

# Mathematical Structure of Chemical Reaction Networks: Toward a Theory of the Origin of Life

M. Flashman

*Physics Department, Reed College, 3203 SE Woodstock Blvd. Portland OR, 97202 and  
Santa Fe Institute, 1399 Hyde Park Rd. Santa Fe NM, 87501*

(Dated: 20 October 2007)

The following is an account of my research efforts while at the Santa Fe Institute during the summer of 2007. The questions considered in this paper stem from work on the origin of life and in particular the problem of sparsity of core metabolic pathways in living organisms. This research aims at providing a firm mathematical foundation to the rigorous investigation of such a problem. Precisely, we consider the theoretical properties of random driven chemical reaction networks. While the problem of consideration is in its original form one of biology and biochemistry, it and other related problems will be treated whenever possible and without apology from a purely mathematical perspective. Some interesting work relevant to this problem is summarized and a few new results are presented. Strong emphasis is placed upon methodology and the wide variety of interesting if not promising approaches.

## I. INTRODUCTION

Eric Smith and Harold Morowitz, in collaboration with Shelley Copley, argue that life emerged on earth according to a metabolism-first scenario. In this scenario a few core metabolic pathways, ubiquitous to all living things, are understood as a likely or even necessary result of early-earth's energetic environment and transport mechanisms of geochemistry, very much "like a lightning bolt through the graph of geochemical reactions." [1] From this lowest level of chemical organization, genetic inheritance and Darwinian selection may perturb core pathways to achieve higher levels of complexity and variation. For an in-depth look at this argument, see the aforementioned reference.

Within the context of this scenario, the question that we begin to approach is Why, given the great variety of possible geochemical reactions, are so few core metabolic pathways observed in life today? Smith has suggested that a graph theoretic analysis of the system, removed from the details of geochemistry, may be sufficient for explaining the observed sparsity. Under this approach, the earth's energetics and geochemistry are reduced to driven random chemical reaction networks subject to appropriate constraints. With this well defined construction it is then possible to ask such question as: What kinds of global structures can and cannot emerge? What is the density, dominance and diffuseness/sparsity of emergent transport structures on such networks? How stable are such structures to perturbation?. The model also provides a convenient stage for studying such important topics in complexity science as modularity, functionality, emergence, and generally, complex dynamical processes. Finally, from the mathematical perspective, the model has a rich mathematical architecture, incorporating such areas of pure mathematics as the theory of differential equations, graph/hypergraph theory, operations research, abstract algebra, and combinatorics. These different mathematical facets can be explored in isolation, with respect to the model. But interestingly, the model

is ideal for studying their interplay.

The questions of primary concern described above turn out to be nontrivial questions to answer for such a model. Here we will explore some of the preliminary and foundational aspects of such a model and more generally the mathematics used to describe such a model. It is our hope that such a foundation will benefit the future investigation of this and other related problems. In addition we hope to demonstrate that this problem contains a rich and varied mathematical structure, much of it unexplored, ideal for investigation by the applied or pure mathematician, or the mathematically inclined scientist.

This paper is arranged as follows. In Section II, we introduce the model and necessary notions of chemical kinetics. Section III contains an preliminary analysis of linear aspects of the model, with emphasis upon conservation laws for the network. In Section IV, graph theoretic elements of this model are considered, precisely in terms of hypergraphs. Section V contains a brief survey of steady state operation of reaction networks and the algebraic structure employed in its analysis. The second half of this paper consist of a spattering of different mathematical and methodological approaches. With Section VI, we work toward a realization of the analogy between chemical reaction networks and electrical circuits. Section VII contains some results from simple numerical simulations of the model. We conclude with a summery and comment on directions for future inquiry.

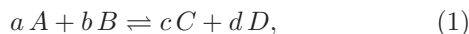
## II. MODELING THE ORIGIN OF LIFE

Before the model is introduced, we review some basic elements from the theory of chemical kinetics. Along the way, much of the formalisms of the model will be described. For a review of the theory of chemical kinetics, see [2].

### A. Chemical kinetics

Let  $\mathcal{C}$  be a set of  $m$  chemical species and  $\mathcal{E}$  be the set of  $l$  atomic species. Each molecule of a chemical species is made up of several different atomic species in varying amounts. We say that molecule  $A_i \in \mathcal{C}$  contains  $Z_{hi}$  atoms of  $\alpha_h \in \mathcal{E}$ .  $\mathbf{Z}$  is called the *chemical composition matrix*. In a chemical system, each chemical species  $A_i$  is ascribed a concentration denoted  $[A_i]$ . So the system is described by the concentration state vector  $[\mathbf{A}] = (A_1 \dots A_m)^T$ . It will be this state vector that develops with time. By assumption  $[A_i] \geq 0$  for all  $i$  and  $\sum_{i=1}^m A_i > 0$ . The second assumption ensures that the system is not empty.

By an *elementary reaction* we mean a chemical reaction of the form



where  $A, \dots, D \in \mathcal{C}$  and  $a, \dots, d \in \mathbb{N}$ , for which no chemical interactions/processes occur that are not represented by Eq. (1). This is to say that everything happening during the chemical reaction is described by Eq. (1). For example, a reaction such as



would be elementary if in fact  $2A$ 's and one  $B$  simultaneously came together to form  $D$ , but would not be elementary if  $A$  and  $B$  first came together to form  $AB$  and then  $A$  and  $AB$  came together to form  $D$ . Although reactions considered here will generally be reversible, we arbitrarily assign a positive direction to the reaction (from left to right in Eq. (1)) and speak of species entering into the reaction as *reactants* and species leaving the reaction as *products* (so  $A$  and  $B$  are reactants and  $C$  and  $D$  are products).

When a reaction such as Eq. (1) is elementary, the rate  $f$  at which the reaction proceeds (in the positive direction) is described by the *rate equation* (or *mass action law*):

$$f = k[A]^a[B]^b - \bar{k}[C]^c[D]^d, \quad (2)$$

where  $k$  and  $\bar{k}$  are rate ‘‘constants’’. Actually, each rate constant is a function of temperature and reactant and product chemical potentials, which are in turn functions of reactant and product chemical concentrations. For a discussion and derivation of the appropriate functional form of the rate constants, see [2]. In the present analysis, rate constants will be treated as independent model parameters.

The rate at which a chemical species concentration changes is equal to the sum of the rates at which each reaction produces the species, each reaction rate weighted by the amount of the species produced, minus the sum of the rates at which each reaction uses the species. If Eq. (1) is the only reaction of the system, then

$$[\dot{A}] = -af, \quad [\dot{B}] = -bf, \quad [\dot{C}] = cf, \quad \text{and} \quad [\dot{D}] = df. \quad (3)$$

For a chemical system of  $m$  species  $\mathcal{S} = \{A_1, \dots, A_m\}$  and  $n$  reactions  $\mathcal{R} = \{R_1, \dots, R_n\}$ , Eq. (1) generalizes to

$$\sum_{i=1}^m a_{ij}A_i \rightleftharpoons \sum_{i=1}^m b_{ij}A_i \quad (j = 1, \dots, n) \quad (4)$$

and Eq. (2) becomes

$$f_j = k_j \prod_{i=1}^m [A_i]^{a_{ij}} - \bar{k}_j \prod_{i=1}^m [A_i]^{b_{ij}} \quad (j = 1, \dots, n). \quad (5)$$

And lastly, Eq. (3) becomes

$$[\dot{\mathbf{A}}] = \mathbf{S} \mathbf{f} \quad (6)$$

where

$$S_{ij} = b_{ij} - a_{ij} \quad \text{and} \quad \mathbf{f} = (f_1 \dots f_n)^T.$$

We refer to  $\mathbf{S}$  as the *stoichiometric matrix*.

We say that a reaction is *well formed* if it conserves the amount of each atomic species participating in the reaction through the reaction process. So the number of atoms of carbon, for instance, entering into a well formed reaction will equal the number of atoms of carbon leaving the reaction. It has been assumed in the previous discussion that all reaction were well formed. The property of a reaction being well formed constitutes a proper conservation law for that reaction. If all of the reactions in a system are well formed (which must be the case in any real chemical system), then the set of all conservation laws associated with the well formed property are given by the matrix equation:

$$\mathbf{ZS} = 0. \quad (7)$$

Let  $\zeta = \text{Rank}(\mathbf{Z})$ . It follows immediately from Eq. (7) that  $\text{Rank}(\mathbf{S}) \leq \min(m - \zeta, n)$ , but we will have occasion to discuss these matters further in Section III.

### B. The model

We represent the geochemistry of early-earth with a random chemical reaction network, consisting of  $m$  species and  $n$  reactions. The numbers  $m$  and  $n$  will generally be considered to be large, as the number of chemical species and reactions of early geochemistry were likely on the order of a hundred thousand. A chemical composition matrix is also assigned to the system such that Eq. (7) holds, that is, random networks are drawn from a population of well formed networks. Forward and reverse rate constants are ascribed randomly to each reaction by way of random assignments of free energies of formation and activation energies to species and reactions, respectively. For the purposes of theoretical investigation, rate constants will generally be treated as independent model parameters. Temperature will be treated as fixed during

reactions. This would ofcourse be the case if the system were in contact with a heat reservoir such as a geothermal vent. However, it could be an interesting extension of the model to include temperature variations, or to keep track of the internal energy of the system as concentrations of species with different heat capacities evolve. The exact details of the random graph generation will be laid out in a later work. For the present theoretical work, they will in any case be of no consequence.

There are two mechanism by which the network may be driven. First, certain species concentrations may be fixed, either at high values or low values. The physical interpretation of such driving would be that certain species appear with great or insignificant concentrations with respect to other species concentrations so that in the finite window of time during which the reaction network is allowed to evolve the species concentrations, those species with exceptional concentrations experience no significant change in concentration. Mathematically, this driving mechanism renders certain of  $[A_i]$  constant in Eq. (5) (and so in Eq. (6)). Second, certain species may be sourced or drained by external reactions having some fixed reaction rate. The physical interpretation of such driving would be that certain processed external to the chemical reaction network create or use up certain species within the network. Let  $f_{ext,i}$  be the rate at which species  $A_i$  is added (or removed if  $f_{ext,i} < 0$ ) to(from) the network, ( $i = 1, \dots, m$ ). We define  $\mathbf{f}_{ext} = (f_{ext,1} \dots f_{ext,m})^T$ . Under the influence of this driving mechanism, Eq. (6) becomes

$$[\dot{\mathbf{A}}] = \mathbf{S} \mathbf{f} + \mathbf{f}_{ext}. \quad (8)$$

Note that Eq. (8) amounts to a flux conservation law for the system, in differential form. We refer to both driving mechanisms as *boundary conditions* of the network. The combination of driving mechanisms that would most accurately describe the state of early geochemistry have yet to be considered in great detail.

### III. CONSERVATION LAWS AND FREE CURRENTS

This sections follows a line of thought pursued by E. Smith and the author in the previous weeks, a similar treatment is given in [3].

#### A. Null space dimension and basis

As mentioned in Section II A,

$$d \equiv \text{Rank}(\mathbf{S}) \leq \min(m - \zeta, n). \quad (9)$$

It follows that the left null space of  $\mathbf{S}$ ,  $\text{LNS}(\mathbf{S})$ , has dimension  $m - d$  and the right null space of  $\mathbf{S}$ ,  $\text{RNS}(\mathbf{S})$ , has dimension  $n - d$ . These results follow from the rank-nullity theorem and the fact that  $\text{Rank}(\mathbf{S}) = \text{Rank}(\mathbf{S}^T)$ .

As a side question, we may ask, When does equality hold? This question will be left to Section IV the graph theory associated with the problem is considered in greater detail.

Introduce a complete set  $\beta = \{\mathbf{u}_1^T, \dots, \mathbf{u}_m^T\}$  of independent (row) vectors of length  $m$ , such that  $\mathbf{u}_i^T \mathbf{S} = 0$ , ( $i = d + 1, \dots, m$ ). The vectors of  $\beta$  form a basis for  $\mathbb{R}^m$ , and the last  $m - d$  components form a basis for  $\text{LNS}(\mathbf{S})$ . Because the entries in  $\mathbf{S}$  are integers, it is always possible to write the basis vectors of  $\beta$  with integer entries. Similarly, introduce a complete set  $\gamma = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  of (column) vectors of length  $m$ , such that  $\mathbf{S} \mathbf{v}_j = 0$ , ( $j = d + 1, \dots, n$ ). Again, it is always possible to do this with integer entries, the  $\gamma$  forms a basis for  $\mathbb{R}^n$ , and the last  $n - d$  components form a basis for  $\text{RNS}(\mathbf{S})$ . From  $\beta$  and  $\gamma$ , we define matrices

$$\mathbf{U}^T = \begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_m^T \end{pmatrix} \text{ and } \mathbf{V} = (\mathbf{v}_1 \dots \mathbf{v}_n) \quad (10)$$

When we relate the present discussion to the graph topology it will convenient that basis vectors are of minimal length (while still having integer entries) and maximally orthogonal. This side problem is left for consideration at a latter time.

It would also be convenient if we could show that the  $d \times d$  non zero block of  $\mathbf{U}^T \mathbf{S} \mathbf{V}$  could be made diagonal with integer valued transformations  $\mathbf{U}^T$ , the matrix formed from the (row) vectors of  $\beta$ , and  $\mathbf{V}$ , the matrix formed from the (column) vectors of  $\gamma$ .

Write

$$\mathbf{U}^T = \begin{pmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{pmatrix} \text{ and } (\mathbf{V}_1 \ \mathbf{V}_2), \quad (11)$$

where  $\mathbf{U}_1^T$  is the matrix of the first  $d$  rows of  $\mathbf{U}$ ,  $\mathbf{U}_2^T$  is the matrix of the remaining rows of  $\mathbf{U}$ , and analogously for  $\mathbf{V}$ . Then

$$\begin{aligned} \mathbf{U}^T \mathbf{S} \mathbf{V} &= \begin{pmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{pmatrix} \mathbf{S} (\mathbf{V}_1 \ \mathbf{V}_2) \\ &= \begin{pmatrix} \mathbf{U}_1^T \mathbf{S} \mathbf{V}_1 & \mathbf{U}_1^T \mathbf{S} \mathbf{V}_2 \\ \mathbf{U}_2^T \mathbf{S} \mathbf{V}_1 & \mathbf{U}_2^T \mathbf{S} \mathbf{V}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1^T \mathbf{S} \mathbf{V}_1 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned} \quad (12)$$

where  $\mathbf{U}_1^T \mathbf{S} \mathbf{V}_1$  is a  $d \times d$  non-degenerate matrix. When it is possible to choose orthogonal bases  $\beta$  and  $\gamma$  from  $\mathbb{Z}$ , where either set could be scaled by an additional real number, such that the previously defined matrices are unitary, then after appropriately scaling the matrices under consideration the result that we seek follows from the proof of the existence and uniqueness of the singular value decomposition. Because the orthogonalization process generally requires the use of inner products and norms, it would generally seem to be the case that sufficient condition for diagonalization with integer valued matrices  $\mathbf{U}$  and  $\mathbf{V}$  will only be met under exceptional circumstances.

## B. Interpretations of the null space

For the moment we will set aside the non null space basis vectors. For latter convenience, map the internal and external rate equations to the basis defined by  $\beta$  and  $\gamma$ :

$$\hat{\mathbf{f}} \equiv \mathbf{V}^{-1} \mathbf{f} \text{ and } \hat{\mathbf{f}}_{ext} \equiv \mathbf{U}^T \mathbf{f}_{ext}. \quad (13)$$

We can then work with the conservation law Eq. (8), transformed to

$$\mathbf{U}^T [\dot{\mathbf{A}}] = \mathbf{U}^T \mathbf{S} \mathbf{V} \hat{\mathbf{f}} + \hat{\mathbf{f}}_{ext}. \quad (14)$$

Notice that the last  $m - d$  equation of Eq. (14) may be written

$$\mathbf{u}_h^T [\dot{A}_i] = f_{ext,h} \quad (h = d + 1, \dots, m). \quad (15)$$

In the absence of external source  $f_{ext,h}$ ,  $\mathbf{u}_h^T [\mathbf{A}]$  is a conserved quantity of the internal flow on the network,  $h \in \{d+1, \dots, m\}$ . If all such external sources are absent, then the collection of conserved quantities are a maximal independent set. From Eq. (15), it is also clear that while the bottom most  $m - d$  components of  $\mathbf{f}_{ext}$  can change the total amount of the conserved quantities previously mentioned, no internal current can compensate or redirect this change (since internal currents do not appear in Eq. (15)).

Although the fact was used to obtain Eq. (15), since the last  $n - d$  columns of  $\mathbf{V}$  span  $\text{RNS}(\mathbf{S})$  the values of the last  $n - d$  components of  $\hat{\mathbf{f}}$  are unspecified by either  $[\dot{\mathbf{A}}]$  or  $\mathbf{f}_{ext}$ . Hence they must be free internal currents of the network. The free currents will be considered further in relation to graph topology in Section V. It is proved in [4] that under certain conditions of consistency between kinetics and thermodynamics (stated precisely in said reference),

$$\lim_{t \rightarrow \infty} \hat{f}_j = 0 \quad (j = d + 1, \dots, n). \quad (16)$$

## IV. GRAPH THEORY OF CHEMICAL REACTION NETWORKS

In this section, the graph structure of the chemical reaction network is restated explicitly in terms of graph theory, precisely hypergraph theory. The main goal of the section is to understand graphically the results of Section III. For example what is the graphical interpretation of the null space of the stoichiometric coefficient, and what about its orthogonal complement. To that end we review the theory of flows on graphs. We then consider the possibility of such a theory for flows on hypergraphs. A few formal results are obtained; we also pose a number of problems for future consideration.

## A. Hypergraph Basics and the Model

A *weighted hypergraph* is an ordered triple  $\mathcal{H}(V, E, w)$  (or just  $\mathcal{H}$ ).  $V$  or  $V(\mathcal{H})$  is a set of *vertices*, and  $|V| = m$ .  $E$  or  $E(\mathcal{H})$  is a set of *edges*, subsets of  $V$  each of which contains atleast two vertices, and  $|E| = m$ ;  $w : V \times E \rightarrow \mathbb{N}^+$  is a function which assigns a weight  $w_{ij}$  to each  $v_i \in e_j$ . If  $v_i \notin e_j$ , then  $w_{ij}$  is by default zero. Note that  $\mathcal{H}$  contains no self loops. A *regular hypergraph* is just a weighted hypergraph with all weights set to zero or one. Note that a *graph* is a hypergraph  $\mathcal{G}$  such that  $\forall e_j \in E, |e_j| = 2$ .

A *directed weighted hypergraph* is a weighted hypergraph  $\mathcal{H}$  together with an orientation  $\sigma$ , and is denoted  $\mathcal{H}^\sigma$ .  $\sigma : E \rightarrow H \times T$  is a function which partitions every  $e_j \in E$  into a head set  $h_j$  and a tail set  $t_j$ .

$\mathcal{H}^\sigma$  admits the incidence matrix representation  $\mathbf{S} \in M_{m,n}(\mathbb{Z})$ :

$$S_{ij} = \begin{cases} w_{ij} & \text{if } v_i \in h_j \\ -w_{ij} & \text{if } v_i \in t_j \end{cases} \quad (i = 1, \dots, n; j = 1, \dots, m). \quad (17)$$

Ofcourse,  $\mathcal{H}$  admits an incidence matrix representation given by  $|\mathbf{S}|$ .

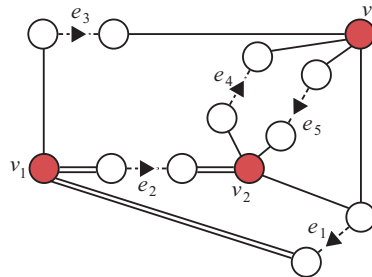


FIG. 1: A directed weighted hypergraph, represented according to the Sinanoğlu formalism. Weights appear in the figure as multiple connections.

An example of a directed weighted hypergraph appears in Figure I. Precisely, the figure is a Sinanoğlu representation of a hypergraph (this formalism provides the clearest graphical depiction of the type of hypergraphs presently under consideration). According to this formalism, a directed edge is represented by a pair of white nodes connected by a dashed line. The error along the dashed line points from the “tail set node” of the edge to the “head set node” of the edge. Of course, when a vertex is connected to a tail or head set node, the vertex is said to be in that set. The weight associated with that vertex of the edge can either appear as a numerical label on its connection or graphically as multiple connections (the second formalism is used in Figure I). The incidence matrix associated with Figure I is

$$\begin{pmatrix} 2 & -2 & -1 & 0 & 0 \\ -1 & 2 & 0 & -1 & 1 \\ -1 & 0 & 1 & 1 & -1 \end{pmatrix}.$$

We relate the chemical reaction network and the directed weighted hypergraph by making the associations between vertices and chemical species, edges and reactions, head and tail sets and reaction products and reactants, respectively, and weights and stoichiometric coefficients. Under this association (in so far as the given definition of a hypergraph allows), catalysists are ignored. It is evident that the incidence matrix is precisely the stoichiometric matrix of the reaction network.

The theory of hypergraphs has not been elaborated in the great detail that it has for the theory of graphs. However, a few important references do exist. For a somewhat technical introduction to the field, see [5].

## B. Cuts and Flows on Hypergraphs

The material of this subsection is adapted with modification from Chapter 14 in [6]. It also draws upon [7].

Let  $\mathbb{R}^E$  denote the real vector space with coordinates indexed by the edges (precisely, their indices) of  $\mathcal{H}$ . The row space of  $\mathbf{S}$ , that is the subspace of  $\mathbb{R}^E$  spanned by the rows of  $\mathbf{S}$ , is known as the *cut space of  $\mathcal{H}^\sigma$* . And the orthogonal complement of the cut space is called the *flow space of  $\mathcal{H}^\sigma$* . So the cut space is the set of all vectors  $\mathbf{x} \in \mathbb{R}^E$  satisfying  $\mathbf{S}\mathbf{x} = \mathbf{0}$ . By an abuse of notation we will simply refer to the cut and flow space of  $\mathcal{H}$  with the understanding that  $\sigma$  is fixed. In the following, these terms will be justified.

### 1. The cut space

If  $(U, V)$  is a partition of  $V(\mathcal{H})$  (the vertex set of  $\mathcal{H}$ ), into two nonempty subsets. The set of edges  $e \in E(\mathcal{H})$  with  $e \cap U \neq \emptyset$  and  $e \cap V \neq \emptyset$  is a *cut*, denoted  $C$ . We shall call  $U$  and  $V$  the *shores* of a cut. A nonempty cut that is of minimal size is called a *bond*.

An *oriented cut* is a cut with one shore declared as positive  $V(+)$ , and one shore declared as negative  $V(-)$ . Using the orientation of  $\mathcal{H}$ , that is using  $\mathcal{H}^\sigma$ , an oriented cut  $C$  determines a vector  $\mathbf{z} \in \mathbb{R}^E$  as follows:

$$z_j \equiv \begin{cases} 0 & \text{if } e_j \notin C \\ \sum_{i: v_i \in e_j \cap V(+)} S_{ij} & \text{if } e_j \in C \end{cases} \quad (18)$$

We refer to  $\mathbf{z}$  as the *signed characteristic vector* of the oriented cut  $C$ ; we will have reason to introduce the more complete notation  $\mathbf{z}(C)$ . For a given oriented cut, reversing the direction on the direction of the edges of  $\mathcal{H}^\sigma$  changes the sign of  $\mathbf{z}$ . However, it is not necessarily the case that the signed characteristic vector is invariant if all signs in Eq. (18) are switched, that is, if we define the sign characteristic vector in terms of  $V(-)$  instead of  $V(+)$ .

We call an edge  $e_j$  *balanced* if

$$\sum_{i: v_i \in t_j} w_{ij} = \sum_{i: v_i \in t_j} w_{ij}.$$

We speak of a *balanced hypergraph* every every edge  $\mathcal{H}$  is balanced. The components of the sign characteristic vector associated with balanced edges are invariant with respect to the two definitions of the characteristic vector described above. So, in a balanced hypergraph the sign characteristic vector is invariant with respect to the two definitions.

Each vertex  $v_i$  determines an oriented cut  $C(v_i)$  with positive shore  $\{v_i\}$  and negative shore  $V(\mathcal{H}) \setminus v_i$ . The  $i^{\text{th}}$  row of  $\mathbf{S}$  is the signed characteristic vector of the cut  $C(v_i)$ , so these vectors lie in the cut space of  $\mathcal{H}$ .

**Theorem 1.** *If  $\mathcal{H}$  is a balanced hypergraph, then for every cut  $C$*

$$\mathbf{z}(C) = \sum_{v \in V(+)} \mathbf{z}(C(v)) \quad (19)$$

*Proof.* We consider the sum on the RHS of Eq. (19). Let  $p \equiv |V(+)|$  and  $I = (i_1, \dots, i_p)$  an index set for the elements of  $V(+)$ . Then, by the definition of the signed characteristic vector

$$\sum_{v \in V(+)} z_j(C(v)) = \sum_{k=1}^p S_{i_k j}.$$

If  $e_j \subseteq V(-)$  (so  $e_j \notin C$ ), then  $S_{i_k j} = 0$  ( $k = 1, \dots, p$ ) and the prior sum evaluates to zero. Otherwise,

$$= \sum_{i: v_i \in e_j \cap V(+)} S_{ij}.$$

If  $e_j \subseteq V(+)$  (so  $e_j \in C$ ), then by the fact that  $\mathcal{H}$  is balanced, the prior sum evaluates to zero. Otherwise,  $e_j \in C$ . The desired result follows from the definition of the signed characteristic vector.  $\square$

**Corollary 1.** *If  $\mathcal{H}$  is a balanced hypergraph, then the signed characteristic vector of each cut lies in the cut space of  $\mathcal{H}$ .*

I suspect that more can be said of the cut space. In particular, there should exist upper and lower bounds on the dimension of the cut space, if not a precise value, that can be determined from basic topological properties of the hypergraph. And, some minimal forest on a hypergraph should act as an alternative basis for the cut space. This basis would have the advantage that it more accurately describes “fundamental” transport structures of the hypergraph. These issues will be consider in future work, but for the moment we move on.

### 2. The flow space

The flow space of  $\mathcal{H}$  is the orthogonal complement of the cut space, so consists of all vectors  $\mathbf{x} \in \mathbb{R}^E$  such that  $\mathbf{S}\mathbf{x} = \mathbf{0}$  [15]. We would like to extend the results from the theory of flows on graphs just as we have done for cuts

of graphs. However the manner by which to proceed becomes less clear than in the previous case. Whereas in the previous section we were able to extend the definition of a cut, in this case we must appropriately restrict the definition of a cycle [16]. The precise definition of a cycle, which in its present form will seem like less of a restriction than it aught, is quite cumbersome and requires additional formalism. Lacking followup results to justify such formalism, we refrain from said tedium.

### C. Other Aspects in the Theory of Hypergraphs

We define the linearization of  $\mathcal{H}$  to be the directed weighted graph  $\Lambda(\mathcal{H})$  obtained by collapsing tail sets and head sets to single nodes of a graph. This construction, then admits analysis according to the usual techniques of graph theory. It should be noted that this linearization resembles in some ways the technique of principle component analysis in graph theory.

A few obvious remarks can be made of this construction. First, the time that it takes to perform the linearization ostensibly scales with the number of edges in the hypergraph. Second the dimension of the flow space of  $\Lambda(\mathcal{H})$  provides an immediate upper bound on the dimension of the flow space of  $\mathcal{H}$ . It would also be nice if there were a simple algebraic matrix operation that could immediately yield the incidence matrix of  $\Lambda(\mathcal{H})$ .

## V. STEADY STATE OPERATION

A reaction network is said to be in steady state if  $[\dot{\mathbf{A}}] = 0$ . Under this condition, Eq. (14) yields

$$\mathbf{U}^T \mathbf{S} \mathbf{V} \hat{\mathbf{f}} = -\hat{\mathbf{f}}_{ext}. \quad (20)$$

Evidently, in steady state there is a non-degenerate linear mapping between the first  $d$  internal and external reaction rates. Notice that if a reaction network is in steady state, then the last  $n - d$  components of  $\hat{\mathbf{f}}$  are necessarily zero. But it is possible to say much more.

Recall that each reaction rate is a polynomial function in the species concentrations Eq. (5). By breaking each reaction into a forward and reverse reactions, the number of reactions in the system is doubled and each reaction rate becomes a monomial. So reaction rates represent the image of a family of monomial mappings from the concentration space into the space of reaction rates. Since the number of chemical species is generally less than the number of monomials, the image of the mapping is not the full reaction space, so there must be some dimensional constraint on reaction rates. These constraints are given by the *variety of a deformed toric ideal*. They enable a significant reduction in the dimension of the stationary reaction rates. In addition, the nonnegative stationary reaction rates must lie in the nonnegative kernel of the linear mapping given by the stoichiometric matrix (since all

reactions have been split in to forward and reverse reactions), that is, reaction rates lie in  $\text{RNS}(\mathbf{S}) \cap \mathbb{R}_{\geq}^n$ . This intersection set of a convex cone and a toric variety represents a curve, which can be mapped uniquely onto the set of stationary solutions in the concentration space under certain conditions. A complete introduction to and analysis of the algebraic and geometric structure described above can be found in [8].

The aforementioned reference also includes a discussion of chemical network structure and its relation the structure in the kinetic equations and in the Jacobian matrix. A graph describing stability is derived. It encodes important structures determining a network's stability. An application of such algebraic and geometric methods appears in [9].

For other work on the algebraic structure of equations describing chemical reaction networks, see [10] or [9].

## VI. FUNCTIONAL ANALYSIS OF CHEMICAL REACTION NETWORKS

Up to this point, the model has been approached from a level of complete generality, that is from the top down. However, there is significant reason to approach the model from the ground up. A bottom up approach is motivated primarily by the apparent similarity between chemical reaction networks and electrical circuits and the great success of such an approach in understanding electrical circuits. In this section, we describe preliminary efforts (our own as well as others) to realize analytically this similarity by way of a methodological shift, as well as some of the difficulties encountered in such a pursuit.

Precisely, we shift our attention to approximate/limited operation and modular architecture of simple chemical reaction networks. The constructing of complete electrical network does, of course, rely upon a thorough understanding of network architecture and global properties like Kirchoff's laws, among others. We will have occasion to speak briefly of efforts to apply Kirchoff style analysis of electrical circuits to chemical networks. Functional analysis of individual "chemical circuit elements" and linear approximations thereof provides a interesting, effective, and perhaps complementary approach to understanding the model under consideration.

### A. Analysis of simple circuits

We begin our analytic investigation of reaction circuit element behavior by considering the simplest reaction network that can be formed:



Eq. (21) is governed by the rate equation:

$$f = k u_A - \bar{k} u_B, \quad (22)$$

where for visual clarity, the concentrations of species  $A$  and  $B$  are written as  $u_A$  and  $u_B$  instead of  $[A]$  and  $[B]$ . From Eq. (22), we obtain two differential equations governing reaction species concentration:

$$\begin{aligned} \dot{u}_A &= -f = -k u_A + \bar{k} u_B \\ \dot{u}_B &= f = k u_A - \bar{k} u_B \end{aligned} \quad (23)$$

or in matrix form,

$$\dot{\mathbf{u}} = \mathbb{K}\mathbf{u} \quad (24)$$

where

$$\mathbf{u} \equiv \begin{pmatrix} u_A \\ u_B \end{pmatrix} \text{ and } \mathbb{K} \equiv \begin{pmatrix} -k & \bar{k} \\ k & -\bar{k} \end{pmatrix}.$$

Eq. (24) has the immediate solution [11],

$$\mathbf{u}(t) = e^{\mathbb{K}t} \mathbf{u}_0 = \frac{1}{\kappa} \begin{pmatrix} (k u_A - \bar{k} u_B) \exp(-\kappa t) + \bar{k}(u_A + u_B) \\ (\bar{k} u_B - k u_A) \exp(-\kappa t) + k(u_A + u_B) \end{pmatrix}, \quad (25)$$

where  $\mathbf{u}_0$  is an initial condition and  $\kappa \equiv k/\bar{k}$ . [17]

In the pervious analysis, it was assumed that  $u_A$  and  $u_B$  were both free variables. Alternatively, we might suppose that the quantity of  $A$  far exceeds that of  $B$ , thereby rendering  $u_A$  effectively constant. Under this assumption Eq. (23) become

$$\begin{aligned} \dot{u}_A &= 0 \\ \dot{u}_B &= k u_A - \bar{k} u_B \end{aligned} \quad (26)$$

from whence we obtain

$$\dot{u}_B + \bar{k} u_B = (D + \bar{k}) u_B = k u_A. \quad (27)$$

Eq. (27) has the homogeneous solution

$$u_B^{(h)} = C e^{-\bar{k}t}, \quad (28)$$

where  $C$  is a constant. Noting that a  $u_B = k u_A / \bar{k}$  trivially satisfies Eq. (27), we obtain the general solution to Eq. (27):

$$u_B(t) = \frac{k}{\bar{k}} u_A + (u_{B,0} - \frac{k}{\bar{k}} u_A) e^{-\bar{k}t} \quad (29)$$

where  $u_{B,0}$  is an initial condition.

The previous analysis suggests one final generalization in which  $A$  is treated as an arbitrarily controlled species, that is,  $u_A$  plays the role of an arbitrary driving force in the system. Our problem is then to solve Eq. (27) given such a  $u_A$ . To that end, we attempt to solve,

$$G(\dot{t}, t') + \bar{k} G(t, t') = k \delta(t - t'). \quad (30)$$

where  $t' > 0$  is some constant and the dot denotes differentiation with respect to  $t$ . The solution to such an equation is a Green's function, and is the integral kernel for the general solution to Eq. (27). We assume  $G(0, t') = 0$ . For  $t < t'$ , by the homogeneous solution,

$$G(t, t') = C e^{-\bar{k}t} \Theta(t - t'), \quad (31)$$

where  $C$  is a constant to be determined and  $\Theta$  is the step function. Now integrate both sides of Eq. (30) from  $t' - \epsilon$  to  $t' + \epsilon$  ( $\epsilon > 0$ ): on the right hand side,

$$\int_{t'-\epsilon}^{t'+\epsilon} k \delta(t - t') dt = k \quad (32)$$

and on the left hand side,

$$\begin{aligned} & \int_{t'-\epsilon}^{t'+\epsilon} (G(\dot{t}, t') + \bar{k} G(t, t')) dt \\ &= C e^{-\bar{k}t} \Theta(t - t') \Big|_{t'-\epsilon}^{t'+\epsilon} + \bar{k} \int_{t'-\epsilon}^{t'+\epsilon} C e^{-\bar{k}t} \Theta(t - t') dt. \end{aligned}$$

then integrating by parts

$$\begin{aligned} &= C e^{-\bar{k}(t'+\epsilon)} - C e^{-\bar{k}t} \Theta(t - t') \Big|_{t'-\epsilon}^{t'+\epsilon} \\ &+ C \int_{t'-\epsilon}^{t'+\epsilon} e^{-\bar{k}t} \delta(t - t') dt \\ &= C e^{-\bar{k}t'}. \end{aligned}$$

It follows that,

$$C = k e^{\bar{k}t'}.$$

Hence,

$$G(t, t') = k e^{\bar{k}(t'-t)} \Theta(t - t') \quad (33)$$

Finally, we obtain the general solution

$$\begin{aligned} u_B &= u_{B,0} e^{-\bar{k}t} + \int_0^\infty G(t, t') u_A(t') dt' \\ &= u_{B,0} e^{-\bar{k}t} + k e^{-\bar{k}t} \int_0^t e^{\bar{k}t'} u_A(t') dt' \end{aligned} \quad (34)$$

With this general solution in hand, define  $u_A = \text{Re}(z_A)$  and  $u_B = \text{Re}(z_B)$ , and consider the complex driving concentration

$$z_A \equiv a e^{i\omega t}.$$

According to Eq. (34),

$$\begin{aligned} z_B &= z_{B,0} e^{-\bar{k}t} + k e^{-\bar{k}t} \int_0^t e^{\bar{k}t'} a e^{i\omega t'} dt' \\ &= z_{B,0} e^{-\bar{k}t} + \frac{k}{\bar{k} + i\omega} a e^{i\omega t} \end{aligned} \quad (35)$$

Define  $\tau = 1/\bar{k}$  to be the characteristic time for the reaction. For  $t \gg \tau$ ,

$$z_B \approx \frac{k}{\bar{k} + i\omega} z_A.$$

This observation motivates the following definition. For a linear reaction we define a complex transfer coefficient

$$T \equiv \frac{k}{\bar{k} + i\omega}. \quad (36)$$

It follows that the gain of the reaction is

$$g(\omega) = |T| = \frac{k}{\sqrt{\bar{k}^2 + \omega^2}}. \quad (37)$$

For completeness, the phase function for the reaction is

$$\phi(\omega) = -\arctan(\omega/\bar{k}).$$

From these results, we see that a driven reactor acts as low pass filter in a closed system. It might be more informative to include such boundary conditions in the analysis, namely fixed concentrations (effectively driving concentrations), currents into the reaction and currents out of the reaction. Unfortunately, such considerations must be left to later consideration.

Although the analogy between chemical reactions and electrical circuits can be carried over to a limited extent, the diversity of species and the fact that only reactions can mediated concentration differences place the analogous notions of “potential” (chemical concentration) and “current” (reaction rate) on questionable footing. Some approaches have been proposed in which the analogy is carried out using chemical potentials and reaction rates. Such a formalism allows for a proper discussion and use of Kirchoff’s laws (see [12], [13], or [14]). However, the functional relation between chemical potential and reaction rate seems to lack the simplicity and elegance of that for electrical circuits.

## VII. NUMERICAL SIMULATIONS OF CHEMICAL REACTION NETWORK DYNAMICS

In order to gain qualitative insight into the possible behavior of concentration and current dynamics on the model, numerical simulations of the model were performed. It should be mentioned that the primary thrust of our investigation was toward an analytic analysis of the model. As such, the program used for the simulations is more of a toy model than a precise thorough and complete simulation, and the results of such a toy model should be taken at most as suggestive.

### A. The program

The Mathematic code for the simulation is given in Appendix A. Control parameters for the simulation appear in paragraphs two and three of the code. The program produces a  $25 \times 25 \times 60$  (precisely, a  $\{\text{NumberOfRuns} \times \text{MinorMax} \times \text{NumMus}\}$  entropy array, the  $\{i, j, k\}$  entry

of which is the entropy of the  $i^{\text{th}}$  random chemical network with the  $j^{\text{th}}$  random sink–source pair, driven with a normalized amplitude:  $k \text{ StepSize}$ . It should also be mentioned that other dynamic quantities present in the model, such as reaction rates, are also accessible. This program was run with the number of chemical species in the model ranging in five species steps from five to 45 chemical species.

### B. Results

Few presentable results were obtained from these simulations, nonetheless, we briefly sketch the behavior demonstrated by the model. Figure 2 displays the average of all entries in the entropy array described in Section VII A as it varies with the number of reactions in the chemical network. The logarithmic growth in Fig-

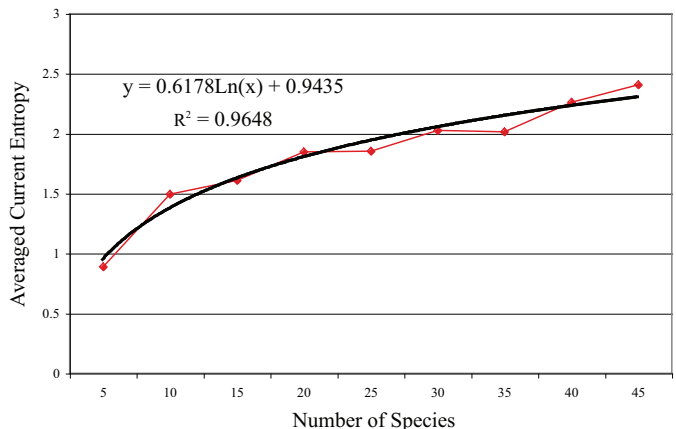


FIG. 2: Plot of average current entropy with respect to chemical reaction size. Averages are taken over all three array coordinates of the resulting data set for each run with a given number of species.

ure 2 should not be to surprising as it is characteristic of flows on growing random network. Figure 3 displays the averaged current entropies of several different random reactions of 45 chemical species as they vary with driving strength. Averages were taken over the second index of the resulting data set for the simulation with 45 chemical species. The feature to notice in Figure 3 is the approximate dip that occurs in entropy when the system is driven with a strength between 10 and 20. Although we will not include similar data for simulations of other sizes, it appears that this dip becomes more pronounced as the size of the reaction increases. The main problem in making this observation precise is normalizing current entropy and driving strengths between runs of different sizes. Drawing analogy with other nonlinear systems, this feature suggests that the the model undergoes some form of self organization when driven. This is precisely the behavior we would hope to see in the model if the metabolism first scenario outlined in Section I is to have



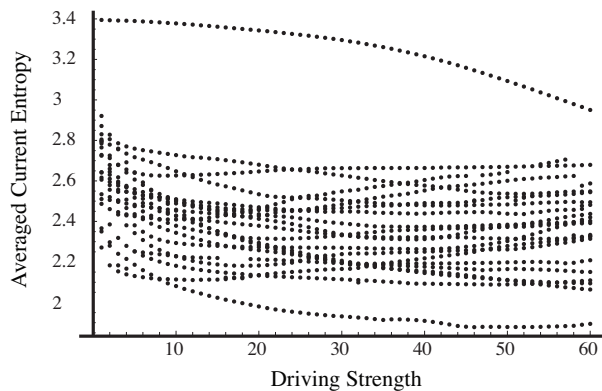


FIG. 3: Plot of average current entropy with respect to driving strength, for several random reactions having 45 chemical species. Averages were taken over the second index of the resulting data set for the simulation.

any explanatory strength.

### VIII. CONCLUSION

In this report, we have presented several basic approaches to a principled study of the metabolism first scenario to the origin of life. From the simplest conceivable model for such a system, we have presented a preliminary analytic investigation of the structure of the governing

```
Needs["DiscreteMathCombinatorica"]
Needs["StatisticsDiscreteDistributions"]
Needs["StatisticsContinuousDistributions"]
Needs["DiscreteMathGraphPlot"]
Needs["GraphicsColors"]
Needs["GraphicsGraphics"]
```

```
zeromatrix[m_, n_] := Table[Table[0, {j, n}], {i, m}]
onesmatrix[m_, n_] := Table[Table[1, {j, n}], {i, m}]
randint[m_, n_, o_] := RandomArray[DiscreteUniformDistribution[o], {m, n}]
randval[m_, n_, l_, h_] := RandomArray[UniformDistribution[l, h], {m, n}]
sprand[m_, n_, d_, max_] := Table[Table[UnitStep[d - Random[]] Random[DiscreteUniformDistribution[max]], {j, n}], {i, m}]
SpecialLog[x_] := If[x > 0, Log[x], 0]
```

```
T = 350;
kB = 8.61734310(-5);
kT = kB * T;
```

```
NumberOfRuns = 25;
MinorMax = 25;
ReacPerSpecies = 2;
Indegree = 2;
ZRange = 4;
NSpec = 45;
NReac = NSpec * ReacPerSpecies;
NumMus = 60;
```

differential equations, graph theoretic interpretations of their structure, as well as functional analysis of reaction substructure and some simple simulations of the model. While few results were obtained from this work, it is our sincere hope that these efforts have laid a basic foundation for subsequent investigations of the metabolism first scenario for the origin of life. In addition, we hope to have conveyed some of the rich mathematical structure sprung from this model. These preliminary investigation have uncovered some general areas in need of further investigation, most notably within the theory of flows on hypergraphs and the functional analysis of chemical reaction networks.

### Acknowledgments

I give many thanks to my mentor, Eric Smith, for bringing me into this research project, for his time and support through the summer. I would also like to thank Chris Moore and Aaron Clauset for useful discussions. This work was funded by the National Science Foundation through the Research Experience for Undergraduates program.

### APPENDIX A: MATHEMATICA SOURCE CODE

```

StepSize = 1;
ConvNum = 2000;

Entropy = Table[0, {i, 1, NumberOfRuns}, {j, 1, MinorMax}, {k, 1, NumMus}];
Entropy2 = Table[0, {i, 1, NumberOfRuns}, {j, 1, MinorMax}, {k, 1, NumMus}];
ShortestPathSizeList = Table[0, {i, 1, NumberOfRuns}, {j, 1, MinorMax}];

For[MasterIndex = 1, MasterIndex ≤ NumberOfRuns, MasterIndex++,
  MinRz = 0;
  While[MinRz < 1,
    InAdj = sprand[NReac, NSpec, Indegree/(NSpec), Indegree];
    ZsIntoReac = InAdj.ZForm;
    MinRz = Min[ZsIntoReac];
    MaxRz = Max[ZsIntoReac];

    OutAdj = zeromatrix[NReac, NSpec];
    For[RzInd = MinRz, RzInd ≤ MaxRz, RzInd++,
      reacts = Position[ZsIntoReac, RzInd][[All, 1]];
      SzCands = 0;
      While[SzCands < Length[reacts],
        {TmpAdj = sprand[NReac, NSpec, Indegree/(NSpec), Indegree];
        ZsOutOfTmp = TmpAdj.ZForm;
        Cands = Position[ZsOutOfTmp, RzInd][[All, 1]];
        SzCands = Length[Cands];
        For[ReacInd = 1, ReacInd ≤ Length[reacts], ReacInd++,
          {OutAssn = reacts[[ReacInd]];
          OutAdj = ReplacePart[OutAdj, TmpAdj[[Cands[[ReacInd]], All], OutAssn]};
        ];
      ];

    AdjMat = Table[Which[(i ≤ NSpec)&&(j ≤ NSpec), 0, (i ≤ NSpec)&&(j > NSpec), Transpose[InAdj]
      [[i, j - NSpec]], (i > NSpec)&&(j ≤ NSpec), OutAdj[[i - NSpec, j]], (i > NSpec)&&(j > NSpec), 0],
      {i, 1, NSpec + NReac}, {j, 1, NSpec + NReac}];

    GForm = randval[NSpec, 1, -1/4, 0];
    ZForm = randint[NSpec, 1, ZRange];
    GBarr = randval[NReac, 1, 0, .1];
    EsIntoReac = InAdj.GForm;
    EsOutOfReac = OutAdj.GForm;
    GReac = EsIntoReac - EsOutOfReac;

    GTrans = GBarr + Table[Max[EsIntoReac[[i, 1]], EsOutOfReac[[i, 1]], {i, 1, NReac}];
    KFwd = onesmatrix[NReac, 1];
    KRev = onesmatrix[NReac, 1];
    For[i = 1, i ≤ NReac, i++,
      KFwd[[i, 1]] = e-(GTrans[[i,1]]-EsIntoReac[[i,1]])/kT;
      KRev[[i, 1]] = e-(GTrans[[i,1]]-EsOutOfReac[[i,1]])/kT];

    CarbPerSpecies = 1;
    ConcCarb = CarbPerSpecies * NSpec;
    Concs0 = ConcCarb * onesmatrix[NSpec, 1]/Total[ZForm][[1]];
    Concs = Table[{conc[i][t]}, {i, 1, NSpec}];
    Currs = zeromatrix[NReac, 1];

    RateEq = KFwd * eInAdj.Log[Concs] - KRev * eOutAdj.Log[Concs];
    ConcEq = Flatten[Transpose[OutAdj - InAdj].FullSimplify[Flatten[RateEq]]];

    SystemOfEqs = Join[Table[∂tconc[i][t] == ConcEq[[i]], {i, 1, NSpec}],

```

```

Table[conc[i][0] == Concs0[[i, 1]], {i, 1, NSpec}]];
solution = NDSolve[SystemOfEqs, Flatten[Concs], {t, -0.1, 1000000}];
For[i = 1, i ≤ NSpec, i++,
  Concs0[[i, 1]] = conc[i][t]/.solution[[1, i]]/.t → 1000000];

ConcsFinal = zeromatrix[NSpec, NumMus];
CurrentFinal = zeromatrix[NReac, NumMus];

For[MinorIndex = 1, MinorIndex ≤ MinorMax, MinorIndex++,
  Source = Random[Integer, {1, NSpec}];
  Sink = Random[Integer, {1, NSpec}];
  While[Sink == Source, Sink = Random[Integer, {1, NSpec}]];
  HiConc[x.]:=Concs0[[Source, 1]]eZForm[[Source,1]]xkT/2;
  LoConc[x.]:=Concs0[[Sink, 1]]e-ZForm[[Sink,1]]xkT/2;

  ShortestPathSizeList[[MasterIndex, MinorIndex]] = Length
    [ShortestPath[FromAdjacencyMatrix[AdjMat, Type → Directed], Source, Sink]];
  PotentialList[MasterIndex][MinorIndex] = Table[HiConc[jStepSize]
    -LoConc[jStepSize], {j, 1, NumMus}];

For[MuInd = 1, MuInd ≤ NumMus, MuInd++,
  Clear[conc, Concs, RateEq, ConcEq, ConcSystemOfEqs];
  Concs = Table[{conc[i][t]}, {i, 1, NSpec}];
  conc[Sink][t] = LoConc[MuIndStepSize];
  conc[Source][t] = HiConc[MuIndStepSize];
  RateEq = KFwd * eInAdj.Log[Concs] - KRev * eOutAdj.Log[Concs];
  ConcEq = Flatten[Transpose[OutAdj - InAdj].FullSimplify[Flatten[RateEq]]];
  ConcSystemOfEqs = Join[Delete[Table [∂tconc[i][t] == ConcEq[[i]], {i, 1, NSpec}],
    {{Source}, {Sink}}], Delete
    [Table[conc[i][0] == Concs0[[i, 1]], {i, 1, NSpec}], {{Source}, {Sink}}]];
  ConcSolution[MuInd] = Flatten[NDSolve[ConcSystemOfEqs, Flatten
    [Delete[Concs, {{Source}, {Sink}}]], {t, -0.1, ConvNum}]];
  Clear[conc];
  Concs = Table[{conc[i][t]}, {i, 1, NSpec}];
  ConcSolution[MuInd] = If[Source < Sink, Insert[Insert[ConcSolution[MuInd],
    conc[Source][t] → HiConc[MuIndStepSize], Source], conc[Sink][t]
    → LoConc[MuIndStepSize], Sink], Insert[Insert[ConcSolution[MuInd],
    conc[Sink][t] → LoConc[MuIndStepSize], Sink],
    conc[Source][t] → HiConc[MuIndStepSize], Source]];
  For[i = 1, i ≤ NSpec, i++,
    ConcsFinal[[i, MuInd]] = conc[i][t]/.ConcSolution[MuInd]/.t → ConvNum];
  For[i = 1, i ≤ NReac, i++,
    CurrentFinal[[i, MuInd]] =
      (KFwd[[i, 1]] ∏j=1NSpec ((Evaluate[conc[j][t]/.ConcSolution[MuInd]/.t →
        ConvNum])InAdj[[i,j]]) - KRev[[i, 1]] ∏j=1NSpec ((Evaluate[conc[j][t]
        /.ConcSolution[MuInd]/.t → ConvNum])OutAdj[[i,j]]));
    J = ∑z=1NReac Abs[CurrentFinal[[z, MuInd]]];
    Entropy[[MasterIndex, MinorIndex, MuInd]] =
      - ∑p=1NReac (  $\frac{\text{Abs[CurrentFinal][p, MuInd]]}{J}$  SpecialLog  $\left[ \frac{\text{Abs[CurrentFinal][p, MuInd]]}{J} \right]$  );
  ];
];
];

```

[1] E. Smith and H. Morowitz (2006), in SFI Working Pa-

[2] J. E. House, *Principles of Chemical Kinetics* (Wm. C.

- Brown, 1997).
- [3] E. Smith (2007), research notes.
- [4] G. R. Galvalas, *Nonlinear Differential Equations of Chemically Reacting Systems*, Springer Tracts in Natural Philosophy (Springer-Verlag, 1968).
- [5] C. Berge, *Hypergraphs*, vol. 45 (North-Holland Mathematical Library, 1989).
- [6] C. Godsil and G. Royle, *Algebraic Graph Theory*, vol. 207 of *Graduate Texts in Mathematics* (Springer-Verlag, 2001).
- [7] R. Diestel, *Graph Theory* (Springer-Verlag, 2005), electrical edition ed.
- [8] A. Sensse, Dissertation, Ottp-von-Guericke-University (2005).
- [9] K. Gatermann and B. Huber, *J. Sym. C.* **33**, 275 (2002).
- [10] K. Gatermann, M. Eiswirth, and A. Sensse, *J. Sym. C.* **40**, 1361 (2005).
- [11] V. I. Arnold and Y. S. Il'yashenko, *Dynamical Systems* *I: Ordinary Differential Equations*, vol. 1 of *Encyclopaedia of Mathematical Sciences* (Springer-Verlag, 1988).
- [12] H. Quin, D. A. Beard, and S. Liang, *Eur. J. Biochem* **270**, 415 (2003).
- [13] D. A. Beard, S. Liang, and H. Qian, *Biophys J.* **83**, 79 (2002).
- [14] M. Ederer and E. D. Gilles, *Biophys J.* **92**, 1846 (2007).
- [15] In fact, this is equivalent to a statement of Kirchoff's junction law at vertices:
- $$\sum_{j:v_i \in H_j} w_{ij} x_j - \sum_{j:v_i \in T_j} w_{ij} x_j = 0.$$
- [16] Recall that in graph theory the signed characteristic vectors (appropriately defined) of the cycles of the graph generate the flow space.
- [17] The second equality of Eq. (25) is not exactly trivial, but it is mechanical and is left in this case to *Mathematica*.