

## Reaction Route Graphs. I. Theory and Algorithm

Ilie Fishtik, Caitlin A. Callaghan, and Ravindra Datta\*

Fuel Cell Center, Department of Chemical Engineering, Worcester Polytechnic Institute,  
Worcester, Massachusetts 01609

Received: November 7, 2003; In Final Form: February 11, 2004

A theory and algorithm for reaction route (RR) network analysis is developed in analogy with electrical networks and is based on the combined use of RR theory, graph theory, and Kirchhoff's laws. The result is a powerful new approach of "RR graphs" that is useful in not only topological representation of complex reactions and mechanisms but, when combined with techniques of electrical network analysis, is able to provide revealing insights into the mechanism as well as the kinetics of the overall reactions involving multiple elementary reaction steps including the effect of topological constraints. Unlike existing graph theory approaches of reaction networks, the approach developed here is suitable for linear as well as nonlinear kinetic mechanisms and for single and multiple overall reactions. The theoretical approach for the case of a single overall reaction involving minimal kinetic mechanisms (unit stoichiometric numbers) is developed in Part I of this series followed by its application to examples of heterogeneous and enzyme catalytic reactions in Part II.

### 1. Introduction

Reaction schematics of one kind or another are universally employed to depict reaction pathways in chemistry and cellular biology, and are invaluable in the study of reaction mechanisms. Typically, species, often showing molecular structure, are drawn and interconnected via arrows to show the reactions. Such a scheme, while well-suited for monomolecular reactions, becomes complicated when more than one species is involved in a reaction and especially when there are parallel pathways.

The term "reaction graphs" in the literature alludes to the topology of reaction mechanisms and derives from the fact that reaction schematics are structurally similar to graphs, with nodes or vertices denoting species, and branches or edges representing their reactions.<sup>1,2</sup> The graph theoretical viewpoint of mechanisms, besides its intuitive appeal, allows the results of graph theory to be used for describing reaction topology, and is especially useful for computer-assisted enumeration.<sup>3</sup> The main use of chemical graph theory has been in the study of molecular structure;<sup>4</sup> however, it is being increasingly utilized in the elucidation<sup>1,2</sup> and generation of reaction mechanisms.<sup>5–8</sup>

Christiansen<sup>9</sup> proposed the use of reaction diagrams for open (noncatalytic) or closed (catalytic) sequences by depicting reaction intermediates at nodes connected by branches. For linear kinetic mechanisms, i.e., those comprising elementary reactions that are first-order with respect to the intermediates, in both the forward and the reverse directions, he derived rate expressions for sequential steps at steady state. King and Altman<sup>10</sup> utilized the approach to derive the kinetics of linear enzyme-catalyzed reactions. M. I. Temkin<sup>11–13</sup> further developed Christiansen's approach to analyze the kinetics of steady-state sequential and parallel catalytic reactions. Yatsimirskii<sup>14</sup> has discussed the application of graph theory to catalytic kinetics. Balaban et al.<sup>1,15–16</sup> proposed "reaction graphs" for studying the topology of isomerization reactions. These early studies were followed by numerous other graph-theoretic studies, as sum-

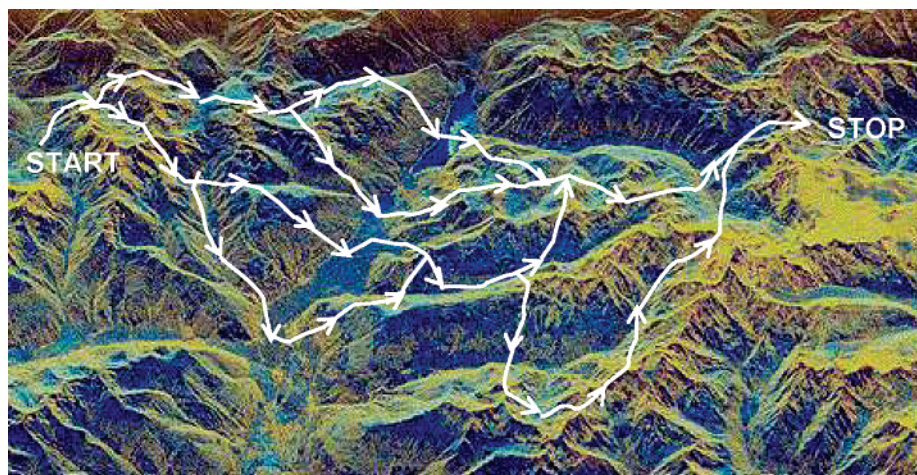
marized in the book by O. N. Temkin et al.<sup>2</sup> More recently, these methods have been applied to metabolic networks.<sup>17–21</sup>

Unfortunately, the universal practice of depicting reaction intermediates individually as vertices or nodes connected by branches or edges restricts their application to linear kinetic mechanisms. Temkin et al.<sup>2</sup> proposed the use of bipartite graphs for nonlinear kinetic mechanisms, i.e., those with elementary reactions with more than one intermediate on either side of the reaction. In their approach, one set of vertices on one side of the graph represents intermediates, while another set on the opposite side represents terminal species. A third set of markers in the middle denotes elementary reaction steps. Arrows connect various species and reactions and their directions show whether species are being consumed or produced, with double arrows denoting nonunit stoichiometric coefficients. This, unfortunately, results in jumbled graphs that are devoid of intuitive appeal, even for simple reactions.

Oster et al.<sup>22</sup> and Oster and Perelson<sup>23</sup> utilized the analogy between reaction networks and electrical networks in their development of graph-theory-based "network thermodynamics." They represented the topological structure of a reaction network as a directed graph, with species representing tree branches and links representing the reactions interconnecting the species. The forward and reverse steps were represented separately by two links. The nonunit stoichiometric coefficient was represented by an ideal transformer. Kirchhoff's laws of current and potential were applied to the resulting network, with reaction affinity corresponding to potential and reaction rate corresponding to current. However, their chemical reaction network representation is particularly cumbersome, resulting in complicated networks, even for simple reaction systems. Surprisingly, few other researchers have quantitatively utilized the evident analogy between reaction systems and electrical circuits;<sup>24</sup> however, the results are either similarly cumbersome<sup>25</sup> or only simple monomolecular examples have been treated.<sup>21</sup>

We develop a new graph-theoretical approach here that overcomes many of the limitations of the current methodologies and may be a powerful tool in the graphical depiction, as well

\* Author to whom correspondence should be addressed. E-mail address: rdatta@wpi.edu.



**Figure 1.** Reaction route (RR) network, presented conceptually as a mountain trek between two camps.

as mechanistic and kinetic interpretation, of reaction networks based on reaction route theory and their analogy with electrical circuits. These “reaction route graphs” are distinct from “reaction graphs” in that the nodes do not represent a given species but, rather, simply show how the elementary steps, or graph branches, are connected to depict the various reaction routes. Thus, there is a direct analogy to conventional graphs. The approach has broad applicability including elementary reactions of arbitrary stoichiometry and multiple overall reactions. However, here, we will concern ourselves with only a single overall reaction that comprises a minimal kinetic mechanism. Subsequent work will address their generalization.

With the advent of quantum mechanical calculations of catalytic molecular events and their energetics,<sup>26</sup> along with the availability of powerful semi-theoretical methods,<sup>27</sup> such reaction pathway analyses will become increasingly indispensable. It is only a matter of time before we have a very fundamental understanding of the molecular steps involved in important catalytic processes along with their reliable kinetics. However, this is only the first step toward unraveling the mechanism and kinetics of the overall catalytic reactions of interest. The predicted steps must be organized into a coherent mechanism that illustrates the kinetics of the overall reaction. The objective of this work is to develop such a framework.

There are currently available two alternate ways in which one might use the kinetic information on elementary reaction steps: (i) the conventional Langmuir–Hinshelwood–Hougen–Watson (LHHW) approach,<sup>28,29</sup> in which an explicit rate expression might be derived based on the universal, but arbitrary, assumptions such as the rate-determining step (RDS), quasi-steady-state (QSS) approximation, and the most abundant reactive intermediate (MARI), and (ii) the so-called microkinetic approach,<sup>30,31</sup> wherein no assumptions are made, but only numerical analysis of the resulting large system of ordinary or partial differential equations of intermediate and terminal species material balances for a given reactor configuration is possible. Although it does involve numerical solution, much information about the controlling steps and the reaction network can be discerned from microkinetic models.

Here, we propose a complimentary approach<sup>32</sup> to that of microkinetic analysis. More specifically, we begin with a microkinetic analysis based on a priori predicted elementary reaction kinetics, but then utilize a systematic graph-theoretic approach in conjunction with reaction route theory and an analogy to electrical networks to elucidate the major pathways followed by systematic reduction of the network to arrive at

simpler mechanisms including, when possible, precise rate expressions involving predicted rate constants. The reaction route approach that we previously developed<sup>33–36</sup> is a generalization of the conventional reaction route theory<sup>13,37–39</sup> and forms the basis of our work here.

The theoretical approach developed in Part I is illustrated with application to heterogeneous and enzyme catalyzed reactions in Part II of this series.<sup>40</sup> In fact, this approach may also be utilized for the analysis of noncatalytic reactions as well as the functioning of the cellular metabolic machinery, topics to be subsequently addressed.

## 2. Reaction Routes and Stoichiometry

As a prelude to the description of our approach to reaction route graphs, it is useful to imagine a reaction route network as a trek across a mountain range with many peaks and valleys that correspond to individual elementary reaction steps (Figure 1). In keeping with the transition-state theory, the valleys are viewed conceptually as the energy level of reactants or products, whereas a single elementary reaction is viewed as the hike from one valley to an adjacent one over a mountain pass, representative of the energy barrier for a unit reaction. Many such excursions constitute the overall trek and correspond to the reaction route network for an overall reaction. Clearly, different hikers embarking on the trek would follow different routes. In fact, a lost hiker might go around in a loop or a cycle before making forward progress. Thus, an infinite variety of reaction routes (RRs) exists, in principle, for a given trek in going from the reactants to the products, including those involving loops or cycles, which are called empty routes (ERs). The direct RRs are defined as those that are the shortest, not involving any cycles, and are finite in number, although this may still be large. Because there is a natural tendency to do things with the greatest economy, given a large number of hikes, the fraction following a particular route would be inversely related to its difficulty. With this picture in mind, we consider some definitions first, followed by further details.

We consider a set of  $p$  elementary reaction steps  $s_\rho$  ( $\rho = 1, 2, \dots, p$ ):  $\sum_{i=1}^l \nu_{\rho i} B_i = 0$ , involving species  $B_i$  ( $i = 1, 2, \dots, l$ ). The stoichiometric coefficient  $\nu_{\rho i}$  of species  $i$  in the reaction  $s_\rho$  is, by convention, positive for a product, negative for a reactant, and zero for an inert. Alternately, the reaction may be written in a more conventional format as

$$s_\rho: \sum_{i=1}^l (-\bar{\nu}_{\rho i}) B_i \rightleftharpoons \sum_{i=1}^l \bar{\nu}_{\rho i} B_i \quad (\text{for } \rho = 1, 2, \dots, p) \quad (1)$$

where the stoichiometric coefficients of the reactants and products are differentiated by the crowning arrows. All reactions are, of course, considered to be reversible. The degree of reversibility and the direction of reaction flux of each reaction is determined by the sign and magnitude of its affinity,<sup>41</sup> which is a state function characteristic of the reaction and its distance from equilibrium, defined for step  $s_p$  as the negative of its Gibbs free energy change, or, alternately, as the difference of the forward affinity  $\bar{A}_p$  minus the reverse affinity  $\bar{A}_p$ .<sup>41</sup>

$$A_p = \sum_{i=1}^l (-\bar{\nu}_{\rho i}) \mu_i - \sum_{i=1}^l \bar{\nu}_{\rho i} \mu_i = \bar{A}_p - \bar{A}_p \quad (2)$$

where  $\mu_i$  is the chemical potential of species  $B_i$ . This provides the condition for the reaction equilibrium ( $A_p = 0$ ), as well as the direction of spontaneous reaction rate  $r_p$  (i.e., forward for  $A_p > 0$ , or for  $\bar{A}_p > \bar{A}_p$ , and reverse for  $A_p < 0$ , or for  $\bar{A}_p < \bar{A}_p$ ), as expressed succinctly by the De Donder inequality,  $P_p \equiv A_p r_p \geq 0$ .<sup>41</sup>

The rate of an elementary reaction step is given by<sup>43</sup>

$$r_p = \bar{r}_p - \bar{r}_p = \bar{k}_p \prod_{i=1}^l a_i^{-\bar{\nu}_{\rho i}} - \bar{k}_p \prod_{i=1}^l a_i^{\bar{\nu}_{\rho i}} \quad (3)$$

where  $a_i$  is the activity of species  $B_i$ . With the use of  $\mu_i = \mu_i^0 + RT \ln a_i$ , eq 3 may be written in the form

$$\frac{\bar{r}_p}{\bar{r}_p} = \exp(-\mathcal{A}_p)$$

or

$$r_p = \bar{r}_p [1 - \exp(-\mathcal{A}_p)] \quad (4)$$

which is the so-called De Donder relation. Here,  $\mathcal{A}_p = A_p/(RT)$  is the dimensionless affinity.

Unfortunately, in this relation, both the affinity  $\mathcal{A}_p$  and the forward rate  $\bar{r}_p$  are functions of temperature and composition; thus, there is not a one-to-one correspondence between the rate  $r_p$  and the affinity  $\mathcal{A}_p$ . As an example,<sup>22</sup> consider the elementary isomerization reaction  $B_1 \rightleftharpoons B_2$ . If the activities of both  $B_1$  and  $B_2$  are doubled, the net rate  $r_p$  doubles due to the doubling of the forward rate  $\bar{r}_p$ , even though the affinity remains the same. The exception is the case close to equilibrium. When  $\mathcal{A}_p = 0$ , the net rate is  $r_p = 0$ , so that  $\bar{r}_p = \bar{r}_{p,0} = \bar{r}_p = \bar{r}_{p,0} = r_{p,0}$ , which is the absolute value of the forward or reverse reaction rate, as required by the principle of microscopic reversibility. In the vicinity of equilibrium, then<sup>44</sup>

$$r_p \approx r_{p,0} \mathcal{A}_p \quad (5)$$

where  $r_{p,0}$  is the exchange rate of the elementary reaction step  $s_p$ .

We will find it useful to write an elementary reaction step  $s_p$  for a catalytic (or enzyme) reaction more explicitly, in terms of the reaction intermediates  $I_i$  and the terminal species  $T_i$ :

$$s_p: \alpha_{p0} I_0 + \sum_{k=1}^q \alpha_{pk} I_k + \sum_{i=1}^n \beta_{pi} T_i = 0 \quad (\text{for } p = 1, 2, \dots, p) \quad (6)$$

For simplicity, a single type of active site  $I_0$  (denoted by  $S$  for a heterogeneous catalyst, and  $E$  for an enzyme) is assumed here, excluded from consideration among  $I_k$  by virtue of site balance. The stoichiometric coefficients of the intermediates  $I_k$  are  $\alpha_{pk}$

( $k = 1, 2, \dots, q$ ) and for the terminal species  $T_i$  are  $\beta_{pi}$  ( $i = 1, 2, \dots, n$ ). The De Donder affinity, thus, becomes

$$\mathcal{A}_p = \ln K_p - \alpha_{p0} \ln a_0 - \sum_{k=1}^q \alpha_{pk} \ln a_k - \sum_{i=1}^n \beta_{pi} \ln a_i \quad (7)$$

where  $K_p = \bar{k}_p/\bar{k}_p$  is the equilibrium constant for the elementary reaction. The species activities  $a_i$  may be replaced by a suitable composition measure, e.g., the site fraction  $\theta_i$  for the intermediates in heterogeneous catalysis and partial pressure or concentration for terminal species. Thus, the affinity may be computed, e.g., from elementary reaction energetics and numerical results of a microkinetic analysis for a given set of conditions and a specified reactor configuration.<sup>35</sup>

The overall stoichiometric matrix  $\nu$  is written with the rows corresponding to reactions (including the overall reaction (OR),  $\sum_{i=1}^n \nu_{OR,i} T_i$ , as the first row) and the columns to the species with the intermediates followed by the terminal species:

$$\nu = \begin{bmatrix} 0 & 0 & \dots & 0 & \nu_{OR,1} & \nu_{OR,2} & \dots & \nu_{OR,n} \\ \alpha_{11} & \alpha_{12} & \dots & \alpha_{1q} & \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2q} & \beta_{21} & \beta_{22} & \dots & \beta_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pq} & \beta_{p1} & \beta_{p2} & \dots & \beta_{pn} \end{bmatrix} \quad (8)$$

in which the active sites are excluded because of the site mass balance and, of course, the stoichiometric coefficients of the intermediates in the OR are zero. In general, we find that rank  $\nu = m \leq p$ . We further define the stoichiometric submatrix of  $\nu$ , involving only reaction intermediates,

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1q} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pq} \end{bmatrix} \quad (9)$$

referred to as the intermediates stoichiometric matrix.<sup>35</sup> It is assumed here that the surface intermediates are linearly independent, i.e., rank  $\alpha = q$ , which would generally be smaller than  $p$ . If this is not so, then the linearly dependent columns may be arbitrarily omitted.

A reaction route (RR) is defined as an appropriate linear combination of the sequential elementary reaction steps  $s_1, s_2, \dots, s_p$  that eliminates all of the reaction intermediates, thus resulting in the OR.<sup>35</sup> As mentioned previously, an infinite variety of RRs is possible if no additional restrictions are imposed. Thus, following Milner,<sup>37</sup> we require the RRs to be *direct*: that is, the elementary reaction steps involved in a direct RR (henceforth denoted simply as RR) are minimal. This means that if an elementary reaction is omitted from the given RR, then it becomes impossible to eliminate all of the reaction intermediates by linearly combining the remaining elementary reactions. Milner<sup>37</sup> showed that a RR involves several elementary reaction steps less than or equal to the rank  $\alpha + 1 = q + 1$ , because the number of elementary reaction steps in a RR are independent. Thus, if the selected  $q + 1$  steps from among the given set of  $p$  elementary reactions are  $s_{i_1}, s_{i_2}, \dots, s_{i_q}, s_{i_{q+1}}$  where  $i_1, i_2, \dots, i_q, i_{q+1}$ , is a set of integers satisfying the conditions,  $1 \leq i_1 < i_2 < \dots < i_q < i_{q+1} \leq p$ , then the  $k$ th reaction route yields<sup>35</sup> OR =  $\sum_{h=1}^{q+1} \sigma_{kh} s_{i_h}$ , or



$$\text{RR}_k: -\text{OR} + \sum_{h=1}^{q+1} \sigma_{kh} s_{i_h} = 0 \quad (\text{for } k = 1, 2, \dots, Q) \quad (10)$$

The total number of conceivable RRs ( $Q$ ) is less than or equal to the total number of ways of selecting  $q + 1$  independent elementary reactions from among a total of  $p$ :<sup>35</sup>

$$Q \leq \frac{p!}{(q+1)!(p-(q+1))!} \quad (11)$$

Usually, the number of distinct RRs is much smaller than this value, because some reactions (e.g., adsorption, desorption steps) must be involved in all RRs, and many of the resulting RRs are indistinct. Nonetheless, they can be large in number.

Fishtik and Datta<sup>35</sup> have shown that the stoichiometric numbers  $\sigma_{kh}$  may be obtained from the intermediates stoichiometric matrix of the chosen  $q + 1$  elementary reaction steps from among  $p$  for the  $k$ th RR, i.e.,

$$\sigma_{kh} = \begin{vmatrix} \alpha_{i_1,1} & \alpha_{i_1,2} & \dots & \alpha_{i_1,q} & 0 \\ \alpha_{i_2,1} & \alpha_{i_2,2} & \dots & \alpha_{i_2,q} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{i_{h-1},1} & \alpha_{i_{h-1},2} & \dots & \alpha_{i_{h-1},q} & 0 \\ \alpha_{i_h,1} & \alpha_{i_h,2} & \dots & \alpha_{i_h,q} & 1 \\ \alpha_{i_{h+1},1} & \alpha_{i_{h+1},2} & \dots & \alpha_{i_{h+1},q} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{i_q,1} & \alpha_{i_q,2} & \dots & \alpha_{i_q,q} & 0 \\ \alpha_{i_{q+1},1} & \alpha_{i_{q+1},2} & \dots & \alpha_{i_{q+1},q} & 0 \end{vmatrix} \quad (12)$$

Within this RR enumeration algorithm, thus, the following scenarios are conceivable:

(a) The selected  $q + 1$  elementary reactions are linearly independent. The resulting RR is a *full* RR, i.e., a RR that produces an OR. Of course, not all of the full RRs are independent, as discussed below.

(b) The selected  $q + 1$  elementary reactions are linearly dependent. The resulting RR, then, is an empty route (ER) or a cycle that produces a zero OR, i.e., an OR in which the stoichiometric coefficients of all the terminal species are zero.

(c) The selected  $q + 1$  elementary reactions are linearly dependent, and there are two or more full RRs or ERs that involve a subset of the same  $q + 1$  elementary reactions. The resulting RR then is a zero RR, i.e., a RR in which all of the stoichiometric numbers are equal to zero. A zero RR, of course, also produces a zero OR.

Equation 10 may be alternately written in matrix notation, i.e.,

$$\sigma s = 0$$

or

$$\begin{bmatrix} \sigma_{\text{OR},1} & \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{\text{OR},2} & \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{\text{OR},Q} & \sigma_{Q1} & \sigma_{Q2} & \dots & \sigma_{Qp} \end{bmatrix} \begin{pmatrix} s_{\text{OR}} \\ s_1 \\ s_2 \\ \vdots \\ s_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (13)$$

where  $\sigma$  is the reaction route matrix and  $s = (s_{\text{OR}}, s_1, s_2, \dots, s_p)^T$

is the reaction vector. The discussion in this paper is limited to RRs in which the stoichiometric numbers, i.e., the elements of  $\sigma$ , are  $+1$ ,  $-1$ , or  $0$ . Furthermore, the entries of the first column are  $\sigma_{\text{OR},k} = -1$  for a full RR, and  $\sigma_{\text{OR},k} = 0$  for an ER.

The thermodynamic affinities of the OR and the elementary reactions follow the same linear combination, i.e.,<sup>35</sup>

$$-A_{\text{OR}} + \sum_{\rho=1}^p \sigma_{k\rho} A_{\rho} = 0$$

or

$$\sigma A = 0 \quad (14)$$

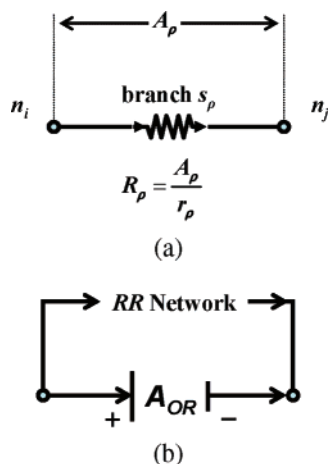
where  $A$  is the affinity vector, i.e.,  $A = (A_{\text{OR}}, A_1, A_2, \dots, A_p)^T$  and  $A_{\text{OR}}$  represents the affinity of the OR. Alternately, this may be written in terms of the dimensionless affinity,  $\mathcal{A}_{\rho}$ .

### 3. Reaction Route Graphs and Their Topology

Because alternate terminology and representation for reaction networks in graph theory exists in the literature, some basic terms are defined below, within the context of our approach, for the sake of specificity, albeit at the peril of some redundancy. The terminology adopted here is more common in electrical networks<sup>45,46</sup> (e.g., “branches” rather than “edges,” “nodes” instead of “vertices,” and “trees” instead of “spanning trees”), because we extensively use the electrical analogy for subsequent RR analysis.

The RR topological structure of a reaction network is represented by the so-called reaction route graph  $G_R$ , which is defined as an ordered, connected, directed (i.e., a digraph), planar or nonplanar, cycle graph that is comprised of  $B$  branches, each representing an elementary or overall reaction, and  $N$  nodes, which illustrates how the reactions are interconnected to allow for enumeration of all direct RRs. If all given reactions are represented only once, the  $G_R$  is termed a *minimal* RR graph,  $G_{R,\text{min}}$ . We will concern ourselves here only with the case of a single OR and the development of minimal RR graphs, their generalizations being considered in subsequent work. As per the convention in graph theory, when a branch originates at a node  $n_j$ , it is said to be *incident from* the node  $n_j$ . When a branch terminates at a node  $n_j$ , it is said to be *incident to* the node  $n_j$ . Two or more branches are *parallel* if they have the same pair of starting and ending nodes.

Because graph theory is concerned with topological features of a network, the specific nature of the branch is unimportant and, thus, may be represented simply by a directed line segment. However, because we will be later using the electrical analogy of reaction networks, it is convenient to represent an elementary reaction as a resistor (Figure 2a), which is appropriate for the steady-state case under consideration here, although more complex electrical representation (e.g., involving capacitors) may be required for a time-dependent case.<sup>24</sup> Thus, in what follows, a RR graph is synonymous to a RR network. The oriented branch variables (Figure 2a) are the reaction affinity  $A_{\rho}$  (much as the electrical potential *difference* or branch voltage) and its rate  $r_{\rho}$  (much as electrical current). The orientation shown is simply assumed, and the reaction may actually proceed in either direction. Branch arrows depict the assumed direction of reaction progress and, thus, are directed from reactants to products, including the OR. In a given reaction route, of course,  $s_{\rho}$  may proceed in either the assumed direction ( $\sigma_{\rho} = +1$ ) or in the opposite direction ( $\sigma_{\rho} = -1$ ). The OR is depicted as a voltage



**Figure 2.** Transformation of a RR graph into a reaction network. In panel a, an elementary reaction is viewed as a resistor ( $R_\rho = A_\rho/r_\rho$ ) connecting two nodes in a RR graph; the arrows indicate the direction of the forward reaction. In panel b, the affinity of the overall reaction is viewed as a voltage source; by the convention adopted here, the arrows indicate that the rates of the overall reaction (OR) via the RR network and via the voltage source have opposite signs.

source when the electrical analogy is utilized (Figure 2b). Otherwise, a directed line segment with assumed orientation suffices.

It may be remarked that the nodes in the RR graphs previously defined do not denote individual intermediates or terminal species, as is almost universally the case in the so-called "reaction graphs",<sup>2</sup> but simply indicate the interconnection of reactions involved in RRs (hence, the reason for labeling these "reaction route graphs"). The interconnection of the elementary reaction steps at a node in a RR graph is subject to QSS conditions for the intermediates. For linear mechanisms, the physicochemical meaning of the nodes is transparent; each node represents a species and its QSS condition. For RR graphs, however, the nodes represent a certain linear combination of species and their QSS conditions. Thus, a node  $n_j$  represents the linear combination of species and their properties that are associated with the products of the reactions incident to the node plus those associated with the reactants of the reactions incident from the node with an appropriate sign (negative for reactions incident to a node, and positive for reactions incident from a node, as discussed later in the context of the incidence matrix). The forward and reverse affinities ( $\bar{A}_\rho$  and  $A_\rho$ ) of the reactions (branches) at a node follow a similar additive rule, so that each node is associated with a unique value of Gibbs free energy, which is a state variable. In fact, any other thermodynamic state function, e.g., enthalpy, may be similarly associated with each node. This is useful, for instance, in obtaining a reaction energy diagram directly from  $G_R$ , as shown in Part II.<sup>40</sup>

The  $k$ th walk from node  $n_i$  to node  $n_j$  is an alternating sequence of nodes and branches, beginning with a starting node  $n_i$  and ending with terminal node  $n_j$ , and is given by  $w_k = \sum_{n_i \leftarrow n_j} \sigma_{k\rho} s_\rho$ . If the branch is oriented along the direction of the walk,  $\sigma_{k\rho} = +1$ ; otherwise,  $\sigma_{k\rho} = -1$ . If a branch does not occur in the walk,  $\sigma_{k\rho} = 0$ . A given branch may not be crossed more than once in a walk, although a node may be traversed more than once. A closed walk begins and ends at the same node; otherwise, it is termed an open walk. A closed walk in which no node except the terminal node appears more than once is called a cycle. If a cycle includes the OR, it is considered a

full RR; if only elementary reactions steps are traversed, then the cycle represents an empty route (ER).

A (spanning) reaction tree  $T_R$  is a connected subgraph of the  $G_R$  that has all of the nodes of the original  $G_R$  but is without some of its branches so that it has no cycles. Thus,  $T_R$  is a connected graph of  $N$  nodes and  $N - 1$  branches. The branches of a given tree are called its twigs,  $s_\rho^{(t)}$ . Those branches of the  $G_R$  that are excluded in a given tree are called its links,  $s_\rho^{(l)}$ . For a RR graph with  $N$  nodes and  $B$  branches, the number of twigs is  $N - 1$ , which is called the tree rank, and the number of links is  $L = B - N + 1$ . In fact, we shall see that twigs are simply a chosen set of independent elementary reaction steps (or branches).

If a link is added back to a tree, the resulting graph contains one cycle, which is given the orientation of the link. The addition of each subsequent link forms one or more additional cycles. Cycles that contain only one link are independent and are called the fundamental cycles or fundamental RRs of the  $G_R$ . Clearly, these include at least one full RR. There are  $B - N + 1$  fundamental cycles in a  $G_R$ . Therefore, in a minimal RR graph, the number of branches is  $B = p + 1$  (i.e., the number of elementary reactions plus the number of ORs), the number of fundamental cycles is equal to the number of independent RRs  $p - q$ , and the number of nodes  $N$  are given as  $N = q + 2$ .

A cut set ( $X_R$ ) is a set of branches which, when removed from  $G_R$ , leave it disconnected as two components or connected subgraphs, provided no subset of  $X_R$  does so, and is assigned an orientation from one component to another. Furthermore, because the removal of any twig from a tree cuts it into two parts, every twig of a tree is included in a cut-set. A cut set that contains exactly one twig from the tree  $T_R$ , whereas the rest of the removed branches are links from the corresponding  $G_R$ , is called a fundamental cut set ( $X_{R,f}$ ), which is given the same orientation as the twig that defines it.

The incidence of branches at the nodes in the  $G_R$  are given by the  $N \times B$  incidence matrix  $\mathbf{M} = [m_{j\rho}]$ , with rows corresponding to the  $N$  nodes and columns to the  $B$  branches (reactions) of the  $G_R$ . The elements of the incidence matrix are defined by

$$m_{j\rho} = \begin{cases} +1 & \text{if branch } s_\rho \text{ is incident from node } n_j \\ -1 & \text{if branch } s_\rho \text{ is incident to node } n_j \\ 0 & \text{if branch } s_\rho \text{ is not incident at node } n_j \end{cases}$$

Because every branch  $s_\rho$  of  $G_R$  is incident to one node and incident from another node, each column of  $\mathbf{M}$  has exactly one  $+1$  value and one  $-1$  value; the remaining values are zeros. Consequently, the sum of each column in the incidence matrix is zero, and  $\text{rank}(\mathbf{M}) = N - 1$ . Thus, any row corresponding to a reference node may be deleted to obtain a  $(N - 1) \times B$  matrix called the reduced incidence matrix,  $\mathbf{M}_f$ . Because a tree  $T_R$  is a connected subgraph of  $N$  nodes and  $(N - 1)$  branches, its reduced incidence matrix is a square matrix of order and rank  $N - 1$ . Thus, an  $(N - 1) \times (N - 1)$  submatrix of  $\mathbf{M}$  is nonsingular if and only if the  $(N - 1)$  branches of this matrix correspond to tree twigs,  $s_\rho^{(t)}$ . As a result, the reduced incidence matrix may be rearranged in the form

$$\mathbf{M}_f = [\mathbf{M}_t : \mathbf{M}_l] \quad (15)$$

where  $\mathbf{M}_t$  is a square submatrix with  $(N - 1)$  columns, which correspond to the twigs  $s_\rho^{(t)}$  of the chosen tree  $T_R$  and the

submatrix  $\mathbf{M}_l$  has ( $L = B - N + 1$ ) columns, which correspond to the links  $s_\rho^{(l)}$  of the tree. Thus,  $\mathbf{M}_l$  is nonsingular and its inverse exists.

The RR matrix  $\sigma = [\sigma_{k\rho}]$  of a  $G_R$ , akin to the cycle matrix in conventional graph theory, is a  $Q \times B$  matrix (as defined in eq 13), with rows that correspond to the reaction routes (full RRs as well as ERs) and columns to the branches of the  $G_R$ . The elements of the RR matrix are

$$\sigma_{k\rho} = \begin{cases} +1 & \text{if the } k\text{th cycle includes the } \rho\text{th branch} \\ & \text{and their orientations coincide} \\ -1 & \text{if the } k\text{th cycle includes the } \rho\text{th branch} \\ & \text{and their orientations do not coincide} \\ 0 & \text{if the } k\text{th cycle does not include the} \\ & \rho\text{th branch} \end{cases}$$

In other words, in this paper, we are concerned only with RRs in which the stoichiometric numbers are equal to +1, -1, or 0.

The total number of linearly independent RRs, as per the Horiuti–Temkin theorem,<sup>13,38</sup> is well-known to be given by  $L = p - \text{rank } \alpha = p - q = B - N + 1$ . Any submatrix of  $\sigma$  that is comprised of an arbitrary set of  $L$  linearly independent RRs is denoted by  $\sigma_f$  and is called the fundamental RR matrix. If we further rearrange the columns in  $\sigma_f$  in the same order as in the reduced incidence matrix, i.e., twigs followed by links, and rearrange the rows so that the first row corresponds to the fundamental cycle in the first column, and so on, then  $\sigma_f$  takes the form

$$\sigma_f = [\sigma_f; \sigma_l] = [\sigma_f; \mathbf{I}_L] \quad (16)$$

i.e.,  $\sigma_l = \mathbf{I}_L$  is an identity submatrix of the order  $L$ , and  $\sigma_f$  is the remaining  $L \times (N - 1)$  submatrix corresponding to the twigs of the tree  $T_R$ . Thus, each fundamental cycle contains only one link.

The cut-set matrix  $\mathbf{X} = [x_{h\rho}]$  of a  $G_R$  is a  $(N - 1) \times B$  matrix, in which rows correspond to cut sets and columns correspond to the branches of the  $G_R$ , and is defined by

$$x_{h\rho} = \begin{cases} +1 & \text{if the } h\text{th cut set includes the } \rho\text{th branch} \\ & \text{and their orientations coincide} \\ -1 & \text{if the } h\text{th cut set includes the } \rho\text{th branch} \\ & \text{and their orientations do not coincide} \\ 0 & \text{if the } h\text{th cut set does not include the} \\ & \rho\text{th branch} \end{cases}$$

The rank of the cut-set matrix is equal to the number of twigs, i.e.,  $N - 1$ . A cut-set matrix that contains only the fundamental cut sets is called the fundamental cut-set matrix,  $\mathbf{X}_f$  with respect to  $T_R$ . Furthermore, it may be written in the form

$$\mathbf{X}_f = [\mathbf{I}_{N-1}; \mathbf{X}_l] \quad (17)$$

where the first  $(N - 1)$  columns form an identity matrix corresponding to the  $(N - 1)$  twigs of the tree, because such a cut set includes only one twig each, and the final  $(B - N + 1)$  columns form the submatrix  $\mathbf{X}_l$  that corresponds to the links,  $s_\rho^{(l)}$ .

The well-known interrelationships in conventional graph theory among the reduced incidence matrix  $\mathbf{M}_f$ , the fundamental

cycle or RR matrix  $\sigma_f$ , and the fundamental cut-set matrix  $\mathbf{X}_f$  for a given tree  $T_R$ <sup>45,46</sup> prove to be of great utility in RR graphs. Therefore, key relationships are provided below.  $\mathbf{M}_f \sigma_f^T = 0$ , i.e.,  $\mathbf{M}$  and  $\sigma$  are orthogonal; thus,

$$[\mathbf{M}_l; \mathbf{M}_t] \begin{bmatrix} \sigma_f^T \\ \dots \\ \mathbf{I}_L \end{bmatrix} = 0$$

or

$$\sigma_f^T = -\mathbf{M}_t^{-1} \mathbf{M}_l \quad (18)$$

Similarly, because  $\mathbf{X}_f \sigma_f^T = 0$ , the submatrix  $\mathbf{X}_l = -\sigma_f^T$ , or

$$\mathbf{X}_f = [\mathbf{I}_{N-1}; -\sigma_f^T] \quad (19)$$

In other words, given  $\mathbf{M}_f$ , and starting with an arbitrary tree  $T_R$  and its subgraph in  $\mathbf{M}_t$  and  $\mathbf{M}_l$ , the matrixes  $\sigma_f$  and  $\mathbf{X}_f$  can be readily found. For the converse problem, however, i.e., given  $\sigma_f$  or  $\mathbf{X}_f$ ,  $\mathbf{M}_f$  generally cannot be readily determined. One useful relation in this regard, however, is obtained as follows. Solving for  $\mathbf{M}_l$  from eq 18, using it in the relation  $\mathbf{M}_f = [\mathbf{M}_l; \mathbf{M}_t]$ , and comparing the result with eq 19 provides the relation  $\mathbf{M}_f = \mathbf{M}_t [\mathbf{I}_{N-1}; -\sigma_f^T]$ , i.e.,

$$\mathbf{M}_f = \mathbf{M}_t \mathbf{X}_f \quad (20)$$

Thus,  $\mathbf{M}_t$  acts as a nonsingular transformation matrix, transforming the cut-set matrix to the reduced incidence matrix. In other words, the reduced incidence matrix of a graph is row-equivalent to the fundamental cut-set matrix for a given tree. Thus, the rows of  $\mathbf{M}_f$  may, in principle, be obtained from linear combination of the rows of  $\mathbf{X}_f$ . This is used later in the realization of RR graphs from the RR matrix.

#### 4. Network Conservation Laws and Electrical Circuit Analogy

The previous discussion is concerned with the topological features of RR graphs that are considered in a different light than the previous literature for reaction graphs. When combined with conservation laws, along with an analogy with electrical circuits,<sup>22,45,46</sup> it provides a powerful new methodology for analyzing RR networks. These conservation laws impose constraints on branch rates and affinities by virtue of the topology of the  $G_R$ .

##### 4.1. Kirchhoff's Current Law (Conservation of Mass).

Assuming the node  $n_j$  to have zero capacity, the net rates of reactions (akin to branch current) incident at the node  $n_j$  sum to zero, i.e.,

$$\mathbf{M}_f \mathbf{r} = 0 \quad (21)$$

where  $\mathbf{r}$  is the vector of branch rates in the same order as the branches in  $\mathbf{M}_f$ , i.e.,  $\mathbf{r} = (-r_{\text{OR}}, r_1, r_2, \dots, r_p)^T$ . Notice that, by the convention adopted here, the rates of the OR via the RR network and via the voltage source have equal but opposite values (see Figure 2b). Equation 21 expresses the mass balance for the terminal species or a group of species around a control volume. This is the equivalent to Kirchhoff's Current Law

(KCL) of electrical circuits. If use is made of eq 20 in this methodology,

$$\mathbf{X}_f \mathbf{r} = 0 \quad (22)$$

which is an alternate form of KCL. If eq 19 is used in this expression for  $\mathbf{X}_f$ , it provides the final form of KCL:

$$\mathbf{r}^{(i)} = \boldsymbol{\sigma}_t^T \mathbf{r}^{(l)} \quad (23)$$

where  $\mathbf{r}^{(i)}$  is the vector of twig rates and  $\mathbf{r}^{(l)}$  is that of link rates.

In other words, the twig rates can be determined in terms of the  $L$  linearly independent rates of links for a chosen tree of  $G_R$ . Furthermore, because each link is present in only one of the fundamental cycles, the link rates are also equal to the rates of the fundamental RRs of a RR graph, akin to loop currents in an electrical circuit. Of course, the fundamental RRs include both full and empty RRs. Denoting the rates (fluxes) of these independent RRs by  $J_I$  for  $\text{RR}_1$ ,  $J_{II}$  for  $\text{RR}_2, \dots$ , and  $J_L$  for  $\text{RR}_L$ , then  $\mathbf{r}^{(l)} = \mathbf{J}$ , where  $\mathbf{J} = (J_I, J_{II}, \dots, J_L)^T$  and

$$\mathbf{r}^{(i)} = \boldsymbol{\sigma}_t^T \mathbf{J} \quad (24)$$

which also provides the rate of a single OR as the sum of the fluxes  $J_k$  of the independent RRs:

$$r_{\text{OR}} = - \sum_{k=1}^L \sigma_{\text{OR},k} J_k \quad (25)$$

**4.2. Kirchhoff's Voltage Law (Thermodynamic Consistency).** Being a state function, the sum of reaction affinities (akin to branch voltages) around the  $k$ th full or empty RR is zero, i.e.,  $\boldsymbol{\sigma}_f \mathbf{A} = 0$ , or

$$\mathbf{A}^{(l)} = -\boldsymbol{\sigma}_f \mathbf{A}^{(i)} \quad (26)$$

For a RR graph  $G_R$  with  $N$  nodes and  $B$  branches, Kirchhoff's Voltage Law (KVL) thus provides  $L$  independent equations for affinities of individual reactions (analogous to branch voltage drops). If the constitutive equations (rate laws) were available relating  $r_\rho$  to  $A_\rho$ , the individual reaction rates and affinities could, in principle, be determined. However, note that KCL and KVL are entirely independent of the specific form of the constitutive equation and are applicable to linear as well as nonlinear constitutive laws,<sup>47</sup> and both steady-state and unsteady-state cases.

Furthermore, KVL provides thermodynamic consistency of the rates of reactions involved in a cycle, which is an important check on the consistency of the given or calculated kinetics of the elementary reaction steps. Thus, when eq 4 is used in eq 26,

$$\prod_{\text{ER}_k} \left( \frac{\vec{r}_\rho}{\bar{r}_\rho} \right)^{\sigma_{k\rho}} = 1 \quad (27a)$$

and

$$\prod_{\text{RR}_k} \left( \frac{\vec{r}_\rho}{\bar{r}_\rho} \right)^{\sigma_{k\rho}} = \frac{\vec{r}_{\text{OR}}}{\bar{r}_{\text{OR}}} \quad (27b)$$

where  $\vec{r}_{\text{OR}}$  and  $\bar{r}_{\text{OR}}$  are the rates of the forward and reverse OR, respectively. Of course, the use of eq 3 in eq 27 results in the alternate form

$$\prod_{\text{ER}_k} \left( \frac{\vec{k}_\rho}{\bar{k}_\rho} \right)^{\sigma_{k\rho}} = 1 \quad (28a)$$

and

$$\prod_{\text{RR}_k} \left( \frac{\vec{k}_\rho}{\bar{k}_\rho} \right)^{\sigma_{k\rho}} = K_{\text{OR}} \quad (28b)$$

**4.3. Tellegen's Theorem (Conservation of Energy).** The reaction power dissipated by an elementary reaction, or, alternatively, the entropy production, is  $P_\rho = A_\rho r_\rho$ .<sup>41</sup> Notice that, for the overall reaction, considered as a voltage source (akin to, e.g., a battery as shown in Figure 2b), the power dissipated is given by  $P_{\text{OR}} = A_{\text{OR}} r_{\text{OR}}$ . From the conservation of energy,

$$\mathbf{A}^T \mathbf{r} = 0$$

or

$$-A_{\text{OR}} r_{\text{OR}} + \sum_{\rho=1}^p A_\rho r_\rho = 0 \quad (29)$$

which is equivalent to Tellegen's theorem of electrical circuits.<sup>48</sup> In fact, this follows as a consequence of the applicability of KCL and KVL to a network and, thus, does not provide any new information.

## 5. Realization of Reaction Route Graph

The problem of realization of a RR graph for a minimal mechanism is the converse of the problem previously discussed, namely that of the analysis of a given RR graph. When the incidence matrix of a RR graph is available, it is, of course, straightforward to construct a RR graph. Therefore, the essential problem in RR graph realization is the determination of the incidence matrix. There are two alternate ways of obtaining the incidence matrix for a RR graph: (i) from the RRs matrix  $\boldsymbol{\sigma}$  and (ii) from the overall stoichiometric matrix  $\boldsymbol{\nu}$ , as described below.

**5.1. The Incidence Matrix from the Fundamental RR Matrix.** In principle, it is possible to construct the RR graph directly from the RR matrix  $\boldsymbol{\sigma}$  by trial and error, because it contains all the graph connectivity information, although this is more directly accomplished from the incidence matrix  $\mathbf{M}$ . Thus, the following algorithm may be followed:

Step 1: Derive and select a fundamental RR matrix  $\boldsymbol{\sigma}_f$  using literature methods.<sup>33–35,39</sup> Rearrange the matrix to the form  $\boldsymbol{\sigma}_f = [\boldsymbol{\sigma}; \mathbf{I}_t]$  by column interchange. This procedure, thus, simultaneously identifies the links (corresponding to the columns of the identity matrix), with the remaining branches being twigs of the resulting tree selected by the choice of the full and empty RRs. Clearly, there are other possible trees and, thus,  $\boldsymbol{\sigma}_f$  is not unique.

Step 2: Obtain the corresponding fundamental cut-set matrix from the relation  $\mathbf{X}_f = [\mathbf{I}_{N-1}; -\boldsymbol{\sigma}_f^T]$  for this tree.

Step 3: In view of the fact that  $\mathbf{M}_f$  and  $\mathbf{X}_f$  are row equivalent (eq 20),  $\mathbf{M}_f$  is obtained from  $\mathbf{X}_f$  via elementary row operations such that, at most, each column of the resulting matrix consists of one +1 value and one -1 value, where the remaining values are zeros.



Step 4: Obtain the complete incidence matrix  $\mathbf{M}$  from  $\mathbf{M}_f$  by adding the missing row, so that the sum of elements is zero in each column.

**5.2. The Incidence Matrix from the Overall Stoichiometric Matrix.** According to the QSS condition, the rates of formation and consumption of the intermediates are equal. The QSS conditions for the intermediates when further combined with the material balances for the terminal species results in

$$\mathbf{v}^T \mathbf{r} = 0$$

or

$$\begin{bmatrix} 0 & \alpha_{11} & \alpha_{21} & \dots & \alpha_{p1} \\ 0 & \alpha_{12} & \alpha_{22} & \dots & \alpha_{p2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \alpha_{1q} & \alpha_{2q} & \dots & \alpha_{pq} \\ \nu_{\text{OR},1} & \beta_{11} & \beta_{21} & \dots & \beta_{p1} \\ \nu_{\text{OR},2} & \beta_{12} & \beta_{22} & \dots & \beta_{p2} \\ \dots & \dots & \dots & \dots & \dots \\ \nu_{\text{OR},n} & \beta_{1n} & \beta_{2n} & \dots & \beta_{pn} \end{bmatrix} \begin{pmatrix} r_{\text{OR}} \\ r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (30)$$

It is evident that this is analogous to KCL, namely,

$$\mathbf{M} \mathbf{r} = 0 \quad (31)$$

and, hence, may be used to determine the incidence matrix  $\mathbf{M}$ . For the case of monomolecular reactions, the overall stoichiometric matrix  $\mathbf{v}^T$  is, in fact, the same as the reduced incidence matrix  $\mathbf{M}_f$ .<sup>49</sup> In that case, of course, the nodes do indeed represent the individual chemical species and the branches represent the reactions. For the more general case, a node  $n_j$  represents the mass balance for a group of species obtained by adding specified rows from  $\mathbf{v}^T$ , i.e.,

$$\sum_i \left( \sum_\rho \nu_{\rho i} r_\rho \right) = 0 = \sum_\rho r_\rho \sum_i \nu_{\rho i} = \lambda \left( \sum_\rho r_\rho m_{j\rho} \right) \quad (\text{for } j = 1, 2, \dots, n) \quad (32)$$

where  $\lambda$  is a constant. This is, thus, equivalent to KCL. More generally, therefore, the two matrixes,  $\mathbf{M}$  and  $\mathbf{v}^T$ , are equivalent, and one may be obtained from the other by a finite number of elementary row operations. To first determine the incidence matrix  $\mathbf{M}$  from the stoichiometric matrix  $\mathbf{v}^T$ , the following steps are recommended:

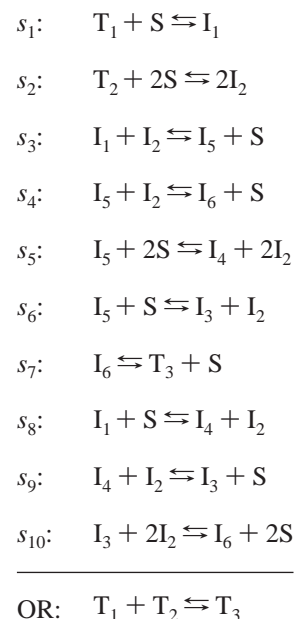
Step 1: Perform elementary row operations with  $\mathbf{v}^T$  involving (a) an interchange of rows or columns, (b) a mutual addition of rows, and (c) multiplication of a row by a scalar constant, until each column has at most one +1 value and one -1 value, with the rest of the values being zero.

Step 2: Eliminate the rows that contain all zeros; the resulting matrix is  $\mathbf{M}$ .

## 6. An Example

Consider the following hypothetical catalytic mechanism involving 10 elementary reaction steps on a catalyst (where

S is a vacant catalyst site):



One may notice that this mechanism is nonlinear, in that the elementary reactions are not first order in the intermediates. Although this mechanism is hypothetical, it could conceivably correspond to that of olefin hydrogenation, for example, with  $T_1$  = olefin,  $T_2$  = hydrogen, and  $T_3$  = alkane. The intermediates then might correspond to  $I_1$  = adsorbed olefin,  $I_2$  = adsorbed hydrogen atom,  $I_6$  = adsorbed alkane product, and  $I_3, I_4$ , and  $I_5$  = adsorbed intermediates of different structure or state of hydrogenation or dehydrogenation. More specific examples are illustrated in Part II.

**6.1. The Incidence Matrix from the Fundamental RR Matrix.** The aforementioned mechanism is comprised of  $p = 10$  elementary reaction steps,  $q = 6$  reaction intermediates (i.e.,  $I_1$ – $I_6$ , besides the catalyst site S), and three terminal species ( $T_1$ – $T_3$ ). Thus, according to the Horiuti–Temkin theorem, the number of linearly independent RRs is equal to  $L = p - q = 10 - 6 = 4$ . An appropriate set of linearly independent RRs—that is, the fundamental RR matrix—may be generated, using the algorithm presented in refs 34 and 35:

$$\sigma_f = \begin{array}{c} \text{OR} \quad s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6 \quad s_7 \quad s_8 \quad s_9 \quad s_{10} \\ \text{RR}_1 \begin{bmatrix} -1 & +1 & +1 & +1 & +1 & 0 & 0 & : & +1 & 0 & 0 & 0 \end{bmatrix} \\ \text{ER}_1 \begin{bmatrix} 0 & 0 & 0 & -1 & 0 & -1 & 0 & : & 0 & +1 & 0 & 0 \end{bmatrix} \\ \text{ER}_2 \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & +1 & -1 & : & 0 & 0 & +1 & 0 \end{bmatrix} \\ \text{ER}_3 \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 & +1 & : & 0 & 0 & 0 & +1 \end{bmatrix} \end{array} \quad (33)$$

which is already arranged in the form  $\sigma_f = [\sigma_i : \mathbf{I}_L]$ . Therefore, the identity submatrix in  $\sigma_f$  identifies the links as steps  $s_7$ – $s_{10}$ , whereas the twigs are steps  $s_1$ – $s_6$ , in addition to the OR. Thus,  $\sigma_i$  is the submatrix that is composed of the first seven columns of  $\sigma_f$ .

Next, the fundamental cut-set matrix is obtained from the relation  $\mathbf{X}_f = [\mathbf{I}_{N-1} : -\sigma_i^T]$ :



$\mathbf{X}_f =$ 

$$\begin{array}{c}
 \begin{array}{c} \text{twigs} \end{array} \quad \begin{array}{c} \text{links} \end{array} \\
 \begin{array}{c} \text{OR} \quad s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6 \quad s_7 \quad s_8 \quad s_9 \quad s_{10} \end{array} \\
 \begin{array}{l} X_{f,1} \\ X_{f,2} \\ X_{f,3} \\ X_{f,4} \\ X_{f,5} \\ X_{f,6} \\ X_{f,7} \end{array} \left[ \begin{array}{cccccccccccc}
 +1 & 0 & 0 & 0 & 0 & 0 & 0 & : & +1 & 0 & 0 & 0 \\
 0 & +1 & 0 & 0 & 0 & 0 & 0 & : & -1 & 0 & 0 & 0 \\
 0 & 0 & +1 & 0 & 0 & 0 & 0 & : & -1 & 0 & 0 & 0 \\
 0 & 0 & 0 & +1 & 0 & 0 & 0 & : & -1 & +1 & 0 & 0 \\
 0 & 0 & 0 & 0 & +1 & 0 & 0 & : & -1 & 0 & 0 & +1 \\
 0 & 0 & 0 & 0 & 0 & +1 & 0 & : & 0 & +1 & -1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & +1 & : & 0 & 0 & +1 & -1
 \end{array} \right]
 \end{array} \quad (34)$$

Performing elementary row operations on this matrix, so that each column of the resulting matrix consists of, at most, one +1 value and one -1 value, with the balance of elements being zero, and finally adding the missing row, the complete incidence matrix is obtained:

 $\mathbf{M} =$ 

$$\begin{array}{c}
 \begin{array}{c} \text{twigs} \end{array} \quad \begin{array}{c} \text{links} \end{array} \\
 \begin{array}{c} \text{OR} \quad s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6 \quad s_7 \quad s_8 \quad s_9 \quad s_{10} \end{array} \\
 \begin{array}{l} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \\ n_7 \\ n_8 \end{array} \left[ \begin{array}{cccccccccccc}
 +1 & +1 & 0 & 0 & 0 & 0 & 0 & : & 0 & 0 & 0 & 0 \\
 0 & -1 & +1 & 0 & 0 & 0 & 0 & : & 0 & 0 & 0 & 0 \\
 0 & 0 & -1 & +1 & 0 & 0 & 0 & : & 0 & +1 & 0 & 0 \\
 0 & 0 & 0 & -1 & +1 & +1 & +1 & : & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & -1 & 0 & 0 & : & +1 & 0 & 0 & -1 \\
 0 & 0 & 0 & 0 & 0 & -1 & 0 & : & 0 & -1 & +1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & -1 & : & 0 & 0 & -1 & +1 \\
 -1 & 0 & 0 & 0 & 0 & 0 & 0 & : & -1 & 0 & 0 & 0
 \end{array} \right]
 \end{array} \quad (35)$$

The RR graph may now be drawn based on this incidence matrix and is shown in Figure 3a, which identifies the fundamental cycles chosen (eq 33). The corresponding tree is shown in Figure 3b. The RR graph could alternatively have been constructed by a graphical algorithm, namely by drawing  $\text{RR}_1$  and then adding the three remaining cycles ( $\text{ER}_1$ ,  $\text{ER}_2$ , and  $\text{ER}_3$ ) given in eq 33. However, this algorithm may be undesirable for large numbers of independent fundamental cycles.

**6.2. The Incidence Matrix from the Overall Stoichiometric Matrix.** The overall stoichiometric coefficient matrix, in this case, is

$$\begin{array}{c}
 \begin{array}{c} \text{OR} \quad s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6 \quad s_7 \quad s_8 \quad s_9 \quad s_{10} \end{array} \\
 \begin{array}{l} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ \dots \\ T_1 \\ T_2 \\ T_3 \end{array} \left[ \begin{array}{cccccccccccc}
 0 & +1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
 0 & 0 & +2 & -1 & -1 & +2 & +1 & 0 & +1 & -1 & -2 \\
 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & +1 & -1 \\
 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & +1 & -1 & 0 \\
 0 & 0 & 0 & +1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & +1 & 0 & 0 & -1 & 0 & 0 & +1 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 +1 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0
 \end{array} \right]
 \end{array} \quad (36)$$

The determination of the incidence matrix from this is ac-

complished via a series of elementary row operations. The resulting incidence matrix is the same as that previously given and, therefore, the corresponding RR graph is the same as that shown in Figure 3a.

**6.3. The Complete RR Matrix from the RR Graph.** Once a RR graph is available, e.g., as illustrated previously for the example, the conventional graph theory can be used to enumerate a complete set of direct RRs (that is, the RR matrix,  $\sigma$ ). For our example,  $\sigma$  can be easily generated by a simple inspection of the RR graph. The result is

$$\begin{array}{c}
 \begin{array}{c} \text{OR} \quad s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6 \quad s_7 \quad s_8 \quad s_9 \quad s_{10} \end{array} \\
 \begin{array}{l} \text{RR}_1 \\ \text{RR}_2 \\ \text{RR}_3 \\ \text{RR}_4 \\ \text{RR}_5 \\ \text{RR}_6 \\ \text{RR}_7 \\ \dots \\ \text{ER}_1 \\ \text{ER}_2 \\ \text{ER}_3 \\ \text{ER}_4 \\ \text{ER}_5 \\ \text{ER}_6 \end{array} \left[ \begin{array}{cccccccccccc}
 -1 & +1 & +1 & +1 & +1 & 0 & 0 & +1 & 0 & 0 & 0 \\
 -1 & +1 & +1 & 0 & +1 & -1 & 0 & +1 & +1 & 0 & 0 \\
 -1 & +1 & +1 & +1 & 0 & 0 & +1 & +1 & 0 & 0 & +1 \\
 -1 & +1 & +1 & 0 & 0 & 0 & 0 & +1 & +1 & +1 & +1 \\
 -1 & +1 & +1 & +1 & 0 & +1 & 0 & +1 & 0 & +1 & +1 \\
 -1 & +1 & +1 & 0 & 0 & -1 & +1 & +1 & +1 & 0 & +1 \\
 -1 & +1 & +1 & 0 & +1 & 0 & -1 & +1 & +1 & +1 & 0 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & +1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & +1 & -1 & 0 & 0 & +1 & 0 \\
 0 & 0 & 0 & 0 & -1 & 0 & +1 & 0 & 0 & 0 & +1 \\
 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & +1 & +1 & 0 \\
 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & +1 & +1 \\
 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & +1 & +1 & +1
 \end{array} \right]
 \end{array} \quad (37)$$

This RR matrix is equivalent to that generated using the conventional enumeration procedures.<sup>34,35</sup>

## 7. Reaction Network Analysis and Reduction

The construction of a RR graph, even without a subsequent kinetic analysis, provides a powerful visualization tool and a deeper understanding of the reaction mechanism, as compared to the traditional and computationally expensive analyses. A further use of the electrical circuit analogy and Kirchhoff's laws provides a powerful new tool for a quantitative analysis and reduction of the RR network. Obviously, not all of the possible RRs are equally important. There are, in fact, many alternate methods available for the reduction of kinetic systems.<sup>50</sup> We propose a different approach here based on the concept of reaction resistance in analogy to Ohm's Law, followed by utilization of the techniques of electrical network analysis, to analyze and reduce RR networks. This turns out to be a particularly powerful method, allowing us not only to discriminate among the various RRs, but also to determine the slow (or rate-determining) steps, as well as the quasi-equilibrium steps and, ultimately, when possible, the derivation of a simple and accurate algebraic rate equation suitable for reactor analysis and design.

Because of its nonlinearity, the basic constitutive equation (i.e., the De Donder relation in eq 4) makes it difficult to utilize the techniques of linear circuit analysis, based on Ohm's Law. We, thus, define a reaction resistance as the mean value of the reciprocal rate (conductance) between the limiting values characterizing the reaction, i.e., the reverse and forward rate ( $\bar{r}_\rho$  and  $\bar{r}_\rho$ , respectively), i.e.,

$$R_\rho = \left\langle \frac{1}{r_\rho} \right\rangle = \frac{1}{\bar{r}_\rho - \bar{r}_\rho} \int_{\bar{r}_\rho}^{\bar{r}_\rho} \frac{1}{r_\rho} dr_\rho = \frac{\ln(\bar{r}_\rho/\bar{r}_\rho)}{\bar{r}_\rho - \bar{r}_\rho} \quad (38)$$

where  $\langle r_\rho \rangle$  is a logarithmic mean conductance and reduces to



$$R_{k,n_i \rightarrow n_j} = \sum_{n_i \rightarrow n_j} \sigma_{kp}^2 R_p \quad (42)$$

and, of course, the reaction flux through the sequential branches is

$$J_k = \frac{A_{n_i \rightarrow n_j}}{R_{k,n_i \rightarrow n_j}} = \frac{\sum_{n_i \rightarrow n_j} \sigma_{kp} A_p}{R_{k,n_i \rightarrow n_j}} \quad (43)$$

For parallel paths between two nodes  $n_i$  and  $n_j$ , the total resistance is

$$\frac{1}{R_{\text{Tot},n_i \rightarrow n_j}} = \frac{1}{R_{\text{I},n_i \rightarrow n_j}} + \frac{1}{R_{\text{II},n_i \rightarrow n_j}} + \dots \quad (44)$$

If the resistance along one of the paths is much larger than that in the other, it would be safe to assume that the path contributes little to the OR rate and, hence, that path may be eliminated from the RR network. This, of course, concomitantly simplifies the mechanism by pointing out the reaction steps that are kinetically inconsequential. As a corollary, if both pathways are significant, it implies that each of the parallel branches has a resistance, or slow steps, of comparable order. The procedure involves analysis of the resistances in all ERs and may result in a very significant simplification of the RR network. Recall that each link that is eliminated reduces the number of independent RRs by one.

The next step is to determine the relative resistances of the steps in a given sequence. At QSS, the rates of all reactions in a sequence are the same; therefore, this may be accomplished by a comparison of the affinity drop over each of the branches in a sequence

$$\frac{A_1}{A_2} = \frac{R_1}{R_2} \quad (45)$$

The order of series steps in a branch is unimportant, from the point of view of this analysis, although it is usually significant mechanistically. The slowest steps are those with significant affinity drops, whereas those with affinity drops that approach zero may be considered to be at pseudo-equilibrium. This might allow the development of explicit rate expressions for the OR. However, numerically, the rate of the OR may be alternately calculated from  $r_{\text{OR}} = A_{\text{OR}}/R_{\text{OR}}$ .

Clearly, the aforementioned simplification and reduction algorithm should provide a quantitative value for the tolerance that is appropriate for retaining or discarding various reaction steps. The problem may be simply formulated in terms of the overall resistance of the RR network. Thus, a reaction step may be neglected if the change in the overall resistance of the RR network does not exceed a chosen tolerance.

## 8. Concluding Remarks

The determination of the chemical kinetics and mechanism for an overall reaction (OR) based on a realistic set of elementary steps with predicted rate constants is an important current goal. Frequently, the OR is composed of sequential as well as parallel pathways and more than one significant reaction route (RR). All of these RRs, in principle, affect the OR rate, because frequencies of all unit steps are finite. Furthermore, the slow steps and the significant RRs may change with reaction conditions. However, the traditional approach has been to

assume a rate-determining step (RDS), with all other steps being at quasi-equilibrium.<sup>28</sup> Furthermore, parallel pathways are rarely considered. More recently, the approach of microkinetics<sup>30</sup> has allowed these assumptions to be abandoned. However, the numerical approach obscures an understanding of the key steps involved in the reaction mechanism and their interconnections.

What is proposed here is a new algorithm for the rigorous visualization and analysis of RR networks that is based on the combined use of graph theory, RR theory, microkinetic analysis, and an analogy with electrical circuits. The topological features of the RR graph are determined from an enumeration of the RRs<sup>35</sup> or simply from the application of the quasi-steady state (QSS) condition. The approach is very different from that for reaction graphs used so far in the literature. In our approach, the reactions are represented by graph branches, whereas the nodes simply represent the manner in which these reaction steps are connected to provide the various pathways followed by reactants in formation of the products, whereas in the literature, the nodes represent individual intermediate or terminal species. Our approach results in intuitively obvious graphs that are also amenable to quantitative analysis based on the use of electrical circuit analogy and application of Kirchhoff's laws of current and potential (adapted to reaction rate and affinity). Finally, making use of numerical microkinetic results, a systematic and rigorous analysis and simplification of the resulting mechanism and kinetics can be performed. Such an understanding of significant elementary steps and RRs is crucial not only in developing practical rate expressions for performance analysis and design of reactors, but also in forming the basis of rational design of selective and active catalysts. Furthermore, the resulting RR graphs are amenable to depicting energy diagrams, with nodes representing unique values of thermodynamic state variables.

The theoretical approach and algorithm described in Part I is illustrated via application to the cases of an enzyme reaction and a surface-catalyzed reaction in Part II of this series of papers.<sup>40</sup>

## References and Notes

- (1) Balaban, A. T. In *Reaction Graphs*; Bonchev, D., Mekenyan, O., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994; p 137.
- (2) Temkin, O. N.; Zeigarnik, A. V.; Bonchev, D. G. *Chemical Reaction Networks: A Graph-Theoretical Approach*; CRC Press: New York, 1996.
- (3) Fan, L. T.; Bertok, B.; Friedler, F. *Comput. Chem.* **2002**, 26, 265.
- (4) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983.
- (5) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. *Ind. Eng. Chem. Res.* **1994**, 33, 790.
- (6) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. *Comput. Chem. Eng.* **1996**, 20, 113.
- (7) Susnow, R. G.; Dean, M. A.; Green, W. H.; Peczak, P.; Broadbelt, L. *J. Phys. Chem. A* **1997**, 101, 3731.
- (8) Ratkiewicz, A.; Truong, T. N. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 36.
- (9) Christiansen, J. A. *Adv. Catal.* **1953**, 5, 311.
- (10) King, E. L.; Altman, C. *J. Phys. Chem.* **1956**, 93, 6676.
- (11) Temkin, M. I. *Dokl. Akad. Nauk SSSR* **1963**, 152, 156.
- (12) Temkin, M. I. *Dokl. Akad. Nauk SSSR* **1965**, 165, 615.
- (13) Temkin, M. I. *Adv. Catal.* **1979**, 28, 173.
- (14) Yatsimirskii, K. B. *Applications of the Graph Method in Chemistry*; Khimiya: Moscow, 1971.
- (15) Balaban, A. T.; Farcasiu, D.; Banica, R. *Rev. Roum. Chim.* **1966**, 11, 1205.
- (16) Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1985**, 25, 334.
- (17) Clarke, B. L. *Cell Biophys.* **1988**, 12, 237.
- (18) Sen, A. K. *Biochem. J.* **1991**, 275, 253.
- (19) Heinrich, R.; Schuster, S. *The Regulation of Cellular Systems*; Chapman and Hall: New York, 1996.
- (20) Schilling, C. H.; Palsson, B. O. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 4193.
- (21) Beard, D. A.; Liang, S.-D.; Qian, H. *Biophys. J.* **2002**, 83, 79.



- (22) Oster, G. F.; Perelson, A. L.; Katchalsky, A. *Q. Rev. Biophys.* **1973**, 6, 1.
- (23) Oster, G. F.; Perelson, A. L. *IEEE Trans. Circuits Syst.* **1974**, CAS-21, 709.
- (24) Bockris, J. O'M.; Srinivasan, S. *Fuel Cells: Their Electrochemistry*; McGraw-Hill: New York, 1969.
- (25) Shiner, J. S. *Adv. Thermodyn.* **1992**, 6, 248.
- (26) Broadbelt, L. J.; Snurr, R. Q. *Appl. Catal., A* **2000**, 200, 23.
- (27) Shustorovich, E.; Sellers, H. *Surf. Sci. Rep.* **1998**, 31, 1.
- (28) Hougen, O. A.; Watson, K. M. *Chemical Process Principles*, Vol. III; Wiley: New York, 1947.
- (29) Boudart, M.; Djega-Mariadassou, G. *Kinetics of Heterogeneous Catalytic Reactions*; Princeton University Press: Princeton, NJ, 1984.
- (30) Dumesic, J. A.; Rudd, D. F.; Aparicio, L. M.; Rekoske, J. E.; Trevino, A. A. *The Microkinetics of Heterogeneous Catalysis*; American Chemical Society: Washington, DC, 1993.
- (31) Stoltze, P. *Prog. Surf. Sci.* **2000**, 65, 65.
- (32) Fishtik, I.; Callaghan, C. A.; Datta, R. Reaction Network Analysis: The Kinetics and Mechanism of Water-Gas-Shift Reaction on Cu(111). In *Computational Material Science 15*; Leszczynski, J., Ed.; Elsevier: Amsterdam, 2004.
- (33) Fishtik, I.; Datta, R. *Chem. Eng. Sci.* **2000**, 55, 4029.
- (34) Fishtik, I.; Datta, R. *Ind. Eng. Chem. Res.* **2001**, 40, 2416.
- (35) Fishtik, I.; Datta, R. *Surf. Sci.* **2002**, 512, 229.
- (36) Callaghan, C. A.; Fishtik, I.; Datta, R.; Carpenter, M.; Chmielewski, M.; Lugo, A. *Surf. Sci.* **2003**, 541, 21.
- (37) Milner, P. C. *J. Electrochem. Soc.* **1964**, 111, 228.
- (38) Horiuti, J.; Nakamura, T. *Adv. Catal.* **1967**, 17, 1.
- (39) Happel, J.; Sellers, P. H. *Adv. Catal.* **1983**, 32, 272.
- (40) Fishtik, I.; Callaghan, C. A.; Datta, R. *J. Phys. Chem. B* **2004**, 108, 5683.
- (41) De Donder, T.; Ryssselberghe, P. *Thermodynamic Theory of Affinity*; Stanford University Press: Palo Alto, CA, 1936.
- (42) Ryssselberghe, P. *Chem. Eng. Sci.* **1967**, 22, 706.
- (43) Laidler, K. J. *Chemical Kinetics*, 3rd ed.; Addison-Wesley: Reading, MA, 1987.
- (44) Wagner, C. *Adv. Catal.* **1970**, 21, 323.
- (45) Deo, N. *Graph Theory with Applications to Engineering and Computer Science*; Prentice Hall: Englewood Cliffs, NJ, 1974.
- (46) Balabanian, N.; Bickart, T. A. *Electrical Network Theory*; Wiley: New York, 1969.
- (47) Clay, R. *Nonlinear Networks and Systems*; Wiley: New York, 1971.
- (48) Tellegen, B. D. H. *Philips Res. Rep.* **1952**, 7, 259.
- (49) Gleiss, P. M.; Stadler, P. F.; Wagner, A.; Fell, D. A. *Adv. Complex Syst.* **2001**, 1, 1.
- (50) Okino, M. S.; Mavrovouniotis, M. L. *Chem. Rev.* **1998**, 98, 391.
- (51) Boudart, M. *AIChE J.* **1972**, 18, 465.