Tangling single-stranded (ss)RNA.

Stephen Hyde. University of Sydney

Absence of knots in known RNA structures

Cristian Micheletti[a,1], Marco Di Stefano[a], and Henri Orland[b,c]

[a]Physics, Scuola Internazionale Superiore di Studi Avanzati, 34136 Trieste, Italy; [b]Institut de Physique Théorique, Commissariat à l'énergie atomique-Saclay, CEA, 91191 Gif-sur-Yvette, France; and [c]Beijing Computational Science Research Center, Haidian District Beijing, 100084, China

Edited by Mehran Kardar, Massachusetts Institute of Technology, Cambridge, MA, and accepted by the Editorial Board December 24, 2014 (received for review September 24, 2014)

The ongoing effort to detect and characterize physical entanglement in biopolymers has so far established that knots are present in many globular proteins and also, abound in viral DNA packaged inside bacteriophages. RNA molecules, however, have not yet been systematically screened for the occurrence of physical knots. We have accordingly undertaken the systematic profiling of the several thousand RNA structures present in the Protein Data Bank (PDB). The search identified no more than three deeply knotted RNA molecules. These entries are rRNAs of about 3,000 nt solved by cryo-EM. Their genuine knotted state is, however, doubtful based on the detailed structural comparison with homologs of higher resolution, which are all unknotted. Compared with the case of proteins and viral DNA, the observed incidence of knots in available RNA structures is, therefore, practically negligible. This fact suggests that either evolutionary selection or thermodynamic and kinetic folding mechanisms act toward minimizing the entanglement of RNA to an extent that is unparalleled by other types of biomolecules. A possible general strategy for designing synthetic RNA sequences capable of self-tying in a twist-knot fold is finally proposed.

RNA structure | RNA knots | physical knots | PDB-wide topological profiling

25) when it is packaged inside a viral particle, where it cannot be simplified by topoisomerases (26, 27).

As for the case of proteins, the discovery of knots in packaged viral DNA raised several questions about their functional implications, particularly for the expected difficulty of ejecting the knotted genome from the narrow capsid exit pore. More recent studies have shown that this conundrum can be solved by considering the working of topological friction at the molecular scale (28) and especially, the ordering effect of DNA self-interactions (25), which favors the untying of DNA knots inside the capsid during ejection (29).

Nowadays, the occurrence of physical entanglement in proteins and DNA is documented and characterized well enough that novel knotted proteins and short DNA molecules have been successfully designed, and the average entanglement of DNA filaments can be created or relaxed in a controlled manner (30–32).

These topological profiling efforts, however, have not been paralleled for the third and last kind of strand of life (33), namely RNA. To the best of our knowledge, no systematic survey of physical knots in RNAs has been carried out so far, and no genuine physical knots have been reported in naturally occurring individual RNA structures.

What are the simplest designs for knotted RNA ?

The simplest “tangle” is a double helix (duplex) - Watson-Crick DNA

complementary nucleotides are H-bonded:

(C…G) & (A…T) in DNA

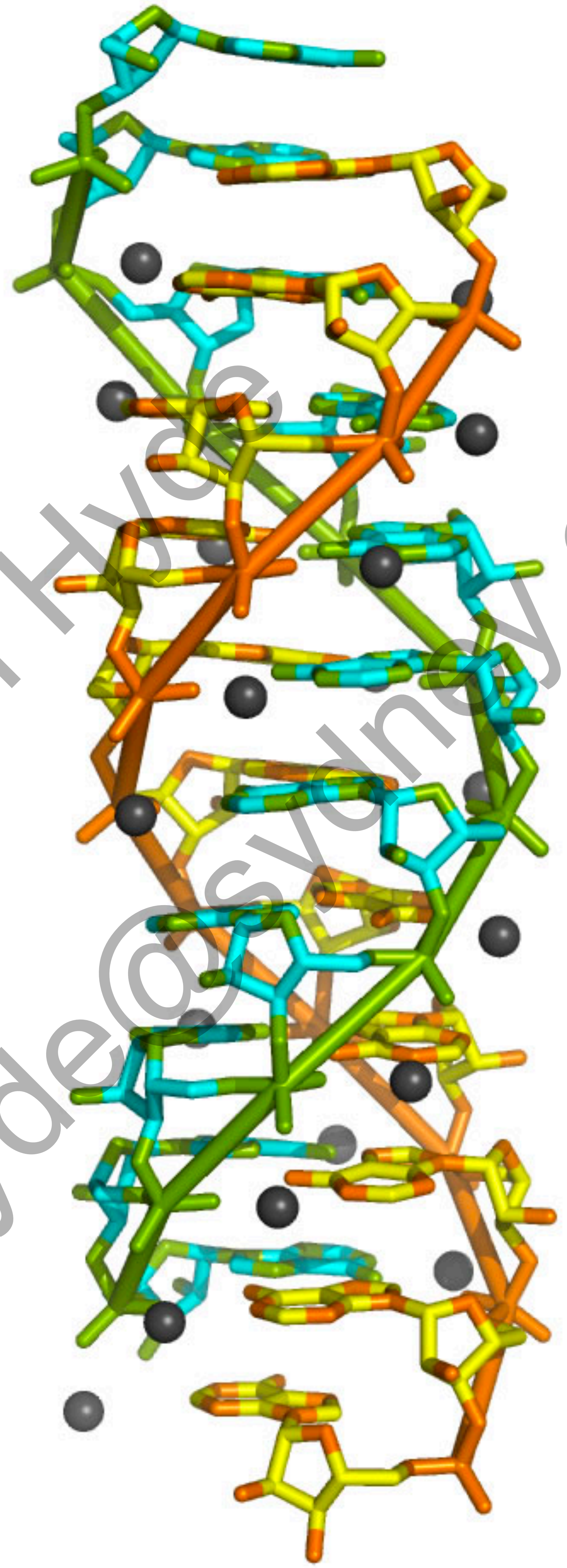
Wikipedia

RNA also builds double helices

complementary nucleotides are H-bonded:

(C…G) & (A…T) in DNA

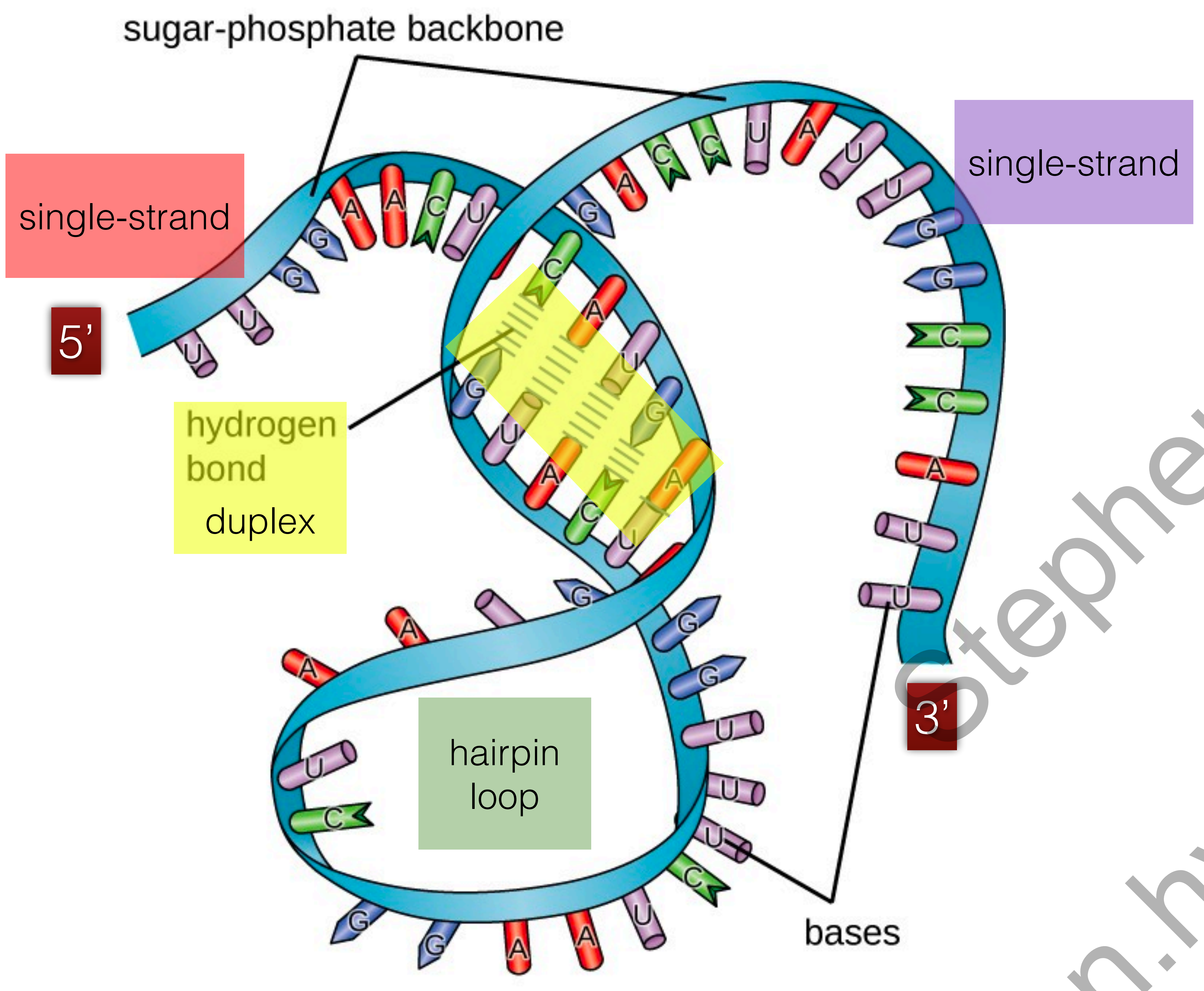
(C…G) & (A…U) in RNA



Kathryn Janzen, Canadian LightSource

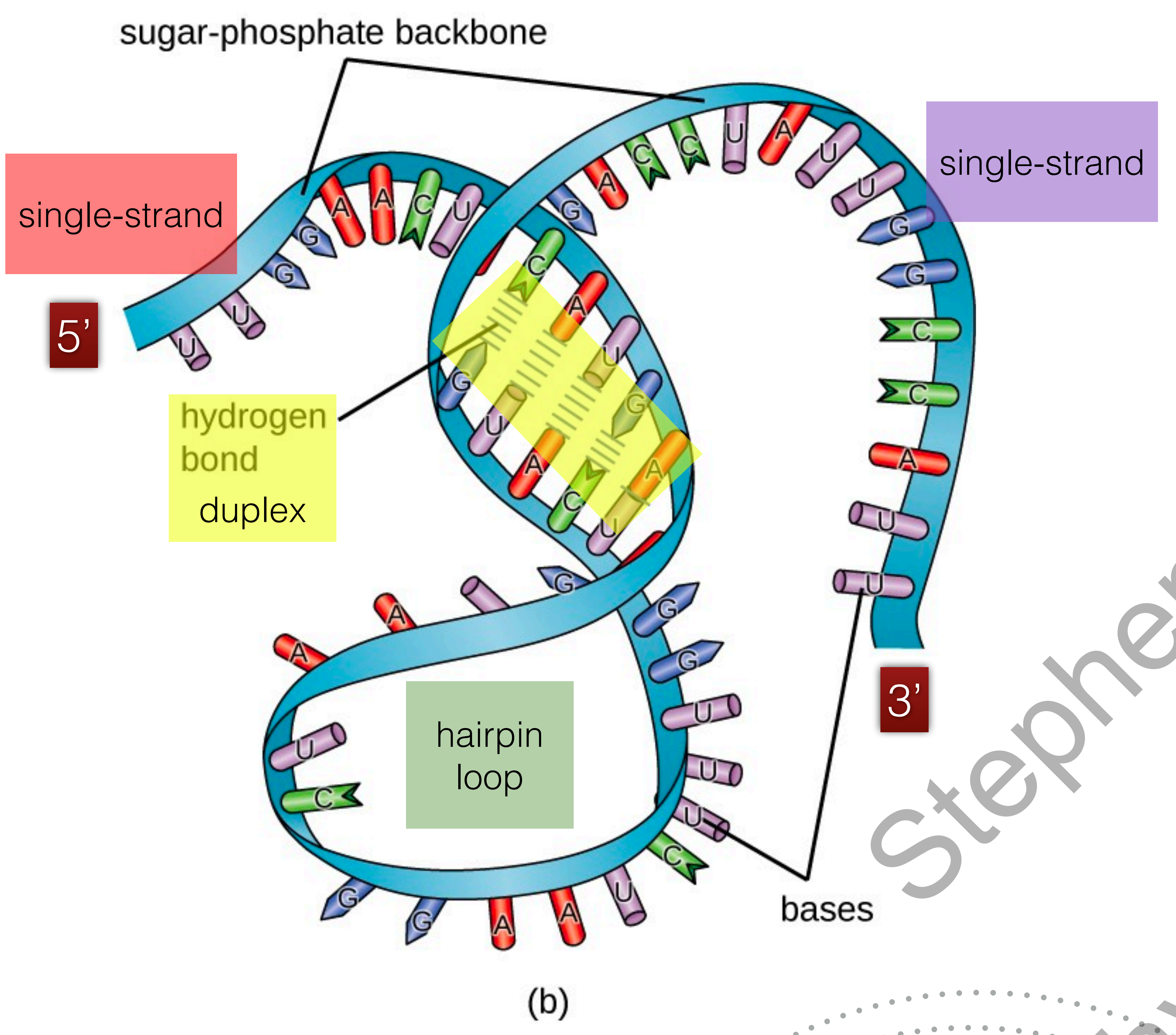
…often by folding on itself

ssRNA secondary fold topology is governed combinatorics of complementary nucleotide sequences



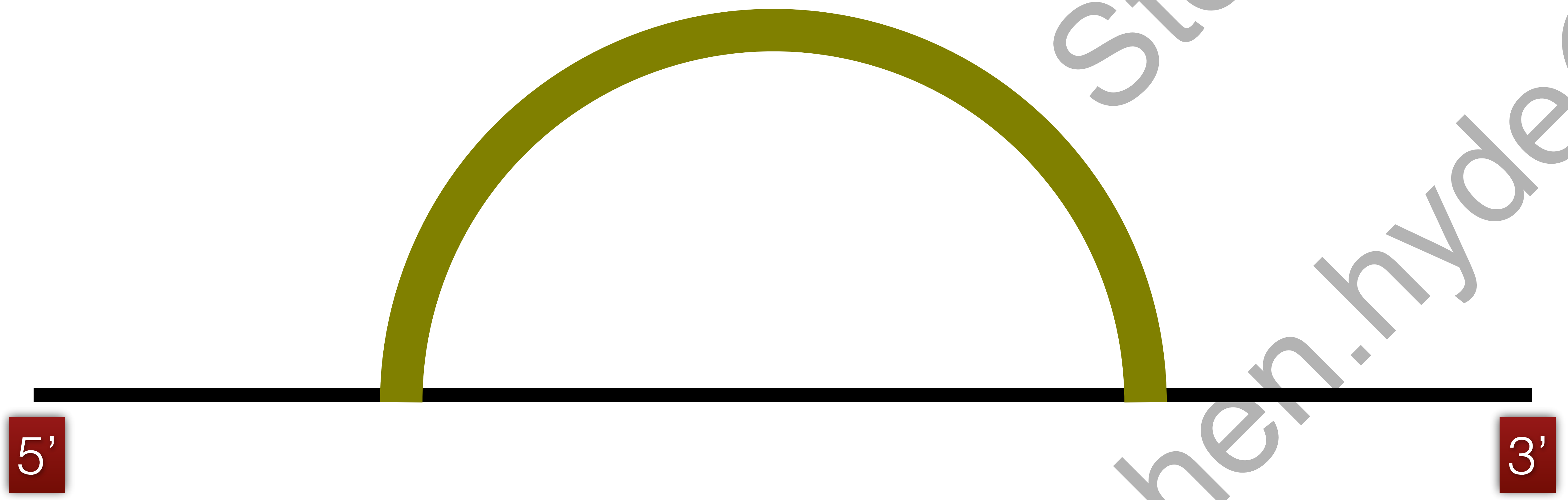
(b)


sugar-phosphate backbone
single-strand
5’
hydrogen bond
duplex
hairpin loop
bases
single-strand
3’
(b)
5’
single-strand
duplex
hairpin loop
duplex
single-strand
3’

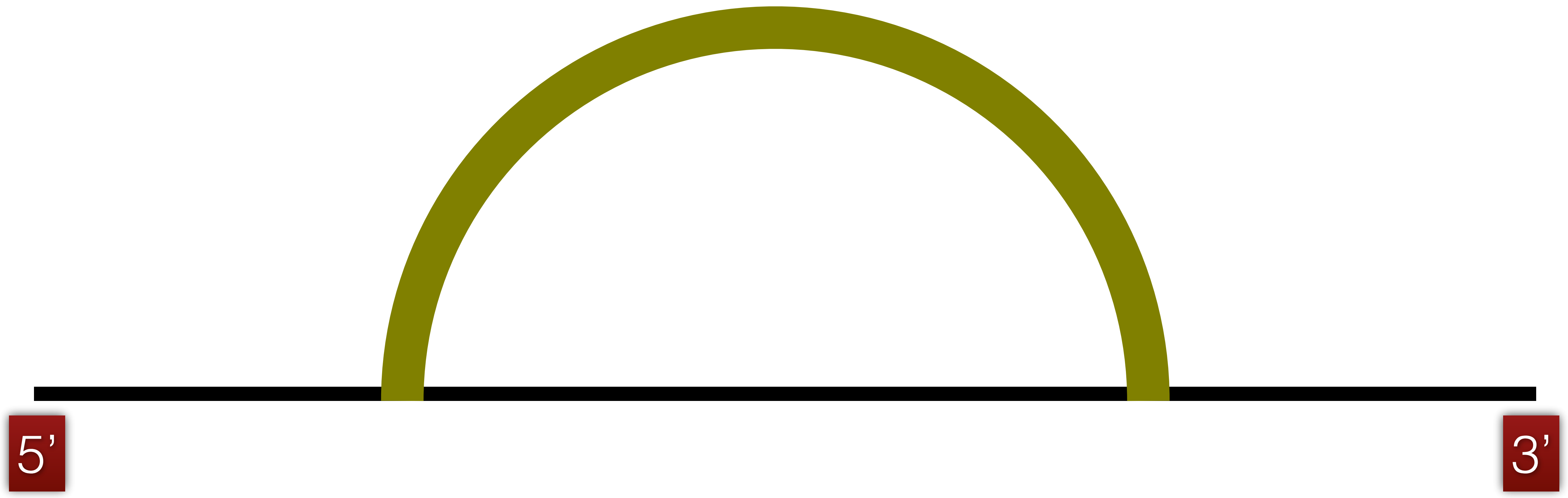

sugar-phosphate backbone
single-strand
5'
hydrogen bond
duplex
hairpin loop
single-strand
3'
bases
(b)


hydrogen bonding
5'
single-strand
duplex
hairpin loop
duplex
single-strand
3'

This arc diagram encodes the folding -
1 antiparallel duplex only



capture this antiparallel duplex fold *topology* with a simplified LINEAR RIBBON diagram



capture single antiparallel duplex fold topology with a simplified diagram

RIBBON DIAGRAM
1 antipar duplex

form closed single strand (bring 5’ and 3’ ends together)

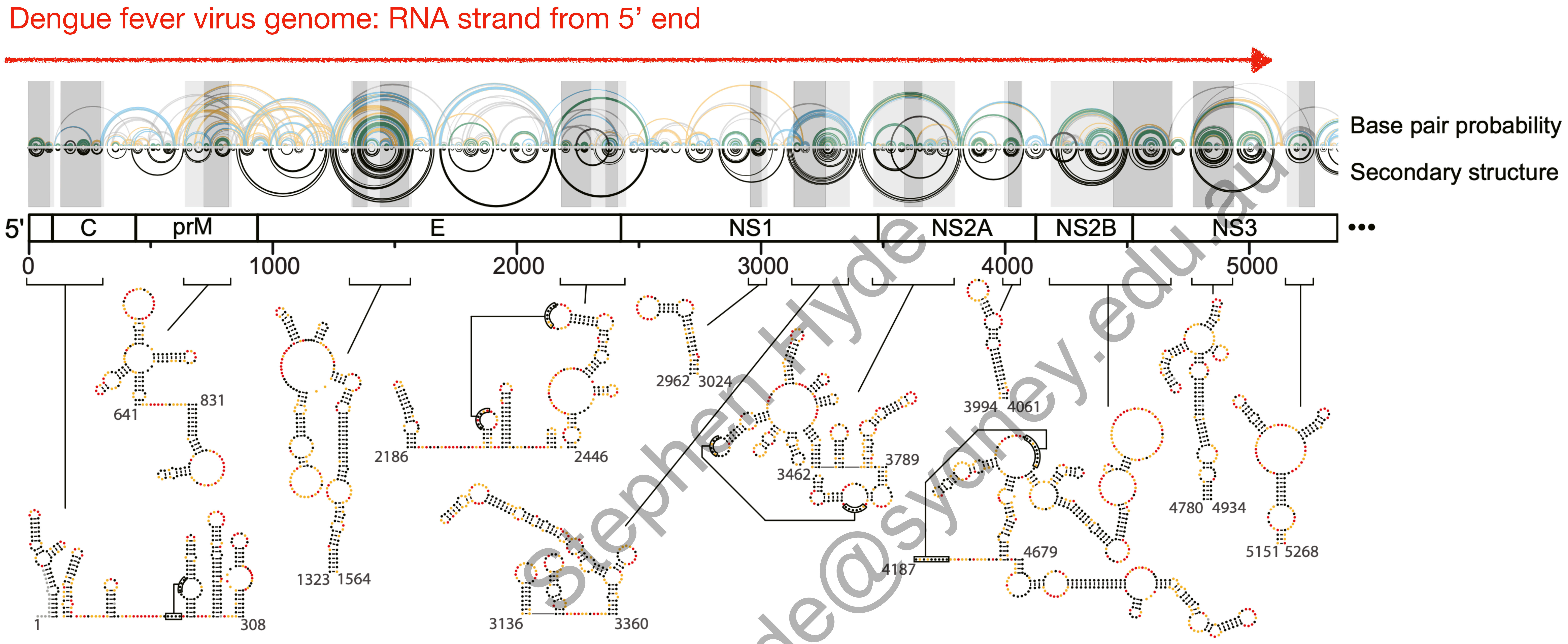
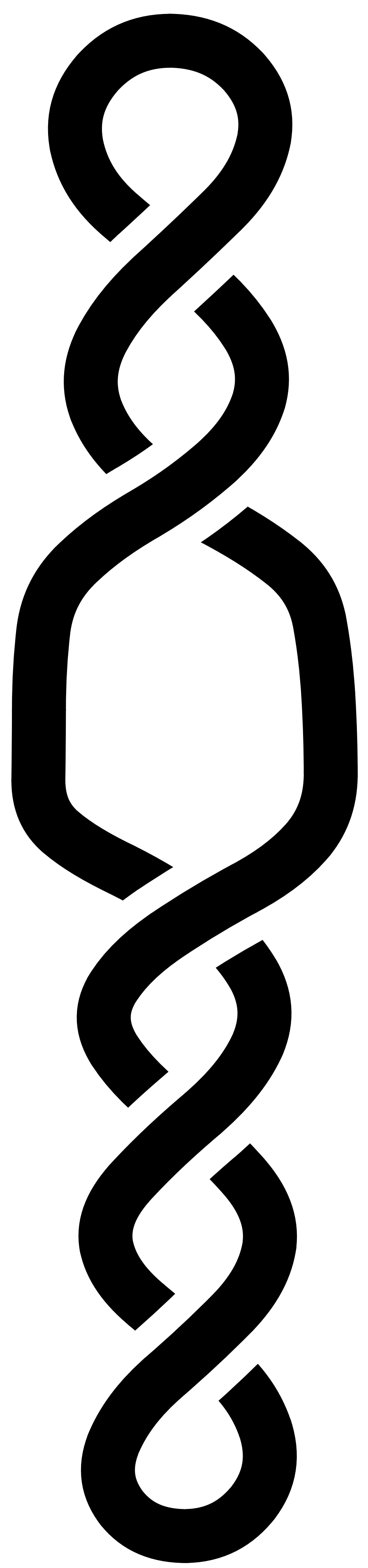
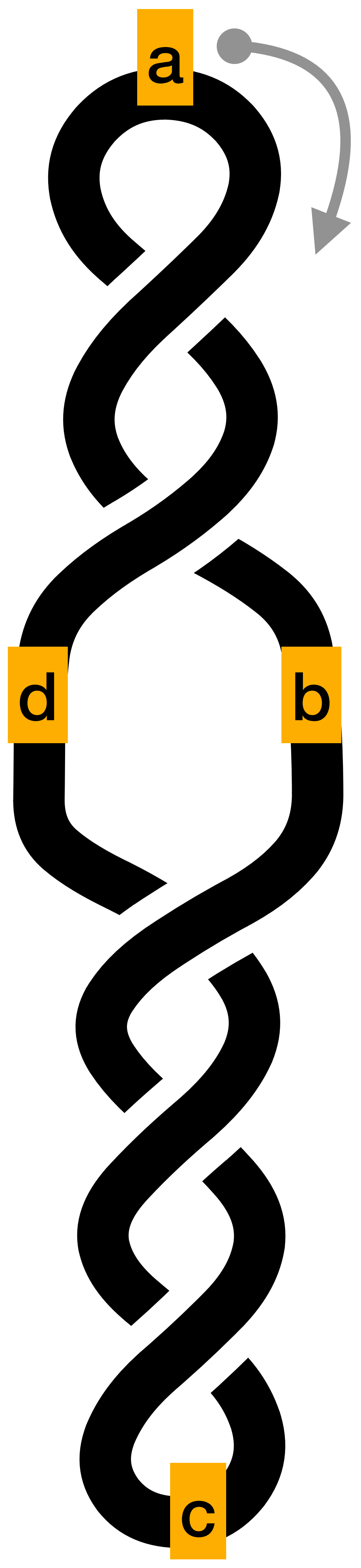


Fig. 1. Well-determined secondary structure elements in the DENV2 RNA genome. The first half of the genome is shown (the entire genome is shown in *SI Appendix*, Fig. S1, panel 1). Median ex virion 1M7 SHAPE reactivities (black) and Shannon entropies (dark blue) are plotted over centered 55-nt windows. Regions with both low SHAPE and low Shannon entropy are highlighted by dark-gray shading, with light-gray shading extended to encompass entire intersecting helices. The first 12 elements (out of 24 total elements) with well-determined structures are numbered. Base pair probability arcs are colored by probability (see scale), with green arcs indicating the most probable base pairs; black arcs indicate plausible pseudoknots (PK). The minimum free-energy secondary structure (inverted black arcs) was obtained using both 1M7 and differential SHAPE reactivities as constraints (1, 12, 13). Secondary structures of elements 1–12 are colored by SHAPE reactivity; high-resolution structures are provided in *SI Appendix*, Fig. S1, panel 2.

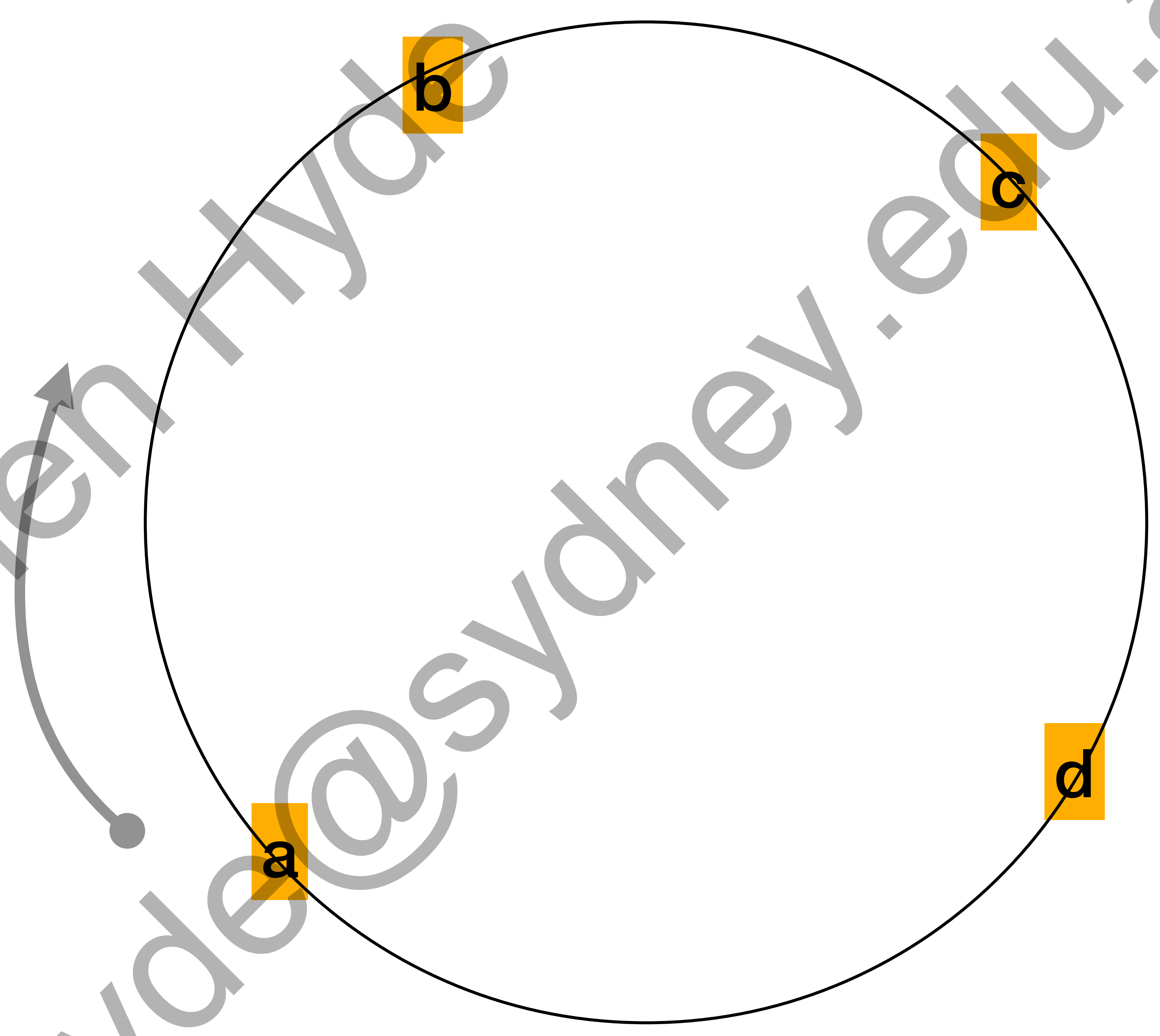
… assume the ssRNA string is closed….



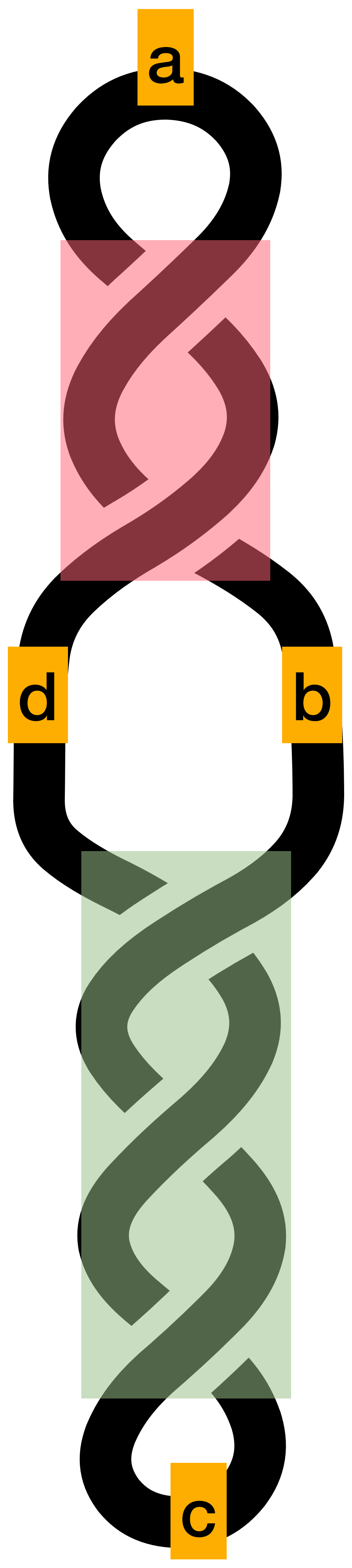
a ‘fold’



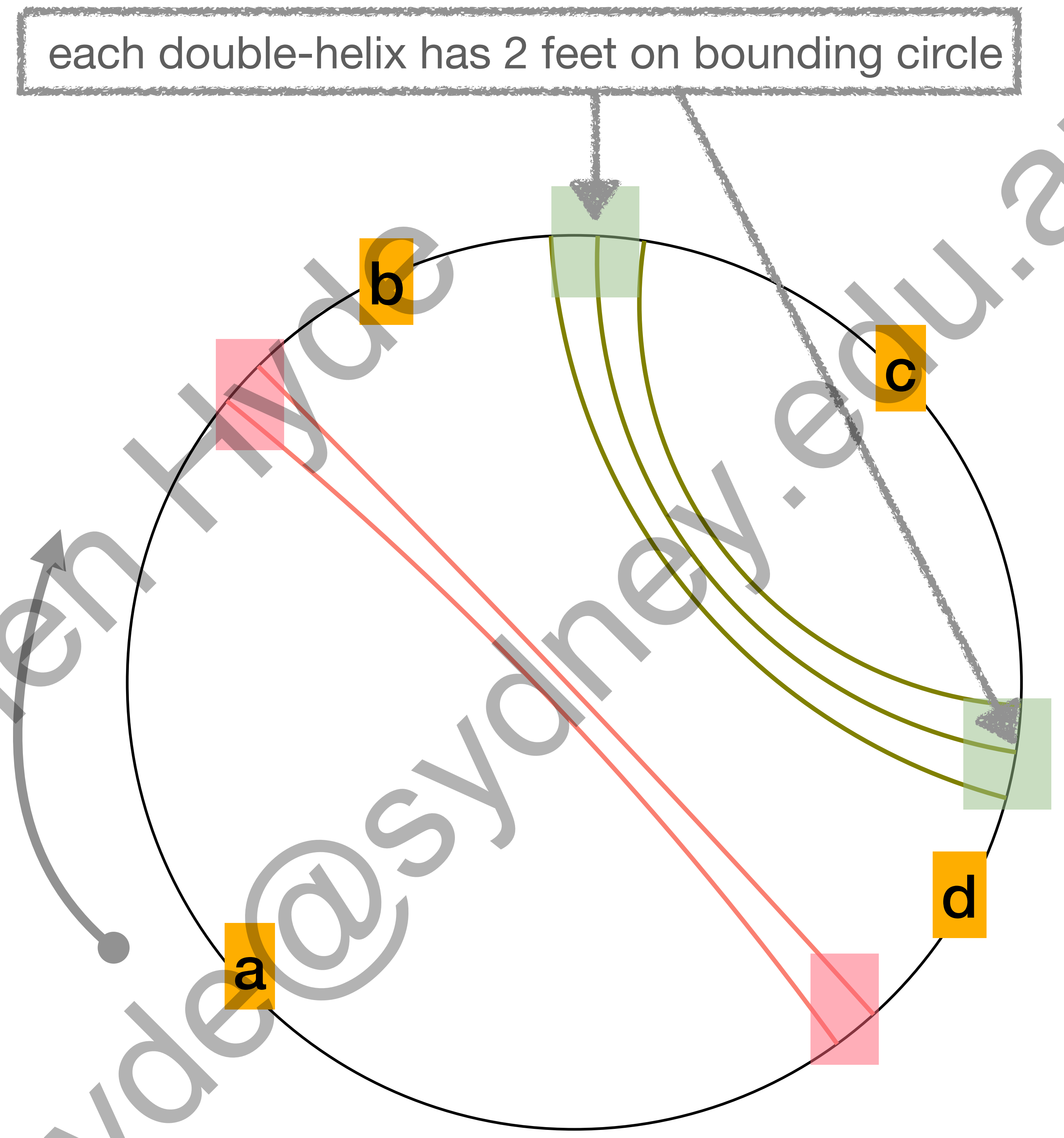
a fold



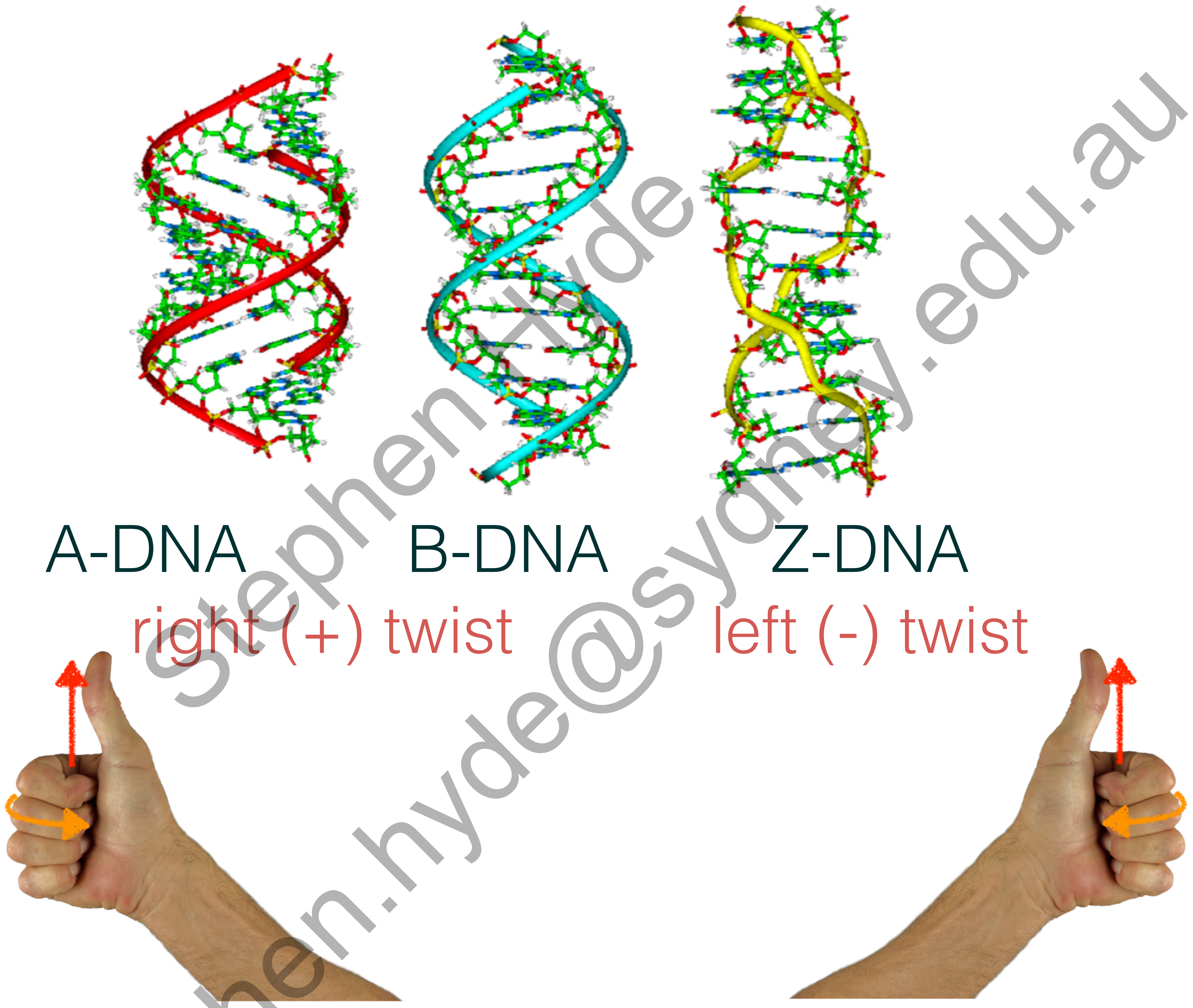
circular diagram of the denatured fold

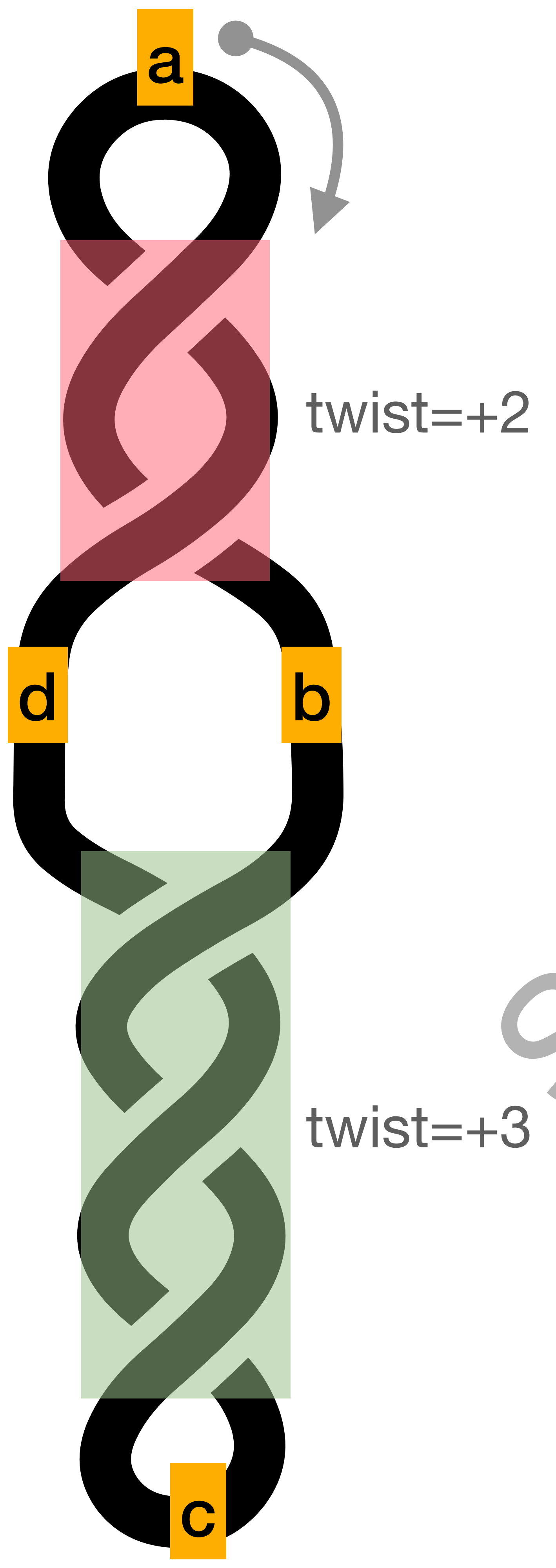


a fold

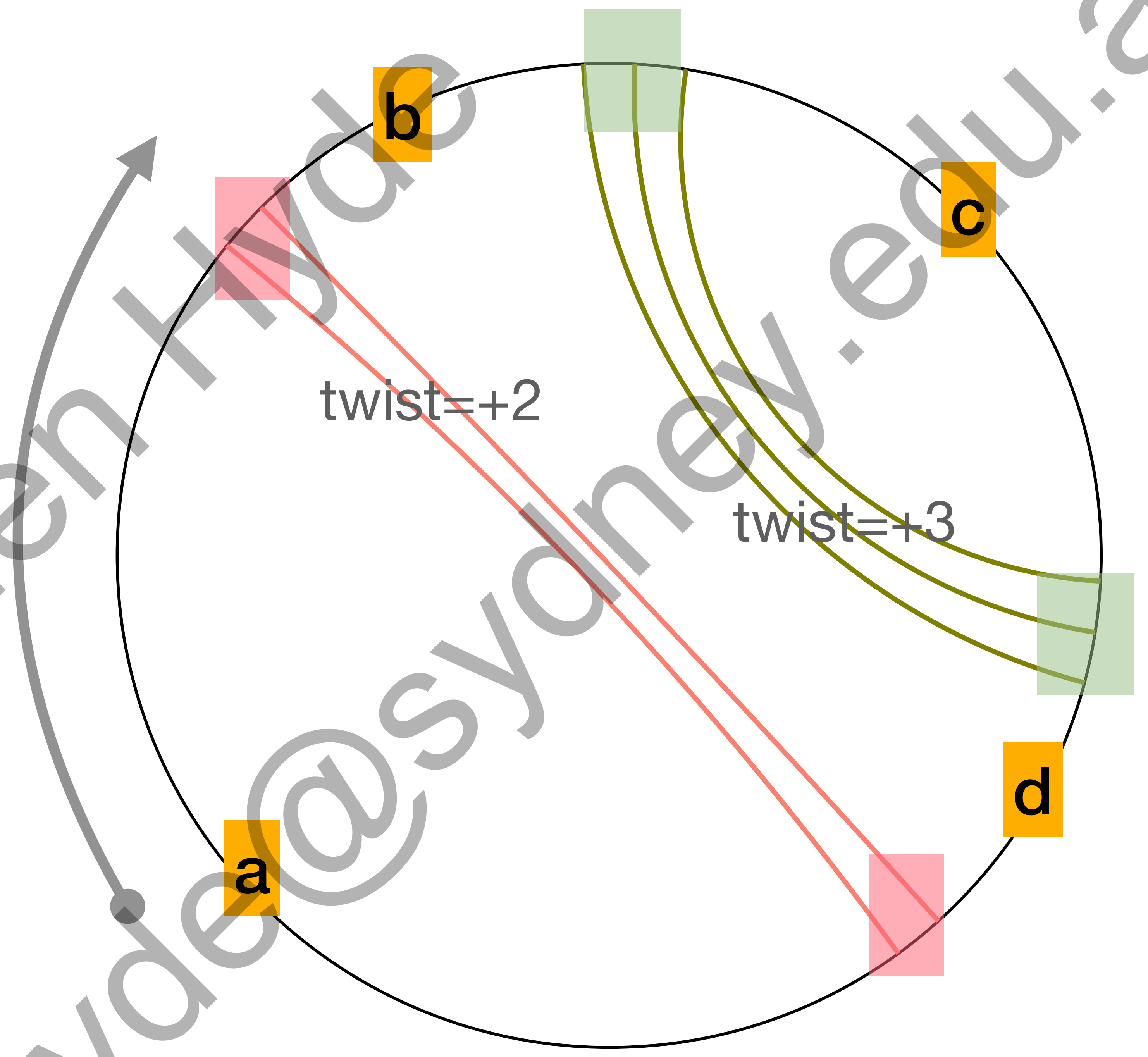


circular diagram of the fold

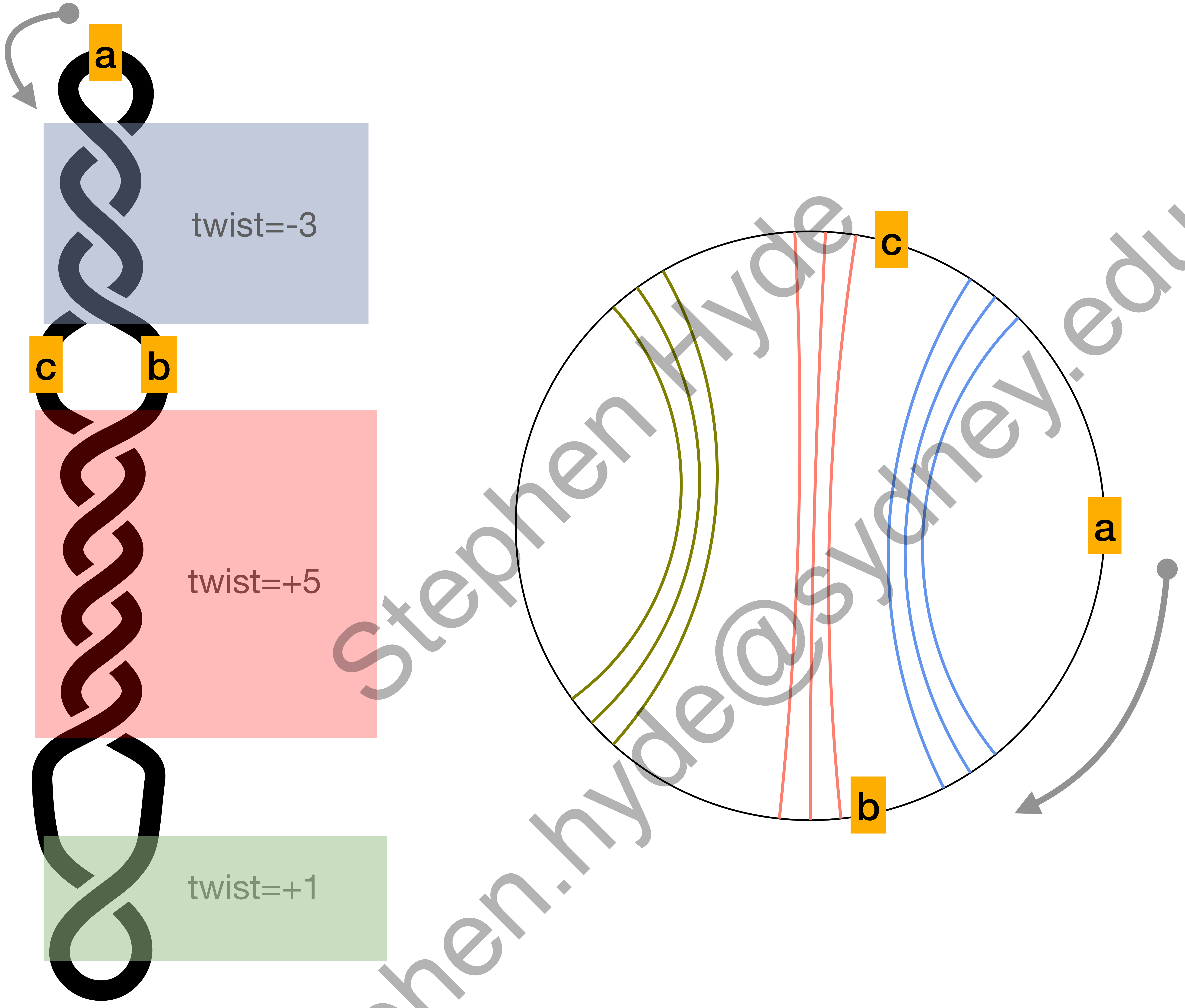

A-DNA
B-DNA
Z-DNA
right (+) twist
left (-) twist

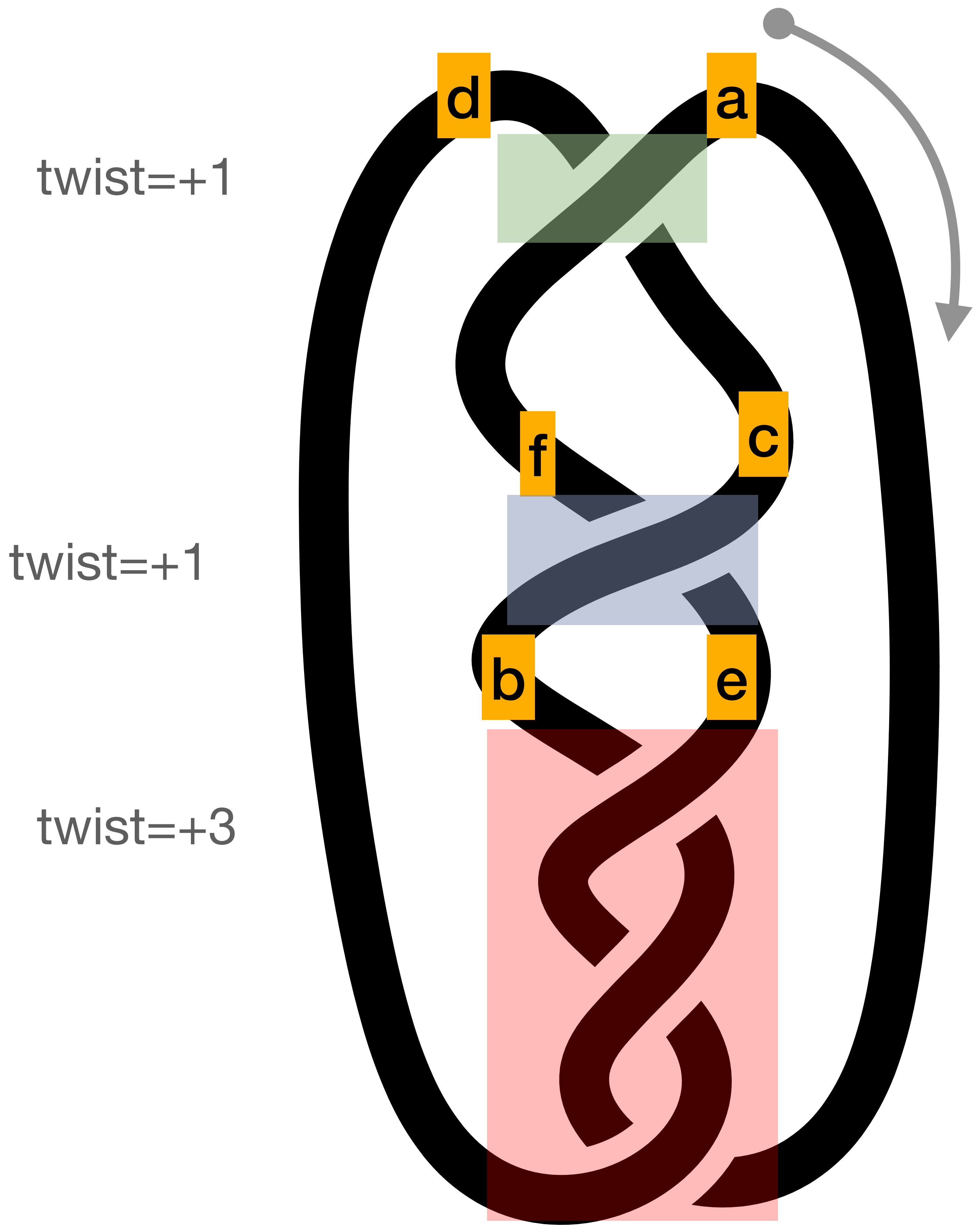


a fold

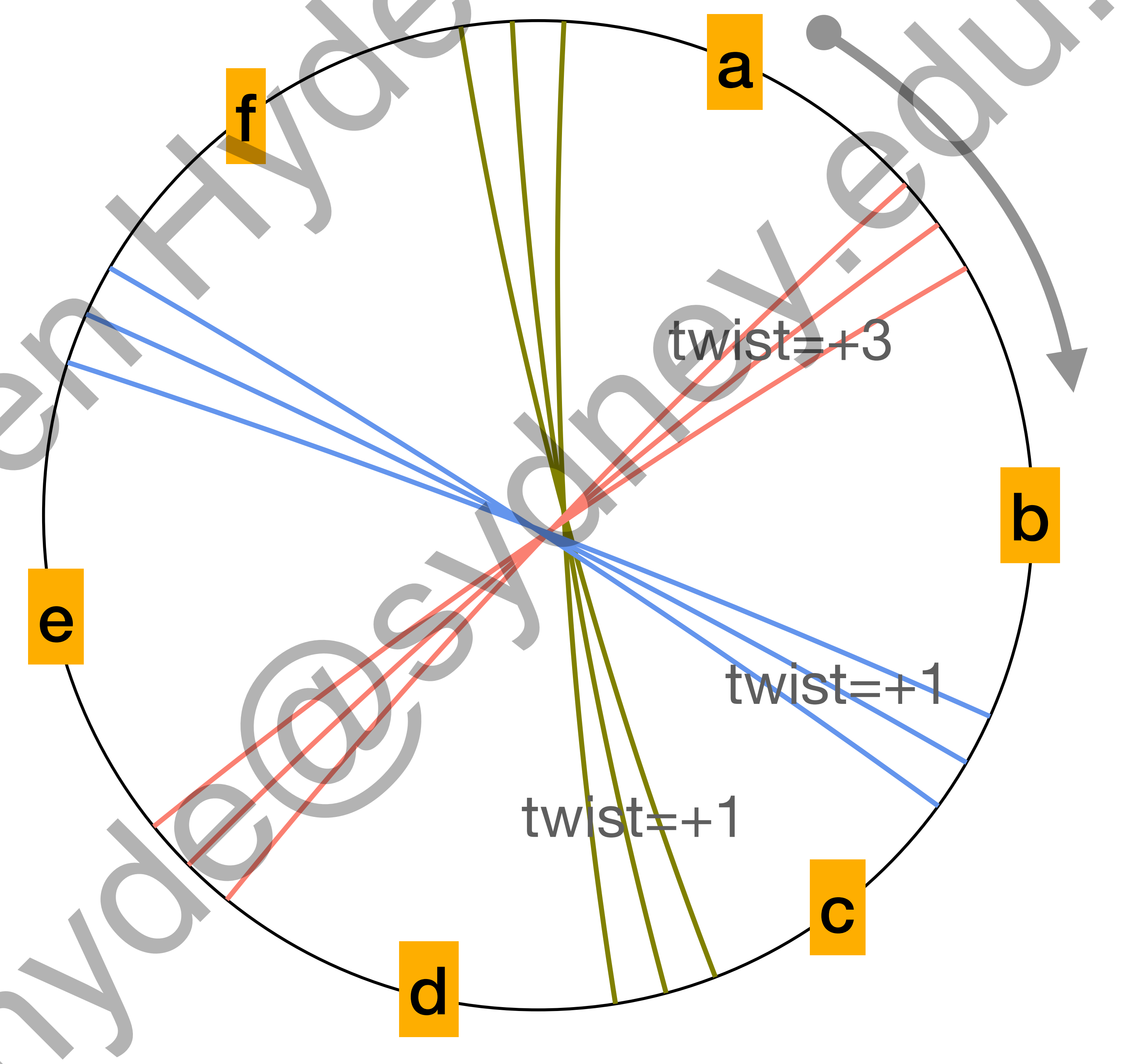


circular diagram of the fold

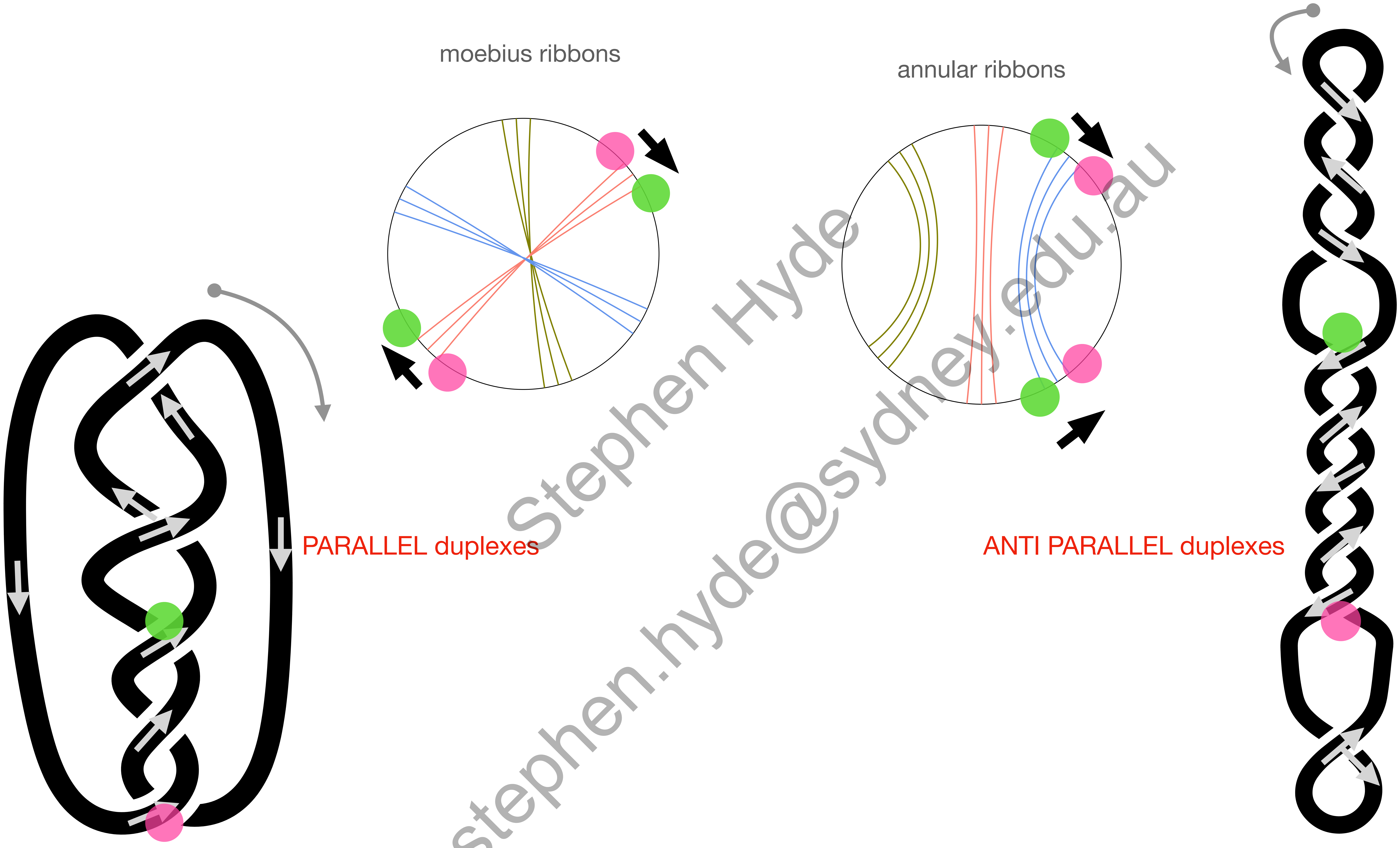

a
twist=-3
c
b
twist=+5
twist=+1
c
a
b



twist=+1
twist=+1
twist=+3
d
a
f
c
b
e

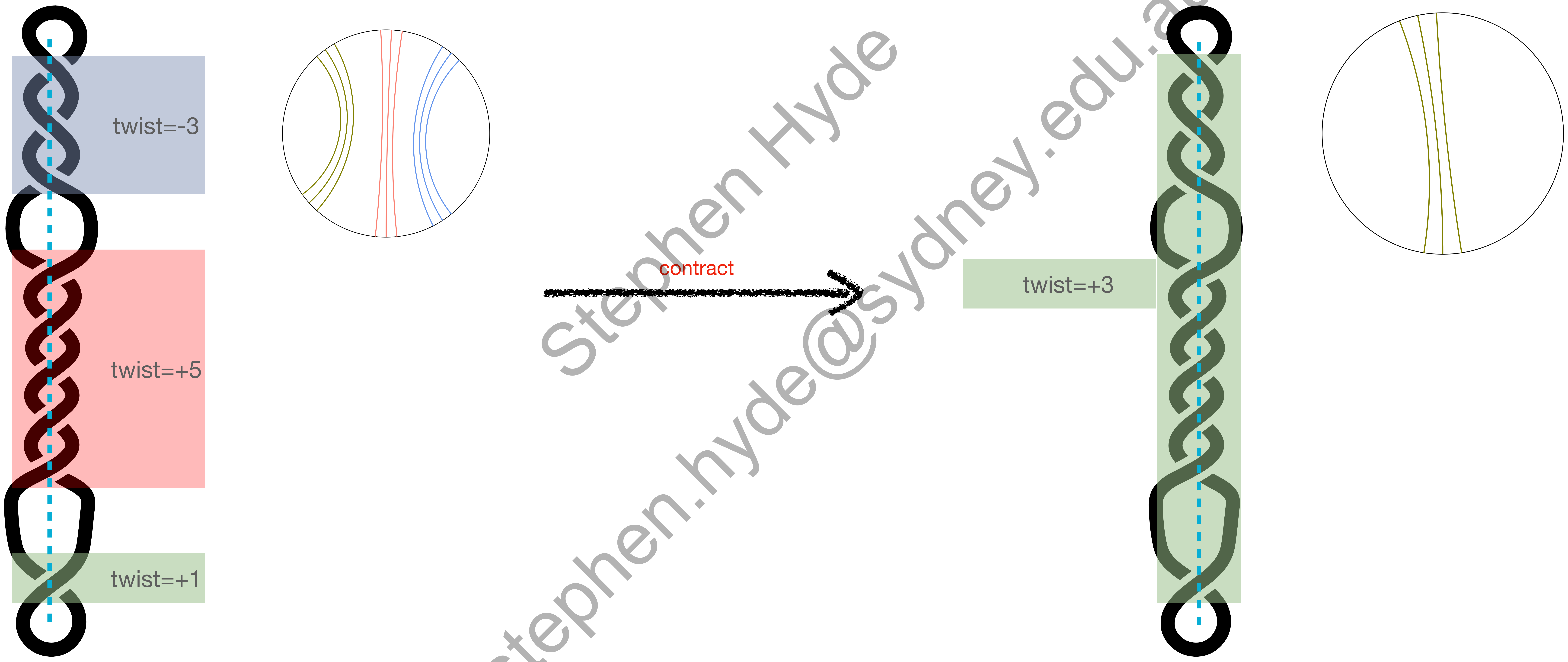


a
f
b
e
c
d
twist=+3
twist=+1
twist=+1


moebius ribbons
annular ribbons
PARALLEL duplexes
ANTI PARALLEL duplexes

CONTRACT *multi* antiparallel duplexes along common “plumbline” axis

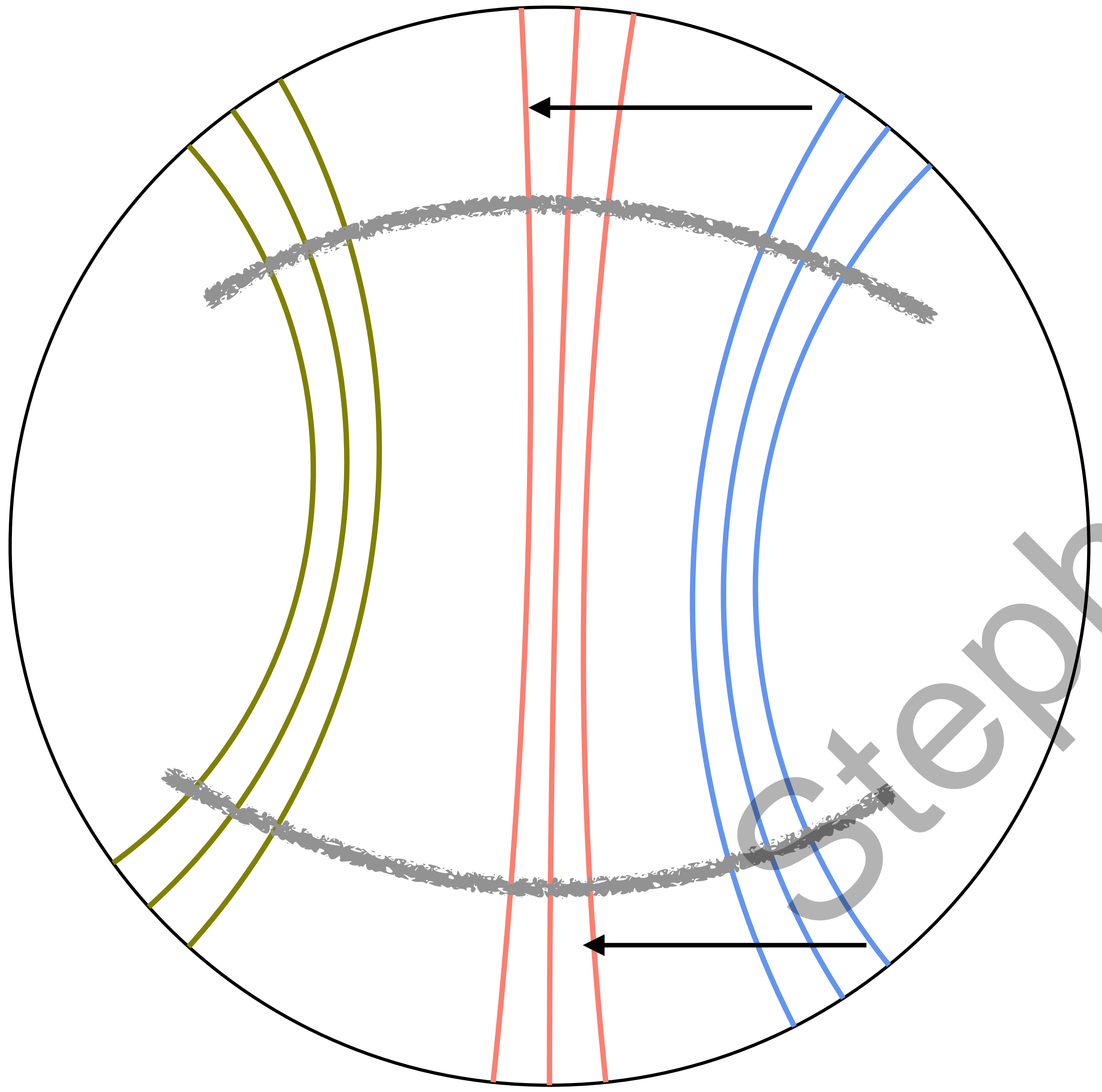
-> *single* antiparallel duplex, twist = sum of local twists

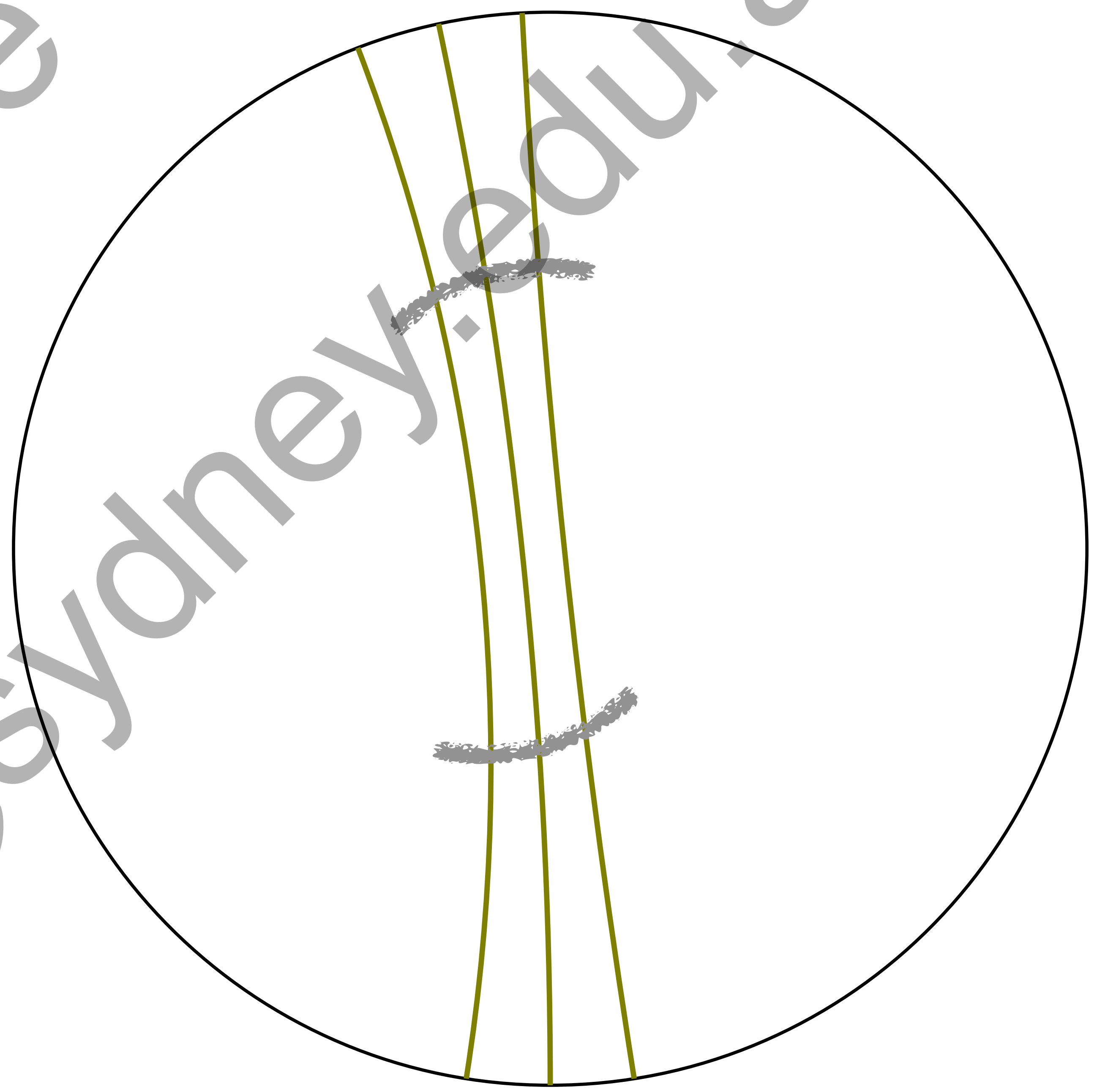


multi parallel duplexes along common "plumbline" axis

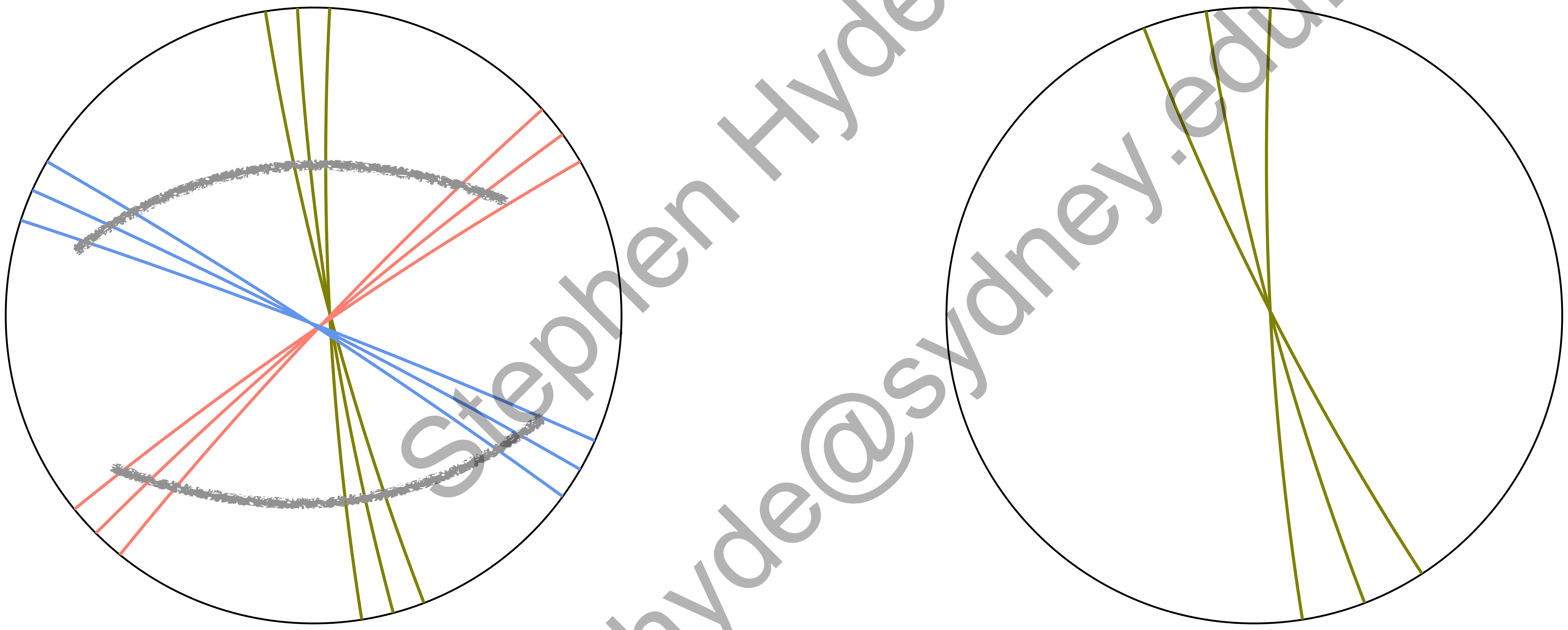
-> single parallel duplex

RULE 1: nested annular ribbons collapse to a single annular ribbon



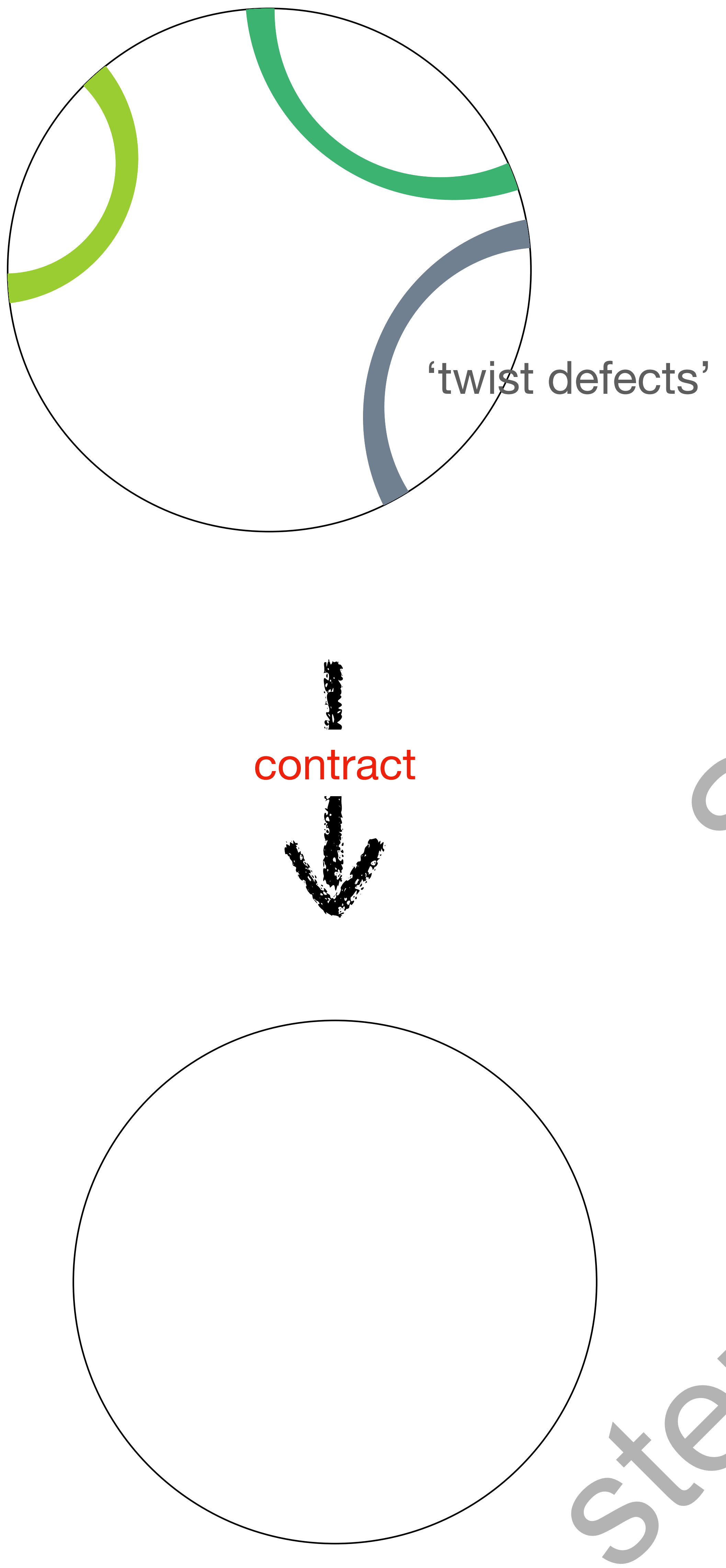


RULE 2: fans of moebius ribbons fold up to a single moebius ribbon

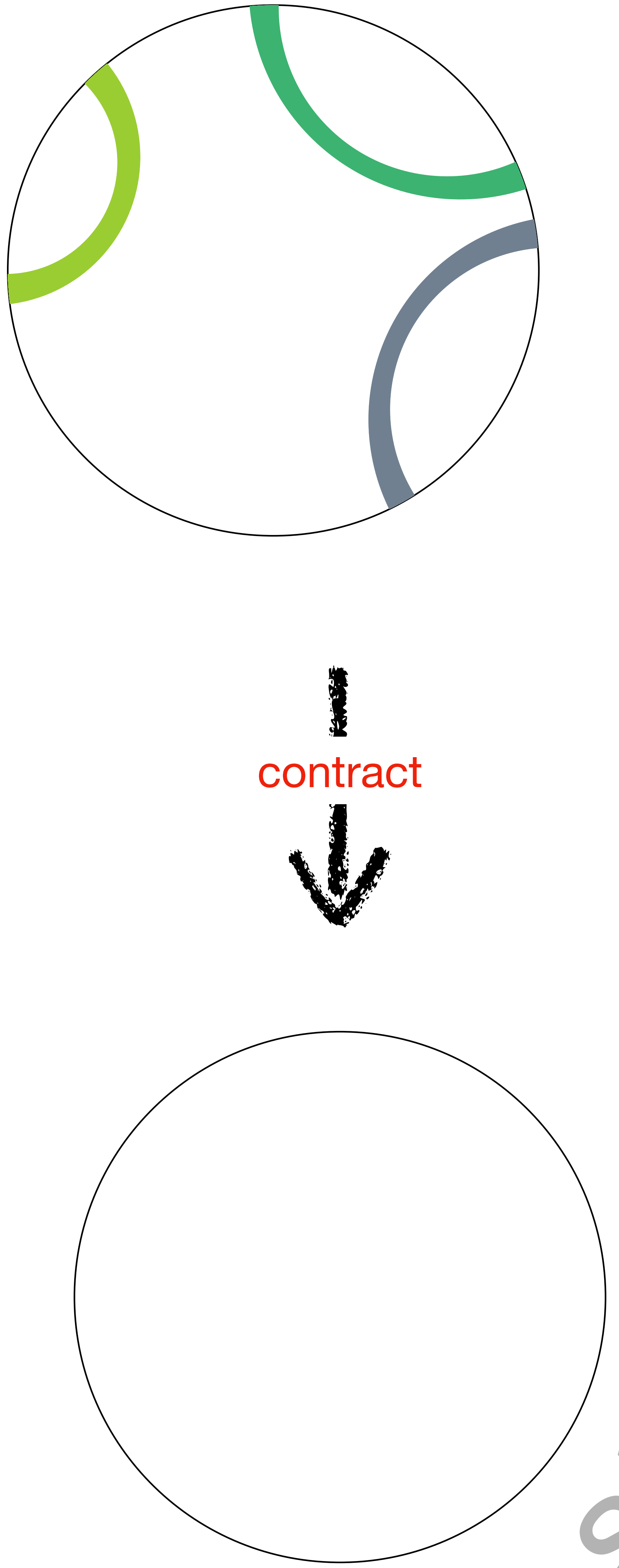


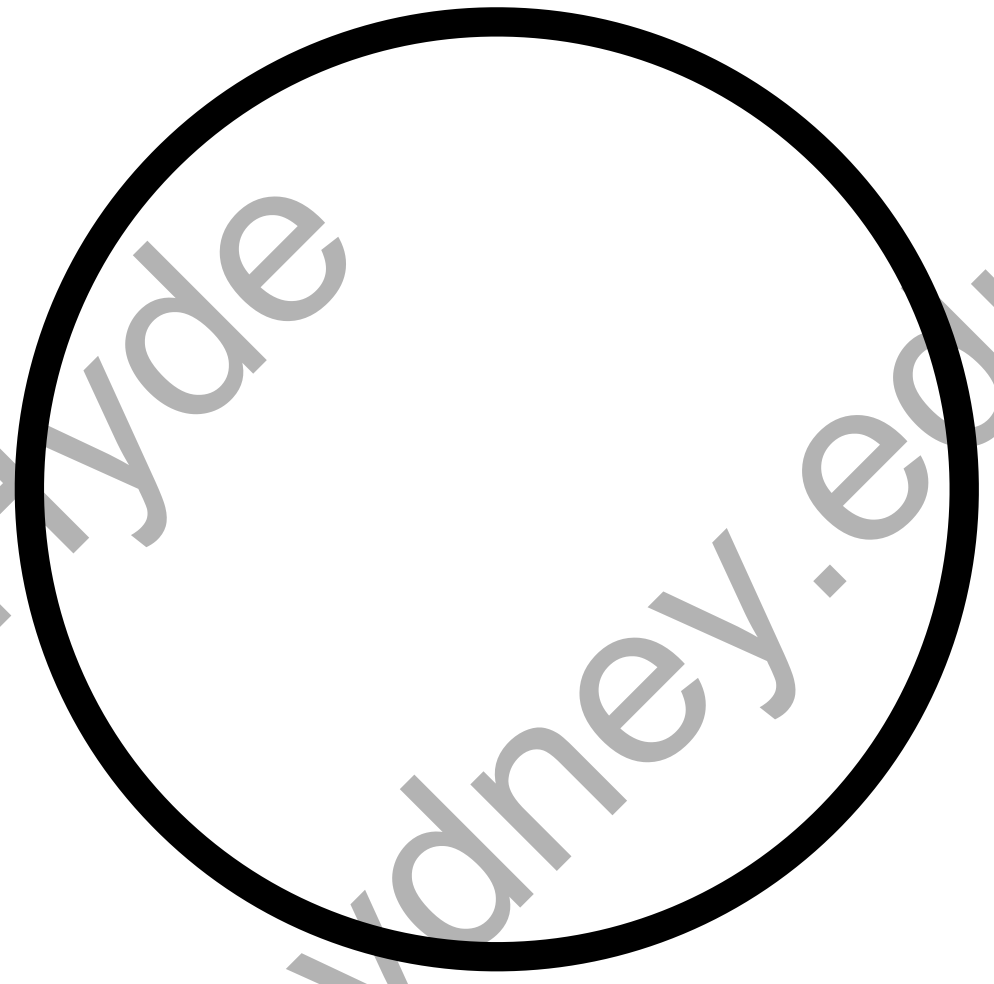
closing up the fan = CONTRACTING multi parallel double-helices

RULE 3. free annular ribbons are deleted

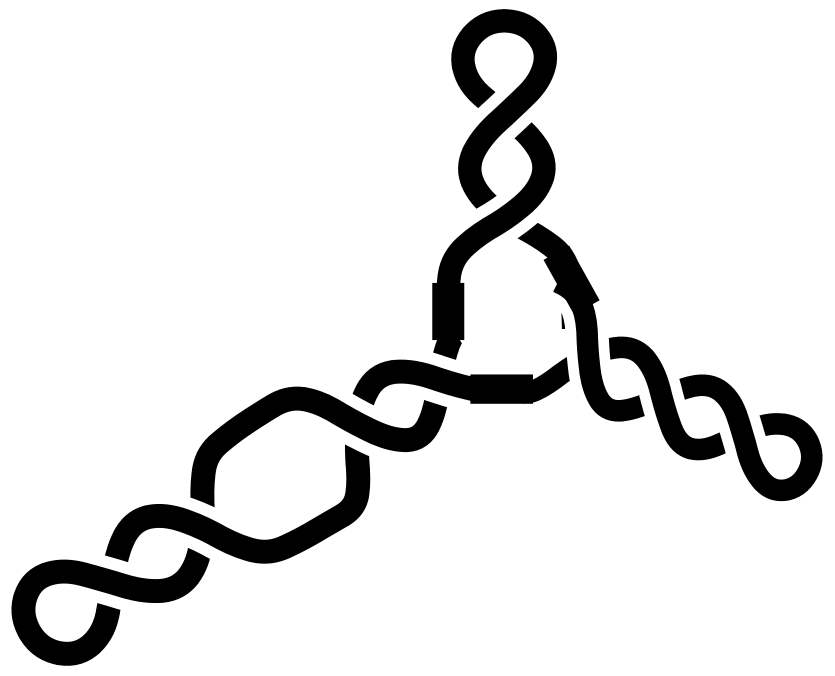


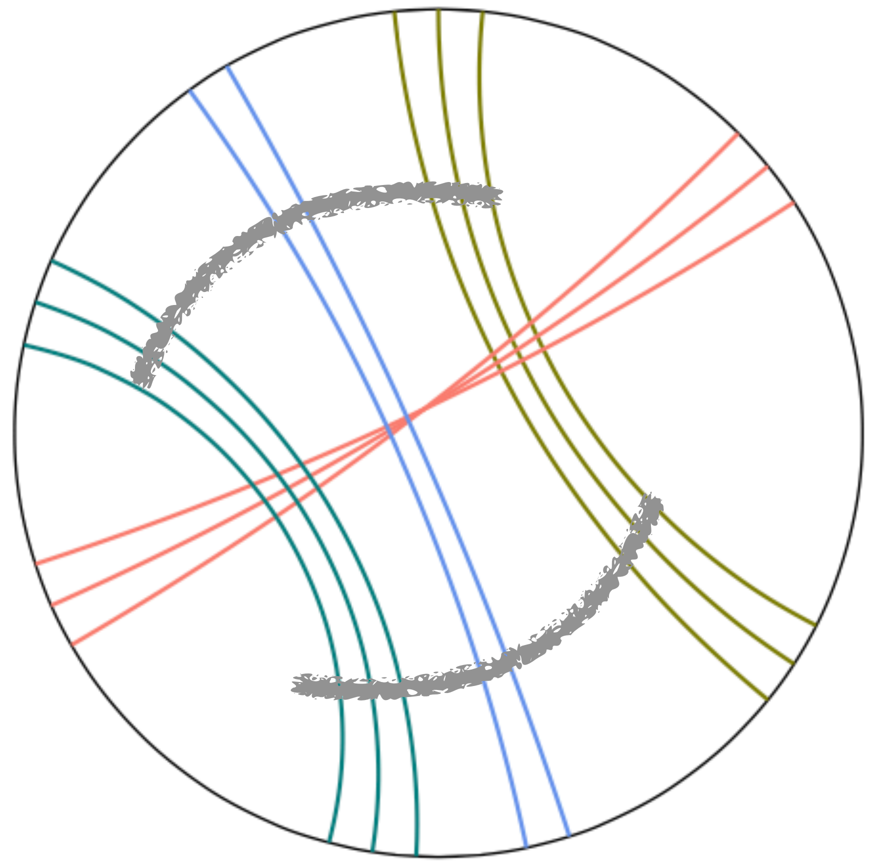
(RULE 3. free annular ribbons are deleted)





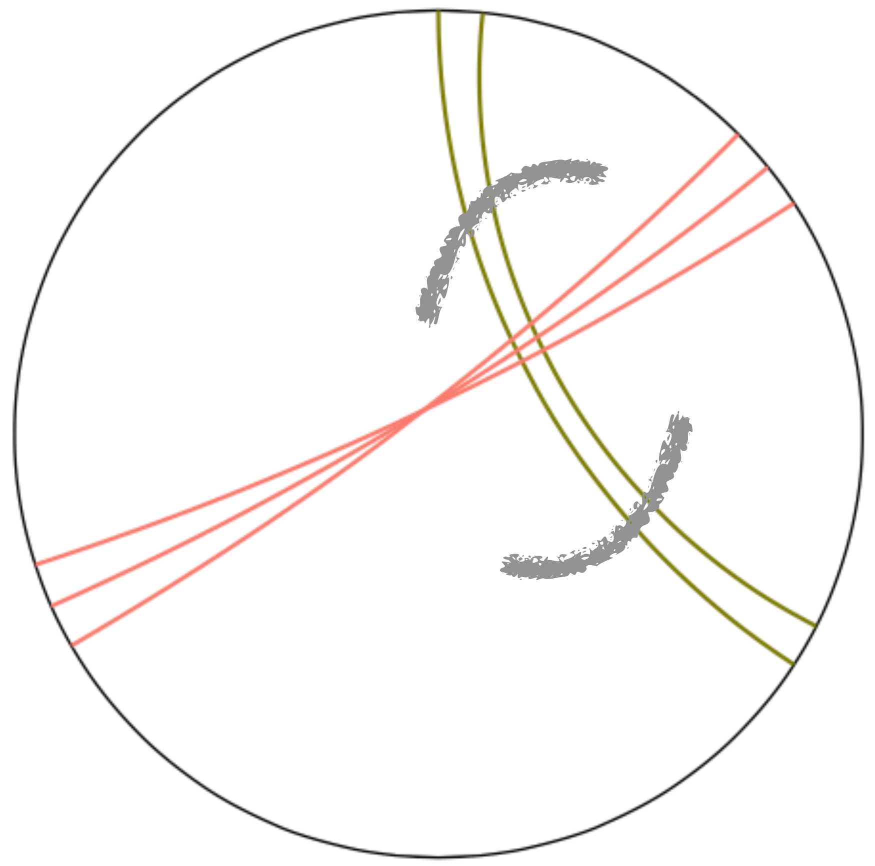
- we can intercalate an arbitrary number of these defects into any fold ad nauseam…





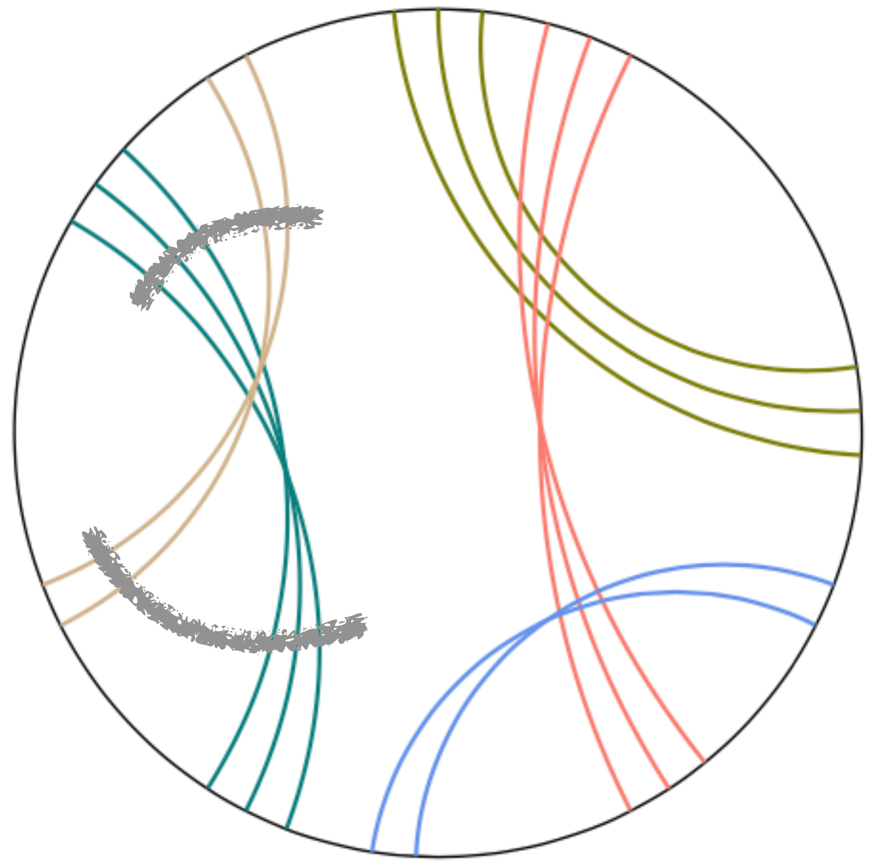
$$\{1_o\overline{2}_o13_e4_o243\}$$

$$\downarrow$$



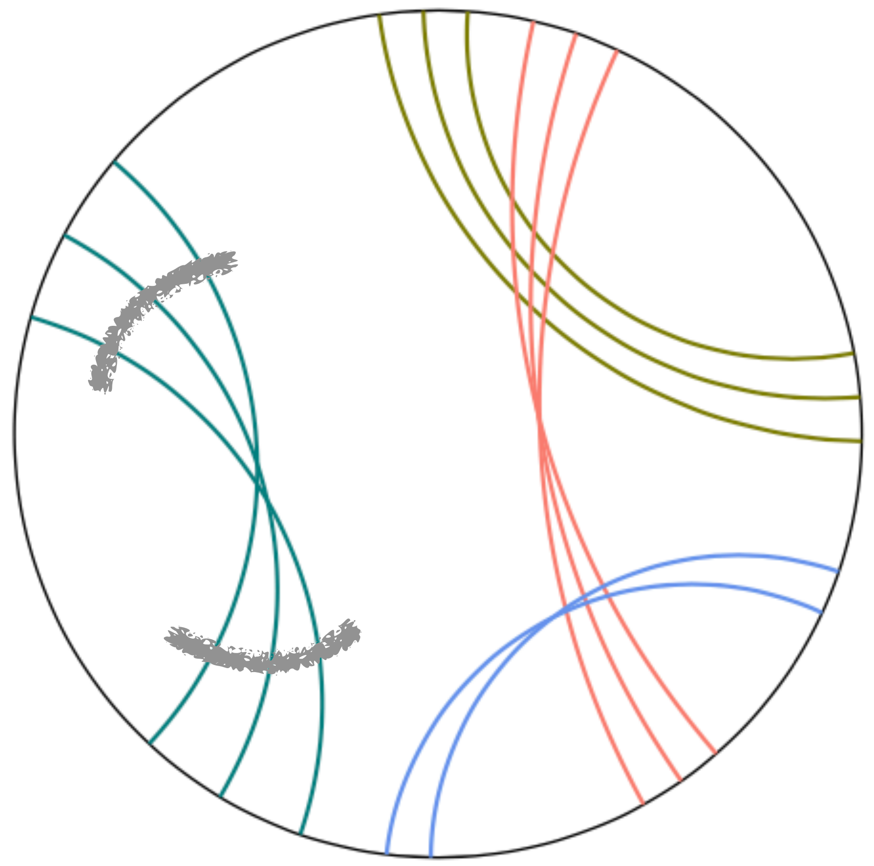
$$\{1_e\overline{2}_o12\}$$

contraction is always possible provided there are no ribbons in the way…..



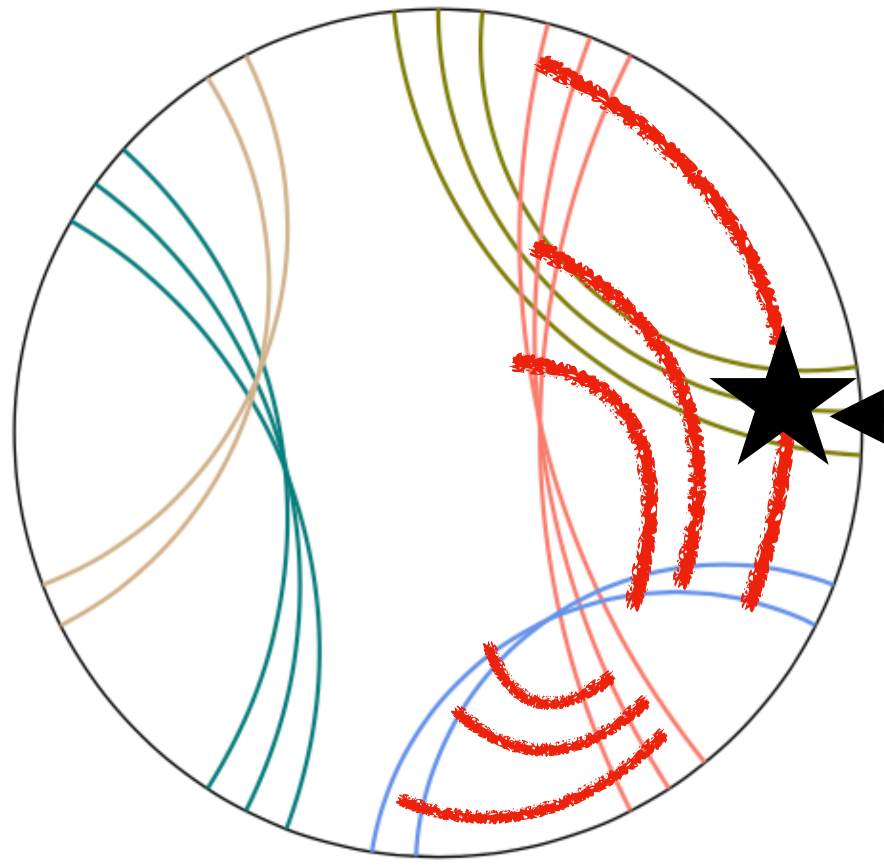
$$\{1_o\overline{2}_o1\overline{3}_e23\overline{4}_o\overline{5}_e45\}$$

$$\downarrow$$



$$\{1_o\overline{2}_o1\overline{3}_e23\overline{4}_o4\}$$

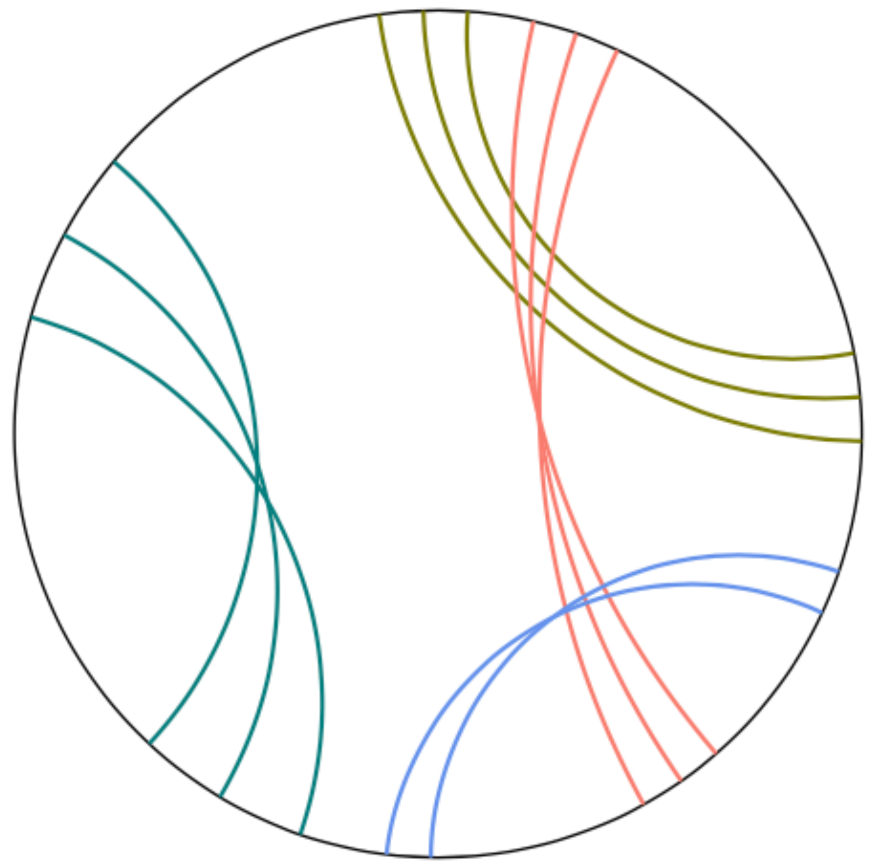
contraction is always possible provided there are no ribbons in the way.....



...cannot contract if perimeter is blocked

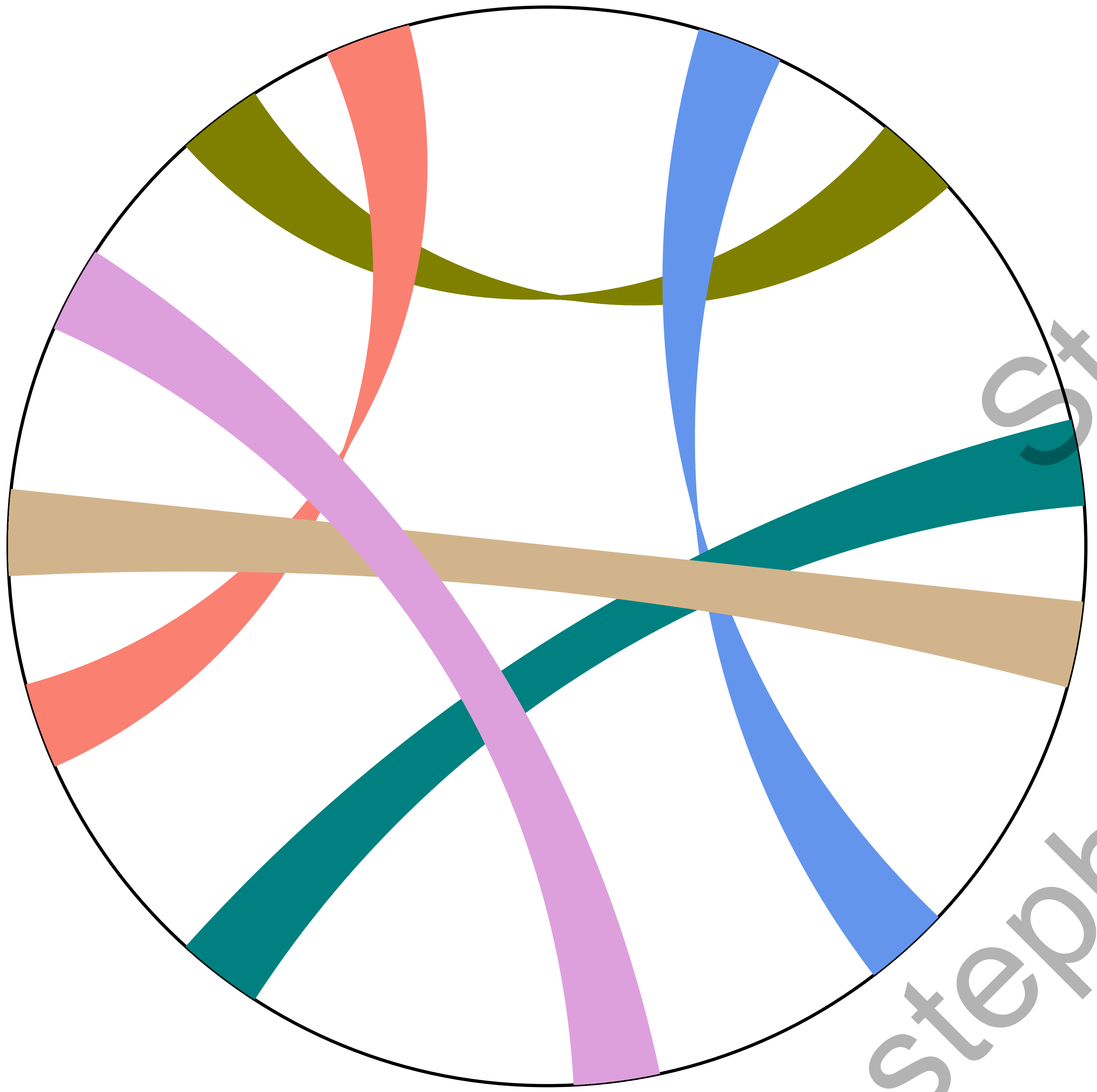
$$\{1_o\overline{2}_o1\overline{3}_e23\overline{4}_o\overline{5}_e45\}$$

$$\downarrow$$

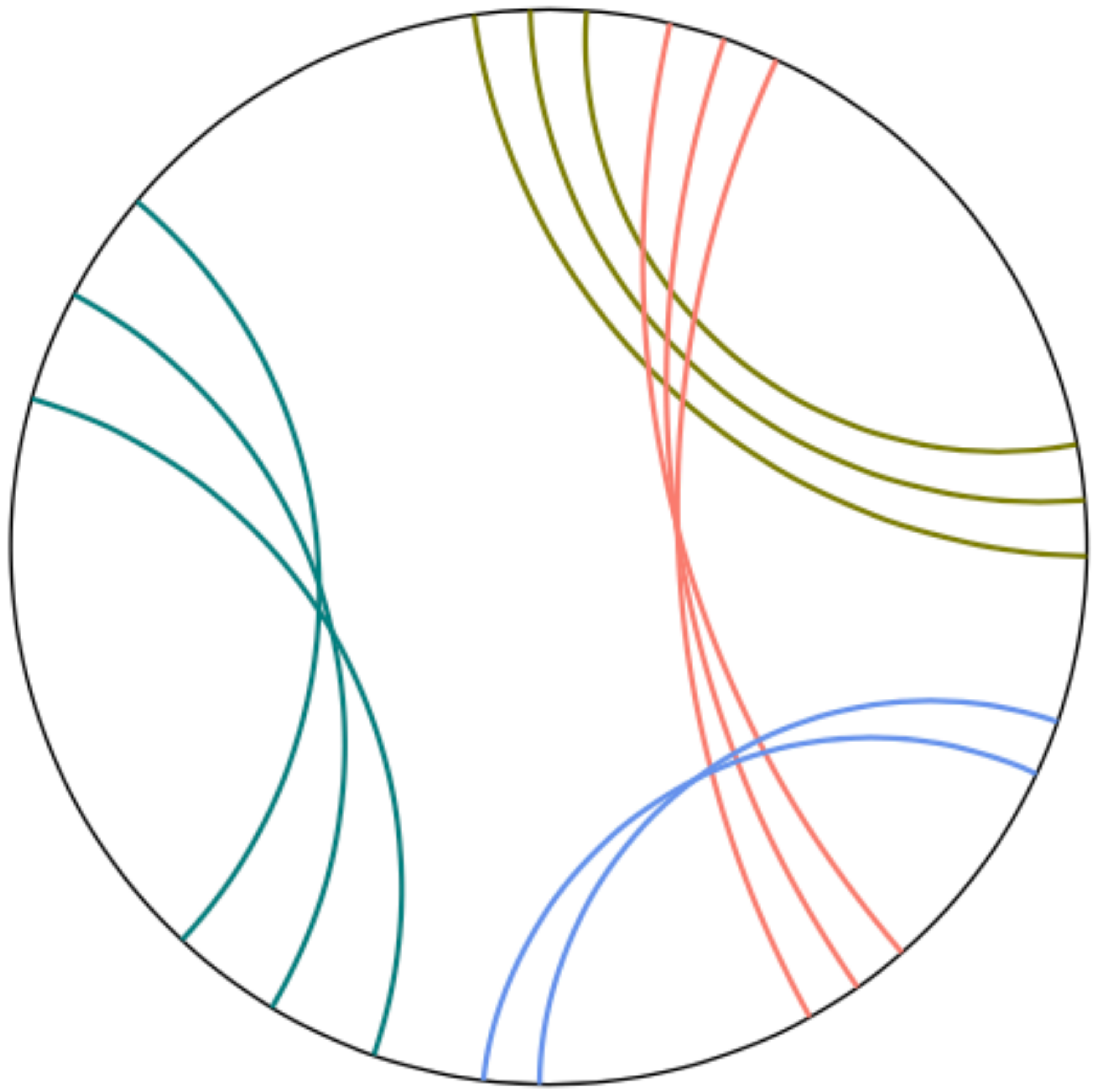


$$\{1_o\overline{2}_o1\overline{3}_e23\overline{4}_o4\}$$

any circular ribbon diagram with annular and/or moebius ribbons describes a duplexed fold



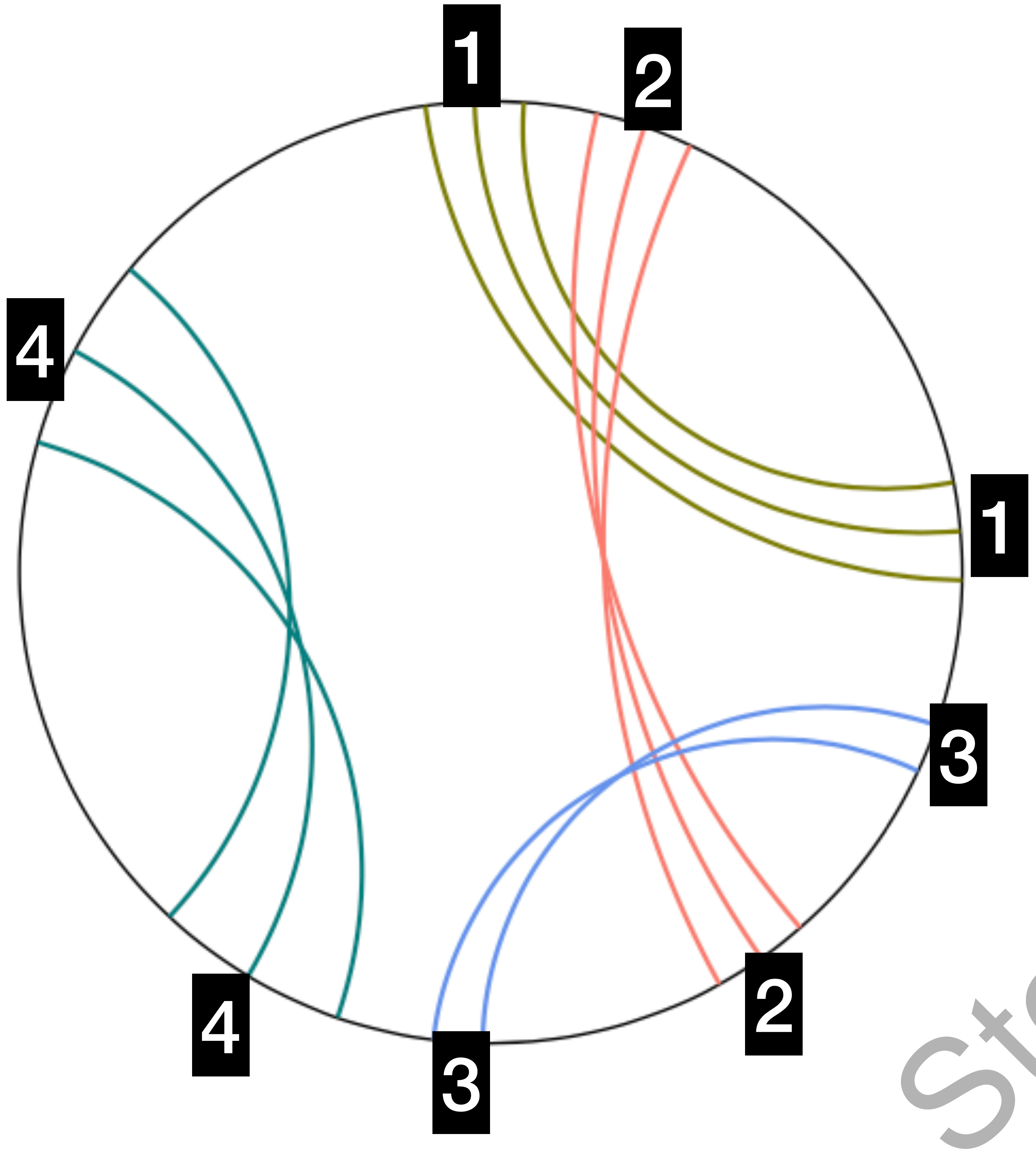
contract all nested annular ribbons and fans of moebius ribbons to form contracted fold



every circular diagram has a canonical fully-flagged FOLD LABEL

$$\mathcal{L}_{\Pi}^{\otimes}$$

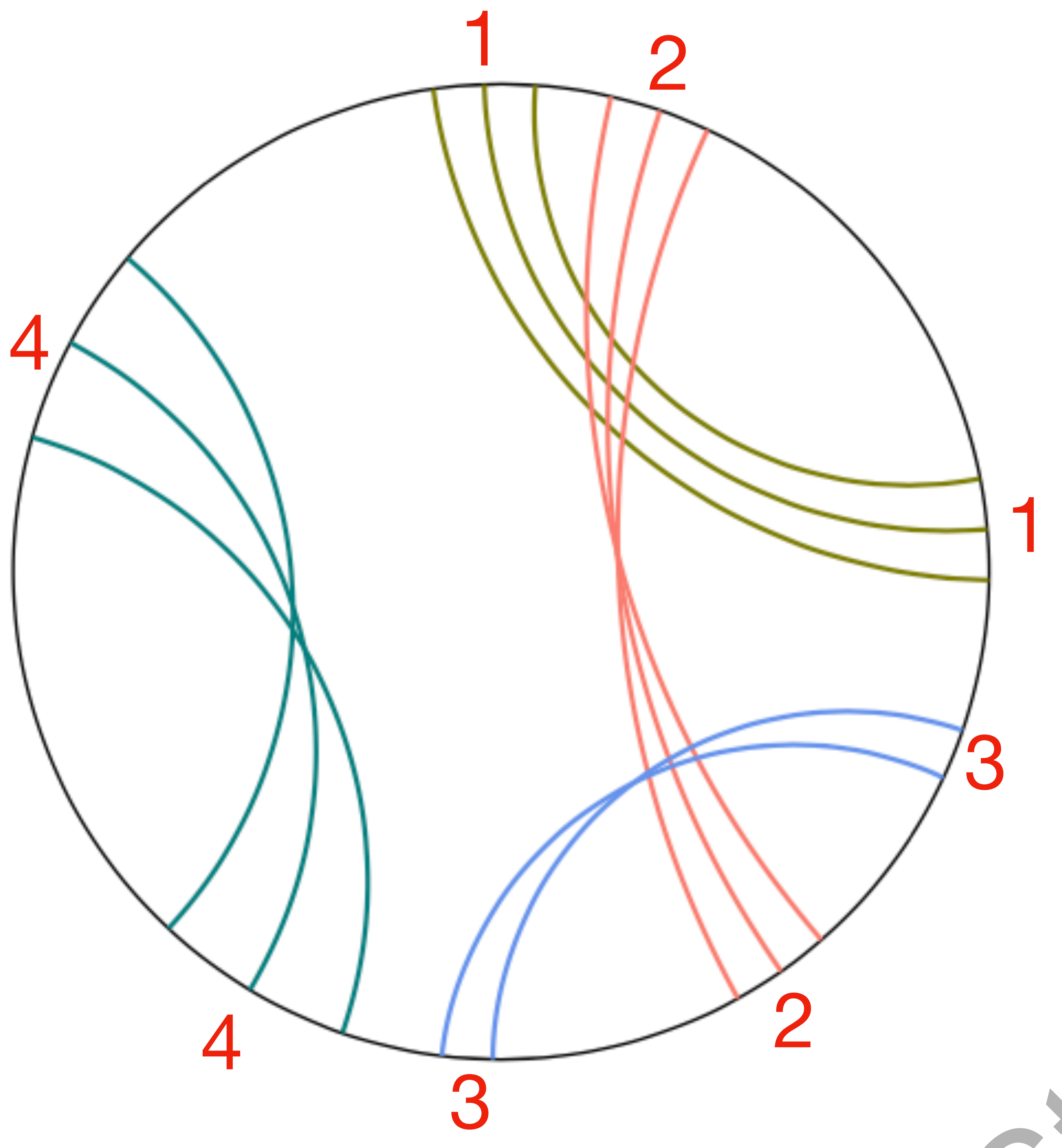
$$\{1_o\bar{2}_o1\bar{3}_e23\bar{4}_o4\}$$



label each ribbon 1,2,3,…

$$\{1_o\bar{2}_o1\bar{3}_e23\bar{4}_o4\}$$

bare fold label $\mathcal{L}^{\otimes}_{\Pi}$

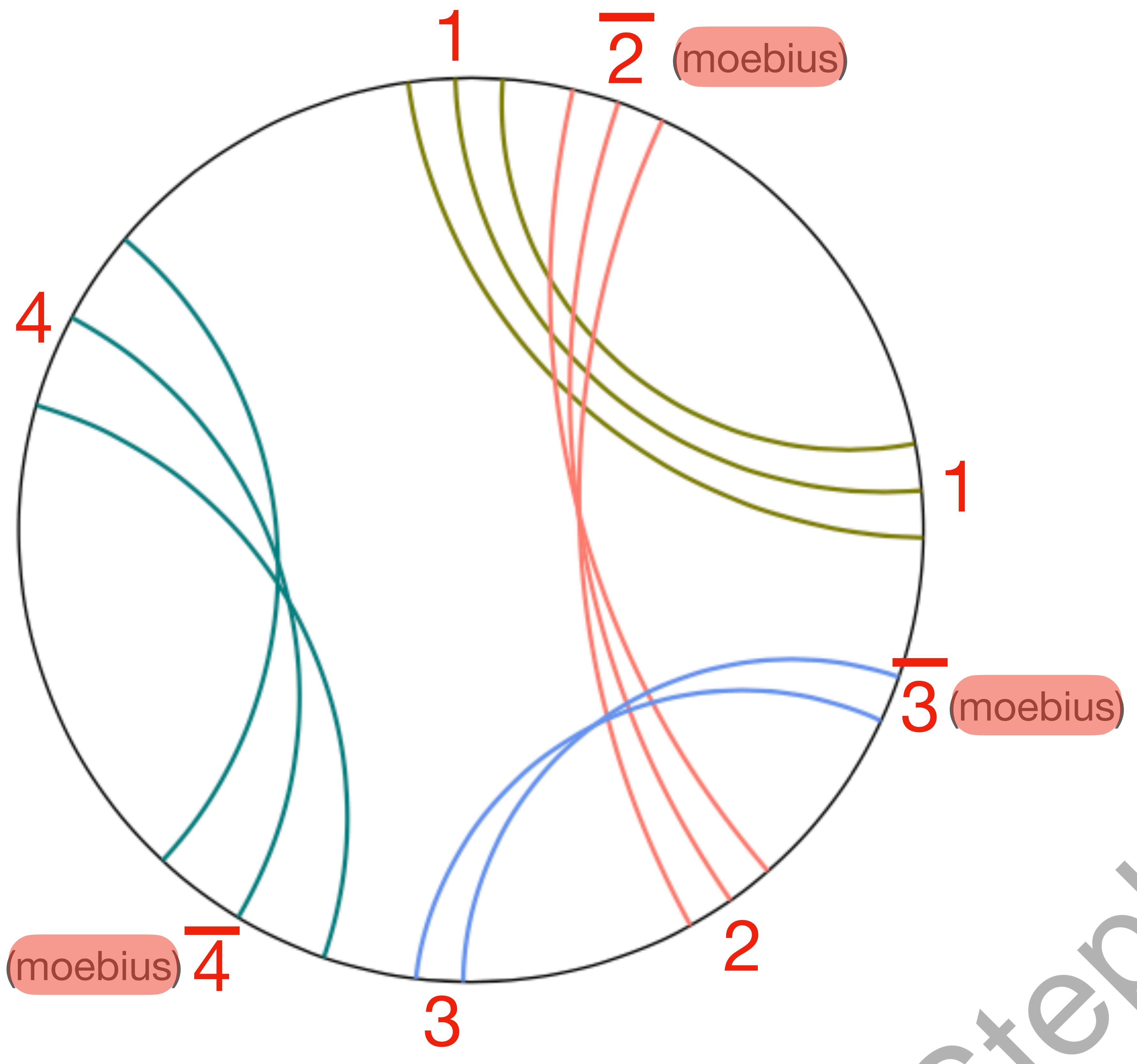


every circular diagram has a canonical fully-flagged FOLD LABEL

bare fold label $\mathcal{L}^{\otimes}_{\Pi}$

canonical label: order all possible clockwise, anticlockwise circuits around perimeter, choose smallest label

$$\{1_o\bar{2}_o1\bar{3}_e23\bar{4}_o4\}$$

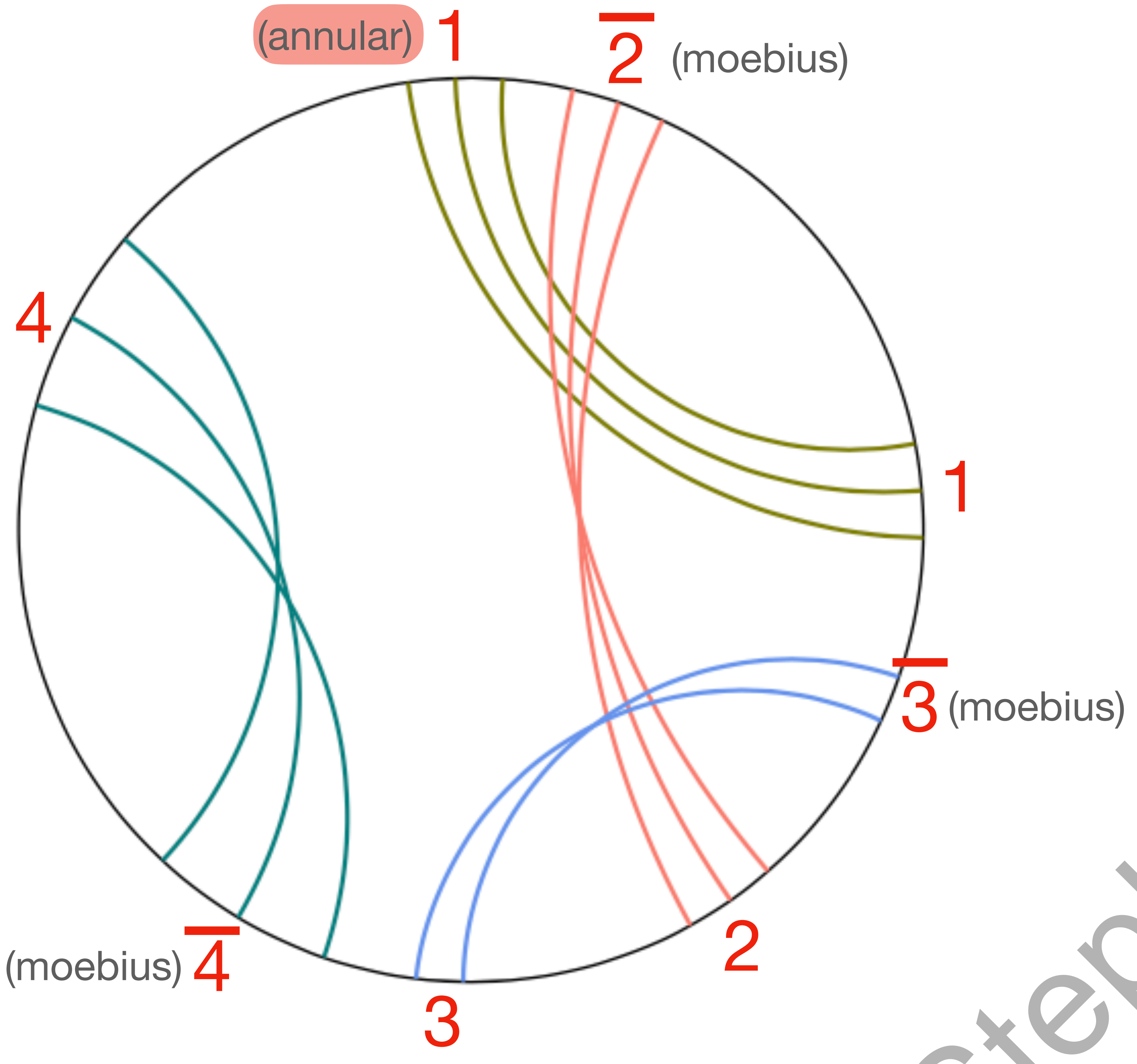


every circular diagram has a canonical fully-flagged FOLD LABEL

$$\mathcal{L}_{\Pi}^{\otimes}$$

annular/moebius flag
(anti/parallel duplex)

$$\{1_o\overline{2}_o 1\overline{3}_e 23\overline{4}_o 4\}$$

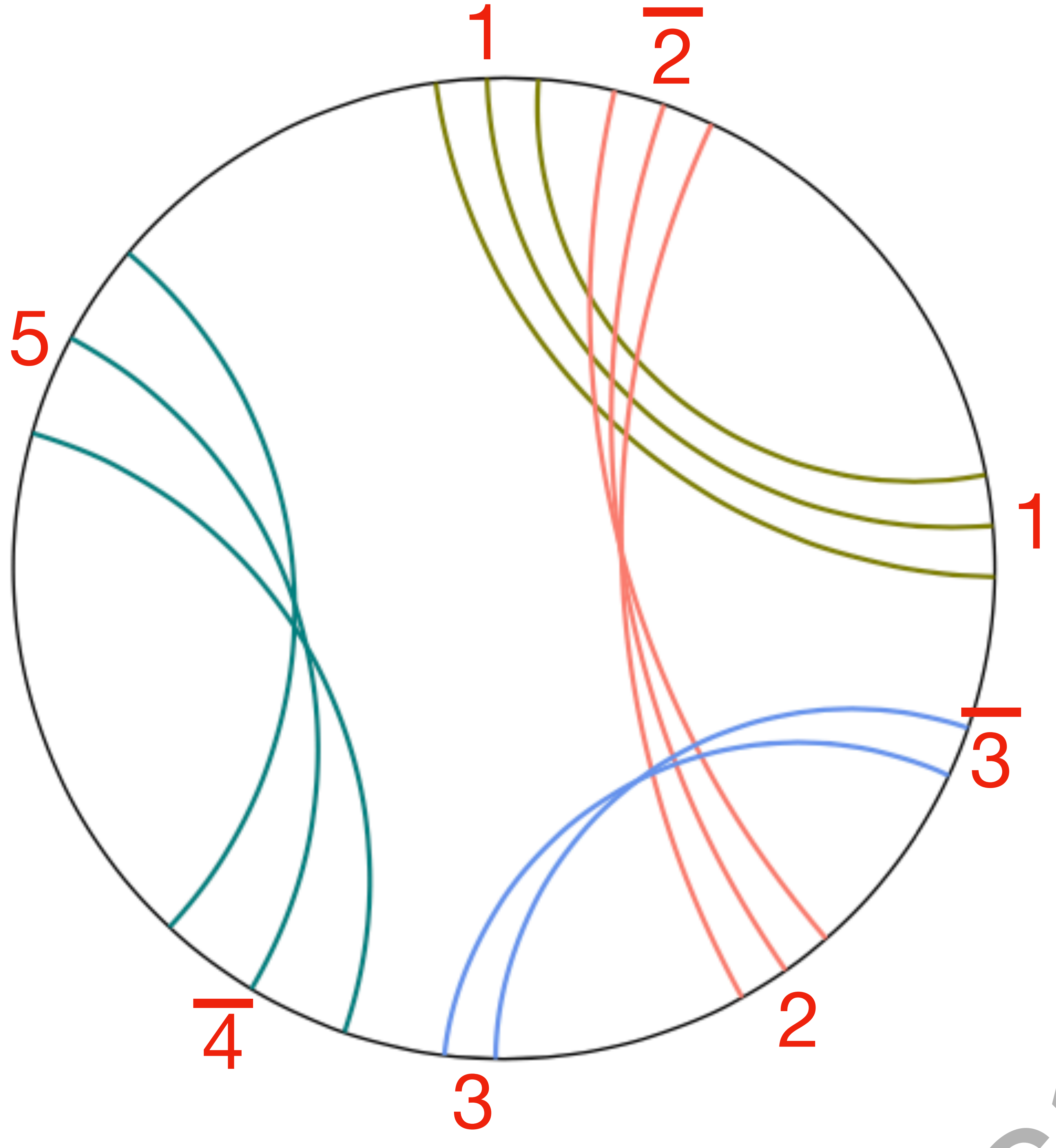


every circular diagram has a canonical fully-flagged FOLD LABEL

$$\mathcal{L}_{\Pi}^{\otimes}$$

annular/moebius flag

$$\{1_o\bar{2}_o1\bar{3}_e23\bar{4}_o4\}$$



every circular diagram has a canonical
fully-flagged FOLD LABEL

$$\mathcal{L}^{\otimes}_{\Pi}$$

parity flag
(even/odd half-twists in duplex)

$$\{1_o\bar{2}_o1\bar{3}_e23\bar{4}_o4\}$$

Any digit string with each digit listed twice describes an unflagged (likely uncontracted) fold

$$\mathcal{L} := \{1122\}$$

$$\mathcal{L} := \{123132\}$$

$$\mathcal{L} := \{121345326456\}$$

possibly UNCONTRACTED

Table 1: Distinct unflagged canonical fold labels, $\mathcal{L}$, for folds containing up to 5 duplexes.

n	*nmbr of labels*	*smallest* $\mathcal{L}$	*largest* $\mathcal{L}$
0	1	{0}	{0}
1	1	{11}	{11}
2	2	{1122}	{1212}
3	5	{112233}	{123123}
4	17	{11223344}	{12341234}
5	79	{1122334455}	{1234512345}

Any combination of **orientation** flags is allowed:

$$\mathcal{L}_{\bullet}^{\otimes} = \{\overline{1}21345326456\}$$

$$\mathcal{L}_{\bullet}^{\otimes} = \{1\overline{2}1345326456\}$$

$$\vdots$$

$$\mathcal{L}_{\bullet}^{\otimes} = \{\overline{12}1345326456\}$$

$$\vdots$$

$$\mathcal{L}_{\bullet}^{\otimes} = \{\overline{12}1\overline{345}32\overline{6}456\}$$

Any combination of **parity** flags is allowed:

$$\mathcal{L}_{\dot{\Pi}} = \{1_e 2_e 13_e 4_e 5_e 326_e 456\}$$

⋮

$$\mathcal{L}_{\dot{\Pi}} = \{1_o 2_e 13_e 4_o 5_e 326_o 456\}$$

⋮

Any flagged label with each digit listed twice is an allowed flagged (contracted) fold label

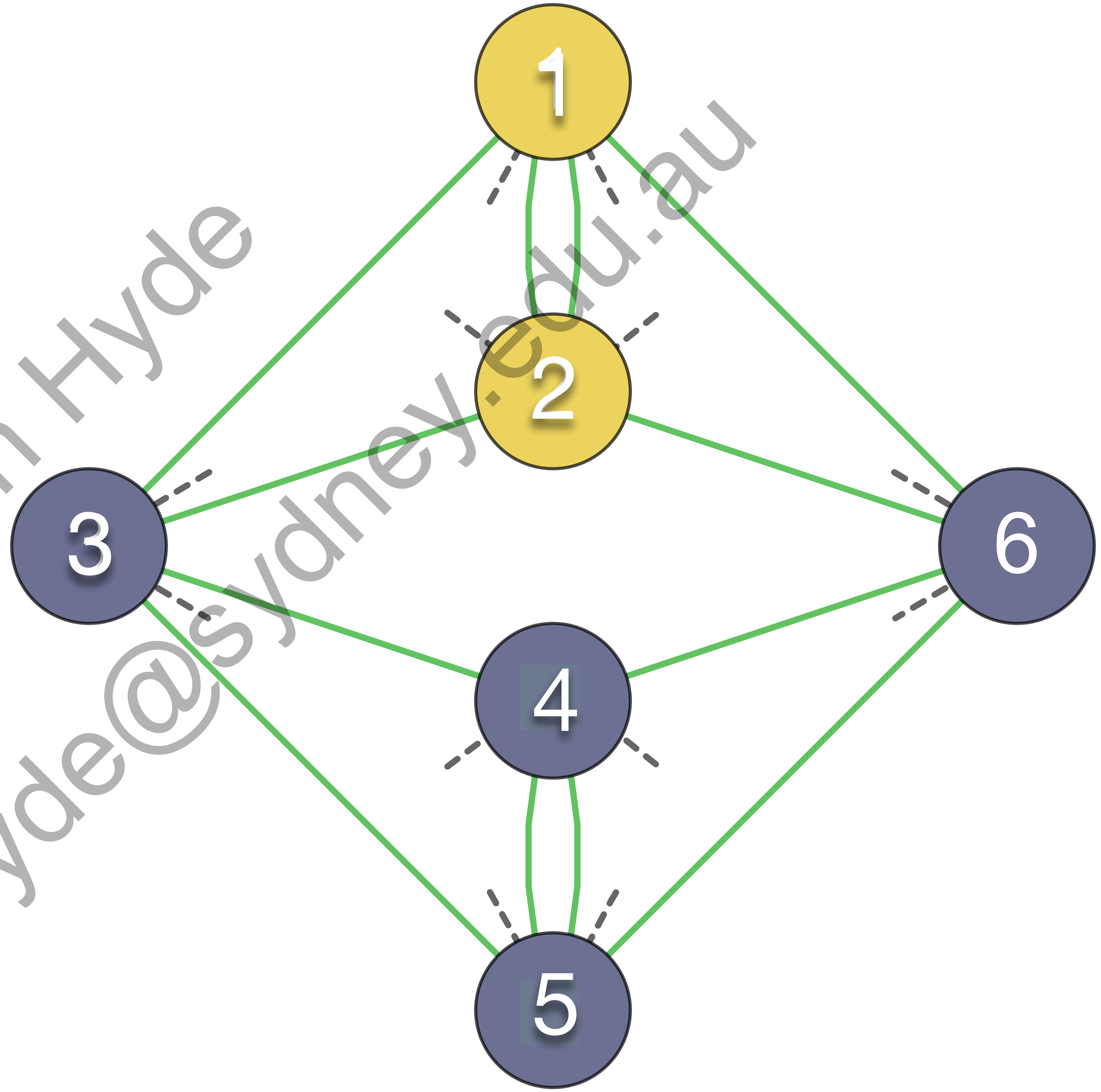
$$\mathcal{L}_{\Pi}^{\otimes} = \{1_o\overline{2}_e 13_e 4_o\overline{5}_e 326_o 456\}$$

Folds are equivalent if and only if they have identical contracted, fully-flagged canonical fold labels

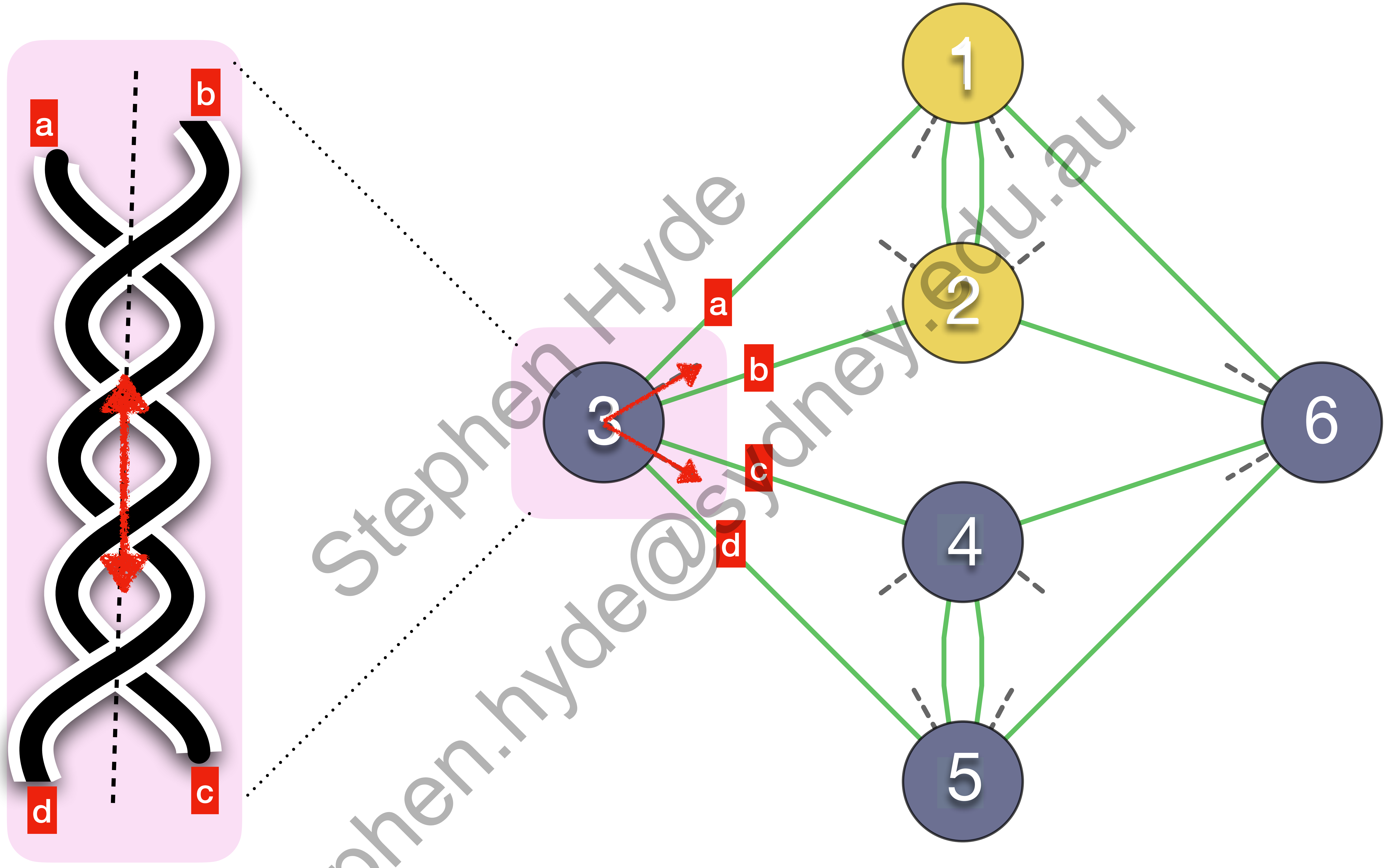
A fully-flagged canonical fold label uniquely defines a degree-4 'polarised strand graph'

$$\{\overline{1}_o\overline{2}_e13_e4_o5_o326_o456\}$$

THE FOLD

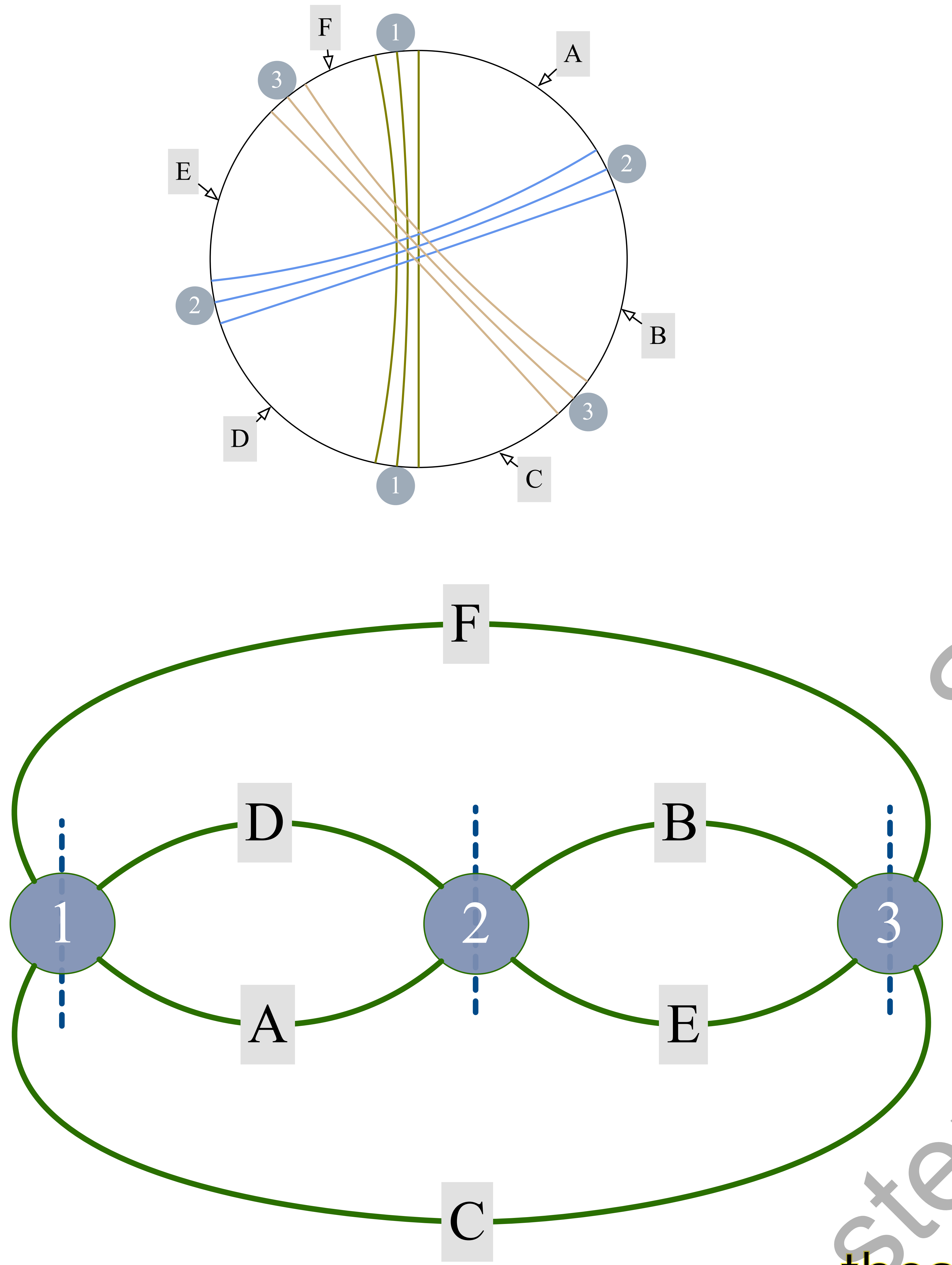


THE POLARISED (RIGID VERTEX) STRAND GRAPH each vertex hosts a double-helix, whose helical axis is set by the **plumbline polarisation**

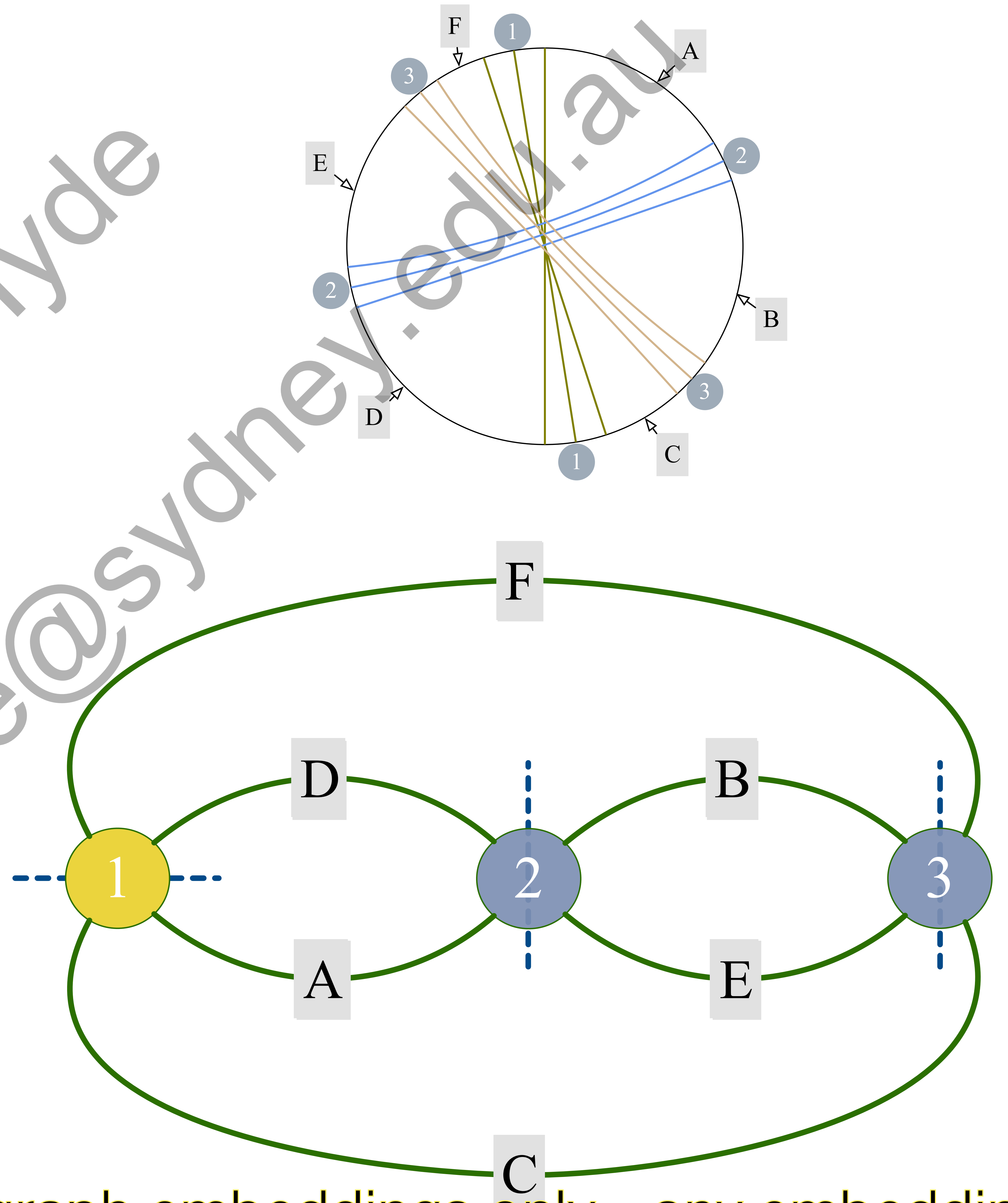


THE POLARISED STRAND GRAPH

$\{1_o 2_o 3_o 123\}$

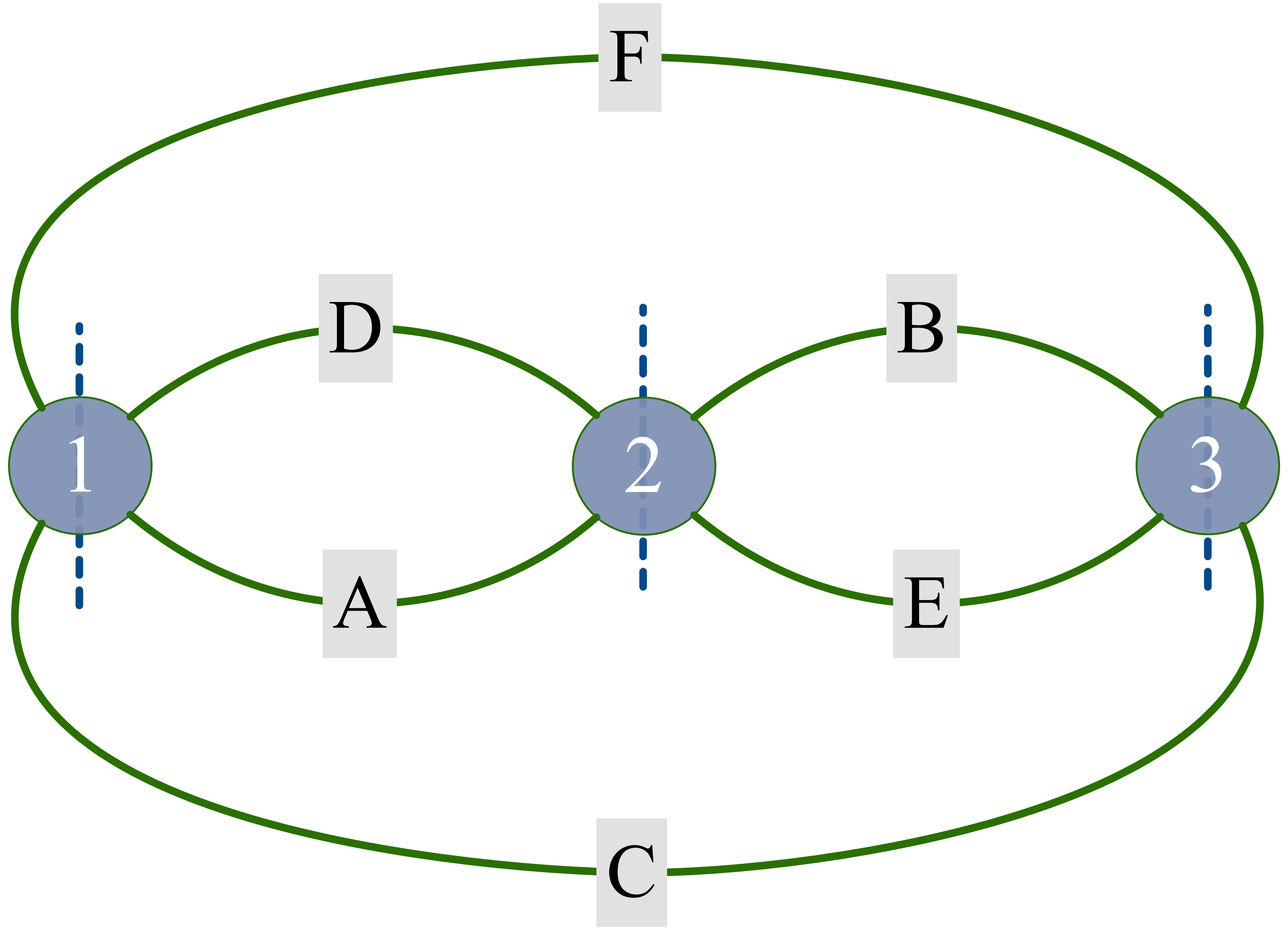


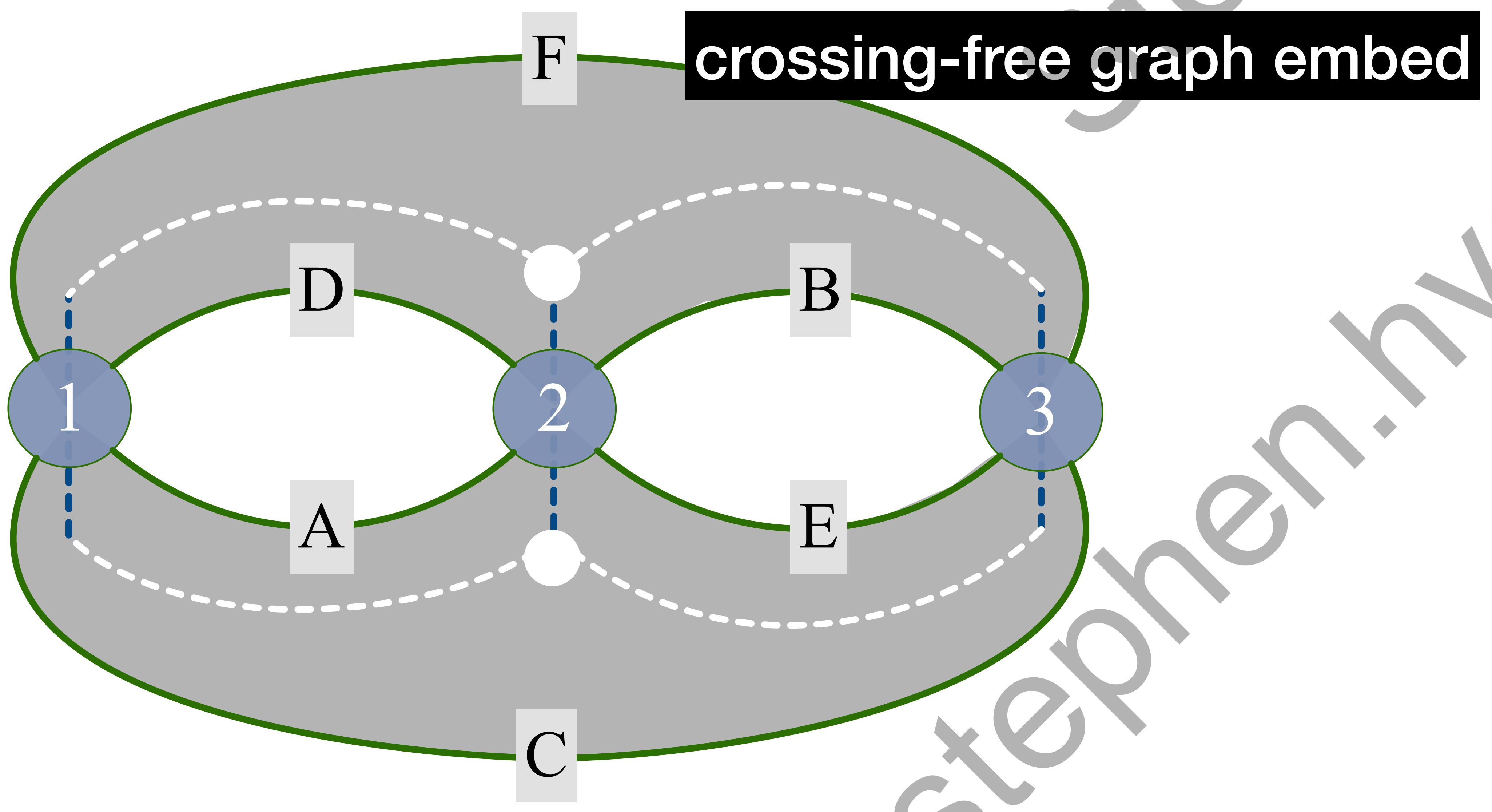
$\{\bar{1}_o 2_o 3_o 123\}$



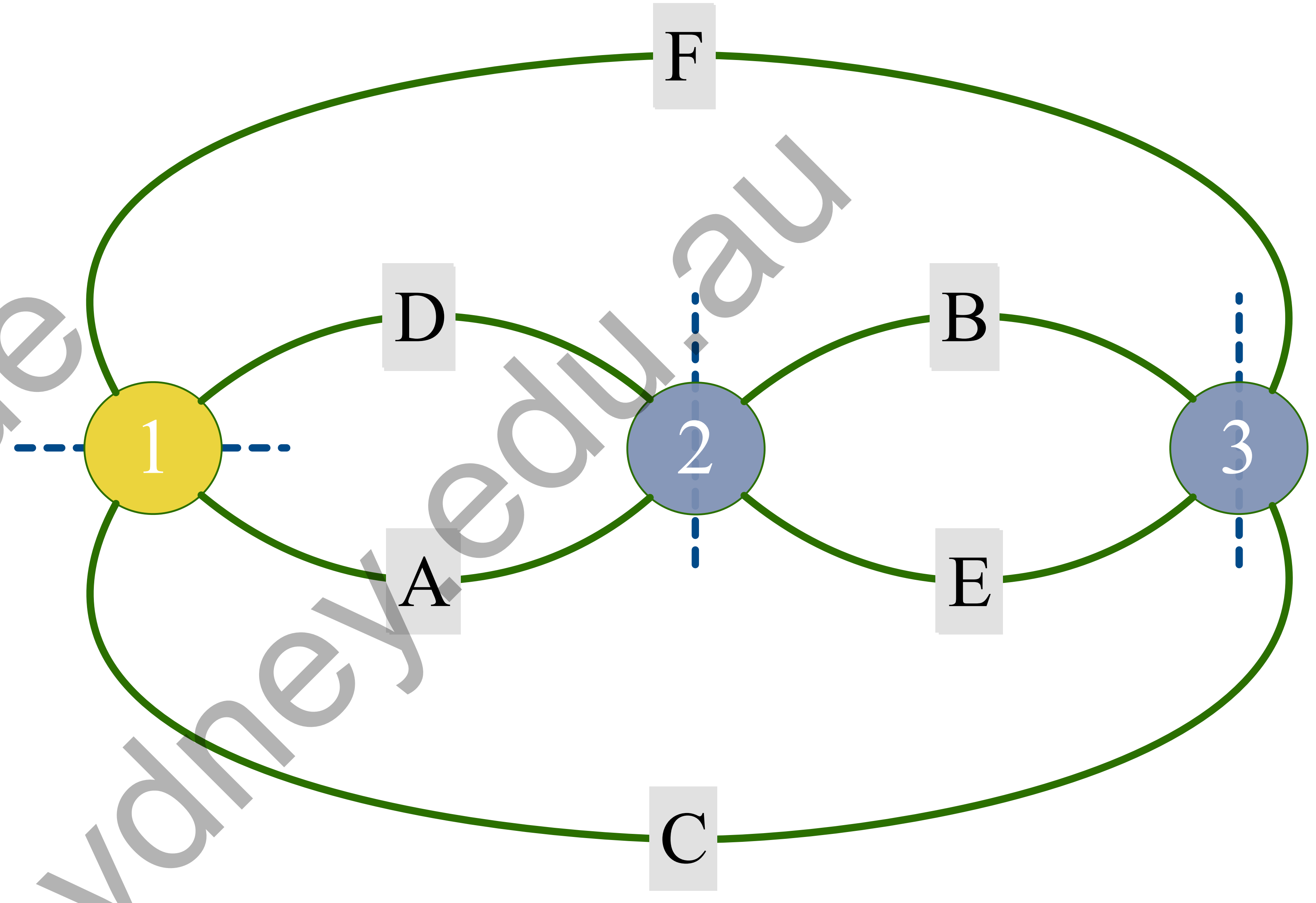
these are simplest graph embeddings only - *any* embedding is OK!

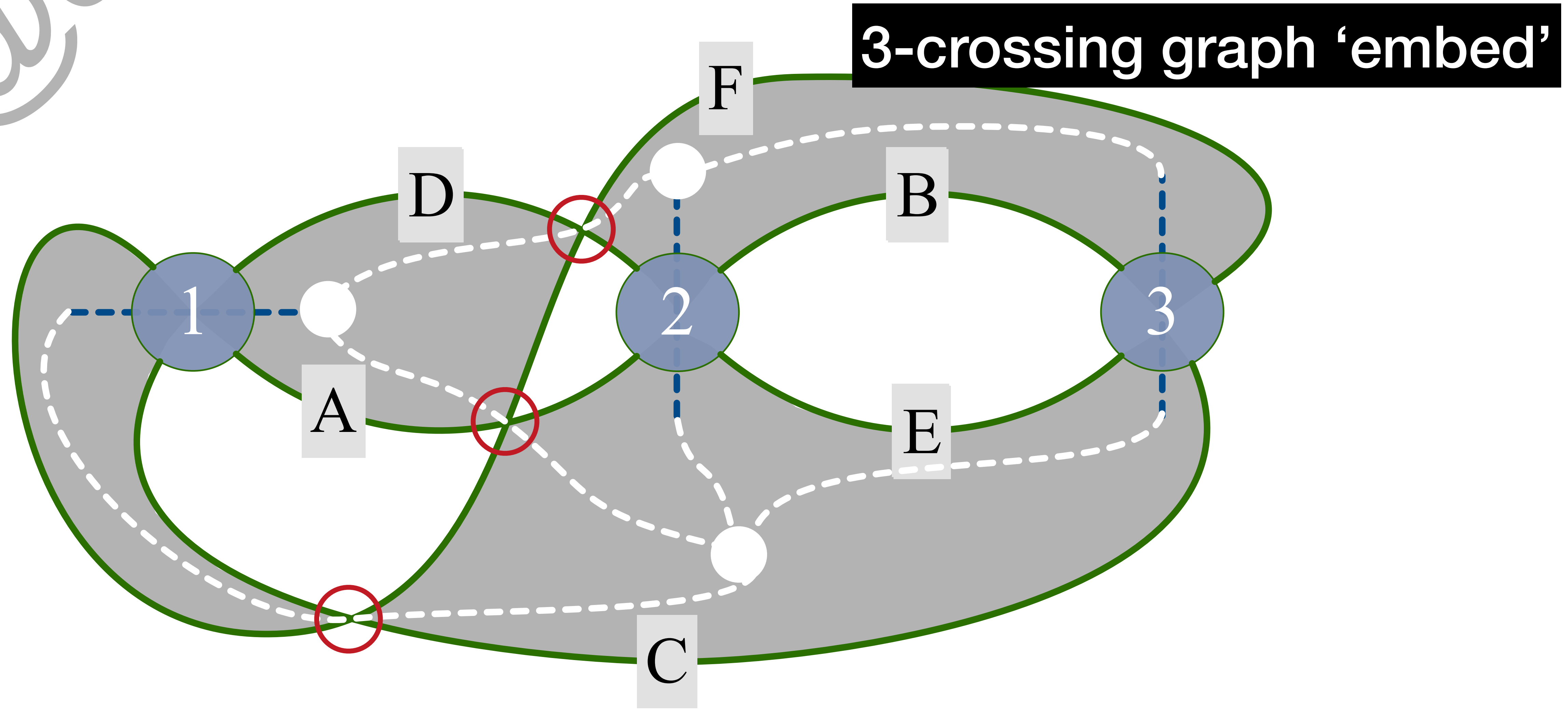
$\{1_o 2_o 3_o 123\}$



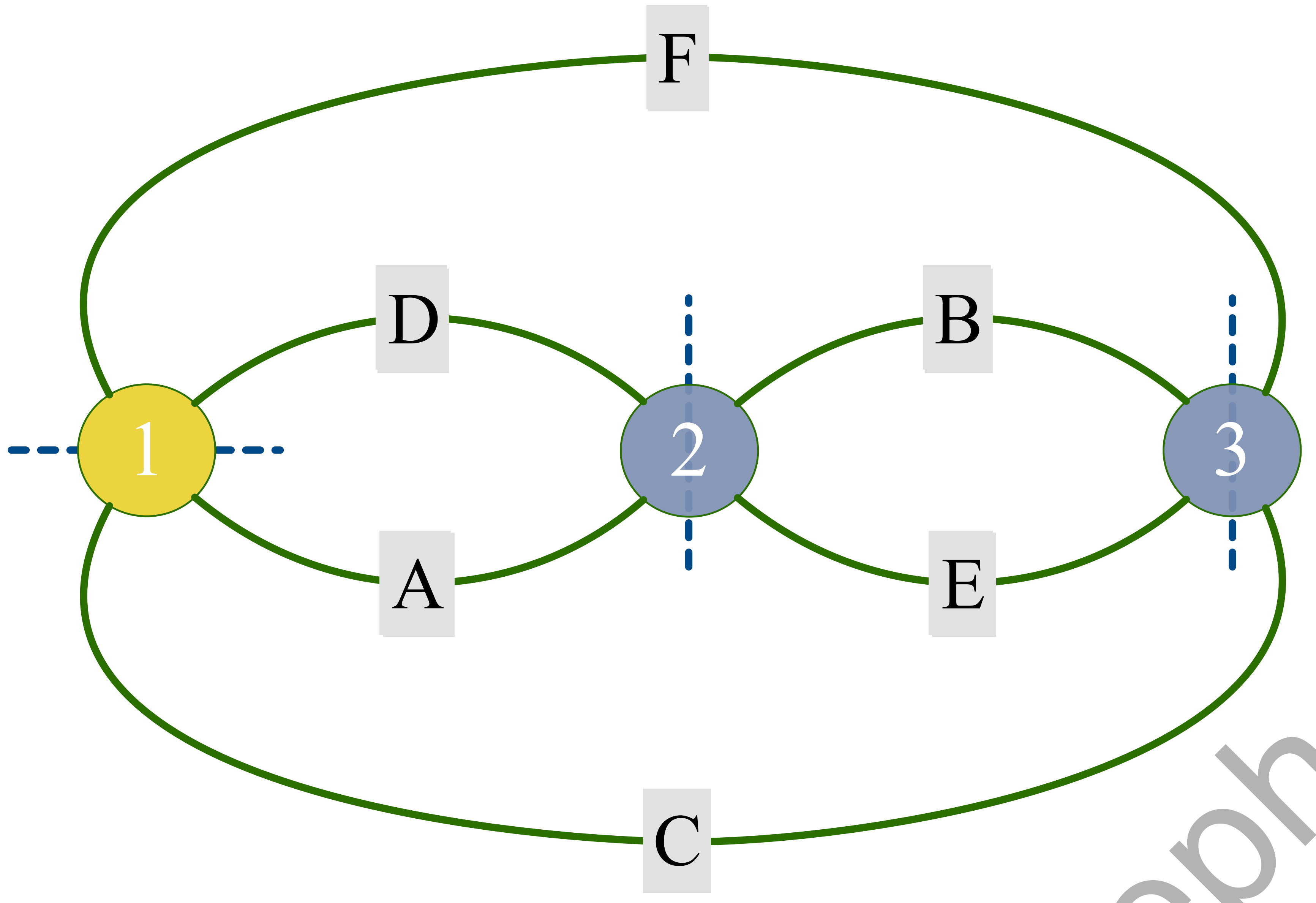


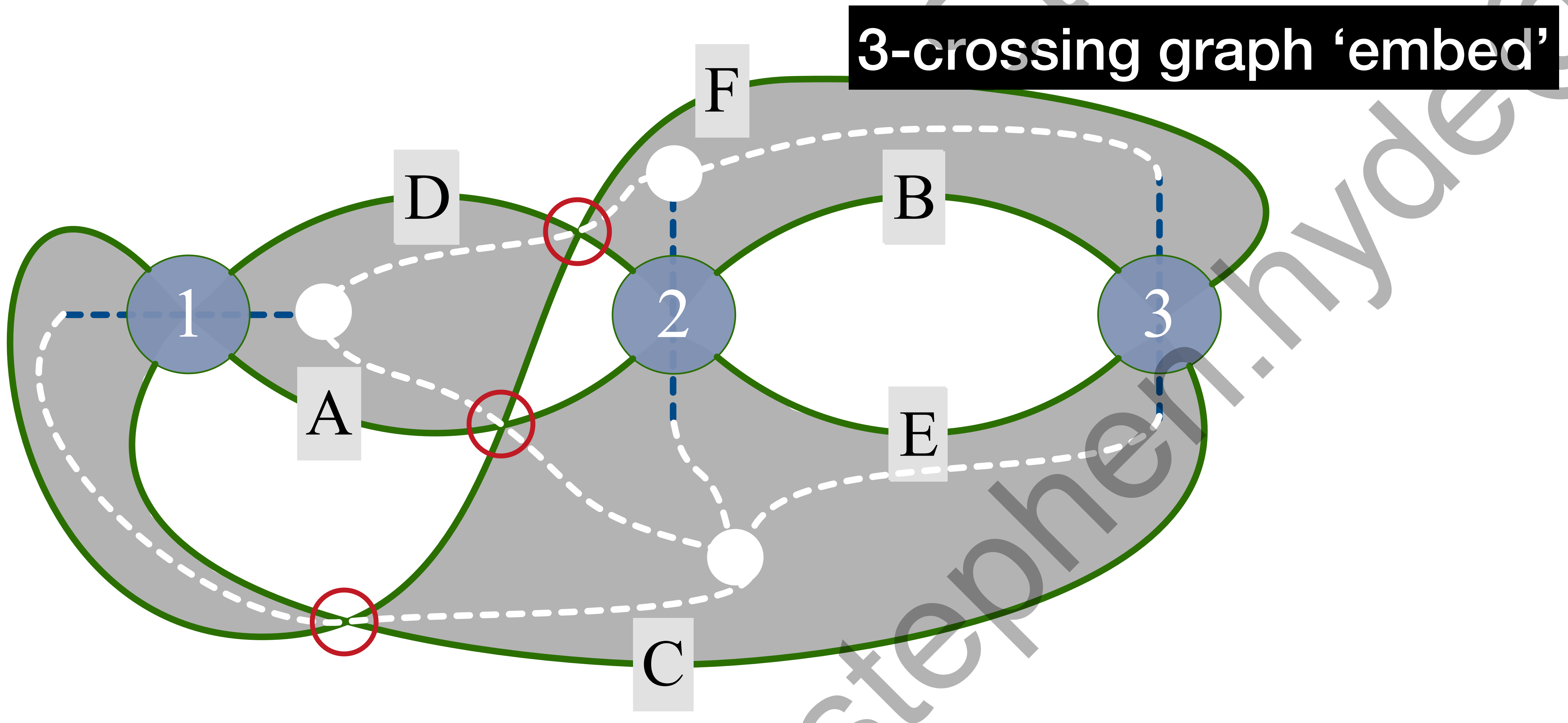
$\{\bar{1}_o 2_o 3_o 123\}$

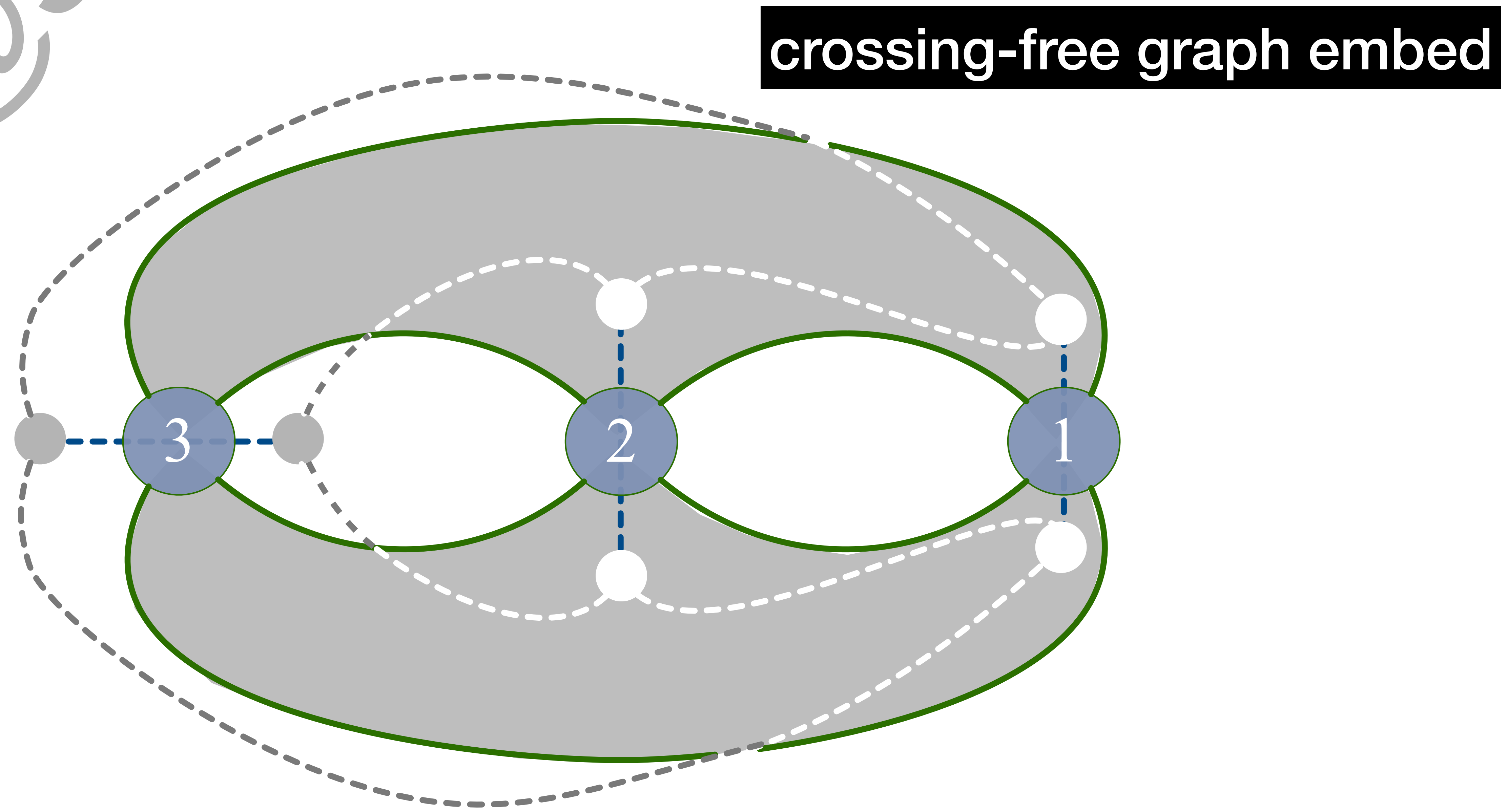




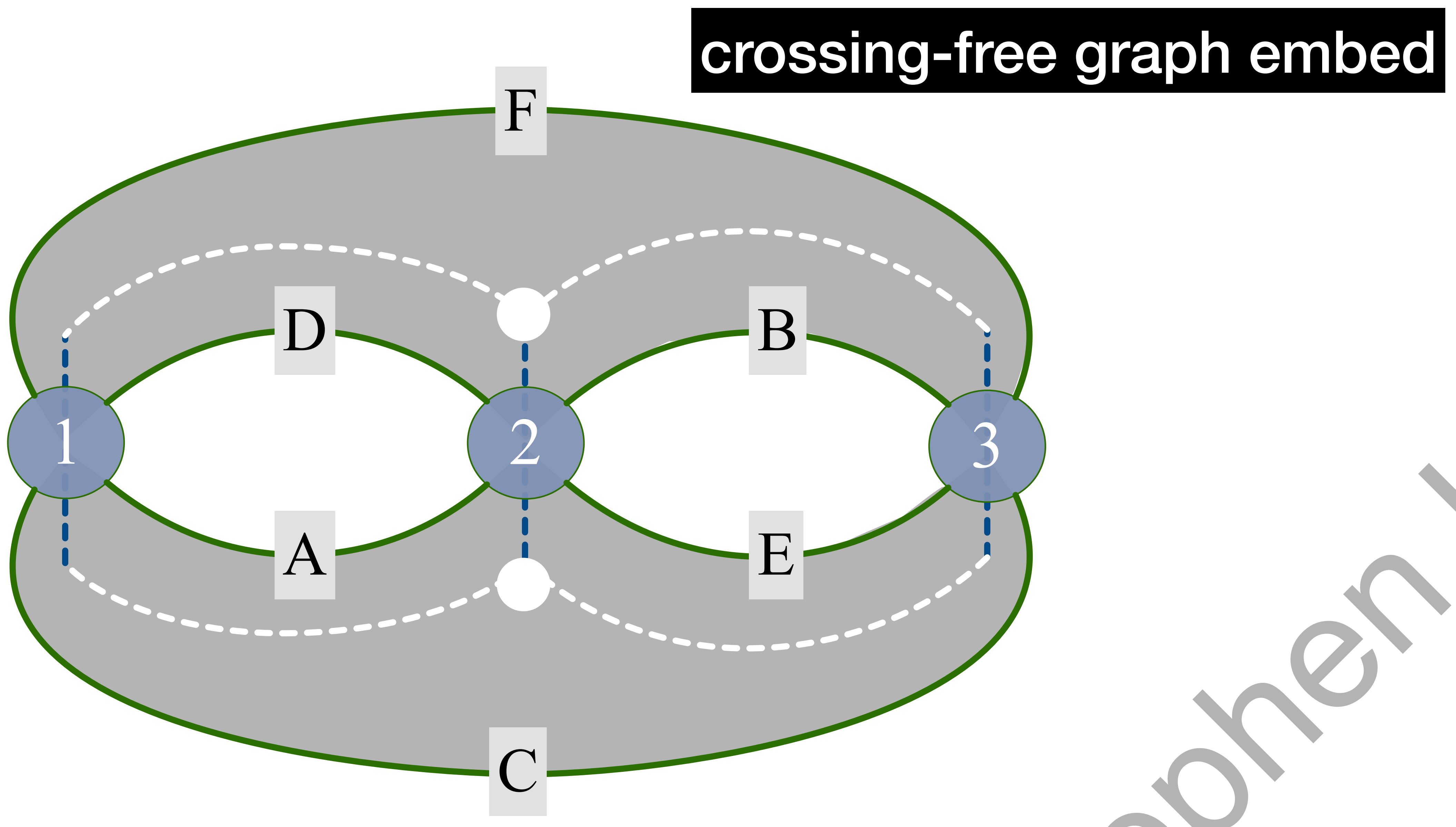
$$\{\overline{1}_o 2_o 3_o 123\}$$





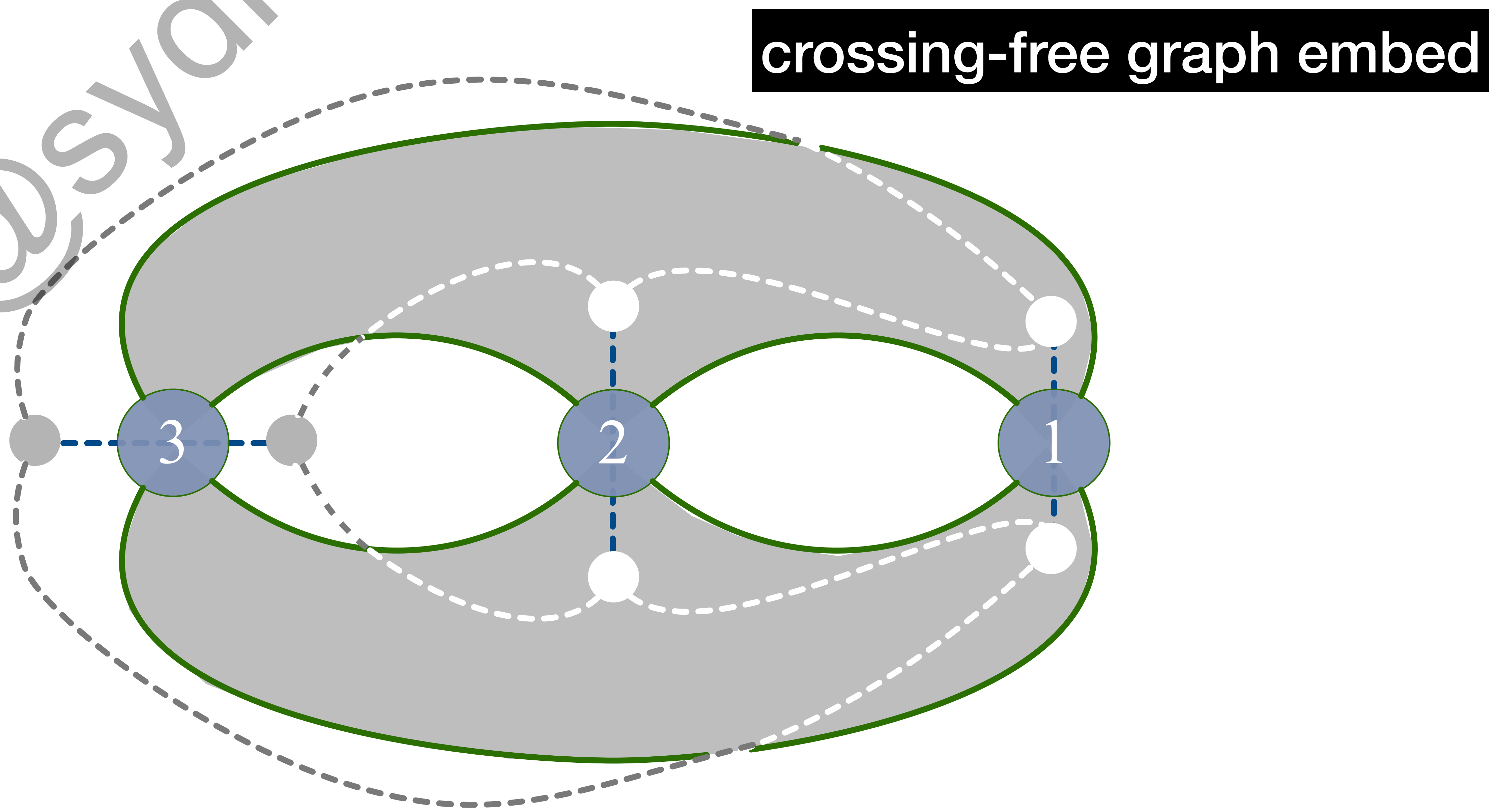


$$\{1_o 2_o 3_o 123\}$$



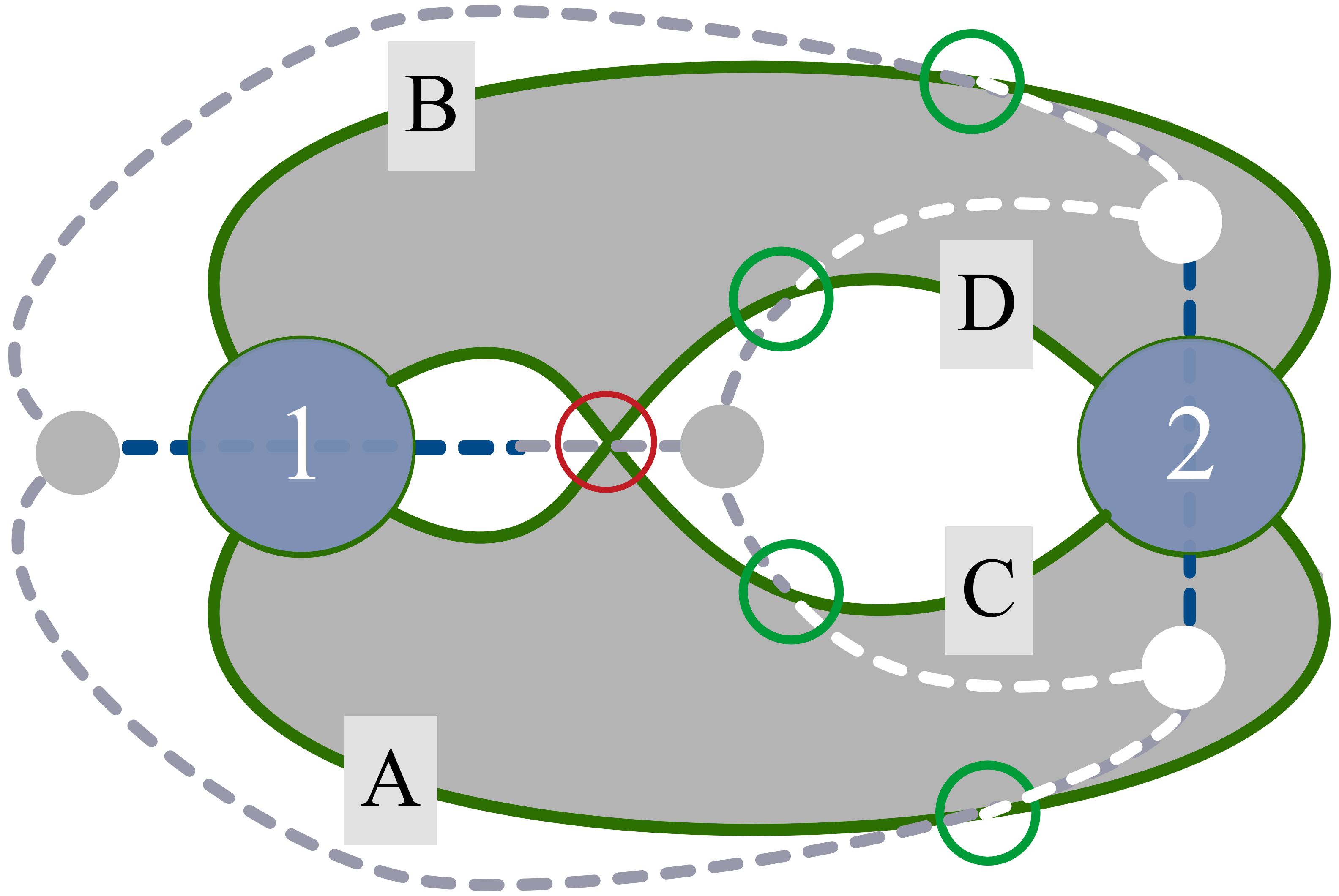
(a) uncrossed, one-colored embed

$$\{\bar{1}_o 2_o 3_o 123\}$$

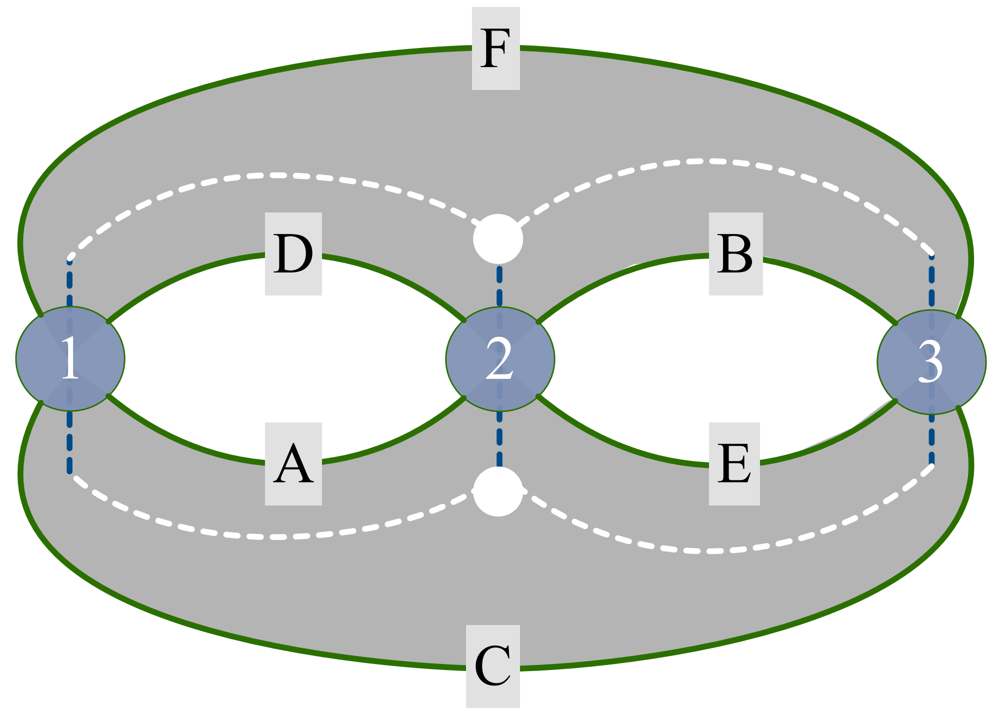


(b) uncrossed two-colored embed

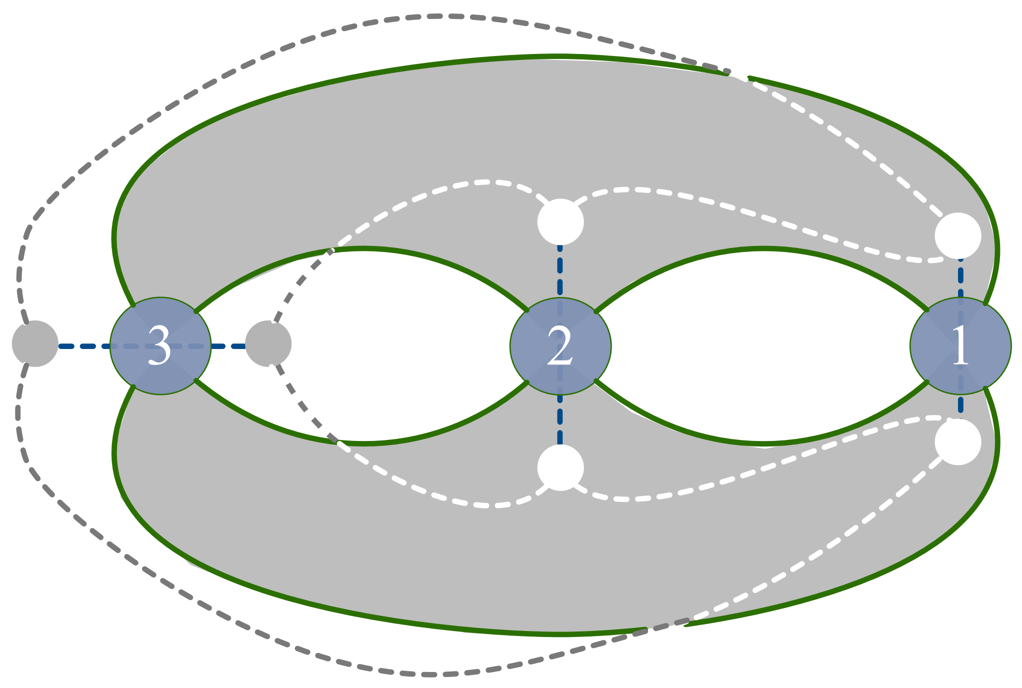
$$\{1_o 2_e 12\}$$



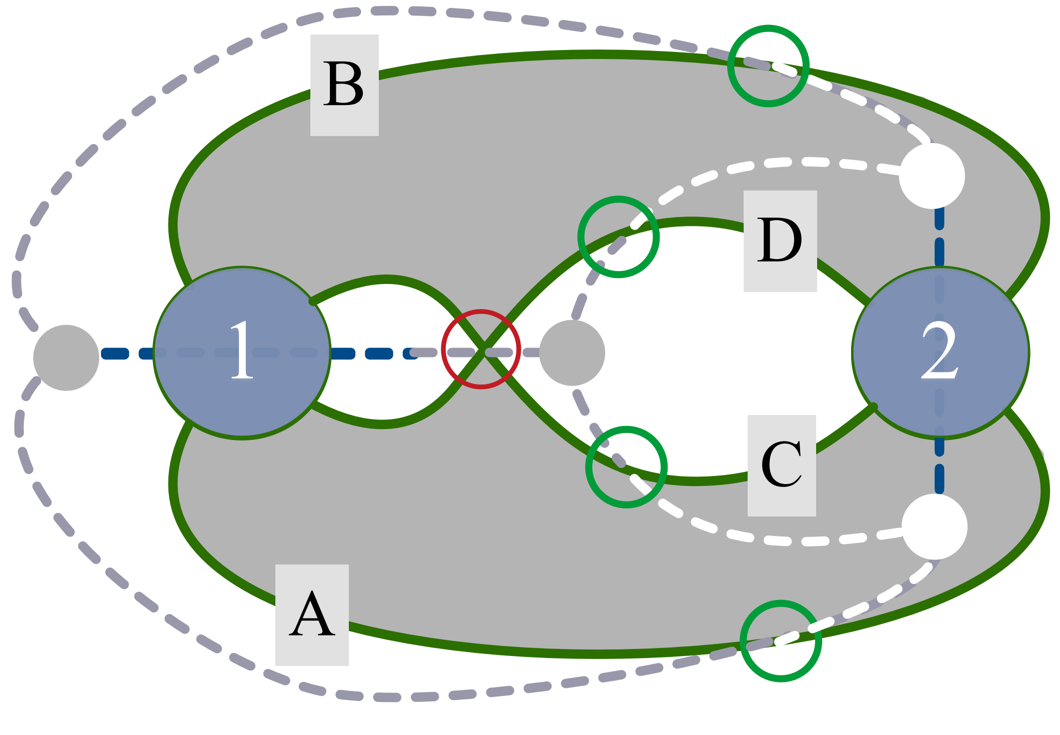
(c) all graph embeds contain edge-crossings



(a) uncrossed embed is one-colored



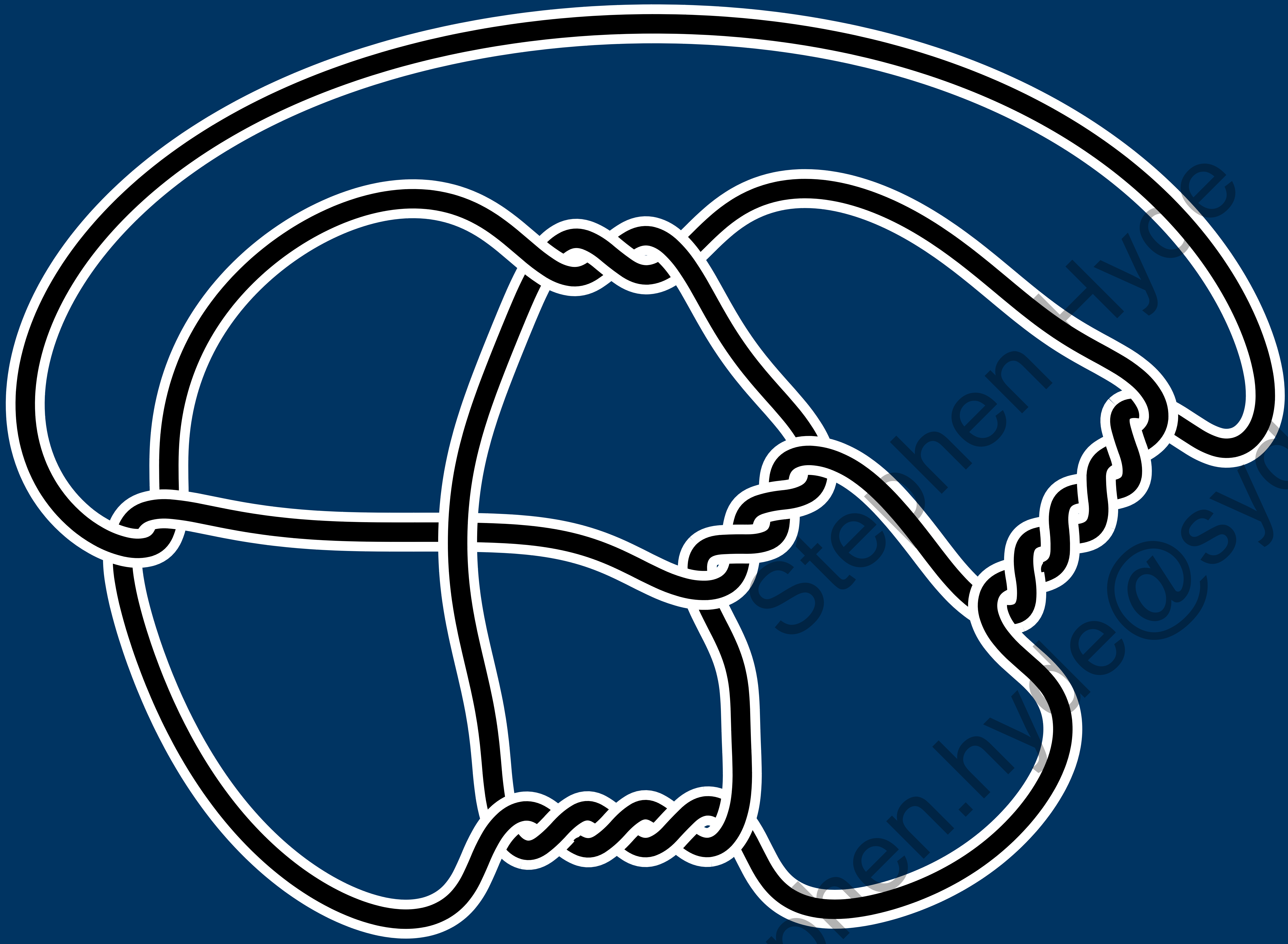
(b) uncrossed embed is two-colored

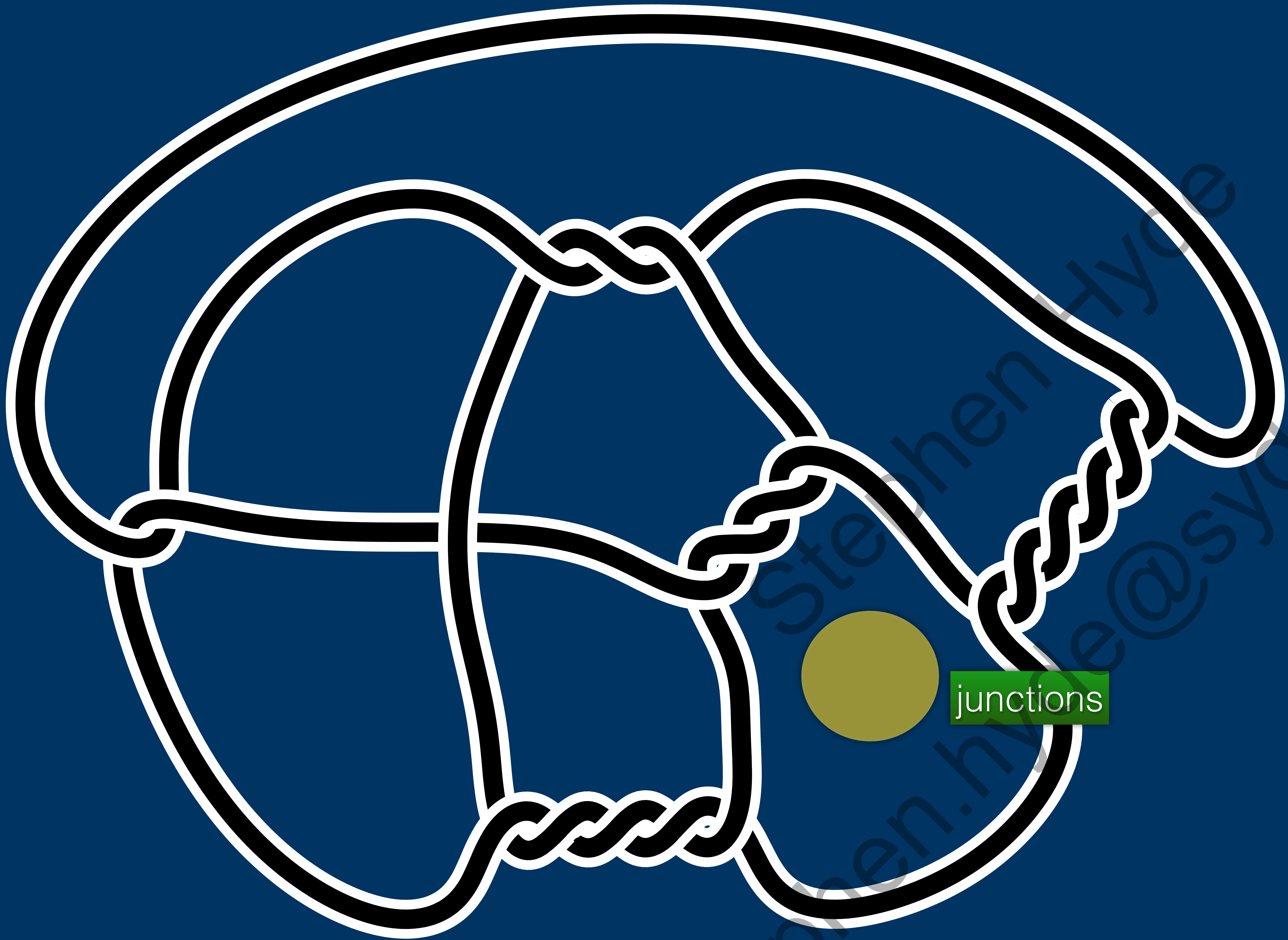


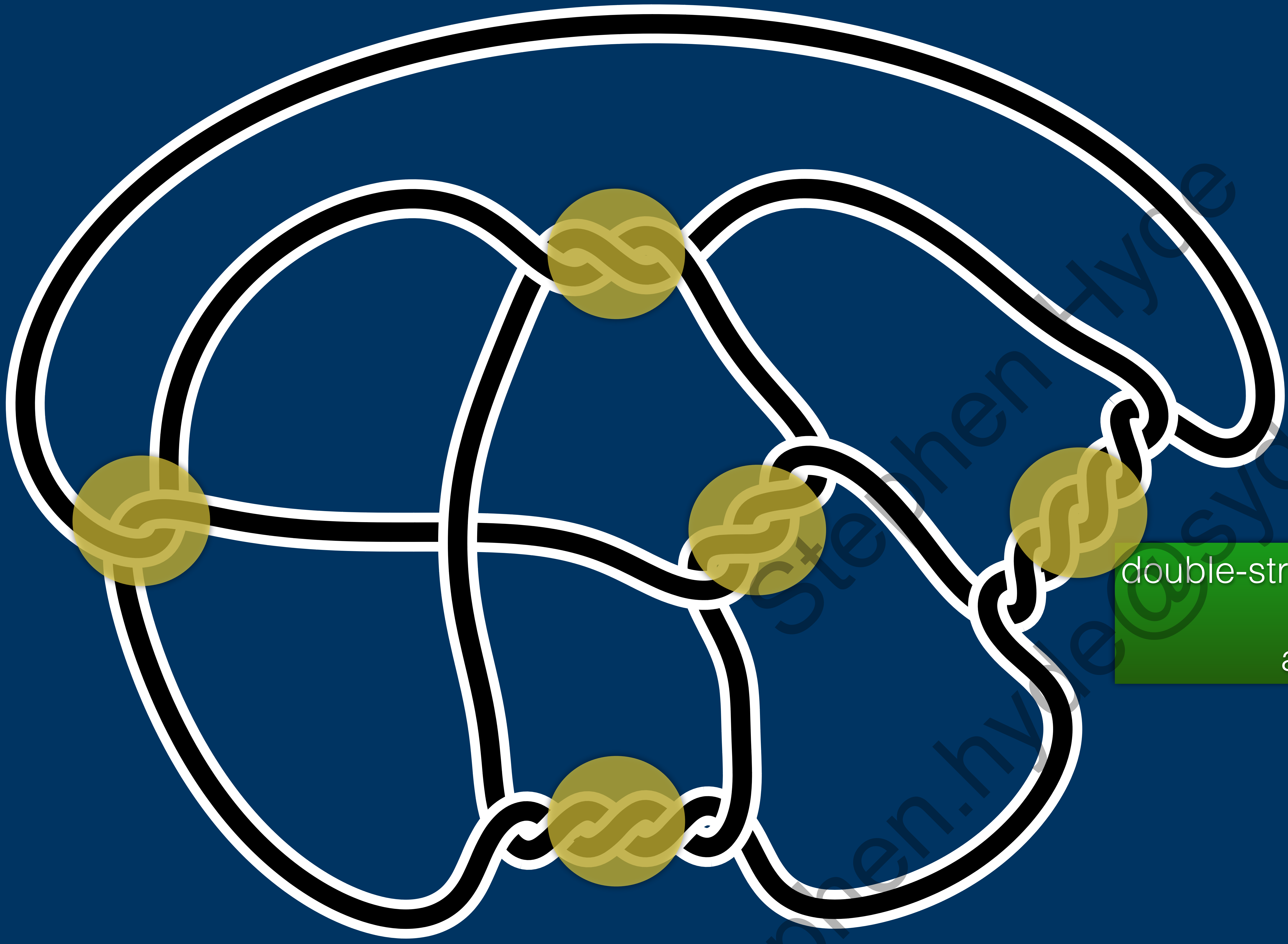
(c) all graph embeds contain edge-crossings

all folds are either (a), (b) or (c)

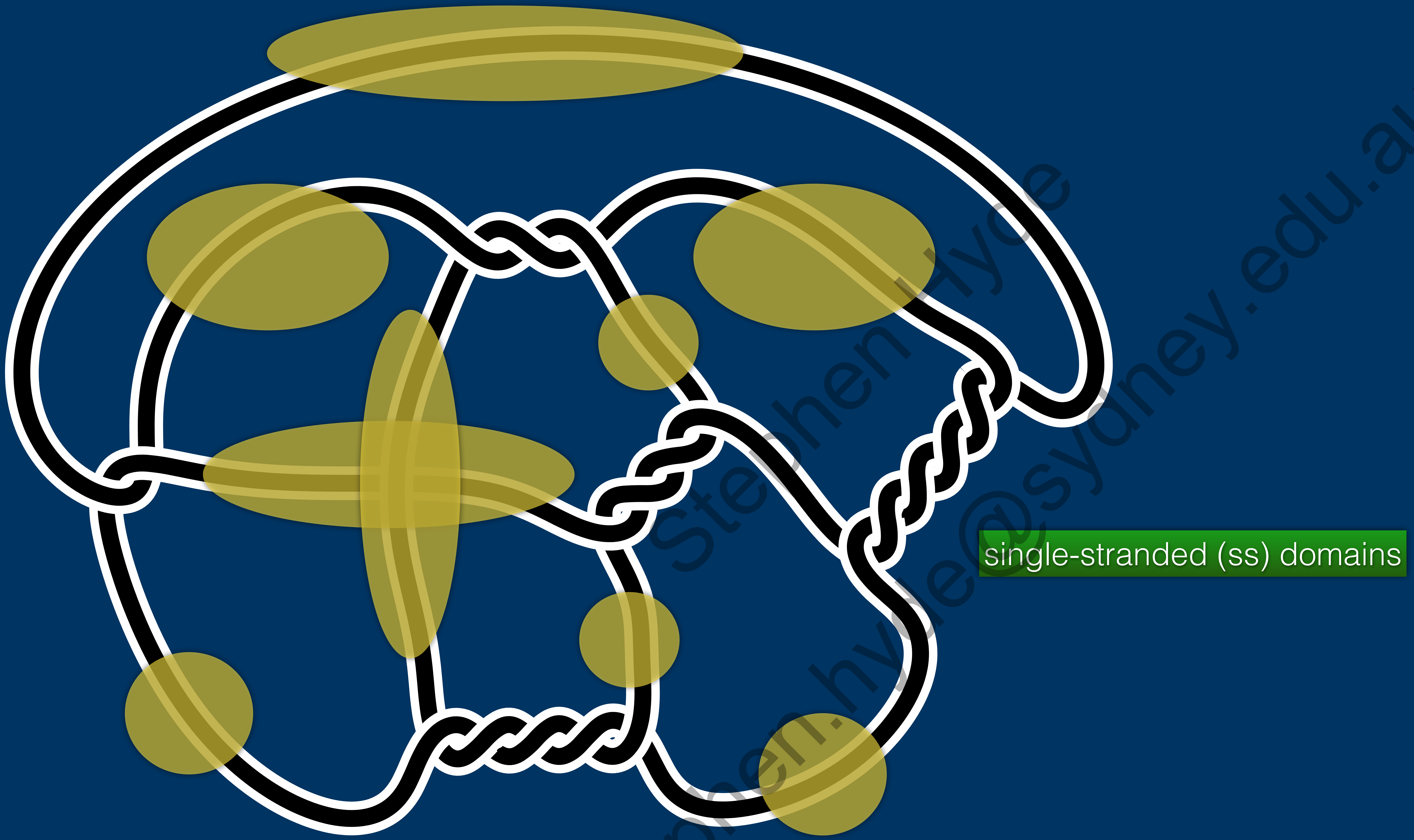
generic contracted folds can be complex

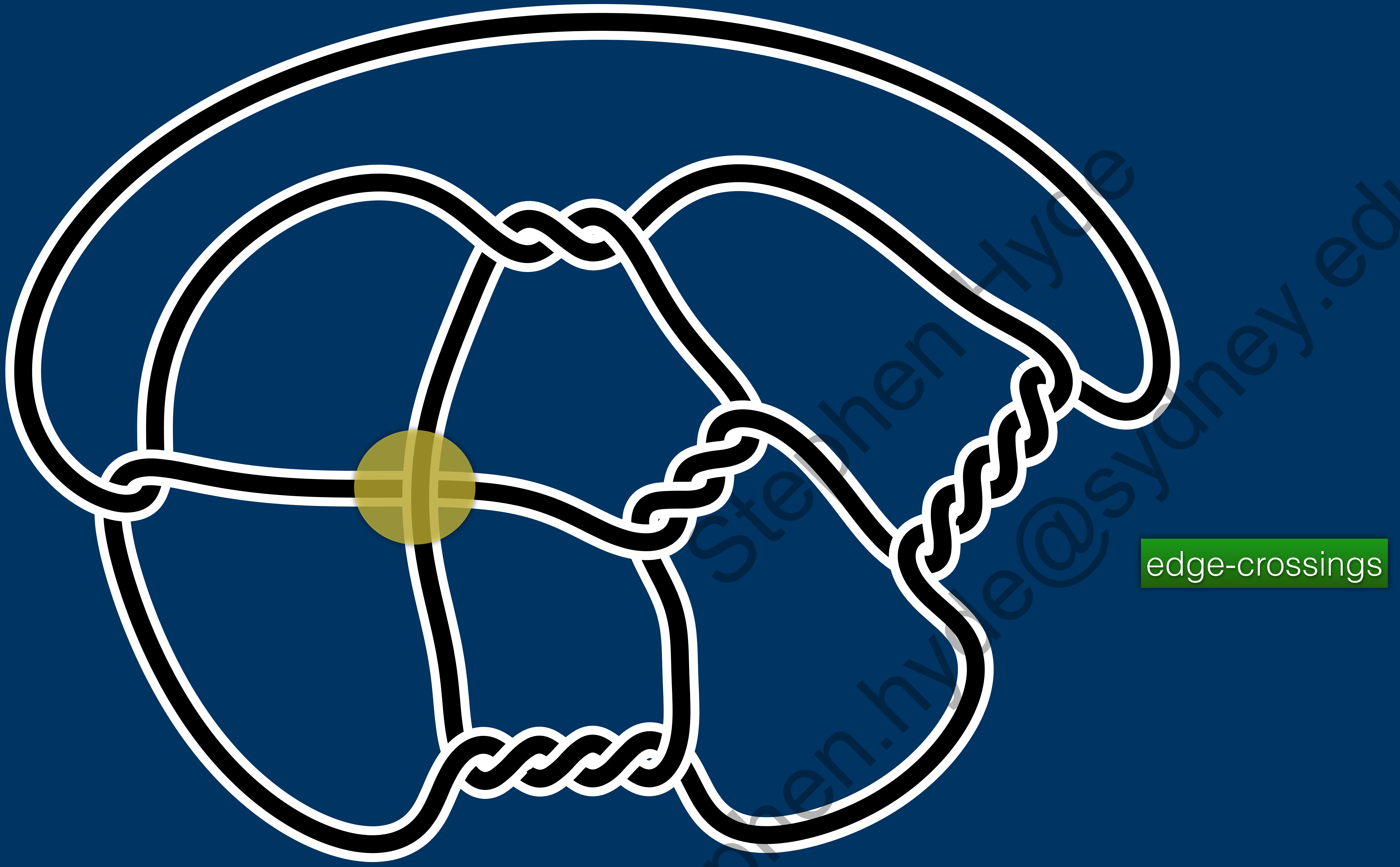



junctions



double-stranded (ds) domains:
parallel
anti-parallel


single-stranded (ss) domains


edge-crossings

All contracted (a)-type folds up to 6 ribbons

Y-junctions only

diagram	$\mathcal{L}^{\otimes}$	*genus*	*strand graph**	*skeleton**
	$\{0\}$	0		
	$\{\bar{1}1\}$	1		
	$\{\bar{1}12\bar{3}32\}$	2		
	$\{\bar{1}21\bar{3}23\}$	2		
	$\{123123\}$	2		
	$\{\bar{1}12\overline{34}35\bar{6}6542\}$	3		
	$\{\bar{1}123\bar{4}435\bar{6}652\}$	3		
	$\{\bar{1}12345364562\}$	3		

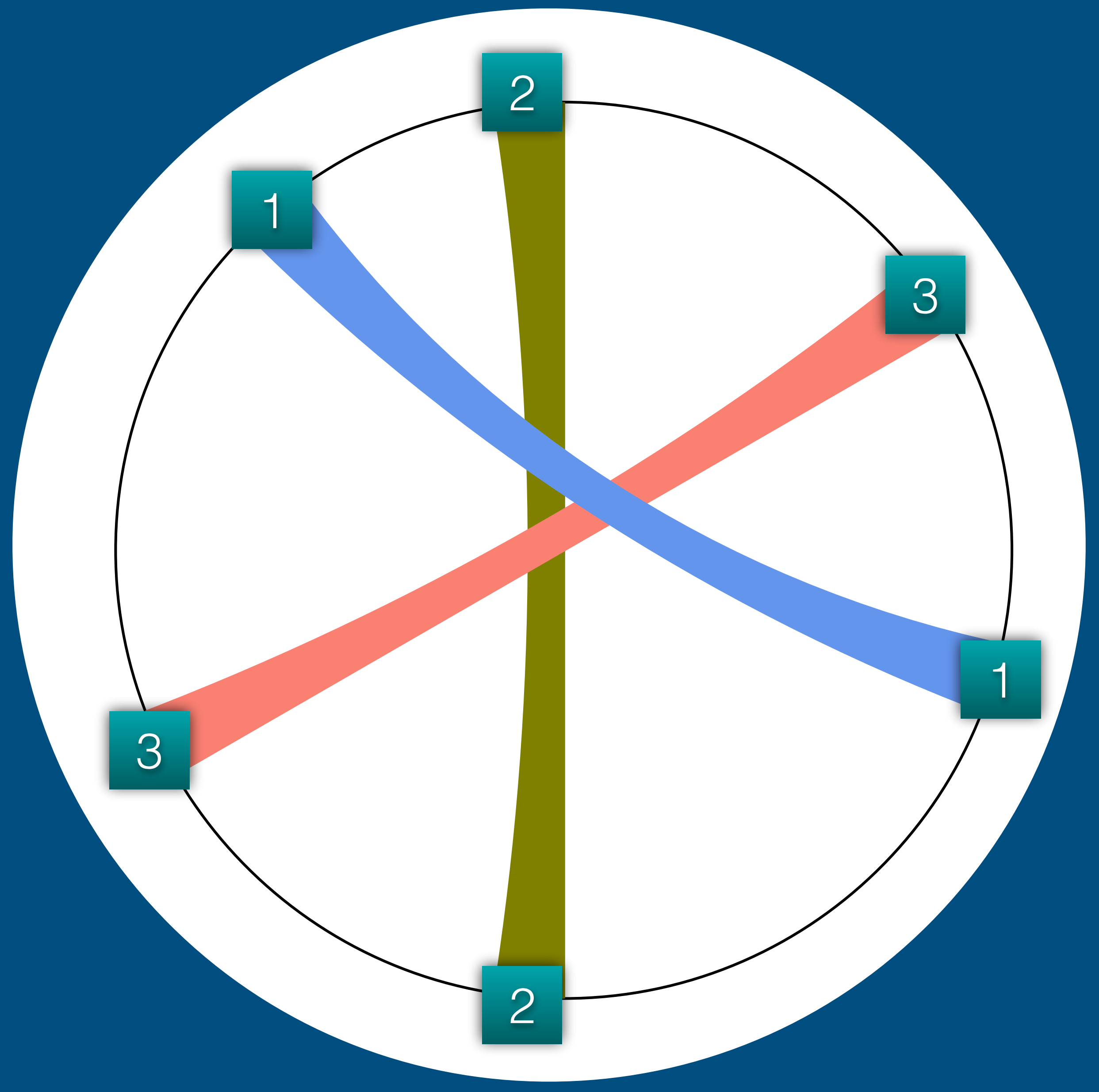
diagram	$\mathcal{L}^{\otimes}$	*genus*	*strand graph**	*skeleton**
	$\{\bar{1}12\bar{3}4\bar{5}3\bar{6}4652\}$	3		
	$\{\bar{1}123\overline{45}365462\}$	3		
	$\{\bar{1}21\bar{3}4\bar{5}23\bar{6}465\}$	3		
	$\{\overline{12}1345326456\}$	3		
	$\{\overline{12}13\overline{45}326546\}$	3		
	$\{\bar{1}231\bar{4}524\bar{6}356\}$	3		
	$\{\overline{123}145364256\}$	3		
	$\{\overline{12}314\bar{5}3\bar{6}4265\}$	3		

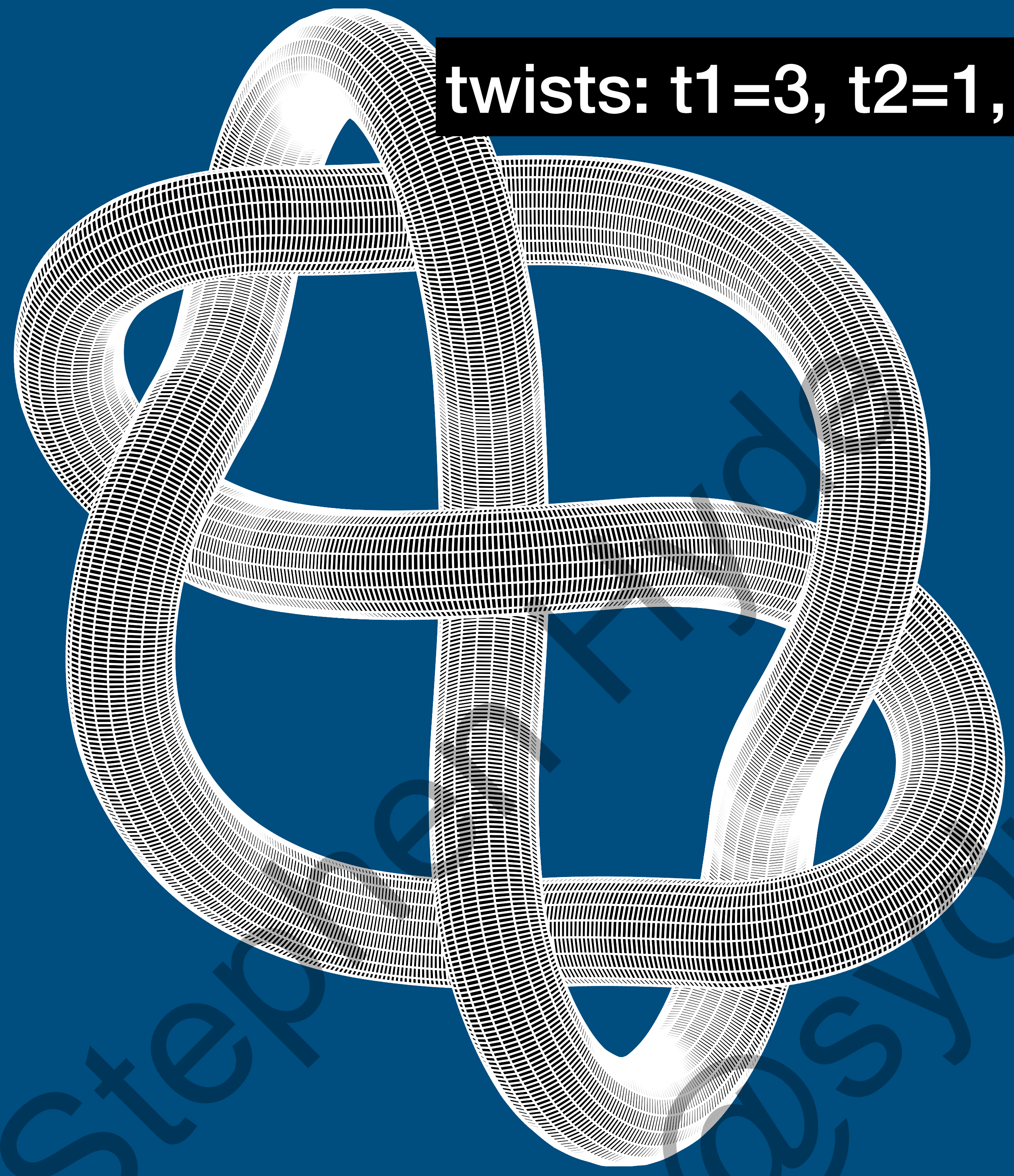
Y-, X-,… junctions

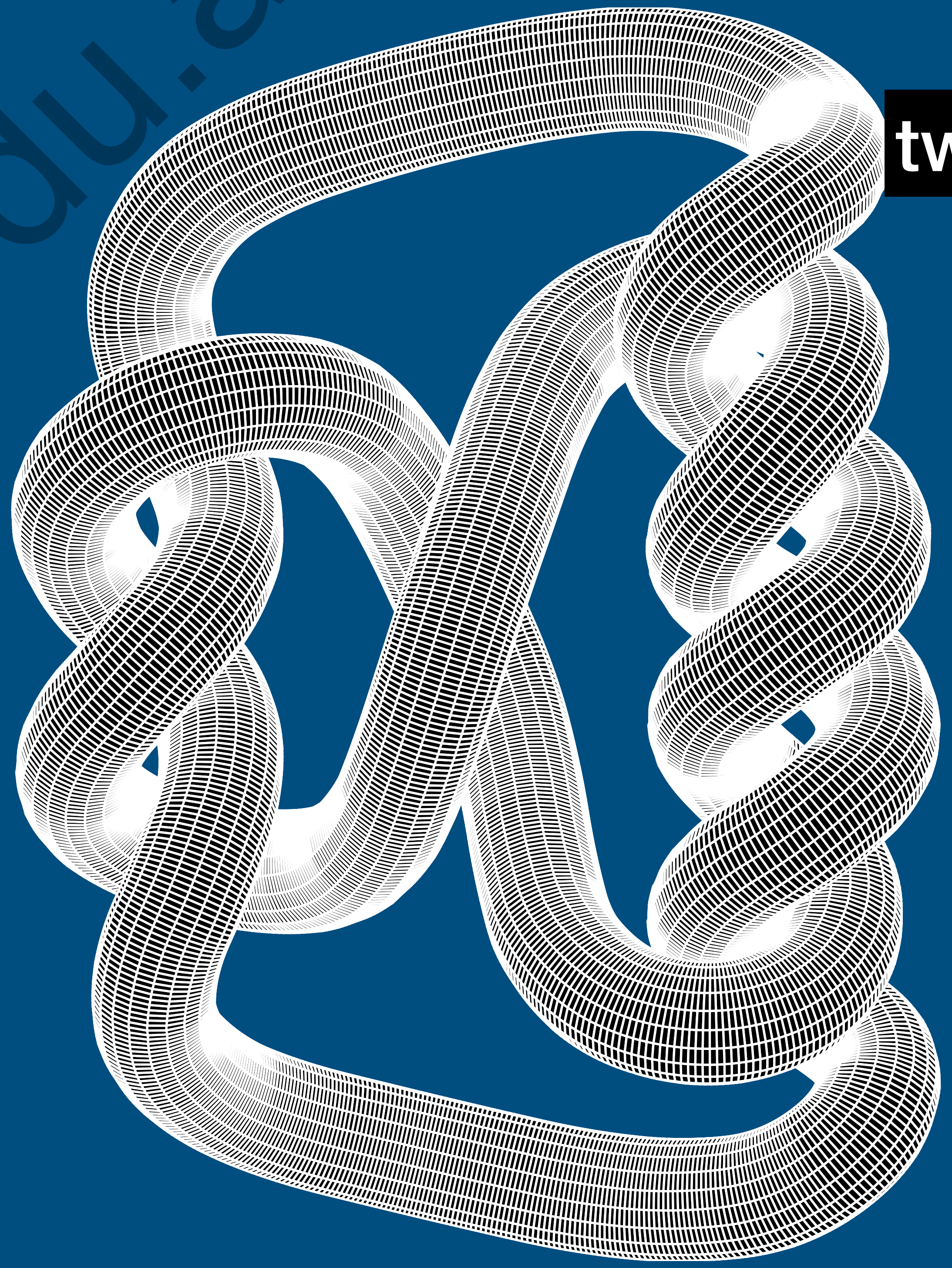
diagram	$\mathcal{L}^{\otimes}$	*genus*	*strand graph*	*skeleton*
	$\{\bar{1}1\bar{2}2\}$	2		
	$\{\bar{1}1\bar{2}2\bar{3}3\}$	3		
	$\{\bar{1}1\overline{23}32\}$	3		
	$\{\bar{1}1\bar{2}23\bar{4}43\}$	3		
	$\{\bar{1}1\bar{2}32\bar{4}34\}$	3		
	$\{\bar{1}1\overline{23}2\bar{4}43\}$	3		
	$\{\bar{1}1234234\}$	3		
	$\{\bar{1}12\overline{34}243\}$	3		
	$\{\bar{1}12\overline{34}432\}$	3		
	$\{\overline{12}1\overline{34}243\}$	3		
	$\{\bar{1}1\overline{23}24\bar{5}543\}$	3		

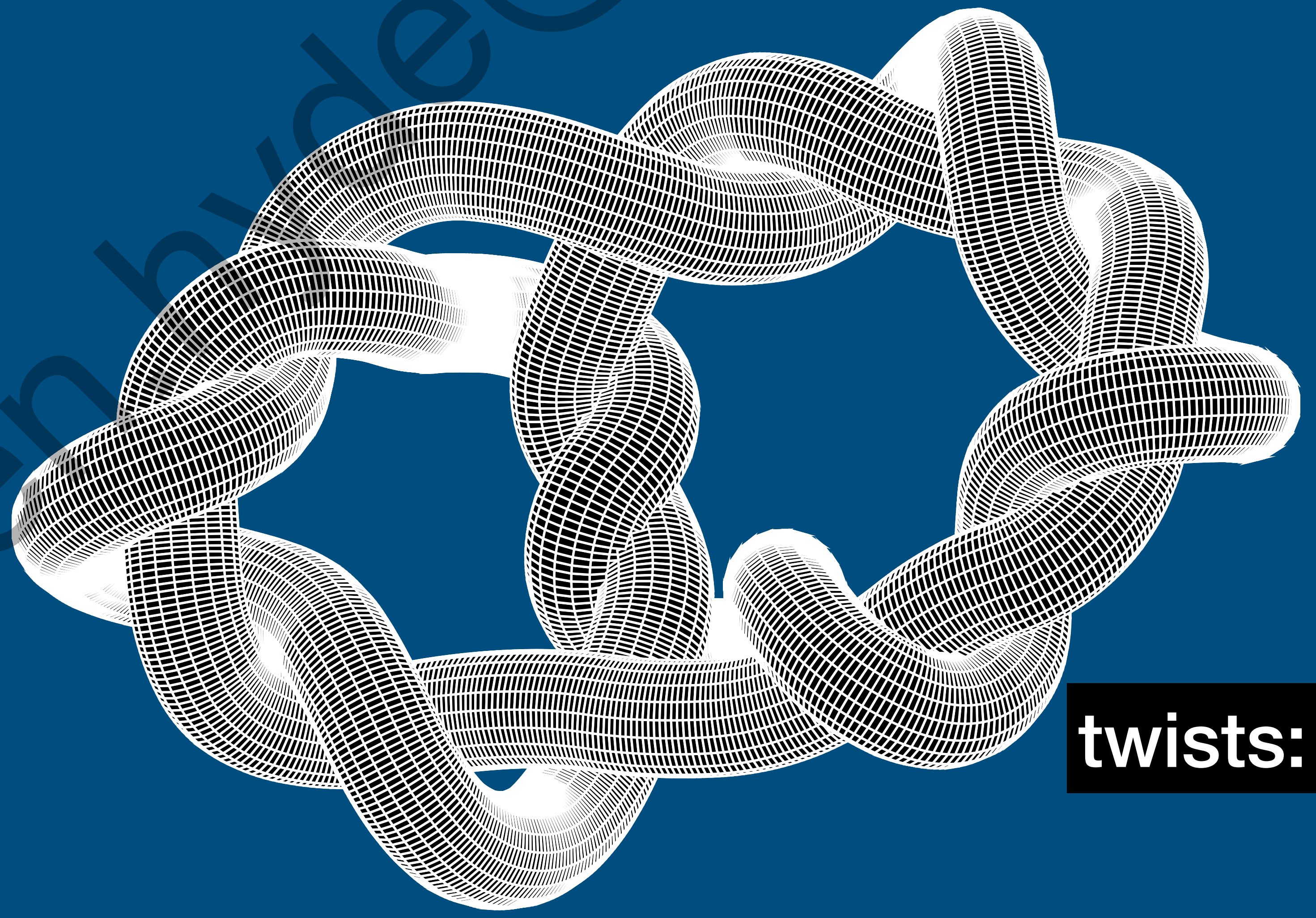
diagram	$\mathcal{L}^{\otimes}$	*genus*	*strand graph*	*skeleton*
	$\{\bar{1}12\bar{3}324\bar{5}54\}$	3		
	$\{\bar{1}123425345\}$	3		
	$\{\bar{1}12\bar{3}\bar{4}2\bar{5}354\}$	3		
	$\{\bar{1}12\overline{34}25435\}$	3		
	$\{\bar{1}12\bar{3}43\bar{5}452\}$	3		
	$\{\bar{1}12\overline{34}3\bar{5}542\}$	3		
	$\{\bar{1}123453452\}$	3		
	$\{\bar{1}123\overline{45}3542\}$	3		
	$\{\overline{12}13425345\}$	3		
	$\{\bar{1}213\bar{4}2\bar{5}354\}$	3		
	$\{\bar{1}21\bar{3}423\bar{5}45\}$	3		
	$\{\overline{12}1\overline{34}25435\}$	3		

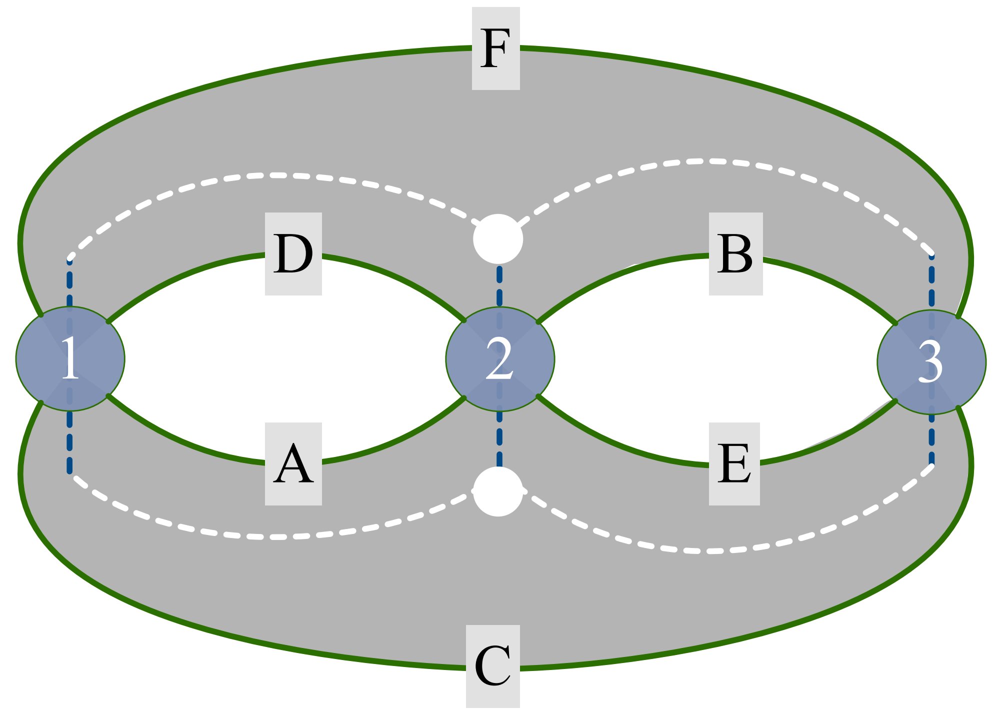
generic folds are KNOTTED



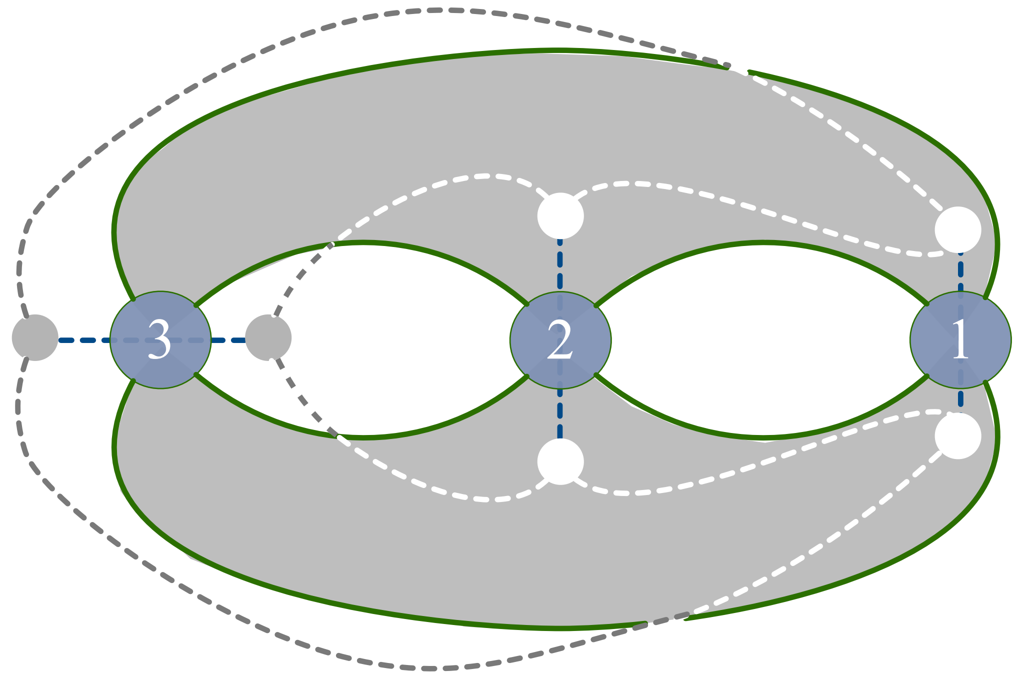






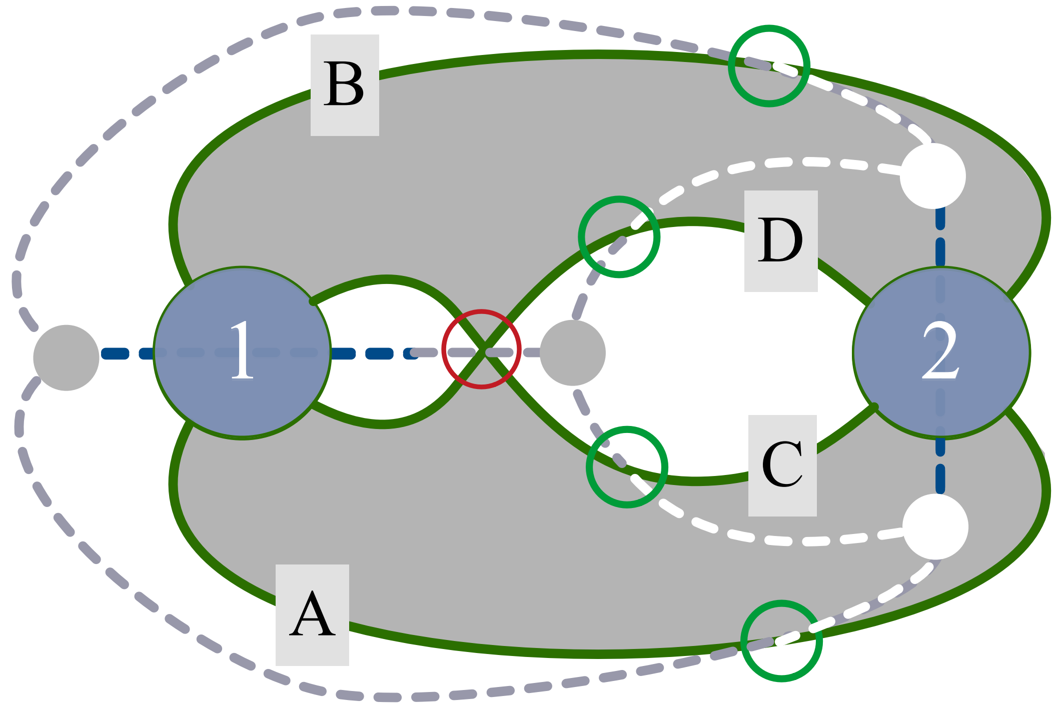


(a) uncrossed embed is one-colored



(b) uncrossed embed is two-colored

unique simplest strand knots set by duplex twists



multiple simplest strand knots set by duplex twists, ±1 sign of edge-crossings

(c) all graph embeds contain edge-crossings

(a) uncrossed embed is one-colored

(b) uncrossed embed is two-colored

Idea: An **uncrossed** polarised strand graph is **‘topologically rigid’**

embedding of a fold with no edge crossings describes a unique (un)knot

Knot depends on nmbr of twists in each duplex: all (simpler?) cases are algebraic tangles

$\mathcal{L}_T^{\otimes}$	*Conway rational link*
$\{1_a 2_b 12\}$	$[a\bar{b}]$
$\{1_a 2_b 3_c 123\}$	$[a, b, c]$
$\{1_a 2_b 123_c 4_d 34\}$	$[b(-a)]\#[d(-c)]$
$\{1_a 2_b 13_c 24_d 34\}$	$[\bar{c} d \bar{a} b]$

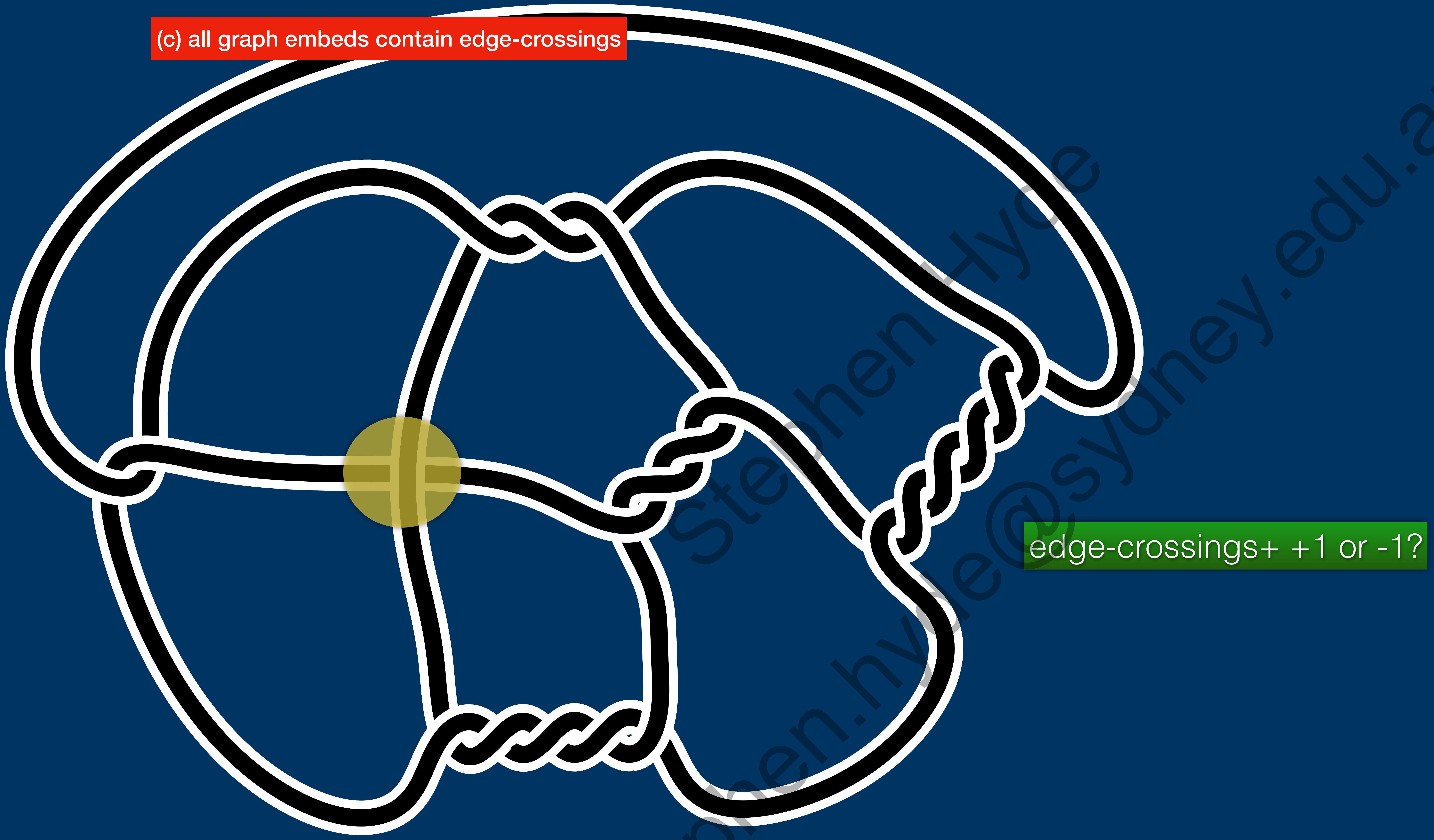
$\mathcal{L}_{\amalg}^{\otimes}$	t_1	t_2	t_3	t_4	*pseudoknot*	*knot*	*knot ID*
$\{1_e 2_e 12\}$	0	0	–	–	×		
	0	2	–	–	×		
unknot	0	4	–	–	×		
trefoil	2	2	–	–		×	3_1
	2	4	–	–		×	5_2
	4	4	–	–		×	7_4
$\{1_o 2_o 3_o 123\}$	1	1	1	–		×	3_1
	1	1	3	–		×	5_2
	1	3	3	–		×	7_4
	3	3	3	–		×	9_{35}
$\{1_e 2_e 123_e 4_e 34\}$	0	0	0	0	×		
	0	0	0	2	×		
	0	0	0	4	×		
	0	0	2	2		×	3_1
	0	0	2	4		×	5_2

⋮

(c) all graph embeds contain edge-crossings

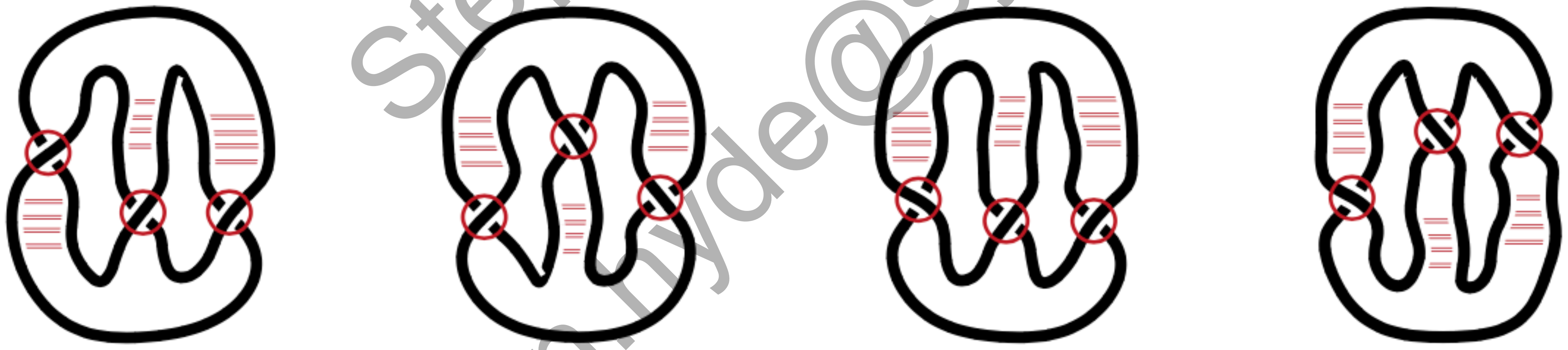
Idea: A ***crossed*** polarised strand graph embedding of a fold is ‘topologically nonrigid’

Knot depends on edge-crossing twist (±1)


(c) all graph embeds contain edge-crossings
edge-crossings+ +1 or -1?

Idea: A ***crossed*** polarised strand graph embedding of a fold is ‘topologically nonrigid’

Knot depends on edge-crossing twist (±1)

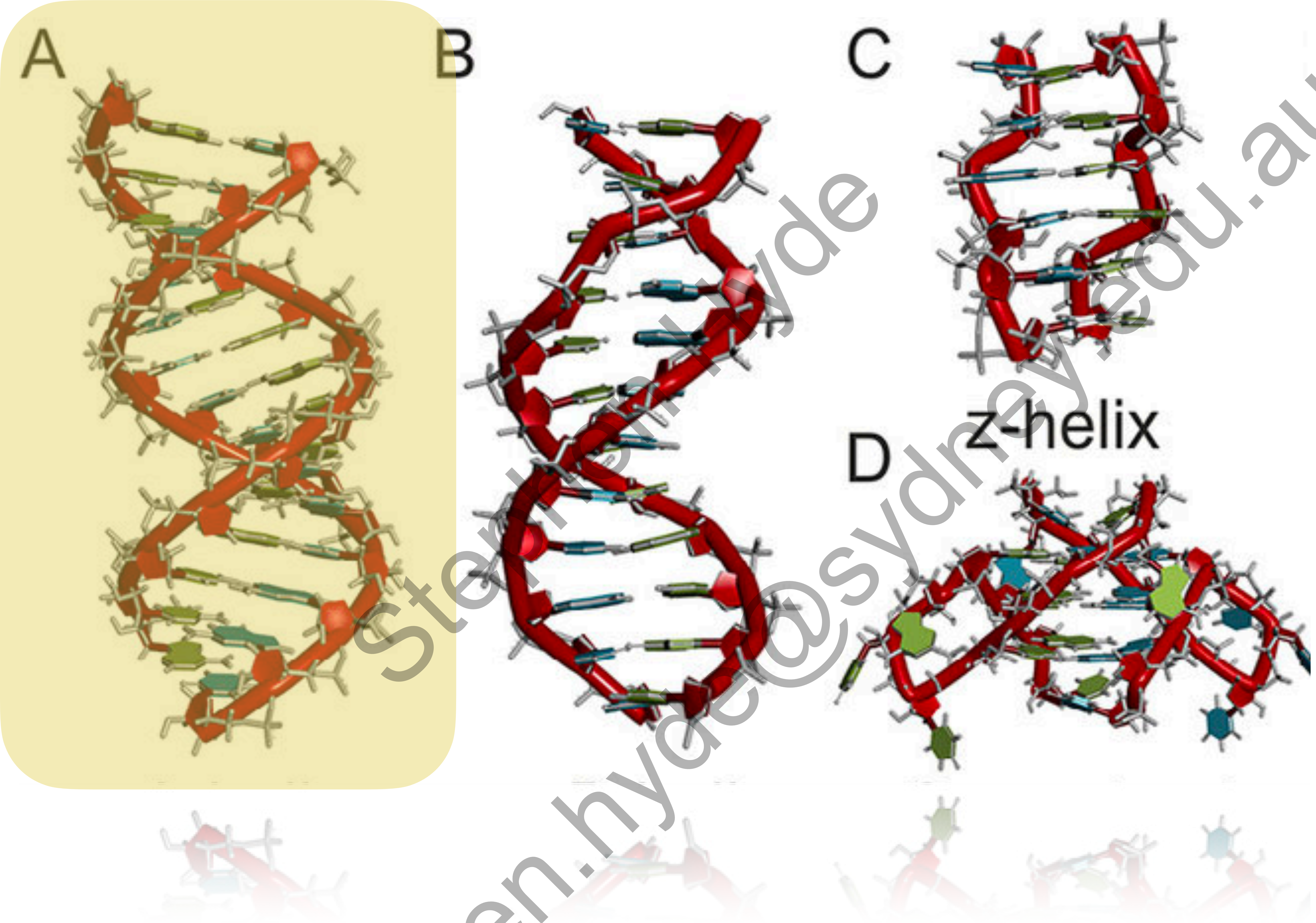


3_1^+ 0_1 0_1 3_1^-

$$\{1_0 2_0 3_0 123\}$$

In absence of other interactions and entropy 50% chance of knotting!

ssRNA duplexes are usually A-form: antiparallel and right-handed, ca 5 nucleotides per half-twist



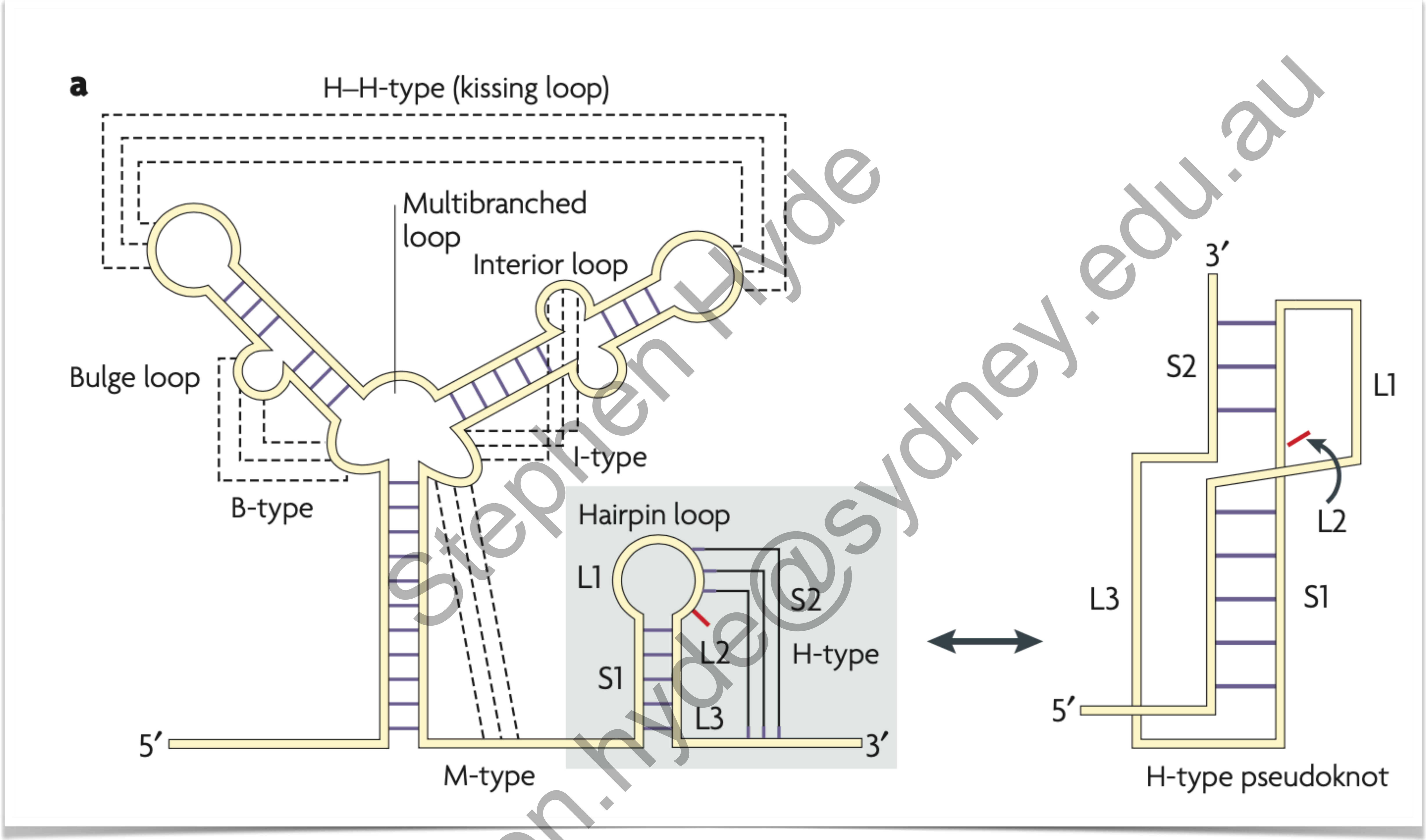
Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

Table 2: Distinct semiflagged, contracted canonical fold labels, $\mathcal{L}$, for folds containing up to 5 duplexes, assuming all ribbons are annular.

all-antiparallel folds

n	*nmbr of labels*	*smallest* $\mathcal{L}$	*largest* $\mathcal{L}$
0	1	{0}	{0}
1	0	–	–
2	1	{1212}	{1212}
3	1	{123123}	{123123}
4	4	{12123434}	{12341234}
5	19	{1212343545}	{1234512345}

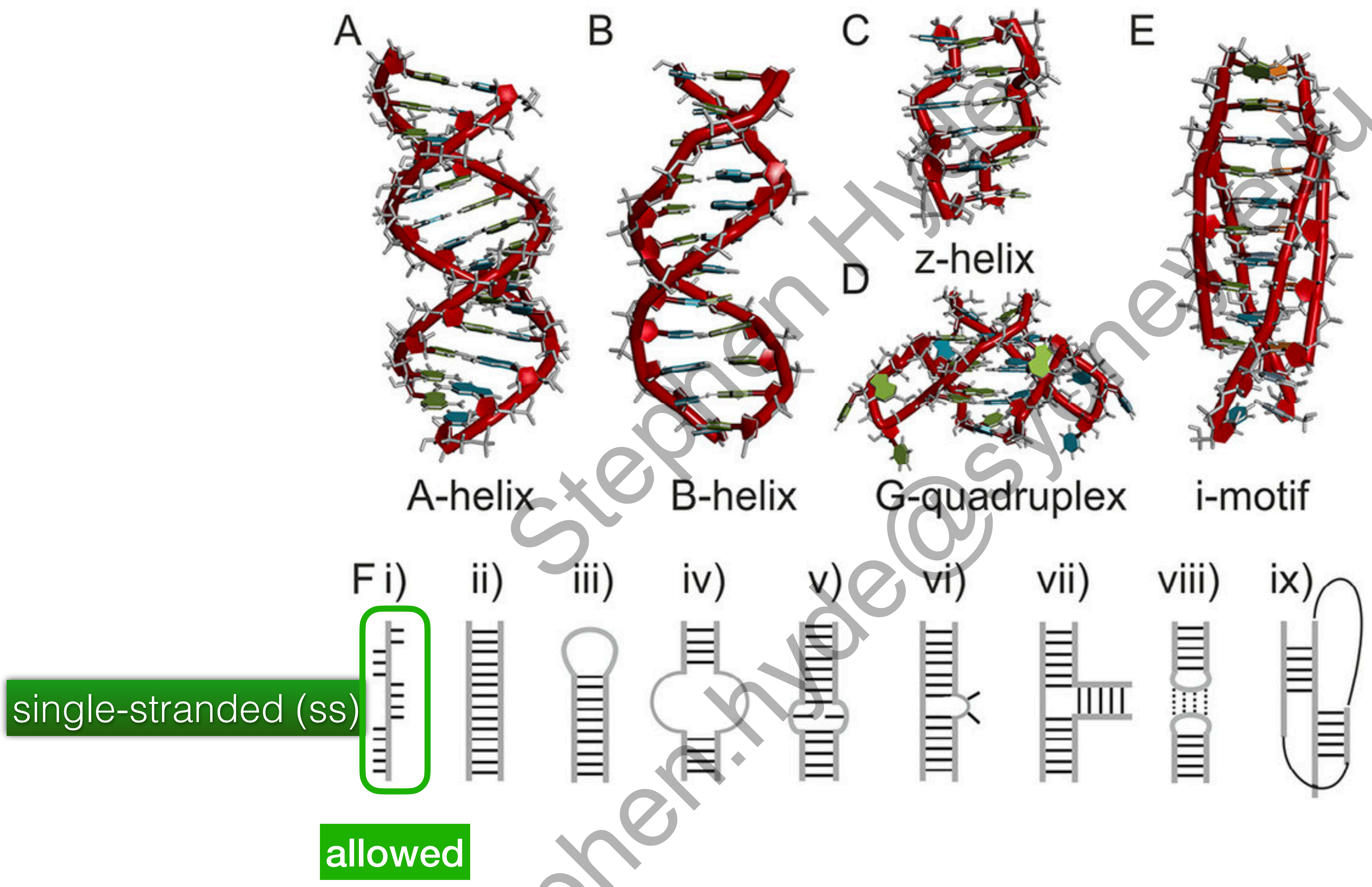
RNA folds can be far more complex than simple linear double-helices…



Pseudoknots

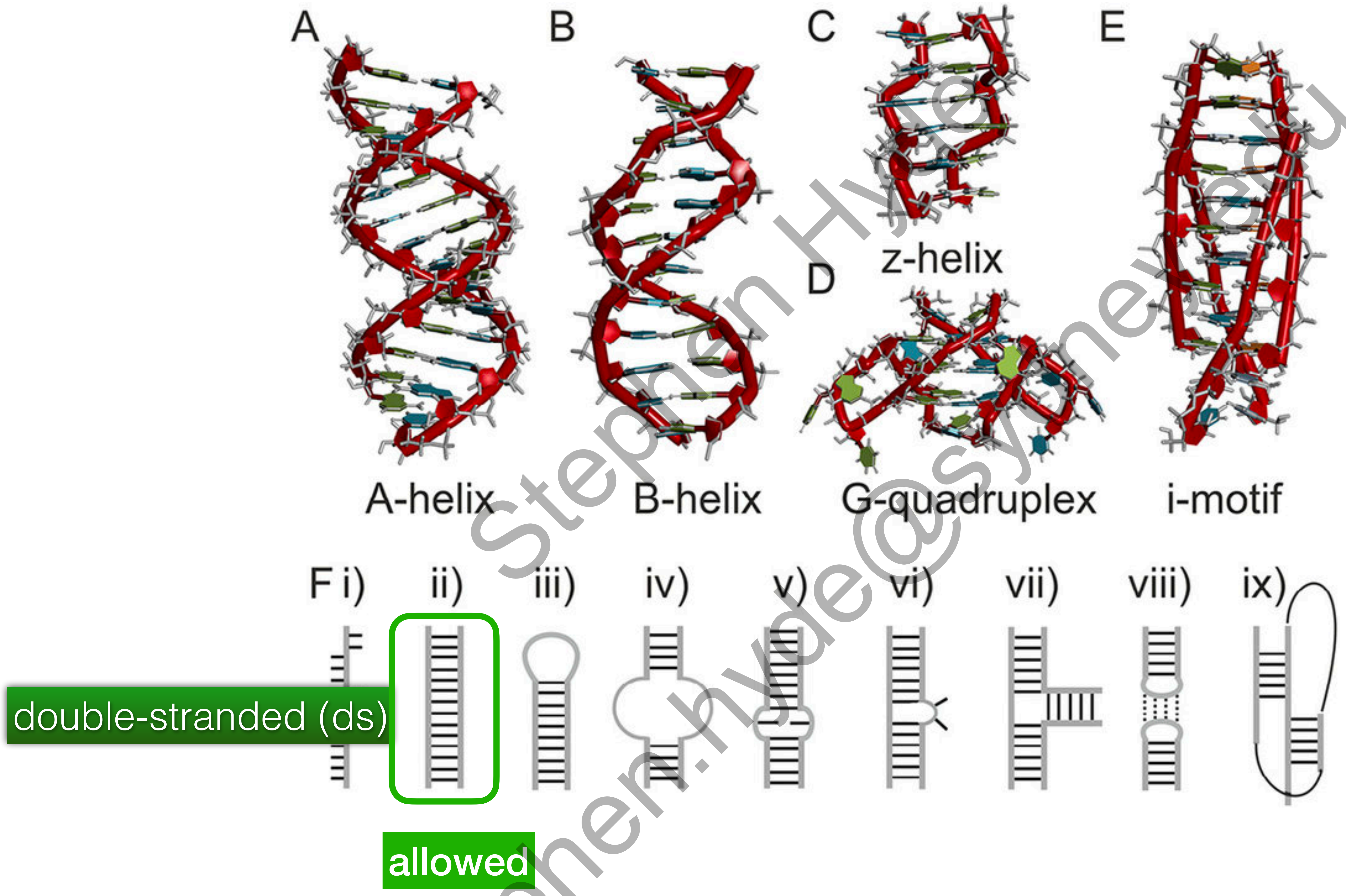
Brierley, Ian, Simon Pennell, and Robert JC Gilbert. "Viral RNA pseudoknots: versatile motifs in gene expression and replication." *Nature Reviews Microbiology* 5.8 (2007): 598-610.

RNA folds can be far more complex than simple linear double-helices…



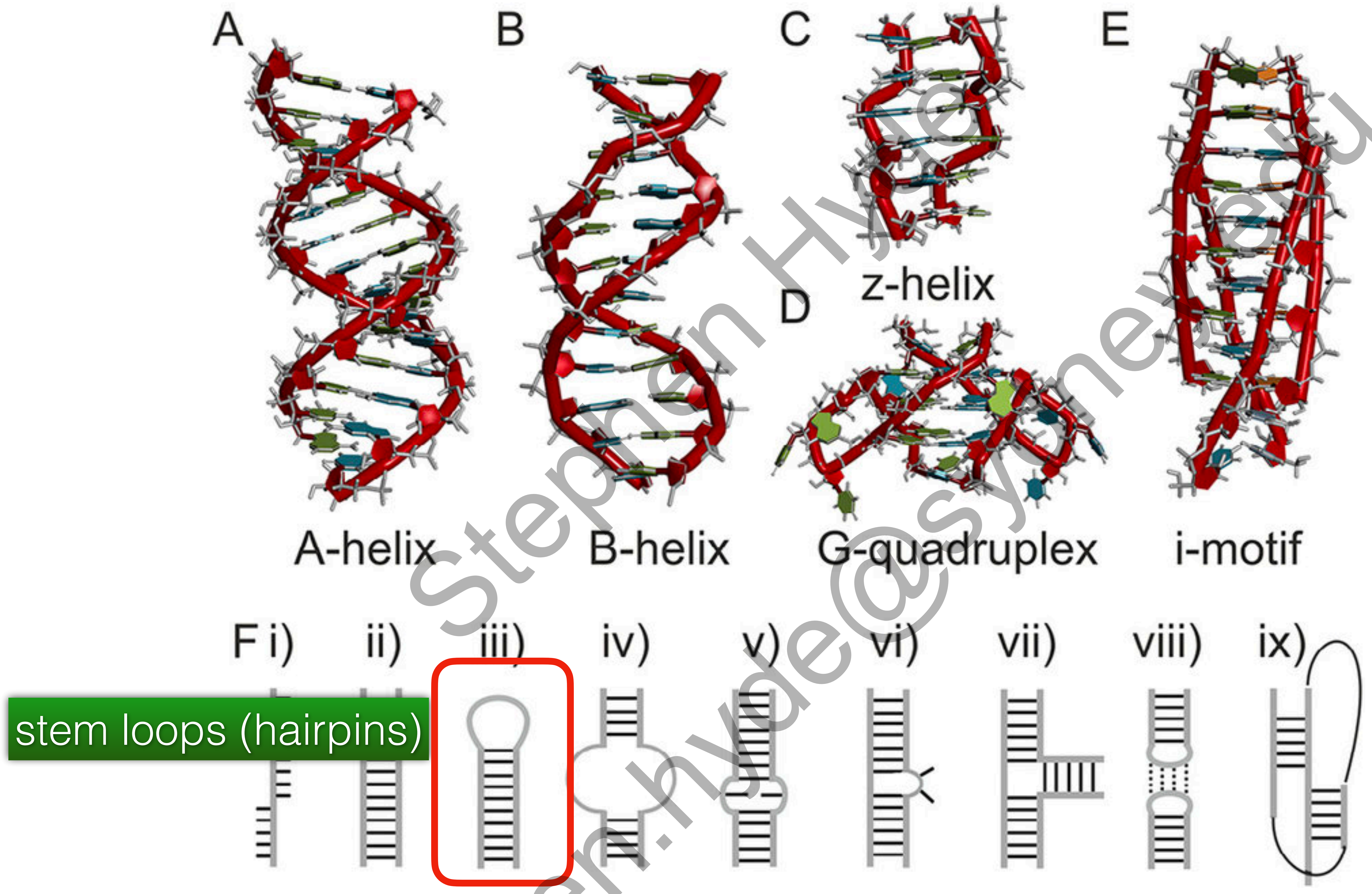
Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

RNA folds can be far more complex than simple linear double-helices…



Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

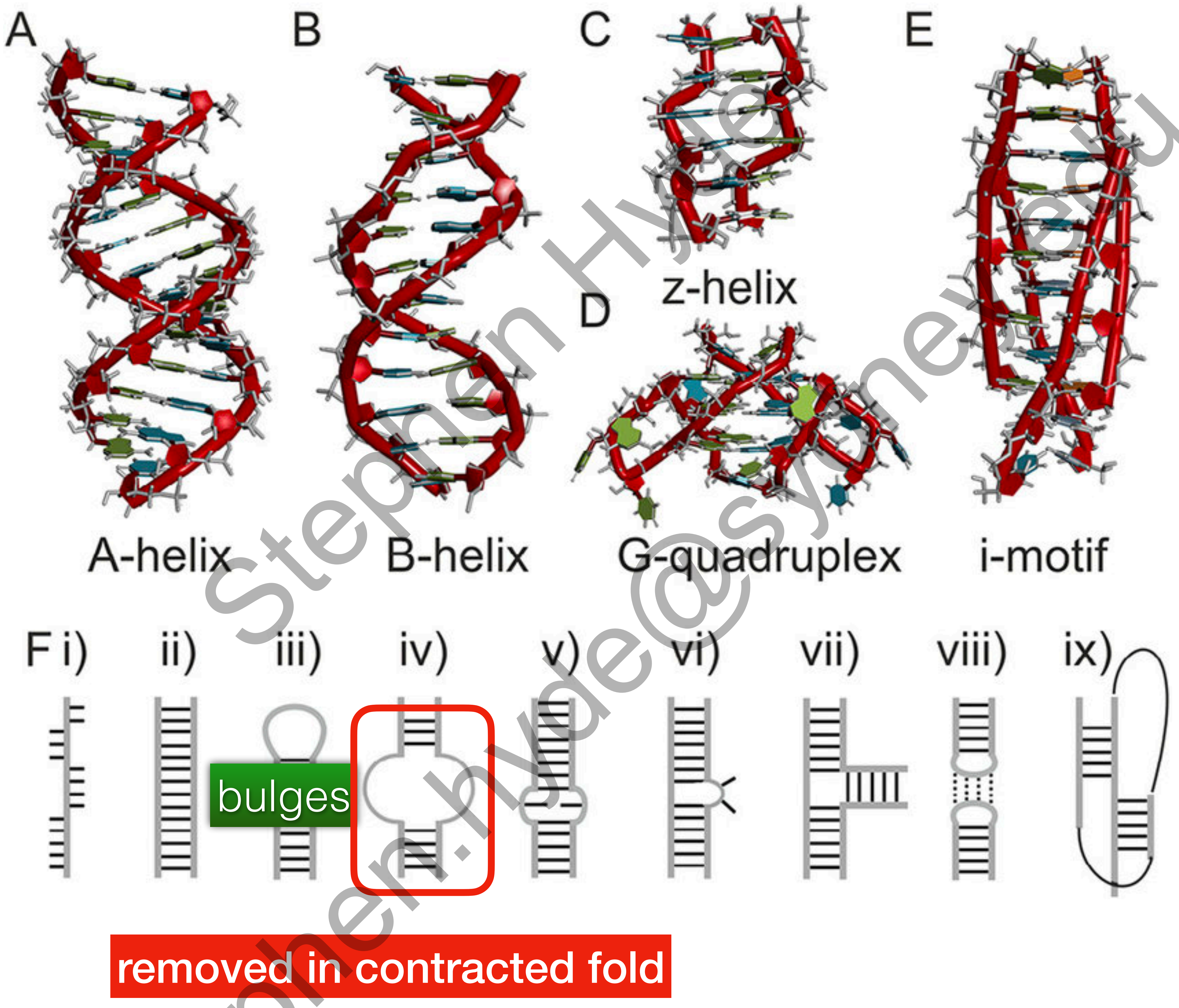
RNA folds can be far more complex than simple linear double-helices…



'twist defects': deleted in contracted fold

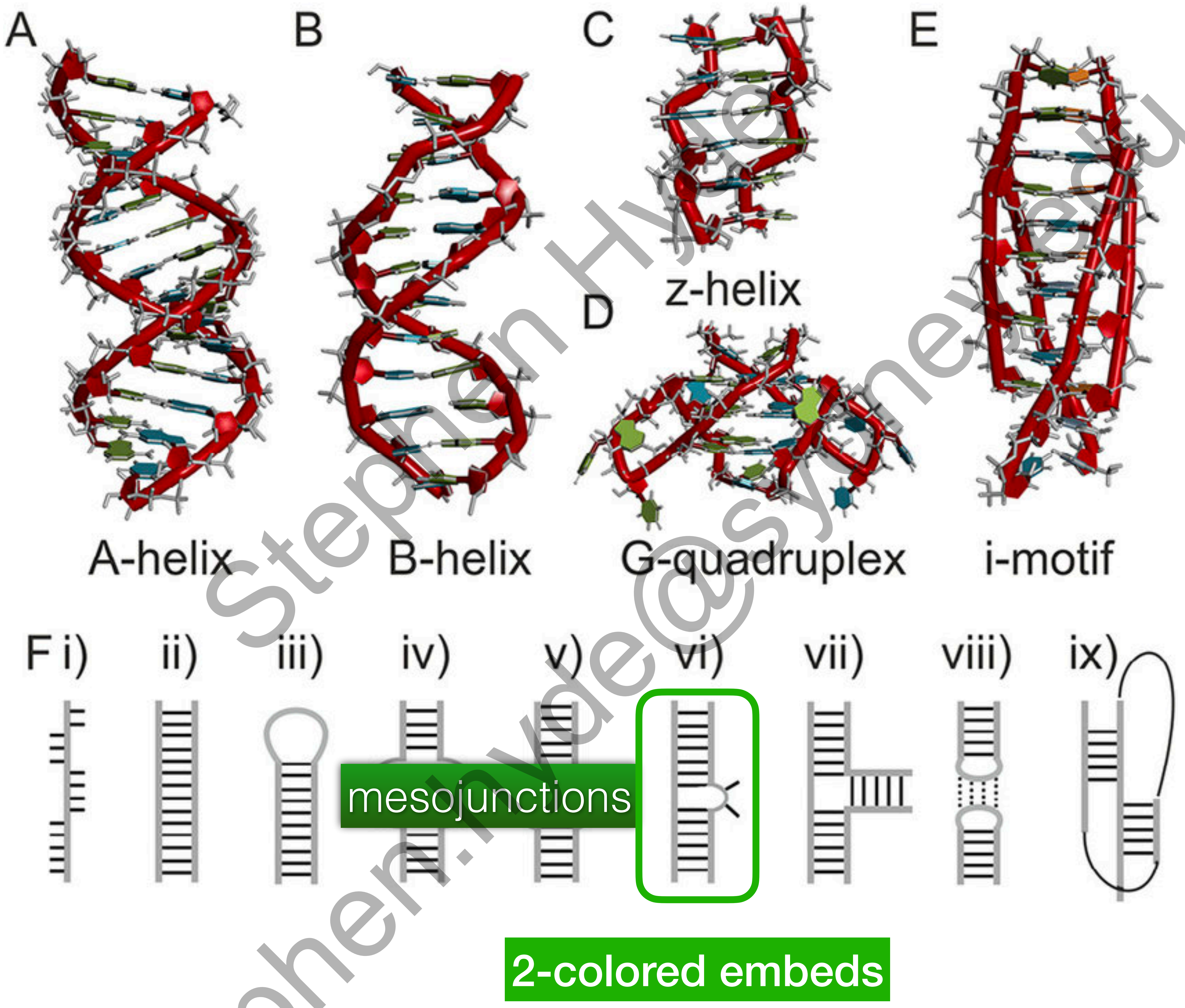
Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

RNA folds can be far more complex than simple linear double-helices…



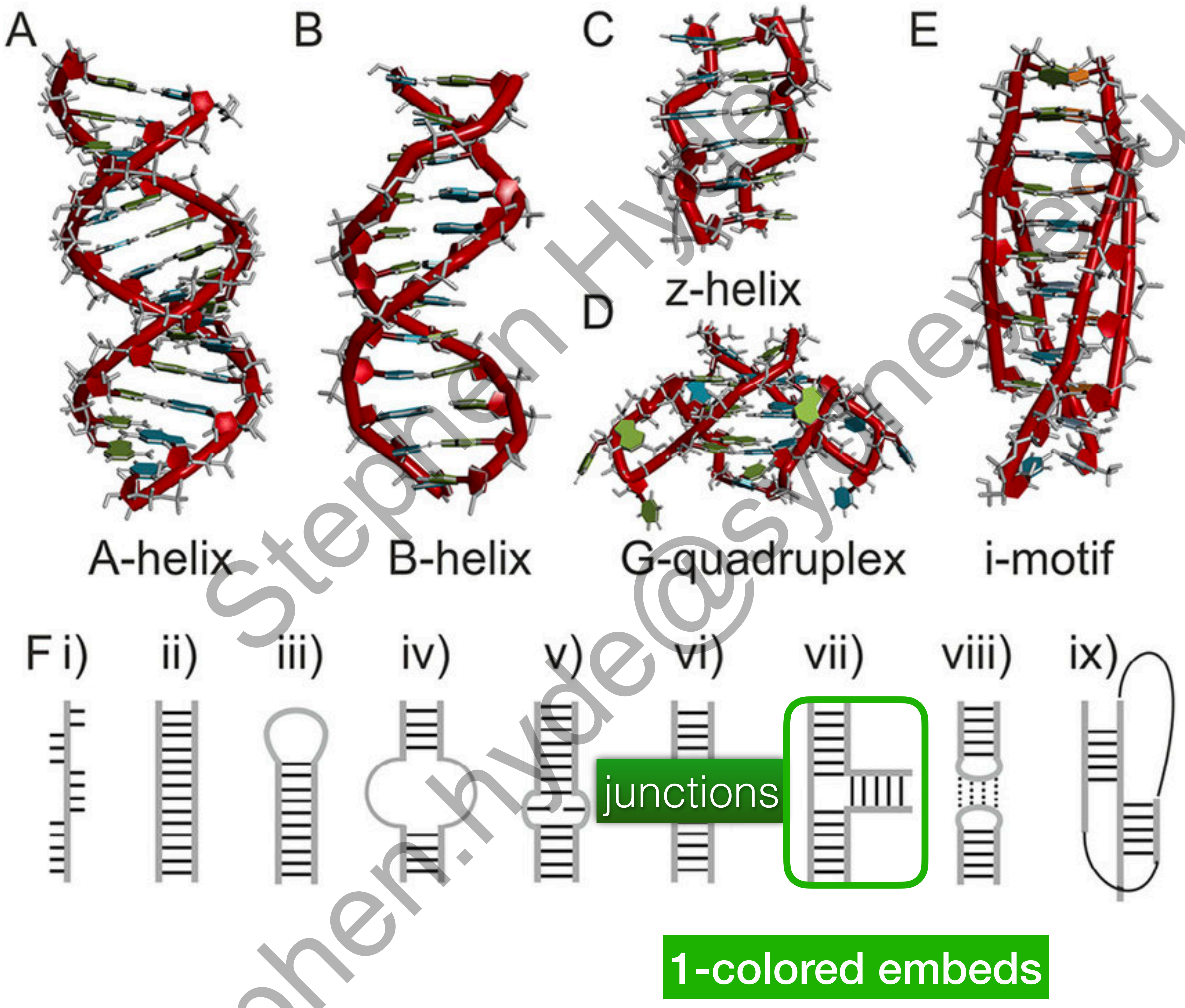
Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

RNA folds can be far more complex than simple linear double-helices…



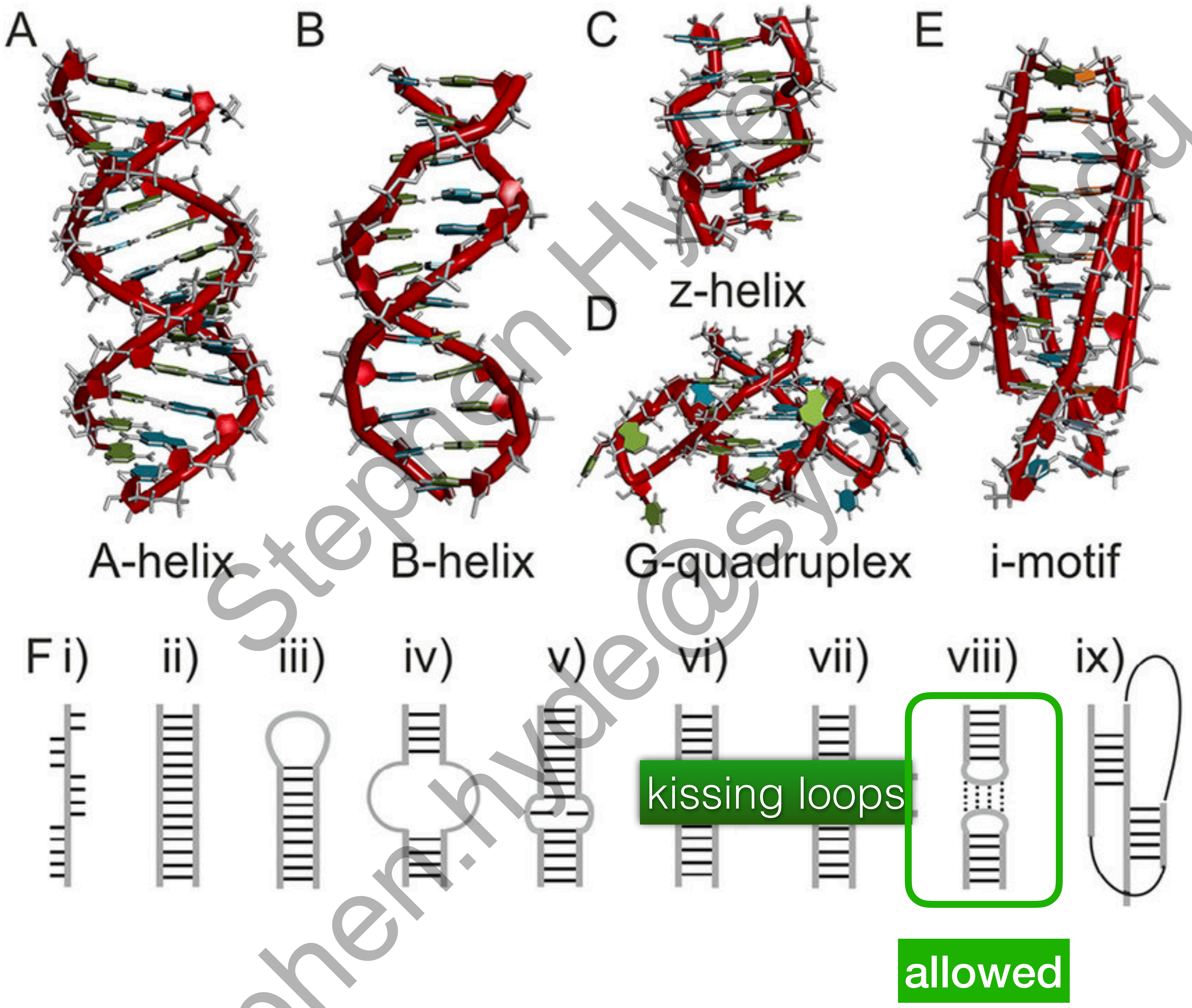
Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

RNA folds can be far more complex than simple linear double-helices…



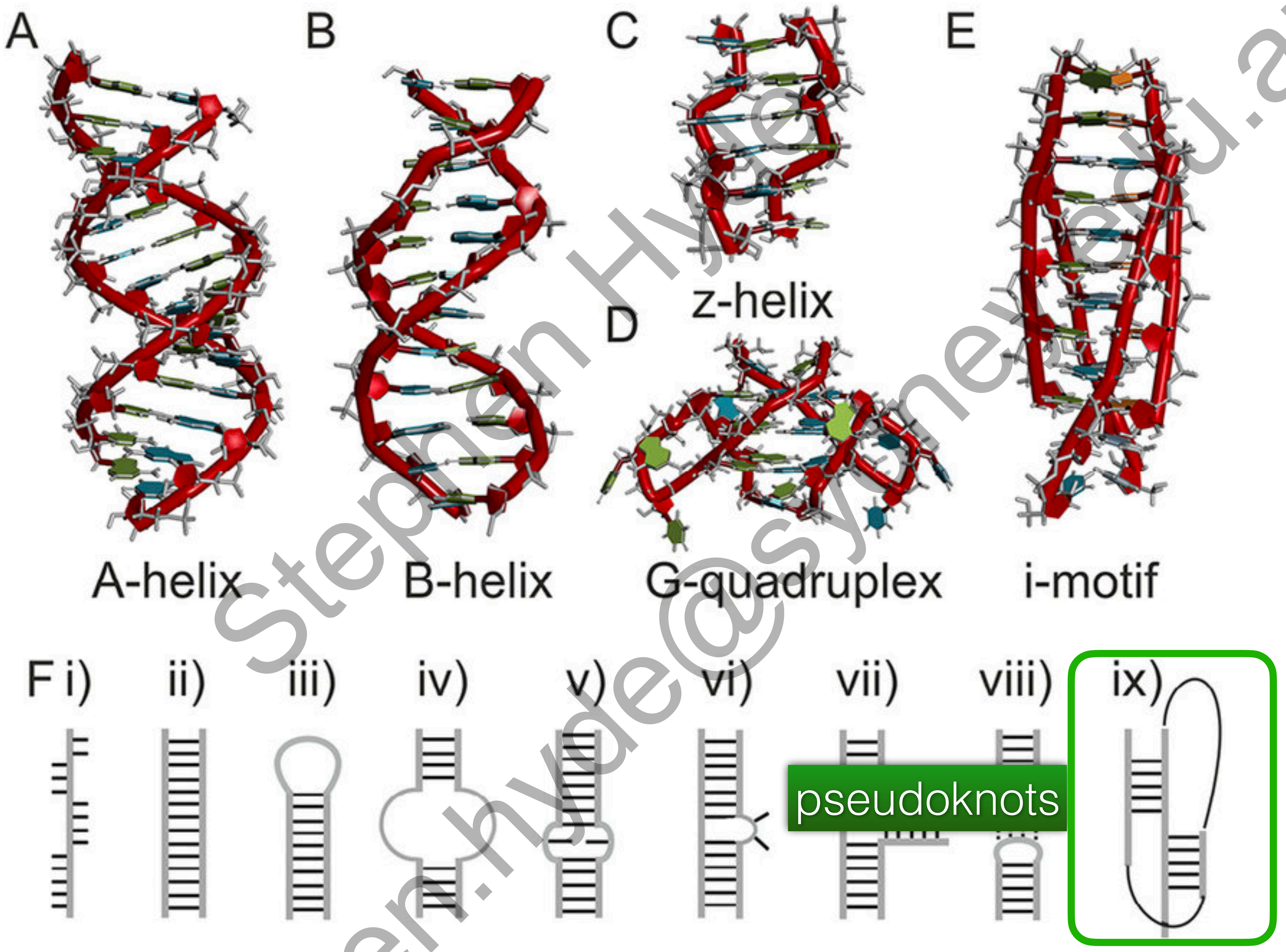
Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

RNA folds can be far more complex than simple linear double-helices…



Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

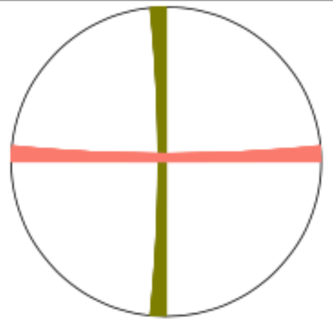
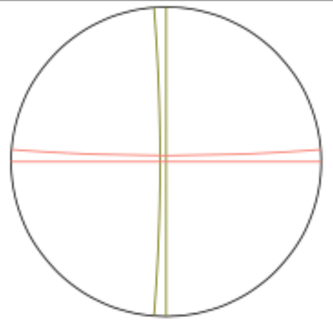
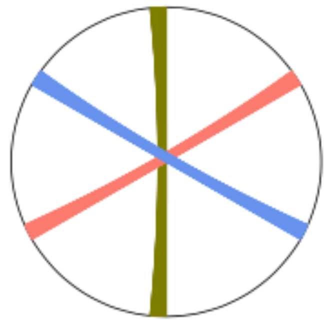
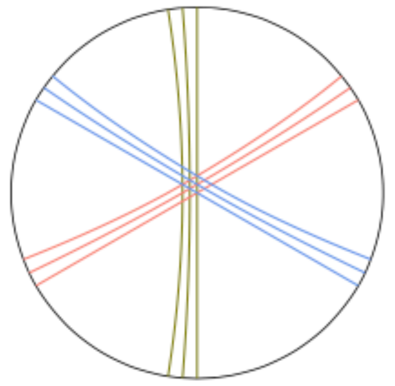
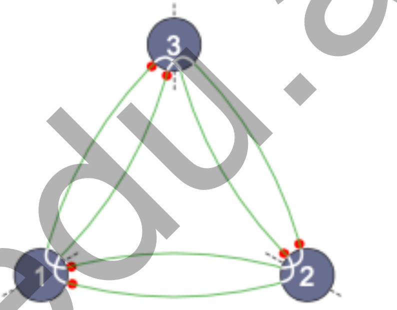
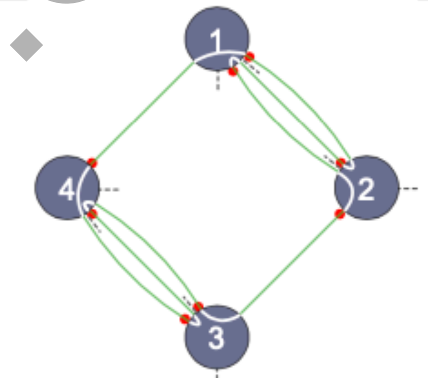
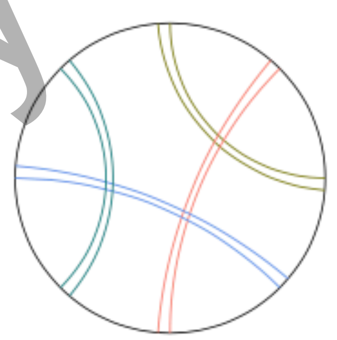
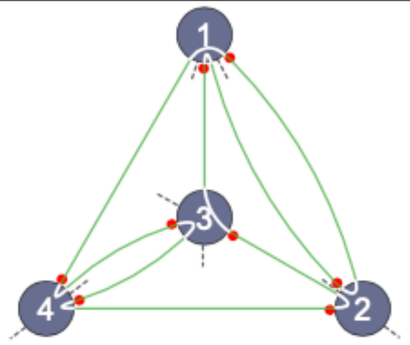
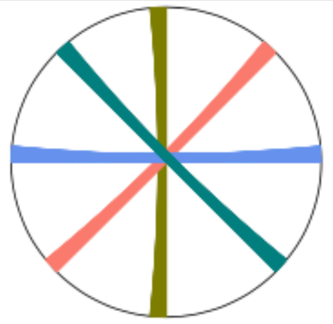
RNA folds can be far more complex than simple linear double-helices…



allowed: ALL contracted folds are pseudo knotted

Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F., & Sigel, R. K. (2016). Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coordination Chemistry Reviews*, *327*, 123-142.

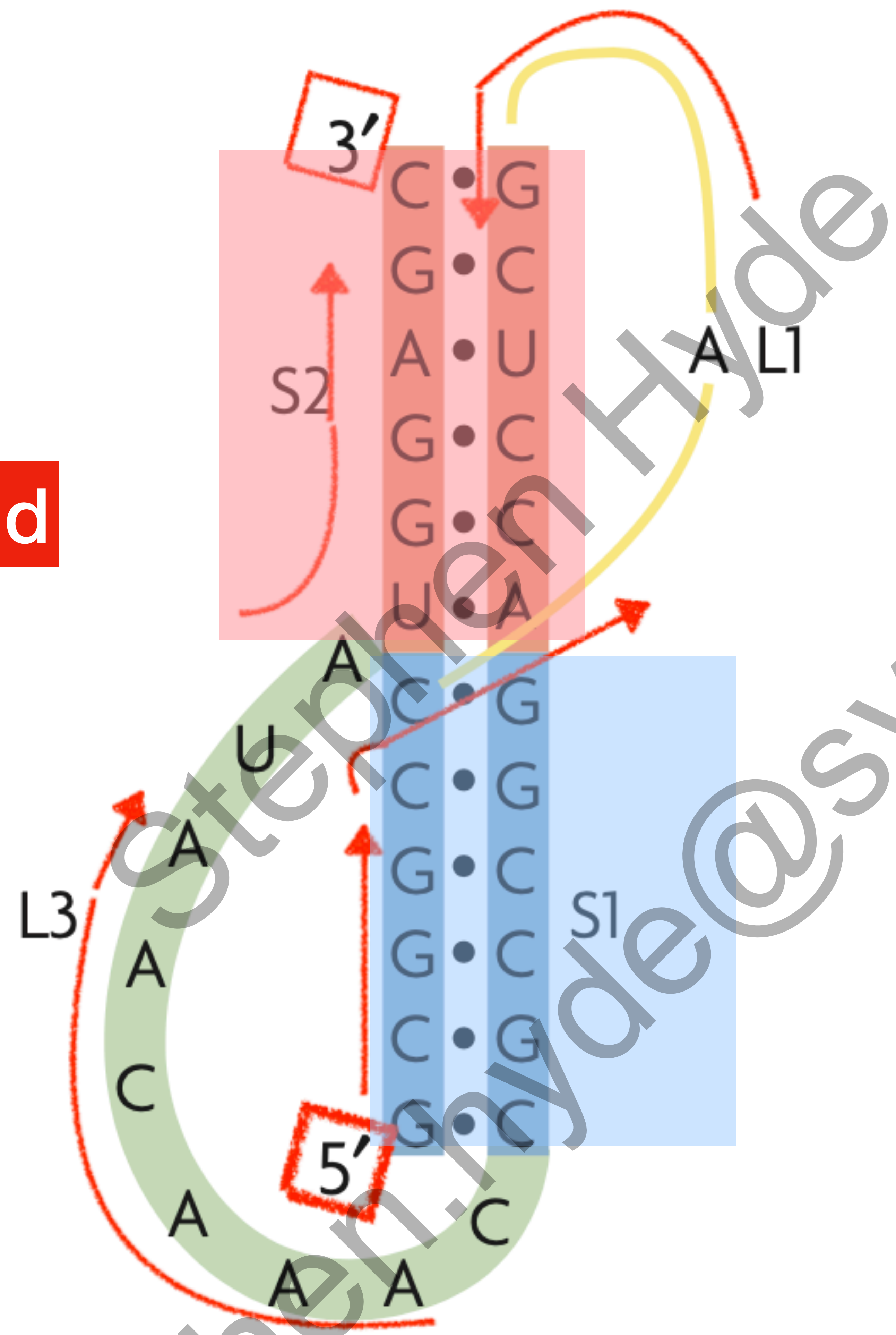
All contracted (a), (b), (c) antiparallel folds - up to 4 ribbons:

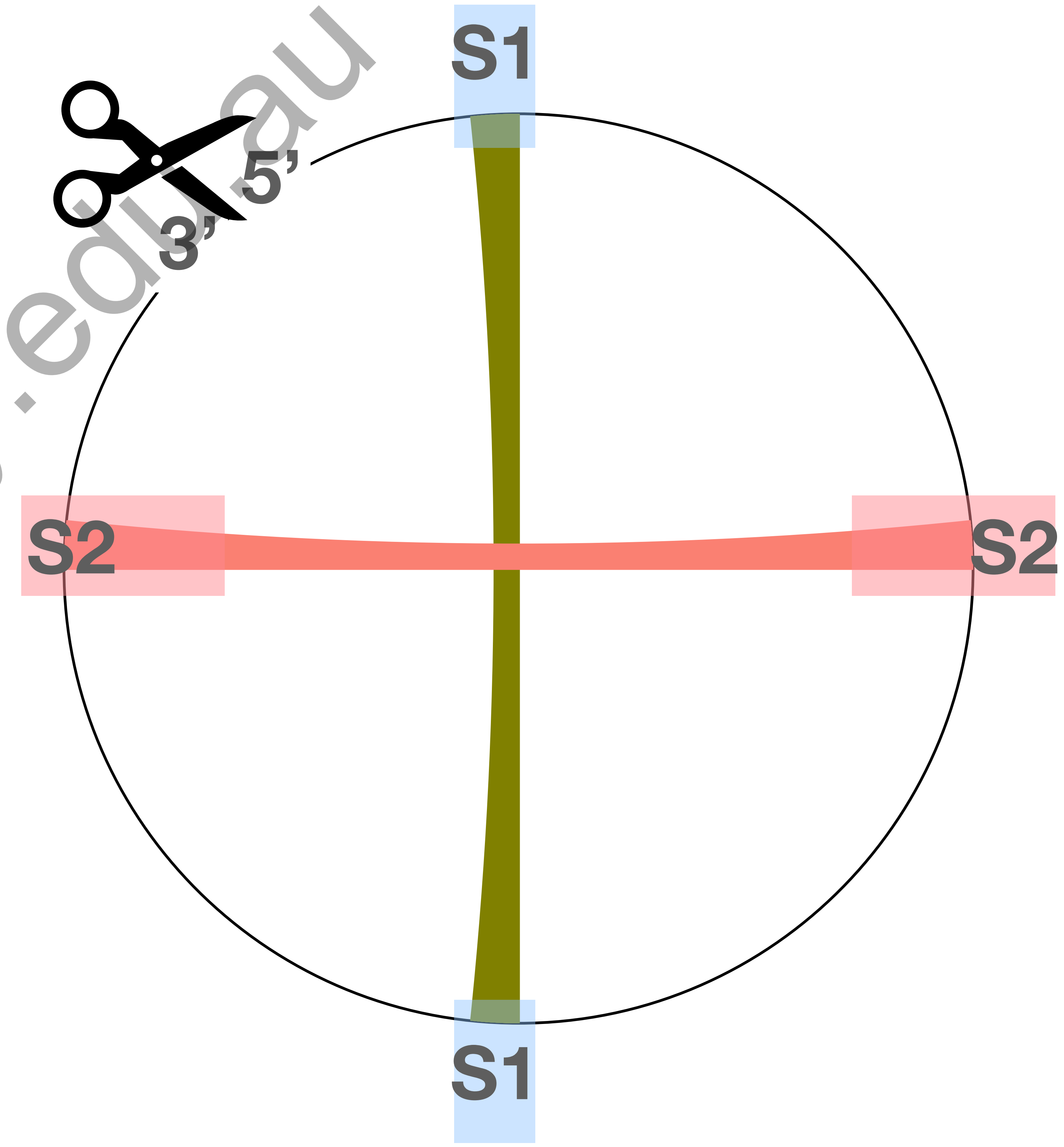
fold diag	$\mathcal{L}^{\otimes}_{\cdot}$	*fold class*	$\mathcal{L}^{\otimes}_{\Pi}\ (N_{\times} = 0)$	*fold diag*	*strand graph*	*genus*
	$\{1212\}$	(b)	$\{1_e2_e12\}$			3
	$\{123123\}$	(a)	$\{1_o2_o3_o123\}$			2
	$\{12123434\}^{\dagger}$	(b)	$\{1_e2_e123_e4_e34\}$			5
	$\{12132434\}^{\dagger}$	(b)	$\{1_e2_e13_e24_e34\}$			5
	$\{12134234\}$	(c)	–			
	$\{12341234\}$	(c)	–			

…… with 5 ribbons:

{1212343545}†	(b)	$\{1_e2_e123_e4_e35_e45\}$	6
{1212345345}†	(b)	$\{1_e2_e123_o4_o5_o345\}$	5
{1212345453}†	(b)	$\{1_e2_e123_o4_e5_e453\}$	6
{1213243545}	(b)	$\{1_e2_e13_e24_e35_e45\}$	5
{1213245345}	(c)	–	
{1213245435}	(b)	$\{1_e2_e13_e24_e5_e435\}$	6
{1213423545}	(b)	$\{1_o2_e13_o4_e235_o45\}$	5
{1213424535}	(b)	$\{1_e2_e13_e4_e245_e35\}$	6
		$\{1_o2_e13_e4_o245_e35\}$	6
{1213425345}	(b)	$\{1_e2_e13_o4_o25_o345\}$	5
{1213425354}	(b)	$\{1_o2_e13_e4_o25_o354\}$	6
{1213425435}	(b)	$\{1_e2_e13_e4_e25_e435\}$	5
		$\{1_e2_e13_o4_o25_e435\}$	5
{1213452345}	(b)	$\{1_o2_e13_o4_o5_o2345\}$†	6
{1213452534}	(c)	–	
{1231245345}	(c)	–	
{1231425345}*	(c)	–	
{1231425435}*	(c)	–	
{1231435425}*	(c)	–	
{1231452345}	(c)	–	
{1234512345}	(a)	$\{1_o2_o3_o4_o5_o12345\}$	4

“H-type” pseudoknotted fold





linear diagram for “H-type” fold

Brierley, Ian, Simon Pennell, and Robert JC Gilbert. "Viral RNA pseudoknots: versatile motifs in gene expression and replication." *Nature Reviews Microbiology* 5.8 (2007): 598-610.

detected in ssRNA

All possible crossing-free linear antiparallel folds (formed by cutting the circular diagram)

	fold ID	$\lambda_{\Pi}^{\otimes}$	lin diagram	circ diagram	$\mathcal{L}_{\Pi}^{\otimes}$
H-fold	1 (*H*)	$[1_e2_e12]$			$\{1_e2_e12\}$
K-fold	2 (*K*)	$[1_e2_e13_e23]$			$\{1_e2_e12\}$
K-fold	3 (*K*)	$[1_o2_e13_o23]$			$\{1_e2_e12\}$
L-fold	4 (*L*)	$[1_o2_o3_o123]$			$\{1_o2_o3_o123\}$
?	5	$[1_e2_e123_e4_e34]^{\dagger}$			$\{1_e2_e123_e4_e34\}$
	6	$[1_e2_e13_e24_e34]$			$\{1_e2_e13_e24_e34\}$
	7	$[1_e2_e13_e4_e342]$			$\{1_e2_e123_e4_e34\}$
M-fold	8 (*M*)	$[1_e2_o3_o14_o234]$			$\{1_o2_o3_o123\}$
M-fold	9 (*M*)	$[1_o2_o3_o14_e234]$			$\{1_o2_o3_o123\}$
	10	$[1_e2_e3_e14_e342]$			$\{1_e2_e13_e24_e34\}$
	11	$[1_e2_e3_e234_e14]$			$\{1_e2_e123_e4_e34\}$
	12	$[1_e2_e3_e24_e143]$			$\{1_e2_e13_e24_e34\}$
	13	$[1_e2_e3_e24_e314]$			$\{1_e2_e13_e24_e34\}$
	14	$[1_e2_e3_e4_e2413]$			$\{1_e2_e13_e24_e34\}$
	15	$[1_e2_e3_e4_e3142]$			$\{1_e2_e13_e24_e34\}$

H-pseudoknot region in SARS-CoV-2 (nt 13418-13488 within 30000 nt genome) is *almost* knotted ….

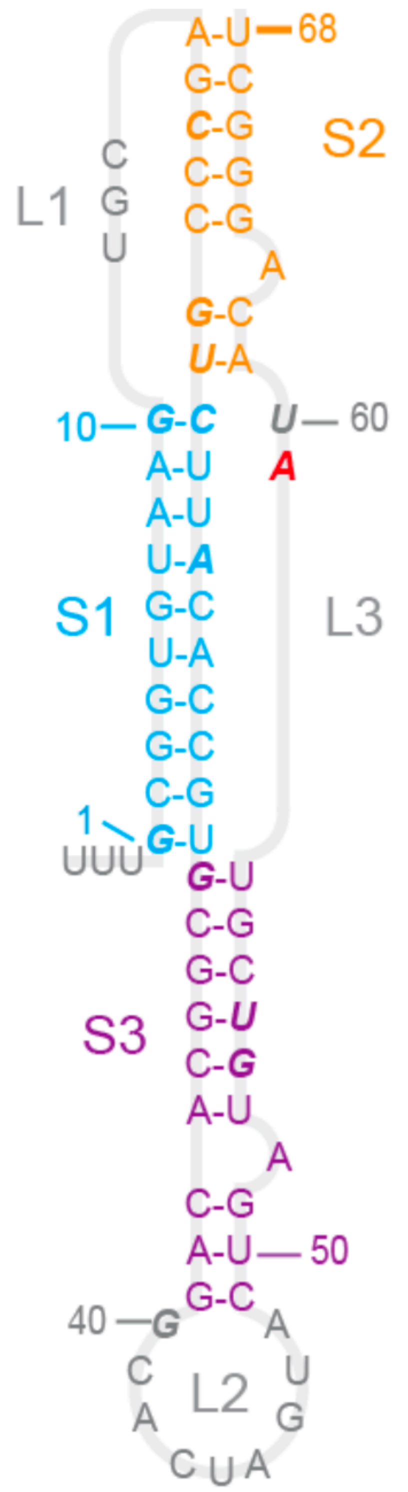
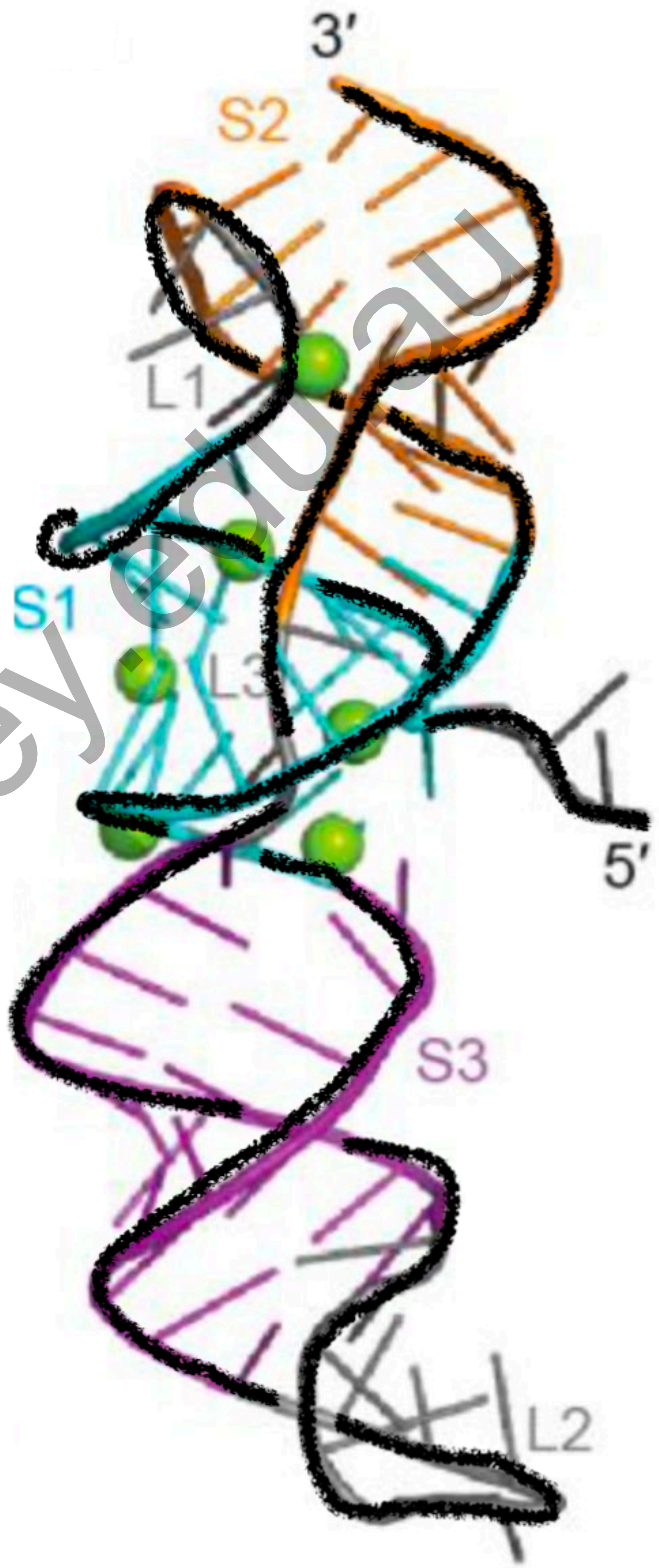


Figure 1: SARS-CoV-2 pseudoknot primary and secondary structure. The sequence is color-coded by secondary structure (S1: cyan, S2: orange, S3: purple, loops: grey). The only difference from SARS-CoV-1 is that A59 (red) is changed to C59 in the latter. Bases shown in italic are protected against nuclease digestion in SARS-CoV-1.



(presumably replication needs the unknot)

Modeling the structure of the frameshift stimulatory pseudoknot in SARS-CoV-2 reveals multiple possible conformers, Woodside et al, BioRXiv https://doi.org/10.1101/2020.06.08.141150 June 2020

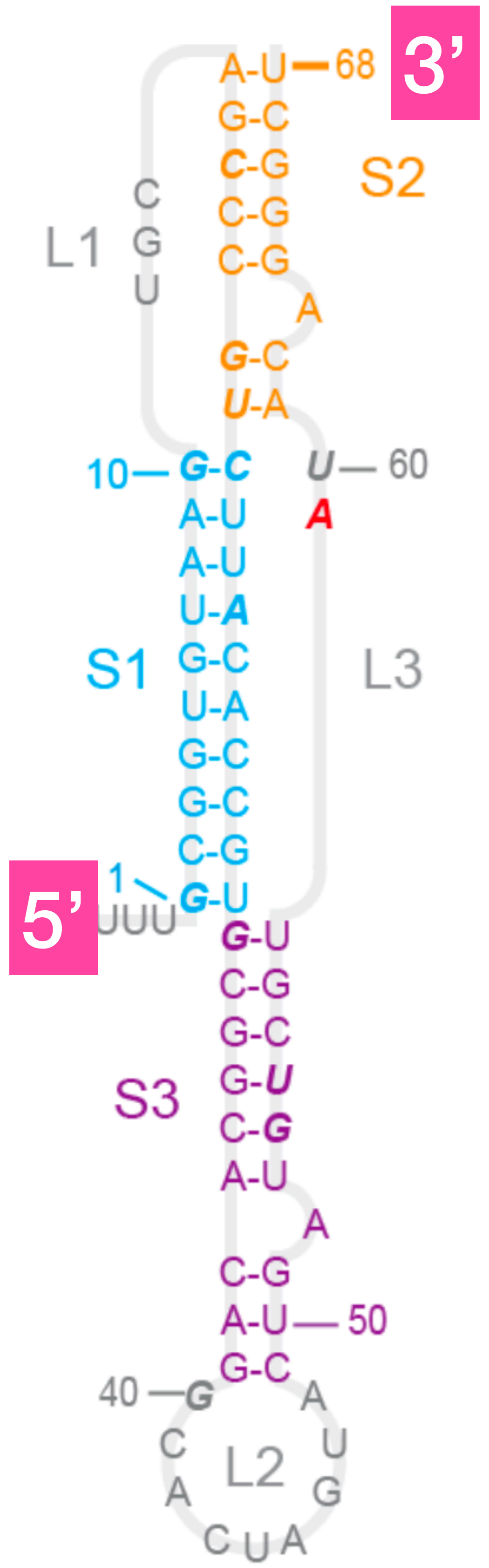
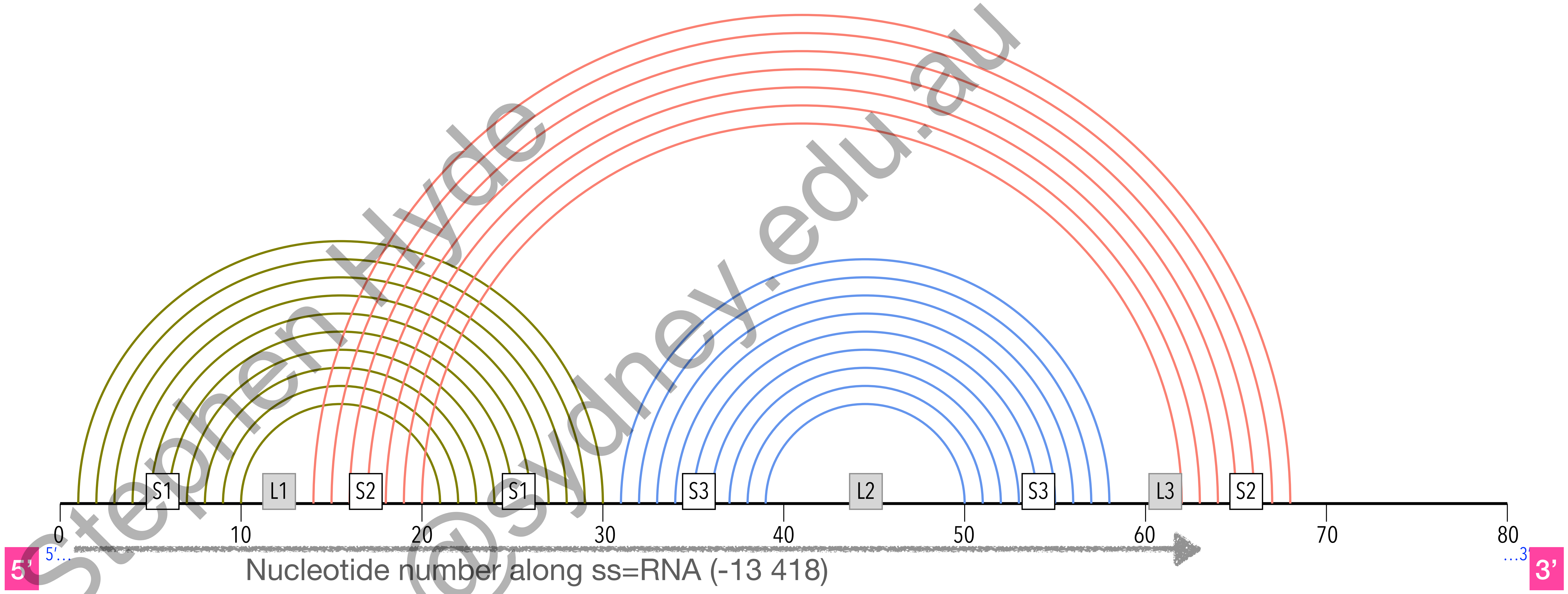


Figure 1: SARS-CoV-2 pseudoknot primary and secondary structure. The sequence is color-coded by secondary structure (S1: cyan, S2: orange, S3: purple, loops: grey). The only difference from SARS-CoV-1 is that A59 (red) is changed to C59 in the latter. Bases shown in italic are protected against nuclease digestion in SARS-CoV-1.

approx metric realisation of rectified linear fold



Woodside et al, BioRXiv https://doi.org/10.1101/2020.06.08.141150 June 2020 from metric to topological realisation of fold

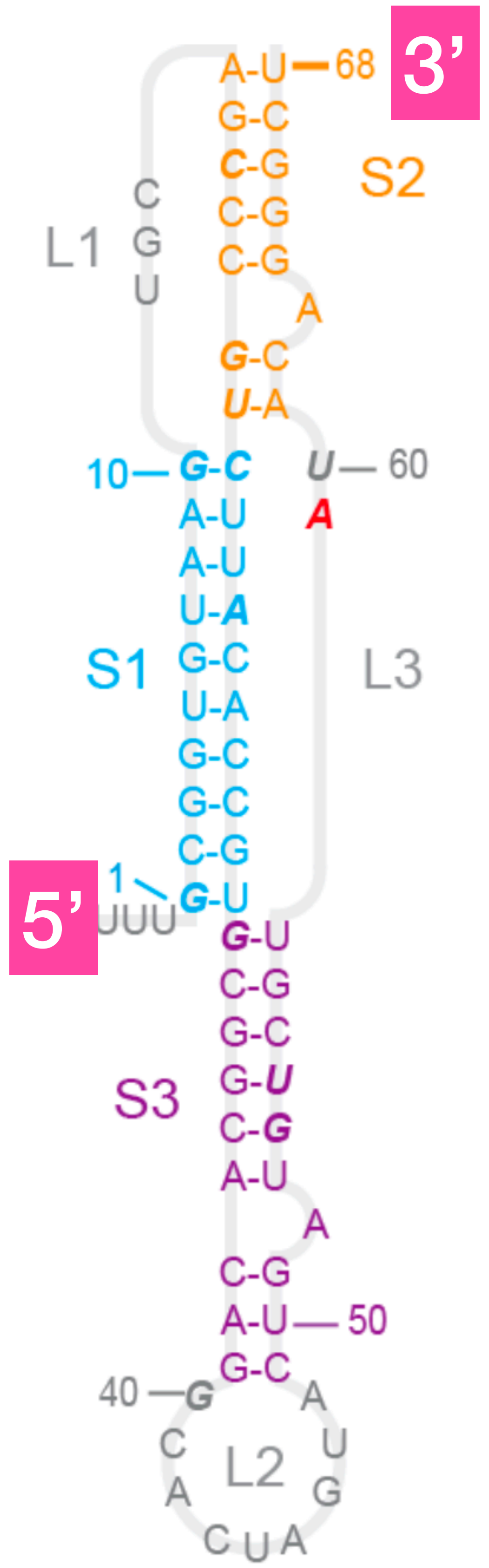
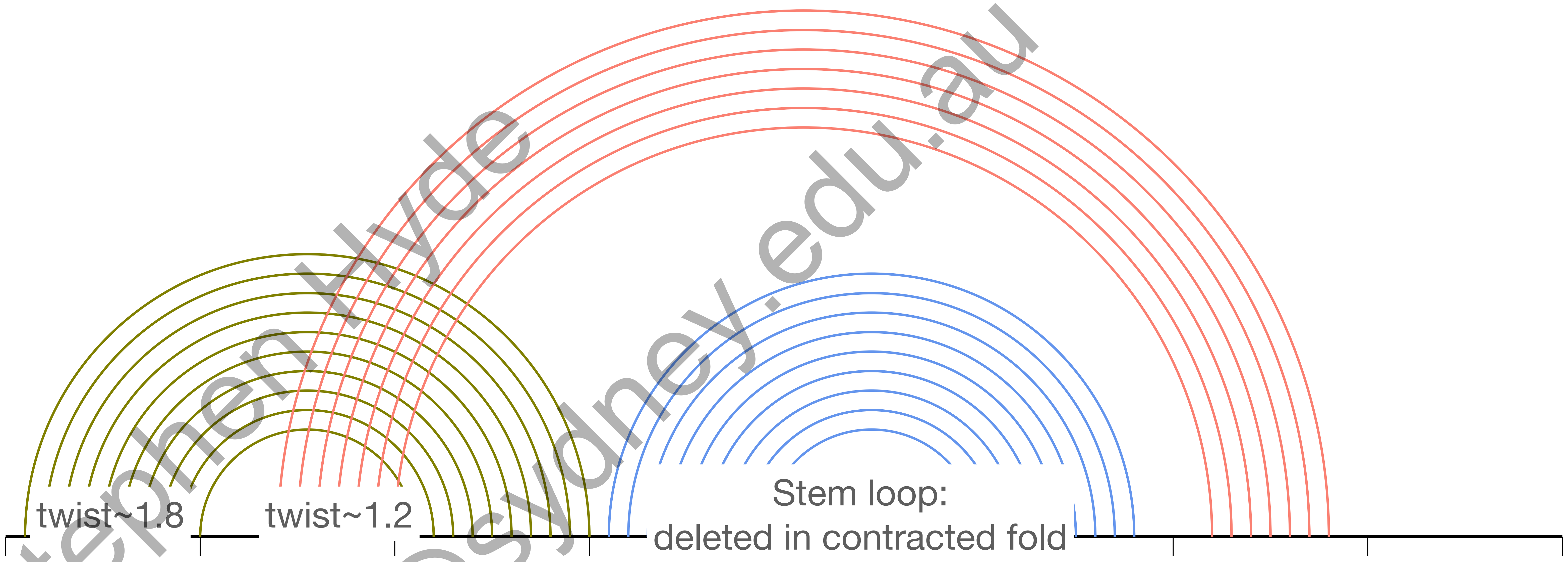


Figure 1: SARS-CoV-2 pseudoknot primary and secondary structure. The sequence is color-coded by secondary structure (S1: cyan, S2: orange, S3: purple, loops: grey). The only difference from SARS-CoV-1 is that A59 (red) is changed to C59 in the latter. Bases shown in italic are protected against nuclease digestion in SARS-CoV-1.



$$[1_{(1.8)}2_{(1.2)}12] \rightarrow [1_1 2_1 12]$$

Woodside et al, BioRXiv https://doi.org/10.1101/2020.06.08.141150 June 2020

$[1_1 2_1 12]$ is a (c)-type fold

$\{1_1 2_1 12\} \subset \{1_o 2_o 12\}$

edge-crossings

$\mathcal{L}^{\otimes}_{\amalg}$	$t_\times$	t_1	t_2	t_3	t_4	*pseudoknot*	*knot*	*knot ID*
$\{1_o 2_o 12\}$	$\bar{1}, \bar{1}$	1	1	–	–	×		
		1	3	–	–	×		
		3	1	–	–	×		
		3	3	–	–		×	3_1
	$\bar{1}, 1$	1	1	–	–	×		
		1	3	–	–	×		
		3	1	–	–		×	3_1
		3	3	–	–		×	5_2
	$1, 1$	1	1	–	–		×	3_1
		1	3	–	–		×	5_2
		3	1	–	–		×	5_2
		3	3	–	–		×	7_4

In absence of other interactions and entropy 33% chance of knotting!

…so, secondary interactions alone will induce **true knotting** in pseudoknot domain of CoV-2 RNA

likely suppressed by tertiary interactions, disfavouring +1, +1 edge-crossings in edge graph

H-pseudoknot region in SARS CoV-2 is a potential target for viral deactivation… modify the fold by nt mutations?

Schlick, Tamar, Qiyao Zhu, Swati Jain, and Shuting Yan. "Structure-Altering Mutations of the SARS-CoV-2 Frameshifting RNA Element." *Biophysical Journal* (2020).

OR, modify the fold by altering tertiary interactions?

Absence of knots in known RNA structures

Cristian Micheletti[a,1], Marco Di Stefano[a], and Henri Orland[b,c]

PNAS

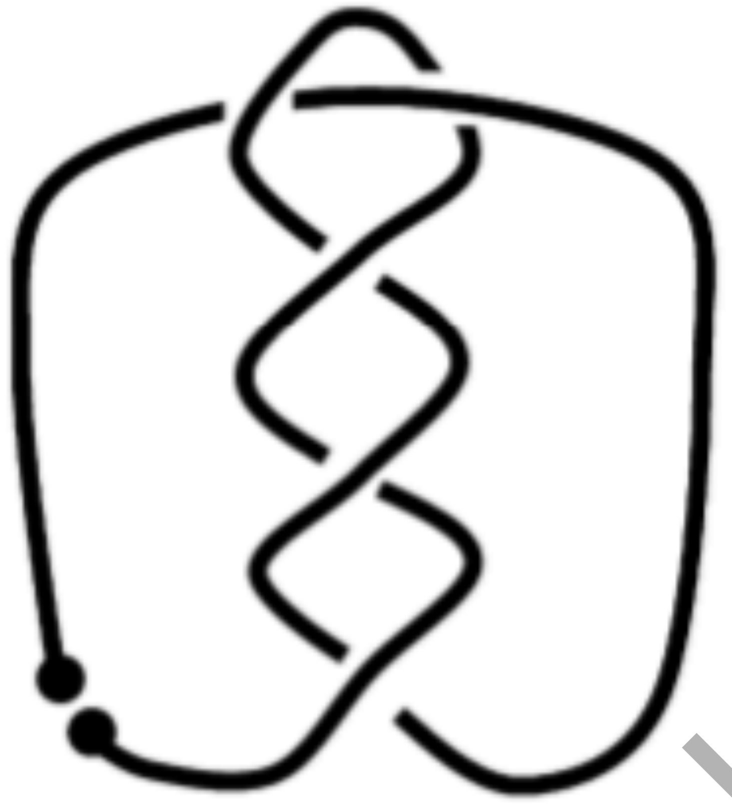


Fig. 6. Design of RNA twist knots. Twist knots (such as the shown 5_2 knot) can be formed by RNA sequences designed to fold into a helix with an unpaired loop large enough to be threaded by one of two termini. The knot could be stabilized by base pairing at the helix apex or annealing of the two complementary termini.

the concentration and type of counterions in solution that could affect both the geometry of the helix (64) and the electrostatic persistence length controlling the knot size (65).

The systematic design of twist or other types of RNA knots could be significantly aided by suitable structure prediction algorithms. As a prerequisite, these methods need to be capable of handling conformations of genus different from zero. A number of such methods based on free-energy minimization (50, 51, 66–69) or kinetic folding approaches (70) have been developed in recent years. Their predicted fold is typically encoded by a graph representation, which carries information about the succession of RNA strands contacts but is oblivious to associated sequence of strands over- and underpasses. Consequently, genuine knots and pseudoknots may share the same graph. Accordingly, we believe that a most interesting research avenue would be to extend the scope of current RNA structure–prediction methods to distinguish these different forms of entanglement.

To appear in bioRxiv
&
“The Structure of Tangles”
(Clarendon Press)