q-bio Summer School

Albuquerque, NM
July 28 - August 12, 2014

Welcome

q-bio Welcome and Introductions

Colorado State University

- A brief history of q-bio
- Educational and Professional Goals
- Coursework
  - Course Selection
  - Projects and Presentations
  - Schedules (Lectures, Breakouts, Student Presentations)
  - Software
- Contacts and Sponsors
  - Course Leaders
  - Administrative Support
  - Sponsors
- Weekend Activities
  - Car Rentals and Trip Ideas.
  - Weather and Lightening Safety.
- Conference Registration and Lodging.
- Other and Questions.

QB1 - Stochastic Gene Regulation

Introduction and Course Overview

About me: Brian E Munsky

Colorado State University

Education:

B.S. (2000) and M.S. (2002) in Aerospace Engineering,
Pennsylvania State University

Ph.D. (2008) in Mechanical Engineering,
University of California at Santa Barbara

1st Annual q-bio Summer School (Student, 2007).

Experience:

2008-2010, Director's Postdoctoral fellow — Los Alamos National Lab
2010-2013, Richard P Feynman Distinguished Postdoctoral Fellow in Theory and Computing — Los Alamos National Lab
2014- Assistant Professor — Colorado State University, Chemical and Biological Engineering

Contact Information:

EMAIL: munsky@engr.colostate.edu
MOBILE: 805-252-0712

Course Outline - Week 1

- Monday, July 28
  - 09:00 - 10:15 — Introduction to stochastic effects in gene regulation (Munsky).
  - 14:40 - 17:00 — Crash Course in Stochastic Processes 1 (Flores).
  - 19:30 - 20:30 — Introduction to course Projects (Munsky)
- Tuesday, July 29
  - 08:30 - 10:00 — Modeling Evolution of the Myelodysplastic Syndrome (Kimmel)
  - 14:40 - 17:00 — Crash Course in Stochastic Processes 2 (Flores)
  - 19:30 - 21:00 — Stochastic Simulation Algorithms and other Kinetic Monte Carlo approaches (Munsky)
- Wednesday, July 30
  - 08:30 - 10:00 — Stochastic models of stem cell renewal and dedifferentiation in cancer (Jilkine)
  - 14:40 - 17:00 — Finite State Projection Analyses (Munsky)
- Thursday, July 31
  - 10:30 - 12:00 — TBA (Lidke)
- Friday, August 1
  - 14:00 - 17:00 — Parameter/Model Inference using Single-Cell Data (Munsky)

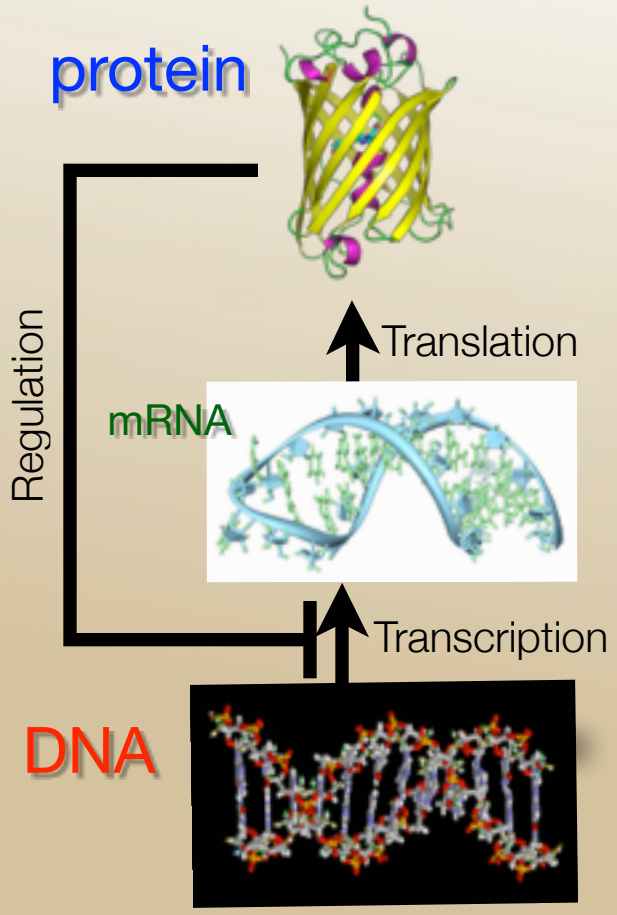
Course Outline - Week 2

- Monday, August 4
  - 08:30 - 10:00 — TBA (Ostheimer).
  - 10:30 - 12:00 — TBA (Shepherd).
  - 14:00 - 17:00 — Partial Least Square Regression (Ostheimer) — or — Spectroscopy techniques (Werner)
- Tuesday, August 5
  - 10:30 - 12:00 — TBA (Bellesia)
  - 14:00 - 17:00 — Spatial Statistics and emerging experimental/computational tools (Shepherd/Wilson) — or — TBA (Bellesia)
- Wednesday, August 6
  - 08:30 - 10:00 — Computation with Molecular Systems (Klavins).
  - 14:40 - 17:00 — Programming Multicellular Systems with gro (Klavins)
- Thursday, August 7
  - 08:30 - 10:00 — Quantitative tools to study signaling and gene regulation in single cells (Neuert)
  - 14:40 - 17:00 — Identifying gene regulatory models through variations in mRNA expression. (Neuert)

Stochastic Biochemistry: Theme Overview

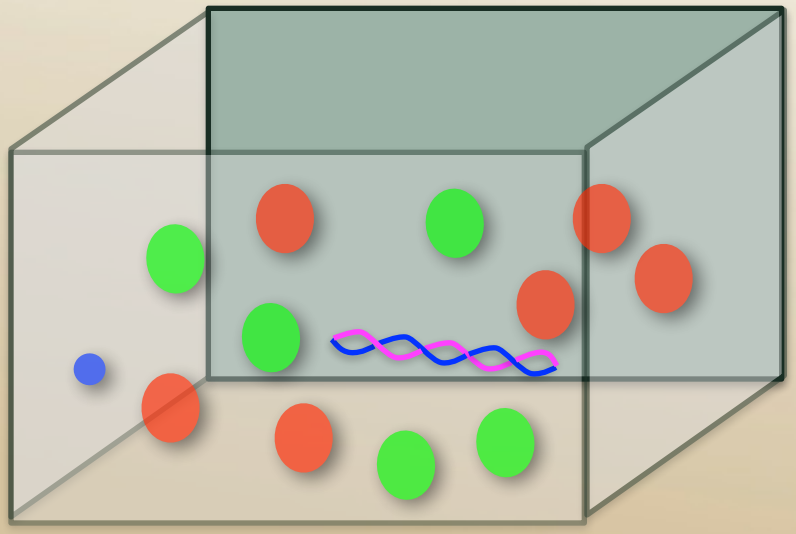
- Origins of Stochastic Phenomena
- Consequences of Stochastic Phenomena
- Observations of Stochastic Phenomena
- The Markov Description of Stochastic Biochemical Processes

Small numbers of important molecules



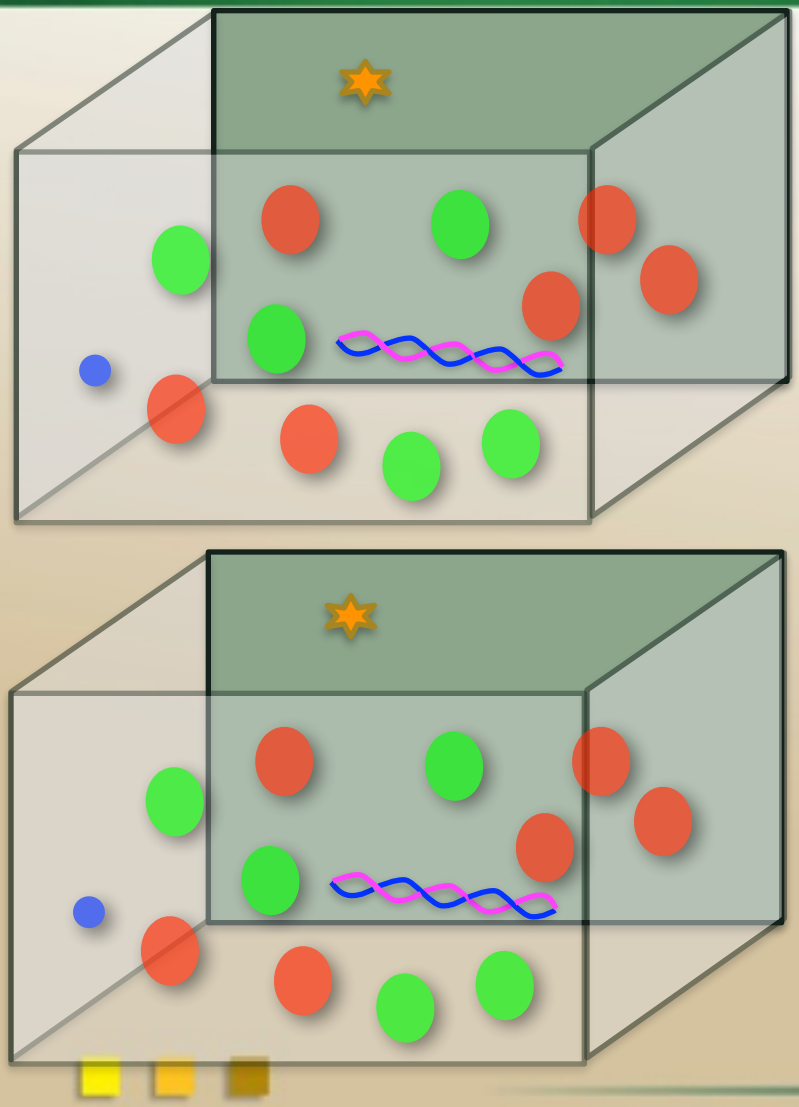
- Proteins build cellular structures, pass cellular information and regulate cellular activities. Variable copy numbers (~0-100,000/cell).
- mRNA transfer instructions for creating specific proteins. Low copy numbers (~0-100/cell).
- DNA contains all of the genetic instructions. Extremely low copy numbers (~0-5/cell).

Spatial fluctuations of cellular constituents



Thermal fluctuations can lead to randomness in times between reactions.

Competition of exclusive events



Different reactions lead to different consequences.

A molecular race may define the final outcome.

“Extrinsic” fluctuations

Changes in temperature, nutrients, radiation, chemicals, pressure, etc...

Fluctuations of upstream genes, transcriptional or translational machinery (polymerases, ribosomes), intercellular signals.

Unknown elements left out from the current model (i.e., everything else).

Stochastic Biochemistry: Theme Overview

- Origins of Stochastic Phenomena
- Consequences of Stochastic Phenomena
- Observations of Stochastic Phenomena
- The Markov Description of Stochastic Biochemical Processes

Stochastic Effects Lead to Phenotypical Differences

Fingerprints of identical twins

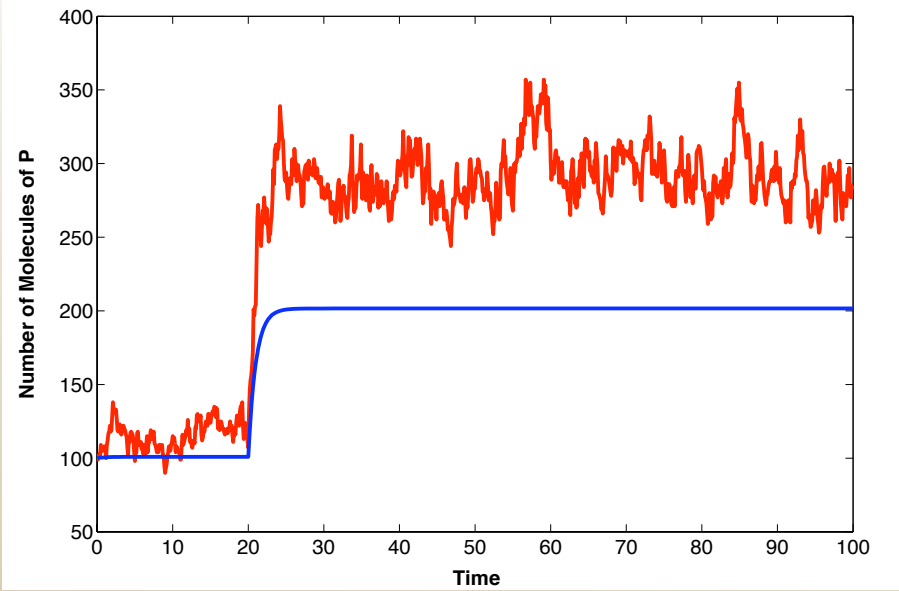
Cc, the first cloned cat and her genetic mother, Rainbow

J. Raser and E. O'Shea, "Noise in Gene Expression: Origins, Consequences, and Control", Science, 2005

Signal Amplification and Damping

$$\phi \underset{k_a S}{\overset{k}{\rightleftharpoons}} I \xrightarrow{k_p} P \xrightarrow{1} \phi$$

$$\phi \underset{k_d}{\overset{k_s}{\rightleftharpoons}} S$$

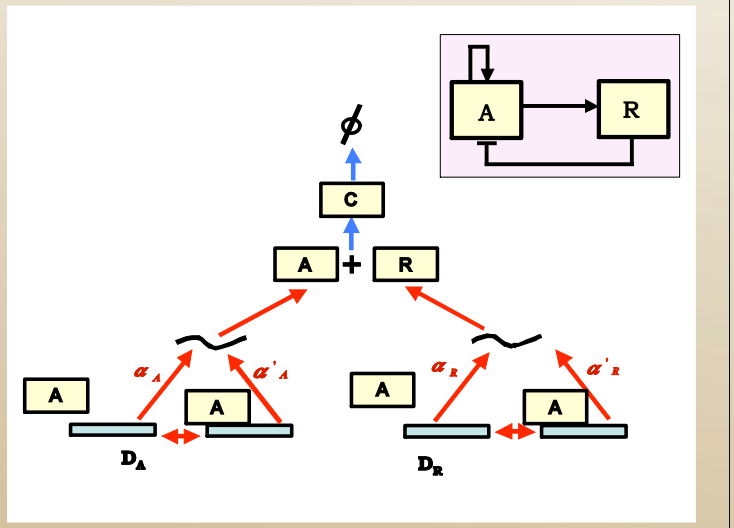


Johan Paulsson , Otto G. Berg , and Måns Ehrenberg, "Stochastic Focusing: Fluctuation-enhansed sensitivity of intracellular regulation" PNAS 2000

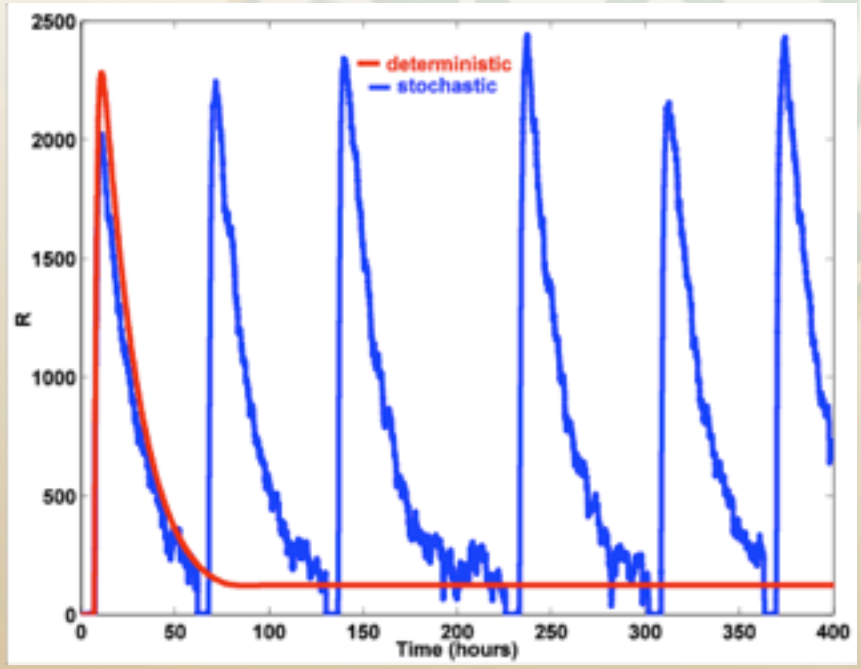
- Stochastic mean value different from deterministic steady state
- Noise *enhances* signal!

Noise Induced Oscillations

Circadian Rhythms

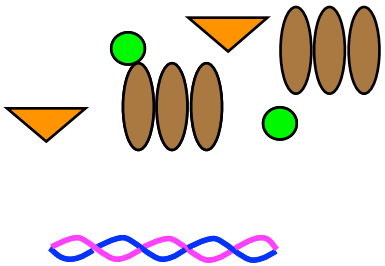


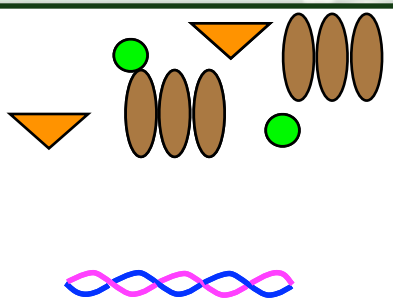
Vilar, Kueh, Barkai, Leibler, PNAS 2002



- Oscillations disappear from deterministic model after a small change in one parameter.
  - These oscillations may be restored by noise.
  - Oscillation Regularity is altered by tuning the noise level [El-Samad, Khammash]

Stochastic Switches

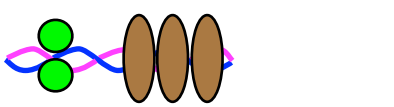


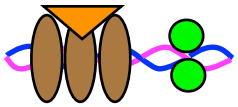


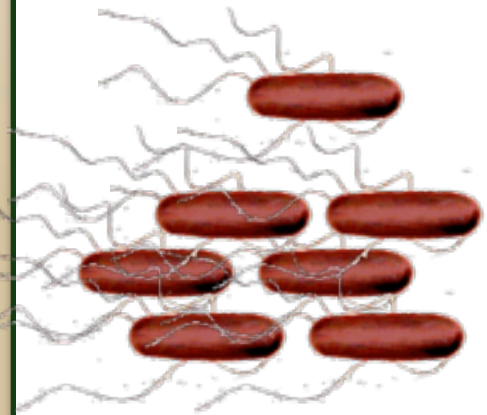
Same chemical environment.

Same genetic code.

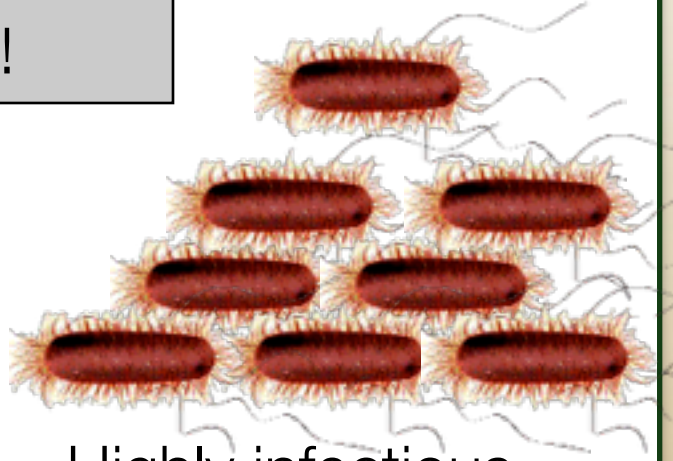
Random reactions can lead to vastly different results!





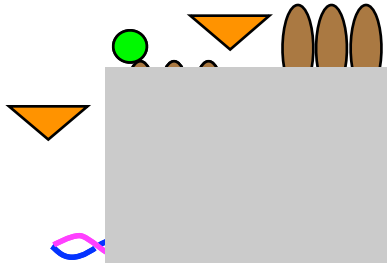


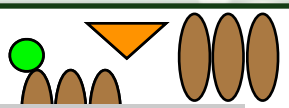
Harmless phenotype.



Highly infectious phenotype.

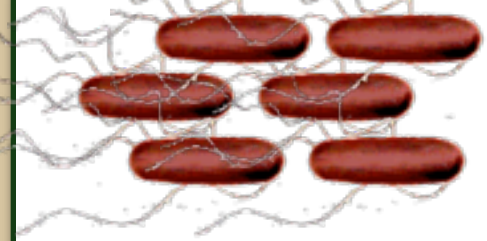
Stochastic Switches



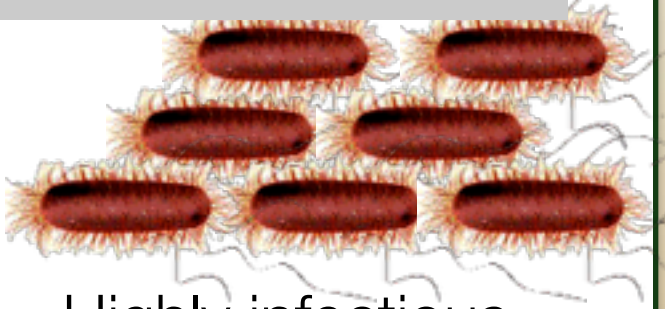


For these systems, we need single cell analyses to answer:

- ★ What will happen?
- ★ How frequently?
- ★ Why does it happen?
- ★ Under what conditions?
- ★ What advantages does it provide?
- ★ How can we prevent it?
- ★ How can we cause it?



Harmless phenotype.

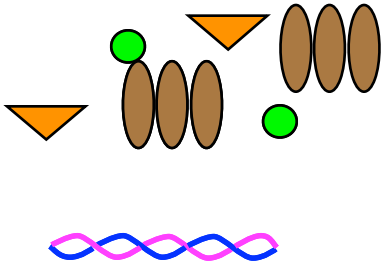


Highly infectious phenotype.

Stochastic Biochemistry: Theme Overview

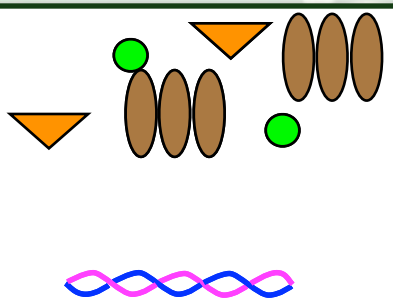
- Origins of Stochastic Phenomena
- Consequences of Stochastic Phenomena
- Observations of Stochastic Phenomena
- The Markov Description of Stochastic Biochemical Processes

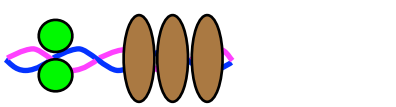
The Importance of Single Cell Analyses



Same chemical environment.

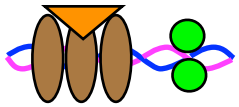
Same genetic code.

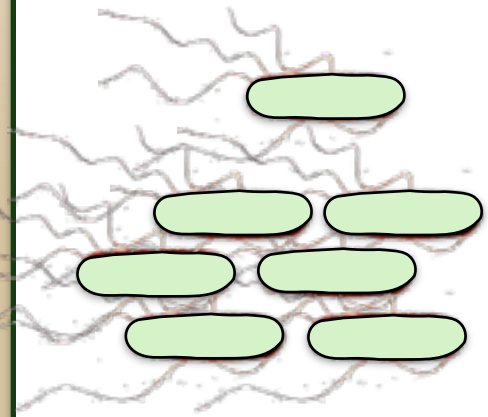




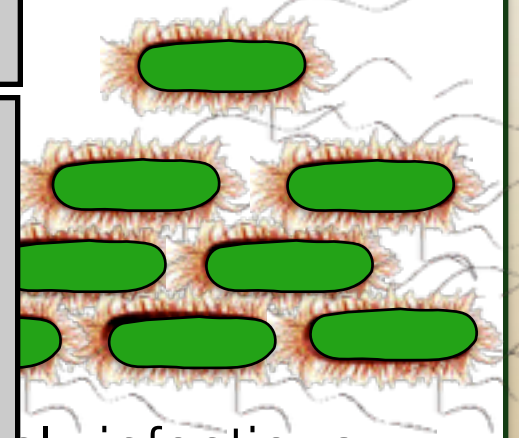
Random reactions can lead to vastly different results!

Fluorescent labels and specific genetic mutations make it possible to observe the dynamics of single-cell heterogeneity.





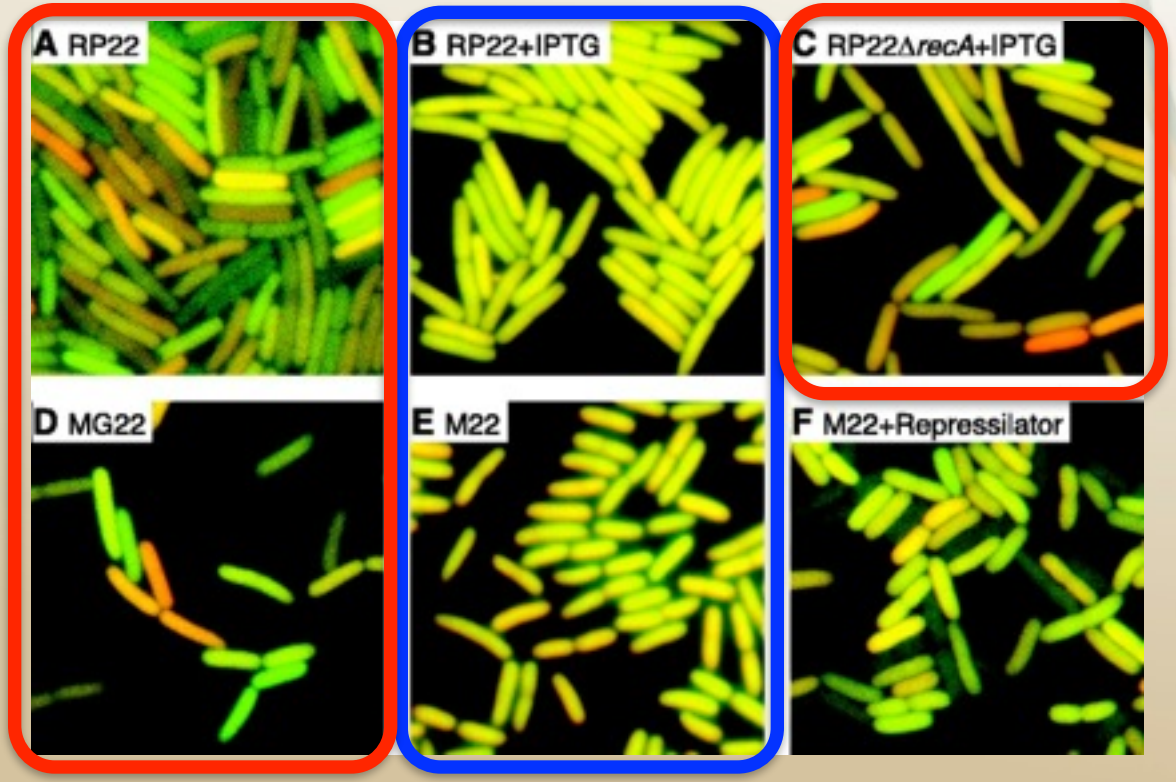
Harmless phenotype.



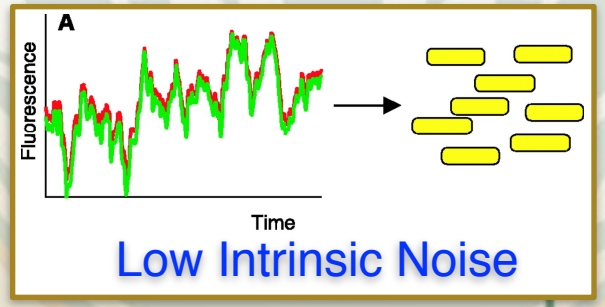
Highly infectious phenotype.

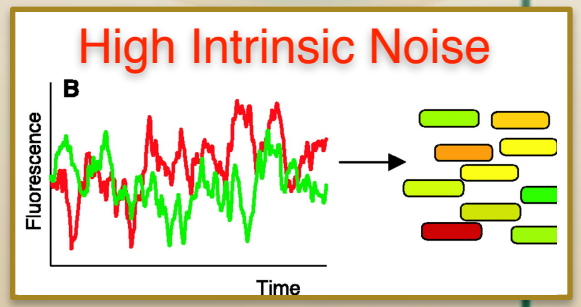
“Intrinsic” versus “Extrinsic” Noise

- Variability is present and can be measured



Elowitz et al, “Stochastic Gene Expression in a Single Cell”, Science 2002



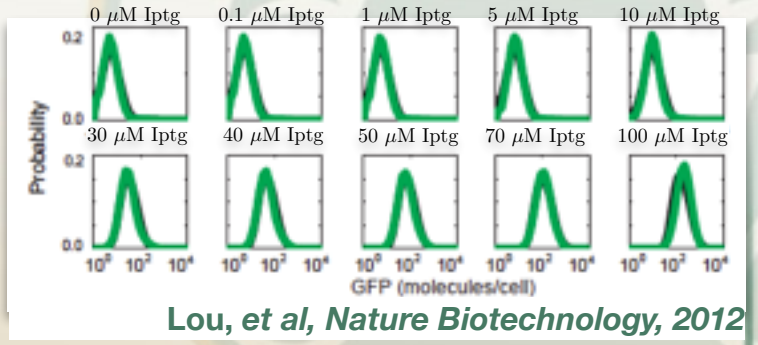


- Inserted two reporters on the chromosome (cfp, yfp)
- Each was controlled by the same promoter
- Expression of cfp shown in green, yfp in red

Experimental tools for single-cell analyses

Flow Cytometry

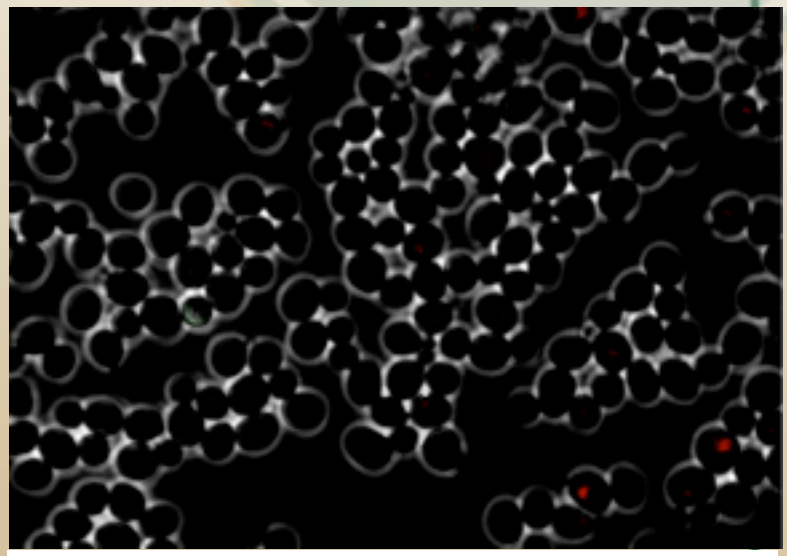
- Measure expression with fluorescent proteins or antibody labels for thousands of cells per second.



Lou, *et al, Nature Biotechnology, 2012*

Time Lapse Fluorescence Microscopy

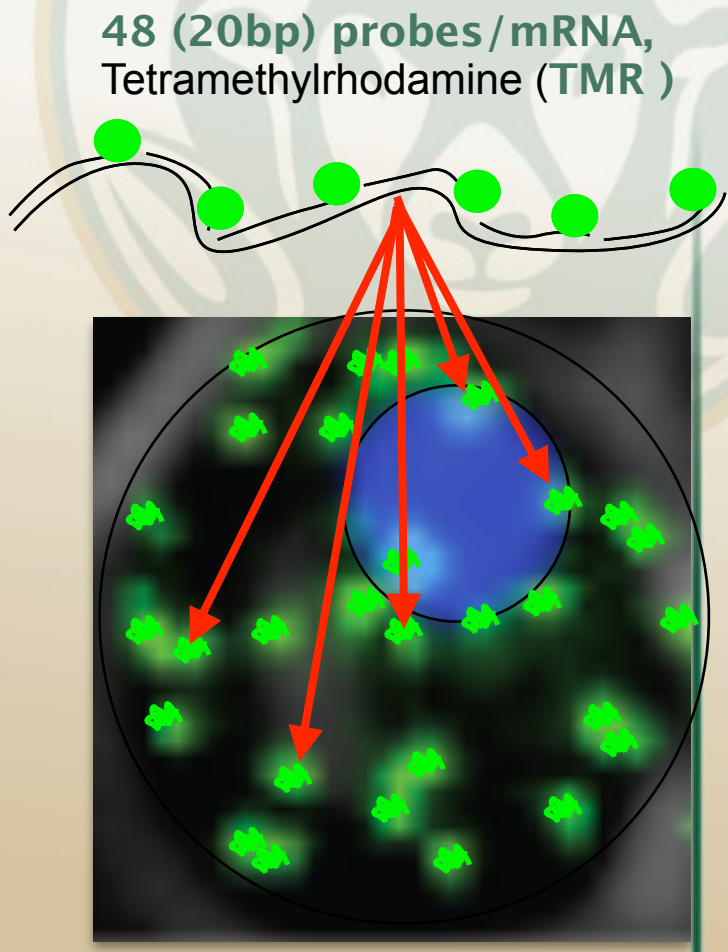
- Measure spatial and temporal properties of fluorescent protein responses.



(Neuert, Munsky, *et al*, 2013)

Single-Molecule FISH (smFISH)

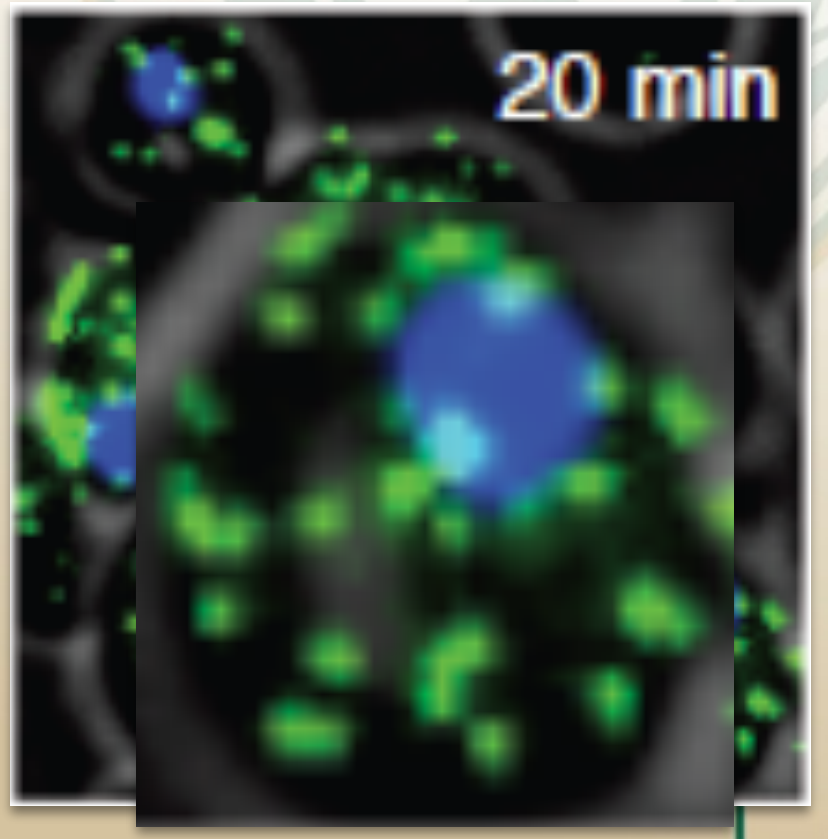
- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.



(Neuert, Munsky, *et al*, 2013)

Single-Molecule FISH (smFISH)

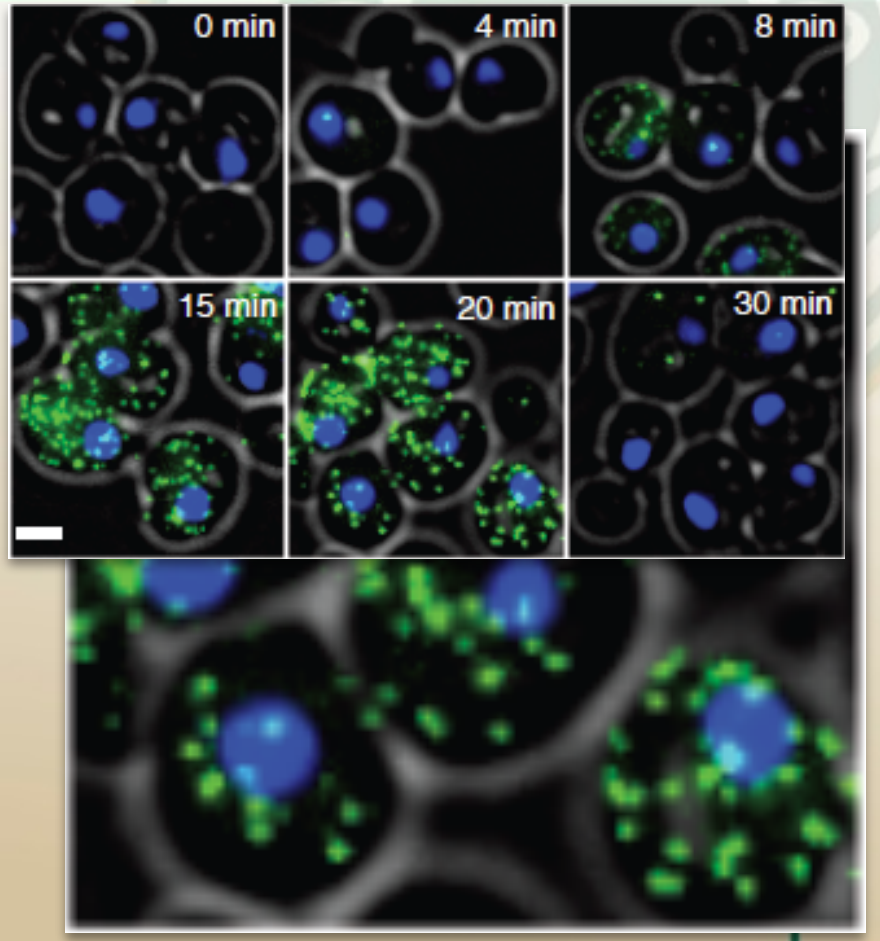
- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.
- Spatial localization enable inter- and intra-nuclear detection.



(Neuert, Munsky, *et al*, 2013)

Single-Molecule FISH (smFISH)

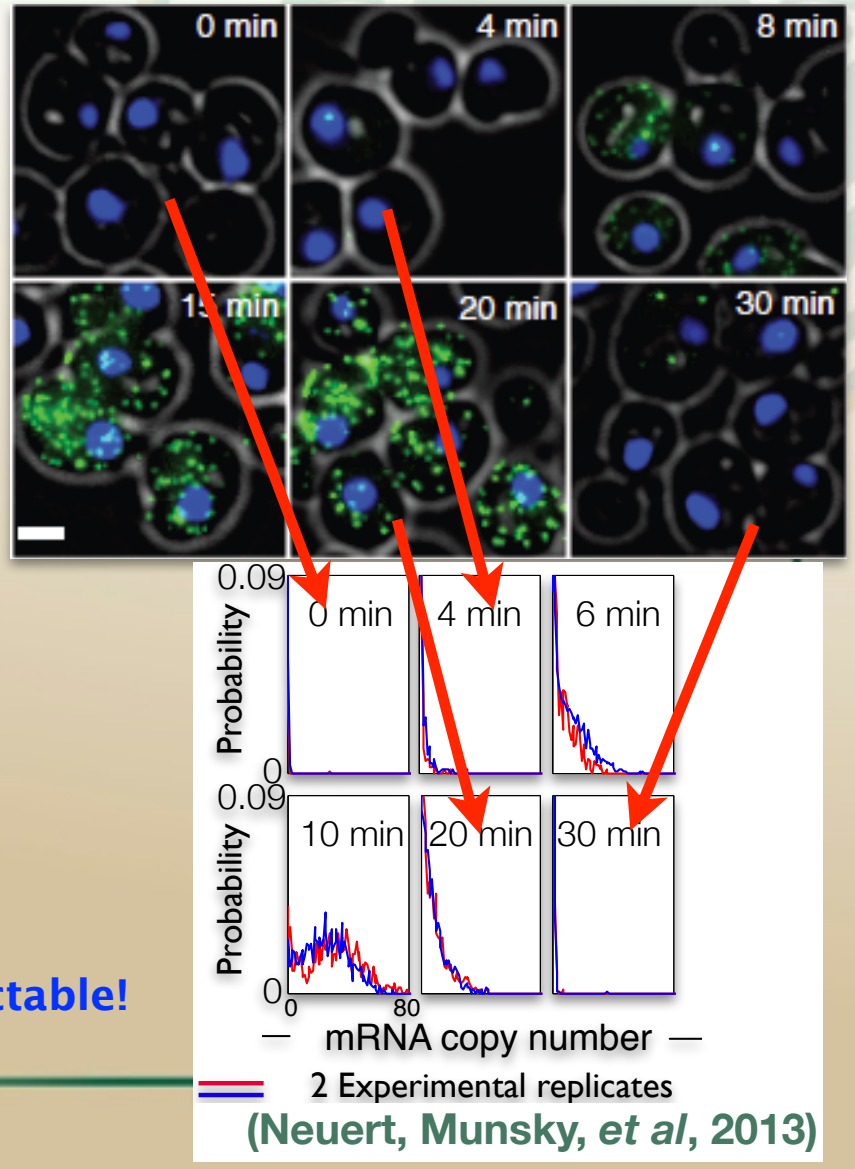
- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.
- Spatial localization enable inter- and intra-nuclear detection.
- Fast time resolution (1-2 min).



(Neuert, Munsky, *et al*, 2013)

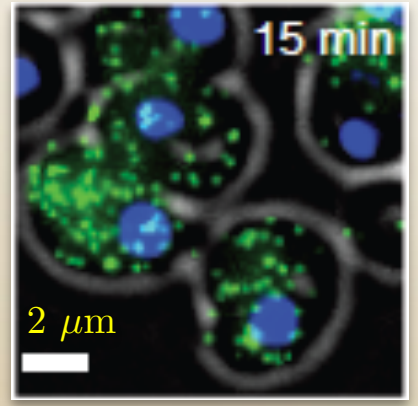
Single-Molecule FISH (smFISH)

- Endogenous mRNA's can be labeled with single molecule Fluorescence *in situ* Hybridization (smFISH--Femino, 1998, Raj, 2008).
- Many probes (~50) are attached to endogenous mRNA.
- High signal-to-noise ratio enables single-molecule detection.
- Spatial localization enable inter- and intra-nuclear detection.
- Fast time resolution (1-2 min).

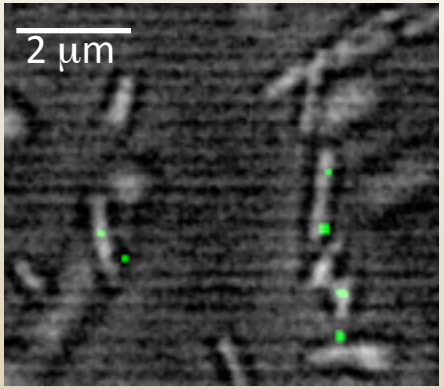


Statistics are repeatable and therefore predictable!

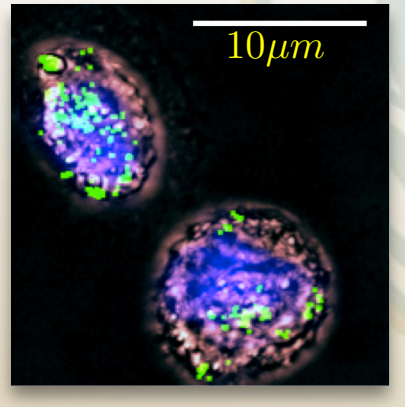
Single-Molecule FISH (smFISH)



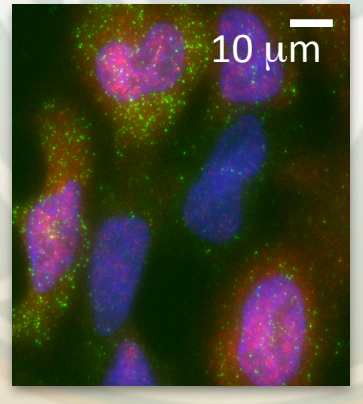
STI1 mRNA in *Saccharomyces cerevisiae* (budding yeast)
-G. Neuert (MIT)



Ysr35 sRNA in *Yersinia Pseudotuberculosis* (339nt)
-D. Shepherd (LANL)



IL1a mRNA in THP1 cells
-D. Shepherd (LANL)



c-Fos mRNA (green) and p-p38 kinase (red) in U2OS cells
-A. Senecal (CNRS)

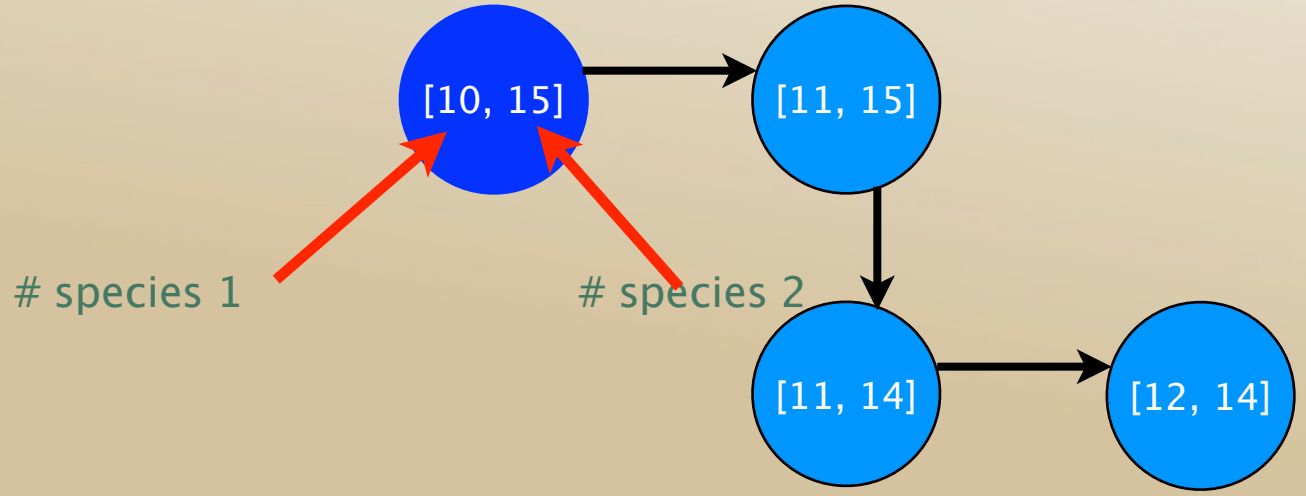
smFISH has been applied to many different RNA in many different organisms

Stochastic Biochemistry: Theme Overview

- Origins of Stochastic Phenomena
- Consequences of Stochastic Phenomena
- Observations of Stochastic Phenomena
- The Markov Description of Stochastic Biochemical Processes

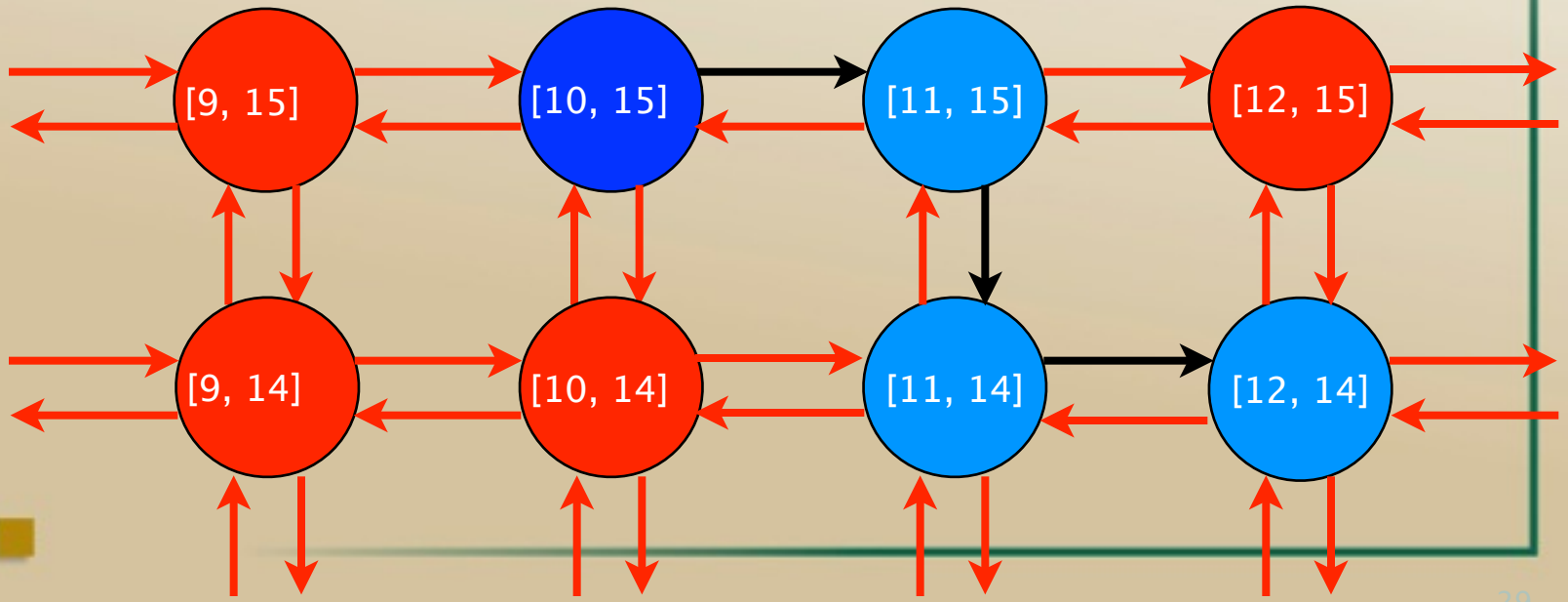
A Markov description of single-cell gene regulation

- At any time, the state of the system is defined by its integer population vector: $\mathbf{x} \in \mathbb{Z}^N$
- Reactions are transitions from one state to another:

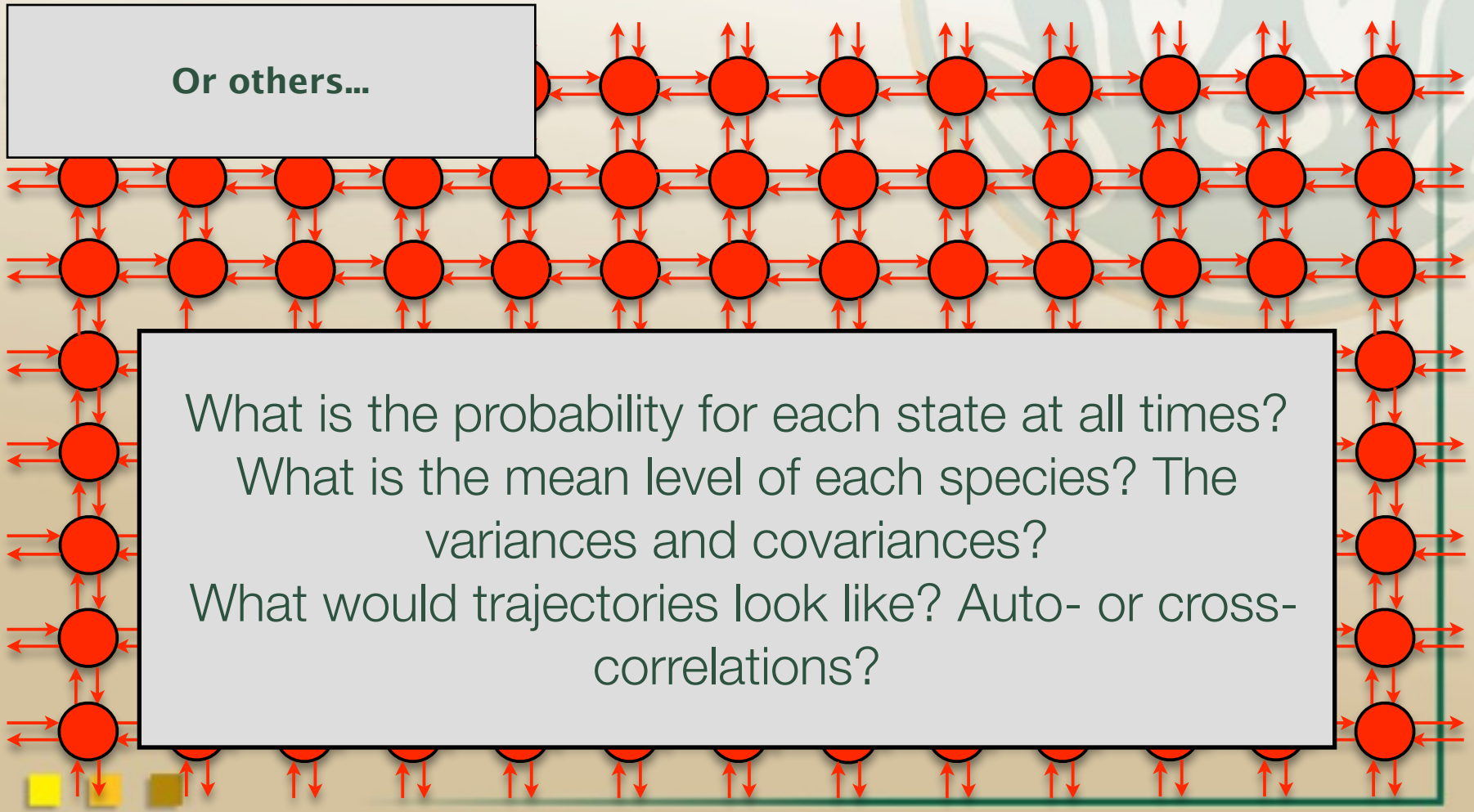


A Markov description of single-cell gene regulation

- At any time, the state of the system is defined by its integer population vector: $\mathbf{x} \in \mathbb{Z}^N$
- Reactions are transitions from one state to another.
- These reactions are random, others could have occurred:



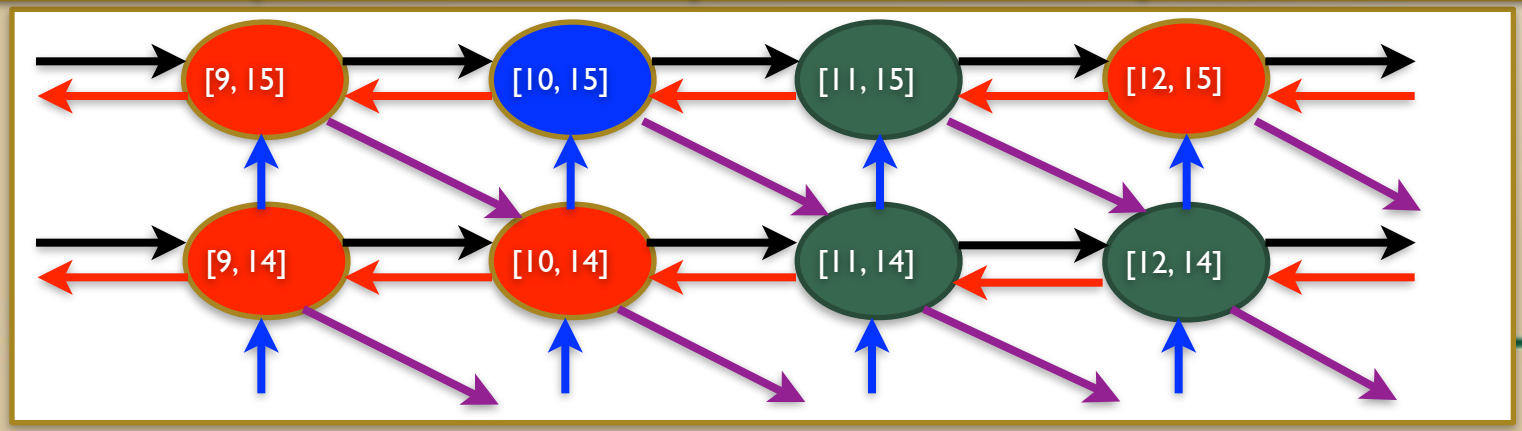
A Markov description of single-cell gene regulation



Reaction Stoichiometry

- The Stoichiometric vector, s, refers to the relative change in the population vector after a reaction.
- There may be many different reactions for a given stoichiometry.

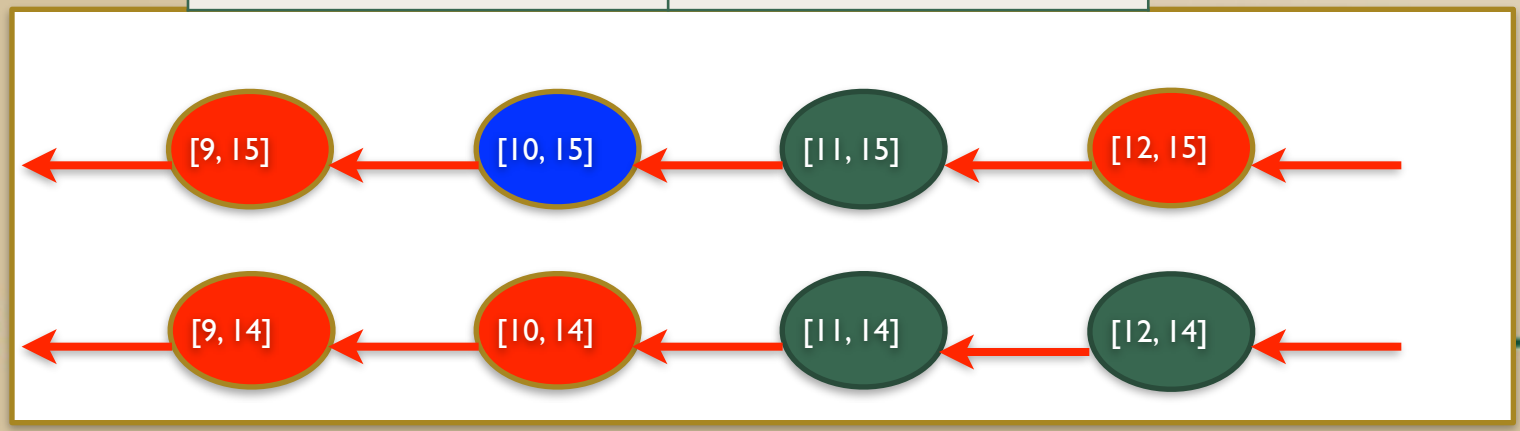
$\mathbf{s}_1 = [1, 0]^T$	$\mathbf{s}_2 = [-1, 0]^T$	$\mathbf{s}_3 = [0, 1]^T$	$\mathbf{s}_4 = [1, -1]^T$
$\mathcal{S}_1 \rightarrow \mathcal{S}_1 + \mathcal{S}_1$	$\mathcal{S}_1 + \mathcal{S}_1 \rightarrow \mathcal{S}_1$	$\mathcal{S}_2 \rightarrow \mathcal{S}_2 + \mathcal{S}_2$	$\mathcal{S}_2 \rightarrow \mathcal{S}_1$
$\mathcal{S}_2 \rightarrow \mathcal{S}_2 + \mathcal{S}_1$	$\mathcal{S}_1 + \mathcal{S}_2 \rightarrow \mathcal{S}_2$	$\mathcal{S}_1 \rightarrow \mathcal{S}_1 + \mathcal{S}_2$	$\mathcal{S}_1 + \mathcal{S}_2 \rightarrow \mathcal{S}_1 + \mathcal{S}_1$
$\emptyset \rightarrow \mathcal{S}_1$	$\mathcal{S}_1 \rightarrow \emptyset$	$\emptyset \rightarrow \mathcal{S}_2$	$\mathcal{S}_2 + \mathcal{S}_2 \rightarrow \mathcal{S}_1 + \mathcal{S}_2$



Reaction Propensities

- The propensity, $\mathbf{w}$, of a reaction is its rate.
- $\mathbf{w}_\mu dt$ is the probability that the μ^{th}reaction will occur in a time step of length dt.
- Typically, propensities depend only upon reactant populations.

$\mathbf{s}_2 = [-1, 0]^T$	$w_2(x_1, x_2)$
$\mathcal{S}_1 + \mathcal{S}_1 \rightarrow \mathcal{S}_1$	$k_1 x_1 (x_1 - 1)/2$
$\mathcal{S}_1 + \mathcal{S}_2 \rightarrow \mathcal{S}_2$	$k_2 x_1 x_2$
$\mathcal{S}_1 \rightarrow \emptyset$	$k_3 x_1$



The (Chemical) Master Equation

- The CME Description
- Example: Transcription as a Birth-Death Process.
- Kinetic Monte Carlo Approaches
- Finite State Projection Approaches
- Moment Computations

See notes online

The Chemical Master Equation

Prob. that no reactions fire in $[t, t+dt] = 1 - \sum_k w_k(x)dt + \mathcal{O}(dt^2)$

Prob. that reaction R_k fires once in $[t, t+dt] = w_k(x)dt + \mathcal{O}(dt^2)$

Prob. that more than one reaction fires in $[t, t+dt] = \mathcal{O}(dt^2)$

$$p(x, t+dt) = \underbrace{p(x,t)}_{\text{at } x}\underbrace{\left(1 - \sum_k w_k(x)dt + \mathcal{O}(dt^2)\right)}_{\text{No reaction fires}} + \sum_k \underbrace{p(x-s_k,t)}_{\substack{R_k \text{ reaction} \\ \text{away from } x}}\underbrace{\left(\sum_k w_k(x)dt + \mathcal{O}(dt^2)\right)}_{R_k \text{ fires once}} + \underbrace{\mathcal{O}(dt^2)}_{\substack{\text{more than one} \\ \text{reaction in } dt}}$$

$$p(x, t+dt) - p(x,t) = -p(x,t)\sum_k w_k(x)dt + \sum_k p(x-s_k,t)w_k(x)dt + \mathcal{O}(dt^2)$$

The Chemical Master Equation

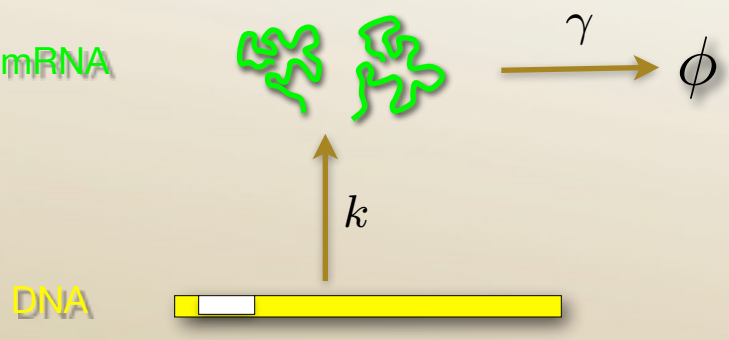
$$\frac{dp(x,t)}{dt} = -p(x,t)\sum_k w_k(x) + \sum_k p(x-s_k,t)w_k(x-s_k)$$

The (Chemical) Master Equation

- The CME Description
- Example: Transcription as a Birth-Death Process.
- Kinetic Monte Carlo Approaches
- Finite State Projection Approaches
- Moment Computations

See notes online

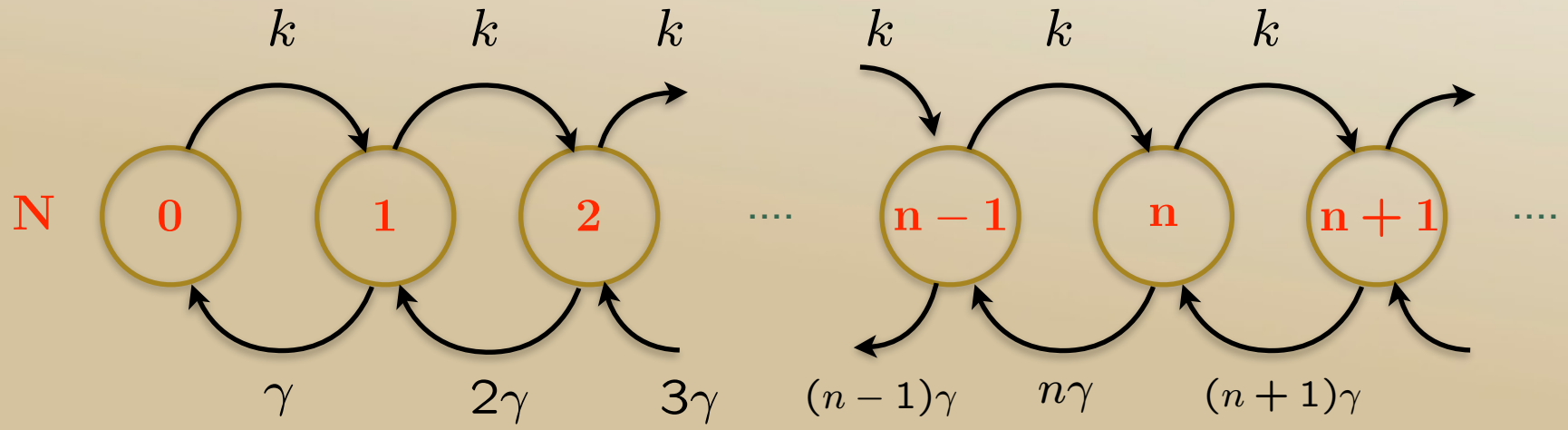
RNA Copy Number as a Random Variable



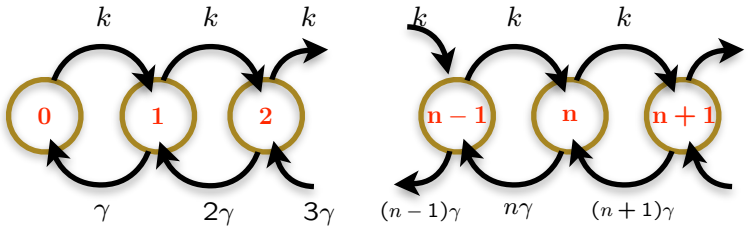
mRNA copy number $N(t)$ is a random variable

Transcription: Probability a single mRNA is transcribed in time dt is kdt

Degradation: Probability a single mRNA is degraded in time dt is $n\gamma dt$



Key Question:



Find $p(n,t)$, the probability that $N(t) = n$.

$$P(n, t+dt) = P(n-1, t) \cdot k dt$$ Prob.$\{N(t) = n-1$ and mRNA created in [t,t+dt)$\}$

$$+ \; P(n+1, t) \cdot (n+1)\gamma dt$$ Prob.$\{N(t) = n+1$ and mRNA degraded in [t,t+dt)$\}$

$$+ \; P(n, t) \cdot (1 - kdt)(1 - n\gamma dt)$$ Prob.$\{N(t) = n$ and mRNA not created nor degraded in [t,t+dt)$\}$

$$P(n, t+dt) - P(n,t) = P(n-1,t)kdt + P(n+1,t)(n+1)\gamma dt - P(n,t)(k + n\gamma)dt + O(dt^2)$$

Dividing by dt and taking the limit as $dt \to 0$

The Chemical Master Equation

$$\frac{d}{dt}P(n,t) = kP(n-1,t) + (n+1)\gamma P(n+1,t) - (k + n\gamma)P(n,t)$$

mRNA Stationary Distribution

We look for the stationary distribution $P(n,t) = p(n)\ \forall t$

The stationary solution satisfies: $\frac{d}{dt}P(n,t) = 0$

From the Master Equation ...

$$(k + n\gamma)p(n) = kp(n-1) + (n+1)\gamma p(n+1)$$

$n = 0$ $\quad kp(0) = \gamma p(1)$

$n = 1$ $\quad kp(1) = 2\gamma p(2)$

$n = 2$ $\quad kp(2) = 3\gamma p(3)$

$\vdots$

$$kp(n-1) = n\gamma\ p(n)$$

mRNA Stationary Distribution

$kp(n-1) = n\gamma\ p(n)$ We can express $p(n)$ as a function of $p(0)$:

$$\begin{aligned} p(n) &= \frac{k}{\gamma}\frac{1}{n}\ p(n-1) \\ &= \left(\frac{k}{\gamma}\right)^2 \frac{1}{n}\ \frac{1}{n-1}\ p(n-2) \\ &\vdots \\ &= \left(\frac{k}{\gamma}\right)^n \frac{1}{n!}\ p(0) \end{aligned}$$

We can solve for $p(0)$ using the fact $\sum_{n=0}^{\infty} p(n) = 1$

$$\begin{aligned} 1 &= \sum_{n=0}^{\infty} \left(\frac{k}{\gamma}\right)^n \frac{1}{n!}\ p(0) \\ &= e^{k/\gamma}\ p(0) \quad \Rightarrow \quad p(0) = e^{-k/\gamma} \end{aligned}$$

$$p(n) = e^{-a}\frac{a^n}{n!} \qquad a = \frac{k}{\gamma}$$

Poisson Distribution

We can compute the mean and variance of the Poisson RV $\bar{N}$ with density $p(n) = e^{-a}\frac{a^n}{n!}$:

$$\mu = E[\bar{N}] = \sum_{n=0}^{\infty} np(n) = e^{-a} \sum_{n=0}^{\infty} n\frac{a^n}{n!} = a$$

The second moment

$$E[\bar{N}^2] = \sum_{n=0}^{\infty} n^2 p(n) = a^2 + a$$

Therefore,

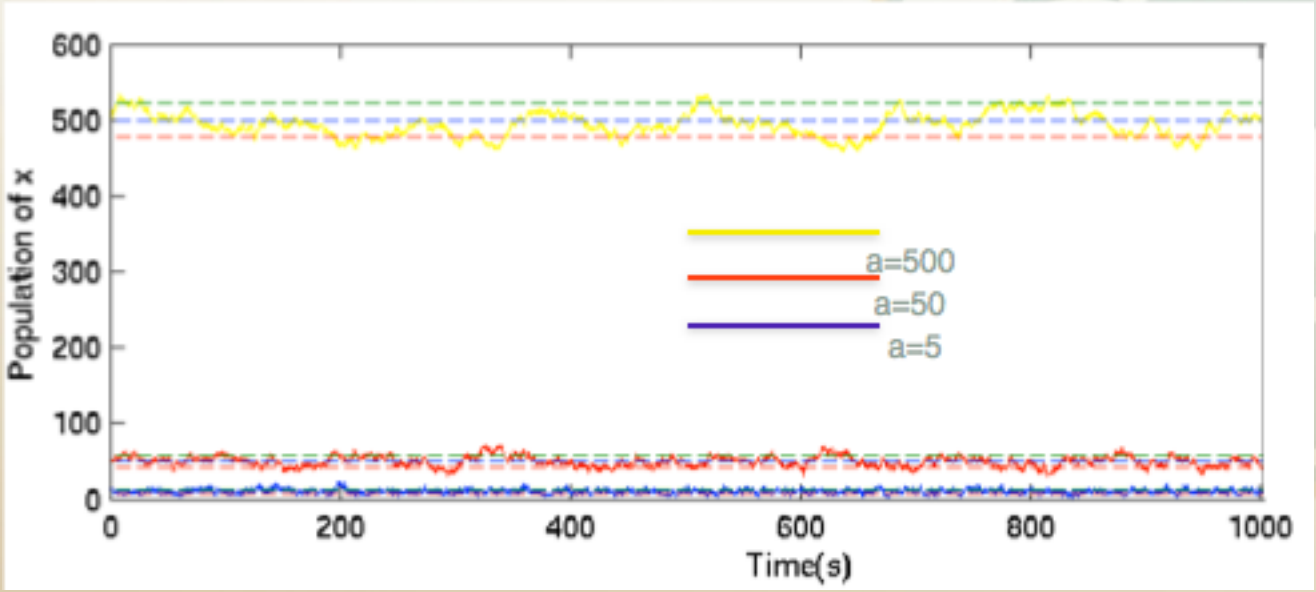
$$\sigma^2 = E[\bar{N}^2] - E[\bar{N}]^2 = a$$

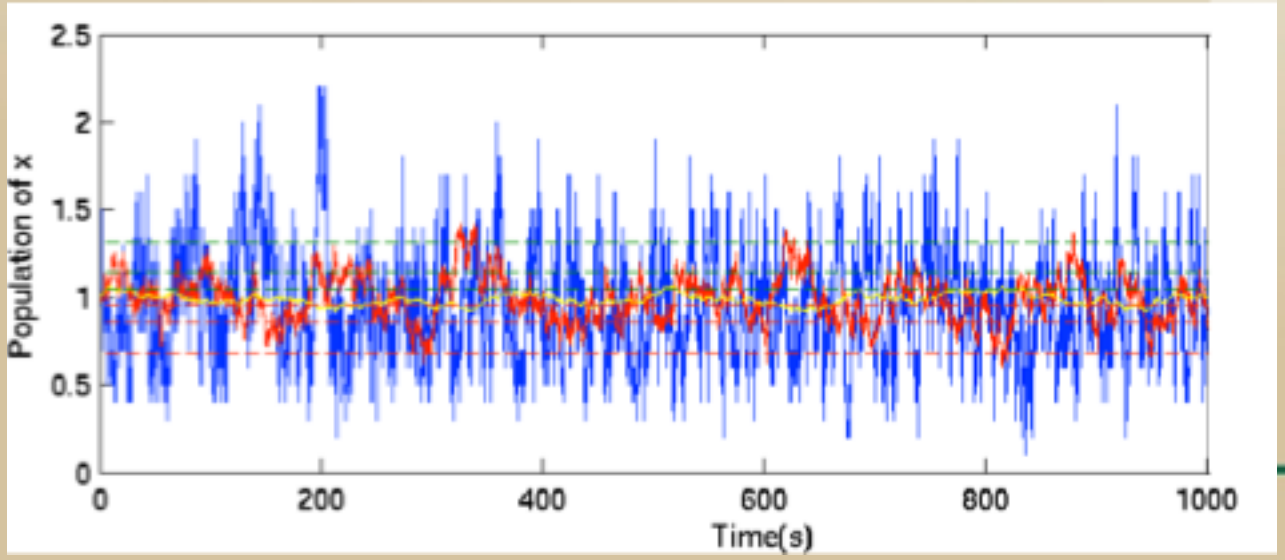
mean = variance = a

The coefficient of variation $C_v = \sigma/\mu$ is

$$C_v = \frac{1}{\sqrt{a}} = \frac{1}{\sqrt{\mu}}$$

Relative noise decreases as system size increases.



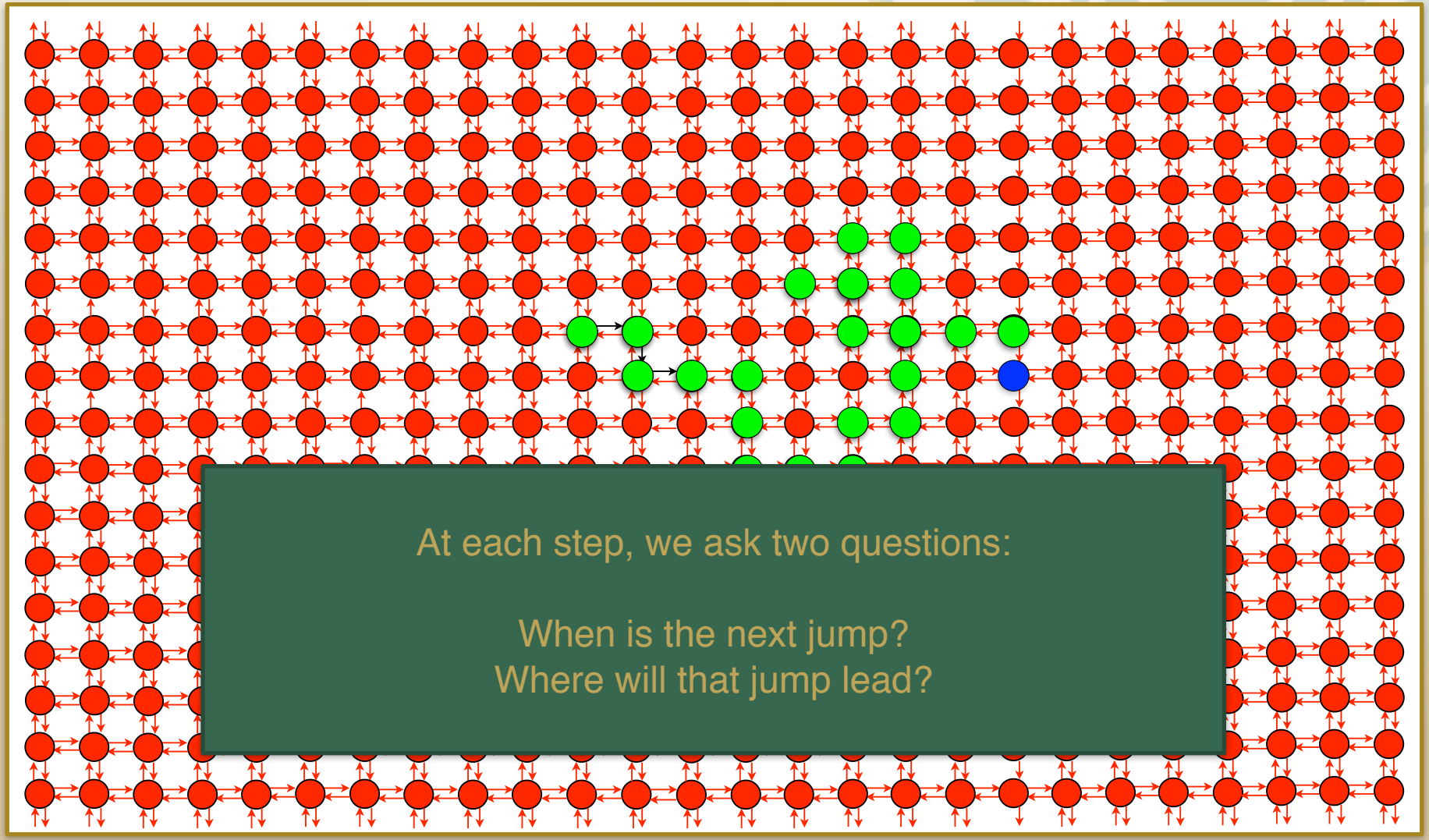


The (Chemical) Master Equation

- The CME Description
- Example: Transcription as a Birth-Death Process.
- Kinetic Monte Carlo Approaches
- Finite State Projection Approaches
- Moment Computations

See notes online

Kinetic Monte Carlo Algorithms



Kinetic Monte-Carlo Simulation Methods

- Stochastic Simulation Algorithm
  - D.T. Gillespie, J. Phys. Chem. A 81, 2340 (1977)
  - M. Gibson and J. Bruck, J. Phys. Chem. 104, 1876 (2000)
- τ leaping
  - D. Gillespie, J. Chem. Phys. 115, 1716 (2001); 119, 8229 (2003)
  - M. Rathinam et al., J. Chem. Phys. 119, 12784 (2003)
  - T. Tian and K. Burrage, J. Chem. Phys. 121, 10356 (2004)
  - A. Chatterjee, et al. J. Chem. Phys. 122, 054104 (2005)
  - Y. Cao, D. Gillespie and L. Petzold, J. Chem. Phys. 123, 054104 (2005)
- Chemical Langevin Equations
  - D. Gillespie, J. Chem. Phys. 113, 1716 (2000)
- System Partitioning Methods
  - C. Rao and A. Arkin, J. Chem. Phys. 118, 4999 (2003)
  - Y. Cao et al., J. Chem. Phys. 122, 014116 (2005)
- Hybrid Methods
  - E. Haseltine and J. Rawlings, J. Chem. Phys. 117, 6959 (2002)
  - H. Salis and Y. Kaznessis, J. Chem. Phys. 122, 054103 (2005)

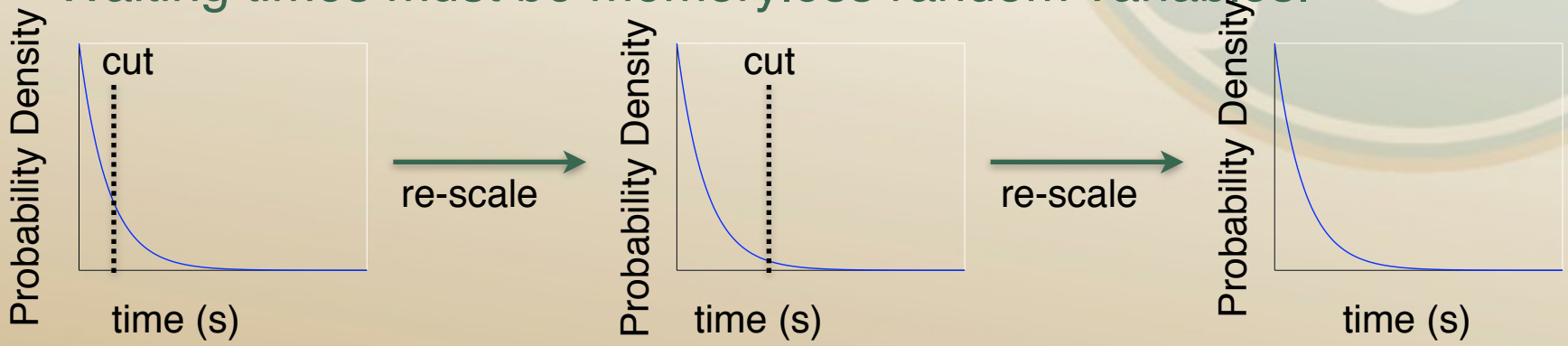
Kinetic Monte Carlo Methods

- Exponential waiting times between events
- Stochastic Simulation Algorithm
- Tau leaping
- Chemical Langevin (Stochastic Differential Equation)
- System Partitioning Methods
- Relationship between stochastic and deterministic trajectories

Online notes to be covered in this evening's lab

When is the next jump?

- We have assumed that the system is fully described by the population vectors.
- If no reaction occurs, then nothing will have changed.
- Waiting times must be memoryless random variables.



- Wherever we cut and scale the distribution, it looks the same.

$$\frac{f(t-\tau)}{\int_\tau^\infty f(\hat{t})d\hat{t}} = \frac{we^{-w(t-\tau)}}{e^{-w\tau}} = we^{-w(t)} = f(t)$$

The exponential is the only continuous r.v. with this property.

When is the next jump?

- To generate an exponentially distributed random number, all we need is a uniform random number generator.

- Find the cumulative distribution,

$$F(t) = 1 - \exp(-\lambda t)$$

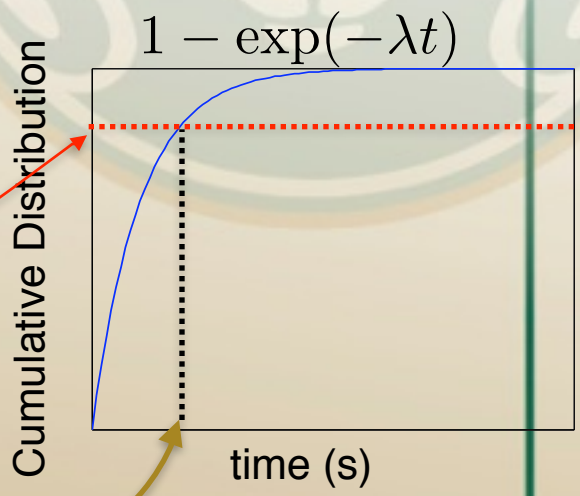
- Generate uniform random number,

$$r \in \mathrm{U}[0, 1]$$

- Find intersection where $F(t) = r$

$$\tau = \frac{1}{\lambda} \log \frac{1}{1 - r}$$

- This is the time of the next reaction.

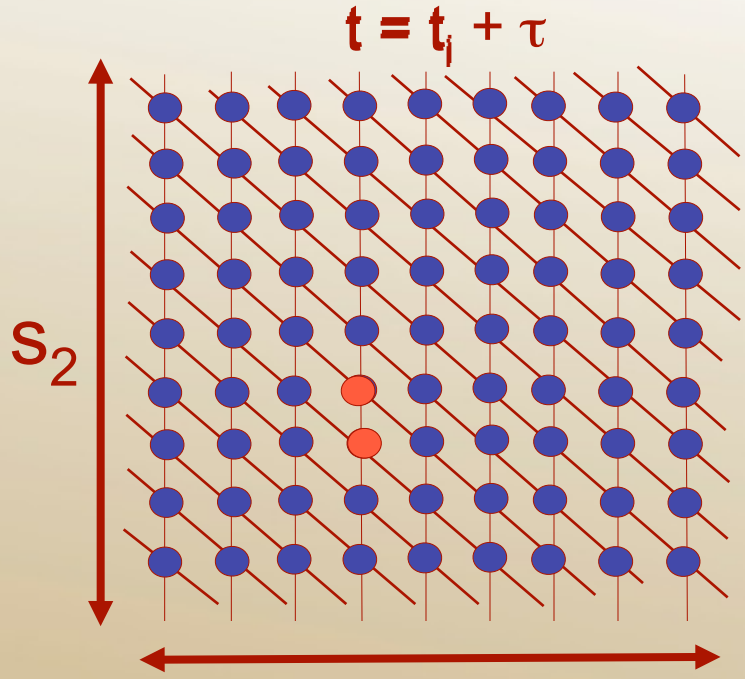


Kinetic Monte Carlo Methods

- Exponential waiting times between events
- Stochastic Simulation Algorithm
- Tau leaping
- Chemical Langevin (Stochastic Differential Equation)
- System Partitioning Methods
- Relationship between stochastic and deterministic trajectories

Online notes to be covered in this evening's lab

Stochastic Simulation Algorithm



Step 1. Generate the time of the next reaction.

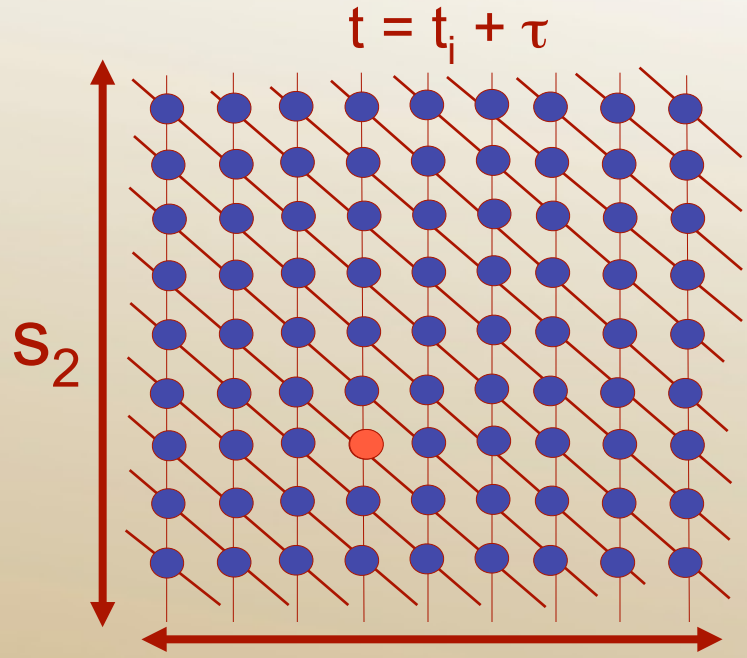
Step 2. Decide which reaction has occurred.

Step 3. Update current Time ($t=t+\tau$) and State ($\mathbf{x} = \mathbf{x}+s_k$).

Possible SSA methods:

- First Reaction Method (Gillespie ‘77)
- Next Reaction Method (Gibson and Bruck ‘00)
- Direct Method (Gillespie ‘77)

The First Reaction Method (FRM)



Step 1. Generate the time of the next reaction of each type.

The time until the next reaction is a random variable of exponential distribution:

$$P_{\tau_\mu}(t) = w_\mu(\mathbf{x})\mathrm{e}^{-w_\mu(\mathbf{x})t}$$

To generate each next reaction time, generate r_1 from a uniform distribution on (0,1) and use the equation:

$$\tau_\mu = \frac{1}{w_\mu(\mathbf{x})}\log\frac{1}{r_\mu}$$

Step 2. Decide which reaction has occurred.

This is simply the reaction with the smallest τ_μ:

$$k = \arg\left\{\min_{\mu\in\{0,\ldots,M\}} \tau_\mu\right\}$$

Step 3. Update current Time (t=t+τ_k) and State (**x** = **x**+$\mathbf{s}_k$).

In the FRM each reaction requires M rv's.

The First Reaction Method SSA in Matlab.


```
clear all
t=0;tstop = 2000;                                   %% Specify initial and final times
x = [0; 0];                                         %% Specify initial conditions
S = [1 -1 0  0; 0  0 1 -1];                         %% Specify stoichiometry
while t<tstop
    w = [10; 1*x(1); 10*x(1); 1*x(2)];             %% Specify Propensity functions
    tpos = 1./w.*log(1./rand(4,1));                % possible times until first reaction
    [tpos,i]=min(tpos);                            % find which is first reaction
    t=t+tpos;
    if t<=t_stop
        x = x+S(:,i);                               % update the configuration
    end
end
```

Possible SSA methods:

- First Reaction Method (Gillespie ‘77)
- Next Reaction Method (Gibson and Bruck ‘00)
- Direct Method (Gillespie ‘77)

The Next Reaction Method (NRM)

- In the FRM, we generate times, $\{\tau_\mu\}$, for all *M* reactions and choose the reaction, *k*, with the smallest time, τ_k.
- Only a few species will change population as a result of this reaction--the rest will remain constant.
- For most reactions, the propensity functions will remain constant.
  - For these, the times can be reused in the subsequent step to find the next reaction: $\{\tau_\mu\} \rightarrow \{\tau_\mu - \tau_k\}$
- When there are many different species and reactions, this NRM approach can be done with far fewer random number than the FRM.
- Particularly useful for compartmental or Reaction-Diffusion processes.

Possible SSA methods:

- First Reaction Method (Gillespie ‘77)
- Next Reaction Method (Gibson and Bruck ‘00)
- Direct Method (Gillespie ‘77)

Minimum of two Exponential Random Variables

Let $\{\tau_1, \tau_2, \ldots, \tau_M\}$ be a set of exponentially distributed random variables: $\tau_\mu \in \text{EXP}\,(w_\mu)$

The minimum of $\{\tau_\mu\}$ is an exponentially distributed random variable given by:

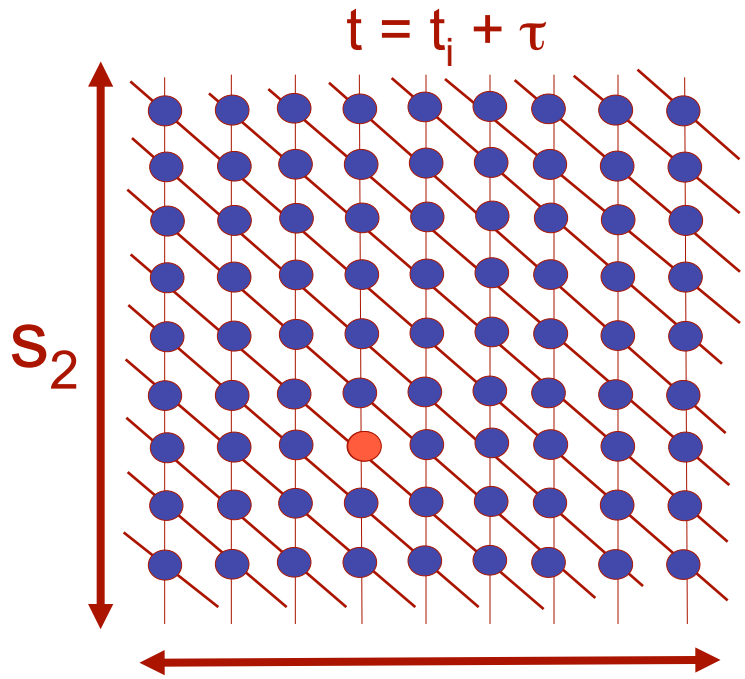
$$\min_{\mu \in \{0,\ldots,M\}} \tau_\mu \in \text{EXP}\,(|\mathbf{w}|_1)$$

The argument, *k*, of this distribution is also a random variable with distribution:

$$P(k = \mu) = \frac{w_\mu}{|\mathbf{w}|_1}$$

In the DM, we only need to generate 2 rv's.

The Direct Method (DM)



Step 1. Generate the time of the next reaction.

The time until the next reaction is a random variable of exponential distribution:

$$P_\tau(t) = |\mathbf{w}(\mathbf{x})|_1 e^{-|\mathbf{w}(\mathbf{x})|_1 t}$$

To generate the next reaction time, generate r_1 from a uniform distribution on (0,1) and use the equation:

$$\tau = \frac{1}{|\mathbf{w}|_1} \log \frac{1}{r_1}$$

Step 2. Decide which reaction has occurred.

To obtain a realization of which reaction will occur, generate a second uniform random number, r_2, and find the smallest *k* such that:

$$\sum_{\mu=1}^{k-1} w_\mu(\mathbf{x}) \le r_2 |\mathbf{w}|_1 \le \sum_{\mu=1}^{k} w_\mu(\mathbf{x})$$

Step 3. Update current Time ($t=t+\tau$) and State ($\mathbf{x} = \mathbf{x}+s_k$).

The Direct Method (SSA) in Matlab.


```
clear all
t=0;tstop = 2000;                                   %% Specify initial and final times
x = [0; 0];                                         %% Specify initial conditions
S = [1 -1 0  0; 0  0 1 -1];                         %% Specify stoichiometry
while t<tstop
    w = [10; 1*x(1); 10*x(1); 1*x(2)];              %% Specify Propensity functions
    w0 = sum(w);                                    %% Compute the sum of the prop. functions
     t = t+1/w0*log(1/rand);                        %% Update time of next reaction
    if t<=t_stop
       r2w0=rand*w0;                %% generate second random number and multiply by prop. sum
       i=1;                         %% initialize reaction counter
       while sum(w(1:i))<r2w0              % increment counter until sum(w(1:i)) exceeds r2w0
          i=i+1;
       end
       x = x+S(:,i);                                % update the configuration
   end
end
```

Kinetic Monte Carlo Methods

- Exponential waiting times between events
- Stochastic Simulation Algorithm
- Tau leaping
- Chemical Langevin (Stochastic Differential Equation)
- System Partitioning Methods
- Relationship between stochastic and deterministic trajectories

Online notes to be covered in this evening's lab

- τ-leaping

  - D. Gillespie, J. Chem. Phys. **115**, 1716 (2001)
  - D. Gillespie, L. Petzold, J. Chem. Phys. **119**, 8229 (2003)
  - M. Rathinam *et al.*, J. Chem. Phys. **119**, 12784 (2003)
  - T. Tian and K. Burrage, J. Chem. Phys. **121**, 10356 (2004)
  - Y. Cao, D. Gillespie and L. Petzold, J. Chem. Phys. **123**, 054104 (2005)

τ Leaping

Step 0. Specify length of each time step, τ.

Assume that all propensity functions are constant over the time interval $(t,t+\tau)$.

The number of times each reaction will fire is a Poisson* random number with mean $w_\mu \tau$:

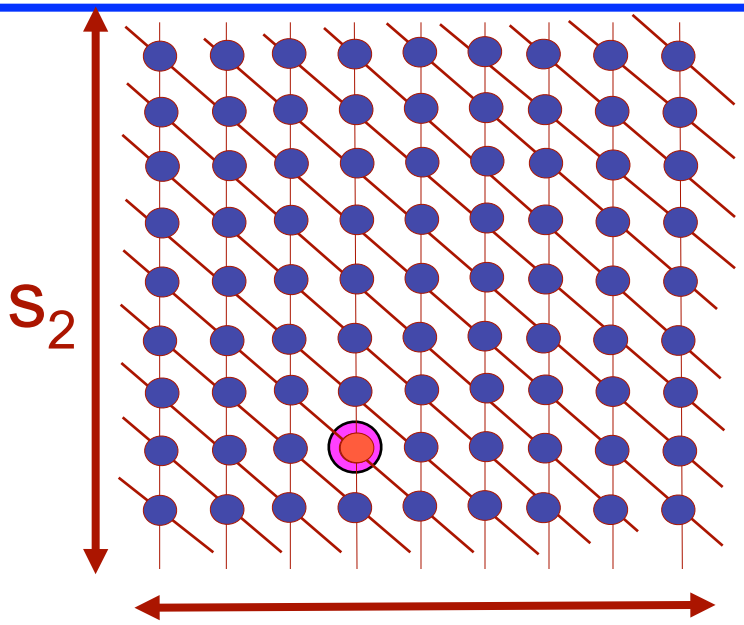
$$P_{k_\mu}(n) = \frac{[w_\mu(\mathbf{x})\tau]^n}{n!} e^{w_\mu(\mathbf{x})\tau}$$

Step 1. For each μ, generate k_μ.

Step 2. Update the time: $t = t + \tau$

Update the state: $\mathbf{x} = \mathbf{x} + \sum_{\mu=1}^{M} k_\mu \mathbf{s}_\mu$

τ Leaping



t = t_i +τ Update Time

$$k_1 = 4; \mathbf{s}_1 = [0, 1]^T$$

$$k_2 = 2; \mathbf{s}_1 = [-1, 1]^T$$

$$k_3 = 3; \mathbf{s}_1 = [0, -1]^T$$

$$k_4 = 4; \mathbf{s}_1 = [1, -1]^T$$

The number of times each reaction will fire is a Poisson random number with mean $w_\mu \tau$: $P_{k_\mu}(n) = \frac{[w_\mu(\mathbf{x})\tau]^n}{n!} e^{w_\mu(\mathbf{x})\tau}$

Step 1. For each μ, generate k_μ.

Step 2. Update the state: $\mathbf{x} = \mathbf{x} + \sum_{\mu=1}^{M} k_\mu \mathbf{s}_\mu$

Update the time: $t = t + \tau$

Limitations of τ leaping

- For many situations τ leaping significantly speeds up the Monte Carlo simulation, but:
  - Poisson r.v.'s are unbounded
    - Propensity functions may change dramatically over small time intervals.
    - May result in negative populations.

Note that these concerns are most important when the population of some species are very small.

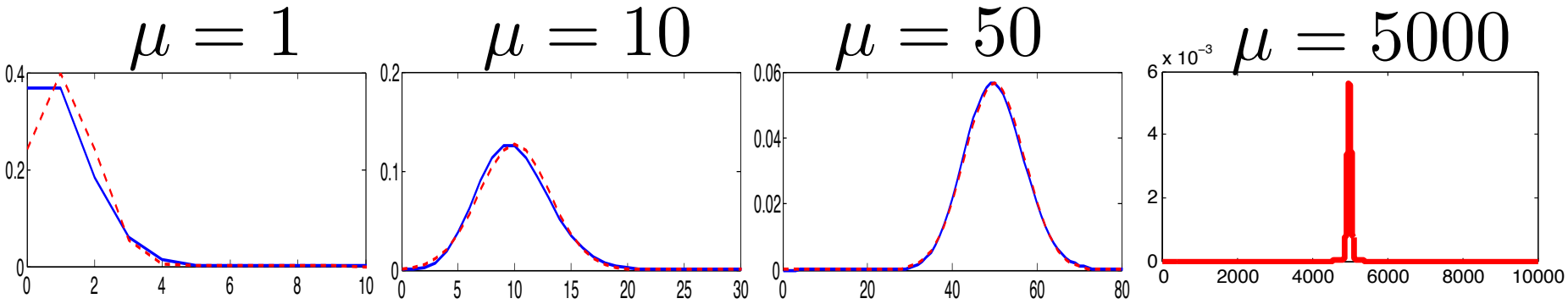
Precisely the circumstance where stochastic models are most important!

Kinetic Monte Carlo Methods

- Exponential waiting times between events
- Stochastic Simulation Algorithm
- Tau leaping
- Chemical Langevin (Stochastic Differential Equation)
- System Partitioning Methods
- Relationship between stochastic and deterministic trajectories

Online notes to be covered in this evening's lab

Comparison of step updates for SSA, tau-leap, Langevin and ODE's



- In SSA, every step has exactly one update.
- tau leaping has a Poisson number of updates per step.
- For large numbers of reactions, replace the Poisson distribution with a normal distribution (same mean and variance), which are cheaper to generate -- this is the chemical Langevin equation.
- For very large numbers of reactions, the update number approaches a Delta distribution -- this is an ODE!

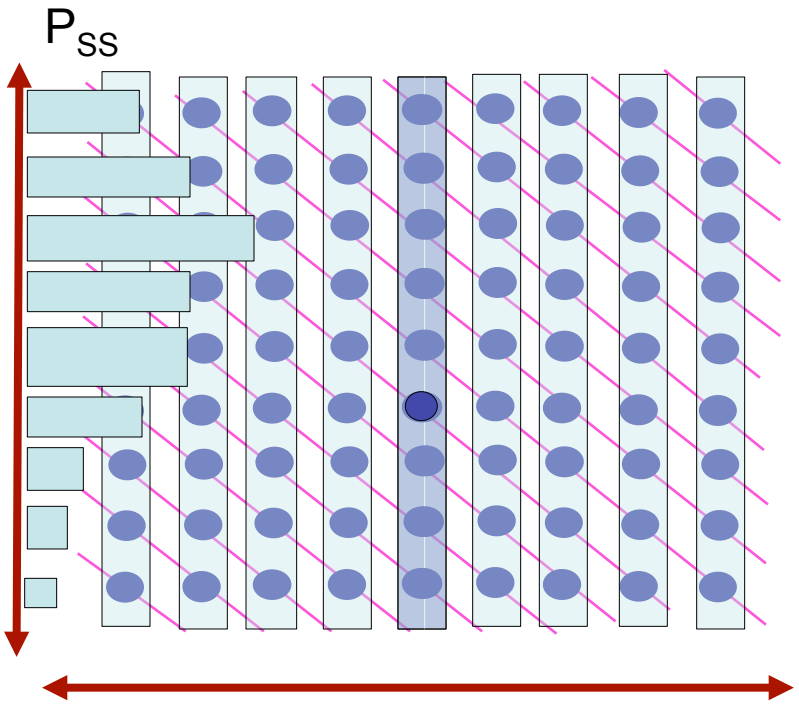
Kinetic Monte Carlo Methods

- Exponential waiting times between events
- Stochastic Simulation Algorithm
- Tau leaping
- Chemical Langevin (Stochastic Differential Equation)
- System Partitioning Methods
- Relationship between stochastic and deterministic trajectories

Online notes to be covered in this evening's lab

- System Partitioning Methods
  - Fast--Slow Partitions
    - C. Rao and A. Arkin, J. Chem. Phys. **118**, 4999 (2003)
    - Y. Cao *et al.*, J. Chem. Phys. **122**, 014116 (2005)
  - Continuous--Discrete Partitions
    - E. Haseltine and J. Rawlings, J. Chem. Phys. **117**, 6959 (2002)
    - H. Salis and Y. Kaznessis, J. Chem. Phys. **122**, 054103 (2005)

Fast--Slow partitions.

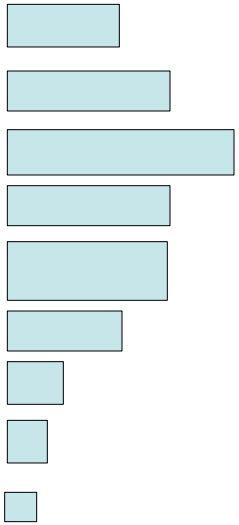


Separate into “fast” and “slow” partitions.

Assume that the “fast” partitions reach probabilistic equilibrium before a slow reaction occurs.

Fast--Slow partitions.

P_{SS}



Slow Reaction Propensities

$$\begin{bmatrix} w_\mu(\mathbf{x}_1) \\ w_\mu(\mathbf{x}_2) \\ w_\mu(\mathbf{x}_3) \\ \vdots \end{bmatrix}$$

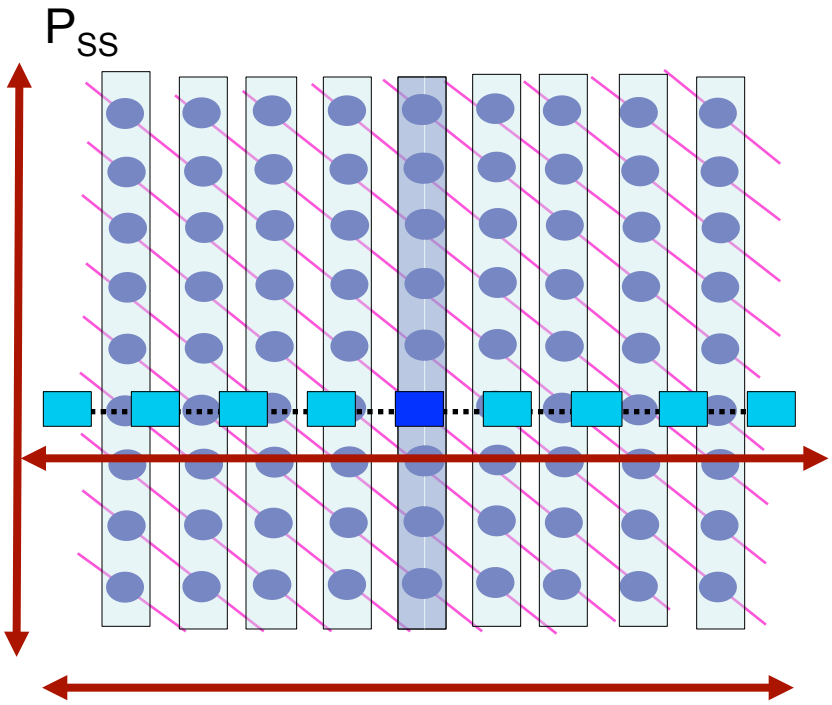
Average Slow Reaction Propensities

$$= \bar{w}_\mu, \text{ for } \mu = \{1, 2, \ldots, M\}$$

Use the fast sets' steady state probability distributions to scale the propensity functions of the slow reactions.

Results in a vector of average propensity functions, $\bar{\mathbf{w}}$, for the slow reactions.

Fast--Slow partitions.



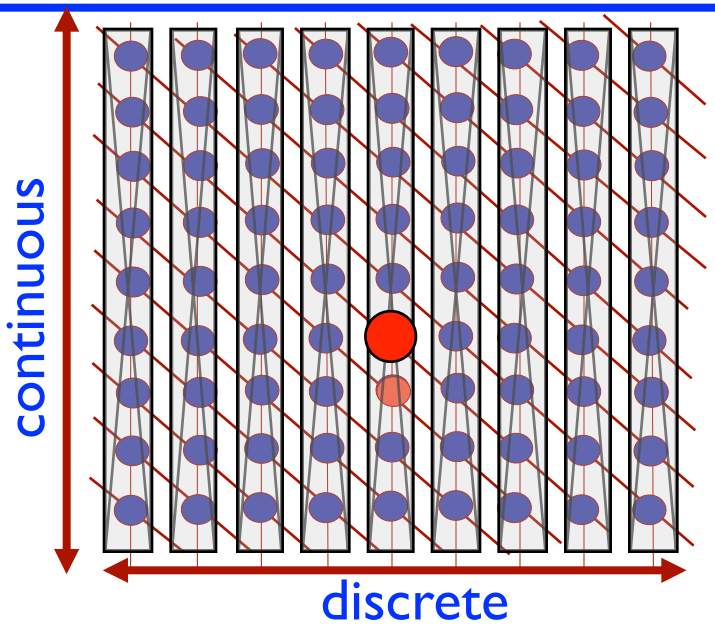
The projection to the slow manifold results in a new lower dimensional Markov chain.

This is simulated with SSA.

Continuous--Discrete partitions.

- In some systems, there are great differences in scale:
  - Large populations (continuous)
  - Small populations (discrete)
- All discrete models take too long.
- All continuous models are inaccurate.
- Hybrid models are necessary.

Separate into "continuous" and "discrete" partitions.



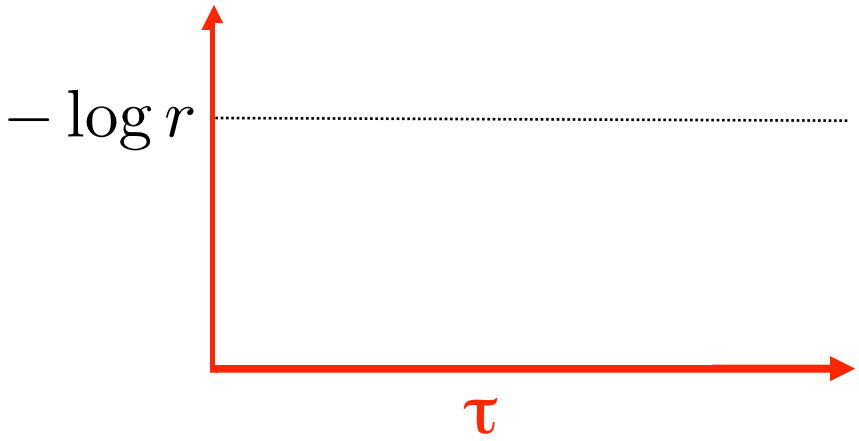
Simulate the continuous part with ordinary or stochastic differential equations.

Choose uniform rv, *r*.

Numerically integrate propensity functions until:

$$\int_{t_0}^{t_0+\tau} \sum_{\mu=1}^{M} w_\mu(\mathbf{x}(t))dt = -\log r$$

Choose next discrete reaction.



Kinetic Monte Carlo Methods

- Exponential waiting times between events
- Stochastic Simulation Algorithm
- Tau leaping
- Chemical Langevin (Stochastic Differential Equation)
- System Partitioning Methods
- Relationship between stochastic and deterministic trajectories

Online notes to be covered in this evening's lab

Relationship of Stochastic (X) and Deterministic (Φ) Descriptions

Given N species $X_1, \ldots, X_N$ and M elementary reactions. Let $\Phi_i := [X_i]$.

A deterministic description can be obtained from mass-action kinetics:

$$\frac{d\Phi}{dt} = Sf(\Phi)$$

where $f(\cdot)$ is at most a second order monomial. It depends on the type of reactions and their rates.

Example:

$$A + B \xrightarrow{k_1} C$$
$$A \xrightarrow{k_2} B$$

$$\begin{aligned}
\frac{d\Phi_A}{dt} &= -k_1\Phi_A\Phi_B - k_2\Phi_A \\
\frac{d\Phi_B}{dt} &= -k_1\Phi_A\Phi_B + k_2\Phi_A \\
\frac{d\Phi_C}{dt} &= k_1\Phi_A\Phi_B
\end{aligned}$$

or

$\frac{d\Phi}{dt} = Sf(\Phi)$ where

$$S = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & 0 \end{bmatrix}, \quad f(\Phi) = \begin{bmatrix} k_1\Phi_A\Phi_B \\ k_2\Phi_A \end{bmatrix}$$

Relationship of Stochastic (X) and Deterministic (Φ) Descriptions

Define $X^{\Omega}(t) = \frac{X(t)}{\Omega}$.

Question: How does $X^{\Omega}(t)$ relate to $\Phi(t)$?

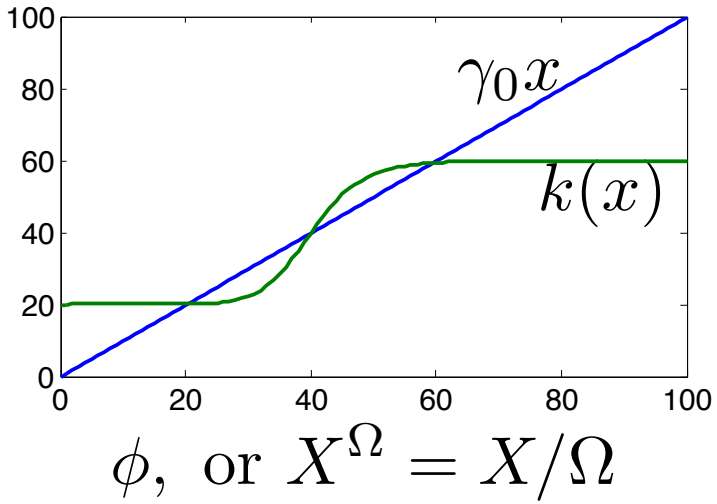
Fact: Let $\Phi(t)$ be the deterministic solution to the reaction rate equations

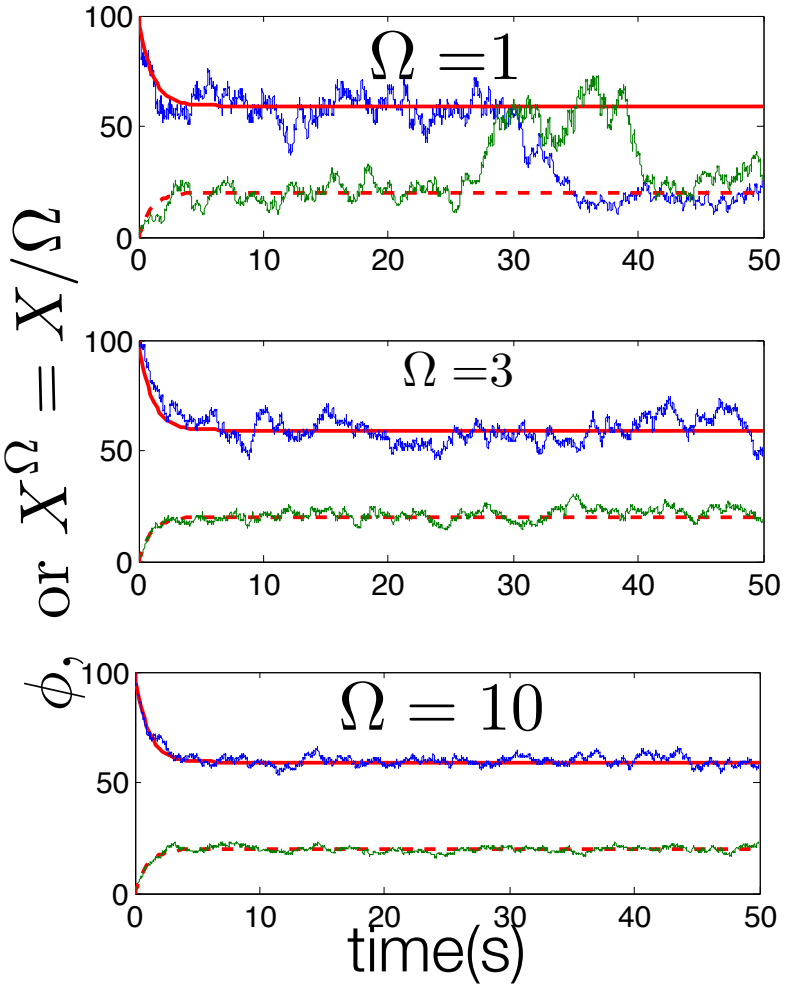
$$\frac{d\Phi}{dt} = Sf(\Phi), \ \Phi(0) = \Phi_0.$$

Let $X^{\Omega}(t)$ be the stochastic representation of the same chemical systems with $X^{\Omega}(0) = \Phi_0$. Then for every $t \geq 0$:

$$\lim_{\Omega \to \infty} \sup_{s \leq t} \left| X^{\Omega}(s) - \Phi(s) \right| = 0 \ a.s.$$

x produced with rate $k(x)$
and degraded with rate $\gamma_0 x$.





$$w_1(\phi) = \gamma\phi$$

$$w_2(\phi) = \left(20 + 40\frac{\phi^{10}}{40^{10} + \phi^{10}}\right)$$

Deterministic

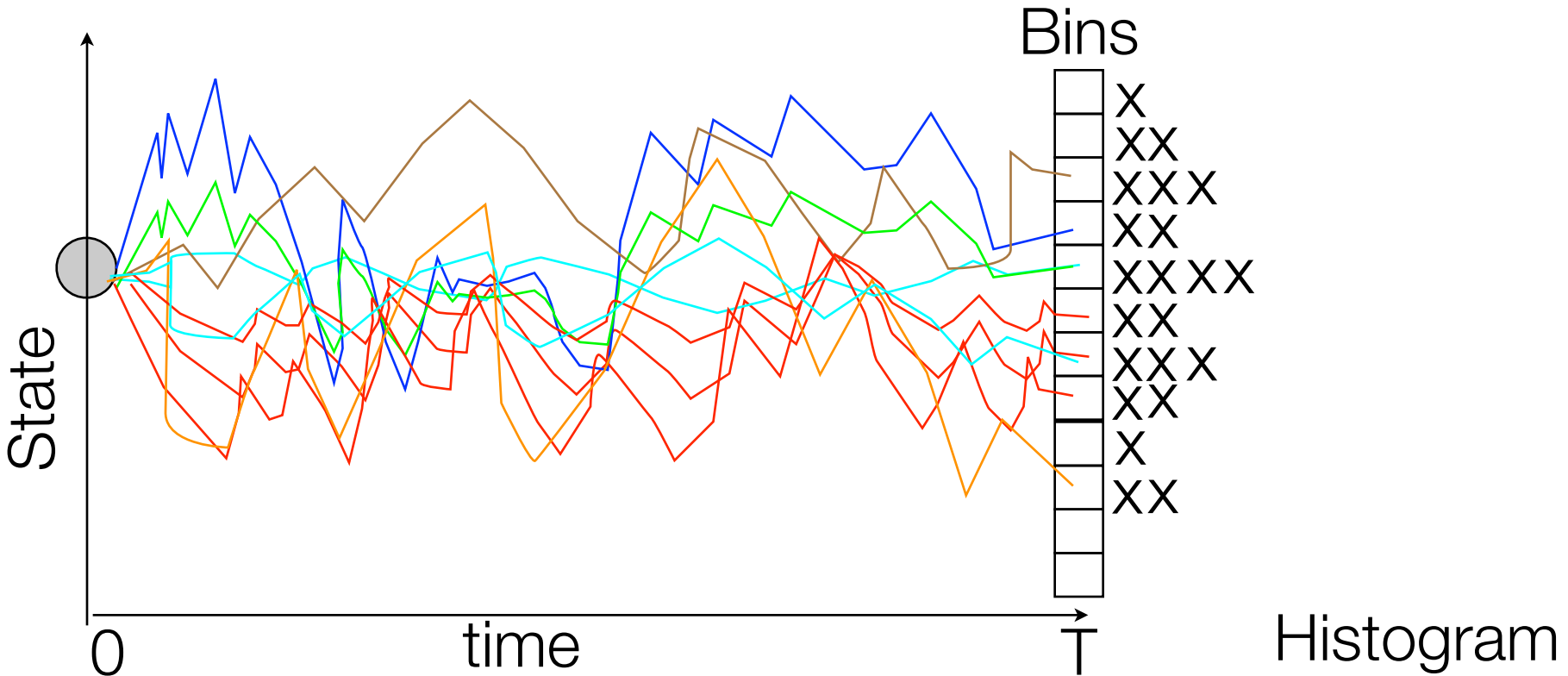
$$w_1(X) = \Omega\gamma_0 X/\Omega = \gamma_0 X$$

$$w_2(X) = \Omega\left(20 + 40\frac{(X/\Omega)^{10}}{40^{10} + (X/\Omega)^{10}}\right)$$

Stochastic

Using Simulations to Find Distributions

- The SSA does an excellent job of producing possible trajectories.
- Sometimes one might want to compute probability distributions at certain times.
- This is done in the SSA by *binning* results of several trajectories.



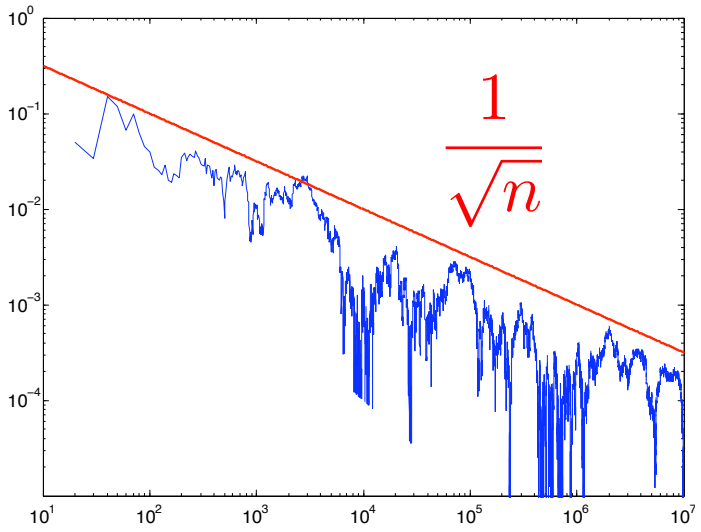
Convergence of KMC Methods

- To get more accurate distributions, one needs more SSA runs.
- Unfortunately, the convergence rate of any Monte Carlo algorithm is fundamentally limited: $error = \mathcal{O}(n^{-\frac{1}{2}})$
- If very high precision is required, then MC methods will be very inefficient.

Convergence for Coin Toss

error:

$$\left| \frac{Heads}{n} - 0.5 \right|$$



After 10^7 tosses there is still an error of about 3×10^{-4}.

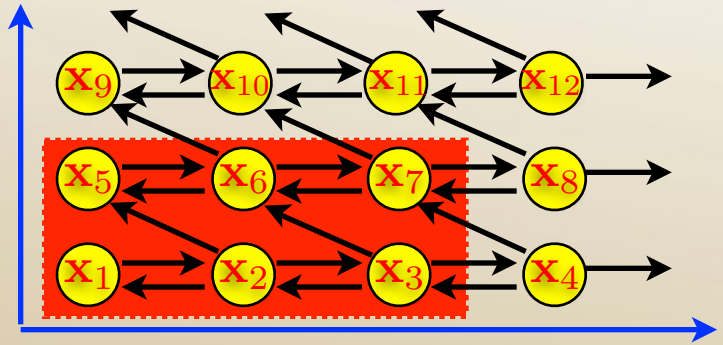
The (Chemical) Master Equation

- The CME Description
- Example: Transcription as a Birth-Death Process.
- Kinetic Monte Carlo Approaches
- Finite State Projection Approaches
- Moment Computations

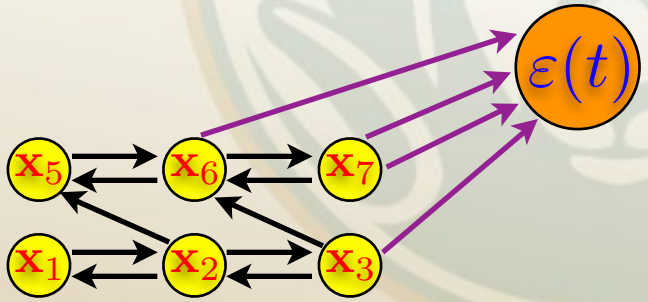
See notes online

The finite state projection approach

The Full System



The Projected System (FSP)



Full Master Equation

$$\begin{bmatrix} \dot{\mathbf{P}}_J \\ \dot{\mathbf{P}}_{J'} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_J & \mathbf{A}_{JJ'} \\ \mathbf{A}_{J'J} & \mathbf{A}_{J'} \end{bmatrix} \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix}$$

Dimension $= \#(J) + \#(J') =$ Infinite

FSP Master Equation

$$\begin{bmatrix} \dot{\mathbf{P}}_J^{FSP} \\ \dot{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_J & \mathbf{0} \\ -\mathbf{1}^T\mathbf{A}_J & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \varepsilon(t) \end{bmatrix}$$

Dimension $= \#(J) + 1 = 7$

The FSP Theorem

(Munsky, JCP '06)

$$\mathbf{P}_J(t) \geq \mathbf{P}_J^{FSP}(t) \text{ and}$$

$$\left\| \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'} \end{bmatrix} - \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \mathbf{0} \end{bmatrix} \right\|_1 = \varepsilon(t)$$

Download software and tutorial available at:
http://www.engr.colostate.edu/~munsky/Software.html

FSP -- a quick and easy proof

split full probability into portions that 'stay' or 'leave' J:

$$\begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix}_{\text{stay}} + \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix}_{\text{leave}}$$

apply FSP definition:

$$\begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_J^{\text{FSP}}(t) \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix}_{\text{leave}}$$

compute approximation error:

$$\left\| \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix} - \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \mathbf{0} \end{bmatrix} \right\|_1 = ||\mathbf{P}_{\text{leave}}(t)||_1 = \varepsilon(t)$$

Download software and tutorial available at:
http://www.engr.colostate.edu/~munsky/Software.html

The FSP Algorithm

Inputs: Initial Conditions, System Parameters, Final time (t_f), Allowable error ($\varepsilon_{\max}$)

Step 1: Choose initial projection space, $\mathbf{X}_{J_0}$.

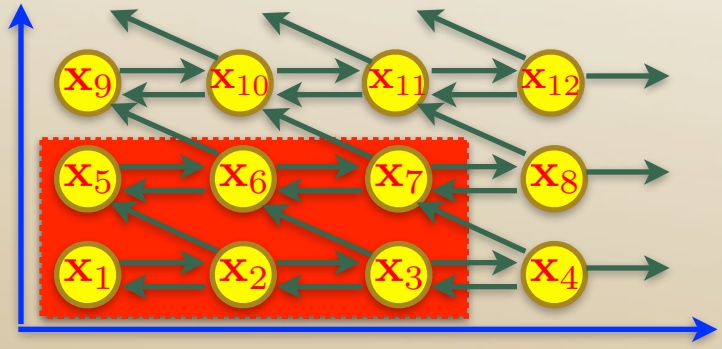
Step 2: Use projection $\mathbf{X}_{J_i}$ to find corresponding error, $\varepsilon_i(t_f)$.

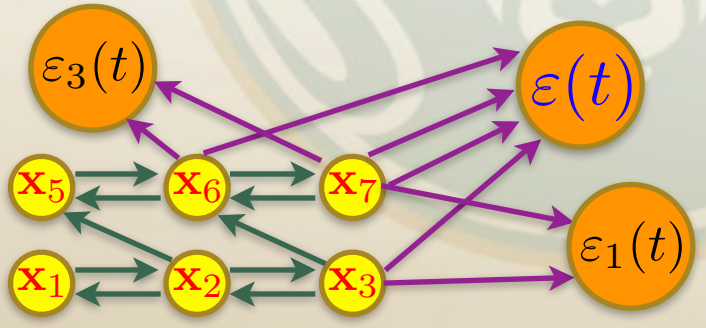
Step 3: If $\varepsilon_i(t_f) \leq \varepsilon_{\max}$, **Stop**.
$\mathcal{P}_{J_i}^{FSP}(t_f)$ approximates $\mathcal{P}(t_f)$ to within $\varepsilon_{\max}$.

Step 4: Expand projection, $\mathbf{X}_{J_{i+1}} \supset \mathbf{X}_{J_i}$,
Increment i and return to Step 2.

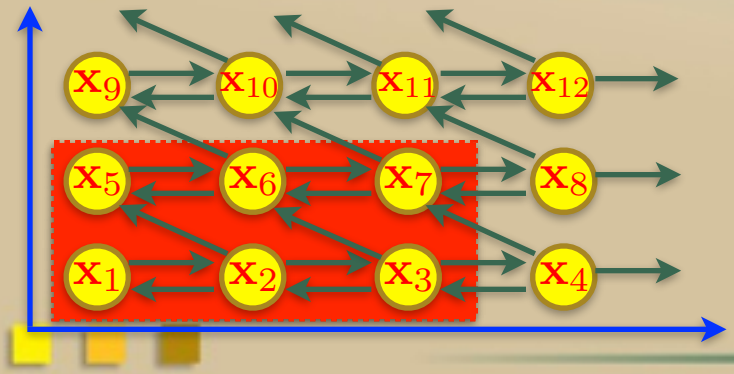
FSP - Expanding the projection space

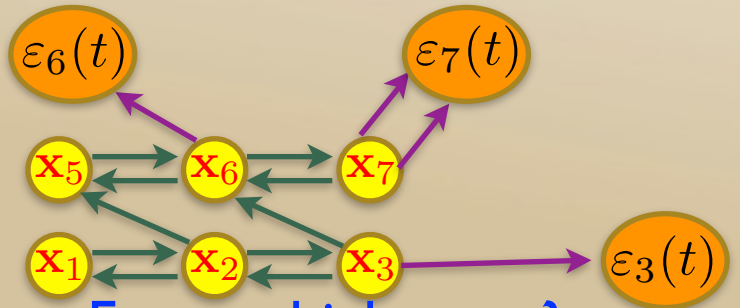
By using multiple sinks, one can determine how the probability measure exits $\mathbf{X}_J$.





Which Reaction Leaves $\mathbf{X}_J$?





From which state?

Finite State Projection Analyses

- Forming the Infinitesimal Generator
- Interpreting and Using FSP Sinks
- Advantages and Limitations of the FSP
- System Reductions to Improve FSP Efficiency
- Examples for Using the FSP

Online notes to be covered in tomorrows lab

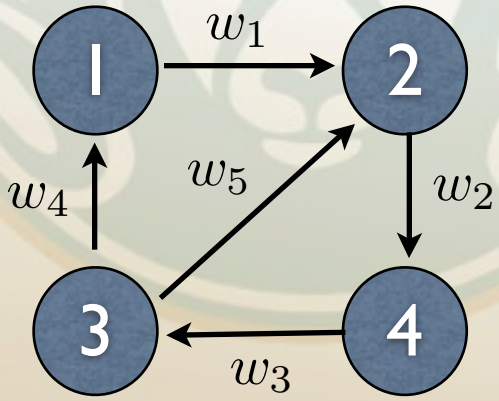
Forming the Infinitesimal Generator

A has one row/column for each state.

Each transition, $\mathbf{x}_i \rightarrow \mathbf{x}_j$, contributes to **A** in two locations:

$-w_\mu(\mathbf{x}_i)$ goes in the *diagonal* element $A_{i,i}$

$+w_\mu(\mathbf{x}_i)$ goes in the *off-diagonal* element $A_{j,i}$



$$\mathbf{A} = \begin{bmatrix} -w_1 & 0 & w_4 & 0 \\ w_1 & -w_2 & w_5 & 0 \\ 0 & 0 & -w_4 - w_5 & w_3 \\ 0 & w_2 & 0 & -w_3 \end{bmatrix}$$

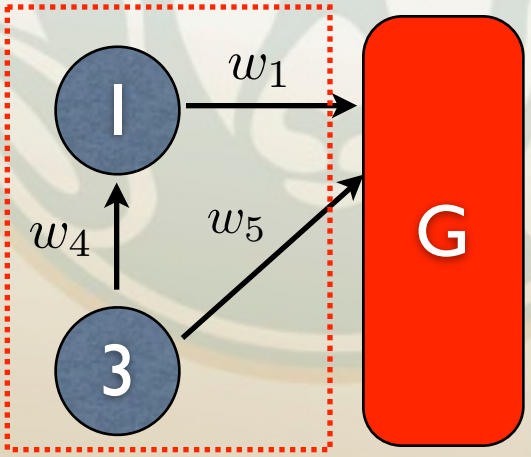
Applying the Finite State Projection

Select the states to keep.

Find the corresponding projection matrix:

$$\mathbf{A}_{[1,3]} = \begin{bmatrix} -w_1 & w_4 \\ 0 & -w_4 - w_5 \end{bmatrix}$$

Collapse remaining states into a single absorbing state



$$\mathbf{A}^{FSP}_{[1,3]} = \begin{bmatrix} -w_1 & w_4 & 0 \\ 0 & -w_4 - w_5 & 0 \\ w_1 & w_5 & 0 \end{bmatrix}$$

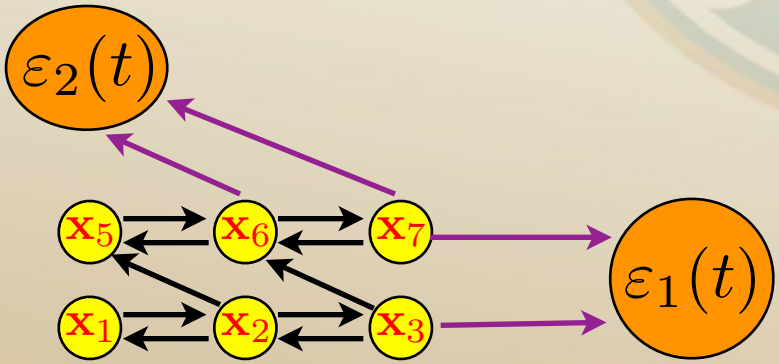
This is the generator for the new Markov chain.

$$\mathbf{A} = \begin{bmatrix} -w_1 & 0 & w_4 & 0 \\ w_1 & -w_2 & w_5 & 0 \\ 0 & 0 & -w_4 - w_5 & w_3 \\ 0 & w_2 & 0 & -w_3 \end{bmatrix}$$

Finite State Projection Analyses

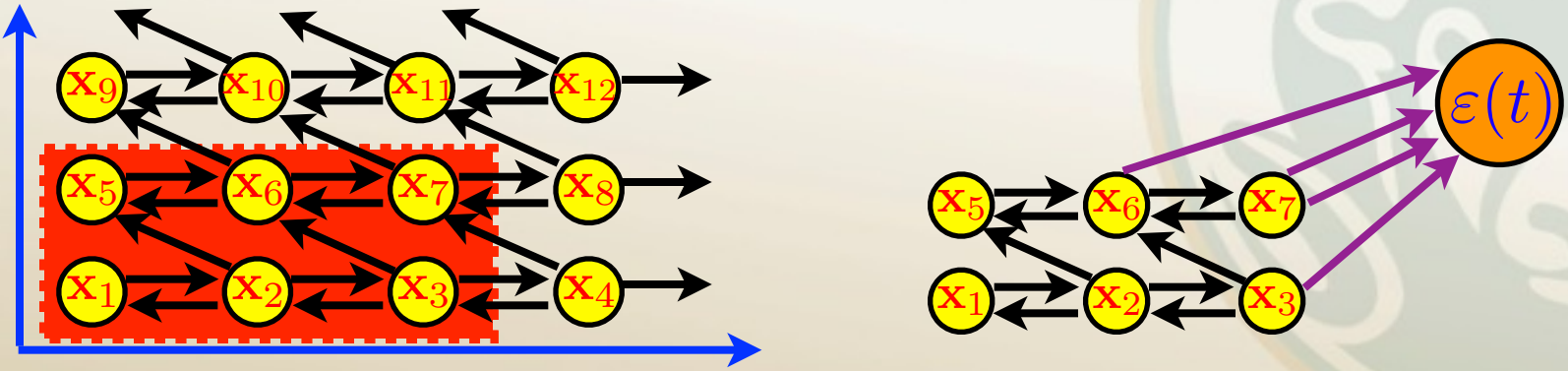
- Forming the Infinitesimal Generator
- Interpreting and Using FSP Sinks
- Advantages and Limitations of the FSP
- System Reductions to Improve FSP Efficiency
- Examples for Using the FSP

A Test...



What do $\varepsilon_1(t)$ and $\varepsilon_2(t)$ mean?

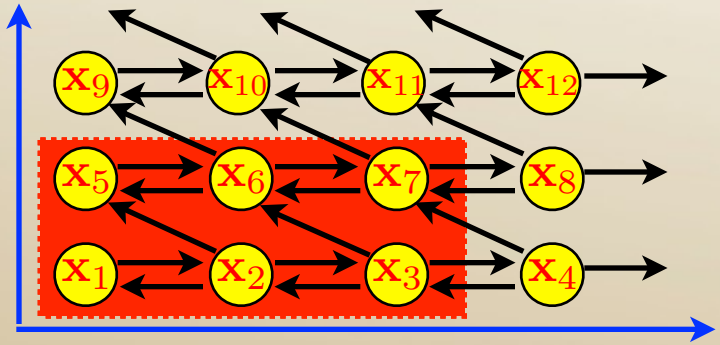
Interpreting the FSP Error Sink

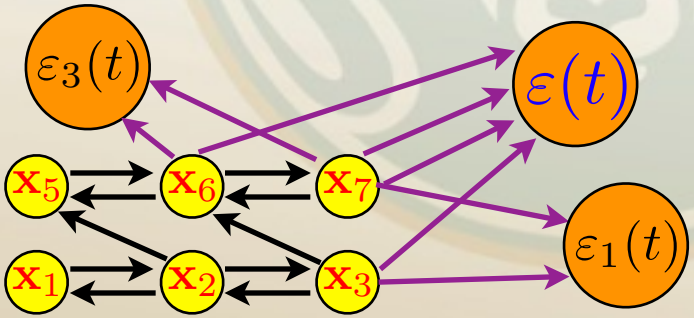


- In the original FSP, $\varepsilon(t)$ is the amount of the probability measure that exits the projection region $\mathbf{X}_J$.
- Median exit time: $t_{50} = t, \text{ s.t. } \varepsilon(t) = 0.5$
- In this form $\varepsilon(t)$ gives information as to when the system exits $\mathbf{X}_J$, but not how.

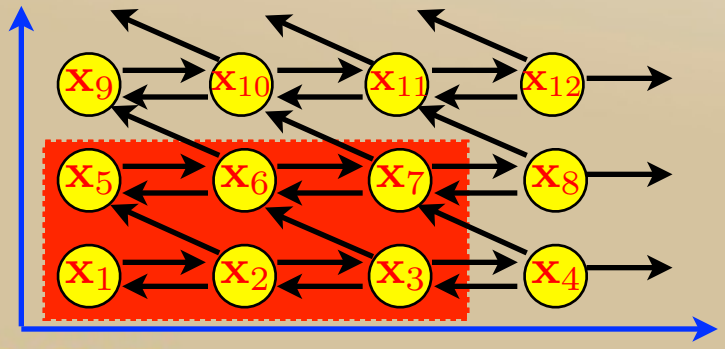
Using Multiple Sink to Track FSP Expansion

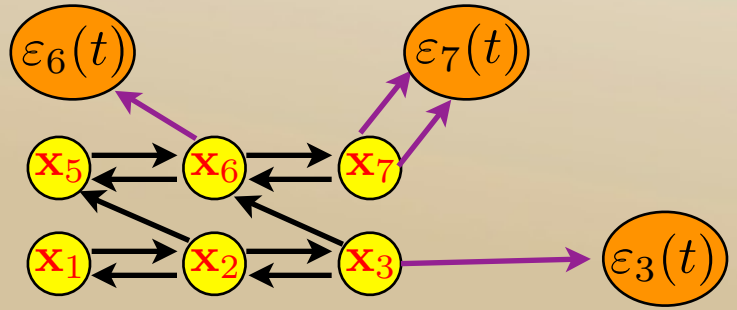
- By using multiple sinks, one can determine how the probability measure exits $\mathbf{X}_J$.





Which Reaction Leaves $\mathbf{X}_J$?

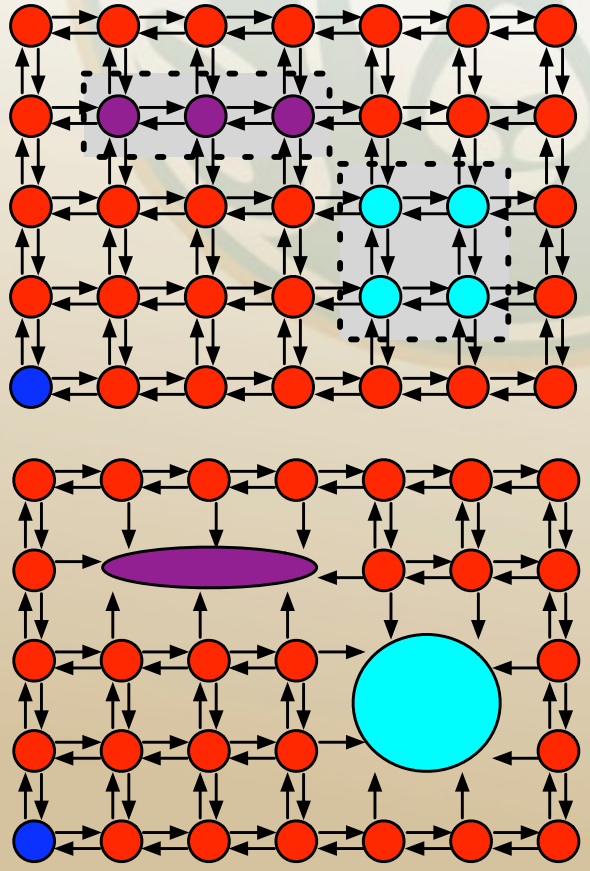




From which state?

Using Multiple Sinks to Analyze Switch Decisions

- Multiple sinks can also be used to get precise analyses of complex switches:
  - Does the cell reach phenotype A before phenotype B?
  - How long until the cell exhibits phenotype A and then B?
  - What is the likelihood of an observed trajectory from A to B and back to A at specific time points?



Finite State Projection Analyses

- Forming the Infinitesimal Generator
- Interpreting and Using FSP Sinks
- Advantages and Limitations of the FSP
- System Reductions to Improve FSP Efficiency
- Examples for Using the FSP

Advantages of the FSP Approach

- Deterministic.
  - Every run of the FSP yields the same result.
  - Enables easier comparisons of different systems (sensitivity analysis and system identification).
- Provides accuracy guarantees.
  - Can be made as precise as required.
  - Allows for analysis of rare events.
- Does not depend upon initial conditions.
- Is open to many subsequent model reductions.

Limitations of the FSP Approach

- Numerical stiffness may lead to computational inefficiency.
- Systems may become very large as distributions cover large regions of the configuration space.
  - Compact distributions may drift over time.
  - Dilute distributions may spread over large regions.
  - Dimension grows exponentially with the number of species.
- For these problems, the original FSP may not suffice,
  - BUT, with additional model reductions and systematic techniques, many of these problems may be alleviated.

Finite State Projection Analyses

- Forming the Infinitesimal Generator
- Interpreting and Using FSP Sinks
- Advantages and Limitations of the FSP
- System Reductions to Improve FSP Efficiency
- Examples for Using the FSP

Reductions to the FSP

- Model Reduction Through Observability
- Time Interval Discretization
- Slow Manifold Projection
- Coarse Meshes for the CME

Using Input & Output Relations for FSP Reduction.

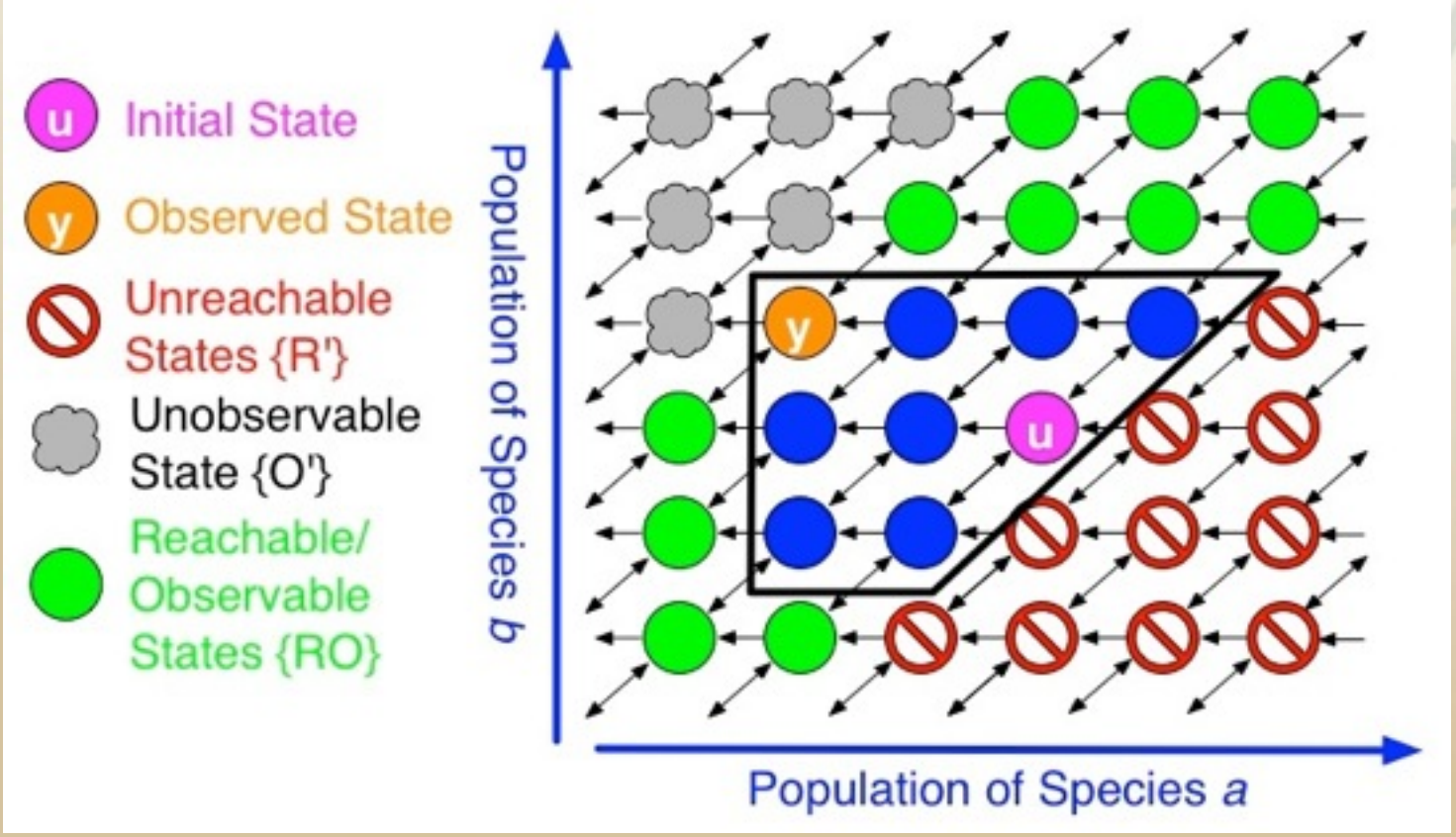
- Often one is not interested in the entire probability distribution.
- Instead one may wish only to estimate:
  - a statistical summary of the distribution (e.g. means, variances, or higher moments)
  - probability of certain traits (switch rate, extinction, specific trajectories, etc…)
- In each of these cases, one can define an output **y**(t):

$$\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t)$$
$$\mathbf{y}(t) = \mathbf{C}\mathbf{P}(t)$$

Munsky/Khammash, CDC, 2006

Using Input & Output Relations for FSP Reduction.

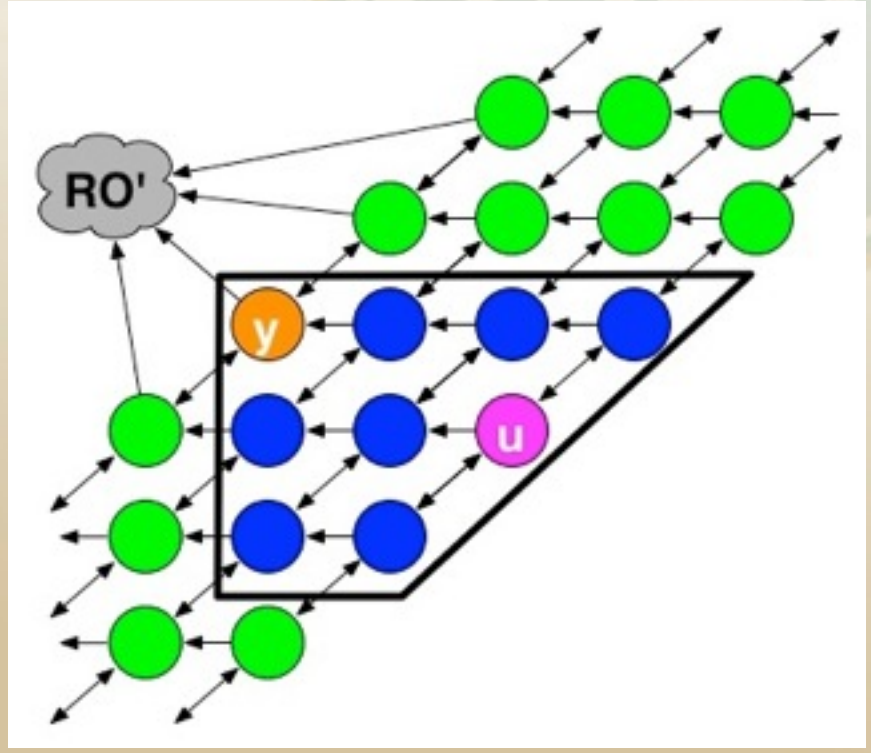
- Begin with a Full Integer Lattice Description of the System States.



Munsky/Khammash, CDC, 2006

Using Input & Output Relations for FSP Reduction.

- Remove Unreachable States and Aggregate the Unobservable States.

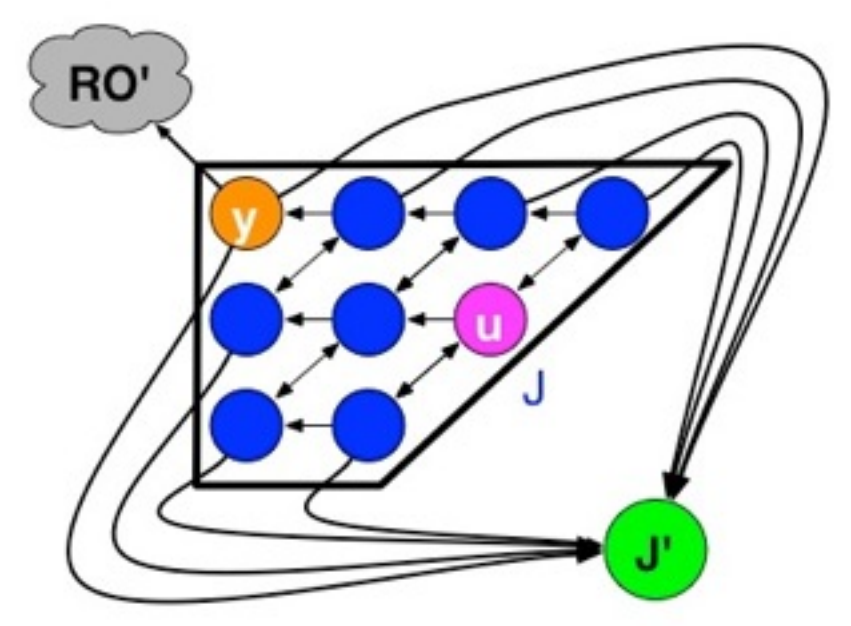


Munsky/Khammash, CDC, 2006

Using Input & Output Relations for FSP Reduction.

We now have a solvable approximation, for which the FSP gives bounds on the approximation's accuracy.

Even stronger reductions can be achieved using balanced truncations.



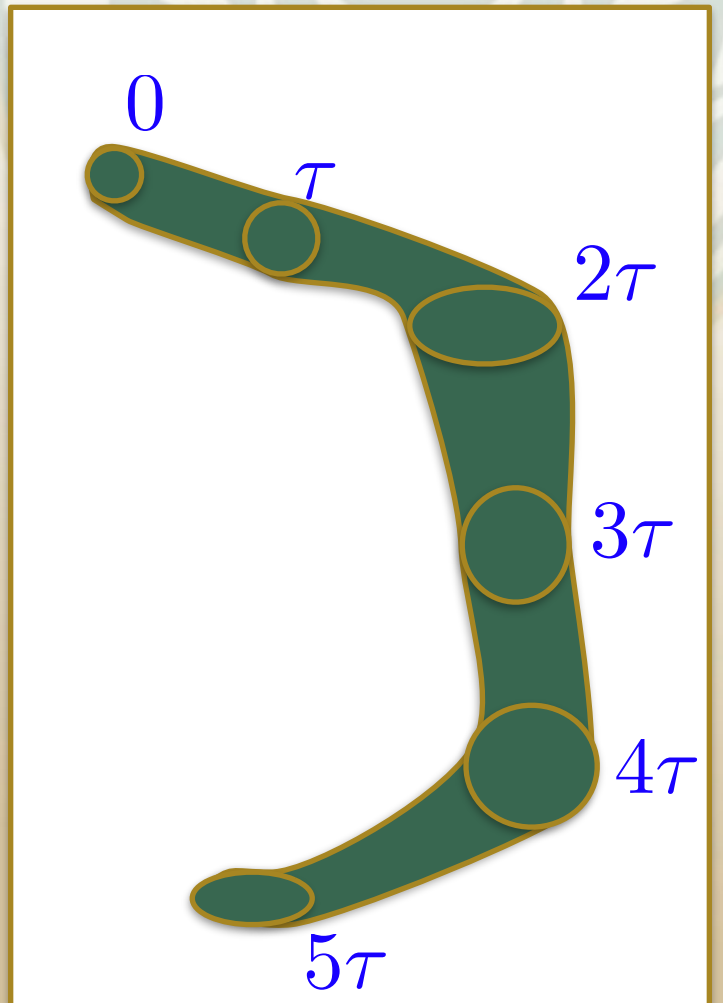
Munsky/Khammash, CDC, 2006

Reductions to the FSP

- Model Reduction Through Observability
- Time Interval Discretization
- Slow Manifold Projection
- Coarse Meshes for the CME

Time Interval Discretization for the FSP

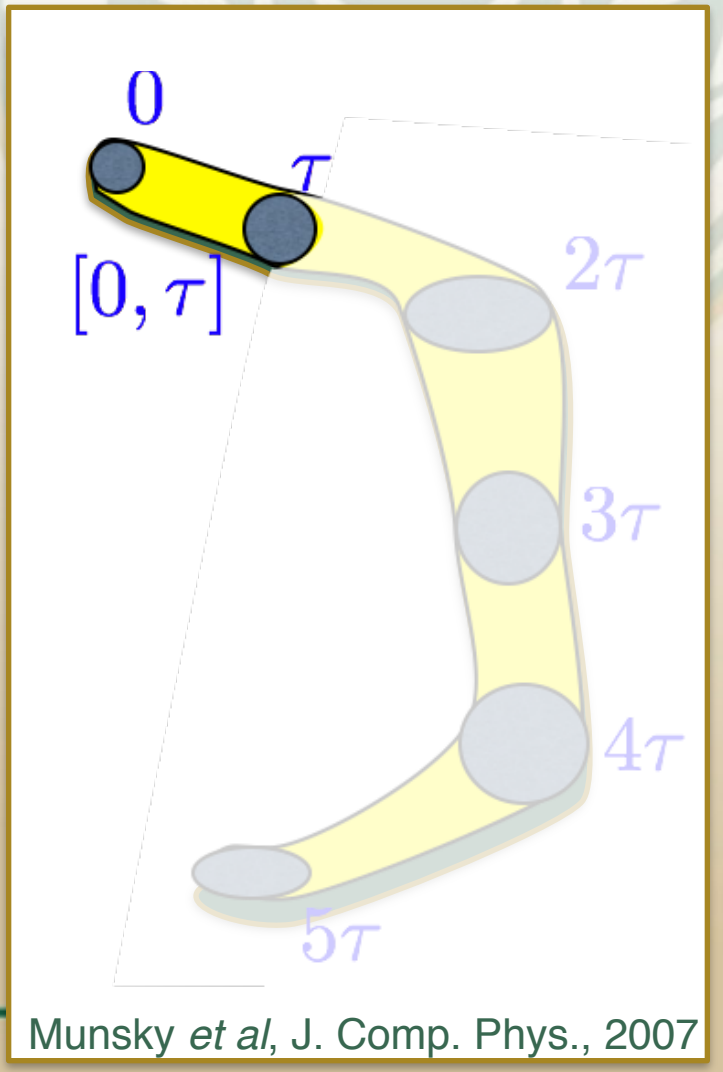
- For many systems, the distribution may drift over time.
- At any one time, the distribution may have a limited support, but...
- The FSP solution must include all intermediate configurations.
- This may lead to an exorbitantly large system of ODEs.



Munsky *et al*, J. Comp. Phys., 2007

Time Interval Discretization for the FSP

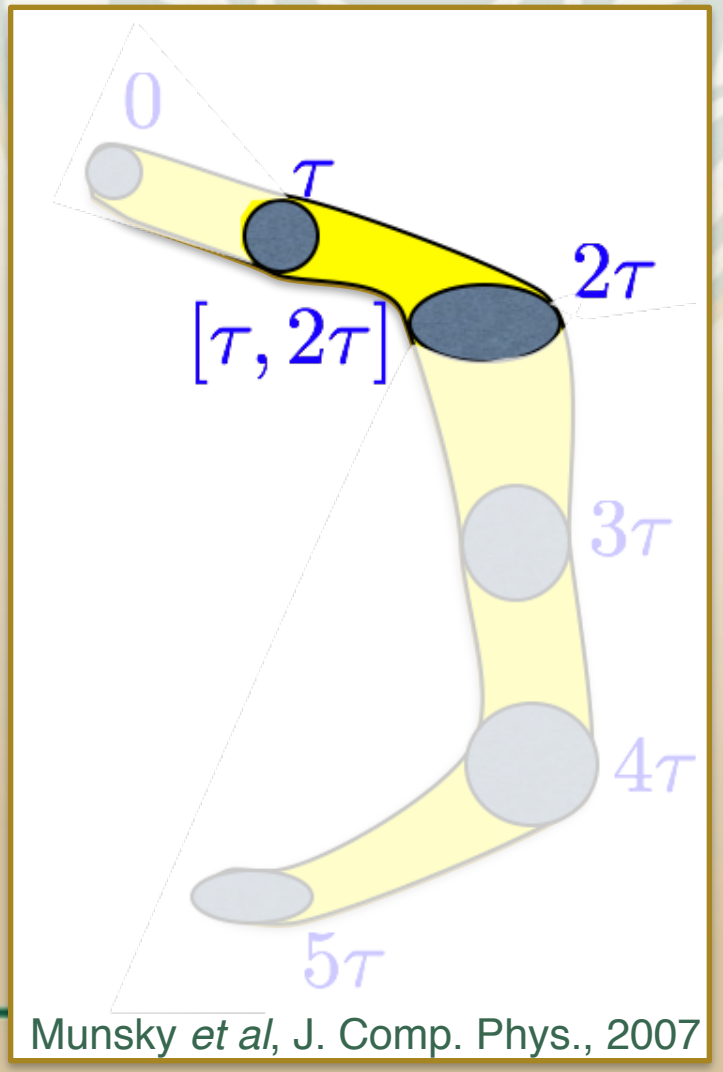
- Instead:
  - Discretize the time interval into smaller steps and solve a separate projection for each interval.



Munsky *et al*, J. Comp. Phys., 2007

Time Interval Discretization for the FSP

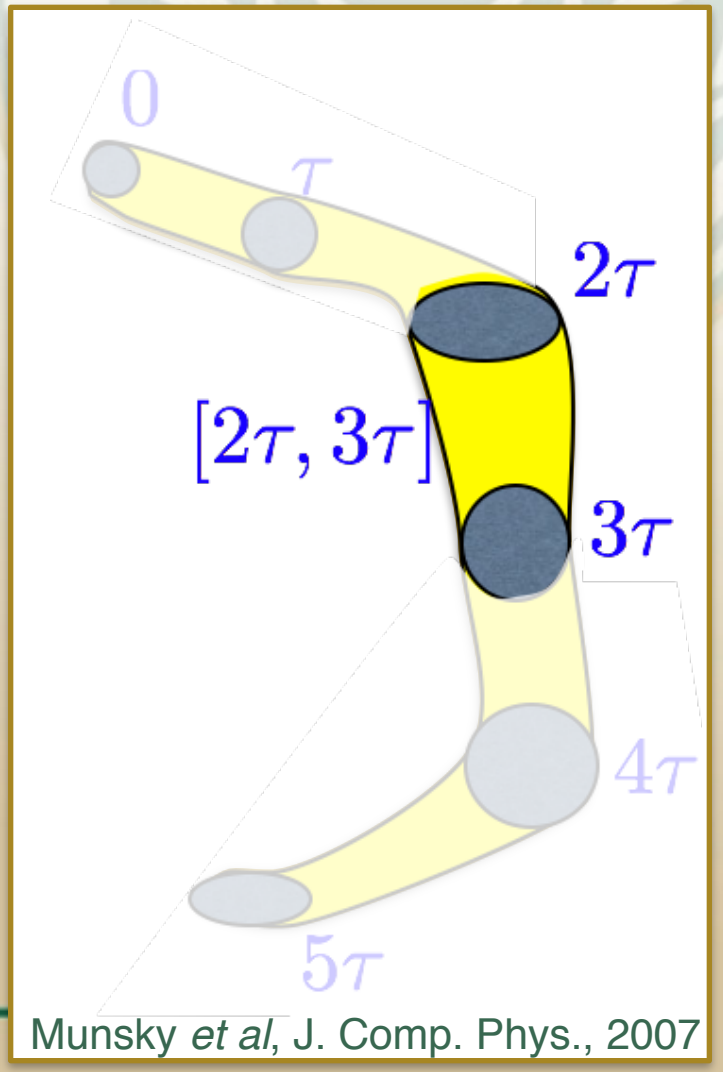
- Instead:
  - Discretize the time interval into smaller steps and solve a separate projection for each interval.



Munsky *et al*, J. Comp. Phys., 2007

Time Interval Discretization for the FSP

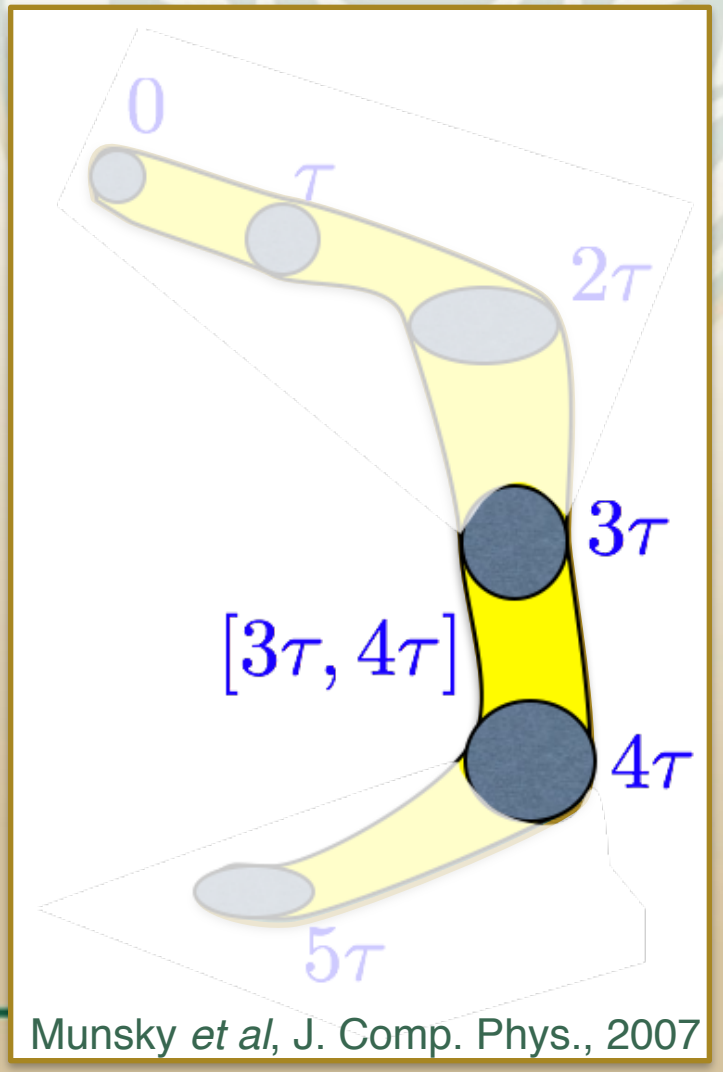
- Instead:
  - Discretize the time interval into smaller steps and solve a separate projection for each interval.



Munsky *et al*, J. Comp. Phys., 2007

Time Interval Discretization for the FSP

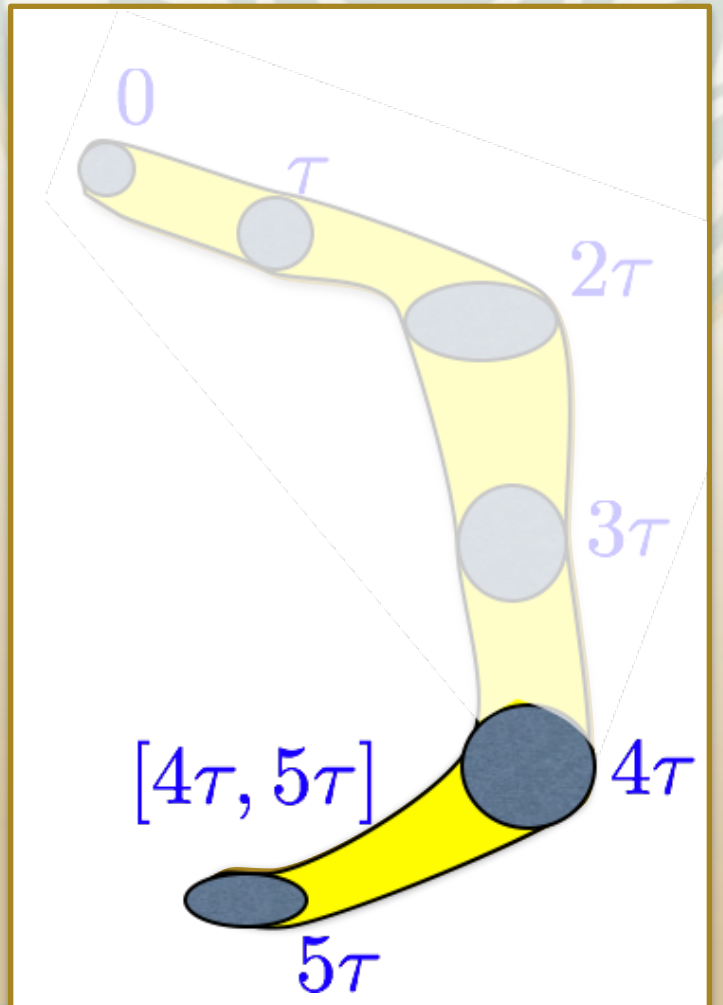
- Instead:
  - Discretize the time interval into smaller steps and solve a separate projection for each interval.



Munsky *et al*, J. Comp. Phys., 2007

Time Interval Discretization for the FSP

- Solving many small systems can be much faster than solving a single large system.
- Control the error at each step to obtain a guaranteed final error.
- Caching and reusing information from one step to the next may further reduce effort.



Munsky *et al*, J. Comp. Phys., 2007

Reductions to the FSP

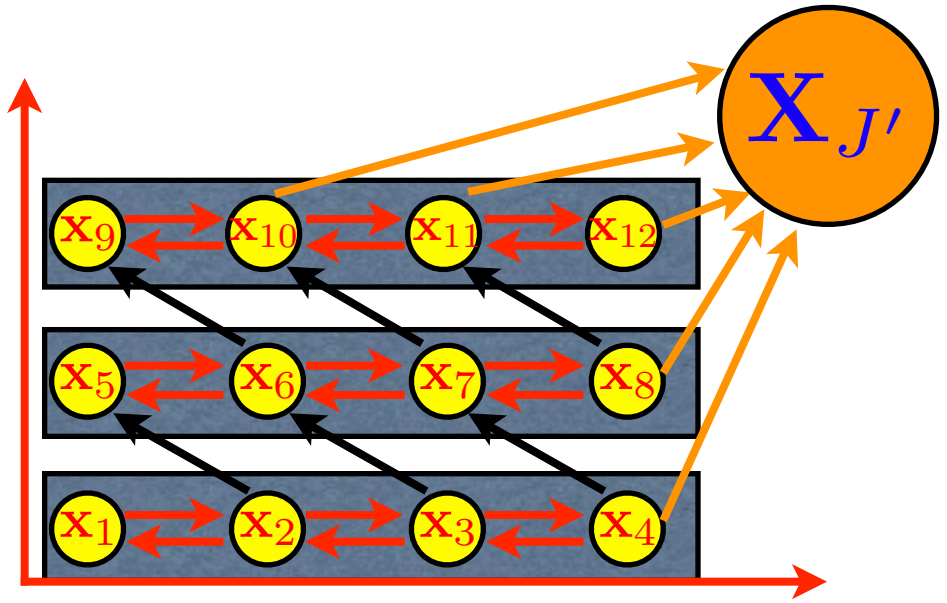
- Model Reduction Through Observability
- Time Interval Discretization
- Slow Manifold Projection
- Coarse Meshes for the CME

Perturbation Theory and the FSP

- Some reactions occur faster and more frequently than others.
- This can result in a separation of time-scales in the CME.
  - **Disadvantages**: Often results in numerical stiffness and increased computational complexity.
  - **Advantage**: May be able to apply perturbation theory to reduce computational effort.

Intuition (Slow Manifold Projection)

1. Begin with a finite state (projected) Markov process.
2. Group states connected by frequent reactions.



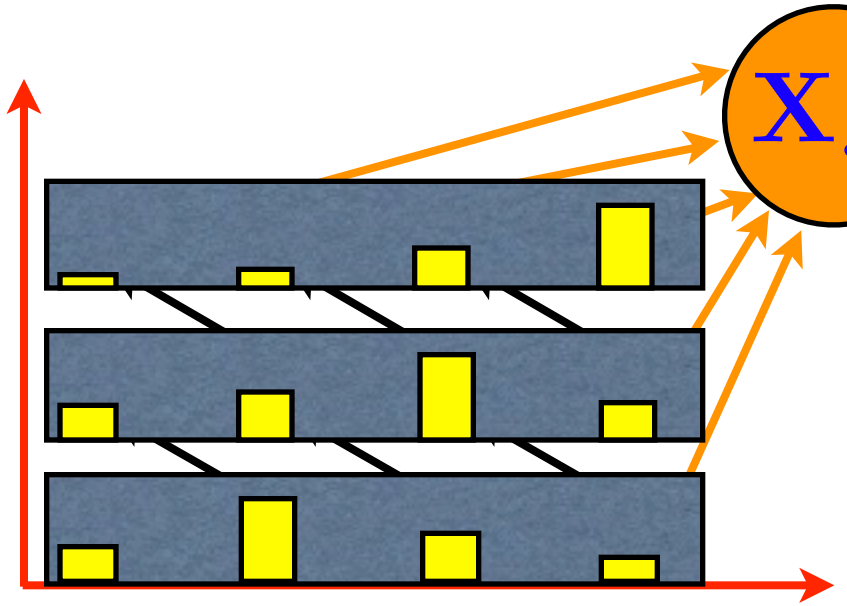
Red Arrows = Fast (Frequent) Reactions

Black Arrows = Slow (Rare) Reactions

Orange Arrows = (Rare) Transitions to Sink

Intuition (Slow Manifold Projection)

1. Begin with a finite state (projected) Markov process.
2. Group states connected by frequent reactions.
3. Find invariant distribution for each group.



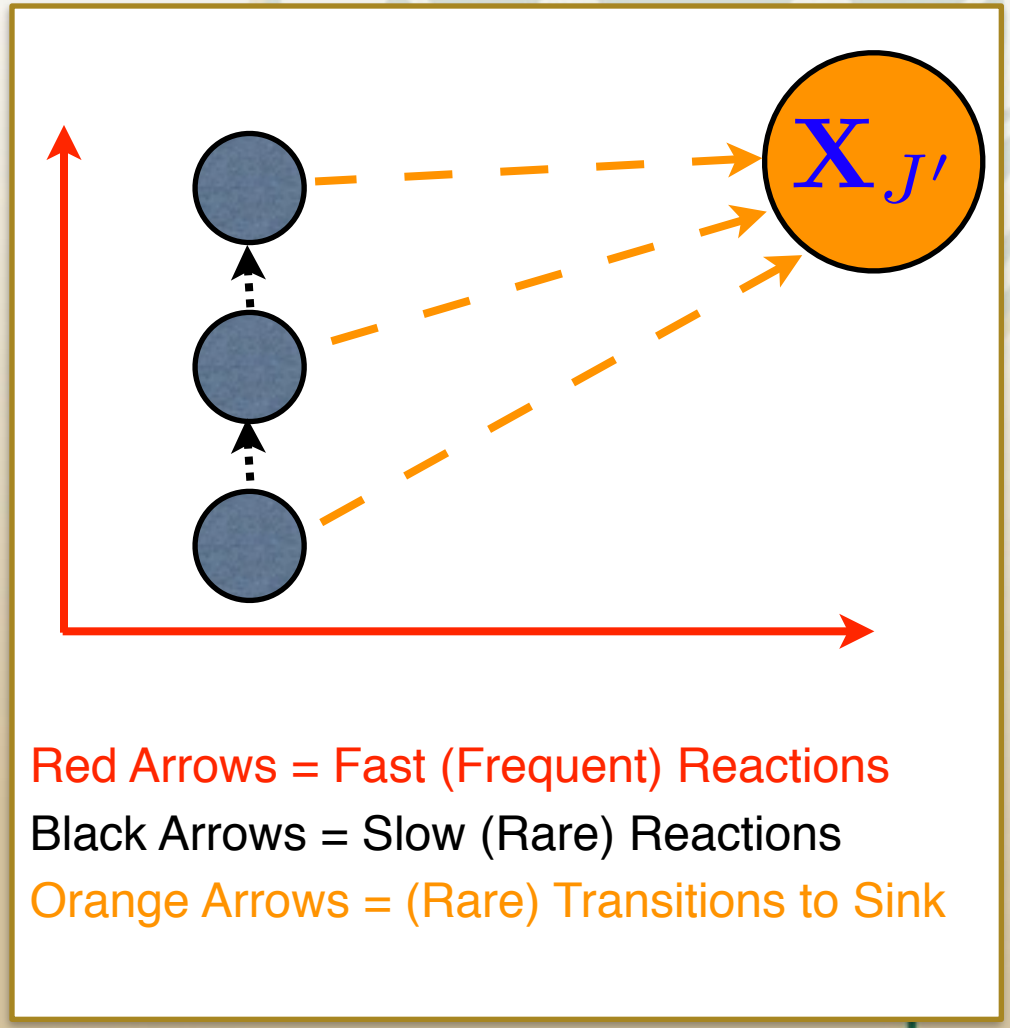
Red Arrows = Fast (Frequent) Reactions

Black Arrows = Slow (Rare) Reactions

Orange Arrows = (Rare) Transitions to Sink

Intuition (Slow Manifold Projection)

1. Begin with a finite state (projected) Markov process.
2. Group states connected by frequent reactions.
3. Find invariant distribution for each group.
4. Average to find the rates of the slow reactions.
5. Solve for the solution on the slow-manifold.
6. Lift solution to original coordinate system.



Reductions to the FSP

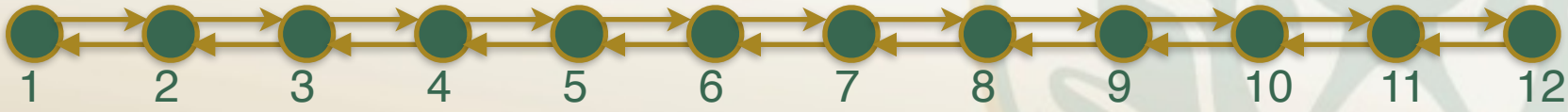
- Model Reduction Through Observability
- Time Interval Discretization
- Slow Manifold Projection
- Coarse Meshes for the CME

Coarse Mesh Approximation of the CME

- Precision requirements may change for different regions of the configurations space.
  - Small populations require great precision.
  - High populations require far less precision.
- By choosing a good coarse approximation of the CME, we can take advantage of this.
  - The general idea is similar to discretization for the numerical solution of a PDE.

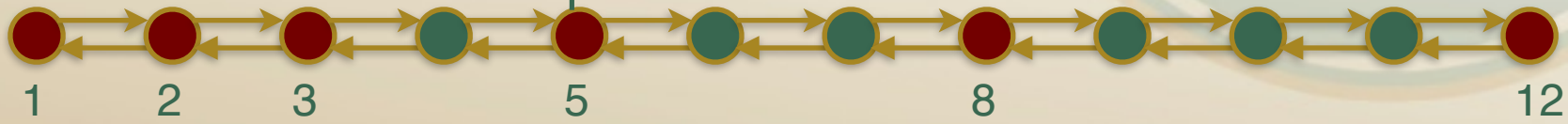
Coarse Mesh: One-species problem.

Start with the full 1-dimensional Markov lattice.



$$\dot{\mathbf{P}} = \mathbf{A} \cdot \mathbf{P}(t)$$ Original CME

Choose a subset of mesh points.



and specify an approximate relation for the probability of the removed points:

$$\mathbf{P} \approx \mathbf{\Phi q}(t)$$

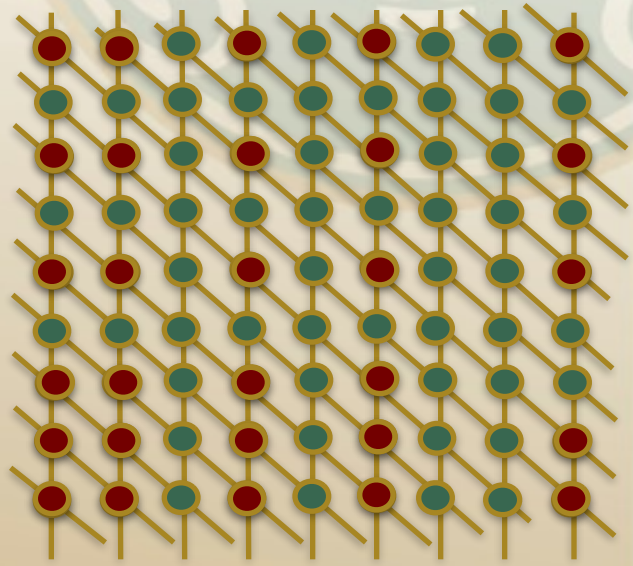
Solve the reduced system ODE: $\dot{\mathbf{q}} = \mathbf{\Phi}^{-L}\mathbf{A}\mathbf{\Phi}\mathbf{q}(t)$

and lift back to the original system coordinates:

$$\mathbf{P}(t) \approx \mathbf{\Phi}\exp(\mathbf{\Phi}^{-L}\mathbf{A}\mathbf{\Phi}t)\mathbf{\Phi}^{-L}\mathbf{P}(0)$$

Coarse Mesh: Multiple-Species Problems.

1. Begin with original lattice.
2. Choose interpolation points.
3. Form interpolation (shape) function: $\mathbf{P}(t) \approx \mathbf{\Phi q}(t)$
4. Project system to find reduced system of ODEs: $\dot{\mathbf{q}}(t) = \mathbf{\Phi}^{-L}\mathbf{A}\mathbf{\Phi q}(t)$
6. Solve reduced system.
7. Lift back to original coordinates.



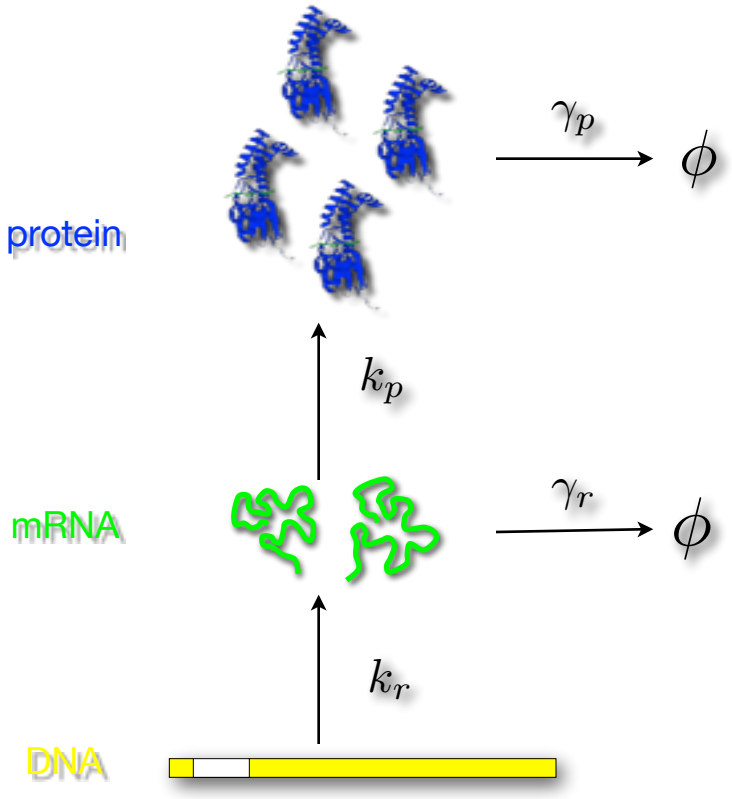
Finite State Projection Analyses

- Forming the Infinitesimal Generator
- Interpreting and Using FSP Sinks
- Advantages and Limitations of the FSP
- System Reductions to Improve FSP Efficiency
- Examples for Using the FSP

Examples Using the FSP:

- Transcription and Translation
- Feedback for Noise Suppression
- Stochastic Focussing / Stochastic Damping
- Stochastic Switches
- Stochastic Resonance
- Lambda Phage
- Bacterial heat shock

Gene Transcription and Translation



Reactants

$X_1(t)$ is # of mRNA; $X_2(t)$ is # of protein

Reactions

$$R_1 : \phi \xrightarrow{k_r} mRNA$$

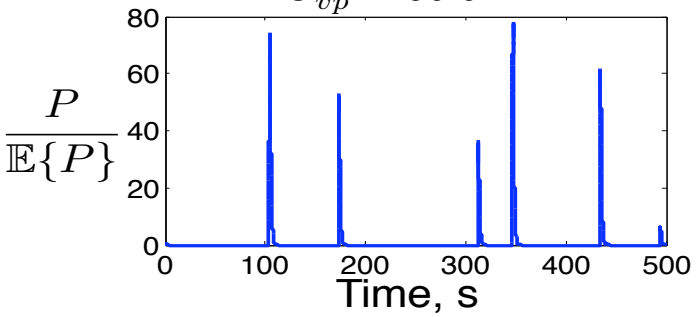
$$R_2 : mRNA \xrightarrow{\gamma_r} \phi$$

$$R_3 : mRNA \xrightarrow{k_p} protein + mRNA$$

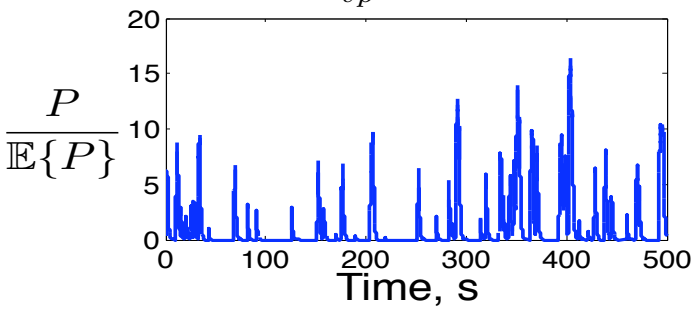
$$R_4 : protein \xrightarrow{\gamma_p} \phi$$

$$\mathbb{E}\{P\} = 100, \quad \gamma_r = \gamma_p = 1$$

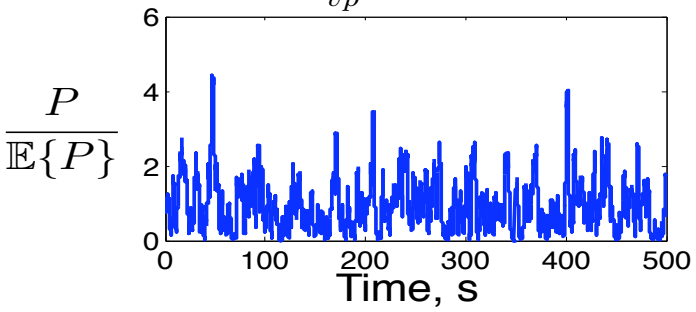
$k_r = 0.01 \quad k_p = 10,000$
$C_{vp}^2 = 50.01$



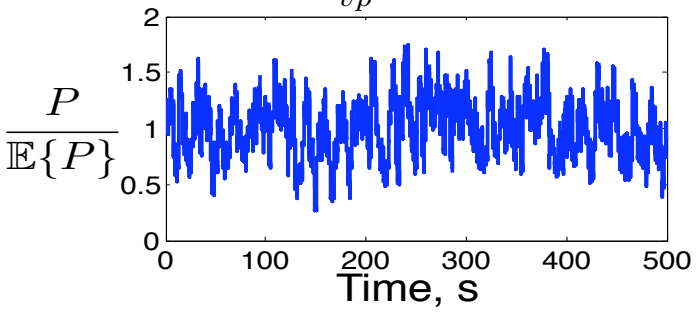
$k_r = 0.1 \quad k_p = 1000$
$C_{vp}^2 = 5.01$



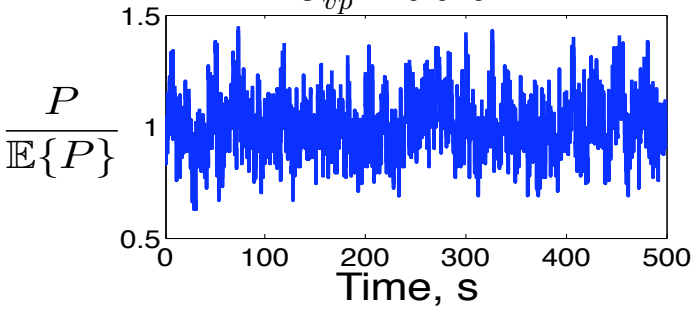
$k_r = 1 \quad k_p = 100$
$C_{vp}^2 = 0.51$



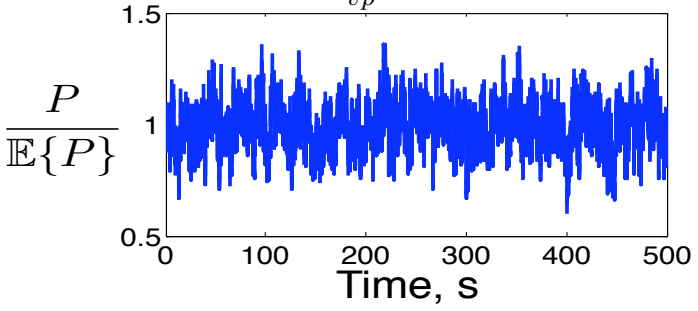
$k_r = 10 \quad k_p = 10$
$C_{vp}^2 = 0.06$



$k_r = 100 \quad k_p = 1$
$C_{vp}^2 = 0.015$



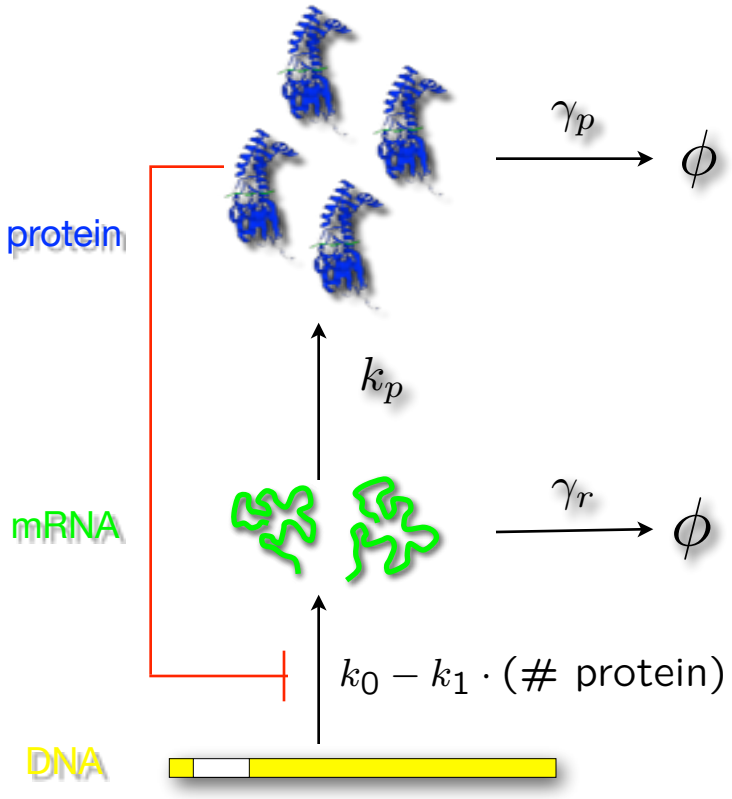
$k_r = 1000 \quad k_p = 0.1$
$C_{vp}^2 = 0.0105$



Examples:

- Transcription and Translation
- Feedback for Noise Suppression
- Stochastic Focussing / Stochastic Damping
- Stochastic Switches
- Stochastic Resonance
- Lambda Phage
- Bacterial heat shock

Noise Attenuation through Negative Feedback



Reactants

$X_1(t)$ is # of mRNA; $X_2(t)$ is # of protein

Reactions

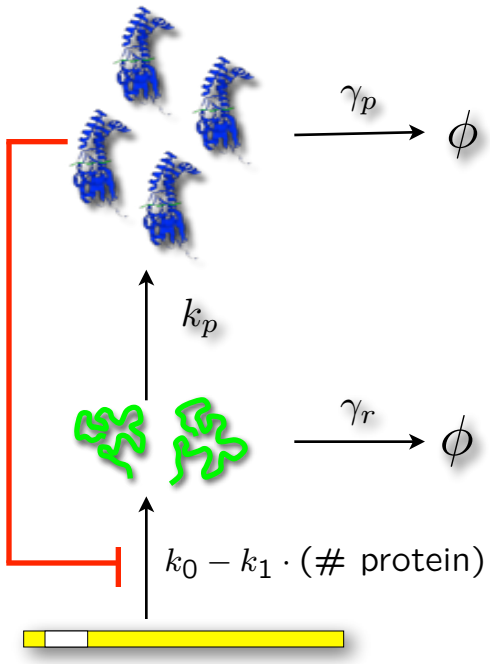
$$R_1 : \phi \xrightarrow{k_r} mRNA \qquad k_r = k_0 - k_1 \cdot (\#\ \text{protein})$$

$$R_2 : mRNA \xrightarrow{\gamma_r} \phi$$

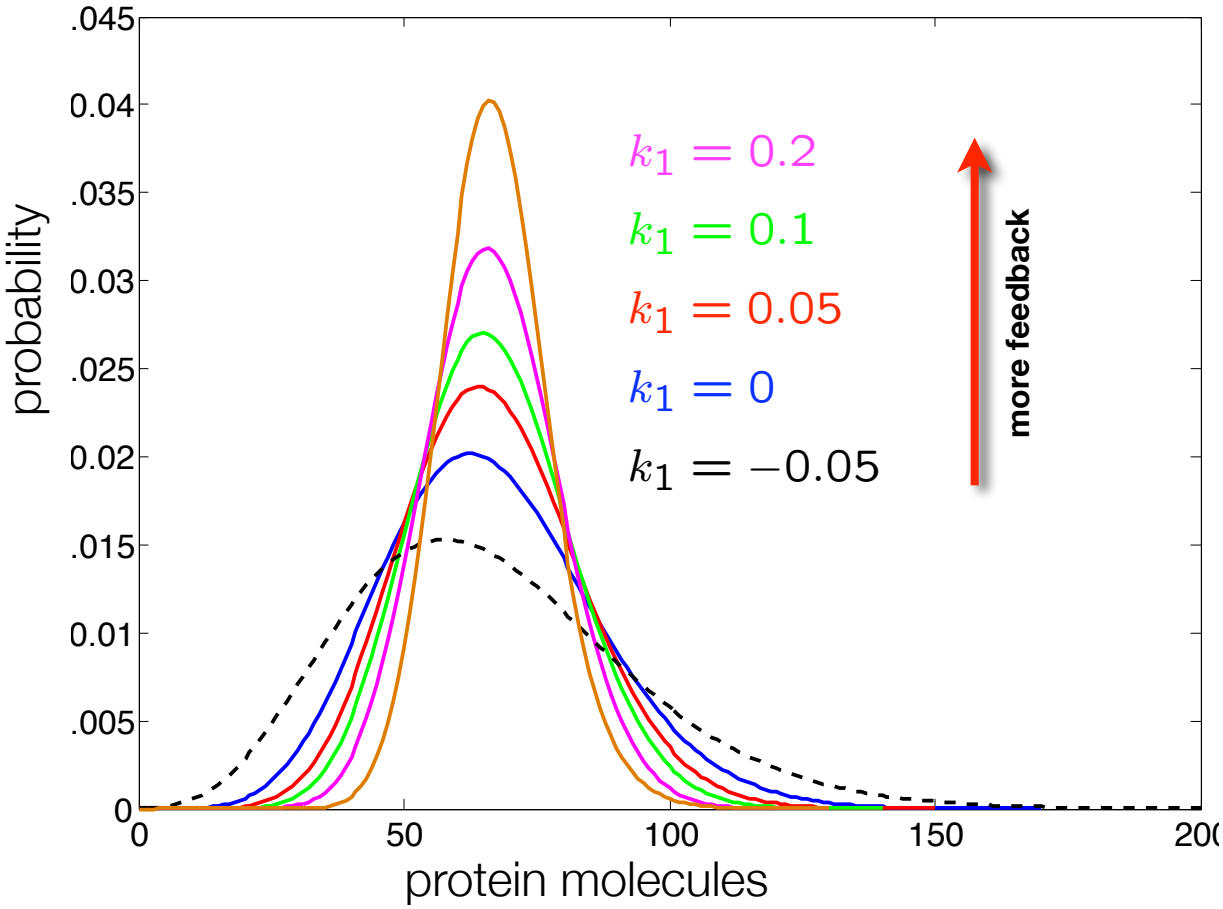
$$R_3 : mRNA \xrightarrow{k_p} protein + mRNA$$

$$R_4 : protein \xrightarrow{\gamma_p} \phi$$

Example



$\gamma_p = \gamma_r = 1 \qquad k_p = 10;$

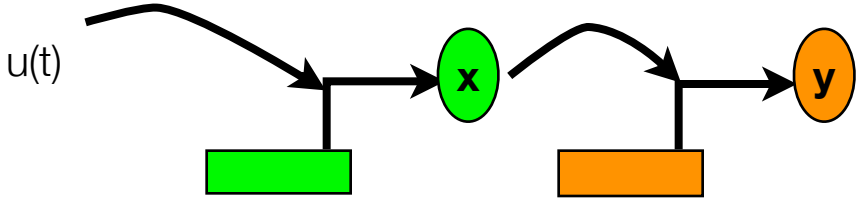


Note that these distributions are NOT Gaussian.

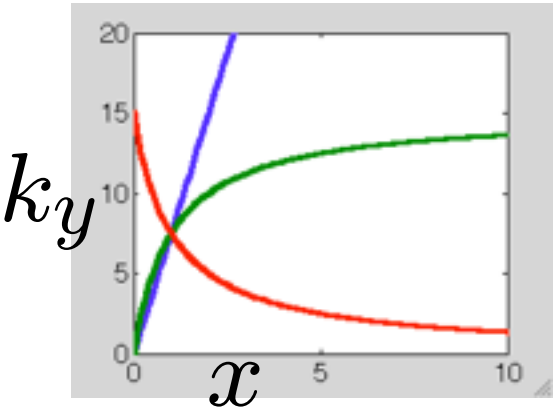
Examples:

- Transcription and Translation
- Feedback for Noise Suppression
- Stochastic Focussing / Stochastic Damping
- Stochastic Switches
- Stochastic Resonance
- Lambda Phage
- Bacterial heat shock

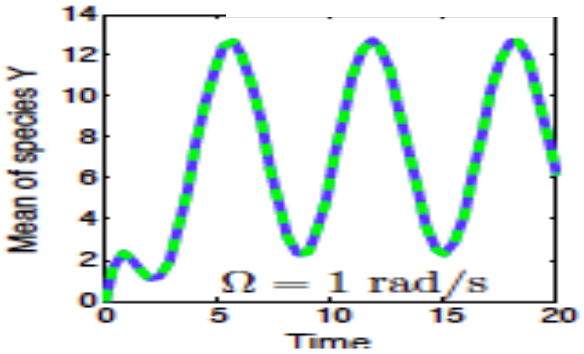
Effects of Nonlinearities



u(*t*) activates production of *X, which regulates Y*

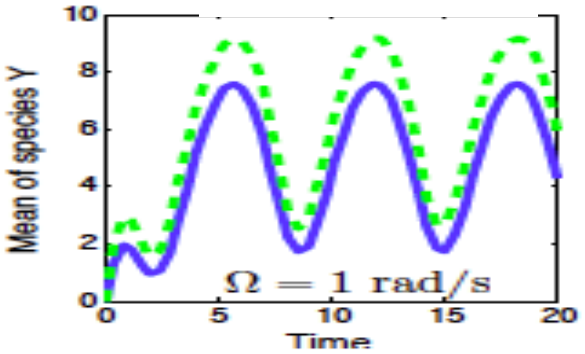


Species *X activates Y* in linear fashion.



The stochastic mean is *equal to* the deterministic model.

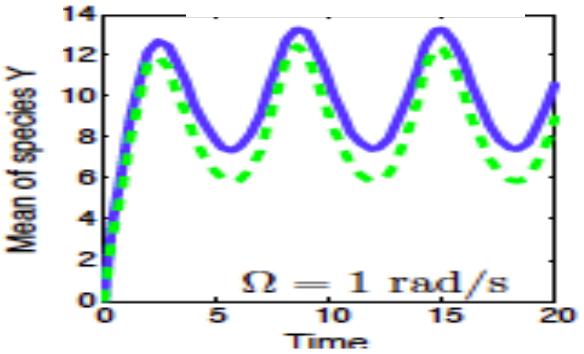
Species *X activates Y* in concave function.



The stochastic mean is *less than* the deterministic model.

stochastic damping

Species *X represses Y* in convex fashion.



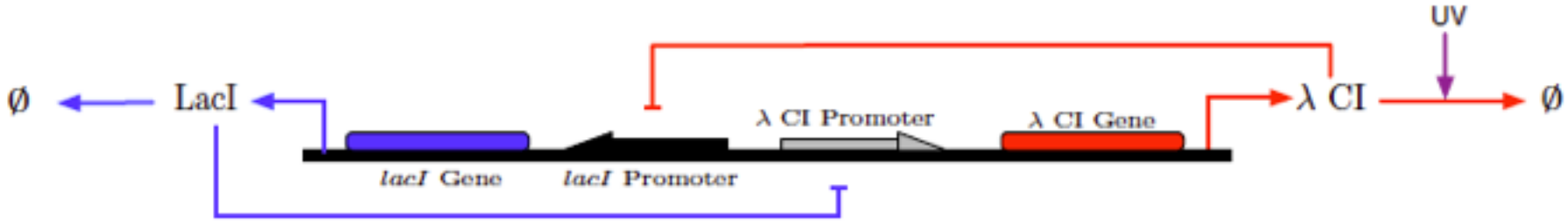
The stochastic mean is *more than* the deterministic model.

stochastic focussing

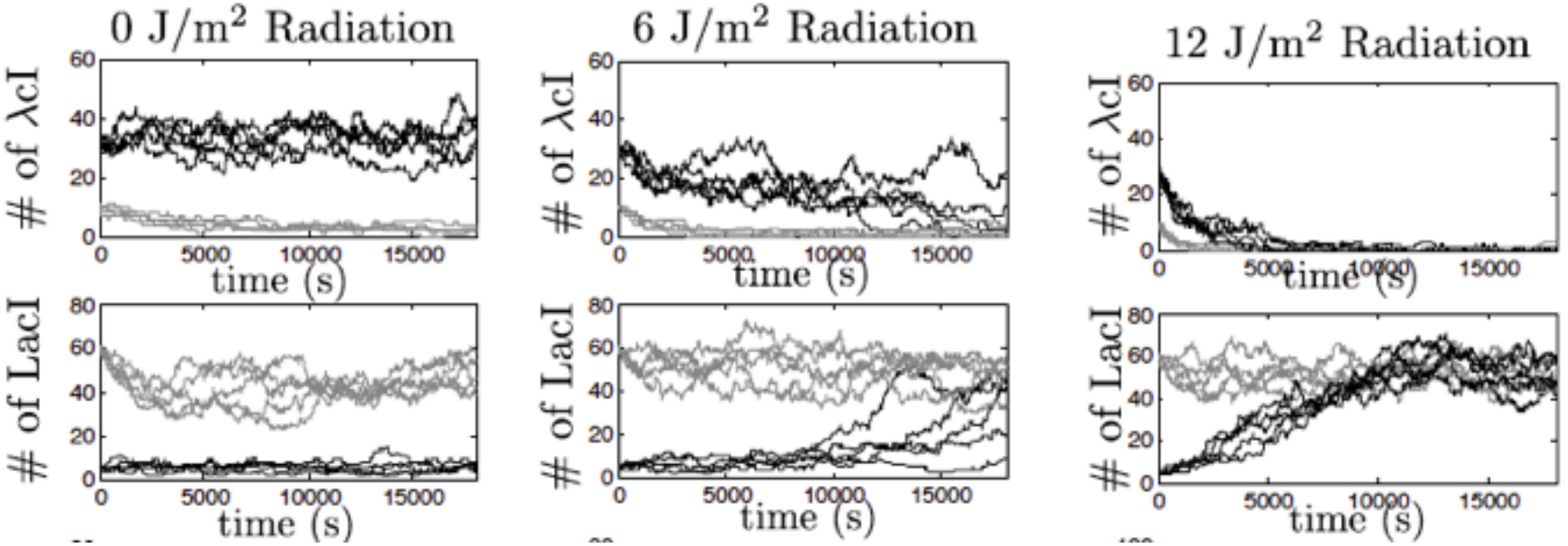
Examples:

- Transcription and Translation
- Feedback for Noise Suppression
- Stochastic Focussing / Stochastic Damping
- Stochastic Switches
- Stochastic Resonance
- Lambda Phage
- Bacterial heat shock

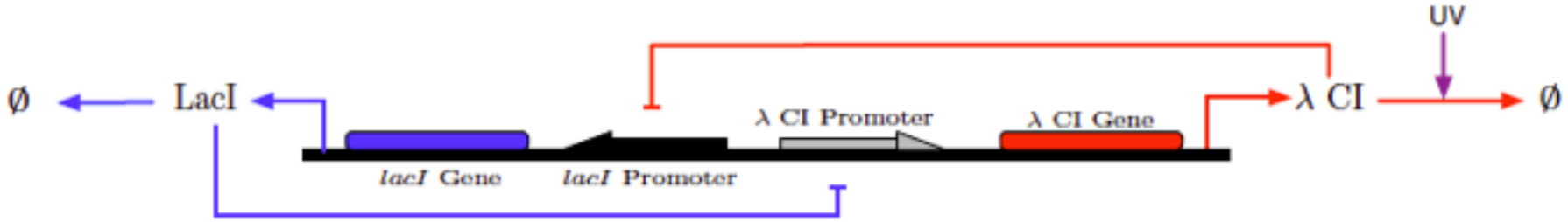
The Genetic Toggle Switch

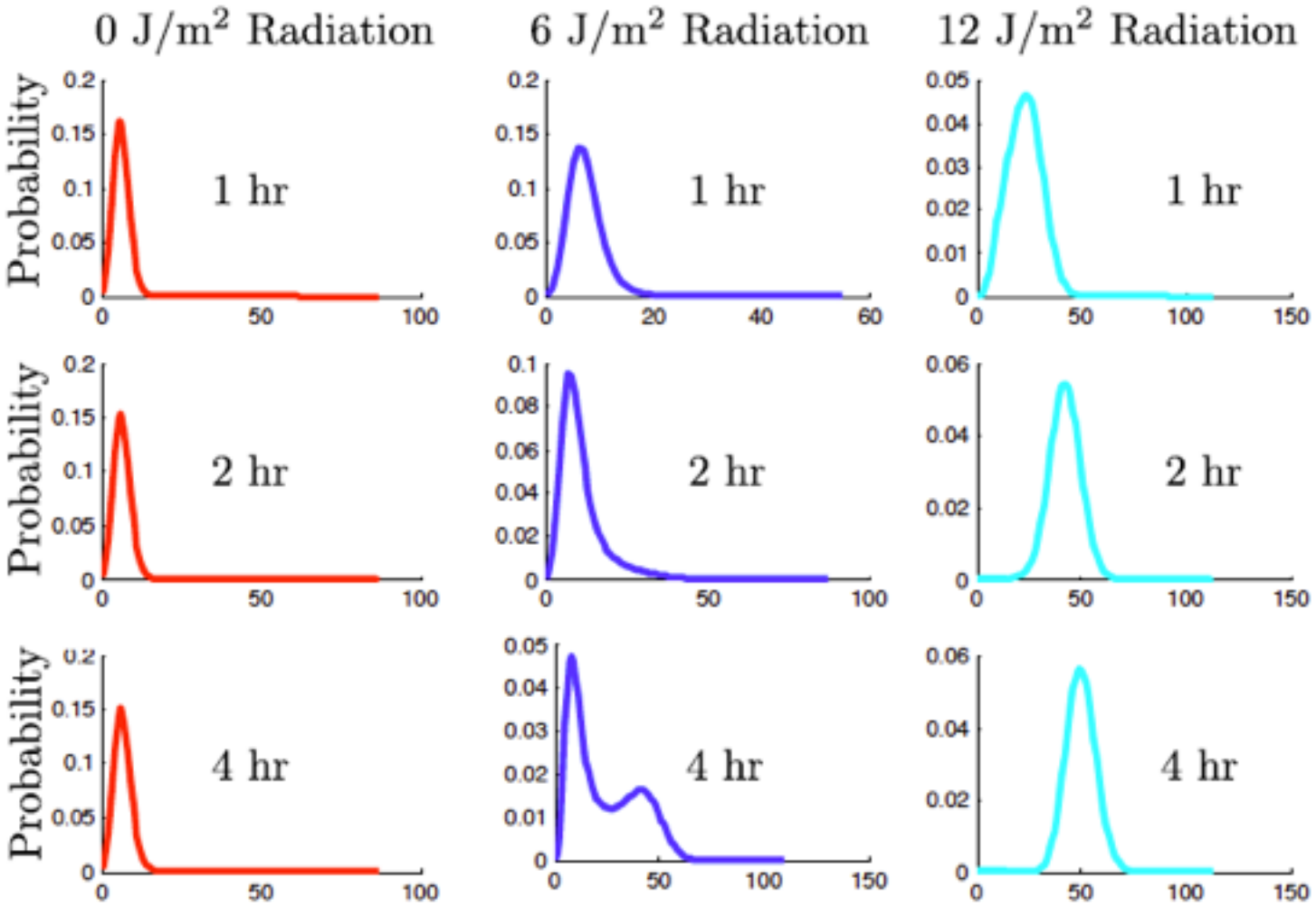


LacI inhibits production of λcI
λcI inhibits production of LacI
UV Radiation increases degradation of λcI



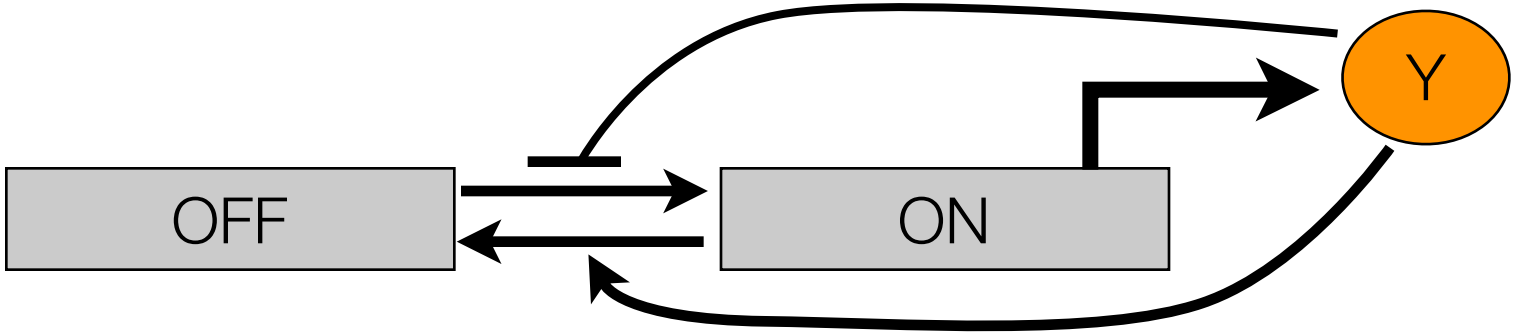
The Genetic Toggle Switch





Examples:

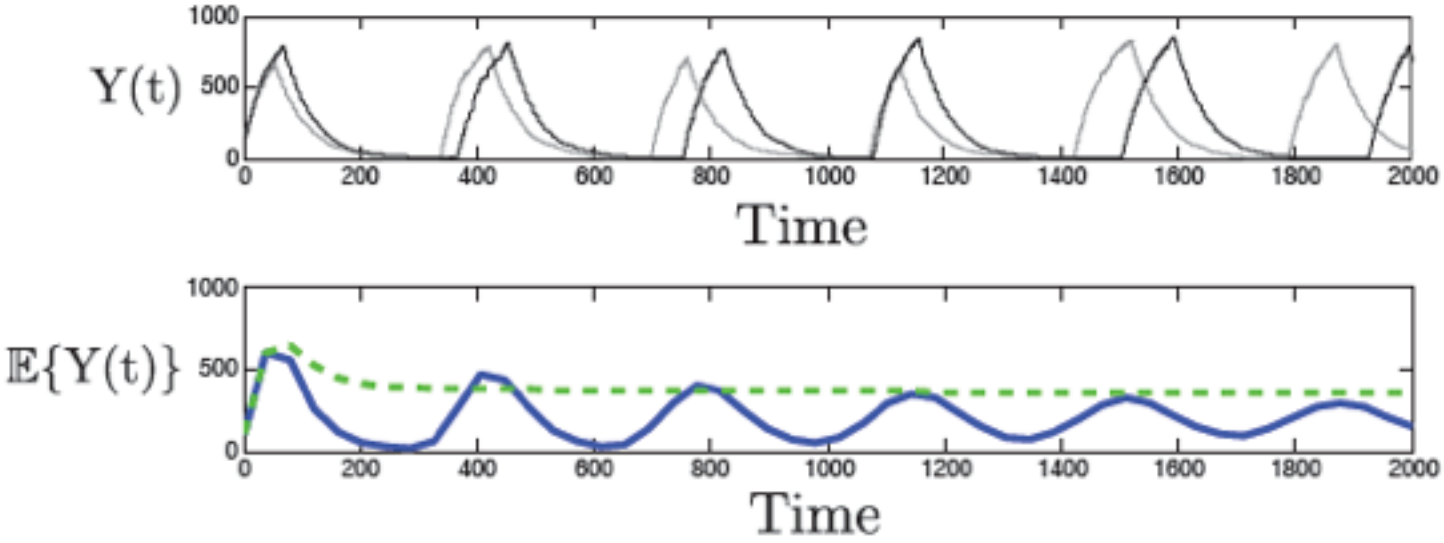
- Transcription and Translation
- Feedback for Noise Suppression
- Stochastic Focussing / Stochastic Damping
- Stochastic Switches
- Stochastic Resonance
- Lambda Phage
- Bacterial heat shock



The *y gene* switches between ON and OFF states.

Y strongly inhibits activation.

High Y concentrations also increase deactivation.

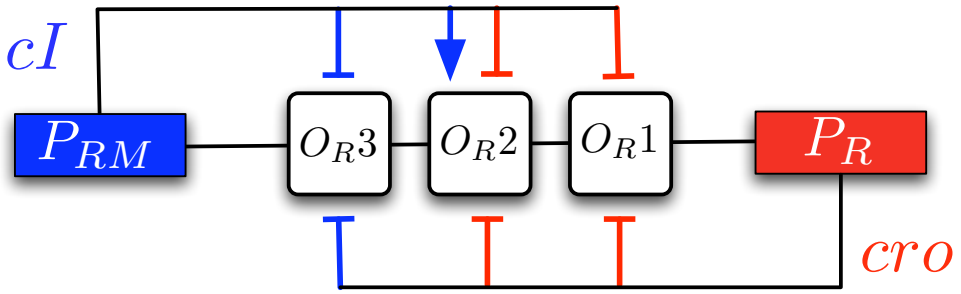


Stochasticity enables sustained oscillations.

Examples:

- Transcription and Translation
- Feedback for Noise Suppression
- Stochastic Focussing / Stochastic Damping
- Stochastic Switches
- Stochastic Resonance
- Lambda Phage
- Bacterial heat shock

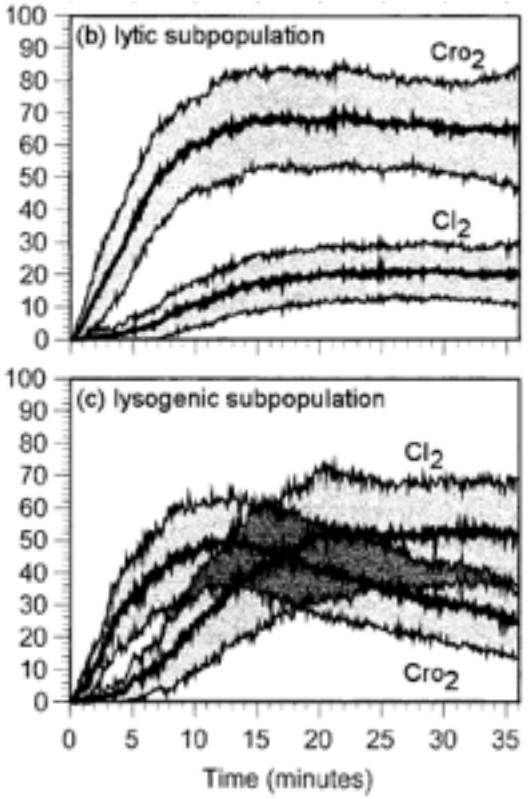
A toy model of phage lambda



- We consider only the core of the lambda switch.
- Two proteins, cI and cro.
- These activate and repress the P_R and P_{RM}promoters according to the model of Shea and Ackers, 1985.

The Phage Lambda Lysis-Lysogeny Decision

Arkin, Ross, McAdams, 1998.
Full Model



Lytic fate

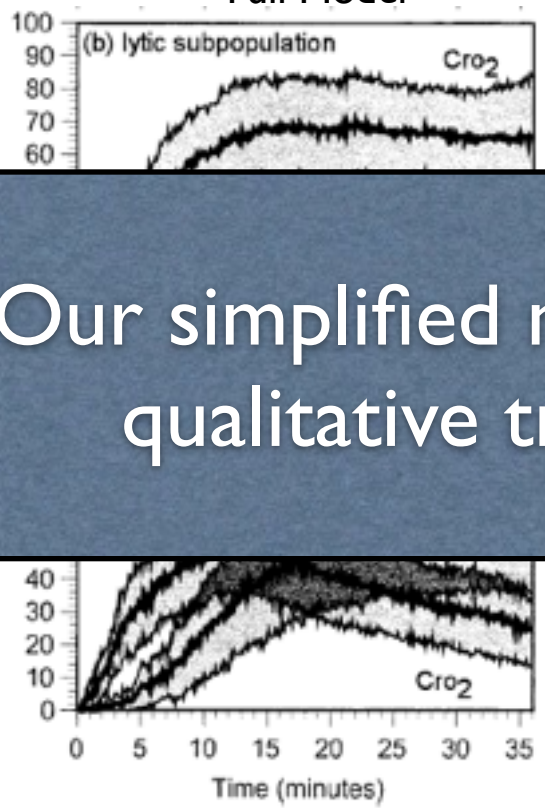
- ★ Cro reaches a high level before CI is produced in much quantity.
- ★ Cro represses transcription of CI.

Lysogenic fate

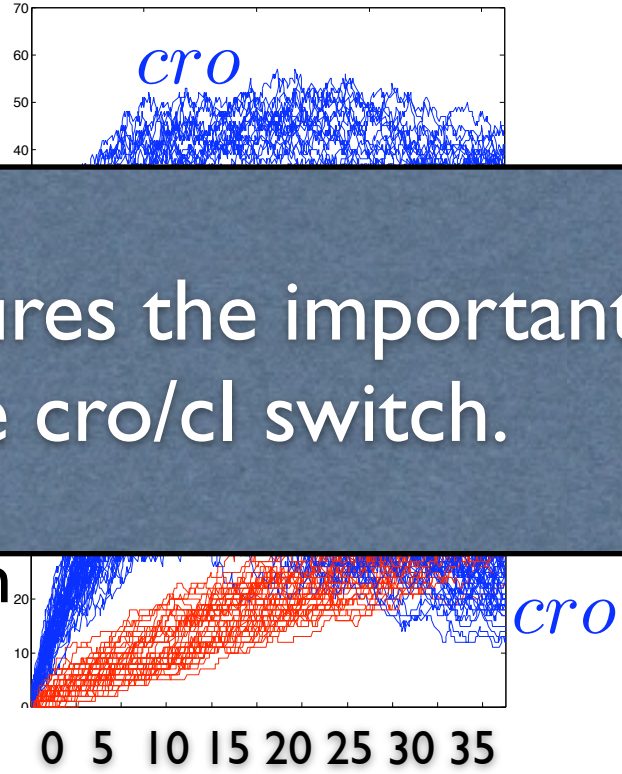
- ★ CI increases a little earlier.
- ★ CI represses transcription of Cro.
- ★ CI is free to increase even further.

Relevance of Simplified Model

Arkin, Ross, McAdams, 1998.
Full Model



Simplified model



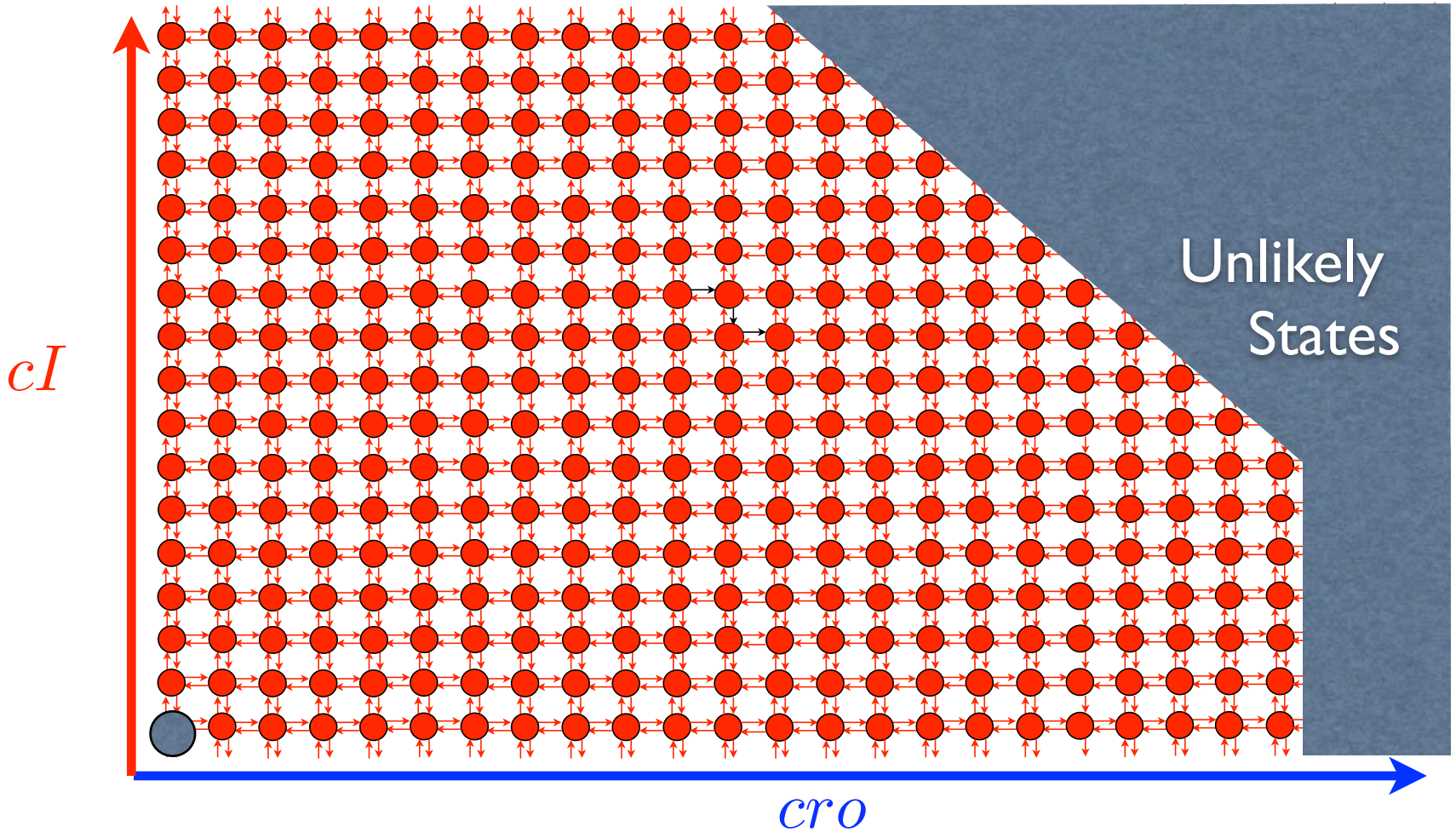
Lytic

subpopulation

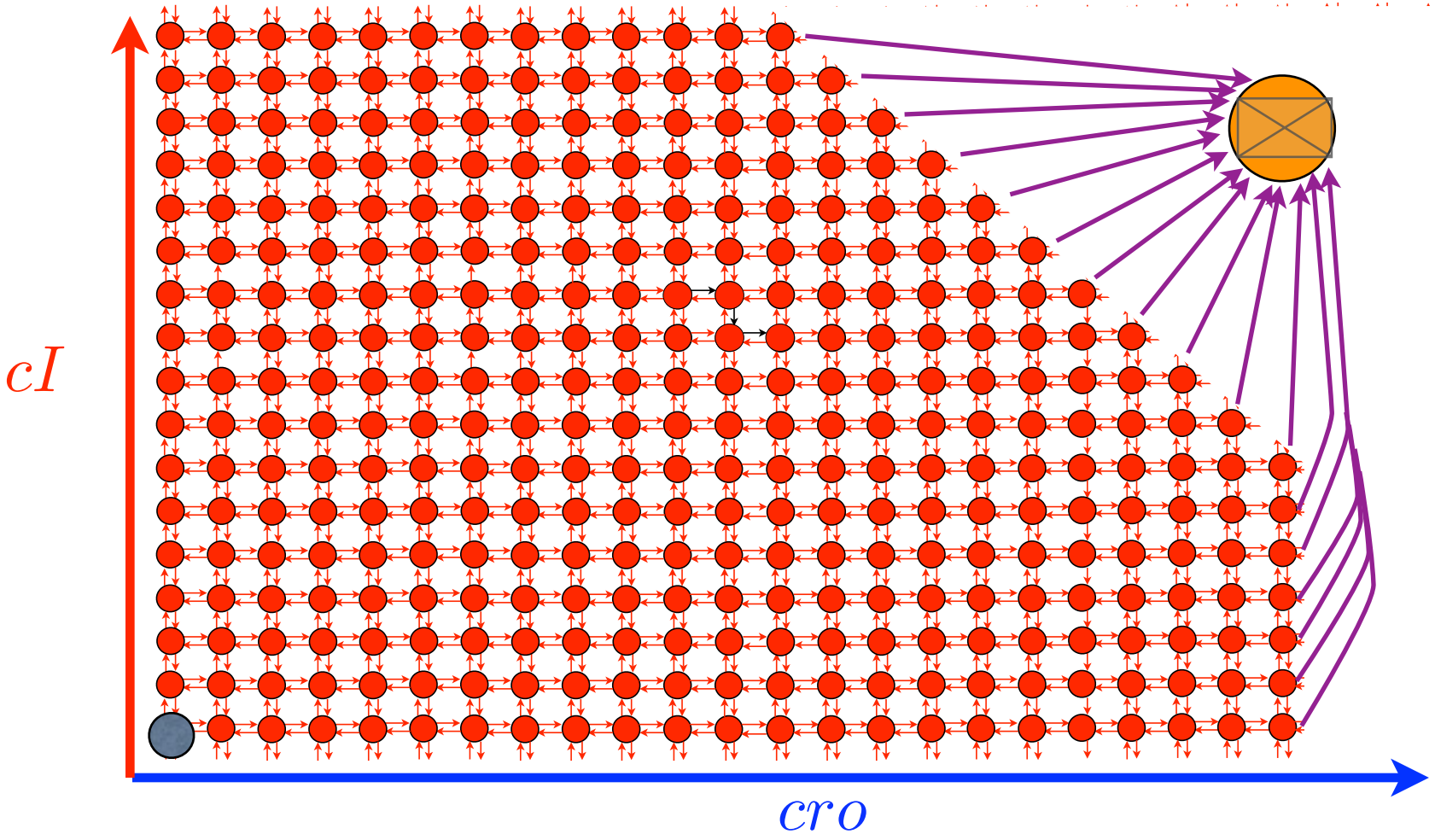
Our simplified model captures the important qualitative trends of the cro/cI switch.

Computations done using Gillespie's SSA.

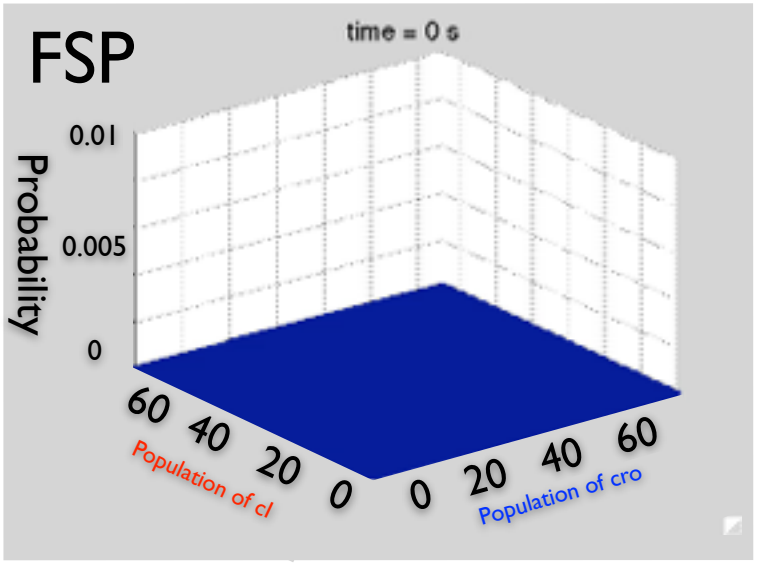
Applying the FSP to the Phage Lambda Switch

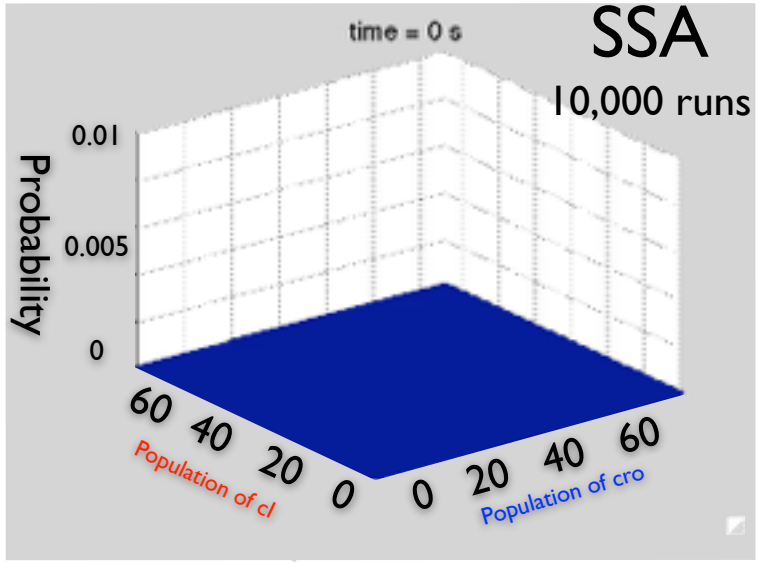


Applying the FSP to the Phage Lambda Switch



Efficiency and Accuracy of FSP Results





Method	# Simulations	Time (s)	$\|\|\text{Error}\|\|_1$
FSP	– [a]	163	$\leq 5.3 \times 10^{-3}$

Guaranteed

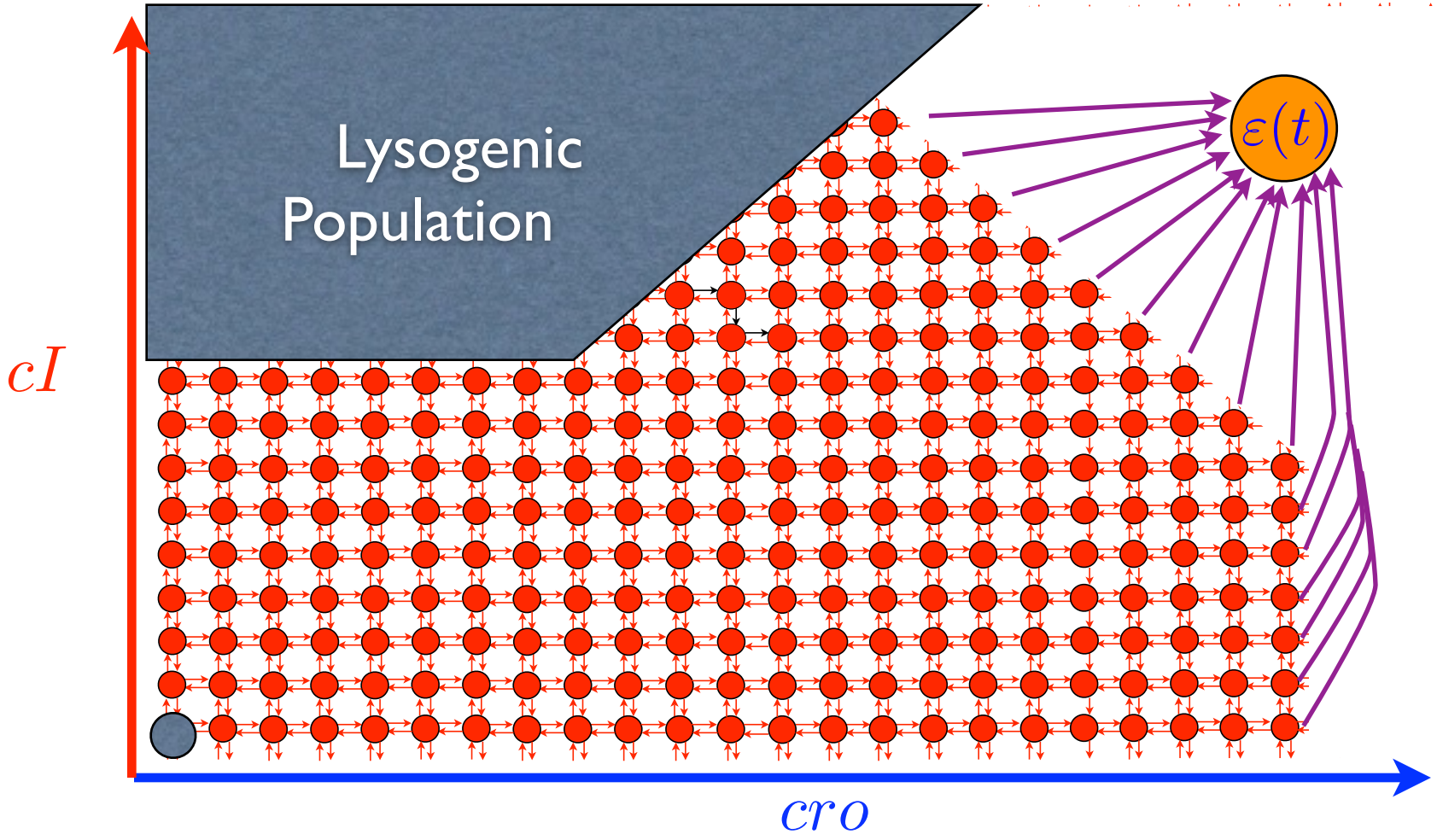
No Guarantees

[a] The FSP algorithm is run only once.

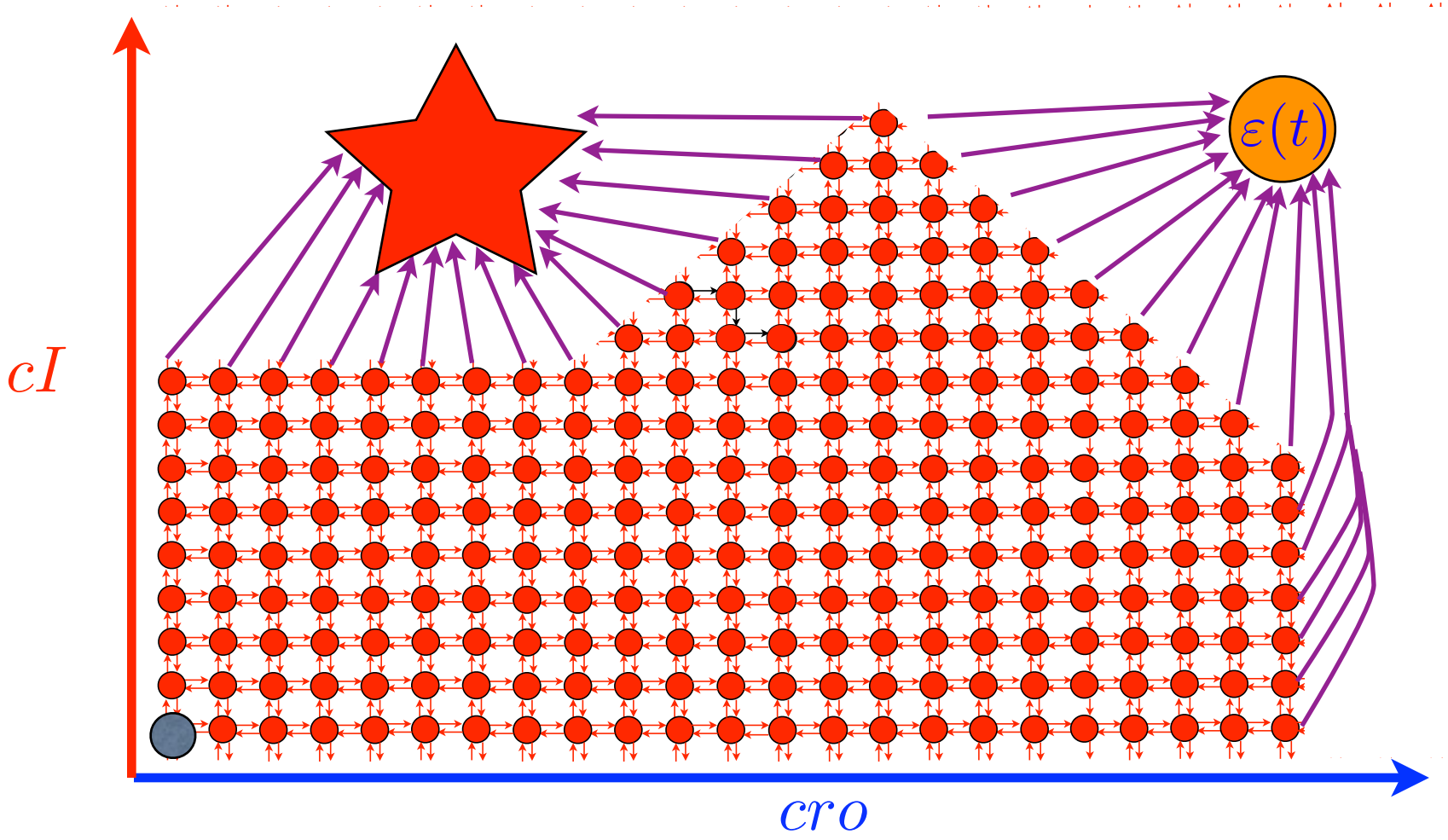
Additional information available with the FSP solution

- In many cases the FSP is faster and more accurate the Monte Carlo methods.
- Higher precision allows greater flexibility.
  - ★ Direct Computation of Switch Rates.

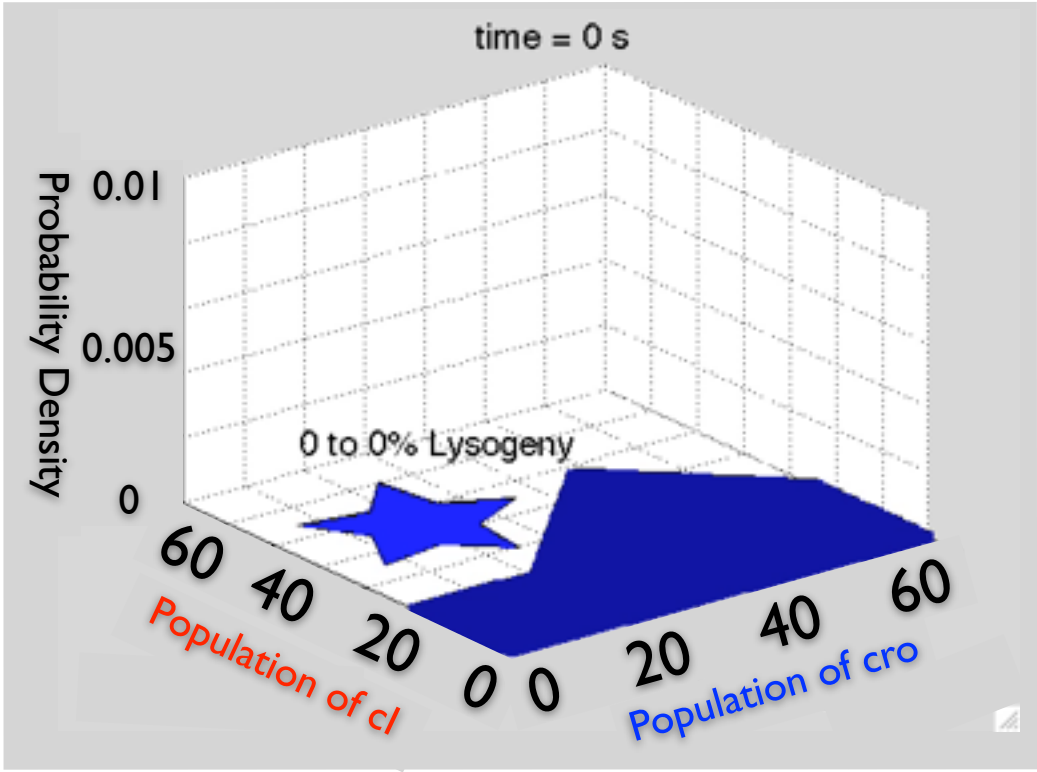
Using the FSP to Compute Switch Rates



Using the FSP to Compute Switch Rates



Using the FSP to Compute Switch Rates



Method	Time (s)	Relative Error	Guarantee?
FSP	25.5 s	< 0.08 %	yes
10^4 SSA runs	440.0 s	≈ 0.90 %	no

Additional information available with the FSP solution

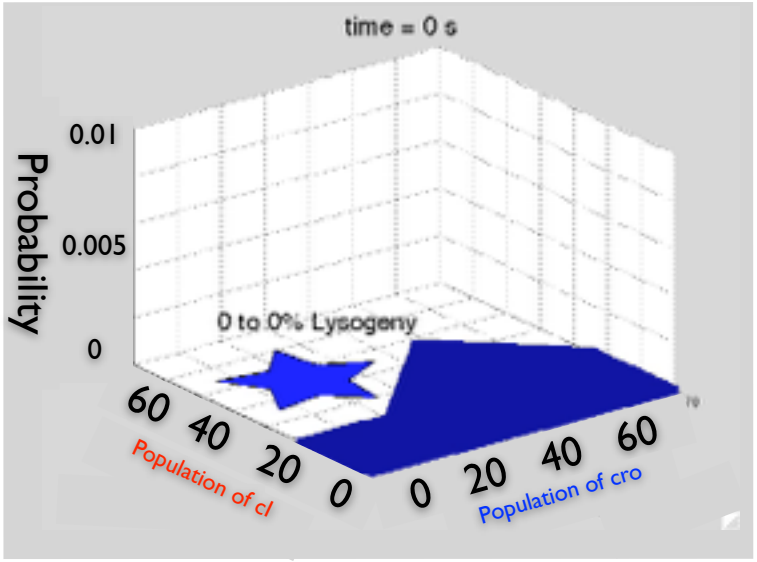
- In many cases the FSP is faster and more accurate the Monte Carlo methods.
- Higher precision allows greater flexibility.
  - ★ Direct Computation of Switch Rates.
  - ★ Simultaneous consideration of many different initial conditions.

Comparing different initial conditions.

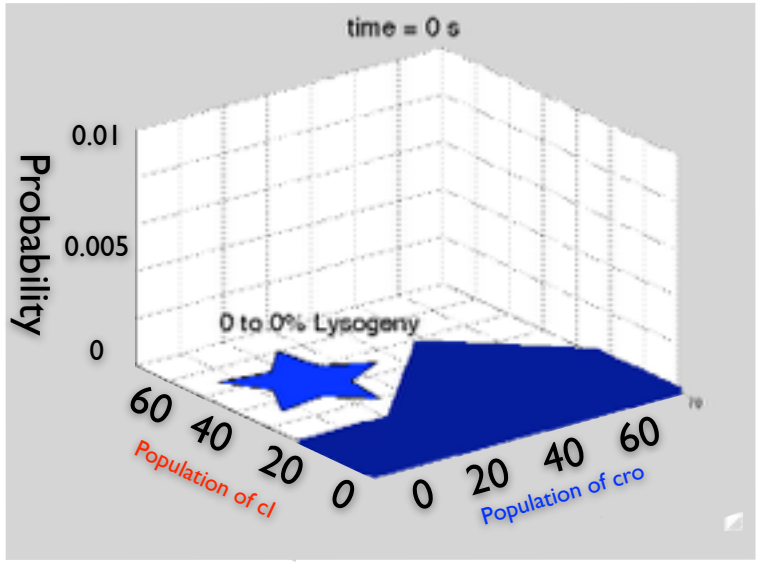
- The FSP is an approximate map of distributions from one time to another.
- This map is valid for any initial distribution.
  - ★ Once computed, this map is cheap to apply again and again.
  - ★ The map automatically provides error bounds for any initial condition!

Comparing different initial conditions. (Increase in)

$cI_0 = 0$
$cro_0 = 0$



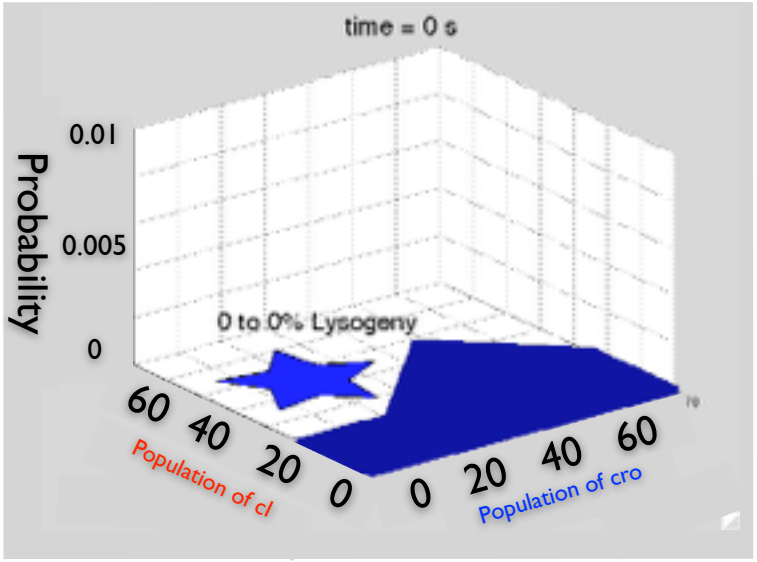
$cI_0 = 0$
$cro_0 = 5$



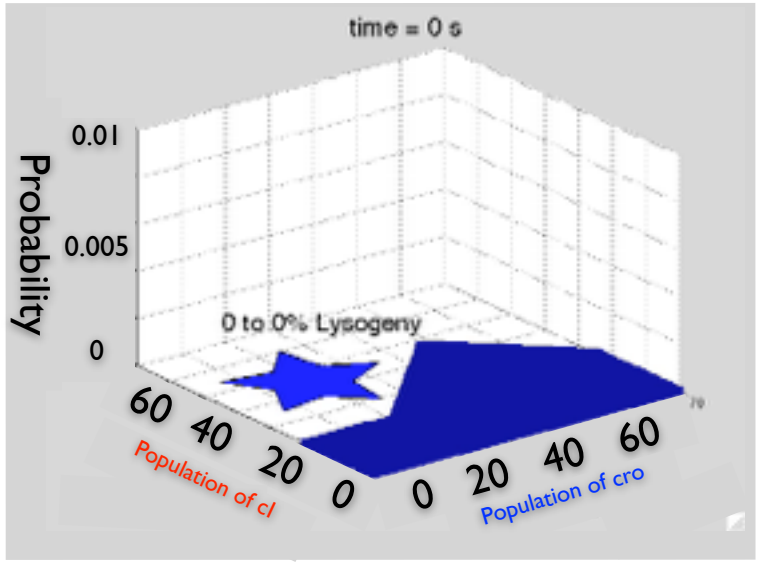
Increasing the initial amount of cro yields a slight decrease in the lysogeny rate.

Comparing different initial conditions. (Increase in cI)

$cI_0 = 0$
$cro_0 = 0$

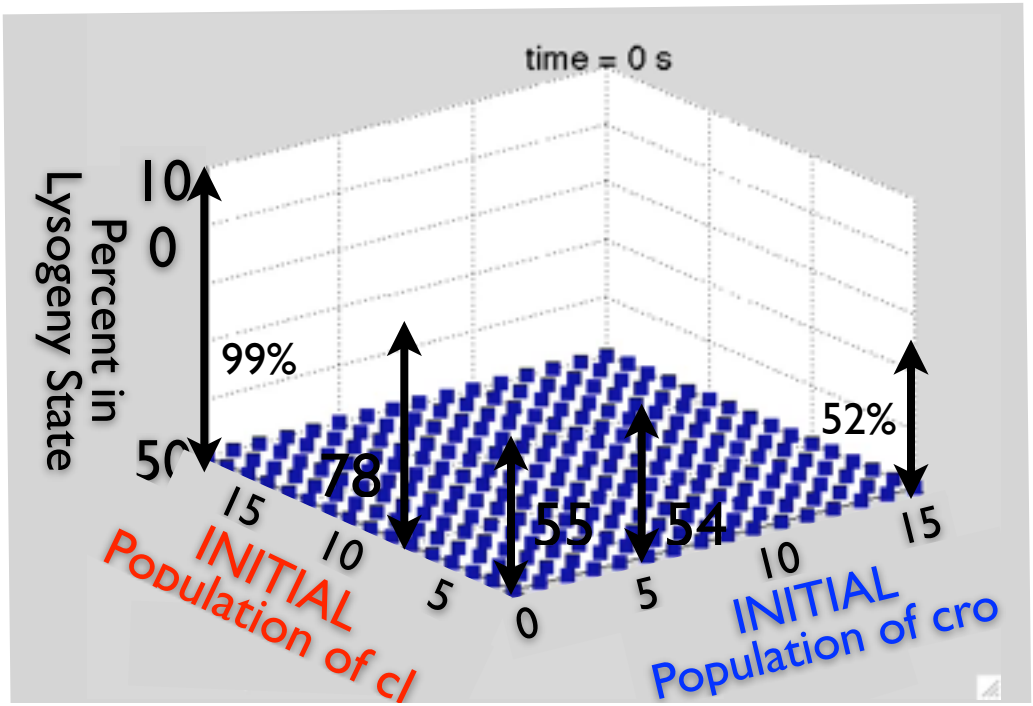


$cI_0 = 5$
$cro_0 = 0$



Increasing the initial amount of cI yields a significant increase in lysogeny rate.

Simultaneous comparison of an array of initial condition.)



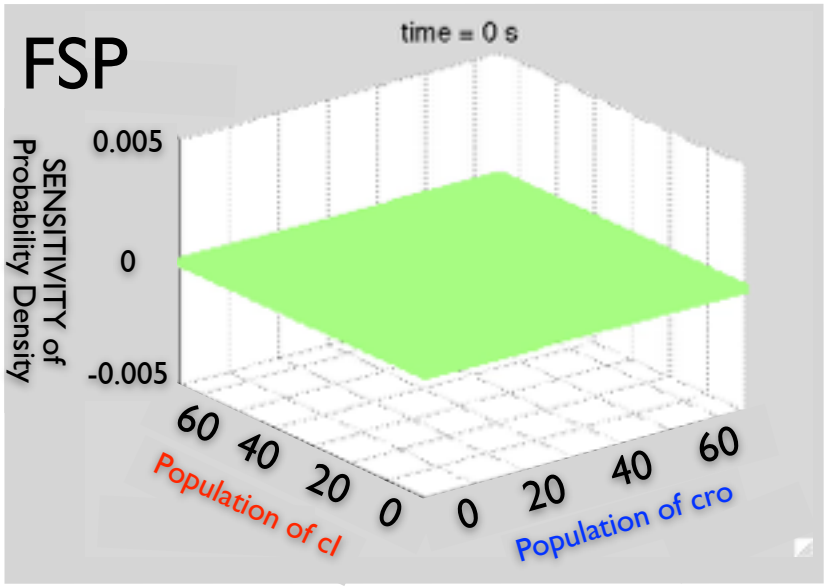
Method	Time (s)	# I.C.'s	$\|\|Error\|\|_1$	Guarantee?
FSP	66.9 s	2000	$< 1 \times 10^{-4}$	yes
10^4 SSA runs	440.0 s	1	≈ 0.09	no
10^{13} SSA runs	$\approx$ 14,000 years!	2000	$\approx 1 \times 10^{-4}$	no

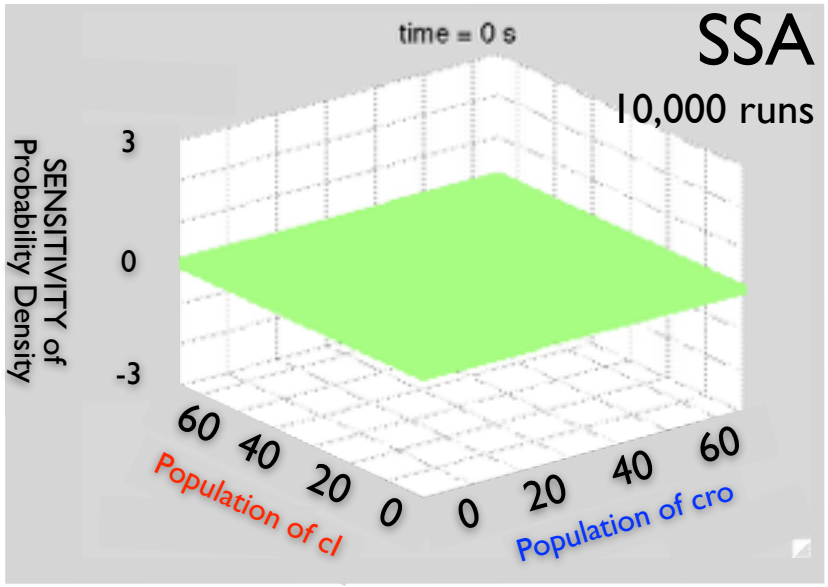
Additional information available with the FSP solution

- In many cases the FSP is both faster and more accurate than other available methods.
- Higher precision allows greater flexibility.
  - ★ Direct Computation of Switch Rates.
  - ★ Simultaneous consideration of many different initial conditions.
  - ★ Sensitivity to parameter changes.

Parametric Sensitivity of Probability Distributions.

Sensitivity to a small increase in cell Volume.





★ Sensitivity analysis requires a huge degree of accuracy.

★ Monte Carlo methods would require hundreds of millions of runs!!

Examples:

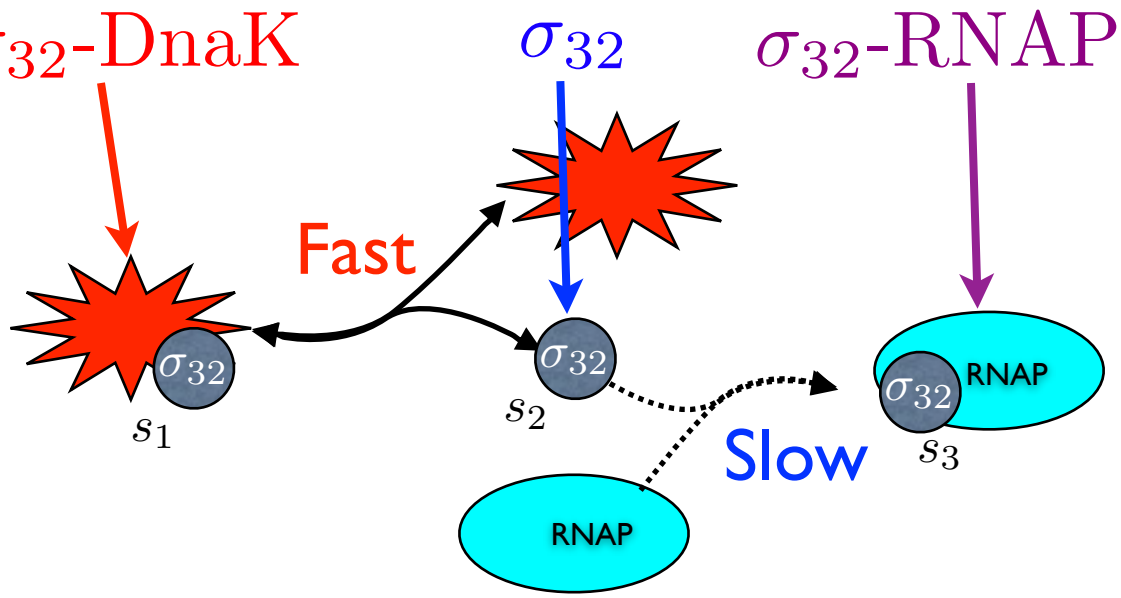
- Transcription and Translation
- Feedback for Noise Suppression
- Stochastic Focussing / Stochastic Damping
- Stochastic Switches
- Stochastic Resonance
- Lambda Phage
- Bacterial heat shock

Toy Heat Shock Model in *E. coli*

3 forms for σ_{32}:

$$S_1 \underset{k_2}{\overset{k_1}{\rightleftarrows}} S_2$$

$$S_2 \xrightarrow{k_3} S_3$$

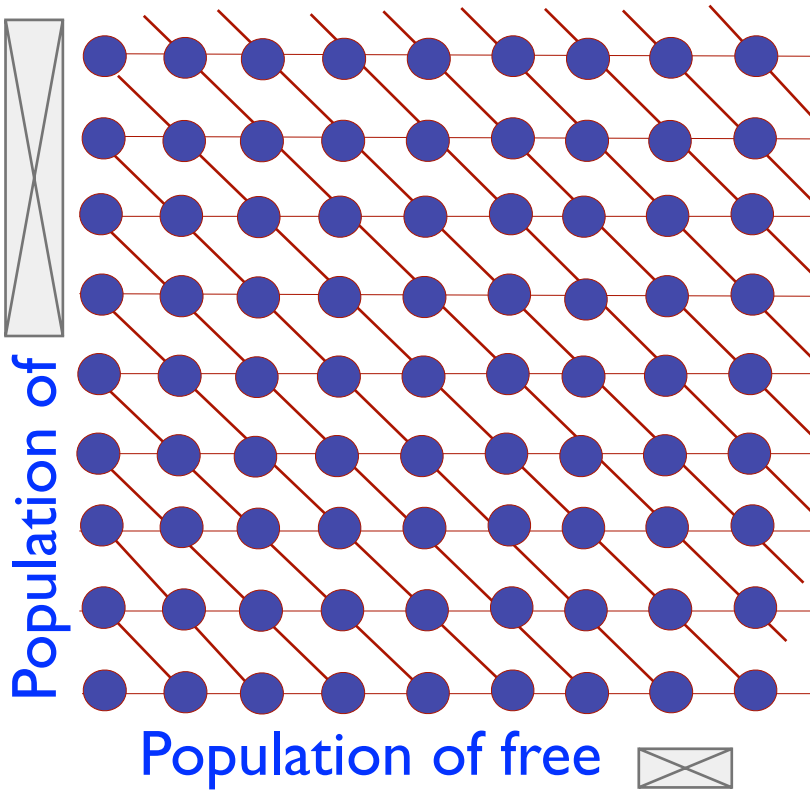


El Samad et al, *PNAS*, vol. 102, No. 8, 2005

Toy Heat Shock Model in *E. coli* (cont.)

Five Different FSP Solution Schemes:

1. Full FSP

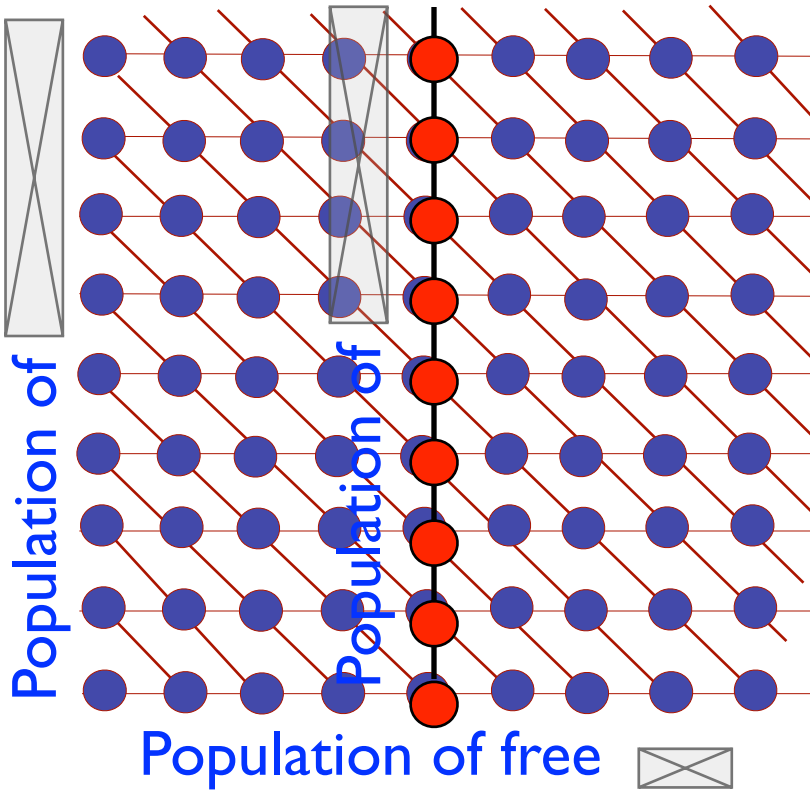


4459 ODEs

Toy Heat Shock Model in *E. coli* (cont.)

Five Different FSP Solution Schemes:

1. Full FSP
2. Slow manifold (FSP-SM)

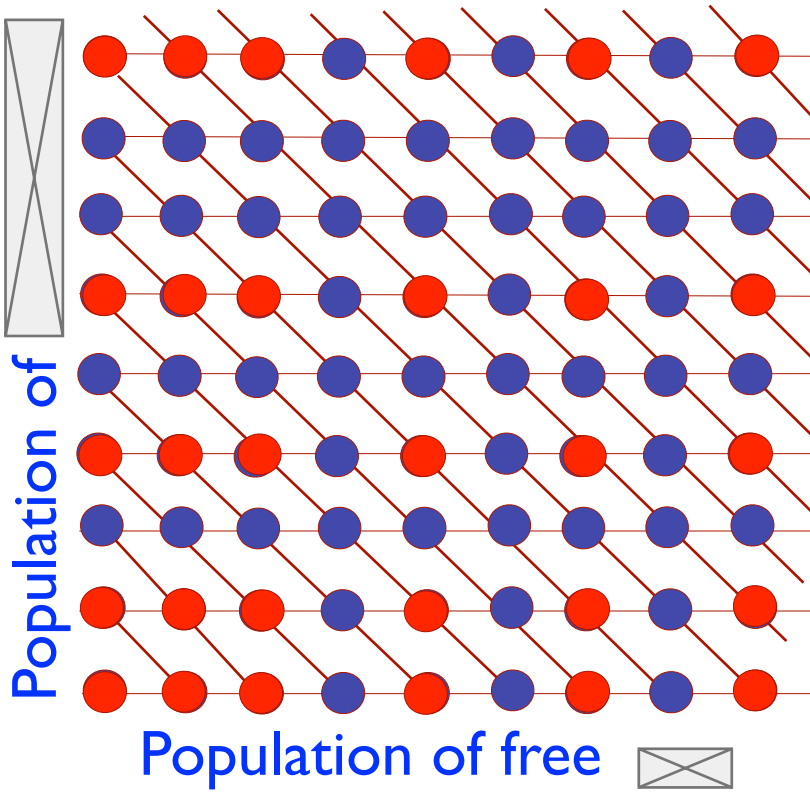


343 ODEs

Toy Heat Shock Model in *E. coli* (cont.)

Five Different FSP Solution Schemes:

1. Full FSP
2. Slow manifold (FSP-SM)
3. Interpolated (FSP-I)

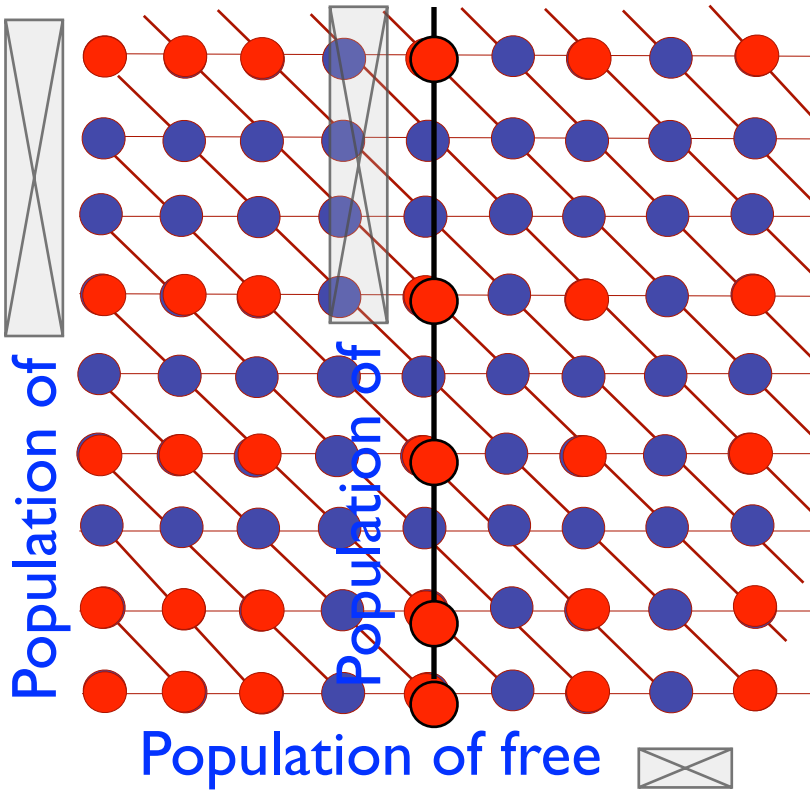


539 ODEs

Toy Heat Shock Model in *E. coli* (cont.)

Five Different FSP Solution Schemes:

1. Full FSP
2. Slow manifold (FSP-SM)
3. Interpolated (FSP-I)
4. Hybrid (FSP-SM/I)

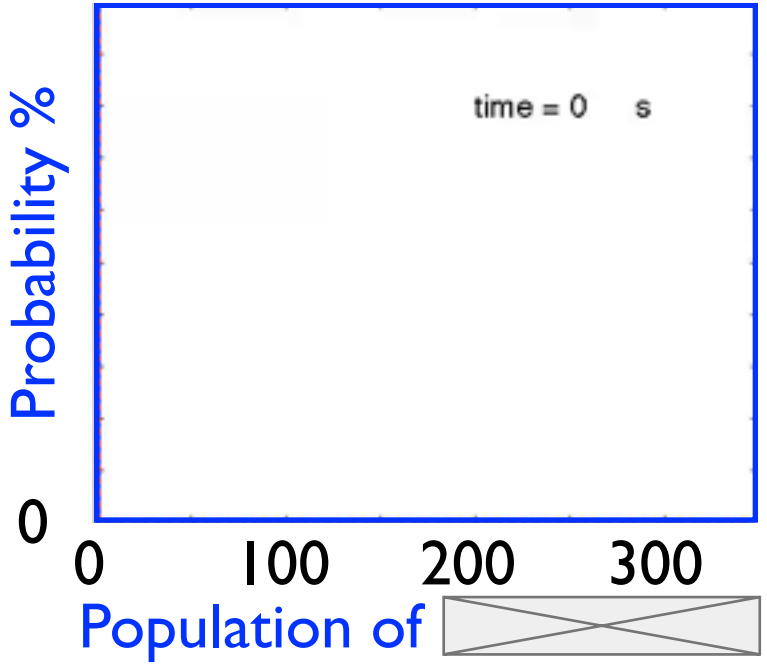


49 ODEs

Toy Heat Shock Model in *E. coli* (cont.)

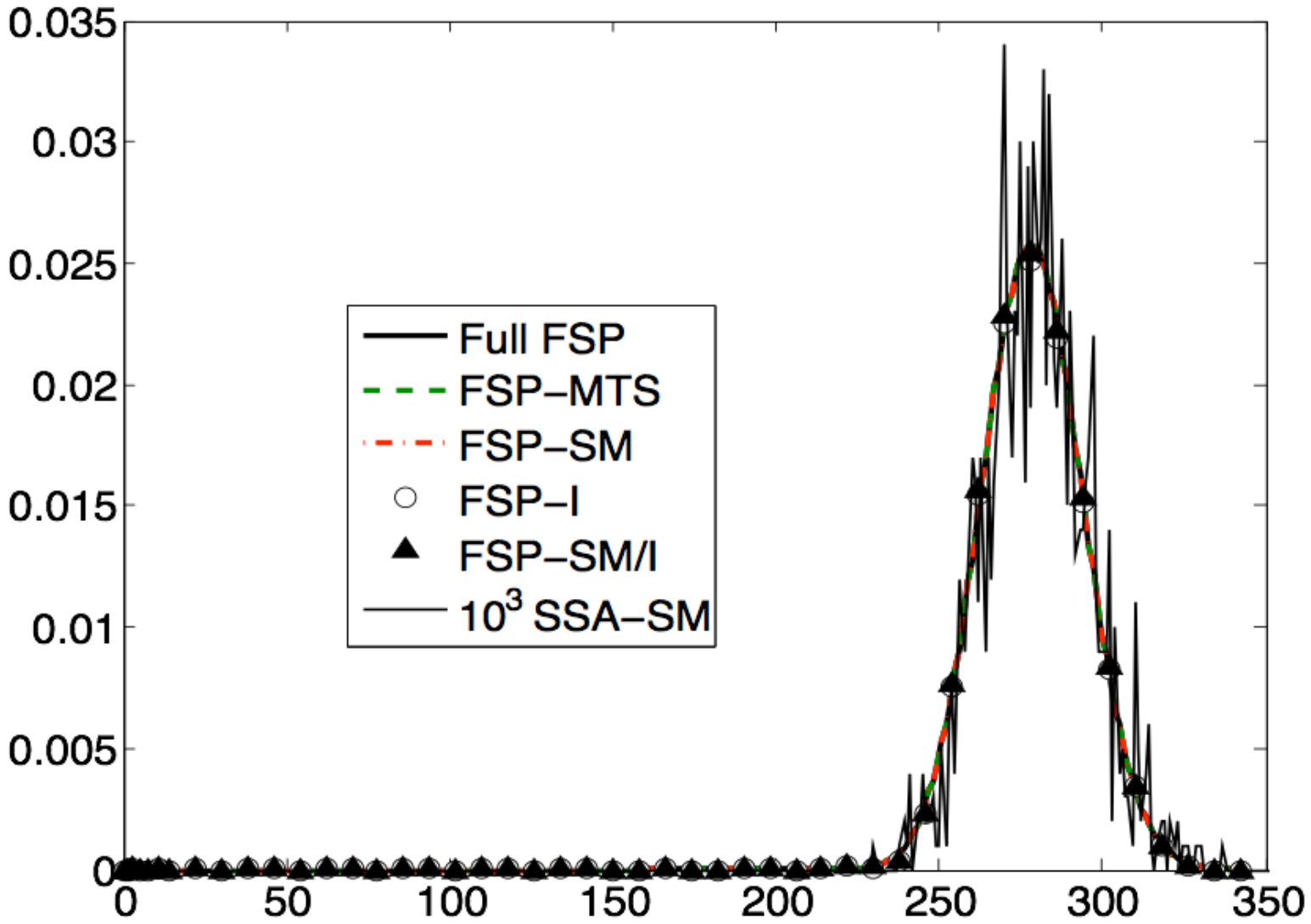
Five Different FSP Solution Schemes:

1. Full FSP
2. Slow manifold (FSP-SM)
3. Interpolated (FSP-I)
4. Hybrid (FSP-SM/I)
5. Multiple time interval (FSP-MTI)



70 sets of 195 or fewer ODEs.

Efficiency and accuracy of the reduced FSP methods



Efficiency and accuracy of the reduced FSP methods

For final time $t_f = 300s$				
Method	Matrix Size	J_{solve}	J_{total}	∞-norm Error
FSP	4459	750s	750s	$< 3.0 \times 10^{-5}$
FSP-MTS	195[1]	-	40.2s	$< 1.68 \times 10^{-4}$
FSP-SM	343	0.25s	0.94s	$\approx 5.1 \times 10^{-4}$
FSP-I	539	5.1s	6.1s	$\approx 7.7 \times 10^{-4}$
FSP-SM/I	49	0.04s	0.78s	$\approx 8.2 \times 10^{-4}$
10^4 SSA	Results would take more than 55 hours.			
10^3 SSA-SM	-	-	84.1s	≈ 0.0116
10^4 SSA-SM	-	-	925s	$\approx 3.4 \times 10^{-3}$
10^5 SSA-SM	-	-	9360s	$\approx 1.6 \times 10^{-3}$

The Reduced FSP approaches can be much faster and more accurate than alternative approaches!

The (Chemical) Master Equation

- The CME Description
- Example: Transcription as a Birth-Death Process.
- Kinetic Monte Carlo Approaches
- Finite State Projection Approaches
- Moment Computations

See notes online

Moment Computations

For the first moment $E[X_i]$, multiply the CME by x_i and sum over all $(x_1, \ldots, x_N) \in \mathbb{N}^N$

$$\frac{dE[X_i]}{dt} = \sum_{k=1}^{M} s_{ik} E[w_k(X)]$$

For the second moment $E[X_i X_j]$, multiply the CME by $x_i x_j$ and sum over all $(x_1, \ldots, x_N) \in \mathbb{N}^N$

$$\frac{dE[X_i X_j]}{dt} = \sum_{k=1}^{M} (s_{ik} E[X_j w_k(X)] + E[X_i w_k(X)] s_{jk} + s_{ik} s_{jk} E[w_k(X)])$$

Let $w(x) = [w_1(x), \ldots, w_M(x)]^T$

In matrix notation:

$$\frac{dE[X]}{dt} = SE[w(X)]$$

$$\frac{dE[XX^T]}{dt} = SE[w(X)X^T] + E[w(X)X^T]^T S^T + S\{diagE[w(X)]\}S^T$$

Moment Computations

- Affine Propensity
- Moment Closures

Affine Propensity

Suppose the propensity function is affine:

$$w(x) = Wx + w_0, \qquad (W \text{ is } N \times N,\ w_0 \text{ is } N \times 1)$$

Then $E[w(X)] = WE[X] + w_0$, and $E[w(X)X^T] = WE[XX^T] + w_0 E[X^T]$.

This gives us the moment equations:

$$\begin{aligned} \frac{d}{dt}E[X] &= SWE[X] + Sw_0 && \text{First Moment} \\ \frac{d}{dt}E[XX^T] &= SWE[XX^T] + E[XX^T]W^T S^T + S\ diag(WE[X] + w_0)S^T \\ &+ Sw_0 E[X^T] + E[X]w_0^T S^T && \text{Second Moment} \end{aligned}$$

These are linear ordinary differential equations and can be easily solved!

Affine Propensity (cont.)

Define the covariance matrix $\Sigma = E[(X - E[X])(X - E(X)]^T]$.
We can also compute mean and covariance equations:

$$\frac{d}{dt}E[X] = SWE[X] + Sw_0 \qquad \text{First Moment}$$

$$\frac{d}{dt}\Sigma = SW\Sigma + \Sigma W^T S^T + S\ diag(WE[X] + w_0)S^T \qquad \text{Covariance}$$

Steady-state Case

The steady-state moments and covariances can be obtained by solving linear algebraic equations:

Let $\bar{X} = \lim_{t\to\infty} E[X(t)]$ and $\bar{\Sigma} = \lim_{t\to\infty} \Sigma(t)$. Then

$$SW\bar{X} = -Sw_0$$

$$SW\bar{\Sigma} + \bar{\Sigma}W^T S^T + S\ diag(W\bar{X} + w_0)S^T = 0$$

Affine Propensity (cont.)

Define $A = SW$, and $B = S\sqrt{diag(W\bar{X} + w_0)}$.
The steady-state covariances equation

$$SW\bar{\Sigma} + \bar{\Sigma}W^T S^T + S\ diag(W\bar{X} + w_0)S^T = 0$$

becomes

$$A\bar{\Sigma} + \bar{\Sigma}A^T + BB^T = 0 \qquad \text{Lyapunov Equation}$$

The Lyapunov equation characterizes the steady-state covariance of a output of the linear dynamical system

$$\dot{y} = Ay + B\omega$$

where ω is a unit intensity white Gaussian noise!

More precisely, the solution of the vector SDE:

$$dy = Ay\ dt + B\ dW_t$$

where W_t is Brownian motion. This is also called **Ornstein-Uhlenbeck process**.

Moment Computations

- Affine Propensity
- Moment Closures

Moment Closures

From before, the mean level changes as:

$$\frac{dE[X]}{dt} = SE[w(X)]$$

- When Second and Higher order terms exist in the propensity functions, each moment depends upon higher moments.
  - For example, if $w(X) = \mathbf{u}^T \mathbf{X}\mathbf{X}^T \mathbf{v}$, then

$$\frac{dE[\mathbf{X}]}{dt} = S\mathbf{u}^T E[\mathbf{X}\mathbf{X}^T]\mathbf{v}$$

- The first moment depends upon the second; the second upon the third; and so on…
- "Moment closures" are approximations that attempt to remove this infinite dependency structure.

Moment Closures

$$\frac{dE[X_i]}{dt} = \sum_{k=1}^{M} s_{ik} E[w_k(X)]$$

$$\frac{dE[X_i X_j]}{dt} = \sum_{k=1}^{M} (s_{ik} E[X_j w_k(X)] + E[X_i w_k(X)] s_{jk} + s_{ik} s_{jk} E[w_k(X)])$$

$$\frac{d}{dt}\begin{bmatrix} \{\mu_i\} \\ \{\sigma_{ij}\} \end{bmatrix} = \begin{bmatrix} f_1(\{\mu_i\},\{\sigma_{ij}\}) + u_1(\{\mu_i\},\{\sigma_{ij}\},\{\sigma_{ijk}\},\ldots) \\ f_2(\{\mu_i\},\{\sigma_{ij}\}) + u_2(\{\mu_i\},\{\sigma_{ij}\},\{\sigma_{ijk}\},\ldots) \end{bmatrix},$$

$$\frac{d}{dt}\begin{bmatrix} \{\mu_i\} \\ \{\sigma_{ij}\} \end{bmatrix} = \begin{bmatrix} f_1(\{\mu_i\},\{\sigma_{ij}\}) + \hat{u}_1(\{\mu_i\},\{\sigma_{ij}\}) \\ f_2(\{\mu_i\},\{\sigma_{ij}\}) + \hat{u}_2(\{\mu_i\},\{\sigma_{ij}\}) \end{bmatrix},$$

where the choice of $\hat{u}_1$ and $\hat{u}_2$
depends upon the chosen moment closure.

Gaussian Moment Closure

- For Gaussian distributions, the closure is simple:

$$\sigma_{ijk} = \mathbb{E}\{(X_i - \mathbb{E}\{X_i\})(X_j - \mathbb{E}\{X_j\})(X_k - \mathbb{E}\{X_k\})\} = 0$$

- which yields:

$$\mathbb{E}\{(X_i X_j X_k\} = -\mathbb{E}\{X_i X_j\}\mathbb{E}\{X_k\} - \mathbb{E}\{X_j X_k\}\mathbb{E}\{X_i\} - \mathbb{E}\{X_k X_i\}\mathbb{E}\{X_j\} + 2\mathbb{E}\{X_i\}\mathbb{E}\{X_j\}\mathbb{E}\{X_k\}$$

- Higher moments are easy to derive with a moment generating function:

$$M_{\mathbf{x}}(\mathbf{t}) = \exp\left(\mu^T \mathbf{t} + 1/2 \mathbf{t}^T \mathbf{\Sigma} \mathbf{t}\right),$$

$$\mathbb{E}\{x_1^{n_1} \dots x_4^{n_4}\} = \frac{d^{n_1+\dots+n_4}}{dx_1^{n_1} \dots dx_4^{n_4}} M_x(\mathbf{t})\Big|_{\mathbf{t}=\mathbf{0}}.$$

Many other closures are possible:

- If one assumes that the distributions are Log-Normal, a different closure is used:

$$\mathbb{E}[X_i X_j X_k] = \frac{\mathbb{E}[X_i X_j]\mathbb{E}[X_j X_k]\mathbb{E}[X_i X_k]}{\mathbb{E}[X_i]\mathbb{E}[X_j]\mathbb{E}[X_k]}.$$

- One of the most common closures is the Linear Noise Approximation.
- In this, all moments are written in terms of themselves and lower moments:
  - the mean is set equal to the deterministic process.
  - the second moments are assumed to be Gaussian, and depend upon the mean and themselves:

$$\frac{d}{dt}\begin{bmatrix} \{\mu_i\} \\ \{\sigma_{ij}\} \end{bmatrix} = \begin{bmatrix} f_1(\{\mu_i\}) \\ f_2(\{\mu_i\},\{\sigma_{ij}\}) \end{bmatrix}$$

A 5 minute break, and when we return...

Colorado State University

Integrating Single-Cell Experiments and Stochastic Analyses to Predict Gene Expression Dynamics

Brian Munsky
Chemical and Biological Engineering
Colorado State University

munsky@engr.colostate.edu

Outline

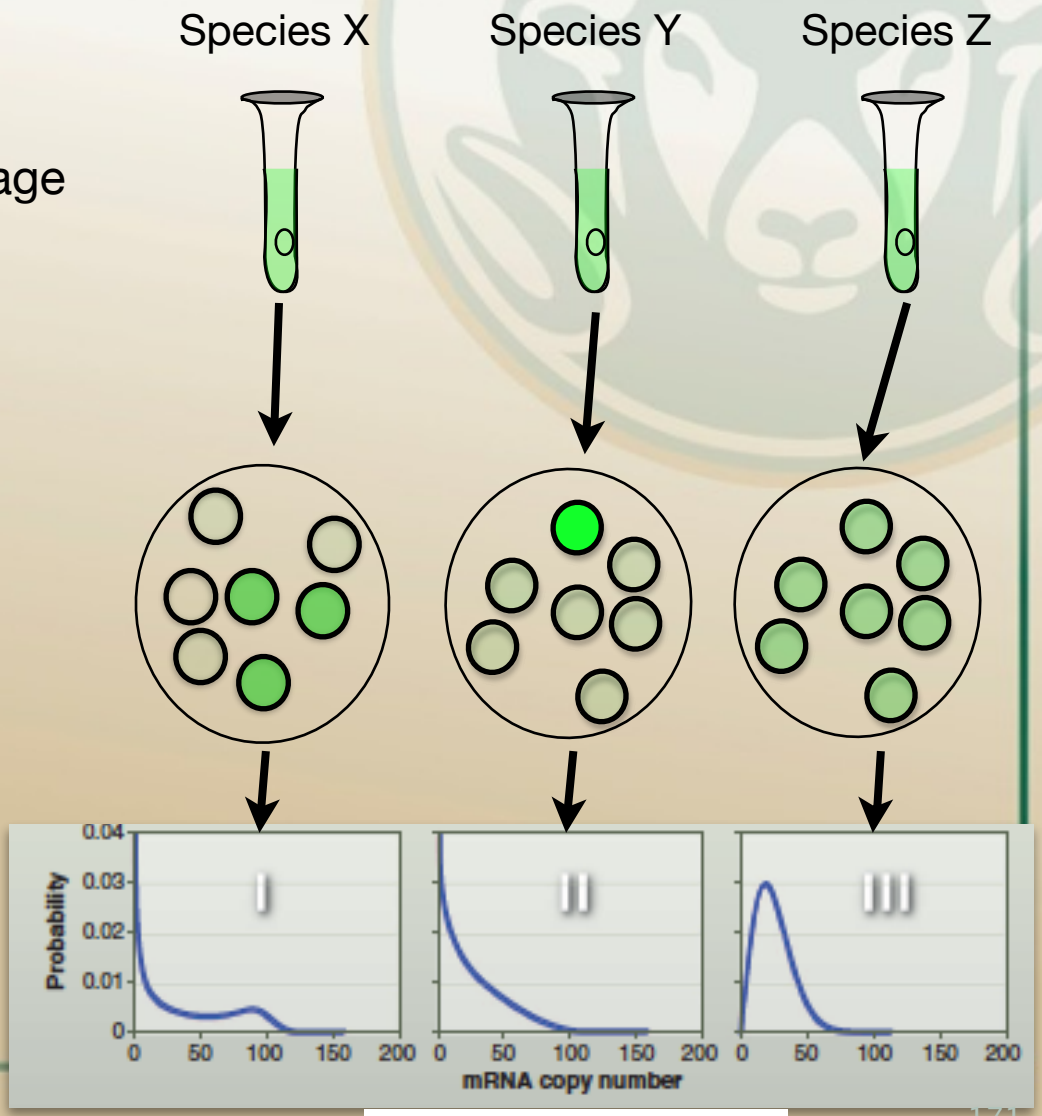
1. **Information from single-cell fluctuation**
2. Analyzing stochastic dynamics in gene regulation
3. Case studies:
    a) Predicting kinase-activated gene regulation dynamics in Saccharomyces cerevisiae (budding yeast).
    b) Predicting multi-generation stochastic behavior of the Pap epigenetic switch in E. coli
    c) Predictable design of synthetic circuits in E. coli
    d) sRNA regulation in Yersinia Pestis and Yersinia Pseudotuberculosis
    e) Examining multiscale spatiotemporal mRNA fluctuations in human THP1 cells
4. Concluding remarks

Information from fluctuation

Several different species may express a gene at the same average level.

Single-cell measurements may reveal hidden differences in the species.

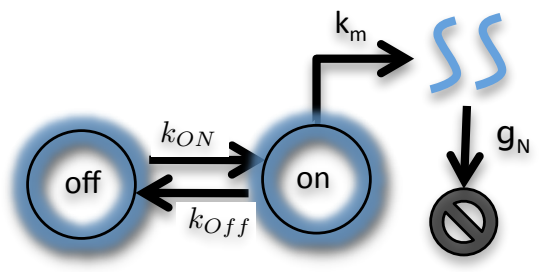
Each species has a distinctive "fluctuation fingerprint".

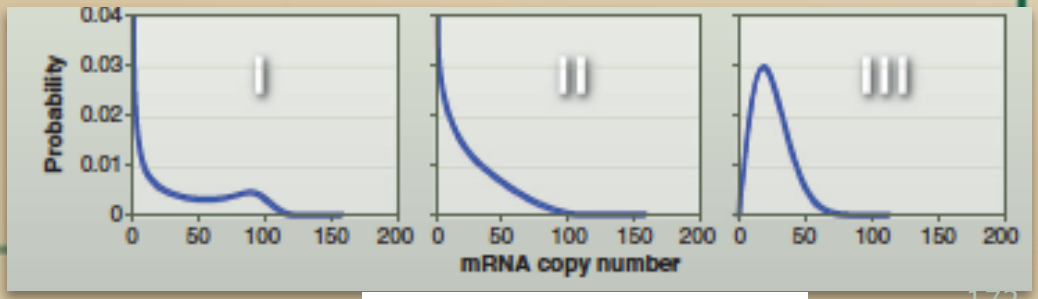


Munsky et al, *Science* 2012

Fluctuations may indicate gene regulation mechanisms

- Consider the bursting gene expression model:

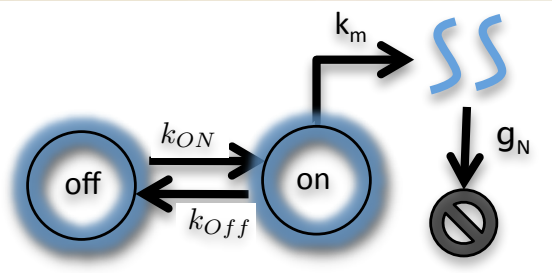




Munsky et al, *Science* 2012

Fluctuations may indicate gene regulation mechanisms

- Consider the bursting gene expression model:



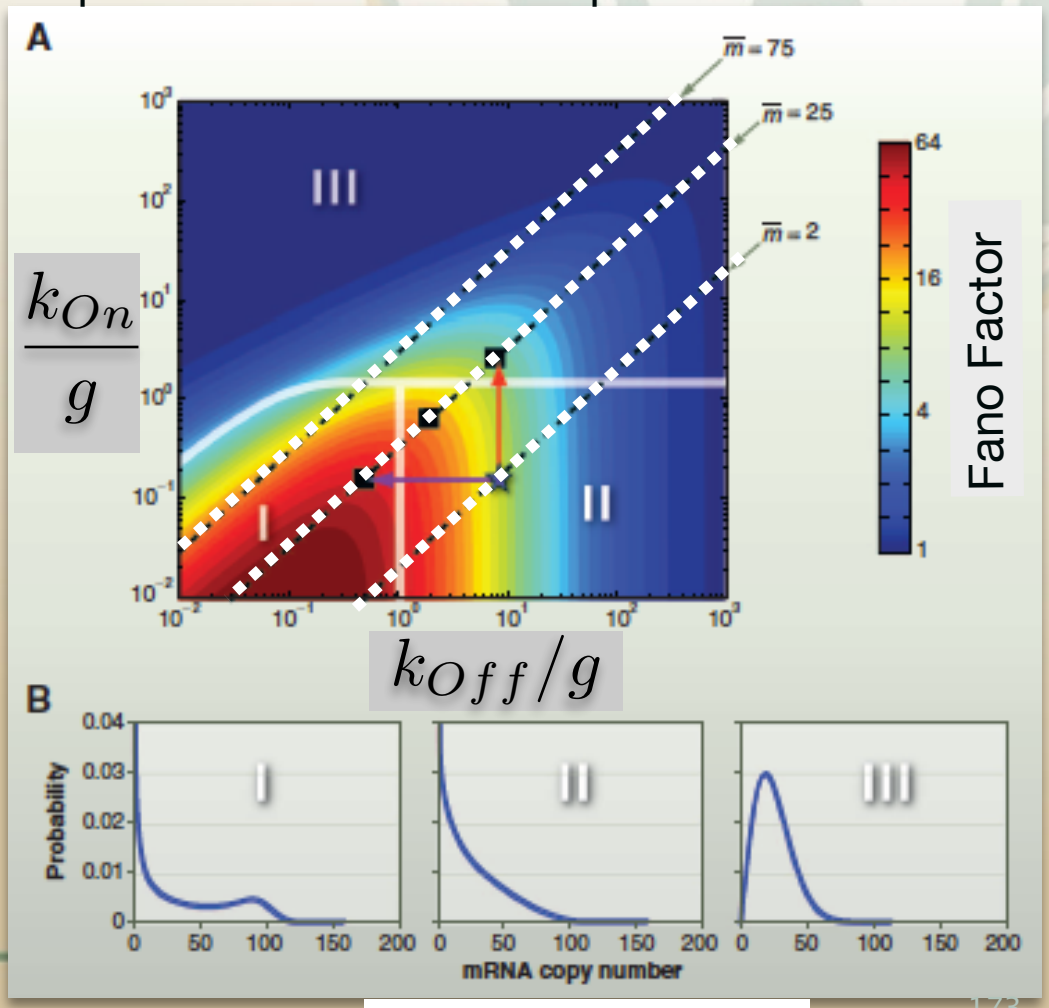
- Compute the expression mean and variability as functions of all parameters.

$$f_{on} = \frac{k_{ON}}{k_{ON} + k_{OFF}}$$

$$\mu = f_{on}\frac{k_m}{g_m}$$

$$\frac{\sigma^2}{\mu} = 1 + \frac{(1 - f_{on})\, k_m}{k_{ON} + k_{OFF} + g_m}$$

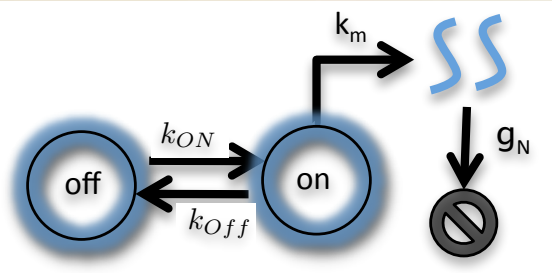
Expression 'Noise' versus parameters



Munsky et al, *Science* 2012

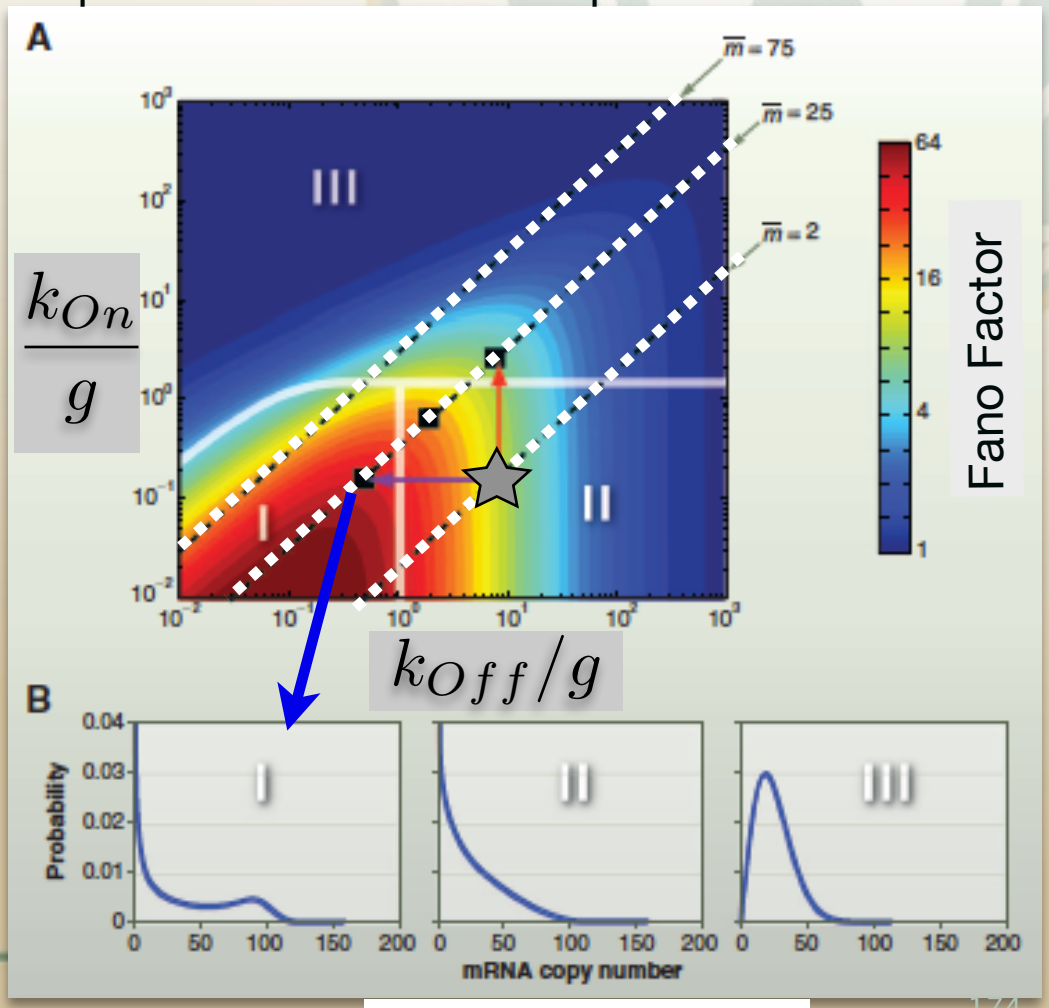
Fluctuations may indicate gene regulation mechanisms

- Consider the bursting gene expression model:



- Compute the expression mean and variability as functions of all parameters.
- Tuning k_{Off} or k_{On} can increase expression, but:
- **Tuning k_{off} increases variability.**

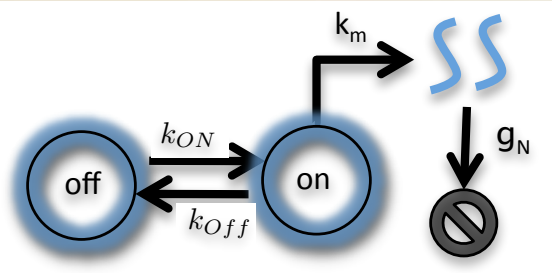
Expression 'Noise' versus parameters



Munsky et al, *Science* 2012

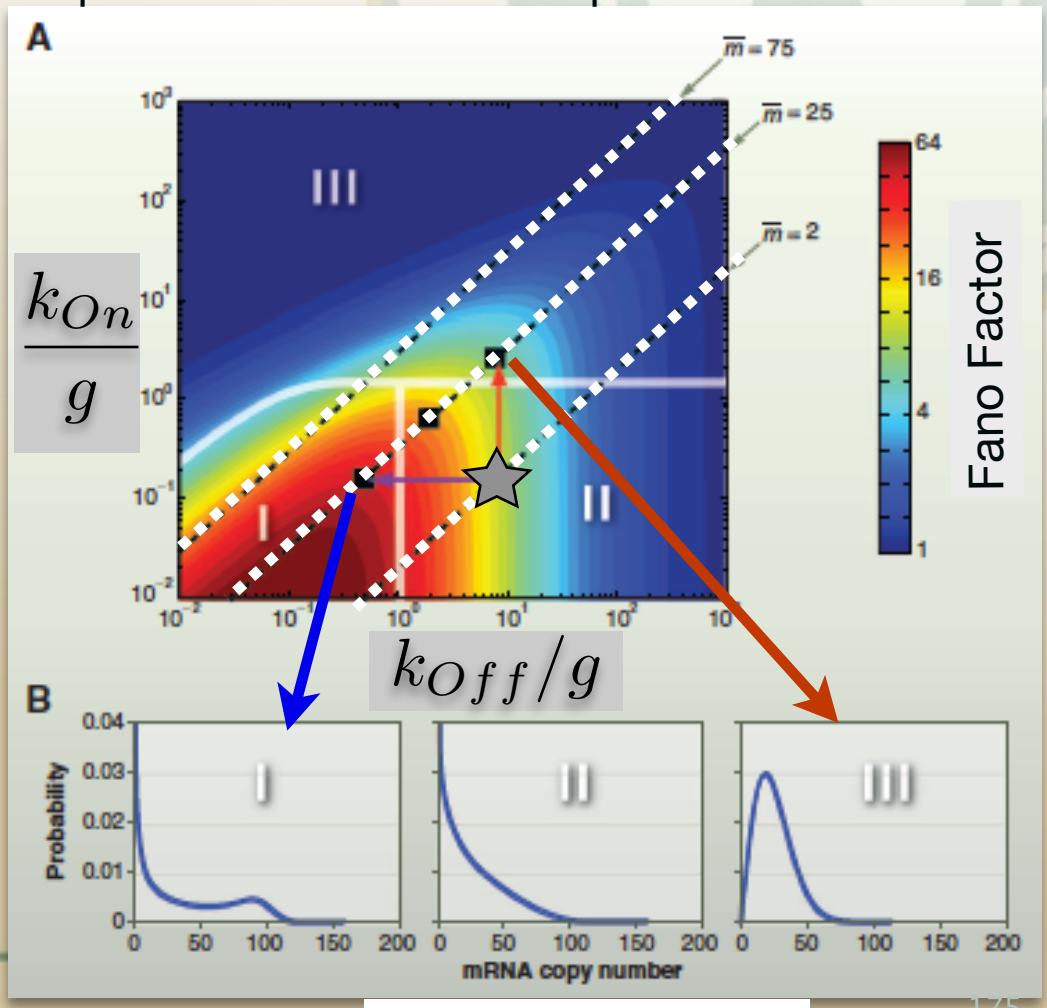
Fluctuations may indicate gene regulation mechanisms

- Consider the bursting gene expression model:



- Compute the expression mean and variability as functions of all parameters.
- Tuning k_{Off} or k_{On} can increase expression, but:
- **Tuning k_{Off} increases variability.**
- **Tuning k_{On} decreases variability.**

Expression 'Noise' versus parameters



Munsky et al, *Science* 2012

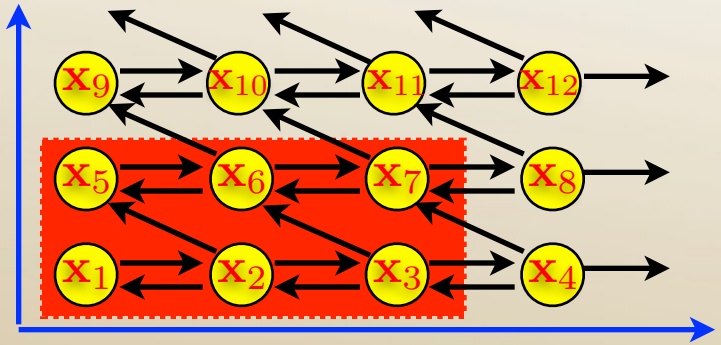
Outline

1. Information from single-cell fluctuation
2. **Analyzing stochastic dynamics in gene regulation**

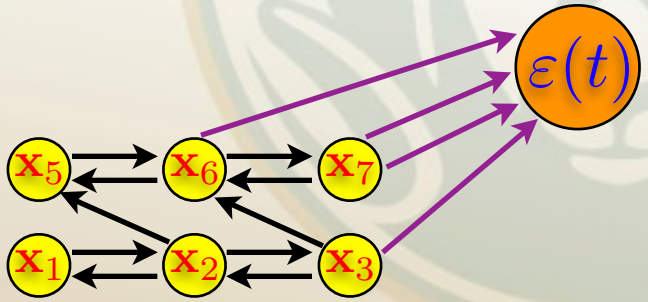
The finite state projection approach

Colorado State University

The Full System



The Projected System (FSP)



Full Master Equation

$$\begin{bmatrix} \dot{\mathbf{P}}_J \\ \dot{\mathbf{P}}_{J'} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_J & \mathbf{A}_{JJ'} \\ \mathbf{A}_{J'J} & \mathbf{A}_{J'} \end{bmatrix} \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'}(t) \end{bmatrix}$$

Dimension = $\#(J) + \#(J')$ Infinite

FSP Master Equation

$$\begin{bmatrix} \dot{\mathbf{P}}_J^{FSP} \\ \dot{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_J & \mathbf{0} \\ -\mathbf{1}^T\mathbf{A}_J & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \varepsilon(t) \end{bmatrix}$$

Dimension = $\#(J) + 1$ = 7

The FSP Theorem

(Munsky, JCP '06)

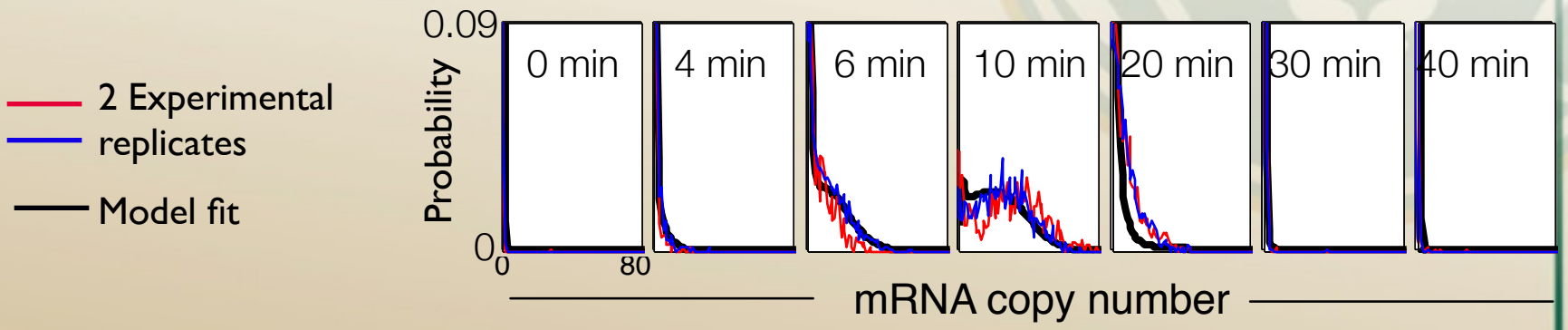
$$\mathbf{P}_J(t) \geq \mathbf{P}_J^{FSP}(t) \text{ and}$$

$$\left\| \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'} \end{bmatrix} - \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \mathbf{0} \end{bmatrix} \right\|_1 = \varepsilon(t)$$

Download software and tutorial available at: http://cnls.lanl.gov/~munsky

Inferring parameters from single-cell measurements.

Although single-cell reactions may be ***Stochastic***, their statistics follow a ***Deterministic*** set of ODE's *(i.e., the CME).*



We can fit and potentially predict these statistics.

Fitting metrics:

High cell counts (>10^5 cells) --> Kullback Leibler Divergence.*

Low cell counts (<10^3 cells) --> maximum likelihood.*

**Equivalent up to a constant that depends upon sample sizes.*

Outline

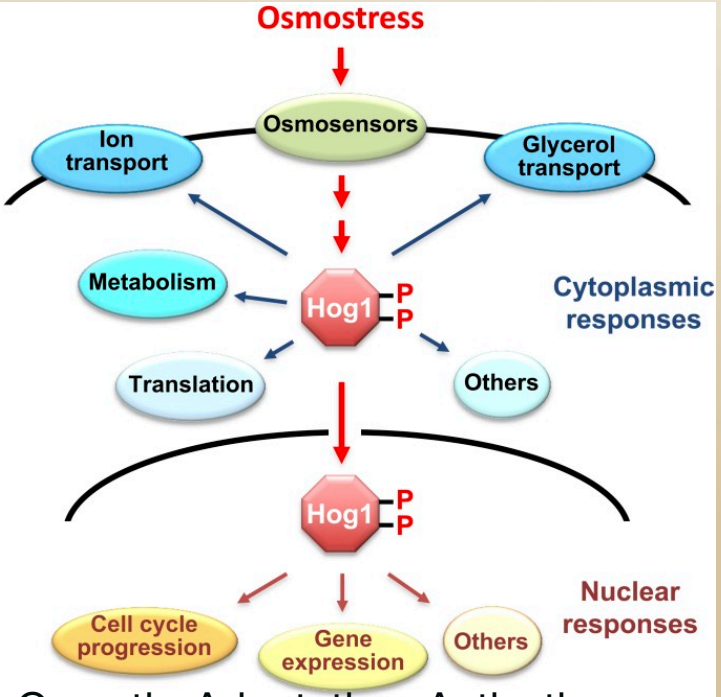
1. Information from single-cell fluctuation
2. Analyzing stochastic dynamics in gene regulation
3. **Case studies:**
    a. **Predicting kinase-activated gene regulation dynamics in *Saccharomyces cerevisiae* (budding yeast).**
    b. Predicting multi-generation stochastic behavior of the Pap epigenetic switch in *E. coli*
    c. Predictable design of synthetic circuits in *E. coli*
    d. sRNA regulation in *Yersinia Pestis* and *Yersinia Pseudotuberculosis*

Signal-activated gene regulation

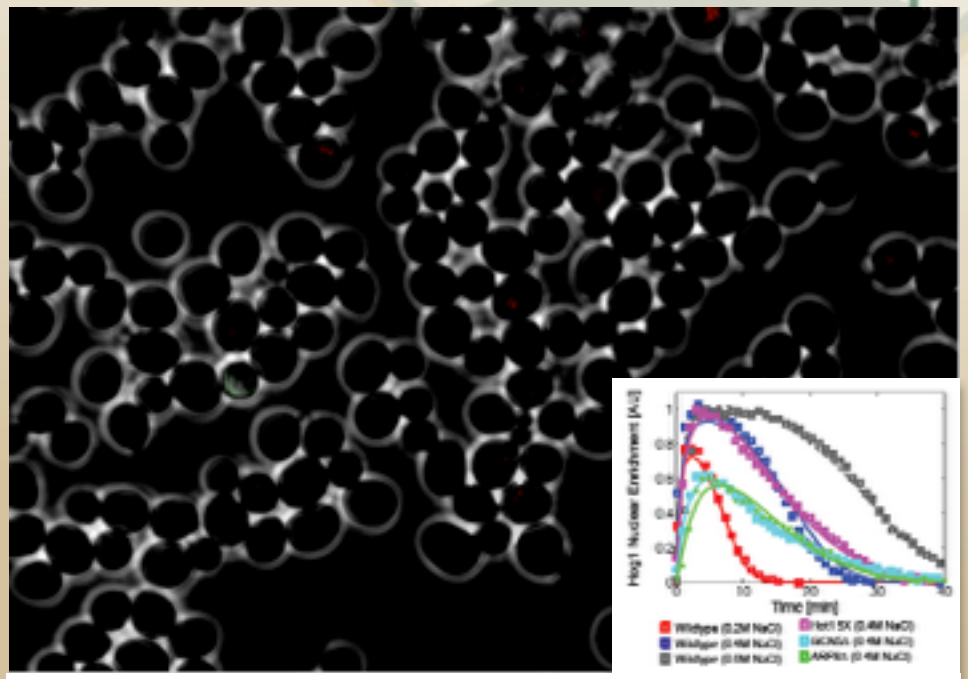
(Osmotic shock response in yeast)

Colorado State University

- 0.2M NaCl is added at t=0.
- Hog1 (red) activates in 1-2 min.
- ... and remains active for ~12 min.



Osmotic-Adaptation, Activation and Localization of Hog1p, (Saito, Posas, *Genetics*, 2012)



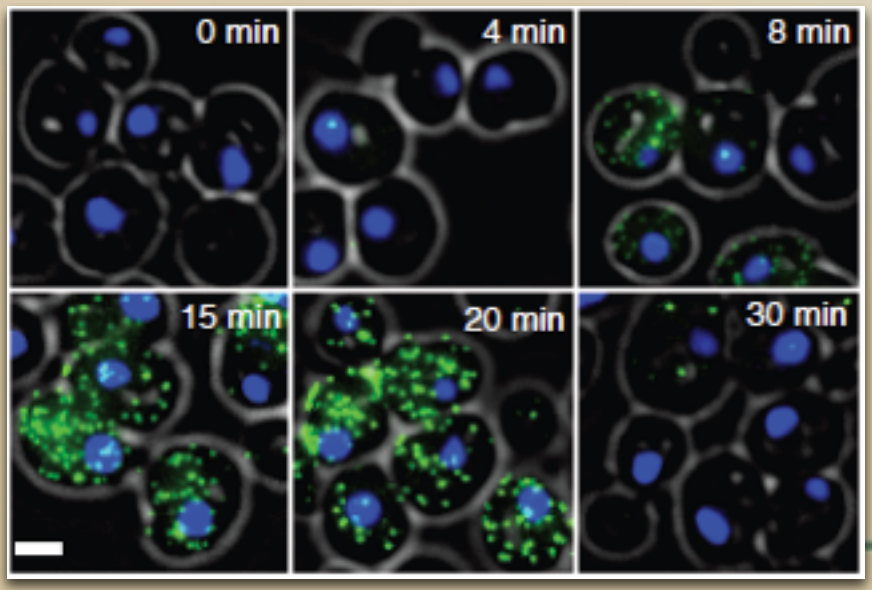
0-20 minutes after 0.2M NaCl shock (Neuert, Munsky, et al, 2013)

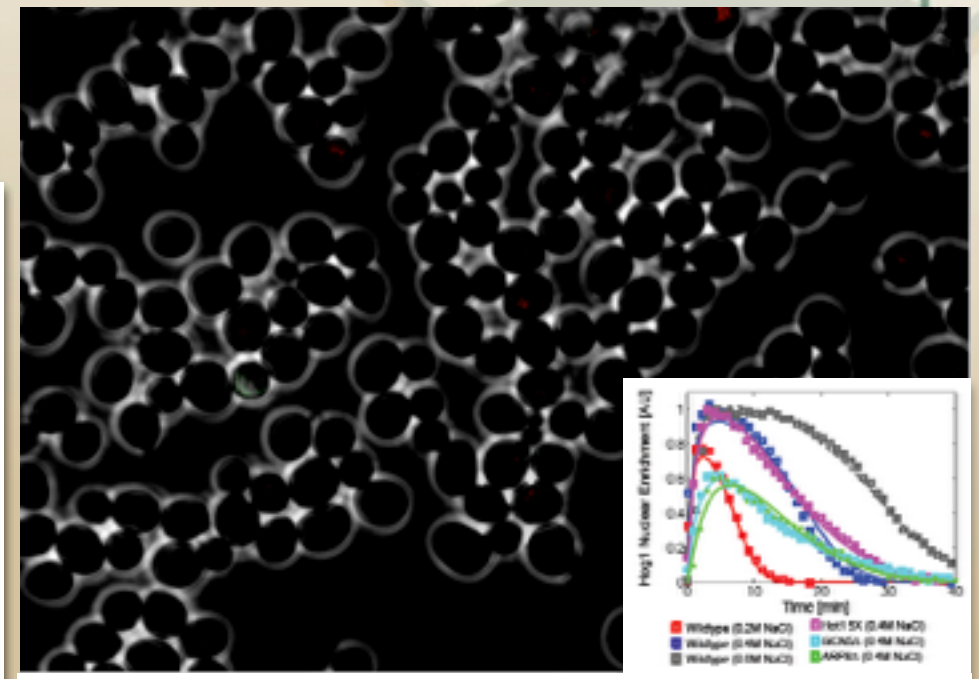
Signal-activated gene regulation

(Osmotic shock response in yeast)

Colorado State University

- 0.2M NaCl is added at t=0.
- Hog1 (red) activates in 1-2 min.
- ... and remains active for ~12 min.
- Stl1 mRNA appear at 4 min.
- ... and are gone by 25 min.





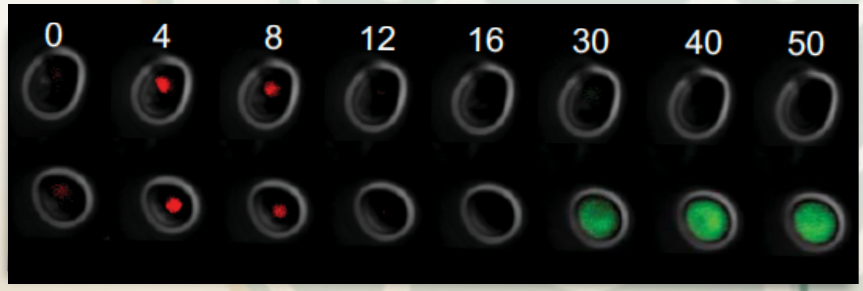
0-20 minutes after 0.2M NaCl shock
(Neuert, Munsky, et al, 2013)

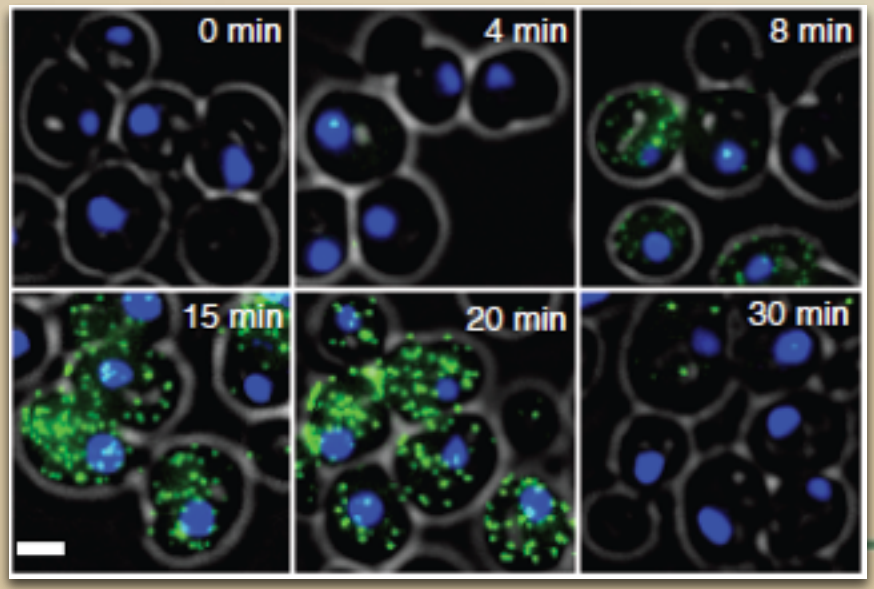
Signal-activated gene regulation

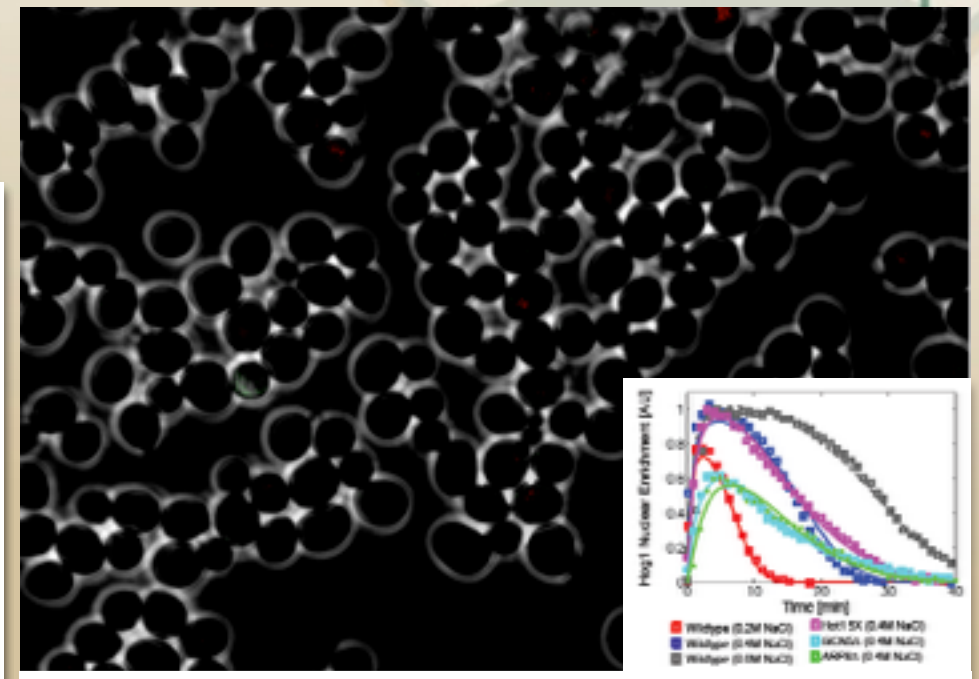
(Osmotic shock response in yeast)

Colorado State University

- 0.2M NaCl is added at t=0.
- Hog1 (red) activates in 1-2 min.
- ... and remains active for ~12 min.
- Stl1 mRNA appear at 4 min.
- ... and are gone by 25 min.
- Stl1-GFP appear at ~30 min.





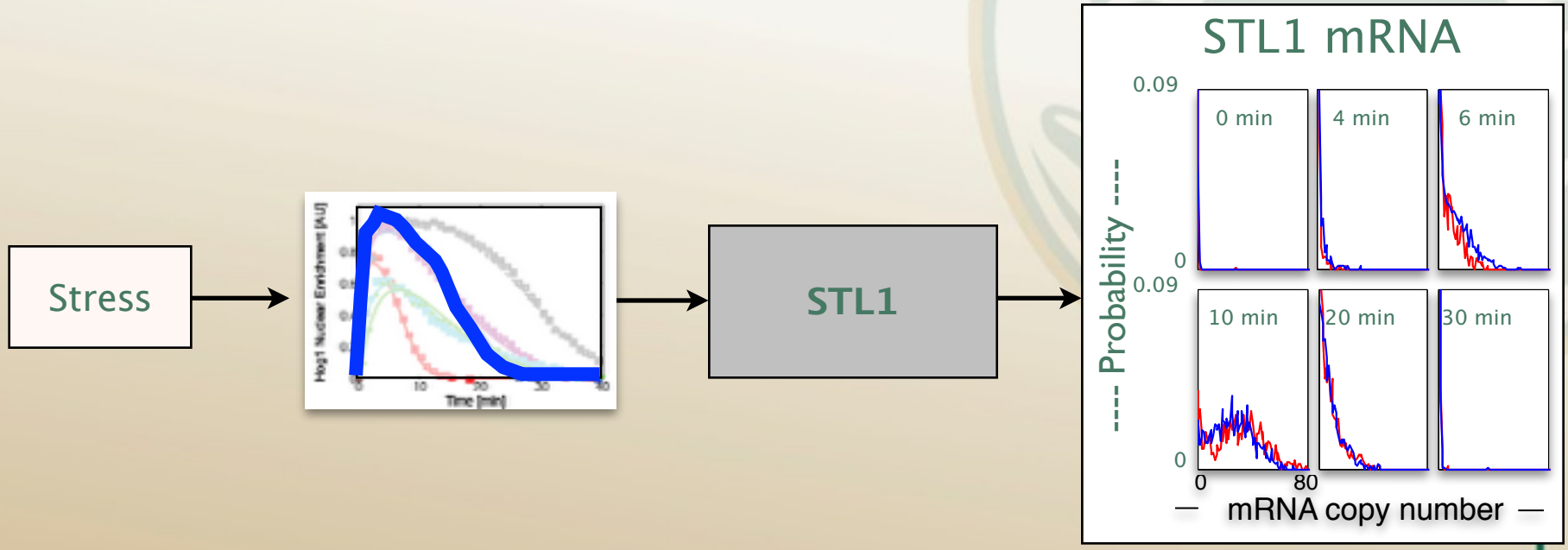


0-60 minutes after 0.2M NaCl shock
(Neuert, Munsky, et al, 2013)

Signal-activated gene regulation

(Osmotic shock response in yeast)

Colorado State University

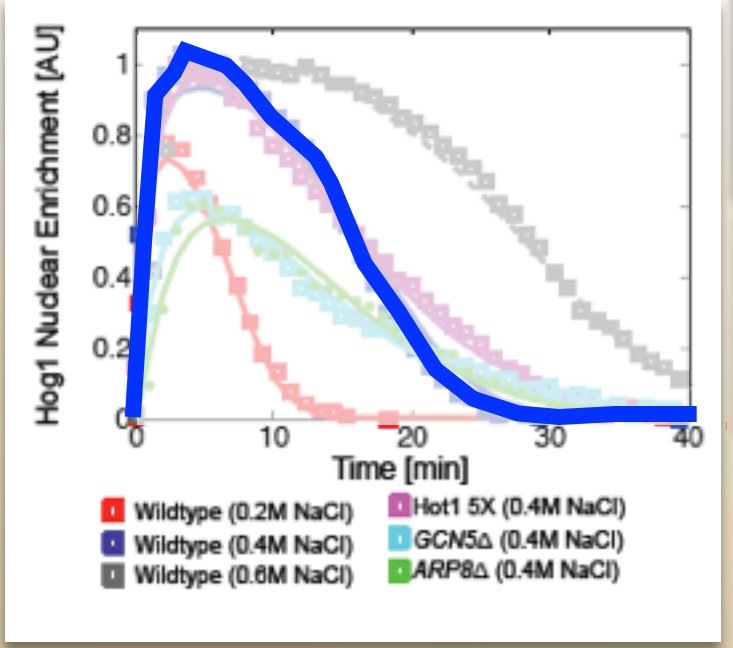


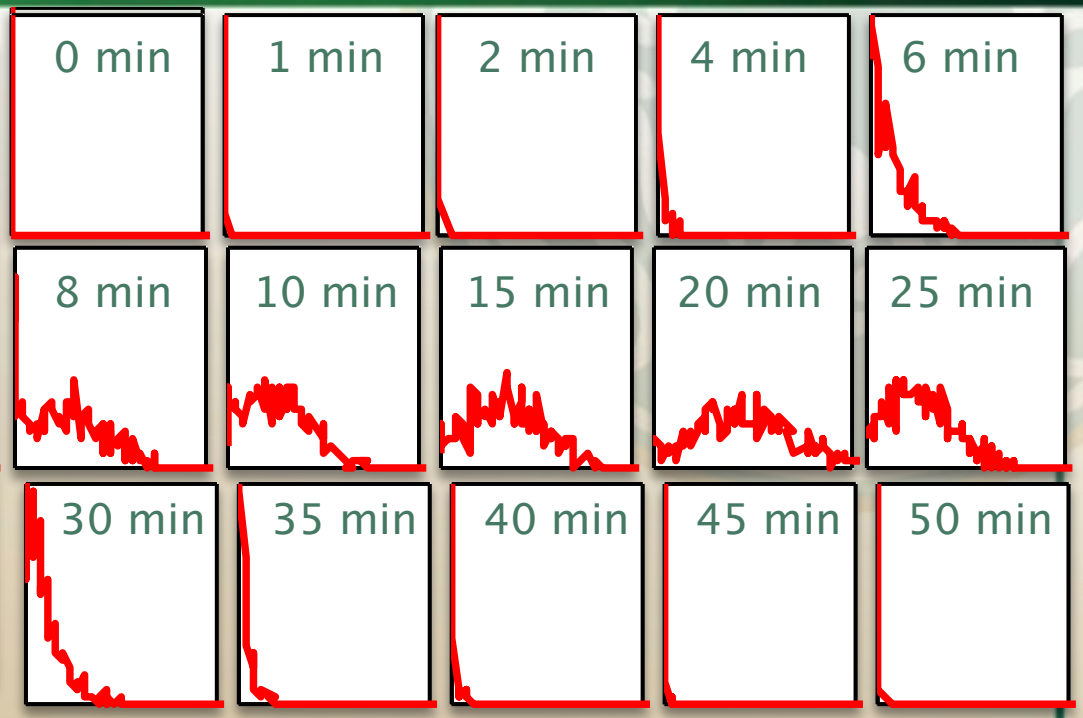
Our goal is to identify the mechanisms and parameters of STL1

Features of the data

(Osmotic shock response in yeast)

Colorado State University





At. 0.4M NaCl:

1) Hog1 localizes immediately (<2 min)
2) Hog1 remains active for 20 min.
3) Transcription starts in 2 min.
4) Cells activate at different times (bimodality at 8 min).
5) mRNA levels reach maximum at 20 min.
6) Most mRNA are cleared by 35 min.

Possible model structures:

The Hog1 kinase (activates STL1, but how?

Colorado State University

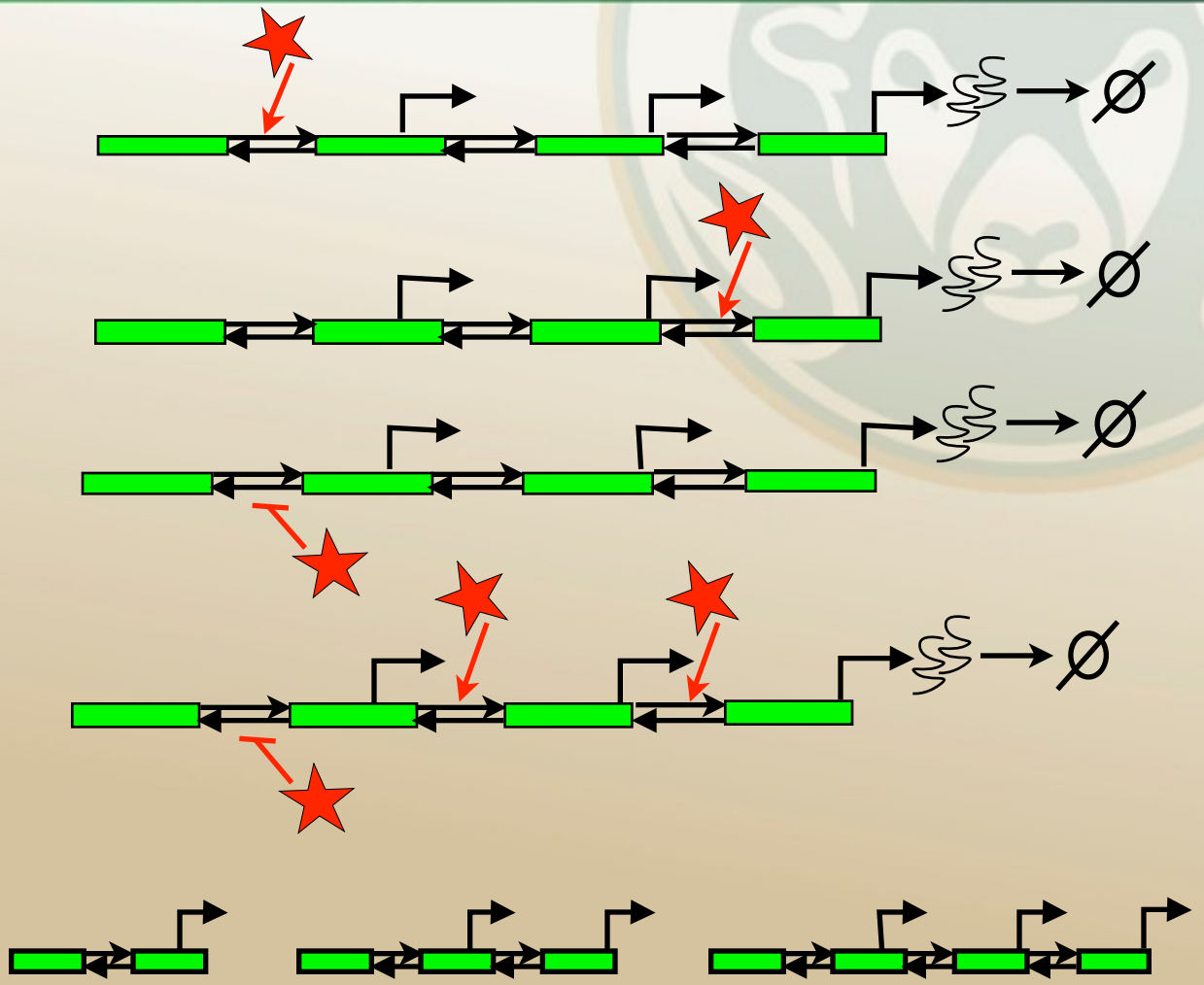
Is it the first of a cascade of activation events?

...the last activation event?

Does it repress a deactivation event?

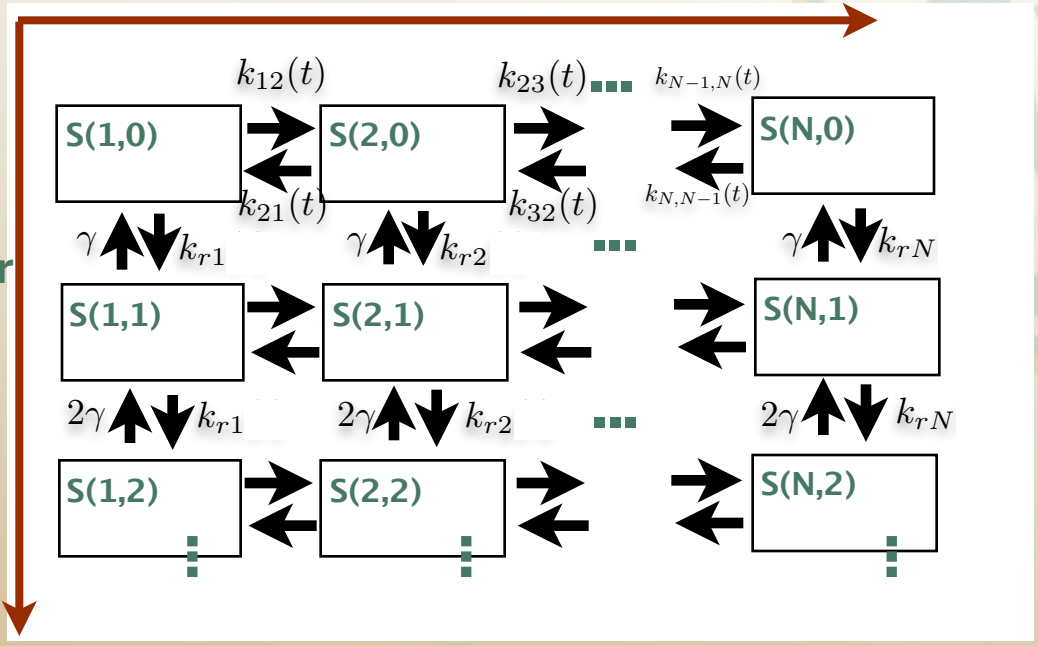
Are there multiple effects?

How many states are needed?



Each structure defines a hidden Markov Model

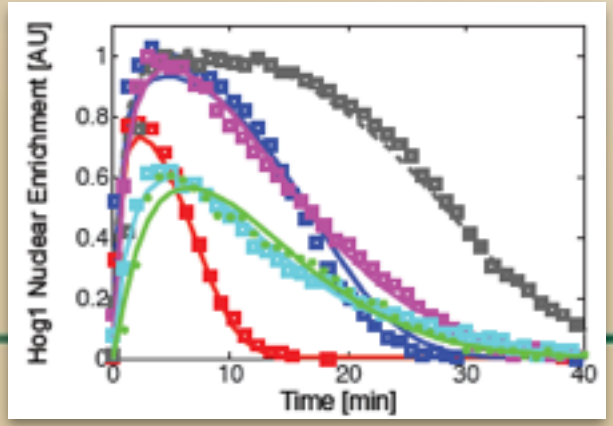
HIDDEN: N ={2,3,...} possible gene states



OBSERVABLE: Integer number of mRNA

State-transition rates may vary in time, with experimental conditions, and/or with genetic mutations.

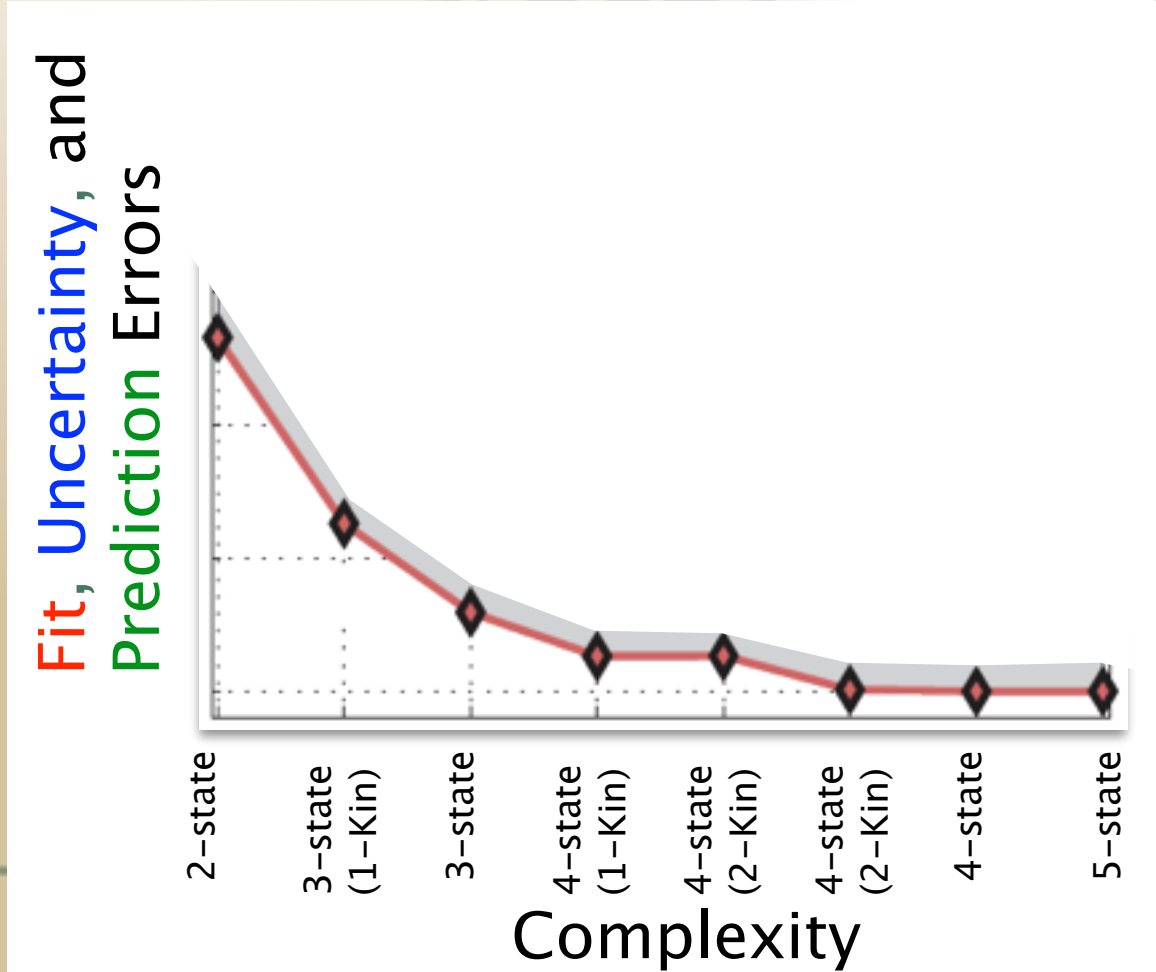
$$k_{ij} = k_{ij}(\text{Hog1}) = k_{ij}(t)$$



Evaluating model structures of varying complexity

We fit different 2-, 3-, 4- and 5- state model structures to wild-type data at 0.4M osmotic shock.

More states (and parameters) yield better fits,...

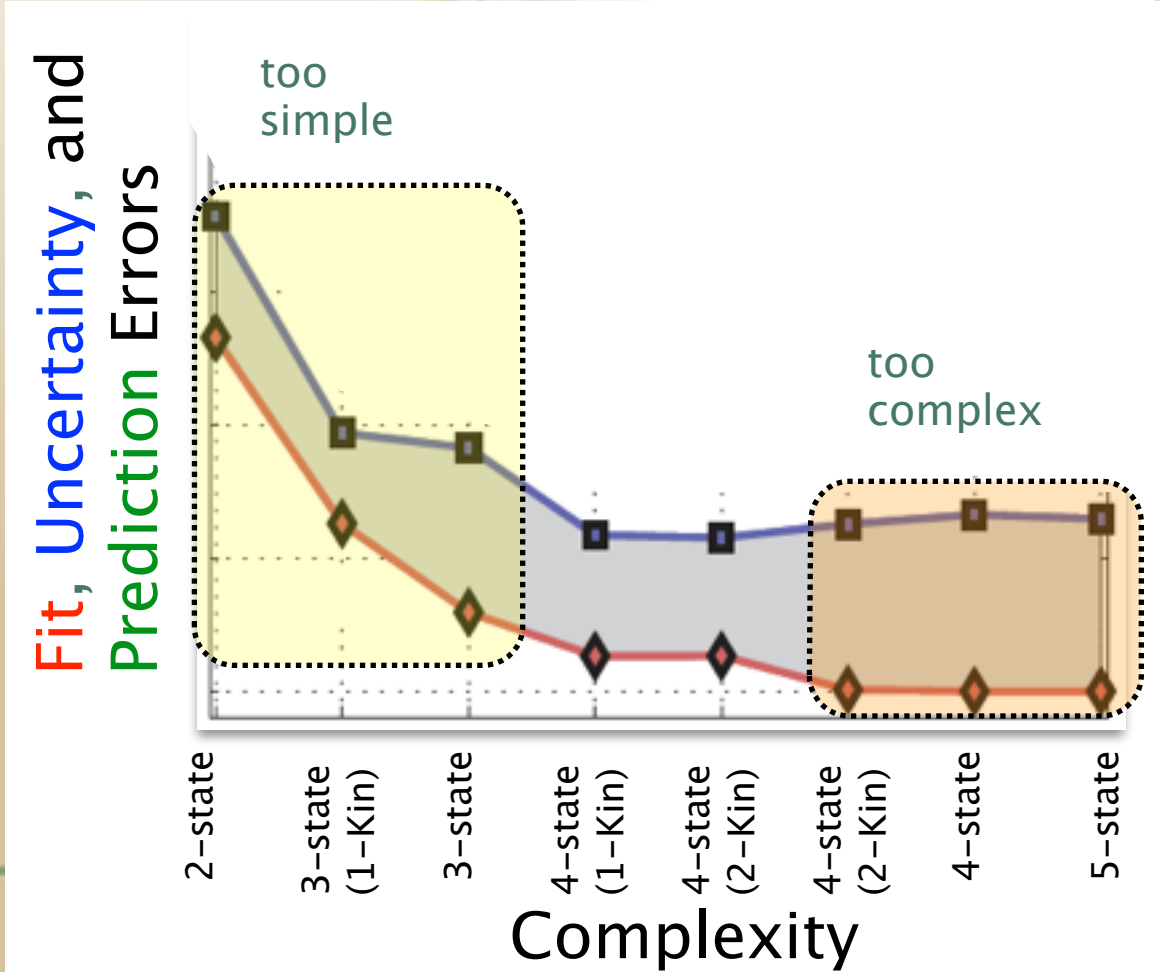


Evaluating model structures of varying complexity

We fit different 2-, 3-, 4- and 5- state model structures to wild-type data at 0.4M osmotic shock.

More states (and parameters) yield better fits,...

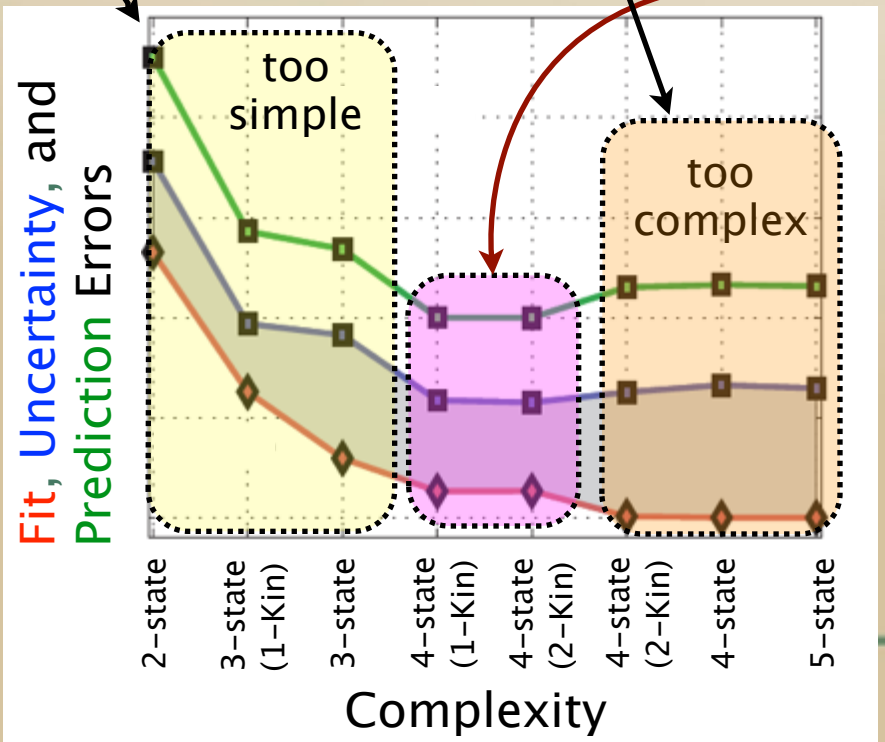
but they also give rise to greater uncertainty.



Evaluating model structures of varying complexity

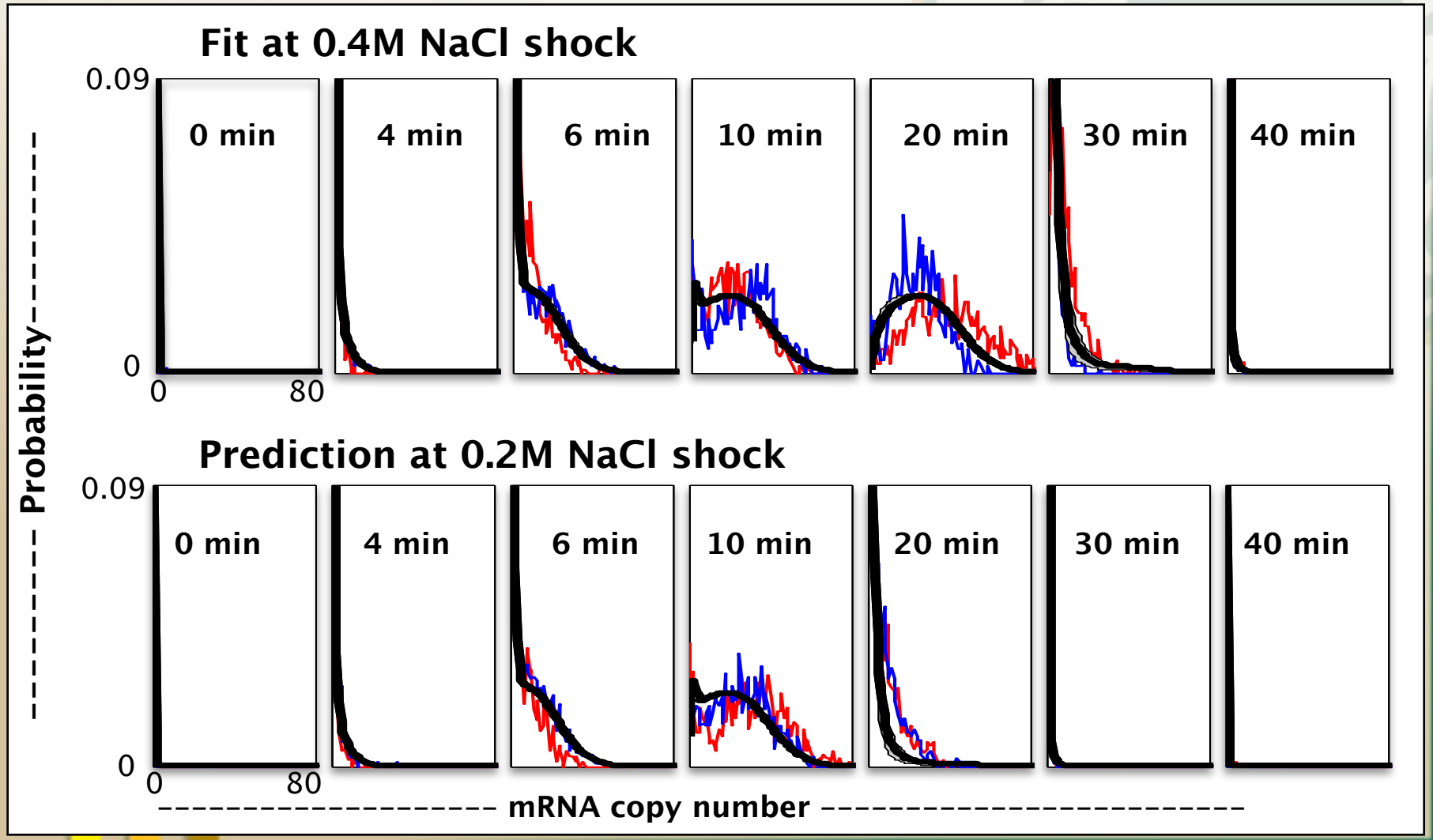
Overly-simple models cannot match the data. → Inaccurate predictions.

Overly-complex models are poorly constrained. → Imprecise predictions.



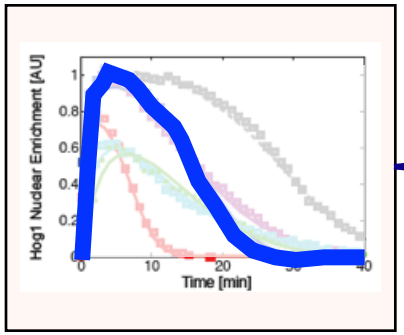
Cross-validation analysis provides an excellent a priori estimate of predictive power.

Fits and predictions for STL1 regulation

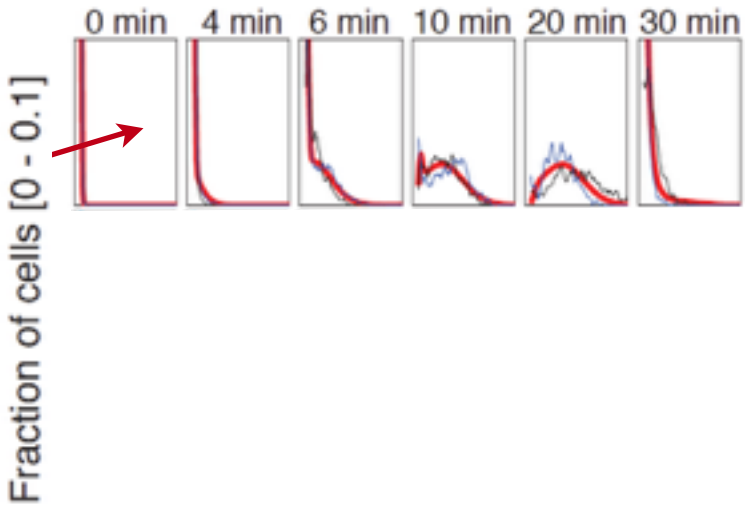


The model can capture and predict WT mRNA dynamics for *STL1*

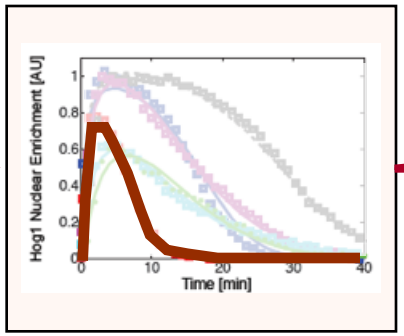
0.4M NaCl Stress



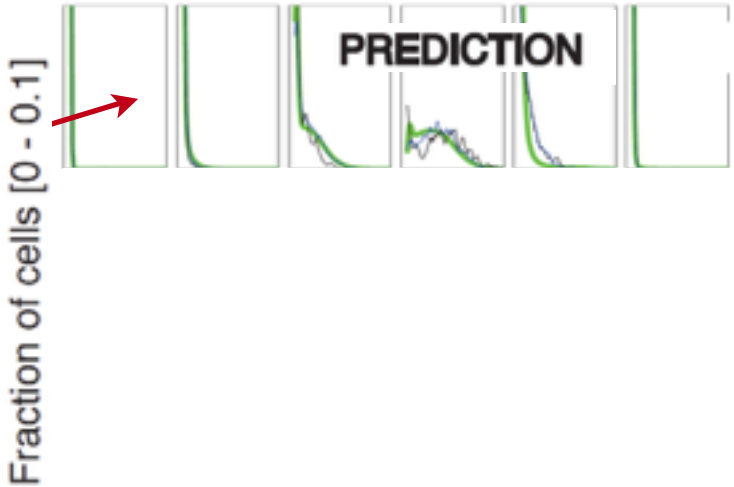
STL1



0.2M NaCl Stress

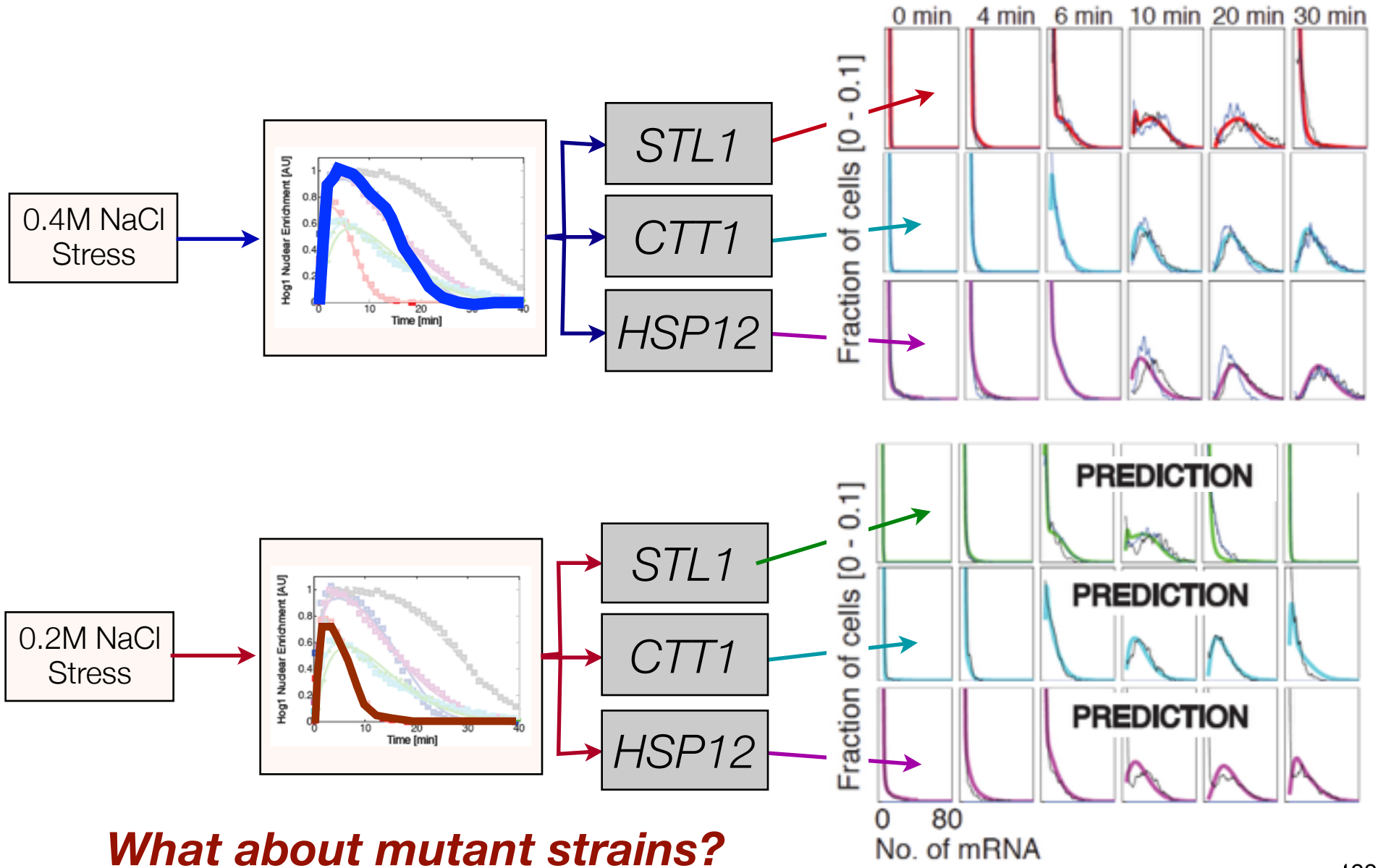


STL1



What about other genes?

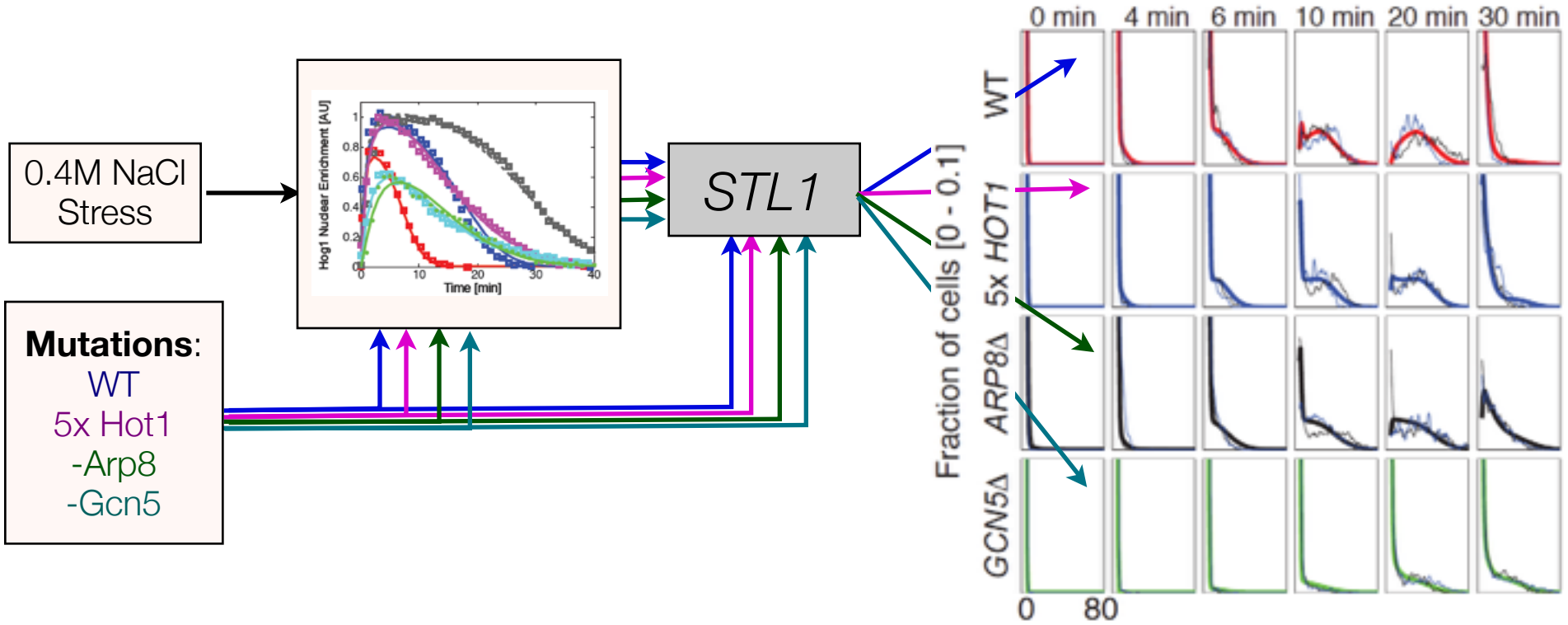
The model can capture and predict WT mRNA dynamics for *STL1*, *CTT1* and *HSP12*



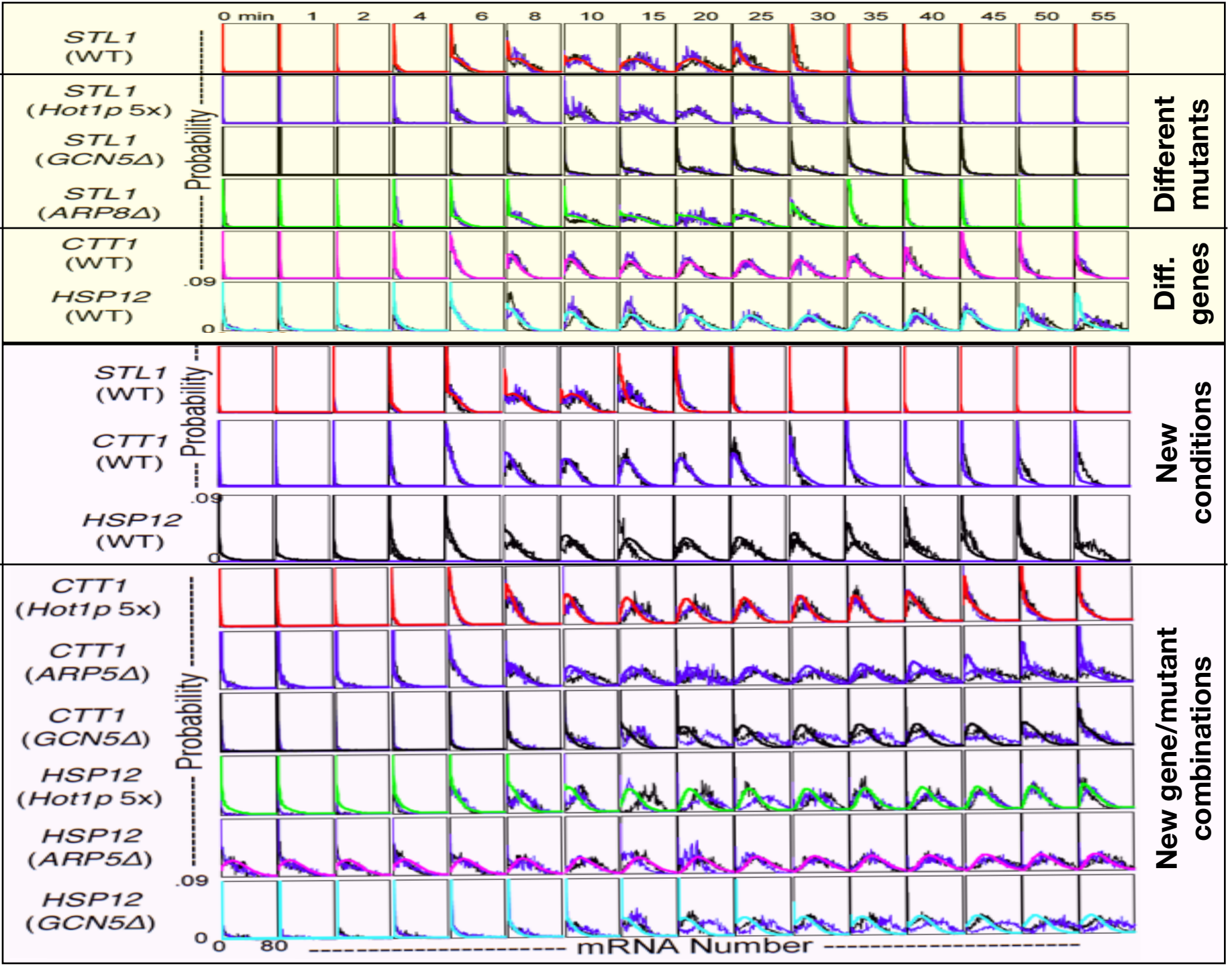
What about mutant strains?

The model can capture and predict WT mRNA dynamics for *STL1*, *CTT1* and *HSP12*

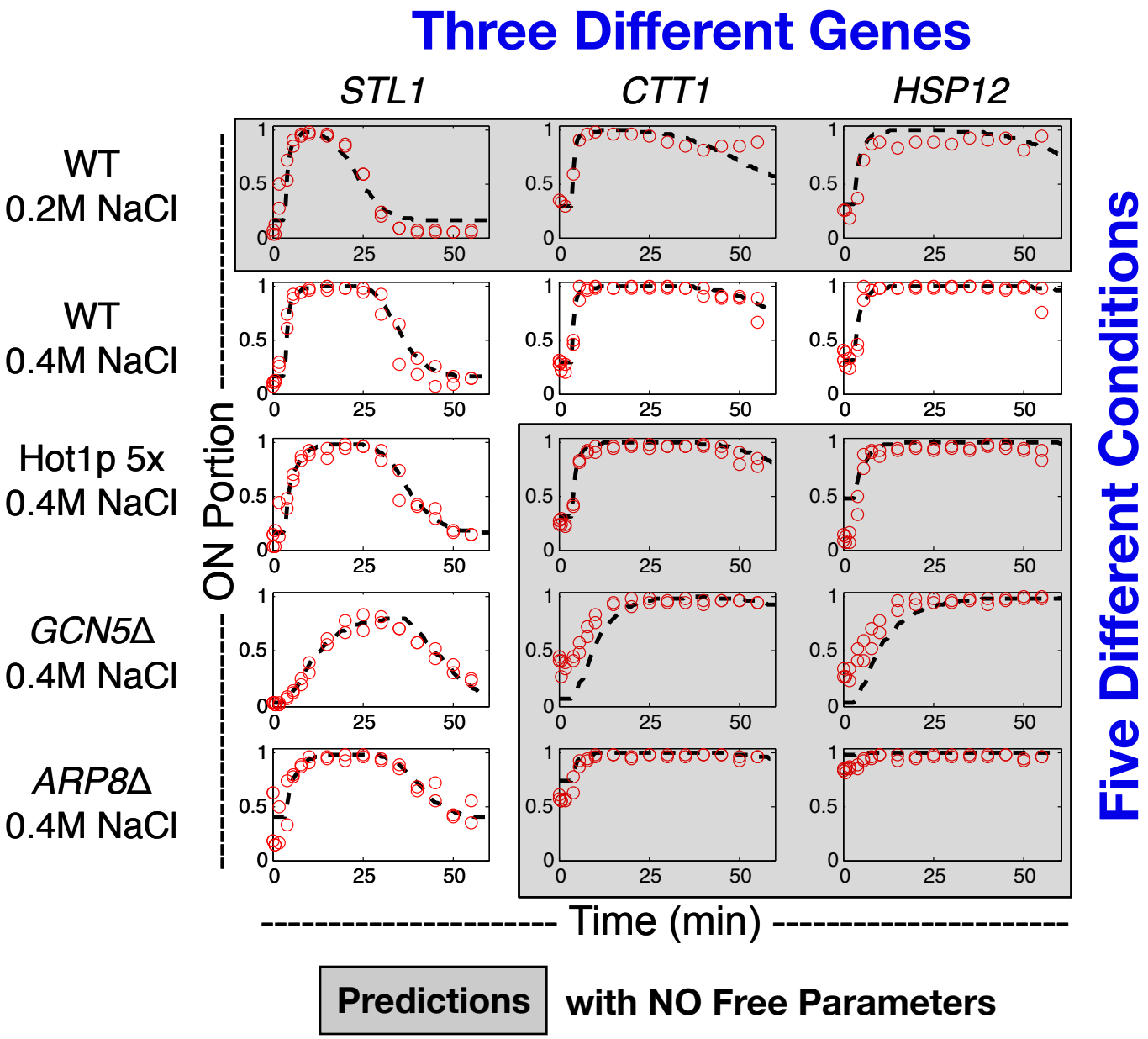
It also captures *STL1* mRNA dynamics in Wild Type, Hot1 over expression and Arp8 or Gcn5 deletion strains



What about new combinations of different genes and mutant strains?

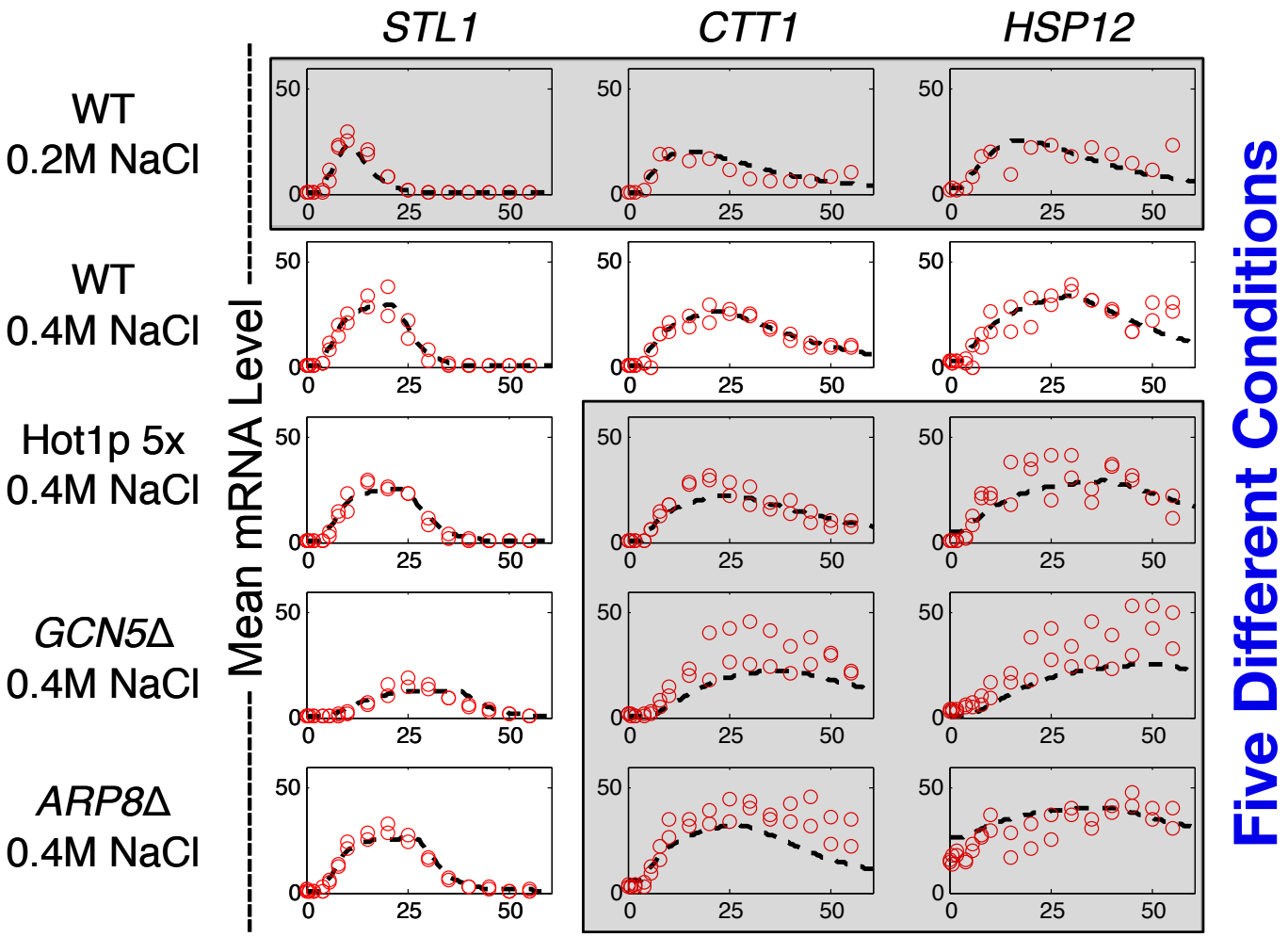

0 min
1
2
4
6
8
10
15
20
25
30
35
40
45
50
55
STL1 (WT)
STL1 (Hot1p 5x)
STL1 (GCN5Δ)
STL1 (ARP8Δ)
CTT1 (WT)
HSP12 (WT)
Probability
.09
0
Different mutants
Diff. genes
Fits
STL1 (WT)
CTT1 (WT)
HSP12 (WT)
Probability
.09
0
New conditions
CTT1 (Hot1p 5x)
CTT1 (ARP5Δ)
CTT1 (GCN5Δ)
HSP12 (Hot1p 5x)
HSP12 (ARP5Δ)
HSP12 (GCN5Δ)
Probability
.09
0
0
80
mRNA Number
New gene/mutant combinations
Predictions

Fitting and Predicting the Probability of ON Cells



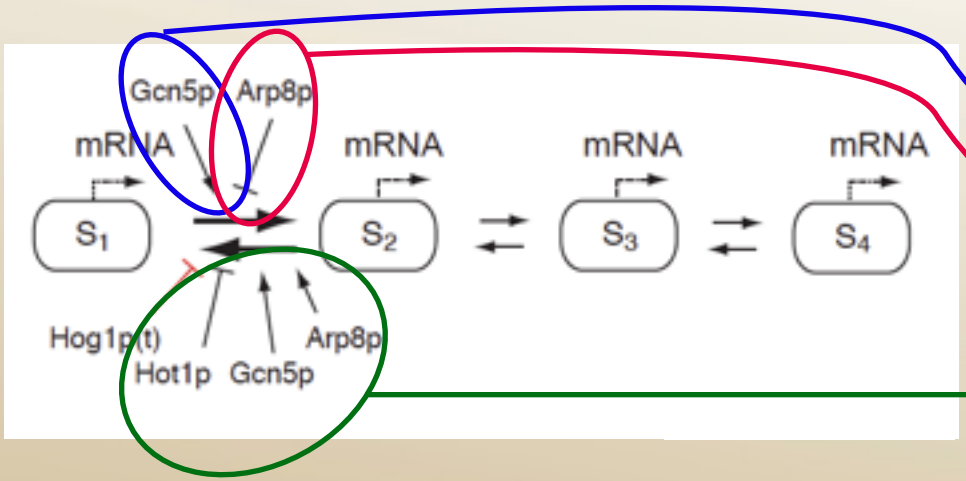
Fitting and Predicting the Mean Expression Level

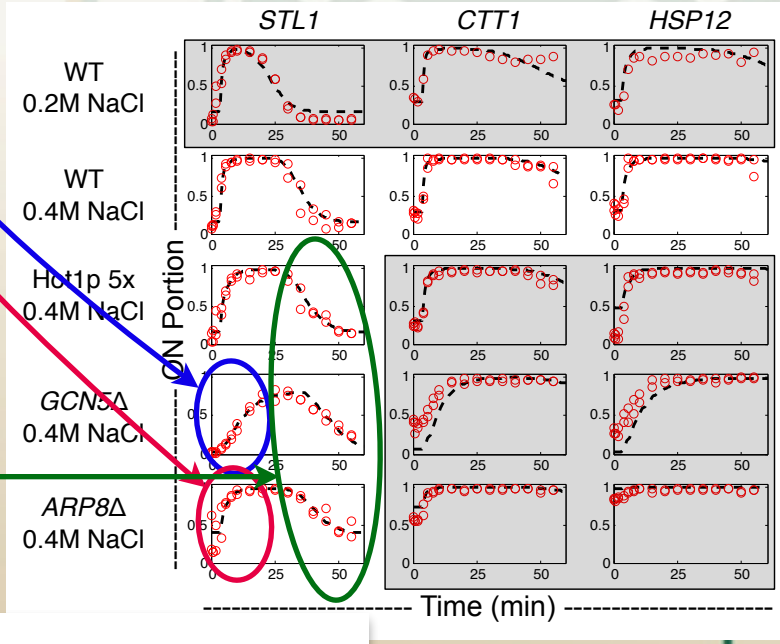
Three Different Genes

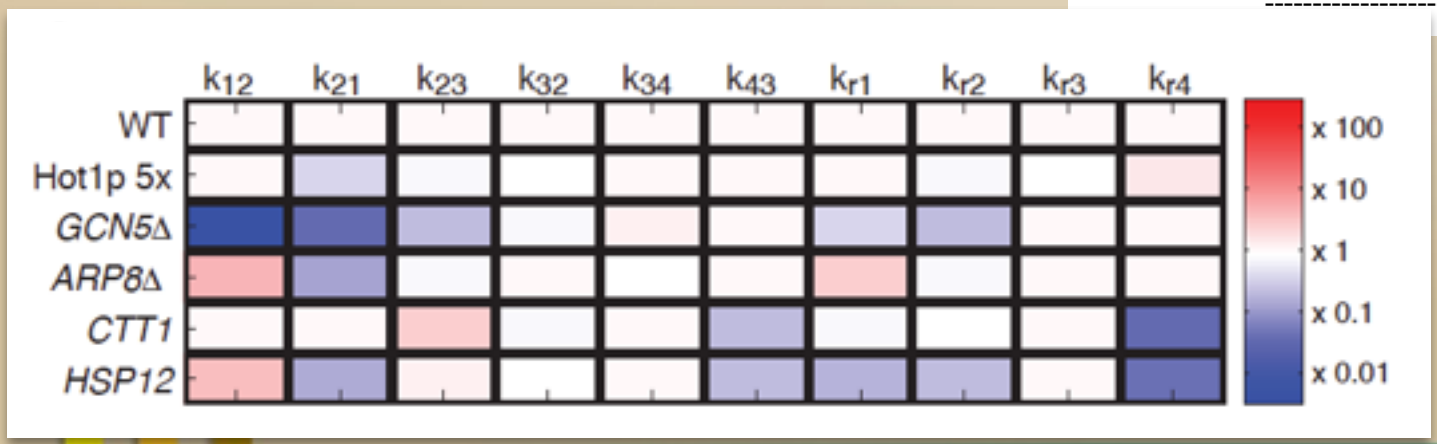


Predictions **with NO Free Parameters**

Final Model Structure:







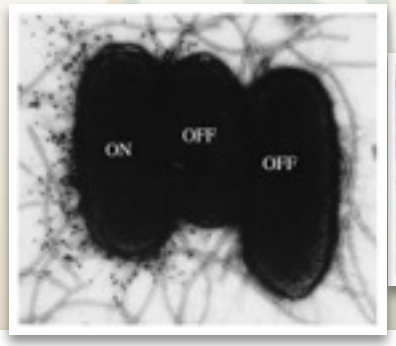
Outline

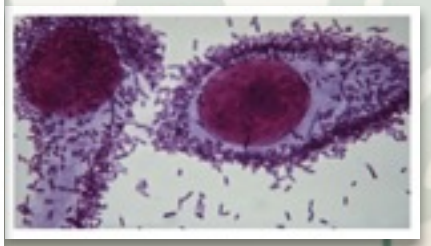
1. Information from single-cell fluctuation
2. Analyzing stochastic dynamics in gene regulation
3. **Case studies:**
    a. Predicting kinase-activated gene regulation dynamics in *Saccharomyces cerevisiae* (budding yeast).
    b. **Predicting multi-generation stochastic behavior of the Pap epigenetic switch in *E. coli***
    c. Predictable design of synthetic circuits in *E. coli*
    d. sRNA regulation in *Yersinia Pestis* and *Yersinia Pseudotuberculosis*

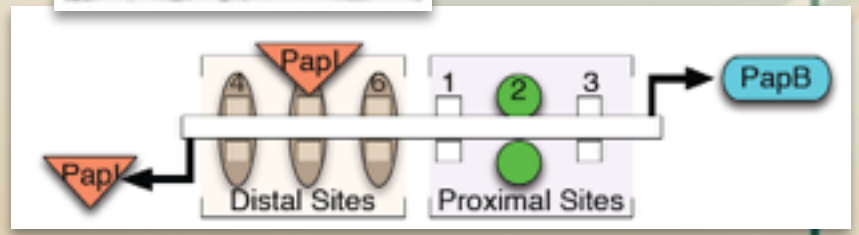
Predicting rare epigenetic switches in *E. coli.*

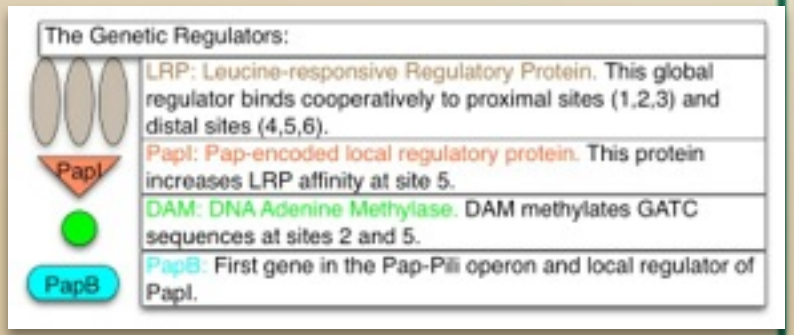
Colorado State University

- Pyelonephritis-Associated Pili are hair-like structures that enable some *E. coli* bind host cells and establish infection.
- Pap express distinct ON and OFF states.
- The *pap* operon contains 6 sites that interact with global regulators DAM and LRP and local regulator PapI.
- Gene activation and DNA methylation pass epigenetic information from mother to daughter cell and stabilize the ON and OFF states.



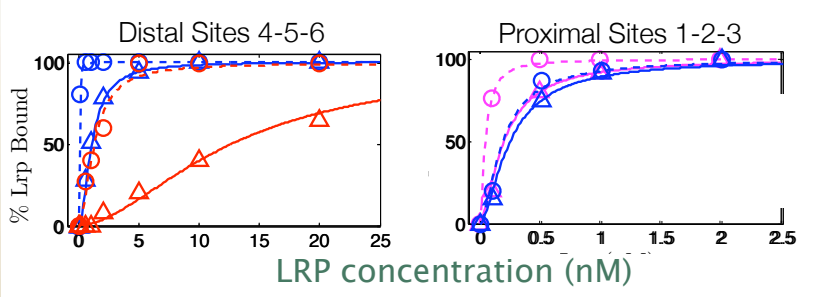




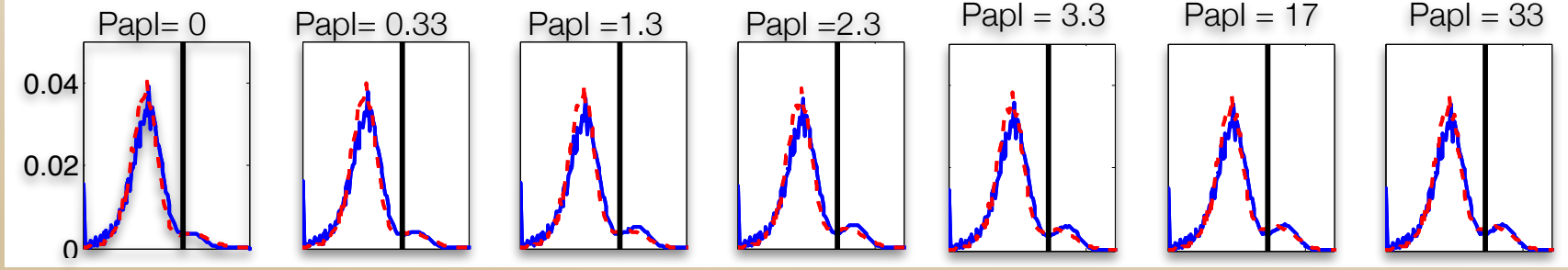


Predicting rare epigenetic switches in *E. coli.*

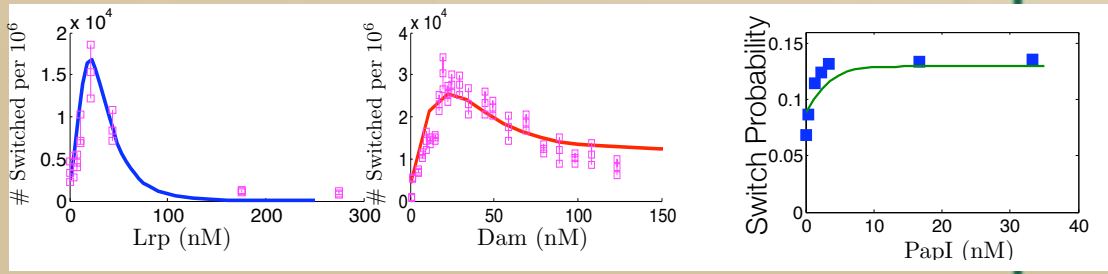
- LRP binding affinities events were fit to previously measured EMSA (*in vitro*) data in various methylation patterns and at various PapI concentrations.



- Remaining parameters were fit to *in vivo* flow cytometry measurements under varying LRP, DAM or PapI titrations



- Switch rates are accurately predicted in all conditions and for several genetic mutations.



Outline

1. Information from single-cell fluctuation

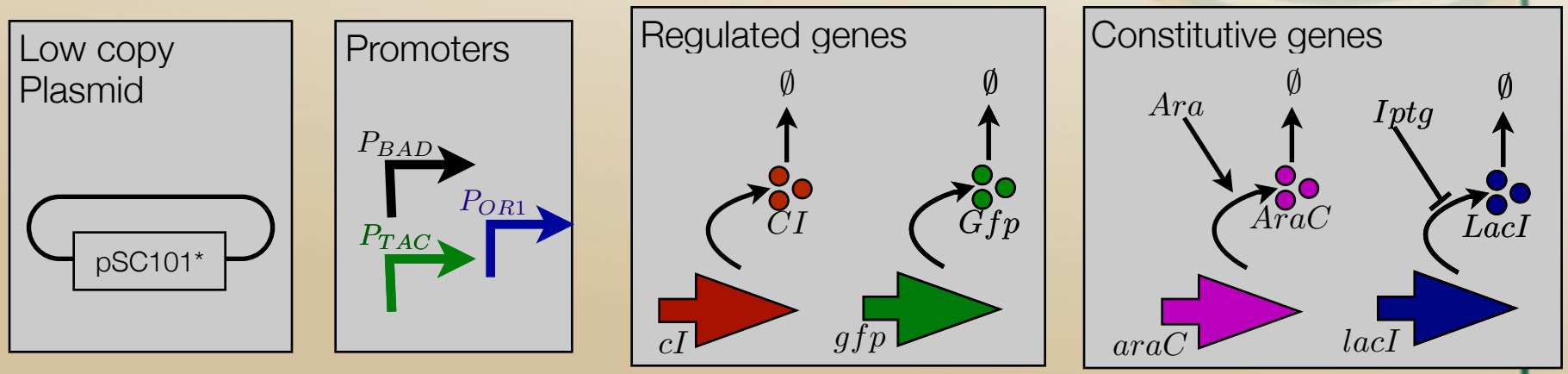
2. Analyzing stochastic dynamics in gene regulation

3. **Case studies:**
    a. Predicting kinase-activated gene regulation dynamics in *Saccharomyces cerevisiae* (budding yeast).
    b. Predicting multi-generation stochastic behavior of the Pap epigenetic switch in *E. coli*
    c. **Predictable design of synthetic circuits in *E. coli***
    d. sRNA regulation in *Yersinia Pestis* and *Yersinia Pseudotuberculosis*

Designing predictable parts for synthetic biology

- Synthetic biology requires genetic building blocks, which...
  1. can be characterized independently of final context, and
  2. behave in a predictable manner when assembled.

Synthetic building blocks

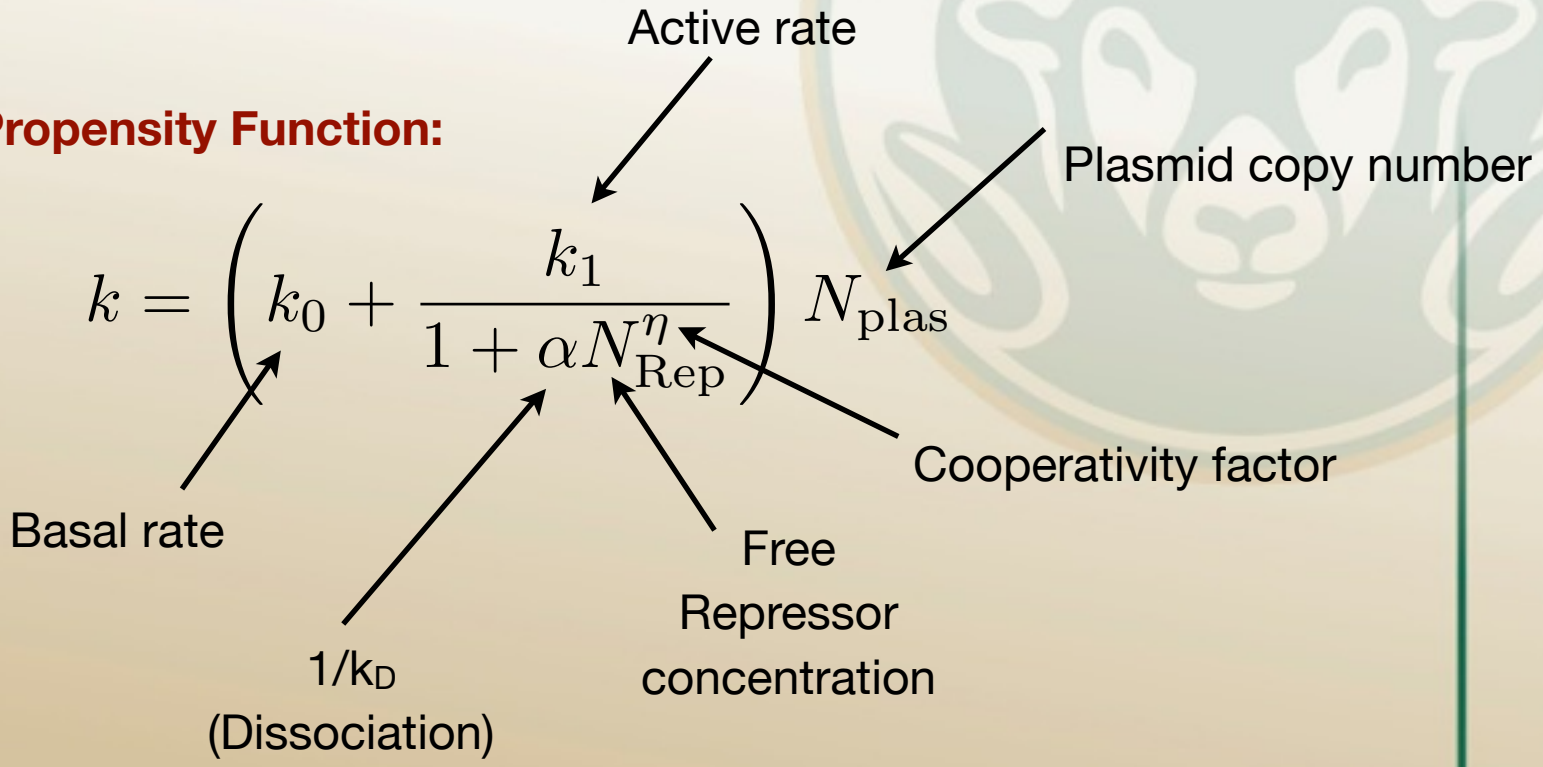


Can we account for noise and predict responses when when we mix-and-match these parts?

Lou, et al, *Nature Biotechnology, 2012*

Parameterization of reactions

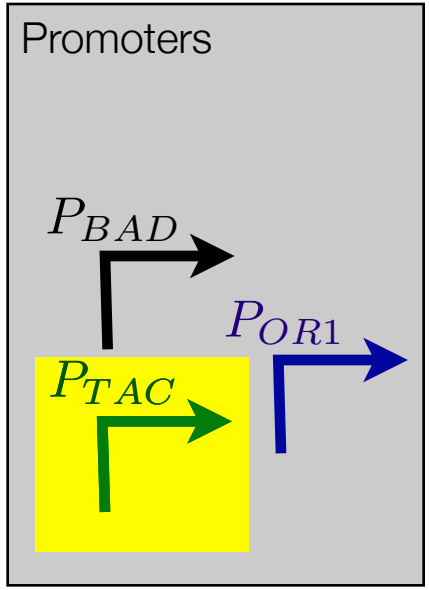
Production Propensity Function:

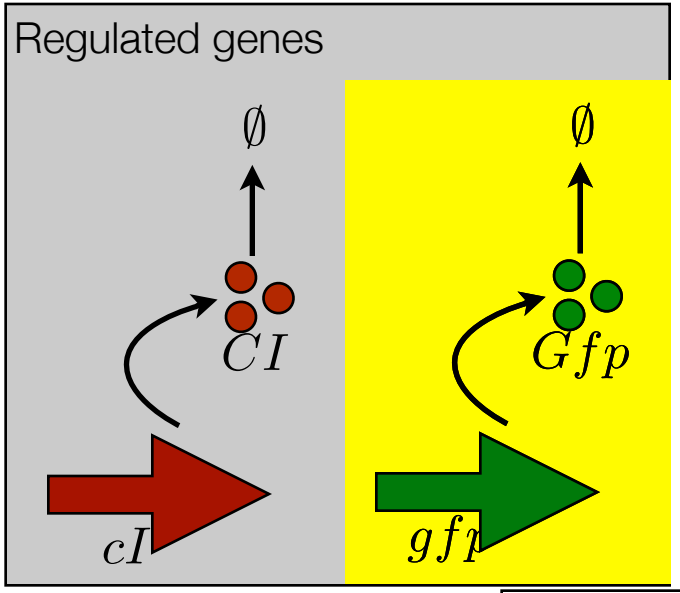


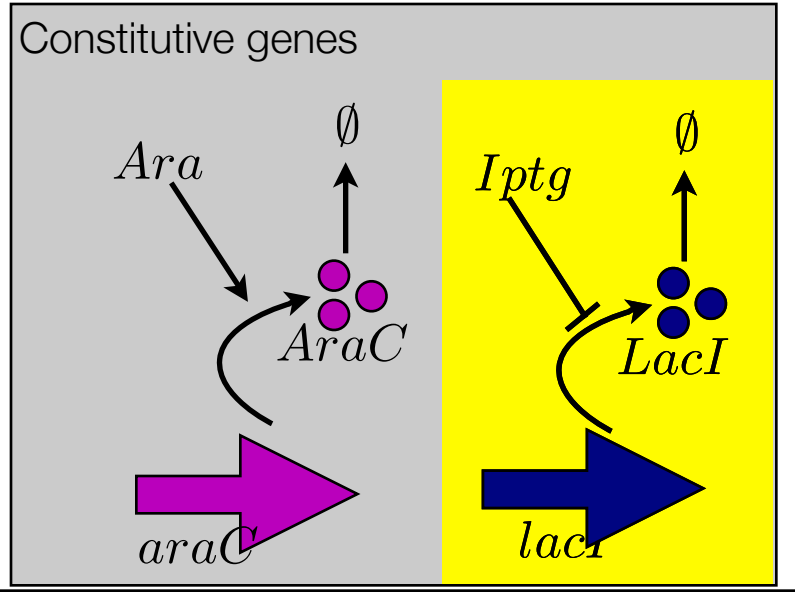
Degradation Propensity Function:

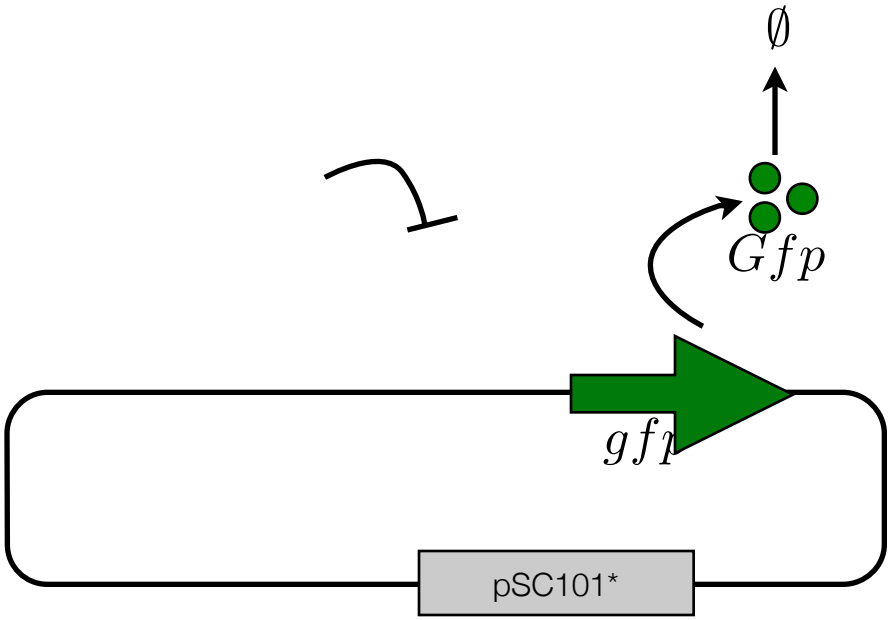
$$\gamma = \gamma_1 N$$

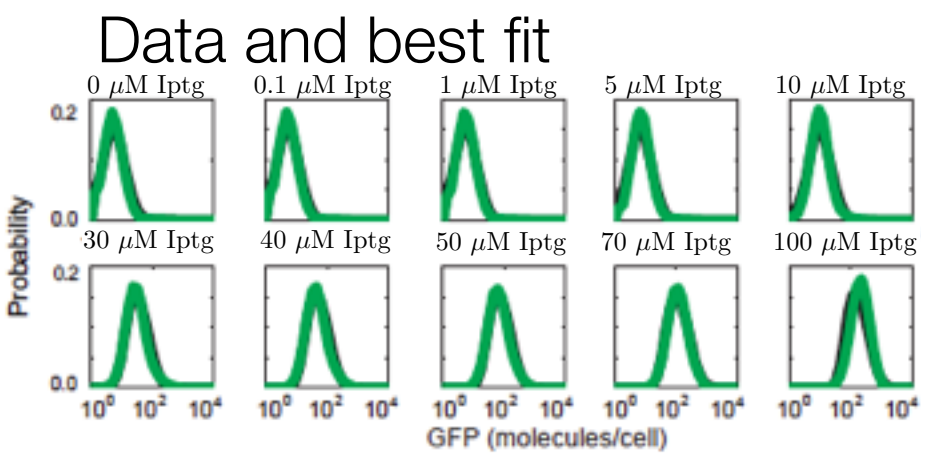
Lou, et al, *Nature Biotechnology, 2012*

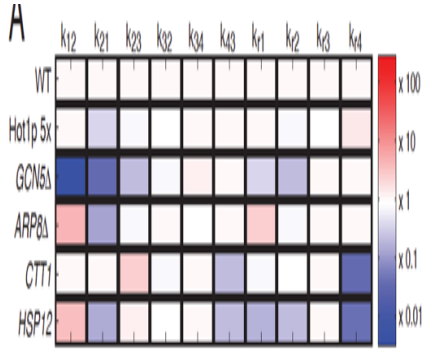




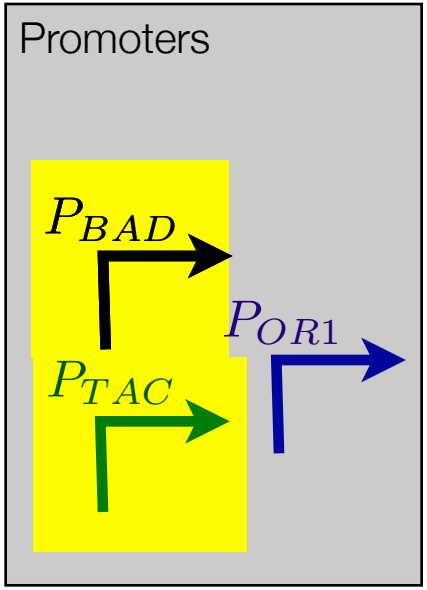


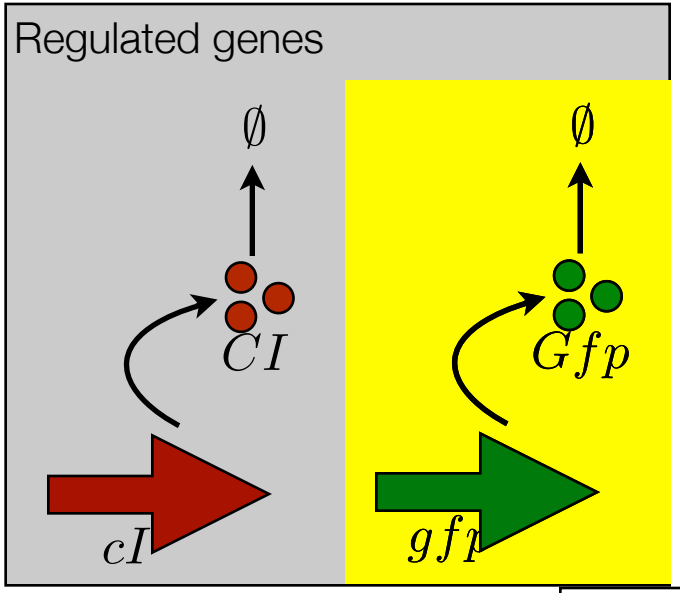


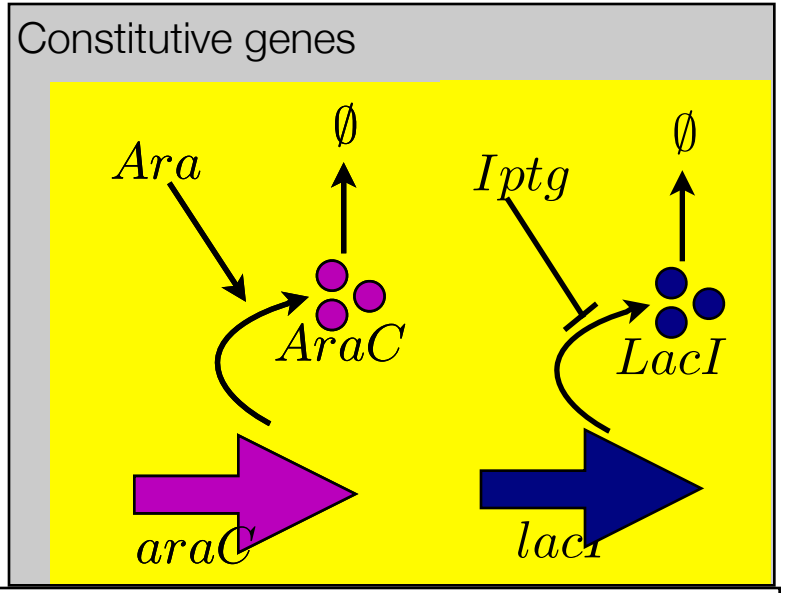


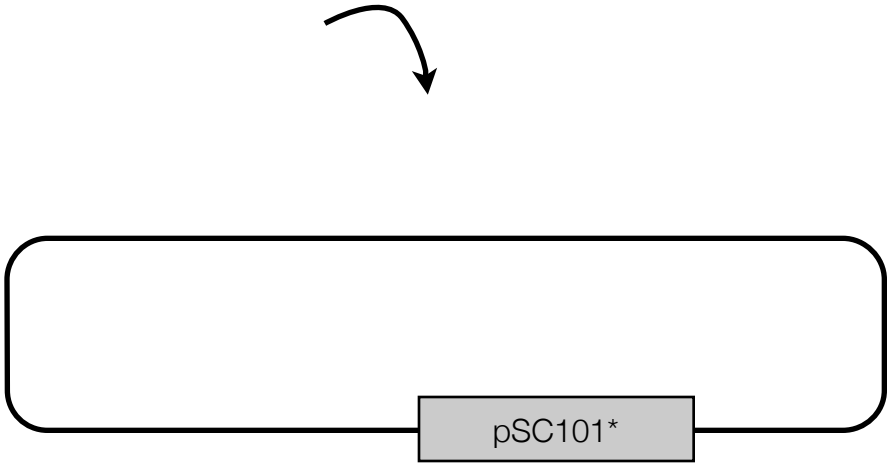


Lou, et al, *Nature Biotechnology, 2012*

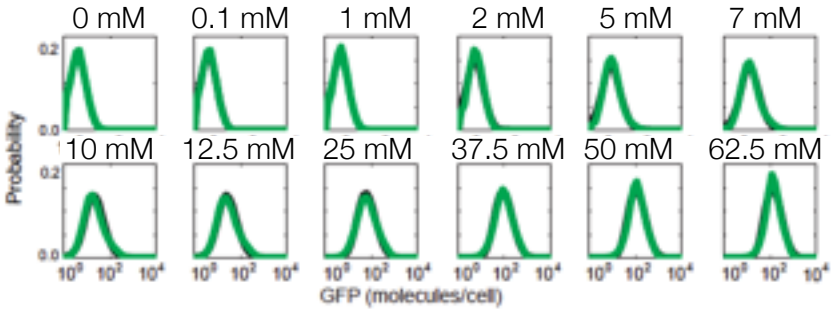


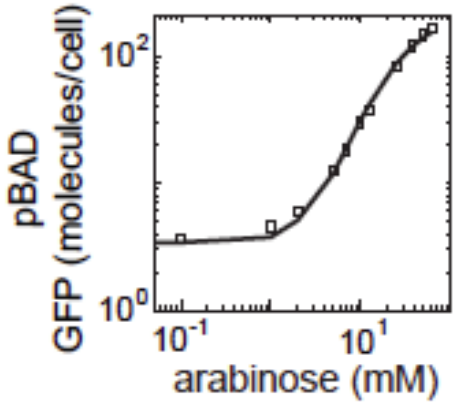




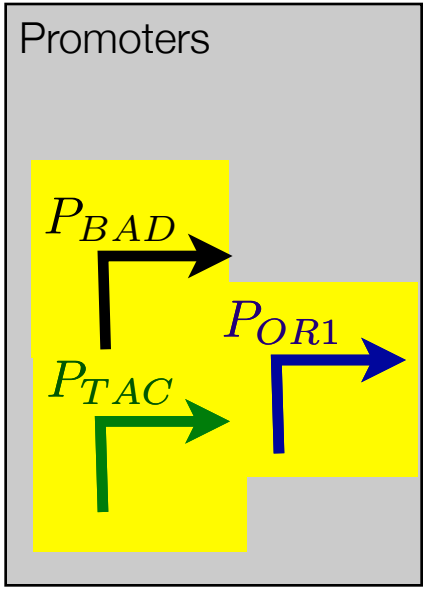


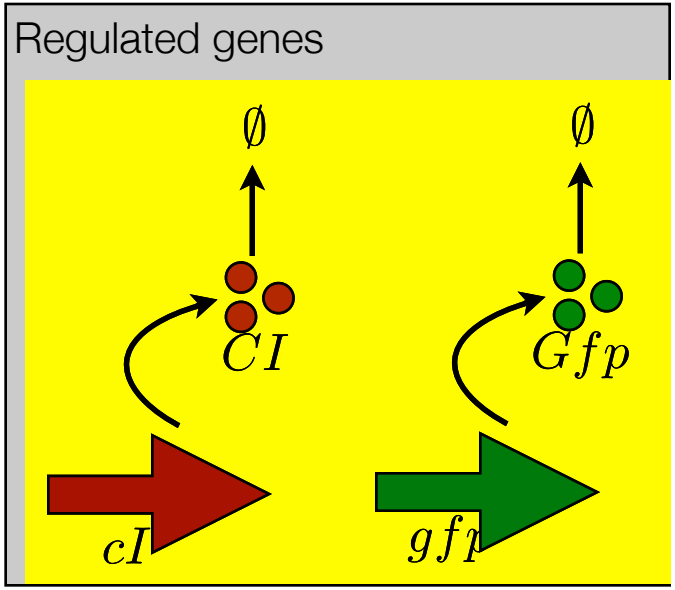
Data and best fit

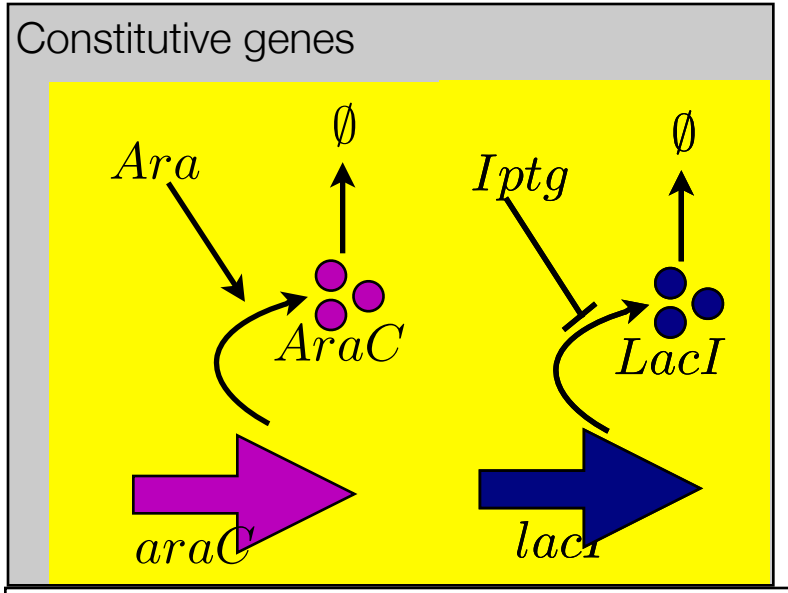


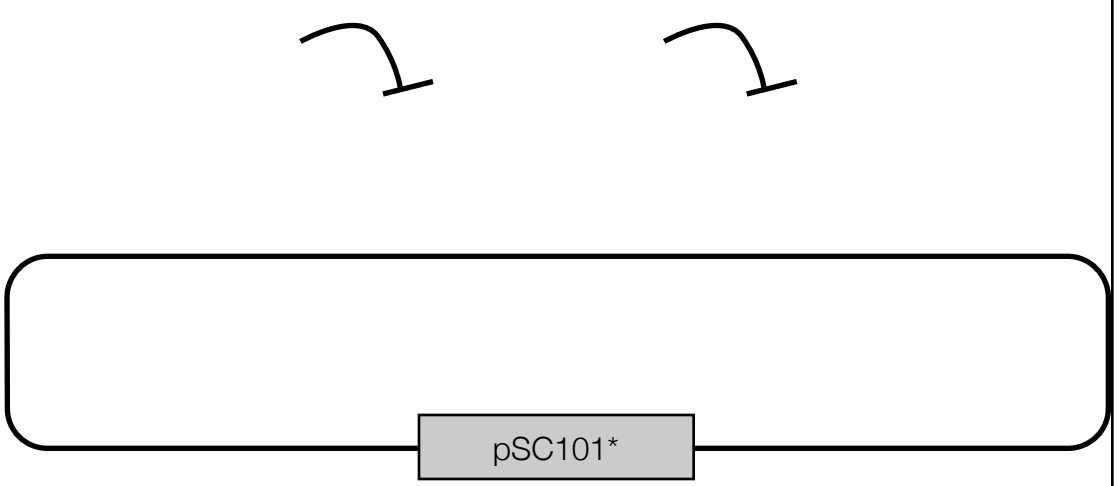


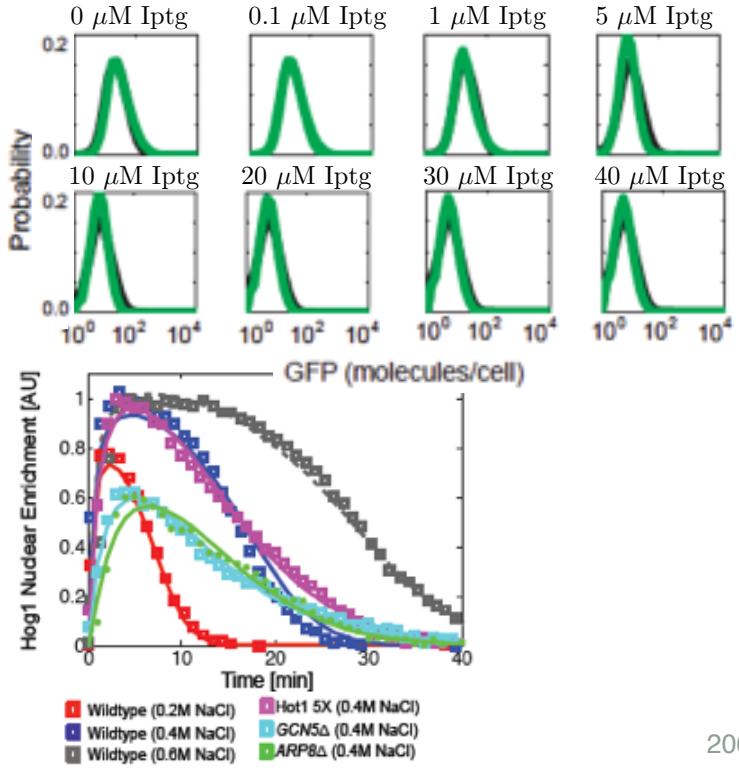
Lou, et al, *Nature Biotechnology, 2012*



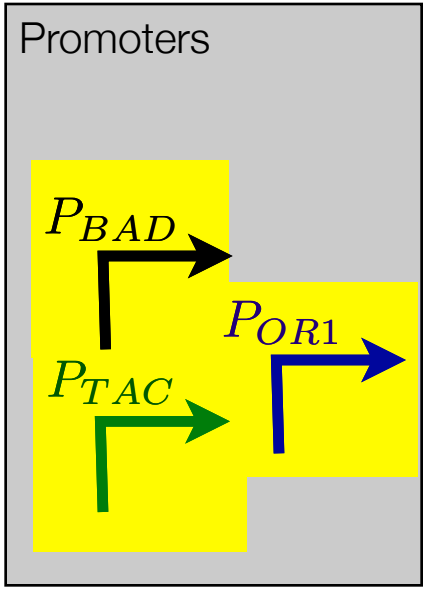


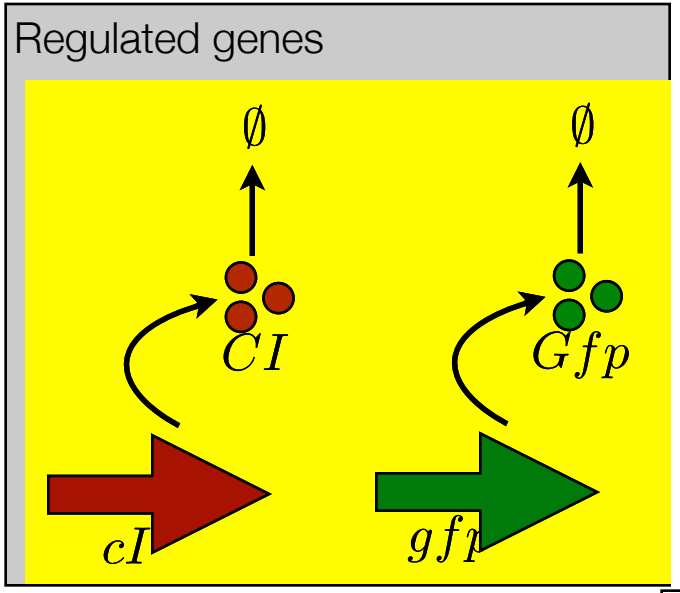


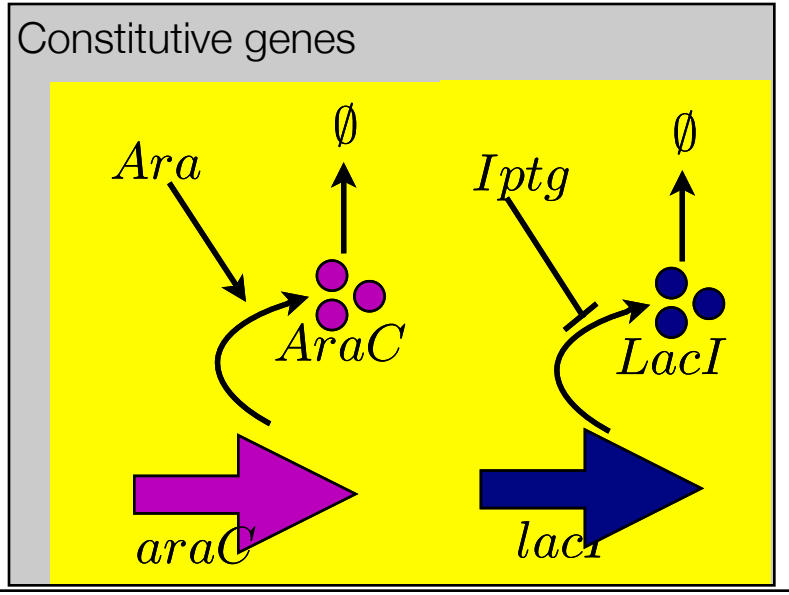


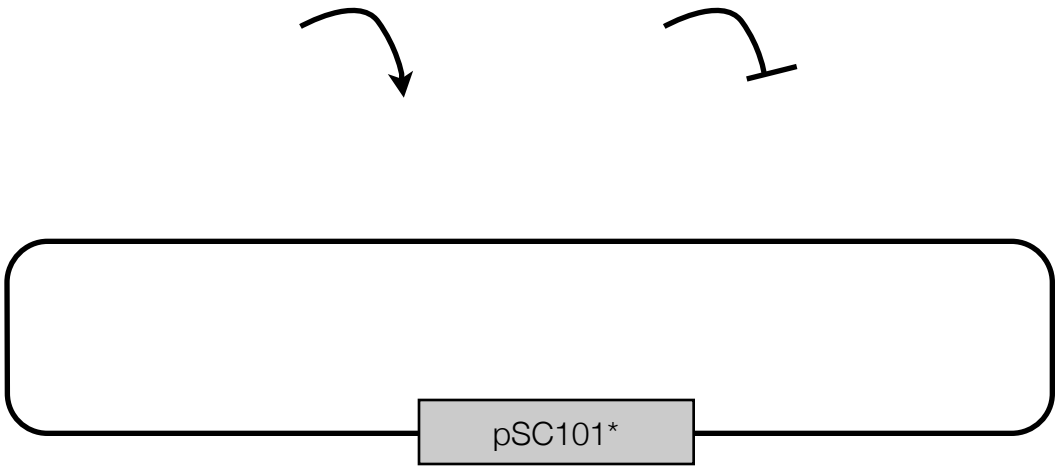


Lou, et al, *Nature Biotechnology, 2012*

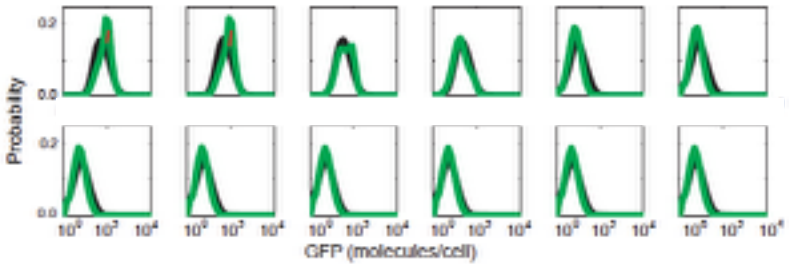


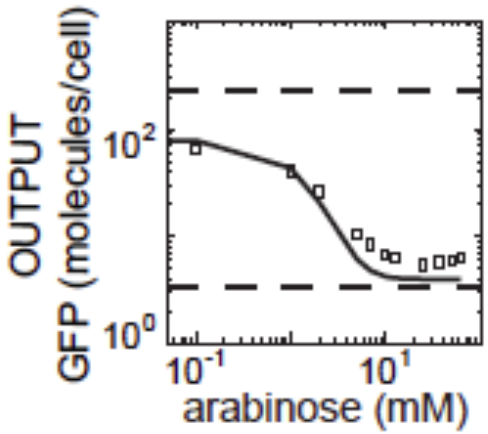






Prediction and Data



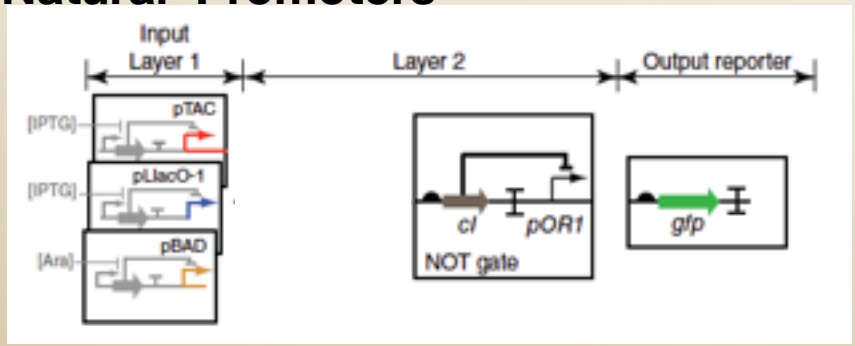


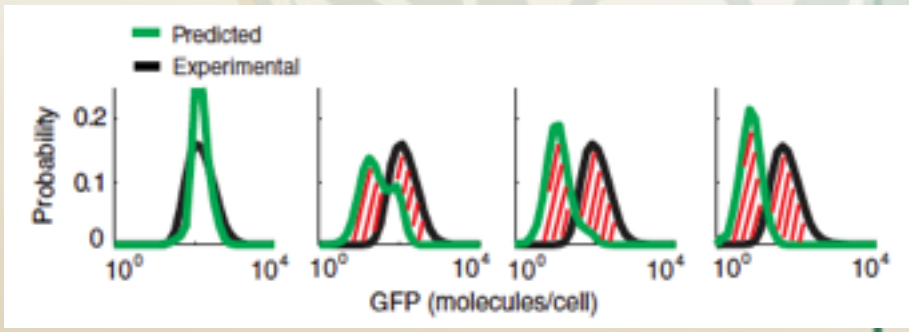
Lou, et al, *Nature Biotechnology, 2012*

Eliminating part-junction interference for synthetic design.

Transcribed sequences between promotors and output genes may disrupt modular behavior.

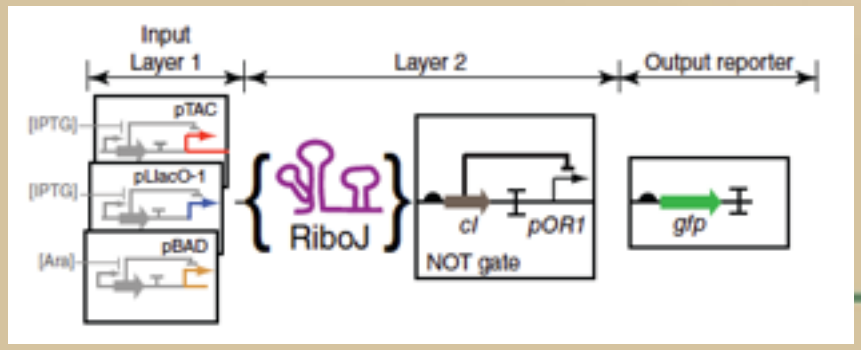
'Natural' Promoters

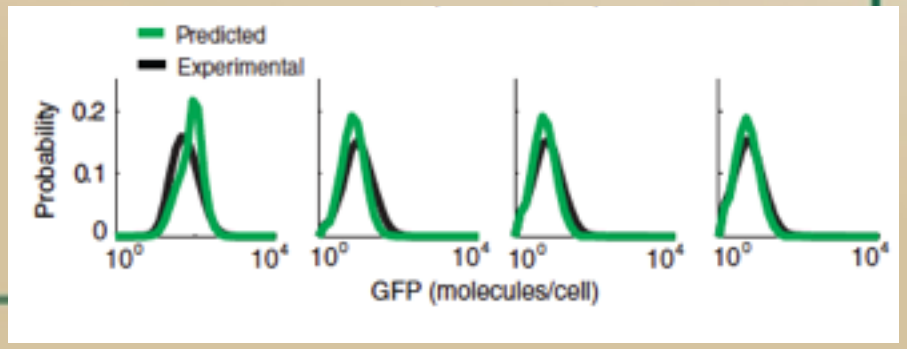




Ribozyme buffers remove these sequences to restore plug-and-play modularity.

'Buffered' Promoters





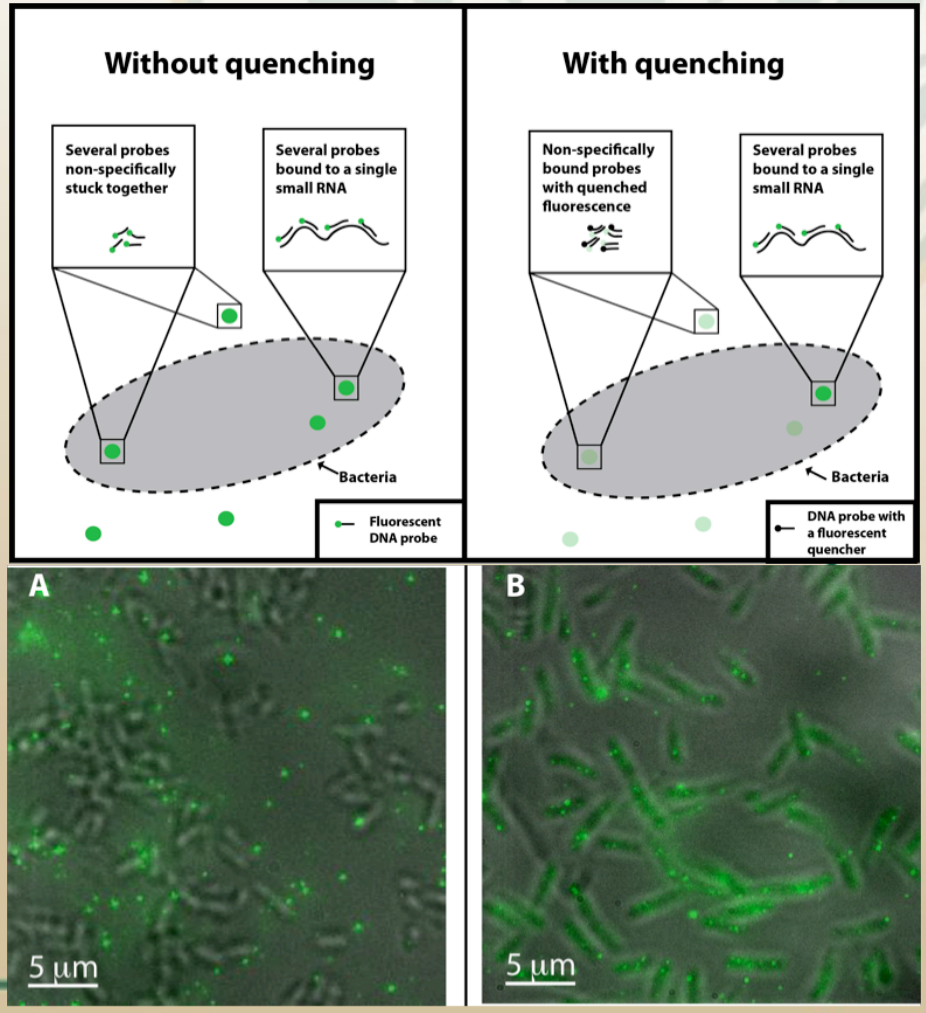
Outline

1. Information from single-cell fluctuation
2. Analyzing stochastic dynamics in gene regulation
3. **Case studies:**
    a. Predicting kinase-activated gene regulation dynamics in *Saccharomyces cerevisiae* (budding yeast).
    b. Predicting multi-generation stochastic behavior of the Pap epigenetic switch in *E. coli*
    c. Predictable design of synthetic circuits in *E. coli*
    d. **sRNA regulation in *Yersinia Pestis* and *Yersinia Pseudotuberculosis***

Analyzing sRNA regulation in *Yersinia Pestis* and *Yersinia Pseudotuberculosis*

Colorado State University

- Small RNA (sRNA) are too short for so many probes.
- Nonspecifically bound probes dominate the fluorescence signal, and new labeling approaches are required.
- NEW complementary quenchers silence non-specifically bound probes, and allow localization and counting of smaller RNA molecules.



YSR35 small RNA in Yersinia Pseudotuberculosis in the absence (A) or presence (B) of quenchers

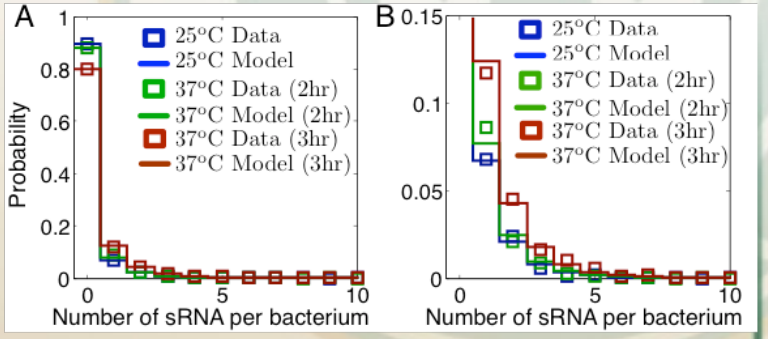
Shepherd, et al, *Analytical Chemistry, 2013*

Analyzing sRNA regulation in *Yersinia Pestis* and *Yersinia Pseudotuberculosis*

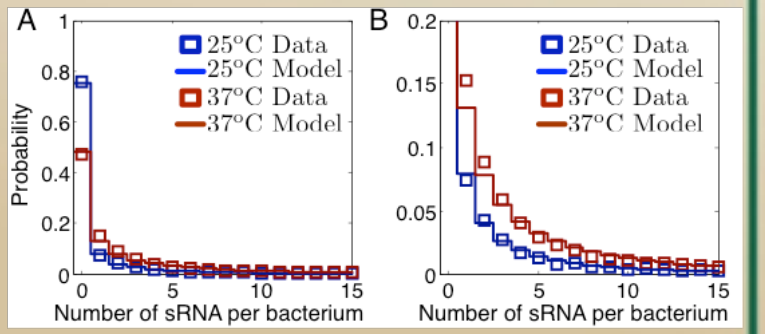
Colorado State University

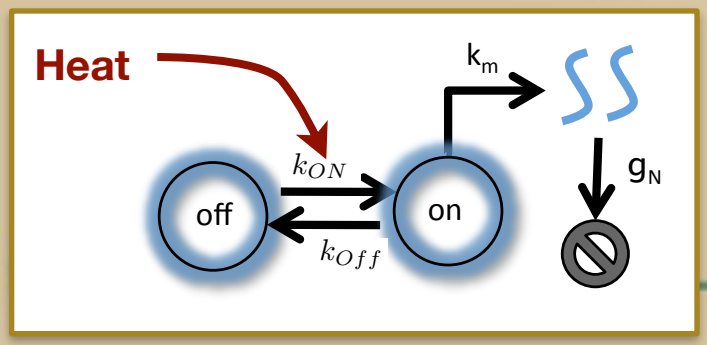
- We measured two temperature sensitive sRNA in YPE and YPSE.
- Both are fit by the bursting gene expression model, but does temperature affect the frequency or amplitude of bursts?
- Testing both hypotheses, we found that YSP8 and YSR35 dynamics match burst frequency modulation.

YSP8 sRNA in Yersinia Pestis



YSR35 sRNA in Y. Pseudotuberculosis





Shepherd, et al, *Analytical Chemistry, 2013*

Outline

1. Information from single-cell fluctuation
2. Analyzing stochastic dynamics in gene regulation
3. Case studies:
    a. Predicting kinase-activated gene regulation dynamics in Saccharomyces cerevisiae (budding yeast).
    b. Predicting multi-generation stochastic behavior of the Pap epigenetic switch in E. coli
    c. Predictable design of synthetic circuits in E. coli
    d. sRNA regulation in Yersinia Pestis and Yersinia Pseudotuberculosis
4. **Concluding remarks**

Summary and Conclusions

- Fluctuations of single cells are stochastic:
  - can complicate modeling and disrupt the design of synthetic systems.
- Statistics of single-cell fluctuations are deterministic:
  - Cells may exhibit distinct repeatable ‘fluctuation fingerprints’, which can be measured with single-cell and single-molecule approaches.
  - Fluctuation statistics may reveal subtle mechanisms and parameters of gene regulation.
  - **Fluctuation statistics can be predicted with high accuracy.**
- Uncertainty Quantification can reveal when models are too simple, too complex, or just right.
- We have identified predictive quantitative models of transcriptional regulation for many natural and synthetic genes in several organisms.
- **Prediction is a first step toward design, optimization and control in systems and synthetic biology.**

References:

wherein *Single-Cell fluctuations reveal gene regulation mechanisms*

Colorado State University

1. Munsky, Trinh, Khammash, Listening to the Noise: Random Fluctuations Reveal Gene Network Parameters, ***Molecular Systems Biology,*** **5**:318, 2009.
2. Munsky, Neuert, van Oudenaarden, Using Gene Expression Noise to Understand Gene Regulation, ***Science***, **336**:6078, 183–187, 2012.
3. Munsky, Modeling Cellular Variability, in ***Quantitative Biology From Molecular to Cellular Systems***, M. Wall, Ed. (Taylor & Francis Group, New York, 2012).
4. Lou, Stanton, Chen, Munsky, Voigt, Ribozyme-based insulator parts buffer synthetic circuits from genetic context, ***Nature Biotechnology***, **30**:11, 2012.
5. Shepherd, Li, Hong-Geller, Munsky, Werner, New tools for discovering the role sRNA plays in cellular regulation, ***Proc. SPIE,*** 822808, 2012.
6. Neuert, Munsky, Tan, Teytelman, Khammash, van Oudenaarden, Systematic Identification of Signal-Activated Stochastic Gene Regulation, ***Science***, **339**:6119, 2013.
7. Shepherd, Li, Micheva-Viteva, Munsky, Hong-Geller, and Werner, Counting small RNA in pathogenic bacteria, ***Analytical Chemistry***, **85**:10, 2013.
8. *Shepherd, Munsky, Information from Fluctuation: Multiscale Stochastic Analyses to Improve Efficiency and Effectiveness of Single-Cell Studies, in revision.*
9. *A Senecal, B Munsky, et al, Transcription Factors Modulate c-Fos Transcriptional Bursts,* ***Cell Reports, 8****:1, 2014.*

Acknowledgments

Colorado State University

Hog Signaling:
Gregor Neuert, Vanderbilt
Alexander van Oudenaarden, Hubrecht
Rui Zhen Tan, MIT
Leonid Teytelman, MIT
Mustafa Khammash, ETH

Modules for Synthetic Biology:
Chunbo Lou,
Chris Voigt, MIT
Brynne Stanton, MIT
Ying-Ja Chen, MIT

Activation of c-Fos
Adrien Senecal, Albert Einstein
Xavier Darzacq, UC Berkeley
Florian Mueller, Institut Pasteur
Christophe Zimmer, Institut Pasteur

sRNA Dynamics:
Douglas Shepherd, UC Denver
James Werner, LANL
Elizabeth Hong-Geller, LANL
Nan Li, LANL
Sofiya N. Michva-Viteva

The PAP Switch:
Brooke Trinh, UCSB
David Low, UCSB
Mustafa Khammash, ETH

Collaborators on Similar Projects:
Kumkum Ganguly, LANL
Babetta Marrone, LANL
Golan Bel, BIDR
James Faeder, Pitt
Jose Juan Tapia, Pitt
Ilya Nemenman, Emory

Funding: NIH, LANL-LDRD, CSU Start-up funds

munsky@engr.colostate.edu