

Theory and algorithms for clusters of cycles in graphs for material networks

Perrin E. Ruth^{*1} and Maria Cameron^{†1}

¹Department of Mathematics, University of Maryland, College Park, MD 20742, USA

November 14, 2025

Abstract

Analysis of complex networks, particularly material networks such as the carbon skeleton of hydrocarbons generated in hydrocarbon pyrolysis in carbon-rich systems, is essential for effectively describing, modeling, and predicting their features. An important and the most challenging part of this analysis is the extraction and effective description of cycles, when many of them coalesce into complex clusters. A deterministic minimum cycle basis (MCB) is generally non-unique and biased to the vertex enumeration. The union of all MCBs, called the set of relevant cycles, is unique, but may grow exponentially with the graph size. To resolve these issues, we propose a method to sample an MCB uniformly at random. The output MCB is statistically well-defined, and its size is proportional to the number of edges. We review and advance the theory of graph cycles from previous works of Vismara, Gleiss et al., and Kolodzik et al. In particular, we utilize the polyhedron-interchangeability (**pi**) and short loop-interchangeability (**sli**) classes to partition the relevant cycles. We introduce a postprocessing step forcing pairwise intersections of relevant cycles to consist of a single path. This permits the definition of a dual graph whose nodes are cycles and edges connect pairs of intersecting cycles. The **pi** classes identify building blocks for crystalline structures. The **sli** classes group together sets of large redundant cycles. We present the application to an amorphous hydrocarbon network, where we (i) theorize how the number of relevant cycles may explode with system size and (ii) observe small polyhedral structures related to diamond.

Contents

1	Introduction	2
1.1	Motivation	2
1.2	An overview	4
1.3	The goal and a summary of main results	5
2	Background	6
2.1	Cycle space and fundamental cycle bases	6
2.2	Minimum cycle bases and relevant cycles	8
2.2.1	Vismara’s families of cycles	10
2.3	Lemmas about exchanging basis cycles	13
2.4	Partitions of the relevant cycles	15

^{*}pruth@umd.edu

[†]mariakc@umd.edu

3	Theory: Concepts	16
3.1	Polyhedron-interchangeability: definition and basic concepts	17
3.1.1	Theoretical properties	18
3.1.2	pi classes as polyhedron clusters	19
3.2	Short loop-interchangeability: definition, equivalence classes, & computation	21
4	Algorithms for sampling pi and sli classes	23
4.1	Modified Vismara’s cycle families	23
4.1.1	Definition	23
4.1.2	An algorithm for computing V’-families	24
4.1.3	The computational cost	28
4.2	Minimum cycle basis	29
4.3	Change of basis	31
4.4	pi and sli classes	32
5	Theory and algorithms for sampling minimum cycle bases	34
5.1	Uniform random sampling of minimum cycle bases	34
5.1.1	The basic algorithm	34
5.1.2	The efficient algorithm	36
5.2	Pairwise intersections in a minimum cycle basis	37
6	Applications to hydrocarbon pyrolysis systems	40
6.1	Generating data	40
6.2	Relevant cycles versus minimum cycle basis	41
6.2.1	C ₁₀ H ₁₆ simulation data	41
6.2.2	Random geometric graphs	42
6.2.3	A potential explosion of the relevant cycle count	43
6.3	Sources of additional cycles	43
6.4	Polyhedra and diamond	43
7	Discussion	44
7.1	Sampling cycles (ring perception)	44
7.2	Polyhedra.	45
8	Conclusion	45
A	Proofs for Section 3	46
A.1	Mutual relevance and useful lemmas	46
A.2	Polyhedron-interchangeability	48
A.3	Short loop-interchangeability	52
B	Proofs for Section 5.2	53

1 Introduction

1.1 Motivation

Amorphous material networks appear in numerous contexts such as polymers [2], energetic carbon [3, 4], soot [5], and silicate melts [6]. Understanding the network topology of these systems is

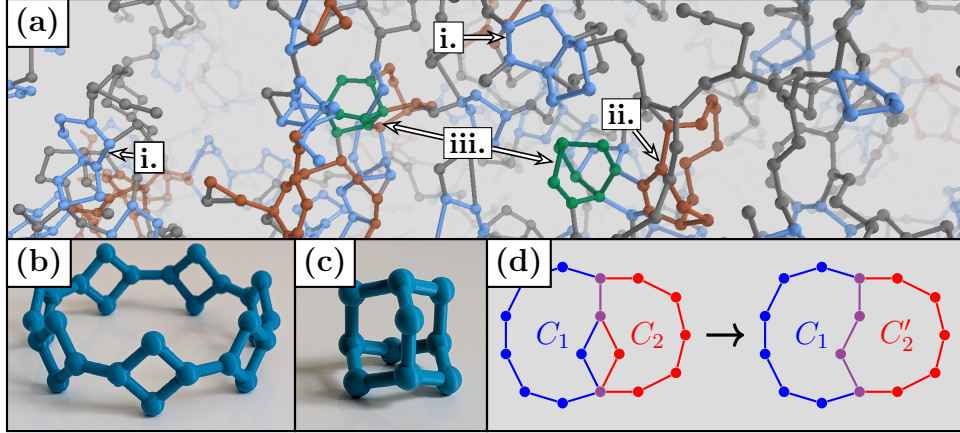


Figure 1: **(a)** Snapshot of the giant biconnected component in a simulation initialized to adamantane ($C_{10}H_{16}$) at 4000K and 40.5GPa. We highlight (i) *essential* [1] pentagonal and hexagonal rings belonging to every MCB in blue, (ii) a *sli* class composed of two 11-membered rings separated by a hexagonal ring in orange, and (iii) two *pi* classes with the same shape as barallene (C_8H_8) in green. **(b)** A graph made of diamond-shaped cycles chained into a loop like a bracelet. An MCB of this graph includes all of the small diamond-shaped loops and one large loop. **(c)** Carbon skeleton of adamantane, the smallest diamondoid, i.e., diamond is formed by repeatedly tiling adamantane. An MCB of adamantane includes any 3 of the 4 hexagonal faces. **(d)** Depiction of the process of modifying a cycle $C_2 \rightarrow C'_2$ such that the cycles C_1 and C_2 overlap over a single path.

critical because it influences their material properties. The local atomistic structures within these networks can be leveraged to make predictions about their thermodynamic limits and to reduce combinatorial complexity. Here, we review, discuss, and extend theory and algorithms for the seemingly simple task of sampling rings from material networks. The assumptions under which these algorithms are efficient are consistent with the carbon skeletons of large hydrocarbons arising in hydrocarbon pyrolysis, a complex chemical reaction system composed of carbon and hydrogen at extreme temperatures and pressures.

Conventional approaches for studying hydrocarbon pyrolysis include molecular dynamics (MD) [7, 8, 9, 10] simulations and kinetic Monte Carlo [11, 9]. MD is flexible and can be used to simulate various systems and thermodynamic conditions. However, MD simulations are expensive. They may take days or even weeks to achieve nanosecond timescales. Furthermore, they need to be performed anew for every set of initial conditions. Kinetic Monte Carlo [12, 13] simulates a chemical system using a set of known chemical reactions. These reactions and their reaction rates can be learned from a set of molecular dynamics simulations. This approach is limited by the combinatorial growth in the number of reactions, e.g., a simulation with $\sim 10^3$ atoms results in $\sim 10^4$ potential reactions [14, 15, 8]. The combinatorial growth of hydrocarbons has been studied as early as 1875, where Cayley used generating functions to estimate the number of alkanes of a given length [16, 17]. Due to this growth, only small molecules counts can be predicted accurately by this approach. To address this limitation, Refs. [11, 9] propose kinetic models with only local reactions accounting for chemical changes near the reaction site. This approach reduces the number of reactions and can be used to make predictions about the size of the largest molecule. However, like MD simulations, the local kinetic Monte Carlo requires a new run for each set of initial conditions.

Our research group utilized random graphs as equilibrium models for hydrocarbon pyrolysis [10, 18]. For hydrogen-rich systems, the configuration model [19, 20] – a classical random graph model

– was used to estimate the molecule size distribution [10]. This model is particularly effective for small alkanes (≤ 20 carbon atoms) or fully saturated acyclic hydrocarbons. However, as the hydrogen-to-carbon ratio decreases, the configuration model significantly exaggerates the size of the largest molecule, mostly because it does not promote the formation of small carbon rings abundant in hydrocarbon molecules. Using Refs. [21, 22] as building blocks, we introduced a random graph model with disjoint loops that accurately predicted the size of the largest molecule [18]. This disjoint loop model also predicted the phase transition between the states with and without a giant molecule, i.e., a molecule that contains a macroscopic fraction of the total number of atoms. In carbon-rich systems, the rings cluster together, and the disjoint loop assumption fails. See Figure 1 (a). In this setting, proper statistics for carbon rings are required.

1.2 An overview

At a macroscopic scale, a ring cluster can be identified as a biconnected component or a collection of nodes where each pair of nodes is connected by two or more disjoint paths [23, 24]. A method for computing biconnected components is given in the 1971 paper by Hopcroft and Tarjan [25]. At a microscopic scale, we want to obtain the building blocks of these clusters, i.e., their loops. In many cases, the naïve approach of computing all cycles is computationally intractable. The total number of cycles grows exponentially with graph size, many of which are redundant. Instead, most methods for sampling loops utilize the cycle space or the vector space of cycles.

Classically, the loops of a graph are sampled from a *minimum cycle basis* (MCB) or a basis of the cycle space that minimizes the sum of cycle lengths [26]. In chemical systems, a minimum cycle basis is also called a smallest set of smallest rings, as introduced by Plotkin [27]. MCBs are not without their flaws. Most notably, they are not unique, so the loops sampled via an MCB depend on node labels and algorithmic implementation. This issue has led to several competing definitions for sampling cycles in chemistry literature [28].

Here, we focus on the approach introduced by Vismara, 1997 [29], where the set of *relevant cycles* defined as the union of all MCBs is considered. Equivalently, the relevant cycles are the unique cycles that cannot be decomposed into smaller cycles. In the worst case, the number of relevant cycles grows exponentially with respect to the number of nodes, but in practice the number of relevant cycles is typically much smaller [30]. Nevertheless, this is an undesirable property of the relevant cycles. To guarantee properties such as the number of relevant cycles and their lengths can be computed in polynomial time, Vismara partitions the relevant cycles into families of related cycles [29].

Another approach would be to investigate Vismara’s cycle families directly. However, these families are not unique, thus returning us to our original issue. To resolve this issue, Gleiss et al., 2000, [1] introduce the *interchangeability relation*, an equivalence relation that partitions the relevant cycles into unique interchangeability classes. This relation helps to characterize the lack of uniqueness in MCBs. Two relevant cycles are interchangeable if one cycle can be swapped for the other in a cycle basis, and if the two cycles only differ by equal or shorter length cycles. Often, this relation groups together large cycles that differ by smaller cycles. For example, the cycles around the perimeter of the bracelet graph in Figure 1 (b) form a single interchangeability class. In other cases, this relation groups together cycles that belong to three-dimensional structures. For instance, the four hexagonal faces of adamantane in Figure 1 (c) form a single interchangeability class.

Kolodzik et al., 2012 [31], argue these interchangeability classes are too coarse and introduce the *unique ring family (URF)-pair-relation*. This relation is more restrictive, where two relevant cycles are URF-pair-related if they share edges and if they differ by *strictly shorter* cycles. This separates

the cases above. The large cycles in the bracelet graph are URF-pair-related because they differ by smaller diamond cycles. The hexagonal faces in adamantane are not URF-pair-related because they differ by equal length hexagonal faces. The URF-pair-relation is not an equivalence relation. Instead, the unique ring families are given by the transitive closure of the URF-pair-relation.

1.3 The goal and a summary of main results

The goal of this work is to review and extend theory and algorithms that describe and extract complex amorphous ring clusters in carbon-rich simulation data of hydrocarbon pyrolysis. To achieve this, we adapt and advance concepts and methods from Refs. [1, 31]. Our theoretical developments are the following:

- At a coarse level, we introduce the *polyhedron-interchangeability* (**pi**) relation, a modification of the interchangeability relation from Ref. [1]. Unlike interchangeability, **pi** relates pairs of cycles that can be swapped between MCBs. We prove **pi** is an equivalence relation (Theorem 6) and the resulting equivalence classes can be computed using a given MCB (Theorem 8). The theory of **pi** also characterizes how we can modify MCBs by swapping cycles.
- At a granular level, we introduce the *short loop-interchangeability* (**sli**) relation, adapted from the URF-pair-relation from Ref. [31]. **sli** removes the requirement that cycles share edges from the URF-pair-relation. As a result, **sli** is an equivalence relation. **sli** was first introduced in the dissertation of Gleiss [32] as stronger interchangeability, where it was shown to be an equivalence relation. We prove the **sli** classes can be computed using a given MCB. We suggest to use the **pi** and **sli** relations together to analyze ring clusters.

Using this theory we develop the following methods for sampling cycles in a graph.

- *Random sampling of minimum cycle bases.* We introduce a method for sampling MCBs uniformly at random. Of course, a random MCB is not deterministic, but it does have a concrete definition as a probabilistic object. Thus, the practitioner can confidently sample loops from a random MCB knowing it is statistically well-defined.
- *Pairwise intersection of cycles.* A graph is defined by nodes and by edges denoting connections between nodes. Similarly, we are interested in loops and their intersections that resulting in the formation of loop cluster. In this work, we fully characterize the way pair of cycles in MCBs can intersect. Let the intersection of two cycles contain more than one path as in Fig. 1 (d). We introduce a method for modifying one of these cycles to make the intersection consist of a single path. This property is desirable as individual paths are conceptually simple and they are easy to represent computationally as node lists.

Using this, we define a dual graph, which can be interpreted as a graph of cycles. The notion of dual graph has a long history in planar graphs [23], and graphs of cycles have been considered for modeling chemical systems as well [2, 33]. Our contribution is defining the structure of these dual graphs for MCBs.

We apply this theory to the hydrocarbon pyrolysis system. In most cases, we find that lack of uniqueness of MCBs is largely caused by rings at varying lengths as in Figure 1 (b). We find ring statistics vary greatly between the relevant cycles and an MCB. In particular, the relevant cycles introduce redundant large rings. Furthermore, our analysis of the relevant cycles naturally uncovers three-dimensional structures. These structures take the form of polyhedra with non-flat

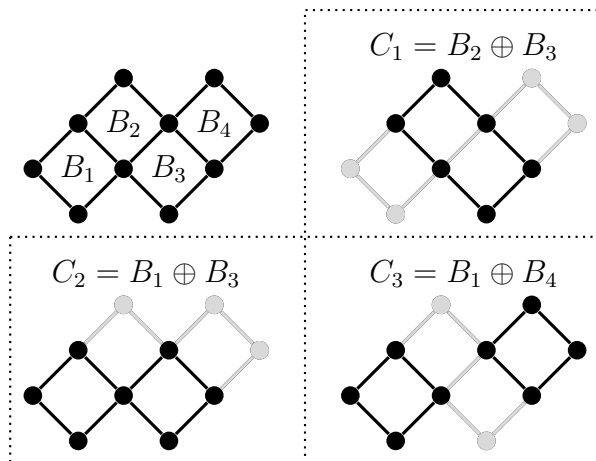


Figure 2: A graph with four face cycles B_1 , B_2 , B_3 , and B_4 forming a basis. Three additional cycles C_1 , C_2 , and C_3 are shown. C_1 is a single loop while C_2 and C_3 are not.

faces as in Figure 1 (c). These polyhedra are useful for describing crystalline materials. Proper sampling of these polyhedra will require further theoretical advancements.

This manuscript is organized as follows. In Section 2, we provide the necessary background for the cycle space including the relevant cycles from Ref. [29] and the partitions of the relevant cycles from Refs. [1, 31]. In Section 3, we introduce the `pi` and `sli` relations used to partition the relevant cycles. In Section 4, we provide algorithms for computing the `pi` and `sli` classes. Our Python codes are available on Github [34]. These codes take a simple graph from the NetworkX package [35] as input. In Section 5, we provide an algorithm for sampling MCBs uniformly at random, and we characterize the intersection of cycles in MCBs. In Section 6, we show how this approach can be used to characterize ring clusters from hydrocarbon pyrolysis simulations. These molecular dynamics simulations are obtained via LAMMPS [36, 37] using the ReaxFF [38, 39, 40, 41] force field permitting chemical reactions.

2 Background

2.1 Cycle space and fundamental cycle bases

Cycle spaces have a rich history in electronic circuit theory, chemistry, and other contexts. We start with a brief discussion of classical results in this field. We also introduce notation, which is merged from Refs. [1, 26, 29]. For a more detailed history and discussion of these topics we recommend the books on graph theory [24, 23], review of cycle bases [26], review of cycles in chemical systems [28], and references therein.

Let $G = (V, E)$ be a graph with $n := |V|$ nodes and $m := |E|$ edges. Formally, a *cycle* is an edge-induced subgraph of G where all nodes have even degree. More intuitively, a cycle is a collection of loops that do not share edges. See Fig. 2. We note, nearly all cycles we consider are simple loops, including all cycles introduced later in this section in fundamental and minimum cycle bases.

The utility of this definition is that the set of cycles – denoted by $\mathcal{C}$ – forms a vector space called the *cycle space*. The operation we apply to cycles is the symmetric difference $C_1 \oplus C_2 = (C_1 \setminus C_2) \cup (C_2 \setminus C_1)$. The scalars of this vector space, 0 and 1, comprise the field $GF(2)$ with addition and multiplication modulo 2. Note, a cycle is necessarily its own additive inverse, $C \oplus C = \emptyset$.

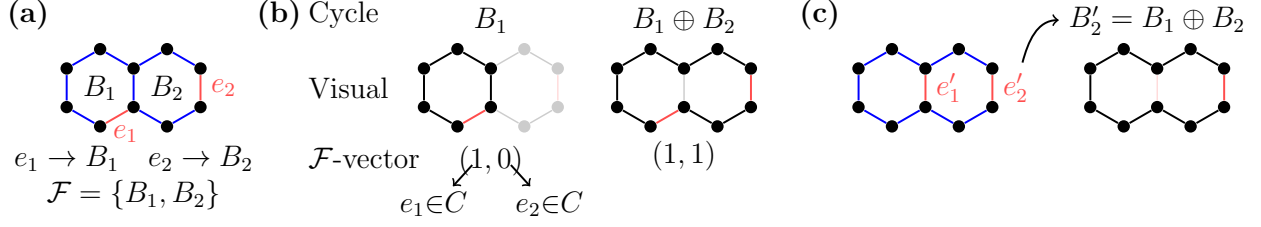


Figure 3: **(a)** A graph G with edges colored according to whether they are inside (blue) or outside (red) the spanning tree T . The face cycles B_1 and B_2 form a fundamental cycle basis $\mathcal{F}$, where they are composed of an edge in $E(G) \setminus E(T)$ joined by a path in T . **(b)** A depiction of the cycles B_1 and $B_1 \oplus B_2$ alongside their binary representation as fundamental cycle basis $\mathcal{F}$ vectors. **(c)** Another spanning tree T' resulting in a different fundamental cycle basis that contains $B'_2 = B_1 \oplus B_2$.

Since a cycle $C \in \mathcal{C}$ is a set of edges, it may be represented as an *incidence vector*: a vector in $\{0, 1\}^m$ where element i indicates whether edge i is contained in C . Not all incidence vectors correspond to cycles. In vector form, the symmetric difference of cycles can be computed via an element-wise XOR operation (or by vector addition modulo 2). We distinguish that the cycle space is only a subspace of the set of incidence vectors, and ultimately we want a lower-dimensional representation of cycles.

This notion of a vector space leads to a natural definition of a basis. Specifically, a *cycle basis* is a set of linearly independent cycles $\mathcal{B} = \{B_1, \dots, B_\nu\}$ that span the cycle space, where ν is the dimension of the cycle space. Then, we may expand a cycle C as follows,

$$C = b_1 B_1 \oplus \dots \oplus b_\nu B_\nu, \quad b_1, \dots, b_\nu \in \{0, 1\}. \quad (1)$$

Note, lower case letters are used for scalars, upper case for cycles, and math italics (e.g. $\mathcal{B}$) for sets of cycles. Computationally, it is easier to store cycles as binary vectors $(b_1, \dots, b_\nu)$ where the addition operator $\oplus$ is still interpreted as the element-wise XOR operation.

Now that we know how to use a cycle basis, the next question is “how do we construct such a basis?” A classical approach is to utilize a spanning tree.

Definition 1. Let T be a spanning tree of G , or a forest if G is disconnected, and $\{e_i\}$ the edges in G outside of T , i.e., $e_i \in E(G) \setminus E(T)$. The fundamental cycle basis induced by T is the set of cycles $\mathcal{F} = \{B_i\}$, where B_i is the cycle composed of the edge e_i and the path that connects its endpoints in T .

An example of a fundamental cycle basis is given in Fig. 3 (a). To get the expansion of a cycle C with respect to $\mathcal{F}$ we focus on the edges $\{e_i\}$. Note, there exists a unique basis cycle $B_i \in \mathcal{F}$ for each edge e_i . So, B_i is in the expansion of C with respect to $\mathcal{F}$ if and only if $\{e_i\}$ is contained in C . Furthermore, the coefficient b_i as defined in Eq. (1) is the indicator that $e_i \in C$ and $(b_1, \dots, b_\nu)$ is the incidence vector of C restricted to edges in $E(G) \setminus E(T)$. Following this procedure, we find the rank of the cycle space is the number of edges in $E(G) \setminus E(T)$:

$$\nu = m - n + (\# \text{ Connected Components}), \quad (2)$$

where a tree with n nodes has $n - 1$ edges, and T has a spanning tree of each connected component of G . However, we have not yet verified that $\mathcal{F}$ is a basis.

It is easy to see that $\mathcal{F}$ is linearly independent, since each cycle B_i has an edge e_i that the other cycles in $\mathcal{F}$ do not. To see that $\mathcal{F}$ spans $\mathcal{C}$, we take an arbitrary cycle $C \in \mathcal{C}$. Then, for each basis

cycle $B_i \in \mathcal{F}$ we add it to C if the corresponding edge $e_i \in E(G) \setminus E(T)$ is in C :

$$C' = C \oplus \mathbf{1}_{[e_1 \in C]} B_1 \oplus \dots \oplus \mathbf{1}_{[e_\nu \in C]} B_\nu \quad (3)$$

where $\mathbf{1}$ is the indicator function. By cancellation from the $\oplus$ -operator, C' contains no edges in $E(G) \setminus E(T)$. In other terms, C' is contained in the spanning forest T . Trees lack loops, so C' is necessarily the empty cycle $\emptyset$, and C is contained in the span of $\mathcal{F}$.

2.2 Minimum cycle bases and relevant cycles

Fundamental cycle bases are useful for analyzing the cycle space as a vector space. However, these bases should not be used when sampling loop statistics. Specifically, fundamental cycle bases are not unique and depend on the spanning tree used to construct them. Furthermore, a given fundamental cycle basis may contain large cycles that are not intuitively relevant.

To elaborate on this issue consider the graph depicted in Figure 3 with two hexagons that share an edge. The spanning tree in Figure 3 (a) results in a fundamental cycle basis composed of these hexagons, as desired. However, the spanning tree in Figure 3 (c) produces a fundamental cycle basis containing the larger cycle with 10 nodes surrounding the perimeter of this graph. We argue, that it is more natural to describe this larger cycle as the sum of smaller cycles.

A common approach taken to address this concern is to take a cycle basis that contains only short cycles:

Definition 2. A minimum cycle basis (MCB) $\mathcal{M}$ is a cycle basis that minimizes the sum of cycle lengths

$$\text{Cost}(\mathcal{M}) = \sum_{B \in \mathcal{M}} |B|$$

where $|B|$ is the number of edges in cycle B .

As one might hope, the two hexagons in Figure 3 form a unique MCB.

This definition has an interesting history in many fields. Ref. [26] offers a survey of algorithms to compute MCBs. Ref. [28] reviews methods for counting cycles in chemical systems. We remark that, in chemical systems, a cycle is typically called a ring, and an MCB is referred to as the “Smallest Set of Smallest Rings”. In Section 4.2, we review an algorithm from Ref. [42] for computing MCBs. This algorithm runs in $O(mn^2)$ time. In Section 4.3, we show how one can expand an arbitrary cycle in terms of an MCB output by this algorithm.

Unfortunately, MCBs are not always unique – see Figure 4. Therefore, if we evaluate summary statistics from cycles in a given MCB, then we obtain inconsistent results depending on which MCB is used. For instance, consider the number of times node 2 in Figure 4 participates in a cycle. If we compute this statistic using the MCB $\mathcal{M}_1 = \{C_1, C_2\}$ then node 2 participates in two cycles C_1 and C_2 , but if we take $\mathcal{M}_2 = \{C_2, C_3\}$ as an MCB then node 2 participates in only one cycle C_2 . As mentioned in Ref. [31], this is an issue with pattern matching of chemical rings in SMARTS, a programming library for identifying molecular patterns and properties [43]. Additionally, the OEChem library for chemistry and cheminformatics avoids using MCBs in chemical ring perception citing lack of uniqueness, and states “Smallest Set of Smallest Rings (SSSR) considered Harmful” [44].

Because of this issue, several works in computational chemistry consider definitions that extend MCBs to larger sets of cycles. We do not review all of these definitions here, and refer the interested reader to Ref. [28]. Instead, we review the notion of relevant cycles from Ref. [29], which serves as a foundation for the definitions discussed in this manuscript. The relevant cycles have two equivalent definitions:

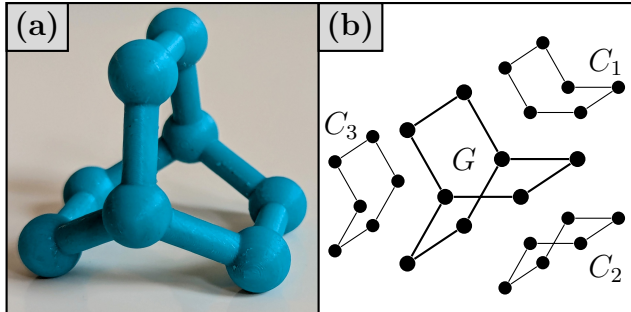


Figure 4: A graph G that does not have a unique MCB. Ignoring double bonds, G is the carbon skeleton of barallene with formula C_8H_8 . **(a)** 3D printed visualization of G . **(b)** Two-dimensional representation of G . The three hexagonal relevant cycles C_1 , C_2 , and C_3 are shown separately. Any pair of these cycles forms an MCB, where the third cycle is the sum of the other two, e.g., $\mathcal{M}_1 = \{C_1, C_2\}$ is an MCB and $C_3 = C_1 \oplus C_2$.

Definition 3. The set of relevant cycles $\mathcal{C}_R$ is the union of all MCBs.

Definition 4. A cycle $C \in \mathcal{C}$ is relevant ($C \in \mathcal{C}_R$) if and only if it is not generated by smaller cycles, i.e., there does not exist a set of smaller cycles $\mathcal{X}$, $|C'| < |C|$ for $C' \in \mathcal{X}$, such that

$$C = \bigoplus_{C' \in \mathcal{X}} C' \quad (4)$$

where $\bigoplus$ is the series representation of the $\oplus$ -operator.

We include a proof of the equivalence of these definitions after introducing additional theory about the cycle space – see Corollary 5 in Sec. 2.3. For the graph depicted in Figure 4, the relevant cycles are the three hexagons $\mathcal{C}_R = \{C_1, C_2, C_3\}$. This set of relevant cycles has several useful properties [28].

- *Uniqueness:* The set $\mathcal{C}_R$ is unique, so statistics on the relevant cycles are well-defined.
- *Completeness:* The relevant cycles can be used to generate the cycle space since $\mathcal{C}_R$ contains a cycle basis.
- *Computability:* There exists a polynomial time algorithm for computing the set of relevant cycles. [29].
- *Chemical relevance:* The set of relevant cycles contains all building blocks of chemical structures present in molecules.

One downside is that the number of relevant cycles may grow exponentially for certain types of graphs. For example, in the bracelet-like graph composed of k diamonds chained together into a loop shown in Figure 1 (b). All MCBs in this graph are composed of the k diamonds and one large loop of length $3k$. Each large loop either takes the top or bottom path of a given diamond, so the number of large relevant cycles grows exponentially as 2^k , while the number of nodes grow linearly as $4k$. [29]

It is remarkable that the algorithm from Ref. [29] for computing the relevant cycles of a graph runs in polynomial time. In broad steps, the relevant cycles are computed by (1) generating a candidate set of cycles (a superset of $\mathcal{C}_R$) and (2) filtering out the cycles that are irrelevant. The key idea that makes this procedure feasible in polynomial time is that the cycles are grouped into families of equal length.

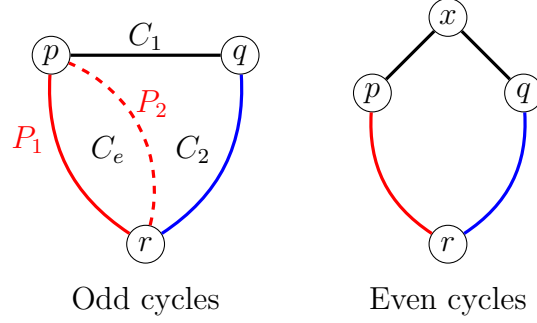


Figure 5: Depiction of the descriptors used to organize the relevant cycles into V-families as in Ref. [29]. For an odd length cycle (left), we fix its largest index node r and the edge (p, q) at the opposite end of the cycle. For an even length cycle (right), we fix its largest index node r and the triple (p, x, q) at the opposite end of the cycle. In the left, we also draw the exchange of the path P_1 for P_2 by adding the cycle $C_e = P_1 \oplus P_2$ to C_1 to obtain $C_2 = C_1 \oplus C_e$.

2.2.1 Vismara's families of cycles

Here we review the construction of cycle families from Ref. [29]. The key idea to make the algorithm for computing the relevant cycles polynomial is to group the relevant cycles into a polynomial number of families. At the same time, it has a theoretical drawback that it is attached to a given enumeration of the vertices.

Consider a simple cycle C . We select antipodal parts of C to determine the family it belongs to. First, we fix the node r with largest index in C . The largest index node is selected because it is unique up to vertex labels. Second, if C is an odd length cycle, then we fix the edge (p, q) at the opposite end of the cycle. If C has even length, then we fix the triple (p, x, q) at the opposite end of the cycle. In either case, the paths from r to p and q in C are of equal length. See Figure 5. We call these families V-families.

Definition 5 (Vismara cycle family).

- The odd V-family with descriptors r, p, q is the set of cycles with largest index node r consisting of shortest paths from r to p and q joined by the edge (p, q) .
- The even V-family with descriptors r, p, x, q is the set of cycles with largest index node r consisting of shortest paths from r to p and q joined by the triple (p, x, q) .

All cycles in a V-family are of equal length. We only consider shortest paths from r to p and q because it is a necessary condition for a cycle to be relevant – see Lemma 1.

As an example, consider the graph depicted in Figure 6. This graph has eight large relevant cycles of odd length nine, distinguished by the path they take through each diamond. Four of these cycles take the path $(10, 12, 11)$ in the diamond $(9, 11, 12, 10)$ making their largest index node $r = 12$. All four cycles share the same opposing edge $(2, 5)$ and belong to the same family. The remaining large relevant cycles take $(10, 9, 11)$ instead of $(10, 12, 11)$ in the diamond $(9, 11, 12, 10)$, so they have largest index node $r = 11$. The opposing edge for these remaining cycles is $(5, 6)$ if they take the path $(5, 6, 8)$ in the diamond $(5, 6, 8, 7)$ or $(5, 7)$ if they take the path $(5, 7, 8)$. This results in two additional V-families, each of which has two cycles, one for each path in the diamond $(1, 2, 4, 3)$.

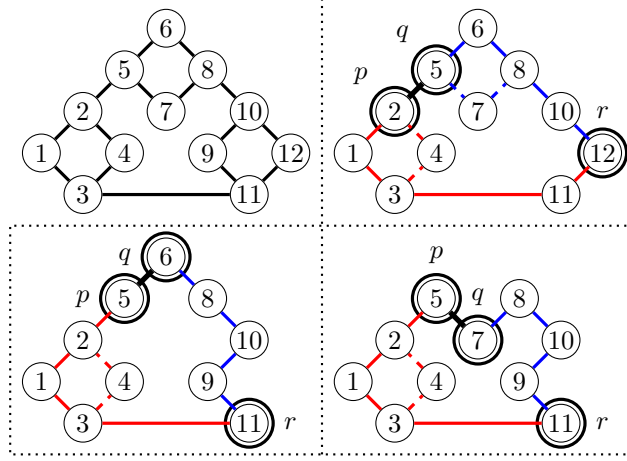


Figure 6: A labeled graph composed of three diamonds chained into a loop. The relevant cycles of this graph are the three diamonds and eight large loops of length nine. For the given node labels, the large cycles form three V-families depicted separately with their descriptors.

Ultimately, splitting cycles into V-families as in Definition 5 is effective because shortest paths can be exchanged via cycle addition. For instance, let us consider two cycles C_1 and C_2 in a V-family that differ by taking disjoint paths P_1 and P_2 from r to p respectively. See, e.g., Figure 5, left. Additionally, consider the cycle C_e obtained by merging these paths, given by

$$C_e := P_1 \oplus P_2. \quad (5)$$

Then, we exchange C_1 for C_2 as follows

$$C_1 = C_2 \oplus C_e \quad (6)$$

where the path P_2 cancels between C_2 and C_e and is replaced by P_1 .

In general, one cycle C_1 in a family can be exchanged for another cycle C_2 in the same family by adding at most two cycles. Specifically, one cycle is used to swap paths from r to p , and the other is used to swap paths from r to q . Note, these exchange cycles are shorter than the equal length cycles C_1, C_2 , e.g., for C_e given by Eq. (5)

$$|C_e| \leq |P_1| + |P_2| = 2|P_1| < |C_1|. \quad (7)$$

This exchange procedure can be used to show two useful properties for computing cycle families.

First, this can be used to show that the paths from r to p and q in a relevant cycle are shortest paths, because otherwise they can be exchanged for shortest paths. The proof of this claim relies on the fact that the paths from r to p and q are less than half the length of the cycle.

Lemma 1 (Ref. [29], Lemma 2). *Let C be a relevant cycle and P a path in C that is at most half the length of C , $|P| \leq |C|/2$. Then, P is a shortest path.*

Proof. Suppose towards a contradiction that P is not a shortest path. Let P' be a shortest path with the same endpoints, such that $|P'| < |P|$. Let C_e be the cycle used to exchange paths P and P' given by

$$C_e = P \oplus P'. \quad (8)$$

Let C' be given by

$$C' := C \oplus C_e = (C \setminus P) \oplus P'. \quad (9)$$

where the second equality uses associativity and that the path P belongs to C . Then, our initial cycle is given by the sum

$$C = C' \oplus C_e. \quad (10)$$

Both C' and C_e are strictly smaller than C

$$|C_e| \leq |P| + |P'| < 2|P| \leq |C|, \quad (11)$$

$$|C'| \leq |C| - |P| + |P'| < |C|. \quad (12)$$

Therefore, C is not relevant by Definition 4, because it is the sum of shorter cycles. This contradicts our initial assumptions, so P is a shortest path. $\square$

Shortest paths from a given root node r to the other vertices can be obtained by breadth-first search. Neglecting some detail – how to avoid nodes with index greater than r – the algorithm for generating candidate cycle families is to perform breadth-first search on all vertices $r \in V$ and identify node pairs p and q that form structures as in Fig. 5. We discuss this algorithm and our modifications to it in Sec. 4.1.

Another benefit of this exchange procedure is that we only need to test if one cycle in a V-family is relevant. In Ref. [29], a single cycle *prototype* is taken arbitrarily from a V-family. If this prototype is relevant, then all cycles in the V-family are relevant. This is computationally feasible to test because there are a polynomial number of cycle prototypes. In the following lemma, C_1 and C_2 are cycles in a family, and $\mathcal{X}$ is a set of two cycles used to exchange the paths from r to p and q . The cycles $\mathcal{X}$ used to exchange paths are shorter than C_1 and C_2 . See Eq. (7).

Lemma 2. *Let $C_1, C_2 \in \mathcal{C}$ be equal length cycles that differ by smaller cycles, i.e., there exists a set of cycles $\mathcal{X}$, $|C'| < |C_1|$ for $C' \in \mathcal{X}$, such that*

$$C_1 = C_2 \oplus \bigoplus_{C' \in \mathcal{X}} C'. \quad (13)$$

Then, C_1 is relevant if and only if C_2 is relevant.

Proof. Let C_2 not relevant. We want to show that C_1 is not relevant. By Definition 4, C_2 is the sum of smaller cycles $\mathcal{Y}$, $|C'| < |C_2|$ for $C' \in \mathcal{Y}$, such that

$$C_2 = \bigoplus_{C' \in \mathcal{Y}} C'. \quad (14)$$

Then, by Eq. (13), C_1 is given by the pair of sums

$$C_1 = \bigoplus_{C' \in \mathcal{Y}} C' \oplus \bigoplus_{C'' \in \mathcal{X}} C''. \quad (15)$$

Cycles in the intersection $C' \in \mathcal{X} \cap \mathcal{Y}$ cancel in the above double sum, since $C' \oplus C'$. Thus, C_1 is given by the single sum

$$C_1 = \bigoplus_{C' \in \mathcal{X} \Delta \mathcal{Y}} C' \quad (16)$$

where Δ is the symmetric difference $\mathcal{X} \Delta \mathcal{Y} = (\mathcal{X} \setminus \mathcal{Y}) \cup (\mathcal{Y} \setminus \mathcal{X})$ for more general sets. Finally, C_1 is not relevant by Definition 4, because it is the sum of smaller cycles $\mathcal{X} \Delta \mathcal{Y} \subseteq \mathcal{X} \cup \mathcal{Y}$.

To see the reverse statement, we note that the relationship given by Eq. (13) is symmetric. In particular, we can add cycles $C' \in \mathcal{Y}$ to both sides of Eq. (13) to move them to the left-hand side

$$C_1 \oplus \bigoplus_{C' \in \mathcal{Y}} C' = C_2. \quad (17)$$

By the same argument as above, if C_1 is not relevant, then C_2 is not relevant. $\square$

2.3 Lemmas about exchanging basis cycles

Much of the theory in this text relies on the procedure of swapping a given cycle in a basis for another cycle. This notion of basis exchange can be used to define a matroid, but it is easier to understand these properties as a consequence of the cycle space being a vector space. In this section, we state this notion of basis exchange in Lemma 3; construct conditions for a cycle to be relevant in Lemma 4; and verify that Definitions 3 and 4 are relevant in Corollary 5.

For these lemmas, it is useful to introduce further notation for expanding a cycle with respect to a basis. Recall from Eq. (1), a cycle C can be expanded as the sum of basis cycles B_i with binary coefficients b_i . Thus, C is simply the sum of cycles B_i with coefficients $b_i = 1$. We define the following function to output these cycles.

Definition 6. *Let $\mathcal{B}$ be a cycle basis. The expansion operator $\mathcal{E}_{\mathcal{B}}(\cdot)$ is the function on cycles $C \in \mathcal{C}$ that outputs the unique subset $\mathcal{E}_{\mathcal{B}}(C) \subseteq \mathcal{B}$ that sums to C*

$$C = \bigoplus_{B \in \mathcal{E}_{\mathcal{B}}(C)} B, \quad \forall C \in \mathcal{C}. \quad (18)$$

In the following lemma we show a cycle C can be exchanged for cycles in its expansion $\mathcal{E}_{\mathcal{B}}(C)$. This well-known lemma has many equivalent forms, and is often referred to as the matroid property of the cycle space, see e.g., Proposition 4 in Ref. [1].

Lemma 3 (Basis exchange). *Let $\mathcal{B}$ be a cycle basis, $C \in \mathcal{C}$ be a cycle, and $B \in \mathcal{E}_{\mathcal{B}}(C)$ be an arbitrary cycle in the expansion of C with respect to $\mathcal{B}$. If we replace B with C in $\mathcal{B}$ then we obtain a new cycle basis $\mathcal{B}' = (\mathcal{B} \setminus \{B\}) \cup \{C\}$.*

Proof. To start, we expand $C \in \mathcal{C}$ with respect to the basis $\mathcal{B}$ as follows

$$C = B_1 \oplus B_2 \oplus \dots \oplus B_k, \quad B_i \in \mathcal{B}, \quad 1 \leq i \leq k. \quad (19)$$

Note, in the above equation we fix $\mathcal{E}_{\mathcal{B}}(C) = \{B_1, \dots, B_k\}$. We want to show that if we swap B_i with C in $\mathcal{B}$, then we obtain a new cycle basis. Without loss of generality, we will show that $\mathcal{B}' := (\mathcal{B} \setminus \{B_1\}) \cup \{C\}$ is a cycle basis. To show this, we add C and B_1 to both sides of Eq. (19) to obtain

$$B_1 = C \oplus B_2 \oplus \dots \oplus B_k. \quad (20)$$

Equation (20) shows that B_1 is in the span of $\mathcal{B}'$, and that our original set $\mathcal{B}$ is contained in the span of $\mathcal{B}'$. $\mathcal{B}$ is a cycle basis, so $\mathcal{B}'$ must span the full cycle space. C is linearly independent of $\{B_2, \dots, B_k\}$, otherwise B_1 would be in the span of $\{B_2, \dots, B_k\}$ by Eq. (20). Thus, $\mathcal{B}'$ is a basis. $\square$

In the following lemma, we introduce an alternate definition of relevance in terms of the expansion of a cycle with respect to a MCB. We will use this lemma as a backbone for our theoretical results in this manuscript and for testing the relevance of cycles in V-families.

Lemma 4 (Relevance criterion). *Let $C \in \mathcal{C}$ be a cycle, $\mathcal{M}$ an MCB, and*

$$L = \max\{|C'| : C' \in \mathcal{E}_{\mathcal{M}}(C)\}$$

the length of the largest cycle in the expansion of C with respect to $\mathcal{M}$. Then, the following are true

1. $L \leq |C|$, i.e., no cycles in $\mathcal{E}_{\mathcal{M}}(C)$ are larger than C ,
2. $L = |C|$ if and only if C is relevant $C \in \mathcal{C}_{\mathcal{R}}$. Furthermore, if $C \in \mathcal{C}_{\mathcal{R}}$, then for any $C' \in \mathcal{E}_{\mathcal{M}}(C)$, $|C'| = |C|$, the set $\mathcal{M}' = (\mathcal{M} \setminus \{C'\}) \cup \{C\}$ is an MCB.

Proof. (1) Let $C' \in \mathcal{E}_{\mathcal{M}}(C)$ be an arbitrary cycle in the expansion of C with respect to $\mathcal{M}$. By Lemma 3, $\mathcal{B} = (\mathcal{M} \setminus \{C'\}) \cup \{C\}$ is a cycle basis. The sum of cycle lengths in $\mathcal{B}$ is

$$\sum_{B \in \mathcal{B}} |B| = \sum_{B \in \mathcal{M}} |B| - |C'| + |C|. \quad (21)$$

$\mathcal{M}$ is an MCB, so the sum of cycle lengths in $\mathcal{B}$ is greater than or equal to the sum of cycle lengths in $\mathcal{M}$, and by the above equation

$$\begin{aligned} \sum_{B \in \mathcal{M}} |B| - |C'| + |C| &\geq \sum_{B \in \mathcal{M}} |B| \\ \implies |C| &\geq |C'|. \end{aligned} \quad (22)$$

Thus, C is not smaller than any of the cycles in its expansion with respect to $\mathcal{M}$ as claimed.

(2, $\Rightarrow$) Let $\mathcal{M}'$ be an MCB containing C . Let $C' \in \mathcal{E}_{\mathcal{M}}(C)$ be independent of $\mathcal{M} \setminus \{C\}$. By item 1, $|C'| \leq |C|$. Then, $\mathcal{B} = \mathcal{M}' \setminus \{C\} \cup \{C'\}$ is another cycle basis.

$$\sum_{B \in \mathcal{M}'} |B| \leq \sum_{B \in \mathcal{B}} |B| = \sum_{B \in \mathcal{M}'} |B| - |C| + |C'|. \quad (23)$$

Hence $|C'| \geq |C|$. Thus $|C'| = |C|$. $\square$

Corollary 5 (Ref [29], Lemma 1). *A cycle C is not relevant if and only if it is generated by smaller cycles. In other terms, Definition 4 is equivalent to Definition 3.*

Proof. ($\Rightarrow$) Let C not be relevant. Let $\mathcal{M}$ be a given MCB. C is the sum of shorter cycles

$$C = \bigoplus_{B \in \mathcal{E}_{\mathcal{M}}(C)} B \quad (24)$$

where $|B| < |C|$ for $B \in \mathcal{E}_{\mathcal{M}}(C)$ by Lemma 4.

($\Leftarrow$) Let C be the sum of smaller cycles $\mathcal{X}$, i.e., $C = \bigoplus_{C' \in \mathcal{X}} C'$ where $|C'| < |C|$ for $C' \in \mathcal{X}$. Then, the expansion of C with respect to an MCB $\mathcal{M}$ satisfies

$$\mathcal{E}_{\mathcal{M}}(C) \subseteq \bigcup_{C' \in \mathcal{X}} \mathcal{E}_{\mathcal{M}}(C'). \quad (25)$$

By construction of $\mathcal{X}$ and Statement 1 of Lemma 4, the cycles in $\mathcal{E}_{\mathcal{M}}(C)$ are smaller

$$|B| \leq |C'| < |C|, \quad C' \in \mathcal{X}, B \in \mathcal{E}_{\mathcal{M}}(C') \quad (26)$$

as claimed. $\square$

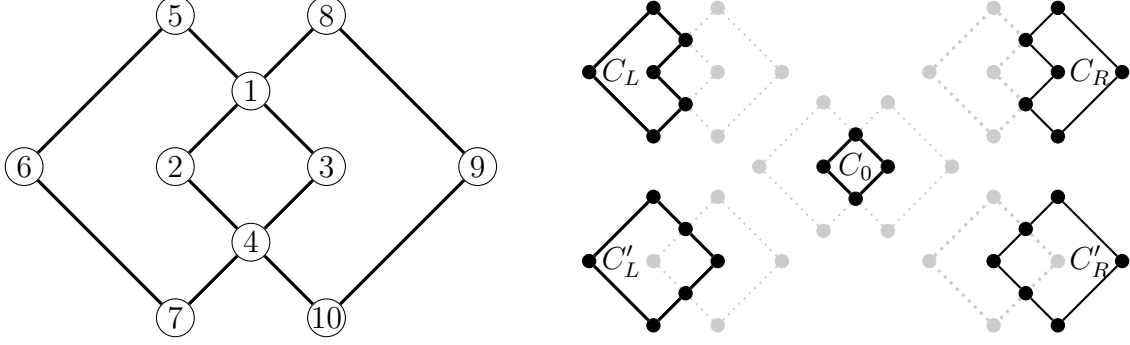


Figure 7: (Adapted from Ref. [1], Figure 4). A graph consisting of two overlapping diamonds. The graph has five relevant cycles $\mathcal{C}_R = \{C_L, C'_L, C_0, C_R, C'_R\}$, depicted separately. There are four MCBs $\mathcal{M}_{1,1} = \{C_L, C_0, C_R\}$, $\mathcal{M}_{1,2} = \{C_L, C_0, C'_R\}$, $\mathcal{M}_{2,1} = \{C'_L, C_0, C_R\}$, $\mathcal{M}_{2,2} = \{C'_L, C_0, C'_R\}$, where each MCB has one left cycle in $\{C_L, C'_L\}$, one right cycle in $\{C_R, C'_R\}$, and center cycle C_0 .

2.4 Partitions of the relevant cycles

The V-families output by Vismara's algorithm [29] are often more useful than the relevant cycles themselves. Unfortunately, these families are not unique, as they rely on a particular enumeration of the vertices. Here we review two works that extend V-families to unique partitions of the relevant cycles [1, 31].

Gleiss et al. [1] extend the V-families to include cycles that can be obtained by the basis exchange property. They achieve this with the following.

Definition 7 (Ref. [1], Definition 6). *Two relevant cycles $C_1, C_2 \in \mathcal{C}_R$ of equal length $|C_1| = |C_2|$ are interchangeable if there exist cycles $\mathcal{X} \subset \mathcal{C}_R$, $C_1, C_2 \notin \mathcal{X}$, such that*

1. $\mathcal{X} \cup \{C_2\}$ is a linearly independent set of relevant cycles,
2. $C_1 = C_2 \oplus \bigoplus_{C' \in \mathcal{X}} C'$,
3. $|C'| \leq |C_1|$ for $C' \in \mathcal{X}$.

In words, C_1 and C_2 are interchangeable if they are of equal length, and C_1 is a sum of C_2 and cycles of smaller or equal length.

By linear independence (condition 1), there exists a basis $\mathcal{B}_2$ such that $\mathcal{X} \cup \{C_2\} \subseteq \mathcal{B}_2$. The expansion of C_1 is $\mathcal{E}_{\mathcal{B}_2}(C_1) = \mathcal{X} \cup \{C_2\}$ by condition 2. By Lemma 3, the exchange $\mathcal{B}_1 = (\mathcal{B}_2 \setminus \{C_2\}) \cup \{C_1\}$ is a basis.

Interchangeability is an equivalence relation that partitions the relevant cycles into classes. Cycles in a V-family are interchangeable, so the number of interchangeability classes is polynomial in number. For the chained diamonds in Figure 6, all large relevant cycles belong to the same interchangeability class. Interchangeability classes are tricky to compute because they can not be obtained using a single cycle prototype from each V-family [1]. With additional theory, interchangeability classes can be computed in polynomial time [45].

C_1, C_2 being interchangeable does not guarantee the existence of an MCB $\mathcal{M}_2$ where $\mathcal{M}_1 = (\mathcal{M}_2 \setminus \{C_2\}) \cup \{C_1\}$ is also minimal. To see this, consider the graph in Figure 7. First, we observe that $C_L = C_0 \oplus C'_L$ and $C_0 = C_L \oplus C'_L$. Similarly, for the right cycles

$$\begin{aligned} C_R &= C_0 \oplus C'_R = (C_L \oplus C'_L) \oplus C'_R \\ &= C_L \oplus (C'_L \oplus C'_R) \end{aligned} \tag{27}$$

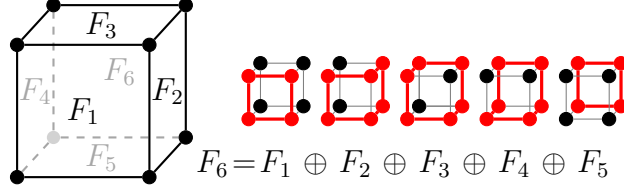


Figure 8: Depiction of the cube graph with labeled face cycles $F_1, F_2, \dots, F_6$. The partial sums of $F_1 \oplus F_2 \oplus \dots \oplus F_5$ are depicted in red to show how they sum to F_6 .

where $|C_L| = |C'_L| = |C'_R| = |C_R|$. Thus, C_L and C_R are interchangeable, as in Definition 7, where $\mathcal{X} = \{C'_L, C'_R\}$. However, each MCB has exactly one left and right cycle, so we cannot exchange C_L for C_R between MCBs.

The interchangeability relation splits cycles into classes that may be too coarse for some applications. [31] As an illustrative example, consider the cube graph, where the relevant cycles are the six face cycles $F_1, F_2, \dots, F_6$. Any face is the sum of the other face cycles, e.g.,

$$F_6 = F_1 \oplus F_2 \oplus F_3 \oplus F_4 \oplus F_5. \quad (28)$$

See Figure 8. Thus, all face cycles belong to the same interchangeability class. Kolodzik et al. [31] argue that this interchangeability class is too coarse and propose the face cycles should be placed in separate groups. They achieve this using a more restrictive pairwise relation that closely resembles the relation of cycles in Vismara's families.

Definition 8 (Ref. [31], Definition 1). *Let $C_1, C_2 \in \mathcal{C}_{\mathcal{R}}$, where $|C_1| = |C_2|$, then C_1 and C_2 are unique ring family (URF)-pair-related if*

1. C_1 and C_2 share at least one common edge
2. There exists a set of strictly smaller cycles $\mathcal{X}$ such that $C_1 = C_2 \oplus \bigoplus_{C' \in \mathcal{X}} C'$.

The key distinction here is the cycles $\mathcal{X}$ used to exchange C_1 and C_2 are strictly smaller. The faces of the cube are of the same length, so Eq. (28) does not satisfy the URF-pair-relation. For the graph composed of chained diamonds in Figure 6, the large cycles are URF-pair-related because they are exchanged by adding shorter diamonds. More generally, cycles in a V-family satisfy the URF-pair-relation.

The URF-pair-relation is not transitive, so it is not an equivalence relation. Instead, the URFs are given by the transitive closure of the URF-pair-relation. The URFs can be computed in polynomial time using the V-families. [31]

3 Theory: Concepts

In this section, we introduce two partitions of the relevant cycles and related theory. The definitions of these partitions are inspired by those in Refs. [1] and [31]. We summarize the granularity of these partitions in Figure 9.

At a coarse level, we define the polyhedron-interchangeability (**pi**) relation, a modification of the interchangeability relation from Ref. [1]. Unlike the interchangeability relation, the **pi** relation is defined using the basis exchange property between MCBs. This allows us to more precisely characterize the lack of uniqueness in MCBs. We define the **pi** relation and elaborate on its

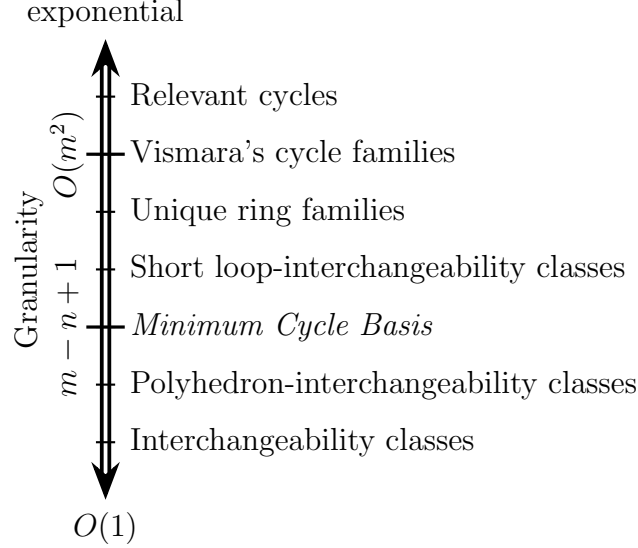


Figure 9: Relative granularity of each partition of the relevant cycles. The more granular partitions are refinements of the coarser ones. The axis represents the number of sets in each partition. Notably, a **pi** class is a collection of **sli** classes, and a **sli** class is a collection of V-families. An MCB is not a partition of the relevant cycles, but it is useful for comparing the granularity of each partition. A **pi** class has a fixed number of cycles in each MCB; a **sli** class contains at most one cycle in any given MCB.

advantages in Section 3.1. Also, we introduce a physical interpretation of pairs of cycles that are **pi-interchangeable** as being shared faces of an abstract polyhedron. We describe this in Section 3.1.2. Thus, we suggest the intuition to interpret **pi** classes as polyhedron clusters.

At a granular level, we define the short loop-interchangeability (**sli**) relation, which modifies the URFs from Ref. [31]. In particular, the **sli** relation is equivalent to the URF-pair-relation except that cycles do not need to share edges. Our definition is more amenable to theoretical results, e.g., **sli** is an equivalence relation. In computation, the **sli** classes are used in place of cycles since they are polynomial in number. The **sli** relation is introduced in Section 3.2.

3.1 Polyhedron-interchangeability: definition and basic concepts

Here, we define and investigate the **pi** relation.

Definition 9. $C_1, C_2 \in \mathcal{C}_{\mathcal{R}}$ are polyhedron-interchangeable if there exists an MCB $\mathcal{M}_2$, $C_2 \in \mathcal{M}_2$, such that $\mathcal{M}_1 = (\mathcal{M}_2 \setminus \{C_2\}) \cup \{C_1\}$ is also an MCB.

We define the operator $C_1 \xleftrightarrow{\text{pi}} C_2$ to denote C_1 is polyhedron-interchangeable for C_2 .

An equivalent definition was introduced in Ref. [1] as a “stronger interchangeability relation”, but it was not analyzed in detail. In particular, they argue this relation should be avoided because it is not symmetric. However, we find this is incorrect and that **pi** is an equivalence relation. Furthermore, the **pi** relation has two important benefits:

- (*Interpretability*) Two cycles C_1 and C_2 such that $C_1 \xleftrightarrow{\text{pi}} C_2$ are related by the basis exchange property between MCBs. This is not the case for interchangeability. C_L and C_R in Figure 7

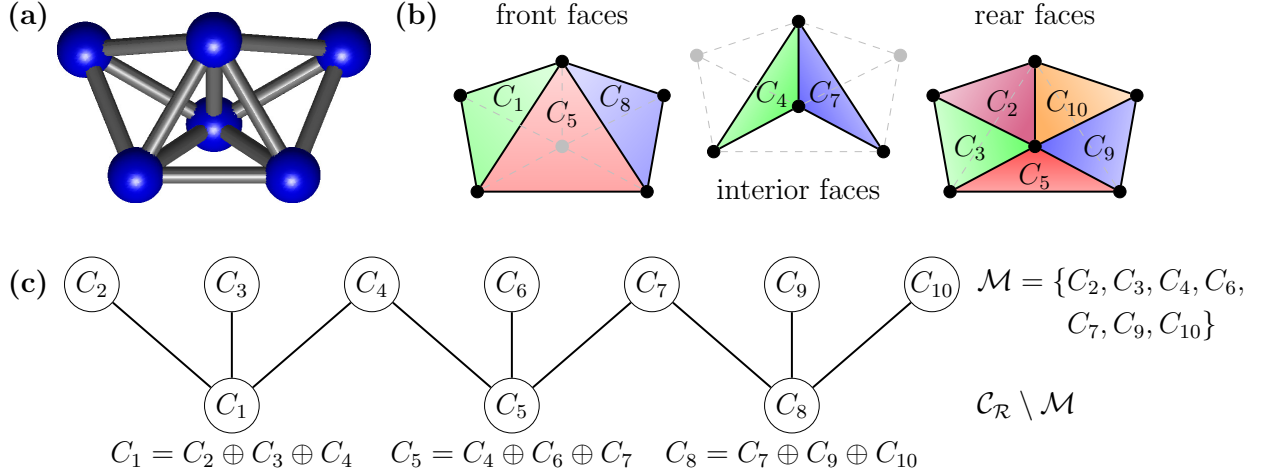


Figure 10: An example showing relevant cycles and the **pi** classes. **(a)** Graph with 6 nodes composed of 3 overlapping tetrahedra. **(b)** Depiction of the relevant cycles $\mathcal{C}_{\mathcal{R}} = \{C_1, \dots, C_{10}\}$. **(c)** Bipartite graph representation used to compute the **pi** classes.

are interchangeable by Definition 7, but they are not related by the exchange of cycles in MCBs.

- *(Computability)* The **pi** classes are computed in polynomial time using the individual cycle prototypes from the V-families. See Theorem 8 and the following discussion. Interchangeability classes are also computable in polynomial time, but they require additional theory [45].

3.1.1 Theoretical properties

Here, we state our theoretical results for the **pi** relation and their implications.

Theorem 6. ***pi** is an equivalence relation on $\mathcal{C}_{\mathcal{R}}$.*

The proofs of Theorem 8 and the other results in Section 3.1.1 are in Appendix A.2.

As an equivalence relation, **pi** partitions the relevant cycles into equivalence classes. Although **pi** satisfies the mathematical definition of an equivalence relation, we do not consider a pair of **pi** cycles to be equivalent. For instance, the faces of the cube are **pi**, but they are visually distinct. We find the **pi** classes are better interpreted as polyhedron clusters as discussed in Section 3.1.2.

Definition 9 for **pi** is valuable because it explicitly uses the basis exchange property. However, Definition 9 is hard to test in practice. Thus, it is also useful to have an equivalent, algebraic statement of this definition, which we state below.

Lemma 7. *Let $C_1, C_2 \in \mathcal{C}_{\mathcal{R}}$ where $|C_1| = |C_2|$. $C_1 \xrightarrow{\text{pi}} C_2$ if and only if there exist cycles $\mathcal{X} \subseteq \mathcal{C}_{\mathcal{R}}$, $C_1, C_2 \notin \mathcal{X}$, such that*

1. *there exists an MCB $\mathcal{M}_2$ such that $\mathcal{X} \cup \{C_2\} \subseteq \mathcal{M}_2$ and,*
2. *C_1 is given by the sum*

$$C_1 = C_2 \oplus \bigoplus_{C' \in \mathcal{X}} C', \quad (29)$$

i.e., the expansion of C_1 in $\mathcal{M}_2$ is given by $\mathcal{E}_{\mathcal{M}_2}(C_1) = \mathcal{X} \cup \{C_2\}$.

Note, the cycles in $\mathcal{X}$ are bounded in length, $|C'| < |C_1|$ for $C' \in \mathcal{X}$, by Statement 1 of Lemma 4. Moreover, $\mathcal{X}$ satisfies the conditions for C_1, C_2 to be interchangeable as in Ref. [1] (Definition 7) with the added constraint that $\mathcal{X} \cup \{C_2\}$ belongs to a MCB. Thus, pi implies interchangeability. The reverse is not true, e.g., C_L and C_R in Figure 7 are interchangeable but not pi -interchangeable.

In brief, Lemma 7 states a cycle C_1 is pi -related to the equal length cycles in its expansion with respect to a MCB $\mathcal{M}_2$. One may wonder, do we need many MCBs to construct the pi classes? Fortunately, it is sufficient to generate a single MCB, as stated below.

Theorem 8. *Let $\mathcal{M}$ be a given MCB. Let $C_2 \in \mathcal{M}$ and $C_1 \in \mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}$. If $C_2 \in \mathcal{E}_{\mathcal{M}}(C_1)$ and $|C_1| = |C_2|$ then $C_1 \xleftrightarrow{\text{pi}} C_2$. The pi classes are given by the transitive closure of these relations.*

Theorem 8 reduces the task of computing the pi classes to breadth-first search. As input, we compute an MCB $\mathcal{M}$ and the relevant cycles $\mathcal{C}_{\mathcal{R}}$. Then, we expand the relevant cycles in $\mathcal{M}$. Using this, we create a bipartite graph, where nodes represent the relevant cycles split by if they are in $\mathcal{M}$. Two cycles $C_1 \in \mathcal{M}$ and $C_2 \in \mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}$ are joined by an edge if $C_2 \in \mathcal{E}_{\mathcal{M}}(C_1)$. The pi classes are the connected components of this bipartite graph. See Figure 10 (c) for an instance of this bipartite graph.

The pi relation can be used to construct a path from one MCB to another by only swapping one pair of cycles at a time.

Lemma 9. *Let $\mathcal{M}_1, \mathcal{M}_2$ be distinct MCBs. There exists $C_1 \in \mathcal{M}_1 \setminus \mathcal{M}_2$ and $C_2 \in \mathcal{M}_2 \setminus \mathcal{M}_1$ such that $\mathcal{M}'_1 = (\mathcal{M}_1 \setminus \{C_1\}) \cup \{C_2\}$ is an MCB.*

In this procedure, each cycle in $\mathcal{M}_1$ also belongs to $\mathcal{M}_2$ or it is pi -interchangeable for a cycle in $\mathcal{M}_2 \setminus \mathcal{M}_1$. This results in a bijective mapping from $\mathcal{M}_1$ and $\mathcal{M}_2$ between pi -interchangeable cycles. From this, we obtain the following result.

Corollary 10. *Let $\mathcal{M}_1, \mathcal{M}_2$ be MCBs and $\mathcal{PI} \subseteq \mathcal{C}_{\mathcal{R}}$ be a pi class. Then, $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same number of cycles in $\mathcal{PI}$:*

$$|\mathcal{M}_1 \cap \mathcal{PI}| = |\mathcal{M}_2 \cap \mathcal{PI}|. \quad (30)$$

This introduces a notion of rank for the pi classes, given by the number of cycles in the pi class that belong to a MCB $|\mathcal{M} \cap \mathcal{PI}|$.

3.1.2 pi classes as polyhedron clusters

The partitions of the relevant cycles were introduced as a means of addressing the lack of uniqueness in MCBs. In the examples considered so far, lack of uniqueness can be caused by symmetries or by three-dimensional data. Here, we will discuss the latter. Specifically, we will discuss how pi classes may be used to identify three-dimensional structures.

First, consider a graph composed of a convex polyhedron with face cycles $F_1, \dots, F_k$ ordered by decreasing length $|F_1| \geq |F_2| \geq \dots \geq |F_k|$. For instance, our graph could be one of the platonic solids, e.g., the tetrahedron or cube. We observe, that each edge in the polyhedron participates in two face cycles. Thus, all edges cancel in the sum $F_1 \oplus \dots \oplus F_k$. In particular, the face cycles sum to the empty set

$$F_1 \oplus \dots \oplus F_k = \emptyset. \quad (31)$$

Next, we re-arrange Eq. (31) to express F_1 as the sum of the remaining face cycles

$$F_1 = F_2 \oplus (F_3 \oplus \dots \oplus F_k). \quad (32)$$

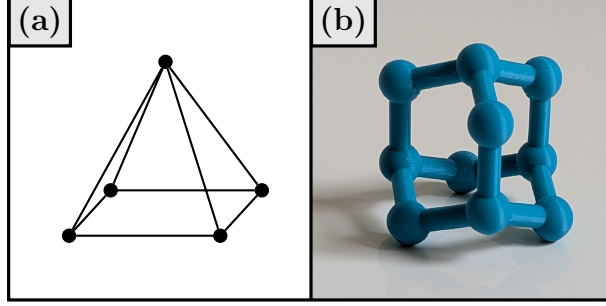


Figure 11: **(a)** The square pyramid, a polyhedron whose faces do not form a **pi** class. The unique MCB of this graph is the set of triangular faces. **(b)** Carbon skeleton of adamantane, $C_{10}H_{16}$, the smallest diamondoid, or a small molecule with the same structure as diamond. The faces of adamantane are a **pi** class, but they are not flat as in a classical polyhedron.

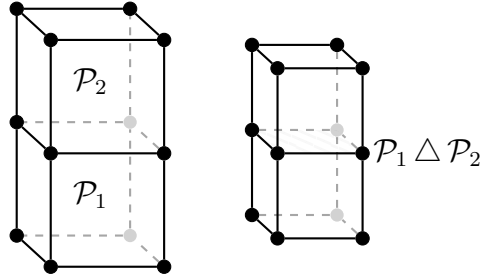


Figure 12: Graph composed of two cubes $\mathcal{P}_1$ and $\mathcal{P}_2$. Both $\mathcal{P}_1$ and $\mathcal{P}_2$ are sets of six face cycles. The symmetric difference $\mathcal{P}_1 \triangle \mathcal{P}_2$ (depicted right) gives the ten faces of the larger rectangular prism containing $\mathcal{P}_1$ and $\mathcal{P}_2$. The interior face cycle (shaded with gray stripes) belonging to both $\mathcal{P}_1$ and $\mathcal{P}_2$ is not present in $\mathcal{P}_1 \triangle \mathcal{P}_2$.

See Figure 8 for the case of the cube. Eq. (32) resembles Eq. (29) in the definition of the **pi** relation given by Lemma 7. Indeed, if we add two constraints (i) $|F_1| = |F_2|$ and (ii) $F_2, \dots, F_k$ form an MCB, then F_1 and F_2 are **pi**-interchangeable by Lemma 7.

We remark, not all polyhedra can be used to obtain a pair of face cycles F_1, F_2 such that $F_1 \xleftrightarrow{\text{pi}} F_2$. For example, if F_1 is strictly longer than the other faces $|F_1| > |F_i|$, $i = 2, \dots, k$, then F_1 is not relevant as the sum of smaller cycles by Eq. (32). This is the case for the square pyramid, where the square base is the largest cycle in Figure 11 (a). Nevertheless, polyhedra are useful for constructing examples for the **pi** relation.

The **pi** classes can be used to identify three-dimensional structures. If C_1, C_2 are **pi**-interchangeable, then we can use Lemma 7 to obtain a set of cycles $\mathcal{P} = \{C_1, C_2\} \cup \mathcal{X}$ with null sum

$$C_1 \oplus C_2 \oplus \bigoplus_{C' \in \mathcal{X}} C' = \emptyset. \quad (33)$$

This matches the sum of face cycles for polyhedra, Eq. (31). Thus, we propose to interpret these sets $\mathcal{P}$ with null sums obtained by Lemma 7 as polyhedra.

It is not necessarily the case that we can draw these null sums as polyhedra, e.g., adamantane in Figure 11 (b) does not have flat faces. Instead, the sets $\mathcal{P}$ obtained via **pi** relations might be more appropriately defined as abstract polyhedra [46] with possibly curved faces and edges.

We remark, a **pi** class is often not a single polyhedron. For instance, if we glue two cubes together so that they share a face as shown in Figure 12, then all square faces are **pi**-interchangeable

by transitivity. Thus, in the same way that a biconnected component is a cluster of loops, we argue that a **pi** class should be interpreted as a cluster of polyhedra.

Furthermore, we may extend the polyhedra obtained by a **pi** class to a vector space. As before, the addition operator of this vector space is the symmetric difference, and the scalars are the binary GF(2) field. This “polyhedron space” is simply the null space of the relevant cycles, i.e., if polyhedra $\mathcal{P}_1$ and $\mathcal{P}_2$ have null sums then so does their symmetric difference

$$\bigoplus_{C' \in \mathcal{P}_1 \triangle \mathcal{P}_2} C' = \bigoplus_{C' \in \mathcal{P}_1} C' \oplus \bigoplus_{C'' \in \mathcal{P}_2} C''. \quad (34)$$

A polyhedron space is likely useful for graphs representing material structures, where there are three intrinsic spatial dimensions. It appears pertinent to define a notion of “minimum polyhedron basis” to serve as a summary statistic for this space. We lack the algorithms and data to justify such a construction.

3.2 Short loop-interchangeability: definition, equivalence classes, & computation

Here, we introduce the second **sli** relation used to partition the relevant cycles. To motivate this, we discuss where the **pi** classes are less effective as a descriptive tool.

In Section 3.1.2, the **pi** relation was related to three-dimensional structures, i.e., polyhedra with non-flat faces. Consider the graph shown in Figure 13 (a). The polyhedra are sampled using an MCB $\mathcal{M}$ by merging each cycle $C \in \mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}$ with its expansion $\mathcal{E}_{\mathcal{M}}(C)$ as in Figure 13 (b). These polyhedra describe three-dimensional structures like beads of a bracelet, as in Figure 13 (c).

Visually, the third polyhedron $\mathcal{P}_3$ appears redundant, where its two beads are a combination of the other polyhedra. This is not resolved by treating the polyhedra as a vector space since $\mathcal{P}_3 \neq \mathcal{P}_1 \triangle \mathcal{P}_2$. As the bracelet grows in length, the number of visually redundant polyhedra grows exponentially with the number of relevant cycles.

To address this, we observe the cycles in the examples above vary in length. We argue that larger cycles that differ by shorter ones should be treated as equivalent. For instance, we would group the circumference cycles around the carbon nanotube into an equivalence class without sampling polyhedra. We formalize this equivalence with the **sli** relation.

Definition 10. $C_1, C_2 \in \mathcal{C}_{\mathcal{R}}$, $|C_1| = |C_2|$, are short loop-interchangeable if

$$C_1 = C_2 \oplus (\text{smaller cycles}), \quad (35)$$

i.e., there exist cycles $\mathcal{Y}$, $|C| < |C_1|$ for $C \in \mathcal{Y}$, such that

$$C_1 = C_2 \oplus \bigoplus_{C \in \mathcal{Y}} C. \quad (36)$$

We use the operator $C_1 \xleftrightarrow{\text{sli}} C_2$ to denote C_1 is short loop-interchangeable for C_2 .

The numerous cycles around the circumference of the carbon nanotube in Figure 14 (a) form a single **sli** class.

sli is a stronger relation than **pi**. A pair of **pi**-related cycles may differ by equal length cycles. Therefore, the faces of a polyhedron with uniform lengths (e.g., the cube) form a **pi** class but not a **sli** class. We sample polyhedra from **pi** classes that do not reduce to a **sli** class. This is discussed in more detail in Section 4.4.

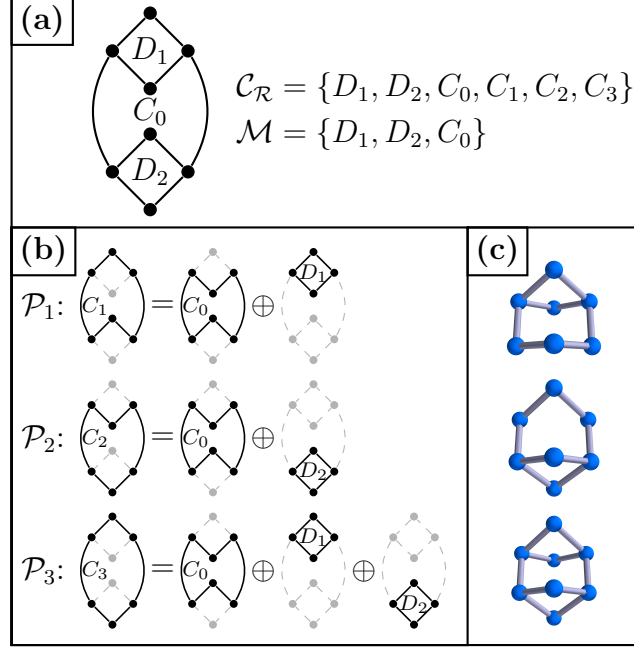


Figure 13: (a) Bracelet graph composed of two diamonds chained together. The planar faces form an MCB $\mathcal{M}$. (b) Three polyhedra $\mathcal{P}_1 = \{C_1, C_0, D_1\}$, $\mathcal{P}_2 = \{C_2, C_0, D_2\}$, and $\mathcal{P}_3 = \{C_3, C_0, D_1, D_2\}$ constructed using the cycles not in $\mathcal{M}$ and their expansions in $\mathcal{M}$. (c) Three-dimensional representation of the polyhedra as the union of their face cycles.

Definition 10 was first introduced in the dissertation by Gleiss [32] as a stronger interchangeability relation. Two of the results shown here, Lemma 11 and Corollary 14, were first shown there. We include them for completeness. Here, we add a method of computing the **sli** classes, Lemma 12. Also, our two-fold partition using both **pi** and **sli** classes is novel. The proofs of these results are in Appendix A.3

Our approach of equivocating cycles that differ by shorter cycles is inspired by the URFs in Ref. [31]. The **sli** relation is weaker than the URF-pair-relation. Specifically, cycles do not need to share edges to be **sli**-interchangeable. For example, the two hexagonal cycles around the perimeter of the prism shown in Figure 14 (b) are **sli**-interchangeable but not URF-pair-related. In MD data, we find these two definitions behave similarly. We use the **sli** relation because it is simpler, making it easier to develop theory.

Lemma 11. ***sli** is an equivalence relation on $\mathcal{C}_{\mathcal{R}}$.*

By Lemma 11, the **sli** relation partitions the relevant cycles into equivalence classes. To compute these classes, we redefine **sli** using an MCB.

Lemma 12. *Let $\mathcal{M}$ be a given MCB. $C_1, C_2 \in \mathcal{C}_{\mathcal{R}}$, $|C_1| = |C_2|$, are **sli**-interchangeable if and only if their expansions in $\mathcal{M}$ are equal up to shorter cycles.*

Cycles in Vismara's families differ by shorter cycles. Therefore, the relevant cycles can be broken into **sli** classes in polynomial time using Vismara's cycle families.

Note, if a cycle C_1 belongs to the MCB $\mathcal{M}_1$, then its expansion in $\mathcal{M}_1$ is $\mathcal{E}_{\mathcal{M}_1}(C_1) = \{C_1\}$. If $C_2 \xleftrightarrow{\text{sli}} C_1$ then the expansion of C_2 in $\mathcal{M}_1$ is C_1 and shorter cycles by Lemma 12. Thus, we may exchange C_2 for C_1 in $\mathcal{M}_1$ to obtain a new MCB by Lemma 4. We state this as a corollary.

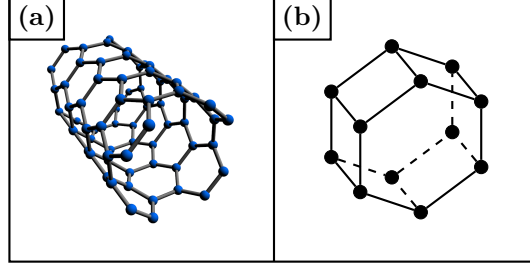


Figure 14: (a) Small carbon nanotube. The large cycles around the circumference of the nanotube are both a **sli** class and a URF. (b) Hexagonal prism graph. The two hexagonal faces form a **sli** class but not a URF.

Corollary 13. *Let C_1, C_2 be cycles such that $C_1 \xleftrightarrow{\text{sli}} C_2$. Let $\mathcal{M}_1$ be an MCB where $C_1 \in \mathcal{M}_1$. Then, $\mathcal{M}_2 = (\mathcal{M}_1 \setminus \{C_1\}) \cup \{C_2\}$ is an MCB.*

Also, observe C_2 is not in the MCB $\mathcal{M}_1$ since it is already in the span of C_1 and shorter cycles in $\mathcal{M}_1$. Therefore, C_1 is the only cycle from its **sli** class that belongs to $\mathcal{M}_1$:

Corollary 14. *Let $\mathcal{M}$ be an MCB and $\mathcal{S} \subseteq \mathcal{C}_{\mathcal{R}}$ a **sli** class. At most one cycle in $\mathcal{S}$ belongs to $\mathcal{M}$, i.e., $|\mathcal{M} \cap \mathcal{S}| \leq 1$.*

Note, the intersection may be empty, i.e., $|\mathcal{M} \cap \mathcal{S}| = 0$. For instance, consider the cube where each face is its own **sli** class, and only five faces belong to each MCB.

4 Algorithms for sampling **pi** and **sli** classes

In this section, we review and upgrade algorithms for computing the relevant cycle families, **pi** classes, and **sli** classes. In Section 4.1, we modify the relevant cycle families from Ref. [29] to be easier to compute in our setting. In Section 4.2, we show how a MCB can be extracted from these families using the algorithms from Ref. [42]. In Section 4.3, we show how to compute the expansion of cycles with respect to a MCB. In Section 4.4, we compute the **pi** and **sli** classes using the MCB representation of the cycle families.

4.1 Modified Vismara's cycle families

4.1.1 Definition

Here, we describe a method for computing candidate families of relevant cycles. We modify the construction of the V-families as introduced by Vismara [29] by rooting at an edge belonging to a fundamental cycle basis rather than a node. Compare Figures 5 and 15. We call the modified cycle families V'-families.

Definition 11 (V'-family). *Let G be a graph and T a given spanning tree of G . Let $e_j \in E(G) \setminus E(T)$ be an enumeration of the edges outside of the spanning tree T .*

- The odd V'-family with descriptors $e_j = (u_j, v_j)$ and node p , where $d(u_j, p) = d(v_j, p)$, is the set of cycles with largest-indexed edge e_j in $E(G) \setminus E(T)$ composed of e_j and shortest path pairs from u_j and v_j to p . See Figure 15, left.

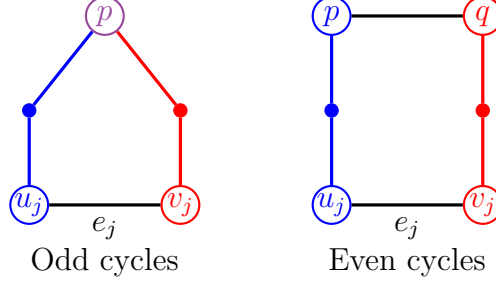


Figure 15: (Left) A cycle of length 5 whose descriptors are the largest-indexed edge $e_j \in E(G) \setminus E(T)$ for some spanning tree T and the node p antipodal to e_j . (Right) A cycle of length 6 whose descriptors are its largest-indexed edge $e_j = (u, v)$ and antipodal edge (p, q) . Nodes are colored red or blue if they are closer to u or v respectively. The left node p is purple because it is equidistant to u and v .

- *The even V' -family with descriptors $e_j = (u_j, v_j)$ and edge (p, q) , where $d(u_j, p) = d(v_j, q)$, is the set of cycles with largest-indexed edge e_j composed of e_j , (p, q) , and shortest path pairs from u_j to p and from v_j to q . See Figure 15, right.*

This has the following benefits:

- **Node labeling.** The shortest paths for the V' -families are computed using breadth-first search starting at the root edge $e_j = (u_j, v_j) \in E(G) \setminus E(T)$. By rooting at an edge, we may recursively label whether a node is closer to u_j or v_j . See the blue and red nodes in Figure 15.
- **Computation time.** Repeatedly traversing the graph to construct V' -families is the primary bottleneck of our algorithm. To compute the V' -families, breadth-first search is run once per root edge in $E(G) \setminus E(T)$, the number of which is equal to the dimension of the cycle space ν . For the original Vismara families, breadth-first search is run once per node. In our simulations of hydrocarbon pyrolysis, the dimension of the cycle space ranges from $0.01n \leq \nu \leq 0.2n$. This reduces the number of times we traverse the graph by a scaling factor that depends on initial conditions.

For diamond, with mean degree $\langle k \rangle = 4$, the dimension of the cycle space is $\nu = m - n + 1 \sim n$ since $n\langle k \rangle = 2m$. Thus, even for reasonably dense materials, the number of times we traverse the graph does not increase. This result does not extend to arbitrarily dense graphs where the assumption $m = O(n)$ fails.

- **Simple descriptors.** By modifying the original V -families, the descriptors of even cycles change from a node and connected triple to a pair of edges. Connected triples are harder to work with, e.g., the number of edges m is given whereas the number of connected triples as a function of the degree sequence $\{k_i\}$ is given by $\sum_i \binom{k_i}{2}$.

All relevant cycles belong to a V' -family, but not all V' -families contain relevant cycles. The method for filtering out irrelevant V' -families is discussed later in Section 4.4.

4.1.2 An algorithm for computing V' -families

We assume a spanning tree T is given and the edges $E(G) \setminus E(T)$ are enumerated. Algorithm 1 computes the V' -families by looping over each edge $e_j \in E(G) \setminus E(T)$.

Algorithm 1: Compute V' -families

Input: • G = an undirected, unweighted graph,

- $\mathcal{F}$ = fundamental cycle basis of G with cycles associated with edges $e_j \in E(G) \setminus E(T)$,
 $j = 1, \dots, \nu$, where T is a spanning tree of G

Output: • V' -families represented by their descriptors,

- D_j = a directed, acyclic network for computing shortest paths to e_j for $j = 1, \dots, \nu$,
- node properties ancestor_j , distance_j , num_paths_j for $j = 1, \dots, \nu$;

```
1 for  $e_j = (u_j, v_j)$  in  $E(G) \setminus E(T)$  do
2   Initialize root nodes  $\text{ancestor}_j[u_j] \leftarrow u_j, \text{distance}_j[u_j] \leftarrow 0, \text{num\_paths}_j[u_j] \leftarrow 1$ , and
    $\text{ancestor}_j[v_j] \leftarrow v_j, \text{distance}_j[v_j] \leftarrow 0, \text{num\_paths}_j[v_j] \leftarrow 1$ ;
3   For each remaining node  $x$ , mark  $\text{observed}[x] \leftarrow \text{False}$  and  $\text{valid}[x] \leftarrow \text{True}$ ;
4   Initialize (FIFO)  $\text{queue} \leftarrow [u_j, v_j]$ ;
5   for  $p$  in  $\text{queue}$  do
6     #  $p$  is a descriptor for an odd  $V'$ -family if it has two ancestors  $u_j$  and  $v_j$ ;
7     if  $\text{ancestor}_j[p] = \{u_j, v_j\}$  and  $\text{valid}[p]$  then
8       Count shortest paths from  $p$  to  $u_j$ :  $n_1 = \sum_{x, p \rightarrow x \in D_j} \text{num\_paths}_j[x] \cdot \mathbf{1}[\text{ancestor}_j[x] = u_j]$ ;
9       Count shortest paths from  $p$  to  $v_j$ :  $n_2 = \sum_{x, p \rightarrow x \in D_j} \text{num\_paths}_j[x] \cdot \mathbf{1}[\text{ancestor}_j[x] = v_j]$ ;
10      if  $n_1 n_2 > 0$  then
11        Store the odd  $V'$ -family with descriptors  $e_j$  and  $p$  representing  $n_1 n_2$  cycles of length
           $2\text{distance}_j[p] + 1$ ;
12      for  $q$  neighboring  $p$  in  $G$  do
13        #  $p$  is  $q$ 's parent and  $q$  is unobserved, initialize  $q$ 's node properties;
14        if not  $\text{observed}[q]$  then
15          Mark  $\text{observed}[q] \leftarrow \text{True}$  and add  $q$  to  $\text{queue}$ ;
16           $\text{ancestor}_j[q] \leftarrow \text{ancestor}_j[p]$ ;
17           $\text{distance}_j[q] \leftarrow \text{distance}_j[p] + 1$ ;
18          if  $(p, q) = e_k \in E(G) \setminus E(T)$  where  $k > j$  then
19             $\text{num\_paths}_j[q] \leftarrow 0$ ;
20          else
21             $\text{num\_paths}_j[q] \leftarrow \text{num\_paths}_j[p]$ ;
22            Add edge  $q \rightarrow p$  to  $D_j$  # child to parent;
23          #  $p$  is  $q$ 's parent and  $q$  has been observed, update  $q$ 's node properties;
24          else if  $\text{distance}_j[q] = \text{distance}_j[p] + 1$  then
25             $\text{ancestor}_j[q] \leftarrow \text{ancestor}_j[q] \cup \text{ancestor}_j[p]$ ;
26            if  $\text{ancestor}_j[p] = \{u_j, v_j\}$  then
27              Mark  $q$  invalid;
28            if  $\text{num\_paths}_j[p] > 0$  and not  $(p, q) = e_k \in E(G) \setminus E(T)$  where  $k > j$  then
29               $\text{num\_paths}_j[q] \leftarrow \text{num\_paths}_j[q] + \text{num\_paths}_j[p]$ ;
30              Add edge  $q \rightarrow p$  to  $D_j$ ;
31          #  $(p, q)$  is a descriptor of an even  $V'$ -family if  $p$  and  $q$  have different ancestors and are
          equidistant to  $e_j$ ,  $\text{ancestor}_j[p] = u_j$  to break symmetry;
32          else if  $\text{ancestor}_j[p] = u_j$  and  $\text{ancestor}_j[q] = v_j$  then
33            if not  $(p, q) = e_k \in E(G) \setminus E(T)$  where  $k > j$  then
34              if  $\text{num\_paths}_j[p] \cdot \text{num\_paths}_j[q] > 0$  then
35                Include the even  $C'$ -family with descriptors  $e_j$  and  $(p, q)$  representing
                   $\text{num\_paths}_j[p] \cdot \text{num\_paths}_j[q]$  cycles of length  $\text{distance}_j[p] + \text{distance}_j[q] + 2$ 
36          end
37        end
38    end
```

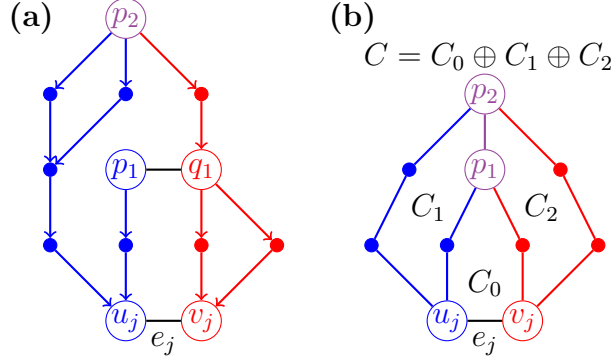


Figure 16: **(a)** The directed acyclic graph with edges pointing along shortest paths towards the root edge $e_j = (u_j, v_j)$. There are two V'-families: one with descriptors e_j and (p_1, q_1) representing 2 cycles of length 6, and one with descriptors e_j and p_2 representing 4 cycles of length 9. **(b)** An odd V'-family with descriptors e_j and p_2 where p_2 is a descendant of p_1 equidistant to u_j and v_j . The paths from p_2 to u_j and v_j in the outer cycle C are exchanged for paths through p_1 by adding C_1 and C_2 to C respectively. This shows C is not relevant because it is the sum of shorter cycles C_0, C_1, C_2 .

Algorithm 1 outputs V'-families represented by their descriptors from Definition 11. The largest-indexed edge $e_j = (u_j, v_j)$ is given, and the p and (p, q) descriptors are computed. Nodes p for odd V'-families are those that are equidistant to u_j and v_j , obtained in lines 7-11 of Algorithm 1. Edges (p, q) for even V'-families are such that p is closer to u_j than v_j , q is closer to v_j than u_j , and $d(p, u_j) = d(q, v_j)$, obtained in lines 32-35.

The cycles in a V'-family are constructed by joining the pair of descriptors by shortest paths as in Figure 15. Algorithm 1 computes shortest paths from all nodes to $e_j = (u_j, v_j)$ using breadth-first search. Here, we define the distance to e_j as

$$d(x, e_j) = \min(d(x, u_j), d(x, v_j)), \quad (37)$$

and a shortest path from x to e_j as a minimum length path from x to either u or v . As a data structure, we store a directed acyclic graph whose edges point along shortest paths to e_j – see Figure 16 (a). We call a neighbor q of node p a child of p if it is further from the root edge e_j . Edges in the directed acyclic graph point backwards from child to parent, which are assigned in lines 22 and 30.

We compute three additional node properties that further characterize the V'-families:

- **Ancestor.** A node's ancestor is the node u_j or v_j in the root edge e_j closest to it. If a node is equidistant to u_j and v_j , then both u_j and v_j are its ancestors. The ancestors of a child node q is the union of its parent's ancestors, computed in lines 16 and 25. Descendants of a node with both ancestors produce irrelevant cycles that can be decomposed into smaller cycles. See, Figure 16 (b). We mark these descendants invalid in lines 26-27.
- **Distance.** A node q will first be observed by a neighbor p lying on a shortest path to e_j . Thus, q is one edge further from e_j than its parent node p , as in line 15. This distance is used to compute the lengths of cycles in a family, as in lines 11 and 35.
- **Number of paths.** We count shortest paths from each node q to e_j as the sum of paths through its parents. See lines 19 and 26. This is used to compute the number of cycles in

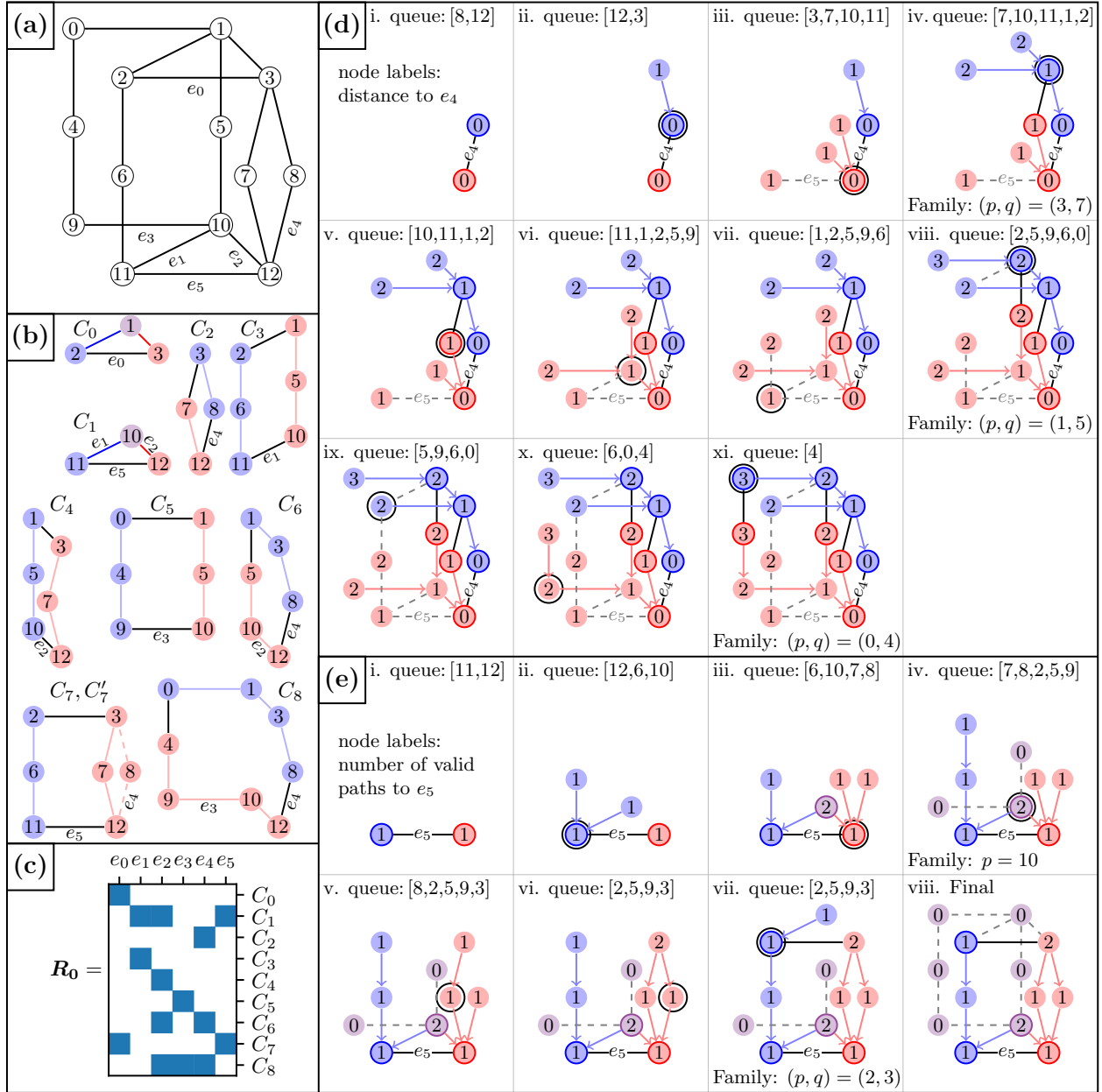


Figure 17: **(a)** A graph composed of a six-sided loop joined to a triangular prism with two paths (3, 7, 12) and (3, 8, 12) connecting 3 and 12. The enumerated edges $e_0, \dots, e_5$ are used to construct a fundamental cycle basis. **(b)** V'-families sorted by length output by Algorithm 1 iterated over $e_0, \dots, e_5$. The only V'-family with multiple cycles is $\{C_7, C'_7\}$. The secondary cycle C'_7 is plotted using dashed edges. **(c)** Matrix of representative cycles from each V'-family as fundamental cycle basis vectors. **(d)** Depiction of Algorithm 1 rooted at e_4 . Each frame represents a step through the central for loop over nodes p (circled in black). Steps $p = 5, 6, 4$ are omitted. Node colors denote ancestry, node labels denote distance to e_4 , and directed edges point along valid shortest paths to e_4 . **(e)** Depiction of Algorithm 1 rooted at e_5 , where node labels denote the number of valid shortest paths to e_5 .

a V'-family, which is equal to the number of shortest path pairs to u_j and v_j . The size of a cycle family is computed in lines 8-11 and 35.

As a final note, in each loop of Algorithm 1 we compute cycles with largest-indexed edge e_j . We reject any path that includes an edge $e_k \in E(G) \setminus E(T)$ with index $k > j$ as in lines 16-17, 23-24, and 29. We still traverse edges e_k , $k > j$, in Algorithm 1 to accurately compute the distance and ancestor node properties.

As a post-processing step, we sample a representative cycle from each V'-family. We order these cycles, and the associated V'-families, by length, i.e., $|C_i| \leq |C_j|$ for $i \leq j$. See Figure 17 (b). We convert each cycle C_i into a fundamental cycle basis vector using the non-tree edges $e_j \in E(G) \setminus E(T)$ that lie in C_i . Then, we construct a matrix whose rows are representative cycles from each V'-family as fundamental cycle basis vectors:

$$\mathbf{R}_0(i, :) = (\mathbf{1}[e_1 \in C_i], \mathbf{1}[e_2 \in C_i], \dots, \mathbf{1}[e_\nu \in C_i]), \quad (38)$$

where C_i represents family i . See Figure 17 (c).

4.1.3 The computational cost

The complexity of algorithms in this section depend on the size of $\mathbf{R}_0$. The number of columns in $\mathbf{R}_0$ is equal to the dimension of the cycle space ν . The number of rows in $\mathbf{R}_0$ is equal to the number of V'-families. $\mathbf{R}_0$ has at least ν linearly independent rows that can be used to construct an MCB. The number of V'-families is bounded by

$$\begin{aligned} \nu &\leq \# \text{ of V'-families} \\ &\leq |E(G) \setminus E(T)| (\# \text{ V'-families per edge}) \leq \nu^2 \end{aligned}$$

We justify that Algorithm 1 outputs at most ν cycles in each of its ν iterations below. This bound is tight, where the complete bipartite graph has $O(\nu^2) = O(n^4)$ cycle families.

To bound the number of V'-families per iteration of Algorithm 1, consider a breadth-first spanning tree T_j rooted at e_j . Note, T_j is not the spanning tree T used to construct a fundamental cycle basis. T_j is a subgraph of the directed acyclic graph D_j used to construct paths to e_j , as in Figure 16, where each node has out-degree 1. The edge descriptor (p, q) for an even V'-family lies outside T_j . The node descriptor p for an odd V'-family has disjoint shortest paths to u_j and v_j , so it has out-degree greater than or equal to 2 in D_j . Hence, one of these outgoing edges lies outside of T_j . In both cases, we associate an edge outside of the tree T_j with the descriptor p or (p, q) , and so the number of V'-families with root edge e_j is bounded by the number of non-tree edges ν .

In summary, $\mathbf{R}_0$ is a $\{0, 1\}^{O(\nu^2) \times \nu}$ matrix. Assuming dense matrix operations, the computational cost of the algorithms in the following sections is $O(\nu^4)$, i.e., the cost of naïvely multiplying $\mathbf{R}_0$ by a $\nu \times \nu$ matrix. However, we find $\mathbf{R}_0$ is not too large and typically very sparse. The number of rows in this matrix is typically closer to the dimension of the cycle space. In a loop cluster sampled from a carbon-rich simulation of hydrocarbon pyrolysis (initial composition $\text{C}_{10}\text{H}_{16}$), the dimension of the cycle space is $\nu = 1304$ and the number of cycle families is 1604. This observation agrees with other chemical systems where the number of relevant cycles is approximately equal to ν , see Table 8 in Ref. [30]. Also, relevant cycles are typically short, bounding the number of fundamental cycle basis edges $e_j \in E(G) \setminus E(T)$ they have. As a result, their fundamental cycle basis vectors and the matrix $\mathbf{R}_0$ are very sparse. The 1604×1304 $\mathbf{R}_0$ matrix described above has 7280 non-zero entries. In the case where $\mathbf{R}_0$ is dense, breadth-first search may be used to compute matrix vector products $\mathbf{R}_0 x$ in $O(\nu^2)$ time, as discussed in Section 5.2 in Ref. [47], which can be used to reduce the complexity of matrix multiplication to $O(\nu^3)$.

Using our Python codes on a Ryzen 9 9900X CPU, the computation of the matrix $\mathbf{R}_0$ for the system mentioned above takes approximately 5 seconds. For comparison, a sparse matrix-matrix product $\mathbf{R}_0\mathbf{W}$ is computed in less than 0.5 seconds. Indeed, Algorithm 1 is the main bottleneck of our approach. The complexity of a single iteration of Algorithm 1 is equal to the complexity of breadth-first search, i.e., $O(m)$. The total cost of Algorithm 1 is $O(\nu m)$. For the $\text{C}_{10}\text{H}_{16}$ system, $\nu m = 0.32n^2$. For diamond, as a proxy for dense material systems, $\nu m \approx 2n^2$. Therefore, it is reasonable to write the complexity as $O(n^2)$ for material networks. This procedure can be further accelerated by running Algorithm 1 in parallel, because no communication between iterations over e_j in Algorithm 1 is required.

4.2 Minimum cycle basis

Our tests for relevance and for the `pi` and `sli` relations rely on expanding cycles in terms of an MCB. See Lemma 4, Theorem 8, and Lemma 12 respectively. Here, we construct an MCB.

There is a large body of literature dedicated to the computation of MCBs. See Refs. [26, 47] for a review. We opt for the approach in Ref. [42], which we review below. This approach is computationally efficient, easy to implement in our present setting, and can be used to expand cycles in terms of the MCB.

This approach relies on an orthogonality argument using the inner product of binary vectors in the $\text{GF}(2)$ field:

$$\langle A, B \rangle_{\mathcal{F}} := \bigoplus_i a_i b_i. \quad (39)$$

We emphasize that this inner product is defined in terms of a fundamental cycle basis $\mathcal{F}$. In this context, a cycle C should be interpreted as a binary vector in $\{0, 1\}^\nu$ where entry i indicates whether C contains the i th edge outside of the spanning tree used to construct $\mathcal{F}$. See Figure 3 (b) for a reminder of this decomposition.

This construction of an MCB is iterative. Suppose cycles $B_1, \dots, B_{j-1}$ belonging to an MCB are given. To obtain the next basis cycle, we utilize a *witness vector* defined as a non-zero vector $S_j \in \{0, 1\}^\nu$ satisfying the orthogonality condition

$$\langle B_i, S_j \rangle_{\mathcal{F}} = 0, \quad i < j. \quad (40)$$

We construct the vectors S_j later in this section. The next cycle B_j in the MCB is given by the optimization problem

$$\min_{B \in \mathcal{C}} |B|, \text{ such that } \langle B, S_j \rangle_{\mathcal{F}} = 1 \quad (41)$$

where $\mathcal{C}$ is the set of all cycles. We obtain this cycle from the first row of the matrix $\mathbf{R}_0$ from Eq. (38) with inner product 1.

The last ingredient of this algorithm is to construct the witness vectors S_j . These vectors are computed dynamically, where the initial vectors $S_j, j = 1, \dots, \nu$, are the unit vector with value 1 at coordinate j . The witness vectors are updated each time a new basis vector B_j is constructed. The future witness vectors are assumed to be orthogonal to the previous cycles, i.e., $\langle B_j, S_k \rangle = 0$ for $j < k$. At step j if $\langle B_j, S_k \rangle = 1$ for $k > j$ then we update S_k as

$$S_k \leftarrow S_k \oplus S_j. \quad (42)$$

This ensures $\langle B_j, S_k \rangle = 0$ for $k > j$.

Algorithm 2: Compute MCB

Input: • G = simple, connected graph
 with ν independent cycles
 • Matrix of cycles $\mathbf{R}_0$ from V' -
 families as fundamental cycle
 basis vectors

Output: • MCB $\mathcal{M} = \{B_1, \dots, B_\nu\}$
 • Witness vectors $\{S_1, \dots, S_\nu\}$

```

1 Initialize  $S_i$  to the standard unit vector in  $\{0, 1\}^\nu$  that is 1 at coordinate  $i$  and 0 elsewhere;
2 for  $j = 1, \dots, \nu$  do
3   # Basis cycle  $B_j$  as first cycle  $C_k$  from  $\mathbf{R}_0$  with  $\langle C_k, S_j \rangle_{\mathcal{F}} = 1$ ;
4    $B_j \leftarrow$  row  $k$  of  $\mathbf{R}_0$  where  $k$  is the first non-zero entry of  $\mathbf{R}_0 S_j$ ;
5   for  $k = j + 1, \dots, \nu$  do
6     # Ensure  $\langle C_j, S_k \rangle_{\mathcal{F}} = 0$  for  $j < k$ ;
7     if  $\langle C_j, S_k \rangle_{\mathcal{F}} = 1$  then
8       |  $S_k \leftarrow S_k \oplus S_j$ ;
9   end
10 end
```

To implement the approach above, we utilize the V' -families constructed in Section 4.1. This is shown in Algorithm 2. The cycles B_j are drawn from the matrix $\mathbf{R}_0$ whose row vectors are the expansions of representative cycles from each family expanded with respect to our fundamental cycle basis $\mathcal{F}$. If C_k is the cycle represented by the k th row of $\mathbf{R}_0$, then the inner product $\langle C_k, S_j \rangle_{\mathcal{F}}$ is the k th entry in the matrix-vector product $\mathbf{R}_0 S_j$. Also, the cycles in $\mathbf{R}_0$ are sorted by length. Thus, the optimization problem in Eq. (41) is solved by the row vector in $\mathbf{R}_0$ corresponding to the first non-zero entry in $\mathbf{R}_0 S_j$, as in line 4 of Algorithm 2.

The main operations in Algorithm 2 are matrix-vector products $\mathbf{R}_0 S_j$, inner products $\langle C_j, S_k \rangle$, and vector additions $S_k \oplus S_j$. If the entries are dense, the time complexity of Algorithm 2 is $O(\nu^4)$. However, the actual cost of these operations is much less because $\mathbf{R}_0$ is very sparse. Additionally, the matrix-vector products $\mathbf{R}_0 S_j$ only need to be computed up to the first non-zero entry.

Lemma 15. *The cycles $B_1, \dots, B_\nu$ computed by Algorithm 2 form an MCB.*

Proof. The correctness of Algorithm 2 is proved via induction. Suppose $B_1, \dots, B_{j-1}$ for $j \geq 1$ belong to an MCB $\mathcal{M}_{j-1}$. The following identity is obtained by expanding B_j in $\mathcal{M}_{j-1}$ and taking the inner product with S_j

$$1 = \langle B_j, S_j \rangle_{\mathcal{F}} = \bigoplus_{B \in \mathcal{M}_{j-1}(B_j)} \langle B, S_j \rangle_{\mathcal{F}}. \quad (43)$$

The cycles $B_1, \dots, B_{j-1}$ are orthogonal to S_j . Hence, there exists $B \in \mathcal{M}_{j-1} \setminus \{B_1, \dots, B_{j-1}\}$ in the expansion of B_j in $\mathcal{M}_{j-1}$ such that $\langle B, S_j \rangle_{\mathcal{F}} = 1$. $|B| \leq |B_j|$ because B is in the expansion of B_j in $\mathcal{M}_{j-1}$. $|B_j| \leq |B|$ because B_j is a minimum length cycle such that $\langle B_j, S_j \rangle_{\mathcal{F}} = 1$. Therefore, $\mathcal{M}_j = (\mathcal{M}_{j-1} \setminus \{B\}) \cup \{B_j\}$ is an MCB by Lemma 4. By induction, $\mathcal{M}_\nu = \{B_1, \dots, B_\nu\}$ is an MCB. $\square$

Algorithm 3: Compute MCB representation

Input: • Matrix of cycles $\mathbf{R}_0$ from $\mathbf{V}'$ -families as fundamental cycle basis vectors

• MCB $\mathcal{M} = \{B_1, \dots, B_\nu\}$

• Witness vectors $S_1, \dots, S_\nu$

Output: • Matrix of cycles $\mathbf{R}_1$ from $\mathbf{V}'$ -families as MCB vectors

```

1 Initialize  $\tilde{S}_i \leftarrow S_i$  for  $i = 1, \dots, \nu$ ;
2 for  $k = \nu, \nu - 1, \dots, 2$  do
3   #  $\tilde{S}_k$  is known, update  $\tilde{S}_j$  for  $j < k$ ;
4   for  $j = 1, \dots, k - 1$  do
5     if  $\langle B_k, S_j \rangle_{\mathcal{F}} = 1$  then
6        $\tilde{S}_j \leftarrow \tilde{S}_j \oplus \tilde{S}_k$ ;
7   end
8 end
9  $\mathbf{W} \leftarrow [\tilde{S}_1, \dots, \tilde{S}_\nu]$ ;
10  $\mathbf{R}_1 \leftarrow \mathbf{R}_0 \mathbf{W}$ ;

```

4.3 Change of basis

Next, we describe our procedure to compute the expansion of a cycle C in terms of an MCB $\mathcal{M} = \{B_1, \dots, B_\nu\}$ computed by Algorithm 2. We assume C is given in vector form corresponding to its expansion with respect to the fundamental cycle basis $\mathcal{F}$ used to construct $\mathcal{M}$. Let the expansion of C with respect to $\mathcal{M}$ be given by

$$C = b_1 B_1 \oplus \dots \oplus b_\nu B_\nu, \quad b_i \in \{0, 1\}, \quad (44)$$

where the coefficients b_i are unknown. We will compute these coefficients using the witness vectors S_i .

We start with the coefficient b_ν . By construction, S_ν is orthogonal to every cycle in $\mathcal{M}$ except B_ν . Thus,

$$\langle C, S_\nu \rangle_{\mathcal{F}} = b_\nu. \quad (45)$$

For the coefficient $b_{\nu-1}$ we use the witness vector $S_{\nu-1}$ which is orthogonal to $B_1, \dots, B_{\nu-2}$ by Eq. (40). We take the inner product of Eq. (44) with $S_{\nu-1}$ and substitute $\langle C, S_\nu \rangle_{\mathcal{F}}$ for b_ν using Eq. (45)

$$\langle C, S_{\nu-1} \rangle_{\mathcal{F}} = b_{\nu-1} \oplus \langle C, S_\nu \rangle_{\mathcal{F}} \langle B_\nu, S_{\nu-1} \rangle_{\mathcal{F}}. \quad (46)$$

Then, we rearrange and solve for $b_{\nu-1}$

$$b_{\nu-1} = \langle C, S_{\nu-1} \rangle_{\mathcal{F}} \oplus \langle C, S_\nu \rangle_{\mathcal{F}} \langle B_\nu, S_{\nu-1} \rangle_{\mathcal{F}}. \quad (47)$$

We move the inner product $\langle C, S_{\nu-1} \rangle_{\mathcal{F}}$ into the right term of $\langle C, S_\nu \rangle_{\mathcal{F}}$ by linearity. After doing so – and combining similar terms – we find that Eq. (47) simplifies into a single inner product

$$b_{\nu-1} = \langle C, \tilde{S}_{\nu-1} \rangle_{\mathcal{F}}, \quad (48)$$

$$\tilde{S}_{\nu-1} = S_{\nu-1} \oplus \langle B_\nu, S_{\nu-1} \rangle_{\mathcal{F}} S_\nu. \quad (49)$$

We call $\tilde{S}_{\nu-1}$ a *modified witness vector*.

In general, the coefficients b_j can be computed via the inner product of C with a modified witness $\tilde{S}_j$. This can be verified by induction where Eq. (45) is the base case. If we assume that $b_k = \langle C, \tilde{S}_k \rangle$ for $k > j$ then the inner product of Eq. (44) with S_j is

$$\langle C, S_j \rangle_{\mathcal{F}} = b_j \oplus \bigoplus_{k=j+1}^{\nu} \underbrace{\langle C, \tilde{S}_k \rangle_{\mathcal{F}}}_{b_k} \langle B_k, S_j \rangle_{\mathcal{F}}. \quad (50)$$

Using the linearity of the inner product, we can combine terms to obtain b_j and the associated modified witness $\tilde{S}_j$

$$b_j = \langle C, \tilde{S}_j \rangle_{\mathcal{F}}, \quad (51)$$

$$\tilde{S}_j = S_j \oplus \bigoplus_{k=j+1}^{\nu} \langle B_k, S_j \rangle_{\mathcal{F}} \tilde{S}_k. \quad (52)$$

Once the modified witness vectors $\tilde{S}_j$ are computed using Eq. (52), it is useful to collect them as the columns of a matrix

$$\mathbf{W} = [\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_{\nu}]. \quad (53)$$

Then if we left-multiply $\mathbf{W}$ by our cycle C (as a vector expanded in terms of a fundamental cycle basis) we obtain its expansion in terms of $\mathcal{B}$

$$C \cdot \mathbf{W} = [\langle C, \tilde{S}_1 \rangle_{\mathcal{F}}, \langle C, \tilde{S}_2 \rangle_{\mathcal{F}}, \dots, \langle C, \tilde{S}_{\nu} \rangle_{\mathcal{F}}]. \quad (54)$$

Recall, in Section 4.1 we computed the matrix $\mathbf{R}_0$ whose rows are representative cycles from the V' -families as fundamental cycle basis vectors. If we right-multiply this matrix by $\mathbf{W}$,

$$\mathbf{R}_1 = \mathbf{R}_0 \mathbf{W}, \quad (55)$$

then we obtain the expansion of these cycles in terms of our MCB $\mathcal{M}$.

Algorithm 3 computes the matrix $\mathbf{R}_1$. Lines 2-8 compute the modified witness vectors according to Eq. (52). The inner products and vector sums in these lines are $O(\nu^3)$. In line 10, $\mathbf{R}_1$ is computed by Eq. (55). Since $\mathbf{R}_0$ is a $O(\nu^2) \times \nu$ matrix, a simple implementation of this matrix multiplication is $O(\nu^4)$. In practice, Algorithm 3 is very efficient due to sparsity.

4.4 pi and sli classes

Here we present a method for computing the **pi** and **sli** classes of an arbitrary unweighted, undirected graph. In Figure 18 (a-d), we show the major steps for this procedure for the graph shown in Figure 17 (a). We describe these steps below.

Inputs. The inputs for our algorithm are (i) the V' -families output by Algorithm 1 and (ii) the matrix $\mathbf{R}_1$ of representative cycles from each V' -family as MCB vectors from Algorithm 3. We sort the rows and columns of $\mathbf{R}_1$ by increasing cycle length. We utilize the block structure of this matrix, where the row and column cycles have fixed length. See Figure 18 (a). This block matrix is lower triangular because the MCB representation of a cycle only includes equal or shorter length cycles – see Lemma 4.

Remove irrelevant cycles. By Lemma 4, Statement 2, a cycle is relevant if and only if its MCB representation includes an equal length cycle. In Figure 18 (a), the eight-sided cycle C_8 is not

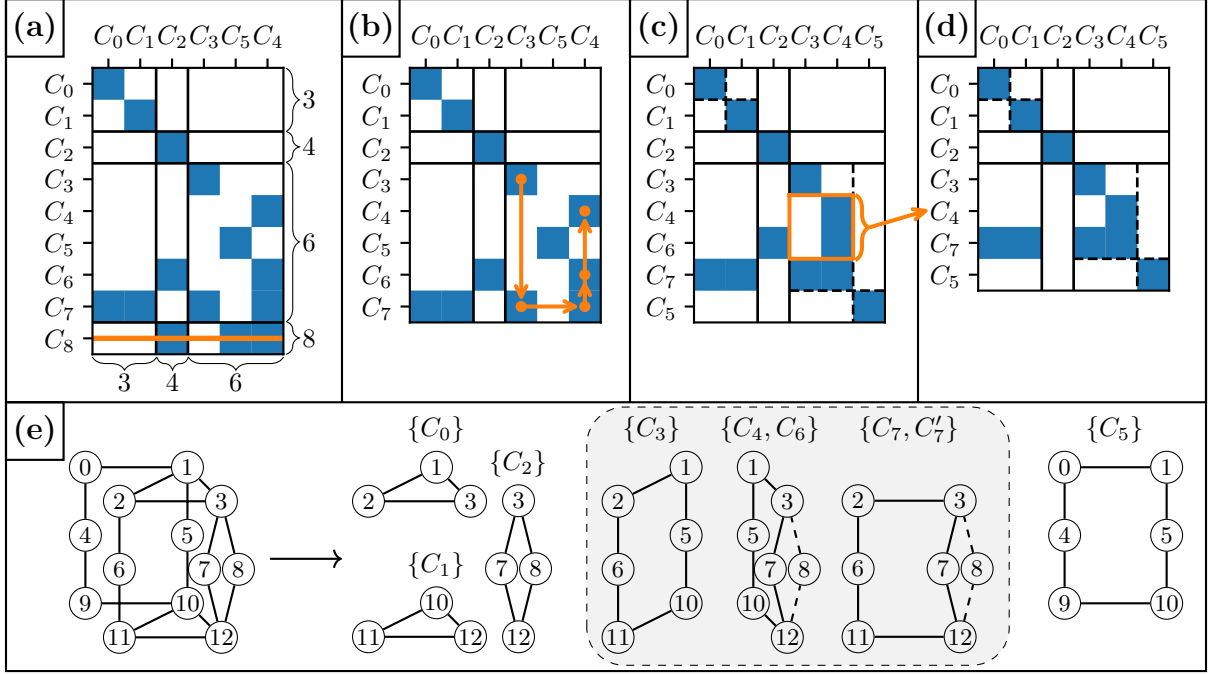


Figure 18: **(a)** Input matrix of representative cycles from each V' -family as MCB vectors. The blocks of this matrix separate cycles of differing lengths. **(b)** Reduced matrix without the irrelevant cycle C_8 since it is the sum of shorter cycles. The orange arrows depict the procedure for traversing rows and columns to compute the π classes. **(c)** The same matrix where the rows and columns are sorted by π classes. The block matrices for the π classes are separated by dashed lines. **(d)** Output matrix $\mathbf{R}$ with C_4 and C_6 merged into a single row representing a sli class, since $C_4 = C_6 \oplus$ (smaller cycles). **(e)** The graph (left) and its relevant cycles separated into sli classes (right). The representative and secondary cycles in the sli classes are drawn using solid and dashed edges. The gray box highlights the large π class $\{C_3, C_4, C_6, C_7, C_7'\}$ with multiple sli classes.

relevant as the sum of two six-sided cycles C_4, C_5 and a four-sided cycle C_2 . Algorithmically, we loop through each row to find the largest cycle in each MCB vector to check if the row is relevant and remove it otherwise.

π classes. From Lemma 7, a relevant cycle is π -interchangeable for the equal length cycles in its MCB representation. These relations are identified by the non-zero entries in the diagonal blocks of $\mathbf{R}_2$, e.g., C_7 is π -interchangeable for C_3 and C_4 in Figure 18 (b). From Theorem 8, the π classes are computed using the transitive closure of these relations. This is done by traversing the rows and columns in a diagonal block of $\mathbf{R}_2$ using a queue – see the orange arrows in Figure 18 (b).

We permute the rows and columns of $\mathbf{R}_2$ to group the π classes together. We also permute rows within a π class by

1. number of equal length cycles in the MCB vector, then by
2. lexicographic ordering of the MCB vectors read *right-to-left*.

In Figure 18 (c), the cycle C_7 follows C_6 because it has one more equal length cycle in its MCB representation. The cycle C_6 follows C_4 since $[0, 1, 0, 1, 0, 0]$ follows $[0, 1, 0, 0, 0, 0]$ in lexicographic ordering. We denote this sorted matrix by $\mathbf{R}_3$.

sli classes. Two cycles belong to the same **sli** class if their MCB vectors are equal up to shorter cycles by Lemma 12. By using reverse lexicographic ordering, cycles in a **sli** class will appear as consecutive rows of $\mathbf{R}_3$. See the cycles C_4 and C_6 in Figure 18 (c). We merge cycles in a **sli** class into a single row, and we keep track of the relevant V' -families they represent. We denote this final matrix as $\mathbf{R}$.

Time-Complexity. Most steps of this procedure can be done using a single pass through the working matrix $\mathbf{R}_i$. This has time-complexity $O(\text{nnz}(\mathbf{R}_i)) = O(\text{nnz}(\mathbf{R}_1))$ where $\text{nnz}(\mathbf{X})$ is the number of nonzero entries in $\mathbf{X}$. The largest exception to this is sorting the rows of each **pi** class using reverse lexicographic ordering. Using dense operations, this has $O(\nu^3 \log \nu)$ time-complexity, where $\mathbf{R}_2$ has $O(\nu^2)$ rows and comparing two rows takes $O(\nu)$ time. Using sparse operations, this has $O(\text{nnz}(\mathbf{R}_1) \log \nu)$ time-complexity, where the time to compare rows reduces to the average number of nonzero entries per row. For our MD simulations, this procedure is an order of magnitude faster than computing the relevant cycle families and an MCB.

Output. The output relevant cycle matrix $\mathbf{R}$ can be interpreted as follows. Each row of this matrix is a representative cycle for a **sli** class as an MCB vector. The choice of representative cycle only changes these MCB vectors by shorter cycles. Moreover, the entries in the diagonal blocks of $\mathbf{R}$ do not depend on the choice of representative cycles.

The **pi** classes form diagonal blocks in $\mathbf{R}$. The rows in these blocks are sorted so that the cycles belonging to the MCB appear first. Because of this, the format of each block is the identity matrix followed by additional rows $[I, X']'$. The polyhedra are obtained via these additional rows X . The number of polyhedra obtained is equal to the dimension of the (left) null space of $\mathbf{R}$, given by the number of rows minus the number of columns in $\mathbf{R}$.

For instance, from the row corresponding to C_7 in Figure 18 (c), we obtain the polyhedron

$$C_7 = C_3 \oplus C_4 \oplus (\text{smaller cycles}). \quad (56)$$

Here, $\{C_3, C_4, C_7\}$ are the large faces of the triangular prism. The smaller cycles used to construct the polyhedron depend on which representative cycles are sampled from each family. For the cycles given, the full polyhedron is $\{C_0, C_1, C_3, C_4, C_7\}$. Codes for computing the decomposition of this graph are given in the notebook `cycle_decomposition.ipynb` on Github [18].

If a **pi** class has multiple polyhedra, then the polyhedra sampled depend on the MCB used to represent the columns of $\mathbf{R}$. For instance, for the graph shown in Figure 12, this procedure may identify either two cubes or a cube and a large rectangular prism. Intuitively, the most desirable polyhedra are the smallest ones, i.e., the cubes in Figure 12. We leave the challenge of computing smallest polyhedra for future work.

5 Theory and algorithms for sampling minimum cycle bases

5.1 Uniform random sampling of minimum cycle bases

Here, we introduce a method for sampling MCBs uniformly at random using a Markov Chain Monte Carlo approach. We assume the reader is familiar with the basic theory of Markov chains [48].

5.1.1 The basic algorithm

A given MCB $\mathcal{M}$ is randomly modified as follows.

1. Randomly select a cycle $C' \in \mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}$ to exchange with a cycle in $\mathcal{M}$.
2. Construct a polyhedron using C' and $\mathcal{M}$, i.e., $\mathcal{P} = \{C'\} \cup \mathcal{E}_{\mathcal{M}}(C')$.

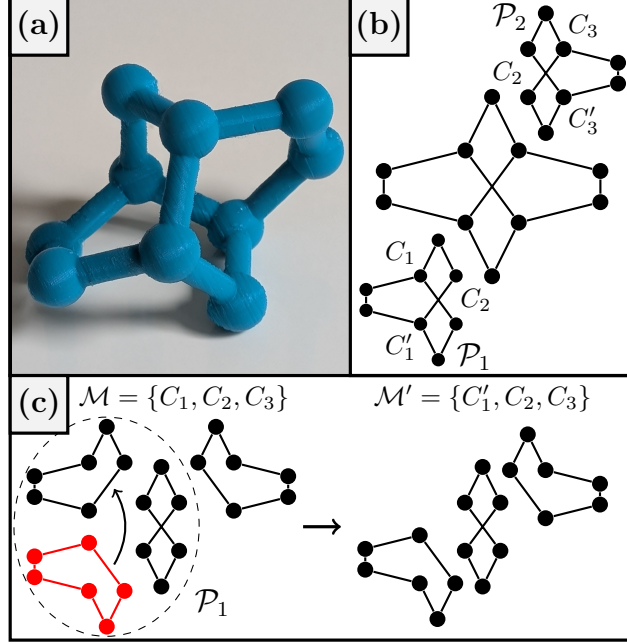


Figure 19: (a) 3D print of the carbon skeleton of twistane, a molecule with formula $C_{10}H_{16}$. (b) The carbon skeleton of twistane as a graph and its smallest polyhedra $\mathcal{P}_1 = \{C_2, C_3, C'_3\}$ and $\mathcal{P}_2 = \{C_1, C'_1, C_2\}$. (c) Transition from the MCB $\mathcal{M} = \{C_1, C_2, C_3\}$ to $\mathcal{M}' = \{C'_1, C_2, C_3\}$ by swapping C'_1 for C_1 using the polyhedron $\mathcal{P}_1$.

3. Randomly select a cycle $C \in \mathcal{P}$, $|C| = |C'|$, and return $\mathcal{M}' = (\mathcal{M} \setminus \{C\}) \cup \{C'\}$. If $C = C'$, return the same basis $\mathcal{M} = \mathcal{M}'$.

See Figure 19 (c) for an example of this procedure in a graph with multiple polyhedra. These steps are repeated towards equilibrium.

To validate this approach, we make the following observation.

Proposition 1. *The above Markov chain is ergodic.*

Proof. First, we need the above procedure to be aperiodic. This is true because there is nonzero holding probability, i.e., the probability we return to the same state $P(\mathcal{M}' = \mathcal{M})$. These holding probabilities cause the MCB to leave its current state at a random (geometric) time, which breaks any otherwise periodic behavior in the system.

Second, the system must be irreducible, i.e., any MCB must be reachable by any other. Consider two distinct MCBs $\mathcal{M}_1$ and $\mathcal{M}_2$. By Lemma 9, we can exchange a cycle in $\mathcal{M}_1$ for another in $\mathcal{M}_2$ to obtain a new MCB $\mathcal{M}_1 \rightarrow \mathcal{M}'_1$ that shares an additional cycle with $\mathcal{M}_2$. We may repeat this procedure to obtain $\mathcal{M}_2$ from $\mathcal{M}_1$. $\square$

By ergodicity, this Markov chain approaches a stationary distribution over the set of MCBs. This stationary distribution is the uniform distribution because the Markov chain has symmetric transition probabilities

$$P(\mathcal{M} \rightarrow \mathcal{M}') = P(\mathcal{M}' \rightarrow \mathcal{M}) = \frac{1}{|\mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}|} \frac{1}{|\mathcal{P}|}. \quad (57)$$

The first term $1/|\mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}| = 1/(|\mathcal{C}_{\mathcal{R}}| - \nu)$ is the constant probability with which we choose a given cycle $C' \in \mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}$ in step 1 of the procedure above. The second term is the probability we select

Algorithm 4: Random MCB

Input: $\mathcal{M}$ = input MCB $\mathbf{R}$ = representative cycles for
sli classes as vectors in $\mathcal{M}$ **Output:** $\mathcal{M}_r$ = random MCB

```

1 for  $i = 1, \dots, N_{\text{steps}}$  do
2   Randomly sample  $C' \notin \mathcal{M}$  from  $\mathbf{R}$ ;
3   Using the row for  $C'$  in  $\mathbf{R}$  construct the polyhedron  $\mathcal{P} = \{C'\} \cup \mathcal{E}_{\mathcal{M}}(C')$ ;
4   Sample  $C \in \mathcal{P}, |C| = |C'|$ , with probability proportional to  $1/|\mathcal{S}(C)|$ ;
5   if  $C \neq C'$  then
6      $\mathcal{M} \leftarrow (\mathcal{M} \setminus \{C\}) \cup \{C'\}$ ;
7     # Update  $\mathbf{R}$  to new basis;
8     for Row cycle  $C_i = \mathbf{R}[i, :]$  do
9       if  $C \in \mathcal{E}_{\mathcal{M}}(C_i)$  then
10        Set  $\mathbf{R}[i, :]$  to  $\mathcal{E}_{\mathcal{M}'}(C_i) = (\mathcal{E}_{\mathcal{M}}(C_i) \setminus \{C\}) \triangle \mathcal{E}_{\mathcal{M}'}(C)$ ;
11      end
12 end
13  $\mathcal{M}_r \leftarrow []$  # empty list;
14 for  $C \in \mathcal{M}$  do
15   Randomly sample a cycle from the sli class  $\mathcal{S}(C)$  and add it to  $\mathcal{M}_r$ ;
16 end

```

the cycle $C \in \mathcal{P}$ to exchange for C' . The same polyhedron $\mathcal{P}$ is used in both directions $\mathcal{M} \rightarrow \mathcal{M}'$ and $\mathcal{M}' \rightarrow \mathcal{M}$.

5.1.2 The efficient algorithm

We simplify this procedure by drawing cycles from the matrix $\mathbf{R}$ computed in Section 4.4. Each cycle in $\mathbf{R}$ represents a sli class. The MCB we will sample from $\mathbf{R}$ is used to select a set of sli classes to draw cycles from. It is valuable to sample cycles from $\mathbf{R}$ since there is only a polynomial number of sli classes. Cycles can be efficiently sampled from a sli class directly, which we discuss at the end of this subsection.

By Corollary 13, we can exchange pairs of cycles within a sli class arbitrarily between MCBs. Thus, the total number of MCBs we can obtain using valid sli classes $\mathcal{S}_1, \dots, \mathcal{S}_\nu$ is $|\mathcal{S}_1| \cdots |\mathcal{S}_\nu|$. We want the stationary distribution to be proportional to the number of MCBs that $\mathcal{M}$ drawn from $\mathbf{R}$ represents,

$$\pi_{\mathcal{M}} \propto \prod_{C \in \mathcal{M}} |\mathcal{S}(C)| \quad (58)$$

where $\mathcal{S}(C)$ is the sli class containing C .

To achieve the stationary distribution given by Eq. (58), we bias the transition probabilities to satisfy the detailed balance condition

$$P(\mathcal{M} \rightarrow \mathcal{M}')\pi_{\mathcal{M}} = P(\mathcal{M}' \rightarrow \mathcal{M})\pi_{\mathcal{M}'}. \quad (59)$$

We re-arrange Eq. (59) to obtain the desired ratio between the forward and backward transition

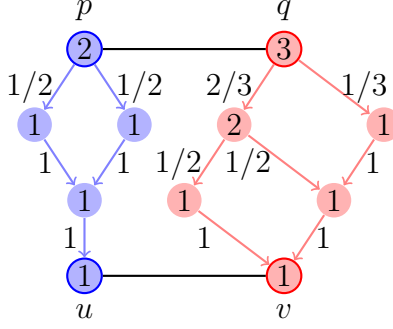


Figure 20: Procedure used to randomly sample a cycle from an even cycle family. A cycle is sampled by joining two random paths $p \rightarrow u$ and $q \rightarrow v$ to the descriptors (p, q) and (u, v) . Node labels represent the number of valid shortest paths from a node to (u, v) . Edge labels represent the probability a node is selected to construct a shortest path.

probabilities

$$\frac{P(\mathcal{M} \rightarrow \mathcal{M}')}{P(\mathcal{M}' \rightarrow \mathcal{M})} = \frac{\pi_{\mathcal{M}'}}{\pi_{\mathcal{M}}} = \frac{|\mathcal{S}(C')|}{|\mathcal{S}(C)|} \quad (60)$$

where $\mathcal{M}' = (\mathcal{M} \setminus \{C\}) \cup \{C'\}$. This is achieved by removing the cycle C from our polyhedron with probability proportional to $1/|\mathcal{S}(C)|$, as in Algorithm 4, line 4. Then, the transition probabilities are given by

$$\begin{aligned} P(\mathcal{M} \rightarrow \mathcal{M}') \\ = \frac{1}{\dim(\text{coker}(\mathbf{R}))} \frac{1/|\mathcal{S}(C)|}{\sum_{C' \in \mathcal{P}} 1/|\mathcal{S}(C')|} \end{aligned} \quad (61)$$

where $\dim(\text{coker}(\mathbf{R}))$ is the dimension of the left null space of $\mathbf{R}$, also the number of row cycles in $\mathbf{R}$ that do not belong to $\mathcal{M}$. Only the numerator $1/|\mathcal{S}(C)|$ changes in Eq. (61) between the forward $(\mathcal{M} \rightarrow \mathcal{M}')$ and reverse $(\mathcal{M}' \rightarrow \mathcal{M})$ exchange. The same polyhedron $\mathcal{P}$ is used in both directions.

After performing the exchange $\mathcal{M} \rightarrow \mathcal{M}'$ we update the rows of $\mathbf{R}$ to be vectors in $\mathcal{M}'$. We only need to update the representation of a cycle $C_i = \mathbf{R}[i, :]$ if its expansion $\mathcal{E}_{\mathcal{M}}(C_i)$ includes the cycle C swapped for C' . If $C \in \mathcal{E}_{\mathcal{M}}(C_i)$, we replace C with its expansion in $\mathcal{M}'$

$$\mathcal{E}_{\mathcal{M}'}(C_i) = (\mathcal{E}_{\mathcal{M}}(C_i) \setminus \{C\}) \triangle \mathcal{E}_{\mathcal{M}'}(C). \quad (62)$$

Last, we describe how cycles are sampled from each **sli** class. A **sli** class is the union of disjoint **V'**-families output by Algorithm 1. We choose a **V'**-family with probability proportional to the number of cycles it has. A cycle is sampled from the **V'**-family by backtracking from either the node p or edge (p, q) to the root edge (u, v) . At each step, the probability we step through a given node is proportional to the number of paths it has – see Figure 20. The final MCB is obtained by sampling a random cycle from each **sli** class. The whole procedure is summarized in Algorithm 4.

5.2 Pairwise intersections in a minimum cycle basis

The overarching goal of this manuscript is to develop mathematical theory that is capable of describing ring clusters in hydrocarbon pyrolysis. So far, we have discussed how to obtain the relevant cycles and how to partition these cycles into a polynomial number of **pi** and **sli** classes.

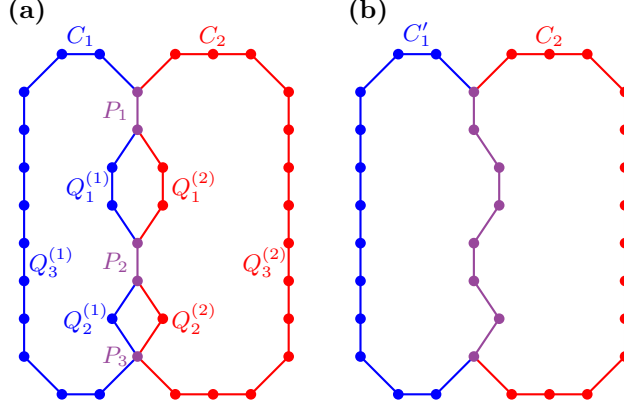


Figure 22: **(a)** Two cycles C_1, C_2 that intersect over three paths. Blue nodes belong to C_1 , red nodes to C_2 , and purple nodes to both. If we traverse C_1 clockwise and C_2 counterclockwise, then the intersection paths appear in the same order $P_1 \rightarrow P_2 \rightarrow P_3$ and with the same orientation. The shorter separating paths have equal lengths $|Q_1^{(1)}| = |Q_1^{(2)}|$ and $|Q_2^{(1)}| = |Q_2^{(2)}|$. **(b)** A cycle C'_1 that can be exchanged for C_1 such that C'_1 and C_2 intersect over a single path. C'_1 is constructed from C_1 by swapping the paths $Q_1^{(1)}$ for $Q_1^{(2)}$ and $Q_2^{(1)}$ for $Q_2^{(2)}$.

In summary, Theorem 16 states that a pair of cycles belonging to some MCB may intersect over a long path broken up by shorter cycles. See Figure 22 (a). In general, we may define the intersection of two cycles using a series of paths, which can be stored as a list of lists. We emphasize, that Theorem 16 does not apply to arbitrary pairs of cycles. See Figure 23.

Our theory allows us to go further. Specifically, we can stitch together a pair of cycles in an MCB so that they intersect over a single path.

Theorem 17. *Let $\mathcal{M}$ be an MCB and let cycles $C_1, C_2 \in \mathcal{M}$ share at least two nodes. There exists a cycle C'_1 such that*

1. $\mathcal{M}' = (\mathcal{M} \setminus \{C_1\}) \cup \{C'_1\}$ is an MCB and,
2. C'_1 and C_2 intersect over a single path.

The cycle C'_1 is obtained by exchanging paths in C_1 for C_2 . See Figure 22 (b) for an example of this procedure.

It is desirable that C_1 and C_2 intersect over a single path because a path can be stored as a single list. We propose to postprocess an MCB so that all pairs of cycles share at most one path, as in Algorithm 5. Then, the cluster of cycles may be characterized using a set of cycles and the paths that connect them. This is the dual graph representation shown in Figure 21. We remark that the dual graph as shown in Figure 21 is not lossless because it only includes path lengths, not the positions of these paths in the cycles.

At this point, we cannot guarantee that Algorithm 5 always outputs an MCB with every pair of cycles intersecting at most over a single path. It always succeeds on our hydrocarbon data in 10-20 iterations. We leave the investigation of this issue for future work.

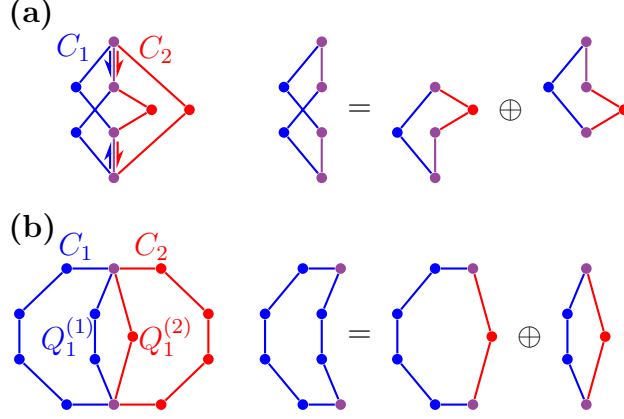


Figure 23: Two pairs of cycles that do not satisfy the format in Theorem 16. In both cases, the left cycle C_1 is not relevant because it is the sum of shorter cycles. **(a)** Two cycles C_1, C_2 that intersect over two paths. The red and blue arrows depict an example pair of orientations of C_1 and C_2 . There are no orientations of C_1 and C_2 where the intersection paths have the same orientation in C_1 and C_2 . Hence, C_1 appears twisted next to C_2 . **(b)** Two cycles C_1, C_2 that are separated by shorter paths $Q_1^{(1)}$ and $Q_1^{(2)}$ with different lengths $|Q_1^{(1)}| \neq |Q_1^{(2)}|$.

Algorithm 5: Postprocess MCB

Input: MCB $\mathcal{M}$

Output: MCB $\mathcal{M}'$, where all pairs $C_1, C_2 \in \mathcal{M}'$ intersect over a single path

- 1 Compute the intersection paths for all pairs of cycles $C_i, C_j \in \mathcal{M}$;
 - 2 **for** $i = 1, \dots, N_{\max}$ **do**
 - 3 Randomly sample a pair $C_i, C_j \in \mathcal{M}$ that intersect over multiple paths;
 - 4 Exchange $\mathcal{M} \leftarrow (\mathcal{M} \setminus \{C_i\}) \cup \{C'_i\}$ by Theorem 17 where C'_i, C_j intersect over a single path;
 - 5 Recompute intersection paths between C'_i and other cycles;
 - 6 **end**
-

6 Applications to hydrocarbon pyrolysis systems

6.1 Generating data

Here, we apply the theory for the **pi** and **sli** relations to analyze loop clusters in hydrocarbon pyrolysis. This data comes from a single simulation using a ReaxFF force field [41] in LAMMPS [36, 37, 49]. The simulation was initialized with 800 adamantane molecules, $C_{10}H_{16}$, resulting in $N_C = 8000$ carbon and $N_H = 12800$ hydrogen atoms. The system was ramped to a temperature of 4000K and pressure of 40.5 GPa. This was followed by an NPT simulation for a total of 1.2 ns. Bonds were sampled using a length and duration criterion as discussed in previous works [18, 10]. The accuracy of the ReaxFF force field [41] at these thermodynamic conditions is questionable, but it provides sufficiently rich data to investigate our ring statistics.

We started sampling after 0.6 ns to allow the system to reach equilibrium. Every 1.2 ps (500 samples total) we decompose the loop cluster into **pi** and **sli** classes as discussed in Section 4.4. From this decomposition we collect the following:

- A random MCB.

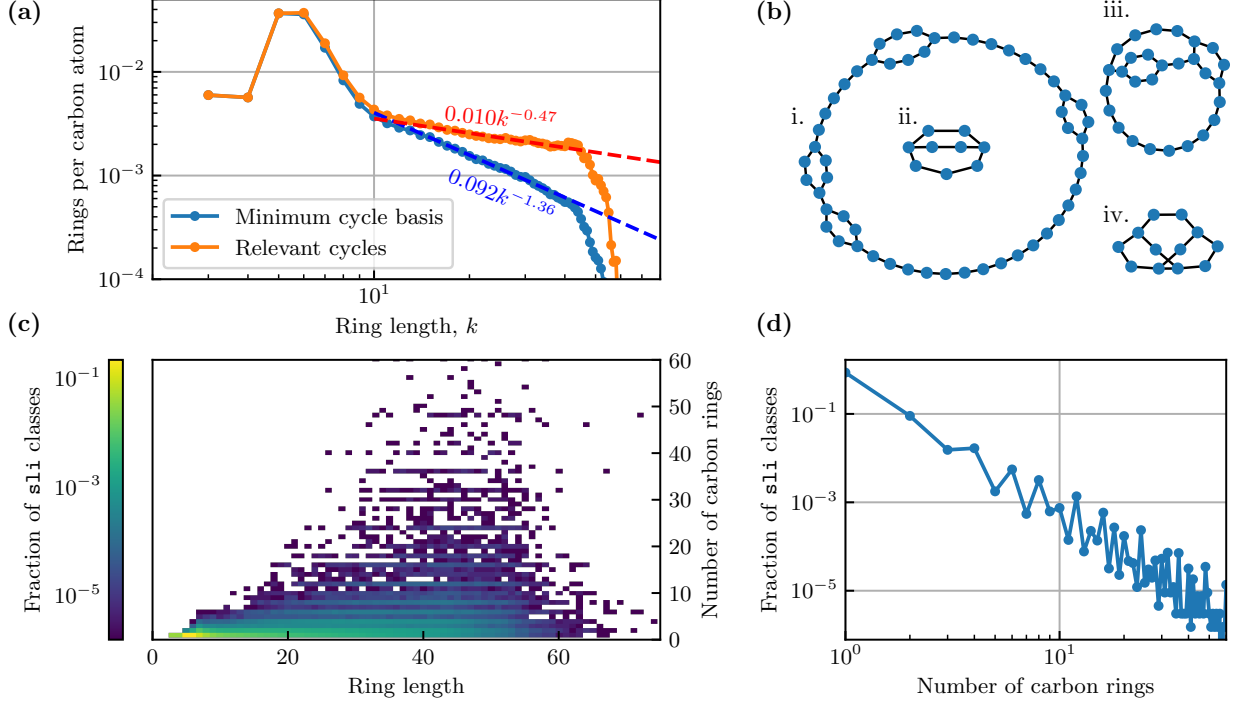


Figure 24: Analysis of cycles for a system initialized to adamantane with formula $C_{10}H_{16}$. **(a)** Rate distribution of carbon rings obtained via an MCB (blue) and the relevant cycles (orange). The dashed lines represent a power law fit of these distributions for rings of length $15 \leq k \leq 40$. **(b)** Four examples of **sli** classes: (i) 16 rings of length 45, (ii) 2 rings of length 7, (iii) 4 rings of length 19, and (iv) 3 rings of length 8. **(c)** Joint distribution of ring lengths and number of rings in **sli** classes. **(d)** Distribution of the number of rings in **sli** class.

- The length and number of rings within each **sli** class.
- The length, number of rings, and number of independent polyhedra within each **pi** class.

6.2 Relevant cycles versus minimum cycle basis

6.2.1 $C_{10}H_{16}$ simulation data

In Section 5.1, we introduced Algorithm 4 for sampling MCBs uniformly at random. To highlight the value of this approach, we compare counts of large carbon rings from random MCBs and from the relevant cycles. Specifically, we focus on the ring rate distribution, or the number of rings per carbon atom as a distribution over ring length – see Figure 24 (a). This statistic has also been considered in computational studies of amorphous carbon [3, 4].

Most rings are short with five or six carbon atoms. The amorphous structure of these hydrocarbon networks results in many large rings, as shown by the heavy tail of the ring rate distribution. This tail approximately fits to a power law distribution $Ck^{-\alpha}$ with a cutoff.

Importantly, the power law constant α varies with the method of sampling loops. When the loops are sampled from an MCB we obtain $\alpha \approx 1.36$. In contrast, sampling loops from the relevant cycles results in a power of $\alpha \approx 0.47$.

Without the cutoff, the power law tail $k^{-0.47}$ of the loop rate distribution for the relevant cycles is not summable. We observe this cutoff increases with respect to both system size and the carbon-

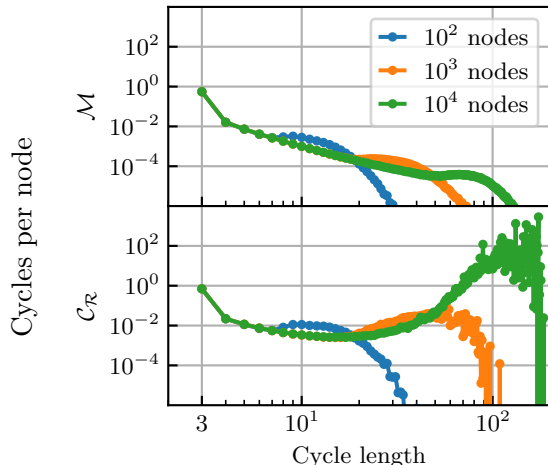


Figure 25: Analysis of the cycle rate distribution in random geometric graphs as a function of graph size. Cycles are sampled using an MCB $\mathcal{M}$ (top) and the relevant cycles $\mathcal{C}_{\mathcal{R}}$ (bottom). Nodes are assigned random positions in the unit cube $[0, 1]^3$ and the cutoff radius for edges is chosen so that the expected degree is 3. The random geometric graph with 10^2 , 10^3 , and 10^4 nodes is sampled 10^5 , 10^4 and 10^3 times respectively.

to-hydrogen ratio. Larger simulations are needed to see if this cutoff continues to increase with system size. If so, then the power law tail $k^{-0.47}$ will cause the relevant cycles to grow asymptotically faster than the dimension of the cycle space ν :

$$\frac{|\mathcal{C}_{\mathcal{R}}|}{\nu} \rightarrow \infty \quad \text{as} \quad N_C \rightarrow \infty. \quad (67)$$

For our simulation, the time-averaged number of relevant cycles is $|\mathcal{C}_{\mathcal{R}}| = 1736$ and the average dimension of the cycle space is $\nu = 1311$.

6.2.2 Random geometric graphs

To further investigate the scaling of $|\mathcal{C}_{\mathcal{R}}|$ with system size, we consider random geometric graphs [50]. A random geometric graph is a graph with n nodes randomly positioned in the d -dimensional cube $[0, 1]^d$. A pair of nodes are joined by an edge if their distance is less than a given fixed radius. The cycle statistics from these random graphs share some qualitative characteristics with our simulation because they have a spatial component. We choose a dimension of $d = 3$ and periodic boundary conditions for sampling edges.

We choose the radius so that the average degree is 3. For comparison, the average degree of the carbon skeleton in our ReaxFF simulation is 2.25. We choose a higher mean degree of 3 because it results in more large loops. This reduces the number of samples needed to observe enough large loops.

As in molecular dynamics simulation, the tail of the cycle rate distribution of MCBs in random geometric graphs roughly follows a power law. The cutoff for this power law distribution increases with system size. See Figure 25, top. For the random geometric graph with 10^2 nodes, the cycle rate distribution for the relevant cycles has the same shape as the cycle rate distribution given by MCBs. See the blue curve in Figure 25, bottom. However, as system size increases, the number of large relevant cycles increases much faster than the number of large cycles in MCBs. For the

random geometric graphs with 10^4 nodes, we observed a shocking 9900 relevant cycles per node. See the green curve in Figure 25, bottom.

6.2.3 A potential explosion of the relevant cycle count

It is unclear if the ring rate distribution in molecular dynamics will follow this same trend with system size. Given this ambiguity, we argue that the relevant cycles should not be used to count rings in large amorphous chemical systems. Our studies of random geometric graphs suggest the number of large rings may explode at scales achievable using multiple GPUs in a molecular dynamics simulation. Beyond the relevant cycles, many other unique sets of rings in the chemistry literature are also worst case exponential in size. See, e.g., the review [28]. This is not an issue for our approach, where the size of an MCB grows linearly with the number of edges.

6.3 Sources of additional cycles

The `pi` and `sli` classes can be used to determine where the extra relevant cycles are coming from. For the given initial conditions, we find that the `sli` classes are the most common. Over 99% of `pi` classes reduce to a single `sli` class. Most `sli` classes consist of a large ring with nested smaller rings – see Figure 24 (b.i.). In some cases, there is only a minor distinction between large and small rings, e.g., the two larger 7-rings in Figure 24 (b.ii.) are separated by a smaller 6-ring. In other cases, the structure of a `sli` class is more complex – see Figure 24 (b.iii-iv.).

In Figure 24 (c), we show the joint distribution of the lengths and number of rings within a `sli` class. From this distribution, we observe most (roughly 85%) of the `sli` classes consist of a single cycle, and the average size of the `sli` classes increases with ring length. In Figure 24 (d), we show the size distribution of the `sli` classes, which roughly follows a power law distribution. Interestingly, the oscillations in this size distribution are not noise, and they can be seen as horizontal bands in the joint length size distribution in Figure 24 (c). The number of cycles in a `sli` class tends more often to a composite than a prime number.

In summary, the relevant cycles have an increased number of large cycles forming `sli` classes. At the same time, these large cycles are highly dependent, e.g., the cycles in the `sli` classes in Figure 24 (b.i-iii.) share many edges. Our codes can be used to isolate and visualize these `sli` classes.

6.4 Polyhedra and diamond

Last, we investigate the `pi` classes that do not reduce to `sli` classes. As discussed in Section 3.1.2, these `pi` classes can be described as polyhedra clusters. These are relatively rare, there are roughly eight of these `pi` classes per time step. Since more theory is needed to describe these structures, we will consider individual polyhedra that are output by these codes in a case-by-case basis. We label the `pi` classes by the small hydrocarbons with the same carbon skeleton. Note, the `pi` classes are not molecules but substructures within the loop cluster.

The majority of the `pi` classes are composed of highly organized short rings. Roughly 75% of these are composed of rings of length 6. See Figure 26 (b-d) for some examples of these `pi` classes. Most of these `pi` classes are small, consisting of a single polyhedron. See Figure 26 (a). Sometimes, the faces of a polyhedron are not uniquely defined. For instance, the observed polyhedron in Figure 26 (d) has two hexagonal cycles that can be chosen for its right face. However, polyhedra with non-unique faces are more rare.

The most obvious polyhedron to investigate is adamantane. Adamantane is the smallest diamondoid, a nanoscale structure with the same network topology as diamond, and it has been ob-

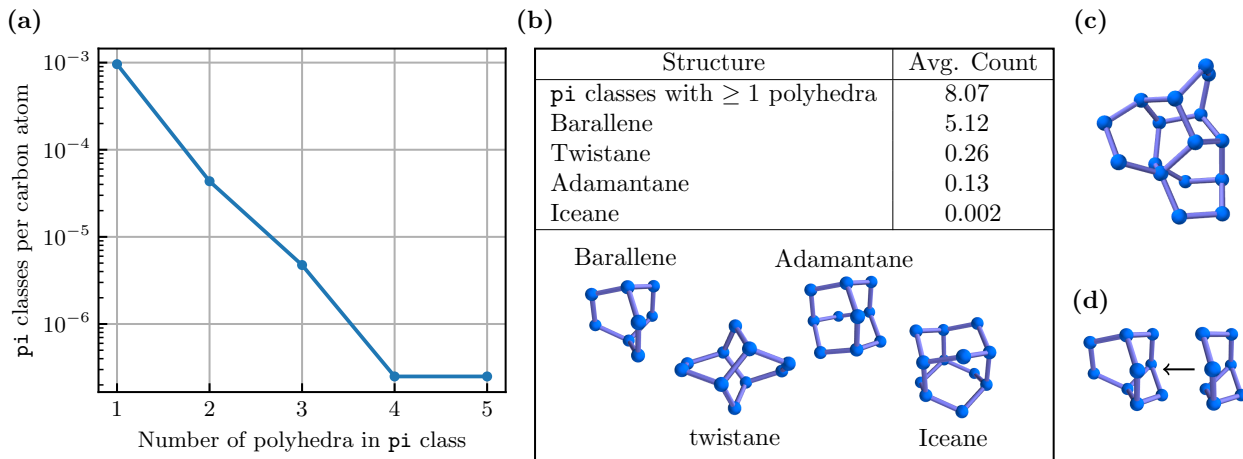


Figure 26: Analysis of polyhedra for a system initialized to adamantane with formula $C_{10}H_{16}$. (a) Size distribution of the π classes given by the number of polyhedra in the class. Only π classes with at least one polyhedron are considered, otherwise, they reduce to sli classes. (b) Average counts of common π classes per time step and their visualizations. (c) The largest observed π class composed of five barallene polyhedra glued together by faces. (d) A π class composed of a polyhedron that has a hexagonal face that is not uniquely defined because it is a sli class.

served in petroleum [51]. Even though our simulation was initialized with adamantane molecules, adamantane was not common in our simulation. See Figure 26 (b).

Instead, the most commonly observed polyhedron is barallene*. Barallene can be used to form hexagonal diamond – also called lonsdaleite – by gluing its faces to iceane. Iceane is relatively rare in our simulations, so it seems unlikely that this crystal structure will form. Instead, the majority of the larger π classes are composed of multiple barallene polyhedra glued together by faces. See twistane in Figure 26 (b) and the larger structure in Figure 26 (c).

7 Discussion

7.1 Sampling cycles (ring perception)

In Section 5.1, we introduced a method for sampling MCBs uniformly at random, Algorithm 4. This differs from other approaches because we do not sample a deterministic set of rings from our chemical data. Regardless, as a probabilistic structure, the random MCB is well-defined, and can be used to sample summary statistics from a chemical network. This approach results in fewer cycles, as discussed in Section 6. In particular, the size of the random MCB grows linearly with the number of bonds in the system.

A potential critique of using a random MCB is that it may not uncover all useful cycles. Unfortunately, it is challenging to define such a set of cycles that is always small in size. The set of relevant cycles can be exponential in number, and it only includes cycles from MCBs. However, all cycles are within the span of a MCB by definition. Thus, we argue it is better to interpret additional cycles as a consequence of the overlapping structure of a cycle cluster.

*It is likely that many of these polyhedra more closely match the molecule bicyclo[2,2,2]octane with the same topology as barallene without double bonds. However, we do not store bond order or infer double bonds. Hence, we use the simpler name barallene in both cases.

In Theorem 16 in Section 5.2, we fully characterized how a pair of cycles in a given MCB may intersect as a series of paths that appear with the same order and orientation in both cycles. This gives insight on what data structures are needed to properly store these pairwise intersections of cycles. Furthermore, in Algorithm 5 we introduced a method of modifying an MCB so that all pairs of cycles intersect over a single path. This permits a dual graph representation of the loop cluster, where nodes represent cycles in an MCB and edges represent their intersection paths. These intersection paths are easy to store as a list of nodes.

Theorem 16 is specialized to pairs of cycles in an MCB. Its proof largely relies on the fact that these cycles preserve distances. Because of this, Theorem 16 can be easily adapted to the full set of relevant cycles. The largest set of cycles this proof technique applies to is the set of shortest path rings from Ref. [52].

7.2 Polyhedra.

In Section 3.1.2, we related lack of uniqueness of MCBs to polyhedral structures. These generalized polyhedra were defined as sets of non-flat faces obtained via a cycle and its expansion with respect to an MCB

$$\mathcal{P} = \{C\} \cup \mathcal{E}_{\mathcal{M}}(C), \quad C \in \mathcal{C}_{\mathcal{R}} \setminus \mathcal{M}.$$

The **pi** classes were interpreted as clusters of these polyhedra.

A potential application of these polyhedra is in crystallography. A crystal can be represented by a periodic tiling of these polyhedra glued together by faces, e.g., diamond is a tetrahedral tiling of adamantane. This representation has a benefit when compared to the periodic unit cell representation of a crystal; the **pi** relation is defined for general graphs. There is no requirement that the chemical network or its substructures are large. Small **pi** classes might be interpreted as nanocrystals. Also, the input graph does not need to be periodic. In an inhomogeneous crystal (e.g., a mixture of cubic and hexagonal diamond) it may be useful to analyze how these polyhedra vary within the material.

Before investigating these applications, a pair of limitations must be addressed. Algorithmically, the polyhedra are left-kernel vectors of the diagonal blocks of the matrix **R** from Section 4.4. These correspond to **sli** classes that sum to shorter cycles. This notion of polyhedra is abstract. In some cases, this may be ok, where the **sli** classes can be interpreted as generalized cycles and visualized as the union of their cycles (e.g., Figure 26 (d)). If not, then a cycle may be sampled from each **sli** class to represent a polyhedron as a set of face cycles. Polyhedra sampled using this latter approach are not unique.

The second limitation is that these polyhedra are the basis of a vector space. As with cycles, we are only interested in small polyhedra. Ultimately, one would want to properly define a minimum polyhedron basis, and to find an algorithm to compute it. This task is likely complex. A greedy approach may be to bias Algorithm 4 to obtain a matrix **R** that outputs small polyhedra.

8 Conclusion

We reviewed and advanced theory and algorithms for identifying and grouping cycles in graphs. We proposed a new method for sampling MCBs uniformly at random. The **pi** relation helps to extract generalized polyhedra, that may be start-ups for forming crystalline structures. The **sli** classes may comprise large numbers of relevant cycles as those around the perimeter carbon nanotubes. We propose to include only one representative of each **sli** class in an MCB. We also provide theory

for the format in which pairs of cycles in an MCB intersect, and we introduce a postprocessing step that makes these intersections single paths. This enables a dual graph representation for a cluster of cycles. We propose using this dual graph approach to model complex material networks. We illustrate this theory and algorithms by applying them to a reactive simulation of hydrocarbon pyrolysis MD simulation data initialized with 800 adamantane, $\text{C}_{10}\text{H}_{16}$, molecules.

Acknowledgments

This work was partially supported by ONR MURI Grant No. N000142412547 (M.C.), AFOSR MURI Grant No. FA9550-20-1-0397 (M.C.), and NSF GRFP Grant No. DGE 2236417 (P.R.).

A Proofs for Section 3

A.1 Mutual relevance and useful lemmas

To start, Lemma 18 below generalizes Statement 1 of Lemma 4 to sets of cycles. This lemma is useful for many of the results in this appendix.

Lemma 18. *Let $\mathcal{Y}$ be a set of cycles bounded in length by $k \in \mathbb{N}$, $|C| \leq k$ for $C \in \mathcal{Y}$. Let $\mathcal{M}$ be a given MCB. There exists a subset $\mathcal{Y}' \subseteq \mathcal{M}$ also bounded in length by k , $|C'| \leq k$ for $C' \in \mathcal{Y}'$, such that*

$$\bigoplus_{C' \in \mathcal{Y}'} C' = \bigoplus_{C \in \mathcal{Y}} C. \quad (68)$$

Proof. Let $\mathcal{Y}$ be a set of cycles bounded in length by k and $\mathcal{M}$ an MCB. By Definition 6 of the expansion operator, the set of cycles

$$\mathcal{Y}' = \mathcal{E}_{\mathcal{M}} \left(\bigoplus_{C \in \mathcal{Y}} C \right) \quad (69)$$

is the unique subset $\mathcal{Y}' \subseteq \mathcal{M}$ that satisfies Eq. (68). We want to show the cycles $\mathcal{Y}'$ are also bounded in length by k . We observe, the cycles $\mathcal{Y}'$ satisfy the inclusion

$$\mathcal{Y}' \subseteq \bigcup_{C \in \mathcal{Y}} \mathcal{E}_{\mathcal{M}}(C). \quad (70)$$

Thus, the cycles $\mathcal{Y}'$ are bounded in length by

$$|C'| \leq |C| \leq k, \quad C \in \mathcal{Y}, C' \in \mathcal{E}_{\mathcal{M}}(C) \quad (71)$$

where the first inequality is Statement 1 of Lemma 4. $\square$

Next, we introduce mutual relevance, which generalizes the notion of relevance to sets of cycles. We will use mutual relevance in our proofs regarding the pi relation.

Definition 12. *A set of cycles $\mathcal{X}$ is mutually relevant if there exists an MCB $\mathcal{M}$ such that $\mathcal{X} \subseteq \mathcal{M}$.*

We emphasize linearly independent, relevant cycles are not necessarily mutually relevant. See, e.g., the cycles C_L, C'_L, C_R in Figure 7 in Section 2.4.

Recall, vectors $v_1, \dots, v_k$ from a vector space are linearly independent if the only linear combination that achieves

$$c_1 v_1 + \dots + c_k v_k = 0$$

is when $c_1 = \dots = c_k = 0$. In Lemma 19 below, we define mutual relevance in a way that resembles this notion of linear independence.

Lemma 19. *A set of cycles $\mathcal{X} \subseteq \mathcal{C}_R$ of equal length are not mutually relevant if and only if there exist cycles $C_1, C_2, \dots, C_k \in \mathcal{X}$ such that*

$$C_1 \oplus C_2 \oplus \dots \oplus C_k = (\text{smaller cycles}) \quad (72)$$

where (smaller cycles) is the sum of a set of smaller cycles $\mathcal{Y}$, $|C| < |C_1|$ for $C \in \mathcal{Y}$.

Proof. ($\Rightarrow$) Let $\mathcal{X} \subseteq \mathcal{C}_R$ not be mutually relevant. The individual cycles in $\mathcal{X}$ are relevant. Hence, there exists a mutually relevant subset $\mathcal{Y} \subset \mathcal{X}$ and a cycle $C \in \mathcal{X} \setminus \mathcal{Y}$ such that $\mathcal{Y} \cup \{C\}$ is not mutually relevant. By the definition of mutual relevance, let $\mathcal{M}$ be an MCB where $\mathcal{Y} \subseteq \mathcal{M}$.

We want to construct a subset of $\mathcal{X}$ that sums to smaller cycles. First, we expand C with respect to $\mathcal{M}$,

$$C = \bigoplus_{C' \in \mathcal{E}_{\mathcal{M}}(C)} C' = \bigoplus_{\substack{C' \in \mathcal{E}_{\mathcal{M}}(C) \\ |C'| = |C|}} C' \oplus \bigoplus_{\substack{C'' \in \mathcal{E}_{\mathcal{M}}(C) \\ |C''| < |C|}} C''. \quad (73)$$

We move the cycles of length $|C| = |C'|$ to the left-hand side, so that the right-hand side contains only smaller cycles:

$$C \oplus \bigoplus_{\substack{C' \in \mathcal{E}_{\mathcal{M}}(C) \\ |C'| = |C|}} C' = \bigoplus_{\substack{C'' \in \mathcal{E}_{\mathcal{M}}(C) \\ |C''| < |C|}} C''. \quad (74)$$

We argue the cycles in the left-hand side of Eq. (74) belong to $\mathcal{Y} \cup \{C\} \subseteq \mathcal{X}$.

Assume towards a contradiction that there exists $C' \in \mathcal{E}_{\mathcal{M}}(C)$ of length $|C'| = |C|$ such that $C' \notin \mathcal{Y}$. By Statement 2 of Lemma 4, we can exchange C for C' to obtain a new minimum cycle basis $\mathcal{M}' = (\mathcal{M} \setminus \{C\}) \cup \{C'\}$. Indeed, this contradicts our initial assumptions, as it implies $\mathcal{Y} \cup \{C\} \subseteq \mathcal{M}'$ is mutually relevant. Thus, the cycles in the left-hand side of Eq. (74) form a subset of $\mathcal{X}$ that sums to smaller cycles as desired.

($\Leftarrow$) Let $C_1, C_2, \dots, C_k$ satisfy Eq. (72). We want to show that these cycles are not mutually relevant. We assume $C_2, \dots, C_k$ are mutually relevant, otherwise our proof is complete. Let $\mathcal{M}$ be an MCB such that $C_2, \dots, C_k \in \mathcal{M}$. By Lemma 18 we can choose smaller cycles $\mathcal{Y}' \subseteq \mathcal{M}$ satisfying Eq. (72). Re-arranging Eq. (72) and substituting in $\mathcal{Y}'$ gives

$$C_1 = C_2 \oplus \dots \oplus C_k \oplus \bigoplus_{C' \in \mathcal{Y}'} C'. \quad (75)$$

Hence, C_1 is in the span of cycles in $\mathcal{M}$, and C_1 does not belong to $\mathcal{M}$. Here, $\mathcal{M}$ is an arbitrary MCB containing $C_2, \dots, C_k$, so there is no MCB containing $C_1, C_2, \dots, C_k$. In other terms, $C_1, C_2, \dots, C_k$ are not mutually relevant. $\square$

A.2 Polyhedron-interchangeability

Here we prove the results for the pi relation from Section 3.1.1. These proofs rely on the lemmas and definitions from Appendix A.1.

First, we prove an equivalent definition of the pi relation. Instead of proving Lemma 7, we prove Lemma 20 below. Lemma 20 has weaker requirements than Lemma 7. In particular, the shorter cycles used to exchange C_1 for C_2 no longer need to belong to an MCB. It is straightforward to prove Lemma 7 from Lemma 20. Lemma 20 is often easier to verify in proofs.

Lemma 20. *Let $C_1, C_2 \in \mathcal{C}_{\mathcal{R}}$ where $|C_1| = |C_2|$. $C_1 \xleftrightarrow{\text{pi}} C_2$ if and only if there exists equal length cycles $\mathcal{X}$, $|C| = |C_1|$ for $C \in \mathcal{X}$ where $C_1, C_2 \notin \mathcal{X}$, and shorter cycles $\mathcal{Y}$, $|C| < |C_1|$ for $C \in \mathcal{Y}$, such that*

1. (mutual relevance) there exists an MCB $\mathcal{M}$ where $\mathcal{X} \cup \{C_2\} \subseteq \mathcal{M}$,
2. (expansion) C_1 is given by the sum

$$C_1 = C_2 \oplus \bigoplus_{C' \in \mathcal{X}} C' \oplus \bigoplus_{C'' \in \mathcal{Y}} C''. \quad (76)$$

The set of equal length cycles in the expansion of C_1 in $\mathcal{M}$ is $\mathcal{X}$.

Proof. ($\Leftarrow$) Let $\mathcal{X}$ and $\mathcal{Y}$ satisfy the given hypothesis. We want to show $C_1 \xleftrightarrow{\text{pi}} C_2$. By mutual relevance, let $\mathcal{M}_2$ be an MCB such that $\mathcal{X} \cup \{C_2\} \subseteq \mathcal{M}$. By Lemma 18, there exists shorter cycles $\mathcal{Y}' \subseteq \mathcal{M}_2$ with the same sum as $\mathcal{Y}$. We replace $\mathcal{Y}'$ for $\mathcal{Y}$ in Eq. (76)

$$C_1 = C_2 \oplus \bigoplus_{C' \in \mathcal{X}} C' \oplus \bigoplus_{C'' \in \mathcal{Y}'} C'', \quad (77)$$

The right-hand side of Eq. (77) is the expansion of C_1 in $\mathcal{M}_2$. By Lemma 4, Statement 2, we may exchange to a new MCB $\mathcal{M}_1 = (\mathcal{M}_2 \setminus \{C_2\}) \cup \{C_1\}$. Thus, $C_1 \xleftrightarrow{\text{pi}} C_2$.

($\Rightarrow$) Let $C_1 \xleftrightarrow{\text{pi}} C_2$ via the MCB $\mathcal{M}_2$. We want to construct sets $\mathcal{X}$ and $\mathcal{Y}$ that satisfy the given hypothesis. First, we show that C_2 belongs to the expansion of C_1 in $\mathcal{M}_2$, or $C_2 \in \mathcal{E}_{\mathcal{M}_2}(C_1)$. To do so, consider the MCB $\mathcal{M}_1 = (\mathcal{M}_2 \setminus \{C_2\}) \cup \{C_1\}$ given by the pi relation. If $C_2 \notin \mathcal{E}_{\mathcal{M}_2}(C_1)$, then the full expansion $\mathcal{E}_{\mathcal{M}_2}(C_1)$ belongs to $\mathcal{M}_1$. This is not possible, since the set $\{C_1\} \cup \mathcal{E}_{\mathcal{M}_2}(C_1)$ is linearly dependent,

$$C_1 = \bigoplus_{C' \in \mathcal{E}_{\mathcal{M}_2}(C_1)} C',$$

so they cannot all belong to $\mathcal{M}_1$. Since C_2 is in the expansion of C_1 in $\mathcal{M}_2$, we may decompose the expansion as follows

$$C_1 = C_2 \oplus \bigoplus_{\substack{C' \in \mathcal{E}_{\mathcal{M}_2}(C_1) \setminus \{C_2\} \\ |C'| = |C_1|}} C' \oplus \bigoplus_{\substack{C'' \in \mathcal{E}_{\mathcal{M}_2}(C_1) \\ |C''| < |C_1|}} C''. \quad (78)$$

Thus, the cycles $\mathcal{X}$ are the cycles of length $|C_1|$ in $\mathcal{E}_{\mathcal{M}_2}(C_1)$ other than C_2 , and the cycles $\mathcal{Y}$ are the shorter cycles in $\mathcal{E}_{\mathcal{M}_2}(C_1)$. $\square$

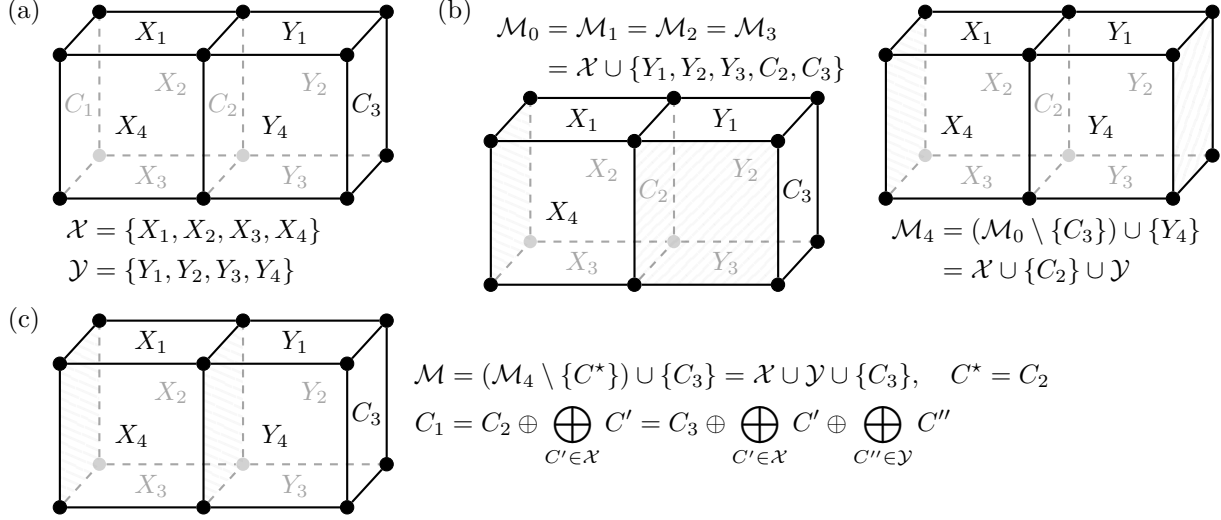


Figure 27: Construction of transitivity for the pi relation. (a) A graph with relevant cycles $\mathcal{C}_{\mathcal{R}} = \mathcal{X} \cup \mathcal{Y} \cup \{C_1, C_2, C_3\}$, where $C_1 \xleftrightarrow{\text{pi}} C_2$ and $C_2 \xleftrightarrow{\text{pi}} C_3$ via $\mathcal{X}$ and $\mathcal{Y}$ respectively. (b) MCBs $\mathcal{M}_j$ such that the mutually relevant sets $\mathcal{Z}_j = \mathcal{X} \cup \{C_2\} \cup \{Y_1, \dots, Y_j\}$ satisfy $\mathcal{Z}_j \subseteq \mathcal{M}_j$. The relevant cycles not in $\mathcal{M}_j$ are shaded with gray stripes. (c) A final MCB $\mathcal{M}$ obtained by exchanging C_3 for C_2 in $\mathcal{M}_4$. C_3 is in the expansion of C_1 in $\mathcal{M}$ so $C_1 \xleftrightarrow{\text{pi}} C_3$.

Next, we prove that pi is an equivalence relation, Theorem 6. First, consider the following property about the expansion operator. Let $\mathcal{B}$ be a cycle basis. If we sum C_1 and C_2 and expand in $\mathcal{B}$ then we obtain

$$C_1 \oplus C_2 = \bigoplus_{C' \in \mathcal{E}_{\mathcal{B}}(C_1)} C' \oplus \bigoplus_{C'' \in \mathcal{E}_{\mathcal{B}}(C_2)} C'' = \bigoplus_{C' \in \mathcal{E}_{\mathcal{B}}(C_1) \triangle \mathcal{E}_{\mathcal{B}}(C_2)} C' \quad (79)$$

where the reduced sum in the final equality is over the symmetric difference $\mathcal{E}_{\mathcal{B}}(C_1) \triangle \mathcal{E}_{\mathcal{B}}(C_2)$ because of cancellations of the form $C \oplus C = \emptyset$. Thus, the expansion operator is linear

$$\mathcal{E}_{\mathcal{B}}(C_1 \oplus C_2) = \mathcal{E}_{\mathcal{B}}(C_1) \triangle \mathcal{E}_{\mathcal{B}}(C_2) \quad (80)$$

where addition is given by the symmetric difference operator. We will use linearity of the expansion operator in the following proofs.

Proof of Theorem 6. We show that the pi relation satisfies three properties. The reflexive property, $C \xleftrightarrow{\text{pi}} C$, is immediate from an MCB $\mathcal{M}, C \in \mathcal{M}$, since $\mathcal{M} = (\mathcal{M} \setminus \{C\}) \cup \{C\}$ is already an MCB.

(Symmetry) Here we use the definition of pi as in Definition 9. Let $C_1 \xleftrightarrow{\text{pi}} C_2$ via $\mathcal{M}_2$, $C_2 \in \mathcal{M}_2$, an MCB such that the exchange $\mathcal{M}_1 = (\mathcal{M}_2 \setminus \{C_2\}) \cup \{C_1\}$ is an MCB. Then, the reverse exchange $\mathcal{M}_2 = (\mathcal{M}_1 \setminus \{C_1\}) \cup \{C_2\}$ also results in a MCB. Hence, $C_2 \xleftrightarrow{\text{pi}} C_1$.

(Transitivity) Transitivity is trickier to verify for the pi relation. As an intuition leading up to this result, recall from Section 3.1.2 that pi relations can be interpreted as polyhedra. For instance, consider the graph depicted in Figure 27 (a). The cycles C_1 and C_2 satisfy $C_1 \xleftrightarrow{\text{pi}} C_2$ since they belong to the same cube $\mathcal{P}_1 = \{C_1, C_2\} \cup \mathcal{X}$. Similarly, $C_2 \xleftrightarrow{\text{pi}} C_3$ since they belong to the cube $\mathcal{P}_2 = \{C_2, C_3\} \cup \mathcal{Y}$. The transitive relation can be verified using the rectangular prism $\mathcal{P}_3 = \mathcal{P}_1 \triangle \mathcal{P}_2 = \{C_1, C_3\} \cup \mathcal{X} \cup \mathcal{Y}$. Unfortunately, the symmetric difference $\mathcal{P}_3$ does not define a pi

relation in general, as the cycles $\mathcal{P}_3 \setminus \{C_1\}$ may not be mutually relevant. Instead, we reproduce this procedure using mutually relevant sets below. We illustrate these methods in Figure 27 as well.

To verify transitivity, we use the definition of the pi relation given by Lemma 20. Let $C_1 \xleftrightarrow{\text{pi}} C_2$ and $C_2 \xleftrightarrow{\text{pi}} C_3$ via $\mathcal{X}$ and $\mathcal{Y}$ up to smaller cycles respectively:

$$C_1 = C_2 \oplus \bigoplus_{C \in \mathcal{X}} C \oplus (\text{smaller cycles}), \quad (81)$$

$$C_3 = C_2 \oplus \bigoplus_{C \in \mathcal{Y}} C \oplus (\text{smaller cycles}), \quad (82)$$

where (i) $|C| = |C_1|$ for $C \in \mathcal{X} \cup \mathcal{Y}$, (ii) $\mathcal{X} \cup \{C_2\}$ and $\mathcal{Y} \cup \{C_2\}$ are mutually relevant, and (iii) $C_1, C_2 \notin \mathcal{X}$ and $C_2, C_3 \notin \mathcal{Y}$. Note, the set $\mathcal{X} \cup \{C_1\}$ is mutually relevant, as it belongs to the MCB obtained by the exchange $\mathcal{M}_1 = (\mathcal{M}_2 \setminus \{C_2\}) \cup \{C_1\}$, where $\mathcal{M}_2$ is an MCB such that $\mathcal{X} \cup \{C_2\} \subseteq \mathcal{M}_2$. $\mathcal{Y} \cup \{C_3\}$ is mutually relevant by a similar argument.

Motivated by the polyhedron $\mathcal{P}_3$ combining $\mathcal{P}_1$ and $\mathcal{P}_2$ in our example, we iteratively merge the mutually relevant sets $\mathcal{X} \cup \{C_2\}$ and $\mathcal{Y}$. First, initialize $\mathcal{Z}_0 = \mathcal{X} \cup \{C_2\}$ and arbitrarily order the cycles $\mathcal{Y} = \{Y_1, \dots, Y_k\}$. Then, for $j = 1, \dots, k$ we do the following

1. if there exists $\tilde{Y}_j \in \text{span}(\mathcal{Z}_{j-1})$ such that $Y_j = \tilde{Y}_j \oplus (\text{smaller cycles})$ then $\mathcal{Z}_j = \mathcal{Z}_{j-1}$,
2. otherwise, set $\mathcal{Z}_j = \mathcal{Z}_{j-1} \cup \{Y_j\}$.

Recall from Lemma 19, a set of cycles is mutually relevant if it does not have a subset that sums to smaller cycles. Thus, at each step the cycles $\mathcal{Z}_j$ are mutually relevant. Furthermore, all cycles $C \in \mathcal{X} \cup \mathcal{Y} \cup \{C_2\}$ are within the span of $\mathcal{Z}_k \subseteq \mathcal{X} \cup \mathcal{Y} \cup \{C_2\}$ up to smaller cycles.

By definition of mutual relevance, let $\mathcal{M}_k$ be an MCB containing $\mathcal{Z}_k$. Consider the expansion of C_3 with respect to $\mathcal{M}_k$. By Eq. (82), C_3 is in the span of $\mathcal{Z}_k$ up to smaller cycles. By Lemma 19, the mutually relevant cycles $\mathcal{Y} \cup \{C_3\}$ cannot be combined to sum to smaller cycles. Thus, there exists an additional cycle in the expansion $C^* \in \mathcal{E}_{\mathcal{M}_k}(C_3)$, $C^* \in \mathcal{Z}_k \setminus \mathcal{Y}$, of length $|C^*| = |C_3|$. By Statement 2 of Lemma 4 we can exchange C_3 for C^* to obtain a new MCB $\mathcal{M} = (\mathcal{M}_k \setminus \{C^*\}) \cup \{C_3\}$.

Finally, we want to show that $C_1 \xleftrightarrow{\text{pi}} C_3$ in $\mathcal{M}$ by showing $C_3 \in \mathcal{E}_{\mathcal{M}}(C_1)$, $|C_1| = |C_3|$. By linearity of the expansion operator

$$\mathcal{E}_{\mathcal{M}}(C_1) = \mathcal{E}_{\mathcal{M}}(C_1 \oplus C_3) \triangle \{C_3\}. \quad (83)$$

where $C_1 = (C_1 \oplus C_3) \oplus C_3$ and $\mathcal{E}_{\mathcal{M}}(C_3) = \{C_3\}$. Next, consider the identity

$$\mathcal{E}_{\mathcal{M}_k}(C_1 \oplus C_3) = \mathcal{E}_{\mathcal{M}_k}(C_1) \triangle \mathcal{E}_{\mathcal{M}_k}(C_3). \quad (84)$$

By construction, C^* belongs to both terms in the right-hand side of Eq. (84), where the cycles of length $|C_1|$ in $\mathcal{E}_{\mathcal{M}_k}(C_1)$ are $\mathcal{X} \cup \{C_2\}$ by Eq. (81). Hence, C^* cancels in the symmetric difference in Eq. (84) and

$$\mathcal{E}_{\mathcal{M}_k}(C_1 \oplus C_3) \subseteq \mathcal{M}_k \setminus \{C^*\} \subset \mathcal{M}. \quad (85)$$

Thus, $\mathcal{E}_{\mathcal{M}_k}(C_1 \oplus C_3)$ is the expansion of $C_1 \oplus C_3$ in $\mathcal{M}$ as well. We substitute $\mathcal{E}_{\mathcal{M}_k}(C_1 \oplus C_3)$ for $\mathcal{E}_{\mathcal{M}}(C_1 \oplus C_3)$ in Eq. (83):

$$\mathcal{E}_{\mathcal{M}}(C_1) = \mathcal{E}_{\mathcal{M}_k}(C_1 \oplus C_3) \cup \{C_3\} \quad (86)$$

where $C_3 \notin \mathcal{M}_k$. To conclude $C_3 \in \mathcal{E}_{\mathcal{M}}(C_1)$ so $C_1 \xleftrightarrow{\text{pi}} C_3$ in $\mathcal{M}$ as desired. $\square$

Next, we prove the **pi** classes can be computed using a single MCB.

Proof of Theorem 8. Let $\mathcal{M}$ be a given MCB. We have already shown that if $C_1 \in \mathcal{C}_{\mathcal{R}}$ and $C_2 \in \mathcal{E}_{\mathcal{M}}(C_1)$, $|C_1| = |C_2|$, then $C_1 \xleftrightarrow{\text{pi}} C_2$. See Lemma 4, Statement 2. We want to show the **pi** classes are obtained from the transitive closure of these relations.

Let $C_1 \xleftrightarrow{\text{pi}} C_2$. By Lemma 20, there exists a set of cycles $\mathcal{X}$, $C_1, C_2 \notin \mathcal{X}$, such that

$$C_1 = C_2 \oplus \bigoplus_{C \in \mathcal{X}} C \oplus (\text{smaller cycles}) \quad (87)$$

where $|C| = |C_1|$ for $C \in \mathcal{X}$ and $\mathcal{X} \cup \{C_2\}$ is mutually relevant. In the following identity, we expand Eq. (87) in $\mathcal{M}$ and use linearity of the expansion operator

$$\begin{aligned} \mathcal{E}_{\mathcal{M}}(C_1) \triangle \mathcal{E}_{\mathcal{M}}(C_2) \triangle \left(\bigtriangleup_{C \in \mathcal{X}} \mathcal{E}_{\mathcal{M}}(C) \right) \\ = (\text{smaller cycles}) \end{aligned} \quad (88)$$

In Eq. (88), (smaller cycles) represents a set of smaller cycles $\mathcal{Y}$, $|C'| < |C_1|$ for $C' \in \mathcal{Y}$, which belonging to $\mathcal{M}$ by Lemma 18. We utilize Eq. (88) to obtain **pi** relations from $\mathcal{M}$ which we can stitch together to show that $C_1 \xleftrightarrow{\text{pi}} C_2$.

Consider the following procedure

1. Initialize $\mathcal{X}_0 = \emptyset$ and index $j = 1$.
2. Let $B_j \in \mathcal{M}$ be a cycle of length $|B_j| = |C_1|$ in the expansion of $\mathcal{X}_{j-1} \cup \{C_2\}$ in $\mathcal{M}$,

$$B_j \in \mathcal{E}_{\mathcal{M}}(C_2) \triangle \left(\bigtriangleup_{C \in \mathcal{X}_{j-1}} \mathcal{E}_{\mathcal{M}}(C) \right).$$

3. Let $X_j \in (\mathcal{X} \cup \{C_1\}) \setminus \mathcal{X}_{j-1}$ such that $B_j \in \mathcal{E}_{\mathcal{M}}(X_j)$ and set $\mathcal{X}_j = \mathcal{X}_{j-1} \cup \{X_j\}$.
4. If $X_j = C_1$ then exit, otherwise repeat steps 2-4 with $j \leftarrow j + 1$.

The feasibility of this procedure is as follows. In step 2, the cycles $\mathcal{X}_{j-1} \cup \{C_2\} \subseteq \mathcal{X} \cup \{C_2\}$ are mutually relevant, so they do not sum to shorter cycles by Lemma 19. Thus, there exists an equal length cycle B_j , $|B_j| = |C_1|$, in the expansion of $\mathcal{X}_{j-1} \cup \{C_2\}$ with respect to $\mathcal{M}$. For step 3, we observe that B_j is not contained in the right-hand side of Eq. (88) because it is not a smaller cycle. Therefore, there exists an additional cycle X_j such that $B_j \in \mathcal{E}_{\mathcal{M}}(X_j)$ which cancels with the B_j in the expansion of $\mathcal{X}_{j-1} \cup \{C_2\}$ in the left-hand side of Eq. (88).

We show by an inductive argument the cycles $\mathcal{X}_j \cup \{C_2\}$ are in the same **pi** class via the transitive closure of the **pi** relations given by $\mathcal{M}$. Initially, $\mathcal{X}_0 \cup \{C_2\}$ is a single cycle, so there is nothing to prove. For the inductive step, let $\mathcal{X}_{j-1} \cup \{C_2\}$ belong to the same **pi** class by the transitive closure of **pi** relations from $\mathcal{M}$. By construction, B_j is an equal length cycle in the expansion of both $\mathcal{X}_{j-1} \cup \{C_2\}$ and X_j in $\mathcal{M}$. Thus, B_j is **pi**-related to a cycle in $\mathcal{X}_{j-1} \cup \{C_2\}$ and to X_j via $\mathcal{M}$. Therefore, the cycles $\mathcal{X}_j \cup \{C_2\} = \mathcal{X}_{j-1} \cup \{C_2\} \cup \{X_j\}$ are in the same **pi** class by the transitive closure of **pi** relations from $\mathcal{M}$.

The stopping condition of this procedure is that $C_1 \in \mathcal{X}_j \cup \{C_2\}$. Hence, $C_1 \xleftrightarrow{\text{pi}} C_2$ by the transitive closure of the **pi** relations from $\mathcal{M}$. The cycles C_1, C_2 are arbitrary, so $\mathcal{M}$ can be used to identify all **pi** relations. $\square$

Last, we prove we can always exchange cycles between distinct MCBs.

Proof of Lemma 9. Let $\mathcal{M}_1, \mathcal{M}_2$ be distinct MCBs and $C_2 \in \mathcal{M}_2 \setminus \mathcal{M}_1$. We want to construct a cycle $C_1 \in \mathcal{M}_1 \setminus \mathcal{M}_2$ such that $(\mathcal{M}_1 \setminus \{C_1\}) \cup \{C_2\}$ is an MCB.

We expand C_2 in $\mathcal{M}_1$ as follows

$$C_2 = \bigoplus_{\substack{C' \in \mathcal{E}_{\mathcal{M}_1}(C_2) \\ |C'| = |C_2|}} C' \oplus \bigoplus_{\substack{C'' \in \mathcal{E}_{\mathcal{M}_1}(C_2) \\ |C''| < |C_2|}} C''. \quad (89)$$

Next, we move the equal length cycles to the left-hand side.

$$C_2 \oplus \bigoplus_{\substack{C' \in \mathcal{E}_{\mathcal{M}_1}(C_2) \\ |C'| = |C_2|}} C' = \bigoplus_{\substack{C'' \in \mathcal{E}_{\mathcal{M}_1}(C_2) \\ |C''| < |C_2|}} C''. \quad (90)$$

Importantly, the cycles of length $|C_2|$ in $\{C_2\} \cup \mathcal{E}_{\mathcal{M}_1}(C_2)$ sum to shorter cycles, so they are not mutually relevant by Lemma 19.

By definition, the cycles of length $|C_2|$ in $\mathcal{M}_2$ are mutually relevant. A subset of a mutually relevant set is mutually relevant, so the cycles of length $|C_2|$ in $\{C_2\} \cup \mathcal{E}_{\mathcal{M}_1}(C_2)$ are not contained in $\mathcal{M}_2$. In particular, there exists $C_1 \in \mathcal{E}_{\mathcal{M}_1}(C_2)$ of length $|C_1| = |C_2|$ such that $C_1 \in \mathcal{M}_1 \setminus \mathcal{M}_2$. By Lemma 4, Statement 2, we may exchange C_2 for C_1 in $\mathcal{M}_1$ as desired. $\square$

A.3 Short loop-interchangeability

Here, we prove the theoretical results for the **sli** relation from Section 3.2. Recall, two equal length cycles C_1 and C_2 satisfy $C_1 \xleftrightarrow{\text{sli}} C_2$ if there exist shorter cycles $\mathcal{Y}$, $|C| < |C_1|$ for $C \in \mathcal{Y}$, such that

$$C_1 = C_2 \oplus \bigoplus_{C \in \mathcal{Y}} C. \quad (91)$$

First, we prove **sli** is an equivalence relation.

Proof of Lemma 11. The reflexive property is immediate by $C = C \oplus \emptyset$.

(symmetry) Let $C_1 \xleftrightarrow{\text{sli}} C_2$ via shorter cycles $\mathcal{Y}$. By adding cycles $\mathcal{Y}$ to both sides of Eq. (91) we find that $C_2 \xleftrightarrow{\text{sli}} C_1$:

$$C_1 = C_2 \oplus \bigoplus_{C \in \mathcal{Y}} C \Rightarrow C_2 = C_1 \oplus \bigoplus_{C \in \mathcal{Y}} C.$$

(transitivity) Let $C_1 \xleftrightarrow{\text{sli}} C_2$ and $C_2 \xleftrightarrow{\text{sli}} C_3$ by sets of smaller cycles $\mathcal{Y}$ and $\mathcal{Y}'$ respectively,

$$C_1 = C_2 \oplus \bigoplus_{C \in \mathcal{Y}} C, \quad C_2 = C_3 \oplus \bigoplus_{C' \in \mathcal{Y}'} C'.$$

Substituting the right-hand side of the second equation for C_2 in the first equation gives

$$C_1 = \left(C_3 \oplus \bigoplus_{C' \in \mathcal{Y}'} C' \right) \oplus \bigoplus_{C \in \mathcal{Y}} C = C_3 \oplus \bigoplus_{C'' \in \mathcal{Y} \triangle \mathcal{Y}'} C''.$$

Hence, $C_1 \xleftrightarrow{\text{sli}} C_3$ via shorter cycles $\mathcal{Y} \triangle \mathcal{Y}'$. $\square$

Next, we prove C_1 and C_2 satisfy $C_1 \xleftrightarrow{\text{sli}} C_2$ if and only if their expansions with respect to a given MCB differ by shorter cycles.

Proof of Lemma 12. ($\Leftarrow$) Let $\mathcal{M}$ be an MCB and $C_1, C_2 \in \mathcal{C}_{\mathcal{R}}$ such that $|C_1| = |C_2|$. Let the expansions of C_1 and C_2 in $\mathcal{M}$ differ by shorter cycles, i.e., there exist cycles $\mathcal{Y} \subseteq \mathcal{M}$, $|C| < |C_1|$ for $C \in \mathcal{Y}$, such that

$$\mathcal{E}_{\mathcal{M}}(C_1) = \mathcal{E}_{\mathcal{M}}(C_2) \triangle \mathcal{Y}. \quad (92)$$

By summing over cycles in Eq. (92) we obtain

$$C_1 = C_2 \oplus \bigoplus_{C \in \mathcal{Y}} C.$$

Thus, $C_1 \xleftrightarrow{\text{sli}} C_2$ via shorter cycles $\mathcal{Y}$ as desired.

($\Rightarrow$) Let $C_1 \xleftrightarrow{\text{sli}} C_2$ by shorter cycles $\mathcal{Y}$ as in Eq. (91). Then, by linearity of the expansion operator applied to Eq. (91) we obtain

$$\begin{aligned} \mathcal{E}_{\mathcal{M}}(C_1) &= \mathcal{E}_{\mathcal{M}}(C_2) \triangle \mathcal{E}_{\mathcal{M}}\left(\bigoplus_{C \in \mathcal{Y}} C\right) \\ &= \mathcal{E}_{\mathcal{M}}(C_2) \triangle \mathcal{Y}' \end{aligned} \quad (93)$$

where $\mathcal{Y}' \subseteq \mathcal{M}$ is a set of shorter cycles in $\mathcal{M}$ by Lemma 18. Thus, the expansions of C_1 and C_2 in $\mathcal{M}$ differ by shorter cycles $\mathcal{Y}'$ as desired. $\square$

B Proofs for Section 5.2

We prove these results using a distance-based approach. The distance $d(u, v)$ between nodes of a graph G is the length of a shortest path connecting u and v . Recall, a path P belonging to a relevant cycle C of length $|P| \leq |C|/2$ is a shortest path [29] – see Lemma 1. A pair of nodes u and v in C are connected by two paths in C , one for each orientation of C . The shorter of these paths, $P_{u,v}$, is bounded in length by $|P_{u,v}| \leq |C|/2$. Thus, every pair of nodes in C are connected by a shortest path of G in C . We restate this as a result:

Corollary 21 (distance preservation). *Let C be a relevant cycle of a graph G and let u, v be nodes belonging to C . The distance between u and v with respect to C is equal to their distance with respect to G .*

Using Corollary 21, we employ the notion of resolving sets. A resolving set [53] is a set of nodes $R = \{r_1, \dots, r_k\} \subseteq V$ such that each node $x \in V$ is uniquely identified by its distances to the nodes R . For general graphs, an efficient resolving set is hard to construct. However, a resolving set of a cycle requires only two nodes.

Lemma 22. *Let C be a cycle graph and u, v nodes in C such that $d(u, v) < |C|/2$. Then, every node x in C is uniquely identified by its distances to u and v , i.e., the function*

$$\varphi(x) = (d(x, u), d(x, v)), \quad x \in V(C) \quad (94)$$

is one-to-one.

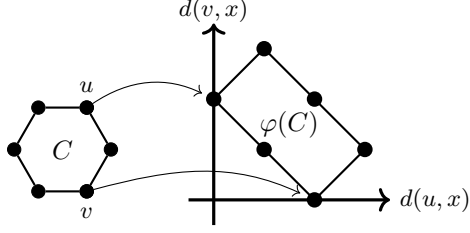


Figure 28: Even cycle for Lemma 22.

Proof. Let u, v be nodes belonging to a cycle graph C and φ given by Eq. (94). We want to show φ is bijective. We also want to describe the shape given by $\varphi(C)$ which will be useful in the proof of Theorem 16 as well. When plotting φ , the x and y coordinates correspond to the distance of a given node to u and v respectively. This is shown for an even cycle in Figure 28. We observe that φ maps this cycle to the integer coordinates of a rectangle.

In contrast, φ maps an odd cycle to a rectangle with two of its corners cut off as shown in Figure 29 below.

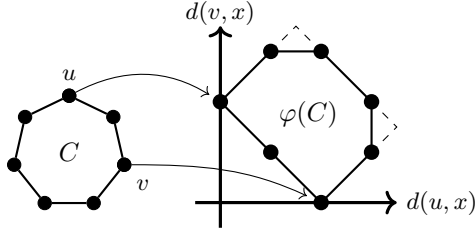


Figure 29: Odd cycle for Lemma 22.

In either case, the function φ maps our cycle to the integer coordinates of a polygon. This shape is not self-intersecting, so φ is one-to-one. The only exception to this is if the nodes u and v are antipodal ($d(u, v) = |C|/2$) as shown in Figure 30 below. In this case, the symmetries of u and

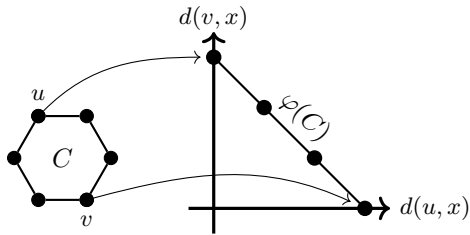


Figure 30: Degenerate case for Lemma 22.

v cause the mapping $\varphi(C)$ to collapse from a rectangle into a line. Because this line doubles back on itself the function φ is not one-to-one when $d(u, v) = |C|/2$. $\square$

Next, we use the resolving set representation of cycles to describe the intersection format of two cycles in an MCB.

Proof of Theorem 16. Let $C_1, C_2, |C_1| \leq |C_2|$, be cycles in an MCB $\mathcal{M}$. We split this proof into three cases, a main case where the methods in Lemma 22 apply, and two degenerate cases.

We remark that a path can be a single node, e.g., the path P_3 in Figure 32. The length of a path is the number of edges it has. A path with one node has length 0.

Main case. Let u and v be nodes belonging to both C_1 and C_2 such that $d(u, v) \leq |C_1|/2$. Then, we may use the function $\varphi(x) = (d(u, x), d(v, x))$ to embed C_1 and C_2 as overlapping rectangles $\varphi(C_1)$ and $\varphi(C_2)$ (or rectangles with cut corners as in Figure 29). By Lemma 22, the mapping φ is one-to-one when its domain is restricted to C_1 or to C_2 . This shared embedding is shown in Figure 31. In Figure 31 blue nodes belong to C_1 , red nodes to C_2 , and purple nodes to both. Dual

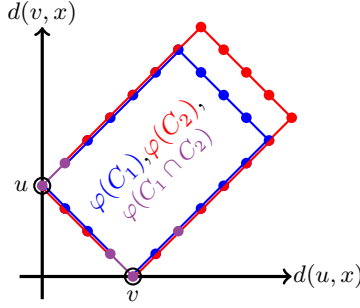


Figure 31: Embedding of C_1 and C_2 for the main case of Theorem 16.

blue and red points denote a pair of nodes $x_1 \in V(C_1)$, $x_2 \in V(C_2)$ that share the same embedding $\varphi(x_1) = \varphi(x_2)$. Edges are colored similarly.

The mapping φ couples the cycles C_1 and C_2 by stacking them on top of each other. We orient C_1 and C_2 by traversing $\varphi(C_1)$ and $\varphi(C_2)$ in the counter-clockwise direction. If C_1 and C_2 are equal in length, then the rectangles $\varphi(C_1)$ and $\varphi(C_2)$ overlap completely. If C_2 is longer than C_1 , then the rectangle $\varphi(C_2)$ extends past $\varphi(C_1)$, as in Figure 31.

We use this shared embedding to decompose C_1 and C_2 into paths, starting with their intersection paths $P_1, \dots, P_k$. Since we traverse both $\varphi(C_1)$ and $\varphi(C_2)$ in the counter-clockwise direction, these intersection paths have the same order and orientation in C_1 and C_2 . See Figure 32. In

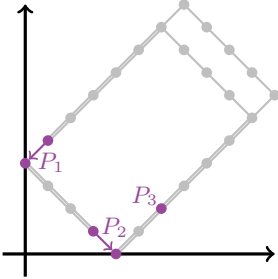


Figure 32: Intersection paths of C_1 and C_2 and their directions in the main case of Theorem 16.

Figure 32, we mark the orientation of P_1 and P_2 using purple arrows. The orientation of P_3 is arbitrary, because it is a single point. We remark that we have not selected a starting point when traversing $\varphi(C_1)$ and $\varphi(C_2)$. Because of this, the order of the intersection paths is circular, e.g., in Figure 32 P_2 follows P_1 , P_3 follows P_2 , and P_1 follows P_3 .

Next, we discuss the paths $Q_i^{(j)}$ that separate C_1 and C_2 . All of the path pairs $Q_i^{(1)}, Q_i^{(2)}$ fully overlap – as in Figure 33 (a) – unless $|C_1| < |C_2|$. If C_1 is shorter than C_2 , then the pair that includes the northeast lines of the rectangles $\varphi(C_1), \varphi(C_2)$ do not overlap, as in Figure 33 (b). We prove three identities regarding the lengths of these pairs.

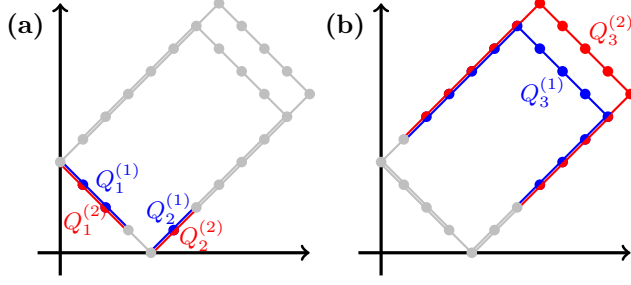


Figure 33: Separated path pairs $Q_i^{(1)}, Q_i^{(2)}$ for the main case of Theorem 16.

1. Let $|C_1| < |C_2|$, and order the path pair with different lengths last: $|Q_k^{(1)}| \neq |Q_k^{(2)}|$. This is the pair shown in Figure 33 (b). We will show this pair satisfies

$$|Q_k^{(2)}| = |Q_k^{(1)}| + |C_2| - |C_1| \quad (95)$$

$$|Q_k^{(j)}| \geq |C_j|/2, \quad j = 1, 2. \quad (96)$$

Eq. (95) is obvious because $Q_k^{(1)}$ and $Q_k^{(2)}$ account for the difference in lengths between C_1 and C_2 :

$$|Q_k^{(2)}| - |Q_k^{(1)}| = |C_2| - |C_1|$$

As for Eq. (96), consider the cycle C'_2 obtained by swapping the path $Q_k^{(2)}$ in C_2 for $Q_k^{(1)}$,

$$C'_2 = C_2 \oplus C_k^e, \quad C_k^e = Q_k^{(1)} \cup Q_k^{(2)}.$$

C_2 cannot be broken into shorter cycles because it is relevant. C'_2 is shorter than C_2 because $Q_k^{(1)}$ is shorter than $Q_k^{(2)}$. Hence, the cycle C_k^e used to exchange C_2 for C'_2 satisfies

$$|C_k^e| = |Q_k^{(1)}| + |Q_k^{(2)}| \geq |C_2|. \quad (97)$$

Eq. (96) is obtained by combining Eqs. (95) and (97).

2. Let $|C_1| = |C_2|$. We will show there exists a unique large path pair satisfying

$$|Q_k^{(1)}|, |Q_k^{(2)}| \geq |C_1|/2. \quad (98)$$

This pair is ordered last to match the previous item.

Suppose towards a contradiction that each path pair is shorter,

$$|Q_i^{(1)}| = |Q_i^{(2)}| < |C_1|/2,$$

for $i = 1, \dots, k$. Then C_2 can be obtained from C_1 by swapping each path $Q_i^{(1)}$ for $Q_i^{(2)}$,

$$C_2 = C_1 \oplus (C_1^e \oplus \dots \oplus C_k^e)$$

$$C_i^e = Q_i^{(1)} \cup Q_i^{(2)}, \quad i = 1, \dots, k.$$

The cycles C_i^e are shorter

$$|C_i^e| = |Q_i^{(1)}| + |Q_i^{(2)}| < |C_1|.$$

Thus, $C_1 \xrightarrow{\text{sli}} C_2$. This is a contradiction because at most one cycle in a **sli** class may belong to an MCB by Corollary 13.

3. Last, we show the remaining path pairs are bounded above in length by

$$|Q_i^{(1)}|, |Q_i^{(2)}| < |C_1|/2, \quad (99)$$

for $i = 1, \dots, k-1$. Consider a $Q_i^{(j)}$ that lies on a shortest path from u to v , such as $Q_1^{(1)}$ or $Q_1^{(2)}$ in Figure 33 (a). Such a path is bounded in length by

$$|Q_i^{(j)}| \leq d(u, v) < |C_1|/2.$$

Next, consider a $Q_i^{(j)}$ that does not lie on a shortest path from u to v , e.g., $Q_2^{(1)}$ and $Q_2^{(2)}$ in Figure 33 (a). Then, $Q_i^{(j)}, i \neq k$ is bounded in length by

$$\begin{aligned} |Q_i^{(j)}| &= |Q_i^{(1)}| \leq |C_1| - |Q_k^{(1)}| - d(u, v) \\ &< |C_1|/2 \end{aligned}$$

because (i) $|Q_i^{(1)}| = |Q_i^{(2)}|$ for $i \neq k$, (ii) the path $Q_k^{(1)}, |Q_k^{(1)}| > |C_1|/2$ also does not lie on a shortest path from u to v , and (iii) $d(u, v) > 0$ because $u \neq v$. In both cases $Q_i^{(j)}$ satisfies the length bound in Eq. (99).

In summary, these paths satisfy the following:

$$|Q_i^{(1)}| = |Q_i^{(2)}|, \quad i = 1, \dots, k-1, \quad (100)$$

$$|Q_i^{(j)}| < |C_1|/2, \quad i=1, \dots, k-1, j=1, 2, \quad (101)$$

$$|Q_k^{(2)}| = |Q_k^{(1)}| + |C_2| - |C_1|, \quad (102)$$

$$|Q_k^{(j)}| \geq |C_j|/2, \quad j = 1, 2. \quad (103)$$

We note that, Eq. (101) is a strict inequality, which is a stronger result than we intended to show. This observation will be useful in the proof of Theorem 17. This inequality does not extend to degenerate case 2 below.

Degenerate case 1. Let C_1 and C_2 share only a single node u . See Figure 34. Then they overlap

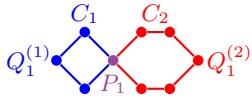


Figure 34: Degenerate case 1 for Theorem 16.

over a single path $P_1 = (u)$. They are separated by a pair of paths $Q_1^{(1)}$ and $Q_1^{(2)}$ connecting u to itself. These paths are of length $|Q_1^{(1)}| = |C_1|$ and $|Q_1^{(2)}| = |C_2|$ which satisfies the required length requirements of the larger path pairs $Q_k^{(1)}, Q_k^{(2)}$.

Degenerate case 2. Let C_1 and C_2 share two nodes u and v which are antipodal in C_1 , i.e., $d(u, v) = |C_1|/2$. See Figure 35. In this case, C_1 and C_2 overlap over two paths consisting of single nodes $P_1 = (u)$ and $P_2 = (v)$. C_1 is composed of two shortest paths from u to v . Label these paths $Q_1^{(1)}$ and $Q_2^{(1)}$ arbitrarily. Also, C_2 has a shortest path connecting u and v . Label this path $Q_1^{(2)}$ where $|Q_1^{(1)}| = |Q_2^{(1)}| = |C_1|/2$ satisfying Eqs. (63) and (64). Let $Q_2^{(2)}$ be the other path connecting u and v in C_2 . This path has length $|Q_2^{(2)}| = |C_2| - |C_1|/2$ which satisfies Eqs. (65) and (66). $\square$

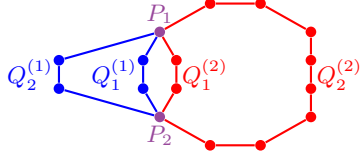


Figure 35: Degenerate case 2 for Theorem 16.

Next, we prove that if C_1 and C_2 belong to an MCB and they intersect over multiple paths, then we may exchange C'_1 for C_1 such that C'_1 and C_2 intersect over a single path.

Proof of Theorem 17. Here, we assume the cycles C_1, C_2 intersect over multiple paths, otherwise, no modification is needed and we may set $C'_1 = C_1$. By Theorem 16, let C_1 and C_2 intersect over paths $P_1, \dots, P_k$ and deviate over path pairs $Q_i^{(j)}$, $j = 1, 2$, for $i = 1, \dots, k$. We split the proof into two cases. Note, we do not assume $|C_1| < |C_2|$.

Main case. Let C_1 and C_2 belonging to an MCB $\mathcal{M}$ share two nodes u and v of distance $d(u, v) < |C_1|/2$. As mentioned in the main case in the proof of Theorem 16, the paths $Q_i^{(1)}$ separating C_1 from C_2 satisfy the length bound $|Q_i^{(1)}| < |C_1|/2$ for $i = 1, \dots, k-1$. Also by Theorem 16, the paths $Q_i^{(2)}$ satisfy $|Q_i^{(2)}| = |Q_i^{(1)}|$ for $i = 1, \dots, k-1$. Thus, we may exchange the paths in C_1 for those in C_2 by adding cycles

$$C_i^e = Q_i^{(1)} \cup Q_i^{(2)}, \quad i = 1, \dots, k-1, \quad (104)$$

resulting in the cycle C'_1 shown in Figure 36 below.

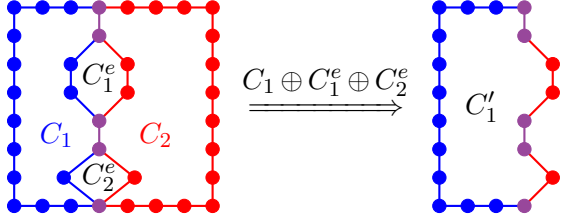


Figure 36: Main case of Theorem 17.

Importantly, the exchange cycles are shorter than C_1

$$|C_i^e| = |Q_i^{(1)}| + |Q_i^{(2)}| < |C_1|, \quad (105)$$

Thus, the cycle

$$C'_1 = C_1 \oplus (C_1^e \oplus \dots \oplus C_{k-1}^e) \quad (106)$$

satisfies $C'_1 \xleftrightarrow{\text{sl}} C_1$. By Corollary 13, we may exchange C'_1 for C_1 to obtain a new MCB $\mathcal{M}' = (\mathcal{M} \setminus \{C_1\}) \cup \{C'_1\}$. Moreover, C'_1 and C_2 intersect over a single larger path.

Degenerate case. Let C_1 and C_2 belonging to an MCB $\mathcal{M}$ intersect at two nodes u and v of distance $d(u, v) = |C_1|/2$. In this case, C_1 is composed of two shortest paths $Q_1^{(1)}$ and $Q_2^{(1)}$ joining u to v . Also, C_2 has a shortest path $Q_1^{(2)}$ connecting u to v . We consider two candidate cycles C'_1 and C''_1 obtained by switching one of the paths from u to v in C_1 for $Q_2^{(1)}$ in C_2 . The cycles C'_1 and C''_1 and their relation to C_1 are shown in Figure 37 below.

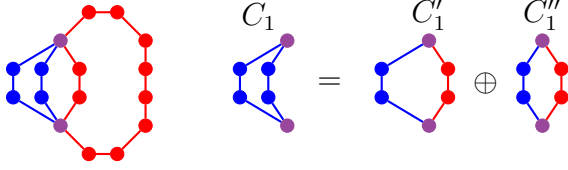


Figure 37: Degenerate case of Theorem 17.

Algebraically, C'_1 and C''_1 are given by

$$C'_1 = C_1 \oplus (Q_1^{(2)} \cup Q_2^{(1)}) = Q_1^{(1)} \cup Q_1^{(2)}, \quad (107)$$

$$C''_1 = C_1 \oplus (Q_1^{(1)} \cup Q_1^{(2)}) = Q_1^{(2)} \cup Q_2^{(1)}, \quad (108)$$

and C_1 is equal to their sum

$$C_1 = C'_1 \oplus C''_1 = Q_1^{(1)} \cup Q_2^{(1)}. \quad (109)$$

We remark that, C'_1 and C''_1 deviate from C_1 by equal length cycles, so we can not exchange them for C_1 using the `sli` relation as in the main case. However, by linearity of the expansion operator, we observe C_1 is in the expansion of C'_1 or C''_1

$$\mathcal{E}_{\mathcal{M}}(C'_1) \triangle \mathcal{E}_{\mathcal{M}}(C''_1) = \mathcal{E}_{\mathcal{M}}(\underbrace{C'_1 \oplus C''_1}_{C_1}) = \{C_1\}.$$

Without loss of generality, let $C_1 \in \mathcal{E}_{\mathcal{M}}(C'_1)$. Then, by Lemma 4, $\mathcal{M}' = (\mathcal{M} \setminus \{C_1\}) \cup \{C'_1\}$ is an MCB. Furthermore, C'_1 and C_2 intersect over a single path $Q_2^{(1)}$, completing our proof. $\square$

References

- [1] P. M. Gleiss, J. Leydold, and P. F. Stadler, “Interchangeability of Relevant Cycles in Graphs,” *The Electronic journal of combinatorics*, vol. 7, no. 1, 2000.
- [2] A. Torres-Knoop, I. Kryven, V. Schamboeck, and P. D. Iedema, “Modeling the free-radical polymerization of hexanediol diacrylate (HDDA): A molecular dynamics and graph theory approach,” *Soft Matter*, vol. 14, pp. 3404–3414, May 2018.
- [3] A. Dernov, M. Kowalik, A. C. T. van Duin, and T. Dumitrică, “Mapping the structural–mechanical landscape of amorphous carbon with ReaxFF molecular dynamics,” *Journal of Applied Physics*, vol. 137, p. 065107, Feb. 2025.
- [4] V. L. Deringer, “Machine learning based interatomic potential for amorphous carbon,” *Physical Review B*, vol. 95, no. 9, 2017.
- [5] E. Villa-Aleman, J. R. Darvin, M. H. Nielsen, and T. M. Willey, “Raman signatures of detonation soot,” *Journal of Raman Spectroscopy*, vol. 53, no. 9, pp. 1571–1579, 2022.
- [6] A. N. Clark, J. M. D. Lane, J.-P. Davis, A. R. Sarafian, K. R. Cochrane, J. P. Townsend, and S. D. Jacobsen, “Shock-ramp of SiO₂ melt,” *AIP Conference Proceedings*, vol. 2844, p. 330002, Sept. 2023.

- [7] B. Cheng, S. Hamel, and M. Bethkenhagen, “Thermodynamics of diamond formation from hydrocarbon mixtures in planets,” *Nature Communications*, vol. 14, no. 1, p. 1104, 2023.
- [8] E. Chen, Q. Yang, V. Dufour-Décieux, C. A. Sing-Long, R. Freitas, and E. J. Reed, “Transferable Kinetic Monte Carlo Models with Thousands of Reactions Learned from Molecular Dynamics Simulations,” *The Journal of Physical Chemistry A*, vol. 123, pp. 1874–1881, Mar. 2019.
- [9] V. Dufour-Décieux, B. Ransom, A. D. Sendek, R. Freitas, J. Blanchet, and E. J. Reed, “Temperature Extrapolation of Molecular Dynamics Simulations of Complex Chemistry to Microsecond Timescales Using Kinetic Models: Applications to Hydrocarbon Pyrolysis,” *Journal of Chemical Theory and Computation*, vol. 18, pp. 7496–7509, Dec. 2022.
- [10] V. Dufour-Décieux, C. Moakler, E. J. Reed, and M. Cameron, “Predicting molecule size distribution in hydrocarbon pyrolysis using random graph theory,” *The Journal of Chemical Physics*, vol. 158, p. 024101, Jan. 2023.
- [11] V. Dufour-Décieux, R. Freitas, and E. J. Reed, “Atomic-Level Features for Kinetic Monte Carlo Models of Complex Chemistry from Molecular Dynamics Simulations,” *The Journal of Physical Chemistry A*, vol. 125, pp. 4233–4244, May 2021.
- [12] D. T. Gillespie, “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions,” *Journal of Computational Physics*, vol. 22, pp. 403–434, Dec. 1976.
- [13] D. J. Higham, “Modeling and Simulating Chemical Reactions,” *SIAM Review*, vol. 50, pp. 347–368, Jan. 2008.
- [14] Q. Yang, C. A. Sing-Long, and E. J. Reed, “Learning reduced kinetic Monte Carlo models of complex chemistry from molecular dynamics,” *Chemical Science*, vol. 8, pp. 5781–5796, July 2017.
- [15] Q. Yang, C. A. Sing-Long, E. Chen, and E. J. Reed, “Data-Driven Methods for Building Reduced Kinetic Monte Carlo Models of Complex Chemistry from Molecular Dynamics Simulations,” in *Computational Approaches for Chemistry Under Extreme Conditions* (N. Goldman, ed.), pp. 209–227, Cham: Springer International Publishing, 2019.
- [16] E. Cayley, “Ueber die analytischen figuren, welche in der mathematik bäume genannt werden und ihre anwendung auf die theorie chemischer verbindungen,” *Berichte der deutschen chemischen Gesellschaft*, vol. 8, no. 2, pp. 1056–1059, 1875.
- [17] E. M. Rains and N. J. Sloane, “On cayley’s enumeration of alkanes (or 4-valent trees),” *Journal of Integer Sequences*, vol. 2, no. 99.1, p. 1, 1999.
- [18] P. E. Ruth, V. Dufour-Décieux, C. Moakler, and M. K. Cameron, “Cyclic random graph models predicting giant molecules in hydrocarbon pyrolysis,” *Physical Review E*, vol. 111, p. 034303, Mar. 2025.
- [19] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, “Random graphs with arbitrary degree distributions and their applications,” *Physical Review E*, vol. 64, p. 026118, July 2001.
- [20] M. Molloy and B. Reed, “The Size of the Giant Component of a Random Graph with a Given Degree Sequence,” *Combinatorics, Probability and Computing*, vol. 7, pp. 295–305, Sept. 1998.

- [21] B. Karrer and M. E. J. Newman, "Random graphs containing arbitrary distributions of sub-graphs," *Physical Review E*, vol. 82, p. 066118, Dec. 2010.
- [22] M. E. J. Newman, "Assortative Mixing in Networks," *Physical Review Letters*, vol. 89, p. 208701, Oct. 2002.
- [23] R. Diestel, *Graph Theory*, vol. 173 of *Graduate Texts in Mathematics*. Berlin, Heidelberg: Springer, 2017.
- [24] B. Bollobás, *Modern Graph Theory*, vol. 184 of *Graduate Texts in Mathematics*. New York, NY: Springer, 1998.
- [25] J. Hopcroft and R. Tarjan, "Algorithm 447: Efficient algorithms for graph manipulation," *Commun. ACM*, vol. 16, pp. 372–378, June 1973.
- [26] T. Kavitha, C. Liebchen, K. Mehlhorn, D. Michail, R. Rizzi, T. Ueckerdt, and K. A. Zweig, "Cycle bases in graphs characterization, algorithms, complexity, and applications," *Computer Science Review*, vol. 3, pp. 199–243, Nov. 2009.
- [27] M. Plotkin, "Mathematical Basis of Ring-Finding Algorithms in CIDS," *Journal of Chemical Documentation*, vol. 11, pp. 60–63, Feb. 1971.
- [28] F. Berger, C. Flamm, P. M. Gleiss, J. Leydold, and P. F. Stadler, "Counterexamples in Chemical Ring Perception," *Journal of Chemical Information and Computer Sciences*, vol. 44, pp. 323–331, Mar. 2004.
- [29] P. Vismara, "Union of all the Minimum Cycle Bases of a Graph," *The Electronic Journal of Combinatorics*, pp. R9–R9, Jan. 1997.
- [30] J. W. May and C. Steinbeck, "Efficient ring perception for the Chemistry Development Kit," *Journal of Cheminformatics*, vol. 6, p. 3, Jan. 2014.
- [31] A. Kolodzik, S. Urbaczek, and M. Rarey, "Unique Ring Families: A Chemically Meaningful Description of Molecular Ring Topologies," *Journal of Chemical Information and Modeling*, vol. 52, pp. 2013–2021, Aug. 2012.
- [32] P. M. Gleiss, *Short Cycles Minimum Cycle Bases of Graphs from Chemistry and Biochemistry*. PhD thesis, Universität Wien, 2001.
- [33] S. Nouleho Ilemo, D. Barth, O. David, F. Quessette, M.-A. Weisser, and D. Watel, "Improving graphs of cycles approach to structural similarity of molecules," *PloS one*, vol. 14, no. 12, p. e0226680, 2019.
- [34] P. E. Ruth, "Github: Enhanced sampling of minimum cycle bases," 2025.
- [35] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring Network Structure, Dynamics, and Function using NetworkX," in *Proceedings of the 7th Python in Science Conference* (G. Varoquaux, T. Vaught, and J. Millman, eds.), (Pasadena, CA USA), pp. 11–15, 2008.
- [36] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *Journal of Computational Physics*, vol. 117, pp. 1–19, Mar. 1995.

- [37] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, "LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," *Computer Physics Communications*, vol. 271, p. 108171, Feb. 2022.
- [38] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "ReaxFF: A Reactive Force Field for Hydrocarbons," *The Journal of Physical Chemistry A*, vol. 105, pp. 9396–9409, Oct. 2001.
- [39] K. Chenoweth, A. C. T. van Duin, and W. A. Goddard, "ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation," *The Journal of Physical Chemistry A*, vol. 112, pp. 1040–1053, Feb. 2008.
- [40] S. G. Srinivasan, A. C. T. van Duin, and P. Ganesh, "Development of a ReaxFF Potential for Carbon Condensed Phases and Its Application to the Thermal Fragmentation of a Large Fullerene," *The Journal of Physical Chemistry A*, vol. 119, pp. 571–580, Jan. 2015.
- [41] C. Ashraf and A. C. van Duin, "Extension of the ReaxFF Combustion Force Field toward Syngas Combustion and Initial Oxidation Kinetics," *The Journal of Physical Chemistry A*, vol. 121, pp. 1051–1068, Feb. 2017.
- [42] T. Kavitha, K. Mehlhorn, D. Michail, and K. Paluch, "A Faster Algorithm for Minimum Cycle Basis of Graphs," in *Automata, Languages and Programming* (J. Díaz, J. Karhumäki, A. Lepistö, and D. Sannella, eds.), (Berlin, Heidelberg), pp. 846–857, Springer, 2004.
- [43] Daylight Chemical Information Systems, "SMARTS - A Language for Describing Molecular Patterns." <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>, 2022. (accessed February 2nd, 2025).
- [44] OpenEye Scientific Software, "Smallest Set of Smallest Rings (SSSR) considered Harmful." <https://ics.uci.edu/~dock/manuals/oechem/pyprog/node123.html>, 2004. (accessed February 2nd, 2025).
- [45] F. Berger, P. Gritzmann, and S. de Vries, "Computing cyclic invariants for molecular graphs," *Networks*, vol. 70, no. 2, pp. 116–131, 2017.
- [46] P. McMullen, *Abstract Regular Polytopes*. Cambridge ; New York : Cambridge University Press, 2002.
- [47] K. Mehlhorn and D. Michail, "Minimum cycle bases: Faster and simpler," *ACM Trans. Algorithms*, vol. 6, pp. 8:1–8:13, Dec. 2010.
- [48] R. Durrett and R. Durrett, *Essentials of stochastic processes*, vol. 1. Springer, 2012.
- [49] H. M. Aktulga, J. C. Fogarty, S. A. Pandit, and A. Y. Grama, "Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques," *Parallel Computing*, vol. 38, pp. 245–259, Apr. 2012.
- [50] M. Penrose, *Random Geometric Graphs*, vol. 5. OUP Oxford, 2003.
- [51] C. Fang, Y. Xiong, Y. Li, Y. Chen, J. Liu, H. Zhang, T. A. Adedosu, and P. Peng, "The origin and evolution of adamantanes and diamantanes in petroleum," *Geochimica et Cosmochimica Acta*, vol. 120, pp. 109–120, Nov. 2013.

- [52] D. Franzblau, “Computation of ring statistics for network models of solids,” *Physical Review B*, vol. 44, no. 10, p. 4925, 1991.
- [53] R. C. Tillquist, R. M. Frongillo, and M. E. Lladser, “Getting the lay of the land in discrete space: A survey of metric dimension and its applications,” *SIAM Review*, vol. 65, no. 4, pp. 919–962, 2023.