Loop homology of bi-secondary structures II

Andrei C. Bura · Qijun He · Christian M. Reidys

Received: 3 December 2020 / Accepted: 20 March 2022 / Published online: 27 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In this paper, we analyze the homology of the simplicial complex induced by a given pair of RNA secondary structures, \( R = (S, T) \). Such a pair induces a bi-secondary structure, whose associated loop nerve \( X \) is the simplicial complex obtained by loop intersections. We will provide an algebraic proof of the fact that \( H_1(X) = 0 \). We will provide a combinatorial interpretation for the generators of \( H_2(X) \) in terms of crossing components of the bi-structure and establish that the rank of \( H_2(X) \) equals the total number of such crossing components. Finally, we shall prove that each crossing component naturally encodes a triangulation of a 2-sphere and provide an analysis of the geometric realization of \( X \).

Keywords RNA · Bi-secondary structure · Loop · Nerve · Simplicial homology · Bijection

Mathematics Subject Classification 05E45 · 55U10

1 Introduction

An RNA secondary structure encodes the base pairing interactions of the folded structure of an RNA molecule [17] and can be represented as a planar arc diagram [12, 30]. Vertices in the diagram correspond to nucleotides on the RNA backbone, while arcs correspond to pairings between nucleotides. Secondary structures, viewed as abstract
diagrams or trees, have been studied in enumerative combinatorics [15, 16, 27, 31], algebraic combinatorics [20], matrix models [2, 23] and topology [3, 7, 18].

In [27], explicit formulae for the number of secondary structures on \( n \) vertices having exactly \( k \) arcs were derived, based on a beautiful bijection between secondary structures and linear trees. The number of \( k \)-non-crossing RNA structures was computed in [20]. This enumeration was derived from the bijection between \( k \)-non-crossing partial matchings and walks in \( \mathbb{Z}^{k-1} \), which remain in the interior of the Weyl chamber \( C_0 \) [10]. The bijection between oscillating tableaux and matchings originated with Stanley [28] and was generalized by Sundaram [29].

Harer and Zagier computed the generating function of the number of linear chord diagrams of genus \( g \) with \( n \) chords, in the course of computing the Euler characteristic of the Moduli space of a curve [14]. Based on this line of work, the number of chord diagrams of fixed genus with specified numbers of backbones and chords was computed in [2]. The number of secondary structures of a given genus, as well as the number of cells in Riemann’s moduli spaces for bordered surfaces, was derived from such an enumeration, using the Hermitian matrix model techniques and topological recursions along the lines of [1].

Penner and Waterman studied the space of RNA secondary structures from a topological perspective. They showed that the geometrical realization of the associated complex of secondary structures is a sphere [24]. A topological classification of secondary structures was given in [7], based on matrix models. Huang et al employed an augmented version of the topological recursion on unicellular maps of Chapuy [9] to derive explicit expressions for the coefficients of the generating polynomial of topological shapes of RNA structures and the generating function of RNA structures of genus \( g \) [18]. This lead to uniform sampling algorithms for structures of fixed topological genus as well as a natural way to resolve crossings in pseudoknotted structures [19].

When an RNA secondary structure is interpreted as an orientable fatgraph [19, 26], the structure has a loop decomposition, where a loop consists of vertices (nucleotides) that are visited subsequently when a boundary component of the fatgraph is traversed. The loop decomposition of secondary structures facilitates polynomial time minimum free energy folding algorithms [22, 32, 34] and Boltzmann sequence sampling algorithms [4, 5, 21]. The efficiency of these algorithms is rooted in the fact that any two loops intersect either trivially or in exactly two vertices. As a result, the nerve [13] of the loops of a secondary structure is a tree.

The notion of bi-secondary structures was introduced by Haslinger and Stadler to classify RNA pseudoknots [15]. Bi-secondary structures also emerge naturally in the context of evolutionary transitions and RNA riboswitch design, since they are closely connected to RNA sequences that are simultaneously energetically compatible with two different secondary structures [25].

A central question in biology is the characterization of RNA structural changes occurring in evolution and regulatory processes. In practice, emphasis is placed on overall structural changes rather than local ones at the base-pair level. A framework is needed that can qualitatively capture such non-local shape changes. As such, in [8], we studied the loop complex (loop nerve) \( K(R) \) obtained from loop intersections for a bi-secondary structure \( R \). We proved that \( H_2(R) \), the second homology group
of the geometrical realization $|K(R)|$, is free abelian. Remarkably, we found that all known riboswitch structure pairs in the SwiSpot database [6] exhibit $H_2(R)$ rank of 1. A structural interpretation of this rank would thus be valuable. In the following, we will identify the precise substructures of a given bi-secondary structure $R$, that, when considered within the loop nerve $K(R)$, correspond to subcomplexes that are triangulations of 2-spheres. These substructures are in fact the crossing components (CCs) of the bi-structure. We will show that there is a bijective correspondence between a minimal-generating set of $H_2(R)$ and the set of CCs of $R$, and thus, the rank of $H_2(R)$ equals the number of CCs in $R$.

2 Secondary and bi-secondary structures

In this paper, we use notation and nomenclature established in [8]. For a loop $s$ of a secondary structure (RNA diagram) $S$ over nucleotide set $[n]$, its maximal arc is denoted by $\alpha(s)$ with its beginning and end points denoted by $b(\alpha_s)$ and $e(\alpha_s)$, respectively. For any secondary structure $S$, there is a bijective correspondence between loops and their maximal arcs. Furthermore, any non-rainbow arc appears in exactly two loops and any unpaired nucleotide is contained in exactly one loop. For $s_1, s_2 \in S$, we write $s_1 \prec_s s_2$ if $b(\alpha_{s_2}) \leq b(\alpha_{s_1}) \leq e(\alpha_{s_1}) \leq e(\alpha_{s_2})$. The Hasse diagram of $(S, \prec_s)$ is a tree $Tr(S)$, rooted at the rainbow loop. A bi-structure is a pair of secondary structures $R = (S, T)$ over $[n]$, see Fig. 1.

3 The loop complex of a bi-structure

Following [8], for $R = \{r_0, r_1, \ldots, r_m\}$ the collection of loops (as sets of nucleotides) in $R = (S, T)$, we let $\prec_R$, the simplicial order on $R$, be a linearization of the extension

---

**Fig. 1** A reference bi-structure $R = (S, T)$ on nucleotide set [11], with formal vertices 0, 12 corresponding to the rainbow arcs. Arcs in the same secondary structure are non-crossing, while two arcs $(i, j) \in S, (p, q) \in T$ are crossing if $i < p < j < q$ or $p < i < q < j$. We have $(3, 5) \prec_s (1, 7) = \alpha_s$ and $(4, 6) \prec_T (2, 8)$.
Fig. 2 LHS: the reference bi-structure $R = (S, T)$. RHS: the complex $X$ associated to $R$

Fig. 3 LHS: case 1. RHS: case 2

of $<_S$ and $<_T$ via $r_1 <_R r_2$ for any $r_1 \in S, r_2 \in T$. The complex (nerve) of $R$ is $X = \bigcup_{d=0}^{\infty} K_d(R)$, where $K_d(R) = \{\sigma = [r_{i_0}, \ldots, r_{i_d}] | 0 \neq \omega(\sigma) := | \cap_{k=0}^d r_{i_k} \cap_{i \neq k} r_{i_i} \in R\}$, see Fig. 2. We call $\tau = [r_{i_0}, r_{i_1}] \in K_1$ pure if $r_{i_0}$ and $r_{i_1}$ are loops in the same secondary structure and mixed, otherwise. By construction, any $\Delta \in K_2(R)$ contains exactly one pure edge and two mixed edges. Furthermore, for any $\sigma \in K_3(R)$, $1 \leq \omega(\sigma) \leq 2$ and $K_{\geq 4}(R) = \emptyset$. In particular, for any secondary structure $S$, it is not difficult to show that $K_1(S) \cong Tr(S)$.

4 Simple bi-structures

For fixed $p \in P = \{p \in \{1, \ldots, n\} | \deg(p) = 4\}$, we will investigate the effect on $X$, of splicing $p$ into two adjacent nucleotides $q_1, q_2$, such that the two arcs incident to $p$ are resolved into two non-crossing arcs, having endpoints $q_1$ and $q_2$. In case $\omega(\sigma) = 2$, splitting does not change $X$, and if $\omega(\sigma) = 1$, splitting induces a specific alteration: the removal of $\sigma$ from $X$ as well as a distinguished free edge, $\tau_\sigma$, and exactly two free triangles $\Delta_1, \Delta_2$ glued at $\tau_\sigma$. We refer to $(\tau_\sigma, \Delta_1, \Delta_2)$ as a $\sigma$-butterfly. We depict the cases $p = e(\alpha_{s_0}) < b(\alpha_{t_0})$ and $p = b(\alpha_{s_0}) = b(\alpha_{t_0})$, in Fig. 3. A bi-structure is called simple, if $P = \emptyset$. By construction, simplicity implies $K_{\geq 3}(R) = \emptyset$. We formalize this in Lemma 1.

Lemma 1 For a bi-structure $R = (S, T)$ with complex $X$, there exists a simple bi-structure $R' = (S', T')$ having complex $X' < X$ where $X'$ is derived from $X$ by successively removing any 3-simplex, $\sigma$, together with an associated $\sigma$-butterfly.

Proof It suffices to prove the cases in Fig. 3, all others following analogously.

Case 1: $b(\alpha_{s_0}) < e(\alpha_{s_0}) = b(\alpha_{t_0}) < e(\alpha_{t_0})$.

Splicing $p$, we obtain $b(\alpha_{s_0}) < e(\alpha_{s_0}) < b(\alpha_{t_0}) < e(\alpha_{t_0})$ with the new loops $s_0 = (s_0 \setminus \{p\}) \cup \{q_1\}$, $t_0 = (t_0 \setminus \{p\}) \cup \{q_2\}$, $s_1 = (s_1 \setminus \{p\}) \cup \{q_1, q_2\}$ and $t_1 = (t_1 \setminus \{p\}) \cup \{q_1, q_2\}$.
\(\vec{t} = (t_1 \setminus \{p\}) \cup \{q_1, q_2\}\), see Fig. 3 LHS. This results in \(\vec{R}\) with associated \(\vec{X}\) given by the embedding

\[
\epsilon : \vec{X} \longrightarrow X, \quad \epsilon(\cdots, \vec{t}, \cdots, \vec{r}', \cdots) = [\cdots, r, \cdots, r', \cdots],
\]

where \(r \in \{s_0, s_1, t_0, t_1\}\) and \(r' \notin \{s_0, s_1, t_0, t_1\}\).

By construction, no \(X\)-simplices that have \(\tau^\sigma = [s_0, t_0]\) as a face have a preimage under \(\epsilon\) and any \(X\)-simplex that does not have a preimage under \(\epsilon\) must contain \(\tau^\sigma\) as a face. As \(\tau^\sigma\) is free so are \(\Delta^\sigma_1 = [s_0, t_0, t_1]\) and \(\Delta^\sigma_2 = [s_0, s_1, t_0]\). Therefore, \(\vec{X}\) is obtained from \(X\) by removing \(\sigma\) and \((\tau^\sigma, \Delta^\sigma_1, \Delta^\sigma_2)\).

\[\text{Case 2: } b(\alpha_{s_0}) = b(\alpha_{t_0}) < e(\alpha_{s_0}) < e(\alpha_{t_0}).\]

Splicing \(p\) produces \(b(\alpha_{s_0}) < b(\alpha_{t_0}) < e(\alpha_{s_0}) < e(\alpha_{t_0})\) with the new loops \(\vec{s_0} = (s_0 \setminus \{p\}) \cup \{q_2\}, \vec{t_0} = (t_0 \setminus \{p\}) \cup \{q_1, q_2\}, \vec{r_1} = (r_1 \setminus \{p\}) \cup \{q_1\}\), and finally \(\vec{r} = (t_1 \setminus \{p\}) \cup \{q_1\}\), see Fig. 3 RHS. The argument proceeds analogously as in Case 1 but with \(\tau^\sigma = [s_0, t_1]\), \(\Delta^\sigma_1 = [s_0, s_1, t_1]\) and \(\Delta^\sigma_2 = [s_0, t_0, t_1]\).

Thus, successively splicing all \(p \in P\) induces a bi-structure \(R'\) with associated complex \(X' < X\) obtained by removing any \(\sigma \in K_3(R)\), together with an associated \(\sigma\)-butterfly.

Note that the complexes \(X'\) can depend on a choice of \(\sigma\)-butterfly for some given \(\sigma\). By abuse of notation, we simply write \(X'\), with a choice implied.

Removing a \(\sigma\)-butterfly and its corresponding \(\sigma\) is tantamount to \(X \setminus^\sigma X'\), a simplicial collapse consisting of two elementary collapses [11, 33]: the first removing \(\sigma\) and \(\Delta^\sigma_1\) and the second \(\tau^\sigma\) and \(\Delta^\sigma_2\). This immediately yields

**Proposition 1** Let \(X\) be the loop complex of a bi-structure \(R = (S, T)\) and let \(X' < X\) be an \(X\)-sub complex obtained by removing from \(X\) all 3-simplices, \(\sigma\), together with corresponding \(\sigma\)-butterflies. Then, \(X'\) is the complex of a bi-structure \(R' = (S', T')\) that satisfies \((X')^2 = X'\) and for any \(k \geq 0\), \(H_k(X') \cong H_k(X)\).

Note that in a simple bi-structure, an endpoint of a non-rainbow arc corresponds to a unique 2-simplex in the loop complex and any 2-simplex corresponds to at least one of these endpoints. Consider the multi-set \((\Delta_i)_{i \in M}\) where \(M\) is the set of endpoints of all non-rainbow arcs in \(R\) and the \(\Delta_i\)'s are their corresponding 2-simplices. The \(\Delta_i\)'s can then be arranged, from left to right on the backbone, with their pure \(S\)- or \(T\)-face top or bottom facing, respectively, see Fig. 4.

For each \(\Delta_i\), we assign a *parity* \(\pi(\Delta_i)\) as follows: if when traversing counter clockwise the boundary of a \(\Delta_i\) we match \(\prec_R\) on its pure edge, then \(\pi(\Delta_i) = 1\), and otherwise \(\pi(\Delta_i) = -1\). The following is immediate

**Lemma 2** Let \(\Delta_i\) and \(\Delta_j\) be a pair of 2-simplices that share a pure edge \(e\). Then, the coefficient of \(e\) in \(\partial_2(\pi(\Delta_i)\Delta_i + \pi(\Delta_j)\Delta_j)\) equals 0. Furthermore, for any \(\Delta_i\) and \(\Delta_j\) that are adjacent on the backbone, the coefficient of their shared mixed edge \(e'\) in \(\partial_2(\pi(\Delta_i)\Delta_i + \pi(\Delta_j)\Delta_j)\) equals 0.
5 The simplicial homology

Clearly $H_0(X) = \mathbb{Z}$. Furthermore, $H_1(X) = 0$ has been established in [8] via a complicated combinatorial construction. Here we present a shorter, more algebraic proof of this fact.

**Proposition 2** Let $R = (S, T)$ be a bi-structure with complex $X$. Then,

$$H_1(X) = 0.$$  

**Proof** By Proposition 1 w.l.o.g. we may assume $X^2 = X$. We shall build $X$ by inductively adding $T$-loops, $t_i$, via inserting maximal arcs. By construction, this insertion process adds new simplices step by step, not affecting previously existing gluings. As a result, we obtain a sequence of complexes $X_0, X_1, \ldots, X_n = X$, where $X_0 = K(S)$ and $X_{i+1}$ is obtained from $X_i$ by adding some maximal loop $t_i$. At the final step, $X_n = X$ is obtained by adding the $T$-rainbow. Note that for $i < n$ none of the $X_i$ are complexes of bi-structures. We proceed by proving by induction on $m \leq n$, that $H_1(X_m) = 0$ for $1 \leq m \leq n$. For the induction basis $X_0 = K(S)$ is a tree, whence $H_1(X_0) = 0$. By construction, $X_m \prec X_{m+1}$ and we have the short exact sequence of chain complexes

$$0 \longrightarrow C_n(X_m) \overset{i}{\longrightarrow} C_n(X_{m+1}) \overset{j}{\longrightarrow} C_n(X_{m+1}, X_m) \longrightarrow 0$$

which induces the long exact sequence

$$0 = H_1(X_m) \overset{i_*}{\longrightarrow} H_1(X_{m+1}) \overset{j_*}{\longrightarrow} H_1(X_{m+1}, X_m) \overset{\delta_1}{\longrightarrow} H_0(X_m) \overset{i_*}{\longrightarrow} H_0(X_{m+1})$$
Fig. 5 LHS: the reference bi-structure $R = (S, T)$. RHS: $G(R)$ has two non-trivial (size two or higher) connected components and three trivial ones (size one), each corresponding to an IC of $R$, the former to CCs and the latter to trivial ICs

where by induction hypothesis, $H_1(X_m) = 0$. It suffices then to show that $H_1(X_{m+1}, X_m) = \text{Ker}(\delta_1)/\text{Im}(\delta_2) = 0$, where

$$C_2(X_{m+1}, X_m) \xrightarrow{\delta_2} C_1(X_{m+1}, X_m) \xrightarrow{\partial_1} C_0(X_{m+1}, X_m).$$

Note that $C_1(X_{m+1}, X_m) = \langle [a, t] | a \in X^0_m, a \cap t \neq \emptyset \rangle$. Furthermore, $\text{Ker}(\delta_1) = \langle [a, t] - [b, t] | a \prec_R b; [a, t], [b, t] \in C_1(X_{m+1}, X_m) \rangle$. It thus suffices to prove $[a, t] - [b, t] = -\delta_2([a, b, t])$. Otherwise $a \cap b \cap t = \emptyset$, and we distinguish nucleotides $u \in a \cap t$ and $v \in b \cap t$ where w.l.o.g. $u < v$. Then, there exist $u'$ and $v'$, $u \leq u' < v' \leq v$, endpoints of $X_{m+1}$-arcs, such that $[a, t]$ is an edge of a leftmost triangle $\Delta_{u'}$, and $[b, t]$ is an edge of a rightmost triangle $\Delta_{v'}$, of a block of subsequently glued triangles $\Delta_{u' \leq i \leq v'}$ on the backbone (via their mixed edges), which is of minimal length. By Lemma 2, it is then not difficult to see that

$$[a, t] - [b, t] = \delta_2 \left( \sum_{u' \leq i \leq v'} \pi(\Delta_i) \Delta_i \right).$$

As such, $H_1(X_{m+1}, X_m) = 0$ and in view of the long exact sequence, the proposition follows by induction.

Let $R = (S, T)$ and let $G(R)$ denote the graph induced by the sets of $S$- and $T$-arcs in which two arcs form an edge if they are crossing. A set of arcs, $A$, associated with a $G(R)$-component is called an irreducible component (IC) of $R$. In case of $|A| > 1$, we refer to $A$ as a crossing component (CC). We also identify $A$ with the set of loops whose maximal arcs are in $A$. An IC consists of either a single arc or it is non-trivial and hence a CC. The ICs partition the arcs of $R$, see Fig. 5. Finally, note that splicing does not affect CCs.

We next show that the $H_2(X)$-generators are in bijection with non-trivial $G(R)$ components. This can simplify the computation of the $\text{rnk}(H_2(X))$ invariant [8], which in view of the theorem below will simply reduce to a depth-first search algorithm that
can count the connected components of $G(R)$, in linear time in the number of loops of $R$.

**Theorem 1** (Main Theorem) *Let $R = (S, T)$ be a bi-structure with complex $X$; then, the following assertions hold:*

(a) $H_2(X)$ is freely generated by sums of 2-simplices where each such sum corresponds uniquely to an $(S, T)$-crossing component

(b) $H_3(X) = 0$ and $H_2(X^2) \cong \hat{\delta}_3(C_3(X)) \oplus H_2(X)$.

**Proof** *Claim 1. $H_2(X)$ is free.*

Consider the short exact sequence of chain complexes

$$
0 \longrightarrow C_n(X^1) \xrightarrow{I} C_n(X) \xrightarrow{J} C_n(X, X^1) \longrightarrow 0
$$

which induces the long exact sequence

$$
H_2(X^1) \xrightarrow{I_*} H_2(X) \xrightarrow{J_*} H_2(X, X^1) \xrightarrow{\hat{\delta}_2} H_1(X^1) \xrightarrow{I_*} H_1(X) \longrightarrow H_1(X, X^1).
$$

Since $H_2(X^1) = 0$, $J_*$ is an embedding into $H_2(X, X^1) = C_2(X)$ which is free.

*Claim 2. $H_2(X)$ is freely generated by elements corresponding to CC’s.*

W.l.o.g. we assume $X^2 = X$ and $R$ simple. $\text{Im}(J_*) = \text{Ker}(\hat{\delta}_2)$ follows from

$$
\text{H}_2(X^1) \xrightarrow{I_*} \text{H}_2(X) \xrightarrow{J_*} \text{H}_2(X, X^1) \xrightarrow{\hat{\delta}_2} \text{H}_1(X^1) \xrightarrow{I_*} \text{H}_1(X) = 0
$$

(2)

We proceed by identifying particular generators of $\text{Ker}(\hat{\delta}_2)$. For any $M' \subset M$, a subset of arc end-points, call $(\Delta_j)_{j \in M'}$ a $(\Delta_j)_{j \in M}$-subsequence, and call it *closed* iff the left 2-face of $\Delta_{\min(M')}^i$ and the right 2-face of $\Delta_{\max(M')}^j$ are the same in $X_1(X)$. Furthermore, call such a closed subsequence a *block* if it does not contain any closed subsequences, and call a $M'$-block *trivial* iff $|M'| = 2$.

We process $(\Delta_i)_{i \in M}$ recursively into a tree of blocks, $\mathcal{T}_R$ as follows: Firstly, let $U = \{(\Delta_i)_{i \in M}\}$ and $P = \emptyset$. While $U \neq \emptyset$ execute the following

- Fix $\omega \in U$.
- If $\omega$ is a block, update $U := U \setminus \{\omega\}$, $P := P \cup \{\omega\}$.
- Otherwise, fix $\omega' < \omega$ a proper, closed subsequence of $\omega$ of maximal length and update $U := U \setminus \{\omega\} \cup \{\omega', \omega \setminus \omega'\}$.

This algorithm is well defined: at each step $U$ consists of closed subsequences. This is since $(\Delta_i)_{i \in M}$ is closed and, if both $\omega$ and its subsequence $\omega'$ are closed, so is $\omega \setminus \omega'$. Also, $|M|$ is finite; hence, the algorithm eventually terminates. At that point, by construction $P$ will be the set of blocks in $R$. The $\mathcal{T}_R$ construction starts with
a formal root \( \beta_0 \) corresponding to \((\Delta_i)_{i \in M'} \), where an edge from \((\Delta_i)_{i \in M''} \in P\) to \((\Delta_i)_{i \in M''} \in P\) is added whenever \([\min(M''), \max(M'')] \] is maximally included in \([\min(M'), \max(M')] \setminus M'\), see Fig. 6.

Let \( \beta \) be a non-trivial block and \( \Delta_i \in \beta \). Then, by definition, there does not exist \( \Delta_j \in \beta \) with \( \Delta_j = \Delta_i \) and \( i \neq j \), whence \( \{\nabla_\beta = \sum_{\Delta_i \in \beta} \pi(\Delta_i)\Delta_i \mid \beta \in T_R, \beta \neq \beta_0\} \) are free in \( H_2(X, X^1) \).

Any pure \( S \)- and \( T \)-faces appearing in \( \beta \) occur as complementary pairs w.r.t. the orientation induced by \( \prec_R \) and so by Lemma 2 \( \hat{\partial}_2(\nabla_\beta) = 0 \), i.e.,

\[
\langle \{\nabla_\beta \mid \beta \in T_R, \beta \neq \beta_0\} \rangle \subseteq \text{Ker}(\hat{\partial}_2).
\]

For the converse: note that \( H_1(X^1) \) is free and hence projective as a \( \mathbb{Z} \) module. Thus, sequence (2) is split exact, i.e., \( H_2(X, X^1) \cong H_2(X) \oplus H_1(X^1) \). Let \( f, e, v \) denote the numbers of 2- 1- and 0-simplices, respectively. As \( H_2(X, X^1) = C_2(X) \) and \( H_2(X) \) is free,

\[
\text{rnk}(H_2(X)) = \text{rnk}(C_2(X)) - \text{rnk}\left(H_1(X^1)\right), \tag{3}
\]

where, since \( X^1 \) is a connected graph, \( \text{rnk}(H_1(X^1)) = (e - v + 1) \).

Removing a non-crossing arc from \( R \) and passing to \( R^* \) complex \( X^* \) has the effect of removing one 2-simplex, two 1-faces and one 0-simplex, while potentially relabeling some of the remaining vertices. Hence, \((e - v + 1) \) changes to \((e - 2) - (v - 1) + 1 = (e - v + 1) - 1 \), i.e.,

\[
\text{rnk}(C_2(X)) - \text{rnk}\left(H_1(X^1)\right) = \text{rnk}(C_2(X^*)) - \text{rnk}\left(H_1(X^*)\right).
\]

Accordingly, removing all non-crossing arcs from \( R \) does not affect \( \text{rnk}(H_2(X)) \) and so, w.l.o.g. we can assume that \( R \) does not contain any non-crossing arcs, i.e., the \( R \)-arcs partition into CCs. We can process \( T_R \) from left to right and leaves to root, by recursively extracting blocks \( \beta \).

Note that in general each block \( \beta = (\Delta_j)_{j \in M'} \) corresponds uniquely to the IC’s of a bi-structure \( R \) as follows: firstly, any arcs corresponding to a non-trivial block \( \beta \) must be crossing, otherwise a trivial block corresponding to a non-crossing arc would be a proper subsequence of \( \beta \). Secondly, any arc corresponding to \( \beta \) has both its endpoints
in $M'$, otherwise the left edge of $\Delta_{\min(M')}^\prime$ and the right edge of $\Delta_{\max(M')}^\prime$ would differ in a 0-face.

As top and bottom faces of a block represent a $S$- and $T$-substructure, respectively, by inserting formal rainbow arcs as needed, each block $\beta$ induces a sub-complex $X_\beta < X$ corresponding to a bi-structure $R_\beta^*$ consisting of the single unique CC determined by $\beta$ and the inserted rainbows, see Fig. 7.

As we assume that $R$ has no trivial arcs, any $X_\beta$ exhibits one additional edge and two additional vertices and $e_{X_\beta} = e_\beta + 1$, $v_{X_\beta} = v_\beta + 2$, where $v_\beta$ denotes the number of vertices, i.e., loops corresponding to arcs, whose endpoints generate $\beta$ and $e_\beta$ is the number of distinct edges with at least one $v_\beta$-vertex in $\beta$. Then, $e_{X_\beta} - v_{X_\beta} + 1 = e_\beta - v_\beta$, and we proceed recursively until $T_R$ is processed to the $\beta_0$-root. With each extraction, the remaining complex has $e_\beta$ fewer edges and $v_\beta$ fewer vertices. In view of $e - v + 1 = \sum_\beta (e_\beta - v_\beta) + 1$,

$$\sum_\beta \rk(H_1(X_\beta^1)) = \rk(H_1(X^1)),$$

Since 2-simplices are not affected in the construction of the $X_\beta$, we have

$$\rk(C_2(X)) - \rk(H_1(X^1)) = \sum_{\beta \in T_R, \beta \neq \beta_0} \rk(C_2(X_\beta)) - \rk(H_1(X_\beta^1)),$$

whence

$$\rk(H_2(X)) = \sum_{\beta \in T_R, \beta \neq \beta_0} \rk(H_2(X_\beta)).$$

We next compute $\rk(H_2(X_\beta))$. Suppose $|P = \{\Delta \mid \Delta \in \beta\}| = 2k$. These 2-simplices have $6k$ 1-faces, each of which appears twice. The set $P$ is generated by the endpoints of $k$ mutually crossing arcs together with two minimal $S$ and $T$ arcs, respectively, covering them. Replacing the two minimal $S$ and $T$ arcs by rainbows, we see that $P$ has the complex $X_\beta$ and by sequence (2) we have

$\textcircled{3}$ Springer
\[
\begin{align*}
0 & \longrightarrow H_2(X_\beta) \longrightarrow H_2 \left( X_\beta, X_\beta \right) \longrightarrow H_1 \left( X_\beta^1 \right) \longrightarrow 0 \quad (4)
\end{align*}
\]

and

\[
\text{r}n\text{k} \left( H_2(X_\beta) \right) = |\{ \Delta \mid \Delta \in \beta \}| - \text{r}n\text{k} \left( H_1(X_\beta^1) \right).
\]

There are \( k + 2 \) distinct vertices in \( P \) with \( k \) of them manifesting as loops whose maximal arcs are the \( k \) mutually crossing arcs that give rise to \( P \) itself and the final two arise from the two corresponding minimal \( S \) and \( T \) arcs. As each of the \( 6k \) 1-faces of \( X_\beta \) appears exactly twice, \( X_\beta^1 \) contains \( 3k \) distinct edges and as many vertices as there are vertices in \( P \), namely \( k + 2 \). As a result,

\[
\text{r}n\text{k} \left( H_2(X_\beta) \right) = |\{ \Delta \mid \Delta \in \beta \}| - \text{r}n\text{k} \left( H_1(X_\beta^1) \right) = 2k - (3k - (k + 2) + 1) = 1.
\]

This implies

\[
\text{r}n\text{k} (H_2(X)) = \sum_{\beta \in I, \beta \neq \beta_0} \text{r}n\text{k} \left( H_2(X_\beta) \right) = |\{ \nabla_\beta \mid \beta \in \mathcal{T}, \beta \neq \beta_0 \}|
\]

and in view of \( \text{r}n\text{k}(\text{Ker}(\hat{\partial}_2)) = \text{r}n\text{k}(H_2(X)) \) and \( \hat{\partial}_2(\nabla_\beta) = 0 \),

\[
\text{Ker}(\hat{\partial}_2) = \langle \{ \nabla_\beta \mid \beta \in \mathcal{T}, \beta \neq \beta_0 \} \rangle.
\]

Claim 3. \( H_3(X) = 0 \) and \( H_2(X^2) \cong \hat{\partial}_3(C_3(X)) \oplus H_2(X) \).

Suppose now that \( X \) might contain 3-simplices. By Proposition 1, we can immediately conclude \( H_3(X) = 0 \) since \( K_3(R') = \emptyset \). Note, however, that the 2-skeleton, \( X^2 \), is not necessarily the complex of a bi-structure, rather \( X^2 \) contains \( X' \) as a subcomplex, together with all removed \( \sigma \)-butterflies.

To prove the second part of the claim, we consider the short exact sequence of chain complexes

\[
\begin{align*}
0 & \longrightarrow C_n(X^2) \overset{I}{\longrightarrow} C_n(X) \overset{J}{\longrightarrow} C_n(X, X^2) \longrightarrow 0 \\
0 & \longrightarrow C_{n-1}(X^2) \overset{I}{\longrightarrow} C_{n-1}(X) \overset{J}{\longrightarrow} C_{n-1}(X, X^2) \longrightarrow 0
\end{align*}
\]

which produces the long exact sequence

\[
\text{r}n\text{k} \left( H_2(X_\beta) \right) = |\{ \Delta \mid \Delta \in \beta \}| - \text{r}n\text{k} \left( H_1(X_\beta^1) \right).
\]
Fig. 8 LHS: a bi-structure with one $CC = \{\alpha_b, \alpha_c, \alpha_2\}$, and its corresponding non-trivial block $\beta$ MLS: the polygonal region $P(\beta)$, $y_m = 1 = y_M$ and $z_m = a = z_M$ MRS: the annular region obtained from $P(\beta)$ by the identification of $m = [1, a] = M$ RHS: the complex corresponding to the block $\beta$

$$H_2(X)$$ is free and so $H_2(X)$ is a projective module, whence the exact sequence

$$0 \rightarrow J^* \rightarrow H_3(X, X^2) \xrightarrow{\partial_3} H_2(X^2) \xrightarrow{I^*} H_2(X) \xrightarrow{J^*} H_2(X, X^2) \rightarrow 0$$

is split exact, i.e., $H_2(X^2) \cong \hat{\partial}_3(C_3(X)) \oplus H_2(X)$. \hfill\qed

6 The geometric realization

Proposition 3 Let $\beta \in T_R$ be a block of a simple bi-structure $R = (S, T)$ with geometric realization $|\beta|$. If $\beta$ is a non-trivial block, then we have $|\beta| \cong S^2$ and otherwise $|\beta| \cong D^2$ holds.

Proof Suppose $\beta$ is trivial. Then, it corresponds to a non-crossing arc which gives rise to two copies of the 2-simplex $\Delta$, whence $|\beta| \cong D^2$, a 2-disk as claimed.

Suppose then $\beta$ is non-trivial. Any two subsequent $\Delta, \Delta' \in \beta$ are distinct 2-simplices and share a single mixed edge since, if $\Delta$ and $\Delta'$ are not adjacent on the backbone, then any gap in $\beta$ is induced by the removal of a closed subsequence. We can construct the $\sim_R$-oriented polygonal presentation of $|\beta|$, $P_\beta$, obtained by pairwise consecutive gluing of the $\beta$-2-faces from left to right, along the backbone, see Fig. 8. Let $Y = [x_m, y_m, z_m]$ and $Y' = [x_M, y_M, z_M]$ be the leftmost and rightmost 2-simplices in $\beta$. W.l.o.g. we can assume that $m = [y_m, z_m]$ and $M = [y_M, z_M]$ are the mixed leftmost and rightmost edges of $Y$ and $Y'$, respectively. $P_\beta$ thus contains only pure $S$-edges, pure $T$-edges and the two distinguished, mixed edges $m = M$, which are complementary, see Fig. 8.

Let $\gamma$ be an $S$-arc corresponding to $\beta$, with $\Delta, \Delta' \in \beta$ being its two endpoint 2-simplices and $[x, y]$ being the pure edge of $\Delta, \Delta'$, i.e., $\gamma = \alpha_x$. Then, $[x, y]$ appears twice as an edge in the boundary of $P_\beta$ and we denote these two instances by $m'$ and $M'$. By construction of $P_\beta$, traversing its boundary clockwise is tantamount to traversing

\[\text{Springer}\]
the backbone from left to right. As such, we first encounter $\Delta$ and then $\Delta'$, which translates to encountering the vertex $y$ of $m'$ first, followed by the vertex $x$ of $m'$. As for $M'$, we first encounter $x$ and then $y$, whence pure edges appear as complementary pairs in $P_\beta$. Analogously pure $T$-edges appear as complementary pairs.

In $\text{Tr}(S)$ and $\text{Tr}(T)$, respectively, any leaf corresponds to a loop and gives rise to the aforementioned complementary edge pairs, whence $P_\beta$ is a polygonal presentation of a sphere and the proposition follows.

\section*{Acknowledgements}
We would like to thank Robert Penner for his input and helpful remarks on the manuscript. We also gratefully acknowledge the comments from Fenix Huang. Many thanks to Thomas Li and Ricky Chen for discussions.

\section*{Funding}
The authors received no financial support for the research, authorship, and/or publication of this article.

\section*{Data availability statement}
Not applicable.

\section*{Declarations}

\subsection*{Conflict of interest}
None.

\subsection*{Code availability}
Not applicable.

\section*{References}

1. Ambjørn, J., Chekhov, L., Kristjansen, C.F., Makeenko, Y.: Matrix model calculations beyond the spherical limit. Nucl. Phys. B 404(1–2), 127–172 (1993)
2. Andersen, J.E., Chekhov, L.O., Penner, R., Reidys, C.M., Sulkowski, P.: Topological recursion for chord diagrams, RNA complexes, and cells in moduli spaces. Nucl. Phys. B 866(3), 414–443 (2013)
3. Andersen, J.E., Penner, R.C., Reidys, C.M., Waterman, M.S.: Topological classification and enumeration of RNA structures by genus. J. Math. Biol. 67(5), 1261–1278 (2013)
4. Barrett, C., He, Q., Huang, F.W., Reidys, C.M.: An efficient dual sampling algorithm with hamming distance filtration. J. Comput. Biol. 25(11), 1179–1192 (2018)
5. Barrett, C., Huang, F.W., Reidys, C.M.: Sequence–structure relations of biopolymers. Bioinformatics 33(3), 382–389 (2017)
6. Barsacchi, M., Novoa, E.M., Kellis, M., Bechini, A.: Swispot: modeling riboswitches by spotting out switching sequences. Bioinformatics 32(21), 3252–3259 (2016)
7. Bon, M., Vernizzi, G., Orland, H., Zee, A.: Topological classification of RNA structures. J. Mol. Biol. 379(4), 900–911 (2008)
8. Bura, A.C., He, Q., Reidys, C.M.: Loop homology of bi-secondary structures. Discrete Math. 344(6), 112371 (2021)
9. Chapuy, G.: A new combinatorial identity for unicellular maps, via a direct bijective approach. Adv. Appl. Math. 47(4), 874–893 (2011)
10. Chen, W., Deng, E., Du, R., Stanley, R., Yan, C.: Crossings and nestings of matchings and partitions. Trans. Amer. Math. Soc. 359(4), 1555–1575 (2007)
11. Cohen, M.M.: A Course in Simple-Homotopy Theory, 2nd edn. Springer, Berlin (1973)
12. Fresco, J.R., Alberts, B.M., Doty, P., et al.: Some molecular details of the secondary structure of ribonucleic acid. Nature 188, 98–101 (1960)
13. Ghrist, R.W.: Elementary Applied Topology, vol. 1. Createspace, Seattle (2014)
14. Harer, J., Zagier, D.: The Euler characteristic of the moduli space of curves. Invent. Math. 85(3), 457–485 (1986)
15. Haslinger, C., Stadler, P.F.: RNA structures with pseudo-knots: graph-theoretical, combinatorial, and statistical properties. Bull. Math. Biol. 61(3), 437–467 (1999)
16. Hofacker, I.L., Schuster, P., Stadler, P.F.: Combinatorics of RNA secondary structures. Discret. Appl. Math. **88**(1–3), 207–237 (1998)
17. Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., Zamir, A.: Structure of a ribonucleic acid. Science 1462–1465 (1965)
18. Huang, F.W., Reidys, C.M.: Shapes of topological RNA structures. Math. Biosci. **270**, 57–65 (2015)
19. Huang, F.W., Reidys, C.M.: Topological language for RNA. Math. Biosci. **282**, 109–120 (2016)
20. Jin, E.Y., Qin, J., Reidys, C.M.: Combinatorics of RNA structures with pseudoknots. Bull. Math. Biol. **70**(1), 45–67 (2008)
21. Levin, A., Lis, M., Ponty, Y., O'donnell, C.W., Devadas, S., Berger, B., Waldispühl, J.: A global sampling approach to designing and reengineering RNA secondary structures. Nucl. Acids Res. **40**(20), 10041–10052 (2012)
22. Nussinov, R., Jacobson, A.B.: Fast algorithm for predicting the secondary structure of single-stranded RNA. Proc. Natl. Acad. Sci. **77**(11), 6309–6313 (1980)
23. Orland, H., Zee, A.: RNA folding and large n matrix theory. Nucl. Phys. B **620**(3), 456–476 (2002)
24. Penner, R., Waterman, M.S.: Spaces of RNA secondary structures. Adv. Math. **101**(1), 31–49 (1993)
25. Reidys, C.M.: Neutral networks of RNA secondary structures (1995)
26. Reidys, C.M., Huang, F.W., Andersen, J.E., Penner, R.C., Stadler, P.F., Nebel, M.E.: Topology and prediction of RNA pseudoknots. Bioinformatics **27**(8), 1076–1085 (2011)
27. Schmitt, W.R., Waterman, M.S.: Linear trees and RNA secondary structure. Discret. Appl. Math. **51**(3), 317–323 (1994)
28. Stanley, R.P.: Enumerative Combinatorics. Wadsworth Publ Co., Belmont (1986)
29. Sundaram, S.: The Cauchy identity for sp (2n) (1990)
30. Thirumalai, D., Lee, N., Woodson, S.A., Klimov, D.: Early events in RNA folding. Annu. Rev. Phys. Chem. **52**(1), 751–762 (2001)
31. Waterman, M.S.: Secondary structure of single-stranded nucleic acids. Adv. Math. Suppl. Stud. 1, 167–212 (1978)
32. Waterman, M.S., Smith, T.F.: Rapid dynamic programming algorithms for RNA secondary structure. Adv. Appl. Math. **7**(4), 455–464 (1986)
33. Whitehead, J.: Simplicial spaces, nuclei and m-groups. Proc. Lond. Math. Soc. **s2–45**(1), 243–327 (1939)
34. Zuker, M., Sankoff, D.: RNA secondary structures and their prediction. Bull. Math. Biol. **46**(4), 591–621 (1984)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.