

ZSE Manuscript Outline—Needs a Title!!!!

Jerry T. Crum^a, Justin R. Crum^b, Cameron Taylor^a, William F. Schneider^{a,c}

^a Department of Chemical and Biomolecular Engineering, University of Notre Dame, 250 Nieuwland Science Hall, Notre Dame, IN 46556, USA

^b Department of Applied Mathematics, University of Arizona, 617 N Santa Rita Ave, Tucson, AZ 85721, USA

^c Department of Chemistry and Biochemistry, University of Notre Dame, 251 Nieuwland Science Hall, Notre Dame, IN 46556, USA

1 Introduction

Zeolites are three dimensional crystalline structures containing tetrahedral Si or Al atoms connected by oxygen bridges. The International Zeolite Association (IZA) lists over 200 known zeolite frameworks that are described by dimensionality, pore shape and size, and Si/Al ratios.¹ It is natural to characterize differences in zeolites by the size and shapes of the features present in the crystal. Rings are one common type of feature widely reported and used, both to characterize a zeolite crystal and as descriptors of the individual tetrahedral sites (T-sites) of a zeolite. Rings have been used to identify feasible zeolites topologies,¹⁶ to describe the similarity and differences between zeolites,^{2,4} to identify sites or voids of catalytic relevance,^{13,15} and as machine learning finger prints of physicochemical properties.¹¹

Many researchers have presented methods to define and count the rings present in zeolites.^{7–9,14,20–22} Zeolites can naturally be represented using graph theory, where atoms are nodes, and bonds are edges. In some cases researchers will use T-sites as nodes and oxygen atoms as edges, since the oxygen atoms are only connected to two T-sites.⁸ Some basic definitions of graphs that will be used going forward are highlighted in a cutout of the chabazite (CHA) framework shown in Fig. 1, and described in Table 1. The size of each of these features can either be defined by the total number of atoms contained in them, or by the number of T-sites contained in them. The latter method is the standard convention which we will use. A path is a sequence of edges which joins a sequence of nodes in which every edge is distinct (shown with purple highlighted bonds in Fig. 1). A cycle is a path that starts and ends at the same node, and does not repeat any other nodes (shown by blue bonds in Fig. 1). In the most basic form a ring is a cycle that does not contain any shortcuts.^{8,9} Fig. 1 shows an 8-membered ring and a 4-MR filled in with pink and green respectively.

A shortcut is defined as a path connecting two nodes of a cycle that decomposes the cycle into smaller rings,^{8,9} or in other words, the path is a shortcut if it is shorter than both paths between these nodes along the cycle. An example is the path connecting T17 and T18 in Fig. 1, which is a shortcut of the blue cycle. This path contains two T-sites, splitting the 8-MR cycle into two 4-MRs. Many definitions build off this idea that rings do not contain shortcuts. In a zeolite, rings are used to define the topology of the framework, such as the channels and void environments that are present. Counting the rings that pass through a T-site can also be used to describe and differentiate the local void environment around symmetry distinct T-sites. Like T-sites, ring counts can be used to differentiate symmetry distinct oxygen atoms.

Vertex symbols are a common way to describe the T-sites of a zeolite framework, and were first used to do so in 1997 by O’Keeffe and Hyde in 1997.¹⁸ The vertex symbol of a T-site contains the shortest rings associated with each of the six edges of the tetrahedron formed by the T-atom and its four bounded oxygen atoms. The ring sizes, and their multiplicity, for opposite edges (edges of the tetrahedron that do not connect to the same oxygen atom) are grouped together. These grouped pairs are listed from smallest to largest forming the vertex symbol. Fig. 2(a) shows an example of the tetrahedron formed at the single symmetry distinct T-site in the CHA framework. The edges of the tetrahedron are labeled to aid in identification. Fig. 2(b-d) show the rings associated with each opposite pair of edges. The vertex symbol of T1 in CHA is thus able to be determined as 4·4·4·8·6·8. For a T-site that contains a multiplicity of rings at one edge, that multiplicity would be represented as a subscript in the vertex symbol. An example would be the vertex symbol of 4·6₂·6·6₃·6₂·6₃ for T1 in AFI.

Another ring counting convention presented by Sastre and Corma is to count only the shortest path connecting an O-T-O in a framework. With this definition, they can find and count all the rings in a framework that are the shortest path for at least one set of O-T-O along the cycle.¹⁹ Using AFI as an example again (??), we see that both a 12-MR (purple) and a 6-MR (blue) pass through the

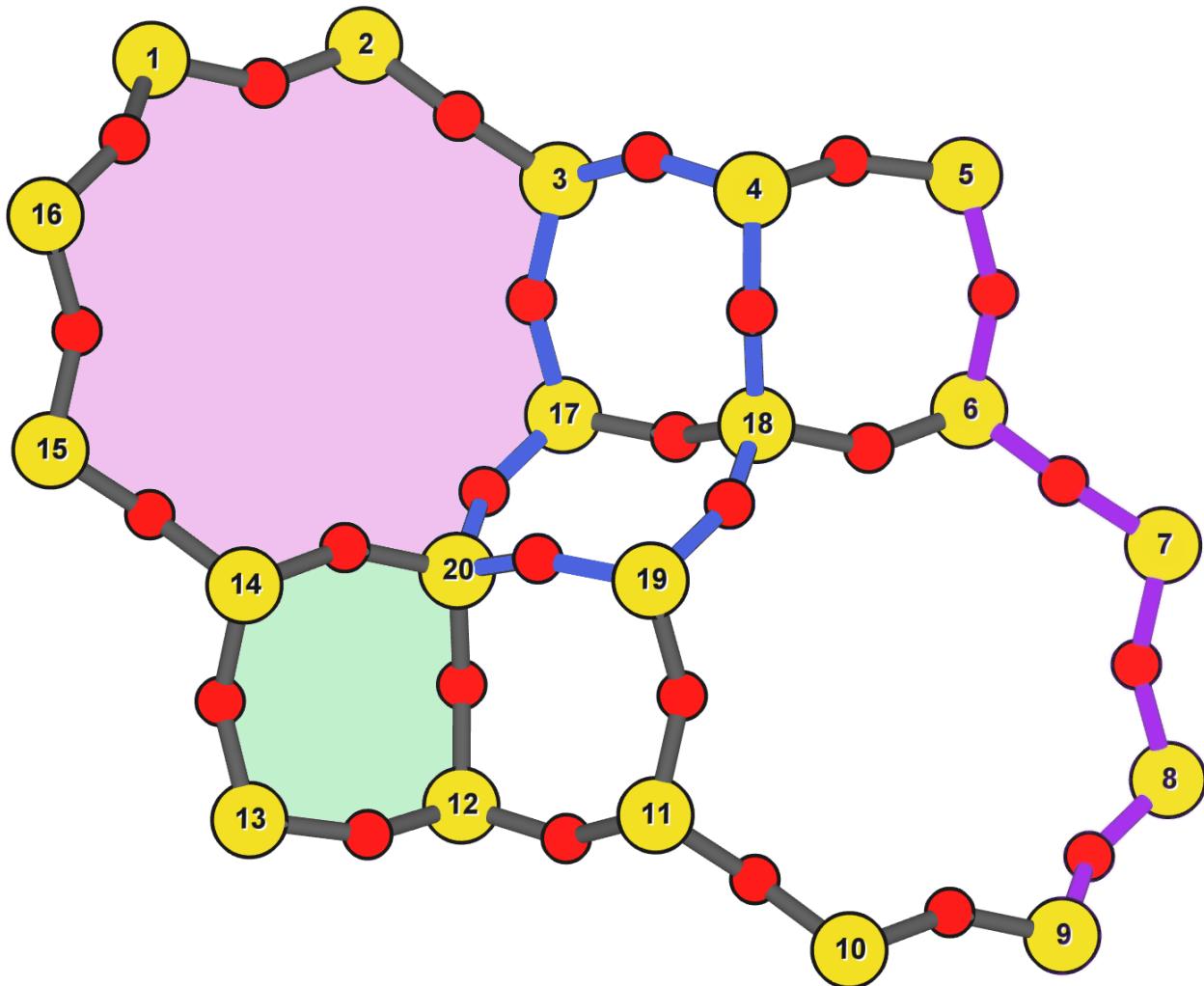


Figure 1: Cutout of the Chabazite framework showing a path (5-6-7-8-9) highlighted with purple bonds, a cycle (3-4-18-19-20-17) highlighted with blue bonds, an 8-MR filled in with pink, and a 4-MR filled in with green. Yellow atoms are Si (T-sites), and red atoms are oxygen.

Table 1: List of graph based features, their descriptions, and how they apply to frameworks, T-sites, and oxygen atoms.

Feature	Description	Application to Frameworks	Application to T-sites	Application to Oxygen Atoms
Node	T-site or oxygen atom	Contains some set of symmetry distinct nodes		
Path	A sequence of nodes that are connected to each other	Traverses through a framework	T-sites and oxygen atoms connected in series forms a path	T-sites and oxygen atoms connected in series forms a path
Cycle	A path that starts and ends at the same node, and does not repeat any other nodes	Contains some set of cycles	Formed by alternating connected T-sites and oxygen atoms	Formed by alternating connected T-sites and oxygen atoms
Rings	A cycle that does not contain a shortcut ^{8,9}	Can describe channels of a framework, or other distinct void environments	Can be described by counting the number of rings that pass through it	Can be described by the number of rings that pass through it
Modified Shortcut Rings	A cycle that does not contain a modified shortcut	A subset of the rings will be found in the framework	Used to describe a T-site	Used to describe an oxygen atom
Shortest Path Rings	A ring that is the shortest ring for at one least one set of O-T-O along the cycle	Another subset of the rings of a framework	Used to describe a T-site	Used to describe an oxygen atom
Vertex Symbol Rings	The rings making up the vertex symbol of a T-site	Yet another subset of the rings of the framework	Used to describe a T-site	Not applicable to oxygen atoms

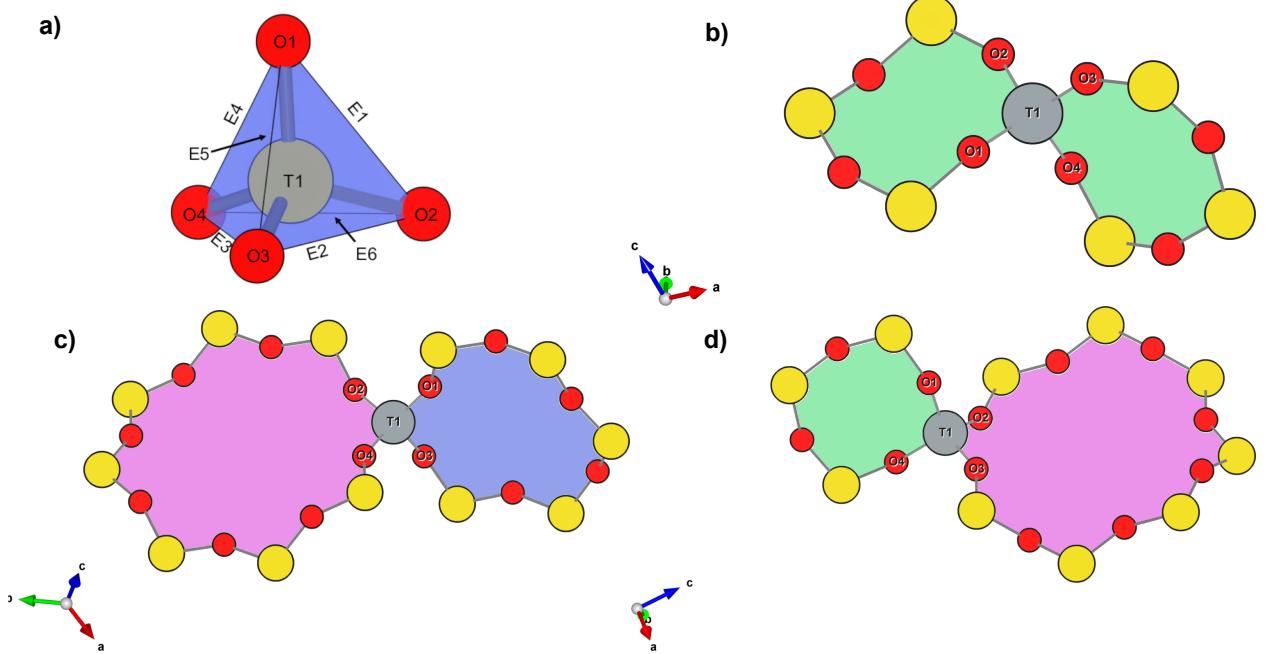


Figure 2: Cutout of the CHA framework, showing the rings that make up the vertex symbol of the single symmetry distinct T-site. a) Example of the tetrahedron formed by the T-site and four connected oxygen atoms, with labeled edges of the tetrahedron. b) Rings associated with opposite edges E1 and E3. c) Rings associated with opposite edges E5 and E6. d) Ring associated with opposed edges E2 and E4. Rings are colored as: 4-MR (green), 6-MR (blue), and 8-MR (pink).

labeled T1 atom. The 12-MR would not be included because for every O-T-O along the 12-MR, the shortest path connecting them is not the 12-MR. The difference between this method and the vertex symbol rings is subtle, but with this shortest path convention any ring belonging to the vertex symbol of any T-site in a framework will be included in the ring count for each of the T-sites that ring passes through. Fig. 4 shows a cutout of the TON framework including a 6- (blue) and 10-MR (orange). For T1, only the 6-MR would be counted in the vertex symbol because it is the shortest path connecting O2 and O3. The 10-MR is part of the vertex symbol for T3 because it is the shortest path connecting O2 and O14. Since this 10-MR is the shortest path for at least one set of O-T-O along the ring, and passes through T1, it does get counted in the shortest path rings for T1.

We can use ring counts to characterize entire zeolite frameworks, T-sites that make up these frameworks, or even the oxygen atoms that connect the T-sites. Since various conventions exist that can reduce the set of rings in a zeolite to more strictly defined properties, the ring counts returned by the various conventions will differ. Differences in ring counts leads to differences in how we might describe the topological environment of a zeolite. Therefore, when using rings to determine the properties of a framework, T-site, or oxygen atom, it is important to know the difference in the conventions, and use one that determines the features of interest.

Here we present the Zeolite Simulation Environment (ZSE, <https://www.github.com/jtcrum/zse>), a Python package that implements the ring finding algorithm presented by Goetzke and Klein⁸ to find rings up to a user defined cutoff size, and can implement the previously published ring set reduction conventions. We use ZSE to provide an analysis of rings returned from each conventions on the entire set of IZA Database of zeolite frameworks¹ to compare how they results in different characterizations. We highlight rings that are found by these conventions but not typically discussed in the literature for a number of frameworks. We also show that the vertex symbol, a common approach used to characterize T-sites misses important parts of the stereochemistry around a T-site.

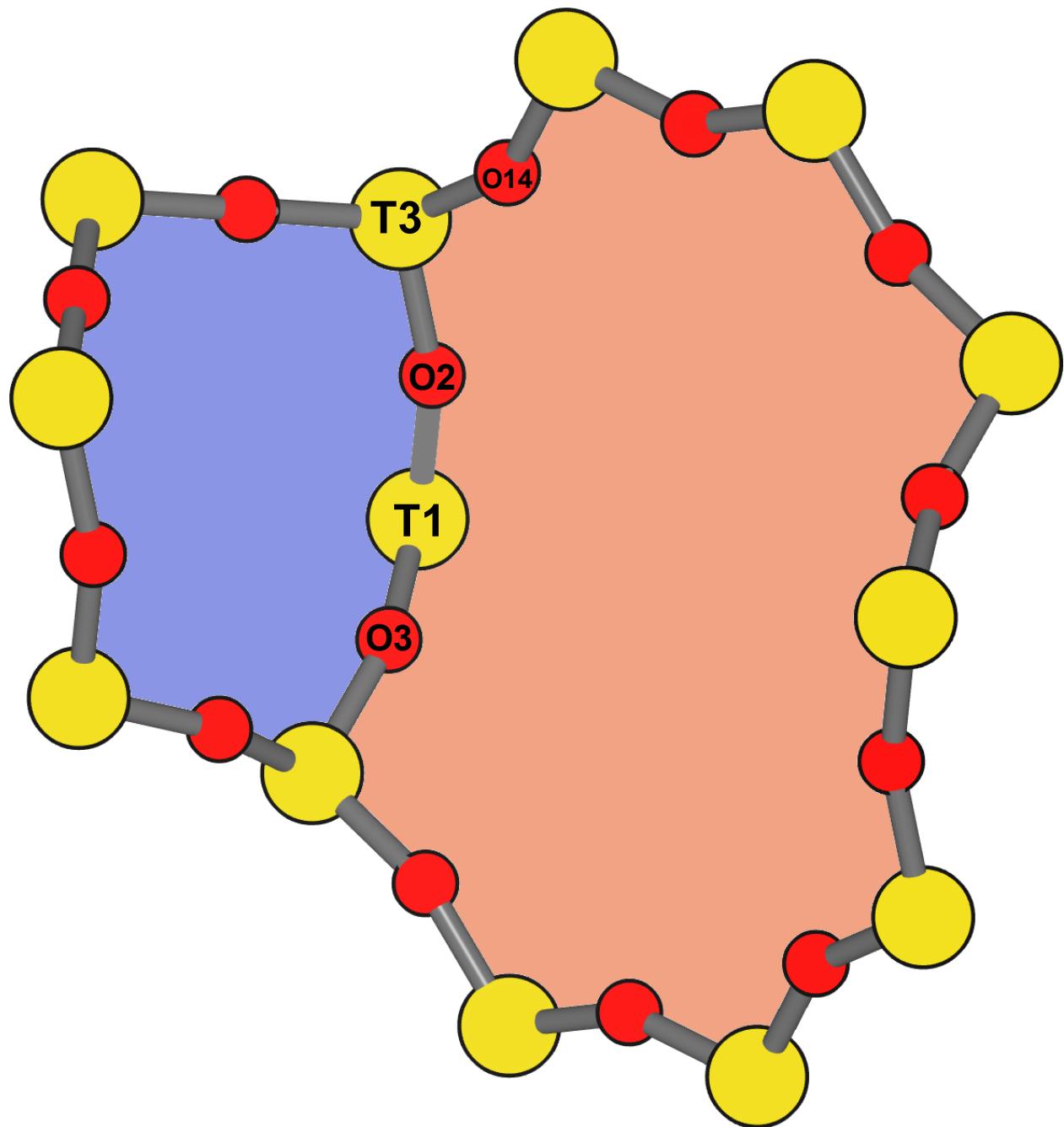


Figure 4: Cutout of the TON framework showing a 6- (blue) and 10-MR (orange). The 10-MR is the shortest path connecting O14-T3-O2, and passes through T1, so it is counted in the shortest path rings for T1.

2 Methods

2.1 Finding Rings That do not Contain Shortcuts

In this work we implement an efficient algorithm that was presented by Goetzke and Klein to find all the rings associated with a T-site that do not contain a shortcut⁸ in a Python package called the Zeolite Simulation Environment (ZSE). In ZSE we use the framework put in place by the Atomic Simulation Environment (ASE)¹² to handle routine analysis zeolite crystal structures. All graph theory functions are performed used the NetworkX Python package.¹⁰

First, we convert the ASE atoms object into a connectivity matrix which represents every atom across the columns and rows. If two atoms are bound together, their respective entry in the connectivity matrix contains a 1, else a 0. This connectivity matrix is then converted to a NetworkX graph object, and then a distance dictionary using NetworkX built in functions. Then we implement Step 3 from Geotzke and Klein's algorithm⁸ summarized here: to find the rings that pass through a T-site, we iteratively search for every size ring between 3-MR and a maximum ring value that is user specified. For this work we set a cutoff of 18-MRs. A schematic showing the evolution of the ring search is shown in Fig. 5. For ring size λ we start at the T-site of interest (labeled 1) Fig. 5, and search the distance matrix for any T-sites that are $\lambda/2$ (even λ) or $(\lambda-1)/2$ (for odd λ) distance from the starting T-site (labeled 2). Next we attempt to create to distinct paths from $1 \rightarrow 2$ and from $2 \rightarrow 1$ alternating adding a node to each path as indicated by FIGURE. Each node added to each of the paths must be $\lambda/2$ (even λ) or $(\lambda-1)/2$ (odd λ) from the head of the other path. Also each node added to each path needs to be the correct distance from 1 and 2 for the given step respectively. If either of the previous two conditions are not met, a ring cannot be formed of length λ along the given paths, we backtrack and repeat until all possible options have been explored for λ . Then we increase λ and continue until the cutoff ring size is completed.

2.2 Finding Vertex Symbol Rings

Starting from the set of all rings found in Section 2.1, we can prune the ring list to the set of vertex symbol rings. We find the shortest ring in the set that connects each pair of oxygens bound to our initial T-site. It is possible for there to be multiple rings of the same size connecting each oxygen, in which case all the rings of that size are kept.

2.3 Finding Shortest Path Rings

Here we prune the set of all rings from Section 2.1 to a subset of rings the meets the shortest path definition published by Sastre and Corma.¹⁹ For each ring, we iterate over every group of O-T-O atoms in the ring, and check if this ring is the shortest path connecting the two oxygen atoms. If so, the loop is broke, because the ring need only be the shortest path for one group of O-T-O atoms to fit the definition. This is the most time consuming process out of all the ring finding conventions we have implemented.

2.4 Modified Shortcut Rings

In this work we present a modified definition of shortcut to capture a different subset of rings from any of the other ring finding conventions. Traditionally a shortcut is a path connecting two nodes of a cycle that decomposes the cycle into smaller rings.^{8,9} We propose that a shortcut is a path connecting two nodes of a cycle that decomposes the cycle into at least one smaller ring. This definition does not require that the shortcut between two nodes be shorter than both paths connecting those nodes along the cycle. Our shortcut only needs to be shorter than one of the

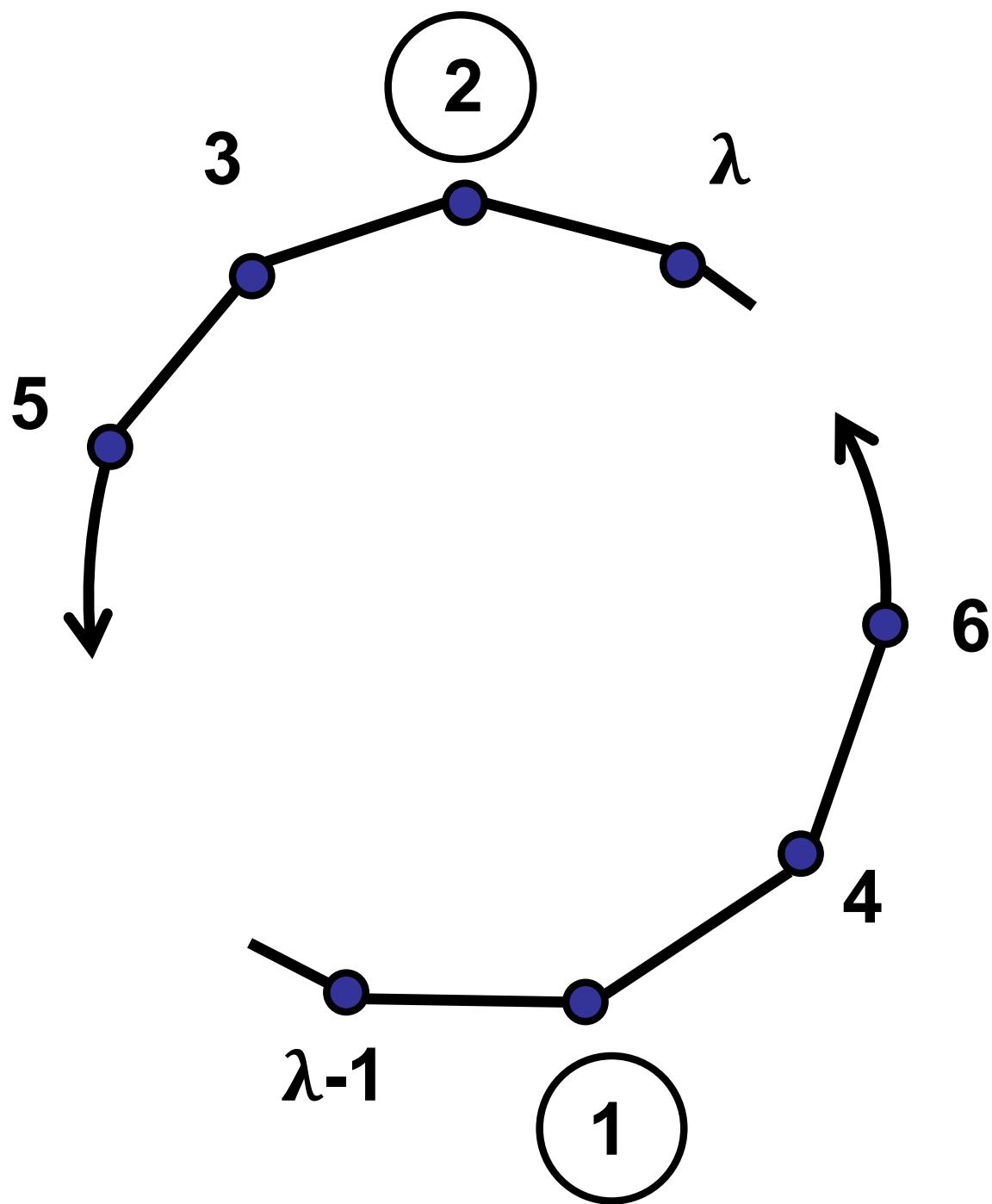


Figure 5: Diagram showing how the ring finding algorithm evolves. Adapted from Goetzke and Klein.⁸

paths connecting those two nodes along the cycle.

This definition is explained graphically in Fig. 6, where we present a cutout of the AFI framework showing a portion of the 12-MR channel. There is a 14-MR that traverses seven T-sites in each of the 12-MRs, through the combination of Path 1 (blue) and Path 2 (purple). With the classical definition of a shortcut, this cycle is considered a ring. However, Path 3 (in green) connecting T1 and T2 produces a 12-MR when combined with Path 1, making Path 3 a modified shortcut. This 14-MR would not be counted under our new definition.

To remove rings containing modified shortcuts from the full set of rings, we iterate over every T-site pair of the ring and check for the shortest path connecting them. If that shortest path is shorter than either of the two paths along the ring connecting the two T-sites, we check if the combination of this shorter path and the shorter of the two ring paths forms a new smaller ring. If so the iteration is broken, and the ring is removed from the counted set.

2.5 Ordered Vertex Symbols

To add information about the spatial orientation of the rings around a T-site to the vertex symbol, we have developed a method to order the edges in the vertex symbol. We systematically list the rings by following the edges of the tetrahedron such that each ring listed is connected to one of the oxygens of the next ring listed. After removing all the rings that are not a part of the vertex symbol (Section 2.2) we use the following process to order them.

1. List all the possible arrangements of the oxygens bound to the T-site ($4! = 24$ possible arrangements).
2. Use a predetermined order of edges: $[[0,2],[0,1],[1,2],[2,3],[3,0],[1,3]]$.
 - (a) Where each of those values represents the index of the oxygen to use.
3. Find the ring size (and multiplicity) connecting each pair of oxygens in this predetermined order.
4. Make a list of weights, where for each pair of oxygens the weight is the ring size \times multiplicity.
5. Reverse sort the list of all possible oxygen arrangements by the correlating list of weights.
6. Use the first oxygen arrangement coupled with the predetermined edge order to list the rings and multiplicity for each edge.

2.6 Determining All Ring Sizes Contained in a Zeolite Framework

Finally, to determine all the ring sizes exhibited with in a zeolite framework, we take advantage of T-site symmetry. The rings of a framework are made of T-sites, and if two T-sites are symmetrically identical they will have the same set of rings passing through them. Therefore, we only need to find the rings associated with each symmetry distinct T-sites to know all the possible ring sizes within a framework. For example, AFI only contains one symmetry distinct T-site (T1). Using the basic definition of a shortcut, T1 is a part of 4-, 6-, 12-, and 14-MRs when using a cutoff of 18-MR. Every other T-site in the AFI framework is also a T1, thus the only possible ring sizes in AFI are 4-, 6-, 12-, and 14-MRs.

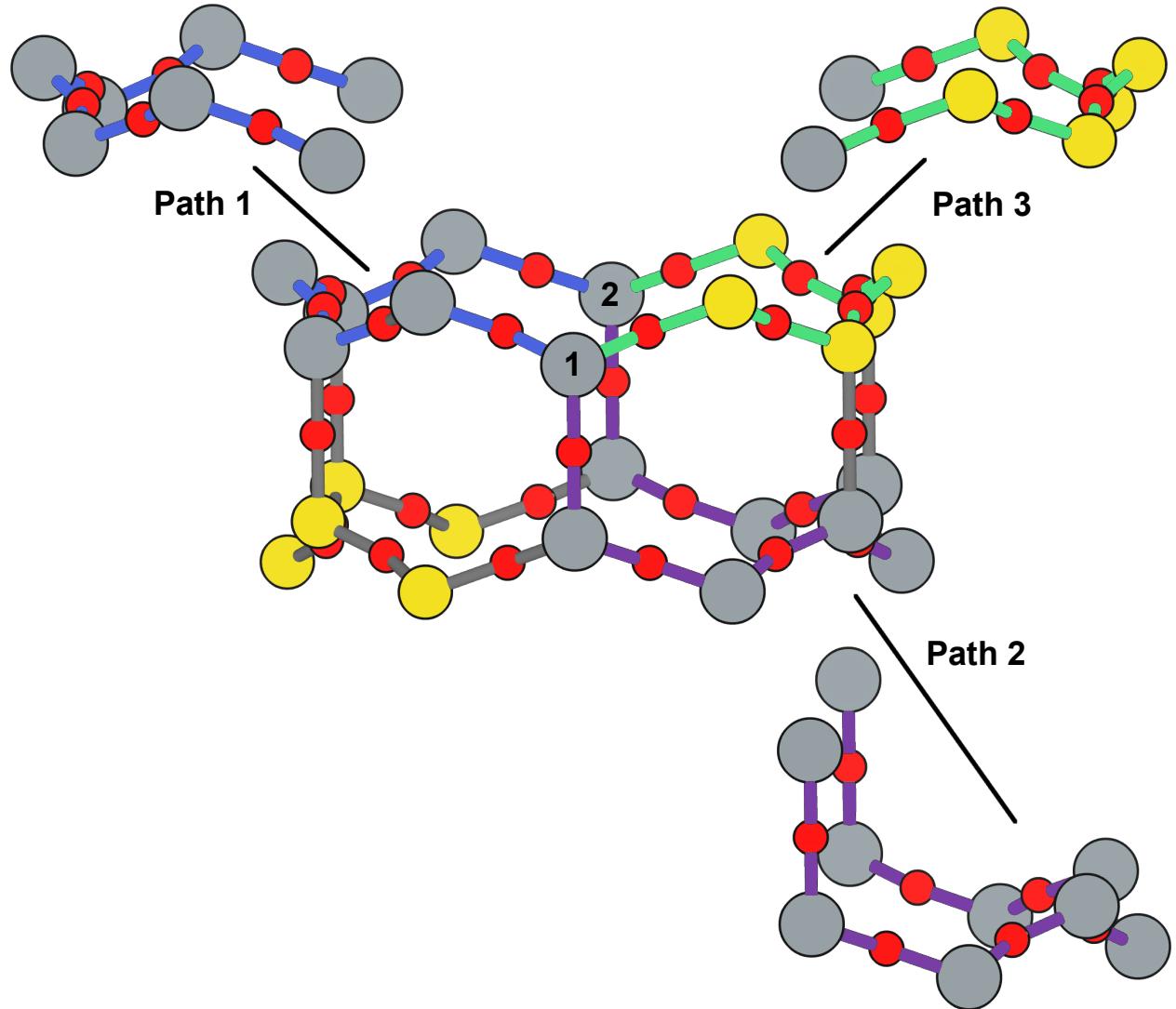


Figure 6: Cutout of the AFI framework showing two stacked 12-MRs from the channel. A 14-MR is shown as T-sites replaced with aluminum atoms in gray. The two paths making connecting Al 1 and Al2 that make this 14-MR are highlighted with blue and purple bonds. Path 3 highlighted with green bonds is a modified shortcut connecting Al1 and Al2.

3 Results

3.1 Characterizing Rings in a Zeolite Graph

The IZA Database¹ is a common reference used to identify all the rings in a zeolite framework, however it only lists the rings that define a channel (ex: 12-MR in AFI), or rings associated with the symbol of a T-site. These rings listed by the IZA are referred to as tabulated rings in the literature.⁴ In some frameworks, other rings (cycles not containing shortcuts) exist that are not included in the list of tabulated rings. These 'untabulated' rings may still provide important topological information about a zeolite framework, or the local void environment around a T-site. Fig. 7 shows counts of frameworks containing each size ring from 3- to 18-MR using the Goetzke algorithm and the listed rings on the IZA database.¹ There are slight differences in the counts up to 6-MRs, but the main divergence takes place as we get to ring sizes >6-MR.

Taking a closer look at some of these untabulated rings, highlights rings not typically listed for some frameworks, but still relevant to describing their topology. Using CHA as an example, Fig. 8 displays a 12-MR (in purple) that exists in CHA that circumferences the cage. This ring is not associated with the vertex symbol of the single symmetry distinct T-site in CHA and does not define a channel. Thus, this ring is not included in the list of tabulated rings. We would argue that this is still a ring that provides an important topological descriptor of CHA because none of the tabulated rings provides information about the size of the CHA cage.

Using AFI as another example, we find another type of ring that arises from traversing a pair of stacked rings and is not included in the list of tabulated rings. AFI, like CHA, contains one symmetry distinct T-site. According to the IZA, the AFI framework contains 4-, 6-, and 12-MRs.¹ When we search for rings using the Goetzke algorithm,⁸ we also find that it contains 14-MRs created by using seven T-sites from two 12-MRs that are separated by a distance of one oxygen (Fig. 9). Rings of this nature are prevalent in many frameworks, another example can be seen in the bottom right of Fig. 8, where an 8-MR is highlighted traversing the two 6-MRs of the D6R. These types of rings may not be of interest depending on which topological feature one intends to describe. This has led us to create a modified definition of a shortcut as explained in METHODS REFERENCE, which excludes these types of rings. The benefit of this new shortcut definition is that larger rings that are missed by the vertex symbol or shortest path rings (i.e., 12-MR in AFI) are still captured, while excluding rings that arise from convolution of staked rings found with the typical shortcut definition.

3.2 Characterizing Frameworks by Rings

With the addition of our modified shortcut definition, we have four ring finding conventions to compare, as well as including the tabulated rings from the IZA Database. Fig. 10 shows how many frameworks contain each size ring found using the various ring counting conventions from 3- to 18-MRs. This plot highlights the differences in the conventions and shows that a topological description of a framework based on rings will depend on the way that you define a ring. In general, a hierarchy of ring sizes found by each convention is: all rings not containing a shortcut >this work >shortest path rings >vertex symbol rings. While the IZA listed rings includes the vertex symbol rings, and a selection of general rings.¹

One drawback to using a ring convention based on connectivity and shortcuts is the exclusion of cycles that don't fit this definition, but still exhibit properties like rings. This is a trade-off between well-defined connectivity rules, and the inclusion of particular void environments that may still have important applications. These shortcut containing cycles can display chemical and/or geometric properties consistent with rings and are of interest to catalysis researchers even though they are

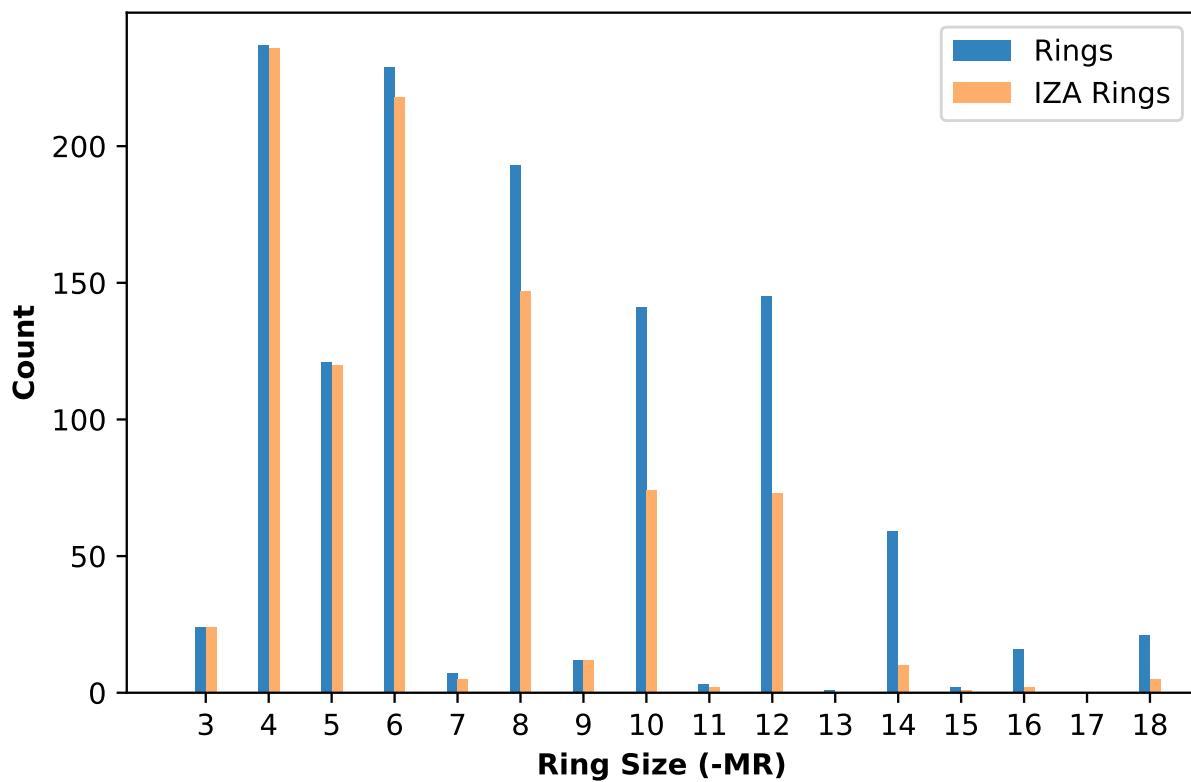


Figure 7: Counts of IZA frameworks containing each size ring between 3- and 18-MR using the Goetzke algorithm and the tabulated rings listed by the IZA.¹

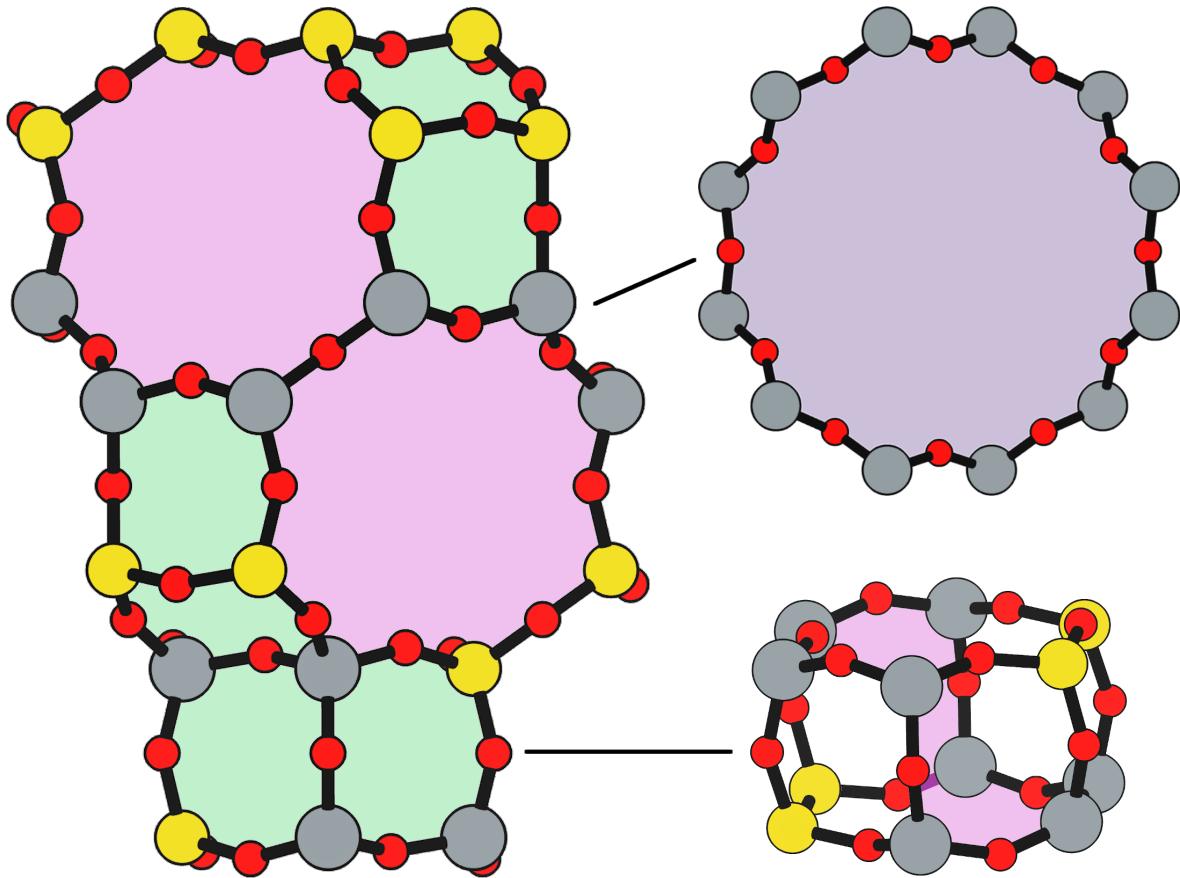


Figure 8: Chabazite cage and double 6-MR (D6R) with highlighted rings: 4-MR in green, 8-MR in pink, and 12-MR in purple. The 8-MR in the D6R and the 12-MR are rings not typically discussed in literature. Si atoms have been replaced with Al atoms to help identify those rings in the overall cage strcutre.

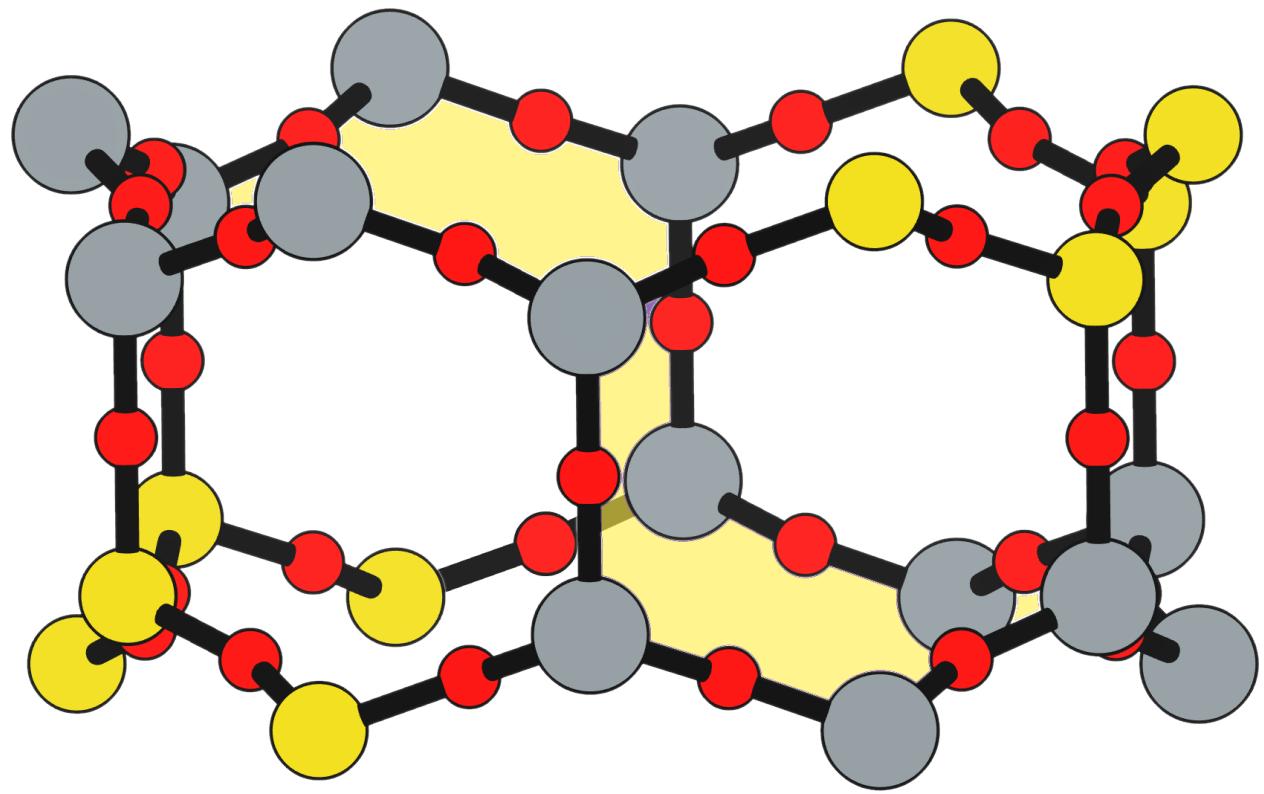


Figure 9: Cutout of the 12-MR channel in AFI with a 14-MR (yellow) traversing seven T-sites of each 12-MR. The T-sites comprising the 14-MR have been replaced with Al for visibility.

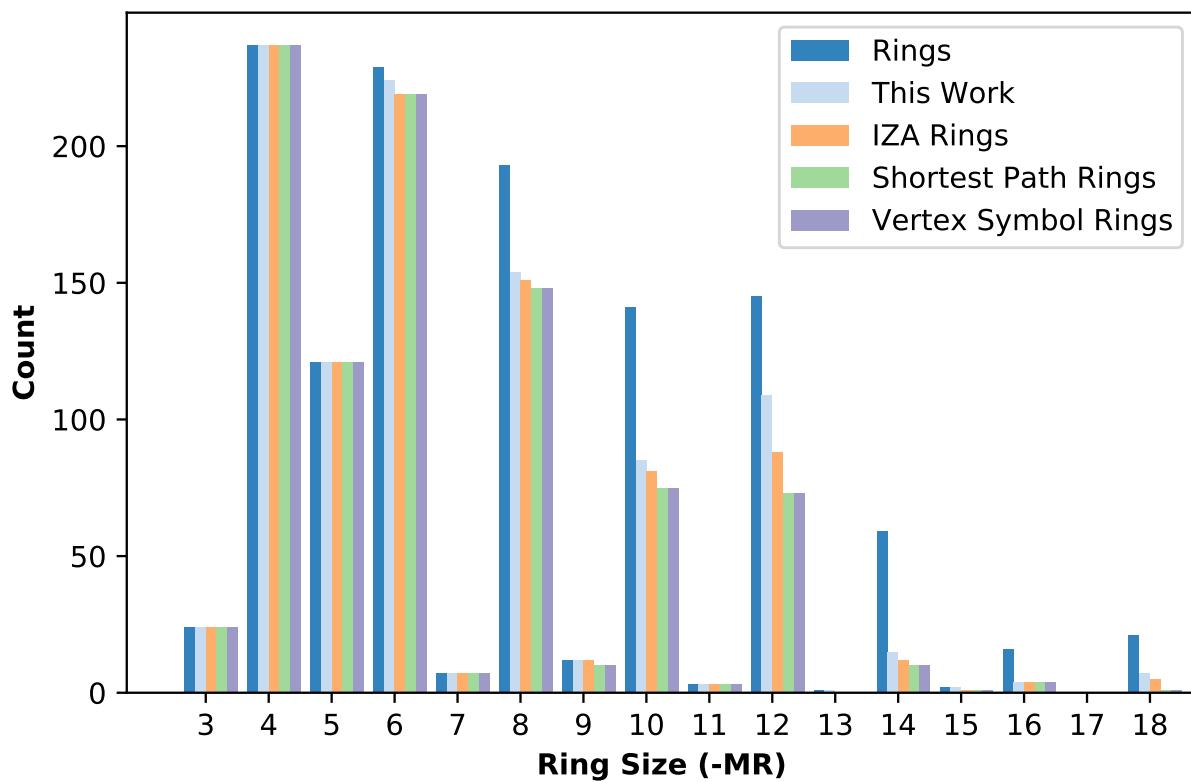


Figure 10: Number of IZA zeolite frameworks containing each size ring, using the various ring counting conventions, as well as the rings listed by the IZA Database.¹

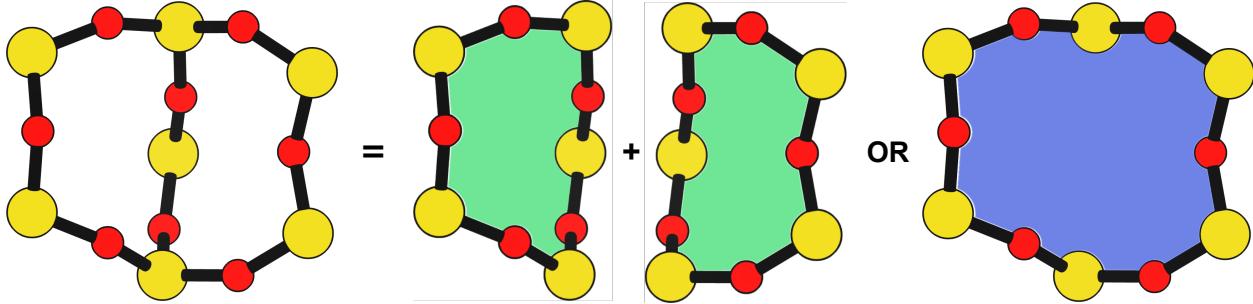


Figure 11: Cutout of MFI framework showing the structure referred to as an α -6-MR in blue, and the two 5-MRs that compose it in green. The 6-membered cycle would not be found by any of the connectivity ring rules outlined in this work.

not classically considered rings. One example is the 6-membered cycle referred to as the α -6-MR in literature (Fig. 11) and is present in a number of frameworks including but not limited to MOR, FER, MFI, and BEA,^{3,5} which is a potential location for Co^{2+} uptake when two Al atoms are 3rd nearest neighbor (NN) in the cycle.^{3,17} Similar to Co^{2+} uptake at 3NN Al atoms in 6-MRs in other frameworks such as CHA.⁶ Fig. 11 shows that this particular structure would be considered two 5-MRs using connectivity rules based on a shortcut.

3.3 Characterizing T-sites by Rings

Considering that zeolite frameworks are comprised of one or more symmetry distinct T-sites, it may be of interest to describe those T-sites by the rings that pass through them. Most often the vertex symbol is used to make such a classification.¹⁸ Sastre and Corma also provided characterization of T-sites using the shortest path rings that pass through them.¹⁹ In their work, they presented the ring index, which lists all the rings passing through a T-site from smallest to largest, and a subscript for each size representing its multiplicity. The rings associated with a T-site can provide information about the local void environments around the T-site, and could potentially be correlated to other physicochemical properties of the T-site once enough once sufficient data on those physicochemical properties exists.

Take for example the AFI framework, containing one symmetry distinct T-site. AFI contains 4-, 6-, 12-, and 14-MRs. To describe that T-site we can count how many of each of those rings pass through the T-site. We can also prune this list using our modified definition of a shortcut, the shortest path rings definition,¹⁹ or the rings contained within the vertex symbol of this T-site.¹⁸ Using the ring index outlined above, each of these conventions will provide a different description of the zeolite (highlighted in Fig. 12):

- Rings: $4\cdot6_{13}\cdot12\cdot14_7$
- This work: $4\cdot6_{13}\cdot12$
- Shortest Path Rings: $4\cdot6_{13}$
- Vertex Symbol Rings: $4\cdot6_{11}$

With an understanding of how we characterize T-site by counting the rings that pass through, Table 2 shows the ring index for a selection of T-sites from uninodal (containing only one symmetry distinct T-site) frameworks. This table highlights the differences in rings counts found with each convention, and shows that in general as you move from left to right across the table the largest ring found decreases. The results in the shortest path column were found using ZSE, but agree

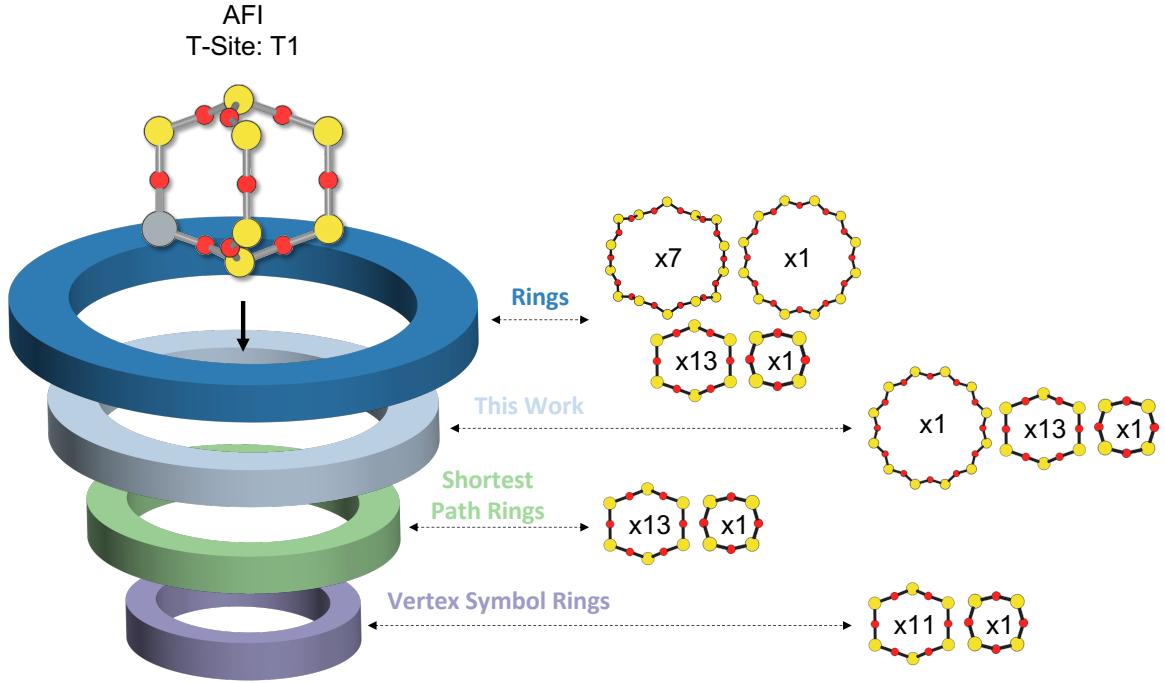


Figure 12: Diagram showing the ring counts of each size ring that pass through the single symmetry distinct T-site in AFI for each of the various ring finding conventions.

directly with the results shown by Sastre and Corma.¹⁹ The results in the vertex symbol rings column were also found with zse, and agree directly with the vertex symbols listed on the IZA Database website.¹

Next, we take an in-depth look at the ring counts for a framework with multiple symmetry distinct T-sites, to show how a ring index can provide information about the local environment around a T-site, and help differentiate them. MOZ is a zeolite framework containing 4-, 6-, 8-, 10-, 12-, 14-, and 18-MRs, 6 symmetry distinct T-sites, and two distinct 12-MR channels. Table 3 shows the ring index for each T-site using each ring finding method.

Fig. 13 shows the T-site locations inside a 2-dimensional view of the framework. If you were interested in which T-sites have access to the 12-MR channels, the shortest path rings and vertex

Table 2: Comparison of ring indices for the T-sites in various uninodal zeolite frameworks.

Framework	Rings	This Work	Shortest Path Rings ¹⁹	Vertex Symbol Rings ¹
ABW	42·63·84	42·63·84	42·63·84	42·63·82
ACO	43·63·86·10 ₁₅	43·86	43·86	43·86
AFI	4 ₁ ·13·12 ₁ ·14 ₇	4 ₁ ·6 ₁₃ ·12 ₁	4 ₁ ·6 ₁₃	4 ₁ ·6 ₁₁
ANA	4 ₂ ·6 ₂ ·8 ₁₆	4 ₂ ·6 ₂ ·8 ₁₆	4 ₂ ·6 ₂ ·8 ₁₆	4 ₂ ·6 ₂ ·8 ₈
ATO	4 ₁ ·6 ₉ ·8 ₈ ·12 ₂₀	4 ₁ ·6 ₉ ·12 ₂₀	4 ₁ ·6 ₉	4 ₁ ·6 ₉
BCT	4 ₁ ·6 ₆ ·8 ₂₀	4 ₁ ·6 ₆ ·8 ₁₂	4 ₁ ·6 ₆	4 ₁ ·6 ₆
CHA	4 ₃ ·6 ₁ ·8 ₆ ·12 ₁	4 ₃ ·6 ₁ ·8 ₂ ·12 ₁	4 ₃ ·6 ₁ ·8 ₂	4 ₃ ·6 ₁ ·8 ₂
DFT	4 ₂ ·6 ₆ ·8 ₁₀ ·10 ₁₀	4 ₂ ·6 ₆ ·8 ₁₀	4 ₂ ·6 ₆ ·8 ₁₀	4 ₂ ·6 ₄ ·8 ₆
GIS	4 ₃ ·8 ₄	4 ₃ ·8 ₄	4 ₃ ·8 ₄	4 ₃ ·8 ₄
GME	4 ₃ ·6 ₁ ·8 ₆ ·12 ₇	4 ₃ ·6 ₁ ·8 ₂ ·12 ₁	4 ₃ ·6 ₁ ·8 ₂	4 ₃ ·6 ₁ ·8 ₂
MER	4 ₃ ·8 ₄ ·10 ₁₀ ·14 ₁₄	4 ₃ ·8 ₄	4 ₃ ·8 ₄	4 ₃ ·8 ₄
MON	4 ₁ ·5 ₅ ·8 ₆	4 ₁ ·5 ₅ ·8 ₆	4 ₁ ·5 ₅ ·8 ₆	4 ₁ ·5 ₄ ·8 ₄
NPO	3 ₁ ·6 ₆ ·12 ₄₀	3 ₁ ·6 ₆ ·12 ₄₀	3 ₁ ·6 ₆	3 ₁ ·6 ₆

been represented in ring index format for ease of comparison.

Vertex symbols have

Table 3: Ring indices for each distinct T-site in the MOZ framework using each ring counting convention.

T-Site	Rings	This Work	Shortest Path Rings	Vertex Symbol Rings	
T1	$4_3 \cdot 6_2 \cdot 8_7 \cdot 10_7 \cdot 18_5$	$4_3 \cdot 6_2 \cdot 8_3$	$4_3 \cdot 6_2 \cdot 8_3$	$4_3 \cdot 6_2 \cdot 8$	
T2	$4_3 \cdot 6_2 \cdot 8_7 \cdot 10_7 \cdot 14_5$	$4_3 \cdot 6_2 \cdot 8_3$	$4_3 \cdot 6_2 \cdot 8_3$	$4_3 \cdot 6_2 \cdot 8$	
T3	$4_3 \cdot 6_2 \cdot 8_5 \cdot 10_4 \cdot 12_4 \cdot 14_5$	$4_3 \cdot 6_2 \cdot 8 \cdot 12_4$	$4_3 \cdot 6_2 \cdot 8$	$4_3 \cdot 6_2 \cdot 8$	Vertex symbols have been
T4	$4_2 \cdot 6 \cdot 8_6 \cdot 10_6 \cdot 12 \cdot 18_2$	$4_2 \cdot 6 \cdot 8_6 \cdot 12$	$4_2 \cdot 6 \cdot 8_6 \cdot 12$	$4_2 \cdot 6 \cdot 8_6 \cdot 12$	
T5	$4_2 \cdot 6 \cdot 8_7 \cdot 10_6 \cdot 14_{18}$	$4_2 \cdot 6 \cdot 8_7$	$4_2 \cdot 6 \cdot 8_7$	$4_2 \cdot 6 \cdot 8_7$	
T6	$4_2 \cdot 6 \cdot 8_3 \cdot 10_2 \cdot 12_8 \cdot 14_{18}$	$4_2 \cdot 6 \cdot 8_3 \cdot 12_8$	$4_2 \cdot 6 \cdot 8_3$	$4_2 \cdot 6 \cdot 8_3$	

represented in ring index format for ease of comparison.

symbol rings would only suggest T3 participates in the 12-MR rings. However, all rings and this work both identify T4 and T6 as participating in the 12-MR channels as highlighted in Fig. 13.

We next used ZSE to find the rings associated with every symmetry distinct T-site in every framework across the IZA Database using each of the four ring counting conventions. For each T-site we used the rings to generate a ring index, and Fig. 14 shows how many unique ring indices are present when using each of the ring counting conventions. The plot follows intuition with the number of unique ring indices decreasing as we use more restrictive ring counting conventions, because less rings are found and provides less room for differentiation. This raises the question, if you want to ascertain chemical or physical properties about a T-site based on its ring count, and differentiate these T-sites from other similar but distinct T-site, which ring counting convention will suffice? The answer will depend on what level of detail is desired. Larger rings can be found with the standard shortcut definition, rings traversing other stacked rings will be excluded with our modified shortcut definition. The shortest path definition and vertex symbol rings will provide the most localized information about a T-site.

To further compare the ring counting conventions, we show a distribution of the number of T-sites containing each size ring between 3- and 18-MR in Fig. 15 (right). This plot highlights that more T-sites contain larger sized rings when using the basic definition of a shortcut, and at smaller rings sizes (<6-MR) all the ring counting conventions return the same results. To further emphasize this results, we have provided a cumulative distribution of the same data normalized to the maximum 'rings' value in Fig. 15. At 6-MRs is where we see the cumulative distribution functions deviate from each other. The largest deviation takes place at 12-MRs, and the cumulative distributions start to level out at larger ring sizes.

To complete the comparison of ring counting conventions, we have developed method to determine how similar the ring counts of one convention are to another. We do this by comparing the ring index for a T-site using each convention, where the similarity of the ring indices is scored with Eq. (1). In this equation, sr is the number of similar rings that are found in both counting conventions, and mr is the maximum number of rings found by either convention. For example: the ring index of AFI using the classic shortcut definition and the shortest path definition are $4 \cdot 6_{13} \cdot 12 \cdot 14_7$ and $4 \cdot 6_{13} \cdot$. The number of similar rings found by both conventions is 14, and the maximum number of rings found by either convention is 22. This would lead to a similarity score of 0.636. We do this for every T-site between two conventions and average the similarity score to get the results in Fig. 16. Down the diagonal each method is compared to itself and clearly has a similarity of 1. The remainder of the table follows intuition, in that the most restrictive ring counting convention (vertex symbol rings) compare to the least restrictive convention (rings) has the lowest similarity score. The two most similar ring counting methods are our modified shortcut definition and the shortest path rings.

$$s = \frac{sr}{mr} \quad (1)$$

One final point we would like to make about T-site characterization, is that the vertex symbol and

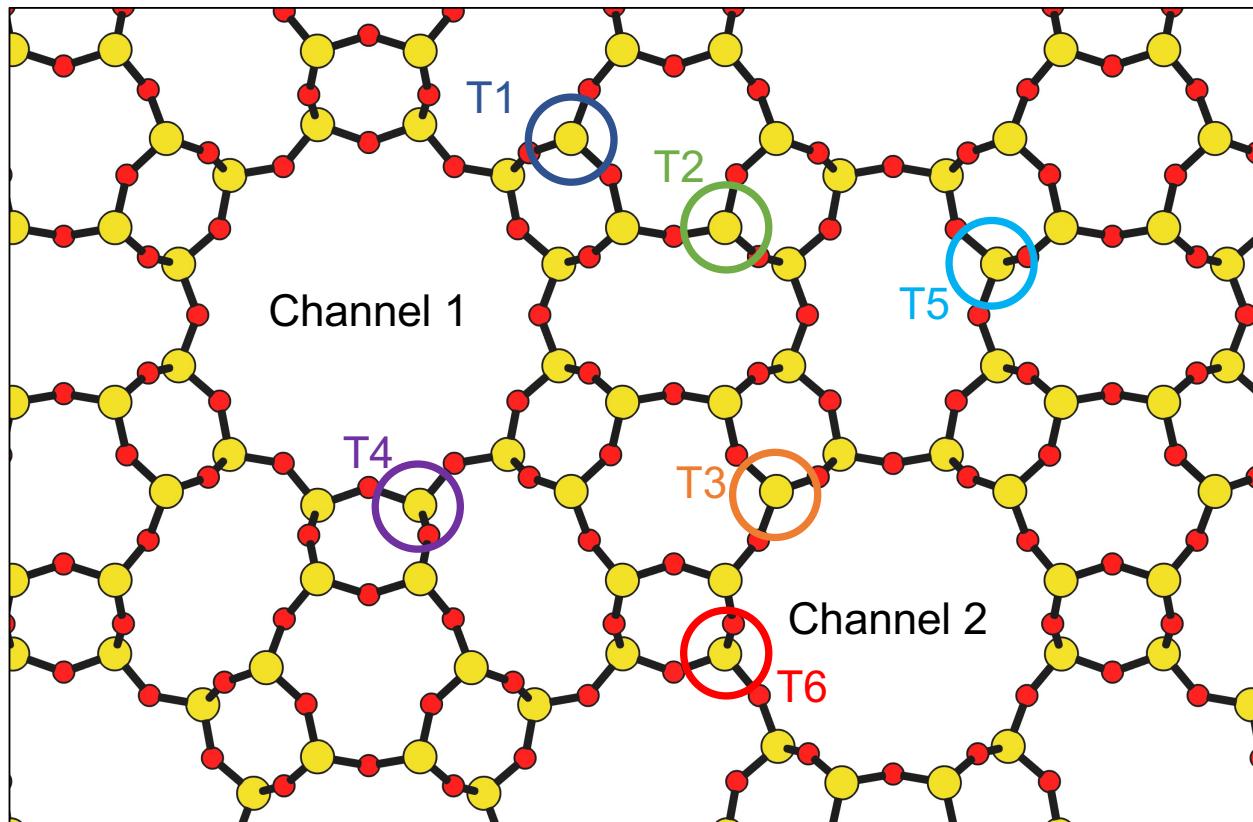


Figure 13: Cutout of the MOZ framework showing two 12-MR channels, with an example of each distinct T-site highlighted. T1: navy, T2: green, T3: orange, T4: purple, T5: blue, and T6: red. As shown, T3, T4, and T6 are all associated with the 12-MR channels, while T1, T2, and T5 are not connected to the 12-MR channels.

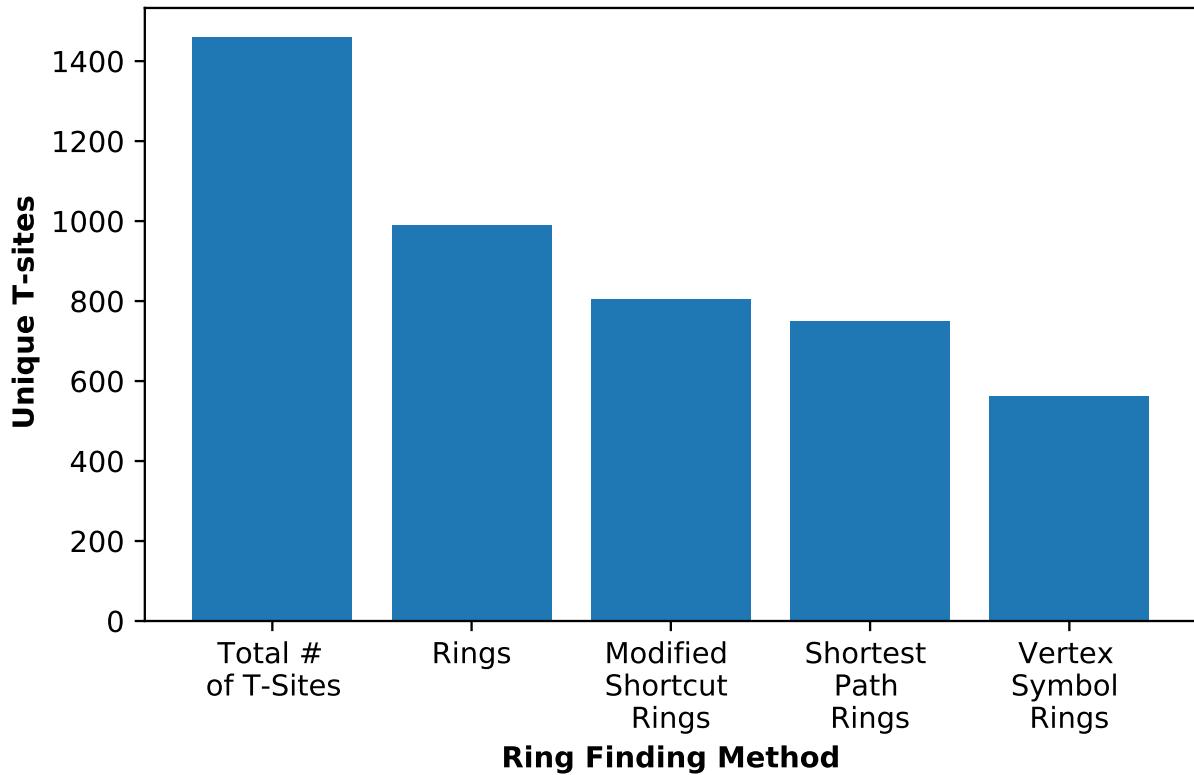


Figure 14: Number of unique T-site ring indices when classified by the rings passing through them using the various ring counting conventions. There are 1460 T-sites across all the frameworks in the IZA Database. As we move from less restrictive to more restrictive (left to right) ring counting conventions, the number of unique ring indices decreases.

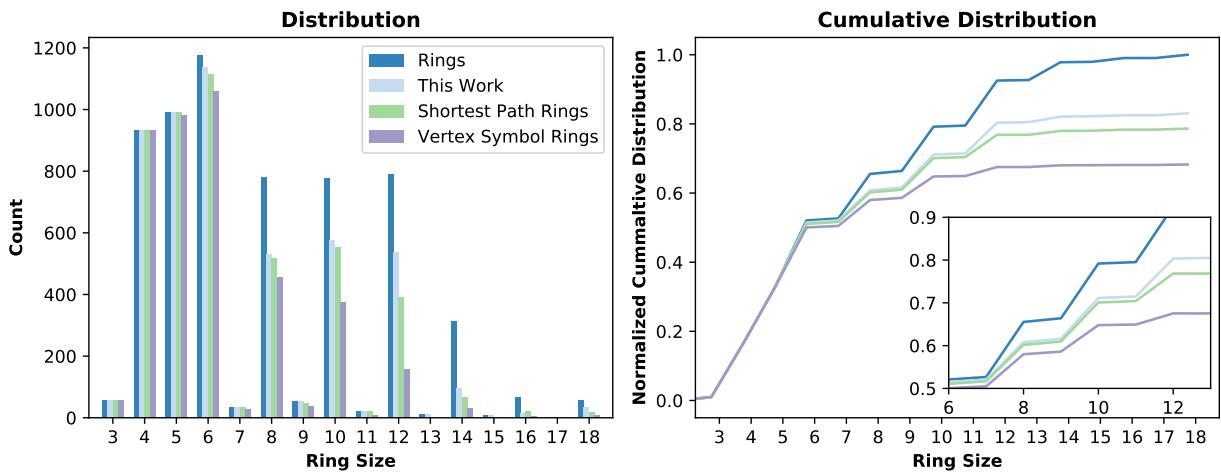


Figure 15: Frequency of T-sites across all IZA frameworks containing ring sizes between 3- and 18-MR (left), and cumulative distribution of T-sites containing each ring size normalized to the final 'rings' value (right).

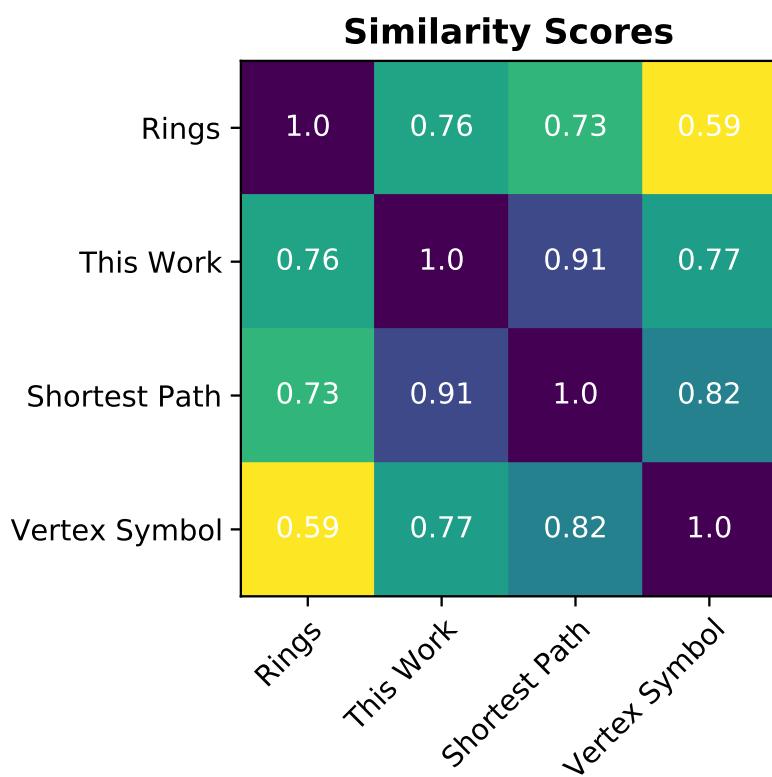


Figure 16: Heat map showing the similarity score for four ring counting methods. Similarity score of 1 means identical set of rings returned, while a similarity of 0 would mean no matching rings are returned.

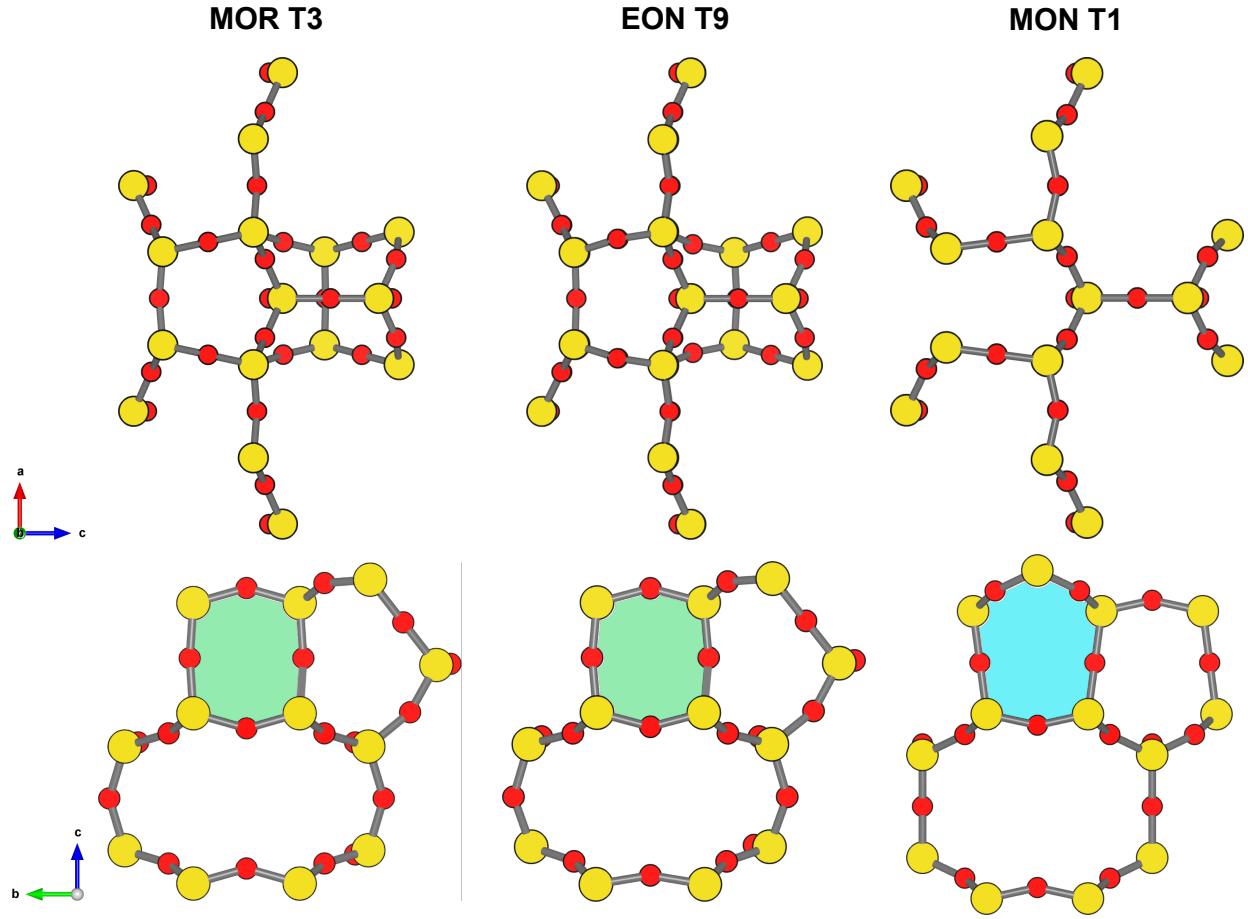


Figure 17: Cutout of the MOR, EON, and MON frameworks that only shows the rings associated with the vertex symbol of T3, T9, and T1 respectively. The 4-MR (green) and $2 \times$ 5-MR (teal) that are in swapped positions are highlighted for emphasis. The 4-MR for MON, and the 25-MRs for MOR and EON are into the plane, and not easily shown.

ring indices only provide information about rings, and do not give any information about the spatial orientation of those rings around the T-site. The vertex symbol comes close to accomplishing that, by finding rings associated with opposite edge pairs of the tetrahedron, and then listing those pairs from smallest to largest. However, the vertex symbol does not capture subtle but distinct differences in the orientation of the rings around the T-site that can lead to varying local void environments. For example: MOR T3, MON T1, and EON T9 all have the same vertex symbol of: $45_2 \cdot 5 \cdot 8_2 \cdot 5 \cdot 8_2$. However, the orientation of those rings around each of those T-sites are not identical. Fig. 17 shows a cutout of each of these frameworks that only includes the atoms that make up the rings of the vertex symbol around the specified T-sites. We can see that MOR T3 and EON T9 have the same ring orientation, and that orientation is different from the rings making up the vertex symbol of MON T1. The main difference is the location of the 5_2 - and 4-MR edges. They are highlighted in Fig. 17.

The structural differences shown in Fig. 17 that are not able to be captured by the vertex symbol leads us to believe that the vertex symbol is not a complete descriptor, and there is room to define a new descriptor that takes into consideration ring orientation. This has led us to create a new method for listing the rings in the vertex symbol that considers the structural connection of the rings.

Table 4: List of three vertex symbols, the T-sites associated with them, and the representative ordered vertex symbol for those T-sites.

Vertex Symbol	Framework T-site	Ordered Vertex Symbol
5·5·5·5 ₂ ·5·10	EWS T3	10·5 ₂ ·5·5·5·5
	ITN T9	10·5 ₂ ·5·5·5·5
	ITN T21	5 ₂ ·10·5·5·5·5
	OKO T2	10·5 ₂ ·5·5·5·5
	OKO T5	5 ₂ ·10·5·5·5·5
	PCS T2	10·5 ₂ ·5·5·5·5
	PCS T3	5 ₂ ·10·5·5·5·5
	SFS T10	5 ₂ ·10·5·5·5·5
	SFV T3	5 ₂ ·10·5·5·5·5
	SFV T7	5 ₂ ·10·5·5·5·5
4·5 ₂ ·5·8·5·8	TUN T10	10·5 ₂ ·5·5·5·5
	DAC T3	5 ₂ ·8·8·5·5·4
	DAC T4	5 ₂ ·8·8·5·5·4
	EON T10	5 ₂ ·8·8·5·5·4
	EPI T1	5 ₂ ·8·8·5·5·4
	MOR T4	5 ₂ ·8·8·5·5·4
	RSN T4	5 ₂ ·8·5·5·8·4
	VNI T2	5 ₂ ·8·5·5·8·4
	VSV T2	5 ₂ ·8·5·5·8·4
4·6·4·6·6·8	YFI T9	5 ₂ ·8·8·5·5·4
	ATN T1	8·6·4·4·6·6
	JSN T3	8·6·6·4·4·6
	PON T1	8·6·4·4·6·6
	SAS T1	8·6·4·4·6·6
	ZON T3	8·6·6·4·4·6

DEFINE ORDERED VERTEX SYMBOL HERE

With this new descriptor MOR T3 and EON T9 would be labeled as: 8₂·8₂·5₂·5·4·5, and MON T1 as: 8₂·8₂·4·5·5₂·5. The difference is subtle but highlights the distinct structural difference between the two types of T-sites that is not otherwise captured by a vertex symbol. Stereochemistry of the rings associated with a T-site could influence the chemical properties we care about such as deprotonation energy, T-site substitution energy, or catalytic properties of reactions happening at that T-site.

We have used this new ordered vertex symbol to characterize all 1460 T-sites in the IZA Database. We found that using the standard vertex symbol to characterize T-sites there are 649 unique vertex symbols present. In contrast, there are 666 unique ordered vertex symbols across every T-site. This would imply that not a large amount of vertex symbols contain stereochemical differences. In Table 4 we provide a list of some common T-site vertex symbols, and their representative ordered vertex symbols.

4 Conclusions

Rings of a graph are well defined; here we identified all rings up to 18-MR in every zeolite framework listed on the IZA Structure Database.¹ We find that the commonly reported ring sizes in literature and on the IZA website leave out many rings that fit the classical definition of a cycle that does not contain a shortcut. To completely describe the topology of a zeolite these rings are required, however there are often cases where someone might want to consider only a subset of rings of interest.

We have shown a comparison of three different conventions used to count rings, and highlighted the differences in rings that are found by each convention. The classic definition of a ring identifies the largest set of ring sizes across all the zeolite frameworks, while the shortest path ring and vertex symbol rings only identify smaller ring sizes. We have provided a modified definition of a shortcut

that when used to find rings still finds larger rings defining channel openings, but excludes rings that are able to be decomposed into at least one smaller ring. It is important to understand the difference of ring sizes and types found by each convention when discussing the rings of a zeolite framework. A disadvantage to using purely connectivity based definitions of rings is the exclusion of cycles in a framework that behave physicochemically like a ring but contain a shortcut. We have displayed an example case of these geometric rings, and in the future it would be beneficial to develop a computation method of identifying these cycles.

This same methodology was used to describe T-sites in zeolite frameworks by counting all of the rings that pass through the T-site using each of the ring counting conventions described. When using ring finding conventions that find larger rings, we see more diversity in the descriptions of T-sites, which can aid researchers who want to identify similar T-sites across multiple frameworks. We have also shown that the vertex symbol used to describe T-sites leaves out subtle but distinct stereochemical differences in the spatial orientation of the rings around a T-site. To address this shortcoming we have provided a new method for ordering the rings of a vertex symbol that takes into consideration the ring stereochemistry and is able to identify differences in T-sites that have the same vertex symbol. In the future, correlating physicochemical properties of T-sites to the ring descriptors identified with each ring counting convention can help identify sets of frameworks with desired T-site properties.

5 Acknowledgments

We acknowledge financial support provided by the National Science Foundation under Cooperative Agreement No. EEC-1647722, which is an Engineering Research Center for the Innovative and Strategic Transformation of Alkane Resources. JTC thanks the Arthur J. Schmitt Foundation for financial aid in the form of a PhD fellowship. We thank Dr. Christian Baerlocher for the numerous discussions about the topology of zeolites, and methodologies used by the IZA Structure Database. We thank Dr. German Sastre for providing a copy of zeoTsites to compare results, and for the helpful conversions about use cases and zeolite topology. This research was supported in part by the Notre Dame center for Research Computing through access to high performance computing clusters.

References

- [1] C. Baerlocher and L. McCusker. Database of Zeolite Structures. URL <http://www.iza-structure.org/databases/>.
- [2] D. Bermúdez and G. Sastre. *Theoretical Chemistry Accounts*, 136(10):116, Oct. 2017. ISSN 1432-2234. doi: 10.1007/s00214-017-2143-6. URL <https://doi.org/10.1007/s00214-017-2143-6>.
- [3] M. Bernauer, E. Tabor, V. Pashkova, D. Kaucký, Z. Sobalík, B. Wichterlová, and J. Dedecek. *Journal of Catalysis*, 344:157–172, Dec. 2016. ISSN 00219517. doi: 10.1016/j.jcat.2016.09.025. URL <https://linkinghub.elsevier.com/retrieve/pii/S002195171630197X>.
- [4] R. A. Curtis and M. W. Deem. *Journal of Physical Chemistry B*, 107:8612–8620, 2003.
- [5] J. Dedeček, Z. Sobalík, and B. Wichterlová. *Catalysis Reviews*, 54(2):135–223, Apr. 2012. ISSN 0161-4940, 1520-5703. doi: 10.1080/01614940.2012.632662. URL <http://www.tandfonline.com/doi/abs/10.1080/01614940.2012.632662>.
- [6] J. R. Di Iorio, S. Li, C. B. Jones, C. T. Nimlos, Y. Wang, E. Kunkes, V. Vattipalli, S. Prasad, A. Moini, W. F. Schneider, and R. Gounder. *Journal of the American Chemical Society*, 142

- (10):4807–4819, Mar. 2020. ISSN 0002-7863, 1520-5126. doi: 10.1021/jacs.9b13817. URL <https://pubs.acs.org/doi/10.1021/jacs.9b13817>.
- [7] D. S. Franzblau. *Physical Review B*, 44(10):4925–4930, Sept. 1991. ISSN 0163-1829, 1095-3795. doi: 10.1103/PhysRevB.44.4925. URL <https://link.aps.org/doi/10.1103/PhysRevB.44.4925>.
- [8] K. Goetzke and H.-J. Klein. *Journal of Non-Crystalline Solids*, 127(2):215–220, Feb. 1991. ISSN 00223093. doi: 10.1016/0022-3093(91)90145-V. URL <https://linkinghub.elsevier.com/retrieve/pii/002230939190145V>.
- [9] L. Guttmann. *Journal of Non-Crystalline Solids*, 116(2-3):145–147, Feb. 1990. ISSN 00223093. doi: 10.1016/0022-3093(90)90686-G. URL <https://linkinghub.elsevier.com/retrieve/pii/002230939090686G>.
- [10] 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*, pages 11–15, Pasadena, CA USA, Aug. 2008.
- [11] B. A. Helfrecht, R. Semino, G. Pireddu, S. M. Auerbach, and M. Ceriotti. *The Journal of Chemical Physics*, 151(15):154112, Oct. 2019. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.5119751. URL <http://aip.scitation.org/doi/10.1063/1.5119751>.
- [12] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen. *Journal of Physics: Condensed Matter*, 29(27):273002, July 2017. ISSN 0953-8984, 1361-648X. doi: 10.1088/1361-648X/aa680e. URL <http://stacks.iop.org/0953-8984/29/i=27/a=273002?key=crossref.20f9751653d872507bf6c0cc5737032c>.
- [13] P. M. Kester, J. T. Crum, S. Li, W. F. Schneider, and R. Gounder. *Journal of Catalysis*, 395:210–226, Mar. 2021. ISSN 00219517. doi: 10.1016/j.jcat.2020.12.038. URL <https://linkinghub.elsevier.com/retrieve/pii/S0021951721000191>.
- [14] S. Le Roux and P. Jund. *Computational Materials Science*, 49(1):70–83, June 2010. ISSN 09270256. doi: 10.1016/j.commatsci.2010.04.023. URL <https://linkinghub.elsevier.com/retrieve/pii/S0927025610002363>.
- [15] S. Li, H. Li, R. Gounder, A. Debellis, I. B. Müller, S. Prasad, A. Moini, and W. F. Schneider. *The Journal of Physical Chemistry C*, 122(41):23564–23573, Oct. 2018. ISSN 1932-7447, 1932-7455. doi: 10.1021/acs.jpcc.8b07213. URL <https://pubs.acs.org/doi/10.1021/acs.jpcc.8b07213>.
- [16] X. Li and M. W. Deem. *The Journal of Physical Chemistry C*, 118(29):15835–15839, July 2014. ISSN 1932-7447, 1932-7455. doi: 10.1021/jp504143r. URL <https://pubs.acs.org/doi/10.1021/jp504143r>.
- [17] C. T. Nimlos, A. J. Hoffman, Y. Gul Hur, B. Jin Lee, J. R. Di Iorio, D. Hibbitts, and R. Gounder. *Chemistry of Materials*, 2020. doi: <https://doi.org/10.1021/acs.chemmater.0c03154>.
- [18] M. O’Keeffe and S. Hyde. *Zeolites*, 19(5-6):370–374, Nov. 1997. ISSN 01442449. doi: 10.1016/S0144-2449(97)00133-4. URL <https://linkinghub.elsevier.com/retrieve/pii/S0144244997001334>.

- [19] G. Sastre and A. Corma. *The Journal of Physical Chemistry C*, 113(16):6398–6405, Apr. 2009. ISSN 1932-7447, 1932-7455. doi: 10.1021/jp8100128. URL <https://pubs.acs.org/doi/10.1021/jp8100128>.
- [20] G. Sastre and J. D. Gale. *Microporous and Mesoporous Materials*, 43(1):27–40, Mar. 2001. ISSN 13871811. doi: 10.1016/S1387-1811(00)00344-9. URL <https://linkinghub.elsevier.com/retrieve/pii/S1387181100003449>.
- [21] F. Wooten. *Acta Crystallographica Section A: Foundations of Crystallography*, 58(4):346–351, July 2002. ISSN 0108-7673. doi: 10.1107/S0108767302006669. URL <http://scripts.iucr.org/cgi-bin/paper?bk0109>. Number: 4 Publisher: International Union of Crystallography.
- [22] X. Yuan and A. Cormack. *Computational Materials Science*, 24(3):343–360, June 2002. ISSN 09270256. doi: 10.1016/S0927-0256(01)00256-7. URL <https://linkinghub.elsevier.com/retrieve/pii/S0927025601002567>.