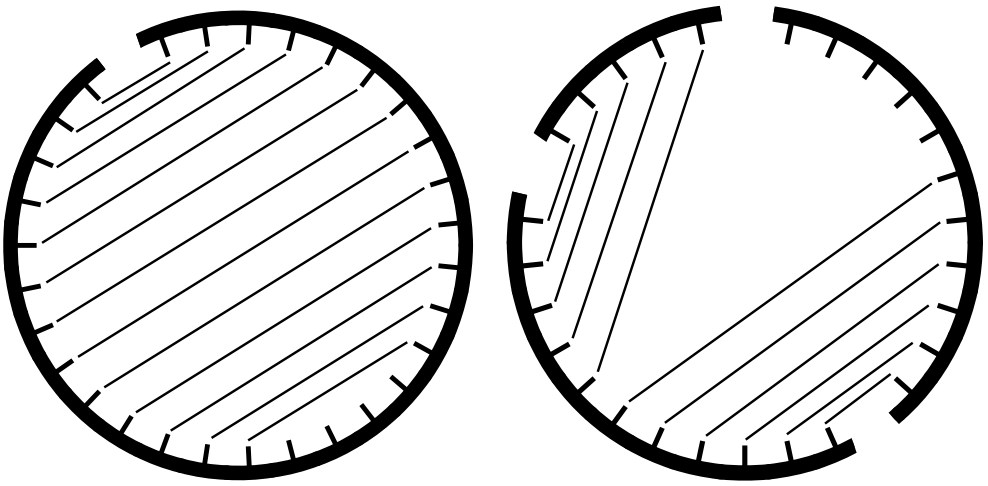
Appendix D

Strand Orderings for Pseudoknot-Free Representations

Given a complex microstate c, we can draw a *polymer graph* representation (also called the *circle-chord* representation) by laying out the strands on a circle, in the ordering $\pi^*(c)$, and representing the base pairs by chords connecting the appropriate locations (Figure D.1). In the case of a complex microstate with a single strand, we call the secondary structure *pseudoknotted* if there are crossing chords. However, the case where a complex microstate contains multiple strands requires a slightly more complex definition. We note that in this case, a strand ordering $\pi^*(c)$ corresponds to one particular way of arranging the strands on the circle; the circular permutations of $\pi^*(c)$ are the $(L-1)!$ permutations of the strands which are distinct when arranged on a circle, e.g. for three strands labeled A, B, C, there are only two distinct circular permutations: (A, B, C) and (A, C, B). With that in mind, we call an arbitrary secondary structure pseudoknotted if every circular permutation of the strand ordering has a polymer graph representation that contains a crossed chord.

While our simulator is not constrained to using a strand ordering $\pi^*(c)$ whose polymer graph representation does not contain a crossed chord, it is convenient for us to do so, as the output representation is easier to generate in these cases. The following heuristics allow us to maintain the property that our complex microstates always use a strand ordering $\pi^*(c)$ whose polymer graph does not contain a crossed chord:

The initial strand orderings we generate are based on a dot-paren structure, which naturally translate to a polymer graph with no crossed chords. The only time strand orderings can change is when performing a bimolecular move (either a break move or join move). For a break move, the resulting pair of complexes maintain the same orderings

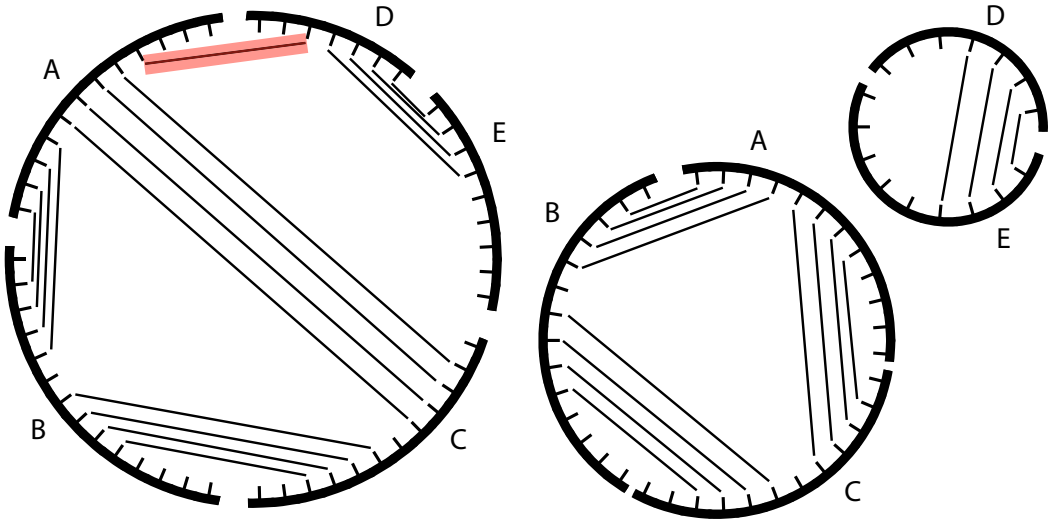


(a) A hairpin with long stem (b) Short three-way branch migration system

Figure D.1: Two different secondary structures using polymer graph representation. Strands are always arranged with $5' \rightarrow 3'$ orientation being clockwise around the circle.

the strands originally had in the original complex – e.g. if we had a complex of 5 strands (A through E) with ordering (A, D, E, C, B) which broke apart into two complexes with strands A, B, C and D, E, the resulting pair of orderings would be (A, C, B) and (D, E) (Figure D.2).

For a join move, we first note that a complex's open loops correspond to sequential pairs of strands in the circular strand ordering (this corresponds to the strands on either side of the "nick" in the dot-paren representation), so that a complex with ordering (A, B, D, C) has open loops corresponding to the pairs (A, B), (B, D), (D, C) and (C, A). When we perform a join move, we take each complex's strand ordering and find the cyclic permutation (these are the permutations that are identical when arranged on a circle, e.g. (A, B, C), (B, C, A) and (C, A, B) are cyclic permutations on the strand ordering (A, B, C)) which places the affected open loops at the edge of the permutation. For example, if we are joining (A, B, D, C) with (E, G, F) by the open loops around (B, D) and (E, G), we use the cyclic permutations (D, C, A, B) and (G, F, E), ending up with the strand ordering (D, C, A, B, G, F, E) for the resulting complex (Figure D.3). Why does this ordering have no crossing chords (assuming the starting ones did not)? We note that this join move was joining a base on either the $5'$ end of D or the $3'$ end of B, with a base on either the $5'$ end of G or the $3'$ end of E - in all four of these cases, the resulting chord cannot cross



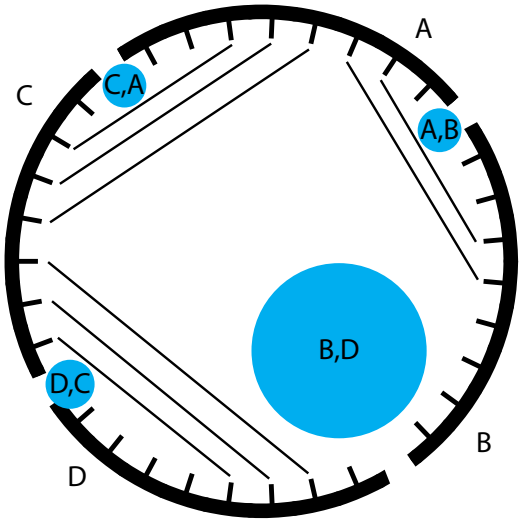
(a) (A,D,E,C,B) before break move

(b) (A,C,B) and (D,E) after break move

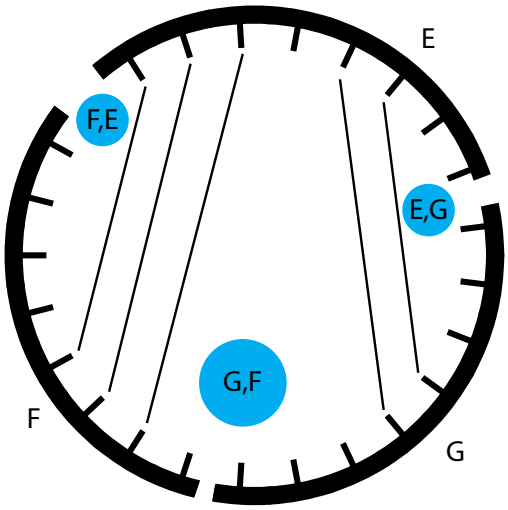
Figure D.2: Polymer graph representation before and after a break move (base pair highlighted in red). Note that the ordering is consistent, but we now have two separate complexes and thus two separate polymer graphs.

any existing chords if we use the given strand ordering. Note that in general, the bases available for a join move are not necessarily from the $5'$ and $3'$ edges near the nick but the same argument applies. For example, if we used a join from the (G, F) open loop (from the previous example), there is a single base on strand E that could be used to make the join, but since it's in the same open region as the $5'$ end of G and $3'$ end of F, it also could not create a crossed chord.

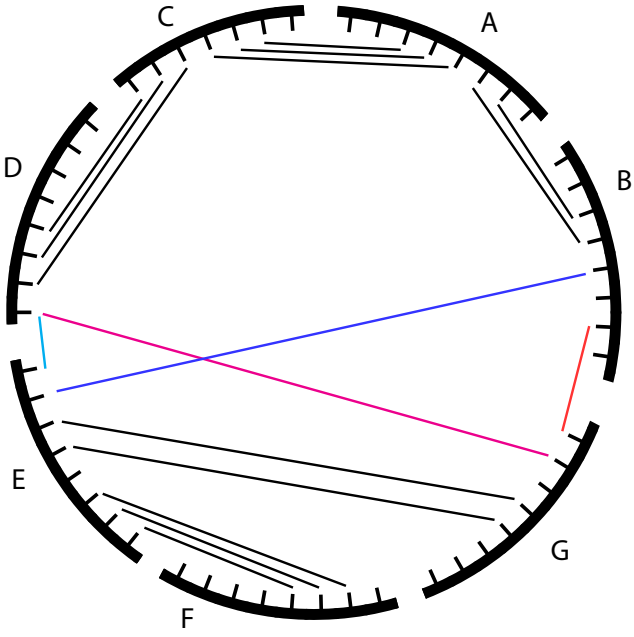
So we have shown that the strand ordering $\pi^*(c)$ maintained by our simulator for a complex microstate c has a polymer graph with no crossed chords. This leads naturally to the question of whether there is a different circular permutation of $\pi^*(c)$ which also has no crossed chords. The (surprising) answer is no - every other circular permutation has at least one crossed chord! This is stated in the following theorem:



(a) (A,B,D,C) before join move



(b) (E,G,F) before join move



(c) (D,C,A,B,G,F,E) after cyclic permutations and join

Figure D.3: Polymer graph representation before and after a join move. Open loop regions are noted with a cyan circle marker. Four of the (many) possible join moves between open loops (B, D) and (E, G) are shown in (c), using red, blue, magenta, and cyan.

D.1 Representation Theorem

For every non-pseudoknotted complex microstate c, there is exactly one circular permutation $\pi^*(c)$ *whose polymer graph has no crossed chords.*

While the above heuristic can be expanded on to prove this theorem via induction on the number of strands in a complex, a more thorough proof can be found in [5] and so we will not reproduce it here.