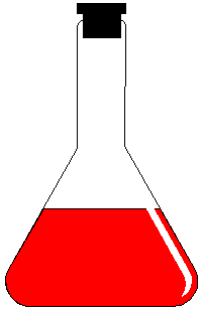
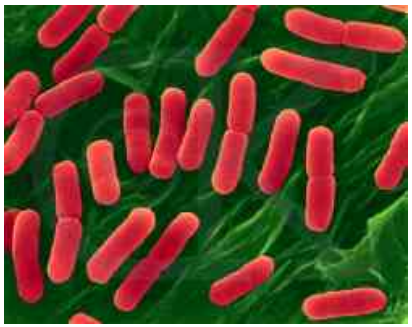
Introduction to Stochastic Simulation with the Gillespie Method

David Karig

April 18, 2005

Stochastic Systems

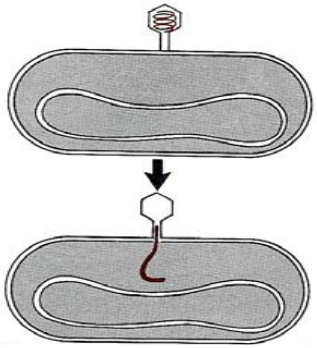




- Many systems driven by random, discrete interactions
- Traditional deterministic models may not accurately describe such systems

Example: The Lambda Switch

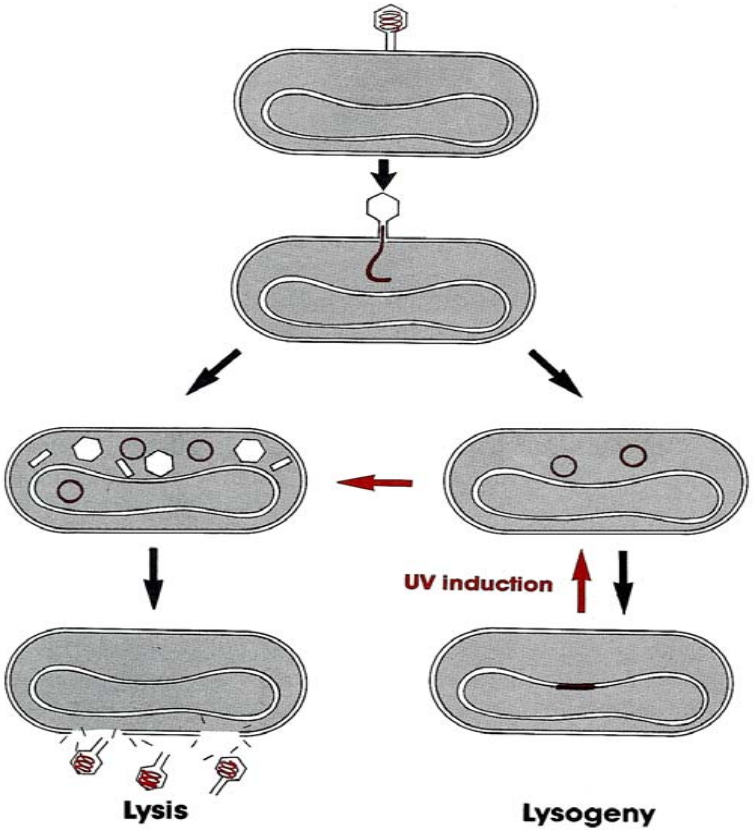
Virus Decision Dictated by Noise



Arkin, Ross, McAdams, *Genetics* (1998)

Example: The Lambda Switch

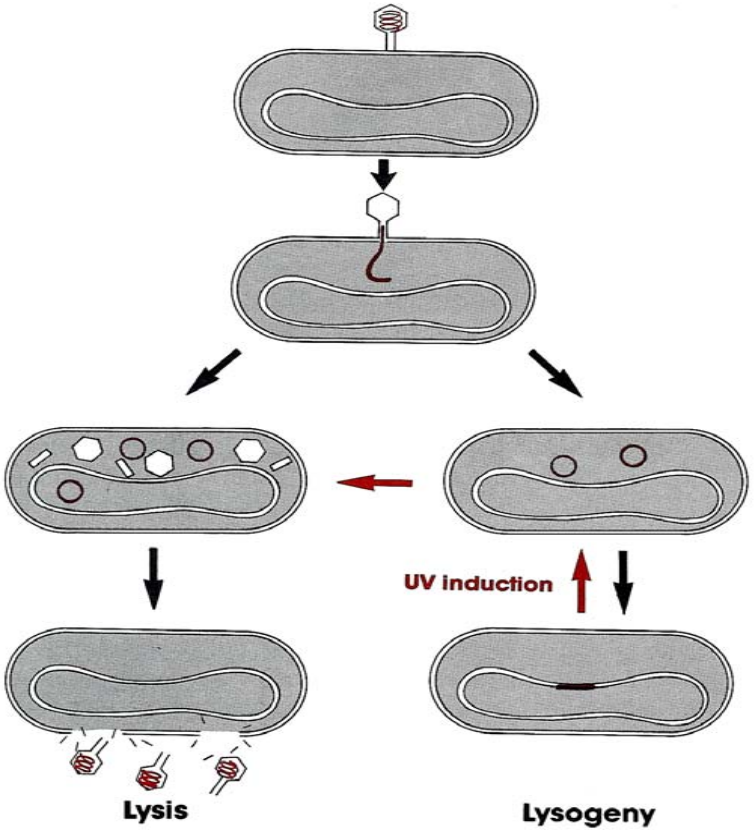
Virus Decision Dictated by Noise

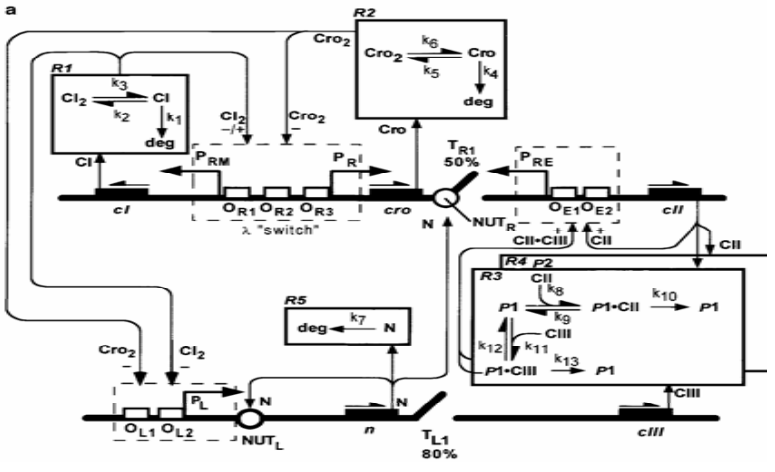


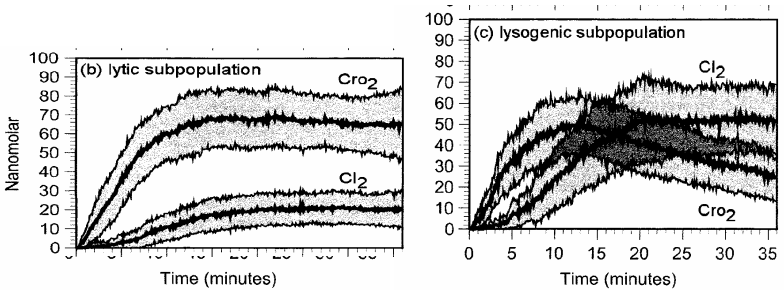
Arkin, Ross, McAdams, *Genetics* (1998)

Example: The Lambda Switch

Virus Decision Dictated by Noise







Arkin, Ross, McAdams, *Genetics* (1998)

Outline

- Deterministic rate reaction model
- Gillespie method
- Examples
  - Lambda phage
  - Epidemiology
- Optimizations

Deterministic Model

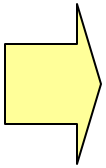
Reactions:

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

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

$$A \xrightarrow{k_3} B$$

$$C \xrightarrow{k_4}$$



ODE's:

$$dA/dt = -k_1*A*B + k_2*C - k_3*A$$

$$dB/dt = -k_1*A*B + k_2*C + k_3*A$$

$$dC/dt = k_1*A*B - k_2*C - k_4*C$$

- Given initial conditions, integrate the coupled equations for some period of time

Problem Statement

If we start with N species which can interact through one of M reactions at a given time,

what will be the population levels of species after a given period of time?

Deterministic Solution

$A + B \rightarrow C$

$C \rightarrow A + B$

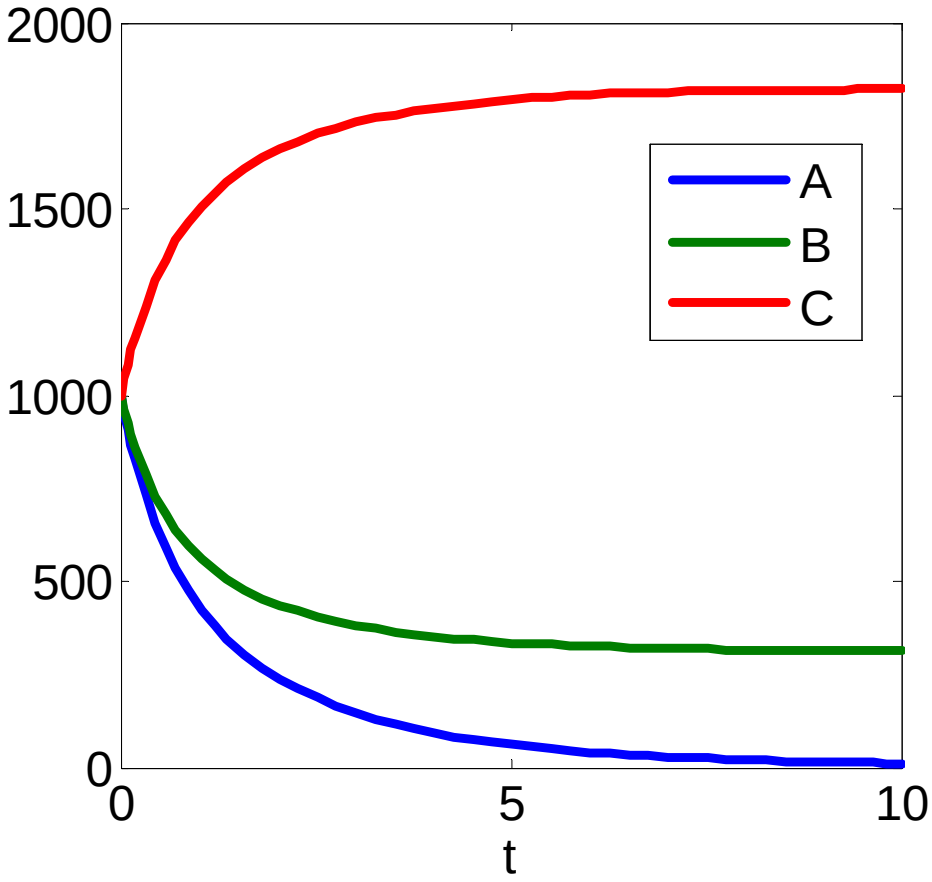
$A \rightarrow B$

$C \rightarrow$

$dA/dt = -k_1*A*B + k_2*C - k_3*A$

$dB/dt = -k_1*A*B + k_2*C + k_3*A$

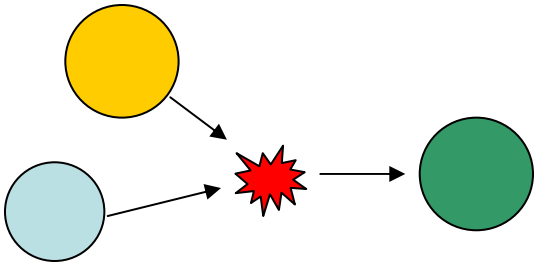
$dC/dt = k_1*A*B - k_2*C - k_4*C$

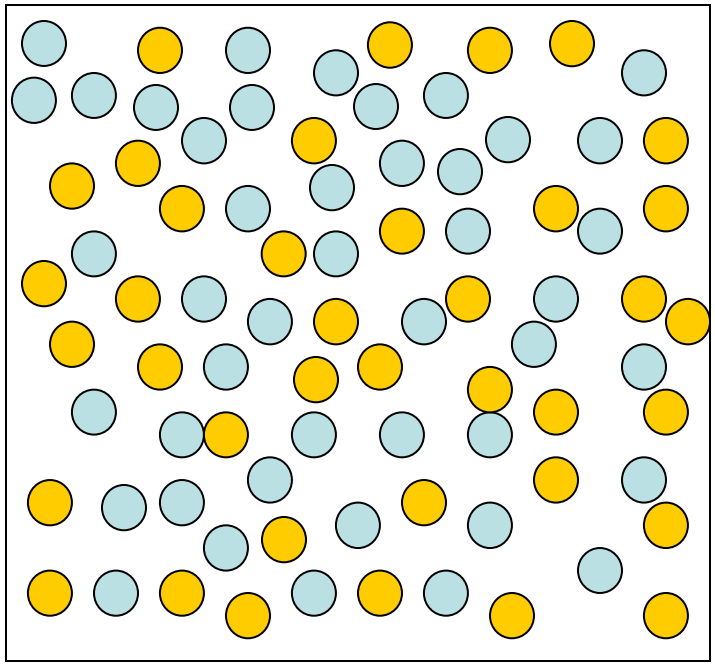


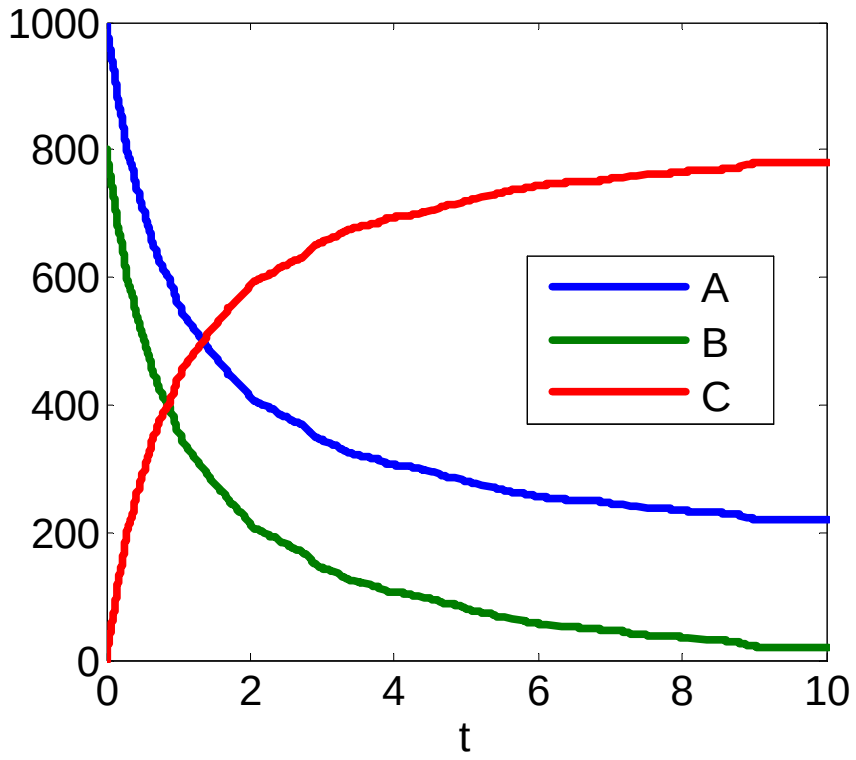
- Continuous
- Average kinetic rates represent reaction probabilities

Validity of Deterministic Solution

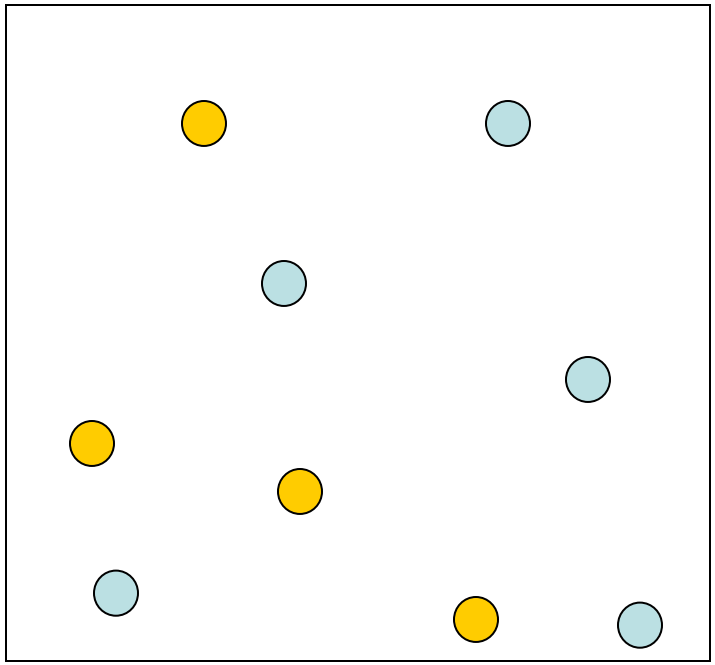
$A + B \rightarrow C$

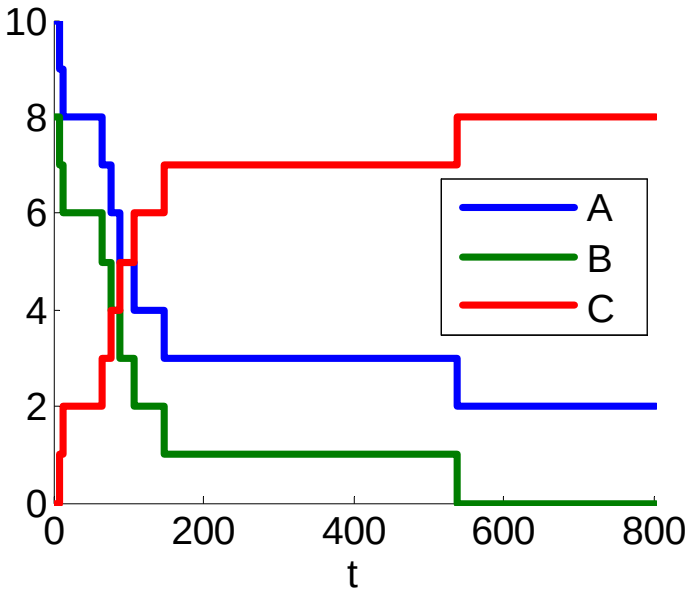


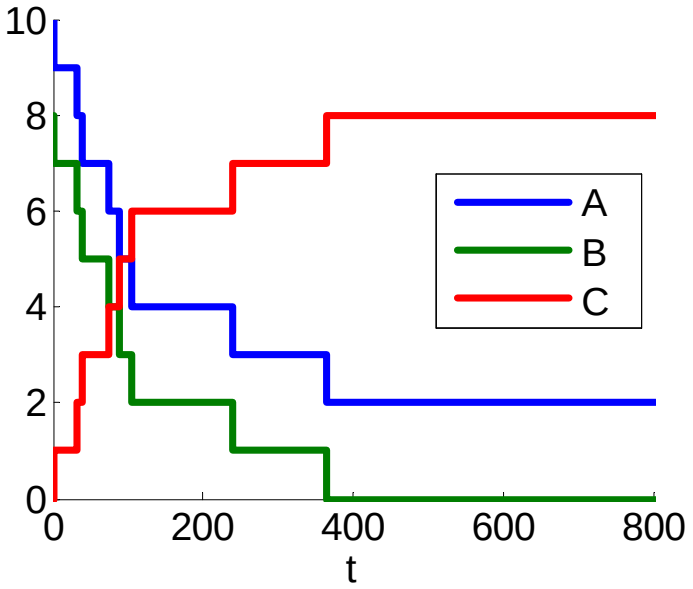




Validity of Deterministic Solution







Reaction Probabilities

- $c_\mu \, dt$ = average probability that a *particular* combination of reactants will react according to R_μ in the next time interval dt
- h_μ = number of reactant combinations
- $h_\mu c_\mu \, dt = a_\mu \, dt$ = average probability that an R_μ reaction will occur somewhere inside V in the next time interval dt

Reaction Probabilities

- $c_\mu \, dt$ = average probability that a *particular* combination of reactants will react according to R_μ in the next time interval dt
- h_μ = number of reactant combinations
- $h_\mu c_\mu \, dt = a_\mu \, dt$ = average probability that an R_μ reaction will occur somewhere inside V in the next time interval dt

R_1: $A + B \rightarrow C$

X molecules of A
Y molecules of B

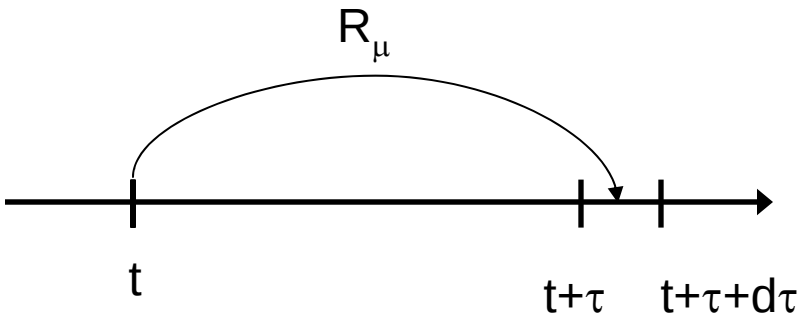
} h_1 = XY reactant combinations

$XYc_1 \, dt$ = probability that an R_1 reaction will occur somewhere inside V in the next time interval dt

Exact Stochastic Simulation

- Avoid averaging assumptions
- Probabilistic formulation
  - *When* does next reaction occur?
  - *Which* reaction occurs next?
- Reaction probability density function:

$P(\tau,\mu)d\tau$ = probability at time *t* that the next reaction is R_μ *and* occurs in interval $(t+\tau, t+\tau+d\tau)$

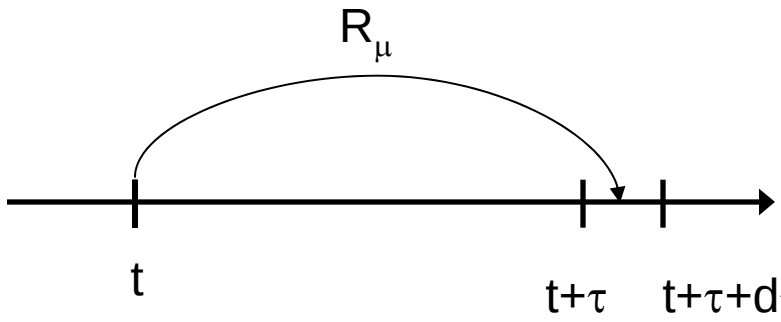


R_1: A + B → C
R_2: C → A + B
R_3: A → B
R_4: C →

System State:

	t	t+τ
A	1	2
B	1	2
C	8	7

Deriving Reaction PDF



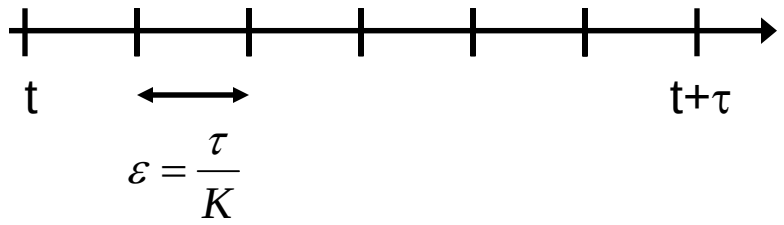
$$P(\tau,\mu)d\tau = P_0(\tau)\, h_\mu c_\mu d\tau$$

$h_\mu c_\mu d\tau$: probability that reaction μ occurs during $(t+\tau, t+\tau+d\tau)$

$P_0(\tau)$: probability that no reaction occurs during $(t,t+\tau)$

Deriving $P_0(\tau)$ for Reaction PDF

- Divide $(t,t+\tau)$ into K subintervals of width ε



- Probability that none of the reactions occur in any of the K subintervals:

$$P_0(\tau) = \left[1 - \sum_{i=1}^{M} h_i c_i \varepsilon\right]^K = \lim_{K\to\infty}\left[1 - \frac{\sum_{i=1}^{M} h_i c_i \tau}{K}\right]^K = e^{-\sum_{i=1}^{M} h_i c_i \tau}$$

Direct Method for Generating P(τ,μ)

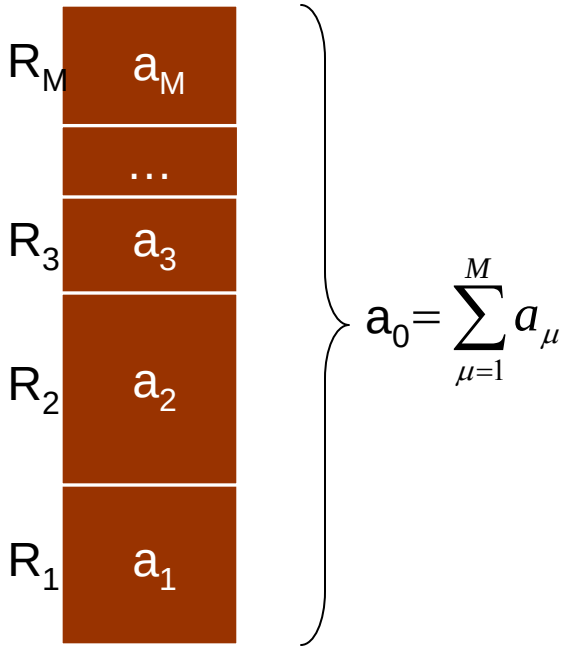
$$P(\tau,\mu) = h_\mu c_\mu P_0(\tau) = h_\mu c_\mu e^{-\sum_{i=1}^{M} h_i c_i \tau}$$

$$= a_\mu e^{-a_0 \tau}$$

$$= \left(a_0 e^{-a_0 \tau}\right)\left(\frac{a_\mu}{a_0}\right) = P(\tau)\cdot P(\mu|\tau)$$

when next reaction occurs

which reaction occurs



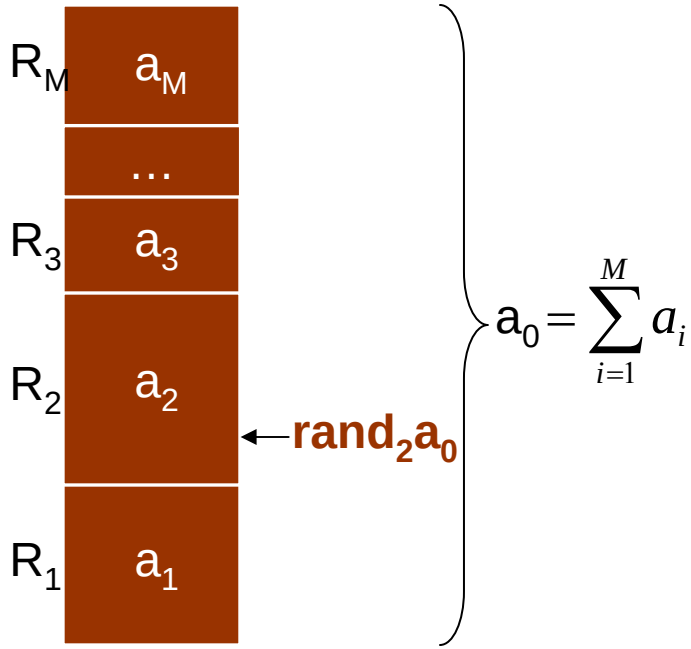
Direct Method for Generating P(τ,μ)

$$P(\tau,\mu) = h_\mu c_\mu P_0(\tau) = h_\mu c_\mu e^{-\sum_{i=1}^{M} h_i c_i \tau}$$

$$= a_\mu e^{-a_0 \tau}$$

$$= \left(a_0 e^{-a_0 \tau}\right)\left(\frac{a_\mu}{a_0}\right) = P(\tau)\cdot P(\mu|\tau)$$

when next reaction occurs — which reaction occurs



$\tau = (1/a_0)\ln(1/rand_1)$

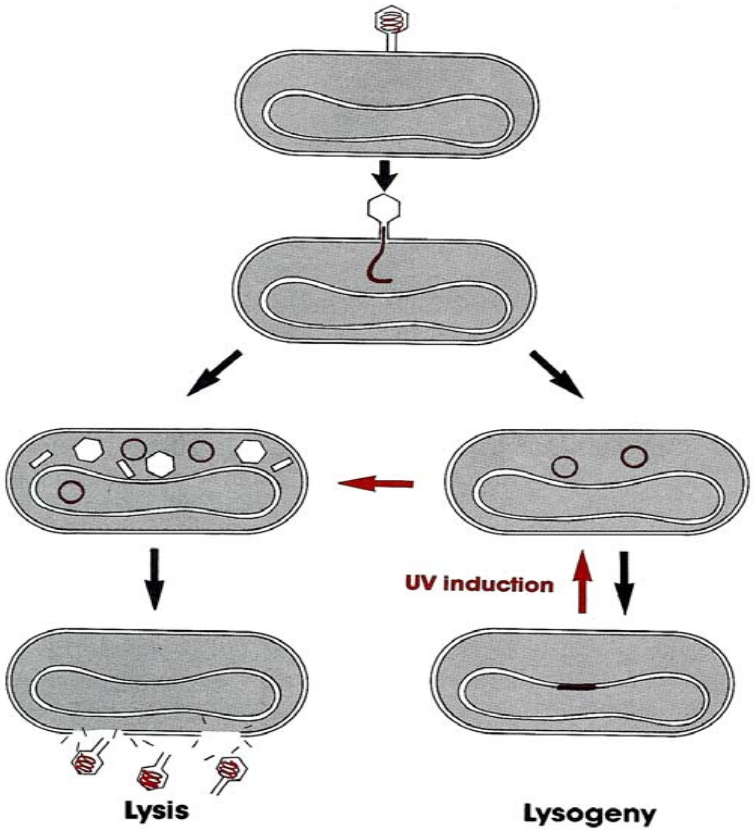
μ is the integer for which $\sum_{i=1}^{\mu-1} a_i < rand_2 a_o < \sum_{i=1}^{\mu} a_i$

Stochastic Simulation Algorithm

1. Initialization
   - Set values of c_μ for the M reactions.
   - Set initial population sizes
2. Calculate the M values a_μ and $a_0 = \Sigma\, a_\mu$.
3. Generate (τ,μ) based on $P(\tau,\mu)$
4. Adjust population levels according to the reaction R_μ, and increase t by τ
5. Return to Step 2

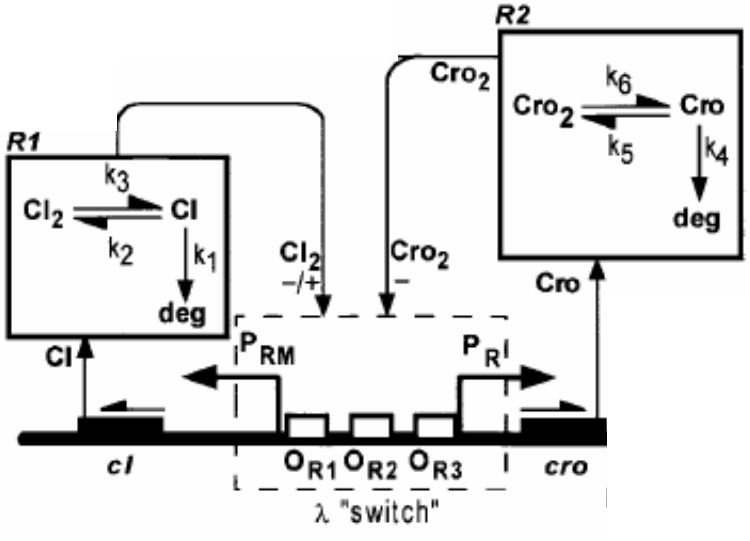
Lambda Phage Developmental Pathway

Arkin, Ross, McAdams, *Genetics* (1998)



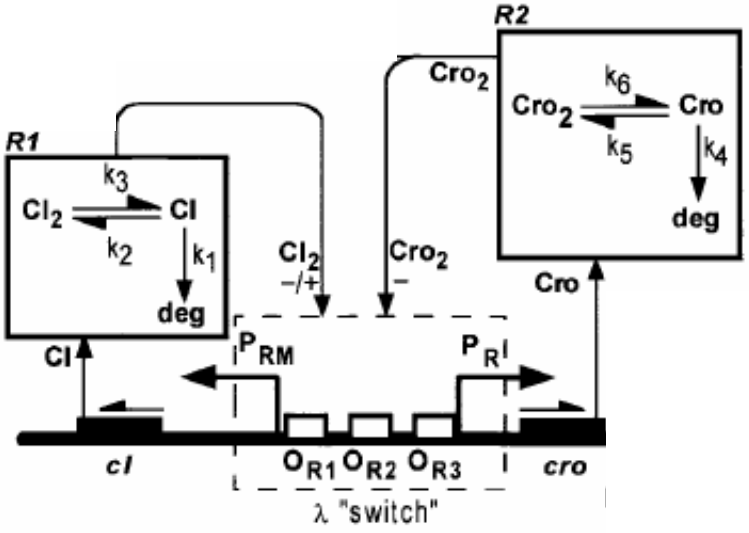
- Regulatory circuit exploits stochastic noise to produce different outcomes
- Stochastic model can predict statistics of regulatory outcomes

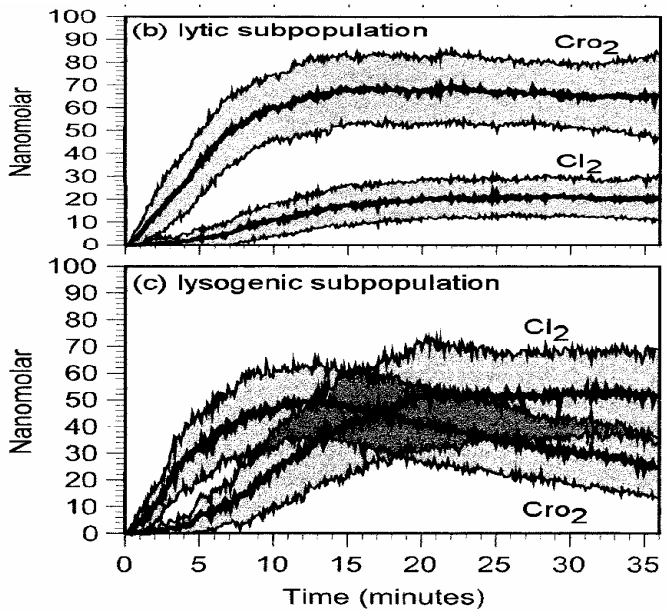
Lambda Phage: Regulatory Circuit



- CI and Cro competitively bind O_{R1}, O_{R2}, O_{R3}
- Cro represses P_R and P_{RM}
- CI represses P_R, can activate P_{RM}

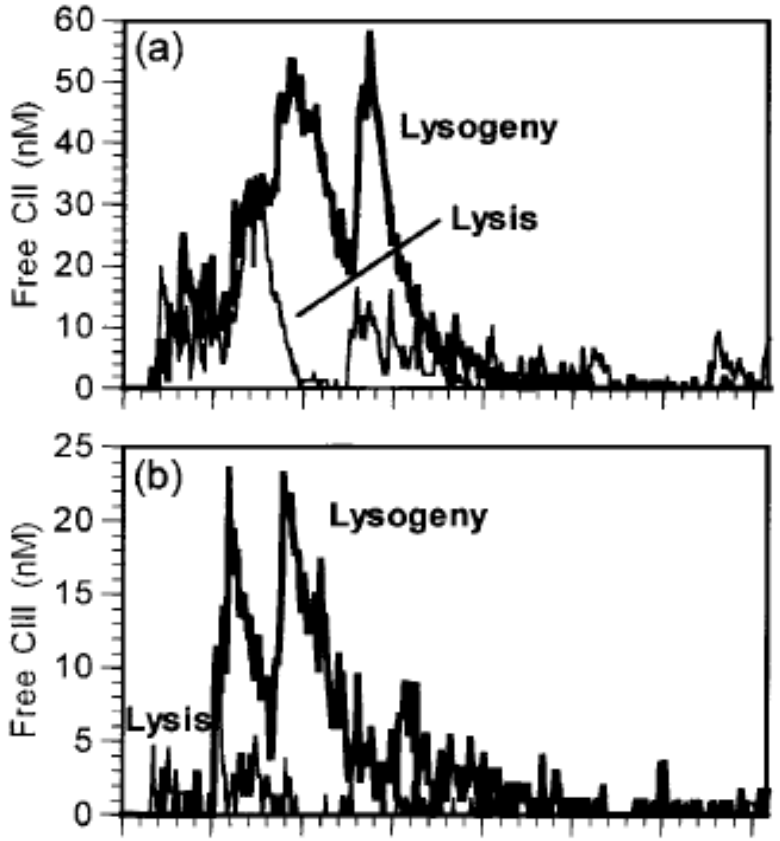
Lambda Phage: Regulatory Circuit

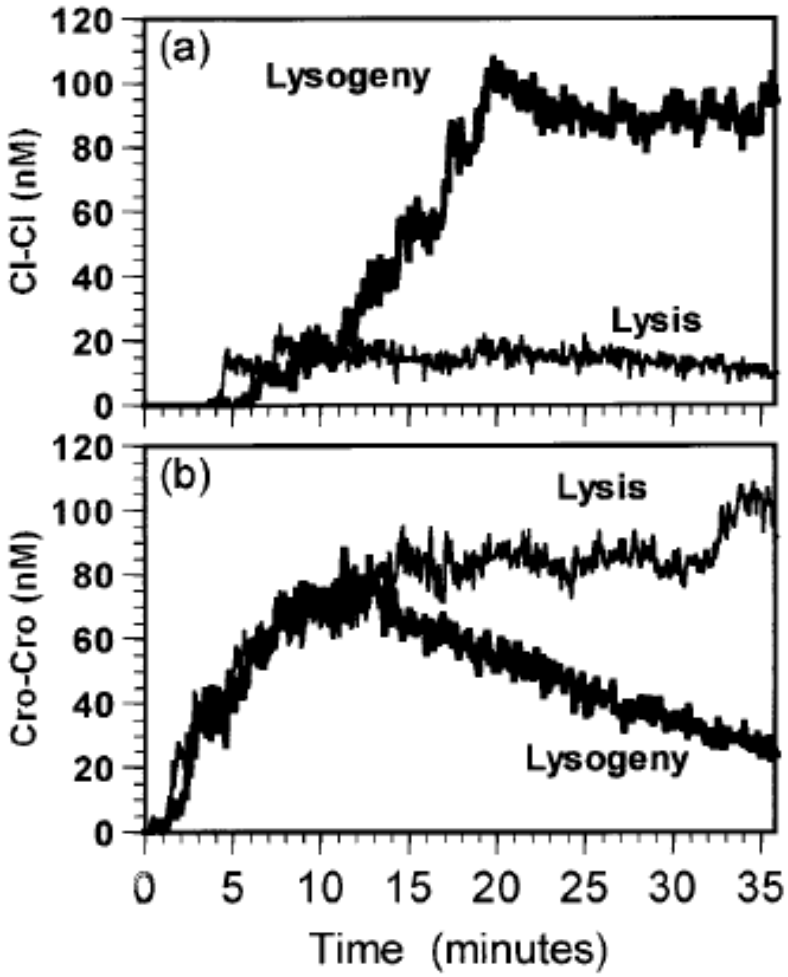




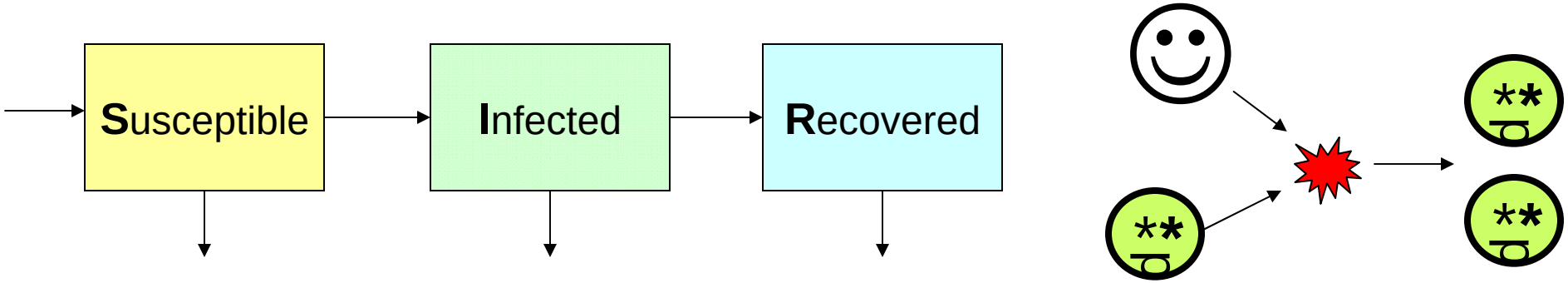
- CI and Cro competitively bind O_{R1}, O_{R2}, O_{R3}
- Cro represses P_R and P_{RM}
- CI represses P_R, can activate P_{RM}

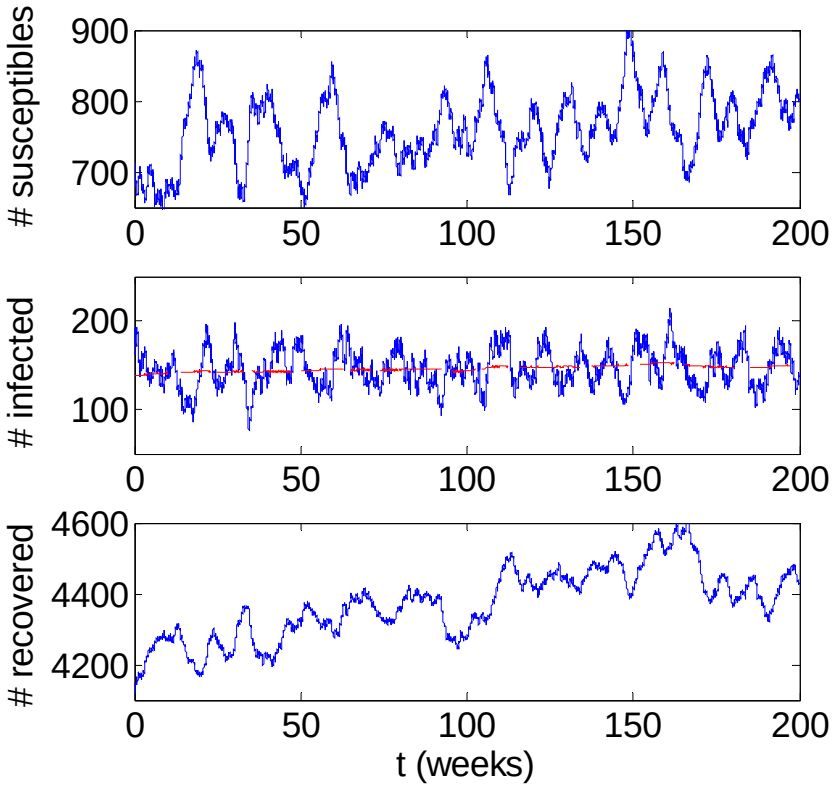
Time Evolution of Two Runs

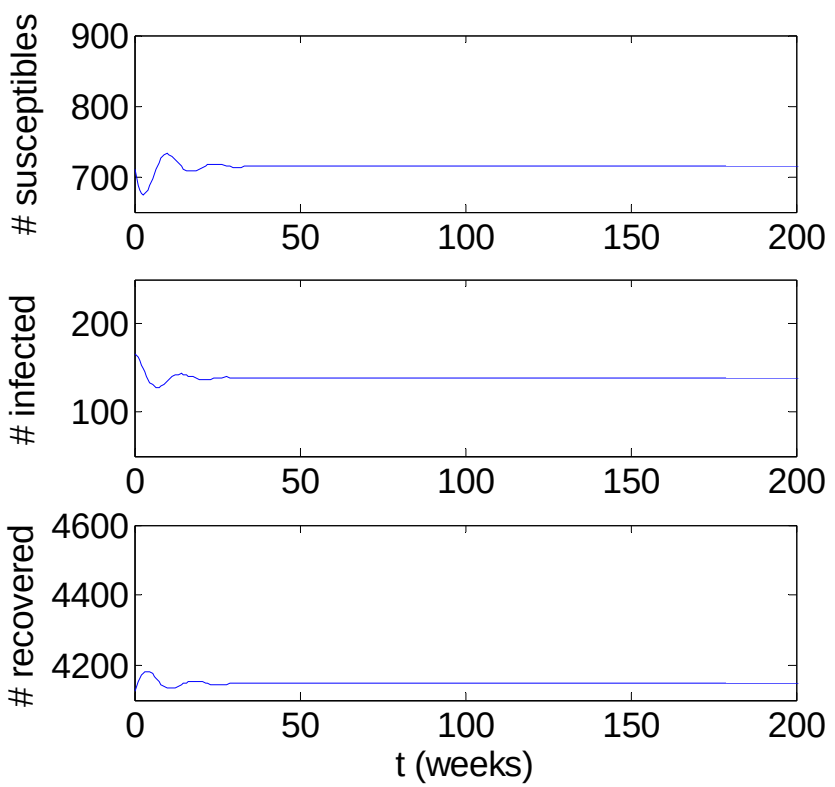




Epidemiology Example







Evaluation of Stochastic Simulation

- Advantages
  - continuous time, discrete population changes
  - captures effects of noise
  - simple implementation
  - small memory requirements
- Disadvantages
  - CPU intensive
  - typically must simulate many runs
  - must use good random number generator
    - periodicity affects size of simulation
    - resolution limits range of probabilities

Computational Requirements

- Memory (N + 2M + 1)
  - N species populations
  - *c* and *a* values for each of M reactions; a_0
- Total time scales with number of reactions that occur
- Operations per reaction:

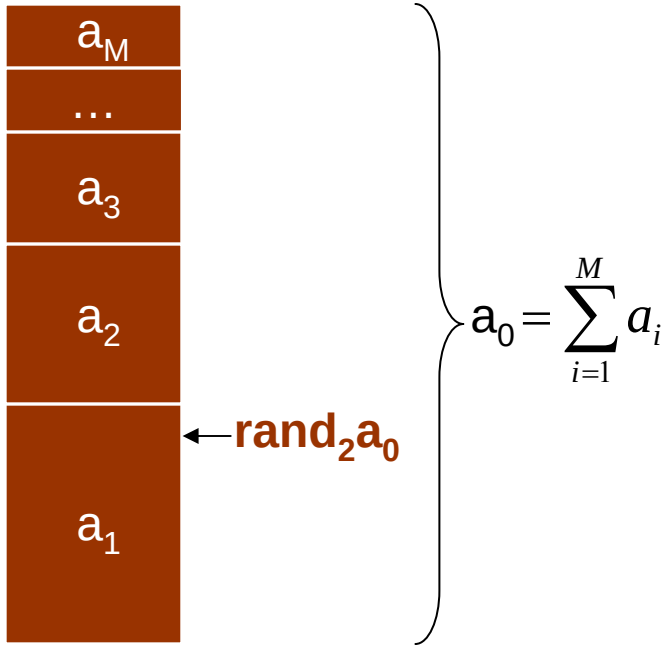
Generate 2 random numbers	
Calculate a_0	**O(M)**
Calculate τ	
Search μ	**O(M)**
Calculate *a* values	**O(M)**

Optimized Direct Method (ODM)

Y. Cao, H. Li, L. Petzold. 2004. J. Chem. Phys. 121:4059-4067

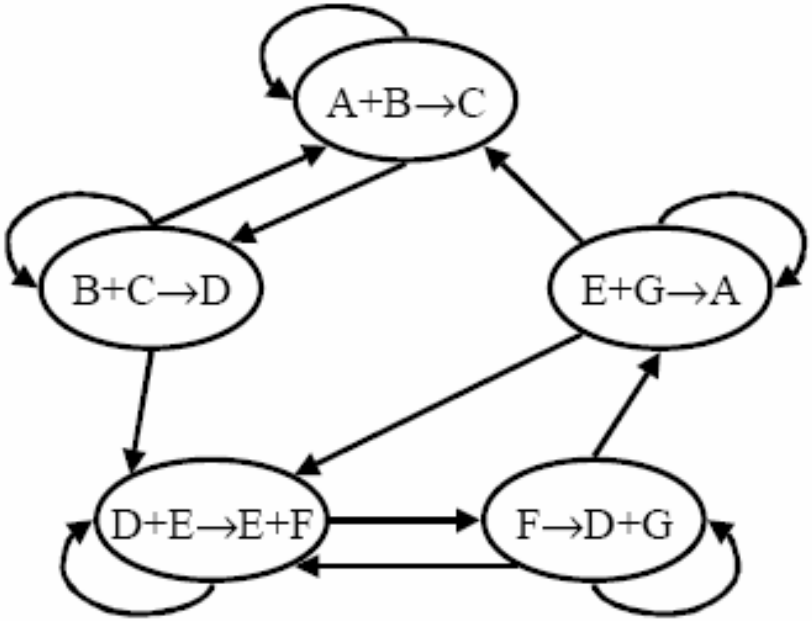
- Reduce cost of searching for index μ
- Observation: Reactions are typically multiscale in a large system. Subset will frequently occur.
- Sort index of reactions based on how often they occur

$$\sum_{i=1}^{\mu-1} a_i < rand_2 a_o < \sum_{i=1}^{\mu} a_i$$



Optimized Direct Method (ODM)

- Reduce cost of calculating all a_μ
- Dependency graph
- Reduce cost of summing all a_μ to calculate a_0
- Modify a_0 by subtracting old values, adding new



Gibson and Bruck, 2000

Conclusion

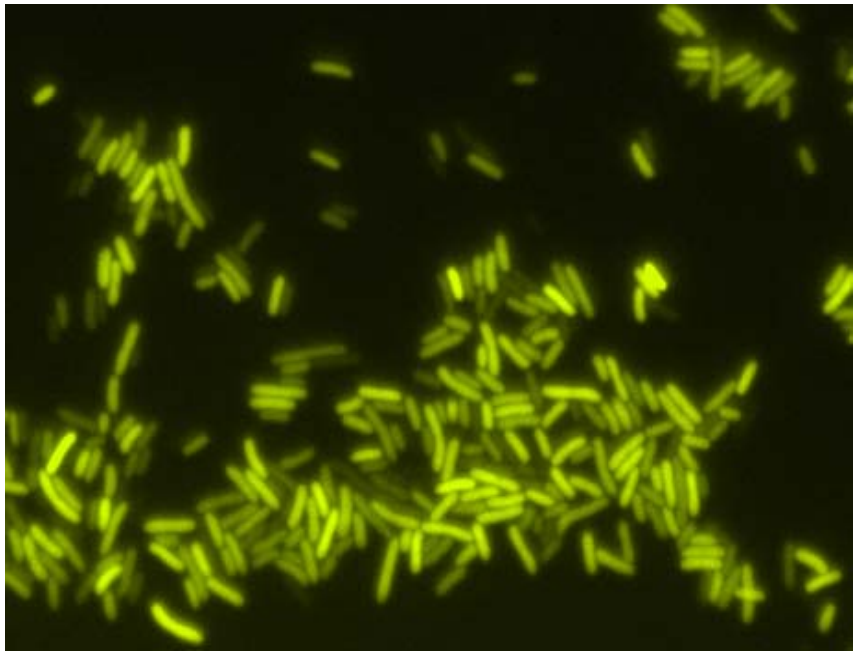
- Provides means of studying role of noise in complex systems
- Can predict statistics (Lambda phage)
- Can depict behavior that deterministic simulations do not capture (epidemiology example)
- Enhancing performance an active area of investigation

Useful References / Further Study

- **Implementations**
  - STOCKS: http://www.sysbio.pl/stocks/stocks1.html
  - BioNetS: https://users.biospice.org/toolsumm.php?id=2
  - http://www.staff.ncl.ac.uk/d.j.wilkinson/software/
- **Algorithms/optimizations**
  - D.T. Gillespie. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. 1976. *J Comput Phys* 22:403-434.
  - D.T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. 1977. *J Phys Chem* 81:2340-2361
  - M.A. Gibson and J. Bruck. 2000. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *J Phys Chem* 104:1876-1889
  - D.T. Gillespie. 2001. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys*. 115:1716-1733.
  - Y. Cao, H. Li and L. Petzold, Efficient formulation of the stochastic simulation algorithm for chemically reacting system. 2004. *J Chem Phys* 121:4059-4067
  - Y. Cao, D. Gillespie, L. Petzold, The slow-scale stochastic simulation algorithm. 2005. *J Chem Phys* 122(1).
  - A. Chatterjee and D.G. Vlachos. Binomial distribution based τ-leap accelerated stochastic simulation. 2005. *J Chem Phys* 122:024112
- **Example Uses**
  - A. Arkin, J. Ross, H. McAdams. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in λ Phage-Infected *Escherichia coli* Cells. 1998. *Genetics* 149:1633-1648
  - J. Dushoff, J.B. Plotkin, S.A. Levin, D.J.D. Earn. Dynamical resonance can account for seasonality of influenza epidemics. 2004 *PNAS*.
  - S. Hooshangi, S. Thiberge, R. Weiss. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. 2005. PNAS 102:3581-3586

Acknowledgements

- Ron Weiss and the Weiss group
- PICASso
  - Steven Kleinstein
  - J.P. Singh



Comparing to Experimental Results

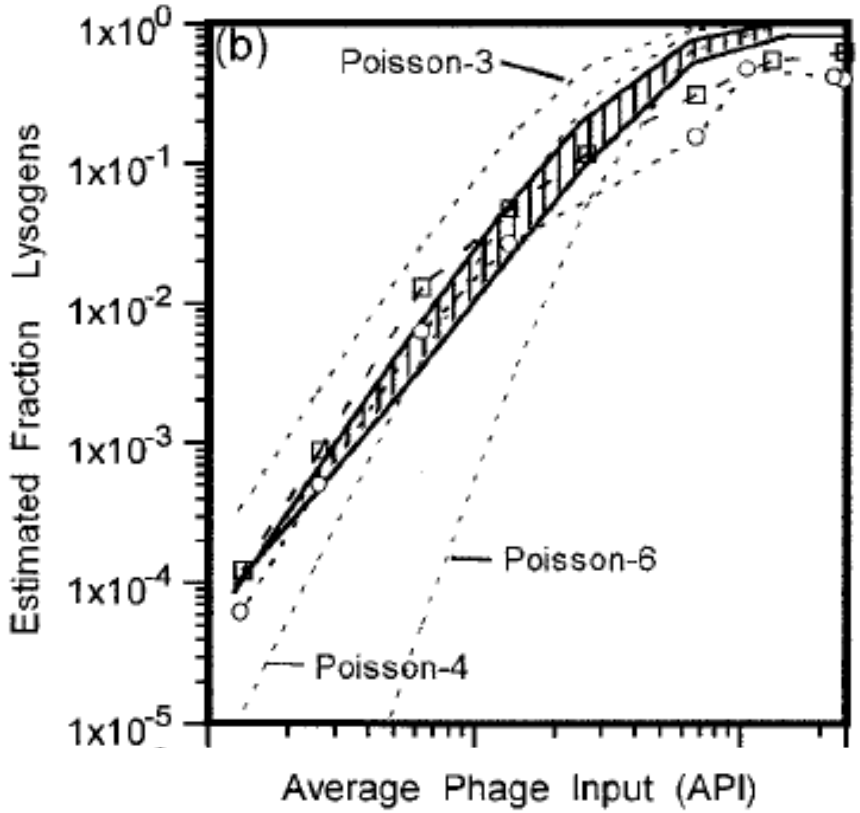
- Compare predicted percent lysogenization to experimental results at different infection levels
- Average phage input (API) = ratio of phage particles to cells at time of infection
- Multiplicity of infection (MOI) = phage particles per cell
- Poisson probability that a given cell will be infected with MOI=M when API=A

$$P(M, A) = \frac{A^M}{A!} e^{-A}$$

- Expected fraction of lysogens:

$$F_{lysogens}(A) = \sum_M P(M, A) \cdot F(M)$$

Comparing to Experimental Results



Example (deterministic assumptions)

$$\mathrm{R}_1: \quad \mathrm{S}_1 + \mathrm{S}_2 \rightarrow 2\mathrm{S}_1 \quad \Longrightarrow \quad \frac{\partial x_1}{\partial t} = k_1 x_1 x_2 \qquad \frac{\partial x_2}{\partial t} = -k_1 x_1 x_2$$

Average rate of R_1 in d*t* per unit volume*:*

$$\frac{\langle X_1 X_2 c_1 \rangle}{V} = \frac{\langle X_1 X_2 \rangle c_1}{V}$$

Using concentrations $x_i = X_i/V$ and dividing by density of reactants:

$$k_1 = \frac{\langle x_1 x_2 \rangle c_1 V}{\langle x_1 \rangle \langle x_2 \rangle}$$

Deterministic assumption:

$$k_1 = c_1 V$$

Master equation

$$\frac{\partial}{\partial t} P(X_1,\ldots,X_N;t) = \sum_{\mu=1}^{M} \left[B_\mu - a_\mu P(X_1,\ldots,X_N;t) \right]$$

- $B_\mu \, dt$ = probability that single R reaction brings us to state $X_1,\ldots,X_N$
- $a_\mu \, dt = h_\mu c_\mu \, dt$, where h_μ is the number of reactant combinations
- Often difficult to solve analytically and even numerically
- →Instead simulate individual trajectories using Monte Carlo algorithm