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:
  \[ A + B \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} C \]
  \[
  \begin{bmatrix}
  A_o \\
  B_o \\
  C_o
  \end{bmatrix} = \begin{bmatrix}
  10 \\
  50 \\
  0
  \end{bmatrix} \text{ molecules}
  \]
- Scale probabilities by reaction rates
  - \( r_1 = k_1AB \)
  - \( r_2 = k_{-1}C \)
  - \( r_{\text{tot}} = r_1 + r_2 \)
- We randomly select:
  1. \textbf{When} the next reaction occurs
     \[ \tau = -\frac{\log(p_1)}{r_{\text{tot}}} \]
  2. \textbf{Which} reaction occurs

\[ \begin{array}{c|c|c|c}
  & 0 & \frac{r_1}{r_{\text{tot}}} & \frac{r_2}{r_{\text{tot}}} & 1 \\
  \hline
  p_2 & & & & \\
  \end{array} \]
Stochastic Simulation

One simulation

Average of many simulations

Extents and times for deterministic and stochastic simulations.
Connection to Deterministic Kinetics ($\Omega = \text{System Size}$)

As $\Omega$ increases:
Stochastic → Deterministic
Computing burden increases!
What can be done when reactions occur over drastically different time scales?

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?

Make a stochastic approximation...
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

\[ 2A \xrightarrow{k_1} B, \quad \varepsilon_1 \]
\[ A + C \xrightarrow{k_2} D, \quad \varepsilon_2 \]

\[ k_1 = k_2 = 1E-7 \]

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

- Approximate \( \varepsilon_1 \) deterministically
- Reconstruct the mean and standard deviation with 10,000 simulations
- Stochastic simulation CPU time: \( \sim 8\) sec per simulation
- Hybrid simulation CPU time: \( \sim 0.6\) sec per simulation
**Exact Stochastic-Deterministic Results**

**Mean for A and B**

- **Stochastic A**
- **Exact A**
- **Stochastic B**
- **Exact B**

**Standard Deviation for A and B**

- **Stochastic A**
- **Exact A**
- **Stochastic B**
- **Exact B**

---

**C**

- **Stochastic Mean**
- **Stochastic +/-1 SD**
- **Exact Mean**
- **Exact +/-1 SD**

**D**

- **Stochastic Mean**
- **Stochastic +/-1 SD**
- **Exact Mean**
- **Exact +/-1 SD**
Approximate Stochastic-Deterministic Results, $a_0 = 4E-2$

Mean for A and B

Standard Deviation for A and B

Stochastic Mean
Stochastic +/-1 SD
Exact Mean
Exact +/-1 SD
Squared Error Trends

Error in Mean

Error in Standard Deviation

Squared Error

\( a_0 \)

Approximate C
Exact C

Squared Error

\( a_0 \)
Hepatitis B Infection

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

\[
\begin{align*}
\text{nucleotides} & \xrightarrow{cccDNA} \text{rcDNA} & (1) \\
\text{nucleotides} + \text{rcDNA} & \rightarrow \text{cccDNA} & (2) \\
\text{nucleotides} + \text{amino acids} & \xrightarrow{cccDNA} \text{envelope} & (3) \\
\text{cccDNA} & \rightarrow \text{degraded} & (4) \\
\text{envelope} & \rightarrow \text{secreted} & (5) \\
\text{rcDNA} + \text{envelope} & \rightarrow \text{virus} & (6)
\end{align*}
\]

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

- Reduced state

\[ x = \begin{bmatrix} \text{cccDNA} \\ \text{rcDNA} \\ \text{envelope} \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \end{bmatrix} \]

- System reaction channels \( H \):

\[ H(x) = \begin{bmatrix} k_1A \\ k_2B \\ k_3A \\ k_4A \\ k_5C \\ k_6BC \end{bmatrix}, \quad \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \\ k_5 \\ k_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.025 \\ 1000 \\ 0.25 \\ 1.9985 \\ 7.5E-6 \end{bmatrix} \text{ day}^{-1} \]

- After \( \sim 200 \) days:

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

- 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)
An ODE model gets the mean wrong!

Comparison of the Mean

Comparison of the Standard Deviation
Envelope Comparisons

Comparison of the Mean

Comparison of the Standard Deviation
Why the deterministic solution is wrong...

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
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

\[ \frac{d\hat{C}}{dt} = -3 \rho_c k_v h G \int_0^\infty f L^2 dL \]

\[ \rho V C_p \frac{dT}{dt} = -3 \Delta H_c \rho_c k_v V G \int_0^\infty f L^2 dL - UA(T - T_j(t)) \]

Nucleation and growth in the bulk

\[ B = k_b \left( \frac{\hat{C} - \hat{C}_{sat}(T)}{\hat{C}_{sat}(T)} \right)^b = k_b S^b \]

\[ G = k_g \left( \frac{\hat{C} - \hat{C}_{sat}(T)}{\hat{C}_{sat}(T)} \right)^g = k_g S^g \]
Stochastic Formulation

Reaction Mechanism for Nucleation and Growth

\[ 2M \xrightarrow{k_n} N_{l_1} \quad \Delta H_{\text{rxn}}^n \]

\[ N_{l_n} + M \xrightarrow{k_g} N_{l_{n+1}} \quad \Delta H_{\text{rxn}}^g \]

\( N_{l_n} \) 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

Deterministic Solution
Via Orthogonal Collocation

Discrete particle sizes
Integer-valued particle accounting

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

\[
N_{lp} + N_{lq} \xrightarrow{k_a} N_{lp+q}
\]

Stochastic Solution
Add one reaction: $N_{lp} + N_{lq} \xrightarrow{k_a} N_{lp+q}$
Average of 500 Simulations

Deterministic Solution
Via Adaptive Mesh Methods?
Large time investment!

TWMCC—27 September 2001
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