ISSN 2820-6657 (electronic edn.)
DISCRETE MATHEMATICAL CHEMISTRY 1 (2025) #P1.01
https://doi.org/10.26493/2820-6657.8.d8b
(Also available at http://dmc-journal.eu)

Every atom-atom map for neutral molecules can be explained by electron pushing diagrams*

Christoph Flamm
Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria

Stefan Müller
Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Wien, Austria

Peter F. Stadler †
Bioinformatics Group, Department of Computer Science & Interdisciplinary Center for Bioinformatics and
Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig and
School of Embedded Composite Artificial Intelligence, Leipzig University, Härtelstraße 16–18, D-04107 Leipzig, Germany and
Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany and
Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria and
Facultad de Ciencias, Universidad National de Colombia; Bogotá, Colombia and
Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe NM 87501, USA

Received 22 November 2023, accepted 13 June 2024, published online 8 September 2025

Abstract

Chemical reactions can be understood as transformations of multigraphs (molecules) that preserve vertex labels (atoms) and degrees (sums of bonding and non-bonding electrons), thereby implying the atom-atom map of a reaction. The corresponding reaction mechanism is often described by an electron pushing diagram that explains the transformation by consecutive local relocations of individual edges (electron pairs). Here, we show that every degree-preserving map between multigraphs, and thus every atom-atom map, can be generated by cyclic electron pushing. Moreover, it is always possible to decompose such

*This research was funded in part by the Novo Nordisk Foundation, grant NNF21OC0066551 "MATOMICS" to CF and PFS, and the Austrian Science Fund (FWF), grant 10.55776/P33218 to SM.
†Corresponding author.

This work is licensed under https://creativecommons.org/licenses/by/4.0/

an explanation into electron pushing diagrams involving only four electron pairs. This in turn implies that every reaction can be decomposed into a sequence of elementary reactions that involve at most two educt molecules and two product molecules. Hence, the requirement of a mechanistic explanation in terms of electron pushing and imaginary transition states involving only a small number of bonds does not impose a combinatorial constraint on the feasibility of hypothetical chemical reactions.

Keywords: Chemical reaction networks, graph transformations, reaction mechanisms.

Math. Subj. Class. (2020): 05C92, 92E10

1 Introduction

Chemical reaction networks can be viewed abstractly as directed hypergraphs with compounds as vertices and reactions as directed hyperedges. However, not every directed hypergraph has a chemical interpretation. In particular, the conservation of mass, and in more detail of atoms in chemical reactions, implies non-trivial constraints. Most importantly, they ensure the existence of atom-atom maps that guarantee that reactions can be written as re-assignments of bonds [21]. Here, we ask whether there are constraints on chemically feasibly atom-atom maps that derive from the concept of electron pushing diagrams, a low-level mechanistic description of chemical reactions as stepwise local relocations of electron pairs. We shall see that this is not the case: at least mathematically, every atom-atom map admits such a low-level mechanistic explanation.

A structural formula represents a chemical species as a (connected) graph, whose vertices are labeled by atom types and edges refer to chemical bonds. *Lewis formulas* [19] are equivalent to vertex-labeled multigraphs in which each bonding electron pair is represented as an individual edge, and each non-bonding electron pair as a (self-)loop. The electron pair of a bond is considered to be divided up between the two atoms that it connects, while a non-bonding pair is localized at a single atom. This representation of molecules agrees with the matrix formalism of Dugundji & Ugi [7]. Since the number of electrons in the outer shell is usually preserved, the atom type defines the degree of a vertex in the multigraph, matching Frankland's "atomicity" and conforming to the IUPAC term "valency" [20].

When molecules are modeled as Lewis formulas, chemical reaction mechanisms can be described as *electron pushing diagrams* (EPD) [18]. Their simplest form describes the movement of electrons in terms of stepwise local movements of electron pairs. In terms of conventional chemical notation, this amounts to considering only "curved arrows" in the conventional chemical notation [2]. The elementary step in an EPD is the replacement of an electron pair (edge) xy by an electron pair (edge) yz. Of course, we require $x \neq z$ since otherwise no change in the molecule would have incurred. However, we may have either $x = y$ or $y = z$, corresponding to transitions that convert a non-bonding to a bonding electron pair or *vice versa*. Since this operation increases the degree at z (and decreases the degree at x), an electron pair at z needs to move in the next step in order to re-establish valency at z. This process progresses until the degree deficit at x is eventually compensated, see Figure 1. The reaction therefore can be understood as a cyclic sequence of alternating steps of deleting and inserting consecutive edges in the molecular graph. By construction,

E-mail addresses: xtof@tbi.univie.ac.at (Christoph Flamm), st.mueller@univie.ac.at (Stefan Müller), studla@bioinf.uni-leipzig.de (Peter F. Stadler)

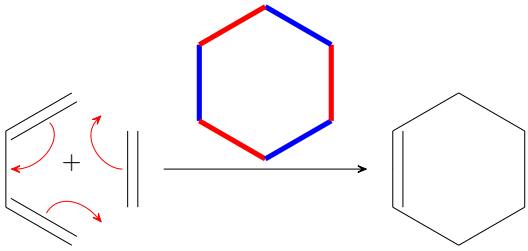


Figure 1: Electron pushing representation of a Diels-Alder reaction. Above the reaction arrow, the difference multigraph (DMG) or Imaginary Transition State (ITS) graph Δ is shown with blue edges for $\sigma(e) = +1$ and red edges for $\sigma(e) = -1$. Tails and heads of electron pushing arrows translate to red and blue edges in the ITS, respectively.

the application of an EPD preserves the total number of unique valence electrons at each atom, i.e., the vertex degree, with each loop accounting for 2 electrons.

In a non-mechanistic setting, chemical reactions are modeled in terms of their atom-atom maps (AAMs), i.e., the one-to-one correspondence of atoms in the educt (reactant) multigraph and the product multigraph. The AAM implicitly describes the net change of bonds between educts and products. In the terminology of chemical reaction network theory, the educt and product graphs G and H correspond to *complexes* [17], and their connected components are the educt and product molecules. AAMs thus are fine-grained descriptions of the (directed) edges in the complex-reaction graph. The net effect of a chemical reaction is captured by a "difference multigraph" (DMG) in which changes of bond orders are recorded. In essence, this graph is a graph-theoretic version of the R-matrix in the formalism of Dugundji and Ugi [7] and can be seen as a variation of the Imaginary Transition State (ITS) graph [11] and the Condensed Graph of a Reaction (CGR) [16], from which it differs only by the use of signed multiple edges instead of specific edge-labels.

The application of an EPD to an educt multigraph G produces a multigraph H on the same vertex set V, and the corresponding AAM is the identity on $V = V(G) = V(H)$. Since every push (movement of an electron pair) is eventually compensated, the EPD model of a reaction preserves the degree (valency) at each vertex. In this contribution we ask whether the converse is also true: Is there always a sequence of EPDs that explains the bond changes in a given degree-preserving AAM? Indeed, we shall show that the DMG of a degree-preserving AAM can always be explained by a set of disjoint EPDs.

In the following, we restrict ourselves to neutral molecules and electron pair pushing. It therefore suffices to model molecules as multigraphs and reactions as degree-preserving relocations of edges.

2 Mathematical preliminaries

Let $G = (V, E)$ be a finite undirected multigraph and denote by $\mathfrak{C}(G)$ its connected components. We consider the elements of E as distinct; nevertheless, we will write "an edge xy" to mean an "element $e \in E$ incident to $x, y \in V$". For a pair of vertices $x, y \in V$ we define $m(x, y) = m(y, x)$ as the number of edges connecting x and y. In particular, $m(x, y) = 0$ if x and y are not adjacent. Similarly, $m(x, x)$ denotes the multiplicity of loops at vertex x. Thus $m(x, x) = 0$ if there is not loop at $x \in V$. A simple graph is a multigraph without loops and multiple edges, i.e., $m(x, x) = 0$ and $m(x, y) \in \{0, 1\}$ for all $x \neq y$. Following the usual convention for multigraphs, the degree of a vertex is defined

as

$$\deg(x) \coloneqq \sum_{y \in V \setminus \{x\}} m(x,y) + 2m(x,x). \tag{2.1}$$

Loops count twice because both of their endpoints are incident with x. Chemically, we can interpret $m(x,y)$ with $x \neq y$ as the bond order, while $m(x,x)$ denotes the number of non-bonding electron pairs at x. Thus, the degree $\deg(x)$ describes the number of electrons at an atom, which we consider to be a constant.

Lewis formulas have vertex labels designating the atom type and atom-atom maps (AAMs) are defined as label-preserving bijections $\alpha\colon V(G) \to V(H)$ between two Lewis formulas G and H. For the following mathematical results, however, the labels themselves are not relevant. We will only make use of the fact that the total number of unique electrons in the outer (valency) shell of an atom is considered to be fixed and thus reactions preserve the vertex degree. Instead of atom-atom maps, we therefore consider degree-preserving bijections $\alpha\colon V(G) \to V(H)$ such that $\deg_H(\alpha(x)) = \deg_G(x)$ for all $x \in V(G) = V(H)$.

Our key construction is a graph-theoretical reformulation of the R-matrix in the theory of Dugundji & Ugi [7]:

Definition 2.1. Let G and H be multigraphs and $\alpha\colon V(G) \to V(H)$ be a degree-preserving bijection. Moreover, for vertices x and y, let $\delta(x,y) \coloneqq m_H(\alpha(x),\alpha(y)) - m_G(x,y)$, $m(x,y)_\Delta \coloneqq |\delta(x,y)|$, and $\sigma_\Delta(x,y) \coloneqq \operatorname{sgn}(\delta(x,y))$, where $\operatorname{sgn}\colon \mathbb{R} \to \{-1,0,+1\}$ is the sign function. Then, the *difference multigraph (DMG)* $\Delta(G,H,\alpha)$ has vertex set $V = V(G) = V(H)$ and $m_\Delta(x,y)$ edges connecting x and y. An edge xy is labeled by $\sigma_\Delta(x,y)$.

In particular, $\Delta(G,H,\alpha)$ contains no edge between x and y if the multiplicity of the edges between two vertices x and y in G coincides with the multiplicity of the edges between the images $\alpha(x)$ and $\alpha(y)$ in H. The DMG describes the net change of bond orders that result from the chemical reaction $\mathfrak{C}(G) \to \mathfrak{C}(H)$. The multiplicities in G, H, and $\Delta \coloneqq \Delta(G,H,\alpha)$ satisfy

$$m_H(\alpha(x),\alpha(y)) - m_G(x,y) = \sigma_\Delta(x,y) \cdot m_\Delta(x,y), \tag{2.2}$$

for all x, y. We note that setting $B_{xy} \coloneqq m_G(x,y)$, $E_{xy} \coloneqq m_H(\alpha(x),\alpha(y))$, and $R_{xy} \coloneqq \sigma_\Delta(x,y) \cdot m_\Delta(x,y)$ recovers the representation of a reaction as $B + R = E$ in [7].

In the following, it will be useful to count the increasing and decreasing bond orders separately. To this end we introduce

$$d'_+(x) \coloneqq \sum_{\substack{y \neq x \\ \sigma_\Delta(x,y) = +1}} m_\Delta(x,y) \quad \text{and} \quad d'_-(x) \coloneqq \sum_{\substack{y \neq x \\ \sigma_\Delta(x,y) = -1}} m_\Delta(x,y). \tag{2.3}$$

To account also for the loops, we introduce

$$\begin{aligned} d_+(x) &\coloneqq \begin{cases} d'_+(x) & \text{if } \sigma_\Delta(x,x) \leq 0 \\ d'_+(x) + 2m_\Delta(x,x) & \text{if } \sigma_\Delta(x,x) > 0, \end{cases} \\ d_-(x) &\coloneqq \begin{cases} d'_-(x) & \text{if } \sigma_\Delta(x,x) \geq 0 \\ d'_-(x) + 2m_\Delta(x,x) & \text{if } \sigma_\Delta(x,x) < 0. \end{cases} \end{aligned} \tag{2.4}$$

We first state a simple property of DMGs:

Lemma 2.2. *Every DMG Δ satisfies $d_+(x) = d_-(x)$ for all $x \in V$, and hence $\deg_\Delta(x)$ is always even.*

Proof. Since α is a degree-preserving bijection, we have $\deg_G(x) - \deg_H(\alpha(x)) = 0$ and thus

$$\begin{aligned} 0 &= \sum_{y \in V\setminus\{x\}} m(\alpha(x),\alpha(y)) - m(x,y) + 2\big(m(\alpha(x),\alpha(x)) - m(x,x)\big) \\ &= \sum_{y \in V\setminus\{x\}} \sigma_\Delta(x,y)\cdot m_\Delta(x,y) + 2\sigma_\Delta(x,x)\cdot m_\Delta(x,x) \\ &= d'_+(x) - d'_-(x) + 2\sigma_\Delta(x,x)\cdot m_\Delta(x,x), \end{aligned}$$

where we separate the positive and negative edges in the sum. If $d'_+(x) > d'_-(x)$, then $\sigma_\Delta(x,x) = -1$ and further $d_+(x) = d'_+(x)$ and $d_-(x) = d'_-(x) + 2m_\Delta(x,x)$, which yields $d_+(x) = d_-(x)$. An analogous argument applies if $d'_+(x) < d'_-(x)$. Finally, if $d'_+(x) = d'_-(x)$, then $m_\Delta(x,x) = 0$. Taken together, we always have $d_+(x) = d_-(x)$. Moreover, $\deg_\Delta(x) = d_+(x) + d_-(x) = 2d_+(x)$, and hence $\deg_\Delta(x)$ is even. □

We define a *walk* on a multigraph as a sequence of distinct edges such that any two successive edges in the sequence share a vertex. A walk on a multigraph with sign function $\sigma\colon E \to \{+1,-1\}$ is *alternating* if consecutive edges have opposite signs, i.e., $\sigma(e_k)\cdot\sigma(e_{k+1}) = -1$ for all k (mod n).

3 Existence of alternating closed walks

In order to investigate the DMG in some more detail, it will be useful to consider an equivalent simple graph.

Definition 3.1. For every signed multigraph Δ denote by $A(\Delta)$ the simple graph obtained by subdiving each edge of Δ by the insertion of two *subdivision vertices*. All other vertices of $A(\Delta)$ will be a referred to as Δ-*vertices*. The sign $\sigma(e')$ of an edge in $A(\Delta)$ that is incident to a Δ-vertex equals the sign $\sigma(e)$ of the edge from which e' derives by subdivision. The sign $\sigma(e'')$ of an edge e'' connecting two subdivision vertices is the opposite of the edge from which it derives, i.e., $\sigma(e'') = -\sigma(e)$.

In other words, each edge e in Δ is replaced by a path $P_3(e)$ of length 3 in $A(\Delta)$. Moreover, the two terminal edges of $P_3(e)$ are labeled with $\sigma(e)$, while the middle edge of $P_3(e)$ is assigned $-\sigma(e)$, see Figure 2 for an illustration. From a mathematical point of view, the construction of $A(\Delta)$ is related to dominating induced matchings (see [1] and the references therein). Interestingly, $A(\Delta)$ also corresponds in chemical terms to constructing "full" carbomers, see e.g. [5].

Lemma 3.2. *Let Δ be a signed multigraph. If W^* is a closed walk in $A(\Delta)$ then there is uniquely defined closed walk W in Δ such that W^* is obtained from W by replacing each edge $e \in W$ by the path $P_3(e)$. Moreover, if W^* is alternating, then W is also alternating.*

Proof. First we note that the edge set of $A(\Delta)$ can be partioned into the edge sets of $P_3(e)$ for $e \in E(\Delta)$. Furthermore, since the subdivision vertices have degree two in $A(\Delta)$, we have either $P_3(e) \subseteq W^*$ or $P_3(e) \cap W^* = \emptyset$ (with sequences treated as sets). Since $P_3(e)$

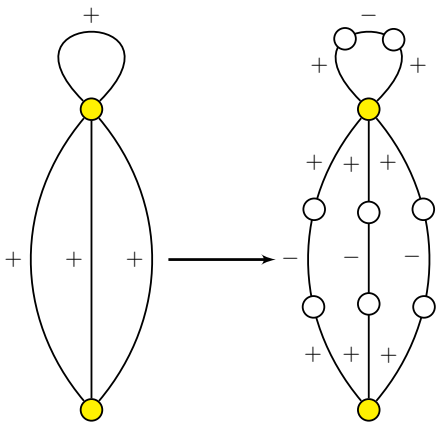


Figure 2: Transformation of a signed multigraph Δ into a signed auxiliary graph $A(\Delta)$. Every edge of Δ is replaced by a a signed path P_3. At each Δ-vertex, the number of positive and the number of negative edges remains unchanged.

is incident with Δ-vertex x in $A(\Delta)$ if and only if e is incident with x in Δ, every closed walk in $A(\Delta)$ corresponds to a closed walk in Δ by contracting $P_3(e)$ to the single edge e.

Consider two consecutive edges in W^*. By assumption, they have opposite sign. If they lie in the same $P_3(e)$, then one of them is the middle edge of $P_3(e)$. Otherwise, they lie in two different paths P_3, say e'_1 in $P_3(e_1)$ and e'_2 in $P_3(e_2)$, and the vertex x between them is a Δ-vertex. Thus, $\sigma(e'_1) = \sigma(e_1)$ and $\sigma(e'_2) = \sigma(e_2)$, and $\sigma(e_1) = -\sigma(e_2)$ implies $\sigma(e'_1) = -\sigma(e'_2)$. That is, W is an alternating closed walk in Δ. □

As an immediate consequence, we observe:

Corollary 3.3. *Let Δ be a signed multigraph and $A(\Delta)$ be the corresponding simple graph. The edge set of Δ is the disjoint union of the edge sets of alternating closed walks on Δ if and only if the edge set of $A(\Delta)$ is the disjoint union of the edge sets of alternating closed walks on $A(\Delta)$.*

The auxiliary graph $A(\Delta)$ retains the signs of the edges in the following sense:

Lemma 3.4. *Let Δ be a signed multigraph and $A(\Delta)$ be the corresponding simple graph. Every Δ-vertex x in $A(\Delta)$ is incident with $d_+(x)$ positive edges and $d_-(x)$ negative edges. Moreover, every subdivision vertex in $A(\Delta)$ is incident with exactly one positive and one negative edge.*

Proof. By construction, if e is an edge between x and $y \neq x$ in Δ, then there is exactly one edge in $P_3(e)$ incident with x in $A(\Delta)$, and if e is loop at x, then there are exactly two edges incident with x. In each case, the edges incident with x in $A(\Delta)$ have the same sign as those in Δ. Since every loop in Δ contributes two to $d_+(x)$ resp. $d_-(x)$, cf. Equation (2.4), the assertion follows. □

Corollary 3.5. *Let G and H be multigraphs, $\alpha\colon V(G) \to V(H)$ be a degree-preserving bijection, and Δ be the DMG. Every vertex of the corresponding simple graph $A(\Delta)$ is incident to as many positive as negative edges.*

In the following we derive a characterization of simple graphs with signed edges in terms of alternating closed walks that is reminiscent of Euler's solution of the Königsberg Bridges Problem [8]. We start with a simple observation:

Lemma 3.6. *Let A be a simple graph with non-empty edge set E and edge labels $\sigma\colon E \to \{+1,-1\}$ such that every vertex is incident to as many positive as negative edges. Then there exists an alternating closed walk in G.*

Proof. We construct an alternating closed walk in A starting at the vertex x_0 with a positive edge. At each step, we extend the walk by traversing an edge that has not been used before. If a vertex $y \neq x_0$ has been entered along an edge e with sign $\sigma(e)$, then by assumption there is a previously unused edge e' with $\sigma(e') = -\sigma(e)$ along which the walk can leave y again. Since A is finite, the walk eventually encounters x_0 again. If it enters x_0 along a negative edge, we have obtained the desired alternating closed walk. If x_0 is entered along a positive edge, then the walk so far contains two more positive than negative edges incident with x_0. Repeating the above construction (starting at x_0 this time with a remaining negative edge), we conclude that the walk will eventually return to x_0 along a negative edge. Thus, there is always an alternating closed walk. □

Theorem 3.7. *Let A be a simple graph with non-empty edge set E and edge labels $\sigma\colon E \to \{+1,-1\}$. Then E is the disjoint union of alternating closed walks if and only if every vertex is incident to as many positive as negative edges.*

Proof. First we note that the condition is necessary. Suppose there is a vertex x with $d_+(x) \neq d_-(x)$. Since every alternating closed walk W_i necessarily traverses the same number of positive and negative edges, successively removing all edges in the walks W_i leaves an excess of positive or negative edges at x. Thus $E \setminus \bigcup_i W_i \neq \emptyset$, contradicting the assumption that the W_i form a partition of E.

In order to see that the condition is sufficient, we use Lemma 3.6 to establish the existence of an alternating closed walk in A, whose edge set we denote by W. Denote by $A' = (V, E \setminus W)$ the graph obtained by deleting the edges along the alternating closed walk. Since every vertex $x \in V$ is incident to as many positive as negative edges in both $E(A)$ and W, this is also true for $E(A') = E \setminus W$. The graph A' therefore contains an alternating closed walk W'. Repeating the argument yields a partition of E into the edge sets of alternating closed walks. □

The connection to Eulerian graphs is that the vertex degree of the graph in Theorem 3.7 is even, and thus the graph is Eulerian if and only if it connected. Assuming that A is connected, one can combine the alternating closed walks to a single alternating closed walk, see e.g. [10], arriving at the following variant of Hierholzer's algorithm [15]: If two alternating closed walks W_1 and W_2 share a vertex x, we first follow W_1 from an arbitrary starting point to x, and then traverse W_2 choosing the direction such that the combined walk is alternating. After returning to x, we follow the unused part of W_1 from x to the starting point. This results in a single alternating closed walk W' covering exactly the edges of W_1 and W_2. If W' shares a vertex x' with some alternating closed walk W_3, the procedure is repeated, resulting in a single alternating closed walk W'' comprising the edges of W_1, W_2, and W_3. One continues until all W_i are absorbed into a single alternating closed walk. Connectedness of A implies that this final walk contains all edges and thus is an *alternating Euler tour*. Otherwise, Hierholzer's algorithm produces an alternating Euler tour for each connected component of A. This argument implies:

Corollary 3.8. *Let A be a simple graph with non-empty edge set E and edge labels $\sigma\colon E \to \{+1,-1\}$. Then A admits an alternating Euler tour if and only if A is connected and every vertex is incident to as many positive as negative edges.*

4 Explanations by electron pushing diagrams

Returning to atom-atom maps (AAMs) and corresponding DMGs $\Delta(G, H, \alpha)$, we can use Corollary 3.5 to apply Theorem 3.7. An essentially equivalent statement can be found in [9, Proposition 1] for a slightly different setting with a sketched proof.

Corollary 4.1. *The edge set of a DMG Δ can be partitioned into the edge sets of alternating closed walks. Moreover, if Δ is connected, it admits an alternating Euler tour.*

Every alternating closed walk can be directly interpreted as an electron pushing diagram (EPD) that involves only electron pairs. To this end, choose an arbitrary vertex x and a starting edge $e = xy$ with $\sigma(e) = -1$. The corresponding electron pair is considered to be "pushed" from xy to the next edge yz along the walk. Clearly, every EPD interpretation of an alternating closed walk applied to the educt graph G yields the same graph G', describing a reaction intermediate. Thus, we can derive an *EPD explanation* for every AAM as follows: First, construct the DMG $\Delta := \Delta(G, H, \alpha)$, then choose a sequence of alternating closed walks whose edge sets partition $E(\Delta)$. The sequences of negative and positive edges of each walk, starting from any of its vertices, form an EPD. We say that the AAM is *explained* by applying any of these EPDs to the educt multigraph G. Of course, the order of the closed walks can be chosen arbitrarily and yields the same product graph.

It is important to note that this notion of an EPD explanation is a formal mathematical definition. It does not imply by any means that a particular EPD explanation is chemically reasonable. However, we shall see below, that a given AAM admits many EPD explanations, in general.

By definition, alternating closed walks have the property that every (multi-)edge is traversed only once. By construction, all edges of Δ connecting two given vertices x and y have the same sign. Hence, the EPD explanation of the AAM is monotone in the sense that the bond order, i.e., the multiplicity $m(x, y)$, is either always increased or always decreased when an alternating closed walk traverses xy. Reactions in which several vertex pairs are visited twice along a single alternating closed walk do exist. A rather spectacular example is the Möbius type coarctate double cyclization in [14, Figure 11].

Each alternating closed walk in the construction used to prove Corollary 4.1 traverses only edges of the DMG, i.e., edges that are present in the educt multigraph, the product multigraph, or both. The EPD explanations produced in this manner therefore never insert a new bond and later delete it again.

However, the walks in this construction may be very long. Such walks are often chemically unrealistic when interpreted as the mechanism of a single-step reaction. They could appear, however, as the "superposition" of multiple elementary reactions. We ask, therefore, if it is always possible to produce any union of alternating closed walks as a sequence of EPD explanations with restricted lengths of their underlying alternating closed walks. In other words, we ask whether every AAM can be explained as a sequence of reactions with mechanisms that are simple as far as their EPDs are concerned. Clearly, the minimum length of an alternating cycle that corresponds to a change in bonds is $2k = 4$. A cyclic EPD of length 2 would correspond to inserting and subsequently deleting the same edge. This does not change the graph to which it is applied, and hence does not correspond to a reaction.

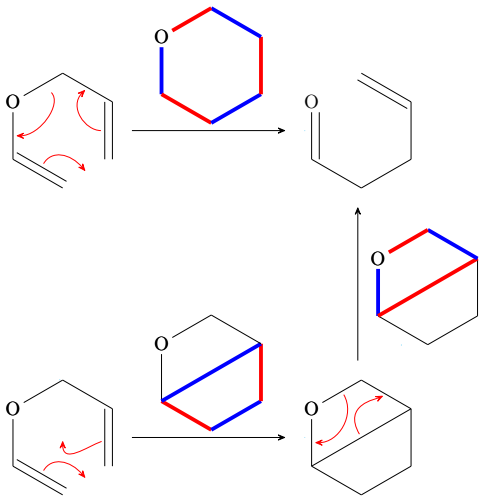


Figure 3: Claisen rearrangement of vinyl-allyl ethers. The 6-cyclic imaginary transition state (upper path), can be decomposed into two 4-cyclic imaginary transition states (lower path) to yield the γ,δ-unsaturated carbonyl compound.

Lemma 4.2. *Let W be an alternating closed walk of length $|W| = 2k \geq 6$ with the first edge being negative, let $\hat{W} = (x_0, x_1, \ldots, x_{2k-1}, x_0)$ be the sequence of traversed vertices, and set*

$$\begin{aligned} \hat{W}_1 &\coloneqq (x_0, x_1, \ldots, x_{2i-1}, x_0) \\ \hat{W}_2 &\coloneqq (x_0, x_{2i-1}, x_{2i}, \ldots, x_{2k-1}, x_0) \end{aligned} \tag{4.1}$$

with $2 \leq i < k$. Then the product graph H obtained by applying W to an educt graph G is also obtained by either first applying W_1 and then W_2 or by first applying W_2 and then W_1.

Proof. First, note that edges $x_{2j}x_{2j+1}$ are negative, and edges $x_{2j+1}x_{2j+2}$ (mod $2k$) are positive. Further, $\hat{W}_1$ and $\hat{W}_2$ are alternating closed walks with length $2i$ and $2(k-i+1)$, respectively. In particular, $\hat{W}_1$ has the fictitious last step $x_{2i-1}x_0$, which is positive (since the first step x_0x_1 is negative), even if there is no such edge in Δ or if this edge has a negative sign. Analogously, the fictitious first step x_0x_{2i-1} of $\hat{W}_2$ is negative (since the last step $x_{2k-1}x_0$ is positive). Now, if x_0x_{2i-1} is not an edge of G, then $\hat{W}_1$ is applicable to G since all changes caused by $\hat{W}_1$ (other than x_0x_{2i-1}) yield changes in multiplicities that are smaller than the changes caused by $\hat{W}$. Moreover, $\hat{W}_1$ adds the edge x_0x_{2i-1} to G. Hence $\hat{W}_2$ is applicable to the resulting graph. Furthermore, since the union of $\hat{W}_1$ and $\hat{W}_2$ equals $\hat{W}$ except for the insertion and deletion of the fictitious edge x_0x_{2i-1}, the consecutive application of $\hat{W}_1$ and $\hat{W}_2$ recovers the graph H. Similarly, if x_0x_{2i-1} is an edge in G, then $\hat{W}_2$ is applicable to G, and it reduces the corresponding bond order by 1. The original bond order is then re-established by applying $\hat{W}_1$. Again, the consecutive application of $\hat{W}_1$ and $\hat{W}_2$ recovers the graph H. □

Lemma 4.2 shows that we can split an EPD $\hat{W}$ of length $2k$ with $k \geq 3$ into two shorter EPDs $\hat{W}_1$ and $\hat{W}_2$ of length l_1 and l_2 with $l_1 + l_2 = 2(k+1)$. This splitting operation first forms an additional bond and later breakes this bond again, or *vice versa*. Since a splitting operation applies to every alternating closed path of length at least 6, Corollary 4.1 and Lemma 4.1 imply the first main result of this contribution:

Corollary 4.3. *Every atom-atom map can be explained by a finite sequence of EPDs of length* 4*.*

For an illustration, see Figure 3. We emphasize that there is no reason to assume that an explanation of an AAM by a sequence of length 4 EPDs is chemically plausible.

A chemical reaction in which the number of non-bonding electron pairs, i.e., the loops in the multigraph representing the Lewis formula, does not change at any atom is called *homovalent* [13]. Otherwise, the reaction is called *ambivalent*. The DMG Δ is loop-free if and only if the reaction is homovalent. The resulting alternating closed walks therefore are also loop-free. Corollary 4.3 immediately implies:

Corollary 4.4. *Every atom-atom map of a homovalent reaction can be explained by a finite sequence of EPDs each of which is a simple four cycle.*

5 Decomposition into elementary reactions

In an *elementary reaction*, one or more chemical species react directly to form products in a single step and with a single transition state. Collision theory implies that the probability of three or more molecules reacting simultaneously is negligible. Hence almost all reactions will be uni-molecular (isomerizations or decompositions) or bi-molecular. Even classical examples of termolecular reactions such as the oxidation of nitrogen monoxide, $2\,NO{+}O_2 \to 2\,NO_2$, are likely explained by sequences of uni- and bi-molecular steps [22]. Reversibility then implies that elementary reactions produce no more than two product molecules.

Definition 5.1. A chemical reaction is *combinatorially elementary* if it is of the form $A \to B$, $A \to B + C$, $A + B \to C$, or $A + B \to C + D$, where A, B, C, and D are not necessarily distinct.

In chemical terms, the combinatorially elementary reactions correspond to isomerization $A \to B$, elimination $A \to B + C$, addition $A + B \to C$, and metathesis $A + B \to C + D$.

This begs the question whether an arbitrary (complex) reaction can always be decomposed into a sequence of combinatorially elementary reactions. Next we show that this is always possible at the abstract level:

Proposition 5.2. *Every formal reaction* $\sum a_i X_i \to \sum b_i X_i$ *can be written as a sum of combinatorially elementary reactions. Let* $a = \sum_i a_i$ *and* $b = \sum_i b_i$*. Then, the number of additional intermediate species is bounded above by* $\max(a-2,0) + \max(b-2,0)$*.*

Proof. If $a, b \leq 2$, the reaction itself is combinatorially elementary. Now, let $a, b > 2$, and let T_a and T_b be rooted binary trees with a and b leaves, respectively. The leaves of the two trees can be labeled such that the chemical species X_i appears a_i times as a leaf label of T_a and b_i times as a leaf label of T_b. Each non-leaf vertex of T_a and T_b corresponds to a chemical intermediate such that in T_a a parent p is formed from its two children c_1 and c_2, i.e., by the reaction $c_1 + c_2 \to p$, while in T_b two children are formed from their parent, i.e., by the reaction $p \to c_1 + c_2$. Clearly, the trees have $a - 1$ and $b - 1$ non-leaf vertices. At the roots of the two trees, we have the intermediates r_a and r_b, respectively, and the corresponding reactions $c_1 + c_2 \to r_a$ and $r_b \to c_3 + c_4$. We link the two sets of reactions by adding a reaction between the roots of the two trees, $r_a \to r_b$, or by directly

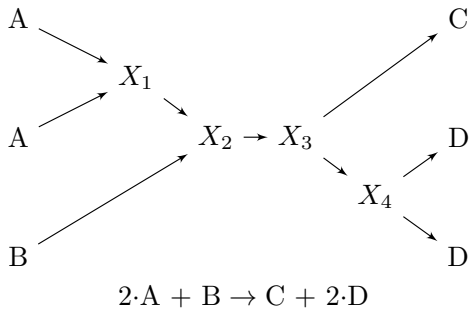


Figure 4: Representation of a complex overall reaction by elementary steps. Introducing intermediate species X_i, every left-hand side of an overall reaction (e.g. 2·A + B → C + 2·D) can be transformed into the respective right-hand side. In the proof of Proposition 5.2, the educt species are first aggregated into a single intermediate (X_2 in the shown example) by means of 2-to-1 reactions. The corresponding intermediate in the product species (X_3) is then disassembled by 1-to-2 reactions.

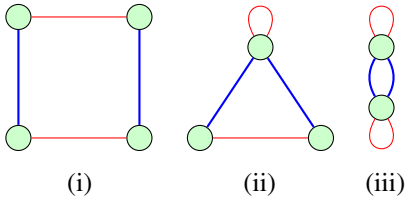


Figure 5: There are three types of alternating closed walks with four edges. (Positive edges in blue, negative edges in red.) The simple 4-cycle (left) corresponds to the simplest homovalent reaction mechanism. The triangular walk (middle) is the simplest example of an ambivalent mechanism, changing the oxidation state of the atom at the top.

adding $c_1 + c_2 \to c_3 + c_4$, thereby omitting two intermediates. In the latter case, we obtain a possible decomposition of the formal reaction into combinatorially elementary reactions with $n = a - 2 + b - 2 = a + b - 4$ intermediates (and $n + 1$ reactions among them). If either $a > 2$ and $b \leq 2$ or vice versa, then we need to consider only one tree. See also Figure 4. □

Proposition 5.2 prompts the question whether an arbitrary AAM also can be decomposed into combinatorially elementary reactions. As we shall see, Corollary 4.3 can be used to establish a stronger version of Proposition 5.2. To this end, we define the *active parts* G^a and H^a of the educt and product graphs G and H under an AAM α as the union of the connected components of G and H that contain an edge contributing to the DMG $\Delta(G, H, \alpha)$.

Lemma 5.3. *Let $\alpha\colon V(G) \to V(H)$ be an AAM explained by a single EPD of length 4. Then the active parts G^a and H^a each comprise at most two connected components.*

Proof. An EPD of length 4 corresponds to an alternating closed path of length 4. It is easy to see that there are exactly three types of alternating closed walks of length 4, see Figure 5.

(i) Δ is a simple cycle of length 4 with alternating signs. In particular, both the positive and the negative edges form a matching.

(ii) Δ is a triangle with one vertex carrying a loop. The edge not incident to that vertex has the same sign as the loop.

(iii) Δ is a pair of vertices connected by two edges and with a loop at each vertex. Both, the two edges and the two loops, respectively, have the same sign.

Clearly, the negative edges must be present in the educt graph G, while positive edges imply that the incident vertices are connected in the product graph G. In case (i), the active vertices form two pairs in both G^a and H^a, and hence they fall into at most two connected components. In case (ii), at least one pair of vertices is connected, and hence the three vertices are located in at most two connected components. In case (iii), there are only two active vertices, and hence the statement is trivial. □

Case (i) in the proof corresponds to the common pattern of a cyclic transition state, which occur in this form e.g. in [2+2] cycloadditions. Case (ii) is the simplest example of an ambivalent reaction mechanism, see e.g. [13]. It appears e.g. in cyclopropanation. Mechanism (iii) does not appear in Hendrickson's classification [13] although carbene coupling can be understood in this manner [6].

We are are not aware of a clear example of mechanism (iii) for "stable" molecules. As a net reaction, it is possible, however, to represent (iii) by two consecutive reactions from Hendrickson's classification, provided at least one of the two focal atoms is part of a larger molecule. Suppose the lower atom x in Figure 5(iii) is connected to a third atom z. Then we can explain (iii) as the superposition of two steps of type (ii) as shown in Figure 6. Here, the bond-order of xz is decreased by one in the first step and restored in the second step, hence the two-step mechanism can be used as an explanation only if x and z are connected by a bond in the reactants.

Since the connected components of G and H are the educt and product molecules, respectively, Lemma 5.3 can be rephrased in the following form:

Corollary 5.4. *Let $\alpha\colon V(G) \to V(H)$ be an AAM explained by a single EPD of length 4. Then the restriction $\alpha_a\colon V(G^a) \to V(H^a)$ is a combinatorially elementary reaction.*

Corollaries 5.4 and 4.3 imply our main result, which can be seen as a *mechanistic* refinement of Proposition 5.2.

Theorem 5.5. *Let $\alpha\colon V(G) \to V(H)$ be an AAM for the reaction $\mathfrak{C}(G) \to \mathfrak{C}(H)$. Then there is a finite sequence of intermediate graphs $G = G_0, G_1, \ldots, G_k = H$ and AAMs $\alpha_i\colon V(G_{i-1}) \to V(G_i)$ such that (i) each α_i is explained by an EPD of length 4 and hence is a combinatorially elementary reaction, and (ii) $\alpha = \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_k$.*

6 Conclusions

In [21] we have shown that every reversible reaction network has a "chemical representation" in terms of Lewis formulas (i.e., multigraphs representing molecules and AAMs for

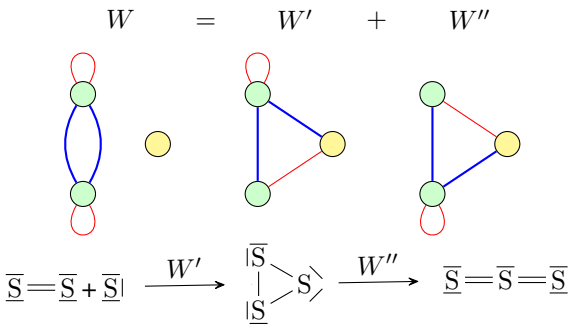


Figure 6: The 4-walk W with 2 atoms (leftmost ITS graph) can be explained as a superposition of two 4-walks W' and W'' with 3 atoms (middle). This requires that at least one of the two atoms (green) is connected to a third atom (yellow) since in the two-step mechanism one bond order is first decreased and restored in the second step. In the hypothetical trisulfur example, the yellow atom is one of the two sulfur atoms in S_2. Application of the first alternating closed 4-walk yields the cyclic S_3 as intermediate.

each reactions) if and only if it is conservative.[1] Here we extend this representability result by showing that that every degree-preserving bijection, and thus every AAM α between Lewis formulas G and H, can be explained by a sequence of (edge-disjoint) alternating closed walks on the DMG $\Delta(G, H, \alpha)$. Each of these walks in turn can be obtained by a suitably ordered sequence of EPDs of length 4, which finally correspond to combinatorially elementary reactions. In particular, the "abstract" Proposition 5.2 becomes a corollary of the "mechanistic" Theorem 5.5. We conclude that every conservative reaction network admits a representation in terms of multigraphs such that all reactions can be decomposed into elementary reactions with short alternating closed walks, i.e., short cyclic EPDs. Most known elementary reactions can be described by EPDs of lenght 6, with most other cases having length 4 or 8 [12].

The proofs given here are constructive and hence translate directly into algorithms for constructing such a representation. First, Hierholzer's algorithm can be used to decompose $E(\Delta)$ into a set of (edge-disjoint) alternating closed walks W_i with $O(|E(\Delta)|)$ effort [10]. The decomposition of the W_i into EPDs of length 4 can also be achieved in $O(|E(\Delta)|)$ total time. Applying each EPD requires only constant time, and their number is again bounded by $O(|E(\Delta)|)$. Thus the entire decomposition can be computed in linear time.

Of course, a decomposition computed in this manner will in general not be a chemically plausible explanation of the atom-atom map of a complex reaction. For instance, it is possible to formally decompose the EPD of the Diels-Alder reaction in Figure 1 and the Claisen rearrangement in Figure 3 into two 4-cycles. These shorter EPDs, however, are chemically infeasible. It is even possible that *none* of the EPD explanations that can be derived from the DMG corresponds to the chemically correct mechanism. This is in particular the case for reaction mechanisms that involve charged intermediates or free radicals. Long alternating closed walks, on the other hand, appear e.g. in EPDs of coarctate reactions [14], and thus are not always chemically unrealistic.

[1] A chemical reaction network is conservative if its stochiometric matrix has a strictly positive left kernel vector [17].

The key message of this contribution, in fact, is not to provide a method for inferring the true reaction mechanism or to help with the inference of atom-atom maps. Instead, our results imply that the basic *combinatorial* properties of chemical reaction mechanisms do not impose any fundamental constraints on the feasibility of transformations as long as the preservation of atoms (and charges) is guaranteed.

We considered here only transformations of multigraphs with loops that can be described by relocating individual edges, i.e., electron pairs, since these are already sufficient to handle reactions between neutral molecules. In principle, the formalism could also be extended to transformations that change vertex degrees and thereby introduce charges at individual atoms:

$$(CH_3)_3C-Br \longrightarrow (CH_3)_3C^{\oplus} + Br^{\ominus}$$

In general, the charge $\mathrm{ch}(x)$ at atom $x \in V$ is given as $\mathrm{ch}(x) := \mathrm{val}(x) - \deg(x)$, where $\mathrm{val}(x)$ is number of outer-shell electrons of atom x. Reaction mechanisms such Huisgen's dipolar cycloaddition [4] also require such a framework in which charges (and thus vertex degree) may change over the course of the reaction. In order to account for radicals, i.e., unpaired electrons, multigraphs can be extended to allow semi-edges, which in contrast to loops have only one end attached to a vertex. Such constructs have been studied in particular in the context of graph covers, see e.g. [3] and the references therein. However, a detailed study of charged molecules and radicals is beyond the scope of this contribution.

We note, finally, that the description of reaction mechanisms employed here is limited to alternating sequences of bond formation and bond breaking. While this approach can account e.g. for the difference between non-coarctate and coarctate reactions [14], it does not model the distinction between concerted mechanisms and non-concerted mechanism in which electron pushes are executed consecutively.

ORCID iDs

Christoph Flamm https://orcid.org/0000-0001-5500-2415
Stefan Müller https://orcid.org/0000-0002-3541-7856
Peter F. Stadler https://orcid.org/0000-0002-5016-5191

References

[1] S. Akbari, H. Baktash, A. Behjati, A. Behmaram and M. Roghani, Some results on dominating induced matchings, *Graphs Comb.* **38** (2022), 73, doi:10.1007/s00373-022-02470-6, https://doi.org/10.1007/s00373-022-02470-6.

[2] S. Alvarez, Chemistry: A panoply of arrows, *Angew. Chem. Int. Ed.* **51** (2012), 590–600, doi:10.1002/anie.201101767, https://doi.org/10.1002/anie.201101767.

[3] J. Bok, J. Fiala, N. Jedličková and P. Kratochvíl, Jan Rzążewski, List covering of regular multigraphs with semi-edges, *Algorithmica* (2023), doi:10.1007/s00453-023-01163-7, https://doi.org/10.1007/s00453-023-01163-7.

[4] M. Breugst and H. Reissig, The Huisgen reaction: Milestones of the 1,3-dipolar cycloaddition, *Angew Chem Int Ed Engl.* **59** (2020), 12293–12307, doi:10.1002/anie.202003115, https://doi.org/10.1002/anie.202003115.

[5] K. Cocq, C. Barthes, A. Rives, V. Maraval and R. Chauvin, Synthesis of functional carbobenzenes with functional properties: The C2 tether key, *Synlett* **30** (2019), 30–43, doi:10.1055/s-0037-1610269, https://doi.org/10.1055/s-0037-1610269.

[6] D. S. Davas, S. Bhardwaj, R. Sen, D. K. Gopalakrishnan and J. Vaitla, Synthesis of olefins by formal carbene coupling, *Adv. Synth. Catal.* **364** (2022), 3122–3142, doi:10.1002/adsc.202200444, https://doi.org/10.1002/adsc.202200444.

[7] J. Dugundji and I. K. Ugi, An algebraic model of constitutional chemistry as a basis for chemical computer programs, *Topics Curr. Chem.* **39** (1973), 19–64, doi:10.1007/BFb0051317, https://doi.org/10.1007/BFb0051317.

[8] L. Euler, Solutio problematis ad geometriam situs pertinentis, *Comm. acad. sci. Petropolitanae* **8** (1741), 128–140.

[9] C. Flamm, D. Merkle, P. F. Stadler and U. Thorsen, Automatic inference of graph rewriting rules using the cyclic nature of chemical reactions, in: R. Echahed and M. Minas (eds.), *Graph Transformation, ICGT 2016*, Springer Verlag, Berlin, Heidelberg, D, volume 9761 of *Lecture Notes Comp. Sci.*, 2016 pp. 208–222, doi:10.1007/978-3-319-40530-8_13, https://doi.org/10.1007/978-3-319-40530-8_13.

[10] H. Fleischner, *Eulerian graphs and related topics*, volume 45 of *Ann. Discr. Math.*, North Holland, Amsterdam, 1990.

[11] S. Fujita, Description of organic reactions based on imaginary transition structures. 1. Introduction of new concepts, *J. Chem. Inf. Comput. Sci.* **26** (1986), 205–212, doi:10.1021/ci00052a009, https://doi.org/10.1021/ci00052a009.

[12] S. Fujita, Description of organic reactions based on imaginary transition structures. 2. Classification of one-string reactions having an even-membered cyclic reaction graph, *J. Chem. Inf. Comput. Sci.* **26** (1986), 212–223, doi:10.1021/ci00052a010, https://doi.org/10.1021/ci00052a010.

[13] J. B. Hendrickson, Comprehensive system for classification and nomenclature of organic reactions, *Journal of Chemical Information and Computer Science* **37** (1997), 852–860, doi:10.1021/ci970040v, https://doi.org/10.1021/ci970040v.

[14] R. Herges, Coarctate and pseudocoarctate reactions: Stereochemical rules, *J. Org. Chem.* **80** (2015), 11869–11876, doi:10.1021/acs.joc.5b01959, https://doi.org/10.1021/acs.joc.5b01959.

[15] C. Hierholzer, Ueber die Möglichkeit, einen Linienzug ohne Wiederholung und ohne Unterbrechung zu umfahren, *Mathematische Annalen* **6** (1873), 30–32, doi:10.1007/BF01442866, https://doi.org/10.1007/BF01442866.

[16] F. Hoonakker, N. Lachiche, A. Varnek and A. Wagner, A representation to apply usual data mining techniques to chemical reactions – illustration on the rate constant of SN_2 reactions in water, *Int. J. Artif. Intelligence Tools* **20** (2011), 253–270, doi:10.1142/S0218213011000140, https://doi.org/10.1142/S0218213011000140.

[17] F. Horn and R. Jackson, General mass action kinetics, *Arch. Rational Mech. Anal.* **47** (1972), 81–116, doi:10.1007/BF00251225, https://doi.org/10.1007/BF00251225.

[18] W. O. Kermack and R. Robinson, An explanation of the property of induced polarity of atoms and an interpretation of the theory of partial valencies on an electronic basis, *J. Chem. Soc., Transactions* **121** (1922), 427–440, doi:10.1039/CT9222100427, https://doi.org/10.1039/CT9222100427.

[19] G. N. Lewis, The atom and the molecule, *J. Am. Chem. Soc.* **38** (1916), 762–785, doi:10.1021/ja02261a002, https://doi.org/10.1021/ja02261a002.

[20] P. Muller, Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994), *Pure Appl. Chem.* **66** (1994), 1077–1184, doi:10.1351/pac199466051077, https://doi.org/10.1351/pac199466051077.

[21] S. Müller, C. Flamm and P. F. Stadler, What makes a reaction network "chemical"?, *J. Cheminformatics* **14** (2022), 63, doi:10.1186/s13321-022-00621-8, https://doi.org/10.1186/s13321-022-00621-8.

[22] J. Olbregts, Termolecular reaction of nitrogen monoxide and oxygen: A still unsolved problem, *Chemical Kinetics* **17** (1985), 835–848, doi:10.1002/kin.550170805, https://doi.org/10.1002/kin.550170805.