Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

Article

On the loop homology of a certain complex of RNA structures

Thomas J. X. Li [1,†] and Christian M. Reidys [2,3,†*]

[1] Biocomplexity Institute and Initiative, UVA; jl9gx@virginia.edu
[2] Biocomplexity Institute and Initiative, UVA;
[3] Mathematics Department, UVA; cmr3hk@virginia.edu
* Correspondence: cmr3hk@virginia.edu
† Current address: Physical Address: Biocomplexity Institute, Town Center Four, 994 Research Park Boulevard, Charlottesville, VA 22911 | Mailing Address: Biocomplexity Institute, P.O. Box 400298, Charlottesville, VA 22904-4298

Abstract: In this paper we establish a topological framework of τ-structures to quantify the evolutionary transitions between two RNA sequence-structure pairs. τ-structures developed here consist of a pair of RNA secondary structures together with a non-crossing partial matching between the two backbones. The loop complex of a τ-structure captures the intersections of loops in both secondary structures. We compute the loop homology of τ-structures. We show that only the zeroth, first and second homology groups are free. In particular, we prove that the rank of the second homology group equals the number γ of certain arc-components in a τ-structure, and the rank of the first homology is given by $\gamma - \chi + 1$, where χ is the Euler characteristic of the loop complex.

Keywords: topology; simplicial complex; homology; Mayer-Vietoris sequence; RNA; secondary structure

1. Introduction

In this paper we study triples, (S, T, ϕ), consisting of an ordered pair of RNA secondary structures, (S, T), [1–6] together with a non-crossing partial matching, ϕ, between the two backbones. We shall denote such a triple a transition structure or τ-*structure* and note that the mapping ϕ relates homologous bases between two underlying RNA sequences.

The topological framework of τ-structures developed here is designed to quantify the probability of evolutionary transitions between RNA sequence-structure pairs subject to specific sequence constraints. In an algorithmic guise, τ-structures play a central role in solving the following computational problem: given two arbitrary RNA secondary structures, determine a pair of specifically related sequences that minimizes the total free energies of the two sequence-structure pairs.

Our main objective is to compute the simplicial homology of the *loop complex* of a τ-structure. In order to state our main theorem we next introduce the notions of ϕ-crossing and arc-component.

Two arcs (i_1, i_2) and (j_1, j_2) are ϕ-crossing, i.e. if either
(1) both i_1 and i_2 are incident to ϕ-arcs such that $\phi(i_1) < j_1 < \phi(i_2) < j_2$ or $j_1 < \phi(i_1) < j_2 < \phi(i_2)$, or
(2) both j_1 and j_2 are incident to ϕ-arcs such that $i_1 < \phi^{-1}(j_1) < i_2 < \phi^{-1}(j_2)$ or $\phi^{-1}(j_1) < i_1 < \phi^{-1}(j_2) < i_2$.

ϕ-crossing induces an equivalence relation whose nontrivial equivalence classes are called *arc-components*. An arc-component is *of type* 1 if both endpoints of any arc are incident to ϕ-arcs.

We shall prove the following main result:

© 2021 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

Main Theorem. *Let $I = (S, T, \phi)$ be a τ-structure and $H_n(I)$ the n-th homology group of its loop complex, $K(I)$. Then*

$$H_3(I) = 0,$$
$$H_2(I) \cong \bigoplus_{k=1}^{\gamma} \mathbb{Z},$$
$$H_1(I) \cong \bigoplus_{k=1}^{\gamma-\chi+1} \mathbb{Z},$$

where γ denotes the number of arc-components of type 1 and χ the Euler characteristic of $K(I)$.

Our framework generalizes the work of [7,8] on bi-structures. Bi-structures can be viewed as particular τ-structures, namely those over two identical sequences. The loop complex of a bi-structure exhibits only a nontrivial zeroth and second homology, the latter being freely generated by crossing components [7,8]. On the algorithmic side τ-structures play a similar role as bi-structures in [9], where a Boltzmann sampler of sequences that are simultaneously compatible with two structures is presented.

The loop complex of τ-structures exhibits a variety of new features. First it has in general a nontrivial first homology group. Secondly, it features two types of arc-components. Those of type 1, that correspond to spheres and arc-components of type 2, that correspond to surfaces with nontrivial boundary, see Section 7.

To prove the main theorem, we first establish basic properties of the loop complex $X = K(I)$. Secondly, we manipulate X, by means of simplicial collapses and then dissect K_1, a certain sub-complex of dimension one. As a result we derive $\tilde{X}_2$, a specific X-sub-complex, which we organize into its crossing components. We note that $\tilde{X}_2$ is not unique, as its construction depends on certain choices. We then proceed with a combinatorial analysis of $\tilde{X}_2$-crossing components, that lays the foundation for proving that their geometric realizations are either spheres or surfaces with boundary.

The combinatorics of crossing components, specifically Lemma 6, controls furthermore the way crossing components are glued and this observation facilitates to prove the main result in two stages. First computing the homology of components in isolation and secondly computing the homology of the loop complex by Mayer-Vietoris sequences. Several of our arguments are reminiscent of ideas of classical facts, as for instance the Jordan curve theorem in the proof of Lemma 7.

The paper is organized as follows: in Section 2, we recall the definitions of RNA secondary structures, bi-structures and their loop complexes. In Section 3, we introduce τ-structures and present some basic facts of their loop complexes. In Section 4, we simplify the loop complex in multiple rounds via simplicial collapses and topological stratification. In Section 5, we investigate the combinatorics of crossing components in the simplified complex. We then compute the homology of components in Section 6 and integrate this information in order to obtain the homology of τ-structures in Section 7. Finally we discuss our results in Section 8.

2. Secondary structures, bi-structures and the loop complex

An RNA secondary structure encapsulates the nucleotide interactions within a single RNA sequence. It can be represented as a *diagram*, a labeled graph over the vertex set $[n] := \{1, \ldots, n\}$, whose vertices are arranged in a horizontal line and arcs are drawn in the upper half-plane. Each vertex corresponds to a nucleotide in the primary sequence and each *arc*, denoted by (i, j), represents the base pairing between the i-th and j-th nucleotides in the RNA structure. Two arcs (i_1, j_1) and (i_2, j_2) are *crossing* if $i_1 < i_2 < j_1 < j_2$. An RNA *secondary structure* is defined as a diagram satisfying the following three conditions [1,3]: (1) if (i, j) is an arc, then $j - i \geq 2$, (2) any two arcs do not have a common vertex, (3) any two arcs are non-crossing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

Given a secondary structure S, the set $[1,n] := \{1,\ldots,n\}$ is called the *backbone* and an *interval* $[i,j] \subset [1,n]$ denotes the set of consecutive vertices $\{i,i+1,\ldots,j-1,j\}$. A vertex k is *covered* by an arc (i,j) if $i \leq k \leq j$ and there exists no other arc (p,q) such that $i < p < k < q < j$. The set of vertices covered by an arc (i,j) is called a *loop*, s, and we refer to (i,j) as the *covering arc* of s. We shall equip each diagram with two "formal" vertices at positions 0 and $n+1$ together with the arc $(0,n+1)$, which we call the *rainbow*. Its associated distinguished loop is called the exterior loop. Each loop, s, can be represented as a disjoint union of intervals on the backbone of S, $s = \bigcup_{k=1}^{l}[i_k,j_k]$, such that (i_1,j_l) and (j_k,i_{k+1}) for $1 \leq k \leq l-1$ form arcs and any other vertices in s are unpaired. By construction, a diagram S is in one-to-one correspondence with its set of arcs as well as its set of loops $\{s_i\}_i$.

Let $\mathcal{X}$ be a collection of finite sets. A subset $\{X_0,X_1,\ldots,X_d\} \subset \mathcal{X}$ is defined to be a *d-simplex* of $\mathcal{X}$ if the set intersection $\cap_{k=0}^{d} X_k \neq \varnothing$. Let $K_d(\mathcal{X})$ be the set of all d-simplices of $\mathcal{X}$. The *nerve* of $\mathcal{X}$ is a simplicial complex given by $K(\mathcal{X}) = \bigcup_{d=0}^{\infty} K_d(\mathcal{X}) \subseteq 2^{\mathcal{X}}$.

The nerve formed by the collection $\{s_i\}_i$ of S-loops is referred to as the *loop complex* $K(S)$ of a secondary structure S. The fact that secondary structures correspond to non-crossing diagrams immediately implies that the loop complex of a secondary structure is a tree.

Let S and T be two secondary structures. The pair of secondary structures S and T over $[n]$, $R = (S,T)$, is called a *bi-structure*. We represent a bi-structure as a diagram on a horizontal backbone with the S-arcs drawn in the upper and the T-arcs drawn in the lower half plane. The nerve formed by S-loops and T-loops of a bi-structure, R, is called the *loop complex*, $K(R)$.

In bi-structures [7], two arcs α,α' are equivalent if there exists a sequence of arcs $\alpha_1 = \alpha,\alpha_2,\ldots,\alpha_l = \alpha'$ such that two consecutive arcs α_k and α_{k+1} are crossing for $1 \leq k \leq l-1$. A *crossing component* of $R = (S,T)$ is an equivalence class that contains at least two crossing arcs.

The loop complex of a bi-structure has a nontrivial zeroth and second homology group. In particular, the first homology $H_1(R)$ is zero, and the second homology $H_2(R)$ is free and its rank equals the number of crossing components of the bi-structure R [8].

3. Some basic facts

Let S and T be two secondary structures over $[n]$ and $[m]$, respectively. We shall refer to their backbones as $[n]_S = \{1_S,2_S,\ldots,n_S\}$ and $[m]_T = \{1_T,2_T,\ldots,m_T\}$. We refer to the pair $(i,\phi(i))$ as a *ϕ-arc* between $i \in [n]_S$ and $\phi(i) \in [m]_T$. Two ϕ-arcs $(i_1,\phi(i_1))$ and $(i_2,\phi(i_2))$ are *crossing* if $i_1 < i_2$ and $\phi(i_1) > \phi(i_2)$.

Definition 1. *A τ-structure I is a triple (S,T,ϕ), consisting of an ordered pair of secondary structures, (S,T), together with a partial matching ϕ between their vertex sets $[n]_S$ and $[m]_T$ such that any two ϕ-arcs are non-crossing.*

In analogy to the case of bi-structures, we shall represent the diagram of a τ-structure as a labeled graph over two horizontal backbones, in which S-arcs are drawn in the upper and T-arcs are drawn in the lower halfplane, with the edges drawn in between the two backbones, representing ϕ, see Fig. 1.

A bi-structure is a particular τ-structure, having two backbones of the same length and ϕ being the identity map.

The partial matching ϕ induces an equivalence relation $\sim$ on the set of vertices of the two backbones by identifying for each pair $(i,\phi(i))$ the vertices i and $\phi(i)$. The ϕ-arcs then correspond one-to-one to nontrivial equivalence classes of size two. This equivalence relation gives rise to the definition of a loop in a τ-structure.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

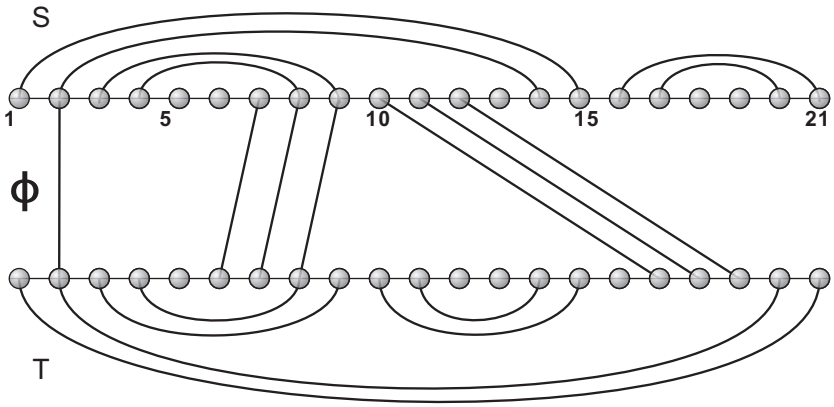

Figure 1. A τ-structure.

Definition 2. *Given a S- or T-vertex $v \in [n]_S \cup [m]_T$, an I-*vertex*, $\bar{v}$, is an equivalence class induced by v, i.e. $\bar{v} \in [n]_S \cup [m]_T / \sim$. Given a loop q, the I-*loop*, $\bar{q}$, is the set of equivalence classes induced by the set of q-vertices.*

Now we are in position to introduce the loop complex of a τ-structure:

Definition 3. *The nerve formed by I-loops is called the* loop complex *$K(I) = \bigcup_{d=0}^{\infty} K_d(I)$, where $K_d(I)$ denotes the set of d-simplices i.e. the non-empty intersections of d distinct loops. $H_n(I)$ shall denote the n-th homology group of the loop complex of I.*

Let us proceed by collecting some basic properties of the loop complex.

Proposition 1. *Let $I = (S, T, \phi)$ be a τ-structure and v_1, v_2 be two distinct S-vertices. Then $\bar{v}_1 \neq \bar{v}_2$. In particular there exists a bijection between vertices in an S-loop s and vertices in the corresponding $\bar{S}$-loop $\bar{s}$.*

Proof. By definition, any two ϕ-arcs in the partial matching ϕ do not share common endpoints. Thus ϕ maps v_1 and v_2 to different T-vertices. Therefore the equivalence classes $\bar{v}_1$ and $\bar{v}_2$ are different. □

Proposition 2. *Let $I = (S, T, \phi)$ be a τ-structure and $\bar{s}_1$, $\bar{s}_2, \bar{s}_3$ be three different $\bar{S}$-loops. Then*
(1) either $\bar{s}_1 \cap \bar{s}_2 = \varnothing$ or $|\bar{s}_1 \cap \bar{s}_2| = 2$.
(2) $\bar{s}_1 \cap \bar{s}_2 \cap \bar{s}_3 = \varnothing$.

Proof. For the secondary structure S, any two S-loops s_1 and s_2 either have trivial intersection or intersect at two endpoints of one arc, i.e., either $s_1 \cap s_2 = \varnothing$ or $|s_1 \cap s_2| = 2$. Proposition 1 guarantees that $|\bar{s}_1 \cap \bar{s}_2| = |s_1 \cap s_2|$, whence assertion (1). Any S-vertex is contained in at most two distinct S-loops. Thus $s_1 \cap s_2 \cap s_3 = \varnothing$. By Proposition 1, $\bar{s}_1 \cap \bar{s}_2 \cap \bar{s}_3 = \varnothing$. □

Proposition 3. *Let I be a τ-structure with loop complex $K(I)$. Then any 2-simplex $\sigma \in K_2(I)$ has the form $\{s_1, s_2, t_1\}$ or $\{t_1, t_2, s_1\}$, where s_1, s_2 are $\bar{S}$-loops and t_1, t_2 are $\bar{T}$-loops.*

Proof. By construction, the three loops in a 2-simplex σ have a non-empty intersection. In view of Proposition 2, not all these three loops are from the same secondary structure. Therefore σ is of the form $\{s_1, s_2, t_1\}$ or $\{t_1, t_2, s_1\}$. □

A 1-simplex $\sigma = \{r_1, r_2\} \in K(I)$ is called *pure* if both r_1 and r_2 belong to the same secondary structure and *mixed*, otherwise. Proposition 3 shows that any 2-simplex contains exactly one pure edge and two mixed edges.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

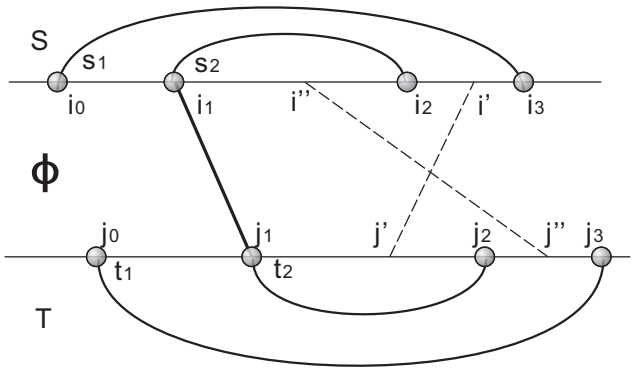

Figure 2. Any 3-simplex contains a free mixed 1-face. The loops s_1, s_2, t_1, t_2 have the covering arcs (i_0, i_3), (i_1, i_2), (j_0, j_3), (j_1, j_2), respectively. In view of the ϕ-arc (i_1, j_1), the covering arcs form a 3-simplex $\sigma = \{s_1, s_2, t_1, t_2\}$. If the mixed edges $\{s_1, t_2\}$ and $\{s_2, t_1\}$ were not σ-free, then there would exist two ϕ-arcs (i', j') and (i'', j''), which are crossing.

Proposition 4. *Let I be a τ-structure with loop complex $K(I)$. Then any 3-simplex $\sigma \in K_3(I)$ has the form $\{s_1, s_2, t_1, t_2\}$, where s_1, s_2 are $\bar{S}$-loops and t_1, t_2 are $\bar{T}$-loops.*

Proof. Suppose that $\sigma = \{r_1, r_2, r_3, r_4\}$. As a face of σ, $\{r_1, r_2, r_3\} \in K_2(I)$. By Proposition 3, we can w.l.o.g. set that $r_1 = s_1$, $r_2 = s_2$, and $r_3 = t_1$. Similarly, the face $\{r_1, r_2, r_4\} \in K_2(I)$ implies that r_4 is necessarily a $\bar{T}$-loop, i.e., $r_4 = t_2$. □

By abuse of notation we refer to $\bar{S}$- and $\bar{T}$-loops simply as S- and T-loops, respectively.

Definition 4. *Given a simplex $\sigma \in K(I)$, a face $\tau \subset \sigma$ is called σ-free if, for any simplex $\sigma_1 \in K(I)$ containing τ, σ_1 is a face of σ. Clearly, σ is maximal in $K(I)$. Moreover, σ is the unique maximal simplex that contains τ.*

Proposition 5. *Let I be a τ-structure with loop complex $K(I)$. Then any 3-simplex $\sigma \in K_3(I)$ contains a σ-free, mixed 1-face τ.*

Proof. Let $\sigma = \{s_1, s_2, t_1, t_2\}$. Suppose that the covering arcs of the loops s_1, s_2, t_1, t_2 are (i_0, i_3), (i_1, i_2), (j_0, j_3), (j_1, j_2), respectively and that we have the distinguished ϕ-arc (i_1, j_1) with $\phi(i_1) = j_1$. Then the loops s_1, s_2, t_1, t_2 intersect at the equivalence class of $\bar{i_1} = \bar{j_1}$ and w.l.o.g. we may assume $i_0 < i_1 < i_2 < i_3$ and $j_0 < j_1 < j_2 < j_3$, see Fig. 2.

We shall prove the proposition by contradiction. Suppose that none of the mixed 1-faces of σ are free, we consider the two particular mixed edges $\tau_1 = \{s_1, t_2\}$ and $\tau_2 = \{s_2, t_1\}$. By assumption, there exist 2-simplices Δ_1 and Δ_2, such that $\tau_1 \subset \Delta_1$, $\tau_2 \subset \Delta_2$, neither of which being faces of σ.

By construction, Δ_1 corresponds to a ϕ-arc (i', j') distinct from (i_1, j_1). Since $\tau_1 = \{s_1, t_2\} \subset \Delta_1$, the ϕ-arc (i', j') connects two intervals $[i_0, i_1)_S \cup [i_2, i_3]_S$ and $(j_1, j_2]_T$, i.e., $i' \in [i_0, i_1)_S \cup [i_2, i_3]_S$ and $j' \in (j_1, j_2]_T$. Note that if $i' \in [i_0, i_1)_S$, then $i' < i_1$ and $j' > j_1$, i.e. the ϕ-arcs (i', j') and (i_1, j_1) are crossing, a contradiction. Hence we derive that the ϕ-arc (i', j') connects $[i_2, i_3]_S$ and $(j_1, j_2]_T$, see Fig. 2.

Similarly, Δ_2 corresponds to a ϕ-arc (i'', j'') distinct from (i_1, j_1). Since $\tau_2 \subset \Delta_2$, it follows that the ϕ-arc (i'', j'') connects two intervals $(i_1, i_2]_S$ and $[j_2, j_3]_T$, see Fig. 2. In view of $i'' < i'$ and $j'' > j'$, the ϕ-arcs (i', j') and (i'', j'') are crossing, contradicting the fact that ϕ-arcs are non-crossing.

Accordingly, the non-crossing property of ϕ-arcs implies that either one, $\{s_1, t_2\}$ or $\{s_2, t_1\}$ is σ-free. □

Proposition 6. *Given a loop complex $K(I)$, then any 3-simplex $\sigma \in K_3(I)$ contains at least two σ-free 2-faces.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

Proof. By Proposition 5, σ contains at least one σ-free mixed edge τ. Let Δ_1 and Δ_2 denote the 2-faces of σ containing τ. Suppose that there exists a simplex σ_1 satisfying $\Delta_1 \subset \sigma_1$ and $\sigma_1 \not\subset \sigma$. Then σ_1 also contains τ, contradicting τ being σ-free. Therefore Δ_1 is σ-free, and analogously we derive that Δ_2 is σ-free. □

Proposition 7. *Let I be a τ-structure with loop complex $K(I)$. Then $H_0(I) \cong \mathbb{Z}$ and $H_d(I) = 0$ for $d \geq 4$.*

Proof. By assumption, ϕ induces at least one ϕ-arc, which in turn gives rise to at least a 1-simplex connecting an S-loop and a T-loop. Therefore $K(I)$ is path-connected and $H_0(I) \cong \mathbb{Z}$. By the pigeonhole principle, for any set of $d+1$ I-loops with $d \geq 4$, at least three belong to the same secondary structure. By Proposition 2, these three loops intersect trivially. Thus, for $d \geq 4$, the intersection of any $d+1$ loops is trivial. As a result, $K(I)$ does not contain simplices of dimension four or higher. Therefore $K_d(I) = \varnothing$ and $H_d(I) = 0$ for $d \geq 4$. □

4. From X to $\tilde{X}_2$

We now begin computing the homology of the loop complex of a τ-structure. Directly constructing its discrete Morse function [10,11] or gradient vector field seems rather involved. We shall take a slightly different approach along the lines of discrete Morse theory. We shall simplify the loop complex in multiple rounds making it amendable to its homological analysis. One of these simplifications is obtained via elementary collapses, a reduction introduced by J.H.C. Whitehead [12]. These collapses can be understood in terms of non-critical points of the Morse function.

4.1. Simplicial collapses

The notion of elementary collapse is closely related to the freeness of a simplex. Given a complex K, let σ be a maximal simplex and Δ be a σ-free face with codimension 1. The removal of σ and Δ constitutes *elementary collapse* [12,13] and gives rise of a new complex $K \setminus \{\sigma, \Delta\}$. We say that K *simplicial collapses* to a sub-complex K', denoted by $K \searrow K'$, if there is a sequence of complexes $K_0 = K, K_1, \ldots, K_q = K'$, such that K_i is obtained from K_{i-1} via an elementary collapse. Clearly, K and K' are homotopy equivalent.

In the following, we shall employ simplicial collapses to reduce the loop complex in two steps: first eliminating all tetrahedra, and then removing triangles having a free edge.

Step 1: given a 3-simplex $\sigma_1 = \{s_1, s_2, t_1, t_2\}$, Proposition 5 guarantees that σ_1 always has a free mixed edge τ and two free triangles Δ_1, Δ_1' containing τ. The triple $(\tau, \Delta_1, \Delta_1')$ is called a σ_1-*butterfly*. In analogy to the analysis in case of bi-structures [7], we perform the butterfly removal on the loop complex. Specifically, the simplicial collapse consists of two elementary collapses with respect to (σ_1, Δ_1) and (Δ_1', τ) gives us a sub-complex $K_1 = K(I) \setminus \{\sigma_1, \Delta_1, \Delta_1', \tau\}$, which is homotopy equivalent to $K(I)$.

We observe that K_1 is again the loop complex of a τ-structure $I_1 = (S_1, T_1, \phi_1)$, obtained by splicing vertices in I as follows: for any I-vertex $q \in s_1 \cap s_2 \cap t_1 \cap t_2$, suppose that q is the equivalence class of an S-vertex i and a T-vertex j. Then there is a ϕ-arc connecting i and j, i.e., $\phi(i) = j$. We now splice i into two consecutive vertices i_1, i_2, and splice j into two consecutive vertices j_1, j_2. The ϕ-arc $\phi(i) = j$ is replaced with two new ϕ-arcs $\phi_1(i_1) = j_1$ and $\phi_1(i_2) = j_2$. The arcs are connected in a way such that the mixed edge τ is not formed in K_1. In case of $\tau = \{s_1, t_2\}$, we depict the resulting τ-structure I_1 in Fig. 3, and note that all other cases are treated similarly. Furthermore we remove any other ϕ-arcs that give rise to Δ_1, Δ_1' or τ. It is straightforward to verify that the loop complex of I_1 is exactly K_1.

By iteratively eliminating all 3-simplices σ and the associated σ-butterflies, we derive a sequence of sub-complexes $K_0 = K(I), K_1, \ldots, K_q = X$ such that X does not contain any 3-simplices. Accordingly, we obtain a sequence of τ-structures $I_0 = I, I_1, \ldots, I_q = I'$ by splicing the corresponding vertices of σ such that the loop complex of I_j is given by K_j. It is clear that the sub-complex X satisfies the the following property.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

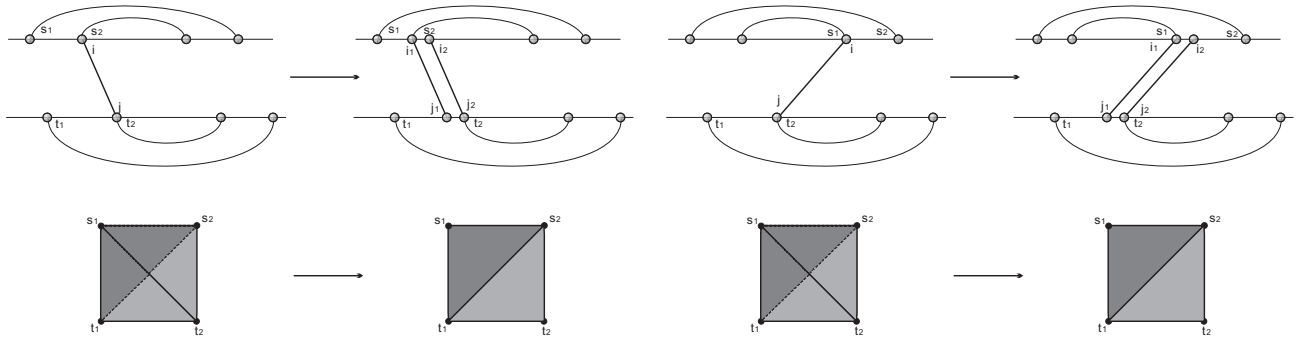

Figure 3. The σ_1-butterfly removal. The simplicial collapse of the butterfly $(\tau, \Delta_1, \Delta_1')$ corresponds to splicing vertices in the τ-structure. Here, $\sigma_1 = \{s_1, s_2, t_1, t_2\}$, $\Delta_1 = \{s_1, t_1, t_2\}$, $\Delta_1' = \{s_1, s_2, t_2\}$, and $\tau = \{s_1, t_2\}$. The vertex i is split into two consecutive vertices i_1, i_2, and j is split into two consecutive vertices j_1, j_2. The ϕ-arc $\phi(i) = j$ is replaced with two new ϕ-arcs $\phi_1(i_1) = j_1$ and $\phi_1(i_2) = j_2$. Depending on the situation, the arcs are connected in a way such that the loop s_1 and the loop t_2 do not intersect.

Lemma 1. *Let $K(I)$ be the loop complex of a τ-structure I, and X be the complex obtained by removing from $K(I)$ all 3-simplices σ and the associated σ-butterflies. Then*
(1) X is the complex of a τ-structure I' such that each ϕ-arc is incident to at most one arc,
(2) X is a sub-complex of $K(I)$ such that the 0-th skeleton $X^{(0)} = K(I)^{(0)}$ and X does not contain any 3-simplices,
(3) $H_n(X) \cong H_n(I)$, for $n \geq 0$.

A complex having the properties of Lemma 1 is called *lean*. As an immediate consequence we obtain:

Corollary 1. *Let I be a τ-structure, then $H_3(I) = 0$.*

Proof. In view of Proposition 1, as a sub-complex of $K(I)$, X does not contain any 3-simplices. For X holds $\ker \partial_3 = 0$, whence $H_3(I) \cong H_3(X) = 0$. □

Step 2: suppose that Δ_1 is a 2-simplex in X having a free 1-face τ_1. By collapsing (Δ_1, τ_1), we obtain a sub-complex $X_1 = X \setminus \{\Delta_1, \tau_1\}$. By successively deleting 2-simplices τ together with τ-free edges, we derive a sequence of sub-complexes $X_0 = X, X_1, \ldots, X_m = \tilde{X}$ such that each 2-simplex of $\tilde{X}$ does not contain any free 1-face.

Proposition 8. *Let $\tilde{X}$ be a complex obtained from X by iteratively removing all 2-simplices τ together with τ-free 1-face. Then*
(1) $\tilde{X}$ is a sub-complex of X such that the 0-th skeleton $\tilde{X}^{(0)} = X^{(0)}$ and each 2-simplex of $\tilde{X}$ does not have a free 1-face,
(2) for $n \geq 0$, we have $H_n(\tilde{X}) \cong H_n(X)$.

Fig. 4 displays the simplicial collapse from X to $\tilde{X}$.

Remark: The sub-complex $\tilde{X}$ is not unique. By construction, $\tilde{X}$ depends on the particular sequence of the simplicial collapses, as well as which free face of a simplex is being removed, see Fig. 5.

Lemma 1 guarantees that $\tilde{X}$ is derived from a lean complex.

Lemma 2. *Let $\tilde{X}$ be the sub-complex derived from a lean complex, then*
(1) for any 2-simplex $\Delta = \{s_1, s_2, t_1\} \in \tilde{X}$, holds $s_1 \cap s_2 \cap t_1 = \{\bar{i}\}$, i.e. any 2-simplex Δ corresponds to the ϕ-arc, $(i, \phi(i))$,
(2) any pure 1-simplex τ of $\tilde{X}$ is either maximal or contained in exactly two 2- simplices Δ, Δ'. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

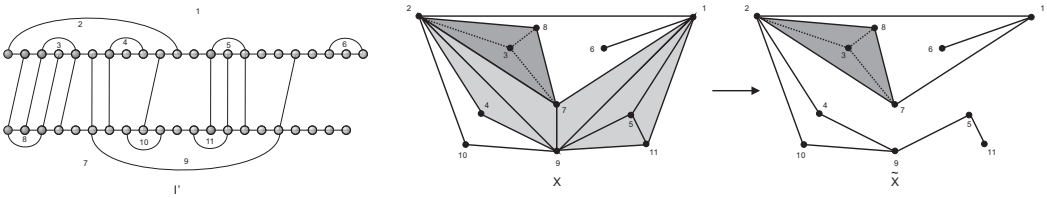

Figure 4. Simplicial collapse. LHS: a τ-structure I' in which each ϕ-arc is incident to at most one arc. RHS: X, the complex of I' and $\tilde{X}$. The underlying sequence of elementary collapses that generates $\tilde{X}$ is given by $(\{1,5,11\},\{1,11\})$, $(\{5,9,11\},\{9,11\})$, $(\{1,5,9\},\{1,5\})$, $(\{1,7,9\},\{1,9\})$, $(\{2,7,9\},\{7,9\})$, $(\{2,4,9\},\{2,9\})$. The distinguished, empty tetrahedron (darker gray level and dashed lines) cannot be eliminated by simplicial collapses.

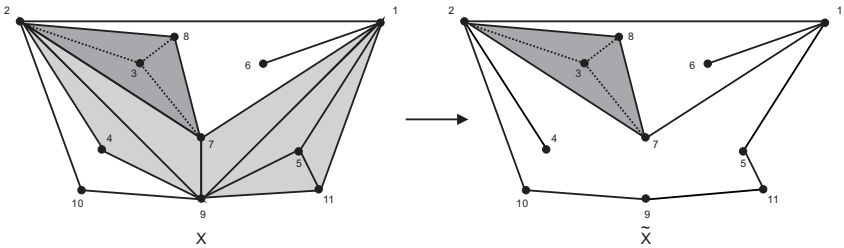

Figure 5. $\tilde{X}$ is not unique. Performing the following sequence of elementary collapses on the complex X, displayed in Fig. 4, $(\{2,4,9\},\{4,9\})$, $(\{2,7,9\},\{2,9\})$, $(\{1,7,9\},\{7,9\})$, $(\{1,5,9\},\{1,9\})$, $(\{1,5,11\},\{1,11\})$, $(\{5,9,11\},\{5,9\})$ produces a different $\tilde{X}$.

the latter case, τ is associated with an arc (i,i') in I, such that vertices i,i' are incident to ϕ-arcs (i,j) and (i',j'), respectively.

Proof. For any $\Delta = \{s_1,s_2,t_1\} \in \tilde{X}$, we have $|s_1 \cap s_2 \cap t_1| \leq |s_1 \cap s_2| = 2$. By construction, in $\tilde{X}$, τ is not Δ-free, whence there exists another 2-simplex $\Delta' = \{s_1,s_2,t_1'\} \neq \Delta$ containing τ. Lemma 1 guarantees that $\tilde{X}$ does not contain any 3-simplices, i.e. we have

$$(s_1 \cap s_2 \cap t_1) \cap (s_1 \cap s_2 \cap t_1') = s_1 \cap s_2 \cap t_1 \cap t_1' = \varnothing.$$

Therefore sets $s_1 \cap s_2 \cap t_1$ and $s_1 \cap s_2 \cap t_1'$ are disjoint and both have cardinality 1, whence assertion (1) follows. Let $s_1 \cap s_2 \cap t_1 = \{\bar{i}\}$, then Δ is in one-to-one correspondence with the vertex i as well as its induced ϕ-arc $(i,\phi(i))$ in I.

Suppose the pure 1-simplex τ is not maximal, then there exists at least one 2-simplex Δ, having τ as a face. Since we work in the complex $\tilde{X}$, τ cannot be Δ-free, whence there exists another 2-simplex Δ' containing τ. Assertion (1) then implies that Δ and Δ' are the only such 2-simplices, completing the proof. □

4.2. Topological stratification

We shall next show that the complex $\tilde{X}$ is topologically *stratified*, decomposing into two sub-complexes $\tilde{X}_2$ and K_1, where $\tilde{X}_2$ is induced by all $\tilde{X}$-2-simplices and K_1 by all 1-simplices of $\tilde{X} \setminus \tilde{X}_2$, see Fig. 6.

Proposition 9. *We have $H_2(\tilde{X}) \cong H_2(\tilde{X}_2)$ and $H_1(\tilde{X}) \cong H_1(\tilde{X}_2) \bigoplus \oplus_{i=1}^{k}\mathbb{Z}$, where the non-negative integer k depends on K_1 and $\tilde{X}_2 \cap K_1$.*

Proof. By construction, $\tilde{X}_2$ is a sub-complex of $\tilde{X}$. We have the short exact sequence of chain complexes $0 \to C_*(\tilde{X}_2) \xrightarrow{i} C_*(\tilde{X}) \xrightarrow{j} C_*(\tilde{X},\tilde{X}_2) \to 0$, which induces the long exact sequence

$$\cdots \to H_{n+1}(\tilde{X},\tilde{X}_2) \xrightarrow{\partial} H_n(\tilde{X}_2) \xrightarrow{i_*} H_n(\tilde{X}) \xrightarrow{j_*} H_n(\tilde{X},\tilde{X}_2) \to \cdots.$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

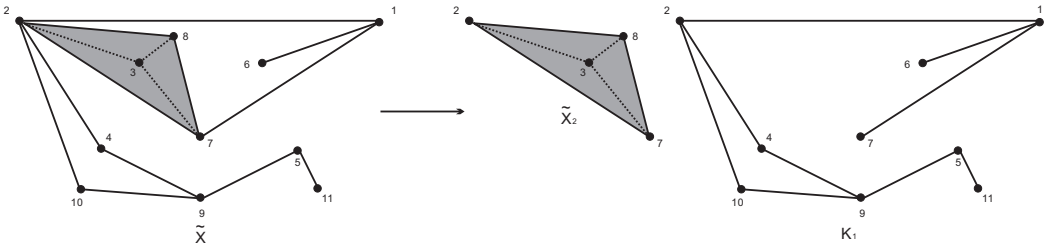

Figure 6. Topological stratification. $\tilde{X}$ stratifies into a sub-complex $\tilde{X}_2$ induced by all 2-simplices and K_1 induced by the remaining 1-simplices.

By construction, $\tilde{X}$ does not have any 3-simplices, which implies that $C_3(\tilde{X}, \tilde{X}_2) = 0$. Since $\tilde{X}_2$ contains all 2-simplices in $\tilde{X}$, we have $C_2(\tilde{X}, \tilde{X}_2) = 0$. Then we derive $H_3(\tilde{X}, \tilde{X}_2) = H_2(\tilde{X}, \tilde{X}_2) = 0$. In view of the long exact sequence, we have

$$0 = H_3(\tilde{X}, \tilde{X}_2) \xrightarrow{\partial} H_2(\tilde{X}_2) \xrightarrow{i_*} H_2(\tilde{X}) \xrightarrow{j_*} H_2(\tilde{X}, \tilde{X}_2) = 0,$$

i.e. $H_2(\tilde{X}) \cong H_2(\tilde{X}_2)$. For $n = 1$, the long exact sequence reads

$$0 = H_2(\tilde{X}, \tilde{X}_2) \xrightarrow{\partial} H_1(\tilde{X}_2) \xrightarrow{i_*} H_1(\tilde{X}) \xrightarrow{j_*} H_1(\tilde{X}, \tilde{X}_2),$$

inducing the exact sequence

$$0 \xrightarrow{\partial} H_1(\tilde{X}_2) \xrightarrow{i_*} H_1(\tilde{X}) \xrightarrow{j_*} \operatorname{Im} j_* \longrightarrow 0.$$

The relative homology $C_1(\tilde{X}, \tilde{X}_2)$ is freely generated by the 1-simplices in $\tilde{X} \setminus \tilde{X}_2$. By definition, we have $H_1(\tilde{X}, \tilde{X}_2) = \ker(\bar{\partial}_1)/\operatorname{Im}(\bar{\partial}_2)$, where $C_2(\tilde{X}, \tilde{X}_2) \xrightarrow{\bar{\partial}_2} C_1(\tilde{X}, \tilde{X}_2) \xrightarrow{\bar{\partial}_1} C_0(\tilde{X}, \tilde{X}_2)$. In view of $\operatorname{Im}(\bar{\partial}_2) = 0$, $H_1(\tilde{X}, \tilde{X}_2) \cong \ker(\bar{\partial}_1)$ and $H_1(\tilde{X}, \tilde{X}_2)$ is free. As a subgroup of the free group $H_1(\tilde{X}, \tilde{X}_2)$, $\operatorname{Im} j_*$ is free and thus projective. Then the short exact sequence is split exact, and

$$H_1(\tilde{X}) \cong H_1(\tilde{X}_2) \oplus \operatorname{Im} j_* \cong H_1(\tilde{X}_2) \bigoplus \oplus_{i=1}^{k} \mathbb{Z},$$

where the non-negative integer $k = \operatorname{rnk}(\operatorname{Im} j_*)$ depends on K_1 and $\tilde{X}_2 \cap K_1$. $\square$

Proposition 9 shows that the sub-complex K_1 gives rise to only free generators in the first homology of $\tilde{X}$.

5. Some combinatorics of crossing components

To compute the homology of $\tilde{X}$, we introduce an equivalence relation partitioning the set of 2-simplices. To this end, we observe that 2-simplices of $\tilde{X}$ appear in pairs: given a 2-simplex Δ with pure edge, τ, Lemma 2 guarantees that there exists a unique 2-simplex Δ' such that $\Delta \cap \Delta' = \tau$. Each pair (Δ, Δ') is associated with a unique arc (i_1, i_2) and i_1 and i_2 are incident to ϕ-arcs, $(i_1, \phi(i_1))$ and $(i_2, \phi(i_2))$. We refer to (Δ, Δ') together with their pure and ϕ-arcs as a *couple*.

Two couples (Δ_1, Δ_1') and (Δ_2, Δ_2') are *ϕ-crossing* if their corresponding arcs (i_1, i_2) and (j_1, j_2) are ϕ-crossing, i.e. if either
(1) both i_1 and i_2 are incident to ϕ-arcs such that $\phi(i_1) < j_1 < \phi(i_2) < j_2$ or $j_1 < \phi(i_1) < j_2 < \phi(i_2)$, or
(2) both j_1 and j_2 are incident to ϕ-arcs such that $i_1 < \phi^{-1}(j_1) < i_2 < \phi^{-1}(j_2)$ or $\phi^{-1}(j_1) < i_1 < \phi^{-1}(j_2) < i_2$.
This notion induces an equivalence relation $\sim_\phi$ on all $\tilde{X}$-couples to be the transitive closure with respect to ϕ-crossing. Two couples (Δ_1, Δ_1') and (Δ_k, Δ_k') are ϕ-equivalent, denoted by $(\Delta_1, \Delta_1') \sim_\phi (\Delta_k, \Delta_k')$, if there exists a sequence of pairs $(\Delta_1, \Delta_1'), (\Delta_2, \Delta_2'), \ldots, (\Delta_k, \Delta_k')$, such

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

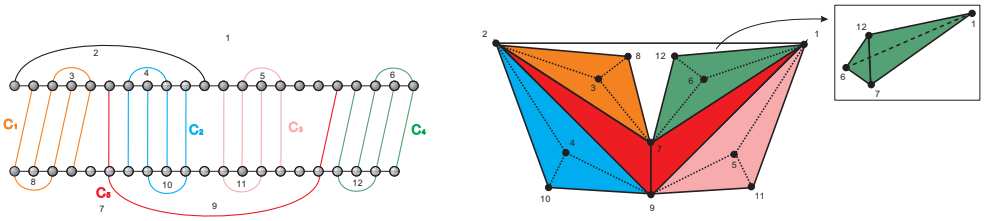

Figure 7. Components. LHS: a τ-structure I' with labeled loops. RHS: the complex $K(I')$ of I' embedded in the Euclidean space $\mathbb{R}^3$. Note that $K(I')$ identical to $\tilde{X}$. The 2-simplices in $K(I')$ are partitioned into five components. The insert (RHS) depicts the component in green, which is an empty tetrahedron.

that (Δ_i, Δ'_i) and $(\Delta_{i+1}, \Delta'_{i+1})$ are ϕ-crossing for $1 \leq i \leq k-1$. The relation $\sim_\phi$ partitions the set of $\tilde{X}$-pairs into equivalence classes, $[(\Delta_1, \Delta'_1)]$. A *ϕ-crossing component* or *component*, C, of $\tilde{X}$ is the set of 2-simplices contained in an equivalence class $[(\Delta, \Delta')]$. By construction, the set of all 2-simplices in $\tilde{X}$ is partitioned into components, C, and each component C induces a $\tilde{X}$-sub-complex. By abuse of notation, we shall denote a component, as well as its induced $\tilde{X}$-sub-complex, by C. We refer to the ϕ-arcs of a component C as *C-ϕ-arcs*.

Suppose that $(i_1, j_1), (i_2, j_2), \ldots, (i_l, j_l)$, where $i_1 < i_2 < \cdots < i_l$ and $\phi(i_k) = j_k$ are C-ϕ-arcs. These partition the two backbones into *blocks*, the outer block $\beta_1 = [1, i_1)_S \cup (i_l, n]_S \cup [1, j_1)_T \cup (j_l, m]_T$ and the inner blocks $\beta_r = (i_{r-1}, i_r)_S \cup (j_{r-1}, j_r)_T$, $2 \leq r \leq l$. A block, β, *contains* a couple (Δ, Δ') if all endpoints of ϕ-arcs associated with (Δ, Δ') are contained in β and the component C_1 if β contains all C_1-couples.

Fig. 7 demonstrates a decomposition of $\tilde{X}$ into components. In this example, the complex $K(I')$ does *a priori* not contain any free 1-face, whence $K(I') = \tilde{X}$. It is straightforward to verify that $\tilde{X}$ features the five components:

$$\begin{aligned}
C_1 =& \{\{2,3,8\}, \{2,3,7\}, \{2,7,8\}, \{3,7,8\}\},\\
C_2 =& \{\{2,4,9\}, \{2,4,10\}, \{2,9,10\}, \{4,9,10\}\},\\
C_3 =& \{\{1,5,9\}, \{1,5,11\}, \{1,9,11\}, \{5,9,11\}\},\\
C_4 =& \{\{1,6,7\}, \{1,6,12\}, \{1,7,12\}, \{6,7,12\}\},\\
C_5 =& \{\{2,7,9\}, \{1,7,9\}\}.
\end{aligned}$$

The ϕ-arcs of C_5 induce the outer block $[1,5)_S \cup (19,22]_S \cup [1,5)_T \cup (18,21]_T$ and the inner block $(7,17)_S \cup (7,16)_T$. The components C_1, C_4 are contained in the outer and C_2, C_3 in the inner block, respectively.

Lemma 3. *Let C be a $\tilde{X}$-component with blocks $\beta_1, \beta_2, \ldots, \beta_l$. Then, for any other component $C_1 \neq C$, there exists a unique block, β_k, that contains C_1.*

Proof. Suppose that C-ϕ-arcs are given by $(i_1, j_1), (i_2, j_2), \ldots, (i_l, j_l)$, where $i_1 < i_2 < \cdots < i_l$ and $\phi(i_k) = j_k$.

Claim: Given a C_1-couple (Δ, Δ') with ϕ-arcs (i, j) and (i', j'), then there exists a block β_k that contains both (i, j) and (i', j').

Firstly, both endpoints i, j are contained in the same C-block, as otherwise (i, j) would cross a C-ϕ-arc. Suppose (i, j) and (i', j') are contained in different C-blocks. Since C and C_1 are distinct components, C- and C_1-couples are mutually non-crossing. As a result C_1 partitions the C-couples into to two non-empty subsets, one contained in $(i, i')_S \cup (j, j')_T$ and the other in $[1, i)_S \cup (i', n]_S \cup [1, j)_T \cup (j', m]_T$. By construction any couple contained in the former subset and any pair in the latter are non-crossing, which is impossible since C is a component.

It follows that there exists a C-block β_k that contains both: (i, j) and (i', j').

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

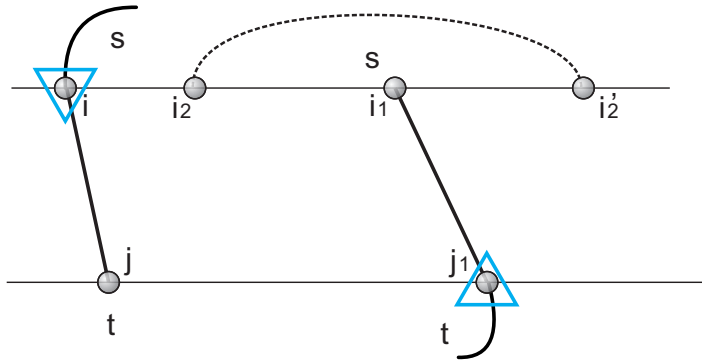

Figure 8. The planarity of a τ-structure. Triangles Δ and Δ_1 share the mixed edge, $\tau = \{s, t\}$ and correspond to the ϕ-arcs (i, j) and (i_1, j_1). If an S-arc (i_2, i_2') has its endpoints contained in $(i, i_1)_S$ and $[1, i)_S \cup (i_1, n]_S$, then i, i_1 belong to distinct S-loops, a contradiction.

Any C_1-couple (Δ_1, Δ_1') that crosses (Δ, Δ') has at least one ϕ-arc contained in β_k. By the above argument, both of its ϕ-arcs are then contained in β_k, i.e. β_k contains (Δ_1, Δ_1'). Since C_1 is a component this implies that any C_1-couple is contained in β_k. $\square$

Lemma 3 gives rise to a relation over the set of $\tilde{X}$-components as follows: $C_1 \prec C_2$ if all C_1-ϕ-arcs are contained in an inner block of C_2. In view of Lemma 3, we can verify that the relation $\prec$ is a well-defined partial order. In Fig. 7, $C_2 \prec C_5$, $C_3 \prec C_5$ and C_1, C_4, C_5 are maximal.

We next identify how two 2-simplices, that share a mixed edge affect the location of couples along the two backbones. It reflects the planarity of τ-structures and has important consequences for how components are glued in the loop complex.

Lemma 4. *Let Δ and Δ_1 be $\tilde{X}$-2-simplices that share the mixed edge, τ and that correspond to the ϕ-arcs (i, j) and (i_1, j_1), where $i < i_1$. Then any couple (Δ_2, Δ_2') where $\{\Delta_2, \Delta_2'\} \cap \{\Delta, \Delta_1\} = \varnothing$ is contained in either*

$$(i, i_1)_S \cup (j, j_1)_T \quad \text{or} \quad [1, i)_S \cup (i_1, n]_S \cup [1, j)_T \cup (j_1, m]_T.$$

Proof. Without loss of generality we may assume that (Δ_2, Δ_2') features the S-arc (i_2, i_2'). Let $\tau = \{s, t\}$, since $\Delta \cap \Delta_1 = \tau$, by construction, $i, i_1 \in s$ and $j, j_1 \in t$. In case one of the vertices i_2, i_2' is contained in $(i, i_1)_S$ and the other is contained in $[1, i)_S \cup (i_1, n]_S$, the arc (i_2, i_2') organizes i, i_1 into distinct S-loops, see Fig. 8, a contradiction. Thus, i_2, i_2' are contained in $(i, i_1)_S$ or $[1, i)_S \cup (i_1, n]_S$, i.e. the ϕ-arcs associated with (Δ_2, Δ_2') are contained in $(i, i_1)_S \cup (j, j_1)_T$ or $[1, i)_S \cup (i_1, n]_S \cup [1, j)_T \cup (j_1, m]_T$. $\square$

Lemma 5. *Let C be a $\tilde{X}$-component. Then for any 2-simplex $\Delta \in C$ and any 1-face τ, there exists at most one 2-simplex $\Delta_1 \in C$ such that $\Delta \cap \Delta_1 = \tau$.*

Proof. Clearly, the statement holds when τ is pure. For mixed τ, w.l.o.g. let $\Delta = \{s_1, s_2, t\}$, $\tau = \{s_2, t\}$ and assume the S-loop s_2 corresponds to the arc (i, i') where $i < i'$.

Suppose there exist distinct 2-simplices $\Delta_1, \Delta_2 \in C$, having τ as a 1-face. We shall denote the Δ, Δ_1 and Δ_2-ϕ-arcs by (i, j), (i_1, j_1), (i_2, j_2), respectively. Without loss of generality, we may assume that $i < i_1 < i_2$. Since any ϕ-arcs are non-crossing, we have $j < j_1 < j_2$.

We now apply Lemma 4: since $\Delta \cap \Delta_2 = \tau$, the couple (Δ_1, Δ_1') is contained in $(i, i_2)_S \cup (j, j_2)_T$ or $[1, i)_S \cup (i_2, n]_S \cup [1, j)_T \cup (j_2, m]_T$. Since $i < i_1 < i_2$, (Δ_1, Δ_1') is contained in $(i, i_2)_S \cup (j, j_2)_T$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

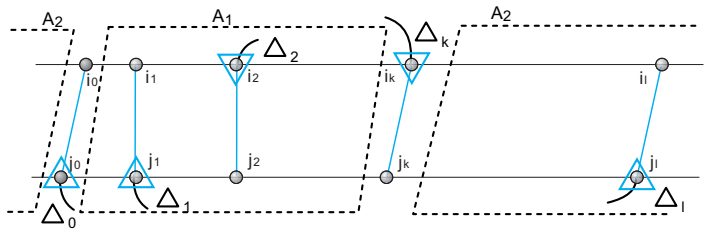

Figure 9. The fact $\Delta_0 \cap \Delta_k = \tau$ organizes C-couples into two subsets $A_1 \subset (i_0, i_k)_S \cup (j_0, j_k)_T$ and $A_2 \subset [1, i_0)_S \cup (i_k, n]_S \cup [1, j_0)_T \cup (j_k, m]_T$, marked by dashed parallelograms.

In view of (Δ, Δ') and (Δ_1, Δ_1') belonging to C, there exists a sequence of pairs $(\alpha_1, \alpha_1') = (\Delta_1, \Delta_1'), (\alpha_2, \alpha_2'), \ldots, (\alpha_k, \alpha_k') = (\Delta, \Delta')$ such that (α_i, α_i') and $(\alpha_{i+1}, \alpha_{i+1}')$ are crossing.

Lemma 4 applies to any couple (α_i, α_i'), locating it in $(i, i_2)_S \cup (j, j_2)_T$ or $[1, i)_S \cup (i_2, n]_S \cup [1, j)_T \cup (j_2, m]_T$. In addition, since (α_2, α_2') and (Δ_1, Δ_1') are crossing, (α_2, α_2') is contained in $(i, i_2)_S \cup (j, j_2)_T$, as is $(\alpha_1, \alpha_1') = (\Delta_1, \Delta_1')$. Analogously, it follows that $(\alpha_k, \alpha_k') = (\Delta, \Delta')$ is contained in $(i, i_2)_S \cup (j, j_2)_T$, which is, by construction, impossible.

Accordingly, there exists at most one 2-simplex $\Delta_1 \in C$ such that $\Delta \cap \Delta_1 = \tau$. □

Let C be a component and τ a C-1-face. Lemma 5 implies that τ is contained in either one or two C-2-simplices. We call a 1-face τ that is contained in a unique C-2-simplex *C-free*. Furthermore we call a component C *complete* if, for any of its 1-faces τ, there exist exactly two 2-simplices $\Delta, \Delta' \in C$ such that $\Delta \cap \Delta' = \tau$ and *incomplete*, otherwise.

Let C be a component with ϕ-arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_l, j_l)$, where $i_1 < i_2 < \cdots < i_l$ and $\phi(i_k) = j_k$. Let $\Delta_1, \Delta_2, \ldots, \Delta_l$ denote the 2-simplices associated with the ϕ-arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_l, j_l)$, respectively. Suppose that C-ϕ-arcs partition the two backbones into the sequence of blocks $\beta_1, \beta_2, \ldots, \beta_l$, ordered from left to right and where $\beta_1 = [1, i_1)_S \cup (i_l, n]_S \cup [1, j_1)_T \cup (j_l, m]_T = (i_l, i_1)_S \cup (j_l, j_1)_T$ denotes the outer-block. Let τ_k and τ_k' denote the mixed 1-faces of Δ_k that are associated with β_k and β_{k+1}, respectively. We call the mixed edges τ_1 and τ_l' associated with the outer block the *C-boundary*. Note that τ_1 and τ_l' coincide iff Δ_1 and Δ_l share a common mixed edge.

The next lemma constitutes a key observation which facilitates the computation of the homology of τ-structures via the Mayer-Vietoris sequence, see Theorem 2.

Lemma 6. *Let C be an $\tilde{X}$-component having boundary τ_1 and τ_l', and let C_1 be an $\tilde{X}$-component with the property $C_1 \not\prec C$. Suppose that τ is a 1-simplex shared by C and C_1, then $\tau = \tau_1$ or $\tau = \tau_l'$.*

Proof. Let $\Delta_1, \Delta_2, \ldots, \Delta_l$ denote the 2-simplices of C associated with the ϕ-arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_l,$ ordered from left to right. Suppose that $\Delta_0 \in C_1$ and $\Delta_k \in C$ share a common mixed edge τ and have the ϕ-arcs (i_0, j_0) and (i_k, j_k), respectively. In view of $C_1 \not\prec C$, C_1 is contained in the outer C-block and we may, w.l.o.g. assume that $i_0 < i_1$.

The fact $\Delta_0 \cap \Delta_k = \tau$, now organizes C relative to Δ_k in a particular fashion, see Fig. 9. Employing Lemma 4 we obtain that, except of (Δ_k, Δ_k'), all C-couples partition into

$$A_1 \subset (i_0, i_k)_S \cup (j_0, j_k)_T \quad \text{and} \quad A_2 \subset [1, i_0)_S \cup (i_k, n]_S \cup [1, j_0)_T \cup (j_k, m]_T.$$

Now we use that C is a component: firstly any A_1-couple and any A_2-couple are by construction non-crossing. Secondly, by construction, (Δ_k, Δ_k') is crossing into either only one A_1 or A_2, which implies that either one, A_1 or A_2 is trivial.

Suppose $A_1 = \varnothing$. Then all C-ϕ-arcs are contained in $[i_k, n]$ which implies $\Delta_k = \Delta_1$. In case $A_2 = \varnothing$ all C-ϕ-arcs are contained in $(i_0, i_k]$ which implies $\Delta_k = \Delta_l$. Therefore for the shared 1-edge, τ, holds $\tau = \tau_1$ or $\tau = \tau_l'$. □

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

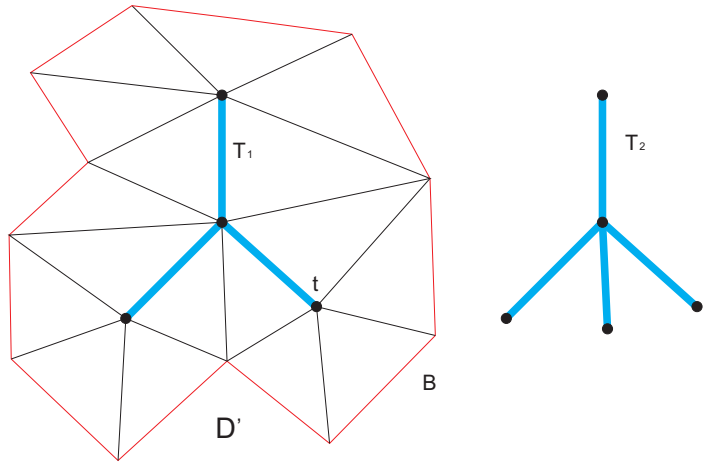

Figure 10. The boundary of D'. The trees T_1 and T_2 (blue) represent the two connected components of T_{D_i}. The sub-complex D' is formed by 2-simplices of D_i containing at least one T_1-vertex. Then the boundary B of D' (red) consists of only S-loops.

6. The homology of components

In order to compute the homology of τ-structures, we shall first compute the homology of components separately and then integrate this information via the Mayer-Vietoris sequence [14,15]. To this end we adopt a topological perspective of crossing components, taking a closer look at the manifold formed by 2-simplices of a component obtained via gluing along their edges.

Given a component C, two couples $(\Delta_1, \Delta_1') \sim_1 (\Delta_k, \Delta_k')$ if there exists a sequence of pairs $(\Delta_1, \Delta_1'), (\Delta_2, \Delta_2'), \ldots, (\Delta_k, \Delta_k')$ such that (Δ_i, Δ_i') and $(\Delta_{i+1}, \Delta_{i+1}')$ share at least one 1-face for $1 \leq i \leq k-1$. Clearly, $\sim_1$ is an equivalence relation, and partitions all pairs in C into equivalence classes $D_1, D_2, \ldots, D_k$, which we refer to as *C-ribbons*.

Lemma 7. *The geometric realization of a complete $\tilde{X}$-component, C is a sphere.*

Proof. Suppose that $\sim_1$ partitions all C-pairs into the ribbons $D_1, D_2, \ldots, D_k$.

Claim 1: The geometric realization of a ribbon, D_i, is a surface without boundary.

Since C is complete, for each 2-simplex $\Delta \in D_i$ and its 1-face τ there exists a unique 2-simplex $\Delta_1 \in C$ such that $\Delta \cap \Delta_1 = \tau$. By construction, Δ_1 is furthermore contained in D_i. This implies that the geometric realization of D_i is a compact, connected 2-manifold, i.e., a surface without boundary.

Let T_{D_i} denote the D_i-sub-complex consisting of all pure edges whose loops are T-loops. We shall employ T_{D_i} in order to relate the crossing relation with $\sim_1$.

Claim 2: T_{D_i} is connected, i.e., T_{D_i} is a subtree of T.

Suppose that T_{D_i} contains at least two connected components T_1 and T_2. Let D' denote the subset of 2-simplices of D_i, containing at least one T_1-vertex. Since all 2-simplices of D' are incident to the connected component T_1, the complex formed by D' is connected. D' does not contain any T_2-vertices, since otherwise there would, by construction of D', exist a pure edge connecting a T_1-vertex and a T_2-vertex, contradicting the assumption that T_{D_i} is disconnected. Thus, D' is a connected, proper subset of D_i, and its geometric realization is a surface having boundary B, see Fig. 10.

The set of 2-simplices containing a fixed T_1-vertex, t, forms a neighborhood of t in the D_i-induced surface, whence any T_1-vertex is contained in the interior of D'. Consequently, the boundary does not contain any T_1-vertex. From this we conclude that the boundary, being a cycle, consists of only S-loops. This, however, is impossible since the complex of a secondary structure S is a tree.

Claim 3: C consists of a unique ribbon.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

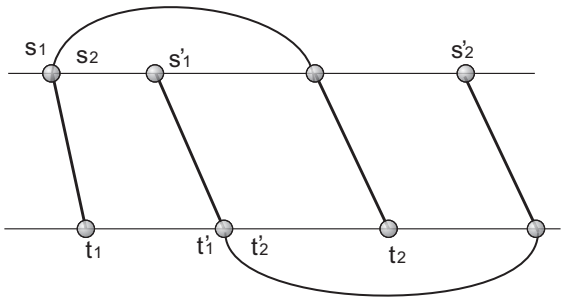

Figure 11. A couple $(\Delta, \Delta') \in D_1$ and $(\Delta_0, \Delta_0') \notin D_1$ that are crossing, where $\Delta = \{s_1, s_2, t_1\}$, $\Delta' = \{s_1, s_2, t_2\}$, $\Delta_0 = \{t_1', t_2', s_1'\}$ and $\Delta_0' = \{t_1', t_2', s_2'\}$.

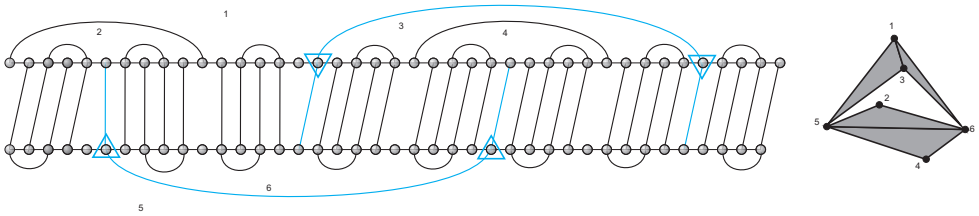

Figure 12. LHS: a τ-structure having an incomplete component (colored in blue, with its loops labeled by numbers and its 2-simplices labeled by triangles). RHS: the complex induced by the incomplete component indicates that it contains two ribbons.

Suppose that C contains at least two ribbons D_1 and D_2. Since C is a component, there exists a couple (Δ, Δ') in D_1 and (Δ_0, Δ_0') not contained in D_1, such that (Δ, Δ') and (Δ_0, Δ_0') are crossing.

Without loss of generality, we may consider $\Delta = \{s_1, s_2, t_1\}$, $\Delta' = \{s_1, s_2, t_2\}$, $\Delta_0 = \{t_1', t_2', s_1'\}$ and $\Delta_0' = \{t_1', t_2', s_2'\}$, see Fig. 11. Since (Δ, Δ') and (Δ_0, Δ_0') are crossing, the T-loops t_1', t_2' are on the path connecting T-loops t_1 and t_2 in the complex of the secondary structure T.

By Claim 2, T_{D_1} is connected and thus $t_1, t_2 \in T_{D_1}$ implies that $\{t_1', t_2'\} \in T_{D_1}$. Claim 1 guarantees that each 1-face of D_1 is not D_1-free, whence there exist two D_1-2-simplices having the face $\{t_1', t_2'\}$. However, (Δ_0, Δ_0') is the unique couple, whose 2-simplices, Δ_0, Δ_0' both contain the pure 1-simplex $\{t_1', t_2'\}$. As a result the fact that D_1 is a surface without boundary guarantees, that (Δ_0, Δ_0') is contained in D_1, which is a contradiction.

Therefore C is organized as a single ribbon, whose geometric realization is a surface without boundary. We shall proceed by computing its Euler characteristic χ.

Suppose that C contains n pairs (Δ, Δ'). Then the complex C features $2n$ 2-simplices and $3n$ 1-simplices. Claim 2 stipulates the connectivity of S_C and T_C, whence both: S_C and T_C are connected subsets of trees and as such trees themselves. We shall use this in order to count 0-simplices as follows: each pair (Δ, Δ') corresponds to one pure 1-simplex in S_C or T_C. Thus S_C and T_C contain n 1-simplices, which implies that C contains $n + (1 + 1)$ 0-simplices. From this follows $\chi = 2$ and C is homeomorphic to a sphere. $\square$

Remark: the key in the proof is to show that the projection of D_i onto each secondary structure, T_{D_i}, is connected. If T_{D_i} is disconnected, then we can construct a cycle separating different connected components and consisting of only S-loops, leading to a contradiction. The proof is reminiscent of Jordan curve theorem in the plane [16,17].

While any complete component is organized as a distinguished ribbon, an incomplete component can consist of multiple ribbons. In Fig. 12, the two couples $(\{2, 5, 6\}, \{4, 5, 6\})$ and $(\{1, 3, 5\}, \{1, 3, 6\})$ form an incomplete component which contains two ribbons, each of which being a couple.

Lemma 8. *Given an incomplete $\tilde{X}$-component C, having the ribbons $D_1, D_2, \ldots, D_k$. Then, for each D_i, there exists some 1-face $\tau \in D_i$ that is D_i-free.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

Proof. Suppose there exists a ribbon D_i, in which each 1-face $\tau \in D_i$ is not D_i-free. Lemma 5 guarantees that there exist exactly two 2-simplices $\Delta, \Delta' \in D_i$ such that $\Delta \cap \Delta' = \tau$. Using the argument of Claim 2 of Lemma 7, we can conclude that T_{D_i} is connected.

C consists of at least two distinct ribbons since otherwise $C = D_i$ contains no C-free 1-face, contradicting the assumption that C is incomplete. Thus there exist a couple $(\Delta, \Delta') \in D_i$ and $(\Delta_0, \Delta_0') \notin D_i$, such that (Δ, Δ') and (Δ_0, Δ_0') are crossing. In analogy to Claim 3 of Lemma 7, we can conclude that both Δ_0 and Δ_0' are contained in D_i, which is impossible.

Therefore there exists a 1-face $\tau \in D_i$ that is D_i-free. □

Theorem 1. *Let C be a component of $\tilde{X}$. Then*

$$H_2(C) \cong \begin{cases} \mathbb{Z} & \text{if } C \text{ is complete} \\ 0 & \text{if } C \text{ is incomplete} \end{cases} \tag{1}$$

and

$$H_1(C) \cong \begin{cases} 0 & \text{if } C \text{ is complete} \\ \oplus_{k=1}^{r} \mathbb{Z} & \text{if } C \text{ is incomplete,} \end{cases} \tag{2}$$

where the non-negative integer r depends on C. Furthermore, a complete component, $H_2(C)$ is freely generated by the sum of all C-2-simplices.

Proof. In case C is complete, by Lemma 7, C is homeomorphic to a sphere, whence $H_2(C) \cong \mathbb{Z}$ and $H_1(C) \cong 0$. Clearly, $H_2(C)$ is freely generated by the sum of 2-simplices of C.

In case C is incomplete, suppose that $D_1, D_2, \ldots, D_k$ are the C-ribbons. In the following we prove eqs. (1) and (2) by induction on the number k of ribbons.

For $k = 1$, C is a ribbon and as such is connected. Since C is incomplete, it contains by Lemma 8 at least one C-free 1-face. This implies that the geometric realization of C is a surface with boundary. Therefore $H_2(C) \cong 0$ and $H_1(C)$ is free.

For the induction step, we shall combine ribbons in order to compute the homology of an incomplete component. In view of the fact that C is the union of the sub-complexes D_k and $D' = \cup_{j<k} D_j$, we have the following inclusion maps:

$$\begin{tikzcd}
 & D_k \arrow[dr, "h"] & \\
D_k \cap D' \arrow[ur, "i"] \arrow[dr, "j"'] & & C = D_k \cup D' \\
 & D' \arrow[ur, "l"'] &
\end{tikzcd}$$

Each inclusion map induces a chain map on the corresponding simplicial chain groups and a homomorphism between the corresponding homology groups.

$$\begin{tikzcd}
 & H_n(D_k) \arrow[dr, "h_*"] & \\
H_n(D_k \cap D') \arrow[ur, "i_*"] \arrow[dr, "j_*"'] & & H_n(C) \\
 & H_n(D') \arrow[ur, "l_*"'] &
\end{tikzcd}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

Accordingly we obtain the Mayer-Vietoris sequence

$$\cdots \to H_n(D_k \cap D') \xrightarrow{\Phi} H_n(D_k) \oplus H_n(D') \xrightarrow{\Psi} H_n(C) \xrightarrow{\partial} H_{n-1}(D_k \cap D') \to \cdots$$

where $\Phi = i_* \oplus j_*$, $\Psi = h_* - l_*$ and ∂ is the connecting homomorphism given by the zig-zag lemma, see [18,19] for more details.

By Lemma 8, D_k contains at least one D_k-free 1-face, whence its geometric realization of D_k is a surface with boundary. As a result $H_2(D_k) \cong 0$ and $H_1(D_k)$ is free.

From the definition of ribbon follows, that the intersection $D_k \cap D_j$ of any two ribbons cannot contain any 2-simplices. Furthermore, $D_k \cap D_j$ does not contain any 1-simplices, whence $D_k \cap D_j$ contains only 0-simplices.

In view of $D_k \cap (\cup_{j<k} D_j) = \cup_{j<k}(D_k \cap D_j)$, we conclude that $D_k \cap D'$ consists of only 0-simplices. Therefore $H_2(D_k \cap D') \cong H_1(D_k \cap D') \cong 0$ and $H_0(D_k \cap D')$ is free.

In case of $n = 2$, the Mayer-Vietoris sequence reads

$$0 = H_2(D_k \cap D') \xrightarrow{\Phi} H_2(D_k) \oplus H_2(D') \xrightarrow{\Psi} H_2(C) \xrightarrow{\partial} H_1(D_k \cap D') = 0.$$

Thus Ψ is an isomorphism

$$H_2(C) \cong H_2(D_k) \oplus H_2(D').$$

By induction hypothesis, we have $H_2(D') \cong 0$ and due to the fact that D_k contains a free D_k-1-face, $H_2(D_k) \cong 0$. Accordingly we obtain $H_2(C) \cong 0$.

In case of $n = 1$, we observe

$$0 = H_1(D_k \cap D') \xrightarrow{\Phi} H_1(D_k) \oplus H_1(D') \xrightarrow{\Psi} H_1(C) \xrightarrow{\partial} H_0(D_k \cap D')$$

which gives rise to the following exact sequence

$$0 \xrightarrow{\Phi} H_1(D_k) \oplus H_1(D') \xrightarrow{\Psi} H_1(C) \xrightarrow{\partial} \operatorname{Im}\partial \longrightarrow 0.$$

As a subgroup of the free group $H_0(D_k \cap D')$, $\operatorname{Im}\partial$ is free and thus projective. As a result, the short exact sequence is split exact, and

$$H_1(C) \cong H_1(D_k) \oplus H_1(D') \oplus \operatorname{Im}\partial$$

holds. Since both $H_1(D_k)$ and $H_1(D')$ are free, we conclude that $H_1(C)$ is free, which completes the proof. □

Theorem 1 shows that, while the complete components contribute only to the second homology, the incomplete components provide generators of the first homology. In terms of discrete Morse theory, each complete component contains a critical point of dimension 2 and each incomplete component can feature multiple, critical points of dimension 1.

7. The main theorem

In this section, we compute the homology of a τ-structure. The key tool here is the Mayer-Vietoris sequence, which allows us to connect and compose the homology data of the sub-complexes.

A certain ordering by which the components are glued in combination with Lemma 6 are critical for the application of the Mayer-Vietoris sequence, since they constitute the determinants of how components intersect.

Theorem 2. *Let $\tilde{X}$ be the complex obtained from the loop complex of a τ-structure I. Then $H_1(\tilde{X})$ is free and*

$$H_2(\tilde{X}) \cong \oplus_{i=1}^{M} \mathbb{Z},$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

where M denotes the number of complete $\tilde{X}$-components. Furthermore, $H_2(\tilde{X})$ is freely generated by $\{c_i\}_i^M$, where c_i denotes the sum of 2-simplices contained in a complete component, C_i.

Proof. Let $\tilde{X}_2$ be the $\tilde{X}$-sub-complex induced by the $\tilde{X}$-2-simplices. Suppose that $\tilde{X}_2$ is partitioned into j components $C_1, \ldots, C_j$. The set $\{C_i\}$ and any of its subsets are partially ordered and recursively removing maximal components we can obtain a "descending" sequence $(C_{i_1}, \ldots, C_{i_j})$ such that

$$C_{i_h} \not< C_{i_k} \quad \text{for } h < k.$$

This sequence gives rise to a sequence of sub-complexes, $(Y_0, Y_1, \ldots, Y_j)$ obtained by recursively adding a component, as follows: $Y_0 = \varnothing$ and $Y_k = Y_{k-1} \cup C_{i_k}$. By construction, we have $Y_j = \tilde{X}_2 = \cup_{i=1}^{j} C_i$.

We next prove by induction on the number of components, k, that $H_1(Y_k)$ is free and $H_2(Y_k) \cong \oplus_{i=1}^{M_k} \mathbb{Z}$, where M_k denotes the number of complete components in $C_{i_1}, \ldots, C_{i_k}$.

For the induction basis $k = 1$, $Y_1 = C_{i_1}$ itself is a component and its homology has been computed in Theorem 1. Clearly, when C_{i_1} is complete, $H_2(Y_1)$ is freely generated by the sum of 2-simplices of C_{i_1}.

For the induction step we consider Y_k as the union of two sub-complexes C_{i_k} and Y_{k-1} and shall combine C_{i_k} and Y_{k-1} by means of the Mayer-Vietoris sequence:

$$\cdots \to H_n(C_{i_k} \cap Y_{k-1}) \to H_n(C_{i_k}) \oplus H_n(Y_{k-1}) \to H_n(Y_k) \to H_{n-1}(C_{i_k} \cap Y_{k-1}) \to \cdots$$

By construction $C_{i_h} \not< C_{i_k}$ for any $h < k$, which enables the application of Lemma 6. We accordingly conclude that the 1-skeleton of $C_{i_k} \cap C_{i_h}$ is contained in $\{\tau_1, \tau_l'\}$.

In view of $C_j \cap (\cup_{i<j} C_i) = \cup_{i<j}(C_i \cap C_j)$ for any i and j, we derive that the 1-skeleton of $C_{i_k} \cap Y_{k-1}$ is contained in $\{\tau_1, \tau_l'\}$. This severely constraints $C_{i_k} \cap Y_{k-1}$, consisting of 0-simplices and at most two 1-simplices.

As a result, $H_2(C_{i_k} \cap Y_{k-1}) \cong H_1(C_{i_k} \cap Y_{k-1}) \cong 0$ and $H_0(C_{i_k} \cap Y_{k-1})$ is free.

In case $n = 2$, we have

$$0 = H_2(C_{i_k} \cap Y_{k-1}) \xrightarrow{\Phi} H_2(C_{i_k}) \oplus H_2(Y_{k-1}) \xrightarrow{\Psi} H_2(Y_k) \xrightarrow{\partial} H_1(C_{i_k} \cap Y_{k-1}) = 0.$$

Thus Ψ is an isomorphism and $H_2(Y_k) \cong H_2(C_{i_k}) \oplus H_2(Y_{k-1})$. By induction hypothesis, $H_2(Y_{k-1}) \cong \oplus_{k=1}^{M_{k-1}} \mathbb{Z}$, where M_{k-1} denotes the number of complete components in $C_{i_1}, \ldots, C_{i_{k-1}}$. Due to Theorem 1, $H_2(C_{i_k}) \cong \mathbb{Z}$ in case of C_{i_k} being complete and $H_2(C_{i_k}) \cong 0$ in case of C_{i_k} being incomplete. Thus we can conclude by induction

$$H_2(Y_k) \cong \oplus_{i=1}^{M_k} \mathbb{Z}.$$

Clearly, $H_2(Y_k)$ is freely generated by $\{c_{i_h}\}_h^{M_k}$, where c_{i_h} is given by the sum of 2-simplices of the i_h-th complete component.

In case $n = 1$, we have

$$0 = H_1(C_{i_k} \cap Y_{k-1}) \xrightarrow{\Phi} H_1(C_{i_k}) \oplus H_1(Y_{k-1}) \xrightarrow{\Psi} H_1(Y_k) \xrightarrow{\partial} H_0(C_{i_k} \cap Y_{k-1})$$

and the following short exact sequence

$$0 \xrightarrow{\Phi} H_1(C_{i_k}) \oplus H_1(Y_{k-1}) \xrightarrow{\Psi} H_1(Y_k) \xrightarrow{\partial} \operatorname{Im} \partial \longrightarrow 0.$$

As a subgroup of the free group $H_0(C_{i_k} \cap Y_{k-1})$, $\operatorname{Im} \partial$ is free and thus projective. Then the short exact sequence is split exact, and

$$H_1(Y_k) \cong H_1(C_{i_k}) \oplus H_1(Y_{k-1}) \oplus \operatorname{Im} \partial.$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

Since both $H_1(C_{i_k})$ and $H_1(Y_{k-1})$ are free, we can conclude that $H_1(Y_k)$ is free.

Therefore it follows by induction that $H_1(\tilde{X}_2) = H_1(Y_j)$ is free and $H_2(\tilde{X}_2) = H_2(Y_j) \cong \oplus_{i=1}^{M}\mathbb{Z}$, where M denotes the number of complete components in $C_1, \ldots, C_j$.

In view of Proposition 9,

$$H_2(\tilde{X}) \cong H_2(\tilde{X}_2) \cong \oplus_{i=1}^{M}\mathbb{Z}$$

and $H_1(\tilde{X}) \cong H_1(\tilde{X}_2) \bigoplus \oplus_{i=1}^{k}\mathbb{Z}$ is free, completing the proof. □

The transitive closure with respect to ϕ-crossing produces an equivalence relation, i.e. $(i_1, i_2) \sim_\phi (j_1, j_2)$, if there exists a sequence of arcs

$$\alpha_1 = (i_1, i_2), \alpha_2, \ldots, \alpha_l = (j_1, j_2)$$

such that two consecutive arcs α_k and α_{k+1} are ϕ-crossing for $1 \leq k \leq l-1$.

An *arc-component*, A, is an $\sim_\phi$-equivalence class of arcs such that A contains at least two arcs. A is *of type* 1 if both endpoints of any of its arcs are incident to ϕ-arcs and *of type* 2, otherwise.

Suppose that $(i_1, j_1), (i_2, j_2), \ldots, (i_l, j_l)$, where $i_1 < i_2 < \cdots < i_l$ and $\phi(i_k) = j_k$ are ϕ-arcs associated with A. These partition the two backbones into blocks, the outer block $\beta_1 = [1, i_1)_S \cup (i_l, n]_S \cup [1, j_1)_T \cup (j_l, m]_T = (i_l, i_1)_S \cup (j_l, j_1)_T$ and the inner blocks $\beta_r = (i_{r-1}, i_r)_S \cup (j_{r-1}, j_r)_T, 2 \leq r \leq l$.

Lemma 9. *In I', let A be an arc-component of type 1 with blocks $\beta_1, \beta_2, \ldots, \beta_l$. Then, for any arc $(i, i') \notin A$, there exists a unique block, β_k, that contains i and i'.*

Proof. Without loss of generality, we may assume that (i, i') is an S-arc. Since each ϕ-arc is incident to at most one arc due to Lemma 1, (i, i') is not incident to any ϕ-arc associated with A, and thus its endpoints are contained in A-blocks.

By construction, any T-arc of A, (j_p, j_q), and (i, i') are non-crossing, since otherwise (i, i') would belong to the arc-component A. Accordingly, $\phi^{-1}(j_p)$ and $\phi^{-1}(j_q)$ are either contained in the interval $(i, i')_S$ or $[1, i)_S \cup (i', n]_S$ and we can conclude that A-arcs contained in either $Y_0 \subset (i, i')_S$ or $Y_1 \subset [1, i)_S \cup (i', n]_S$.

Suppose that i and i' are contained in different blocks, then Y_0 and Y_1 are nontrivial. By construction, any Y_0- and any Y_1-arc are non-crossing, which is impossible since A is an arc-component. As a result i and i' are contained in a single A-block. □

Remark: Lemma 9 is the "arc"-analogue of Lemma 3 for arc-components of type 1. Note that the statement does not hold for arc-components of type 2.

Given an arc-component A of type 1 in the interaction structure I', each ϕ-arc is associated with exactly one 2-simplex, as Lemma 1 guarantees that each ϕ-arc is incident to at most one arc. Let $\Delta_1, \Delta_2, \ldots, \Delta_l$ denote the 2-simplices associated with the ϕ-arcs $(i_1, j_1), (i_2, j_2), \ldots, (i_l, j_l)$, respectively. Let τ_k and τ'_k denote the mixed 1-faces of Δ_k that are associated with β_k and β_{k+1}, respectively.

Lemma 10. *Let A be an I'-arc-component of type 1 with blocks $\beta_1, \beta_2, \ldots, \beta_l$. Let $\Delta_1, \Delta_2, \ldots, \Delta_l$ denote the 2-simplices associated with A, and τ_k and τ'_k be the mixed 1-faces of Δ_k that are associated with β_k and β_{k+1}, respectively. Then $\tau'_k = \tau_{k+1}$ for $1 \leq k \leq l-1$ and $\tau'_l = \tau_1$.*

Proof. Let $(i_1, j_1), (i_2, j_2), \ldots, (i_l, j_l)$ denote the ϕ-arcs associated with A, ordered from left to right. It suffices to prove $\tau'_1 = \tau_2$. Set $\tau'_1 = \{s, t\}$, where s and t are S- and T-loops associated with β_2. Clearly, $i_1 \in s$ and $j_1 \in t$.

By Lemma 9, any I'-arc is either contained in β_2 or its complement. This guarantees that i_1 and i_2 belong to the same S-loop s, and furthermore that j_1 and j_2 belong to the same T-loop t.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

As a result the 2-simplex Δ_2 associated with (i_2, j_2) contains $\{s, t\}$ as a mixed face, i.e., $\tau_1' = \tau_2$. $\square$

Now we are in position to prove the Main Theorem.

Proof of the Main Theorem. The triviality of the third homology group, $H_3(I)$, follows from Corollary 1. In view of Lemma 1, Proposition 8 and Theorem 2, we have

$$H_2(I) \cong H_2(\tilde{X}) \cong \oplus_{k=1}^{M} \mathbb{Z},$$

where M denotes the number of complete $\tilde{X}$-components.

First we observe that any σ-butterfly removal does not change the crossing status of any two arcs. Thus there is a natural bijection between the set of I-arc-components and that of I'-arc-components. Moreover, each I-arc-component of type 1 corresponds to an I'-arc-component of type 1.

It suffices to show that when passing from the complex $X = K(I')$ to $\tilde{X}$, the number of I'-arc-components of type 1 equals the number M of complete $\tilde{X}$-components.

Claim: there exists a bijection between the set of I'-arc-components of type 1 and that of complete components of $\tilde{X}$.

Given an I'-arc-component of type 1, A, let C denote the set of couples (Δ, Δ') associated with arcs of A. In view of Lemma 10, any 1-simplex that appears as a face in C-couples is shared by at least two 2-simplices. As a result, C does not contain any free 1-simplex and passing from X to $\tilde{X}$, simplicial collapses do not affect C-2-simplices.

Consequently, all C-couples in I' simply persist when passing to $\tilde{X}$, where C itself becomes a $\tilde{X}$-component. Lemma 10 guarantees furthermore, that C does not contain any free 1-face, whence Lemma 8 implies that C is complete.

Accordingly, the Ansatz $\Phi(A) = C$ produces a well-defined mapping between I' arc-components of type 1 and complete $\tilde{X}$-components.

Φ is by construction injective since the mapping constitutes a mere reinterpretation of an arc-component of type 1.

To establish surjectivity, let C_1 be a complete $\tilde{X}$-component. C_1 induces a distinguished, unique set of I'-arcs, A_1. By construction, any two A_1-arcs $\alpha_1, \alpha_2 \in A_1$ satisfy $\alpha_1 \sim_\phi \alpha_2$, whence A_1 is contained in a nontrivial, distinguished arc-component A'.

We proceed by proving that $A_1 = A'$ is an arc-component of type 1. Suppose $A_1 \subsetneq A'$, then there exist an arc $(i, i') \in A' \setminus A_1$ and $(j, j') \in A_1$ such that (i, i') and (j, j') are ϕ-crossing. Note that (j, j') is associated to a C_1-couple $(\Delta_j, \Delta_{j'})$. Let $\{s, s'\}$ be the pure edge associated with (i, i'). We showed in the proof of Lemma 7, that S_{C_1} is connected.

Since (i, i') and (j, j') are ϕ-crossing, the S-loops s_1, s_2 are vertices of $\Delta_j, \Delta_{j'}$ from which follows that both s and s' are on a path of two S-loops in C_1. Accordingly $\{s, s'\}$ is contained in S_{C_1}. Lemma 7 guarantees that C_1 is a sphere, whence there exists a C_1-couple (Δ_0, Δ_0') for which $\{s, s'\}$ is a pure face. As a result we obtain $(i, i') \in A_1$, a contradiction from which $A' = A_1$ follows.

By Lemma 2, both endpoints of an A_1-arc are incident to ϕ-arcs, whence A_1 is an arc-component of type 1 and $\Phi(A_1) = C_1$, whence Φ is surjective and thus bijective.

Therefore $\gamma = M$ and $H_2(I) \cong \oplus_{k=1}^{\gamma} \mathbb{Z}$.

The Euler characteristic χ of $K(I)$ can be expressed as the alternating sum of ranks of its homology groups, i.e. $\chi = \text{rnk}(H_0(I)) - \text{rnk}(H_1(I)) + \text{rnk}(H_2(I)) - \text{rnk}(H_3(I))$, whence $\text{rnk}(H_1(I)) = \gamma - \chi + 1$. Theorem 2 shows that $H_1(I) \cong H_1(\tilde{X})$ is free from which we conclude $H_1(I) \cong \oplus_{k=1}^{\gamma-\chi+1} \mathbb{Z}$. $\square$

8. Discussion

In this paper we computed the simplicial homology of τ-structures. τ-structures represent a meaningful generalization of bi-structures [7,8] as they allow to study transitions between sequence-structure pairs, where the underlying sequences differ in specific ways. Bi-structures would only facilitate the analysis of such pairs, where the underlying

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

sequences are equal. Intuitively, the free energy of such sequence-structure pairs, a quantity that is straightforward to compute, characterizes the probability of finding such a pair of sequences in the course of evolution.

We are currently extending the results of this paper by computing the weighted homology of τ-structures over a discrete valuation ring, R. This means to study weighted complexes in which simplices are endowed with weights [20–22]. These weighted complexes feature a new boundary map, $\partial_n^v : C_{n,R}(X) \to C_{n-1,R}(X)$, where $C_{n,R}(X)$ denotes the free R-module generated by all n-simplices contained in X. ∂_n^v is given by

$$\partial_n^v(\sigma) = \sum_{i=0}^{n} \frac{v(\hat{\sigma}_i)}{v(\sigma)} \cdot (-1)^i \hat{\sigma}_i.$$

As it is the case for bi-structures [22], the weighted homology plays a crucial role in the Boltzmann sampling of sequence-structure pairs that minimize the free energy of the τ-structure.

Our approach differs from the purely algebraic proofs for the simplicial loop homology of bi-structures [8]. In the case of bi-structures, the computation for the second homology employs the fact that their first homology group is trivial. This allows to understand the second homology group via a long exact sequence of relative homology groups. However, for τ-structures the first homology is in general nontrivial and therefore requires a different approach.

We first reduce the loop complex $K(I)$ to the sub-complex $\tilde{X}$ via simplicial collapses, retaining the homology of the original space. We then dissect a certain 1-dimensional sub-complex in Proposition 9 and then decompose $\tilde{X}$ into components based on the ϕ-crossing of couples. This decomposition allows us to identify different generators of the homology, with incomplete and complete components contributing to the first and second homologies, respectively. Finally, we compute the homology of $\tilde{X}$ by gluing components in a particular way. This makes use of the planarity of τ-structures in Lemma 6 and assures that we encounter particularly simple intersections, when applying the Mayer-Vietoris sequence. It is worth pointing out that in the proof of Theorem 2 the particular ordering, in which the components are glued is crucial.

In view of Proposition 9, Theorem 1 and Theorem 2, the generators of the first homology of the loop complex originate from the combining of the sub-complex K_1, incomplete components and the gluing of different components. The detailed description of all these generators is work in progress and here we restrict ourselves to using the Euler characteristic to express the rank of the first homology group. Only complete crossing components contribute to the second homology and their geometric realizations are spheres. We give a combinatorial characterization of the generators in I in terms of arc-components of type 1.

As for applications of this framework, we currently employ τ-structures to investigate the evolutionary trajectories of viruses, such as flu and coronavirus. Specifically, we compute the loop homology of evolutionary transitions to gain deeper insight into sequence-structure-function relationships of the virus.

Author Contributions: All authors have contributed equally to this work. All authors have read and agreed to the possible publication of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We gratefully acknowledge the comments and discussions from Andrei Bura, Qijun He and Fenix Huang.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Waterman, M. Secondary Structure of Single-Stranded Nucleic Acids. Studies on foundations and combinatorics, Advances in mathematics supplementary studies; Rota, G.C., Ed. Academic Press N.Y., 1978, Vol. 1, pp. 167–212.
2. Waterman, M. Combinatorics of RNA Hairpins and Cloverleaves. *Stud. Appl. Math.* **1979**, *60*, 91–98.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2021 doi:10.20944/preprints202107.0191.v1

3. Smith, T.F.; Waterman, M.S. RNA secondary structure. *Math. Biol.* **1978**, *42*, 31–49.
4. Howell, J.; Smith, T.; Waterman, M. Computation of Generating Functions for Biological Molecules. *SIAM J. Appl. Math.* **1980**, *39*, 119–133.
5. Schmitt, W.; Waterman, M. Linear trees and RNA secondary structure. *Disc. Appl. Math.* **1994**, *51*, 317–323.
6. Penner, R.; Waterman, M. Spaces of RNA secondary structures. *Adv. Math.* **1993**, *217*, 31–49.
7. Bura, A.C.; He, Q.; Reidys, C.M. Loop homology of bi-secondary structures. *Discrete Mathematics* **2021**, *344*, 112371.
8. Bura, A.C.; He, Q.; Reidys, C.M. Loop Homology of Bi-secondary Structures II. *arXiv preprint arXiv:1909.01222* **2019**.
9. Huang, F.W.D.; Barrett, C.; Reidys, C.M. The energy-spectrum of bicompatible sequences. *arXiv:1910.00190v1* **2021**.
10. Forman, R. A user's guide to discrete Morse theory. *ÉMINAIRE LOTHARINGIEN DE COMBINATOIRE* **2002**, p. B48c.
11. Forman, R. Morse Theory for Cell Complexes. *Advances in Mathematics* **1998**, *134*, 90–145.
12. Whitehead, J.H.C. Simplicial Spaces, Nuclei and m-Groups. *Proceedings of the London Mathematical Society* **1939**, *s2*, 243–327.
13. Cohen, M.M. *A Course in Simple-Homotopy Theory*; Graduate Texts in Mathematics, Springer-Verlag: New York, 1973.
14. Mayer, W. Über abstrakte Topologie. *Monatshefte für Mathematik und Physik* **1929**, *36*, 1–42.
15. Vietoris, L. Über die Homologiegruppen der Vereinigung zweier Komplexe. *Monatshefte für Mathematik und Physik* **1930**, *37*, 159–162.
16. Jordan, C. *Cours d'analyse de l'École polytechnique*; Gauthier-Villars: Paris, 1893.
17. Thomassen, C. The Jordan-Schönflies theorem and the classification of surfaces. *Amer. Math. Monthly* **1992**, *99*, 116–130.
18. Fulton, W. *Algebraic Topology: A First Course*; Graduate Texts in Mathematics, Springer-Verlag: New York, 1995.
19. Hatcher, A. *Algebraic Topology*; Cambridge University Press, 2005.
20. Dawson, R.J.M. Homology of weighted simplicial complexes. *Cahiers de Topologie et Géométrie Différentielle Catégoriques* **1990**, *31*, 229–243. Publisher: Dunodéditeur, publié avec le concours du CNRS.
21. Ren, S.; Wu, C.; Wu, J. Weighted persistent homology. *Rocky Mountain Journal of Mathematics* **2018**, *48*, 2661–2687.
22. Bura, A.; He, Q.; Reidys, C. Weighted Homology of Bi-Structures over Certain Discrete Valuation Rings. *Mathematics* **2021**, *9*, 744.