1
Group theory

1.1 Introduction and basic notation

In this book we are assuming that the reader has studied group theory at undergraduate level, and is familiar with its fundamental results, including the basic theory of free groups and group presentations. However, in many of the interactions between group theory and formal language theory, it is convenient to consider group presentations as special cases of semigroup and monoid presentations, so we describe them from that aspect here.

We refer the reader to one of the standard textbooks on group theory, such as [223] or [221] for the definitions and basic properties of nilpotent, soluble (solvable) and polycyclic groups,

We also include some specific topics, mainly from combinatorial group theory, that will be required later. The normal form theorems for free products with amalgamation and HNN-extensions are used in the proofs of the insolubility of the word problem in groups, and we summarise their proofs. We introduce Cayley graphs and their metrical properties, and the idea of *quasi-isometry* between groups, which plays a central role in the area and throughout geometric group theory, and we define the small cancellation properties of presentations and describe related results.

The final section of the chapter is devoted to a brief introduction to some of the specific families of groups, such as Coxeter groups and braid groups, that arise frequently as examples throughout the book. The informed reader may prefer not to read this chapter in detail, but to refer back to it as necessary.

1.1.1 Some basic notation For g, h in a group, we define the conjugate of g by h, often written as g^h, to be hgh^{-1} and the commutator $[g, h]$ to be $ghg^{-1}h^{-1}$. But we note that some authors use the notations g^h and $[g, h]$ to mean $h^{-1}gh$ and $g^{-1}h^{-1}gh$, respectively.

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

We recall that a *semigroup* is a set with an associative binary operation, usually written as multiplication, a *monoid* is a semigroup with an identity element, and a *group* is a monoid G in which every element is invertible.

We extend the multiplication of elements of a semigroup S to its subsets, defining $TU = \{tu : t \in T, u \in U\}$ and we frequently shorten $\{t\}U$ to tU, as we do for cosets of subgroups of groups.

1.1.2 Strings and words Strings over a finite set are important for us, since they are used to represent elements of a finitely generated group.

Let A be a finite set: we often refer to A as an *alphabet*. We call the elements of A its *letters*, and we call a finite sequence $a_1a_2 \cdots a_k$ of elements from A a *string* or *word* of length k over A. We use these two terms interchangeably. We denote by ε the string of length 0, and call this the *null string* or *empty word*. For a word w, we write $|w|$ for the length of w.

We denote by A^k the set of all strings of length k over A, by A^* the set (or monoid) of all strings over A, and by A^+ the set (or semigroup) of all nonempty strings over A; that is

$$A^* = \bigcup_{k=0}^{\infty} A^k, \quad A^+ = \bigcup_{k=1}^{\infty} A^k = A^* \setminus \{\varepsilon\}.$$

For $w = a_1a_2 \cdots a_k$ and $i \in \mathbb{N}_0$, we write $w(i)$ for the prefix $a_1a_2 \cdots a_i$ of w when $0 < i \le k$, $w(0) = \varepsilon$ and $w(i) = w$ for $i > k$.

In this book, A often denotes the set $X \cup X^{-1}$ of generators and their inverses for a group G; we abbreviate $X \cup X^{-1}$ as $X^{\pm}$. In this situation, we often refer to words in A^* as *words over* X even though they are really words over the alphabet A.

For $g \in G$, a word w over X of minimal length that represents g is called a *geodesic word* over X, and we denote the set of all such geodesic words by $\mathcal{G}(G, X)$. If w is an arbitrary word representing $g \in G$, then we write $|g|$ or $|w|_G$ (or $|g|_X$ or $|w|_{G,X}$ if X needs to be specified) for the length of a geodesic word over X that represents g. Similarly, we use $v = w$ to mean that the words v and w are identical as strings of symbols, and $v =_G w$ to mean that v and w represent the same element of the group.

We call a set of strings (i.e. a subset of A^*) a *language*; the study of languages is the topic of Chapter 2. It is convenient at this stage to introduce briefly the notation of a language for a group.

1.1.3 Languages for groups For a group G generated by X, we call a subset of $(X^{\pm})^*$ that contains at least one representative of each element in G a *language for* G; if the set contains precisely one representative of each element we call it a *normal form* for G. We shall be interested in finding *good* languages for a group G; clearly we shall need to decide what constitutes a good language. Typically we find good examples as the minimal representative words under a word order, such as word length or *shortlex*, $<_{\mathrm{slex}}$, defined below in 1.1.4. The *shortlex normal form* for a group selects the least representative of each group element under the shortlex ordering as its normal form word. The set $\mathcal{G}(G, X)$ of all geodesic words provides a natural language that is not in general a normal form.

1.1.4 Shortlex orderings *Shortlex* orderings (also known as *lenlex* orderings) of A^* arise frequently in this book. They are defined as follows. We start with any total ordering $<_A$ of A. Then, for $u, v \in A^*$, we define $u <_{\mathrm{slex}} v$ if either (i) $|u| < |v|$ or (ii) $|u| = |v|$ and u is less than v in the lexicographic (dictionary) ordering of strings induced by the chosen ordering $<_A$ of A.

More precisely, if $u = a_1 \cdots a_m$, $v = b_1 \cdots b_n$, then $u <_{\mathrm{slex}} v$ if either (i) $m < n$ or (ii) $m = n$ and, for some k with $1 \le k \le m$, we have $a_i = b_i$ for $i < k$ and $a_k <_A b_k$.

Note that $<_{\mathrm{slex}}$ is a well-ordering whenever $<_A$ is, which of course is the case when A is finite.

1.2 Generators, congruences and presentations

1.2.1 Generators If X is a subset of a semigroup S, monoid M or group G, then we define $\mathrm{Sgp}\langle X\rangle$, $\mathrm{Mon}\langle X\rangle$ or $\langle X\rangle$ to be the smallest subsemigroup, submonoid or subgroup of S, M or G that contains X. Then X is called a semigroup, monoid or group *generating set* if that substructure is equal to S, M or G respectively, and the elements of X are called *generators*.

We say that a semigroup, monoid or group is *finitely generated* if it possesses a finite generating set X.

1.2.2 Congruences If S is a semigroup and $\sim$ is an equivalence relation on S, then we say that $\sim$ is a *congruence* if

$$s_1 \sim s_2,\ t_1 \sim t_2 \implies s_1 t_1 \sim s_2 t_2.$$

We then define the semigroup $S/\!\sim$ to be the semigroup with elements the equivalence classes $[s] = \{t \in S : t \sim s\}$ of $\sim$, where $[s_1][s_2] = [s_1 s_2]$.

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

1.2.3 Presentations for semigroups, monoids and groups For a semigroup S generated by a set X, let $\mathcal{R} = \{(\alpha_i, \beta_i) : i \in I\}$ be a set of pairs of words from X^+ with $\alpha_i =_S \beta_i$ for each i. The elements of $\mathcal{R}$ are called *relations* of S. If $\sim$ is the smallest congruence on X^+ containing $\mathcal{R}$, and S is isomorphic to $X^+/\sim$, then we say that $\mathcal{R}$ is a *set of defining relations* for S, and that $\mathrm{Sgp}\langle X \mid \mathcal{R}\rangle$ is a *presentation* for S. In practice, we usually write $\alpha_i = \beta_i$ instead of (α_i, β_i). (This is an abuse of notation but the context should make it clear that we do not mean identity of words here.) Similarly the monoid presentation $\mathrm{Mon}\langle X \mid \mathcal{R}\rangle$ defines the monoid $X^*/\simeq$, for which $\simeq$ is the smallest congruence on X^* containing $\mathcal{R}$.

For groups the situation is marginally more complicated. If G is a group generated by a set X and $A = X^{\pm}$, then G is isomorphic to $A^*/\sim$, where $\sim$ is some congruence on A^* containing $\{(aa^{-1}, \varepsilon), (a^{-1}a, \varepsilon) : a \in X\}$. We define a *relator* of G to be a word $\alpha \in A^*$ with $\alpha =_G \varepsilon$. Let $R = \{\alpha_i : i \in I\}$ be a set of relators of G. If $\sim$ is the smallest congruence on A^* containing

$$\{(\alpha, \varepsilon) : \alpha \in R\} \cup \{(aa^{-1}, \varepsilon) : a \in X\} \cup \{(a^{-1}a, \varepsilon) : a \in X\},$$

and if G is isomorphic to $A^*/\sim$, then we say that R is a *set of defining relators* for G and that $\langle X \mid R\rangle$ is a *presentation* for G. Rather than specifying a relator α, so that α represents the identity, we can specify a *relation* $\beta = \gamma$ (as in the case of monoids or semigroups), which is equivalent to $\beta\gamma^{-1}$ being a relator.

We say that a semigroup, monoid or group is *finitely presented* (or, more accurately, *finitely presentable*) if it has a presentation in which the sets of generators and defining relations or relators are both finite.

1.2.4 Exercise Let $G = \langle X \mid R\rangle$ and let $A = X^{\pm}$. Show that

$$G \cong \mathrm{Mon}\langle A \mid \mathcal{I}_X \cup \mathcal{R}\rangle,$$

where $\mathcal{I}_X = \{(xx^{-1}, \varepsilon) : x \in X\} \cup \{(x^{-1}x, \varepsilon) : x \in X\}$ and $\mathcal{R} = \{(w, \varepsilon) : w \in R\}$.

1.2.5 Free semigroups, monoids and groups If S is a semigroup with presentation $\mathrm{Sgp}\langle X \mid \emptyset\rangle$ (which we usually write as $\mathrm{Sgp}\langle X \mid\rangle$), then we say that S is the *free semigroup* on X; we see that S is isomorphic to X^+ in this case. Similarly, if M is a monoid with presentation $\mathrm{Mon}\langle X \mid\rangle$, then we say that M is the *free monoid* on X, and we see that M is then isomorphic to X^*. If $S = X^+$ and $L \subseteq S$, then $\mathrm{Sgp}\langle L\rangle = L^+$; similarly, if $M = X^*$ and $L \subseteq M$, then $\mathrm{Mon}\langle L\rangle = L^*$.

If F is a group with a presentation $\langle X \mid\rangle$, then we say that F is the *free group* on X; if $|X| = k$, then we say that F is the free group of *rank* k (any two free groups of the same rank being isomorphic). We write $F(X)$ for the free group on X and F_k to denote a free group of rank k.

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

1.2.6 Exercise Let $G = \langle X \mid R \rangle$ be a presentation of a group G. Show that the above definition of G, which is essentially as a monoid presentation, agrees with the more familiar definition $\langle X \mid R \rangle = F(X)/\langle R^{F(X)} \rangle$, where $\langle R^{F(X)} \rangle$ denotes the normal closure of R in $F(X)$.

1.2.7 Reduced and cyclically reduced words In $F(X)$, the free group on X, every element has a unique representation of the form $w = x_1^{\epsilon_1} x_2^{\epsilon_2} \dots x_n^{\epsilon_n}$, where $n \geq 0$, $x_i \in X$ and $\epsilon_i \in \{1, -1\}$ for all i, and where we do not have both $x_i = x_{i+1}$ and $\epsilon_i = -\epsilon_{i+1}$ for any i; in this case, we say that the word w is *reduced*. Each word $v \in A^*$ is equal in $F(X)$ to a unique reduced word w.

If w is a reduced word and w is not of the form $x^{-1}vx$ or xvx^{-1} for some $x \in X$ and $v \in A^*$, then we say that w is *cyclically reduced*. Since replacing a defining relator by a conjugate in $F(X)$ does not change the group defined, we may (and often do) assume that all defining relators are cyclically reduced words.

1.3 Decision problems

In his two well-known papers in 1911 and 1912 [75, 76], Dehn defined and considered three decision problems in finitely generated groups, the word, conjugacy and isomorphism problems. While the word problem in groups is one of the main topics studied in this book, the other two will only be fleetingly considered. A good general reference on these and other decision problems in groups is the survey article by Miller [192].

1.3.1 The word problem A semigroup S is said to have *soluble word problem* if there exists an algorithm that, for any given words $\alpha, \beta \in X^+$, decides whether $\alpha =_S \beta$. The solubility of the word problem for a monoid or group generated by X is defined identically except that we consider words α, β in X^* or $(X^\pm)^*$. For groups, the problem is equivalent to deciding whether an input word is equal to the identity element. The word problem for groups is discussed further in Chapter 3 and in Part Three of this book. Examples of finitely presented semigroups and groups with insoluble word problem are described in Theorems 2.9.7 and 10.1.1.

1.3.2 The conjugacy and isomorphism problems The conjugacy problem in a semigroup S is to decide, given two elements $x, y \in S$, whether there exists $z \in S$ with $zx = yz$. Note that this relation is not necessarily symmetric in x and

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

y, but in a group G it is equivalent to deciding whether x and y are conjugate in G.

Since the word problem in a group is equivalent to deciding whether an element is conjugate to the identity, the conjugacy problem is at least as hard as the word problem, and there are examples of groups with soluble word problem but insoluble conjugacy problem. A number of such examples are described in the survey article by Miller [192], including Theorem 4.8 (an extension of one finitely generated free group by another), Theorem 4.11 (examples showing that having soluble conjugacy problem is not inherited by subgroups or overgroups of index 2), Theorem 5.4 (residually finite examples), Theorem 6.3 (a simple group), Theorem 7.7 (asynchronously automatic groups), and Theorem 7.8 (groups with finite complete rewriting systems) of that article.

The isomorphism problem is to decide whether two given groups, monoids or semigroups are isomorphic. Typically the input is defined by presentations, but could also be given in other ways, for example as groups of matrices. There are relatively few classes for which the isomorphism problem is known to be soluble. These classes include polycyclic and hyperbolic groups [232, 234, 72].

1.3.3 The generalised word problem Given a subgroup H of a group G, the generalised word problem is to decide, given $g \in G$, whether $g \in H$. So the word problem is the special case in which H is trivial. We shall encounter some situations in which this problem is soluble in Chapter 8. As for the conjugacy problem, the survey article [192] is an excellent source of examples (in particular in Theorems 5.4 and 7.8 of that article), in this case of groups with soluble word problem that have finitely generated subgroups with insoluble generalised word problem.

1.4 Subgroups and Schreier generators

Let H be a subgroup of a group $G = \langle X \rangle$, and let U be a right transversal of H in G. For $g \in G$, denote the unique element of $Hg \cap U$ by $\overline{g}$. Define

$$Z := \left\{ux\overline{ux}^{-1} : u \in U, x \in X\right\}.$$

Then $Z \subseteq H$.

1.4.1 Theorem *With the above notation, we have $H = \langle Z \rangle$.*

Our proof needs the following result.

1.4.2 Lemma *Let $S = \{ux^{-1}\overline{ux^{-1}}^{-1} : u \in U, x \in X\}$. Then $Z^{-1} = S$.*

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

Proof Let $g \in Z^{-1}$, so $g = (ux\overline{ux}^{-1})^{-1} = \overline{ux}x^{-1}u^{-1}$. Let $v := \overline{ux} \in U$. Then, since the elements vx^{-1} and u are in the same coset of H, we have $\overline{vx^{-1}} = u$, and $g = vx^{-1}\overline{vx^{-1}}^{-1} \in S$.

Conversely, let $g = ux^{-1}\overline{ux^{-1}}^{-1} \in S$, so $g^{-1} = \overline{ux^{-1}}xu^{-1}$. Let $v := \overline{ux^{-1}}$. Then $\overline{vx} = u$, so $g^{-1} = vx\overline{vx}^{-1} \in Z$ and $g \in Z^{-1}$. □

Proof of Theorem 1.4.1 Let $U \cap H = \{u_0\}$. (We usually choose $u_0 = 1$, but this is not essential.) Let $h \in H$. Then we can write $u_0^{-1}hu_0 = a_1 \cdots a_l$ for some $a_i \in A := X^{\pm}$. For $1 \leq i \leq l$, let $u_i := \overline{a_1 \cdots a_l}$. Since $u_0^{-1}hu_0 \in H$, we have $u_l = u_0$. Then

$$h =_G (u_0a_1u_1^{-1})(u_1a_2u_2^{-1}) \cdots (u_{l-1}a_lu_l^{-1}).$$

Note that $u_{i+1} = \overline{a_1 \cdots a_{l+1}}$ is in the same coset of H as u_ia_{i+1}, so $\overline{u_ia_{i+1}} = u_{i+1}$, and

$$h =_G (u_0a_1\overline{u_0a_1}^{-1})(u_1a_2\overline{u_1a_2}^{-1}) \cdots (u_{l-1}a_l\overline{u_{l-1}a_l}^{-1}). \qquad (\dagger)$$

Each bracketed term is in Z if $a_i \in X$, and in Z^{-1} if $a_i \in X^{-1}$ by Lemma 1.4.2. So $H = \langle Z \rangle$. □

1.4.3 Corollary *A subgroup of finite index in a finitely generated group is finitely generated.*

1.4.4 Rewriting The process described in the above proof of calculating a word v over Z from a word w over X that represents an element of H is called *Reidemeister–Schreier rewriting*. We may clearly omit the identity element from the rewritten word, which results in a word over $Y = Z \setminus \{1\}$, which we denote by $\rho_{X,Y}(w)$. From the proof, we see immediately that:

1.4.5 Remark If $1 \in U$, then $|\rho_{X,Y}(w)| \leq |w|$.

1.4.6 Schreier generators and transversals The above set Y of non-identity elements of Z is called the set of *Schreier generators* of H in G. Of course, this set depends on X and on U.

The set U is called a *Schreier transversal* if there is a set of words over X representing the elements of U that is closed under taking prefixes. Note that such a set must contain the empty word, and hence $1 \in U$. By choosing the least word in each coset under some *reduction ordering* of A^* (where $A = X^{\pm}$), it can be shown that Schreier transversals always exist. Reduction orderings are defined in 4.1.5. They include the shortlex orderings defined in 1.1.4.

It was proved by Schreier [228] that, if G is a free group and U is a Schreier transversal, then the Schreier generators freely generate H.

The following result, known as the *Reidemeister–Schreier Theorem*, which we shall not prove here, provides a method of computing a presentation of the subgroup H from a presentation of the group G. Note that it immediately implies the celebrated *Nielsen–Schreier Theorem*, that any subgroup of a free group is free. As with many of the results stated in this chapter, we refer the reader to the standard textbook on combinatorial group theory by Lyndon and Schupp [183] for the proof.

1.4.7 Theorem (Reidemeister–Schreier Theorem [183, Proposition II.4.1]) *Let $G = \langle X \mid R\rangle = F/N$ be a group presentation, where $F = F(X)$ is the free group on X, and let $H = E/N \le G$. Let U be a Schreier transversal of E in F and let Y be the associated set of Schreier generators. Then $\langle Y \mid S\rangle$ with $S = \left\{\rho_{X,Y}(uru^{-1}) : u \in U, r \in R\right\}$ is a presentation of H.*

1.4.8 Corollary *A subgroup of finite index in a finitely presented group is finitely presented.*

1.5 Combining groups

In this section we introduce various constructions that combine groups. We leave the details of many of the proofs of stated results to the reader, who is referred to [183, Chapter IV] for details.

1.5.1 Free products Informally, the *free product* $G * H$ of the groups G, H is the largest group that contains G and H as subgroups and is generated by G and H. Formally, it can be defined by its universal property:

(i) there are homomorphisms $\iota_G\colon G \to G * H$ and $\iota_H\colon H \to G * H$;
(ii) if K is any group and $\tau_G\colon G \to K$, $\tau_H\colon H \to K$ are homomorphisms, then there is a unique homomorphism $\alpha\colon G * H \to K$ with $\alpha\iota_G = \tau_G$ and $\alpha\iota_H = \tau_H$.

As is often the case with such definitions, it is straightforward to prove uniqueness, in the sense that any two free products of G and H are isomorphic, and it is not hard to show that $G * H$ is generated by $\iota_G(G)$ and $\iota_H(H)$. But the existence of the free product is not immediately clear.

To prove existence, let $G = \langle X \mid R\rangle$ and $H = \langle Y \mid S\rangle$ be presentations of G and H. Then we can take

$$G * H = \langle X \cup Y \mid R \cup S\rangle,$$

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

where ι_G and ι_H are the homomorphisms induced by the embeddings $X \to X \cup Y$ and $Y \to X \cup Y$; we tacitly assumed that X and Y are disjoint.

It is not completely obvious that ι_G and ι_H are monomorphisms. This follows from another equivalent description of $G * H$ as the set of alternating products of arbitrary length (including length 0) of non-trivial elements of G and H, with multiplication defined by concatenation and multiplications within G and H. With this description, ι_G and ι_H are the obvious embeddings, and G and H are visibly subgroups of $G * H$, known as the *free factors* of $G * H$. The equivalence of the two descriptions follows immediately in a more general context from Proposition 1.5.12.

The definition extends easily to the free product of any family of groups. The following result, which we shall not prove here, is used in the proof of the special case of the Muller–Schupp Theorem (Theorem 11.1.1) that torsion-free groups with context-free word problem are virtually free.

1.5.2 Theorem (Grushko's Theorem [183, IV.1.9]) *For a group G, let $d(G)$ denote the minimal number of generators of G. Then $d(G * H) = d(G) + d(H)$.*

1.5.3 Direct products The *direct product* $G \times H$ of two groups G, H is usually defined as the set $G \times H$ with component-wise multiplication. We generally identify G and H with the component subgroups, which commute with each other, and are called the *direct factors* of $G \times H$. Then each element has a unique representation as a product of elements of G and H. It can also be defined by a universal property:

(i) there are homomorphisms $\pi_G \colon G \times H \to G$ and $\pi_H \colon G \times H \to H$;
(ii) if K is any group and $\tau_G \colon K \to G$ and $\tau_H \colon K \to H$ are homomorphisms, then there is a unique homomorphism $\varphi \colon K \to G \times H$ with $\tau_G = \pi_G \circ \varphi$ and $\tau_H = \pi_H \circ \varphi$.

If $G = \langle X \mid R \rangle$ and $H = \langle Y \mid S \rangle$ are presentations, then $G \times H$ has the presentation

$$G \times H = \langle X \cup Y \mid R \cup S \cup \{[x, y] : x \in X, y \in Y\} \rangle.$$

We can extend this definition to direct products of families of groups as follows. Let $\{G_\omega : \omega \in \Omega\}$ be a family of groups. Then the *(full) direct product*, also known sometimes as the *Cartesian product*, $\prod_{\omega\in\Omega} G_\omega$ of the family consists of the set of functions $\beta \colon \Omega \to \cup_{\omega\in\Omega} G_\omega$ for which $\beta(\omega) \in G_\omega$ for all $\omega \in \Omega$, where the group operation is component-wise multiplication in each G_ω; that is, $\beta_1\beta_2(\omega) = \beta_1(\omega)\beta_2(\omega)$ for all $\omega \in \Omega$.

The elements of $\prod_{\omega\in\Omega} G_\omega$ consisting of the functions β with finite support (i.e. $\beta(\omega) = 1_G$ for all but finitely many $\omega \in \Omega$) form a normal subgroup of $\prod_{\omega\in\Omega} G_\omega$. We call this subgroup the *restricted direct product* of the family $\{G_\omega : \omega \in \Omega\}$. It is also sometimes called the *direct sum* of the family to distinguish it from the direct product.

1.5.4 Semidirect products Let N and H be groups, and let $\phi : H \to \mathrm{Aut}(N)$ be a right action of H on N. We define the *semidirect product* of H and N, written $N \rtimes_\phi H$ or just $N \rtimes H$, to be the set $\{(n,h) : n \in N, h \in H\}$ equipped with the product

$$(n_1, h_1)(n_2, h_2) = (n_1 n_2^{\phi(h_1^{-1})}, h_1 h_2).$$

We leave it as an exercise to the reader to derive a presentation of $N \rtimes_\phi H$ from presentations of H and N and the action ϕ. We note that sometimes the notation $H \ltimes N$ is used for the same product. We identify the subgroups $\{(n, 1) : n \in N\}$ and $\{(1, h) : h \in H\}$ with N and H, and hence (n, h) with nh, so that the expression above reads

$$n_1 h_1 n_2 h_2 = n_1 n_2^{\phi(h_1^{-1})} h_1 h_2.$$

The direct product $N \times H$ is the special case when ϕ is the trivial action. The semidirect product is itself a special case of a *group extension*, which is a group G with normal subgroup N and $G/N \cong H$. Unfortunately roughly half of the set of mathematicians refer to this as an extension of N by H, and the other half call it an extension of H by N. An extension is isomorphic to a semidirect product if and only if N has a complement in G (that is, G has a subgroup K, with $N \cap K = \{e\}$, $G = NK$), in which case it is also called a *split extension*.

Note that we can also define a semidirect product of two groups N and H, from a left action of H on N.

1.5.5 Wreath products Let G and H be groups and suppose that we are given a right action $\phi \colon H \to \mathrm{Sym}(\Omega)$ of H on the set Ω. We define the associated *(full) permutational wreath product* $G \wr H = G \wr_\phi H$ as follows.

Let $N = \prod_{\omega\in\Omega} G_\omega$, where the groups G_ω are all equal to the same group G. So the elements of N are functions $\gamma \colon \Omega \to G$. We define a right action $\psi \colon H \to \mathrm{Aut}(N)$ by putting $\gamma^{\psi(h)}(\omega) = \gamma(\omega^{\phi(h^{-1})})$ for each $\gamma \in N$, $h \in H$, and $\omega \in \Omega$. We then define $G \wr_\phi H$ to be the semidirect product $N \rtimes_\psi H$. So the elements have the form (γ, h) with $\gamma \in N$ and $h \in H$. As in 1.5.4, we identify $\{(\gamma, 1) : \gamma \in N\}, \{(1, h) : h \in H\}$ with N and H, and hence (γ, h) with the product γh.

If we restrict elements of N to the functions $\gamma \colon \Omega \to G$ with finite support, then we get the *restricted wreath product*, which we shall write as $G \wr_R H$ or $G \wr_{R\phi} H$.

The special case in which ϕ is the right regular action of H (i.e. $\Omega = H$ and $h_1^{\phi(h_2)} = h_1h_2$ for $h_1, h_2 \in H$) is known as the *standard* or *restricted standard* wreath product. This is the default meaning of $G \wr H$ or $G \wr_R H$ when the action ϕ is not specified.

Finally we mention that, if we are given a right action $\rho\colon G \to \mathrm{Sym}(\Delta)$, then we can define an action $\psi\colon G \wr_\phi H \to \mathrm{Sym}(\Delta \times \Omega)$ by setting $(\delta, \omega)^{\psi(\gamma,h)} = (\delta^{\rho\circ\gamma(\omega)}, \omega^{\phi(h)})$. This right action plays a central role in the study of imprimitive permutation groups, but it will not feature much in this book.

1.5.6 Exercise Show that the restricted standard or permutational wreath product $G \wr_R H$ is finitely generated if both G and H are finitely generated. Verify also that $G \wr H$ is not finitely generated unless H is finite and G is finitely generated.

1.5.7 Graph products Let Γ be a simple undirected graph with vertices labelled from a set I, and let G_i $(i \in I)$ be groups. Then the *graph product* of the G_i with respect to Γ can be thought of as the largest group G generated by the G_i such that $[G_i, G_j] = 1$ whenever $\{i, j\}$ is in the set $E(\Gamma)$ of edges of Γ.

If $\langle X_i \mid R_i\rangle$ is a presentation of G_i for each i, then

$$\langle \cup_{i\in I} X_i \mid \cup_{i\in I} R_i \cup \{[x_i, x_j] : x_i \in X_i, x_j \in X_j, i, j \in I, \{i, j\} \in E(\Gamma)\}\rangle$$

is a presentation of the graph product.

Note that the right-angled Artin groups (see 1.10.4) can be described equivalently as graph products of copies of $\mathbb{Z}$.

1.5.8 Free products with amalgamation The amalgamated free product generalises the free product. Suppose that G and H are groups with subgroups $A \le G$, $B \le H$, and that there is an isomorphism $\phi\colon A \to B$.

Informally, the free product $G *_A H$ of G and H amalgamated over A (via ϕ) is the largest group P with $G, H \le P$, $\langle G, H\rangle = P$, and $a = \phi(a)$ for all $a \in A$.

1.5.9 Example Suppose that $\Gamma = \langle G, H\rangle$ and $G \cap H = A$, where A is a subgroup of both G and H with $|G : A| \ge 3$ and $|H : A| \ge 2$. Suppose also that Γ acts on the left on a set Ω and that Ω_1, Ω_2 are subsets of Ω with $\Omega_1 \not\subseteq \Omega_2$, such that

(1) $(G \setminus A)(\Omega_1) \subseteq \Omega_2$ and $(H \setminus A)(\Omega_2) \subseteq \Omega_1$;
(2) $A(\Omega_i) \subseteq \Omega_i$ for $i = 1, 2$.

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

Then $\Gamma \cong G *_A H$.

This result is often known as the *ping-pong lemma*. It is proved in [74, IIB.24] for the case $A = 1$ but essentially the same proof works for general A. The reader could attempt it as an exercise, using Corollary 1.5.13 below.

1.5.10 Exercise Let $\Gamma = \mathrm{SL}(2,\mathbb{Z}) = \langle x, y\rangle$ with

$$x = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Let $G = \langle y\rangle$, $H = \langle x\rangle$, and $A = G \cap H = \langle y^3\rangle = \langle x^2\rangle$. Show that $\Gamma \cong G *_A H$ by taking $\Omega = \mathbb{Z}^2$, $\Omega_1 = \{(x,y) \in \Omega : xy < 0\}$ and $\Omega_2 = \{(x,y) \in \Omega : xy > 0\}$.

1.5.11 Example As a consequence of the Seifert–van Kampen Theorem [189, Chapter 4, Theorem 2.1], we see that, for a topological space $X = Y \cup Z$ for which Y, Z and $Y \cap Z$ are open and path-connected, and the fundamental group $\pi_1(Y \cap Z)$ embeds naturally into $\pi_1(Y)$ and $\pi_1(Z)$, the fundamental group of X is isomorphic to the free product with amalgamation $\pi_1(Y) *_{\pi_1(Y\cap Z)} \pi_1(Z)$ (see [183, IV.2]).

Formally, $G *_A H$ can be defined by the following universal property:

(i) there are homomorphisms $\iota_G\colon G \to G *_A H$ and $\iota_H\colon H \to G *_A H$ with $\iota_G(a) = \iota_H(\phi(a))$ for all $a \in A$;
(ii) if K is any group and $\tau_G\colon G \to K$, $\tau_H\colon H \to K$ are homomorphisms with $\tau_G(a) = \tau_H(\phi(a))$ for all $a \in A$, then there is a unique homomorphism $\alpha\colon G *_A H \to K$ with $\alpha\iota_G = \tau_G$ and $\alpha\iota_H = \tau_H$.

The uniqueness of $G *_A H$ up to isomorphism follows easily, but not its existence, which is most conveniently established using presentations, as follows. Let $G = \langle X \mid R\rangle$ and $H = \langle Y \mid S\rangle$ be presentations of G and H. For each element of $a \in A$, let w_a and v_a be words over X and Y representing a and $\phi(a)$, respectively, and put $T := \{w_a = v_a : a \in A\}$. Then, as in [183, IV.2], we define

$$G *_A H := \langle X \cup Y \mid R \cup S \cup T\rangle,$$

and it is straightforward to show, using standard properties of group presentations, that $G *_A H$ has the above universal property, where ι_G and ι_H are defined to map words in G and in H to the same words in $G *_A H$.

Note that, in the definition of T, it would be sufficient to restrict a to the elements of a generating set of A so, if G and H are finitely presented and A is finitely generated, then $G *_A H$ is finitely presentable.

But we have still not proved that G and H are subgroups of $G *_A H$; that is, that ι_1 and ι_2 are embeddings. We do that by finding a normal form for the

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

elements of $G *_A H$. Let U and V be left transversals of A in G and $B \in H$, respectively, with $1_G \in U$, $1_H \in V$. From now on, we shall suppress the maps ι_G, ι_H and just write g rather than $\iota_G(g)$.

1.5.12 Proposition *Every element of $G *_A H$ has a unique expression as $t_1 \cdots t_k a$ for some $k \geq 0$, where $a \in A$, $t_i \in (U \setminus \{1_G\}) \cup (V \setminus \{1_H\})$ for $1 \leq i \leq k$ and, for $i < k$, $t_i \in U \Leftrightarrow t_{i+1} \in V$.*

In particular, since distinct elements of G and of H give rise to distinct expressions of this form, G and H embed into $G *_A H$ as subgroups, and $G \cap H = A = \phi(A)$.

Proof By definition, each $f \in G *_A H$ can be written as an alternating product of elements of G and H, and working from the left and writing each such element as a product of a coset representative and an element of A (which has been identified with $\phi(A) = B$), we can write f in the specified normal form.

Let Ω be the set of all normal form words. We define a right action of $G *_A H$ on Ω, which corresponds to multiplication on the right by elements of $G *_A H$. To do this, it is sufficient to specify the actions of G and of H, which must of course agree on the amalgamated subgroup.

Let $\alpha = t_1 \cdots t_k a \in \Omega$ and $g \in G$. If $k = 0$ or $t_k \in V$, then we define $\alpha^g = t_1 \cdots t_k t_{k+1} a'$, where $t_{k+1} a' =_G ag$. Otherwise, $k > 0$ and $t_k \in U$, and we put $\alpha^g = t_1 \cdots t_{k-1} t_{k+1} a'$, where $t_{k+1} a' =_G t_k ag$. In both cases, $t_{k+1} \in U$, $a' \in A$ and we omit t_{k+1} if it is equal to 1. We define the action of H on Ω similarly.

It is easy to see that these definitions do indeed define actions of G and H on Ω that agree on the amalgamated subgroup A, so we can use them to define the required action of $G *_A H$ on Ω. This follows from the universal property of $G *_A H$. It is also clear from the definition that, taking $\alpha = \varepsilon \in \Omega$, and f to be the element of $G *_A H$ defined by the normal form word $t_1 \cdots t_{k-1} t_k a$, we have $\alpha^f = t_1 \cdots t_{k-1} t_k a$. So the elements of G represented by distinct normal form words have distinct actions on Ω, and hence they cannot represent the same element of $G *_A H$. $\square$

1.5.13 Corollary *Suppose that $f = f_1 f_2 \cdots f_k \in G *_A H$ with $k > 0$, where $f_i \in (G \setminus A) \cup (H \setminus B)$ for $1 \leq i \leq k$ and, for $i < k$, $f_i \in G \Leftrightarrow f_{i+1} \in H$. Then f is not equal to the identity in $G *_A H$.*

Conversely, suppose that the group F is generated by subgroups (isomorphic to) G and H with $G \cap H = A$, where $a =_F \phi(a)$ for all $a \in A$, and that $f \neq 1$ for every element $f = f_1 f_2 \cdots f_k \in F$ with $k > 0$, $f_i \in (G \setminus A) \cup (H \setminus B)$ for $1 \leq i \leq k$ and, for $i < k$, $f_i \in G \Leftrightarrow f_{i+1} \in H$. Then $F \cong G *_A H$.

Proof The assumptions ensure that, when we put f into normal form as described in the above proof, the resulting expression has the form $t_1 \cdots t_k a$ with

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

the same k and, since we are assuming that $k > 0$, this is not the representative ε of the identity element.

For the converse, observe that the hypothesis implies that the normal form expressions for elements of F described in Proposition 1.5.12 represent distinct elements of F, and so the map $\alpha\colon G*_A H \to F$ specified by (ii) of the definition of the $G *_A H$ is an isomorphism. □

A product $f = f_1 f_2 \cdots f_k$ as in the above corollary is called a *reduced form* for f. It is called *cyclically reduced* if all of its cyclic permutations are reduced forms. Every element of $G *_A H$ is conjugate to an element $u = f_1 \cdots f_n$ in cyclically reduced form, and every cyclically reduced conjugate of u can be obtained by cyclically permuting $f_1 \cdots f_n$ and then conjugating by an element of the amalgamated subgroup A [183, page 187].

If $f_1 f_2 \cdots f_k$ is a reduced form with $k > 1$ and $(f_1 f_2 \cdots f_k)^n$ is not a reduced form for some $n > 0$, then $f_k f_1$ cannot be reduced, and so $f_1 f_2 \cdots f_k$ cannot be cyclically reduced. This proves the following result.

1.5.14 Corollary [183, Proposition 12.4] *An element of finite order in $G*_A H$ is conjugate to an element of G or to an element of H.*

1.5.15 HNN-extensions Suppose now that A and B are both subgroups of the same group G, and that there is an isomorphism $\phi\colon A \to B$. The corresponding HNN-*extension*, (due to Higman, Neumann and Neumann [141]) with *stable letter* t, *base group* G and *associated subgroups* A and B, is roughly the largest group $G*_{A,t}$ that contains G as a subgroup, and is generated by G and an extra generator t such that $t^{-1}at = \phi(a)$ for all $a \in A$.

1.5.16 Example Analogously to Example 1.5.11, suppose that we have a path-connected topological space Y with two homeomorphic open subspaces U and V, of which the fundamental groups embed into that of Y, and suppose that we form a new space X by adding a handle that joins U to V using the homeomorphism between them. Then it is a consequence of the Seifert–van Kampen Theorem that the fundamental group $\pi_1(X)$ of X is isomorphic to the HNN-extension $\pi_1(Y) *_{\pi_1(U),t}$; see [183, IV.2].

We can also define an HNN-extension of $G = \langle X \mid R\rangle$ via a presentation. Again, for $a \in A$, we let w_a and v_a be words over X representing a and $\phi(a)$, respectively, and define

$$G*_{A,t} := \langle X, t \mid R \cup T\rangle, \quad \text{where} \quad T := \{t^{-1}w_a t = v_a : a \in A\}.$$

Again, we can restrict the elements a in T to a generating set of A, so $G*_{A,t}$ is finitely presentable if G is finitely presented and A is finitely generated.

There is a homomorphism $\iota\colon G \to G*_{A,t}$ that maps each word over X to the same word in $G*_{A,t}$. Once again, we can use a normal form to prove that ι embeds G into $G*_{A,t}$, and we shall henceforth suppress ι and write g rather than $\iota(g)$. Let U and V be left transversals of A and B in G, respectively, with $1_G \in U$, $1_H \in V$.

1.5.17 Proposition *Every element of $G*_{A,t}$ has a unique expression as $t^{j_0}g_1t^{j_1}g_2\cdots g_kt^{j_k}g_{k+1}$ for some $k \geq 0$, where*

(i) *$g_i \in G$ for $1 \leq i \leq k+1$ and $g_i \neq 1$ for $1 \leq i \leq k$;*
(ii) *$j_i \in \mathbb{Z}$ for $0 \leq i \leq k$ and $j_i \neq 0$ for $1 \leq i \leq k$;*
(iii) *for $1 \leq i \leq k$, we have $g_i \in U$ if $j_i > 0$, and $g_i \in V$ if $j_i < 0$.*

In particular, since distinct elements of G give rise to distinct expressions of this form with $k = 0$ and $j_0 = 0$, G embeds as a subgroup of $G*_{A,t}$.

Proof Clearly each $f \in G*_{A,t}$ can be written as $t^{j_0}g_1t^{j_1}g_2\cdots g_kt^{j_r}g_{k+1}$ for some $k \geq 0$ such that (i) and (ii) are satisfied. If $j_1 > 0$, then we write g_1 as $g_1'a$ with $g_1' \in U$ and $a \in A$ and, using the relation $t^{-1}at = \phi(a)$, replace at in the word by $t\phi(a)$. Similarly, if $j_1 < 0$, then we write g_1 as $g_1'b$ with $g_1' \in V$ and $b \in B$, and replace bt^{-1} in the word by $t^{-1}\phi^{-1}(b)$. By working through the word from left to right making these substitutions, we can bring f into the required normal form (i.e. satisfying (i), (ii) and (iii)).

Let Ω be the set of all normal form words. We define a right action of $G*_{A,t}$ on Ω, which corresponds to multiplication on the right by elements of $G*_{A,t}$. To do this, it is sufficient to specify the actions of G and of t provided that, for each $a \in A$ the action of a followed by that of t is the same as the action of t followed by that of $\phi(a)$. In fact it is more convenient to specify the actions of t and t^{-1} separately and then check that they define inverse mappings. Let $\alpha = t^{j_0}g_1t^{j_1}g_2\cdots g_kt^{j_k}g_{k+1} \in \Omega$.

If $g \in G$, then we define $\alpha^g := t^{j_0}g_1t^{j_1}g_2\cdots g_kt^{j_k}g_{k+1}'$, where $g_{k+1}' = g_{k+1}g$. We need to subdivide into three cases for the action of t.

(a) If $g_{k+1} \notin A$, then we write $g_{k+1} = g_{k+1}'a$ with $1 \neq g_{k+1}' \in U$ and $a \in A$, and $\alpha^t := t^{j_0}g_1t^{j_1}g_2\cdots g_kt^{j_k}g_{k+1}'t\phi(a)$;
(b) If $g_{k+1} \in A$ and $j_k \neq -1$, then $\alpha^t := t^{j_0}g_1t^{j_1}g_2\cdots g_kt^{j_k+1}\phi(g_{k+1})$;
(c) If $g_{k+1} \in A$ and $j_k = -1$, then $\alpha^t := t^{j_0}g_1t^{j_1}g_2\cdots t^{j_{k-1}}g_k'$, where $g_k' = g_k\phi(g_{k+1})$ (and $g_k = 1$ if $k = 0$).

We have the corresponding three cases for the action of t^{-1}.

(a) If $g_{k+1} \notin B$, then we write $g_{k+1} = g_{k+1}'b$ with $1 \neq g_{k+1}' \in V$ and $b \in B$, and $\alpha^{t^{-1}} := t^{j_0}g_1t^{j_1}g_2\cdots g_kt^{j_k}g_{k+1}'t^{-1}\phi^{-1}(b)$;

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

(b) If $g_{k+1} \in B$ and $j_k \neq 1$, then $\alpha^{t^{-1}} := t^{j_0} g_1 t^{j_1} g_2 \cdots g_k t^{j_k - 1} \phi^{-1}(g_{k+1})$;
(c) If $g_{k+1} \in B$ and $j_k = 1$, then $\alpha^{t^{-1}} := t^{j_0} g_1 t^{j_1} g_2 \cdots t^{j_{k-1}} g'_k$, where $g'_k = g_k \phi^{-1}(g_{k+1})$ (and $g_k = 1$ if $k = 0$).

We leave it to the reader to verify that the action of $t^{\pm 1}$ followed by that of $t^{\mp 1}$ is the identity map on Ω and that, for each $a \in A$, the action of a followed by that of t is the same as the action of t followed by that of $\phi(a)$. This shows that we do indeed have an action of $G{*}_{A,t}$ on Ω.

As with the corresponding proof for free products with amalgamation, taking $\alpha = \varepsilon \in \Omega$, and f to be the element of $G{*}_{A,t}$ defined by the normal form word $t^{j_0} g_1 t^{j_1} g_2 \cdots g_k t^{j_k} g_{k+1}$, we have $\alpha^f = t^{j_0} g_1 t^{j_1} g_2 \cdots g_k t^{j_k} g_{k+1}$. So the elements of G represented by distinct normal form words have distinct actions on Ω, and hence they cannot represent the same element of $G{*}_{A,t}$. □

The following corollary is known as *Britton's Lemma*, and is used in many arguments involving HNN-extensions. It will play a crucial role in the construction of groups with insoluble word problems in Chapter 10.

1.5.18 Corollary (Britton's Lemma (1963)) *Let $f = g_1 t^{j_1} g_2 \cdots g_k t^{j_r} g_{k+1}$ for some $k \geq 1$, where $g_i \in G$ for $1 \leq i \leq k+1$, $g_i \neq 1$ for $2 \leq i \leq k$, and each $j_i \in \mathbb{Z} \setminus \{0\}$. Suppose also that there is no subword in this expression of the form $t^{-1} g_i t$ with $g_i \in A$, or $t g_i t^{-1}$ with $g_i \in B$. Then f is not equal to the identity in $G{*}_{A,t}$.*

Proof The assumptions ensure that, when we put f into normal form as described in the above proof, there is no cancellation between t and t^{-1}, so the resulting expression is nonempty. □

A product $f = g_1 t^{j_1} g_2 \cdots g_k t^{j_r} g_{k+1}$ as in the above corollary is called a *reduced form* for f [183, pages 181–186]. A subword of the form $t^{-1} g_i t$ with $g_i \in A$, or $t g_i t^{-1}$ with $g_i \in B$ is known as a *pinch*, so a reduced word contains no pinches.

It is easily seen that the presentation for the surface group

$$\langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1][a_2, b_2] \ldots [a_g, b_g] \rangle$$

can be rewritten as

$$\langle a_1, b_1, \ldots, a_g, b_g \mid a_1 b_1 a_1^{-1} = [b_g, a_g] \ldots [b_2, a_2] b_1 \rangle,$$

and hence this group can be expressed as an HNN-extension of the free group $F(b_1, a_2, b_2, \ldots, a_g, b_g)$. We can find similar decompositions for the other surface groups.

One of the earliest applications of HNN-extensions was the following result of Higman, Neumann and Neumann [141].

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

1.5.19 Theorem *Every countable group can be embedded in a 2-generator group.*

Sketch of proof (see [223] for details) Let $G = \{1 = g_0, g_1, g_2, \ldots\}$ be a countable group and $H = G * F$ with F free on $\{x, y\}$. Then the subgroups $A = \langle x, g_i y^{-i} x y^i \ (i \geq 1)\rangle$ and $B = \langle y, x^{-i} y x^i \ (i \geq 1)\rangle$ of G are both freely generated by their generators, so there is an HNN-extension K of H in which the stable letter t conjugates x to y, and $g_i y^{-i} x y^i$ to $x^{-i} y x^i$ for all $i \geq 1$. The 2-generator subgroup $L = \langle t, x\rangle$ of K contains g_i for all i, which generate a subgroup isomorphic to G. □

Another application from [141] is the construction of torsion free groups in which all non-identity elements are conjugate.

1.5.20 Multiple HNN-extensions More generally, suppose that the group G has subgroups A_i and B_i with $i \in I$ for some indexing set I, with isomorphisms $\phi_i \colon A_i \to B_i$. Then we can define a corresponding multiple HNN-extension with stable letters t_i and relations $t_i^{-1} a_i t_i = \phi_i(a_i)$ for all $i \in I$, $a_i \in A_i$. We generally refer to this just as an HNN-extension, but with multiple stable letters. By ordering I, it can be regarded as an ascending chain of standard HNN-extensions. Groups of this type will arise in the the construction of groups with insoluble word problems in Chapter 10.

1.5.21 Graphs of groups The following construction, due to Bass and Serre, is a generalisation of both HNN-extensions and free products with amalgamation. Let Γ be a finite graph, possibly with loops at some vertices. For convenience we assign an orientation to each edge. Then to each vertex $v \in V(\Gamma)$ we assign a group G_v, and to each edge $e \in E(\Gamma)$ a group G_e, together with monomorphisms

$$\phi_{e,\iota(e)} \colon G_e \to G_{\iota(e)} \quad \text{and} \quad \phi_{e,\tau(e)} \colon G_e \to G_{\tau(e)}$$

of each edge group into the vertex groups corresponding to its initial and terminal vertices.

We call the system $\mathcal{G}$, consisting of the graph together with its associated groups and monomorphisms, a *graph of groups*.

Now suppose that T is a spanning tree of Γ. We define the *fundamental group* of $\mathcal{G}$ with respect to T, written $\pi_1(\mathcal{G}, T)$, to be the group generated by all of the vertex subgroups G_v ($v \in V(\Gamma)$) together with generators y_e, one for each oriented edge e, and subject to the conditions that

(1) $y_e \phi_{e,\iota(e)}(x) y_e^{-1} = \phi_{e,\tau(e)}(x)$ for each edge e and each $x \in G_e$,

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

(2) $y_e = 1$ for each edge e in T.

It turns out that the resulting group is independent of the choice of spanning tree, and hence we call it the *fundamental group of* $\mathcal{G}$, written $\pi_1(\mathcal{G})$. The groups G_v and G_e embed naturally as subgroups.

It can be seen from the definition that $\pi_1(\mathcal{G})$ can be constructed by first taking free products of the groups G_v with subgroups amalgamated along the edges of the spanning tree, and then taking HNN-extensions of the result, with one extra generator for each remaining edge. In particular, when Γ is a graph with 2 vertices v, w and a single edge e, we get the free product with amalgamation $G_v *_{G_e} G_w$; and when Γ is a graph with a single vertex v and a loop e on v, we get the HNN-extension $G_v *_{G_e,t}$.

We can construct an infinite graph on which the fundamental group acts. The vertex set consists of the union of all of the sets of left cosets in $\pi(\mathcal{G})$ of the subgroups G_v for $v \in V(\Gamma)$, while the edge set consists of the union of all of the sets of left cosets in $\pi(\mathcal{G})$ of the subgroups G_e, and incidence between vertices and edges is defined by inclusion. This graph turns out to be a tree, which is called the *Bass–Serre tree*. The fundamental group $\pi_1(\mathcal{G})$ acts naturally as a group of automorphisms on the tree, and the vertices of the original graph Γ form a fundamental domain.

1.6 Cayley graphs

Here we define the Cayley graph and collect some basic results. A metric space (Γ, d) is called *geodesic* if for all $x, y \in \Gamma$ there is an isometry (distance preserving bijection) $f\colon [0, l] \to \Gamma$ with $f(0) = x$ and $f(l) = y$, where $l = d(x, y)$. The image of such a map f is called a *geodesic path* from x to y, often written $[x, y]$ (and hopefully never confused with a commutator). A familiar example is $\mathbb{R}^n$ with the Euclidean metric.

1.6.1 Cayley graphs Let G be a group with finite generating set X. Then we define the Cayley graph $\Gamma(G, X)$ of G over X to be the graph with vertex set G and with a directed edge labelled by x leading from g to gx, for each $g \in G$ and each $x \in X^{\pm}$. The vertex labelled 1 is often referred to as the *base point* or the *origin* of $\Gamma(G, X)$. It is customary (and convenient, particularly when drawing pictures of the Cayley graph), to regard the edges labelled x and x^{-1} that connect the same two vertices as a single edge with two orientations. If x has order 2 then the edge is shown undirected. More generally, one can define the Cayley graph $\Gamma(G, A)$ for an arbitrary subset A of G that generates G as a

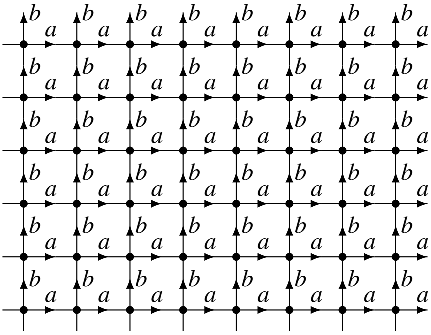


Figure 1.1 Cayley graph of $\mathbb{Z}^2 = \langle a, b \mid ab = ba \rangle$

monoid, but we will restrict to inverse-closed monoid generating sets in this book. Some examples of Cayley graphs are depicted in Figures 1.1–1.3.

1.6.2 Paths and geodesics The graph metric in which each edge has length 1 makes $\Gamma(G, X)$ into a geodesic metric space, on which G acts vertex transitively on the left as a group of isometries.

Each word w over X labels a distinct path of length $|w|$ in Γ from each vertex g to gw. We denote this path by ${}_g w$, or sometimes just by w when $g = 1$. A word w is geodesic if and only if it labels a geodesic path in Γ. In particular, the shortest paths in Γ from 1 to g have length $|g|$ and, for $g, h \in G$, we have $d(g, h) = |g^{-1}h|$.

1.6.3 Schreier graphs The Cayley graph generalises to the *Schreier graph* $\Gamma(G, H, X)$, which is defined for a subgroup $H \leq G$. Here the vertices are labelled by the right cosets Hg of H in G, with an edge labelled x from Hg to Hgx.

1.6.4 Embeddings of Cayley graphs in spaces The facts that the Cayley graphs for $\mathbb{Z}^2$ and the Coxeter group of type $\tilde{A}_2$ (see 1.10.2) embed naturally in $\mathbb{R}^2$, and those for F_2 and the 237-Coxeter group in the Poincaré disk are consequences of the natural cocompact actions of the groups on those spaces as groups of isometries. An embedding of the graph in the appropriate space is found by selecting a point v of the space to represent the vertex 1, and then setting the image of v under g in the space to represent the vertex g for each

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

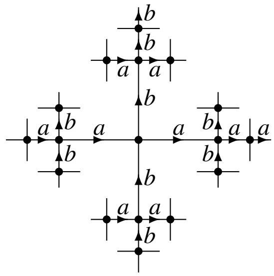


Figure 1.2 Cayley graph of $F_2 = \langle a, b \mid \rangle$

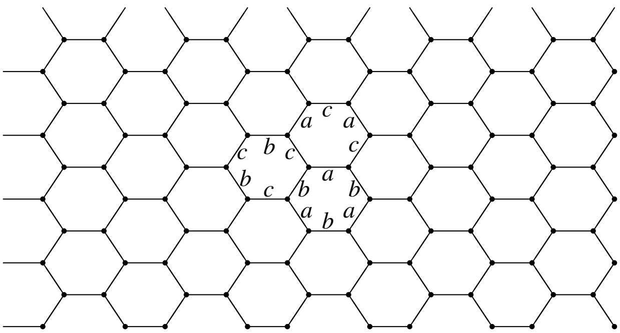


Figure 1.3 Cayley graph of $\tilde{A}_2 = \langle a, b, c \mid a^2 = b^2 = c^2 = (ab)^3 = (bc)^3 = (ca)^3 = 1 \rangle$

$g \in G$. In each of these cases the graph metric d_Γ on Γ and the natural metric $d_\mathcal{X}$ on the space $\mathcal{X}$ in which the graph embeds are related by inequalities

$$d_\Gamma(g, h) \leq \lambda d_\mathcal{X}(g, h) + \epsilon, \quad d_\mathcal{X}(g, h) \leq \lambda d_\Gamma(g, h) + \epsilon,$$

for fixed λ, ϵ and all $g, h \in G$.

An embedding satisfying these conditions is called a *quasi-isometric embedding* (formally defined in 1.7). We shall discuss its relevance to the theory of hyperbolic groups in Section 6.2.

1.6.5 Growth of groups For a group $G = \langle X \rangle$ with X finite, we define the growth function $\gamma_{G,X} : \mathbb{N}_0 \to \mathbb{N}$ by

$$\gamma_{G,X}(n) = |\{g \in G : |g|_X \leq n\}|.$$

Equivalently, $\gamma_{G,X}(n)$ is the number of vertices of $\Gamma(G, X)$ in the closed ball of radius n centred at the origin.

It is straightforward to show that, for two finite generating sets X and Y of G, there is a constant $C \geq 1$ such that

$$\gamma_{G,Y}(n)/C \leq \gamma_{G,X}(n) \leq C\gamma_{G,Y}(n)$$

for all $n \in \mathbb{N}_0$. Hence we can meaningfully refer to groups as having, for example, linear or polynomial growth without reference to the (finite) generating set.

It is also not difficult to see that, if $H = \langle Y \rangle \leq G$ with $|G : H|$ finite, then

$$\gamma_{H,Y}(n)/C \leq \gamma_{G,X}(n) \leq C\gamma_{H,Y}(n),$$

for some constant $C \geq 1$ and all $n \in \mathbb{N}_0$.

By a deep theorem of Gromov [124], a group has polynomial growth if and only if it is virtually nilpotent. We shall use an argument based on growth in our proof that groups with 1-counter word problem are virtually cyclic (Theorem 11.2.1), but for that we only need the more elementary result that groups with linear growth are virtually cyclic.

The *growth series* of the group G with respect to the generating set X is defined to be the power series $\sum_{n=0}^{\infty} \gamma_{G,X}(n)t^n$.

1.7 Quasi-isometries

Let (X_1, d_1) and (X_2, d_2) be metric spaces. A map $f : X_1 \to X_2$ is called a *quasi-isometric embedding* if there exist real constants $\lambda \geq 1$, $\epsilon \geq 0$ such that, for all $x, y \in X_1$,

$$\frac{d_1(x,y)}{\lambda} - \epsilon \leq d_2(f(x), f(y)) \leq \lambda d_1(x,y) + \epsilon.$$

The map f is called a *quasi-isometry* if, in addition, there exists $\mu > 0$ such that for all $y \in X_2$ there exists $x \in X_1$ with $d_2(f(x), y) \leq \mu$. The Cayley graph embeddings mentioned in 1.6.4 above are examples of quasi-isometries.

Note that an isometry is a quasi-isometry with $\lambda = 1$ and $\epsilon = \mu = 0$. Note also that, despite the names, a quasi-isometric embedding is not necessarily injective, and a quasi-isometry is not necessarily bijective.

We say that the spaces (X_1, d_1) and (X_2, d_2) are *quasi-isometric* if there is a quasi-isometry between them. So, for example, any space with finite diameter is quasi-isometric to the one-point space.

1.7.1 Exercise Show that being quasi-isometric is an equivalence relation on metric spaces.

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

1.7.2 Quasi-isometric groups We say that the two finitely generated groups $G = \langle X \rangle$ and $H = \langle Y \rangle$ are *quasi-isometric* if (the 1-skeletons of) their associated Cayley graphs $\Gamma(G, X)$ and $\Gamma(H, Y)$ are. The next result implies that this notion does not depend on the choice of finite generating sets of G and H, and so we can henceforth talk about two finitely generated groups being quasi-isometric.

1.7.3 Proposition *Let X and Y be finite generating sets of the group G. Then $\Gamma(G, X)$ and $\Gamma(G, Y)$ are quasi-isometric with $\epsilon = 0$ and $\mu = 1/2$.*

Proof For each $x \in X$ let w_x^Y be a word over Y with $x =_G w_x^Y$, and define w_y^X similarly. We define $f\colon \Gamma(G, X) \to \Gamma(G, Y)$ to map the vertex labelled g in $\Gamma(G, X)$ to the vertex with the same label in $\Gamma(G, Y)$, and an edge labelled $x^{\pm 1}$ continuously to the corresponding path labelled $(w_x^Y)^{\pm 1}$ in $\Gamma(G, Y)$. It is straightforward to verify that f is a quasi-isometry with

$$\lambda = \max(\{|w_x^Y| : x \in X\} \cup \{|w_y^X| : y \in Y\}),$$

$\epsilon = 0$, and $\mu = 1/2$. (Note that the edges of $\Gamma(G, Y)$ are not necessarily in the image of f.) □

1.7.4 Proposition *Let $H \leq G$ be finitely generated groups with $|G : H|$ finite. Then H and G are quasi-isometric.*

Proof Let $G = \langle X \rangle$, and let Y be the associated Schreier generators of H (see Section 1.4) for some right transversal T of H in G. So each $y \in Y$ is equal in G to a word w_y over X, and we define $f\colon \Gamma(H, Y) \to \Gamma(G, X)$ by mapping each vertex of $\Gamma(H, Y)$ to the vertex of $\Gamma(G, X)$ that represents the same group element, and mapping each edge of $\Gamma(H, Y)$ labelled y continuously to the path with the same initial and final vertices labelled w_y in $\Gamma(G, X)$.

Since, by Remark 1.4.5, any word w over X that represents an element of H can be rewritten as a word v over Y with $v =_H w$ and $|v| \leq |w|$, we see that f is a quasi-isometry with $\lambda = \max\{|w_y| : y \in Y\}$, $\epsilon = 0$ and $\mu = \ell + 1/2$, where ℓ is the maximum X-length of an element of the transversal T. □

1.7.5 Quasigeodesics Let p be a rectifiable path in a metric space. For real numbers $\lambda \geq 1$ and $\epsilon \geq 0$, we say that p is a (λ, ϵ)-*quasigeodesic* if $d_p(x, y) \leq \lambda d(x, y) + \epsilon$ for all $x, y \in p$. The path is a *quasigeodesic* if it is a (λ, ϵ)-quasigeodesic for some λ and ϵ. In this book we shall be only concerned with quasigeodesics in the Cayley graphs of groups. A word w over X that labels a quasigeodesic path in $\Gamma(G, X)$ is called a *quasigeodesic word*.

We should like to be able to say that the image of a geodesic path under a quasi-isometry is a quasigeodesic, but the lack of any continuity requirements

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

on quasi-isometries prevents this being true. There is however a remedy that is sufficient for our applications. The arguments here are based on Bowditch's notes [36, Section 6.9].

The *Hausdorff distance* between two subsets X, Y of a metric space is defined to be the infimum of $r \in \mathbb{R}_{\geq 0}$ such that every point of X is at distance at most r from some point in Y and vice versa.

1.7.6 Lemma *Let $f\colon X \to Y$ be a (λ, ϵ)-quasi-isometric embedding between geodesic metric spaces X, Y, and let p be a geodesic path in X. Then, for some λ', ϵ', r that depend only on λ and ϵ, there is a (λ', ϵ')-quasigeodesic path q in Y that is at Hausdorff distance at most r from $f(p)$.*

Proof Choose $h \in \mathbb{R}_{>0}$ (such as $h = 1$), and let $r := \lambda h + \epsilon$. If $\ell(p) < h$, then let $q = f(p)$, which is a $(0, r)$-quasigeodesic. Otherwise, choose points $x_0, x_1, \ldots, x_n$ in order along p, where x_0 and x_n are its two endpoints, and $h/2 \leq d(x_{i-1}, x_i) \leq h$ for $1 \leq i \leq n$. Let $y_i = f(x_i)$ for $0 \leq i \leq n$, Choose any geodesic path in Y from y_{i-1} to y_i for $1 \leq i \leq n$, and let q be the union of these n geodesic paths taken in order.

We claim that q is a (λ', ϵ')-quasigeodesic, with $\lambda' := 4r\lambda/h$ and $\epsilon' = 4(\epsilon + 2r)\lambda r/h + 2r$. Note that, by choice of the x_i, we have $d(y_{i-1}, y_i) \leq r$ and $d(y_{i-1}, y_j) > (j - i)h/(2\lambda) - \epsilon$ for $1 \leq i \leq j \leq n$. Let t, u be points on q and suppose that $t \in [y_i, y_{i+1}]$ and $u \in [y_j, y_{j+1}]$ with $i \leq j$. Then $d_q(t, u) \leq r(j-i-1)+2r$ and $d(t, u) \geq (j - i - 1)h/(2\lambda) - \epsilon - 2r$. So, if $j - i - 1 \leq 4(\epsilon + 2r)\lambda/h$ then $d_q(t, u) \leq \epsilon'$, and otherwise $d_q(t, u) \leq \lambda' d(t, u) + 2r$, which proves the claim.

Any point z of p is at distance at most h from some x_i and hence, since $y_i \in q$, we have $d(f(z), q) \leq r$. Similarly, any point of q is at distance at most r from $f(p)$. So the Hausdorff distance between $f(p)$ and q is at most r. □

1.8 Ends of graphs and groups

For a connected graph Γ in which every vertex has finite degree, we define the number of ends of Γ to be the smallest integer n such that, whenever a set F of finitely many vertices and their incident edges is deleted from Γ, the graph $\Gamma \setminus F$ that remains has at most n infinite connected components; if there is no such n, then Γ has infinitely many ends. We may take F to be a finite ball within Γ and, if Γ is a Cayley graph of a group, then we may take F to be a finite ball about the vertex representing the identity element.

For a finitely generated group $G = \langle X \rangle$, it turns out that the number of ends of the Cayley graph $\Gamma(G, X)$ is the same for any finite generating set X, and we call this the *number of ends of G*, and write it $e(G)$.

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

When G is not finitely generated the above definition does not work. In that case we define $e(G)$ to be the dimension of the cohomology group $H^2(G, \mathbb{F}_2 G)$ of G acting on its group algebra over the field with 2 elements; when G is finitely generated this gives the same value as above.

This topic was first studied by Hopf [160], where the following results are proved for finitely generated groups G.

(1) $e(G) \in \{0, 1, 2, \infty\}$.
(2) $e(G) = 0$ if and only if G is finite.
(3) $e(G) = 2$ if and only if G is virtually infinite cyclic.
(4) $e(\mathbb{Z}^k) = 1$ and $e(F_k) = \infty$ for all $k > 1$.

1.8.1 Theorem (Stalling's Ends Theorem [241]) *If G is finitely generated, then $e(G) > 1$ if and only if either $G = H *_A K$ or $G = H*_{A,t}$ for subgroups H, A and (in the first case) K of G, with A finite and distinct from H (and K).*

In particular, if G is torsion-free, and so cannot have a finite subgroup, Stallings' theorem implies that $e(G) > 1$ if and only if G decomposes non-trivially as a free product. The proof by Muller and Schupp that groups with context-free word problem are virtually free, which we shall present in Theorem 11.1.1, uses this result.

1.9 Small cancellation

Small cancellation theory extends the ideas of Dehn, who exploited properties of the natural presentations of surface groups in order to solve the word and conjugacy problems in these groups. The theory identifies particular features of a presentation that indicate restricted interaction between relators, from which various properties of groups possessing such presentations can be deduced.

For a set R of words over an alphabet $(X^{\pm})^*$, we define the *symmetric closure* $\hat{R}$ of R to be the set of all cyclic conjugates of the elements of R and their inverses. The symmetric closure of a group presentation $\langle X \mid R \rangle$ is defined to be $\langle X \mid \hat{R} \rangle$, which of course defines the same group. For example, the symmetric closure of the presentation $\langle a, b \mid aba^{-1}b^{-1} \rangle$ of $\mathbb{Z}^2$ is

$$\langle a, b \mid abAB, bABa, ABab, BabA, baBA, aBAb, BAba, AbaB \rangle,$$

where $A := a^{-1}$, $B := b^{-1}$. We normally assume that the words in R are all cyclically reduced before taking the symmetric closure.

We define a *piece* in a presentation $\langle X \mid R \rangle$ to be a word u that is a prefix of at least two distinct elements of $\hat{R}$. Now, for p, q positive integers and $\lambda \in (0, 1)$, the presentation is said to satisfy:

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

$C(p)$ (with $p \in \mathbb{N}$) if no element of $\hat{R}$ is a product of fewer than p pieces;
$C'(\lambda)$ (with $\lambda \in (0,1)$) if whenever a piece u is a prefix of $r \in \hat{R}$, then $|u| < \lambda|r|$;
$T(q)$ (with $q \in \mathbb{N}$) if, whenever $3 \le h < q$ and $r_1, r_2, \ldots, r_h \in \hat{R}$ with $r_i \ne r_{i+1}^{-1}$ for $1 \le i < h$ and $r_1 \ne r_h^{-1}$, then at least one of the products $r_1r_2, r_2r_3, \ldots, r_hr_1$ is freely reduced without cancellation.

Note that $C'(\lambda)$ implies $C(p+1)$ whenever $\lambda \le 1/p$, and that all presentations satisft $T(q)$ for $q \le 3$. The conditions $C(p)$ and $T(q)$ might be best understood in the context of van Kampen diagrams, which will be defined in Section 3.2. The condition $C(p)$ requires that, in a reduced van Kampen diagram for the presentation, no internal region has fewer than p consolidated edges (where edges separated by a vertex of degree 2 are regarded as being part of the same consolidated edge), while $T(q)$ requires that in such a diagram no internal vertex has degree less than q.

1.9.1 Example In the above presentation of $\mathbb{Z}^2$, the pieces are a, b, a^{-1}, b^{-1}, and the presentation satisfies $C'(1/3), C(4)$ and $T(4)$. More generally, all right angled Artin groups (see 1.10.4) have presentations satisfying $C(4)$ and $T(4)$, and in particular free groups do so vacuously, since they have no relations.

1.9.2 Example The involutory relations in Coxeter groups (see 1.10.2) obstruct good small cancellation conditions, but they have index 2 subgroups that do better. For example,

$$\mathrm{Cox}_{4,4,4} = \langle x_1, x_2, x_3 \mid x_1^2 = x_2^2 = x_3^2 = 1, (x_1x_2)^4 = (x_1x_3)^4 = (x_2x_3)^4 \rangle$$

has an index 2 subgroup generated by $a = x_1x_2$, $b = x_2x_3$,

$$\langle a, b \mid a^4 = b^4 = (ab)^4 = 1 \rangle$$

that satisfies $C(4)$ and $T(4)$.

1.9.3 Results We call a group G a $C(p)$-*group*, or a $C(p) + T(q)$-*group*, if it has a presentation that satisfies $C(p)$, or both $C(p)$ and $T(q)$, respectively.

The earliest results on groups satisfying small cancellation conditions are due to Greendlinger, who used purely combinatorial arguments to prove in [110] that $C'(1/6)$-groups have Dehn presentations (Section 3.5) (and so their word problem is soluble in linear time), and in [111] that they have soluble conjugacy problem. In [112] he proved further that $C'(1/4)+T(4)$-groups have Dehn presentations and soluble conjugacy problem.

Lyndon [182] (or see [183, Section V.4]), used curvature arguments on van Kampen diagrams to prove the stronger results that $C(6)$-groups, $C(4) + T(4)$-groups and $C(3) + T(6)$-groups all have soluble word problem. Schupp [229]

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

(or see [183, Section V.7]) proved that such groups also have soluble conjugacy problem.

We shall be studying the more recently defined classes of (bi)automatic and (word-)hyperbolic groups later in this book, in Chapters 5 and 6. If we restrict attention to finite presentations, then a variety of small cancellation conditions have been shown to imply biautomaticity, which implies word problem soluble in quadratic time and soluble conjugacy problem, and hyperbolicity, which implies word and conjugacy problems soluble in linear time. We summarise these results in 5.5.2.

1.10 Some interesting families of groups

In the final section of this introductory chapter, we introduce some of the specific families of groups that arise in combinatorial and geometric group theory, and which occur as examples in the book. We describe some of their properties, which in some cases, such as *automaticity*, may not be familiar to the reader at this stage, but will be discussed in more detail later in the book.

1.10.1 Surface groups The fundamental group T_k of the orientable surface of genus k can be presented as

$$T_k = \langle a_1, b_1, a_2, b_2, \ldots, a_k, b_k \mid [a_1, b_1][a_2, b_2] \cdots [a_k, b_k]\rangle.$$

In particular the fundamental group T_1 of the torus is isomorphic to $\mathbb{Z}^2$.

The fundamental group of a non-orientable surface of genus n (a sphere with n attached crosscaps) has an orientable surface of genus n as a double cover. The non-orientable surfaces exhibit some differences according to the parity of the genus. The genus $2k$ surface is a k-fold Klein bottle, and its group K_k has presentation

$$\langle a_1, b_1, a_2, b_2, \ldots, a_k, b_k \mid [a_1, b_1][a_2, b_2] \cdots [a_{k-1}, b_{k-1}]a_k b_k a_k^{-1} b_k\rangle,$$

while the genus $2k + 1$ surface has group P_{k+1} with presentation

$$\langle a_1, b_1, a_2, b_2, \ldots, a_k, b_k, c \mid [a_1, b_1][a_2, b_2] \cdots [a_k, b_k]c^2\rangle.$$

By putting $c = a_k$ and $d = a_k^{-1} b_k$, we see that K_k has the alternative presentation

$$\langle a_1, b_1, a_2, b_2, \ldots, a_{k-1}, b_{k-1}, c, d \mid [a_1, b_1][a_2, b_2] \cdots [a_{k-1}, b_{k-1}]c^2 d^2\rangle.$$

From the small cancellation properties of these presentations and the results summarised in 5.5.2, we find that the surface groups are all automatic, and that T_k ($k \geq 2$), P_k ($k \geq 1$), and K_k ($k \geq 2$) are hyperbolic.

1.10.2 Coxeter groups Coxeter groups are groups with presentations

$$\langle x_1, x_2, \ldots, x_n \mid x_i^2 = 1\,(1 \le i \le n),\ \ (x_i x_j)^{m_{ij}} = 1\,(1 \le i < j \le n)\rangle.$$

They can be described by their associated Coxeter matrices (m_{ij}); these are defined to be symmetric matrices with entries from $\mathbb{N} \cup \{\infty\}$, with $m_{ii} = 1$ for all i and $m_{ij} \ge 2$ for all $i \ne j$. When $m_{ij} = \infty$, the relation $(x_i x_j)^{m_{ij}} = 1$ is omitted.

A Coxeter matrix can also be represented by a Coxeter diagram, with n nodes, indexed by $\{1, \ldots, n\}$, and with an edge labelled m_{ij} joining i to j for each $i \ne j$. The diagram is generally simplified by modifying edges labelled 2, 3, 4 as follows: the edges labelled 2 are deleted, the edges labelled 3 are shown unlabelled and the edges labelled 4 are shown as double edges. The group is called *irreducible* if its (simplified) Coxeter diagram is connected as a graph; otherwise it is a direct product of such groups. The reader is warned that there are alternative labelling conventions in which edges labelled ∞ are omitted.

Any Coxeter group has a faithful representation [34, Chapter V, 4.3] as a group of linear transformations of $\mathbb{R}^n = \langle e_i : i = 1, \ldots, n\rangle$ that preserves the symmetric bilinear form B defined by $B(e_i, e_j) = -\cos(\pi/m_{ij})$. The image of x_i is the involutory map $x \mapsto x - 2B(x_i, x)e_i$. The Coxeter group is finite, and then called *spherical*, if and only if B is positive definite [34, Chapter V, 4.8], which is equivalent to the eigenvalues of the associated Cartan matrix $(-2\cos(\pi/m_{ij}))$ all being positive. It is *Euclidean* (and virtually abelian) if it is not spherical and the Cartan matrix has no negative eigenvalues, and *hyperbolic* if the Cartan matrix has $n - 1$ positive and 1 negative eigenvalue. These are the only three possibilities for 3-generator Coxeter groups, but not in general.

In the 3-generator case, putting $a := m_{12}$, $b := m_{13}$, and $c := m_{23}$, the Cartan matrix is

$$2\begin{pmatrix} 1 & -\cos\frac{\pi}{a} & -\cos\frac{\pi}{b} \\ -\cos\frac{\pi}{a} & 1 & -\cos\frac{\pi}{c} \\ -\cos\frac{\pi}{b} & -\cos\frac{\pi}{c} & 1 \end{pmatrix},$$

which has positive trace 6, and determinant

$$1 - \cos^2\frac{\pi}{a} - \cos^2\frac{\pi}{b} - \cos^2\frac{\pi}{c} - 2\cos\frac{\pi}{a}\cos\frac{\pi}{b}\cos\frac{\pi}{c}.$$

The type of the group now just depends on the angle sum

$$\frac{\pi}{a} + \frac{\pi}{b} + \frac{\pi}{c}.$$

The group is hyperbolic if and only if the angle sum is less than π, Euclidean if the angle sum is equal to π, and spherical if the angle sum is greater than π.

The reader is undoubtedly familiar with Escher's artwork displaying tessellations of the Poincaré disc (e.g. by angels and devils); these tessellations are preserved by 3-generated hyperbolic Coxeter groups. Similarly, tessellations of the Euclidean plane (e.g. by equilateral triangles or by squares) are preserved by 3-generated Euclidean Coxeter groups, and tessellations of the sphere (or, equivalently, Platonic solids) are preserved by 3-generated spherical Coxeter groups.

The subgroups of a Coxeter group generated by subsets of the standard generating set $\{x_1, \ldots, x_n\}$ are called its *standard subgroups*, and are also Coxeter groups, with presentations given by the appropriate subdiagrams. The term *parabolic subgroup* is also used, with exactly the same meaning.

The list of spherical Coxeter diagrams, of types A_n, $B_n = C_n$, D_n, E_6, E_7, E_8, F_4, G_2, H_3, H_4, $I_2(m)$, that correspond to the finite Coxeter groups is well-known. The Coxeter groups of type A_n are the symmetric groups Sym_{n+1}. The related affine diagrams, of types $\tilde{A}_n$, $\tilde{B}_n$, $\ldots$, correspond to the affine Coxeter groups, which are all infinite and virtually abelian.

The reader is warned that hyperbolicity of a Coxeter group does not correspond to *word-hyperbolicity*, which is the subject of Chapter 6. It is proved by Moussong [197] that a Coxeter group is word-hyperbolic if and only if it contains no subgroup isomorphic to $\mathbb{Z}^2$ or, equivalently, if and only if it contains no affine standard subgroup of rank greater than 2, and there are no pairs of infinite commuting standard subgroups. The word-hyperbolic Coxeter groups are all hyperbolic as Coxeter groups, and in the 3-generator cases the two properties coincide, but in higher dimensions there are hyperbolic Coxeter groups that are not word-hyperbolic.

In general, for letters x, y, define $(x, y)_m$ to be the word of length m that starts with x and then alternates between y and x. So, for example, $(x, y)_5 = xyxyx$. Then the Coxeter group defined by (m_{ij}) has the alternative presentation

$$\langle x_1, x_2, \ldots, x_n \mid x_i^2 = 1 \ \forall i,\ (x_i, x_j)_{m_{ij}} = (x_j, x_i)_{m_{ji}} \ \forall i, j, i < j \rangle.$$

The relations $x_i^2 = 1$ are called the *involutory relations*, and the relations $(x_i, x_j)_{m_{ij}} = (x_j, x_i)_{m_{ji}}$ the *braid relations*.

In a Coxeter group any word is clearly equivalent in the group to a *positive word*; that is, to a word that contains no inverses of generators. It was shown by Tits [246] that any positive word can be reduced to an equivalent geodesic word by application of braid relations and deletion of subwords of the form x_i^2. Hence the word problem is soluble in Coxeter groups. The conjugacy problem was proved to be soluble (in cubic time) by Krammer [173], using the action of the group on a suitable cubical complex. Brink and Howlett [49] proved that all Coxeter groups are shortlex automatic.

1.10.3 Artin groups The presentations formed by deleting the involutory relations from presentations for Coxeter groups define the Artin groups:

$$\langle x_1, x_2, \ldots, x_n \mid (x_i, x_j)_{m_{ij}} = (x_j, x_i)_{m_{ji}} \forall i, j, i < j \rangle.$$

These groups are all infinite (there is a homomorphism to $\mathbb{Z}$ mapping all generators to 1), and they map onto their corresponding Coxeter groups.

An Artin group is said to be of dihedral type if the associated Coxeter group is dihedral, of finite or spherical type if the associated Coxeter group is finite, and of large or extra-large type if all of the m_{ij} are at least 3 or at least 4, respectively.

It is proved by van der Lek [247] that a subgroup of any Artin group that is generated by a subset of the standard generating set is also an Artin group, with presentation defined by the appropriate subdiagram. Such a subgroup is called a *standard* or *parabolic subgroup*, and shown by Charney and Paris [67] to be convex within the original group; that is, all geodesic words in the group generators that lie in the subgroup are words over the subgroup generators.

It is proved by Brieskorn and Saito [47] that Artin groups of spherical type are torsion-free; but in general the question of whether or not Artin groups are torsion-free is open. It is not known which Artin groups are word-hyperbolic; but certainly an Artin group whose Coxeter diagram contains two vertices that are not connected cannot be hyperbolic, since it contains a standard $\mathbb{Z}^2$ subgroup. A number of open questions for Artin groups are listed by Godelle and Paris [104].

Artin groups of spherical and of extra-large type are known to be biautomatic, and Artin groups of large type are automatic; see 5.5.4 for references.

1.10.4 Right angled Artin groups The free group of rank n, F_n, and the free abelian group $\mathbb{Z}^n$ are Artin groups. In fact they are examples of right-angled Artin groups: those Artin groups for which all entries in the Coxeter matrix are in the set $\{2, \infty\}$, where 2 indicates a commuting relation, and ∞ the lack of a relation between two generators. Such groups are also called graph groups, since the commuting relations in the group can be indicated using a finite graph. They can be be described equivalently as graph products (see 1.5.7) of infinite cyclic groups.

1.10.5 Braid groups The spherical Artin groups of type A_n are equal to the braid groups B_{n+1}, found as the mapping class groups of $(n+1)$-punctured disks, with presentations

$$\begin{aligned} B_{n+1} = \langle x_1, x_2, \ldots, x_n \mid\ & x_i x_{i+1} x_i = x_{i+1} x_i x_{i+1}\,(1 \le i \le n), \\ & x_i x_j = x_j x_i\,(1 \le j < i \le n,\ i - j \ge 2)\rangle. \end{aligned}$$

1.10.6 Garside groups Spherical type Artin groups provide examples of Garside groups. Indeed a Garside group is defined by Dehornoy and Paris [78] to be the group of fractions of a monoid with certain properties of divisibility that had already been identified in spherical type Artin monoids. Many other examples are described in [78, Section 5]. They include torus knot groups (i.e. $\langle x, y \mid x^m = y^n \rangle$ for $m, n > 1$), the fundamental groups of complements of systems of complex lines through the origin in $\mathbb{C}^2$, and some 'braid groups' associated with complex reflection groups.

The definition of a Garside group depends on the concept of an *atomic monoid*. An element m of a monoid M is called *indivisible* if $m \neq 1$ and $m = ab$ implies $a = 1$ or $b = 1$. Then M is called *atomic* if it is generated by its indivisible elements and, for each $m \in M$, the supremum of the lengths of words $a_1 a_2 \cdots a_r$ equal to m in M and with each a_i atomic is finite. We can define a partial order with respect to left divisibility on any atomic monoid by $a \preceq_L b$ if $ac = b$ for some $c \in M$. The finiteness of the supremum of lengths of words for elements in M implies that we cannot have $a = acd$ for $a, c, d \in M$ unless $c = d = 1$, and hence $a \preceq_L b$ and $b \preceq_L a$ if and only if $a = b$. We can do the same for right divisibility, and so define a second partial order $\preceq_R$. A *Garside group* G is now defined to be a group having a submonoid G^+ that is atomic, and for which the following additional properties hold:

(i) any two elements of G^+ have least common left and right multiples and greatest common left and right divisors;
(ii) there exists an element Δ of G^+ with the property that the sets of left and right divisors of Δ coincide; that set forms a finite generating set for G^+ as a monoid and G as a group. We call Δ the *Garside element*.

We note that a group may have more than one Garside structure, and indeed any Artin group of spherical type has at least two quite distinct such structures.

An elementary example of a Garside group is provided by the braid group on three strings generated by a and b, for which $\Delta = aba = bab$ and $X = \{a, b, ab, ba, aba\}$. Notice that the set X of divisors of Δ is generally much larger than the natural minimal generating set of the group. For the braid group B_n on n strings generated by $n-1$ simple crossings, we have $|X| = n! - 1$. The $n-1$ simple crossings are of course the atomic elements.

1.10.7 Knot groups A *knot* is the image of an embedding of the circle into $\mathbb{R}^3$; we generally envisage the knot as being contained within a portion of $\mathbb{R}^3$ of the form

$$\{(x, y, z); -\delta < z < \delta\}.$$

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

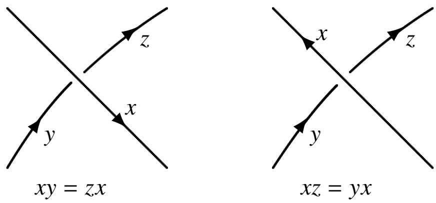


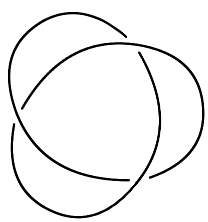
Figure 1.4 Wirtinger relations

A *link* is the image of an embedding of finitely many disjoint circles into $\mathbb{R}^3$. The *fundamental group*, commonly called the *knot group*, of a knot, or link L, is defined to be the fundamental group of its complement in $\mathbb{R}^3$. Its *Wirtinger presentation* is defined as follows.

Let D be an oriented diagram for L, with n arcs. The arcs are labelled with the symbols from X, a set of cardinality n. Then for each of the n crossings, R contains the rule as shown in Figure 1.4.

Basically these are the relations satisfied by homotopy classes of loops in the complement of the knot, where the loop corresponding to a given oriented arc is the loop from a base point above the knot (e.g on the positive z-axis) that comes down to the slice of $\mathbb{R}^3$ containing the knot, crosses under the arc from left to right (these being defined by the orientation) and then returns to the base point. We call a loop that corresponds to an oriented arc in this way a *latitudinal loop*.

1.10.8 Exercise Show that the trefoil knot, depicted below, has fundamental group isomorphic to the braid group on three strands.



1.10.9 Flats in knot groups We can always find a $\mathbb{Z}^2$ subgroup within a knot group, generated by a longitudinal and a latitudinal loop. A longitudinal loop is defined by a path that goes from the base point above the knot to a point just above the knot, then follows the knot sitting just above it until it reaches the point at which it originally reached the knot, and then returns to the base point. For example, we can take the latitudinal loop b together with the longitudinal

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press

loop $a^{-1}b^{-1}c^{-1} = (cba)^{-1}$ that first reaches the knot just after the lowest of the three crossings. Now we see easily from our three relations that

$$cba.b = cb(ab) = cb(ca) = c(bc)a = c(ab)a = (ca)ba = (bc)ba = b.cba.$$

Hence the two generators commute, and we have a $\mathbb{Z}^2$ subgroup. The existence of this subgroup means that a knot group can never be word-hyperbolic, even when the knot complement has a hyperbolic structure; that structure has finite volume but is not compact. However it was proved by Epstein [84, Theorem 11.4.1] that the group of a hyperbolic knot is biautomatic, and more generally that the fundamental group of any geometrically finite hyperbolic manifold is biautomatic.

1.10.10 Baumslag–Solitar groups The Baumslag–Solitar groups are defined by the presentations

$$\mathrm{BS}(m,n) := \langle x, y \mid y^{-1}x^m y = x^n \rangle,$$

where $m, n \in \mathbb{Z} \setminus \{0\}$. They are examples of HNN-extensions, which were the topic of 1.5.15. This means, in particular, that their elements have normal forms, a fact that facilitates their study.

The group $\mathrm{BS}(1,n)$ is metabelian, and isomorphic to the subgroup

$$\left\{ \begin{pmatrix} n^k & an^l \\ 0 & 1 \end{pmatrix} : a, k, l \in \mathbb{Z} \right\}$$

of $\mathrm{GL}_2(\mathbb{Q})$.

The groups $\mathrm{BS}(m,n)$ were first studied by Baumslag and Solitar [21], who proved that, for certain m and n, they are *non-Hopfian*; that is, they are isomorphic to proper quotients of themselves. To be precise, they are non-Hopfian provided that $|m|, |n| > 1$ and m and n do not have the same sets of prime divisors. For example, it is not hard to show that $G := \mathrm{BS}(2,3) \cong G/N$, where N is the normal closure of $r := (x^{-1}y^{-1}xy)^2x^{-1}$ in G. The normal form for HNN-extensions (Proposition 1.5.17) can be used to show that $r \neq_G 1$.

When $|m| \neq |n|$, the groups are asynchronously automatic but not automatic; see 5.5.7. The complexity of the word and co-word problems of these groups is addressed in Proposition 12.4.4, in 13.1.7 and in 14.2.10.

1.10.11 Higman–Thompson groups The *Higman–Thomson* groups $G_{n,r}$, for $n \geq 2$ and $r \geq 1$, are finitely presented infinite groups, whose commutator subgroups are simple and of index 2 if n is odd and equal to the whole group otherwise. They were defined as automorphism groups of certain algebras by

Higman [142], as generalisations of Richard Thompson's group V, which is $G_{2,1}$.

These groups have a number of different but equivalent definitions, and we summarise here the definition that is most convenient for their connections with formal language theory. Let $Q = \{q_1, q_2, \ldots, q_r\}$ and $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ be disjoint sets. Define Ω to be the set of all infinite sequences of the form $\omega = q_i\sigma_{i_1}\sigma_{i_2}\sigma_{i_3}\cdots$, where $q_i \in Q$ and $\sigma_{i_j} \in \Sigma$ for all $j \geq 1$. Henceforth, by a *prefix* of ω, we mean a nonempty finite prefix. Then, roughly speaking, the group $G_{n,r}$ is the group that acts on Ω by replacing prefixes.

More precisely, a finite complete anti-chain in Ω is a finite subset C of $Q\Sigma^*$ such that every element of Ω has precisely one element of C as a prefix. Any bijection $\phi\colon B \to C$ between two finite complete anti-chains B and C induces a permutation of Ω, also called ϕ, by setting $\phi(\omega) = \phi(b)\omega'$ where $\omega = b\omega'$ with $b \in B$. The group $G_{n,r}$ is the group of all such permutations of Ω.

It is important to note that each element of $G_{n,r}$ has infinitely many representations by bijections between finite complete anti-chains, because we can replace B by $B' := (B \setminus \{b\}) \cup b\Sigma$, similarly C by $C' := (C \setminus \{\phi(b)\}) \cup \phi(b)\Sigma$ and ϕ by $\phi' : B' \to C'$, where ϕ' agrees with ϕ on $(B \setminus \{b\}) \subseteq B'$ and $\phi'(b\sigma) = \phi(b)\sigma$ for $\sigma \in \Sigma$. Using this process of expansion, the group operations can be carried out within the set of bijections between finite complete anti-chains, because any two finite complete anti-chains have a common expansion. This is easy to see, if one views the set $Q\Sigma^*$ as a disjoint union of r rooted trees with roots in Q and edges between v and $v\sigma$ for all $v \in Q\Sigma^*$ and $\sigma \in \Sigma$. Then Ω is the set of ends of this forest and finite complete anti-chains are in bijection with leaf sets of finite subforests that are finite with set of roots equal to Q. In order to find the common expansion of two finite complete anti-chains, take the leaves of the forest that is the union of the two given forests.

If we equip Ω with the lexicographic order induced by total orders on Q and Σ (see 1.1.4), then the group $G_{n,r}$ has an order preserving subgroup $F_{n,r}$ and a cyclic order preserving subgroup $T_{n,r}$. The famous Thompson group F equals $F_{2,1}$. The standard reference for F is still [62]. We will encounter some of these groups again in 9.1.4, Section 12.2, 14.2.5 and 14.2.9.

1.10.12 Exercise Given that $\langle x_i\,(i \in \mathbb{N}_0) \mid x_i x_j x_i^{-1} = x_{j+1} \text{ for } 0 \leq i < j\rangle$ is a presentation for Thompson's group F, show that F is an HNN-extension of a group isomorphic to F. Show also that $F = \langle x_0, x_1 \mid [x_1^{x_0}, x_1^{-1}x_0], [x_1^{x_0^2}, x_1^{-1}x_0]\rangle$.

https://doi.org/10.1017/9781316588246.002 Published online by Cambridge University Press