Stochastic Chemical Kinetics

Dan T. Gillespie
GillespieDT@mailaps.org

Abstract: The time evolution of a well-stirred chemically reacting system is traditionally modeled by a set of coupled ode's called the *reaction rate equations* (RRE). The resulting picture of continuous deterministic evolution is, however, valid only for infinitely large systems. That condition is adequately approximated in most "macroscopic" chemical systems. But in biological systems formed by single living cells, the small population numbers of some reactant species can result in dynamical behavior that is noticeably *discrete* rather than continuous, and noticeably *stochastic* rather than deterministic. In that case, a more accurate mathematical modeling is obtained by using the machinery of Markov process theory, specifically, the *chemical master equation* (CME) and the *stochastic simulation algorithm* (SSA). This talk will review the theoretical foundations of stochastic chemical kinetics, and then discuss some recent efforts to (1) approximate the SSA by a faster simulation procedure, and (2) establish the formal connection between the CME/SSA description and the RRE description.

STOCHASTIC CHEMICAL KINETICS

Dan Gillespie

GillespieDT@mailaps.org

Current Support: **AFOSR** (#F30602-01-2-0558) and **Caltech**

Past Support: **ONR**

Collaborators:

John Doyle (Caltech)
Linda Petzold (UCSB)
Muruhan Rathinam (UMBC)
Adam Arkin (UCB, LBNL)
Harley McAdams (Stanford)
Yang Cao (UCSB)
Andrew Hall (UCSB)

CHEMICAL KINETICS

- N ***chemical species*** $S_1,\ldots,S_N$. M ***reaction channels*** $R_1,\ldots,R_M$.

  E.g., the $N = M = 2$ system: $S_1 + S_2 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} 2S_1$

- Assume system has constant volume Ω, and constant temperature T.
- Also assume system is ***well-stirred*** (spatially homogeneous).
- Let $X_i(t) \triangleq$ the *number* of S_i molecules in the system at time t.
- $\big(X_1(t),\ldots,X_N(t)\big) \triangleq \mathbf{X}(t)$, the ***state*** of the system at time t.

➢ ***The Problem:*** Given $\mathbf{X}(t_0) = \mathbf{x}_0$, find $\mathbf{X}(t)$ for $t > t_0$.

The Traditional Approach

- Asserts that $\mathbf{X}(t)$ evolves in time according to a set of coupled, first-order, *ordinary differential equations* of the form

$$\frac{dX_i}{dt} = f_i(X_1,\ldots,X_N) \quad (i=1,\ldots,N),$$

  where the f_i are determined by the forms of the M reaction channels.

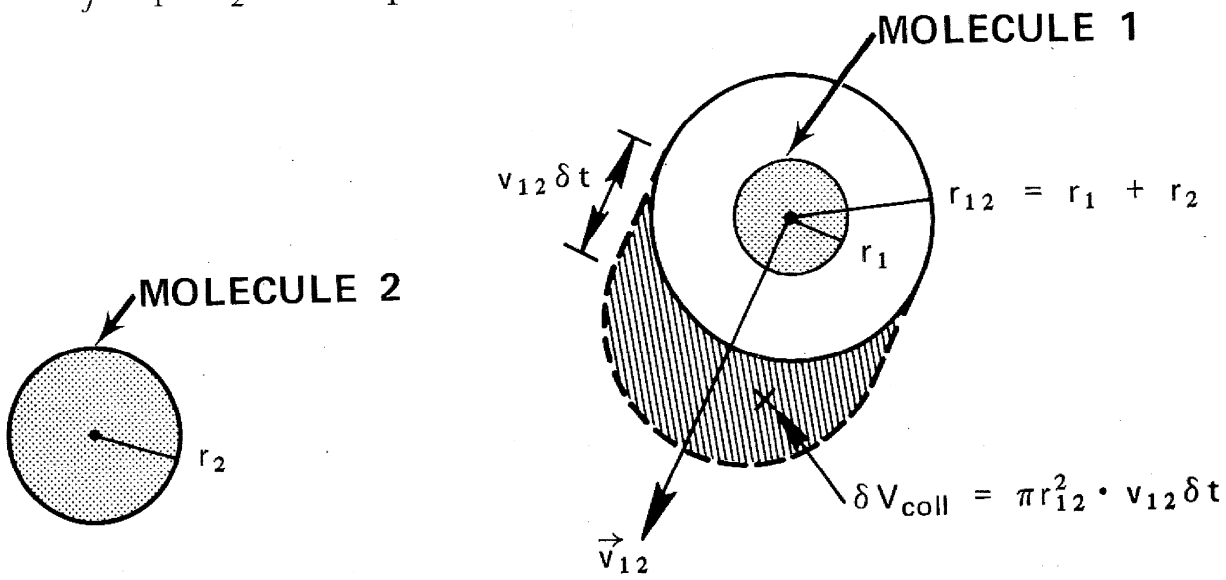
- Called the ***reaction rate equations*** (RRE).
- Usually written in terms of the *concentrations*, $Z_i \equiv X_i/\Omega$.
- RRE $\Rightarrow \mathbf{X}(t)$ is a ***continuous***, ***deterministic*** process.

But in fact . . .

- $\mathbf{X}(t)$ is *not continuous*; it's ***discrete***:
  - Molecules come in *whole numbers*.
  - Molecular populations change only by *integer* amounts.

- $\mathbf{X}(t)$ is *not deterministic*; it's ***stochastic***:
  - Only if we were to define the system's state as the positions and the velocities of all the molecules (and assume Newtonian mechanics) could we regard the system as being "deterministic".
  - But even then, the extreme *sensitivity to initial conditions* will render the system *effectively* stochastic – like a tossed coin.
  - Chemical reactions occur as discrete events, as a result of molecular collisions that cannot be precisely predicted.
  - At best, we can predict only the *probability* that a reaction event will occur.

$$R_j:\ S_1 + S_2 \xrightarrow{c_j} \text{products}$$



$$\underbrace{\left\{\left(\frac{(\bar{v}_{12}dt)(\pi r_{12}^2)}{\Omega}\right)\times\exp\left(-E_j/k_BT\right)\right\}}_{\text{Probability that a randomly chosen } S_1\text{-}S_2 \text{ pair will react according to } R_j \text{ in next } dt.}\times x_1x_2 = \underbrace{\left(\frac{\bar{v}_{12}\pi r_{12}^2}{\Omega}\exp\left(-E_j/k_BT\right)\right)}_{c_j} x_1x_2\,dt$$

$$\triangleq \underbrace{\left(c_j\,x_1x_2\right)}_{a_j(\mathbf{x})}dt \triangleq a_j(\mathbf{x})\,dt$$

STOCHASTIC CHEMICAL KINETICS

Each elemental reaction channel R_j is defined by two quantities:

- a ***propensity function*** $a_j(\mathbf{x})$, where $a_j(\mathbf{x})\,dt \triangleq$ the *probability*, given $\mathbf{X}(t) = \mathbf{x}$, that one R_j reaction event will occur in $[t, t+dt)$;
- a ***state change vector*** $\boldsymbol{\nu}_j \equiv \left(\nu_{1j}, \ldots, \nu_{Nj}\right)$, where $\nu_{ij} \triangleq$ the *change* in the S_i population caused by one R_j reaction event.

E.g.,

$$S_1 + S_2 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} 2S_1 \quad \Rightarrow \quad \begin{cases} a_1(\mathbf{x}) = c_1 x_1 x_2, & \boldsymbol{\nu}_1 = (+1, -1, 0, \ldots, 0) \\ a_2(\mathbf{x}) = c_2 \dfrac{x_1(x_1 - 1)}{2}, & \boldsymbol{\nu}_2 = (-1, +1, 0, \ldots, 0) \end{cases}$$

➢ **Implication:** $\mathbf{X}(t)$ is a ***jump Markov process*** (a continuous-time, discrete-state, past-forgetting, stochastic process).

Two Approaches to Stochastic Chemical Kinetics

- An *analytical* approach, and a *simulation* approach.
- They are *logically equivalent* – both follow rigorously from the same stochastic premise.
- They are *exact* for ***well-stirred*** (or ***self-stirring***) systems.

THE ANALYTICAL APPROACH

Focuses on the *probability density function* of the random variable $\mathbf{X}(t)$, namely

$$P(\mathbf{x},t\,|\,\mathbf{x}_0,t_0) \triangleq \text{Prob}\left\{\mathbf{X}(t)=\mathbf{x},\ \text{given that } \mathbf{X}(t_0)=\mathbf{x}_0\right\}.$$

Can prove that P obeys the time-evolution equation

$$\frac{\partial P(\mathbf{x},t\,|\,\mathbf{x}_0,t_0)}{\partial t} = \sum_{j=1}^{M}\left[a_j(\mathbf{x}-\boldsymbol{\nu}_j)P(\mathbf{x}-\boldsymbol{\nu}_j,t\,|\,\mathbf{x}_0,t_0) - a_j(\mathbf{x})P(\mathbf{x},t\,|\,\mathbf{x}_0,t_0)\right]$$

- Called the ***chemical master equation*** (CME).
- In principle, it completely determines $P(\mathbf{x},t\,|\,\mathbf{x}_0,t_0)$, and hence $\mathbf{X}(t)$.
- In practice, it's impossible to solve for all but the simplest of systems.

Derivation of the Chemical Master Equation

Write Prob$\{\mathbf{X}(t+dt) = \mathbf{x}\}$ as the sum of the probabilities of all (mutually exclusive) paths leading to state $\mathbf{x}$ at time $t+dt$ via possible states at time t:

$$P(\mathbf{x}, t+dt \,|\, \mathbf{x}_0, t_0) = P(\mathbf{x}, t \,|\, \mathbf{x}_0, t_0) \times \left[1 - \sum_{j=1}^{M} \left(a_j(\mathbf{x}) dt \right) \right]$$
$$+ \sum_{j=1}^{M} P(\mathbf{x} - \boldsymbol{\nu}_j, t \,|\, \mathbf{x}_0, t_0) \times \left(a_j(\mathbf{x} - \boldsymbol{\nu}_j) dt \right)$$

Subtract $P(\mathbf{x}, t \,|\, \mathbf{x}_0, t_0)$, divide through by dt, then let $dt \to 0 \;\Rightarrow$ CME.

Note: Invocation of the Addition Law here relies on the fact that dt is so small that *no more than one* reaction event will ever occur in any dt.

THE SIMULATION APPROACH

Focuses on the function $p(\tau, j|\mathbf{x},t)$, defined by

$p(\tau, j|\mathbf{x},t)\,d\tau \triangleq$ the *probability*, given $\mathbf{X}(t)=\mathbf{x}$, that the *next* reaction in the system will occur in the infinitesimal time interval $[t+\tau, t+\tau+d\tau)$, *and* will be an R_j reaction.

Can prove that p is given by

$$p(\tau, j|\mathbf{x},t) = a_j(\mathbf{x}) \exp\left(-a_0(\mathbf{x})\tau\right), \quad \text{where } a_0(\mathbf{x}) \triangleq \sum_{j'=1}^{M} a_{j'}(\mathbf{x}).$$

- *Implies*: The time τ to the *next reaction event* is an exponentially distributed random variable with mean $1/a_0(\mathbf{x})$, and the channel index j of that reaction is an integer random variable with prob $a_j(\mathbf{x})/a_0(\mathbf{x})$.
- It's easy to generate on a computer (pseudo)random samples of τ and j according to this prescription.

Derivation of the Next-Reaction Probability Density Function

Let $P_0(\tau|\mathbf{x},t) \triangleq$ the probability, given $\mathbf{X}(t)=\mathbf{x}$, that *no* reactions will occur in time interval $[t, t+\tau)$. Then

$$p(\tau, j|\mathbf{x},t)\,d\tau = P_0(\tau|\mathbf{x},t)\times\Big(a_j(\mathbf{x})d\tau\Big),$$

so

$$p(\tau, j|\mathbf{x},t) = a_j(\mathbf{x})\,P_0(\tau|\mathbf{x},t).$$

To calculate $P_0(\tau|\mathbf{x},t)$, observe that it must satisfy

$$P_0(\tau+d\tau|\mathbf{x},t) = P_0(\tau|\mathbf{x},t)\times\left[1-\sum_{j'=1}^{M}\Big(a_{j'}(\mathbf{x})d\tau\Big)\right] \equiv P_0(\tau|\mathbf{x},t)\big[1-a_0(\mathbf{x})d\tau\big],$$

whence

$$\frac{dP_0(\tau|\mathbf{x},t)}{d\tau} = -a_0(\mathbf{x})\,P_0(\tau|\mathbf{x},t).$$

The solution to this ODE for the initial condition $P_0(\tau=0|\mathbf{x},t)=1$ is

$$P_0(\tau|\mathbf{x},t) = \exp\big(-a_0(\mathbf{x})\tau\big).$$

QED

The Stochastic Simulation Algorithm (SSA)

An explicit, exact procedure for constructing a *numerical realization* of the stochastic process $\mathbf{X}(t)$. The "direct" version of the SSA is:

1. With the system in state $\mathbf{x}$ at time t, compute $a_0(\mathbf{x}) \triangleq \sum_{j'=1}^{M} a_{j'}(\mathbf{x})$.

2. Draw two *unit-interval uniform* random numbers r_1 and r_2, and compute τ and j according to

- $\tau = \dfrac{1}{a_0(\mathbf{x})} \ln\left(\dfrac{1}{r_1}\right)$,
- j = the *smallest integer* satisfying $\sum_{j'=1}^{j} a_{j'}(\mathbf{x}) > r_2\, a_0(\mathbf{x})$.

3. *Effect the next reaction*: Replace $t \leftarrow t+\tau$ and $\mathbf{x} \leftarrow \mathbf{x}+\boldsymbol{\nu}_j$.

4. Record $(\mathbf{x},t)$. Then return to Step **1**, or else end the simulation.

The Stochastic Simulation Algorithm . . .

- Numerically simulates the time evolution of a ***well-stirred*** chemically reacting system in a way that accurately reproduces all effects of naturally occurring randomness.
- Does *not* entail approximating a "dt" by a "Δt".
- Is *procedurally simple*, even for systems whose CME is intractable.
- Has been redesigned to be faster and more efficient (though more complicated to code) by **M. Gibson** and **J. Bruck** [*J. Phys. Chem. A* **104**, 1876 (2000)]
- ***Remains too slow for most practical problems***: Since it simulates *every* reaction event, *one* at a time, it can take a very long time to execute if any reactant is present in very large numbers – as is usually the case.

WHY BOTHER?

- For most *practical* chemical systems the molecular populations are *very large*, and the phenomenological RRE,

$$\frac{d\mathbf{X}(t)}{dt} = \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j\big(\mathbf{X}(\mathbf{t})\big),$$

  is *accurate* and *fast*.

- But in ***small biochemical systems*** (like the inside of a living cell) it can happen that *some* critical reactant species are present in *low numbers*. Then the inherent stochasticity of the system becomes noticeable, and the deterministic RRE can be *inaccurate* and *misleading*.

➢ For such systems, we must instead use either the CME or the SSA.

TWO QUESTIONS:

- **1.** Is there a way to *speed up* the SSA for large systems, even at the cost of some exactness?

- **2.** How does it happen that the *rigorous* but computationally intractable *C*ME,

$$\frac{\partial P(\mathbf{x},t|\mathbf{x}_0,t_0)}{\partial t}=\sum_{j=1}^{M}\Big[a_j(\mathbf{x}-\boldsymbol{\nu}_j)P(\mathbf{x}-\boldsymbol{\nu}_j,t|\mathbf{x}_0,t_0)-a_j(\mathbf{x})P(\mathbf{x},t|\mathbf{x}_0,t_0)\Big],$$

segues for "large" systems to the *heuristic* but computationally efficient RRE,

$$\frac{d\mathbf{X}(t)}{dt}=\sum_{j=1}^{M}\boldsymbol{\nu}_j a_j\big(\mathbf{X}(\mathbf{t})\big)\,?$$

Tau-Leaping

An *approximate*, *accelerated* stochastic simulation procedure.

- Advances the process by a *pre-selected* time τ, which may encompass *more than one* reaction event.
- The size of τ is limited by the ***Leap Condition***: *The changes in the propensity function values during the leap must be "small".*
- Accelerated simulation occurs whenever τ can satisfy the Leap Condition, yet encompass *many* reaction events.
- ***But must use with caution!*** Leap over only the "unimportant" reactions, not the "important" ones.

Key Ideas Underlying Tau-Leaping

- For positive *constants* a and τ, the *Poisson random variable* $\mathcal{P}(a,\tau)$ is defined as the *number* of "events" that will occur in time τ, given that the *probability* of an event occurring in any *infinitesimal* dt is adt.
- Therefore, if $\mathbf{X}(t)=\mathbf{x}$, and τ is such that $a_j(\mathbf{x}) \approx$ constant in $[t,t+\tau]$, then the number of R_j reactions that will occur in $[t,t+\tau]$ is (approximately) $\mathcal{P}\left(a_j(\mathbf{x}),\tau\right)$.
- Numerical procedures exist for generating sample values of $\mathcal{P}(a,\tau)$ [Press, et al., *Numerical Recipes: The Art of Scientific Computing*].
- A way has been developed to estimate in advance the *largest* τ that is consistent with the Leap Condition.

A Method for Choosing τ

If $\mathbf{X}(t) = \mathbf{x}$, then the *state change* induced by a time leap τ is

$$\mathbf{X}(t+\tau) - \mathbf{x} \triangleq \mathbf{\Lambda}(\tau, \mathbf{x}) = \sum_{j=1}^{M} K_j(\mathbf{x}, \tau)\boldsymbol{\nu}_j \approx \sum_{j=1}^{M} \mathcal{P}_j\left(a_j(\mathbf{x}), \tau\right)\boldsymbol{\nu}_j \ .$$

The resultant change in propensity function a_j is

$$\Delta a_j(\tau, \mathbf{x}) \triangleq a_j(\mathbf{x} + \mathbf{\Lambda}(\tau, \mathbf{x})) - a_j(\mathbf{x}) \approx \sum_{i=1}^{N} \frac{\partial a_j(\mathbf{x})}{\partial x_i} \Lambda_i(\tau, \mathbf{x})$$

$$\approx \sum_{i=1}^{N} \frac{\partial a_j(\mathbf{x})}{\partial x_i} \sum_{j'=1}^{M} \mathcal{P}_{j'}(a_{j'}(\mathbf{x}), \tau)\nu_{ij'}$$

$$\Delta a_j(\tau, \mathbf{x}) \approx \sum_{j'=1}^{M} f_{jj'}(\mathbf{x})\, \mathcal{P}_{j'}(a_{j'}(\mathbf{x}), \tau), \quad \text{where} \quad f_{jj'}(\mathbf{x}) \triangleq \sum_{i=1}^{N} \frac{\partial a_j(\mathbf{x})}{\partial x_i} \nu_{ij'} \ .$$

$$\text{So,} \quad \langle \Delta a_j(\tau, \mathbf{x}) \rangle \approx \sum_{j'=1}^{M} f_{jj'}(\mathbf{x}) \langle \mathcal{P}_{j'}(a_{j'}(\mathbf{x}), \tau) \rangle = \sum_{j'=1}^{M} f_{jj'}(\mathbf{x}) \left(a_{j'}(\mathbf{x})\tau\right),$$

$$\text{var}\{\Delta a_j(\tau, \mathbf{x})\} \approx \sum_{j'=1}^{M} f_{jj'}^2(\mathbf{x})\, \text{var}\{\mathcal{P}_{j'}(a_{j'}(\mathbf{x}), \tau)\} = \sum_{j'=1}^{M} f_{jj'}^2(\mathbf{x}) \left(a_{j'}(\mathbf{x})\tau\right).$$

- Satisfy the **Leap Condition** by requiring the *absolute mean* and the *standard deviation* of $\Delta a_j(\tau, \mathbf{x})$ to be $\leq \varepsilon\, a_0(\mathbf{x})\ \ \forall j$, where $\varepsilon \ll 1$ is an *accuracy control parameter*. Find that the largest τ satisfying this requirement is

$$\tau = \underset{j\in[1,M]}{\text{Min}} \left\{ \frac{\varepsilon\, a_0(\mathbf{x})}{\left|\mu_j(\mathbf{x})\right|}, \frac{\varepsilon^2 a_0^2(\mathbf{x})}{\sigma_j^2(\mathbf{x})} \right\},$$

where, with $f_{jj'}(\mathbf{x}) \triangleq \sum_{i=1}^{N} \frac{\partial a_j(\mathbf{x})}{\partial x_i} \nu_{ij'}$,

$$\mu_j(\mathbf{x}) \triangleq \sum_{j'=1}^{M} f_{jj'}(\mathbf{x})\, a_{j'}(\mathbf{x}), \quad \sigma_j^2(\mathbf{x}) \triangleq \sum_{j'=1}^{M} f_{jj'}^2(\mathbf{x})\, a_{j'}(\mathbf{x}).$$

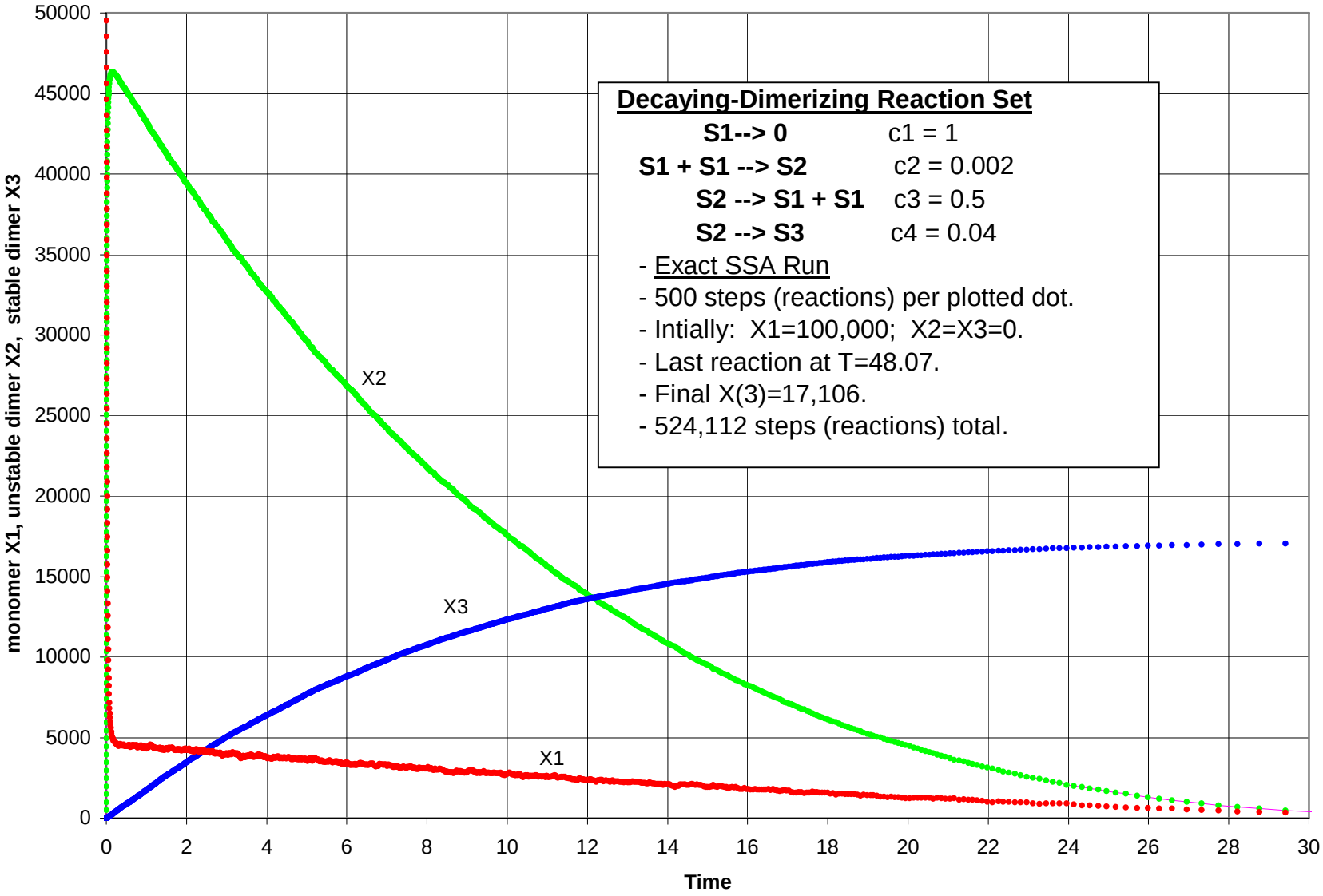
The functions $f_{jj'}$, μ_j, and σ_j^2 can be explicitly computed *prior* to simulation, and they will usually be easy to evaluate for a given $\mathbf{x}$.

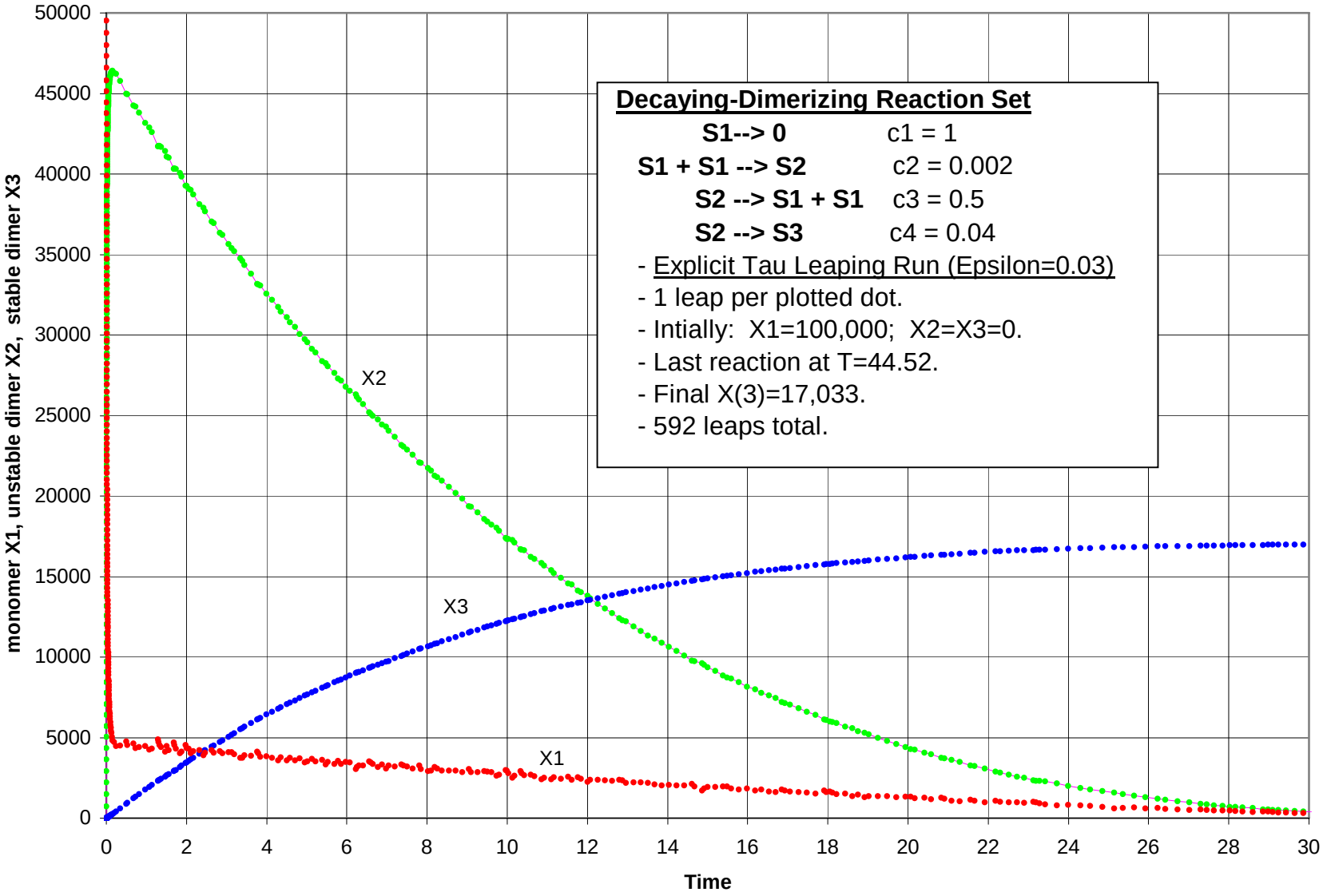
The Tau-Leap Simulation Algorithm

- **1**. In state $\mathbf{x}$ at time t, choose τ so that the expected change in *every* propensity function in $[t,t+\tau]$ is $\leq \varepsilon a_0(\mathbf{x})$.
- **2**. Generate the number of firings k_j of channel R_j in $[t,t+\tau]$ as

$$k_j = \mathcal{P}\left(a_j(\mathbf{x}),\tau\right) \quad (j=1,\ldots,M).$$

- **3**. *Leap*: Replace $t \leftarrow t+\tau$ and $\mathbf{x} \leftarrow \mathbf{x} + \sum_{j=1}^{M} k_j \boldsymbol{\nu}_j$.
- **4**. Record $(\mathbf{x},t)$. Then return to Step **1**, or else end the simulation.


Decaying-Dimerizing Reaction Set
S1--> 0 c1 = 1
S1 + S1 --> S2 c2 = 0.002
S2 --> S1 + S1 c3 = 0.5
S2 --> S3 c4 = 0.04
- Exact SSA Run
- 500 steps (reactions) per plotted dot.
- Intially: X1=100,000; X2=X3=0.
- Last reaction at T=48.07.
- Final X(3)=17,106.
- 524,112 steps (reactions) total.
X2
X3
X1
monomer X1, unstable dimer X2, stable dimer X3
Time
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30


Decaying-Dimerizing Reaction Set
S1--> 0 c1 = 1
S1 + S1 --> S2 c2 = 0.002
S2 --> S1 + S1 c3 = 0.5
S2 --> S3 c4 = 0.04
- Explicit Tau Leaping Run (Epsilon=0.03)
- 1 leap per plotted dot.
- Intially: X1=100,000; X2=X3=0.
- Last reaction at T=44.52.
- Final X(3)=17,033.
- 592 leaps total.
X2
X3
X1
monomer X1, unstable dimer X2, stable dimer X3
Time
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30

Tau-Leaping usually speeds things up, *but* . . .

- Selecting the largest τ consistent with the Leap Condition entails some computational overhead.
- Generating *Poisson* random numbers is more time consuming than generating the *exponential* random numbers used by the SSA.
- Care must be taken not to leap over "important" reactions.
- Additional strategies are needed to cope with ***stiffness*** (widely varying time-scales).

➢ *By itself, it's **not** the final answer. But it provides a **context** for such.*

How does the CME/SSA segue to the RRE for "large" systems?

- We saw in tau-leaping that, if $\mathbf{X}(t) = \mathbf{x}$ and τ is *macroscopically infinitesimal* (in that none of the a_j's change "noticeably" during τ), then to a good approximation,

$$\mathbf{X}(t+\tau) = \mathbf{x} + \sum_{j=1}^{M} \mathcal{P}_j\left(a_j(\mathbf{x}), \tau\right)\boldsymbol{\nu}_j ,$$

  where the $\mathcal{P}_j$'s are *statistically independent Poisson* random variables.

- Suppose it **also** happens that *every* reaction channel R_j fires *many more times than once* in the next τ. (This nearly always happens whenever *all* the reactant molecular populations are *large* enough.)

- This would mean, since $\left\langle \mathcal{P}_j\left(a_j(\mathbf{x}),\tau\right)\right\rangle = a_j(\mathbf{x})\tau$, that $a_j(\mathbf{x})\tau \gg 1 \;\; \forall j$.
- But when $a\tau \gg 1$, can approximate $\mathcal{P}(a,\tau) \approx \mathcal{N}(a\tau, a\tau)$; hence,

$$\mathbf{X}(t+\tau) = \mathbf{x} + \sum_{j=1}^{M} \mathcal{N}_j\left(a_j(\mathbf{x})\tau, a_j(\mathbf{x})\tau\right)\boldsymbol{\nu}_j$$

$$= \mathbf{x} + \sum_{j=1}^{M} \left\{ a_j(\mathbf{x})\tau + \left[a_j(\mathbf{x})\tau\right]^{1/2} \mathcal{N}_j(0,1) \right\} \boldsymbol{\nu}_j$$

- Collecting terms, and remembering that $\mathbf{x} = \mathbf{X}(t)$, this is

$$\mathbf{X}(t+\tau) = \mathbf{X}(t) + \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j\left(\mathbf{X}(t)\right)\tau + \sum_{j=1}^{M} \boldsymbol{\nu}_j \sqrt{a_j\left(\mathbf{X}(t)\right)}\, \mathcal{N}_j(0,1)\, \sqrt{\tau}$$

We have just proved the following . . .

Theorem: If $\bar{dt}$ is a ***macroscopic infinitesimal,*** in that during $\bar{dt}$,

- *no* propensity function changes its value significantly, yet
- *every* reaction channel fires many more times that once,

then we can *approximate* the t to $t+\bar{dt}$ system update by

$$\mathbf{X}(t+\bar{dt}) = \mathbf{X}(t) + \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j\left(\mathbf{X}(t)\right)\bar{dt} + \sum_{j=1}^{M} \boldsymbol{\nu}_j \sqrt{a_j\left(\mathbf{X}(t)\right)}\, N_j(t)\, \sqrt{\bar{dt}} \ .$$

Here, the $N_j(t)$ are statistically independent, temporally uncorrelated, *normal* random variables with means 0 and variances 1.

- This is the ***Chemical Langevin Equation (CLE)***.
- It *approximates* $\mathbf{X}(t)$ as a <u>*continuous*</u> (versus a *jump*) Markov process.
- It is mathematically the same as the SDE

$$\frac{d\mathbf{X}(t)}{dt} = \sum_{j=1}^{M} \boldsymbol{\nu}_j\, a_j\left(\mathbf{X}(t)\right) + \sum_{j=1}^{M} \boldsymbol{\nu}_j \sqrt{a_j\left(\mathbf{X}(t)\right)}\, \Gamma_j(t) \ ,$$

where $\left\langle \Gamma_j(t)\, \Gamma_{j'}(t') \right\rangle = \delta_{jj'}\, \delta(t-t') \quad (j, j' = 1, \ldots, M)$.

- In the theory of continuous Markov processes, every *Langevin equation* for $\mathbf{X}(t)$ implies a *Fokker-Planck equation* for $P(\mathbf{x},t|\mathbf{x}_0,t_0)$.

- For the ***chemical Langevin equation*** (CLE)

$$\mathbf{X}(t+\bar{dt}) = \mathbf{X}(t) + \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j\left(\mathbf{X}(t)\right)\bar{dt} + \sum_{j=1}^{M} \boldsymbol{\nu}_j \sqrt{a_j\left(\mathbf{X}(t)\right)}\, N_j(t)\sqrt{\bar{dt}}\ ,$$

the corresponding ***chemical Fokker-Planck equation*** (CFPE) can be shown to be

$$\frac{\partial P(\mathbf{x},t|\mathbf{x}_0,t_0)}{\partial t} = -\sum_{i=1}^{N}\frac{\partial}{\partial x_i}\left[\left(\sum_{j=1}^{M}\nu_{ij}a_j(\mathbf{x})\right)P(\mathbf{x},t|\mathbf{x}_0,t_0)\right]$$
$$+\frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2}\left[\left(\sum_{j=1}^{M}\nu_{ij}^2 a_j(\mathbf{x})\right)P(\mathbf{x},t|\mathbf{x}_0,t_0)\right]$$
$$+\sum_{\substack{i,i'=1\\(i<i')}}^{N}\frac{\partial^2}{\partial x_i \partial x_{i'}}\left[\left(\sum_{j=1}^{M}\nu_{ij}\nu_{i'j}a_j(\mathbf{x})\right)P(\mathbf{x},t|\mathbf{x}_0,t_0)\right]$$

Now consider . . . ***The Thermodynamic Limit.***

Definition: All $X_i \to \infty$, and $\Omega \to \infty$, with $X_i/\Omega \to$ constant .

- Can *prove* that, in this limit, *all* propensity functions grow *linearly* with the system size.
- Therefore, in the CLE

$$\mathbf{X}(t+\bar{dt}) = \mathbf{X}(t) + \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j\left(\mathbf{X}(t)\right)\bar{dt} + \sum_{j=1}^{M} \boldsymbol{\nu}_j \sqrt{a_j\left(\mathbf{X}(t)\right)}\, N_j(t)\, \sqrt{\bar{dt}} \ ,$$

the *deterministic* term grows like $\mathbf{X}(t)$, while the *stochastic* term grows like $\sqrt{\mathbf{X}(t)}$. So in the thermodynamic limit, the CLE reduces to

$$\mathbf{X}(t+\bar{dt}) = \mathbf{X}(t) + \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j\left(\mathbf{X}(t)\right)\bar{dt} \ .$$

- **This is how the RRE arises in stochastic chemical kinetics.**

A Hierarchy of Schemes for Modeling Chemical Kinetics

1. ***Molecular Dynamics*** (MD)
   - Tracks the position and velocity of *every* molecule.
   - Simulates *every* collision, non-reactive as well as reactive.
   - Shows changes in species populations *and* spatial concentrations.
   - Is essentially *exact*.
   - Is extremely *slow* for realistic systems.
2. ***Stochastic Simulation Algorithm*** (SSA)
   - Assumes that the *non-reactive* collisions merely keep the molecules *well-stirred*; hence, assumes the system is *spatial homogeneous*.
   - Simulates only the *reactive* collisions, i.e., every chemical reaction event.
   - Tracks only the species *populations*.
   - Is exactly equivalent to the Chemical Master Equation (CME).
   - Is much faster than MD.

3. ***Tau-Leaping***
   - *An approximation to the SSA.*
   - Advances time by a *pre-selected* τ during which no propensity function changes its value "noticeably" and more than one reaction may occur.
   - Number of R_j firings in τ is approximated as $\mathcal{P}\left(a_j(\mathbf{x}),\tau\right)$.
   - Is faster than the SSA if *many* reactions occur in τ.

4. ***Chemical Langevin Equation*** (CLE)
   - *A special case of Tau-Leaping.*
   - Applies whenever τ is such that $a_j(\mathbf{x})\tau \gg 1$ for *every* R_j.
   - Number of R_j firings in τ is approximated as $\mathcal{N}\left(a_j(\mathbf{x})\tau, a_j(\mathbf{x})\tau\right)$.
   - Is an SDE that approximates the *jump* Markov process $\mathbf{X}(t)$ of the CME/SSA by a *continuous* Markov process.
   - Is faster than ordinary Tau-Leaping.

5. ***Reaction Rate Equation*** (RRE)
   - *The thermodynamic limit of the CLE.*
   - The (random) diffusion term in the CLE becomes negligibly small compared to the (deterministic) drift term.
   - The CLE then reduces from an SDE to an ODE.
   - Defines a continuous *deterministic* process.
   - Computationally the fastest.

<u>The Challenge</u>:

To build a seamless, easy-to-use software package that automatically selects the optimal appropriate method to advance the state in time, and then executes that method correctly and efficiently.