6

Modular group representations throughout the realm

There are two aspects to Moonshine. The more general one is the unexpected presence of modular group actions over a wide range of algebraic settings, and is now fairly well understood. We have seen instances of this already with, for example, the characters of affine algebras and VOAs. This chapter completes our treatment of these modular actions. The more specific aspect – the association of Hauptmoduls to the Monster – is still poorly understood and is the subject of the following chapter.

Much of this chapter is orthogonal to Monstrous Moonshine. For example, we discuss here fusion rings and modular data; both the fusion ring and modular data of the Moonshine module $V^\natural$ are maximally trivial. Nevertheless, this chapter helps to paint the general context of Monstrous Moonshine. In Section 7.2.4 we build on some of the lessons from this chapter to speculate on a possible second proof of Monstrous Moonshine.

6.1 Combinatorial rational conformal field theory

Recall the semi-simple Lie algebras: we study their structure and obtain their classification by abstracting out combinatorial features (e.g. roots, Coxeter–Dynkin diagrams). Of course this is easy to do with a finite-dimensional linear structure. RCFTs are infinite-dimensional, but by definition their infinite-dimensional symmetry and implicit rigidity again effectively reduces them to certain discrete structures. As we see next section, those discrete structures are remarkable for their ubiquity in modern mathematics. See [**208**], [**207**], [**33**], [**131**], [**437**], [**236**] for further background. As with all other chapters except Chapter 7, we've tended to avoid giving original references, as these are voluminous and can be recovered from the numerous review articles and books.

6.1.1 Fusion rings

Recall that the eigenvalues of a self-adjoint (equivalently, Hermitian) matrix are all real. Consider the following scenario. Let A, B and C be $n \times n$ Hermitian matrices with eigenvalues $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n, \beta_1 \geq \cdots \geq \beta_n, \gamma_1 \geq \cdots \geq \gamma_n$. What are the conditions on these eigenvalues so that $C = A + B$? The answer consists of a number of inequalities involving the numbers $\alpha_i, \beta_j, \gamma_k$. Now discretise this problem:

Theorem 6.1.1 *Let* $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0, \beta_1 \geq \cdots \geq \beta_n \geq 0, \gamma_1 \geq \cdots \geq \gamma_n \geq 0$, *all be integers. Then the following are equivalent:*

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

(i) *Hermitian matrices A, B and $C = A + B$ exist with eigenvalues α, β, γ, respectively;*

(b) *the $\mathfrak{gl}_n(\mathbb{C})$ tensor product multiplicity $T^{\gamma}_{\alpha\beta}$ is nonzero.*

Recall from Section 1.5.1 that the finite-dimensional unitary irreducible modules of the Lie algebra $\mathfrak{gl}_n(\mathbb{C}) \cong \mathbb{C} \oplus \mathfrak{sl}_n(\mathbb{C})$ are naturally labelled by pairs $(a, \lambda) \in \mathbb{R} \times \mathbb{N}^{n-1}$, where $z \mapsto \mathrm{i}az$ is a representation of the abelian Lie algebra $\mathbb{C}$, and $\lambda = (\lambda_1, \dots, \lambda_{n-1})$ is a highest weight for the simple Lie algebra $\mathfrak{sl}_n$. The eigenvalues α correspond to labels $a = \alpha_n$ and $\lambda_i = \alpha_i - \alpha_{i+1}$. The number $T^{\gamma}_{\alpha\beta}$ is the number of times the $\mathfrak{gl}_n$-module $L(\gamma)$ appears in the tensor product $L(\alpha) \otimes L(\beta)$ of modules. This remarkable theorem and related results are discussed in the review article [**218**].

Now consider instead $n \times n$ unitary matrices with determinant 1. Any such matrix $D \in \mathrm{SU}_n(\mathbb{C})$ can be assigned a unique n-tuple $\delta = (\delta_1, \dots, \delta_n)$ as follows. Write its eigenvalues as $e^{2\pi \mathrm{i}\delta_i}$, where $\delta_1 \geq \cdots \geq \delta_n$, $\sum_{i=1}^n \delta_i = 0$ and $\delta_1 - \delta_n \leq 1$. Let Δ_n be the set of all such n-tuples δ, as D runs through $\mathrm{SU}_n(\mathbb{C})$. Note that D will have finite order iff all $\delta_i \in \mathbb{Q}$, and that D will be a scalar matrix dI iff all differences $\delta_i - \delta_j \in \mathbb{Z}$. Of course, a sum of Hermitian matrices corresponds here to a product of unitary matrices.

Theorem 6.1.2 [**4**] *Choose any rational n-tuples $\alpha, \beta, \gamma \in \Delta_n \cap \mathbb{Q}^n$. Then the following are equivalent:*

(i) *there exist matrices $A, B, C \in \mathrm{SU}_n(\mathbb{C})$, with $C = AB$, with n-tuples α, β, γ;*

(ii) *there is a positive integer k such that all differences $k\alpha_i - k\alpha_j$, $k\beta_i - k\beta_j$, $k\gamma_i - k\gamma_j$ are integers, and the fusion multiplicity $\mathcal{N}^{(k)\,k\gamma}_{k\alpha,k\beta}$ of ${\mathfrak{sl}_n}^{(1)}$ at level k is nonzero.*

We met the affine algebra ${\mathfrak{sl}_n}^{(1)} = {A_{n-1}}^{(1)}$ and its modules in Section 3.2. Here, $k\alpha$ corresponds to the level-k integrable highest weight $\lambda \in P^k_+(A^{(1)}_{n-1})$ with Dynkin labels $\lambda_i = k\alpha_i - k\alpha_{i+1}$. The ${\mathfrak{sl}_n}^{(1)}$ fusion multiplicities are studied in Section 6.2.1. Theorems 6.1.1 and 6.1.2 provide one instance of a general principle:

> *A result or construction valid for $\mathfrak{gl}_n$ or $\mathfrak{sl}_n$ tensor products should have an interesting analogue for the ${\mathfrak{sl}_n}^{(1)}$ fusion product.*

The $\mathfrak{gl}_n$ tensor product multiplicities are classical quantities, appearing in numerous and varied contexts. The ${\mathfrak{sl}_n}^{(1)}$ fusion multiplicities are equally fundamental, equally ubiquitous, but less well understood.

Just as the tensor product multiplicities are structure constants of the character ring of the Lie algebra, so do fusion multiplicities define a *fusion ring*, an aspect of Moonshine complementary to Monstrous Moonshine.

Definition 6.1.3 *A fusion ring $R = R(\beta, \mathcal{N})$ is a commutative ring R with unity 1, together with a finite basis $\beta = \{x_a \,|\, a \in \Phi\}$ (over $\mathbb{Z}$) containing $1 = X_0$, such that:*

F1. *The structure constants $\mathcal{N}^c_{ab}$, defined by $x_a x_b = \sum_{c\in\Phi} \mathcal{N}^c_{ab} x_c$, are all nonnegative integers.*

F2. *There is a ring homomorphism $x \mapsto x^*$ stabilising the basis Φ (we write $(x_a)^* = x_{a^*}$).*

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

F3. $\mathcal{N}_{ab}^{1} = \delta_{b,a^*}$.
F4. *'$S = S^t$' (we'll explain this shortly, but it says R is self-dual in a strong sense).*
The numbers $\mathcal{N}_{ab}^c$ are called fusion multiplicities*, the labels $a \in \Phi$ are called* primaries*, $0 \in \Phi$ is called the* vacuum *and '$*$' is called* charge-conjugation.

The only reason for distinguishing the basis β from the labels Φ is that for fusion rings the multiplicative notation (e.g. unit 1) is natural, but in the traditional examples of modular data additive notation is used. The terminology here comes from RCFT.

An important ingredient of fusion rings, as with character rings, is their preferred basis β. Abstract rings don't come with a basis. Forgetting the basis β, fusion rings aren't interesting: for example, the algebra $R \otimes_{\mathbb{Z}} \mathbb{C}$ over $\mathbb{C}$ (i.e. the span over $\mathbb{C}$ of β, retaining the same multiplication and addition) is isomorphic as a $\mathbb{C}$-algebra to $\mathbb{C}^{\|\Phi\|}$ with operations defined component-wise (see Lemma 6.1.4 below). This is reminiscent of the character ring of the Lie algebra X_r, which is isomorphic (as a $\mathbb{C}$-algebra) to a polynomial algebra in r variables.

For each $a \in \Phi$, define the *fusion matrix* $\mathcal{N}_a$ by

$$(\mathcal{N}_a)_{b,c} = \mathcal{N}_{ab}^c.$$

Note that the fusion matrix $\mathcal{N}_0$ equals the identity matrix I, and $\mathcal{N}_{a^*} = (\mathcal{N}_a)^t$ (Question 6.1.1). The fusion matrices can be simultaneously diagonalised:

Lemma 6.1.4 *(a) Given any fusion ring $R = R(\Phi, \mathcal{N})$, there is a unique (up to ordering of the columns) unitary matrix S, with rows parametrised by Φ and columns by say Φ', obeying both*

$$S_{0i} > 0, \tag{6.1.1a}$$

$$\mathcal{N}_{ab}^c = \sum_i \frac{S_{ai}\, S_{bi}\, \overline{S_{ci}}}{S_{0i}}, \tag{6.1.1b}$$

for all $a, b, c \in \Phi$ and $i \in \Phi'$.
(b) All simultaneous eigenspaces of all the fusion matrices are of dimension 1*, and are spanned by each column $S_{\updownarrow,b}$.*

The proof of Lemma 6.1.4 only involves F1–F3. The condition F4 can now be expressed by requiring that the S of Lemma 6.1.4 (for some ordering of the columns) obey $S = S^t$ (so $\Phi' = \Phi$). The proof of Lemma 6.1.4 is elementary – the fusion matrices commute with each other and hence with their transposes, and so are simultaneously diagonalisable – and analogues hold in much greater generality. Equation (6.1.1b) says that the bth column $S_{\updownarrow,b}$ of S is an eigenvector of each $\mathcal{N}_a$, with eigenvalue $\frac{S_{ab}}{S_{0b}}$. From the unitarity of S, we know that $\frac{S_{ab}}{S_{0b}} = \frac{S_{ac}}{S_{0c}}$ can hold for all $a \in \Phi$, only if $b = c$, which gives us part (b).

> *The matrix S acts a lot like the character table of a finite group; a general theorem valid for character tables has a fusion ring analogue.*

Note that *a priori* the rows (parametrising basis vectors) and columns (parametrising eigenvectors) of S in Lemma 6.1.4 play entirely different roles. In a natural sense [**236**],

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

the dual ring to R has structure constants given by replacing S in (6.1.1b) with its transpose S^t. This is what underlies calling F4 a self-duality condition. In contrast, the character ring of a finite group is fusion-like, is diagonalised by the character table, but its dual involves multiplying conjugacy classes and is isomorphic to the character ring only for abelian groups. The appearance of self-duality here may seem somewhat mysterious, but some sort of self-duality is pervasive in the mathematics of this chapter. In particular, Drinfel'd's 'quantum double' construction (Section 6.2.3) generates algebraic structures possessing fusion rings and modular data, by combining a given (inadequate) algebraic structure with its dual in some way. An example is provided by Section 6.2.4, where the true (self-dual) fusion ring of a finite group is built up out of the character ring and its dual.

Fusion rings arise naturally in RCFT (Sections 4.3.2 and 6.1.4). The 'primaries' are the chiral primaries, parametrising the irreducible modules of the chiral algebra $\mathcal{V}$. The fusion multiplicities $\mathcal{N}_{ab}^{c}$ are the dimension of the space of chiral blocks $\mathfrak{B}_{a,b}^{(0,3)c}$ on a sphere with three punctures (two 'incoming' and 1 'outgoing'), where we label those punctures with the primaries a, b, c. Equation (6.1.1b) is called *Verlinde's formula* [**542**], and S has an interpretation in terms of modular transformations of the characters (4.3.9a). A similar formula gives the dimension of any space of chiral blocks:

$$\begin{aligned}\dim \mathfrak{B}_{a_1,\ldots,a_n}^{(g,n+m)\,b_1,\ldots,b_m} &:= \mathcal{N}_{a_1,\ldots,a_n}^{(g,n+m)\,b_1,\ldots,b_m} \\ &= \sum_{c\in\Phi}(S_{0c})^{2(1-g)}\,\frac{S_{a_1c}}{S_{0c}}\cdots\frac{S_{a_nc}}{S_{0c}}\,\frac{\overline{S_{b_1c}}}{S_{0c}}\cdots\frac{\overline{S_{b_mc}}}{S_{0c}}. \qquad (6.1.2)\end{aligned}$$

The depth of Verlinde's formula (6.1.1b), (6.1.2), which is considerable, lies in this modular interpretation given to S. The S matrix is called the *modular matrix* for this reason. Historically [**50**], the fusion ring arose directly by interpreting the chiral OPE symbolically in terms of products of $\mathcal{V}$-families of chiral fields (see e.g. section 7.3 of [**131**]).

Recall Perron–Frobenius theory from Section 2.5.2. The fusion matrices $\mathcal{N}_a$ are non-negative, and it is indeed natural to multiply them:

$$\mathcal{N}_a\mathcal{N}_b = \sum_{c\in\Phi}\mathcal{N}_{ab}^{c}\mathcal{N}_c.$$

So we can expect Perron–Frobenius to tell us something interesting. By (6.1.1a), the Perron–Frobenius eigenvalue of $\mathcal{N}_a$ is $\frac{S_{a0}}{S_{00}}$; hence we obtain the important inequality

$$S_{a0}S_{0b} \geq |S_{ab}|\,S_{00}. \qquad (6.1.3a)$$

Unitarity of S applied to (6.1.3a) forces

$$\min_{a\in\Phi}S_{a0} = S_{00}. \qquad (6.1.3b)$$

The *quantum-dimension* $\mathcal{D}_{(a)}$ of (5.3.12) equals $\frac{S_{a0}}{S_{00}}$, and so is bounded below by 1.

The borderline case of (6.1.3b) are those primaries $a \in \Phi$, called *simple-currents* in RCFT, obeying $S_{a0} = S_{00}$. To any such simple-current $j \in \Phi$, there is a phase

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

$\varphi_j : \Phi \to \mathbb{C}$ and a permutation J of Φ such that $j = J0$ and

$$S_{Ja,b} = \varphi_j(b)\, S_{a,b}, \tag{6.1.4a}$$

$$\mathcal{N}_{j,a}^{b} = \delta_{b,Ja}. \tag{6.1.4b}$$

For example, we see from (4.3.11e) that ϵ is a simple-current for the Ising model, with phases $\varphi_\epsilon(0) = \varphi_\epsilon(\epsilon) = 1$ and $\varphi_\epsilon(\sigma) = -1$.

It is clear what plays the role of the endomorphism '$*$' in the character ring of a finite group: complex conjugation. So take the complex conjugate of (6.1.1b). We find that $\overline{S}$ also simultaneously diagonalises the fusion matrices $\mathcal{N}_a$. Hence from Lemma 6.1.4(b) there is a permutation of Φ, which we denote by C, and some $\alpha_b \in \mathbb{C}$, such that

$$\overline{S_{ab}} = \alpha_b\, S_{a,Cb}.$$

Unitarity of S forces each $|\alpha_b| = 1$. Looking at $a = 0$ and applying (6.1.1a), we see that the α_b must be positive. Hence

$$\overline{S_{ab}} = S_{Ca,b} = S_{a,Cb}, \tag{6.1.5}$$

so as a permutation matrix, $C = S^2$. Comparing F3 to Verlinde's formula (6.1.1b), we find that C is charge-conjugation: $Ca = a^*$. Note that C, like complex conjugation, is an involution, and that $C_{00} = 1$.

More generally, recall our discussion of cyclotomic fields and their Galois automorphisms from Section 1.7. The character values $\mathrm{ch}(g)$ of a finite group G lie in the cyclotomic field $\mathbb{Q}[\xi]$, for the root of unity $\xi = \xi_{\|G\|}$. Write σ_ℓ for the automorphism of $\mathbb{Q}[\xi]$ defined by $\sigma_\ell(\xi) = \xi^\ell$, for some integer ℓ coprime to $\|G\|$. Then σ_ℓ acts on the character table by

$$\sigma_\ell(\mathrm{ch}(g)) = \mathrm{ch}(g^\ell) = \mathrm{ch}^{\sigma_\ell}(g), \tag{6.1.6}$$

for some character $\mathrm{ch}^{\sigma_\ell}$ of G (to see which one, use the fact [**308**] that every G-representation is equivalent to a matrix representation with all entries in $\mathbb{Q}[\xi_{\|G\|}]$).

Theorem 6.1.5 [**114**] *Choose any fusion ring, and let S be the associated modular matrix. The entries S_{ab} of the matrix S lie in some cyclotomic field $\mathbb{Q}[\xi_N]$. Given any Galois automorphism $\sigma \in \mathrm{Gal}(\mathbb{Q}[\xi_N]/\mathbb{Q})$,*

$$\sigma(S_{ab}) = \epsilon_\sigma(a)\, S_{a^\sigma,b} = \epsilon_\sigma(b)\, S_{a,b^\sigma} \tag{6.1.7}$$

for some permutation $b \mapsto b^\sigma$ of Φ, and some signs (parities) $\epsilon_\sigma : \Phi \to \{\pm 1\}$.

This is a fundamental symmetry of fusion rings, or rather their modular matrices. For example, for σ equal to complex conjugation, (6.1.7) reduces to (6.1.5). Equation (6.1.7) is essentially the statement that the fusion multiplicities are rational numbers; the cyclotomicity follows from Theorem 1.7.1 and depends crucially on self-duality F4. Any property of charge-conjugation seems to have an analogue for any of these Galois symmetries, although it is usually more complicated.

What has a fusion ring to do with 'modular stuff'? That is explained next.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

6.1.2 Modular data

Choose any even integer $n > 0$. The matrix

$$S = \left(\frac{1}{\sqrt{n}} e^{-2\pi \mathrm{i}\, mm'/n} \right)_{0 \le m, m' < n} \tag{6.1.8}$$

is the finite Fourier transform. Define the diagonal matrix T by $T_{mm} = \exp(\pi \mathrm{i} m^2/n - \pi \mathrm{i}/12)$. The assignment

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mapsto S, \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \mapsto T \tag{6.1.9}$$

defines an n-dimensional representation ρ of $\mathrm{SL}_2(\mathbb{Z})$, using (2.2.1a). This is the simplest (and least interesting) example of *modular data*. Verlinde's formula (6.1.1b) associates a fusion ring with (6.1.8). Here the labels are $\Phi = \{0, 1, \ldots, n-1\}$ and the fusion ring is the ring of integers $\mathbb{Z}[\xi_n]$ with preferred basis

$$\beta = \left\{ 1, \xi_n, \ldots, \xi_n^{n-1} \right\}.$$

The fusion multiplicities are given by addition mod n.

This $\mathrm{SL}_2(\mathbb{Z})$-representation (6.1.9) is realised by modular functions in the following sense. For each $a \in \{0, 1, \ldots, n-1\}$, define the functions

$$\chi_a(\tau) = \frac{1}{\eta(\tau)} \sum_{k=-\infty}^{\infty} q^{n\,(k+a/n)^2/2},$$

where as always $q = e^{2\pi \mathrm{i}\tau}$ and $\eta(\tau)$ is the Dedekind eta function (2.2.6b). Then (4.3.9) hold. Thus $\vec{\chi} = (\chi_a)^t_{a \in \Phi}$ is a vector-valued modular function with multiplier ρ for $\mathrm{SL}_2(\mathbb{Z})$ (Definition 2.2.2).

Definition 6.1.6 *Let Φ be a finite set of labels, one of which – denote it 0 – is distinguished. Modular data are matrices $S = (S_{ab})_{a,b \in \Phi}$, $T = (T_{ab})_{a,b \in \Phi}$ of complex numbers such that:*

MD1. *S, T are unitary and symmetric, and T is diagonal and of finite order. That is, $T^N = I$ for some N.*

MD2. *$S_{0a} > 0$ for all $a \in \Phi$.*

MD3. *$S^2 = (ST)^3$.*

MD4. *The numbers $\mathcal{N}_{ab}^c$ defined by* (6.1.1*b*) *are nonnegative integers.*

From the presentation (2.2.1a) of the modular group $\mathrm{SL}_2(\mathbb{Z})$, we see that modular data defines a representation of $\mathrm{SL}_2(\mathbb{Z})$, as in (6.1.9). Modular data abstracts out the $\mathrm{SL}_2(\mathbb{Z})$ action arising in unitary RCFT (for non-unitary RCFT, MD2 should be weakened). It is a significant refinement of fusion rings. In particular, most fusion rings are not realised by any modular data (Question 6.1.5), but those that are are always realised by at least three sets of modular data.

We can generalise (6.1.8) using lattices (recall Section 1.2.1). If we write L for the lattice $\sqrt{n}\mathbb{Z}$, then $L^* = \frac{1}{\sqrt{n}}\mathbb{Z}$ is the dual lattice, the labels $\{0, \ldots, n-1\}$ parametrise the cosets L^*/L, and the modular function χ_a is the theta series of the ath coset, normalised

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

by η. More generally, any even lattice L defines modular data in this way. The vacuum '0' will be $[0] = L$. The fusion multiplicities $\mathcal{N}_{[a],[b]}^{[c]}$ equal the Kronecker delta $\delta_{[c],[a+b]}$, so the fusion product is given by addition in the finite group L^*/L. All primaries $[a] \in \Phi$ are simple-currents (6.1.4), corresponding to permutation $J_{[a]}([b]) = [a+b]$ and phase $\varphi_{[a]}([b]) = e^{2\pi i a\cdot b}$. Charge-conjugation (6.1.5) is given by $C[a] = [-a]$. The Galois action (6.1.7) here is also simple: there is a Galois automorphism σ_ℓ for any integer ℓ coprime to the determinant $|L|$; σ_ℓ takes $[a]$ to $[\ell a]$, and all parities $\epsilon_\ell([a])$ equal $+1$. From our point of view, however, this lattice example is a little too trivial.

In RCFT (Section 4.3.2), the labels $a \in \Phi$ are the chiral primaries and '0' is the vacuum state. The matrix T equals (4.3.10). Charge-conjugation C is a symmetry in quantum field theory that interchanges particles with their anti-particles (and so reverses charge, hence the name). The modular data S, T arise through (4.3.9), where χ_a are the one-point functions on a torus. The above lattice example corresponds to the string theory of m free bosons compactified on the torus $\mathbb{R}^m/L$, where $m = \dim L$.

Every property of fusion rings should have an analogue in modular data. For example, the analogue of (6.1.5) is

$$T_{Ca,Cb} = T_{ab}, \tag{6.1.10a}$$

which says that T and $C = S^2 = (ST)^3$ commute. The analogue of (6.1.4) is

$$T_{Ja,Ja}\overline{T_{aa}} = \overline{\varphi_j(a)}\, T_{jj}\, \overline{T_{00}}. \tag{6.1.10b}$$

In all known examples, including all those associated with RCFT [**37**], Galois is intimately connected with the existence of characters χ_a realising the modular data as in (4.3.9), which are modular functions for a congruence subgroup (recall (2.2.4)). In particular, for all these examples, we get the remarkable property:

Definition 6.1.7 (congruence property) *Let S, T be modular data, and let ρ be the associated $\mathrm{SL}_2(\mathbb{Z})$-representation. Let N be the order of the matrix T, so $T^N = I$. Then we say S, T obey the congruence property if the following are all satisfied: ρ is trivial (i.e. with value I) on the congruence subgroup $\Gamma(N)$, and so defines a representation of the finite group $\mathrm{SL}_2(\mathbb{Z}_N)$; we have characters χ_a realising the modular data in the sense of* (4.3.9)*, and those characters are modular functions for $\Gamma(N)$; the entries S_{ab} all lie in the cyclotomic field $\mathbb{Q}[\xi_N]$; and finally, the Galois automorphism σ_ℓ corresponds to the modular transformation $\begin{pmatrix} \ell & 0 \\ 0 & \ell^{-1} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}_N)$, and so we get*

$$\left(\rho\begin{pmatrix} \ell & 0 \\ 0 & \ell^{-1} \end{pmatrix}\right)_{ab} = \epsilon_\ell(a)\,\delta_{b,a^{\sigma_\ell}}, \qquad \forall a, b \in \Phi, \tag{6.1.11a}$$

$$T_{a^{\sigma_\ell},a^{\sigma_\ell}} = (T_{aa})^{\ell^2}, \qquad \forall a \in \Phi. \tag{6.1.11b}$$

The finite group $\mathrm{SL}_2(\mathbb{Z}_N)$ arises as $\mathrm{SL}_2(\mathbb{Z})/\Gamma(N)$. The quantity '$\ell^{-1}$' denotes the multiplicative inverse of ℓ (mod N), and exists because $\gcd(\ell, N) = 1$. We return to the congruence property in Section 6.3.3. Probably Definition 6.1.6 is so weak that some 'sick' S, T are examples. It is expected, however, that all reasonably healthy modular

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

data, for example, modular data associated with nice CFTs, VOAs or modular categories, would obey the congruence property (or at least something close to it). It is known [**169**] that modular data obeying the congruence property will typically (always?) be realised by some vector-valued modular function as in (4.3.9).

6.1.3 Modular invariants

Modular data axiomatises the appearance of $\mathrm{SL}_2(\mathbb{Z})$ in unitary RCFT. Two places modular data directly impacts on RCFT are Verlinde's formula (6.1.2) and the partition function (4.3.8b).

Definition 6.1.8 *Choose any modular data S, T. A modular invariant is a matrix $\mathcal{Z}$, with rows and columns labelled by Φ, obeying:*

MI1. $\mathcal{Z}S = S\mathcal{Z}$ *and* $\mathcal{Z}T = T\mathcal{Z}$*;*

MI2. $\mathcal{Z}_{ab} \in \mathbb{N}$ *for all* $a, b \in \Phi$*; and*

MI3. $\mathcal{Z}_{00} = 1$.

It will be convenient at times to rewrite $\mathcal{Z}S = S\mathcal{Z}$ as $S\mathcal{Z}\overline{S} = \mathcal{Z}$ (recall that S is unitary). The easiest modular invariants are the identity $\mathcal{Z} = I$ and charge-conjugation $\mathcal{Z} = C$. More generally, $\mathcal{Z}$ is a modular invariant iff $C\mathcal{Z}$ is.

Modular invariants axiomatise the 1-loop partition functions $\mathcal{Z}(\tau)$ (4.3.8b) of RCFT. More precisely, an RCFT consists of two VOAs, called chiral algebras. For convenience we will take them to be isomorphic, though this is not necessary (when they aren't isomorphic, the theory is called 'heterotic'). The modular invariant describes how these VOAs act on the state space $\mathcal{H}$, that is how $\mathcal{H}$ decomposes into modules of the chiral algebras:

$$\mathcal{H} = \oplus_{a,b\in\Phi}\mathcal{Z}_{ab}\mathcal{H}_a \otimes \overline{\mathcal{H}_b}.$$

MI2 holds because the $\mathcal{Z}_{ab}$ are multiplicities. The adjoint module $\mathcal{H}_0 \otimes \overline{\mathcal{H}_0}$ contains the vacuum $\mathbf{1} \otimes \overline{\mathbf{1}}$, and MI3 says there should be only one vacuum. Finally, the 1-loop partition function $\mathcal{Z}(\tau)$, being a physical correlation function defined on the torus, must be invariant with respect to the modular group $\mathrm{SL}_2(\mathbb{Z})$ of the torus. Equivalently, $\mathcal{Z}(\tau) = \mathcal{Z}(-1/\tau) = \mathcal{Z}(\tau + 1)$. Applying (4.3.9) and the unitarity of S and T gives the modular invariance condition MI1.

Perhaps it is because of their basic importance to RCFT, but the lists of modular invariants associated with affine algebras (Section 6.2.1) are quite remarkable. They also play natural roles for subfactors and VOAs, as we'll see.

A second partition function, playing the same role for boundary CFT (the open string) that $\mathcal{Z}(\tau)$ plays for bulk CFT (the closed string), is that corresponding to a cylinder. Its coefficient matrices $\mathcal{M}_{ax}^{y}$ define a fusion ring representation (6.2.6), called a NIM-rep [**47**], [**236**]. Although they are a fascinating part of the bigger picture, we'll say little about them in this book.

Fix a choice of modular data. Commutation MI1 of $\mathcal{Z}$ with T is trivial to solve, since T is diagonal: it yields the selection rule

$$\mathcal{Z}_{ab} \neq 0 \;\Rightarrow\; T_{aa} = T_{bb}. \tag{6.1.12}$$

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

More subtle and valuable is commutation with S. In particular, each symmetry of S yields a symmetry of $\mathcal{Z}$, a selection rule telling us certain entries of $\mathcal{Z}$ must vanish, and a way to construct new modular invariants.

First consider simple-currents $j,\ j'$. Equation (6.1.4a) and positivity tell us

$$\mathcal{Z}_{j,j'} = \left| \sum_{c,d\in\Phi} \varphi_j(c)\, S_{0c}\, \mathcal{Z}_{cd}\, \overline{S_{d0}\, \varphi_{j'}(d)} \right| \le \sum_{c,d} S_{0c} \mathcal{Z}_{cd} S_{d0} = \mathcal{Z}_{00} = 1.$$

Thus $\mathcal{Z}_{j,j'} \neq 0$ implies $\mathcal{Z}_{j,j'} = 1$, as well as the selection rule

$$\mathcal{Z}_{cd} \neq 0 \ \Rightarrow\ \varphi_j(c) = \varphi_{j'}(d). \tag{6.1.13a}$$

A similar calculation yields the symmetry

$$\mathcal{Z}_{J0,J'0} \neq 0 \ \Rightarrow\ \mathcal{Z}_{Ja,J'b} = \mathcal{Z}_{ab}, \qquad \forall a, b \in \Phi. \tag{6.1.13b}$$

The most useful application of simple-currents to modular invariants is to their construction. In particular, let $j = J_0$ be a simple-current of order n. Then (by Question 6.1.7(b)) we can find integers r_j and $Q_j(a)$ such that

$$\varphi_j(a) = \exp\left[2\pi\mathrm{i}\,\frac{Q_j(a)}{n}\right], \qquad T_{jj}\,\overline{T_{00}} = \exp\left[2\pi\mathrm{i}\, r_j\, \frac{n-1}{2n}\right].$$

Now define the matrix $\mathcal{Z}[j]$ by [**489**]

$$\mathcal{Z}[j]_{ab} = \sum_{\ell=1}^{n} \delta_{J^\ell a,b}\, \delta\left(Q_j(a) + \frac{\ell}{2n} r_j\right), \tag{6.1.14}$$

where $\delta(x) = 1$ when $x \in \mathbb{Z}$ and is 0 otherwise. This matrix will be a modular invariant iff $T_{jj}\overline{T_{00}}$ is an nth root of 1. For instance, $\mathcal{Z}[0] = I$.

Now look at the consequences of Galois. Applying the Galois automorphism σ to $\mathcal{Z} = S\mathcal{Z}\overline{S}$ yields, from (6.1.7) and $\mathcal{Z}_{ab} \in \mathbb{Q}$, the equation

$$\mathcal{Z}_{ab} = \sum_{c,d\in\Phi} \epsilon_\sigma(a)\, S_{\sigma a,c}\, \mathcal{Z}_{cd}\, \overline{S_{d,\sigma b}}\, \epsilon_\sigma(b) = \epsilon_\sigma(a)\, \epsilon_\sigma(b)\, \mathcal{Z}_{\sigma a,\sigma b}.$$

(Why must σ commute with complex conjugation?) Because $\mathcal{Z}_{ab} \ge 0$, this implies the selection rule and symmetry

$$\mathcal{Z}_{ab} \neq 0 \ \Rightarrow \epsilon_\sigma(a) = \epsilon_\sigma(b), \tag{6.1.15a}$$

$$\mathcal{Z}_{\sigma a,\sigma b} \ = \ \mathcal{Z}_{ab}, \tag{6.1.15b}$$

valid for any σ. Of all the equations (6.1.13) and (6.1.15), (6.1.15a) is the most useful. The reader can try to construct modular invariants from certain special σ_ℓ.

6.1.4 The generators and relations of RCFT

In fundamental and influential work of the late 1980s, Moore and Seiberg [**436**], [**437**] isolated the data (finite-dimensional vector spaces and linear transformations) defining each chiral half of RCFT, and provided a complete set of relations they satisfy. Roughly,

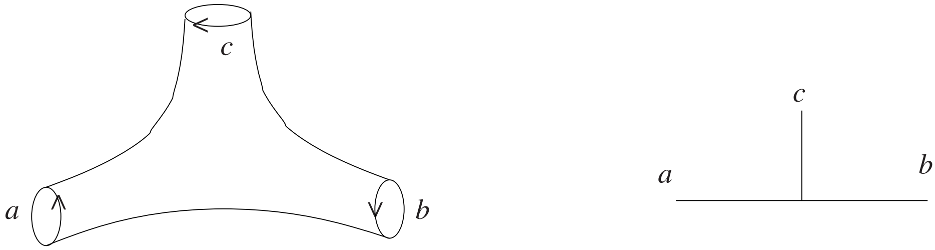

Fig. 6.1 A vertex.

they do for topological field theories in 2 + 1 dimensions what Theorem 4.4.4 does in 1 + 1 dimensions. Most of their work has been rigorously clarified in the important book [**32**]. This section sketches the basic ideas.

Their goal is to understand the spaces $\mathfrak{B}(\Sigma)$ of chiral blocks (Section 4.3.2). As in Section 4.4.1, incoming strings are those boundary circles oriented oppositely to the surface. We can change the orientation of a boundary circle provided we also replace its label (a module $M \in \Phi(\mathcal{V})$) with its charge-conjugate $M^\star$ (5.3.4a). Thus, for instance, the spaces $\mathfrak{B}^{(g,n+m)\;b_1,\dots,b_m}_{a_1,\dots,a_n}$ and $\mathfrak{B}^{(g,n+m)}_{a_1,\dots,a_n,b_1^*,\dots,b_m^*}$ are naturally isomorphic in this way.

We know from the proof of Theorem 4.4.4 that we can build up an arbitrary surface with boundary by sewing together discs, cylinders and pairs-of-pants. Hence the basic building block is the vertex in Figure 6.1. In the spirit of the diagrams of Section 1.6.2, it can be written as the graph on the right. This vertex represents an *intertwining operator* – the $\mathcal{I}_i$ in (4.3.7). They are a natural generalisation of vertex operators (in fact they are often called that), and they generate the chiral blocks $\mathcal{F}$ in exactly the same way that quantum fields generate correlation functions (4.3.1a).

Definition 6.1.9 [**199**], [**436**] *Let $\mathcal{V}$ be a VOA, and let (M^i, Y^i), for labels $i \in \Phi$, be its irreducible modules. For any $a, b, c \in \Phi$, an* intertwining operator of type $\binom{c}{a\,b}$ *is a linear map*

$$w \mapsto \mathcal{Y}(w, z) = \sum_{n\in\mathbb{Q}} w_{(n)} z^{-n-1} \tag{6.1.16}$$

for each $w \in M^a$, where each mode $w_{(n)} \in \mathrm{Hom}(M^b, M^c)$ (hence the name 'intertwiner'), such that for all $w^a \in M^a$, $w^b \in M^b$ and $v \in \mathcal{V}$, $w^a_{(n)}(w^b) = 0$ for all sufficiently large n (depending on both w^a, w^b), and we have both

$$\begin{aligned} z_0^{-1}\delta\left(\frac{z_1 - z_2}{z_0}\right) Y^c(v, z_1)\,\mathcal{Y}(w^a, z_2)w^b - z_0^{-1}\delta\left(\frac{z_2 - z_1}{-z_0}\right) \mathcal{Y}(w^a, z_2)\,Y^b(v, z_1)w^b \\ = z_2^{-1}\delta\left(\frac{z_1 - z_0}{z_2}\right) \mathcal{Y}(Y^b(v, z_0)w^a, z_2)w^b, \end{aligned}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\mathcal{Y}(w^a, z) = \mathcal{Y}(L_{-1}w^a, z).$$

Let $\mathcal{V}\binom{c}{a\,b}$ denote the space of all $\mathcal{Y}$ of the given type.

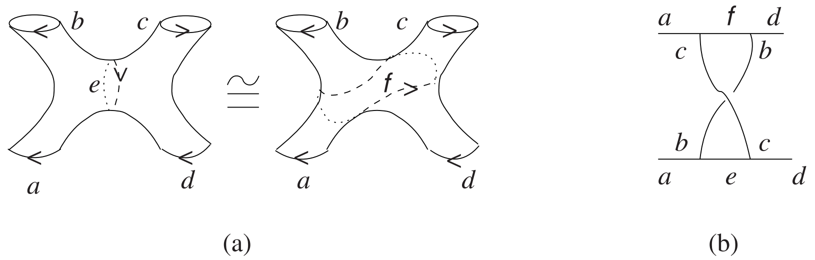

Fig. 6.2 The braiding operator $B_{ef}\begin{bmatrix} b & c \\ a & d \end{bmatrix}$.

In short, the intertwining operator obeys all the properties the vertex operator Y_M obeys in Definition 5.3.1. Of course the z-derivative in the definition completely specifies the z-dependence of an intertwining operator. Note that the defining vertex operator $Y(v, z)$ of a VOA is an intertwining operator of type $\binom{0}{0\,0}$, while the vertex operator Y_{M^a} of the module M^a is of type $\binom{a}{0\,a}$. Summing the formal power series in (6.1.16) over $\mathbb{Q}$ is a little lazy here: the sum really is over $n \in r + \mathbb{Z}$, where $r = \mathrm{wt}\, w^c - \mathrm{wt}\, w^a - \mathrm{wt}\, w^b \in \mathbb{Q}$. The analogue of the grading VA1 here is that $\mathrm{wt}\, w^a_{(n)} = \mathrm{wt}\, w^a - n - 1$.

The dimension of the space of intertwiners is just the fusion multiplicities:

$$\dim\left(\mathcal{V}\binom{c}{a\,b}\right) = \dim \mathfrak{B}\left(\Sigma^{(0,3)\,c}_{ab}\right) = \mathcal{N}^c_{ab} < \infty. \tag{6.1.17}$$

Given a surface Σ with $m + n$ boundary circles, finding a basis for the space $\mathfrak{B}(\Sigma^{b_1,\ldots,b_n}_{a_1,\ldots,a_m})$ is now trivial, at this formal level: simply perform the following Feynman rules.

(i) Fix a basis for each space $\mathcal{V}\binom{c}{a\,b}$ of intertwining operators.
(ii) Fix some dissection of Σ into pairs-of-pants, as in Figure 4.12 (it is more convenient but not necessary to draw the corresponding trivalent graph).
(iii) Assign to each internal cut, or equivalently each internal edge of the trivalent graph, a dummy label.
(iv) To each vertex in your dissection, bounded by labels $a, b, c \in \Phi$ (appropriately oriented), choose an intertwining operator from the basis of the appropriate space of intertwiners.
(v) 'Evaluate' the corresponding chiral block in (4.3.7) – this is a desired basis vector.
(vi) Repeat for each operator in your basis, and each possible value of all dummy labels.

For example, consider the left-most dissection in Figure 6.2(a) of a sphere with four boundary components. Let $\mathcal{Y}$ and $\mathcal{Y}'$ be any intertwining operators in $\mathcal{V}\binom{b}{a\,e}$ and $\mathcal{V}\binom{e}{d\,c}$, respectively. Then we get a chiral block

$$\mathcal{F} = \langle w^b, \mathcal{Y}(w^a, z)\,\mathcal{Y}'(w^d, z')w^c\rangle, \tag{6.1.18}$$

where Möbius invariance was used to send the b- and c-marked points to 0 and ∞. Section 9.3 of [**253**] gives a more physical description of sewing. Incidentally, each dissection corresponds to moving towards a 'maximally degenerate' boundary point on

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

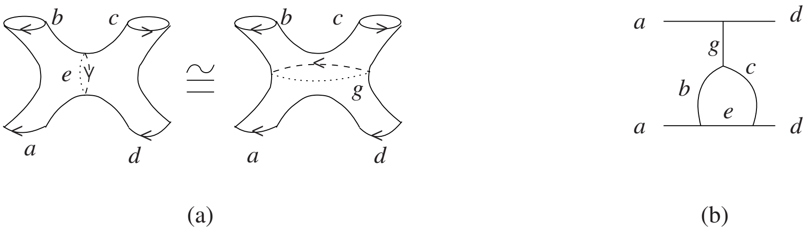

Fig. 6.3 The fusing operator $F_{eg}\begin{bmatrix} b & c \\ a & d \end{bmatrix}$.

$\overline{\mathfrak{M}}_{g,n}$ (recall Section 2.1.4), that is deforming the surface ever more closely to a trivalent graph.

For each dissection, the chiral blocks of (v) are linearly independent and form a basis for the desired space $\mathfrak{B}(\Sigma^{b_1,\ldots,b_n}_{a_1,\ldots,a_m})$. This linear independence implies a product formula for fusion multiplicities, for any pair of dissections of each labelled surface. For instance, the dissections in Figures 6.2(a) and 6.3(a) tell us the nontrivial fact that

$$\mathcal{N}^{(0,4)\,b}_{acd} = \dim \mathfrak{B}\left(\Sigma^{(0,4)\,b}_{acd}\right) = \sum_{e\in\Phi} \mathcal{N}^{b}_{ae}\mathcal{N}^{e}_{cd} = \sum_{f\in\Phi} \mathcal{N}^{f}_{ac}\mathcal{N}^{b}_{df} = \sum_{g\in\Phi} \mathcal{N}^{g}_{ad}\mathcal{N}^{b}_{cg}. \qquad (6.1.19)$$

These identities imply that the fusion ring of an RCFT, defined here formally to have structure constants $\mathcal{N}^{c}_{ab}$, is both commutative and associative. All of these product formulae can be quickly deduced from Verlinde's formula (6.1.2).

As we've repeatedly mentioned, a given surface can be dissected in different ways. *Duality* here is the statement that although each dissection of Σ produces a different basis of chiral blocks, they must be bases for the same space $\mathfrak{B}(\Sigma)$, that is there must be invertible linear maps relating the chiral blocks of different dissections. Consider the easy examples in Figures 6.2 and 6.3. There we've given three dissections of the $(g, n) = (0, 4)$ surface. The corresponding linear maps (actually matrices, given our explicit but noncanonical bases) are denoted $B\begin{bmatrix} b & c \\ a & d \end{bmatrix} = \oplus_{e,f\in\Phi} B_{ef}\begin{bmatrix} b & c \\ a & d \end{bmatrix}$ and $F\begin{bmatrix} b & c \\ a & d \end{bmatrix} = \oplus_{e,g\in\Phi} F_{eg}\begin{bmatrix} b & c \\ a & d \end{bmatrix}$. For the purposes of manipulating identities, it is convenient to represent these operators pictorially as in (b) (recall Section 1.6.2). Because of these pictures, they are usually called *braiding* and *fusing*. They play the same role here as the *Clebsch–Gordon* and *Racah* coefficients (or 3j- and 6j-symbols), respectively, play in the Lie theory of the quantum mechanics literature. See also the treatment in chapter 16 of [**214**].

The proposition at the end of [**278**] gives us four basic 'moves' from which any two dissections can be related. These occur for surfaces with

$$(g, n) = (0, 1), (0, 2), (0, 4), (1, 1) \qquad (6.1.20)$$

(namely, the surfaces that need at most one cut to unfold them into discs, cylinders or pairs-of-pants). The one for (1,1) is given in Figure 6.4. The corresponding operator is called $S(a)$ because it corresponds to the modular transformation $\tau \mapsto -1/\tau$. The result

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

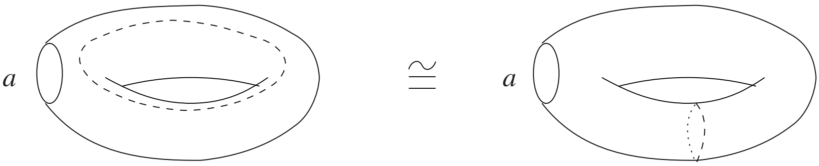

Fig. 6.4 The S-operator $S(a)$.

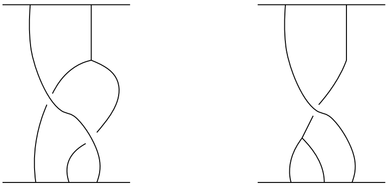

Fig. 6.5 A typical identity.

of [**278**] is the key to proving that a few dualities generate all others. In particular, all duality transformations can be written in terms of F, B, S, $e^{2\pi \mathrm{i}c/24}$.

These duality operators obey several identities, coming from surfaces (0, 5) and (1, 2) (those requiring two cuts to decompose into pairs-of-pants). An example is Figure 6.5; another is the Yang–Baxter equation (Figure 1.29). The reader is encouraged to write these identities down explicitly. Figure 6.5 has the shape $FBB = BF$, while the Yang–Baxter equation looks like $BBB = BBB$. Other identities are given in section 3 of [**437**].

[**436**] argue, and [**32**] prove, that all mapping class group actions on the spaces $\mathfrak{B}(\Sigma)$ can be deduced from these relations. They also argue that Verlinde's formula (6.1.2) follows, by considering the space $\mathfrak{B}^{(1,2)}_{aa^*}$.

For example, consider the Ising model (Section 4.3.2). Here $\Phi = \{1, \epsilon, \sigma\}$. Its modular data S, T is given in (4.3.11), and a basis for the space of chiral blocks in $\mathfrak{B}^{(0,4)\sigma}_{\sigma\sigma\sigma}$ is given in (4.3.13). Its fusion ring is defined by $\epsilon \boxtimes \epsilon = 1$, $\epsilon \boxtimes \sigma = \sigma$ and $\sigma \boxtimes \sigma = 1 \oplus \epsilon$. Recall that these blocks assume that the four points $z_1, \ldots, z_4$ have been mapped to $0, w, 1, \infty$, respectively (so w goes to the cross-ratio). To find the fusing matrix, one way is to note that this duality interchanges the roles of $z_1 = 0$ and $z_3 = 1$, and therefore corresponds to the Möbius transformation $w \mapsto (1 - w)/(1 - 0) = 1 - w$. Likewise, braiding interchanges z_2 with z_3, and so corresponds to the Möbius transformation $w \mapsto (0 - 1)/(0 - w) = 1/w$. When applying Möbius transformations to chiral blocks, recall (4.3.5); equivalently, chiral blocks (of quasi-primaries) are often written as differential forms: here they are $\mathcal{F}_i \, \mathrm{d}w^{-1}$. The braiding and fusing matrices here become

$$B\begin{bmatrix} \sigma & \sigma \\ \sigma & \sigma \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} y & y^{-3} \\ y^{-3} & y \end{pmatrix}, \tag{6.1.21a}$$

$$F\begin{bmatrix} \sigma & \sigma \\ \sigma & \sigma \end{bmatrix} = \frac{y^2 + y^{-2}}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{6.1.21b}$$

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

for some primitive 16th root y of 1. Also, $S(\sigma)$ is 0×0 (since $\mathcal{N}_\sigma^{(1,1)} = 0$ by (6.1.2)), and $S(\epsilon) = (y^{-2})$.

Question 6.1.1. (a) Directly from Definition 6.1.3, prove that the fusion ring homomorphism $*$ in F2 is an involution (i.e. $*^2 = id$).
(b) Again directly from the definition, prove that $\mathcal{N}_{a^*} = (\mathcal{N}_a)^t$ for any fusion matrix $\mathcal{N}_a$.
(c) Directly from the definition, prove that the numbers $\mathcal{N}_{abc} := \mathcal{N}_{ab}^{c^*}$ in any fusion ring are completely symmetric in a, b, c.

Question 6.1.2. Choose your favourite character table theorem in, for example, [**308**] and find and prove the fusion ring analogue.

Question 6.1.3. Prove that a fusion ring $R(\beta, \mathcal{N}) \otimes_{\mathbb{Z}} \mathbb{Q}$, considered as an algebra over $\mathbb{Q}$, is isomorphic to a direct sum of number fields. Construct these number fields explicitly, from the matrix S. (*Hint*: (6.1.7) may be helpful.)

Question 6.1.4. Prove Theorem 6.1.5.

Question 6.1.5. Classify all one- and two-dimensional fusion rings and modular data.

Question 6.1.6. What happens to the modular data of the lattice example when the lattice is integral but not even (i.e. it has odd norm-squared vectors).

Question 6.1.7. (a) Prove (6.1.13b).
(b) Prove that if $j = J_0$ is order n, then $\varphi_j(a)$ is an nth root of unity, and for n odd $T_{Ja,Ja}\overline{T_{aa}}$ is also an nth root of 1, while for n even it is a $2n$th root of 1.
(c) Prove that the set of all simple-currents forms an abelian group (with respect to composition of the permutations J).
(d) Prove that $\mathcal{N}_{Ja,J'b}^{JJ'c} = \mathcal{N}_{ab}^{c}$. Describe σj and $\epsilon_\sigma(j)$ of simple-currents, for any $\sigma \in \mathrm{Gal}(\mathbb{Q}[\xi_N]/\mathbb{Q})$.

Question 6.1.8. Suppose all $a \in \Phi$ are simple-currents. Prove that any modular invariant is of the form (6.1.14).

Question 6.1.9. Suppose we have four sets of functions, namely $a_i(z)$ and $b_i(z)$ (for $1 \le i \le n$), and $c_j(z)$ and $d_j(z)$ (for $1 \le j \le m$), and they are all holomorphic in some common domain (e.g. the unit disc). Suppose the equality

$$\sum_{i=1}^{n} a_i(z)\,\overline{b_i(z)} = \sum_{j=1}^{m} c_j(z)\,\overline{d_j(z)}$$

holds throughout that domain. Then $n = m$ and there is an invertible $n \times n$ matrix M such that both

$$a_i(z) = \sum_{j=1}^{n} M_{ij} c_j(z), \qquad b_i(z) = \sum_{j=1}^{n} \overline{(M^{-1})_{ij}} d_j(z).$$

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

6.2 Examples

6.2.1 Affine algebras

The mathematical riches of CFT go far beyond Lie theory, but CFT would have remained an esoteric part of mathematical physics, unknown to mathematics proper, if its deep connection to Lie theory hadn't been discovered.

The source of some of the most interesting modular data are the nontwisted affine Kac–Moody algebras $\mathfrak{g} = X_r{}^{(1)}$ (Section 3.2). We are interested in its integral highest weights $\lambda \in P_+^k(\mathfrak{g})$ with a given fixed level $k \in \mathbb{N}$.

Recall that the $\mathfrak{g}$-character $\chi_\lambda(\tau)$ (3.2.11c) is essentially a lattice theta function, and transforms nicely under the modular group $\mathrm{SL}_2(\mathbb{Z})$. In fact, the $\mathrm{SL}_2(\mathbb{Z})$-representation ρ of Theorem 3.2.3 defines modular data. The 'vacuum' is $0 = k\omega_0$, and the set of 'primaries' Φ are the highest weights $P_+^k(\mathfrak{g})$ given in (3.2.8). The matrix T is related to the eigenvalues of the second Casimir operator of $\overline{\mathfrak{g}} = X_r$, and S to elements of finite order in the Lie group of X_r [**333**]:

$$T_{\lambda\mu} = \exp\left[\frac{-\pi\mathrm{i}\,(\rho|\rho)}{h^\vee}\right] \exp\left[\frac{\pi\mathrm{i}\,(\overline{\lambda}+\rho|\overline{\lambda}+\rho)}{k+h^\vee}\right] \delta_{\lambda,\mu}, \tag{6.2.1a}$$

$$S_{\mu\nu} = \alpha \sum_{w\in\overline{W}} \det(w) \exp\left[-2\pi\mathrm{i}\,\frac{(w(\overline{\mu}+\rho)|\overline{\nu}+\rho)}{k+h^\vee}\right], \tag{6.2.1b}$$

$$\frac{S_{\lambda\mu}}{S_{0\mu}} = \mathrm{ch}_{L(\overline{\lambda})}\left(\exp\left[-2\pi\mathrm{i}\,\frac{(\overline{\lambda}\,|\,\overline{\mu}+\rho)}{k+h^\vee}\right]\right). \tag{6.2.1c}$$

The unimportant number α is given explicitly in theorem 13.8(a) of [**328**]. The inner-product is the usual Killing form of $\overline{\mathfrak{g}}$, $\overline{W}$ is the (finite) Weyl group of $\overline{\mathfrak{g}}$, ρ is the Weyl vector $\sum_{i=1}^r \omega_i$ and $h^\vee$ is the dual Coxeter number (the sum $\sum_{i=0}^r a_i^\vee$ of the colabels in Figure 3.2). Also, $\overline{\lambda}$ denotes the projection $\lambda_1\omega_1 + \cdots + \lambda_r\omega_r$, and '$\mathrm{ch}_{L(\overline{\lambda})}$' is the appropriate finite-dimensional Lie group character.

The combinatorics of Lie group characters at elements of finite order, that is the ratios (6.2.1c), are quite rich and have been studied by many people. For instance, [**431**] show that they lead to quick algorithms for computing, for example, tensor product multiplicities. Kac [**327**] used them in a Lie theoretic proof of quadratic reciprocity.

For example, for $A_1{}^{(1)}$ at level k, we may take $P_+^k = \{0, 1, \ldots, k\}$ (the value of λ_1), and then the S and T matrices and fusion multiplicities are given by

$$S_{ab} = \sqrt{\frac{2}{k+2}}\,\sin\left(\pi\,\frac{(a+1)\,(b+1)}{k+2}\right), \tag{6.2.2a}$$

$$T_{aa} = \exp\left[\frac{\pi\mathrm{i}(a+1)^2}{2(k+2)} - \frac{\pi\mathrm{i}}{4}\right], \tag{6.2.2b}$$

$$\mathcal{N}_{ab}^c = \begin{cases} 1 & \text{if } c \equiv a+b \ (\mathrm{mod}\ 2) \text{ and } |a-b| \le c \le \min\{a+b, 2k-a-b\} \\ 0 & \text{otherwise} \end{cases}. \tag{6.2.2c}$$

For $A_1{}^{(1)}$ the matrix S is real and so charge-conjugation $C = id$. More generally, for $X_r{}^{(1)}$ C corresponds to a symmetry of the Coxeter–Dynkin diagram of X_r. For $A_1{}^{(1)}$, there

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

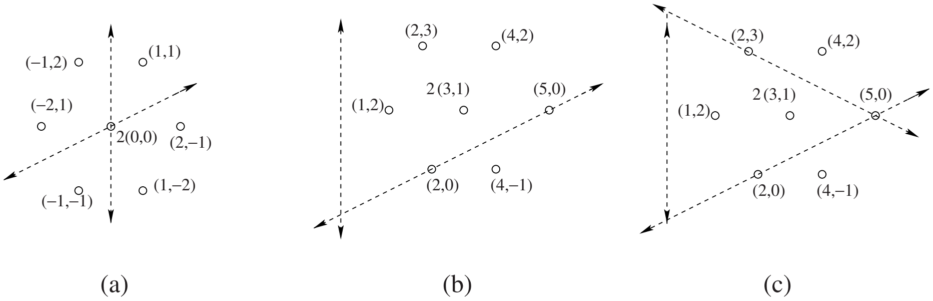

Fig. 6.6 Tensor and fusion products $L(\overline{2,0}) \otimes L(\overline{1,1})$ and $L(0,2,0) \boxtimes {}_2L(0,1,1)$.

is precisely one nontrivial simple-current, namely $j = k$, corresponding to $Ja = k - a$ and $\varphi_j(a) = (-1)^a$. More generally, to any affine algebra (except for $E_8{}^{(1)}$ at $k = 2$), the simple-currents correspond to symmetries of the extended Coxeter–Dynkin diagram. For $A_1{}^{(1)}$ this symmetry interchanges the zeroth and first nodes, that is $J(\lambda_0\omega_0 + \lambda_1\omega_1) = \lambda_1\omega_0 + \lambda_0\omega_1$ (recall $a = \lambda_1$ and $k = \lambda_0 + \lambda_1$).

The fusion multiplicities $\mathcal{N}_{\lambda\mu}^{\nu}$, defined by (6.1.1b), are essentially the tensor product multiplicities $T_{\overline{\lambda}\overline{\mu}}^{\overline{\nu}} := \text{mult}_{\overline{\lambda}\otimes\overline{\mu}}(\overline{\nu})$ for $\overline{\mathfrak{g}}$ (as opposed to the unrelated and less interesting tensor product multiplicities of $\mathfrak{g}$), except 'folded' in a way depending on the level k. This is seen explicitly by the Kac–Walton formula (see [**328**] page 288, [**552**], though there are other co-discoverers):

$$\mathcal{N}_{\lambda\mu}^{\nu} = \sum_{w\in W} \det(w)\, T_{\overline{\lambda}\overline{\mu}}^{w.\overline{\nu}}, \tag{6.2.3a}$$

where $w.\gamma := w(\gamma + \rho) - \rho$ and W is the affine Weyl group of $X_r{}^{(1)}$ (the dependence on k arises through this action of W). The proof follows quickly from (6.2.1c). This practical formula is also described in Section 16.2 of [**131**] and Section 4.9 of [**553**].

Equation (6.2.3a) looks more natural when viewed as follows. The Racah–Speiser formula (there are other co-discoverers) for tensor product multiplicities says

$$T_{\overline{\lambda}\overline{\mu}}^{\overline{\nu}} = \sum_{w\in\overline{W}} \det(w) \dim L(\overline{\mu})_{w.\overline{\nu}-\overline{\lambda}}. \tag{6.2.3b}$$

Combining (6.2.3) gives the 'affinisation' of Racah–Speiser:

$$\mathcal{N}_{\lambda\mu}^{\nu} = \sum_{w\in W} \det(w) \dim L(\overline{\mu})_{w.\overline{\nu}-\overline{\lambda}}. \tag{6.2.3c}$$

For example, the weights for the eight-dimensional A_2-module $L(\overline{1,1})$ are given in Figure 6.6(a). In Figure 6.6(b), we translate this weight space by $\rho + \overline{\lambda} = (3,1)$. Equation (6.2.3b) now tells us to Weyl-reflect each dot not in the A_2 alcove $P_+^k + \rho$. Two of these dots are fixed by a Weyl reflection and so cancel themselves. Weight $(4, -1)$ gets Weyl reflected to $(3,1)$ and so reduces the multiplicity there by 1. Shifting back by $\rho = (1,1)$, we thus get the tensor product

$$L(\overline{2,0}) \otimes L(\overline{1,1}) = L(\overline{0,1}) \oplus L(\overline{2,0}) \oplus L(\overline{1,2}) \oplus L(\overline{3,1}).$$

The calculation of the ${A_2}^{(1)}$ fusion multiplicity at, for example, level 2 (Figure 6.6(c)) is identical, except we now have extra Weyl reflections and the alcove is much smaller. The weight (4, 2) now lies outside the alcove, and reflects to (3, 1) where it reduces that multiplicity to 0. Thus we obtain the fusion product (writing the level as subscript)

$$L(0,2,0) \boxtimes_2 L(0,1,1) = L(1,0,1).$$

Equation (6.2.3a) has the flaw that, although the $\mathcal{N}_{\lambda\mu}^{\nu}$ are manifestly integral, it is not clear why they are positive. An open problem in the theory is the discovery of a combinatorial rule, for example, in the spirit of the well-known Littlewood–Richardson rule [**217**], for the affine algebra fusions. Such a rule for ${A_r}^{(1)}$ is conjectured in [**88**], although it is quite complicated even for ${A_1}^{(1)}$.

Identical numbers $\mathcal{N}_{\lambda\mu}^{\nu}$ appear in several other contexts, many of which we'll see below. Because of these isomorphisms, we know that the $\mathcal{N}_{\lambda\mu}^{\nu}$ defined by (6.1.1b) and (6.2.1b) do indeed lie in $\mathbb{N}$, for any affine algebra, as predicted by RCFT.

As mentioned before, the fusion product here is *not* the usual tensor product of affine algebra modules. However, the fusion product has been interpreted algebraically (with much effort) as a new kind of tensor product of affine algebra modules, in a series of papers by Kazhdan and Lusztig; it was proved equivalent to fusions in [**190**].

Fusion multiplicities arise in the quantum cohomology or Gromov–Witten invariants of Grassmannians [**565**], [**57**], often called the 'quantum Schubert calculus'. Recall that 'points' in the projective plane consist of lines through the origin; more generally, the Grassmannian Gr(m,n) consists of m-dimensional subspaces in $\mathbb{R}^n$. The (classical) Schubert calculus (see e.g. [**217**]) uses the cohomology ring of Gr(m,n) to solve problems in enumerative geometry such as 'How many lines in projective 3-space $\mathbb{P}^3(\mathbb{R})$ meet four given lines?'. On the other hand, the Gromov–Witten invariants count surfaces lying in the Grassmannian, which satisfy certain conditions (see e.g. [**359**]). The quantum cohomology ring (which counts spheres) of Gr(m,n) is isomorphic to the fusion ring of ${\mathfrak{gl}_m}^{(1)} = (\mathfrak{u}_1 \oplus A_{m-1})^{(1)}$ at level $(nm, n-m)$, 'orbifolded' with a 'projection/field-identification' given by the order-m simple-current (J^{-n}, J); the Gromov–Witten invariants are the fusion multiplicities. Now, there is a classical isomorphism Gr$(m,n) \cong$ Gr$(n-m,n)$ (why?); this implies that there is a close relation ('rank–level duality') between the fusion rings of ${A_r}^{(1)}$ level k and ${A_{k-1}}^{(1)}$ level $r+1$. There are analogous rank–level dualities for the other classical algebras [**428**]. This is one of many symmetries of the $\mathfrak{g}$ fusion multiplicities that has no analogue for the $\overline{\mathfrak{g}}$ tensor product multiplicities. Another example is that any symmetry of the extended Coxeter–Dynkin diagram is a symmetry of fusion multiplicities. In short, *affine algebra fusion multiplicities are mathematically more interesting than their classical counterparts.*

We have long known that the representation theory of a Lie group G is related to K-theory. For example, the equivariant K-theory $K_G^{dim\,G}(p)$ of the (trivial) action of G on a point p is the representation ring (over $\mathbb{Z}$). The analogue of this for fusion rings is due to Freed–Hopkins–Teleman [**193**]: the fusion ring of ${X_r}^{(1)}$ at level k is the twisted equivariant K-theory ${}^hK_{\mathcal{L}G}^{dim\,G}(p) := {}^{k+h^\vee}K_G^{dim\,G}(G)$, where G is the compact simply-connected Lie group corresponding to X_r, G acts on itself by conjugation and

$k + h^\vee \in \mathbb{Z} = H^3_G(G, \mathbb{Z})$ is the twist h. The strength of this important formulation is also its weakness: it pushes most technical difficulties under the carpet, but what remains is a clean conceptual characterisation of the fusion ring.

Fusion multiplicities also arise as dimensions of spaces of generalised theta functions [**179**] (see also the discussion in [**565**]), as tensor product multiplicities in Hecke algebras at roots of 1 [**255**] and modular representations for, for example, the Lie algebra $\overline{\mathfrak{g}}$ for fields $\mathbb{F}_p$ (see e.g. [**392**]). In Section 6.1.1 we give another appearance of the $A_{n-1}{}^{(1)}$ fusion multiplicities.

The Galois action for the affine algebras can be expressed geometrically using the action of the affine Weyl group on the weight lattice of X_r. The parity $\epsilon_\sigma(\lambda)$ is quite interesting (see e.g. [**7**] for cohomological and number-theoretic interpretations). For a concrete example, consider $A_1{}^{(1)}$: (6.2.2a) shows explicitly that S_{ab} lies in the cyclotomic field $\mathbb{Q}[\xi_{4(k+2)}]$. Write $\{x\}$ for the number congruent to x mod $2(k+2)$ satisfying $0 \leq \{x\} < 2(k+2)$. Choose any Galois automorphism $\sigma = \sigma_\ell$. Then if $\{\ell(a+1)\} < k+2$, we will have $a^\sigma = \{\ell(a+1)\} - 1$, while if $\{\ell(a+1)\} > k+2$, we will have $a^\sigma = 2(k+2) - \{\ell(a+1)\} - 1$. The parity $\epsilon_\sigma(a)$ depends on a contribution from $\sqrt{\frac{2}{k+2}}$ (which can usually be ignored), as well as the sign $+1$ or -1, respectively, depending on whether or not $\{\ell(a+1)\} < k+2$.

Affine algebra modular data corresponds to Wess–Zumino–Witten RCFT [**245**], where a closed string lives on a Lie group manifold G. The action is given by the sum of two terms: one is an integral over the world-sheet and corresponds to a so-called sigma model [**343**] of a bosonic field living on G; the other is a topological Wess–Zumino term, an integral over the volume bounded by the (compactified) world-sheet. Classically, the sigma model by itself would be conformally invariant, but quantisation breaks this. It was Witten who realised that conformal invariance would be retained if the Wess–Zumino term was added. For topological reasons the Wess–Zumino term comes with an integral prefactor (or coupling constant), which we call the level k.

Why is the level k always shifted by the dual Coxeter number $h^\vee$ in the formulae, and the weights by the Weyl vector ρ? The ρ-shift appears even for the simple finite-dimensional algebras (1.5.11), and arises from the combinatorics of geometric series. The algebraic explanation of the $h^\vee$-shift was given after (3.2.15). Physically, in the Wess–Zumino–Witten model, these ρ- and $h^\vee$-shifts also arise automatically: the former as a quantum effect, due to normal-ordering or regularisation, much like the $q^{1/24}$ shift in the Dedekind eta; the latter as an effect of latent supersymmetry caused by decoupling fermions (see e.g. section 8 of [**248**], or [**206**]).

The modular data (6.2.2) of $A_1{}^{(1)}$ level k is related to the dilogarithm by the remarkable formula

$$\frac{1}{L(1)} \sum_{b=1}^{k} L\left(\frac{S_{0a}^2}{S_{ba}^2}\right) = c_k - 24h_a + 6a \tag{6.2.4a}$$

for each $a \in P_+^k$, where $c_k = 3k/(k+2)$ is the central charge and $h_a = \frac{a(a+2)}{4(k+2)}$ the conformal weight (recall (3.2.9)). $L(x)$ here is *Roger's dilogarithm*, which for $0 < x < 1$ is

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

given by

$$L(x) = \sum_{n=1}^{\infty} \frac{x^2}{n^2} + \frac{1}{2} \log x \log (1 - x). \tag{6.2.4b}$$

We put $L(1) := \lim_{x \to 1^-} L(x) = \pi/6$. $L(x)$ is strictly increasing, real-analytic, and obeys $L(x) + L(1 - x) = L(1)$ and

$$L(x) + L(y) = L(xy) + L\left(\frac{x - xy}{1 - xy}\right) + L\left(\frac{y - xy}{1 - xy}\right). \tag{6.2.4c}$$

As was discovered by Lobachevsky and Schläffli in the nineteenth century, the dilogarithm is related to volumes of tetrahedra, and several other appearances have been uncovered since. Equation (6.2.4a) is the tip of the iceberg; see [**347**] for several other identities and some history. (6.2.4a) can be proved by studying the $\tau \to 0$ asymptotics of certain character formulae. For a simple example, the two $k = 1$ ${A_1}^{(1)}$ characters can be written

$$\chi_{i\omega_1 + (1-i)\omega_0}(\tau) = \sum_{\substack{M+N+i \text{ even} \\ M,N \in \mathbb{N}}} \frac{q^{(M+N)^2/2}}{(q)_M (q)_N}, \tag{6.2.5a}$$

where $(q)_N$ is the q-deformed factorial $\prod_{n=1}^{N}(1 - q^n)$. Similar expressions exist for all other affine algebras and conjecturally all RCFT – see [**14**] for the state-of-the-art, and below for a conjecture. Actually, (6.2.4a) is obtained from the asymptotics of these character identities for certain non-unitary RCFTs, which have essentially the same S matrix as (6.2.2a). An explanation of some of these identities (at least mod 1) has been made by [**164**], who use the dilogarithm to express a natural map from $H_3(\widetilde{\mathrm{SL}}_2(\mathbb{R}), \mathbb{Z})$ to $\mathbb{R}/\mathbb{Z}$.

Choose any $r \times r$ rational positive-definite matrix $A = A^t$, $b \in \mathbb{Q}^r$ and $d \in \mathbb{Q}$. Define

$$f_{A,b,d}(\tau) := \sum_{n \in \mathbb{N}^n} \frac{\exp[2\pi i \tau\, (n^t A n/2 + b^t n + d)]}{(q)_{n_1} \cdots (q)_{n_r}}. \tag{6.2.5b}$$

Conjecture 6.2.1 (Nahm [444]) *Let A be any $n \times n$ rational positive-definite matrix. Then there are finitely many vectors $b_1, \ldots, b_m \in \mathbb{Q}^n$ and numbers $d_1, \ldots, d_m \in \mathbb{Q}$ such that the functions $\chi_i(\tau) := f_{A,b_i,d_i}(\tau)$ are the entries of a vector-valued modular function for $SL_2(\mathbb{Z})$, iff these $\chi_i(\tau)$ are the graded-dimensions of the m primaries of some (not necessarily unitary) RCFT where $d_i = h_i - c/24$, iff there is a corresponding element of finite order in the Bloch group.*

The precise statement involving the Bloch group would take us too far afield, but see [**444**] for details. This beautiful conjecture has been verified only for $r = 1$ (which has three different A). A plausibility argument suggesting that RCFT characters should always be of that form involves considering their massive integrable perturbations [**444**]. Torsion in the Bloch group has known connections with modularity.

The affine algebra $\mathfrak{g}$ arises in the Wess–Zumino–Witten model, for the same reason the Virasoro does (recall the discussion around (4.3.4)): to each $g \in G$ we get a conserved

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

current, and its conserved charges define the level-k representation of $\mathfrak{g}$. As before, we get two commuting actions of $\mathfrak{g}$ on the state-space $\mathcal{H}$, recovering the finite decomposition (4.3.6b).

For affine algebra modular data, the classification of modular invariants seems to be just barely possible, and the answer is that (generically) the only modular invariants are constructed in straightforward ways from symmetries of the Coxeter–Dynkin diagrams. For instance, consider $A_1{}^{(1)}$:

Theorem 6.2.2 [**91**] *Recall that $P_+^k = \{0, 1, \dots, k\}$, and the simple-current is given by $Ja = k - a$. Then the complete list of $A_1{}^{(1)}$ modular invariants is*

$$\begin{aligned}
\mathcal{A}_{k+1} &= \sum_{a=0}^{k} |\chi_a|^2 && \text{for all } k \geq 1,\\
\mathcal{D}_{\frac{k}{2}+2} &= \sum_{a=0}^{k} \chi_a\, \chi_{J^a a}^* && \text{when } \frac{k}{2} \text{ is odd},\\
\mathcal{D}_{\frac{k}{2}+2} &= |\chi_0 + \chi_{J0}|^2 + |\chi_2 + \chi_{J2}|^2 + \cdots + 2|\chi_{\frac{k}{2}}|^2 && \text{when } \frac{k}{2} \text{ is even},\\
\mathcal{E}_6 &= |\chi_0 + \chi_6|^2 + |\chi_3 + \chi_7|^2 + |\chi_4 + \chi_{10}|^2 && \text{for } k = 10,\\
\mathcal{E}_7 &= |\chi_0 + \chi_{16}|^2 + |\chi_4 + \chi_{12}|^2 + |\chi_6 + \chi_{10}|^2 \\
&\quad + \chi_8\,(\chi_2 + \chi_{14})^* + (\chi_2 + \chi_{14})\,\chi_8^* + |\chi_8|^2 && \text{for } k = 16,\\
\mathcal{E}_8 &= |\chi_0 + \chi_{10} + \chi_{18} + \chi_{28}|^2 + |\chi_6 + \chi_{12} + \chi_{16} + \chi_{22}|^2 && \text{for } k = 28.
\end{aligned}$$

A simple proof is given in [**234**]. The modular invariants $\mathcal{A}_n$ and $\mathcal{D}_n$ are generic, given by (6.1.14), and correspond respectively to the order 1 (i.e. identity) and order 2 (i.e. simple-current J) Coxeter–Dynkin diagram symmetries. Physically, $\mathcal{A}_n$ and $\mathcal{D}_n$ are the partition functions (4.3.8b) of Wess–Zumino–Witten models on the $\mathrm{SU}_2(\mathbb{C})$ and $\mathrm{SO}_3(\mathbb{R})$ group manifolds, respectively. The exceptionals $\mathcal{E}_6$ and $\mathcal{E}_8$ correspond to strings living on Sp_4 and G_2 manifolds, at level 1. The $\mathcal{E}_7$ exceptional is harder to interpret, but is the first in an infinite series of exceptionals involving rank–level duality and D_4 triality.

Around Christmas 1985, Zuber wrote to Kac about the $A_1{}^{(1)}$ modular invariant problem, and mentioned the modular invariants they knew at that point (what we now call $\mathcal{A}_\star$ and $\mathcal{D}_{even}$). A few weeks later, Kac wrote back saying he found one more invariant, and jokingly pointed out that it must indeed be quite exceptional as the exponents of E_6 appeared in it. By summer 1986, Cappelli–Itzykson–Zuber found $\mathcal{E}_7$, $\mathcal{D}_{odd}$ and then $\mathcal{E}_8$, and at some point recalled by chance Kac's cryptic remark. They rushed to the library to find a list of the exponents of the other algebras, and were delighted to discover that they all matched. Thus the *A*–*D*–*E* pattern (Section 2.5.2) to their modular invariants was discovered!

The modular invariants for $A_1{}^{(1)}$ realise the *A*–*D*–*E* pattern, in the following sense [**91**]. The (dual) Coxeter number $h = h^\vee$ of the name $\mathcal{X}_n$ equals $k + 2$, and the exponents m_i of $\mathcal{X}_n$ equal 1 plus those $a \in P_+^k$ for which $\mathcal{Z}_{aa} \neq 0$ (for the algebras A_n, D_n, E_n, the integers m_i are defined by writing the eigenvalues of the corresponding Cartan matrix

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

(Definition 1.4.5) as $4\sin^2(\frac{\pi m_i}{2h})$). Probably what first led Kac to his observation about the E_6 exponents was that $k+2$ (this is how k enters most formulae), for his exceptional, equals the Coxeter number 12 for E_6. More recently, deeper connections between *A–D–E* and the ${A_1}^{(1)}$ modular invariants have been found, notably in subfactor theory (Section 6.2.6). This modular invariant classification, however, has never been directly reduced to the suggestion of Section 2.5.2.

The modular invariants have also been classified, for example, for ${A_2}^{(1)}$ [**232**], and they too seem quite interesting (Section 6.3.2). We are almost at the point where we can safely conjecture the complete list of modular invariants for ${X_r}^{(1)}$ at any k, for X_r a simple algebra (see e.g. [**236**]). The most surprising thing about these affine algebra modular invariant classifications is that there are so few surprises: almost every modular invariant is 'generic', that is constructable using a few simple uniform methods such as Coxeter–Dynkin diagram symmetries. Unfortunately, the classification for semi-simple algebras $X_{r_1} \oplus \cdots \oplus X_{r_s}$ does not reduce to that for simple ones, and will be hopeless.

Has *A–D–E* been discovered in the other modular invariant classifications? No, only in those classifications trivially reducible to Theorem 6.2.2. There is, however, a rather natural way to assign (multi-di)graphs to modular invariants, generalising the *A–D–E* pattern for ${A_1}^{(1)}$. It is called a NIM-rep, and is a *rep*resentation of the fusion ring by *n*onnegative *i*nteger *m*atrices. More precisely, for each weight $a \in P_+^k(A_1^{(1)})$ we want a nonnegative integer matrix $\mathcal{M}_a$ such that

$$\mathcal{M}_a\mathcal{M}_b = \sum_{c=0}^{k} \mathcal{N}_{ab}^c \mathcal{M}_c, \tag{6.2.6}$$

where $\mathcal{N}_{ab}^c$ are the fusion multiplicities of (6.2.2c). We also require $\mathcal{M}_0 = I$, and all these matrices to be symmetric: $\mathcal{M}_a = (\mathcal{M}_a)^t$. In Question 6.2.2 you are asked to find all such assignments $a \mapsto \mathcal{M}_a$. Surprisingly, there is a near-perfect correspondence between the ${A_1}^{(1)}$ modular invariants, and these NIM-reps. Physically, NIM-reps are associated with boundary conformal field theory or D-branes in string theory. See [**47**], [**236**] and references therein for the basic theory and examples of NIM-reps. They are an integral part of the combinatorial data of RCFTs. However, the simplicity of the correspondence for ${A_1}^{(1)}$ is an accident due to the small size of the relevant Perron–Frobenius eigenvalue here. In particular there appear to be far more NIM-reps for ${A_2}^{(1)}$ than modular invariants.

Hanany–He [**271**] suggest that the ${A_1}^{(1)}$ *A–D–E* pattern can be related to subgroups $G \subset \mathrm{SU}_2(\mathbb{C})$ by orbifolding four-dimensional $N = 4$ supersymmetric gauge theory by G, resulting in an $N = 2$ superconformal field theory whose 'matter matrix' can be read off from the Coxeter–Dynkin diagram corresponding to G. The same game can be played with finite subgroups of $\mathrm{SU}_3(\mathbb{C})$, resulting in $N = 1$ superconformal field theories whose matter matrices resemble the NIM-reps of ${A_2}^{(1)}$. [**271**] use this to conjecture optimistically a McKay-type correspondence between singularities of type $\mathbb{C}^n/G$, for $G \subset \mathrm{SU}_n(\mathbb{C})$, and the modular invariants of ${A_{n-1}}^{(1)}$. This in their view would be the form *A–D–E* takes for higher-rank modular invariants. Their conjecture is still too vague to be probed.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

So far we have considered only integrable modules, which are necessarily at level $k \in \mathbb{N}$. But their modular behaviour can be mimicked at certain fractional levels, by the so-called *admissible modules* [**335**]. It is tempting to guess that there should be natural CFT and VOA interpretations for these, analogous to the integrable ones. The matrix S there is symmetric, but has no column of constant phase and thus naively putting it into Verlinde's formula (6.1.1b) will necessarily produce some negative numbers (it appears that they'll always be integers though). A legitimate fusion ring has been obtained for $A_1{}^{(1)}$ at fractional level in other ways [**26**], [**184**], and initial steps for $A_2{}^{(1)}$ have been made in [**221**]. VOA interpretations for $A_1{}^{(1)}$ admissible modules are given in [**2**], [**148**]. Serious doubt, however, on the relevance of these efforts has been cast by [**225**], [**378**]. Sorting this out is a high priority.

Related roles for other Kac–Moody algebras are slowly being found. The *twisted* affine algebras also have modular-like data, and arise naturally in the data for NIM-reps [**58**], [**226**]. *Lorentzian* Kac–Moody algebras have been proposed [**171**], [**285**] as the symmetries of 'M-theory', the conjectural 11-dimensional theory underlying superstrings. Relations between strings and Borcherds–Kac–Moody algebras are discussed in [**275**], [**276**], [**134**].

6.2.2 Vertex operator algebras

Let $\mathcal{V}$ be a 'nice' VOA (more on this shortly). The primaries $a \in \Phi$ label the finitely many irreducible $\mathcal{V}$-modules M^a. The relation between VOAs and $\mathrm{SL}_2(\mathbb{Z})$ given in (4.3.9) was anticipated by RCFT, and proved by Zhu (Theorem 5.3.8). It gives (among other things) the modular matrices S and T. Do they define modular data? If so, does Verlinde's formula (6.1.1b) compute the dimensions of intertwiner spaces (6.1.17)?

Definition 6.2.3 *By a rational vertex operator algebra (RVOA) we mean a weakly rational vertex operator algebra $\mathcal{V}$ (Definition 5.3.2) obeying in addition*

(i) *$\mathcal{V}$ is simple (that is is an irreducible module for itself) and the contragredient $\mathcal{V}^\star$ is isomorphic to $\mathcal{V}$ as a $\mathcal{V}$-module;*
(ii) *$M_0 = \{0\}$ for all irreducible modules $M \neq \mathcal{V}$;*
(iii) *every $\mathbb{N}$-graded weak module is completely reducible;*
(iv) *$\mathcal{V}$ is C_2-cofinite (Definition 5.3.5).*

C_2-cofiniteness is a technical condition with many consequences. As we know, every VOA is a module for itself; the contragredient of a module is discussed around (5.3.4a). In any unitary RCFT, all conformal weights h_a, $a \in \Phi$, are positive except for $a = 0$, so condition (ii) is then automatic. Condition (iii) is a little stronger than the usual complete reducibility requirement.

This use of the term 'rational' is not standard, and different definitions of 'RVOA' can be found in the literature (some of these are listed in appendix A of [**224**]). But the term 'rational VOA' should be limited to those VOAs that possess some variant of modular data. The justification for our use of the term is the following recent theorem:

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

Theorem 6.2.4 (Huang [297]) *Let $\mathcal{V}$ be a VOA, rational in the sense of Definition* 6.2.3. *Let Φ label its (finitely many) irreducible modules, let $\mathcal{N}_{ab}^{c}$ be the dimension of the space $\mathcal{V}\binom{c}{a\,b}$ of intertwiners, and let S be the matrix defined in Theorem 5.3.8, satisfying* (4.3.9*a*). *Then Verlinde's formula* (6.1.1*b*) *holds and S is symmetric. Also, the category Rep $\mathcal{V}$ of $\mathcal{V}$-modules has a natural structure as a modular category.*

The objects of the category Rep $\mathcal{V}$ are $\mathcal{V}$-modules, and the morphisms are $\mathcal{V}$-module homomorphisms. A modular category is described in Section 6.2.5 and is (among many other things) a braided monoidal category. Theorem 6.2.4 is a corollary to Huang's programme of constructing geometric VOAs (Section 5.4.1) in genus ≤ 1 from an algebraic VOA. It appears that additional minor conditions on the VOA $\mathcal{V}$ will be needed [**296**] in order that the higher-genus chiral blocks be constructed – once identified, these restrictions should be included in the definition of rationality for VOAs. Extending this work to genus > 1 would be the final step in associating a modular functor – that is, a chiral half of an RCFT, including all the Moore–Seiberg data – to a nice VOA.

Equation (6.1.1b) can be defined only if all $S_{M0} \neq 0$, so Theorem 6.2.4 certainly implies that. Some RVOAs (e.g. those associated with non-unitary RCFTs) won't possess modular data in the narrow sense of Definition 6.1.6. However, suppose in addition to being rational that $\mathcal{V}$ has the (common) property that any irreducible module $M \neq \mathcal{V}$ has positive conformal weight h_M (recall $h_M - c/24$ is the smallest power of q in the Fourier expansion of the graded dimension $\chi_M(\tau) = q^{-c/24} \sum_{n=0}^{\infty} a_n^M q^{n+h_M}$). This holds for instance in all VOAs associated with unitary RCFTs. Then consider the behaviour of $\chi_M(\tau)$ for $\tau \to 0$ along the positive imaginary axis: since each Fourier coefficient a_n^M is nonnegative, $\chi_M(\tau)$ will go to $+\infty$. But this is equivalent to considering the limit of $\sum_N S_{MN}\, \chi_N(\tau)$ as $\tau \to \mathrm{i}\infty$ along the positive imaginary axis. By hypothesis, this latter limit is dominated by $S_{M0}\, a_0^0 q^{-c/24}$, at least when $S_{M0} \neq 0$. So what we find is that, under this hypothesis, the 0-column of S consists of nonnegative real numbers (and also that the central charge c is positive). But Verlinde's formula certainly requires that all numbers in the 0-column of S be nonzero. Thus we get:

Corollary 6.2.5 *Suppose $\mathcal{V}$ is a rational VOA and for all irreducible modules M, $M_n = 0$ for all $n < 0$. Then* (4.3.9) *(more precisely Theorem 5.3.8) define modular data.*

Of course the affine algebra modular data discussed in Section 6.2.1 is a special case of that considered here, corresponding to the integrable affine VOA $\mathcal{V}(\mathfrak{g}, k)$ constructed in Section 5.2.2.

Verlinde's formula (6.1.1b) is only a genus-0 special case of (6.1.2). What makes the proof of Theorem 6.2.4 difficult is the difficulty in constructing chiral blocks in genus > 0. At the time of writing, only special cases have been worked out in arbitrary genus (see, e.g., theorem 6.2 in [**573**]). Moore–Seiberg bypassed this difficulty by assuming the chiral blocks all exist and have all the required properties.

As mentioned in Section 5.3.5, one direction Huang's Theorem could possibly be extended is to 'quasi-rational' CFT [**436**]. These are VOAs with infinitely many

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

irreducible modules, but with finite fusion products (5.3.3). They would correspond to a 'C_1-cofiniteness' condition and typically have infinite-dimensional Zhu's algebra. The easiest example is the Heisenberg VOA (5.2.5), associated with the oscillator algebra $\mathfrak{u}_1{}^{(1)}$ (3.2.12). We find directly from (3.2.12c) that the graded dimension of $V(\lambda)$ obeys

$$\chi_\lambda(\tau+1) = e^{\pi \mathrm{i}(\lambda^2-\frac{1}{12})}\chi_\lambda(\tau), \tag{6.2.7a}$$

$$\chi_\lambda(-1/\tau) = \int_{-\infty}^{\infty} e^{2\pi\mathrm{i}\lambda\mu}\chi_\mu(\tau)\,\mathrm{d}\mu. \tag{6.2.7b}$$

In other words, on the Hilbert space $L^2(\mathbb{R})$ of square-integrable functions $f(\alpha)$, let $S(f)$ be the Fourier transform of f, and $T(f)$ the function given by

$$T(f)(\alpha) = e^{\pi\mathrm{i}(\alpha^2-\frac{1}{12})}f(\alpha)$$

Then S and T define a unitary representation of $\mathrm{SL}_2(\mathbb{Z})$ on the space $L^2(\mathbb{R})$ spanned by the χ_λ (more precisely, they act on the space of functions $\chi_f(\tau) = \int_{-\infty}^{\infty} f(\alpha)\,\chi\alpha(\tau)\mathrm{d}\alpha$ for $f \in L^2(\mathbb{R})$). In Verlinde's formula (6.1.1b), the sum over Φ becomes an integral over $\mathbb{R}$, and yields the distribution

$$\mathcal{N}_{\lambda\mu}^{\nu} = \delta(\nu-\lambda-\mu),$$

in other words $L(\lambda) \boxtimes L(\mu) = L(\nu)$, so the 'fusion ring' $L^2(\mathbb{R})$ is given a convolution product.

It can be hoped that this modular behaviour would be typical for a wide class of other quasi-rational theories. The generalisation of Zhu's Theorem 5.3.8 and Huang's Theorem 6.2.4 to such quasi-rational theories would be wonderful to see.

Modular invariants have a VOA interpretation. Let M^a and M'^i be the irreducible modules of RVOAs $\mathcal{V} \subset \mathcal{V}'$ sharing the same conformal vector ω. Then each M'^i is a $\mathcal{V}$-module. An RVOA is completely reducible, so each M'^i should be expressible as a direct sum of M^a's – these are called the branching rules. The sum of $\sum_{i\in\Phi'} |\chi'_{M'^i}|^2$ is invariant under that $\mathrm{SL}_2(\mathbb{Z})$-action; rewriting the $\chi'_{M'^i}$'s there in terms of the χ_{M^a}'s via the branching rules yields a nontrivial modular invariant for $\mathcal{V}$.

For instance, the VOA $L(\omega_0)'$ corresponding to the affine algebra $G_2{}^{(1)}$ at level 1 contains the VOA $L(28\omega_0) = L(0)$ for $A_1{}^{(1)}$ at level 28. We get the branching rules

$$L(\omega_0)' = L(0) \oplus L(10) \oplus L(18) \oplus L(28),$$
$$L(\omega_2)' = L(6) \oplus L(12) \oplus L(16) \oplus L(22).$$

Thus the $\mathcal{Z}' = I$ modular invariant for $G_2{}^{(1)}$ level 1 yields the $A_1{}^{(1)}$ modular invariant $\mathcal{E}_8$ in Theorem 6.2.2.

So knowing the modular invariants for an RVOA $\mathcal{V}$ gives considerable information concerning its possible 'nice' extensions $\mathcal{V}'$. For instance, we are learning from this that the only finite extensions of a generic integrable affine algebra VOA are those studied in [**147**] ('simple-current extensions'), and whose modular data is given in [**212**].

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

6.2.3 Quantum groups

The chiral data of affine algebras and Wess–Zumino–Witten models is also recovered by quantum groups (deformations of the universal enveloping algebra $U(\overline{\mathfrak{g}})$), though the reasons are still somewhat mysterious (i.e. indirect).

Over the years large numbers of two-dimensional models in statistical mechanics were found that are exactly solvable (completely integrable). Gradually it became clear that the underlying reason was the so-called *(quantum) Yang–Baxter equation* [**394**]:

$$R^{12}\,R^{13}\,R^{23} = R^{23}\,R^{13}\,R^{12}, \tag{6.2.8}$$

where $R : V \otimes V \to V \otimes V$ is linear and where, for example, $R^{13} : V \otimes V \otimes V \to V \otimes V \otimes V$ sends $v_1 \otimes v_2 \otimes v_3 \in V \otimes V \otimes V$ to $\sum_i a_i \otimes v_2 \otimes b_i$, where $R(v_1 \otimes v_3) = \sum_i a_i \otimes b_i$. (Generalisations of (6.2.8) exist but this is enough for us.) The Yang–Baxter equation should make us think of braids (recall Figure 1.29) and indeed an easy result is:

Proposition 6.2.6 *Given a solution R to* (6.2.8)*, we obtain a representation of the braid group $\mathcal{B}_n$ on $V \otimes \cdots \otimes V$ (n times) by sending the braid generator σ_i to $(\tau R)^{i,i+1}$, defined by $(\tau R)^{i,i+1}(v_1 \otimes \cdots \otimes v_n) = v_1 \otimes \cdots v_{i-1} \otimes (\sum_j b_j \otimes a_j) \otimes v_{i+1} \otimes \cdots \otimes v_n$, where $R(v_i \otimes v_{i+1}) = \sum_j a_j \otimes b_j$.*

The 'transpose' τ in Proposition 6.2.6 is the flip of the two copies of V; we see it again in Definition 6.2.8. The reader should try to prove the proposition, but it's also proved in section 15.2A of [**98**].

We are interested in families $R = R(q)$ of solutions to (6.2.8), depending on a complex parameter q. Write $q = e^{\mathrm{i}\hbar}$. If we Taylor expand $R(e^{\mathrm{i}\hbar}) = \sum_{n=0}^{\infty} \hbar^n r_n$ and retain only the first-order terms in $\hbar$, we obtain the classical Yang–Baxter equation for $r := r_1$:

$$[r^{12}, r^{13}] + [r^{12}, r^{23}] + [r^{13}, r^{23}] = 0. \tag{6.2.9}$$

Being a sum of commutators, it's reminiscent of Lie algebras and indeed Lie theory provides classes of solutions [**98**], [**394**]. Roughly, *quantum groups* were proposed by Drinfel'd and Jimbo around 1985 as a Lie-like symmetry underlying (6.2.8), that is, as providing a way to solve the quantum Yang–Baxter equation using q-deformations of Lie theory.

The idea of deformations [**279**] is a beautiful one. For example, consider n-space $\mathbb{R}^n$ and fix a vector $q \in \mathbb{R}^n$ (the 'deformation parameter'). Define the new multiplication by scalars to be $k \cdot_q x := kx + (1-k)q$ and vector addition to be $x +_q y := x + y - q$ (where the operations on the right sides are the usual $\mathbb{R}^n$ ones). The zero-vector here is $0_q := q$. This defines a new vector-space structure on the same underlying space. However, it is of course isomorphic (as a vector space) to the original one, since the dimension hasn't changed.

The finite-dimensional complex semi-simple Lie algebras $\overline{\mathfrak{g}}$ are also rigid in this sense (see Question 6.2.3(b)). However, nontrivial deformations of their universal enveloping algebras $U(\overline{\mathfrak{g}})$ (Section 1.5.3) do exist.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

Consider for concreteness $\overline{\mathfrak{g}} = A_1$, with basis e, f, h of (1.4.2b). Define

$$[e, f] = \frac{q^h - q^{-h}}{q - q^{-1}}, \tag{6.2.10a}$$

$$q^h e = q^2 e q^h, \tag{6.2.10b}$$

$$q^h f = q^{-2} f q^h. \tag{6.2.10c}$$

Here by, for example, 'q^h' we mean the Taylor expansion in powers of h. These equations define the *quantum group* $U_q(A_1)$, a one-parameter deformation of $U(A_1)$. Given this, we get a solution $R(q)$ to (6.2.8):

$$R(q) = \sum_{n=0}^{\infty} q^{\frac{n(n+1)}{2}} \frac{(1 - q^{-2})^n}{\lfloor n \rfloor_q!} (q^{-h} e)^n \otimes (q^h f)^n e^{\frac{h \otimes h}{2}}, \tag{6.2.10d}$$

where $\lfloor n \rfloor_q! = \lfloor n \rfloor_q \lfloor n-1 \rfloor_q \cdots \lfloor 1 \rfloor_q$ for $\lfloor k \rfloor_q = (q^k - q^{-k})/(q - q^{-1})$. Nevertheless, these equations look random and opaque (to this author at least). The next few paragraphs aim to make some sense out of them.

Definition 6.2.7 *Let k be a ring (take $k = \mathbb{C}$ if this generality makes you uncomfortable). A Hopf algebra A is:*

(i) *An associative algebra over k with unit $\mathbf{1}$ and multiplication μ.*

(ii) *A co-associative co-algebra over k, i.e. with co-multiplication $\Delta : A \to A \otimes A$ and co-unit $\epsilon : A \to k$.*

(iii) *The algebra and co-algebra structures are compatible, i.e. Δ and ϵ are algebra homomorphisms, and μ and $\mathbf{1}$ (regarded as a map $\iota : k \to A$ sending $x \mapsto x\mathbf{1}$) are co-algebra homomorphisms.*

(iv) *A has a map $S : A \to A$, called the antipode, which obeys*

$$\mu \circ (id \otimes S) \circ \Delta = \iota \circ \epsilon = \mu \circ (S \otimes id) \circ \Delta.$$

We've seen 'algebra' before. A Hopf algebra may or may not be commutative as an algebra. A 'co-algebra' is an 'algebra with the arrows reversed': just as an algebra has a bilinear map $A \otimes A \to A$ (multiplication), so a co-algebra has a linear map $A \to A \otimes A$ (co-multiplication), and similarly for unit and co-unit.

Perhaps [**51**] or the introduction to [**398**] can help make this definition seem more natural. Hopf algebras are algebras with a rich representation theory. If M, N are modules of a generic algebra A, then their usual vector-space tensor product $M \otimes N$ always has a natural structure as an $A \otimes A$-module, but generally not an A-module. But if A has a co-product, we get the A-module structure by the formula $a.(m \otimes n) := \Delta(a).(m \otimes n)$. The antipode converts left modules into right modules, and is used to define the representation M^* dual to a given representation M. It plays the role of inverse in the algebra. See also Question 6.2.4.

For example, a universal enveloping algebra $U(\overline{\mathfrak{g}})$ forms a Hopf algebra with co-product given by $\Delta(x) = x \otimes \mathbf{1} + \mathbf{1} \otimes x$ for $x \in \overline{\mathfrak{g}}$ and $\Delta(\mathbf{1}) = \mathbf{1} \otimes \mathbf{1}$; co-unit $\epsilon(x) = 0$ for $x \in \overline{\mathfrak{g}}$ and $\epsilon(\mathbf{1}) = \mathbf{1}$; and antipode $S(x) = -x$ for $x \in \overline{\mathfrak{g}}$ and $S(\mathbf{1}) = \mathbf{1}$. In a similar way, the space $F(G)$ of functions on a Lie group G is also a Hopf algebra (in fact a dual

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

of $U(\overline{\mathfrak{g}})$). $U(\overline{\mathfrak{g}})$ is co-commutative, whereas $F(G)$ is commutative; in fact, these $U(\overline{\mathfrak{g}})$ are the only co-commutative, and $F(G)$ the only commutative, Hopf algebras (modulo certain technical assumptions). This is in fact why Drinfel'd [**160**] cooked up the name 'quantum group' for these q-deformations. $U_q(\overline{\mathfrak{g}})$ is a non-co-commutative deformation of $U(\overline{\mathfrak{g}})$, so we could imagine that just as the dual of $U(\overline{\mathfrak{g}})$ consists of the functions on a group G, the dual of $U_q(\overline{\mathfrak{g}})$, which will be a non-commutative Hopf algebra, should correspond to something like the functions on a group-like object G_q, which would be some sort of q-deformed version of G. This picture is in the same spirit as Connes' non-commutative geometry. In any case the term 'quantum group' has inappropriately slipped from G_q to apply directly to $U_q(\overline{\mathfrak{g}})$.

The co-product, etc. for these $U_q(\overline{\mathfrak{g}})$ are explicitly given in proposition 6.5.1 of [**98**] in full generality. Although $U_q(\overline{\mathfrak{g}})$ is not co-commutative, it is nearly so:

Definition 6.2.8 *A quasi-triangularisable Hopf algebra A is a Hopf algebra with invertible element $\mathcal{R} \in A \otimes A$ such that $\tau(\Delta(a)) = \mathcal{R}\,\Delta(a)\,\mathcal{R}^{-1}$ for all $a \in A$, as well as*

$$(\Delta \otimes id)(\mathcal{R}) = \mathcal{R}^{13}\mathcal{R}^{23} \in A \otimes A \otimes A,$$
$$(id \otimes \Delta)(\mathcal{R}) = \mathcal{R}^{13}\mathcal{R}^{12} \in A \otimes A \otimes A.$$

This element $\mathcal{R}$ is called the *universal R-matrix* (or braiding) of A. Of course if A is co-commutative, then $\mathcal{R} = \mathbf{1} \otimes \mathbf{1}$ works. The point: the element $\mathcal{R}$ satisfies the quantum Yang–Baxter equation (6.2.8). This is the origin of the word 'triangular' in Definition 6.2.8: an alternate name for the Yang–Baxter equation is the star–triangle relation. So given any representation of A, $\mathcal{R}$ maps to a matrix satisfying (6.2.8) – this representation-independent aspect of $\mathcal{R}$ justifies the word 'universal'. Any non-co-commutative quasi-triangularisable Hopf algebra is now called a quantum group.

Drinfel'd [**160**] found a remarkable way, independent of the Yang–Baxter equation, to construct quantum groups from any Hopf algebra A. The quasi-triangular Hopf structure is put on the space $A \otimes (A^*)^{op}$, where $(A^*)^{op}$ is the dual Hopf algebra A^* except that its co-multiplication is changed from Δ^* to its transpose $\tau \circ \Delta^*$. A nice discussion is in [**480**]; a general categorical interpretation is the 'centre construction' [**338**]. In particular, the quantum group $U_q(\overline{\mathfrak{g}})$ of (6.2.10) arises as a simple quotient of the quantum double of $U_q(B^+)$, where B^+ is the Borel subalgebra of $\overline{\mathfrak{g}}$, generated by h_i and e_i. See section 4.6 of [**207**] , where this is discussed very explicitly. The point is that $U_q(B^+)$ is very easy to understand, so this gives an explicit way to compute $\mathcal{R}$ for $U_q(\overline{\mathfrak{g}})$.

As usual we're interested in representation theory. Recall that the modules of A_1 and $U(A_1)$ are identical. There is only one one-dimensional A_1-module: everything gets sent to 0. However, there are exactly two one-dimensional representations of the quantum group $U_q(A_1)$: $e.v = f.v = 0$ and $q^h.v = \pm v$. Call these $\psi_\pm$. ψ_+ is just the deformation of the trivial $U(A_1)$-representation, but ψ_- has no classical (i.e. $q \to 1$) analogue. The existence of ψ_- is the only difference between the representation theory of $U_q(A_1)$ and $U(A_1)$ (or A_1): every finite-dimensional irreducible $U_q(A_1)$ module is uniquely expressible as the tensor product of a one-dimensional representation $\psi_\pm$ with

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

some highest-weight representation $L_q(m)$, for $m \in \mathbb{N}$, where $L_q(m)$ is a deformation of $L(m)$ with the same Weyl character. This generalises to any $U_q(\overline{\mathfrak{g}})$.

We're more interested in $U_q(\overline{\mathfrak{g}})$ 'at a root of unity'. The meaning of this is very subtle, but is explained very thoroughly in chapter 9 of [**98**] (we are interested in their second construction, the 'restricted integral form' $U_q^{res}(\overline{\mathfrak{g}})$, which is a quotient of $U_q(\overline{\mathfrak{g}})$); see also [**10**], [**392**]. The representation theory is also subtle, and most treatments (e.g. that of [**98**]) assume from the start that the order of the root of unity must be odd. See, for example, [**392**], [**10**] for their modules. There are now indecomposable modules that are not irreducible, a common situation in algebra (recall Question 1.1.6). The trick of how to proceed was discovered by physicists: throw the sick modules away! In particular, when we evaluate the Weyl characters at the root of unity q, the result is called the *quantum dimension* of the module. We keep those modules with nonzero quantum dimension, and discard the others. This prescription works because the direct product of any $U_q^{res}(\overline{\mathfrak{g}})$-module with any sick one is a direct sum of sick ones. We can call this 'the reduced representation ring of the quantum group $U_q(\overline{\mathfrak{g}})$ specialised to the root of unity q'. See section 4.5 of [**207**] for examples (though note that his q is the square of ours).

The result is somewhat surprising: this reduced representation ring, for $q = e^{\pi \mathrm{i}/m(k+h^\vee)}$ (where m is defined below), is isomorphic to that of the fusion ring of $\overline{\mathfrak{g}}^{(1)}$ at level k [**190**]. Here, $m = 1$ for $\overline{\mathfrak{g}} = A_r, D_r, E_6, E_7, E_8$; $m = 2$ for $\overline{\mathfrak{g}} = B_r, C_r, F_4$; and $m = 3$ for $\overline{\mathfrak{g}} = G_2$.

More generally, much of the chiral data of the Wess–Zumino–Witten theories are recovered by the corresponding quantum group at a root of unity [**253**], [**207**]: along with the fusion multiplicities, also the braiding and fusing matrices of Section 6.1.4, and the associated knot invariants of Section 6.2.5. Explanations for these 'coincidences' are given in, for example, chapter 11 of [**253**], but they are all unsatisfying in that they are so indirect.

6.2.4 Twisted #6: finite group modular data

In many respects, a finite group G behaves much like a compact connected Lie group, and so we may hope that they possess an analogue of Section 6.2.1. Indeed that is beautifully the case.

For any finite group G (Section 1.1), let $K_1, \ldots, K_h$ be its conjugacy classes, and write k_i for $\sum_{g \in K_i} g \in \mathbb{C}G$. These k_i's form a basis for the centre of $\mathbb{C}G$. Write

$$k_i\, k_j = \sum_\ell c_{ij}^\ell k_\ell; \tag{6.2.11a}$$

then the structure constants c_{ij}^ℓ are nonnegative integers, and we obtain

$$c_{ij}^\ell = \frac{\|K_i\|\,\|K_j\|}{\|G\|} \sum_{\mathrm{ch} \in \mathrm{Irr}\, G} \frac{\mathrm{ch}(g_i)\,\mathrm{ch}(g_j)\,\overline{\mathrm{ch}(g_\ell)}}{\mathrm{ch}(e)}, \tag{6.2.11b}$$

where $g_i \in K_i$. This resembles (6.1.1b), with S_{ab} replaced by $S_{i,\mathrm{ch}} = \mathrm{ch}(g_i)$ and the vacuum 0 by the identity e. Unfortunately, the other axioms of modular data fail.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

However, the group algebra $\mathbb{C}G$ is a Hopf algebra, with co-multiplication $\Delta(g) = g \otimes g$, co-unit $\epsilon(g) = 1$ and antipode $S(g) = g^{-1}$. The way to obtain true modular data is to take the quantum double of $\mathbb{C}G$. Its Hopf dual, the space $F[G]$ of functions $G \to \mathbb{C}$, is also a Hopf algebra, for example, with co-product $\Delta(f)(g_1, g_2) = f(g_1 g_2)$. The construction of the double $\mathcal{D}(G)$ is described nicely in [**406**]; we will simply describe its modular data.

Let Φ be the set of all pairs (a, ch), where the a are representatives of the conjugacy classes of G and ch is the character of an irreducible representation of the centraliser $C_G(a)$. (Recall that $C_G(a)$ is the set of all $g \in G$ commuting with a.) Φ parametrises the irreducible modules of the double $\mathcal{D}(G)$. Put [**393**], [**136**]

$$S_{(a,\mathrm{ch}),(a',\mathrm{ch}')} = \frac{1}{\|C_G(a)\| \, \|C_G(a')\|} \sum_{g \in G(a,a')} \overline{\mathrm{ch}'(g^{-1}ag)}\, \overline{\mathrm{ch}(ga'g^{-1})}, \qquad (6.2.12a)$$

$$T_{(a,\mathrm{ch}),(a',\mathrm{ch}')} = \delta_{a,a'} \delta_{\mathrm{ch},\mathrm{ch}'} \frac{\mathrm{ch}(a)}{\mathrm{ch}(e)}, \qquad (6.2.12b)$$

where $G(a, a') = \{g \in G \,|\, aga'g^{-1} = ga'g^{-1}a\}$ and $e \in G$ is the identity. For the ‘vacuum’ 0 take $(e, 1)$. Then (6.2.12) is modular data. Manifestly, $\mathbb{N}$-valued descriptions of the fusion multiplicity $\mathcal{N}_{(a,\mathrm{ch}),(b,\mathrm{ch}')}^{(c,\mathrm{ch}'')}$ exist (see section 2 of [**391**], who realises the fusion ring as the Grothendieck ring for G-equivariant vector bundles). For Lusztig, (6.2.12) arose in his determination of irreducible characters of Chevalley groups. The higher-genus fusion multiplicities in (6.1.2) also have interpretations as multiplicities of representations of $\mathcal{D}(G)$ in $\mathcal{D}(G) \otimes \cdots \otimes \mathcal{D}(G)$ [**35**].

For instance, the modular data associated with the finite group $\mathcal{S}_3$ is

$$S = \frac{1}{2} \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 \\ 1 & 1 & 2 & 2 & 2 & 2 & -3 & -3 \\ 2 & 2 & 4 & -2 & -2 & -2 & 0 & 0 \\ 2 & 2 & -2 & 4 & -2 & -2 & 0 & 0 \\ 2 & 2 & -2 & -2 & -2 & 4 & 0 & 0 \\ 2 & 2 & -2 & -2 & 4 & -2 & 0 & 0 \\ 3 & -3 & 0 & 0 & 0 & 0 & 3 & -3 \\ 3 & -3 & 0 & 0 & 0 & 0 & -3 & 3 \end{pmatrix}, \qquad (6.2.13a)$$

$$T = \mathrm{diag}(1, 1, 1, 1, e^{2\pi \mathrm{i}/3}, e^{-2\pi \mathrm{i}/3}, 1, -1). \qquad (6.2.13b)$$

See [**115**] for several more explicit examples.

This modular data can be twisted [**138**], [**135**], [**34**], [**115**] by a 3-cocycle $\alpha \in H^3(G, \mathbb{C}^\times)$. Indeed this twisted modular data is absolutely as fundamental as (6.2.12) – recall the discussion in Sections 4.3.4 and 5.3.6. This cocycle α plays the same role here that level does in affine algebra modular data, as $H^3(G, \mathbb{C}^\times) \cong \mathbb{Z}$ when G is simply-connected and simple. This sort of twist has a generalisation to arbitrary chiral data [**118**].

One of the remarkable features of affine algebra modular data – its ubiquity – is shared by finite group modular data. Most important for us, it arises in the orbifold of holomorphic VOAs (recall Section 5.3.6). Let G be a finite group of automorphisms

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

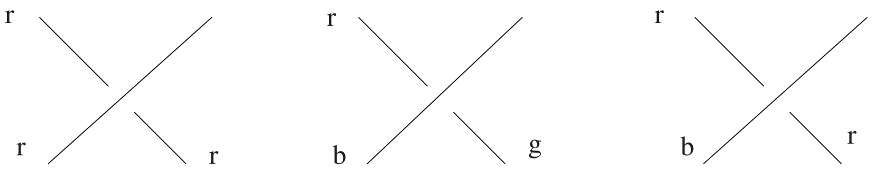

Fig. 6.7 Colourings at a crossing.

of a holomorphic VOA $\mathcal{V}$ – all finite groups arise in this way (Question 6.2.7). Let $\mathcal{V}^G$ be the space of fixed points of G; it inherits a VOA structure from $\mathcal{V}$. Then the modular data of $\mathcal{V}$ is trivial but that of $\mathcal{V}^G$ is expected to be (6.2.12) or some twisted version (see Conjecture 5.3.10). This modular data also appears in the crossed-product construction in von Neumann algebras (Section 6.2.6). In physics, it arises in $(2+1)$-dimensional Chern–Simons theory with finite gauge group G [**138**], [**194**], as well as $(2+1)$-dimensional quantum field theories where a continuous gauge group has been spontaneously broken to a finite group [**31**] (adding a Chern–Simons term here corresponds to the cohomological twist).

This modular data is quite interesting for nonabelian G, and deserves more study. It seems very effective at distinguishing groups – in fact, it is known to distinguish all groups of order < 128. Conversely, there are non-isomorphic groups of order $2^{15} \cdot 3^4 \cdot 5 \cdot 7$ with identical modular data up to reordering primaries [**175**]. Finite group modular data behaves very differently from the affine algebra data (see e.g. [**115**], [**457**], [**178**]). For instance, Eiichi Bannai has found that the alternating group $\mathcal{A}_5$, which has only 22 primaries, has a remarkably high number (8719) of modular invariants. By contrast, affine algebras have relatively few modular invariants.

6.2.5 Knots

The Jordan curve theorem states that all knots in $\mathbb{R}^2$ are trivial. Are there any nontrivial knots in $\mathbb{R}^3$?

In Figures 1.9 and 1.10 are some knots in $\mathbb{R}^3$, flattened into the plane of the paper. A moment's consideration will confirm that the second knot of Figure 1.9 is indeed trivial. What about the trefoil?

A knot diagram cuts the knotted S^1 into several connected components (*arcs*), whose endpoints lie at the various *crossings* (double-points of the projection). By a *3-colouring*, we mean to colour each arc in the knot diagram either red, blue or green, so that at each crossing either one or three distinct colours are used. For example, the first two colourings in Figure 6.7 are allowed, but the third isn't. By considering the 'Reidemeister moves' (Figure 1.12), which tell us how to move between equivalent knot diagrams, different diagrams for equivalent knots (such as the two in Figure 1.9) are seen to have equal numbers of distinct 3-colourings. Hence, the number of 3-colourings is a knot invariant.

For example, consider the diagrams in Figure 1.9 for the trivial knot: clearly, all arcs must be given the same colour, and thus there are precisely three distinct 3-colourings.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

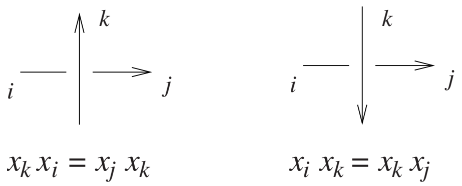

Fig. 6.8 The Wirtinger presentation of the knot group.

On the other hand, the trefoil has nine distinct 3-colourings – the bottom two arcs of Figure 1.10 can be assigned arbitrary colour, and that choice fixes the colour of the top arc. Thus the trefoil is nontrivial!

Essentially what we are doing here is counting the number of homomorphisms φ from the *knot group* $\pi_1(\mathbb{R}^3 \setminus K)$ of knot K to the symmetric group $\mathcal{S}_3$. The reason is that any (oriented) knot diagram gives a presentation for $\pi_1(\mathbb{R}^3 \setminus K)$, where there is a generator x_i for each arc and a relation of the form $x_i x_j = x_k x_i$ for each crossing (Figure 6.8). See section 3.D of [**478**] for more details and a proof. For example, the knot group of the right knot of Figure 1.9 has presentation

$$\langle x_1, \ldots, x_7 \,|\, x_1x_2 = x_4x_1,\, x_5x_1 = x_3x_5,\, x_5x_4 = x_3x_5,\, x_2x_1 = x_5x_2,$$
$$x_2x_7 = x_2x_2,\, x_2x_7 = x_6x_2,\, x_2x_5 = x_6x_2\rangle,$$

which is isomorphic to $\mathbb{Z}$. By contrast, the knot group of the trefoil is $\mathcal{B}_3$ (Question 6.2.8). Incidentally, the complement $\mathbb{R}^3 \setminus K$ of a knot determines the knot, and the extent to which the knot group determines the knot is also understood (see section 1 of [**61**]). Therefore, in this sense the trefoil and $\mathcal{B}_3$ are intimately connected (recall Section 2.4.3).

$\mathcal{S}_3$ is generated by the transpositions (12), (23), (13). The homomorphism $\varphi : \pi_1(\mathbb{R}^3 \setminus K) \to \mathcal{S}_3$ is defined using, for example, the identification $r \leftrightarrow (12)$, $b \leftrightarrow (23)$, $g \leftrightarrow (13)$, and the above 3-colouring condition at each crossing is equivalent to requiring that φ obeys each relation in the Wirtinger presentation. Our homomorphism φ will be onto iff at least two different colours are used. By considering more general (non-abelian) colourings, the target ($\mathcal{S}_3$ here) can be made to be any other group G, resulting in a different knot invariant.

In the early 1980s, knot theory was dormant; by the late 1980s it was flourishing. But as a consequence, we suddenly had too many knot invariants. Reshitikhin and Turaev [**473**] brought order to this chaos, by proving that whenever we have a ribbon category **V**, we get invariants of (framed) knots and links, that is of knotted and linked ribbons. The reason for their result, as we explain in Section 1.6.2, is the universality of the topological category **Ribbon** of ribbons (Theorem 1.6.2). Given any knotted link, coloured with the objects of **V**, their functor associates the link with some morphism Hom(∅, ∅) of **V**, and isotopic links get assigned the same morphism. This morphism is the desired link invariant. For example, the 3-colouring invariant comes from a ribbon category associated with the modular data (6.2.13).

We can express their result slightly differently. Suppose we have a representation of every braid group $\mathcal{B}_n$ (e.g. Proposition 6.2.6 says we get this from a solution to

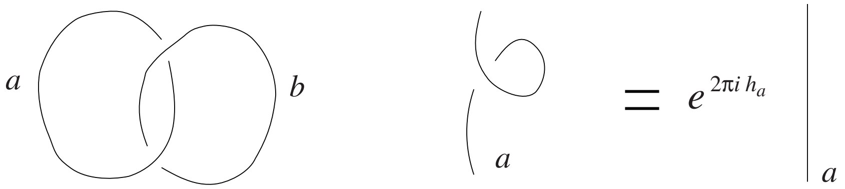

Fig. 6.9 The S_{ab} and T_{aa} matrix entries in modular categories.

the quantum Yang–Baxter equation). To every braid we get a link by closing it up, as in Figure 1.14. Unfortunately, different braids can get assigned the same link. As we explain in Section 1.2.3, the two Markov moves capture precisely this redundancy. Thus we get a link invariant from our braid representations if we can construct a quantity invariant with respect to these two moves. The first move $\beta' \leftrightarrow \beta\beta'\beta^{-1}$ suggests we assign to the braid the trace of its representing matrix; unfortunately, that usually won't respect the second move, $\beta \leftrightarrow \beta T_m^{\pm 1}$.

However, [**473**] explain how to enhance the braid representation coming from any quasi-triangularisable Hopf algebra (Definition 6.2.8), to get link invariants. See section XI.3.1 of [**534**] for details. Thus, combining their construction with the Drinfel'd double, which associates a quasi-triangularisable Hopf algebra with any Hopf algebra, we can construct (or recover) enormous numbers of link invariants.

So far we have discussed invariants of links embedded in $\mathbb{R}^3$ (equivalently, S^3). Much more difficult is to construct invariants of links in arbitrary 3-manifolds, but it is precisely this that is relevant to our story. There are (at least) two ways to do this: one uses 'Dehn surgery' to construct the manifold from S^3 [**474**], and the other uses triangulation by tetrahedra [**535**]. We allude to the *Turaev–Viro theory* [**535**] elsewhere. In the early 1960s Lickorish and Wallace established that any closed compact oriented 3-manifold M can be obtained by surgery on the 3-sphere S^3 along some framed link L (see section II.2.1 of [**534**] for details). The idea is to construct an invariant for M from the link invariant of L in S^3. For instance, the 3-manifold $S^1 \times S^2$ arises from S^3 by surgery along the trivial ribbon. The problem is that different links give rise to the same manifold. However, this redundancy is completely captured by the so-called 'Kirby moves' (see section II.3.1 of [**534**] for details). Once again, Reshitikhin and Turaev [**474**] find the necessary refinement to ribbon categories, as well as the precise expression for the 3-manifold invariant, which will make the quantity invariant under the Kirby moves. The result is called a *modular category* (see chapter 2 of [**534**] for complete details). Roughly speaking, it is a ribbon category with the additional property of direct sum, with a finite set of 'simple objects' (closed under $*$) and a complete reducibility property, whose Hopf link invariant (Figure 6.9) is nondegenerate. More generally, this procedure gives us link invariants in any 3-manifold. Again, the ultimate source of these topological invariants is a universality property of the appropriate topological category. All of these universalities have as their source the universality of **Braid** for braided monoidal categories (Theorem 1.6.1).

Any RCFT gives a modular category (in fact two of them, one for each chiral half). For an RCFT, the simple objects are the objects that are the chiral primaries, the monoidal

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

structure is the fusion product and duality is charge-conjugation. Modular data is obtained directly from the Hopf link and twist, as in Figure 6.9. There are thus three different incarnations of the S-matrix in RCFT: the modular transformation (4.3.9a), Verlinde's formula (6.1.2), and the Hopf link. In fact, the notion of a modular category is equivalent to that of Segal's modular functor (Section 4.4.1) [**534**], [**32**]. For a sufficiently nice VOA $\mathcal{V}$, the simple objects are the irreducible $\mathcal{V}$-modules. The 3-colouring invariant of Figure 6.7 comes from a holomorphic orbifold VOA, and as such can be modified to yield a link invariant in any 3-manifold.

For instance, we get S^3 knot invariants from the quantum group $U_q(X_r)$ with generic parameter, but to get invariants for any closed 3-manifold requires specialising q to a root of unity. Modular categories are far less common than ribbon categories, but they can be obtained by an analogue of the Drinfel'd double.

6.2.6 Subfactors

The final general source of modular data that we discuss is from subfactor theory. The relations of subfactors to knots is reviewed in, for example, [**317**], [**318**], [**319**], while reviews of the relation between subfactors and CFT can be found in [**177**], [**66**].

Recall the definitions in Section 1.3.2. Let $N \subset M$ be an inclusion of type II$_1$ factors. We call N a *subfactor*, provided N includes the identity of M. Jones' motivation for looking at subfactors came from their formal similarity with Galois theory. After all, the very notation $\dim_M(\mathcal{H})$ for the 'coupling constant' of Section 1.3.2 suggests thinking of a type II$_1$ factor as a non-commutative analogue of 'field of scalars'.

In particular, let G be a finite group acting on some type II$_1$ factor N. Then the crossed-product $N \rtimes G$ is also a type II$_1$ factor, iff each $g \in G$, $g \neq e$, is 'outer'. By an outer automorphism g of N we mean that there are no unitary operators $u \in N$ such that $g.x = uxu^*$ for all $x \in N$. Any locally compact (e.g. finite) group G acts on, for example, the hyperfinite type II$_1$ factor by outer automorphisms, so this isn't a major restriction. This yields a Galois correspondence between subgroups H of G, and subalgebras of M containing the algebra M^G of fixed points, given by $H \leftrightarrow M^H$. This is analogous to the relation between subfields $\mathbb{K} \subset \mathbb{L}$ and Galois groups in Section 1.7.2. So what is the subfactor analogue of the index $[\mathbb{L} : \mathbb{K}]$?

Jones' answer is the *Jones index* of the subfactor $N \subseteq M$:

$$[M : N] := \dim_N(L^2(M)) \geq 1, \tag{6.2.14}$$

where $L^2(M)$ is the Hilbert space of Question 1.3.6. For instance, for any $n \geq 1$, $[N \otimes M_n(\mathbb{C}) : N] = n^2$. If $H \leq G$ are finite groups of outer automorphisms, then $[M \rtimes G : M \rtimes H] = \|G\|/\|H\| = [M^H : M^G]$, where the crossed-product $M \rtimes H$ and fixed-point M^H factors are discussed in Section 1.3.2.

The following theorem was completely unexpected.

Theorem 6.2.9 [**316**] *For any number*

$$d \in \{4\cos^2(\pi/n)\}_{n=3}^{\infty} \cup [4, \infty],$$

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

there is a subfactor $N \subseteq M$ of the unique hyperfinite type II_1 *factor M, with index $[M : N] = d$. Conversely, the index of any subfactor of a (not necessarily hyperfinite) type* II_1 *factor will be in that set.*

In fact, the following rigidity is true: if M is the hyperfinite type II_1 factor, then at most four inequivalent subfactors $N \subseteq M$ can possess the same index < 4. The reader, with Section 2.5.2 fresh in mind, may recognise the discrete sequence of indices in Theorem 6.2.9 as the square of the Perron–Frobenius eigenvalues of the *A–D–E* graphs – is this a coincidence?

The key to proving Theorem 6.2.9, as well as the further developments, is the so-called *basic construction*, which appears to have been found independently by a number of people in the late 1970s. Let $N \subseteq M$ be an inclusion of type II_1 factors. Even though M and N are isomorphic as factors, there is rich combinatorics surrounding how N is embedded in M. The Hilbert space $L^2(N)$ is naturally contained in $L^2(M)$. Let e_N be the orthogonal projection of $L^2(M)$ onto $L^2(N)$. Then M and e_N generate the von Neumann algebra $\langle M, e_N\rangle''$ acting on the space $L^2(M)$. If the index $[M : N]$ is finite, then $\langle M, e_N\rangle''$ will also be a type II_1 factor, with index $[\langle M, e_N\rangle'' : M] = [M : N]$. Moreover, since the trace (normalised so that $\mathrm{tr}(1) = 1$) on a type II_1 factor is unique, we can unambiguously speak of the trace $\mathrm{tr}(e_N)$, and we find it equals $1/[M : N]$. For later convenience define $\tau := 1/[M : N]$.

For example, taking N to be the fixed points M^G, for some finite group G of outer automorphisms, then $e_N = (1/\|G\|)\sum_g g$, $\mathrm{tr}(e_N) = 1/\|G\|$ and $\langle M, e_N\rangle'' = M \rtimes G$. This demonstrates the naturalness of this construction. What is the von Neumann algebra generated by M and e? The answer is the crossed-product $M \rtimes G$.

We can repeat the basic construction indefinitely. Put $M_0 := N$, $M_1 := M$ and define inductively

$$M_{i+1} := \langle M_i, e_{i-1}\rangle'',$$

where $e_i := e_{M_{i-1}}$ is the orthogonal projection from $L^2(M_i)$ onto $L^2(M_{i-1})$. We thus get a tower $M_0 \subset M_1 \subset \cdots$ of type II_1 factors, and a sequence $e_1, e_2, \ldots$ of projections. The limit $M_\infty := \cup_{n=0}^{\infty} M_n$ is also a type II_1 factor, with a unique (normalised) trace tr, which restricts to the unique trace on each M_n. Thus each $\mathrm{tr}(e_n) = \tau$. The algebra $\mathcal{A}_{\infty,\tau}$ spanned by the projections e_i obeys the relations

$$e_i^2 = e_i^* = e_i, \tag{6.2.15a}$$

$$e_i e_{i\pm1} e_i = \tau e_i, \tag{6.2.15b}$$

$$e_i e_j = e_j e_i \quad \text{if } |i - j| \geq 2, \tag{6.2.15c}$$

$$\mathrm{tr}(x e_{n+1}) = \mathrm{tr}(x)\,\tau, \tag{6.2.15d}$$

where x is in the (finite-dimensional semi-simple) algebra $\mathcal{A}_{n,\tau}$ generated by $1, e_1, \ldots, e_{n-1}$. In fact these are the complete list of relations for $\mathcal{A}_{n,\tau}$, because the (normalised) trace tr on any type II_1 factor obeys $\mathrm{tr}(xx^*) \geq 0$ with equality only if $x = 0$. The (easy) proofs of all these statements are in [**319**]. The point is that the tower

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

$M_0 \subset M_1 \subset \cdots$ and the projections $e_1, e_2, \ldots$ depend only on the original subfactor. Positive-definiteness of the trace on $\mathcal{A}_{n,\tau}$ gives the discrete values of Theorem 6.2.9.

Of course we are now trained to recognise (6.2.15b) and (6.2.15c) as having to do with the braid groups. In particular, if we try to send the braid group generator σ_i to $ae_i + b$, we obtain the solution $a = t + 1$, $b = -1$, where t satisfies $t + t^{-1} + 2 = \tau^{-1}$. Thus to any finite index type II_1 subfactor, we get a representation of the braid group!

We know how to go from a braid group representation to a link invariant: we need to associate a number with each braid that is invariant under the two Markov moves (Section 1.2.3). For a braid $\beta \in \mathcal{B}_n$, the combination

$$J_\beta(t) = \left(-\left(\sqrt{t} + \frac{1}{\sqrt{t}}\right)\right)^{n-1} \sqrt{t}^{\deg \beta} \operatorname{tr}(\beta) \tag{6.2.16}$$

works (verify this), where 'deg β' is defined in Section 1.1.4 and 'tr(β)' means the trace of the corresponding element in M_n. This function J_β is the famous Jones polynomial.

Witten showed that the Jones polynomial can be recovered from the topological field theory (or modular category) associated with affine algebra ${A_1}^{(1)}$ at level $k \in \mathbb{N}$, when the highest weight $\omega_1 + (k-1)\omega_0$ is assigned to each strand of the link. Of course, there is no need to restrict to ${A_1}^{(1)}$ or that weight, and other choices yield other link invariants.

Can the subfactor approach also recover these other link polynomials, or is it inherently 'rank 1'? Is the full topological field theory (or if you prefer, the CFT or modular category) obtainable from the subfactor, or does the subfactor only see the link polynomials? The answer to both questions is yes; the construction was originally due to Ocneanu, and is explained carefully in [**177**] (see also [**354**] for a very accessible treatment of certain parts of the theory). The starting point is the realisation that the projections e_i are only a small part of the full tower $M_0 \subset M_1 \subset M_2 \subset \cdots$.

Subtleties in any representation theory arise through the interplay of addition with multiplication, and with contragredient (dual). Addition (direct sum) of modules comes for free here. Unfortunately, the modules of factors (which we briefly described at the end of Section 1.3.2) don't have an obvious tensor product, and in any case are rather colourless (e.g. there is a unique nontrivial module for type III factors).

The right objects to study here are *bimodules*. We call a Hilbert space $X = {}_M X_N$ an M–N bimodule if M acts on the left and N on the right. The point is that they have a natural multiplication: the relative tensor product ('Connes fusion') ${}_M X_N \otimes_N Y_P$ will be an M–P bimodule. The multiplicative identity (playing the role of the trivial one-dimensional module) is ${}_M L^2(M)_M$, usually abbreviated to ${}_M M_M$. Given any bimodule ${}_M X_N$, the conjugate Hilbert space $\overline{X}$ is naturally an N–M bimodule: $n\overline{x}m := m^* x n^*$. Moreover, the possibilities for bimodules are far richer than for modules.

Let $N \subset M$ be an inclusion of II_1 factors with finite Jones index $[M : N]$. Recall the tower $M_0 = N \subset M = M_1 \subset M_2 \subset \cdots$ arising from the basic construction. Let Φ_M denote the set of equivalence classes of irreducible M–M submodules of $\oplus_{n \geq 1}\, {}_M L^2(M_n)_M$, and Φ_N that for the irreducible N–N submodules of $\oplus_{n \geq 0}\, {}_N L^2\,(M_n)_N$. We require these sets to be finite ('finite depth'). Write $\mathcal{H}_{AB}^C$ for the

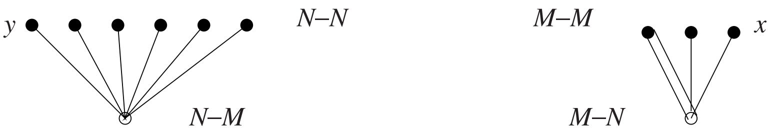

Fig. 6.10 The principal and dual principal graphs associated with $\mathcal{S}_3$.

(finite-dimensional) intertwiner space $\mathrm{Hom}_{M-M}(C, A \otimes_M B)$. For any $A, B \in \Phi_M$, the product $A \otimes_M B$ can be decomposed into a finite sum $\sum_{C \in \Phi_M} \mathcal{N}_{AB}^{C} C$, where $\mathcal{N}_{AB}^{C} = \dim \mathcal{H}_{AB}^{C} \in \mathbb{N}$ are the multiplicities. Indeed, all axioms of a fusion ring will be obeyed, except usually commutativity and self-duality.

Returning to the Galois theory analogy, the Jones index merely corresponds to the degree of the field extension. To what corresponds the Galois group? Ocneanu's answer is an intricate subfactor invariant called a *paragroup* [**453**] (see especially chapter 10 of [**177**]). It consists of two graphs (the *principal* and *dual principal*), whose vertices are bimodules for M and N; an order-2 involution of the vertices corresponding to the contragredient map $A \mapsto \overline{A}$; and a 'connection', that is an assignment of complex numbers to closed paths in the graphs, reminiscent of $6j$-symbols, describing the change between natural bases. The graphs are obtained from the fusion rings; their Perron–Frobenius eigenvalues equal the square-roots of the Jones index. For example, when the Jones index is < 4 (corresponding to eigenvalue < 2), those two graphs are equal, and are one of A_n, D_{even}, E_6 or E_8 (recall Figure 1.4) – it cannot be the tadpole T_n for elementary reasons, but D_{odd} and E_7 are excluded for their inability to support a connection. Two inequivalent connections are possible on the E_6 and E_8 graphs, corresponding to different subfactors. Thus Theorem 6.2.9 indeed constitutes another realisation of *A–D–E*, and for the ultimate reason suggested in Section 2.5.2.

A paragroup is a generalised ('quantised') sort of group. Figure 6.10 gives the graphs for $R \subset R \rtimes G$ (for R the hyperfinite II_1 factor and $G = \mathcal{S}_3$). The $M–M$ bimodules are parametrised by the irreducible characters ch_i of G, with precisely $\mathrm{ch}_i(e)$ edges connecting the ith node to the root of the graph. The $N–N$ bimodules are parametrised by elements of the group. The contragredient involution and fusion rings are the ones familiar to aficionados of character tables: complex-conjugate and the character ring, and $g \mapsto g^{-1}$ and the group ring $\mathbb{C}G$. The connection explicitly recovers the group structure, much as in the topological field theory of Section 4.4.2. On the otherhand, the graphs for $R^G \subset R$ are switched. More generally, given any subgroup $H < G$, we get subfactors $R^G \subset R^H$ and $R \rtimes H \subset R \rtimes G$, and their paragroups give a group-like interpretation to G/H even when H is not normal.

We say subfactors $N_i \subset M_i$ are equivalent if there is an isomorphism $\theta : M_1 \to M_2$ with $\theta(N_1) = N_2$. When M is hyperfinite type II_1, the paragroup identifies $N \subset M$ up to equivalence. Hence, when G is a finite abelian group, $R^G \subset R$ is equivalent to $R \subset R \rtimes G$ (when instead G is nonabelian, they are merely dual).

The paragroup yields a topological invariant for manifolds, generalising the Turaev–Viro one [**535**] (see [**354**] for a very readable treatment of this part of the theory).

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

However, it doesn't directly correspond to the data of an RCFT (e.g. the fusion rings of Figure 6.10 aren't self-dual). To get RCFT data, we must pass from $N \subset M$ to the 'asymptotic inclusion' $\langle M, M' \cap M_\infty \rangle \subset M_\infty$, where M_∞ is the (weak completion of the) union of all M_n. Asymptotic inclusion plays the role of Drinfel'd's quantum-double here, and corresponds physically to taking the continuum limit of the lattice model, yielding the CFT from the underlying statistical mechanical model (see section 12.6 of [**177**]). All chiral data of the VOA or RCFT, including the link invariants, are obtainable from the asymptotic inclusion. For instance, the Jones index $[M : N]$ equals $1/S_{00}^2$.

A very similar (but simpler) theory has been developed for type III factors. Bimodules now are equivalent to 'sectors', that is equivalence classes of endomorphisms $\lambda : N \to N$ (the corresponding subfactor is $\lambda(N) \subset N$). This use of endomorphisms is the key difference (and simplification) between the type II and type III fusion theories. Given $\lambda, \mu \in \mathrm{End}(N)$, we define $\langle \lambda, \mu \rangle$ to be the dimension of the vector space of intertwiners, that is all $t \in N$ such that $t\lambda(n) = \mu(n)t \; \forall n \in N$. The endomorphism $\lambda \in \mathrm{End}(N)$ is irreducible if $\langle \lambda, \lambda \rangle = 1$. Let Φ be a finite set of irreducible sectors. The fusion product is given by composition $\lambda \circ \mu$; addition can also be defined, and the fusion multiplicity $\mathcal{N}_{\lambda\mu}^{\nu}$ is then the dimension $\langle \lambda \circ \mu, \nu \rangle$. The 'vacuum' 0 is the identity id_N. Restricting to a finite set Φ of irreducible sectors, closed under fusion, the result is again a (noncommutative non-self-dual) fusion ring (after all, why should the compositions $\lambda \circ \mu$ and $\mu \circ \lambda$ be related). The missing ingredients are nondegenerate braidings $\epsilon^{\pm}(\lambda, \mu) \in \mathrm{Hom}(\lambda \circ \mu, \mu \circ \lambda)$, which say roughly that λ and μ nearly commute (the $\epsilon^{\pm}$ must also obey some analogue of the Yang–Baxter equation (6.2.8)). Provided we have a nondegenerate braiding (which we can obtain from asymptotic inclusion as before), Rehren [**470**] proved that we will automatically have modular data. When we have a hyperfinite type III$_1$ subfactor $N \subset M$ with a braided system of endomorphisms, there is a simple expression (see [**65**] and references therein) for the corresponding modular invariant (Definition 6.1.8) using 'α-induction' (a process of inducing an endomorphism from N to M using the braiding $\epsilon^{\pm}$): we get $\mathcal{Z}_{\lambda\mu} = \langle \alpha_\lambda^+, \alpha_\mu^- \rangle$. The NIM-rep is defined similarly [**65**].

Wassermann and collaborators (see e.g. [**554**]) have explicitly constructed the affine algebra subfactors, recovering the affine algebra modular data, at least for $A_r{}^{(1)}$ and $B_r{}^{(1)}$. To any subgroup–group pair $G < H$, the subfactor $R \rtimes G \subset R \rtimes H$ of crossed-products has a (in general non-commutative) fusion-like ring. But sometimes it will have a braiding – for example, the diagonal embedding $G < G \times G$ recovers the finite group modular data of Section 6.2.4.

These approaches cannot reconstruct the full RCFT or VOA. To give a simple example, the VOA associated with any even self-dual lattice or the Moonshine module corresponds to the trivial subfactor $N = M$, where M is the unique hyperfinite type II$_1$ factor. The way to get more information uses nets of subfactors.

There are two standard axiomatisations of quantum field theory (Section 4.2.4). The Wightman axioms, applied to two-dimensional CFT, yield quite naturally a VOA (see chapter 1 of [**330**]). Algebraic quantum field theory [**269**], on the other hand, leads to

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

subfactors. In particular, to any open set $\mathcal{O}$ in Minkowski space $\mathbb{R}^{1,1}$ we are to assign a von Neumann algebra $\mathcal{A}(\mathcal{O}) \subset \mathcal{L}(\mathcal{H})$ of observables localised to $\mathcal{O}$, obeying various properties (such as $\mathcal{O}_1 \subset \mathcal{O}_2$ implies $\mathcal{A}(\mathcal{O}_1) \subset \mathcal{A}(\mathcal{O}_2)$). The axioms imply these $\mathcal{A}(\mathcal{O})$ will all be type III$_1$ factors. In two dimensions, choosing 'light-cone' coordinates $x_0 \pm x_1$, we can take these $\mathcal{O}$ to be the product $\mathcal{I} \times \mathcal{J}$ of open intervals $\mathcal{I}, \mathcal{J} \subset \mathbb{R}$. This means that for most purposes the theory decomposes into a one-dimensional net $\mathcal{A}(\mathcal{I})$ – the chiral theory. The one-dimensional 'space-time' $\mathbb{R}$ is compactified to S^1, and requiring the theory to be covariant with respect to Diff(S^1), the result is called a *local conformal net*. The theory of these one-dimensional nets should be equivalent to that of VOAs, and that of the two-dimensional ones to the full RCFT, though most details of this equivalence are still to be established. Nevertheless, some aspects of the theory will likely remain much more accessible using, for example, subfactors than VOAs (in particular, orbifolds seem simpler in subfactor theory). For references and results, see, for example, [**341**], [**340**], [**568**], [**332**] and references therein.

Question 6.2.1. Prove equation (6.2.3a).

Question 6.2.2. Find all NIM-reps for $A_1{}^{(1)}$ at each level $k = 1, 2, 3, \ldots$ (*Hint*: Verify that the Perron–Frobenius eigenvalue of $\mathcal{M}_1$ is $S_{10}/S_{00} = 2\cos(\pi/(k+2)) < 2$.)

Question 6.2.3. (a) Find a continuous one-parameter deformation of the three-dimensional complex Lie algebra span$\{x, y, z\}$ with brackets $[xy] = x, [xz] = [yz] = 0$.
(b) Verify that any continuous deformation of A_1 is trivial.

Question 6.2.4. Let M, N be left A-modules, where A is a Hopf algebra. Prove that $\mathrm{Hom}_K(M, N)$ is a left A-module.

Question 6.2.5. (a) When does the character table of a finite group, with rows and columns appropriately normalised and ordered, equal the S-matrix of modular data?
(b) Let G be finite and abelian. Is the fusion ring for the quantum double $\mathcal{D}(G)$ (see Section 6.2.4) isomorphic to the group ring of $G \times G$?

Question 6.2.6. Let G be any finite group and consider the modular data of (6.2.12). Find the conjugation C, the simple-currents J and their action and monodromy φ_J, and identify the group of all simple-currents. Identify the Galois action and parities.

Question 6.2.7. Prove that any finite group can be realised as a subgroup of the group of automorphisms of a holomorphic VOA. (*Hint*: think of self-dual lattices.)

Question 6.2.8. Identify the knot group $\pi_1(\mathbb{R}^3 \setminus T)$ of the trefoil, using the Wirtinger presentation of Figure 6.8.

Question 6.2.9. Prove, using the Reidemeister moves, that the Wirtinger presentation yields the same group no matter which knot diagram is chosen for the given knot.

Question 6.2.10. Recall (6.2.15). Find all values a, b such that $\sigma_i \mapsto ae_i + b$, $i = 1, \ldots, n-1$, yields a representation of the braid group $\mathcal{B}_n$ in $\mathcal{A}_{n,\tau}$.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

6.3 Hints of things to come

String theory has profoundly affected geometry (e.g. elliptic genus and mirror symmetry), algebra (e.g. VOAs) and topology (e.g. knot invariants), but so far it has had little impact on number theory. That may have something to do with the knowledge and interests of the individuals who have developed its mathematical side. There are in fact several indications of deep relations with number theory, waiting to be developed. In this section we sketch some of these.

6.3.1 Higher-genus considerations

String theory tells us that CFT can live on any surface Σ. The VOAs, including the geometric VOAs of Section 5.4.1, capture CFT in genus 0. The graded dimensions and traces considered above concern CFT quantities ('chiral blocks') at genus 1: $\tau \mapsto e^{2\pi i\tau}$ maps $\mathbb{H}$ onto a cylinder, and the trace identifies the two ends. But there are analogues of all this at higher genus [**573**] (though the formulae can rapidly become awkward). We have alluded to this throughout the book so will only add some quick remarks here. Our main point is that this is surely the direction for important future research, with direct implications to Moonshine.

For example, the graded dimension of the $V^\natural$ CFT in genus 2 is computed in [**533**], and involves, for example, Siegel theta functions. The higher-genus mapping class group representations coming from the $A_1{}^{(1)}$ RCFT are studied in [**220**]. A more radical suggestion, using projective limits, is given in Section 4.3.3.

The orbifold theory in Sections 5.3.6 and 7.3.2 is genus 1: each sector (g, h) corresponds to a homomorphism from the fundamental group $\mathbb{Z}^2$ of the torus into the orbifold group G (e.g. $G = \mathbb{M}$) – g and h are the targets of the two generators of $\mathbb{Z}^2$ and so must commute. More generally, the sectors correspond to homomorphisms $\varphi : \pi_1(\Sigma) \to G$, and for each we get a higher-genus trace $\mathcal{Z}(\varphi)$, which are functions on the Teichmüller space $\mathfrak{T}_g$ (generalising the upper half-plane $\mathbb{H}$ for genus 1). The action (7.3.3) of $\mathrm{SL}_2(\mathbb{Z})$ on $N_{(g,h)}$ generalises to the action of the mapping class group on $\pi_1(\Sigma)$ and $\mathfrak{T}_g$.

For example, we can count the number of inequivalent homomorphisms $\pi_1(\Sigma) \to G$, for G a compact genus-g surface. This number is given by Verlinde's formula (6.1.2) together with the expression (6.2.12a) [**194**]:

$$\mathcal{N}^{(g,0)} = \sum_h \sum_{\mathrm{ch}\in\mathrm{Irr}(C_G(h))} \left(\frac{\|C_G(h)\|}{\mathrm{ch}(e)}\right)^{2(g-1)}, \tag{6.3.1}$$

where we sum over representatives h of the various conjugacy classes of G.

6.3.2 Complex multiplication and Fermat

A few years ago Philippe Ruelle was walking in a library in Dublin. He spotted a yellow book in the mathematics section, called *Complex Multiplication* [**367**]. A strange title for a book by Lang! Ruelle flipped it to a random page, which turned out to be 26. There

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

he found what we would call the Galois selection rule (6.1.15a) for ${A_2}^{(1)}$, analysed and solved for the cases where $k+3$ is coprime to 6. Lang, however, knew nothing of modular invariants; he was reviewing work by Koblitz–Rohrlich [**351**] on decomposing the Jacobians of the Fermat curve $x^n + y^n + z^n = 0$ into their prime pieces, called 'simple factors'.

Fix $n > 3$. Let F_n denote the nth Fermat curve, that is the projective complex curve $x^n + y^n + z^n = 0$. We will describe some similarities with the modular data of ${A_2}^{(1)}$ at level $k = n - 3$.

First, let's review some ${A_2}^{(1)}$ chiral data. Call a pair $(r, s) \in \mathbb{N} \times \mathbb{N}$ *admissible* if $1 \le r, s$ and $r + s < n$. The integrable highest weights $\lambda \in P_+^k(A_2^{(1)})$ are in one-to-one correspondence with the admissible pairs, given by $\lambda_{(r,s)} := (n - r - s - 1)\omega_0 + (r - 1)\omega_1 + (s - 1)\omega_2$. For any admissible (r, s), define

$$H_{r,s} = \{\ell \in \mathbb{Z}_n^\times \mid \langle \ell r\rangle + \langle \ell s\rangle < n\},$$

where $\mathbb{Z}_N^\times$ is (as always) the multiplicative group (mod N) of integers coprime to N, and $\langle a\rangle$ is the unique integer $0 \le \langle a\rangle < n$ congruent to a (mod n). Then $\mathbb{Z}_{3n}^\times$ is the Galois group over $\mathbb{Q}$ of the field generated by all entries $S_{\lambda\mu}$ of the ${A_2}^{(1)}$ level-k matrix S. The Galois selection rule (6.1.15a) says that if $\mathcal{Z}$ is a modular invariant, then

$$\mathcal{Z}_{\lambda_{(r,s)},\lambda_{(r',s')}} \neq 0 \qquad \Rightarrow \qquad H_{r,s} = H_{r',s'}.$$

The hard part of the ${A_2}^{(1)}$ modular invariant classification involves solving this condition $H_{r,s} = H_{r',s'}$ [**232**].

Before we compare this to F_n, let's introduce some geometric terminology. An *abelian variety* is a torus of the form $\mathbb{C}^m/L$, where L is a $2m$-dimensional lattice in $\mathbb{C}^m$, which admits an embedding into projective space. This means there is a Hermitian form on $\mathbb{C}^m$ (defined in Section 1.1.3), whose imaginary part takes integer values when restricted to L. Most tori (when $m > 1$) don't satisfy this Hermitian form condition, though it is automatic when $m = 1$. We say two abelian varieties $\mathbb{C}^m/L$ and $\mathbb{C}^m/L'$ are *isogenous* if there exists a continuous group homomorphism from one to the other that is surjective; equivalently, if there is an invertible complex-linear endomorphism of $\mathbb{C}^m$ taking the lattice L onto a sublattice of L'. Isogeny is an equivalence relation preserving most things of interest.

Now suppose an abelian variety $\mathbb{C}^m/L$ contains another, $\mathbb{C}^n/L'$, of dimension $n < m$. Then the Hermitian form can be used to show that the original variety is isogenous to the product of $\mathbb{C}^n/L'$ with some $\mathbb{C}^{m-n}/L''$ (roughly, L'' is the orthogonal complement of L' in L). Continuing in this way, we get that any abelian variety is isogenous to the product of simple factors, where *simple factor* means an abelian variety containing no proper abelian subvariety.

A very special property that an abelian variety may possess is *complex multiplication*. The general definition is a little too complicated to get into here (see chapter 1.4 of [**367**]), so let's restrict to one-dimensional abelian varieties, that is the torus $A_\tau = \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$. We say A_τ has complex multiplication if its endomorphism ring $\mathrm{End}(A_\tau)$ is strictly greater than $\mathbb{Z}$; equivalently, if there is a non-integer $z \in \mathbb{C}$ such that

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

$z(\mathbb{Z}+\tau\mathbb{Z}) \subset \mathbb{Z}+\tau\mathcal{Z}$ (hence the name). It turns out that if A_τ has complex multiplication, then (among other things) $j(\tau)$ is an algebraic integer. This illustrates just how rare complex multiplication is: only countably many A_τ have it. It also illustrates its number-theoretic significance, which only becomes more profound as the dimension rises.

We get an abelian variety from any complex projective curve, by taking the Jacobian (Section 2.1.4), which is of complex dimension equal to the genus. In the case of the Fermat curve F_n, the genus is $\binom{n-1}{2}$, which equals the cardinality $\|P_+^k(A_2^{(1)})\|$. A bijection between $P_+^k(A_2^{(1)})$ and a basis of holomorphic 1-forms is

$$\lambda_{(r,s)} \leftrightarrow \omega_{(r,s)} := x^{r-1}y^{s-1}\frac{\mathrm{d}x}{y^{n-1}},$$

for any admissible (r, s). For each (r, s) let $[r, s]$ denote the $H_{r,s}$-orbit $\{(\langle \ell r\rangle, \langle \ell s\rangle)\}_{\ell\in H_{r,s}}$. Then the Jacobian $\mathrm{Jac}(F_n)$ is isogenous to the product, over all orbits $[r, s]$, of a $\|\mathbb{Z}_m^\times\|/2$-dimensional abelian variety $A_{[r,s]}$, for $m = n/\gcd(r, s, n-r-s)$. All $A_{[r,s]}$ have complex multiplication, which simplifies the following analysis.

We wish to decompose $\mathrm{Jac}(F_n)$ into a product of simple factors. Thus we need to know when the $A_{[r,s]}$ are isogenous to one another, and also when they are simple. Both questions reduce to knowing when $H_{r,s} = H_{r',s'}$, which as we mentioned earlier is also the key step in the ${A_2}^{(1)}$ modular invariant classification.

Similarly, Itzykson discovered traces of the ${A_2}^{(1)}$ exceptionals – these occur when $k + 3 = 8, 12, 24$ – in the Jacobian of F_{24}. See [**46**] for additional observations.

The point is that the combinatorial heart of two very different problems – the decomposition of the Jacobian of Fermat curves into simple factors, and the classification of RCFT associated with $A_2^{(1)}$ – are identical. Nevertheless, this must seem a little ad hoc. What is needed are other independent probes of this (still hypothetical) relationship. One possibility, suggested by the presence of complex multiplication, is the following.

Basic data associated with an algebraic variety V is its zeta-function $L(V, s)$, which counts its points over various finite fields. Isogenous varieties have equal zeta-functions. The Mellin transform of the zeta-function (Section 2.3.1) formally gives a q-series $f_V(\tau) = \sum_n a_n q^n$. For a typical variety V, f_V won't have any special properties, but when V has complex multiplication, the zeta-function decomposes into a product of Hecke L-functions, and their q-series do have modularity properties [**505**], [**506**].

Thus, associated with the abelian varieties $A_{[r,s]}$ – by virtue of complex multiplication – are various sorts of modular forms. And associated with the weights $\lambda_{(r,s)}$ – by virtue of being integrable highest weights of an affine algebra – are various sorts of modular forms.

Problem *How are the modular forms associated with the zeta-functions of the factors $A_{[rs]}$ in the Jacobian of the Fermat curve F_n related to the modular forms associated with integrable highest-weight modules of ${A_2}^{(1)}$ at level $n-3$?*

The easiest n to check will be $n = 4, 6, 8, 12$, since for them $\mathrm{Jac}(F_n)$ is isogenous to a product of elliptic curves. A somewhat related project, concerning ${A_1}^{(1)}$, is proposed in [**490**], though nothing definite has been achieved there yet.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

In any case, these Fermat ↔ ${A_2}^{(1)}$ 'coincidences' are still not understood. It is tempting to guess that, more generally, the ${A_r}^{(1)}$ level-k modular invariant classification is somehow related to the hypersurface $x_1^n + \cdots + x_r^n = z^n$, for $n = k + r + 1$, but this is probably too naive. As with other meta-patterns, the most realistic hope wouldn't be to find a direct connection between Fermat curves and the RCFTs associated with $\mathfrak{sl}_3$. Rather, the idea is to identify the combinatorial nugget common to both. The real hope would be that this 'coincidence' lies in a series: *A–D–E* for $\mathfrak{sl}_2$, Fermat for $\mathfrak{sl}_3, \ldots,$ and that this would lead to insights into $\mathfrak{sl}_4$ RCFT and beyond.

Complex multiplication in CFT has been the subject of other work – see [**435**] for several references. Let's mention two examples. Arithmetic varieties related to number fields seem to be naturally selected in the study of black holes in Calabi–Yau compactifications of string theory [**435**]. It has been conjectured [**268**] that superconformal field theory with target space given by a Calabi–Yau manifold M will be rational iff both M and its mirror have complex multiplication.

6.3.3 Braided # 6: the absolute Galois group

The absolute Galois group of the rationals is the group of symmetries of the field of algebraic numbers. It is the most important, and poorest understood, group in algebraic number theory. But it also has deep contacts with geometry (through the generalised Riemann existence theorem), and there have been several proposals conjecturing its relevance to RCFT (see e.g. [**128**], [**435**], [**268**] and references therein), and even quantum field theory [**106**], [**93**].

Recall the discussion of algebraic numbers and Galois groups in Section 1.7. The algebraic closure $\overline{\mathbb{Q}}$ of the rationals is the set of all algebraic numbers, or equivalently the union of all finite-dimensional field extensions of $\mathbb{Q}$. The absolute Galois group of $\mathbb{Q}$ is $\Gamma_{\mathbb{Q}} := \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. It's uncountably infinite, and extremely complicated. Only two of its elements have names: the identity and complex conjugation. If $\mathbb{K}$ is any finite Galois extension of $\mathbb{Q}$, then its Galois group $G = \mathrm{Gal}(\mathbb{K}/\mathbb{Q})$, which will be a finite group, is a homomorphic image of $\Gamma_{\mathbb{Q}}$ and so is a quotient $\Gamma_{\mathbb{Q}}/N$ of $\Gamma_{\mathbb{Q}}$. Much effort has been devoted to discovering which groups G can arise as Galois groups over $\mathbb{Q}$ (see [**548**] for a review of the so-called inverse Galois problem).

Conjecture 6.3.1 *Any finite group G is a quotient of* $\Gamma_{\mathbb{Q}}$.

This conjecture shows just how complicated $\Gamma_{\mathbb{Q}}$ is. Incidentally, there are many nontrivial points of contact between braid groups and inverse Galois theory (see e.g. [**549**]).

$\Gamma_{\mathbb{Q}}$ is an example of a *profinite* group, that is a projective limit of finite groups (here, of the Galois groups G). We define projective limit in Section 2.4.1 – the indexing set here are the fields $\mathbb{K}$, ordered by inclusion, to which is attached its Galois group G. This just means that $\sigma \in \Gamma_{\mathbb{Q}}$ consists of a choice of Galois automorphism $\sigma_{\mathbb{K}}$ for each finite extension $\mathbb{K} \supseteq \mathbb{Q}$, which obeys the obvious compatibility constraint (if $\mathbb{K} \subset \mathbb{L}$, then $\sigma_{\mathbb{L}}$ restricted to $\mathbb{K}$ must equal $\sigma_{\mathbb{K}}$). Thus, if the conjecture is true, $\Gamma_{\mathbb{Q}}$ would be the limit

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

$\lim_{\leftarrow} G$ of all finite groups, in this sense. Of course any finite group is also a quotient of some free group $\mathcal{F}_n$, and so we may wonder if $\Gamma_{\mathbb{Q}}$ and $\mathcal{F}_n$ are somehow related.

Thanks to their realisations as fundamental groups, the braid group $\mathcal{B}_n$ acts faithfully on $\mathcal{F}_n$ (Question 6.3.5) – in other words, $\mathcal{B}_n$ can be regarded as a subgroup of Aut($\mathcal{F}_n$). This can be seen as follows. Recall the space $\mathfrak{C}_n$ of (1.2.6). We have the obvious projection $\pi : \mathfrak{C}_{n+1} \to \mathfrak{C}_n$, given by forgetting the $(n+1)$th point. Hence π induces an action of the fundamental group $\pi_1(\mathfrak{C}_n)$ of the base on the fundamental group of the fibre $\pi^{-1}(z_1, \ldots, z_n) = \mathbb{C} \setminus \{z_1, \ldots, z_n\}$, that is an action of the pure braid group $\mathcal{P}_n$ on $\mathcal{F}_n$. The action of $\mathcal{B}_n$ is obtained similarly. We will find that similar reasoning allows us to replace $\mathcal{B}_n$ by $\Gamma_{\mathbb{Q}}$, and $\mathcal{F}_n$ by its profinite completion.

Let X be an algebraic variety defined over $\mathbb{Q}$ – that is, X is defined as the set of solutions $(z_1, \ldots, z_n) \in \mathbb{C}^n$ to a collection of polynomials $p_i(z_1, \ldots, z_n) = 0$, and the polynomials have coefficients in $\mathbb{Q}$. Let $X(\mathbb{Q})$ be the set of points $(z_1, \ldots, z_n) \in X$ with all coordinates $z_i \in \mathbb{Q}$. Fix a base-point $p \in X(\mathbb{Q})$ (assuming one exists).

Let N be a finite-index normal subgroup of $\pi_1(X, p)$. Then by the geometric Galois correspondence (Section 1.7.2), N corresponds to a finite Galois cover $f_N : X_N \to X$ of X, with $\pi_1(X_N) \cong N$ and the quotient $\pi_1(X, p)/N$ can be identified with the set of homeomorphisms $\gamma : X_N \to X_N$ satisfying $f_N \circ \gamma = f_N$. Each γ, restricted to the finite set $f_N^{-1}(p)$, will be a permutation, and this permutation uniquely determines it.

By the *generalised Riemann existence theorem* (Grauert–Remmert, 1958), each finite cover X_N of X is an algebraic variety defined over $\overline{\mathbb{Q}}$. Thus each automorphism $\sigma \in \Gamma_{\mathbb{Q}}$ permutes the finite covers of X (or if you prefer, the normal subgroups N): it acts on X_N by acting simultaneously on the coefficients of all the defining polynomials of X_N.

Grothendieck [**267**] explained that $\Gamma_{\mathbb{Q}}$ acts on the profinite completion $\widehat{\pi_1}(X, p)$ of the fundamental group of X, called the *algebraic fundamental group* of X. This means the following. The profinite completion $\widehat{G}$ of a group G is the projective limit $\lim_{\leftarrow} G/N$ over all finite quotients G/N (i.e. N runs over all normal subgroups of finite index in G). An element $g \in \widehat{G}$ consists of a choice $g_N N$ of coset in G/N for each such N, such that whenever N_1 is a subgroup of N_2 then $g_{N_1} N_2 = g_{N_2} N_2$. This should remind us of the construction of the p-adic integers $\widehat{\mathbb{Z}}_p$ – indeed, $\widehat{\mathbb{Z}} = \prod_p \widehat{\mathbb{Z}}_p$ is the profinite completion of $\mathbb{Z}$. Profinite completion is the algebraic analogue of the topological completion of a space by Cauchy sequences (as in the construction of $\mathbb{R}$ from $\mathbb{Q}$). Its purpose is the same: just as $\mathbb{R}$ fills in the 'gaps' in $\mathbb{Q}$, so does $\widehat{G}$ supply the missing elements in G. For example, $\sqrt{2}$ exists in $\widehat{\mathbb{Z}}_7$ but not in $\mathbb{Z}$. Of course, being a projective limit, the profinite completion is also an 'integration' of all G/N, that is a way of treating them all simultaneously. A solution in $\widehat{\mathbb{Z}}$ to a polynomial equation gives us simultaneously a solution modulo any n.

For example, $\widehat{\ell} \in \widehat{\mathbb{Z}}$ corresponds, for each $n \in \mathbb{N}$, to an integer $\widehat{\ell}_n$ defined modulo n, subject to the obvious compatibility condition. Then an element $\widehat{\ell}$ is invertible, written $\widehat{\ell} \in \widehat{\mathbb{Z}}^\times$, iff for each $n > 1$, $\widehat{\ell}_n$ is invertible mod n. Hence any $\widehat{\ell} \in \widehat{\mathbb{Z}}^\times$ has a well-defined action on finite-order roots of unity: given any nth root of unity ξ, $\xi^{\widehat{\ell}}$ is defined to be $\xi^{\widehat{\ell}_n}$. In fact, consider the field $\overline{\mathbb{Q}}^{ab}$ obtained by taking the union of all cyclotomic fields (or equivalently, by Theorem 1.7.1, all abelian extensions of $\mathbb{Q}$). Its Galois group Gal($\overline{\mathbb{Q}}^{ab}/\mathbb{Q}$) can be naturally identified with the multiplicative group $\widehat{\mathbb{Z}}^\times$ in this way. This is just the

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

action of $\Gamma_{\mathbb{Q}}$ restricted to cyclotomic fields – call this restriction the *cyclotomic character* $\chi^{cyclo} : \Gamma_{\mathbb{Q}} \to \widehat{\mathbb{Z}}^{\times}$ (this is a 'character' in the sense of a one-dimensional representation, not as a trace of a higher-dimensional character). This action has a large kernel – in fact, $\widehat{\mathbb{Z}}^{\times}$ is isomorphic to the abelianisation $\Gamma_{\mathbb{Q}}/[\Gamma_{\mathbb{Q}}\Gamma_{\mathbb{Q}}]$.

Let $\widehat{\gamma} \in \widehat{\pi_1}(X, p)$, that is for each finite-index normal subgroup N of $\pi_1(X, p)$, we have a coset representative $\widehat{\gamma}_N$ of some coset $\widehat{\gamma}_N N \in \pi_1(X, p)/N$ and these $\widehat{\gamma}_N$ – which we are to think of as permutations of finite sets $f_N^{-1}(p)$ – are compatible in the appropriate way. Then for any $\sigma \in \Gamma_{\mathbb{Q}}$ and $\widehat{\gamma} \in \pi_1(X, p)/N$, the action $\sigma.\widehat{\gamma}$ is defined by

$$(\sigma.\widehat{\gamma})_N = \sigma \circ \widehat{\gamma}_{\sigma^{-1}N} \circ \sigma^{-1}, \tag{6.3.2}$$

where σ acts on the points in $f_N^{-1}(p) \subset \overline{\mathbb{Q}}^n$ component-wise, and acts on the normal subgroups N as above. As we will see, choosing the variety X appropriately, (6.3.2) includes the profinite analogue of the braid group action on $\mathcal{F}_n$ mentioned earlier: the image of $\Gamma_{\mathbb{Q}}$ in Aut $\widehat{\mathcal{F}_n}$ lies in this image of $\widehat{\mathcal{B}}_n$. Equation (6.3.2) generalises to an action of $\Gamma_{\mathbb{Q}}$ on the fundamental groupoids $\pi_1(X, p, q)$ of (homotopy equivalence classes of) paths in X with endpoints $p, q \in X(\mathbb{Q})$.

Now, generically $\pi_1(X, p)$ is isomorphic to the mapping class group $\Gamma_{g,n}$, when X is a surface of genus g with n punctures. By the *modular tower* we mean the collection of moduli spaces $\mathfrak{M}_{g,n}$, where the different spaces are related by the obvious topological actions such as forgetting marked points, or sewing surfaces together ('tower' means a family of objects linked by homomorphisms). In Section 2 of his *Esquisse d'un Programme*, Grothendieck conjectured that $\Gamma_{\mathbb{Q}}$ acts on the profinite completion of this tower (i.e. on the profinite completion of all $\Gamma_{g,n}$, and respecting those topological actions), and is in fact the full automorphism group of this completion, and that this provides an effective, almost combinatorial, way to study $\Gamma_{\mathbb{Q}}$, not directly related to its action on algebraic numbers. He conjectured that his profinite modular tower could be reconstructed from $\mathfrak{M}_{0,3}$, $\mathfrak{M}_{0,4}$, $\mathfrak{M}_{1,1}$, with all relations obtained from $\mathfrak{M}_{0,5}$ and $\mathfrak{M}_{1,2}$.

For example, the ordered moduli space $\mathfrak{M}_{0,4}$ is the thrice-punctured sphere $\mathbb{P}^1(\mathbb{C}) \setminus \{0, 1, \infty\}$ and can be defined over $\mathbb{Q}$ – indeed, it is just $\Gamma(2)\backslash\mathbb{H}$ and has defining equation $z_1 z_2^2 (z_2 - 1)^2 = z_2^2 - z_2 + 1$. Its fundamental group is $\mathcal{F}_2$, the free group on two generators. Therefore, $\Gamma_{\mathbb{Q}}$ acts on $\widehat{\mathcal{F}}_2$. In fact, this action is known to be faithful (Belyi, 1987), so $\Gamma_{\mathbb{Q}}$ is a subgroup of Aut $\widehat{\mathcal{F}}_2$. Similarly, we get an action of $\Gamma_{\mathbb{Q}}$ on $\widehat{\mathcal{B}}_n$, which we will give shortly. This action yields one on $\widehat{\mathcal{B}}_n/Z(\widehat{\mathcal{B}}_n)$, and for $n = 3$ the latter equals the completion $\widehat{\mathrm{PSL}_2(\mathbb{Z})}$ of the modular group (recall (1.1.10b)).

Does Moonshine (or if you prefer, RCFT or VOAs) see this same $\Gamma_{\mathbb{Q}}$-action? After all, modular data possesses a nice Galois action (6.1.7), as does the spectrum of the theory (6.1.15b). Also, Grothendieck's modular tower, with generators (0, 3), (0, 4), (1, 1) and relations (0, 5) and (1, 2), reminds one of the Moore–Seiberg data of Section 6.1.4. There are a few difficulties with this hope. For instance, we should take profinite limits of these actions – for example, lift our action on $\mathrm{SL}_2(\mathbb{Z})$ to one on $\widehat{\mathrm{SL}_2(\mathbb{Z})}$. Can that have any natural meaning to RCFT? Also, and most disappointingly, the modular data always lies in cyclotomic fields, so the Galois action (6.1.7) in RCFT really only sees the rather uninteresting action of the abelianisation $\Gamma_{\mathbb{Q}}/[\Gamma_{\mathbb{Q}}\Gamma_{\mathbb{Q}}] \cong \widehat{\mathbb{Z}}^{\times}$, as explained earlier.

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

The first difficulty is easy to address. Subject to Conjecture 6.1.7, we obtain the following universal actions of $\Gamma_{\mathbb{Q}}$ on modular data: for any $\sigma \in \Gamma_{\mathbb{Q}}$,

$$\sigma.T = T^{\chi^{cyclo}(\sigma)}, \tag{6.3.3a}$$

$$\sigma.S = T^{\chi^{cyclo}(\sigma)} S T^{\chi^{cyclo}(\sigma^{-1})} S T^{\chi^{cyclo}(\sigma)} S^2. \tag{6.3.3b}$$

In order for (6.3.3) to make sense, these equations must live in the profinite completion of $\mathrm{SL}_2(\mathbb{Z})$. This is the meaning of the profinite completions here: the 'integration' of the data of all RCFT (or VOAs) necessary for universal formulae. The generators S, T of $\mathrm{SL}_2(\mathbb{Z})$ also generate $\widehat{\mathrm{SL}_2(\mathbb{Z})}$, though in the topological sense (i.e. just as 1 topologically generates $\widehat{\mathbb{Z}}$). Since the action (6.3.3) is continuous, it defines a $\Gamma_{\mathbb{Q}}$-action on $\widehat{\mathrm{SL}_2(\mathbb{Z})}$. It is very natural, in the sense that there is a map $\Gamma_{\mathbb{Q}} \to \widehat{\mathrm{SL}_2(\mathbb{Z})}$ given by

$$\sigma \mapsto G_\sigma := T^{\chi^{cyclo}(\sigma)} S T^{\chi^{cyclo}(\sigma^{-1})} S T^{\chi^{cyclo}(\sigma)} S = \begin{pmatrix} \chi(\sigma) & 0 \\ 0 & \chi(\sigma^{-1}) \end{pmatrix} \in \widehat{\mathrm{SL}_2(\mathbb{Z})}, \tag{6.3.4}$$

and $\sigma.S$ equals the matrix multiplication $G_\sigma S$. This map (6.3.4) is also what gives the Galois action (2.3.14) on modular functions for $\Gamma(N)$ or, in more suggestive language, the meromorphic functions on $\lim_{\leftarrow} \Gamma(N)\backslash\overline{\mathbb{H}}$ (see Section 2.4.1). Of course, in RCFT there is a preferred basis for this $\widehat{\mathrm{SL}_2(\mathbb{Z})}$-representation (namely, that given by the VOA characters), and in that basis the matrices become signed permutation matrices $\epsilon_\sigma(a)\,\delta_{a^\sigma,b}$. It will be extremely interesting to find universal formulae for the Galois action on the remaining Moore–Seiberg data. The difficulty is that, in obtaining (6.3.3), we were guided by the presence of a preferred basis, and so (6.3.3) reduces to the usual Galois action on the corresponding matrices. For the braiding and fusing matrices, typically there isn't a preferred basis, and so other principles must be our guide.

Why do cyclotomic fields exhaust RCFT, hence demanding that the RCFT Galois action, unlike that on Grothendieck's modular tower, be far from faithful? Is it trying to tell us something? What other principles can guide us to a Galois action on the remaining Moore–Seiberg data?

Those questions lead us to Drinfel'd [**161**]. Recall from Section 1.6.2 that the pure braid group $\mathcal{P}_n$ acts on each set $\mathrm{Hom} A_1 \oplus \cdots \oplus A_n, V)$ in any braided monoidal category. In particular, we can ask which subgroup of $\mathcal{P}_3 \times \mathcal{P}_2$ acts on the set of all braided monoidal categories, where $\beta \in \mathcal{P}_3$ and $\gamma \in \mathcal{P}_2$ send the associativity constraint $a : (A \otimes B) \otimes C \to A \otimes (B \otimes C)$ and the commutativity constraint $c : A \otimes B \to B \otimes A$, respectively, of one such category to that of another. We require that $\beta.a$ and $\gamma.c$ satisfy the various axioms, most importantly the pentagon and hexagon equations.

Dualising this, Drinfel'd suggested to act with $\mathcal{P}_3 \times \mathcal{P}_2$ on the data of *quasi-triangular quasi-Hopf algebras* A (defined in e.g. [**98**]). These algebras are co-commutative up to conjugation by the R-matrix $\mathcal{R} \in A \otimes A$ (as in Definition 6.2.8), and co-associative up to conjugation by the *associator* $\Phi \in A \otimes A \otimes A$ (Φ measures how A fails to be Hopf). Φ and $\mathcal{R}$ are required to obey the triangle, pentagon and hexagon equations of Section 1.6.2. We met quasi-triangular Hopf algebras in Definition 6.2.8; it will be clear shortly why Drinfel'd prefers quasi-Hopf algebras. Identify $\mathcal{P}_2$ with $\mathbb{Z}$ and $\mathcal{P}_3$ with $\mathcal{F}_2 \times \mathbb{Z}$ (1.1.10c);

then $m \in \mathcal{P}_2$ acts on the R-matrix by $m.\mathcal{R} = \mathcal{R}.(\mathcal{R}_{21}\mathcal{R})^m$ and, for example, a word $f(x, y) \in \mathcal{F}_2 < \mathcal{P}_3$ acts on the associator by $f.\Phi = f(\mathcal{R}_{21}\mathcal{R}_{12}, \Phi\mathcal{R}_{32}\mathcal{R}_{23}\Phi^{-1})^{-1}\Phi$. The other quantities in the algebra A are left unchanged. Unfortunately, this nice idea fails: only the two elements $(\pm 1, 1) \in \mathcal{P}_2 \times \mathcal{P}_3$ satisfy the constraints and thus permute quasi-triangular quasi-Hopf algebras (the nontrivial one sending $\mathcal{R}$ to $\mathcal{R}_{21}$ and fixing everything else).

Drinfel'd then proposed that there would be more solutions if we take profinite completions (indeed, this is a *raison d'être* of completions), so in place of $\mathcal{P}_2 \cong \mathbb{Z}$ and $\mathcal{P}_3 \cong \mathbb{Z} \times \mathcal{F}_2$ we take $\widehat{\mathcal{P}_2} \cong \widehat{\mathbb{Z}}$ and $\widehat{\mathcal{P}_3} \cong \widehat{\mathbb{Z}} \times \widehat{\mathcal{F}_2}$. To get these profinite actions on the $\mathcal{R}$ and Φ, it suffices to take the scalars of the algebras A to be formal power series $\widehat{\mathbb{Q}}[[h]]$ rather than $\mathbb{C}$. The hope is that by completing the groups, there is more chance of nontrivial solutions to the triangle, pentagon and hexagon equations. The details would take us too far afield, but the result is that there are indeed several solutions.

Drinfel'd was interested in this because, in an earlier paper, he had found, for each choice of simple Lie algebra $\mathfrak{g}$, a universal formula for one solution $(\Phi, \mathcal{R})$ to those equations, using Kohno's monodromy theorem for the KZ connection. Unfortunately this formula for Φ is quite complicated. In [**161**] he investigates two commuting actions on the set of all solutions $(\Phi, \mathcal{R})$, which he uses to deduce the existence of a simpler solution (see [**39**]). One of these actions was this pure braid group action.

Let $\widehat{GT}$, the *Grothendieck–Teichmüller group*, be the group of all pairs $(\lambda, f) \in \widehat{\mathbb{Z}} \times \widehat{\mathcal{F}_2}$ (the $\widehat{\mathbb{Z}}$ of $\widehat{\mathcal{P}}_3$ can't contribute) satisfying those equations and thus permuting those quasi-triangular quasi-Hopf algebras. $\widehat{GT}$ is large, in fact as we will see $\Gamma_{\mathbb{Q}}$ embeds as a subgroup in it. Drinfel'd conjectured that $\widehat{GT}$ should act on the profinite completion of Grothendieck's tower. For example, on $\widehat{\mathcal{B}_n}$, topologically generated as we know by $\sigma_1, \ldots, \sigma_{n-1}$, we get the action by $(\lambda, f) \in \widehat{GT}$ given by

$$(\lambda, f).\sigma_i = f\big(y_i, \sigma_i^2\big)^{-1} \sigma_i^{\lambda} f\big(y_i, \sigma_i^2\big), \tag{6.3.5a}$$

$$(\lambda, f).Z = Z^{\lambda}, \tag{6.3.5b}$$

where $Z = (\sigma_{n-1} \cdots \sigma_1)^n$ topologically generates the centre of $\widehat{\mathcal{B}_n}$ (just as it does that of $\mathcal{B}_n$) and $y_i = \sigma_{i-1} \cdots \sigma_1^2 \cdots \sigma_{i-1}$. This element y_i arises in presentations of the genus-0 mapping class groups $\Gamma_{0,n}$ or braid groups of the sphere [**59**]. The 'profinite word' $f(y_i, \sigma_i^2) \in \widehat{\mathcal{F}_2}$ means the value $\varphi(f)$ of the homomorphism $\varphi : \widehat{\mathcal{F}_2} \to \widehat{\mathcal{B}_n}$ defined by $\varphi(x) = y_i$ and $\varphi(y) = \sigma_i^2$.

Moreover, $\Gamma_{\mathbb{Q}}$ maps injectively into $\widehat{GT}$ and so can be identified with some subgroup of $\widehat{GT}$. Conjecturally, $\Gamma_{\mathbb{Q}}$ equals $\widehat{GT}$. For example, $(-1, 1)$ corresponds to complex-conjugation. See [**305**], [**494**], [**493**], [**39**] and section 16.4 of [**98**] for reviews of $\widehat{GT}$ and its action on, for example, the modular tower; [**128**] speculates on its relation to RCFT.

This is brought one step closer to RCFT by Kassel–Turaev [**339**]. It is relatively straightforward to extend Drinfel'd's action to certain braided monoidal categories. In [**339**] a 'pro-unipotent completion' $\widehat{\mathbf{R}}$ is defined for any ribbon category $\mathbf{R}$. $\widehat{\mathbf{R}}$ is itself a ribbon category, with the same objects as $\mathbf{R}$, but with each $\mathrm{Hom}(A, B)$ replaced by some projective limit of its linearisation over $\widehat{\mathbb{Q}} = \prod_p \widehat{\mathbb{Q}}_p$. For example, for the choice

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press

$\mathbf{R} = \mathbf{ribbon}$, Hom$(\emptyset, \emptyset)$ can be identified with the space of formal finite linear combinations over $\widehat{\mathbb{Q}}$ of framed oriented links in $\mathbb{R}^3$. Drinfel'd's work yields an action of $\Gamma_{\mathbb{Q}}$ on the collection of these ribbon categories.

This category $\widehat{\mathbf{ribbon}}$ obeys a universality property as in Theorem 1.6.2. Now, any automorphism $\sigma \in \Gamma_{\mathbb{Q}}$ acts on the data of $\widehat{\mathbf{ribbon}}$ to produce a new ribbon category $\widehat{\mathbf{ribbon}}^{\sigma}$. Its objects and Hom$(\emptyset, \emptyset)$ are unchanged. By universality, there is a functor from $\widehat{\mathbf{ribbon}}$ to $\widehat{\mathbf{ribbon}}^{\sigma}$, sending Hom$(\emptyset, \emptyset)$ to itself. That is, we get an action of $\Gamma_{\mathbb{Q}}$ on the $\widehat{\mathbb{Q}}$-span of links: a (framed oriented) link L is taken to some linear combination (over $\widehat{\mathbb{Q}}$) of links. $\Gamma_{\mathbb{Q}}$ also acts on related spaces, such as $\widehat{\mathbb{Q}}$-valued Vassiliev invariants [**339**].

For example, complex-conjugation sends a link L to its mirror reflection (in general a link is not isotopic to its mirror reflection – see footnote 6 in chapter 1). However, [**339**] show that this $\Gamma_{\mathbb{Q}}$ action is trivial on the commutator $[\Gamma_{\mathbb{Q}}\Gamma_{\mathbb{Q}}]$, and thus really is an action of $\widehat{\mathbb{Z}}^{\times}$.

This action is clearly very similar to that of RCFT. As we know, RCFT attaches the matrix S to the Hopf link (Figure 6.9). Complex-conjugation ($\lambda = -1 \in \widehat{\mathbb{Z}}^{\times}$) sends the Hopf link to its mirror image; the mirror image corresponds to $\overline{S}$, which is what (6.3.3b) reduces to for $\lambda = -1$.

Problem *Identify the relation between* [**339**] *and the action (6.3.3) in RCFT. Can this be used somehow to identify the Galois action on arbitrary Moore–Seiberg data?*

We conjecture these actions are identical or at least very close. After all, they both factor through to $\widehat{\mathbb{Z}}^{\times}$ and agree with complex-conjugation applied to the Hopf link. Theorem 4 of [**40**] should make it possible to compute the [**339**] action on the Hopf link for any $\lambda \in \widehat{\mathbb{Z}}^{\times}$, thus allowing us to compare it directly to (6.3.3a). As we've learned, there are topological underpinnings of chiral RCFT data (e.g. the modular categories of [**534**], [**32**]) as well as full RCFT (see e.g. [**211**]); this seems the obvious way to attack this problem.

At least as interesting as this Galois action on the Moore–Seiberg data is that we can also hope that $\Gamma_{\mathbb{Q}}$ (or at least $\widehat{\mathbb{Z}}^{\times}$) will act on the spaces $\mathfrak{B}^{(g,n)}$ of chiral blocks, since they do on $\mathfrak{B}^{(1,1)}$, i.e. on the characters, which are modular functions (recall Section 2.3.3).

The Galois action (6.3.3) of RCFT is not directly related to Grothendieck's (6.3.5). The RCFT action would seem to be intimately related to Congruence Property 6.1.7, so more relevant to RCFT than $\widehat{\mathrm{SL}_2(\mathbb{Z})}$ should be the much simpler $\lim_{\leftarrow}\mathrm{SL}_2(\mathbb{Z})/\,\Gamma(N) = \mathrm{SL}_2(\hat{\mathbb{Z}})$.

So far in this subsection we've only addressed CFT 'in the bulk'. What if anything does Galois do to, for example, D-branes? Indeed, an action persists in boundary RCFT, though it is more complicated [**235**]. In particular, *this Galois action will no longer be abelian* – the algebraic numbers involved belong to exponent-2 extensions of the cyclotomic field $\overline{\mathbb{Q}}^{ab}$. This complication opens the door to much more interesting mathematics.

It will be interesting to see if the $\hat{\mathbb{Z}}^{\times}$ action in [**106**] can be related to that of RCFT. We are to think of RCFT as being to generic quantum field theory what semi-simple finite-dimensional Lie algebras are to generic ones. In this spirit, this Galois action on RCFT, and its relation to $\Gamma_{\mathbb{Q}}$ and Grothendieck's *Esquisse*, can be regarded perhaps as a toy model for the much more ambitious Cosmic Galois Group of [**93**], which conjecturally underlies the multiple zeta values found by Kreimer and others in more physical quantum field theories.

As a final remark, it is quite possible that the Galois actions explored in this subsection are related to the Fermat remarks of last subsection (see in particular section II of [**304**]). The Fermat curve $F_N = \{x^N + y^N = 1\}$ $x^N + y^N = 1$ is an abelian cover of $\mathbb{P}^1(\mathbb{C}) \setminus \{0, 1, \infty\}$; in turn, its abelian covers are controlled by torsion points on its Jacobian $\mathrm{Jac}(F_N)$, and in [**304**] the action of $\Gamma_{\mathbb{Q}}$ on $\widehat{\mathcal{F}_2}$ is studied via those torsion points, with results somewhat reminiscent of Section 6.3.2.

Question 6.3.1. Use the fact that the S and T matrices of (6.1.8) define modular data to compute the sum $\sum_{m=1}^{n} e^{2\pi \mathrm{i} m^2/n}$. (*Note*: This is called a Gauss sum. A similar calculation yields a generalisation of Gauss sums for any modular data.)

Question 6.3.2. Find all $\tau \in \mathbb{H}$ such that the torus $\mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau)$ is isogenous to $\mathbb{C}/(\mathbb{Z} + \mathbb{Z}\mathrm{i})$.

Question 6.3.3. Prove that the elliptic curves $y^2 = x^3 + ax$ and $y^2 = x^3 + b$ both have complex multiplication for any a, b.

Question 6.3.4. What is the profinite completion $\widehat{G}$ for finite groups G?

Question 6.3.5. (a) Define $\sigma_i.x_j = x_j$ if $j \neq i, i+1$, and $\sigma_i.x_i = x_{i+1}$ and $\sigma_i.x_{i+1} = x_{i+1}^{-1} x_i x_{i+1}$. Verify that this is a well-defined action. (It turns out that this action is faithful.)
(b) Verify that for any $\beta \in \mathcal{B}_n$, β fixes $x_1 \cdots x_n$, and there is a permutation π_β and words $a_i \in \mathcal{F}_n$ such that $\beta.x_i = A_i x_{\pi_\beta} A_i^{-1}$. (It turns out that, conversely, any automorphism β obeying those two conditions must come from this braid group action. This gives a way to solve the word problem in $\mathcal{B}_n$.)

Question 6.3.6. Choosing X to be a sphere with two punctures, describe the associated $\Gamma_{\mathbb{Q}}$-action (6.3.2).

https://doi.org/10.1017/9781009401548.007 Published online by Cambridge University Press