# 1 Foundations: Chemical Reaction Networks
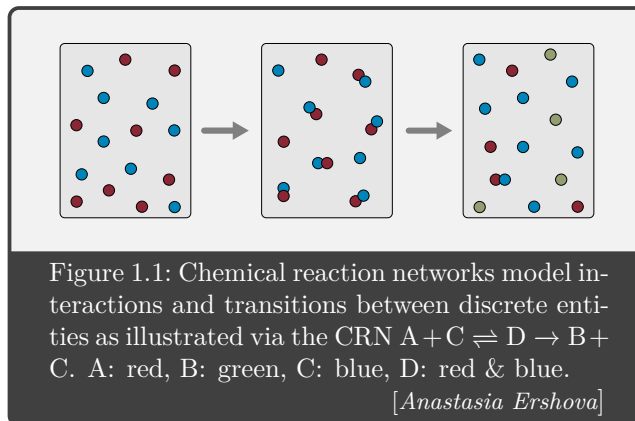
—William Poole

1.1 Chemical Reaction Networks Model Interactions between Discrete Entities

Scientists and engineers use mathematical models in order to quantitatively describe and reason about the behavior of complex systems involving many interacting parts. Indeed, modeling is in many ways its own discipline with a variety of principles, nuances, and methods. To the molecular programmer, modeling provides a way to translate understanding and hypotheses about how a system functions into mathematical expressions which can then be compared to experiments, analyzed theoretically, or optimized computationally.

One fundamental model in molecular programming is a **chemical reaction network** (CRN). CRNs model the interactions between a finite set of species which interact via a finite set of reactions converting one multiset of species to another. Historically, as the name CRN suggests, these systems were invented to study chemical reactions where species are molecules made up of atoms. For example, the burning of hydrogen can be written as

$$2\,\mathrm{H_2} + \mathrm{O_2} \to 2\,\mathrm{H_2O}.$$
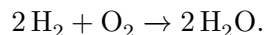
Here, two hydrogen gas molecules come together with one oxygen molecule in order to produce two water molecules. This example exposes the implicit assumptions of the CRN model. Species in the model are assumed to be **well-mixed**, meaning the hydrogen and oxygen gasses are evenly and constantly stirred together. Spatial structures such as flames, heat transfer, and diffusion are neglected. Similarly, internal structure of each species, such as the internal energy of each hydrogen and oxygen atom, is assumed to be homogeneous.

Figure 1.1: Chemical reaction networks model interactions and transitions between discrete entities as illustrated via the CRN $A + C \rightleftharpoons D \rightarrow B + C$. A: red, B: green, C: blue, D: red & blue.
[*Anastasia Ershova*]

More recently, scientists have used CRNs to model many systems beyond well-mixed atoms and molecules. Species can also be abstract entities where the specific atomic structure is obscured, for example rabbits:

$$\text{Black Rabbit} + \text{White Rabbit} \rightarrow \text{Black Rabbit} + \text{White Rabbit} + 5\,\text{Spotted Rabbits}.$$

Importantly, CRNs typically involve multiple coupled reactions. For example, the activity of an enzyme C binding to a substrate A to form a complex D and then catalytically converting A to B illustrated in figure 1.1. These interactions can be represented via three reactions:

$$A + C \xrightarrow{k_b} D, \qquad D \xrightarrow{k_u} A + C, \qquad D \xrightarrow{k_{\text{cat.}}} B + C.$$
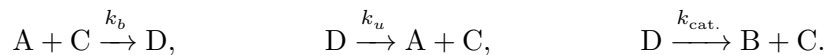
These three reactions can be written more succinctly by noticing that the first two reactions are reverses of each other:

$$A + C \underset{k_u}{\overset{k_b}{\rightleftharpoons}} D \xrightarrow{k_{\text{cat.}}} B + C.$$
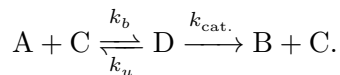
Here, we have added rates, k_u, k_b, and $k_{\text{cat.}}$ to the reactions to denote how fast they occur. Formally, a CRN is written as a set of species $\{S_i\}$ each with amount s_i and a set of reactions of the form:

$$\sum_i I_{i,r}\, \mathrm{S}_i \xrightarrow{\rho_r(s)} \sum_i O_{i,r}\, \mathrm{S}_i . \tag{1.1}$$

where $I_{i,r}$ and $O_{i,r}$ are non-negative integers describing the number of species S_i that are inputs and outputs to the reaction r, respectively. The function $\rho_r(s) \geq 0$ is the rate at which each reaction occurs, also known as a **propensity**. Here s is a vector of the amounts of all the species and is also referred to as the **state** of the system. Sometimes,

it is convenient to consider the net change of species that occur in each reaction as a matrix, called the **stoichiometric matrix**:

$$\Delta = O - I \tag{1.2}$$

Finally, in this text, when only a rate constant k_r is given for a reaction, the propensity is by convention assumed to be **mass action**:

$$\rho_r^{\text{det}}(s) = k_r \prod_i s_i^{I_{i,r}} \qquad\qquad \rho_r^{\text{stoch}}(s) = \frac{k_r}{V^{|I_r|-1}} \prod_i \frac{s_i!}{(s_i - I_r^i)!}. \tag{1.3}$$

Here, $\prod$ is the product operator, V is the volume of the CRN. The two different forms of this function depend on whether CRNs are simulated stochastically or deterministically, which will be elaborated on the proceeding sections. Mass action propensities are rooted in the physics of ideal gasses and ideal solutions. For example, consider the stochastic propensity for the binding of A to B:

$$\text{A} + \text{C} \xrightarrow{\text{k}_b} \text{D} \qquad\qquad \rho^{\text{stoch}}(s) = \frac{k_b}{V} ab.$$

Intuitively, the stochastic propensity means that the probability of a molecule A colliding with a molecule B is proportional to the number of molecules a times the number of molecules b divided by the total volume V that any two molecules need to search, on average, in order to bump into each other. Then, once the molecules collide, they bind together at some rate k_b. In statistical physics, this is analogous to the collisions that occur between particles in an ideal gas []. However, most CRNs in this book model chemicals dissolved in aqueous solution. An ideal solution is extension of the physics of ideal gasses to include a non-reactive solvent which keeps the total volume and pressure of the system constant such that the non-solvent species obey the same physics as an ideal gas []. The deterministic mass action propensity is derived as an approximation of the stochastic propensity in very large volumes with very large numbers of molecules []. Other propensities may also used to model more complex phenomenological behavior; in general, these can be thought of as simplifications or reductions of more complex mass action systems []. This idea is explored further in Problem 1.4.

Process	CRN	Description
One Step Catalysis	$\text{A} + \text{C} \rightarrow \text{B} + \text{C}$	C converts A to B in one step.
Molecular Binding	$\text{A} + \text{B} \rightleftharpoons \text{C}$	A and B reversible form a complex C.
Michaelis-Menten Catalysis	$\text{A} + \text{C} \rightleftharpoons \text{D} \rightarrow \text{B}$	C binds with A before converting it to B.
Conformation Changes	$\text{S}_1 \rightleftharpoons \text{S}_2 \rightleftharpoons \text{S}_3$	S reversibly changes between three conformations.
Diffusion Between Compartments	$\text{X}_{\text{in}} \rightleftharpoons \text{X}_{\text{out}}$	X diffuses between the inside an outside of a compartment.
Amplification	$\text{S} + \text{A} \rightarrow 2\,\text{S} + \text{A}$	A exponentially amplifies S.

Table 1.1: CRN motifs used in many processes common to molecular programming. These reactions are frequently combined together to represent diverse molecular processes. Note, however, that these motifs are just a small subset of the possible reactions molecular programmers want to consider.

By combining abstract species with highly customizable interactions, CRNs embody a general framework that has been studied in diverse fields ranging from physics [] to chemistry [] to biology [] to ecology [] to computer science [] to mathematics []. Molecular programmers use CRNs because they succinctly represent interactions between elements of a system at various levels of abstraction: atoms, molecular structures, cells, or even entire organisms can be modeled using the same formalism. Table 1.1 shows some common CRN motifs that represent processes relevant to molecular programming. In particular, note that the well-mixed assumption still allows for discrete conformations and limited spatial structure. Indeed, CRNs excel at modeling finite systems where all the species and reactions can be easily enumerated. This makes CRNs numerically tractable meaning that these models to be easily simulated on a computer. However, the finite nature of CRNs makes them less suited for modeling continuous systems such as the 3D arrangement of atoms including molecular structures and continuous motion in space.

In the following sections, we will discuss two different ways simulate CRNs: stochastic and deterministic semantics. These sections assume some familiarity with ordinary differential equations, which are briefly explained in box 1.1

Pop Quiz 1.1: Check your understanding!

Identify whether each of the following situations can be modeled using a CRN. Why or why not?

1. Cars driving on a freeway.

2. The entry and replication of a virus in a cell.

3. People moving between different rooms at a party.

4. Cooking a scrambled egg.

5. A seed growing into a flower.

6. All of the ways to put 10 lego blocks together.

Solution on page 15

Box 1.1: Ordinary Differential Equations Model a System's Dynamics

Molecular programs often consist of systems that change over time. Such programs may respond to user inputs, constantly adapt to environmental cues, and/or carry out a series of steps in order to produce an output. The way a system changes over time is called the **dynamics** of the system. In order to successfully understand

Box 1.1: Ordinary Differential Equations Model a System's Dynamics (cont.)

and program the dynamics of molecular systems, molecular programmers make use of a variety of mathematical models and techniques. At the heart of many of these techniques are differential equations,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, t).$$

This equation is incredibly general and describes how the state of the system, described by a continuous variable x, changes in time, t, according to a bounded function f. When x is a scalar or a finite vector, this equation is called an **ordinary differential equation** (ODE). Integrating the right hand side from some initial condition $x(t_0)$ at t_0 to a final time t produces a solution $x(t; x(t_0))$ to the ODE which describes how x changes in time given initial condition $x(t_0)$:
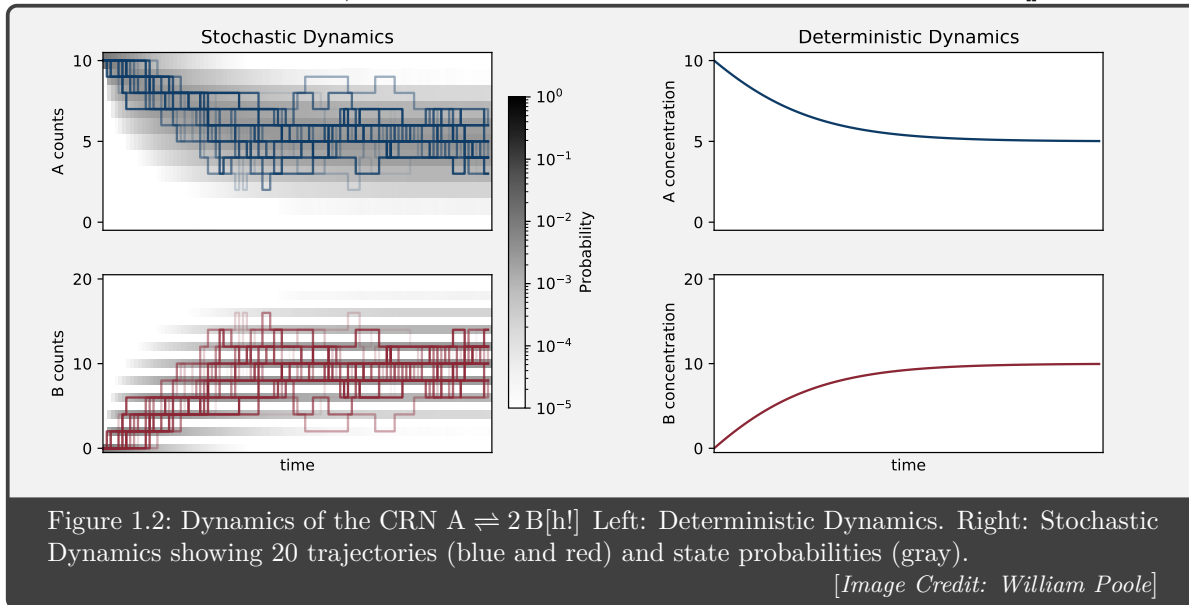
$$x(t; x(t_0)) = \int_{t_0}^{t} f(x, t')\, \mathrm{d}t'\,.$$

Typically, this equation is solved numerically using a computer although in some special cases it can be solved analytically. Choosing the state x and function f in order to understand and program molecular behavior will be a recurring theme in this textbook. In general, allowing f to be *any* function is unnecessary and instead we restrict ourselves to specific models. In some cases, f will be rooted in physical descriptions of how molecules behave. In other cases, f will describe the phenomenological or idealized behavior we wish to program into a system.

1.2 Stochastic Chemical Reaction Networks

The stochastic description of CRNs is deeply rooted in the physics of interacting molecules. **Stochastic** is a word phycists and mathematicians use to refer to mathematically quantifiable randomness. Molecules are discrete, finite, and countable, meaning that the amount of any molecule will always be a non-negative integer. You cannot have a quarter of a molecule, at least not without breaking the molecule into some different kind of molecule! Stochastic CRNs describe how the **counts**, s_i, of each molecule S_i changes in time. As these molecules interact by crashing into each other and participating in reactions, discrete sets of molecules are interconverted. If the exact positions and velocities of every individual molecule were known, the order of collisions would be predetermined. However, the well-mixed assumption of CRNs implicitly neglects the detailed positions of individual molecules. Instead, molecular collisions are modeled as random events be-

tween uniformly distributed species in a volume V which occur at random intervals and cause reactions to occur proportionally to that reaction's propensity []. Due to the fact that the waiting time between reactions as well as which reaction occurs are stochastic, this process is modeled using what mathematicians call a time Markov chain (also called a Markov jump process) which will be illustrated intuitively via an example [].

Figure 1.2: Dynamics of the CRN $A \rightleftharpoons 2B$[h!] Left: Deterministic Dynamics. Right: Stochastic Dynamics showing 20 trajectories (blue and red) and state probabilities (gray).

[Image Credit: William Poole]

Consider dynamics of a CRN which reversibly converts one molecule A into two molecules of B:

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 2B \tag{1.4}$$

Denote the state of the system $s = (a, b)$. The forward reaction can written in terms of the state of the system $s \rightarrow s + (-1, 2)$ meaning one molecule of A is consumed and two molecules of B are created. Similarly, the reverse can be written as $s \rightarrow s + (1, -2)$. If the system starts in any state s, after some random time, τ, one of these reactions will occur. The random time $\tau \sim \text{Exponential}(1/\sum_r \rho_r(s))$, meaning τ is sampled from an exponential distribution with mean of one divided by the sum of the propensities. Intuitively, this scaling makes sense; the larger the reaction propensities, the less time there is between reactions. Similarly, the probability of any given reaction occurring within an arbitrary window of time is proportional to its propensity, $\mathbb{P}(\text{reaction } i \text{ fires}) = \frac{\rho_i(s)}{\sum_j \rho_j(s)}$. After a reaction occurs, the system will be in a new state s' and the process will repeat. This results in each state having a probability $\mathbb{P}(s, t)$ which in general depends on the initial condition $s(0)$ and changes in time as reactions occur. Figure 1.2 (left) illustrate stochastic trajectories of this CRN and the dynamics of the probability vector

in time. Box 1.2 generalizes this description of the dynamics of a stochastic CRN into an algorithm called the Gillespie stochastic simulation algorithm (SSA).

An alternative method for understanding the dynamics of a stochastic CRN is to write out an ordinary differential equation for the probability vector of each state: $\mathbb{P}(s,t)$. This equation is called the chemical master equation (CME):

$$\frac{\mathrm{d}\,\mathbb{P}(s,t)}{\mathrm{d}t} = \sum_r [\mathbb{P}(s-\Delta_r,t)\rho_r(s-\Delta_r) - \mathbb{P}(s)\rho_r(s)]. \tag{CME}$$

This equation, although frequently long and complicated, is easy to understand. For each state s and each reaction r there is incoming probability from the state $s - \Delta_r$ at a rate $\rho_r(s-\Delta_r)$ and outgoing probability from s at rate $\rho_r(s)$. For example, the CME for the CRN (1.4) is given by:

$$\frac{\mathrm{d}\,\mathbb{P}(s,t)}{\mathrm{d}t} = \mathbb{P}(s+(1,-2),t)\rho_1(s+(1,-2)) + \mathbb{P}(s+(-1,2),t)\rho_2(s+(-1,2)) - \mathbb{P}(s,t)(\rho_1(s)+\rho_2(s))$$
$$= \mathbb{P}((a+1,b-2),t)k_1(a+1) + \mathbb{P}((a-1,b+2),t)k_2(b+2)(b+1) - \mathbb{P}((a,b),t)(k_1a+k_2b(b-1)).$$

Here, the first two terms represent incoming probability into the state $s = (a,b)$ and the third term represent outgoing probability from the state s. Notice that this ODE has a probability for every possible value of a and b. Indeed, you should convince yourself that the total number of states is in fact $|\Gamma| = a(0) + \lfloor \frac{b(0)}{2} \rfloor$. In general, the number of states, and hence the number of probabilities in the chemical master equation, may grow exponentially in the number of species or even be infinite. Very large or infinite state spaces make the CME description of a stochastic CRN difficult to integrate numerically, so for many applications stochastic CRNs are simulated many times using SSA and probability distributions are empirically approximated by this ensemble of trajectories.

Box 1.2: Gillespie Stochastic Simulation Algorithm (SSA) Pseudocode

$s \leftarrow$ initial state
$t \leftarrow$ initial time
while $t \leq t_{\max}$ **do**
    $\rho_r \leftarrow \rho_r(s)$ *compute each propensity function*
    $\lambda \leftarrow \sum_r \rho_r$ *compute total reaction rate*
    $\tau \sim \mathrm{Exponential}(1/\lambda)$ *sample waiting time*
    $r \sim \mathrm{Categorical}(\rho_r/\lambda)$ *sample a reaction, $\mathbb{P}(r) \propto \rho_r$*
    $s \leftarrow s + \Delta_r$ *update the state*
    $t \leftarrow t + \tau$ *advance the time*
end while

- -

Notes: Sampling an exponential distribution with parameter $\frac{1}{\lambda}$ can be done by sampling a uniformly distributed random number, $u \in [0,1]$, and computing $-\frac{1}{\lambda}\log u$. Sampling a categorical distribution can be done by picking a uni-

Pop Quiz 1.2: Check your understanding!

Write down all the states that can be reached via any sequence of reactions by the following CRNs from the given initial condition $s(0)$. This set is called the **reachability class** and is denoted $\Gamma_{s(0)}$. For example, in the CRN $\varnothing \rightleftharpoons X$ with $x(0) = 10$ the reachability class is $\Gamma = \{0, 1, 2...\} = \mathbb{Z}_{\geq 0}$ (all the non-negative integers).

1. $A \rightleftharpoons B$ with $a(0) = 10$ and $b(0) = 0$

2. $X \rightleftharpoons 2Y$, $Y \rightleftharpoons 2Z$ with $x(0) = 100$ and $y(0) = z(0) = 0$.

3. $2A \rightarrow 2B$, $2B \rightarrow 2C$, $2C \rightarrow 2A$, with $a(0) = b(0) = c(0) = 1$.

1.3 Deterministic Chemical Reaction Networks

In typical laboratory settings, it can be difficult to measure or control the precise counts of molecules. Instead, scientists often use concentrations: the number of molecules divided by the volume in which the reaction occurs. Unlike stochastic CRNs, this dynamic description of concentrations includes no randomness in the values of s, hence being called deterministic. Mathematically, deterministic CRNs are an ordinary differential equation model for understanding how the concentrations of chemical species vary in time due to reactions:

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = \sum_r \Delta_{i,r} \rho_r(s)$$

Here, $s_i = \frac{n_i}{V}$ is the concentration of species S_i which has n_i molecules in a volume V. Physically, this ODE approximates the behavior of a stochastic CRN in the limit of infinite volume, $V \to \infty$, and infinite count, $n \to \infty$, allowing species to be treated as continuous concentrations instead of discrete counts []. Similar to the reachability class of a stochastic CRN, deterministic CRNs have a **stoichiometric compatibility class** $\gamma_{s(0)} = s(0) + \mathrm{span}(\Delta)$. This space includes all the possible concentrations of the trajectory $s(t, s(0))$ for any positive or negative value of t.

Although an physical approximation, deterministic CRNs are more mathematically tractable than stochastic CRNs. Consider a CRN with N species. The deterministic dynamics will be governed by an ODE with N variables, one representing each species. On the other hand, the stochastic dynamics will be governed by $|\Gamma_{s(0)}|$-dimensional probability vector, where $\Gamma_{s(0)}$ is the set of all reachable states and may even be infinite in size. For example, the deterministic dynamics of the CRN (1.4) are given by just two equations:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = k_2 b^2 - k_1 a \qquad\qquad \frac{\mathrm{d}b}{\mathrm{d}t} = k_1 a - k_2 b^2.$$

Notice that this ODE is considerably simpler than the chemical master equation for the same CRN. This illustrates an important concept in molecular programming: compilation. A CRN can be viewed as an abstract specification, or programming language, which is compiled into a mathematical representation. The same set of species and reactions can be compiled into either a stochastic representation or a deterministic representation. Many of the later sections in this book will build on this general idea of molecular compilation such as examining what kinds of programs can be compiled into CRNs and compiling a specification CRN into more detailed implementation CRNs.

Box 1.3: When to use stochastic versus deterministic CRNs

Determining whether the stochastic or deterministic description of a CRN is the appropriate model for a given problem can be challenging. As a rule of thumb, if the absolute number of molecules in the system is large, the deterministic semantics is appropriate. For example, in a laboratory it is difficult to accurately pipette less than $1\,\mu\mathrm{L}$ ($10^{-6}\,\mathrm{L}$) by hand; $1\,\mu\mathrm{L}$ of $1\,\mu\mathrm{M}$ ($10^{-6}\,\mathrm{M}$) solution is 6.022×10^{11} molecules—so for most hand-pipetted laboratory experiments, deterministic models are probably more appropriate. However, using state of the art techniques, scientists are increasingly able to work with incredibly small volumes and hence very low molecular counts. The volume of a single eukaryotic cell is about $1\,\mathrm{pL}$ ($10^{-12}\,\mathrm{L}$), and of a bacterium is about $1\,\mathrm{fL}$ ($10^{-15}\,\mathrm{L}$). Volumes this small can also be created artificially using lithography **??** or droplets **??**. When working with small molecule counts, noise may become important to the functionality of the system in which case a stochastic model is appropriate. On the other hand, a deterministic model still may capture the phenomenological behavior and is typically much easier to work with analytically and numerically. Finally it is worth noting that there exist exceptions to these rules of thumb, rooted in the exceptional computational power of CRNs **??**.

Pop Quiz 1.3: Check your understanding!

For each of the following cases, describe whether stochastic or deterministic CRNs would be more appropriate.

a. Modeling how single piece of DNA in a vesicle interacts with proteins.

b. Modeling how synthetic single stranded RNA interacts with itself (such as folding into secondary structures).

c. Modeling the dissociation of table salt in water.

d. Modeling the spread of infections between people in different countries during a pandemic

Box 1.4: CRN Mathematical Definitions:

Definition 1.1: Chemical Reaction Network (CRN)

A CRN consists a finite set of species, reactions, and propensities each of which is defined below. CRNs can be simulated using both stochastic and deterministic dynamics.

Definition 1.2: Species

The abstract chemical (or otherwise) elements which interact in the CRN. Denoted by an uppercase letter such as S_i, X, Y, etc. Subscripts or different symbols may be used to denote different species. The amount of a species (count or concentration) will be denoted by the lower case letter e.g. s_i, x, y, etc.

Definition 1.3: Reaction

Species inter-convert via reactions of the form
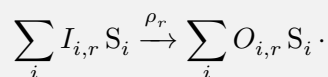
$$\sum_i I_{i,r} \, S_i \xrightarrow{\rho_r} \sum_i O_{i,r} \, S_i \,.$$

$I_{i,r}$ and $O_{i,r}$ are non-negative integers describing the number of species S_i that are inputs and outputs to the reaction r, respectively.

Box 1.4: CRN Mathematical Definitions: (cont.)

Definition 1.4: Propensity

The rate at which a reaction occurs. In general, propensities reactions can be any function $\rho_r(s) \geq 0$.

Definition 1.5: Mass action

The most commonly used propensity in CRN models derived from the behavior of ideal gasses. This propensity has different forms for stochastic and deterministic semantics:

$$
\rho_r^{\text{det}}(s) = k_r \prod_i s_i^{I_{i,r}} \qquad \rho_r^{\text{stoch}}(s) = \frac{k_r}{V^{|I_r|-1}} \prod_i \frac{s_i!}{(s_i - I_r^i)!}. \qquad (1.5)
$$

Here, $\prod$ is the product operator, V is the volume of the CRN, and $|I_r| = \sum_i I_{i,r}$ denotes the total number of inputs into the reaction r. In this text, if only a rate constant k_r is given, the propensity is assumed to be mass action.

Definition 1.6: Stoichiometric Matrix

Commonly denoted $\Delta = O - I$, the element $\Delta_{i,r}$ denotes how many of species i are produced or consumed in reaction r.

Definition 1.7: Chemical Master Equation:

Determines the the time evolution of the probability, $\mathbb{P}(s,t)$, that the species of a CRN have the counts s at time t:

$$
\frac{\mathrm{d}\,\mathbb{P}(s,t)}{\mathrm{d}t} = \sum_r [\rho_r(s - \Delta_r)\,\mathbb{P}(s - \Delta_r, t) - \rho_r(s)\,\mathbb{P}(s,t)]. \qquad (1.6)
$$

The first term is the probability flowing into the state with counts s and the second term is the is the probability flowing out of state with counts s.

Box 1.4: CRN Mathematical Definitions: (cont.)

Definition 1.8: Steady State Distribution

A time-independent solution to stochastic CRN dynamics $\mathbb{P}_{\text{ss}}(s)$. This is found by setting the chemical master equation $\frac{\mathrm{d}\,\mathbb{P}(s,t)}{\mathrm{d}t} = 0$ and can be thought of as the limiting behavior of the system when the dynamics are averaged over infinite time.

Definition 1.9: Reachability Class

The set of discrete states (counts) $\Gamma_{s(0)}$ reachable by a stochastic CRN via any sequence of reactions starting at initial condition $s(t = 0)$. Note that the reachability class is always a subset of the positive integer lattice, $\Gamma_{s(0)} \subseteq \mathbb{Z}_{\geq 0}^{N}$ where N is the number of species.

Definition 1.10: Deterministic Dynamics

Are given by an ordinary differential equation governing the concentration s of the species:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \sum_{r} \Delta_r \rho_r(s).$$

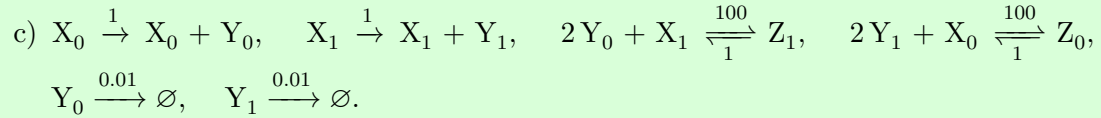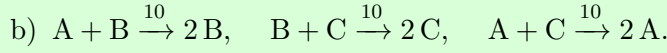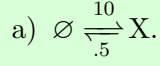### Definition 1.11: Fixed Points

Points s^* where the concentrations no longer change in time, $\frac{\mathrm{d}s^*}{\mathrm{d}t} = 0$. Also called stationary points.

Definition 1.12: Stoichiometric Compatability Class

The set of continuous states (concentrations), $\gamma_{s(0)}$ spanned by the forward and reverse dynamics of a deterministic CRN given an initial condition $s(t = 0)$. Mathematically, $\gamma_{s(0)}$ can be computed by calculating the kernel of the stoichiometric matrix and restricting it to the positive real numbers: $\gamma_{s(0)} \subseteq \mathbb{R}_{\geq 0}^{N}$.

1.4 Problems

1.1[quantitative]   Implement the Gillespie SSA algorithm from Box 1.2 in your language of choice and simulate the following CRNs (for simplicity you may assume the volume $V = 1$). Describe the behavior of each model as you change the initial conditions.

a) $\varnothing \underset{.5}{\overset{10}{\rightleftharpoons}} X$.

b) $A + B \xrightarrow{10} 2\,B, \quad B + C \xrightarrow{10} 2\,C, \quad A + C \xrightarrow{10} 2\,A$.

c) $X_0 \xrightarrow{1} X_0 + Y_0, \quad X_1 \xrightarrow{1} X_1 + Y_1, \quad 2\,Y_0 + X_1 \underset{1}{\overset{100}{\rightleftharpoons}} Z_1, \quad 2\,Y_1 + X_0 \underset{1}{\overset{100}{\rightleftharpoons}} Z_0,$

$\quad Y_0 \xrightarrow{0.01} \varnothing, \quad Y_1 \xrightarrow{0.01} \varnothing$.

1.2[quantitative]   Simulate the CRNs from the previous question deterministic using any ODE integrator (or by writing a simple one yourself). Describe the behavior of each model as you change the initial conditions.

1.3[quantitative]   The **fixed point** of a deterministic CRN is a point s^* such that $\frac{\mathrm{d}s}{\mathrm{d}t}\big|_{s^*} = 0$. Similarly, the **steady state distribution** of a stochastic CRN is a probability vector $\mathbb{P}_{ss}(s)$ such that $\frac{\mathrm{d}\,\mathbb{P}(s)}{\mathrm{d}t}\big|_{ss} = 0$. Consider the following CRN,

$$\varnothing \underset{k_1}{\overset{k_2}{\rightleftharpoons}} X.$$
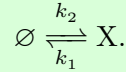
a. show that $s = \frac{k_1}{k_2}$ is a fixed point for the deterministic CRN.

b. show that $\mathbb{P}(s) = \frac{1}{s!}\exp\left(-s\log\frac{k_1}{k_2}\right)$ is the steady state distribution for the stochastic CRN.
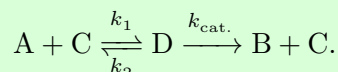
1.4[quantitative]   Hill Functions are a common kind of propensity of the forms

$$\rho_{\mathrm{hill+}} = k\frac{S^n}{K_D^n + S^n}, \text{ a positive hill function, and}$$

$$\rho_{\mathrm{hill-}} = k\frac{1}{K_D^n + S^n}, \text{ a negative hill function.}$$

These functions are used to model saturating behaviors and can be thought of as simpli-

fications to more complex mass action systems. Consider the deterministic CRN

$$A + C \underset{k_2}{\overset{k_1}{\rightleftharpoons}} D \xrightarrow{k_{\text{cat.}}} B + C.$$

a. Assume $k_1, k_2 \gg k_{\text{cat.}}$. This means that the first two reactions can be approximated as being at **quasi-equilibrium** relative to the third reaction. To make this approximation, assume $k_{\text{cat.}} = 0$ and solve for the steady state concentration of D.

b. The third reaction can then be reintroduced individually as a hill function $A \xrightarrow{\rho(s)} B$. Using the result from (a), determine the function $\rho(s)$.

c. Assume that $\frac{dD}{dt} = 0$. This is called the quasi-steady-state approximation because the concentration of the bound complex D is assumed to be in steady state. Furthermore, assume and $D \ll A + B$. Show that under these assumptions the entire CRN can be modeled as a single reaction $A \xrightarrow{\rho(s)} B$ with a positive hill function different than the one in part (b).

d. Simulate the full CRN as well as the reduced hill function CRNs for different parameters of k_1, k_2 and $k_{\text{cat.}}$ and initial conditions $D(0)$, and $A(0)$. When do each of these approximations work? [Hint: try varying the order of magnitude of different parameters independently].

1.5[advanced]   Find a CRN where the mean of the stochastic steady state distribution is not equal to the fixed point of the deterministic CRN.

1.5 Further Reading

[1] *(Leonard M Adleman. "Molecular computation of solutions to combinatorial problems". In:* Science *266.5187 (1994), pp. 1021–1024)*

Read this to find out more about the origins of dna computing.

[2] *(Erik Winfree et al. "Design and self-assembly of two-dimensional DNA crystals". In:* Nature *394.6693 (1998), pp. 539–544)*

Read this to find out more about tile assembly.

1.6 Solutions

Pop Quiz 1.1: Solution(s)

1. No, because cars move in continuous space. However, CRNs could model how quickly cars go from one city to another.

2. Yes, because a virus consists of discrete molecules and the entry to a cell and replication can be thought of as a series of discrete binding, transcription, translation, and self-assembly reactions.

3. Yes, because there are a finite number of people in a finite number of rooms.

4. No, this process is not well mixed and involves phase transition as the egg goes from liquid to solid.

5. No, the spatial structure of a flower exists in continuous space and is subject to mechanical forces such as pressure.

6. Yes, although there may be many possible structures made from 10 lego bricks, this is still a finite number and some structures can be built from other structures.

Pop Quiz 1.2: Solution(s)

1. $a \in \{0, 1, ..., 10\}$ with $b = 10 - a$.

2. Any non-negative integer values (x, y, z) such that $x + 2y + 4z = 100$.

3. No reactions can fire with fewer than two inputs of a given species so $\Gamma = \{(1, 1, 1)\}$

Pop Quiz 1.3: Solution(s)

a. Stochastic is probably appropriate here because there is a single piece of DNA in a small volume.

b. Both stochastic and deterministic models could be used in this case because the RNA only interacts with itself.

> **Pop Quiz 1.3: Solution(s) (cont.)**
>
> c. A deterministic model is best here because there are many salt molecules and ions.
>
> d. Both stochastic and determinsitic models could be used. In the stochastic case, the number of of infected individuals in each country would be modeled. In the deterministic case, the percentage of infected people would be modeled.
>

1.7 Bibliography

[1] Leonard M Adleman. "Molecular computation of solutions to combinatorial problems". In: *Science* 266.5187 (1994), pp. 1021–1024.

[2] Erik Winfree et al. "Design and self-assembly of two-dimensional DNA crystals". In: *Nature* 394.6693 (1998), pp. 539–544.