THE GILLESPIE ALGORITHM

Andrea Giansanti
Dipartimento di Fisica, Sapienza Università di Roma

Andrea.Giansanti@roma1.infn.it

CBIO_19_L30 3/6/2019

DIPARTIMENTO DI FISICA

SAPIENZA
UNIVERSITÀ DI ROMA

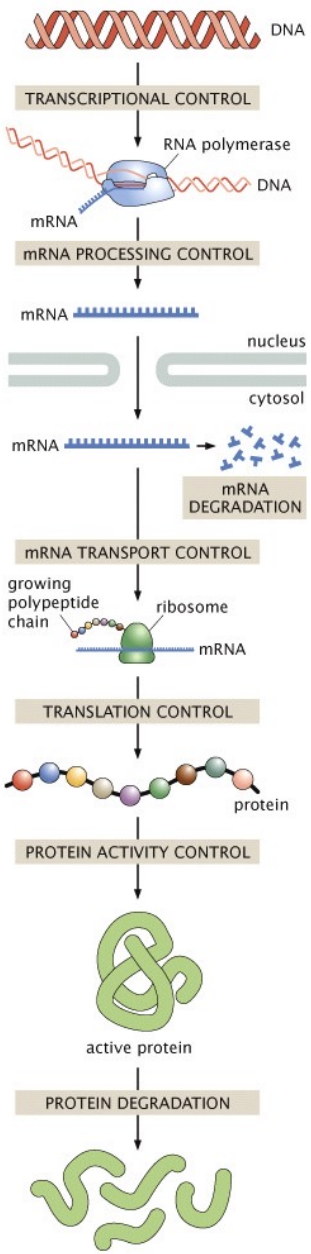


Figure 6.7 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Statistical mechanics of gene expression

- Gene expression can be measured at the mRNA level: Transcriptome
- CONTROLED BY TRANSCRIPTION FACTORS

- Gene expression can be measured at the protein level: proteome
- CONTROLLED By various mechanisms at the ribosome

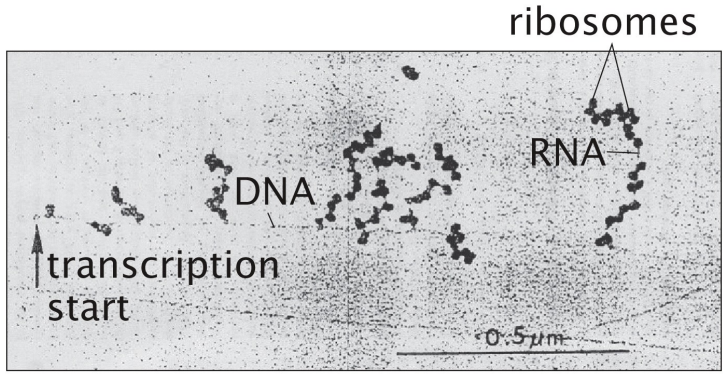


Figure 3.13 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Figure 3.9 Electron microscopy image of simultaneous transcription and translation. The image shows bacterial DNA and its associated mRNA transcripts, each of which is occupied by ribosomes. (Adapted from O. L. Miller et al., *Science* 169:392, 1970.)

Heterogeneity of phenotypes: noise

Pictoresquely, originating from cellular "decisions"

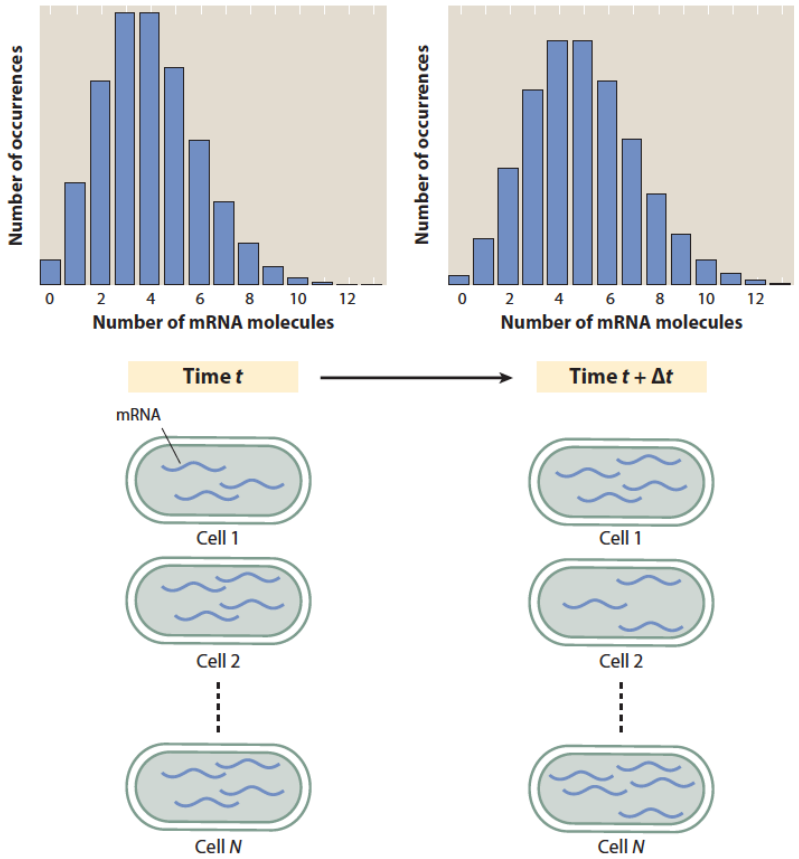


Figure 2

Transcription process resulting in change in mRNA census between times t and $t + \Delta t$. The schematic histogram shows the distribution of the number of mRNA molecules found per cell. We refer to the average number of mRNA at time t as $m(t)$; it is found by adding up the total number of mRNA over all cells and dividing by the number of cells. The number of mRNA per cell increases because of transcription and decreases because of mRNA degradation.

From Phillips2015

Focus on trascriptional control

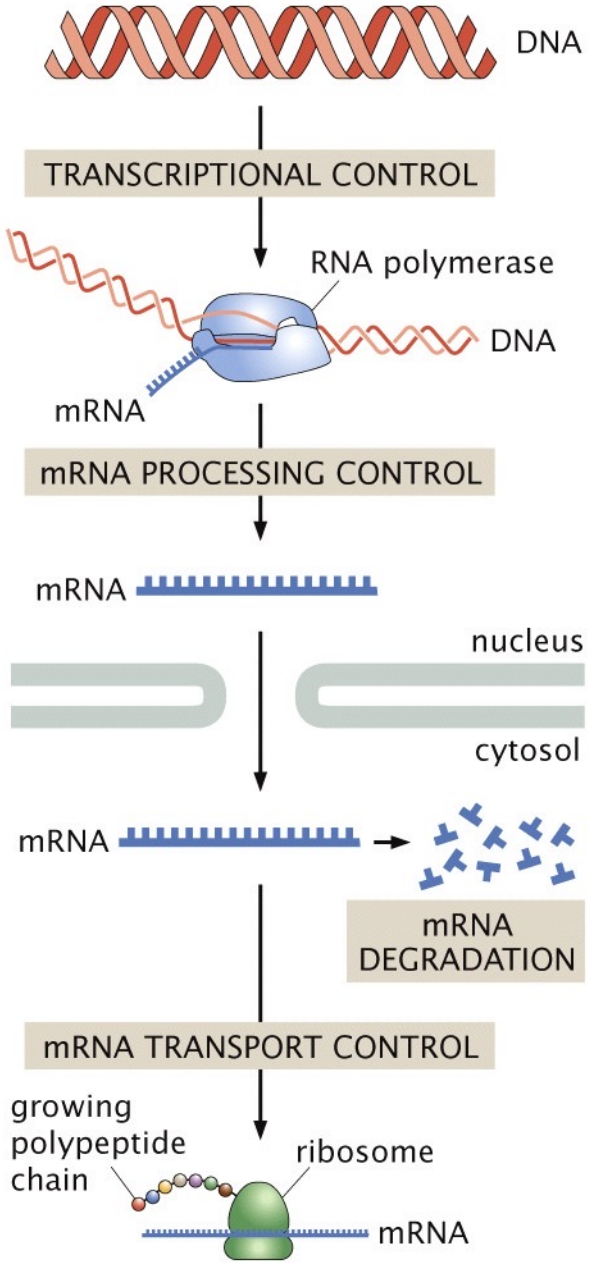


Figure 6.7 (part 1 of 2) Physical Biology of the Cell, 2ed. (© Garland Science 2013)

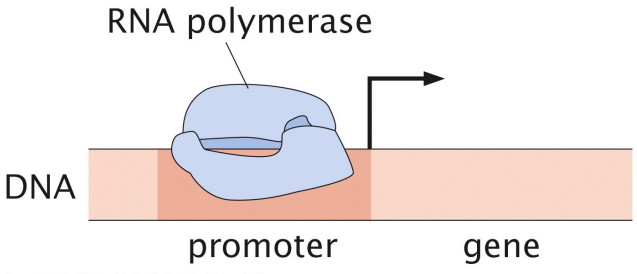


Figure 6.8 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

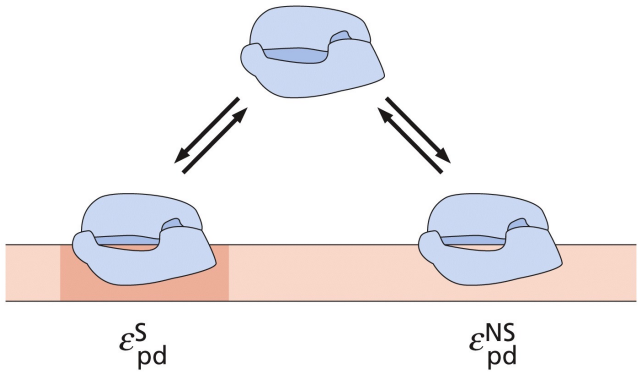


Figure 6.10 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Figure 6.10 Polymerase binding energies. Illustration of the difference in binding energy for RNA polymerase when it is bound specifically (ε^{S}_{pd}) and nonspecifically (ε^{NS}_{pd}).

Exact Stochastic Simulation of Coupled Chemical Reactions

Daniel T. Gillespie*

Research Department, Naval Weapons Center, China Lake, California 93555 *(Received May 12, 1977)*

Publication costs assisted by the Naval Weapons Center

There are two formalisms for mathematically describing the time behavior of a spatially homogeneous chemical system: The *deterministic approach* regards the time evolution as a continuous, wholly predictable process which is governed by a set of coupled, ordinary differential equations (the "reaction-rate equations"); the *stochastic approach* regards the time evolution as a kind of random-walk process which is governed by a single differential-difference equation (the "master equation"). Fairly simple kinetic theory arguments show that the stochastic formulation of chemical kinetics has a firmer physical basis than the deterministic formulation, but unfortunately the stochastic master equation is often mathematically intractable. There is, however, a way to make exact numerical calculations within the framework of the stochastic formulation without having to deal with the master equation directly. It is a relatively simple digital computer algorithm which uses a rigorously derived Monte Carlo procedure to *numerically simulate* the time evolution of the given chemical system. Like the master equation, this "stochastic simulation algorithm" correctly accounts for the inherent fluctuations and correlations that are necessarily ignored in the deterministic formulation. In addition, unlike most procedures for numerically solving the deterministic reaction-rate equations, this algorithm never approximates infinitesimal time increments dt by finite time steps Δt. The feasibility and utility of the simulation algorithm are demonstrated by applying it to several well-known model chemical systems, including the Lotka model, the Brusselator, and the Oregonator.

In this paper we shall be concerned with the following general problem: If a fixed volume V contains a spatially uniform mixture of N chemical species which can interreact through M specified chemical reaction channels, then given the numbers of molecules of each species present at some initial time, what will these molecular population levels be at any later time?

The traditional way of treating this problem begins by translating it into the mathematical language of ordinary differential equations. More specifically, if we assume that the number of molecules of the ith species in V at time t can be represented by a continuous, single-valued function $X_i(t)$ $(i = 1,\ldots,N)$, and if we further assume that each of the M chemical reactions can be regarded as a continuous rate process, then we can easily construct a set of coupled, first-order, ordinary differential equations of the form

$$\begin{aligned} dX_1/dt &= f_1(X_1,\ldots,X_N) \\ dX_2/dt &= f_2(X_1,\ldots,X_N) \\ &\cdots \\ dX_N/dt &= f_N(X_1,\ldots,X_N) \end{aligned} \tag{1}$$

The specific forms of the functions f_i on the right (which are usually nonlinear in the X_i's) are determined by the structures and rate constants of the M chemical reaction channels. These equations are called the "reaction-rate equations"; solving them for the functions $X_1(t),\ldots,X_N(t)$, subject to the prescribed initial conditions, is tantamount to solving the time-evolution problem posed earlier.

Continuous/deteministic/discrete/stochastic

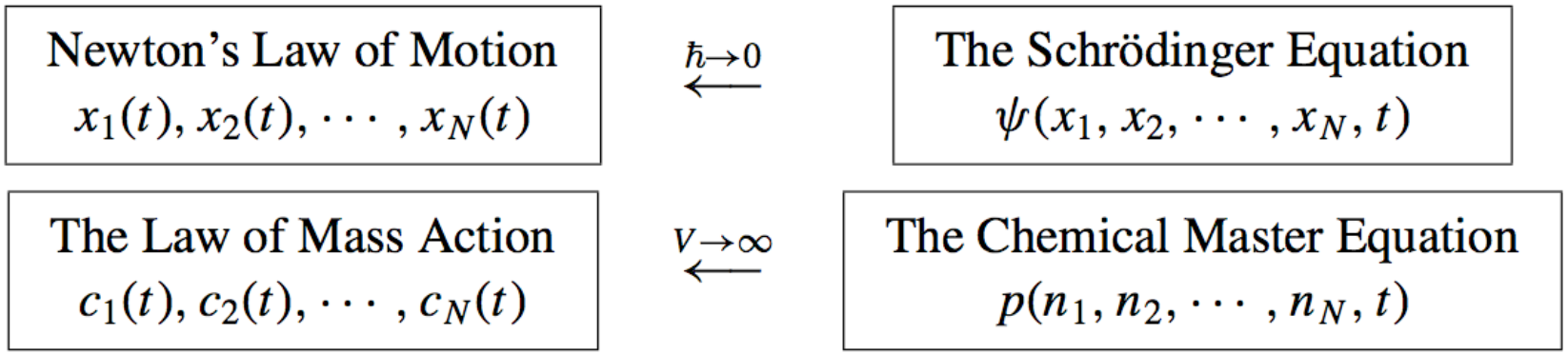


Figure 11.1 A schematic that illustrates the analogy between the theories for mechanical motions and for chemical dynamics. Newton's law of motion, governing a collection of particles with positions $x_1(t), x_2(t), \cdots, x_N(t)$, arises from Schrödinger's equation for the wave function ψ in the limit $\hbar \to 0$. Similarly, the chemical master equation for $p(n_1, n_2, \cdots, n_N, t)$ yields the law of mass action in the limit $V \to \infty$.

The master Chemical Equation IN GENERAl
(see D. T. Gillespie, Physica A188 404-425 (1992)

Let us consider a system made of chemical reacting species,
contained in a thermally isolated stirred vessel

- Composed of n chemical species $\{S_1, \ldots, S_n\}$ and m chemical reaction channels $\{R_1, \ldots, R_m\}$.
- Assume species contained within constant volume Ω.
- Assume system is *well-stirred* to neglect spatial effects.
- Assume system is in *thermal equilibrium* (i.e., at a constant temperature), but not necessarily *chemical equilibrium*.

- $X_i(t)$ is the number of molecules of S_i at time t.
- $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ is the state of a system at time t.
- $\mathbf{X}(t_0) = \mathbf{x}_0$ is initial number of molecules at initial time t_0.
- After R_μ, the new state is $\mathbf{x}' = \mathbf{x} + v_\mu$ where $v_\mu = (v_{1\mu}, \ldots, v_{n\mu})$ is the *state-change vector* and $v_{i\mu}$ is the change in S_i due to R_μ.
- The 2-dimensional array $\{v_{i\mu}\}$ is known as the *stoichiometric matrix*.
- R_μ is *elemental* if it can be considered a distinct physical event that happens nearly instantaneously.
- For elemental R_μ, values of $v_{i\mu}$ are constrained to $0, \pm 1, \pm 2$.

- Every R_μ has a *specific probability rate constant*, c_μ, which is related to the reaction rate constant, k_μ.
- $c_\mu dt$ is the probability that a randomly chosen combination of reactant molecules react as defined by R_μ inside Ω in $[t, t+dt)$.
- Multiplying c_μ by the number of possible combinations of reactant molecules for R_μ in a state $\mathbf{x}$ yields the *propensity function*, a_μ.
- $a_\mu(\mathbf{x})dt$ is the probability that R_μ occurs in the state $\mathbf{x}$ within Ω in the next infinitesimal time interval $[t, t+dt)$.

- Not possible to know the exact state $\mathbf{X}(t)$.
- Only can know probability of being in a given state at time t starting from a state $\mathbf{X}(t_0) = \mathbf{x}_0$ (i.e., $\mathcal{P}(\mathbf{x}, t|\mathbf{x}_0, t_0)$).
- Probability using a time-evolution of step dt is shown below:

$$\begin{aligned}\mathcal{P}(\mathbf{x}, t+dt|\mathbf{x}_0, t_0) &= \mathcal{P}(\mathbf{x}, t|\mathbf{x}_0, t_0) \times \left[1 - \sum_{j=1}^{m} (a_j(\mathbf{x})dt)\right] \\ &+ \sum_{j=1}^{m} \mathcal{P}(\mathbf{x} - v_j, t|\mathbf{x}_0, t_0) \times (a_j(\mathbf{x} - v_j)dt)\,.\end{aligned}$$

- dt is small enough that at most one reaction occurs during dt.

- *Chemical master equation* (CME) defines time evolution of state probabilities, $\mathcal{P}(\mathbf{x}, t|\mathbf{x}_0, t_0)$:

$$\begin{aligned}\frac{\partial \mathcal{P}(\mathbf{x}, t|\mathbf{x}_0, t_0)}{dt} &= \lim_{dt \to 0} \frac{\mathcal{P}(\mathbf{x}, t+dt|\mathbf{x}_0, t_0) - \mathcal{P}(\mathbf{x}, t|\mathbf{x}_0, t_0)}{dt} \\ &= \sum_{j=1}^{m} [a_j(\mathbf{x} - v_j)\mathcal{P}(\mathbf{x} - v_j, t|\mathbf{x}_0, t_0) \\ &\quad - a_j(\mathbf{x})\mathcal{P}(\mathbf{x}, t|\mathbf{x}_0, t_0)]\end{aligned}$$

- Typically cannot be solved analytically since it represents a set of equations as large as the number of molecules in the system.

$\overline{\delta V_{\text{coll}}/V} = V^{-1}\pi r_{12}^2 \overline{v_{12}}\,\delta t$ Rate constants that are probabilities: from molecular collisions

= average probability that a particular 1–2 molecular pair will collide in the next vanishingly small time interval δt (2a)

For Maxwellian velocity distributions the average relative speed $\overline{v_{12}}$ will be equal to $(8kT/\pi m_{12})^{1/2}$, where k is Boltzmann's constant, T the absolute temperature, and m_{12} the reduced mass $m_1 m_2/(m_1 + m_2)$. In any case, if we are given that at time t there are in VX_1 of the S_1 molecules and X_2 of the S_2 molecules, making a total of X_1X_2 distinct 1–2 molecular pairs, then it follows from (2a) that

$X_1X_2V^{-1}\pi r_{12}^2\overline{v_{12}}\,dt$ = probability that a 1–2 collision will occur somewhere inside V in the next infinitesimal time interval $(t, t + dt)$ (2b)

We see then that, although we *cannot* rigorously calculate the *number* of 1–2 collisions occurring in V in any infinitesimal time interval, we *can* rigorously calculate the *probability* of a 1–2 collision occurring in V in any infinitesimal time interval. This means that we really ought to characterize a system of thermally equilibrized molecules by a "collision probability per unit time"—namely, the coefficient of dt in (2b)—instead of by a "collision rate". This is why these collisions constitute a stochastic Markov process instead of a deterministic rate process.

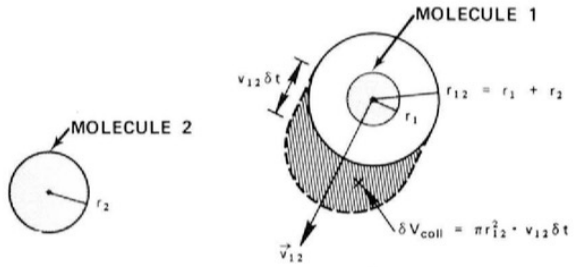


Figure 1. The "collision volume" δV_{coll} which molecule 1 will sweep out relative to molecule 2 in the next small time interval δt.

More generally, suppose the volume V contains a spatially homogeneous (or thermally equilibrized) mixture of X_i molecules of chemical species S_i ($i = 1,\ldots,N$), and suppose further that these N species can interreact through M specified chemical reaction channels R_μ ($\mu = 1,\ldots,M$). Then we may assert the existence of M constants c_μ ($\mu = 1,\ldots,M$), which depend only on the physical properties of the molecules and the temperature of the system, such that

$$c_\mu\, dt = \text{average probability that a particular combination of } R_\mu \text{ reactant molecules will react accordingly in the next infinitesimal time interval } dt \qquad (4)$$

By "average" here we mean simply that, if we multiply $c_\mu\, dt$ by the total number of distinct combinations of R_μ reactant molecules in V at time t, we will obtain the probability that an R_μ reaction will occur somewhere inside V in the next infinitesimal time interval $(t, t + dt)$.

Equation 4 may be regarded both as the *definition* of the stochastic reaction constant c_μ, and also as the *fundamental hypothesis* of the stochastic formulation of chemical kinetics. As is shown by more detailed arguments in ref 1, we may expect this hypothesis to be valid for any molecular system that is kept "well-mixed", either by

IIIA. *The Master Equation Approach.* The traditional method of calculating the stochastic time evolution of a chemically reacting system is to set up and solve a so-called "master equation" for the system. A good review of the master equation approach to chemical kinetics has been given by McQuarrie.[3] Here we would merely like to summarize briefly the main features of the master equation formalism in order to provide a conceptual setting for our subsequent presentation of the stochastic simulation approach. It should be emphasized, though, that the master equation itself plays no role whatsoever in either the derivation or the implementation of the stochastic simulation algorithm.

The key element of the master equation formalism is the "grand probability function"

$$P(X_1,X_2,\ldots,X_N;t) \equiv \text{probability that there will be in } V \text{ at time } t \; X_1 \text{ molecules of species } S_1, \text{ and } X_2 \text{ molecules of species } S_2,\ldots, \text{ and } X_N \text{ molecules of species } S_N \tag{7}$$

A knowledge of this function would evidently provide a fairly complete characterization of the "stochastic state" of the system at time t. Thus, for example, the kth moment of P with respect to X_i

$$X_i^{(k)}(t) \equiv \sum_{X_1=0}^{\infty} \cdots \sum_{X_N=0}^{\infty} X_i^k P(X_1,\ldots,X_N;t)$$
$$(i = 1,\ldots,N; \quad k = 0, 1, 2,\ldots) \tag{8}$$

gives the "average (number)^k of S_i molecules in V at time t". By "average" here we mean an average taken over many repeated "runs" from time 0 to time t of the stochastic process defined by (4), each run having the same initial numbers of molecules.[4] The number X_i of S_i molecules found at time t will vary from run to run, but the average of the kth power of these numbers will approach $X_i^{(k)}(t)$ in the limit of infinitely many runs. Especially useful are the $k = 1$ and $k = 2$ moments; this is because $X_i^{(1)}(t)$ and

$$\Delta_i(t) \equiv \{X_i^{(2)}(t) - [X_i^{(1)}(t)]^2\}^{1/2} \tag{9}$$

The master equation is simply the time-evolution equation for the function $P(X_1,\ldots,X_N; t)$. It may be derived from (4) by using the addition and multiplication laws of probability theory to write $P(X_1,\ldots,X_N; t + dt)$ as the *sum* of the probabilities of the $1 + M$ different ways in which the system can arrive at the state $(X_1,\ldots,X_N)$ at time $t + dt$:[5]

$$P(X_1,\ldots,X_N; t + dt) = P(X_1,\ldots,X_N; t)\left[1 - \sum_{\mu=1}^{M} a_\mu\, dt\right] + \sum_{\mu=1}^{M} B_\mu\, dt \qquad (11)$$

Here we have defined the quantities a_μ by

$$a_\mu\, dt \equiv c_\mu\, dt \times \{\text{number of distinct } R_\mu \text{ molecular combinations in the state } (X_1,\ldots,X_N)\}$$

$$= \text{probability that an } R_\mu \text{ reaction will occur in } V \text{ in } (t,\ t + dt), \textit{given} \text{ that the system is in the state } (X_1,\ldots,X_N) \text{ at time } t \qquad (12)$$

$t + dt$). The quantity $B_\mu\, dt$ gives the probability that the system is one R_μ reaction *removed* from the state $(X_1,\ldots,X_N)$ at time t, and then undergoes an R_μ reaction in $(t,\ t + dt)$. Thus, B_μ will be the product of P evaluated at the appropriate once-removed state at t, times c_μ, times the number of R_μ molecular reactant combinations available in that once-removed state. Without going into any further details, it is sufficient here to simply observe that (11) leads directly to the "master equation"

$$\frac{\partial}{\partial t}P(X_1,\ldots,X_N; t) = \sum_{\mu=1}^{M} [B_\mu - a_\mu P(X_1,\ldots,X_N; t)] \qquad (13)$$

In any particular case, the master equation is fairly easy to write; however, *solving* it is quite another matter. The number of problems for which the master equation (13) can be solved analytically is even fewer than the number of problems for which the deterministic reaction-rate equations (1) can be solved analytically. In addition, unlike

Prompted by these considerations, we introduce the function $P(\tau, \mu)$ defined by[1]

$$P(\tau, \mu)\, d\tau \equiv \text{probability that, given the state } (X_1,\ldots,X_N) \text{ at time } t, \text{ the } \textit{next} \text{ reaction in } V \text{ will occur in the infinitesimal time interval } (t+\tau,\ t+\tau+d\tau), \textit{ and} \text{ will be an } R_\mu \text{ reaction} \qquad (14)$$

We call $P(\tau, \mu)$ the "reaction probability density function", because in mathematical terminology it is a joint probability density function on the space of the continuous variable τ ($0 \leq \tau < \infty$) and the discrete variable μ ($\mu = 1, 2,\ldots,M$). Notice that the variables τ and μ are quantities whose respective values would give us answers to the two questions mentioned above. Our first step toward finding a legitimate method for assigning numerical values to τ and μ is to derive, from the fundamental hypothesis (4), an analytical expression for $P(\tau, \mu)$.

To this end, we begin by defining for each reaction R_μ a function h_μ according to

$$h_\mu \equiv \text{number of distinct } R_\mu \text{ molecular reactant combinations available in the state } (X_1,X_2,\ldots,X_N) \quad (\mu = 1,\ldots,M) \qquad (15)$$

Thus, if R_μ has the form $S_1 + S_2 \rightarrow$ anything, then we will have $h_\mu = X_1X_2$; if R_μ has the form $2S_1 \rightarrow$ anything, then we will have $h_\mu = {}^1/_2X_1(X_1 - 1)$. In general, h_μ will be some combinatorial function of the variables $X_1, X_2,\ldots,X_N$. With h_μ so defined, then (4) implies that [cf. (12)]

$$a_\mu\, dt \equiv h_\mu c_\mu\, dt = \text{probability that an } R_\mu \text{ reaction will occur in } V \text{ in } (t,\ t+dt), \textit{given} \text{ that the system is in the state } (X_1,\ldots,X_N) \text{ at time } t \quad (\mu = 1,\ldots,M) \qquad (16)$$

We now calculate the probability in (14) as the *product* of: $P_0(\tau)$, the probability that, given the state $(X_1,\ldots,X_N)$ at time t, no reaction will occur in the time interval $(t,\ t+\tau)$; *times* $a_\mu\, d\tau$, the subsequent probability that an R_μ reaction will occur in the time interval $(t+\tau,\ t+\tau+d\tau)$:[5]

$$P(\tau, \mu)\, d\tau = P_0(\tau) \cdot a_\mu\, d\tau \qquad (17a)$$

We now calculate the probability in (14) as the *product* of: $P_0(\tau)$, the probability that, given the state $(X_1,\ldots,X_N)$ at time t, no reaction will occur in the time interval $(t, t + \tau)$; *times* $a_\mu\, d\tau$, the subsequent probability that an R_μ reaction will occur in the time interval $(t + \tau, t + \tau + d\tau)$:[5]

$$P(\tau, \mu)\, d\tau = P_0(\tau)\cdot a_\mu\, d\tau \tag{17a}$$

To find an expression for $P_0(\tau)$, we first note that $[1 - \Sigma_\nu a_\nu\, d\tau']$ is the probability that no reaction will occur in time $d\tau'$ from the state $(X_1,\ldots,X_N)$. Therefore

$$P_0(\tau' + d\tau') = P_0(\tau')\cdot[1 - \sum_{\nu=1}^{M} a_\nu\, d\tau'] \tag{17b}$$

from which it is readily deduced that

$$P_0(\tau) = \exp[-\sum_{\nu=1}^{M} a_\nu \tau] \tag{17c}$$

Inserting (17c) into (17a), we conclude that the reaction probability density function defined in (14) is given by[7]

$$P(\tau, \mu) = \begin{cases} a_\mu \exp(-a_0\tau) & \text{if } 0 \leqslant \tau < \infty \text{ and} \\ & \mu = 1,\ldots,M \\ 0 \quad \text{otherwise} \end{cases} \tag{18}$$

where

$$a_\mu \equiv h_\mu c_\mu \qquad (\mu = 1,\ldots,M) \tag{19a}$$

and

$$a_0 \equiv \sum_{\nu=1}^{M} a_\nu \equiv \sum_{\nu=1}^{M} h_\nu c_\nu \tag{19b}$$

function $P(\tau, \mu)$ in (18), the construction procedure turns out to be as follows:

With r_1 and r_2 two random numbers from the unit-interval uniform distribution, take

$$\tau = (1/a_0) \ln (1/r_1) \tag{21a}$$

and take μ to be that integer for which

$$\sum_{\nu=1}^{\mu-1} a_\nu < r_2 a_0 \leqslant \sum_{\nu=1}^{\mu} a_\nu \tag{21b}$$

Our algorithm for simulating the stochastic time evolution of a chemically reacting system should now be rather obvious (see Figure 2):

Step 0 (Initialization). Input the desired values for the M reaction constants $c_1,\ldots,c_M$ and the N initial molecular population numbers $X_1,\ldots,X_N$. Set the time variable t and the reaction counter n both to zero. Initialize the unit-interval uniform random number generator (URN).

Step 1. Calculate and store the M quantities $a_1 = h_1c_1,\ldots,a_M = h_Mc_M$ for the current molecular population numbers, where h_ν is that function of $X_1,\ldots,X_N$ defined in (15). Also calculate and store as a_0 the sum of the M a_ν values.

Step 2. Generate two random numbers r_1 and r_2 using the unit-interval uniform random number generator, and calculate τ and μ according to (21a) and (21b).

Step 3. Using the τ and μ values obtained in step 2, increase t by τ, and adjust the molecular population levels to reflect the occurrence of one R_μ reaction; e.g., if R_μ is the reaction in (3a), then increase X_1 by 1 and decrease X_2 by 1. Then increase the reaction counter n by 1 and return to step 1.

In returning to step 1 from step 3, notice that it is necessary to recalculate only those quantities a_ν corresponding to reactions R_ν whose *reactant* population levels were just altered in step 3; also, a_0 may be recalculated simply by adding to a_0 the *difference* between each newly changed a_ν value and its corresponding old value.

Of course, somewhere in the 1–2–3 loop one will want to provide for writing out or plotting the $(X_1,\ldots,X_N, t)$ values at regular intervals of either t or n. Also, one will want to make provisions for halting the calculations when either t or n reaches some predetermined value, or if a_0 should ever reach zero.

distribution in the unit interval, and take

$$\tau = \frac{1}{a_0(\mathbf{x})} \ln\left(\frac{1}{r_1}\right), \tag{10a}$$

$$j = \text{the smallest integer satisfying } \sum_{j'=1}^{j} a_{j'(\mathbf{x})} > r_2\, a_0(\mathbf{x}). \tag{10b}$$

Stochastic simulation algorithm (SSA): a Monte Carlo procedure for numerically generating time trajectories of the molecular populations in exact accordance with the CME

With this generating method (or any mathematically equivalent one), we have the following stochastic simulation algorithm (SSA) for constructing an exact numerical realization of the process $\mathbf{X}(t)$ (8, 9):

0. Initialize the time $t = t_0$ and the system's state $\mathbf{x} = \mathbf{x}_0$.
1. With the system in state $\mathbf{x}$ at time t, evaluate all the $a_j(\mathbf{x})$ and their sum $a_0(\mathbf{x})$.
2. Generate values for τ and j using Equations 10a,b (or their equivalent).
3. Effect the next reaction by replacing $t \leftarrow t + \tau$ and $\mathbf{x} \leftarrow \mathbf{x} + \boldsymbol{\nu}_j$.
4. Record $(\mathbf{x}, t)$ as desired. Return to Step 1, or else end the simulation.

The $\mathbf{X}(t)$ trajectory produced by the SSA may be thought of as a stochastic version of the trajectory that would be obtained by solving the RRE (Equation 6). But note that the time step τ in the SSA is exact and not a finite approximation to some infinitesimal dt, as is the time step in a typical ODE solver. If it is found that every SSA-generated trajectory is practically indistinguishable from the RRE trajectory, then we may conclude that microscale randomness is ignorable. But if the SSA trajectories are found to deviate significantly from the RRE trajectory, or from each other, then we must conclude that microscale randomness is not ignorable, and the deterministic RRE does not provide an accurate description of the system's true behavior.

2. STOCHASTIC CHEMICAL KINETICS: THE CHEMICAL MASTER EQUATION AND THE STOCHASTIC SIMULATION ALGORITHM

Let us consider a well-stirred system of molecules of N chemical species $\{S_1, \ldots, S_N\}$, which interact through M chemical reactions $\{R_1, \ldots, R_M\}$. We assume that the system is confined to a constant volume Ω and is in thermal (but not chemical) equilibrium at some constant temperature. We let $X_i(t)$ denote the number of molecules of species S_i in the system at time t. Our goal is to estimate the state vector $\mathbf{X}(t) \equiv (X_1(t), \ldots, X_N(t))$, given that the system was in state $\mathbf{X}(t_0) = \mathbf{x}_0$ at some initial time t_0.

The justification for the tacit assumption that we can describe the system's state by specifying only the molecular populations, ignoring the positions and velocities of the individual molecules, lies in the conditions responsible for the system being well stirred. The fundamental assumption being made is that the overwhelming majority of molecular collisions that take place in the system are elastic (nonreactive), and further that the net effect of these elastic collisions is twofold: First, the positions of the molecules become uniformly randomized throughout Ω; second, the velocities of the molecules become thermally randomized to the Maxwell-Boltzmann distribution. To the extent that this happens, we can ignore the nonreactive molecular collisions, the simulation of which would occupy most of the computation time in a molecular dynamics simulation, and concern ourselves only with events that change the populations of the chemical species. This simplifies the problem enormously.

The changes in the species populations are of course a consequence of the chemical reactions. Each reaction channel R_j is characterized mathematically by two quantities. The first is its state-change vector $\boldsymbol{\nu}_j \equiv (\nu_{1j}, \ldots, \nu_{Nj})$, where ν_{ij} is the change in the S_i molecular population caused by one R_j reaction, so if the system is in state $\mathbf{x}$ and one R_j reaction occurs, the system immediately jumps to state $\mathbf{x} + \boldsymbol{\nu}_j$. The other characterizing quantity for R_j is its propensity function a_j, which is defined so that

$$a_j(\mathbf{x})\,dt \stackrel{\Delta}{=} \text{the probability, given } \mathbf{X}(t) = \mathbf{x}, \text{ that one } R_j \text{ reaction will occur somewhere inside } \Omega \text{ in the next infinitesimal time interval } [t, t + dt). \tag{2}$$

Definition 2 can be regarded as the fundamental premise of stochastic chemical kinetics because everything else in the theory follows from it via the laws of probability. The physical rationale for Definition 2 for unimolecular and bimolecular reactions can be briefly summarized as follows.

Alcune slides tratte da una lezione di Marco S. Nobile

- Poiché descriviamo il sistema in termini probabilistici, possiamo essere interessati a determinare $P(\mathbf{X}, t|\mathbf{X}_0, t_0)$
  - Ovvero, la probabilità di trovarci nello stato $\mathbf{X}$ al tempo t, partendo dallo stato iniziale $\mathbf{X}_0$ al tempo t_0

- Un metodo: Chemical Master Equation (CME)

$$\frac{\partial P(\mathbf{X}, t|\mathbf{X}_0, t_0)}{\partial t} = \sum_{\mu=1}^{M} [a_\mu(\mathbf{X} - \boldsymbol{\nu}_\mu) P(\mathbf{X} - \boldsymbol{\nu}_\mu, t|\mathbf{X}_0, t_0) - a_\mu(\mathbf{X}) P(\mathbf{X}, t|\mathbf{X}_0, t_0)]$$

  - a_μ è la propensity function della reazione R_μ
  - $\boldsymbol{\nu}_\mu$ è il vettore di variazione di stato legato alla reazione R_μ

- La CME determina la probabilità di trovarci in uno stato specifico
  - Domanda: in quanti stati diversi può trovarsi un sistema biochimico?

- $c_\mu dt$ = probabilità media che una particolare combinazione di molecole che compaiono come reagenti nella reazione R_μ reagiscano nell'intervallo di tempo infinitesimale dt
  - Ragionamento «al limite»
  - Stiamo negando ogni considerazione «spaziale»

- $a_\mu(\mathbf{X}(t)) = c_\mu h_\mu(t) =$ propensity function della reazione R_μ
  - Stato $\mathbf{X}(t) \equiv (X_1, \dots, X_N)$ ovvero l'esatto numero di molecole delle N specie al tempo t
  - $h_\mu(t)$ è il numero di combinazioni distinte di reagenti di R_μ al tempo t

- $a_\mu(\mathbf{X}(t))dt =$ probabilità che una reazione R_μ avverrà nell'intervallo infinitesimale $[t, t + dt)$, per via del fatto che ci troviamo nello stato $\mathbf{X}(t)$

- SSA è un algoritmo di simulazione stocastica di sistemi biochimici
  - Introdotto da Gillespie nel 1976
  - Consente la simulazione dell'evoluzione stocastica temporale del sistema, per ottenere realizzazioni della CME
  - Impostazione Markoviana: il sistema evolve una reazione alla volta

- Dato uno stato iniziale del sistema, SSA procede iterativamente:
  1. Calcola le propensity functions delle reazioni, sulla base dello stato attuale del sistema $\mathbf{X}$
  2. Calcola $P(\tau, \mu|\mathbf{X}, t)$ ovvero la probabilità congiunta del tempo «di attesa» τ prima della prossima reazione R_μ
  3. Determina i valori τ e μ
  4. Aggiorna $\mathbf{X}$ in base al vettore di variazione $\boldsymbol{\nu}_\mu$ associato a R_μ
  5. Aggiorna il tempo della simulazione calcolando $t = t + \tau$
  6. Se non possono più essere eseguite reazioni, l'algoritmo termina; altrimenti, itera da 1

Determinazione di τ

- Esiste un modo per campionare τ dalla distribuzione esponenziale

$$\tau = \frac{1}{a_0(\mathbf{X})} \ln \frac{1}{r_1}$$

  - Dove r_1 è un numero pseudo-casuale campionato con distribuzione uniforme in [0,1)

- Osservazione: τ è inversamente proporzionale a a_0 dunque più cresce e più i time step dell'algoritmo saranno brevi
  - Implica una simulazione più lunga

Determinazione di μ

- Data la propensity function a_μ di una reazione R_μ e il valore cumulativo a_0 è possibile determinare la probabilità che R_μ avvenga:

$$P(\mu) = \frac{a_\mu}{a_0}$$

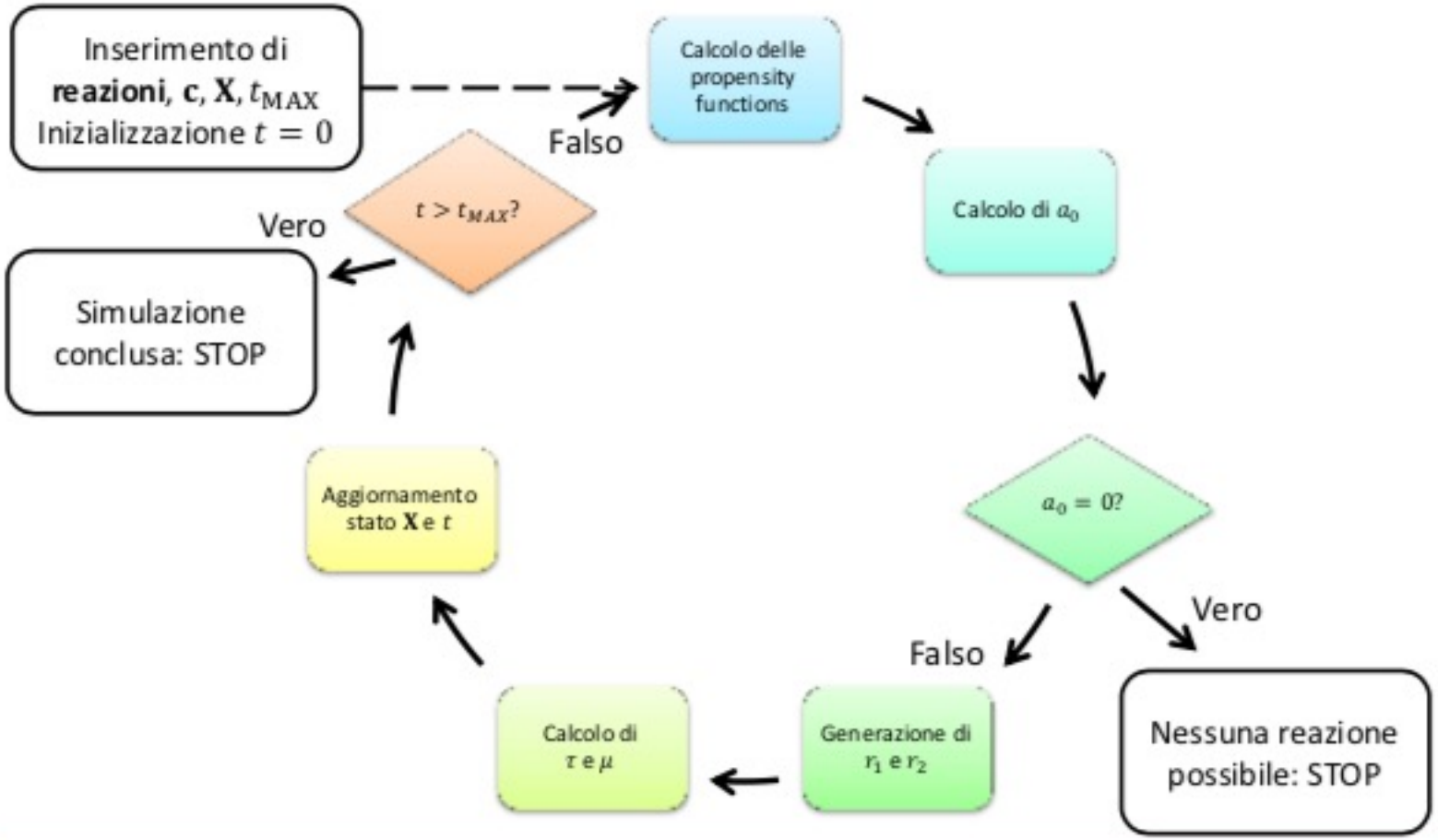
- Possiamo determinare la reazione che sarà eseguita scegliendo l'indice μ tale per cui:

$$\sum_{j=1,\dots,\mu-1} a_j(\mathbf{X}) < r_2 \cdot a_0(\mathbf{X}) \leq \sum_{j=1,\dots,\mu} a_j(\mathbf{X})$$

  - Dove r_2 è un secondo numero pseudo-casuale campionato con distribuzione uniforme in [0,1)

- Osservazione: la probabilità di R_μ è proporzionale a a_μ
  - Ma $a_\mu = c_\mu \cdot h_\mu(t)$ dunque la probabilità di una reazione è proporzionale alla sua costante stocastica e al numero di molecole dei reagenti

L'algoritmo SSA



SUMMARY POINTS

1. The SSA is a procedure for numerically simulating well-stirred chemically reacting systems by stepping in time to successive molecular reaction events in exact accord with the premises of the CME.
2. The ability of the SSA to take proper account of the discrete, stochastic nature of chemical reactions makes it better suited to cellular chemical kinetics than the traditional RRE because in cellular systems the small numbers of molecules of some key reactants can amplify the effects of discreteness and randomness.
3. Because the SSA simulates every successive molecular reaction event that occurs in the system, it is often too slow for practical simulation of realistic cellular systems.
4. An approximate speedup to the SSA is provided by tau-leaping, in which time is advanced by a preselected amount τ and the numbers of firings of the individual reaction channels are approximated by Poisson random numbers.
5. If the expected number of firings of each reaction channel during a tau-leap is much greater than one, the Poisson random numbers are well approximated by normal random numbers, and the result is equivalent to a Langevin-type equation called the CLE.
6. In the thermodynamic (macroscopic) limit, the noise terms in the CLE become negligibly small and the CLE reduces to the conventional RRE, thereby establishing deterministic chemical kinetics in the context of stochastic chemical kinetics.
7. For stiff systems—which evolve on both fast and slow timescales with the fastest modes being stable—accuracy in tau-leaping requires τ to be small on the fastest timescale, which makes even tau-leaping seem too slow.
8. Two acceleration procedures for stiff systems are implicit tau-leaping, which mirrors the implicit Euler method in ODE theory, and the ssSSA, in which the fast reactions are skipped over and only the slow reactions are directly simulated using specially modified propensity functions.

12. GILLESPIE ALGORITHM AND ENZYME KINETICS

Rate equations vs chemical master equations
Gillespie' algorithm (see Gillespie1976, Gillespie1977, Gillespie2007)
The Michaelis-Menten approximation for the rate equations of enzymic catalysis
Rate equations for the rapid equilibrium model (PBoC 15.2.6)
Michaelis-Menten Kinetics of enzymatic catalysis (PBoC 15.2.7)
Reverse Michaelis-Menten (Fabrini2011, Tang2015)
List of computational tools for chemical kinetic (Deterministic and Stochastic Simulations)
COPASI (www.copasi.org)
StochPy (Marleeveld2013)
You can download the programming tools here below from Sapienza with your INFOSTUD ID (https://campus3.uniroma1.it/campus/indexlogin.php)
in particular: Mathematica & Matlab