Algebra & Number Theory

www.jant.org

See inside back cover or www.jant.org for submission instructions.

K3 surfaces with Picard number one and infinitely many rational points

Ronald van Luijk

In general, not much is known about the arithmetic of K3 surfaces. Once the geometric Picard number, which is the rank of the Néron–Severi group over an algebraic closure of the base field, is high enough, more structure is known and more can be said. However, until recently not a single explicit K3 surface was known to have geometric Picard number one. We give explicit examples of such surfaces over the rational numbers. This solves an old problem that has been attributed to Mumford. The examples we give also contain infinitely many rational points, thereby answering a question of Swinnerton-Dyer and Poonen.

1. Introduction

K3 surfaces are two-dimensional analogues of elliptic curves in the sense that their canonical sheaf is trivial. However, as opposed to elliptic curves, little is known about the arithmetic of K3 surfaces in general. It is for instance an open question if there exists a K3 surface X over a number field such that the set of rational points on X is neither empty, nor dense (which throughout this paper will always refer to the Zariski topology). We will answer a longstanding question regarding the Picard group of a K3 surface. The Picard group of any variety is the group of line bundles on it, up to isomorphism. For a K3 surface X over a field k this is a finitely generated free abelian group, the rank of which is called the Picard number of X. The Picard number of $\overline{X} = X \times_k \overline{k}$, where $\overline{k}$ denotes an algebraic closure of k, is called the geometric Picard number of X. We will give the first known examples of explicit K3 surfaces shown to have geometric Picard number 1.

Bogomolov and Tschinkel [2000] showed an interesting relation between the geometric Picard number of a K3 surface X over a number field K and the arithmetic of X. They proved that if the geometric Picard number is at least 2, then in most cases the rational points on X are potentially dense, which means that there exists a finite field extension L of K such that the set $X(L)$ of L-rational points is Zariski dense in X. However, it is not yet known whether there exists

any K3 surface over a number field and with geometric Picard number 1 on which the rational points are potentially dense. Neither do we know if there exists a K3 surface over a number field and with geometric Picard number 1 on which the rational points are *not* potentially dense!

In December 2002, at the AIM workshop on rational and integral points on higher-dimensional varieties in Palo Alto, Swinnerton-Dyer and Poonen asked a related question. They asked whether there exists a K3 surface over a number field and with Picard number 1 that contains infinitely many rational points. In this article we will show that such K3 surfaces do indeed exist. It follows from our main theorem.

Theorem 1.1. *In the moduli space of K3 surfaces polarized by a very ample divisor of degree* 4, *the set of points parametrizing surfaces defined over $\mathbb{Q}$ with geometric Picard number* 1 *and infinitely many rational points is Zariski dense.*

As important as this result is the strategy of its proof. It contains a new way of finding sharp bounds for the geometric Picard number of a surface. This new method is widely applicable. It is based on the older idea that the Néron–Severi group of a surface X defined over a number field injects into the Néron–Severi group of its reduction $X_{\mathfrak{p}}$ at a prime $\mathfrak{p}$ of good reduction. By the Tate conjecture (proven in many cases for K3 surfaces), the geometric Picard number of a K3 surface in positive characteristic is even, and therefore at least 2. We will lower this upper bound for the geometric Picard number of X by comparing the lattice structure on the geometric Néron–Severi group of the reduction of X at two different primes of good reduction. If these both have rank 2 and their discriminants do not differ by a square factor, then there is no 2-dimensional lattice that injects into both, and we may conclude that the geometric Picard number of X equals 1.

Note that a polarization of a K3 surface is a choice of an ample divisor H. The degree of such a polarization is H^2. A K3 surface polarized by a very ample divisor of degree 4 is a smooth quartic surface in $\mathbb{P}^3$. We will prove the main theorem by exhibiting an explicit family of quartic surfaces in $\mathbb{P}^3_{\mathbb{Q}}$ with geometric Picard number 1 and infinitely many rational points. Proving that these surfaces contain infinitely many rational points is the easy part. It is much harder to prove that the geometric Picard number of these surfaces equals 1. It has been known since Max Noether that a general hypersurface in $\mathbb{P}^3_{\mathbb{C}}$ of degree at least 4 has geometric Picard number 1. A modern proof of this fact is given in [Deligne and Katz 1973, Theorem XIX.1.2]. Despite this fact, it has been an old challenge, attributed to Mumford and disposed of in this article, to find even one explicit quartic surface, defined over a number field, of which the geometric Picard number equals 1. Deligne's result does not actually imply that such surfaces exist, as "general" means "up to a countable union of closed subsets of the moduli space." A priori, this could

exclude all surfaces defined over $\overline{\mathbb{Q}}$. Although they do not give explicit surfaces with geometric Picard number 1 over number fields either, Terasoma and Ellenberg have proved that they do exist.

Theorem 1.2 [Terasoma 1985]. *For any positive integers* $(n; a_1, \ldots, a_d)$ *not equal to* $(2; 3)$, $(n; 2)$, *or* $(n; 2, 2)$, *and with* n *even, there is a smooth complete intersection* X *over* $\mathbb{Q}$ *of dimension* n *defined by equations of degrees* $a_1, \ldots, a_d$ *such that the middle geometric Picard number of* X *is* 1.   □

Theorem 1.3 [Ellenberg 2004]. *For every even integer* d *there exists a number field* K *and a polarized K3 surface* X/K *of degree* d, *with geometric Picard number* 1.
                                                                                                              □

The proofs of Terasoma and Ellenberg are ineffective in the sense that they do not give explicit examples. In principle it might be possible to extend their methods to test whether a given explicit K3 surface has geometric Picard number 1. In practice however, it is an understatement to say that the amount of work involved is not encouraging. The explicit examples we will give to prove the main theorem also prove the case $(n; a_1, \ldots, a_d) = (2; 4)$ of Theorem 1.2 and the case $d = 4$ of Theorem 1.3.

Shioda did find explicit examples of surfaces with geometric Picard number 1. In fact, he has shown [1981] that for every prime $m \geq 5$ the surface in $\mathbb{P}^3$ given by

$$w^m + xy^{m-1} + yz^{m-1} + zx^{m-1} = 0$$

has geometric Picard number 1. However, for $m = 4$ this equation determines a K3 surface with maximal geometric Picard number 20, i.e., a singular K3 surface.

Before we prove the main theorem in Section 3, we will recall some definitions and results in Section 2.

2. Prerequisites

A *lattice* is a free $\mathbb{Z}$-module L of finite rank, endowed with a symmetric, bilinear, nondegenerate map $\langle \_\,, \_ \rangle \colon L \times L \to \mathbb{Q}$, called the *pairing* of the lattice. An *integral lattice* is a lattice with a $\mathbb{Z}$-valued pairing. A lattice L is called *even* if $\langle x, x \rangle \in 2\mathbb{Z}$ for every $x \in L$. A *sublattice* of L is a submodule L' of L, such that the induced bilinear pairing on L' is nondegenerate. The *Gram matrix* of a lattice L with respect to a given basis $x = (x_1, \ldots, x_n)$ is $I_x = (\langle x_i, x_j \rangle)_{i,j}$. The *discriminant*

of L is defined by disc $L = \det I_x$ for any basis x of L. For any sublattice L' of finite index in L we have disc $L' = [L : L']^2$ disc L. The image of disc L and disc L' in $\mathbb{Q}^*/\mathbb{Q}^{*2}$ is the discriminant of the inner product space $L_{\mathbb{Q}}$, where the inner product is induced by the pairing of L.

Let X be a smooth, projective, geometrically integral surface over a field k and set $\overline{X} = X \times_k \overline{k}$, where $\overline{k}$ denotes an algebraic closure of k. As mentioned in the introduction, the *Picard group* Pic X of X is the group of line bundles on X up to isomorphism, or equivalently, the group of divisor classes modulo linear equivalence. The divisor classes that become algebraically equivalent to 0 over $\overline{k}$ (see [Hartshorne 1977, exercise V.1.7]) form a subgroup $\text{Pic}^0 X$ of Pic X. The quotient is the *Néron–Severi group* $\text{NS}(X) = \text{Pic } X / \text{Pic}^0 X$, which is a finitely generated abelian group, see [Hartshorne 1977, exercise V.1.7–8], or [Milne 1980, Theorem V.3.25], for surfaces or [Grothendieck et al. 1971, exposé XIII, théorème 5.1] in general. The intersection pairing endows the group $\text{NS}(X)/\text{NS}(X)_{\text{tors}}$ with the structure of a lattice. Its rank is called the *Picard number* of X. The Picard number of $\overline{X}$ is called the *geometric Picard number* of X.

By definition a smooth, projective, geometrically integral surface X is a *K3 surface* if the canonical sheaf ω_X on X is trivial and $H^1(\overline{X}, \mathcal{O}_{\overline{X}}) = 0$. Examples of K3 surfaces are smooth quartic surfaces in $\mathbb{P}^3$. The Betti numbers of a K3 surface are $b_0 = 1$, $b_1 = 0$, $b_2 = 22$, $b_3 = 0$, and $b_4 = 1$.

Lemma 2.1. *If X is a K3 surface, then $\text{Pic}^0 X$ is trivial, the Néron–Severi group $\text{NS}(X) \cong \text{Pic } X$ is torsion free, and the intersection pairing on $\text{NS}(X)$ is even.*

Proof. See [Barth et al. 1984, p. 21 and Proposition VIII.3.2] for characteristic 0 and [Bombieri and Mumford 1977, Theorem 5] for positive characteristic.   □

For any scheme Z over $\mathbb{F}_q$, any prime $l \nmid q$, and any integer m, we will use the étale cohomological groups $H^i_{\text{ét}}(Z, \mathbb{Q}_l)$ and their Tate twists $H^i_{\text{ét}}(Z, \mathbb{Q}_l)(m)$ as defined in for instance [Tate 1965, p. 94], Proposition 2.2 describes the behavior of the Néron–Severi group under good reduction. Its corollary will be used to show that the geometric Picard number of a certain surface is equal to 1.

Proposition 2.2 [van Luijk 2007, Proposition 6.2]. *Let A be a discrete valuation ring of a number field L with residue field $k \cong \mathbb{F}_q$. Let S be an integral scheme with a morphism $S \to \text{Spec } A$ that is projective and smooth of relative dimension 2. Assume that the surfaces $\overline{S} = S_{\overline{L}}$ and $\tilde{S} = S_{\overline{k}}$ are integral. Let $l \nmid q$ be a prime number. Then there are natural injective homomorphisms*

$$\text{NS}(\overline{S}) \otimes \mathbb{Q}_l \hookrightarrow \text{NS}(\tilde{S}) \otimes \mathbb{Q}_l \hookrightarrow H^2_{\text{ét}}(\tilde{S}, \mathbb{Q}_l)(1) \tag{1}$$

of finite dimensional inner product spaces over $\mathbb{Q}_l$. The first injection is induced by a natural injection $\text{NS}(\overline{S}) \otimes \mathbb{Q} \hookrightarrow \text{NS}(\tilde{S}) \otimes \mathbb{Q}$. The second injection respects the Galois action of $G(\overline{k}/k)$.

Recall that for any scheme Z over $\mathbb{F}_q$ with $q = p^r$ and p prime, the *absolute Frobenius* $F_Z \colon Z \to Z$ of Z acts as the identity on points, and by $f \mapsto f^p$ on the structure sheaf. Set $\Phi_Z = F_Z^r$ and $\bar{Z} = Z \times \bar{\mathbb{F}}_q$. Let Φ_Z^* denote the automorphism on $H^2_{\text{ét}}(\bar{Z}, \mathbb{Q}_l)$ induced by $\Phi_Z \times 1$ acting on $Z \times \bar{\mathbb{F}}_q = \bar{Z}$.

Corollary 2.3. *With the notation as in [Proposition 2.2](), the ranks of $\mathrm{NS}(\tilde{S})$ and $\mathrm{NS}(\bar{S})$ are bounded from above by the number of eigenvalues λ of $\Phi_{S_k}^*$ for which λ/q is a root of unity, counted with multiplicity.*

Proof. By [Proposition 2.2]() any upper bound for the rank of $\mathrm{NS}(\tilde{S})$ is an upper bound for the rank of $\mathrm{NS}(\bar{S})$. Let σ denote the q-th power Frobenius map, i.e., the canonical topological generator of $G(\bar{k}/k)$. For any positive integer m, let σ^* and $\sigma^*(m)$ denote the automorphisms induced on $\mathrm{NS}(\tilde{S}) \otimes \mathbb{Q}_l$ and $H^2_{\text{ét}}(\tilde{S}, \mathbb{Q}_l)(m)$ respectively. As all divisor classes are defined over some finite extension of k, some power of Frobenius acts as the identity on $\mathrm{NS}(\tilde{S})$, so all eigenvalues of σ^* acting on $\mathrm{NS}(\tilde{S})$ are roots of unity. It follows from [Proposition 2.2]() that the rank of $\mathrm{NS}(\tilde{S})$ is bounded from above by the number of roots of $\sigma^*(1)$ that are a root of unity. As the eigenvalues of $\sigma^*(0)$ differ from those of $\sigma^*(1)$ by a factor of q, this equals the number of roots λ of $\sigma^*(0)$ for which λq is a root of unity. The Corollary follows from the fact that $\Phi_{S_k}^*$ acts on $H^2_{\text{ét}}(\bar{Z}, \mathbb{Q}_l)$ as the inverse of $\sigma^*(0)$. See also [van Luijk 2007, Corollary 6.3]. $\qquad\square$

Remark 2.4. Tate's conjecture [1965] states that the upper bound mentioned is actually equal to the rank of $\mathrm{NS}(\tilde{S})$. Tate's conjecture has been proven for ordinary K3 surfaces over fields of characteristic $p \geq 5$; see [Nygaard and Ogus 1985, Theorem 0.2].

To find the characteristic polynomial of Frobenius as in [Corollary 2.3](), we will compute the traces of powers of Frobenius and use Newton's identities, which for convenience we state here (see [Borwein and Erdélyi 1995, p. 5]):

Lemma 2.5 (Newton's identities). *Let V be a vector space of dimension n and T a linear operator on V. Let t_i denote the trace of T^i. Then the characteristic polynomial of T is equal to*

$$f_T(x) = \det(x \cdot \mathrm{Id} - T) = x^n + c_1 x^{n-1} + c_2 x^{n-2} + \cdots + c_n,$$

with the c_i given recursively by

$$c_1 = -t_1 \qquad \text{and} \qquad -k c_k = t_k + \sum_{i=1}^{k-1} c_i t_{k-i}.$$

3. Proof of the main theorem

First we will give a family of smooth quartic surfaces in $\mathbb{P}^3$ with Picard number 1. Let $R = \mathbb{Z}[x, y, z, w]$ be the homogeneous coordinate ring of $\mathbb{P}^3_{\mathbb{Z}}$. Throughout the rest of this article, for any homogeneous polynomial $h \in R$ of degree 4, let $\mathfrak{X}_h$ denote the scheme in $\mathbb{P}^3_{\mathbb{Z}}$ given by

$$wf_1 + 2zf_2 = 3g_1g_2 + 6h, \tag{2}$$

with $f_1, f_2, g_1, g_2 \in R$ equal to

$$
\begin{aligned}
f_1 = x^3 &- x^2y - x^2z + x^2w - xy^2 - xyz + 2xyw + xz^2 + 2xzw + y^3 \\
&+ y^2z - y^2w + yz^2 + yzw - yw^2 + z^2w + zw^2 + 2w^3,
\end{aligned}
$$
$$f_2 = xy^2 + xyz - xz^2 - yz^2 + z^3,$$
$$g_1 = z^2 + xy + yz,$$
$$g_2 = z^2 + xy.$$

Its base extensions to $\mathbb{Q}$ and $\overline{\mathbb{Q}}$ are denoted X_h and $\overline{X}_h$ respectively.

Theorem 3.1. *Let $h \in R$ be a homogeneous polynomial of degree* 4. *Then the quartic surface X_h is smooth over $\mathbb{Q}$ and has geometric Picard number* 1. *The Picard group* $\mathrm{Pic}\,\overline{X}_h$ *is generated by a hyperplane section.*

Proof. For $p = 2, 3$, let $X_p/\mathbb{F}_p$ denote the fiber of $\mathfrak{X}_h \to \mathrm{Spec}\,\mathbb{Z}$ over p. As they are independent of h, one easily checks that X_p is smooth over $\mathbb{F}_p$ for $p = 2, 3$. As the morphism $\mathfrak{X}_h \to \mathrm{Spec}\,\mathbb{Z}$ is flat and projective, it follows that the generic fiber X_h of $\mathfrak{X}_h \to \mathrm{Spec}\,\mathbb{Z}$ is smooth over $\mathbb{Q}$ as well; compare [Hartshorne 1977, exercise III.10.2].

We will first show that X_2 and X_3 have geometric Picard number 2. For $p = 2, 3$, let Φ_p denote the absolute Frobenius of X_p. Set $\overline{X}_p = X_p \times \overline{\mathbb{F}}_p$ and let $\Phi_p^*(i)$ denote the automorphism on $H^i_{\text{ét}}(\overline{X}_p, \mathbb{Q}_l)$ induced by $\Phi_p \times 1$ acting on $\overline{X}_p = X_p \times_{\mathbb{F}_p} \overline{\mathbb{F}}_p$. Then by Corollary 2.3 the geometric Picard number of X_p is bounded from above by the number of eigenvalues λ of $\Phi_p^*(2)$ for which λ/p is a root of unity. We will find the characteristic polynomial of $\Phi_p^*(2)$ from the traces of its powers. These traces we will compute with the Lefschetz formula

$$\#X_p(\mathbb{F}_{p^n}) = \sum_{i=0}^{4} (-1)^i \, \mathrm{Tr}(\Phi_p^*(i)^n), \tag{3}$$

for which see [Milne 1980, Theorem VI.12.3]. Since X_p is a smooth hypersurface in $\mathbb{P}^3$ of degree 4, it is a K3 surface and its Betti numbers are $b_0 = 1$, $b_1 = 0$, $b_2 = 22$, $b_3 = 0$, and $b_4 = 1$. It follows that $\mathrm{Tr}(\Phi_p^*(i)^n) = 0$ for $i = 1, 3$, and for $i = 0$ and $i = 4$ the automorphism $\Phi_p^*(i)^n$ has only one eigenvalue, which by the

Weil conjectures equals 1 and p^{2n} respectively (see [Deligne 1974, théorème 1.6]). From the Lefschetz formula (3) we conclude $\mathrm{Tr}(\Phi_p^*(2)^n) = \#X_p(\mathbb{F}_{p^n}) - p^{2n} - 1$. After counting points on X_p over $\mathbb{F}_{p^n}$ for $n = 1, \ldots, 11$, this allows us to compute the traces of the first 11 powers of $\Phi_p^*(2)$. With Lemma 2.5 we can then compute the first coefficients of the characteristic polynomial f_p of $\Phi_p^*(2)$, which has degree $b_2 = 22$. We write $f_p = x^{22} + c_1 x^{21} + \cdots + c_{22}$, which by construction is independent of the choice of h, and find this table:

p	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}
2	-3	-2	12	0	-32	64	-128	128	256	0	-2048
3	-5	-6	72	27	-891	0	9477	-4374	-78732	19683	708588

The Weil conjectures give a functional equation $p^{22} f_p(x) = \pm x^{22} f_p(p^2/x)$. As in our case (both for $p = 2$ and $p = 3$) the middle coefficient c_{11} of f_p is nonzero, the sign of the functional equation is positive. This functional equation allows us to compute the remaining coefficients of f_p.

If λ is a root of f_p then λ/p is a root of $\tilde{f}_p(x) = p^{-22} f_p(px)$. Hence, the number of roots of $\tilde{f}_p(x)$ that are also a root of unity gives an upper bound for the geometric Picard number of X_p. After full factorization, we find

$$\tilde{f}_2 = \tfrac{1}{2}(x-1)^2\left(2x^{20} + x^{19} - x^{18} + x^{16} + x^{14} + x^{11} + 2x^{10} + x^9 + x^6 + x^4 - x^2 + x + 2\right),$$

$$\tilde{f}_3 = \tfrac{1}{3}(x-1)^2\left(3x^{20} + x^{19} - 3x^{18} + x^{17} + 6x^{16} - 6x^{14} + x^{13} + 6x^{12} - x^{11}\right.$$
$$\left. - 7x^{10} - x^9 + 6x^8 + x^7 - 6x^6 + 6x^4 + x^3 - 3x^2 + x + 3\right).$$

When $p = 2, 3$, the roots of the irreducible factor of $\tilde{f}_p$ of degree 20 are not integral. Therefore these roots are not roots of unity and we conclude that $\tilde{f}_p$ has only two roots that are roots of unity, counted with multiplicities. By Corollary 2.3 this implies that the geometric Picard number of X_p is at most 2.

Note that besides the hyperplane section H, the surface X_2 also contains the conic C given by $w = g_2 = z^2 + xy = 0$. We have $H^2 = \deg X_2 = 4$ and $H \cdot C = \deg C = 2$. As the genus $g(C)$ of C equals 0 and the canonical divisor K on X_2 is trivial, the adjunction formula $2g(C) - 2 = C \cdot (C + K)$ yields $C^2 = -2$. Thus H and C generate a sublattice of $\mathrm{NS}(\bar{X}_2)$ with Gram matrix

$$\begin{pmatrix} 4 & 2 \\ 2 & -2 \end{pmatrix}.$$

We conclude that the inner product space $\mathrm{NS}(\bar{X}_2)_{\mathbb{Q}}$ has rank 2 and discriminant $-12 \in \mathbb{Q}^*/\mathbb{Q}^{*2}$. Similarly, X_3 contains the line L given by $w = z = 0$, also with

genus 0 and thus $L^2 = -2$. The hyperplane section on X_3 and L generate a sublattice of $\mathrm{NS}(\bar{X}_3)$ of rank 2 with Gram matrix

$$\begin{pmatrix} 4 & 1 \\ 1 & -2 \end{pmatrix}.$$

We conclude that the inner product space $\mathrm{NS}(\bar{X}_3)_{\mathbb{Q}}$ also has rank 2, and discriminant $-9 \in \mathbb{Q}^*/\mathbb{Q}^{*2}$.

Let ρ denote the geometric Picard number $\rho = \mathrm{rk}\, \mathrm{NS}(\bar{X}_h)$. It follows from Proposition 2.2 that there is an injection $\mathrm{NS}(\bar{X}_h)_{\mathbb{Q}} \hookrightarrow \mathrm{NS}(\bar{X}_p)_{\mathbb{Q}}$ of inner product spaces for $p = 2, 3$. Hence we get $\rho \leq 2$. If equality held, then both these injections would be isomorphisms and $\mathrm{NS}(\bar{X}_2)_{\mathbb{Q}}$ and $\mathrm{NS}(\bar{X}_3)_{\mathbb{Q}}$ would be isomorphic as inner product spaces. This is not the case because they have different discriminants. We conclude $\rho \leq 1$. As a hyperplane section H on X_h has self intersection $H^2 = 4 \neq 0$, we find $\rho = 1$. Since $\mathrm{NS}(\bar{X}_h)$ is a 1-dimensional even lattice (see Lemma 2.1), the discriminant of $\mathrm{NS}(\bar{X}_h)$ is even. The sublattice of finite index in $\mathrm{NS}(\bar{X}_h)$ generated by H gives

$$4 = \mathrm{disc}\langle H \rangle = [\mathrm{NS}(\bar{X}_h) : \langle H \rangle]^2 \cdot \mathrm{disc}\, \mathrm{NS}(\bar{X}_h).$$

Together with $\mathrm{disc}\, \mathrm{NS}(\bar{X}_h)$ being even this implies $[\mathrm{NS}(\bar{X}_h) : \langle H \rangle] = 1$, so H generates $\mathrm{NS}(\bar{X}_h)$, which is isomorphic to $\mathrm{Pic}\, \bar{X}_h$ by Lemma 2.1. $\square$

Remark 3.2. Corollary 2.3 was pointed out to the author by Jasper Scholten and people have used it before to bound the geometric Picard number of a surface. However, since all nonreal roots of the characteristic polynomial of Frobenius come in conjugate pairs, the upper bound has the same parity as the second Betti number of the surface. For K3 surfaces this means that the upper bound is even, and therefore at least 2. Note that by Tate's conjecture (see Remark 2.4) the actual geometric Picard number of any K3 surface over a field of positive characteristic is at least 2. It is a complete mystery where this second cycle should come from. The strategy of the proof of Theorem 3.1 allows us to sharpen the upper bound in characteristic zero. If the reductions modulo two different primes give the same upper bound r, but the corresponding Néron–Severi groups have discriminants that do not differ by a square factor, then in fact $r - 1$ is an upper bound.

Kloosterman [2005] has used our method to construct an elliptic K3 surface with Mordell–Weil rank 15 over $\bar{\mathbb{Q}}$. In the proof of Theorem 3.1 we were able to compute the discriminant up to squares of the Néron–Severi lattice of $\bar{X}_p$ because we knew a priori a sublattice of finite index. Kloosterman realized that it is not always necessary to know such a sublattice. For an elliptic surface Y over $\bar{\mathbb{F}}_p$, the image in $\mathbb{Q}^*/\mathbb{Q}^{*2}$ of the discriminant of the Néron–Severi lattice can also be deduced from

the Artin–Tate conjecture, which has been proved for ordinary K3 surfaces in characteristic $p \geq 5$; see [Nygaard and Ogus 1985, Theorem 0.2] and [Milne 1975, Theorem 6.1]. It allows one to compute the ratio disc $NS(Y) \cdot \# Br(Y)/(\# NS(Y)_{tors})^2$ from the characteristic polynomial of Frobenius acting on $H^2_{\text{ét}}(Y, \mathbb{Q}_l)$. For an elliptic surface the Brauer group has square order, so this ratio determines the same element in $\mathbb{Q}^*/\mathbb{Q}^{*2}$ as disc $NS(Y)$. Of course this relies on heavy machinery, while our method is essentially elementary.

Remark 3.3. In the proof we counted points over $\mathbb{F}_{p^n}$ for $p = 2, 3$ and $n = 1, \ldots, 11$ in order to find the traces of powers of Frobenius up to the 11-th power. We could have got away with less counting. In both cases $p = 2$ and $p = 3$ we already know a 2-dimensional, Frobenius stable subspace W of $NS(\overline{X}_p)_{\mathbb{Q}_l} \subset H^2_{\text{ét}}(\overline{X}_p, \mathbb{Q}_l)(1)$, generated by the hyperplane section H and another divisor class. Therefore it suffices to find out the characteristic polynomial of Frobenius acting on the quotient $V = H^2_{\text{ét}}(\overline{X}_p, \mathbb{Q}_l)(1)/W$. This implies it suffices to know the traces of powers of Frobenius acting on V up to the 10-th power.

An extra trick was used for $p = 3$. The family of planes through the line L given by $w = z = 0$ cuts out a fibration of curves of genus 1. We can give all nonsingular fibers the structure of an elliptic curve by quickly looking for a point on it. There are efficient algorithms available in for instance MAGMA to count the number of points on these elliptic curves.

Using these few speed-ups we let a computer run to compute the characteristic polynomial of several random surfaces given by an equation of the form $wf_1 = zf_2$ over $\mathbb{F}_3$ or $wf_1 = g_1g_2$ over $\mathbb{F}_2$, as in (2). If the middle coefficient of the characteristic polynomial was zero, no more effort was spent on trying to find the sign of the functional equation (see proof of Theorem 3.1) and the surface was discarded. After one night two examples over $\mathbb{F}_3$ were found with geometric Picard number 2 and one example over $\mathbb{F}_2$. With the Chinese Remainder Theorem this allows us to construct two families of surfaces with geometric Picard number 1. One of these families consists of the surfaces X_h. A program written in MAGMA that checks the characteristic polynomial of Frobenius on X_2 and X_3 is electronically available from the author upon request.

Remark 3.4. For $p = 2, 3$, let $A_p \subset NS(\overline{X}_p)$ denote the lattice as described in the proof of Theorem 3.1, i.e., A_2 is generated by a hyperplane section and a conic, and A_3 is generated by a hyperplane section and a line. Then in fact A_p equals $NS(\overline{X}_p)$ for $p = 2, 3$. Indeed, we have

$$\text{disc } A_p = [NS(\overline{X}_p) : A_p]^2 \cdot \text{disc } NS(\overline{X}_p).$$

For $p = 2$ this implies disc $NS(\overline{X}_2) = -12$ or disc $NS(\overline{X}_2) = -3$. The latter is impossible because modulo 4 the discriminant of an even lattice of rank 2 is

congruent to 0 or -1. We conclude disc $\mathrm{NS}(\bar{X}_2) = -12$, and therefore $[\mathrm{NS}(\bar{X}_2) : A_2] = 1$, so $A_2 = \mathrm{NS}(\bar{X}_2)$.

For $p = 3$ we find disc $\mathrm{NS}(\bar{X}_3) = -9$ or disc $\mathrm{NS}(\bar{X}_3) = -1$. Suppose the latter equation held. By the classification of even unimodular lattices we find that $\mathrm{NS}(\bar{X}_3)$ is isomorphic to the lattice with Gram matrix

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

By a theorem of Van Geemen [2005, 5.4], this is impossible. From this contradiction we conclude disc $\mathrm{NS}(\bar{X}_3) = -9$ and thus $[\mathrm{NS}(\bar{X}_3) : A_3] = 1$, so $A_3 = \mathrm{NS}(\bar{X}_3)$.

Since there are $\binom{4+3}{3} = 35$ monomials of degree 4 in $\mathbb{Q}[x, y, z, w]$, the quartic surfaces in $\mathbb{P}^3_{\mathbb{Q}}$ are parametrized by the space $\mathbb{P}^{34}_{\mathbb{Q}}$, which we will denote by M. More explicitly, with these 35 monomials as coordinates, the surface given by $F = 0$ with F a homogeneous quartic in $\mathbb{Q}[x, y, z, w]$ corresponds to the point in M whose coordinates are the coefficients in F of the corresponding monomials. Let $M' \cong \mathbb{P}^{27} \subset M$ denote the subvariety of those surfaces X for which the coefficients of the monomials x^4, $x^3 y$, $x^3 z$, y^4, $y^3 x$, $y^3 z$, and $x^2 z^2$ in the defining polynomial of X are all zero. Note that the vanishing of the coefficients of the first six of these monomials is equivalent to the tangency of the plane H_w given by $w = 0$ to the surface X at the points $P = [1:0:0:0]$ and $Q = [0:1:0:0]$. Thus, the vanishing of these coefficients yields a singularity at P and Q in the plane curve $C_X = H_w \cap X$. If the singularity at P in C_X is not worse than a double point, then the vanishing of the coefficient of $x^2 z^2$ is equivalent to the fact that the line given by $y = w = 0$ is one of the limit-tangent lines to C_X at P.

Proposition 3.5. *There is a nonempty Zariski open subset $U \subset M'$ such that every surface $X \in U$ defined over $\mathbb{Q}$ is smooth and has infinitely many rational points.*

Proof. The singular surfaces in M' form a closed subset of M'. So do the surfaces X for which the intersection $H_w \cap X$ has worse singularities than just two double points at P and Q. Leaving out these closed subsets we obtain an open subset V of M'. Let $X \in V$ be given. The plane quartic curve $C_X = X \cap H_w$ has two double points, so the geometric genus g of the normalization $\tilde{C}_X$ of C_X equals $p_a - 2$, where p_a is the arithmetic genus of C_X; see [Hartshorne 1977, exercise IV.1.8]. As we have $p_a = \frac{1}{2}(4 - 1)(4 - 2) = 3$, we get $g = 1$. Now assume X is defined over $\mathbb{Q}$. One of the limit-tangents to C_X at P is given by $w = y = 0$. Its slope, being rational, corresponds to a rational point P' on $\tilde{C}_X$ above P. Fixing this point as the unit element $\mathbb{O} = P'$, the curve $\tilde{C}_X$ obtains the structure of an elliptic curve. Let $D \in \mathrm{Pic}^0(\tilde{C}_X)$ be the pull back under normalization of the divisor $P - Q \in \mathrm{Pic}^0(C_X)$. By the theory of elliptic curves there is a unique point T on $\tilde{C}_X$ such that D is linearly equivalent to $T - \mathbb{O}$; see [Silverman 1986, Proposition

III.3.4]. As D is defined over $\mathbb{Q}$, so is T. By Mazur's theorem (see [Silverman 1986, Theoremm III.7.5] for statement and [Mazur 1977, Theorem 8] for a proof), the point T has finite order if and only if $mT = \mathbb{O}$ for some $m \in \{1, 2, \ldots, 10, 12\}$. Note that we have $\mathrm{lcm}(1, 2, \ldots, 10, 12) = 2520$. Take for U the complement in V of the closed subset of those X for which we have $2520T = \mathbb{O}$ for the corresponding point T on $\tilde{C}_X$. Then each $X \in U$ contains an elliptic curve with infinitely many rational points. By choosing a Weierstrass equation, one verifies easily that if we take $X = X_h$ with $h = 0$, then the corresponding point T on $\tilde{C}_X$ satisfies $mT \neq \mathbb{O}$ for $m \in \{1, 2, \ldots, 10, 12\}$. Therefore, we find $X \in U$, so U is nonempty.   □

Remark 3.6. If $\tilde{C}_X$ is the normalization of C_X as in the proof of Proposition 3.5, then generically there is another rational point P'' on $\tilde{C}_X$ above P, besides P'. Generically this point also has infinite order and the Mordell–Weil rank of $\tilde{C}_X$ is at least 2 with independent points P'' and T as in the proof of Proposition 3.5. For $X = X_h$ with $h = 0$ however, the curve $\tilde{C}_X$ is given by

$$3x^2y^2 + xy^2z + 4xyz^2 + 2xz^3 + 5yz^3 + z^4 = 0.$$

As the point $P = [1 : 0 : 0]$ is a cusp, there is only one point above P on $\tilde{C}_X$ here. The conductor of this elliptic curve equals 686004. Both points on $\tilde{C}_X$ above $Q = [0 : 1 : 0]$ are rational and we have an extra rational point $[1 : 1 : -1]$. These generate the full Mordell–Weil group of rank 3.

Remark 3.7. Besides the family X_h (with $h \in U$ as in Proposition 3.5) of surfaces containing an elliptic curve with positive Mordell–Weil rank, we can also find surfaces with infinitely many points on some curve of genus 0. By requiring other coefficients to vanish than is required for M', we can find quartic surfaces Y for which the plane H_w given by $w = 0$ is tangent at $[1 : 0 : 0 : 0]$, $[0 : 1 : 0 : 0]$, and $[0 : 0 : 1 : 0]$. Then the intersection $H_w \cap Y$ has geometric genus 0 and if its normalization has a point defined over $\mathbb{Q}$, then this intersection is birational to $\mathbb{P}^1$. The quartic surface Z given by

$$w(x^3 + y^3 + z^3 + x^2z + xw^2) = 3x^2y^2 - 4x^2yz + x^2z^2 + xy^2z + xyz^2 - y^2z^2 \quad (4)$$

is an example of such a surface. As in the proof of Theorem 3.1, modulo 3 the surface Z contains the line $z = w = 0$. Also, the reduction of Z at $p = 2$ contains a conic again, as the right-hand side of (4) factors over $\mathbb{F}_4$ as $(xy + xz + \zeta yz)(xy + xz + \zeta^2 yz)$, with $\zeta^2 + \zeta + 1 = 0$. An argument very similar to the one in the proof of Theorem 3.1 then shows that Z also has geometric Picard number 1 with the Picard group generated by a hyperplane section. The only difference is that Frobenius does not act trivially on the conic $w = xy + xz + \zeta yz = 0$. The hyperplane section $H_w \cap Z$ is a curve of geometric genus 0, parametrized by

$$[x : y : z : w] = [-(t^2 + t - 1)(t^2 - t - 3) : 2(t + 2)(t^2 + t - 1) : 2(t + 2)(t^2 - t - 3) : 0].$$

The Cremona transformation $[x : y : z : w] \mapsto [yz : xz : xy]$ gives a birational map from this curve to a nonsingular plane curve of degree 2. Coincidentally, it turns out that the curve on Z given by $x = 0$ has a triple point at $[0 : 0 : 0 : 1]$, so it is birational to $\mathbb{P}^1$ as well. It can be parametrized by

$$[x : y : z : w] = [0 : 1 + t^3 : t(1 + t^3) : -t^2].$$

From the local and global Torelli theorem for K3 surfaces [Pyatetskii-Shapiro and Shafarevich 1971] one can find a very precise description of the moduli space of polarized K3 surfaces in general; see [Beauville 1985]. A polarization of a K3 surface Z by a very ample divisor of degree 4 gives an embedding of Z as a smooth quartic surface in $\mathbb{P}^3$ with the very ample divisor corresponding to a hyperplane section. An isomorphism between two smooth quartic surfaces in $\mathbb{P}^3$ that sends one hyperplane section to an other hyperplane section comes from an automorphism of $\mathbb{P}^3$. As any two hyperplane sections are linearly equivalent, we conclude that the moduli space of K3 surfaces polarized by a very ample divisor of degree 4 is isomorphic to the open subset in $M = \mathbb{P}^{34}$ of smooth quartic surfaces modulo the action of PGL(4) by linear transformations of $\mathbb{P}^3$. We are now ready to prove the main theorem of this article.

Proof Theorem 1.1. By the description of the moduli space of K3 surfaces polarized by a very ample divisor of degree 4 given above, it suffices to prove that the set $S \subset M(\mathbb{Q})$ of smooth surfaces with geometric Picard number 1 and infinitely many rational points is Zariski dense in M.

We will first show that $S \cap M'$ is dense in M'. Note that the coefficients of the monomials x^4, x^3y, x^3z, y^4, y^3x, y^3z, and x^2z^2 in $wf_1 + 2zf_2 - 3g_1g_2$ in (2) are zero, so if the coefficients of these monomials in a homogeneous polynomial $h \in R$ of degree 4 are all zero, then X_h is contained in M'. It follows that the set

$$T = M' \cap \{X_h : h \in R, h \text{ homogeneous of degree } 4\}$$

is dense in M'. Let U be as in Proposition 3.5. Then U is a dense open subset of M', so $T \cap U$ is also dense in M'. By Theorem 3.1 and Proposition 3.5 every surface in $T \cap U$ has geometric Picard number 1 and infinitely many rational points. Thus we have an inclusion $T \cap U \subset S \cap M'$, so $S \cap M'$ is dense in M' as well.

Let W denote the vector space of 4×4–matrices over $\mathbb{Q}$ and let T denote the dense open subset of $\mathbb{P}(W)$ corresponding to elements of PGL(4). Let $\varphi \colon T \times M' \to M$ be given by sending (A, X) to $A(X)$. Note that $T(\mathbb{Q}) \times (S \cap M')$ is dense in $T \times M'$ and φ sends $T(\mathbb{Q}) \times (S \cap M')$ to S. Hence, in order to prove that S is dense in M, it suffices to show that φ is dominant, which can be checked after extending to the algebraic closure. A general quartic surface in $\mathbb{P}^3$ has a one-dimensional family of bitangent planes, i.e., planes that are tangent at two different

points. This is closely related to the theorem of Bogomolov and Mumford; see the appendix to [Mori and Mukai 1983]. In fact, for a general quartic surface $Y \subset \mathbb{P}^3$, there is such a bitangent plane H, for which the two tangent points are ordinary double points in the intersection $H \cap Y$. Let Y be such a quartic surface and H such a plane, say tangent at P and Q. Then there is a linear transformation that sends H, P, and Q to the plane given by $w = 0$, and the points $[1:0:0:0]$ and $[0:1:0:0]$. Also, one of the limit-tangent lines to the curve $Y \cap H$ at the singular point P can be sent to the line given by $y = w = 0$. This means that there is a linear transformation B that sends Y to an element X in M'. Then $\varphi(B^{-1}, X) = Y$, so φ is indeed dominant.   □

Remark 3.8. The explicit polynomials f_1, f_2, g_1, g_2 for X_h in (2) were found by letting a computer pick random polynomials modulo $p = 2$ and $p = 3$ such that the surface X_h with $h = 0$ is contained in M' as in Proposition 3.5. The computer then computed the characteristic polynomial of Frobenius and tested if there were only 2 eigenvalues that were roots of unity, see Remark 3.3.

Remark 3.9. In finding the explicit surfaces X_h not much computing power was needed, as we constructed the surface to have good reduction at small primes p so that counting points over $\mathbb{F}_{p^n}$ was relatively easy. Based on ideas of for instance Alan Lauder, Daqing Wan, Kiran Kedlaya, and Bas Edixhoven, it should be possible to develop more efficient algorithms for finding characteristic polynomials of (K3) surfaces. Together with these algorithms, the method used in the proof of Theorem 3.1 becomes a strong tool in finding Picard numbers of K3 surfaces over number fields.

4. Open problems

We end with the remark that still very little is known about the arithmetic of K3 surfaces, especially those with geometric Picard number 1. We reiterate three questions that remain unsolved.

Question 1. Does there exist a K3 surface over a number field such that the set of rational points is neither empty nor dense?

Question 2. Does there exist a K3 surface over a number field with geometric Picard number 1, such that the set of rational points is potentially dense?

Question 3. Does there exist a K3 surface over a number field with geometric Picard number 1, such that the set of rational points is not potentially dense?

The surfaces exhibited in this paper are candidates to yield affirmative answers to all of these questions, most notably Questions 2 and 3.

References

[Barth et al. 1984] W. Barth, C. Peters, and A. van de Ven, *Compact complex surfaces*, Ergebnisse der Math. **4**, Springer, Berlin, 1984. MR 86c:32026 Zbl 0718.14023

[Beauville 1985] A. Beauville, "Application aux espaces de modules", pp. 141–152 in *Géométrie des surfaces K3: modules et périodes* (Palaiseau, 1981/1982), Astérisque **126**, Soc. math. de France, Paris, 1985. MR MR785231 Zbl 0577.14007

[Bogomolov and Tschinkel 2000] F. A. Bogomolov and Y. Tschinkel, "Density of rational points on elliptic K3 surfaces", *Asian J. Math.* **4**:2 (2000), 351–368. MR 2002b:14025 Zbl 0983.14008

[Bombieri and Mumford 1977] E. Bombieri and D. Mumford, "Enriques' classification of surfaces in char. *p*, II", pp. 23–42 in *Complex analysis and algebraic geometry*, edited by W. L. Baily and T. Shioda, Iwanami Shoten, Tokyo, 1977. MR 58 #10922a Zbl 0348.14021

[Borwein and Erdélyi 1995] P. Borwein and T. Erdélyi, *Polynomials and polynomial inequalities*, Graduate Texts in Mathematics **161**, Springer, New York, 1995. MR 97e:41001 Zbl 0840.26002

[Deligne 1974] P. Deligne, "La conjecture de Weil, I", *Inst. Hautes Études Sci. Publ. Math.* 43 (1974), 273–307. MR 49 #5013 Zbl 0287.14001

[Deligne and Katz 1973] P. Deligne and N. Katz, *Groupes de monodromie en géométrie algébrique, II: Séminaire de géométrie algébrique du Bois-Marie 1967–1969* (SGA 7 II), Lecture Notes in Math. **340**, Springer, Berlin, 1973. MR 50 #7135

[Ellenberg 2004] J. S. Ellenberg, "K3 surfaces over number fields with geometric Picard number one", pp. 135–140 in *Arithmetic of higher-dimensional algebraic varieties* (Palo Alto, CA, 2002), edited by B. Poonen and Y. Tschinkel, Progr. Math. **226**, Birkhäuser, Boston, 2004. MR 2001f: 17055 Zbl 02158897

[Grothendieck et al. 1971] A. Grothendieck et al., *Théorie des intersections et théorème de Riemann–Roch: Séminaire de géométrie algébrique du Bois-Marie 1966–1967* (SGA 6), Lecture Notes in Math. **225**, Springer, Berlin, 1971. MR 50 #7133

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Math **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Kloosterman 2005] R. Kloosterman, "An explicit example of an elliptic K3 surface with Mordell–Weil rank 15", preprint, 2005. to be published in *Can. Math. Bull.* math.AG/0502439

[Mazur 1977] B. Mazur, "Modular curves and the Eisenstein ideal", *Inst. Hautes Études Sci. Publ. Math.* 47 (1977), 33–186. MR 80c:14015 Zbl 0394.14008

[Milne 1975] J. S. Milne, "On a conjecture of Artin and Tate", *Ann. of Math.* (2) **102**:3 (1975), 517–533. MR 54 #2659 Zbl 0343.14005

[Milne 1980] J. S. Milne, *Étale cohomology*, Princeton Mathematical Series **33**, Princeton University Press, Princeton, N.J., 1980. MR 81j:14002 Zbl 0433.14012

[Mori and Mukai 1983] S. Mori and S. Mukai, "The uniruledness of the moduli space of curves of genus 11", pp. 334–353 in *Algebraic geometry* (Tokyo/Kyoto, 1982), edited by A. Dold and B. Eckmann, Lecture Notes in Math. **1016**, Springer, Berlin, 1983. MR 85b:14033 Zbl 0557.14015

[Nygaard and Ogus 1985] N. Nygaard and A. Ogus, "Tate's conjecture for K3 surfaces of finite height", *Ann. of Math.* (2) **122**:3 (1985), 461–507. MR 87h:14014 Zbl 0591.14005

[Pyatetskii-Shapiro and Shafarevich 1971] I. I. Pyatetskii-Shapiro and I. R. Shafarevich, "Torelli's theorem for algebraic surfaces of type K3", *Izv. Akad. Nauk SSSR Ser. Mat.* **35** (1971), 530–572. MR 44 #1666 Zbl 0219.14021

[Shioda 1981] T. Shioda, "On the Picard number of a complex projective variety", *Ann. Sci. École Norm. Sup.* (4) **14**:3 (1981), 303–321. MR 83i:14005 Zbl 0567.14021

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Math. **106**, Springer, New York, 1986. MR 87g:11070 Zbl 0585.14026

[Tate 1965] J. T. Tate, "Algebraic cycles and poles of zeta functions", pp. 93–110 in *Arithmetical algebraic geometry* (Purdue Univ., 1963), edited by O. F. G. Schilling, Harper & Row, New York, 1965. MR 37 #1371 Zbl 0213.22804

[Terasoma 1985] T. Terasoma, "Complete intersections with middle Picard number 1 defined over ℚ", *Math. Z.* **189**:2 (1985), 289–296. MR 86f:14010 Zbl 0579.14006

[van Geemen 2005] B. van Geemen, "Some remarks on Brauer groups of K3 surfaces", *Adv. Math.* **197**:1 (2005), 222–247. MR 2006h:14025 Zbl 1082.14040

[van Luijk 2007] R. van Luijk, "An elliptic K3 surface associated to Heron triangles", *J. Number Theory* **123** (2007), 92–119. Zbl 05129748

rmluijk@gmail.com   *Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada*
http://www.cecm.sfu.ca/~rluijk

Realising fusion systems

Ian J. Leary and Radu Stancu

We show that every fusion system on a p-group S is equal to the fusion system associated to a discrete group G with the property that every p-subgroup of G is conjugate to a subgroup of S.

1. Introduction

Let p be a prime number. By a p-group we shall mean a finite group whose order is a power of p. A *fusion system* on a p-group S is a category $\mathcal{F}$ whose objects are the subgroups of S, and whose morphisms are injective group homomorphisms, subject to certain axioms. The notion of a fusion system is intended to axiomatize the p-local structure of a discrete group $G \geq S$ in which every p-subgroup is conjugate to a subgroup of S. Every such G gives rise to a fusion system $\mathcal{F}_S(G)$ on S, and we say that G realises $\mathcal{F}$ if $\mathcal{F}_S(G) = \mathcal{F}$.

The notion of a *saturated fusion system* is intended to axiomatize the p-local structure of a finite group in which S is a Sylow p-subgroup. It is known that there are saturated fusion systems $\mathcal{F}$ which are not realised by any finite group G, although showing that this is the case is very delicate. In the case when $p = 2$, the only known examples are certain systems discovered by Solomon [1974] (see also [Benson 1994; Levi and Oliver 2002]).

In contrast, we show that every fusion system on any p-group S is realised by some discrete group $G \geq S$ in which every maximal p-subgroup is conjugate to S. The groups G that are used in our proofs are constructed as graphs of finite groups. In particular each of our groups G contains a free subgroup of finite index. In an appendix we give a brief account of those parts of the theory of graphs of groups that we use.

While preparing this paper, we learned that Robinson [2006] has proved a similar, but not identical result. Since his article was already submitted when we started to write this paper, we have taken it upon ourselves to compare and contrast the two results. Robinson's construction realises a large class of fusion systems, including all saturated fusion systems, but does not realise all fusion systems. The groups

that Robinson constructs are iterated free products with amalgamation, whereas the groups that we construct are iterated HNN extensions. In both cases the groups may be viewed as graphs of finite groups.

We state and outline the proof of a version of Robinson's theorem, along the lines of the proof of our main result. We also give examples of fusion systems that cannot be realised by Robinson's method, we give examples of nonsaturated fusion systems that are realised by Robinson's method, and we prove an analogue of Cayley's theorem for fusion systems.

2. Definitions and results

Let p be a prime, and let G be a discrete group. The p-Frobenius category $\Phi_p(G)$ of the group G is a category whose objects are the p-subgroups of G. If P and Q are p-subgroups of G, or equivalently objects of $\Phi_p(G)$, the morphisms from P to Q are the group homomorphisms $f : P \to Q$ that are equal to conjugation by some element of G. Thus $f : P \to Q$ is in $\Phi_p(G)$ if and only if there exists $g \in G$ with $f(u) = g^{-1}ug$ for all $u \in P$. (Note that the element g is *not* part of the morphism. If $g' = zg$ for some element z in the centralizer of P, then g and g' define the same morphism.)

Now suppose that S is a p-subgroup of G, and let $\mathscr{F}_S(G)$ denote the full subcategory of $\Phi_p(G)$ with objects the subgroups of S. Such categories as these are examples of *fusion systems on S*. According to Puig, a fusion system or 'Frobenius system' on S is a category $\mathscr{F}$. The objects of $\mathscr{F}$ are the subgroups of S, and the morphisms from P to Q form a subset of the set $\operatorname{Inj}(P, Q)$ of injective group homomorphisms from P to Q. These are subject to the following axioms:

(1) For any $s \in S$, and any $P, Q \le S$ with $s^{-1}Ps \le Q$, the morphism $\phi : P \to Q$ defined by $\phi : u \mapsto s^{-1}us$ is in $\mathscr{F}$. (Equivalently, $\mathscr{F}_S(S) \subseteq \mathscr{F}$.)

(2) If $f : P \to Q$ is in $\mathscr{F}$, with $R = f(P) \le Q$, then so are $f : P \to R$ and $f^{-1} : R \to P$.

It is easily checked that these axioms are satisfied in the case when $\mathscr{F} = \mathscr{F}_S(G)$ as defined above.

Now consider the special case in which S is a p-subgroup of G that is maximal, and further suppose that every p-subgroup of G is conjugate to a subgroup of S. In this case, every object of $\Phi_p(G)$ is isomorphic within the category $\Phi_p(G)$ to a subgroup of S. It follows that the full subcategory $\mathscr{F}_S(G)$ is equivalent to $\Phi_p(G)$. This was one of the main motivating examples for Puig's definition.

We say that the pair (G, S) realises the fusion system $\mathscr{F}$ if S is a p-subgroup of G, $\mathscr{F}$ is a fusion system on S, and $\mathscr{F}_S(G) = \mathscr{F}$. If G is a group in which every p-subgroup is conjugate to a subgroup of some p-subgroup S, we say that G realises $\mathscr{F}$ if the pair (G, S) realises $\mathscr{F}$.

Remark 1. Another source of fusion systems on a p-group S is the Brauer category of a p-block b [Alperin and Broué 1979; Linckelmann 2006]. Here H is a finite group, S is the defect group of the p-block b, and the morphisms in the category are those conjugations by elements of H that preserve some extra structure associated to b. In the case when b is the principal block, S is the Sylow p-subgroup of H and this fusion system is just $\mathscr{F}_S(H)$. One corollary of our Theorem 2 is that every such fusion system is realised by some group G in which every p-subgroup is conjugate to a subgroup of S.

Let $\mathscr{F}$ be a fusion system on S and let $\mathscr{F}'$ be a fusion system on S'. A morphism $\alpha : \mathscr{F} \to \mathscr{F}'$ consists of a group homomorphism $\alpha_0 : S \to S'$ and a functor α from $\mathscr{F}$ to $\mathscr{F}'$ such that

(1) for all $P \leq S$, $\alpha_0(P) = \alpha(P)$;

(2) for each $P, Q \leq S$ and each $\phi \in \mathrm{Hom}_{\mathscr{F}}(P, Q)$, $\alpha(\phi) \circ \alpha_0 = \alpha_0 \circ \phi$.

With this notion of morphism, the class of all fusion systems on p-groups becomes a category. If S is a p-subgroup of G and S' is a p-subgroup of H, then any group homomorphism $f : G \to H$ with the property that $f(S) \leq S'$ gives rise to a morphism of fusion systems $f_* : \mathscr{F}_S(G) \to \mathscr{F}_{S'}(H)$.

Define the category of pairs to have objects the pairs (G, S), where G is a group and S is a p-subgroup, where the morphisms from the pair (G, S) to the pair (H, S') are the group homomorphisms $f : G \to H$ such that $f(S) \leq S'$. With these definitions, there is a realisation functor from the category of pairs to the category of fusion systems, which takes the object (G, S) to the fusion system $\mathscr{F}_S(G)$ on S, and takes the homomorphism f to the morphism f_*.

There is a fusion system $\mathscr{F}_S^{\max}$ on S, in which the set of morphisms from P to Q consists of all injective group homomorphisms from P to Q. Any fusion system on S is a subcategory of $\mathscr{F}_S^{\max}$, and the intersection of a family of fusion systems on S is itself a fusion system. If $\Phi = \{\phi_1, \ldots, \phi_r\}$ is a collection of morphisms in $\mathscr{F}_S^{\max}$, where $\phi_i : P_i \to Q_i$, the fusion system generated by Φ is defined to be the smallest fusion system that contains each ϕ_i.

Theorem 2. *Suppose $\mathscr{F}$ is the fusion system on S generated by $\Phi = \{\phi_1, \ldots, \phi_r\}$. Let T be a free group with free generators $t_1, \ldots, t_r$, and define G as the quotient of the free product $S * T$ by the relations $t_i^{-1} u t_i = \phi_i(u)$ for all i and for all $u \in P_i$. Then S embeds as a subgroup of G, every p-subgroup of G is conjugate to a subgroup of S, and $\mathscr{F}_S(G) = \mathscr{F}$. Moreover, every finite subgroup of G is conjugate to a subgroup of S, and G has a free normal subgroup of index dividing $|S|!$.*

As was pointed out to us by the referee, the group constructed in Theorem 2 enjoys a universal property.

Corollary 3. *Suppose that H is a group containing S as a subgroup, and that the fusion system $\mathcal{F}$ as in the statement of Theorem 2 is realised by the pair (H, S). Let $h_1, \ldots, h_r$ be any elements of H such that conjugation by h_i induces the morphism $\phi_i : P_i \to Q_i$. For G as defined in the statement of Theorem 2 there is a unique group homomorphism $f : G \to H$ such that $f(s) = s$ for all $s \in S$ and such that $f(t_i) = h_i$. Furthermore $f_* : \mathcal{F}_S(G) \to \mathcal{F}_S(H)$ is an isomorphism.*

Corollary 4. *The category of fusion systems is a retract of the category of pairs as defined above. In other words, there is a functor from the category of fusion systems to the category of pairs which is a preinverse to the realisation functor.*

If $f : S' \to S$ is an injective group homomorphism between p-groups, and $\mathcal{F}'$ is a fusion system on S', then there is a functor $f_!$ from $\mathcal{F}'$ to $\mathcal{F}_S^{\max}$, which sends $P' \leq S'$ to $f(P')$ and $\phi' : P' \to Q'$ to

$$f \circ \phi' \circ f^{-1} : f(P') \to f(Q').$$

Theorem 5. *Suppose that $\mathcal{F}$ is the fusion system on S generated by the images $(f_i)_!(\mathcal{F}_{S_i}(G_i))$ for injective group homomorphisms $f_i : S_i' \to S$ for $1 \leq i \leq r$, where G_i is a finite group with S_i' as a Sylow p-subgroup. Define G as the quotient of the free product $S * G_1 * \cdots * G_r$ by the relations $s = f_i(s)$ for all i and for all $s \in S_i'$. Then S embeds as a subgroup of G, every p-subgroup of G is conjugate to a subgroup of S, and $\mathcal{F}_S(G) = \mathcal{F}$. Moreover, every finite subgroup of G is conjugate to a subgroup of one of the G_i, or to a subgroup of S, and G has a free normal subgroup of index dividing $N!$, where N is the least common multiple of $|S|$ and the $|G_i|$.*

Remark 6. The theorem can be obtained from Theorem 1 of [Robinson 2006] by induction. The main result of that paper is Theorem 2, which is similar to the statement above except that extra conditions are put on the G_i. These extra conditions allow Robinson to improve the bound on the index of a free normal subgroup, and to deduce some information about the finite quotient by such a subgroup. Another slight difference is that Robinson describes his group as a free product with amalgamation $G_1 * \cdots * G_r$, where G_1 has S as a Sylow p-subgroup. The groups that arise in this way are the same groups as those that arise from our statement, since if S is a subgroup of G_1, then $S *_S G_1 = G_1$.

Theorem 7. *Let Σ denote the group of all permutations of the elements of a p-group S, and identify S with a subgroup of Σ via the Cayley embedding. Every fusion system on S is equal to a subcategory of the Frobenius category $\Phi_p(\Sigma)$ of Σ.*

3. Saturated fusion systems

In this section we present the definition of a saturated fusion system, due to Puig [2002], although we shall describe an equivalent definition due to Broto, Levi and

Oliver [Broto et al. 2003]. There are two additional axioms as well as the axioms for a fusion system. These axioms necessitate some preliminary definitions.

As usual, if G is a group and H is a subgroup of G, we write $C_G(H)$ for the centralizer of H in G and $N_G(H)$ for the normalizer of H in G.

Suppose that $\mathcal{F}$ is a fusion system on S. Say that $P \leq S$ is fully $\mathcal{F}$-centralized if

$$|C_S(P)| \geq |C_S(P')|$$

for every P' which is isomorphic to P as an object of $\mathcal{F}$. Suppose that $\mathcal{F} = \mathcal{F}_S(G)$ for some discrete group G in which every p-subgroup is conjugate to a subgroup of S. In this case, if P is fully $\mathcal{F}$-centralized, one sees that $C_S(P)$ is a p-subgroup of $C_G(P)$ of maximal order.

Similarly, say that P is fully $\mathcal{F}$-normalized if

$$|N_S(P)| \geq |N_S(P')|$$

for every P' which is isomorphic to P as an object of $\mathcal{F}$. If $\mathcal{F} = \mathcal{F}_S(G)$ as above and P is fully $\mathcal{F}$-normalized, one sees that $N_S(P)$ is a p-subgroup of $N_G(P)$ of maximal order.

Now suppose that $\mathcal{F} = \mathcal{F}_S(G)$ for some finite group G, and that $P \leq S$ is fully $\mathcal{F}$-normalized. In this case, $N_S(P)$ must be a Sylow p-subgroup of the finite group $N_G(P)$. Moreover, $C_G(P) \cap N_S(P) = C_S(P)$ must be a Sylow p-subgroup of $C_G(P)$, and $\mathrm{Aut}_S(P) = N_S(P)/C_S(P)$ must be a Sylow p-subgroup of $\mathrm{Aut}_{\mathcal{F}}(P) = N_G(P)/C_G(P)$. This gives the first of two extra axioms for a saturated fusion system:

(3) If P is fully $\mathcal{F}$-normalized, then P is also fully $\mathcal{F}$-centralized, and $\mathrm{Aut}_S(P)$ is a Sylow p-subgroup of $\mathrm{Aut}_{\mathcal{F}}(P)$.

Next, suppose that $\mathcal{F} = \mathcal{F}_S(G)$ for some finite group G and that $f : P \to Q \leq S$ is an isomorphism in $\mathcal{F}$ such that Q is fully $\mathcal{F}$-centralized. This implies that $C_S(Q)$ is a Sylow p-subgroup of $C_G(Q)$. Pick an element $h \in G$ so that f is equal to conjugation by h, i.e., so that $f(u) = c_h(u) = h^{-1}uh$ for all $u \in P$. The image $c_h(C_S(P))$ is a p-subgroup of $C_G(c_h(P)) = C_G(Q)$, and so there exists $h' \in C_G(Q)$ so that $c_{h'} \circ c_h(C_S(P)) \leq C_S(Q)$. Since $c_{h'}$ acts as the identity on Q, if we define $k = hh'$, we see that c_k extends f and $c_k(C_S(P)) \leq C_S(Q)$.

The map c_k clearly extends to a map from $N_f = N_S(P) \cap c_k^{-1}(N_S(Q))$ to $N_S(Q)$. But since $C_S(P)$ is a subgroup of $c_k^{-1}(N_S(Q))$, we may rewrite this as

$$N_f = \{g \in N_S(P): c_k \circ c_g \circ c_k^{-1} \in \mathrm{Aut}_S(Q)\} = \{g \in N_S(P): f \circ c_g \circ f^{-1} \in \mathrm{Aut}_S(Q)\},$$

which does not depend on choice of k. This leads to the second extra axiom:

(4) If $f : P \to Q$ is an isomorphism in $\mathcal{F}$ and Q is fully $\mathcal{F}$-centralized, then f extends in $\mathcal{F}$ to a map from N_f to $N_S(Q)$, where

$$N_f = \{g \in N_S(P) \colon f \circ c_g \circ f^{-1} \in \mathrm{Aut}_S(Q)\}.$$

Remark 8. It was shown in [Kessar and Stancu 2007] that the axioms for a saturated fusion system can be simplified to:

(3′) $\mathrm{Aut}_S(S)$ is a Sylow p-subgroup of $\mathrm{Aut}_{\mathcal{F}}(S)$.

(4′) If $f : P \to Q$ is an isomorphism in $\mathcal{F}$ and Q is fully $\mathcal{F}$-normalized, then f extends in $\mathcal{F}$ to a map from N_f to $N_S(Q)$, where N_f is as defined in axiom 4.

Remark 9. In the case when S is abelian, axioms 3 and 4 simplify. In this case, every subgroup of S is fully $\mathcal{F}$-centralized and fully $\mathcal{F}$-normalized for any fusion system $\mathcal{F}$, and for any $f \in \mathcal{F}$, $N_f = S$. Hence a fusion system $\mathcal{F}$ on an abelian p-subgroup S is saturated if and only if $\mathrm{Aut}_{\mathcal{F}}(S)$ is a p'-group and every morphism $f : P \to S$ in $\mathcal{F}$ extends to an automorphism of S.

Remark 10. As mentioned in the introduction, there are saturated fusion systems which are not realised by any finite group. One source of saturated fusion systems is the fusion systems associated to p-blocks of finite groups [Alperin and Broué 1979; Linckelmann 2006]. The question of whether every such fusion system can be realised by a finite group is a long-standing open problem.

4. Examples

Let E be an elementary abelian p-group of rank at least three, i.e., a direct product of at least three copies of the cyclic group of order p. Let $A = \mathrm{Aut}(E)$ be the full group of automorphisms of E, which is of course isomorphic to a general linear group over the field of p elements. Let B be a subgroup of A of order a power of p, and let C be a nontrivial subgroup of A of order coprime to p. Note that A is generated by its subgroups of order coprime to p.

Each of A, B and C may be viewed as a collection of morphisms in the fusion system $\mathcal{F}_E^{\max}$. For $X = A$, B or C, let $\mathcal{F}_E(X)$ denote the fusion system generated by all the morphisms in X.

Example 11. The fusion system $\mathcal{F}_E(C)$ is saturated, and is equal to the fusion system $\mathcal{F}_E(G)$, where G is the semidirect product $G = E \rtimes C$.

Example 12. The fusion system $\mathcal{F}_E(A)$ is not saturated, since in $\mathcal{F}_E(A)$ the automorphism group of the object E does not have $E/Z(E)$ as a Sylow p-subgroup. However, $\mathcal{F}_E(A)$ can be realised by the procedure of Theorem 5. Let $C_1, \ldots, C_r$ be p'-subgroups of A that together generate A. If we put $G_i = E \rtimes C_i$ with f_i the identity map of E, then the fusion system generated by all of the $(f_i)_!(\mathcal{F}_E(G_i))$ is equal to $\mathcal{F}_E(A)$.

Example 13. The fusion system $\mathcal{F}_E(B)$ cannot be realised by the procedure used in Theorem 5. For suppose that $G_1, \ldots, G_r$ are finite groups with Sylow p-subgroups $E_1, \ldots, E_r$, each of which is isomorphic to a subgroup of E, and suppose that $\mathcal{F}_E(B)$ is generated by the fusion systems $(f_i)_!\mathcal{F}_{E_i}(G_i)$. Those G_i for which $f_i : E_i \to E$ is not an isomorphism do not contribute any morphisms to $\mathrm{Aut}_{\mathcal{F}}(E)$. If $f_i : E_i \to E$ is an isomorphism, then either $\mathrm{Aut}_{G_i}(E_i)$ contains nonidentity elements of p' order, implying that $\mathcal{F} \neq \mathcal{F}_E(B)$, or E_i is central in G_i and G_i does not contribute any morphisms to $\mathrm{Aut}_{\mathcal{F}}(E)$.

Next we consider some examples of fusion systems $\mathcal{F}$ on an abelian p-group E in which $\mathrm{Aut}_{\mathcal{F}}(E)$ is a p'-group, but for which some isomorphisms between proper subgroups of E do not extend to elements of $\mathrm{Aut}_{\mathcal{F}}(E)$.

Example 14. Let F and F' be distinct order p subgroups of E, and let $\phi : F \to F'$ be an isomorphism. Let $\mathcal{F}_E(\phi)$ be the fusion system generated by ϕ. Every morphism in $\mathcal{F}_E(\phi)$ is equal to either an inclusion map or the composite of either ϕ or ϕ^{-1} with an inclusion map. In particular, in $\mathcal{F}_E(\phi)$, the automorphism group of each object $E' \leq E$ is trivial. The fusion system $\mathcal{F}_E(\phi)$ cannot be realised by the procedure of Theorem 5, as will be explained below.

In view of Remark 9, $\mathcal{F}_E(\phi)$ is not a saturated fusion system, since the morphism $\phi : F \to F'$ does not extend to an automorphism in $\mathcal{F}_E(\phi)$ of the group E.

Now suppose $\mathcal{F}$ is a fusion system on E generated by the images $(f_i)_!\mathcal{F}_{E_i}(G_i)$ of some fusion systems for finite groups. If $\phi : F \to F'$ is a morphism in $\mathcal{F}$, then there exists i so that $F, F' \leq f_i(E_i)$ and $\phi \in (f_i)_!\mathcal{F}_{E_i}(G_i)$. But then (by the same argument as used above) there is a morphism $\tilde{\phi} : f_i(E_i) \to f_i(E_i)$ extending $\phi : F \to F'$. Thus $\mathcal{F}$ cannot be equal to the fusion system $\mathcal{F}_E(\phi)$, since this fusion system contains no such $\tilde{\phi}$.

Example 15. Let F be a proper subgroup of E, and suppose that D is a nontrivial p'-group of automorphisms of F. Let $F \rtimes D$ denote the semidirect product of F and D, let G be the free product with amalgamation $G = E *_F (F \rtimes D)$, and let $\mathcal{F}$ be the fusion system $\mathcal{F}_E(G)$. From this definition one sees that $\mathcal{F}$ can be obtained by the procedure of Theorem 5. On the other hand, since $\mathrm{Aut}_{\mathcal{F}}(E)$ is trivial, one sees that the nontrivial automorphisms of F do not extend to automorphisms of E, and hence $\mathcal{F}$ is not saturated.

As remarked earlier, Robinson does not consider all fusion systems that can be built by the procedure of Theorem 5, but only those that he calls Alperin fusion systems [Robinson 2006]. With the notation of Theorem 5 (and bearing in mind Remark 6), a fusion system is Alperin if the following conditions hold:

(1) Inside each G_i there is a subgroup E_i which is the largest normal p-subgroup of G_i, and the centralizer of this subgroup is as small as possible, in the sense that $C_{G_i}(E_i) = Z(E_i)$;

(2) The quotient G_i/E_i is isomorphic to $\mathrm{Out}_{\mathscr{F}}(E_i) := \mathrm{Aut}_{\mathscr{F}}(E_i)/\mathrm{Aut}_{E_i}(E_i)$;

(3) Inside S, the image of the subgroup S_i' (the Sylow p-subgroup of G_i which is to be identified with a subgroup of S) is equal to the normalizer of the image of E_i, i.e., $f_i(S_i') = N_S(f_i(E_i))$.

In terms of this definition, the content of Alperin's fusion theorem [1967], with some later embellishments [Goldschmidt 1970], is that the fusion system for any finite group is Alperin. Robinson [2006] remarks that work of Broto, Castellana, Grodal, Levi and Oliver implies that every saturated fusion system is Alperin [Broto et al. 2005]. It is easy to see that a fusion system on an abelian p-group is Alperin if and only if it is saturated. We finish this section by giving an example of a fusion system that is Alperin but not saturated.

Example 16. Let p be an odd prime, let $A = (C_p)^3$, and let B be a subgroup of $\mathrm{Aut}(A)$ of order p such that A is indecomposable as a B-module. (Equivalently, the action of a generator for B on A should be a single Jordan block.) Let S be the semidirect product $S = A \rtimes B$. The centre Z of S has order p. Let $E = Z \times B \leq S$, a subgroup isomorphic to $C_p \times C_p$. It is readily seen that $C_S(E) = E$ and that $P = N_S(E)$ is isomorphic to a semidirect product $(C_p)^2 \rtimes C_p$, the unique non-abelian group of order p^3 and exponent p. Let G_1 be the semidirect product $G_1 = E \rtimes \mathrm{Aut}(E)$. Since the Sylow p-subgroups of $\mathrm{Aut}(E)$ are cyclic of order p, there is an isomorphism between P and a Sylow p-subgroup of G_1 that extends the inclusion of E.

By construction, the fusion system $\mathscr{F}$ for the free product with amalgamation $S *_P G_1$ is Alperin in the sense of [Robinson 2006], but this fusion system is not saturated. For example, there are nonidentity self-maps of Z inside $\mathscr{F}$, and if $\mathscr{F}$ were saturated, any self-map of Z inside $\mathscr{F}$ would extend to a self-map of S. But in $\mathscr{F}$, S has only inner automorphisms, and these restrict to Z as the identity.

5. Proofs

Proof of Theorem 7. As in the statement, let Σ be the group of all permutations of S, and identify S with a subgroup of Σ. Let P and Q be subgroups of $S \leq \Sigma$, and let $\phi : P \to Q$ be any injective group homomorphism. It suffices to show that there is some $\sigma \in \Sigma$ such that for all $u \in P$, $\sigma^{-1}u\sigma = \phi(u)$. Let Ω denote the group S viewed as a set with a left S-action. There are two ways to view Ω as a set with a left P-action, via $P \leq S$ and via $\phi : P \to Q \leq S$. Denote these two P-sets by Ω and $^\phi\Omega$ respectively. Each of Ω and $^\phi\Omega$ is isomorphic as a P-set to the disjoint union of $|S : P|$ copies of P. In particular, there is an isomorphism of P-sets $\sigma : {}^\phi\Omega \to \Omega$. Viewing σ as an element of Σ, one has that $\sigma\phi(u)\omega = u\sigma\omega$ for all $u \in P$ and $\omega \in \Omega$. Hence $\sigma^{-1}u\sigma = \phi(u)$ for all u as required. $\square$

Remark 17. A version of Theorem 7 appeared in [Leary et al. 1997], although fusion systems were not mentioned there.

Before proving Theorem 2 we give a result concerning extending group homomorphisms, and two corollaries, one of which will be used in the proof.

Lemma 18. *Let S and G be as in the statement of Theorem 2, let $j : S \to G$ be the natural map from S to G, let H be a group and let $f : S \to H$ be a group homomorphism. There is a group homomorphism $\tilde{f} : G \to H$ with $f = \tilde{f} \circ j$ if and only if for each i, the homomorphisms $f : P_i \to H$ and $f \circ \phi_i : P_i \to H$ differ by an inner automorphism of H.*

Proof. Given a homomorphism $\tilde{f}$ as in the statement, one has that for each i and for each $u \in P_i$, $f\phi_i(u) = h_i^{-1} f(u) h_i$, where $h_i = \tilde{f}(t_i)$. For the converse, suppose that there exists, for each i, an element h_i satisfying the equation $f\phi_i(u) = h_i^{-1} f(u) h_i$ for all $u \in P_i$. In this case one may define $\tilde{f}$ on the generators of G by $\tilde{f}(s) = f(s)$ for all $s \in S$ and $\tilde{f}(t_i) = h_i$. $\qquad\square$

Corollary 19. *With notation as in the statement of Theorem 2, there is a homomorphism from G to Σ, the group of all permutations of the set S, extending the Cayley representation of S.*

Proof. The argument used in the proof of Theorem 7 shows that the conditions of Lemma 18 hold. $\qquad\square$

Remark 20. Corollary 19 gives an alternative way to prove Corollary 26, at least in the special case of a rose-shaped graph.

Corollary 21. *With notation as in the statement of Theorem 2, a complex representation of S with character χ extends to a complex representation of G if and only if for each i and for each $u \in P_i$, $\chi(u) = \chi(\phi_i(u))$.*

Remark 22. Of course, a representation of S will extend to G in many different ways if it extends at all.

Proof of Theorem 2. As in Appendix 6.2, one sees that the group G presented in the statement is the fundamental group of a graph of groups with one vertex group, S, and one edge group P_i for each ϕ_i, $1 \le i \le r$. From Corollary 26 it follows that S is a subgroup of G. From Corollary 30, it follows that any finite subgroup of G, and in particular any p-subgroup of G, is conjugate to a subgroup of S. By Theorem 28, there is a cellular action of G on a tree T, with one orbit of vertices and r orbits of edges. By suitable choice of orbit representatives, we may choose a vertex v whose stabilizer is S, and edges $e_1, \ldots, e_r$ so that the stabilizer of e_i is P_i, and so that the initial vertex of e_i is v while the final vertex is $t_i \cdot v$.

Since every p-subgroup of G is conjugate to a subgroup of S, there is a fusion system $\mathcal{F}_S(G)$ associated to G. By construction $\mathcal{F}_S(G)$ contains each ϕ_i, which corresponds to conjugation by t_i.

Conversely, suppose that $g \in G$ has the property that $g^{-1}Pg \le Q$ for some subgroups P, Q of S. It suffices to show that conjugation by g, as a map from P to Q, is equal to a composite of (restrictions of) the maps ϕ_j and their inverses with conjugation maps by elements of S.

Consider the action of P on the tree T. By hypothesis, the action of P fixes both the vertex v and the vertex $g \cdot v$. Since T is a tree, P must fix all the vertices and edges on the unique shortest path from v to $g \cdot v$. Let this path have length n. Define $g_0 = 1_G$, $g_n = g$, and for $1 \le i \le n - 1$, choose $g_i \in G$ so that $g_0 \cdot v, g_1 \cdot v, \ldots, g_n \cdot v$ is the shortest path in T from v to $g \cdot v$. For each i, P is contained in the stabilizer of the vertex $g_i \cdot v$, and so $P \le g_i S g_i^{-1}$, or equivalently $g_i^{-1} P g_i \le S$.

The edge joining $g_i \cdot v$ and $g_{i+1} \cdot v$ is an edge of the form $g_i \cdot e_j$ or $g_{i+1} \cdot e_j$ for some j depending on i. Consider the two cases separately, first supposing that the edge is of the form $g_i \cdot e_j$. In this case it follows that $P \le g_i P_i g_i^{-1}$, since P stabilizes the edge $g_i \cdot e_j$. Also one sees that $g_{i+1} \cdot v = g_i t_j \cdot v$, and hence $g_{i+1}^{-1} g_i t_j \in S$. Hence conjugation by $g_i^{-1} g_{i+1}$, viewed as a map from $g_i^{-1} P g_i$ to $g_{i+1}^{-1} P g_{i+1}$ is equal to the composite of the map ϕ_j (restricted to $g_i^{-1} P g_i \le P_i$) followed by conjugation by an element of S.

The other case is similar. Here it follows that $P \le g_{i+1} P_{i+1} g_{i+1}^{-1}$, and one has that $g_i \cdot v = g_{i+1} t_j \cdot v$, from which $g_i^{-1} g_{i+1} t_j = s \in S$. In this case conjugation by $g_i^{-1} g_{i+1}$, as a map from $g_i^{-1} P g_i$ to $g_{i+1}^{-1} P g_{i+1}$, is equal to the composite map given by conjugation by s followed by the map ϕ_j^{-1} (restricted to $s^{-1} g_i^{-1} P g_i s \le \phi_j(P_{i+1})$).

Thus conjugation by $g = g_n$ as a map from P to Q can be expressed as a composite of maps inside the fusion system generated by the ϕ_i, and so $\mathcal{F}_S(G)$ is equal to this fusion system.

It remains to show that the group G contains a free normal subgroup of index at most $|S|!$. Let Σ denote the symmetric group on the set S. By Corollary 19, there is a homomorphism $G \to \Sigma$ which extends the natural injection $S \to \Sigma$. By Corollary 31, the kernel of this homomorphism is a free normal subgroup of G, and its index is a factor of $|\Sigma| = |S|!$. □

Proof of Corollary 3. Define a function f on the union of S and the generators of T by $f(s) = s$ and $f(t_i) = h_i$. This extends uniquely to a group homomorphism f from G to H by an argument similar to that used in the proof of Lemma 18. Since the pairs (G, S) and (H, S) both realise the same fusion system, it is immediate that f_* is an isomorphism of fusion systems. □

Proof of Corollary 4. For $\mathcal{F}$ a fusion system on a p-group S, let $\Phi(\mathcal{F})$ be the (finite) set of all morphisms in $\mathcal{F}$, and define $G_m(\mathcal{F})$ to be the group constructed as

in Theorem 2 using the set $\Phi(\mathcal{F})$ as the chosen generators for $\mathcal{F}$. Any morphism of fusion systems $\alpha : \mathcal{F} \to \mathcal{F}'$ will give rise to a function from $\Phi(\mathcal{F})$ to $\Phi(\mathcal{F}')$ and hence a group homomorphism from $G_m(\mathcal{F})$ to $G_m(\mathcal{F}')$. Hence the map sending $\mathcal{F}$ to the pair $(G_m(\mathcal{F}), S)$ is a functor from fusion systems to pairs. It is easily checked that the fusion system on S realised by $G_m(\mathcal{F})$ is equal to $\mathcal{F}$, which shows that this functor is a preinverse to the realisation functor. $\qquad\square$

Sketch of proof of Theorem 5. In this case, the group G is the fundamental group of a star-shaped graph of groups, with one central vertex labelled S and r outer vertices labelled $G_1, \ldots, G_r$. The edge from G_i to S is labelled by the group S_i'. By Theorem 28, there is a cellular action of G on a tree T, with $r+1$ orbit of vertices and r orbits of edges. We may choose orbit representatives $v_0, v_1, \ldots, v_r$ of vertices and $e_1, \ldots, e_r$ of edges so that the stabilizer of v_0 is S, and for $1 \le i \le r$, the stabilizer of v_i is G_i (resp. of e_i is S_i'). Moreover, we may assume that e_i has initial vertex v_i and terminal vertex v_0.

In this case, one sees that any finite subgroup of G is conjugate to either a subgroup of S or to a subgroup of G_i for some i. Since S_i' is a Sylow p-subgroup of G_i, any p-subgroup of G is conjugate to a subgroup of S as required.

As in the previous proof, it is clear that the fusion system $\mathcal{F}_S(G)$ contains the image of each $\mathcal{F}_{S_i'}(G_i)$, but an argument is needed to show that these images generate $\mathcal{F}_S(G)$. Given $g \in G$ and $P, Q \le S$ so that $g^{-1}Pg \le Q$, one argues that the action of P fixes the vertices v_0 and $g \cdot v_0$ in the tree T, and hence fixes the shortest path (necessarily of even length, say $2n$) that joins these vertices.

Let $g_0 = 1_G$, $g_{2n} = g$, and pick group elements so that the vertices on the shortest path from v_0 to $g \cdot v_0$ are

$$g_0 \cdot v_0, \ g_1 \cdot v_{j(1)}, \ g_2 \cdot v_0, \ g_3 \cdot v_{j(2)}, \ \ldots, \ g_{2n-1} \cdot v_{j(n)}, \ g_{2n} \cdot v_0,$$

for some function $j : \{1, \ldots, n\} \to \{1, \ldots, r\}$. If i is even, then $g_i^{-1}Pg_i \le S$, and if i is odd then $g_i^{-1}Pg_i \le G_{j((i+1)/2)}$. Since P stabilizes each edge, one sees that $P \le g_i^{-1}S_kg_i$, where S_k denotes the image of S_k' inside S, and $k = j((i+1)/2)$ if i is odd and $k = j(i/2)$ if i is even. In particular, each $g_i^{-1}Pg_i$ is a subgroup of S.

One may show that in the case when i is odd, $g_i^{-1}g_{i+1} \in G_{j((i+1)/2)}$ and that in the case when i is even, $g_i^{-1}g_{i+1} \in S$. Thus the map from $g_i^{-1}Pg_i$ to $g_{i+1}^{-1}Pg_{i+1}$ given by conjugation by $g_i^{-1}g_{i+1}$ is a map inside the fusion system generated by the images of the $\mathcal{F}_{S_i'}(G_i)$, and conjugation by $g = g_{2n}$ as a map from P to $Q \le S$ is expressed as a composite of maps of the required form.

Finally, if Ω is a finite set so that $|\Omega|$ is divisible by $|S|$ and by each $|G_i|$, one may define free actions of S and each G_i on Ω which give rise to the same (free) action of $S_i = f_i(S_i')$. This gives rise to a group homomorphism from G to Σ, the symmetric group on Ω, whose kernel is free by Corollary 31. $\qquad\square$

6. Appendix: graphs of groups

In this section we give proofs of those results about graphs of groups that we use. Our treatment of graphs of groups is topological and follows that of Scott and Wall [1979]; an alternative, more algebraic, treatment of this subject can be found in [Serre 1980]. There is no direct correspondence between the two treatments but we give references to the closest results following Serre's approach.

For the purposes of this paper, a graph Γ consists of two sets, the vertices V and the directed edges E, together with two functions $\iota, \tau : E \to V$. For $e \in E$, $\iota(e)$ is called the initial vertex of e and $\tau(e)$ is the terminal vertex of e. Multiple edges and loops are allowed in this definition. Γ is connected if the only equivalence relation on V that contains all pairs $(\iota(e), \tau(e))$ is the relation with just one class.

A graph Γ may be viewed as a category, with objects the disjoint union of V and E and two nonidentity morphisms with domain e for each $e \in E$, one morphism $e \to \iota(e)$ and one morphism $e \to \tau(e)$.

A graph of groups is a connected graph Γ together with groups G_v, G_e for each vertex and edge, and injective group homomorphisms $f_{e,\iota} : G_e \to G_{\iota(e)}$ and $f_{e,\tau} : G_e \to G_{\tau(e)}$ for each edge e. If Γ is viewed as a category, this is just a functor from Γ to the category of groups and injective group homomorphisms. Without loss of generality, one may assume that each map $f_{e,\iota} : G_e \to G_{\iota(e)}$ is the inclusion of a subgroup.

6.1. *The fundamental group of a graph of groups.* For a topologist, and arguably for anybody, the easiest way to define the fundamental group of a graph of groups is via the notion of a graph of spaces.

A graph of spaces is a connected graph Γ together with topological spaces X_v, X_e for each vertex and edge, and continuous maps $f_{e,\iota} : X_e \to X_{\iota(e)}$ and $f_{e,\tau} : X_e \to X_{\tau(e)}$. A graph of spaces is just a functor from the category Γ to the category of topological spaces and continuous functions. A graph of based spaces is defined similarly: each X_e and X_v is equipped with a base point, and the maps must preserve base points. Let I denote the closed unit interval $[0, 1]$. The total space of a graph of spaces is the space X made from the disjoint union

$$\coprod_{v \in V} X_v \sqcup \coprod_{e \in E} X_e \times I$$

by identifying $(x, 0) \in X_e \times I$ with $f_{e,\iota}(x) \in X_{\iota(e)}$ and identifying $(x, 1) \in X_e \times I$ with $f_{e,\tau}(x) \in X_{\tau(e)}$. As an example, consider the graph of spaces in which each X_e and X_v is a single point. For this graph of spaces the total space is the usual topological realization of the graph as a 1-dimensional CW-complex. The reader who is familiar with the homotopy colimit construction will note that if one views a graph of spaces as a functor $X_{(-)}$ on the category Γ, then the total space

X is naturally homeomorphic to the homotopy colimit of the functor $X_{(-)}$, or in symbols, $X = \mathrm{hocolim}_\Gamma X_{(-)}$.

Given a graph of groups, one may define a graph of connected based spaces by taking classifying spaces as the spaces $X_e = BG_e = K(G_e, 1)$ and $X_v BG_v = K(G_v, 1)$. For the continuous map $f_{e,\iota} : X_e \to X_{\iota(e)}$ (resp. $f_{e,\tau} : X_e \to X_{\tau(e)}$) one may take any continuous map that induces the given map $G_e \to G_{\iota(e)}$ (resp. $G_e \to G_{\tau(e)}$) on fundamental groups. Define a total space X as the realization of this graph of spaces.

For discrete groups K and H, the space BK is unique up to based homotopy, and homotopy class of based maps from BK to BH are in bijective correspondence with group homomorphisms from K to H. It follows that the homotopy type of the space X defined above depends only on the graph of groups, rather than on the particular choices of classifying spaces and maps between them. The fundamental group G of the graph of groups can now be defined as the fundamental group of X. This describes the fundamental group of the graph of groups up to isomorphism. The inclusion of each X_v in X defines a conjugacy class of homomorphism $G_v \to G$ (which will be shown to be injective, below). For many purposes one wants a more precise description of G, together with a single choice of homomorphism $G_v \to G$. This can be done by choosing a basepoint for the space X, and for each v, a path in X from the basepoint for X to the basepoint for $X_v \subseteq X$.

6.2. *Presentations for graphs of groups.* We shall only consider presentations for graphs of groups where the underlying graph is either a *rose*, meaning a graph with only one vertex (so every edge has the same initial and terminal vertices) or a *star*, which is a connected graph with $n + 1$ vertices and n edges, for some $n > 0$, with one central vertex, such that all the edges have this vertex as their terminal vertex and every other vertex is the initial vertex of exactly one edge.

Suppose one is given a p-group S, subgroups $P_i, Q_i \leq S$, and injective group homomorphisms $\phi_i : P_i \to Q_i$ for $1 \leq i \leq r$, as in the statement of Theorem 2. Use this data to make a rose-shaped graph of groups with r edges. Let S be the vertex group, let P_i be the ith edge group, with the inclusion map $P_i \leq S$ (resp. the composite $\phi_i : P_i \to Q_i \leq S$) as the ith initial (resp. terminal) homomorphism. There is a model for BP_i having just one 0-cell and one 1-cell for each element of P_i. Take a model for BS having just one 0-cell and take this 0-cell as the base point. To make a CW-complex of the homotopy type of the total space of the graph of groups, it suffices to add to BS one 1-cell t_i for each i (with both ends at the unique 0-cell), one 2-cell $D_{i,u}$ for $1 \leq i \leq r$ and for each $u \in P_i$, and higher dimensional cells (which will not affect the fundamental group). The attaching map for the 2-cell $D_{i,u}$ spells out the word $u\, t_i\, \phi_i(u)\, t_i^{-1}$, and so the presentation coming from this CW-structure is the presentation given in the statement of Theorem 2.

Next suppose that one is given a p-group S, groups G_i for $1 \leq i \leq r$ with Sylow p-subgroups S_i, and injective group homomorphisms $f_i : S_i \to S$, i.e., the data found in the statement of Theorem 5. In this case, define a star of groups with central vertex group P, other vertex groups $G_1, \ldots, G_r$, and edge groups $S_1, \ldots, S_r$. The map of each edge group into its initial vertex group is the inclusion $S_i \to G_i$, and the map of each edge group into its terminal vertex group is $f_i : S_i \to S$. An argument similar to that given in the previous paragraph shows that the fundamental group of this graph of groups has the presentation given in the statement of Theorem 5. Note that here one can make a space homotopy equivalent to the total space of the graph of spaces by starting from the one-point union of BS and the BG_i, without adding any extra 1-cells. This is reflected in the fact that the vertex groups generate the fundamental group of the graph of groups.

6.3. *Properties of graphs of groups.*

Proposition 23. *Let G be the fundamental group of a graph of groups based on a graph Γ. Every subgroup $H \leq G$ is itself the fundamental group of a graph of groups, indexed by a graph Δ equipped with a map $f : \Delta \to \Gamma$ which does not collapse any edges. For each v and $e \in \Delta$, the group H_v (resp. H_e) is a subgroup of $G_{f(v)}$ (resp. $G_{f(e)}$).*

This proposition appears as [Scott and Wall 1979, Theorem 3.7] in the special case when the graph is either an interval or a loop, i.e., the case when the fundamental group of the graph of groups is either a free product with amalgamation or an HNN extension.

Proof. Use the bijection between connected covering spaces of a connected CW-complex (with a choice of base point) and subgroups of its fundamental group. Let X be the total space of the graph of spaces used in the definition of G, so that there is a covering space of X whose fundamental group is H. Any connected covering space of X can be expressed as the total space of a graph of spaces indexed by some Δ as in the statement. This gives an expression for the fundamental group of any connected covering space of X as the fundamental group of a graph of groups as claimed. $\square$

Theorem 24. *Let X be the total space of the graph of spaces used in the definition of the fundamental group G of a graph of groups. The universal covering space of X is contractible, and hence X is homotopy equivalent to BG.*

Proof. We shall build a space Y in such a way that it is clear that Y is contractible and a covering space of X. For v a vertex, define the subspace X'_v of X by

$$X'_v = X_v \cup \bigcup_{\iota(e)=v} X_e \times [0, 0.5) \cup \bigcup_{\tau(e)=v} X_e \times (0.5, 1].$$

Similarly, define for e an edge, $X'_e = X_e \times (0, 1)$. The inclusions $X_v \to X'_v$ and $X_e \cong X_e \times \{0.5\} \to X'_e$ are homotopy equivalences, and it may be useful to think of X'_v as a nice open neighbourhood of X_v in X. Let Y_v, Y'_v, Y_e, and Y'_e be the universal covering spaces of X_v, X'_v, X_e and X'_e respectively. Each Y'_v (resp. Y'_e) is contractible since it is the universal covering space of the classifying space BG_v (resp. BG_e).

The definition of the space X'_v lifts to a description of the space Y'_v. The complement $Y'_v - Y_v$ is identified with a collection of disjoint copies of $Y_e \times (0, 0.5)$, and $Y_e \times (0.5, 1)$, for different edges e. There are copies of $Y_e \times (0, 0.5)$ if and only if $\iota(e) = v$. In this case the copies are in bijective correspondence with the cosets of $f_{e,\iota}(G_e)$ in G_v. Similarly, there are copies of $Y_e \times (0.5, 1)$ for each e with $\tau(e) = v$, and these copies are indexed by cosets of $f_{e,\tau}(G_e)$ in G_v.

By induction, we shall construct a sequence $Y_0 \subseteq Y_1 \subseteq Y_2 \cdots$ of spaces so that: each Y_n is contractible; there is a map $\pi : Y_n \to X$ which is locally a covering map except at some points of X; for any $x \in X$ and any $n \geq 0$, at least one of $\pi : Y_n \to X$ and $\pi : Y_{n+1} \to X$ is locally a covering map at x.

Pick a vertex v of the graph Γ, and define Y_0 to be the space Y'_v. Define a map $\pi : Y'_v \to X$ as the composite of the map $Y'_v \to X'_v$ and the inclusion $X'_v \subseteq X$. As remarked earlier, $Y'_v - Y_v$ consists of lots of subspaces of the form $Y_e \times (0, 0.5)$ for $\iota(e) = v$ and lots of subspaces of the form $Y_e \times (0.5, 1)$ for $\tau(e) = v$. Define Y_1 by attaching to each such subspace a copy of Y'_e. The map $\pi : Y_0 \to X$ extends uniquely to $\pi : Y_1 \to X$ by insisting that on each newly-added Y'_e subspace, π is equal to the composite map $Y'_e \to X'_e \subseteq X$. From the construction of Y_1, it is apparent that Y_1 is contractible.

In constructing Y_1, one attached to Y_0 many spaces of the form Y'_e, by identifying one end of Y'_e with part of Y_0. For each copy of Y'_e that was attached via its initial end, take a copy of $Y'_{\tau(e)}$, and attach this at the other end of Y'_e. Similarly, for each copy of Y'_e that was attached to Y_0 by its terminal end, take a copy of $Y'_{\iota(e)}$ and attach this at the other end of Y'_e. This defines a space Y_2, which is clearly contractible, and the map π extends uniquely to a map $Y_2 \to X$ which agrees with the covering map $Y'_e \to X'_e$ or $Y'_v \to X'_v$ on each such subspace.

Now suppose that n is even, and that Y_n has been constructed from Y_{n-1} by attaching subspaces Y'_v in such a way that the intersection of Y_{n-1} and each new Y'_v is equal to one of the components of $Y'_v - Y_v$. Furthermore, suppose that the map π on each new Y'_v is equal to the map $Y'_v \to X'_v \subseteq X$. Form Y_{n+1} by attaching a copy of Y'_e to each other component of $Y'_v - Y_v$ for each of the copies of Y'_v. Extend the map π as before.

In the case when n is odd, suppose that Y_n has been constructed from Y_{n-1} by attaching subspaces Y'_e in such a way that the intersection of Y_{n-1} and each new Y'_e is equal to one of the two components of $Y'_e - Y_e \times \{0.5\}$. Suppose also that

the map π on each of the new Y_e' is equal to the map $Y_e' \to X_e' \subseteq X$. Form Y_{n+1} by attaching a copy of Y_v' to the other component of each $Y_e' - Y_e \times \{0.5\}$, where v is either $\iota(e)$ or $\tau(e)$ depending which component of $Y_e' - Y_e \times \{0.5\}$ was used. Extend the map π in the same way as before.

By construction, each Y_n is contractible, and comes equipped with a map $\pi : Y_n \to X$. If n is even, this map is locally a covering except possibly at points of X contained in the union of the images of the X_v. If n is odd, this map is a covering except possibly at point of X contained in the union of the images of the $X_e \times \{0.5\}$. Now define Y by $Y = \bigcup_n Y_n$. This space Y is contractible, and the map $\pi : Y \to X$ is a covering map, since it is locally a covering map at every point of X. It follows that Y is the universal covering space of X. Since the universal covering space of X has been shown to be contractible, it follows that X is a model for BG. $\square$

Remark 25. This proof relies on the fact that the edge groups map injectively to the vertex groups. The theorem can be found in [Scott and Wall 1979, Proposition 3.2 (ii)]. There is no direct analogue in the more algebraic treatment of [Serre 1980]. The closest result to this theorem is arguably Serre's Theorem 12.

Corollary 26. *Each vertex group G_v maps injectively into the fundamental group of a graph of groups.*

Proof. Given a vertex v, construct the universal covering space as in the proof of Theorem 24, with $Y_0 = Y_v'$. The group of all deck transformations of Y is naturally isomorphic to G, the fundamental group of X. Under this isomorphism, the subgroup of those deck transformations that preserve Y_0 maps to G_v. $\square$

Remark 27. Corollary 26 is [Scott and Wall 1979, Proposition 3.2(i)]. There is also an algebraic proof that each G_v embeds in G; see for instance [Serre 1980, Corollary 1 to Theorem 11]. In the case when the graph is a rose, this argument is given in Corollary 19.

6.4. *The action on a tree.* Say that an action of a group on a tree is cellular if no element of the group exchanges the ends of any edge. The following theorem is implicit in [Scott and Wall 1979, Section 4].

Theorem 28. *Let G be the fundamental group of a graph of groups indexed by the graph Γ. There is a tree T with a cellular G-action and an isomorphism $f : T/G \cong \Gamma$. If $\tilde{x}$ is either a vertex or edge of T, and $x = f(\tilde{x})$ is the image of $G \cdot \tilde{x}$ under f, then the stabilizer of $\tilde{x}$ is conjugate to G_x.*

Proof. Let X be the total space of the graph of spaces used in defining G. As remarked earlier, the underlying topological space of the graph Γ can be identified with the total space of the constant graph of 1-point spaces indexed by Γ. The unique map from each X_v and X_e to a point induces a map from X to Γ.

Now let Y be the universal covering space of X, as constructed in the proof of Theorem 24. This Y can be viewed as a graph of spaces over some graph Δ, with vertex spaces copies of the spaces Y_v and edges spaces copies of the spaces Y_e. The group G acts on Y in such a way that the setwise stabilizer of each copy of Y_v is a conjugate of G_v, and similarly the setwise stabilizer of each copy of $Y_e \times (0, 1)$ is a conjugate of G_e. Define T to be the total space of the graph of 1-point spaces over the graph Δ. By construction, T is a graph equipped with a G-action, an equivariant map $\phi : Y \to T$, and an isomorphism $f : T/G \to \Gamma$. To check that T is a tree, let $T_n = \phi(Y_n)$. As in the proof of Theorem 24, one shows inductively that T_n is contractible, and $T = \bigcup_n T_n$. □

Lemma 29. *Any cellular action of a finite group H on a tree T fixes a vertex.*

Proof. Take any vertex $t \in T$, and define a finite subtree T' to be the union of all the shortest paths between elements of the orbit $H \cdot t$. If T' is not itself fixed by H, remove an H-orbit of 'leaves' (i.e., vertices of valency one) from T', and continue this process until a subtree fixed by H is all that remains. □

As a consequence of the previous two results we get a very useful corollary, which is stated as [Serre 1980, Corollary to Theorem 8] in the special case of an interval of groups. (This is the case when the fundamental group of the graph of groups is a free product with amalgamation.)

Corollary 30. *Every finite subgroup of the fundamental group of a graph of groups is conjugate to a subgroup of a vertex group.*

Proof. Let G be the fundamental group of the graph of groups and let T be the corresponding tree. If H is a finite subgroup of G then H fixes some vertex of T. The stabilizer of each vertex of T is a conjugate of one of the vertex groups G_v. □

The following corollary is stated as [Serre 1980, Proposition 18] in the special case of an interval of groups.

Corollary 31. *Let H be a subgroup of a graph of groups whose intersection with each conjugate of each vertex group is trivial. Then H is a free group.*

Proof. The hypotheses imply that H acts freely on the tree T, and so the quotient space T/H is a 1-dimensional classifying space for H. □

Acknowledgements

References

[Alperin 1967] J. L. Alperin, "Sylow intersections and fusion", *J. Algebra* **6** (1967), 222–241. MR 35 #6748 Zbl 0168.27202

[Alperin and Broué 1979] J. Alperin and M. Broué, "Local methods in block theory", *Ann. of Math.* (2) **110**:1 (1979), 143–157. MR 80f:20010 Zbl 0416.20006

[Aschbacher and Chermak 2005] M. Aschbacher and A. Chermak, "A group-theoretic approach to a family of 2-local finite groups constructed by Levi and Oliver", preprint, 2005.

[Benson 1994] D. Benson, "Conway's group Co$_3$ and the Dickson invariants", *Manuscripta Math.* **85**:2 (1994), 177–193. MR 95h:55018 Zbl 0853.55018

[Broto et al. 2003] C. Broto, R. Levi, and B. Oliver, "The homotopy theory of fusion systems", *J. Amer. Math. Soc.* **16**:4 (2003), 779–856. MR 2004k:55016 Zbl 1033.55010

[Broto et al. 2005] C. Broto, N. Castellana, J. Grodal, R. Levi, and B. Oliver, "Subgroup families controlling p-local finite groups", *Proc. London Math. Soc.* (3) **91**:2 (2005), 325–354. MR 2007e: 20111 Zbl 02212604

[Goldschmidt 1970] D. M. Goldschmidt, "A conjugation family for finite groups", *J. Algebra* **16** (1970), 138–142. MR 41 #5489 Zbl 0198.04306

[Kessar and Stancu 2007] R. Kessar and R. Stancu, "A reduction theorem for fusion systems of blocks", *J. Algebra* (2007).

[Leary et al. 1997] I. J. Leary, B. Schuster, and N. Yagita, "On universally stable elements", *Quart. J. Math. Oxford Ser.* (2) **48**:192 (1997), 493–498. MR 99b:20084 Zbl 0901.20041

[Levi and Oliver 2002] R. Levi and B. Oliver, "Construction of 2-local finite groups of a type studied by Solomon and Benson", *Geom. Topol.* **6** (2002), 917–990. MR 2003k:55016 Zbl 1021.55010

[Linckelmann 2006] M. Linckelmann, "Simple fusion systems and the Solomon 2-local groups", *J. Algebra* **296**:2 (2006), 385–401. MR 2006i:20024 Zbl 05024287

[Puig 2002] L. Puig, "Full Frobenius systems and their localizing categories", preprint, 2002.

[Robinson 2006] G. Robinson, "Amalgams, blocks, weights, fusion systems and finite simple groups", preprint, 2006, Available at www.maths.abdn.ac.uk/~grr/bio/fusionsubmit.pdf.

[Scott and Wall 1979] P. Scott and C. T. C. Wall, "Topological methods in group theory", pp. 137–203 in *Homological group theory* (Durham, 1977), edited by C. T. C. Wall, London Math. Soc. Lecture Note Ser. **36**, Cambridge Univ. Press, Cambridge, 1979. MR 81m:57002 Zbl 0423.20023

[Serre 1980] J.-P. Serre, *Trees*, Springer, Berlin, 1980. MR 82c:20083 Zbl 0548.20018

[Solomon 1974] R. Solomon, "Finite groups with Sylow 2-subgroups of type 3", *J. Algebra* **28** (1974), 182–198. MR 49 #9077 Zbl 0293.20022

leary@math.ohio-state.edu   *Department of Mathematics, The Ohio State University, 231 W 18th Avenue, Columbus, OH 43210, United States*

stancu@math.ohio-state.edu   *Department of Mathematics, The Ohio State University, 231 W 18th Avenue, Columbus, OH 43210, United States*

A topological quantum field theory of intersection numbers on moduli spaces of admissible covers

Renzo Cavalieri

We construct a two-level weighted topological quantum field theory whose structure coefficients are equivariant intersection numbers on moduli spaces of admissible covers. Such a structure is parallel (and strictly related) to the local Gromov–Witten theory of curves of Bryan and Pandharipande. We compute explicitly the theory using techniques of localization on moduli spaces of admissible covers of a parametrized $\mathbb{P}^1$. The Frobenius algebras we obtain are one-parameter deformations of the class algebra of the symmetric group S_d. In certain special cases we are able to produce explicit closed formulas for such deformations in terms of the representation theory of S_d.

Introduction

We study a large class of (equivariant) intersection numbers on moduli spaces of admissible covers. For a smooth algebraic curve X, ramified covers of X of a given degree by smooth curves of a given genus are parametrized by moduli spaces called Hurwitz schemes. A smooth compactification of a Hurwitz scheme can be obtained by allowing suitable degenerations, called admissible covers.

Moduli spaces of admissible covers were introduced in [Harris and Mumford 1982]. Intersection theory on these spaces was for a long time hard and mysterious, mostly because they are in general not normal, even if the normalization is always smooth. It was only recently that Abramovich, Corti and Vistoli [Abramovich et al. 2003] exhibited this normalization as the stack of balanced stable maps of degree 0 from twisted curves to the classifying stack $\mathcal{B}S_d$. This way they attained both the smoothness of the stack and a nice moduli-theoretic interpretation of it. We abuse terminology and refer to Abramovich–Corti–Vistoli spaces as *admissible covers*.

At about the same time, Ionel [2002] developed a parallel theory in the symplectic category and used push-pull techniques on admissible covers to produce new relations in the tautological ring of $M_{g,n}$. (See also [Ionel 2005].)

In [Graber and Vakil 2003b], admissible cover loci within the boundary of moduli spaces of stable maps play a key role in establishing the result that the degree $3g - 3$ part of the tautological ring of $\overline{M}_g$ has dimension 1, providing further evidence for a conjecture by Faber, stating that $R(\overline{M}_g)$ is a Gorenstein algebra with socle in degree $3g - 3$.

Bryan, Graber and Pandharipande have shown in [Bryan et al. 2005] that the orbifold Gromov–Witten potential of a Gorenstein orbifold can be computed in terms of intersection theory on moduli spaces of admissible covers. With a subtle use of WDVV techniques, they are able to explicitly compute the Gromov–Witten potential for the orbifold $[\mathbb{C}^2/\mathbb{Z}_3]$. Such a computation provides evidence for the crepant resolution conjecture [Bryan and Graber ≥ 2008].

We give a few basic definitions and a working description of moduli spaces of admissible covers in Section 1.

For all choices of:

- an r-pointed curve $(X, p_1, \ldots, p_n)$;
- a rank two vector bundle $N = L_1 \oplus L_2$ on X, endowed with a natural $\mathbb{C}^* \times \mathbb{C}^*$ action (page 46);
- a vector of partitions $\underline{\eta} = (\eta_1, \ldots, \eta_n)$ of a fixed integer d,

we describe the invariants

$$A_d^h(N) := \int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} X, (\eta_1 p_1, \ldots, \eta_r p_r))} e^{\mathrm{eq}}(-R^\bullet \pi_* f^*(L_1 \oplus L_2)).$$

The motivation for studying these invariants is twofold. They are natural and interesting intersection numbers on their own, giving rise to a beautiful structure. Secondly, in the context of Gromov–Witten theory, invariants of this form are known as "local" invariants: roughly speaking, they represent the contribution to the Gromov–Witten invariants of a threefold given by rigid curves.

Theorem 3.1. (See page 48.) *The invariants $A_d^h(N)$ can be organized to be the structure coefficients of a $2-$level, semisimple, weighted topological quantum field theory (TQFT).*

Section 2 is dedicated to presenting these structures to the unfamiliar reader, while in Section 3 the specific TQFT $\mathcal{U}$ is constructed.

The generators for the TQFT are explicitly computed in Section 4. The techniques involved are basic dimension counting, reduction to classical intersection theory on moduli spaces of curves, and Atiyah–Bott localization on moduli spaces of admissible covers of a parametrized $\mathbb{P}^1$.

An interesting feature of this theory is that the degree 0 part is constructed from Hurwitz numbers. The embedded (see page 44) Frobenius algebras induced on

the Hilbert space by $\mathcal{U}$ are one-parameter deformations of the class algebra of the symmetric group, whose TQFT-theoretic description in terms of Hurwitz numbers was studied in the 1990s in [Dijkgraaf and Witten 1990] and [Freed and Quinn 1993]. An explicit description of such deformations is in general quite complicated. By specializing to the antidiagonal action of $\mathbb{C}^*$ inside $\mathbb{C}^* \times \mathbb{C}^*$, it is possible to diagonalize the theory: closed formulas for our invariants and for the deformation are described in Section 5 in terms of the representation theory of the symmetric group S_d (Theorem 5.2).

This work is closely connected to and follows recent work of Jim Bryan and Rahul Pandharipande [2004; 2005], describing the local Gromov–Witten theory of curves.

There, analogous intersection numbers on moduli spaces of (relative) stable maps are organized in a TQFT that we denote by $\mathcal{BP}$. Theorem 4.1 shows that the two theories coincide in level $(0,0)$. In all other levels, $\mathcal{U}$ is a normalization of $\mathcal{BP}$ via appropriate powers of a universal generating function factor, which should be understood as the contribution of maps containing contracting components to the Gromov–Witten invariants.

This result, the most technical in this paper, is established by computing the genus 0, one-pointed invariants via localization, together with the use of some beautiful Hodge integral computations from [Faber and Pandharipande 2000; Ekedahl et al. 2001; Graber and Vakil 2003a]. The explicit statement is this:

Theorem 4.3 (See page 56.). *The coefficients for the one-pointed invariants of $\mathcal{U}$ in level $(0,-1)$ are given by the generating functions*

$$A_d(0|0,-1)_\eta = (-1)^{d-\ell(\eta)} \frac{\left(2\sin\frac{u}{2}\right)^d}{s_1^{\ell(\eta)} \mathfrak{z}(\eta) \prod 2\sin\frac{\eta_i u}{2}},$$

Notation. Here and throughout the paper $\ell(\eta)$ denotes the length r of a partition $\eta = (\eta_1, \ldots, \eta_r)$.

A direct check in the one-pointed case, together with the semisimplicity of both theories, yields:

Corollary 0.1. *The coefficients of the theories $\mathcal{U}$ and $\mathcal{BP}$ are related by*

$$A_d(g \mid k_1, k_2)_{\underline{\eta}} = (d!)^{k_1+k_2} s_1^{dk_2} s_2^{dk_1} \mathcal{BP}_d(g, \mid k_1, k_2)_{\underline{\eta}} \mathcal{BP}_d(0 \mid 0, -1)_{(1,\ldots,1)}^{k_1+k_2}.$$

This close proximity to Gromov–Witten theory reinforces our interest in moduli spaces of admissible covers, as it anticipates the possibility of a fertile exchange of information between the two contexts. In particular, embedded in the theory $\mathcal{U}^\circ$ (the circle superscript indicates we are restricting our attention to connected covers) we rediscover the classical result:

Aspinwall–Morrison formula.

$$\int_{[\overline{M}_{0,0}(\mathbb{P}^1,d)]} R^1\pi_* f^*(\mathcal{O}(-1)\oplus\mathcal{O}(-1)) = \left(\frac{A_d^{\circ,0}(0\mid -1,-1)}{u^{2d-2}}\right)_{|u=0} = \frac{1}{d^3}$$

The technique of Atiyah–Bott localization suits very well the spaces of admissible covers of a parametrized $\mathbb{P}^1$; the fact that these spaces are smooth (as DM stacks) requires no need for a virtual fundamental class in order to do intersection theory on them. The modularity of the boundary-fixed loci naturally produces topological recursions that live completely within the realm of admissible covers.

1. Admissible covers

Moduli spaces of admissible covers are a "natural" compactification of the Hurwitz schemes, parametrizing ramified covers of smooth Riemann Surfaces. The fundamental idea is that, in order to understand limit covers, we allow the base curve to degenerate together with the cover. Branch points are not allowed to come together; as two or more branch points tend to collide, a new component of the base curve sprouts from the point of collision, and the points transfer onto it. Similarly, upstairs the cover splits into a nodal cover.

More formally: let $(X, p_1, \ldots, p_r)$ be an r-pointed nodal curve of genus g.

Definition 1.1. An *admissible cover* $\pi : E \to X$ of degree d is a finite morphism satisfying the following:

(1) E is a nodal curve.

(2) Every node of E maps to a node of X.

(3) The restriction of $\pi : E \to X$ to $X \setminus (p_1, \ldots, p_r)$ is étale of constant degree d.

(4) Nodes can be smoothed. This means that given an admissible cover $\pi : E \to X$, and a node of E, we can find a family of admissible covers $\pi' : E' \to X'$ such that $\pi : E \to X$ is the central fiber of the family, and there are local analytic coordinates and a positive integer $n \leq d$ such that X', E' and π' are described by

$$E : e_1 e_2 = a, \quad X : x_1 x_2 = a^n, \quad \pi : x_1 = e_1^n, \ x_2 = e_2^n.$$

We recall here the notation we use in this paper, and refer the reader to [Cavalieri 2005] for a more extensive discussion.

Let $(X, p_1, \ldots, p_r)$ be as before, and let $\eta_1, \ldots, \eta_r$ be partitions of the fixed integer d. We denote by

$$\overline{\mathrm{Adm}}(h \xrightarrow{d} X, (\mu_1 p_1, \ldots, \mu_r p_r))$$

the stack of *possibly disconnected*, degree d admissible covers of the curve X by curves of genus h, such that

- the ramification profile over the base point p_i is described by the partition η_i;
- all other ramification is simple (and is not marked).

The following variations are also used:

Connected admissible covers: We add the superscript $\circ$ to restrict our attention to admissible covers by connected curves. *Admissible covers of a genus g curve:* We denote by

$$\overline{\mathrm{Adm}}(h \xrightarrow{d} g)$$

the stack of admissible covers of a curve of genus g. This means that also the base curve is allowed to vary in families.

Admissible covers of a parametrized $\mathbb{P}^1$: When we intend to fix a parametrization on the base $\mathbb{P}^1$, we write

$$\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1).$$

Moduli spaces of admissible covers admit forgetful maps to (quotients of) configuration spaces of points on the base curve (or to $\overline{M}_{g,n}$ in the case of admissible covers of a genus g curve), recording the information about branch points that are free to move. Tautological ψ classes on admissible covers are defined by pulling back the ψ classes downstairs via these maps.

There is also a natural map from a moduli space of admissible covers of genus h to the corresponding moduli space of curves $\overline{M}_h$, obtained by forgetting the cover map and only remembering the source curve. Tautological λ classes on admissible covers are defined by pulling back λ classes (the Chern classes of the Hodge bundle on the moduli space of stable curves) via these maps.

Admissible covers of a nodal curve. Admissible covers of a nodal curve can be described combinatorially in terms of admissible covers of the irreducible components of the curve. This is extremely useful because it opens the way to the use of degeneration techniques and induction. Crucial to this work are the following identities [Li 2002] taking place in the Chow ring with rational coefficients.

Reducible nodal curve: Let

$$X = X_1 \bigcup_{x_1 = x_2} X_2$$

be a nodal curve of genus g, obtained by attaching at a point two irreducible curves of genus g_1 and g_2. Then

$$[\overline{\mathrm{Adm}}(h \xrightarrow{d} X)] = \sum_{\eta, h_1, h_2} \mathfrak{z}(\eta)[\overline{\mathrm{Adm}}(h_1 \xrightarrow{d} X_1, (\eta))] \times [\overline{\mathrm{Adm}}(h_2 \xrightarrow{d} X_2, (\eta))], \quad (1\text{-}1)$$

where $\eta = ((\eta^1)^{m_1}, \ldots, (\eta^k)^{m_k})$ runs over all partitions of d, the numbers h_1 and h_2 satisfy $h_1 + h_2 + \ell(\eta) - 1 = h$, and we have defined the combinatorial factor

$$\mathfrak{z}(\eta) := \prod m_i! (\eta^i)^{m_i}. \tag{1-2}$$

In particular, $\mathfrak{z}(\eta)$ is the order of the centralizer in S_d of any group element in the conjugacy class of η.

Note. If we are dealing with admissible cover spaces with also a prescribed vector of ramification conditions μ, analogous formulas hold; the μ_i need to be distributed on the two twigs X_1 and $\overline{X}_2$ in all possible ways.

Irreducible nodal curve: Let

$$X = X'/\{x_1 = x_2\}$$

be a nodal curve of genus g, obtained by gluing two distinct points of an irreducible curve X' of genus $g - 1$. As an element in the Chow ring with rational coefficients, we can then express

$$[\overline{\mathrm{Adm}}(h \xrightarrow{d} X)] = \sum_{\eta} \mathfrak{z}(\eta)[\overline{\mathrm{Adm}}(h' \xrightarrow{d} X', (\eta, \eta))], \tag{1-3}$$

where the sum is over all partitions η of d, and h' is determined by

$$h' + \ell(\eta) = h.$$

2. Topological quantum field theories

As an excellent and elementary reference for two-dimensional topological quantum field theories in mathematics we mention [Kock 2004].

Definition 2.1. A $(1+1)$-*dimensional topological quantum field theory* is a functor of tensor categories:

$$\mathcal{T} : 2\mathsf{Cob} \longrightarrow \mathsf{Free\ Rmod}.$$

On the right-hand side is the category of free modules over a commutative ring R, and on the left is the category 2Cob described thus:

- The *objects* are one-dimensional oriented closed manifolds (finite disjoint unions of oriented circles).
- The *morphisms* are (equivalence classes of) oriented cobordisms between two objects. We can think of them as oriented topological surfaces with oriented boundary components.
- We *compose* two morphisms by concatenation; equivalently, we glue negatively oriented boundary components of one surface to positively oriented boundary components of the other.

– The *tensor operation* is the disjoint union.

The free module $H := \mathcal{T}(S^1)$ is called the *Hilbert space* of the TQFT.

All topological surfaces can be decomposed into discs, annuli, and pairs of pants. Therefore, the structure of a TQFT is completely determined if it is described on these basic building blocks.

Tensor notation. It is convenient, for explicit computations, to use tensor notation for TQFTs. We choose a basis $e_1, \ldots, e_r$ for the Hilbert space H, and denote the dual basis by $e^1, \ldots, e^r$. Let $W_m^n(g)$ be a genus-g cobordism from m to n circles. The map

$$\mathcal{T}(W_m^n(g)) : H^{\otimes m} \to H^{\otimes n}$$

can be thought of as a vector in $(H^*)^{\otimes m} \otimes H^{\otimes n}$. We denote by

$$\Gamma(W_m^n(g))_{i_1, \ldots, i_m}^{j_1, \ldots, j_n}$$

the coefficient of $\mathcal{T}(W_m^n(g))$ in the direction of the basis element $e^{i_1} \otimes \cdots \otimes e^{i_m} \otimes e_{j_1} \otimes \cdots \otimes e_{j_n}$ (see Figure 1):

$$\mathcal{T}(W_m^n(g)) = \sum \Gamma(W_m^n(g))_{i_1, \ldots, i_m}^{j_1, \ldots, j_n} e^{i_1} \otimes \cdots \otimes e^{i_m} \otimes e_{j_1} \otimes \cdots \otimes e_{j_n}.$$

Frobenius algebras. A TQFT gives the Hilbert space H the structure of a commutative Frobenius algebra. This means it defines an associative and commutative multiplication $\cdot$ and an inner product (also called the metric of the TQFT) $\langle \, , \, \rangle$ on H such that

$$\langle h_1 \cdot h_2, h_3 \rangle = \langle h_1, h_2 \cdot h_3 \rangle \tag{2-1}$$

for all h_1, h_2, h_3 in the Hilbert space H. It is easy to see how the structure is induced: multiplication is the map associated to the $(-, -, +)$ pair of pants, the inner product is the scalar map associated to the $(-, -)$ annulus. As a consequence, we see immediately that the cap with positively oriented boundary corresponds
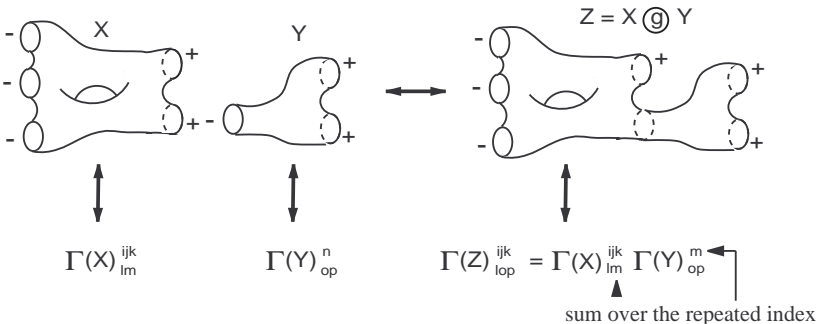

Figure 1. Gluing in tensor notation.

to the unit vector for the multiplication map just defined, whereas the $(-)$ cap corresponds to the counit operator in the Frobenius algebra.

Definition 2.2. A TQFT $\mathcal{T}$ is *semisimple* if the Frobenius algebra induced on the Hilbert space H is semisimple; equivalently, if there is an orthonormal basis $e_1, \ldots e_r$ for H such that

$$e_i \cdot e_j = \delta_{ij} e_i.$$

Yet another equivalent condition is that $\mathcal{T}$ be a direct sum

$$\mathcal{T} = \mathcal{T}_1 \oplus \cdots \oplus \mathcal{T}_r,$$

where all $\mathcal{T}_i$ are TQFTs with Hilbert space equal to the ground ring.

Let $e_1, \ldots, e_r$ be a semisimple basis for H. We can think of each e_i as the identity vector for the space H_i. Let $e^1, \ldots, e^r$ be the dual basis. Then semisimplicity is equivalent to asking all nondiagonal coefficients to vanish:

$$\Gamma^{j_1, \ldots, j_m}_{i_1, \ldots, i_n}(W^n_m(g)) = 0,$$

unless $i_1 = i_2 = \cdots = i_n = j_1 = \cdots = j_m$.

There are now r universal constants $\lambda_1, \ldots, \lambda_r$ that govern the structure of the TQFT. They can be defined in many equivalent ways. Here are two equivalent descriptions that we will be using later on:

(1) $1/\lambda_i$ is the image of the basis vector e_i via the counit operator.

(2) λ_i is the i-th eigenvalue of the genus adding operator (this is the linear map associated to the torus with a negative and a positive puncture, represented in Figure 3).

Structure Theorem 2.3. *Let $\mathcal{T}$ be a semisimple TQFT, and all notation as above. Denote by $W^n_m(g)$ a genus g surface with m input and n output holes. Then*

$$\mathcal{T}(W^n_m(g)) = \sum_{i=1}^r \lambda_i^{g+n-1} \underbrace{e^i \otimes \cdots \otimes e^i}_{m \text{ times}} \otimes \underbrace{e_i \otimes \cdots \otimes e_i}_{n \text{ times}}.$$

In particular,

$$\mathcal{T}(W^0_0(g)) = \sum_{i=1}^r \lambda_i^{g-1}. \tag{2-2}$$

The TQFT of Hurwitz numbers. Dijkgraaf and Witten [1990] used the TQFT approach to give a beautiful and elegant solution to a classical mathematical problem: counting ramified and unramified covers of a topological surface, as follows.

Let $(X, p_1, \ldots, p_r, q_1, \ldots, q_s)$ be an $(r+s)$-marked smooth topological surface. Let $\underline{\eta} = (\eta_1, \ldots, \eta_r)$ be a vector of partitions of the integer d. We define the

Hurwitz number

$$H_d^{h,X}(\underline{\eta})$$

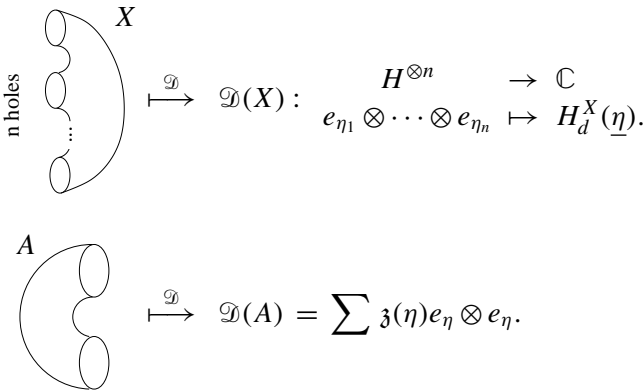as the weighted number of degree d covers $C \xrightarrow{\pi} X$ such that

- C is a surface of genus h;
- π is unramified over $X \setminus \{p_1, \ldots, p_r, q_1, \ldots, q_s\}$;
- π ramifies with profile η_i over p_i;
- π has simple ramification over q_i.

The weight is the number of automorphisms of such covers.

For a Hurwitz number to be nonzero, s, h and $\underline{\eta}$ must satisfy the Riemann–Hurwitz formula. This is why we omit s from the notation. In particular, if we require $s = 0$, then (at most one value of) h is determined by $\underline{\eta}$. We denote by $H_d^X(\underline{\eta})$ the corresponding Hurwitz number.

We define the TQFT $\mathscr{D}$ as follows:

(1) the ground field is $\mathbb{C}$;

(2) the Hilbert space is $H = \bigoplus_{\eta \vdash d} \mathbb{C}e_\eta$, where $\eta \vdash d$ means that η is a partition of d;

(3) morphisms are assigned according to the prescription

$$\mathscr{D}(X): \quad \begin{array}{ccc} H^{\otimes n} & \to & \mathbb{C} \\ e_{\eta_1} \otimes \cdots \otimes e_{\eta_n} & \mapsto & H_d^X(\underline{\eta}). \end{array}$$

$$\mathscr{D}(A) = \sum \mathfrak{z}(\eta) e_\eta \otimes e_\eta.$$

Fact 2.4 (Dijkgraaf, Witten/Freed, Quinn). *The assignment above defines a semisimple TQFT $\mathscr{D}$. Let η be a partition of d, representing a conjugacy class of the symmetric group, and let h be an element in this conjugacy class. Via the identification*

$$e_\eta = \frac{1}{d!} \sum_{g \in S_d} g^{-1} h g,$$

the Hilbert space is isomorphic, as a Frobenius algebra, to the class algebra of the symmetric group in d letters, $\mathscr{Z}(\mathbb{C}[S_d])$.

A semisimple basis is indexed by irreducible representations ρ of S_d. If ρ is such a representation and $\mathscr{X}_\rho$ its character function, then

$$e_\rho = (\dim \rho) \sum_{\eta \vdash d} \mathscr{X}_\rho(\eta) e_\eta. \tag{2-3}$$

This allows one to recover the classical Burnside formula expressing the number of unramified covers of a genus g curve:

$$H_d^{gd-d+1,g}(\phi) = \sum_\rho \left(\frac{d!}{\dim \rho} \right)^{2g-2}. \tag{2-4}$$

Weighted TQFTs. A weighted TQFT contains some extra structure with respect to an ordinary TQFT. Every cobordism comes equipped with a sequence of weights, or levels. When you concatenate two cobordisms, you add the levels componentwise. We are in particular interested in the theory with 2 levels.

Define the category $2\mathsf{Cob}^{k_1,k_2}$ as follows:

(1) Objects and tensor structure are the same as in 2Cob.

(2) Morphisms are given by triples (W, k_1, k_2), where W is an oriented cobordism as in 2Cob and k_1, k_2 are two integers called levels.

(3) Composition of morphisms consists in concatenating cobordisms and adding levels componentwise.

Definition 2.5. A weighted TQFT is a functor of tensor categories

$$\mathscr{WT} : 2\mathsf{Cob}^{k_1,k_2} \longrightarrow \mathsf{FRMod}.$$

Clearly, if we restrict our attention to cobordisms with weight $(0, 0)$, we obtain an ordinary TQFT. More generally, there exists a $\mathbb{Z} \times \mathbb{Z}$ worth of ordinary TQFTs embedded in a weighted TQFT. Denote by $\mathscr{X}$ the Euler characteristic of a cobordism W. For any $(a, b) \in \mathbb{Z} \times \mathbb{Z}$, restricting the weighted TQFT to cobordisms with level

$$(a\mathscr{X}, b\mathscr{X})$$

yields an ordinary TQFT.

Generation results. There are several possible ways to generate a weighted TQFT. A natural one consists in generating the level $(0, 0)$ TQFT, and then giving natural operators that allow one to shift the levels. These elements can be chosen to be, for example, the cylinders with weight $(\pm 1, 0)$ and $(0, \pm 1)$. These operators change the levels of the cobordisms without altering its topology. An equivalent, and equally natural choice, is given by the caps, as illustrated in Figure 2.
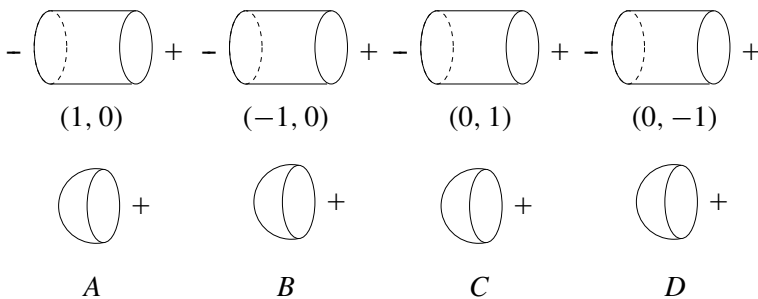
Figure 2. Level-changing objects.

In particular, A is the inverse of B and C is the inverse of D in the level $(0, 0)$ Frobenius algebra. Hence the following generation result.

Fact 2.6 [Bryan and Pandharipande 2004, 4.1]. *A weighted TQFT $\mathcal{WT}$ is uniquely determined by a commutative Frobenius algebra over k for the level* $(0, 0)$ *theory and by two distinguished invertible elements in the Frobenius algebra*:

Semisimple weighted TQFTs. A weighted TQFT of rank r is semisimple if there is a basis for the Hilbert space such that all the nonzero tensors in the theory are diagonal. This is equivalent to requiring all embedded ordinary TQFTs to be semisimple (possibly with rescaled semisimple bases). Let $\lambda_1, \ldots, \lambda_r$ be the eigenvalues of the level $(0, 0)$ genus adding operator. Let $\mu_1, \ldots, \mu_r$ be the eigenvalues of the level $(-1, 0)$ annulus, and $\bar{\mu}_1, \ldots, \bar{\mu}_r$ be the eigenvalues for the level $(0, -1)$ annulus, as illustrated in Figure 3.

Fact 2.7 [Bryan and Pandharipande 2004, 5.2]. *Let $\mathcal{WT}$ be a semisimple TQFT. Denote by $W_m^n(g|k_1, k_2)$ a cobordism of genus g between m input and n output*
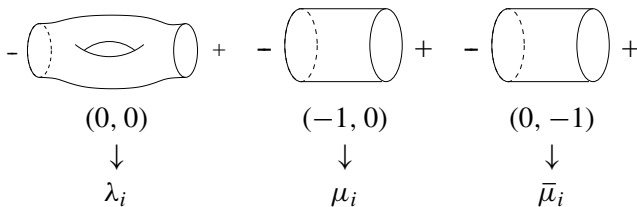

Figure 3. The genus-adding and level-changing operators.

holes, *of level* (k_1, k_2). *Then*

$$\mathcal{T}(W_m^n(g|k_1, k_2)) = \sum_{i=1}^{r} \lambda_i^{g+n-1} \mu_i^{-k_1} \bar{\mu}_i^{-k_2} \underbrace{e^i \otimes \cdots \otimes e^i}_{m \text{ times}} \otimes \underbrace{e_i \otimes \cdots \otimes e_i}_{n \text{ times}}.$$

In particular,

$$\mathcal{T}(W_0^0(g|k_1, k_2)) = \sum_{i=1}^{r} \lambda_i^{g-1} \mu_i^{-k_1} \bar{\mu}_i^{-k_2}.$$

Note. Equivalent definitions can be given for the quantities λ_i, μ_i and $\bar{\mu}_i$. Denote by $e_1, \ldots, e_r$ the vectors of a semisimple basis for the weighted TQFT $\mathcal{W}\mathcal{T}$:

- λ_i^{-1} is the value of the level $(0, 0)$ counit on e_i:

$$\mathcal{W}\mathcal{T}\left(\begin{matrix} - \bigcirc \\ (0,0) \end{matrix} \right)(e_i) = \lambda_i^{-1}.$$

- μ_i is the coefficient of e_i in the level $(-1, 0)$ +disc vector:

$$\mathcal{W}\mathcal{T}\left(\begin{matrix} \bigcirc + \\ (-1,0) \end{matrix} \right) = \sum \mu_i e_i.$$

- $\bar{\mu}_i$ is the coefficient of e_i in the level $(0, -1)$ +disc vector:

$$\mathcal{W}\mathcal{T}\left(\begin{matrix} \bigcirc + \\ (0,-1) \end{matrix} \right) = \sum \bar{\mu}_i e_i.$$

3. Construction of the theory

The admissible covers invariants. Let $(X, p_1, \ldots, p_r)$ be a smooth, irreducible, projective curve of genus g with r distinct marked points, and let $N = L_1 \oplus L_2$ be a rank-2 vector bundle on X. The torus $T = \mathbb{C}^* \times \mathbb{C}^*$ acts naturally on N: the first coordinate scales (with weight one) the fiber of L_1, the second coordinate scales the fiber of L_2.

The T-equivariant cohomology of a point is a polynomial ring in two indeterminates, which we denote by

$$H_T^*(pt) = \mathbb{C}[s_1, s_2].$$

We are interested in the following class of intersection numbers:

$$A_d^h(N) := \int_{\overline{\mathrm{Adm}(h \xrightarrow{d} X, (\eta_1 p_1, \ldots, \eta_r p_r))}} e^{\mathrm{eq}}(-R^\bullet \pi_* f^*(L_1 \oplus L_2)),$$

where

- $\overline{\mathrm{Adm}}(h \xrightarrow{d} X(\eta_1 p_1, \ldots, \eta_r p_r))$ is as defined on page 38;
- e^{eq} is the equivariant Euler class of the virtual bundle in question;
- π is the universal family over the space of admissible covers;
- f is the universal cover map followed by the canonical contraction map to X.

By [Bryan and Pandharipande 2001], this integral only depends on the genus g of the curve X and on the degrees k_1 and k_2 of the line bundles L_1 and L_2. In the TQFT formulation about to be given it will be useful to emphasize this fact, so we choose to denote these invariants by

$$A_d^h(N)_{\underline{\eta}} = A_d^h(g|k_1, k_2)_{\underline{\eta}}.$$

We consider these invariants for all genera h, and organize them in generating function form:

$$A_d(g|k_1, k_2)_{\underline{\eta}} := \sum_{h \in \mathbb{Z}} u^{\star(h)} A_d^h(g|k_1, k_2)_{\underline{\eta}}, \tag{3-1}$$

where the exponent for the generating function is defined by

$$\star(h) = \dim(\overline{\mathrm{Adm}}(h \xrightarrow{d} X, (\eta_1 p_1, \ldots, \eta_r p_r))) = 2h - 2 + d(2 - 2g - r) + \sum_{i=1}^{r} \ell(\eta_i).$$

By expanding the equivariant Euler class in terms of ordinary Chern classes and equivariant parameters, we can express these invariants in terms of nonequivariant integrals. Let $h \in \mathbb{Z} \cup \phi$ be a function of b_1, b_2 determined by the equation

$$b_1 + b_2 = \dim(\overline{\mathrm{Adm}}(h \xrightarrow{d} X, (\eta_1 x_1, \ldots, \eta_r x_r))) = 2h - 2 + d(2 - 2g - r) + \sum_{i=1}^{r} \ell(\eta_i).$$

Define

$$A_d^{b_1,b_2}(g|k_1, k_2)_{\underline{\eta}} := \int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} X, (\eta_1 x_1, \ldots, \eta_r x_r))} c_{b_1}(-R^\bullet \pi_* f^*(L_1)) c_{b_2}(-R^\bullet \pi_* f^*(L_2)).$$

Then the relative invariants are

$$A_d(g|k_1, k_2)_{\underline{\eta}} := \sum_{b_1 + b_2 = 0}^{\infty} u^{b_1 + b_2} i s_1^{r_1 - b_1} s_2^{r_2 - b_2} A_d^{b_1, b_2}(g|k_1, k_2)_{\underline{\eta}}. \tag{3-2}$$

This shows that the partition function for our invariants is a Taylor series in u, whose coefficients are rational functions in s_1 and s_2. The degree of these rational functions is independent of h. It is equal to

$$r_1 + r_2 - b_1 - b_2 = d(2g - 2 - r) - \sum_{i=1}^{r} \ell(\eta_i).$$

The weighted TQFT $\mathcal{U}$. We construct a weighted TQFT $\mathcal{U}$, whose structure coefficients encode the invariants just presented.

The ground ring is defined to be $R = \mathbb{C}[[u]](s_1, s_2)$.

The Hilbert space of the theory is a free R-module of rank equal to the number of partitions of the integer d. A privileged basis is indexed by such partitions η:

$$\mathcal{H} = \bigoplus_{\eta \vdash d} R e_\eta.$$

We denote the dual space by $\mathcal{H}^*$, and the dual basis vectors by e^η.

To construct our TQFT we reason topologically. We think of the marked points on a curve $(X, x_1, \ldots, x_{r+s})$ as of punctures that we can enlarge into loops. We can assign positive or negative orientation to such loops, and arrange the negatively oriented loops $x_1, \ldots, x_r$ to the left, the positively oriented ones to the right (after relabelling $x_{r+i} = y_i$). We now have an oriented cobordism.

To completely determine the structure of the theory we define the scalar maps associated to arbitrary cobordisms into the empty set, and the coproduct, which allows us to move boundary components from left to right:

$$\xmapsto{\mathcal{U}} \mathcal{U}(X) : \quad \begin{array}{ccc} H^{\otimes r} & \to & \mathbb{C}[[u]](s_1, s_2) \\ e_{\eta_1} \otimes \cdots \otimes e_{\eta_r} & \mapsto & A_d(g|k_1, k_2)_{\underline{\eta}}. \end{array}$$

$$\xmapsto{\mathcal{U}} \mathcal{U}(A) = \sum_{\eta \vdash d} \mathfrak{z}(\eta)(s_1 s_2)^{\ell(\eta)} e_\eta \otimes e_\eta.$$

The combinatorial factor $\mathfrak{z}(\eta)$ is defined in (1-2).

Theorem 3.1. *The structure $\mathcal{U}$ defined in the previous paragraph is a two-level, weighted semisimple TQFT.*

In practical terms, it is convenient to adopt the conventional tensor notation of riemannian geometry. If $X = (X, x_1, \ldots, x_r, y_1, \ldots, y_s | k_1, k_2)$ represents a cobordism of genus g and level k_1, k_2 from r circles to s circles, then $\mathcal{U}(X)$ is an element of $(\mathcal{H}^*)^{\otimes r} \otimes \mathcal{H}^{\otimes s}$. We denote by

$$A_d(g|k_1, k_2)_{\eta_1, \ldots, \eta_r}^{\mu_1, \ldots, \mu_s}$$

the coordinate of $\mathcal{U}(X)$ in the direction of the basis element

$$e^{\eta_1} \otimes \cdots \otimes e^{\eta_r} \otimes e_{\mu_1} \otimes \cdots \otimes e_{\mu_s}.$$

With this notation, the coproduct gives the following formula for raising and lowering indices:

$$A_d(g|k_1, k_2)_{\eta_1,\ldots,\eta_r}^{\mu_1,\ldots,\mu_s} = \left(\prod_{i=1}^{s} \mathfrak{z}(\mu_i)(s_1 s_2)^{\ell(\mu_i)} \right) A_d(g|k_1, k_2)_{\eta_1,\ldots,\eta_r,\mu_1,\ldots,\mu_s}. \quad (3\text{-}3)$$

Proof of Theorem 3.1. Proving that $\mathcal{U}$ is indeed a weighted TQFT amounts to verifying three statements:

(*Identity*) The tensor associated to the level $(0, 0)$ trivial cobordism from the circle to the circle is the identity morphism of the Hilbert space $\mathcal{H}$.

(*Gluing two curves*) For any two vectors η, μ of partitions of d, and integers satisfying $g = g' + g''$, $k_1 = k_1' + k_1''$, $k_2 = k_2' + k_2''$,

$$A_d(g|k_1, k_2)_{\eta_1,\ldots,\eta_r}^{\mu_1,\ldots,\mu_s} = \sum_{\nu \vdash d} A_d(g'|k_1', k_2')_{\eta_1,\ldots,\eta_r}^{\nu} A_d(g''|k_1'', k_2'')_{\nu}^{\mu_1,\ldots,\mu_s}. \quad (3\text{-}4)$$

(*Self-gluing*) For any vector of partitions $\underline{\eta}$ and integers g, k_1, k_2,

$$A_d(g+1|k_1, k_2)_{\eta_1,\ldots,\eta_r} = \sum_{\nu \vdash d} A_d(g|k_1, k_2)_{\eta_1,\ldots,\eta_r,\nu}^{\nu}. \quad (3\text{-}5)$$

Identity. This fact is easily proven. One very clever way to do it, pursued in [Bryan and Pandharipande 2005], is to notice that the degree-0 coefficients in our TQFT agree with the classical TQFT of Hurwitz numbers constructed in [Dijkgraaf and Witten 1990] and recalled on page 43. The vanishing of all higher-degree terms can be obtained as a straightforward consequence of the gluing laws, or simply by showing that the dimensions of the moduli spaces in question exceed the maximum degree of a nonequivariant class in the integrand.

Gluing two curves. To minimize bookkeeping, we prove the result when $r = s = 0$ (that is, when the resulting glued curve is not marked). In the general case, the proof follows exactly the same steps, and all the extra indices are simply carried along for the ride.

Consider a one-parameter family of genus g curves W, and the corresponding map to the moduli space,

$$\begin{array}{c} W \\ \downarrow \\ \varphi : \mathbb{A}^1 \longrightarrow \overline{M}_g, \end{array}$$

such that all fibers are smooth curves of genus g, apart from the central fiber

$$W_0 = X_1 \bigcup_{b_1 = b_2} X_2,$$

which is a nodal curve obtained by attaching at a point two smooth curves of genus g' and g'' (with $g' + g'' = g$).

Consider the moduli space $\overline{\mathrm{Adm}}(h \xrightarrow{d} g)$ of admissible covers of a genus g curve by a genus h curve, all ramification simple. By [Abramovich et al. 2003], there is a flat morphism

$$\overline{\mathrm{Adm}}(h \xrightarrow{d} g) \to \overline{M}_g,$$

We can construct the cartesian diagram

$$
\begin{array}{ccccc}
\mathcal{A}_s = \overline{\mathrm{Adm}}(h \xrightarrow{d} W_s) & \hookrightarrow & \mathcal{A} & \longrightarrow & \overline{\mathrm{Adm}}(h \xrightarrow{d} g) \\
\downarrow & & \downarrow & & \downarrow \\
\{s\} & \hookrightarrow & \mathbb{A}^1 & \longrightarrow & \overline{M}_g
\end{array}
\tag{3-6}
$$

The stack $\mathcal{A}$ must be thought of as the stack of relative admissible covers of the family W. For $s \neq 0$, we obtain admissible covers of a smooth genus g curve; for $s = 0$, we recover admissible covers of the nodal curve W_0.

It is possible to construct two line bundles $\mathcal{L}_1$ and $\mathcal{L}_2$ on W with the following properties:

(1) $\mathcal{L}_i$ restricted to any fiber W_s is a line bundle $L_{i,s}$ of degree k_i.

(2) Over the central fiber W_0, $\mathcal{L}_i$ restricts to a line bundle $L'_{i,s}$ of degree k'_i on X_1, and restricts to a line bundle $L''_{i,s}$ of degree k''_i on X_2.

(3) $\mathbb{C}^*$ acts naturally on $\mathcal{L}_i$ by scaling the fibers (with weight one).

Consider the diagram

$$
\begin{array}{ccc}
\mathcal{U}_{\mathcal{A}} & \xrightarrow{f} & \mathcal{W} \longrightarrow W \\
\pi \downarrow & \swarrow & \\
\mathcal{A} & &
\end{array}
$$

where $\mathcal{U}_{\mathcal{A}}$ is the universal family of the moduli space $\mathcal{A}$, $\mathcal{W}$ is the universal target and f the universal admissible cover map.

The pull-push

$$\mathcal{I} = -R^\bullet \pi_* f^*(\mathcal{L}_1 \oplus \mathcal{L}_2)$$

is a virtual bundle of virtual rank $r = 2g - 2 - d(k_1 + k_2)$.

By the flatness of the family $\mathcal{A}$ over $\mathbb{A}^1$, the integral of the top Chern class $c_r(\mathcal{I})$ restricted to a fiber $\mathcal{A}_s$ is independent of the fiber. For $s \neq 0$, we obtain

$$\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} W_s)} c_r(\mathcal{I}|_s) = A_d^h(g|k_1, k_2).$$

We want to evaluate the same expression restricted to $s = 0$, and show it equals the right-hand side of (3-4). We choose to show the equality at the generic genus

h degree of the generating function, to emphasize the geometric nature of the construction. We hence need to establish the following claim, which consists of expanding the genus *h* term in Equation (3-4), and lowering indices as in (3-3).

Claim 3.2.

$$\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} W_0)} c_r(\mathcal{I} \mid_0) = \sum_{\nu \vdash d} \mathfrak{z}(\nu)(s_1 s_2)^{\ell(\nu)} \sum_{h_1, h_2} A_d^{h_1}(g' | k_1', k_2')_\nu A_d^{h_2}(g'' | k_1'', k_2'')_\nu,$$

where the second sum is over pairs of indices such that $h_1 + h_2 + \ell(\nu) - 1 = h$.

Proof. Recall that, by (1-1),

$$[\overline{\mathrm{Adm}}(h \xrightarrow{d} W_0)] = \sum_{\nu \vdash d} \mathfrak{z}(\nu) \sum_{h_1, h_2} [\overline{\mathrm{Adm}}(h_1 \xrightarrow{d} X_1, (\nu b_1))] \times [\overline{\mathrm{Adm}}(h_2 \xrightarrow{d} X_2, (\nu b_2))],$$

where $h_1 + h_2 + \ell(\nu) - 1 = h$ and

$$\dim(\overline{\mathrm{Adm}}(h_1 \xrightarrow{d} X_1, (\nu b_1))) + \dim(\overline{\mathrm{Adm}}(h_2 \xrightarrow{d} X_2, (\nu b_2))) = \dim(\overline{\mathrm{Adm}}(h \xrightarrow{d} W_0)).$$

Consider the pullback of the normalization sequence associated to the restriction of $\mathcal{L}_i$ to W_0:

$$0 \to f^*(L_{i,0}) \to f^*(L'_{i,0}) \oplus f^*(L''_{i,0}) \to f^*(L_{i,0}) \mid_{X_1 \cap X_2} \to 0.$$

This sequence yields a long exact sequence of higher direct image sheaves

$$0 \to R^0 \pi_* f^*(L_{i,0}) \to R^0 \pi_* f^*(L'_{i,0}) \oplus R^0 \pi_* f^*(L''_{i,0}) \to R^0 \pi_* f^*(L_{i,0}) \mid_{X_1 \cap X_2}$$
$$\to R^1 \pi_* f^*(L_{i,0}) \to R^1 \pi_* f^*(L'_{i,0}) \oplus R^1 \pi_* f^*(L''_{i,0}) \to 0.$$

Notice that $(L_{i,0}) \mid_{X_1 \cap X_2}$ is a skyscraper sheaf $\mathbb{C}_b$, on which $\mathbb{C}^*$ acts with weight 1.

We now restrict our attention to a connected component of $\mathcal{A}_0$ on which the covers split as two smooth covers of genus h_1 and h_2, with ramification profile ν over the shadows of the node. Here, $f^*(L_{i,0}) \mid_{X_1 \cap X_2}$ is a trivial vector bundle of rank $\ell(\nu)$, endowed with a natural $\mathbb{C}^*$ action. The preceding exact sequence then leads to

$$c_{r_i}(-R^\bullet \pi_* f^*(L_{i,0})) = s_i^{\ell(\nu)} c_{r_i'}(-R^\bullet \pi_* f^*(L'_{i,0})) c_{r_i''}(-R^\bullet \pi_* f^*(L''_{i,0})),$$

and finally

$$c_r(\mathcal{I} \mid_0) = (s_1 s_2)^{\ell(\nu)} c_{r'}(\mathcal{I} \mid_0') c_{r''}(\mathcal{I} \mid_0'').$$

Putting everything together yields the claim:

$$\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} W_0)} c_r(\mathcal{I}|_0)$$

$$= \sum_\nu \mathfrak{z}(\nu) \sum_{h_1,h_2} \int_{\overline{\mathrm{Adm}}(h_1 \xrightarrow{d} X_1,(\nu b_1)) \times \overline{\mathrm{Adm}}(h_2 \xrightarrow{d} X_2,(\nu b_2))} c_r(\mathcal{I}|_0)$$

$$= \sum_\nu \mathfrak{z}(\nu)(s_1 s_2)^{\ell(\nu)} \sum_{h_1,h_2} \int_{\overline{\mathrm{Adm}}(h_1 \xrightarrow{d} X_1,(\nu b_1))} c_{r'}(\mathcal{I}|_0') \int_{\overline{\mathrm{Adm}}(h_2 \xrightarrow{d} X_2,(\nu b_2))} c_{r''}(\mathcal{I}|_0'')$$

$$= \sum_\nu \mathfrak{z}(\nu)(s_1 s_2)^{\ell(\nu)} \sum_{h_1,h_2} A_d^{h_1}(g'|k_1', k_2')_\nu A_d^{h_2}(g''|k_1'', k_2'')_\nu. \qquad \square$$

Self-gluing. The structure of the proof is very similar to the previous case. Again, we simplify the notation by assuming $r = 0$. Consider a one-parameter family of genus g curves W, and the corresponding map into the moduli space,

$$W$$
$$\downarrow$$
$$\varphi : \mathbb{A}^1 \longrightarrow \overline{M}_g,$$

such that all fibers are smooth curves of genus g, apart from the central fiber

$$W_0 = X/\{b_1 = b_2\},$$

which is a nodal curve obtained by identifying two distinct points on an irreducible smooth curve X of genus $g - 1$.

As before, we construct a cartesian diagram of the form (3-6) and two line bundles $\mathcal{L}_1$ and $\mathcal{L}_2$ on W with properties (1) and (3) from page 50, plus

(2) Over the central fiber W_0, $\mathcal{L}_i$ pulls back to a line bundle $L_{i,s}'$ of degree k_i on the normalization X.

We now consider the equivariant top Chern class of the pull-push

$$\mathcal{I} = -R^\bullet \pi_* f^*(\mathcal{L}_1 \oplus \mathcal{L}_2).$$

For $s \neq 0$,

$$\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} W_s)} c_r(\mathcal{I}|_s) = A_d^h(g|k_1, k_2).$$

Again, we can show that the corresponding integral over the central fiber yields exactly the genus h expansion of the right-hand side of Equation (3-5).

Claim 3.3. $\displaystyle\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} W_{0,})} c_r(\mathcal{I}|_0) = \sum_\nu \mathfrak{z}(\nu)(s_1 s_2)^{\ell(\nu)} A_d^{h'}(g-1|k_1, k_2)_{\nu,\nu}$, *where*
$h' + \ell(\nu) = h.$

Proof. By (1-3), we have $[\overline{\mathrm{Adm}}(h \overset{d}{\to} W_0)] = \sum_{\nu \vdash d} \mathfrak{z}(\nu)[\overline{\mathrm{Adm}}(h' \overset{d}{\to} X, (\nu b_1, \nu b_2))]$, with $h' + \ell(\nu) = h$.

As in the previous claim, after chasing the normalization sequence for the curve W_0 we obtain, over a connected component of $\mathscr{A}_0$ characterized by covers with ramification profile ν over the shadows of the node, the following decomposition:

$$c_r(\mathscr{I} \mid_0) = (s_1 s_2)^{\ell(\nu)} c_{r'}(\mathscr{I} \mid_0'). \tag{3-7}$$

With this in hand, it is easy to obtain the claim and so conclude the proof of Theorem 3.1:

$$\int_{\overline{\mathrm{Adm}}(h \overset{d}{\to} W_0)} c_r(\mathscr{I} \mid_0) = \sum_\nu \mathfrak{z}(\nu) \int_{\overline{\mathrm{Adm}}(h' \overset{d}{\to} X, (\nu b_1, \nu b_2))} c_r(\mathscr{I} \mid_0)$$

$$= \sum_\nu \mathfrak{z}(\nu)(s_1 s_2)^{\ell(\nu)} \int_{\overline{\mathrm{Adm}}(h' \overset{d}{\to} X, (\nu b_1, \nu b_2))} c_{r'}(\mathscr{I} \mid_0')$$

$$= \sum_\nu \mathfrak{z}(\nu)(s_1 s_2)^{\ell(\nu)} A_d^{h'}(g - 1 | k_1', k_2')_{\nu, \nu}. \qquad \square$$

4. Computing the theory

In order to determine the whole weighted TQFT it is sufficient to compute a small number of invariants, as seen in Fact 2.6. Among the many possible choices for a set of generators, we choose as the generators for the level $(0, 0)$ TQFT

(1) the coefficients $A_d(0|0, 0)_\eta$ of the open $(-)$ disc,

(2) the coefficients $A_d(0|0, 0)^{\eta, \mu}$ of the $(+, +)$ annulus, and

(3) the coefficients $A_d(0|0, 0)_{\eta, \mu, \nu}$ associated to the $(-, -, -)$ pair of pants,

and as the generators for level shifting

(4) the coefficients of the Calabi–Yau caps $A_d(0| - 1, 0)_\eta$ and $A_d(0|0, -1)_\eta$.

Theorem 4.1. *The level $(0, 0)$ TQFT coincides with the level $(0, 0)$ theory of* [Bryan and Pandharipande 2004].

The significant difference in the theories lies in the Calabi–Yau caps, which we will compute (starting on page 55) by localization on the moduli spaces of admissible covers.

Proof of Theorem 4.1. It is simple to compute independently the coefficients for the cap. Dimension counts show they are degenerate, in the sense that only the constant term of the series is nonzero. The coefficients for the $(+, +)$ cylinder agree by definition. We will conclude the proof by showing that the coefficients for the pair of pants are the same.

The level (0, 0) pair of pants. The invariants $A_d^\circ(0|0,0)_{\eta,\nu,\mu}$ of the level $(0,0)$ pair of pants are computed by the integrals:

$$\int_{\overline{\mathrm{Adm}}^\circ(h \xrightarrow{d} \mathbb{P}^1, (\eta 0, \mu 1, \nu \infty))} c_{2h-2}^{\mathrm{eq}}(-R^\bullet \pi_* f^*(\mathbb{O}_{\mathbb{P}^1} \oplus \mathbb{O}_{\mathbb{P}^1})).$$

The dimension of the moduli space in question is

$$2h - d - 2 + \ell(\eta) + \ell(\mu) + \ell(\nu).$$

Hence, if $\ell(\eta) + \ell(\mu) + \ell(\nu) > d + 2$, the relative connected integrals vanish. The disconnected integrals are then obtained inductively from invariants of lower degree d.

All other invariants have contributions from connected components, and hence need to be computed directly.

In [Bryan and Pandharipande 2004, Appendix] it is shown that all invariants can be recursively determined from $A_d(0|0,0)_{(d),(d),(2)}$, the invariant corresponding to full ramification over two points, and a simple transposition over the third point. Their proof uses only TQFT formalism; hence it suffices to prove the following statement.

Lemma 4.2. *For $d \geq 2$,*

$$A_d(0|0,0)_{(d),(d),(2)} = \frac{s_1 + s_2}{2 s_1 s_2} \left(d \cot \frac{du}{2} - \cot \frac{u}{2} \right).$$

(This result differs from the analogous one in [Bryan and Pandharipande 2004] by a factor of $-i$, reflecting a normalization in their generating function conventions that we have not adopted.)

Proof. The full ramification conditions force our covers to be connected. Hence the connected and disconnected invariants coincide.

According to (3-2), we have

$$A_d(0|0,0)_{(d),(d),(2)}$$

$$= \sum_{b_1+b_2=0}^{\infty} u^{b_1+b_2} s_1^{h-1-b_1} s_2^{h-1-b_2} \int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)0,(d)1,(2)\infty))} c_{b_1}(\mathbb{E}^*) c_{b_2}(\mathbb{E}^*),$$

with $b_1 + b_2$ equal to the dimension of the moduli space, which is

$$\dim(\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)0,(d)1,(2)\infty))) = 2h - 1.$$

For a given value of h, the only nonvanishing terms in the expression above are those where $(b_1, b_2) = (h, h-1)$ or $(b_1, b_2) = (h-1, h)$. Adding the two, we obtain

$$A_d^h(0|0,0)_{(d),(d),(2)} = \frac{s_1 + s_2}{s_1 s_2} \int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)0,(d)1,(2)\infty))} -\lambda_h \lambda_{h-1}$$

and consequently, the generating function

$$A_d(0|0, 0)_{(d),(d),(2)} = \frac{s_1 + s_2}{s_1 s_2} \sum_{h=0}^{\infty} u^{2h-1} \int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)0, (d)1, (2)\infty))} -\lambda_h \lambda_{h-1},$$

where λ_k denotes the k-th Chern class of the (pullback of the) Hodge bundle $\mathbb{E}$.

Recall that we defined the λ classes on moduli spaces of admissible covers simply by pulling them back from the appropriate moduli spaces of stable curves. In particular we have the diagram

$$\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)0, (d)1, (2)\infty)) \xrightarrow{\rho} \overline{M}_{h,2}$$
$$\searrow \qquad \downarrow \pi$$
$$\overline{M}_h$$

The map ρ is defined by marking on the admissible covers the unique preimages of the branch points 0 and 1. The Hodge bundle on $\overline{M}_h$ pulls back to the Hodge bundle on $\overline{M}_{h,2}$, hence we can think of the λ classes on the moduli space of admissible covers as pulled back from $\overline{M}_{h,2}$.

Denote by $H_d \subset M_{h,2}$ the locus of curves admitting a degree d map to $\mathbb{P}^1$ which is totally ramified at the marked points. Let

$$\overline{H}_d \subset \overline{M}_{h,2}$$

be the closure of H_d, consisting of possibly nodal curves admitting a degree d map to a tree of rational curves, fully ramified over the two marked points. The image of the map

$$\rho : \overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)0, (d)1, (2)\infty)) \longrightarrow \overline{M}_{h,2}$$

is precisely $\overline{H}_d$, and ρ is a degree $2h$ map onto its image.

From this we conclude that

$$\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)0,(d)1,(2)\infty))} -\lambda_h \lambda_{h-1} = 2h \int_{[\overline{H}_d]} -\lambda_h \lambda_{h-1}.$$

This is the integral computed in [Bryan and Pandharipande 2004, pages 28–29]; hence the result follows. This proves Lemma 4.2 and therefore Theorem 4.1.   □

The Calabi–Yau cap. We can obtain $A_d(0|-1, 0)_\eta$ from $A_d(0|0, -1)_\eta$ by simply interchanging the roles of s_1 and s_2. Further:

Theorem 4.3. *Let d be a positive integer, and $\eta = (\eta_1, \ldots, \eta_{\ell(\eta)})$ a partition of d. The degree-d Calabi–Yau invariants are*

$$A_d(0|0, -1)_\eta = (-1)^{d-\ell(\eta)} \frac{\left(2\sin\frac{u}{2}\right)^d}{(s_1)^{\ell(\eta)} \mathfrak{z}(\eta) \prod 2\sin\frac{\eta_i u}{2}}.$$

In [Cavalieri 2004], this formula is computed via localization on moduli spaces of (connected) admissible covers in degree 1, 2, 3. The result is obtained by finding relations between the Calabi–Yau cap invariants and generating functions for simple Hurwitz numbers. Two types of obstructions arise in degrees beyond 3. First, fixed loci inside moduli spaces of connected admissible covers are in principle easily described as finite products and quotients of moduli spaces of connected admissible covers, but the combinatorial complexity grows fast. Second, generating functions for simple Hurwitz numbers are not readily available beyond degree 3.

To circumvent the first problem we interpret the fixed loci in the localization as simpler products of disconnected admissible cover spaces. Then all possible Calabi–Yau invariants, not only the fully ramified ones, appear in the recursions. There is one subtlety to be aware of: Calabi–Yau cap invariants are defined as intersection numbers on moduli spaces of admissible covers of a parametrized $\mathbb{P}^1$, whereas the fixed loci are in terms of admissible covers of unparametrized projective lines. Another localization computation, with an appropriate choice of linearizations for the bundles, gives an expression for the invariants in terms of the unparametrized $\mathbb{P}^1$ admissible covers.

To deal with the lack of explicit generating functions for general simple Hurwitz numbers, we notice that the recursive relation that we need to prove is in fact determined by a virtual localization computation on moduli spaces of stable maps. This is yet more evidence of how intimately related this theory and Gromov–Witten theory are.

Proof of Theorem 4.3. We prove the following formula for the connected Calabi–Yau cap invariants:

$$A_d^\circ(0|0, -1)_\eta = \begin{cases} \dfrac{(-1)^{d-1}}{s_1} \dfrac{1}{d} \dfrac{\left(2\sin\frac{u}{2}\right)^d}{2\sin\frac{du}{2}} & \text{for } \eta = (d), \\[2em] 0 & \text{otherwise.} \end{cases}$$

Theorem 4.3 follows from this via exponentiation.

The vanishing of the connected invariants for all partitions but (d) is a dimension count. By (3-1) and (3-2), the genus-h contribution to the connected Calabi–Yau invariants is

$$A^{\circ h}_d(0|0, -1)_\eta = \int_{\overline{\mathrm{Adm}}^\circ (h \xrightarrow{d} \mathbb{P}^1, (\eta\infty))} c^{\mathrm{eq}}_{2h+d-1}(-R^\bullet\pi_* f^*(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-1)))$$

$$= \sum_{b_1,b_2} s_1^{r_1-b_1} s_2^{r_2-b_2} \int_{\overline{\mathrm{Adm}}^\circ (h \xrightarrow{d} \mathbb{P}^1, (\eta\infty))} c_{b_1}(\mathbb{E}^*) c_{b_2}(R^1\pi_* f^*(\mathcal{O}_{\mathbb{P}^1}(-1))),$$

where

- $b_1 + b_2 = \dim(\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, (\eta\infty))) = 2h + d + \ell(\eta) - 2$;
- $r_1 = h - 1$ is the virtual rank of the virtual bundle $-R^\bullet\pi_* f^*(\mathcal{O}_{\mathbb{P}^1})$;
- $r_2 = h + d - 1$ is the virtual rank of the virtual bundle $-R^\bullet\pi_* f^*(\mathcal{O}_{\mathbb{P}^1}(-1))$.

Since $-R^\bullet\pi_* f^*(\mathcal{O}_{\mathbb{P}^1}(-1)) = R^1\pi_* f^*(\mathcal{O}_{\mathbb{P}^1}(-1))$ is in fact a vector bundle of rank $h + d - 1$, we also have the constraint

$$b_1 + b_2 \leq 2h + d - 1.$$

The only possibly nonvanishing integrals occur when $\ell(\eta) = 1$, i.e. when $\eta = (d)$. The indices b_1 and b_2 are forced to be, respectively, h and $h + d - 1$.

Note. The full ramification condition forces all covers to be connected; the fully ramified connected and disconnected invariants coincide, thus allowing us to drop the superscript $\circ$.

Finally, our task is to prove:

$$\frac{1}{s_1} \sum_{h=0}^\infty u^{2h+d-1} \int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} \mathbb{P}^1, ((d)\infty))} c_h(\mathbb{E}^*) c_{h+d-1}(R^1\pi_* f^*(\mathcal{O}_{\mathbb{P}^1}(-1)))$$

$$= \frac{(-1)^{d-1}}{s_1} \frac{1}{d} \frac{\left(2\sin\frac{u}{2}\right)^d}{2\sin\frac{du}{2}}.$$

Calabi–Yau cap invariants: parametrized to unparametrized. We evaluate via localization the Calabi–Yau cap invariant $A^h_d(0|0, -1)_\eta$, for a general partition η.

We linearize the $(\mathbb{C}^*$ action on the) two bundles by assigning to $\mathcal{O}_{\mathbb{P}^1}$ weight 0 over both 0 and ∞, and assigning $\mathcal{O}_{\mathbb{P}^1}(-1)$ weight 0 over 0 and weight 1 over ∞:

weight	over 0	over ∞
$\mathcal{O}_{\mathbb{P}^1}(-1)$	0	1
$\mathcal{O}_{\mathbb{P}^1}$	0	0

There are a priori many fixed loci in the localization computation. However it is possible to rule out a vast majority of them using either dimension counts

or linearization considerations (see [Cavalieri 2004] or [Bryan and Pandharipande 2005] for a discussion of these standard localization tricks).

Ultimately, the only possibly contributing fixed loci are those whose general element consists of $\ell(\eta)$ spheres S_i, mapping to the main $\mathbb{P}^1$ with degree η_i, all fully ramified over 0 and ∞. A genus 0 twig sprouts from the point ∞ on the main $\mathbb{P}^1$, covered by $\ell(\eta)$ curves C_i of genus h_i. The curve C_i is attached to S_i at a fully ramified point. The h_i are such that

$$h_1 + \ldots + h_{\ell(\eta)} = h + \ell(\eta) - 1.$$

Finally, if we denote by $F_{\eta,h}$ the disjoint union of all such fixed loci as the h_i vary, and by N the normal bundle to such fixed loci, we obtain from localization:

$$A_d^h(0|0, -1)_\eta = \int_{F_{\eta,h}} \frac{e^{\mathrm{eq}}(-R^\bullet \pi_* f^*(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-1))) \mid_{F_{h,\eta}}}{e^{\mathrm{eq}}(N)}.$$

Recursion via localization on admissible covers. We now suppose $d > 1$ and consider the auxiliary integral

$$I^h = \int_{\overline{\mathrm{Adm}}^\circ(h \xrightarrow{d} \mathbb{P}^1)} e^{\mathrm{eq}}(-R^\bullet \pi_* f^*(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-1))), \tag{4-1}$$

computed on the space of connected admissible covers. It vanishes for dimension reasons: we are integrating a class whose highest nonequivariant factor has codimension $(2h + d - 1)$ on a space of dimension $2h + 2d - 2$.

On the other hand, if we evaluate the integral via localization we get a relation among Calabi–Yau cap invariants. We let a one-dimensional torus act naturally on the moduli space and denote the equivariant parameter by s. We choose to linearize the two bundles with the following weights:

weight	over 0	over ∞
$\mathcal{O}_{\mathbb{P}^1}(-1)$	-1	0
$\mathcal{O}_{\mathbb{P}^1}$	1	1

The possibly contributing fixed loci E_{η,h_0,h_∞} are represented by connected localization graphs such that any vertex over ∞ has valence 1; see [Cavalieri 2004]. They can be indexed by triples (η, h_0, h_∞), where

- $\eta = (d_1, \ldots, d_{\ell(\eta)})$ is a partition of d representing the configuration of the spheres over the main $\mathbb{P}^1$;

- h_0 is the genus of the curve lying over 0;

- h_∞ is the genus of the curve lying over ∞ (considered as a disconnected curve);

- $h_0 + h_\infty = h - \ell(\eta) + 1$.

We recognize that a general element in the fixed locus E_{η,h_0,h_∞} is obtained by gluing together an element in the fixed locus F_{η,h_∞} with a connected admissible covers of a genus 0 curve, with a special point of ramification η. Keeping in account the stacky contribution from the gluing, our integral I on E_{η,h_0,h_∞} reduces to

$$I^h_{\eta,h_0,h_\infty}$$
$$= \mathfrak{z}(\eta) \int_{\overline{\mathrm{Adm}}^\circ(h_0 \xrightarrow{d} \mathbb{P}^1, \eta) \times F_{\eta,h_\infty}} \frac{e^{\mathrm{eq}}(-R^\bullet \pi_* f^*(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-1))) \mid_{\overline{\mathrm{Adm}}^\circ(h \xrightarrow{d} \mathbb{P}^1, \eta) \times F_{\eta,h_\infty}}}{e^{\mathrm{eq}}(N)}$$
$$= \mathfrak{z}(\eta) s^{2\ell(\eta)} A^{h_\infty}_d(0|0,-1)_\eta \int_{\overline{\mathrm{Adm}}^\circ(h_0 \xrightarrow{d} \mathbb{P}^1, \eta)} \frac{c_{h_0}(\mathbb{E}^* \otimes \mathbb{C}_1) c_{h_0}(\mathbb{E}^* \otimes \mathbb{C}_{-1})}{s(s - \psi_\eta)},$$

where $\mathbb{C}_a$ is a trivial line bundle where the torus acts on the fibers with weight a.

After expanding and using Mumford's relation [1983] saying that $c(\mathbb{E})c(\mathbb{E}^*)$ equals 1, we obtain

$$I^h_{\eta,h_0,h_\infty} = \mathfrak{z}(\eta) s^{\ell(\eta)+2-d} A^{h_\infty}_d(0|0,-1)_\eta \int_{\overline{\mathrm{Adm}}^\circ(h_0 \xrightarrow{d} \mathbb{P}^1, \eta)} (-1)^{h_0} \psi_\eta^{2h_0+d+\ell(\eta)-4}$$
$$= \mathfrak{z}(\eta) s^{\ell(\eta)+2-d} A^{h_\infty}_d(0|0,-1)_\eta \frac{(-1)^{h_0} H^{h_0}_d(\eta)}{(2h_0 + d + \ell(\eta) - 2)!}.$$

The quantity $H^{h_0}_d(\eta)$ is a simple Hurwitz number, as defined on page 43.

The integral I is evaluated by adding up the contributions from all fixed loci E_{η,h_0,h_∞}:

$$0 = I^h = \sum_{\eta \vdash d} \sum_{h_0+h_\infty=h-\ell(\eta)+1} I^h_{\eta,h_0,h_\infty}. \tag{4-2}$$

This holds for all genera h, and can be expressed in a very compact form in the language of generating functions. Define

$$\mathcal{H}_{d,\eta(u)} := \sum \frac{(-1)^h H^h_d(\eta)}{(2h+d+\ell(\eta)-2)!} u^{(2h+d+\ell(\eta)-2)}.$$

Then formulas (4-2), for all genera h, are encoded in the relation

$$0 = \sum_{\eta \vdash d} \mathfrak{z}(\eta) s^{\ell(\eta)+2-d} A_d(0|0,-1)_\eta(u) \mathcal{H}_{d,\eta}(u). \tag{4-3}$$

This relation determines $A_d(0|0,-1)_{(d)}$ in terms of generating functions for simple Hurwitz numbers and of the invariants $A_d(0|0,-1)_\eta$, for $\ell(\eta) \geq 2$, which can be inductively determined via exponentiation if we assume the theory up to degree $d-1$. The theory has been explicitly computed up to degree 3 in [Cavalieri 2004]; hence the induction can start.

To prove Theorem 4.3 it therefore suffices to show that (4-3) holds for the conjectured values of the Calabi–Yau invariants. After substituting and simplifying,

this amounts to proving that

$$0 = \sum_{\eta \vdash d} (-1)^{\ell(\eta)} \frac{\mathcal{H}_{d,\eta}(u)}{\displaystyle\prod_{\eta_i \in \eta} 2 \sin \frac{\eta_i u}{2}}. \qquad (4\text{-}4)$$

Virtual localization on stable maps. Relation (4-4) is the result of explicitly evaluating via virtual localization the auxiliary integrals

$$J^h = \int_{\overline{M}_h(\mathbb{P}^1,d)} e^{\mathrm{eq}}(-R^\bullet \pi_* f^*(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-1))).$$

Again dimension reasons grant us the vanishing of this integral. We proceed to linearize the bundles by assigning weights as follows:

weight	over 0	over ∞
$\mathcal{O}_{\mathbb{P}^1}(-1)$	-1	0
$\mathcal{O}_{\mathbb{P}^1}$	1	1

The analysis of the possibly contributing fixed loci is parallel to the previous section. The contribution by the fixed locus E_{η,h_0,h_∞} is

$$\sum_{h_1+\ldots+h_{\ell(\eta)}=h_\infty+\ell(\eta)-1} J_{\eta,h_0,h_1,\ldots,h_{\ell(\eta)}},$$

with

$$J_{\eta,h_0,h_1,\ldots,h_{\ell(\eta)}} = \frac{1}{\mathfrak{z}(\eta)} \int_{\overline{M}_{h_0,\ell(\eta)}} \frac{c_{h_0}(\mathbb{E}^* \otimes \mathbb{C}_1) c_{h_0}(\mathbb{E}^* \otimes \mathbb{C}_1) c_{h_0}(\mathbb{E}^* \otimes \mathbb{C}_{-1})}{\prod \left(\eta_i^{-1} - \psi_i \right)}$$

$$\times \prod_{i=1}^{\ell(\eta)} \frac{\eta_i^{\eta_i}}{\eta_i!} \int_{\overline{M}_{h_i,1}} \frac{c_{h_i}(\mathbb{E}^* \otimes \mathbb{C}_1) c_{h_i}(\mathbb{E}^* \otimes \mathbb{C}_1) c_{h_i}(\mathbb{E}^*)}{-\eta_i^{-1} - \psi_1}.$$

(See [Hori et al. 2003, Chapter 27] for a clear and detailed explanation of how to compute these terms, or [Bryan and Pandharipande 2004, proof of Theorem 5.1] for a very similar computation.)

After simplifying via Mumford's relation and rearranging things, the preceding formula becomes

$$\frac{(-1)^{h_0}}{\mathrm{Aut}(\eta)} \prod_{i=1}^{\ell(\eta)} \frac{\eta_i^{\eta_i}}{\eta_i!} \int_{\overline{M}_{h_0,\ell(\eta)}} \frac{1 - \lambda_1 + \ldots \pm \lambda_{h_0}}{\prod(1 - \eta_i \psi_i)} \prod_{i=1}^{\ell(\eta)} -\eta_i^{2h_i-1} \int_{\overline{M}_{h_i,1}} \lambda_{h_i} \psi_1^{2h_i-2}. \qquad (4\text{-}5)$$

We recognize in formula (4-5) two famous results in the field. The first is the *ELSV formula*, establishes the connection between Hurwitz numbers and Hodge

integrals [Ekedahl et al. 2001; Graber and Vakil 2003a]:

$$H_d^h(\eta) = \frac{(2h+d+\ell(\eta)-2)!}{\mathrm{Aut}(\eta)} \prod_{i=1}^{\ell(\eta)} \frac{\eta_i^{\eta_i}}{\eta_i!} \int_{\overline{M}_{h,\ell(\eta)}} \frac{1-\lambda_1+\ldots \pm \lambda_h}{\prod(1-\eta_i\psi_i)}.$$

The second is Faber and Pandharipande's formula [2000], expressing in generating function form the following class of integrals:

$$\mathcal{L}(u) := \sum u^{2h-1} \int_{\overline{M}_{h,1}} \lambda_h \psi_1^{2h-2} = \frac{1}{2\sin\dfrac{u}{2}}.$$

Now it is a matter of careful bookkeeping to translate all this information in the language of generating functions. Doing so concludes the proof of Theorem 4.3 by establishing that

$$0 = \sum_{h\in\mathbb{Z}} J^h u^{2h+2d-2} = \sum_{\eta\vdash d}(-1)^{\ell(\eta)}\mathcal{H}_{d,\eta}(u) \prod_{\eta_i\in\eta} \mathcal{L}(\eta_i u)$$

$$= \sum_{\eta\vdash d}(-1)^{\ell(\eta)} \frac{\mathcal{H}_{d,\eta}(u)}{\prod_{\eta_i\in\eta} 2\sin\dfrac{\eta_i u}{2}}. \qquad \square$$

5. A specialization of the theory

We now discuss a specialization of the theory, obtained by embedding a one-dimensional torus inside the two-dimensional torus T, and considering the theory as depending from one equivariant parameter instead of two.

We specialize to the antidiagonal action, and notice that the coefficients for the product simplify dramatically. It is possible to obtain nice closed formulas for our theory, and to view our TQFT as a one-parameter deformation of the classical TQFT of Hurwitz numbers studied in [Dijkgraaf and Witten 1990; Freed and Quinn 1993]. Our formulas show connections to the representation theory of the symmetric group S_d.

The antidiagonal action. Embed $\mathbb{C}^*$ in the two-dimensional torus T via the map

$$\alpha \mapsto \left(\alpha, \frac{1}{\alpha}\right).$$

$\mathbb{C}^*$ acts on N by composing this embedding with the natural action of T constructed in page 46. If we set

$$H_{\mathbb{C}^*}^*(pt) = \mathbb{C}[s],$$

the one-parameter theory obtained with this action corresponds to setting

$$s = s_1 = -s_2.$$

The Q-dimension of an irreducible representation. Let ρ be an irreducible representation of the symmetric group on d letters S_d. Classically, a partition of d, and hence a Young diagram, can be canonically associated to ρ; see [Fulton and Harris 1991, Chapter 4], for example. Recall that the *hook length* $h(\square)$ of a cell $\square$ in a Young diagram is the number of cells in the L-shaped strip, or "hook", having the given cell as its northwest corner (see figure on the next page). We now define the Q-dimension of the representation ρ by setting

$$\frac{\dim_Q \rho}{d!} := \prod_{\square \in \rho} \frac{1-Q}{1-Q^{h(\square)}} = \prod_{\square \in \rho} \frac{1}{1+Q+\cdots+Q^{h(\square)-1}} \quad (5\text{-}1)$$

The classical hook-length formula says that

$$\dim \rho = d! \Big/ \prod_{\square \in \rho} h(\square).$$

$h(\blacksquare) = 6$
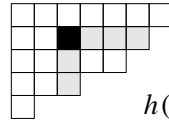
Thus formula (5-1) specializes to the ordinary dimension of ρ when $Q = 1$.

The level $(0, 0)$ TQFT. The main result is that the level $(0, 0)$ TQFT completely collapses to the Dijkgraaf TQFT $\mathscr{D}$. In particular, we have explicit formulas for the semisimple basis of the Frobenius algebra. The basis vectors are indexed by irreducible representations of the symmetric group S_d.

Lemma 5.1. *For the antidiagonal action, the level $(0, 0)$ series have no nonzero terms of positive degree in u.*

Proof. (Essentially by Bryan and Pandharipande.) Endow $\mathbb{C}$ with the $\mathbb{C}^*$ action

$$\alpha \cdot z = \alpha^n z.$$

This corresponds to considering $\mathbb{C}$ as an equivariant line bundle over a point, whose first equivariant Chern class is ns. We denote such an equivariant line bundle by $\mathbb{C}_{ns}$.

The level $(0, 0)$ partition functions are, up to some pure weight factor, constructed from integrals of the form

$$\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} X, (\eta^1 x_1, \ldots, \eta^r x_r))} e^{\mathrm{eq}}(\mathbb{E}^* \otimes \mathbb{C}_s) e^{\mathrm{eq}}(\mathbb{E}^* \otimes \mathbb{C}_{-s})$$

$$\int_{\overline{\mathrm{Adm}}(h \xrightarrow{d} X, (\eta^1 x_1, \ldots, \eta^r x_r))} (-1)^h e^{\mathrm{eq}}((\mathbb{E}^* \oplus \mathbb{E}) \otimes \mathbb{C}_s).$$

Equivariant Chern classes of a bundle also are products of ordinary Chern classes times the appropriate factor of s. But by Mumford's relation $c(\mathbb{E})c(\mathbb{E}^*) = 1$, all Chern classes (but the 0-th) of the bundle $\mathbb{E}^* \oplus \mathbb{E}$ vanish. Hence the only possibly nonvanishing integrals occur when the dimension of the moduli space is 0, which then constitutes the degree 0 term in our generating functions. $\square$

Thus we have already found a semisimple basis for the corresponding Frobenius algebra in (2-3). All we need to do is adjust for the equivariant parameter. Let ρ be an irreducible representation of the symmetric group S_d, with character function χ_ρ; a *semisimple basis* for the level $(0,0)$ TQFT is given by the vectors

$$e_\rho = \frac{\dim \rho}{d!} \sum_{\eta \vdash d} (s)^{\ell(\eta)-d} \chi_\rho(\eta) e_\eta.$$

Notation. If $\eta = (\eta_1, \ldots, \eta_r)$ is a partition, we define

$$n(\eta) := 0\eta_1 + 1\eta_2 + \cdots + (r-1)\eta_r.$$

Theorem 5.2. *The partition functions corresponding to surfaces without boundary in the weighted TQFT are given in closed form by*

$$A_d(g|k_1, k_2) = (-1)^a s^b \sum_\rho \left(\frac{d!}{\dim \rho}\right)^{2g-2} \left(\frac{\dim \rho}{\dim_Q \rho}\right)^{k_1+k_2} Q^{n(\rho)k_1 + n(\rho')k_2},$$

where $a := d(g-1-k_2)$, $b := d(2g-2-k_1-k_2)$, $Q := e^{iu}$, *and* ρ *is an irreducible representation of the symmetric group* S_d, *with dual representation* ρ'.

Note. By setting $Q = 1$, which corresponds to $u = 0$, we recover the classical formula (2-4) counting unramified covers of a genus g topological surface. Thus any TQFT naturally embedded in our weighted TQFT constitutes a one-parameter deformation of the Dijkgraaf TQFT.

Proof of Theorem 5.2. By Fact 2.7, to completely describe the structure of a semisimple weighted TQFT it suffices to evaluate the following quantities:

- the e_ρ-eigenvalue λ_ρ of the genus-adding operator, or, equivalently, the inverse of the counit evaluated on e_ρ;
- the e_ρ-eigenvalue μ_ρ of the left level-subtracting operator, or, equivalently, the coefficient of e_ρ in the $(0, -1)$ Calabi–Yau cap;
- the e_ρ-eigenvalue $\bar{\mu}_\rho$ of the right level-subtracting operator, or, equivalently, the coefficient of e_ρ in the $(-1, 0)$ Calabi–Yau cap.

The computation of λ_ρ coincides exactly with the one in [Bryan and Pandharipande 2004]. We therefore omit it.

To compute μ_ρ and $\bar{\mu}_\rho$ we first observe that the tensors associated to the Calabi–Yau caps in our theory are scalar multiples of the tensors in Bryan and Pandharipande's theory:

$$\mathcal{U}(CYcap) = 2^d \left(\sin \frac{u}{2}\right)^d \mathcal{BP}(CYcap) = \frac{(1-Q)^d}{Q^{(d/2)}(-i)^d} \mathcal{BP}(CYcap).$$

This, together with the formulas in [Bryan and Pandharipande 2004, page 36], implies that

$$\mu_\rho = s^d \frac{d!}{\dim \rho}(1-Q)^d s_\rho(Q), \qquad \bar{\mu}_\rho = (-s)^d \frac{d!}{\dim \rho}(1-Q)^d s_{\rho'}(Q),$$

where s_ρ denotes the Schur function of the representation ρ, and is defined to be (see [Macdonald 1995])

$$s_\rho(Q) := Q^{n(\rho)} \prod_{\square \in \rho} \frac{1}{1 - Q^{h(\square)}}.$$

Plugging this in, we obtain

$$\mu_\rho = s^d \left(\frac{d!}{\dim \rho}\right)(1-Q)^d Q^{n(\rho)} \prod_{\square \in \rho} \frac{1}{1 - Q^{h(\square)}}$$

$$= s^d \left(\frac{d!}{\dim \rho}\right) Q^{n(\rho)} \prod_{\square \in \rho} \frac{1 - Q}{1 - Q^{h(\square)}} = s^d \left(\frac{\dim_Q \rho}{\dim \rho}\right) Q^{n(\rho)},$$

$$\bar{\mu}_\rho = (-s)^d \left(\frac{d!}{\dim \rho}\right)(1-Q)^d Q^{n(\rho')} \prod_{\square \in \rho'} \frac{1}{1 - Q^{h(\square)}}$$

$$= s^d \left(\frac{d!}{\dim \rho}\right) Q^{n(\rho')} \prod_{\square \in \rho'} \frac{1 - Q}{1 - Q^{h(\square)}} = s^d \left(\frac{\dim_Q \rho}{\dim \rho}\right) Q^{n(\rho')}.$$

The theorem is then obtained by using these coefficients in the formula given by Fact 2.7. $\square$

References

[Abramovich et al. 2003] D. Abramovich, A. Corti, and A. Vistoli, "Twisted bundles and admissible covers", *Comm. Algebra* **31**:8 (2003), 3547–3618. Special issue in honor of Steven L. Kleiman. MR 2005b:14049 Zbl 1077.14034

[Bryan and Graber ≥ 2008] J. Bryan and T. Graber, "The crepant resolution conjecture". In preparation.

[Bryan and Pandharipande 2001] J. Bryan and R. Pandharipande, "BPS states of curves in Calabi-Yau 3-folds", *Geom. Topol.* **5** (2001), 287–318. MR 2002h:14092 Zbl 1063.14068

[Bryan and Pandharipande 2004] J. Bryan and R. Pandharipande, "The local Gromov–Witten theory of curves", preprint, 2004. math.AG/0411037

[Bryan and Pandharipande 2005] J. Bryan and R. Pandharipande, "Curves in Calabi–Yau threefolds and topological quantum field theory", *Duke Math. J.* **126**:2 (2005), 369–396. MR 2005k:14117 Zbl 1084.14053

[Bryan et al. 2005] J. Bryan, T. Graber, and R. Pandharipande, "The orbifold quantum cohomology of $\mathbb{C}^2/\mathbb{Z}_3$ and Hurwitz–Hodge integrals", preprint, 2005. math.AG/0510335

[Cavalieri 2004] R. Cavalieri, "Hodge-type integrals on moduli spaces of admissible covers", preprint, 2004. math.AG/0411500

[Cavalieri 2005] R. Cavalieri, *A topological quantum field theory of intersection numbers for moduli spaces of admissible covers*, Ph.D. thesis, Univ. of Utah, 2005, http://www.math.lsa.umich.edu/~crenzo/thesis.pdf.

[Dijkgraaf and Witten 1990] R. Dijkgraaf and E. Witten, "Topological gauge theories and group cohomology", *Comm. Math. Phys.* **129**:2 (1990), 393–429. MR 91g:81133 Zbl 0703.58011

[Ekedahl et al. 2001] T. Ekedahl, S. Lando, M. Shapiro, and A. Vainshtein, "Hurwitz numbers and intersections on moduli spaces of curves", *Invent. Math.* **146**:2 (2001), 297–327. MR 2002j:14034 Zbl 1073.14041

[Faber and Pandharipande 2000] C. Faber and R. Pandharipande, "Hodge integrals and Gromov–Witten theory", *Invent. Math.* **139**:1 (2000), 173–199. MR 2000m:14057 Zbl 0960.14031

[Freed and Quinn 1993] D. S. Freed and F. Quinn, "Chern-Simons theory with finite gauge group", *Comm. Math. Phys.* **156**:3 (1993), 435–472. MR 94k:58023 Zbl 0788.58013

[Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory: a first course*, vol. 129, Graduate Texts in Mathematics, Springer-Verlag, New York, 1991. A first course, Readings in Mathematics. MR 93a:20069 Zbl 0744.22001

[Graber and Vakil 2003a] T. Graber and R. Vakil, "Hodge integrals and Hurwitz numbers via virtual localization", *Compositio Math.* **135**:1 (2003), 25–36. MR 2004c:14108 Zbl 1063.14032

[Graber and Vakil 2003b] T. Graber and R. Vakil, "Relative virtual localization and vanishing of tautological classes on moduli spaces of curves", preprint, 2003. math.AG/0309227

[Harris and Mumford 1982] J. Harris and D. Mumford, "On the Kodaira dimension of the moduli space of curves", *Invent. Math.* **67**:1 (1982), 23–88. With an appendix by William Fulton. MR 83i:14018 Zbl 0506.14016

[Hori et al. 2003] K. Hori, S. Katz, A. Klemm, R. Pandharipande, R. Thomas, C. Vafa, R. Vakil, and E. Zaslow, *Mirror symmetry*, Clay Mathematics Monographs **1**, American Math. Soc., Providence, 2003. MR 2004g:14042 Zbl 1044.14018

[Ionel 2002] E.-N. Ionel, "Topological recursive relations in $H^{2g}(\mathcal{M}_{g,n})$", *Inventiones Math.* **148**:3 (2002), 627–658. MR 2003d:14065 Zbl 1056.14076

[Ionel 2005] E.-N. Ionel, "Relations in the tautological ring of $\mathcal{M}_g$", *Duke Math. J.* **129**:1 (2005), 157–186. MR 2006c:14040 Zbl 1086.14023

[Kock 2004] J. Kock, *Frobenius algebras and 2D topological quantum field theories*, London Math. Soc. Student Texts **59**, Cambridge Univ. Press, 2004. MR 2005a:57028 Zbl 1046.57001

[Li 2002] J. Li, "A degeneration formula of GW-invariants", *J. Differential Geom.* **60**:2 (2002), 199–293. MR 2004k:14096 Zbl 1063.14069

[Macdonald 1995] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Oxford Univ. Press, New York, 1995. MR 96h:05207 Zbl 0824.05059

[Mumford 1983] D. Mumford, "Towards an enumerative geometry of the moduli space of curves",
  pp. 271–328 in *Arithmetic and geometry*, vol. II, edited by M. Artin and J. Tate, Progr. Math. **36**,
  Birkhäuser, Boston, 1983. MR 85j:14046 Zbl 0554.14008

crenzo@umich.edu   *University of Michigan, Department of Mathematics,*
                                    2074 East Hall, 530 Church Street,
                                    Ann Arbor, MI 48109-1043, United States

Multiplicities of Galois representations of weight one

Gabor Wiese
Appendix by Niko Naumann

We consider mod p modular Galois representations which are unramified at p such that the Frobenius element at p acts through a scalar matrix. The principal result states that the multiplicity of any such representation is bigger than 1.

1. Introduction

A continuous odd irreducible Galois representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is said to be of weight one if it is unramified at p. According to Serre's conjecture (with the minimal weight as defined in [Edixhoven 1992]), all such representations should arise from Katz modular forms of weight 1 over $\overline{\mathbb{F}}_p$ for the group $\Gamma_1(N)$ with N the (prime to p) conductor of ρ. Assuming the modularity of ρ, this is known if $p > 2$ or if $p = 2$ and the restriction of ρ to a decomposition group at 2 is not an extension of twice the same character. A weight 1 Katz modular form over $\overline{\mathbb{F}}_p$ can be embedded into weight p and the same level in two different ways: by multiplication by the Hasse invariant of weight $p-1$ and by applying the Frobenius (see [Edixhoven 2006, Section 4]). Hence, the corresponding eigenform(s) in weight p should be considered as old forms; they lie in the ordinary part.

A modular Galois representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ of conductor N can be realised with a certain multiplicity (see Proposition 4.1) on the p-torsion of $J_1(Np)$ or $J_1(N)$. In this article we prove that this multiplicity is bigger than 1 if ρ is of weight one and Frob_p acts through a scalar matrix. If $p = 2$, we also assume that the corresponding weight 1 form exists. Together with [Buzzard 2001, Theorem 6.1], this completely settles the question of multiplicity one for modular Galois representations. Its study had been started by Mazur and continued among others by Ribet, Gross, Edixhoven and Buzzard. The first example of a modular Galois representation not satisfying multiplicity one was found in [Kilford 2002]. See [Kilford and Wiese 2006] for a more detailed exposition.

A systematic computational study of the multiplicity of Galois representations of weight one has been carried out in [Kilford and Wiese 2006]. The data gathered suggest that the multiplicity always seems to be 2 if it is not 1. Moreover, the local factors of the Hecke algebras are becoming astonishingly large.

Overview. We give a short overview over the article with an outline of the proof. In Section 2 an isomorphism between a certain part of the p-torsion of a Jacobian of a modular curve with a local factor of a mod p Hecke algebra is established (Proposition 2.2). As an application one obtains a mod p version of the Eichler–Shimura isomorphism (Corollary 2.3). Together with a variant of a well-known theorem by Boston, Lenstra and Ribet (Proposition 4.1) one also gets an isomorphism between a certain kernel in the local mod p Hecke algebra and a part of the corresponding Galois representation. This gives for instance a precise link between multiplicities and properties of the Hecke algebra (Corollary 4.2). In Section 3 it is proved (Theorem 3.1) that under the identification of Section 2, the geometric Frobenius at p on the part of the Galois representation corresponds to the Hecke operator T_p in the Hecke algebra. This relation is exploited in Section 4 to obtain the principal result (Corollary 4.5), a reformulation and a possible application to weight lowering.

Notation. For integers $N \geq 1$ and $k \geq 1$, we let $S_k(\Gamma_1(N))$ be the $\mathbb{C}$-vector space of holomorphic cusp forms and $S_k(\Gamma_1(N), \mathbb{F}_p)$ the $\mathbb{F}_p$-vector space of Katz cusp forms on $\Gamma_1(N)$ of weight k. Whenever $S \subseteq R$ are rings, m is an integer and M is an R-module on which the Hecke and diamond operators act, we let $\mathbb{T}_S^{(m)}(M)$ be the S-subalgebra inside the R-endomorphism ring of M generated by the Hecke operators T_n with $(n, m) = 1$ and the diamond operators. If $\phi : S \to S'$ is a ring homomorphism, we let $\mathbb{T}_\phi^{(m)}(M) := \mathbb{T}_S^{(m)}(M) \otimes_S S'$ or with ϕ understood $\mathbb{T}_{S \to S'}^{(m)}(M)$. If $m = 1$, we drop the superscript.

Every maximal ideal $\overline{\mathfrak{m}} \subseteq \mathbb{T}_{\mathbb{Z} \to \mathbb{F}_p}(S_k(\Gamma_1(N)))$ corresponds to a Galois conjugacy class of cusp forms over $\overline{\mathbb{F}}_p$ of weight k on $\Gamma_1(N)$. One can attach to $\overline{\mathfrak{m}}$ by work of Shimura and Deligne a continuous odd semisimple Galois representation $\rho_{\overline{\mathfrak{m}}} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ which is unramified outside Np and satisfies $\mathrm{Tr}(\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_l)) \equiv T_l \mod \overline{\mathfrak{m}}$ and $\mathrm{Det}(\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_l)) \equiv \langle l \rangle l^{k-1} \mod \overline{\mathfrak{m}}$ for all primes $l \nmid Np$ via an embedding $\mathbb{T}_{\mathbb{Z} \to \mathbb{F}_p}(S_k(\Gamma_1(N)))/\overline{\mathfrak{m}} \hookrightarrow \overline{\mathbb{F}}_p$. All Frobenius elements Frob_l are arithmetic ones.

For all the article we fix an isomorphism $\mathbb{C} \cong \overline{\mathbb{Q}}_p$ and a ring surjection $\overline{\mathbb{Z}}_p \to \overline{\mathbb{F}}_p$. If K is a field, we denote by $K(\epsilon) = K[\epsilon]/(\epsilon^2)$ the dual numbers. For a finite flat group scheme G, the Cartier dual is denoted by ${}^t G$. The maximal unramified extension of $\mathbb{Q}_p$ (inside $\overline{\mathbb{Q}}_p$) is denoted by $\mathbb{Q}_p^{\mathrm{nr}}$ and its integer ring by $\mathbb{Z}_p^{\mathrm{nr}}$.

For the conventions on modular curves we follow [Gross 1990]; in particular, we work with μ_N-level structures.

Situations. We shall often assume one of the following two situations. In the applications, the second part will be taken for $p = 2$.

Situation I. Let p be an odd prime and N a positive integer not divisible by p. Define the Hecke algebras

$$\mathbb{T}_{\mathbb{Z}_p} := \mathbb{T}^{(1)}_{\mathbb{Z} \to \mathbb{Z}_p}(S_2(\Gamma_1(Np))), \qquad \mathbb{T}'_{\mathbb{Z}_p} := \mathbb{T}^{(p)}_{\mathbb{Z} \to \mathbb{Z}_p}(S_2(\Gamma_1(Np))).$$

Let $\mathfrak{m}$ be an ordinary (i.e. $T_p \notin \mathfrak{m}$) maximal ideal of $\mathbb{T}_{\mathbb{Z}_p}$ with residue field $\mathbb{F} = \mathbb{T}_{\mathbb{Z}_p}/\mathfrak{m}$ such that the p-diamond operators give a nontrivial character

$$(\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{F}^\times, \quad a \mapsto \langle a \rangle_p.$$

Let $\mathfrak{m}' = \mathfrak{m} \cap \mathbb{T}'_{\mathbb{Z}_p}$ and, more generally, $\mathfrak{m}^{(m)} = \mathfrak{m} \cap \mathbb{T}^{(m)}_{\mathbb{Z} \to \mathbb{Z}_p}(S_2(\Gamma_1(Np)))$ for $m \in \mathbb{N}$. Let $\mathbb{T}_{\mathbb{F}_p} := \mathbb{T}_{\mathbb{Z}_p} \otimes_{\mathbb{Z}_p} \mathbb{F}_p$ and $\mathbb{T}'_{\mathbb{F}_p} := \mathbb{T}'_{\mathbb{Z}_p} \otimes_{\mathbb{Z}_p} \mathbb{F}_p$. Denote the image of $\mathfrak{m}$ in $\mathbb{T}_{\mathbb{F}_p}$ by $\overline{\mathfrak{m}}$ and similarly for $\overline{\mathfrak{m}}'$. Assume that $\rho_{\overline{\mathfrak{m}}}$ is irreducible.

Let furthermore $K = \mathbb{Q}_p(\zeta_p)$ and $\mathcal{O} = \mathbb{Z}_p[\zeta_p]$ with a primitive p-th root of unity ζ_p. Also let $J := J_1(Np)_{\mathbb{Q}}$ be the Jacobian of $X_1(Np)$ over $\mathbb{Q}$.

Situation II. Let p be any prime and N a positive integer not divisible by p. Define the Hecke algebras

$$\mathbb{T}_{\mathbb{Z}_p} := \mathbb{T}^{(1)}_{\mathbb{Z} \to \mathbb{Z}_p}(S_2(\Gamma_1(N))), \qquad \mathbb{T}'_{\mathbb{Z}_p} := \mathbb{T}^{(p)}_{\mathbb{Z} \to \mathbb{Z}_p}(S_2(\Gamma_1(N))).$$

Let $\mathfrak{m}$ be an ordinary maximal ideal of $\mathbb{T}_{\mathbb{Z}_p}$ with residue field $\mathbb{F} = \mathbb{T}_{\mathbb{Z}_p}/\mathfrak{m}$. Let $\mathfrak{m}' = \mathfrak{m} \cap \mathbb{T}'_{\mathbb{Z}_p}$ and, more generally for $m \in \mathbb{N}$, let

$$\mathfrak{m}^{(m)} = \mathfrak{m} \cap \mathbb{T}^{(m)}_{\mathbb{Z} \to \mathbb{Z}_p}(S_2(\Gamma_1(N))).$$

Let $\mathbb{T}_{\mathbb{F}_p} := \mathbb{T}_{\mathbb{Z}_p} \otimes_{\mathbb{Z}_p} \mathbb{F}_p$ and $\mathbb{T}'_{\mathbb{F}_p} := \mathbb{T}'_{\mathbb{Z}_p} \otimes_{\mathbb{Z}_p} \mathbb{F}_p$. Denote the image of $\mathfrak{m}$ in $\mathbb{T}_{\mathbb{F}_p}$ by $\overline{\mathfrak{m}}$ and similarly for $\overline{\mathfrak{m}}'$. Assume that $\rho_{\overline{\mathfrak{m}}}$ is irreducible.

Let furthermore $K = \mathbb{Q}_p$ and $\mathcal{O} = \mathbb{Z}_p$. Also let $J := J_1(N)_{\mathbb{Q}}$ be the Jacobian of $X_1(N)$ over $\mathbb{Q}$.

2. Hecke algebras, Jacobians and p-divisible groups

Assume we are in Situation I or II. The maximal ideal $\mathfrak{m}$ of $\mathbb{T}_{\mathbb{Z}_p}$ corresponds to an idempotent $e_{\mathfrak{m}} \in \mathbb{T}_{\mathbb{Z}_p}$, in the sense that applying $e_{\mathfrak{m}}$ to any $\mathbb{T}_{\mathbb{Z}_p}$-module is the same as localising the module at $\mathfrak{m}$. Let $\mathcal{G}$ be the p-divisible group $J[p^\infty]_{\mathbb{Q}}$ over $\mathbb{Q}$. Consider the Tate module $T_p J = T_p \mathcal{G} = \varprojlim J[p^n](\overline{\mathbb{Q}})$. It is a $\mathbb{T}_{\mathbb{Z}_p}[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$-module. The Hecke algebra $\mathbb{T}_{\mathbb{Z}_p}$ acts on $T_p J$ and on $\mathcal{G}$, hence so does the idempotent $e_{\mathfrak{m}}$. We put $G = e_{\mathfrak{m}} \mathcal{G}$ and say that this is the p-divisible group over $\mathbb{Q}$ attached to $\mathfrak{m}$. We shall mainly be interested in the p-torsion of G. However, making the detour via p-divisible groups allows us to quote the following theorem by Gross.

Theorem 2.1 (Gross). *Assume we are in [Situation I or II]. Let G be the p-divisible group over $\mathbb{Q}$ attached to $\mathfrak{m}$, as explained above. Let $h = \mathrm{rk}_{\mathbb{Z}_p} \mathbb{T}_{\mathbb{Z}_p, \mathfrak{m}}$, where $\mathbb{T}_{\mathbb{Z}_p, \mathfrak{m}}$ denotes the localisation of $\mathbb{T}_{\mathbb{Z}_p}$ at $\mathfrak{m}$.*

(a) *The p-divisible group G acquires good reduction over $\mathbb{O}$. We write $G_{\mathbb{O}}$ for the corresponding p-divisible group over $\mathbb{O}$. It sits in the exact sequence*

$$0 \to G_{\mathbb{O}}^0 \to G_{\mathbb{O}} \to G_{\mathbb{O}}^e \to 0,$$

where $G_{\mathbb{O}}^e$ is étale and $G_{\mathbb{O}}^0$ is of multiplicative type, i.e. its Cartier dual is étale. The exact sequence is preserved by the action of the Hecke correspondences.

(b) *Over $\mathbb{O}[\zeta_N]$ the p-divisible group $G_{\mathbb{O}[\zeta_N]}$ is isomorphic to its Cartier dual ${}^t G_{\mathbb{O}[\zeta_N]}$. This gives isomorphisms of p-divisible groups over $\mathbb{O}[\zeta_N]$*

$$G_{\mathbb{O}[\zeta_N]}^e \cong {}^t G_{\mathbb{O}[\zeta_N]}^0 \quad and \quad G_{\mathbb{O}[\zeta_N]}^0 \cong {}^t G_{\mathbb{O}[\zeta_N]}^e.$$

(c) *We have $G_{\bar{\mathbb{F}}_p}^e[p] \cong (\mathbb{Z}/p\mathbb{Z})_{\bar{\mathbb{F}}_p}^h$ and $G_{\bar{\mathbb{F}}_p}^0[p] \cong \mu_{p, \bar{\mathbb{F}}_p}^h$.*

Proof. The references in this proof are to [Gross 1990].

(a) The statement on the good reduction is Propositions 12.8 (1) and 12.9 (1). The exact sequence is proved in Propositions 12.8 (4) and 12.9 (3). That it is preserved by the Hecke correspondences is a consequence of the fact that there are no nontrivial morphisms from a connected group scheme to an étale one, whence any Hecke correspondence on G can be restricted to G^0.

(b) The Cartier self-duality of G over $K(\zeta_N)$ is also proved in Propositions 12.8 (1) and 12.9 (1). It extends to a self-duality over $\mathbb{O}[\zeta_N]$. The second statement follows as in (a) from the nonexistence of nontrivial morphisms from G^0 to G^e over $\mathbb{O}[\zeta_N]$; this argument gives $G^0 \cong {}^t G^e$. Applying Cartier duality to this, we also get $G^e \cong {}^t G^0$.

(c) By part (b), G^e and G^0 have equal height. That height is equal to h by Propositions 12.8 (1) and 12.9 (1). The statement is now due to the fact that up to isomorphism the given group schemes are the only ones of rank p^h which are killed by p and which are étale or of multiplicative type, respectively. $\qquad\square$

The last point makes the ordinarity of $\bar{\mathfrak{m}}$ look like the ordinarity of an abelian variety.

Proposition 2.2. *Assume we are in [Situation I or II] and let G be the p-divisible group attached to $\mathfrak{m}$. Then we have an isomorphism $G^0[p](\bar{\mathbb{Q}}_p) \cong \mathbb{T}_{\mathbb{F}_p, \bar{\mathfrak{m}}}$ of $\mathbb{T}_{\mathbb{F}_p, \bar{\mathfrak{m}}}$-modules.*

Proof. Taking the p-torsion of the p-divisible groups in Theorem 2.1 (a), one obtains the exact sequence

$$0 \to G_{\mathbb{O}}^0[p](\overline{\mathbb{Q}}_p) \to G_{\mathbb{O}}[p](\overline{\mathbb{Q}}_p) \to G_{\mathbb{O}}^e[p](\overline{\mathbb{Q}}_p) \to 0 \tag{1}$$

of $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$-modules with Galois action. We also spell out the dualities in part (b) of Theorem 2.1, restricted to the p-torsion on $\overline{\mathbb{Q}}_p$-points:

$$\begin{aligned}
G_{\mathbb{O}[\zeta_N]}^0[p](\overline{\mathbb{Q}}_p) &\cong \mathrm{Hom}_{\mathrm{gr.sch.}/\overline{\mathbb{Q}}_p}(G^e[p]_{\overline{\mathbb{Q}}_p}, \mu_{p, \overline{\mathbb{Q}}_p}) \\
G_{\mathbb{O}[\zeta_N]}^e[p](\overline{\mathbb{Q}}_p) &\cong \mathrm{Hom}_{\mathrm{gr.sch.}/\overline{\mathbb{Q}}_p}(G^0[p]_{\overline{\mathbb{Q}}_p}, \mu_{p, \overline{\mathbb{Q}}_p}).
\end{aligned} \tag{2}$$

These are isomorphisms of $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$-modules, i.e. in particular of $\mathbb{F}_p$-vector spaces. We will from now on identify $\mu_{p, \overline{\mathbb{Q}}_p}(\overline{\mathbb{Q}}_p)$ with $\mathbb{F}_p$ and the group homomorphisms on $\overline{\mathbb{Q}}_p$-points above with $\mathbb{F}_p$-linear ones.

The final ingredient in the proof is that $G^e(\overline{\mathbb{Q}}_p)[\mathfrak{m}] = G^e[p](\overline{\mathbb{Q}}_p)[\overline{\mathfrak{m}}]$ is a one-dimensional $L := \mathbb{T}_{\mathbb{F}_p}/\overline{\mathfrak{m}}$-vector space; see [Gross 1990, Propositions 12.8 (5) and 12.9 (4)]. We quotient the first isomorphism of Equation (2) by $\overline{\mathfrak{m}}$ and obtain

$$G^0[p](\overline{\mathbb{Q}}_p)/\overline{\mathfrak{m}} \cong \mathrm{Hom}_{\mathbb{F}_p}(G^e[p](\overline{\mathbb{Q}}_p)[\overline{\mathfrak{m}}], \mathbb{F}_p) \cong \mathrm{Hom}_{\mathbb{F}_p}(L, \mathbb{F}_p),$$

which is a 1-dimensional L-vector space. Consequently, Nakayama's Lemma applied to the finitely generated $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$-module $G^0[p](\overline{\mathbb{Q}}_p)$ yields a surjection $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}} \twoheadrightarrow G^0[p](\overline{\mathbb{Q}}_p)$. Next we invoke a result from Section 3 of [Kilford and Wiese 2006]. We point out explicitly that all of that section is independent of Section 2 of the same paper, in which Corollary 2.3 is used. From Proposition 3.7 of that paper, it follows that

$$2 \dim_{\mathbb{F}_p} \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}} = \dim_{\mathbb{F}_p} H^1_{\mathrm{par}}(\Gamma, \mathbb{F}_p)_{\overline{\mathfrak{m}}}$$

with $\Gamma = \Gamma_1(Np)$ in Situation I and $\Gamma = \Gamma_1(N)$ in Situation II. At the same time,

$$H^1_{\mathrm{par}}(\Gamma, \mathbb{F}_p)_{\overline{\mathfrak{m}}} \cong J(\mathbb{C})[p]_{\overline{\mathfrak{m}}} \cong G[p](\overline{\mathbb{Q}}_p)$$

(see [Wiese 2007, Proposition 5.3], for example), so we obtain $\dim_{\mathbb{F}_p} \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}} = \dim_{\mathbb{F}_p} G^0[p](\overline{\mathbb{Q}}_p)$ and, thus, $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}} \cong G^0[p](\overline{\mathbb{Q}}_p)$. $\square$

The following result, together with very helpful hints on its proof (amounting to the preceding proposition), was suggested by Kevin Buzzard. See also the discussion before [Emerton 2002, Proposition 6.3] and the letter by Mazur reproduced in the Appendix to [Tilouine 1997].

Corollary 2.3. *Assume we are in Situation I or II and let G be the p-divisible group attached to $\mathfrak{m}$. Then there is an exact sequence*

$$0 \to \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}} \to G[p](\overline{\mathbb{Q}}) \to \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}^{\vee} \to 0$$

of $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$-modules, where the dual is the $\mathbb{F}_p$-linear dual.

Proof. Substituting the isomorphism of Proposition 2.2 into the second isomorphism of Equation (2) gives

$$G^e[p](\overline{\mathbb{Q}}_p) \cong \mathrm{Hom}(\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}, \mathbb{F}_p)$$

as $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$-modules, whence the corollary follows from Equation (1).   □

The following proposition is similar in spirit to Proposition 2.2. It will not be needed in the sequel.

Proposition 2.4. *Assume we are in Situation I or II and let $G = G_{\mathbb{O}}$ be the p-divisible group over $\mathbb{O}$ attached to $\mathfrak{m}$. Then $G^0[p](\mathbb{F}_p(\epsilon))$ and $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$ are isomorphic as $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$-modules.*

Proof. We only give a sketch. Since $G^0[p](\mathbb{F}_p)$ consists of the origin as unique point, $G^0[p](\mathbb{F}_p(\epsilon))$ coincides with the tangent space at 0 of $G^0_{\mathbb{F}_p}[p]$. The latter, however, is equal to the tangent space at 0 of $G_{\mathbb{F}_p}[p]$. On the other hand, its dual, the cotangent space at 0 of $G_{\mathbb{F}_p}[p]$, is isomorphic to $S_k(\Gamma_1(N), \mathbb{F}_p)_{\overline{\mathfrak{m}}}$ for some $k \in \{2, \ldots, p+1\}$. In Situation II, $k = 2$ and this result is well-known. In Situation I, we quote [Edixhoven 1992, Equations 6.7.1 and 6.7.2], as well as [Gross 1990, Proposition 8.13] (note that the ordinarity assumption kills the second summand in that proposition). Consequently, $G^0[p](\mathbb{F}_p(\epsilon))$ is isomorphic to the Hecke algebra on $S_k(\Gamma_1(N), \mathbb{F}_p)_{\overline{\mathfrak{m}}}$ as a Hecke module. In [Kilford and Wiese 2006, Proposition 2.3], it is shown that this algebra is $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$.   □

From Proposition 2.2 and part of the direct proof of Theorem 3.1 we can also derive an isomorphism between $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$ and the image of the reduction map (4).

3. Comparing Frobenius and the Hecke operator T_p

The aim of this section is to discuss and prove the following theorem, which turns out to be an important key to the principal result of this article.

Theorem 3.1. *Assume we are in Situation I or II and let $G^0 = G^0_{\mathbb{O}}$ be the p-divisible group of Theorem 2.1. The action of the geometric Frobenius on the points*

$$G^0_{\mathbb{O}}[p](\mathbb{Q}^{\mathrm{nr}}_p(\zeta_p))$$

is the same as the action of the Hecke operator T_p.

This result is in fact contained in [Gross 1990]. Apart from giving the appropriate citations, we include two more proofs, in the hope that the chosen approaches may find applications in other contexts, too. Due to the Eichler–Shimura congruence relation in Situation II and the reduction of a well-known semistable model of the modular curve in Situation I, for both of these proofs it suffices to compare the geometric Frobenius and Verschiebung on the special fibre of $G^0[p]$. For the

first alternative proof, such a comparison is carried out elementarily — roughly speaking — by working with the tangent space at 0 over $\bar{\mathbb{F}}_p$, in order to have an injective reduction map from characteristic zero to the finite field. On the special fibre elementary computations then suffice. For the second alternative proof, a comparison between geometric Frobenius and Verschiebung has been worked out conceptually by Niko Naumann in the Appendix in the context of Fontaine's theory of Honda systems.

Proof by citation. In Situation I, we cite [Gross 1990, Proposition 12.9 (3)], which says that $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$ acts on the Tate module of G^0 through the p-adic cyclotomic character times $\lambda(T_p^{-1})$, where $\lambda(T_p^{-1})$ is the character sending Frob_p to T_p^{-1}. As we are restricting to $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p(\zeta_p))$ and to $G^0[p]$, the cyclotomic character is trivial and the Galois action on

$$G^0[p](\bar{\mathbb{Q}}_p) = \mathrm{Hom}({}^t G^0[p] \times \bar{\mathbb{Q}}_p, \mu_{p,\bar{\mathbb{Q}}_p})$$

is unramified since ${}^t G^0[p]$ is étale.

In Situation II, we cite [Gross 1990, Proposition 12.8 (4)], and argue as above. Note that by the Eichler–Shimura congruence relation (see the end of the direct proof) the unit u in the citation equals T_p divided by the diamond operator $\langle p \rangle_N$. $\qquad \square$

Direct proof. In the proof we prefer to work with the étale Cartier dual of $G^0[p]$ since we find it more convenient for making formulae explicit. So, ${}^t G^0[p] = \mathrm{Spec}(A)$ is a finite étale group scheme over $\mathbb{O}$ such that

$$ {}^t G^0[p] \times_{\mathbb{O}} \mathbb{Z}_p^{\mathrm{nr}}[\zeta_p] \cong (\mathbb{Z}/p\mathbb{Z})_{\mathbb{Z}_p^{\mathrm{nr}}[\zeta_p]}^h, $$

i.e. $A \otimes_{\mathbb{O}} \mathbb{Z}_p^{\mathrm{nr}}[\zeta_p] \overset{\alpha}{\cong} \prod \mathbb{Z}_p^{\mathrm{nr}}[\zeta_p]$. If $p = 2$, we put $\zeta_2 = -1$. We consider the commutative diagram

$$
\begin{array}{ccccc}
\mathbb{Z}_p^{\mathrm{nr}}[\zeta_p][X]/(X^p-1) & \xleftarrow{\zeta_p \leftarrow Y} & \mathbb{Z}_p^{\mathrm{nr}}[X,Y]/(X^p-1,Y^p-1) & \xrightarrow{Y \mapsto 1+\epsilon} & \bar{\mathbb{F}}_p(\epsilon)[X]/(X^p-1) \\
\downarrow & & \downarrow & & \downarrow \\
\prod \mathbb{Z}_p^{\mathrm{nr}}[\zeta_p] & \xleftarrow{\zeta_p \leftarrow Y} & \prod \mathbb{Z}_p^{\mathrm{nr}}[Y]/(Y^p-1) & \xrightarrow{Y \mapsto 1+\epsilon} & \prod \bar{\mathbb{F}}_p(\epsilon).
\end{array}
$$

Any morphism of group schemes ${}^t G^0[p] \times_{\mathbb{O}} \mathbb{Z}_p^{\mathrm{nr}}[\zeta_p] \to \mu_{p,\mathbb{Z}_p^{\mathrm{nr}}[\zeta_p]}$ corresponds to a Hopf algebra homomorphism as in the left column. Suppose that it maps X to $(\zeta_p^{i_1}, \ldots, \zeta_p^{i_{hp}})$ for $i_j \in \{0, \ldots, p-1\}$. It has a unique lifting to a Hopf algebra homomorphism as in the central column if we impose that X maps to $(Y^{i_1}, \ldots, Y^{i_{hp}})$. As the referee pointed out, this lift gives the first map in the exact sequence

$$0 \to G^0[p](\mathbb{Z}_p^{\mathrm{nr}}[\zeta_p]) \to G^0[p](\mathbb{Z}_p^{\mathrm{nr}}[Y]/(Y^p-1)) \to G^0[p](\mathbb{Z}_p^{\mathrm{nr}}).$$

From the homomorphism in the centre of the diagram we obtain a Hopf algebra homomorphism in the right column, which sends X to $(1 + i_1\epsilon, \ldots, 1 + i_{hp}\epsilon)$. It should be said that the detour via the central column is only necessary for $p = 2$, as for $p > 2$ one can pass directly from the left hand side column to the right hand side via the map $\mathbb{Z}_p^{nr}[\zeta_p] \to \bar{\mathbb{F}}_p(\epsilon)$, sending ζ_p to $1 + \epsilon$.

This process gives us an injective reduction map

$$\text{Hom}_{\text{gr.sch.}/\mathbb{Z}_p^{nr}[\zeta_p]}({}^t G^0[p] \times_{\mathbb{O}} \mathbb{Z}_p^{nr}[\zeta_p], \mu_{p,\mathbb{Z}_p^{nr}[\zeta_p]})$$

$$\to \text{Hom}_{\text{gr.sch.}/\bar{\mathbb{F}}_p(\epsilon)}({}^t G^0[p] \times_{\mathbb{O}} \bar{\mathbb{F}}_p(\epsilon), \mu_{p,\bar{\mathbb{F}}_p(\epsilon)}). \quad (3)$$

In terms of points of $G^0[p]$, the reduction map is the composition

$$G^0[p](\mathbb{Z}_p^{nr}[\zeta_p]) \hookrightarrow G^0[p](\mathbb{Z}_p^{nr}[Y]/(Y^p - 1)) \to G^0[p](\bar{\mathbb{F}}_p[\epsilon]). \quad (4)$$

The reduction map is compatible for the action induced by the Hecke correspondences.

Next, we describe the geometric Frobenius on the points $G^0[p](\mathbb{Q}_p^{nr}(\zeta_p))$ and $G^0[p](\bar{\mathbb{F}}_p(\epsilon))$. We consider the commutative diagram

$$\begin{array}{ccc}
\text{Hom}_{\text{gr.sch.}/\mathbb{Z}_p^{nr}[\zeta_p]}({}^t G^0[p] \times \mathbb{Z}_p^{nr}[\zeta_p], \mu_{p,\mathbb{Z}_p^{nr}[\zeta_p]}) & \xrightarrow{\sim} (A \otimes \mathbb{Z}_p^{nr}[\zeta_p])^{\text{gl}} \xrightarrow{\sim} & G^0[p](\mathbb{Z}_p^{nr}[\zeta_p]) \\
\downarrow \sim & & \\
\text{Hom}_{\mathbb{Z}_p^{nr}[\zeta_p]-\text{HA}}(\mathbb{Z}_p^{nr}[\zeta_p][X]/(X^p - 1), A \otimes \mathbb{Z}_p^{nr}[\zeta_p]) & & \\
& A \otimes \mathbb{Z}_p^{nr}[\zeta_p] \xrightarrow{\sim \text{ ev}} \text{Hom}_{\mathbb{O}}({}^t A, \mathbb{Z}_p^{nr}[\zeta_p]).
\end{array}$$

It is well-known that a Hopf algebra homomorphism

$$\psi : \mathbb{Z}_p^{nr}[\zeta_p][X]/(X^p - 1) \to A \otimes_{\mathbb{O}} \mathbb{Z}_p^{nr}[\zeta_p]$$

is uniquely given by the "group-like element" $\psi(X) = \sum a_i \otimes s_i$, giving the upper left bijection. On the bottom right, we have the evaluation isomorphism

$$A \otimes_{\mathbb{O}} \mathbb{Z}_p^{nr}[\zeta_p] \to \text{Hom}_{\mathbb{O}}(\text{Hom}_{\mathbb{O}}(A, \mathbb{O}), \mathbb{Z}_p^{nr}[\zeta_p])$$

which is defined by $\text{ev}(a \otimes s)(\varphi) = \varphi(a)s$. We use that as $\mathbb{O}$-modules ${}^t A = \text{Hom}_{\mathbb{O}}(A, \mathbb{O})$ with $G^0[p] = \text{Spec}({}^t A)$, as well as the freeness of A. It is also well-known that the evaluation map gives rise to the upper right bijection.

Let now ϕ be the geometric Frobenius in $\text{Gal}(\mathbb{Q}_p^{nr}(\zeta_p)/\mathbb{Q}_p(\zeta_p))$. Its action on $\text{Hom}_{\mathbb{O}}({}^t A, \mathbb{Z}_p^{nr}[\zeta_p])$ is by composition. Via the evaluation map it is clear that ϕ acts on an element $a \otimes s \in A \otimes_{\mathbb{O}} \mathbb{Z}_p^{nr}[\zeta_p]$ by sending it to $a \otimes \phi(s)$. Consequently, the morphism ψ^ϕ which is obtained by applying ϕ to ψ is uniquely determined by

$\psi^\phi(X) = \sum a_i \otimes \phi(s_i)$. A similar statement holds for the reduction. We note that this implies the compatibility of the reduction map with the ϕ-action.

Next we show that the action of geometric Frobenius on the image of (4) inside the tangent space $G^0[p](\bar{\mathbb{F}}_p(\epsilon))$ coincides with the action induced by Verschiebung on $G^0_{\mathbb{F}_p}[p]$. The étale algebra $A \otimes_{\mathcal{O}} \mathbb{F}_p$ can be written as a product of algebras of the form $\mathbb{F}_p[X]/(f)$ with f an irreducible polynomial. An elementary calculation on the underlying rings gives the commutativity of the diagram

$$
\begin{array}{ccccc}
\mathbb{F}_p[X]/(f) \otimes_{\mathbb{F}_p} \bar{\mathbb{F}}_p(\epsilon) & \xrightarrow{F\otimes 1} & \mathbb{F}_p[X]/(f) \otimes_{\mathbb{F}_p} \bar{\mathbb{F}}_p(\epsilon) & \xrightarrow{1\otimes\phi^{-1}} & \mathbb{F}_p[X]/(f) \otimes_{\mathbb{F}_p} \bar{\mathbb{F}}_p(\epsilon) \\
\downarrow{\alpha} & & & & \downarrow{\alpha} \\
\prod_{i=1}^{d} \bar{\mathbb{F}}_p(\epsilon) & & \xrightarrow{\qquad\prod \phi^{-1}\qquad} & & \prod_{i=1}^{d} \bar{\mathbb{F}}_p(\epsilon),
\end{array}
\tag{5}
$$

where F denotes the absolute Frobenius on ${}^tG^0_{\mathbb{F}_p}[p]$ (defined by $X \mapsto X^p$), which by duality gives the Verschiebung on $G^0_{\mathbb{F}_p}[p]$. We point out that ϕ leaves ϵ invariant.

Let now $\psi : \bar{\mathbb{F}}_p(\epsilon)[X]/(X^p - 1) \to A \otimes_{\mathcal{O}} \bar{\mathbb{F}}_p(\epsilon)$ be an $\bar{\mathbb{F}}_p(\epsilon)$-Hopf algebra homomorphism in the image of (3). It is uniquely given by $\psi(X) = \sum_i a_i \otimes s_i$, and under the identification

$$
A \otimes_{\mathcal{O}} \bar{\mathbb{F}}_p(\epsilon) \overset{\alpha}{\cong} \prod_{j=1}^{hp} \bar{\mathbb{F}}_p(\epsilon)
$$

we get $\psi(X) = (1 + i_1\epsilon, \ldots, 1 + i_{hp}\epsilon)$ with $i_j \in \mathbb{F}_p$ as we have seen above, which is invariant under the arithmetic Frobenius of the bottom row of (5). Hence, $\phi^{-1}(F(\sum_i a_i \otimes s_i)) = \sum_i a_i \otimes s_i$, so that $F(\sum_i a_i \otimes s_i) = \sum_i a_i \otimes \phi(s_i)$. This proves that the geometric Frobenius and Verschiebung coincide on the image of (4) inside $G^0[p](\bar{\mathbb{F}}_p(\epsilon))$.

We now finish the proof. In Situation II, the Eichler–Shimura relation $T_p = \langle p \rangle F + V$ holds on the special fibre of $G[p]$ (see the proof of [Gross 1990, Proposition 12.8 (2)]). Since F is zero on $G^0_{\mathbb{F}_p}[p]$, we get $T_p = V$ on it. We obtain the theorem in this situation since V coincides with ϕ on the image of (4), as we just saw.

In Situation I, we know that $G^0_{\mathbb{F}_p}[p]$ is naturally part of the p-torsion of the Jacobian of the Igusa curve $I_1(N)_{\mathbb{F}_p}$; but on the Igusa curve Verschiebung acts as T_p (see the proof of [Gross 1990, Proposition 12.9 (2)] for both these facts). Hence, we can argue as above and get the theorem also for $p > 2$. $\qquad\square$

More conceptual proof. In both situations, Theorem A.1 of Naumann gives an isomorphism between $G^0[p](\mathbb{Q}_p^{\mathrm{nr}}(\zeta_p))$ and the Dieudonné module M attached to the special fibre $G^0_{\mathbb{F}_p}[p]$. Under this isomorphism the geometric Frobenius $\phi \in \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{nr}}(\zeta_p)/\mathbb{Q}_p(\zeta_p))$ on $G^0[p](\mathbb{Q}_p^{\mathrm{nr}}(\zeta_p))$ is identified with Verschiebung on the Dieudonné module. The isomorphism is compatible with the Hecke action. Using

the same citations as at the end of the direct proof one immediately concludes that the equality $T_p = V$ holds on the Dieudonné module M, finishing the proof. $\qquad \square$

Remark 3.2. (a) Conceptually, taking $\mathbb{Z}_p^{\mathrm{nr}}[\zeta_p]$-points is the same as taking $\mathbb{Z}_p^{\mathrm{nr}}$-points of the Weil restriction from $\mathbb{O}$ to $\mathbb{Z}_p$ and similarly for $\mathbb{Q}_p^{\mathrm{nr}}(\zeta_p)$. So, we could have formulated Theorem 3.1 in terms of the Weil restriction.

(b) We point out again that we are using the conventions of [Gross 1990]. Hence, the representation on the Jacobian must be tensored by the corresponding Dirichlet character ϵ (the nebentype) in order to obtain $\rho_{\overline{\mathfrak{m}}}$ (see [Gross 1990, p. 486]).

(c) A theorem by Deligne (see, for instance, [Edixhoven 1992, Theorem 2.5] or [Gross 1990, Proposition 12.1]) describes the restriction of $\rho_{\overline{\mathfrak{m}}}$ to a decomposition group at p in the ordinary case as

$$\begin{pmatrix} \chi_p^{k-1}\lambda(\epsilon(p)/a_p) & * \\ 0 & \lambda(a_p) \end{pmatrix},$$

where χ_p is the mod p cyclotomic character, $\lambda(u)$ is the unramified character sending the arithmetic Frobenius Frob_p to u and $a_p \equiv T_p \mod \overline{\mathfrak{m}}$. When we restrict to $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p(\zeta_p))$, the cyclotomic character acts trivially and we see that Theorem 3.1 is in accordance with Deligne's description.

Let f be a Katz eigenform of weight 1 over $\mathbb{F}_p$ with eigenvalue $a_p^{(1)}$ for the weight 1 Hecke operator $T_p^{(1)}$. As explained in [Edixhoven 2006, Section 4], one can embed f into weight p in two different ways. On the span in weight p, the Hecke operator T_p has the eigenvalues a_p and $\epsilon(p)/a_p$ and they satisfy $a_p^{(1)} = a_p + \epsilon(p)/a_p$ (see [Wiese 2007, Proposition 8.4]). The mod p Galois representation attached to f coincides with the one attached to a weight p form. We suppose that this representation is of weight one, which is known for $p > 2$ and for many cases with $p = 2$ and is expected to be true without any exception. Then the characteristic polynomial of Frob_p acting on that representation equals $X^2 - a_p^{(1)}X + \epsilon(p)$ and is thus like any characteristic polynomial of a modular Galois representation at any unramified prime.

4. Application to multiplicities

We first state a slight strengthening of a well-known theorem by Boston, Lenstra and Ribet.

Proposition 4.1 (Boston, Lenstra, Ribet). *Assume we are in Situation I or II. Let m be an integer. Then the $\mathbb{F}[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$-module $J(\overline{\mathbb{Q}})[\mathfrak{m}^{(m)}]$ is the direct sum of r copies of $\rho_{\overline{\mathfrak{m}}} \otimes \epsilon^{-1}$ for some $r \geq 1$ and Dirichlet character ϵ corresponding to $\overline{\mathfrak{m}}$.*

The integer r is called the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}^{(m)}]$.

Proof. The same proof as in the original proposition works. More precisely, one considers the two representations $\rho_{\overline{\mathfrak{m}}} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F})$ and $\sigma : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(J(\overline{\mathbb{Q}})[\mathfrak{m}^{(m)}])$. By Chebotarev's density theorem we know that every conjugacy class of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})/\ker(\sigma \otimes \epsilon)$ is hit by a Frobenius element Frob_l for some $l \nmid Npm$.

The Eichler–Shimura congruence relation $T_l = \langle l \rangle F + V$ holds on $J_{\mathbb{F}_l}$ (taking J here over $\mathbb{Z}[\frac{1}{Np}]$) for all primes $l \nmid Npm$. Hence, the minimal polynomial of Frob_l on the Jacobian divides $X^2 - T_l/\langle l \rangle \cdot X + l/\langle l \rangle$. But T_l acts as a_l on $J(\overline{\mathbb{Q}})[\mathfrak{m}^{(m)}]$ and $X^2 - a_l X + \epsilon(l)l$ (with $T_l \equiv a_l \mod \overline{\mathfrak{m}}$) is the characteristic polynomial of $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_l)$. Consequently, $(\sigma \otimes \epsilon)(g)$ is annihilated by the characteristic polynomial of $\rho_{\overline{\mathfrak{m}}}(g)$ for all $g \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Hence, Theorem 1 of [Boston et al. 1991] gives the result. $\square$

The notion of multiplicity is sometimes formulated in a way only depending on the representation and not on a particular piece of one Jacobian; see for example [Ribet and Stein 2001, Definition 3.3]. The next corollary says that one can read off multiplicities from properties of Hecke algebras.

Corollary 4.2. *Assume we are in Situation I or II. Let r be the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$. Then*

$$r = \tfrac{1}{2}(\dim_{\mathbb{F}} \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}] + 1).$$

Proof. Buzzard [2001] explains the exactness of the sequence

$$0 \to G^0(\overline{\mathbb{Q}}_p)[\mathfrak{m}] \to G(\overline{\mathbb{Q}}_p)[\mathfrak{m}] \to G^e(\overline{\mathbb{Q}}_p)[\mathfrak{m}] \to 0.$$

Via Corollary 2.3 we obtain the exact sequence

$$0 \to \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}] \to J(\overline{\mathbb{Q}}_p)[\mathfrak{m}] \to \left(\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}/\overline{\mathfrak{m}}\right)^{\vee} \to 0,$$

from which one reads off the claim by counting dimensions. $\square$

In [Buzzard 2001] Buzzard proved that the multiplicity on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ of $\rho_{\overline{\mathfrak{m}}}$ of weight one is 1 if $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is nonscalar. We include this as a lemma.

Lemma 4.3. *Assume we are in Situation I or II and $\rho_{\overline{\mathfrak{m}}}$ is of weight one. If $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is not a scalar matrix, the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is 1.*

Proof. We first record that T_p acts as a scalar (in $\mathbb{F}$) on $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}]$. Suppose that the multiplicity r of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is greater than 1. Then $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}] = G^0(p)(\overline{\mathbb{Q}})[\overline{\mathfrak{m}}]$ has dimension $2r - 1 > 1$ by (the proof of) Corollary 4.2. Hence, $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ does not act as a scalar on $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}]$, as it is nonscalar on $J(\overline{\mathbb{Q}})[\mathfrak{m}] \cong \rho_{\overline{\mathfrak{m}}}^r$ by assumption. From Theorem 3.1 we obtain a contradiction, since it implies that T_p does not act as a scalar on $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}]$ either. $\square$

Theorem 4.4. *Assume we are in Situation I or II and $\rho_{\overline{\mathfrak{m}}}$ has weight one and $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is conjugate to $\left(\begin{smallmatrix} a & * \\ 0 & a \end{smallmatrix}\right)$. The following statements are equivalent:*

(a) *The representation $\rho_{\overline{\mathfrak{m}}}$ comes from a Katz cusp form of weight 1 on $\Gamma_1(N)$ over $\overline{\mathbb{F}}_p$ and the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is 1.*

(b) $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}] \subsetneq \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}']$.

(c) *T_p does not act as a scalar on $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}']$ (inside $J(\overline{\mathbb{Q}})[\mathfrak{m}'] \cong \oplus \rho_{\overline{\mathfrak{m}}}$ as an $\mathbb{F}$-vector space).*

(d) *The multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is 1, its multiplicity on $J(\overline{\mathbb{Q}})[\mathfrak{m}']$ is 2, and $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is nonscalar.*

Proof. (a) $\Rightarrow$ (b): By Corollary 4.2, the $\mathbb{F}$-dimension of $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}]$ is 1, hence, so is the dimension of $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}^{\vee}/\overline{\mathfrak{m}}$. Thus, Nakayama's Lemma yields that $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}$ is Gorenstein, i.e. that it is isomorphic to its dual as a module over itself. By the q-expansion principle, the dual is $S_p(\Gamma_1(N), \mathbb{F}_p)_{\overline{\mathfrak{m}}}$. By [Edixhoven 2006, Propostion 6.2] or [Wiese 2007, Proposition 8.4] the existence of a corresponding weight 1 form is equivalent to $S_p(\Gamma_1(N), \mathbb{F}_p)_{\overline{\mathfrak{m}}}[\overline{\mathfrak{m}}']$ being 2-dimensional. This establishes (b), since $S_p(\Gamma_1(N), \mathbb{F}_p)_{\overline{\mathfrak{m}}}[\overline{\mathfrak{m}}]$, which is isomorphic to $(\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}/\overline{\mathfrak{m}})^{\vee}$, is 1-dimensional as an $\mathbb{F}$-vector space.

(b) $\Rightarrow$ (c): This is evident.

(c) $\Rightarrow$ (d): First of all, $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}']$ is at least 2-dimensional (as an $\mathbb{F}$-vector space). From Theorem 3.1 we know that T_p acts as the inverse of Frob_p on $G^0[p](\overline{\mathbb{Q}})$. We conclude that $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ cannot be scalar. Hence, Lemma 4.3 yields that the multiplicity r of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is equal to 1. If the multiplicity s of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}']$ were bigger than 2, then $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\mathfrak{m}']$ would be at least 4-dimensional by an argument as in the proof of Corollary 4.2. Then it follows that it must contain at least two linearly independent eigenvectors for T_p corresponding to at least two copies of $\rho_{\overline{\mathfrak{m}}}$, contradicting the fact that $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}[\overline{\mathfrak{m}}]$ is 1-dimensional.

(d) $\Rightarrow$ (a): Clearly, $\overline{\mathfrak{m}} \neq \overline{\mathfrak{m}}'$. Hence, $\mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}/\overline{\mathfrak{m}} \neq \mathbb{T}_{\mathbb{F}_p, \overline{\mathfrak{m}}}/\overline{\mathfrak{m}}'$ and, dually,

$$S_p(\Gamma_1(N), \mathbb{F}_p)_{\overline{\mathfrak{m}}}[\overline{\mathfrak{m}}] \subsetneq S_p(\Gamma_1(N), \mathbb{F}_p)_{\overline{\mathfrak{m}}}[\overline{\mathfrak{m}}'],$$

which implies the existence of a corresponding weight 1 form, again by [Edixhoven 2006, Proposition 6.2] or [Wiese 2007, Proposition 8.4]. $\square$

We now state and prove the principal result of this article.

Corollary 4.5. *Assume we are in Situation I or II and $\rho_{\overline{\mathfrak{m}}}$ is of weight one. If $p = 2$, also assume that a weight 1 Katz form of level N exists which gives rise to $\rho_{\overline{\mathfrak{m}}}$.*

Then the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is 1 if and only if $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is nonscalar.

Proof. By [Edixhoven 1992, Theorem 4.5] together with the remark at the end of the introduction to that article, the existence of the corresponding weight 1 form

is also guaranteed for $p > 2$. First suppose that the multiplicity is 1. If $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ has two distinct eigenvalues, then it clearly is nonscalar. If $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is conjugate to $\left(\begin{smallmatrix} a & * \\ 0 & a \end{smallmatrix}\right)$, then Theorem 4.4 shows that $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is nonscalar. On the other hand, if $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is nonscalar, Lemma 4.3 implies that the multiplicity is 1. □

The following corollary gives a different, somewhat cleaner formulation of the results on multiplicities. It suggests that instead of working with the full Hecke algebra, one should restrict to the prime-to-p one.

Corollary 4.6. *Assume we are in Situation I or II. If $p = 2$, also assume that $\rho_{\overline{\mathfrak{m}}}$ is of weight one if and only if there exists a weight 1 Katz form of level N which gives rise to $\rho_{\overline{\mathfrak{m}}}$.*

Then the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}']$ is 1 if and only if $\rho_{\overline{\mathfrak{m}}}$ is ramified at p.

Proof. As in the previous proof, for $p > 2$ by [Edixhoven 1992, Theorem 4.5], together with the remark at the end of the introduction to that article, the existence of a weight 1 form is equivalent to the attached representation being of weight one. If $\rho_{\overline{\mathfrak{m}}}$ is ramified at p, the result follows from Theorem 6.1 of [Buzzard 2001]. For, it gives that $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is isomorphic to precisely one copy of $\rho_{\overline{\mathfrak{m}}}$. Moreover, the localisation at $\overline{\mathfrak{m}}'$ of $\mathbb{T}_{\mathbb{F}_p}$ (as a $\mathbb{T}'_{\mathbb{F}_p}$-module) is equal to $\mathbb{T}'_{\mathbb{F}_p, \overline{\mathfrak{m}}'}$, as otherwise a weight one form would exist e.g. by [Wiese 2007, Proposition 8.1]. Hence, $\overline{\mathfrak{m}} = \overline{\mathfrak{m}}'$ and $J(\overline{\mathbb{Q}})[\mathfrak{m}] = J(\overline{\mathbb{Q}})[\mathfrak{m}']$.

Suppose now that $\rho_{\overline{\mathfrak{m}}}$ is unramified at p. If $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is scalar, it suffices to apply Corollary 4.5. If $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is conjugate to $\left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right)$ with $b \neq 0$, then the result is obtained from Corollary 4.5 together with the implication $(a) \Rightarrow (d)$ of Theorem 4.4. If, finally, $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ has two distinct eigenvalues, then there are two maximal ideals $\mathfrak{m} = \mathfrak{m}_1, \mathfrak{m}_2$ with $\rho_{\overline{\mathfrak{m}}_1} \cong \rho_{\overline{\mathfrak{m}}_2}$, since the operator T_p has two distinct eigenvalues on $S_p(\Gamma_1(N), \mathbb{F}_p)[\overline{\mathfrak{m}}']$ by the formula in [Wiese 2007, Proposition 8.4], namely the same as $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)^{-1}$. Consequently, $J(\overline{\mathbb{Q}})[\mathfrak{m}_1] \oplus J(\overline{\mathbb{Q}})[\mathfrak{m}_2] = J(\overline{\mathbb{Q}})[\mathfrak{m}']$, finishing this proof. □

Corollary 4.7. *Assume we are in Situation I or II and $\rho_{\overline{\mathfrak{m}}}$ is of weight one. Assume also that the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}']$ is 2. Then the following statements are equivalent.*

(a) *The multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is 1 and a weight 1 Katz form of level N exists which gives rise to $\rho_{\overline{\mathfrak{m}}}$.*

(b) *$\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ is nonscalar.*

Proof. We have seen the implication $(a) \Rightarrow (b)$ in Corollary 4.5. By Lemma 4.3, we obtain from $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ being nonscalar that the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}]$ is 1. From the assumption the inequality $\overline{\mathfrak{m}} \neq \overline{\mathfrak{m}}'$ follows, implying the existence of the weight 1 form as above by [Edixhoven 2006, Proposition 6.2] or [Wiese 2007, Proposition 8.4]. □

If one could prove that the multiplicity of $\rho_{\overline{\mathfrak{m}}}$ on $J(\overline{\mathbb{Q}})[\mathfrak{m}']$ is always equal to 2 in the unramified situation, Corollary 4.7 would extend weight lowering for $p = 2$ to $\rho_{\overline{\mathfrak{m}}}(\mathrm{Frob}_p)$ being nonscalar.

Appendix

by Niko Naumann

Let p be a prime, $A := \mathbb{Z}_p$, $A' := \mathbb{Z}_p[\zeta_p]$, $K := \mathbb{Q}_p$, $K' := \mathbb{Q}_p(\zeta_p)$ and $K' \subseteq \overline{K}$ an algebraic closure. We have the inertia subgroup $I \subseteq G_{K'} := \mathrm{Gal}(\overline{K}/K')$ and for a $G_{K'}$-module V we denote by τ the geometric Frobenius acting on the inertia invariants V^I. If G/A' is a finite flat group-scheme, always assumed to be commutative, we denote by M the Dieudonné module of its special fiber and by $V : M \to M$ the Verschiebung.

Theorem A.1. *Let G/A' be a finite flat group-scheme which is connected with étale Cartier dual and annihilated by multiplication with p. Then $G(\overline{K})^I = G(\overline{K})$ and there is an isomorphism $\phi : G(\overline{K})^I \to M$ of $\mathbb{F}_p$-vector spaces such that $\phi \circ \tau = V \circ \phi$.*

The assumption that $pG = 0$ cannot be dropped:

Proposition A.2. *For every $n \geq 2$ there is a finite flat group-scheme G/A' of order p^n which is connected with an étale dual and such that $G(\overline{K})^I \simeq \mathbb{Z}/p\mathbb{Z}$ with τ acting trivially and $V \neq 1$ on the Dieudonné module of the special fiber of G.*

Proof of Theorem A.1. Denoting by G' the Cartier dual of G/A' we have an isomorphism of $G_{K'}$-modules

$$G(\overline{K}) \simeq \mathrm{Hom}(G'(\overline{K}), \mu_{p^\infty}(\overline{K})) \overset{(pG'=0)}{=} \mathrm{Hom}(G'(\overline{K}), \mu_p(\overline{K})).$$

Since $G'(\overline{K})$ is unramified because G/A' is étale and $\mu_p(\overline{K})$ is unramified because $\zeta_p \in K'$ we see that $G(\overline{K})^I = G(\overline{K})$. Letting p^n denote the order of G we have $\dim_{\mathbb{F}_p} G(\overline{K})^I = \dim_{\mathbb{F}_p} G(\overline{K}) = n = \dim_{\mathbb{F}_p} M$.

In the rest of the proof we use the explicit quasi-inverse to J.-M. Fontaine's functor associating with G a finite Honda system in order to determine the action of τ on $G(\overline{K})^I$ [Fontaine 1977; Conrad 1999].

Let (M, L) be the finite Honda system over A' associated with G/A'. Recall that M is the Dieudonné module of the special fiber of G and $L \subseteq M_{A'}$ is an A'-submodule where $M_{A'}$ is an A'-module functorially associated with M [Fontaine 1977, Chapter IV, Section 2]. We claim that $L = M_{A'}$: Let $\mathfrak{m} \subseteq A'$ denote the maximal ideal. Using the notation of [Conrad 1999, Section 2], the defining epimorphism of A'-modules $M_{A'} \to \mathrm{coker}(\mathscr{F}_M)$ factors through an epimorphism $M_{A'}/\mathfrak{m}M_{A'} \to \mathrm{coker}(\mathscr{F}_M)$ because $\mathfrak{m} \cdot \mathrm{coker}(\mathscr{F}_M) = 0$ [Conrad 1999, Lemma 2.4].

Denoting by l the length of a module we have

$$l_{A'}(\operatorname{coker}(\mathscr{F}_M)) = l_A(\ker F) = l_A(\ker(p : M \to M)) = l_A(M) = n$$

where the first equality follows from [Conrad 1999, 2.4], the second because $\ker F = \ker(p : M \to M)$ since V is bijective, and the third since $pM = 0$. On the other hand, the canonical morphism of A'-modules $\iota_M : M \otimes_A A' \to M_{A'}$ is an isomorphism by [Fontaine 1977, Chapter IV, Proposition 2.5] using again that V is bijective. Thus

$$l_{A'}(M_{A'}/\mathrm{m}M_{A'}) = l_{A'}(M \otimes_A A'/\mathrm{m}) = l_{A'}(M/pM) = l_A(M) = n$$

and $M_{A'}/\mathrm{m}M_{A'} \xrightarrow{\simeq} \operatorname{coker}(\mathscr{F}_M)$. Since $L/\mathrm{m}L \xrightarrow{\simeq} \operatorname{coker}(\mathscr{F}_M)$ holds for every finite Honda system we see that the inclusion $L \subseteq M_{A'}$ induces an isomorphism $L/\mathrm{m}L \xrightarrow{\simeq} M_{A'}/\mathrm{m}M_{A'}$ and Nakayama's lemma implies that $L = M_{A'}$.

Fix $\pi \in \overline{K}$ with $\pi^{p-1} = -p$, then $K' = K(\pi)$: This is obvious for $p = 2$ and for $p \neq 2$ it follows from local class field theory and the norm computation $\mathrm{N}_K^{K'}(\zeta_p - 1) = \mathrm{N}_K^{K(\pi)}(\pi) = p$. Note that $\pi \in A'$ is a local uniformizer. Let K'^{ur} denote the completion of the maximal unramified extension of K' inside $\overline{K}$ and $\mathbb{O} \subseteq K'^{ur}$ its ring of integers.

By [Fontaine 1977, Remarque on p. 218] and the fact that $L = M_{A'}$ we see that reduction induces an isomorphism

$$G(\overline{K})^I = G(K'^{ur}) = G(\mathbb{O})$$
$$\xrightarrow{\simeq} \left\{ \phi \in \operatorname{Hom}_{\mathrm{D}_{\mathbb{F}_p}}(M, \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})) \mid w'^c \circ \phi_{A'} = 0 \right\} \quad (1)$$

where $\mathrm{D}_{\mathbb{F}_p} = \mathbb{F}_p[F, V]$ is the Dieudonné ring, CW denotes Witt covectors [Fontaine 1977, Chapter II, Section 1],

$$w'^c : \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})_{A'} \to K'^{ur}/\pi^2\mathbb{O}$$

is as in [Fontaine 1977, Chapter IV, Section 3] and $\phi_{A'} : M_{A'} \to \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})_{A'}$ is induced by ϕ. By construction of w'^c we have for $\phi \in \operatorname{Hom}_{\mathrm{D}_{\mathbb{F}_p}}(M, \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O}))$ a commutative diagram

in which

$$w^c((x_{-n})_{n \geq 0}) = \sum_{n=0}^{\infty} p^{-n} \hat{x}_{-n}^{p^n}$$

with $\hat{x}_{-n} \in \pi\mathbb{O}$ lifting x_{-n}, $\tilde{w} = w^c \otimes 1$ is the A'-linear extension of w^c and $\iota_{\mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})}$ is surjective by [Fontaine 1977, Chapter IV, Proposition 2.5] since $\mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})$ is V-divisible. It is easy to see that we have

$$w'^c \circ \phi_{A'} = 0 \Leftrightarrow w^c \circ \phi = 0. \tag{2}$$

Combining (2) and (1) we obtain an isomorphism

$$G(\overline{K})^I \overset{\sim}{\to} \{\phi \in \mathrm{Hom}_{\mathrm{D}_{\mathbb{F}_p}}(M, \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})) \mid w^c \circ \phi = 0\}. \tag{3}$$

Now we need to study $\ker(w^c)$. We will use the isomorphism of $\overline{\mathbb{F}}_p$-vector spaces

$$\pi\mathbb{O}/\pi^2\mathbb{O} \overset{\cdot\pi}{\to} \mathbb{O}/\pi\mathbb{O} \simeq \overline{\mathbb{F}}_p \tag{4}$$

to describe elements of $\mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})$ as covectors $(y_{-n})_{n \geq 0}$ with $y_{-n} \in \overline{\mathbb{F}}_p$. Of course, since (4) is not multiplicative, some care has to be taken with this. We denote by $\sigma : \overline{\mathbb{F}}_p \to \overline{\mathbb{F}}_p$, $\sigma(x) = x^p$ the absolute Frobenius and claim that

$$\ker(w^c) = \{(y_{-n})_n \mid y_{-n} \in \overline{\mathbb{F}}_p, \ y_{-1} = y_0^{\sigma^{-1}}\}. \tag{5}$$

To see this, let $(x_{-n})_n \in \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})$ be given, choose $\hat{x}_{-n} \in \pi\mathbb{O}$ lifting x_{-n} and write $\hat{x}_{-n} = \pi\hat{y}_{-n}$ with $\hat{y}_{-n} \in \mathbb{O}$. Then we compute in $K'^{ur}/\pi^2\mathbb{O}$:

$$w^c((x_{-n})) = \sum_{n=0}^{\infty} p^{-n}(\pi\hat{y}_{-n})^{p^n} \overset{(\pi^{p-1} = -p)}{=} \sum_{n=0}^{\infty} (-1)^n \pi^{p^n - n(p-1)} \hat{y}_{-n}^{p^n} = \pi(\hat{y}_0 - \hat{y}_{-1}^p),$$

using that $p^n - n(p-1) \geq 2$ for all $n \geq 2$. Now (5) is obvious.

Next, we claim that the subset

$$\mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O}) \supseteq \mathcal{M} := \{(y_0^{\sigma^{-n}})_{n \geq 0} \mid y_0 \in \overline{\mathbb{F}}_p\} \tag{6}$$

is a $\mathrm{D}_{\mathbb{F}_p}$-submodule. First note that $F = 0$ on $\mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})$ so we will consider it as a $\mathrm{D}_{\mathbb{F}_p}/F = \mathbb{F}_p[V]$-module in the following. Since all products in $\pi\mathbb{O}/\pi^2\mathbb{O}$ are zero we have

$$(x_{-n}) + (y_{-n}) = (x_{-n} + y_{-n})$$

in $\mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})$ and $\mathcal{M}$ is indeed a $\mathbb{F}_p$-submodule, visibly stable under V.

We claim that the inclusion (6) induces an isomorphism

$$\mathrm{Hom}_{\mathbb{F}_p[V]}(M, \mathcal{M}) \overset{\sim}{\to} \{\phi \in \mathrm{Hom}_{\mathrm{D}_{F_p}}(M, \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})) \mid w^c \circ \phi = 0\}. \tag{7}$$

Since $M \subseteq \ker(w^c)$ by (5) we only need to see that an $\mathbb{F}_p[V]$-linear morphism

$$\phi : M \to \mathrm{CW}_{\mathbb{F}_p}(\pi\mathbb{O}/\pi^2\mathbb{O})$$

with $\phi(M) \subseteq \ker(w^c)$ factors through $\mathcal{M}$: For every $m \in M$ and $n \geq 0$ we have, writing $\phi(m) =: (y_{-n})$ with $y_{-n} \in \bar{\mathbb{F}}_p$,

$$0 = w^c(\phi(V^n m)) = w^c(V^n(\phi(m))) = w^c((\dots, y_{-n-1}, y_{-n})),$$

thus $y_{-n-1} = y_{-n}^{\sigma^{-1}}$ by (5) and as this is true for every $n \geq 0$ we get $\phi(m) \in \mathcal{M}$.

To proceed, note that

$$\mathcal{M} \to \bar{\mathbb{F}}_p \ , \ (y_0^{\sigma^{-n}}) \mapsto y_0 \tag{8}$$

is an isomorphism of $\mathbb{F}_p[V]$-modules if one defines $V(\alpha) := \alpha^{\sigma^{-1}}$ for $\alpha \in \bar{\mathbb{F}}_p$. Denoting by $\Phi : G(\bar{K})^I \xrightarrow{\sim} \mathrm{Hom}_{\mathbb{F}_p[V]}(M, \bar{\mathbb{F}}_p)$ the isomorphism obtained by combining (3), (7) and (8), by construction we have a commutative diagram

$$\begin{array}{ccc}
G(\bar{K})^I & \xrightarrow{\ \Phi\ } & \mathrm{Hom}_{\mathbb{F}_p[V]}(M, \bar{\mathbb{F}}_p) \\
\downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle \mathrm{Hom}(V, \bar{\mathbb{F}}_p)} \\
G(\bar{K})^I & \xrightarrow{\ \Phi\ } & \mathrm{Hom}_{\mathbb{F}_p[V]}(M, \bar{\mathbb{F}}_p).
\end{array} \tag{9}$$

Let e_i (resp. ϕ_i) ($1 \leq i \leq n$) be an $\mathbb{F}_p$-basis of M (resp. $\mathrm{Hom}_{\mathbb{F}_p[V]}(M, \bar{\mathbb{F}}_p)$) and define $V e_i =: \sum_j a_{ij} e_j$, hence $A := (a_{ij}) \in \mathrm{Gl}_n(\mathbb{F}_p)$, $\psi_i := \mathrm{Hom}(V, \bar{\mathbb{F}}_p)(\phi_i) =: \sum_j b_{ij} \phi_j$, hence $B := (b_{ij}) \in \mathrm{Gl}_n(\mathbb{F}_p)$ and $C := (\phi_i(e_j)) \in \mathrm{Gl}_n(\bar{\mathbb{F}}_p)$. By definition, A is a representing matrix of $V : M \to M$ and by (9) B is a representing matrix for τ. So we will be done if we can show that A and B are conjugate over $\mathbb{F}_p$.

From the computation $\psi_i(e_j) = \phi_i(V e_j) = \sum_k a_{jk} \phi_i(e_k) = \sum_k b_{ik} \phi_k(e_j)$ we obtain ${}^t A = C^{-1} B C$. Now recall that over every field κ two square matrices with coefficients in κ which are conjugate over an algebraic closure of κ are conjugate over κ and, furthermore, that every square matrix with coefficient in κ is conjugate, over κ, to its transpose. Hence A is indeed conjugate to B over $\mathbb{F}_p$. $\qquad\square$

Remark A.3. Inspecting the proof we see that for G/A' connected with étale dual (not necessarily annihilated by p) we have a commutative diagram

$$\begin{array}{ccc}
G(\bar{K})^I & \xrightarrow[\simeq]{\ \Phi\ } & \mathrm{Hom}_{\mathbb{F}_p[V]}(M/FM, \bar{\mathbb{F}}_p) \\
\downarrow{\scriptstyle \tau} & & \downarrow{\scriptstyle \mathrm{Hom}(V, \bar{\mathbb{F}}_p)} \\
G(\bar{K})^I & \xrightarrow[\simeq]{\ \Phi\ } & \mathrm{Hom}_{\mathbb{F}_p[V]}(M/FM, \bar{\mathbb{F}}_p).
\end{array}$$

Proof of Proposition A.2. Define a finite Honda system over A' by

$$M := \mathbb{Z}/p^n\mathbb{Z}, \quad 1 \neq V \in 1 + p(\mathbb{Z}/p^n\mathbb{Z}) \subseteq (\mathbb{Z}/p^n\mathbb{Z})^* = \mathrm{Aut}_{\mathbb{Z}_p}(M),$$
$$F := pV^{-1},$$
$$L := M_{A'}.$$

It is easy to see that this is indeed a finite Honda system. For the corresponding group G/A' we have by Remark A.3

$$G(\overline{K})^I \simeq \mathrm{Hom}_{\mathbb{F}_p[V]}(M/FM, \overline{\mathbb{F}}_p) = \overline{\mathbb{F}}_p^{V=1} = \mathbb{F}_p$$

with trivial geometric Frobenius. Note that V is the identity on M/FM, but $V \neq 1$. $\qquad\square$

Acknowledgements

References

[Boston et al. 1991] N. Boston, H. W. Lenstra, Jr., and K. A. Ribet, "Quotients of group rings arising from two-dimensional representations", *C. R. Acad. Sci. Paris Sér. I Math.* **312**:4 (1991), 323–328. MR 92c:11057 Zbl 0718.16018

[Buzzard 2001] K. Buzzard, appendix to [Ribet and Stein 2001]. MR 2002h:11047 Zbl 01698970

[Conrad 1999] B. Conrad, "Finite group schemes over bases with low ramification", *Compositio Math.* **119**:3 (1999), 239–320. MR 2001c:14071 Zbl 0984.14015

[Edixhoven 1992] B. Edixhoven, "The weight in Serre's conjectures on modular forms", *Invent. Math.* **109**:3 (1992), 563–594. MR 93h:11124 Zbl 0777.11013

[Edixhoven 2006] B. Edixhoven, "Comparison of integral structures on spaces of modular forms of weight two, and computation of spaces of forms mod 2 of weight one", *J. Inst. Math. Jussieu* **5**:1 (2006), 1–34. MR 2007f:11046 Zbl 05009932

[Emerton 2002] M. Emerton, "Supersingular elliptic curves, theta series and weight two modular forms", *J. Amer. Math. Soc.* **15**:3 (2002), 671–714. MR 2003b:11038 Zbl 01739913

[Fontaine 1977] J.-M. Fontaine, *Groupes p-divisibles sur les corps locaux*, Société Mathématique de France, Paris, 1977. Astérisque, No. 47-48. MR 58 #16699 Zbl 0377.14009

[Gross 1990] B. H. Gross, "A tameness criterion for Galois representations associated to modular forms (mod p)", *Duke Math. J.* **61**:2 (1990), 445–517. MR 91i:11060 Zbl 0743.11030

[Kilford 2002] L. J. P. Kilford, "Some non-Gorenstein Hecke algebras attached to spaces of modular forms", *J. Number Theory* **97**:1 (2002), 157–164. MR 2003j:11046 Zbl 01891434

[Kilford and Wiese 2006] L. J. P. Kilford and G. Wiese, "On the failure of the Gorenstein property for Hecke algebras of prime weight", preprint, 2006. math.NT/0612317

[Ribet and Stein 2001] K. A. Ribet and W. A. Stein, "Lectures on Serre's conjectures", pp. 143–232 in *Arithmetic algebraic geometry* (Park City, UT, 1999), edited by B. Conrad and K. Rubin, IAS/Park City Math. Ser. **9**, Amer. Math. Soc., Providence, RI, 2001. MR 2002h:11047 Zbl 01698970

[Tilouine 1997] J. Tilouine, "Hecke algebras and the Gorenstein property", pp. 327–342 in *Modular forms and Fermat's last theorem* (Boston, 1995), edited by G. Cornell et al., Springer, New York, 1997. MR 1638483 Zbl 01098985

[Wiese 2007] G. Wiese, "On the faithfulness of parabolic cohomology as a Hecke module over a finite field", *J. reine angew. Math.* **606** (2007), 79–103.

gabor.wiese@uni-due.de   *Universität Duisburg-Essen, Institut für Experimentelle Mathematik, Ellernstraße 29, D-45326 Essen, Germany*
http://maths.pratum.net

niko.naumann@mathematik.uni-regensburg.de
NWF I - Mathematik, Universität Regensburg, D-93040 Regensburg, Germany
http://homepages.uni-regensburg.de/~nan25776/

Functional equations for Mahler measures of genus-one curves

Matilde N. Lalin and Mathew D. Rogers

In this paper we will establish functional equations for Mahler measures of families of genus-one two-variable polynomials. These families were previously studied by Beauville, and their Mahler measures were considered by Boyd, Rodriguez Villegas, Bertin, Zagier, and Stienstra. Our functional equations allow us to prove identities between Mahler measures that were conjectured by Boyd. As a corollary, we also establish some new transformations for hypergeometric functions.

1. History and introduction

The goal of this paper is to establish identities between the logarithmic Mahler measures of polynomials with zero varieties corresponding to genus-one curves. Recall that the logarithmic Mahler measure (which we shall henceforth simply refer to as the Mahler measure) of an n-variable Laurent polynomial $P(x_1, x_2, \ldots, x_n)$ is defined by

$$\mathrm{m}\left(P(x_1, \ldots, x_n)\right) = \int_0^1 \cdots \int_0^1 \log \left| P\left(\mathrm{e}^{2\pi \mathrm{i}\theta_1}, \ldots, \mathrm{e}^{2\pi \mathrm{i}\theta_n}\right) \right| \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_n.$$

Many difficult questions surround the special functions defined by Mahler measures of elliptic curves.

The first example of the Mahler measure of a genus-one curve was studied in [Boyd 1998; Deninger 1997]. Boyd found that

$$\mathrm{m}\left(1 + x + \frac{1}{x} + y + \frac{1}{y}\right) \stackrel{?}{=} L'(E, 0), \tag{1-1}$$

where E denotes the elliptic curve of conductor 15 that is the projective closure of $1 + x + 1/x + y + 1/y = 0$. As usual, $L(E, s)$ is its L-function, and the question mark above the equals sign indicates numerical equality verified up to 28 decimal places.

Deninger [1997] gave an interesting interpretation of this formula. He obtained the Mahler measure by evaluating the Bloch regulator of an element $\{x, y\}$ from a certain K-group. In other words, the Mahler measure is given by a value of an Eisenstein–Kronecker series. Therefore Bloch's and Beilinson's conjectures predict that

$$\text{m}\left(1 + x + \frac{1}{x} + y + \frac{1}{y}\right) = cL'(E, 0),$$

where c is some rational number. Let us add that, even if Beilinson's conjectures were known to be true, this would not suffice to prove equality (1-1), since we still would not know the height of the rational number c.

This picture applies to other situations as well. Boyd [1998] performed extensive numerical computations within the family of polynomials $k + x + 1/x + y + 1/y$, as well as within some other genus-one families. Boyd's numerical searches led him to conjecture identities such as

$$\text{m}\left(5 + x + \frac{1}{x} + y + \frac{1}{y}\right) \stackrel{?}{=} 6\text{m}\left(1 + x + \frac{1}{x} + y + \frac{1}{y}\right),$$

$$\text{m}\left(8 + x + \frac{1}{x} + y + \frac{1}{y}\right) \stackrel{?}{=} 4\text{m}\left(2 + x + \frac{1}{x} + y + \frac{1}{y}\right).$$

Boyd conjectured conditions predicting when formulas like (1-1) should exist for the Mahler measures of polynomials with integral coefficients. This was further studied by Rodriguez Villegas [1999], who interpreted these conditions in the context of Bloch's and Beilinson's conjectures. He also used modular forms to express the Mahler measures as Kronecker–Eisenstein series in more general cases. In turn, this allowed him to prove some equalities such as

$$\text{m}\left(4\sqrt{2} + x + \frac{1}{x} + y + \frac{1}{y}\right) = L'\left(E_{4\sqrt{2}}, 0\right), \tag{1-2}$$

$$\text{m}\left(3\sqrt{2} + x + \frac{1}{x} + y + \frac{1}{y}\right) = qL'\left(E_{3\sqrt{2}}, 0\right), \tag{1-3}$$

where q is a rational number that is (numerically) equal to $5/2$. The first equality can be proved using the fact that the corresponding elliptic curve has complex multiplication, and therefore the conjectures are known for this case due to Bloch [2000]. The second equality depends on the fact that one has the modular curve $X_0(24)$, and the conjectures then follow from a result of Beilinson.

Rodriguez Villegas [2002] subsequently used the relationship between Mahler measures and regulators to prove a conjecture of Boyd [1998]:

$$\text{m}(y^2 + 2xy + y - x^3 - 2x^2 - x) = \tfrac{5}{7}\text{m}(y^2 + 4xy + y - x^3 + x^2).$$

He proved this identity without actually expressing the Mahler measures in terms of L-series. Bertin [2004] has also proved similar identities using these ideas.

Although the conjecture in (1-1) remains open, we will in fact prove two of Boyd's other conjectures this paper.

Theorem 1.1. *Assume that $q = 5/2$ in* (1-3). *Then*

$$\mathrm{m}\left(2 + x + \frac{1}{x} + y + \frac{1}{y}\right) = L'\left(E_{3\sqrt{2}}, 0\right), \tag{1-4}$$

$$\mathrm{m}\left(8 + x + \frac{1}{x} + y + \frac{1}{y}\right) = 4L'\left(E_{3\sqrt{2}}, 0\right). \tag{1-5}$$

Our proof of this combines two interesting functional equations for the function

$$m(k) := \mathrm{m}\left(k + x + \frac{1}{x} + y + \frac{1}{y}\right).$$

Kurokawa and Ochiai [2005] recently proved the first functional equation, which says that, if $k \in \mathbb{R}\backslash\{0\}$,

$$m(4k^2) + m\left(\frac{4}{k^2}\right) = 2m\left(2\left(k + \frac{1}{k}\right)\right). \tag{1-6}$$

In Section 3 we use regulators to give a new proof of Equation (1-6). We will also prove a second functional equation in Section 2.1 using q-series. In particular, if k is nonzero and $|k| < 1$,

$$m\left(2\left(k + \frac{1}{k}\right)\right) + m\left(2\left(\mathrm{i}k + \frac{1}{\mathrm{i}k}\right)\right) = m\left(\frac{4}{k^2}\right). \tag{1-7}$$

Theorem 1.1 follows from setting $k = 1/\sqrt{2}$ in both identities, and then showing that $5m(\mathrm{i}\sqrt{2}) = 3m(3\sqrt{2})$. We have proved this final equality in Section 3.6.

This paper is divided into two sections of roughly equal length. In Section 2 we prove more identities like (1-7), which arise from expanding Mahler measures in q-series. In particular, we look at identities for four special functions defined by the Mahler measures of genus-one curves (see Equations (2-1) through (2-4) for notation). Equation (2-14) is undoubtedly the most important result in this part of the paper, since it implies that infinitely many identities like (1-7) exist. Sections 2.1 and 2.2 are mostly devoted to transforming special cases of (2-14) into interesting identities between the Mahler measures of rational polynomials. While the theorems in those subsections rely heavily on Ramanujan's theory of modular equations to alternative bases, we have attempted to maximize readability by eliminating q-series manipulation wherever possible. Finally, we have devoted Section 2.3 to proving some useful computational formulas. As a corollary we

establish several new transformations for hypergeometric functions, including

$$\sum_{n=0}^{\infty} \left(\frac{k(1-k)^2}{(1+k)^2} \right)^n \sum_{j=0}^{n} \binom{n}{j}^2 \binom{n+j}{j}$$

$$= \frac{(1+k)^2}{\sqrt{(1+k^2)\big((1-k-k^2)^2 - 5k^2\big)}}$$

$$\times {}_2F_1\left(\frac{1}{4}, \frac{3}{4}; 1; \frac{64k^5(1+k-k^2)}{(1+k^2)^2\big((1-k-k^2)^2-5k^2\big)^2} \right). \quad (1\text{-}8)$$

We have devoted Section 3 to further studying the relationship between Mahler measures and regulators. We show how to recover the Mahler measure q-series expansions and the Kronecker–Eisenstein series directly from Bloch's formula for the regulator. This in turn shows that the Mahler measure identities can be viewed as consequences of functional identities for the elliptic dilogarithm.

Many of the identities in this paper can be interpreted from both a regulator perspective and from a q-series perspective. The advantage of the q-series approach is that it simplifies the process of finding new identities. The fundamental result in Section 2, Equation (2-14), follows easily from the Mahler measure q-series expansions. Unfortunately the q-series approach does not provide an easy way to explain identities like (1-6). Unlike most of the other formulas in Section 2, Kurokawa's and Ochiai's result *does not* follow from (2-14). An advantage of the regulator approach, is that it enables us to construct proofs of both (1-6) and (1-7) from a unified perspective. Additionally, the regulator approach seem to provide the only way to prove the final step in Theorem 1.1, namely to show that $5m(i\sqrt{2}) = 3m(3\sqrt{2})$. Thus, a complete view of this subject matter should incorporate both regulator and q-series perspectives.

2. Mahler measures and q-series

We will consider four important functions defined by Mahler measures:

$$\mu(t) = m\left(\frac{4}{\sqrt{t}} + x + \frac{1}{x} + y + \frac{1}{y} \right), \quad (2\text{-}1)$$

$$n(t) = m\left(x^3 + y^3 + 1 - \frac{3}{t^{1/3}}xy \right), \quad (2\text{-}2)$$

$$g(t) = m\left((x+y)(x+1)(y+1) - \frac{1}{t}xy \right), \quad (2\text{-}3)$$

$$r(t) = m\left((x+y+1)(x+1)(y+1) - \frac{1}{t}xy \right). \quad (2\text{-}4)$$

Throughout Section 2 we will use the notation $\mu(t) = m(4/\sqrt{t})$ for convenience. Recall from [Rodriguez Villegas 1999] and [Stienstra 2006] that each of these functions has a simple q-series expansion when t is parameterized correctly. To summarize, if we let $(x; q)_\infty = (1 - x)(1 - xq)\left(1 - xq^2\right)\ldots$, and

$$M(q) = 16q \frac{(q; q)_\infty^8 \left(q^4; q^4\right)_\infty^{16}}{\left(q^2; q^2\right)_\infty^{24}}, \tag{2-5}$$

$$N(q) = \frac{27q \left(q^3; q^3\right)_\infty^{12}}{(q; q)_\infty^{12} + 27q \left(q^3; q^3\right)_\infty^{12}}, \tag{2-6}$$

$$G(q) = q^{1/3} \frac{(q; q^2)_\infty}{\left(q^3; q^6\right)_\infty^3}, \tag{2-7}$$

$$R(q) = q^{1/5} \frac{(q; q^5)_\infty \left(q^4; q^5\right)_\infty}{\left(q^2; q^5\right)_\infty \left(q^3; q^5\right)_\infty}, \tag{2-8}$$

then for $|q|$ sufficiently small,

$$\mu(M(q)) = -\operatorname{Re}\left(\frac{1}{2}\log(q) + 2\sum_{j=1}^\infty j\chi_{-4}(j)\log(1 - q^j)\right), \tag{2-9}$$

$$n(N(q)) = -\operatorname{Re}\left(\frac{1}{3}\log(q) + 3\sum_{j=1}^\infty j\chi_{-3}(j)\log(1 - q^j)\right), \tag{2-10}$$

$$g(G^3(q)) = -\operatorname{Re}\left(\log(q) + \sum_{j=1}^\infty (-1)^{j-1} j\chi_{-3}(j)\log(1 - q^j)\right), \tag{2-11}$$

$$r(R^5(q)) = -\operatorname{Re}\left(\log(q) + \sum_{j=1}^\infty j\operatorname{Re}\left((2 - \mathrm{i})\chi_r(j)\right)\log(1 - q^j)\right). \tag{2-12}$$

In particular, $\chi_{-3}(j)$ and $\chi_{-4}(j)$ are the usual Dirichlet characters, and $\chi_r(j)$ is the character of conductor five with $\chi_r(2) = \mathrm{i}$. We have used the notation $G(q)$ and $R(q)$, as opposed to something like $\tilde{G}(q) = G^3(q)$, in order to preserve Ramanujan's notation. As usual, $G(q)$ corresponds to Ramanujan's cubic continued fraction, and $R(q)$ corresponds to the Rogers–Ramanujan continued fraction [Andrews and Berndt 2005].

The first important application of the q-series expansions is that they can be used to calculate the Mahler measures numerically. For example, we can calculate $\mu(1/10)$ with Equation (2-9), provided that we can first determine a value of q for which $M(q) = 1/10$. Fortunately, the theory of elliptic functions shows that if

$\alpha = M(q)$, then

$$q = \exp\left(-\pi \frac{{}_2F_1\left(\frac{1}{2}, \frac{1}{2}; 1; 1 - \alpha\right)}{{}_2F_1\left(\frac{1}{2}, \frac{1}{2}; 1; \alpha\right)}\right). \tag{2-13}$$

Using Equation (2-13) we easily compute $q = .01975\ldots$, and it follows that $\mu(1/10) = 2.524718\ldots$ The function defined in Equation (2-13) is called the *elliptic nome*, and is sometimes denoted by $q_2(\alpha)$. Theorem 2.6 provides similarly explicit inversion formulas for Equations (2-5) through (2-8).

The second, and perhaps more significant fact that follows from these q-series, is that linear dependencies exist between the Mahler measures. In particular, if

$$f(q) \in \left\{\mu(M(q)), n(N(q)), g(G^3(q)), r(R^5(q))\right\},$$

then for an appropriate prime p

$$\sum_{j=0}^{p-1} f\left(e^{2\pi i j/p} q\right) = (1 + p^2 \chi(p)) f(q^p) - p\chi(p) f(q^{p^2}), \tag{2-14}$$

where $\chi(j)$ is the character from the relevant q-series. The prime p satisfies the restriction that $p \neq 2$ when $f(q) = g(G^3(q))$, and $p \not\equiv 2, 3 \pmod{5}$ when $f(q) = r\left(R^5(q)\right)$. The astute reader will immediately recognize that (2-14) is essentially a Hecke eigenvalue equation. A careful analysis of the exceptional case that occurs when $p = 2$ and $f(q) = g(G^3(q))$ leads to the important and surprising inverse relation:

$$\begin{aligned}
3n(N(q)) &= g(G^3(q)) - 8g(G^3(-q)) + 4g(G^3(q^2)), \\
3g(G^3(q)) &= n(N(q)) + 4n(N(q^2)).
\end{aligned} \tag{2-15}$$

In the next two subsections we discuss methods for transforming (2-14) and (2-15) into so-called functional equations.

2.1. *Functional equations from modular equations.* Since the primary goal of this paper is to find relations between the Mahler measures of *rational* (or at least algebraic) polynomials, we will require modular equations to simplify our results. For example, consider (2-14) when $f(q) = \mu(M(q))$ and $p = 2$:

$$\mu(M(q)) + \mu(M(-q)) = \mu(M(q^2)). \tag{2-16}$$

For our purposes, Equation (2-16) is only interesting if $M(q)$, $M(-q)$, and $M(q^2)$ are all simultaneously algebraic. Fortunately, it turns out that $M(q)$ and $M(q^2)$ (hence also $M(-q)$ and $M(q^2)$) satisfy a well known polynomial relation.

Definition 2.1. Suppose that $F(q) \in \{M(q), N(q), G(q), R(q)\}$. An n–th degree modular equation is an algebraic relation between $F(q)$ and $F(q^n)$.

We will not need to derive any new modular equations in this paper. Berndt proved virtually all of the necessary modular equations while editing Ramanujan's notebooks; see [Andrews and Berndt 2005; Berndt 1989; 1991; 1998]. Ramanujan seems to have arrived at most of his modular equations through complicated q-series manipulations (of course this is speculation since he did not write down any proofs!). Modular equations involving $M(q)$ correspond to the classical modular equations [Berndt 1991], relations for $N(q)$ correspond to Ramanujan's signature three modular equations [Berndt 1998], and most of the known modular equations for $G(q)$ and $R(q)$ appear in [Andrews and Berndt 2005].

Now we can finish simplifying Equation (2-16). Since the classical second-degree modular equation shows that whenever $|q| < 1$,

$$\frac{4M(q^2)}{\left(1+M(q^2)\right)^2} = \left(\frac{M(q)}{M(q)-2}\right)^2,$$

we easily obtain the parameterizations:

$$M(q) = \frac{4k^2}{(1+k^2)^2}, \quad M(-q) = \frac{-4k^2}{(1-k^2)^2}, \quad \text{and} \quad M(q^2) = k^4.$$

Substituting these parametric formulas into Equation (2-16) yields:

Theorem 2.2. *The following identity holds whenever $|k| < 1$:*

$$\mathrm{m}\left(\frac{4}{k^2} + x + \frac{1}{x} + y + \frac{1}{y}\right) = \mathrm{m}\left(2\left(k+\frac{1}{k}\right) + x + \frac{1}{x} + y + \frac{1}{y}\right)$$
$$+ \mathrm{m}\left(2\mathrm{i}\left(k-\frac{1}{k}\right) + x + \frac{1}{x} + y + \frac{1}{y}\right).$$

We need to make a few remarks about working with modular equations before proving the main theorem in this section. Suppose that for some algebraic function $P(X, Y)$:

$$P(F(q), F(q^p)) = 0,$$

where $F(q) \in \{M(q), N(q), G(q), R(q)\}$. By an elementary change of variables $q \to \mathrm{e}^{2\pi \mathrm{i} j/p}q$, it follows that $P(F(\mathrm{e}^{2\pi \mathrm{i} j/p}q), F(q^p)) = 0$ for every $j \in \{0, 1, \dots, p-1\}$. If $P(X, Y)$ is symmetric in X and Y, it also follows that $P(F(q^{p^2}), F(q^p))$ vanishes. Therefore, if $P(X, Y)$ is sufficiently simple (for example a symmetric genus-zero polynomial), we can find simultaneous parameterizations for $F(q^p)$, $F(q^{p^2})$, and $F(\mathrm{e}^{2\pi \mathrm{i} j/p}q)$ for all j. In such an instance, (2-14) reduces to an interesting functional equation for one of the four Mahler measures $\mu(t), n(t), g(t), r(t)$. Five basic functional equations follow from applying these ideas to (2-14).

Theorem 2.3. *For* $|k| < 1$ *and* $k \neq 0$, *we have*

$$\mu\left(\frac{4k^2}{(1+k^2)^2}\right) + \mu\left(\frac{-4k^2}{(1-k^2)^2}\right) = \mu(k^4). \tag{2-17}$$

The following identities hold for $|u|$ *sufficiently small but nonzero:*

$$n\left(\frac{27u(1+u)^4}{2(1+4u+u^2)^3}\right) + n\left(-\frac{27u(1+u)}{2(1-2u-2u^2)^3}\right)$$
$$= 2n\left(\frac{27u^4(1+u)}{2(2+2u-u^2)^3}\right) - 3n\left(\frac{27u^2(1+u)^2}{4(1+u+u^2)^3}\right). \tag{2-18}$$

If $\zeta_3 = e^{2\pi i/3}$ *and* $Y(t) = 1 - \left(\dfrac{1-t}{1+2t}\right)^3$, *then*

$$n(u^3) = \sum_{j=0}^{2} n(Y(\zeta_3^j u)). \tag{2-19}$$

If $\zeta_3 = e^{2\pi i/3}$ *and* $Y(t) = t\left(\dfrac{1-t+t^2}{1+2t+4t^2}\right)$, *then*

$$g(u^3) = \sum_{j=0}^{2} g(Y(\zeta_3^j u)). \tag{2-20}$$

If $\zeta_5 = e^{2\pi i/5}$ *and* $Y(t) = t\left(\dfrac{1-2t+4t^2-3t^3+t^4}{1+3t+4t^2+2t^3+t^4}\right)$, *then*

$$r(u^5) = \sum_{j=0}^{4} r(Y(\zeta_5^j u)). \tag{2-21}$$

Proof. We have already sketched a proof of (2-17) in the discussion preceding Theorem 2.2.

Proving (2-18) requires the second-degree modular equation from Ramanujan's theory of signature 3. If $\beta = N(q^2)$ and α is either $N(q)$, $N(-q)$, or $N(q^4)$, then

$$27\alpha\beta(1-\alpha)(1-\beta) - (\alpha+\beta-2\alpha\beta)^3 = 0. \tag{2-22}$$

If we choose u so that $N(q^2) = 27u^2(1+u)^2/(4(1+u+u^2)^3)$, we can use (2-22) to verify easily that

$$N(q) = \frac{27u(1+u)^4}{2(1+4u+u^2)^3}, \qquad N(-q) = -\frac{27u(1+u)}{2(1-2u-2u^2)^3},$$
$$N(q^4) = \frac{27u^4(1+u)}{2(2+2u-u^2)^3}.$$

The proof of (2-18) follows from applying these parameterizations to (2-14) when $f(q) = n(N(q))$, and $p = 2$.

The proof of (2-19) requires Ramanujan's third-degree, signature 3 modular equation. In particular, if $\alpha = N(q)$ and $\beta = N(q^3)$, then

$$\alpha = 1 - \left(\frac{1-\beta^{1/3}}{1+2\beta^{1/3}}\right)^3 = Y\left(\beta^{1/3}\right). \tag{2-23}$$

Since $N^{1/3}(q^3) = q \times \{\text{power series in } q^3\}$, a short computation shows that for all $j \in \{0, 1, 2\}$, we have $N(\zeta_3^j q) = Y(\zeta_3^j N^{1/3}(q^3))$. Choosing u such that $N(q^3) = u^3$, we must have $N(\zeta_3^j q) = Y(\zeta_3^j u)$. Equation (2-19) follows from applying these parametric formulas to (2-14) when $f(q) = n(N(q))$, and $p = 3$.

Since the proofs of Equations (2-20) and (2-21) rely on similar arguments to the proof of (2-19), we will simply state the prerequisite modular equations. In particular, (2-20) follows from Ramanujan's third-degree modular equation for the cubic continued fraction. If $\alpha = G(q)$ and $\beta = G(q^3)$, then

$$\alpha^3 = \beta\left(\frac{1-\beta+\beta^2}{1+2\beta+4\beta^2}\right). \tag{2-24}$$

Similarly, (2-21) follows from the fifth-degree modular equation for the Rogers–Ramanujan continued fraction. In particular, if $\alpha = R(q)$ and $\beta = R\left(q^5\right)$,

$$\alpha^5 = \beta\left(\frac{1-2\beta+4\beta^2-3\beta^3+\beta^4}{1+3\beta+4\beta^2+2\beta^3+\beta^4}\right). \tag{2-25}$$

$\square$

The functional equations in Theorem 2.3 only hold in restricted subsets of $\mathbb{C}$. To explain this phenomenon we will go back to (2-14). As a general rule, we have to restrict q to values for which *none* of the Mahler measure integrals in (2-14) vanish on the unit torus. In other words, we can only consider the set of q's for which each term in (2-14) can be calculated from the appropriate q-series. Next, we may need to further restrict the domain of q depending on where the relevant parametric formulas hold. For example, parameterizations such as $N(q) = 27u(1+u)^4/(2(1+4u+u^2)^3)$ and $N(q^2) = 27u^2(1+u)^2/(4(1+u+u^2)^3)$ hold for $|q|$ sufficiently small, but fail when q is close to 1. After determining the domain of q, we can calculate the domain of u by solving a parametric equation to express u in terms of a q-series.

Theorem 2.4. *For $|p|$ sufficiently small but nonzero,*

$$3g(p) = n\left(\frac{27p}{(1+4p)^3}\right) + 4n\left(\frac{27p^2}{(1-2p)^3}\right). \tag{2-26}$$

Furthermore, for $|u|$ sufficiently small but nonzero,

$$3n\left(\frac{27u(1+u)^4}{2(1+4u+u^2)^3}\right)$$

$$= g\left(\frac{u}{2(1+u)^2}\right) - 8g\left(-\frac{u(1+u)}{2}\right) + 4g\left(\frac{u^2}{4(1+u)}\right). \quad (2\text{-}27)$$

Proof. We will prove (2-27) first. Recall that (2-15) shows that

$$3n(N(q)) = g(G^3(q)) - 8g\left(G^3(-q)\right) + 4g(G^3(q^2)).$$

Suppose that $q = q_2(u(2+u)^3/(1+2u)^3)$, where $q_2(\alpha)$ is the elliptic nome. Classical eta function inversion formulas (which we omit) show that for $|u|$ sufficiently small: $G^3(q) = u/(2(1+u)^2)$, $G^3(-q) = -u(1+u)/2$, $G^3(q^2) = u^2/(4(1+u))$, $N(q) = 27u(1+u)^4/(2(1+4u+u^2)^3)$, and $N(q^2) = 27u^2(1+u)^2/(4(1+u+u^2)^3)$.

To prove (2-26) first recall that

$$3g(G^3(q)) = n(N(q)) + 4n\left(N(q^2)\right).$$

If we let $p = u/(2(1+u)^2)$, then it follows that $G^3(q) = p$, $N(q) = 27p/(1+4p)^3$, and $N(q^2) = 27p^2/(1+2p)^3$. □

Theorem 2.4 shows that $g(t)$ and $n(t)$ are essentially interchangeable. In Section 2.3 we will use (2-26) to derive an extremely useful formula for calculating $g(t)$ numerically.

2.2. *Identities arising from higher modular equations.* The functional equations presented in Section 2.1 are not the only interesting formulas that follow from (2-14). Rather those results represent the subset of functional equations in which every Mahler measure depends on a rational argument (possibly in a cyclotomic field). If we consider the higher modular equations, then we can establish formulas involving the Mahler measures of the modular polynomials themselves. Equation (2-31) is the simplest formula in this class of results.

Consider (2-14) when $p = 3$ and $f(q) = \mu(M(q))$:

$$\sum_{j=0}^{2} \mu\left(M(\zeta_3^j q)\right) = -8\mu\left(M(q^3)\right) + 3\mu\left(M(q^9)\right). \quad (2\text{-}28)$$

By the third-degree modular equation, if $\alpha \in \left\{M(q), M\left(\zeta_3 q\right), M(\zeta_3^2 q), M(q^9)\right\}$ and $\beta = M(q^3)$, then

$$G_3(\alpha, \beta) := (\alpha^2 + \beta^2 + 6\alpha\beta)^2 - 16\alpha\beta\left(4(1+\alpha\beta) - 3(\alpha + \beta)\right)^2 = 0. \quad (2\text{-}29)$$

Since $G_3(\alpha, \beta) = 0$ defines a curve with genus greater than zero, it is impossible to find simultaneous rational parameterizations for all four zeros in α. For example, if

we let $\beta = M(q^3) = p(2+p)^3/(1+2p)^3$, then we can obtain the rational expression $M(q^9) = p^3(2+p)/(1+2p)$, and three messy formulas involving radicals for the other zeros. Despite this difficulty, Equation (2-28) still reduces to an interesting formula if we recall the factorization

$$G_3\left(\alpha, M(q^3)\right) = \left(\alpha - M(q^9)\right) \prod_{j=0}^{2}\left(\alpha - M(\zeta_3^j q)\right), \qquad (2\text{-}30)$$

and then use the fact that Mahler measure satisfies $m(P) + m(Q) = m(PQ)$.

Theorem 2.5. *If* $G_3(\alpha, \beta)$ *is as defined in* (2-29), *then for* $|p|$ *sufficiently small but nonzero,*

$$m\left(G_3\left(\frac{(x+x^{-1})^2(y+y^{-1})^2}{16}, \frac{1}{p}\left(\frac{1+2p}{2+p}\right)^3\right)\right)$$
$$= -16\log(2) - 16\mu\left(p\left(\frac{2+p}{1+2p}\right)^3\right) + 8\mu\left(p^3\left(\frac{2+p}{1+2p}\right)\right). \quad (2\text{-}31)$$

Proof. First notice that from the elementary properties of Mahler's measure

$$\mu(t) = \frac{1}{2}m\left(\frac{16}{(x+x^{-1})^2(y+y^{-1})^2} - t\right) - \frac{1}{2}\log|t|.$$

Applying this identity to (2-28) and appealing to (2-30) yields

$$m\left(G_3\left(\frac{16}{(x+x^{-1})^2(y+y^{-1})^2}, M(q^3)\right)\right)$$
$$= \log\left|M(q)M(\zeta_3 q)M(\zeta_3^2 q)M(q^9)\right| - 16\mu(M(q^3)) + 8\mu(M(q^9)).$$

Elementary q-product manipulations show that

$$M^4(q^3) = M(q)M(\zeta_3 q)M(\zeta_3^2 q)M(q^9),$$

and since $\alpha^4\beta^4 G_3(1/\alpha, 1/\beta) = G_3(\alpha, \beta)$, we obtain

$$m\left(G_3\left(\frac{(x+x^{-1})^2(y+y^{-1})^2}{16}, \frac{1}{M(q^3)}\right)\right) = -16\log 2 - 16\mu(M(q^3)) + 8\mu(M(q^9)).$$

Finally, if we choose p so that $M(q^3) = p((2+p)/(1+2p))^3$, then $M(q^9) = p^3((2+p)/(1+2p))$, and the theorem follows. $\qquad\square$

Although we completely eliminated the q-series expressions from (2-31), this is not necessarily desirable (or even possible) in more complicated examples. Take

the identity involving resultants which follows from (2-14) (and some manipulation) when $p = 11$ and $f(q) = r\left(R^5(q)\right)$:

$$m\left(\operatorname*{Res}_{z}\left(z^5 - \frac{xy}{(x+1)(y+1)(x+y+1)}, P\left(z, R^5(q)\right)\right)\right)$$
$$= -12m(1+x+y) + 12\log|R^5(q)| + 122r(R^5(q)) - 11r(R^5(q^{11})). \quad (2\text{-}32)$$

In this formula $P(u, v)$ is the polynomial

$$P(u, v) = uv(1 - 11v^5 - v^{10})(1 - 11u^5 - u^{10}) - (u - v)^{12},$$

which also satisfies $P(R(q), R(q^{11})) = 0$ [Rogers 1920]. Even if rational parameterizations existed for $R(q)$ and $R(q^{11})$, substituting such formulas into (2-32) would probably just make the identity prohibitively complicated.

2.3. Computationally useful formulas and a few related hypergeometric transformations. While many methods exist for numerically calculating each of the four Mahler measures $\{\mu(t), n(t), g(t), r(t)\}$, two simple and efficient methods are directly related to the material discussed so far.

   The first computational method relies on the q-series expansions. For example, we can calculate $\mu(\alpha)$ with Equation (2-9), provided that a value of q exists for which $M(q) = \alpha$. Amazingly, the elliptic nome function, defined in Equation (2-13), furnishes a value of q whenever $|\alpha| < 1$. Similar inversion formulas exist for all of the q-products in Equations (2-5) through (2-8). Suppose that for $j \in \{2, 3, 4, 6\}$

$$q_j(\alpha) = \exp\left(-\frac{\pi}{\sin(\pi/j)} \frac{{}_2F_1(1/j, 1 - 1/j; 1; 1 - \alpha)}{{}_2F_1(1/j, 1 - 1/j; 1; \alpha)}\right), \quad (2\text{-}33)$$

then we have the following theorem:

Theorem 2.6. *With α and q appropriately restricted, the following table gives inversion formulas for Equations (2-5) through (2-8):*

α	q	α	q
$M(q)$	$q_2(\alpha)$	$G(q)$	$q_2\left(\dfrac{u(2+u)^3}{(1+2u)^3}\right)$ with $\alpha^3 = \dfrac{u}{2(1+u)^2}$
$N(q)$	$q_3(\alpha)$	$R(q)$	$q_4\left(\dfrac{64k(1+k-k^2)^5}{(1+k^2)^2((1+11k-k^2)^2-125k^2)^2}\right)$ with $\alpha^5 = \dfrac{k(1-k)^2}{(1+k)^2}$

For example, if $|q| < 1$ and $\alpha = M(q)$, then $q = q_2(\alpha)$.

Proof. The inversion formulas for $M(q)$ and $G(q)$ follow from classical eta function identities, and the inversion formula for $N(q)$ follows from eta function identities in Ramanujan's theory of signature three.

The inversion formula for $R(q)$ seems to be new, so we will prove it. Let us suppose that $\alpha = R(q)$ and $k = R(q)R^2(q^2)$, where q is fixed. A formula of Ramanujan [Andrews and Berndt 2005] shows that $\alpha^5 = k(1-k)^2/(1+k)^2$, which establishes the second part of the formula. Now suppose that $q = q_2(\alpha_2)$, where $\alpha_2 = M(q)$. A classical identity shows that

$$q\,(-q; q)_\infty^{24} = \frac{\alpha_2}{16(1 - \alpha_2)^2},$$

and comparing this to Ramanujan's identity

$$q\,(-q; q)_\infty^{24} = \left(\frac{k}{1 - k^2}\right)\left(\frac{1 + k - k^2}{1 - 4k - k^2}\right)^5,$$

we deduce that

$$\frac{\alpha_2}{(1 - \alpha_2)^2} = 16\left(\frac{k}{1 - k^2}\right)\left(\frac{1 + k - k^2}{1 - 4k - k^2}\right)^5. \tag{2-34}$$

Now recall that the theory of the signature 4 elliptic nome shows that

$$q = q_2(\alpha_2) = q_4\left(\frac{4\alpha_2}{(1 + \alpha_2)^2}\right) = q_4\left(\frac{4\alpha_2/(1 - \alpha_2)^2}{1 + 4\alpha_2/(1 - \alpha_2)^2}\right).$$

Substituting (2-34) into this final result yields

$$q = q_4\left(\frac{64k(1 + k - k^2)^5}{(1 + k^2)^2\big((1 + 11k - k^2)^2 - 125k^2\big)^2}\right),$$

which completes the proof. $\qquad\qquad\square$

The second method for calculating the four Mahler measures, $\mu(t)$, $n(t)$, $g(t)$, and $r(t)$ depends on reformulating them in terms of hypergeometric functions. For example, Rodriguez Villegas [1999] proved the formula

$$\mu(t) = -\frac{1}{2}\,\mathrm{Re}\left(\log(t/16) + \int_0^t \frac{{}_2F_1\left(\frac{1}{2}, \frac{1}{2}; 1; u\right) - 1}{u}\,du\right).$$

Translated into the language of generalized hypergeometric functions, this becomes

$$\mu(t) = -\,\mathrm{Re}\left(\frac{t}{8}{}_4F_3\left(\begin{smallmatrix}\frac{3}{2},\frac{3}{2},1,1\\2,2,2\end{smallmatrix}; t\right) + \frac{1}{2}\log(t/16)\right). \tag{2-35}$$

He also proved a formula for $n(t)$ which is equivalent to

$$n(t) = -\operatorname{Re}\left(\frac{2t}{27}{}_4F_3\left({}^{\frac{4}{3},\frac{5}{3},1,1}_{\quad 2,2,2}; t\right) + \frac{1}{3}\log(t/27)\right). \tag{2-36}$$

Formulas like (2-35) and (2-36) hold obvious appeal. From a computational perspective they are useful because most mathematics programs have routines for calculating generalized hypergeometric functions. For example, when $|t| < 1$ the Taylor series for the ${}_4F_3$ function easily gives better numerical accuracy than the Mahler measure integrals. Combining Equation (2-36) with (2-26) also yields a useful formula for calculating $g(t)$ whenever $|t|$ is sufficiently small:

$$\begin{aligned}
g(t) = -\operatorname{Re}&\left(\frac{2t}{(1+4t)^3}{}_4F_3\left({}^{\frac{4}{3},\frac{5}{3},1,1}_{\quad 2,2,2}; \frac{27t}{(1+4t)^3}\right)\right.\\
&\left.+ \frac{8t^2}{(1-2t)^3}{}_4F_3\left({}^{\frac{4}{3},\frac{5}{3},1,1}_{\quad 2,2,2}; \frac{27t^2}{(1-2t)^3}\right) + \log\left(\frac{t^3}{(1+4t)(1-2t)^4}\right)\right). \tag{2-37}
\end{aligned}$$

So far we have been unable to find a similar expression for $r(t)$.

Open Problem 1. Express $r(t)$ in terms of generalized hypergeometric functions.

Besides their computational importance, identities like (2-35) allow for a reformulation of Boyd's conjectures in the language of hypergeometric functions. For example, the conjecture

$$\mathrm{m}\left(1 + x + \frac{1}{x} + y + \frac{1}{y}\right) \stackrel{?}{=} L'(E, 0),$$

where E is an elliptic curve with conductor 15, becomes

$$L'(E, 0) \stackrel{?}{=} -2\operatorname{Re}\left({}_4F_3\left({}^{\frac{3}{2},\frac{3}{2},1,1}_{\quad 2,2,2}; 16\right)\right).$$

A proof of this identity would represent an important addition to the vast literature concerning transformations and evaluations of generalized hypergeometric functions.

In the remainder of this section we will apply our results to deduce a few interesting hypergeometric transformations. For example, differentiating (2-37) leads to an interesting corollary:

Corollary 2.7. *For $|t|$ sufficiently small,*

$$\omega(t) := \sum_{n=0}^{\infty} t^n \sum_{k=0}^{n} \binom{n}{k}^3 = \frac{1}{1-2t}{}_2F_1\left(\frac{1}{3}, \frac{2}{3}; 1; \frac{27t^2}{(1-2t)^3}\right), \tag{2-38}$$

and furthermore

$$\omega\left(\frac{p}{2(1+p)^2}\right) = (1+p)\omega\left(\frac{p^2}{4(1+p)}\right), \qquad (2\text{-}39)$$

whenever $|p|$ is sufficiently small.

Proof. We can prove (2-38) by differentiating each side of (2-37), and then by appealing to Stienstra's formulas [2006]. A second possible proof follows from showing that both sides of (2-38) satisfy the same differential equation.

The shortest proof of (2-39) follows from a formula of Zagier [Stienstra 2006]:

$$\omega(G^3(q)) = \prod_{n=0}^{\infty} \frac{(1-q^{2n})(1-q^{3n})^6}{(1-q^n)^2(1-q^{6n})^3}.$$

First use Zagier's identity to verify that $G^2(q)\omega(G^3(q)) = G(q^2)\omega(G^3(q^2))$, and then apply the parameterizations for $G^3(q)$ and $G^3(q^2)$ from 2.4. $\qquad\square$

We will also make a few remarks about the derivative of $r(t)$. Stienstra has shown that

$$r(t) = -\operatorname{Re}\left(\log t + \int_0^t \frac{\phi(u)-1}{u}\,du\right), \qquad (2\text{-}40)$$

where $\phi(t)$ is defined by

$$\phi(t) = \sum_{n=0}^{\infty} t^n \sum_{k=0}^{n} \binom{n}{k}^2 \binom{n+k}{k}. \qquad (2\text{-}41)$$

Even though we have not discovered a formula for $r(t)$ involving hypergeometric functions, we can still express $\phi(t)$ in terms of the hypergeometric function.

Theorem 2.8. *Let $\phi(t)$ be defined by (2-41). For $|k|$ sufficiently small,*

$$\phi\left(k\left(\frac{1-k}{1+k}\right)^2\right) = \frac{(1+k)^2}{\sqrt{(1+k^2)\left((1-k-k^2)^2 - 5k^2\right)}}$$

$$\times {}_2F_1\left(\frac{1}{4},\frac{3}{4};1;\frac{64k^5(1+k-k^2)}{(1+k^2)^2\left((1-k-k^2)^2-5k^2\right)^2}\right), \quad (2\text{-}42)$$

$$\phi\left(k^2\left(\frac{1+k}{1-k}\right)\right) = \frac{(1-k)}{\sqrt{(1+k^2)\left((1+11k-k^2)^2 - 125k^2\right)}}$$

$$\times {}_2F_1\left(\frac{1}{4},\frac{3}{4};1;\frac{64k(1+k-k^2)^5}{(1+k^2)^2\left((1+11k-k^2)^2-125k^2\right)^2}\right). \quad (2\text{-}43)$$

Furthermore, $\phi(t)$ *satisfies the functional equation*

$$\phi\left(k^2\left(\frac{1+k}{1-k}\right)\right) = \frac{1-k}{(1+k)^2}\phi\left(k\left(\frac{1-k}{1+k}\right)^2\right). \tag{2-44}$$

Proof. We prove (2-44) first. A result from [Verrill 2001] shows that

$$\phi^2\left(R^5(q)\right) = \frac{q}{R^5(q)}\frac{(q^5;q^5)_\infty^5}{(q;q)_\infty}. \tag{2-45}$$

Combining (2-45) with the trivial formula $(q^2, q^2)_\infty = (q;q)_\infty(-q;q)_\infty$, we get

$$\frac{\phi^2\left(R^5(q)\right)}{\phi^2\left(R^5(q^2)\right)} = \frac{R^5(q^2)}{R^5(q)}\frac{\{q^{1/24}\left(-q;q\right)_\infty\}}{\{q^{5/24}\left(-q^5;q^5\right)_\infty\}^5}. \tag{2-46}$$

We will apply four of Ramanujan's formulas to finish the proof. If $k = R(q)R^2(q^2)$, we have for $|q|$ sufficiently small (see [Andrews and Berndt 2005])

$$R^5(q) = k\left(\frac{1-k}{1+k}\right)^2, \tag{2-47}$$

$$R^5(q^2) = k^2\left(\frac{1+k}{1-k}\right), \tag{2-48}$$

$$q^{1/24}\left(-q;q\right)_\infty = \left(\frac{k}{1-k^2}\right)^{1/24}\left(\frac{1+k-k^2}{1-4k-k^2}\right)^{5/24}, \tag{2-49}$$

$$q^{5/24}\left(-q^5;q^5\right)_\infty = \left(\frac{k}{1-k^2}\right)^{5/24}\left(\frac{1+k-k^2}{1-4k-k^2}\right)^{1/24}. \tag{2-50}$$

Equation (2-44) follows immediately from substituting these parametric formulas into (2-46).

Next we prove (2-42). Combining Equation (2-47) with Entry 3.2.15 in [Andrews and Berndt 2005], we easily obtain

$$q^{5/24}\left(q^5;q^5\right)_\infty = \left(\frac{k(1-k^2)^2}{(1+k-k^2)\left(1-4k-k^2\right)^2}\right)^{1/6}q^{1/24}\left(q;q\right)_\infty. \tag{2-51}$$

Now we evaluate the eta product $q^{1/24}\left(q;q\right)_\infty$. Recall that if $q = q_4(z)$, then

$$q^{1/24}(q;q)_\infty = 2^{-1/4}z^{1/24}(1-z)^{1/12}\sqrt{{}_2F_1\left(\tfrac{1}{4},\tfrac{3}{4};1;z\right)}.$$

In Theorem 2.6 we showed that if $k = R(q)R^2(q^2)$ then

$$q = q_4\left(\frac{64k(1+k-k^2)^5}{(1+k^2)^2\left((1+11k-k^2)^2-125k^2\right)^2}\right);$$

hence

$$q^{1/24}(q;q)_\infty = \left(\frac{k(1-k^2)^2(1+k-k^2)^5(1-4k-k^2)^{10}}{(1+k^2)^6\left((1+11k-k^2)^2-125k^2\right)^6}\right)^{1/24}$$

$$\times \sqrt{{}_2F_1\left(\frac{1}{4},\frac{3}{4};1;\frac{64k(1+k-k^2)^5}{(1+k^2)^2\left((1+11k-k^2)^2-125k^2\right)^2}\right)}. \quad (2\text{-}52)$$

Substituting (2-52), (2-51), and (2-47) into (2-45) completes the proof of (2-42). The proof of (2-43) also follows from an extremely similar argument. $\quad\square$

We conclude this section by recording a few formulas which do not appear in [Andrews and Berndt 2005], but which were probably known to Ramanujan. We point out that Maier obtained several results along these lines in [Maier 2006]. The functional equation for $\phi(t)$ (after substituting $z = k/(1-k^2)$) implies a new hypergeometric transformation:

$$\sqrt{\frac{(1+11z)^2-125z^2}{(1-z)^2-5z^2}}\,{}_2F_1\left(\frac{1}{4},\frac{3}{4};1,\frac{64z^5(1+z)}{(1+4z^2)((1-z)^2-5z^2)^2}\right)$$

$$= {}_2F_1\left(\frac{1}{4},\frac{3}{4};1;\frac{64z(1+z)^5}{(1+4z^2)\left((1+11z)^2-125z^2\right)^2}\right). \quad (2\text{-}53)$$

Perhaps not surprisingly, we can also use the arguments in this section to deduce that

$$q_4^5\left(\frac{64z(1+z)^5}{(1+4z^2)\left((1+11z)^2-125z^2\right)^2}\right) = q_4\left(\frac{64z^5(1+z)}{(1+4z^2)\left((1-z)^2-5z^2\right)^2}\right), \quad (2\text{-}54)$$

which implies a rational parametrization for the fifth-degree modular equation in Ramanujan's theory of signature 4.

3. A regulator explanation

Now we will reinterpret our identities in terms of the regulators of elliptic curves. The elliptic curves in question are defined by the zero varieties of the polynomials whose Mahler measure we studied. First we explain the relationship between Mahler measures and regulators. Then we use regulators to deduce formulas involving Kronecker–Eisenstein series, including Equations (2-9), (2-10), (2-11), and (2-12).

We will follow some of the ideas from [Rodriguez Villegas 2002].

3.1. *The elliptic regulator.* Let F be a field. By Matsumoto's Theorem, $K_2(F)$ is generated by the symbols $\{a, b\}$ for $a, b \in F^*$, which satisfy the bilinearity relations

$\{a_1 a_2, b\} = \{a_1, b\}\{a_2, b\}$ and $\{a, b_1 b_2\} = \{a, b_1\}\{a, b_2\}$, and the Steinberg relation $\{a, 1-a\} = 1$.

Recall that for a field F, with discrete valuation v, and maximal ideal $\mathcal{M}$, the tame symbol is given by

$$(x, y)_v \equiv (-1)^{v(x)v(y)} \frac{x^{v(y)}}{y^{v(x)}} \mod \mathcal{M}$$

(see [Rodriguez Villegas 1999]). Note that this symbol is trivial if $v(x) = v(y) = 0$. In the case when $F = \mathbb{Q}(E)$ (from now on E denotes an elliptic curve), a valuation is determined by the order of the rational functions at each point $S \in E(\overline{\mathbb{Q}})$. We will denote the valuation determined by a point $S \in E(\overline{\mathbb{Q}})$ by v_S.

The tame symbol is then a map $K_2(\mathbb{Q}(E)) \to \mathbb{Q}(S)^*$.

We have

$$0 \to K_2(E) \otimes \mathbb{Q} \to K_2(\mathbb{Q}(E)) \otimes \mathbb{Q} \to \coprod_{S \in E(\overline{\mathbb{Q}})} \mathbb{Q}Q(S)^* \times \mathbb{Q},$$

where the last arrow corresponds to the coproduct of tame symbols.

Therefore an element $\{x, y\} \in K_2(\mathbb{Q}(E)) \otimes \mathbb{Q}$ can be seen as an element in $K_2(E) \otimes \mathbb{Q}$ whenever $(x, y)_{v_S} = 1$ for all $S \in E(\overline{\mathbb{Q}})$. All of the families considered in this paper are tempered according to [Rodriguez Villegas 1999], and therefore they satisfy the triviality of tame symbols.

The regulator map (defined by Beilinson, after work of Bloch) is given by

$$\mathrm{r} : K_2(E) \to H^1(E, \mathbb{R})$$

$$\{x, y\} \mapsto \left\{ \gamma \to \int_\gamma \eta(x, y) \right\}$$

for $\gamma \in H_1(E, \mathbb{Z})$, and

$$\eta(x, y) := \log|x| \, \mathrm{d} \arg y - \log|y| \, \mathrm{d} \arg x.$$

Here we think of $H^1(E, \mathbb{R})$ as the dual of $H_1(E, \mathbb{Z})$. The regulator is well defined because $\eta(x, 1-x) = \mathrm{d}D(x)$, where

$$D(z) = \mathrm{Im}(\mathrm{Li}_2(z)) + \arg(1-z) \log|z|$$

is the Bloch–Wigner dilogarithm.

In terms of the general formulation of Beilinson's conjectures this definition is not completely correct. One needs to go a step further and consider $K_2(\mathcal{E})$, where $\mathcal{E}$ is a Néron model of E over $\mathbb{Z}$. In particular, $K_2(\mathcal{E})$ is a subgroup of $K_2(E)$. It seems (see [Rodriguez Villegas 1999]) that a power of $\{x, y\}$ always lies in $K_2(\mathcal{E})$.

Assume that E is defined over $\mathbb{R}$. Because of the way that complex conjugation acts on η, the regulator map is trivial for the classes in $H_1(E, \mathbb{Z})^+$. In particular,

these cycles remain invariant under complex conjugation. Therefore it suffices to consider the regulator as a function on $H_1(E, \mathbb{Z})^-$.

We write $E(\mathbb{C}) \cong \mathbb{C}/\mathbb{Z} + \tau\mathbb{Z}$, where τ is in the upper half-plane. Then $\mathbb{C}/\mathbb{Z} + \tau\mathbb{Z} \cong \mathbb{C}^*/q^{\mathbb{Z}}$, where $z \bmod \Lambda = \mathbb{Z} + \tau\mathbb{Z}$ is identified with $e^{2i\pi z}$. Bloch [2000] defines the regulator function in terms of a Kronecker–Eisenstein series

$$R_\tau\left(e^{2\pi i(a+b\tau)}\right) = \frac{y_\tau^2}{\pi} {\sum_{m,n\in\mathbb{Z}}}' \frac{e^{2\pi i(bn-am)}}{(m\tau+n)^2(m\bar{\tau}+n)}, \tag{3-1}$$

where y_τ is the imaginary part of τ.

Let $J(z) = \log|z|\log|1-z|$, and let

$$D(x) = \text{Im}(\text{Li}_2(x)) + \arg(1-x)\log|x|$$

be the Bloch–Wigner dilogarithm.

Consider the function

$$J_\tau(z) = \sum_{n=0}^{\infty} J(zq^n) - \sum_{n=1}^{\infty} J(z^{-1}q^n) + \frac{1}{3}\log^2|q|B_3\left(\frac{\log|z|}{\log|q|}\right) \tag{3-2}$$

on $E(\mathbb{C}) \cong \mathbb{C}^*/q^{\mathbb{Z}}$, where $B_3(x) = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x$ is the third Bernoulli polynomial. If we recall that the elliptic dilogarithm is defined by

$$D_\tau(z) := \sum_{n\in\mathbb{Z}} D(zq^n), \tag{3-3}$$

then the regulator function (see [Bloch 2000]) is given by

$$R_\tau = D_\tau - iJ_\tau. \tag{3-4}$$

By linearity, R_τ extends to divisors with support in $E(\mathbb{C})$. Let x and y be nonconstant functions on E with divisors

$$(x) = \sum m_i(a_i), \qquad (y) = \sum n_j(b_j).$$

Following [Bloch 2000] and the notation in [Rodriguez Villegas 1999], we recall the diamond operation $\mathbb{C}(E)^* \otimes \mathbb{C}(E)^* \to \mathbb{Z}[E(\mathbb{C})]^-$

$$(x) \diamond (y) = \sum m_i n_j(a_i - b_j).$$

Here $\mathbb{Z}[E(\mathbb{C})]^-$ means that $[-P] \sim -[P]$.

Because R_τ is an odd function, we obtain a map

$$\mathbb{Z}[E(\mathbb{C})]^- \to \mathbb{R}.$$

Theorem 3.1 [Beĭlinson 1980]. *Let $E/\mathbb{R}$ be an elliptic curve, x, y nonconstant functions in $\mathbb{C}(E)$, and $\omega \in \Omega^1$. Then*

$$\int_{E(\mathbb{C})} \overline{\omega} \wedge \eta(x, y) = \Omega_0 R_\tau((x) \diamond (y)),$$

where Ω_0 is the real period.

Although a more general version of Beilinson's Theorem exists for elliptic curves defined over the complex numbers, the above version has a simpler formulation.

Corollary 3.2 (after an idea of Deninger). *If x and y are nonconstant functions in $\mathbb{C}(E)$ with trivial tame symbols, then*

$$-\int_\gamma \eta(x, y) = \operatorname{Im}\left(\frac{\Omega}{y_\tau \Omega_0} R_\tau((x) \diamond (y))\right), \quad \text{where } \Omega = \int_\gamma \omega.$$

Proof. Notice that $i\eta(x, y)$ is an element of the two-dimensional vector space $H^2_{\mathscr{D}}(E(\mathbb{C}), \mathbb{R}(2))$ generated by ω and $\overline{\omega}$. Then we may write

$$i\eta(x, y) = \alpha[\omega] + \beta[\overline{\omega}],$$

from which we obtain

$$\int_\gamma i\eta(x, y) = \alpha\Omega + \beta\overline{\Omega}.$$

On the other hand, we have

$$\int_{E(\mathbb{C})} i\eta(x, y) \wedge \overline{\omega} = \alpha \int_{E(\mathbb{C})} \omega \wedge \overline{\omega} = \alpha i 2\Omega_0^2 y_\tau,$$

and

$$\int_{E(\mathbb{C})} i\eta(x, y) \wedge \omega = -\beta i 2\Omega_0^2 y_\tau.$$

By Beilinson's Theorem

$$\int_\gamma i\eta(x, y) = -\frac{R_\tau((x) \diamond (y))\Omega}{2\Omega_0 y_\tau} + \frac{\overline{R_\tau((x) \diamond (y))}\overline{\Omega}}{2\Omega_0 y_\tau},$$

and the statement follows. $\square$

3.2. *Regulators and Mahler measure.* From now on, we will set $k = 4/\sqrt{t}$ in the first family (2-1).

Rodriguez Villegas [1999] proved that if $P_k(x, y) = k + x + 1/x + y + 1/y$ does not intersect the torus $\mathbb{T}^2$, then

$$m(k) \sim_{\mathbb{Z}} \frac{1}{2\pi} \mathrm{r}(\{x, y\})(\gamma). \tag{3-5}$$

Here the $\sim_{\mathbb{Z}}$ stands for "up to an integer", and γ is a closed path that avoids the poles and zeros of x and y. In particular, γ generates the subgroup $H_1(E, \mathbb{Z})^-$ of $H_1(E, \mathbb{Z})$ where conjugation acts by -1.

We would like to use this property, however we need to exercise caution. In particular, $P_k(x, y)$ intersects the torus whenever $|k| \le 4$ and $k \in \mathbb{R}$. Let us recall the idea behind the proof of (3-5) for the special case of $P_k(x, y)$. Writing

$$y P_k(x, y) = (y - y_{(1)}(x))(y - y_{(2)}(x)),$$

we have

$$m(k) = m(y P_k(x, y)) = \frac{1}{2\pi i} \int_{\mathbb{T}^1} \left(\log^+ |y_{(1)}(x)| + \log^+ |y_{(2)}(x)| \right) \frac{dx}{x}.$$

This last equality follows from applying Jensen's formula with respect to the variable y. When the polynomial does not intersect the torus, we may omit the $+$ sign on the logarithm, since each $y_{(i)}(x)$ is always inside or outside the unit circle. Indeed, there is always a branch inside the unit circle and a branch outside. It follows that

$$m(k) = \frac{1}{2\pi i} \int_{\mathbb{T}^1} \log |y| \frac{dx}{x} = -\frac{1}{2\pi} \int_{\mathbb{T}^1} \eta(x, y), \qquad (3\text{-}6)$$

where $\mathbb{T}^1$ is interpreted as a cycle in the homology of the elliptic curve defined by $P_k(x, y) = 0$, namely $H_1(E, \mathbb{Z})$.

If $k \in [-4, 4]$, then we may also assume that $k > 0$ since this particular Mahler measure does not depend on the sign of k. The equation

$$k + x + \frac{1}{x} + y + \frac{1}{y} = 0$$

certainly has solutions when $(x, y) \in \mathbb{T}^2$. However, for $|x| = 1$ and k real, the number $k + x + 1/x$ is real, and therefore $y + 1/y$ must be real. This forces two possibilities: either y is real or $|y| = 1$. Let $x = e^{i\theta}$, then for $-\pi \le \theta \le \pi$ we have

$$-k - 2\cos\theta = y + \frac{1}{y}. \qquad (3\text{-}7)$$

The limiting case occurs when $|k + 2\cos\theta| = 2$. Since we have assumed that k is positive, this condition becomes $k + 2\cos\theta = 2$, which implies that $y = -1$. When $k + 2\cos\theta > 2$ one solution for y, say, $y_{(1)}$, becomes a negative number less than -1, thus $|y_{(1)}| > 1$ (the other solution $y_{(2)}$ is such that $|y_{(2)}| < 1$). When $k + 2\cos\theta < 2$, y_i lies inside the unit circle and never reaches 1. What is important is that $|y_{(1)}| \ge 1$ and $|y_{(2)}| \le 1$, so we can still write (3-6) even if there is a nontrivial intersection with the torus.

3.3. *Functional identities for the regulator.* We recall a result by Bloch [2000] on the modularity of R_τ:

Proposition 3.3. *Take* $\left(\begin{smallmatrix}\alpha & \beta \\ \gamma & \delta\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, *and define* $\tau' = \dfrac{\alpha\tau+\beta}{\gamma\tau+\delta}$. *If we set*

$$\begin{pmatrix} b' \\ a' \end{pmatrix} = \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix},$$

then

$$R_{\tau'}(\mathrm{e}^{2\pi\mathrm{i}(a'+b'\tau')}) = \frac{1}{\gamma\bar{\tau}+\delta} R_\tau(\mathrm{e}^{2\pi\mathrm{i}(a+b\tau)}). \tag{3-8}$$

We will need to use some functional equations for J_τ. Recall the trivial property

$$J(z) = p \sum_{x^p = z} J(x). \tag{3-9}$$

Proposition 3.4. *Let p be an odd prime, let $q = \mathrm{e}^{2\pi\mathrm{i}\tau}$, and let $q_j = \mathrm{e}^{2\pi\mathrm{i}(\tau+j)/p}$ for $j \in \{0, 1, \dots, p-1\}$. Suppose that $(N, k) = 1$, and $p \equiv \pm 1$ or $0 \pmod N$. Then*

$$(1 + \chi_{-N}(p)p^2)J_{N\tau}(q^k) = \sum_{j=0}^{p-1} pJ_{N(\tau+j)/p}(q_j^k) + \chi_{-N}(p)J_{Np\tau}(q^{pk}), \tag{3-10}$$

and for any z,

$$(\chi_{-N}(p) + p^2)J_{N\tau}(z) = \sum_{j=0}^{p-1} pJ_{N(\tau+j)/p}(z) + \chi_{-N}(p)J_{Np\tau}(z). \tag{3-11}$$

Proof. Notice that

$$\sum_{j=0}^{p-1} J_{N(\tau+j)/p}(q_j^k) = \sum_{n=0}^{\infty}\sum_{j=0}^{p-1} J(q_j^{Nn+k}) - \sum_{n=1}^{\infty}\sum_{j=0}^{p-1} J(q_j^{Nn-k}) + \frac{4\pi^2 y_\tau^2 N^2}{3p} B_3\left(\frac{k}{N}\right).$$

By (3-9) this equals

$$\sum_{\substack{n=0 \\ p\nmid Nn+k}}^{\infty} \frac{1}{p}J(q^{Nn+k}) - \sum_{\substack{n=1 \\ p\nmid Nn-k}}^{\infty} \frac{1}{p}J(q^{Nn-k})$$

$$+ \sum_{\substack{n=0 \\ p\mid Nn+k}}^{\infty} pJ(q^{(Nn+k)/p}) - \sum_{\substack{n=1 \\ p\mid Nn-k}}^{\infty} pJ(q^{(Nn-k)/p}) + \frac{4\pi^2 y_\tau^2 N^2}{3p} B_3\left(\frac{k}{N}\right)$$

$$= \sum_{n=0}^{\infty} \frac{1}{p}J(q^{Nn+k}) - \sum_{n=1}^{\infty} \frac{1}{p}J(q^{Nn-k}) - \sum_{\substack{n=0 \\ p\mid Nn+k}}^{\infty} \frac{1}{p}J(q^{Nn+k}) + \sum_{\substack{n=1 \\ p\mid Nn-k}}^{\infty} \frac{1}{p}J(q^{Nn-k})$$

$$+ \sum_{\substack{n=0 \\ p\mid Nn+k}}^{\infty} pJ(q^{(Nn+k)/p}) - \sum_{\substack{n=1 \\ p\mid Nn-k}}^{\infty} pJ(q^{(Nn-k)/p}) + \frac{4\pi^2 y_\tau^2 N^2}{3p} B_3\left(\frac{k}{N}\right).$$

Upon rearranging, this expression becomes

$$\frac{1}{p}J_{N\tau}(q^k) - \frac{4\pi^2 y_\tau^2 N^2}{3p}B_3\left(\frac{k}{N}\right)$$

$$-\sum_{\substack{n=0\\p|Nn+k}}^{\infty}\frac{1}{p}J((q^p)^{(Nn+k)/p}) + \sum_{\substack{n=1\\p|Nn-k}}^{\infty}\frac{1}{p}J((q^p)^{(Nn-k)/p})$$

$$+\sum_{\substack{n=0\\p|Nn+k}}^{\infty}pJ(q^{(Nn+k)/p}) - \sum_{\substack{n=1\\p|Nn-k}}^{\infty}pJ(q^{(Nn-k)/p}) + \frac{4\pi^2 y_\tau^2 N^2}{3p}B_3\left(\frac{k}{N}\right),$$

or again

$$\frac{1}{p}J_{N\tau}(q^k) - \frac{\chi_{-N}(p)}{p}J_{Np\tau}(q^{pk}) + \chi_{-N}(p)pJ_{N\tau}(q^k).$$

This proves the assertion.

The second equality follows in a similar fashion. $\qquad\square$

It is possible to prove analogous identities for D_τ and R_τ.

Proposition 3.5. $\quad J_{(2\mu+1)/2}(e^{\pi i\mu}) = J_{2\mu}(e^{\pi i\mu}) - J_{2\mu}(-e^{\pi i\mu}).$

Proof. Let $z = e^{\pi i\mu}$. then

$$J_{2\mu}(z) - J_{2\mu}(-z)$$

$$= J(z) - J(-z) + \sum_{n=1}^{\infty}\left(J(zq^n) - J(-zq^n) - J(z^{-1}q^n) + J(-z^{-1}q^n)\right)$$

$$= \sum_{n=0}^{\infty}\left(J(e^{\pi i\mu(4n+1)}) - J(-e^{\pi i\mu(4n+1)}) - J(e^{\pi i\mu(4n+3)}) + J(-e^{\pi i\mu(4n+3)})\right).$$

On the other hand,

$$J_{(2\mu+1)/2}(z) = \sum_{n=0}^{\infty}\left(J((-1)^n e^{\pi i\mu(2n+1)}) - J((-1)^{n+1}e^{\pi i\mu(2n+1)})\right),$$

which proves the equality. $\qquad\square$

3.4. *The first family.* First we write the equation

$$x + \frac{1}{x} + y + \frac{1}{y} + k = 0$$

in Weierstrass form. Consider the rational transformation

$$X = \frac{k+x+y}{x+y} = -\frac{1}{xy}, \qquad Y = \frac{k(y-x)(k+x+y)}{2(x+y)^2} = \frac{(y-x)\left(1+\frac{1}{xy}\right)}{2xy},$$

which leads to

$$Y^2 = X\left(X^2 + (\tfrac{1}{4}k^2 - 2)X + 1\right).$$

It is useful to state the inverse transformation:

$$x = \frac{kX - 2Y}{2X(X - 1)}, \qquad y = \frac{kX + 2Y}{2X(X - 1)}.$$

Notice that E_k contains a torsion point of order 4 over $\mathbb{Q}(k)$, namely $P = (1, k/2)$. Indeed, this family is the modular elliptic surface associated to $\Gamma_0(4)$.

We can show that $2P = (0, 0)$, and $3P = (1, -k/2)$.

Now we have

$$(X) = 2(2P) - 2O,$$

$$(x) = (2(P) + (2P) - 3O) - (2(2P) - 2O) - ((P) + (3P) - 2O)$$
$$= (P) - (2P) - (3P) + O,$$

$$(y) = (2(3P) + (2P) - 3O) - (2(2P) - 2O) - ((P) + (3P) - 2O)$$
$$= -(P) - (2P) + (3P) + O.$$

Computing the diamond operation between the divisors of x and y yields

$$(x) \diamond (y) = 4(P) - 4(-P) = 8(P).$$

Now assume that $k \in \mathbb{R}$ and $k > 4$. We will choose an orientation for the curve and compute the real period. Because P is a point of order 4 and $\int_0^1 \omega$ is real, we may assume that P corresponds to $3\Omega_0/4$.

The next step is to understand the cycle $|x| = 1$ as an element of $H_1(E, \mathbb{Z})$. We would like to compute the value of $\Omega = \int_\gamma \omega$. First recall that

$$\omega = \frac{dX}{2Y} = \frac{dx}{x(y - y^{-1})}.$$

When $k > 4$, consider conjugation of ω. This sends x to x^{-1} and dx/x and $-dx/x$. There is no intersection with the torus, so y remains invariant. Therefore we conclude that Ω is the complex period, and $\Omega/\Omega_0 = \tau$, where τ is purely imaginary.

Therefore, for k real and $|k| > 4$,

$$m(k) = \frac{4}{\pi} \operatorname{Im}\left(\frac{\tau}{y_\tau} R_\tau(-i)\right).$$

Now take $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. By Proposition 3.3

$$R_\tau(-i) = R_\tau(e^{-2\pi i/4}) = \bar{\tau} R_{-1/\tau}(e^{-2\pi i/(4\tau)}),$$

therefore

$$m(k) = -\frac{4|\tau|^2}{\pi y_\tau} J_{-1/\tau}(e^{-2\pi i/(4\tau)}).$$

If we let $\mu = -1/(4\tau)$, then for $k \in \mathbb{R}$ we obtain

$$m(k) = -\frac{1}{\pi y_\mu} J_{4\mu}(e^{2\pi i\mu}) = \operatorname{Im}\left(\frac{1}{\pi y_\mu} R_{4\mu}(e^{2\pi i\mu})\right)$$

$$= \operatorname{Re}\left(\frac{16 y_\mu}{\pi^2} {\sum_{m,n}}' \frac{\chi_{-4}(m)}{(m+4\mu n)^2(m+4\bar\mu n)}\right),$$

thus recovering a result of Rodriguez Villegas. We can extend this result to all $k \in \mathbb{C}$, by arguing that both $m(k)$ and $-(1/(\pi y_\mu)) J_{4\mu}(e^{2\pi i\mu})$ are the real parts of holomorphic functions that coincide at infinitely many points; see [Rodriguez Villegas 1996].

Now we show how to deduce (1-7) and (1-6). Applying (3-10) with $N = 4$, $k = 1$, and $p = 2$, we have

$$J_{4\mu}(q) = 2J_{2\mu}(q_0) + 2J_{2(\mu+1)}(q_1),$$

which translates into

$$\frac{1}{y_{4\mu}} J_{4\mu}(e^{2\pi i\mu}) = \frac{1}{y_{2\mu}} J_{2\mu}(e^{\pi i\mu}) + \frac{1}{y_{2\mu}} J_{2\mu}(-e^{\pi i\mu}).$$

This is the content of (1-7). Setting $\tau = -1/(2\mu)$, we may also write

$$D_{\tau/2}(-i) = D_\tau(-i) + D_\tau(-ie^{\pi i\tau}). \tag{3-12}$$

Next we use Proposition 3.5:

$$J_{(2\mu+1)/2}(e^{\pi i\mu}) = J_{2\mu}(e^{\pi i\mu}) - J_{2\mu}(-e^{\pi i\mu}),$$

which translates into

$$\frac{1}{y_{(2\mu+1)/2}} J_{(2\mu+1)/2}(e^{\pi i\mu}) = \frac{2}{y_{2\mu}} J_{2\mu}(e^{\pi i\mu}) - \frac{2}{y_{2\mu}} J_{2\mu}(-e^{\pi i\mu}).$$

Setting $\tau = -1/(2\mu)$, and using $\left(\begin{smallmatrix} 1 & 0 \\ -2 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ on the left-hand side, we have

$$D_{(\tau-1)/2}(-i) = D_\tau(-i) - D_\tau(-ie^{\pi i\tau}). \tag{3-13}$$

Combining Equations (3-12) and (3-13), we see that

$$2D_\tau(-i) = D_{\tau/2}(-i) + D_{(\tau-1)/2}(-i).$$

This is the content of (1-6).

Similarly, we may deduce (2-14) from (3-10) when $k = 1$, $N = 4$, and p is an odd prime.

3.5. *A direct approach.* It is also possible to prove (1-6) and (1-7) directly, without considering the μ-parametrization or the explicit form of the regulator.

For those formulas, it is easy to explicitly write the isogenies at the level of the Weierstrass models. By using the well-known isogeny of degree 2

$$\phi : \left\{ E : y^2 = x(x^2 + ax + b) \right\} \to \left\{ \widehat{E} : \hat{y}^2 = \hat{x}(\hat{x}^2 - 2a\hat{x} + (a^2 - 4b)) \right\}$$

given by (see for example [Cassels 1991; Silverman 1992])

$$(x, y) \mapsto \left(\frac{y^2}{x^2}, \frac{y(b - x^2)}{x^2} \right),$$

where we require that $a^2 - 4b \neq 0$, we find

$$\phi_1 : E_{2n+(2/n)} \to E_{4n^2}, \quad (X, Y) \mapsto \left(\frac{X(n^2 X + 1)}{X + n^2}, -\frac{n^3 Y(X^2 + 2n^2 X + 1)}{(X + n^2)^2} \right),$$

$$\phi_2 : E_{2n+(2/n)} \to E_{4/n^2}, \quad (X, Y) \mapsto \left(\frac{X(X + n^2)}{n^2 X + 1}, -\frac{Y(n^2 X^2 + 2X + n^2)}{n(n^2 X + 1)^2} \right).$$

Write x_1, y_1, X_1, Y_1 for the rational functions and r_1 for the regulator in E_{4n^2}, and x_2, y_2, X_2, Y_2, r_2 for the corresponding objects in E_{4/n^2}. It follows that

$$\pm m(4n^2) = r_1(\{x_1, y_1\}) = \frac{1}{2\pi} \int_{|X_1|=1} \eta(x_1, y_1) = \frac{1}{4\pi} \int_{|X|=1} \eta(x_1 \circ \phi_1, y_1 \circ \phi_1)$$

$$= \tfrac{1}{2} r(\{x_1 \circ \phi_1, y_1 \circ \phi_1\}),$$

where the factor of 2 follows from the degree of the isogeny. Similarly, we find

$$\pm m\left(\frac{4}{n^2} \right) = r_2(\{x_2, y_2\}) = \tfrac{1}{2} r(\{x_2 \circ \phi_2, y_2 \circ \phi_2\}).$$

Now we need to compare the values of

$$r(\{x_1 \circ \phi_1, y_1 \circ \phi_1\}), \quad r(\{x_2 \circ \phi_2, y_2 \circ \phi_2\}), \quad \text{and} \quad r(\{x, y\}).$$

Recall that $(x) \diamond (y) = 8(P)$, where $P = (1, k/2)$. When $k = 2(n+1/n)$, we will also consider the point $Q = (-1/n^2, 0)$, which has order 2 (then $P + Q = (-1, n - 1/n)$, $2P + Q = (-n^2, 0)$, etc).

Let P now denote the point in $E_{2n+(2/n)}$, and let P_1 denote the corresponding point in E_{4n^2}. We have the following table:

$$\phi_1 : \quad \begin{array}{rcl} 3P, \ P+Q & \to & P_1, \\ 2P, \quad Q & \to & 2P_1, \\ P, \ 3P+Q & \to & 3P_1, \\ O_0, \ 2P+Q & \to & O_1. \end{array}$$

Using this table, and the divisors (x_1) and (y_1) in E_{4n^2}, we can compute $(x_1 \circ \phi_1) \diamond (y_1 \circ \phi_1)$. We find that

$$(x_1 \circ \phi_1) \diamond (y_1 \circ \phi_1) = -16(P) + 16(P + Q),$$

and similarly

$$(x_2 \circ \phi_2) \diamond (y_2 \circ \phi_2) = -16(P) - 16(P + Q).$$

These computations show that

$$\tfrac{1}{2}\, r_0(\{x_1 \circ \phi_1,\, y_1 \circ \phi_1\}) + \tfrac{1}{2}\, r_0(\{x_2 \circ \phi_2,\, y_2 \circ \phi_2\}) = 2\, r_0(\{x_0, y_0\}), \qquad \text{(3-14)}$$

and therefore

$$r_1(\{x_1, y_1\}) + r_2(\{x_2, y_2\}) = 2\, r_0(\{x_0, y_0\}). \qquad \text{(3-15)}$$

We can conclude the proof of (1-6) by inspecting signs.

To prove (1-7), it is necessary to use the isomorphism ϕ from (3-16).

3.6. Relations among $m(2)$, $m(8)$, $m(3\sqrt{2})$, and $m(i\sqrt{2})$. Setting $n = 1/\sqrt{2}$ in (1-7), we obtain

$$m(3\sqrt{2}) + m(i\sqrt{2}) = m(8).$$

Doing the same in (1-6), we find that

$$m(2) + m(8) = 2m(3\sqrt{2}).$$

In this section we will establish the identity

$$3m(3\sqrt{2}) = 5m(i\sqrt{2}),$$

from which we can deduce expressions for $m(2)$ and $m(8)$.

Consider the functions f and $1 - f$, where $f = (\sqrt{2}Y - X)/2 \in \mathbb{C}(E_{3\sqrt{2}})$. Their divisors are

$$\left(\frac{\sqrt{2}Y - X}{2}\right) = (2P) + 2(P + Q) - 3O,$$

$$\left(1 - \frac{\sqrt{2}Y - X}{2}\right) = (P) + (Q) + (3P + Q) - 3O.$$

The diamond operation yields

$$(f) \diamond (1 - f) = 6(P) - 10(P + Q).$$

But $(f) \diamond (1 - f)$ is trivial in K-theory, hence

$$6(P) \sim 10(P + Q).$$

Now consider the isomorphism

$$\phi : E_{2n+(2/n)} \to E_{2(in+1/in)}, \quad (X, Y) \mapsto (-X, iY). \qquad \text{(3-16)}$$

This isomorphism implies that

$$r_{i\sqrt{2}}(\{x, y\}) = r_{3\sqrt{2}}(\{x \circ \phi, y \circ \phi\}).$$

But we know that

$$(x \circ \phi) \diamond (y \circ \phi) = 8(P + Q).$$

This implies

$$6\, r_{3\sqrt{2}}(\{x, y\}) = 10\, r_{i\sqrt{2}}(\{x, y\}) \quad \text{and} \quad 3m(3\sqrt{2}) = 5m(i\sqrt{2}).$$

We conclude that

$$m(8) = \tfrac{8}{5}m(3\sqrt{2}), \quad m(2) = \tfrac{2}{5}m(3\sqrt{2}),$$

and finally

$$m(8) = 4m(2).$$

3.7. *The Hesse family.* We will now sketch the case of the Hesse family:

$$x^3 + y^3 + 1 - \frac{3}{t^{1/3}}xy.$$

This family corresponds to $\Gamma_0(3)$. The diamond operation yields

$$(x) \diamond (y) = 9(P) + 9(A) + 9(B), \tag{3-17}$$

where P is a point of order 3, defined over $\mathbb{Q}(t^{1/3})$, and A, B are points of order 3 such that $A + B + P = O$.

For $0 < t < 1$, we have

$$n(t) = \frac{9}{2\pi} \operatorname{Im}\left(\frac{\tau}{y_\tau}\left(R_\tau(e^{4\pi i/3}) + R_\tau(e^{4\pi i(1+\tau)/3}) + R_\tau(e^{2\pi i(2+\tau)/3}) \right) \right).$$

If we let $\mu = -1/\tau$, we obtain, after several steps,

$$n(t) = \operatorname{Re}\left(\frac{27\sqrt{3}y_\mu}{4\pi^2} {\sum_{k,n}}' \frac{\chi_{-3}(n)}{(3\mu k + n)^2(3\bar{\mu}k + n)} \right).$$

As in the previous example, this result may be extended to the complement of κ (the set of t where the polynomial intersects the torus) by comparing holomorphic functions.

3.8. *The $\Gamma_0^0(6)$ example.* We will now sketch a treatment of Stienstra's example [2006]:

$$(x + 1)(y + 1)(x + y) - \frac{1}{t}xy.$$

Applying the diamond operation, we have

$$(x) \diamond (y) = -6(P) - 6(2P),$$

where P is a point of order 6.

For t small, one can write

$$g(t) = \frac{3}{\pi} \operatorname{Im}(\frac{\tau}{y_\tau} R_\tau(\xi_6^{-1}) + R_\tau(\xi_3^{-1})).$$

Eventually, one reaches the expression for $g(t)$ found in [Stienstra 2006]:

$$\operatorname{Re}\left(\frac{36 y_\mu}{\pi^2} \sideset{}{'}\sum_{m,n} \frac{\chi_{-3}(m)}{(m+6\mu n)^2 (m+6\bar\mu n)}\right) + \operatorname{Re}\left(\frac{9 y_\mu}{2\pi^2} \sideset{}{'}\sum_{m,n} \frac{\chi_{-3}(m)}{(m+3\mu n)^2 (m+3\bar\mu n)}\right),$$

3.9. The $\Gamma_0^0(5)$ example. Our final example is

$$(x + y + 1)(x + 1)(y + 1) - \frac{1}{t} xy.$$

Applying the diamond operation, we find that

$$(x) \diamond (y) = 10(P) + 5(2P),$$

where P is a torsion point of order 5.

For $t > 0$,

$$r(t) = \frac{5}{2\pi} \operatorname{Im}\left(\frac{\tau}{y_\tau} \left(2 R_\tau(e^{8\pi i/5}) + R_\tau(e^{6\pi i/5})\right)\right).$$

Finally,

$$r(t) = -\operatorname{Re}\left(\frac{25 i y_\mu}{4\pi^2} \sideset{}{'}\sum_{m,n} \frac{2(\zeta_5^m - \zeta_5^{-m}) + \zeta_5^{2m} - \zeta_5^{-2m}}{(m + 5\mu n)^2 (m + 5\bar\mu n)}\right).$$

In conclusion, we see that the modular structure comes from the form of the regulator function, and the functional identities are consequences of the functional identities of the elliptic dilogarithm.

4. Conclusion

We have used both regulator and q-series methods to prove a variety of identities between the Mahler measures of genus-one polynomials. We will conclude this paper with a final open problem.

Open Problem 2. How do you characterize all the functional equations of $\mu(t)$?

We have seen that there are identities like (1-6), stating that

$$2\mathrm{m}\left(2\left(k + \frac{1}{k}\right) + x + \frac{1}{x} + y + \frac{1}{y}\right)$$
$$= \mathrm{m}\left(4k^2 + x + \frac{1}{x} + y + \frac{1}{y}\right) + \mathrm{m}\left(\frac{4}{k^2} + x + \frac{1}{x} + y + \frac{1}{y}\right).$$

While this formula does not follow from (2-14), it can be proved with regulators.

Indeed, the last section showed us that we can obtain functional identities for the Mahler measures by looking at functional equations for the elliptic dilogarithm.

Now, understanding these identities is a very hard problem. To give an idea of the dimensions of this problem, we note that (3-10) corresponds to the integration of an identity for the Hecke operator T_p. This suggests that more identities will follow from looking at the general operator T_n. And this is just the beginning of the story...

Acknowledgements

References

[Andrews and Berndt 2005] G. E. Andrews and B. C. Berndt, *Ramanujan's lost notebook, Part I*, Springer, New York, 2005. MR 2005m:11001 Zbl 1075.11001

[Beĭlinson 1980] A. A. Beĭlinson, "Higher regulators and values of *L*-functions of curves", *Funktsional. Anal. i Prilozhen.* **14**:2 (1980), 46–47. In Russian; translated in *Funct. Anal. Appl.* **14**:2 (1980), 116–118. MR 81k:14020 Zbl 0475.14015

[Berndt 1989] B. C. Berndt, *Ramanujan's notebooks, Part II*, Springer, New York, 1989. MR 90b:01039 Zbl 0716.11001

[Berndt 1991] B. C. Berndt, *Ramanujan's notebooks, Part III*, Springer, New York, 1991. MR 92j:01069 Zbl 0733.11001

[Berndt 1998] B. C. Berndt, *Ramanujan's notebooks, Part V*, Springer, New York, 1998. MR 99f:11024 Zbl 0886.11001

[Bertin 2004] M. J. Bertin, "Mesure de Mahler d'une famille de polynômes", *J. Reine Angew. Math.* **569** (2004), 175–188. MR 2005g:11204 Zbl 1048.11081

[Bloch 2000] S. J. Bloch, *Higher regulators, algebraic K-theory, and zeta functions of elliptic curves*, CRM Monograph Series **11**, American Mathematical Society, Providence, RI, 2000. MR 2001i:11082 Zbl 0958.19001

[Boyd 1998] D. W. Boyd, "Mahler's measure and special values of L-functions", *Experiment. Math.* **7**:1 (1998), 37–82. MR 99d:11070 Zbl 0932.11069

[Cassels 1991] J. W. S. Cassels, *Lectures on elliptic curves*, London Mathematical Society Student Texts **24**, Cambridge University Press, Cambridge, 1991. MR 92k:11058 Zbl 0752.14033

[Deninger 1997] C. Deninger, "Deligne periods of mixed motives, K-theory and the entropy of certain $\mathbf{Z}^n$-actions", *J. Amer. Math. Soc.* **10**:2 (1997), 259–281. MR 97k:11101 Zbl 0913.11027

[Kurokawa and Ochiai 2005] N. Kurokawa and H. Ochiai, "Mahler measures via the crystalization", *Comment. Math. Univ. St. Pauli* **54**:2 (2005), 121–137. MR 2006j:11145 Zbl 05017542

[Maier 2006] R. S. Maier, "Algebraic hypergeometric transformations of modular origin", preprint, 2006. math.NT/0501425

[Rodriguez Villegas 1996] F. Rodriguez Villegas, "Modular Mahler measures", preprint, Princeton University, 1996, Available at http://www.math.utexas.edu/~villegas/mahler.dvi.

[Rodriguez Villegas 1999] F. Rodriguez Villegas, "Modular Mahler measures, I", pp. 17–48 in *Topics in number theory: in honor of B. Gordon and S. Chowla* (University Park, PA, 1997), edited by S. D. Ahlgren et al., Math. Appl. **467**, Kluwer, Dordrecht, 1999. MR 2000e:11085 Zbl 0980.11026

[Rodriguez Villegas 2002] F. Rodriguez Villegas, "Identities between Mahler measures", pp. 223–229 in *Number theory for the millennium, III* (Urbana, IL, 2000), edited by M. A. Bennett, A K Peters, Natick, MA, 2002. MR 2003m:11177 Zbl 1029.11054

[Rogers 1920] L. J. Rogers, "On a type of modular relation", *Proc. London Math. Soc.* **19** (1920), 387–397. Zbl 48.0151.02

[Silverman 1992] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer, New York, 1992. MR 95m:11054 Zbl 0585.14026

[Stienstra 2006] J. Stienstra, "Mahler measure variations, Eisenstein series and instanton expansions", pp. 139–150 in *Mirror symmetry, V*, edited by N. Yui et al., AMS/IP Stud. Adv. Math. **38**, Amer. Math. Soc., Providence, RI, 2006. MR MR2282958 Zbl 05153032

[Verrill 2001] H. A. Verrill, "Picard–Fuchs equations of some families of elliptic curves", pp. 253–268 in *Proceedings on Moonshine and related topics* (Montreal, 1999), edited by J. McKay and A. Sebbar, CRM Proc. Lecture Notes **30**, Amer. Math. Soc., Providence, RI, 2001. MR 2003k:11065 Zbl 1082.14503

mlalin@math.ubc.ca   *Department of Mathematical and Statistical Sciences, 632 Central Academic Building, University of Alberta, Edmonton, AB T6G 2G1, Canada* http://www.math.ubc.ca/~mlalin

matrogers@math.ubc.ca   *Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada* http://www.math.ubc.ca/~matrogers

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in ANT are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory