THE SLOW-SCALE STOCHASTIC SIMULATION ALGORITHM

Yang Cao (VT), Dan Gillespie, Linda Petzold (UCSB)

***

DG Current Support: **U.C. Santa Barbara**
Caltech (NIGMS , NIH)

DG Past Support: **Caltech** (DARPA/AFOSR, Beckman)
Molecular Sciences Institute (Sandia / DOE)
ONR

ASSUMPTIONS & DEFINITIONS

- A well-stirred chemical system at constant volume and temperature.
- N species $\{S_1,\ldots,S_N\}$. System state is $\mathbf{X}(t)=\left(X_1(t),\ldots,X_N(t)\right)$,

  $X_i(t) \equiv$ number of S_i molecules at time t.
- M reactions $\{R_1,\ldots,R_M\}$. Each R_j is described by two quantities:
  - State change vector: $\boldsymbol{\nu}_j=\left(\nu_{1j},\ldots,\nu_{Nj}\right)$, where

    $\nu_{ij} \equiv$ change induced in X_i by one R_j event.

    So R_j induces $\mathbf{x} \rightarrow \mathbf{x}+\boldsymbol{\nu}_j$.
  - Propensity function: $a_j(\mathbf{x})$, where

  $a_j(\mathbf{x})dt \equiv$ probability, given $\mathbf{X}(t)=\mathbf{x}$, that R_j will fire in $[t,t+dt)$.

The Stochastic Simulation Algorithm (SSA)

- If the system is in state $\mathbf{x}$ at time t, then with $a_0(\mathbf{x}) \equiv \sum_{k=1}^{M} a_k(\mathbf{x})$,
  - the time τ to the next reaction will be the exponential random variable with mean $1/a_0(\mathbf{x})$;
  - the index j of the next reaction will be the integer random variable with probability $a_j(\mathbf{x})/a_0(\mathbf{x})$.
- By generating such random numbers for τ and j, we can advance the system to the next reaction:

$$t \leftarrow t + \tau \text{ and } \mathbf{x} \leftarrow \mathbf{x} + \boldsymbol{\nu}_j.$$

A "Multi-scale" Problem

- If the various reaction channels fire at *vastly different rates*, the SSA will spend most of its time simulating the fastest reactions.

Example: $S_1 \underset{c_2}{\overset{c_1}{\rightleftarrows}} S_2 \xrightarrow{c_3} S_3$

(fast) (slow)

- If, in addition, the fastest reactions are "less important" than the slower ones, the SSA simulation will be slow and inefficient.
  - Is there an approximate but accurate way to *skip over* the fast reactions and simulate *only the slow* ones?
  - And how can we tell when it's okay to do that?

- STIFF SYSTEMS -

- In ODE theory, a ***stiff*** system is one that evolves on *wide range of time scales*, with the *fastest* mode being *stable*.
  - Approximate methods have been devised to efficiently simulate stiff ODEs.

- **But how can we efficiently simulate *stochastic* stiff systems?**

SOME ISSUES

- Do "fast" and "slow" apply to *reactions* or to *species*? Or to both?
- How should fast and slow reactions/species be *defined*?
- Since the fast and slow components of a system are *interconnected*, can they be teased apart without seriously altering their combined effects?

The SLOW-SCALE SSA

- A computational recipe for simulating the stochastic evolution of a stiff system *one **slow** reaction at a time*.
- Proceeds in a series of *steps*, which also serve to *authenticate* the procedure.

1st Provisionally partition the reactions $\{R_1,\ldots,R_M\}$:

- Fast reactions $\left\{R_1^{\mathrm{f}},\ldots,R_{M_{\mathrm{f}}}^{\mathrm{f}}\right\}$. Slow reactions $\left\{R_1^{\mathrm{s}},\ldots,R_{M_{\mathrm{s}}}^{\mathrm{s}}\right\}$.
- Tentative criterion: Want $a_j^{\mathrm{f}}(\mathbf{x}) \gg a_{j'}^{\mathrm{s}}(\mathbf{x})$ $(\forall j, j')$ "most of the time".
- ***The partitioning will later be subjected to an acceptance test.***
- If *no* partitioning is acceptable, then we're SOL.

2nd Partition the species $\{S_1,\dots,S_N\}$:

- Fast species $\{S_1^{\mathrm{f}},\dots,S_{N_{\mathrm{f}}}^{\mathrm{f}}\} \Rightarrow \mathbf{X}^{\mathrm{f}}(t) \equiv \left(X_1^{\mathrm{f}}(t),\dots,X_{N_{\mathrm{f}}}^{\mathrm{f}}(t)\right)$.
- Slow species $\{S_1^{\mathrm{s}},\dots,S_{N_{\mathrm{s}}}^{\mathrm{s}}\} \Rightarrow \mathbf{X}^{\mathrm{s}}(t) \equiv \left(X_1^{\mathrm{s}}(t),\dots,X_{N_{\mathrm{s}}}^{\mathrm{s}}(t)\right)$.
  - ➢ ***Criterion***: A species is *fast* if its population *gets changed* by at least one fast reaction; otherwise, the species is *slow*.

Some subtle points:

- ✓ $\nu_{ij} \to \nu_{ij}^{\sigma\rho}$, where $\sigma = \mathrm{f,s}$ (species) and $\rho = \mathrm{f,s}$ (reaction).
- ✓ A slow species cannot get changed by a fast reaction ($\nu_{ij}^{\mathrm{sf}} \equiv 0$), but a fast species *might* get changed by a slow reaction.
- ✓ $a_j^{\mathrm{s}}(\mathbf{x}) = a_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})$, $a_j^{\mathrm{f}}(\mathbf{x}) = a_j^{\mathrm{f}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})$.
- ✓ The population of a fast species *need not be large*.
- ✓ There might be *no slow species* (in which case $\mathbf{x}^{\mathrm{f}} = \mathbf{x}$).

3rd Define the "virtual fast process" $\hat{\mathbf{X}}^{\mathrm{f}}(t)$.

- $\hat{\mathbf{X}}^{\mathrm{f}}(t) \equiv$ the fast species populations driven by *only* the fast reactions.
- Thus, $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ is $\mathbf{X}^{\mathrm{f}}(t)$ with all slow reactions "turned off".
- $\mathbf{X}^{\mathrm{f}}(t)$ is a *physically real* process. But it's *non-Markovian*.
  - Determining $\mathbf{X}^{\mathrm{f}}(t)$ is no easier than determining $\mathbf{X}(t)$.
- $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ is a *physically fictitious* process. But it's Markovian.
  - It satisfies an ordinary master equation – for the fast species driven by only the fast reactions (so all the slow species populations will be constant).
- The CME for $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ will be *simpler* than the CME for $\mathbf{X}(t)$.
  It determines

$$\hat{P}(\mathbf{x}^{\mathrm{f}},t\,|\,\mathbf{x}_0^{\mathrm{f}},\mathbf{x}_0^{\mathrm{s}},t_0) \triangleq \Pr\left\{\hat{\mathbf{X}}^{\mathrm{f}}(t) = \mathbf{x}^{\mathrm{f}} \,|\, \mathbf{X}(t_0) = (\mathbf{x}_0^{\mathrm{f}},\mathbf{x}_0^{\mathrm{s}})\right\}.$$

$\underline{4^{\text{th}}}$ Require the system to be "stiff":

- ***Stability Condition.*** $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ must be ***stable***; i.e., the limit

$$\lim_{t\to\infty} \hat{P}(\mathbf{x}^{\mathrm{f}},t\,|\,\mathbf{x}_0,t_0) \equiv \hat{P}(\mathbf{x}^{\mathrm{f}},\infty\,|\,\mathbf{x}_0)$$

must *exist*. $\hat{P}(\mathbf{x}^{\mathrm{f}},\infty\,|\,\mathbf{x}_0)$ will be the solution of

$$0 = \sum_{j=1}^{M_{\mathrm{f}}} \left\{ a_j^{\mathrm{f}}(\mathbf{x}^{\mathrm{f}} - \boldsymbol{\nu}_j^{\mathrm{f}}, \mathbf{x}_0^{\mathrm{s}}) \hat{P}(\mathbf{x}^{\mathrm{f}} - \boldsymbol{\nu}_j^{\mathrm{f}}, \infty\,|\,\mathbf{x}_0) - a_j^{\mathrm{f}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}_0^{\mathrm{s}}) \hat{P}(\mathbf{x}^{\mathrm{f}}, \infty\,|\,\mathbf{x}_0) \right\}.$$

- ***Timescale Separation Condition.*** $\hat{\mathbf{X}}^{\mathrm{f}}(t) \to \hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ in a time that is ***small*** compared to the *expected* time to the *next slow* reaction.

- If these conditions are satisfied, the fast reactions should be "less important" than the slow ones. Skipping over them should be okay.
- If these conditions are not satisfied, *and cannot be satisfied by any repartitioning*, then the fast reactions are no less important than the slow ones. Skipping over them is not a good idea.

$\underline{5^{\text{th}}}$ The Slow-Scale Approximation

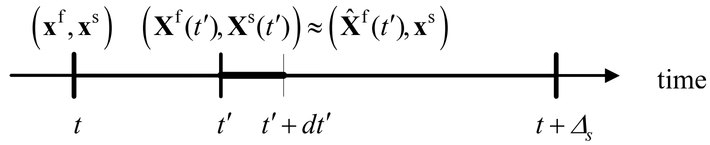
Lemma: With $\mathbf{X}(t) = (\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$, let Δ_{s} be a time increment that is *large* compared to the time for $\hat{\mathbf{X}}^{\mathrm{f}}(t\to\infty) \to \hat{\mathbf{X}}^{\mathrm{f}}(\infty)$, but *small* compared to the expected time to the next *slow* reaction. Then the probability that R_j^{s} will fire in $[t, t+\Delta_{\mathrm{s}})$ is *approximately* $\Delta_{\mathrm{s}} \times \bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$, where

$$\bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}) \equiv \left\langle a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(\infty), \mathbf{x}^{\mathrm{s}}\right)\right\rangle = \sum_{\mathbf{y}^{\mathrm{f}}} \hat{P}(\mathbf{y}^{\mathrm{f}}, \infty\,|\,\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})\, a_j^{\mathrm{s}}(\mathbf{y}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}).$$

Why this lemma does the trick for us:

- For a "true" dt, $a_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}) \times dt = \mathrm{Prob}\left\{R_j^{\mathrm{s}} \text{ in } [t, t+dt)\right\}$.
  This is the ***defining property*** of the propensity function!
- Δ_{s}, though large on the "fast" scale, is *very small* on the "slow" scale, and the Lemma $\Rightarrow$ $\bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}) \times \Delta_{\mathrm{s}} \approx \mathrm{Prob}\left\{R_j^{\mathrm{s}} \text{ in } [t, t+\Delta_{\mathrm{s}})\right\}$.
- $\therefore$ $\bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$ can be regarded as the **effective** propensity function of R_j^{s} on the time scale of the **slow** reactions.

Proof of the lemma:



$$\mathrm{Prob}\left\{R_j^{\mathrm{s}}\text{ in }[t',t'+dt')\right\}=a_j^{\mathrm{s}}\left(\mathbf{X}^{\mathrm{f}}(t'),\mathbf{X}^{\mathrm{s}}(t')\right)dt'\approx a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(t'),\mathbf{x}^{\mathrm{s}}\right)dt'.$$

$$\begin{aligned}\mathrm{Prob}\left\{R_j^{\mathrm{s}}\text{ in }[t,t+\Delta_{\mathrm{s}})\right\}&\approx\int_{t'=t}^{t+\Delta_{\mathrm{s}}}a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(t'),\mathbf{x}^{\mathrm{s}}\right)dt'\\
&=\left\{\frac{1}{\Delta_{\mathrm{s}}}\int_{t}^{t+\Delta_{\mathrm{s}}}a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(t'),\mathbf{x}^{\mathrm{s}}\right)dt'\right\}\Delta_{\mathrm{s}}\\
&\approx\left\{\textbf{temporal}\text{ average of }a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(t'),\mathbf{x}^{\mathrm{s}}\right)\right\}\Delta_{\mathrm{s}}\\
&\approx\left\{\textbf{ensemble}\text{ average of }a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(\infty),\mathbf{x}^{\mathrm{s}}\right)\right\}\Delta_{\mathrm{s}}\\
&=\left\{\sum_{\mathbf{y}^{\mathrm{f}}}\hat{P}(\mathbf{y}^{\mathrm{f}},\infty\,|\,\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})\,a_j^{\mathrm{s}}(\mathbf{y}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})\right\}\Delta_{\mathrm{s}}.\quad\textbf{QED}\end{aligned}$$

- The Slow-Scale SSA -

- In $(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})$ at t, evaluate $\bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})=\left\langle a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(\infty),\mathbf{x}^{\mathrm{s}}\right)\right\rangle$, $j=1,\ldots,M_{\mathrm{s}}$.
  This is the hard part. Approximations may be necessary to get the first and second moments of $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$; e.g., $\left\langle\hat{\mathbf{X}}^{\mathrm{f}}(\infty)\right\rangle\approx\hat{\mathbf{X}}^{\mathrm{RRE}}(\infty)$.
- With $\bar{a}_0^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})\equiv\sum_{j=1}^{M_{\mathrm{s}}}\bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})$,
  - take the time τ to the *next slow* reaction to be the exponential random variable with mean $1/\bar{a}_0^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})$;
  - take the index j of that reaction to be the integer random variable with probability $\bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})/\bar{a}_0^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}},\mathbf{x}^{\mathrm{s}})$.
- Implement the next slow reaction by *first* updating

$$t\leftarrow t+\tau\quad\text{and}\quad\mathbf{x}\leftarrow\mathbf{x}+\boldsymbol{\nu}_j^{\mathrm{s}},$$

  and *then* "relaxing" the *fast* variables,

$$\mathbf{x}^{\mathrm{f}}\leftarrow\text{sample of }\hat{\mathbf{X}}^{\mathrm{f}}(\infty).$$

- A Variation -

Recall in our proof of the Lemma that

$$\bar{a}_j^{\rm s}(\mathbf{x}^{\rm f},\mathbf{x}^{\rm s}) = \left\langle a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(\infty),\mathbf{x}^{\rm s}\right)\right\rangle \approx \frac{1}{\Delta_{\rm s}}\int_t^{t+\Delta_{\rm s}} a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(t'),\mathbf{x}^{\rm s}\right)dt'$$

- If the ***ensemble*** *average* cannot be conveniently and accurately calculated, then compute the ***temporal*** *average* instead:

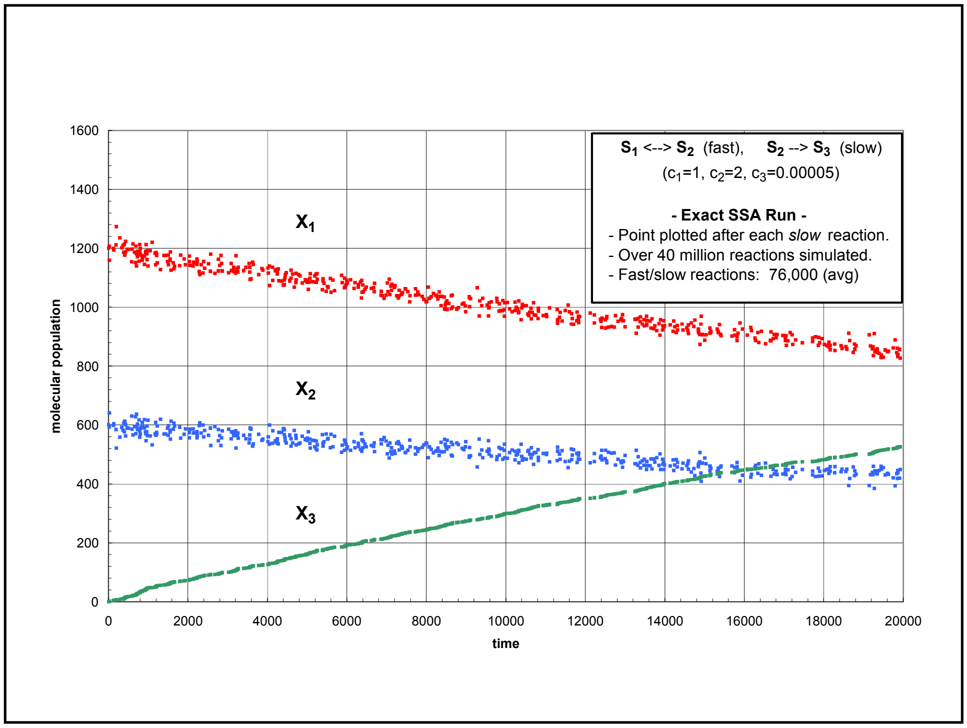
  - Use the SSA to *simulate* $\hat{\mathbf{X}}^{\rm f}(t')$ from time t to time $t+T_{\rm f}$, where $T_{\rm f}$ is "large" compared to the relaxation time of $\hat{\mathbf{X}}^{\rm f}(t')$.
  - Numerically evaluate $\int_t^{t+T_{\rm f}} a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(t'),\mathbf{x}^{\rm s}\right)dt' \Big/ T_{\rm f} \approx \bar{a}_j^{\rm s}(\mathbf{x}^{\rm f},\mathbf{x}^{\rm s})$.
  - Take the random sample of $\hat{\mathbf{X}}^{\rm f}(\infty)$ to be $\hat{\mathbf{X}}^{\rm f}(t+T_{\rm f})$.

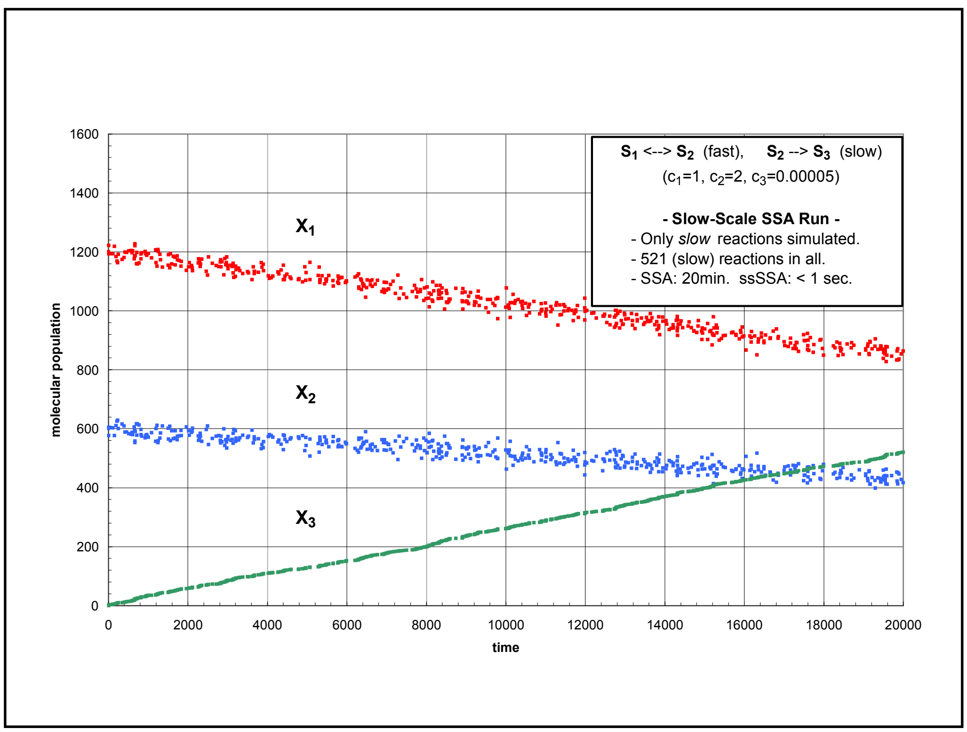
This is the "nested SSA" of …

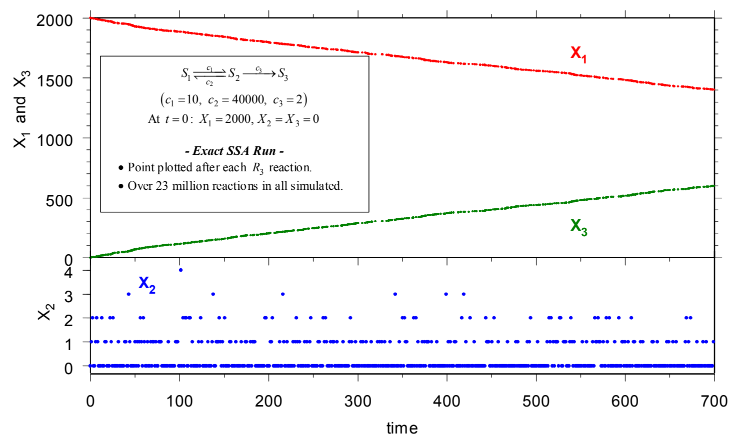
- E W, Liu D, Vanden-Eijnden E. 2005. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *J. Chem. Phys.* **123**:194107.

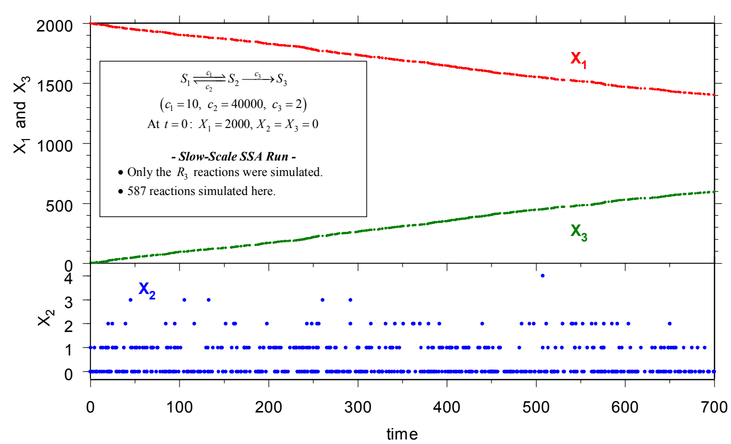
Simplest Example: $\begin{cases} S_1 \underset{c_2}{\overset{c_1}{\rightleftarrows}} S_2 & \text{(fast reactions)} \\ S_2 \xrightarrow{c_3} S_3 & \text{(slow reaction)} \end{cases}$

- $\mathbf{X}^{\rm f}(t) = \left(X_1(t), X_2(t)\right)$, $\mathbf{X}^{\rm s}(t) = X_3(t)$.
- The *virtual fast process* is $\hat{\mathbf{X}}^{\rm f}(t) = \left(\hat{X}_1(t),\hat{X}_2(t)\right)$ for $S_1 \underset{c_2}{\overset{c_1}{\rightleftarrows}} S_2$.
  - $\hat{X}_1(t)+\hat{X}_2(t) = x_{12}$. $\hat{\mathbf{X}}^{\rm f}(\infty)$ can be calculated exactly:
  - $\hat{X}_2(\infty) = \mathrm{Bin}\left(\dfrac{c_1}{c_1+c_2}, x_{12}\right)$, $\hat{X}_1(\infty) = x_{12} - \hat{X}_2(\infty)$.
- It follows that: $\bar{a}_3(\mathbf{x}) \equiv c_3\left\langle \hat{X}_2(\infty)\right\rangle = \dfrac{c_3 c_1 x_{12}}{c_1+c_2}$.
- The *relaxation time* for $\hat{X}^{\rm f}(t) \to \hat{X}^{\rm f}(\infty)$ is $1/(c_1+c_2)$; thus, the condition for using the ssSSA is: $1/(c_1+c_2) \ll (c_1+c_2)/c_3c_1x_{12}$. This condition will be satisfied if c_2 is sufficiently $\gg c_3$.


S1 <--> S2 (fast), S2 --> S3 (slow)
(c1=1, c2=2, c3=0.00005)
- Exact SSA Run -
- Point plotted after each slow reaction.
- Over 40 million reactions simulated.
- Fast/slow reactions: 76,000 (avg)
X1
X2
X3
molecular population
time


S1 <--> S2 (fast), S2 --> S3 (slow)
(c1=1, c2=2, c3=0.00005)
- Slow-Scale SSA Run -
- Only slow reactions simulated.
- 521 (slow) reactions in all.
- SSA: 20min. ssSSA: < 1 sec.
X1
X2
X3
molecular population
time


$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \xrightarrow{c_3} S_3$
$(c_1 = 10,\ c_2 = 40000,\ c_3 = 2)$
At $t = 0$: $X_1 = 2000$, $X_2 = X_3 = 0$
- Exact SSA Run -
• Point plotted after each R_3 reaction.
• Over 23 million reactions in all simulated.
X_1
X_3
X_2
X_1 and X_3
time


$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \xrightarrow{c_3} S_3$
$(c_1 = 10,\ c_2 = 40000,\ c_3 = 2)$
At $t = 0$: $X_1 = 2000$, $X_2 = X_3 = 0$
- Slow-Scale SSA Run -
• Only the R_3 reactions were simulated.
• 587 reactions simulated here.
X_1
X_3
X_2
X_1 and X_3
time

The Enzyme-Substrate Reaction

$$E + S \underset{c_2}{\overset{c_1}{\rightleftharpoons}} ES \xrightarrow{c_3} E + P$$

- Fast reactions: R_1 and R_2. Slow reaction: R_3.
- Fast species: $\mathbf{X}^{\mathrm{f}} = (X_{\mathrm{E}}, X_{\mathrm{S}}, X_{\mathrm{ES}})$. Slow species: $\mathbf{X}^{\mathrm{s}} = X_{\mathrm{P}}$.
- Virtual fast process: $\hat{\mathbf{X}}^{\mathrm{f}} = \left(\hat{X}_{\mathrm{E}}, \hat{X}_{\mathrm{S}}, \hat{X}_{\mathrm{ES}}\right)$ for $E + S \underset{c_2}{\overset{c_1}{\rightleftharpoons}} ES$.

  Two conservation eqns: $\hat{X}_{\mathrm{E}}(t) + \hat{X}_{\mathrm{ES}}(t) = x_{\mathrm{E}*}$, $\hat{X}_{\mathrm{S}}(t) + \hat{X}_{\mathrm{ES}}(t) = x_{\mathrm{S}*}$.

  $\Rightarrow$ one independent variable, say $\hat{X}_{\mathrm{ES}}(t)$. For it we have:

$$a_1(x_{\mathrm{ES}}) = c_1(x_{\mathrm{E}*} - x_{\mathrm{ES}})(x_{\mathrm{S}*} - x_{\mathrm{ES}}), \quad \nu_1 = +1;$$
$$a_2(x_{\mathrm{ES}}) = c_2 x_{\mathrm{ES}}, \quad \nu_2 = -1.$$

  The resulting $t \to \infty$ master equation for $\hat{X}_{\mathrm{ES}}(t)$ implies the "detailed balance condition".

$$c_2(x_{\mathrm{ES}}+1)\hat{P}(x_{\mathrm{ES}}+1,\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*}) = c_1(x_{\mathrm{E}*} - x_{\mathrm{ES}})(x_{\mathrm{S}*} - x_{\mathrm{ES}})\hat{P}(x_{\mathrm{ES}},\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*}), \ \forall x_{\mathrm{ES}}$$

  . . . which gives us a *recursion relation* for $\hat{P}(x_{\mathrm{ES}},\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*})$:

- $$\hat{P}(x_{\mathrm{ES}}+1,\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*}) = \frac{c_1(x_{\mathrm{E}*} - x_{\mathrm{ES}})(x_{\mathrm{S}*} - x_{\mathrm{ES}})}{c_2(x_{\mathrm{ES}}+1)}\hat{P}(x_{\mathrm{ES}},\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*})$$
$$\left(x_{\mathrm{ES}} = 0, 1, \ldots, x_{\mathrm{ES}}^{\max} - 1\right)$$

  where $x_{\mathrm{ES}}^{\max} \equiv \min(x_{\mathrm{E}*}, x_{\mathrm{S}*})$.

- In principle, we can compute $\hat{P}(x_{\mathrm{ES}},\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*})$ from this recursion relation *exactly*. Then we can *exactly*:
  - compute $\bar{a}_3(\mathbf{x}) = c_3\left\langle \hat{X}_{\mathrm{ES}}(\infty)\right\rangle = c_3 \sum_{x_{\mathrm{ES}}=0}^{x_{\mathrm{ES}}^{\max}} x_{\mathrm{ES}}\hat{P}(x_{\mathrm{ES}},\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*})$;
  - generate $\hat{X}_{\mathrm{ES}}(\infty)$ by the rule $\sum_{x_{\mathrm{ES}}=0}^{\hat{X}_{\mathrm{ES}}(\infty)} \hat{P}(x_{\mathrm{ES}},\infty \,|\, x_{\mathrm{E}*}, x_{\mathrm{S}*}) \geq r$,

    and then get $\hat{X}_{\mathrm{E}}(\infty) = x_{\mathrm{E}*} - \hat{X}_{\mathrm{ES}}(\infty)$ and $\hat{X}_{\mathrm{S}}(\infty) = x_{\mathrm{S}*} - \hat{X}_{\mathrm{ES}}(\infty)$.

- **For "small"** $x_{\mathrm{ES}}^{\max}$, all this is feasible. So we do it.
- **For "large"** $x_{\mathrm{ES}}^{\max}$, we have to make some *approximations*.

- Estimating $\bar{a}_3(\mathbf{x}) = c_3\langle \hat{X}_{\mathrm{ES}}(\infty)\rangle$ for "large" $x_{\mathrm{ES}}^{\max}$ -

- Tests show that in this case, $\langle \hat{X}_{\mathrm{ES}}(\infty)\rangle \approx \bar{x}_{\mathrm{ES}}$, where:

$$0 = c_1(x_{\mathrm{E}*} - \bar{x}_{\mathrm{ES}})(x_{\mathrm{S}*} - \bar{x}_{\mathrm{ES}}) - c_2\bar{x}_{\mathrm{ES}}.$$

Solving this gives

$$\bar{x}_{\mathrm{ES}} = \frac{1}{2}\left\{\left(x_{\mathrm{E}*} + x_{\mathrm{S}*} + \frac{c_2}{c_1}\right) - \sqrt{\left(x_{\mathrm{E}*} + x_{\mathrm{S}*} + \frac{c_2}{c_1}\right)^2 - 4x_{\mathrm{E}*}x_{\mathrm{S}*}}\right\}.$$

- So for large $x_{\mathrm{ES}}^{\max}$, we approximate $\bar{a}_3(\mathbf{x}) \approx c_3\bar{x}_{\mathrm{ES}}$.

➢ **Connection to Michaelis-Menten:** Can prove that

$$c_3\bar{x}_{\mathrm{ES}} = \frac{c_3 x_{\mathrm{E}*}\bar{x}_{\mathrm{S}}}{(c_2/c_1) + \bar{x}_{\mathrm{S}}} \left(\equiv v_{\mathrm{MM}}\right)$$

when $\bar{x}_{\mathrm{S}} = x_{\mathrm{S}*} - \bar{x}_{\mathrm{ES}}$ *and* we invoke the ***rapid equilibrium assumption***

$$c_1\left(\bar{x}_{\mathrm{E}*} - \bar{x}_{\mathrm{ES}}\right)\bar{x}_{\mathrm{S}} = c_2\bar{x}_{\mathrm{ES}}.$$

We have thus **derived** the Michaelis-Menten formula as a *large population approximation to* $\bar{a}_3(\mathbf{x})$, but *without* having to make any *assumption* of "rapid equilibrium" or "quasi-steady state".

- Generating a random sample of $\hat{X}_{\mathrm{ES}}(\infty)$ for "large" $x_{\mathrm{ES}}^{\max}$ -

- For this, tests show that we can approximate $\hat{X}_{\mathrm{ES}}(\infty)$ as a *normal* random variable whose *mean* is $\bar{x}_{\mathrm{ES}}$ **…**

… and whose *variance* is the variance $\bar{\sigma}_{\mathrm{G}}^2$ of the Gaussian curve that best fits the *peak* of the function $\hat{P}(x_{\mathrm{ES}},\infty \mid x_{\mathrm{E}*}, x_{\mathrm{S}*})$ defined by the recursion relation. Can show that

$$\bar{\sigma}_{\mathrm{G}}^2 = \frac{c_2\tilde{x}_{\mathrm{ES}}}{-2c_1\tilde{x}_{\mathrm{ES}} + c_1(x_{\mathrm{E}*} + x_{\mathrm{S}*} + 2) + c_2},$$

where

$$\tilde{x}_{\mathrm{ES}} = \frac{1}{2}\left\{\left(x_{\mathrm{E}*} + x_{\mathrm{S}*} + \frac{c_2}{c_1} + 2\right) - \sqrt{\left(x_{\mathrm{E}*} + x_{\mathrm{S}*} + \frac{c_2}{c_1} + 2\right)^2 - 4(x_{\mathrm{E}*}+1)(x_{\mathrm{S}*}+1)}\right\}.$$

- So we get $\hat{X}_{\mathrm{ES}}(\infty)$ by Monte Carlo sampling $\mathcal{N}(\bar{x}_{\mathrm{ES}}, \bar{\sigma}_{\mathrm{G}}^2)$. Then we take $\hat{X}_{\mathrm{E}}(\infty) = x_{\mathrm{E}*} - \hat{X}_{\mathrm{ES}}(\infty)$ and $\hat{X}_{\mathrm{S}}(\infty) = x_{\mathrm{S}*} - \hat{X}_{\mathrm{ES}}(\infty)$.

➢ ***Note:*** Any errors we make in generating the fast species populations will **not** affect the accuracy of the slow reaction simulations!

- The **stochastic stiffness condition** for this reaction set is:

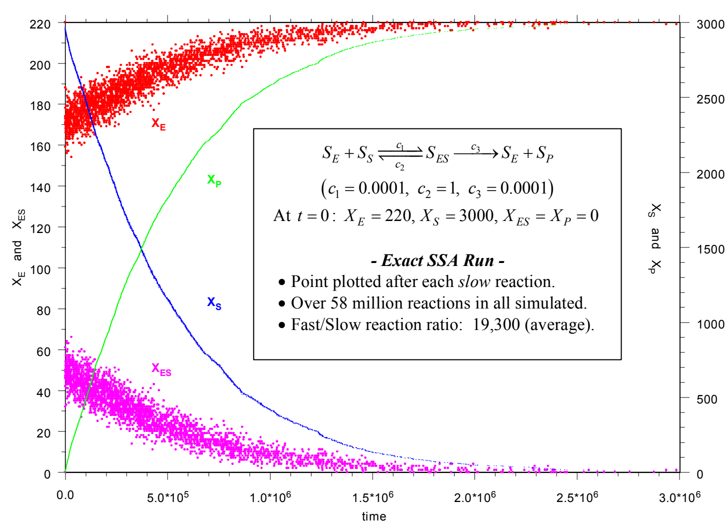
{relaxation time of $\hat{X}_{\mathrm{ES}}(t)$ } $\ll$ {expected time to the next R_3 reaction}.

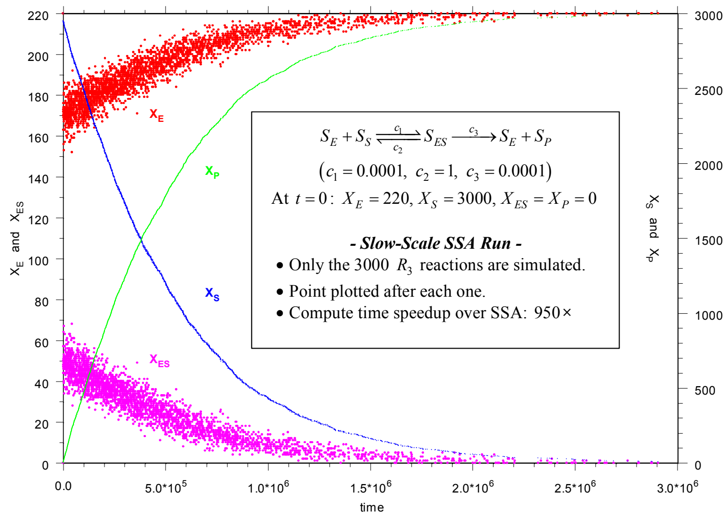
  - We can expect the ssSSA to give reliable results **if and only if** this condition is satisfied.
  - This condition turns out to be:

$$\frac{1}{-2c_1\tilde{x}_{\mathrm{ES}} + c_1(x_{\mathrm{E}*} + x_{\mathrm{S}*} + 2) + c_2} \ll \frac{1}{c_3\tilde{x}_{\mathrm{ES}}}.$$

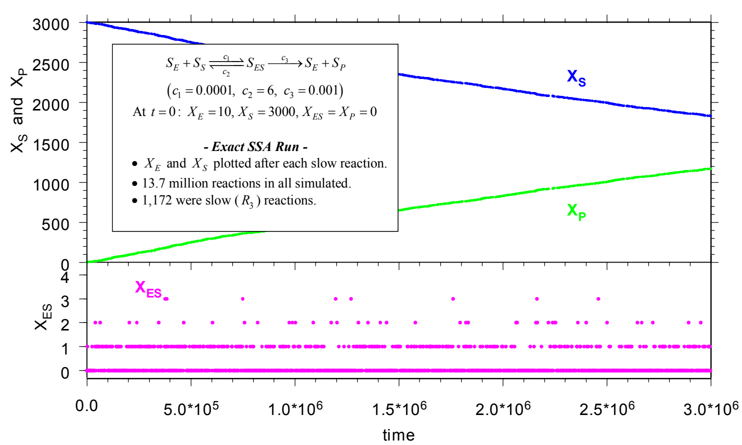
  - It will be satisfied if c_2 is sufficiently larger that c_3.

- **Physical Interpretation:** When $c_2 \gg c_3$, ES will decay to $E+S$ much more often than to $E+P$. $\Rightarrow$ Between two *successive* R_3 reactions there will be *very many* R_1 and R_2 reactions. $\Rightarrow$ R_1 and R_2 will "equilibrate" before the next R_3 reaction occurs. $\Rightarrow$ Lemma.

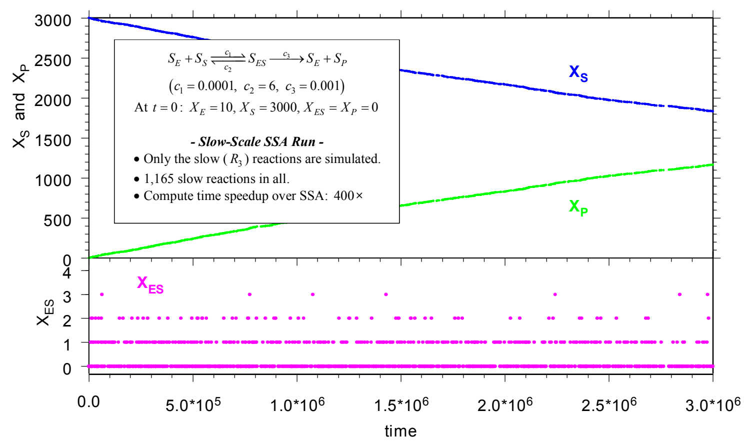




$S_E + S_S \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_{ES} \xrightarrow{c_3} S_E + S_P$
$(c_1 = 0.0001,\ c_2 = 1,\ c_3 = 0.0001)$
At $t = 0$: $X_E = 220,\ X_S = 3000,\ X_{ES} = X_P = 0$
- Slow-Scale SSA Run -
• Only the 3000 R_3 reactions are simulated.
• Point plotted after each one.
• Compute time speedup over SSA: 950×
X_E
X_P
X_S
X_{ES}
X_E and X_{ES}
X_S and X_P
time



$S_E + S_S \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_{ES} \xrightarrow{c_3} S_E + S_P$
$(c_1 = 0.0001,\ c_2 = 6,\ c_3 = 0.001)$
At $t = 0$: $X_E = 10,\ X_S = 3000,\ X_{ES} = X_P = 0$
- Exact SSA Run -
• X_E and X_S plotted after each slow reaction.
• 13.7 million reactions in all simulated.
• 1,172 were slow (R_3) reactions.
X_S
X_P
X_{ES}
X_S and X_P
time



Summary: The ssSSA …

- Provides a detailed operational procedure for determining
  - the fast and slow reactions,
  - the fast and slow species,
  - a virtual fast process $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ that is tractable.
- Requires the *stiffness conditions* to be satisfied:
  - $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ must *exist*;
  - $\hat{\mathbf{X}}^{\mathrm{f}}(t) \to \hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ *rapidly* on the time scale of the *slow* reactions.
- Is based on the ***Slow-Scale Approximation Lemma:***

  On the time scale of the slow reactions, $a_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$ can be

  approximately replaced by $\bar{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}) \triangleq \left\langle a_j^{\mathrm{s}}(\hat{\mathbf{X}}^{\mathrm{f}}(\infty), \mathbf{x}^{\mathrm{s}}) \right\rangle$.
- Simulates *only the slow reactions*, skipping over the numerous and uninteresting fast reactions, yet accurately replicates the trajectories of *all* the species.