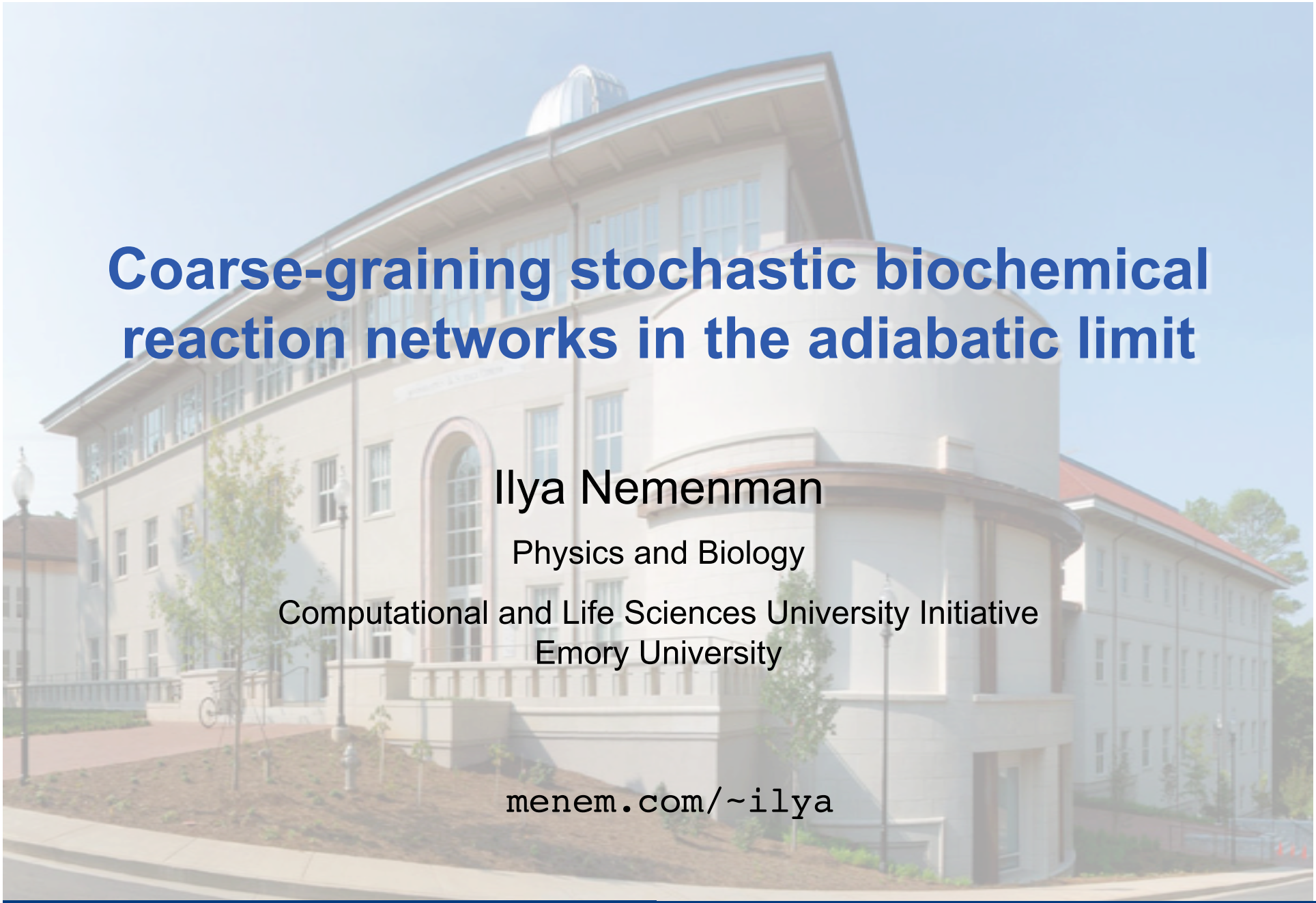


Coarse-graining stochastic biochemical reaction networks in the adiabatic limit

Ilya Nemenman

Physics and Biology

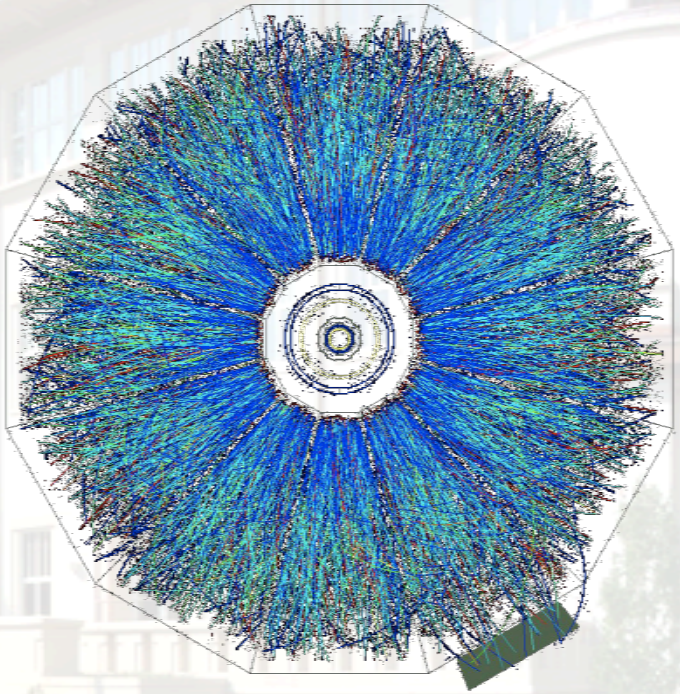
Computational and Life Sciences University Initiative
Emory University

menem.com/~ilya

Thanks

- LANL
  - Nikolai Sinitsyn
  - Brian Munsky
  - Golan Bel (Ben Gurion)
  - Nicolas Hengartner
  - Andrew Mugler (Columbia/ AMOLF)
  - Bryan Daniels (Cornell/ Wisconsin)
  - Wiet de Ronde (AMOLF)
- Emory
  - Sorin Tanase Nicola (Uppsala)
  - Xiang Cheng
  - Lina Merchan
- NIH, NSF, DOE
- LANL, Emory

Hairballs aren't unique to biology



Wikipedia RHIC event image

My prejudices

- Large hairballs of interaction networks are scary and hard to interprete
- Need coarse-grained models – but need rigorous methods to decide what matters and what doesn't
- There is no reason to solve an approximate problem exactly
- Need approximate methods for coarse-graining – need to find small parameters with the most bang for the buck

Goals

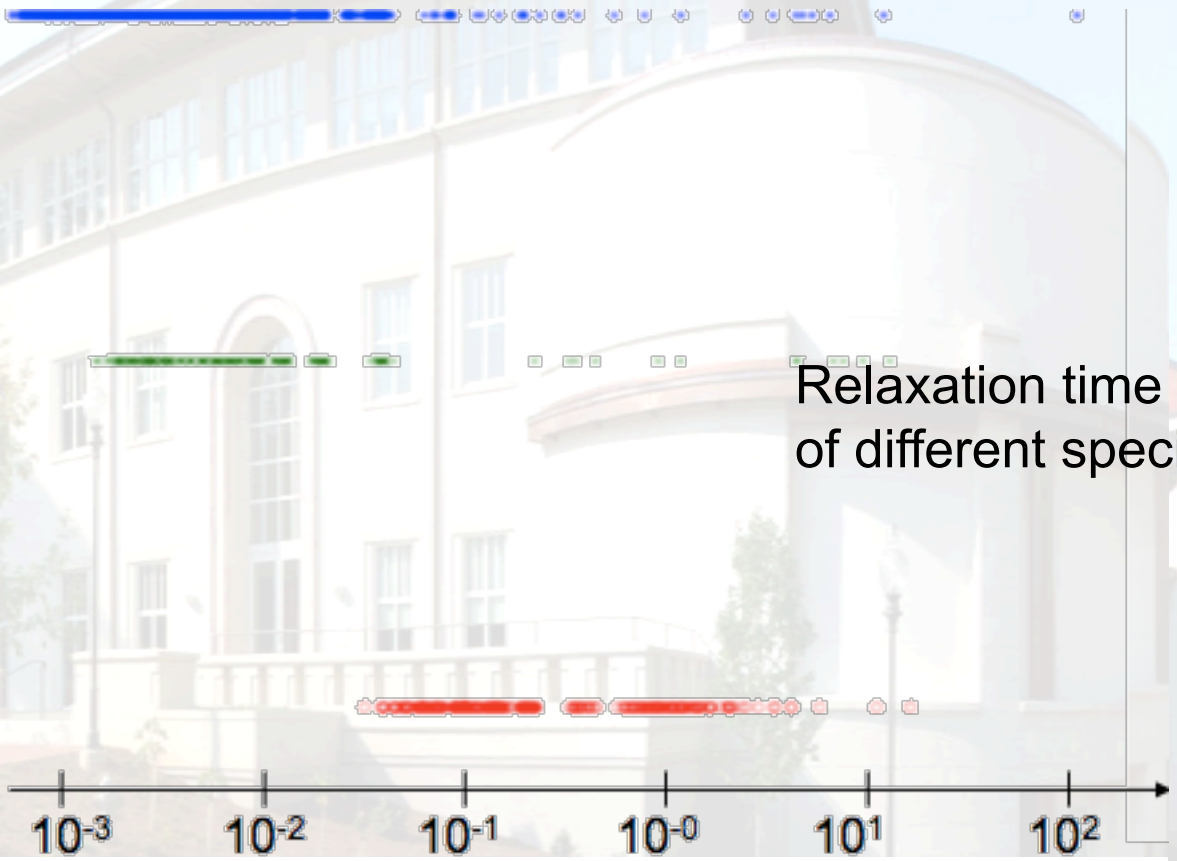
- Is functionally equivalent coarse-graining possible on long temporal/spatial scales?
- Can we build faster simulation tools based on coarse-grained systems?
- Do interesting phenomena emerge due to coarse-graining?

Small parameter in biochemical networks

FcεRI (trimer)
2954 states

FcεRI (dimer)
354 states

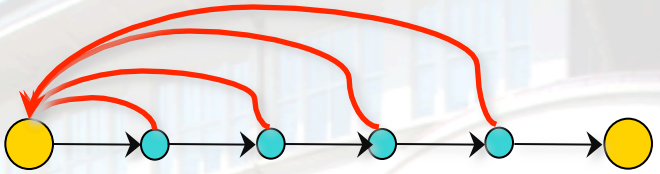
EGFR
356 states



Relaxation time scales
of different species

Faeder et al.

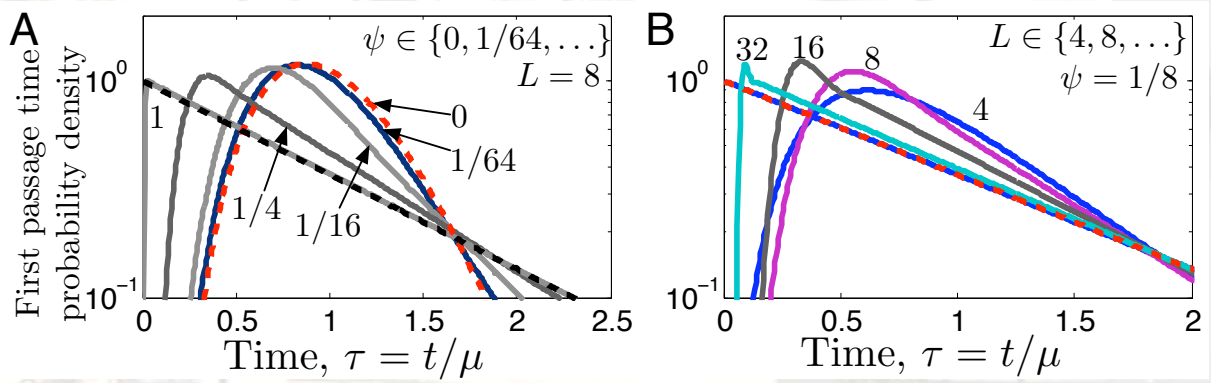
Example of time scales separation: Kinetic proofreading



k forward

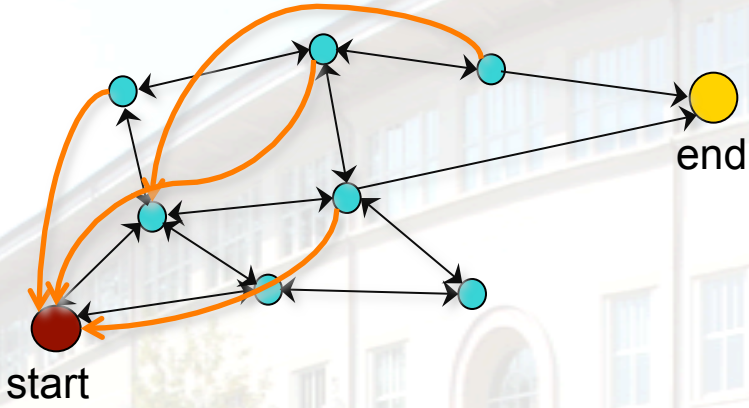
γ proofreading

$\psi = \gamma / k$



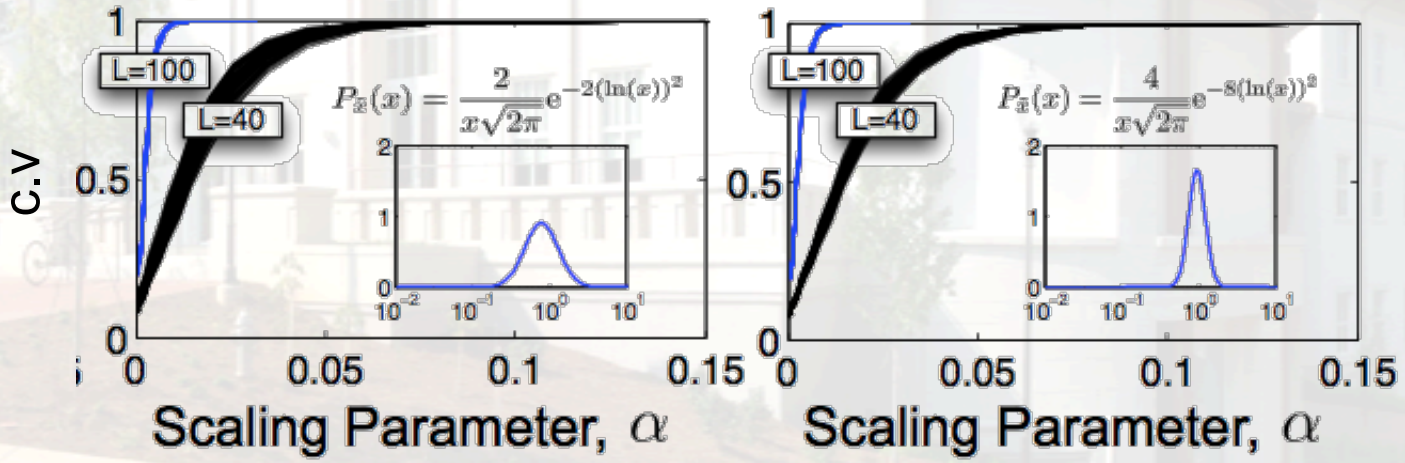
Bel, Munsky, 2009-10

Examples of time scales separation: Random netowkrs



α – relative strength of "shortcuts"

Similar to ligand dissociation



Bel, Munsky, *PB, JCP,* 2009-10

The simplest coarse-grainable system with time scale separation

- A Michaelis-Menten enzyme

$$\text{Substrate+Enzyme} \underset{k_{-1}(t)}{\overset{k_1(t)}{\rightleftharpoons}} \text{Substrate*Enzyme} \underset{k_{-2}(t)}{\overset{k_2(t)}{\rightleftharpoons}} \text{Product+Enzyme}$$

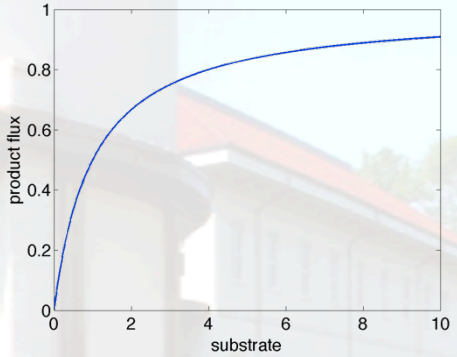
$$L \underset{k_{-1}(t)}{\overset{k_1(t)}{\rightleftharpoons}} \text{bin, pore} \underset{k_{-2}(t)}{\overset{k_2(t)}{\rightleftharpoons}} R$$

$$L \underset{}{\overset{Q}{\rightleftharpoons}} R$$

$$P\left[Q|T,\vec{k}(t)\right] = ?$$

Left-Right particle flux

$$k_\mu \gg \frac{d \log k_\mu}{dt}$$



- How do we coarse-grain these systems?
- Separation of relaxation time scales = different concentration scales

Some notes

- Need to get rid of internal degrees of freedom, not simulate them with a different protocol
- Approximation needs to work at the knee of MM curve, not only at the (easier) linear and the saturated limits
- Approximation needs to work when rates change with time
- Want a substantial speed-up in terms of Langevin or tau-leaping like approach, not simulate reaction by reaction. But nonetheless want to have arbitrary moments of reactions approximately correct.
- What are the long-time, large-current probabilities?

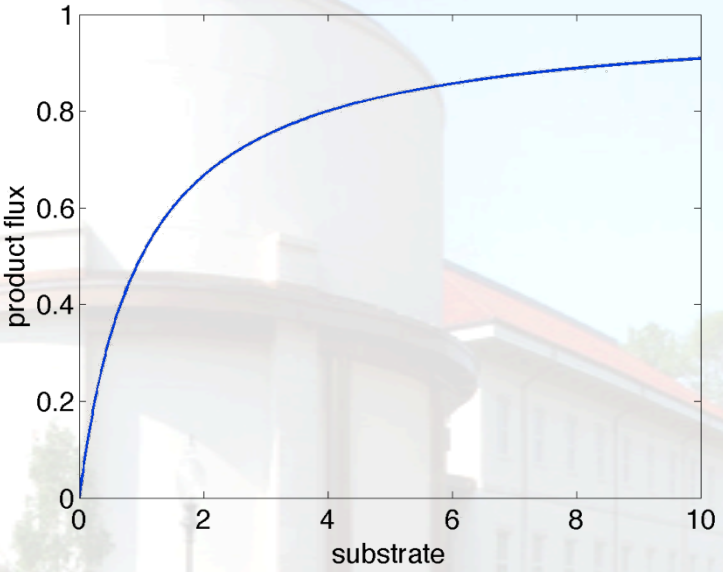
Adiabaticity in the deterministic case

Slow modulation

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

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

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



- Adiabatic approximation
  - Many enzyme turnovers for small (fractional) change in [*P*], [*S*]
- How to do the same with fluctuations?

Let's do the algebra: MGF approach

$$\frac{dP_n^E}{dt} = -(k_1 s + k_{-2} p)P_n^E + k_{-1}P_n^{ES} + k_2 P_{n-1}^{ES}$$

$$\frac{dP_n^{ES}}{dt} = -(k_{-1} + k_2)P_n^{ES} + k_1 s P_n^E + k_{-2} p P_{n+1}^{ES}$$

$$|G\rangle = \{G^E, G^{ES}\} = \sum_n e^{i\chi n}\{P_n^E, P_n^{ES}\}$$

$$G = \langle \mathbf{1} | G\rangle = G^E + G^{ES}$$

$$\left.\frac{\partial \log G}{\partial(i\chi)}\right|_{\chi=0} = \bar{J}$$

$$\left.\frac{\partial^2 \log G}{\partial(i\chi)^2}\right|_{\chi=0} = \sigma^2$$

...

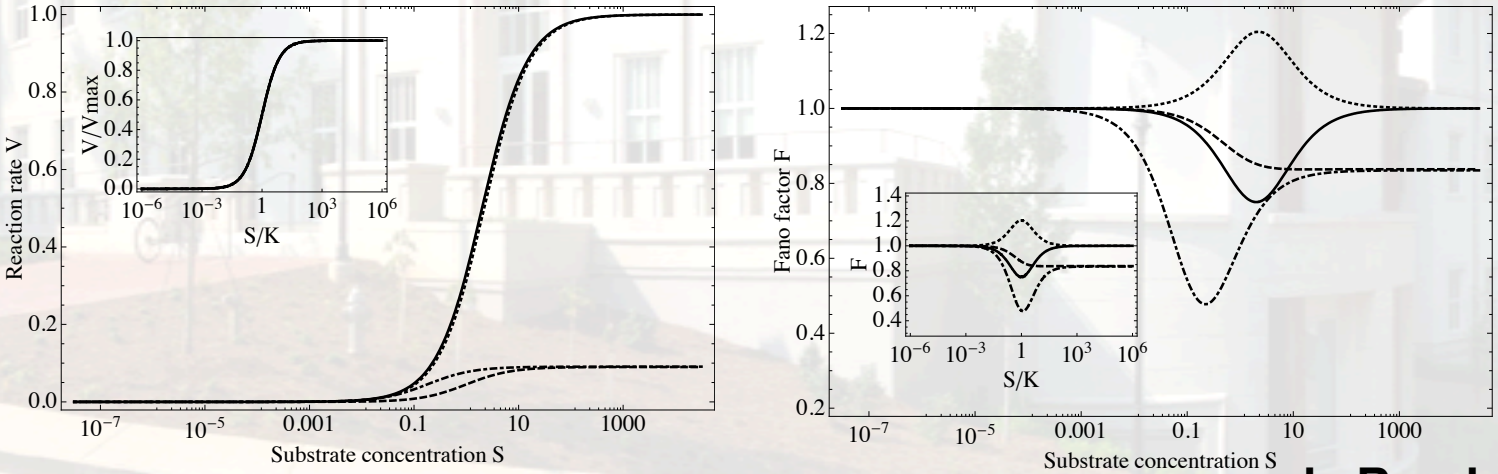
$$\frac{d|G\rangle}{dt} = H|G\rangle = \begin{bmatrix} k_1 + k_{-2} & -k_{-1} - k_2 e^{i\chi} \\ -k_1 - k_{-2} e^{-i\chi} & k_{-1} + k_2 \end{bmatrix} |G\rangle$$

A simple form: introduce counting fields

de Ronde et al, 2009

Results: can find the *coarse-grained* CGF

- For large t, solving this system of linear ODEs is equivalent to finding the largest real part eigenvalue.
- Treating χ as small, can use perturbation theory; e.g., two orders give first two cumulants. Sometimes can find $S=\log G$ completely.
- Easy to do for more complex reactions, e.g., allowing to disambiguate some reaction schemes from mesoscopic fluctuations data.



de Ronde et al, 2009

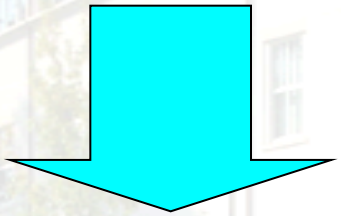
Fast simuluations of coarse-grained reactions

$$\frac{dp(L,R,N)}{dt} = \sum_i k_{i,\ \text{other}\to\text{p}}\, p(\text{other}) - \sum_i k_{i,\ \text{p}\to\text{other}}\, p$$

Chemical master equation

$R \to R \pm 1$ at random time t

Monte-Carlo 4-reaction Gillespie SSA algorithm



Large copy number

$$\frac{dR}{dt} = \frac{k_1 k_2 - k_{-1} k_{-2}}{\sum k} + \text{geom} + \eta$$

$$\langle \eta(0)\eta(t) \rangle = \text{correct} \times \delta(t) \quad \text{[or use more cumulants]}$$

1-reaction coarse "Langevin"

Simulating near-Gaussian η with known higher cumulants is doable

Sinitsyn et al., 2009

Generating non-Gaussian noises

- Assumptions needed when using higher order cumulants
- Gram-Charlier expansion

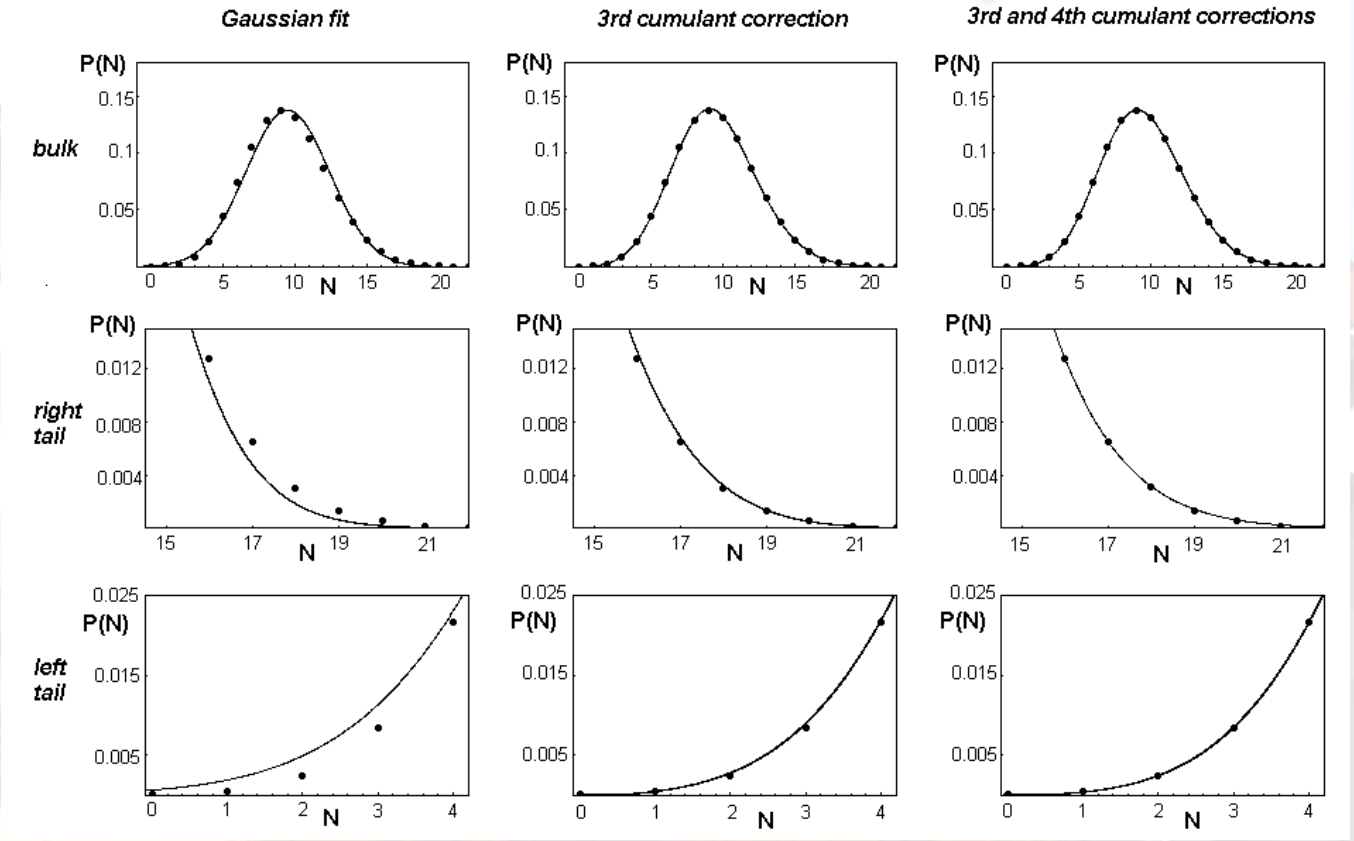
$$P(J \mid c_1, c_2, c_3, \ldots) = N(J \mid c_1, c_2)\left(1 + \frac{c_3\left(y^3 - y\right)}{6c_2^{3/2}} + \frac{c_4\left(y^4 - 6y^2 + 3\right)}{24c_2^2} + \frac{c_3^2\left(y^6 - 15y^4 + 45y^2 - 15\right)}{72c_2^3} + \ldots\right)$$

$$y = \frac{J - c_1}{\sqrt{c_2}}$$

- Higher order terms progressvely smaller since $c_n \sim dt \gg 1$
- Importance or rejection sampling from this near-gaussian distribution

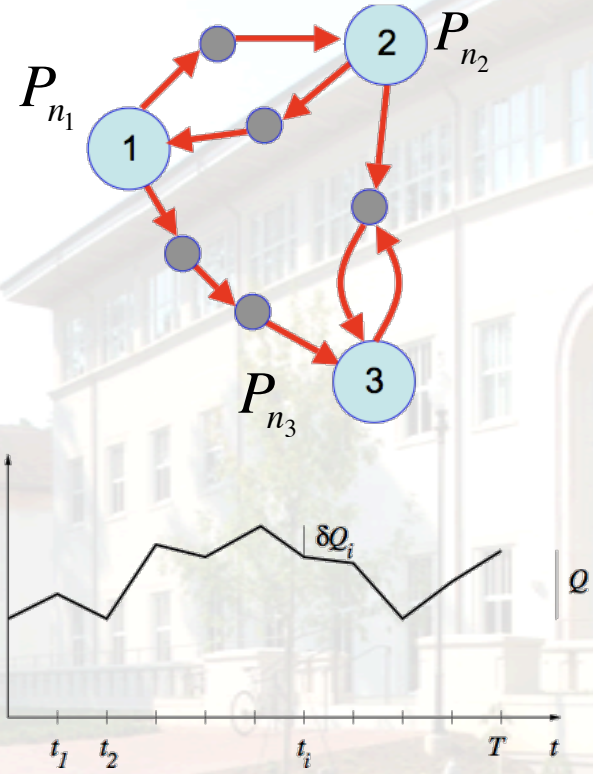
Examples of simulations

$$\varnothing \underset{q}{\overset{k_0(t)}{\rightleftharpoons}} ATP \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ATPase \xrightarrow{k_2} ADP \quad \Rightarrow \quad \varnothing \xrightarrow[\text{class+geom}]{\text{complex reaction}} ADP$$



10^4 speed-up compared to Gillespie

How to deal with larger networks?
Stochastic path integral



- Probability of δn_i particles created at node i over time δt is conditional all on n_j, which are also probabilistic, P_{n_j}
- If δt is small, each n_i is almost constant over δt.
- Overall future state probability is convolution of conditional probabilities with current state probabilities.
- Convolution of probabilities – addition of CGFs (integration in the limit of small n_i)
- Conservation laws couple n_i at different times, and overall CGF is

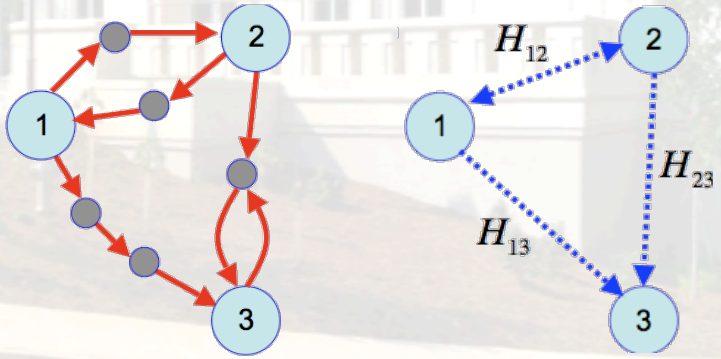
$$S(\chi,T)=\log G=\int_0^T dt\left[\sum_\mu i\chi_\mu \dot{N}_\mu+\sum_{\mu<\nu}H_{\mu\nu}\right]+O\left(\frac{\text{fast}}{\text{slow}}\right)$$

Cont'd

- Extra adiabaticity enforced by saddle point equations

$$\frac{\partial}{\partial \chi_\mu} S(\chi, T) = 0; \qquad \frac{\partial}{\partial \chi_\mu} \sum_\nu H_{\mu\nu} = -i\dot{N}_\mu \approx 0$$

- But what to do if $n_i \sim 1$ and small δt cannot be chosen?
  - Solve for the CGF of entire pathway involving this node as for MM before
  - Small n ensures that the perturbation theory is solvable



Concluding thoughts

- Approximate rigorous coarse-graining is possible in adidabatic limit using MGF techniques
- Simulations in coarse-grained models are many orders of magnitude faster
- Universal properties may emerge

Hopes for the future (Any takers?)

- Simulations
  - Build the ability to simulate non-Poisson nearly-Langevin variables into software packages
  - Build a library of coarse-grained models of various network motifs
  - Build tools to automatically coarse-grain models using the library
- Explore possible dynamics with RBMs
  - What is the set of all possible dynamics that appear with high probability in (semi-)random RBMs?
  - Which rules affect dynamics on long time scales and which don't?
  - For large rule sets, how likely are the dynamics to simplify or become universal? (see, e.g., proofreading or sloppy models)

The q-bio conference

- Bridging together quantitative experiments, theory, modeling
- Santa Fe, NM, Aug 8-12, 2012