Stochastic multi-scale models of biomolecular networks

Heinz Koeppl

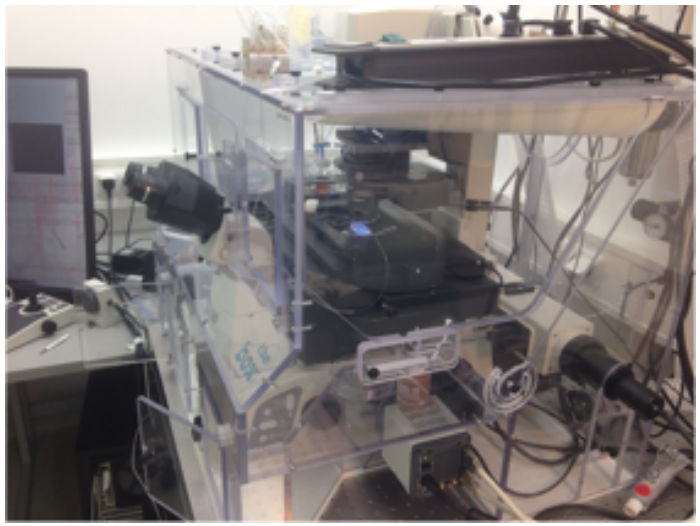
Department of Electrical Engineering and Information Technology
and Department of Biology
Technische Universität Darmstadt, Germany

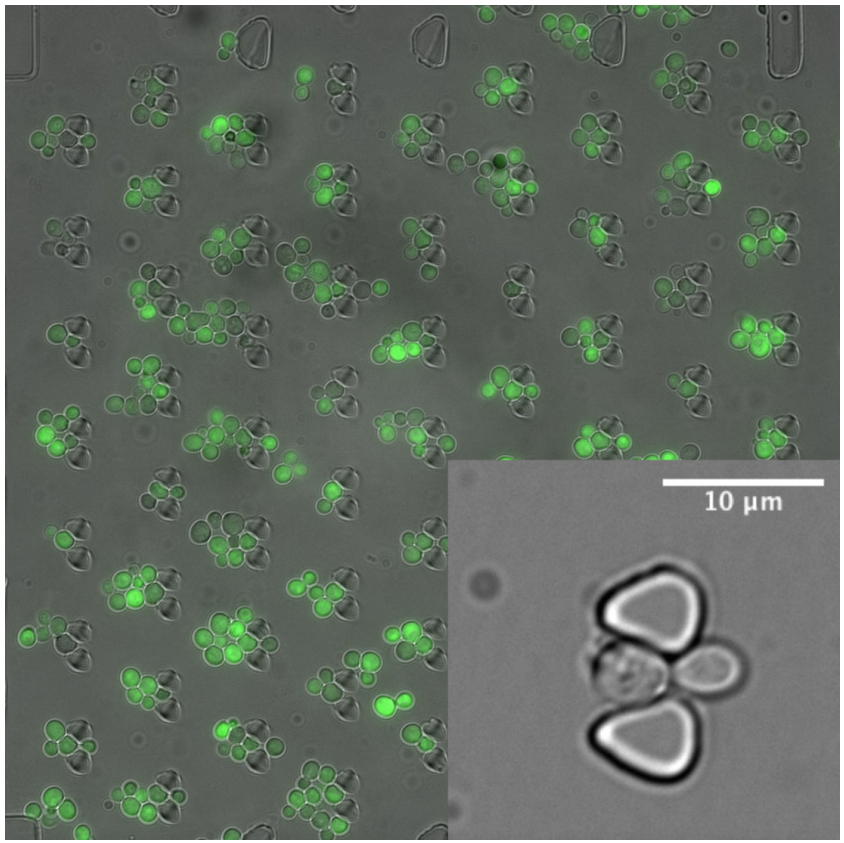
...joint work with A. Ganguly and D. Altintan
(SIAM Multiscale Modeling and Simulation, in press, 2015/arxiv)

Outline

Our research

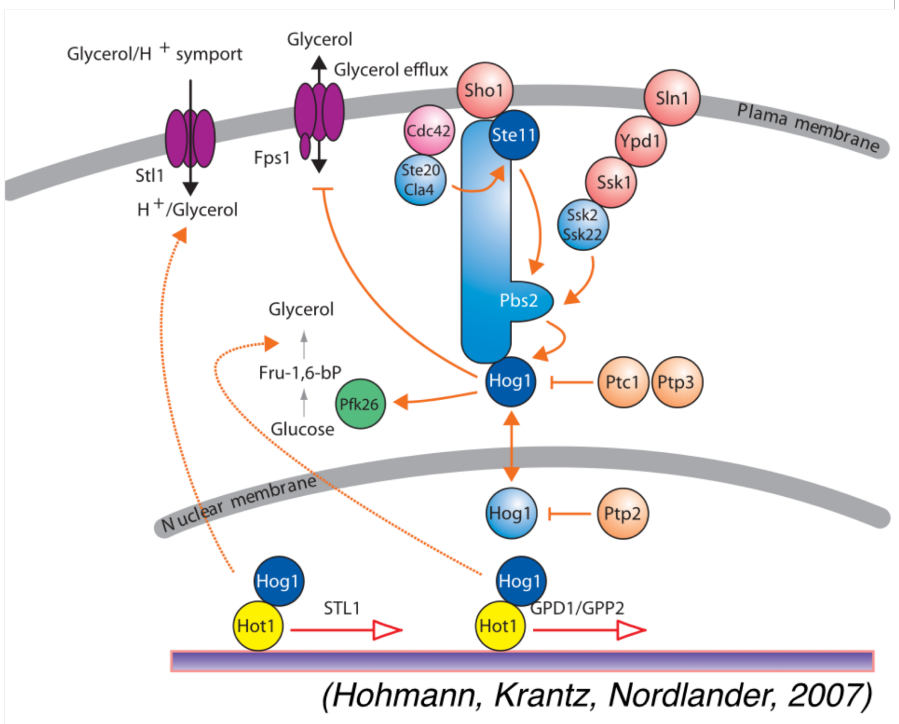
- Statistical inference (ML) methods
- Stochastic analysis of systems
- Single-cell experiments
- Microfluidic chips

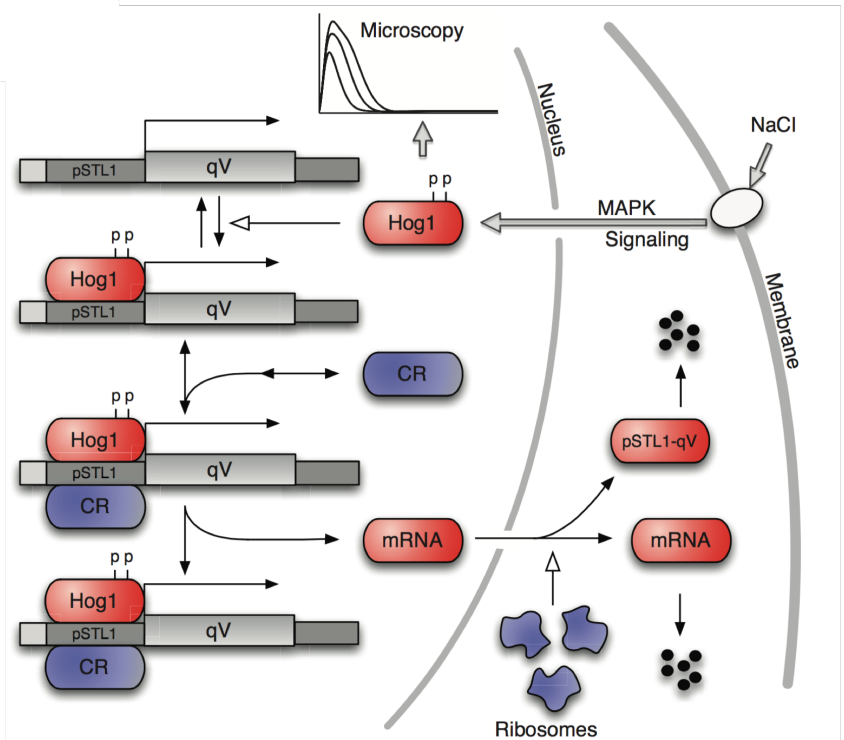




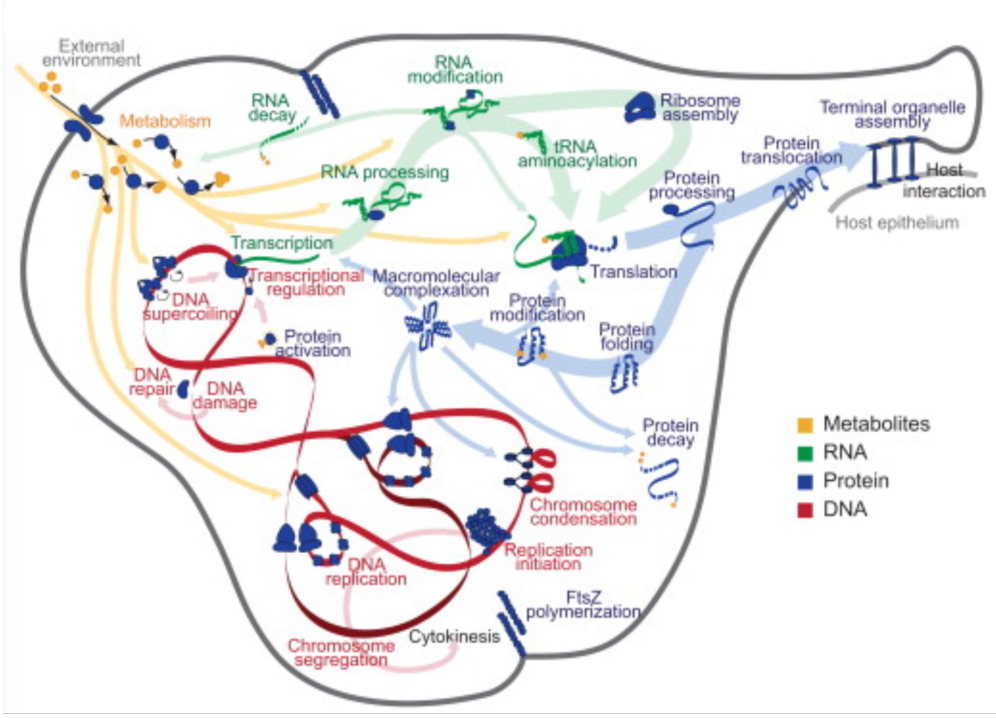
Example: osmotic stress response

- High-osmolarity induced gene expression in yeast *(PNAS'12)*





Motivation: Cell biology is multi-scale (...and multi-level)



Karr et al, Cell'12

- System-level understanding requires coupling of cellular processes.
- Cross-regulation between signaling, gene expression metabolism.
- Different time-scales results in computational bottleneck.
- On top of that: spatial effects (e.g. membrane trafficking) often significant - additional scale of resolution.

Population models (aka Petri nets)

$$R_k : \nu_{k1}^- S_1 + \nu_{k2}^- S_2 + \ldots + \nu_{km}^- S_m \xrightarrow{c_k} \nu_{k1}^+ S_1 + \nu_{k2}^+ S_2 + \ldots + \nu_{km}^+ S_m$$

$\nu_k = (\nu_{k1}^+ - \nu_{k1}^-, \nu_{k2}^+ - \nu_{k2}^-, \ldots, \nu_{km}^+ - \nu_{km}^-)^T$ Stoichiometric change vector

$X(t) = (X_1(t), X_2(t), \ldots, X_m(t))^T$ State vector , $c_k \in \mathbb{R}_+$: Rate constant

$X(t) = X(t-) + \nu_k$ occurrence of R_k updates the state vector

- Jump process formulation (CTMC)
  Continuous time, countable states - stochastic simulation or chemical master equation.
- Diffusion approximation
  Continuous time, continuous state - stochastic differential equation, Fokker-Planck equation.
- Mean-field approximation - thermodynamic limit
  Continuous time, continuous state - ordinary differential equation (RRE).

Multi-scale models

Combine different model description to exploit multi-scale nature.
Addressed questions:

1. *How do we partition a given CTMC model into discrete and approximate, continuous parts?*
2. *How can we bound the error for a certain partition?*
3. *How can we simulate such a multi-scale model?*

Research agenda

- Propose partitioning according to Poisson approximation.
- Perform path-wise, strong error analysis.
- Design dynamic partitioning algorithm to simulate multi-scale systems efficiently.

Random time-change representation

$$R_k : \nu_{k1}^- S_1 + \nu_{k2}^- S_2 + \ldots + \nu_{km}^- S_m \xrightarrow{c_k} \nu_{k1}^+ S_1 + \nu_{k2}^+ S_2 + \ldots + \nu_{km}^+ S_m$$

Propensity function example: $a_k(x) = c_k \prod_{i=1}^{m} \binom{x_i}{\nu_{ki}^-}, \quad k = 1, \ldots, r$

e.g. $R_k : S_1 + S_2 \xrightarrow{c_k} *$, $a_k(X) = c_k X_1 X_2$ or $S_1 \xrightarrow{c_k} *$, $a_k(X) = c_k\, X_1$
With $a_k(x)dt \equiv$ the probability of reaction R_k to occur in time interval $(t, t + dt]$.
A path-wise representation of the CTMC is as follows

$$X(t) = X(0) + \sum_{k=1}^{r} \xi_k \left(\int_0^t a_k(X(s))\mathrm{d}s \right) \nu_k,$$

where ξ_k's are independent Poisson processes.

Scaling

The abundance of species and time-scales of reactions can vary over different orders of magnitude. We introduce additional scaling variables to make that explicit (other variables are $O(1)$).

$$
\begin{array}{rrcl}
 & \bar{X}_i^N & = & X_i/N^{\alpha_i}, \quad \bar{X}_i^N = O(1), \\
 & d_k & = & c_k/N^{\beta_k}, \quad d_k = O(1), \\
\text{Example} & a_k(X) & = & c_k X_i = N^{\beta_k} d_k N^{\alpha_i} \bar{X}_i^N \equiv N^{\beta_k+\alpha_i} \lambda(\bar{X}_i^N) \\
 & a_k(X) & = & N^{\beta_k+\nu_k^- \cdot \alpha} \lambda_k(\bar{X}^N), \quad \Longrightarrow \ \lambda_k(\cdot) = O(1), \\
 & t & \to & t\, N^\gamma, \\
 & X^N(t) & = & \bar{X}^N(t\, N^\gamma).
\end{array}
$$

We obtain random time-change model for normalized variables

$$
X^N(t) = X^N(0) + \sum_{k=1}^{R} \xi_k \left(N^{\rho_k} \int_0^t \lambda_k(X^N(s))\mathrm{d}s \right) \nu_k^N,
$$

where we define $\rho_k = \gamma + \beta_k + \nu_k^- \cdot \alpha$ and $\nu_{ki}^N = \nu_{ki}/N^{\alpha_i}$.

Idea

- High intensity Poisson process can well be approximated by Brownian motion (plus dift), i.e. $\xi(t) \approx t + W(t)$.

Theorem (Kolmos, Major, Tusnady)

There exists a Brownian motion $W(t)$ on the same probability space as $\xi(t)$ such that

$$\Gamma = \sup_t \frac{\tilde{\xi}(t) - W(t)}{\log(2 \vee t)} < \infty \quad (a.s.),$$

where $\tilde{\xi}(t) = \xi(t) - t$ is the centered Poisson process.

- That is

$$\left[\frac{1}{\sqrt{n}} \tilde{\xi}(nt) - \frac{1}{\sqrt{n}} W(nt) \right] \leq \frac{\log(2 \vee nt)}{\sqrt{n}} \Gamma,$$

  where $\frac{1}{\sqrt{n}} W(nt)$ is a standard Brownian motion $W(t)$.
- Strategy: Replace reaction count process of high scaling factor N^{ρ_k} by an diffusion approximation.

Multi-scale approximation

- Jump-Diffusion Approximating process (first reaction replaced)

$$Y^N(t) = X^N(0) + N^{\rho_1} \int_0^t \lambda_1(Y^N(s))\mathrm{d}s\, \nu_1^N + W_1\left(N^{\rho_1} \int_0^t \lambda_1(Y^N(s))\mathrm{d}s\right) \nu_1^N + \sum_{k>1} \xi_k \left(N^{\rho_k} \int_0^t \lambda_k(Y^N(s))\mathrm{d}s\right)\ \nu_k^N$$

- Piece-wise deterministic Markov process (Jump-ODE)

$$Y^N(t) = X^N(0) + N^{\rho_1} \int_0^t \lambda_1(Y^N(s))\mathrm{d}s\, \nu_1^N + \sum_{k>1} \xi_k \left(N^{\rho_k} \int_0^t \lambda_k(Y^N(s))\mathrm{d}s\right)\ \nu_k^N$$

- Between two jumps overall system state $Y^N(t)$ evolves deterministically.

Bounding the error

- Compute bound on path-wise error for some $T \geq 0$

$$\sup_{t \leq T} \mathrm{E}|X^N(t) - Y^N(t)| = \sup_{t \leq T} \sum_{i=1}^{m} \mathrm{E}|X_i^N(t) - Y_i^N(t)|$$

for different reactions replaced by an approximation.

Theorem

Let $X^N(t)$ be the exact jump process and $Y^N(t)$ its jump-diffusion approximation with reaction $k = 1$ replaced then for $T \geq 0$

$$\sup_{t \leq T} \mathrm{E}|X^N(t) - Y^N(t)| \leq C_T(C' \log N^{\rho_1}/N^{m_1} + K''/N^{2\rho_1 + m_1})$$

where $|v_k^N| = O(N^{-m_k}) = O(\sum_{i \in R_k} N^{-\alpha_i})$ with $R_k = \{i \in \mathbb{N} \mid v_{ki} \neq 0\}$ and with the reaction non-specific constant $C_T = \exp\left(2 \sum_{k=1}^{r} N^{\rho_k} |\nu_k^N| L_k T\right)$

From a bound to an algorithm

- Convert reaction-specific error terms involving scaling variables back to states and propensities

$$\delta_k = \log N^{\rho_k}/N^{m_k} + 1/N^{2\rho_k + m_k}$$

- Fix time horizon $\Delta = O(N^{\gamma})$ and recall that $a_k(X)\Delta = O(N^{\rho_k})$ and $(N^{-m_k}) = O(\sum_{i \in R_k} N^{-\alpha_i}) = O(\sum_{i \in R_k} 1/X_i)$
- Then the jump-diffusion error criterion is

$$\delta_k(\Delta) = \sum_{i \in R_k} \frac{\log(a_k(X)\Delta)}{X_i} + \frac{1}{(a_k(X)\Delta)^2 X_i}$$

- Fix approximation accuracy ϵ and check at every Δ for each reaction k whether the incurred error of approximating it by diffusion is below ϵ.

Example: Bursty birth-death process

Consider the single species S

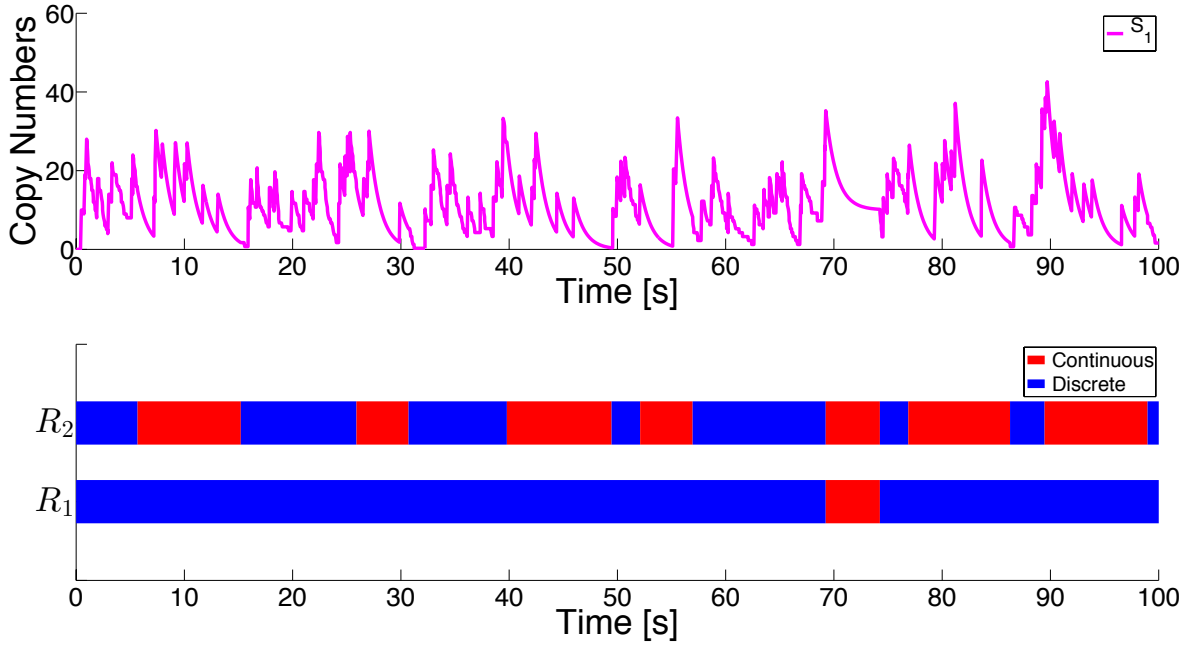
$$\emptyset \xrightarrow{c_1} 10S, \quad S \xrightarrow{c_2} \emptyset. \tag{1}$$

with the number of molecules of S at time t is denoted by $X(t)$. we use

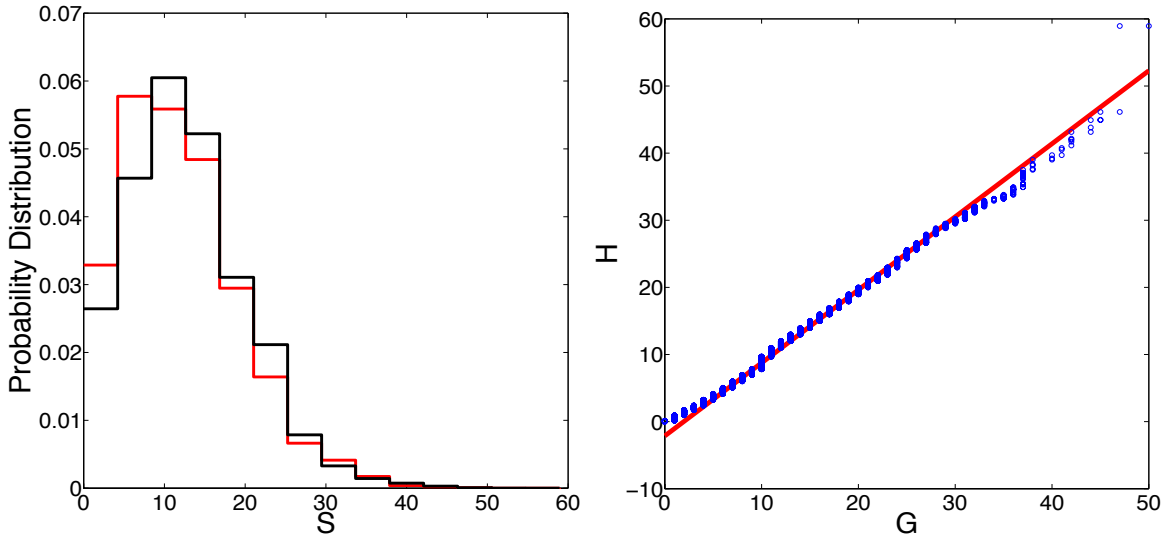
$$X(0) = 0, \quad \Delta = 0.1, \quad \varepsilon = 0.09, \quad P = 50.$$

Reaction constants of R_1, R_2 are given as $c_1 = 1\,\mathrm{molec\,s^{-1}}$, $c_2 = 1\mathrm{s}^{-1}$.
Death process is expected to be approximated by diffusion.

Dynamic partitioning



Weak error analysis by simulation



The Lotka-Volterra model

Let S_1 and S_2 denote the prey and the predator, respectively. The corresponding Lotka-Volterra prey-predator model can be depicted as

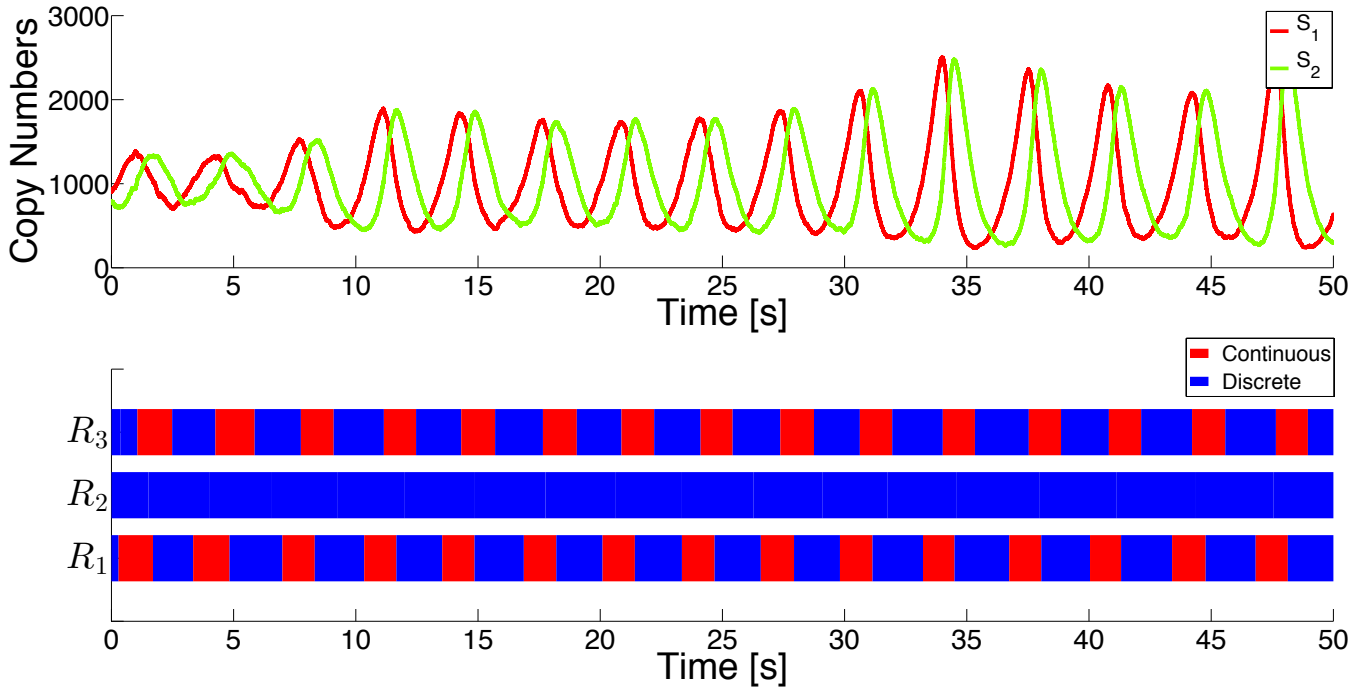
$$S_1 \xrightarrow{c_1} 2S_1, \quad S_1 + S_2 \xrightarrow{c_2} 2S_2 \quad S_2 \xrightarrow{c_3} \emptyset. \tag{2}$$

Let $X_1(t)$ and $X_2(t)$ denote the number of the prey and the predator at time $t > 0$, respectively, then, the state of the system is defined by $X(t) = (X_1(t), X_2(t))^T \in \mathbb{N}^2_{\geq 0}$. We use

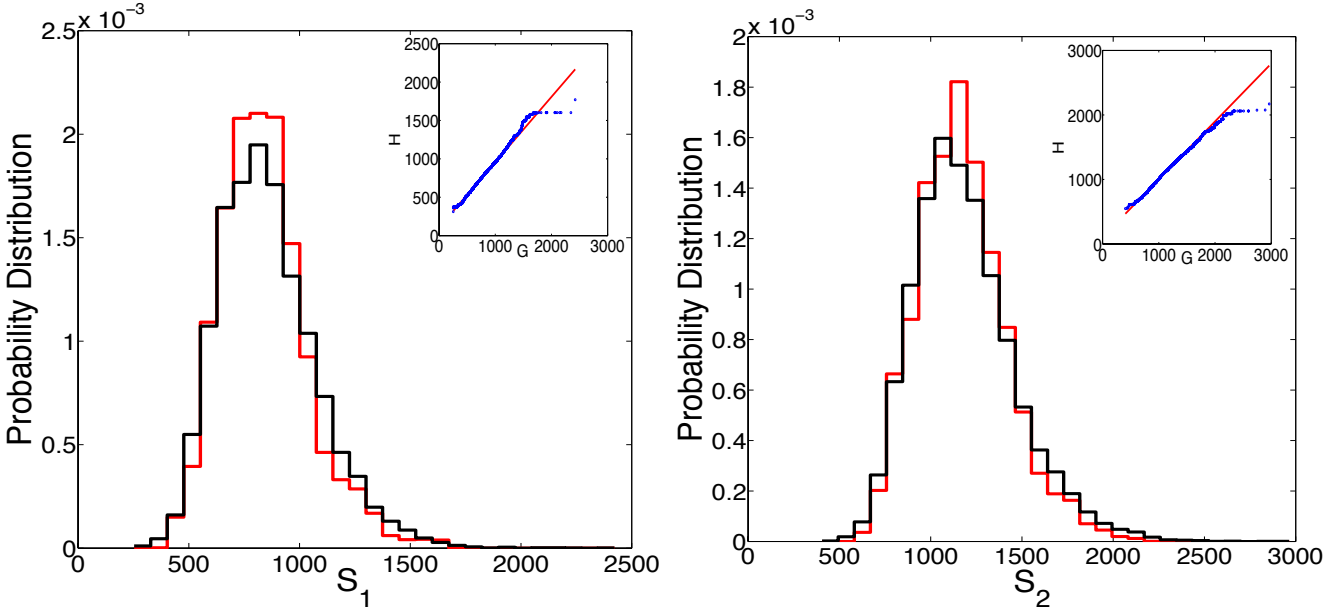
$$X(0) = (900, 800)^T,\ c_1 = 2\mathrm{s}^{-1},\ c_2 = 0.002\mathrm{molec}^{-1}\mathrm{s}^{-1},\ c_3 = 2\mathrm{s}^{-1}.$$

$$\Delta = 0.5,\ \varepsilon = 0.03,\ P = 50,\ t \in (0, 50)$$

Dynamic partitioning

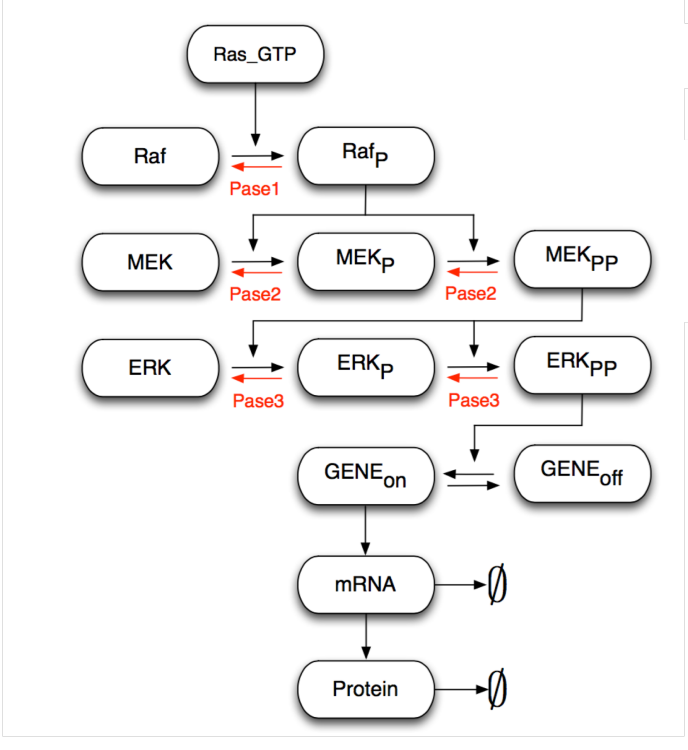


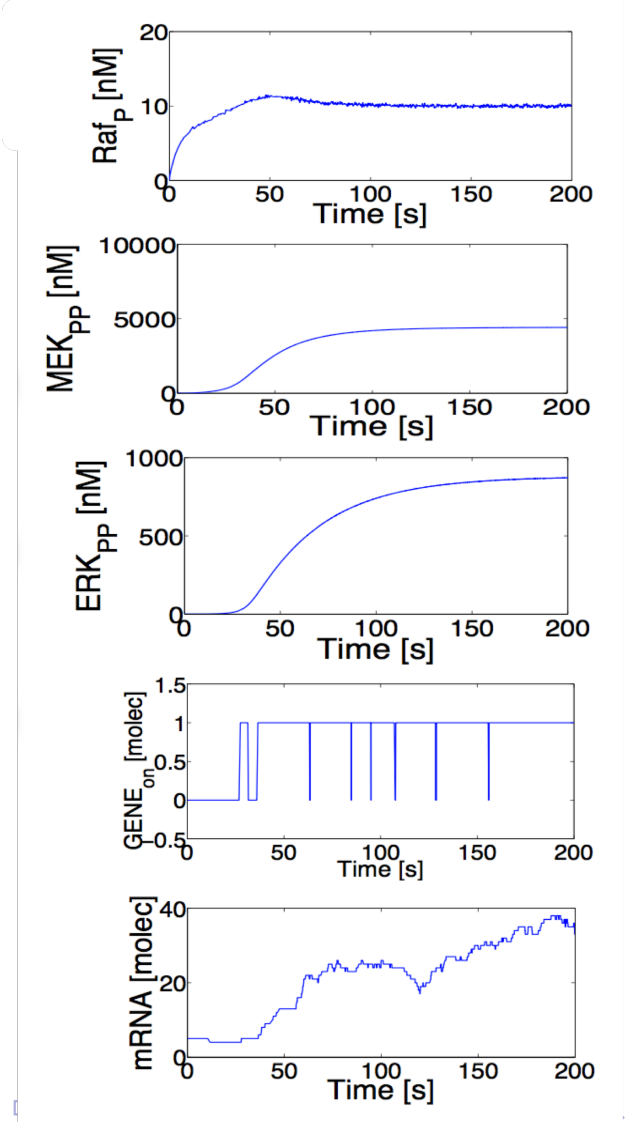
Weak error analysis by simulation



EGFR signaling and gene expression

- Rate constants and abundances based on *Schoeberl et al., Science Sci'09*
- 30 signaling reactions, 6 gene expression reactions
- Runge-Kutta strong order 2 SDE integrator





Conclusions

- Multi-scale models essential for systems biology.
- Traditional hybrid models often involve ad-hoc partitioning of species.
- Reaction partitioning leveraging existing approximation results for point processes.
- Generally, every state becomes a jump-diffusion process (i.e. no species partitioning).
- Explicit bound for finite-time error for approximating specific reaction channels.
- Real gain requires higher-order integration schemes for SDEs.
- Bounds for ODE-Jump and ODE-SDE-Jump.