# Robust Signal Restoration
# in Chemical Reaction Networks [*]

Titus H. Klinge
Iowa State University, Ames, IA 50010, USA
tklinge@iastate.edu

## ABSTRACT

Molecular computing systems that are contained in well-mixed volumes are often modeled using chemical reaction networks. In these systems, concentrations of molecules are treated as signals and used for both communication and memory storage. A common design challenge for such a system is to avoid memory corruption caused by noise in the input signals. In this paper, we analyze two signal restoration algorithms for molecular systems modeled with chemical reaction networks. These algorithms are designed to prevent a memory signal from degrading over time, and we show that under modest conditions these algorithms will maintain the memory indefinitely. We also present an exact solution of the running time of the first algorithm which demonstrates that it converges in logarithmic time.

## CCS Concepts

•**Theory of computation** → **Models of computation;**
•**Mathematics of computing** → *Differential equations;*

## Keywords

Molecular programming, chemical reaction networks, signal restoration, robustness

## 1. INTRODUCTION

Molecular programming is a relatively new computing paradigm aiming to manipulate the function and structure of matter at the molecular level. By utilizing properties of bio-molecules, researchers have developed computing systems that naturally interface with biological systems. One such system is a programmable nanodevice capable of precisely delivering cargo into specific cells [6]. Although individual nanodevices such as these are limited to relatively simple computations, these systems can achieve more complex behavior by networking many nanodevices together.

Chemical reaction networks (CRNs) have recently become a popular programming language for developing molecular computing systems. A CRN is a mathematical model of how well-mixed populations of molecules interact and have been used for over a half century. CRNs also have multiple semantics (i.e. operational meanings) to model particular environments more accurately. The two most common semantics are stochastic mass action and deterministic mass action, and CRNs under such semantics are commonly referred to as stochastic CRNs and deterministic CRNs, respectively. Stochastic CRNs model molecular systems in small volumes (e.g. a cell) by discretely modeling the number of molecules in the system. In contrast, deterministic CRNs are suited for relatively larger volumes and model the number of molecules as continuous concentrations.

CRNs are a popular molecular programming language for two reasons: (1) they are able to compute any algorithm; and (2) they can be easily implemented with DNA molecules. In 2008, Soloveichik, Cook, Winfree, and Bruck showed that stochastic CRNs are Turing universal [12], and in 2010, Soloveichik, Seelig, and Winfree showed that CRNs can be compiled into DNA strands that simulate them [13]. These results demonstrate that even completely disorganized molecules can collaborate via simple reactions to generate complex behavior. They also make CRNs a promising design tool for developing molecular computing systems.

When designing a molecular system, it is crucial to consider its robustness. Since the CRN model only approximates reality, CRNs that depend on precise initial conditions and system parameters are not achievable in practice. Furthermore, many molecular programming applications are safety critical and must be reliable even in adverse conditions. Consequently, techniques for developing robust molecular programs and proving their correctness has become an important research subject [10, 4, 11, 7].

One recurring design challenge is creating systems that are robust with respect to communication signals. CRNs use concentrations of molecules both for communication and memory storage, and any noise present in communication signals can cause memory to degrade over time. A bit of memory is usually implemented in a CRN by contrasting high and low concentrations of a particular molecule. Maintaining the separation between "high" and "low" is essential to prevent data loss.

Signal restoration is a memory refresh technique for deterministic CRNs which utilizes bistability to prevent memory from deteriorating over time. In this paper, we give a thorough analysis of two signal restoration algorithms and show

that under modest conditions they can maintain memory indefinitely even in the presence of adversarial reactions due to noise. The first algorithm consists of two termolecular reactions (reactions with three reactants, i.e., input molecules) and is the simpler of the two algorithms [10, 9]. The simplicity of this algorithm makes it easy to use and relatively easy to analyze. We also give an exact analytical solution for the runtime of this algorithm that demonstrates that the algorithm converges in logarithmic time.

The second signal restoration algorithm consists of four bimolecular reactions and was developed in 2008 by Angluin, Aspnes, and Eisenstat for population protocols. (Population protocols were designed to model networks of finite state sensors and are equivalent to stochastic CRNs [5, 1].) This second algorithm is significantly harder to analyze, but it has the advantage that bimolecular reactions are easier to physically implement.

The CRN model is reviewed in section 2, and our results are presented in sections 3 and 4. We conclude the paper with a brief discussion of the results in section 5.

## 2. CHEMICAL REACTION NETWORKS

Let $S$ be a finite set whose elements we call *species*. A *reaction* over the set $S$ is a triple

$$\rho = (\mathbf{r}, \mathbf{p}, k) \in \mathbb{N}^S \times \mathbb{N}^S \times (0, \infty),$$

where $X^Y$ denotes the set of all functions mapping $Y$ into $X$. We regard the elements of $\mathbb{N}^S$ as finite vectors of natural numbers indexed by the elements of $S$. We call $\mathbf{r}$ the *reactant vector*, $\mathbf{p}$ the *product vector*, and $k$ the *rate constant* of the reaction $\rho = (\mathbf{r}, \mathbf{p}, k)$. To avoid excessive use of subscripts, we use the notation $\mathbf{r}(\rho) = \mathbf{r}$, $\mathbf{p}(\rho) = \mathbf{p}$, and $k(\rho) = k$ to access the individual components of a reaction $\rho = (\mathbf{r}, \mathbf{p}, k)$.

We also borrow notation from chemistry to specify reactions. If $S = \{X, Y, Z\}$, the notation

$$X + Y \xrightarrow{k} X + 2Z \tag{1}$$
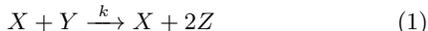
specifies the reaction $\rho = (\mathbf{r}, \mathbf{p}, k)$ where $\mathbf{r}$ is the vector defined by $\mathbf{r}(X) = 1$, $\mathbf{r}(Y) = 1$, $\mathbf{r}(Z) = 0$, and $\mathbf{p}$ is the vector defined by $\mathbf{p}(X) = 1$, $\mathbf{p}(Y) = 0$, and $\mathbf{p}(Z) = 2$.

The *net effect* of a reaction $\rho = (\mathbf{r}, \mathbf{p}, k)$ is the vector $\Delta\rho = \mathbf{p} - \mathbf{r} \in \mathbb{Z}^S$. A species $Y \in S$ is called a *catalyst* of a reaction $\rho = (\mathbf{r}, \mathbf{p}, k)$ if $\mathbf{r}(Y) > 0$ and $\Delta\rho(Y) = 0$. For example, the species $X$ from reaction (1) is a catalyst.

A *chemical reaction network* is an ordered pair $N = (S, R)$ where $S$ is a finite set of species and $R$ is a finite set of reactions over $S$. This completes the specification of the *syntax* of a CRN. The remainder of this section specifies the *deterministic mass action semantics* of CRNs.

Under deterministic mass action semantics, a *state* of a CRN $N = (S, R)$ is a vector $\mathbf{x} \in [0, \infty)^S$. For a state $\mathbf{x} \in [0, \infty)^S$ and species $Y \in S$, we call $\mathbf{x}(Y)$ the *concentration* of $Y$ in the state $\mathbf{x}$. If a state is obvious from context, we denote the concentrations of species with their corresponding lowercase letters. For example, we use $x$, $y$, and $z$ to denote the concentrations of the species $X$, $Y$, and $Z$ in a state, respectively.

For a state $\mathbf{x} \in [0, \infty)^S$ and a reaction $\rho = (\mathbf{r}, \mathbf{p}, k)$, we define the *rate of $\rho$ in $\mathbf{x}$* to be

$$\text{rate}_{\mathbf{x}}(\rho) = k(\rho) \prod_{Y \in S} \mathbf{x}(Y)^{\mathbf{r}(\rho)(Y)}. \tag{2}$$

For example, the rate of reaction (1) in the state $\mathbf{x} \in [0, \infty)^S$ is $k \cdot \mathbf{x}(X) \cdot \mathbf{x}(Y)$. Intuitively, the rate of the reaction is proportional to the frequency of $X$ molecules colliding with $Y$ molecules, and the constant $k$ encapsulates features of the reaction that do not depend on the state (e.g. salinity, temperature, etc.).

The *deterministic mass action function* of a species $Y \in S$ is the function $F_Y : [0, \infty)^S \to \mathbb{R}$ defined by

$$F_Y(\mathbf{x}) = \sum_{\rho \in R} \Delta(\rho)(Y)\text{rate}_{\mathbf{x}}(\rho) \tag{3}$$

for all $\mathbf{x} \in [0, \infty)^S$. The real value $F_Y(\mathbf{x})$ is the total rate of change of the concentration of $Y$ in the state $\mathbf{x}$. Therefore, if the CRN $N = (S, R)$ is in state $\mathbf{x}(t)$ at time $t$, the species $Y \in S$ must obey the ordinary differential equation (ODE)

$$y'(t) = F_Y(\mathbf{x}(t)). \tag{4}$$

A state $\mathbf{x} \in [0, \infty)^S$ is called an *equilibrium state* (or *equilibrium point*) of a CRN $N = (S, R)$ if $F_Y(\mathbf{x}) = 0$ for all $Y \in S$.

If we let $\mathcal{E}_Y$ be the ODE shown above, the *deterministic mass action system* of a CRN $N = (S, R)$ is the coupled system of ODEs

$$(\mathcal{E}_Y \mid Y \in S). \tag{5}$$

An *initialized chemical reaction network* (*ICRN*) is an ordered pair $(N, \mathbf{x}_0)$ where $N = (S, R)$ is a CRN and $\mathbf{x}_0 \in [0, \infty)^S$ is an initial state of $N$. The ICRN $(N, \mathbf{x}_0)$ then specifies an initial value problem consisting of the deterministic mass action system (5) and the initial condition $y(0) = \mathbf{x}_0(Y)$ for all $Y \in S$. By the standard existence-uniqueness theory for ODEs [3, 14], this initial value problem has a unique solution $\mathbf{x}(t)$ defined for some interval $[0, b)$ where $b \in (0, \infty]$. In this paper, all CRNs are well-behaved and have solutions defined for all $t \in [0, \infty)$.

Under deterministic mass action semantics, the solution $\mathbf{x}(t)$ of an ICRN $(N, \mathbf{x}_0)$ completely specifies the meaning of the CRN $N$ initialized with the state $\mathbf{x}_0$.

## 3. TERMOLECULAR RESTORATION

The termolecular signal restoration algorithm presented here is used in [10, 9] and consists of the two reactions

$$2X + Y \xrightarrow{1} 3X \tag{6}$$

$$2Y + X \xrightarrow{1} 3Y. \tag{7}$$

Intuitively, both species $X$ and $Y$ are competing for total majority by converting the opposing species to itself. The algorithm is designed so that when the system is initialized with some concentrations of $X$ and $Y$, the species with the initial majority will asymptotically annihilate the other.

The CRN $N = (S, R)$ consisting of the reactions (6) and (7) specifies a system of ODEs. Using (4) to derive these ODEs, we see that the system consists of the equations

$$x'(t) = x(t)^2 y(t) - x(t)y(t)^2 \tag{8}$$

$$y'(t) = y(t)^2 x(t) - y(t)x(t)^2. \tag{9}$$

For clarity, we omit the references to $t$ in the ODEs and use Leibniz's notation for differentiation. Therefore we rewrite

the coupled system as

$$\frac{dx}{dt} = x^2y - xy^2 \tag{10}$$

$$\frac{dy}{dt} = y^2x - yx^2. \tag{11}$$

Nonlinear systems of ODEs are usually difficult to solve analytically. However, the simplicity of these equations makes the system easily solvable. If $N$ is the CRN consisting of the reactions (6) and (7), and $\mathbf{x}_0$ is an initial state of $N$ with $\mathbf{x}_0(X) + \mathbf{x}_0(Y) = 1$, it is easy to show that the mass action IVP $(N, \mathbf{x}_0)$ has solution

$$x(t) = \begin{cases} \frac{1}{2}(1-u), & \text{if } x(0) < \frac{1}{2} \\ \frac{1}{2}, & \text{if } x(0) = \frac{1}{2} \\ \frac{1}{2}(1+u), & \text{if } x(0) > \frac{1}{2} \end{cases} \tag{12}$$

$$y(t) = 1 - x(t), \tag{13}$$

where $u = \frac{1}{\sqrt{1 + C \cdot e^{-t}}}$ and $C = \frac{4x(0)y(0)}{(x(0)-y(0))^2}$.

By inspection of (12) and (13), we can see that the CRN has the behavior we expect. The species with the majority initial concentration eliminates the minority at an exponential rate.

The drawback to this solution is that it depends on the following assumptions.

1. The rate constants of the reactions are *exactly* equal.

2. The initial concentrations sum *exactly* to 1.

3. No other reactions interfere with the system.

In practice, we do not have infinitely precise control over rate constants and initial conditions. Furthermore, the algorithm is most useful in systems with other reactions affecting $X$ and $Y$.

We now specify a more general system that avoids these dependencies.

CONSTRUCTION 3.1. *Given the real valued constants $a > 0$, $b > 0$, and $c > 0$, let $N = (S, R)$ be a CRN with $S = \{X, Y\}$ and $R$ consisting of the reactions*

$$2X + Y \xrightarrow{a} 3X \tag{14}$$

$$2Y + X \xrightarrow{b} 3Y \tag{15}$$

$$X \xrightarrow{c} Y. \tag{16}$$
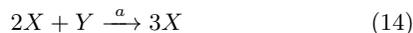
This CRN has unrelated rate constants and one additional reaction (16) that models interference that biases the system toward $Y$. Without this extra reaction, our analysis would only be modeling the system in isolation. We do not include a reaction $Y \to X$ because of the symmetry of the species $X$ and $Y$. Therefore this system captures the worst case scenario for the species $X$.

THEOREM 3.2. *If $a > 0$, $b > 0$, and $c > 0$ are real valued constants, $N = (S, R)$ is constructed according to Construction 3.1, $\mathbf{x}_0$ is an initial state of $N$ with $p = \mathbf{x}_0(X) + \mathbf{x}_0(Y)$, and $c < \frac{p^2 a^2}{4(a+b)}$, then the mass action IVP $(N, \mathbf{x}_0)$ has the*

equilibrium points $\hat{\mathbf{x}}_0$, $\hat{\mathbf{x}}_1$, and $\hat{\mathbf{x}}_2$ defined by

$$\begin{aligned} \hat{\mathbf{x}}_0(X) &= 0, & \hat{\mathbf{x}}_0(Y) &= p, \\ \hat{\mathbf{x}}_1(X) &= E_1, & \hat{\mathbf{x}}_1(Y) &= p - E_1, \\ \hat{\mathbf{x}}_2(X) &= E_2, & \hat{\mathbf{x}}_2(Y) &= p - E_2, \end{aligned}$$

where $E_1$ and $E_2$ are real valued constants defined by

$$E_1 = p\left(\frac{b}{a+b}\right) + A, \qquad E_2 = p - A,$$

and where $A = \frac{p}{2}\left(\frac{a}{a+b}\right)\left(1 - \sqrt{1 - c^*}\right)$ and $c^* = c \cdot \frac{4(a+b)}{p^2 a^2}$. Moreover, $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{x}}_2$ are exponentially stable, and $\hat{\mathbf{x}}_1$ is unstable.

PROOF. Assume the hypothesis. Then the ODEs for the species $X$ and $Y$ can be derived from the reactions (14)-(16) and are

$$\frac{dx}{dt} = ax^2y - bxy^2 - cx, \tag{17}$$

$$\frac{dy}{dt} = by^2x - ayx^2 + cx, \tag{18}$$

respectively. Note that $\frac{dx}{dt} + \frac{dy}{dt} = 0$, so the concentrations $x(t)$ and $y(t)$ differ by a constant. Since $p = \mathbf{x}_0(X) + \mathbf{x}_0(Y)$, it is easy to show that for all $t \in [0, \infty)$

$$y(t) = p - x(t). \tag{19}$$

It immediately follows from (17) that

$$\frac{dx}{dt} = ax^2(p - x) - bx(p - x)^2 - cx. \tag{20}$$

The equilibrium points of $(N, \mathbf{x}_0)$ can now be derived from the roots of the right-hand side of (20). It is routine to verify that these roots are 0, $E_1$, and $E_2$, and since $c < \frac{p^2 a^2}{4(a+b)}$ these roots are distinct and real. It follows from (19) that $\hat{\mathbf{x}}_0$, $\hat{\mathbf{x}}_1$, and $\hat{\mathbf{x}}_2$ are the equilibrium points of $(N, \mathbf{x}_0)$.

We now examine the stability of the points. It is well known that the stability of an equilibrium point can be determined by examining the eigenvalues of the Jacobian matrix evaluated at that point [8, 14]. A point is exponentially stable if each eigenvalue has a negative real part, and it is unstable if each eigenvalue has a positive real part.

Since (20) is now 1-dimensional, there is only a single eigenvalue for each equilibrium point. It is not difficult to show that the eigenvalues for the points $\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2$ are

$$\begin{aligned} \lambda_0 &= (a + b)(-E_1 E_2), \\ \lambda_1 &= (a + b)E_1(E_2 - E_1), \\ \lambda_2 &= (a + b)E_2(E_1 - E_2), \end{aligned}$$

respectively. Since $a > 0$, $b > 0$, and $E_2 > E_1 > 0$, we see that $\lambda_0$ and $\lambda_2$ are negative and $\lambda_1$ is positive. It immediately follows that $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{x}}_2$ are exponentially stable and $\hat{\mathbf{x}}_1$ is unstable. $\square$

The above theorem shows that the CRN is indeed a bistable switch if $c$ is sufficiently small. It is also clear that $E_1$ is the deciding threshold. If $x(0) > E_1$, then $x(t)$ converges to $E_2$, and if $x(0) < E_1$, then $x(t)$ converges to 0.

THEOREM 3.3. *Under the assumptions of Theorem 3.2, if $x_1$, $x_2$, and $t$ are constants such that $t \geq 0$ and $E_1 < x_1 < x_2 < E_2$, with $x(0) = x_1$ and $x(t) = x_2$, then*

$$t = \frac{1}{ap\sqrt{1 - c^*}}\left(\frac{1}{E_2}\log u + \frac{1}{E_1}\log v\right), \tag{21}$$

where $u = \frac{x_2(E_2 - x_1)}{x_1(E_2 - x_2)}$ and $v = \frac{x_1(x_2 - E_1)}{x_2(x_1 - E_1)}$.

PROOF. Assume the hypothesis. It is not difficult to show that the ODE (20) of $x(t)$ can be rewritten

$$\frac{dx}{dt} = (a + b)x(x - E_1)(E_2 - x). \tag{22}$$

By separation of variables and integration, we have

$$\int_0^t dt = \frac{1}{a + b} \int_{x_1}^{x_2} \frac{1}{x(x - E_1)(E_2 - x)} dx,$$

whence

$$t = \frac{1}{a + b} \left( \frac{1}{E_1 E_2 (E_2 - E_1)} \right) \log \left( u^{E_1} v^{E_2} \right).$$

The theorem immediately follows from the fact that $E_2 - E_1 = \frac{ap}{a+b}\sqrt{1 - c^*}$. $\square$

The above theorem gives the exact time required to move from an inital concentration $x(0) > E_1$ to a concentration closer to the equilibrium point $E_2$. It is useful to note that the only components in equation (21) that are not bounded by a constant are $\frac{1}{x_1 - E_1}$ and $\frac{1}{E_2 - x_2}$. Therefore it is clear that the concentration of $X$ converges to its stable equilibrium in logarithmic time.

## 4. BIMOLECULAR RESTORATION

The bimolecular signal restoration algorithm we present here was originally given by Angluin et al. in [2] as an approximate majority algorithm for population protocols. In the context of CRNs, the algorithm consists of the four reactions

$$X + Z \xrightarrow{1} 2X \tag{23}$$

$$X + Y \xrightarrow{1} X + Z \tag{24}$$

$$Y + Z \xrightarrow{1} 2Y \tag{25}$$

$$X + Y \xrightarrow{1} Y + Z. \tag{26}$$
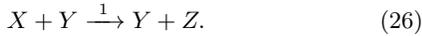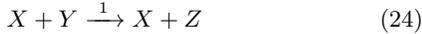
Again, the species $X$ and $Y$ are opposed to one another and are competing for total majority. The additional species $Z$ is intuitively a *neutral* species that acts as a buffer between $X$ and $Y$. Angluin et al. proved that this algorithm effeciently computes approximate majority in the context of population protocols, and here we prove that it can be effeciently used for signal restoration in deterministic CRNs.

The CRN $N = (S, R)$ consisting of the reactions (23)-(26) specifies the system of ODEs

$$\frac{dx}{dt} = xz - xy \tag{27}$$

$$\frac{dy}{dt} = yz - xy \tag{28}$$

$$\frac{dz}{dt} = 2xy - xz - xy. \tag{29}$$

This system is already considerably more complex than the termolecular system in section 3, and a solution cannot be trivially found. The task becomes especially difficult with arbitrary rate constants and interference. However, if the rate constants meet a few constraints, the behavior of this algorithm is extremely similar to that of the termolecular algorithm from section 3. We leverage this relationship in the fully general proof.

CONSTRUCTION 4.1. *Given the strictly positive, real valued constants $k_1, \ldots, k_6$, let $N = (S, R)$ be a CRN with $S = \{X, Y\}$ and $R$ consisting of the reactions*

$$X + Z \xrightarrow{k_1} 2X \tag{30}$$

$$X + Y \xrightarrow{k_2} X + Z \tag{31}$$

$$Y + Z \xrightarrow{k_3} 2Y \tag{32}$$

$$X + Y \xrightarrow{k_4} Y + Z \tag{33}$$

$$Z \xrightarrow{k_5} Y \tag{34}$$

$$X \xrightarrow{k_6} Z. \tag{35}$$

This fully generalized CRN contains the new reactions (34) and (35). These reactions serve the same purpose of representing outside interference from other reactions in the system. The new reactions still bias the algorithm in favor of $Y$, but do so in two steps instead of one. Rather than $X$ molecules being converted immediately into $Y$ molecules, they are converted to $Z$ molecules first.

Before we analyze this CRN in its full generality, we assume that the rate constants have the following relationships. For $a > 0$, $b > 0$, and $c > 0$, the rate constants $k_1, \ldots, k_6$ are related by

$$a = k_1 = \frac{k_2}{2} \tag{36}$$

$$b = k_3 = \frac{k_4}{2} \tag{37}$$

$$c = k_5 = \frac{k_6}{2}. \tag{38}$$

If these relationships are satisfied, the system of ODEs for $N = (S, R)$ from Construction 4.1 are

$$\frac{dx}{dt} = axz - 2bxy - 2cx \tag{39}$$

$$\frac{dy}{dt} = byz - 2axy + cz \tag{40}$$

$$\frac{dz}{dt} = 2axy + 2bxy - axz - byz + 2cx - cz. \tag{41}$$

Note that $\frac{dx}{dt} + \frac{dy}{dt} + \frac{dz}{dt} = 0$, so the solutions to $x(t)$, $y(t)$, and $z(t)$ differ only by a constant. If $p = x(0) + y(0) + z(0)$, then it is easy to show that for all $t \in [0, \infty)$

$$z(t) = p - x(t) - y(t). \tag{42}$$

To help us analyze a system of this complexity, we introduce several new variables. Let $\bar{x}(t)$, $\bar{y}(t)$, $v(t)$, and $w(t)$ be the functions defined by

$$\bar{x}(t) = \sqrt{x(t)}, \tag{43}$$

$$\bar{y}(t) = \sqrt{y(t)}, \tag{44}$$

$$v(t) = 2\bar{x}(t)\bar{y}(t) - z(t), \tag{45}$$

$$w(t) = \bar{x}(t) + \bar{y}(t). \tag{46}$$

LEMMA 4.2. *The ODEs of $\bar{x}$, $\bar{y}$, $v$, and $w$ are*

$$\frac{d\bar{x}}{dt} = \bar{x}\bar{y}\left(a\bar{x} - b\bar{y} - \frac{c}{\bar{y}}\right) - \frac{v}{2}a\bar{x} \tag{47}$$

$$\frac{d\bar{y}}{dt} = -\bar{x}\bar{y}\left(a\bar{x} - b\bar{y} - \frac{c}{\bar{y}}\right) - \frac{v}{2}\left(b\bar{y} + \frac{c}{\bar{y}}\right) \tag{48}$$

$$\frac{dv}{dt} = -v(\bar{x} + \bar{y})\left(a\bar{x} + b\bar{y} + \frac{c}{\bar{y}}\right) \tag{49}$$

$$\frac{dw}{dt} = -\frac{v}{2}\left(a\bar{x} + b\bar{y} + \frac{c}{\bar{y}}\right). \tag{50}$$

PROOF. Since $\bar{x} = \sqrt{x}$, its derivative is

$$\frac{d\bar{x}}{dt} = \frac{1}{2\bar{x}} \cdot \frac{dx}{dt},$$

and from (39) we obtain

$$\frac{d\bar{x}}{dt} = \frac{a}{2}\bar{x}z - b\bar{x}\bar{y}^2 - c\bar{x}.$$

Since $v = 2\bar{x}\bar{y} - z$, it follows that

$$\frac{d\bar{x}}{dt} = \frac{a\bar{x}}{2}(2\bar{x}\bar{y} - v) - b\bar{x}\bar{y}^2 - c\bar{x},$$

and therefore (47) holds. By symmetry, (48) also holds.

Since $v = 2\bar{x}\bar{y} - z$, its derivative is

$$\frac{dv}{dt} = 2\left(\bar{y}\frac{d\bar{x}}{dt} + \bar{x}\frac{d\bar{y}}{dt}\right) - \frac{dz}{dt}.$$

By substituting (41), (47), (45), and (48) into the above equation, we obtain (49).

Finally, since $w = \bar{x} + \bar{y}$, its derivative is

$$\frac{dw}{dt} = \frac{d\bar{x}}{dt} + \frac{d\bar{y}}{dt},$$

which clearly simplifies to (50) from (47) and (48). □

LEMMA 4.3. *If $\alpha = \min\{a, b\}$ and $\hat{v}$ is the function*

$$\hat{v}(t) = \frac{p \cdot v(0)}{(p + v(0))e^{\alpha p t} - v(0)}, \tag{51}$$

*then for all $t \in [0, \infty)$,*

$$v(t) \geq \hat{v}(t), \qquad \text{if } v(0) < 0 \tag{52}$$
$$v(t) \leq \hat{v}(t), \qquad \text{if } v(0) > 0 \tag{53}$$
$$v(t) = 0, \qquad \text{if } v(0) = 0. \tag{54}$$

PROOF. Assume the hypothesis. By (49), we see that if $v(t) = 0$ then $\frac{dv}{dt} = 0$. It follows that (54) holds.

If $v(0) < 0$, then

$$\frac{dv}{dt} \geq -v(\bar{x} + \bar{y})(a\bar{x} + b\bar{y})$$
$$\geq -v\alpha(\bar{x} + \bar{y})^2.$$

Since $p = \bar{x}^2 + \bar{y}^2 + z$ and $v = 2\bar{x}\bar{y} - z$,

$$(\bar{x} + \bar{y})^2 = \bar{x}^2 + \bar{y}^2 + 2\bar{x}\bar{y}$$
$$= p - z + 2\bar{x}\bar{y}$$
$$= p + v,$$

and therefore

$$\frac{dv}{dt} \geq -v\alpha(p + v).$$

This ODE can be trivially solved with separation of variables and has solution $\hat{v}(t)$. Therefore $v(t) \geq \hat{v}(t)$ for all $t \in [0, \infty)$, so (52) holds.

Similarly, if $v(t) > 0$, then

$$\frac{dv}{dt} \leq -v\alpha(p + v),$$

and so (53) holds. □

COROLLARY 4.4. *The limits of $v(t)$ and $w(t)$ are*

$$\lim_{t \to \infty} v(t) = 0 \tag{55}$$
$$\lim_{t \to \infty} w(t) = \sqrt{p}. \tag{56}$$

THEOREM 4.5. *If $a > 0$, $b > 0$, and $c > 0$ are real valued constants, $N = (S, R)$ is constructed according to Construction 4.1 such that constraints (36)-(38) are satisfied, $\mathbf{x}_0$ is an initial state of $N$ with $p = \mathbf{x}_0(X) + \mathbf{x}_0(Y) + \mathbf{x}_0(Z)$, and $c < \frac{pa^2}{4(a+b)}$, then the mass action IVP $(N, \mathbf{x}_0)$ has the equilibrium points $\hat{\mathbf{x}}_0$, $\hat{\mathbf{x}}_1$, and $\hat{\mathbf{x}}_2$ defined by*

$$\hat{\mathbf{x}}_0(X) = 0, \qquad \hat{\mathbf{x}}_0(Y) = p,$$
$$\hat{\mathbf{x}}_1(X) = E_1^2, \qquad \hat{\mathbf{x}}_1(Y) = (\sqrt{p} - E_1)^2,$$
$$\hat{\mathbf{x}}_2(X) = E_2^2, \qquad \hat{\mathbf{x}}_2(Y) = (\sqrt{p} - E_2)^2,$$

*and for all $0 \leq i \leq 2$*

$$\hat{\mathbf{x}}_i(Z) = p - \hat{\mathbf{x}}_i(X) - \hat{\mathbf{x}}_i(Y),$$

*where $E_1$ and $E_2$ are real valued constants defined by*

$$E_1 = \sqrt{p}\left(\frac{b}{a+b}\right) + A, \qquad E_2 = \sqrt{p} - A,$$

*and where $A = \frac{\sqrt{p}}{2}\left(\frac{a}{a+b}\right)\left(1 - \sqrt{1 - c^*}\right)$, and $c^* = c \cdot \frac{4(a+b)}{pa^2}$. Moreover, $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{x}}_2$ are exponentially stable, and $\hat{\mathbf{x}}_1$ is a saddle node.*

PROOF. Assume the hypothesis. We prove the theorem using the variables $\bar{x}(t)$ and $w(t)$. By Lemma 4.2 and the fact that $v = w^2 - p$, the ODEs for $\bar{x}$ and $w$ can be written as

$$\frac{d\bar{x}}{dt} = a\bar{x}^2(w - \bar{x}) - b\bar{x}(w - \bar{x})^2 - c\bar{x} - \frac{a}{2}\bar{x}(w^2 - p)\bar{x} \tag{57}$$

$$\frac{dw}{dt} = \frac{p - w^2}{2}\left(a\bar{x} + b(w - \bar{x}) + \frac{c}{w - \bar{x}}\right), \tag{58}$$

respectively. Corollary 4.4 tells us that all equilibrium points are states in which $w = \sqrt{p}$. Therefore we need only find the values for $\bar{x}$ by finding the roots of the right-hand side of

$$\frac{d\bar{x}}{dt} = a\bar{x}^2(\sqrt{p} - \bar{x}) - b\bar{x}(\sqrt{p} - \bar{x})^2\bar{y} - c\bar{x}. \tag{59}$$

The ODE (47) is identical in structure to (20), and it follows that these roots are 0, $E_1$, and $E_2$. Since $x = \bar{x}^2$, we have $\hat{\mathbf{x}}_0(X) = 0$, $\hat{\mathbf{x}}_1(X) = E_1^2$, and $\hat{\mathbf{x}}_2(X) = E_3^2$. Similarly, since $\bar{y} = w - \bar{x}$ and $y = \bar{y}^2$, it follows that $\hat{\mathbf{x}}_0(Y) = p$, $\hat{\mathbf{x}}_1(Y) = (\sqrt{p} - E_1)^2$, and $\hat{\mathbf{x}}_2(Y) = (\sqrt{p} - E_2)^2$. Finally, since $z = p - x - y$, it follows that $\hat{\mathbf{x}}_i(Z) = p - \hat{\mathbf{x}}_i(X) - \hat{\mathbf{x}}_i(Y)$ for $0 \leq i \leq 2$.

What remains to be shown is the stability of the three points. Using the ODEs (57) and (50), it is routine to show that the two eigenvalues of the Jacobian matrix evaluated at $\hat{\mathbf{x}}_0$ are

$$\lambda_1 = \lambda_2 = -c - bp.$$

Since both are negative, it follows that $\hat{\mathbf{x}}_0$ is exponentially stable.

Similarly, the eigenvalues at $\hat{\mathbf{x}}_1$ are

$$\lambda_3 = \frac{a\sqrt{p}}{2(a+b)}\sqrt{1-c^*}\left(a(1-\sqrt{1-c^*})+2b\right),$$

$$\lambda_4 = -\frac{a\sqrt{p}}{a+b}\left(a(1-\sqrt{1-c^*})+2b\right).$$

Since $0 \le \sqrt{1-c^*} < 1$, then $\lambda_3 > 0$ and $\lambda_4 < 0$, and so $\hat{\mathbf{x}}_1$ is a saddle node.

Finally, the eigenvalues of $\hat{\mathbf{x}}_2$ are

$$\lambda_5 = -\frac{ap}{2(a+b)}\sqrt{1-c^*}\left(a(1+\sqrt{1-c^*})+2b\right),$$

$$\lambda_6 = -\frac{ap}{a+b}\left(a(1-\sqrt{1+c^*})+2b\right),$$

which are both clearly negative, so $\hat{\mathbf{x}}_2$ is exponentially stable. $\square$

The above theorem not only demonstrates that the bimolecular algorithm has bistable behavior but that its behavior is closely related to the termolecular algorithm from section 3. In fact, if the CRN is initialized such that $z(0) = 2\sqrt{x(0)y(0)}$, then the solutions have a precise relationship.

From the theorem it is also clear that if $x(0) > E_1^2$ and $y(0) < (\sqrt{p} - E_1)^2$, then with certainty the system converges to the equilibrium point $\hat{\mathbf{x}}_2$.

We conclude this section with the following theorem that bounds the fully general bimolecular algorithm.

THEOREM 4.6. *If $N = (S, R)$ is constructed according to Construction 4.1 with the strictly positive real-valued constants $k_1, \ldots, k_6$ and $\hat{N} = (\hat{S}, \hat{R})$ is constructed according to Construction 4.1 with the constants $\hat{k}_1, \ldots, \hat{k}_6$ defined by $\hat{k}_1 = \frac{k_2}{2} = \min\{k_1, \frac{k_2}{2}\}$, $\hat{k}_3 = \frac{\hat{k}_4}{2} = \max\{k_3, \frac{k_4}{2}\}$, and $\hat{k}_5 = \frac{\hat{k}_6}{2} = \max\{k_5, \frac{k_6}{2}\}$, then the behavior of $X$ in $N$ is lower bounded by $\hat{X}$ in $\hat{N}$.*

PROOF. The theorem immediately follows from the fact that we are only reducing the rate constants responsible for generating $X$ and increasing the rate constants responsible for destroying $X$. $\square$

The above theorem shows that we can bound the behavior of a fully general CRN from Construction 4.1 by a slightly modified CRN. Modifying the CRN in this way may not always be possible, but in the most common uses of the algorithm, the rate constants $k_1, \ldots, k_6$ are "close" to satisfying the constraints (36)-(38), therefore the modifications are minor.

## 5. CONCLUSION

The results presented in section 3 make the termolecular algorithm easy to use in deterministic CRNs with many other reactions. Theorems 3.2 and 3.3 give exact analytical descriptions of the behavior and running time of the algorithm under the worst case conditions.

The results in section 4 demonstrate an interesting relationship between the termolecular and bimolecular algorithms. However, the results are significantly less complete. We only provide bounds on the behavior of the bimolecular algorithm and provide no analysis of the running time. Giving tight bounds on the runtime of the algorithm remains an open problem at this time.

## 6. REFERENCES

[1] Dana Angluin, James Aspnes, Zoë Diamadi, Michael J. Fischer, and René Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed Computing*, 18(4):235–253, 2006.

[2] Dana Angluin, James Aspnes, and David Eisenstat. A simple population protocol for fast robust approximate majority. *Distributed Computing*, 21(2):87–102, 2008.

[3] Tom M Apostol. *Calculus, Volume 2: Multi-Variable Calculus and Linear Algebra with Applications to Differential Equations and Probability.* John Wiley & Sons, 1969.

[4] Ho-Lin Chen, David Doty, and David Soloveichik. Rate-independent computation in continuous chemical reaction networks. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, pages 313–326. ACM, 2014.

[5] David Doty. Timing in chemical reaction networks. In *Proceedings of the 25th Symposium on Discrete Algorithms*, pages 772–784, 2014.

[6] Shawn M. Douglas, Ido Bachelet, and George M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335(6070):831–834, 2012.

[7] Samuel J. Ellis, Eric R. Henderson, Titus H. Klinge, James I. Lathrop, Jack H. Lutz, Robyn R. Lutz, Divita Mathur, and Andrew S. Miner. Automated requirements analysis for a molecular watchdog timer. In *Proceedings of the 29th International Conference on Automated Software Engineering*, pages 767–778. ACM, 2014.

[8] Irving Robert Epstein and John Anthony Pojman. *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos.* Oxford University Press, 1998.

[9] Thomas Hinze, Raffael Fassler, Thorsten Lenser, and Peter Dittrich. Register machine computations on binary numbers by oscillating and catalytic chemical reactions modelled using mass-action kinetics. *International Journal of Foundations of Computer Science*, 20(3):411–426, 2009.

[10] Titus H. Klinge, James I. Lathrop, and Jack H. Lutz. Robust biomolecular finite automata. Technical Report 1505.03931, arXiv.org e-Print archive, 2015.

[11] David Soloveichik. Robust stochastic chemical reaction networks and bounded tau-leaping. *Journal of Computational Biology*, 16(3):501–522, 2009.

[12] David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, 2008.

[13] David Soloveichik, Georg Seelig, and Erik Winfree. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences*, 107(12):5393–5398, 2010.

[14] Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140 of *Graduate Studies in Mathematics.* American Mathematical Society, 2012.