Stochastic Modeling of Chemical Reactions

Eric Haseltine
Department of Chemical Engineering
University of Wisconsin-Madison

TWMCC

27 September 2001

Outline

- Why stochastic chemical kinetics?
- Stochastic simulation
- Handling reactions with different time scales
- Examples
  1. Simple motivating example
  2. Hepatitis B infection
  3. Particle engineering
- Conclusions
- So what can you do with these tools?

Why stochastic chemical kinetics?

- Stochastic kinetic models treat reactions as molecular events
- Consider the well-mixed reaction:

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C$$

$$\begin{bmatrix} A_o \\ B_o \\ C_o \end{bmatrix} = \begin{bmatrix} 10 \\ 50 \\ 0 \end{bmatrix} \text{ molecules}$$

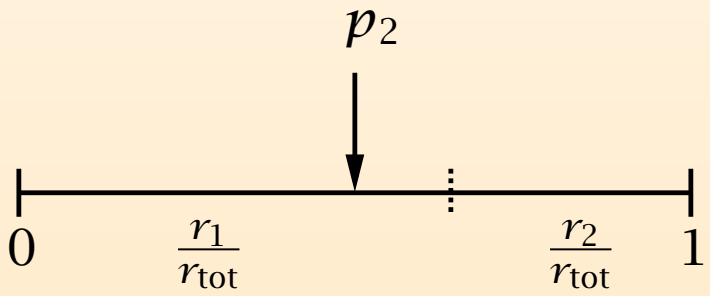
- Scale probabilities by reaction rates
  - $r_1 = k_1 AB$
  - $r_2 = k_{-1} C$
  - $r_{\text{tot}} = r_1 + r_2$

- We randomly select:

1. When the next reaction occurs

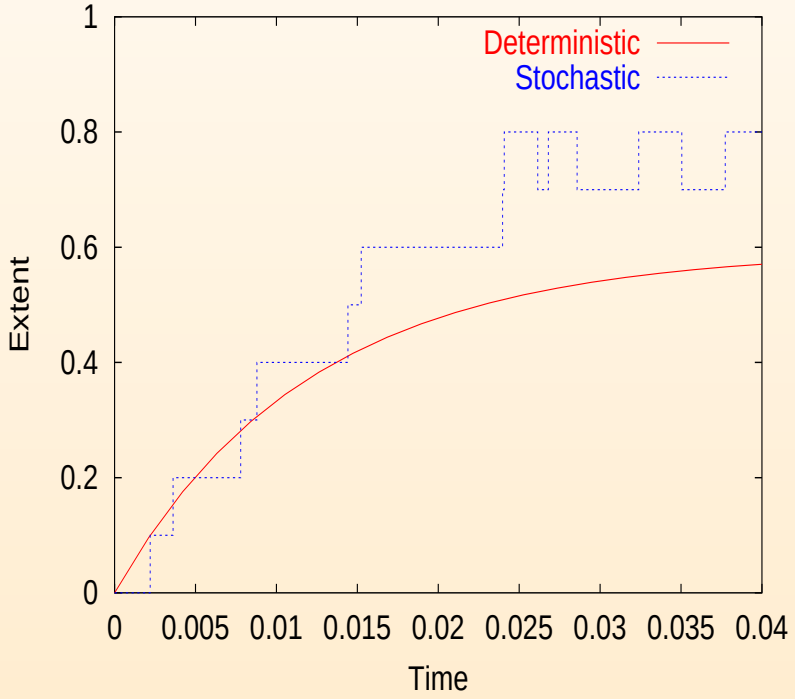
$$\tau = -\frac{\log(p_1)}{r_{\text{tot}}}$$

2. Which reaction occurs

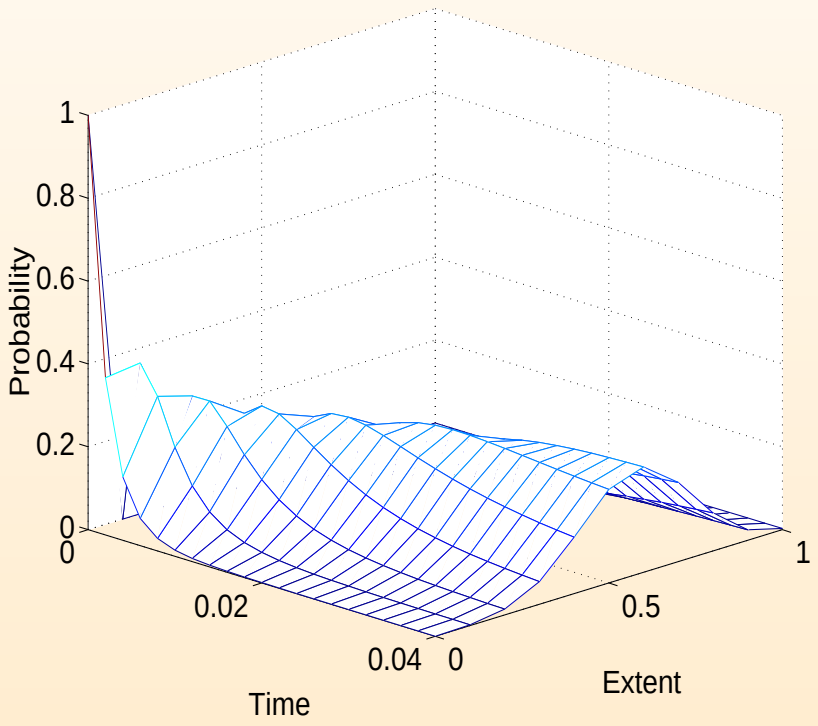


Stochastic Simulation

One simulation

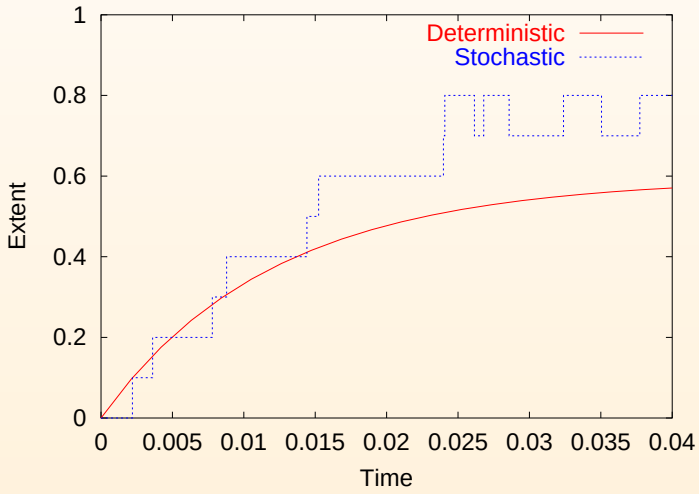


Average of many simulations

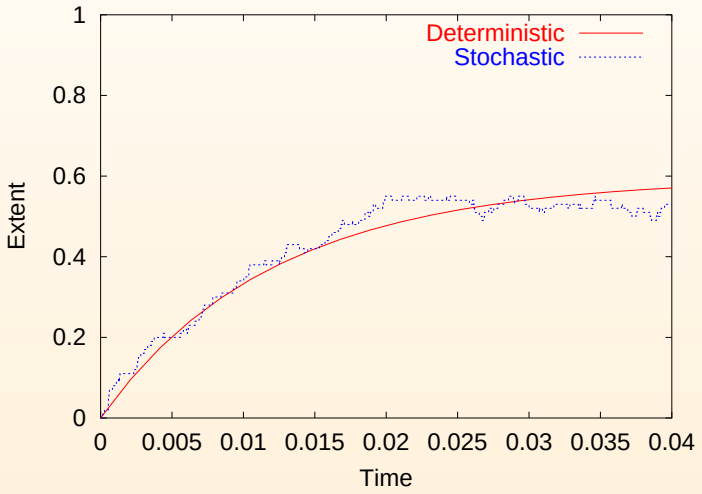


Connection to Deterministic Kinetics (Ω = System Size)

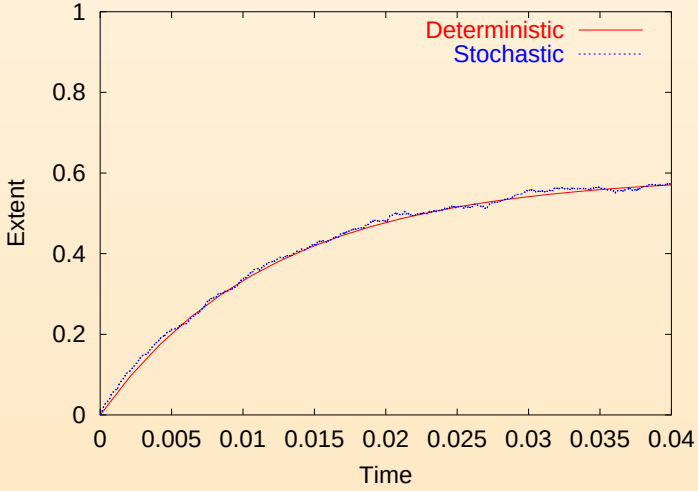
Ω



$\Omega \times 10$

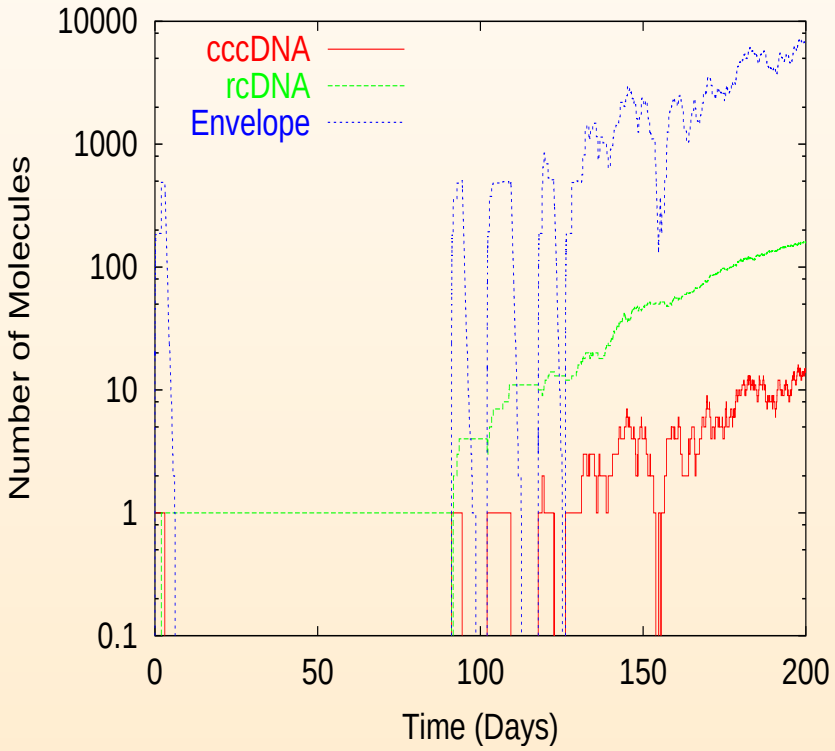


$\Omega \times 100$



As Ω increases:
Stochastic → Deterministic
Computing burden increases!

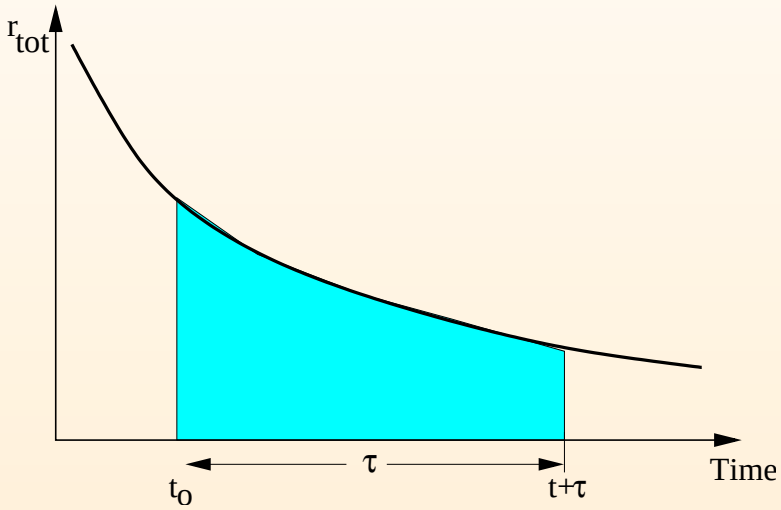
What can be done when reactions occur over drastically different time scales?

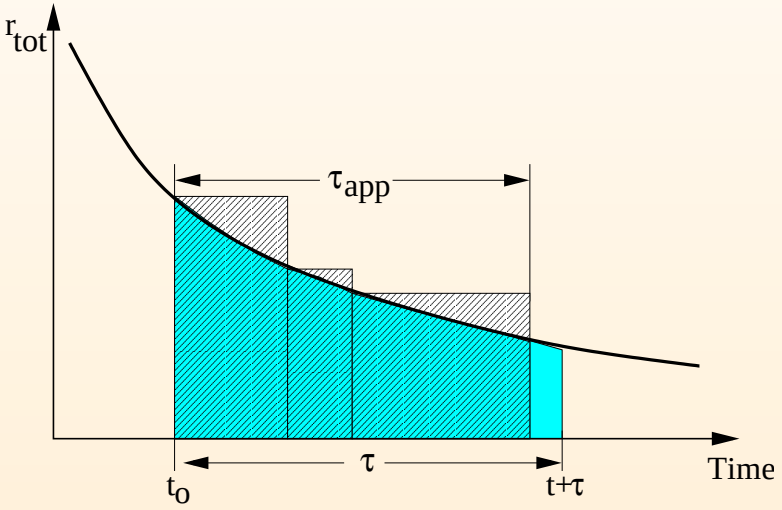


Make a stochastic approximation...

1. Partition the reactions into two subsets:
   - a fast reaction subset
   - a slow reaction subset
2. Fast reactions
   - ODE approximation
   - Eliminates fluctuations
3. Slow reactions
   - Time-varying reaction rates
   - Exact solution or approximate solution?

Solution Techniques





- Exact solution

$$\int_t^{t+\tau} r_{\text{tot}}(t')dt' + \log(p_1) = 0$$

May be computationally expensive to solve!

- Approximate solution
  - Artificially introduce a probability of no reaction, $a_0 dt$
  - Let $\tau \approx -\frac{\log(p_1)}{r_{\text{tot}}}$
  - As $a_0 \to \infty$, this simulation method becomes exact

A Simple Example

Reaction System

$$2A \xrightarrow{k_1} B, \quad \varepsilon_1$$

$$A + C \xrightarrow{k_2} D, \quad \varepsilon_2$$

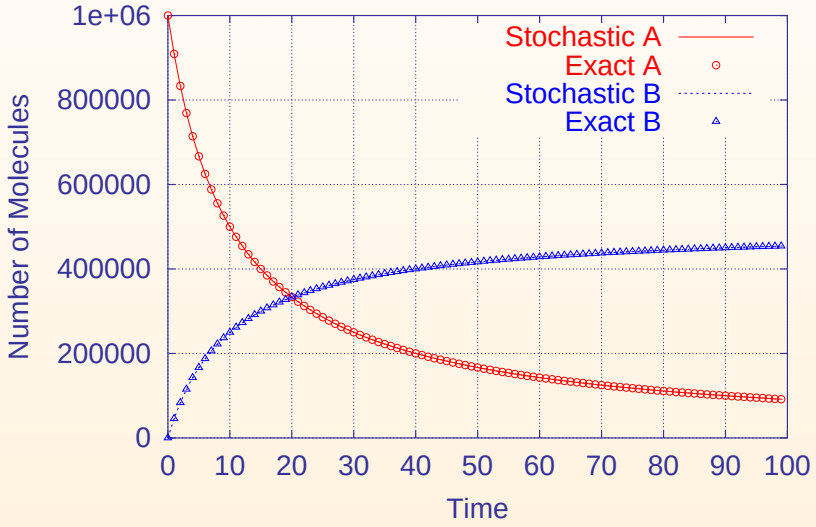
$$k_1 = k_2 = 1\text{E-}7$$

$$\begin{bmatrix} A_0 \\ B_0 \\ C_0 \\ D_0 \end{bmatrix} = \begin{bmatrix} 1\text{E+}6 \\ 0 \\ 10 \\ 0 \end{bmatrix}$$

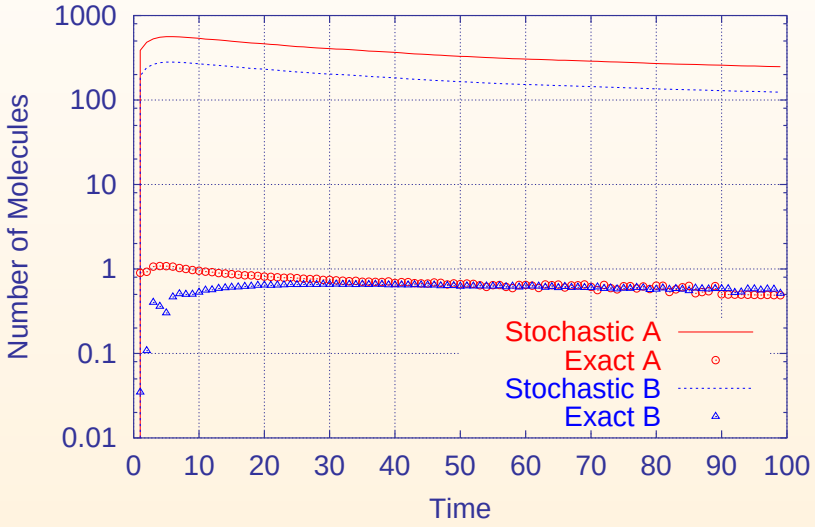
- Approximate ε_1 deterministically
- Reconstruct the mean and standard deviation with 10,000 simulations
- Stochastic simulation CPU time: ~ 8 sec per simulation
- Hybrid simulation CPU time: ~ 0.6 sec per simulation

Exact Stochastic-Deterministic Results

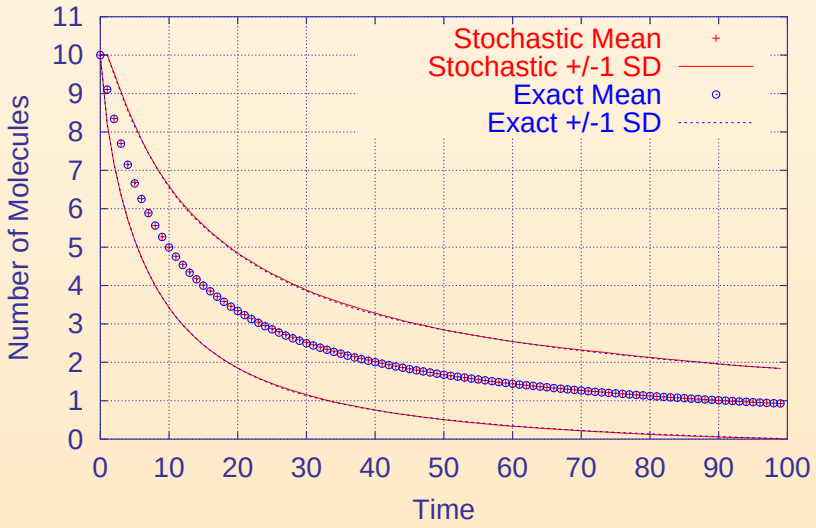
Mean for A and B



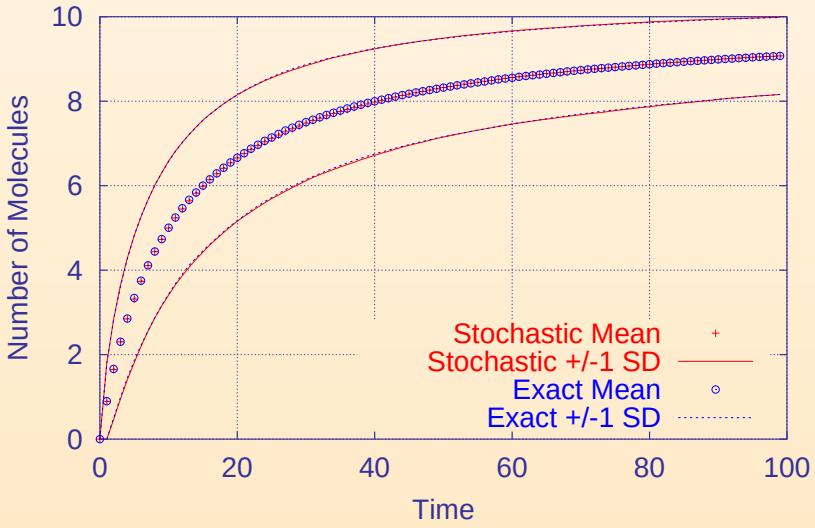
Standard Deviation for A and B



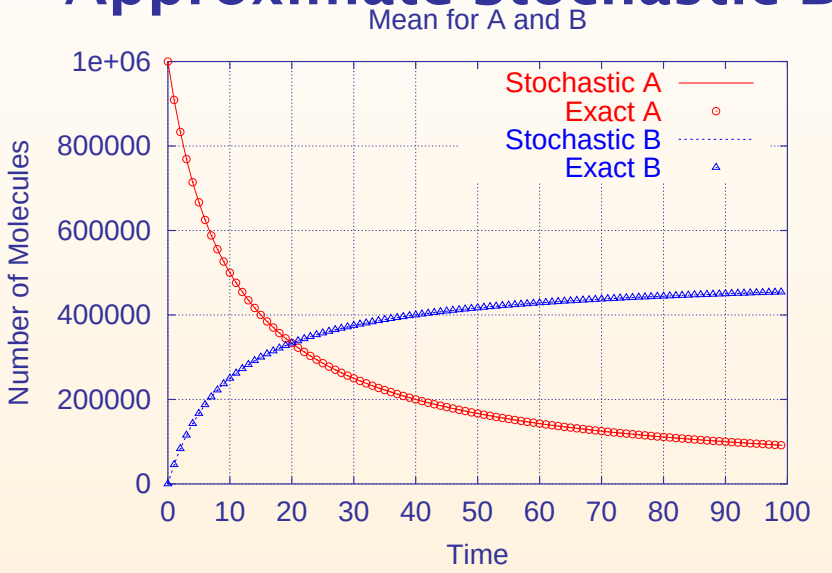
C

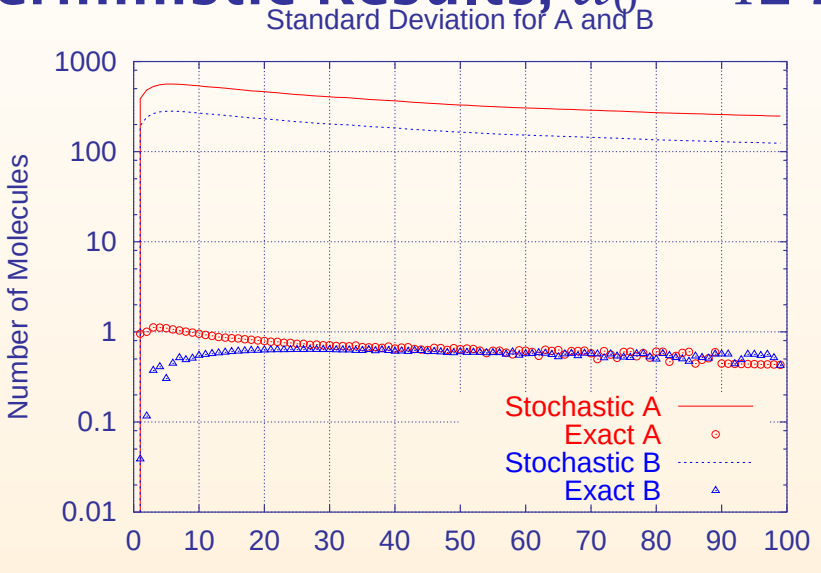


D

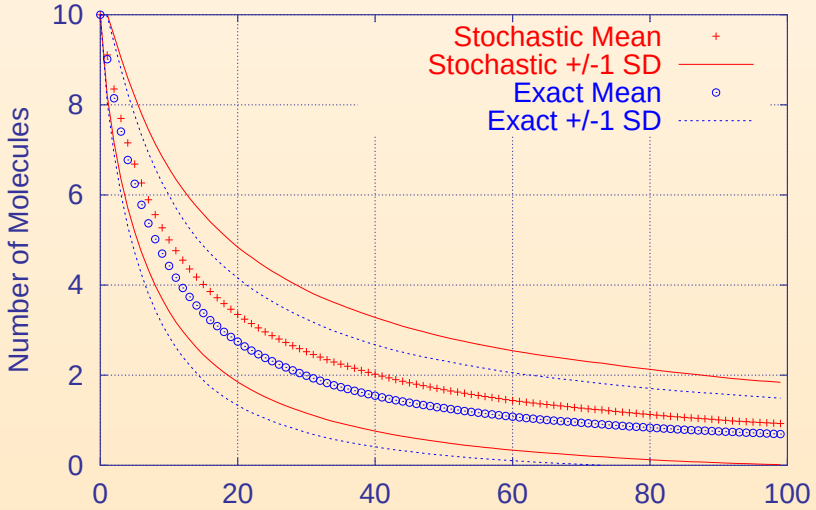


Approximate Stochastic-Deterministic Results, $a_0 = 4\text{E-}2$

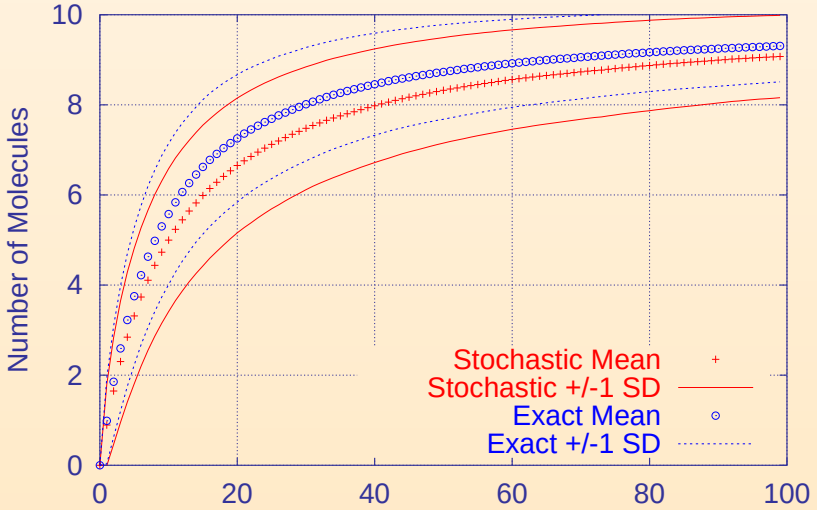




C

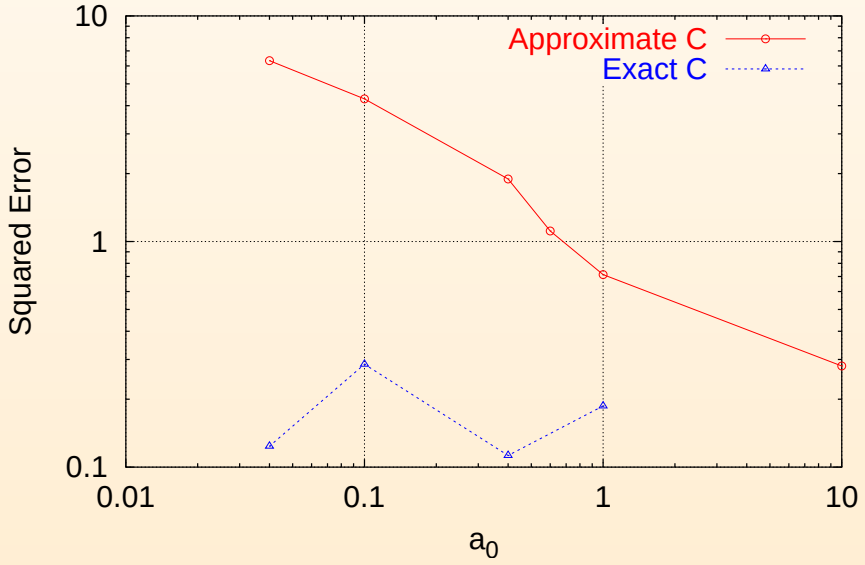


D

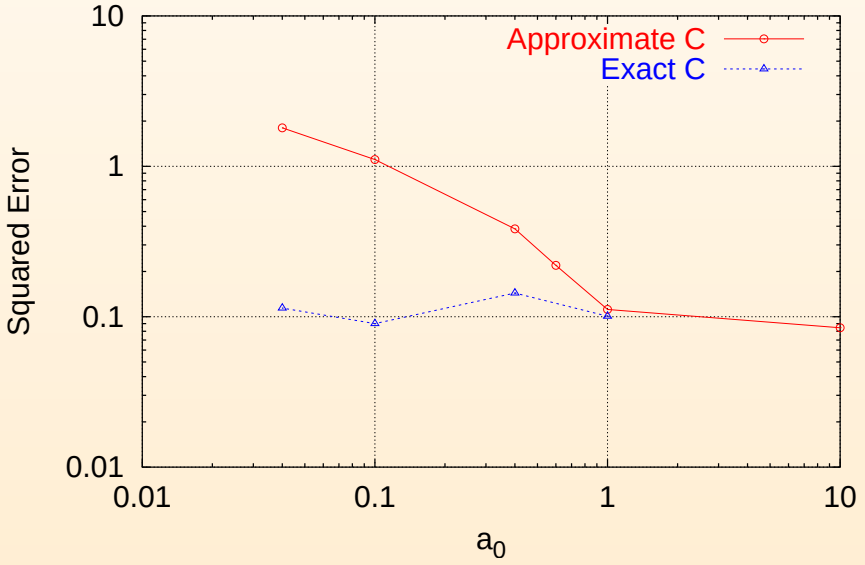


Squared Error Trends

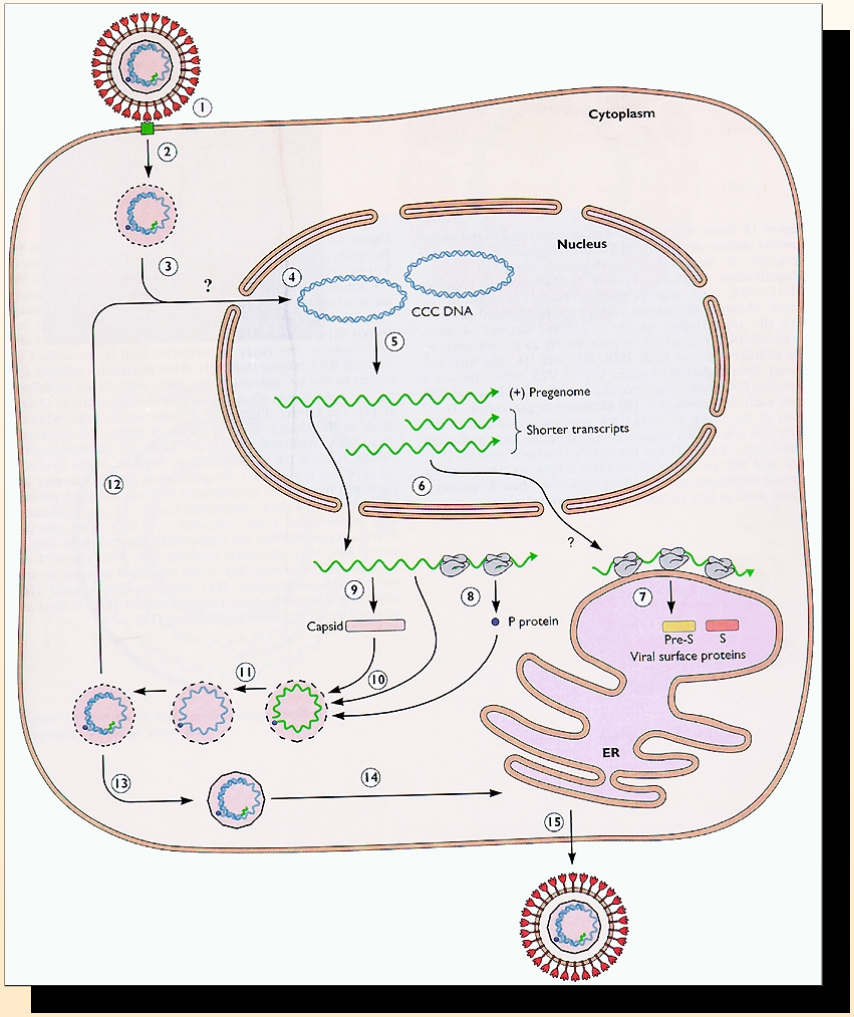
Error in Mean



Error in Standard Deviation



Hepatitis B Infection



- Consider the infection of a cell by a virus
- System model:

$$\text{nucleotides} \xrightarrow{\text{cccDNA}} \text{rcDNA} \quad (1)$$

$$\text{nucleotides} + \text{rcDNA} \longrightarrow \text{cccDNA} \quad (2)$$

$$\text{nucleotides} + \text{amino acids} \xrightarrow{\text{cccDNA}} \text{envelope} \quad (3)$$

$$\text{cccDNA} \longrightarrow \text{degraded} \quad (4)$$

$$\text{envelope} \longrightarrow \text{secreted} \quad (5)$$

$$\text{rcDNA} + \text{envelope} \longrightarrow \text{virus} \quad (6)$$

- Assume:
  1. nucleotides and amino acids are available at constant concentrations
  2. cccDNA catalyzes reactions (1) and (3)

Hepatitis B Infection

- Reduced state
  - $x = \begin{bmatrix} \text{cccDNA} \\ \text{rcDNA} \\ \text{envelope} \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \end{bmatrix}$
  - $x_o = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
  - After ~ 200 days:

$$x_{ss} = \begin{bmatrix} 20 \\ 200 \\ 10000 \end{bmatrix}$$

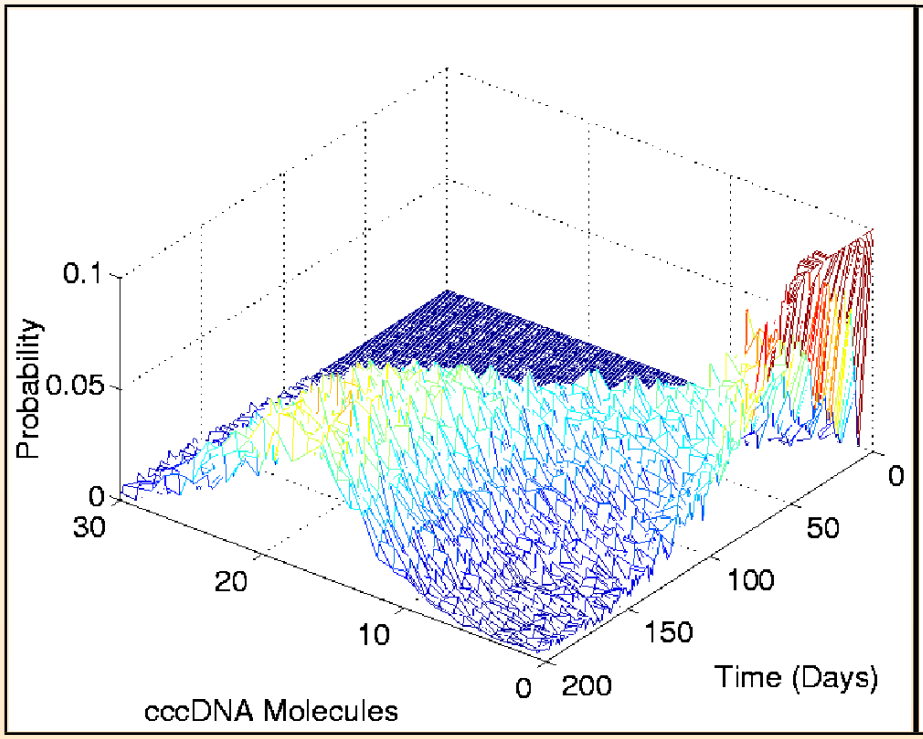
- System reaction channels H:

$$H(x) = \begin{bmatrix} k_1 A \\ k_2 B \\ k_3 A \\ k_4 A \\ k_5 C \\ k_6 BC \end{bmatrix}, \quad \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \\ k_5 \\ k_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.025 \\ 1000 \\ 0.25 \\ 1.9985 \\ 7.5\text{E-}6 \end{bmatrix} \text{day}^{-1}$$

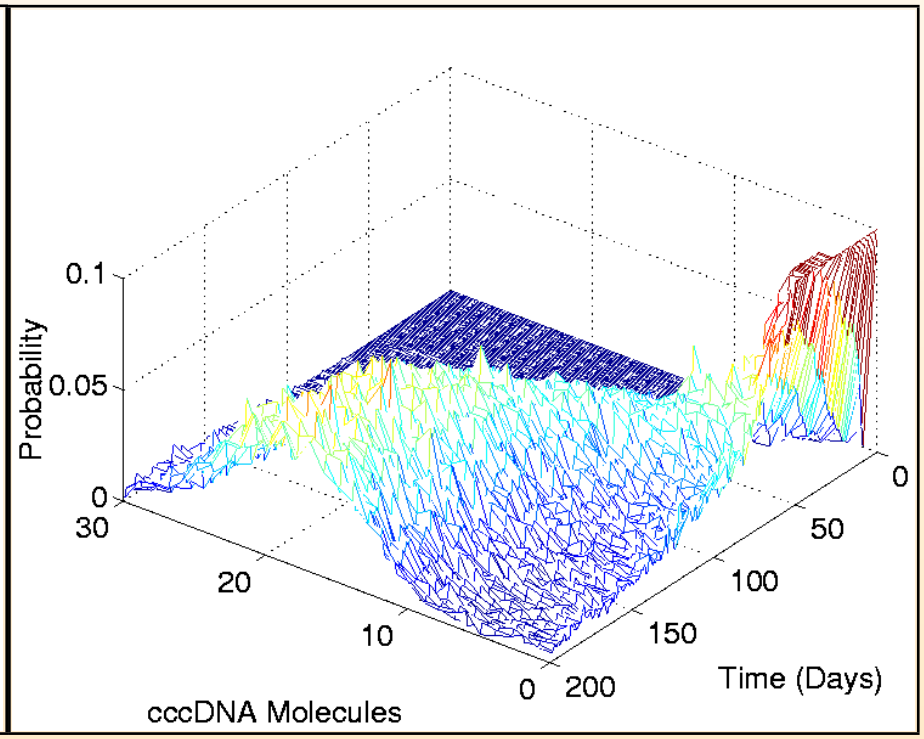
- When $A > 0$ and $C > 100$, reactions (3) and (5) dominate. Partition the system accordingly.

cccDNA Comparisons

Stochastic Results
Average of 1000 Simulations

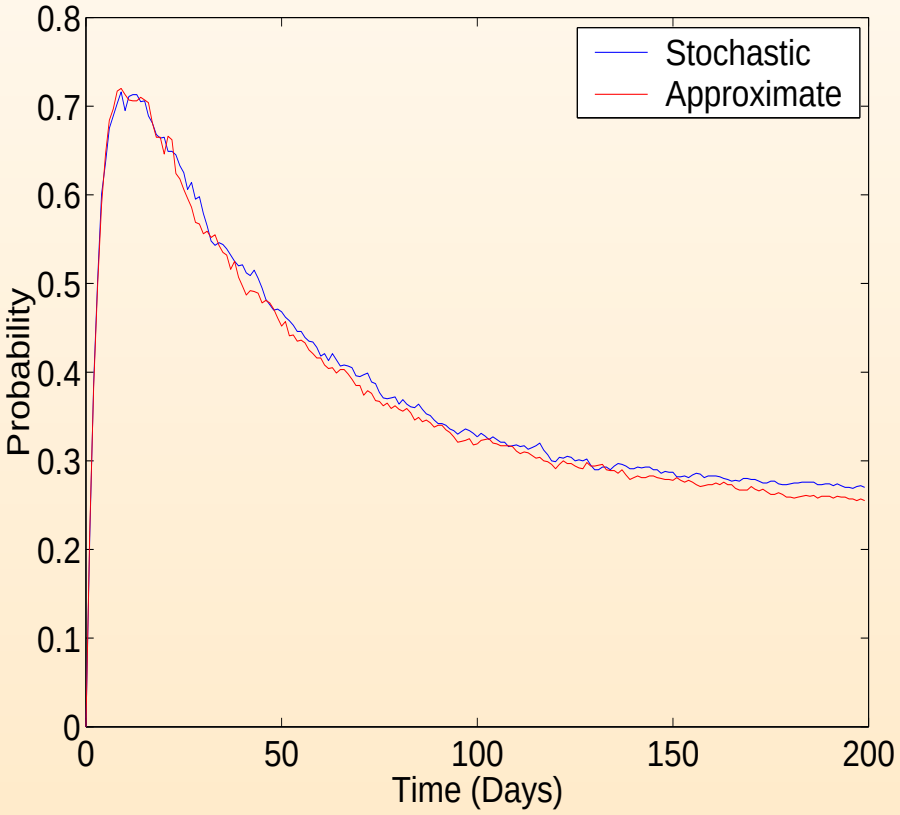


Approximate
Stochastic-Deterministic Results
Average of 1000 Simulations

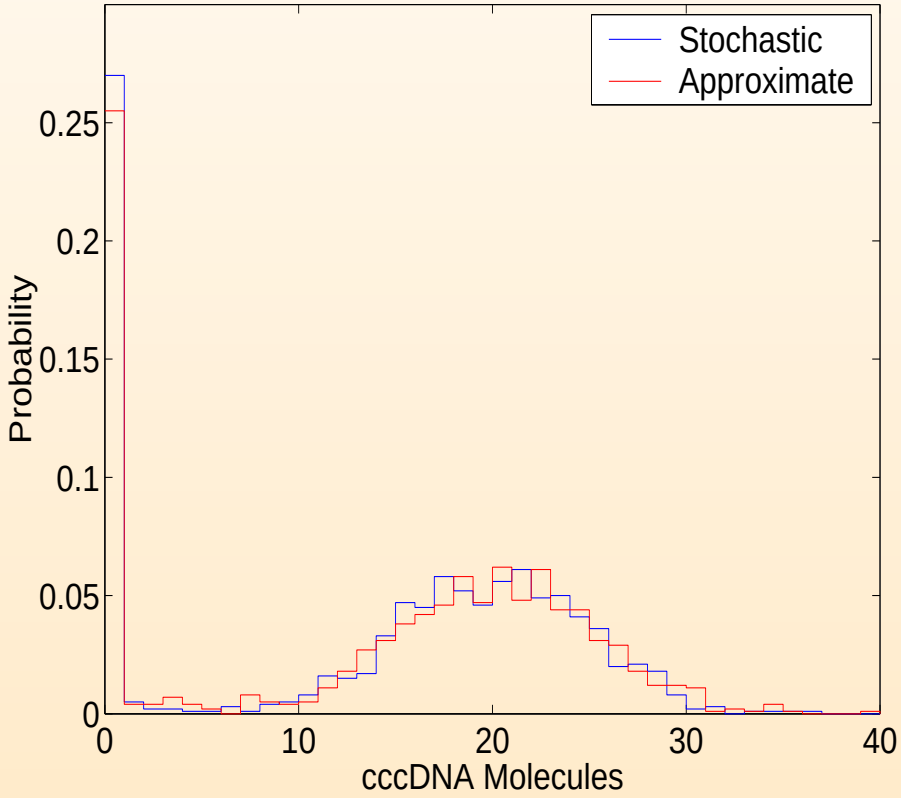


cccDNA Comparisons

Prob(cccDNA = 0,t)

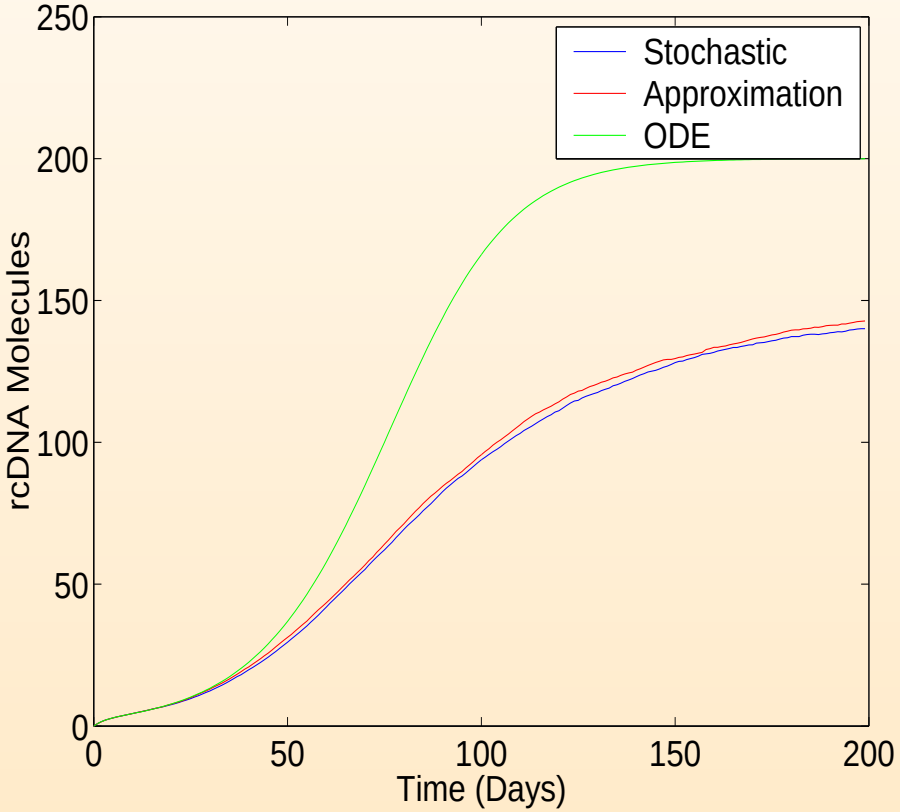


Prob(cccDNA,t = 199 days)

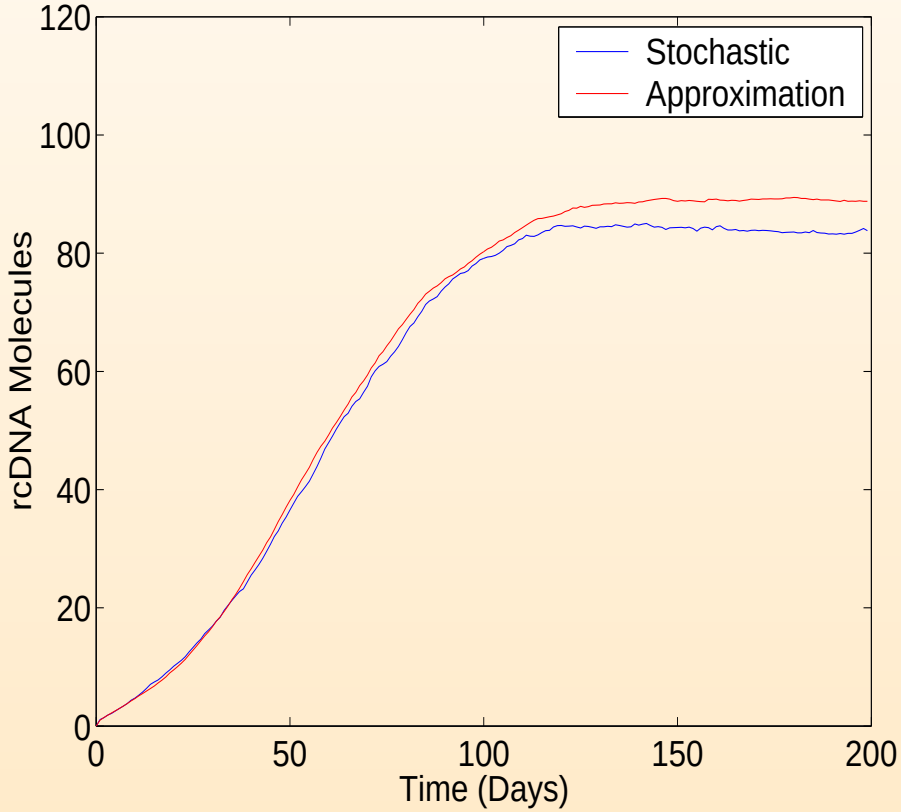


rcDNA Comparisons

Comparison of the Mean



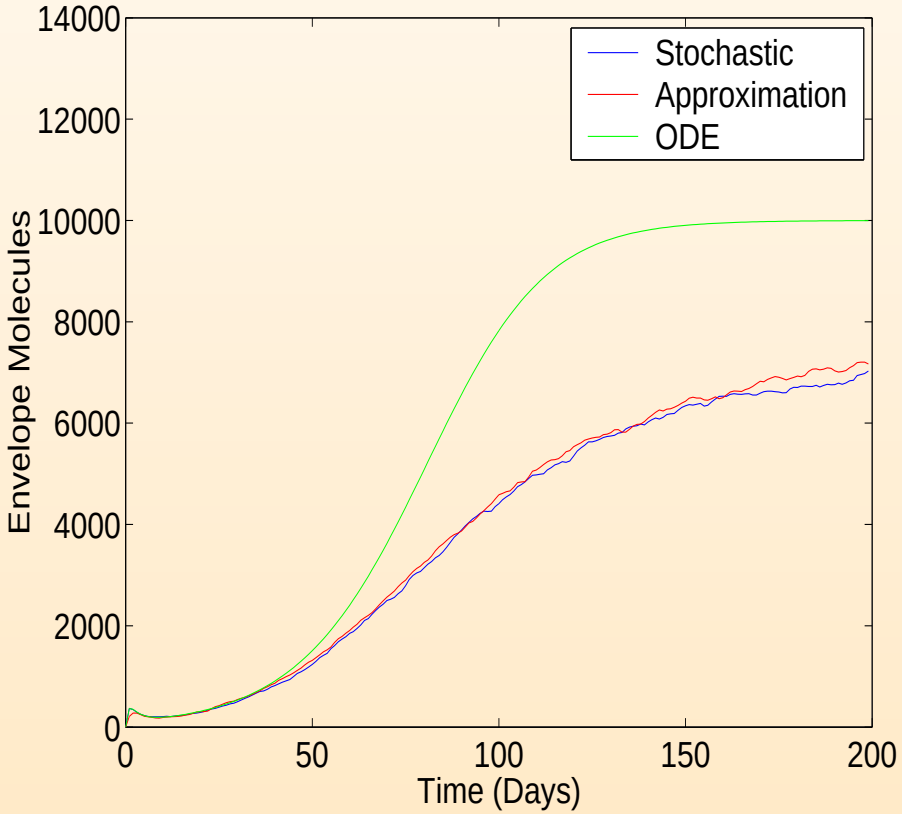
Comparison of the Standard Deviation



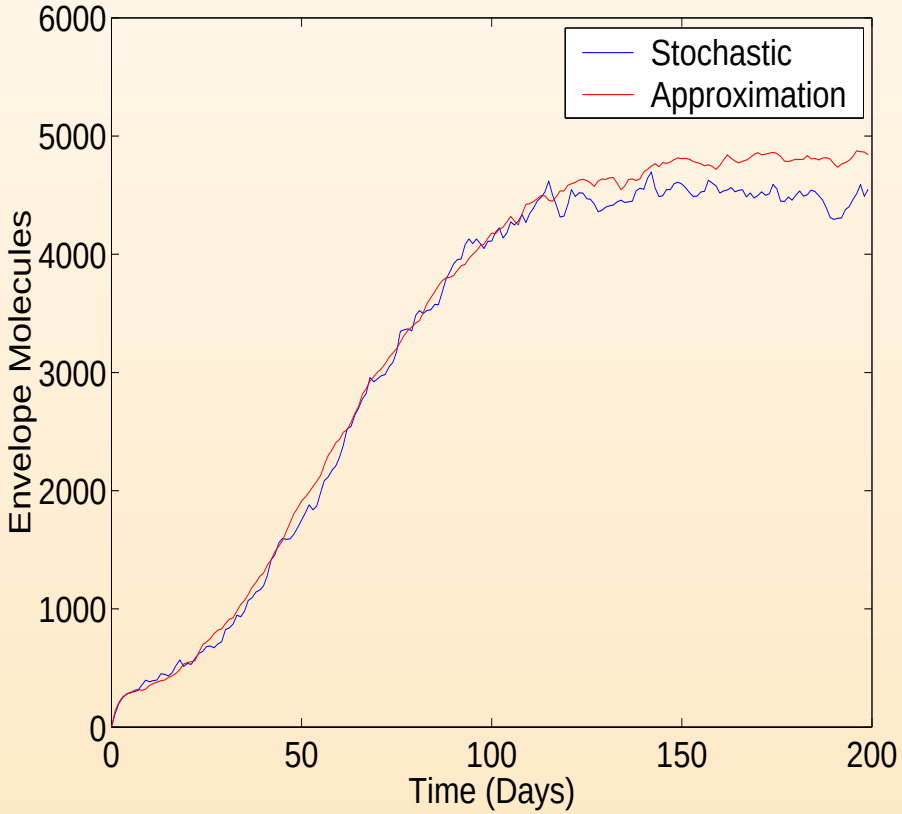
An ODE model gets the mean wrong!

Envelope Comparisons

Comparison of the Mean

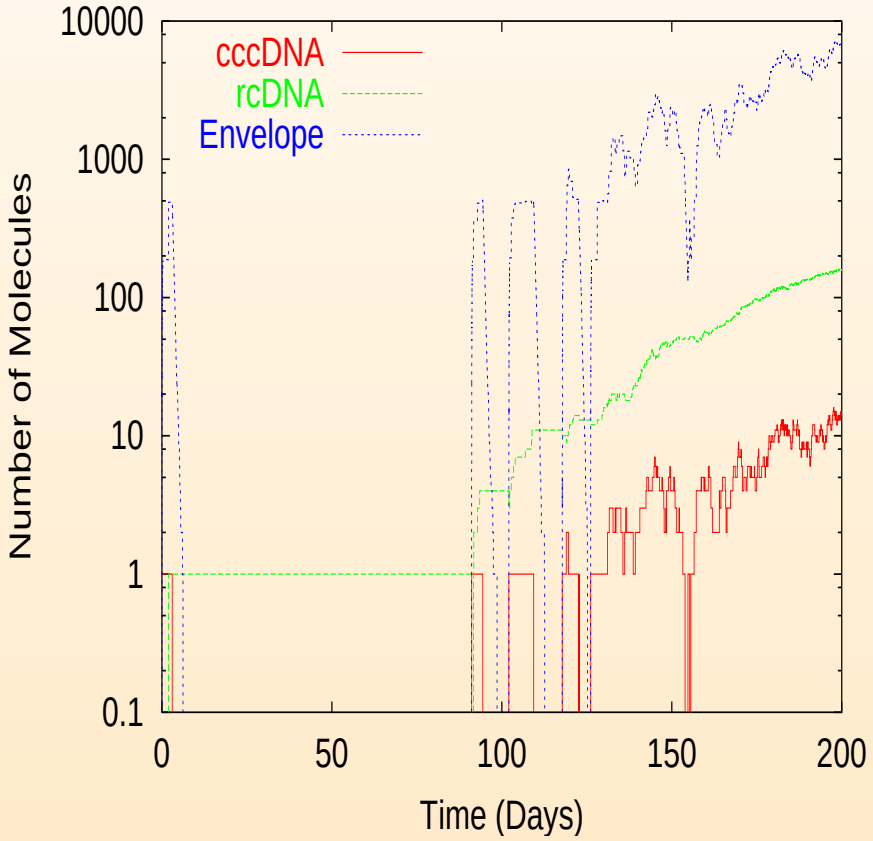


Comparison of the Standard Deviation

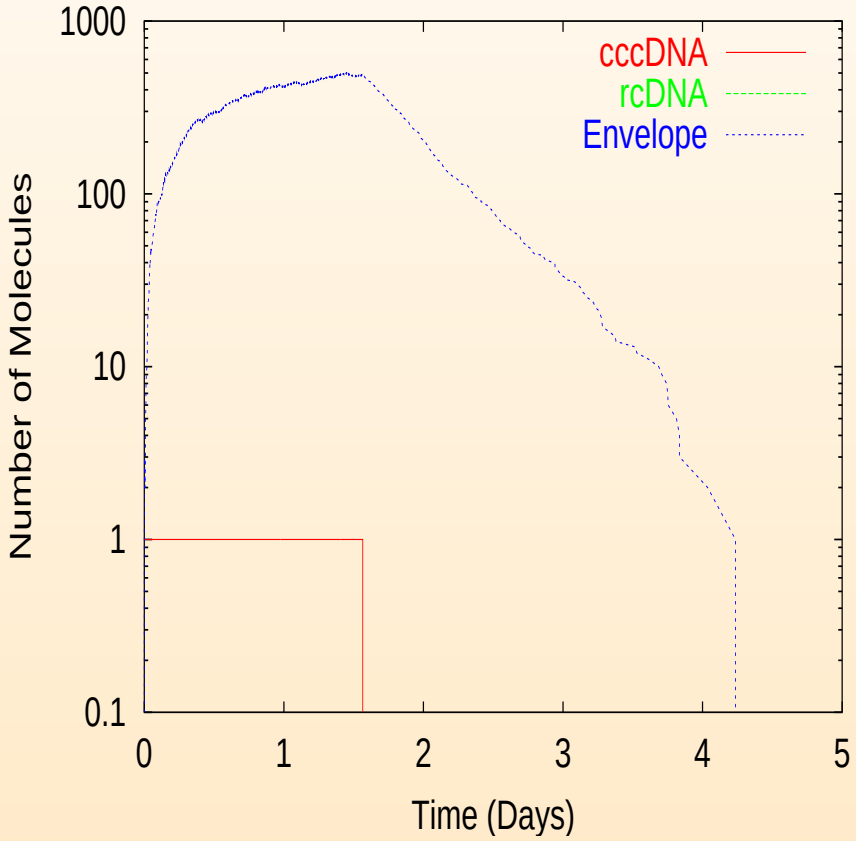


Why the deterministic solution is wrong...

Typical Infection



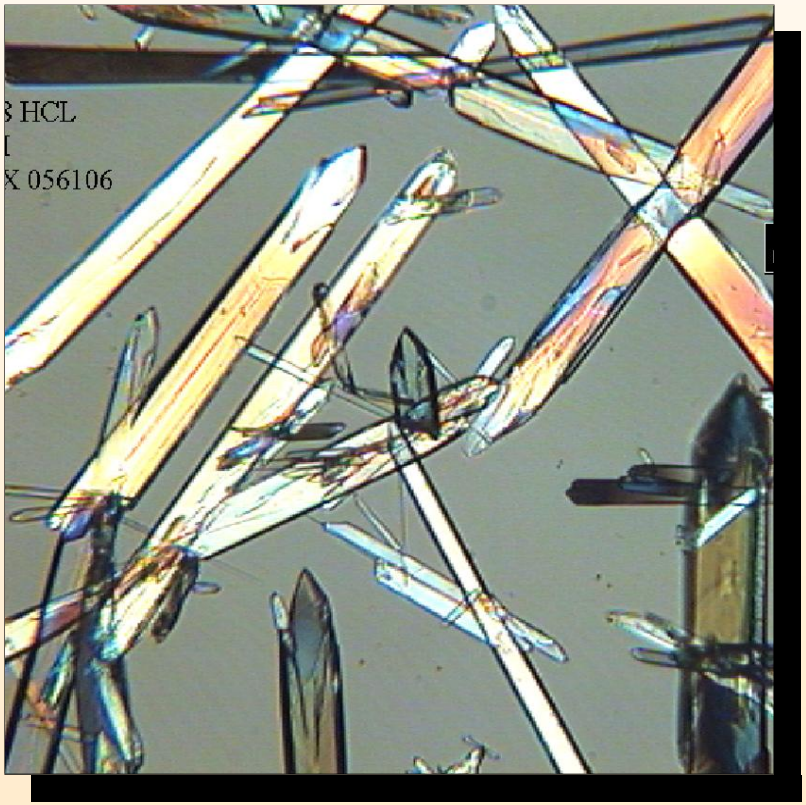
Aborted Infection



The deterministic model cannot express both cases.

Particle Engineering

- Size-independent nucleation and growth
  - Empirical deterministic formulation
  - Stochastic formulation
  - Examples
- Size-independent
  nucleation, growth and agglomeration



more

more

more

more

more

more

more

Empirical Deterministic Formulation

Population Balance

$$\frac{\partial f(L,t)}{\partial t} = -G\frac{\partial f(L,t)}{\partial L}$$

in which $f(L,t)$ is the number of crystals of size L and G is the crystal growth rate.

Mass and Energy Balances

$$\frac{d\hat{C}}{dt} = -3\rho_c k_v h G \int_0^\infty f L^2 dL \qquad \rho V C_p \frac{dT}{dt} = -3\Delta H_c \rho_c k_v V G \int_0^\infty f L^2 dL - UA(T - T_j(t))$$

Nucleation and growth in the bulk

$$B = k_b \left(\frac{\hat{C} - \hat{C}_{sat}(T)}{\hat{C}_{sat}(T)}\right)^b = k_b S^b \qquad G = k_g \left(\frac{\hat{C} - \hat{C}_{sat}(T)}{\hat{C}_{sat}(T)}\right)^g = k_g S^g$$

Stochastic Formulation

Reaction Mechanism for Nucleation and Growth

$$2M \xrightarrow{k_n} N_{l_1} \qquad \Delta H^n_{\text{rxn}}$$

$$N_{l_n} + M \xrightarrow{k_g} N_{l_{n+1}} \qquad \Delta H^g_{\text{rxn}}$$

N_{l_n} is the number of crystals of size l_n per volume.

Mass Balance

$$M_{\text{tot}} = M_{\text{sat}} + M$$

Energy Balance

$$\frac{dT}{dt} = \frac{UA}{\rho C_p V}(T_j - T)$$

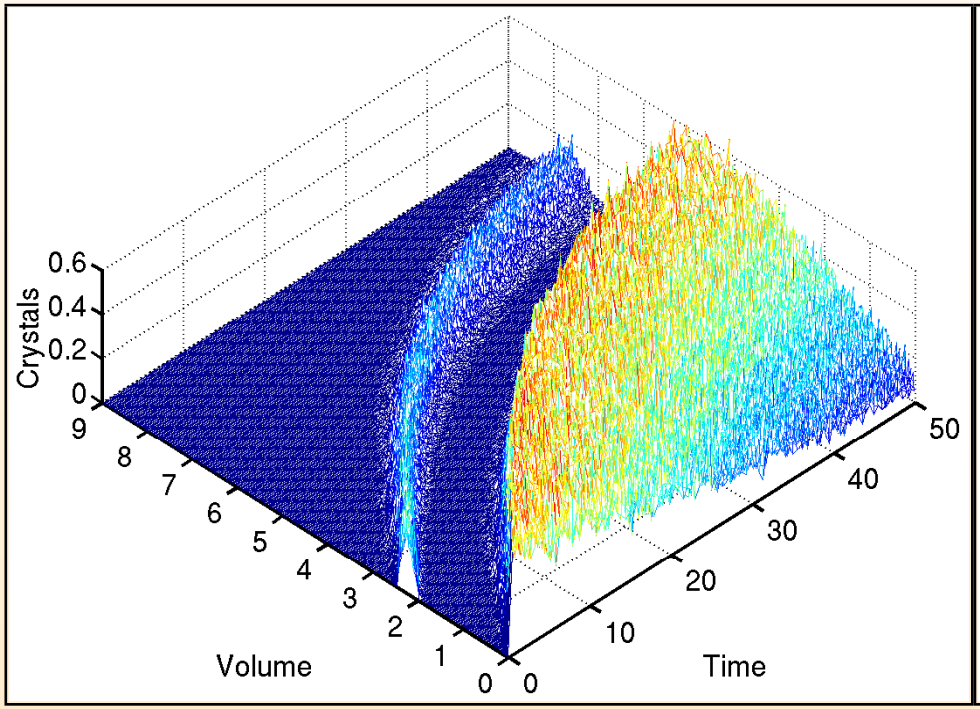
(between reaction events)

Solubility Relationship

$$\log_{10} M_{\text{sat}} = a \log_{10} T + \frac{b}{T} + c$$

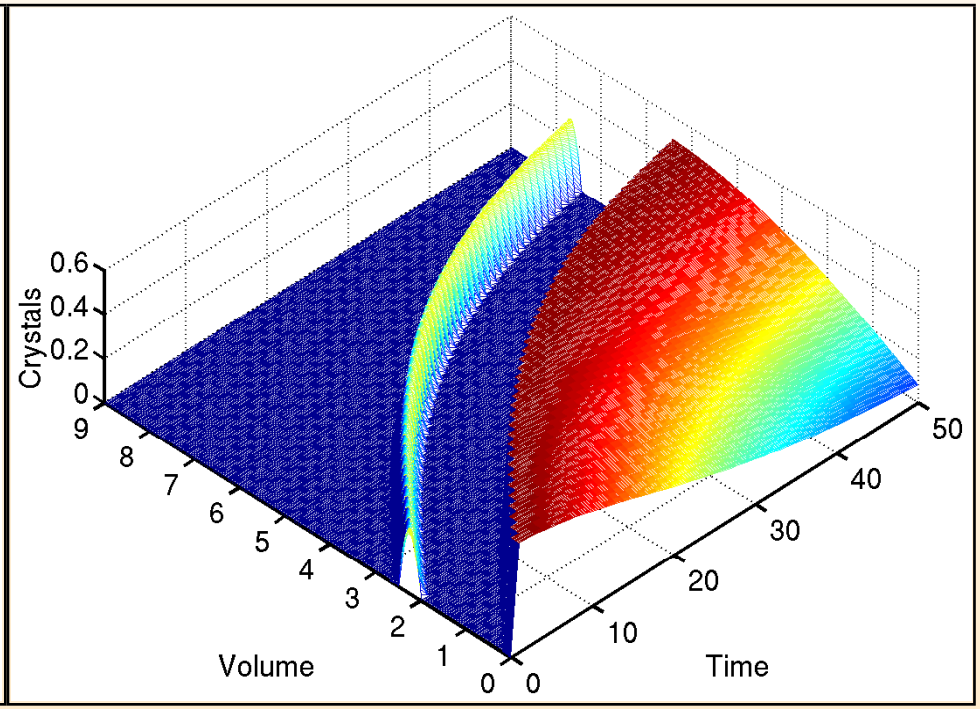
Isothermal, Size-Independent Nucleation and Growth

Stochastic Solution
Average of 100 Simulations



Discrete particle sizes
Integer-valued particle accounting

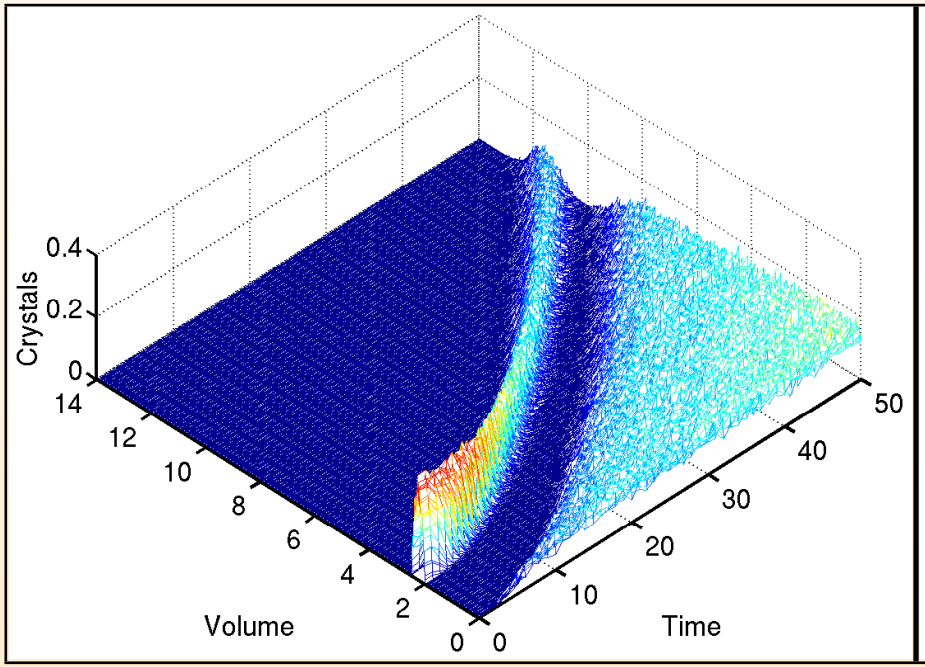
Deterministic Solution
Via Orthogonal Collocation



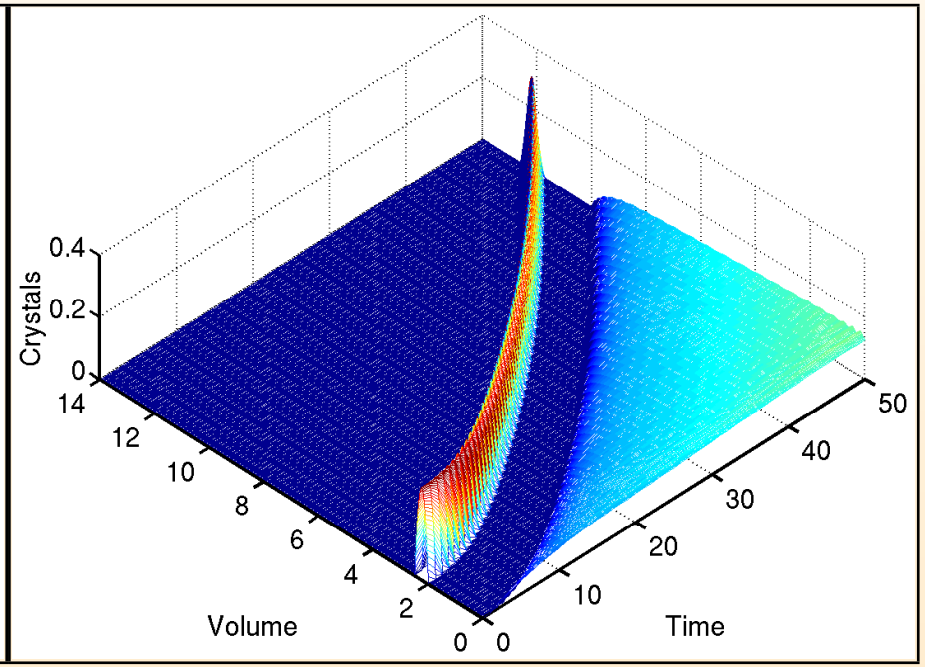
Continuous particle sizes
Real-valued particle accounting

Nonisothermal, Size-Independent Nucleation and Growth

Stochastic Solution
Average of 500 Simulations



Deterministic Solution
Via Orthogonal Collocation

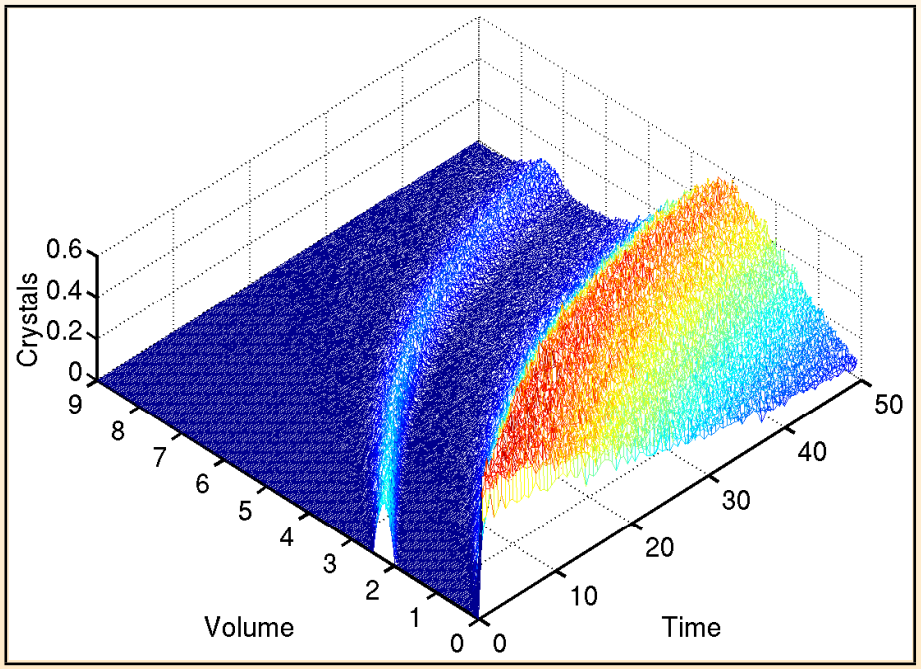


Isothermal, Size-Independent Nucleation, Growth, and Agglomeration

Stochastic Solution

Add one reaction: $N_{l_p} + N_{l_q} \xrightarrow{k_a} N_{l_{p+q}}$

Average of 500 Simulations



Deterministic Solution

Via Adaptive Mesh Methods?

Large time investment!

Conclusions

Stochastic models are important when:

1. Fluctuations in the numbers of particles are important

2. In regions where multiple steady states are possible

3. Solution or formulation of the deterministic problem is difficult

Our advances to the current technology:

1. New modeling approximation for handling reactions occurring over different time scales

2. Application of modeling techniques

So what can you do with these tools?

- Wide array of possible applications
  - Biotechnology
  - Particle engineering
- Bridging the gap from the microscopic to the macroscopic
  - Engineering at interfaces
  - Modeling site interactions on catalysts
  - Nanomaterials
- Further insight into the estimation problem
  - Better understanding of:
    1. State estimation for fluctuating systems
    2. Chemical reaction models
  - Other approaches for computing conditional densities of nonlinear dynamical systems