Stochastic Simulation of Multiscale Intracellular Reacting Networks

Di Liu

http://www.math.msu.edu/~richardl

Department of Mathematics, Michigan State University

Support: National Science Foundation WHERE DISCOVERIES BEGIN

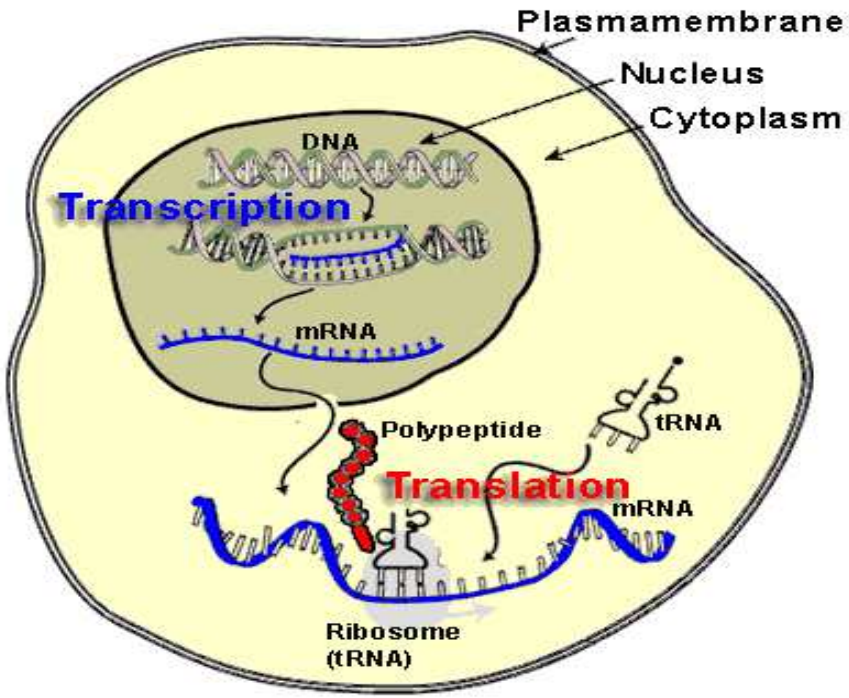
DMS CAREER 0845061, DMS 1418959

Multi-scale stochastic chemical kinetic systems

Genetic Regulatory Network (GRN)

1. Transcription $\{\text{RNA polymerase} \longrightarrow \text{DNA}\} + \text{Nucleotides} \implies \text{RNA}$

2. Translation $\{\text{Ribosome} \longrightarrow \text{MRNA}\} + \text{Amino Acids} \implies \text{Proteins}$



3. Regulation Transcription factors $\longrightarrow$ Regulatory DNA sequences

$\implies$ Activate or Repress Transcription in response to different stimuli

Multi-scale stochastic chemical kinetic systems

Multiscale modeling of Genetic Regulatory Network (GRN)

1. GRN: Genes, proteins, small molecules within cells

   LOW concentrations in SMALL volumes

2. Macro-scale model: ODE/PDE/SDE dynamics in terms of concentrations

   Valid with large concentrations

3. Micro-scale model: Molecular dynamics or First principle

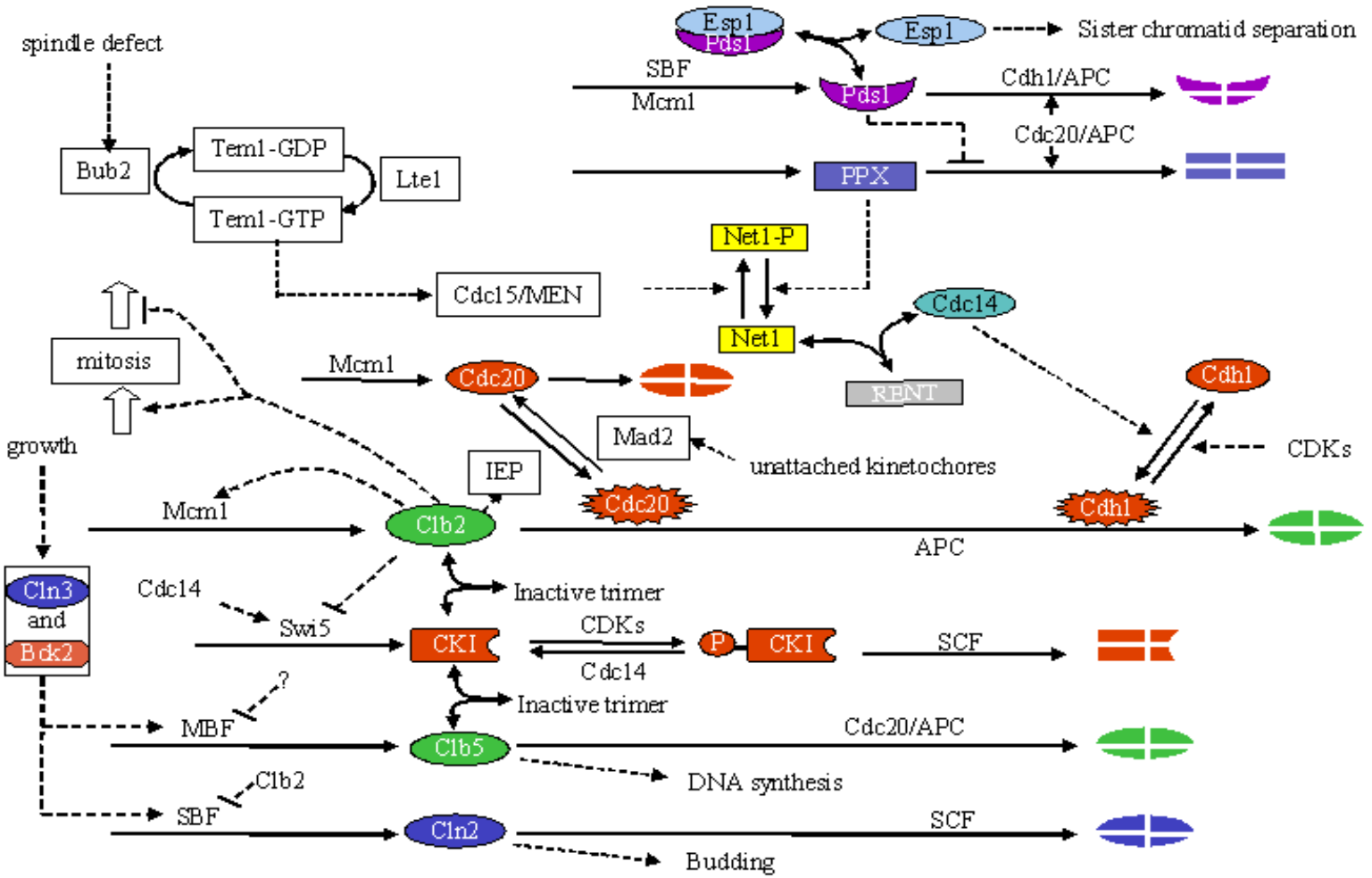
   Numerically / Analytically intractable

4. Meso-scale model: Molecular events omitting details (position, momentum)

   Stochastic dynamics / Discrete molecular numbers

Multi-scale stochastic chemical kinetic systems

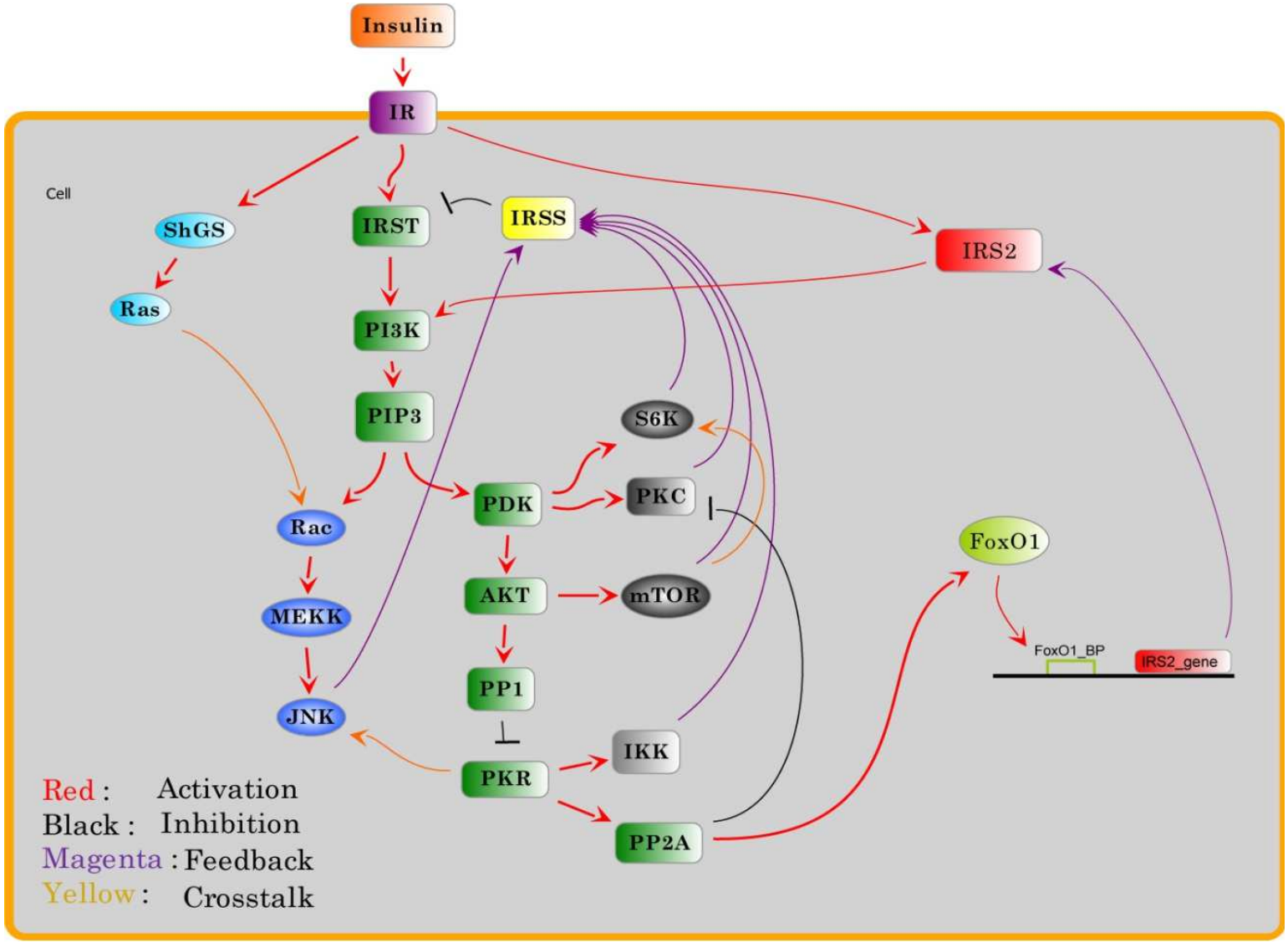
Cell cycle model of Budding Yeast (Tyson et. al. 04)



>80 reactions and >50 reacting species

Multi-scale stochastic chemical kinetic systems

Insulin Response Model (Huang, Wu, Du, Chan and Liu, *J. Theo. Bio.*, 14)



$\sim$40 reactions and species

Multi-scale stochastic chemical kinetic systems

More on Modeling and Simulation

1. Modularized structure — built collectively on previous models

2. Large amount of parameters and initial conditions
   —expensive to do experiments

3. Logic models v.s. Kinetic models
   (a) Multidimensional pathways — forward and backward feedback loops, etc
   (b) Time scales and delays
   (c) Crosstalks between different pathways

4. Numerical output
   (a) Hypothesis test on important pathways
   —IRS1 and IRS2 plays a central role in functioning of the Insulin response network
   (b) Sensitivity analysis$\rightarrow$potential therapeutic targets
   —IRS and JNK

Multi-scale stochastic chemical kinetic systems

Molecular numbers of reacting species :

$$x = (x_1, \cdots, x_d) \in \mathbb{N}^d$$

Reaction : $R_j = \big(a_j(x), \nu_j\big)$

reaction rate : $a_j(x)$ — probability of reaction j in unit time interval

change of the state : $x \longrightarrow x + \nu_j$

Example: $S_1 \underset{a_2}{\overset{a_1}{\rightleftarrows}} S_2 \qquad 2S_2 \xrightarrow{a_3} S_3$

$$\nu_1 = (-1, 1, 0) \qquad \nu_2 = (1, -1, 0) \qquad \nu_3 = (0, -2, 1)$$

$$(a_1,\ a_2,\ a_3) = (c_1 x_1,\ c_2 x_2,\ c_3 x_2(x_2 - 1))$$

Forward master equation :

$$\frac{\partial P(x,t)}{\partial t} = \sum_j \Big(a(x - \nu_j) P(x - \nu_j, t) - a_j(x) P(x,t) \Big)$$

Multi-scale stochastic chemical kinetic systems

SSA (Gillespie's algorithm, BKL, Kinetic Monte Carlo, ...)

$$R = (a, \nu) \qquad\qquad a_0(x) = \sum_j a_j(x)$$

1. At (t_n, x_n), generate r_1 and r_2 with uniform dist. on unit internal

$$\delta t_{n+1} = \frac{1}{a_0(x_n)} \ln(\frac{1}{r_1})$$

Skip the time when no reaction happens

$$k_{n+1}: \qquad \sum_{i=1}^{k_{n+1}-1} a_i(x_n) < r_2 a_0(x_n) \leq \sum_{i=1}^{k_{n+1}} a_i(x_n)$$

Pick up a reaction

2. $t_{n+1} = t_n + \delta t_{n+1}, \qquad x_{n+1} = x_n + \nu_{k_{n+1}}$

Update the system

Multi-scale stochastic chemical kinetic systems

Multiscale nature of GRNs:

1. Multiple time scales

   eg. Binding of RNAP v.s. Transcription/Translation

   Nested SSA (Weinan E, Di Liu and Eric Vanden-Eijnden)

2. Multiple population scales

   eg. Transcription/Translation v.s. Protein-Protein reactions

   Sampling invariant measures with τ-leaping method (Can Huang and Di Liu)

3. Metastability

   Different sable gene expression profiles $\Longrightarrow$ Different phenotypic states

   eg. Cell fates : proliferation, apoptosis, . . .

   Minimum Action Method for Chemical Kinetic Systems (Di Liu)

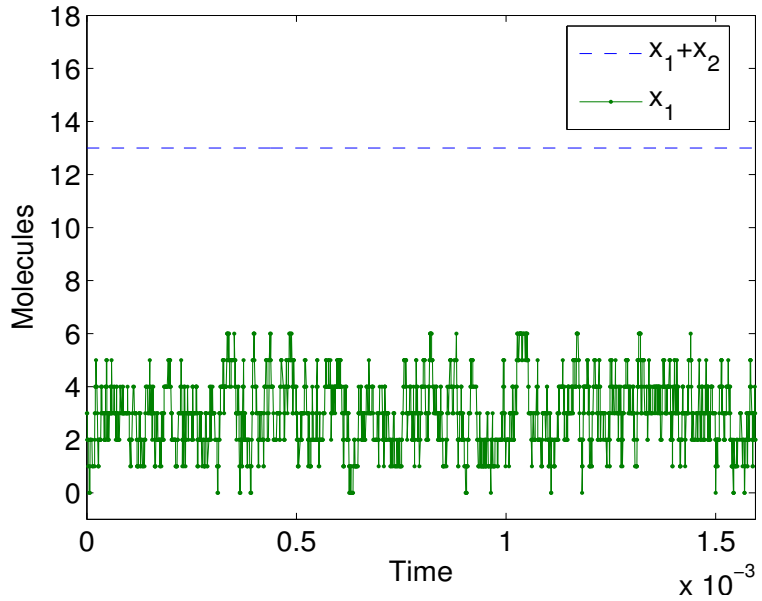
Nested stochastic simulation algorithm

Example: $S_1 \underset{a_2}{\overset{a_1}{\rightleftarrows}} S_2 \qquad S_2 \xrightarrow{a_3} S_3$

$$\nu_1 = (-1, 1, 0) \qquad \nu_2 = (1, -1, 0) \qquad \nu_3 = (0, -1, 1)$$

Time-scale separation due to the separation of the reaction rates

$$(a_1,\ a_2,\ a_3) = (\frac{1}{\epsilon}x_1,\ \frac{1}{\epsilon}x_2,\ x_2)$$



Nested stochastic simulation algorithm

Effective dynamics on slow time-scale

$$R = (R^s, R^f)$$

$$R^s = \left(a_j^s(x), \nu_j^s\right) \qquad R^f = \left(\frac{1}{\epsilon} a_j^f(x), \nu_j^f\right)$$

Assuming an equilibrium distribution P for the fast reactions

Effective dynamics for the slow variables: (Principle of Averaging)

$$\bar{R}_j = \left(\bar{a}_j^s(x), \nu^s\right) \qquad \bar{a}_j^s(x) = P a_j^s(x)$$

— Singular perturbation for the backward equation

Nested stochastic simulation algorithm

Example revisited:

$$S_1 \underset{a_2}{\overset{a_1}{\rightleftarrows}} S_2 \qquad\qquad S_2 \xrightarrow{a_3} S_3$$

Fast reaction (Birth-Death process)

$$S_1 \underset{a_2}{\overset{a_1}{\rightleftarrows}} S_2$$

Equilibrium distribution: $p(x) = p(x_2 = x) = \frac{(N-x_3)!}{x!(N-x_3-x)!} q^x (1-q)^{N-x_3-x}$

Effective dynamics in terms of the slow variable x_3:

$$\bar{R} = \big(Px_2,\ (0,-1,1)\big) = \big(\frac{1}{2}(N-x_3),\ (0,-1,1)\big)$$

— One direction Birth-Death process.

Nested stochastic simulation algorithm

Identification of the slow variables:

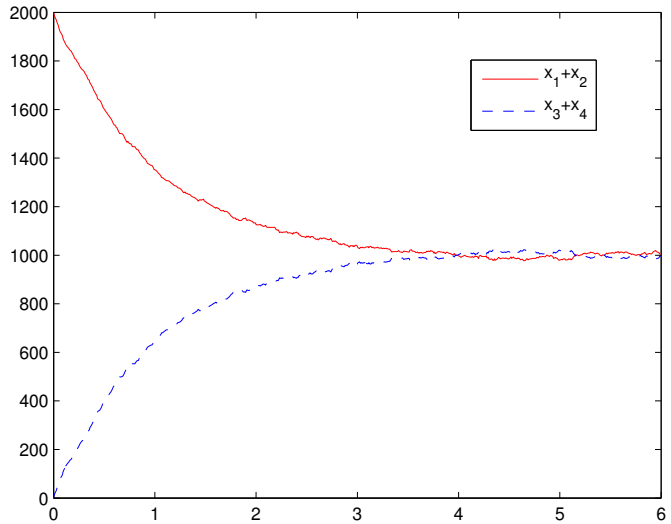
$$S_1 \underset{a_2}{\overset{a_1}{\rightleftarrows}} S_2 \qquad S_2 \underset{a_4}{\overset{a_3}{\rightleftarrows}} S_3 \qquad S_3 \underset{a_6}{\overset{a_5}{\rightleftarrows}} S_4$$

$$(a_1, a_2, a_3, a_4, a_5, a_6) = (\frac{x_1}{\epsilon}, \frac{x_2}{\epsilon}, x_2, x_3, \frac{x_3}{\epsilon}, \frac{x_4}{\epsilon})$$

Every variable is involved in one fast reaction

$$\bar{R}_1 = (\ \frac{1}{2}(x_1 + x_2),\ (0, -1, 1, 0)\ ) \qquad \bar{R}_2 = (\ \frac{1}{2}(x_3 + x_4),\ (0, 1, -1, 0)\ )$$

Slow variables : Conversed linear functions in fast reactions



Nested stochastic simulation algorithm

(W. E, Liu and Vanden-Eijnden, *J. Chem. Phys.*, 05; *J. Comp. Phys.*, 07):

$$\bar{R}_j = \left(\bar{a}_j^s(x), \nu^s\right)$$

1. Inner SSA: Direct SSA for fast reactions $R^f = (a^f, \nu^f)$

   $x_k(t), \quad k = 1, \ldots, N$

   N — ensemble number (parallel implementation)

2. Estimate the macro data (time-ensemble average)

$$\bar{a}_i^s(x) \approx \tilde{a}_i^s(x) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{T_f} \int_0^{T_f} a^s(x_k(t))dt$$

   T_f — time for averaging

3. Outer SSA Direct SSA for slow reactions with modified rates:

$$\tilde{R}^s = (\tilde{a}^s, \ \nu^s)$$

Nested stochastic simulation algorithm

Exponential mixing for fast reactions:

$$| \mathbb{E}f(y_t) - Pf | \leq Re^{-\alpha t/\epsilon}$$

Error Estimate: **Weak** (E, Liu and Vanden-Eijnden, 05, 07)
Strong (Huang and Liu, *Comm. Comp. Phys.*, 14)

$$\mathbb{E}\left|x_t - \tilde{x}_t\right| \leq C\left(\epsilon + \frac{1}{1+T_f/\epsilon} + \frac{1}{\sqrt{N(1+T_f/\epsilon)}}\right)$$

Principle of averaging + Relaxation + Sampling

Designing: Given error tolerance λ, choose the parameters such that

$$Error \leq \lambda$$

Computational cost: $O\left(\frac{1}{\lambda^2}\right)$ when $\epsilon \ll \lambda$

Nested stochastic simulation algorithm

Reiterated Averaging

$$R = (R^s,\ \frac{1}{\epsilon}R^f,\ \frac{1}{\epsilon^2}R^{uf})$$

$$R^s = (a^s, \nu^s) \qquad R^f = (\frac{1}{\epsilon}a^f, \nu^f) \qquad a^{uf} = (a^{uf}, \frac{1}{\epsilon^2}\nu^{uf})$$

Suppose the following is ergodic:

$$\tilde{R}^f = (\mathbb{P}a^f, \nu^f)$$

$\mathbb{P}$ equilibrium of R^{uf}

Effective dynamics:

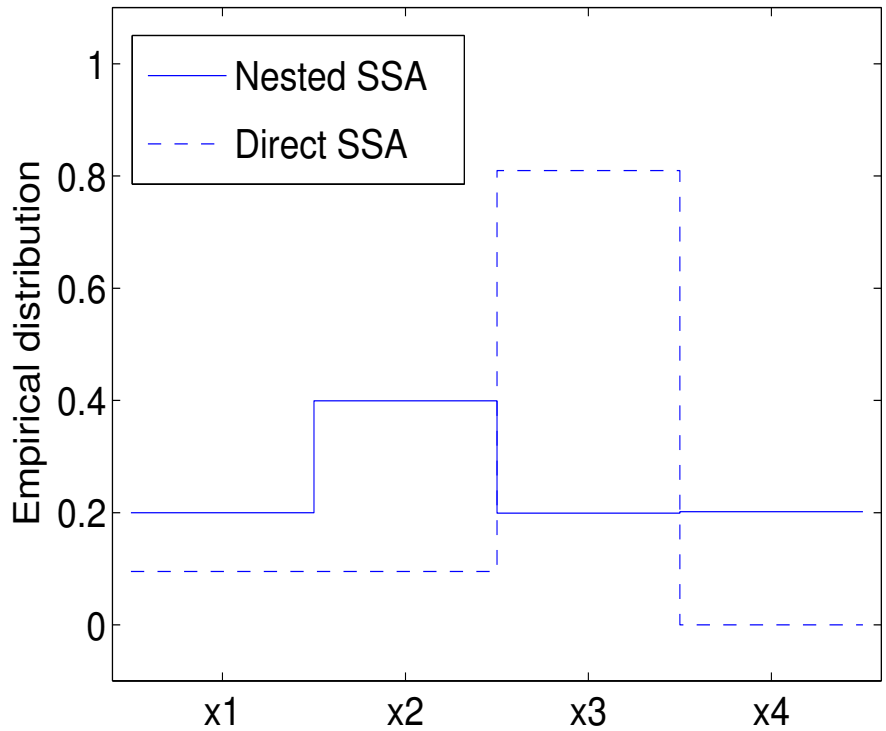
$$\bar{R}^s = (\mathcal{Q}\mathbb{P}a^s, \nu^s)$$

$\mathcal{Q}$ equilibrium of $\tilde{R}^f$

Nested stochastic simulation algorithm

$$S_1 \underset{a_2}{\overset{a_1}{\rightleftarrows}} S_2, \qquad S_2 \underset{a_4}{\overset{a_3}{\rightleftarrows}} S_3, \qquad S_3 \underset{a_6}{\overset{a_5}{\rightleftarrows}} S_4.$$

$$(a_1, a_2, a_3, a_4, a_5, a_6) = \left(\frac{2x_1}{\epsilon^2}, \frac{x_2}{\epsilon^2}, \frac{x_2}{\epsilon}, \frac{2x_3}{\epsilon}, x_3, x_4\right)$$

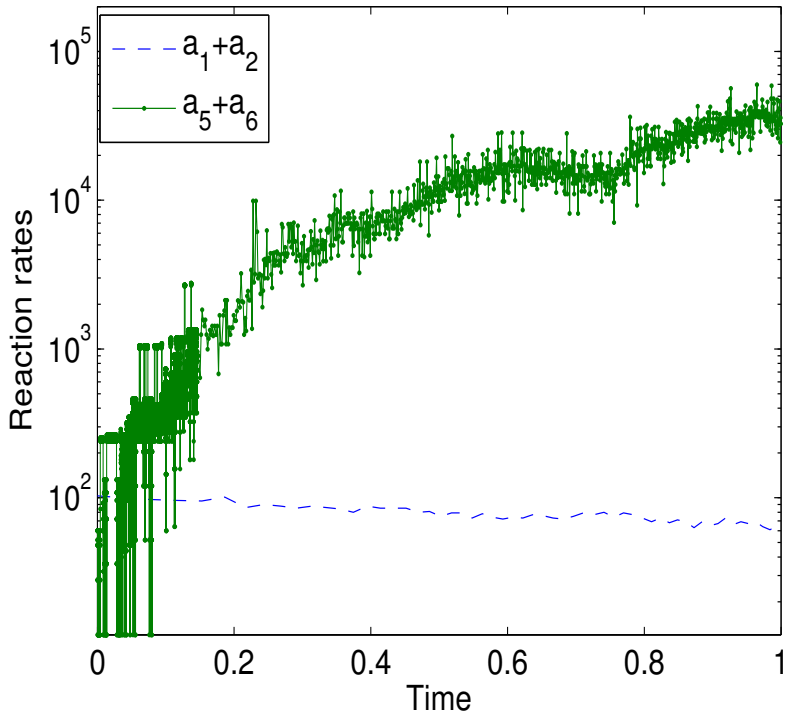


Nested stochastic simulation algorithm

Adaptivity

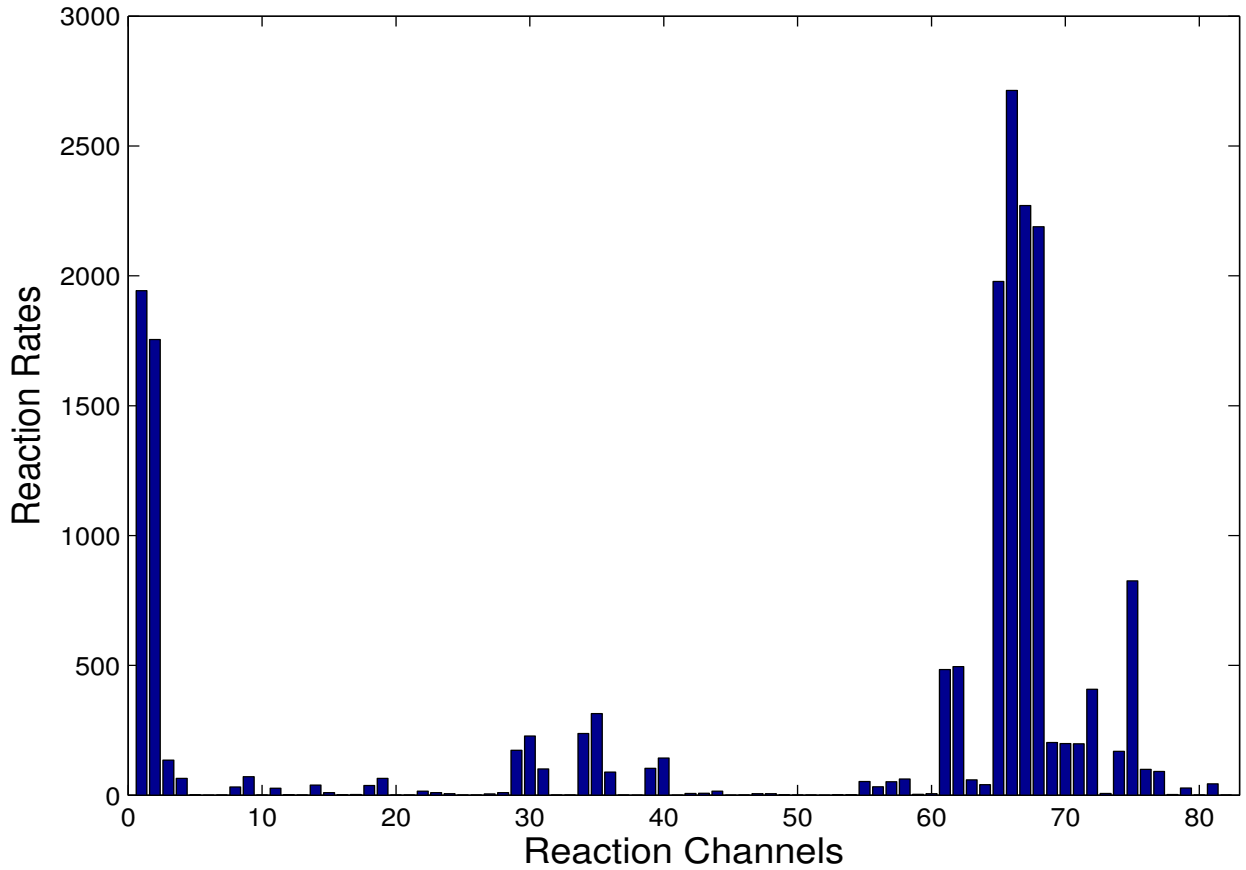
$$S_1 \underset{a_2}{\overset{a_1}{\rightleftarrows}} S_2, \qquad S_2 \underset{a_4}{\overset{a_3}{\rightleftarrows}} S_3, \qquad 2S_2 + S_3 \underset{a_6}{\overset{a_5}{\rightleftarrows}} 3S_4.$$

$$\begin{aligned}(a_1,\ a_2) &= (x_1,\ x_2)\\ (a_3,\ a_4) &= (10^4 x_2,\ 10^4 x_3)\\ (a_5,\ a_6) &= \big(2x_2(x_2-1)x_3,\ 2x_4(x_4-1)(x_4-2)\big)\end{aligned}$$



Cell cycle model of budding yeast

Time scale separation is $O(10^3)$.

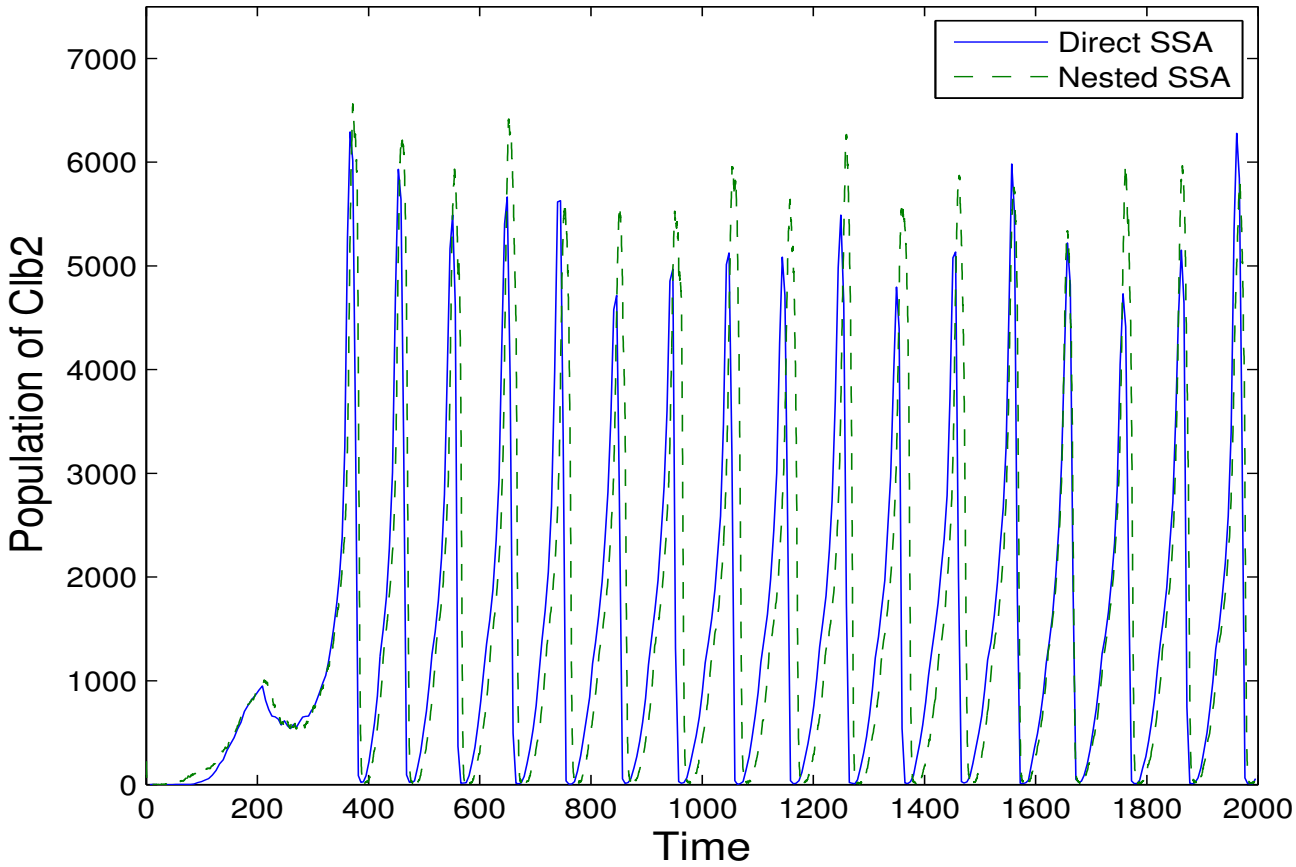


Efficiency gain:

$$\frac{\sum a^s + \frac{1}{\epsilon}\sum a^f}{\sum a^s + \sum a^f + \text{sampling}}$$

Cell cycle model of budding yeast

(Liu, *Comm. Comp. Phys.*, 11)
NSSA is 4 times faster with only $.02\%$ relative error in the averaged period.

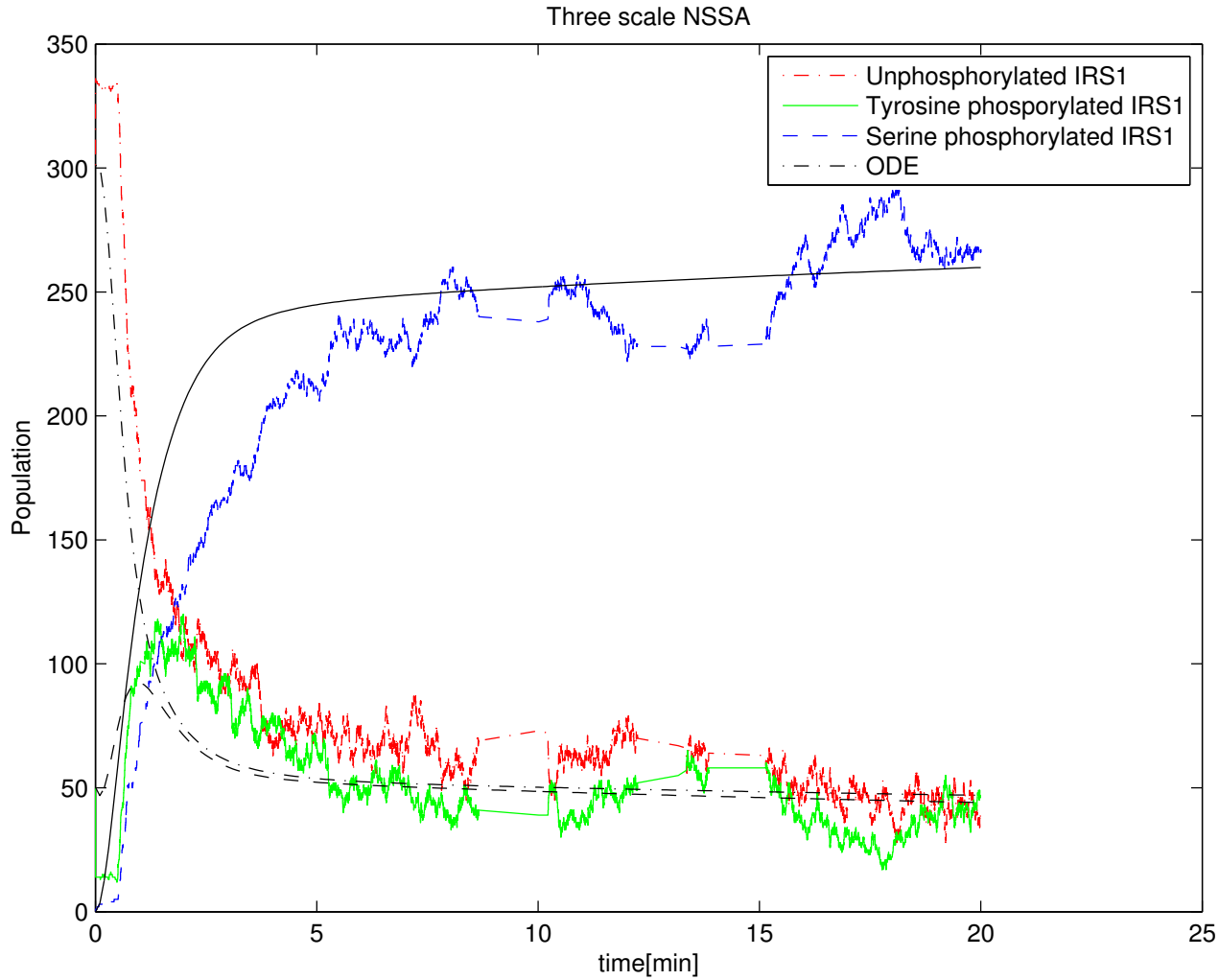


Adaptivity is indispensable

Insulin response model

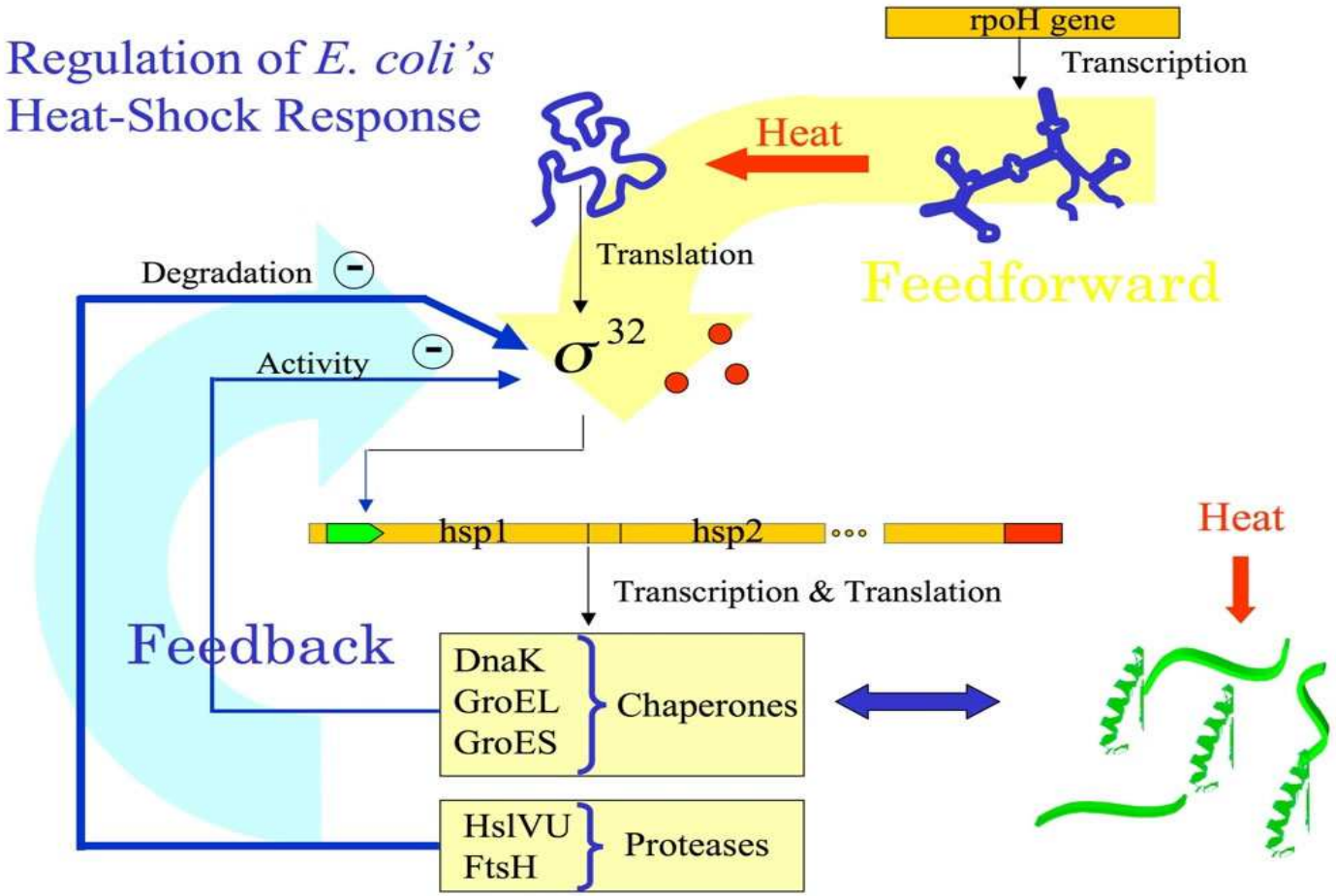
(Huang and Liu, *Comm. Comp. Phys.*, 14)

- There are 3 time scales in the model.
- The slow-fast-ultrafast partition of reactions dynamically changes.
- NSSA is 60 faster than Direct SSA.



Heat shock response of E. Coli

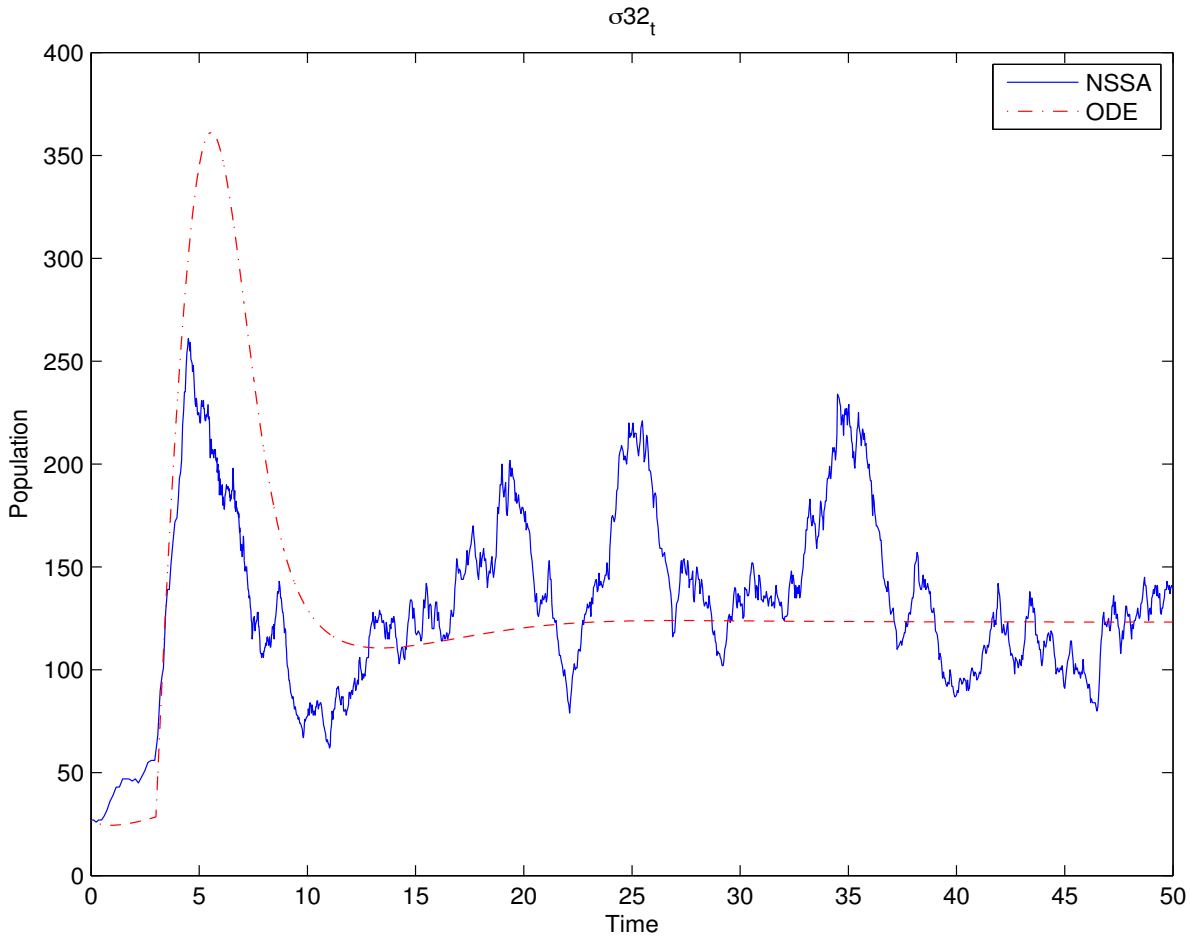
(El-Samad, Kurata, Doyle, Gross and Khammash 01)



Heat shock response of E. Coli

(Huang and Liu, *Comm. Comp. Phys.*, 14)

- The original model is Differential-Algebraic Equations.
- In the full stochastic model, there are 3 time scales.
- Direct SSA is impossible.



Sampling Invariant Measures with τ-leaping

SDE for SSA

$$dX_t = \sum_j \int_0^\infty \nu_j \mathbf{A}_j(q, X_t)\mathcal{P}(dt, dq),$$

where

$$\mathbf{A}_j(q, X_t) = \begin{cases} 1, & q \in (\sum_{i=1}^{j-1} a_i(X_t), \sum_{i=1}^{j} a_i(X_t)) \\ 0, & \text{otherwise.} \end{cases}$$

and $\mathcal{P}(dt, dq)$ is the Poisson random measure with Lebesgue intensity.

τ-leaping (Euler) method (Gillespie 01)

$$X_{n+1} = x + \sum_{j=1}^{M_R} \nu_j P_j(a_j(x), \tau),$$

where $P_j(a_j(x), \tau), j = 1, \cdots M_R$ are independent Poisson r.v.

Sampling Invariant Measures with τ-leaping

Allow large time steps when populations are high: (T. Li 07, D. Anderson et al. 11)

$$\frac{1}{a_0} < \tau \ll 1.$$

Effectiveness on infinite time horizon: (Huang and Liu, *Comm. Comp. Phys.*, 14)

$$\left| \int \phi(y) d\mu_{Z_n}(y) - \mathbb{E}\frac{1}{N}\sum_{i=1}^{N} \phi(Y_{n,i}) \right| \leq C(\tau + \frac{1}{T_f}).$$

Modified Inner SSA

$$X_{n,i+1} = X_{n,i} + \sum_{j=1}^{M_f} \nu_j^f P_j(a_j^f(X_{n,i}), \tau), \quad X_{n,0} = X_n, \quad i = 0, \cdots, N,$$

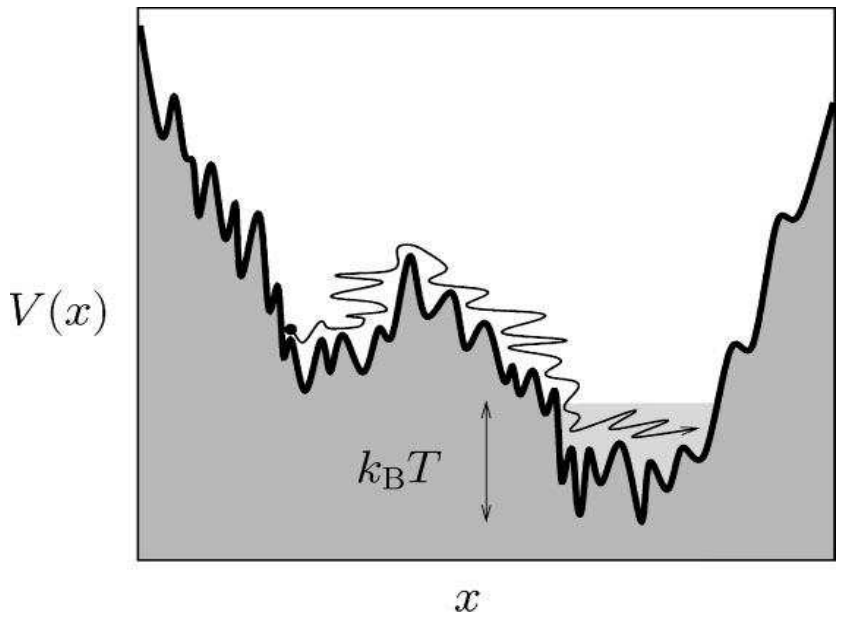
Speed up 30% of computation for the Heat Shock Response model

Minimum Action Method for Chemical Kinetic Systems

Metastability

$$\dot{x} = -\frac{\partial V(x)}{\partial x} + \sqrt{2k_B T}\,\dot{w}_t$$

Ergodicity $\Longrightarrow$ *Exploring the whole configuration space*



Time scale separation: Arrhenius Formula

$$\tau = \nu \exp(\Delta V / k_B T)$$

τ—Mean exit time from a local minimum

Minimum Action Method for Chemical Kinetic Systems

WKB method: (Dykman et. al. 94; Roma et. al. 05)

Forward equation :

$$\frac{\partial P(y,t)}{\partial t} = \Omega \sum_j \Big(b_j\big(y - \nu_j/\Omega\big) P\big(y - \nu_j/\Omega, t\big) - b_j\big(y\big) P\big(y,t\big) \Big)$$

WKB form : $$P(y,t) = C \exp(-\Omega S\big(y,t\big)\big)$$

Taylor expansion in terms of Ω :

$$\frac{\partial P(y,t)}{\partial t} = H P(y,t)$$

$$H(y,p) = \sum_j b_j(y)\big(e^{\nu_j \cdot p} - 1\big)\ , \qquad p^i = \frac{\partial}{\partial y_i} S_0(y,t)$$

Optimal path :

$$\dot{y}_t^i = \frac{\partial H}{\partial p^i}\ , \qquad \dot{p}_t^i = -\frac{\partial H}{\partial y^i}$$

Minimum Action Method for Chemical Kinetic Systems

Large Deviation of Stochastic Processes

Local function :

$$\ell(y, y') = \sup_{\theta \in \mathbb{R}^{N_S}} \left(< \theta, y' > - \sum_j b_j(y) \left(e^{<\theta, \nu_j>} - 1 \right) \right)$$

Action functional :

$$I_{[0,T]}(\varphi) = \int_0^T \ell\Big(\varphi(t), \dot{\varphi}(t)\Big) dt$$

Large Deviation Principle :

$$\mathbb{P}_x \left\{ \| \varphi^\Omega - \varphi \| \right\} \approx \exp\left(-\frac{1}{\Omega} I_{[0,T]}(\varphi) \right)$$

Mean exit time :

$$\tau \approx \exp\left\{ \Omega \inf_T I_T[\psi_T] \right\}$$

Large computing cost for the numerical evaluation of the local function :

$$y' = \sum_j e^{<\theta^*, \nu_j>} b_j(y) \nu_j$$

Minimum Action Method for Chemical Kinetic Systems

Minimum Action Method

(E, Ren and Vanden-Eijnden 04; Vanden-Eijnden and Heymann 07)

$$\frac{\partial \phi}{\partial s} = -\frac{\partial I_T[\phi]}{\partial \phi}$$

—Numerical approach for local rates

Minimum Action Method for Chemical Kinetic Systems —

(Di Liu *J. Chem. Phys.* 06; *J. Comp. Phys.* 08)

—Asymptotic approach for local rates based on Taylor expansions

Reaction advancement coordinates : (Van Kempen)

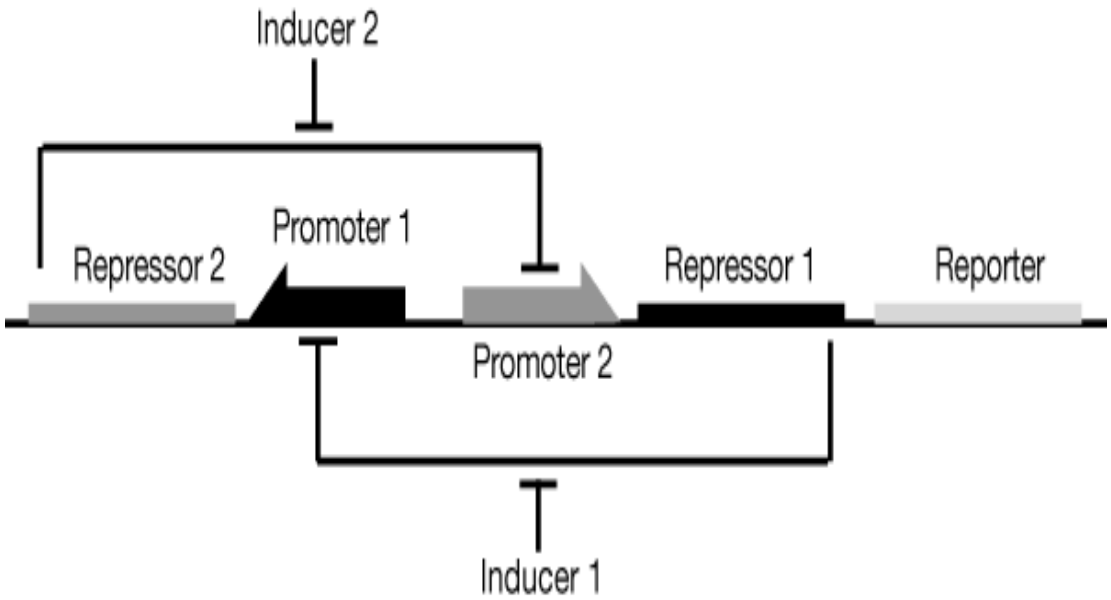
$$y_t = y_0 + \sum_j z_t^j \nu_j$$

Minimum Action Method for Chemical Kinetic Systems

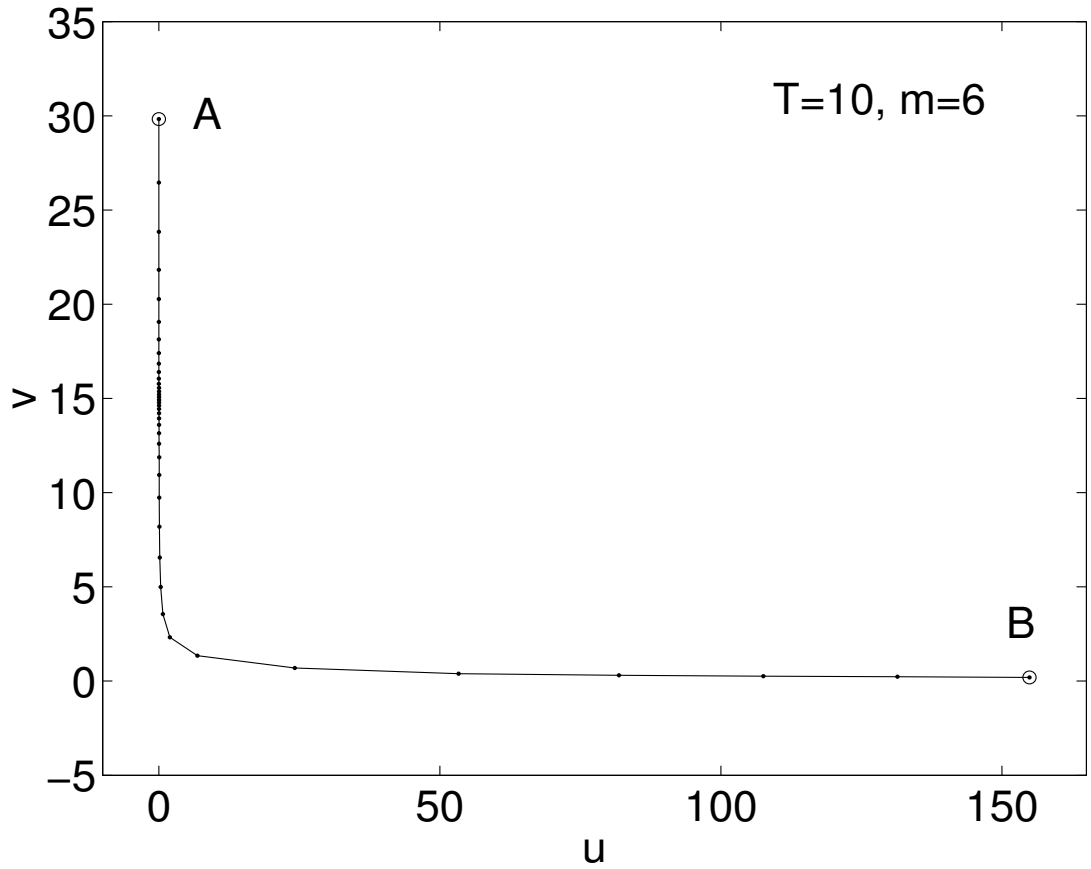
Toggle switch model Bacteriophage lambda virus infection in E. coli

$$R_1 = \left(\frac{\alpha_1}{1+v^\beta}, \ (1,0) \right) \ , \qquad R_2 = (u, \ (-1,0))$$

$$R_3 = \left(\frac{\alpha_2}{1+u^\gamma}, \ (0,1) \right) \ , \qquad R_4 = (v, \ (0,-1))$$



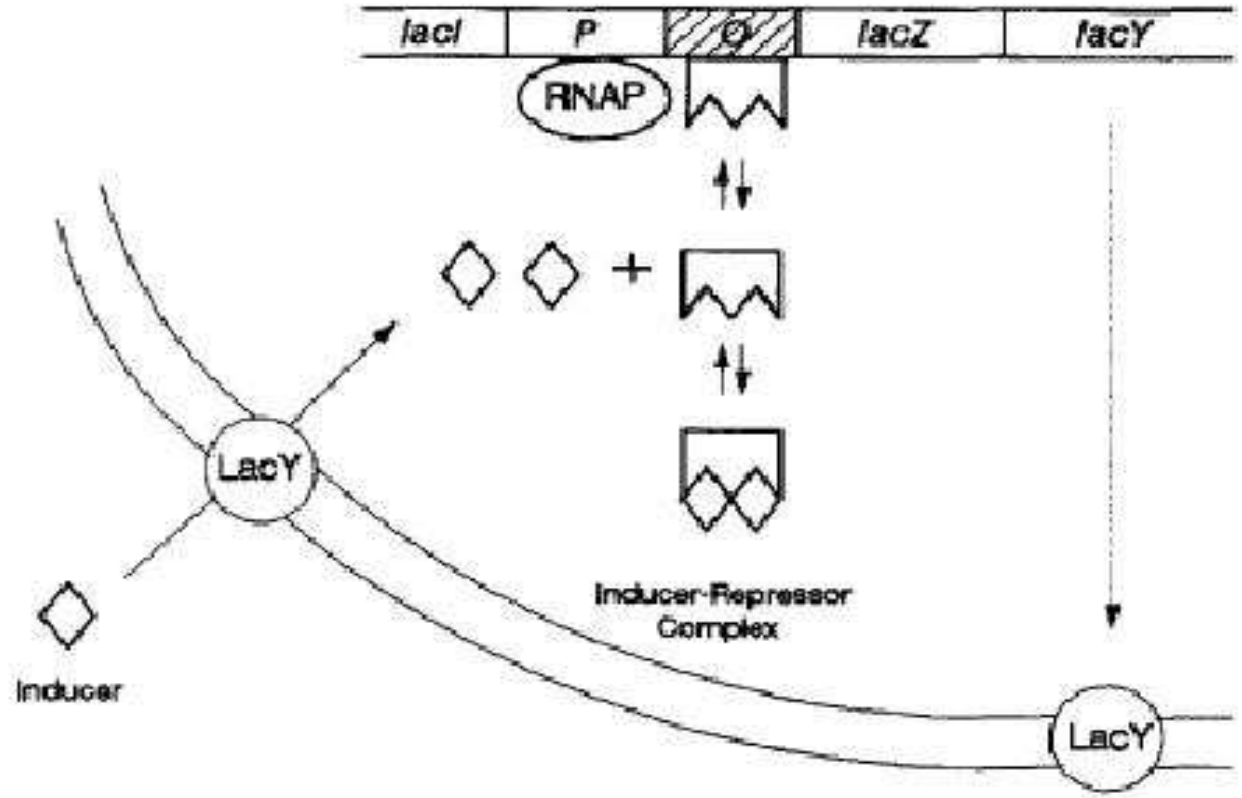
Minimum Action Method for Chemical Kinetic Systems



Qualitatively agrees with WKB method

Minimum Action Method for Chemical Kinetic Systems

Lactose utilization of E. Coli



Minimum Action Method for Chemical Kinetic Systems

Y—cell-associated Lac permease I—Inducer I_{ex}—extracellular Inducer

Generation of Lac permease :

$$R_{GEN} = k_1 O_T \frac{1 + K_1 I^2}{1 + K_1 I^2 + K_2 R_T}$$

Transportation of Inducer by Permease :

$$R_{ACTIVE} = \frac{\alpha I_{ex} Y}{\beta + I_{ex}}$$

Permease independent transport of inducer :

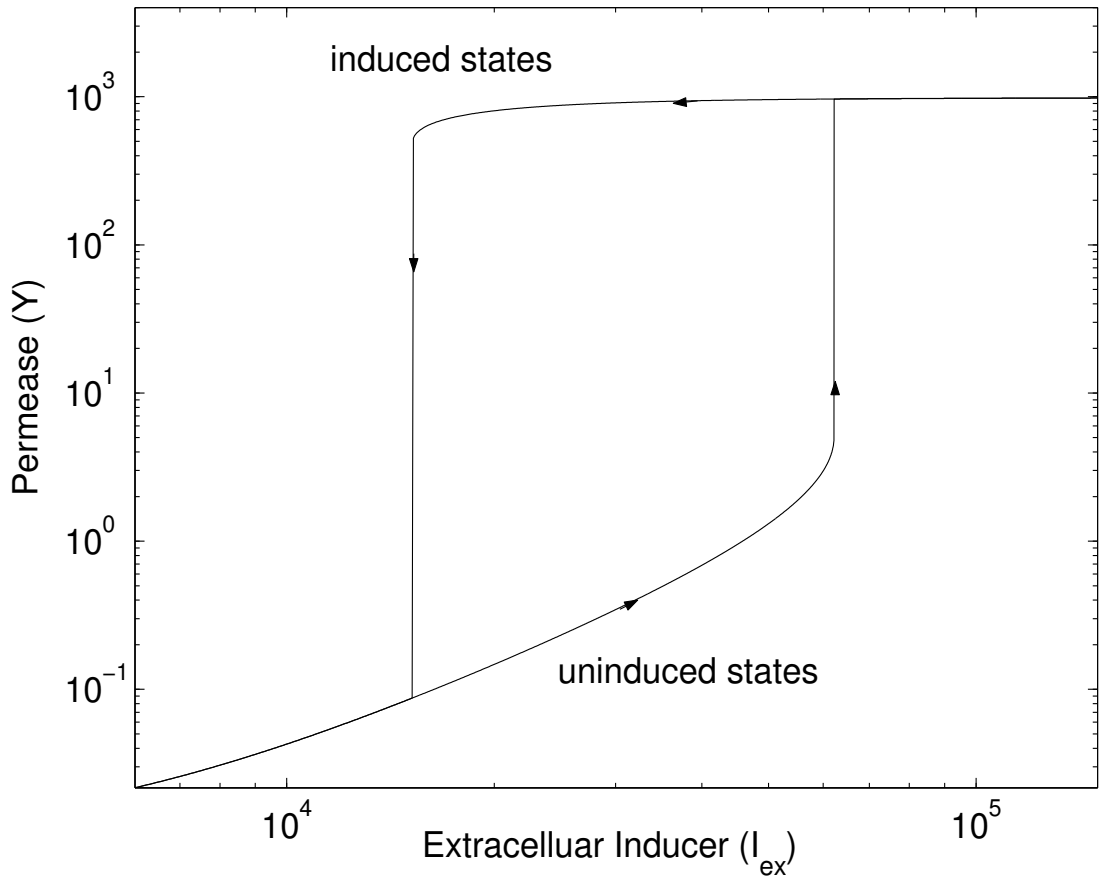
$$R_{FACILITATED} = \delta(I - I_{ex})$$

Dilution of Inducer and Permease due to growth :

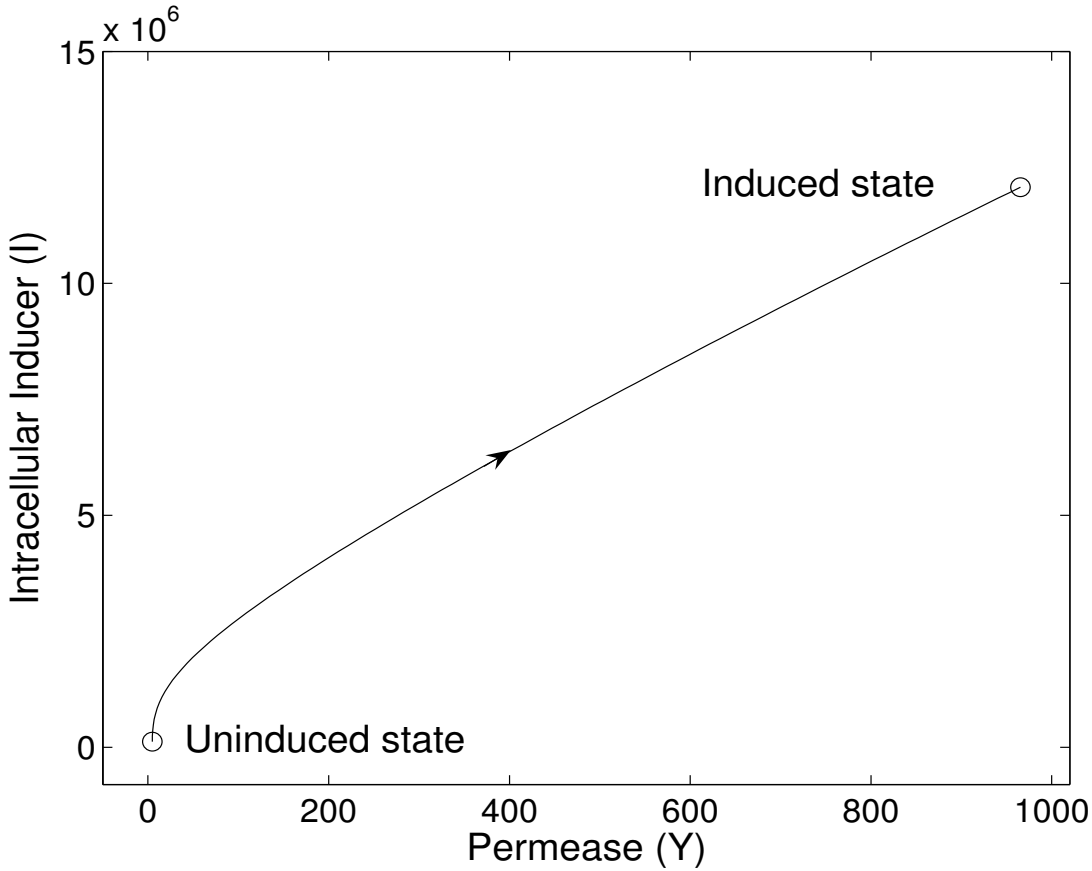
$$R_{dI} = k_2 I \ , \qquad R_{dY} = k_2 Y$$

Minimum Action Method for Chemical Kinetic Systems

Hysteresis loop under different external inducer populations



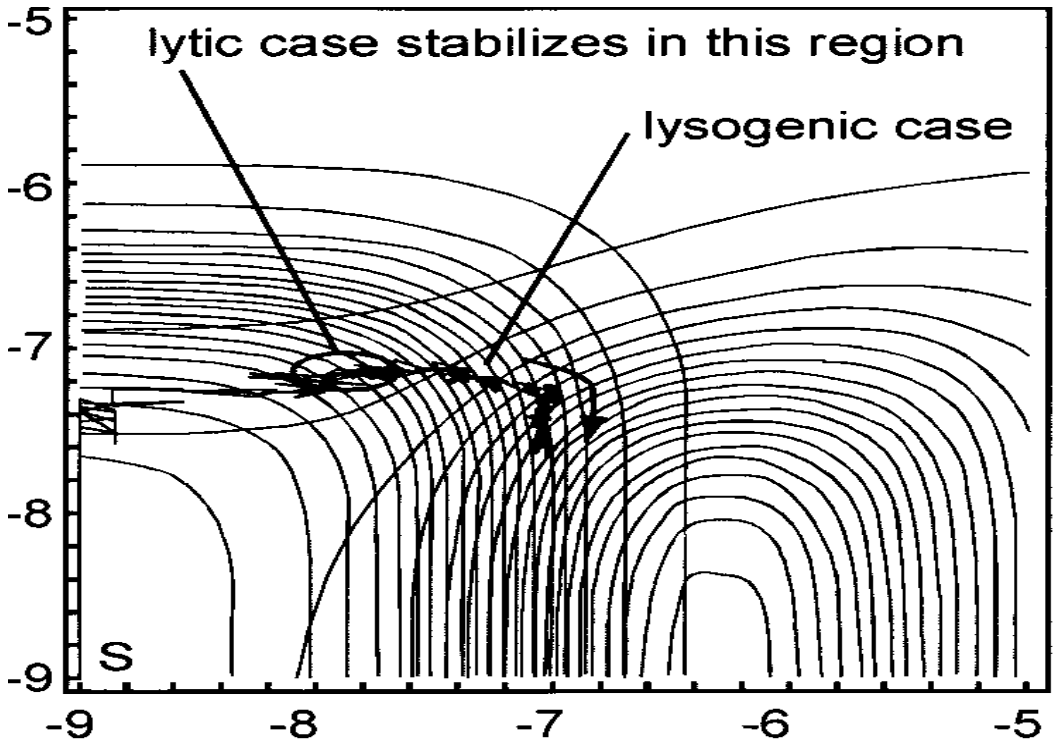
Minimum Action Method for Chemical Kinetic Systems



Transition path from uninduced state to induced state

Minimum Action Method for Chemical Kinetic Systems

(Arkin et al., 98)



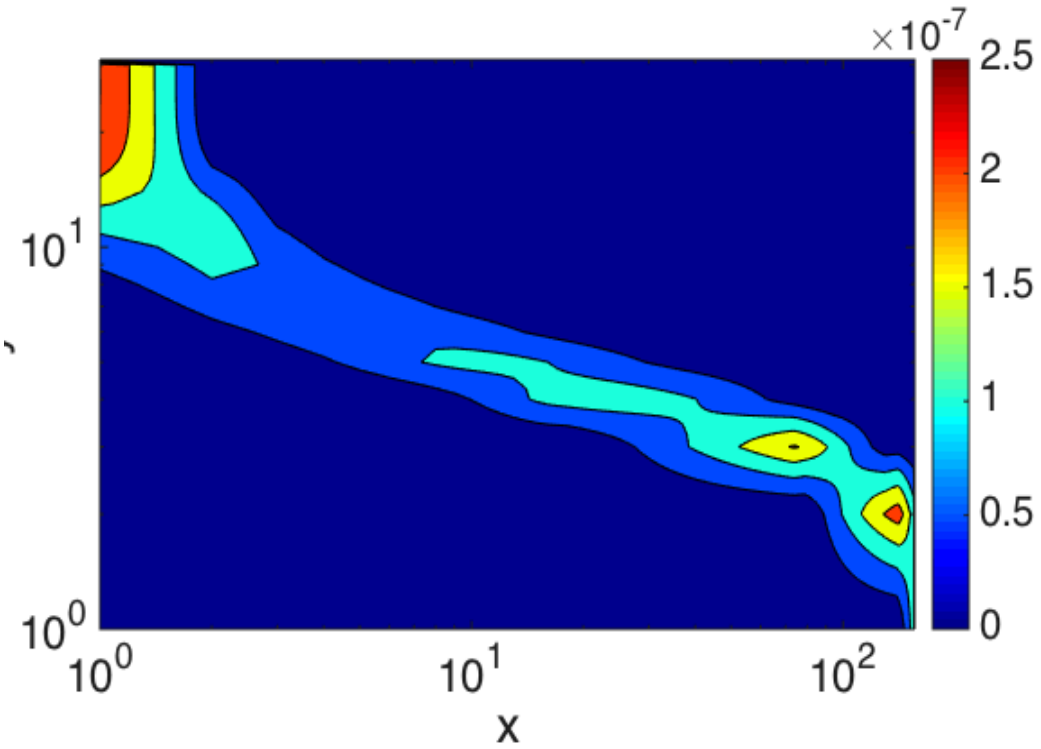
Transition Path Theory (TPT)

(E and Vanden-Eijnden 06; Metzner, Schutte, and Vanden-Eijnden 09)

Minimum Action Method for Chemical Kinetic Systems

(Du and Liu, preprint)

$$f_{ij}^{AB} = \lim_{s \to 0_+} \frac{1}{s} \mathbb{P}\Big(X(t) = i, X(t+s) = j, t \in R, t+s \in R \Big).$$



Concluding remarks

1. Multiscale methods for simulating multi scale stochastic reacting networks
2. Optimal error estimate is proved and efficiency is discussed
3. Applied to realistic Genetic Regulatory Networks
4. Current investigations on transition paths of metastable systems,
   system with delays (Chen and Liu, submitted).