sessions covering a range of topics, and selected sessions will be open to the public.

There will be an opportunity for selected presentations made at the conference to be submitted for review via the conference website.

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, Xiaohong Jiang, Lynn Nemenman, and Michael E. Wall (Los Alamos National Laboratory)

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

Speakers Include:
- Adam P. Arkin
- Lawrence Berkeley National Laboratory

Central 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
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?

Good Circuits

Bad Circuits

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

\[ I(C, G) = \int dc \, dg \hspace{1em} p(c, g) \log \frac{p(c, g)}{p(c)p(g)} \]

\[ 0 \leq I(C, G) \leq \min \{S(C), S(G)\} \]

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 $c \subset C$

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

(actually do for 3 inputs)
How good are circuits?

4. And maximize information.
\[ \hat{\theta} = \arg \max_{\theta \subset \{\text{biologically realistic}\}} I(C,G) \]
\[ \hat{\theta} = \arg \max_{\theta} I(G,C) - \lambda \langle N \rangle \]
\[ \hat{\theta} = \arg \max_{\theta} I(G,C) - \gamma \frac{\tau_{\text{max}}}{\tau_{\text{min}}} \]
\[ \hat{\theta} = \arg \max_{\theta} I(G,C) - \lambda \langle N \rangle - \gamma \frac{\tau_{\text{max}}}{\tau_{\text{min}}} \]
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

- 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)
Specific circuits: 
more than 1 bit, almost optimal

Maxima: analytics and numerics
Is topology important?

All are great! Some are better than others
Positive vs. negative feedback

NF circuits have higher capacity and reach it easier

Explanation available in terms of decreased state variance

negative

positive

\[ p = 0.0002 \]
\[ p = 0.0003 \]
\[ p = 0.01 \]
Multiple functions?

Topography 2

<table>
<thead>
<tr>
<th>Chemical State</th>
<th>000</th>
<th>001</th>
<th>010</th>
<th>011</th>
<th>100</th>
<th>101</th>
<th>110</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak 1</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Peak 2</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Peak 3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Peak 4</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Peak 5</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>
Robust maxima?
Predictions

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

<table>
<thead>
<tr>
<th></th>
<th>Negative Feedback</th>
<th>No Negative Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteolysis</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>No Proteolysis</td>
<td>44</td>
<td>88</td>
</tr>
</tbody>
</table>

$p = 0.013$
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

\[ n_i - 1 \xleftrightarrow{r} n_i \xrightarrow{\text{Hill}} n_i + 1 \]

- 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

\[
S \xrightleftharpoons[k_{-1}]{k_1} SE \xrightarrow{k_2} P
\]

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

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

- 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

\[ S \xrightleftharpoons[k_{-1}]{k_1} SE \xrightleftharpoons[k_{-2}]{k_2} P \]

\[ Q = P(T) - P(0) \]

4 Poisson processes with (almost) constant rates \( k_i \)

\[ P(Q) = \int d\delta Q_1 \int d\delta Q_2 \cdots \int d\delta Q_{T/t} \prod P(\delta Q_i) \delta(Q - \sum \delta Q_i) \]

Functional integral over all paths - can get full MGF

(Simper version of Sinitsyn and Nemenman, 2007)
Michaelis-Menten reaction: Periodic modulation of two rates

\[ MGF = MGF_{cl} + \int \int_{s_c} dSdP \ F \]

Berry curvature
As in adiabatic QM

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

Pump, ratchet

Shielding
Example 1: Bulk

\[ k_1 = 1.5 + R \cos \omega t; \quad k_{-2} = 1.5 + R \sin \omega t; \quad k_{-1} = k_2 = 1 \]
equilibrium, on average: \( J_{cl} = 0 \)

Pump current up to 10% for realistic enzymes
Example 2: Single molecule experiments

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