The Expected Genus of a Random Chord Diagram

Nathan Linial · Tahl Nowik

Abstract To any generic curve in an oriented surface there corresponds an oriented chord diagram, and any oriented chord diagram may be realized by a curve in some oriented surface. The genus of an oriented chord diagram is the minimal genus of an oriented surface in which it may be realized. Let g_n denote the expected genus of a randomly chosen oriented chord diagram of order n. We show that g_n satisfies:

$$g_n = \frac{n}{2} - \Theta(\ln n).$$

I.e., there exist $0 < c_1 < c_2$ and n_0 such that $c_1 \ln n \leq \frac{n}{2} - g_n \leq c_2 \ln n$ for all $n \geq n_0$.

Keywords Gauss code · Curves in surfaces

1 Introduction

The study of plane curves dates back to Gauss [8]. Gauss has assigned to any plane curve with n double points a $2n$ letter word, as follows. To each double point attach a letter, and then register the letters you encounter as you travel along the curve. One obtains a word of length $2n$, where each of the n letters appears precisely twice. Such a word is called a *Gauss code*. Clearly not every Gauss code may be realized by a

N. Linial
School of Computer Science and Engineering, The Hebrew University of Jerusalem,
Jerusalem 91904, Israel
e-mail: nati@cs.huji.ac.il
url: www.cs.huji.ac.il/~nati

T. Nowik (✉)
Department of Mathematics, Bar-Ilan University, Ramat-Gan 52900, Israel
e-mail: tahl@math.biu.ac.il
url: www.math.biu.ac.il/~tahl

plane curve, and Gauss has pointed out a necessary condition for it to be realizable. One can enhance the Gauss code of a curve with a mark on one of the two occurrences of each letter, signifying the order on the pair of strands at the given double point, induced by the orientation of the plane. Various necessary and sufficient conditions for a Gauss code to be realizable in the plane have been eventually given, both in the marked and unmarked settings. See [3, 4, 6, 12, 13, 16–18] and references therein.

To avoid the arbitrariness of the assignment of letters to the different double points, one can replace Gauss codes with chord diagrams. To a marked Gauss code corresponds an oriented chord diagram, which is by definition a splitting of the set $\{1, \ldots, 2n\}$ into n ordered pairs. This may be represented as a circle with $2n$ designated points, and n oriented chords connecting pairs of these points.

Though not every oriented chord diagram may be realized by a curve in the plane, it may be realized by a curve in some oriented surface. Carter in [5] has given a direct construction for the minimal genus surface in which a given diagram may be realized, as follows. Take an annulus, which is thought of as a regular neighborhood of the curve, and identify n pairs of regions along the annulus according to the prescription of the diagram. One obtains an orientable surface F with some d boundary components. Capping off the boundary components with discs produces the required minimal genus surface. The genus of this surface is $g = \frac{1}{2}(n + 2 - d)$ and so $0 \leq g \leq \frac{1}{2}(n+1)$. We will refer to this minimal genus as the genus of the given diagram. A diagram being realizable in the plane is equivalent to it having genus 0.

For fixed n, we are interested in the distribution of genera of the oriented chord diagrams with n chords, and we ask: What is the expected genus g_n of a randomly chosen diagram? (By "random choice" we will always mean that we are choosing with *uniform* probability.) We show that the expected genus is very close to the maximal possible genus $\frac{1}{2}(n + 1)$, in fact, we show that
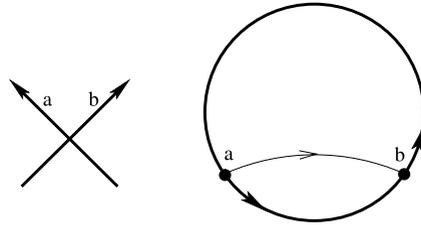
$$g_n = \frac{n}{2} - \Theta(\ln n).$$

Since $g = \frac{1}{2}(n + 2 - d)$, this is equivalent to showing that the expected number of boundary components is $\Theta(\ln n)$. (We say $f(n) = \Theta(h(n))$ if there are $0 < c_1 < c_2$ and n_0 such that $c_1 h(n) \leq f(n) \leq c_2 h(n)$ for all $n \geq n_0$.)

Though the above description is geometric, counting the number of boundary components may be described in a purely combinatorial manner. Traveling along a boundary component of our surface corresponds to a walk on the chord diagram according to the following rule. When moving along the circle of the diagram, and arriving at an end of a chord, continue your motion along the chord to its other side. If your motion along the chord is in the direction (respectively, against the direction) of its orientation, then continue your motion along the circle in the same (respectively, opposite) direction as you have moved before entering the chord. So, to identify a boundary component, one travels along the diagram according to the above rule, until returning to the starting point. Repeating this process, one obtains all boundary components.

This combinatorial walk along a chord diagram is reminiscent of the walk along the cycles of a permutation. The distribution of the number and lengths of cycles of a random permutation on n letters is well understood (see [1]), and the expected number of cycles is also $\Theta(\ln n)$. As will be seen, the setting of the present work

Fig. 1 Chord (a, b)

is substantially more complicated than that of permutations, yet, in the concluding section we will demonstrate that it is for similar reasons that the expected number of cycles in a permutation and the expected number of cycles of the walk along our chord diagrams are both $\Theta(\ln n)$. In the concluding section we will also present a matrix integral for the distribution of the number of boundary components.

Since our problem may be formulated both in a topological and a purely combinatorial manner, it may be of interest to both topologists and combinatorialists, and indeed the text is aimed at both audiences. This work may be viewed as part of the recent expanding interest in probabilistic questions in topology, as appears in [2, 7, 11, 14, 15].

2 Definitions and Statement of Result

Let F be an oriented surface. A *generic curve* in F is an immersion $c : S^1 \to F$ for which the only self intersections are transverse double points. We fix n once and for all, and call a point of $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ a *dot* if it is one of the $2n$ points $\{e^{\pi i k/n} : 1 \le k \le 2n\}$. If a generic curve c has n double points, then there are $2n$ points in S^1 mapped into them, and we will always assume that these $2n$ points are precisely our $2n$ dots. A generic curve with such n double points will be called an n-curve.

An oriented chord diagram of order n is a splitting of the set of dots into n disjoint ordered pairs. One can represent an oriented chord diagram in the plane, by drawing an oriented chord connecting each ordered pair, where the orientation of the chord represents the order of the pair. An oriented chord diagram will also be called simply a *diagram*. (We emphasize that since we have chosen $2n$ fixed points in S^1 as our dots, we distinguish between different splittings obtained from each other by a rotation.)

We denote the set of all diagrams of order n by $\mathcal{D}_n$, and we have $|\mathcal{D}_n| = \frac{(2n)!}{n!}$ (indeed, any ordering $i_1, i_2, \ldots, i_{2n}$ of $1, \ldots, 2n$ determines a splitting into ordered pairs $(i_1, i_2), (i_3, i_4), \ldots$, but in this way each splitting is counted $n!$ times). Any n-curve determines a diagram $D(c) \in \mathcal{D}_n$ as follows. The double points of c split the $2n$ dots into pairs, and the orientation of the surface F induces an ordering on each pair, in the following way. If $c(a) = c(b)$ for dots a, b, and $c'(b), c'(a)$ is a positive basis with respect to the orientation of the surface, then the ordered pair (a, b) is taken. See Fig. 1.

Any diagram may be realized by a curve on some oriented surface, and a regular neighborhood of the curve in all such surfaces is the same, as we now explain (compare [5]). Given a diagram $D \in \mathcal{D}_n$, take the annulus $A = S^1 \times [-\epsilon, \epsilon]$ with a

fixed orientation, and identify S^1 with $S^1 \times \{0\}$. For each dot a let $S_a \subseteq A$ be the $2\epsilon \times 2\epsilon$ square centered at $(a, 0)$. We now identify pairs of dots a, b according to the prescription of D, and we identify the corresponding squares S_a, S_b with a positive or negative $\frac{\pi}{2}$ rotation, so that the self intersection of the curve at the identified square will be as prescribed by the orientation of the chord between a and b in D. We obtain an oriented surface with $d(D) \geq 1$ boundary components, which we denote $F(D)$. Clearly, the embedding $S^1 \to A$ given by $z \mapsto (z, 0)$, composed with the quotient map $A \to F(D)$, is an n-curve $S^1 \to F(D)$ which realizes D, and a regular neighborhood of any curve in any surface realizing D, is identical to $F(D)$. It follows that the genus of the closed surface obtained by capping off the $d(D)$ boundary components of $F(D)$ with discs, is the minimal genus of a surface in which D may be realized.

Definition 2.1 Given a diagram D, we define $g(D)$, the genus of D, to be the minimal genus of a closed oriented surface admitting a curve $c : S^1 \to F$ with $D(c) = D$.

The image in $F(D)$ of our n-curve is a graph with n vertices and $2n$ edges, and so $2 - 2g(D) = n - 2n + d(D)$, or, $g(D) = \frac{1}{2}(n + 2 - d(D))$. Since $d(D) \geq 1$, we deduce that $0 \leq g(D) \leq \frac{1}{2}(n + 1)$.

In this work we study the following question: What is the expected genus of a randomly chosen diagram $D \in \mathcal{D}_n$? We will show that the expected genus is close to the maximal possible genus $\frac{1}{2}(n + 1)$. More precisely we will show:

Theorem 2.2 *The expected genus g_n of a random oriented chord diagram $D \in \mathcal{D}_n$ satisfies*:

$$g_n = \frac{n}{2} - \Theta(\ln n).$$

In fact we show that $\frac{n}{2} - \frac{3}{2} \ln n - 200 \leq g_n \leq \frac{n}{2} - \frac{1}{36} \ln n$, *for* $n \geq 50$.

We think of $g(D)$ and $d(D)$ as random variables defined on a randomly chosen $D \in \mathcal{D}_n$. That is, our sample space is $\mathcal{D}_n$, each diagram having equal probability $\frac{n!}{(2n)!}$. We denote the expected values by $g_n = E[g]$ and $d_n = E[d]$. We will show that $d_n = \Theta(\ln n)$, from which Theorem 2.2 follows via $g_n = \frac{1}{2}(n + 2 - d_n)$.

3 The Random Procedure

We label the dot $e^{\pi i k/n}$, and its corresponding square, simply by k, and so $k \pm 1$ will mean addition mod $2n$. The interval along the boundary of the annulus A between two adjacent squares will be called an *edge*, so we have $4n$ edges. We orient the edges according to the orientation induced on ∂A from that of A, and we denote the oriented edge from square a to square b by $[a, b]$. The edge $[b, a]$ will then be the parallel edge in the other boundary component. So, all edges in $S^1 \times \{\epsilon\}$ are of the form $[a, a + 1]$, and they will be called *positive* edges, and all edges in $S^1 \times \{-\epsilon\}$ are of the form $[a + 1, a]$ and will be called *negative* edges. See Fig. 2.

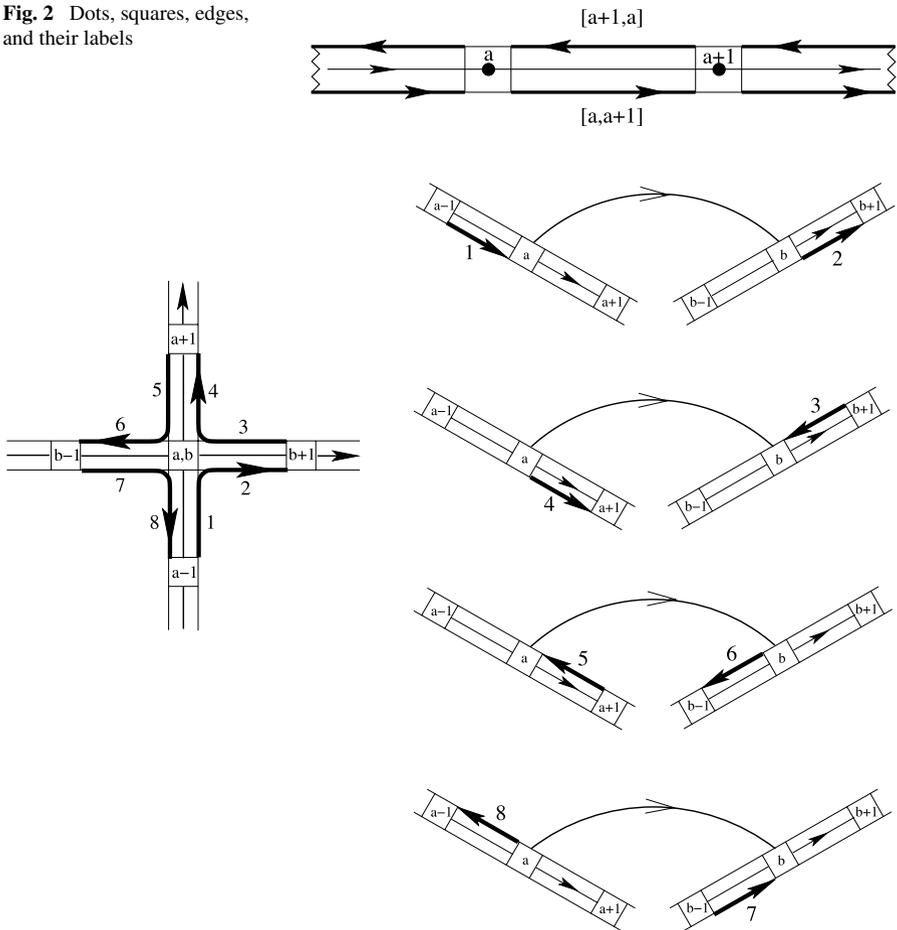**Fig. 2** Dots, squares, edges, and their labels

Fig. 3 Edge attachments due to chord (a, b)

We would like to see how our edges are attached to each other due to the gluing of two squares. We will say an attachment $[a, b] - [c, d]$ takes place if the end point of $[a, b]$ is glued to the beginning point of $[c, d]$. So, say we have identified the two squares a and b according to the oriented chord (a, b). As seen in Fig. 3, the eight edges involved are attached to each other as follows: $[a - 1, a] - [b, b + 1]$, $[b + 1, b] - [a, a + 1]$, $[a + 1, a] - [b, b - 1]$, $[b - 1, b] - [a, a - 1]$. This can be summarized by the following rule.

Rule 3.1 An attachment $[a, b] - [c, d]$ holds if the oriented chord (b, c) exists, and the signs of $[a, b]$ and $[c, d]$ are the same, or if the oriented chord (c, b) exists and the signs of $[a, b]$ and $[c, d]$ are opposite.

Traveling along a boundary component of the surface $F(D)$ corresponds to a walk along the diagram D, which by Rule 3.1 proceeds as follows: When moving along

the circle of the diagram, and arriving at an end of a chord, continue your motion along the chord to its other side. If your motion along the chord is in the direction (respectively, against the direction) of its orientation, then continue your motion along the circle in the same (respectively, opposite) direction as you have moved before entering the chord.

We may thus read all boundary components directly from the diagram D, as follows: Choose an arbitrary edge, and start traveling along the diagram in the above way, alternatingly passing edges and chords, until you return to your initial edge. Then choose some unvisited edge, and similarly travel until you return to it. Continue until all edges have been visited. Notice that when this is done, each chord of the diagram has been visited four times, each visit corresponding to one of the four corners of the glued square.

For $k \leq n$, a *k-n* diagram is a choice of $2k$ out of the $2n$ dots, and a splitting of these $2k$ dots into k oriented chords, i.e. ordered pairs. The remaining $2n - 2k$ dots will be called *vacant* dots. So, an oriented chord diagram of order n is an *n-n* diagram.

Definition 3.2 A *path* in a *k-n* diagram D, is a sequence $[a_1, b_1] - [a_2, b_2] - \cdots - [a_r, b_r]$ of distinct edges, attached via oriented chords of D according to Rule 3.1. A path in D is called a *loop* if the attachment $[a_r, b_r] - [a_1, b_1]$ also holds. A path in D is called a *segment* if a_1 and b_r are vacant dots (perhaps the same dot). Loops and segments are precisely those paths that cannot be further extended.

We now specify our procedure for choosing a random *n-n* diagram. Our procedure will choose the chords one by one. We first fix an ordering $e_1, \ldots, e_{4n}$ of our edges, once and for all. Before the procedure begins, we announce e_1 as the "pointer" edge. Assume that after the $(j-1)$th step, we have already chosen $j-1$ oriented chords, and the pointer edge lies in a segment (rather than a loop) of the given $(j-1)$-n diagram. The next chord is now chosen with one of its dots being the concluding dot of the segment in which the pointer lies, and its other dot is randomly chosen from the other $2n - 2j + 1$ vacant dots. The orientation of the new chord is also randomly chosen. If in the new *j-n* diagram, the pointer's segment continues to be a segment, i.e. it does not close into a loop, then the same edge remains the pointer. If on the other hand, after the jth chord is added, the pointer's segment closes into a loop, then the edge with smallest index which lies in a segment in the new *j-n* diagram, becomes the new pointer.

This procedure indeed produces all *n-n* diagrams with equal probability. To see this note that at the jth step we randomly choose one of the $2n - 2j + 1$ vacant dots, and then randomly choose an orientation for the new chord. So the probability for each choice is $\frac{1}{2 \cdot (2n - 2j + 1)}$, and so the probability for any particular history of the random procedure is $\frac{1}{2^n \cdot (2n-1)(2n-3) \cdots 3 \cdot 1} = \frac{n!}{(2n)!}$. Then note that any diagram may be obtained via a unique history of the random procedure.

Examples of two runs of our random procedure appear in Figs. 4 and 5, demonstrating some of the interesting features of the possible evolution of the pointer's segment. In both figures the edge e_1 is the edge $[a, a+1]$ and is marked by 1.

In Fig. 4 the chords are chosen in the following order: $(a+1, b)$, $(c, b+1)$, $(c-1, c+1)$, $(a, a+2)$. The segment of e_1 after the four steps of this run is $[a+3, a+$
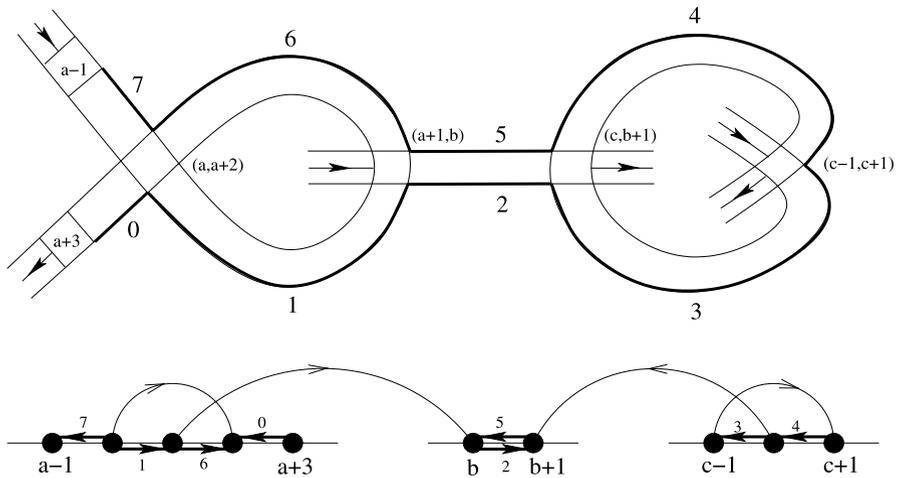
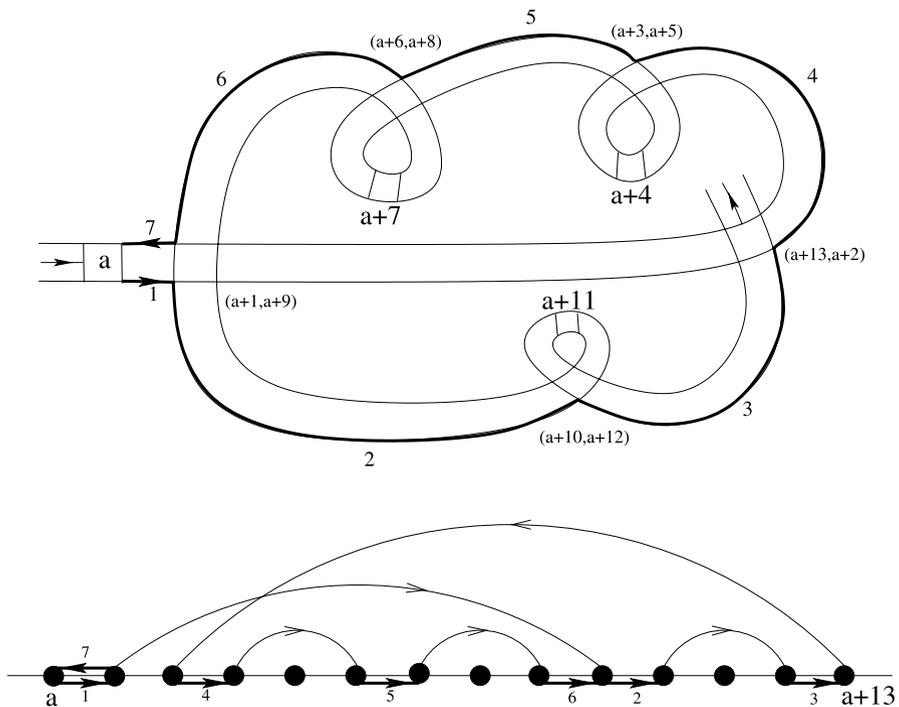**Fig. 4** A four step run of the random procedure

Fig. 5 A five step run of the random procedure

$2] - [a, a + 1] - [b, b + 1] - [c, c - 1] - [c + 1, c] - [b + 1, b] - [a + 1, a + 2] -$
$[a, a - 1]$ and these edges are marked in the figure by $0, \ldots, 7$, in this order. The
evolution of e_1's segment throughout the four steps of this run is $1 - 2$, $1 - 2 - 3$,

$1 - 2 - 3 - 4 - 5 - 6, 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7$. Note that after the third step the segment revisits some chords that were chosen in previous steps. Note that after the fourth step the segment extends in both directions, and so 1 is no longer the first edge in the segment. We point out the following difference between this run, and the run of Fig. 5. In Fig. 4 the initial and final dots of the pointer's segment are distinct, whereas in Fig. 5 it is the same dot.

4 Upper Bound for d_n

In this section we establish our upper bound for d_n. If ℓ is a loop in an n-n diagram D, then the *size* of ℓ is the number of distinct chords visited by ℓ. If k is the size of the loop ℓ and r is the number of edges it visits, then since ℓ alternatingly visits an edge and a chord, and since each chord is visited at most four times, we have $k \leq r \leq 4k$.

For given n, let $L_k = L_k(n)$ denote the expected number of loops of size k in a random n-n diagram, then $d_n = \sum_{k=1}^{n} L_k$. We will show in Proposition 4.5 below, that for $k \leq \frac{n}{100}$, $L_k \leq \frac{3}{k}$. On the other hand, since any chord is visited by at most four different loops, the total number of all loops of size $k > \frac{n}{100}$ is at most 400, and so its expected value $\sum_{k > \frac{n}{100}} L_k$ is at most 400. Together this gives

$$d_n = \sum_{k \leq \frac{n}{100}} L_k + \sum_{k > \frac{n}{100}} L_k \leq \sum_{k \leq \frac{n}{100}} \frac{3}{k} + 400 \leq 3 \ln n + 400.$$

In order to obtain our bound $L_k \leq \frac{3}{k}$ we will need to bound the probability that at a given kth step, the pointer's segment closes into a loop. We now prove the following.

Proposition 4.1 *For $k \leq \frac{n}{100}$, the probability that the pointer's segment will close into a loop at the kth step is at most $\frac{3}{4n}$.*

Let S be the pointer's segment in our $(k-1)$-n diagram after step $k-1$, and let p be its concluding dot. We now need to choose the kth chord, with one end being p. We must determine how many choices will result in closing S into a loop. Let q be the initial dot of S. If $q \neq p$ (that is, the initial and final dots are distinct, as occurs in the example in Fig. 4), then for S to be closed into a loop, we must choose q as the second dot for the kth chord, and so the choice of unoriented chord is unique. Though usually only one of the two choices of orientation for this chord will indeed close the segment S into a loop (as is the case in Fig. 4), it may in fact occur that both orientations accomplish this. Since we are seeking an upper bound for the probability, we will always count both orientations as possible, or in other words we will ignore the choice of orientation in the computation of the probability. Since there are $2n - 2k + 1$ vacant dots from which we may choose the second dot for the new chord, the probability of choosing the correct dot q is $\frac{1}{2n-2k+1}$, and since we assume $k \leq \frac{n}{100}$ we have $\frac{1}{2n-2k+1} \leq \frac{1.1}{2n}$.

If $q = p$ (that is, the initial and final dot is the same, as occurs in the example of Fig. 5), then at first sight it may seem that closing the pointer's segment into a loop
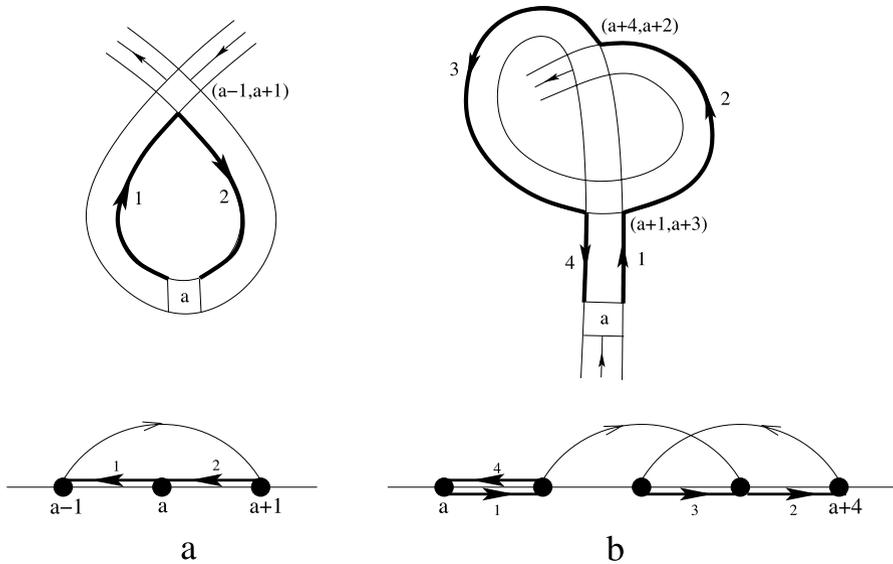
Fig. 6 Plugs

is impossible, and in most cases this is in fact true. But on the other hand, there are cases with $q = p$ where not only does there exist a choice of chord that closes the given segment into a loop, but there are in fact many such choices. The example in Fig. 5 is such a case. The oriented chords $(a, a + 4)$, $(a, a + 7)$ or $(a, a + 11)$ may each be added in the present step to close the pointer's segment into a loop. For the analysis of this phenomenon, we define the following notion.
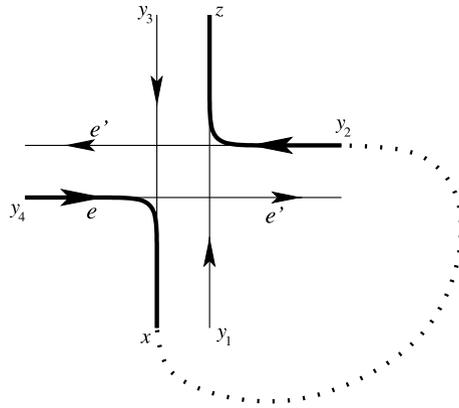
Definition 4.2 A *plug* is a segment $[a_1, b_1] − \cdots − [a_r, b_r]$ for which $a_1 = b_r$. The vacant dot $a_1 = b_r$ is called the *entrance* to the plug.

Examples of two plugs are depicted in Fig. 6. In Fig. 6a, the chord $(a − 1, a + 1)$ produces the plug $[a, a − 1] − [a + 1, a]$ with entrance a. In Fig. 6b, the chords $(a + 1, a + 3)$ and $(a + 4, a + 2)$ produce the plug $[a, a + 1] − [a + 3, a + 4] − [a + 2, a + 3] − [a + 1, a]$ with entrance a. Note that the same vacant dot can be the entrance to two different plugs.

In our present case, where $q = p$, the pointer's segment is itself a plug, but this fact is not of interest to us. What enables us to close the pointer's segment into a loop in Fig. 5 is each one of the *additional* plugs that are present in the given 5-n diagram, namely the three plugs $[a + 4, a + 3] − [a + 5, a + 4]$, $[a + 7, a + 6] − [a + 8, a + 7]$, and $[a + 11, a + 10] − [a + 12, a + 11]$ (each of which is similar to the plug in Fig. 6a). In fact, the next lemma shows that in order to close the pointer's segment into a loop in the case $q = p$, it is necessary that the second dot of the new chord will be an entrance to a plug.

Lemma 4.3 *Let D be a $(j − 1)$-n diagram, and let a be a vacant dot in D. Let e be an edge entering a, (i.e. e is $[a − 1, a]$ or $[a + 1, a]$). Assume Q is an additional*

Fig. 7 Proof of Lemma 4.3

oriented chord with one end at a and the other end at some other vacant dot b, such that in the j-n diagram obtained by adding Q, the path beginning at e leads to an edge e′ which is one of the two exiting edges at a (i.e. e′ is [a, a + 1] or [a, a − 1]). Then b is the entrance to a plug in D.

Proof Assume b is not an entrance to a plug. In Fig. 7, edge e is marked, and the two possibilities for $e′$ are marked. The path beginning at e passes the point x, and in order for it to lead to $e′$, it must eventually arrive back into the region depicted in the figure. It does not arrive at y_1 or y_3 since we have assumed that b is not an entrance to a plug. If it arrives at y_4 then it closes a loop without passing either possibility for $e′$. So, it must arrive at y_2 as depicted, and it then exits our region through point z. This time its only possibility for returning is through y_4, which as before prevents it from ever arriving at either possibility for the exiting edge $e′$.   □

As we have seen, there may be many plugs available for completing our segment into a loop, but fortunately, the *expected* number of available plugs is small. The main technical effort of this work is the following proposition whose proof we defer to Sect. 6.

Proposition 4.4 *For $k \leq \frac{n}{100}$, the expected number of plugs present after k steps of the random procedure is at most $\frac{1}{4}$.*

Back to the proof of Proposition 4.1 for the case $q = p$. By Lemma 4.3, in order for S to close into a loop, the second dot we choose for the new chord must be the entrance to some plug. By Proposition 4.4, the expected number of plugs existing in the present stage of the random procedure (i.e. after $k − 1$ steps), is at most $\frac{1}{4}$. Together with the case $q \neq p$ we have on average at most $1 + \frac{1}{4}$ choices for the new unoriented chord. (As before, we ignore the additional choice of orientation.) Note that we must take the sum and not the maximum of the bounds for the two possibilities $q \neq p$ and $q = p$, since the expectation for the number of plugs that we bound in Proposition 4.4 is not conditional on $q = p$.

We obtain that for $k \leq \frac{n}{100}$, the probability that the pointer's segment closes into a loop at the kth step of the random procedure is at most $(1 + \frac{1}{4}) \cdot \frac{1.1}{2n} \leq \frac{3}{4n}$, which completes the proof of Proposition 4.1.

Proposition 4.5 *Let L_k denote the expected number of loops of size k in a random n-n diagram. Then for $k \leq \frac{n}{100}$ we have $L_k \leq \frac{3}{k}$.*

Proof It is clear from the definition of our random procedure, that if the segment of e_1 closes into a loop at the kth step, then this loop is of size k. So the event that the edge e_1 lies in a loop of size k is the same as the event that e_1 survives as pointer until step k, and then at step k its segment closes into a loop. The probability for this event is at most the probability that at the kth step the pointer's segment closes into a loop, and by Proposition 4.1 this probability is at most $\frac{3}{4n}$. Now, our random procedure produces each n-n diagram with equal probability, and so by the symmetry of our annulus, the probability for *any* given edge to lie in a loop of size k is also at most $\frac{3}{4n}$, or alternatively, the probability P_k that a *randomly* chosen edge will lie in a loop of size k is at most $\frac{3}{4n}$.

We obtain a lower bound for P_k by noting that each loop of size k includes at least k edges, and the total number of edges is $4n$, and so $P_k \geq \frac{kL_k}{4n}$. Together we get $\frac{kL_k}{4n} \leq P_k \leq \frac{3}{4n}$ which proves our claim. $\qquad\square$

As already explained above (before Proposition 4.1), the bound $L_k \leq \frac{3}{k}$ for $k \leq \frac{n}{100}$ implies the following upper bound for d_n:

$$d_n \leq 3 \ln n + 400.$$

5 Lower Bound for d_n

We have asked in the proof of Proposition 4.5, what is the probability P_k that a randomly chosen edge will lie in a loop of size k. We have noticed that this is precisely the probability that in our random procedure, the segment of e_1 survives until the kth step, and then at the kth step it closes into a loop. In this section we will find a lower bound for P_k, for $n \geq 50$ and $k \leq \sqrt{n}$, which in turn will provide a lower bound for d_n.

A run of $j-1$ steps of the random procedure is called *good* if after these $j-1$ steps e_1 is still the pointer, and its segment is of the form $[a_1, b_1] - \cdots - [a_j, b_j]$ with all the dots $a_1, b_1, \ldots, a_j, b_j$ being distinct. When $j-1 = 0$, i.e. before starting the random procedure, then the pointer's segment is simply $e_1 = [a_1, b_1]$, so the run is good. If the run is good after $j-1$ steps, and at the jth step the second dot chosen for the new chord is not adjacent to any of the dots $a_1, b_1, \ldots, a_j, b_j$, then the run is still good after the jth step. This restriction for the choice of the jth chord means that if an edge of the segment is say $[a, a+1]$, then the four dots $a-1, a, a+1, a+2$ are not chosen. So, at the jth step we have at most $4j$ dots which we are forbidden to choose, so the number of allowed choices for a new dot at the jth step is at least

$2n - 4j$. Since the total number of dots from which we choose is $2n - 2j + 1$, the probability for such restricted choice at the jth step is at least $\frac{2n-4j}{2n-2j+1}$.

If after $k - 1$ steps of the random procedure the run is still good, then in particular, the initial and final dots of e_1's segment are distinct. So, at the kth step there exists a choice of oriented chord that closes e_1's segment into a loop, and the probability for this choice is $\frac{1}{2(2n-2k+1)}$. So together, for $n \geq 50$ and $k \leq \sqrt{n}$, the probability that the segment of e_1 survives until the kth step, and then at the kth step closes into a loop satisfies:

$$P_k \geq \frac{1}{2(2n - 2k + 1)} \prod_{j=1}^{k-1} \frac{2n - 4j}{2n - 2j + 1}$$

$$\geq \frac{1}{4n} \prod_{j=1}^{k-1} \left(1 - \frac{2j + 1}{2n - 2j + 1}\right) \geq \frac{1}{4n} \prod_{j=1}^{k-1}\left(1 - \frac{j + 1}{n - k}\right)$$

$$\geq \frac{1}{4n} \prod_{j=1}^{k-1} e^{-\frac{6}{5} \cdot \frac{j+1}{n-k}} = \frac{1}{4n} e^{-\frac{6}{5} \sum_{j=1}^{k-1} \frac{j+1}{n-k}} \geq \frac{1}{4n} e^{-\frac{3}{5} \cdot \frac{k^2+k}{n-k}} \geq \frac{1}{4n} e^{-\frac{3}{5} \cdot \frac{n+\sqrt{n}}{n-\sqrt{n}}} \geq \frac{1}{9n}.$$

We use the assumption $n \geq 50$ in the fourth inequality, observing that $1 - x \geq e^{-\frac{6}{5}x}$ for $0 \leq x \leq \frac{\sqrt{50}}{50-\sqrt{50}}$, and in the last inequality, observing that $e^{-\frac{3}{5}x} \geq \frac{4}{9}$ for $0 \leq x \leq \frac{50+\sqrt{50}}{50-\sqrt{50}}$.

As before, let L_k be the expected number of loops of size k in a random n-n diagram, then since the number of edges in a loop of size k is at most $4k$ we have $P_k \leq \frac{4kL_k}{4n}$. Together, for $n \geq 50$ and $k \leq \sqrt{n}$ we get $\frac{1}{9n} \leq P_k \leq \frac{4kL_k}{4n}$, so $L_k \geq \frac{1}{9k}$. We may now establish our lower bound for d_n, for $n \geq 50$:

$$d_n = \sum_{k=1}^{n} L_k \geq \sum_{k=1}^{\sqrt{n}} \frac{1}{9k} \geq \frac{1}{9} \ln \sqrt{n} = \frac{1}{18} \ln n.$$

Together with the upper bound of Sect. 4 we obtain $d_n = \Theta(\ln n)$, which proves Theorem 2.2, stating that the expected genus g_n of a random diagram of order n satisfies:

$$g_n = \frac{n}{2} - \Theta(\ln n).$$

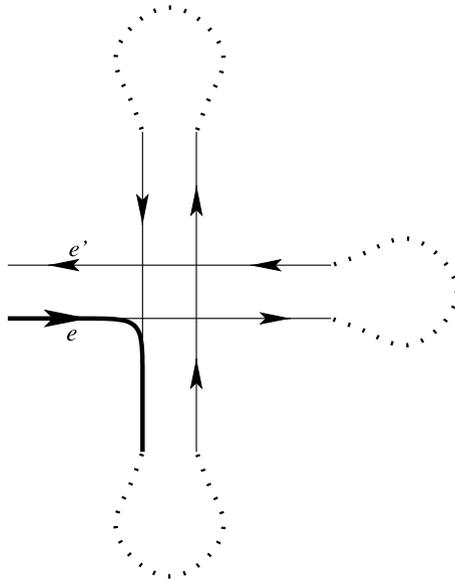## 6 Upper Bound for the Expected Number of Plugs

In this section we prove Proposition 4.4, stating that for $k \leq \frac{n}{100}$, the expected number of plugs present in our k-n diagram after k steps of the random procedure is at most $\frac{1}{4}$.

Definition 6.1 Two vacant dots in a k-n diagram D are called *neighbors*, if they are the two end points of a segment in D.

Fig. 8 Proof of Lemma 6.3

Definition 6.2 A *positive* plug is a plug $[a_1, b_1] - \cdots - [a_r, b_r]$ for which the two edges $[a_1, b_1], [a_r, b_r]$ are of the same sign, that is, they are of the form $[a, a+1]$, $[a-1, a]$ or $[a, a-1], [a+1, a]$ (as in Fig. 6a). A *negative* plug is a plug for which these two edges are of opposite sign, that is, they are of the form $[a, a+1], [a+1, a]$ or $[a, a-1], [a-1, a]$ (as in Fig. 6b). Note that if the same vacant dot is the entrance to two different plugs, then these two plugs must be of the same sign.

Lemma 6.3 *Under the assumptions of Lemma* 4.3, *if e and e′ are of opposite sign, and if b is not the entrance to a positive plug* (*and so by Lemma* 4.3 *it is the entrance to one or two negative plugs*), *then either a and b are neighbors* (*Definition* 6.1), *or a is also an entrance to a plug.*

Proof Assume a and b are not neighbors. In order for us to arrive at e', given that b is not the entrance to a positive plug and a and b are not neighbors, our path must be as in Fig. 8, which shows that a is the entrance to a (negative) plug.   □

Any chord is involved in at most four different segments, and so at each step, when adding a new chord, at most four new plugs can be created. But we will show that in fact the expected number of plugs created at each step $k \le \frac{n}{100}$ is at most $\frac{25}{n}$. This implies that the expected number of plugs present after $k \le \frac{n}{100}$ steps is at most $\frac{1}{4}$. To establish this bound we will in fact need to prove the following more detailed proposition, which distinguishes between positive and negative plugs.

Proposition 6.4 *The following holds for $k \le \frac{n}{100}$:*

(1) *Let G_k^+ (respectively G_k^-) denote the expected number of positive (respectively negative) plugs completed at the kth step. Then $G_k^+ \le \frac{5}{n}$ and $G_k^- \le \frac{20}{n}$.*

(2) *The expected number of plugs present after k steps is at most $\frac{1}{4}$.*

(3) *Let H_k^+ (respectively H_k^-) denote the probability that after the kth step the concluding dot of the pointer's segment is an entrance to a positive (respectively negative) plug. Then $H_k^+ \leq \frac{6}{n}$ and $H_k^- \leq \frac{21}{n}$.*

Proof (1) Say at the kth step we have chosen a chord Q between dots a and b, and a plug has been completed, with dot c being its entrance. This means that after adding Q there is a segment with edges $e_{i_1} - e_{i_2} - \cdots - e_{i_r}$ beginning and ending at the vacant dot c, and before adding Q this segment did not exist. This means that before adding Q, the segment S_1 beginning with e_{i_1} ended at some vacant dot $a \neq c$, and the segment S_2 ending with e_{i_r} began at some vacant dot $a' \neq c$. We now distinguish three cases as follows. If $a \neq a'$ then the new chord Q must be between a and a'. By definition of our random procedure, the concluding vacant dot p of the pointer's segment is one of the dots of the new chord Q, so must be either a or a'. We will refer to this case as Case A. If on the other hand $a = a'$ then the new chord Q must be between a and some other vacant dot b. In this case either $p = a$ or $p = b$, and these two possibilities will be referred to as Case B and Case C, respectively.

For Case A, we note that there are at most four different segments with one end being p. The other end of each such segment is a vacant dot that may be c of the above description. For each such c there is a unique second segment with which a configuration S_1, S_2 as described above may arise for a positive plug, and a unique such second segment for a negative plug. Our assumption is that $a \neq a'$ and so for each such configuration there is a unique choice of unoriented chord with which such a plug may be created. As discussed in Sect. 4, it may be that both choices of orientation for this chord bring to the completion of the plug. So here and in all following cases, we do as we have done in Sect. 4, and include both choices in our count by simply ignoring the choice of orientation. As before, the probability for the correct unoriented chord to be chosen in each case is $\frac{1}{2n-2k+1}$ since there are $2n - 2k + 1$ additional vacant dots, and for $k \leq \frac{n}{100}$ we have $\frac{1}{2n-2k+1} \leq \frac{1.1}{2n}$. So, the contribution of this case to G_k^+ and G_k^- is at most $4 \cdot \frac{1.1}{2n}$. Note that it may be that different configurations in our count are completed into a plug by the same choice of chord, but by the additivity of expectation, the contributions of all configurations may be added regardless of the dependence between them.

In Case B, Q is between the dot $p = a$ and the dot b, and by Lemma 4.3, b must be an entrance to an existing plug. We bound all possible contributions that may be from choosing the second dot of the new chord as the entrance to an existing plug. Any new chord may participate in at most four different segments, and so at most four new plugs may be completed. By induction, we may use (2) of the present proposition for $k - 1$ to conclude that on average we have at most $\frac{1}{4}$ existing plugs available. So, the contribution is on average at most $4 \cdot \frac{1}{4} \cdot \frac{1.1}{2n} = \frac{1.1}{2n}$. We cannot determine how this contribution will divide between G_k^+ and G_k^- and so we add it to both.

In Case C, $p = b$, and our choice of the second dot a for Q is such that a is part of a configuration of segments S_1, S_2 and dots c, a as described above. The segments S_1 and S_2 may or may not pass chords, but there is just one special configuration for S_1, S_2 in which both S_1 and S_2 do not pass any chord, namely, the configuration where a and c are adjacent dots along the annulus, and S_1, S_2 are the two edges

connecting them. If the configuration is not this special one, then necessarily the dot c is adjacent along the annulus to a dot that is the end of one of the $k - 1$ existing chords. So there are at most $4(k - 1)$ possibilities for such dot. For each such dot c there are two possibilities for a pair of segments S_1, S_2 that may give rise to a positive plug, and two possibilities for a negative plug. Together this gives at most $8k$ possible pairs of segments for positive plugs and for negative plugs. Now we note that in order for us to be in Case C, our dot p must be at the entrance to an existing plug after step $k - 1$. By induction we may use (3) of the present proposition for $k - 1$ to conclude that the probability for us being in Case C is at most $\frac{6}{n} + \frac{21}{n}$. And so the contribution of the non-special configurations to G_k^+ and G_k^- is at most $8k(\frac{6}{n} + \frac{21}{n})\frac{1.1}{2n} \leq 8 \cdot \frac{n}{100} \cdot \frac{27}{n} \cdot \frac{1.1}{2n} \leq 3 \cdot \frac{1.1}{2n}$.

For the special configuration, if a plug is completed then it is necessarily a negative plug, so contributes only to G_k^-. If $p = b$ is the entrance to a positive plug, which happens, by induction on part (3), with probability at most $\frac{6}{n}$, then we take our bound to be simply the total number of choices $2n - 2k + 1$ for a. There may be a special configuration on each side of a, and so the contribution to G_k^- is at most $2 \cdot (2n - 2k + 1) \cdot \frac{6}{n} \cdot \frac{1.1}{2n} \leq 24 \cdot \frac{1.1}{2n}$.

If $p = b$ is the entrance to a negative plug, which happens by part (3), by induction, with probability at most $\frac{21}{n}$, then by Lemma 6.3, we must choose a which is either a neighbor of b or the entrance to a plug. The dot b has at most 4 neighbors. For each such neighbor a there is at most one special configuration that may be completed into a plug, since it may not be on the side of a where the segment from b arrives at a. So, the contribution of this case is at most $4 \cdot \frac{21}{n} \cdot \frac{1.1}{2n}$. The second possibility is that a itself is an entrance to a plug, but in Case B above we have already counted all possible contributions from connecting p to a dot which is the entrance to an existing plug, and so we need not count this again here. The contribution to G_k^- is thus at most $4 \cdot \frac{21}{n} \cdot \frac{1.1}{2n} \leq \frac{1.1}{2n}$, since $n \geq 100$ whenever the assumption $k \leq \frac{n}{100}$ is relevant.

We add all contributions for G_k^+:

$$G_k^+ \leq (4 + 1 + 3) \cdot \frac{1.1}{2n} \leq \frac{5}{n},$$

and for G_k^-:

$$G_k^- \leq (4 + 1 + 3 + 24 + 1) \cdot \frac{1.1}{2n} \leq \frac{20}{n}.$$

(2) In each step $j \leq k$ on average at most $\frac{5}{n} + \frac{20}{n}$ plugs are completed, by (1), and so after k steps the expected number of plugs is at most $k(\frac{5}{n} + \frac{20}{n}) \leq \frac{n}{100}(\frac{5}{n} + \frac{20}{n}) = \frac{1}{4}$.

(3) If after the kth step, the final dot of our segment is the entrance to a positive plug, then this plug may either be one that has existed previously, or one that has just been completed. If it is a plug that has existed previously, then in the $(k - 1)$-n diagram we had before the kth step, there is a unique segment S leading to its entrance (which is not the plug itself), and let a denote the vacant dot at the beginning of S. In order for us to end up at the entrance to the given plug after adding the kth chord, this chord must include a. As before, let p denote the concluding dot of the pointer's segment. If $p \neq a$ then we have one choice for such unoriented chord. If $p = a$ then

in order for us to continue into the segment S, then by Lemma 4.3 the other dot b of the new chord must be the entrance to an existing plug. Together we see that in order for us to land at the entrance of an existing plug, we must choose the second dot for the new chord either as a dot a as described above, which is uniquely determined by a plug, or as a dot which is itself the entrance to a plug. By (2) we know that there are on average at most $\frac{1}{4}$ previously existing plugs, and so this contributes at most $2 \cdot \frac{1}{4} \cdot \frac{1.1}{2n}$ to the probability.

On the other hand, the probability that after the kth step we have landed at the entrance of a positive plug that has just been completed, is at most the probability that such a plug has at all been completed at the kth step. By (1) this probability is at most $\frac{5}{n}$, since the expected number of plugs completed is a bound to the probability that at least one plug has been completed. Together we get $H_k^+ \leq 2 \cdot \frac{1}{4} \cdot \frac{1.1}{2n} + \frac{5}{n} \leq \frac{6}{n}$. In the same way, using $G_k^- \leq \frac{20}{n}$ we get $H_k^- \leq \frac{21}{n}$ □

Recall that what we have actually used from Proposition 6.4 is only part (2), which bounds the total number of plugs. The need for this more detailed analysis is due to the large contribution of existing positive plugs to the completion of new negative plugs in Case C with the special configuration. This required that we separate between positive and negative plugs in the inductive proof, with a larger bound for the negative plugs.

7 Similar Problems and Matrix Integrals

We recall our combinatorial rule for traveling along an oriented chord diagram, to produce its boundary components: *When moving along the circle of the diagram, and arriving at an end of a chord, continue your motion along the chord to its other side. If your motion along the chord is in the direction (respectively, against the direction) of its orientation, then continue your motion along the circle in the same (respectively, opposite) direction as you have moved before entering the chord.* The number of cycles of this travel along a diagram D was denoted $d(D)$, and served to compute the genus of the corresponding surface via $2 - 2g = -n + d$. With all cycles present, each arc between adjacent dots is visited twice (once in each direction) and each chord is visited four times (twice in each direction).

If one looks at an *unoriented* chord diagram, i.e., a chord diagram where there is no choice of direction along each chord, then one can consider the following analogous travel: *Always move along the circle of the diagram in the counterclockwise direction. When arriving at an end of a chord, continue your motion along the chord to its other side, and continue moving along the circle.* In this setting one may also count the number of cycles of the travel along an unoriented chord diagram U, and we denote this number by $d'(U)$. Here each arc between adjacent dots is visited once, and each chord is visited twice (once in each direction).

It turns out that this simplified combinatorial setting also has a geometric interpretation in terms of the construction of a surface, as follows. Start with a disc whose boundary is the circle of the given unoriented chord diagram U with n chords. For any chord, attach a band to the boundary of the disc according to the given chord, where

the gluing respects the orientation of the disc. We obtain an orientable surface with $d' = d'(U)$ boundary components, so if we cap off the boundary components with discs we obtain a closed orientable surface of genus g where $2 - 2g = 1 - n + d'$. Indeed, each boundary component of this surface corresponds to a cycle of the travel described above on the unoriented chord diagram.

Another way of describing this construction is as follows. The initial disc is thought of as a $2n$-gon, and instead of gluing bands, one glues pairs of edges of this $2n$-gon, according to the pairing of the chord diagram. The same closed surface is obtained, but now the number $d'(U)$ is the number of vertices in the cell structure that is obtained, which is dual to the handle structure obtained from the previous description.

There are $(2n - 1)!! = (2n - 1)(2n - 3) \cdots 3 \cdot 1$ possible unoriented chord diagrams with n chords, and we may ask, what is the distribution of genera for a randomly chosen diagram, where each diagram is chosen with equal probability $\frac{1}{(2n-1)!!}$. Harer and Zagier in [9] have given a complete answer to this question in terms of the following generating function. If $p_{k,n}$ denotes the probability that $d'(U) = k$ for a random unoriented chord diagram U with n chords, and if $f(x, N) = 1 + 2\sum_{k,n} p_{k,n} x^{n+1} N^k$ then

$$f(x, N) = \left(\frac{1 + x}{1 - x}\right)^N.$$

From this generating function we may in particular deduce the expected number $d'_n = E[d']$, of cycles. Indeed, d'_n is clearly half the coefficient of x^{n+1} in the series $\frac{\partial f}{\partial N}(x, 1)$. Computation of the coefficients gives $d'_n = \sum_{k=0}^{n} \frac{1}{k^*}$ where $k^* = k + 1$ for k even, and $k^* = k$ for k odd. So, $d'_n = \Theta(\ln n)$, or moreover, $d'_n \sim \ln n$, by which we mean $d'_n / \ln n \to 1$.

The proof in [9] of the above explicit expression for $f(x, N)$ uses the technique of matrix integrals. A nice exposition of this proof, including the necessary background on matrix integrals, may be found in [10]. The proof includes two steps. The first step is to express $g_n(N) = \sum_k p_{k,n} N^k$ as a certain integral over a space of matrices. The second step is to compute this integral, and deduce from it the explicit expression for $f(x, N)$.

For the problem of the present work, we now carry out the first step of this program. That is, if $a_{k,n}$ denotes the number of oriented chord diagrams D of order n satisfying $d(D) = k$, then we will present a matrix integral for $\sum_{k=1}^{n+2} a_{k,n} N^k$. (Since $2 - 2g = -n + d$, indeed $d \leq n + 2$.) The integral itself has precisely the same form as that of Harer–Zagier, only it is carried out on a different space of matrices.

We look at $M_N(\mathbb{C}) \otimes M_N(\mathbb{C})$ which we denote $M_N^{\otimes 2}$. We arrange the indices for an element $A \in M_N^{\otimes 2}$ as follows, ${}^i_l A^j_k$, where the upper indices correspond to the first factor of $M_N^{\otimes 2}$ and the bottom indices correspond to the second factor. So if $C = AB$ then

$$ {}^i_l C^j_k = \sum_{p,q} {}^i_l A^{p\,p}_{q\,q} B^j_k. $$

We will have two operations on $M_N^{\otimes 2}$. The first is the usual conjugation on each of the two factors of $M_N^{\otimes 2}$, that is, we define A^* by ${}^i_l (A^*)^j_k = \overline{{}^j_k A^i_l}$. We introduce a

second operation $A^\#$ given by ${}^i_l(A^\#)^j_k = \overline{{}^j_i A^k_l}$. In words: the element is conjugated, and the indices are shifted circularly around the figure. The operation $A^\#$ is quite peculiar from the point of view of linear algebra, but it reflects our geometric setting of gluing pairs of squares in our annulus with a $\frac{\pi}{2}$ rotation. We define the subspace $H_N \subseteq M_N^{\otimes 2}$, over which we will integrate, as follows: $H_N = \{A \in M_N^{\otimes 2} : A^\# = A\}$. That is, H_N is the space of all $A \in M_N^{\otimes 2}$ such that ${}^i_l A^j_k = \overline{{}^j_i A^k_l}$. It follows that also ${}^i_l A^j_k = {}^k_j A^l_i$ and ${}^i_l A^j_k = \overline{{}^l_k A^i_j}$.

The space H_N is a real space of dimension $M = N + (N^2 - N) + (N^4 - N^2)/2 = (N^4 + N^2)/2$. We choose the following coordinates $s_1, \ldots, s_M$ for H_N, which will be of three types. If ${}^i_l x^j_k$ and ${}^i_l y^j_k$ are the real and imaginary parts of ${}^i_l A^j_k$, then the first type of coordinates are ${}^i_i x^i_i$, $1 \le i \le N$ (N coordinates of type 1). The second type of coordinates are ${}^i_j x^j_i$ and ${}^i_j y^j_i$, $1 \le i < j \le N$ ($N^2 - N$ coordinates of type 2). The third type of coordinates are ${}^i_l x^j_k$ and ${}^i_l y^j_k$, where $i \ne k$ or $j \ne l$, and the 4-tuple (i, j, k, l) is the minimal with respect to lexicographical ordering among its four cyclic permutations, ($(N^4 - N^2)/2$ coordinates of type 3). We denote the volume form $\prod_{i=1}^M ds_i$ on H_N by ds. Our integral will be the following:

$$I_{n,N} = \frac{1}{c_N} \int_{H_N} 2^n \operatorname{tr}(A^{2n}) e^{-\operatorname{tr}(AA^*)/2} \, ds,$$

where $c_N = \int_{H_N} e^{-\operatorname{tr}(AA^*)/2} \, ds$, and we claim:

Proposition 7.1 *Let $a_{k,n}$ denote the number of oriented chord diagrams D of order n satisfying $d(D) = k$, then*

$$I_{n,N} = \sum_{k=1}^{n+2} a_{k,n} N^k.$$

Proof We first note that the quadratic form $\operatorname{tr}(AA^*)$ appearing in the integral is simply $\sum_{i,j,k,l} |{}^i_l A^j_k|^2$. The restriction of this form to H_N is given in terms of our coordinates $s_1, \ldots, s_M$ by

$$\sum_{\text{type 1}} s_i^2 + 2 \sum_{\text{type 2}} s_i^2 + 4 \sum_{\text{type 3}} s_i^2.$$

For functions f, g on H_N we define $\langle f, g \rangle = \frac{1}{c_N} \int_{H_N} 2fg e^{-\operatorname{tr}(AA^*)/2} \, ds$. (This differs by factor 2 from the definition in [10].) For any one of our coordinates s_i we have $\langle s_i, s_i \rangle = 2, 1$ or $\frac{1}{2}$, according to whether s_i is of type 1, 2 or 3, respectively, and for any $i \ne j$, $\langle s_i, s_j \rangle = 0$. It follows that $\langle {}^i_j A^j_i, {}^j_i A^i_j \rangle = 2$ (whether or not $i = j$), and if $i \ne k$ or $j \ne l$ then $\langle {}^i_l A^j_k, {}^j_i A^k_l \rangle = \langle {}^i_l A^j_k, {}^l_k A^i_j \rangle = 1$. In all other cases $\langle \cdot, \cdot \rangle = 0$. That is, $\langle \cdot, \cdot \rangle$ is nonzero only when the second term is obtained from the first by a $\frac{\pi}{2}$ or $-\frac{\pi}{2}$ circular shift of the indices. If those indices are symmetric with respect to a π circular shift then $\langle \cdot, \cdot \rangle = 2$, and if they do not possess this symmetry then $\langle \cdot, \cdot \rangle = 1$.

Writing $\mathrm{tr}(A^{2n})$ explicitly gives

$$I_{n,N} = \sum_{1 \leq i_1,\ldots,i_{2n},j_1,\ldots,j_{2n} \leq N} \frac{1}{c_N} \int_{H_N} 2^n \, {}^{i_1}_{j_1} A^{i_2}_{j_2} \, {}^{i_2}_{j_2} A^{i_3}_{j_3} \, {}^{i_3}_{j_3} A^{i_4}_{j_4} \cdots {}^{i_{2n}}_{j_{2n}} A^{i_1}_{j_1} \, e^{-\mathrm{tr}(AA^*)/2} \, ds.$$

Wick's formula for each such integral (see [10]) expresses it as a sum of products of n pairings $\langle \cdot, \cdot \rangle$. (The factor 2^n in the integral corresponds to the factor 2 in our definition of $\langle \cdot, \cdot \rangle$.) We think of the factor ${}^{i_k}_{j_k} A^{i_{k+1}}_{j_{k+1}}$ as representing the kth square of our annulus (Fig. 2), and we think of the indices $i_1, \ldots, i_{2n}, j_1, \ldots, j_{2n}$ as colorings of the $4n$ edges of the annulus in N colors. A choice of such n pairings, and of the $4n$ indices, such that the n pairings are all nonzero, precisely corresponds to an oriented chord diagram D, together with a coloring of the $4n$ edges, which is monochromatic on each boundary component of $F(D)$. Whenever the indices in ${}^{i_k}_{j_k} A^{i_{k+1}}_{j_{k+1}}$ have π symmetry, then the $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ shifts coincide, but this is precisely compensated by the fact that in this case the pairing gives 2. It follows that each oriented chord diagram D contributes the number of colorings that it admits, which is $N^{d(D)}$. $\qquad\square$

We conclude this section by justifying the remark made in the introduction, that it is for similar reasons that the expected number of cycles in a permutation and the expected number of cycles of the walk along an oriented chord diagram are both $\Theta(\ln n)$. We do this by presenting a proof for permutations, which has the same structure as the proof we have presented for the problem of the present work.

We define the procedure for choosing a random permutation $\sigma \in S_n$ as follows. First randomly choose $\sigma(1)$. Then randomly choose $\sigma(\sigma(1))$ from among the remaining available elements. Go on until 1 is chosen, that is, until a cycle is closed. Then repeat the procedure starting with the smallest element that has not been chosen yet. Now we ask, what is the probability that the cycle containing 1 is completed precisely at the kth step, and so its length is k. For this we must make sure not to choose 1 at the first $k-1$ steps, and then make sure we do choose 1 at the kth step. The probability for that is $\frac{n-1}{n} \frac{n-2}{n-1} \cdots \frac{n-k+1}{n-k+2} \cdot \frac{1}{n-k+1} = \frac{1}{n}$. By symmetry, the probability for *any* element to lie in a cycle of length k is $\frac{1}{n}$, and so the probability P_k that a *randomly* chosen element lies in a cycle of length k is $\frac{1}{n}$. Now let L_k be the expected number of cycles of length k, then also $P_k = \frac{kL_k}{n}$. So we have $\frac{1}{n} = \frac{kL_k}{n}$, or $L_k = \frac{1}{k}$, so the expected number of all cycles is $h_n = \sum_{k=1}^{n} \frac{1}{k} = \Theta(\ln n)$, or moreover, $h_n \sim \ln n$.

References

1. Arratia, R., Tavare, S.: The cycle structure of random permutations. Ann. Probab. **20**(3), 1567–1591 (1992)
2. Babson, E., Hoffman, C., Kahle, M.: The fundamental group of random 2-complexes. arXiv:0711.2704
3. Cairns, G., Elton, D.M.: The planarity problem for signed Gauss words. J. Knot Theory Ramif. **2**(4), 359–367 (1993)
4. Cairns, G., Elton, D.M.: The planarity problem. II. J. Knot Theory Ramif. **5**(2), 137–144 (1996)
5. Carter, J.S.: Classifying immersed curves. Proc. Am. Math. Soc. **111**(1), 281–287 (1991)
6. Dowker, C.H., Thistlethwaite, M.B.: Classification of knot projections. Topol. Appl. **16**, 19–31 (1983)

7. Dunfield, N.M., Thurston, W.P.: Finite covers of random 3-manifolds. Invent. Math. **166**(3), 457–521 (2006)
8. Gauss, C.G.: Werke **8**, 271–286 (1900)
9. Harer, J., Zagier, D.: The Euler characteristic of the moduli space of curves. Invent. Math. **85**, 457–485 (1986)
10. Lando, S.K., Zvonkin, A.K.: Graphs on Surfaces and Their Applications. Encyclopaedia of Mathematical Sciences, Low-Dimensional Toplogy II, vol. 141. Springer, Berlin (2004)
11. Linial, N., Meshulam, R.: Homological connectivity of random 2-complexes. Combinatorica **26**(4), 475–487 (2006)
12. Lovász, L., Marx, M.L.: A forbidden substructure characterization of Gauss codes. Acta Sci. Math. (Szeged) **38**, 115–119 (1976)
13. Manturov, V.O.: A proof of V.A. Vassiliev's conjecture on the planarity of singular links. Izv. Math. **69**(5), 1025–1033 (2005)
14. Meshulam, R., Wallach, N.: Homological connectivity of random k-dimensional complexes. Random Struct. Algorithms **34**, 408–417 (2009)
15. Pippenger, N., Schleich, K.: Topological characteristics of random triangulated surfaces. Random Struct. Algorithms **28**(3), 247–288 (2006)
16. Read, R.C., Rosenstiehl, P.: On the Gauss crossing problem. In: Combinatorics. Proc. Fifth Hungarian Colloq., Keszthely, vol. II, pp. 843–876 (1976)
17. Rosenstiehl, P.: Solution algébrique du problème de Gauss sur la permutation des points d'intersection d'une ou plusieurs courbes fermés du plan. C. R. Acad. Sci. Paris Sér. A–B **283**, A551–A553 (1976)
18. Turaev, V.: Curves on surfaces, charts, and words. Geom. Dedic. **116**, 203–236 (2005)