The Mapping Class Group of the Twice Punctured Torus

John R. Parker & Caroline Series

Department of Mathematical Sciences, University of Durham
Durham DH1 3LE, England

Mathematics Institute, University of Warwick
Coventry CV4 7AL, England

8 March 2005

Introduction.

Let Σ be a (possibly punctured) surface of negative Euler characteristic, and let $C(\Sigma)$ be the set of isotopy classes of families of disjoint simple closed curves on Σ. When Σ is a once punctured torus Σ_1, there is a well known recursive structure on $C(\Sigma_1)$ which arises from the relationships between $C(\Sigma_1)$ (identified with the extended rational numbers), continued fractions, and $\mathrm{PSL}(2,\mathbb{Z})$ (the mapping class group of Σ_1) [11], [21]. The results in this paper arose out of a search for an analogous structure on $C(\Sigma_2)$, where Σ_2 is a torus with two punctures. Masur and Minsky [15], [16] have recently described an alternative approach.

Our method is motivated by the Bowen-Series construction [4], [22] of Markov maps for Fuchsian groups. This generalised the relationship between $\mathrm{PSL}(2,\mathbb{Z})$ (now thought of as a Fuchsian group acting in the hyperbolic plane) and continued fractions (now thought of as points in the limit set of $\mathrm{PSL}(2,\mathbb{Z})$), to a large class of Fuchsian groups Γ.

The Markov map was a map on the boundary at infinity, in other words the limit set $\Lambda(\Gamma)$, which generated continued fraction expansions for points in $\Lambda(\Gamma)$, and whose admissible sequences simultaneously gave an elegant solution to the word problem in Γ [22] (see section 1.1 below).

The idea behind this paper rests on the analogy between Γ acting on the hyperbolic plane and the mapping class group $\mathcal{MCG}(\Sigma)$ acting on Teichmüller space $\mathcal{T}(\Sigma)$. In this analogy, the boundary S^1 of the hyperbolic plane (or the limit set of Γ) is replaced by a suitable boundary of $\mathcal{T}(\Sigma)$. We use the **Thurston boundary**, namely the space $\mathcal{PML}(\Sigma)$ of projective measured laminations on Σ (see section 1.4 below). The mapping class group $\mathcal{MCG}(\Sigma)$ (see section 1.5 below) acts on both $\mathcal{T}(\Sigma)$ and $\mathcal{PML}(\Sigma)$. By analogy with the Bowen-Series construction, we define a Markov map f on $\mathcal{PML}(\Sigma)$ which has the same

relation to the action of $\mathcal{MCG}(\Sigma)$ on $\mathcal{PML}(\Sigma)$ as the Bowen-Series map has to the action of Γ on $\Lambda(\Gamma) = S^1$.

Thurston's well known theory of train tracks [18], [23] gives $\mathcal{ML}(\Sigma)$ a piecewise linear cone structure. Here we use the variant of π_1-train tracks introduced by Birman and Series [2]. This special class of train tracks is defined relative to a fixed choice of fundamental domain and associated geometric generators for $\pi_1(\Sigma)$, in such a way that an integer weighting yields not only a (multiple) simple loop but simultaneously allows one to read off a shortest representative as a cyclic word in $\pi_1(\Sigma)$. Thus the space $\mathcal{ML}(\Sigma)$ is partitioned into finitely many maximal cells corresponding to weightings on the (finitely many) possible π_1-train tracks associated to a given fundamental domain for Σ_2.

In this paper we study the special case of the twice punctured torus Σ_2. We use this structure to construct a Markov map f on $\mathcal{ML}(\Sigma_2)$. The Markov partition is essentially the set of maximal cells and the restriction of f to each cell is a specific (rather simple) element of $\mathcal{MCG}(\Sigma_2)$ which acts linearly on the set of weights. Labelling the cells by the corresponding elements of $\mathcal{MCG}(\Sigma_2)$, we show that the f-expansions (that is the labelled orbit paths of f, see section 1.1) give a unique normal form for the elements of $\mathcal{MCG}(\Sigma_2)$. In particular, the labels are a set of generators for $\mathcal{MCG}(\Sigma_2)$, and comparison of normal forms for nearby elements allows us to find a presentation for $\mathcal{MCG}(\Sigma_2)$. Since the map f is Markov, the f-expansions lie in a subshift of finite type which is in fact close to geodesic with respect to the set of generators in question.

In the language of automatic groups (see section 1.2 below), these normal forms for elements of $\mathcal{MCG}(\Sigma_2)$ allow us to construct a word acceptor. If we can show that these normal forms satisfy the fellow traveller property then this gives an explicit automatic structure on $\mathcal{MCG}(\Sigma_2)$. We conclude the paper by showing that this is indeed the case. Our method is rather similar to that given by Mosher [17] who shows that any mapping class group has an automatic structure.

The dimension of $\mathcal{PML}(\Sigma)$ is necessarily odd. We choose to study the case of the twice punctured torus because it is one of the few three dimensional examples. The details of the construction are rather special; we conjecture that the underlying principles are not. One of the main obstacles to finding a complete generalisation of these techniques is the difficulty of finding a map f and a Markov partition of $\mathcal{PML}(\Sigma)$ suitably related to the piecewise linear structure on $\mathcal{PML}(\Sigma)$.

An illuminating discussion can be made for the once punctured torus Σ_1, where the dimension of $\mathcal{PML}(\Sigma_1)$ is 1. Here the Teichmüller space is the upper half plane and the mapping class group is $\mathrm{PSL}(2,\mathbb{Z})$. Although the final results are familiar, the methods may be of interest, and we begin by presenting this example in some detail to explain our ideas.

Presumably similar methods would prove the automaticity result for Bowen-Series

expansions in the Fuchsian group case.

The outline of the paper is as follows. Section 1 draws together the necessary background material from a variety of sources. Section 2 gives the construction for the once punctured torus. Sections 3, 4 and 5 extend this construction to the twice punctured torus. Section 3 contains an explicit development of the Birman-Series construction for Σ_2. Section 4 gives the action of the mapping class group and the construction of the Markov map f. In section 5 we construct a word difference machine which shows that the f-expansions satisfy the fellow traveller property.

The results in this paper arose out of discussions between the authors and Linda Keen as part of our ongoing work to understand the Maskit embedding of the twice punctured torus, [12], [13]. We would also like to thank David Epstein and Sarah Rees for helpful discussions about automatic groups. Part of this research was carried out while the first author was supported by a S.E.R.C./E.P.S.R.C. Research Fellowship held at the University of Warwick in the period 1992–1994.

Index of symbols used.

δ_j	a Dehn twist on the once or twice punctured torus: sections 1.5, 2.2, 3.2.
ι_j	a symmetry of the once or twice punctured torus: sections 2.2, 3.2.
ρ_j	the composition of Dehn twists $\delta_j\delta_0\delta_j$ for $j = 1, 2$: sections 4.6, 5.3, 5.4.
ϕ	an element of the mapping class group $\mathcal{MCG}(\Sigma_j)$.
ψ	a word difference: sections 1.2, 2.5, 5.
e	the identity element of a group.
$\mathbf{e}_j$, $\mathbf{e}_j^i$	irreducible loops on the once or twice punctured torus: sections 2.1, 3.1.
f_j	a Markov map on $\mathcal{ML}(\Sigma_j)$.
I_j	maximal cells for $\mathcal{ML}(\Sigma_1)$: section 2.1.
Δ_j	maximal cells for $\mathcal{ML}(\Sigma_2)$: section 3.1.
Σ_j	the j times punctured torus, for $j = 1, 2$.
A_j, B_j, ...	regions in $\mathcal{ML}(\Sigma_2)$ or $\mathcal{F}$: sections 4.1, 4.5.
Q_j, R_j, ...	the union of several regions in $\mathcal{F}$: sections 2.4, 5.2.
X°	the interior of a set X.
$\mathcal{A}$	the alphabet for a word acceptor: sections 1.2, 2.4, 4.7.
$\mathcal{D}$	the collection of word differences: sections 1.2, 2.5, 5.1, 5.5.
$\mathcal{F}$	the set of Farey blocks (pairs) for Σ_2 (or Σ_1): sections 2.3, 4.3.

1. Background

In this section we gather together all of the background material we need. This is taken from a variety of different areas. Much of the material is expository in nature.

1.1. Markov Maps and the Bowen-Series construction

A **Markov map** on a space X is a map $f\colon X \to X$, together with a finite (or in certain cases infinite) partition of X into sets X_i, such that $f(X_i)$ is an exact union of sets X_j. We say that f satisfies the **Markov property**: if $f(X_i^\circ) \cap X_j^\circ$ is non empty then $X_j \subset f(X_i)$. For $\xi \in X$, let $p(\xi) = j$ if $\xi \in I_j$. The sequence $p(\xi), p(f(\xi)), p(f^2(\xi)), p(f^3(\xi)), \ldots$ is called the **f-expansion** of ξ. Associated to f is a transition matrix of zeros and ones recording which transitions between states can occur. Since f is Markov, all infinite sequences with allowable transitions occur. The finite blocks which occur in these expansions are called **admissible**. Often a Markov map is required to be expanding or to have other differentiability properties. Such questions will not concern us here.

The Bowen-Series construction was modelled on the relationship between $\mathrm{PSL}(2,\mathbb{Z})$ acting in the upper half plane model of the hyperbolic plane $\mathbb{H}^2$ and the continued fraction transformation acting on the extended real line $\mathbb{R} \cup \{\infty\}$. The continued fraction map

$$f(x) = \begin{cases} x-1 & \text{if } x \geq 1 \\ x+1 & \text{if } x \leq -1 \\ \frac{-1}{x} & \text{if } |x| \leq 1 \end{cases}$$

can be regarded as an example of a Markov map with the partition Π into intervals $[\infty,-1]$, $[-1,0]$, $[0,1]$, $[1,\infty]$. (For simplicity here and in what follows we omit details about endpoints. As defined above, the map f is 2-valued at the endpoints.) The f-expansion of a point $\xi \in \mathbb{R}$ is essentially the same as its continued fraction expansion. We note that the restriction of f to each element of Π belongs to the finite subset $\Gamma_0 = \{x \longmapsto x-1, x \longmapsto x+1, x \longmapsto -1/x\} \subset \mathrm{PSL}(2,\mathbb{Z})$. (The well known fact that Γ_0 is a generating set for $\mathrm{PSL}(2,\mathbb{Z})$ may be proved using these expansions [22].) Via f-expansions, $\mathbb{R}\cup\{\infty\}$ may be mapped in an obvious way into $\Pi_{n=0}^{\infty}\Gamma_0$, giving an alternative viewpoint in which points in $\mathbb{R}$ are regarded as infinite words in Γ_0 [20]. Furthermore, the finite admissible blocks which occur in these f-expansions give an elegant and well known solution to the word problem in $\mathrm{PSL}(2,\mathbb{Z})$ [20]: each finite admissible block is a shortest word relative to the generators Γ_0 and every element in $\mathrm{PSL}(2,\mathbb{Z})$ occurs as an admissible block in precisely one way.

This construction was generalised in [4] to the case of an arbitrary Fuchsian group Γ acting in the disc model of the hyperbolic plane with a given geometric set of generators Γ_0. (See [22] for the best exposition.) This involves the construction of a Markov map f on the boundary at infinity, the unit circle S^1. The elements of the partition were

intervals I_j, and, for each j, the restriction $f|_{I_j}$ was in Γ_0. The f-expansions carried full information about the Γ action on S^1, in the sense that two points were in the same Γ orbit if and only if the "tails" of their f-expansions agreed. Furthermore, and this is the point of interest here, these f-expansions simultaneously generate a most elegant solution to the word problem in Γ [22]. If to each partition interval is associated $f|_{I_j} \in \Gamma_0$, the f-expansions map to a set of infinite sequences in Γ_0. The finite admissible blocks in the f-expansion, give unique shortest representatives for words in Γ relative to the generators Γ_0. Clearly, this comes very close to saying they generate an automatic structure for Γ.

1.2 Automatic Groups

In this section we give the properties of automatic groups that will be used later. More general references to this and related material are the books of Epstein et al [7] and Holt [9], to which the reader is referred for more details. See also [10], [19].

An **alphabet** $\mathcal{A}$ is a finite set. A **language** $\mathcal{L}$ over an alphabet $\mathcal{A}$ is a collection of finite sequences of elements of $\mathcal{A}$ (called **words** or **strings**). The **length** of a string $w = (a_1, \ldots, a_n)$ is $|w| = n$.

For the purpose of this paper a **finite state automaton** over an alphabet $\mathcal{A}$ is a finite, directed, edge labelled graph whose vertices are called **states** and whose directed edges are called **arrows**. There is a specified state called the **start state** and a partition of the states into two disjoint sets, the **accept states** and the **non-accept states**. Every arrow from a state is labelled with a symbol from $\mathcal{A}$ and no two arrows from the same state have the same label. Given any string $w = (a_1, \ldots, a_n)$ over $\mathcal{A}$ and any state s there is at most one path of arrows starting at s so that the jth arrow is labelled with a_j. This path terminates at some state s'. We say that w goes from s to s'.

The language **accepted** by this automaton is the collection of strings w over $\mathcal{A}$ which go from the start state to some accept state. A language $\mathcal{L}$ over $\mathcal{A}$ is called **regular** if it is accepted by some finite state automaton over $\mathcal{A}$ and this automaton is said to **recognise** the language $\mathcal{L}$.

Let G be a group with identity element e. Consider an alphabet $\mathcal{A}$ and a map $\mathcal{A} \longrightarrow G$ denoted by $a \longmapsto \overline{a}$. This extends to a map from the collection of strings over $\mathcal{A}$ to G by $w = (a_1, \ldots, a_n) \longrightarrow \overline{w} = \overline{a}_1 \ldots \overline{a}_n$, the product of the image of the letters in w. If every element of G can be described in this way we call $\mathcal{A}$ a finite generating set for G. A language $\mathcal{L} = \mathcal{L}(G)$ over $\mathcal{A}$ is called an **automatic structure** for G if two conditions are satisfied. First, $\mathcal{L}(G)$ is a regular language which maps onto G. That is, there is a finite state automaton so that every element of G may be described by (at least) one path through this automaton. This automaton is called the **word acceptor** $\mathcal{W}(G)$. The second property is known as the fellow traveller property which we explain below. If a group G has an automatic structure then G is called an **automatic group**.

For a group G with finite generating set $\mathcal{A}$, the **word length** of $g \in G$ denoted $|g|$ with respect to $\mathcal{A}$ is the shortest length of any word in $\mathcal{A}$ representing g. The **word metric** on G is $d(g,h) = |g^{-1}h|$. Given a word $w = (a_1, \ldots, a_n)$ over $\mathcal{A}$, for each integer $0 \leq t \leq n$, denote by $w(t) = (a_1, \ldots, a_t)$ the prefix of w of length t, and for integers $t \geq n$ denote $w(t) = w$. Given a constant k, two words w, v over $\mathcal{A}$ are **k-fellow travellers** if $d\big(w(t), v(t)\big) \leq k$ for all $t \geq 0$. Also k is called the fellow traveller constant for w and v. The group G satisfies the **fellow traveller property** if there is a constant k such that for any words $w, v \in \mathcal{L}$ with $d(\overline{w}, \overline{v}) \leq 1$ then w and v are k-fellow travellers.

Let w and v be a pair of words as above and $a \in \mathcal{A} \cup \{e\}$ so that $\overline{wa} = \overline{v}$. If w and v are k-fellow travellers then $\psi(t) = w(t)^{-1}v(t)$ has length at most k for all $t \geq 0$. Thus for all choices of w and v with $d(\overline{w}, \overline{v}) \leq 1$ the $\psi(t)$ lie in a finite set $\mathcal{D}$, the collection of **word differences** and $\mathcal{A} \cup \{e\} \subset \mathcal{D} \subset \mathcal{L}$. Knowledge of the word differences allows us to reconstruct the multiplicative structure of G in an automated way.

More precisely, the fellow traveller property is equivalent to the existence of **multiplier automata** $\mathcal{M}_a$ for each $a \in \mathcal{A} \cup \{\$\}$ for G [7], [10]. Each $\mathcal{M}_a$ is a **2-stringed automaton** whose alphabet is $\mathcal{A}' \times \mathcal{A}'$, where A' is the **padded alphabet** $\mathcal{A} \cup \{\$\}$. It accepts the **padded pair** (w^+, v^+) for strings (w, v) over $\mathcal{A}$ whenever w, v are accept states of $\mathcal{W}(G)$ and $\overline{wa} = \overline{v}$. Here the symbols w^+, v^+ indicate that the padding symbol \$, which maps to the identity in G, may be added to the shorter of w, v to make them have equal length. The automaton $\mathcal{M}_\$$ recognises identity in G, replacing the condition $\overline{wa} = \overline{v}$ by $\overline{w} = \overline{v}$.

The multiplier automata $\mathcal{M}_a$ for $a \in \mathcal{A} \cup \{\$\}$ may be constructed by means of a **word-difference machine**, clearly explained in [10] and summarised here. This is really a collection of new automata, all of which have the same state space, namely the set of triples (s_1, s_2, ψ) such that s_1, s_2 are states of $\mathcal{L}(G)$ and $\psi \in \mathcal{D}$. The start state is (s_0, s_0, e) where s_0 is the start state of $\mathcal{L}(G)$ and e is the identity of G. For $a, b \in \mathcal{A}$ there is an arrow from (s_1, s_2, ψ) to (s_1', s_2', ψ') if and only if there are arrows $s_1 \xrightarrow{a} s_1'$ and $s_2 \xrightarrow{b} s_2'$ in the word acceptor and if $\overline{\psi'} = \overline{x^{-1}\psi y}$. In the automaton $\mathcal{M}_a$, the state (s_1, s_2, d) is a success state if s_1, s_2 are in $\mathcal{L}$ and if $\bar{\psi} = \bar{a}$.

There is an extra technicality needed to deal with the padding symbol. Namely, we have to add an extra state to the word acceptor $\mathcal{W}(G)$ which is reached when $\mathcal{W}(G)$ is in an accept state and the padding symbol is read. If either of s_1 or s_2 is this extra state, then one or other of x, y as above will be replaced by the padding symbols \$ and the condition $\overline{\psi'} = \overline{x^{-1}\psi y}$ will be replaced by $\overline{\psi'} = \overline{\psi y}$ or $\overline{\psi'} = \overline{x^{-1}\psi}$. Frequently we shall think of these conditions as commutative squares or triangles of relations between elements in the group.

If we can construct a word difference machine using a finite set of word differences $\mathcal{D}'$, then we have clearly verified the fellow traveller property and can use the above process to simultaneously construct all the multiplicative automata $\mathcal{M}_a$. The process can be seen

as concatenating squares or triangles to yield a collection of cross paths in $\mathcal{D}$ between all the prefixes $w(t)$ and $wa(t)$ occurring in the normal forms of any two words w and wa for $a \in \mathcal{A}$. The collection of all those squares and triangles which arise is easily seen to give a presentation for the group.

We can often use a Markov map to construct a word acceptor. This is analogous to Mosher's construction of a word acceptor by reversing the combing process [17]. Suppose that we have a (fixed point) free action of a group G on a space X and a Markov map f defined with respect to a partition $\{X_i\}$ of X so that on X_i the map f is some element of G. Suppose that there exists a particular $x \in X$ so that for each $g \in G$ there is a non-negative integer n so that $f^n(gx) = x$, and so that $n = 0$ if and only if g is the identity e [1]. Then we can use f to define a word acceptor for G as follows. There is a special start state corresponding to x and there is one state for each X_i in the Markov partition. For each X_i suppose that $f|_{X_i} = \alpha_i$ and $f|_{X_i} = \alpha_i \colon X_i \longrightarrow X_j \cup \cdots \cup X_k$. We draw an arrow from the each of the states $X_j, \ldots, X_k$ to X_i with the label ${\alpha_i}^{-1}$. This means that all the arrows arriving in each state have the same label and all arrows leaving each state have different labels. (Strictly speaking, there should be arrows leaving each state with every label in the alphabet. If there are letters in the alphabet that do not occur as labels leaving a particular state X_i then we draw arrows from X_i with these labels to a new state called the **fail state**. All arrows leaving the fail state return there. In practice we do not use the fail state and will omit all arrows leading there.) In order to read a normal form for $g \in G$ we consider the word in G obtained by inverting the composition of the particular values of f arising from $f^n(gx) = x$. This is the same as path through the word acceptor corresponding to g.

We shall find a suitable set of word differences $\mathcal{D}$ by starting from the generating set $\mathcal{A}$ and successively adding more words ψ' as dictated by the conditions $\overline{\psi'} = \overline{x^{-1}\psi y}$ until the collection we arrive at becomes closed under further moves of this kind. The method is similar to Mosher's construction of "raising bems" [17]. Since the states of the word acceptor are elements of the partition of X, the states of the difference machine are elements (X_i, X_j, ψ) for X_i, X_j in the Markov partition and $\psi \in \mathcal{D}$. The new relations will be of the form $\overline{\psi'} = \overline{{\alpha_i}^{-1}\psi\alpha_j}$ where $\alpha_i = f|_{X_i}$. We shall also allow degenerate squares or triangles corresponding to pairs of states (X_i, X_j, ψ), (X_i', X_j, e) with $\psi = f|_{X_i}$, giving the trivial relation $\overline{e} = \overline{\psi^{-1}\psi}$. During this process of adding new word differences, it will unfortunately sometimes be necessary to subdivide some of the states X_i. This is because the various word differences $\psi \in \mathcal{D}$ may map the state X_i to a number of different states on which the definition of f varies, thus possibly introducing several different variants of the

[1] This is very close to the property of **orbit equivalence**: f is said to be orbit equivalent to G on X if for any $x,y \in X$ then $x = gy$ for some $g \in G$ if and only if $f^n x = f^m y$ for some $m,n \geq 0$.

relation $\overline{\psi'} = \overline{\alpha_i^{-1}\psi\alpha_j}$. Technically, this means that we have to add new states (Y_{i_k}, Y_{j_l}, ψ) to the word difference automaton, where the new sets Y_{i_k}, Y_{j_l} are certain subsets of X_i and X_j. However, these difficulties are also resolved after a finite number of steps, and it should be clear that from the resulting collection of squares and triangles we can construct automata as required.

1.3 Multiple simple loops and π_1-train tracks.

In what follows we do not use the conventional Thurston theory of train tracks (for which see [18], [23]) but a variant due to Birman and Series [2]. We will only be concerned with punctured surfaces and there the theory is much easier. Thus we restrict our attention to this case. For details see also [12].

A **loop** on a surface Σ is a closed curve. A loop is called **simple** if it has no self intersections. A loop is **boundary parallel** or **peripheral** if it is homotopic to a loop around a puncture. A **multiple simple loop** is a collection of pairwise disjoint simple loops none of which is either homotopically trivial or boundary parallel. For the p times punctured torus Σ_p the maximal number of non-trivial homotopy classes of disjoint, non boundary parallel curves is p. Thus a multiple simple loop γ on Σ_p can be written as $m_1\gamma_1 + \cdots + m_p\gamma_p$ where m_j is a non-negative integer and the γ_j are distinct homotopy classes of simple closed curves on Σ_p.

For definiteness, fix a choice of hyperbolic structure on Σ and let $R \subset \mathbb{H}^2$ be a fundamental region for our surface whose vertices are all at punctures of the surface [1]. Suppose R has sides σ_k and side pairing maps $\mu_k\colon \sigma_k \longmapsto \sigma_{k'}$ where $\mu_{k'} = {\mu_k}^{-1}$ for each k. Let $\overline{R}$ be the closure of R in $\mathbb{H}^2$. A **π_1-train track** τ is a collection of pairwise disjoint arcs $\alpha_j\colon [0,1] \longrightarrow \overline{R}$ so that

(i) $\alpha_j(0) \in \sigma_k$ and $\alpha_j(1) \in \sigma_l$,
(ii) $\alpha_j(\lambda) \in R^\circ$ for $\lambda \in (0,1)$,
(iii) at most one arc joins each pair of sides.
(iv) no arc goes from one side to itself. That is, if k and l are as in (i) then $k \neq l$.

An arc of τ is called a **corner arc** if it joins adjacent sides of R. Each corner arc faces a particular vertex of R and for each vertex cycle in the side pairing of R we have the corresponding **corner cycle** consisting of all corner branches corresponding to the same puncture.

A **weighting** w on a π_1-train track τ is an assignment of a non-negative number $w(\alpha_j)$ to each arc α_j of τ. A weighting is **integral** if each weight is a (non-negative) integer. We

[1] Our results are combinatorial in nature and hence independent of the particular hyperbolic structure chosen. Nevertheless, since the theory of π_1-train tracks involves hyperbolic geometry, some choice needs to be made.

define the **length** of w, denoted $|w|$, as $|w| = \sum w(\alpha_j)$ where the sum is over all arcs α_j of τ.

We now explain how to collapse a multiple simple loop to obtain a π_1-train track with an integral weighting. We begin by lifting the multiple simple loop γ to the fundamental region R. The multiple simple loop becomes a collection of arcs, called **strands**, joining sides of R. We say that a multiple simple loop γ is **supported** on a π_1-train track τ if, for every strand of γ there is an arc of τ joining the same pair of sides. If γ is supported on τ we may give τ an integral weighting w_γ by assigning to each arc of τ the number of strands of γ joining that pair of sides. This weighting has the following properties (see [2], [12]):

(i) For each side pairing $\mu_k\colon \sigma_k \longrightarrow \sigma_{k'}$, the sum of the weights of arcs with endpoints on σ_k is the same as the sum of the weights of arcs with endpoints on $\sigma_{k'}$.

(ii) At least one arc in each corner cycle must have weight zero.

The first condition holds because when we perform the gluing coming from μ_k each endpoint of a strand of γ on σ_k is identified with the endpoint of a strand on $\sigma_{k'}$. Thus the total numbers of endpoints on this pair of sides are the same. The second condition holds because, if not, the strands in corner cycle would join up to give a peripheral loop in γ.

If a (non-negative but not necessarily integral) weighting satisfies (i) and (ii) we call it a **proper weighting**. Conversely, every proper integral weighting w on a π_1-train track τ gives rise to a multiple simple loop γ. This means that in order to study multiple simple loops it is sufficient to study proper integral weightings on π_1-train tracks. Let $W(\tau)$ denote the collection of all proper weightings on the π_1-train track τ and $W_{\mathcal{O}}(\tau)$ the collection of proper integral weightings on τ (see [12]).

1.4 Irreducible loops and $\mathcal{PML}$.

A π_1-train track τ is said to be **recurrent** (see [23]) if there exists a proper integral weighting $w \in W_{\mathcal{O}}(\tau)$ so that $w(\alpha_j)$ is non-zero for all branches α_j of τ. Such a π_1-train track τ is said to be **maximal** if there does not exist a recurrent π_1-train track τ' so that τ is properly contained in τ' in the obvious sense. It follows from Thurston's theory, or as one can directly verify in the special cases of concern to us here, that if τ is a maximal recurrent train track then the dimension of $W(\tau)$ is $6g-6+2p$ where Σ is a surface of genus g with p points removed. We call the collection of all proper weightings $W(\tau)$ on a maximal recurrent π_1-train track τ a **maximal cell**.

Any simple loop γ defines a recurrent π_1-train track $\tau(\gamma)$ with weights $w(\gamma)$ as above. A simple loop γ is said to be **irreducible** if $w(\gamma) \neq w_1 + w_2$ for any $w_1, w_2 \in W_{\mathcal{O}}\big(\tau(\gamma)\big)$ and $w_j \neq 0$ for $j = 1, 2$. Clearly there are only finitely many maximal recurrent π_1-train tracks. We shall see below that if Σ_p is the p-times punctured torus, for $p = 1, 2$, each maximal cell is the linear span of $2p$ irreducible loops.

We denote the collection of all homotopy classes of multiple simple, non-boundary parallel loops on Σ by $\mathcal{ML}_{\mathcal{O}}(\Sigma)$ and the collection of all measured geodesic laminations on Σ by $\mathcal{ML}(\Sigma)$. It is a theorem of Birman and Series [2] that $\mathcal{ML}_{\mathcal{O}}(\Sigma)$ and $\mathcal{ML}(\Sigma)$ can be identified with the collections of proper integral weightings and proper weightings respectively on π_1-train tracks on Σ. For $\mathcal{ML}_{\mathcal{O}}(\Sigma)$ the proof of this follows the outline given above. If $w \in \mathcal{ML}(\Sigma)$ then clearly w is contained in some maximal cell $W(\tau)$. Thus $\mathcal{ML}(\Sigma)$ is the union of maximal cells $W(\tau_i)$ where τ_i runs over all maximal recurrent π_1-train tracks on Σ. This gives $\mathcal{ML}(\Sigma)$ a natural cell structure. In the cases we are interested in, namely Σ_1 and Σ_2, we shall prove (Propositions 2.1.1 and 3.1.2) the following result.

Proposition 1.4.1. *For $p = 1, 2$ let Σ_p denote the p times punctured torus. There are finitely many irreducible loops $\mathbf{e}_1, \ldots, \mathbf{e}_k$ on Σ_p so that for each maximal π_1-train track τ the corresponding maximal cell $W(\tau)$ is the positive linear span of $2p$ irreducible loops: $W(\tau) = \mathrm{sp}^+\{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{2p}}\}$ where $i_j \in \{1, \ldots, k\}$. Also the intersection of two cells is*

$$W(\tau) \cap W(\tau') = \mathrm{sp}^+\Big(\{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{2p}}\} \cap \{\mathbf{e}_{i'_1}, \ldots, \mathbf{e}_{i'_{2p}}\}\Big)$$

where

$$W(\tau) = \mathrm{sp}^+\{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{2p}}\} \text{ and } W(\tau') = \mathrm{sp}^+\{\mathbf{e}_{i'_1}, \ldots, \mathbf{e}_{i'_{2p}}\}.$$

The space $W(\tau)$ may be projectivised in a natural way to obtain $\mathbb{P}W(\tau)$ and similarly $W_{\mathcal{O}}(\tau)$ can be projectivised to obtain the set of rational weightings $\mathbb{P}W_{\mathcal{O}}(\tau)$. The space $\mathcal{PML} = \mathcal{PML}(\Sigma)$ is the union over all τ of the corresponding cones $\mathbb{P}W(\tau)$, which we call π_1-cones, glued along their lower dimensional common simplices as in the proposition. We denote the union of all rational weightings $\mathbb{P}W_{\mathcal{O}}(\tau)$ by $\mathcal{PML}_{\mathcal{O}}(\Sigma)$. Using the Birman-Series identification, the space $\mathcal{PML}(\Sigma)$ can be naturally identified with the space of

projective measured laminations on Σ, shown by Thurston to be a sphere of dimension $6g-7+2p$ [18]. Thus in our two examples, we expect $\mathcal{PML}(\Sigma_1)$ and $\mathcal{PML}(\Sigma_2)$ to be S^1 and S^3 respectively. In each example we shall first determine the maximal cones. Gluing up using Proposition 1.4.1 will allow us to see explicitly how the spheres S^1 and S^3 are formed.

1.5 Dehn twists and the mapping class group

The (orientation preserving) **mapping class group** $\mathcal{MCG} = \mathcal{MCG}(\Sigma)$ of Σ is the group of isotopy classes of (orientation preserving) automorphisms of Σ, see [1] or [17] for example. That is, an element of $\mathcal{MCG}$ is an (orientation preserving) homeomorphism of Σ to itself and two such homeomorphisms give the same element of $\mathcal{MCG}$ if one can be deformed to the other isotopically along a continuous path of homeomorphisms of Σ to itself. There is a natural action of the mapping class group on Teichmüller space of Σ as the Teichmüller modular group. This action can be extended to $\mathcal{PML}(\Sigma)$, [24], and it is this action we will consider here.

Let w be a simple closed curve on a surface Σ parametrised by $\xi \in [0, l_w]$ where l_w is the length of w. Consider a small tubular neighbourhood around w in Σ and denote this by $N_w = [0,1] \times w$. We define a homeomorphism of Σ called (left) **Dehn twist about** w (see [1] for example) denoted δ_w as the identity on $\Sigma - N_w$ and by requiring that $(\eta, \xi) \in [0,1] \times w$ is mapped by δ_w to $(\eta, \xi - \eta l_w)$ where $\eta \in [0,1]$ and $\xi - \eta l_w$ is defined mod l_w. Observe that if $\eta = 0$ or 1 then δ_w is the identity. By a well known result of Dehn [6], the (orientation preserving) mapping class group is generated by Dehn twists.

We will produce a set of Dehn twists which we shall show are generators for the mapping class groups of Σ_1 and Σ_2. We will then investigate the action of these Dehn twists on the piecewise linear structure on $\mathcal{PML}(\Sigma_p)$ given by π_1-train tracks. In particular we show that the Dehn twists act piecewise linearly on $\mathcal{PML}(\Sigma_p)$ with respect to this piecewise linear structure. Moreover, the action also restricts to an action on $\mathcal{PML}_{\mathcal{O}}(\Sigma_p)$. This gives a piecewise linear action of $\mathcal{MCG}$ on $\mathcal{PML}$ and $\mathcal{PML}_{\mathcal{O}}$ respectively.

This action is not free in the sense that there are elements of $\mathcal{PML}_{\mathcal{O}}$ which have non-trivial stabilisers in $\mathcal{MCG}$. For example, performing a Dehn twist about w fixes w and any curve disjoint from w. In order to construct a Markov map whose orbits describe $\mathcal{MCG}$ we need to find a (fixed point) free action of $\mathcal{MCG}$ on a suitable space Y. For the case of the once punctured torus, Y will be the space of (ordered) Farey neighbours, that is, pairs of curves which intersect exactly once. For the twice punctured torus we will generalise this idea by defining quadruples of curves in a special topological configuration which we call Farey blocks. The space Y of Farey blocks will admit a free action of $\mathcal{MCG}(\Sigma_2)$.

2. The mapping class group of the once punctured torus.

In this section we carry out the construction of π_1-train tracks, a Markov map and an automatic structure for the once punctured torus. Much of the material in this section is, to some extent, well known. However, we shall adopt a non-standard view point. The reasons for including this section are two-fold. First, the main structure of the argument is the same as for the twice punctured torus. Thus it will serve as motivation and a guide for what follows, explaining the main ideas with computations of a much more manageable scale. Secondly, when we are dealing with the twice punctured torus there are several steps in the construction of the Markov map and automatic structure. One of these steps is essentially the construction we present in this section. This will save us considerable effort later on.

In sections 2.1 and 2.2 we show that certain elementary Dehn twists act on $\mathcal{ML}(\Sigma_1)$ exactly like the continued fraction map on $\mathbb{R} \cup \{\infty\}$. In section 2.3 we construct the Markov map and in sections 2.4 and 2.5 we explain how it gives the automatic structure for $\mathcal{MCG}(\Sigma_1) = \mathrm{PSL}(2, \mathbb{Z})$.

2.1. π_1-train tracks and the cell structure of $\mathcal{ML}(\Sigma_1)$.

As remarked in section 1.3, we start by fixing a definite hyperbolic structure for Σ_1 and a fundamental domain $R_1 \subset \mathbb{H}^2$ for the action of $\pi_1(\Sigma)$ on the hyperbolic plane.

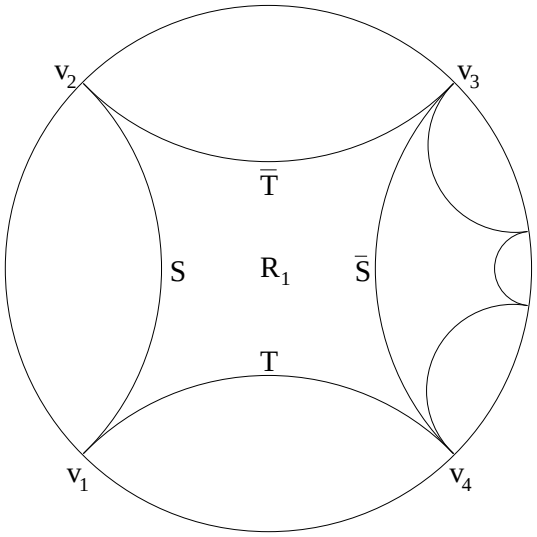


Fig. 2.1.1. A hyperbolic fundamental domain R_1 for Σ_1, where $\overline{S}$, $\overline{T}$ denote S^{-1}, T^{-1}.

The fundamental domain R_1 we choose is the standard rectangular one with opposite sides identified by side pairings which match the midpoints of the sides. The region R_1 has four vertices all of which project to the puncture of Σ_1 (see Fig. 2.1.1). We label these $v_1, \ldots, v_4$ in clockwise order. Writing $v_i v_j$ for the side joining v_i to v_j, the side pairings

will be S carrying v_1v_2 to v_4v_3 and T carrying v_1v_4 to v_2v_3. The maps S and T correspond to homotopy classes of simple closed curves which freely generate the fundamental group.

We now introduce the irreducible loops that will form the basis for $\mathcal{ML}(\Sigma_1)$ as explained in Proposition 1.4.1. They are defined as follows (see Fig. 2.1.2. [1]):

$\mathbf{e}_0$ consists of a single arc joining v_1v_2 and v_3v_4;
$\mathbf{e}_1$ consists of an arc joining v_1v_2 and v_2v_3 and an arc joining v_3v_4 and v_4v_1;
$\mathbf{e}_\infty$ consists of a single arc joining v_2v_3 and v_4v_1;
$\mathbf{e}_{-1}$ consists of an arc joining v_2v_3 and v_3v_4 and an arc joining v_4v_1 and v_1v_2.

We mention in passing that one may also define these loops in terms of the cutting sequences as discussed in [3] or [22]: $\mathbf{e}_0 = S$, $\mathbf{e}_1 = ST$, $\mathbf{e}_\infty = T$ and $\mathbf{e}_{-1} = S^{-1}T$. (Since the loops are unoriented, strictly speaking $\mathbf{e}_0 = S$ or S^{-1} and so on.)

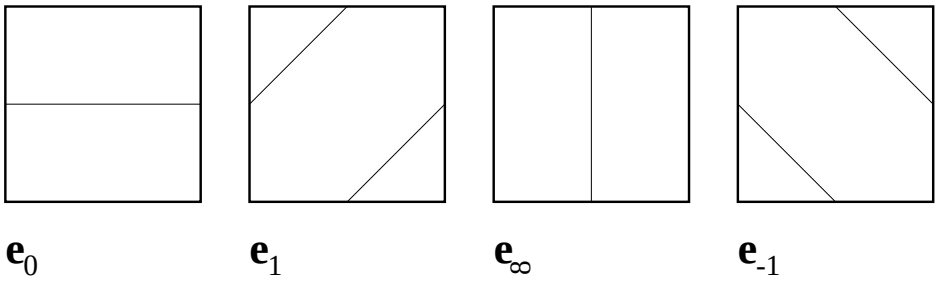


Fig. 2.1.2. The elementary π_1-train tracks.

Next, we define cells in $\mathcal{ML}(\Sigma_1)$. We show below that these are maximal. The cells are:

$$I_0 = \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}_1\}, \quad I_1 = \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}_{-1}\}, \quad I_2 = \mathrm{sp}^+\{\mathbf{e}_\infty, \mathbf{e}_{-1}\}, \quad I_3 = \mathrm{sp}^+\{\mathbf{e}_\infty, \mathbf{e}_1\}.$$

Proposition 2.1.1. *The cells I_0, I_1, I_2, I_3 are maximal and their union is $\mathcal{ML}(\Sigma_1)$.*

Proof. It is sufficient to show that no extra arcs can be added to any of these four π_1-train tracks, and that any loop is supported on one of them. This is carried out in the appendix to [2]. For convenience we reproduce it here.

Clearly, any maximal π_1-train track on Σ_1 must have one of the two forms illustrated in Fig. 2.1.3. Summing the weights over the two pairs of identified sides and cancelling a we obtain two equations

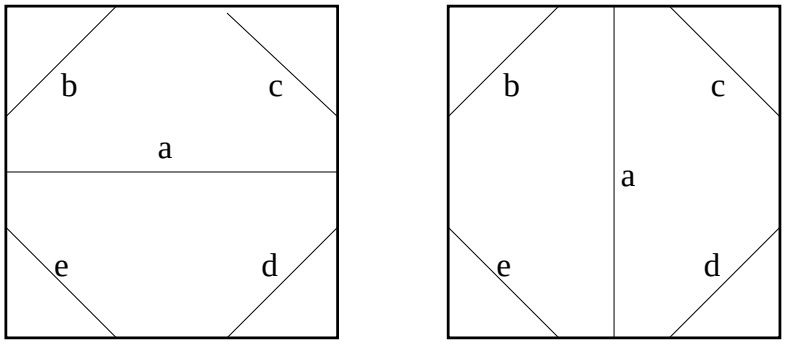
$$b + e = c + d, \quad b + c = e + d.$$

We may solve these to obtain $b = d$ and $c = e$. Now we know that on the corner cycle we cannot have all the weights non-zero. Thus $b = 0$ or $c = 0$. Since all the weights are

[1] For simplicity we draw R_1 as a Euclidean rectangle in what follows.

non-negative, this means that there are four configurations of π_1-train track corresponding to non-boundary parallel, simple loops on Σ_1. □

Notation 2.1.2. By Proposition 2.1.1 any simple closed curve γ may be represented as $a\mathbf{e}_i + b\mathbf{e}_j$ where $w(\gamma) \in I_k = \mathrm{sp}^+\{\mathbf{e}_i, \mathbf{e}_j\}$ for some $k = 0, 1, 2, 3$. *We always write the ordered pair* (a, b) *to represent* $a\mathbf{e}_0 + b\mathbf{e}_1$ *if* $w(\gamma) \in I_0$, $a\mathbf{e}_0 + b\mathbf{e}_{-1}$ *if* $w(\gamma) \in I_1$, $a\mathbf{e}_\infty + b\mathbf{e}_{-1}$ *if* $w(\gamma) \in I_2$ *or* $a\mathbf{e}_\infty + b\mathbf{e}_1$ *if* $w(\gamma) \in I_3$.



(i) (ii)

Fig. 2.1.3. The two possible configurations for a maximal weighted π_1-train track on R_1.

The notation for the $\mathbf{e}_j$ has the following rationale. Regard R_1 as a square with v_1 in the bottom left hand corner, and side pairings which are Euclidean translations. By Proposition 2.1.1 above, any simple closed curve γ on Σ_1 is supported on one of the four maximal cells I_k. Thus, up to homotopy, γ is equivalent to a family of parallel Euclidean straight lines across R_1. Lifting to the Euclidean universal cover of R_1, that is $\mathbb{R}^2$, such a family of lines links to form a line of rational slope on the plane. With this identification, it is clear that the curve we have labelled $\mathbf{e}_j$ has slope j.

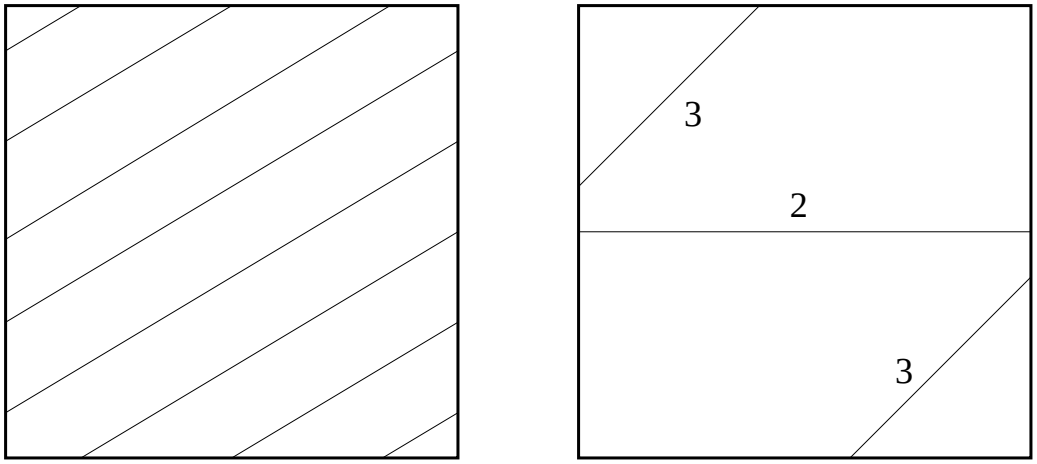


Fig. 2.1.4. The line of slope 3/5 drawn on R_1 and as a weighted π_1-train track.

More generally, we obtain an identification of $\mathcal{PML}(\Sigma_1)$ with the extended real line $\mathbb{R} \cup \{\infty\}$ by mapping $(a, b) \in I_0$ to the point $b/(a + b)$, $(a, b) \in I_1$ to $-b/(a + b)$, $(a, b) \in I_2$ to $-(a + b)/b$ and $(a, b) \in I_3$ to $(a + b)/b$. An example, the curve represented by $(2, 3) = 2\mathbf{e}_0 + 3\mathbf{e}_1 \in I_0$ is shown in Fig. 2.1.4. This corresponds to the line of slope $3/5$ in $\mathbb{R}^2$.

The maximal cells I_0, I_1, I_2, I_3 have their boundaries identified as in Proposition 1.4.1. In this case it is easy to see that $I_0 \cap I_1 = \mathrm{sp}^+\{\mathbf{e}_0\}$, $I_1 \cap I_2 = \mathrm{sp}^+\{\mathbf{e}_{-1}\}$, $I_2 \cap I_3 = \mathrm{sp}^+\{\mathbf{e}_\infty\}$ and $I_3 \cap I_0 = \mathrm{sp}^+\{\mathbf{e}_0\}$. The other two intersections are empty. This is illustrated in Fig. 2.1.5, from which one clearly sees that $\mathcal{PML}(\Sigma_1) \sim S^1$. We remark that in the appendix to [2], Birman and Series considered oriented curves and so found a different cell structure for $\mathcal{ML}(\Sigma_1)$.

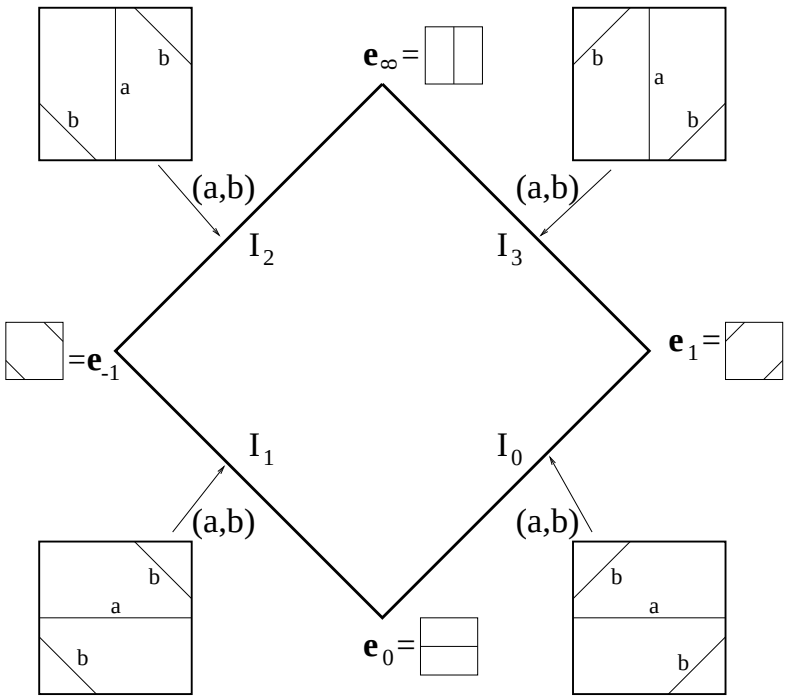


Fig. 2.1.5. The partition of $\mathcal{ML}(\Sigma_1)$ into maximal cells.

2.2 Dehn twists and the mapping class group

Let the Dehn twists about $\mathbf{e}_\infty$ and $\mathbf{e}_0$ be denoted by δ_0 and δ_1 respectively. We now want to investigate the effects of these twists on the projective structure on the space of train tracks constructed in the previous section.

In order to simplify things we will make use of some natural symmetries of R_1 and the π_1 train tracks we constructed above. These symmetries are defined as follows:

ι_1 interchanges the pairs (v_1, v_4), (v_2, v_3);

ι_2 cyclically permutes the vertices sending v_1 to v_2, v_2 to v_3 and so on;

ι_3 fixes v_1 and v_3 and interchanges (v_2, v_4).

In addition, we will sometimes write ι_0 for the identity map.

Observe that applying ι_2 twice we get a rotation of R_1 by 180° which interchanges v_1, v_3 and v_2, v_4. Even though this is not the identity on R_1 it does act as the identity on each of the I_j. (This map is just the map which sends any curve to itself with the opposite orientation.) The ι_j act on $\mathcal{PML}(\Sigma_1)$ as the Klein 4-group. On these irreducible loops this action is given by:

$$\begin{array}{llll} \iota_1\colon \mathbf{e}_0 \longmapsto \mathbf{e}_0, & \mathbf{e}_1 \longmapsto \mathbf{e}_{-1}, & \mathbf{e}_\infty \longmapsto \mathbf{e}_\infty, & \mathbf{e}_{-1} \longmapsto \mathbf{e}_1, \\ \iota_2\colon \mathbf{e}_0 \longmapsto \mathbf{e}_\infty, & \mathbf{e}_1 \longmapsto \mathbf{e}_{-1}, & \mathbf{e}_\infty \longmapsto \mathbf{e}_0, & \mathbf{e}_{-1} \longmapsto \mathbf{e}_1, \\ \iota_3\colon \mathbf{e}_0 \longmapsto \mathbf{e}_\infty, & \mathbf{e}_1 \longmapsto \mathbf{e}_1, & \mathbf{e}_\infty \longmapsto \mathbf{e}_0, & \mathbf{e}_{-1} \longmapsto \mathbf{e}_{-1}. \end{array}$$

Note that the action of the symmetries extends naturally to the cells I_j. Moreover, the action is given by $\iota_j(I_0) = I_j$.

The benefit of applying these symmetries is that we only need consider the action of δ_0 on $\mathcal{ML}(\Sigma_1)$. The action of ${\delta_0}^{-1}$ and ${\delta_1}^{\pm 1}$ will follow by symmetry as follows. We claim that

$$\begin{array}{lll} \iota_1\delta_0\iota_1 = {\delta_0}^{-1} & \iota_2\delta_0\iota_2 = \delta_1 & \iota_3\delta_0\iota_3 = {\delta_1}^{-1} \\ \iota_1\delta_1\iota_1 = {\delta_1}^{-1} & \iota_2\delta_1\iota_2 = \delta_0 & \iota_3\delta_1\iota_3 = {\delta_0}^{-1}. \end{array}$$

This is because ι_1 and ι_3 reverse orientation and so conjugate right Dehn twists to left Dehn twists, while ι_2 preserves orientation but interchanges $\mathbf{e}_0$ and $\mathbf{e}_\infty$.

Applying a Dehn twist to a weighted π_1-train track sometimes results in an **unreduced** π_1-train track, that is a π_1-train track which may have arcs with both ends on the same edge of the fundamental domain. An unreduced π_1-train track satisfies conditions (i)–(iii) given in section 1.3 but fails to satisfy (iv). The process of converting an unreduced (weighted) π_1-train track into a (reduced) π_1-train track is called **pulling tight**. Suppose that the unreduced π_1-train track τ has an arc α from the side σ_k to itself and that this arc has weight $w(\alpha)$. Suppose also that τ has a proper integral weighting. We begin by converting it into a multiple simple loop γ on R_1. This means that we replace each arc α_j with weight $w(\alpha_j)$ by $w(\alpha_j)$ strands joining the same pair of sides as α_j. In particular we have $w(\alpha)$ strands from σ_k to itself. We now perform a homotopy of Σ which will remove

the intersections of all these strands with σ_k. This is done as follows. We can always choose an innermost strand β, which together with an arc of σ_k bounds a disc in R_1 containing no other strands. Suppose that the endpoints of β on σ_k are x_+ and x_-. Now consider the images of x_+ and x_- under the side pairing map μ_k. These are points of $\sigma_{k'}$ that are ends of strands β_+ and β_- respectively. (To find out which strands, put an orientation on σ_k and $\sigma_{k'}$ consistent with μ_k and then count endpoints from the corresponding ends of σ_k and $\sigma_{k'}$.) The other endpoints of β_+ and β_- are points y_+ and y_- on sides σ_+ and σ_-. We replace β, β_+ and β_- by a single strand β' from y_+ to y_-. It is clear that this strand can be drawn disjoint from the other strands of γ. We have reduced the number of strands by two. This process clearly terminates after a finite number of applications, giving a multiple simple loop which has no strands with both endpoints on σ_k. Repeating for all k gives a reduced π_1-train track on R_1 with a proper integral weighting.

Proposition 2.2.1. *Let $(a,b) \in I_0$. The Dehn twists act on I_0 as follows*

$$\begin{aligned}
\delta_0(a,b) &= (b, a+b) \in I_3 \\
{\delta_0}^{-1}(a,b) &= (b,a) \in I_1 \\
\delta_1(a,b) &= \begin{cases} (a-b,b) \in I_0 & \text{if } a \geq b \\ (b-a,a) \in I_3 & \text{if } a \leq b \end{cases} \\
{\delta_1}^{-1}(a,b) &= (a+b,b) \in I_0.
\end{aligned}$$

Proof. This follows directly from the linearity theorems in [2]. We include an alternative proof as an illustration of how we manipulate π_1-train tracks. It is illustrated in Fig. 2.2.1. We begin with the train track for a general integral point $(a,b) \in I_0$. We want to perform the Dehn twist $\delta_j^{\pm 1}$, for $j = 0, 1$, about the curve γ_j which is either $\mathbf{e}_\infty$ or $\mathbf{e}_0$ respectively. We draw a tubular neighbourhood about γ_j as a strip going from one side to the opposite side. This strip is bounded by dotted lines in the figure. The Dehn twist is the identity outside this strip and inside the strip fixes one boundary component of the cylinder about γ_j while rotating the other component one whole turn. In between we interpolate linearly so that an arc of a train track which went straight across is now wrapped once around the cylinder before emerging on the other side. It still carries the same weight which represents the number of strands in the corresponding multiple simple loop. In the cases illustrated in the top or bottom diagrams in Fig. 2.2.1, all that remains is to gather together arcs whose endpoints lie on the same sides and add their weights.

In the middle two cases in Fig. 2.2.1 we need an intermediate step. Namely the image train track is unreduced so that we need to pull tight. In the first case, that is ${\delta_0}^{-1}(a,b)$, the unreduced arc joins the bottom side to itself and has weight b. Convert the train track into a multiple simple loop by drawing $w(\alpha)$ strands joining the same pair of sides as α for each arc α with weight $w(\alpha)$. This gives b strands joining the bottom side to itself.

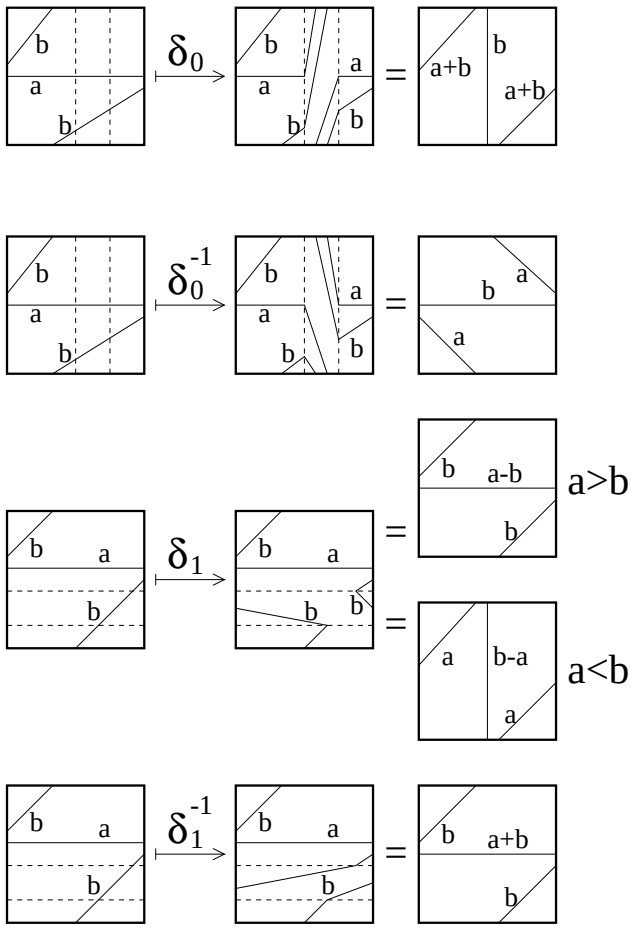


Fig. 2.2.1. The action of Dehn twists on points of I_0.

Reading from the left, the first b endpoints of strands along the bottom edge are joined to the next b strands in the reverse order. Likewise reading from the left, the first b strands on the top side are ends of strands all joining the left hand side and the next b strands all join the right hand side. Thus we may pull all b simple loops tight at once by replacing all $3b$ of these strands by b strands joining the left and right sides. The result after converting back to a π_1-train track is shown in the right hand column.

For $\delta_1(a, b)$ we perform the same process but the result is slightly more complicated. Now we have an arc of weight b joining the right hand side to itself. When we convert to a multiple simple loop this arc becomes b strands which, reading from the bottom, are the first b strands on the right hand side. These are joined to the next b strands on the same side. On the left hand side the first b strands from the bottom have their other endpoint on the bottom side. However we need to take care when finding the next b strands. If $a \geq b$ these next b strands join the right hand side and we may pull these loops tight to obtain b strands joining the bottom side to the right hand side. There were $a - b$ strands joining the left and right sides which we have not used and these remain after pulling tight. In

the case where $a \leq b$ we can pull a strands tight in this way. There remain $b-a$ strands joining the right hand side to itself. Their other ends join the left and top sides. Thus after pulling tight we obtain $b-a$ strands joining the top and bottom sides and a remain joining the left and top sides. After reconverting to π_1-train tracks one obtains the result, again shown in the right hand column.

To complete the proof for general weightings (a, b), note that we can clearly obtain the result for proper rational weightings by clearing denominators in $\mathcal{ML}_{\mathcal{O}}(\Sigma_1)$. The result for general $w(\alpha)$ follows by continuity. □

It is well known that the Dehn twists δ_0 and δ_1 generate the mapping class group $\mathcal{MCG}(\Sigma_1)$. In fact we have $\iota_2 = \delta_1\delta_0\delta_1$. Together with $\iota_2\delta_1\iota_2 = \delta_0$ and the fact that ι_2 has order 2 this immediately gives

$$\delta_1\delta_0\delta_1 = \delta_0\delta_1\delta_0, \quad (\delta_0\delta_1)^3 = e.$$

It turns out that this gives a presentation of $\mathcal{MCG}(\Sigma_1)$:

$$\mathcal{MCG}(\Sigma_1) = \langle \delta_0, \delta_1 | \delta_1\delta_0\delta_1 = \delta_0\delta_1\delta_0, (\delta_0\delta_1)^3 = e \rangle.$$

This may be seen either using standard facts about the modular group or can be deduced from the automatic structure given below, see section 2.5.

In order to obtain the identification of $\mathcal{MCG}(\Sigma_1)$ with $\mathrm{PSL}(2, \mathbb{Z})$, consider the identification of $\mathcal{PML}(\Sigma_1)$ with $\mathbb{R} \cup \{\infty\}$ given in section 2.1. It is easy to see that after making this identification, the action of δ_0 and δ_1 on $\mathbb{R} \cup \{\infty\}$ is given by

$$\delta_0 \colon x \longmapsto x/(-x+1), \quad \delta_1 \colon x \longmapsto x+1.$$

Notice that this is essentially the same as the continued fraction map explained in section 1.1. We see that $\iota_2 \colon x \longmapsto -1/x$. If we had been considering oriented curves as in the appendix of [2] it is clear that ${\iota_2}^2$ would fix each non-trivial, non-peripheral curve but reverse its orientation. This corresponds to the matrix $-I$ and we would have obtained an action of $\mathrm{SL}(2, \mathbb{Z})$ rather than $\mathrm{PSL}(2, \mathbb{Z})$.

2.3 The Markov map and Farey pairs.

We will now define a Markov map f_1 on $\mathcal{ML}(\Sigma_1)$ from which we shall construct the desired automatic structure for $\mathcal{MCG}(\Sigma_1)$. The Markov partition of $\mathcal{ML}(\Sigma_1)$ will consist of the four maximal cells I_0, I_1, I_2 and I_3. These cells are closed and therefore intersect along their boundaries. This gives rise to ambiguities, but this will not present a problem. The map $f_1|_{I_j}$ will be chosen from $\{{\delta_0}^{\pm 1}, {\delta_1}^{\pm 1}\}$ in such a way that f_1 has the required Markov property.

Lemma 2.3.1. $\delta_1(I_0) = I_0 \cup I_3$.

Proof. By Proposition 2.2.1 we see that $\delta_1(I_0) \subset I_0 \cup I_3$. Also ${\delta_1}^{-1}(I_0) \subset I_0$ and ${\delta_1}^{-1}(I_3) = \iota_3\delta_0\iota_3(\iota_3 I_0) = \iota_3\delta_0(I_0) \subset \iota_3(I_3) = I_0$. This gives the result. □

We now define $f_1|_{I_0} = \delta_1$ and $f_1|_{I_j}$ by symmetry. In summary f_1 is defined as

$$\begin{aligned} f_1|_{I_0} &= \delta_1 \quad : I_0 \longmapsto I_0 \cup I_3 \\ f_1|_{I_1} &= {\delta_1}^{-1} : I_1 \longmapsto I_1 \cup I_2 \\ f_1|_{I_2} &= \delta_0 \quad : I_2 \longmapsto I_1 \cup I_2 \\ f_1|_{I_3} &= {\delta_0}^{-1} : I_3 \longmapsto I_0 \cup I_3. \end{aligned}$$

On a boundary $I_i \cap I_j$ the map is considered to be two valued. By Lemma 2.3.1 the map f satisfies the Markov property of section 1.1.

The following lemma will be crucial for constructing the automaton. Recall the definition of length given in section 1.3. In this case we can see by direct inspection that if $w = (a, b) \in I_j$ then $|w| = a + 2b$ for $j = 0, 1, 2, 3$.

Lemma 2.3.2. *Let $w \in \mathcal{ML}_{\mathcal{O}}(\Sigma_1)$ be a proper integral weighting on a π_1-train track τ. Then $|f_1(w)| \leq |w|$ with equality if and only if $w = a\mathbf{e}_0$ or $a\mathbf{e}_\infty$ for $a \in \mathbb{N}$.*

Proof. This is easy to check from Proposition 2.2.1. □

The rough idea of the construction of the word acceptor for $\mathcal{MCG}(\Sigma_1)$ is to use the four cells I_0, I_1, I_2, I_3 as states and to define arrows using the transition matrix associated to the Markov map f_1. However there is a problem with this idea, namely $\mathcal{MCG}(\Sigma_1)$ does not act freely on $\mathcal{ML}(\Sigma_1)$. In other words, $\phi\gamma = \gamma$ for $\phi \in \mathcal{MCG}(\Sigma_1)$ and γ a simple loop on Σ_1 does not imply that ϕ is the identity. We therefore need to consider the action of $\mathcal{MCG}(\Sigma_1)$ on a space of slightly more elaborate objects on which the action is fixed point free. To this end, we introduce the notation of a Farey pair.

Two (homotopy classes of) simple closed curves on Σ_1 are called **Farey neighbours** if they (have representatives that) intersect exactly once. Notice that this condition automatically implies that these curves do not divide the punctured torus so neither of them can

be boundary parallel or homotopically trivial. We consider ordered pairs of Farey neighbours (γ_1, γ_2) which we refer to as **Farey pairs**. It is clear that $(\mathbf{e}_0, \mathbf{e}_\infty)$ and $(\mathbf{e}_\infty, \mathbf{e}_0)$ are both Farey pairs. We denote the set of all Farey pairs by $\mathcal{F}$.

It is easy to see the Farey pair $(\mathbf{e}_0, \mathbf{e}_\infty)$ has trivial stabiliser in $\mathcal{MCG}(\Sigma_1)$ and that, for any other Farey pair (γ_1, γ_2), there is an element ϕ of $\mathcal{MCG}(\Sigma_1)$ sending it to $(\mathbf{e}_0, \mathbf{e}_\infty)$. As $(\mathbf{e}_0, \mathbf{e}_\infty)$ has trivial stabiliser this element is unique. In particular, ι_2 sends $(\mathbf{e}_\infty, \mathbf{e}_0)$ to $(\mathbf{e}_0, \mathbf{e}_\infty)$. Fortunately, the notion of Farey neighbours is compatible with the cell structure of $\mathcal{ML}(\Sigma_1)$ in the following sense.

Proposition 2.3.3. *Let (γ_1, γ_2) be a pair of Farey neighbours. If $\{\gamma_1, \gamma_2\} \neq \{\mathbf{e}_0, \mathbf{e}_\infty\}$ then γ_1 and γ_2 are both contained in I_j for some $j = 0, 1, 2, 3$.*

Proof. The easiest way to see this is to use the well known fact that, using the identification of $\mathcal{PML}(\Sigma_1)$ with $\mathbb{R} \cup \{\infty\}$ given in section 2.1, a pair of Farey neighbours corresponds to a pair of rational numbers p/q and r/s with $ps - qr = \pm 1$, see [20]. Provided we have $\{\pm p/q, \pm r/s\} \neq \{0 = 0/1, \infty = 1/0\}$, it is clear that p/q and r/s are both contained in one of the intervals $[-\infty, -1]$, $[-1, 0]$, $[0, 1]$ or $[1, \infty]$. The result follows from the discussion in section 2.1. □

On a cell I_j, the map f_1 is constantly equal to a fixed element α_j of $\mathcal{MCG}(\Sigma_1)$ with possible ambiguity at the endpoints. Proposition 2.3.3 allows us to extend the action of f_1 to $\mathcal{F} - \big\{(\mathbf{e}_0, \mathbf{e}_\infty), (\mathbf{e}_\infty, \mathbf{e}_0)\big\}$ by defining $f_1(\gamma_1, \gamma_2) = \big(\alpha_j(\gamma_1), \alpha_j(\gamma_2)\big)$ whenever γ_1 and γ_2 are both in I_j. (Notice that this automatically takes care of the ambiguities at the endpoints.) Since the mapping class group preserves Farey neighbours, it follows that $\big(\alpha_j(\gamma_1), \alpha_j(\gamma_2)\big) \in \mathcal{F}$ and we can continue to iterate f_1 until possibly ${f_1}^n(\gamma_1, \gamma_2) \in \big\{(\mathbf{e}_0, \mathbf{e}_\infty), (\mathbf{e}_\infty, \mathbf{e}_0)\big\}$. The following shows that the iteration process will always terminate in this way. It is an immediate consequence of Lemma 2.3.2.

Lemma 2.3.4. *Suppose $(\gamma_1, \gamma_2) \in \mathcal{F}$ and $\{\gamma_1, \gamma_2\} \neq \{\mathbf{e}_0, \mathbf{e}_\infty\}$. Then*

$$|f_1(\gamma_1)| + |f_1(\gamma_2)| < |\gamma_1| + |\gamma_2|.$$

□

Suppose now that $\phi \in \mathcal{MCG}(\Sigma_1)$. Since the condition of being Farey neighbours is topological, the pair $\big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty)\big)$ is always a Farey pair. Our normal form results from the following proposition.

Proposition 2.3.5. *Let $\phi \in \mathcal{MCG}(\Sigma_1)$. Then there exists a non-negative integer n so that ${\iota_2}^\varepsilon {f_1}^n\big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty)\big) = (\mathbf{e}_0, \mathbf{e}_\infty)$ where $\varepsilon = 0$ or 1.*

Proof. This follows immediately from the above discussion and Lemma 2.3.4. □

Remark. This proposition shows that the actions of f_1 and $\mathcal{MCG}(\Sigma_1)$ on the space of Farey neighbours are **orbit equivalent**. In other words, for any pairs of Farey neighbours $\{\gamma_1, \gamma_2\}$ and $\{\gamma_1', \gamma_2'\}$ we have

$$\{\gamma_1', \gamma_2'\} = \big\{\phi(\gamma_1), \phi(\gamma_2)\big\}$$

for $\phi \in \mathcal{MCG}(\Sigma_1)$ if and only if there exist non-negative integers m, n so that

$${f_1}^n\{\gamma_1, \gamma_2\} = {f_1}^m\{\gamma_1', \gamma_2'\}.$$

The concept of orbit equivalence is of considerable importance in ergodic theory since any properties depending on only on the orbit structure, for example invariant measures, can now be studied relative to f_1 rather than the group $\mathcal{MCG}(\Sigma_1)$. In particular, we have shown that the action on the space of Farey neighbours is **hyperfinite**, see [5]. This should be compared with the analogous results for Fuchsian groups in [4], [22].

2.4 The normal form and the word acceptor.

Let us denote the exceptional Farey pairs $(\mathbf{e}_0, \mathbf{e}_\infty)$ and $(\mathbf{e}_\infty, \mathbf{e}_0)$ by K_0 and K_2. We extend the definition of f_1 by setting $f_1|_{K_0} = \iota_0 = e$ and $f_1|_{K_2} = \iota_2 = \delta_1\delta_0\delta_1$. Thus we can write $\mathcal{F} = I_0 \cup I_1 \cup I_2 \cup I_3 \cup K_0 \cup K_2$.

The point of Proposition 2.3.5 is that it allows us to define normal forms for elements of $\mathcal{MCG}(\Sigma_1)$ in the following way. For any $\phi \in \mathcal{MCG}(\Sigma_1)$, the pair $\big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty)\big)$ lies in some cell, U_n say, where U_n is one of $I_0, \ldots, I_3$, K_0, K_2. As we apply the map f_1 we move through a sequence of cells

$$U_n \longrightarrow U_{n-1} \longrightarrow \cdots \longrightarrow U_1 \longrightarrow U_0 = K_0.$$

Here each cell U_j for $j \geq 2$ is one of the four cells $I_0, \ldots, I_3$ and U_1 is one of the five cells $I_0, \ldots, I_3$, K_2. At each stage, $f_1|_{U_j} = \alpha_j$, a fixed element in the set $\{{\delta_0}^{\pm 1}, {\delta_1}^{\pm 1}\}$ (or possibly ι_2 if $U_1 = K_2$). Thus we have

$${f_1}^n\big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty)\big) = \big(\alpha_1\alpha_2\cdots\alpha_n\phi(\mathbf{e}_0), \alpha_1\alpha_2\cdots\alpha_n\phi(\mathbf{e}_\infty)\big) = (\mathbf{e}_0, \mathbf{e}_\infty).$$

Since $\mathcal{MCG}(\Sigma_1)$ acts freely on the space $\mathcal{F}$ this shows that $\alpha_1\alpha_2\cdots\alpha_n\phi = e$ giving the normal form $\phi = {\alpha_n}^{-1}\cdots{\alpha_1}^{-1}$. In particular, we have shown that $\{{\delta_0}^{\pm 1}, {\delta_1}^{\pm 1}\}$ generate $\mathcal{MCG}(\Sigma_1)$.

For example, suppose that $\phi = {\delta_0}^2\delta_1$. We claim that the normal form for ϕ is $\delta_0{\delta_1}^{-1}\iota_2$. From Proposition 2.2.1 we see that $\phi(\mathbf{e}_0) = (1,1) \in I_3$ and $\phi(\mathbf{e}_\infty) = (0,1) \in I_3$. Thus we need to find the f_1-expansion for the Farey pair $\big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty)\big) \in I_3$. Applying $f_1|_{I_3} = {\delta_0}^{-1}$

we obtain $f_1\phi(\mathbf{e}_0) = (0,1) \in I_0$ and $f_1\phi(\mathbf{e}_\infty) = (1,0) \in I_0$. Applying $f_1|_{I_0} = \delta_1$ we obtain K_2. Applying $f_1|_{K_2} = \iota_2$ brings us back to K_0. Thus

$$f_1{}^3\big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty)\big) = \big(\iota_2\delta_1\delta_0{}^{-1}\phi(\mathbf{e}_0), \iota_2\delta_1\delta_0{}^{-1}\phi(\mathbf{e}_\infty)\big) = (\mathbf{e}_0, \mathbf{e}_\infty).$$

From Proposition 2.3.5 we see that $\iota_2\delta_1\delta_0{}^{-1}\phi = e$ and hence $\phi = \delta_0\delta_1{}^{-1}\iota_2$ as claimed.

Now we follow the procedure outlined in section 1.2 and construct a finite state automaton that recognises our normal form. We have just extended the definition of f_1 to the $\mathcal{F}$, the set of Farey pairs:

$$\begin{aligned}
f_1|_{I_0} &= \delta_1 \quad : I_0 \longmapsto I_0 \cup I_3 \cup K_0 \cup K_2 \\
f_1|_{I_1} &= \delta_1{}^{-1} : I_1 \longmapsto I_1 \cup I_2 \cup K_0 \cup K_2 \\
f_1|_{I_2} &= \delta_0 \quad : I_2 \longmapsto I_1 \cup I_2 \cup K_0 \cup K_2 \\
f_1|_{I_3} &= \delta_0{}^{-1} : I_3 \longmapsto I_0 \cup I_3 \cup K_0 \cup K_2 \\
f_1|_{K_0} &= e \quad : K_0 \longmapsto K_0 \\
f_1|_{K_2} &= \iota_2 \quad : K_2 \longmapsto K_0.
\end{aligned}$$

In order to define the word acceptor, we define six states labelled I_0, I_1, I_2, I_3, K_0, K_2 and we draw an arrow from state U to state V labelled by α if $U \subset f_1(V)$ and $f_1|_V = \alpha^{-1}$. The start state is K_0. It is clear that there is at most one arrow with a given label from each state and that all arrows ending at a particular state have the same label. Any path in this graph beginning at K_0 and following arrows in the given direction gives the normal form for an element of $\mathcal{MCG}(\Sigma_1)$ by reading the labels on the arrows in the order given by the path. Moreover any ϕ has a unique normal form given in this way. The word acceptor is shown in Fig. 2.4.1. We remark that because we are dealing with composition of functions we read all strings from right to left. Perhaps it is worth pointing out that the normal forms that this word acceptor produces all have the form $W(\delta_0, \delta_1{}^{-1})\iota_2^{\varepsilon}$ or $W(\delta_0{}^{-1}, \delta_1)\iota_2^{\varepsilon}$ where $W(\alpha, \beta)$ is any string in the letters α and β and ε is either 0 or 1.

For example the element $\phi = \delta_0{}^2\delta_1$ considered in the example above corresponds to the following path in the word acceptor

$$K_0 \xrightarrow{\iota_2} K_2 \xrightarrow{\delta_1{}^{-1}} I_0 \xrightarrow{\delta_0} I_3.$$

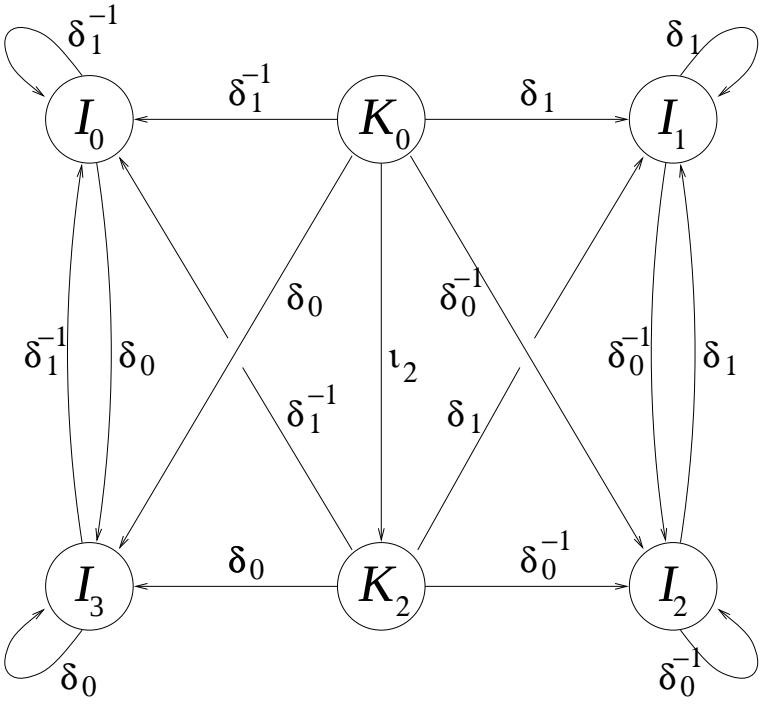


Fig. 2.4.1. The word acceptor.

2.5 The word difference machine.

In order to produce an automatic structure for $\mathcal{MCG}(\Sigma_1)$ we now explain how to construct the word difference machine as explained in section 1.2. We must find a finite set of words $\mathcal{D}$ in $\mathcal{MCG}(\Sigma_1)$ called the **word differences**. This set should have the following properties: First $\mathcal{D}$ should contain the identity e and all the letters of the alphabet $\mathcal{A}$. The second property is slightly more complicated. Suppose that $\psi \in \mathcal{D}$ and U, V are two subsets of the I_j for $j = 0, 1, 2, 3$ or K_j for $j = 0, 2$ with the property that $\psi(U) = V$. Let $f_1|_U = \alpha$ and $f_1|_V = \beta$ denote the restriction of f_1 to U and V. Here α and β are particular elements of the alphabet $\mathcal{A}$. Then $\psi' = \beta\psi\alpha^{-1}$ sends $\alpha(U)$ to $\beta(V)$ [1]. Our second requirement on $\mathcal{D}$ is that for all $\psi \in \mathcal{D}$ we should have $\psi' \in \mathcal{D}$ for all choices of U and V. This means that we get a commutative diagram which we call a **square**:

$$\begin{array}{ccc} U & \xrightarrow{\psi} & V \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\beta} \\ \alpha(U) & \xrightarrow{\psi'} & \beta(V) \end{array} \qquad (*)$$

This square corresponds to the relation $\psi' = \beta\psi\alpha^{-1}$ in G.

We want to be able to concatenate squares vertically. In general $\alpha(U)$ and $\beta(V)$ will contain points in several elements of the partition of $\mathcal{F}$ into I_j and K_j. This means that

[1] This differs from the expression in section 1.2 as we are now reading strings from right to left.

f_1 may not be a fixed element of $\mathcal{MCG}(\Sigma_1)$ on $\alpha(U)$ and $\beta(V)$. Let U' and V' be subsets of $\alpha(U)$ and $\beta(V)$ which each lie in a single set in the partition and satisfy $\psi'(U') = V'$. We are always able to subdivide $\alpha(U)$ and $\beta(V)$ into finitely many pieces for which this property holds. Since U' and V' are each contained in a single set of the partition, the restriction of f_1 to each of these two sets is a fixed element of $\mathcal{MCG}(\Sigma_1)$. In this way we can now construct several squares

$$\begin{array}{ccc} U' & \xrightarrow{\psi'} & V' \\ {\scriptstyle\alpha'}\downarrow & & \downarrow{\scriptstyle\beta'} \\ \alpha'(U') & \xrightarrow{\psi''} & \beta'(V') \end{array} \qquad (*)$$

each of which may each be placed below $(*)$. In other words, we may concatenate squares vertically.

A special case is where the word difference is the identity e

$$\begin{array}{ccc} U & \xrightarrow{e} & U \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\alpha} \\ \alpha(U) & \xrightarrow{e} & \alpha(U) \end{array}$$

For example, if $U = I_0$, $V = I_1$ and $\psi = {\delta_0}^{-1}$ we can construct the following square. This may be verified using Proposition 2.2.1 and the discussion in section 2.4.

$$\begin{array}{ccc} I_0 & \xrightarrow{{\delta_0}^{-1}} & I_1 \\ {\scriptstyle f_1|_{I_0}=\delta_1}\downarrow & & \downarrow{\scriptstyle f_1|_{I_1}={\delta_1}^{-1}} \\ I_0 \cup I_3 \cup K_0 \cup K_2 & \xrightarrow[\iota_2]{} & I_1 \cup I_2 \cup K_0 \cup K_2 \end{array}$$

In order to concatenate vertically, we need to subdivide the sets in the bottom line of this square. It could be followed by squares whose top lines are one of

$$I_0 \xrightarrow{\iota_2} I_2, \qquad I_3 \xrightarrow{\iota_2} I_1, \qquad K_0 \xrightarrow{\iota_2} K_2, \qquad K_2 \xrightarrow{\iota_2} K_0.$$

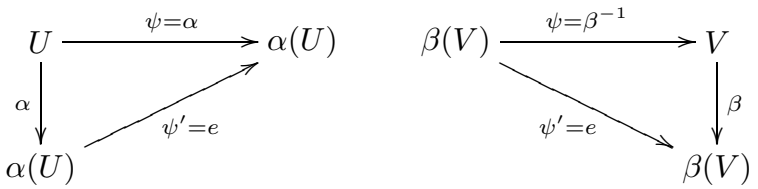
For example, it could be followed by the following square with $U' = I_0$ and $V' = I_2$:

$$\begin{array}{ccc} I_0 & \xrightarrow{\iota_2} & I_2 \\ {\scriptstyle f_1|_{I_0}=\delta_1}\downarrow & & \downarrow{\scriptstyle f_1|_{I_2}=\delta_0} \\ I_0 \cup I_3 \cup K_0 \cup K_2 & \xrightarrow[\iota_2]{} & I_1 \cup I_2 \cup K_0 \cup K_2 \end{array}$$

In order to simplify such diagrams, we make the following definitions

$$Q_0 = I_0 \cup I_3 \cup K_0 \cup K_2, \qquad Q_2 = I_1 \cup I_2 \cup K_0 \cup K_2.$$

If $\psi = \alpha$ or $\psi = \beta^{-1}$ we define degenerate squares, or **triangles** as follows. In the case where $\psi = \alpha$, we do not apply f_1 to V. This means that V may contain points from several sets in the partition, indeed we can take $V = \alpha(U)$. Likewise if $\psi = \beta^{-1}$ we define a triangle by not applying f_1 to U. Again U may contain points from several sets in the partition and we may take $U = \beta(V)$. In both cases ψ' is the identity map e:

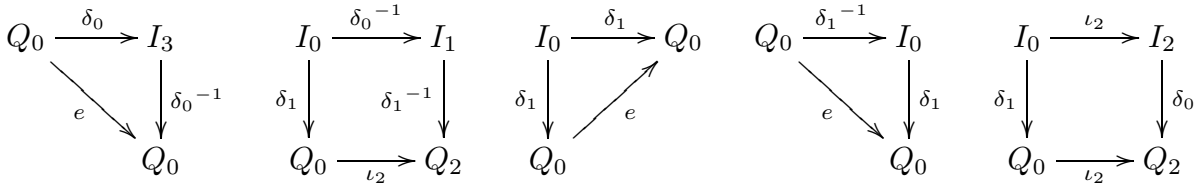


(**)

We now show that we can take the set of word differences to be

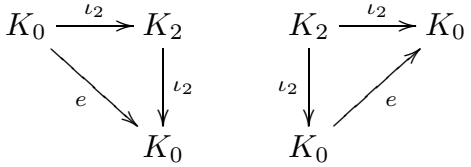
$$\mathcal{D} = \{e, \delta_0, {\delta_0}^{-1}, \delta_1, {\delta_1}^{-1}, \iota_2\}.$$

We first display all squares and triangles for which U is (a subset of) I_0. In the cases of triangles where $\psi = \beta^{-1}$ then we replace U by $\beta(V)$ and include the cases for which $I_0 \subset \beta(V)$ in our list.



The squares and triangles where U is (a subset of) I_1, I_2 or I_3 or where $\beta(V) = Q_2$ may be obtained from these by symmetry.

Finally, suppose that U is either K_0 or K_2. If the word difference ψ is ${\delta_j}^{\pm 1}$ for $j = 0, 1$ the relevant squares and triangles have already been included in the above list. Moreover, for such word differences the new word difference ψ' is e. On the other hand, if the word difference is ι_2 we obtain triangles



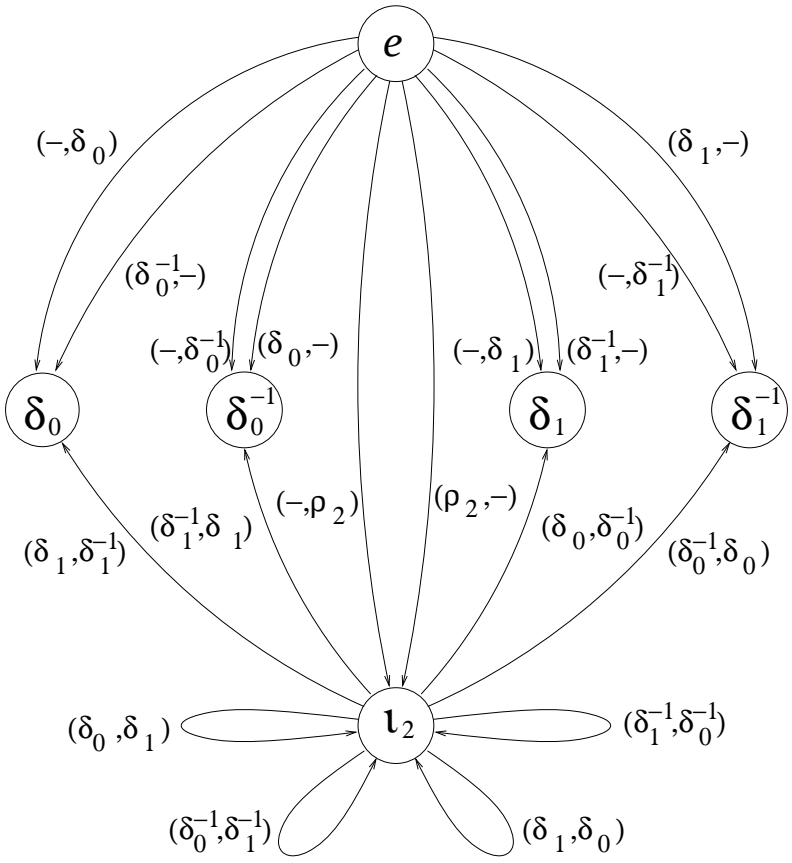


Fig. 2.5.1. The (asynchronous) word difference machine. In addition there are edges from e to itself labelled (δ_0, δ_0), $({\delta_0}^{-1}, {\delta_0}^{-1})$, (δ_1, δ_1), $({\delta_1}^{-1}, {\delta_1}^{-1})$ and (ι_2, ι_2).

We now explain how to construct the word difference machine from these squares and triangles. Following the outline in section 1.2, the states of the machine should be any triple (U, V, ψ) which appears as the top line of one of our squares or triangles. However, since the only function of the choice of U, V is to determine the value of the map f, we may as well take the states of the word difference machine to be the elements of $\mathcal{D}$. The arrows will be ordered pairs $(x, y) \in \mathcal{A} \cup \{-\}$ where x and y are essentially the inverses of α and β. In other words, given a square of the form $(*)$ we draw an arrow from ψ' to ψ and label it $(\alpha^{-1}, \beta^{-1})$. Similarly the triangles $(**)$ correspond to arrows from e to ψ labelled $(\alpha^{-1}, -)$ and $(-, \beta^{-1})$ respectively. We illustrate this in Fig. 2.5.1. In addition there should be arrows labelled (α, α) from e to itself for all $\alpha \in \mathcal{A}$.

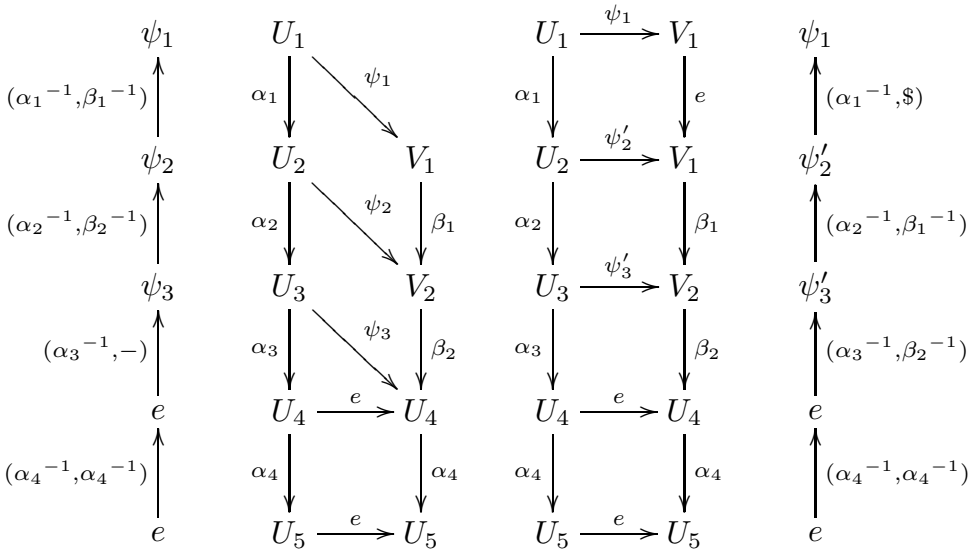
It is automatic from our construction that for any path in the word difference machine with strings of labels $({\alpha_j}^{-1}, {\beta_j}^{-1})$ the strings of labels ${\alpha_j}^{-1}$ and ${\beta_j}^{-1}$ are both paths through the word acceptor.

A result of this construction is that we have verified the presentation for $\mathcal{MCG}(\Sigma_1)$ given in section 2.2 (compare the proof of Theorem 2.3.12 of [7], page 51). In order to see this, observe that any closed path through the Cayley graph can be decomposed into a

union of triangles each of which has one side of length at most one and of which the other two sides are paths in normal form leading back to the identity. This forms a van Kampen diagram for the closed path by covering it with squares and triangles of the form we have constructed above. One can easily verify that each of these squares and triangles corresponds to a relation which may be derived from $\delta_1\delta_0\delta_1 = \delta_0\delta_1\delta_0$ or $(\delta_0\delta_1)^3 = e$.

This has essentially constructed a 2 stringed automaton for the word difference machine. There is still a technical problem to be overcome. Namely, the word difference machine is asynchronous. This is because some of the labels have the form $(\alpha^{-1}, -)$ or $(-, \beta^{-1})$. In fact, this will occur exactly once when we are dealing with pairs of elements of the group which differ by a word of length exactly one. Specifically, the first time the normal forms of the prefixes from the word acceptor differ we see the symbol "$-$" in one of the strings in the word difference machine. This is because all arrows from e to any other state have this form and no other arrow does.

In order to rectify this difficulty we need to **synchronise** the word difference machine. This is done as follows. In the definition of a two stringed automaton, we need to put a padding symbol \$ at the *end* of one of the words to ensure that they have the same length. Thus we need to move "$-$" in the middle of the word to a \$ at the end of the word. This is achieved by adding to our set of word differences the diagonals in each square. That is, for squares of the type $(*)$ we add the diagonal word differences $\beta^{-1}\psi' = \psi\alpha^{-1}$ and $\beta\psi = \psi'\alpha$. (We remark that it is easy to see by inspection that this new word difference can be rearranged to the form $\iota_2\delta_j{}^{\pm 1}$ for $j = 0, 1$.) This has the following effect. Suppose the normal forms from the word acceptor for words differing by ψ are ${\alpha_1}^{-1}{\alpha_2}^{-1}{\alpha_3}^{-1}{\alpha_4}^{-1}$ and ${\beta_1}^{-1}{\beta_2}^{-1}{\alpha_4}^{-1}$. Below we give the path in the word difference machine above (read from bottom to top), the corresponding squares, the amended squares and finally the path in the synchronised difference machine.



Here ψ_2, ψ_3, ψ_2' and ψ_3' are chosen so that the middle two diagrams commute. It is clear how to change the word difference machine in the light of this example. Of course there are now rather more states and arrows in the synchronised difference machine. In particular, the new states are

$$\mathcal{D}' = \{e, \delta_0, {\delta_0}^{-1}, \delta_1, {\delta_1}^{-1}, \iota_2, \iota_2\delta_0, \iota_2{\delta_0}^{-1}, \iota_2\delta_1, \iota_2{\delta_1}^{-1}\}.$$

3. Train tracks for the twice punctured torus.

We now turn our attention to the twice punctured torus Σ_2. We want to mimic the constructions of the previous section. As we shall see, at every stage the basic ideas are the same but the implementation is considerably more complex.

3.1. The cell structure of $\mathcal{ML}(\Sigma_2)$.

The construction of π_1-train tracks for the twice punctured torus was given in [12]. For convenience we go through this briefly.

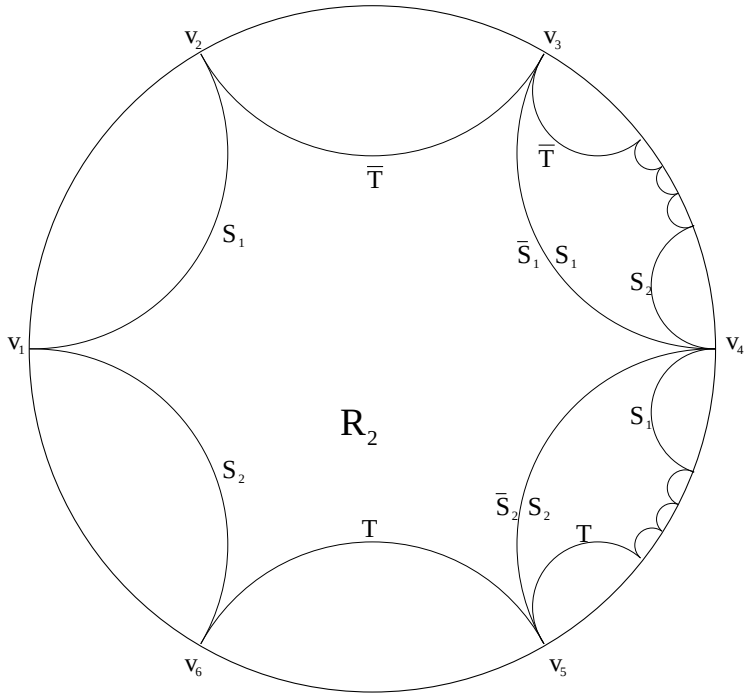


Fig. 3.1.1. A hyperbolic fundamental domain R_2 for Σ_2.

Once again, we fix a hyperbolic structure on Σ_2 by specifying a fundamental polygon for the action of $\pi_1(\Sigma_2)$ on $\mathbb{H}^2$. The fundamental domain R_2 that we choose to work with has six vertices, all of which project to punctures of Σ_2. We label these $v_1, \ldots, v_6$ in clockwise order. The side pairings will be S_1 identifying v_1v_2 to v_4v_3, S_2 identifying v_6v_1 with v_5v_4 and T identifying v_5v_6 with v_3v_2. We assume that S_1, S_2 and T match the endpoints of the respective sides. It is clear that v_1 and v_4 project to one puncture and the other four vertices project to the other. The maps S_1, S_2 and T correspond to homotopy classes of simple closed curves that generate the fundamental group $\pi_1(\Sigma_2)$.

We now introduce the irreducible loops that will form the basis of $\mathcal{ML}(\Sigma_2)$. Fig. 3.1.2 is a schematic picture of the eleven loops as they appear on the fundamental domain R_2. The end of a strand on one side of R_2 is glued by a side-pairing transformation to the corresponding end of the paired side. Thus shortest words representing these loops can be

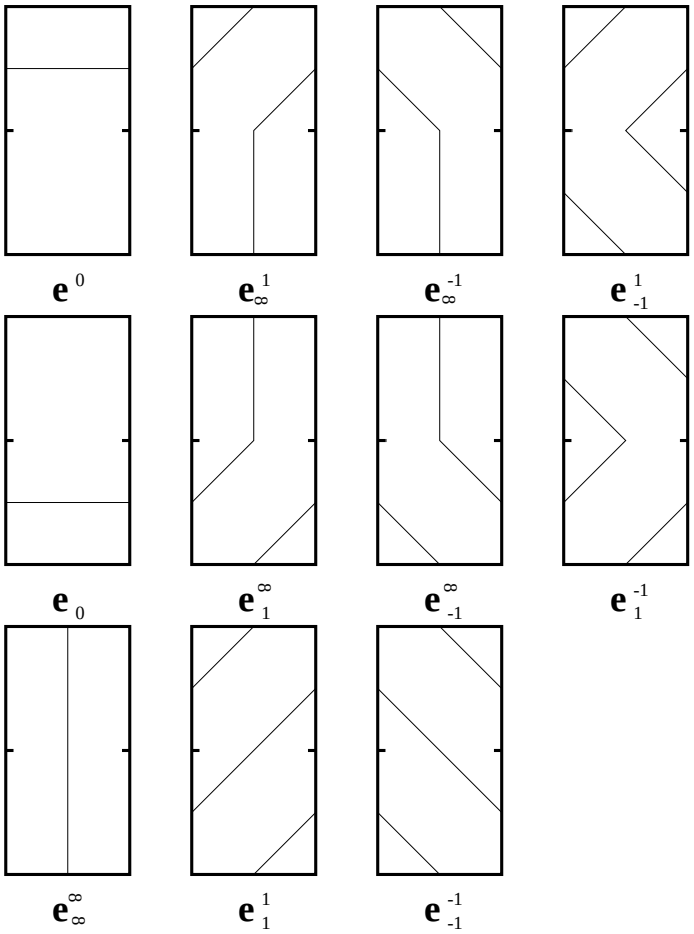


Fig. 3.1.2. The irreducible loops on Σ_2.

either computed directly or read off using the method of cutting sequences, see [2], [3] or [12]. For example, in the loop $\mathbf{e}^1_1$ there are three strands. The end of the strand on v_4v_3 is glued to the end on v_1v_2; the end on v_2v_3 is glued to the end on v_5v_6, and the end on v_5v_4 is glued to the end on v_1v_6. Thus the cutting sequence is S_1TS_2, which as one may easily verify represents this loop in $\pi_1(\Sigma_2)$. Since we are only interested in the unoriented loop up to free homotopy, any cyclic permutation of this sequence or its inverse would work just as well. The full list of cutting sequences for loops is

$$\begin{array}{llll} \mathbf{e}^0 = S_1, & \mathbf{e}^1_\infty = S_1T, & \mathbf{e}^{-1}_\infty = {S_1}^{-1}T, & \mathbf{e}^1_{-1} = S_1T{S_2}^{-1}, \\ \mathbf{e}_0 = S_2, & \mathbf{e}^\infty_1 = TS_2, & \mathbf{e}^\infty_{-1} = T{S_2}^{-1}, & \mathbf{e}^{-1}_1 = {S_1}^{-1}TS_2, \\ \mathbf{e}^\infty_\infty = T, & \mathbf{e}^1_1 = S_1TS_2, & \mathbf{e}^{-1}_{-1} = {S_1}^{-1}T{S_2}^{-1}. & \end{array}$$

The reason for our notation is the following. If we split R_2 into two boxes, the upper one with vertices v_1, v_2, v_3, v_4 and the lower with vertices v_1, v_4, v_5, v_6 (see [12]) then $\mathbf{e}^i_j$ has gradient i in the upper box and j in the lower box. Where there is no superscript (subscript) then the relevant loop has no arcs in the upper (respectively lower) box. This idea is developed further in [12].

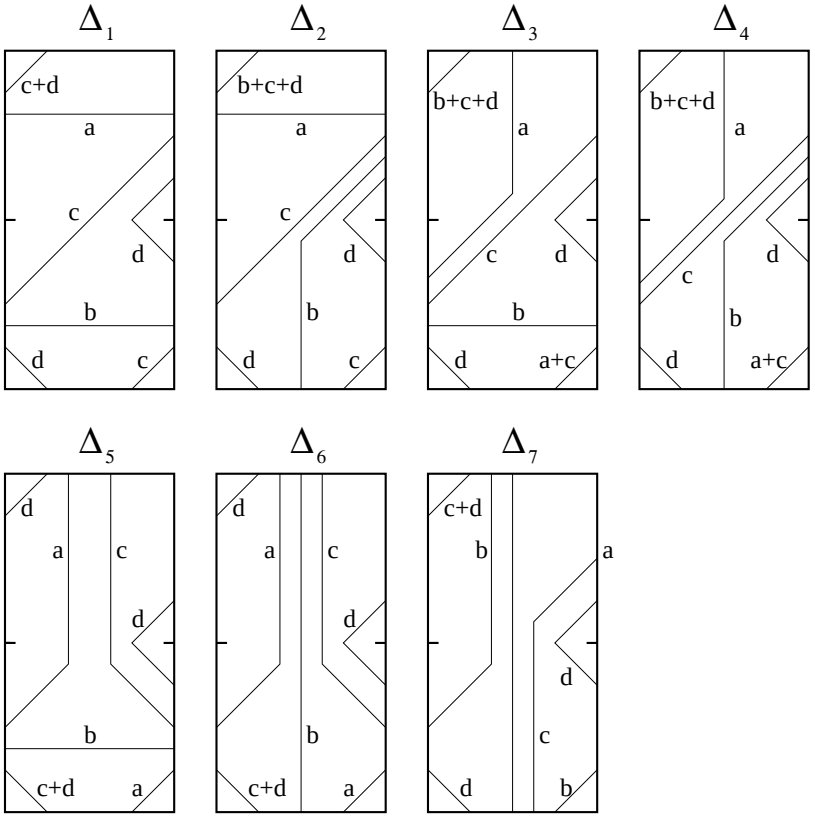


Fig. 3.1.3. Generic points in the maximal cells $\Delta_1, \ldots, \Delta_7$.

We now define 28 cells Δ_j in $\mathcal{ML}(\Sigma_2)$. As we shall show, these cells are maximal, meeting only on lower dimensional faces, and their union is $\mathcal{ML}(\Sigma_2)$.

$$\Delta_1 = \mathrm{sp}^+\{\mathbf{e}^0, \mathbf{e}_0, \mathbf{e}_1^1, \mathbf{e}_{-1}^1\}, \quad \Delta_2 = \mathrm{sp}^+\{\mathbf{e}^0, \mathbf{e}_\infty^1, \mathbf{e}_1^1, \mathbf{e}_{-1}^1\},$$
$$\Delta_3 = \mathrm{sp}^+\{\mathbf{e}_1^\infty, \mathbf{e}_0, \mathbf{e}_1^1, \mathbf{e}_{-1}^1\}, \quad \Delta_4 = \mathrm{sp}^+\{\mathbf{e}_1^\infty, \mathbf{e}_\infty^1, \mathbf{e}_1^1, \mathbf{e}_{-1}^1\},$$
$$\Delta_5 = \mathrm{sp}^+\{\mathbf{e}_1^\infty, \mathbf{e}_0, \mathbf{e}_{-1}^\infty, \mathbf{e}_{-1}^1\}, \quad \Delta_6 = \mathrm{sp}^+\{\mathbf{e}_1^\infty, \mathbf{e}_\infty^\infty, \mathbf{e}_{-1}^\infty, \mathbf{e}_{-1}^1\},$$
$$\Delta_7 = \mathrm{sp}^+\{\mathbf{e}_\infty^\infty, \mathbf{e}_1^\infty, \mathbf{e}_\infty^1, \mathbf{e}_{-1}^1\}.$$

$$\Delta_8 = \mathrm{sp}^+\{\mathbf{e}^0, \mathbf{e}_0, \mathbf{e}_{-1}^{-1}, \mathbf{e}_1^{-1}\}, \quad \Delta_9 = \mathrm{sp}^+\{\mathbf{e}^0, \mathbf{e}_\infty^{-1}, \mathbf{e}_{-1}^{-1}, \mathbf{e}_1^{-1}\},$$
$$\Delta_{10} = \mathrm{sp}^+\{\mathbf{e}_{-1}^\infty, \mathbf{e}_0, \mathbf{e}_{-1}^{-1}, \mathbf{e}_1^{-1}\}, \quad \Delta_{11} = \mathrm{sp}^+\{\mathbf{e}_{-1}^\infty, \mathbf{e}_\infty^{-1}, \mathbf{e}_{-1}^{-1}, \mathbf{e}_1^{-1}\},$$
$$\Delta_{12} = \mathrm{sp}^+\{\mathbf{e}_{-1}^\infty, \mathbf{e}_0, \mathbf{e}_1^\infty, \mathbf{e}_1^{-1}\}, \quad \Delta_{13} = \mathrm{sp}^+\{\mathbf{e}_{-1}^\infty, \mathbf{e}_\infty^\infty, \mathbf{e}_1^\infty, \mathbf{e}_1^{-1}\},$$
$$\Delta_{14} = \mathrm{sp}^+\{\mathbf{e}_\infty^\infty, \mathbf{e}_{-1}^\infty, \mathbf{e}_\infty^{-1}, \mathbf{e}_1^{-1}\}.$$

$$\Delta_{15} = \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}^0, \mathbf{e}_1^1, \mathbf{e}_1^{-1}\}, \quad \Delta_{16} = \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}_1^\infty, \mathbf{e}_1^1, \mathbf{e}_1^{-1}\},$$
$$\Delta_{17} = \mathrm{sp}^+\{\mathbf{e}_\infty^1, \mathbf{e}^0, \mathbf{e}_1^1, \mathbf{e}_1^{-1}\}, \quad \Delta_{18} = \mathrm{sp}^+\{\mathbf{e}_\infty^1, \mathbf{e}_1^\infty, \mathbf{e}_1^1, \mathbf{e}_1^{-1}\},$$
$$\Delta_{19} = \mathrm{sp}^+\{\mathbf{e}_\infty^1, \mathbf{e}^0, \mathbf{e}_\infty^{-1}, \mathbf{e}_1^{-1}\}, \quad \Delta_{20} = \mathrm{sp}^+\{\mathbf{e}_\infty^1, \mathbf{e}_\infty^\infty, \mathbf{e}_\infty^{-1}, \mathbf{e}_1^{-1}\},$$
$$\Delta_{21} = \mathrm{sp}^+\{\mathbf{e}_\infty^\infty, \mathbf{e}_\infty^1, \mathbf{e}_1^\infty, \mathbf{e}_1^{-1}\}.$$

$$\Delta_{22} = \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}^0, \mathbf{e}^{-1}_{-1}, \mathbf{e}^1_{-1}\}, \quad \Delta_{23} = \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}^\infty_{-1}, \mathbf{e}^{-1}_{-1}, \mathbf{e}^1_{-1}\},$$
$$\Delta_{24} = \mathrm{sp}^+\{\mathbf{e}^{-1}_\infty, \mathbf{e}^0, \mathbf{e}^{-1}_{-1}, \mathbf{e}^1_{-1}\}, \quad \Delta_{25} = \mathrm{sp}^+\{\mathbf{e}^{-1}_\infty, \mathbf{e}^\infty_{-1}, \mathbf{e}^{-1}_{-1}, \mathbf{e}^1_{-1}\},$$
$$\Delta_{26} = \mathrm{sp}^+\{\mathbf{e}^{-1}_\infty, \mathbf{e}^0, \mathbf{e}^1_\infty, \mathbf{e}^1_{-1}\}, \quad \Delta_{27} = \mathrm{sp}^+\{\mathbf{e}^{-1}_\infty, \mathbf{e}^\infty_\infty, \mathbf{e}^1_\infty, \mathbf{e}^1_{-1}\},$$
$$\Delta_{28} = \mathrm{sp}^+\{\mathbf{e}^\infty_\infty, \mathbf{e}^{-1}_\infty, \mathbf{e}^\infty_{-1}, \mathbf{e}^1_{-1}\}.$$

The statement that Δ_j is a cell should be interpreted in the following way. One needs to check that the four irreducible loops defining Δ_j are all supported on a common π_1-train track τ_j. This is immediate since one checks that, in each case, all four loops can be drawn in R_2 in such a way that they intersect only on the boundary ∂R_2. The arcs may be homotoped so that their endpoints are at the midpoints of the sides of R_2. Since the midpoints are identified by the side pairings, this exactly gives a π_1-train track in the sense of [2]. The cell Δ_j consists of all proper weightings on the π_1-train track τ_j. Fig. 3.1.4 shows the π_1-train track τ_1 which supports for the cell Δ_1. We normally draw this as in the top left hand corner of Fig. 3.1.3 where it is clearer that the weighting shown is $a\mathbf{e}^0 + b\mathbf{e}_0 + c\mathbf{e}^1_1 + d\mathbf{e}^1_{-1}$. It is easy to check that this is a proper weighting as defined in section 3.1.

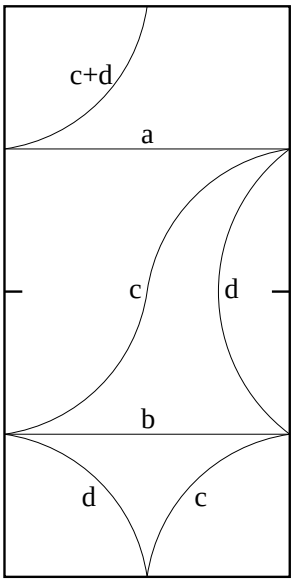


Fig. 3.1.4. The π_1-train track τ_1 corresponding to Δ_1.

Notation 3.1.1. *When we want to speak of a point of one of these cells we write it as an ordered quadruple (a, b, c, d) to represent $a\mathbf{e}_i + b\mathbf{e}_j + c\mathbf{e}_k + d\mathbf{e}_k \in \mathrm{sp}^+\{\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k, \mathbf{e}_l\}$ where the irreducible loops are taken in the order given above.*

Thus, for example $(a, b, c, d) \in \Delta_1$ means $a\mathbf{e}^0 + b\mathbf{e}_0 + c\mathbf{e}^1_1 + d\mathbf{e}^1_{-1}$.

Proposition 3.1.2. *The cells $\Delta_1, \ldots, \Delta_{28}$ are maximal and their union is $\mathcal{ML}(\Sigma_2)$.*

Proof. (Outline) The idea is similar to the proof of Proposition 2.1.1. We will sketch the idea and then illustrate it by performing the computation in one case. All other cases are similar, straightforward, and left to the reader. The idea is the following. It is clear that any multiple simple loop may be homotoped so that it runs along a collection of arcs in R_2 joining midpoints of distinct sides, and which meet only on ∂R_2. Collapsing all arcs joining the same pair of sides yields a properly weighted π_1-train track on R_2. We now reverse this process and investigate what the possibilities for maximal weightings of this kind are.

Take a copy of R_2 and draw strands joining the midpoints of pairs of distinct sides in such a way that no two strands intersect, that no two strands join the same pair of sides and that no more strands can be added without violating the previous two conditions. Now put a weight on each strand. In order to be proper, the weights must satisfy the following conditions as outlined in section 1.3:

(i) all weights should be non-negative and not all zero,

(ii) either the weight on the corner strand separating v_1 from the rest of R should be zero or else the weight on the corner strand separating v_4 should be zero,

(iii) the weight on at least one of the four corner strands separating v_2, v_3, v_5, v_6 from the rest of R_2 should be zero,

(iv) the sum of the weights of all strands ending on a given side should be equal to the sum of weights on the side it is identified with.

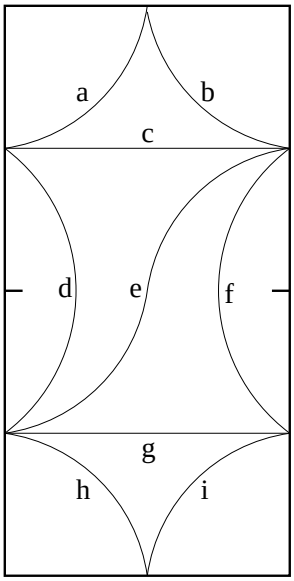


Fig. 3.1.5. Possible weightings on a maximal configuration.

Condition (iv) puts three linear relations between the weights. Solving these relations and inserting conditions (i), (ii) and (iii), we see that we must be in one of the 28 maximal cells

defined above depending on our initial configuration of strands. Detailed computations for the configuration of Fig. 3.1.5 are carried out below. □

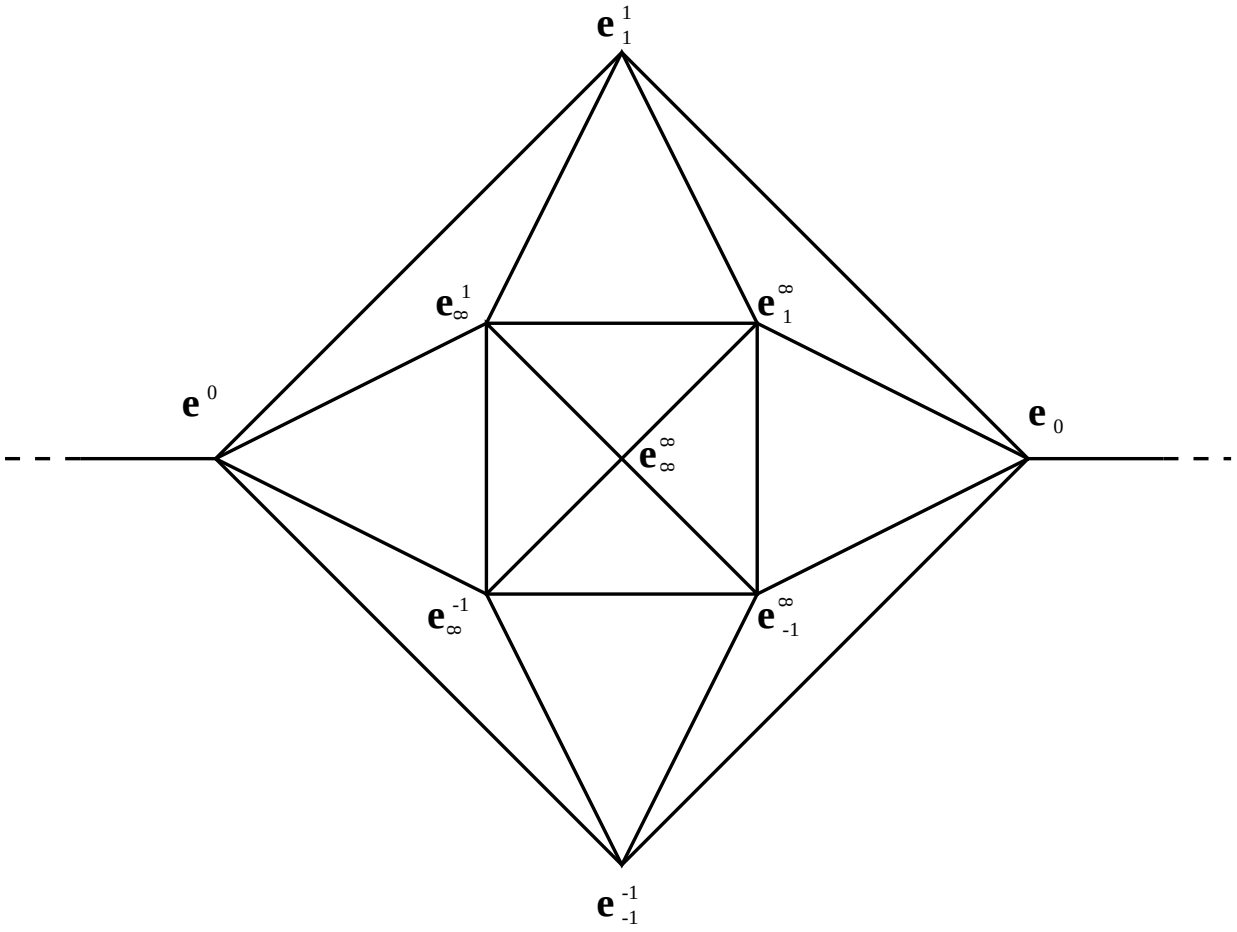


Fig. 3.1.6. Dividing the 2−sphere into 14 three-cells.

An Example. We now perform the computations in the configuration given Fig. 3.1.5 which is a maximal diagram of the type described. By equating the weights on each side we see that

$$a+b=h+i, \quad a+c+d=b+c+e+f, \quad d+e+g+h=f+g+i.$$

At least one of d and f must vanish by condition (ii). (Otherwise there would be a loop homotopic to one of the punctures.) Without loss of generality we suppose $d=0$. This means

$$a+b=h+i, \quad a=b+e+f, \quad e+g+h=f+g+i.$$

Substituting for a in the first equation we obtain

$$2b+e+f=h+i, \quad e+h=f+i.$$

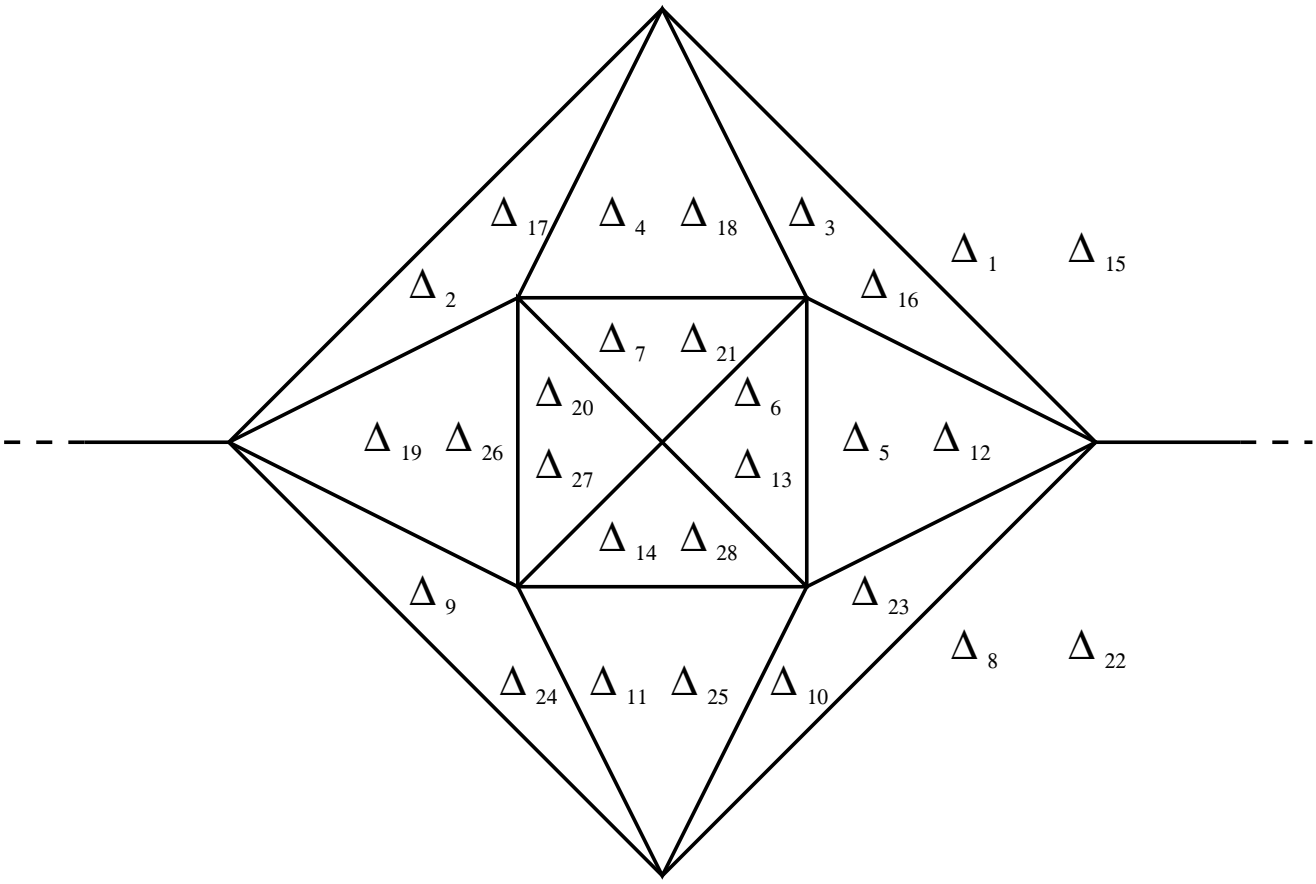


Fig. 3.1.7. Fig. 3.1.6 repeated but indicating the maximal cells.

Adding these and cancelling $f + h$ from each side we get $i = b + e$. Substituting and cancelling once again we find that $h = b + f$. To summarise:

$$a = b + e + f, \quad d = 0, \quad h = b + f, \quad i = b + e.$$

Now by (iii), at least one of a, b, h, i must vanish. By inspection, if any of a, h, i vanish then so must b. (Remember all weights are non-negative.) This means that b must be zero and the train track is:

$$c\mathbf{e}^0 + g\mathbf{e}_0 + e\mathbf{e}_1^1 + f\mathbf{e}_{-1}^1 = (c, g, e, f) \in \Delta_1,$$

corresponding, after changing labels, to the picture shown in Fig. 3.1.4.

We now indicate how the lower dimensional facets in the boundaries of the maximal cells Δ_j for $j = 1, \ldots, 28$ fit together in such a way that the resulting manifold is a 3−sphere. To do this, observe that there are fourteen maximal cells containing the irreducible loop $\mathbf{e}_{-1}^1$ and fourteen containing $\mathbf{e}_1^{-1}$. Moreover, these two irreducible loops never occur together in one of the cells (or else there would be loops around both punctures). Thus each maximal cell is a cone with apex $\mathbf{e}_{-1}^1$ or $\mathbf{e}_1^{-1}$ over the cell spanned by the other

three irreducible loops. One can verify that there are fourteen possibilities for these cells spanned by three loops and that each one arises. Moreover, these fourteen cells may be glued together to form a 2−sphere as indicated in Fig. 3.1.6. Thus the fourteen maximal cells involving $\mathbf{e}^1_{-1}$ form a cone over the 2−sphere, that is a 3−ball. Similarly the other fourteen maximal cells also give a 3−ball. When the boundaries of these two balls are glued together in the obvious manner they form a 3−sphere. We show which maximal cells intersect to give the three-cells on the 2-sphere in Fig. 3.1.7.

3.2. Dehn twists and a presentation for $\mathcal{MCG}(\Sigma_2)$

Let the Dehn twists about $\mathbf{e}^\infty_\infty$, $\mathbf{e}^0$ and $\mathbf{e}_0$ be denoted by δ_0, δ_1 and δ_2 respectively. We now want to investigate the action of these Dehn twists on the cell structure of $\mathcal{ML}(\Sigma_2)$ given in the previous section.

Again we begin by introducing some symmetries that will simplify matters. We are only interested in symmetries which fix the punctures. The symmetry group will be isomorphic to Klein's four group and we describe its non-trivial elements by their action on the vertices of R_2:

ι_1 interchanges the pairs (v_1, v_4), (v_2, v_3), (v_5, v_6);
ι_2 interchanges the pairs (v_1, v_4), (v_2, v_5), (v_3, v_6);
ι_3 fixes v_1, v_4 and interchanges the pairs (v_2, v_6), (v_3, v_5).

When necessary we shall denote the identity by ι_0. It is clear that ι_1 and ι_3 are orientation reversing homeomorphisms of Σ_2 and that ι_2 is orientation preserving. We easily see that ι_j has the following effect on the eleven irreducible loops:

$$\begin{array}{lll} \iota_1\colon \mathbf{e}^i_j \longmapsto \mathbf{e}^{-i}_{-j}, & \mathbf{e}^0 \longmapsto \mathbf{e}^0, & \mathbf{e}_0 \longmapsto \mathbf{e}_0, \\ \iota_2\colon \mathbf{e}^i_j \longmapsto \mathbf{e}^j_i, & \mathbf{e}^0 \longmapsto \mathbf{e}_0, & \mathbf{e}_0 \longmapsto \mathbf{e}^0, \\ \iota_3\colon \mathbf{e}^i_j \longmapsto \mathbf{e}^{-j}_{-i}, & \mathbf{e}^0 \longmapsto \mathbf{e}_0, & \mathbf{e}_0 \longmapsto \mathbf{e}^0 \end{array}$$

where $i, j \in \{\infty, \pm 1\}$ and $-\infty = \infty$.

Thus these actions clearly respect the cell structure of $\mathcal{ML}(\Sigma_2)$. In particular, the maximal cells $\Delta_8, \ldots, \Delta_{28}$ can be expressed as $\Delta_{j+7k} = \iota_k(\Delta_j)$ for $j = 1, \ldots, 7$ and $k = 1, 2, 3$.

The symmetries ι_k conjugate the Dehn twists δ_j to one another. It is easy to check that

$$\begin{array}{lll} \iota_1\delta_0\iota_1 = {\delta_0}^{-1}, & \iota_1\delta_1\iota_1 = {\delta_1}^{-1}, & \iota_1\delta_2\iota_1 = {\delta_2}^{-1}, \\ \iota_2\delta_0\iota_2 = \delta_0, & \iota_2\delta_1\iota_2 = \delta_2, & \iota_2\delta_2\iota_2 = \delta_1, \\ \iota_3\delta_0\iota_3 = {\delta_0}^{-1}, & \iota_3\delta_1\iota_3 = {\delta_2}^{-1}, & \iota_3\delta_2\iota_3 = {\delta_1}^{-1}. \end{array}$$

We can also express ι_2 in terms of the δ_j as

$$\iota_2 = \delta_2\delta_0\delta_2\delta_1\delta_0\delta_2 = (\delta_0\delta_1\delta_2)^2.$$

The proof of this identity will be an easy exercise once the effect of the Dehn twists on $\mathcal{ML}(\Sigma_2)$ has been investigated in the next section. This proof will be left to the reader.

Before we go on to investigate the effect of Dehn twists on π_1-train tracks we will discuss the presentation of $\mathcal{MCG}$ in terms of Dehn twist generators. The first claim is that δ_0, δ_1 and δ_2 generate the (orientation preserving) mapping class group of Σ_2. As in section 2, this will be a consequence of our construction of the Markov map, which will once again produce a unique normal form for every element of $\mathcal{MCG}(\Sigma_2)$ in terms of δ_0, δ_1 and δ_2 (and their inverses).

There are certain relations in G which arise by inspection on Σ_2. For example, since $\mathbf{e}^0$ and $\mathbf{e}_0$ are disjoint, the Dehn twists δ_1 and δ_2 commute. Also, since $\mathbf{e}^0$ and $\mathbf{e}_0$ each intersect $\mathbf{e}^\infty_\infty$ exactly once, for $j = 1, 2$ the Dehn twists δ_0 and δ_j satisfy the braid relation $\delta_0\delta_j\delta_0 = \delta_j\delta_0\delta_j$. Finally, ι_2 is an involution so its square is the identity. Using the form for ι_2 constructed above, we can write $\iota_2 = (\delta_0\delta_1\delta_2)^2$. This gives the relation $(\delta_0\delta_1\delta_2)^4 = e$ (compare [1]).

It turns out that these are all the relations we need to give a presentation of G. This will follow from our construction of the word difference machine (section 5) along the same lines as explained in section 2.5. Hence we obtain the following presentation for $\mathcal{MCG}(\Sigma_2)$. This resembles the presentations for other mapping class groups given by Birman in [1] and could have been derived using methods similar to hers.

Theorem 3.2.1. *The mapping class group of the twice punctured torus admits a presentation:*

$$\mathcal{MCG}(\Sigma_2) = \left\langle \delta_0, \delta_1\delta_2 \middle| \delta_1\delta_2 = \delta_2\delta_1,\ \delta_1\delta_0\delta_1 = \delta_0\delta_1\delta_0,\ \delta_2\delta_0\delta_2 = \delta_0\delta_2\delta_0,\ (\delta_0\delta_1\delta_2)^4 = e \right\rangle.$$

□

In section 5 of [14] Magnus gives the following presentation of $\mathcal{MCG}(\Sigma_2)$. The generators are r, s, ρ, σ, τ subject to the following relations:

$$\begin{aligned}
&s^2 = (r^{-1}s)^3, \quad s\tau s^{-1} = \rho, \quad s\rho s^{-1} = \rho\tau^{-1}\rho^{-1},\\
&r\tau r^{-1} = \tau, \quad r\rho r^{-1} = \rho\tau^{-1}, \quad s^{-4}\rho\tau^{-1}\rho^{-1}\tau = 1,\\
&\sigma^2 = s^{-4}, \quad \sigma r\sigma^{-1} = r, \quad \sigma s\sigma^{-1} = s, \quad \sigma\tau\sigma^{-1} = \tau^{-1}s^4.
\end{aligned}$$

One may pass from our presentation to Magnus' presentation via the substitution

$$\begin{aligned}
&r = \delta_1, \quad s = \delta_1\delta_0\delta_1, \quad \rho = \delta_1\delta_0\delta_1{\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1},\\
&\sigma = {\delta_1}^{-1}{\delta_0}^{-1}{\delta_1}^{-1}{\delta_1}^{-1}{\delta_0}^{-1}{\delta_1}^{-1}, \quad \tau = \delta_1{\delta_2}^{-1}.
\end{aligned}$$

The proof of this is straightforward and is left to the reader.

3.3. The effect of Dehn twists on $\mathcal{PML}$.

We now investigate the effect of the Dehn twists on the maximal cells. Using general results of Birman-Series [2] and Hamidi-Tehrani-Chen [8] we know that these maps are piecewise linear. In fact using reductions similar to those in section 2 we will show this directly. We begin by summarising the results.

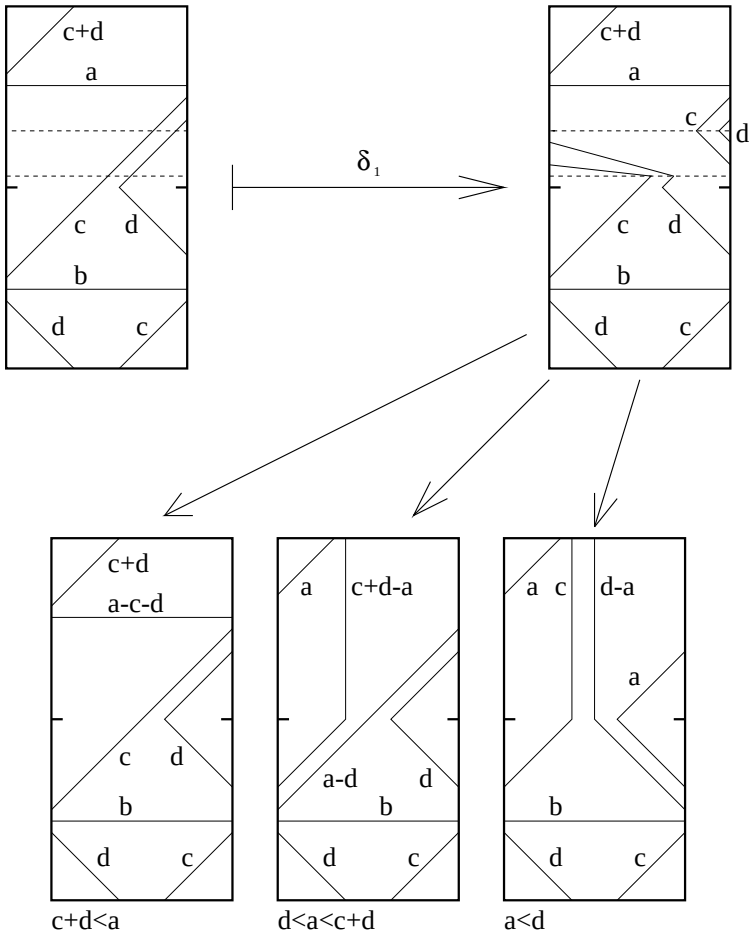


Fig. 3.3.1. The effect of δ_1 on Δ_1.

Proposition 3.3.1. *The Dehn twist δ_1 has the following effect on $\Delta_1, \ldots, \Delta_7$*

(i) δ_1 maps Δ_1 piecewise linearly to $\Delta_1 \cup \Delta_3 \cup \Delta_5$ as follows:

$$\delta_1(a,b,c,d) = \begin{cases} (a-c-d,b,c,d) \in \Delta_1 & \text{if } c+d \leq a \\ (c+d-a,b,a-d,d) \in \Delta_3 & \text{if } d \leq a \leq c+d \\ (c,b,d-a,a) \in \Delta_5 & \text{if } a \leq d. \end{cases}$$

(ii) δ_1 maps Δ_2 piecewise linearly to $\Delta_2 \cup \Delta_4 \cup \Delta_7 \cup \Delta_6$ as follows:

$$\delta_1(a,b,c,d) = \begin{cases} (a-b-c-d,b,c,d) \in \Delta_2 & \text{if } b+c+d \leq a \\ (b+c+d-a,b,a-b-d,d) \in \Delta_4 & \text{if } b+d \leq a \leq b+c+d \\ (b+d-a,c,a-d,d) \in \Delta_7 & \text{if } d \leq a \leq b+d \\ (c,b,d-a,a) \in \Delta_6 & \text{if } a \leq d. \end{cases}$$

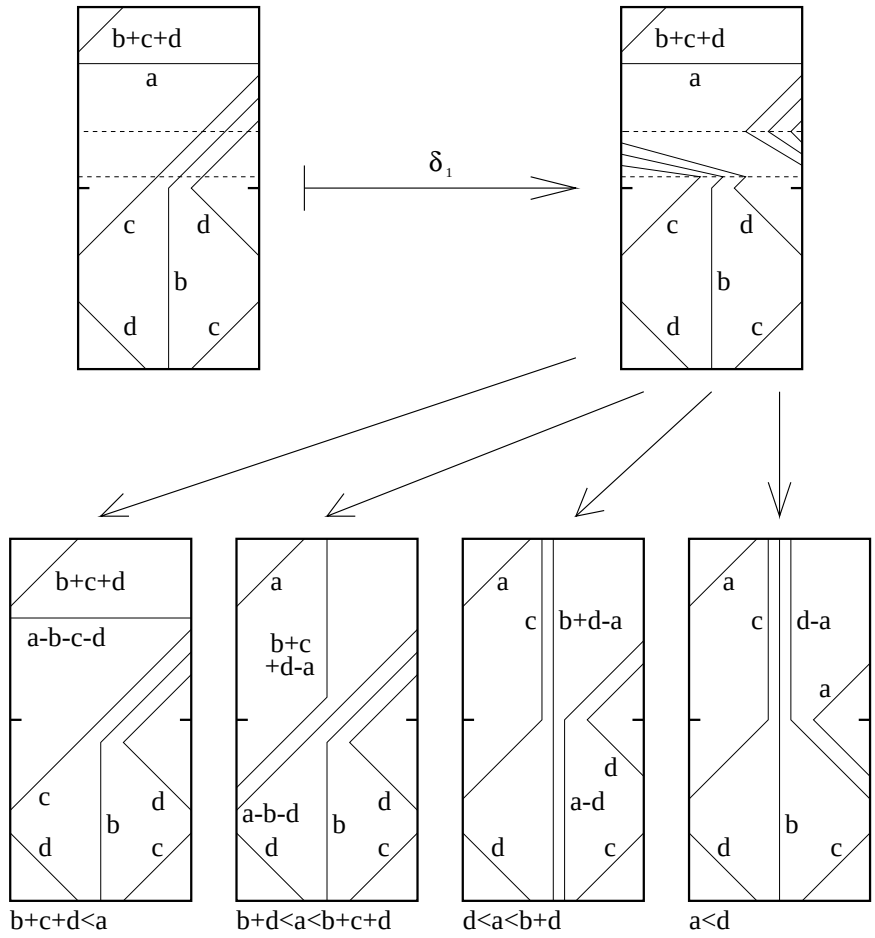


Fig. 3.3.2. *The effect of δ_1 on Δ_2.*

(iii) *δ_1 maps Δ_3 to Δ_{12}, Δ_4 to Δ_{13}, Δ_5 to Δ_{10} and Δ_6 to Δ_{11} as follows:*

$$\delta_1(a,b,c,d) = (d,b,c,a).$$

(iv) *δ_1 maps Δ_7 to Δ_{14} as follows:*

$$\delta_1(a,b,c,d) = (c,d,a,b).$$

We remark that inverting the maps in Propositions 3.3.1(i) and (ii) shows that ${\delta_1}^{-1}$ maps $\bigcup_{j=1}^{7}\Delta_j$ to $\Delta_1 \cup \Delta_2$. We now investigate the action of δ_2.

Proposition 3.3.2. *The Dehn twist δ_2 has the following effect on $\Delta_1, \ldots, \Delta_7$*

(i) *δ_2 maps Δ_1 to $\Delta_1 \cup \Delta_2$ as follows:*

$$\delta_2(a,b,c,d) = \begin{cases} (a,b+d-c,c,d) \in \Delta_1 & \text{if } c \leq b+d \\ (a,c-b-d,b+d,d) \in \Delta_2 & \text{if } b+d \leq c. \end{cases}$$

(ii) *δ_2 maps Δ_2 to $\Delta_1 \cup \Delta_2$ as follows:*

$$\delta_2(a,b,c,d) = \begin{cases} (a,d-c,c,b+d) \in \Delta_1 & \text{if } c \leq d \\ (a,c-d,d,b+d) \in \Delta_2 & \text{if } d \leq c. \end{cases}$$

(iii) δ_2 *maps* Δ_3 *to* $\Delta_4 \cup \Delta_3 \cup \Delta_7$ *as follows:*

$$\delta_2(a,b,c,d) = \begin{cases} (a, b+d-a-c, c, d) \in \Delta_3 & \text{if } a+c \leq b+d \\ (a, a+c-b-d, b+d-a, d) \in \Delta_4 & \text{if } a \leq b+d \leq a+c \\ (a-b-d, b+d, c, d) \in \Delta_7 & \text{if } b+d \leq a. \end{cases}$$

(iv) δ_2 *maps* Δ_4 *to* $\Delta_4 \cup \Delta_3 \cup \Delta_7$ *as follows:*

$$\delta_2(a,b,c,d) = \begin{cases} (a, d-a-c, c, b+d) \in \Delta_3 & \text{if } a+c \leq d \\ (a, a+c-d, d-a, b+d) \in \Delta_4 & \text{if } a \leq d \leq a+c \\ (a-d, d, c, b+d) \in \Delta_7 & \text{if } d \leq a. \end{cases}$$

(v) δ_2 *maps* Δ_5 *to* $\Delta_5 \cup \Delta_6$ *as follows:*

$$\delta_2(a,b,c,d) = \begin{cases} (a, b+c+d-a, c, d) \in \Delta_5 & \text{if } a \leq b+c+d \\ (b+c+d, a-b-c-d, c, d) \in \Delta_6 & \text{if } b+c+d \leq a. \end{cases}$$

(vi) δ_2 *maps* Δ_6 *to* $\Delta_5 \cup \Delta_6$ *as follows:*

$$\delta_2(a,b,c,d) = \begin{cases} (a, c+d-a, b+c, d) \in \Delta_5 & \text{if } a \leq c+d \\ (c+d, a-c-d, b+c, d) \in \Delta_6 & \text{if } c+d \leq a. \end{cases}$$

(vii) δ_2 *maps* Δ_7 *to* $\Delta_5 \cup \Delta_6$ *as follows:*

$$\delta_2(a,b,c,d) = \begin{cases} (b, d-b, a, c+d) \in \Delta_5 & \text{if } b \leq d \\ (d, b-d, a, c+d) \in \Delta_6 & \text{if } d \leq b. \end{cases}$$

We remark that inverting these maps shows that ${\delta_2}^{-1}$ maps $\Delta_1 \cup \Delta_2$ to $\Delta_1 \cup \Delta_2$; $\Delta_3 \cup \Delta_4 \cup \Delta_7$ to $\Delta_3 \cup \Delta_4$ and $\Delta_5 \cup \Delta_6$ to $\Delta_5 \cup \Delta_6 \cup \Delta_7$. We now turn our attention to ${\delta_0}^{-1}$.

Proposition 3.3.3. *The Dehn twist* ${\delta_0}^{-1}$ *has the following effect on* $\Delta_1, \ldots, \Delta_7$

(i) ${\delta_0}^{-1}$ *maps* Δ_1 *to* Δ_{25}, Δ_2 *to* Δ_{24}, Δ_3 *to* Δ_{23} *and* Δ_4 *to* Δ_{22} *as follows:*

$${\delta_0}^{-1}(a,b,c,d) = (a,b,c,d).$$

(ii) ${\delta_0}^{-1}$ *maps* Δ_5 *to* $\Delta_5 \cup \Delta_6$ *as follows:*

$${\delta_0}^{-1}(a,b,c,d) = \begin{cases} (c, a-c, b+c, d) \in \Delta_5 & \text{if } c \leq a \\ (a, c-a, b+c, d) \in \Delta_6 & \text{if } a \leq c. \end{cases}$$

(iii) ${\delta_0}^{-1}$ *maps* Δ_6 *to* $\Delta_5 \cup \Delta_6$ *as follows:*

$${\delta_0}^{-1}(a,b,c,d) = \begin{cases} (b+c, a-b-c, c) \in \Delta_5 & \text{if } b+c \leq a \\ (a, b+c-a, c) \in \Delta_6 & \text{if } a \leq b+c. \end{cases}$$

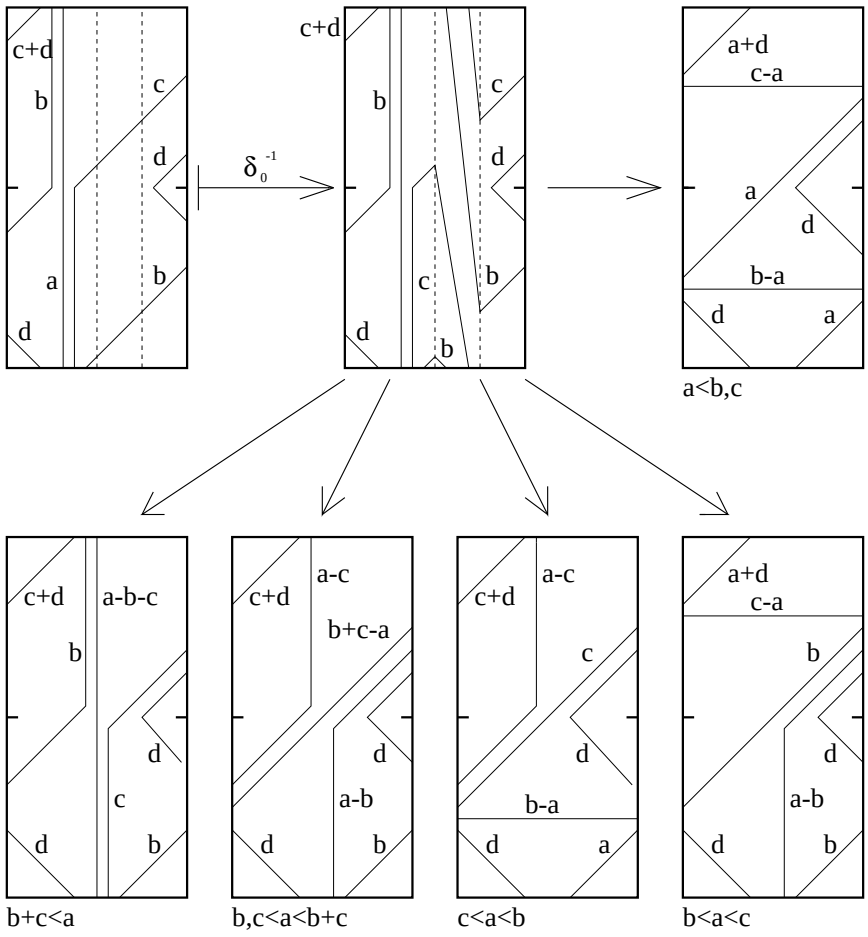


Fig. 3.3.3. *The effect of* $\delta_0{}^{-1}$ *on* Δ_7.

(iv) $\delta_0{}^{-1}$ *maps* Δ_7 *onto* $\Delta_7 \cup \Delta_4 \cup \Delta_3 \cup \Delta_2 \cup \Delta_1$ *as follows:*

$$\delta_0{}^{-1}(a,b,c,d) = \begin{cases} (a-b-c,b,c,d) \in \Delta_7 & \text{if } b+c \leq a \\ (a-c,a-b,b+c-a,d) \in \Delta_4 & \text{if } b,c \leq a \leq b+c \\ (a-c,b-a,c,d) \in \Delta_3 & \text{if } c \leq a \leq b \\ (c-a,a-b,b,d) \in \Delta_2 & \text{if } b \leq a \leq c \\ (c-a,b-a,a,d) \in \Delta_1 & \text{if } a \leq b,c. \end{cases}$$

We remark that inverting these maps we see that δ_0 maps $\Delta_1 \cup \Delta_2 \cup \Delta_4 \cup \Delta_3 \cup \Delta_7$ to Δ_7 and $\Delta_5 \cup \Delta_6$ to $\Delta_5 \cup \Delta_6$.

Remark. We remark that points in the images of the Dehn twists are well defined. In other words if equality holds in one of the conditions above we are on the common boundary of two maximal cells. For example if $(a,b,c,d) \in \Delta_1$ with $a = c+d$ then

$$\delta_1(c+d,b,c,d) = b\mathbf{e}_0 + c\mathbf{e}_1^1 + d\mathbf{e}_{-1}^1 \in \Delta_1 \cap \Delta_3.$$

Proof of Propositions 3.3.1, 3.3.2 and 3.3.3. The method will be the same for each of the propositions and is completely analogous to the proof of Proposition 2.2.1. The

proof of Proposition 3.3.1(i) follows from an analysis of Fig. 3.3.1, Proposition 3.3.1(ii) from Fig. 3.3.2 and Proposition 3.3.3(iv) from Fig. 3.3.3. The proofs of the other parts follow along similar lines but are easier. To avoid repetition we will discuss the proof of Proposition 3.3.1(i) and leave the rest to the reader. These proofs should be compared to the very similar discussion in section 3.2 of [8].

In Fig. 3.3.1, we start off with the π_1-train track for a general point $(a, b, c, d) \in \Delta_1$. We want to perform a Dehn twist δ_1 about $\mathbf{e}^0$. A tubular neighbourhood about $\mathbf{e}^0$ is represented by the pair of dotted lines in the figure. The Dehn twist is the identity everywhere outside these dotted lines and so everything we see here remains the same. In the cylinder between the dotted lines, we do a whole turn to the left. This means that a line crossing this cylinder is wrapped once around the cylinder before exiting in the same place. In the diagram this is represented by the diagonal lines inside the strip (in the train track on the top right). The new train track is a representation of the original train track after the Dehn twist has taken place. Unfortunately, this train track is unreduced. That is, it has strands from one side to itself (the upper right hand side v_4v_3). We need to remove these loops by pulling them tight as explained in section 2.2.

The details of how to pull tight are as follows. The results are shown in the lower part of Fig. 3.3.1. As usual, it is enough to assume that all the weights are integers. Recall that an integral weight m on a strand means that m strands join the same pair of sides of R_2. The ends of these strands are identified, preserving order, with the strands on the paired side. In the top right hand diagram of Fig. 3.3.1, there are $c + d$ strands joining the side v_4v_3 to itself. The side pairing ${S_1}^{-1}$ takes the $2(c + d)$ ends of these loops to the lowest $2(c+d)$ ends of strands emanating from v_1v_2. We begin with the innermost of these loops, that is the strand that, together with an arc of v_4v_3, bounds a disc containing no strands. The endpoints of this loop are the $(c + d)$th and $(c + d + 1)$th ends from the bottom of v_4v_3. These are identified by the side pairing ${S_1}^{-1}$ with the $(c + d)$th and $(c + d + 1)$th ends from the bottom of v_1v_2. Providing a and d are non-zero these are ends of strands joining v_1v_2 with v_5v_4 and v_4v_3 respectively. When we pull this loop tight these three strands become a single strand from v_5v_4 to v_4v_3. Doing this $\min\{a, d\}$ times we get this number of strands joining v_5v_4 and v_4v_3.

In the case where $a \leq d$ we have exhausted all a strands joining the sides v_1v_2 and v_4v_3. We therefore continue using the $c + d$ strands joining v_3v_2 and v_1v_2. After pulling each of the next $d - a$ loops tight we obtain strands from v_3v_2 to v_5v_4. Finally the each of the remaining c loops gives a strand from to v_3v_2 to v_6v_1. There are a strands from v_1v_2 to v_3v_2 remaining that have not been changed. Putting all this information together we obtain the π_1-train track in the bottom right of Fig. 3.3.1. It is then clear that this train track is the point $(c, b, d - a, a) \in \Delta_5$.

The case $a \geq d$ is similar but with further sub-cases $c+d \leq a$ and $d \leq a \leq c+d$. These

give the other train tracks on the lower line of Fig. 3.3.1 and the points of Δ_1 and Δ_3 listed in the statement of Proposition 3.3.1. The rest of the propositions follow similarly. □

The following corollary is an immediate consequence of these results. It may be verified by considering the image of each of the $\mathbf{e}^i_j$ and extending linearly to the whole of $\mathcal{ML}(\Sigma_2)$.

Corollary 3.3.4. *We have* $\iota_2 = (\delta_0\delta_1\delta_2)^2$. □

4. The Markov map

In this section we construct the Markov map which will be the key to constructing our normal form for elements of $\mathcal{MCG}(\Sigma_2)$.

4.1. Canonical Dehn twists for maximal cells.

For each maximal cell Δ_j we now define a **canonical Dehn twist** η_j which is one of ${\delta_0}^{\pm 1}$, ${\delta_1}^{\pm 1}$, ${\delta_2}^{\pm 1}$. For simplicity we work with $\Delta_1, \ldots, \Delta_7$ and then define canonical Dehn twists for the other cells by symmetry. These twists will map maximal cells onto the union of other cells and are the essential step for defining our Markov map.

Definition. The canonical Dehn twist on Δ_1, Δ_2, Δ_5 and Δ_6 is δ_1. The canonical Dehn twist on Δ_3, Δ_4 and Δ_7 is ${\delta_0}^{-1}$.

From Propositions 3.3.1, 3.3.2 and 3.3.3 we see that the images of the canonical Dehn twists are as follows

$$\begin{aligned}
\delta_1 &: \Delta_1 \longrightarrow \Delta_1 \cup \Delta_3 \cup \Delta_5 \\
\delta_1 &: \Delta_2 \longrightarrow \Delta_2 \cup \Delta_4 \cup \Delta_6 \cup \Delta_7 \\
{\delta_0}^{-1} &: \Delta_3 \longrightarrow \Delta_{23} = \iota_3(\Delta_2) \\
{\delta_0}^{-1} &: \Delta_4 \longrightarrow \Delta_{22} = \iota_3(\Delta_1) \\
\delta_1 &: \Delta_5 \longrightarrow \Delta_{10} = \iota_1(\Delta_3) \\
\delta_1 &: \Delta_6 \longrightarrow \Delta_{11} = \iota_1(\Delta_4) \\
{\delta_0}^{-1} &: \Delta_7 \longrightarrow \Delta_1 \cup \Delta_2 \cup \Delta_3 \cup \Delta_4 \cup \Delta_7.
\end{aligned}$$

We now show that the canonical Dehn twists map maximal cells onto the union of other maximal cells listed above. The **interior** of a maximal cell is defined to be Δ_j°, the collection of points $(a, b, c, d) \in \Delta_j$ with a, b, c, d all positive.

Proposition 4.1.1. *Denote the canonical Dehn twist on the cell Δ_k by η_k. For each $j, k \in \{1, \ldots, 28\}$, if $\Delta_j^\circ \cap \eta_k(\Delta_k^\circ)$ is non-empty then $\Delta_j \subset \eta_k(\Delta_k)$.*

Proof. Since the maximal cells only overlap on their boundaries and we are assuming that $\Delta_j^\circ \cap \eta_k(\Delta_k^\circ)$ is non-empty it is sufficient to consider only the maximal cells which appear in the images of canonical Dehn twists listed above. In order to check that $\Delta_j \subset \eta_k(\Delta_k)$ we need only show ${\eta_k}^{-1}(\Delta_j) \subset \Delta_k$. This may be checked using the propositions of section 3.3 by inspection on a case by case basis. □

4.2. The Markov map on $\mathcal{ML}(\Sigma_2)$.

We are going to define the Markov map on the partition of $\mathcal{ML}(\Sigma_2)$ into maximal cells by taking the shortest word in canonical Dehn twists that maps the maximal cell Δ_j onto a union of at least two maximal cells. This will ensure that the resulting mapping has enough "expansion" to strictly reduce length. We will see that composition of at most three canonical Dehn twists has this property. As usual there will be an ambiguity as to how the Markov map is defined on the intersections of maximal cells (that is on their common boundary faces).

It turns out that certain pairs of maximal cells, for example $\Delta_1 \cup \Delta_2$, have the same canonical Dehn twist and always occur together in the image of a canonical twist. Others, for example Δ_7, occur on their own in the image. This means that we can make our partition of $\mathcal{ML}(\Sigma_2)$ coarser. That is, we group the maximal cells Δ_j in pairs or on their own according to how they behave under these canonical Dehn twists. These groupings will be called **regions**. The regions are defined as follows:

$$\begin{array}{cccc} A_0 = \Delta_1 \cup \Delta_2, & B_0 = \Delta_3 \cup \Delta_4, & C_0 = \Delta_5 \cup \Delta_6, & D_0 = \Delta_7, \\ A_1 = \Delta_8 \cup \Delta_9, & B_1 = \Delta_{10} \cup \Delta_{11}, & C_1 = \Delta_{12} \cup \Delta_{13}, & D_1 = \Delta_{14}, \\ A_2 = \Delta_{15} \cup \Delta_{16}, & B_2 = \Delta_{17} \cup \Delta_{18}, & C_2 = \Delta_{19} \cup \Delta_{20}, & D_2 = \Delta_{21}, \\ A_3 = \Delta_{22} \cup \Delta_{23}, & B_3 = \Delta_{24} \cup \Delta_{25}, & C_3 = \Delta_{26} \cup \Delta_{27}, & D_3 = \Delta_{28}. \end{array}$$

There are four different types of region which we call A_j, B_j, C_j, D_j where $j = 0, 1, 2, 3$ is determined by $A_j = \iota_j(A_0)$ and so on.

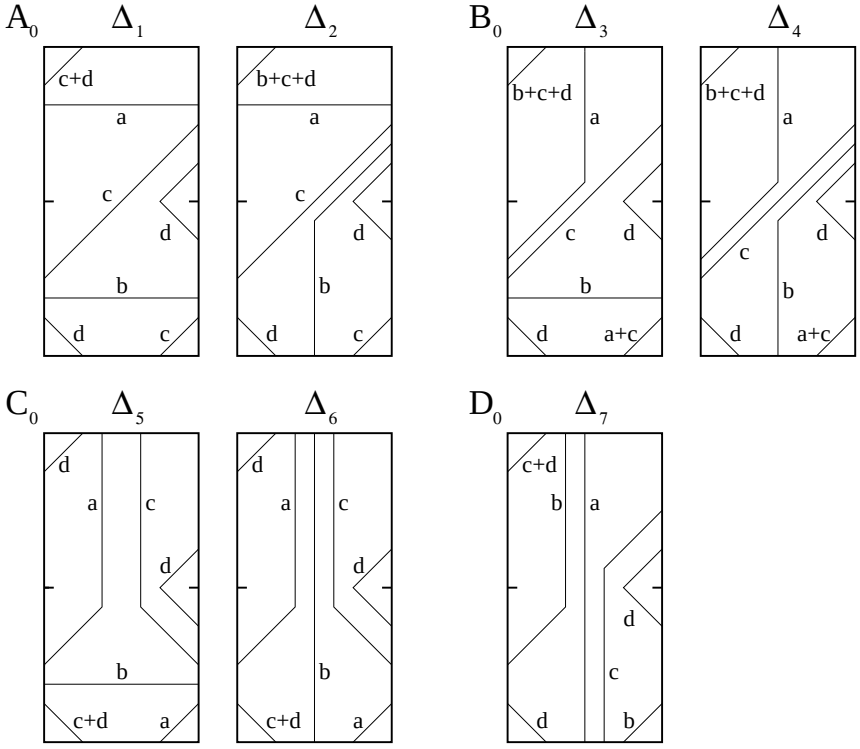


Fig. 4.2.1. Generic points in the regions A_0, B_0, C_0 and D_0.

Lemma 4.2.1. *The canonical Dehn twists map regions onto unions of other regions. For A_0, B_0, C_0 and D_0 these regions are given below.*

$$\begin{aligned}
\delta_1 &: A_0 \longmapsto A_0 \cup B_0 \cup C_0 \cup D_0 \\
{\delta_0}^{-1} &: B_0 \longmapsto A_3 = \iota_3(A_0) \\
\delta_1 &: C_0 \longmapsto B_1 = \iota_1(B_0) \\
{\delta_0}^{-1} &: D_0 \longmapsto A_0 \cup B_0 \cup D_0.
\end{aligned}$$

The results for the other regions may be obtained by symmetry. □

We are now ready to define the Markov map f_2 on $\mathcal{ML}(\Sigma_2)$. This is defined to be the shortest word in the canonical Dehn twists that maps each region onto at least two other regions. It is given as follows:

$$\begin{aligned}
f_2|_{A_0} = \delta_1 &: A_0 \longmapsto A_0 \cup B_0 \cup C_0 \cup D_0 \\
f_2|_{B_0} = {\delta_2}^{-1}{\delta_0}^{-1} &: B_0 \longmapsto A_3 \cup B_3 \cup C_3 \cup D_3 \\
f_2|_{C_0} = \delta_2\delta_0\delta_1 &: C_0 \longmapsto A_2 \cup B_2 \cup C_2 \cup D_2 \\
f_2|_{D_0} = {\delta_0}^{-1} &: D_0 \longmapsto A_0 \cup B_0 \cup D_0 \\
f_2|_{A_1} = {\delta_1}^{-1} &: A_1 \longmapsto A_1 \cup B_1 \cup C_1 \cup D_1 \\
f_2|_{B_1} = \delta_2\delta_0 &: B_1 \longmapsto A_2 \cup B_2 \cup C_2 \cup D_2 \\
f_2|_{C_1} = {\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1} &: C_1 \longmapsto A_3 \cup B_3 \cup C_3 \cup D_3 \\
f_2|_{D_1} = \delta_0 &: D_1 \longmapsto A_1 \cup B_1 \cup D_1 \\
f_2|_{A_2} = \delta_2 &: A_2 \longmapsto A_2 \cup B_2 \cup C_2 \cup D_2 \\
f_2|_{B_2} = {\delta_1}^{-1}{\delta_0}^{-1} &: B_2 \longmapsto A_1 \cup B_1 \cup C_1 \cup D_1 \\
f_2|_{C_2} = \delta_1\delta_0\delta_2 &: C_2 \longmapsto A_0 \cup B_0 \cup C_0 \cup D_0 \\
f_2|_{D_2} = {\delta_0}^{-1} &: D_2 \longmapsto A_2 \cup B_2 \cup D_2 \\
f_2|_{A_3} = {\delta_2}^{-1} &: A_3 \longmapsto A_3 \cup B_3 \cup C_3 \cup D_3 \\
f_2|_{B_3} = \delta_1\delta_0 &: B_3 \longmapsto A_0 \cup B_0 \cup C_0 \cup D_0 \\
f_2|_{C_3} = {\delta_1}^{-1}{\delta_0}^{-1}{\delta_2}^{-1} &: C_3 \longmapsto A_1 \cup B_1 \cup C_1 \cup D_1 \\
f_2|_{D_3} = \delta_0 &: D_3 \longmapsto A_3 \cup B_3 \cup D_3.
\end{aligned}$$

It is clear from Proposition 4.1.1 and Lemma 4.2.1 that f_2 satisfies the Markov property with respect to the partition of $\mathcal{ML}(\Sigma)$ into the sixteen regions $A_0, \ldots, D_3$.

Recall that the **length** $|\gamma|$ of a multiple simple loop γ supported on a weighted π_1-train track τ is the sum of the weights on all its strands. For a proper integral weighting

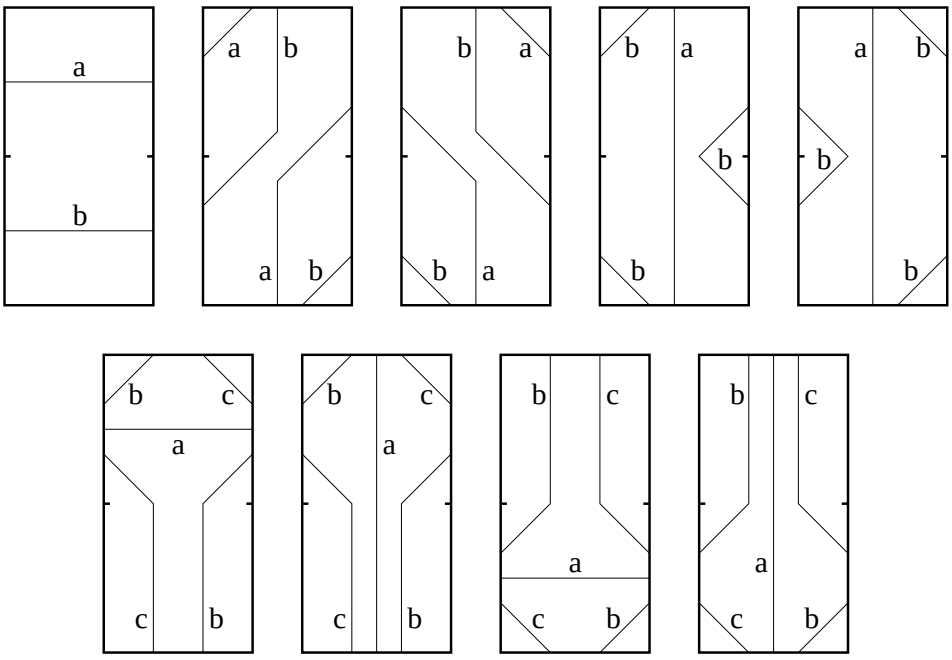


Fig. 4.2.2. The nine exceptional configurations of Proposition 4.2.2.

this is a positive integer. Let $(a, b, c, d) \in \Delta_j$ with the standard bases. We now compare the length of the π_1-train track represented by this point with the length of its image under the Markov map f_2 as defined above. We have

Proposition 4.2.2. *The Markov map f_2 does not increase length. Moreover f_2 strictly decreases length for every train track except those of the form:*

$$a\mathbf{e}^0 + b\mathbf{e}_0, \quad a\mathbf{e}^1_\infty + b\mathbf{e}^\infty_1, \quad a\mathbf{e}^{-1}_\infty + b\mathbf{e}^\infty_{-1}, \quad a\mathbf{e}^\infty_\infty + b\mathbf{e}^1_{-1}, \quad a\mathbf{e}^\infty_\infty + b\mathbf{e}^{-1}_1, \qquad (*)$$

$$a\mathbf{e}^0 + b\mathbf{e}^1_\infty + c\mathbf{e}^{-1}_\infty, \quad a\mathbf{e}^\infty_\infty + b\mathbf{e}^1_\infty + c\mathbf{e}^{-1}_\infty, \quad a\mathbf{e}_0 + b\mathbf{e}^\infty_1 + c\mathbf{e}^\infty_{-1}, \quad a\mathbf{e}^\infty_\infty + b\mathbf{e}^\infty_1 + c\mathbf{e}^\infty_{-1}. \ (\dagger)$$

Proof. This follows from an analysis of lengths for the action of f_2 on each maximal cell Δ_j:

cell	$\lvert(a,b,c,d)\rvert$	f_2	$\lvert f_2(a,b,c,d)\rvert$	$\lvert(a,b,c,d)\rvert - \lvert f_2(a,b,c,d)\rvert$
Δ_1	$a+b+3c+3d$	δ_1	$a+b+2c+2d$	$c+d$
Δ_2	$a+2b+3c+3d$	δ_1	$a+b+2c+2d$	$b+c+d$
Δ_3	$2a+b+3c+3d$	${\delta_2}^{-1}{\delta_0}^{-1}$	$a+b+2c+2d$	$a+c+d$
Δ_4	$2a+2b+3c+3d$	${\delta_2}^{-1}{\delta_0}^{-1}$	$a+b+2c+3d$	$a+b+c$
Δ_5	$2a+b+2c+3d$	$\delta_2\delta_0\delta_1$	$2a+b+2c+d$	$2d$
Δ_6	$2a+b+2c+3d$	$\delta_2\delta_0\delta_1$	$2a+b+2c+d$	$2d$
Δ_7	$a+2b+2c+3d$	${\delta_0}^{-1}$	$a+b+c+3d$	$b+c.$

□

In what follows we will be particularly interested in the four train tracks (†) (as illustrated in the lower line of Fig. 4.2.2) which are $\Delta_{19} \cap \Delta_{26}$, $\Delta_{20} \cap \Delta_{27}$ (which together form $C_2 \cap C_3$) and $\Delta_5 \cap \Delta_{12}$, $\Delta_6 \cap \Delta_{13}$ (which together form $C_0 \cap C_1$).

4.3. The space of Farey blocks

As for the once punctured torus, in order to construct a normal form and word acceptor from the Markov map we constructed in the previous section, we need to find a space on which $\mathcal{MCG}(\Sigma_2)$ acts without fixed points. For the once punctured torus we could take the space of Farey pairs. For the twice punctured torus we will generalise this to the space of **Farey blocks**. A Farey block is an ordered quadruple of (homotopy classes of) simple loops which lie in a certain topological configuration on Σ_2.

Let γ_i and γ_j be (homotopy classes of) simple closed curves on Σ_2. Define the intersection number $i(\gamma_i, \gamma_j)$ to be the minimal number of points in $\gamma_i \cap \gamma_j$ as γ_i, γ_j vary through all elements of their free homotopy class. The idea behind the construction of Farey blocks is the following. Consider a pair of Farey neighbours on the twice punctured torus, that is an ordered pair of curves (γ_1, γ_2) that intersect exactly once, so that $i(\gamma_1, \gamma_2) = 1$. Both of these curves are necessarily non-dividing and there exists a curve β that separates $\gamma_1 \cup \gamma_2$ from the punctures and which is unique up to isotopy. The curve β is the commutator $[\gamma_1, \gamma_2]$. On an unpunctured or once punctured torus β would be homotopically trivial or peripheral respectively. Moreover, in that case (γ_1, γ_2) would have trivial stabiliser in the mapping class group. However, for the twice punctured torus this is not the case. There is a non-trivial homeomorphism which preserves γ_1, γ_2 and β and which interchanges the punctures. The square of this map is the Dehn twist about β. We get around this difficulty by considering an extra curve γ_3 that is disjoint from γ_1 and intersects γ_2 and β once and twice respectively. Here is the precise definition.

Definition. A **Farey block** is an ordered quadruple of (homotopy classes of) curves $(\gamma_1, \gamma_2, \beta; \gamma_3)$ with the properties that:
(i) γ_j is non-dividing for $j = 1, 2, 3$ and β is dividing
(ii) $i(\gamma_j, \beta) = 0$ for $j = 1, 2$ and $i(\gamma_3, \beta) = 2$
(iii) $i(\gamma_2, \gamma_j) = 1$ for $j = 1, 3$ and $i(\gamma_1, \gamma_3) = 0$.
The collection of all Farey blocks will be denoted $\mathcal{F}$.

Since Farey blocks are only defined using topological data, namely the intersection number and the separation properties of (homotopy class of) simple closed curves, the image of a Farey block under an element of $\mathcal{MCG}(\Sigma_2)$ is also a Farey block. This defines an action of $\mathcal{MCG}(\Sigma_2)$ on $\mathcal{F}$. We claim that this action is free. In order to see this, consider the Farey block

$$\left(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \mathbf{e}_1^\infty + \mathbf{e}_{-1}^\infty; \mathbf{e}^0\right).$$

It is clear that this has trivial stabiliser. (It is easy to see that this block is mapped to itself by ι_1 but we are only considering orientation preserving automorphisms of Σ_2.)

The idea behind the normal form for the twice punctured torus is similar to that for the once punctured torus. Namely we apply a general element $\phi \in \mathcal{MCG}(\Sigma_2)$ to the Farey

block $\big(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \mathbf{e}_1^\infty + \mathbf{e}_{-1}^\infty; \mathbf{e}^0\big)$ to obtain a new Farey block

$$\phi\big(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \mathbf{e}_1^\infty + \mathbf{e}_{-1}^\infty; \mathbf{e}^0\big) = \big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty^\infty), \phi(\mathbf{e}_1^\infty + \mathbf{e}_{-1}^\infty); \phi(\mathbf{e}^0)\big).$$

The idea is to apply the Markov map f_2 repeatedly to $\phi\big(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \mathbf{e}_1^\infty + \mathbf{e}_{-1}^\infty; \mathbf{e}^0\big)$ until we get back our original Farey block. The resulting f_2-expansion should be the normal form for ϕ. In practice it is slightly more complicated than this.

Crucial to the construction for the once punctured torus, in section 2, was the following fact. If (γ_1, γ_2) was a Farey pair, then γ_1 and γ_2 both lay in the same cell I_j in $\mathcal{ML}(\Sigma_1)$ so that we could take $f_1(\gamma_1) = \alpha_j(\gamma_1)$, $f_1(\gamma_2) = \alpha_j(\gamma_2)$ for the *same* element $\alpha_j \in \mathcal{MCG}(\Sigma_2)$. In consequence, $\big(f_1(\gamma_1), f_1(\gamma_2)\big)$ was again a Farey pair and applying f_1 to a sequence of such pairs gave a well-defined sequence of elements in $\mathcal{MCG}(\Sigma_1)$, which defined our normal forms.

The analogous statement about Farey blocks is almost, but unfortunately not quite, correct. In fact we can observe that if $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is a Farey block and if β is in the interior of a region R then γ_1 and γ_2 are also in R. This is because γ_1 and γ_2 are disjoint from β and so the π_1-train tracks representing these γ_i and β are both contained in the same maximal cell (compare this with Proposition 2.3.3). In addition, if γ_1 is in the R° then γ_3 is also in R. If, however, γ_1 is on the boundary of R then γ_3 may be in an adjacent region. In the next section we characterise those weighted π_1-train tracks which represent dividing curves. This will enable us to determine those exceptional cases which cause difficulty in extending the map f_2.

4.4. Dividing curves.

The following lemma will allow us to determine which dividing curves do not lie in the interior of any region.

Lemma 4.4.1. *If a proper integral weighting w on a π_1-train track τ on R_2 represents a connected dividing loop β on Σ_2 then*
(i) for each side σ of R_2, the sum of the weights on arcs with endpoints on σ is even, and
(ii) the weights of w have no common factor.

Proof. Let τ be a π_1-train track carrying weight $w(\alpha)$ on strand α. As usual, the weighting can be expanded to the loop β by replacing each strand α with $w(\alpha)$ disjoint strands joining the same pair of sides and gluing their ends together with side pairings. Since β is connected, condition (ii) is clear.

Any dividing curve β separates Σ_2 into a torus with a hole and a sphere with 2 punctures and a hole. Colouring the components of $\Sigma_2 - \beta$ with distinct colours, we see that both punctures lie in a region of the same colour. The colouring lifts to a colouring of $R_2 - \beta$ in such a way that colours alternate along the sides of R_2, changing each time an endpoint of a strand of β meets the side. Since β avoids the punctures, the two segments of side which meet in a puncture have the same colour. The condition (i) is now clear. □

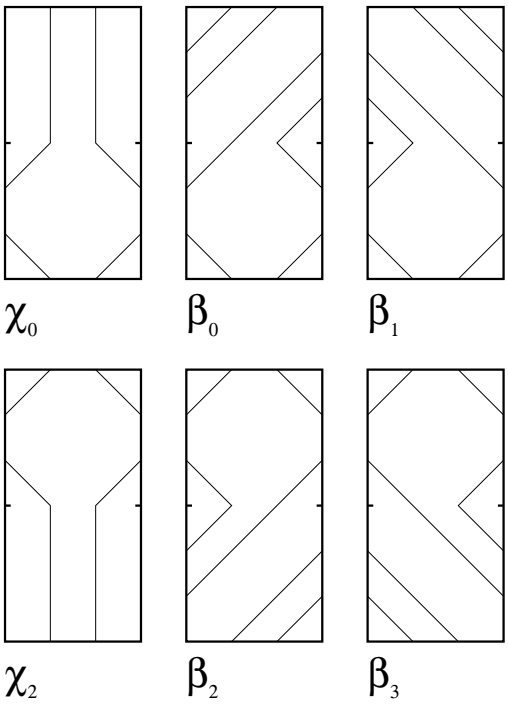


Fig. 4.4.1. The six curves of Proposition 4.4.2.

Proposition 4.4.2. *Every dividing curve is in the interior of a region with the following*

six exceptions, see Fig. 4.4.1:

$$\chi_0 = \mathbf{e}_1^\infty + \mathbf{e}_{-1}^\infty \in C_0 \cap C_1, \quad \beta_0 = \mathbf{e}_1^1 + \mathbf{e}_{-1}^1 \in A_0 \cap B_0, \quad \beta_1 = \mathbf{e}_1^{-1} + \mathbf{e}_{-1}^{-1} \in A_1 \cap B_1,$$
$$\chi_2 = \mathbf{e}_\infty^1 + \mathbf{e}_\infty^{-1} \in C_2 \cap C_3, \quad \beta_2 = \mathbf{e}_1^1 + \mathbf{e}_1^{-1} \in A_2 \cap B_2, \quad \beta_3 = \mathbf{e}_{-1}^1 + \mathbf{e}_{-1}^{-1} \in A_3 \cap B_3.$$

Proof. Any weighting w representing a dividing curve must satisfy conditions (i) and (ii) of Lemma 4.4.1.

Recall that in section 3.1 we mentioned the idea of splitting R_2 into two boxes by drawing a horizontal line from v_1 to v_4. We claim that if w represents a dividing curve then, in at least one of these boxes, there are arcs with non-zero weights across each of the four corners. We say that such a box **contains a cross**. (This condition plays a crucial role in [12]. As can be seen in Fig. 4.4.1, the curves χ_0, β_0, β_1 contain a cross in the lower box and χ_2, β_2, β_3 have one in the upper box.) Let w be a weighting representing a connected dividing curve where neither box contains a cross. If a box does not contain a cross, then one can see by inspection of the irreducible loops in Fig. 3.1.2 (or the generic π_1-train tracks in Fig. 3.1.3), that it contains a certain number a of corner strands across one pair of opposite corners and a number b strands going across a pair of opposite sides. Thus the number of strands ending at each pair of opposite sides is a and $a+b$ respectively. By (i) both a and $a+b$ are even. Thus so is b. Doing this for both boxes we obtain a contradiction to (ii).

Now we refer to Fig. 4.2.1 which shows the generic configurations for weights in any of the four regions A_0, B_0, C_0, D_0. Let us take the region $A_0 = \Delta_1 \cup \Delta_2$; the other cases are similar. The observation that w contains a cross translates into the statement that $c > 0$ and $d > 0$, where, as shown in the top left two diagrams, c and d are the weights on $\mathbf{e}_1^1$ and $\mathbf{e}_{-1}^1$ respectively. Since Δ_1 and Δ_2 are glued across the face $b = 0$ to form the region A_0, any such weighting is in the interior of A_0 unless a (the coefficient of $\mathbf{e}^0$) vanishes. It is however easy to see that any weight $b\mathbf{e}_0 + c\mathbf{e}_1^1 + d\mathbf{e}_{-1}^1$ or $b\mathbf{e}_\infty^1 + c\mathbf{e}_1^1 + d\mathbf{e}_{-1}^1$ represents a multiple loop, one of whose components is $\mathbf{e}_1^1 + \mathbf{e}_{-1}^1$. This contradicts the hypothesis that w represents a connected curve unless $b = 0$ and $c = 1 = d$, in which case $w = \beta_0$.

For $B_0 = \Delta_3 \cup \Delta_4$, the weighting w contains a cross provided $a + c > 0$ and $d > 0$. Again $b = 0$ gives $\Delta_3 \cap \Delta_4$ in the interior of B_0. Thus we have to check two cases, $c = 0$ and $a = 0$. When $a = 0$ we obtain a contradiction as above unless $w = \beta_0$. When $c = 0$, Lemma 4.4.1(i) shows that all the weights are even, in contradiction to Lemma 4.4.1(ii).

The cases of C_0 and D_0 are similar and can be left to the reader. □

The following lemma will be needed in the next section

Lemma 4.4.3. *Let β be a simple closed dividing curve on Σ_2.*

(i) If β is in the interior of A_0 then $f_2|_{A_0}(\beta) = \delta_1(\beta)$ is either in the interior of one of A_0, B_0, C_0, D_0 or else $\delta_1(\beta) = \beta_0$.

(ii) If β is in the interior of D_0 then $f_2|_{D_0}(\beta) = {\delta_0}^{-1}(\beta)$ is either in the interior of one of A_0, B_0, D_0 or else ${\delta_0}^{-1}(\beta) = \beta_0$.

Proof. Both parts are similar. We only consider (i). Consider δ_1 acting on $\mathcal{ML}(\Sigma_2)$. We know that β is in the interior of A_0 and $\delta_1(A_0) = A_0 \cup B_0 \cup C_0 \cup D_0$. Therefore $\delta_1(\beta)$ is in the interior of $A_0 \cup B_0 \cup C_0 \cup D_0$. In other words, $\delta_1(\beta)$ is in the interior of one of A_0, B_0, C_0, D_0 or else it is in the common boundary of at least 2 of these regions. By Proposition 4.4.2 we see that the only possibility is that $\delta_1(\beta) = \beta_0$. □

We will need to characterise the curve γ_3 in the Farey block $(\gamma_1, \gamma_2, \beta; \gamma_3)$ when $(\gamma_1, \gamma_2, \beta)$ is either $(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0)$ or $(\mathbf{e}_\infty^\infty, \mathbf{e}_0, \chi_0)$.

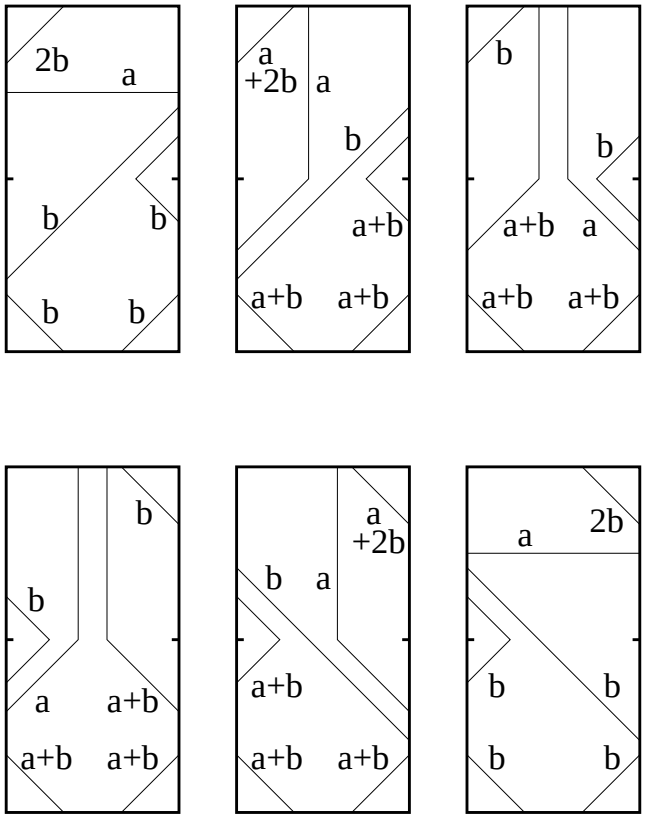


Fig. 4.4.2. The curves disjoint from $\mathbf{e}_0$.

Proposition 4.4.4. *If $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is a Farey block with $(\gamma_1, \gamma_2, \beta) = (\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0)$ then, for some non-negative integer a, γ_3 has one of the following forms:*

(i) $\mathbf{e}^0$,

(ii) $\mathbf{e}_1^\infty + \mathbf{e}_{-1}^1 + a\chi_0 = (a+1, 0, a, 1) \in \Delta_5 \cap \Delta_6$,

(iii) $\mathbf{e}_{-1}^\infty + \mathbf{e}_1^{-1} + a\chi_0 = (a+1, 0, a, 1) \in \Delta_{12} \cap \Delta_{13}$.

If $(\gamma_1, \gamma_2, \beta) = (\mathbf{e}_\infty^\infty, \mathbf{e}_0, \chi_0)$ *then, for some non-negative integer* a, γ_3 *has one of the following forms:*

(iv) $$\mathbf{e}_{-1}^1 + a\chi_0 = (a, 0, a, 1) \in \Delta_5 \cap \Delta_6,$$

(v) $$\mathbf{e}_1^{-1} + a\chi_0 = (a, 0, a, 1) \in \Delta_{12} \cap \Delta_{13}.$$

Remark. Observe that, apart from the case $(\gamma_1, \gamma_2, \beta; \gamma_3) = (\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0)$, in this situation all four curves $\gamma_1, \gamma_2, \gamma_3, \beta$ are in either $\Delta_5 \cup \Delta_6 = C_0$ or $\Delta_{12} \cup \Delta_{13} = C_1$.

Proof. We begin with the case $(\gamma_1, \gamma_2, \beta) = (\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0)$. We know γ_3 must be disjoint from $\gamma_1 = \mathbf{e}_0$. There is a one parameter family of simple closed curves on Σ_2 disjoint from $\mathbf{e}_0$. (To see this, observe that $\Sigma_2 - \{\mathbf{e}_0\}$ is topologically a four times punctured sphere. It is well known that simple closed curves on the four punctured sphere are parametrised by $\mathbb{Q} \cup \{\infty\}$.) It is not hard to show that all curves disjoint from $\mathbf{e}_0$ must have one of the following six types, see Fig. 4.4.2.

$$\begin{array}{ccc} (a, 0, b, b) \in \Delta_1, & (a, 0, b, a+b) \in \Delta_3, & (a+b, 0, a, b) \in \Delta_5, \\ (a+b, 0, a, b) \in \Delta_{12}, & (a, 0, b, a+b) \in \Delta_{10}, & (a, 0, b, b) \in \Delta_8. \end{array}$$

It is easy to see that the intersection numbers of these curves with $\gamma_2 = \mathbf{e}_\infty^\infty$ are: $a + 2b$, $a + 2b$, b, b, $a + 2b$, $a + 2b$ respectively. Since γ_3 should intersect $\gamma_2 = \mathbf{e}_\infty^\infty$ exactly once we obtain the result.

Now we turn our attention to the case $(\gamma_1, \gamma_2, \beta) = (\mathbf{e}_\infty^\infty, \mathbf{e}_0, \chi_0)$. All curves disjoint from $\gamma_1 = \mathbf{e}_\infty^\infty$ must have one of the following four forms, see Fig. 4.4.3

$$(a, 0, a, b) \in \Delta_6, \quad (a, 0, a, b) \in \Delta_{13}, \quad (a, 0, a, b) \in \Delta_{20}, \quad (a, 0, a, b) \in \Delta_{27}.$$

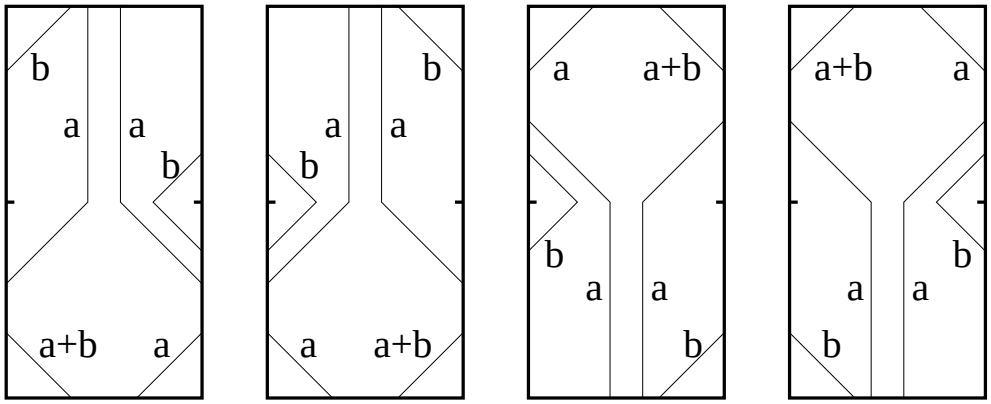


Fig. 4.4.3. The curves disjoint from $\mathbf{e}_\infty^\infty$.

The intersection numbers of these curves with $\gamma_2 = \mathbf{e}_0$ are b, b, $2a + b$, $2a + b$ respectively. The result follows as above. □

The next proposition is a similar characterisation of γ_3 when $(\gamma_1, \gamma_2, \beta)$ is $(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2)$ or $(\mathbf{e}_\infty^\infty, \mathbf{e}^0, \chi_2)$.

Proposition 4.4.5. *If $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is a Farey block with $(\gamma_1, \gamma_2, \beta) = (\mathbf{e}^0, \mathbf{e}^\infty_\infty, \chi_2)$ then, for some non-negative integer a, γ_3 has one of the following forms:*

(i) $\mathbf{e}_0$,

(ii) $\mathbf{e}^1_\infty + \mathbf{e}^{-1}_1 + a\chi_2 = (a+1, 0, a, 1) \in \Delta_{19} \cap \Delta_{20}$,

(iii) $\mathbf{e}^{-1}_\infty + \mathbf{e}^1_{-1} + a\chi_2 = (a+1, 0, a, 1) \in \Delta_{26} \cap \Delta_{27}$.

If $(\gamma_1, \gamma_2, \beta) = (\mathbf{e}^\infty_\infty, \mathbf{e}^0, \chi_2)$ then, for some non-negative integer a, γ_3 has one of the following forms:

(iv) $\mathbf{e}^{-1}_1 + a\chi_2 = (a, 0, a, 1) \in \Delta_{19} \cap \Delta_{20}$,

(v) $\mathbf{e}^1_{-1} + a\chi_2 = (a, 0, a, 1) \in \Delta_{26} \cap \Delta_{27}$.

Proof. This follows by applying ι_2 to the results in Proposition 4.4.4. □

Observe that in this case we have $(\gamma_1, \gamma_2, \beta; \gamma_3) = (\mathbf{e}^0, \mathbf{e}^\infty_\infty, \chi_2; \mathbf{e}_0)$ or else all four curves are in $\Delta_{19} \cup \Delta_{20} = C_2$ or $\Delta_{26} \cup \Delta_{27} = C_3$.

4.5. Subdividing $\mathcal{F}$: the states of the word acceptor.

In the next section we shall extend the Markov map f_2 to a map on the space of Farey blocks $\mathcal{F}$. Inverting this map will give the word acceptor. This construction will resemble that given in section 2.3 but will involve some extra steps. The first part of this process is to divide $\mathcal{F}$ into subsets which will form the states of the word acceptor.

We begin by stratifying $\mathcal{F}$ into three subsets. Define

$$\mathcal{F}_0 = \left\{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta = \chi_0,\ \{\gamma_1, \gamma_2\} = \{\mathbf{e}_0, \mathbf{e}^\infty_\infty\} \text{ or } \beta = \chi_2,\ \{\gamma_1, \gamma_2\} = \{\mathbf{e}^0, \mathbf{e}^\infty_\infty\}\right\},$$
$$\mathcal{F}_1 = \left\{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta = \chi_0,\ \{\gamma_1, \gamma_2\} \neq \{\mathbf{e}_0, \mathbf{e}^\infty_\infty\} \text{ or } \beta = \chi_2,\ \{\gamma_1, \gamma_2\} \neq \{\mathbf{e}^0, \mathbf{e}^\infty_\infty\}\right\},$$
$$\mathcal{F}_2 = \left\{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta \notin \{\chi_0, \chi_2\}\right\}.$$

Our goal is to define a map f on $\mathcal{F}$ so that for each Farey block $(\gamma_1, \gamma_2, \beta; \gamma_3)$ there is a non-negative integer n so that

$$f^n(\gamma_1, \gamma_2, \beta; \gamma_3) = (\mathbf{e}_0, \mathbf{e}^\infty_\infty, \chi_0; \mathbf{e}^0).$$

At each stage, we want f to equal a specific one of the sixteen possible values of f_2, so that for any Farey block we have

$$f(\gamma_1, \gamma_2, \beta; \gamma_3) = \big(f_2(\gamma_1), f_2(\gamma_2), f_2(\beta); f_2(\gamma_3)\big) = \big(\phi(\gamma_1), \phi(\gamma_2), \phi(\beta); \phi(\gamma_3)\big)$$

for some $\phi \in \mathcal{MCG}(\Sigma_2)$. The difficulty is that, if the members of the block $(\gamma_1, \gamma_2, \beta; \gamma_3)$ lie in different regions, it is not clear which value for ϕ to pick. This happens, for example, in the case of the Farey block $(\mathbf{e}^1_{-1}, \mathbf{e}_0, \beta_0; \mathbf{e}^\infty_\infty)$.

We resolve this difficulty by using the results of the previous section and the stratification above. Roughly speaking, f will be defined as follows. On $\mathcal{F}_2$, the map f will take the value of f_2 on the region which contains the dividing curve β. Applying f will decrease the length $|\beta|$ of β. We continue applying f until $\beta = \chi_0$ or χ_2. Thus, there will be a non-negative integer n_2 so that $f^{n_2}(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F}_1 \cup \mathcal{F}_0$. On $\mathcal{F}_1$ the map f will be the Markov map f_1 on $\mathcal{ML}(\Sigma_1)$ constructed in section 2.3. This map will fix β and reduce $|\gamma_1| + |\gamma_2|$. Thus, there will be a non-negative integer n_1 so that $f^{n_1+n_2}(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F}_0$. Finally, on $\mathcal{F}_0$ the map f will fix $|\gamma_1| + |\gamma_2| + |\beta|$ and decrease $|\gamma_3|$. There will be a non-negative integer n_0 so that $f^{n_0+n_1+n_2}(\gamma_1, \gamma_2, \beta; \gamma_3) = (\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0)$.

In order to carry out the details we shall introduce a number of extra regions which will become the states of the word acceptor. These regions will all be subsets of the space $\mathcal{F}$ of Farey blocks.

We begin by considering $\mathcal{F}_2$. By definition, if the Farey block $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is in $\mathcal{F}_2$ then the dividing curve β is neither χ_0 nor χ_2. Using Proposition 4.4.2, we see that either β is in the interior of one of the sixteen regions $A_0, \ldots, D_3$ or else β is one of the four dividing curves β_0, β_1, β_2, β_3. (Recall that for these four curves $\beta_j \in A_j \cap B_j$). We divide $\mathcal{F}_2$ into twenty regions as follows. We call these $\mathbf{A}_j$, $\mathbf{B}_j$, $\mathbf{C}_j$, $\mathbf{D}_j$, $\mathbf{E}_j$ for $j = 0, 1, 2, 3$. The sixteen subsets $\mathbf{A}_0, \ldots, \mathbf{D}_3$ of $\mathcal{F}_2$ are defined to consist of all Farey blocks for which β is in the interior of the region $A_0, \ldots, D_3$ in $\mathcal{ML}(\Sigma_2)$, respectively. The four subsets $\mathbf{E}_j$ of $\mathcal{F}_2$ are defined to consist of all Farey blocks for which $\beta = \beta_j$ where $j = 0, 1, 2, 3$. That is when $j = 0$:

$$\begin{aligned}
\mathbf{A}_0 &= \{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta \in (\Delta_1 \cup \Delta_2)^\circ\}, \\
\mathbf{B}_0 &= \{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta \in (\Delta_3 \cup \Delta_4)^\circ\}, \\
\mathbf{C}_0 &= \{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta \in (\Delta_5 \cup \Delta_6)^\circ\}, \\
\mathbf{D}_0 &= \{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta \in \Delta_7^\circ\}, \\
\mathbf{E}_0 &= \{(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F} \colon \beta = \beta_0\}.
\end{aligned}$$

For $k = 1, 2, 3$ apply the symmetry ι_k to the above five regions in order to obtain $\mathbf{A}_k$, $\mathbf{B}_k$, $\mathbf{C}_k$, $\mathbf{D}_k$ and $\mathbf{E}_k$.

By abuse of notation we will frequently drop the bold face notation $\mathbf{A}_j$ and simply write A_j when the meaning is clear from the context. Thus the reader should keep clearly in mind that A_j may either denote a region of $\mathcal{ML}(\Sigma_2)$ as defined in section 4.2 or the subset $\mathbf{A}_j \subset \mathcal{F}_2$ of Farey blocks.

The subsets $\mathbf{A}_j, \mathbf{B}_j, \ldots, \mathbf{E}_j$ will be states in the word acceptor we are aiming to construct. Before defining the map f on each of these twenty regions (whose inverse will give the arrows in the word acceptor), we will proceed to define the states for the strata $\mathcal{F}_1$ and $\mathcal{F}_0$.

We consider first the Farey blocks in $\mathcal{F}_1$. Here we have $\beta = \chi_0$ or χ_2. For the sake of definiteness we will describe the situation for χ_0 in detail. In order to perform the same constructions for χ_2 it is necessary to apply the symmetry ι_2. The curve χ_0 divides Σ_2 into two components, one of which is a twice punctured disc and the other is a torus with a hole. All homotopically non-trivial, non-peripheral simple closed curves on Σ_2 that are disjoint from χ_0 are contained in the one holed torus component of $\Sigma_2 - \chi_0$. In particular, this is true for γ_1 and γ_2. The stabiliser of χ_0 in $\mathcal{MCG}(\Sigma_2)$ is the group generated by δ_0 and δ_2. The action of the group they generate on the one-holed-torus component of $\Sigma_2 - \chi_0$ exactly corresponds to the action of $\mathcal{MCG}(\Sigma_1)$ on Σ_1 considered in section 2. Therefore, we divide $\mathcal{F}_1$ into states which correspond to the intervals I_0, I_1, I_2 and I_3 for the space of Farey pairs on the once punctured torus. This is done as follows:

$$
\begin{aligned}
F_0 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_0;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}_1^\infty\}\right\},\\
G_0 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_0;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}_\infty^\infty, \mathbf{e}_1^\infty\}\right\},\\
F_1 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_0;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}_0, \mathbf{e}_{-1}^\infty\}\right\},\\
G_1 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_0;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}_\infty^\infty, \mathbf{e}_{-1}^\infty\}\right\},\\
F_2 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_2;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}^0, \mathbf{e}_\infty^1\}\right\},\\
G_2 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_2;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}_\infty^\infty, \mathbf{e}_\infty^1\}\right\},\\
F_3 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_2;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}^0, \mathbf{e}_\infty^{-1}\}\right\},\\
G_3 &= \left\{(\gamma_1,\gamma_2,\beta;\gamma_3) \in \mathcal{F}\colon \beta = \chi_2;\ \gamma_1,\gamma_2 \in \mathrm{sp}^+\{\mathbf{e}_\infty^\infty, \mathbf{e}_\infty^{-1}\}\right\}.
\end{aligned}
$$

Finally, we consider $\mathcal{F}_0$. Here either $\beta = \chi_0$ and $\{\gamma_1,\gamma_2\} = \{\mathbf{e}_0, \mathbf{e}_\infty^\infty\}$ or $\beta = \chi_2$ and $\{\gamma_1,\gamma_2\} = \{\mathbf{e}^0, \mathbf{e}_\infty^\infty\}$. In Propositions 4.4.4 and 4.4.5 we analysed the different possibilities for γ_3. We divide $\mathcal{F}_3$ into states according to these different possibilities:

$$
\begin{aligned}
H_0 &= \{(\mathbf{e}_\infty^\infty, \mathbf{e}_0, \chi_0; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a,0,a,1) \in \Delta_5 \cap \Delta_6 \text{ where } a \in \mathbb{N}\cup\{0\}\},\\
I_0 &= \{(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a+1,0,a,1) \in \Delta_5 \cap \Delta_6 \text{ where } a \in \mathbb{N}\cup\{0\}\},\\
H_1 &= \{(\mathbf{e}_\infty^\infty, \mathbf{e}_0, \chi_0; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a,0,a,1) \in \Delta_{12} \cap \Delta_{13} \text{ where } a \in \mathbb{N}\cup\{0\}\},\\
I_1 &= \{(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a+1,0,a,1) \in \Delta_{12} \cap \Delta_{13} \text{ where } a \in \mathbb{N}\cup\{0\}\},\\
H_2 &= \left\{(\mathbf{e}_\infty^\infty, \mathbf{e}^0, \chi_2; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a,0,a,1) \in \Delta_{19} \cap \Delta_{20} \text{ where } a \in \mathbb{N}\cup\{0\}\right\},\\
I_2 &= \left\{(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a+1,0,a,1) \in \Delta_{19} \cap \Delta_{20} \text{ where } a \in \mathbb{N}\cup\{0\}\right\},\\
H_3 &= \left\{(\mathbf{e}_\infty^\infty, \mathbf{e}^0, \chi_2; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a,0,a,1) \in \Delta_{26} \cap \Delta_{27} \text{ where } a \in \mathbb{N}\cup\{0\}\right\},\\
I_3 &= \left\{(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2; \gamma_3) \in \mathcal{F}\colon \gamma_3 = (a+1,0,a,1) \in \Delta_{26} \cap \Delta_{27} \text{ where } a \in \mathbb{N}\cup\{0\}\right\},\\
J_0 &= \left\{(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0) \in \mathcal{F}\right\},\\
J_2 &= \left\{(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2; \mathbf{e}_0) \in \mathcal{F}\right\}.
\end{aligned}
$$

In each case, notice that $(a,0,a,0)$ corresponds to the loop β, that is χ_0 or χ_2.

4.6. The definition of the map f.

Having defined the states which partition the space $\mathcal{F}$ of Farey blocks, we now turn our attention to the definition of the map f. First we consider states in $\mathcal{F}_2$.

For Farey blocks in $\mathbf{A}_j, \ldots, \mathbf{D}_j$ we define f to be the same as the Markov map f_2 on the corresponding regions $A_j, \ldots, D_j$ in $\mathcal{ML}(\Sigma_2)$. We need to be slightly more careful in computing the images of these blocks.

Lemma 4.6.1. *On $\mathcal{F}_2$ the Dehn twist δ_1 maps A_0 onto $A_0 \cup B_0 \cup C_0 \cup D_0 \cup E_0$ and ${\delta_0}^{-1}$ maps D_0 onto $A_0 \cup B_0 \cup D_0 \cup E_0$.*

Proof. This is an immediate consequence of Lemma 4.4.3. □

To define f in $\mathbf{E}_j$ we have a choice since, as in Proposition 4.4.2, $\beta_j \in A_j \cap B_j$. We choose $f|_{\mathbf{E}_j} = f_2|_{A_j}$ so that, for example, $f|_{\mathbf{E}_0} = f_2|_{A_0} = \delta_1$. It is easy to check that $\delta_1(\beta_0) = \chi_0$ so that, with this definition, f maps $\mathbf{E}_0$ into $\mathcal{F}_1 \cup \mathcal{F}_0$. (We remark that $f_2|_{B_0}(\beta_0) = {\delta_2}^{-1}{\delta_0}^{-1}(\beta_0) = \chi_2$ so this would also be true if we had made the other choice.) Since ${\delta_1}^{-1}(\chi_0) = \beta_0$ it is easy to see that f maps $\mathbf{E}_0$ onto that subset of $\mathcal{F}_1 \cup \mathcal{F}_0$ consisting of all Farey blocks with $\beta = \chi_0$. Applying symmetries, corresponding results are true for $f(\mathbf{E}_j)$ for $j = 1, 2, 3$. We can now summarise the effect of the map f on all states in $\mathcal{F}_2$. In each case, the arrow indicates that f maps the gives state *onto* the union of the states listed on the right. Note that for simplicity, we have now replaced the bold $\mathbf{A}_j, \ldots, \mathbf{E}_j$ with $A_j, \ldots, E_j$ for $j = 0, 1, 2, 3$.

$$
\begin{aligned}
f|_{A_0} = \delta_1 &: A_0 \longmapsto A_0 \cup B_0 \cup C_0 \cup D_0 \cup E_0,\\
f|_{B_0} = {\delta_2}^{-1}{\delta_0}^{-1} &: B_0 \longmapsto A_3 \cup B_3 \cup C_3 \cup D_3 \cup E_3,\\
f|_{C_0} = \delta_2\delta_0\delta_1 &: C_0 \longmapsto A_2 \cup B_2 \cup C_2 \cup D_2 \cup E_2,\\
f|_{D_0} = {\delta_0}^{-1} &: D_0 \longmapsto A_0 \cup B_0 \cup D_0 \cup E_0,\\
f|_{E_0} = \delta_1 &: E_0 \longmapsto F_0 \cup G_0 \cup H_0 \cup I_0 \cup J_0 \cup F_1 \cup G_1 \cup H_1 \cup I_1,\\
f|_{A_1} = {\delta_1}^{-1} &: A_1 \longmapsto A_1 \cup B_1 \cup C_1 \cup D_1 \cup E_1,\\
f|_{B_1} = \delta_2\delta_0 &: B_1 \longmapsto A_2 \cup B_2 \cup C_2 \cup D_2 \cup E_2,\\
f|_{C_1} = {\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1} &: C_1 \longmapsto A_3 \cup B_3 \cup C_3 \cup D_3 \cup E_3,\\
f|_{D_1} = \delta_0 &: D_1 \longmapsto A_1 \cup B_1 \cup D_1 \cup E_1,\\
f|_{E_1} = {\delta_1}^{-1} &: E_1 \longmapsto F_0 \cup G_0 \cup H_0 \cup I_0 \cup J_0 \cup F_1 \cup G_1 \cup H_1 \cup I_1,\\
f|_{A_2} = \delta_2 &: A_2 \longmapsto A_2 \cup B_2 \cup C_2 \cup D_2 \cup E_2,\\
f|_{B_2} = {\delta_1}^{-1}{\delta_0}^{-1} &: B_2 \longmapsto A_1 \cup B_1 \cup C_1 \cup D_1 \cup E_1,\\
f|_{C_2} = \delta_1\delta_0\delta_2 &: C_2 \longmapsto A_0 \cup B_0 \cup C_0 \cup D_0 \cup E_0,
\end{aligned}
$$

$$
\begin{aligned}
f|_{D_2} = {\delta_0}^{-1}\colon D_2 &\longmapsto A_2 \cup B_2 \cup D_2 \cup E_2,\\
f|_{E_2} = \delta_2\colon E_2 &\longmapsto F_2 \cup G_2 \cup H_2 \cup I_2 \cup J_2 \cup F_3 \cup G_3 \cup H_3 \cup I_3,\\
f|_{A_3} = {\delta_2}^{-1}\colon A_3 &\longmapsto A_3 \cup B_3 \cup C_3 \cup D_3 \cup E_3,\\
f|_{B_3} = \delta_1\delta_0\colon B_3 &\longmapsto A_0 \cup B_0 \cup C_0 \cup D_0 \cup E_0,\\
f|_{C_3} = {\delta_1}^{-1}{\delta_0}^{-1}{\delta_2}^{-1}\colon C_3 &\longmapsto A_1 \cup B_1 \cup C_1 \cup D_1 \cup E_1,\\
f|_{D_3} = \delta_0\colon D_3 &\longmapsto A_3 \cup B_3 \cup D_3 \cup E_3,\\
f|_{E_3} = {\delta_2}^{-1}\colon E_3 &\longmapsto F_2 \cup G_2 \cup H_2 \cup I_2 \cup J_2 \cup F_3 \cup G_3 \cup H_3 \cup I_3.
\end{aligned}
$$

Having defined f on all of $\mathcal{F}_2$, we now consider $\mathcal{F}_1$. As we described in the previous section, we may regard (γ_1, γ_2) as lying on a one-holed torus (one of the components of $\Sigma_2 - \beta$). Moreover, as γ_1 and γ_2 intersect exactly once they correspond to Farey neighbours. Therefore, we define f to agree with the map f_1 on the space of Farey pairs and described in section 2.4. Results about the image of this map follow as in that section. Thus we have:

$$
\begin{aligned}
f|_{F_0} = \delta_2\colon F_0 &\longmapsto F_0 \cup G_0 \cup H_0 \cup I_0 \cup J_0 \cup H_1 \cup I_1,\\
f|_{G_0} = {\delta_0}^{-1}\colon G_0 &\longmapsto F_0 \cup G_0 \cup H_0 \cup I_0 \cup J_0 \cup H_1 \cup I_1,\\
f|_{F_1} = {\delta_2}^{-1}\colon F_1 &\longmapsto F_1 \cup G_1 \cup H_0 \cup I_0 \cup J_0 \cup H_1 \cup I_1,\\
f|_{G_1} = \delta_0\colon G_1 &\longmapsto F_1 \cup G_1 \cup H_0 \cup I_0 \cup J_0 \cup H_1 \cup I_1,\\
f|_{F_2} = \delta_1\colon F_2 &\longmapsto F_2 \cup G_2 \cup H_2 \cup I_2 \cup J_2 \cup H_3 \cup I_3,\\
f|_{G_2} = {\delta_0}^{-1}\colon G_2 &\longmapsto F_2 \cup G_2 \cup H_2 \cup I_2 \cup J_2 \cup H_3 \cup I_3,\\
f|_{F_3} = {\delta_1}^{-1}\colon F_3 &\longmapsto F_3 \cup G_3 \cup H_2 \cup I_2 \cup J_2 \cup H_3 \cup I_3,\\
f|_{G_3} = \delta_0\colon G_3 &\longmapsto F_3 \cup G_3 \cup H_2 \cup I_2 \cup J_2 \cup H_3 \cup I_3.
\end{aligned}
$$

Finally, we consider Farey blocks in $\mathcal{F}_0$. If we were in the case of the once punctured torus we would need to apply a power of the involution $\delta_1\delta_0\delta_1$. For the twice punctured torus, this element has infinite order (its fourth power is Dehn twist about χ_2). Therefore we need to investigate the effect of powers of $\delta_1\delta_0\delta_1$ and $\delta_2\delta_0\delta_2$. For simplicity we denote $\delta_j\delta_0\delta_j$ by ρ_j for $j = 1, 2$. We have the following lemma.

Lemma 4.6.2. *If the Farey block $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is in H_0 then its image under ${\rho_2}^{-1}$ is in I_0 if $a \geq 1$ or is in J_0 if $a = 0$. If $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is in I_0 then its image under ${\rho_2}^{-2}$ is in I_0 if $a \geq 1$ or is in J_0 if $a = 0$. Moreover any Farey block in I_0 or J_0 arises as the image of a Farey block in H_0 or I_0 in this way.*

Proof. The curve χ_0 is fixed under application of δ_0 or δ_2 and so under any power of ρ_2. It follows by a similar argument to those given in section 2.1 that ρ_2 interchanges $\mathbf{e}_0$ and $\mathbf{e}_\infty^\infty$. It remains to check the effect of ρ_2 on the possible curves γ_3.

If the Farey block $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is in I_0 then $\gamma_3 = (a+1, 0, a, 1) \in \Delta_5 \cap \Delta_6$. We claim that ${\rho_2}^{-1}$ sends $(a+1, 0, a, 1) \in \Delta_5 \cap \Delta_6$ to $(a, 0, a, 1) \in \Delta_5 \cap \Delta_6$. Likewise, if $(\gamma_1, \gamma_2, \beta; \gamma_3)$ is in H_0 then $\gamma_3 = (a, 0, a, 1) \in \Delta_5 \cap \Delta_6$ and we claim that ${\rho_2}^{-1}$ sends $(a, 0, a, 1) \in \Delta_5 \cap \Delta_6$ to $(a, 0, a-1, 1) \in \Delta_5 \cap \Delta_6$ if $a \geq 1$ or to $\mathbf{e}^0$ if $a = 0$.

This claim is proved using Propositions 3.3.2 and 3.3.3 as follows:

$$\begin{array}{cc}
(a+1, 0, a, 1) \in \Delta_5 \cap \Delta_6 & (a, 0, a, 1) \in \Delta_5 \cap \Delta_6 \\
\downarrow {\delta_2}^{-1} & \downarrow {\delta_2}^{-1} \\
(a+1, 0, a, 1) \in \Delta_5 \cap \Delta_6 & (a, 1, a-1, 1) \in \Delta_6 \\
\downarrow {\delta_0}^{-1} & \downarrow {\delta_0}^{-1} \\
(a, 1, a, 1) \in \Delta_5 & (a, 0, a-1, 1) \in \Delta_5 \cap \Delta_6 \\
\downarrow {\delta_2}^{-1} & \downarrow {\delta_2}^{-1} \\
(a, 0, a, 1) \in \Delta_5 \cap \Delta_6 & (a, 0, a-1, 1) \in \Delta_5 \cap \Delta_6
\end{array}$$

where we assume $a \geq 1$ in the right hand column. If $a = 0$ then

$$\rho_2{}^{-1}(\mathbf{e}^1_{-1}) = \delta_2{}^{-1}\delta_0{}^{-1}\delta_2{}^{-1}(\mathbf{e}^1_{-1}) = \delta_2{}^{-1}\delta_0{}^{-1}(\mathbf{e}^1_{\infty}) = \delta_2{}^{-1}(\mathbf{e}^0) = \mathbf{e}^0.$$

□

Corollary 4.6.3. *Applying ${\rho_2}^{-j}$ to Farey blocks in H_0, I_0 decreases the length of γ_3 by $2j$.* □

Therefore we define f on $\mathcal{F}_0$ as follows.

$$\begin{aligned}
f|_{H_0} &= {\rho_2}^{-1}\colon H_0 \longrightarrow I_0 \cup J_0, & f|_{I_0} &= {\rho_2}^{-2}\colon I_0 \longrightarrow I_0 \cup J_0, \\
f|_{H_1} &= \rho_2\colon H_1 \longrightarrow I_1 \cup J_0, & f|_{I_1} &= {\rho_2}^{2}\colon I_1 \longrightarrow I_1 \cup J_0, \\
f|_{H_2} &= {\rho_1}^{-1}\colon H_2 \longrightarrow I_2 \cup J_2, & f|_{I_2} &= {\rho_1}^{-2}\colon I_2 \longrightarrow I_2 \cup J_2, \\
f|_{H_3} &= \rho_1\colon H_3 \longrightarrow I_3 \cup J_2, & f|_{I_3} &= {\rho_1}^{2}\colon I_3 \longrightarrow I_3 \cup J_2, \\
f|_{J_0} &= e\colon J_0 \longrightarrow J_0, & f|_{J_2} &= \iota_2\colon J_2 \longrightarrow J_0.
\end{aligned}$$

This completes the definition of the map f.

4.7. The word acceptor.

We have now subdivided the space $\mathcal{F}$ of Farey blocks into states and defined the map f on each state, in such a way that

(i) for each state U, we have $f|_U \equiv \phi$ for some $\phi \in \mathcal{MCG}(\Sigma_2)$, and

(ii) for any states U and V if $f(U) \cap V^\circ \neq \emptyset$ then $V \subset f(U)$.

In order for f to define a normal form leading to a suitable word acceptor, we now only need to verify that successive applications of the map f always eventually terminate in the end state J_0. Of course, this requirement was central to our choice of definition of the map f. In analogy with the case of the once punctured torus Σ_1 we have:

Proposition 4.7.1. *Let $(\gamma_1, \gamma_2, \beta; \gamma_3)$ be any Farey block in $\mathcal{F}$. There exists a non-negative integer n so that*

$$f^n(\gamma_1, \gamma_2, \beta; \gamma_3) = (\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0).$$

Proof. This is similar to Proposition 2.3.5. First, if $(\gamma_1, \gamma_2, \beta; \gamma_3) \in \mathcal{F}_2$ it follows from Proposition 4.2.2 and Proposition 4.4.2. that f strictly decreases $|\beta|$. Thus there is a non-negative integer n_2 so that $f^{n_2}(\gamma_1, \gamma_2, \beta; \gamma_3)$ is in $\mathcal{F}_1$ or $\mathcal{F}_0$. For any Farey block in $\mathcal{F}_1$ the map f strictly decreases $|\gamma_1| + |\gamma_2|$. This follows from Corollary 2.3.4. Thus there is a non-negative integer n_1 so that $f^{n_2+n_1}(\gamma_1, \gamma_2, \beta; \gamma_3)$ is in $\mathcal{F}_0$. For any Farey block in $\mathcal{F}_0$ other than $(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0)$ or $(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2; \mathbf{e}_0)$, using Corollary 4.6.3 we see that the map strictly decreases $|\gamma_3|$. Thus there is a non-negative integer n_0 so that $f^{n_2+n_1+n_0}(\gamma_1, \gamma_2, \beta; \gamma_3)$ is either $(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0)$ or $(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2; \mathbf{e}_0)$. Finally

$$f|_{J_2}(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2; \mathbf{e}_0) = \iota_2(\mathbf{e}^0, \mathbf{e}_\infty^\infty, \chi_2; \mathbf{e}_0) = (\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0).$$

□

Remark. Just as in section 2 this proposition proves that the actions of f and $\mathcal{MCG}(\Sigma_2)$ are **orbit equivalent**, in other words, for any $(\gamma_1, \gamma_2, \beta; \gamma_3)$ and $(\gamma_1', \gamma_2', \beta'; \gamma_3')$ in $\mathcal{F}$ we have

$$(\gamma_1', \gamma_2', \beta'; \gamma_3') = \big(\phi(\gamma_1), \phi(\gamma_2), \phi(\beta); \phi(\gamma_3)\big)$$

for $\phi \in \mathcal{MCG}(\Sigma_2)$ if and only if there exist non negative integers m, n so that

$$f^n(\gamma_1, \gamma_2, \beta; \gamma_3) = f^m(\gamma_1', \gamma_2', \beta'; \gamma_3').$$

See the remark following Proposition 2.3.5 for the significance of this observation.

Proposition 4.7.1 allows us to construct a normal form for elements of $\mathcal{MCG}(\Sigma_2)$. Namely, for any $\phi \in \mathcal{MCG}(\Sigma_2)$, we determine the map f^n for which

$$f^n\big(\phi(\mathbf{e}_0), \phi(\mathbf{e}_\infty^\infty), \phi(\chi_0); \phi(\mathbf{e}^0)\big) = (\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0).$$

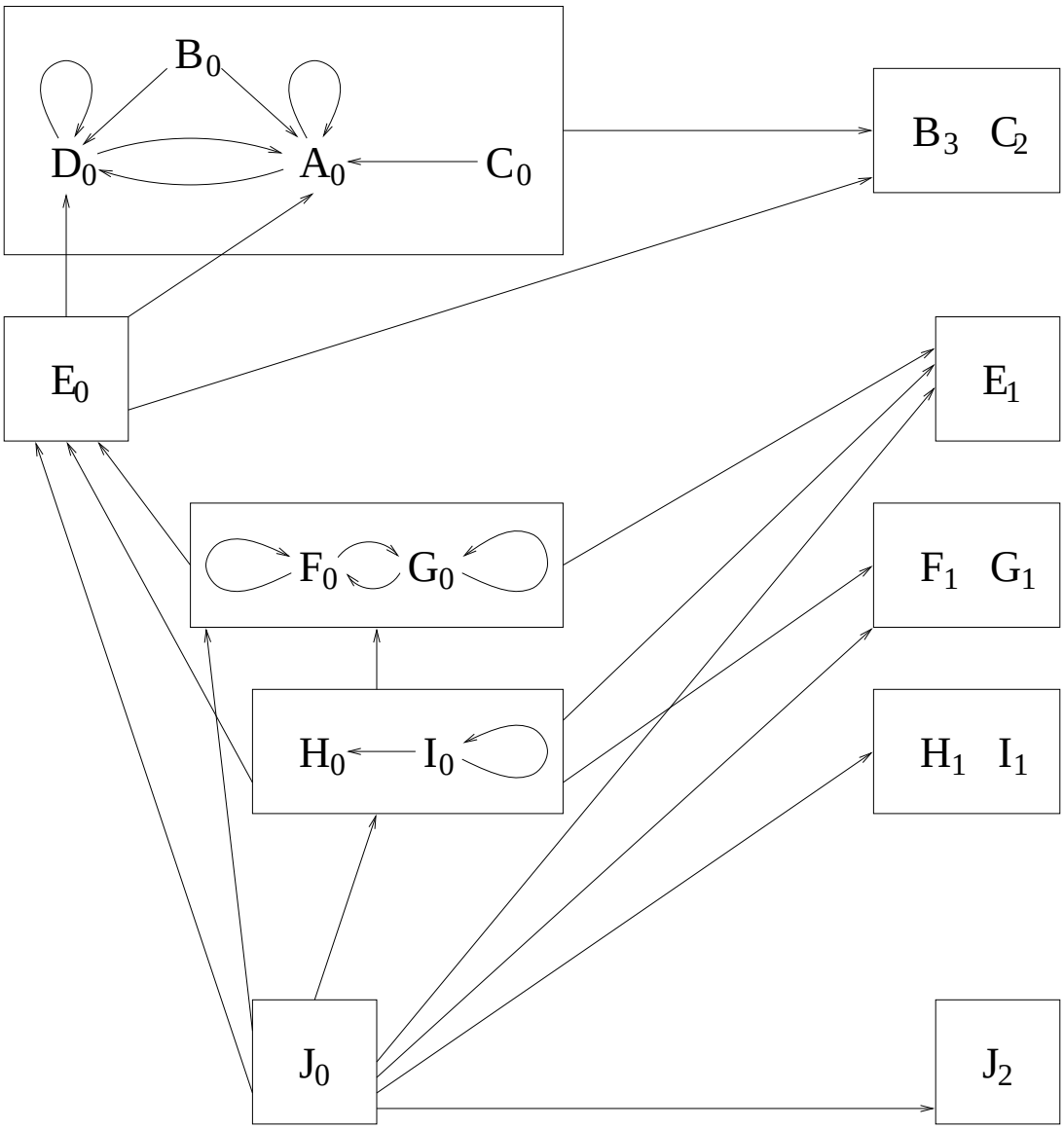


Fig. 4.7.1. A diagram of the arrows from states $A_0, \ldots, J_0$. An arrow from (or to) a box means that there should be an arrow from (or to) each state in that box. Each arrow should be labelled with the inverse of f on the state it points towards.

Since at each stage f is a fixed element of $\mathcal{MCG}(\Sigma_2)$ this, together with the fact that $\mathcal{MCG}(\Sigma_2)$ acts freely on $\mathcal{F}$, gives a unique expression for ϕ. The details of the normal form are now rather complicated and are best described in terms of a **word acceptor** for $\mathcal{MCG}(\Sigma_2)$.

The states of the word acceptor are obviously $A_j, \ldots, I_j$ for $j = 0, 1, 2, 3$ and J_0, J_2. To get the arrows we need to invert the map f on each separate state. The alphabet $\mathcal{A}$

which labels these arrows will consist of all possible values of f^{-1}, namely:

$$\mathcal{A} = \left\{ \begin{array}{c} e, \iota_2, {\delta_0}^{\pm 1}, {\delta_1}^{\pm 1}, {\delta_2}^{\pm 1}, \delta_1\delta_0, {\delta_1}^{-1}{\delta_0}^{-1}, \delta_2\delta_0, {\delta_2}^{-1}{\delta_0}^{-1}, \\ (\delta_1\delta_0\delta_2)^{\pm 1}, (\delta_2\delta_0\delta_1)^{\pm 1}, {\rho_1}^{\pm 1}, {\rho_2}^{\pm 1}, {\rho_1}^{\pm 2}, {\rho_2}^{\pm 2}, \end{array} \right\}.$$

We list the labelled arrows leading from each state. We begin with J_0. It has the following arrows which may be read off from the definition of f. The arrows from J_0 are.

$$\begin{array}{lllll} J_0 \xrightarrow{{\delta_1}^{-1}} E_0, & J_0 \xrightarrow{{\delta_2}^{-1}} F_0, & J_0 \xrightarrow{\delta_0} G_0, & J_0 \xrightarrow{\rho_2} H_0, & J_0 \xrightarrow{{\rho_2}^2} I_0, \\ J_0 \xrightarrow{\delta_1} E_1, & J_0 \xrightarrow{\delta_2} F_1, & J_0 \xrightarrow{{\delta_0}^{-1}} G_1, & J_0 \xrightarrow{{\rho_2}^{-1}} H_1, & J_0 \xrightarrow{{\rho_2}^{-2}} I_1, \\ & & J_0 \xrightarrow{\iota_2} J_2. & & \end{array}$$

The arrows from J_2 may be found from those from J_0 by applying the symmetry ι_2, with one exception: there is no arrow from J_2 to J_0.

The arrows from I_j may be found by applying ι_j to the following arrows from I_0:

$$\begin{array}{lllll} I_0 \xrightarrow{{\delta_1}^{-1}} E_0, & I_0 \xrightarrow{{\delta_2}^{-1}} F_0, & I_0 \xrightarrow{\delta_0} G_0, & I_0 \xrightarrow{\rho_2} H_0, & I_0 \xrightarrow{{\rho_2}^2} I_0, \\ I_0 \xrightarrow{\delta_1} E_1, & I_0 \xrightarrow{\delta_2} F_1, & I_0 \xrightarrow{{\delta_0}^{-1}} G_1. & & \end{array}$$

The arrows from H_j may be found by applying ι_j to the following arrows from H_0:

$$\begin{array}{lll} H_0 \xrightarrow{{\delta_1}^{-1}} E_0, & H_0 \xrightarrow{{\delta_2}^{-1}} F_0, & H_0 \xrightarrow{\delta_0} G_0, \\ H_0 \xrightarrow{\delta_1} E_1, & H_0 \xrightarrow{\delta_2} F_1, & H_0 \xrightarrow{{\delta_0}^{-1}} G_1. \end{array}$$

The arrows from F_j and G_j are very similar. They may be found by applying ι_j to the following arrows from F_0 and G_0:

$$\begin{array}{llll} F_0 \xrightarrow{{\delta_1}^{-1}} E_0, & F_0 \xrightarrow{\delta_1} E_1, & F_0 \xrightarrow{{\delta_2}^{-1}} F_0, & F_0 \xrightarrow{\delta_0} G_0, \\ G_0 \xrightarrow{{\delta_1}^{-1}} E_0, & G_0 \xrightarrow{\delta_1} E_1, & G_0 \xrightarrow{{\delta_2}^{-1}} F_0, & G_0 \xrightarrow{\delta_0} G_0. \end{array}$$

The arrows from A_0, B_0, D_0 and E_0 are all similar. They are

$$\begin{array}{llll} A_0 \xrightarrow{\delta_0} D_0, & A_0 \xrightarrow{{\delta_1}^{-1}} A_0, & A_0 \xrightarrow{{\delta_0}^{-1}{\delta_1}^{-1}} B_3, & A_0 \xrightarrow{{\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1}} C_2, \\ B_0 \xrightarrow{\delta_0} D_0, & B_0 \xrightarrow{{\delta_1}^{-1}} A_0, & B_0 \xrightarrow{{\delta_0}^{-1}{\delta_1}^{-1}} B_3, & B_0 \xrightarrow{{\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1}} C_2, \\ D_0 \xrightarrow{\delta_0} D_0, & D_0 \xrightarrow{{\delta_1}^{-1}} A_0, & D_0 \xrightarrow{{\delta_0}^{-1}{\delta_1}^{-1}} B_3, & D_0 \xrightarrow{{\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1}} C_2, \\ E_0 \xrightarrow{\delta_0} D_0, & E_0 \xrightarrow{{\delta_1}^{-1}} A_0, & E_0 \xrightarrow{{\delta_0}^{-1}{\delta_1}^{-1}} B_3, & E_0 \xrightarrow{{\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1}} C_2. \end{array}$$

Finally the arrows from C_0 are

$$C_0 \xrightarrow{{\delta_1}^{-1}} A_0, \qquad C_0 \xrightarrow{{\delta_0}^{-1}{\delta_1}^{-1}} B_3, \qquad C_0 \xrightarrow{{\delta_2}^{-1}{\delta_0}^{-1}{\delta_1}^{-1}} C_2.$$

The arrows from A_j, B_j, C_j, D_j and E_j for $j = 1, 2, 3$ may be found by applying the symmetry ι_j to A_0, B_0, C_0, D_0 or E_0 respectively.

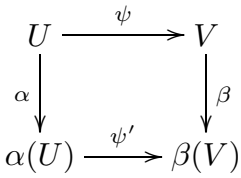
In Fig. 4.7.1 we have given a schematic representation of the arrows from $A_0, \ldots, J_0$ listed above. In order to simplify the diagram we have drawn a single arrow to represent several between different pairs of states. In order to reconstruct the word acceptor, the diagram should be reproduced with all suffices $j = 0, 1, 2, 3$ by applying symmetries ι_j. The arrows to the right hand columns indicate how these four different diagrams are linked. An arrow between two of the rectangular boxes should be replaced with arrows between all the states in each of the two boxes. Finally, the arrows from state U_j to V_k should be labelled with the inverse of $f|_{V_k}$. Observe that all arrows between boxes either go upwards or across but never downwards. This gives the word acceptor the structure of a partially ordered set.

5. The word difference machine

In this final section, following the procedures of sections 1.2 and 2.5, we construct a word difference machine for the word acceptor of $\mathcal{MCG}(\Sigma_2)$.

5.1. Outline of the construction

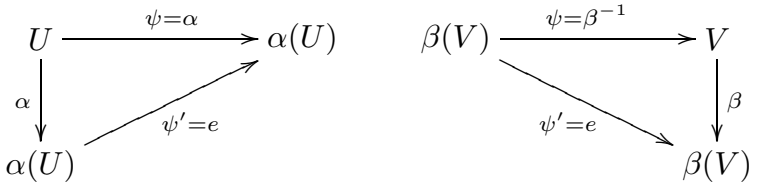
The construction of the word difference machine is very similar to the construction for the once punctured torus given earlier. Our notation will follow that established in the introduction to section 2.5. As before, the word difference machine is a 2-stringed finite state automaton. Its states are the elements of a set of word differences $\mathcal{D}$. As before, the basic building blocks are squares



$(*)$

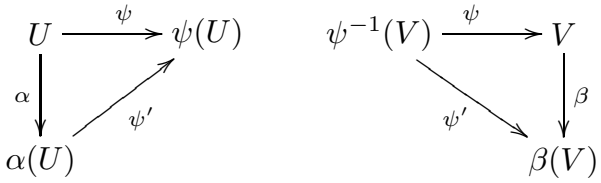
where now U, V are subsets of the states $A_j, \ldots, I_j$, J_0, $J_2 \subset \mathcal{F}$. As usual, in such a square $\psi, \psi \in \mathcal{D}$, $\psi(U) = V$, $\alpha = f|_U$, $\beta = f|_V$ and $\psi' = \beta\psi\alpha^{-1}$.

In addition to degenerate squares, or triangles, of type $(**)$



$(**)$

used un section 2.5 we also introduce further degenerate squares where we only apply f to either U or to V. If we do not apply f to U, then we may relax the requirement that U be contained in a single state in the partition of $\mathcal{F}$. In this case we write $\psi^{-1}(V)$ in place of U. Likewise when we do not apply f to V we write $\psi(U)$ instead of V and allow it to contain points in more than one state.



$(***)$

We require that $\psi' = \psi\alpha^{-1}$ or $\psi' = \beta\psi$ respectively is in $\mathcal{D}$. We need to be careful that only finitely many such triangles occur in each path through the word difference machine. This is a key point which we will discuss later.

The arrows in the word difference machine will consist of pairs $(\alpha^{-1}, \beta^{-1})$ corresponding to the diagram $(*)$ or $(\alpha^{-1}, -)$, $(-, \beta^{-1})$ corresponding to $(**)$ or $(***)$. For example, a square of type $(*)$ will give an arrow $(\alpha^{-1}, \beta^{-1})$ from $\psi' \in \mathcal{D}$ to $\psi \in \mathcal{D}$.

The set $\mathcal{D}$ should contain all elements of the alphabet $\mathcal{A}$ constructed in the previous section. In fact we will begin by considering the set

$$\mathcal{D}_0 = \left\{e, \iota_2, {\delta_0}^{\pm 1}, {\delta_1}^{\pm 1}, {\delta_2}^{\pm 1}\right\}.$$

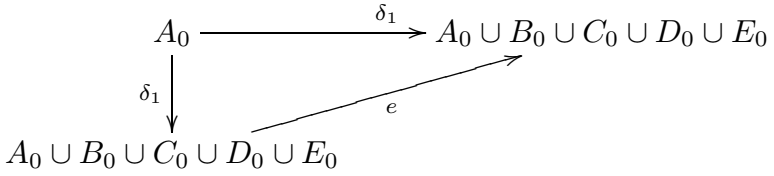
As all the elements of $\mathcal{A}$ have length at most six in these letters we can break squares involving word differences in $\mathcal{A}$ down into at most six squares (placed horizontally) involving word differences in $\mathcal{D}_0$ (see section 5.5 below). During our construction we will add to the list $\mathcal{D}_0$. Recall that in Section 3.3 we found various relations in $\mathcal{MCG}(\Sigma_2)$. We will use these when constructing the squares. The fact that we need to use no more relations is the proof that the presentation for $\mathcal{MCG}(\Sigma_2)$ given in Theorem 3.2.1 works.

Because of the stratification of states in $\mathcal{F}$ described in section 4.5, the normal form given by the word acceptor for each element ϕ in $\mathcal{MCG}(\Sigma_2)$ may be broken down as $\phi = \phi_2\phi_1\phi_0$ where $\phi_j(\mathbf{e}_0, \mathbf{e}_\infty^\infty, \chi_0; \mathbf{e}^0)$ is in $\mathcal{F}_j$ for $j = 0, 1, 2$. We will break the word difference machine into subgraphs which correspond to these pieces. There will additionally be a fourth subgraph which will correspond to certain special word differences, which we call exceptional (see section 5.4 below).

5.2. Squares and triangles arising from states in $\mathcal{F}_2$.

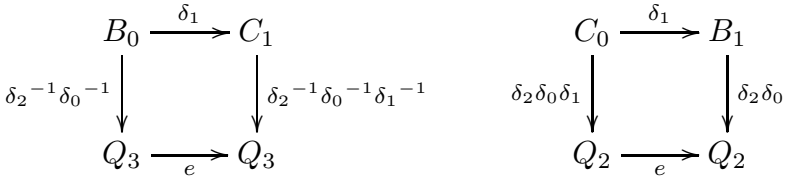
We begin by constructing squares and triangles where U and V are contained in $\mathcal{F}_2$ and where $\psi \in \mathcal{D}_0 = \{{\delta_j}^{\pm 1} | j = 0, 1, 3\} \cup \{\iota_2\}$. By use of the symmetries ι_j for $j = 1, 2, 3$ we may restrict our attention to the case where U and V are subsets of $\mathbf{A}_0$, $\mathbf{B}_0$, $\mathbf{C}_0$, $\mathbf{D}_0$ and $\mathbf{E}_0$. As usual, we drop the distinction between $\mathbf{A}_0, \ldots, \mathbf{E}_0$ and $A_0, \ldots, E_0$.

Since $\delta_1 = f|_{A_0}$ we already know its effect on A_0. Thus we can write down the triangle:



In order to simplify things further, we define $Q_j = A_j \cup B_j \cup C_j \cup D_j \cup E_j$ for $j = 0, 1, 2, 3$.

Now suppose that U is one of B_0 or C_0. Using Proposition 3.3.1, we obtain two squares:



When considering D_0 and E_0 we make the following definitions which again simplify the notation.

$$R_j = A_j \cup B_j \cup D_j \cup E_j \quad \text{for } j = 0, 1, 2, 3;$$
$$S_j = F_j \cup G_j \quad \text{for } j = 0, 1, 2, 3;$$
$$T_j = H_j \cup I_j \cup J_j \cup H_{j+1} \cup I_{j+1} \quad \text{for } j = 0, 2.$$

We obtain

$$\begin{array}{ccc} D_0 & \xrightarrow{\delta_1} & D_1 \\ {\scriptstyle \delta_0{}^{-1}}\downarrow & & \downarrow{\scriptstyle \delta_0} \\ R_0 & \xrightarrow{\delta_1\delta_0\delta_1} & R_1 \end{array} \qquad \begin{array}{ccc} E_0 & \xrightarrow{\delta_1} & S_0 \cup S_1 \cup T_0 \\ {\scriptstyle \delta_1}\downarrow & \nearrow_{e} & \\ S_0 \cup S_1 \cup T_0 & & \end{array}$$

Now we do the same for $\psi = {\delta_1}^{-1}$. We obtain one triangle:

$$\begin{array}{ccc} Q_0 & \xrightarrow{\delta_1{}^{-1}} & A_0 \\ & {\scriptstyle e}\searrow & \downarrow{\scriptstyle \delta_1} \\ & & Q_0 \end{array}$$

Now we consider squares where $\psi = \delta_2$. From Proposition 3.3.2 we know the effect of δ_2 on $\Delta_1, \ldots, \Delta_7$. We know that $\delta_2 \colon A_0 \longrightarrow A_0$ and $\delta_2 \colon B_0 \longrightarrow B_0 \cup D_0$. We divide $U = B_0$ into maximal subsets $U_1 = {\delta_2}^{-1}(B_0)$ or $U_2 = {\delta_2}^{-1}(D_0)$. This gives:

$$\begin{array}{ccc} A_0 & \xrightarrow{\delta_2} & A_0 \\ {\scriptstyle \delta_1}\downarrow & & \downarrow{\scriptstyle \delta_1} \\ Q_0 & \xrightarrow{\delta_2} & Q_0 \end{array} \qquad \begin{array}{ccc} {\delta_2}^{-1}(B_0) & \xrightarrow{\delta_2} & B_0 \\ {\scriptstyle {\delta_2}^{-1}\delta_0{}^{-1}}\downarrow & & \downarrow{\scriptstyle {\delta_2}^{-1}\delta_0{}^{-1}} \\ C_3 \cup D_3 & \xrightarrow{\delta_0} & Q_3 \end{array} \qquad \begin{array}{ccc} {\delta_2}^{-1}(D_0) & \xrightarrow{\delta_2} & D_0 \\ {\scriptstyle {\delta_2}^{-1}\delta_0{}^{-1}}\downarrow & & \downarrow{\scriptstyle \delta_0{}^{-1}} \\ A_3 \cup B_3 \cup E_3 & \xrightarrow{\delta_2\delta_0} & Q_0 \end{array}$$

Now $\delta_2(C_0 \cup D_0) = C_0$. Therefore we divide $V = C_0$ into subsets $V_1 = \delta_2(C_0)$ and $V_2 = \delta_2(D_0)$.

$$\begin{array}{ccc} C_0 & \xrightarrow{\delta_2} & \delta_2(C_0) \\ {\scriptstyle \delta_2\delta_0\delta_1}\downarrow & & \downarrow{\scriptstyle \delta_2\delta_0\delta_1} \\ Q_2 & \xrightarrow{\delta_0} & C_2 \cup D_2 \end{array} \qquad \begin{array}{ccc} D_0 & \xrightarrow{\delta_2} & \delta_2(D_0) \\ {\scriptstyle \delta_0{}^{-1}}\downarrow & & \downarrow{\scriptstyle \delta_2\delta_0\delta_1} \\ R_0 & \xrightarrow{\iota_2{\delta_2}^{-1}} & A_2 \cup B_2 \cup E_2 \end{array} \qquad \begin{array}{ccc} E_0 & \xrightarrow{\delta_2} & E_0 \\ {\scriptstyle \delta_1}\downarrow & & \downarrow{\scriptstyle \delta_1} \\ S_0 \cup S_1 \cup T_0 & \xrightarrow{\delta_2} & S_0 \cup S_1 \cup T_0 \end{array}$$

We remark that it is not immediately obvious that the bottom lines of these squares are as claimed. We now do an example which illustrates how the bottom line is found. The rest are simple and are left to the reader.

The example we choose is

$$\begin{array}{ccc} {\delta_2}^{-1}(B_0) \cap B_0 & \xrightarrow{\delta_2} & B_0 \\ \downarrow{\scriptstyle {\delta_2}^{-1}{\delta_0}^{-1}} & & \downarrow{\scriptstyle {\delta_2}^{-1}{\delta_0}^{-1}} \\ C_3 \cup D_3 & \xrightarrow{\delta_0} & Q_3 \end{array}$$

We do this using the following result

Proposition 5.2.1. *If $(a,b,c,d) \in \Delta_3 \cup \Delta_4 = B_0$ and $\delta_2(a,b,c,d) \in \Delta_3 \cup \Delta_4 = B_0$ then*

$$ {\delta_2}^{-1}{\delta_0}^{-1}(a,b,c,d) \in \Delta_{26} \cup \Delta_{27} \cup \Delta_{28} = C_3 \cup D_3$$

and

$$ {\delta_2}^{-1}{\delta_0}^{-1}\delta_2(a,b,c,d) = \delta_0{\delta_2}^{-1}{\delta_0}^{-1}(a,b,c,d) \in Q_3.$$

Moreover, any point of Q_3 arises in this way.

Proof. The word difference can of course be found without reference to the states on the bottom line of the diagram. It is clear that in the present case the new word difference ψ' is

$$\big({\delta_2}^{-1}{\delta_0}^{-1}\big)\delta_2\big({\delta_2}^{-1}{\delta_0}^{-1}\big)^{-1} = {\delta_2}^{-1}{\delta_0}^{-1}\delta_2\delta_0\delta_2 = \delta_0$$

using the relation $\delta_2\delta_0\delta_2 = \delta_0\delta_2\delta_0$.

We now justify the claims about the regions involved in this diagram. We begin by identifying $U_1 = B_0 \cap {\delta_2}^{-1}(B_0)$. We know that $B_0 = \Delta_3 \cup \Delta_4$. By Proposition 3.3.2 (iii) and (iv) we see that $(a,b,c,d) \in \Delta_3$ is sent to $\Delta_3 \cup \Delta_4 = B_0$ provided $a \leq b+d$ and that $(a,b,c,d) \in \Delta_4$ is sent to B_0 provided $a \leq d$. Thus,

$$U_1 = \{(a,b,c,d) \in \Delta_3 \colon a \leq b+d\} \cup \{(a,b,c,d) \in \Delta_4 \colon a \leq d\}.$$

By definition $\delta_2(U_1) \subset B_0$, and it is not hard to check that $\delta_2(U_1) = B_0$.

We now investigate the effect of ${\delta_2}^{-1}{\delta_0}^{-1}$ on U. We know from Proposition 3.4.3 (i) that ${\delta_0}^{-1}$ sends Δ_3 and Δ_4 to Δ_{23} and Δ_{22} respectively by ${\delta_0}^{-1}(a,b,c,d) = (a,b,c,d)$. Thus U is sent to the appropriate subset of A_3. Applying ι_3 to Proposition 3.3.1 (i) we see that if $(a,b,c,d) \in \Delta_{22}$ with $a \leq d$ then ${\delta_2}^{-1}(a,b,c,d)$ is in Δ_{26}. Similarly, if $(a,b,c,d) \in \Delta_{23}$ with $a \leq b+d$ then ${\delta_2}^{-1}(a,b,c,d)$ is in Δ_{27} or Δ_{28}. It is not hard to show that this map is surjective and so ${\delta_2}^{-1}{\delta_0}^{-1}(U) = C_3 \cup D_3$.

Finally, we know that ${\delta_2}^{-1}{\delta_0}^{-1}$ maps B_0 onto Q_3. □

By similar reasoning we construct all the squares and triangles for which $\psi = {\delta_2}^{-1}$.

$$\begin{array}{ccc} A_0 & \xrightarrow{\delta_2^{-1}} & A_0 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle \delta_1} \\ Q_0 & \xrightarrow{\delta_2^{-1}} & Q_0 \end{array} \qquad \begin{array}{ccc} B_0 & \xrightarrow{\delta_2^{-1}} & {\delta_2}^{-1}(B_0) \cap B_0 \\ \downarrow{\scriptstyle {\delta_2}^{-1}{\delta_0}^{-1}} & & \downarrow{\scriptstyle {\delta_2}^{-1}{\delta_0}^{-1}} \\ Q_3 & \xrightarrow{\delta_0^{-1}} & C_3 \cup D_3 \end{array}$$

$$\begin{array}{ccc} \delta_2(C_0)\cap C_0 & \xrightarrow{{\delta_2}^{-1}} & C_0 \\ \downarrow{\scriptstyle \delta_2\delta_0\delta_1} & & \downarrow{\scriptstyle \delta_2\delta_0\delta_1} \\ C_2\cup D_2 & \xrightarrow{{\delta_0}^{-1}} & Q_2 \end{array} \qquad \begin{array}{ccc} \delta_2(D_0)\cap C_0 & \xrightarrow{{\delta_2}^{-1}} & D_0 \\ \downarrow{\scriptstyle \delta_2\delta_0\delta_1} & & \downarrow{\scriptstyle {\delta_0}^{-1}} \\ A_2\cup B_2\cup E_2 & \xrightarrow{\iota_2\delta_1} & R_0 \end{array}$$

$$\begin{array}{ccc} D_0 & \xrightarrow{{\delta_2}^{-1}} & {\delta_2}^{-1}(D_0)\cap B_0 \\ \downarrow{\scriptstyle {\delta_0}^{-1}} & & \downarrow{\scriptstyle {\delta_2}^{-1}{\delta_0}^{-1}} \\ R_0 & \xrightarrow{{\delta_0}^{-1}{\delta_2}^{-1}} & A_3\cup B_3\cup E_3 \end{array} \qquad \begin{array}{ccc} E_0 & \xrightarrow{{\delta_2}^{-1}} & E_0 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle \delta_1} \\ S_0\cup S_1\cup T_0 & \xrightarrow{{\delta_2}^{-1}} & S_0\cup S_1\cup T_0 \end{array}$$

Similarly, when $\psi = \delta_0$ or ${\delta_0}^{-1}$ we obtain:

$$\begin{array}{ccc} R_0 & \xrightarrow{\delta_0} & D_0 \\ & {\scriptstyle e}\searrow & \downarrow{\scriptstyle {\delta_0}^{-1}} \\ & & R_0 \end{array} \qquad \begin{array}{ccc} C_0 & \xrightarrow{\delta_0} & C_0 \\ \downarrow{\scriptstyle \delta_2\delta_0\delta_1} & & \downarrow{\scriptstyle \delta_2\delta_0\delta_1} \\ Q_2 & \xrightarrow{\delta_1} & Q_2 \end{array}$$

$$\begin{array}{ccc} A_0 & \xrightarrow{{\delta_0}^{-1}} & B_3 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle \delta_1\delta_0} \\ Q_0 & \xrightarrow[e]{} & Q_0 \end{array} \qquad \begin{array}{ccc} B_0 & \xrightarrow{{\delta_0}^{-1}} & A_3 \\ \downarrow{\scriptstyle {\delta_2}^{-1}{\delta_0}^{-1}} & & \downarrow{\scriptstyle {\delta_2}^{-1}} \\ Q_3 & \xrightarrow[e]{} & Q_3 \end{array} \qquad \begin{array}{ccc} C_0 & \xrightarrow{{\delta_0}^{-1}} & C_0 \\ \downarrow{\scriptstyle \delta_2\delta_0\delta_1} & & \downarrow{\scriptstyle \delta_2\delta_0\delta_1} \\ Q_2 & \xrightarrow{{\delta_1}^{-1}} & Q_2 \end{array}$$

$$\begin{array}{ccc} D_0 & \xrightarrow{{\delta_0}^{-1}} & R_0 \\ \downarrow{\scriptstyle {\delta_0}^{-1}} & \nearrow{\scriptstyle e} & \\ R_0 & & \end{array} \qquad \begin{array}{ccc} E_0 & \xrightarrow{{\delta_0}^{-1}} & E_3 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle {\delta_2}^{-1}} \\ S_0\cup S_1\cup T_0 & \xrightarrow{\iota_2\rho_2} & S_3\cup S_3\cup T_2 \end{array}$$

Finally we consider the word difference ι_2.

$$\begin{array}{ccc} A_0 & \xrightarrow{\iota_2} & A_2 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle \delta_2} \\ Q_0 & \xrightarrow{\iota_2} & Q_2 \end{array} \qquad \begin{array}{ccc} B_0 & \xrightarrow{\iota_2} & B_2 \\ \downarrow{\scriptstyle {\delta_2}^{-1}{\delta_0}^{-1}} & & \downarrow{\scriptstyle {\delta_1}^{-1}{\delta_0}^{-1}} \\ Q_3 & \xrightarrow{\iota_2} & Q_1 \end{array} \qquad \begin{array}{ccc} C_0 & \xrightarrow{\iota_2} & C_2 \\ \downarrow{\scriptstyle \delta_2\delta_0\delta_1} & & \downarrow{\scriptstyle \delta_1\delta_0\delta_2} \\ Q_2 & \xrightarrow{\iota_2} & Q_0 \end{array}$$

$$\begin{array}{ccc} D_0 & \xrightarrow{\iota_2} & D_2 \\ \downarrow{\scriptstyle {\delta_0}^{-1}} & & \downarrow{\scriptstyle {\delta_0}^{-1}} \\ R_0 & \xrightarrow{\iota_2} & R_2 \end{array} \qquad \begin{array}{ccc} E_0 & \xrightarrow{\iota_2} & E_2 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle \delta_2} \\ S_0\cup S_1\cup T_0 & \xrightarrow{\iota_2} & S_2\cup S_3\cup T_2 \end{array}$$

Now consider the diagrams we have constructed above. If the bottom line consists of a word difference between regions we have already constructed then we have no more work. However, there are some diagrams for which this is not the case. First, there are those diagrams where the bottom line involves the regions making up S_j or T_j. We will consider these in the next section. Secondly, there are word differences between regions in Q_j or R_j which we have not yet considered. We consider these individually.

The easiest case is where we have a word difference of $\iota_2\psi$ where ψ is one of the word differences that we can already deal with. These can be analysed as follows. For each of the squares we have constructed with word difference ψ and written in the form $(*)$ we add a new square with word difference $\iota_2\psi$:

$$\begin{array}{ccccc} U & \xrightarrow{\psi} & V & \xrightarrow{\iota_2} & W \\ \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} \\ U' & \xrightarrow{\psi'} & V' & \xrightarrow{\iota_2} & W' \end{array}$$

where $W = \iota_2(V)$, $W' = \iota_2(V')$ and $\gamma = \iota_2\beta\iota_2$ is $f|_W$. If ψ' has the form $\iota_2\psi''$ for some ψ'', then we use ${\iota_2}^2 = e$ to get a word difference of ψ''. This is illustrated in the example given in section 5.5 below.

The remaining word differences we have to consider are

$$\delta_2\delta_0 \colon A_3 \cup B_3 \cup E_3 \longrightarrow R_0, \tag{1}$$

$${\delta_0}^{-1}{\delta_2}^{-1} \colon R_0 \longrightarrow A_3 \cup B_3 \cup E_3, \tag{2}$$

$$\rho_1 = \delta_1\delta_0\delta_1 \colon R_0 \longrightarrow R_1. \tag{3}$$

The word difference (3) is slightly more complicated than the others. We will treat this word difference and those arising from it separately in section 5.4. These will be called **exceptional word differences** and will constitute a separate subgraph of the word difference machine.

We now consider the word difference (1). By examining which point goes to which, we see that this may be broken into four new arrows. Namely

$$\delta_2\delta_0 \colon {\delta_0}^{-1}{\delta_2}^{-1}(B_0) \cap A_3 \longrightarrow B_0, \qquad \delta_2\delta_0 \colon {\delta_0}^{-1}{\delta_2}^{-1}(D_0) \cap A_3 \longrightarrow D_0,$$
$$\delta_2\delta_0 \colon B_3 \longrightarrow A_0, \qquad \delta_2\delta_0 \colon E_3 \longrightarrow E_0$$

Now ${\delta_0}^{-1}{\delta_2}^{-1}(B_0) \subset A_3$ and ${\delta_0}^{-1}{\delta_2}^{-1}(D_0) \subset A_3$. Thus we obtain the following squares

$$\begin{array}{ccc} {\delta_0}^{-1}{\delta_2}^{-1}(B_0) & \xrightarrow{\delta_2\delta_0} & B_0 \\ \downarrow{\scriptstyle{\delta_2}^{-1}} & & \downarrow{\scriptstyle{\delta_2}^{-1}{\delta_0}^{-1}} \\ C_3 \cup D_3 & \xrightarrow{\delta_0} & Q_3 \end{array} \qquad \begin{array}{ccc} {\delta_0}^{-1}{\delta_2}^{-1}(D_0) & \xrightarrow{\delta_2\delta_0} & D_0 \\ \downarrow{\scriptstyle{\delta_2}^{-1}} & & \downarrow{\scriptstyle{\delta_0}^{-1}} \\ A_3 \cup B_3 \cup E_3 & \xrightarrow{\delta_2\delta_0} & R_0 \end{array}$$

$$\begin{array}{ccc} B_3 & \xrightarrow{\delta_2\delta_0} & A_0 \\ \downarrow{\scriptstyle \delta_1\delta_0} & & \downarrow{\scriptstyle \delta_1} \\ Q_0 & \xrightarrow{\delta_2} & Q_0 \end{array} \qquad \begin{array}{ccc} E_3 & \xrightarrow{\delta_2\delta_0} & E_0 \\ \downarrow{\scriptstyle {\delta_2}^{-1}} & & \downarrow{\scriptstyle \delta_1} \\ S_2 \cup S_3 \cup T_2 & \xrightarrow{\iota_2{\delta_0}^{-1}{\delta_1}^{-1}} & S_0 \cup S_1 \cup T_0 \end{array}$$

The word differences on the bottom lines of these diagrams either have been considered above or else involve S_j and T_j. In the latter case we will consider them in the next section.

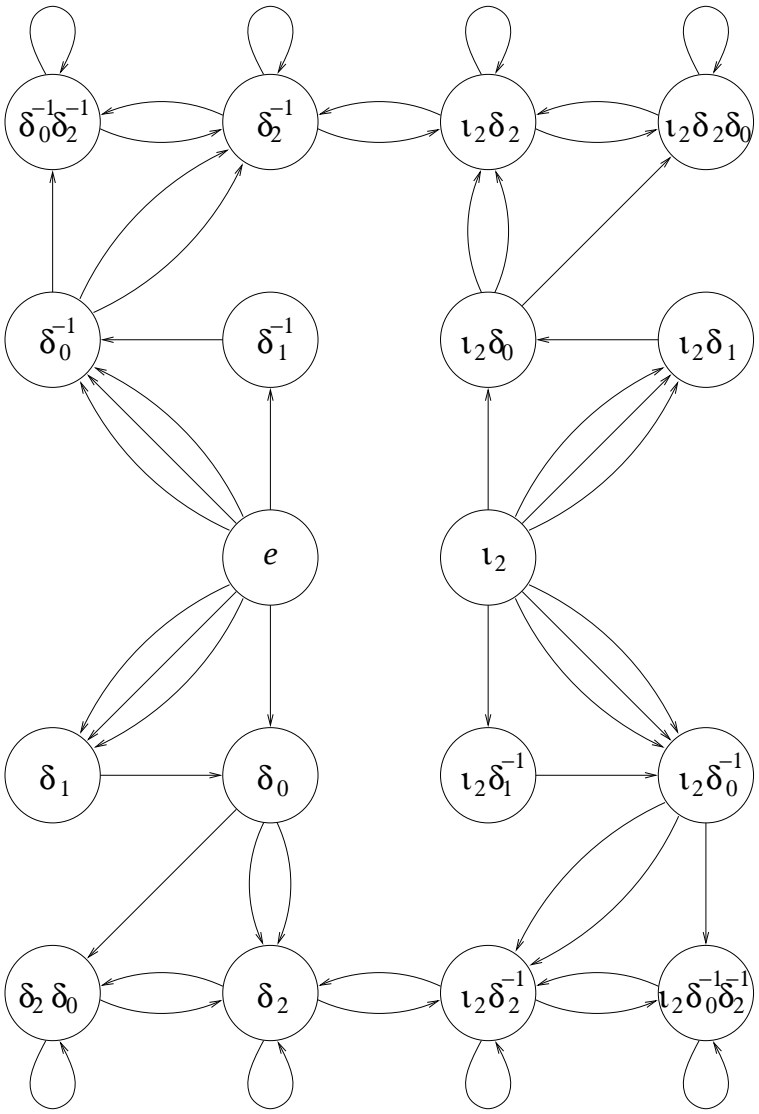
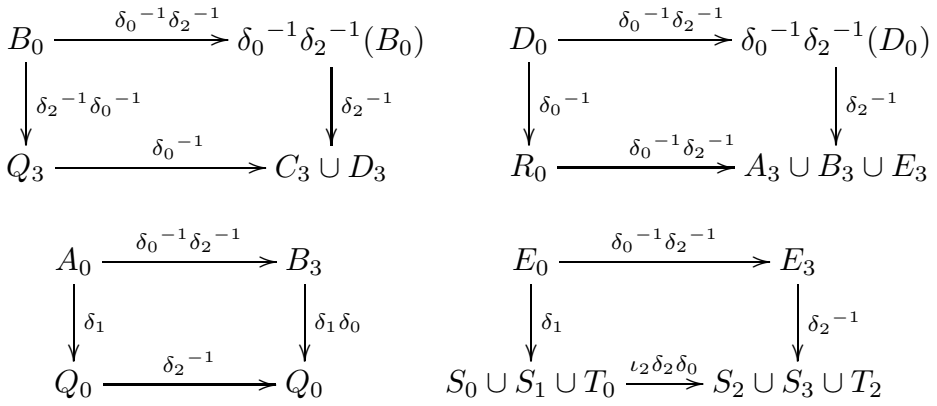


Fig. 5.2.1. Non-exceptional arrows in the word difference machine where U is one of A_0, B_0, C_0, D_0. Each arrow is labelled by a pair $(\alpha^{-1}, \beta^{-1})$ as described in section 5.1.

We now consider the word difference (2). It is rather similar to (1), and splits as

$$\delta_0{}^{-1}\delta_2{}^{-1}\colon B_0 \longrightarrow \delta_0{}^{-1}\delta_2{}^{-1}(B_0) \cap A_3, \qquad \delta_0{}^{-1}\delta_2{}^{-1}\colon A_0 \longrightarrow B_3,$$
$$\delta_0{}^{-1}\delta_2{}^{-1}\colon D_0 \longrightarrow \delta_0{}^{-1}\delta_2{}^{-1}(D_0) \cap A_3, \qquad \delta_0{}^{-1}\delta_2{}^{-1}\colon E_0 \longrightarrow E_3.$$

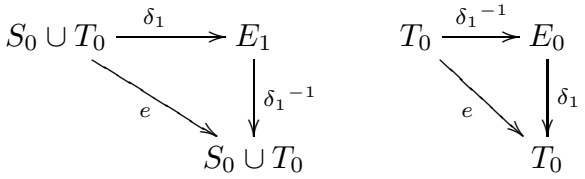


The word differences on the bottom lines of these diagrams either have been considered above or else involve S_j and T_j, again to be treated in the next section.

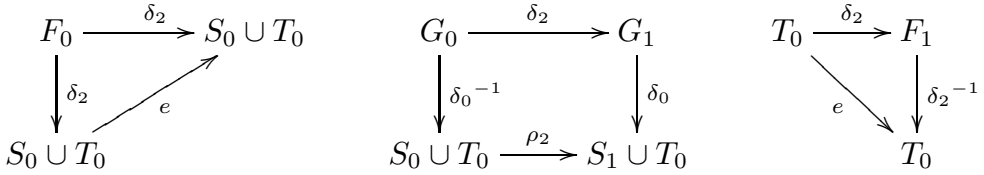
Fig. 5.2.1 shows the non-exceptional arrows constructed above where U is one of A_0, B_0, C_0 or D_0. The labels on the arrows may be obtained from the squares listed above. For example, the three arrows from e to δ_1 are labelled $({\delta_1}^{-1}, -)$, $(\delta_0\delta_2, \delta_1\delta_0\delta_2)$ and $({\delta_1}^{-1}{\delta_0}^{-1}{\delta_2}^{-1}, {\delta_0}^{-1}{\delta_2}^{-1})$.

5.3. Squares and triangles arising from states in $\mathcal{F}_0$ and $\mathcal{F}_1$.

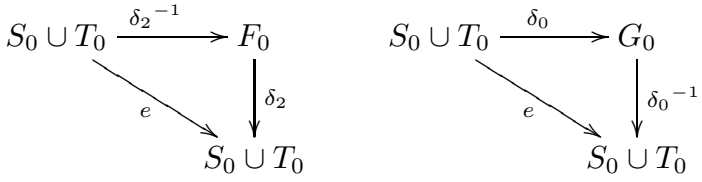
We proceed along the lines of the previous section. What we do now is essentially the same as in section 2.5. Recall that $S_j = F_j \cup G_j$ for $j = 0, 1, 2, 3$ and $T_j = H_j \cup I_j \cup J_j \cup H_{j+1} \cup I_{j+1}$ for $j = 0, 2$. We begin with the word diffences δ_1 and δ_1^{-1}. We have:



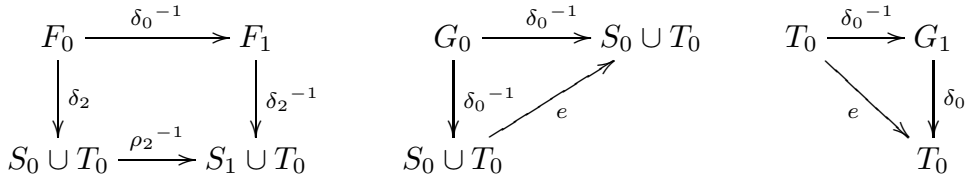
We now consider δ_2:



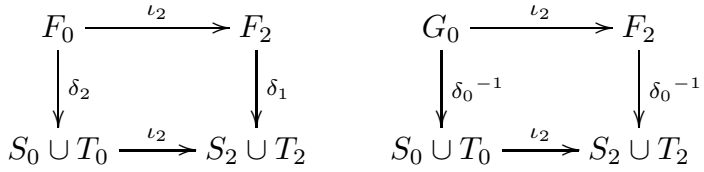
(Recall that $\rho_2 = \delta_2\delta_0\delta_2$.) We now consider ${\delta_2}^{-1}$ and δ_0:



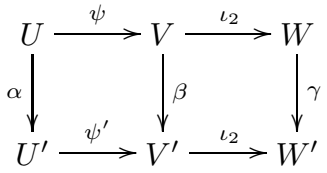
We now consider ${\delta_0}^{-1}$:



We now consider ι_2:



As for word differences between states in $\mathcal{F}_2$, if we have a square for the word difference ψ of the form $(*)$ then we add a new square with the word difference $\iota_2\psi$:



where $W = \iota_2(V)$, $W' = \iota_2(V')$ and $\gamma = \iota_2\beta\iota_2$ is $f|_W$. Now if ψ' were $\iota_2\psi''$ then we have $\iota_2\psi' = {\iota_2}^2\psi'' = \psi''$ since ${\iota_2}^2$ is the identity.

This has completed the construction of squares and triangles for the standard word differences. There remain a few cases that we have not dealt with. These arise is the bottom line in some of the diagrams we have found. The relevant word differences are

$$\rho_2 \colon S_0 \cup T_0 \longrightarrow S_1 \cup T_0, \tag{1}$$

$${\rho_2}^{-1} \colon S_0 \cup T_0 \longrightarrow S_1 \cup T_0, \tag{2}$$

$${\delta_0}^{-1}{\delta_1}^{-1} \colon S_2 \cup S_3 \cup T_2 \longrightarrow S_2 \cup S_3 \cup T_2, \tag{3}$$

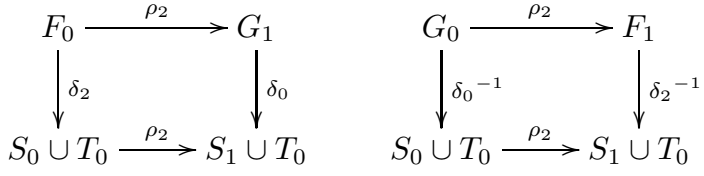
$$\delta_2\delta_0 \colon S_0 \cup S_1 \cup T_0 \longrightarrow S_0 \cup S_1 \cup T_0 \tag{4}$$

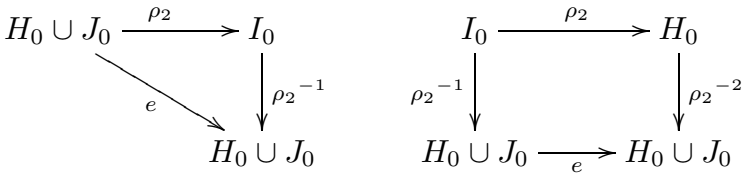
where, as in section 4.6, $\rho_j = \delta_j\delta_0\delta_j$ for $j = 1, 2$. We consider them separately.

The word difference (1) splits as

$$\rho_2 \colon F_0 \longrightarrow G_1, \quad \rho_2 \colon G_0 \longrightarrow F_1, \quad \rho_2 \colon H_0 \cup J_0 \longrightarrow I_0, \quad \rho_2 \colon I_0 \longrightarrow H_0,$$

as well as ρ_2 acting on H_1 and I_1. Applying the symmetry ι_1 this is equivalent to ${\rho_2}^{-1}$ acting on H_0 and I_0 which we consider under (2). Thus we get diagrams

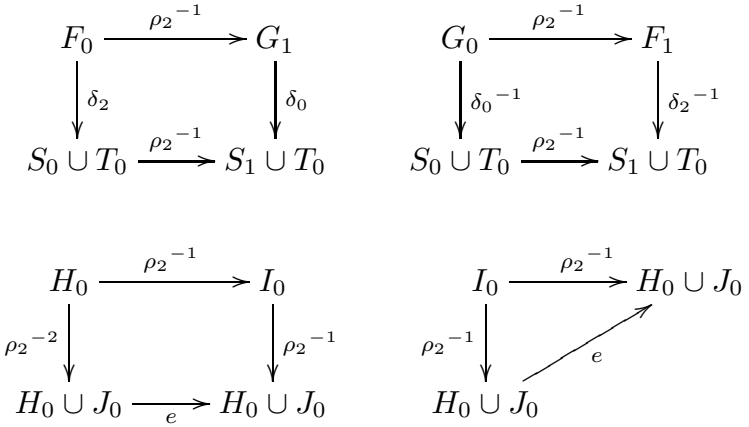




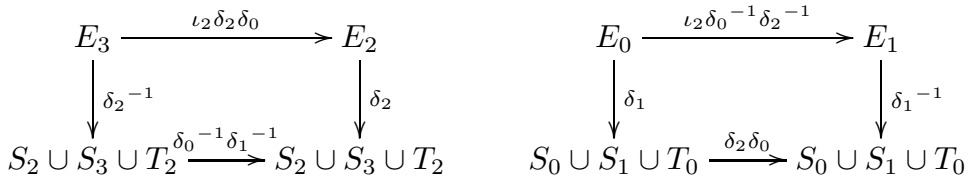
Similarly for (2). We get

$$ {\rho_2}^{-1}\colon F_0 \longrightarrow G_1, \quad {\rho_2}^{-1}\colon G_0 \longrightarrow F_1, \quad {\rho_2}^{-1}\colon H_0 \longrightarrow I_0, \quad {\rho_2}^{-1}\colon I_0 \longrightarrow H_0 \cup J_0. $$

In addition, there is ${\rho_2}^{-1}$ acting on H_1 and I_1. Applying the symmetry ι_1, we can obtain these word differences from ρ_2 acting on H_0 and I_0. This was done above. The remaining word differences give the following diagrams



We now consider the word differences (3) and (4). The only way that these word differences can occur is from one of the following diagrams



Applying the symmetries we see that it is sufficient to consider the following word differences

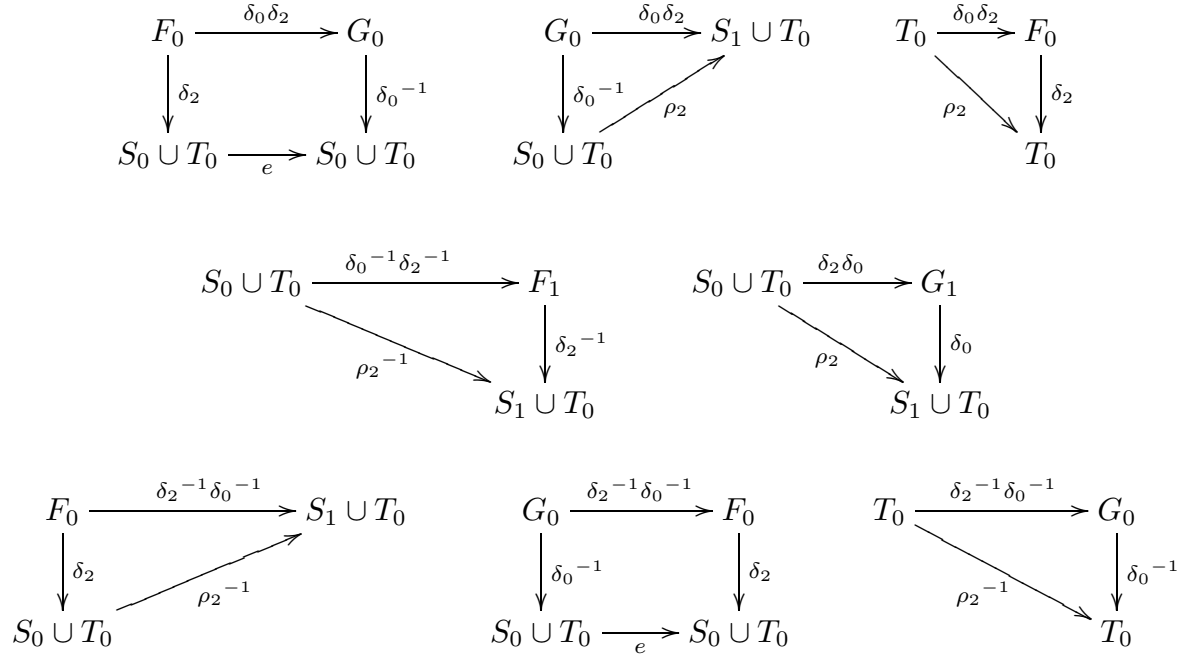
$$ \delta_0\delta_2\colon S_0 \cup T_0 \longrightarrow S_0 \cup S_1 \cup T_0, \tag{5} $$

$$ {\delta_0}^{-1}{\delta_2}^{-1}\colon S_0 \cup T_0 \longrightarrow F_1, \tag{6} $$

$$ \delta_2\delta_0\colon S_0 \cup T_0 \longrightarrow G_1, \tag{7} $$

$$ {\delta_2}^{-1}{\delta_0}^{-1}\colon S_0 \cup T_0 \longrightarrow S_0 \cup S_1 \cup T_0. \tag{8} $$

These break down into diagrams as follows



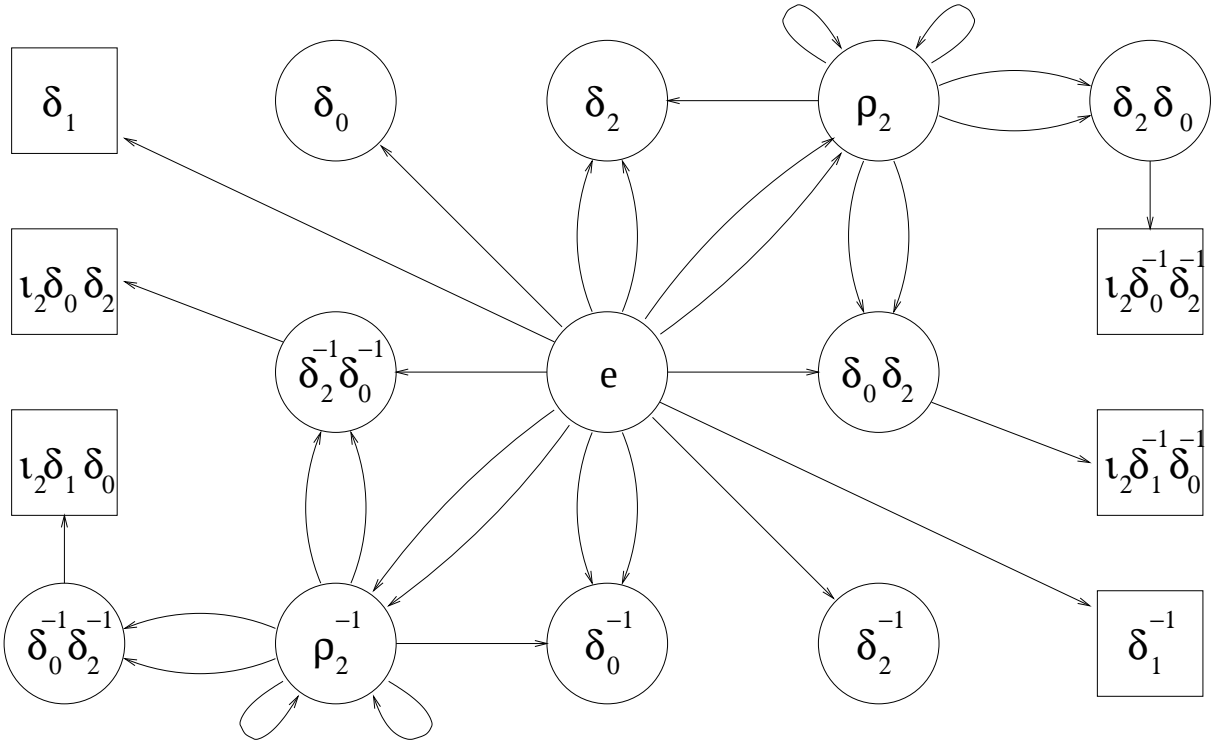


Fig. 5.3.1. Arrows in the word difference machine where U is one of F_0, G_0, H_0, I_0, J_0. In addition there should be word differences $\iota_2\psi$. The square boxes denote word differences where V is E_j.

We remark that all these squares have introduced no new word differences in their bottom lines.

5.4. The exceptional word differences.

We have now completed the word difference machine except that we have not dealt with word differences arising from the following square (which was called (3) in section 5.2):

$$\begin{array}{ccc} D_0 & \xrightarrow{\delta_1} & D_1 \\ {\scriptstyle \delta_0{}^{-1}}\downarrow & & \downarrow{\scriptstyle \delta_0} \\ R_0 & \xrightarrow{\rho_1} & R_1 \end{array}$$

and its symmetric images. Recall that $R_j = A_j \cup B_j \cup D_j \cup E_j$.

Observe that $\delta_1 \colon D_0 \longrightarrow D_1$ and its symmetric images do not arise in the bottom row of any of the squares or triangles we have constructed so far. We call them **exceptional initial states**. There are some states which only arise in paths beginning with an exceptional initial states. We call these **exceptional states**.

Because they never occur in a bottom row, once we have gone from an exceptional state to a non-exceptional state (that is any of the states considered in sections 5.2 and 5.3) we can never return to an exceptional state. In order to get from an exceptional to a non-exceptional state it is usually necessary to pass through a triangle. Thus a triangle of this special type can only occur once in any path through the word difference machine. The exceptional states constitute a separate subgraph of the word difference machine.

We now construct the exceptional states arising from the map: $\rho_1 \colon R_0 \longrightarrow R_1$. Intersecting this with the states of $\mathcal{F}_2$ gives:

$$\rho_1 : {\rho_1}^{-1}(D_1) \cap A_0 \longrightarrow D_1, \tag{1}$$

$$\rho_1 : {\rho_1}^{-1}(B_1) \cap A_0 \longrightarrow B_1, \tag{2}$$

$$\rho_1 : D_0 \longrightarrow \rho_1(D_0) \cap A_1, \tag{3}$$

$$\rho_1 : B_0 \longrightarrow \rho_1(B_0) \cap A_1, \tag{4}$$

$$\rho_1 : E_0 \longrightarrow E_1. \tag{5}$$

Each of these five maps is the top line in a square or triangle of the form $(*)$, $(**)$ or $(***)$. We claim that ${\rho_1}^{-1}(D_1)$ and ${\rho_1}^{-1}(B_1)$ are subsets of A_0 and that $\rho_1(D_0)$ and $\rho_1(B_0)$ are subsets of A_1. This may be checked using Propositions 3.3.1, and 3.3.3. The maps (1) and (3) above give rise to squares of the form $(*)$ for which the new word difference ψ' is again

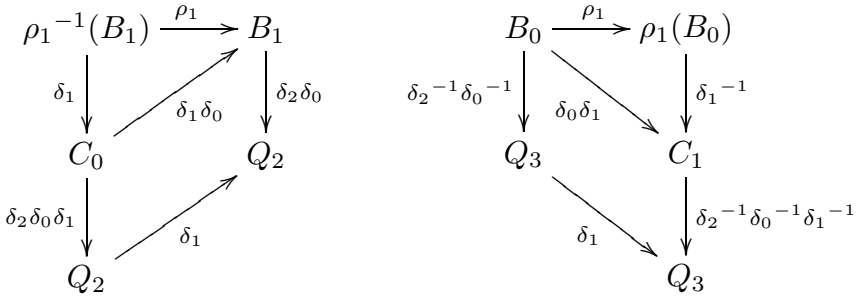
ρ_1. They are

$$\begin{array}{ccc} {\rho_1}^{-1}(D_1) & \xrightarrow{\rho_1} & D_1 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle \delta_0} \\ R_0 & \xrightarrow{\rho_1} & R_1 \end{array} \qquad \begin{array}{ccc} D_0 & \xrightarrow{\rho_1} & \rho_1(D_0) \\ \downarrow{\scriptstyle {\delta_0}^{-1}} & & \downarrow{\scriptstyle {\delta_1}^{-1}} \\ R_0 & \xrightarrow{\rho_1} & R_1 \end{array}$$

The map (5) gives a square which leads directly to a non-exceptional state involving $\mathcal{F}_1$ and $\mathcal{F}_0$. It is

$$\begin{array}{ccc} E_0 & \xrightarrow{\rho_1} & E_1 \\ \downarrow{\scriptstyle \delta_1} & & \downarrow{\scriptstyle {\delta_1}^{-1}} \\ S_0 \cup S_1 \cup T_0 & \xrightarrow{\delta_0} & S_0 \cup S_1 \cup T_0 \end{array}$$

The maps (2) and (4) are immediately followed by a triangle and a square.



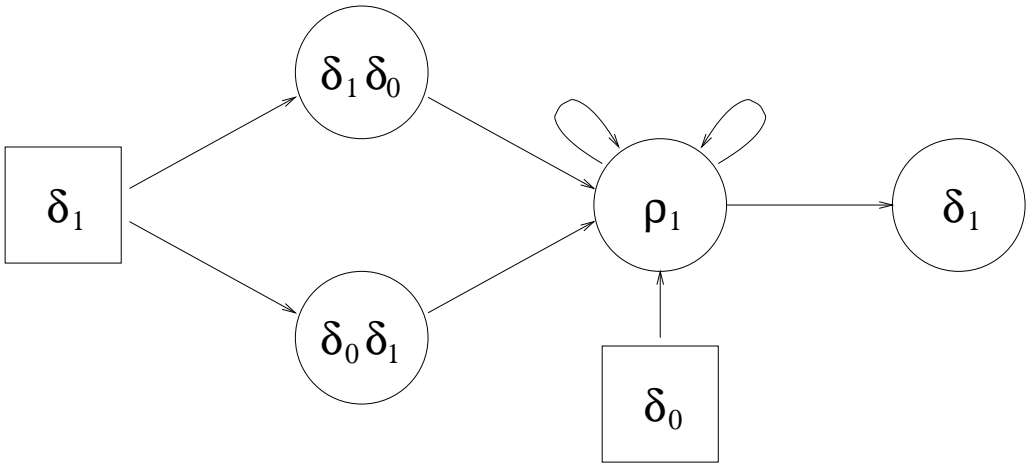


Fig. 5.4.1. Arrows between exceptional word differences ending at the initial difference δ_1. The square boxes denote non-exceptional word differences.

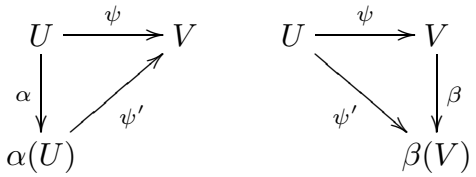
When we form arrows in the word difference machine, we reverse the arrows in each of the squares and triangles listed above. Recall that exceptional initial states never occur

as the bottom line in any of the squares we have constructed. This means that there are no arrows leading out of exceptional initial states. In other words they are dead ends in the word difference machine. Moreover, there are no arrows to non-exceptional states from the subgraph of the word difference machine consisting of exceptional states.

5.5. Synchronising the word difference machine

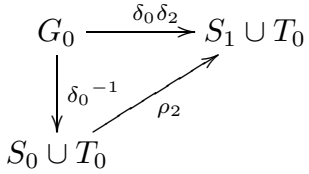
Sections 5.2, 5.3 and 5.4 contain a full list of all word differences. There is one last technical problem because, as in section 2.5, the word difference machine we have constructed using this process is not synchronised. This is because of the presence of triangles rather than squares. We need to check that only finitely many triangles can occur in any path through the difference machine and to then compensate for this by adding padding symbols \$.

In the previous sections we have constructed, up to symmetry, all the squares and triangles that give rise to arrows in the difference machine. As before, we use the following notation for triangles:



There are exactly three ways that triangles can arise. First, there are triangles where ψ' is the identity. Clearly, this can occur at most once in any path through the difference machine. This is equivalent to saying that the only arrows leading to the state e in the difference machine also start at e.

Secondly, there are triangles between non-exceptional states where ψ is $(\delta_j\delta_0)^{\pm 1}$ or $(\delta_0\delta_j)^{\pm 1}$ and ψ' is ${\rho_j}^{\pm 1}$ for $j = 1, 2$. For example



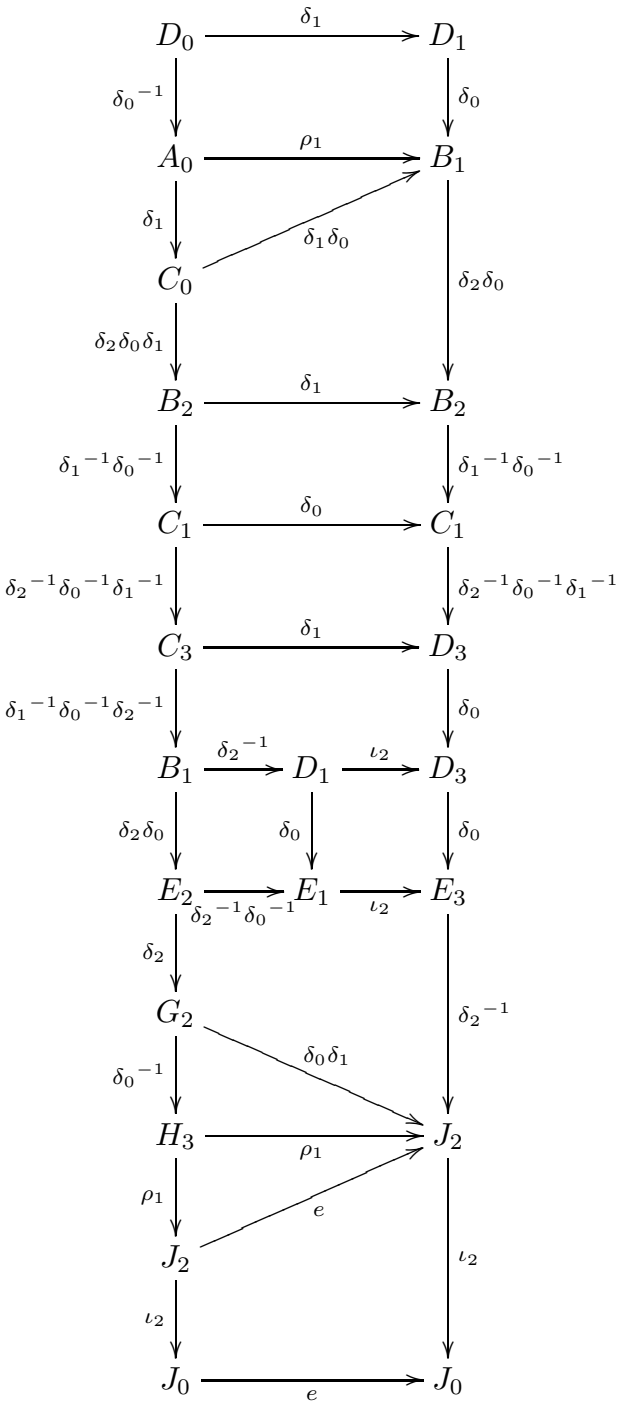
Any subsequent squares have the word difference e or ${\rho_j}^{\pm 1}$. Therefore this type of triangle can occur at most once. This is equivalent to saying that, in Fig. 5.3.1, the only arrows leading to the state ${\rho_2}^{\pm 1}$ begin either at ${\rho_2}^{\pm 1}$ or at e.

Thirdly, there are triangles occurring during transition from exceptional states to non-exceptional states. Once we leave exceptional states we never return and so this can occur at most once. In other words, in Fig. 5.4.1 there are no arrows from exceptional states to non-exceptional states and the triangles occur on the arrows from $\delta_1\delta_0$ and $\delta_0\delta_1$ to ρ_1.

We now give an example which contains all three types of triangles. This is the worst possible case we must deal with. This is the word difference δ_1 for the following Farey block in D_0:

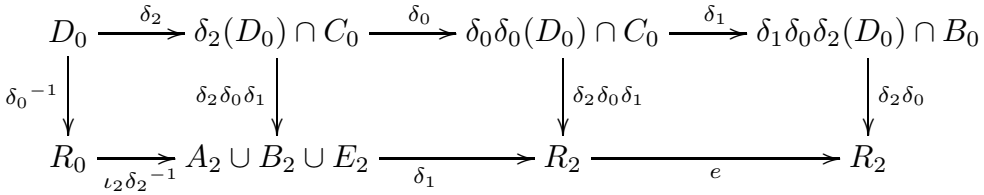
$\gamma_1 = (0,1,2,3) \in \Delta_7,\ \gamma_2 = (1,1,5,8) \in \Delta_7,\ \beta = (3,3,11,15) \in \Delta_7;\ \gamma_3 = (1,1,4,6) \in \Delta_7.$

The squares associated with reducing this Farey block back to J_0 are:

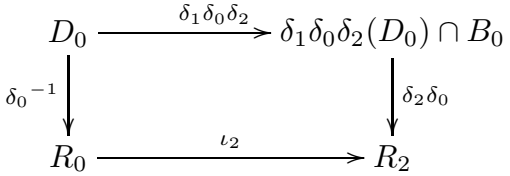


In order to synchronise the word difference machine, as we saw in section 2.5, it is necessary to add to the set of word differences by adding in diagonals to squares. Because we can have more than one triangle these diagonals may carry over several squares. In our example, we need word differences $\delta_2\delta_1\delta_0\delta_1$ between H_3 and E_3; ${\delta_0}^{-1}\delta_2\delta_0\delta_1$ between G_2 and D_3 and so on. Adding all possible diagonals in groups of one, two and three squares adds considerably to the possible word differences.

Finally, we need to make further changes in the collection of word differences. This is because we need to consider initial word differences for all elements of the alphabet $\mathcal{A}$. So far we have only considered initial word differences in $\mathcal{D}_0$ as indicated in section 5.1. Because all words in $\mathcal{A}$ have length at most six in the letters of $\mathcal{D}_0$ this means we need to concatenate up to six word differences. For example, the word difference $\delta_1\delta_0\delta_2$ on the state D_0 gives rise to three squares concatenated horizontally:



In fact, this may be simplified to give the square



All these changes make the final collection of word differences $\mathcal{D}$ rather large but it is still finite. Thus we have a word difference machine in the usual sense. This makes $\mathcal{MCG}(\Sigma_2)$ automatic.

It seems likely that the same structure shows that $\mathcal{MCG}(\Sigma_2)$ is biautomatic (in other words generators can be added at either end) but we have not explored this possibility.

References.

[1] J. S. Birman, *Braids, Links and Mapping Class Groups*, Annals of Maths. Studies **82**, Princeton University Press, 1974.

[2] J. S. Birman & C. Series, Algebraic linearity for an automorphism of a surface group, *Journal of Pure and Applied Algebra* **52** (1988), 227–275.

[3] J. S. Birman & C. Series, Dehn's algorithm revisited, with applications to simple curves on surfaces, *in "Combinatorial Group Theory and Topology"*, ed. S. Gersten & J. Stallings, *Annals of Maths. Studies* **111**, *Princeton University Press* (1987), 451–478.

[4] R. Bowen & C. Series, Markov maps associated to Fuchsian groups, *Inst. Hautes Études Sci. Publ. Math.* **50** (1979), 153–170.

[5] A. Connes, J. Feldman & B. Weiss, An amenable equivalence relation is generated by a single transformation, *Ergodic Theory and Dynamical Systems* **1** (1981), 435–450.

[6] M. Dehn, Die Gruppe der Abbildungklassen, *Acta Math.* **69** (1938), 135–206.

[7] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson & W. P. Thurston, *Word Processing in Groups*, Jones and Bartlett 1992.

[8] H. Hamidi-Tehrani & Z.-H. Chen, Surface diffeomorphisms via train-tracks, *Topology and its Applications* **20** (1996), 1–27.

[9] D. F. Holt, *The Warwick Automatic Groups Software*, DIMACS Series in Mathematics and Computer Science **25**, 1996.

[10] D. F. Holt, Automatic groups, subgroups and cosets, *in "The Epstein Birthday Schrift"*, ed. I. Rivin, C. Rourke & C. Series, *Geometry & Topology Monographs* **1** (1998), 249–260.

[11] L. Keen & C. Series, Pleating coordinates for the Maskit embedding of the Teichmüller space of punctured tori, *Topology* **32** (1993), 719–749.

[12] L. Keen, J. R. Parker & C. Series, Combinatorics of simple closed curves on the twice punctured torus, *Israel J. Maths.* **112** (1999), 29–60.

[13] L. Keen, J. R. Parker & C. Series, Pleating coordinates for the Teichmüller space of the twice punctured torus, (in preparation).

[14] W. Magnus, Über Automorphismen von Fundamentalgruppen berandeter Flächen, *Math. Annalen* **109** (1934), 617–646.

[15] H. A. Masur & Y. N. Minsky, Geometry of the complex of curves I: Hyperbolicity, *Invent. Math.* **138** (1999), 103–149.

[16] H. A. Masur & Y. N. Minsky, Geometry of the complex of curves II: Hierarchical structure, (to appear).

[17] L. Mosher, Mapping class groups are automatic, *Annals of Maths.* **142** (1995), 303–384.

[18] R. C. Penner with J. L. Harer, *Combinatorics of Train Tracks*, Annals of Maths. Studies **125**, Princeton University Press, 1992.

[19] S. Rees, Hairdressing in groups: a survey of combings and formal languages, *in "The Epstein Birthday Schrift"*, ed. I. Rivin, C. Rourke & C. Series, *Geometry & Topology Monographs* **1** (1998), 493–509.

[20] C. Series, The modular surface and continued fractions, *J. London Math. Soc. 2* **31** (1985), 69–85.

[21] C. Series, The geometry of Markoff numbers, *Math. Intelligencer* **7** (1985), 20–29.

[22] C. Series, Geometrical methods of symbolic coding, *in "Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces"*, ed. T. Bedford, M. Keane & C. Series, *Oxford University Press* (1991), 125–151.

[23] W. P. Thurston, *The Geometry and Topology of Three-Manifolds*, Princeton University Press, 1997.

[24] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces, *Bull. American Math. Soc.* **19** (1988), 417–431.