1. Research Plan: Geometry of organic chemistry: polytopes and polyhexes.

1.1. Nature of the Research and Significance. Just a few months ago, in November 2011, a team of biochemists from the Scripps Institute began a search for new compounds to combat mutant strains of drug resistant malaria. They started by combining simple rings of carbon and hydrogen–not physically– instead using a computer program. The number of candidate molecules quickly grew into the millions. To process this combinatorial explosion of potential antivirals the researchers have already used more than 5,000 years of CPU time, all donated from home computers belonging to individual members of the IBM World Community Grid.

"Combinatorial" describes objects that are constructed from basic building blocks according to specified rules. Rather than trial and error in the lab, many drug designers start with a virtual catalogue of the structures that can be formed from basic molecular ingredients. This means that the computers list the possibilities in order based on shared structural characteristics in what is known as a *combinatorial library*. Searching such a library can facilitate the efficient design of organic molecules, whether they are intended to mimic biological tissue, attack cancer cells or suppress the immune system. Hexagons of carbon in a honeycomb pattern are also the basis for sheets of graphite or nanotubes.

Figure 1. Logo from the GO Malaria project: the hexagon is Kekule's model of a carbon ring.

This proposal is for research within a developing area of pure mathematics called *geometric combinatorics*, which combines combinatorial questions with geometrical constructions. This research may lead to finding new and better ways to organize organic molecules.

1.1.1. *Combinatorics.* We plan to study the *polyhexes*, which consist of groups of a certain number of hexagons which share at least one side with another in the group. Three hexagons can be arranged in three ways, allowing for rotations in the plane, as in Figure 2.

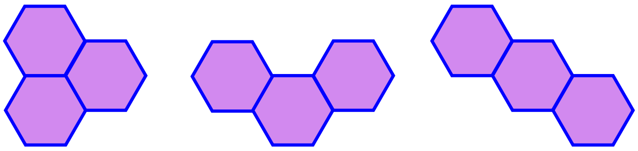


Figure 2. 3-cell polyhexes.

The number of ways to attach the hexagons grows quickly. If we have 14 hexagons to arrange, mathematicians have counted 15,796,897 polyhex arrangements. These arrangements are familiar to an organic chemist, since they are pictures of *polycyclic benzenoid hydrocarbons.* The name refers to the way that carbon often occurs in a molecule as a hexagonal ring of six atoms. One of these rings alone is the molecule benzene, C_6H_6, pictured in Figure 1.

Despite much recent research it is still unknown how to calculate the number of possible hydrocarbons of a given size [1]. In fact there is no known algebraic formula for calculating the total number of polyhexes with precisely n hexagons. Such knowledge would be valuable for anyone planning to search for drug candidates. It would inform the searcher what fraction of the solution space they could cover given time constraints. Important partial results have been discovered, such as when we restrict our attention to special polyhexes like the tree-like ones with a chosen "root" edge [5]. Figure 3 shows the five of these with 2 or fewer hexagons, including the one with zero hexagons.

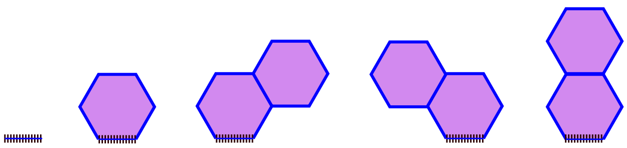


Figure 3. $\leq$ 2-cell rooted polyhexes.

1.1.2. *Geometry.* The PI will not stop at finding a formula for the length of the list of polyhexes. We are also looking for a geometric mapping that organizes them based on similarities. The mathematical principles we plan to focus on involve special shapes: 2-dimensional polygons, 3-dimensional polyhedra, and 4 - dimensional *polytopes*. Recall that dimension can be defined as the number of coordinates needed to describe a point. Polytopes are actually analogs of polygons, but in dimension n for any whole number n. Figure 4 shows some familiar polytopes. The

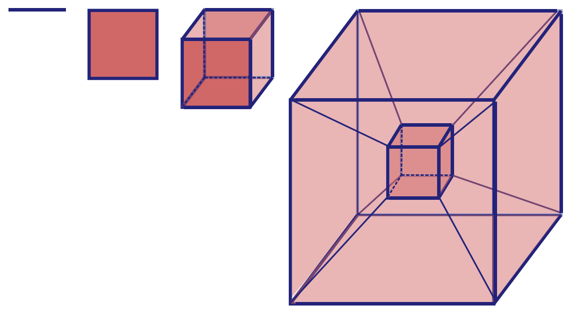


FIGURE 4. Segment, square, cube, hypercube.

4-d *hypercube* is perhaps least familiar. The picture is actually a 3d map of the 4-dimensional shape. Notice the pattern however; the cube is made of 2 squares (front and back) with corners connected, and the hypercube is made of 2 cubes (outer and inner) again with corresponding corners connected. This pattern continues into all higher dimensions: we call the entire list the sequence of n-dimensional cubes. All of these shapes have corners (technically *vertices*) and sides (*facets*) just as in a cut diamond.

The key realization driving the research we propose occurred when the PI discovered several families of new polytope sequences analogous to the n-cubes [3]. One sequence begins with a segment, then a pentagon, then a new 3d shape (the *composihedron* in Figures 6 and 8). Simultaneous with these discoveries was the uncovering of a plethora of combinatorial objects corresponding to the vertices, facets and edges of the polytope sequences.(See Figure 5 where we show the rooted polyhexes from Figure 3 arranged around a pentagon; and then Figure 6.) On top of that came the discovery of natural ways to construct the polytopes in real space.

Most formulas for counting the number of polytopes with n vertices or k facets are still unsolved mysteries. One reason for investigating new polytope families is to add to the collection of data that might reveal patterns about polytopes in general. Our overall goal is to discover how to arrange families of combinatorial structures at the vertices in such a way that the edges and facets take on meaning. Information then can be gained by using facts about polytopes to prove new theorems about polyhexes, or vice versa.

This knowledge should also accelerate the computer processes of building and searching libraries of molecules. The number of facets of the polytopes we are planning to study grows much more slowly than the number of vertices. The sequence of total numbers of tree-like rooted polyhexes starts out 1,2,5,15,51,188... and then eventually grows as quickly as 5^n. The sequence of composihedra however has numbers of facets which start out 0,2,5,10,19,36... and that grow only as quickly as 2^n.

If the facets had meaning in terms of the chemical properties of the molecules, then the search process could be sped up by screening for entire groups of molecules that share a facet. Even better, perhaps only certain facets need be represented in the library. Often there is a general combinatorial object associated to each facet which allows fast generation of the vertex objects. This would help in the building stage, which can be the most time consuming. A database of relatively few facets could implicitly store exponentially more specific compounds– as the mathematically determined vertices of those facets. The vertices would only be generated when actually needed.

Another advantage is that a polytope is *realizable*: it can be thought of as a *convex solid* in space, made of all its interior points rather than just the vertices. Many of our proofs actually construct a realization. This can lead to the solution of problems by use of continuous optimization techniques. For instance a property such as the conductivity of the molecule we are building might be represented by a continuous function on the polytope. If the continuous function is linear, then the convexity of the polytope guarantees that the maximum will be found at a vertex; the process of finding it is known as *linear programming*.

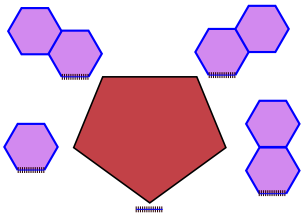


FIGURE 5. Polyhexes around a pentagon: 1^{st} guess.

1.2. Goals and Objectives. Applying geometric combinatorics to the search for cures serves as a grand motivation, and as a goal to be facilitated via interdisciplinary collaboration. Our concrete goals will at first focus on the mathematical structures themselves. Our primary objective is to thoroughly research and publish results about our new discoveries involving polytopes, polyhexes and their algebraic properties in preparation for a federal grant effort. Three main avenues of pure mathematics will be further developed. They are A) discovery and proof of new families of combinatorially generated polytopes; B) research into ways that objects like the polyhexes can be organized by those polytopes; and C) research into the algebraic structure of the faces of those polytopes.

The objective of avenue (A) is to answer questions about the existence of polytopes arising from combinatorics. For instance, the open question: how many polytopes of dimension n with k vertices are there? We will also attack more specific questions, such as: Given any *cell complex*, is there a polytope whose faces model its structure, by corresponding to connected subcomplexes? A cell complex is a construction of a space using multidimensional disks and spheres. There are also analogous questions about several new varieties of combinatorial trees.

Along avenue (B) we will consider some of the polytopes we have already shown to exist, and which happen to have vertex and face counts matching a formula that counts polyhexes or other combinatorial objects. For instance, we are asking: What pattern in the tree-like polyhexes will allow them to be naturally identified with faces of the *composihedra*? See Figures 5 and 6 for conjectures.

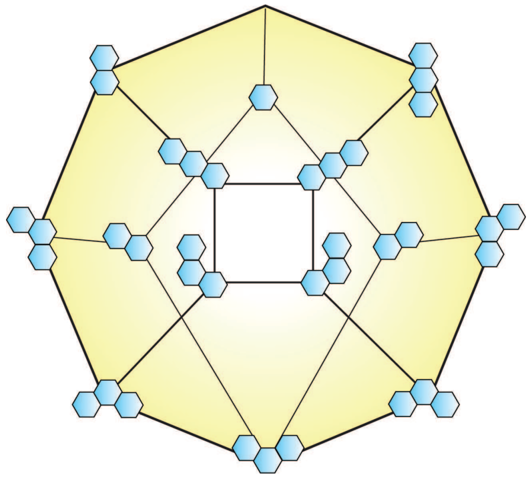


FIGURE 6. Small tree-like polyhexes, on the composihedron.

Finally avenue (C): we will follow up on research of the PI and collaborators which showed that there are natural ways to multiply and divide the combinatorial objects we study. Given a polyhex, there should be a corresponding "inverse" polyhex (or several in combination) called its *antipode*. We ask a question of great interest to algebraists: is there a pattern which will allow us to predict the antipode for any polyhex?

Our second objective is to engage undergraduate math majors and graduate degree candidates in this area of research. It is elementary enough to allow students to quickly be able to participate in the actual research, designing and conducting experiments, looking for patterns in the collected data, and helping to formulate and prove conjectures. The future funding requested of the NSF will include tuition and stipend money for students.

Among the open questions to be researched or assigned to students are: What are all the geometrical properties of the various polytopes–centers, volumes, symmetries, edge lengths and facet areas? Also of interest are the combinatorial properties–number of vertices, numbers of faces, numbers of triangulations, and space tiling.

1.2.1. *Feasability.* This work is the continuation of a series of projects carried out over the last four years by the primary investigator (S. Forcey), four senior collaborators from other universities, and many students. As sole author and jointly the PI has ten papers in this area since 2008 (including 7 in journals, 2 refereed conference proceedings and 1 book chapter). Several more papers are in the initial stages–there is more being discovered than we have time to write up.

There has been very positive interest shown in this project from external grant reviewers. My next step is to apply for external funding from the NSA in October 2012, and this proposal is partly aimed at securing time over the summer to prepare that much longer application. We also plan to try for NSF (DMS Combinatorics) or NIH funding later this year, especially if it can be done as a collaborative effort. More specific plans are detailed in the timeline at the end of this proposal. These include the organization of two special sessions on related topics this October, when the UA Mathematics Department will host the regional meeting of the American Math Society. Myself and collaborators will be hosting two-day sessions in which experts on geometric and algebraic combinatorics will be invited to speak. Students (specifically those working on this project) will also be invited to talk.

The principal investigator has directed six masters theses and seven senior theses. Five of the theses have been on topics described in this proposal. My goal for theses is that they are publication worthy: in the past three of my students have become coauthors on published papers and another is being submitted. Currently two more graduate students at U.A. are busy working with me on polytope problems.

1.3. **Procedures A: Shapes and sizes.** When we speak of a *combinatorial sequence* we mean a list of whole numbers each of which counts the ways of building something given more and more building blocks. For instance the sequence 1,1,3,10,36,137,... gives the numbers of tree-like polyhexes with n hexagons. Another sequence which begins 1,1,2,5,14,42,... gives the numbers of binary trees with n leaves. For example here are the binary trees with 1 or 2 leaves:  . The trees with 3 or 4 leaves can be seen in Figure 7. This sequence of numbers of trees is known as the Catalan numbers, and there is a very large list of other interpretations of (or things you can count with) that same sequence. The fact about the Catalan numbers we want to focus on is the way in which the trees can be organized as the vertices of a polytope. For example in Figure 7 we show the trees with 4 leaves all arranged around the pentagon. The edges (facets)

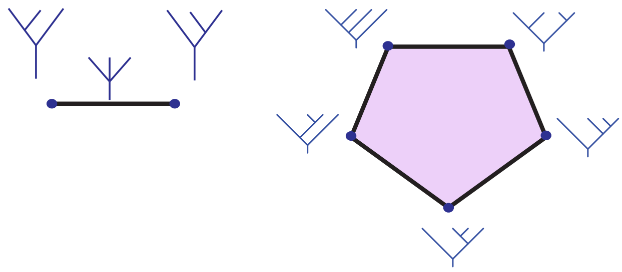


FIGURE 7. Binary trees and their relationships.

of this pentagon can represent simple branch moves from one type of tree to another, or a tree with the branch in mid-position. The latter is seen between the two trees with 3 leaves. When we have five leaves, the trees arrange themselves to make the first shape in Figure 8. It is called the 5th associahedron, or $\mathcal{K}(5)$, and has 14 vertices. The second shape in Figure 8 is called the

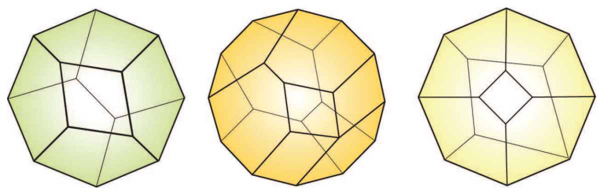
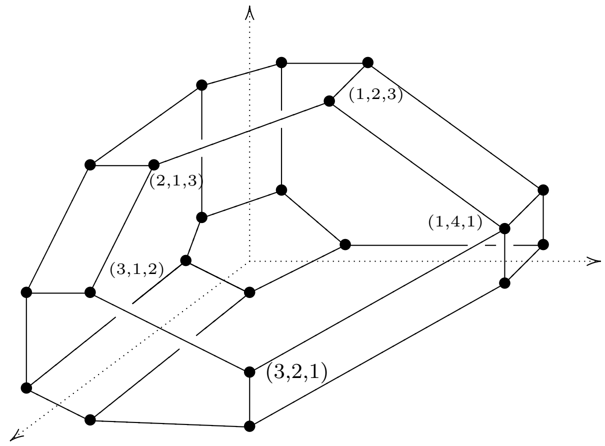


FIGURE 8. Associahedron, multiplihedron and composihedron.

4th multiplihedron, or $\mathcal{J}(4)$. Its vertices correspond to trees with two colors of "paint" applied to their branches. The number of vertices was discovered by the principal investigator to be derived by a simple transformation of the Catalan numbers. The third shape in Figure 8 is also a new invention of the principal investigator. It is called the 4th composihedron, or $\mathcal{CK}(4)$. Its vertices correspond to many combinatorial objects, including the tree-like polyhexes–as discovered by the PI.

First we discuss further polytope research and then we will return to the combinatorial interpretations.

In 2004 it was discovered by J.L. Loday that a simple algorithm existed for finding the actual points in space that are the vertices of the associahedron [6]. This algorithm was generalized to the multiplihedra by the PI in 2008. Here is the actual geometric version of the multiplihedron with some of the points labeled:



Count the number of rectangles, (6), pentagons, (2), and hexagons, (5), to match this to the picture of $\mathcal{J}(4)$ that is in Figure 8 of this paper. This *geometric representation* of the multiplihedra answered the long-standing open question – is every multiplihedron a realizable polytope – in the affirmative.

S. Devadoss, a collaborator of the principal investigator, has also developed a generalization of the associahedra to be based upon any *graph* [2]. Here a graph is just any simple collection of points and connecting lines. Devadoss and the PI collaborated to develop the graph multiplihedra in 2008, and with M.Carr the *pseudograph* associahedra in 2011. The former are based on colored graphs and the latter on graphs with loops and multiple edges. The principal investigator has recently developed a new and more general family of polytopes based on the multidimensional analog of a pseudograph, known as a *cell complex*. Since we have several successful techniques for finding geometric representations, we plan to extend that method of proof to demonstrate that our new polytopes do exist.

Figure 9 shows a graph associahedron and a new pseudograph associahedron discovered by the principal investigator.

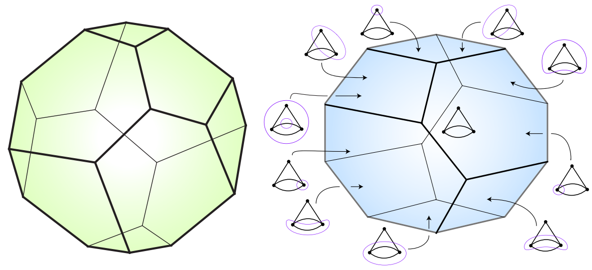


FIGURE 9. The cyclohedron (from a cycle graph) and a new polytope, labeled by subgraphs of a pseudograph.

1.4. **Procedures B: Sorting and counting.** I discovered that the n^{th} composihedron has the same number of vertices as the number of all the rooted tree-like polyhexes with up to n cells. One of the first questions we would like to answer is how the polyhexes might be arranged according to the vertices of $\mathcal{CK}(n)$, and what the edges and facet groupings of the diagrams and their corresponding hydrocarbons might mean. The same sort of question will also be asked for many combinatorial problems involving the other shapes in Figures 8 and 9.

Recall how in Figure 5 and Figure 6 we show the rooted polyhexes with 3 or fewer hexagons arranged on a pentagon and 3-d composihedron. These figures show only possible arrangements. The question is how to choose the “right” arrangement so that it extends meaningfully to an arrangement on the 4-dimensional composihedron of all 51 of the tree-like polyhexes with 4 or fewer hexagons.

In fact, we want to find a recipe for putting the polyhexes with n or fewer hexagons at the vertices of the n-dimensional composihedron. The tools for attacking the problem of finding a meaningful recipe include a list of known one-to-one complete correspondences (bijections) between the polyhexes and other combinatorial objects. Four of these other types of objects are shown in Figure 10 with their correct arrangements around pentagons: strings of words made with a given alphabet, trees with a whole number assigned to each leaf, trees with extra long branches, and branching polyhexes. Others with unknown arrangements (in addition to the rooted tree-like polyhexes) include paths from point to point on

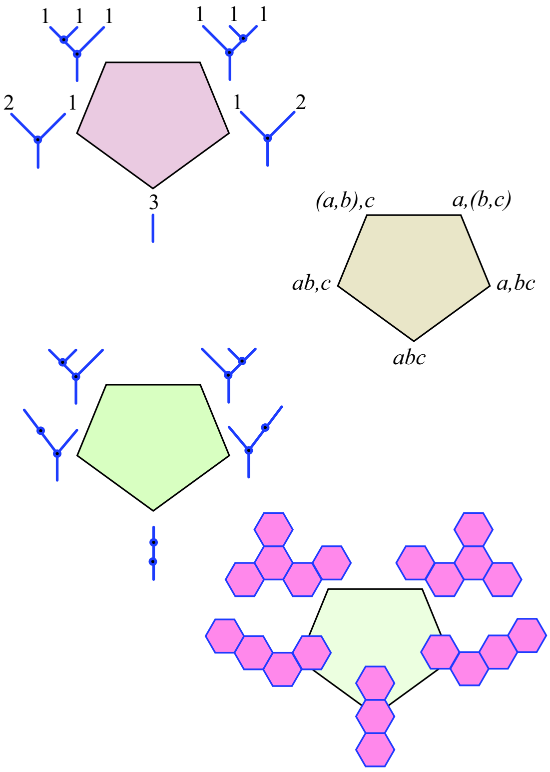


FIGURE 10. More pentagons.

a grid and the symmetric polyhexes with $2n+1$ hexagons. By linking together the various bijections we hope to find useful new ones. Another tool is to understand the polytope edges as moves made between objects. Specific bijections are often also found using generating function techniques.

1.5. **Procedures C: Combining.** Here is just a brief description of some of the results we have achieved regarding the algebraic structure of our combinatorial polytopes. First, the multiplication and division we actually do takes place in what is called a *Hopf algebra* of the polytope. This means, for instance, that the result of multiplying two trees is a sum of several trees [6]. Thus the objects of the algebra are actually collections of objects from the polytope.

In recent papers with F. Sottile and A. Lauve, the PI showed for the first time that the vertices of the composihedron and the multiplihedron can be multiplied in a coherent way [4]. Figure 11 shows the result of multiplying two of the vertices of the composihedron. They are represented by the trees with whole numbers as leaves, as seen in Figure 10. We would very much like to see how this product looks using polyhexes–but that will require a successful discovery of a bijection between the trees and the polyhexes. On top of that, we plan to experimentally search for the inverses, or antipodes, that this Hopf algebra structure implies exist for every polyhex.

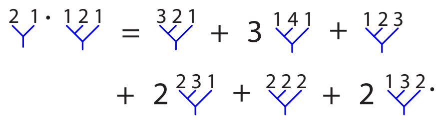


FIGURE 11. Multiplying in the composihedron.

1.6. **Expected Results and Data Analysis.** The experiments we plan to perform are initially all done using either a computer or, for simple cases, on paper. They involve working many examples. For instance, we hypothesize that the bijection pictured in Figure 6 extends to the 4th dimension. This will require checking 51 cases to be certain. We also hypothesize that our constructions using cell complexes and trees do indeed result in polytopes. First we need to test this by constructing more examples and then we need to turn those examples into the plan for a general proof. A successful experiment can either reveal a new pattern (perhaps add confirming evidence for a pattern we already guessed at) or it can provide a counterexample that rules out a false positive.

When the numbers of vertices or faces of a particular sequence of polytopes are known, then there is the opportunity to find other (molecular) interpretations of those numbers which the polytopes also help to organize. Another variety of experiment we plan is to generate all the possible objects in a family–polyhexes or certain tree diagrams– and then develop a table listing the numbers of each type. The Online Encyclopedia of Integer Sequences (OEIS) is the premier database for series of counting numbers that occur in math research. It provides a free way to check our results against millions of known and conjectured patterns. Eventually, if we develop molecular descriptors (numerical predictors of physical properties), there is also a free online database of molecular properties–both physical and mathematical. It is called

"Datasets for molecular descriptor comparisons", and is provided by the International Academy of Mathematical Chemistry.

1.7. **Publication and presentation.** My students and I will be presenting results in the AMS special session I am organizing here at U.A. this October. Also we plan to attend the annual meeting entitled FPSAC (Formal power series and algebraic combinatorics) where our preliminary papers were presented last year and the year before. Journals we plan to utilize for our upcoming papers include the pure math journals Annals of Combinatorics and J. of algebraic combinatorics; as well as the Journal of Chemical Information and Modeling and the Journal of Mathematical Chemistry. All these methods of dissemination are crucially important to achieving the requirements for RTP and merit in the mathematics department. One further method of dissemination we have already begun is a website entitled "Encyclopedia of Combinatorial Polytope Sequences." There are many completed entries already, from the classic sequences to new discoveries.

1.8. **Collaborators.** S. Devadoss from Williams College has coauthored two papers about polytopes with the PI, and is a coauthor on a third in progress. A. Lauve from Loyola Chicago and F. Sottile from Texas A&M have coauthored 2 journal articles and 2 conference proceedings about algebraic combinatorics – and more are planned. M.Carr, now at Brandeis, is a former student of S. Devadoss who is also a coauthor. J.P. Cossey at U.A. works with Catalan numbers and will be consulted. M. Ronco, from Valparaiso Chile, has collaborated on basic research and will be visiting U.A. this spring to work on a new paper with the PI. F. Fisher is a former student from George Washington U. who will be cohosting the special session at U.A. this October. J. Stasheff is a senior researcher at the U. of Pennsylvania who has helped put the PI in touch with many researchers, and who requested the building of the website "Encyclopedia of Combinatorial Polytope Sequences."

References

1. D. Babić, D.J. Klein, J. von Knop, and N. Trinajstić, *Combinatorial enumeration in chemistry*, Chem. Modelling, Appl. and Theory **3** (2004).
2. Satyan L. Devadoss, *A realization of graph associahedra*, Discrete Math. **309** (2009), no. 1, 271–276. MR MR2479448
3. Stefan Forcey, *Quotients of the multiplihedron as categorified associahedra*, Homology, Homotopy Appl. **10** (2008), no. 2, 227–256. MR MR2475611
4. Stefan Forcey, Aaron Lauve, and Frank Sottile, *Hopf structures on the multiplihedra*, SIAM J. Discrete Math. **24** (2010), 1250–1271.
5. Frank Harary and Ronald C. Read, *The enumeration of tree-like polyhexes*, Proc. Edinburgh Math. Soc. (2) **17** (1970), 1–13. MR MR0284379 (44 #1607)
6. Jean-Louis Loday and María O. Ronco, *Hopf algebra of the planar binary trees*, Adv. Math. **139** (1998), no. 2, 293–309. MR **MR1654173 (99m:**16063)

1.9. **Funding Status.** The PI is nearing the end of his start-up funding at U.A.

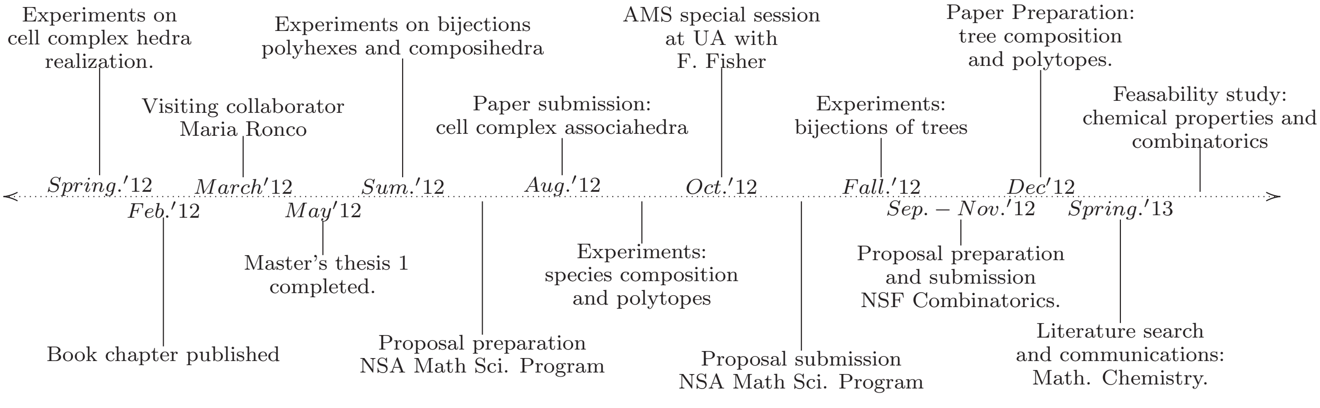


Figure 12. Timeline.