Notes on geometric group theory and the geometry of Baumslag-Solitar groups

Ethan Lake

January 3, 2024

These notes were prepared as part of work on a project done together with Shankar Balasubramanian, Alexey Khudorozhkov, and Sarang Golparakrishnan.

1 Some results geometric group theory

In this section we state and prove some useful facts about the large-scale geometry of finitely presentable discrete groups. A good review of background material can be found in [2].

Geometry of finitely presentable discrete groups: preliminaries

Before starting in earnest, a remark on notation and a reminder of the basic definitions. We will frequently be using asymptotic scaling notation, writing $f \sim g$ when two functions of the same variable (usually a system size L) have the same growth.[1]

Consider a group G with the finite presentation

$$G = \langle a_1, \dots, a_n \,|\, R \rangle, \tag{2}$$

where R are a set of relations among the a_i and their inverses. We will let w denote a word in $\{a_i, a_i^{-1}, e\}^*$, with $|w|$ its legnth. We will write w_g or $w \sim g$ to indicate explicitly that w reduces to the group element g, and in other cases will write $g(w)$ to denote the group element that w reduces to. The symbol $\mathcal{K}_g(L)$ will denote the collection of all length-L words that reduce to g,

$$\mathcal{K}_g(L) \equiv \{w \,|\, w \sim g,\, |w| = L\} \tag{3}$$

(we will often write just $\mathcal{K}_g$, with the dependence on L kept implicit). The length $|g|$ of a group element g will be defined as its geodesic distance in the Cayley graph, viz.

$$|g| \equiv \min_{w \in \mathcal{K}_g} |w|. \tag{4}$$

[1]More precisely, we write $f(L) \sim g(L)$ when there exists a constant $C > 0$ such that

$$f(L/C) \leq g(L) \leq f(LC). \tag{1}$$

Thus e.g. $\mathrm{poly}(L)\alpha^L \sim \beta^L$ for all $\alpha, \beta > 1$, etc.

It is also easiliy checked that $d(g,h) \equiv |gh^{-1}|$ is a metric on the Cayley graph.

We will mostly be interested in infinite groups, since finite ones have trivial large-scale geometry (all are quasi-isometric to the trivial group). Finitely presentable infinite groups are of course very easily constructed; indeed any presentation where the number of generators exceeds the number of nontrivial[2] relations will generate an infinite group. Free groups on $n > 1$ generators and Abelian groups in some sense define opposite geometric extremes, and most of the interesting action will happen in cases where an intermediate amount of Abelian-ness is introduced to a free group.

Time complexity: the Dehn function

Any $w \in \mathcal{K}_e$ defines a closed loop in the Cayley graph (really, Cayley 2-complex, with a 2-cell present for each nontrivial relation) of G. We define the *area* $A(w)$ of a word $w \in \mathcal{K}_e$ by the area of the *minimal* surface in the Cayley 2-complex with boundary w:

$$A(w) \equiv \min_{S\,:\,\partial S = w} \text{Area}(S). \tag{5}$$

It is easy to see that $A(w) = 0$ iff $w = ue^k u^{-1}$ for some word u.

$A(w)$ is one of the functions which allows us to connect geometric group theory with complexity theory, as it is approximately equal to the (non-determininstic) time complexity of an algorithm which uses successive application of group relations to solve the *word problem*, viz. the problem of deciding wether or not a given word is in $\mathcal{K}_e$.[3] This correspondence is approximate as it holds only up to factors going as $|w|$, which happens because $A(ww^{-1}) = 0$ despite the fact that ww^{-1} requires the application of $\sim |w|$ trivial relations (viz. $aa^{-1} \leftrightarrow e$ and $ae \leftrightarrow ea$) to be reduced to the identity word.

We can go beyond single words in $\mathcal{K}_e$ by defining the *distance* $d(w_1, w_2)$ between two words w_1, w_2 with $w_1 \sim w_2$ as the minimum number of relations needed to deform w_1 into w_2 (or vice versa). With this definition, $d(w,e) \sim A(w)$ up to factors of $|w|$. The following proposition is easy to verify:

Proposition 1. *The distance $d : \mathcal{K}_g \times \mathcal{K}_g \to \mathbb{N}$ is a metric on each $\mathcal{K}_g$.*

The above discussion has focused on the complexity of reducing words in $\mathcal{K}_e$ to the trivial word. We can get a handle on the complexity of $\mathcal{K}_e$ as a whole by looking at the worst-case complexity over words in $\mathcal{K}_e$. To this end we define the *Dehn function* $D : \mathbb{N} \to \mathbb{N}$ of a group as

$$D(L) \equiv \max_{w \in \mathcal{K}_e(L)} d(w,e) \sim \max_{w \in \mathcal{K}_e(L)} A(w). \tag{6}$$

In practice we will just identify $D(L)$ with the largest value of $A(w)$ for words in $\mathcal{K}_e$; the factors of L we loose from this are inconsequential since $D(L) \sim L$ is the most trivial scaling a Dehn function can have.

One may wonder to what extent $D(L)$ is presentation-dependent. To this end we have the following result:

[2] viz. not including free reduction or expansion.

[3] This kind of proof system, where each step consists of the application of a group relation, is known as the *Dehn proof system*. $A(w)$ gives the *non*-deterministic time complexity since the algorithm runs by blindly applying all possible sequences of group relations in superposition.

Proposition 2. *Let $D_S(L)$ be the Dehn function of a group with a finite presentation $P = \langle S|R \rangle$. Then if P, P' are any two such finite presentations,*

$$D_P(L) \sim D_{P'}(L). \tag{7}$$

The proof of this fact is essentially identical to the proof of presentation-independence of group growth rates, which is given below in proposition 8. This means that the scaling of $D(L)$ is a well-defined property of the group itself, rather than the presentation.

The definition of the Dehn function relates only to the (worst-case) complexity of deforming a given $w \in \mathcal{K}_e$ into the identity word. While in general we can ask about the complexity of deforming any two words in the same $\mathcal{K}_{g \neq e}$ into one another, doing so does not produce anything which is not already captured by $D(L)$:

Proposition 3. *For a given element g of geodesic distance $|g| \leq L$, define the g-sector Dehn function as*

$$D_g(L) \equiv \max_{w_1, w_2 \in \mathcal{K}_g(L)} d(w_1, w_2). \tag{8}$$

Then $D_g(L) \sim D(L)$ for all g.

Proof. For two words $w_{1,2}$ in the same $\mathcal{K}_g$ sector, any deformation (aka *based homotopy*) of w_1 to w_2 gives a deformation between the length-$2L$ word $w_1 w_2^{-1} \in \mathcal{K}_e(2L)$ and e. Thus the minimal number of steps needed to relate w_1 to w_2 cannot be asymptotically smaller than the minimal number of steps needed to deform $w_1 w_2^{-1}$ to e. This implies

$$A(w_1 w_2^{-1}) \lesssim d(w_1, w_2) \tag{9}$$

up to unimportant contributions linear in L.[4] This means that $D(L) \lesssim D_g(L)$. Conversely, since we can deform e to $w_1^{-1} w_2 \sim e$ in time $\sim A(w_1 w_2^{-1})$, w_1 can be deformed into $w_1(w_1^{-1} w_2) = w_2$ in time $\lesssim D(L)$. Thus we also have $d(w_1, w_2) \lesssim A(w_1 w_2^{-1})$, so up to factors of order L, we have

$$A(w_1 w_2^{-1}) \sim d(w_1, w_2) \implies D_g(L) \sim D(L). \tag{10}$$

□

A large class of groups have rather simple Dehn functions:

Example 1. All finite groups have $D(L) \sim CL$ for some (presentation-dependent) constant C related to the diameter of the group's Cayley graph. Since finite groups have trivial geometry (they are quasi-isometric to the trivial group), this provides another sense in which $O(L)$ contributions to Dehn functions are trivial.

Example 2. $D(L) \sim L^2$ for any Abelian group. Here the L^2 comes from applying relations to move around generators past other ones and next to their inverses.

[4] They are "unimportant" because we need $O(L)$ steps just to get rid of trivial words like ww^{-1}. However, an extra L steps can be very important in certain settings, especially where e.g. we do not allow words of length greater than L to appear at any stage of the deformation (which we very well may want to do on physical grounds!). In situations like this, most of statements in this section need to be critically reexamined.

The above examples show that only infinite non-Abelian groups can have interesting Dehn functions. However, being *too* non-Abelian also does not work:

Example 3. Free groups also have $D(L) \sim L$. This follows from the fact that the Cayley graphs of free groups are trees, and thus any word $w \in \mathcal{K}_e$ must represent a path which retraces its steps at every point and forms no loops. Such paths can always be contracted to the trivial path in $O(L)$ steps.

To look for interesting Dehn functions we thus need infinite groups with a moderate amount of Abelianness, which (roughly speaking) possess nontrivial loops at all length scales. First examples of such groups are provided by those where $D(L)$ scales as a nontrivial polynomial. These turn out to be precisely those groups which are virtually nilpotent, as proven in Ref. [6]. The simplest example is:

Example 4. The Heisenberg group

$$H_3 = \langle x, y, z \,|\, [x, y]z^{-1},\ [x, z],\ [y, z] \rangle \tag{11}$$

has $D(L) \sim L^3$ [6].

In the next section we will study in detail $BS(1, 2)$, the simplest group with exponential Dehn function.

Space complexity: the expansion length

The above distance measures relate to the time complexity of the word problem. We may also construct measures related to the space complexity. The non-deterministic space complexity of the word problem is given by the maximal size of words that one must encounter when reducing a $w \in \mathcal{K}_e$ to the identity. To this end we define the *expansion length* $EL(w)$ of words $w \sim e$ as

$$EL(w) \equiv \min_{\{\Delta_{w\to e}\}} \max_t |\Delta_{w\to e}(t)|, \tag{12}$$

where the minimum is over all homotopies $\Delta_{w\to e}(t)$, viz. sequences of loops in the Cayley graph with $\Delta_{w\to e}(0) = w, \Delta_{w\to e}(t_f) = e$ (here t_f is a nonnegative integer), and with $\Delta_{w\to e}(t+1)$ related to $\Delta_{w\to e}(t)$ by the application of a single relation. If $EL(w) > |w|$, then w must expand while being reduced to the identity.

As we did with the area, we may use the expansion length to define a distance between any two words $w_{1,2}$ that represent the same group element. Instead of doing something like $EL(w_1 w_2^{-1})$ though,[5] we define the *relative expansion length* $EL(w_1, w_2)$ between two words $w_1 \sim w_2$ by replacing e by w_2 in the above definition:

$$EL(w_1, w_2) \equiv \min_{\{\Delta_{w_1\to w_2}\}} \max_t |\Delta_{w_1\to w_2}(t)|, \tag{13}$$

where the $\Delta_{w_1\to w_2}(t)$ are all paths in the Cayley graph with endpoints fixed at e and $g(w_1) = g(w_2)$. When the homotopy is trivial, i.e. when $w_1 = w_2$, we define the relative expansion length to vanish.

[5] which we do not do on account of the homotopy relating w_1 to w_2 needing to be properly based.

Proposition 4. *EL is a metric on each $\mathcal{K}_g$.*

Proof. We have $EL(w,w) = 0$ by definition. The symmetry of $EL(w_1,w_2)$ is similarly clear since we can always reverse any homotopy. The triangle inequality follows because $EL(w_1,w_3)+EL(w_3,w_2) \geq \max(EL(w_1,w_3), EL(w_3,w_2))$. The RHS is the maximum length enountered during the particular homotopy $\Delta_{w_1\to w_3}\circ\Delta_{w_3\to w_2}$, where the $\Delta_{w_i\to w_j}$ here are those based homotopies from w_i to w_j with the smallest value of the maximum length. This length is obvioulsy larger than the length one encounters when minimizing over *all* possible based homotopies,$\max(EL(w_1,w_3), EL(w_3,w_2)) \geq EL(w_1,w_2)$, giving the triangle inequality. □

The expansion length of a *group* can be defined as the worst-case spatial complexity of words in $\mathcal{K}_e$, as was done in our definition of the Dehn function:

$$EL(L) \equiv \max_{w\in\mathcal{K}_e(L)} EL(w). \tag{14}$$

Similarly to the Dehn function, the asymptotic scaling of $EL(L)$ is independent of the choice of (finite) presentation. Also as with the Dehn function, considering expansion in other $\mathcal{K}_g$ sectors does not yield anything new. Just as with proposition 3, one can similarly show that

Proposition 5. *For a given g of geodesic distance $|g| \leq L$, define the expansion lengths*

$$EL_g(L) \equiv \max_{w_1,w_2\in\mathcal{K}_g(L)} EL(w_1,w_2). \tag{15}$$

Then

$$EL_g(L) \sim EL(L) \tag{16}$$

for all g.

Most groups one is familiar with have $EL(L) \lesssim L$. For example, it is easy to see that all Abelian groups $\langle a_1,\ldots,a_n\,|\,[a_i,a_j]\rangle$ have $EL(L) = L$. One can also show

Proposition 6. *All finite groups have $EL(L) < L + C$ for some constant C.*

Proof. The proof proceeds according to the same one that would be used in demonstrating the correctness of example 1. Let $w \in \mathcal{K}_e$ be a length-L closed loop in the Cayley graph of a finite group, and let N be the number of vertices in the Cayley graph. Then since N is finite, we may write

$$w = \prod_j w_j, \qquad w_j \sim e,\ |w_j| \leq N, \tag{17}$$

since a path in the Cayley graph can only reach N different vertices before doubling back on itself. Let $M = EL(N)$. If we append M identity characters to the end of w, we can use them to turn any one of the w_j into e without increasing the length of the appended word. Since we can do this for all of the w_j, we thus have $EL(L) = L + N$ as claimed. □

Furthermore, we will see later that despite having exponential time complexity ($D(L) \sim 2^L$), $BS(1,2)$ only has linear spatial complexity ($EL(L) \sim L$). This is part of a more general result that all so-called *asychronously combable groups* have $EL(L) \lesssim L$ [5].

It is however relatively straightforward to construct examples where $EL(L)$ grows faster than linearly. This is done simply by finding groups with Dehn functions that scale as $D(L) = \omega(2^L)$. This is due to the following "spacetime" bound:

Proposition 7. *For a finitely presentable group generated by n_g generators,*

$$D(L) \lesssim (2n_g + 1)^{EL(L)}. \tag{18}$$

Thus $EL(L) \gtrsim \log_{2n_g+1} D(L)$, which grows superlinearly if $D(L)$ grows super-exponentially.

Proof. For $w \in \mathcal{K}_e$ with expansion length $EL(w)$, the number of words that w can possibly visit as it is reduced to the identity is $|\{a_i, a_i^{-1}, e\}|^{EL(w)} = (2n_g + 1)^{EL(w)}$. Since the quickest reduction of w to e cannot visit a given word w' more than once, the number of steps in the reduction is (often very loosely) upper bounded by $(2n_g + 1)^{EL(w)}$. □

Growth rates of groups

A simple measure of a group's geometry we have not yet touched on is how fast the group grows, viz. how the number of elements within a distance L of the origin of the Cayley graph grows as L is increased. To this end, let $\mathcal{B}(L)$ denote the radius-L ball centered at the origin of the Cayley graph, viz.

$$\mathcal{B}(L) \equiv \{g \,|\, |g| \leq L\}, \tag{19}$$

and define the *growth function* (or *group volume*) as $V(L) \equiv |\mathcal{B}(L)|$. Note that $V(L)$ is constrained to be submultiplicative,

$$V(L + M) \leq V(L)V(M), \tag{20}$$

which follows from noting that $g_L g_M \in \mathcal{B}(L \cdot M)$ for all $g_L \in \mathcal{B}(L)$, $g_M \in \mathcal{B}(M)$. Of course with a finite presentation, we could also argue this from the fact that $V(L) \lesssim \exp(L)$; more precisely, we have the trivial upper bound

$$V(L) \leq 2n_g(2n_g - 1)^{L-1} \tag{21}$$

for a presentation with n_g generators.

Since geodesic distances of group elements are presentation-dependent, $V(L)$ will also depend on the presentation. However, asymptotic scaling of $V(L)$ is presentation-*in*dependent, and thus we can meaningfully talk about the growth function of a group, rather than of a presentation:[6]

[6] In math lingo we would say that changing presentations induces a *quasi-isometry* of the Cayley graph.

Proposition 8. *Let V_S denote the growth function for a particular presentation P of G. Then*

$$V_P \sim V_{P'} \tag{22}$$

for all finite presentations P, P' of G.

Proof. Let $P = \langle a_1, \ldots, a_{|S|} \,|\, R\rangle$ and $P' = \langle a'_1, \ldots, a'_{|S'|} \,|\, R'\rangle$. Then since both P, P' present G, each a'_i can be expressed as a product of a finite number of a_i. Let $n_{PP'}$ denote the maximal number of generators of P that appear when writing the a'_i in terms of these generators. Let also $|g|_P$ denote the geodesic distance of $g \in G$ with respect to the presentation P. Then $|g|_{P'} \leq n_{PP'}|g|_P$. Thus

$$V_{S'}(L) \leq V_S(n_{PP'}L). \tag{23}$$

By symmetry, we may also perform a similar rewriting of the generators of P. Thus there exist $O(1)$ constants $n_{PP'}, n_{P'P}$ such that

$$V_{P'}(L/n_{P'P}) \leq V_P(L) \leq V_{P'}(n_{PP'}L), \tag{24}$$

and hence $V_P \sim V_{P'}$. □

A basic property of a group is the asymptotic scaling of V with L.

Example 5. Obviously $V \sim O(1)$ for finite groups and $V \sim \text{poly}(L)$ for Abelian groups.

Example 6. An important theorem of Gromov [8] states that *all* groups with polynomial growth are virtually nilpotent and have $V \sim L^d$ for integer d.

Example 7. Free groups provide the simplest examples with $V \sim \exp(L)$, and saturate the maximum possible growth rate (21). Generalizing slightly, one can also show that all virtually free groups exhibit exponential growth.

Groups with growth rate intermediate between polynomial and exponential exist — the "simplest" examples are groups arising from automorphisms of trees [7] — but it is conjectured that all such groups have only infinite presentations.

We now ask what implications group expansion has for the time and space complexity measures introduced above. It turns out that exponential growth is needed in order to have $D(L) > \text{poly}(L)$:

Theorem 1. *If a finitely presented*[7] *group G possesses a superpolynomial Dehn function, then G has exponential growth.*

Proof. The contrapositive of this theorem follows from Gromov's aforementioned result that all finitely generated groups with polynomial growth are virtually nilpotent (viz. have a nilpotent subgroup of finite index) [8]. Since virtually nilpotent groups have the same Dehn functions as nilpotent ones, we may combine Gromov's theorem with the fact that all nilpotent groups have $D(L) \sim L^d$ for some d [6] to arrive at the result. □

[7]Striclty speaking, this can be generalized to hold for finitely *generated* groups. We will however only be interested in groups with finite numbers of generators *and* relations.

Note that the convsere to theorem 1 is obviously false, as free groups on more than one generator provide examples of groups with exponential growth but with polynomial (in fact $D(L) \sim L$) Dehn functions.

We would also like to know the sizes of the different sectors $\mathcal{K}_g$. One result along these lines is that the sector sizes must get small as $|g|$ gets large:

Proposition 9. *Define*

$$\mathcal{D} \equiv \sum_{g \,|\, |g| \leq L} |\mathcal{K}_g(L)| = (2n_g + 1)^L \tag{25}$$

as the total number of words of length L. Then the size of $\mathcal{K}_g(L)$ is upper bounded as

$$\frac{|\mathcal{K}_g(L)|}{\mathcal{D}} \leq C \exp\left(-c\frac{|g|^2}{L}\right) \tag{26}$$

for some g, L-independent constants C, c.

Proof. The proof follows from connecting the counting of walks in $\mathcal{K}_g(L)$ with the heat kernel on the Cayley graph. Consider the symmetric simple lazy walk on the Cayley graph, and let $p_L(g,h)$ be the probability that a length-L walk starting at h ends at g. Then

$$\begin{aligned} p_L(g,h) &= \frac{1}{2n_g+1}\left(p_{L-1}(g,h) + \sum_{k\in\partial g} p(k,h)\right) \\ &= p_{L-1}(g,h) + \frac{1}{2n_g+1}\sum_{k\in\partial g}(p(k,h) - p(g,h)), \end{aligned} \tag{27}$$

where the sum over k runs over the neighbors of g in the Cayley graph. By translation invariance of the Cayley graph, we can fix $h = e$ wolog, and will simply write $p_L(g)$ for $p_L(g,e)$.

The preivous equation can be written more succinctly as

$$\delta_L p = \Delta p, \tag{28}$$

where Δ is the normalized graph Laplacian and δ_L denotes the discrete derivative along the "time" direction determined by L. Thus the different sizes of the $\mathcal{K}_g(L)$ are determined by using the heat equation to evolve a delta function concentrated on e for a total time of L. The claim we are trying to prove then follows from estimates of the discrete heat kernel Greens function developed in the graph theory literature, see e.g. [4] for a review. □

Various other facts follow from the observation that $p_L(x)$ obeys the heat equation. For example, it implies that $p_L(x)$ obeys strong maximum and minimum principles, which guarantees that all local maxima and minima of $p_L(x)$ occur on the boundaries of its domain of definition. It also means that since in groups of exponential growth almost all group elements in $\mathcal{B}(L)$ have geodesic diststance close to L, almost all $\mathcal{K}_g(L)$ will be exponentially smaller than $\mathcal{D}$.

In addition to the above asymptotic bound, we also know that $\mathcal{K}_e$ is always the largest sector:

Proposition 10. *For all L and all g, $|g| \leq L$, we have*[8]

$$|\mathcal{K}_e| \geq |\mathcal{K}_g|. \tag{29}$$

Proof. We aim to show that $p_L(e,e) \geq p_L(e,g)$ for all g, L. For simplicity of notation, let $L \in 2\mathbb{N}$. Then using $p_L(g,h) = p_L(gk,hk)$ for all g,h,k on account of translation invariance of the Cayley graph, we have

$$\begin{aligned} p_L(e,e) &= \sum_g p_{L/2}(e,g)^2 \\ &= \sqrt{\sum_h p_{L/2}(e,h)^2}\sqrt{\sum_{h'} p_{L/2}(g,h')^2} \\ &\geq \sum_h p_{L/2}(e,h) p_{L/2}(h,g) \\ &= p_L(e,g), \end{aligned} \tag{30}$$

where in the third line we used Cauchy-Schwarz. □

It is also possible to make statements about the absolute size of $\mathcal{K}_e$. In particular, $|\mathcal{K}_e|$ admits different bounds depending on the growth rate of the group [9]:

Fact 1. *For a group with growth function $V(L)$ scaling as a polynomial of L,*

$$L^{d_1} \lesssim V(L) \lesssim L^{d_2} \implies (L \log L)^{-d_2/2} \lesssim \frac{|\mathcal{K}_e|}{\mathcal{D}} \lesssim L^{-d_1/2}. \tag{31}$$

For a group with exponential growth,

$$\frac{|\mathcal{K}_e|}{\mathcal{D}} \lesssim e^{-L^{1/3}}. \tag{32}$$

We can also connect the above bound with a previous result to derive:

Corollary 1. *Groups with superpolynomial Dehn functions have exponentially many group sectors for words of a fixed length, and have an identity sector $\mathcal{K}_e$ that contains a fraction of all length-L words that scales at most as $e^{-L^{1/3}}$.*

Proof. This follows by combining fact 1 with theorem 1. □

[8] Note that in order for this proposition to be true, we need the words in $\mathcal{K}_g$ to be drawn from the alphabet $S \cup S^{-1} \cup \{e\}$, where S is the generating set. If we were to just use $S \cup S^{-1}$ then it obviously cannot be strictly true, as in this case one could e.g. have a bipartite Cayley graph such that $|\mathcal{K}_e(L)| = 0$ for odd L.

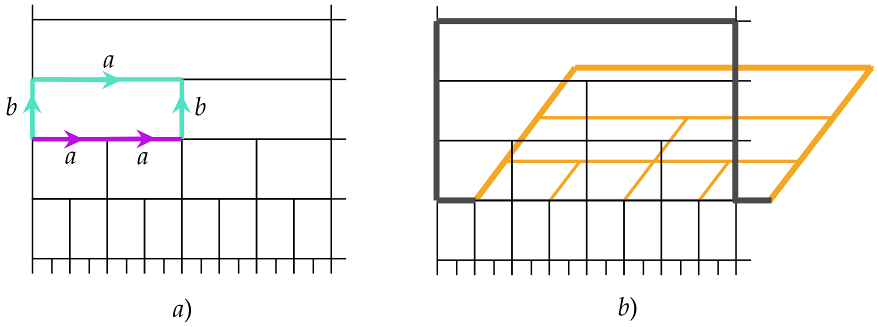


Figure 1: The Cayley graph of $BS(1,2)$. a) A (small portiona of a) single sheet of the Cayley graph. The cyan and purple portions of the outlined cell end at the same location on account of $ab = baa$. b) A second sheet which connects to the one drawn in a) is shown in orange. A word in $\mathcal{K}_e$ which is in the family of those with exponentially large area is drawn as the outlined loop (corresponding to (39) with $n = 3$).

2 The geometry of Baumslag-Solitar groups

In this section we state and prove some facts about the group geometry of the Baumslag-Solitar group $BS(1,2)$ and some of its simple generalizations. We will define this group by the presentation

$$BS(1,2) = \langle a, b \,|\, ab = baa \rangle. \tag{33}$$

Everything we say below can in fact be modified only slightly in order to work for $BS(1,q) = \langle a, b \,|\, ab = ba^q \rangle$, but for concreteness we will specify to $q = 2$ throughout. It will occasionally be useful to employ the matrix representation

$$a = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}, \tag{34}$$

which is useful for e.g. quickly solving the word problem.

Our notation will carry over from the previous section on general aspects of group geometry. For example, $\mathcal{K}_g(L)$ will denote the set of all length-L words w that reduce to the group element g, $g(w)$ will denote the group element that a word reduces to, and we will write $w \sim w'$ if $g(w) = g(w')$. We will also use $n_b(w)$ to denote the net number of bs that appear in w, namely

$$n_b(w) \equiv \sum_{i=1}^{|w|} (\delta_{w_i,b} - \delta_{w_i,b^{-1}}). \tag{35}$$

The "worst-case" geometry of $BS(1,2)$

We will start by analyzing the worst-case time and space complexity of words in $BS(1,2)$, which we do by computing the Dehn and expansion length functions.

To orient ourselves, it is helpful to realize that the Cayley graph of $BS(1,2)$ is homeomorphic to the product of the real line with a 3-regular tree (for $BS(1,q)$ it is a $(q+1)$-tree), with the depth of a given word w along the tree being controlled by the

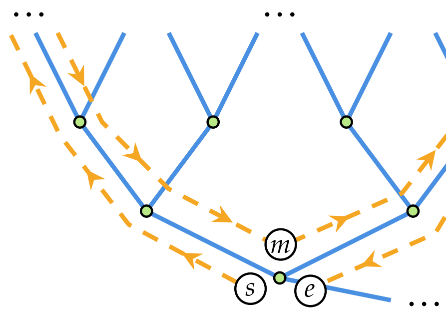


Figure 2: The excursion on the b-tree performed by the word w_{big}. The walk corresponding to w_{big} starts at the node labeled by s, moves to a depth n in the tree, and then returns to depth 0 at the midpoint m. By the time it reaches m, the walk has traveled a distance of 2^n along the a axis of the Cayley graph. In the second half of the walk, from m to e, the walk makes another depth-n excursion along a different branch of the tree.

relative number of bs and b^{-1}s that w contains. We will refer to this tree as the *b-tree*, and will use terminology whereby multiplying by b (b^{-1}) moves one "up" ("down") on the sheet of the b-tree one is currently at, while multiplying by a (a^{-1}) moves one "right" ("left") within the sheet (along the *a-axis*). Because of the heirarchical nature of the graph, multiplying by b — viz. moving deeper into the b-tree — moves one to larger scales, while multiplying by b^{-1} does the opposite.

Because of the tree structure, it is clear that

$$|\mathcal{B}(L)| \sim \lambda^L \tag{36}$$

grows exponentially with L for some λ. A naive guess for the value of λ is as follows. First, we realize that the exponential growth comes from the tree structure, and that motion by one node on this tree is always possible though the use of at most two group generators (since to perform an arbitrary move on the tree one must multiply by either b, b^{-1}, or ab). Since the number of points at depths $d \leq L$ of a 3-regular tree goes like 3^L, we therefore estimate $|\mathcal{B}(L)| \sim 3^{L/2}$, giving $\lambda \approx \sqrt{3}$. This estimate is actually not bad (although interestingly is a slight *over*estimate), as

Fact 2. *The growth rate exponent λ[9] of $BS(1,2)$ is equal to the positive root of $x^3 - x^2 - 2$ [3],*

$$\lim_{L\to\infty} (|\mathcal{B}(L)|)^{1/L} \approx 1.69. \tag{37}$$

Similar but slightly larger exponents arise for $BS(1,q)$.

Dehn function

We now turn to computing the Dehn function and expansion length of $BS(1,2)$. The Dehn function must be large, since it is easy to construct words with $A(w) \sim 2^{|w|}$:

[9] The exact growth rate depends on the presentation (although it being exponential does not), but for $BS(1,q)$ the canonical presentation (viz. the one used here) has the smallest growth rate out of all presentations and in this sense is thus unambiguous.

Proposition 11. *Define* $w_n \equiv b^{-n}ab^n$*, so that* $w_n \sim a^{2^n}$*.*[10] *Then the word*

$$w_{big} = w_n a w_n^{-1} a^{-1} \tag{38}$$

has area

$$A(w_{big}) = 2^{n+1} - 2 \sim 2^{|w_{big}|}. \tag{39}$$

A visual illustration of this statement is given in panel b) of Fig. 1, with the walk given by w_{big} projected onto the b-tree shown in Fig. 2. The proof is as follows:

Proof. To determine $A(w_{big})$, we need to find the minimal number of relations needed to turn w_{big} into the identity word. Geometrically, this corresponds[11] to the number of 2-cells in the Cayley complex that form a minimal spanning surface $S_{w_{big}}$ with w_{big} as its boundary. Note that the loop defined by w_{big} is entirely contained within two sheets of the Cayley graph. It is furthermore clear that if we only consider bounding surfaces $S_{w_{big}}$ contained within these two sheets, the minimal bounding surface has area

$$2\sum_{k=0}^{n} 2^k = 2^{n+1} - 2. \tag{40}$$

Therefore we need only show that this area cannot be reduced by considering surfaces that extend into other sheets. This is however clearly true, as a 2-cell of $S_{w_{big}}$ that lives on any other sheet will necessarily make an unwanted contribution to $\partial S_{w_{big}}$. More formally, we can recognize that, being homeomorphic to the product of $\mathbb{R}$ with a 3-tree, the Cayley 2-complex is contractible, and thus the S_{big} found above is the unique bounding surface. □

Note that by considering a homotopy which shrinks w_{big} down to the identity by first making the loop narrower along the a axis before shrinking it along the b axis (see Fig. 1), the length of the word does not parametrically increase — this is an intuitive explanation for the following quasi-nontrivial fact:

Fact 3. $BS(1,2)$ *has linear expansion length,* $EL(L) \sim L$ *[5].*

The previous proposition immediately implies the existence of exponentially many (in L) elements of $\mathcal{K}_e(L)$ with exponentially large (also in L) area. It also allows us to prove

Proposition 12. $BS(1,2)$ *has exponential Dehn function:*

$$D(L) \sim 2^L. \tag{41}$$

Proof. The construction above tells us that $D(L) \gtrsim 2^L$. A matching upper bound can be proven using an algorithm which converts an input word to a particular standard form. This is done by simply moving all occurances of b in w to the left and all occurances of b^{-1} to the right, dupicating a and a^{-1}s along the way as needed and (optionally, for our present purposes) eliminating bb^{-1} pairs as they are encountered.

[10] w_n is nearly a geodesic for a^{2^n}; the true geodesic with our presentation is instead $b^{-(n-1)}a^2b^{n-1}$, which is one character shorter than w_n.

[11] Up to the aforementioned $O(|w_{big}|^2)$ contributions.

It is clear that at most an exponential in L number of additional as are generated during this process of shuffling the bs and b^{-1}s around (as usual, this means exponential up to a polynomial factor, here linear in L), so that the resulting word is $w' = b^k w_a b^{-m}$, where w_a contains only as and a^{-1}s and has length $|w_a| \lesssim 2^L$ asymptotically. Since $w \sim e$, we know that $k = m$ and that $w_a \sim e$. Thus an additional 2^L relations suffice to reduce w' to e, and 2^L is also an upper bound on $D(L)$. $\square$

Average-case complexity

Distribution of geodesic lengths for random words

One natural question we can ask about the word geometry of $BS(1,2)$ is this: given a random length-L word w, what is the geodesic distance $|g(w)|$ of w? We will answer this in the following way. First, note that any word can always be mapped to a unique standard form

$$w_{knl} = b^k a^n b^{-l}, \qquad k,n,l \in \mathbb{Z},\ n,l \geq 0, \tag{42}$$

where n can be even only if at least one of k,l are zero (if both are nonzero then we can simplify $b^k a^{2n} b^{-l} \sim b^{k-1} a^n b^{-(l-1)}$). The truth of this statement can be seen by looking at the matrix representation (34) or by having a good long stare at the Cayley graph.

Since k,n,l are unique, the w_{knl} serve as a set of canonical representatives for each g. To get the geodesic of an arbitrary word, we first reduce it to this form, and then make use of the fact that [1]

$$|g(w_{knl})| \sim k + l + (1-\delta_{n,0}) \log |n| \tag{43}$$

where we have dropped constant contributions and ignored overall constant prefactors.

All that remains is to find a way of determining k,n,l given an arbitrary word w. This is done using the matrix representation (34), in which w_{knl} becomes

$$w_{knl} = \begin{pmatrix} 2^{l-k} & n2^{-k} \\ 0 & 1 \end{pmatrix}. \tag{44}$$

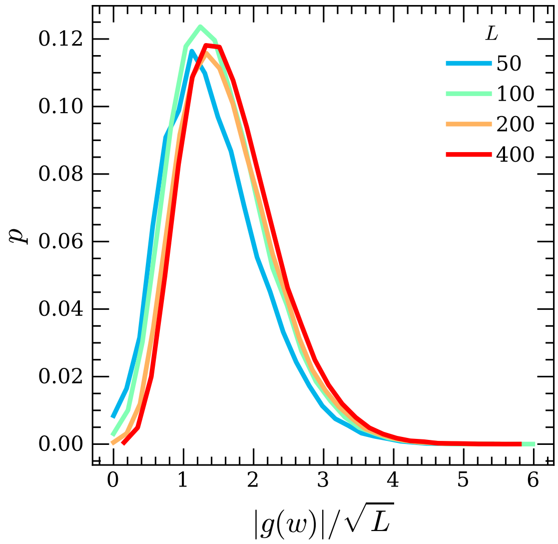
Therefore to find k,n,l for an input string w, we proceed as follows. We first find the matrix corresponding to w by explicit matrix multiplication, yielding a result of the form $\begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix}$. We know right away that $\log_2 A = n_b(w) = k - l$, the number of bs contained in w minus the number of b^{-1}s. Then:

- If $B \in \mathbb{Z}$, either $k = 0$ or $l = 0$. Which one of these scenarios holds depends on $\mathrm{sgn}(n_b)$: if $n_b < 0$ then $k = 0, l = -n_b$ and $n = B$, while if $n_b > 0$ then $l = 0, k = n_b$ and $n = 2^{n_b} B$.

- If $B \notin \mathbb{Z}$, then $k > 0$, and k is determined by the number of significant figures after the decimal point when B is represented in binary,[12] after which both l and n are determined.

[12] There is a small subtlety here, since this approach is incorrect if $l = 0$ and $n = 2^m n_o$ with $m > 0, n_o \in 2\mathbb{Z}+1$. However in this case it is easy to show that the naive values of (k,n,l) one obtains from this procedure in this case are $(k,n,l)_{\text{naive}} = (k-m, n_0, -m)$. Since $l_{\text{naive}} < 0$, this mistake is easy to identify and correct for.

Numerically implementing the above procedure for a few modest values of L gives the following histogram:

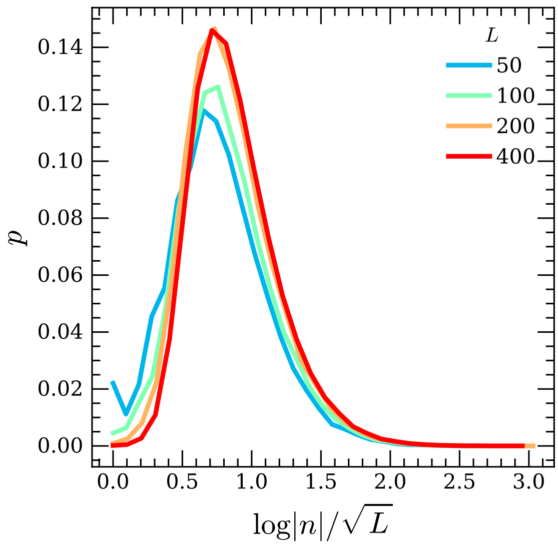


, (45)

where p is the probability of getting a word with a given geodesic length. From this we see that although the probability of getting $|g(w)| = 0$ (viz. getting $w \in \mathcal{K}_e$) is suppressed (we know that it goes as $\sim e^{-L^{1/3}}$), and that the median value for the geodesic distance still goes as $\sim \sqrt{L}$.

Since we are sampling random words, $n_b(w) = k - l$ will converge to a Gaussian of width $\sqrt{L}$; thus a typical word will reach a depth of $\sqrt{L}$ on the b-tree part of the Cayley graph. We also expect the $k + l$ contribution to the geodesic estimate (43) to scale as $\sqrt{L}$; indeed this can be numerically verified to be the case. The fact that a random word typically has $|g(w)| \sim \sqrt{L}$ then means that the $\log |n|$ contribution to the geodesic estimate scales as $\log |n| \lesssim \sqrt{L}$. This means that words which traverse *exponentially* far along the a axis of the Cayley graph (viz. those with $\log |n| \sim L$) are rare, meaning that we should not expect the Dehn time of typical words to be close to the worst-case result.

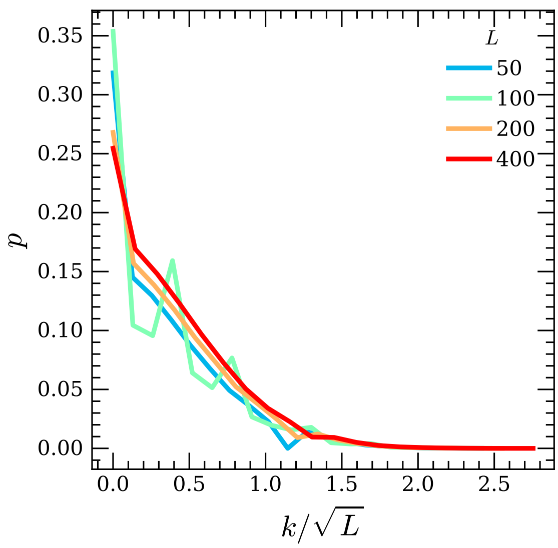
To determine whether $\log |n| < \sqrt{L}$ or $\log |n| \sim \sqrt{L}$, we simply make a histogram of $(\log|n|)/\sqrt{L}$, finding



(46)

Thus the $\log|n|$ part makes an $O(1)$ contribution to the expected geodesic distance, and we learn that a random word typically travels a distance of $2^{\sqrt{L}}$ along the a direction of the Cayley graph. This fact will be important in our following discussion of random walks and typical Dehn times.

We can also look at the expected distribution of k and l individually. For k, we find (l is identical)



(47)

Dehn times of typical words

The Dehn function is a *worst*-case complexity result, and gives the maximal area over all length-L words in $\mathcal{K}_e$. For physics applications we are instead usually interested in the *average*-case complexity, viz. the typical value of $A(w)$ for w chosen uniformly from $\mathcal{K}_e$ (not to be confused with the average of $A(w)$ over $w \in \mathcal{K}_e$!).[13] We might expect

[13]Note that it appears reasonable that sampling words randomly from $\mathcal{K}_e$ is actually exponentially hard (the author is not aware of any efficient sampling algorithm, in any case). Thus we are not

this area to be fairly large due to our previous result on $BS(1,2)$'s Dehn function, but determining exactly how large will require a bit of work.

This question can equivalently be phrased as asking about the typical amount of area enclosed by a random walk on the Cayley graph which returns to the origin after L steps. Despite the exponential growth of the Cayley graph, walks which return to the origin (i.e. words $w_e \in \mathcal{K}_e$) constitute a fraction of all 5^L words which is *not* exponentially small. This would not be the case on a tree, where the set of returning length-L walks is exponentially less than the total number of length-L walks — thus the loops introduced into the $BS(1,2)$ Cayley graph serve to quasi-strongly "attract" random walks to the origin. Interestingly, while the return probability is not exponentially small, neither is it polynomially small. One of the few known results about random walks on the BS groups is that

Fact 4. *The return probability of random walks on $BS(1,2)$ scales as [9]*

$$\frac{|\mathcal{K}_e|}{5^L} \sim e^{-L^{1/3}}, \tag{48}$$

which saturates the upper bound on this quantity for groups with exponential growth as given in fact 1.

We will start by understanding how random returning walks on $BS(1,2)$ behave when they are projected onto the tree. We will assume that the walk on $BS(1,2)$ is symmetric and range-1, meaning that at each step, each move — viz. multiplication by one of a, a^{-1}, b, b^{-1}, e — occurs with an equal probability of $1/5$.

The first thing to understand is that symmetric random walks do not remain symmetric when projected onto the b-tree. Moving upwards while staying on the same sheet can be done by moving directly upwards, or by moving to the left or right by an even number of steps, and then moving up. Moving upwards onto a *different* sheet on the other hand is done by moving left or right by an *odd* number of steps before moving up. Finally, moving downards (on the same sheet) can be done by moving an arbitrary amount either left or right, and then moving downards.

From the discussion above we immediately see that the probability to move downards is actually equal to the probability of moving upwards, despite the fact that moving upwards can be done on either of two sheets. More precisely, let $p^{\nwarrow}, p^{\nearrow}$ and $p^{\downarrow}$ be the probabilities of moving up on the same sheet, up on a different sheet, and down, respectively. Note that the values of these quantities will be the same for both lazy and non-lazy walks (since occurances of e can simply be ignored for the purposes of computing the p^σ), and so we can pretend the walk is lazy. Thus

$$\begin{aligned} p^{\nwarrow} &= \frac{4}{5}\frac{1}{4}\sum_{k\in\mathbb{Z}}\sum_{l=|k|}^{\infty}\binom{2l}{l+|k|}(1/4)^{2l} = \frac{1}{3} \\ p^{\nearrow} &= \frac{1}{2} - p^{\nwarrow} = \frac{1}{6} \\ p^{\downarrow} &= \frac{1}{2}. \end{aligned} \tag{49}$$

interested in the typical value of $A(w)$ because a typical w is easily preparable, but rather because $A(w)$ for typical w determines how a system prepared in e and evolved by randomly applying relations thermalizes across $\mathcal{K}_e$.

The fact that $p^{\downarrow} = 1/2$ means that as far as radial motion in the tree is concerned, the walk will *not* move ballistically upwards or downwards, but will instead move diffusively. Note that the average distance between occurances of bs or b^{-1}s is

$$\frac{2}{5}\sum_j j(3/5)^j = \frac{3}{2}. \tag{50}$$

Thus motion on the tree behaves like a returning length-$2L/3$ symmetric random walk on $\mathbb{Z}$.

We will use the fact that motion on the b-tree is diffusive to prove the following theorem:

Theorem 2. *Let $w \sim e$ be a randomly chosen word in $\mathcal{K}_e$. Then*

$$A(w) \sim 2^{\sqrt{L}} \tag{51}$$

with probability that approaches unity in the limit $L \to \infty$.

Intuitively, this theorem holds because as was seen in (39), it is easy to produce words whose areas are exponentially large in the maximum depth they attain on the tree. For a random length-L word the above result about motion on the tree suggests a typical maximum depth scaling as $\sqrt{L}$, which in turn suggests the correctness of (51).

We begin with a lemma which illustrates the proof strategy.

Lemma 1. *Consider a length-L word w constrained to have the following characteristics, but which is otherwise chosen randomly. First, $n_b(w) = 0$. Second, the returning walk induced on $\mathbb{Z}$ by restricting to the b, b^{-1} characters of w is constrained to $\mathbb{Z}^{\geq 0}$ (i.e. it is a Dyck walk), meaning that the cumulative sums $n_b(x) = \sum_{j=1}^{x}(\delta_{w_i,b} - \delta_{[w_i]_j,b})$ are positive for all x.*

Let $S_{Dyck}(w)$ be the set of all length-L words w' that satisfy $w' \sim w$ and which themselves obey the above Dyck walk property. Then if w' is drawn randomly from $S_{Dyck}(w)$, w and w' are exponentially likely to satisfy $d(w, w') \sim 2^{\sqrt{L}}$.

Proof. The Dyck walk property means that all $w' \in S_{Dyck}(w)$ are reducible to a^{n_w} for some n_w, with n_w shared by all words in $S_{Dyck}(w)$. This can be seen by an inspection of the Cayley graph, or by recalling the canonical form (42) (the walks corresponding to a group element are read right to left, so $b^k a^n b^{-l} \in S_{Dyck}(w)$ only if $l = k = 0$).

Since w was chosen randomly from the set of words obeying the Dyck walk condition, the b-walk defined by w is exponentially likely to reach a height of $O(\sqrt{L})$, viz. to have $\max_x n_b(x) \sim \sqrt{L}$. A random w fullfiling this condition is easily seen to be exponentially likely to reduce to a word a^{n_w} of length $n_w \sim 2^{\sqrt{L}}$, as can be seen by e.g. following the procedure that brings w into canonical form (42).

Supose now that n_w scales as $2^{\sqrt{L}}$, and consider a random element $w' \in S_{Dyck}(w)$. We claim that the distance between w and w' is exponentially likely to scale as $d(w, w') \sim 2^{\sqrt{L}}$. This is because w, w' are exponentially likely to travel on different sheets of the Cayley graph for nearly all of the length of their walks. Suppose that the walks of w, w' go on to distinct sheets starting at depth d (with d exponentially likely to be $O(1)$). The contractibility of the Cayley 2-complex means that the minimal bounding surface linking w to w' must consist of all cells bounded by w and the a axis that lie at a depth greater than d, together with the analogous set of cells for w' (a similar argument arose in the proof of proposition 11, where there we had $d = 0$). Since each of these contributions to the bounding surface consists of $\sim 2^{\sqrt{L}}$ cells, we indeed have $d(w, w') \sim 2^{\sqrt{L}}$. □

We are now in a position to prove Theorem 2. While Dyck walks do not constitute a constant fraction of all returning walks[14], the basic idea is to realize that a generic word in $\mathcal{K}_e$ is likely to contain at least two subwords of size $\sim \sqrt{L}$ obeying the Dyck property, allowing an application of the above lemma.

Proof. When sampling a random word in $\mathcal{K}_e$, we can first sample uniformly from all $5^{L/2}$ words w_L of length $L/2$, and then sample from all words w_R of length $L/2$ such that $w_L w_R \sim e$. Since w_L is chosen randomly, it is easy to show that with unit probaility in the $L \to \infty$ limit, the walk defined by w_L reaches a distance of $\sim 2^{\sqrt{L}}$ along the a axis of the Cayley graph (see also the numerics employed in the discussion of typical geodesic distances which emperically confirm this fact). Since $|w_L| \sim L$, reaching this distance is only possible if w_L contains an excursion along a particular sheet of the b-tree which reaches a maximal depth of $\sim \sqrt{L}$. w_L must therefore contain at least one subword $w_{L,D}$ which performs a Dyck walk of height $\sim \sqrt{L}$.

We now consider sampling words w_R such that $w_L w_R \sim e$. Since w_R must also traverse a distance of $\sim 2^{\sqrt{L}}$ along the a axis, it too must contain a subword $w_{R,D}$ which performs a Dyck walk of height $\sim \sqrt{L}$. By the contractibility of the Cayley 2-complex, the areas that $w_{L,D}$ and $w_{R,D}$ define with respect to the a axis can cancel out only if the excursions that $w_{L,D}$ and $w_{R,D}$ perform occur along the *same* sheets of the b-tree for almost the entirety of their respective walks. The chance for this to occur is however exponentially small in the depth of the walk, which goes as $\sqrt{L}$. Therefore the areas contributed by $w_{L,D}$ and $w_{R,D}$ are exponentially unlikely to cancel, and thus the area of $w = w_L w_R$ will scale as $2^{\sqrt{L}}$ with probability approaching 1 as $L \to \infty$. □

Theorem 2 specifically deals with the number of steps needed to reduce $w \in \mathcal{K}_e$ to e. However, the identity word is not special in this regard, and the same scaling occurs for reduction to any other element in $\mathcal{K}_e$:

Corollary 2. *The typical number of moves needed to connect two randomly chosen length-L words $w_{1,2} \in \mathcal{K}_e$ scales as $d(w_1, w_2) \sim 2^{\sqrt{L}}$.*

Proof. Let $w_{1,2}$ be two random words in $\mathcal{K}_e$. We have seen that $A(w_1 w_2^{-1}) \sim d(w_1, w_2)$ up to factors linear in the $|w_i|$. Now both of the w_i are likely to have area that scales as $2^{\sqrt{L}}$; in what follows we will assume that this is indeed the case. We now use the same argument as was employed during the proof of Theorem 2: because of the contractibility of the Cayley 2-complex, the only way for $A(w_1 w_2^{-1})$ to be parametrically smaller than $\sim \max(A(w_1), A(w_2^{-1}))$ is if the large sheet excursion of w_1 is cancelled by that of w_2^{-1}. By the usual logic, the probability that the large excursions of both

[14] The number of length-L Dyck walks goes as $L^{-3/2} 2^L$, while the number of returning walks goes as $L^{-1/2} 2^L$; thus the fraction of returning walks that are also Dyck walks scales as $1/L$.

w_i occur on the same sheet is however exponentially small, and thus $A(w_1 w_2^{-1}) \sim 2^{\sqrt{L}}$ with probability 1 in the $L \to \infty$ limit. $\square$

By a similar variant on the proof of Theorem 2, a matching result holds when we look beyond $\mathcal{K}_e$:

Corollary 3. *Let w_1 be a random length-L word. Then a random word w_2 drawn uniformly from $\mathcal{K}_{g(w_1)}$ satisfies $d(w_1, w_2) \sim 2^{\sqrt{L}}$ with probability 1 in the $L \to \infty$ limit.*

This result is perhaps the most interesting one so far from a physics point of view, as it gives super-polynomially long lower bounds on thermalization times for systems prepared with generic initial conditions.

A small aside: the above results are concerned with the average case time complexity of mapping random words within their respective sectors. The worst case complexity in the sector defined by a random word w is however always exponential in L:[15]

Proposition 13. *Let w_1 be a random length-L word. Then with probability 1 in the $L \to \infty$ limit, $\mathcal{K}_{g(w_1)}$ contains exponentially many words w_2 which with $d(w_1, w_2) \sim 2^L$.*

Proof. As we saw in our discussion of geodesic lengths, a random w will have $|g(w)| \sim \sqrt{L}$ with probability 1 as $L \to \infty$. Let w_{geo} be a geodesic word in $\mathcal{K}_{g(w)}$. Then we can construct exponentially many words $w_1, w_2 \in \mathcal{K}_g$ with $d(w_1, w_2) \sim 2^L$ simply by letting $w_i = w_{geo} w_i'$, where $|w_i'| \sim L$ are chosen at random from the words in $\mathcal{K}_e(L - |g(w)|)$ with exponentially large area. The distance $d(w_1', w_2')$ is then exponentially likely (in the choice of the w_i') to scale as 2^L. $\square$

Iterated BS groups

We now turn to examining the group $BS^{(2)}(1,2)$, which we define through the presentation[16]

$$BS^{(2)}(1,2) = \langle a, b, c \,|\, ab = baa,\ bc = cbb \rangle. \tag{52}$$

The Cayley graph consists of an infinite heirarchy of $BS(1,2)$ Cayley graphs. Multiplying by c increases the "scale" of the $BS(1,2)$ graph generated by a, b, while multiplying by c decreases the scale. This group is interesting for several reasons, one of which is:

[15] Of course we know that the worst case complexity is upper bounded by 2^L by the fact that $D_{g(w)}(L) \lesssim D(L)$ for all w. Here we showing that this upper bound is in fact almost always (on the choice of w) saturated.

[16] Unlike $BS(1,2)$, I could not find a matrix representation for $BS^{(2)}(1,2)$. This is because the subgroup $\langle a, c \rangle$ is isomorphic to $\mathbb{Z} * \mathbb{Z}$. This complicates things because the only matrix representation of a free group I know about has generators $a_1 = \begin{pmatrix} 1 & \\ m & 1 \end{pmatrix}$, $a_2 = \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix}$; with $|mn| \geq 4$ (as proven by the ping-pong lemma), and it is hard to fit this structure in with the duplication rules required by the interactions between a, b and b, c.

In fact there is good reason for not being able to find such a matrix representation, since it is in fact an outstanding open question if there exist groups with $D > 2^L$ for which the word problem can be solved in linear time (a matrix representation for $BS^{(2)}(1,2)$ would answer this question in the affirmative). Note that this question is specifically about *linear* time algorithms; there are many examples of groups with crazy superexponential Dehn functions which nevertheless admit polynomial time word problem algorithms (such as the iterated BS groups!).

Proposition 14. *$BS^{(2)}(1,2)$ has Dehn function*

$$D(L) \sim 2^{2^L}. \tag{53}$$

The key result we need to complete the proof is known as *Britton's lemma*,[17] which is stated as follows:

Lemma 2. *Let G be a group with presentation S. Further let G contain two isomorphic subgroups $H, K \subset G$, with $\phi : H \to K$ the isomorphism between them. Define the group*

$$G_\phi \equiv \langle S, t \,|\, t^{-1}Ht = \phi(H)\rangle. \tag{54}$$

Now any word w on $\{S,t\}^*$ can be written in the form

$$w = g_0 t^{\varepsilon_1} g_1 t^{\varepsilon_2} g_2 \cdots t^{\varepsilon_{n-1}} g_{n-1} t^{\varepsilon_n} g_n, \qquad g_i \in G \ \varepsilon_i = \pm 1. \tag{55}$$

Britton's lemma states that if $w \sim e$ represents the identity in G_ϕ, then there must be some i such that either

1. *$n = 0$ and $g_0 = e$,*
2. *$\varepsilon_i = -1, \varepsilon_{i+1} = 1$ and $g_i \in H$, or*
3. *$\varepsilon_i = 1, \varepsilon_{i+1} = -1$ and $g_i \in K$.*

This gives us a way of simplifying any word $w \sim e$ representing the identity in G_ϕ. Case 1 above is trivial. In case 2, we can replace the occurance of $t^{-1}ht$ with $\phi(h)$, while in case 3 we may replace tkt^{-1} with $\phi^{-1}(k)$. After doing this reduction, we will be left with a new word w' which still represents e, and can apply the lemma again. This gives us a guarantee that we will always be able to apply a successive series of reductions to eliminate all ts from and $w \sim e$ in $\{S,t\}^*$ to obtain a word $w_G \in \{S\}^*, w_G \sim e$. The application to $BS(1,2)$ is clear, where we take $G = \mathbb{Z}$, $H = G$, and $K = 2\mathbb{Z}$.

We now return to a proof of the Dehn function scaling:

Proof. We first construct a lower bound. Define the word $w_n = c^{-n}bc^n$, so that $w_n \sim b^{2^n}$. Then feed this word into the construction of the large-area word w_{big} for $BS(1,2)$, by defining $w_n' = w_n a w_n^{-1}$. Then we claim the word

$$w_{huge} = (w_n')^{-1} a w_n' a^{-1} \tag{56}$$

has area

$$A(w_{huge}) \sim 2^{2^n}, \tag{57}$$

which is doubly exponential in $|w_{huge}|$. This follows by an argument similar to the one we gave for the area of w_{big} in $BS(1,2)$. The tree-like structure of the sheets of the $BS(1,2)$ Cayley graph give tree-like structures both for words built from b, c and those built from a, b. Letting $\widetilde{w}_{huge} = b^{-2^n} a b^{2^n}$, this means that

$$A(w_{huge}) = A(\widetilde{w}_{huge} a \widetilde{w}_{huge}^{-1} a^{-1}) + O(2^n). \tag{58}$$

[17] I thank Tim Riley for suggesting this proof strategy.

But using our results from our study of $BS(1,2)$, we know that $A(\widetilde{w}_{huge}a\widetilde{w}_{huge}^{-1}a^{-1}) \sim 2^{2^n}$. Thus $D(L) \geq 2^{2^L}$ asymptotically.

We now need to provide a matching upper bound. We do this by combining Britton's lemma with our earlier result about $BS(1,2)$. Note that $BS^{(2)}(1,2)$ can be obtained from $BS(1,2)$ using just the type of extension as appears in Britton's lemma, where $G = BS(1,2), H = \langle b \rangle, K = \langle b^2 \rangle$. Then we know that if we are given $w \sim e$, $|w| = L$ in $BS^{(2)}(1,2)$, we can obtain a word $w' \sim e$ in $BS(1,2)$ after at most $O(L)$ applications of $c^{-1}bc = b^2$. The maximum amount that $|w|$ can grow by under these substitutions is $O(2^L)$. Thus an upper bound on $D(L)$ in $BS^{(2)}(1,2)$ can be obtained by an upper bound on $D(2^L)$ in $BS(1,2)$. Using our previous result on the latter, we conclude

$$D(L) \lesssim 2^{2^L}, \tag{59}$$

and thus when combined with the lower bound, we also have that $D(L) \lesssim 2^{2^L}$. $\square$

$BS^{(2)}(1,2)$ also provides an example with a superlinear expansion length:

Corollary 4. $BS^{(2)}(1,2)$ *has exponential expansion length,*

$$EL(L) \sim 2^L. \tag{60}$$

Proof. From the general bound (18) and our above result about the Dehn function of $BS^{(2)}(1,2)$, we know that $EL(L) \geq 2^L$. The upper bound follows from the above application of Britton's lemma and the fact that the expansion length of $BS(1,2)$ is only $O(L)$. $\square$

It is easy to recursively generalize the above examples to construct groups with even faster growing space and time complexity:

Corollary 5. *Define the group* $BS^{(l)}(1,2)$ *through the presentation*

$$BS^{(l})(1,2) = \langle a_0, \ldots, a_l \,|\, a_{i-1}a_i = a_i a_{i-1} a_{i-1},\ i = 1, \ldots, n \rangle. \tag{61}$$

This group has Dehn function and expansion length

$$D(L) \sim \exp^{(l)}(L), \qquad EL(L) \sim \exp^{(l-1)}(L). \tag{62}$$

Average case complexity of $BS^{(l)}(1,2)$

ethan: *under construction...*

References

[1] J. Burillo and M. Elder. Metric properties of baumslag–solitar groups. International Journal of Algebra and Computation, 25(05):799–811, 2015.

[2] M. Clay and D. Margalit. Office hours with a geometric group theorist. Princeton University Press, 2017.

[3] D. J. Collins, M. Edjvet, and C. Gill. Growth series for the group¡ x, y¦ x- 1 yx= yl¿. Archiv der Mathematik, 62:1–11, 1994.

[4] E. B. Davies. Heat kernel bounds, conservation of probability and the feller property. Journal d'Analyse Mathématique, 58(1):99–119, 1992.

[5] S. M. Gersten. Isoperimetric and isodiametric functions of finite presentations. Geometric group theory, 1:79–96, 1993.

[6] S. M. Gersten, D. F. Holt, and T. R. Riley. Isoperimetric inequalities for nilpotent groups. Geometric & Functional Analysis GAFA, 13:795–814, 2003.

[7] R. Grigorchuk and I. Pak. Groups of intermediate growth: an introduction. Enseign. Math.(2), 54(3-4):251–272, 2008.

[8] M. Gromov. Groups of polynomial growth and expanding maps (with an appendix by jacques tits). Publications Mathématiques de l'IHÉS, 53:53–78, 1981.

[9] W. Woess. Random walks on infinite graphs and groups. Number 138. Cambridge university press, 2000.