DISSERTATION

MOLECULAR CONFIGURATIONS AND PERSISTENCE: BRANCHED ALKANES AND ADDITIVE ENERGIES

Submitted by

Brittany M. Story

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2022

Doctoral Committee:

Advisor: Henry Adams

Patrick Shipman
Jeff Achter
Anders Fremstad

ABSTRACT

MOLECULAR CONFIGURATIONS AND PERSISTENCE: BRANCHED ALKANES AND ADDITIVE ENERGIES

Energy landscapes are high-dimensional functions that encapsulate how certain molecular properties affect the energy of a molecule. Chemists use disconnectivity graphs to find transition paths, the lowest amount of energy needed to transfer from one energy minimum to another. But disconnectivity graphs fail to show not only some lower-dimensional features, such as transition paths with an energy value only slightly higher than the minimum transition path, but also all higher-dimensional features. Sublevelset persistent homology is a tool that can be used to capture other relevant features, including all transition paths. In this paper, we will use sublevelset persistent homology to find the structure of the energy landscapes of branched alkanes: tree-like molecules consisting of only carbons and hydrogens. We derive complete characterizations of the sublevelset persistent homology of the OPLS-UA energy function on two different families of branched alkanes. More generally, we explain how the sublevelset persistent homology of any additive energy landscape can be computed from the individual terms comprising that landscape.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Henry Adams, for all the impromptu meetings and good conversations and my committee for all of their time and energy. My husband, Dustin Story, for supporting me and reminding me that I can, in fact, get a PhD. My friends for long walks, office talks, and fun, non-school adventures. Finally, I'd like to thank my family for believing in me all along, even when I didn't believe in myself.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Chemists want to understand the structure and shape of energy landscapes; real-valued functions that are evaluated over high-dimensional domains of variables that affect the energy of a molecule. As one example, consider carbon-based molecules. Carbon-based molecules can get very large and as a result, their energy landscapes can be extremely high dimensional. Every bond added to a molecule increases the dimension of the domain of the energy landscape. If chemists want to fully understand the behavior of larger molecules, they need information about the structure of the energy landscape. One characteristic chemists want to understand is how molecules transition between energy minimums. Traditionally chemists use merge trees, diagrams that capture the lowest transition path between two minimums, to understand how molecules transition from one minimum energy state to another [3, 4]. But due to randomness or external forces, molecules can sometimes transition through higher minimums not captured by merge trees. Sublevelset persistent homology is a tool from topological data analysis that captures all homological features of the function. Further, since the energy functions are Morse, sublevelset persistent homology can capture all critical points of all degrees, when it can be computed. This provides chemists with all possible transition paths, as well as information about higher dimensional features.

Carbon-based molecules come in a variety of structures. Chains of carbons, n-alkanes, have the most simple structure of the alkane molecules. But, carbon atoms can bond with anywhere from one to four other carbons, and as such, we study more complex molecules. We will call any tree-like carbon structure a *branched alkane*. Carbons can also bond to each other with multiple bonds. *Alkenes* are molecules where two carbons are bonded together by double bonds while *alkynes* have triple bonds [5]. Finally, carbons can also form loops, which can contain three or more carbons; these molecules are called *cyclo-alkanes*. A natural question asks how we can study the homological structure of different carbon-based molecules.

Work has been done on understanding the energy landscapes of n-alkanes in [2]. Other work has also been done on the energy landscape of cyclo-octane in [6], but as we increase the size of the cyclo-alkanes, the domain space gets complicated very quickly. We will restrict our focus to examining the persistent homology of the energy landscapes of branched alkanes.

Topology is often used to examine the geometry of high-dimensional spaces. Several papers, such as those by Pietrucci and Andreoni [7] and Zhou et al. [8], look at topological descriptions of intermolecular interactions. In the context of molecular chemistry, persistent homology, a tool from topology, has been used to examine energy landscapes, namely functions that map an input space of states in a chemical system to a corresponding energy value. The paper by Mirth et al. [2] looks specifically at the energy landscapes of n-alkanes. For one approximation, the domain of any n-alkane chain can be reduced to a product of circles, where each circle S^1 represents the rotation angle of a bond between two carbons. Utilizing this deconstruction, the authors were able to use a formulation of the Künneth formula [9], a result that generalizes the universal coefficient theorem for homology [10], to describe the length and birth time of all bars in the persistent homology barcode. This work allowed the authors to characterize the sublevelset persistent homology of the energy landscape for any n-alkane.

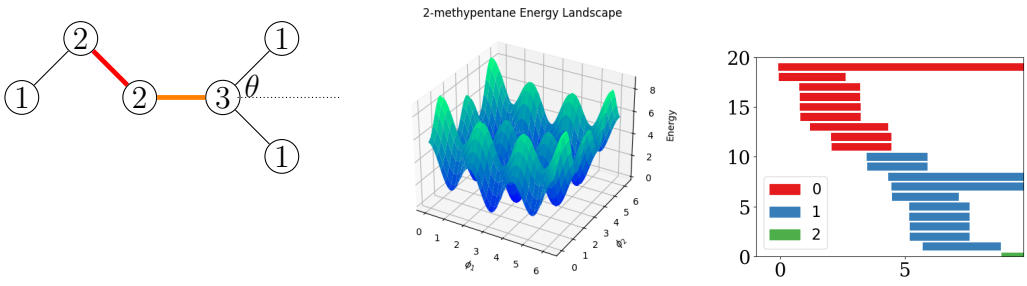

Figure 1.1: [Left] An example of a branched alkane with two dihedral types; 1-2-2-3 (red) and 2-2-3-1 (orange). [Middle] The molecule's corresponding energy landscape. The ϕ_1 axis is the position of the 2-2 internal bond and the ϕ_2 axis is the position of the the 3-2 internal bond. Each pair (ϕ_1, ϕ_2) gives a corresponding energy value. [Right] The sublevelset persistent homology barcode of the energy landscape. The x-axis is the energy value that corresponds to the homological feature and the y-axis is a count of the number of bars.

We would like to do a similar characterization of the sublevelset persistence for branched alkane energy landscapes. Branched alkanes consist of two types of bonds; internal and external. Internal bonds, or bonds that connect carbons of degree 2 or higher, are the bonds we will focus on that contribute to the energy landscape. We will denote *internal bonds* as strings of two numbers, while we will call strings of four numbers the *dihedral type*, where each number is the degree of the carbon surrounding the internal bond. In Figure 1.1, the molecule consists of 6 carbons and 5 bonds, 2 internal bonds (red and orange) and 3 external bonds (black). The internal bonds are 2-2 and 3-2 with corresponding dihedral types 1-2-2-3 and 2-2-3-1. If we rotate around the two internal bonds, the pair of angles (ϕ_1, ϕ_2) corresponds to a single energy value in $\mathbb{R}$. Note, the approximation of the energy landscapes we consider only uses interactions between the bonded atoms. The interactions between non-bonded atoms do have an effect on the energy, but we will ignore that effect in this specific approximation. Hence, the energy contribution of angle ϕ_1 is not dependent on the position of angle ϕ_2. Thus, these functions are additive functions over a product space and as such, in this approximation we can combine the energy functions that correspond to each internal bond to get the energy function of the entire molecule. Then, by applying the persistence formulation of the Künneth formula, we can use the sublevelset persistence of each internal bond to calculate the sublevelset persistent homology of larger molecules.

Throughout this paper, we will explore how the persistent Künneth formula can be used to calculate the sublevelset persistence of branched alkane energy landscapes. First we will give a brief overview of relevant background on chemistry, sublevelset persistence, Morse theory, and the Künneth formula in Chapter 2. Next in Chapter 3, we will look at the sublevelset persistence for six specific dihedral types, which we will use to build larger branched alkanes. Chapter 4 provides results characterizing certain features for all branched alkanes, such as the number of bars in each dimension. In Chapter 5, with an understanding of these six building block dihedral types, we will completely characterize the sublevelset persistence for a few select examples via the Künneth formula. Using that information, we

will generalize our results for a larger class of molecules in Chapter 6, and then explain how to extend these results to any branched alkane.

Over the course of these chapters, we establish several new results. We determine the number of bars in the sublevelset persistence barcode for any branched alkane in Theorem 4.1.2. Theorem 4.2.1 provides the number of k-dimensional bars in each barcode. Then, Theorem 4.3.2 gives the number of bars of each length in the barcode. Finally, we will completely characterize the energy landscapes for two different types of branched alkanes: molecules with all 3-2 type internal bonds (Theorem 5.2.6) and molecules with both 2-2 and 3-2 types of internal bonds (Theorem 6.1.5). Using the results above, we will discuss how to extend them to determine the sublevelset persistent homology of any additive function over a product space by using the persistence barcodes of the product components.

Chapter 2

Preliminaries

To fully explore the homology of branched alkane energy landscapes, we will introduce some necessary notions from chemistry. In particular, we will introduce the family of branched alkane molecules. After we look at the some of the underlying chemistry of branched alkanes, we will provide some background on sublevelset persistent homology. Next, we will look at Morse theory and how it can be combined with sublevelset persistence to provide us with additional structure. Finally, we will look at the main underlying tool of the paper; the persistence Künneth formula, and how we apply it to energy landscapes.

2.1 Branched alkanes

Throughout this paper, we consider branched-chain alkanes, or as we will refer to them, branched alkanes. Alkanes are a type of molecule which consist of carbons and hydrogens connected by single bonds. Branched alkanes are molecules that are made up of only carbons and hydrogens whose bond structure is tree-like; there are no cycles. For more on alkanes and branched alkanes, see Chapter 4 of [5]. One specific subset of branched alkanes, n-alkanes, consist of chains of carbon atoms and hydrogen atoms connected by single bonds. These have already been studied by Mirth et al. in [2], so we will look exclusively at molecules that are not chains.

Note, for the entirety of this paper we will look exclusively at the OPLS-UA (Optimized Potential Liquid Simulation - United Atom) approximation of the energy landscape, as established in the paper by Jorgensen and Tirado-Rives [11]. The OPLS-UA energy landscape is a function that inputs the rotational angle of each internal bond and outputs an energy value. This model equalizes all bond lengths and fixes the angles between any three carbon atoms. Additionally, the OPLS-UA model includes the energy from the hydrogen atoms in the carbon atoms and ignores non-bonded intramolecular interactions. Thus, even though

each carbon has degree four including hydrogens, when looking at pictures of molecules, we will only look at the structure of the carbons.

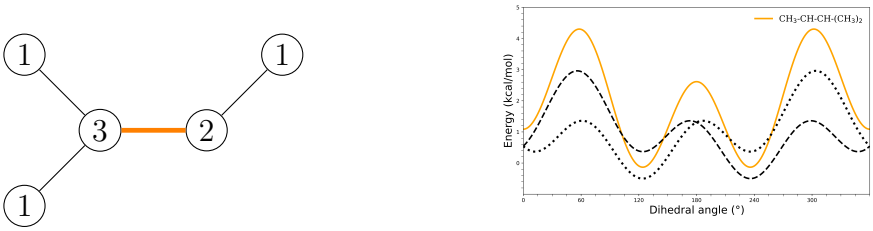

Figure 2.1: [Left] One configuration of isopentane with only the carbons shown. The internal bond is denoted by a thick orange line and the number inside each carbon atom is the degree of that carbon with respect to other carbons. [Right] The corresponding OPLS-UA energy landscape.

For example, consider the branched alkane C_5H_{12}, known commonly as isopentane, shown in Figure 2.1. Isopentane has one internal bond; a bond where the two bounding carbons each have degree greater than or equal to two. As we are looking to characterize the energy landscapes of branched alkanes, first consider the energy landscape of isopentane (Figure 2.1). The OPLS-UA energy landscape of isopentane is given on the right of Figure 2.1. This function is calculated with the coefficients given in [12–14]. When adjacent carbons are close together (for isopentane, those labeled with 1's), they produce a higher energy value. Similarly, when they are far apart, they produce a lower energy value. Hence, as we rotate around the internal bond, we get a changing energy value that is in direct relation to the angle of the internal bond. For the entirety of this paper, we will denote *internal bonds* as strings of two numbers, while we will call strings of four numbers the *dihedral type*, where each number is the degree of the carbon surrounding the internal bond. These dihedral types determine the OPLS-UA energy functions of each molecule. For isopentane, we have only one internal bond; the bond connecting the carbon atoms of degrees 3 and 2. Further, this molecule has two copies of one dihedral type, which we denote as 1-3-2-1.

Different internal bonds and dihedral types affect the energy of the molecule. As we will see in Chapter 3, the energy landscapes of butane (the molecule with one copy of dihedral

type 1-2-2-1) and isopentane (the molecule with two copies of dihedral type 1-3-2-1) have different shapes (Figure 3.1). Our overarching goal is to characterize the geometry of the energy landscapes of different types of branched alkanes, including larger molecules with many atoms. As mentioned above, the natural first step is to examine the sublevelset persistence of n-alkanes, which is done in [2]. The next step, and the purpose of this paper, is to consider all possible tree-shaped molecules consisting of carbons and hydrogens. We will calculate and then utilize the sublevelset persistence of six different dihedral types to approximate the energy landscapes of larger molecules. But first, we discuss exactly what we mean by sublevelset persistent homology.

2.2 Sublevelset persistent homology

Next, we develop some background on sublevelset persistent homology. Sublevelset persistence examines the sublevelsets of a space as the threshold defining the sublevelset varies. The threshold variable could represent time, height, or as in this work, an energy value. Sublevelset persistence is a relatively new mathematical topic that has several applications to chemistry [15–17]. The following explanation will follow the book by Edelsbrunner and Harer [18].

Recall that persistent homology looks at how long homological features, such as components, holes, voids, etc. persist over a given scale. In the context of energy landscapes, chemists are curious about the energy values of critical points of all indices. Roughly speaking, the index of a critical point is the number of linearly independent directions that one could move from a critical point while having the energy value decrease. For example, critical points of index one correspond to saddle points, points where you can flow down in only one direction. Each saddle point corresponds to an energy barrier, or the minimal amount needed to get from one local minimum to another. Identifying these saddle points and other higher-dimensional critical points provides information regarding the homological features of the energy landscape, even without visualization.

For example, chemists are concerned with how molecules transition between energy minimums. Molecules are constantly fluctuating, but they are often found near low energy configurations which correspond to energy minimums. For a molecule to transition between energy minimums, they have to pass through a saddle point that connects the two minimums. Merge trees are one tool chemists have used to find these energy barriers. These graphs identify the lowest energy saddle point a molecule would have to pass through to transition from one minimum to another. One restriction of merge trees is that they only capture the lowest energy value; there could be another path that has a saddle point with a slightly higher energy value that a molecule might take. Hence, having information about all critical points provides further insight into molecular transformations and interactions. Therefore, we are interested in the homology of an energy landscape, X, and how it changes as the height of the energy threshold changes.

Let $f\colon X \to \mathbb{R}$ be a real-valued function with domain X. We define a *sublevelset* for some value $t \in \mathbb{R}$ as $X_t = \{x \in X \mid f(x) \in (-\infty, t]\}$. Note for $t \leq t'$, we have $X_t \subseteq X_{t'}$. To determine the sublevelset persistent homology, we will start by looking at singular homology. The definitions and notation we use originate from the book "Algebraic Topology" by Allen Hatcher [10]. Define $(C_\bullet(X), \partial)$ to be the singular chain complex, where each $C_n(X)$ is the free abelian group where the basis is the set of singular n-simplices in X. A singular n-chain, σ, is a finite sum given by $\sigma = \sum_i n_i\sigma_i \in C_n(X)$, where $n_i \in \mathbb{Z}$ and $\sigma_i\colon \delta^n \to X$. A boundary map $\Delta_n\colon C_n(X) \to C_{n-1}(X)$ is defined to be

$$\delta_n(\sigma) = \sum_i (-1)^i \sigma|[v_0, \ldots, \hat{v}_i, \ldots, v_n]$$

where $\hat{v}_i$ is the removal of the ith vertex. This gives a singular $(n-1)$-chain. Thus, consider the standard homology on $(C_\bullet(X), \partial)$ where $H_n(X) = \ker(\partial_n)/\mathrm{im}(\partial_{n+1})$. The sublevelset persistent homology is given by first evaluating the singular homology on X_t, $H_k(X_t)$, as opposed to all of X. Furthermore, in sublevelset persistence, one also considers the morphisms $H_k(X_t) \to H_k(X_{t'})$ induced by the inclusions $X_t \overset{X}{\hookrightarrow}_{t'}$ for $t \leq t'$.

A common first example of sublevelset persistence looks at a torus standing on its side and examines the changing homology as the height increases; see Figure 2.2. Starting at the base, there is a single connected component isomorphic to the closed disk. As we increase the height of the function, we have that the homology transitions to that of a circle at the second critical point, and then changes to two connected circles (a figure-eight), until at the top, we recover the whole torus.

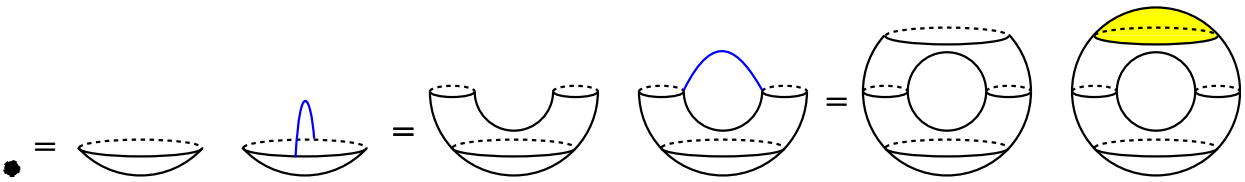

Figure 2.2: A depiction of sublevelset persistent homology for the torus. We start with a point, evolve to a circle, add in a second circle, and finally a void. Figure from [1].

We can capture this information via a persistence barcode. A *persistence barcode* is a collection of horizontal bars, where the birth of a k-homological feature is the value of the left endpoint of the bar and the death (if it exists) of that feature is the right endpoint of the bar. The color of a bar gives the dimension of the feature.

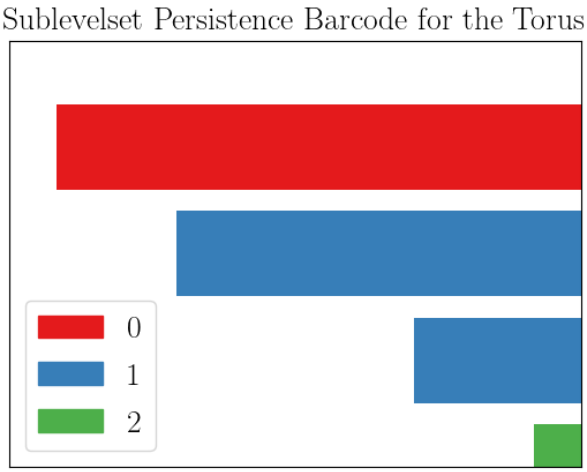

Figure 2.3: Sublevelset persistence barcode for the torus. The color of the bar gives the dimension of the feature and all bars are semi-infinite.

For the example above, the barcode in Figure 2.3 is the sublevelset persistence barcode for the torus in Figure 2.2. In this example, all bars are semi-infinite, which means that they have no right endpoint (death value). We will see in later examples that this is not always the case; there will be homological features that only persist for a finite range. Further, the functions we will be studying are Morse functions, a well-behaved class of functions outlined in the section below.

2.3 Morse theory

In order to talk about Morse homology, we introduce some background on Morse theory. An extensive look can be found in the book by Milnor [19], but we provide a summary below. Morse theory is a classic topic in smooth topology that relates directly to the newer invention of persistent homology. Consider a smooth function $f\colon X \to \mathbb{R}$ where X is a smooth manifold. Throughout this thesis, we use the word manifold to refer to a manifold without boundary. We say $x \in X$ is a critical point of f if and only if $\nabla f(x) = 0$. Further, we call x a non-degenerate critical point of X if the Hessian at x is non-singular. We say f is a *Morse function* if it is smooth and if all critical points are non-degenerate [20]. Throughout this paper, f will be a function that takes any number of bond angles in a molecule as input, and outputs the energy value of the molecule at that configuration. Each bond angle has a circle's worth of positions, hence we denote the domain of each bond with S^1. We could expand this idea to include parameters such as bond length and bond type, which would also give smooth manifolds, but we will restrict to bond angle in this paper. Since we are looking specifically at how the bond angles affect the energy, the domain of the energy function is the n-dimensional torus $(S^1)^n$, which is a smooth manifold. Further, each function we consider in this paper will have non-degenerate critical points, which implies that all functions we consider will be Morse.

Since our functions are Morse, we utilize two results from Morse theory. The first shows that if f is a smooth function on some finite dimensional smooth manifold, then given an interval $[a, b]$ that contains no critical points, X_a is homotopy equivalent to X_b.

Theorem 2.3.1 (Banyaga and Hurtubise [20]). *Let $f\colon X \to \mathbb{R}$ be a smooth function on a finite dimensional smooth manifold. For all $t \in \mathbb{R}$, let*

$$X_t = f^{-1}((-\infty, t]) = \{x \in X \mid f(x) \leq t\}.$$

Let $a < b$ and assume that $f^{-1}([a, b])$ is compact and contains no critical points of f. Then, X_a is diffeomorphic to X_b and X_a is a deformation retract of X_b. Moreover, there is a smooth diffeomorphism $F\colon f^{-1}(a) \times [a, b] \to f^{-1}([a, b])$ such that the diagram

$$\begin{array}{ccc} f^{-1}(a) \times [a,b] & \xrightarrow{F} & f^{-1}([a,b]) \\ & \searrow^{\pi_2} & \downarrow f \\ & & [a,b] \end{array}$$

commutes. In particular, all the level surfaces of f between a an b are diffeomorphic.

This means that the bars in the sublevelset persistent homology can only change when we cross a critical point. The homology of the function stays the same between two height-adjacent critical points. Next, we describe how crossing a critical point with index k in the sublevelset adds a k-cell to the homotopy type.

Theorem 2.3.2 (Banyaga and Hurtubise [20]). *Let $f\colon X \to \mathbb{R}$ be a smooth function. Suppose that for $a < b$, $f^{-1}([a, b])$ is compact and inside $f^{-1}([a, b])$ there is exactly one critical point. Assume that this critical point is non-degenerate and of index k. Then X_b has the homotopy type of X_a with one k-cell attached. In fact, there exists a set $e^k \subseteq X_b$ diffeomorphic to a closed k-disk $D^k = \{x \in \mathbb{R}^k \mid |x| < 1\}$ such that $X_a \cup e^k \subseteq X_b$.*

Combined with Theorem 2.3.1, this tells us that until the sublevelset crosses a critical point, the homology remains the same. Once we cross a critical point, the homology changes

in a predictable way. These two results will be helpful when discussing the sublevelset persistence of Morse functions.

2.4 Sublevelset persistence and Morse theory

We want to completely characterize the sublevelset persistence of the energy landscapes of branched alkanes. As such, we will look at a well-established lemma connecting Morse theory to sublevelset persistence, namely Lemma 2.4.1.

First, define $(C_\bullet(f), \partial)$ to be the chain complex for Morse homology, as used in [2, 20]. Each $C_k(f)$ is the free abeliean group generated by the critical points of index k ($Crit_k$). The boundary map is the function

$$\partial_{k+1}(q) = \sum_{q \in Crit_k(f)} n(q,p)p$$

where $n(q,p)$ is the number of signed gradient flow lines from q to p. Let $\mathcal{M}^f_{p,q}$ be the moduli space of gradient flow paths from p to q (with $\mathbb{Z}/2\mathbb{Z}$ coefficients). Recall, the gradient flow paths refers to the path a point would follow to flow either up or down. Hence, the k^{th} homology of the chain complex $(C_\bullet(f), \partial)$ is given by $H_k(X) = \ker(\partial_k)/\mathrm{im}(\partial_{k+1})$. For the remainder of this thesis, we will use Morse homology. Thus, when we are talking about sublevelset persistent homology, this is the chain complex and homology we are referring to. With this notation in mind, we state the Morse Lemma.

Lemma 2.4.1 (Morse Lemma)**.** *If $f\colon M \to \mathbb{R}$ is a Morse function, then the birth and non-infinite death values in the sublevelset persistent homology correspond precisely to the critical points of f. Each k-dimensional bar has birth time corresponding to a critical point of index k, and death time either equal to infinity or otherwise corresponding to a critical point of index $k+1$. Furthermore, the number of semi-infinite bars in dimension k is given by the k-dimensional homology of M.*

Lemma 2.4.1 gives that the sublevelset persistent homology of a Morse function is completely determined by the critical points of the energy landscape. A proof of this can be found in Appendix A.1 of [2].

As an example, consider the branched alkane isopentane as seen in Figure 2.1 and as discussed in Section 2.1, which we refer to as 1-3-2-1, pictured below. We label the critical points as a_1, a_2, b, c, d_1, and d_2, and their corresponding energy values as $\alpha, \alpha', \beta, \gamma, \delta, \delta' \in \mathbb{R}$ with $\alpha \leq \alpha' < \beta < \gamma < \delta \leq \delta'$, where critical point a_1 corresponds to energy value α, a_2 corresponds to α', b corresponds to β, c corresponds to γ, and d_1 corresponds to δ, and d_2 corresponds to δ'. Note, $\alpha = \alpha'$ and $\delta = \delta'$, but in order to classify critical points properly, we will later define an ε-perturbation to break symmetry and as a result, $\alpha < \alpha'$ and $\delta < \delta'$.

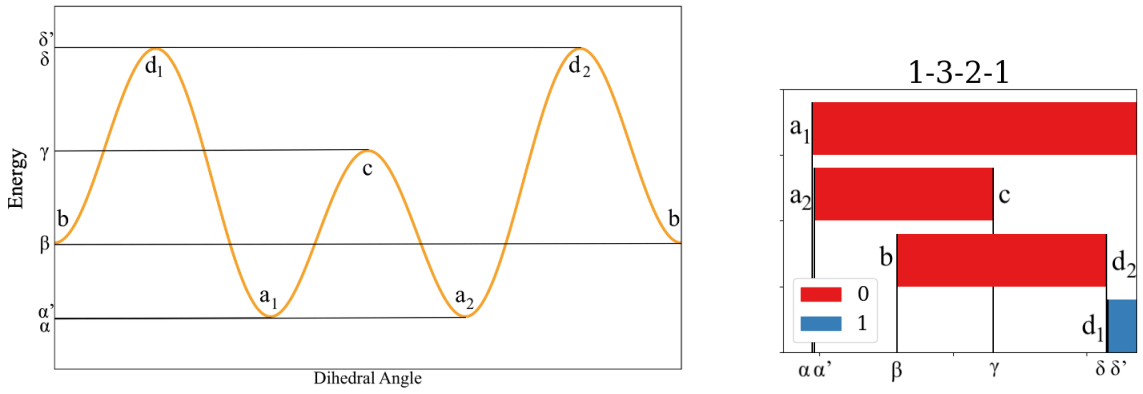

Figure 2.4: An example of the correspondence between the critical points of a function and the birth and death times of the persistence bars.

Each birth and death in the sublevelset persistent homology corresponds to a critical point of the function. All of the dimension 0 critical points correspond to births of 0-dimensional bars. Two of the 1-dimensional critical points correspond to deaths of 0-dimensional bars, while one of them corresponds to the birth of a 1-dimensional bar. This completely describes the sublevelset persistence of this function. Further, we will see in the next section that if we have an additive function over a product space, we can use the sublevelset persistence of each component to determine the sublevelset persistence of the entire function.

2.5 Künneth formula

Suppose we know the sublevelset persistent homology of two energy landscapes over X_1 and X_2. We want to calculate the sublevelset persistent homology of the energy landscape over $X_1 \times X_2$. More precisely, if $f_i \colon X_i \to \mathbb{R}$ is a collection of functions for $i = 1, \ldots, n$, then one can define their sum $f \colon X_1 \times \ldots \times X_n \to \mathbb{R}$ on the product space $X_1 \times \ldots \times X_n$ by $f(x_1, \ldots, x_n) = f_1(x_1) + \ldots + f(x_n)$. We will often refer to these functions as additive functions over a product space.

Additive functions over a product space possess some nice properties, particularly if each component function is a Morse function. One such property is that the sum of Morse functions over a product space is also a Morse function. Another nice property of Morse functions characterizes the critical points of the additive function via the critical points of the component functions.

Lemma 2.5.1. *Let $X_1, \ldots X_n$ be manifolds, let $f_i \colon X_i \to \mathbb{R}$ be Morse functions, and let $f \colon X_1 \times \cdots \times X_n \to \mathbb{R}$ be the additive function over a product space defined by $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i)$. Then f is a Morse function. Further, the point $(x_1, x_2, \ldots, x_n)$ is a critical point of f if and only if each coordinate x_i is a critical point of f_i. Finally, the index of a critical point $(x_1, x_2, \ldots, x_n)$, denoted by $\mu_f(x_1, x_2, \ldots, x_n)$, is equal to the sum of all indices of the component functions,*

$$\mu_f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \mu_{f_i}(x_i).$$

Proof. We prove the case where each X_i is one-dimensional, which is the only case of this lemma we will need. For the proof of the general case, see [2].

To show f is a Morse function, we must show that f is smooth and that all critical points are non-degenerate. First note that each component function, $f_i(x_i)$ is smooth; hence all derivatives exist. Additionally for each i, $\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) = \frac{\partial f_i}{\partial x_i}(x_i)$. Hence when $i \neq j$,

$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \ldots, x_n) = 0$ and for $i = j$, $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \ldots, x_n) = \frac{\partial^2 f_i}{\partial x_1^2}(x_i)$, which exists since each f_i is smooth. This also holds for higher-order derivatives. Therefore f is smooth.

Next, consider the Hessian of f. Above we see that all mixed partial derivatives are 0. As such, the Hessian of f is a diagonal matrix such that each entry on the ith row, ith column is $\frac{\partial^2 f_i}{\partial x_i^2}$. Note, each f_i is a Morse function and as such, the Hessian of each component function is non-singular and is simply the second derivative of each function f_i with respect to x_i. Therefore, the Hessian of f is non-singular and f is a Morse function.

Now, let $(x_1, \ldots, x_n) \in X_1 \times \cdots X_n$. As shown above, since each f_i is a Morse function, f is also Morse. Hence, $(x_1, \ldots, x_n)$ is a critical point of f if and only if $\nabla f(x_1, \ldots, x_n) = 0$. But, the gradient of f is 0 if and only if the gradient of each f_i is 0. Therefore, each coordinate x_i is a critical point of f_i if and only if $(x_1, \ldots, x_n)$ is a critical point of f.

Finally let $(x_1, x_2, \ldots, x_n)$ be a critical point of f and consider the index of that point, denoted by $\mu_f(x_1, x_2, \ldots, x_n)$. The index of a critical point is the number of negative eigenvalues of the Hessian matrix evaluated at the critical point. Since the Hessian is diagonal, the eigenvalues are simply the numbers in the diagonal. Additionally, the Hessian is non-degenerate so each value on the diagonal is either positive or negative. Thus, the index of this critical point is given by the number of entries with negative values. For each f_i, if $\frac{\partial^2 (f_i)}{x_i^2}(x_i)$ is negative, $\mu_{f_i}(x_i) = 1$. Similarly, if $\frac{\partial^2 (f_i)}{x_i^2}(x_i)$ is positive, $\mu_{f_i}(x_i) = 0$. Therefore,

$$\mu_f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \mu_{f_i}(x_i).$$

□

Additive energy functions over a product space of circles are a natural way to view larger energy landscapes of molecules that are determined by bond angles. If we fix one variable, the function over the remaining variables looks like the energy landscape of a smaller molecule. Now, if we have information about the homology of the smaller energy landscapes, we want to use that information to uncover information about the homology of the larger molecule.

The Künneth formula is the main result we need to calculate the sublevelset persistence of the branched alkanes based on the sublevelset persistence of smaller molecules. The algebraic version of this formula generalizes the universal coefficient theorem for homology (see [10] for an in-depth look at its formulation). In this case, we are concerned with the topological version.

Theorem 2.5.2 (Topological Künneth formula [10])**.** *If X and Y are CW complexes and R is a principal ideal domain, then there are natural short exact sequences*

$$0 \to \bigoplus_i \left(H_i(X;R) \otimes_R H_{n-i}(Y;R)\right) \to H_n(X \times Y;R)$$
$$\to \bigoplus_i \operatorname{Tor}_R(H_i(X;R), H_{n-i-1}(Y;R)) \to 0$$

and these sequences split.

Now, let X and Y be filtered spaces. Recall, a filtration of spaces is a nested sequence of spaces $X_0 \subset \ldots \subset X_n$ such that X_0 is the empty space and X_n is the full space. The tensor product of filtered spaces is defined in [9] to be

$$X \otimes_f Y := \left\{ \bigcup_{i+j=k} X_i \times Y_j \right\}_{k \in \mathbb{N}},$$

where the subscript f denotes "filtered". In particular, the k-th level of the filtration is given by $(X \otimes_f Y)_k = \cup_{i+j=k} X_i \times Y_j$.

Note, the additive function structure is modeled well by the tensor product of filtered spaces. In other words, the sublevelset persistence of additive functions is just the tensor product of the filtered spaces. Thus if $f_1 \colon X \to \mathbb{R}$ and $f_2 \colon Y \to \mathbb{R}$ are real-valued functions and $f \colon X \times Y \to \mathbb{R}$ is the corresponding additive function over a product space defined by $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$, then we have $f(x_1, x_2) \leq k$ precisely when $f_1(x_1) \leq i$ and $f_2(x_2) \leq j$ for $i + j \leq k$.

Gakhar and Perea's variation of the Künneth formula allows us to say something about the persistent homology of the product of filtered spaces $X \otimes_f Y$. The persistent homology of $X \otimes_f Y$ is given by barcodes, denoted by $\mathrm{bcd}_n(X \otimes_f Y)$. An interval $[b, d) \in \mathrm{bcd}_n(X \otimes_f Y)$ gives the n-dimensional homological birth and death time of some feature. In the paper by Gakhar and Perea [9], they formulate a version of the Künneth formula for persistent homology. This is the formulation we will use throughout the paper.

Theorem 2.5.3 (Persistent Künneth Formula [9])**.** *There is a natural short exact sequence of graded modules*

$$0 \to \bigoplus_{i+j=n} (PH_i(X) \otimes PH_j(Y)) \to PH_n(X \otimes_f Y) \\ \to \bigoplus_{i+j=n} \mathrm{Tor}(PH_i(X), PH_{j-1}(Y)) \to 0.$$

If $H_i(X)$ and $H_j(Y)$ are point-wise finite, then

$$\begin{aligned} &\mathrm{bcd}_n(X \otimes_f Y) \\ = &\bigsqcup_{i+j=n} \{(\ell_J + I) \cap (\ell_I + J) \mid I \in \mathrm{bcd}_i(X), J \in \mathrm{bcd}_j(Y)\} \\ &\sqcup \bigsqcup_{i+j=n} \{(r_J + I) \cap (r_I + J) \mid I \in \mathrm{bcd}_i(X), J \in \mathrm{bcd}_{j-1}(Y)\} \\ = &\bigsqcup_{i+j=n} \{[\ell_I + \ell_J, \min(\ell_J + r_I, \ell_I + r_J)) \mid I \in \mathrm{bcd}_i(X), J \in \mathrm{bcd}_j(Y)\} \\ &\sqcup \bigsqcup_{i+j=n} \{[\max(\ell_I + r_J, \ell_J + r_I), r_I + r_J) \mid I \in \mathrm{bcd}_i(X), J \in \mathrm{bcd}_{j-1}(Y)\}. \end{aligned}$$

Here ℓ and r are the left and right endpoints of the interval.

This formula allows us to calculate the persistent homology of $X \otimes Y$ using the persistent homology of X and Y. Hence, we can extend this to understand the persistent homology of additive functions over a product space based on the persistent homology of each component. We can obtain n-dimensional barcodes two different ways. The first is through standard combinations, bars with birth and death times given by $[\ell_I + \ell_J, \min(\ell_J + r_I, \ell_I + r_J))$ obtained

by combining i-dimensional bars from X with j-dimensional bars from Y. The second type of n-dimensional bars are called torsion bars with birth and death times given by $[\max(\ell_I + r_J, \ell_J + r_I), r_I + r_J)$ obtained by combining i-dimensional bars from X with $(j-1)$-dimensional bars from Y. The torsion portion originates from the idea that two lower dimensional features can combine in such a way to create a higher dimension feature.

As an example, consider Figure 2.5. The nine 0-dimensional bars (shown in red) are created by combining a 0-dimensional bar from 1-2-2-1 with another 0-dimensional bar from 1-2-2-1. These are examples of standard bars from the Künneth formula. For examples of the torsion portion, consider the four 1-dimensional bars (shown in blue) born earliest. Those 1-dimensional bars at energy height (E) are created upon the death of the four components shown at height (D). For a more precise look at this process, see Section 3.2. In Chapter 3 we use this formula to calculate the persistent homology of large branched alkanes using the persistent homology of smaller molecules. This allows us to compute the sublevelset persistent homology of any branched alkane.

Now that we have an understanding of some chemistry, Morse theory, sublevelset persistence, and the Künneth formula, in the next chapter we discuss how to characterize the energy landscape of branched alkanes. These tools will allow us to determine the sublevelset persistence for any branched alkane.

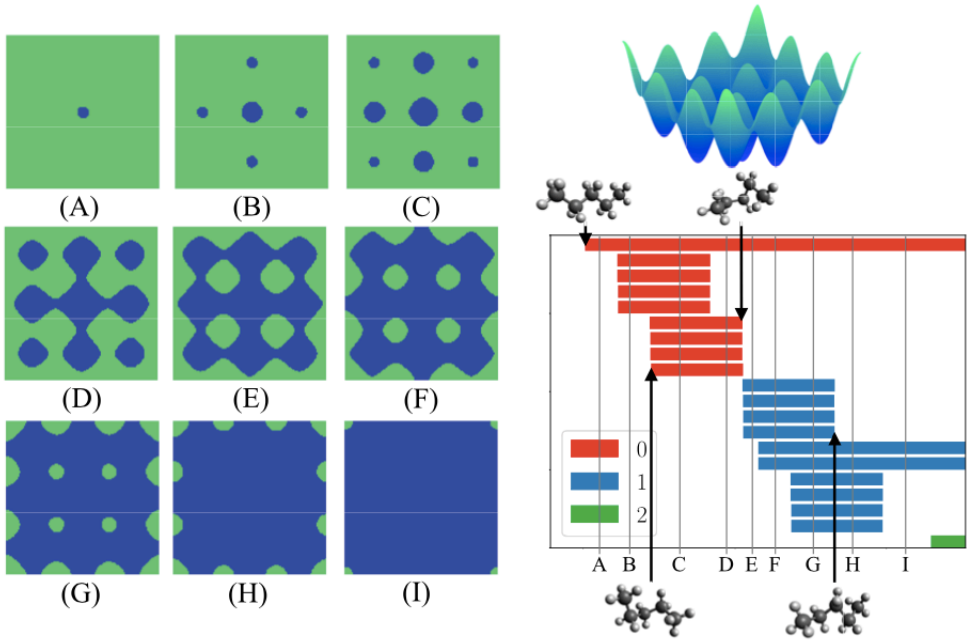

Figure 2.5: [Left] Sublevelsets of the pentane energy landscape, drawn in green, given by $f^{-1}(-\infty, r] := \{x \in (S^1)^2 \mid f(x) \leq r\}$. [Right Top] The energy landscape for pentane. [Right Bottom] The persistence barcodes corresponding to the energy landscape above. Figure from Mirth et al. [2]

Chapter 3

Analytical description of branched alkanes

We would like to deconstruct the energy landscape of any branched alkane into a sum of the energy landscapes of smaller components. As described in Section 2.1, this paper will focus specifically on the Optimized Potentials for Liquid Simulations United Atom or OPLS-UA model of energy landscapes. In this chapter, we will look at a way to approximate the shape of energy landscape of larger branched alkanes.

3.1 Energy landscapes of building block bonds

The energy of any molecule depends on the dihedral angles of the internal bonds. Recall from Section 2.1 that if a molecule is rotated around an internal bond, x-y, the positioning of the adjacent carbons determine whether the energy is higher or lower. Thus, dihedral angles are constructed by paths of four adjacent carbons; we denote each dihedral angle by w-x-y-z where each variable w, x, y, z is the degree of the carbon, and x-y is the internal bond. For examples of different dihedral angles, see Figure 3.1. We denote the carbons labeled with degrees w and z as the adjacent carbons. The energy function associated to dihedral angle w-x-y-z, is $\mathrm{V}_{\text{w-x-y-z}}\colon S^1 \to \mathbb{R}$, which depends on the degrees of the four carbon atoms in the path. Thus, each internal bond i has an energy function, $g_i\colon S^1 \to \mathbb{R}$, which is a sum of all dihedral angles containing the internal bond x-y in the middle of the path, with angular offsets. Then, the OPLS-UA energy landscape of the molecule is given by $f\colon (S^1)^n \to \mathbb{R}$, where n is the number of internal bonds, and where f is given by $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} g_i(x_i)$. We will see an example of $f\colon (S^1)^2 \to \mathbb{R}$ below.

In order to fully characterize each internal bond, we would need the energy function of each dihedral type w-x-y-z where $w, z \in \{1, 2, 3, 4\}$ and $x, y \in \{2, 3, 4\}$. After accounting for symmetry, this gives 78 different dihedral types. Finding each dihedral type is computationally intensive and as such, we will start by using a subset of the dihedral types to

approximate the energy landscapes of larger molecules. To simplify matters, we will assume that the dihedral type does not depend on the degrees of the outer carbons. We will used the term "idealized" to reference this assumption. The approximation will replace dihedral types w-x-y-z with the dihedral types 1-x-y-1. For example, the energy landscape for 1-2-2-1 is very similar to the energy landscape of 2-2-2-2.

Note, this is not an ideal approximation; as the degrees of carbons w and z increase, the energy function of 1-x-y-1 is a worse approximation of w-x-y-z. But, this will serve as a starting point for approximating the energy landscapes of branched alkanes. One could remove this assumption and proceed through the same process we outline below. Therefore, we only consider molecules whose dihedral angles are all of type 1-x-y-1. There are 6 alkanes that consist only of these dihedral types, see Figure 3.1. For each path of length four, we give the energy landscape as a sum of functions $\mathrm{V}_{\text{1-x-y-1}} : S^1 \to \mathbb{R}$ where $V_{1\text{-}x\text{-}y\text{-}1}$ is the energy function for the dihedral angle,

$$V_{1\text{-}x\text{-}y\text{-}1}(\phi) = \sum_{i=0}^{5} c_i \left(1 + (-1)^{i+1} \cos(i(\phi - \theta))\right),$$

and θ is the angle between the leaf carbons and the angle bisector (see Figure 3.4). Here, each c_i is a constant that depends on the internal bond $x - y$.

For example, the molecule isopentane (the upper, center molecule in Figure 3.1) has two paths of length four, both of dihedral type 1-3-2-1. As such we can write its energy function as

$$\begin{aligned} f_{\text{1-3-2-1}}(\phi) =& V_{\text{1-3-2-1}}(\phi - 56^\circ) + V_{\text{1-3-2-1}}(\phi + 56^\circ) \\ =& [c_0 + c_{11}(1 + \cos(\phi - 56^\circ)) + c_{12}(1 - \cos(2(\phi - 56^\circ))) + c_3(1 + \cos(3(\phi - 56^\circ)))] \\ &+ [c_0 + c_{11}(1 + \cos(\phi + 56^\circ)) + c_{12}(1 - \cos(2(\phi + 56^\circ))) + c_3(1 + \cos(3(\phi + 56^\circ)))] \end{aligned}$$

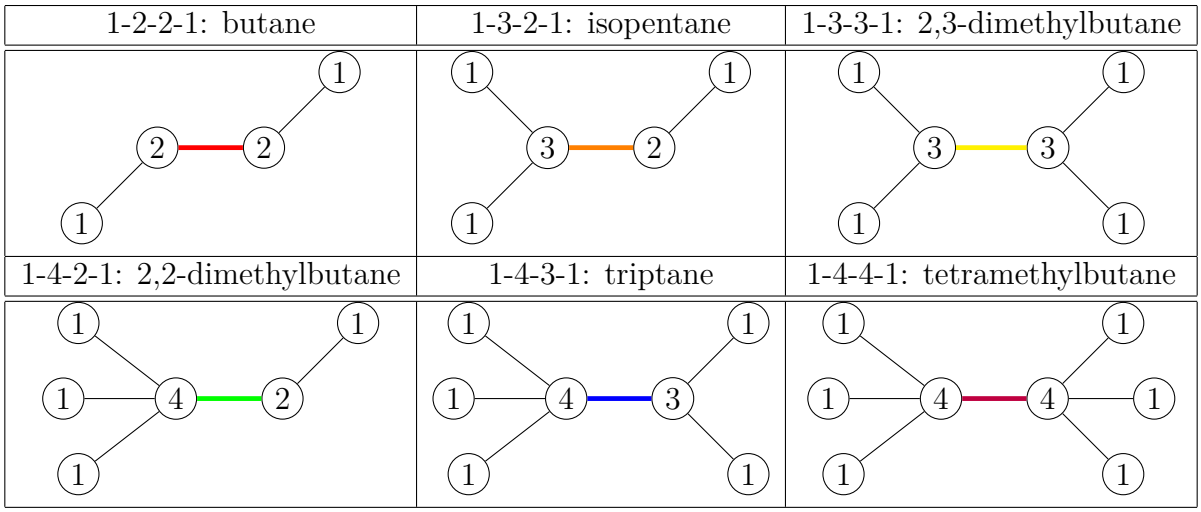

Figure 3.1: Each of these six molecules has an OPLS-UA energy landscape that is a real-valued function $f_{1\text{-}x\text{-}y\text{-}1} : S^1 \to \mathbb{R}$, where the circle S^1 encodes the dihedral angle of a particular type of bond.

with $c_0 = -0.4992$, $c_{11} = 0.8525$, $c_{12} = -0.2224$, and $c_3 = 0.8774$, all in kcal/mol [14]. The 56° is the angle bisector between the two carbons, which roughly comes from the tetrahedral structure of the carbon atom.

Just as butane's energy landscape was used as the single building block for the energy landscapes of the n-alkanes, these six molecules will serve as the building blocks of the branched alkanes. As such, we will refer to these six molecules as the *building block bonds*. For each of the six building block bonds, we plot the energy landscapes $f_{1\text{-}x\text{-}y\text{-}1}$ on the circle in Figure 3.2. Each energy function is generated by code written by Sadhu, which can be found at [21].

The solid colored lines represent the total OPLS-UA energy landscape $f_{1\text{-}x\text{-}y\text{-}1}$ for the molecule. The thinner patterned lines represent the component functions $V_{1\text{-}x\text{-}y\text{-}1}$ that make up the energy landscape. All of the energy landscapes share some characteristics. Each has exactly three minimums and three maximums for a total of six critical points. There are also some additional similarities between a few of the different landscapes. Note, the plot of the 1-2-2-1 molecule looks very similar to that of of the 1-3-3-1 molecule, as each has one global minimum, 2 local minima, 2 local maxima, and 1 global maximum. By contrast, the

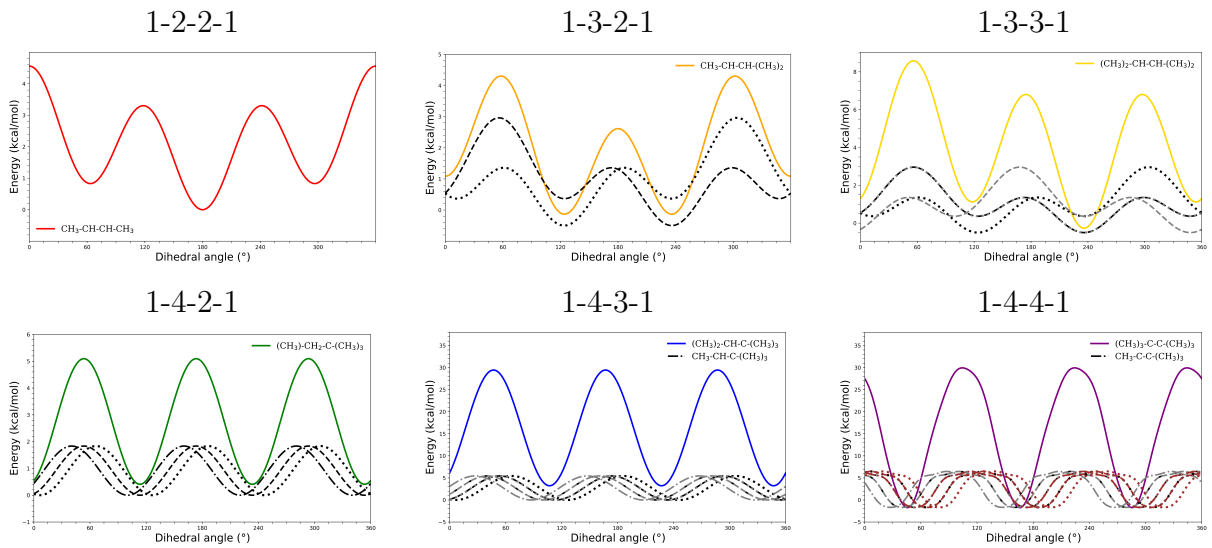

Figure 3.2: Energy landscapes $f_{1\text{-}x\text{-}y\text{-}1}$ for dihedral types 1-x-y-1 where the x-axis is the bond angle and the y-axis is the energy value.

1-4-2-1, 1-4-3-1, and the 1-4-4-1 all have 3 global minima and 3 global maxima. This is due in part to the tetrahedral structure of carbon bonds and the rotational symmetry of carbon.

Next, we compute the sublevelset persistence barcodes of the six building block bond energy landscapes $f_{1\text{-}x\text{-}y\text{-}1}$. We use the GUDHI software package [22]; our code, found here [21], is based off of the code from [2] which can be found at [23].

We discretize each copy of the circle S^1 to be a circular ring with 63 vertices and 63 edges, chosen since 63 is approximately $10 \cdot 2\pi$. Then, we compute the sublevelset persistent homology of the OPLS-UA energy function $f_{1\text{-}x\text{-}y\text{-}1}$ on those vertices and edges. This gives us the barcodes for the sublevelset persistent homology. We can use this same process for the higher dimensional molecules, which we study in later chapters. The similarities between the shapes of the energy landscapes mentioned above can also be seen in the persistence barcodes. The barcodes for 1-2-2-1 and 1-3-3-1 look similar combinatorially, and the barcodes for 1-4-2-1, 1-4-3-1, and 1-4-4-1 do as well. Using the six building block bonds as the components of larger molecules, we can combine their energy landscapes to construct the energy landscape of any branched alkane.

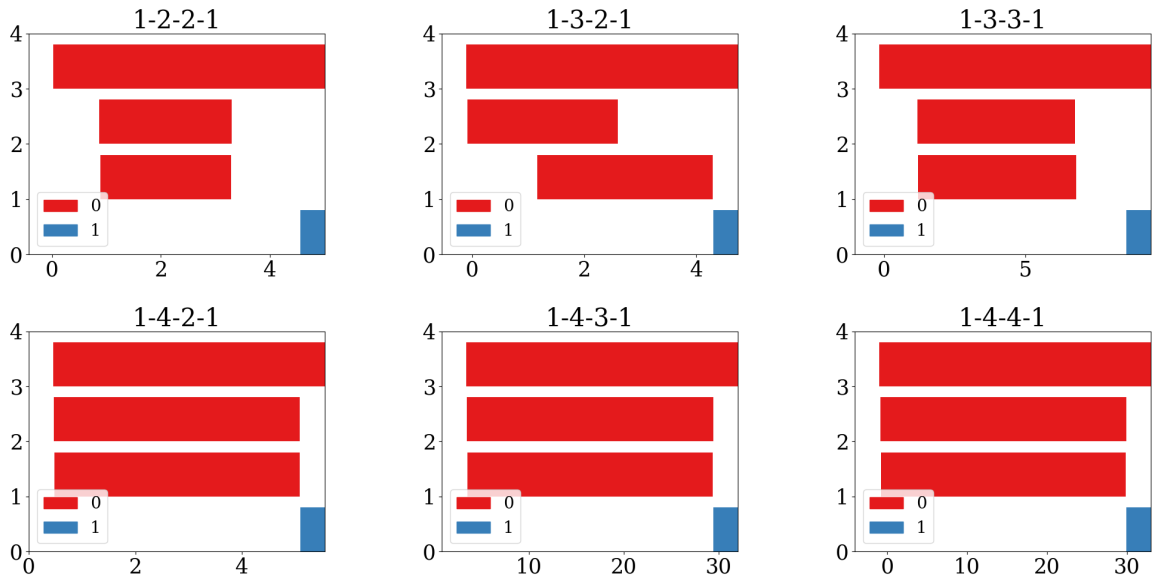

Figure 3.3: Persistent homology bars for dihedral types 1-x-y-1.

3.2 Two examples of branched alkanes

We study two different examples that outline the construction of higher degree energy landscapes from the components of the smaller energy landscapes. First, we will look at 2-methylpentane, the molecule acquired by combining a 3-2 internal bond with a 2-2 internal bond. Second, we will look at 2,2-dimethylpentane, the molecule acquired by combining internal bonds 4-2 and 2-2. Once we study both examples, we compare the sublevelset persistence of the two molecules.

3.2.1 2-methylpentane

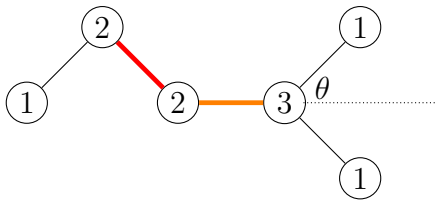

Figure 3.4: A picture of 2-methylpentane, the molecule consisting of building block bonds 1-2-2-3 and 1-3-2-2, whose energy functions we will approximate with the energy functions for 1-2-2-1 and 1-3-2-1, respectively. The colored, thicker bonds correspond to the different dihedral angles. The dotted lines are the angle bisectors between the two leaf carbons, and θ denotes the angle between the bisector and the leaf carbons.

As a first example, consider the molecule 2-methylpentane shown in Figure 3.4. Just like the energy landscape for hexane is a combination of the energy landscapes of pentane and butane, 2-methylpentane can be viewed as a combination of two internal bonds, where one bond has one 1-2-2-3 dihedral type and the other internal bond consists of two 1-3-2-2 dihedral types. We can see this combination in the energy landscape of 2-methylpentane in Figure 3.5. If we take a slice normal to the ϕ_2 axis, we capture the curve given by butane in Figures 3.1 and 3.2. Similarly, if we take a slice normal to the ϕ_1 axis, we capture the isopentane energy landscape given in Figures 3.1 and 3.2. Note, to fully characterize the 3-2 (orange) bond in Figure 3.4, we need two dihedral angles of type 1-3-2-2. It is important to recall from the beginning of this chapter that the energy function for a dihedral angle does not just depend on the degree of the carbons on each side; instead it also depends on the carbons that are one edge away. But as mentioned earlier, we will use 1-x-y-1 to approximate w-x-y-z. Hence, we will use the energy function for 1-3-2-1 to approximate the energy function for dihedral type 1-3-2-2, and similarly we will use 1-2-2-1 for dihedral type 1-2-2-3.

Using our approximation, the idealized energy function for 2-methylpentane, $f\colon (S^1)^2 \to \mathbb{R}$, is the sum of the energy functions of building block bonds 1-2-2-1 and 1-3-2-1.

$$\begin{aligned} f(\phi_1,\phi_2) =& \mathrm{V}_{\text{1-2-2-1}}(\phi_1) + [\mathrm{V}_{\text{1-3-2-1}}(\phi_2+\theta) + \mathrm{V}_{\text{1-3-2-1}}(\phi_2-\theta)] \\ =& c_0 + c_{11}(1+\cos(\phi_1)) + c_{12}(1-\cos(2(\phi_1))) + c_3(1+\cos(3(\phi_1))) \\ &+ \big[c_0' + c_1'(1+\cos(\phi_2-56^\circ)) + c_2'(1-\cos(2(\phi_2-56^\circ))) \\ &+ c_3'(1+\cos(3(\phi_2-56^\circ))) + c_0' + c_1'(1+\cos(\phi_2+56^\circ)) \\ &+ c_2'(1-\cos(2(\phi_2+56^\circ))) + c_3'(1+\cos(3(\phi_2+56^\circ)))\big] \end{aligned}$$

where $V_{x\text{-}y\text{-}z\text{-}w}$ is the energy function for the dihedral angle,

$$V_{x\text{-}y\text{-}z\text{-}w}(\phi) = \sum_{i=0}^{5} c_i(1+(-1)^{i+1}\cos(i(\phi-\theta)),$$

and θ is the angle between the leaf carbons and the angle bisector. For 1-2-2-1 the coefficients are $c_0 = 0$, $c_{11} = 0.7059$, $c_{12} = -0.1355$, and $c_3 = 1.5735$, and for 1-3-2-1 they are $c'_0 = -0.4992$, $c'_1 = 0.8525$, $c'_2 = -0.2224$, and $c'_3 = 0.8774$. The energy landscape is given in Figure 3.5 along with sublevelsets at nine different energy values, which correspond to values right after the birth of certain homological features.

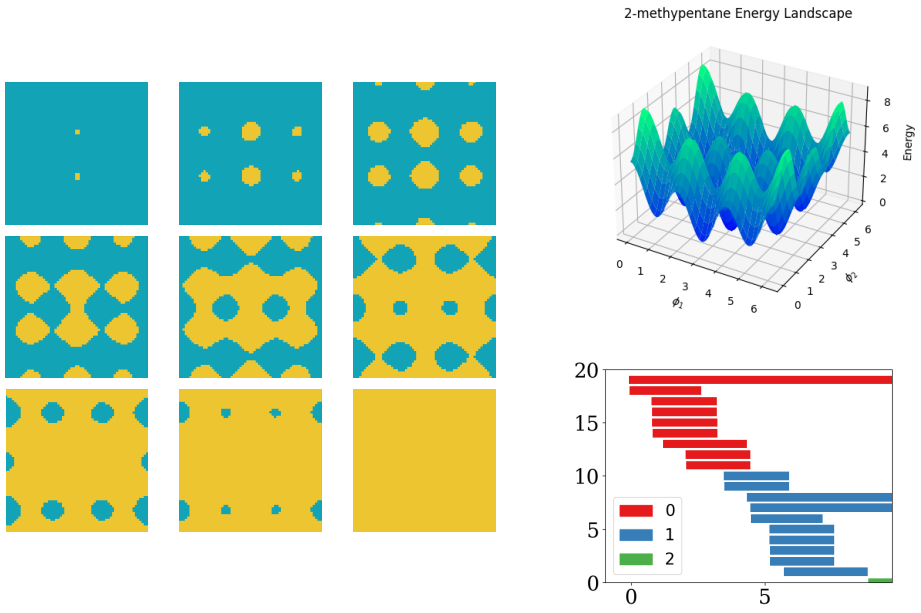

Figure 3.5: [Left] Sublevelsets of the 2-methylpentane energy landscape. [Right Top] The 2-methylpentane energy landscape and [Right Bottom] its persistence barcodes.

Now we introduce some notation. Each building block bond $f_{1\text{-}x\text{-}y\text{-}1}$ is a function on the circle with six critical points. We will write any critical point as p_{ij}, where we replace p with a if it is a global minimum, b if it is a local minimum, c if it is a local maximum, or d if it is a global maximum. This accounts for all types of critical points for the building block bonds. Let i denote the component function of the critical point and let j denote the label of critical point in that component function. Thus, let $i = 1$ correspond to 2-2, $i = 2$ to 3-2, $i = 3$ to 3-3, $i = 4$ to 4-2, $i = 5$ to 4-3, and $i = 6$ to 4-4. Each j depends on the associated component function and whether $p = a$, b, c, or d. For example, for internal bond of type 4-3, (namely $i = 6$), we have critical points a_{6j} with $j = 1, 2, 3$ since there are 3 global minima, d_{6j} with

$j = 1, 2, 3$ since there are 3 global maxima, no critical points of form b_{6j} as there are no local minima, and no critical points of form c_{6j}, as there are no local maxima.

Now that we have an approximation of the energy landscape for 2-methylpentane and a naming convention for critical points, we can examine the sublevelset persistent barcode for this molecule. One bond has a single dihedral angle of type 1-2-2-3, which we simplify to 1-2-2-1. The other bond has two dihedral angles of type 1-3-2-2, both of which we simplify to 1-2-2-1. Let a_{11}, b_{11}, b_{12}, c_{11}, c_{12}, and d_{11} be the critical points of the energy landscape for 1-2-2-1 and let a_{21}, a_{22}, b_{21}, c_{21}, d_{21}, and d_{22} be the critical points of the energy landscape for 1-3-2-1.

Since each energy function is a Morse function, the critical points of that function correspond to birth and death times of bars (Lemma 2.4.1), such as in Figure 2.4 in Section 2.4. Thus, we can identify the length, birth, and death time of each bar. This characterization gives the table of values in Table **??** and the right-bottom barcode in Figure 3.5.

Table 3.1: 2-methylpentane Morse complex computation. By "death r" in the effect column, we mean that the critical point p kills a persistent homology bar that was born at energy r. The notation $f(p)$ denotes the energy value at critical point p.

$p =$	index	∂p	$f(p)$	effect
(a_{11}, a_{21})	0	0	$\alpha_1 + \alpha_2$	birth
(a_{11}, a_{22})	0	0	$\alpha_1 + \alpha_2$	birth
(b_{11}, a_{21})	0	0	$\beta_1 + \alpha_2$	birth
(b_{12}, a_{21})	0	0	$\beta_1 + \alpha_2$	birth
(b_{11}, a_{22})	0	0	$\beta_1 + \alpha_2$	birth
(b_{12}, a_{22})	0	0	$\beta_1 + \alpha_2$	birth
(a_{11}, b_{21})	0	0	$\alpha_1 + \beta_2$	birth
(b_{11}, b_{21})	0	0	$\beta_1 + \beta_2$	birth
(b_{12}, b_{21})	0	0	$\beta_1 + \beta_2$	birth
(a_{11}, c_{21})	1	$(a_{11}, a_{21}) + (a_{11}, a_{22})$	$\alpha_1 + \gamma_2$	death $\alpha_1 + \alpha_2$
(c_{11}, a_{21})	1	$(a_{11}, a_{21}) + (b_{11}, a_{21})$	$\gamma_1 + \alpha_2$	death $\beta_1 + \alpha_2$
(c_{12}, a_{21})	1	$(a_{11}, a_{21}) + (b_{12}, a_{21})$	$\gamma_1 + \alpha_2$	death $\beta_1 + \alpha_2$
(c_{11}, a_{22})	1	$(a_{11}, a_{22}) + (b_{11}, a_{22})$	$\gamma_1 + \alpha_2$	death $\beta_1 + \beta_2$
(c_{12}, a_{22})	1	$(a_{11}, a_{22}) + (b_{12}, a_{22})$	$\gamma_1 + \alpha_2$	death $\beta_1 + \beta_2$
(b_{11}, c_{21})	1	$(b_{11}, a_{21}) + (b_{11}, a_{22})$	$\beta_1 + \gamma_2$	birth
(b_{12}, c_{21})	1	$(b_{12}, a_{21}) + (b_{12}, a_{22})$	$\beta_1 + \gamma_2$	birth
(a_{11}, d_{21})	1	$(a_{11}, a_{21}) + (a_{11}, b_{21})$	$\alpha_1 + \delta_2$	death $\alpha_1 + \beta_2$
(a_{11}, d_{22})	1	$(a_{11}, a_{22}) + (a_{11}, b_{21})$	$\alpha_1 + \delta_2$	birth
(c_{11}, b_{21})	1	$(b_{12}, b_{11}) + (a_{11}, b_{11})$	$\gamma_1 + \beta_2$	death $\beta_1 + \beta_2$
(c_{12}, b_{21})	1	$(b_{12}, b_{12}) + (a_{11}, b_{12})$	$\gamma_1 + \beta_2$	death $\beta_1 + \beta_2$
(d_{11}, a_{21})	1	$(b_{11}, a_{21}) + (b_{12}, a_{21})$	$\beta_1 + \delta_2$	birth
(d_{11}, a_{22})	1	$(b_{11}, a_{22}) + (b_{12}, a_{22})$	$\beta_1 + \delta_2$	birth
(d_{11}, a_{21})	1	$(b_{11}, a_{21}) + (b_{12}, a_{21})$	$\beta_1 + \delta_2$	birth
(d_{11}, a_{22})	1	$(b_{11}, a_{22}) + (b_{12}, a_{22})$	$\beta_1 + \delta_2$	birth
(b_{11}, d_{21})	1	$(b_{11}, b_{21}) + (b_{11}, a_{21})$	$\beta_1 + \delta_2$	birth
(b_{12}, d_{21})	1	$(b_{12}, b_{21}) + (b_{12}, a_{21})$	$\beta_1 + \delta_2$	birth
(d_{11}, b_{21})	1	$(b_{11}, b_{21}) + (b_{12}, b_{21})$	$\delta_1 + \beta_2$	birth
(c_{11}, c_{21})	2	$(a_{11}, c_{21}) + (b_{11}, c_{21}) + (c_{11}, a_{21}) + (c_{11}, a_{22})$	$\gamma_1 + \gamma_2$	death $\beta_1 + \gamma_2$
(c_{12}, c_{21})	2	$(a_{11}, c_{21}) + (b_{12}, c_{21}) + (c_{12}, a_{21}) + (c_{12}, a_{22})$	$\gamma_1 + \gamma_2$	death $\beta_1 + \gamma_2$
(d_{11}, c_{21})	2	$(b_{11}, c_{21}) + (b_{12}, c_{21}) + (d_{11}, a_{21}) + (d_{11}, a_{22})$	$\delta_1 + \gamma_2$	death $\delta_1 + \alpha_2$
(c_{11}, d_{22})	2	$(a_{11}, d_{22}) + (b_{11}, d_{22}) + (c_{11}, a_{22}) + (c_{11}, b_{21})$	$\gamma_1 + \delta_2$	death $\beta_1 + \delta_2$
(c_{12}, d_{22})	2	$(a_{11}, d_{22}) + (b_{12}, d_{22}) + (c_{12}, a_{22}) + (c_{12}, b_{21})$	$\gamma_1 + \delta_2$	death $\beta_1 + \delta_2$
(c_{11}, d_{21})	2	$(a_{11}, d_{21}) + (b_{11}, d_{21}) + (c_{11}, a_{21}) + (c_{11}, b_{21})$	$\gamma_1 + \delta_2$	death $\beta_1 + \delta_2$
(c_{12}, d_{21})	2	$(a_{11}, d_{21}) + (b_{12}, d_{21}) + (c_{12}, a_{21}) + (c_{12}, b_{21})$	$\gamma_1 + \delta_2$	death $\beta_1 + \delta_2$
(d_{11}, d_{21})	2	$(b_{11}, d_{21}) + (b_{12}, d_{21}) + (d_{11}, a_{21}) + (d_{11}, b_{21})$	$\delta_1 + \delta_2$	death $\delta_1 + \beta_2$
(d_{11}, d_{22})	2	$(b_{11}, d_{22}) + (b_{12}, d_{22}) + (d_{11}, a_{22}) + (d_{11}, b_{21})$	$\delta_1 + \delta_2$	birth

3.2.2 2,2-dimethylpentane

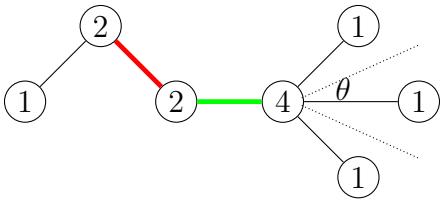

Figure 3.6: A picture of 2,2-dimethylpentane, the molecule consisting of building block bonds 1-2-2-4 and 1-4-2-2 which we will approximate with 1-2-2-1 and 1-4-2-1 respectively.

As a different example, consider 2,2-dimethylpentane, the molecule with internal bonds 4-2 and 2-2. The 4-2 internal bond has three dihedral angles of type 1-4-2-2, each of which we simplify to 1-4-2-1. The 2-2 internal bond has a single dihedral angle of type 1-2-2-4, which we simplify to 1-2-2-1.

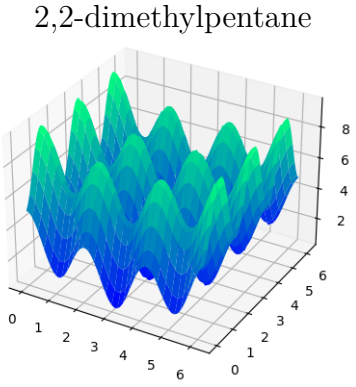

Figure 3.7: The energy landscape for 2,2-dimethylpentane where the x and y axis denote the bond angle for each internal bond and the z axis denotes the corresponding energy value.

The energy landscape of 2,2-dimethylpentane can be found in Figure 3.7. Instead of calculating each birth and death time by hand, we use our code [21] to produce the bottom plot in Figure 3.8. We can also summarize the persistent homology information by looking at how the bars of the building block bonds combine to form the bars for the larger molecule: see the bar labels in Figure 3.8. Each of the bars in the component barcodes are labeled with a letter and number to denote which barcode they originate from and the bar from that

barcode. In the sublevelset persistence barcode for 2,2-dimethylpentane, each bar is labeled with the two component bars that created it in Theorem 2.5.3, the persistent Künneth formula. The bars that come from torsion are labeled with "TOR". For example, the semi-infinite 0-dimensional bar in the 2,2-dimethylpentane barcode (I1+K1) is created by the 0-dimensional semi-infinite bars from 1-2-2-1 (I1) and 1-4-2-1 (K1).

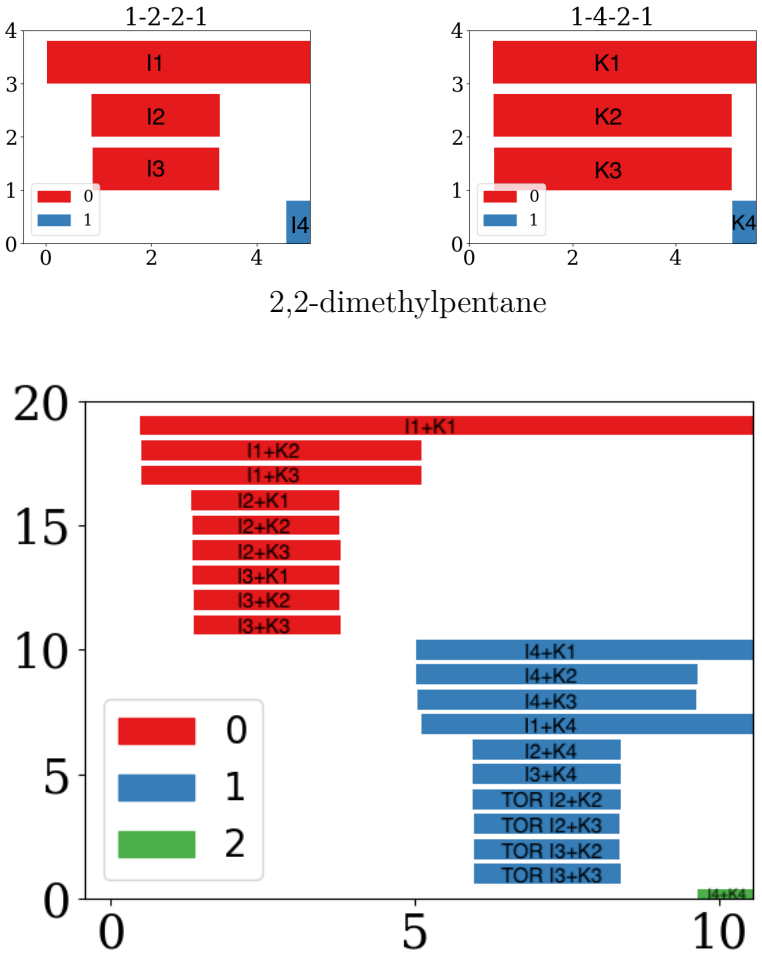

Figure 3.8: Labeled persistence barcodes for 1-2-2-1, 1-4-2-1, and how they combine to form the persistence barcode for 1-2-2-1/1-4-2-1.

Although the birth and death times of each bar vary in the barcodes for 2-methylpentane and 2,2-dimethylpentane, we can see that both of these molecules with two building block bonds have the same number of bars in every dimension. Further, they have the same

number of semi-infinite bars. This is not unique to the two molecules we chose, instead, it is a byproduct of the fact that there are 6 critical points for all six building block bond energy functions on the circle: 3 minimums and 3 maximums for each building block bond. In the next chapter, this type of analysis allows us to determine the exact number of bars, both finite and semi-infinite, for any branched alkane energy landscape. We will also be able to determine the number of bars of each dimension and of each length in a barcode.

Chapter 4

Characterizing the sublevelset persistence of branched alkanes

Now that we have looked at a few examples, we can start to characterize the sublevelset persistence of the branched alkanes. The first results will count the total number of semi-infinite bars and finite bars. Next, we will count the number of bars in each homological dimension. Finally, we will count the number of bars of each length.

4.1 The number of finite and semi-infinite bars

We begin by counting the number of semi-infinite bars. Even though this is a straightforward result, we include it for completeness. See also Mirth et al. [2] for the case of non-branched alkanes.

Lemma 4.1.1. *The energy landscape for any branched alkane with n internal bonds, $f\colon (S^1)^n \to \mathbb{R}$, has $\binom{n}{k}$ semi-infinite bars in dimension k.*

Proof. The domain of our function is $(S^1)^n$. This corresponds to the n-dimensional torus which has k-dimensional homology of rank $\binom{n}{k}$. The semi-infinite bars in a filtration of a space X simply recover the homology of X, and here we have $X = (S^1)^n$. Therefore, the energy landscape also has $\binom{n}{k}$ semi-infinite bars of dimension k. □

Counting the total number of finite and semi-infinite bars depends on the shape of the energy functions and the number of critical points obtained by each energy function. As noted in Chapter 3, for our simplified case each base dihedral type has exactly 6 critical points: 3 minimums and 3 maximums. As a generalization that can be used for more complex cases, we generalize our results to work for any Morse function with p_i critical points.

Theorem 4.1.2. *The energy function of any branched alkane, $f\colon (S^1)^n \to \mathbb{R}$ is given by $f(\phi_1, \ldots, \phi_n) = g_1(\phi_1) + \ldots + g_n(\phi_n)$ with each $g_i\colon S^1 \to \mathbb{R}$ having p_i critical points. Thus, we have that the total number of bars in the sublevelset persistent homology of f is*

$$\textit{total \# of bars} = (2^n)\ \textit{semi-infinite bars} + \left(\frac{\prod_{i=1}^{n} p_i - 2^n}{2} \right)\ \textit{finite bars.}$$

Proof. Since f is a function on $(S^1)^n$, for each $0 \leq k \leq n$, we have $\binom{n}{k}$ semi-infinite bars in dimension k (Lemma 4.1.1). Thus, we have $\sum_{k=0}^{n} \binom{n}{k} = 2^n$ semi-infinite bars.

Each dihedral-type function, g_i is a Morse function. By Lemma 2.5.1, the sum $g_1(x_1) + \ldots + g_n(x_n)$ over a product space of Morse functions is also a Morse function. Hence f is Morse and as such, the birth and death times of the bars correspond directly to the critical points of f. Since 2^n of those points correspond to the birth times of the semi-infinite bars, we have $\sum_{i=1}^{n} p_i - 2^n$ critical points remaining to be the birth and death times of the finite bars. Next, since these bars are finite, we have that our remaining critical points must be split equally between birth times and death times. It is important to note that the number of critical points of a Morse function on the circle will always be even: for every minimum that is introduced, a maximum is created. Formally, the circle is a 1-dimensional manifold and as such, the only possible critical point indices are 0 and 1. Further, the Euler characteristic for the circle is 0 and can be defined as

$$\sum_k (-1)^k C^k = \chi(S^1),$$

where C^k is the number of critical points of index k. Thus the number of 0-dimensional critical points must be the same as the number of 1-dimensional critical points (see also the Morse inequalities, which can be found on page 74 of [20]). Thus the value $\sum_{i=1}^{n} p_i - 2^n$ is even, which gives us $\frac{1}{2} \sum_{i=1}^{n} p_i - 2^n$ finite bars. Therefore we have

$$2^n + \frac{\prod_{i=1}^{n} p_i - 2^n}{2}$$

bars in the sublevelset persistence barcode for the energy function of any branched alkane.

□

This theorem holds when each g_i corresponds to the energy function of some internal bond of type w-x-y-z. We can also use the above theorem for the idealized case where $w = z = 1$. Since every energy function $\mathrm{V}_{\text{1-x-y-1}}$ has 6 critical points, the total number of bars for branched alkanes is exactly that of the n-alkane case, which has $\frac{6^n+2^n}{2}$ bars.

4.2 The number of bars in dimension k

Now that we have looked at the total number of bars, we count the number of bars in each dimension. We look at the idealized case with each w-x-y-z approximated by 1-x-y-1.

Theorem 4.2.1. *The sublevelset persistent homology on any analytical branched alkane with n internal bonds with potential energy landscape $f\colon (S^1)^n \mapsto \mathbb{R}$ has $\binom{n}{k} + (3^n - 1)\binom{n-1}{k}$ persistent homology bars in dimension k.*

This theorem is nearly identical to Theorem 2 in [2]. But instead of induction on k, we use induction on n to prove our result. We consider induction on n to reinforce the view that energy landscapes are nested functions. We build up large molecules via induction by adding one bond at a time, which is a perspective that chemists are interested in. For example, in 2-methylpentane, the molecule consisting of dihedral types 1-3-2-1 and 1-2-2-1, the energy landscapes of both types can be found by slicing the energy landscape of 2-methylpentane parallel to the axes of the domain.

Proof. We proceed by induction on n, the number of bonds.

Base case: $n = 1$

For any building block bond we have that the number of 0-dimensional bars is $3 = \binom{1}{0} + (3^1 - 1)\binom{1-1}{0}$ and the number of 1-dimensional bars is $1 = \binom{1}{1} + (3^1 - 1)\binom{1-1}{1}$.

Inductive step: Assume true for n bonds

Now, suppose that for n bonds, the number of k-dimensional bars is given by $\binom{n}{k} + (3^n - 1)\binom{n-1}{k}$. Hence, we must show that for $n+1$ bonds, the number of bars is $\binom{n+1}{k} + (3^{n+1} - 1)\binom{n}{k}$. Consider a molecule with n bonds to which we attach one new bond. Recall that the second part of the persistent Künneth formula outlines the possible ways to obtain k-dimensional bars from two component barcodes. In particular, we have

$$\sum_{i+j=k} |[\ell_I + \ell_J, \min(\ell_J + r_I, \ell_I + r_J)) \mid I \in \mathrm{bcd}_i(X), J \in \mathrm{bcd}_j(Y)|$$
$$+ \sum_{i+j=k} |[\max(\ell_I + r_J, \ell_J + r_I), r_I + r_J) \mid I \in \mathrm{bcd}_i(X), J \in \mathrm{bcd}_{j-1}(Y)|$$

ways to obtain bars of dimension k. Note, the sublevelset persistent homology of a function on the circle (the single bond) can only have bars in dimensions 0 and 1, which restricts the number of ways to obtain k-dimensional bars. Hence, we can obtain k-dimensional bars by combining 0-dimensional bars with k-dimensional bars, 1-dimensional bars with $(k-1)$-dimensional bars, or $(k-1)$-dimensional bars with 0-dimensional bars via torsion.

There are $\binom{n}{k} + (3^n - 1)\binom{n-1}{k}$ ways to get a dimension k bar by combining each k-dimensional bar from f with a 0-dimensional bar. Since each building block bond has three 0-dimensional bars, we have 3 times that amount. Next, there are $\binom{n}{k-1} + (3^n - 1)\binom{n-1}{k-1}$ ways to obtain a k-dimensional bar by combining the 1-dimensional semi-infinite bar of any building block bond with the $(k-1)$-dimensional bars given by the map $(S^1)^n \to \mathbb{R}$. Finally, to account for the torsion bars, we want to combine the finite length $(k-1)$-dimensional bars given by the map $(S^1)^n \to \mathbb{R}$ with the 0-dimensional finite length bars from any building block bond, of which there are 2. Thus, there are $2\Big(\binom{n}{k-1} + (3^n - 1)\binom{n-1}{k} - \binom{n}{k-1}\Big) = 2\Big((3^n - 1)\binom{n-1}{k}\Big)$ k-dimensional bars created by the torsion part of the Künneth product.

Hence, the number of k dimensional bars for the map $(S^1)^{n+1} \to \mathbb{R}$ is

$$
\begin{aligned}
3\left[\binom{n}{k}+(3^n-1)\binom{n-1}{k}\right] &+ \left[\binom{n}{k-1}+(3^n-1)\binom{n-1}{k-1}\right] + 2\left[(3^n-1)\binom{n-1}{k-1}\right] \\
&= 3\binom{n}{k} + 3(3^n-1)\binom{n-1}{k} + \binom{n}{k-1} + 3(3^n-1)\binom{n-1}{k-1} \\
&= 3\binom{n}{k} + 3^{n+1}\binom{n-1}{k} - 3\binom{n-1}{k} + \binom{n}{k-1} + 3^{n+1}\binom{n-1}{k-1} - 3\binom{n-1}{k-1} \\
&= 3\binom{n}{k} + \binom{n}{k-1} + 3^{n+1}\left[\binom{n-1}{k}+\binom{n-1}{k-1}\right] - 3\left[\binom{n-1}{k}+\binom{n-1}{k-1}\right] \\
&= 3\binom{n}{k} + \binom{n}{k-1} + 3^{n+1}\binom{n}{k} - 3\binom{n}{k} \\
&= 3\binom{n}{k} + \binom{n}{k-1} + (3^{n+1}-1)\binom{n}{k} - 2\binom{n}{k} \\
&= \binom{n}{k} + \binom{n}{k-1} + (3^{n+1}-1)\binom{n}{k} \\
&= \binom{n+1}{k} + (3^{n+1}-1)\binom{n}{k}.
\end{aligned}
$$

We are done by induction on n. Therefore, the branched alkane energy landscape $f\colon (S^1)^n \to \mathbb{R}$ has $\binom{n}{k} + (3^n-1)\binom{n-1}{k}$ persistent homology bars in dimensional k.

□

In the Theorem 4.1.2, we were able to generalize our results for any branched alkane energy function. Here, the generalization of this theorem to the case where w and z are not simplified to 1 depends specifically on the number of critical points for each internal bond and the persistent Künneth formula, and as such, we omit it from this work. For the remainder of this thesis, we will focus specifically on the idealized case with $w = z = 1$.

As an example, consider the four barcodes in Figure 4.1. Even though each landscape consists of two different bonds, all four energy landscapes have ten 0-dimensional bars, nine 1-dimensional bars, and one 2-dimensional bar, as determined by Theorem 4.2.1.

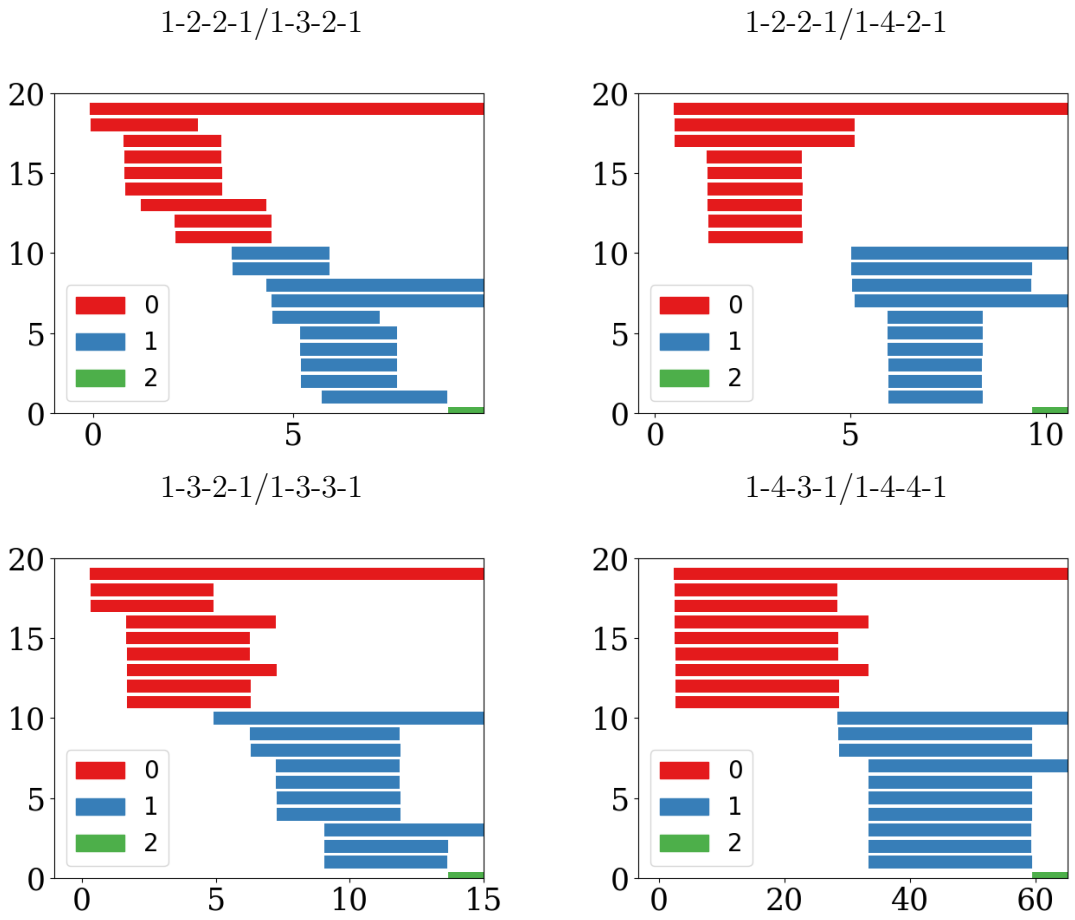

Figure 4.1: The energy landscapes for four different branched alkanes with two internal bonds. As shown in Theorem 4.2.1, each barcode has nine 0-dimensional bars, ten 1-dimensional bars, and one 2-dimensional bar.

4.3 The number of bars of each length

We can also count the number of bars for each possible bar length. The following theorems are based on the persistent Künneth formula, given in Theorem 2.5.3. First, we introduce the case where we have three possible bar lengths over two different component functions. Later, we will generalize the results for n component functions and any number of bar lengths.

Let X_1 and X_2 be two energy landscapes such that their persistence barcodes, $\mathrm{bcd}_n(X_1)$ and $\mathrm{bcd}_n(X_2)$, have two different finite bar lengths $t, u \in \mathbb{R}$ such that $t > u$ along with semi-infinite bars of length ∞. We count the number of bars of each length in $\mathrm{bcd}_n(X_1 \otimes_f X_2)$, the persistence barcode for the product space $X_1 \times X_2$.

Theorem 4.3.1. *Suppose X_1 and X_2 are two energy landscapes such that their persistence barcodes, bcd(X_1) and bcd(X_2), are as follows:*

- *bcd(X_1) has i_1 infinite bars, t_1 finite bars of length t, u_1 finite bars of length u.*
- *bcd(X_2) has i_2 infinite bars, t_2 finite bars of length t, u_2 finite bars of length u.*

Then bcd($X_1 \otimes_f X_2$) has:

- $i_1 i_2$ *infinite bars,*
- $t_1 i_2 + i_1 t_2 + 2t_1 t_2$ *finite bars of length t, and*
- $2u_1 u_2 + u_1 i_2 + i_1 u_2 + 2u_1 t_2 + 2t_1 u_2$ *finite bars of length u.*

Proof. First, the only way to acquire semi-infinite bars in the persistent Künneth formula, Theorem 2.5.3, is from the non-torsion portion by combining semi-infinite bars. Hence, we have $i_1 i_2$ possible combinations and hence, $i_1 i_2$ semi-infinite bars.

Next, consider bars of length t. We have two ways to obtain bars of length t, either through standard (non-torsion) combinations or through torsion. For the standard combinations, note that Theorem 2.5.3 implies that since $t > u$, the only way to obtain a bar of length t is by combining two bars of length t or by combining a bar of length t with a semi-infinite bar. Thus, we get $t_1 t_2 + t_1 i_2 + t_2 i_1$ possible bars of length t. To obtain a torsion bar of length t, we must combine two bars of length t. Hence, there are $t_1 t_2$ possibilities and thus there are $t_1 i_2 + i_1 t_2 + 2t_1 t_2$ bars of length t.

Finally, consider bars of length u. We can obtain non-torsion bars of length u by combining a bar of length u with a bar of any other length. Thus, we have $u_1 u_2 + u_1 i_2 + i_1 u_2 + u_1 t_2 + t_1 u_2$ standard bars of length u. For the torsion bars, we can obtain a length u torsion bar by combining a u length bar with either a t or u length bar. Thus, we get $u_1 t_2 + t_2 u_1 + u_1 u_2$ torsion bars of length u. So in total, we have $2u_1 u_2 + u_1 i_2 + i_1 u_2 + 2u_1 t_2 + 2t_1 u_2$ bars of length u. □

Now we want to generalize the results of Theorem 4.3.1. Before we attempt to count an arbitrary number of bar lengths for an arbitrary number of terms in the product, we first define a function, $\text{count}_n(r, s)$.

Let $\{\text{bcd}(X_q)\}_{q=1}^n$ be a set of barcodes with bar lengths $\{\ell_r\}_{r=0}^m$, i.e. any bar in any barcode $\text{bcd}(X_q)$ has length ℓ_r for some $0 \leq r \leq m$. By convention, we let $\ell_0 = \infty$, and all other lengths are ordered greatest to least (i.e. $\ell_r > \ell_{r+1}$). Let $x_{q,r}$ be the number of bars in $\text{bcd}(X_q)$ with length ℓ_r. This forms a matrix $X_{\text{lengths}} = \{x_{r,q}\}_{r\in\{0,m\}}^{q\in\{1,n\}}$. For integers $0 \leq s, r \leq m$, define

$$\text{count}_n(r, s) := \sum_{i_1=s}^{r} \dots \sum_{i_n=s}^{r} (x_{i_1,1}) \cdot \dots \cdot (x_{i_n,n})$$

to be the nonnegative integer that is the sum of n-fold products. Further, if $s > r$, define $\text{count}_n(r, s) = 0$.

Theorem 4.3.2. *Let $X_1, \dots, X_n$ be a set of energy landscapes. Let $\{bcd(X_q)\}_{q=1}^n$ be the corresponding set of barcodes with bar lengths $\{\ell_r\}_{r=0}^m$, where $\ell_0 = \infty$ and all other lengths are ordered greatest to least (i.e. $\ell_r > \ell_{r+1}$). Let $x_{q,r}$ be the number of bars in $bcd(X_q)$ with length ℓ_r. Then, the number of bars of length ℓ_r in $bcd(X_1) \otimes_f \cdots \otimes_f bcd(X_q)$ is*

$$\text{count}_n(r, 0) - \text{count}_n(r-1, 0) + \text{count}_n(r, 1) - \text{count}_n(r-1, 1).$$

Proof. Recall that the persistent Künneth formula (Theorem 2.5.3) outlines the birth and death times of the barcodes for the product space via the barcodes of the component spaces. We look at two different cases; the semi-infinite case and the finite case.

For the semi-infinite case, the only way to obtain semi-infinite bars is from the non-torsion contribution by combining bars that are all semi-infinite. Thus, to construct a semi-infinite bar in $\text{bcd}(X_1) \otimes_f \cdots \otimes_f \text{bcd}(X_n)$ we take a semi-infinite bar from each $\text{bcd}(X_q)$. This gives

$$\begin{aligned}\binom{x_{0,1}}{1}\binom{x_{0,2}}{1}\cdots\binom{x_{0,n}}{1} &= x_{0,1}x_{0,2}\cdots x_{0,n}\\ &= \sum_{i_1=0}^{0}\cdots\sum_{i_n=0}^{0}(x_{i_0,1})\cdots(x_{i_n,n})\\ &= \mathrm{count}_n(0,0)\\ &= \mathrm{count}_n(0,0)-\mathrm{count}_n(-1,0)+\mathrm{count}_n(0,1)-\mathrm{count}_n(-1,1)\end{aligned}$$

semi-infinite bars.

Now, consider the finite bar case where $1 \leq r \leq m$. For a bar to have length ℓ_r, two conditions must be satisfied. First, at least one component bar must come from the r^{th} row i.e., $x_{i_q,q} = x_{r,q}$ for some q, i.e., $i_q = r$ for some q. Second, all component bar lengths must be greater than or equal to ℓ_r. In other words, for any q we require $i_q \leq r$. Via the persistent Künneth formula, we have two types of finite bars; standard bars and torsion bars. To obtain a standard bar of length ℓ_r, we simply combine any n bars that satisfy the conditions above. Torsion bars can only be created by combining n finite bars with no infinite bars, and hence, we account for these separately.

To count the standard bars, consider all combinations of the form $(x_{i_1,1})(x_{i_2,2})\cdots(x_{i_n,n})$ where $i_q \leq r$ for all q where we choose one bar from each $\mathrm{bcd}(X_q)$ and where at least one bar has length ℓ_r. This is the number of ways to combine n bars of length at least ℓ_r minus the number of ways to combine n bars of length strictly greater than ℓ_r. Hence,

$$\mathrm{count}_n(r,0) - \mathrm{count}_n(r-1,0)$$

is the number of standard bars of length r.

To count the torsion bars, we perform the same counting process as above, except we only allow bars with finite length. This is the number of ways to combine n bars of length at least ℓ_r but less than $\ell_0 = \infty$ minus the number of ways to combine n bars of length strictly greater than ℓ_r but less than $\ell_0 = \infty$. Hence we have

$$\text{count}_n(r,1) - \text{count}_n(r-1,1)$$

torsion bars of length ℓ_r.

Therefore, combining the standard and torsion bars, we have

$$\text{count}_n(r,0) - \text{count}_n(r-1,0) + \text{count}_n(r,1) - \text{count}_n(r-1,1)$$

total bars of length ℓ_r. □

As an example of the theorem above, we can describe the number of bars of each length for all of the molecules above. Specifically, we can look at the sublevelset persistent homology of the energy landscape of the molecule 1-2-2-1/1-3-2-1. There are four different bar lengths: semi-infinite, $\delta_2 - \beta_2$, $\gamma_2 - \alpha_2$, and $\gamma_1 - \beta_1$, where $\delta_2 - \beta_2 > \gamma_2 - \alpha_2 > \gamma_1 - \beta_1$. We use the theorem to determine that there are 4 semi-infinite bars, 2 bars of length $\delta_2 - \beta_2$, 2 bars of length $\gamma_2 - \alpha_2$, and 12 bars of length $\gamma_1 - \beta_1$; see Figure 4.1.

These three theorems (Theorem 4.1.2, Theorem 4.2, and Theorem 4.3.2) provide information about the sublevelset persistence without having to visualize the energy landscape. They also provide results that apply to several different types of energy landscapes and other types of additive functions over product spaces. Next, we will study the sublevelset persistence of branched alkanes that consist only of 3-2 internal bonds.

Chapter 5

An example of sublevelset persistence characterization

Our overarching goal is to characterize any branched alkane by classifying the length and birth time of each bar. As a starting point, we first look at combining n internal bonds of type 3-2. Let $f_{\text{1-3-2-1}}\colon S^1 \to \mathbb{R}$ be the energy function for isopentane, the function we will use to approximate all 3-2 bonds. Denote $f : (S^1)^n \to \mathbb{R}$ as the energy landscape for the molecule with n copies of 3-2 bonds where $f(\phi_1, \ldots, \phi_n) = f_{\text{1-3-2-1}}(\phi_1) + \ldots + f_{\text{1-3-2-1}}(\phi_n)$. Let k denote the index of a critical point in f. Recall, we label the global minima of isopentane as a_1 and a_2, the local minimum with b, the local maximum with c and the global maxima with d_1 and d_2. For reference, see Figure 5.1. Given a critical point $(\phi_1, \ldots, \phi_n)$ of f, let i_1 denote the number of copies of d_1, let i_2 denote the number of copies of d_2, and hence $k - i_1 - i_2$ is the number of copies of c. Similarly, let j_1 denote the number of copies of a_1, let j_2 denote the number of copies of a_2, and so $n - k - j_1 - j_2$ is the number of copies of b. This implies that $k \leq n$, $i_1 + i_2 \leq k$, and $j_1 + j_2 \leq n - k$.

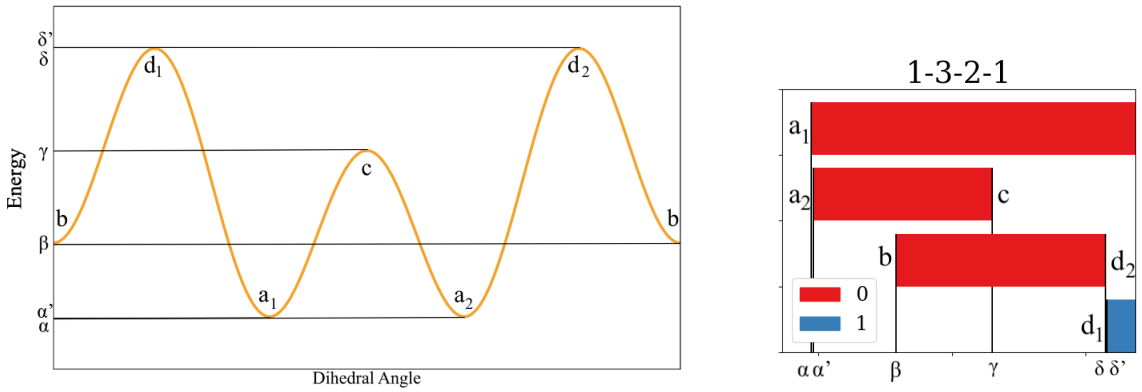

Figure 5.1: The OPLS-UA energy landscape for isopentane and its corresponding sublevelset persistence barcode. Note, this figure is the same as that in Figure 2.4, but we include it here for ease of reference.

To mirror the main theorem from Mirth et al. [2] which characterizes the persistence for any n-alkane (a chain of carbons with type 2-2 internal bonds), we perform the same classification process for the 3-2 internal bonds. Unlike 2-2 internal bonds, 3-2 internal bonds have two different finite bar lengths, $\gamma - \alpha$ and $\delta - \beta$, where $\gamma - \alpha < \delta - \beta$. Further, energy level α corresponds to the birth of a semi-infinite bar and the birth of a finite bar, while energy level δ corresponds to the birth of a 1-dimensional, semi-infinite bar and the death of a 0-dimensional bar. For the 2-2 case, all persistent homology bars born at the same birth value had the same homological dimension and length. Energy level α corresponded to the birth of a semi-infinite bar, energy level β corresponded to the birth of two finite bars of the same length. Similarly, energy level γ corresponded to the death of two finite bars of the same length and energy level δ corresponded to the birth of a semi-infinite bar. These factors make characterizing the sublevelset persistence of any number of 3-2 internal bonds a good starting point, as the 3-2 internal bond is the most combinatorially complex energy landscape of the internal bond types we consider.

An example of the sublevelset persistence barcode and the sublevelset persistence diagram of the molecule with three internal 3-2 bonds can be found in Figure 5.2. The sublevelset persistence diagram is just a different representation of the barcode. The birth time is given by the x-axis and the death time is given by the y-axis. The diagram is a more concise way of sharing the same data and is often used for functions with more homological features. While we can use GUDHI to calculate the sublevelset persistence of branched alkanes for up to nine internal bonds, any larger molecules take a long time (on the order of days). Hence, we look for a different way to calculate the sublevelset persistence of branched alkanes.

To fully characterize the sublevelset persistence of n copies of 3-2 internal bonds, first we will count the number of bars of each length. Then, we will classify all critical points of the energy landscape of the molecule. Next, we partition the classes based on the type of bar born from each class. Finally, for n copies of 3-2 internal bonds, we determine the total number of bars born for any class of critical points i.e., for any given birth energy value and

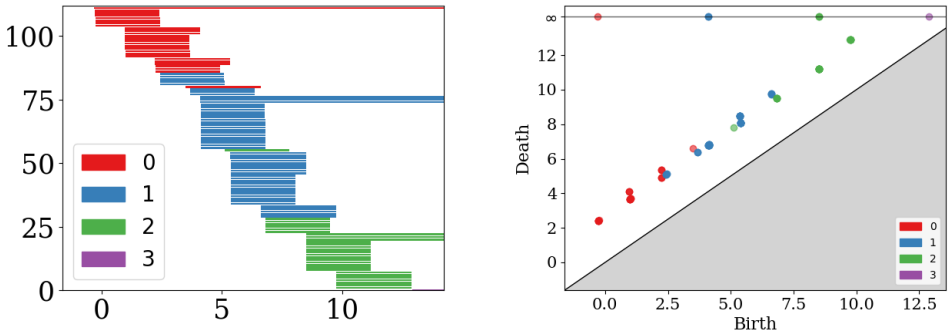

Figure 5.2: The sublevelset persistence barcode and sublevelset persistence diagram for the energy landscape of a molecule with exactly three copies of 3-2 internal bonds.

the length of those bars. Note, the results from Chapter 4 provide information about the number of bars in each dimension and of each length. But that is not enough information to fully describe the sublevelset persistence. The birth time and length of each bar will provide the missing information needed to construct the persistence barcode and allow us to fully characterize the sublevelset persistence.

5.1 Number of bars per length for internal bond 3-2

Based on the results from Chapter 4, we know that each branched alkane has $\frac{6^n+2^n}{2}$ bars and $\binom{n}{k} + (3^n - 1)\binom{n-1}{k}$ bars in dimension k. We can also count the number of bars in all homological dimensions of length $\delta - \beta$, bars of length $\gamma - \alpha$, and semi-infinite bars, the three different length bars for a molecule consisting exclusively of type 3-2 internal bonds. Then, we show the complete characterization of the sublevelset persistence of n copies of 3-2 bonds. Later, we will use our characterization process to generalize the characterization process for any branched alkane.

Theorem 5.1.1. *For n internal bonds of type* 1-3-2-1*, we have $2^{n-1}(2^n - 1)$ bars of length $\delta - \beta$ and $2^{n-1}(3^n - 2^n)$ bars of length $\gamma - \alpha$.*

Proof. Consider n copies of X_1, where X_1 is the persistence diagram of the internal bond 3-2. We calculate the number of bars of length $u := \gamma - \alpha$ and $t := \delta - \beta$. Recall from

Theorem 4.3.1, let i_1 be the number of semi-infinite bars in bcd(X_1), let t_1 be the number of bars of length t in bcd(X_1), and let u_1 be the number of bars of length u in bcd(X_1). Note that for internal bond 3-2, we have $i_1 = 2$, $t_1 = 1$, and $u_1 = 1$. Using the formulas outlined in Theorem 4.3.1, we get

$$i_n = i_1 i_{n-1} = 2i_{n-1} = 2^n$$

infinite bars. As for counting the finite bars, first count the number of bars of length t:

$$\begin{aligned} t_n &= t_1 i_{n-1} + i_1 t_{n-1} + t_1 t_{n-1} + t_1 t_{n-1} \\ &= 1(2^{n-1}) + 4(t_{n-1}). \end{aligned}$$

Since this is a recursion relation, consider the hypothesized closed form, $t_n = 2^{n-1}(2^n - 1)$. We show that this is indeed the correct closed form and satisfies the recursion relation $t_n = 1(2^{n-1})+4(t_{n-1})$. We plug in the closed form to the recursion formula, using $t_n = 2^{n-1}(2^n-1)$ and $t_{n-1} = 2^{(n-1)-1}(2^{n-1} - 1)$, to get Hence,

$$\begin{aligned} 2^{n-1} + 4(t_{n-1}) &= 2^{n-1} + 4(2^{(n-1)-1}(2^{n-1} - 1)) \\ &= 2^{n-1} + 2^2(2^{n-2}(2^{n-1} - 1)) \\ &= 2^{n-1} + 2^n(2^{n-1} - 1) \\ &= 2^{n-1} + 2^{2n-1} - 2^n \\ &= 2^{n-1}(1 + 2^n - 2) \\ &= 2^{n-1}(2^n - 1) \\ &= t_n \end{aligned}$$

as desired, which shows the closed form is indeed the correct solution.

To find u_n, the number of finite bars of length u, subtract the number of bars of length t (t_n) from the total number of finite bars. Hence,

$$
\begin{aligned}
u_n &= \frac{6^n - 2^n}{2} - 2^{n-1}(2^n - 1) \\
&= \frac{3^n 2^n - 2^n - 2^n(2^n - 1)}{2} \\
&= 2^{n-1}3^n - 1 - 1(2^n - 1) \\
&= 2^{n-1}3^n - 1 - 2^n + 1 \\
&= 2^{n-1}(3^n - 2^n)
\end{aligned}
$$

gives the number of bars of length u. Therefore, for n internal bonds of type 1-3-2-1 we have $2^{n-1}(2^n - 1)$ bars of length $t = \delta - \beta$ and $2^{n-1}(3^n - 2^n)$ of length $u = \gamma - \alpha$. □

Knowing how many different features that persist for varying amounts of energy levels gives more insight to the shape of the energy landscape. This alone is not enough to fully understand the structure; we also must identify the birth time and the length of each bar to fully characterize the sublevelset persistent homology. This will provide more information about the shape of the energy landscape, and it will allow us to determine the sublevelset persistence of larger molecules without taking large amounts of computation time.

5.2 Characterizing bar births and lengths for internal bond type 3-2

To completely characterize the sublevelset persistence barcodes for molecules consisting only of internal bonds of type 3-2, we first need to split the critical points of a given energy landscape into classes that correspond to either the birth or death of a homological feature. Classifying critical points for molecules with multiple copies of the 3-2 bond requires some perturbation of the critical points of the 3-2 internal bond energy landscape. As we discussed above, internal 3-2 bonds have different homological feature births and deaths that occur simultaneously. At energy value α we have the birth of a semi-infinite bar and the birth of a finite bar, and at energy value δ we have both a birth of a semi-infinite bar and the

death of a bar born at β (see Figure 5.1). To simplify our critical point classification, we introduce $\varepsilon > 0$ as an arbitrarily small perturbation of the energy values of critical points a_2 and d_2 from the energy landscape of 1-3-2-1, to bump them slightly higher than a_1 and d_1, respectively. This means that the birth of the semi-infinite 0-dimensional bar is slightly before the birth of the finite 0-dimensional bar, now born at $\alpha' := \alpha + \varepsilon$. Additionally, the death of the bar born at energy value β happens slightly before the birth of the 1-dimensional semi-infinite bar, now born at $\delta' := \delta + \varepsilon$.

Note, when constructing our persistence barcode, we want our energy values to be in terms of α and δ as opposed to α' and δ'. To ensure each class corresponds to the proper energy value, we take the limit as $\varepsilon \to 0$ and combine the births and deaths accordingly after we have characterized each class. This gives all energy values in terms of α and δ.

Recall, the internal 3-2 bond has six critical points; see Figure 5.1. Let $\varepsilon > 0$ be an arbitrarily small perturbation to the energy values of a_2 and d_2 such that their corresponding energy values are now $\alpha' := \alpha+\varepsilon$ and $\delta' := \delta+\varepsilon$. Thus, critical point a_1 corresponds to global minima with energy value α, critical point a_2 corresponds to a modified local minimum with energy value α', and critical point b corresponds to the local minimum with energy value β. Similarly, critical point c corresponds to a local maximum with energy value γ, critical point d corresponds to a modified local maximum with energy value δ, and critical point d_2 is the global maximum with energy value δ'. We define the class of critical points of f with these labelings in mind, where we will denote each critical point class using the notation class$(n, k, i_1, i_2, j_1, j_2)$.

Definition 5.2.1. Let $f\colon (S^1)^n \to \mathbb{R}$ be the branched alkane energy function with n internal 3-2 bonds, and let $k \leq n$ be the index of a critical point of f. Let $i_1 + i_2 \leq k$ and let $j_1 + j_2 \leq n - k$. We say that an index k critical point, $(\phi_1, \ldots, \phi_n)$, of f is of class$(n, k, i_1, i_2, j_1, j_2)$ if the list of points, $(\phi_1, \ldots, \phi_n)$, consists of the breakdown of critical points of the 3-2 bond, outlined below.

Type 1-3-2-1		
Critical Point	Feature Type	Number of copies
d_1	Local Max$^\diamond$	i_1
d_2	Global Max	i_2
c	Local Max	$k - i_1 - i_2$
a_1	Global Min	j_1
a_2	Local Min$^\diamond$	j_2
b	Local Min	$n - k - j_1 - j_2$

Note, the $\diamond$ denotes that the critical point from the 3-2 internal bond has been shifted by ε, and hence, has switched from global to local.

The definition above is similar to the class definition found in Mirth et al. [2], as both split up the critical points of f by their homological features. Similarly we will use this definition to identify the energy value associated to each class and count the number of critical points of f in each class. Now consider the following lemmas.

Lemma 5.2.2. *For* $f\colon (S^1)^n \to \mathbb{R}$ *where* $f(\phi_1, \ldots, \phi_n) = \sum_{i=1}^{n} f_{\text{1-3-2-1}}(\phi_i)$*, all critical points of* class$(n, k, i_1, i_2, j_1, j_2)$ *have energy value*

$$E(n, k, i_1, i_2, j_1, j_2) = (j_1)\alpha + (j_2)\alpha' + (n - k - j_1 - j_2)\beta + (k - i_1 - i_2)\gamma + (i_1)\delta + (i_2)\delta'.$$

Proof. The proof follows from the fact that $f(\phi_1, \ldots, \phi_n) = \sum_{k=1}^{n} f_{\text{1-3-2-1}}(\phi_k)$ and $f_{\text{1-3-2-1}}(a_1) = \alpha$, $f_{\text{1-3-2-1}}(a_2) = \alpha'$, $f_{\text{1-3-2-1}}(b) = \beta$, $f_{\text{1-3-2-1}}(c) = \gamma$, $f_{\text{1-3-2-1}}(d_1) = \delta$, and $f_{\text{1-3-2-1}}(d_2) = \delta'$. □

We can also determine the number of critical points of f in each class. Recall that the multinomial coefficient is a tool used to count how x objects can be distributed into n boxes.

Definition 5.2.3. Let $x \in \mathbb{N}^+$ and let $y_1, y_2, \ldots, y_n \in \mathbb{N}^+$ such that $\sum_{\ell=1}^{n} y_\ell = x$. Thus we define the *multinomial coefficient* as

$$\binom{x}{y_1, y_2, \ldots, y_n} := \left(\frac{x!}{y_1! y_2! \ldots y_n!}\right).$$

Using this definition, we can count the number of critical points of f in each class.

Lemma 5.2.4. *The number of critical points of f in each* class$(n, k, i_1, i_2, j_1, j_2)$ *is*

$$\binom{n}{j_1, j_2, n-k-j_1-j_2, i_1, i_2, k-i_1-i_2}.$$

Proof. Among its n entries $(\phi_1, \ldots, \phi_n)$, a critical point of class$(n, k, i_1, i_2, j_1, j_2)$ has i_1 copies of d_1, i_2 copies of d_2, $k-i_1-i_2$ copies of c, j_1 copies of a_1, j_2 copies of a_2, and $n-k-j_1-j_2$ copies of b. Hence the lemma follows from the definition of the multinomial coefficient, given in Definition 5.2.3. □

We are now ready to completely characterize the persistence barcode of a molecule containing n copies of 3-2 bonds. First, we partition our classes by grouping them via homological feature. Then, we use that partition to fully characterize the sublevelset persistence.

Lemma 5.2.5. *For any branched alkane consisting of n internal bonds of type* 3-2*, consider the branched alkane energy landscape, $f\colon (S^1)^n \to \mathbb{R}$. Let $k \leq n$, $i_1 + i_2 \leq k$, and $j_1 + j_2 \leq n - k$. The following restrictions on j_2, j_1, i_1 and i_2 partition the classes into four groups by homological feature type;*

1. *$j_2 = 0$, $j_1 = n - k$, $i_1 = 0$, $i_2 = k$, corresponds to the classes that birth semi-infinite bars,*
2. *$j_2 = 0$, $j_1 > n-k$, $i_1 + i_2 = k$ corresponds to the classes that birth bars of length $\delta - \beta$,*
3. *$j_2 \neq 0$ corresponds to the classes that birth bars of length $\gamma - \alpha$, and*
4. *all other classes correspond to deaths.*

Proof. We show that the restrictions on i_1, i_2, j_1, and j_2 pair each class with the appropriate bar length.

Since the energy landscape, f, is Morse, we know the critical points of f correspond directly to the births and deaths of the persistence bars (Theorem 2.4.1). Hence, we can characterize the bars such that for any n copies of a 3-2 bond, the division of critical point classes of f defined above, (i.e., with the restrictions on j_2, j_1, i_1, and i_2) is indeed the proper division of classes. The persistent Künneth formula (Theorem 2.5.3) dictates the birth and death times of each new bar in the product space based off of the birth and death times of the bars coming from each component.

First, consider semi-infinite length bars (Case 1). Note, semi-infinite bars are only born through standard combinations as opposed to torsion combinations. Hence, we only examine the non-torsion portion of the Künneth formula. If $i_1, j_2 = 0$, then this class contains no critical points $(\phi_1, \ldots, \phi_n)$ of f that have any coefficient $\phi_1, \ldots, \phi_n$ equal to d_1 or to a_2. Critical points d_1 and a_2 of internal bond 3-2 correspond to finite bars, and hence would make the bar finite. Similarly, the additional conditions $j_1 = n - k$ and $i_2 = k$ imply that $n - k - j_1 - j_2 = 0$ and $k - i_1 - i_2 = 0$ and hence there are also no critical points of type b or type c. The number of 1-dimensional semi-infinite bars is given by the i_2 critical points of type d_2 and the number of semi-infinite 0-dimensional bars is given by j_1 critical points of type a_1. Since the only way to get semi-infinite bars is by combining semi-infinite bars, these classes of critical points of f such that $i_1 = 0$, $i_2 = k$, $j_1 = n - k$, $j_2 = 0$, give the semi-infinite bars of dimension k.

Next, consider bars of length $\delta - \beta$ (Case 2). Since $\delta - \beta > \gamma - \alpha$, the only way to obtain bars of length $\delta - \beta$ via standard combination is by combining one bar of length $\delta - \beta$ with either a semi-infinite bar or another bar of length $\delta - \beta$, as outlined in Theorem 4.3.1. Additionally, the only way to obtain bars via torsion of length $\delta - \beta$ is by combining a bar of length $\delta - \beta$ with another bar of length $\delta - \beta$. Hence, the classes that give bars of length $\delta - \beta$ must not contain critical points of internal bond 3-2 that correspond to the birth or death of bars of length $\gamma - \alpha$. The conditions $j_2 = 0$ and $i_1 + i_2 = k$ imply that there are no copies of a_2 or c, the critical points of internal bond 3-2 corresponding to the birth and

death of bars of length $\gamma - \alpha$. Finally, we rule out classes that correspond to the birth of semi-infinite bars. To ensure we are not including classes that give semi-infinite bars, we rule out the cases where $n - k = j_1$. Without that restriction, we can overlap with the semi-infinite class with conditions $i_1 = 0$, $i_2 = k$, $j_1 = n - k$, $j_2 = 0$. Hence, the restrictions $i_1 + i_2 = k$, $j_2 = 0$, $n - k < j_1$ give all critical points of f needed to construct bars of length $\delta - \beta$.

Now consider bars of length $\gamma - \alpha$ (Case 3). To ensure the bar has the proper length, we must guarantee that there is at least one copy of critical point a_2 from the energy landscape of internal bond type 3-2. Hence, we require $j_2 > 0$. This restriction implies we are combining the shortest bar with any number of other bars, and by the persistent Künneth formula (Theorem 2.5.3) the resulting bar must take on the shortest length. Thus $j_2 > 0$ is the only requirement we need to get all classes of critical points of f that correspond to bars with length $\gamma - \alpha$.

Finally, note that all other classes correspond to deaths since we have already accounted for all births (Case 4). Thus, we have accounted for all critical points. □

Now that we have identified the energy value that corresponds to each class and partitioned the critical point classes of f appropriately, we can fully characterize the sublevelset persistence for molecules consisting exclusively of internal bonds of type 3-2.

Theorem 5.2.6. *For any branched alkane consisting of n internal bonds of type* 3-2*, consider the k-dimensional sublevelset persistent homology barcodes of the branched alkane energy landscape, $f\colon (S^1)^n \to \mathbb{R}$. Let $k \leq n$, $i_1 + i_2 \leq k$, and $j_1 + j_2 \leq n - k$. The birth time of all k-dimensional bars created by critical points of f in* $\text{class}(n, k, i_1, i_2, j_1, j_2)$ *is*

$$E(n, k, i_1, i_2, j_1, j_2) = (j_1)\alpha + (j_2)\alpha' + (n - k - j_1 - j_2)\beta + (k - i_1 - i_2)\gamma + (i_1)\delta + (i_2)\delta',$$

where the number of bars in that class is given below by:

1. *$j_2 = 0$, $j_1 = n - k$, $i_1 = 0$, $i_2 = k$, gives*

$$\binom{n}{j_1, j_2, n-k-j_1-j_2, i_1, i_2, k-i_1-i_2}$$

semi-infinite bars,

2. $j_2 = 0$, $j_1 > n-k$, $i_1 + i_2 = k$ *gives*

$$\sum_{\ell=0}^{i_1} (-1)^\ell \binom{n}{j_1, j_2, n-k-j_1-j_2+\ell, i_1-\ell, i_2, k-i_1-i_2}$$

bars of length $\delta - \beta$,

3. $j_2 > 0$ *gives*

$$\sum_{\ell=0}^{k-i_1-i_2} (-1)^\ell \binom{n}{j_1, j_2+\ell, n-k-j_1-j_2, i_1, i_2, k-i_1-i_2-\ell}$$

bars of length $\gamma - \alpha$, *and*

4. *0 bars born for any other type of critical point of* f.

Proof. We proceed by induction on k to determine the number of k-dimensional bars born at the corresponding energy value. First, group all $\text{class}(n, k, i_1, i_2, j_1, j_2)$ via the partition given by Lemma 5.2.5. Now we count the number of bars born in each group.

Base Case: $k = 0$

Let $k = 0$; this implies $i_1, i_2 = 0$. Hence, we consider classes of type $\text{class}(n, 0, 0, 0, j_1, j_2)$, where $j_1 + j_2 \leq n$. Note, since all critical points of f in each of these classes have index 0, each 0-dimensional critical point gives birth to a 0-dimensional bar. Thus, the total number of 0-dimensional bars born from each class is the number of critical points of f in each class, equal to

$$|\text{class}(n, 0, 0, 0, j_1, j_2)| = \binom{n}{j_1, j_2, n-j_1-j_2}.$$

There are three different options for j_1 and j_2. If $j_2 = 0$ and $j_1 = n$, then there is

$$\binom{n}{n,0,0} = 1$$

0-dimensional bar born in class$(n,0,0,0,n,0)$. This verifies we get a single, semi-infinite bar of dimension 0, as we should since the torus has one 0-dimensional semi-infinite bar.

Next, if $j_2 = 0$ and $j_1 < n$, then the persistence barcode has

$$\binom{n}{j_1,0,n-j_1} = \binom{n}{j_1}$$

bars of length $\delta - \beta$ for each class(n,k,i_1,i_2,j_1,j_2).

Finally, if $j_2 > 0$, then the persistence barcode has

$$\binom{n}{j_1,j_2,n-j_1-j_2}$$

bars of length $\gamma - \alpha$ for each class(n,k,i_1,i_2,j_1,j_2).

Summing these three types together gives

$$\binom{n}{n,0,0}+\sum_{j_1=0}^{n-1}\binom{n}{j_1,0,n-j_1}+\sum_{j_1=0}^{n}\sum_{j_2=1}^{n-j_1}\binom{n}{j_1,j_2,n-j_1-j_2} = \sum_{j_1=0}^{n}\sum_{j_2=0}^{n-j_1}\binom{n}{j_1,j_2,n-j_1-j_2},$$

0-dimensional bars which uses all 0-dimensional critical points of f.

Inductive step: Assume true for $k-1$

For the inductive step, suppose our formulas hold for $k-1$. We have

$$|\text{class}(n,k,i_1,i_2,j_1,j_2)| = \binom{n}{j_1,j_2,n-k-j_1-j_2,i_1,i_2,k-i_1-i_2}$$

critical points of f in each class(n,k,i_1,i_2,j_1,j_2). Of these points, some of them must kill bars of length $\delta - \beta$ or length $\gamma - \alpha$. For the bars of length $\delta - \beta$, this corresponds to killing off

$$\sum_{\ell=0}^{i_1-1}(-1)^{\ell}\binom{n}{j_1,j_2,n-k-j_1-j_2+\ell,i_1-1-\ell,i_2,k-i_1-i_2}$$

$(k-1)$-dimensional bars. Note, this is due to the fact that if a critical point of f kills a bar of length $\delta - \beta$, that corresponds to killing off a bar from class$(n, k-1, i_1 - 1, i_2, j_1, j_2)$. Similarly for bars of length $\gamma - \alpha$, this corresponds to killing off

$$\sum_{\ell=0}^{k-i_1-i_2-1} (-1)^\ell \binom{n}{j_1, j_2+\ell, n-k-j_1-j_2, i_1, i_2, k-i_1-i_2-1-\ell}$$

bars. This corresponds to critical points of some class$(n, k, i_1, i_2, j_1, j_2)$ killing off critical points of class$(n, k-1, i_1, i_2, j_1, j_2+1)$. All remaining critical points of class$(n, k, i_1, i_2, j_1, j_2)$ must give birth to new persistent homology bars. Hence there are

$$\begin{aligned} |\text{class}(n, k, i_1, i_2, j_1, j_2)| - \sum_{\ell=0}^{i_1-1} (-1)^\ell \binom{n}{j_1, j_2, n-k-j_1-j_2+\ell, i_1-1-\ell, i_2, k-i_1-i_2} \\ = \sum_{\ell=0}^{i_1} (-1)^\ell \binom{n}{j_1, j_2, n-k-j_1-j_2+\ell, i_1-\ell, i_2, k-i_1-i_2} \end{aligned}$$

bars of length $\delta - \beta$ born from each class$(n, k, i_1, i_2, j_1, j_2)$ such that $j_2 = 0$, $j_1 > n-k$, $i_1 + i_2 = k$ and

$$\begin{aligned} |\text{class}(n, k, i_1, i_2, j_1, j_2)| - \sum_{\ell=0}^{k-i_1-i_2-1} (-1)^\ell \binom{n}{j_1, j_2+\ell, n-k-j_1-j_2, i_1, i_2, k-i_1-i_2-1-\ell} \\ = \sum_{\ell=0}^{k-i_1-i_2} (-1)^\ell \binom{n}{j_1, j_2+\ell, n-k-j_1-j_2, i_1, i_2, k-i_1-i_2-\ell} \end{aligned}$$

bars of length $\gamma - \alpha$ born from each class$(n, k, i_1, i_2, j_1, j_2)$ such that $j_2 > 0$. □

This characterization describes the sublevelset persistence for any molecule consisting of n internal bonds of type 3-2. This is still a small subset of the branched alkanes; we want to expand our characterization to include more molecules. Next, we will look at characterizing all molecules consisting of 3-2 and 2-2 internal bonds and use that as a way to explore the process for (more generally) analyzing the sublevelset persistent homology of any additive function over a product space.

Chapter 6

Generalizing the characterization of the sublevelset persistence of branched alkanes

Now that we have characterized the sublevelset persistence for one type of branched alkane, we outline the process for characterizing the sublevelset persistence for any branched alkane. As an introductory example, we will work through the characterization of the branched alkanes consisting exclusively of 2-2 and 3-2 internal bonds. This will allow us to describe how to extend the sublevelset persistence characterization to any number of bond types with an additive function structure.

6.1 Characterizing bar births and lengths for internal bond types 3-2 and 2-2

We now turn our attention to adapting our results from Section 5.2 to molecules with internal bonds exclusively of types 3-2 and 2-2. As before, first we partition our critical points into classes. We have two building block internal bonds, 2-2 and 3-2, where each base energy landscape has six critical points. This implies we have twelve numbers that determine each class of critical points of f. Hence, we expand our definition of class given in Chapter 5. For an example of a molecule with type 3-2 and 2-2 internal bonds, see Figure 6.1.

Definition 6.1.1. Let $f\colon (S^1)^n \to \mathbb{R}$ be the branched alkane energy function such that $f(\phi_1, \ldots, \phi_n) = \sum_{i=1}^{n_1} f_{2\text{-}2}(\phi_i) + \sum_{i=n_1+1}^{n} f_{3\text{-}2}(\phi_i)$ with $n = n_1 + n_2$ internal bonds; n_1 internal 2-2 bonds and n_2 internal 3-2 bonds. Recall from Lemma 2.5.1, $(\phi_1, \ldots, \phi_n)$ is a critical point of f if and only if ϕ_i is a critical point of f_i for every i. Thus, denote k_1 as the number of critical points among $\phi_1, \ldots, \phi_{n_1}$ with index 1, implying that $n_1 - k_1$ is the number of critical points among $\phi_1, \ldots, \phi_{n_1}$ that have index 0. Similarly, denote k_2 as the number of

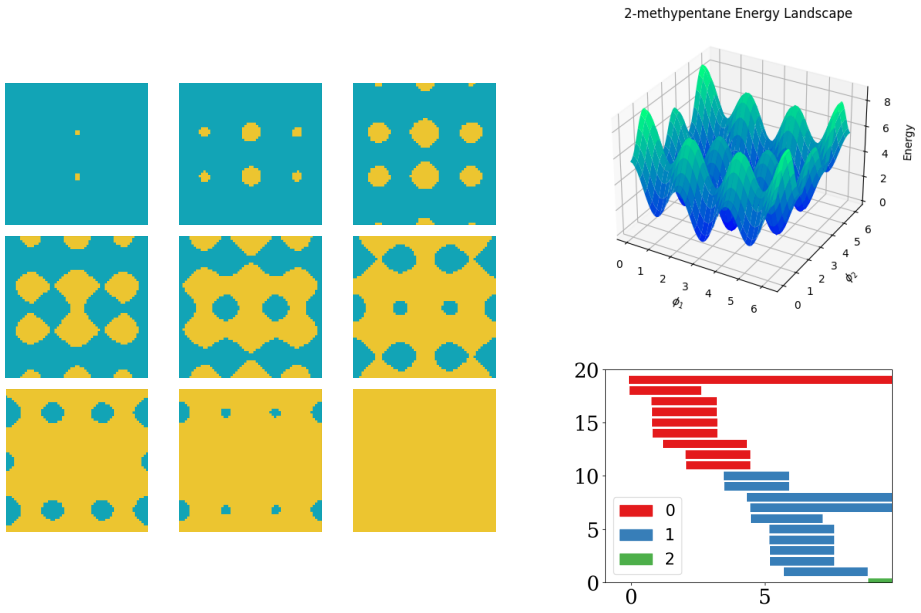

Figure 6.1: (Left) Sublevelsets of the 2-methylpentane energy landscape, the molecule with internal bonds of types 3-2 and 2-2. (Right top) The 2-methylpentane energy landscape and (Right bottom) its persistence barcodes. Note, this figure can also be found in Section 3.2.

critical points among $\phi_{n_1+1}, \ldots, \phi_n$ with index 1, and hence $n_2 - k_2$ is the number of critical points among $\phi_{n_1+1}, \ldots, \phi_n$ with index 0. Thus $k = k_1 + k_2 \leq n = n_1 + n_2$ is the index of a critical point of f. Let $i_{11} + i_{21} + i_{22} \leq k_1 + k_2$ and let $j_{11} + j_{21} + j_{22} \leq (n_1 + n_2) - (k_1 + k_2)$. We say that an index k critical point $(\phi_{11}, \ldots, \phi_{1n_1}, \phi_{21}, \ldots, \phi_{2n_2})$ of f is of class

$$class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)$$

if the ordered list of points, $(\phi_{11}, \ldots, \phi_{2n_2})$, consists of the following number of critical points.

Type 1-2-2-1			Type 1-3-2-1		
Crit. Point	Feature	# of Copies	Crit. Point	Feature	# of Copies
d_{11}	Global Max	i_{11}	d_{21}	Local Max$^\diamond$	i_{21}
			d_{22}	Global Max	i_{22}
c_{11} & c_{12}	Local Max	$k - i_{11}$	c_{21}	Local Max	$k_2 - i_{21} - i_{22}$
a_{11}	Global Min	j_{11}	a_{21}	Global Min	j_{21}
			a_2	Local Min$^\diamond$	j_{22}
b_{11} & b_{12}	Local Min	$n_1 - k_1 - j_{11}$	b	Local Min	$n_2 - k_2 - j_{21} - j_{22}$

Note, the $\diamond$ on Local Max and Local Min denotes that the critical point has been shifted by ε, and hence, has switched from global to local.

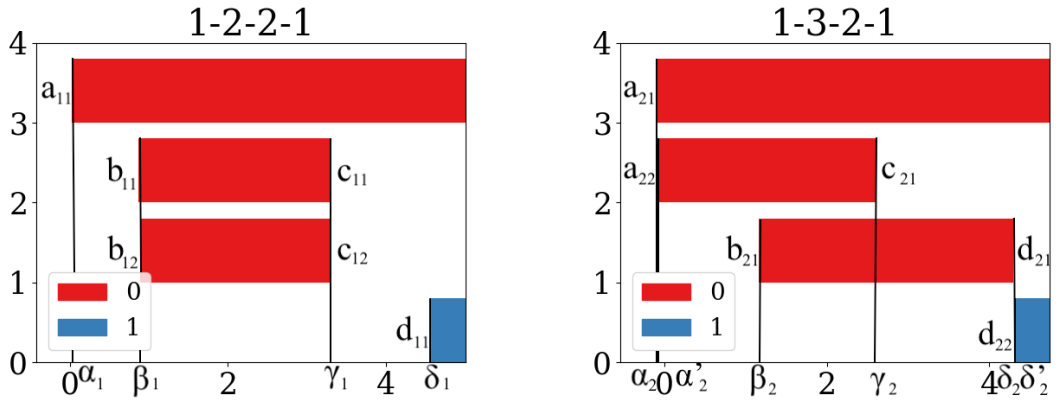

Figure 6.2: The labeled sublevelset persistence barcodes for 1-2-2-1 and 1-3-2-1. Note, b_{11} and b_{12} correspond to the birth of persistence bars of the same length whereas a_{21} and a_{22} correspond to the birth of different length bars. Similarly, the critical points c_{11} and c_{12} correspond to the death of persistence bars of the same length whereas d_{21} and d_{22} correspond to a death and a birth respectively.

For this molecule, we have no need for i_{11} versus i_{12} or for j_{11} versus j_{12}, since there is no need to separate b_{11} from b_{12} or c_{11} from c_{12}. Points b_{11} and b_{12} correspond to the births of $\gamma_1 - \beta_1$ bars and c_{11} and c_{12} correspond to the deaths of $\gamma_1 - \beta_1$, see Figure 6.2. Alternatively, points a_{21} and a_{22} correspond to the birth of a semi-infinite bar and a finite bar, which are

two different homological things. Thus, we simplify our classification process by combining i_{11} and i_{12} as i_{11} and combining j_{11} and j_{12} as j_{11}.

Just like in Chapter 5, we can determine the associated energy value of each class.

Lemma 6.1.2. *For* $f\colon (S^1)^n \to \mathbb{R}$ *where* $f(\phi_1, \ldots, \phi_n) = \sum_{i=1}^{n} f_{\text{1-2-2-1}}(\phi_i) + \sum_{i=n_1+1}^{n_1+n_2} f_{\text{1-3-2-1}}(\phi_i)$, *all critical points of class* $\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix} \right)$ *have energy value*

$$E\left(\text{class}\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)\right) =$$

$$(j_{11})\alpha_1 + (n_{11} - k_{11} - j_{11})\beta_1 + (k_1 - i_{11})\gamma_1 + (i_{11})\delta_1$$
$$+ (j_{21})\alpha_2 + (j_{22})\alpha_2' + (n_2 - k_2 - j_{21} - j_{22})\beta_2 + (k_2 - i_{21} - i_{22})\gamma_2 + (i_{21})\delta_2 + (i_{22})\delta_2'.$$

Proof. The proof follows from the fact that $f(\phi_1, \ldots, \phi_n) = \sum_{i=1}^{n} f_{\text{1-2-2-1}}(\phi_i) + \sum_{i=n_1+1}^{n_1+n_2} f_{\text{1-3-2-1}}(\phi_i)$ and $f_{\text{1-2-2-1}}(a_{11}) = \alpha_1$, $f_{\text{1-2-2-1}}(b_{12}) = \beta_1$, $f_{\text{1-2-2-1}}(b_{12}) = \beta_1$, $f_{\text{1-2-2-1}}(c_{11}) = \gamma_1$, $f_{\text{1-2-2-1}}(c_{12}) = \gamma_1$, and $f_{\text{1-2-2-1}}(d_{11}) = \delta_1$ along with $f_{\text{1-3-2-1}}(a_{21}) = \alpha_2$, $f_{\text{1-3-2-1}}(a_{22}) = \alpha_2'$, $f_{\text{1-3-2-1}}(b_{21}) = \beta_2$, $f_{\text{1-3-2-1}}(c_{21}) = \gamma_2$, $f_{\text{1-3-2-1}}(d_{21}) = \delta_2$, and $f_{\text{1-3-2-1}}(d_{22}) = \delta_2'$. □

Now that we can partition our critical points of f into classes, we can assess how many points are in each class.

Theorem 6.1.3. *The size of a critical point class for any branched alkane energy landscape consisting of internal bonds of types* 1-2-2-1 *and* 1-3-2-1 *is*

$$\left|\text{class}\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)\right| =$$

$$2^{n_1 - i_{11} - j_{11}} \binom{n_1}{i_{11}, k_1 - i_{11}, j_{11}, n_1 - k_1 - j_{11}} \binom{n_2}{i_{21}, i_{22}, k_2 - i_{21} - i_{22}, j_{21}, j_{22}, n_2 - k_2 - j_{21} - j_{22}}$$

critical points of f *in each class.*

Proof. Among its n entries $(\phi_{11}, \ldots, \phi_{1n_1}, \phi_{21}, \ldots, \phi_{2n_2})$, a critical point of $class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)$ has i_{21} copies of d_{21}, i_{22} copies of d_{22}, $k_2 - i_{21} - i_{22}$ copies of c_{21}, j_{21} copies of a_{21}, j_{22} copies of a_{22}, and $n_2 - k_2 - j_{21} - j_{22}$ copies of b_{21}. Additionally, it has i_{11} copies of d_{11}, $k_1 - i_{11}$ copies of c_{11} or c_{12}, j_{11} copies of a_{11}, and $n_1 - k_1 - j_{11}$ copies of b_{11} or b_{12}. Hence the lemma follows from the definition of the multinomial coefficient, see Definition 5.2.3. The extra factor $2^{n_1 - i_{11} - j_{11}}$ comes from the $2^{k_1 - i_{11}}$ choices for the $k - i_{11}$ copies of c_{11} or c_{12} or the $2^{n_1 - k_1 - j_{11}}$ choices for the $n_1 - k_1 - j_{11}$ copies of b_{11} or b_{12}. □

With this partition of critical points of f, we can determine the birth time, number, and length of bars born in each class. This will completely characterize the sublevelset persistence of the OPLS-UA energy landscape of any molecule consisting of 2-2 and 3-2 internal bonds. As before, we prove a lemma that partitions our classes by homological feature type. Then, we provide the full characterization.

Lemma 6.1.4. *For any branched alkane consisting of n_1 internal bonds of type 2-2 and n_2 internal bonds of type 3-2, consider the branched alkane energy landscape, $f\colon (S^1)^n \to \mathbb{R}$. Let $k = k_1 + k_2$, $k_1 + k_2 \leq n_1 + n_2$, $i_{11} + i_{21} + i_{22} \leq k_1 + k_2$, $i_{11} \leq k_1$, $i_{21} + i_{22} \leq k_2$, $j_{11} + j_{21} + j_{22} \leq n_1 + n_2 - k_1 - k_2$, $j_{21} + j_{22} \leq n_2 - k_2$, and $j_{11} \leq n_1 - k_1$. The following restrictions on j_{11}, i_{11} j_{22}, j_{21}, i_{21} and i_{22} partition the classes into five groups by homological feature type;*

1. *$i_{11} + i_{22} = k_1 + k_2$, $i_{21} = 0$, $j_{22} = 0$, $(n_1 + n_2) - (k_1 + k_2) = j_{11} + j_{21}$ corresponds to the classes that give birth to semi-infinite bars,*

2. *$i_{11} + i_{22} = k_1 + k_2$, $j_{22} = 0$, $n_1 - k_1 - j_{11} = 0$, and $(n_1 + n_2) - (k_1 + k_2) > j_{11} + j_{21}$ corresponds to the classes that give birth to bars of length $\delta_2 - \beta_2$,*

3. *$j_{22} > 0$, $i_{11} + j_{11} > 0$, and $n_1 - k_1 - j_{11} = 0$ corresponds to classes that give birth to bars of length $\gamma_2 - \alpha_2$, and*

4. *$n_1 - k_1 - j_{11} > 0$ corresponds to the classes that give birth to bars of length $\gamma_1 - \beta_1$, and*

5. all other classes correspond to deaths.

Proof. We show that the restrictions on i_{11}, j_{11}, i_{21} i_{22}, j_{21}, and j_{22} pair each class with the appropriate bar length. Since f is Morse function, we know the critical points of f correspond directly to the births and deaths of the persistence bars (Theorem 2.4.1). Hence, we can characterize the bars such that for any n_1 copies of a 2-2 bond and n_2 copies of a 3-2 bond, the division of critical point classes of f defined above, (i.e., with the restrictions on i_{11}, j_{11}, i_{21} i_{22}, j_{21}, and j_{22}) is indeed the proper division of classes. The persistent Künneth formula (Theorem 2.5.3) dictates the birth and death times for each new bar in the product space based off of the birth and death times of the bars coming from each component.

First, consider the semi-infinite length bars (Case 1). Recall from the persistent Künneth formula Theorem 2.5.3, the only way to obtain a semi-infinite bar is via standard combinations, hence there is no need to account for torsion. Consider the conditions $i_{21}, j_{22} = 0$, $i_{11} + i_{22} = k_1 + k_2$, and $(n_1 + n_2) - (k_1 + k_2) = j_{11} + j_{21}$. If $i_{21}, j_{22} = 0$, then this class contains no critical points $(\phi_1, \ldots, \phi_n)$ of f that have any coefficient $\phi_1, \ldots, \phi_n$ equal to d_{21} or a_{22}. These critical points are associated with the finite bars originating from the 3-2 internal bond, and if included, would make the bar finite. The births of the 1-dimensional semi-infinite bars are given by the i_{11} and i_{22} critical points of type d_{11} and d_{22} and the births of the semi-infinite 0-dimensional bars are given by j_{11} and j_{21} critical points of type a_{11} and a_{21}. Hence, if $i_{11} + i_{22} = k_1 + k_2$, and $(n_1 + n_2) - (k_1 + k_2) = j_{11} + j_{21}$, then $n_1 - k_1 - j_{11} = 0$ and $n_2 - k_2 - j_{21} - j_{22} = 0$. Thus, there are no points of type b_{11}, c_{11}, b_{21}, or c_{21}, the other critical points from the internal bonds that correspond to births and deaths of finite bars. Hence, these restrictions give us the semi-infinite bars of dimension k.

Next, consider bars of length $\delta_2 - \beta_2$ (Case 2). Since $\delta_2 - \beta_2 > \gamma_2 - \alpha_2 > \gamma_1 - \beta_1$, the only way to get bars of length $\delta_2 - \beta_2$ through standard combination is by combining bars of length $\delta_2 - \beta_2$ with either semi-infinite bars or other bars of length $\delta_2 - \beta_2$, as outlined in Theorem 4.3.1. Additionally, the only way to obtain bars of length $\delta_2 - \beta_2$ via torsion is by combining bars of length $\delta_2 - \beta_2$ with other bars of length $\delta_2 - \beta_2$. Hence, our classes that

give bars of length $\delta_2 - \beta_2$ must not contain critical points of internal bonds that correspond to the birth or death of bars of lengths $\gamma_2 - \alpha_2$ or $\gamma_1 - \beta_1$. Hence, $j_{22} = 0$ and $n_1 - k_1 - j_{11} = 0$ implies that there are no copies of a_{22} or b_{11}, the critical points corresponding to the birth of bars of length $\gamma_2 - \alpha_2$ and bars of length $\gamma_1 - \beta_1$. The condition $i_{11} + i_{22} = k_1 + k_2$ implies that we have $k_1 - i_{11} = 0$ copies of c_{11} and $k_2 - i_{22} - i_{21} = 0$ copies of c_{21} (since all values are greater than or equal to 0), which means that there are no critical points that correspond to the death of bars of length $\gamma_2 - \alpha_2$ or bars of length $\gamma_1 - \beta_1$. Finally, we have to rule out classes that correspond to the birth of semi-infinite bars. To ensure we are not including classes that give semi-infinite bars, we rule out classes where $(n_1 + n_2) - (k_1 + k_2) = j_{11} + j_{21}$. Without that restriction, we can overlap with the semi-infinite class that has conditions $i_{21}, j_{22} = 0$, $i_{11} + i_{22} = k_1 + k_2$, $i_{11} + i_{22} = k_1 + k_2$, and $(n_1 + n_2) - (k_1 + k_2) = j_{11} + j_{21}$. Hence, the restrictions $i_{11} + i_{22} = k_1 + k_2$, $j_{22} = 0$, $n_1 - k_1 - j_{11} = 0$, and $(n_1 + n_2) - (k_1 + k_2) > j_{11} + j_{21}$ give all critical points of f needed to construct all bars of length $\delta - \beta$.

Now consider bars of length $\gamma_2 - \alpha_2$ (Case 3). To ensure the bar has the proper length, we must guarantee that there is at least one copy of critical point a_{22} from internal bond 2-2. Hence, we require $j_{22} > 0$. Additionally, we require that there are no critical points of type b_{11}, the critical point that corresponds to the birth of the $\gamma_1 - \beta_1$ bar (since $\gamma_2 - \alpha_2 > \gamma_1 - \beta_1$). Hence, we require $n_1 - k_1 - j_{11} = 0$ which gives the classes that correspond to the birth of bars of length $\gamma_2 - \alpha_2$.

For Case 4, consider classes that correspond to the birth of bars with length $\gamma_1 - \beta_1$. Consider the restriction $n_1 - k_1 - j_{11} > 0$. This ensures that there is one critical point of type b_{11}, the critical point that gives the birth of a $\gamma_1 - \beta_1$ length bar. This restriction implies we are combining the shortest bar with any number of other bars, and by the persistent Künneth formula (Theorem 2.5.3) the resulting bar must take on the shortest length, in this case, $\gamma_1 - \beta_1$. Thus this is the only requirement we need to get all classes of critical points of f that correspond to the birth of bars with length $\gamma_1 - \beta_1$.

Finally, note that all other classes correspond to deaths since we have already accounted for all births (Case 5). □

Now, we can fully characterize the sublevelset persistence of any branched alkane consisting exclusively of internal bonds of types 3-2 and 2-2.

Theorem 6.1.5. *For any branched alkane consisting of n_1 internal bonds of type* 2-2 *and n_2 internal bonds of type* 3-2*, consider the k-dimensional sublevelset persistent homology barcodes of the branched alkane energy landscape, $f\colon (S^1)^n \to \mathbb{R}$. Let $k = k_1 + k_2$, $k_1 + k_2 \leq n_1 + n_2$, $i_{11} + i_{21} + i_{22} \leq k_1 + k_2$, $i_{11} \leq k_1$, $i_{21} + i_{22} \leq k_2$, $j_{11} + j_{21} + j_{22} \leq n_1 + n_2 - k_1 - k_2$, $j_{21} + j_{22} \leq n_2 - k_2$, and $j_{11} \leq n_1 - k_1$. Hence, for any class* $\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)$, *the birth time of any k-dimensional bars in that class is*

$$
\begin{aligned}
&E(n_1,n_2,k_1,k_2,i_{11},j_{11},i_{21},i_{22},j_{21},j_{22}) = \\
&\quad (j_{11})\alpha_1 + (n_{11} - k_{11} - j_{11})\beta_1 + (k_1 - i_{11})\gamma_1 + (i_{11})\delta_1 \\
&\quad + (j_{21})\alpha_2 + (j_{22})\alpha_2' + (n_2 - k_2 - j_{21} - j_{22})\beta_2 + (k_2 - i_{21} - i_{22})\gamma_2 + (i_{21})\delta_2 + (i_{22})\delta_2',
\end{aligned}
$$

where the number and length of bars born in that class is given below by:

1. $i_{11} + i_{22} = k_1 + k_2$, $i_{21} = 0$, $j_{22} = 0$, $(n_1 + n_2) - (k_1 + k_2) = j_{11} + j_{21}$ *gives*

$$\left| class \left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right) \right|$$

semi-infinite bars,

2. $i_{11} + i_{22} = k_1 + k_2$, $j_{22} = 0$, $n_1 - k_1 - j_{11} = 0$, *and* $(n_1 + n_2) - (k_1 + k_2) > j_{11} + j_{21}$ *gives*

$$\sum_{\ell'=0}^{i_{21}}(-1)^{\ell}\left[\binom{n_1}{i_{11},k_1-i_{11},j_{11},n_1-k_1-j_{11}}\right.$$
$$\left.\binom{n_2}{i_{21}-\ell,i_{22},k_2-i_{21}-i_{22},j_{21},j_{22},n_2-k_2-j_{21}-j_{22}+\ell}\right]$$

bars of length $\delta_2-\beta_2$,

3. $j_{22}>0$, $i_{11}+j_{11}>0$, *and* $n_1-k_1-j_{11}=0$ *gives*

$$\sum_{\ell=0}^{k_2-i_{21}-i_{22}}(-1)^{\ell}\left[\binom{n_1}{i_{11},k_1-i_{11},j_{11},n_1-k_1-j_{11}}\right.$$
$$\left.\binom{n_2}{i_{21},i_{22},k_2-i_{21}-i_{22}-\ell,j_{21},j_{22}+\ell,n_2-k_2-j_{21}-j_{22}}\right]$$

bars of length $\gamma_2-\alpha_2$, *and*

4. $n_1-k_1-j_{11}>0$ *gives*

$$2^{n_1-i_{11}-j_{11}}\sum_{\ell=0}^{k_1-i_{11}}(-1)^{\ell}\left[\binom{n_1}{i_{11},k_1-i_{11}-\ell,j_{11},n_1-k_1-j_{11}+\ell}\right.$$
$$\left.\binom{n_2}{i_{21},i_{22},k_2-i_{21}-i_{22},j_{21},j_{22},n_2-k_2-j_{21}-j_{22}}\right]$$

bars of length $\gamma_1-\beta_1$, *and*

5. *0 bars born for any other classes of critical points of* f.

Proof. In Lemma 6.1.4, we have showed that the restrictions on j_{11}, i_{11}, j_{21}, j_{22}, i_{21}, i_{22} form a proper partition of classes that are grouped by their associated bar lengths. With this partition, we proceed by induction on k to determine the number of k-dimensional bars born at each energy value.

Base Case: $k=0$

For the base case $k=0$, this implies that $k_1,k_2,i_{11},i_{21},i_{22}=0$. Hence, the classes we consider are of type

$$\text{class}\left(\begin{bmatrix} n_1 & 0 & 0 & 0 & j_{11} & 0 \\ n_2 & 0 & 0 & 0 & j_{21} & j_{22} \end{bmatrix}\right),$$

where $j_{11} + j_{21} + j_{22} \leq n_1 + n_2$. Note, since all critical points of f in each of these classes have index 0, each 0-dimensional critical point of f gives birth to a 0-dimensional bar. Thus, the total number of 0-dimensional bars born from each class is the number of critical points of f in each class, equal to

$$\frac{n_1!n_2!}{(j_{11})!(n_1 - j_{11})!(j_{21})!(j_{22})!(n_2 - j_{21} - j_{22})!} = \binom{n_1}{j_{11}}\binom{n_2}{j_{21}, j_{22}, n_2 - j_{21} - j_{22}}.$$

We have four different options for j_{11}, j_{21} and j_{22}. If $n_1 + n_2 = j_{11} + j_{21}$ and $j_{22} = 0$, this implies $j_{11} = n_1$ and $j_{21} = n_2$. Then,

$$\binom{n_1}{n_1}\binom{n_2}{n_2, 0, 0} = 1.$$

This verifies we get a single, semi-infinite bar of dimension 0.

Next, if $j_{11} = n_1$, $j_{21} < n_2$ and $j_{22} = 0$, then we get

$$\binom{n_1}{n_1}\binom{n_2}{j_{21}, 0, n - j_{21}} = \binom{n_2}{j_{21}}$$

bars of length $\delta_2 - \beta_2$.

Then, if $j_{11} = n_1$ and $j_{22} > 0$, we get

$$\binom{n_1}{n_1}\binom{n_2}{j_{21}, j_{22}, n - j_{21} - j_{22}} = \binom{n_2}{j_{21}, j_{22}, n - j_{21} - j_{22}}$$

bars of length $\gamma_2 - \alpha_2$.

Finally, if $j_{11} < n_1$, then we get

$$\binom{n_1}{j_{11}}\binom{n_2}{j_{21}, j_{22}, n_2 - j_{21} - j_{22}}$$

bars of length $\gamma_1 - \beta_1$.

Summing these four types together gives

$$\begin{aligned}
&\binom{n_1}{n_1}\binom{n_2}{n_2,0,0} + \sum_{j_{21}=0}^{n_2-1}\binom{n_1}{n_1}\binom{n_2}{j_{21},0,n_2-j_{21}} \\
&\qquad + \sum_{j_{21}=0}^{n_2}\sum_{j_{22}=1}^{n_2-j_{21}}\binom{n_1}{n_1}\binom{n_2}{j_{21},j_{22},n-j_{21}-j_{22}} \\
&\qquad + \sum_{j_{11}=0}^{n_1-1}\sum_{j_{21}=0}^{n_2}\sum_{j_{22}=0}^{n_2-j_{21}}\binom{n_1}{j_{11}}\binom{n_2}{j_{21},j_{22},n_2-j_{21}-j_{22}} \\
&\qquad = \sum_{j_{11}=0}^{n_1}\sum_{j_{21}=0}^{n_2}\sum_{j_{22}=0}^{n_2-j_{21}}\binom{n_1}{j_{11}}\binom{n_2}{j_{21},j_{22},n_2-j_{21}-j_{22}},
\end{aligned}$$

0-dimensional bars which accounts for all 0-dimensional critical points of f.

Inductive step: Assume true for $k-1$

For the inductive step, suppose our formulas hold for $k-1$. We have

$$2^{n_1-i_{11}-j_{11}}\binom{n_1}{i_{11},k_1-i_{11},j_{11},n_1-k_1-j_{11}}\binom{n_2}{i_{21},i_{22},k_2-i_{21}-i_{22},j_{21},j_{22},n_2-k_2-j_{21}-j_{22}}$$

critical points of f in each class. Of these points, some of them must kill $(k-1)$-dimensional bars of lengths $\delta_2 - \beta_2$, $\gamma_2 - \alpha_2$, or $\gamma_1 - \beta_1$. For the bars of length $\delta_2 - \beta_2$, this corresponds to killing off

$$\begin{aligned}
2^{n_1-i_{11}-j_{11}}\sum_{\ell=0}^{i_{21}-1}(-1)^{\ell}\Bigg[&\binom{n_1}{i_{11},k_1-i_{11},j_{11},n_1-k_1-j_{11}} \\
&\binom{n_2}{j_{21},j_{22},n_2-k_2-j_{21}-j_{22}+\ell,i_{21}-1-\ell,i_{22},k_2-i_{21}-i_{22}}\Bigg]
\end{aligned}$$

$(k-1)$-dimensional bars. Note, this is due to the fact that if a critical point of f kills a bar of length $\delta_2 - \beta_2$, that corresponds to killing off a bar from

$$\text{class}\left(\begin{bmatrix} n_1 & k_1 & i_{11} & i_{12} & j_{11} & j_{12} \\ n_2 & k_2-1 & i_{21}-1 & i_{22} & j_{21} & j_{22} \end{bmatrix}\right).$$

Similarly for bars of length $\gamma_2 - \alpha_2$, this corresponds to killing off

$$2^{n_1-i_{11}-j_{11}} \sum_{\ell=0}^{k_2-i_{21}-i_{22}-1} (-1)^\ell \left[\binom{n_1}{i_{11}, k_1 - i_{11}, j_{11}, n_1 - k_1 - j_{11}} \right.$$
$$\left. \binom{n_2}{j_{21}, j_{22}+\ell, n_2 - k_2 - j_{21} - j_{22}, i_{21}, i_{22}, k_2 - i_{21} - i_{22} - 1 - \ell} \right]$$

$(k-1)$-dimensional bars. This corresponds to critical points of f of a given class killing off critical points of

$$\text{class}\left(\begin{bmatrix} n_1 & k_1 & i_{11} & i_{12} & j_{11} & j_{12} \\ n_2 & k_2 - 1 & i_{21} & i_{22} & j_{21} & j_{22} + 1 \end{bmatrix} \right).$$

Finally, for bars of length $\gamma_1 - \beta_1$, this corresponds to killing off

$$2^{n_1-i_{11}-j_{11}} \sum_{\ell=0}^{k_1-i_{11}-1} (-1)^\ell \left[\binom{n_1}{i_{11}, k_1 - i_{11} - 1, j_{11}, n_1 - k_1 - j_{11} + \ell} \right.$$
$$\left. \binom{n_2}{j_{21}, j_{22}, n_2 - k_2 - j_{21} - j_{22}, i_{21}, i_{22}, k_2 - i_{21} - i_{22}} \right]$$

bars. This corresponds to critical points of f of some class$(n, k, i_1, i_2, j_1, j_2)$ killing off critical points of

$$\text{class}\left(\begin{bmatrix} n_1 & k_1 - 1 & i_{11} & i_{12} & j_{11} & j_{12} \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix} \right).$$

Denote $\left| class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix} \right) \right|$ as $|Class|$. Hence we have

$$|Class| - 2^{n_1-i_{11}-j_{11}} \sum_{\ell=0}^{i_{21}-1} (-1)^\ell \left[\binom{n_1}{i_{11}, k_1 - i_{11}, j_{11}, n_1 - k_1 - j_{11}} \right.$$
$$\left. \binom{n_2}{j_{21}, j_{22}, n_2 - k_2 - j_{21} - j_{22} + \ell, i_{21} - 1 - \ell, i_{22}, k_2 - i_{21} - i_{22}} \right]$$
$$= \sum_{\ell'=0}^{i_{21}} (-1)^\ell \left[\binom{n_1!}{i_{11}, k_1 - i_{11}, j_{11}, n_1 - k_1 - j_{11}} \right.$$
$$\left. \binom{n_2!}{i_{21} - \ell, i_{22}, k_2 - i_{21} - i_{22}, j_{21}, j_{22}, n_2 - k_2 - j_{21} - j_{22} + \ell} \right]$$

bars of length $\delta_2 - \beta_2$ born from $class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)$ such that $i_{11}+i_{22} = k_1+k_2$, $j_{22}=0$, $n_1-k_1-j_{11}=0$, and $(n_1+n_2)-(k_1+k_2) > j_{11}+j_{21}$. Note, $k_1-i_{11}=0$ since there can be no critical points of type c_{11} in these classes. Additionally, $n_1-k_1-j_{11}=0$ since there can be no critical points of type b_{11} in these classes, which together gives that $2^{n_1-i_{11}-j_{11}}=1$.

Next, we have

$$\begin{aligned} |Class| - 2^{n_1-i_{11}-j_{11}} \sum_{\ell=0}^{(k_2-i_{21}-i_{22}-1)} (-1)^\ell \Bigg[&\binom{n_1}{i_{11}, k_1-i_{11}, j_{11}, n_1-k_1-j_{11}} \\ &\binom{n_2}{j_{21}, j_{22}+\ell, n_2-k_2-j_{21}-j_{22}, i_{21}, i_{22}, k_2-i_{21}-i_{22}-1-\ell} \Bigg] \\ = \sum_{\ell=0}^{k_2-i_{21}-i_{22}} (-1)^\ell \Bigg[&\binom{n_1!}{i_{11}, k_1-i_{11}, j_{11}, n_1-k_1-j_{11}} \\ &\binom{n_2}{i_{21}, i_{22}, k_2-i_{21}-i_{22}-\ell, j_{21}, j_{22}+\ell, n_2-k_2-j_{21}-j_{22}} \Bigg] \end{aligned}$$

bars of length $\gamma_2 - \alpha_2$ born from $class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)$ such that $j_{22} > 0$, $i_{11}+j_{11} > 0$, and $n_1-k_1-j_{11}=0$. Again, $2^{n_1-i_{11}-j_{11}}=1$ since these classes have no critical points of types b_{11} or c_{11}.

Finally, we have

$$\begin{aligned} |Class| - 2^{n_1-i_{11}-j_{11}} \sum_{\ell=0}^{k_1-i_{11}-1} (-1)^\ell \Bigg[&\binom{n_1}{i_{11}, k_1-i_{11}-1, j_{11}, n_1-k_1-j_{11}+\ell} \\ &\binom{n_2}{j_{21}, j_{22}, n_2-k_2-j_{21}-j_{22}, i_{21}, i_{22}, k_2-i_{21}-i_{22}} \Bigg] \\ = 2^{n_1-i_{11}-j_{11}} \sum_{\ell=0}^{k_1-i_{11}} (-1)^\ell \Bigg[&\binom{n_1}{i_{11}, k_1-i_{11}-\ell, j_{11}, n_1-k_1-j_{11}+\ell} \\ &\binom{n_2}{i_{21}, i_{22}, k_2-i_{21}-i_{22}, j_{21}, j_{22}, n_2-k_2-j_{21}-j_{22}} \Bigg] \end{aligned}$$

bars of length $\gamma_1 - \beta_1$ born from $class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)$ such that $n_1 - k_1 - j_{11} > 0$.

Thus, we have completely characterized the sublevelset persistent homology for any branched alkane with internal bonds of types 2-2 and 3-2. □

Now that we have characterized a more complicated class of branched alkanes, we outline the process for characterizing any branched alkane. This same procedure can be used to characterize the sublevelset persistence of any additive function over a product space.

6.2 Generalizing for any additive function over a product space

Now that we have looked at a few specific cases, we can outline the process of characterizing the sublevelset persistence for any branched alkane, or more generally, for any additive function over a product space.

First, identify the number of different length bars over all component persistence diagrams. In the example above, we had four different bar lengths: semi-infinite, $\delta_2 - \beta_2$, $\gamma_2 - \alpha_2$, and $\gamma_1 - \beta_1$. Next, order all of the bars from longest to shortest. This gives the layout for the length matrix, $X = x_{q,r}$ discussed in Section 4.3. This information alone will allow us to determine the number of bars of each length, the number of bars in each dimension k, and the total number of bars.

To complete a full characterization, we need to determine the different classes of critical points of f. Thus, construct the class matrix; a matrix where each row ℓ contains the number of copies of each bond (n_ℓ), the index of the critical point of f_ℓ (k_ℓ), and where each letter (i_{i1}, j_{i1}, etc.) denotes a different type of critical point of component function f_i of a given bond type. Let r be the number of different types of components and let cr_i be the number of critical points of f_i in the i^{th} component. It is important to split classes via

ε-perturbations such that each class only corresponds to a birth or death. Later, this will allow us to determine pairings between birth classes and death classes. An example of the class matrix from the previous section is given below.

$$X_{class} = \begin{bmatrix} n_1 & k_1 & i_{11} & i_{12} & j_{11} & j_{12} \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}$$

Note, it is possible that not every component function will possess the same number of critical point classes. For example, most molecules will not need all indices. Bonds of type 1-2-2-1 will have classes of type $(n_1, k_1, i_{11}, i_{12} = 0, j_{11}, j_{12} = 0)$. When we perform these simplifications, we can reduce the number of classes by grouping certain classes together, such as in the 1-2-2-1 with 1-3-2-1 case. Below, we simplify the class matrix accordingly.

$$X_{class} = \begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}$$

Next, determine the number of points in each class. We use the multinomial coefficient to determine which critical points of f are contained in the given class. Recall from Lemma 2.5.1, if f is an additive function over a product space, for a point in f to be a critical point, each coordinate must be a critical point in its component function. Hence, we choose a critical point from each component function and group them accordingly. In the previous example, we had

$$\left| class \left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix} \right) \right| =$$

$$2^{n_1 - i_{11} - j_{11}} \binom{n_1}{i_{11}, k_1 - i_{11}, j_{11}, n_1 - k_1 - j_{11}} \binom{n_2}{i_{21}, i_{22}, k_2 - i_{21} - i_{22}, j_{21}, j_{22}, n_2 - k_2 - j_{21} - j_{22}}$$

points in each class.

If we want to use the simplification of combining i_{11} with i_{12} and j_{11} with j_{12} into the same class, we also have to account for our choice of two different points for each of those groupings. Hence, when counting bars with length $\gamma_1 - \beta_1$, we introduce a factor of $(2^{n_1-k_1-j_{11}})(2^{k_1-i_{11}}) = 2^{n_1-i_{11}-j_{11}}$.

Finally, we identify the birth time and length of the bars in each class. First, we identify the energy value associated with each class of critical points of f. The energy value of the previous example is given by the following function:

$$E\left(class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)\right) =$$
$$(j_{11})\alpha_1 + (n_{11} - k_{11} - j_{11})\beta_1 + (k_1 - i_{11})\gamma_1 + (i_{11})\delta_1$$
$$+ (j_{21})\alpha_2 + (j_{22})\alpha_2' + (n_2 - k_2 - j_{21} - j_{22})\beta_2 + (k_2 - i_{21} - i_{22})\gamma_2 + (i_{21})\delta_2 + (i_{22})\delta_2'.$$

where each energy value corresponds to the appropriate critical point of f.

All that remains to count the number of bars in each class. To do so, recall that each component critical point corresponds to either a birth or death (hence the separation of α and α' and δ and δ') in the previous example. Thus, each class that corresponds to deaths gets paired with a class that corresponds to births. From there, one can identify the appropriate alternating sum, as shown in Theorem 6.1.5 and which length bars they correspond to.

In practice, we turn to computations to guide our conjectures before proving them. For the 3-2 with 2-2 case, we computed the first sublevelset persistence barcode with GUDHI [22], and then by hand via the bar combination rules found in the persistent Künneth formula (Theorem 2.5.3). From here, we were able to label the barcode and identify which bars in the component barcodes were combined to create each bar in the energy landscape. Then we grouped the classes by bar length, births, and deaths, and used that to identify which classes paired together for each bar. After we had a guess as to what the counting functions

should be, we used code found at [21] to verify our guesses for small examples. From there, we prove why the counting formula is correct.

For example, take bars of length $\delta_2 - \beta_2$, as in the example in Section 6.1. We first identify the classes that correspond to the births and deaths of bars of this length. In this case, we looked at the sublevelset persistent barcode of 2-methylpentane and used that to identify the appropriate bars and classes. The restrictions for the classes that produce $\delta_2 - \beta_2$ length bars are

$$i_{11} + i_{22} = k_1 + k_2,\ j_{22} = 0,\ n_1 - k_1 - j_{11} = 0, \text{ and } (n_1 + n_2) - (k_1 + k_2) > j_{11} + j_{21}.$$

Next, for every birth class, we identified the corresponding class that killed each that particular class. Specifically for bars of length $\delta_2 - \beta_2$,

$$class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & 0 & j_{11} & 0 \\ n_2 & k_2 & i_{21} & i_{22} & j_{21} & j_{22} \end{bmatrix}\right) \quad \text{kills} \quad class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & i_{12} & j_{11} & j_{12} \\ n_2 & k_2 - 1 & i_{21} - 1 & i_{22} & j_{21} & j_{22} \end{bmatrix}\right)$$

Using this information, we can determine the alternating sum we need to pair births with deaths. Hence, for the example above we get the following alternating sum.

$$2^{n_1 - i_{11} - j_{11}} \sum_{\ell'=0}^{i_{21}} (-1)^{\ell} \left[\binom{n_1}{i_{11}, k_1 - i_{11}, j_{11}, n_1 - k_1 - j_{11}} \binom{n_2}{i_{21} - \ell, i_{22}, k_2 - i_{21} - i_{22}, j_{21}, j_{22}, n_2 - k_2 - j_{21} - j_{22} + \ell} \right]$$

Note, at $\ell = 0$, this formula gives the number of points in that particular class. Further, at $\ell = 1$, we get

$$\left| class\left(\begin{bmatrix} n_1 & k_1 & i_{11} & i_{12} & j_{11} & j_{12} \\ n_2 & k_2 - 1 & i_{21} - 1 & i_{22} & j_{21} & j_{22} \end{bmatrix}\right) \right|$$

which are the points in the class that corresponds to the death of bars of length $\delta_2 - \beta_2$. This verifies that we are subtracting off the points that correspond to deaths for this particular class.

From here, to finish finding the birth times and the length of bars born from each class, we perform the same process for the rest of the bars as we did for the bars of length $\delta_2 - \beta_2$. In general,

it is easiest to work from the longest bar to the shortest bar. This process gives the procedure to find the full characterization of the sublevelset persistence for any additive function over a product space.

Chapter 7

Conclusion and future work

The goal of this work was to characterize the sublevelset persistent homology for additive functions over a product space in the context of energy landscapes. In Chapter 2, we developed background on branched alkanes, sublevelset persistent homology, Morse theory and its connections to sublevelset persistence, and the Künneth formula. Next in Chapter 3, we looked at analytical descriptions of branched alkanes and considered two examples. Then we established results in Chapter 4 regarding the number of sublevelset persistent homology bars, the number of bars in each homological dimension, and the number of bars per length. Chapter 5 characterized branched alkanes consisting exclusively of dihedral types 1-3-2-1. Finally in Chapter 6, we looked at extending the characterization process to branched alkanes with internal bonds of type 2-2 and 3-2, and used that to show the method to characterize any additive function over a product space.

As a result of this work, we have established several results regarding the sublevelset persistence of additive functions over a product space in the context of energy landscapes. In Theorem 4.1.2, we established the total number of bars in terms of the number critical points in each component function. Next, in Theorem 4.2, we established results regarding the number of bars in any dimension k for any idealized branched alkane. We also established results regarding the number of bars of a given length in Theorem 4.3.2. We characterized two different kinds of branched alkanes, molecules with type 3-2 internal bonds and types 2-2 and 3-2 internal bonds in Theorems 5.2.6 and 6.1.5. Finally, we generalized the characterization process in Chapter 6. All of the results outlined above can be extended to look at different types of molecules. Further, they can be used to characterize additive functions over a product space in different settings. Additionally, these results will also be featured in a paper targeted toward chemists [24].

7.1 Future work

These results can be applied to any additive function over a product space. Hence, the more assumptions we can remove and replace with real data and information, the more realistic our

model becomes. One way to increase the accuracy is to remove the simplification of the dihedral type. For example, 2-methylpentane consists of two dihedral angles of type 1-3-2-2 and one dihedral angle of type 1-2-2-3. Throughout the course of this work, we have replaced these with simpler approximating functions, namely the dihedral types 1-3-2-1 and 1-2-2-1. If we were also given the formulas for the energy landscapes of the bonds 1-3-2-2 and 1-2-2-3, then the characterization process outlined in Chapter 6 would be the same, the only changes would be the component functions of the energy landscape. This process would also work for other appropriate functions.

Another potential path of future work includes looking at how changing the bond length affects the energy landscape. This would give different type of component function with domain $\mathbb{R}$, where the input would be bond length in angstroms, or some other unit of length. Thus, our energy function would be $f\colon (S^1)^n \times \mathbb{R}^m \to \mathbb{R}$ with n angular bonds and m bond lengths, which is still an additive function over a product space. More precisely, since bond lengths are always nonnegative, one might consider an energy function $f\colon (S^1)^n \times \mathbb{R}^m_{\geq 0} \to \mathbb{R}$, where $\mathbb{R}_{\geq 0}$ denotes the nonnegative reals. This would require us to consider manifolds with boundary, and to define critical points that are allowed to live in the boundary of the manifold. However, there do exist versions of Morse theory for manifolds with boundary. For example, Theorem 2.3.1 from [20] also holds for manifolds with boundary.

Cyclo-alkanes, alkane molecules that contain carbon loops, are another potential path of exploration. For example, consider cyclo-octane, a loop with 8 carbon atoms. The domain of the energy landscape of cyclo-octane is the union of a Klein bottle and a sphere along two circles of singularities [6]. This domain is not a product space, and the energy function does not decompose as an additive function over a product space. This prevents us from using most of the machinery described above. To circumvent this problem, we could instead treat the cyclo-alkane confirmation as a single input parameter. This is now a 2-dimensional input, whereas a single bond angle or bond length was a 1-dimensional input parameter. This would allow us to use the results from above to add cyclo-alkanes to chains of molecules. More rigorously, let D be the domain of the cyclo-alkane energy landscape. Then, if we attach a single cyclo-alkane to a branched alkane, then we could model its energy as a function $(S^1)^n \times D \to \mathbb{R}$ that is an additive combination of the energy on each piece. Similarly if we attached m cyclo-alkanes, we could model its energy as a function with

domain $(S^1)^n \times D^m \to \mathbb{R}$, much as we described with changing the bond length. Finally, we can extend this process to other molecules, such as inorganic molecules (non-carbon based), and see how the energy landscapes changes.

Another open problem consists of understanding the energy landscapes when interactions between non-bonded atoms are included. In this situation, not all bond angle arrangements are a possibility. For example, picture a long carbon chain such that the molecule can be arranged where the two ends of the chains touch. This would never happen; the hydrogens that fill in other bonds would repel the two ends away from each other. Thus, certain configurations will never be obtained, and the energy it would take to approach those configurations increases to infinity the closer they get. One way to address these configurations is by adding a function $h(\phi_1, \ldots, \phi_n)$ to the energy landscape, where h incorporates interactions between non-bonded atoms, and therefore can be thought of as describing how much the energy landscape deviates from being an additive function over a product space. If one restricts to situations or to energy regimes where the function h is small in magnitude, then the stability of sublevelset persistent homology [18, 25] would allow us to conclude that the added function does not have a large impact on the sublevelset persistence. Adapting this work to a model that accounts for non-bonded interactions is an open problem.

Energy landscapes are an important part of molecular chemistry. It is difficult to completely understand chemical reactions due to the inherent high dimensionality of the molecules. Topological data analysis, specifically sublevelset persistent homology, provides chemists with a summary of relevant information. But, as with most real-world situations, we either have to sacrifice some accuracy by approximating or have our evaluations take lots of time. The more variables added, the more realistic our model becomes but, even a slight change to improve computation time, such as using 1-2-2-1 to approximate 4-2-2-1, affects the real-word applicability. Going forward, it is important to balance both of these concerns as we look at different molecules, parameters, and configurations.

Bibliography

[1] Atanas Atanasov, Gunnar Carlsson, and Henry Adams. Nudged elastic band in topological data analysis. *Topological Methods in Nonlinear Analysis*, 45(1):247–272, Mar. 2015.

[2] Joshua Mirth, Yanqin Zhai, Johnathan Bush, Enrique G Alvarado, Howie Jordan, Mark Heim, Bala Krishnamoorthy, Markus Pflaum, Aurora Clark, Y Z, and Henry Adams. Representations of energy landscapes by sublevelset persistent homology: An example with n-alkanes, 2020.

[3] Oren M. Becker and Martin Karplus. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *The Journal of Chemical Physics*, 106(4):1495–1517, 1997.

[4] Ryan Gotchy Mullen, Joan-Emma Shea, and Baron Peters. Transmission coefficients, committors, and solvent coordinates in ion-pair dissociation. *Journal of Chemical Theory and Computation*, 10(2):659–667, 2014. PMID: 26580043.

[5] John D. Roberts and Marjorie C. Caserio. Book: Basic Principles of Organic Chemistry (Roberts and Caserio), Mar 5 2021. [Online; accessed 2022-02-07].

[6] Shawn Martin, Aidan Thompson, Evangelos A Coutsias, and Jean-Paul Watson. Topology of cyclo-octane energy landscape. *The journal of chemical physics*, 132(23):234115, 2010.

[7] Fabio Pietrucci and Wanda Andreoni. Graph theory meets ab initio molecular dynamics: atomic structures and transformations at the nanoscale. *Physical review letters*, 107(8):085504, 2011.

[8] Tiecheng Zhou, Ernesto Martinez-Baez, Gregory Schenter, and Aurora E Clark. Pagerank as a collective variable to study complex chemical transformations and their energy landscapes. *The Journal of chemical physics*, 150(13):134102, 2019.

[9] Hitesh Gakhar and Jose A Perea. Künneth formulae in persistent homology. *arXiv preprint arXiv:1910.05656*, 2019.

[10] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

[11] William L. Jorgensen and Julian Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[12] William L Jorgensen, Jeffry D Madura, and Carol J Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of the American Chemical Society*, 106(22):6638–6646, 1984.

[13] Marcus G. Martin and J. Ilja Siepmann. Novel configurational-bias monte carlo method for branched molecules. transferable potentials for phase equilibria. 2. united-atom description of branched alkanes. *The Journal of Physical Chemistry B*, 103(21):4508–4517, 1999.

[14] Jason R. Mick, Mohammad Soroush Barhaghi, Brock Jackman, Loren Schwiebert, and Jeffrey J. Potoff. Optimized mie potentials for phase equilibria: Application to branched alkanes. *Journal of Chemical & Engineering Data*, 62(6):1806–1818, 2017.

[15] Kelin Xia and Guo-Wei Wei. Persistent homology analysis of protein structure, flexibility, and folding. *International journal for numerical methods in biomedical engineering*, 30(8):814–844, 2014.

[16] Kelin Xia, Xin Feng, Yiying Tong, and Guo-Wei Wei. Persistent homology for the quantitative prediction of fullerene stability. *Journal of computational chemistry*, 36(6):408–422, 2015.

[17] Jacob Townsend, Cassie Putman Micucci, John H Hymel, Vasileios Maroulas, and Konstantinos D Vogiatzis. Representation of molecular structures with persistent homology for machine learning applications in chemistry. *Nature Communications*, 11(1):1–9, 2020.

[18] Herbert Edelsbrunner and John L Harer. *Computational Topology: An Introduction*. American Mathematical Society, Providence, 2010.

[19] John Milnor. *Morse theory*, volume 51. Princeton university press, 2016.

[20] Augustin Banyaga and David Hurtubise. *Basic Morse Theory*, pages 45–91. Springer Netherlands, Dordrecht, 2004.

[21] Biswajit Sadhu and Brittany Story. Delta_branched_alkanes, 2022. https://github.com/brimcarr/DELTA_branched_alkanes.

[22] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The Gudhi library: Simplicial complexes and persistent homology. In *International Congress on Mathematical Software*, pages 167–174. Springer, 2014.

[23] Joshua Mirth, Johnathan Bush, Mark Heim, and Henry Adams. deltapersistence, 2020. https://gitlab.com/delta-topology-public/deltapersistence.

[24] Henry Adams, Aurora Clark, Biswajit Sadhu, and Brittany Story. Molecular configurations and persistence: Additive energies and branched alkanes (*In Preparation*). 2022.

[25] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.