Bulletin of Mathematical Biology (2001) **00**, 1–34
doi:10.1006/bulm.2001.0263
Available online at http://www.idealibrary.com on IDEAL®

An Algebraic-combinatorial Model for the Identification and Mapping of Biochemical Pathways

JOSEPH S. OLIVEIRA

Environmental Molecular Sciences Laboratory, Theory,
Modeling and Simulation Group, Pacific Northwest National Laboratories,
Richland, Washington,
U.S.A.
E-mail: jso@pnl.gov

COLIN G. BAILEY

School of Mathematical and Computing Sciences,
Victoria University of Wellington,
Wellington,
New Zealand
E-mail: Colin.Bailey@vuw.ac.nz

JANET B. JONES-OLIVEIRA AND DAVID A. DIXON

Environmental Molecular Sciences Laboratory, Theory,
Modeling and Simulation Group, Pacific Northwest National Laboratories,
Richland, Washington,
U.S.A.
E-mails: jjo@pnl.gov; David.Dixon@pnl.gov

We develop the mathematical machinery for the construction of an algebraic-combinatorial model using Petri nets to construct an oriented matroid representation of biochemical pathways. For demonstration purposes, we use a model metabolic pathway example from the literature to derive a general biochemical reaction network model. The biomolecular networks define a connectivity matrix that identifies a linear representation of a Petri net. The sub-circuits that span a reaction network are subject to flux conservation laws. The conservation laws correspond to algebraic-combinatorial dual invariants, that are called S- (state) and T- (transition) invariants. Each invariant has an associated minimum support. We show that every minimum support of a Petri net invariant defines a unique signed sub-circuit representation. We prove that the family of signed sub-circuits has an implicit order that defines an oriented matroid. The oriented matroid is then used to identify the feasible sub-circuit pathways that span the biochemical network as the positive cycles in a hyper-digraph.

Editor Query:
Accepted year 2000
OK?

0092-8240/01/000001 + 34 $35.00/0

1. BACKGROUND

Biochemical reaction networks are defined by sets of chemical reactions coupled through common reacting species. In general, the actual number of reacting species can be quite large, for example, tens to hundreds of thousands of reactions may be needed to describe the complex, coupled phenomena. Although one could solve the set of reactions by using traditional partial differential equation (PDE) solvers, as is often done in the chemical process industry, this does not provide insights into the various pathways and how biomolecules connect to each other, for example, through feedback loops, oscillators, etc. Furthermore, even though there are many possible connected processes, there are often not large amounts of key molecules present in the cell, sometimes only on the order of thousands to millions. This is in contrast to dealing with molar quantities ($\sim 6 \times 10^{23}$ molecules/mole). Even for picomolar quantities, there are still $\sim 10^{12}$ molecules, far larger than the 10^3 to 10^6 proteins key to the reaction networks. For such small quantities, because molecules react only in integer amounts, discrete mathematical approaches may be more appropriate for the study of such systems as compared to using continuous PDE solvers, cf. Arkin *et al.* (1998), McAdams and Arkin (1997), and Samoilov *et al.* (2001). The integration of a system of chemical kinetic equations numerically could negate several thousand molecules per time step in round-off errors alone. Therefore, the need for combinatorial models to accomplish the task of computationally identifying biochemical pathways is clearly indicated.

In this paper, we describe the use of Petri nets, cf. Reisig (1985), given as generalized hyper-digraphs, to describe biochemical reaction networks. In addition, we define some of the basic mathematical theorems needed in order to use Petri nets in the description of biological networks. Such graph-theoretic approaches will be useful in studying biochemical networks, as they provide insight into how molecules are coupled and how one can selectively interfere with a network to produce a desired biochemical change. Graphical network models can be used to provide a computational framework in the presence of unknowns that allows one to test the operational consistency of a network model.

Various attempts have been made to produce a single graph-theoretic model of biochemical pathways that captures not only the dynamical system's complexity of the chemical kinetics but also generates an algebraic model of biochemical circuits using Boolean switching circuit logic. Examples of such attempts are: the algebraic models of biochemical reaction pathways of Alberty (1991a,b, 1992, 1994, 1996) and Clark (1988); the binary networks models of Kauffman (1971, 1993); the metabolic network graphs of Kohn and Letzkus (1983), Kohn and Lemieux (1991), Karp (1998), Seressiotis and Bailey (1988), and Mavrovouniotis and Stephanopoulos (1990); the Petri net models of metabolic pathways in Reddy *et al.* (1996); and the metabolic reaction network models of Schilling and Palsson (1998, 1999) and Schilling *et al.* (1999). The common thread that runs through these various approaches is that the biochemical system at steady state has a unique linear alge-

braic representation that indicates how to obtain flux balanced systems of chemical reaction equations.

Building on this rich history and synthesizing the best of these approaches, we rigorously define the mathematical framework necessary to identify the circuits that extremally span the biochemical reaction pathways as the positive-valued cycles of hyper-digraphs. This graph-theoretic approach has numerous advantages over the linear algebraic approach, as we will show. We extend these earlier approaches by constructing a combinatorial geometric model of the biochemical reactions, referred to as an oriented matroid. This combinatorial geometry construction defines a finite-dimensional space over the rationals, $\mathbb{Q}$, which in turn defines sets of sub-circuits in a graph that are the spanning trees of a hyper-digraph of the biochemical reaction network.

By constructing the matroid of sub-circuits over the rationals $\mathbb{Q}^n$ instead of the reals $\mathbb{R}^n$, we reduce the computational complexity of the search space. By using integer arithmetic, we avoid the round-off errors introduced by approximating real numbers with rationals, thereby avoiding the problem of introducing spurious cycles. It follows that searching for sub-circuits over the rationals is computationally less expensive than searching over the reals. We are effectively constructing an integer-based lattice representation of a combinatorial geometry that faithfully models the linear algebraic system of the biochemical rate equations as a partially ordered n-set of sub-circuits of a hyper-digraph.

The combinatorics of our model represents biochemical pathways in terms of counting arrangements of colored balls in boxes, where the balls are molecular species with specific attributes that occupy marked places in the network. The partitions of multi-sets of reacting chemical species (the colored balls) are arranged into the set of marking place holders of kinetic reactions referred to as places (the boxes). This operational process generates a cycle decomposition of our hyper-digraph into an n-set of sub-circuits. The complex operational processes associated with biochemical pathways lend themselves to being faithfully modeled by the hyper-digraph abstraction of a Petri net.

In all generality, a Petri net is a directed, simply-connected bi-partite graph, composed of nodes (vertices) and edges, cf. Berge (1973) and Diestel (2000). Every Petri net has two types of nodes: state nodes, called places $\boxed{p}$, and transition nodes, denoted by $\boxed{t}$. State nodes hold information called tokens whereas transition nodes define a set of conditions or rules that regulate the flow of information from one state node to another, i.e., information from state $\boxed{p_1}$ is transferred to state $\boxed{p_2}$ when some set of transition conditions $\boxed{t}$ are satisfied. The transition constraint rules can be quite complicated and may contain, for example, time-dependent probabilistic and/or conditional logic. Thus, they need not only represent reacting chemical species, but can also represent biochemical structures such as vesicles in which different chemistry can occur. Diagrammatically, this is represented as $\boxed{p_1} \rightarrow \boxed{t} \rightarrow \boxed{p_2}$.

From an operational control systems theory point-of-view, the transitions $\boxed{t}$ act as the control laws for the system whereas the states $\boxed{p}$ act as the state variables in the Petri net representations of the biochemical reactions. The Petri net representation is a combinatorial abstraction of the molecular interactions defined over a chemical reaction space where the transitions define the operational conditions or rules that must be satisfied for a reaction to occur. In this graphical representation, a chemical species moves from one state to another subject to a transition rule based on chemical equilibria (thermodynamics) or kinetics. The chemical species do not pass formally through the transition but rather are subject to the rules described by the transition.

Two state nodes are connected subject to a transition node simply if it is possible for the second state to be reached from the first state through some physical/chemical mechanism which is reversible. Of course, the actual amount of chemical system reversibility may be very small and is dependent on the steady-state condition and/or kinetic rate constants. Since all of the reactions under consideration are potentially reversible, then these paired sets of transitions are identified explicitly. The implications of the complexity in the transition rules is deferred to a subsequent paper, and, in this discussion, only the existence of the species moving through the states of the pathways are indicated. In other words, we do not assign weights or probabilities to the paths of the reactions—we only indicate the existence of the reaction paths and their directions.

We can use our combinatorial-graphic network model to design empirical models and test them against flux conservation laws. By construction, each constituent sub-circuit in the hyper-digraph that models a biochemical network satisfies a flux conservation law. This conservation law can be shown to be the analog of Kirchhoff's circuit current law.

Alternative chemical network derivations of analogs to Kirchhoff current and voltage laws, using bond graphs, have been derived in Oster *et al.* (1973), Schnakenberg (1979) and Peusner (1986). It will be shown in a sequel that, with the introduction of Petri net invariants, these results have a natural combinatorial generalization in terms of maximum/minimum flows. This combinatorial generalization recasts the Kirchhoff laws in terms of oriented matroids that define oriented matroid programs subject to objective functions, which are in turn used to identify the chemical network flow sub-circuit paths that either maximize or minimize chemical circuit impedance, capacity, inductance and bulk species transport for a biochemical network.

The flux conservation law governs each chemical equation that defines a biochemical reaction pathway. All combinations of the sub-circuits, up to and including the entire network, therefore satisfy this conservation law. The conservation law also identifies a set of network invariants. These invariants are used in the consistent graphical network model to define a set of tests. The tests are used to search for the set of admissible biochemical outcomes to the hypothetical introduction of unknowns while guaranteeing the preservation of the network conservation

law. Therefore, this process model infers only physically realizable outcomes. An alternative approach is to quantitatively generate numerical solutions to the reaction/diffusion chemical transport equations that model the physical processes. In theory, good numerical approximations for the reaction rates and chemical concentrations can be determined and then used to tune the reaction model. However, this approach does not produce a method for checking the logical consistency of the model itself. In other words, the solution may indicate the need to change a parameter in order to satisfy the hypothesis that generates nonphysical empirical solutions. Continuum-based differential equation solvers potentially have serious drawbacks, considering the fact that continuum-based numerical analysis schemes can wash out hundreds to thousands of molecular species due to round-off errors thereby missing very rare events or introducing sub-circuit paths that do not correspond to physical network paths.

2. BIOCHEMICAL REACTIONS, PETRI NETS, AND GRAPHS

The complex biochemical processes can be considered to be composed of two fundamental types of biochemical building blocks: molecular reaction and complex formation. These fundamental building blocks, summarized in Table 1, can be considered as Petri net stencils. Each of these reactions has associated with it a Petri net representation, which pictorially depicts the communication pathways, and its incidence matrix, which mathematically specifies which states are receiving and/or transmitting information to each other subject to the transition rules. Forward paths are denoted by solid lines with directional arrows, and backward paths are denoted by dashed lines with directional arrows. In the incidence matrix, a state $\boxed{p}$ would have a $+1$ entered for transition $\boxed{t}$ if information is propagating subject to the transition node into the state node; similarly, a state would have an entry -1 for the transition if the information is propagating from the state node subject to the transition node.

The simple reversible reaction path is referred to as either a cycle or a circuit. In this context, reversible means that there is a transition rule governing the return pathway signal with the requisite rules contained in its transition, which complements the forward path. The reaction is composed of two forward path segments and two backward path segments. As an example of this reaction stencil, consider the state $\boxed{p_1}$ to be the enzyme bound state of the substrate and the state $\boxed{p_2}$ to be the enzyme bound state of the product. The bi-directional arrow of the biochemical representation indicates that the reaction is reversible.

In the example, the arrow from $\boxed{p_1}$ to $\boxed{t_1}$ is evaluated to be -1 to indicate that the signal is leaving state $\boxed{p_1}$, whereas the arrow from $\boxed{t_1}$ to $\boxed{p_2}$ is evaluated to be $+1$ to indicate that the signal is entering state $\boxed{p_2}$. Similarly, the arrow from $\boxed{p_2}$ to $\boxed{t_2}$ is evaluated to be -1 to indicate that the signal is leaving state $\boxed{p_2}$;

Table 1. Fundamental Petri net stencils. The transition nodes are physically identified for reversible reactions; if the reaction kinetics are correct, t_1 and t_2 must be identified. This follows from the principle of microscopic reversibility.

Reaction Type	Chemistry Representation	Reaction Representation	Petri Net Representation	Incidence Matrix Representation
Molecular Reaction	$A \leftrightarrow B$			$\begin{array}{c\|cc} & t_1 & t_2 \\ \hline p_1 & -1 & 1 \\ p_2 & 1 & -1 \end{array}$
Complex Formation	$A + B \leftrightarrow A : B$			$\begin{array}{c\|cc} & t_1 & t_2 \\ \hline p_1 & -1 & 1 \\ p_2 & -1 & 1 \\ p_3 & 1 & -1 \end{array}$

whereas the arrow from $\boxed{t_2}$ to $\boxed{p_1}$ is evaluated to be $+1$ to indicate that the signal is entering state $\boxed{p_1}$. The incidence matrix associated with the Petri net follows accordingly. Note that the columns of the incidence matrix always sum to zero for a balanced system in terms of sources and sinks.

A key goal of this work is to find sub-circuit pathways through the Petri net that correspond to biochemically feasible reaction pathways. To accomplish this, we first develop some background theory of cycles, hyper-digraphs and Petri nets.

2.1. *Basic definitions of standard Petri nets.*

DEFINITION 1. A *Petri net* is an ordered 4-tuple, $\Pi = \langle P, T, I, O\rangle$, defined over $\mathbb{Q}$, where

— $P = \{p_1, \ldots, p_n\}$ is a finite set of *places*;
— $T = \{t_1, \ldots, t_m\}$ is a finite set of *transitions*;
— $I\colon P \times T \to \{0, 1\}$;
— $O\colon P \times T \to \{0, 1\}$;
— for all p and t, $I(p, t)O(p, t) = 0$.

I defines the *input* set of arcs from places to transitions and O defines the *output* set of arcs from transitions to places.

For a fixed transition t we define

$$^\bullet t = \{p \in P \mid O(p, t) = 1\}$$
$$t^\bullet = \{p \in P \mid I(p, t) = 1\},$$

where $^\bullet t$ is the set of *input* places to transition t and $t^\bullet$ is the set of *output* places to transition t.

More generally, we can define the set of input/output places for any $X \subseteq T$ by

$$^\bullet X = \{p \in P \mid \text{for some } t \in X\ O(p, t) = 1\}$$
$$X^\bullet = \{p \in P \mid \text{for some } t \in X\ I(p, t) = 1\}.$$

We can also define similar sets for places. If p is a fixed place then

$$^\bullet p = \{t \in T \mid I(p, t) = 1\}$$
$$p^\bullet = \{t \in T \mid O(p, t) = 1\}.$$

DEFINITION 2. A *marking* of a Petri net Π is a function $M\colon P \to \mathbb{N}$.

DEFINITION 3. A transition t is *enabled* by a marking M iff for all $p \in {}^\bullet t$, $M(p) \geq I(p,t)$.

Any transition that is enabled by a marking M can *fire*. This produces a new marking M' defined by

$$M'(p) = M(p) + O(p,t) - I(p,t).$$

We will use the notation

$$M \xrightarrow{t} M'$$

to indicate that t is enabled by M and the firing of t produces M'.

More generally, the firing of a sequence of transitions $\sigma = \langle t_1, t_2, \ldots, t_k \rangle$ such that

$$M_0 = M$$

$$M_i \xrightarrow{t_{i+1}} M_{i+1}$$

$$M_k = M$$

will be denoted by

$$M \xrightarrow{\sigma} M'.$$

Such a σ is called a *firing sequence*.

DEFINITION 4. A marking M is said to be *reachable* from a marking M_0 iff there is a firing sequence σ such that

$$M_0 \xrightarrow{\sigma} M.$$

The set of all markings reachable from M_0 is

$$\mathfrak{R}(M_0).$$

The operation of firing a Petri net is a linear operation on a vector corresponding to the marking. This fact means that we can use all the machinery of linear algebra to deduce facts about Petri nets.

DEFINITION 5. Let $\Pi = \langle P, T, I, O \rangle$ be a Petri net. Let $P = \{p_1, \ldots, p_n\}$ and $T = \{t_1, \ldots, t_m\}$. The *incidence matrix* of Π is the $n \times m$ matrix

$$N_{ij} = O(p_i, t_j) - I(p_i, t_j).$$

It is now clear that if $t = t_j$ is a transition and is enabled by M then firing t produces the marking

$$\begin{aligned} M'(p_i) &= M(p_i) + N_{ij} \\ &= M(p_i) + (N\mathbf{e}_j)_i. \end{aligned}$$

Therefore,

$$M' = M + N\mathbf{e}_j$$

where $\mathbf{e}_j$ is the m-vector

$$\mathbf{e}_j = \langle 0 \quad \cdots \quad \underset{\substack{\uparrow \\ \text{position } j}}{1} \quad \cdots \quad 0 \rangle^t.$$

This means that if $\sigma = \langle t_{j_1}, \ldots, t_{j_k} \rangle$ is a firing sequence enabled by M then we get a *firing vector* $\mathbf{n}_\sigma$ defined inductively by $\mathbf{n}_0 = \langle 0, \ldots, 0 \rangle$ and

$$\mathbf{n}_{i+1} = \mathbf{n}_i + \mathbf{e}_{j_{i+1}}.$$

Therefore,

$$\begin{aligned} M_0 &= M \\ &= M + N\mathbf{n}_0 \\ M_{i+1} &= M_i + Ne_{j_{i+1}} \\ &= M_0 + N\mathbf{n}_i + N\mathbf{e}_{j_{i+1}} \\ &= M + N\mathbf{n}_{i+1} \end{aligned}$$

and so $M' = M + N\mathbf{n}_\sigma$. It is easily seen that the lth-component of $\mathbf{n}_\sigma$ is the same as the number of occurrences of t_{j_l} in σ.

As an application of these ideas, there are two associated invariants of a Petri net Π.

2.2. *S-invariants.*

DEFINITION 6. Let $\Pi = \langle P, T, I, O \rangle$ be a Petri net with incidence matrix N.

(a) An n-vector of rational numbers $\mathbf{v}$ is an *S-invariant* of Π iff

$$N^t\mathbf{v} = \mathbf{0}.$$

(b) An n-vector $\mathbf{v}$ is a *positive S-invariant* of Π iff $\mathbf{v}$ is an S-invariant and all entries in $\mathbf{v}$ are greater than or equal to 0.
(c) Two S-invariants $\mathbf{v}_1$ and $\mathbf{v}_2$ are *equivalent* iff there is some rational number q such that $q\mathbf{v}_1 = \mathbf{v}_2$.

We have used rational numbers to simplify the mathematics. It is clear that the set of all S-invariants is precisely the null space of N^t in $\mathbb{Q}^n$, and as such defines a canonical subspace of the vector space $\mathbb{Q}^n$. Therefore, any rational multiple or sum of S-invariants is also an S-invariant. Since any rational vector has many multiples with purely integer entries, we will ignore the fact that entries may not be integers.

PROPOSITION 1. *An n-vector* $\mathbf{v}$ *is an S-invariant of a Petri net* Π *iff for any marking* M_0 *and any* $M \in \mathfrak{R}(M_0)$ *we have the 'conservation law of fluxes' given by*

$$\mathbf{v}^t M = \mathbf{v}^t M_0.$$

Proof. Suppose that $\mathbf{v}$ is an S-invariant and $M \in \mathfrak{R}(M_0)$. Then we have $M = M_0 + N\mathbf{n}$ for some firing vector $\mathbf{n}$. Hence

$$\begin{aligned}\mathbf{v}^t M &= \mathbf{v}^t M_0 + \mathbf{v}^t N\mathbf{n}\\ &= \mathbf{v}^t M_0 + (N^t\mathbf{v})^t\mathbf{n}\\ &= \mathbf{v}^t M_0 + \mathbf{0}^t\mathbf{n}\\ &= \mathbf{v}^t M_0.\end{aligned}$$

Conversely, suppose that $\mathbf{v}^t M = \mathbf{v}^t M_0$ for all $M \in \mathfrak{R}(M_0)$ and any M_0. Then for each M_0 and $M \in \mathfrak{R}(M_0)$, we must have

$$(N^t\mathbf{v})^t\mathbf{n} = \mathbf{0} \tag{1}$$

Author: Please check the position of proof symbol.

for all firing vectors $\mathbf{n}$. Then, for each $t = t_j \in T$ we let $\lambda_{ij} > I(p_i, t_j)$. Let $M_{ij}(p_l) = \delta_{li}\lambda_{ij}$, where δ_{li} is the Kronecker delta. M_{ij} is a marking that enables t_j and so firing t_j with this marking gives a firing vector $\mathbf{n} = \mathbf{e}_j$. Therefore, [by (1)], we have $(N^t\mathbf{v})^t\mathbf{e}_j = \mathbf{0}$ for all j and so $N^t\mathbf{v} = \mathbf{0}$. □

This result gives us a conservation law in the following sense:

$$\mathbf{v}^t M = \sum_{i=1}^{n} v_i M(p_i),$$

so that a weighted sum of the number $(M_0(p_i))$ of *tokens* located at p_i and moved through the net by repeated firings never changes.

Clearly, this makes most sense when all of the $v_i \geq 0$ and are integers. In our examples, we note that we always expect flux-conservation, which is equivalent to the vector of all ones being an S-invariant. Another way to say this is that the sum of every column of the incidence matrix is zero.

An S-invariant lives only on a certain part of the net.

2.3. *Supports and signed supports.*

DEFINITION 7. Let $\mathbf{v}$ be an S-invariant for a Petri net Π. The *place-support* of $\mathbf{v}$ is the set of places

$$\mathrm{supp}_P(\mathbf{v}) = \{p_i \mid v_i > 0\}.$$

The *transition support* of $\mathbf{v}$ is

$$\mathrm{supp}_T(\mathbf{v}) = \left\{t \in T \;\middle|\; {}^\bullet t \cap \mathrm{supp}_P(\mathbf{v}) \neq \emptyset \text{ and } t^\bullet \cap \mathrm{supp}_P(\mathbf{v}) \neq \emptyset\right\}.$$

This set induces the $\mathbf{v}$-*component* of Π given by

$$\Pi \restriction_S \mathbf{v} = \left\langle \mathrm{supp}_P \mathbf{v}, \mathrm{supp}_T(\mathbf{v}), I \restriction \mathrm{supp}_P(\mathbf{v}) \times \mathrm{supp}_T(\mathbf{v}), O \restriction \mathrm{supp}_P(\mathbf{v}) \times \mathrm{supp}_T(\mathbf{v}) \right\rangle.$$

Note that $\Pi \restriction_S \mathbf{v}$ is also a Petri net.

DEFINITION 8. The place-support of an S-invariant $\mathbf{v}$ is *minimal* iff it is nonempty and does not contain the place-support of any nonequivalent S-invariant.

Minimal supports are useful in that we can recover all supports from the minimal ones.

Place-supports are completely determined by the signs of the S-invariants. We note that if $\mathbf{v}$ is an S-invariant, then so is $-\mathbf{v}$ and the place-supports of these two vectors together 'capture' the use of the invariant. For this reason, it is of some interest to reattach these two supports.

DEFINITION 9. Let $\mathbf{v}$ be an S-invariant. The *signed support* of $\mathbf{v}$ is the signed set

$$\mathrm{ssupp}(\mathbf{v}) = \langle \{i \mid v_i > 0\}, \{i \mid v_i < 0\} \rangle.$$

For technical convenience, we will also assume that $\langle \emptyset, \emptyset \rangle$ is also a signed support.

2.4. *T-invariants.*

DEFINITION 10. Let $\Pi = \langle P, T, I, O \rangle$ be a Petri net with incidence matrix N.

(a) An m-vector of rational numbers $\mathbf{w}$ is a *T-invariant* of Π iff

$$N\mathbf{w} = \mathbf{0}.$$

(b) An m-vector $\mathbf{w}$ is a *positive T-invariant* of Π iff $\mathbf{w}$ is a T-invariant and all entries in $\mathbf{w}$ are greater than or equal to 0.

(c) Two T-invariants $\mathbf{w}_1$ and $\mathbf{w}_2$ are *equivalent* iff there is some rational number q such that $q\mathbf{w}_1 = \mathbf{w}_2$.

REMARK 2.1. If $\Pi = \langle P, T, I, O\rangle$ is a Petri net, then we can form the dual net $\Pi' = \langle T, P, I', O'\rangle$ where $I'(t, p) = I(p, t)$ and $O'(t, p) = O(p, t)$. It is easy to see that the incidence matrix of Π' is the transpose of that for Π and that S-invariants of Π' are the T-invariants of Π. Thus, much of the theory of the two kinds of invariants is the same and we do not need to duplicate it.

As with the S-invariants, we have used rational numbers to simplify the mathematics. The set of all T-invariants is exactly the null space of N in $\mathbb{Q}^n$ and so is a subspace. Therefore, any rational multiple or sum of T-invariants is also a T-invariant.

T-invariants essentially identify those transitions, within the Petri net, which must fire to return the Petri net to an original state. Solutions determine regions of the net where it is possible for a place to lose a token and have it returned by some firing sequence of transitions.

DEFINITION 11. A marking M of a Petri net Π is *reproducible* iff there is some $M' \in \Re(M)$ such that $M \in \Re(M')$.

Notice that a marking M is reproducible iff there is a nontrivial firing sequence σ enabled by M such that $M \xrightarrow{\sigma} M$.

PROPOSITION 2. *A Petri net* Π *has a reproducible marking iff* Π *has a positive T-invariant.*

Proof. Let N be the incidence matrix of Π. Let $\boldsymbol{I} = [I(p_i, t_j)]$ and $\boldsymbol{O} = [O(p_i, t_j)]$ be the input and output matrices. Then $N = \boldsymbol{O} - \boldsymbol{I}$.

If $\mathbf{w}$ is any T-invariant then $N\mathbf{w} = \mathbf{0}$ iff $\boldsymbol{O}\mathbf{w} = \boldsymbol{I}\mathbf{w}$. Let M be a marking defined so that $M(p_i)$ is the ith entry of $\boldsymbol{O}\mathbf{w}$. Let σ be the firing sequence

$$\langle \underbrace{t_1, \ldots, t_1}_{w_1\text{repeats}}, \quad \underbrace{t_2, \ldots, t_2}_{w_2\text{repeats}}, \quad \ldots, \quad \underbrace{t_m, \ldots, t_m}_{w_m\text{repeats}} \rangle.$$

Then the firing vector $\mathbf{n}$ is equal to $\mathbf{w}$, so that $M' = M + N\mathbf{w} = M$. We just need to check that the appropriate transitions are enabled.

Let M_0 be the initial marking and M_i the result of firing the ith element of σ. We argue inductively as follows. M_0 enables t_1 as for all i $M_0(p_i)$ is $\sum_{j=1}^{m} O(p_i, t_j)w_j = \sum_{j=1}^{m} I(p_i, t_j)w_j \geq w_1 I(p_i, t_1)$. By definition, $M_1(p_i) = M_0(p_i) + O(p_i, t_1) - I(p_i, t_1)$.

Suppose we have the result for M_k. Let j be such that $\sum_{l=1}^{j} w_{j-1} \leq k < \sum_{l=1}^{j} w_l$ and $n_k = k - \sum_{l=1}^{j} w_{j-1}$. Then we are attempting to fire t_j. Let $\mathbf{w}_k = \sum_{l=0}^{j-1} w_l \mathbf{e}_l +$

$n_k e_j$. We have $M_k = M_0 + N\mathbf{w}_k$. We have $0 \leq n_k < w_j$. Now we have

$$\begin{aligned}
M_k(p_i) &= M_0(p_i) + \sum_{l=1}^{j-1} N_{il} w_l + N_{ij} n_k \\
&= \sum_{l=1}^{m} I(p_i, t_l) w_l + \sum_{l=1}^{j-1} (O(p_i, t_l) - I(p_i, t_l)) w_l \\
&\quad + (O(p_i, t_j) - I(p_i, t_j)) n_k \\
&= \sum_{l=1}^{j-1} O(p_i, t_l) w_l + \sum_{l=j+1}^{m} I(p_i, t_l) w_l + I(p_i, t_j) w_j \\
&\quad + (O(p_i, t_j) - I(p_i, t_j)) n_k \\
&\geq I(p_i, t_j) w_j + (O(p_i, t_j) - I(p_i, t_j)) n_k \\
&= O(p_i, t_j) n_k + I(p_i, t_j)(w_j - n_k) \\
&\geq I(p_i, t_j).
\end{aligned}$$

Conversely, suppose that σ is a nontrivial firing sequence enabled by M such that $M \xrightarrow{\sigma} M$. Then we have a firing vector $\mathbf{n}$ such that $M = M + N\mathbf{n}$ and so $N\mathbf{n} = \mathbf{0}$; $\mathbf{n}$ is clearly positive. □

A T-invariant also lives only on a certain part of the net.

DEFINITION 12. Let $\mathbf{w}$ be an T-invariant for a Petri net Π. The *transition support* of $\mathbf{w}$ is the set of transitions

$$\operatorname{supp}_T(\mathbf{w}) = \{t_i \in T \mid w_i > 0\}.$$

The *place-support* of $\mathbf{w}$ is

$$\operatorname{supp}_P(\mathbf{w}) = \left\{ p \in P \;\middle|\; {}^\bullet p \cap \operatorname{supp}_T(\mathbf{w}) \neq \emptyset \text{ and } p^\bullet \cap \operatorname{supp}_T(\mathbf{w}) \neq \emptyset \right\}.$$

This set induces the $\mathbf{w}$-*component* of Π given by

$$\Pi \restriction_T \mathbf{w} = \left\langle \operatorname{supp}_P \mathbf{w}, \operatorname{supp}_T(\mathbf{w}), I \restriction \operatorname{supp}_P(\mathbf{w}) \times \operatorname{supp}_T(\mathbf{w}), O \restriction \operatorname{supp}_P(\mathbf{w}) \times \operatorname{supp}_T(\mathbf{w}) \right\rangle.$$

Note that $\Pi \restriction_T \mathbf{w}$ is also a Petri net.

2.5. ***Hypergraphs, hyper-digraphs, graphs and digraphs.*** Petri nets are good models for the flow of objects or data through a network. Our approach to analyzing and understanding these models is primarily graph-theoretic. To do this, we first introduce a graph equivalent of a Petri net and then go on to consider the space of hyper-digraphs. Notably and by definition, all graphs are hypergraphs, but in all generality the converse is not true. For the purposes of this paper, every Petri net is a hypergraph, but not necessarily a graph. This fact can imply the need for graph augmentation of our Petri net model in order to take full advantage of the algebraic-combinatorial properties of graphs.

DEFINITION 13. A *hypergraph* is a pair $\mathscr{H} = \langle V, E\rangle$ where V is a nonempty set of *vertices* and $E \subseteq \wp(V)$ is the set of *edges*.

DEFINITION 14. A *directed hypergraph*, or *hyper-digraph*, is a pair $\mathscr{H} = \langle V, E\rangle$ where V is a nonempty set of *vertices* and $E\colon \|E\| \to {}^{V}\{-1, 0, 1\}$ is the set of *edges*.

In the definition above, we have chosen to have a set of edges indexed by a set $\|E\|$. In practice, only the range of the function E is important and so, by a slight abuse of notation, we often write $e \in E$ to mean that e is in the range of E.

Note that if $\mathscr{H}$ is a directed hypergraph, we can think of the set of edges as a family of signed sets where $E(x)$ corresponds to the signed set

$$\left\langle x^{+}, x^{-}\right\rangle = \langle \{v \mid E(x)(v) = 1\}, \{v \mid E(x)(v) = -1\}\rangle .$$

DEFINITION 15.

(a) Let $\mathscr{H} = \langle V, E\rangle$ be a hypergraph. Then the *incidence matrix* of $\mathscr{H}$ is the $V \times E$ matrix

$$\mathbb{I}(v, e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{otherwise.} \end{cases}$$

(b) Let $\mathscr{H} = \langle V, E\rangle$ be a directed hypergraph. Then the *incidence matrix* of $\mathscr{H}$ is the $V \times E$ matrix

$$\mathbb{I}(v, e) = e(v).$$

DEFINITION 16. A *graph* is a hypergraph $\mathscr{G} = \langle V, E\rangle$ such that $e \in E$ implies e has two elements.

DEFINITION 17. A *digraph* is a directed hypergraph $\mathscr{G} = \langle V, E\rangle$ such that $e \in E$ implies e^{+} and e^{-} both have one element.

Petri nets are equivalent to directed hypergraphs as follows.

DEFINITION 18. Let $\Pi = \langle P, T, I, O\rangle$ be a Petri net. We define the directed hypergraph $\mathscr{H}_P = \langle V, E\rangle$ by:

$$V = P$$

$$\|E\| = T$$

$$\text{if } t \in T \text{ then } E(t) = \langle \{p \mid O(p,t) = 1\}, \{p \mid I(p,t) = 1\}\rangle .$$

DEFINITION 19. Let $\mathscr{H} = \langle V, E\rangle$ be a directed hypergraph. We define the Petri net $\Pi_{\mathscr{H}}$ as follows:

$$P = V$$

$$T = \|E\|$$

$$O(p,t) = E(t)(v)$$

$$I(p,t) = -E(t)(v).$$

It is clear that these two operations are mutual inverses and so the study of Petri nets is equivalent to the study of hypergraphs.

DEFINITION 20. Let X be any set. Then

(a) a *signed subset* of X is a pair $S = \langle A, B\rangle$ where A and B are disjoint subsets of X. We write $S^+ = A$ for the *positive* part of S and $S^- = B$ for the *negative* part of S. $\underline{S} = A \cup B$ is the *support* of S;
(b) the *signed set family* of X is

$$\mathscr{S}(X) = \{\langle A, B\rangle \mid \langle A, B\rangle \text{ is a signed subset of } X\}.$$

DEFINITION 21. If $\langle A_1, A_2\rangle$ and $\langle B_1, B_2\rangle$ are two signed subsets of X, then

$$\langle A_1, A_2\rangle \leq \langle B_1, B_2\rangle \qquad \text{iff } B_1 \subseteq A_1 \qquad \text{and} \qquad B_2 \subseteq A_2.$$

There are a number of possible operations on signed subsets—see Metropolis and Rota (1978) and Bailey and Oliveira (1998) for a general view of such operations. In this context we need the following two operations.

DEFINITION 22. Let $A = \left\langle A^+, A^-\right\rangle$ and $B = \left\langle B^+, B^-\right\rangle$ be two signed subsets of X. Then the *composition* of A and B is

$$A \circ B = \left\langle A^+ \cup (B^+ \setminus A^-), A^- \cup (B^- \setminus A^+)\right\rangle.$$

The *negative* of A is $-A = \left\langle A^-, A^+\right\rangle$.

LEMMA 1. *For any two signed sets A and B, we have $\underline{(A \circ B)} = A \cup B$.*

Proof.

$$\begin{aligned}\underline{(A \circ B)} &= A^+ \cup (B^+ \setminus A^-) \cup A^- \cup (B^- \setminus A^+) \\ &= A^+ \cup A^- \cup B^+ \cup B^- \\ &= A \cup B.\end{aligned}$$

□

This operation is associative but not generally commutative. For a detailed study of this operation, the reader is referred to Bailey and Oliveira (1998, and in preparation)

2.6. ***Oriented matroids and cycles.*** Oriented matroids are a class of structures that generalize many of the properties of families of vectors—particularly, those properties coming from signs. There are a number of equivalent definitions depending on the viewpoint one wishes to emphasize—see Björner *et al.* (1993). The particular use we make of them is their finiteness and the fact that they capture all of the pathways we are interested in finding, as we will show.

DEFINITION 23. An *oriented matroid* is a collection $\mathcal{V}$ of signed subsets of a set X, which is a subset $\mathscr{S}(X)$ satisfying the following axioms:

(a) $\langle \emptyset, \emptyset \rangle \in \mathcal{V}$;
(b) if $A \in \mathcal{V}$ then $-A \in \mathcal{V}$;
(c) if A and B are in $\mathcal{V}$ then so is $A \circ B$;
(d) if A and B are in $\mathcal{V}$ and $e \in A^+ \cap B^-$ and $f \in (\underline{A} \setminus \underline{B}) \cup (\underline{B} \setminus \underline{A}) \cup (A^+ \cap B^+) \cup (A^- \cap B^-)$ then there is some $C \in \mathcal{V}$ such that

$$C^+ \subseteq (A^+ \cup B^+) \setminus e$$

$$C^- \subseteq (A^- \cup B^-) \setminus e$$

$$f \in \underline{C}.$$

The reader is referred to Björner *et al.* (1993) for a more detailed exposition of the theory of oriented matroids.

DEFINITION 24. Let $\mathscr{H} = \langle V, E \rangle$ be a hypergraph. A *cycle* in $\mathscr{H}$ is a sequence $v_0, e_0, \ldots, v_k, e_k$ such that

(a) $v_i \in V$ and $e_i \in E$ for all i;
(b) if $i < k$ then $\{v_i, v_{i+1}\} \subseteq e_i$;
(c) $\{v_0, v_k\} \subseteq e_k$.

DEFINITION 25. Let $\mathscr{H} = \langle V, E\rangle$ be a directed hypergraph.

(a) A *positively directed cycle* in $\mathscr{H}$ is a sequence $v_0, e_0, \ldots, v_k, e_k$ such that
   i. $v_i \in V$ and $e_i \in E$ for all i;
   ii. if $i < k$ then $e_i(v_i) = -1$, $e_i(v_{i+1}) = 1$;
   iii. $e_k(v_k) = -1$, $e_k(v_0) = 1$.

(b) A *negatively directed cycle* is the same but signs are reversed.

(c) A *cycle* or *undirected cycle* in $\mathscr{H}$ is a sequence $v_0, e_0, \ldots, v_k, e_k$ such that
   i. $v_i \in V$ and $e_i \in E$ for all i;
   ii. if $i < k$ then $e_i(v_i) \neq 0$, $e_i(v_{i+1}) \neq 0$;
   iii. $e_k(v_k) \neq 0$, $e_k(v_0) \neq 0$.

DEFINITION 26. A *tight* cycle for any type of hypergraph is a cycle without repeated edges.

We note that a cycle has a natural orientation given by the sequence order. With this in mind, we see that a positively directed cycle is one whose natural orientation agrees with the $\mathscr{H}$-orientation everywhere.

DEFINITION 27. Let $\mathscr{H}$ be a directed hypergraph. Let $\mathbb{I}$ be its incidence matrix.

(a) A *nullvector* of $\mathscr{H}$ is a vector $\mathbf{v} \in {}^{\|E\|}\mathbb{Q}$ such that

$$\mathbb{I}\mathbf{v} = \mathbf{0}.$$

(b) The *signed support* of a vector $\mathbf{v} \in {}^{n}\mathbb{Q}$ is

$$\operatorname{ssup}(\mathbf{v}) = \langle \{i \mid \mathbf{v}(i) > 0\}, \{i \mid \mathbf{v}(i) < 0\}\rangle .$$

(c) A *signed support* of $\mathscr{H}$ is a signed subset S of $\|E\|$ for which there is a nullvector $\mathbf{v} \in^{\|E\|} Q$ such that $S = \operatorname{ssup}(\mathbf{v})$.

(d) $\mathbb{S}(\mathscr{H})$ is the set of signed supports of $\mathscr{H}$.

THEOREM 1. *The set of signed supports of a directed hypergraph is an oriented matroid.*

Proof. Fix a directed hypergraph $\mathscr{H} = \langle V, E\rangle$. We check each of the axioms in turn. Let $\mathcal{P}$ be the set of all signed supports for $\mathscr{H}$.

(a) $\langle \emptyset, \emptyset\rangle \in \mathcal{P}$ since $\mathbf{0}$ is always a nullvector.

(b) If V is the signed support for $\mathbf{v}$ then $-V$ is the signed support for $-\mathbf{v}$.

(c) Let X be the signed support of $\mathbf{x}$ and Y be the signed support of $\mathbf{y}$. We want to find two rational numbers q and r such that if $\mathbf{z} = q\mathbf{x} + r\mathbf{y}$ then

$$z_i > 0 \qquad \text{iff } x_i > 0 \qquad \text{or} \qquad (y_i > 0 \text{ and } \neg(x_i < 0))$$

$$z_i < 0 \qquad \text{iff } x_i < 0 \qquad \text{or} \qquad (y_i < 0 \text{ and } \neg(x_i > 0)).$$

Thus, we want

$$qx_i + ry_i > 0 \qquad \text{iff } x_i > 0 \qquad \text{or} \qquad (y_i > 0 \text{ and } \neg(x_i < 0)) \qquad \text{(c.i)}$$

$$qx_i + ry_i < 0 \qquad \text{iff } x_i < 0 \qquad \text{or} \qquad (y_i < 0 \text{ and } \neg(x_i > 0)). \qquad \text{(c.ii)}$$

We will choose $q > 0$ and $r > 0$. We note that if $x_i = 0$ then z_i and y_i have the same sign so both of these are satisfied.
If $x_i > 0$, then we are using (c.i) so that $qx_i + ry_i > 0$ and, therefore, $\frac{q}{r} > -\frac{y_i}{x_i}$. Note that if $y_i > 0$ this is automatically true for $q > 0$ and $r > 0$, so we need this only for $y_i < 0$.
If $x_i < 0$, then we are using (c.ii) so that $qx_i + ry_i < 0$ and, therefore, $\frac{q}{r} > -\frac{y_i}{x_i}$. Note that if $y_i < 0$ this is automatically true for $q > 0$ and $r > 0$, so we need this only for $y_i > 0$.
Putting all this together, we see that we need merely choose strictly positive q and r to satisfy

$$\frac{q}{r} > \max\left\{ -\frac{y_i}{x_i} \;\middle|\; x_i \neq 0 \neq y_i \text{ and } x_i y_i < 0 \right\}.$$

This is easy to do as we are maximizing over a finite set of strictly positive rational numbers.

(d) Let X be the signed support of $\mathbf{x}$ and Y be the signed support of $\mathbf{y}$. We want to find two rational numbers q and r such that if $\mathbf{z} = q\mathbf{x} + r\mathbf{y}$ then

$$z_e = 0$$

$$z_f \neq 0$$

$$z_i > 0 \rightarrow x_i > 0 \qquad \text{or} \qquad y_i > 0$$

$$z_i < 0 \rightarrow x_i < 0 \qquad \text{or} \qquad y_i < 0.$$

The last two conditions are satisfied by choosing $q > 0$ and $r > 0$. The first condition implies

$$\frac{q}{r} = -\frac{y_e}{x_e}, \tag{4}$$

which is positive as $y_e < 0$ and $x_e > 0$.
Notice that f is such that either exactly one of x_f or y_f is nonzero, or the signs of x_f and y_f are the same.
If we choose any $q > 0$ and $r > 0$ satisfying (4), then $z_f = qx_f + ry_f$ is equal to zero iff both $x_f = 0 = y_f$ (which is impossible) or $x_f \neq 0 \neq y_f$ and $qx_f = -ry_f$. As $q > 0$ and $r > 0$, this implies x_f and y_f have different signs which is also impossible. □

REMARK 2.2. Thinking of a directed hypergraph as a Petri net, we see that this theorem says that the set of signed supports associated with T-invariants forms an oriented matroid. By dualizing, the same is true for the set of signed supports of S-invariants.

DEFINITION 28. Let $\mathscr{G} = \langle V, E \rangle$ be a digraph.

(a) If $C = \langle v_0, e_0, \ldots, v_k, e_k \rangle$ is a cycle for $\mathscr{G}$, define the functions C, C^+, C^- from E to $\mathbb{N}$ by

$$C(e) = \text{the number of times } e \text{ is on } C$$

$$C^+(e) = \text{the number of times } e \text{ is on } C \text{ positively}$$

$$C^-(e) = \text{the number of times } e \text{ is on } C \text{ negatively.}$$

(b) If $C = \langle v_0, e_0, \ldots, v_k, e_k \rangle$ is a cycle for $\mathscr{G}$, define the signed set

$$S(C) = \left\langle \left\{ e_i \mid C^+(e_i) > C^-(e_i) \right\}, \left\{ e_i \mid C^+(e_i) < C^-(e_i) \right\} \right\rangle.$$

Thus S^+ is that set of edges in C with positive orientation, and S^- is that set of edges in C with negative orientation.

(c) $\mathbb{O}(\mathscr{G}) = \{S(C) \mid C \text{ is a cycle of } \mathscr{G}\}$.

It is well-known that $\mathbb{O}(\mathscr{G})$ is an oriented matroid. Furthermore, we have the following relationship between the matroid of cycles and the matroid given by the incidence matrix.

THEOREM 2. *Let $\mathscr{G}$ be a digraph. Then*

$$\mathbb{S}(\mathscr{G}) = \mathbb{O}(\mathscr{G}).$$

This theorem fails for directed hypergraphs, but fortunately we are able to show that associated with any directed hypergraph there is a digraph with more edges and exactly the same cycles. This gives us a way to investigate the cycle matroid of our Petri net. Of course, the full matroid may be very big and so we only wish to find the *minimal elements*.

DEFINITION 29. Let $\mathscr{O}$ be an oriented matroid and $A = \langle A^+, A^- \rangle$ be in $\mathscr{O}$. Then A is *minimal* iff

(a) $A \neq \langle \emptyset, \emptyset \rangle$
(b) if $A \leq B$ then $A = B$ or $B = \langle \emptyset, \emptyset \rangle$.

This may seem the reverse of minimal, but the point is that minimal elements in the cycle matroid are really as small as possible.

2.7. ***Remarks.*** We note that the Petri net model introduced here provides us with an extensive set of combinatorial tools for deducing the qualitative control logic of biochemical networks. This approach defines states in the system to be marked places, that combinatorialists refer to as *boxes*; and the tokens, that are colored with markings that symbolically represent concentrations of biomolecular species such as metabolites, enzymes, and cofactors, etc., are called *colored balls*. The systematic nature of this modeling approach studies the circuit arrangements or partitions of a biochemical network as functions of marked balls (biochemical species) being arranged into marked places, subject to a set of process control rules defined by the transition conditionals of the Petri net. The tokens are symbolic representations of biomolecular concentrations.

Qualitative sets of inferences define the process control logic of a Petri net model of biochemical reaction networks. In every biochemical reaction network, there exist sets of reactions that define the conversion of biochemically reacting species. The replacement and depletion of biochemical species is defined by a set of reactions. The transport of sets of reacting species within a given network is directed by time-ordered sets of operational inferences, the process control logic of the systems' reaction network. We want to identify the extremal (maximal and minimal) sets of sub-circuit paths of a given network in terms of their exchange fluxes, subject to the conservation of system fluxes that act as the balance laws for any chemical reaction pathway. Not every sub-circuit will, in fact, be principal for the regulation of specific staged productions of metabolic species. In a sequel, we will demonstrate this by introducing the biochemical analog of Kirchhoff's law. Some sub-networks will be purely catalytic, and yet others will define maximal and minimal accumulations of metabolites and control pinch points which are analogous to set points in a control system and are referred to as governors.

The boundedness property of the Petri net representation identifies which combinations of paths have intermediate maximum and minimum accumulations. This is the case since we are enumerating multi-sets of molecular species into marked place holders subject to reaction and conservation constraints. The arrangements define a partitioning of kinetic reaction space that corresponds to the set of sub-circuits that span the network. In a future paper, we will develop oriented matroid programs from our oriented matroid of sub-circuit paths to identify minima, maxima, and cyclic metabolic accumulants. We will also, by this method, be able to identify 'pinch points' in the biochemical network.

Every linear system of chemical rate equations is subject to the conservation laws of system flux. Every linear system of chemically balanced equations can be written as an $m \times n$ conservation matrix S, where m corresponds to the number of reacting species in the chemical system and n corresponds to the number of reactions that are taking place within the chemical system.

For every system of biochemical reactions, the set of reactions that defines the system is referred to as a set of fluxes. In deriving a systems analysis model, it is necessary to define a boundary around the set of reactions. This boundary imposes a set of bounding conditions for physically discretizing the system of chemical equations and, as such, defines two classes of system fluxes: the internal set of system fluxes are defined by the set of internal sources and sinks that correspond to our sets of marked places in the Petri net model; and the external fluxes indicate the transport of the biochemical species, i.e., metabolites, that can exist outside of the bounded system as potential input sources to a new sub-system of biochemical reaction equations. Note that each of the external fluxes has a positive value.

If we construct a Petri net model for the system of reactions under consideration, we see that the incidence matrix of the net is exactly this conservation matrix. For every conservation matrix S, there exists a stoichiometric column matrix v, that defines the null space or kernel ($ker(S)$) of S. If S is an $m \times n$ matrix, the null space of S consists of all vectors v in real n-dimensional space, $\mathbb{Q}^n$, such that

$$S \cdot v = 0. \tag{5}$$

The dimension of the null space of S depends on the rank of S. The rank of S is defined as the number of free variables that exist in the system of equations that defines the matrix S.

If equation (5) defines a linear algebraic balance equation that represents the stoichiometric conservation of system flux in the linear system of biochemical reaction equations, then the matrix equation (5) is solved when we obtain the linear combination of linearly independent basis vectors $p_1, \ldots, p_n$ that spans the equational null space, $null(S)$ or $ker(S)$, of the conservation matrix S.

When the dimension of the null space is small, it is a simple exercise to compute the solution to equation (5). If the dimensionality of the system of linear conservation equations and hence the null space of equation (5) is large, then we have to use computational methods to calculate a basis for the null space. The computational hardness associated with computing all of the linearly independent sets of admissible solutions that span $null(S)$ grows by the dimension of S.

The null spaces of equation (5) contain all of the solutions that satisfy the balance equations defined by (5). The principal problem is to be able to optimally generate the entire null space and then search for all of the biochemically meaningful basis sets that span the null space as linear combinations of sub-circuit paths. As indicated above, this null space corresponds to the space of cycles of some graph, and so we take biochemically meaningful to mean a positive cycle or path. In signed sets terms, this means $C^- = \emptyset$.

Ideally, we want all minimal positive cycles. Often, in practice, we may be reduced to a subset of this. There are several ways to find these. The first two approaches look for a basis for the null space.

The first approach requires a reduction in the number of unknown fluxes that define the system's conservation matrix S. This approach leads to an exactly determined system of equations whose solution can be directly obtained as the dimension of the null is iteratively reduced to a subspace of zero-dimensional subspace that defines a point in $\mathbb{Q}^n$. This approach can be found in Smith and Missen (1982), Papoutsakis and Meyer (1985), Pons *et al.* (1996), and Henriksen *et al.* (1996).

The second approach generates extremal subset solutions by the direct application of a linear programming method. This method uses linear objective functions to generate constrained polyhedral bounds on the null space, and then generates a recursive search for the zero-dimensional subspace that satisfies the linear objective function of the search algorithm. The search for optimal linearly independent basis sets, that span the null space as sub-circuit paths, terminates with an optimal basis set that satisfies the conditions defined by a linear extremal objectifying function.

The method used in this paper is based on a combinatorial geometric analysis of cycles in graphs. A variation on this approach has been developed by Schilling and Palsson (1998, 1999) and Schilling *et al.* (1999). Ours is a variation of the first approach in that it first finds a suitable graph, generates the cycle matroid, and translates back to find positive cycles in the Petri net. In a sequel, we will show that this oriented matroid is used to define generalization of linear programming called an oriented matroid program (and its dual).

The computational construction of the null space of interest requires that a decision procedure exist that generates a set of signed vectors, those corresponding to the signed sets of tuples that define an incidence matrix S that characterizes the state representation of the system of equations, that in turn define the biochemical reaction system. The algorithm must efficiently generate a minimum number of linearly independent signed vectors that defines a candidate basis set as a signed sub-circuit. The optimal search of an n-set of points in $\mathbb{Q}^n$ space, that satisfies the dimensional requirements that the number of basis vectors necessary to define the basis set is equal to the dimensions of $null(S) \subseteq \mathbb{Q}^n$, is provably polynomial. This is not true for finding an optimal set of basis vectors over $\mathbb{R}^n$.

The S-invariants can also be used to identify the existence of all non-reachable sub-circuits in the reaction network. We note that the reachability of one marked place to another marked place in a Petri net is subject to the condition that all place markings are reachable when the conservation of fluxes is defined by the network S-invariants.

The T-invariants of the network determine the set of conditional transitions that have to be evaluated to identify the set of sub-circuits that span the entire reaction network as cycles. The T-invariants are defined by an equation that is identified with equation (5).

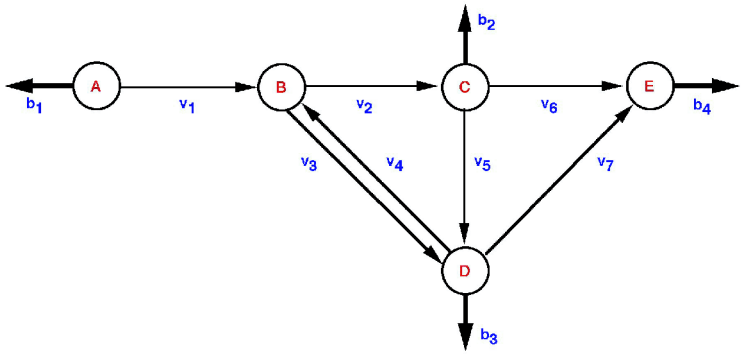


Figure 1. Schematic of a metabolic pathway, Schilling and Palsson (1998).

3. Analysis of Biochemical Reaction Network/Metabolic Pathway Models

The generation, storage and depletion of biomolecular species that define the discrete compartmentalized components in a biochemical reaction network are the corresponding colored balls or tokens that are arranged into marked boxes or places of our Petri net. The pathways that define the 'trajectories' of reacting species in a biochemical reaction network correspond to cycles or sub-network circuits of sets of reacting metabolites.

3.1. ***Model of a metabolic pathway.*** We now construct a combinatorial representation of the simple model metabolic network given by Schilling and Palsson (1998) as shown in Fig. 1. In their example, it is assumed that the system flow is initiated by a sufficient amount of metabolite A, as if there were an input into A.

A combinatorial Petri net of their network, presented in Fig. 2, can be shown to be a canonical generalization that serves as a representation of a biochemical reaction network composed of discrete sets of pathway sub-circuits. Their methodology produces a basis of the null space which implicitly defines the set of minimal cycles. One of the differences between our approach and other approaches is that the matroid explicitly produces all minimal cycles.

A metabolic pathway can be defined by a system of linear reaction equations that contains a set of metabolites. This set of metabolites undergoes sets of reactions that are equivalent to a combinatorial rearrangement of colored balls in marked boxes. A set of arrows indicates the direction of metabolic reactions and products. Together, they define a directed graphical network model of the system of coupled chemical equations.

We construct a partition of the network ensemble and examine the fluxes, i.e., reactions of species, coming into and going out of each block of sub-circuits. The disjoint union of these blocks of sub-circuits defines an equivalence class that de-

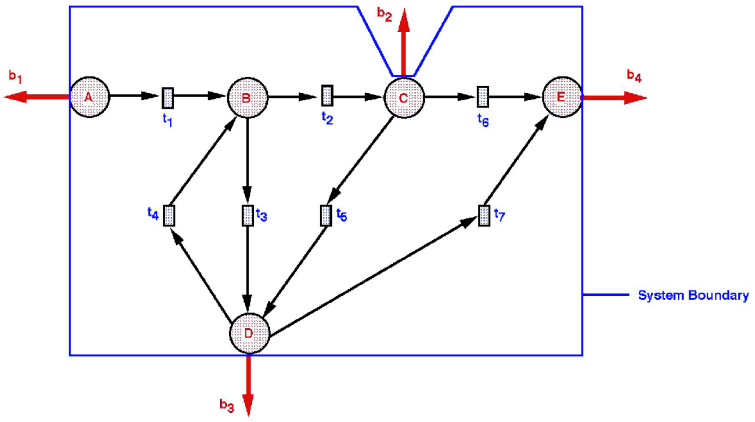


Figure 2. Petri net of the metabolic pathway.

fines the network partition. The biochemical reactions are referred to as internal fluxes defined by a set of vectors $\{v_1, v_2, \ldots, v_n\}$ such that v_n is the nth internal exchange flux. The fluxes that cross the system boundary of the network are precisely those metabolites that undergo transport into and out of the chemical reaction system. The sinks of the system are referred to as exchange fluxes defined by a set of vectors $\{b_1, b_2, \ldots, b_m\}$ such that the b_m is the mth total exchange flux. Note that each exchange flux has a positive value precisely when the metabolites are being transported out of the system.

The digraph of the reaction is transformed into a bounded biochemical reaction network matrix with both internal and exchange metabolic fluxes as follows: for each metabolite in the set of metabolites, define a stoichiometric conservation matrix S, as presented in Table 2a. From Fig. 1, we obtain a set of five metabolites $\{A, B, C, D, E\}$ and their corresponding 11 fluxes, such that there are four exchange fluxes $\{b_1, b_2, b_3, b_4\}$ and seven internal fluxes $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$; the dimensionality of the system corresponds to a 5×11 matrix. There exists a reversible reaction between metabolites B and D, known as a two-cycle. Recall that a two-cycle is a reversible reaction as presented in Table 1. All other metabolites contain either a source or a sink that defines the set of four exchange fluxes.

We next determine the dimension of the system's null space. Since the rank of the system is five, it follows that the dimension of the null space for the system defined is six. This fact follows from a well known theorem of linear algebra that says that $dim(S) = null(S) + rank(S)$. It is both necessary and sufficient to describe the $ker\, S$ of equation (5) via a set of linearly independent vector equations, such that this set of equations is defined in terms of the free parameter variables obtained from the biochemical balance equations that define this space, cf. Schilling and Palsson (1998). To accomplish this task, Schilling and Palsson use the following construction.

Let the set of metabolites $\{A, B, C, D, E\}$ be defined by the following balance

equations:

$$\begin{aligned}
A &: -v_1 - b_1 = 0\\
B &: v_1 + v_4 - v_2 - v_3 = 0\\
C &: v_2 - v_5 - v_6 - b_2 = 0\\
D &: v_3 + v_5 - v_4 - v_7 - b_3 = 0\\
E &: v_6 + v_7 - b_4 = 0
\end{aligned}$$

From equation (5), we obtain the stoichiometric matrix, S, presented in Table 2a. In this form, the stoichiometric matrix gives rise to very few cycles—in fact, only the two-cycle $\{B, D\}$ and the three-cycle $D \xrightarrow{t_4} B \xrightarrow{t_2} C \xrightarrow{t_5} D$ arise from positive vectors.

However, in investigating the chemical kinetics of the reaction pathways, we often assume that all or some of the reactions are potentially reversible or that some reactant is present in such quantities that we may assume it is always available. In Schilling and Palsson (1998), the authors made the latter assumption for reactant A, which modifies the stoichiometric matrix to S', presented in Table 2b.

The basis produced in Schilling and Palsson (1998) contains the two-cycle $\{B, D\}$ and the following six paths, based upon the assumption of sufficient metabolite A as a source input to state A—denoted as c_1:

$$\begin{array}{ll}
\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{b_2} & D \xrightarrow{t_4} B \xrightarrow{t_2} C \xrightarrow{t_5} D\\
\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{t_6} E \xrightarrow{b_4} & \xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_3} D \xrightarrow{b_3}\\
\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{t_5} D \xrightarrow{b_3} & \xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_3} D \xrightarrow{t_7} E \xrightarrow{b_4}
\end{array}$$

Although the matroid, and hence the set of all cycles is implicit in this basis, even determining the number of minimal cycles from the basis is a nontrivial problem. In this example our analysis produced one additional minimal cycle.

3.2. ***The augmented metabolic pathway.*** As our second example, we consider an augmented combinatorial representation of the model metabolic network defined by Schilling and Palsson (1998). The underlying assumption regarding the chemical kinetics of the reaction pathways is that they are all hypothetically reversible reactions. Therefore, the backward pathways have been added, regardless of whether or not they are readily attainable (as would be indicated via weights). The internal flux vector v_n of the original reaction model corresponds to the flux through transition $\boxed{t_1}$, for $n = 1, \ldots, 7$. Note that transition t_m for $m = 8, \ldots, 12$ has been added to complete the Petri net in our augmented example. The augmented reaction model and its Petri net is given in Fig. 3. The solid directed lines indicate Schilling and Palsson's internal fluxes; the dashed directed lines indicate the balancing backward pathways, which are biochemically potentially feasible.

Note that the metabolites $\{A, C, D, E\}$ each depict either a source or a sink indicating a set of four exchange fluxes. The sources of the system are referred to as

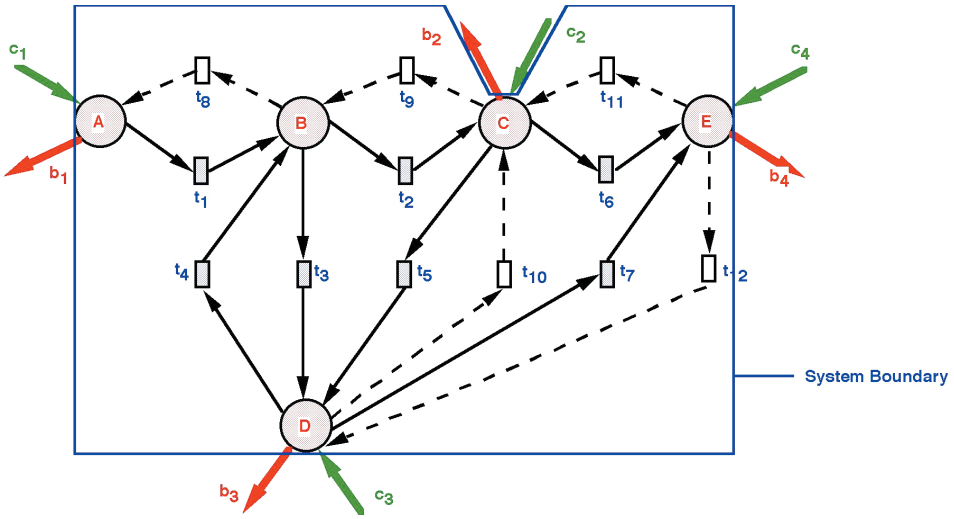


Figure 3. The augmented reversible Schilling and Palsson network example of signaling pathways.

augmented exchange fluxes defined by a set of vectors $\{c_1, c_2, \ldots, c_m\}$ such that c_m is the mth total augmented exchange flux. Note that each augmented exchange flux has a negative value precisely when the metabolites are being transported into the system.

Next note that the digraph presented is transformed into a bounded biochemical reaction network matrix with both internal and exchange metabolic fluxes as shown in Table 2c.

For computational efficiency, we added a single state to represent the space external to this network fragment, and assumed that all of the exchange fluxes connect to the external state. Because we have introduced this state, the cycles that contain the external state correspond to paths that go completely through the network.

The matrix for the original model describing the five metabolite states and seven internal fluxes is 5×7. Our augmented Petri net matrix is balanced, with rows summing to zero, at 5×12. Inclusion of the external fluxes results in matrix dimensions of 5×11 and 5×20, respectively. We next determine the dimension of the systems null space. Since the rank of the system is five, it follows that the dimension of the null space for the system defined is six.

We identified 50 cycles in the augmented model, 44 of which were multi-cycles and six of which were trivial two-cycles.

The six two-cycles are: $\{A, B\}$, $\{\mathbf{B}, \mathbf{D}\}$, $\{B, C\}$, $\{C, D\}$, $\{C, E\}$ and $\{D, E\}$. The single two-cycle identified by Schilling and Palsson is highlighted in bold.

The 44 minimal positive cycles in the space we obtain are provided in Fig. 4. The seven cycles from the first example are, of course, included, listed first.

T-invariants identify the cyclic firing sequences of sub-circuit metabolic processes. These cycles are identified with continuous sets of operations that define

the stability of the system. Here, we identify the minimal spanning set of cyclic circuit paths.

Our hypergraph approach identifies all possible positive paths. Note that all minimal positive paths pass through either state C and/or D, suggesting that they may be 'pinch points'.

3.3. ***A nonsimple example.*** In order to identify the cycle decomposition of a general network, it is necessary that the Petri net model be a graph. Accordingly, the hyper-digraph representation of the Petri net must be augmented in order to satisfy the stronger condition of being a graph. There is an easy repair to obtain a faithful graph representation of the hyper-digraph, as will be demonstrated presently.

In many practical examples, one of the most obvious difficulties which will be encountered is dealing with multiple paths to and/or from the transitions or states. Note that the molecular reaction presented in Table 1 is a simple two-cycle whereas the complex formation is a nonsimple example. As can be seen in the Petri net representation of a complex formation, transition $\boxed{t_1}$ contains the rule combining the flow of information from states $\boxed{p_1}$ and $\boxed{p_2}$ to $\boxed{p_3}$; similarly, transition $\boxed{t_2}$ contains the rule splitting the flow of information from state $\boxed{p_3}$ into states $\boxed{p_1}$ and $\boxed{p_2}$. In our present representation, in which we have assumed that all transitions are simply unary, the regulation cannot be specified.

The issue is that the Petri net, or hypergraph, representation of a nonsimple multiple-path biochemical reaction is not a graph. Therefore, a graph needs to be constructed in order to obtain all minimal paths. We now look at one such example involving nonsimple reactions, the enzyme reaction presented in Table 3a.

The incidence matrix of the Petri net representation yields an associated matroid with only six minimal elements corresponding to the three two-cycles: $\{t_1, t_2\}$, $\{t_3, t_4\}$ and $\{t_5, t_6\}$. Actually, the first and last of these each give rise to two distinct cycles. We are also missing two three-cycles. It is clear that this approach does not capture all of the information.

The readily constructed graph for the enzyme reaction is presented in Fig. 5, with its incidence matrix as presented in Table 3b. We have introduced a slight change in notation from that of the states $\boxed{p}$ to the vertices $\boxed{v}$ and from transitions $\boxed{t}$ to edges e such that $\boxed{p_4} \to \boxed{t_3} \to \boxed{p_5}$ is denoted by e_{10}. In this way, we may distinguish between multiple paths passing through a single transition. For example, the communication of information through transition $\boxed{t_1}$ is now split such that $p_4 \overset{t_1}{\to} p_3$ becomes e_2 and $p_4 \overset{t_1}{\to} p_1$ becomes e_7.

The complete list of correspondences between transitions and and digraph edges is provided in Table 3b.

Table 2a. Schilling and Palsson (1998) stoichiometric incidence matrix.

$$S = \underbrace{\left[\begin{array}{ccccccc} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right.}_{v_i = \text{internal fluxes}, i = 1\text{–}7} \Bigg| \underbrace{\left.\begin{array}{cccc} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right]}_{b_j = \text{exchange fluxes}, j = 1\text{–}4}$$

Table 2b. Modified stoichiometric incidence matrix of the Schilling and Palsson (1998) model.

$$S' = \underbrace{\left[\begin{array}{ccccccc} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right.}_{v_i = \text{internal fluxes}, i = 1\text{–}7} \Bigg| \underbrace{\left.\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right]}_{b_j = \text{exchange fluxes}, j = 1\text{–}4}$$

Table 2c. Augmented stoichiometric incidence matrix.

$$S^* = \underbrace{\left[\begin{array}{ccccccc} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right.}_{\substack{v_i = \text{internal fluxes}, \\ i = 1\text{–}7}} \Bigg| \underbrace{\left.\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & -1 \end{array}\right.}_{\substack{v_i = \text{augmented internal fluxes}, \\ i = 8\text{–}12}} \Bigg| \underbrace{\left.\begin{array}{cccc} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right.}_{\substack{b_j = \text{exchange fluxes}, \\ j = 1\text{–}4}} \Bigg| \underbrace{\left.\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]}_{\substack{c_k = \text{augmented exchange} \\ \text{fluxes}, k = 1\text{–}4}}$$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{b_2}$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{t_6} E \xrightarrow{b_4}$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{t_5} D \xrightarrow{b_3}$

$D \xrightarrow{t_4} B \xrightarrow{t_2} C \xrightarrow{t_5} D$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_3} D \xrightarrow{b_3}$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_3} D \xrightarrow{t_7} E \xrightarrow{b_4}$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{t_5} D \xrightarrow{t_7} E \xrightarrow{b_4}$

$\xrightarrow{c_2} C \xrightarrow{t_9} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_2} C \xrightarrow{t_5} D \xrightarrow{t_4} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_4} E \xrightarrow{t_{11}} C \xrightarrow{t_9} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_2} C \xrightarrow{t_6} E \xrightarrow{b_4}$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_3} D \xrightarrow{t_{10}} C \xrightarrow{t_6} E \xrightarrow{b_4}$

$\xrightarrow{c_2} C \xrightarrow{t_9} B \xrightarrow{t_3} D \xrightarrow{b_3}$

$\xrightarrow{c_3} D \xrightarrow{t_{10}} C \xrightarrow{t_9} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_2} C \xrightarrow{t_5} D \xrightarrow{b_3}$

$\xrightarrow{c_3} D \xrightarrow{t_4} B \xrightarrow{t_2} C \xrightarrow{t_6} E \xrightarrow{b_4}$

$\xrightarrow{c_3} D \xrightarrow{t_{10}} C \xrightarrow{t_6} E \xrightarrow{b_4}$

$\xrightarrow{c_2} C \xrightarrow{t_9} B \xrightarrow{t_3} D \xrightarrow{t_7} E \xrightarrow{b_4}$

$\xrightarrow{c_4} E \xrightarrow{t_{12}} D \xrightarrow{t_{10}} C \xrightarrow{t_9} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_2} C \xrightarrow{t_5} D \xrightarrow{t_7} E \xrightarrow{b_4}$

$D \xrightarrow{t_4} B \xrightarrow{t_2} C \xrightarrow{t_6} E \xrightarrow{t_{12}} D$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_3} D \xrightarrow{t_7} E \xrightarrow{t_{11}} C \xrightarrow{b_2}$

$D \xrightarrow{t_{10}} C \xrightarrow{t_6} E \xrightarrow{t_{12}} D$

$\xrightarrow{c_3} D \xrightarrow{t_7} E \xrightarrow{b_4}$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_2} C \xrightarrow{t_6} E \xrightarrow{t_{12}} D \xrightarrow{b_3}$

$\xrightarrow{c_2} C \xrightarrow{t_6} E \xrightarrow{t_{12}} D \xrightarrow{b_3}$

$D \xrightarrow{t_{10}} C \xrightarrow{t_9} B \xrightarrow{t_3} D$

$\xrightarrow{c_1} A \xrightarrow{t_1} B \xrightarrow{t_3} D \xrightarrow{t_{10}} C \xrightarrow{b_2}$

$\xrightarrow{c_4} E \xrightarrow{t_{11}} C \xrightarrow{b_2}$

$\xrightarrow{c_4} E \xrightarrow{t_{11}} C \xrightarrow{t_5} D \xrightarrow{t_4} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_3} D \xrightarrow{t_4} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_3} D \xrightarrow{t_4} B \xrightarrow{t_2} C \xrightarrow{b_2}$

$\xrightarrow{c_3} D \xrightarrow{t_{10}} C \xrightarrow{b_2}$

$\xrightarrow{c_4} E \xrightarrow{t_{11}} C \xrightarrow{t_9} B \xrightarrow{t_3} D \xrightarrow{b_3}$

$\xrightarrow{c_4} E \xrightarrow{t_{11}} C \xrightarrow{t_5} D \xrightarrow{b_3}$

$\xrightarrow{c_4} E \xrightarrow{t_{12}} D \xrightarrow{t_4} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_4} E \xrightarrow{t_{12}} D \xrightarrow{t_4} B \xrightarrow{t_2} C \xrightarrow{b_2}$

$\xrightarrow{c_4} E \xrightarrow{t_{12}} D \xrightarrow{t_{10}} C \xrightarrow{b_2}$

$D \xrightarrow{t_7} E \xrightarrow{t_{11}} C \xrightarrow{t_9} B \xrightarrow{t_3} D$

$\xrightarrow{c_2} C \xrightarrow{t_6} E \xrightarrow{t_{12}} D \xrightarrow{t_4} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$D \xrightarrow{t_7} E \xrightarrow{t_{11}} C \xrightarrow{t_5} D$

$\xrightarrow{c_4} E \xrightarrow{t_{12}} D \xrightarrow{b_3}$

$\xrightarrow{c_3} D \xrightarrow{t_7} E \xrightarrow{t_{11}} C \xrightarrow{t_9} B \xrightarrow{t_8} A \xrightarrow{b_1}$

$\xrightarrow{c_3} D \xrightarrow{t_7} E \xrightarrow{t_{11}} C \xrightarrow{b_2}$

Figure 4. The augmented pathways.

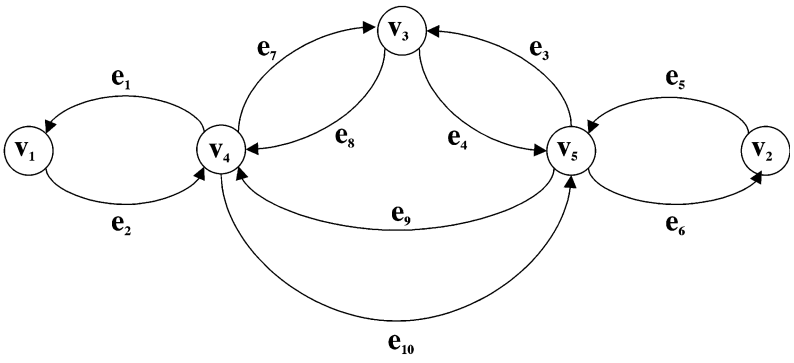


Figure 5. The enzyme reaction digraph.

Table 3a. An enzyme reaction is composed of one molecular reaction and two complex formations.

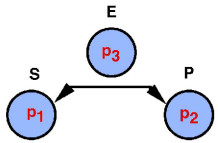
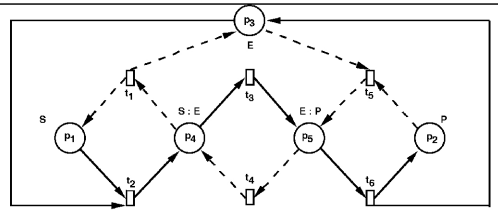
Reaction Type	Chemistry Representation	Reaction Representation	Petri Net Representation	Incidence Matrix Representation
Enzyme Reaction	$S + E \leftrightarrow S : E$ $S : E \leftrightarrow P : E$ $P : E \leftrightarrow P + E$			$\begin{array}{c\|cccccc} & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\ \hline p_1 & 1 & -1 & & & & \\ p_2 & & & & & -1 & 1 \\ p_3 & 1 & -1 & & & -1 & 1 \\ p_4 & -1 & 1 & -1 & 1 & & \\ p_5 & & & 1 & -1 & 1 & -1 \end{array}$

Table 3b. Incidence matrix for the digraph representation of the enzyme reaction.

	Edge index	1	2	3	4	5	6	7	8	9	10
	connectivity	$e_1[4,1]$	$e_2[1,4]$	$e_3[3,5]$	$e_4[5,3]$	$e_5[2,5]$	$e_6[5,2]$	$e_7[4,3]$	$e_8[3,4]$	$e_9[5,4]$	$e_{10}[4,5]$
	transition	t_1	t_2	t_5	t_6	t_5	t_6	t_1	t_2	t_4	t_3
State	Vertex										
p_1	v_1	1	−1	0	0	0	0	0	0	0	0
p_2	v_2	0	0	0	0	−1	1	0	0	0	0
p_3	v_3	0	0	1	−1	0	0	1	−1	0	0
p_4	v_4	−1	1	0	0	0	1	−1	1	1	−1
p_5	v_5	0	0	−1	1	1	−1	0	0	−1	1

$$p_3 \xrightarrow{t_5} p_5 \xrightarrow{t_4} p_4 \xrightarrow{t_1} p_3 \qquad p_3 \xrightarrow{t_2} p_4 \xrightarrow{t_3} p_5 \xrightarrow{t_6} p_3$$
$$p_3 \xrightarrow{t_2} p_4 \xrightarrow{t_1} p_3 \qquad p_1 \xrightarrow{t_2} p_4 \xrightarrow{t_1} p_1$$
$$p_4 \xrightarrow{t_3} p_5 \xrightarrow{t_4} p_4 \qquad p_3 \xrightarrow{t_5} p_5 \xrightarrow{t_6} p_3$$
$$p_2 \xrightarrow{t_5} p_5 \xrightarrow{t_6} p_2$$

Figure 6. The pathways of the augmented enzyme reaction.

The paths we found are presented in Fig. 6.

4. CONCLUSIONS

We selected as a generic example of the linear algebra approach one studied by Schilling and Palsson (1998). The biochemically significant path-circuit information is obtained from a vector space point of view. The biochemical reaction network defined by Fig. 1 is characterized by the matrix representation given in Table 2a. The flux balance matrix S has as its first seven entries a vector $\mathbf{v_i}$ that corresponds to the internal system fluxes, and a set of four vectors $\mathbf{b_j}$ that correspond to the external system fluxes. The incidence matrix S is 5×11. This indicates that the space of all possible basis vectors that must be searched through is 11-dimensional. This follows from a theorem of linear algebra that states that the rank of a matrix plus the nullity of the matrix is equal to the dimension of the vector space that matrix spans as a subspace. Without a significant set of objectifying biochemical constraints or knowing *a priori* which basis transformations are required to obtain the desired sets of optimal sub-circuit paths, searching an n-dimensional space (for $n \geq 3$) for the correct set of basis vectors that produces solutions to the flux balance matrix equation, that in turn generates the flux conserved sub-circuits that span the network over $\mathbb{R}^n$, is worse than NP-complete, cf. Garey and Johnson (1979).

Use of linear programming leads to a choice of constraints driven by heuristic considerations. This fact leads to a decision procedure for setting constraints which is itself NP-complete. In other words, as the dimension of the matrix grows arbitrarily large, so does the computational hardness associated with obtaining the necessary and sufficient basis transformation required to obtain all of the biologically feasible optimal sub-circuit paths. This fact, together with the NP-completeness of finding the basis vectors that optimally span the null space, generates an NP-complexity for finding spanning combinations of biochemically feasible sub-circuits as a function of identifying suitable sets of positive basis vectors. Worst case, it is a $(NP)^{NP}$ PSPACE hard decision problem. If, on the other hand, we restrict ourselves to searching for cycles in a hyper-digraph representation of the biochemical network, then we can reduce the PSPACE complexity that the lin-

ear algebra approach introduces into the search problem. Our decision problem is a bounded satisfiability problem (SAT), which is just NP complete.

Identifying the theoretically feasible pathways is accomplished by inspecting the components of the basis vectors that trace out the sub-circuit pathways. The components of each vector describe the relative flux distribution of each of the reactions in a given network pathway. In fact, only those vector components that are positive are wanted, since any negative vector component points to the existence of a biochemically impossible pathway.

S- and T-invariants of our Petri net representations of cell signaling networks are the first of many algebraic-combinatorial invariants of multi-set arrangements of marked balls placed into the holding boxes called reaction states. As such, the identification of the set of S- and T-invariants in this combinatorial model corresponds to a set of chemical kinetic system conservation laws, e.g., system fluxes. This identification yields a set of fixed points for a specific path algebra model over which path optimizations, cast in the framework of oriented matroids, are well-defined by a set of objectifying functions. The set of S- and T- fixed-point invariants and their supports are used to fix the objective in the optimization metrics over the given path algebra. In subsequent papers, we will use these results to obtain examples of optimal kinetic paths corresponding to optimally weighted edge sets in the multi-digraph representation of the signaling pathway.

Notably, the success of the linear algebraic approach depends largely on finding the necessary and sufficient sets of basis transformations that result in an identification of sub-circuit paths whose flow is indicated by sets of exchange fluxes where the negative values represent reactants and positive values represent products in stoichiometric ratios given by their respective numerical values. Given exact stoichiometry, the Petri net approach can easily incorporate this additional set of constraints. Every flux distribution in a given network is easily represented as a linear combination sub-circuit pathway traced by some new set of basis vectors. If these basis vectors represent specific metabolic functions, either experimentally or computationally determined, then one can obtain new metabolic sub-circuit pathways as a function of linear combinations of functional metabolic sub-circuit paths. These issues will be explored in a sequel to this paper.

To enhance our method, we can consider Petri nets with weighted edges, where the weights come from a path algebra. This allows different optimality criteria to be applied in the search for paths and cycles as discussed by Bailey and Oliveira (1998). So far, the use of oriented matroids has been quite limited. However, they are a powerful tool, giving rise to a number of significantly different viewpoints on the space of cycles. They also allow the use of matroid programming search algorithms similar to those used in linear programming. One possible application of our Petri net hyper-digraph model technique is the anlaysis of genetic networks obtained from gene expression array data. In later work, we will consider these applications and extensions.

ACKNOWLEDGEMENTS

This research was supported by the Laboratory Directed Research and Development program office and performed in the William R Wiley Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory. The EMSL is funded by the Office of Biological and Environmental Research in the U.S. Department of Energy under Contract DE-AC06-76RLO 1830.

REFERENCES

Alberty, R. A. (1992). Conversion of chemical equations to biochemical equations. *J. Chem. Ed.* **69**, 493.

Alberty, R. A. (1994). Review of biochemical thermodynamics. *Biochim. Biophys. Acta* **1207**, 1–11.

Alberty, R. A. (1996). Calculation of biochemical net reactions and pathways by using matrix operations. *Biophys. J.* **71**, 507–515.

Alberty, R. A. (1991a). Chemical equations are actually matrix equations. *J. Chem. Ed.* **68**, 984.

Alberty, R. A. (1991b). Equilibrium compositions of solutions of biochemical species and heats of biochemical reactions. *Proc. Natl Acad. Sci. USA* **88**, 3268–3271.

Arkin, A., J. Ross and H. H. McAdams (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells. *Genetics* **149**, 1633–1648.

Bailey, C. G. and J. S. Oliveira (1998). An axiomatization for cubic algebras, in *Mathematical Essays in Honor of Gian-Carlo Rota*, B. E. Sagan and R. P. Stanley (Eds), Birkhaüser.

Bailey, C. G. and J. S. Oliveira. *Another Universal Axiomatization of Cubes* (in preparation).

Bailey, C. G. and J. S. Oliveira (2001). *MV-Algebras and Cubic Algebras*, submitted to the Annals of Combinatorics.

Berge, C. (1973). *Graphs and Hypergraphs*, New York: American Elsevier Publishing Company.

Björner, A., M. Las Vergnas, B. Sturmfels, N. White and G. Ziegler (1993). *Oriented Matroids*, CUP.

Clark, B. L. (1988). Stoichiometric network analysis. *Cell Biophys.* **12**, 237–253.

Diestel, R. (2000). *Graphs Theory*, 2nd edn, New York: Springer.

Garey, M. R. and D. S. Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Murray Hill, NJ/ San Francisco, CA: Bell Labs./ W. H. Freeman and Co.

Henriksen, C. M., L. H. Christensen, J. Nielsen and J. Villadsen (1996). *J. Biotechnol.* **45**, 149–164.

Karp, P. D. *et al.* (1998). EcoCyc: encyclopedia of Escherichia coli genes and metabolism. *Nucleic Acids Res.* **26**, 43–45.

Author:
Please provide article title

Kauffman, S. A. (1971a). Gene regulation networks: a theory for their global structure and behavior. *Curr. Top. Dev. Biol.* **6**, 145.

Kauffman, S. A. (1993). *The Origins of Order: Self-organization and Selection in Evolution*, New York, NY: Oxford University Press.

Kohn, M. C. and W. J. Letzkus (1983). A graph-theoretical analysis of metabolic regulation. *J. Theor. Biol.* **100**, 293–304.

Kohn, M. C. and D. R. Lemieux (1991). Identification of regulatory properties of metabolic networks by graph theory. *J. Theor. Biol.* **150**, 3–25.

Mavrovouniotis, M. L. and G. Stephanopoulos (1990). Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.* **36**, 1119–1132.

McAdams, H. H. and A. Arkin (1997). Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA, Biochem.* **94**, 814–819.

Metropolis, N. and G-C. Rota (1978). Combinatorial structure of the faces of the n-cube. *SIAM J. Appl. Math.* **35**, 689–694.

Oster, G. S., A. S. Perelson and A. Katchalsky (1973). Network thermodynamics: dynamic modelling of biophysical systems. *Q Rev of Biophys* **6**, 1–134.

Author: Please provide article title

Papoutsakis, E. and C. Meyer (1985). *Biotechnol. Bioeng.* **27**, 50–66.

Peusner, L. (1986). *Studies in Network Thermodynamics*, Vol. 5, Studies in Modern Thermodynamics, Elsevier.

Author: Please provide Publisher's location

Pons, A., C. Dussap, C. Pequignot and J. Gros (1996). *Biotechnol. Bioeng.* **51**, 177–189.

Author: Please provide article title

Reddy, V. N., M. N. Liebman and M. L. Mavrovouniotis (1996). Qualitative analysis of biochemical reaction systems. *Comput. Biol. Med.* **26**, 9–24.

Reisig, W. (1985). *Petri Nets, An Introduction*, W. Brauer, G. Rozenberg and A. Salomaa (Eds), New York, NY: Springer-Verlag.

Samoilov, M., A. Arkin and J. Ross (2001). On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Am. Inst. Phys.* **11**, 108–114.

Schilling, C. H. and B. O. Palsson (1998). The underlying pathway structure of biochemical networks. *Proc. Natl. Acad. Sci., USA* **95**, 4193–4198.

Schilling, C. H. and B. O. Palsson (1999). Towards metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog., Am. Chem. Soc. and Am. Inst. Chem. Engr.* **15**, 288–295.

Schilling, C. H., S. Schuster, B. O. Palsson and R. Heinrichn (1999). Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog., Am. Chem. Soc. and Am. Inst. Chem. Engr.* **15**, 296–303.

Schnakenberg, J. (1979). Simple chemical reaction systems with limit cycle behavior. *J. Theor. Biol.* **81**, 389–400.

Seressiotis, A. and J. E. Bailey (1988). MPS: an artificially intelligent software system for the analysis and synthesis of metabolic pathways. *Biotechnol. Bioeng.* **31**, 587–602.

Smith, W. R. and R. W. Missen (1982). *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*, New York: Wiley.

Received 7 May 2000 and accepted 7 August 2000