WHAT IS A DEHN FUNCTION?

TIMOTHY RILEY

1. The size of a jigsaw puzzle

1.1. Jigsaw puzzles reimagined. I will describe jigsaw puzzles that are somewhat different to the familiar kind. A box, shown in Figure 1, contains an infinite supply of the three types of pieces pictured on its side: one five–sided and two four–sided, their edges coloured green, blue and red and directed by arrows. It also holds an infinite supply of red, green and blue rods, again directed by arrows.

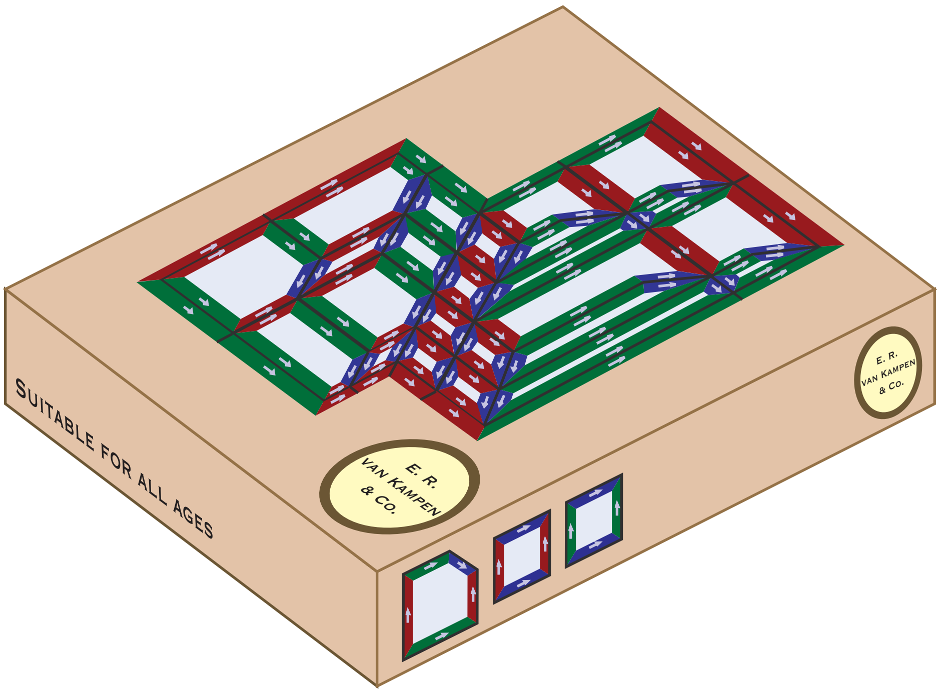

Figure 1. A puzzle kit.

A good strategy for solving a standard jigsaw puzzle is first to assemble the pieces that make up its boundary, and then fill the interior. In our jigsaw

Date: July 13, 2012.

The author gratefully acknowledges partial support from NSF grant DMS–1101651 and from Simons Foundation Collaboration Grant 208567. He also thanks Margarita Amchislavska, Costandino Moraites, and Kristen Pueschel for careful readings of drafts of this chapter, and the editors Matt Clay and Dan Margalit for many helpful suggestions.

puzzles the boundary, a circle of coloured rods end–to–end on a table top, is the starting point. A list, such as that in Figure 2, of boundaries that will make for good puzzles is supplied.

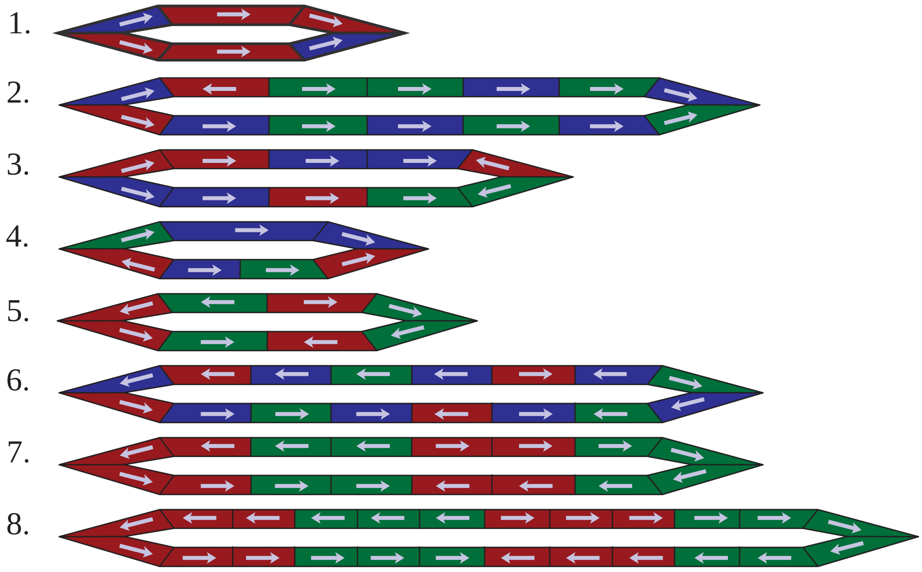

FIGURE 2. A list of puzzles accompanying the puzzle kit shown in Figure 1.

The aim of the puzzle is to fill the interior of the circle with the puzzle pieces in such a way that the edges of the pieces match up in colour and in the direction of the arrows. The way the pieces can be used differs significantly from a standard jigsaw: our pieces can be flipped and can be stretched. Flipping a piece reverses the sequence of coloured edges around its boundary; stretching it does not disturb their order or directions.

Solutions to Puzzles 2, 3 and 4 from the above list are shown in Figure 3. The box in Figure 1 displays a solution to Puzzle 7.

When solving a puzzle you are allowed to push together rods in the boundary circle as happens in our solutions to Puzzles 2 and 3. All that is required for a valid solution is that the completed puzzle be flat on the table top (that is, be *planar*), the colours and arrows should all match up, the boundary should be the prescribed circuit of rods, and the interior should be entirely filled with the supplied pieces.

Solutions are not unique in general — for example, Figure 3 shows two solutions to Puzzle 4. As will become apparent (in the light of Lemma 2.16, especially), there are circles of rods that give puzzles which have no solution; but, assuming the manufacturer has been diligent, all on the supplied list should be solvable.

Exercise 1.1. Solve the remaining puzzles on the list in Figure 2.

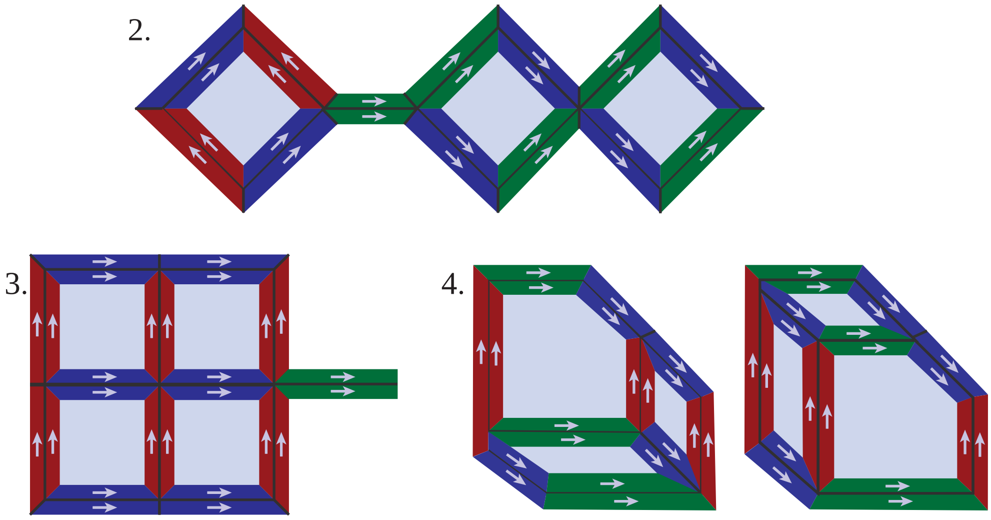

FIGURE 3. Solutions to three of the puzzles listed in Figure 2.

(The flexibility of the pieces and their limitless number would surely prevent these puzzle kits from ever being manufactured, but a computer implementation would seem in range of a skillful programmer.)

The puzzle kit of Figure 1 is one of many possibilities. In general, a puzzle kit will have a finite number of types of pieces, each in infinite supply. Each piece will be a polygonal tile whose edges are coloured and are directed by arrows. A polygonal tile is allowed to have any number of edges greater than or equal to one. We accommodate the possibility of a tile having only one or two edges by allowing the edges to curve: a one– or two–sided tile could, for example, be circular with its perimeter divided into one or two edges, respectively. The kit will also include an infinite supply of rods which are also decorated by arrows and are given one of finitely many possible colours. The set of colours of the edges of the puzzle pieces will always be a subset of the set of colours of the rods.

1.2. **The sizes of puzzles.** So what, then, is a Dehn function? When buying a puzzle kit, you are likely to want to know how hard the puzzles it affords can be. There are many ways of interpreting "hard" here, but, just as for standard jigsaw puzzles, a reasonable first consideration is how many pieces the puzzles require to complete. That is what a Dehn function measures. We look at all circles of at most n rods (there are only finitely many since the rods have only finitely many colours) which give puzzles that admit solutions, and we ask for the minimum number N such that all those puzzles have solutions that use no more than N pieces. The *Dehn function* maps $n \mapsto N$.

In the next section we will see the inspiration for puzzles: they are visual representations of calculations in groups. We will redefine the Dehn function as a measure of the complexity of a group's *Word Problem* and will then reconcile that with the definition in terms of puzzles. In Section 3 we will see how Dehn functions relate to *soap–film geometry*: they record the areas of discs spanning loops in certain spaces associated to groups. Then we will see in Section 4 that Dehn functions are *large–scale invariants* of groups in that the Dehn functions of two groups which look similar on the large–scale grow in the same way. Section 5 is a brief survey of which functions occur as Dehn functions; Section 6 explores some exotic examples of Dehn functions; and Section 7 offers suggestions for further reading.

2. A complexity measure for the Word Problem

2.1. **Words and presentations.** A common way for a group to arise is via a *presentation*. For example, if the group is the fundamental group of some topological space such as a surface, a 3–manifold, a knot complement etc., then you are likely to obtain it as a presentation via the *Seifert–van Kampen Theorem*.

A *word* on a set $A = \{a_1, \ldots, a_n\}$ of symbols (an *alphabet*) is a finite string of the symbols, possibly including repetitions. The set $A^{\pm 1}$ is the union of A with the set $A^{-1} = \left\{{a_1}^{-1}, \ldots, {a_n}^{-1}\right\}$ of corresponding *inverse symbols*; an associated *inverse map* carries $a_i \mapsto {a_i}^{-1}$ and ${a_i}^{-1} \mapsto a_i$, and extends to words on $A^{\pm 1}$ by $x_1 \cdots x_k \mapsto {x_k}^{-1} \cdots {x_1}^{-1}$. A *cyclic permutation of a word* $x_1 \cdots x_i x_{i+1} \cdots x_k$ is a word $x_{i+1} \cdots x_k x_1 \cdots x_i$. The *length* of the word $x_1 \cdots x_k$ is k.

A (finite) *presentation* for a group Γ may be denoted $\langle A \mid R \rangle$ or

$$\langle a_1, \ldots, a_n \mid r_1, \ldots, r_m \rangle,$$

where $A = \{a_1, \ldots, a_n\}$ is a set of symbols known as the *generators* and $R = \{r_1, \ldots, r_m\}$ is a set of words on $A^{\pm 1}$ which we call *defining relations* (or *relators*). A further convenient way to write a presentation is

$$\langle a_1, \ldots, a_n \mid u_1 = v_1, \ \ldots, \ u_m = v_m \rangle,$$

which denotes

$$\langle a_1, \ldots, a_n \mid u_1 {v_1}^{-1}, \ldots, u_m {v_m}^{-1} \rangle.$$

Elements of Γ are represented by words on $A^{\pm 1}$. The defining relations tell us when words w and w' represent the same group element: specifically, when w' can be obtained from w by a finite sequence of the following moves —

(i) *free reduction*: remove a substring $a_i {a_i}^{-1}$ or ${a_i}^{-1} a_i$ from within a word;
(ii) *free expansion*: insert a substring $a_i {a_i}^{-1}$ or ${a_i}^{-1} a_i$ into a word;

(iii) *apply a defining relation*: replace a substring u in a word with a new substring v such that uv^{-1} or vu^{-1} is a cyclic permutation of one of the words in R.

A *null–sequence* for a word w is such a sequence that transforms w to the empty word.

The group operation is concatenation: the product of the group elements represented by the words u and v is the group element represented by uv.

Due diligence requires that we now verify that what we have defined really is a group. The reader can check that concatenation of words gives a well–defined operation. The empty word (the word with no symbols) represents the identity, as do each of the defining relations and, more generally, all words that admit null–sequences. The inverse of a group element represented by a word u is represented by the word u^{-1}. Associativity of the group operation follows from the associativity of the operation of concatenating words.

In the special case where R is empty, the group is the *free group* $F(A)$. However, note that in Magalit and Clay's chapters, elements of $F(A)$ are *reduced* words — that is, words that do not allow any free reductions — whereas for us elements of $F(A)$ are equivalence classes of words.

Example 2.1. The cyclic group of order m is presented by $\langle a \mid a^m \rangle$.

Example 2.2. $\mathbb{Z} \times \mathbb{Z}$ is presented by $\langle a, b \mid a^{-1}b^{-1}ab \rangle$. Here is an example of a null–sequence with respect to this presentation:

$$ba^2ba^{-2}b^{-2} \;\to\; ba^2a^{-1}ba^{-1}b^{-2} \;\to\; baba^{-1}b^{-2} \;\to\; bb^{-1} \;\to\; \textit{empty word}.$$

First a substring ba^{-1} is replaced by an $a^{-1}b$ by applying the defining relation (as $ba^{-1}(a^{-1}b)^{-1} = ba^{-1}b^{-1}a$ is a cyclic permutation of $a^{-1}b^{-1}ab$), then there is a free reduction, then a substring $aba^{-1}b^{-1}$ (also a cyclic permutation of $a^{-1}b^{-1}ab$) is replaced by the empty word, and then there is a final free reduction.

Example 2.3. $\langle\, a, b, c \mid a^{-1}b^{-1}ab = c,\ ac = ca,\ bc = cb \,\rangle$ presents $\mathcal{H}_3$, the three–dimensional integral Heisenberg group, which is the multiplicative group of three–by–three matrices of the form

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix},$$

where $x, y, z \in \mathbb{Z}$. The matrices

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

correspond to a, b, c, respectively.

Exercise 2.4. Find null–sequences for $a^{-2}b^{-1}ab^2ab^{-1}$ and $a^{-4}b^{-2}a^2b^4a^2b^{-2}$ with respect to the presentation for $\mathcal{H}_3$ in Example 2.3.

Exercise 2.5. This exercise establishes an alternative way of viewing the group Γ defined by a presentation $\langle A \mid R \rangle$. Let $F(A)$ denote the free group on the alphabet A and $\langle\langle R \rangle\rangle$ denote the smallest normal subgroup of $F(A)$ containing all the elements represented by the words in R. Show that

$$1 \to \langle\langle R \rangle\rangle \to F(A) \to \Gamma \to 1,$$

with the maps defined in the natural way, is a short exact sequence — in other words, the image of each map in the sequence is the kernel of the next. So $\Gamma \cong F(A)/\langle\langle R \rangle\rangle$.

Exercise 2.6. The previous exercise implies that the words w representing the identity in Γ are precisely those that are equal in $F(A)$ to products

$$\left(u_1{}^{-1} r_{j_1}{}^{\epsilon_1} u_1\right) \cdots \left(u_N{}^{-1} r_{j_N}{}^{\epsilon_N} u_N\right)$$

of conjugates of defining relations or their inverses — that is, each u_i is a word on $A^{\pm 1}$, each r_{j_i} is in R, and each ϵ_i is ± 1. Show the minimal such N for a given w is equal to the minimal N such that there is a null–sequence for w including N application–of–a–relator moves.

2.2. **The Dehn function and the Word Problem.** It would seem that a minimum standard for being able to work with a group given by a finite presentation is that we should be able to tell whether or not two words represent the same group element, or equivalently whether or not a word represents the identity. This is the known as the *Word Problem* for the presentation. It is the first of three problems singled out by Max Dehn in providential writings about a hundred years ago — see [17], which is included in the collection [18] of Stillwell's translations of Dehn's papers.

Well, a word represents the identity when it can be converted to the empty word via a finite sequence of free reductions, free expansions, and applications of defining relations. So counting how many moves this takes gives a natural measure of how hard it is to work with the presentation. This is what the Dehn function does.

To be precise, the *Dehn function* $\mathbb{N} \to \mathbb{N}$ maps n to the minimal number N such that if w is a word of length at most n that represents the identity, then there is a null–sequence for w involving at most N applications–of–defining–relations moves. There are only finitely many words of length at most n since the alphabet A is finite, and so N is well–defined.

(This is essentially how Madlener and Otto introduced the Dehn function, under the name *derivational complexity*, in [30] around the same time as Gromov, in a manuscript which became [27], defined an equivalent geometric invariant in the manner we will discuss in Section 3. The name *Dehn function* was coined by Gersten in [21].)

Remark 2.7. That we only count applications of defining relations here, rather than all the moves, is just a technicality that allows for the cleanest possible statement of Lemma 2.16 below; if we counted all the moves we would get an equivalent function in the sense defined in Section 3.1. This is because if there is a null–sequence for w that uses N applications–of–defining–relations moves, then there is one that uses N applications–of–defining–relations moves and at most $kN + \ell(w)$ free–reduction moves (and no free–expansion moves), where $\ell(w)$ denotes the length of w and k is the length of the longest word in R. (This follows from Lemma 2.16 below and its proof.)

2.3. **Some calculations of Dehn functions.**

Proposition 2.8. *The Dehn function $f(n)$ of the presentation $\langle a \mid a^m \rangle$ of the cyclic group of order m is the greatest integer less than or equal to n/m.*

Here is a proof. A word on $\left\{a, a^{-1}\right\}$ represents the identity in $\langle a \mid a^m \rangle$ when it can be converted to a^{rm} for some $r \in \mathbb{Z}$ by free reductions. Doing so only shortens the word and costs nothing as far as Dehn function is concerned. Then, $|r|$ applications of the defining relation reduce the word to the empty word by deleting $a^{\pm m}$ substrings. So $f(n) \leq n/m$. Consideration of the effect of free reductions, free expansions and applications of the one defining relator on the sum of the exponents of the letters in a word leads to the lower bound. Free reductions and free expansions leave it unchanged and applying $a^m = 1$ changes it by at most m. So it is not possible to reduce a^{rm} to the empty word using fewer than $|r|$ applications of the defining relation.

Exercise 2.9. Show that the Dehn function of a finite presentation of a finite group is always bounded above by Cn for some constant C.

Exercise 2.10. Compute the Dehn function exactly for some finite presentations of finite groups.

Proposition 2.11. *The Dehn function $f(n)$ of the presentation $\langle a, b \mid a^{-1}b^{-1}ab \rangle$ of $\mathbb{Z} \times \mathbb{Z}$ grows quadratically. More precisely, $(n-3)^2 \leq 16 f(n) \leq n^2$ for all n.*

For the upper bound, suppose w is a word on $\{a, b\}^{\pm 1}$ which has length n and represents the identity. Then the number of a present in w equals the number of a^{-1} and the number of b present equals the number of b^{-1}. The defining relation $a^{-1}b^{-1}ab = 1$ can be re–expressed as $ab = ba$ or $ab^{-1} = b^{-1}a$ or $a^{-1}b = ba^{-1}$ or $a^{-1}b^{-1} = b^{-1}a^{-1}$, and so can be used to shuffle an $a^{\pm 1}$ past a $b^{\pm 1}$. So if we collect the $a^{\pm 1}$ together by shuffling them past the past the $b^{\pm 1}$ and then freely reduce we will reach the empty word. There are at most n of each, and so $f(n) \leq n^2$.

Exercise 2.12. Sharpen these estimates in the above paragraph to get $f(n) \leq n^2/16$.

Before we address the lower bound, here are two exercises concerning upper bounds for Dehn function obtainable by the approach we used for $\mathbb{Z} \times \mathbb{Z}$.

Exercise 2.13. Show that for your favorite presentation of any finitely generated abelian group, the Dehn function is at most a constant times n^2.

Exercise 2.14. Use the fact that each element of $\mathcal{H}_3$ can be expressed in a unique way as a word of the form $a^p b^q c^r$ for some integers p, q and r to show that the Dehn function $f(n)$ of the presentation

$$\langle\, a, b, c \mid a^{-1}b^{-1}ab = c,\ ac = ca,\ bc = cb \,\rangle$$

of $\mathcal{H}_3$ satisfies $f(n) \le Cn^3$ for all n and a suitable constant C. (In fact, $f(n)$ also admits a cubic lower bound — see [3].)

The lower bound in Proposition 2.11 comes from the fact that the word $w_k := a^{-k}b^{-k}a^k b^k$ has length $4k$ and represents the identity in $\mathbb{Z} \times \mathbb{Z}$, but any null–sequence carrying it to the empty word requires at least k^2 applications of the defining relation $a^{-1}b^{-1}ab$. A direct proof in terms of null–sequences would be cumbersome and unenlightening. A more natural proof can be given using geometric techniques we will see in Section 6.1. Here is a somewhat surprising alternative approach (which the reader could skip as we will not need it later). I believe it originates in [3].

As per Exercise 2.6, suppose u_i are words on $\{a, b\}^{\pm 1}$ and $\epsilon_i = \pm 1$ so that

$$W_k \;=\; \left(u_1{}^{-1}(a^{-1}b^{-1}ab)^{\epsilon_1}u_1\right)\cdots\left(u_N{}^{-1}(a^{-1}b^{-1}ab)^{\epsilon_N}u_N\right),$$

equals w_k in $F(a, b)$ and N is the minimum number of times the defining relation has to be applied to reduce w_k to the empty word in $\langle a, b \mid a^{-1}b^{-1}ab\rangle$. Then in $\mathcal{H}_3$ the word W_k represents the same element as

$$\left(u_1{}^{-1}c^{\epsilon_1}u_1\right)\cdots\left(u_N{}^{-1}c^{\epsilon_N}u_N\right),$$

and so also as $c^{\epsilon_1}\cdots c^{\epsilon_N}$, since c commutes with a and b. But a calculation using the matrix representation of $\mathcal{H}_3$ given in Example 2.3 shows that w_k and c^{k^2} represent the same element in $\mathcal{H}_3$, and that c has infinite order in $\mathcal{H}_3$. So $N \ge k^2$.

Exercise 2.15. Formulate and prove an analogue of Proposition 2.11 for the presentation $\langle a, b, c \mid abc = 1, b = ac\rangle$ for $\mathbb{Z} \times \mathbb{Z}$.

2.4. **Puzzles kits are presentations.** The beginnings of how to reconcile the definition of Dehn function in terms of null–sequences with that given in terms of puzzles may be evident. A finite presentation corresponds to a puzzle set. The colours of the rods and the edges of the puzzle pieces correspond to the generating set A. Each puzzle piece corresponds to a defining relation $r \in R$: following the boundary of the piece either clockwise or anti-clockwise from some starting point, one reads r on translating the colours to generators and understanding that travel against the direction of an edge should mean an inverse letter. For example $\mathcal{H}_3$, as presented in Example 2.3, corresponds to the puzzle kit of Figure 1.

We claim that for a word w, the problem of finding a sequence of free reductions, free expansions, and applications of defining relations that carries it to the empty word, is equivalent to solving the puzzle where, on translating colours to generators and taking into account the directions, starting from some vertex $*$ one reads w around the initial circle of rods. This is because null–sequences relate to solutions of puzzles, as we will explain.

Here is how to obtain a sequence of free reductions and applications of defining relations carrying w to the empty word from a solution to the corresponding puzzle. Disassemble the completed puzzle, rod–by–rod and piece–by–piece in any way using the following moves until all that remains is the vertex $*$:

(i) remove any pair of rods which form a spike coming out of the puzzle — that is, run side–by–side and only meet the rest of the puzzle at only one end;
(ii) remove any piece which abuts the boundary circle the puzzle, and then reconfigure the boundary circle so as to close up the resulting hole.

An example of such a disassembly of a completed puzzle is shown in Figure 4.

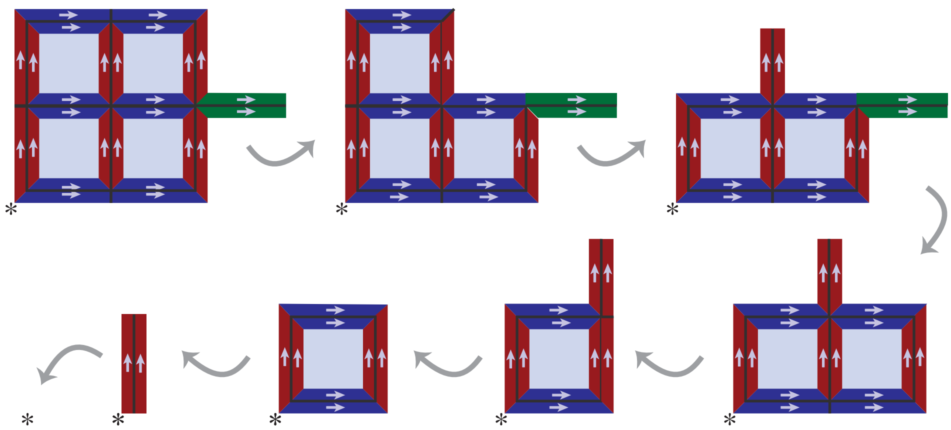

FIGURE 4. A disassembly of a completed puzzle.

Evidently, the word corresponding to the circuit of rods around the perimeter changes by a free reduction in the first type of move and by an application-of-a-relator in the second, and the number of application-of-a-relator moves is equal to the area of the puzzle. In the example of the disassembly in Figure 4, the corresponding null–sequence (reading around the diagrams anticlockwise from $*$) is

$$\begin{aligned} c^2baa^{-1}bc^{-2}b^{-2} &\to c^2baa^{-1}c^{-1}bc^{-1}b^{-2} \to c^2baa^{-1}c^{-1}bb^{-1}c^{-1}b^{-1} \\ &\to c^2bc^{-1}bb^{-1}c^{-1}b^{-1} \to cb^2b^{-1}c^{-1}b^{-1} \to cbc^{-1}b^{-1} \\ &\to bb^{-1} \to \text{empty word}. \end{aligned}$$

The natural way to look to translate from a null–sequence for w to a solution for the corresponding puzzle, is to reverse the procedure: place successive pieces and rods on the table top as dictated by the null–sequence until a solution to the puzzle has been assembled. This is workable and the agreement between the number of puzzle pieces and the number of applications-of-defining-relators becomes evident, but there is a technical problem. It may be that the puzzle cannot be kept in the table–top when assembled in this manner — that is, planarity may break down. (When this problem occurs, in some sense the null-sequence must have been inefficient, and so the difficultly can be avoided by *improving* the null-sequence.)

Seen in this group theoretic light, a completed puzzle is known as a *van Kampen diagram* for w. Modulo the difficulty mentioned above, we have established the following lemma which is closely related (via Exercise 2.6) to a foundational lemma of van Kampen from [44]. (A proof of this lemma which deals carefully with the planarity issue can be found in [10].)

Lemma 2.16. *In a finite presentation for a group, the words that represent the identity are precisely those that correspond to puzzles which admit solutions. Moreover, the Dehn function defined in terms of puzzles agrees exactly with the Dehn function defined in terms of null–sequences.*

The reader may like to revisit Exercise 2.4 in the light of this lemma. The two words in that exercise correspond to Puzzles 5 and 7 of Figure 2.

2.5. **Solving the Word Problem.** Our next result, found for example in Gersten's survey [23], gives a direct connection between the Dehn function of a finite presentation and solving its Word Problem. When we discuss algorithms in what follows, you can think of programs written in any reasonable (pseudo–)programming language, running on any computer you like. But to be formal and precise, we mean an idealized computing device known as a *Turing machine*. A function $g : \mathbb{N} \to \mathbb{N}$ is *recursive* when there is an algorithm which on input n, outputs $g(n)$. There are functions $\mathbb{N} \to \mathbb{N}$ which are not recursive. Indeed, there are finite presentations for which there is no algorithm to decide the Word Problem — we will revisit this in Section 6.4.

Proposition 2.17. *For a finite presentation $\langle A \mid R \rangle$ of a group with Dehn function $f : \mathbb{N} \to \mathbb{N}$, the following are equivalent.*

(i) *There is an algorithm which, on input a word on $A^{\pm 1}$, will declare whether or not that word represents the identity.*
(ii) *There is a recursive function $g : \mathbb{N} \to \mathbb{N}$ such that $f(n) \leq g(n)$ for all n.*
(iii) *f itself* is *a recursive function.*

Here is most of a proof. Given an upper bound $g(n)$ on the Dehn function, it is always possible to reduce a word of length w that represents the identity to the empty word using a null–sequence with at most $g(n) + kg(n) + n$ moves — see Remark 2.7. So if $g(n)$ is recursive, we can test whether a word of length n represents the identity by trying all null–sequences that use at most that number of moves. If, on the other hand, we have an algorithm which solves the word problem, then we can calculate $f(n)$ by the following arduous procedure. First list all words on $A^{\pm 1}$ of length at most n; discard from the list all that fail to represent the identity; and then, for each word w that remains, calculate the minimal number of application–of–a–relator moves necessary to reduce w to the empty word (or the minimal number of pieces in a solution to the puzzle corresponding to w).

Exercise 2.18. Complete this proof by explaining how to do the final step.

2.6. **How hard is the Word Problem really?** To be honest, the Dehn function is not a good measure of the difficulty of the Word Problem. It is a worst–case measure of how long a direct attack on the Word Problem by successively applying defining relations and free reductions and expansions will take. But that attack is *non–deterministic*: in order to reduce a word to the empty word using the shortest possible null–sequence, the right choices need to be made about which moves to apply and where in the word to apply them. Making this a deterministic algorithm — that is, removing the need to make choices — appears to cost an exponential leap in running time. It could be done by exhaustively trying all possible sequences of moves of a given length (specified by the Dehn function). But this rarely seems worth the trouble as there are usually far more efficient ways to tackle the Word Problem as we will now see.

As a simple example, consider $\mathbb{Z}\times\mathbb{Z}$ presented by $\langle a, b \mid ab = ba\rangle$, which has a quadratically growing Dehn function (Proposition 2.11). To tell whether or not a word on $\{a, b\}^{\pm 1}$ represents the identity, it is enough just to add up the exponents of the $a^{\pm 1}$ and $b^{\pm 1}$ present and check whether both are zero.

Another example is the Heisenberg group $\mathcal{H}_3$ of Example 2.3. Its Dehn function grows like $n \mapsto n^3$ (see Exercise 2.14). But viewing $\mathcal{H}_3$ as a matrix group and calculating by multiplying matrices (using techniques like writing the entries in binary) is an efficient means of checking whether a word represents the identity.

The same strategy works for $\langle a, s \mid s^{-1}as = a^2 \rangle$, whose Dehn function grows exponentially fast as we will see in Section 6.1. It can be represented by matrices via $a \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $s \mapsto \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}$.

In Sections 6.2 and 6.3 we will see examples of groups with even faster growing Dehn functions. The Dehn function of Baumslag's group

$$\langle a, t \mid (t^{-1}at)^{-1}a(t^{-1}at) = a^2 \rangle$$

grows like a tower of exponential functions, but a polynomial time solution to its Word Problem was recently found by Miasnikov, Ushakov and Won in [32]. I believe that the Word Problems of the *hydra examples* Γ_k have similarly efficient solutions, despite having Dehn functions growing like Ackermann's fast growing functions A_k (as will be explained in Section 6.3).

There are examples where the gap between the Dehn function and the running time of the most efficient algorithm to solve the Word Problem is similarly large or even greater. Cohen, Madlener & Otto [16] gave examples where the gap is like $n \mapsto A_n(n)$, which is *non–primitive recursive*, and recently Kharlampovich, Myasnikov & Sapir [34] gave examples where the gap is at least any given recursive function. Their technique is to take algorithms which are known to halt always but, on some inputs, take an amount of time comparable to the values of these especially fast–growing functions, and *embed* them in the Word Problem (lots of work goes into making this precise) for a suitable finite presentation of a group. This is done in such a way that checking that certain words represent the identity by using direct applications of generators and relations is similar to running these very slow algorithms, and so makes the Dehn function grow similarly quickly. But this is unnecessary work as the issue of whether those words represent the identity really only hinges on whether those slow algorithms terminate... and they always do. Cohen gives an entertaining description of the phenomenon with the help of a magical salmon (!) in [15].

All this is not to detract from the Dehn function. It represents a compelling link between algebraic computation (in a suitably restricted sense) and, as we will see in Section 3, geometry.

2.7. **The challenge of making demanding puzzles.** Here is an informal question of potential application to cryptography. Recall that each puzzle set comes with a list of suggested puzzles. Assuming they are to be challenging puzzles, we might wonder how the manufacturer can come up with the list. Ideally, generating the list should be much easier than solving the puzzle.

Project 2.19. Are there finite presentations in which it is easy to generate words that represent the identity, but hard to solve the puzzle?

I do not mean that the puzzles necessarily involve a particularly large number of pieces, but that it is hard to describe solutions, or even provide some extrinsic proof why a solution must exist. (Thanks to Sasha Ushakov for conversations on this topic.)

2.8. **Infinite presentations.** It is possible to define groups via infinite presentations, by relaxing the definition in Section 2.1 to allow A or R, or both, to be infinite. The definition of Dehn function given in Section 2.2 remains well–founded for infinite presentations, but is less compelling. The Dehn functions of any two finite presentations of the same group grow in similar ways, as we will discuss in Section 4.2. But for infinite presentation they can be very different; after all, if we go to the extreme of including all the words that represent the identity in the set of defining relations, then the Dehn function will only take values 0 or 1.

Nevertheless, see [25] for an interesting study of Dehn functions of infinite presentations. And there are stark open problems, such as that (for which I thank I. Kapovich) which headlines the following project:

Project 2.20. Which functions are Dehn functions of infinite presentations of $\mathbb{Z}\times\mathbb{Z}$? (It may be natural to work up to *biLipschitz equivalence*, as defined in Thomas' chapter.) In particular:

(i) Can you get $n \mapsto n^3$? What general upper bound can you find for Dehn functions of infinite presentations of $\mathbb{Z} \times \mathbb{Z}$?
(ii) Can you get $n \mapsto n^{1/2}$?
(iii) How do your investigations change if you replace $\mathbb{Z} \times \mathbb{Z}$ with, for example, $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$?

It is possible to vary the definition of Dehn function by *weighting* the defining relations: assign a strictly positive real number to each, which it contributes whenever it is applied. (This is the standard Dehn function when all the weights are 1.) The next exercise show that the weights do not significantly change the scope of Dehn functions of finite presentations.

Exercise 2.21. Show that if $f(n)$ is the standard Dehn function and $f_w(n)$ is a weighted Dehn function of a finite presentation, then there exists $C > 0$ such that $f(n)/C \leq f_w(n) \leq C f(n)$ for all n.

Exercise 2.22. Revisit Project 2.20 admitting weighted Dehn functions.

3. Isoperimetry

3.1. **Ox–hide and soap film.** In Virgil's Aeneid, Dido is described as purchasing land on which to found the city of Carthage. For an agreed price, the sellers allow her all she can enclose with a single ox-hide. She duly cuts the ox–hide into thin strips and arranges it in a semi–circular arc between two points on the (roughly straight) coastline, thereby claiming a far larger parcel of land than the sellers had envisaged. In addition to skillful ox–hide slicing, her success was based on her ability to give the optimal solution to a form of *isoperimetric problem*, namely the problem of finding the arc of a given length which, when connecting two points on a straight line in the plane, will enclose the largest area.

In general, an *isoperimetric problem* concerns determining the maximal area, volume or the like that a shape can have when its boundary is constrained in some way. The Dehn function of a finitely presented group Γ relates to an isoperimetric problem concerning spanning loops with discs in a suitable space. This sort of isoperimetry has a familiar physical manifestation: a wire loop lifted out of soap solution emerges spanned by an area–minimizing surface in the form of a soap film. What can serve as this "suitable space" is the subject of the next section.

3.2. **Spaces associated to finite presentations of groups.** Here are some qualities we should look for in a space associated to Γ if we are to consider the isoperimetry of spanning loops with discs. It should be simply connected — that is, every loop should span a disc, by which we mean that every continuous map from $S^1 = \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\right\}$ to the space should extend to a continuous map from $D^2 = \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\right\}$; there should be reasonable notions of the lengths of paths and the areas of discs; and it should resemble Γ in some strong sense as we shall see.

A group Γ with a finite generating set A is naturally a metric space: the distance between group elements g and h is the length of the shortest word on $A^{\pm 1}$ representing $g^{-1}h$. This distance function is known as the *word metric*. But, as such, Γ appears a sparse cloud of points, and falls short of our requirements. Gromov puts it colourfully in an inspiring introduction to [27]:

> *This space may appear boring and uneventful to a geometer's eye since it is discrete and the traditional local (e.g. topological and infinitesimal) machinery does not run in* Γ.

We will give two ways to flesh out this space when Γ has a finite presentation $\langle A \mid R \rangle$.

The first is combinatorial. The Cayley graph (introduced in Margalit's chapter in this volume) adds some substance to Γ: the vertices are the group elements and there is a directed edge labelled a from g to ga for every $g \in \Gamma$ and $a \in A$. But this space also appears insufficient. It is a graph and so is not usually simply connected and, anyway, to draw a connection with the Dehn function, we surely need to add structure reflecting the set R of defining relations.

So, for every $r \in R$, we add a family of $\ell(r)$–sided polygonal faces (or discs when $\ell(r)$ is 1 or 2) to the Cayley graph, where $\ell(r)$ denotes the length of the word r. For every loop that is made up of a succession of edges along which we read r (traversing edges against their orientations, being the way inverse letters arise), one such face is attached by gluing its boundary edge–by–edge to the loop. The resulting space is named $\widetilde{K}$ and is known as the *Cayley 2–complex* of $\langle A \mid R \rangle$. We will establish that it is simply connected in Section 3.3.

Two examples of $\widetilde{K}$ are shown in Figure 5. The Cayley graph of $\mathbb{Z} \times \mathbb{Z}$ with respect to a two–element generating set $\{a, b\}$ is a grid–like graph and we get the Cayley 2–complex for $\langle a, b \mid a^{-1}b^{-1}ab \rangle$ by filling in squares with faces, with the result that $\widetilde{K}$ is a plane. On the other hand, the Cayley 2–complex of the presentation $\langle a, b \mid \rangle$ of the free group $F(a, b)$ is simply the Cayley graph as there are no defining relations and so no faces.

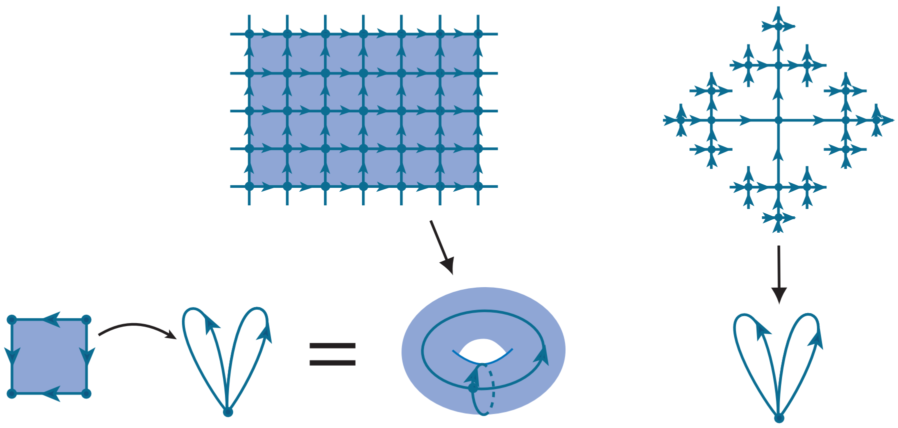

FIGURE 5. Portions of the Cayley 2–complexes for the presentations $\langle a, b \mid a^{-1}b^{-1}ab \rangle$ and $\langle a, b \mid \rangle$ of $\mathbb{Z} \times \mathbb{Z}$ and F_2.

A useful additional perspective on $\widetilde{K}$ is available to those readers with some background in algebraic topology. The reason for the notation $\widetilde{K}$ is that it is the universal cover of a certain finite 2–dimensional complex K, which is illustrated in Figure 6. This K has Γ as its fundamental group and is assembled as dictated by the presentation as follows. Start with a lone vertex. Attach to that vertex, both ends of one directed edge for each element $a \in A$ labelling that edge by a. The result is called a *rose*. Then for each $r \in R$, attach to the rose one $\ell(r)$–sided face, where $\ell(r)$ denotes the length of the word r, along the edge–loop around which we read r. That K has fundamental group Γ is a consequence of the Seifert–van Kampen Theorem. In one of the examples of Figure 5, K is a torus, and in the other it is a rose with two *petals*.

For most presentations, $\widetilde{K}$ is hard to visualize — try the example of the presentation for $\mathcal{H}_3$ in Example 2.3, for instance.

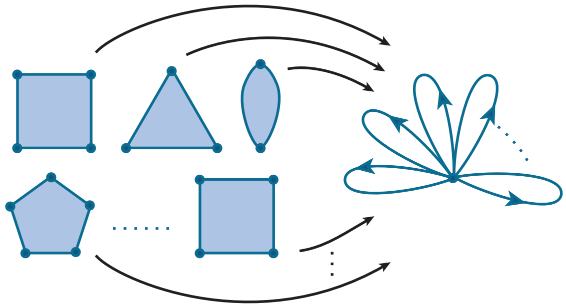

FIGURE 6. The 2–complex K associated to a presentation.

We give $\widetilde{K}$ a metric by taking each of its edges to have length one and each of the faces to be a regular Euclidean polygon whose sides are all of length one. (When we need a one–sided "polygon" use a Euclidean disc of perimeter one, and when we need a two–sided polygon use a Euclidean disc whose perimeter has length two and is divided into two semicircular "edges".) We declare the distance between two points there to be the length of the shortest path connecting them.

The second type of space we associate to a finite presentation of a group Γ is a *Riemannian manifold* $\widetilde{M}$. A *manifold* is a space which on the small scale resembles Euclidean space $\mathbb{R}^n$ for some n; it is *Riemannian* when it is endowed with a certain structure which gives rise to notions such as lengths of paths, angles between paths, areas or volumes of subsets and so on.

It turns out that there is a lot of flexibility over what Riemannian manifolds will do for our purposes. It should be simply connected and should *coarsely resemble* Γ. For example, for $\mathbb{Z} \times \mathbb{Z}$ we can take $\widetilde{M}$ to be the plane — $\mathbb{Z} \times \mathbb{Z}$ and $\widetilde{M}$ bear a coarse resemblance in that $\mathbb{Z} \times \mathbb{Z}$ can be regarded as the set of points in the plane with integer coordinates; if you squint at $\mathbb{Z} \times \mathbb{Z}$ it looks like a plane.

To be more precise (and more technical — skip this paragraph if you like), the universal cover of any compact connected Riemannian manifold M (with no boundary) which has Γ as its fundamental group can serve as $\widetilde{M}$. (For the example of $\mathbb{Z} \times \mathbb{Z}$ presented by $\langle a, b \mid a^{-1}b^{-1}ab \rangle$, a torus can serve as M.) Such an M always exists when Γ is finitely presented: in fact, we can take M to be the boundary of a small neighbourhood of a copy of the complex K embedded in $\mathbb{R}^5$. The manner in which Γ resembles this $\widetilde{M}$ stems from the action of Γ acts on $\widetilde{M}$ by deck transformations. Fix a basepoint $p \in \widetilde{M}$. The map $\Phi : \Gamma \to \widetilde{M}$ taking $g \in \Gamma$ to the translate $g \cdot p$ of p by g, is a *quasi–isometry* — a type of map whose existence captures a precise notion of its domain and target being *coarsely the same*. Quasi–isometries are the subject of Thomas' chapter; we will return to them in Section 4.1.

3.3. **Isoperimetry in the Cayley 2–complex.** Here we will explain the connection between the Dehn function of a finite presentation $P = \langle A \mid R\rangle$ for Γ and isoperimetry in the space $\widetilde{K}$ constructed in the previous section. We will consider $\widetilde{M}$ in Section 3.4.

The outgoing edges from any given vertex in $\widetilde{K}$ are labelled in one–to–one correspondence with the elements of the generating set A. The same is true of the incoming edges. So words w on $A^{\pm 1}$ correspond one–to–one with paths that traverse a succession of edges from any given starting vertex $g \in \Gamma$. Moreover, since gw and g are equal in Γ if and only if w represents the identity —

Lemma 3.1. *Words that represent the identity in* Γ *correspond one–to–one with loops based at some fixed vertex that traverse a succession of edges* (edge–loops).

Suppose ρ is an edge–loop in $\widetilde{K}$ and w is the corresponding word. As w represents the identity, the associated puzzle has a solution; that is, there is a van Kampen diagram Δ for w. View Δ as a 2–dimensional planar complex: the puzzle pieces being faces, and the rods and the edges of the puzzle pieces being edges in the complex. Figures 7, 10, and 11 are examples of van Kampen diagrams viewed as such complexes. The edges in this complex inherit orientations and labellings by generators from the rods and from the sides of the puzzle pieces.

It is possible to regard Δ as a disc spanning ρ in $\widetilde{K}$, as we will now explain. This "disc" may be *singular* in that it may have one–dimensional portions as in Examples 2 and 3 in Figure 3.

Suppose we choose a vertex v in Δ and choose any vertex g in $\widetilde{K}$. We will explain that there is a unique map from all the 1–skeleton of Δ (that is, its edges and vertices) to $\widetilde{K}$ which sends v to g and sends edges to edges in such a way as to match up edge orientations and labels. The point is that the image of any edge–path in Δ emanating from v is determined by the matching of the edge orientations and labels. We might worry that this leads to inconsistencies, but this concern would be unfounded. If two edge–paths p_1 and p_2 in Δ emanate from v and have a common final edge (traversed in the same or in opposite directions), then they enclose a subcomplex of Δ which is itself a van Kampen diagram; so the word w' read around its boundary represents the identity, and therefore w' determines an edge–loop in $\widetilde{K}$ starting from g; therefore the images of the final edges of p_1 and p_2 must agree. Finally, we can extend this map to the whole of Δ by sending faces in Δ to faces in $\widetilde{K}$ — we can do so because the words read around the edge–paths around the faces in Δ are defining relators, and so the images of those edge–paths encircle faces $\widetilde{K}$. Note that the boundary circuit of Δ is carried to an edge–loop in $\widetilde{K}$ around which we read w and that this edge–loop would be ρ if we chose v and g suitably.

As a corollary, we learn that $\widetilde{K}$ is simply connected, as promised in Section 3.2.

In this combinatorial setting, the appropriate notion of area for Δ is the easiest one imaginable: the number of faces it has. Our discussion has established —

Proposition 3.2. *The Dehn function of a finite presentation P, as defined in terms of puzzles, is a* minimal isoperimetric function for $\widetilde{K}$ *in that it maps $n \mapsto N$, where N is the minimal number such that any edge–loop in $\widetilde{K}$ of length at most n can be spanned by a combinatorial disc* (*that is, a van Kampen diagram*) *with at most N faces.*

In light of this proposition, denote the Dehn function of P by $\mathrm{Area}_P : \mathbb{N} \to \mathbb{N}$.

3.4. **Isoperimetry in Riemannian manifolds.** It has long been known that any loop of length ℓ in the plane can be filled with a disc of area at most $\ell^2/(4\pi)$. This bound is realized by a circle of perimeter ℓ. It is no coincidence that the Dehn function of $\mathbb{Z} \times \mathbb{Z}$ also grows quadratically (Proposition 2.11). After all, the plane bears as coarse resemblance to $\mathbb{Z} \times \mathbb{Z}$. A similar connection can be drawn between the Dehn function of *every* finitely presented group and the isoperimetry of the associated Riemannian manifold $\widetilde{M}$ (of Section 3.2).

By a *disc* in $\widetilde{M}$ we mean the image of a continuous map $D^2 \to \widetilde{M}$. Use any reasonable notion of area for discs in $\widetilde{M}$. (*Pulling back the Lebesgue measure* from $\widetilde{M}$ to D^2 is one way to go.) The *minimal isoperimetric function* $\mathrm{Area}_{\widetilde{M}} : [0,\infty) \to [0,\infty)$ for $\widetilde{M}$ is defined so that $\mathrm{Area}_{\widetilde{M}}(\ell)$ is the infimal real number such that every loop of length at most ℓ in $\widetilde{M}$ can be spanned by a disc of area at most $\mathrm{Area}_{\widetilde{M}}(\ell)$. (In fact, this infimum is a minimum... there is a long story here known as *Plateau's Problem*.)

The following theorem is generally attributed to Gromov, who is responsible for richly animating the study of finitely generated groups by drawing on analogies and connections with Riemannian geometry. Detailed proofs can be found in [10] and [14]. The fact that the fundamental group of a compact Riemannian manifold is always finitely presentable will be implicit in this theorem; we will not prove this here but the ideas involved are similar to those that establish Theorem 4.3; see Chapter I.8 of [11] for details.

The word *equivalent* in the theorem refers to the relation $\simeq$ which is commonly used in geometric group theory to capture the notion of functions growing at the same rate. For $f, g : [0,\infty) \to [0,\infty)$ we write $f \preceq g$ when there exists $C > 0$ such that $f(\ell) \leq Cg(C\ell + C) + C\ell + C$ for all $\ell \geq 0$. And $f \simeq g$ when $f \preceq g$ and $g \preceq f$. This relation can be expanded to encompass functions with domain $\mathbb{N}$ by extending their domain to $[0, \infty)$ so that these functions become constant on the intervals $[n, n+1)$ for all $n \in \mathbb{N}$.

Exercise 3.3. Show that all functions which grow at most linearly fast are equivalent. Show that $n \mapsto n^\alpha$ and $n \mapsto n^\beta$, where $\alpha, \beta \geq 1$, are equivalent if and only if $\alpha = \beta$. Show that polynomially growing functions are not equivalent to exponentially growing functions $n \mapsto c^n$ (where $c > 1$). Show that $n \mapsto c^n$ and $n \mapsto d^n$ are equivalent for all $c, d > 1$.

Theorem 3.4 (The Filling Theorem)**.** *Suppose M is a compact Riemannian manifold without boundary and that P is a finite presentation for its fundamental group Γ. The minimal isoperimetric function* $\mathrm{Area}_{\widetilde{M}}(\ell)$ *for $\widetilde{M}$ is* equivalent *to the Dehn function* $\mathrm{Area}_P(n)$ *of* Γ.

For an idea of how to prove the Filling Theorem let us look again at the example of $\mathbb{Z} \times \mathbb{Z}$, presented by $\langle a, b \mid a^{-1}b^{-1}ab \rangle$, with $\widetilde{M}$ being the plane. The *infinite chessboard* complex $\widetilde{K}$ is itself a plane and so maps to $\widetilde{M}$ in the natural way. So, in this instance, the theorem is about comparing filling general loops in the plane with discs to filling edge–loops in the chessboard pattern with chessboard squares. The key points are firsty that an arbitrary loop in the plane can be pushed into the 1–skeleton without increasing its length too much, secondly that the number of squares it then encloses is comparable to the area it originally enclosed, and thirdly that an edge–loop enclosing squares is the same thing as a van Kampen diagram.

This approach works in full generality: a disc spanning a loop in $\widetilde{M}$ is similar to a van Kampen diagram filling an edge–loop in $\widetilde{K}$. To make sense of this we have to relate $\widetilde{K}$ and $\widetilde{M}$. We map $\widetilde{K}$ to $\widetilde{M}$ beginning with its vertices, for which we use $\Phi : \Gamma \to \widetilde{M}$ from Section 3.2, and then we extend to the 1–skeleton by mapping the edge between a pair of vertices in $\widetilde{K}$ to a geodesic (that is, a minimal length path) between their images, and then we extend to the whole of $\widetilde{K}$ by mapping the interiors of faces in $\widetilde{K}$ to minimal area discs spanning the loops that are the images of their boundaries. The resulting map $\widetilde{K} \to \widetilde{M}$ can be used to carry edge–loops and van Kampen diagrams into $\widetilde{M}$, whilst retaining control on their lengths and areas. Moreover, arbitrary loops or discs in $\widetilde{M}$ can be *pushed* to edge–loops or combinatorial discs in the image of the $\widetilde{K}$ whilst maintaining similar control.

4. A large–scale geometric invariant

4.1. **Quasi–isometries.** The chapter by Thomas is devoted to quasi–isometries, so here we will be brief. Quasi–isometries are maps which carry one metric space almost onto another with a bounded amount of stretching and tearing.

To be precise, a map $\Phi : X \to Y$ between metric spaces is a (λ, μ)–*quasi–isometry*, where $\lambda \geq 1$ and $\mu \geq 0$, when

$$\frac{1}{\lambda} d(x,y) - \mu \;\leq\; d(\Phi(x), \Phi(y)) \;\leq\; \lambda d(x,y) + \mu$$

for all $x, y \in X$, and every point of Y is within a distance μ of the image of Φ. Two metric spaces are quasi–isomorphic when there is a quasi–isometry between them.

In an exercise Thomas guides you through why quasi–isometry is an equivalence relation on any given set of metric spaces. And she proves that, whist a finitely generated group can have many different finite generating sets and so many different word metrics, all are equivalent in that the identity map is a quasi–isometry from the group with one word metric to the same group with another word metric.

Exercise 4.1. Show that the map $\Phi : \Gamma \to \widetilde{M}$ given at the end of Section 3.2 and the inclusion map $\Gamma \to \widetilde{K}$ identifying Γ with the set of vertices of $\widetilde{K}$ are both quasi–isometries.

4.2. **The Dehn function is a quasi–isometry invariant.** We defined a Dehn function in terms of a *finite presentation for* a group, rather than simply in terms of a group. But if a group has a finite presentation, then it has many finite presentations. So the nagging question is how the Dehn function depends on the presentation. The relation $\simeq$ defined in Section 3.4 allows us a satisfying answer.

Proposition 4.2. *The Dehn functions of any two finite presentations for the same group are equivalent in the sense of* $\simeq$.

This proposition highlights the need for the "$+Cn$" term in the definition of $\simeq$. Consider the presentations $\langle a \mid \, \rangle$ and $\langle a, b \mid b \rangle$ of $\mathbb{Z}$. The Dehn function of $\langle a \mid \, \rangle$ is constantly zero since there are no defining relations. But the Dehn function of $\langle a, b \mid b \rangle$ is $n \mapsto n$ since it takes n applications of the defining relation to reduce b^n to the empty word.

In fact, the Dehn function is an invariant is a broader sense.

Theorem 4.3. *If finitely generated groups* Γ_1 *and* Γ_2 *are quasi–isometric with respect to some (and so any) word metrics, and* Γ_1 *is finitely presented, then* Γ_2 *is also finitely presentable and their Dehn functions, defined with respect to any finite presentations, are equivalent in the sense of* $\simeq$.

This theorem implies Proposition 4.2 (see Section 4.1). It is proved by a follow–your–nose type of argument, remembering that, as quasi–isometries need not be continuous, it is better to examine their affect on configurations of points than on paths or 2–cells. Here is a sketch. Consider a loop ρ in the Cayley graph of Γ_2. Use a quasi–isometry $\Phi : \Gamma_2 \to \Gamma_1$ to carry the vertices on this loop into the Cayley 2–complex for Γ_1. Join up the points there by geodesics in the order they appeared on ρ, to make a loop. Fill that loop with a minimal area van Kampen diagram Δ, and use a *quasi–inverse* $\Gamma_1 \to \Gamma_2$ to Φ to carry the vertices of Δ back to Γ_2, which gives a configuration of points which *coarsely fill* ρ. Analyzing the distances between the vertices in this configuration by comparing them to the distances between the corresponding points in Δ, leads to the result that Γ_2 is finitely presentable. Moreover, the configuration can be fleshed out to give a genuine van Kampen diagram modeled on Δ filling ρ with respect to any finite presentation for Γ_2, and that shows the Dehn functions are equivalent. See [1] or [11] for details.

Theorem 4.3 tells us that the Dehn function is an invariant of *large–scale* or *coarse* geometry and so is a tool in an influential programme of Gromov to understand discrete groups (such as groups with word metrics) up to quasi–isometry. In fact, the Dehn function is just the first of a number of *filling functions* discussed by Gromov in [27] along with a variety of quasi–isometry invariants. Filling functions record geometric features of discs or other surfaces spanning loops in a space such as area, diameter, radius, the lengths loops grow to in the course of null–homotopies, and so on. They also have higher dimensional analogues, concerning $(n+1)$–dimensional balls spanning n–dimensional spheres rather than just discs (2–dimensional balls) spanning loops (1–dimensional spheres). They can also be recast *homologically* in terms of *cycles* bounding *chains*.

5. The Dehn function landscape

Hyperbolic groups, the subject of Duchin's chapter in this volume, are those finitely presented groups which are *negatively curved*. They can be characterized as the finitely presented groups whose Dehn functions grow $\simeq n$. They stand isolated in that if a finitely presentable group has Dehn function not bounded below by a quadratic function, then that group is hyperbolic [6, 11, 26, 35, 37].

Finitely generated abelian groups with $\mathbb{Z} \times \mathbb{Z}$ subgroups (i.e. those that are not hyperbolic) have Dehn functions $\simeq n^2$. More generally, *semi–hyperbolic* groups, CAT(0) groups (the subject of Ruane's chapter) and *automatic* groups have Dehn function $\preceq n^2$. These groups all display features of non–positive curvature, and we might suspect that having Dehn function $\preceq n^2$ might be a reasonable characterization of non–positive curvature amongst finitely presentable groups. However the class of groups with Dehn function $\simeq n^2$ is broad. It contains $SL_n(\mathbb{Z})$ for $n \geq 5$ (and conjecturally $SL_4(\mathbb{Z})$) [46], Thompson's group F [28], Stallings' group [19], and examples of nilpotent groups of all nilpotency classes [45], and none of these could reasonably be called non–positively curved.

In notes for a summer school in 1996 [24], Gersten wrote the following, which has since been proved prescient as most of the examples just listed were established later —

> *I call this a zoo, because I am unable to see any pattern in this bestiary of groups. It would be striking if there existed a reasonable characterization of groups with quadratic Dehn functions, which was more enlightening than saying that they have quadratic Dehn functions.*

Looking beyond quadratic, we mention next that there are examples of finitely presented groups with Dehn function $\simeq n^{\alpha}$. For example, the two–generator groups which are *free nilpotent of class c* — that is, have only the relations necessary to make them nilpotent of class c — have Dehn functions $\simeq n^{c+1}$ [3, 27, 38].

More generally, there is much known about for which $\alpha \geq 1$, there are finitely presented groups with Dehn function $\simeq n^{\alpha}$. The set of such α is countable as there are only countably many finite presentations, and, as we have indicated, it has no values in the open interval $(1, 2)$. But it is dense in the interval $[2, \infty)$ — see [7, 9]; and for the interval $[4, \infty)$, remarkably detailed information is provided in [43]: conditions on the α in terms of whether there is a Turing Machine capable of writing out the first n digits of the decimal expansion of α within a certain amount of time. Additionally, a wide variety of other functions $f : \mathbb{N} \to \mathbb{N}$ which grow $\geq n^4$ are shown in [43] to be equivalent to Dehn functions — indeed, just about all common such functions that have the *super–additivity* property, $f(n + m) \geq f(n) + f(m)$ for all $n, m \in \mathbb{N}$.

We now turn to the extremes of the Dehn function landscape.

6. Fast growing Dehn functions

6.1. **A group with exponential Dehn function.** Establishing how a Dehn function grows presents two difficulties. The upper bound requires consideration of *all* words that represent the identity. And, for the lower bound, whilst it suffices to consider only a suitable family of words w_n whose lengths grow $\simeq n$, we must argue that *all* van Kampen diagrams for those w_n have at least some given area. The situation is analogous to the struggle to establish the (worst–case) time–complexity of some computational problem: for an upper bound one needs to show there is an algorithm which solves the problem within some given time on *all* inputs, and for the lower bound one has to show that on some "worst" family of inputs, *every* algorithm that solves the problem takes at least some given amount of time.

In this section we will show —

Theorem 6.1. *The Dehn function $f(n)$ of $\langle a, s \mid s^{-1}as = a^2 \rangle$ satisfies $f(n) \simeq 2^n$.*

This group $\langle a, s \mid s^{-1}as = a^2 \rangle$ is often known as $\mathrm{BS}(1, 2)$ and is one of a family of groups discussed in detail in Freden's chapter. *Subgroup distortion* is one of its pertinent geometric features and of those groups to come in Sections 6.2 and 6.3. In general, *subgroup distortion* concerns how a finitely generated subgroup H sits inside an ambient finitely generated group G. There are two natural word metrics on such an H: the intrinsic metric d_H coming from its own generating set and the extrinsic metric d_G coming from the generating set of G. Given $n \in \mathbb{N}$, the *distortion function* supplies the maximal distance $d_H(1, h)$ from the identity of all elements $h \in H$ such that $d_G(1, h) \leq n$. Roughly speaking, this function grows quickly when H sits severely scrunched up on itself inside G. It is, in a sense, a lower dimensional version of Dehn function in that it concerns filling 0–spheres (that is, pairs of points) with 1–discs (that is, paths).

Exercise 6.2. Up to $\simeq$, the growth of the distortion function of a finitely generated subgroup of a finitely generated group does not depend on the finite generating sets.

Proposition 6.3. *The distortion function for the subgroup $\mathbb{Z} = \langle a \rangle$ of $\langle a, s \mid s^{-1}as = a^2 \rangle$ grows exponentially.*

That this distortion function grows $\geq 2^n$ is a consequence of the doubling effect that s has when it conjugates a, which leads to the relations $s^{-n}as^n = a^{2^n}$. We will give some explanation for the upper bound at the end of this section.

The large distortion translates into large Dehn function because a copy of a van Kampen diagram for $s^{-n}as^na^{-2^n}$, displaying the repeated–doubling, can be joined to its mirror image along the side labelled a^{2^n}, off–set by one, to give a van Kampen diagram for

$$w_n \;:=\; as^{-n}as^na^{-1}s^{-n}a^{-1}s^n$$

as illustrated below in the case $n = 5$. (This van Kampen diagram and those to come in Section 6.3 are drawn more economically as 2–complexes rather than as the puzzles of Section 1.) Note that we had to join two mirror–image copies because one on its own would have both area and boundary length $\simeq 2^n$ on account of the exponentially large power of a.

This family of van Kampen diagrams has area $\simeq 2^n$. Here are two strategies for proving the lower bound. (The same strategies can be employed to give the lower bound for Proposition 2.11.) The first uses some concepts from algebraic topology; the second is more elementary.

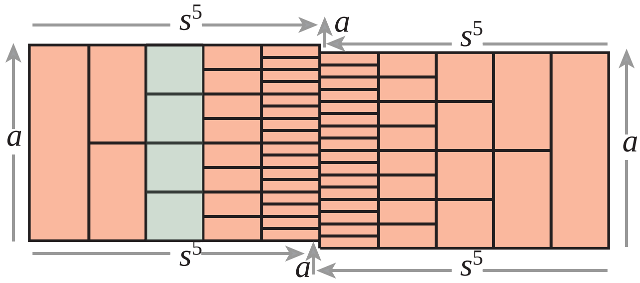

FIGURE 7. A van Kampen diagram for $as^{-5}as^5a^{-1}s^{-5}a^{-1}s^5$ with respect to $\langle a, s \mid s^{-1}as = a^2 \rangle$. All vertical edges are labelled a and directed upwards. All horizontal edges are labelled s; those on the left half of the diagram are oriented to the right and those on the right half are oriented to the left. An example of a *corridor* is shown in green.

The first strategy is to use *Gersten's Lemma*: argue that the *Cayley 2–complex* is contractible and therefore that amongst all van Kampen diagrams for a given word, a diagram Δ that is embedded in the sense of Section 3.1 (or even just embedded on the complement of its 1–skeleton) is of minimal area. The point is that any other van Kampen diagram Δ' for the same word could combine with Δ to make a 2–cycle in the Cayley 2–complex, which is contractible and so all the faces of Δ must cancel with faces in Δ'.

Exercise 6.4. Show that the Cayley 2–complex of $\langle a, s \mid s^{-1}as = a^2 \rangle$ is homeomorphic to the direct product of an infinite trivalent (that is, three edges meet at each vertex) tree with a line, and so is contractible. Show that the maps (in the sense of Section 3.1) from the diagrams of Figure 7 to the Cayley 2–complex are embeddings.

The second strategy uses *corridors* (or *bands*) to understand van Kampen diagrams. Such is the sole defining relation $s^{-1}as = a^2$, adjoining each s in the boundary of a van Kampen diagram, there must be a face which has an edge labelled by s on its far side; and that edge must adjoin another face with same property; and so on. So there is a chain of faces, joined one–to–the–next by edges labelled s, proceeding through the diagram forming a corridor with the edges along its sides all labelled by $a^{\pm 1}$. This corridor must terminate at some other edge labelled s elsewhere on the boundary of the diagram. Corridors therefore pair–off edges labelled by s in the boundary. The four green faces in Figure 7 comprise an example of a corridor.

Corridors cannot cross themselves or each other. So corridors must *stack up* in any van Kampen diagram for w_n and it can be deduced that the exponent sums of the words on $a^{\pm 1}$ along their sides grow exponentially through this stack. In the example of Figure 7, they run vertically through the diagram, and grow exponentially in length towards the centre of the diagram. So one corridor must include $\geq 2^n$ faces. (A complication in the way of making this argument precise is that corridors that form rings instead of emerging on the boundary are possible. But such *annuli* are, in a sense redundant, and do not appear in diagrams of minimal area.)

To complete a proof of Theorem 6.1, we need the exponential upper bound. Corridors are useful here also. Suppose we have a word w representing the identity, and Δ is a van Kampen diagram for w. If w contains any letters $s^{\pm 1}$, then since all such letters are paired off by corridors in Δ and no two corridors cross, there must be a subword $s^{\pm 1} u s^{\mp 1}$ of w such that u is a word on $\left\{a, a^{-1}\right\}$ and the $s^{\pm 1}$ in this subword is joined to the $s^{\mp 1}$ by a corridor. It follows that there is a null-sequence for w which begins by replacing $s^{\pm 1} u s^{\mp 1}$ with a word a^k where $|k|$ is either half or twice the sum of the exponent of the letters in u, and uses at most $|k|$ applications–of–a–relator moves in doing so. (We have just proved and employed *Britton's Lemma*.) Repeat until all $s^{\pm 1}$ have been paired off and removed, and then freely reduce the resulting word on $\left\{a, a^{-1}\right\}$ to the empty word. Summing the applications–of–a–relator moves we use along the way gives our exponential upper bound. The same approach leads to a proof of the exponential upper bound on the distortion function for $\langle a \rangle$ in $\langle a, s \mid s^{-1} a s = a^2 \rangle$.

6.2. **Iterated distortion and Baumslag's one–relator group.** Rename the example from Section 6.1 as $\langle a, s_1 \mid {s_1}^{-1} a s_1 = a^2 \rangle$ and consider embellishing it by distorting $\langle s_1 \rangle$ by introducing a new letter s_2 acting on $\langle s_1 \rangle$ via ${s_2}^{-1} s_1 s_2 = {s_1}^2$. The distortion function for $\mathbb{Z} = \langle a \rangle$ inside the resulting group

$$\langle a, s_1, s_2 \mid {s_1}^{-1} a s_1 = a^2,\ {s_2}^{-1} s_1 s_2 = {s_1}^2 \rangle$$

grows $\succeq \exp^{(2)}$, since

$$({s_2}^{-n} s_1 {s_2}^{n})^{-1} a ({s_2}^{-n} s_1 {s_2}^{n}) \ = \ {s_1}^{-2^n} a {s_1}^{2^n} \ = \ a^{2^{2^n}}.$$

(We write $\exp^{(l)}$ for the l–fold iterate of the exponential function.)

Iterating, we find the distortion function for $\mathbb{Z} = \langle a \rangle$ inside

$$\langle a, s_1, \cdots, s_l \mid {s_1}^{-1} a s_1 = a^2,\ {s_{i+1}}^{-1} s_i s_{i+1} = {s_i}^2 \ (i > 1) \rangle$$

grows $\succeq \exp^{(l)}$. (In fact it grows $\simeq \exp^{(l)}$. We will not explain how the upper bound on distortion is proved, suffice to say the ideas in Section 6.1 can be employed.)

A schematic of a van Kampen diagram illustrating the calculation that leads to the 3–fold iterated exponential distortion is shown in Figure 8.

Exercise 6.5. Draw van Kampen diagrams for

$$({s_2}^{-1} s_1 s_2)^{-1} a^{-1} ({s_2}^{-1} s_1 s_2) a^{-1} ({s_2}^{-1} s_1 s_2)^{-1} a ({s_2}^{-1} s_1 s_2) a$$

and

$$({s_2}^{-2} s_1 {s_2}^2)^{-1} a^{-1} ({s_2}^{-2} s_1 {s_2}^2) a^{-1} ({s_2}^{-2} s_1 {s_2}^2)^{-1} a ({s_2}^{-2} s_1 {s_2}^2) a$$

in $\langle a, s_1, s_2 \mid {s_1}^{-1} a s_1 = a^2,\ {s_2}^{-1} s_1 s_2 = {s_1}^2 \rangle$.

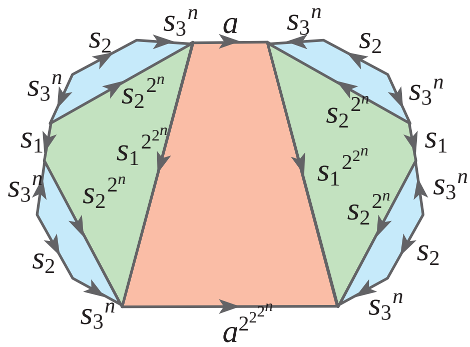

FIGURE 8. Three–fold iterated exponential distortion in $\langle a, s_1, s_2, s_3 \mid {s_1}^{-1} a s_1 = a^2,\ {s_2}^{-1} s_1 s_2 = {s_1}^2,\ {s_3}^{-1} s_2 s_3 = {s_2}^2 \rangle$.

Like the example in Section 6.1, a copy of such a diagram can be glued to its mirror image along the side labelled by the huge power of a and offset by one, and the result is a family of diagrams with perimeter $\simeq n$ and area $\simeq \exp^{(l)}(n)$. This is the beginnings of a proof (along the same lines as that in Section 6.1) that —

Theorem 6.6. *The Dehn function of*

$$\langle a, s_1, \cdots, s_l \mid {s_1}^{-1} a s_1 = a^2,\ {s_{i+1}}^{-1} s_i s_{i+1} = {s_i}^2\ (i > 1) \rangle$$

grows $\simeq \exp^{(l)}$.

This family of groups has a limit (loosely speaking) — a one–relator group

$$\langle a, t \mid (t^{-1}at)^{-1} a (t^{-1}at) \ = \ a^2 \rangle$$

due to Baumslag [2]. Introducing s as shorthand for $t^{-1}at$, we can re–express this presentation as:

$$\langle a, s, t \mid s^{-1} a s = a^2,\ s = t^{-1}at \rangle,$$

and we see that conjugation by s again has a doubling effect on a, and t conjugates a to s. This leads to a feedback effect whereby we get huge distortion of $\mathbb{Z} = \langle a \rangle$ on account of diagrams of the form shown schematically in Figure 9. If the portion of the perimeter of this diagram that excludes the huge power of a is to have length n, then the tree dual to the picture must have depth $\simeq \lfloor \log_2 n \rfloor$ and this suggests the distortion grows like $n \mapsto \exp^{(\lfloor \log_2 n \rfloor)}(1)$, as is indeed the case. This is proved by Platonov in [39], building on work of Gersten [22] and Bernasconi [4], en route to —

Theorem 6.7. *The Dehn function of*

$$\langle a, t \mid (t^{-1}at)^{-1} a (t^{-1}at) \ = \ a^2 \rangle$$

is equivalent to the function $n \mapsto \exp^{(\lfloor \log_2 n \rfloor)}(1)$.

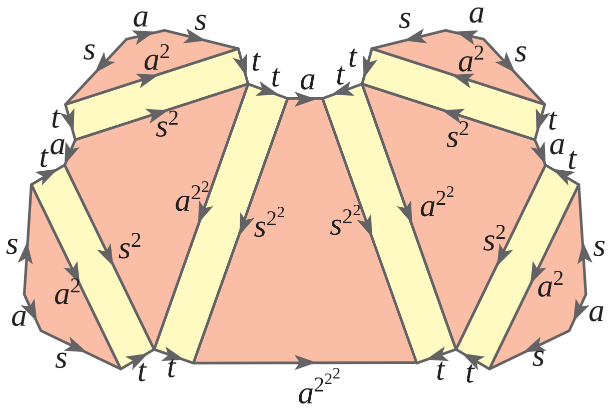

FIGURE 9. Distortion in Baumslag's group $\langle a, s, t \mid s^{-1}as = a^2,\ s = t^{-1}at\rangle$.

Again diagrams with huge area can be obtained by joining two copies of the distortion diagrams along the large power of a, offset by 1.

Exercise 6.8. Draw detailed van Kampen diagrans in the manner of Figure 9 with the power of a along the horizontal path at the bottom being 2, 2^2, 2^{2^2}, etc.

Exercise 6.9. What is the length of a shortest word equalling a^{100} in $\langle a, s, t \mid s^{-1}as = a^2,\ s = t^{-1}at\rangle$? What about a^{1000}?

6.3. **Hydra groups.** Hydra groups were devised by Will Dison and the author. Drawing inspiration from the legend of Hercules' fight with the Lernaean hydra, we defined a *hydra* to be a *positive* word (that means no inverse letters are allowed) on the infinite alphabet $a_1, a_2, \ldots$. Hercules fights this hydra by striking off its first letter. It then regenerates according to the rule that each remaining a_i, where $i > 1$, becomes $a_i a_{i-1}$ (and each remaining a_1 remains as it is). This process — removal of the first letter and then growth — repeats, with Hercules victorious when (not *if*!) the hydra is reduced to the empty word.

Here is an example in which Hercules defeats $a_2a_3a_1$ in five strikes:

$$a_2a_3a_1 \to a_3a_2a_1 \to a_2a_1a_1 \to a_1a_1 \to a_1 \to \textit{empty word}.$$

Exercise 6.10. Prove that Hercules always wins.

Strikingly, battles are of enormous duration, even against simple short hydra. Define $\mathcal{H}(w)$ to be the number of strikes it takes Hercules to defeat the hydra w, and for integers $k \geq 1$, $n \geq 0$, define $\mathcal{H}_k(n) := \mathcal{H}(a_k{}^n)$.

Exercise 6.11. Show that $\mathcal{H}_1(n) = n$ and $\mathcal{H}_2(n) = 2^n - 1$.

Exercise 6.12. Give a formula for $\mathcal{H}_3(n+1)$ in terms of $\mathcal{H}_3(n)$.

Exercise 6.13. For what values of n can you calculate $\mathcal{H}_4(n)$?

Dison and the author showed [20] that these functions $\mathcal{H}_k$ are a variation on Ackermann's famous fast–growing functions $A_k : \mathbb{N} \to \mathbb{N}$ which are defined for integers $k, n \geq 0$ by:

$$\begin{aligned} A_1(n) &= 2n \\ A_{k+1}(n) &= A_k^{(n)}(1). \end{aligned}$$

So, in particular, $A_1(n) = 2n$, $A_2(n) = 2^n$ and $A_3(n) = \exp_2^{(n)}(1)$, where $\exp_2^{(n)}$ denotes the n–fold iterate of $n \mapsto 2^n$. For $k \geq 1$, it turns out that $\mathcal{H}_k \simeq A_k$.

The source of the extreme fast growth of Ackermann's function is the recursion inherent in their definition. Such recursion is apparent in the battle with the hydra in that if ua_k is a hydra that happens to end in the letter a_k, then in the time $\mathcal{H}(u)$ it takes to kill u, there appear $\mathcal{H}(u)$ letters a_{k-1} (and many other letters besides) after the final a_k which then have to be disposed of. So the time it takes to complete that initial task determines the size of the remaining challenge.

Define G_k to be the group presented by

$$\langle\, a_1, \ldots, a_k, t \mid t^{-1}a_1t = a_1,\ t^{-1}a_it = a_ia_{i-1}\ (\forall i > 1)\,\rangle$$

and H_k to be the subgroup $\langle a_1t, \ldots, a_kt\rangle$. The regeneration rules for the hydra are apparent in the defining relations for this presentation. These G_k are well behaved and straight–forward in a number of respects [20] — they are *free–by–cyclic*, CAT(0), *biautomatic*, and they can be presented with only one defining relation. And H_k is a rank–k free subgroup. None–the–less:

Theorem 6.14 ([20])**.** *The distortion function of H_k in G_k grows $\simeq A_k$.*

For an idea of why the distortion function grows so fast, consider this question: for what r (if any) is ${a_k}^n t^r \in H_k$? This is where the battle with the hydra comes in. To see how, look at the case where $k = 2$ and $n = 4$, for example. One can try to convert ${a_2}^4$ to a word on a_1t and a_2t times a power of t by introducing a tt^{-1} to pair the first letter with t, and then carrying the accompanying t^{-1} to the end of the word by conjugating through the intervening letters; then pair off the next a_i likewise, and repeat:

$$\begin{aligned} {a_2}^4 &= a_2t\ t^{-1}{a_2}^3t\ t^{-1} \\ &= a_2t\ a_2a_1a_2a_1a_2a_1\ t^{-1} \\ &= a_2t\ a_2t\ t^{-1}a_1a_2a_1a_2a_1t\ t^{-2} \\ &= a_2t\ a_2t\ a_1a_2a_1a_1a_2a_1a_1\ t^{-2} \\ &\ \ \vdots \end{aligned}$$

The hydra battle

$$a_2a_2a_2a_2 \ \to\ a_2a_1a_2a_1a_2a_1 \ \to\ a_1a_2a_1a_1a_2a_1a_1 \ \to \cdots$$

plays out within this calculation. Pairing off the first letter with a t corresponds to the removal of the first letter of a hydra and conjugating a t^{-1}

through to the right–hand end corresponds to regeneration. So, as Hercules wins after $\mathcal{H}({a_2}^4) = 15$ steps, it eventually arrives at a word on $a_1 t$ and $a_2 t$ times t^{-15}. This calculation, run to its conclusion, gives rise to the van Kampen diagram in Figure 10. There is nothing special about the example $k = 2$ and $n = 4$ here. So, as we know Hercules triumphs in $\mathcal{H}({a_k}^n) = \mathcal{H}_k(n)$ steps, we have the answer $r = \mathcal{H}_k(n)$ to our question.

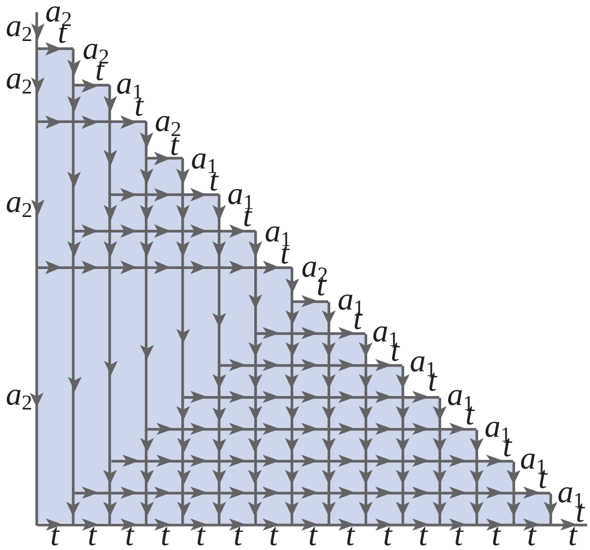

FIGURE 10. A van Kampen diagram showing that ${a_2}^4 t^{15}$ equals a word on $a_1 t$ and $a_2 t$ in $\langle a_1, a_2, t \mid t^{-1} a_2 t = a_2 a_1, t^{-1} a_1 t = a_1 \rangle$. Labels on the interior edges are not shown; those on the horizontal edges are all t, and those on the vertical edges area a_1 and a_2 — which are which should be apparent from the defining relations.

The diagram in Figure 10 can be paired with its mirror image, with three *corridors* of 2–cells arranged between them to give a van Kampen diagram that demonstrates the equality in G_2 of ${a_2}^5 t a_1 {a_2}^{-5}$ and a reduced word on $a_1 t$ and $a_2 t$ of length $2\mathcal{H}_2(4) + 3$. (Two copies of this diagram are shown in blue within the van Kampen diagram in Figure 11.) Similar diagrams can be constructed for all battles between Hercules and the hydra ${a_k}^n$, thereby showing that for all n and k, there are words on $\{a_1, \ldots, a_k, t\}^{\pm 1}$ of length $2n+3$ which represent the same elements in G_k as certain reduced words on $\{a_1 t, \ldots, a_k t\}^{\pm 1}$ of length $2\mathcal{H}_2(n) + 3$. Given that $\mathcal{H}_k \simeq A_k$, this establishes the lower bound of Theorem 6.14. See [20] for more details and a proof of the upper bound.

Now, G_k does not have a large Dehn function — it is a CAT(0) group and so has Dehn function $\preceq n^2$. Distortion does not always lead to large Dehn function. (Even groups with heavily distorted $\mathbb{Z}$ subgroups can have small Dehn functions. For example, $SL_5(\mathbb{Z})$ has exponentially distorted $\mathbb{Z}$ subgroups (see, for example, [29]) and Dehn function $\simeq n^2$ [46].)

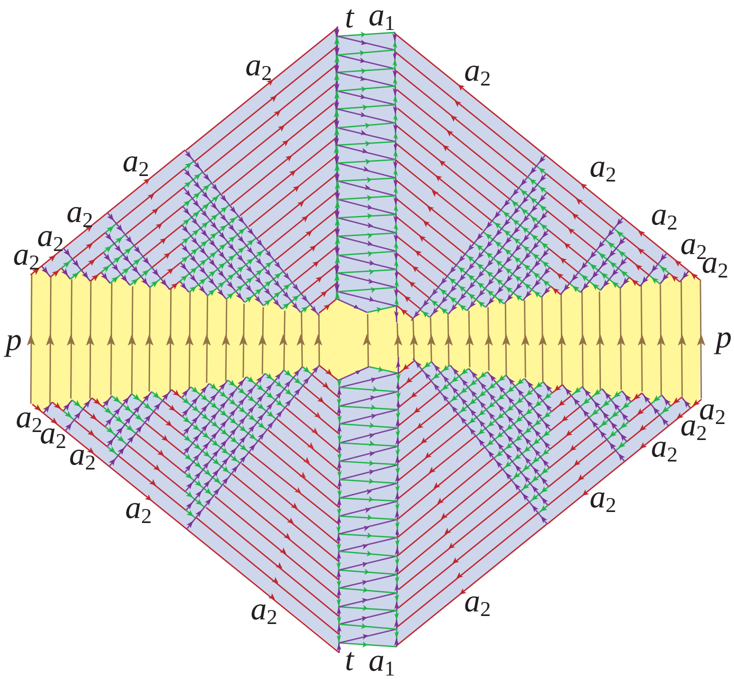

Figure 11. A van Kampen diagram for a word in $\langle G_2, p \mid [H_2, p] \rangle$, illustrating how heavy distortion of H_2 in G_2 leads to a large Dehn function. It is assembled from a pair of mirror–image copies of diagrams that arise due to this distortion; they are separated by a corridor of faces (shown in yellow) which connects the two boundary edges labelled p.

But there are standard methods for translating groups with heavily distorted subgroups to groups with large Dehn function — see Chapter III.Γ in [11] for a general discussion. We define Γ_k to be the HNN–extension of G_k which is presented by $\langle G_k, p \mid [H_k, p] \rangle$ — shorthand for the presentation obtained from our presentation of G_k by adding a new generator p and new defining relations $p\, a_i t = a_i t\, p$ for each i, so that p commutes with all elements of H_k.

Theorem 6.15. *Γ_k presented by $\langle G_k, p \mid [H_k, p] \rangle$ has Dehn function $\simeq A_k$.*

The diagram in Figure 11 indicates how to get the lower bound. *Any* van Kampen diagram with the same boundary must have a corridor of cells connecting one of the edges labelled p to the another, since p only occurs in defining relations of the form $p\, a_i t = a_i t\, p$. (This corridor is shown in yellow in Figure 11.) Along the sides of this corridor we read words on $\{a_1 t, \ldots, a_k t\}^{\pm 1}$ which are necessarily equal to words on $\{a_1, \ldots, a_k, t\}^{\pm 1}$ that we read around part of the boundary of the diagram since killing p maps $\Gamma_k \twoheadrightarrow G_k$. But as the distortion of H_k in G_k is $\succeq A_k$, the length of the corridor (as a function of the length of the boundary circuit of the diagram) must be $\succeq A_k$.

6.4. **Groups with undecidable word problem.** One of the great achievements in twentieth century mathematics was the construction by Boone [5], Britton [12, 13] and Novikov [33] of finitely presented groups for which there can be no algorithm to solve the Word Problem. There is a modern account in [41].

The foundational non–computability result is Turing's proof that there is no algorithm for the *Halting Problem*: that is, no algorithm which will input a program p together with an input i for that program, and will declare whether or not p eventually halts on input i. This can be proved by a Cantor–type diagonal argument; the existence of functions $\mathbb{N} \to \mathbb{N}$ which are not recursive immediately follows. Markoff [31] and Post [40] independently showed that there is no algorithm which solves the Word Problem for finitely presented *semi–groups*. Roughly speaking, the point is that there is a strong analogy between the instructions of a Turing Machine and a finite presentation for a semi–group. Moreover, the play–out of a calculation on the tape of a Turing Machine resembles the manipulation of words in the generators of a semi–group using its defining relations. The idea for groups is the same, but there the result is much harder since the analogy is far more tenuous.

By Proposition 2.17, the Dehn function of a finitely presented group with unsolvable Word Problem is not bounded above by any recursive function. These are therefore examples for which, in this sense, the Dehn function must grow exceptionally quickly. However, there are surprising subtleties here. Recently, Olshanskii [36] constructed an example of a finitely presented group for which there is no algorithm to solve the Word Problem, but on an infinite subset of $\mathbb{N}$ its Dehn function is bounded above by a quadratic function.

7. Further reading

A natural next step after this introduction is Bridson's survey [10], which provides careful proofs of a number of the results discussed here including the Filling Theorem and van Kampen's Lemma, explains other techniques for establishing Dehn functions, and draws a variety of connections with other topics.

My article in [8] on *filling functions* explores the interconnections between and applications of a variety of quasi–isometry invariants, including Dehn functions, that concern the geometry of van Kampen diagrams. The notes by N. Brady and by Short in the same volume respectively address Dehn functions in the context of non–positive curvature and techniques for understanding groups and van Kampen diagrams such as *small–cancellation theory*.

The book [11] by Bridson and Haefliger is a key resource for many topics in geometric group theory including Dehn functions, especially in the context of non–positive curvature.

Gromov instigated and inspired much of the explosion of work over the last thirty years on the geometry of discrete groups. Indeed, a substantial proportion of that research can be viewed as exegesis of his book [27].

Gersten's survey [23] on isoperimetric functions and their analogues, isodiametric functions, which concern the diameters of discs rather than their areas, published in a companion volume of the same conference proceedings as Gromov's book, has also been influential and is where the term *Dehn function* was coined. It remains well worth reading. Also, Gersten's 1996 summer school notes [24] are readily accessible and include discussion of Dehn functions of hyperbolic and automatic groups.

Sapir's [42] is a recent and wide–ranging survey which covers many areas of current research on Dehn function and related topics.

References

[1] J. M. Alonso. Inégalitiés isopérimétriques et quasi–isométries. *C.R. Acad. Sci. Paris Ser. 1 Math.*, 311:761–764, 1990.

[2] G. Baumslag. A non–cyclic one–relator group all of whose finite quotients are cyclic. *J. Austral. Math. Soc.*, 10:497–498, 1969.

[3] G. Baumslag, C. F. Miller, III, and H. Short. Isoperimetric inequalities and the homology of groups. *Invent. Math.*, 113(3):531–560, 1993.

[4] A. A. Bernasconi. *On HNN–extensions and the complexity of the word problem for one-relator groups*. PhD thesis, University of Utah, 1994. http://www.math.utah.edu/~sg/Papers/bernasconi-thesis.pdf.

[5] W. W. Boone. Certain simple unsolvable problems in group theory I, II, III, IV, V, VI. *Nederl. Akad. Wetensch Proc. Ser. A.* **57**, 231–236, 492–497 (1954), **58**, 252–256, 571–577 (1955), **60**, 22-26, 227-232 (1957).

[6] B. H. Bowditch. A short proof that a subquadratic isoperimetric inequality implies a linear one. *Michigan Math. J.*, 42(1):103–107, 1995.

[7] N. Brady, M. R. Bridson, M. Forester, and K. Shankar. Perron-Frobenius eigenvalues, snowflake groups, and isoperimetric spectra. *Geometry and Topology*, 13:141–187, 2009.

[8] N. Brady, T. R. Riley, and H. Short. *The geometry of the word problem for finitely generated groups*. Advanced Courses in Mathematics CRM Barcelona. Birkhäuser–Verlag, 2007.

[9] M. R. Bridson. Fractional isoperimetric inequalities and subgroup distortion. *J. of the A. Math. Soc.*, 12(4):1103–1118, 1999.

[10] M. R. Bridson. The geometry of the word problem. In M. R. Bridson and S. M. Salamon, editors, *Invitations to Geometry and Topology*, pages 33–94. O.U.P., 2002.

[11] M. R. Bridson and A. Haefliger. *Metric Spaces of Non–positive Curvature*. Number 319 in Grundlehren der mathematischen Wissenschaften. Springer Verlag, 1999.

[12] J. L. Britton. The word problem for groups. *Proc. London Math. Soc. (3)*, 8:493–506, 1958.

[13] J. L. Britton. The word problem. *Ann. of Math. (2)*, 77:16–32, 1963.

[14] J. Burillo and J. Taback. Equivalence of geometric and combinatorial Dehn functions. *New York J. Math.*, 8:169–179 (electronic), 2002.

[15] D. E. Cohen. The mathematician who had little wisdom: a story and some mathematics. In *Combinatorial and geometric group theory (Edinburgh, 1993)*, volume 204 of *London Math. Soc. Lecture Note Ser.*, pages 56–62. Cambridge Univ. Press, Cambridge, 1995.

[16] D. E. Cohen, K. Madlener, and F. Otto. Separating the intrinsic complexity and the derivational complexity of the word problem for finitely presented groups. *Math. Logic Quart.*, 39(2):143–157, 1993.

[17] M. Dehn. Über unendliche diskontunuierliche Gruppen. *Math. Ann.*, 71:116–144, 1912.

[18] M. Dehn. *Papers on group theory and topology*. Springer-Verlag, New York, 1987. Translated from the German and with introductions and an appendix by J. Stillwell, With an appendix by O. Schreier.

[19] W. Dison, M. Elder, T.R. Riley, and R. Young. The Dehn function of Stallings' group. *Geom. Funct. Anal.*, 19(2):406–422, 2009.

[20] W. Dison and T. R. Riley. Hydra groups. arXiv:1002.1945.

[21] S. M. Gersten. Isodiametric and isoperimetric inequalities in group extensions. Preprint, University of Utah, 1991.

[22] S. M. Gersten. Dehn functions and l_1-norms of finite presentations. In G. Baumslag and C. Miller, editors, *Algorithms and classification in combinatorial group theory (Berkeley, CA, 1989)*, volume 23 of *Math. Sci. Res. Inst. Publ.*, pages 195–224. Springer-Verlag, 1992.

[23] S. M. Gersten. Isoperimetric and isodiametric functions. In G. Niblo and M. Roller, editors, *Geometric group theory I*, number 181 in LMS lecture notes. Camb. Univ. Press, 1993.

[24] S. M. Gersten. Introduction to hyperbolic and automatic groups. In *Summer School in Group Theory in Banff, 1996*, volume 17 of *CRM Proc. Lecture Notes*, pages 45–70. Amer. Math. Soc., Providence, RI, 1999.

[25] Rostislav I. Grigorchuk and Sergei V. Ivanov. On Dehn functions of infinite presentations of groups. *Geom. Funct. Anal.*, 18(6):1841–1874, 2009.

[26] M. Gromov. Hyperbolic groups. In S. M. Gersten, editor, *Essays in group theory*, volume 8 of *MSRI publications*, pages 75–263. Springer–Verlag, 1987.

[27] M. Gromov. Asymptotic invariants of infinite groups. In G. Niblo and M. Roller, editors, *Geometric group theory II*, number 182 in LMS lecture notes. Camb. Univ. Press, 1993.

[28] V. S. Guba. The Dehn function of Richard Thompson's group F is quadratic. *Invent. Math.*, 163(2):313–342, 2006.

[29] A. Lubotzky, S. Mozes, and M. S. Raghunathan. Cyclic subgroups of exponential growth and metrics on discrete groups. *C.R. Acad. Sci. Paris, Série 1*, 317:723–740, 1993.

[30] K. Madlener and F. Otto. Pseudonatural algorithms for the word problem for finitely presented monoids and groups. *J. Symbolic Comput.*, 1(4):383–418, 1985.

[31] A. Markoff. On the impossibility of certain algorithms in the theory of associative systems. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 55:583–586, 1947.

[32] A. G. Miasnikov, A. Ushakov, and D. W. Won. The word problem in the Baumslag–Gersten group is polynomial time decidable. arXiv:1102.2481.

[33] P. S. Novikov. On the algorithmic unsolvability of the word problem in group theory. *Trudt Mat. Inst. Stkelov*, 44:1–143, 1955.

[34] A. Myasnikov O. Kharlampovich and M. Sapir. Residually finite finitely presented solvable groups. arXiv:1204.6506.

[35] A. Yu. Ol′shanskii. Hyperbolicity of groups with subquadratic isoperimetric inequality. *Internat. J. Algebra Comput.*, 1(3):281–289, 1991.
[36] A. Yu. Ol′shanskii. Groups with undecidable word problem and almost quadratic Dehn function.
arXiv:1104.1476, 2011.
[37] P. Papasoglu. On the sub-quadratic isoperimetric inequality. In *Geometric group theory (Columbus, OH, 1992)*, volume 3 of *Ohio State Univ. Math. Res. Inst. Publ.*, pages 149–157. de Gruyter, Berlin, 1995.
[38] Ch. Pittet. Isoperimetric inequalities for homogeneous nilpotent groups. In R. Charney, M. Davis, and M. Shapiro, editors, *Geometric Group Theory*, volume 3 of *Ohio State University, Mathematical Research Institute Publications*, pages 159–164. de Gruyter, 1995.
[39] A. N. Platonov. An isoperimetric function of the Baumslag–Gersten group. *Vestnik Moskov. Univ. Ser. I Mat. Mekh.*, 3:12–17, 70, 2004. Translation in Moscow Univ. Math. Bull. 59 (2004).
[40] Emil L. Post. Recursive unsolvability of a problem of Thue. *J. Symbolic Logic*, 12:1–11, 1947.
[41] J. J. Rotman. *An Introduction to the Theory of Groups*. Springer Verlag, 4^{th} edition, 1995.
[42] M. Sapir. Asymptotic invariants, complexity of groups and related problems.
arXiv:1012.1325, to appear in the Bulletin of Mathematical Sciences, 2010.
[43] M. V. Sapir, J.-C. Birget, and E. Rips. Isoperimetric and isodiametric functions of groups. *Ann. of Math. (2)*, 156(2):345–466, 2002.
[44] E. R. van Kampen. On some lemmas in the theory of groups. *Amer. J. Math.*, 55:268–273, 1933.
[45] R. Young. Filling inequalities for nilpotent groups. arXiv:math/0608174.
[46] R. Young. The Dehn function of SL(n;Z). arXiv:0912:2697, 2009.

Timothy R. Riley
Department of Mathematics, 310 Malott Hall, Cornell University, Ithaca, NY 14853, USA
tim.riley@math.cornell.edu, http://www.math.cornell.edu/~riley/