

Algebraic Systems Motivated by DNA Origami

James Garrett, Nataša Jonoska, Hwee Kim, Masahico Saito

Department of Mathematics and Statistics, University of South Florida
4202 E. Fowler Ave, Tampa, FL 33620, USA

Abstract We initiate an algebraic approach to study DNA origami structures. We identify two types of basic building blocks and describe a DNA origami structure by their composition. These building blocks are taken as generators of a monoid, called the origami monoid, and motivated by the well studied Temperley-Lieb algebras, we identify a set of relations that characterize the origami monoid. We present several observations about Green's relations for the origami monoid and study the relations to a direct product of Jones monoids, which is a morphic image of an origami monoid.

Key words: DNA origami, Temperley-Lieb algebra, Rewriting system.

1 Introduction

In the past few decades, bottom-up assemblies at the nano scale have introduced new materials and molecular scaffoldings producing structures that have wide ranging applications (e.g. [5, 8]), even materials that seem to violate standard chemistry behavior (e.g. [14]). “DNA origami”, introduced by Rothemund [11] in 2006, significantly facilitated the construction of $\sim 100 \times 100nm$ 2D DNA nanostructures. The method typically involves combining an M13 single-stranded cyclic viral molecule called *scaffold* with 200-250 short *staple strands* to produce about $100nm$ diameter 2D shapes [11], and more recently also to produce a variety of 3D constructs (e.g. [4]). Fig. 1 (left) shows a schematic of an origami structure, where the thick black line represents a portion of the cyclic vector plasmid outlining the shape, and the colored lines are schematics of the short strands that keep the cyclic molecule folded in the shape. Because the chemical construction of DNA origami is much easier than previous methods, this form of DNA nanotechnology has become popular, with perhaps 300 laboratories in the world today focusing on it.

Although numerous laboratories around the world are successful in achieving various shapes with DNA origami, theoretical understanding and characterizations of these shapes is still lacking. With this paper we propose an algebraic system to describe and investigate DNA origami structures. The staple strands usually have 2-4 segments of about 8 bases joining 2-3 locations (folds) of the scaffold. All cross-overs between two staple strands and between two neighboring folds of the scaffold are antiparallel. We divide the DNA origami structure to local scaffold-staples interactions and to such local interactions we associate a

generator of a monoid which we call an *origami monoid*. The origami monoid we present here is closely related to the Jones monoid [2, 3] which is a monoid variant of the well studied Temperley-Lieb algebra [1]. We show that a DNA origami structure can be associated to an element of an origami monoid and propose a set of rewriting rules that are plausible for DNA segments to conform in DNA origami. The number of generators of an origami monoid depends on the number of parallel folds of the scaffold in the DNA origami. We observe that a direct product of two Jones monoids is a surjective image of an origami monoid, and we study the structure of the origami monoids through Green's relations. We characterize the origami monoids for small number of scaffold folds and propose several conjectures for general origami monoids.

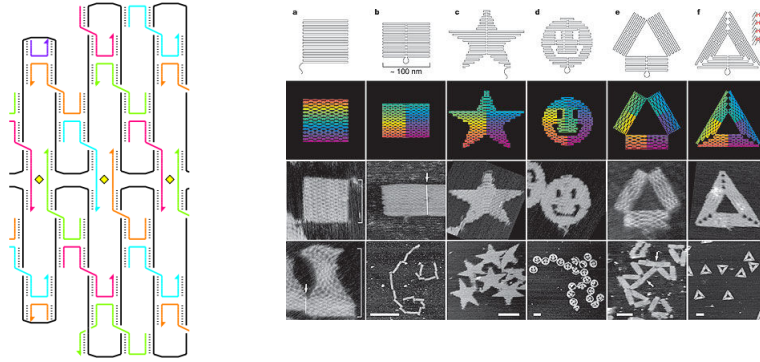


Fig. 1: (Left) A schematic figure of DNA origami structure with scaffold in black and staples in color (edited from [11]). (Right) Various shapes made by DNA origami, from [11]. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature (Folding DNA to create nanoscale shapes and patterns, Paul W. K. Rothemund), [4561420919172] 2006

2 Preliminaries

2.1 Jones monoids

Temperley-Lieb algebras have been used in many fields, particularly in physics and knot theory (see, for example, [1, 2, 7, 9]). The Jones monoid \mathcal{J}_n is derived from the Temperley-Lieb algebras and is defined with generators and relations as follows [2, 9]. The monoid \mathcal{J}_n is generated by h_i , $i = 1, \dots, n-1$, and has relations

$$(B) \ h_i h_j h_i = h_i \text{ for } |i-j| = 1, \ (C) \ h_i h_i = h_i \ (D) \ h_i h_j = h_j h_i \text{ for } |i-j| \geq 2.$$

The elements of \mathcal{J}_n may be represented as planar diagrams with non-crossing lines connecting n points at the top and n points at the bottom of the diagrams. The diagram for the generator h_i is depicted in Fig. 2 (A) [9]. For each h_i ,

parallel vertical lines connect the top j th and bottom j th points ($j \neq i, i+1$) of the diagram for all but the i th and $(i+1)$ st points, while the top i th and $(i+1)$ st points are connected, and the bottom i th and $(i+1)$ st points are connected. Multiplication of two elements is represented by concatenation of diagrams, placing the diagram of the first element on top of the second, and removing closed loops. The diagrammatic representation of the monoid relations are depicted in Fig. 2 (B), (C) and (D). More details can be found in [2, 9].

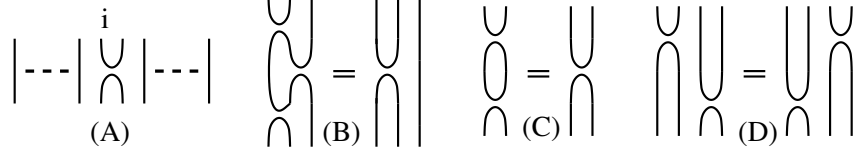


Fig. 2: The generators (A) and relations (B,C,D) of the Jones monoid

2.2 String rewriting systems

An alphabet Σ is a non-empty finite set of symbols. A word over Σ is a finite sequence of elements (symbols) from Σ , and Σ^* is the set of all words over Σ . This set includes the empty string, the word containing no symbols, often written as 1. A word u is called a *factor of a word* v if there exist words x and y , which may be empty, such that $v = xuy$. Note that this is also sometimes referred to as a subword.

A string rewriting system, (Σ, R) consists of an alphabet Σ and a set of rewriting rules, R , which is a binary relation on Σ^* . An element (x, y) of R is called a rewriting rule, and is written $x \rightarrow y$. We extend R to factors of words \xrightarrow{R} , where for any $s, t \in \Sigma^*$, $s \xrightarrow{R} t$ if there exist $x, y, u, v \in \Sigma^*$ such that $s = uxv$, $t = uyv$, and $x \rightarrow y$. We also write $s \rightarrow t$ for simplicity if no confusion arises.

If there is a sequence of words $u = x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_n = v$ in a rewriting system (Σ^*, R) , we write $u \rightarrow_* v$. An element $x \in \Sigma^*$ is *confluent* if for all $y, z \in \Sigma^*$ such that $x \rightarrow_* y$ and $x \rightarrow_* z$, there exists $w \in \Sigma^*$ such that $y \rightarrow_* w$ and $z \rightarrow_* w$. If all words in Σ^* are confluent, then (Σ^*, R) is called *confluent*. In particular, if R is symmetric, then the system (Σ^*, R) is confluent.

2.3 Monoids and Green's relations

A monoid is a pair (M, \cdot) where M is a set and \cdot is an associative binary operation on M that has an identity element 1. The set Σ^* is a (free) monoid generated by Σ with word concatenation as the binary operation, and the empty string as the identity element. Presentations of monoids are defined from the free monoid in a manner similar to presentations of groups. Rewriting systems define the monoid relations by taking the equivalence closure of the rewriting rules, which makes the rewriting system confluent.

For a monoid M , the *principal left (resp. right) ideal* generated by $a \in M$ is defined by $Ma = \{xa \mid x \in M\}$ (resp. aM), and the *principal two-sided ideal*

s MaM . Green's relations \mathcal{L} , \mathcal{R} , and \mathcal{J} are defined for $a, b \in M$ by $a\mathcal{L}b$ if $Ma = Mb$, $a\mathcal{R}b$ if $aM = bM$ and $a\mathcal{J}b$ if $MaM = MbM$. Green's \mathcal{H} relation is defined by $a\mathcal{H}b$ if $a\mathcal{L}b$ and $a\mathcal{R}b$. Green's \mathcal{D} relation is defined by $a\mathcal{D}b$ if there is c such that $a\mathcal{L}c$ and $c\mathcal{R}b$. The equivalence classes of \mathcal{L} are called \mathcal{L} -classes, and similarly for the other relations. In a finite monoid, \mathcal{D} and \mathcal{J} coincide. The \mathcal{D} -classes can be represented in a matrix form called *egg boxes*, where the rows represent \mathcal{R} -classes, columns \mathcal{L} -classes, and each entry is an \mathcal{H} -class. See [10] for more details.

Example 1. In [3], \mathcal{D} -classes are obtained for Jones and related monoids. Here we include an example of a \mathcal{D} -class of \mathcal{J}_3 , which has a \mathcal{D} -class consisting of the identity element and another class of the (2×2) -matrix below, where each element is an \mathcal{H} -class, which in this case are singletons:

$$\begin{bmatrix} h_1 & h_1h_2 \\ h_2h_1 & h_2 \end{bmatrix}$$

where rows $\{h_1, h_1h_2\}$, $\{h_2h_1, h_2\}$, are the \mathcal{R} -classes and columns $\{h_1, h_2h_1\}$, $\{h_1h_2, h_2\}$ are the \mathcal{L} -classes in this \mathcal{D} -class. For instance, we see that multiplying h_1 and h_1h_2 by h_i to the right gives rise to the same right ideal.

3 Origami monoid \mathcal{O}_n

3.1 Generators

Here we identify simple building blocks in DNA origami structures. With each block type we associate a generator of a monoid and derive string rewriting systems to describe DNA structures. We have two motivations for our choices. (1) In Fig. 1 (left), one notices repeated patterns of simple building blocks whose concatenation builds a larger structure. One type of these patterns is a cross-over by the staple strands, and the other is a cross-over of the scaffold strand. Thus, a natural approach to describe DNA origami structures symbolically is to associate generators of an algebraic system to simple building blocks, and to take multiplication in the system to be presented as concatenation of the blocks. (2) In knot theory, a knot diagram is decomposed into basic building blocks of crossings or tangles. For the Kauffman bracket version of the Jones polynomial [7], for example, whose generators resemble building blocks observed in Fig. 1 (left), are used.

For a positive integer n we define a monoid \mathcal{O}_n , where n represents the number of vertical double stranded DNA strands, that is, n is the number of parallel folds of the scaffold. For the structure in Fig. 1, $n = 6$. The generators of \mathcal{O}_n are denoted by α_i (corresponds to anti-parallel staple strands cross-over) and β_i (corresponds to antiparallel scaffold strand cross-over) for $i = 1, \dots, n-1$, as depicted in Fig. 3. The subscript i represents the position of the left scaffold corresponding to α_i and β_i , respectively, by starting at 1 from the leftmost scaffold strand fold and counting right (Fig. 4).

Because DNA is chemically oriented, and the strands in the double stranded DNA are oppositely oriented, we define an orientation within the building blocks corresponding to generators. Because parallel scaffold strands are obtained by folding of the scaffold, consecutive scaffold strands run in alternating directions, while staple strands run in the opposite direction to the scaffold, and for convention we take that the first scaffold runs in an upwards direction. In this way, the direction of the scaffold/staple strands for any particular α_i or β_i depends entirely on the parity of i , as shown in Fig. 3.

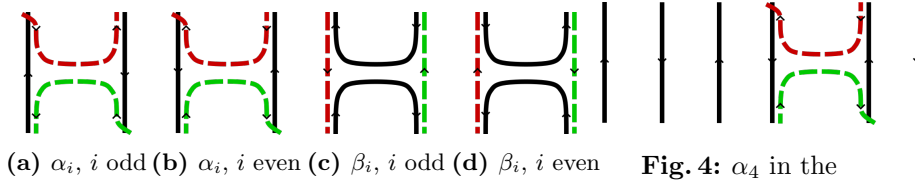


Fig. 3: The generators identified

Fig. 4: α_4 in the context of a 6-fold stranded structure

Fig. 4 shows a diagram corresponding to α_4 as an example of the “full picture” of one of these generators. For the sake of brevity, we neglect to draw the extra scaffold and staple strands in most diagrams, but it may be helpful to imagine them when we describe their concatenation. In addition, we often use α_i and β_i to refer to the corresponding diagrams. As in Fig. 4, parallel scaffolds in generator diagrams do not have counterpart parallel staples.

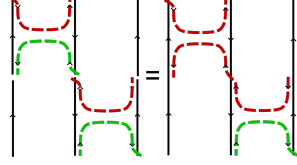
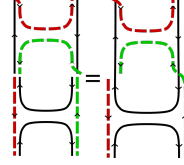
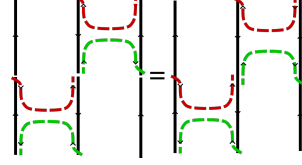
3.2 Concatenation as a monoid operation

To justify modeling DNA origami structures by words over the generators we make a correspondence between concatenations of generators α_i, β_i and concatenations of DNA segments. For a natural number $n \geq 2$, the set of generators of the monoid \mathcal{O}_n is the set $\Sigma_n = \{\alpha_1, \alpha_2, \dots, \alpha_{n-1}, \beta_1, \beta_2, \dots, \beta_{n-1}\}$. For a product of two generators x_i and y_j in Σ_n , we place the diagram of the first generator above the second, lining up the scaffold strings of the two generators, and then we connect each respective scaffold string. If the two generators are adjacent, that is, if for indices i and j it holds $|i - j| \leq 1$, then we also connect their staples as described below. Otherwise, if $|i - j| \geq 2$, no staple connection is performed and the concatenation is finished.

We define a convention of connecting staples for adjacent generators, which is motivated by the manner in which staples connect in Fig. 1. Note how the staples of α -type protrude “outside” of the scaffold in Fig. 3. We refer to these ends of a staple as an “extending staple-ends”, and all other staple ends as “non-extending staple-ends”. We connect staples everywhere *except* when two non-extending staple-ends would have to cross a scaffold to connect (recall that the scaffold strands are connected first), as can be seen in Fig. 6 and Fig. 7.

Our choice of coloring staples in the figure is arbitrary, and we re-color staples in the same color if they get connected when we concatenate generators. By exhausting all possibilities, one can see that under our convention of connection,

the staples remain short by concatenation without joining more than three scaffold folds. Note that concatenation of three or more generators is associative because generators can be connected in an associative manner following the rules described above.

Fig. 5: $\alpha_i \alpha_{i+1}$, i oddFig. 6: $\alpha_i \beta_i$, i oddFig. 7: $\alpha_i \alpha_{i-1}$, i odd

3.3 Relations in \mathcal{O}_n

The rewriting rules (which generate the relations within the monoids) are motivated by similarity between the DNA origami structures as seen in Fig. 1(left) and the diagrams of Jones monoids in Fig. 2. It is deemed that the relations of Jones monoids simplify the DNA origami structure, and may be useful for designing efficient and more solid structures by the rewriting rules proposed below. The figures in this section are for justifying feasibility of corresponding DNA structures, and to represent the rewriting system diagrammatically.

Rewriting rules. For $\Sigma_n = \{\alpha_1, \alpha_2, \dots, \alpha_{n-1}, \beta_1, \beta_2, \dots, \beta_{n-1}\}$, we establish a set of rewriting rules that allows simplification of the DNA structure description. Define a string rewriting system (Σ_n, R) as follows.

To ease the notation, we define a bar on Σ_n by $\overline{\alpha_i} = \beta_i$ and $\overline{\beta_i} = \alpha_i$, and extend this operation to the free monoid by defining \overline{w} for a word w by applying bar to each letter of w . Let $\gamma \in \{\alpha, \beta\}$ and $i \in \{1, \dots, n-1\}$, then we have:

- (1) (Idempotency) $\gamma_i \gamma_i \rightarrow \gamma_i$
- (2) (Left Jones relation) $\gamma_i \gamma_{i+1} \gamma_i \rightarrow \gamma_i$
- (3) (Right Jones relation) $\gamma_i \gamma_{i-1} \gamma_i \rightarrow \gamma_i$
- (4) (Inter-commutation) $\gamma_i \overline{\gamma_j} \rightarrow \overline{\gamma_j} \gamma_i$, for $|i-j| \geq 1$
- (5) (Intra-commutation) $\gamma_i \gamma_j \rightarrow \gamma_j \gamma_i$, for $|i-j| \geq 2$

The rules are extended to Σ_n^* as described in Section 2.2.

The rewriting rules are inspired by Jones monoids, and they are also reflected in the reality of the diagrams of DNA origami, as shown in Figs. 8 and 9. Specifically, a pattern in the left of Fig. 8 (a) has a small staple circle, which is deemed to be simplified by the right side. Staple strands are holding the scaffold in a certain position (obtained to the right of the arrow) and the cyclic staple only reinforces the structure. The small circle of a scaffold in Fig. 8 (b) left cannot form in DNA origami, and therefore is simplified to the structure on the right of the arrow.

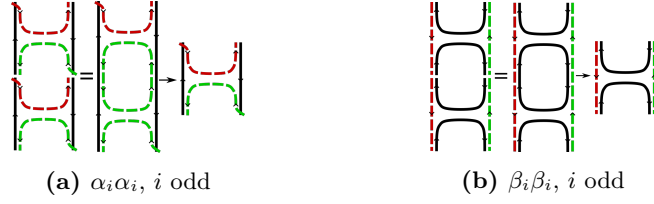


Fig. 8: Two examples of idempotency

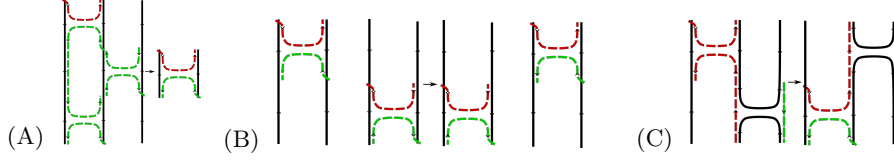


Fig. 9: Examples of (A) Jones relation, (B) Inter-commutation, and (C) Intra-commutation

Deriving additional rewriting rules by substitution. Since DNA origami structure has no internal scaffold loops, applying rewriting rules similar to (1)–(5) to concatenations of generators, that is, products of α 's and β 's, is plausible for DNA origami structures. We extend these rules to more general substitution rules for our specific case of generators α_i and β_i by considering other γ 's, for instance $\gamma = \alpha\beta$. The composition diagrams show that such substitution rules describe the DNA origami staples/scaffold structure in the way we proposed above (see Fig. 10), while these new structures produce rules that cannot be derived from the listed ones in (1)–(5). Therefore we consider rewriting rules for concatenations of generators α 's and β 's. Furthermore, we focus on concatenations of generators with the same or 'neighboring' indexes, because only for these generators can the ends of the staples connect. However, α_i and β_j ($i \neq j$) can commute freely (by the inter-communication rule (4)), so we also do not need to consider substitutions such as $\gamma = \alpha_i \beta_{i+1}$. Further, we observe that by setting $\gamma \in \{\alpha_i \beta_i, \beta_i \alpha_i\}$, the idempotency rule (1) holds as seen in Fig. 10. Therefore there are only four cases to consider: $\gamma \in \{\alpha_i \beta_i, \beta_i \alpha_i, \alpha_i \beta_i \alpha_i, \beta_i \alpha_i \beta_i\}$ and check the plausibility of corresponding DNA diagrams.

First, consider $\gamma \in \{\alpha\beta, \beta\alpha\}$, where γ_i indicates $\alpha_i \beta_i$. Then substituting γ into rewriting rules (1), (2), and (3) gives us new rewriting rules (1a), (2a), and (3a). For example, (1a) consists of $\alpha_i \beta_i \alpha_i \beta_i \rightarrow \alpha_i \beta_i$ and $\beta_i \alpha_i \beta_i \alpha_i \rightarrow \beta_i \alpha_i$. Note that a provisional rewriting rule (5a) could easily be obtained by the rewriting rule (5), so we do not consider it as a new rule. We also do not add rewriting rule (4a) since it conflicts with the structure of the scaffold, as shown in Fig. 11. Notice that the scaffold strand at the top left is connected to the second strand only on the left side of the figure, and on the right hand side of the figure it is connected to the third strand. Next we consider $\gamma \in \{\alpha\beta\alpha, \beta\alpha\beta\}$, which gives us rewriting rules (1b), (2b) and (3b). Similarly as before, rules (4b) and (5b) are not added, (4b) because of incompatible staple strands, and (5b) because it can

be derived from (5). In addition, (1b) can also be derived from (1) and (1a), so it is not considered as a new rule. In the end, we are left with 10 rewriting rules which we use to define the general rewriting rules and the monoids.

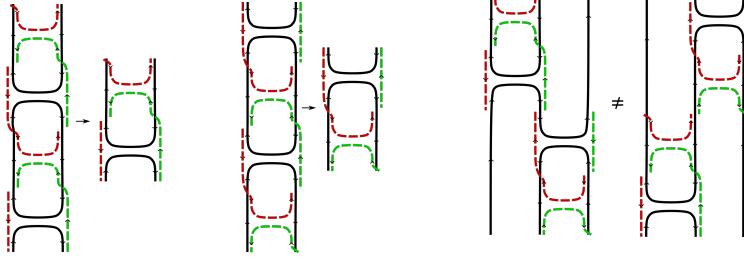


Fig. 10: Substitution of $\alpha\beta$ and $\beta\alpha$ (resp.) into the first rewriting rule (i odd) **Fig. 11:** Substitution of $\gamma = \alpha\beta$ into rewriting rule (4) for i odd.

Definition 1. The *origami monoid* \mathcal{O}_n is the monoid with a set of generators Σ_n and relations generated by the rewriting rules (1) through (5), (1a), (2a), (3a), (2b), (3b).

4 Monoid structure of \mathcal{O}_n

In this section, we present computational results on Green's \mathcal{D} -classes and compare them to those for the Jones monoids obtained in [3]. For comparison, we use the monoid epimorphism from \mathcal{O}_n to the product $\mathcal{J}_n \times \mathcal{J}_n$ defined below.

Let \mathcal{J}_n be the Jones monoid of degree n with generators h_i , $i = 1, \dots, n-1$. We denote the submonoid of \mathcal{O}_n generated by α s (resp. β s), by \mathcal{O}_n^α (resp. \mathcal{O}_n^β). An equivalent description for \mathcal{O}_n^α is the set of all words consisting of only α s (plus the empty word), and similarly for \mathcal{O}_n^β . Let $\mathcal{O}_n^{\alpha\beta} = [\mathcal{O}_n \setminus (\mathcal{O}_n^\alpha \cup \mathcal{O}_n^\beta)] \cup \{1\}$.

Lemma 1. $\mathcal{O}_n^{\alpha\beta}$ is a submonoid of \mathcal{O}_n .

Proof. The left and right hand sides of each rewriting rule show that rewriting a word by these rules does not change the absence, or existence of at least one α in the word, and similarly for β . Thus multiplication of two words in $\mathcal{O}_n^{\alpha\beta}$ does not remove α 's or β 's from the product, hence the product remains in $\mathcal{O}_n^{\alpha\beta}$. \square

Let $p_\alpha : \mathcal{O}_n \rightarrow \mathcal{J}_n$ be the epimorphism defined by 'projections' $p_\alpha(\alpha_i) = h_i$ and $p_\alpha(\beta_i) = 1$, for all $i = 1, \dots, n-1$, and let p_β be defined similarly for β s. Define $p : \mathcal{O}_n \rightarrow \mathcal{J}_n \times \mathcal{J}_n$ by $p(x) = (p_\alpha(x), p_\beta(x))$ for $x \in \mathcal{O}_n$. Since the monoid relations of \mathcal{O}_n hold under p , we have the following:

- $\mathcal{O}_n^\alpha \cong \mathcal{O}_n^\beta \cong \mathcal{J}_n$.
- The map $p : \mathcal{O}_n \rightarrow \mathcal{J}_n \times \mathcal{J}_n$ is a surjective monoid morphism.

In particular, it follows that the order of \mathcal{O}_n is at least $|\mathcal{J}_n|^2$.

4.1 Orders of origami monoids

For $n = 2$ we can determine the order of \mathcal{O}_2 as follows.

Lemma 2. *Every non-empty element of \mathcal{O}_2 can be represented by the rewriting rules as one of the following words: α_1 , β_1 , $\alpha_1\beta_1$, $\beta_1\alpha_1$, $\alpha_1\beta_1\alpha_1$, or $\beta_1\alpha_1\beta_1$.*

Proof. Since $\Sigma_2 = \{\alpha_1, \beta_1\}$, we list the words of length 3 or less exhaustively. After applying rewriting rules to these words, they reduce to those words listed in the statement.

Now consider a word w with length greater than 3. We show that w can be reduced to a word with length 3 or less. If $\alpha_1\alpha_1$ or $\beta_1\beta_1$ are factors of w , we reduce them to α_1 or β_1 , respectively. Repeating this process, we may assume that w is an alternating sequence of α_1 and β_1 . Since $\alpha_1\beta_1$ and $\beta_1\alpha_1$ are idempotent, w reduces to a word of length less than 4. \square

It is known that the elements of the Jones monoid \mathcal{J}_n are in bijection with the linear chord diagrams obtained from the arcs of the diagrams representing them, and the total number of such chord diagrams is equal to the Catalan number $C_n = \frac{1}{n+1} \binom{2n}{n}$ [2]. Thus the numbers of elements of \mathcal{J}_n for $n = 2, \dots, 6$ are 2, 5, 14, 42, 429, respectively. GAP computations show that the number of non-identity elements in \mathcal{O}_3 , \mathcal{O}_4 , \mathcal{O}_5 and \mathcal{O}_6 are 44, 293, 2179, 19086 respectively [12]. This sequence of integers is not listed in the OEIS [13] list of sequences. We observe that the orders of origami monoids are much larger. In fact it is not apparent from the definition whether they are all finite. Thus we conjecture the following.

Conjecture 1. The order of \mathcal{O}_n is finite for all n .

4.2 Green's classes

We have the following observations for Green's classes of \mathcal{O}_n for general n .

Lemma 3. *Let $x \in \mathcal{O}_n^\alpha$, $y \in \mathcal{O}_n^{\alpha\beta}$ be nonempty words and let D_x and D_y be the \mathcal{D} -classes containing x and y , respectively. Then $D_x \neq D_y$.*

Proof. By Lemma 1, if $y\mathcal{L}a$, then $a \in \mathcal{O}_n^{\alpha\beta}$, and if $a\mathcal{R}b$, then $b \in \mathcal{O}_n^{\alpha\beta}$. Thus we cannot have $y\mathcal{D}x$. \square

Corollary 1. *The conclusion of Lemma 3 holds for $x \in \mathcal{O}_n^\beta$, $y \in \mathcal{O}_n^{\alpha\beta}$ and $x \in \mathcal{O}_n^\alpha$, $y \in \mathcal{O}_n^\beta$.*

Remark 1. It follows from the definition of p that every \mathcal{D} -class of \mathcal{O}_n maps into a \mathcal{D} -class of $\mathcal{J}_n \times \mathcal{J}_n$, and by Lemma 1.4 Ch. 5 in [6] the map is also onto. Also, if \mathcal{O}_n is finite, then each \mathcal{D} -class of $\mathcal{J}_n \times \mathcal{J}_n$ is an image of a \mathcal{D} -class of \mathcal{O}_n by p . We conjecture that there is in fact a one-to-one correspondence between the \mathcal{D} -classes of \mathcal{O}_n and those of $\mathcal{J}_n \times \mathcal{J}_n$. We show that this observation is true for $n \leq 6$.

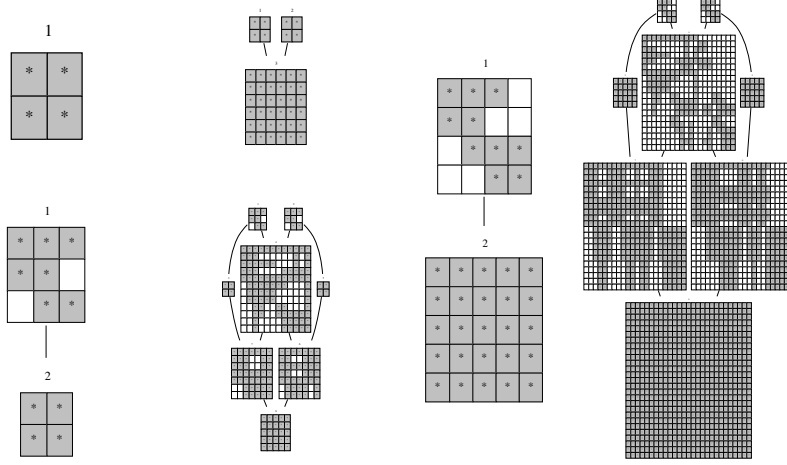


Fig. 12: \mathcal{D} -classes of \mathcal{J}_n (left) and \mathcal{O}_n (right) for $n = 3$ (top left), $n = 4$ (bottom left), and $n = 5$ (right)

4.3 Green's classes for $n \leq 6$

Green's relations for \mathcal{J}_n have been studied in [7]. We show results of GAP computations that determine \mathcal{D} -classes of origami monoids \mathcal{O}_n for $n \leq 6$; the structures are presented in Figs. 12 and 13. Shaded squares represent \mathcal{H} -classes which contain an idempotent. We note that for $n \leq 6$, every \mathcal{H} -class of \mathcal{O}_n is singleton, so each square in the figure represents precisely one element of \mathcal{O}_n .

For $n \leq 6$, since \mathcal{O}_n is finite, the \mathcal{J} and \mathcal{D} relations coincide. A preorder $\leq_{\mathcal{D}}$ is defined on \mathcal{O}_n by $a \leq_{\mathcal{D}} b$ if the two-sided principal ideal generated by a is a subset of the two-sided principal ideal generated by b . This condition is equivalent to the existence of $x, y \in \mathcal{O}_n$ such that $xya = b$. Since any two elements of a \mathcal{D} -class generate the same principal ideal, this preorder may be extended to the set of \mathcal{D} -classes of \mathcal{O}_n such that $D \leq_{\mathcal{D}} D'$ if for $a \in D$ and $b \in D'$, $a \leq_{\mathcal{D}} b$. The lines between \mathcal{D} -classes in the figures represent this preorder.

The relations between \mathcal{O}_n and \mathcal{J}_n described in Section 6.2 can be observed in Fig. 12. We omit the \mathcal{D} -class consisting of only the empty word from the diagrams, which is maximal among \mathcal{D} -classes. For each n , two copies of the \mathcal{D} -classes of \mathcal{J}_n can be found as the \mathcal{D} -classes of \mathcal{O}_n^α and \mathcal{O}_n^β , respectively, in the \mathcal{D} -classes of \mathcal{O}_n . As described in Remark 1, these correspond to the direct product of one identity and one non-identity \mathcal{D} -class of \mathcal{J}_n . The other \mathcal{D} -classes are those of $\mathcal{O}_n^{\alpha\beta}$, and correspond to the direct product of two non-identity \mathcal{D} -classes of \mathcal{J}_n . Which pair of \mathcal{D} -classes of \mathcal{J}_n correspond to which \mathcal{D} -class of \mathcal{O}_n can be better seen in Fig. 13.

In Fig. 13, we arrange the \mathcal{D} -classes of \mathcal{O}_6 to better illustrate the relation between the \mathcal{D} -classes of \mathcal{I}_n , although the same process may be applied to other n . On the right, the preorder of the \mathcal{D} -classes remains, applying left-to-right as well as top-to-bottom. The \mathcal{D} -classes along the top row and left column are the \mathcal{D} -classes of \mathcal{O}_6^α and \mathcal{O}_6^β respectively, which as previously described are isomorphic to \mathcal{I}_n . For any \mathcal{D} -class of $\mathcal{O}_n^{\alpha\beta}$, the \mathcal{D} -classes which it maps onto are greater in the preorder. Thus the grid of \mathcal{D} -classes may be thought of as a table, with the row and column of any entry determining the image of the \mathcal{D} -class by p_α and p_β , respectively. Since rewriting rules are equivalent for α and β , the \mathcal{D} -classes are symmetric up to switching rows and columns. This can be easily seen in the \mathcal{D} -classes in the upper right and lower left corners. However, the rows and columns of any \mathcal{D} -class may be ordered arbitrarily, and are automated by GAP, making the symmetry non-obvious for other \mathcal{D} -classes.

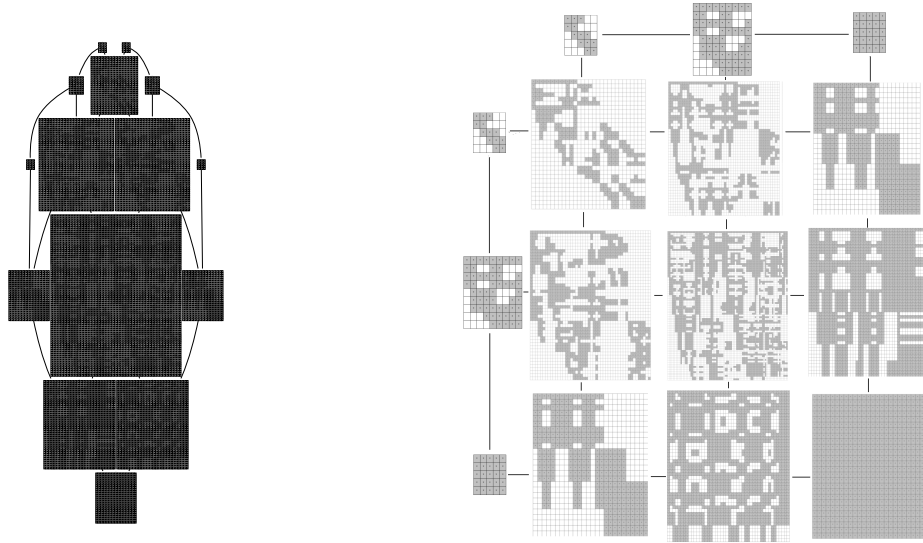


Fig. 13: \mathcal{D} -classes of \mathcal{O}_6 (left) and re-arranged and resized to fit the grid (right)

5 Concluding remarks

In this paper, motivated by similarity to Temperley-Lieb algebras, we introduced an algebraic system that describes DNA origami structures. Generators in this system are defined such that they mimic basic building blocks of DNA origami. Following the structural properties of the DNA origami, we established rewriting rules, as well as monoids whose elements conform to the relations obtained from these rules. To each DNA origami structure we can associate an element from an appropriate monoid. For example, the structure in Fig. 1 corresponds to the

element represented by the word $\alpha_1\alpha_3\alpha_5\beta_2\beta_4$. We hope that such representations of DNA origami may provide a tool for distinguishing constructs.

The monoids introduced here are connected to Jones monoids, and we provide several conjectures with the goal of relating them to known monoids. For example, from our findings for $n \leq 6$, we conjecture that \mathcal{O}_n are finite for all n , and \mathcal{H} -classes are singletons. We also provide conjectures relating to the \mathcal{D} -classes of \mathcal{O}_n and \mathcal{J}_n under the morphism p . Specifically, we conjecture that the \mathcal{D} -classes of \mathcal{O}_n are in one-to-one correspondence with the \mathcal{D} -classes of $\mathcal{J}_n \times \mathcal{J}_n$.

Acknowledgment

This work is partially supported by NIH R01GM109459, and by NSF's CCF-1526485 and DMS-1800443. This research was also partially supported by the Southeast Center for Mathematics and Biology, an NSF-Simons Research Center for Mathematics of Complex Biological Systems, under National Science Foundation Grant No. DMS-1764406 and Simons Foundation Grant No. 594594.

References

1. S. Abramsky. Temperley-Lieb algebra: From knot theory to logic and computation via quantum mechanics. *CoRR*, abs/0910.2737, 2009.
2. M. Borislavljević, K. Došen, and Z. Petric. Kauffman monoids. *Journal of Knot Theory and its Ramifications*, 11(2):127–143, 2002.
3. I. Dolinka and J. East. The idempotent-generated subsemigroup of the Kauffman monoid. *Glasgow Mathematical Journal*, 59(3):673–683, 2017.
4. S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, 459:414–418, 2009.
5. A. K. Geim and K. S. Novoselov. The rise of graphene. *Nature Materials*, 6:183–191, 2007.
6. P. A. Grillet. *Semigroups: An Introduction to the Structure Theory*. CRC Press, 1995.
7. L. H. Kauffman. *Knots and Physics*. World Scientific, 2001.
8. H. Kim, S. Yang, S. R. Rao, S. Narayanan, E. A. Kapustin, H. Furukawa, A. S. Umans, O. M. Yaghi, and E. N. Wang. Water harvesting from air with metal-organic frameworks powered by natural sunlight. *Science*, 356(6336):430–434, 2017.
9. K. W. Lau and D. G. FitzGerald. Ideal structure of the Kauffman and related monoids. *Communications in Algebra*, 34(7):2617–2629, 2006.
10. J. E. Pin. *Varieties of Formal Languages*. North Oxford Academic Publishers, 1986.
11. P. W. K. Rothmund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.
12. Gap – groups, algorithms, and programming, version 4.10.0. <https://www.gap-system.org>.
13. The on-line encyclopedia of integer sequences, id:a047974. <http://www.research.att.com/~njas/sequences/A047974>.
14. W. Zhang, A. R. Oganov, A. F. Goncharov, Q. Zhu, S. E. Boulfelfel, A. O. Lyakhov, E. Stravrou, M. Somayazulu, V. B. Prakapenka, and Z. Konôpková. Unexpected stable stoichiometries of sodium chlorides. *Science*, 342(6165):1502–1505, 2013.