Modular equations and the genus zero property of moonshine functions

C.J. Cummins and T. Gannon*

Department of Mathematics
Concordia University
Montreal, H3GIM8

Canada

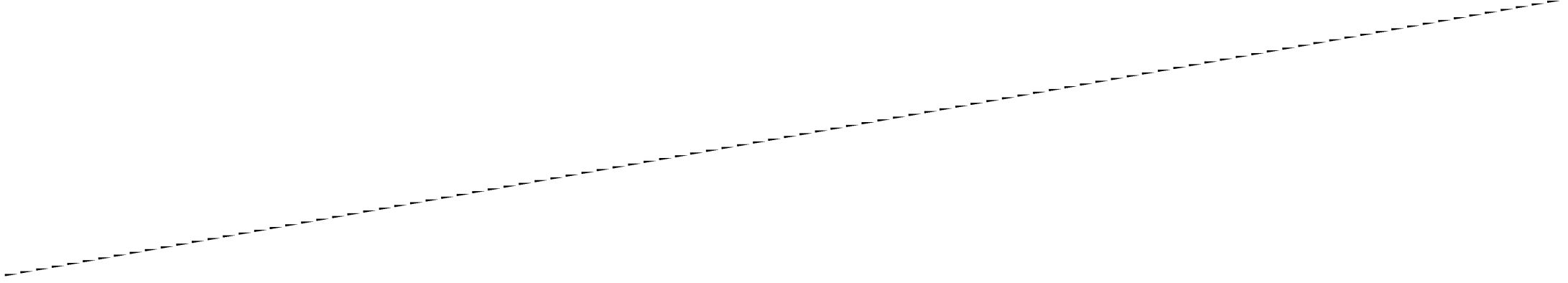
Max-Planck-Institut
für Mathematik
Gottfried-Claren-Str. 26
53225 Bonn

Germany

*Department of Mathematics
York University
North York, M3JIP3

Canada

MPI 96-72



Modular equations and the genus zero property of moonshine functions.

C J Cummins[1] and T Gannon[2,3]

Abstract. *In this paper we obtain the following result: Let f be an analytic function on the upper half plane with Fourier expansion $f(z) = q^{-1} + \sum_{n=1}^{\infty} a_n q^n$, $q = \exp(2\pi i z)$. If $a_i \in \mathbb{Z}[\zeta_K]$, $i = 1, 2, \ldots$, for some integer $K > 0$, then the following are equivalent:*

1 *f satisfies a modular equation of order n for all $n = 1 \pmod K$.*

2 *f is either $q^{-1} + \zeta q$ where $\zeta^{\gcd(24,K)+1} = \zeta$, or is a Hauptmodul for a subgroup G of $SL(2,\mathbb{R})$ satisfying:*

- **a** *G contains $\Gamma_0(N)$ with finite index for some $N \mid K^\infty$.*
- **b** *G contains $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ if and only if $k \in \mathbb{Z}$.*
- **c** *The Riemann surface $X(G)$ has genus zero.*

In Borcherds' proof of the moonshine conjectures the moonshine functions satisfy condition **1** *of this theorem, for a suitable choice of K, by virtue of the twisted denominator formulae of the Monster Lie algebra and hence are Hauptmoduln. A conceptual explanation is thus provided of the genus zero property of these functions.*

§1. Introduction.

Let $\mathbb{M}$ be the Monster simple group. Let $o(g)$, $g \in \mathbb{M}$, be the order of g. Define

$$\Gamma_0(N) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \mid c = 0 \pmod N\},$$

$$\Gamma(N) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod N\}$$

1 email: cummins@abacus.concordia.ca

2 email: gannon@mpim-bonn.mpg.de

3 Address from September 1996:
Department of Mathematics, York University, North York, CANADA, M3J 1P3

Work supported by NSERC and FCAR grants

AMS subject classification: 11F03, 11F22, 30F35

and let $\Lambda(N)$ be the normalizer of $\Gamma_0(N)$ in $SL(2,\mathbb{R})$. We shall call a discrete subgroup Δ of $SL(2,\mathbb{R})$ a congruence group if it contains $\Gamma(N)$ for some N. Necessarily the index of $\Gamma(N)$ in Δ is finite and Δ acts on the extended upper half plane $\mathcal{H}^* = \mathcal{H} \cup \mathbb{Q} \cup \{i\infty\}$ by fractional linear transformations. The quotient $\Delta\backslash\mathcal{H}^*$ has the structure of a compact Riemann surface and will be denoted by $X(\Delta)$. Conway and Norton [CN] conjectured relationships between certain congruence groups and $\mathbb{M}$ known as the moonshine conjectures. These have been proved by Borcherds:

Theorem 1.1. *[B] For each $g \in \mathbb{M}$ there exists a formal q-expansion*

$$j_g(q) = q^{-1} + \sum_{n=1}^{\infty} a_n(g) q^n,$$

$a_n(g) \in \mathbb{Z}$, $n \geq 1$, such that :

1. *For all $n \geq 1$ the map $g \mapsto a_n(g)$ is a character of $\mathbb{M}$.*
2. *For each g there exists $h \,|\, \gcd(24, o(g))$ and a congruence group $\Delta(g)$ of $SL(2,\mathbb{R})$ such that*
$$\Gamma_0(ho(g)) \leq \Delta(g) \leq \Lambda(ho(g))$$
3. *The genus of $X(\Delta(g))$ is zero.*
4. *Replacing q by $\exp(2\pi i z)$ in j_g yields the Fourier expansion of a function $f_g(z)$ which is analytic on $\mathcal{H}$. The field of automorphic functions* of $\Delta(g)$ is $\mathbb{C}(f_g)$.*

Conway and Norton also conjectured that the series of Theorem 1.1 are related by replication formulae. These are also proved by Borcherds:

Theorem 1.2. *[B] Let $\zeta_d = \exp(2\pi i/d)$. For each $g \in \mathbb{M}$ and for all $n \geq 1$ there exists a polynomial $Q_{n,g}$ such that*

$$\sum_{\substack{ad=n \\ 0 \leq b < d}} j_{g^a}(\zeta_d^b q^{\frac{a}{d}}) = Q_{n,g}(j_g(q)). \tag{1.1}$$

Note that $Q_{n,g}$ is the unique polynomial such that

$$Q_{n,g}(j_g(q)) = q^{-n} + \text{ terms of positive degree.}$$

Polynomials with this property are known as the Faber polynomials [D, chapter 4]. We shall use the notation $\zeta_m = \exp(2\pi i/m)$ throughout this paper.

The outline of Borcherds' proof is as follows: First using vertex operator techniques a generalized Kac-Moody algebra, called the Monster Lie algebra, is constructed on which the Monster acts [B]. This Lie algebra has a $\mathbb{Z} \times \mathbb{Z}$ grading which is respected by $\mathbb{M}$. The $(m,n) \neq (0,0)$ graded piece of the Monster Lie algebra affords the representation V_{mn} of $\mathbb{M}$, where V_n is the representation of $\mathbb{M}$ of conformal weight $n+1$ appearing in the

* All automorphic functions in this paper are of weight zero. A modular function of level N is an automorphic function for $\Gamma(N)$.

Monster vertex operator algebra [FLM]. Let $a_n(g)$ be the trace of $g \in \mathbb{M}$ on V_n. By using a generalization of the Weyl denominator formula Borcherds finds a denominator identity for each conjugacy class in $\mathbb{M}$. For the identity element of $\mathbb{M}$ the formula is:

$$p^{-1} \prod_{m>0,\, n\in\mathbf{Z}} (1-p^m q^n)^{c_{mn}} = j(p) - j(q),$$

where the c_n are the coefficients of the q-expansion of

$$j(q) = q^{-1} + 196884\, q + 21493760\, q^2 + 864299970\, q^3 + \ldots,$$

the normalized generator or *Hauptmodul* for $SL(2,\mathbb{Z})$. For the other conjugacy classes the formula is:

$$p^{-1} \exp(-\sum_{i>0} \sum_{m>0,n\in\mathbf{Z}} a_{mn}(g^i) p^{mi} q^{ni}/i) = \sum_{m\in\mathbf{Z}} a_m(g) p^m - \sum_{m\in\mathbf{Z}} a_m(g) q^m. \tag{1.2}$$

Consider next the formal q-series $j_g(q) = \sum_{m\in\mathbf{Z}} a_m(g) q^m$, $g \in \mathbb{M}$. We want to show that these j_g satisfy the conditions of Theorem 1.1. By construction property **1** is satisfied and so it remains to show that each j_g is an automorphic function with the required properties. Borcherds' proof of this is to note firstly that equation (1.2) implies that the j_g, $g \in \mathbb{M}$ satisfy the replication identities of Theorem 1.2 as *formal* series. Koike [Koi] has shown that the modular functions attached to the elements of $\mathbb{M}$ by Conway and Norton satisfy the replication formulae (this is shown in more generality in [CuN]). Any formal series which satisfies the replication formulae must obey recurrence relations which determine the series from a finite number of terms [N1,B; cf. Mah]. Thus, by checking a finite number of coefficients, it can be verified that each j_g is the expected automorphic function.

Unfortunately this last step in the proof does not offer any conceptual understanding of why the series involved are automorphic functions of genus zero. In this paper we address this problem which was first posed by Norton [N1] in terms of completely replicable functions (see section 8).

Let

$$h(q) = q^{-1} + \sum_{n=1}^{\infty} a_n q^n$$

with $a_n \in \mathbb{C}$, be a formal q-series and let $\psi(n) = n \prod_{\substack{p \mid n \\ p \text{ prime}}} (1 + \frac{1}{p})$. A modular polynomial of order $n > 1$ for h is a polynomial $F_n(x,y) \in \mathbb{C}[x,y]$ such that*:

M.1 $F_n(x,y) = F_n(y,x)$.
M.2 $F_n(x,y)$ is a monic polynomial of degree $\psi(n)$ in x (and y).
M.3 For all $a,b,d \in \mathbb{Z}$ such that $ad = n$, $\gcd(a,b,d) = 1$ and $0 \le b < d$, $F_n(h(q), h(\zeta_d^b q^{\frac{a}{d}})) = 0$ as formal q-series.

* Although we have included **M.1** as a separate property, it is in fact a consequence of **M.2** and **M.3**, cf. [K, Proposition 3.2].

We call $F_n(h(q), h(\zeta_d^b q^{\frac{a}{d}})) = 0$ a modular equation of order n for h. The study of modular equations, particularly those of the j function, have a long history, for example see [Sh, section 4.6] and [L, chapter 5 §2]. We note that one consequence of Theorem 1.4 below is that many of the classical results on singular values of the j function can be extended to any Hauptmodul. Some partial results along these lines are contained in [CY].

Mahler [Mah, Theorem 8] has shown that if h satisfies a modular equation for some prime p, then it is the Laurent expansion of a meromorphic function defined in some neighbourhood of $q = 0$. Kozlov [K, Proposition 3.3] shows that if h satisfies a modular equation of order p for infinitely many primes p, then it is the Laurent expansion of a function analytic on the interior of the unit disc $|q| < 1$, except for a pole at $q = 0$ and so, in this case, the function $f(z) = h(\exp(2\pi iz))$ extends to an analytic function on the upper half plane. Moreover f satisfies:

M.3′ For all $a, b, d \in \mathbb{Z}$ such that $ad = n$, $\gcd(a, b, d) = 1$ and $0 \leq b < d$, $F_n(f(z), f((az + b)/d)) = 0$.

Thus in the rest of this paper we may, without loss of generality, consider f to be an analytic function on the upper half plane and we shall set $q = \exp(2\pi iz)$ so that f has Fourier expansion $q^{-1} + \sum_{n=1}^{\infty} a_n q^n$. Let

$$G(f) = \{m \in SL(2, \mathbb{R}) \mid f(m(z)) = f(z)\}$$

we shall call $G(f)$ the symmetry group of f. Our main results are stated in Theorems 1.3 and 1.4. In this paper if a and b are integers such that a divides some power of b, then we write $a \,|\, b^{\infty}$.

Theorem 1.3. *Let f be an analytic function on $\mathcal{H}$ with Fourier expansion $f(z) = q^{-1} + \sum_{n=1}^{\infty} a_n q^n$, where $q = \exp(2\pi iz)$ and $a_i \in \mathbb{C}$, $i = 1, 2, \ldots$. Let $K > 0$ be an integer and suppose f satisfies a modular equation of order n for all $n = 1 \pmod K$. Then*

1. *If $G(f) \neq \{\pm \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \mid t \in \mathbb{Z}\}$, then f is is a Hauptmodul for $G(f)$ which satisfies:*
   - **a** *$G(f)$ contains $\Gamma_0(N)$ with finite index for some $N \,|\, K^{\infty}$.*
   - **b** *$G(f)$ contains $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ if and only if $k \in \mathbb{Z}$.*
   - **c** *$X(G(f))$ has genus zero.*
2. *If $G(f) = \{\pm \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \mid t \in Z\}$ and the coefficients a_n, $n = 1, 2, \ldots$ are algebraic integers, then $f(z) = q^{-1} + \zeta q$ where $\zeta^{\gcd(24,K)+1} = \zeta$.*

As observed by Norton [N1] and shown by Kozlov [K], if $a_m(g) \in \mathbb{Q}$ for all $m \in \mathbb{Z}^*$ and all $g \in \mathbb{M}$ then equation (1.2) implies the existence of a modular polynomial of order n for $j_g(q)$ for all n coprime to $o(g)$. Thus setting $\Delta(g) = G$, Theorem 1.3 implies that

---

* That the coefficients of the moonshine functions are rational can be seen without a case by case analysis, for example from the existence of a $\mathbb{Z}[1/2]$ form of the moonshine vertex operator algebra [BR].

$\Delta(g)$ is a congruence group which satisfies properties **3** and **4** of Theorem 1.1 and contains $\Gamma_0(N)$ for some N, where $N \mid o(g)^\infty$. It has been shown by Thompson [T] that there are only finitely many groups G satisfying conditions **a**, **b** and **c** of Theorem 1.3.

In order to state Theorem 1.4 we must introduce generalized modular equations. If f has rational coefficients, then a generalized modular equation is a modular equation, but in general this is not the case. Let $N \in \mathbb{Z}^{>0}$. For n coprime to N, $*n$ will denote the Galois automorphism of $\mathbb{Q}(\zeta_N)$ such that $\zeta_N * n = \zeta_N^n$. With the same notation as above, if $a_i \in \mathbb{Q}(\zeta_N)$, $i = 1, 2, \ldots$ then we define a generalized modular polynomial of order $n > 1$ for h to be a polynomial $F_n(x, y) \in \mathbb{Q}(\zeta_N)[x, y]$ such that**:

MI.1 $F_n(x, y) = (F_n * n)(y, x)$

MI.2 $F_n(x, y)$ is a monic polynomial of degree $\psi(n)$ in x (and y).

MI.3 For all $a, b, d \in \mathbb{Z}$ such that $ad = n$, $\gcd(a, b, d) = 1$ and $0 \le b < d$, $F_n((h * n)(q), h(\zeta_d^b q^{\frac{a}{d}})) = 0$ as formal q-series.

In this definition $F_n * n$ is the polynomial obtained from F_n by applying $*n$ to each of its coefficients and $h * n$ is obtained from h by applying $*n$ to each of its coefficients. If $f(z) = h(\exp(2\pi i z))$ and $(f * n)(z) = (h * n)(\exp(2\pi i z))$ extend to analytic functions on $\mathcal{H}$, then a generalized modular polynomial for $f(z)$ satisfies **MI.1** and **MI.2** while **MI.3** is replaced by:

MI.3′ For all $a, b, d \in \mathbb{Z}$ such that $ad = n$, $\gcd(a, b, d) = 1$ and $0 \le b < d$, $F_n((f * n)(z), f((az + b)/d)) = 0$.

We then have the following:

Theorem 1.4.

1 *If f be a Hauptmodul for a subgroup G of $SL(2, \mathbb{R})$ satisfying*

- **a** *G contains $\Gamma_0(N)$ with finite index for some N.*
- **b** *G contains $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ if and only if $k \in \mathbb{Z}$.*
- **c** *$X(G)$ has genus zero.*
- **d** *H is a subfield of $\mathbb{Q}(\zeta_N)$, where H is the field generated over $\mathbb{Q}$ by the coefficients of f.*

Then the exponent of the Galois group $\mathrm{Gal}(H/\mathbb{Q})$ is 1 or 2, i.e. H is a composite of quadratic fields, and there exists a generalized modular equation $F_n(x, y)$ for f of order n for all n coprime to N. Also $F_n(f * n, y)$ is irreducible over $\mathbb{C}(f * n)$.

2 *Let $K > 0$ be an integer. If $f = q^{-1} + \zeta q$ where $\zeta^{\gcd(24,K)+1} = \zeta$, then there exists a generalized modular equation for f of order n for all n coprime to K.*

It is perhaps worthwhile to mention that one consequence of Theorem 1.3 is that if f satisfies the hypothesis of Theorem 1.4 **1**, then for all n coprime to N, $f * n$ is a Hauptmodul for the group $G(f * n)$ which satisfies conditions **a**, **b** and **c** of Theorem 1.4.

A full discussion of property **2** of Theorem 1.1 is beyond the scope of this paper. However, if j_g is a Hauptmodul for a group G such that G contains $\Gamma_0(N)$ and $p \nmid N$

** **MI.1** is implied by **MI.2**, **MI.3** and the condition that $\mathrm{Gal}(H/\mathbb{Q})$ has exponent 1 or 2 (H as in Theorem 1.4), see Proposition 6.18.

for some prime p, then, since the Monstrous functions have rational coefficients, Theorem 1.4 implies that j_g satisfies a modular equation of order p. Since the sum of the roots of this modular equation is a polynomial in j_g, by Theorem 1.2 $j_g = j_{g^p}$. In other words, if $j_g \neq j_{g^p}$ and G contains $\Gamma_0(N)$ then $p|N$.

Results related to ours include the work of Martin [Mar], who uses the behaviour of completely replicable functions at "bad" primes to derive invariance under certain congruence groups, and Cohn and McKay [CM], who conjecture, based on computational evidence, that the hypotheses of Theorem 1.3 can be considerably weakened and show, using computer algebra, that if f has integer coefficients and satisfies modular equations of order 2 and 3 then f is either a Hauptmodul or trivial.

In the next four sections we shall take f to be an analytic function on $\mathcal{H}$, with Fourier expansion $f(z) = q^{-1} + \sum_{n=1}^{\infty} a_n q^n$, $q = \exp(2\pi i z)$ and we assume that f has a modular polynomial $F_n(x, y)$ for all $n = 1 \pmod K$. Our aim is to show that f is one of the trivial functions or a Hauptmodul; a result that is obtained in section 7.

The proof is in several stages. In section 2 we show that if $f(z_1) = f(z_2)$, then there is an analytic bijection α from a neighbourhood D_1 of z_1 to a neighbourhood D_2 of z_2 such that $f(\alpha(z)) = f(z)$ for all z in D_1. We call such an α a *local symmetry* of f. If α extends to an automorphism of $\mathcal{H}$, then we say it is a *global symmetry*, or simply a *symmetry*, of f. In section 3 we show that any local symmetry has a unique maximal domain to which it can be extended. Section 4 shows that this maximal domain must be $\mathcal{H}$ and so every local symmetry of f is a global symmetry. In section 5 we address the problem of showing that the symmetry group of f contains $\Gamma_0(N)$ for some N. In section 6 a proof of Theorem 1.4 **1** is given. In section 7 the results are combined to obtain Theorem 1.3 and Theorem 1.4 **2**. In the last section we make some conjectures and comments.

Acknowledgements We thank Yves Martin for many interesting and useful discussions. T.G. appreciates the hospitality shown to him by the Department of Mathematics and Statistics of Concordia University and the Max-Planck-Institut für Mathematik. We also thank Simon Norton and John McKay for comments on an earlier version of this paper and Hershy Kisilevsky for comments on section 6.

§2. Existence of local symmetries.

We shall make repeated use of Definition 2.1 and Lemma 2.2 in the rest of this paper:

Definition 2.1. *Let $\mathcal{N}_K$ be the set of all positive integers $n = 1$ (mod K). We will denote elements of $PGL(2,\mathbb{Q})^+$, where the $+$ denotes positive determinants, by angular brackets. In particular, if m is a nonsingular 2×2 integer matrix with positive determinant, then $\langle m \rangle$ will denote the corresponding element of $PGL(2,\mathbb{Q})^+$. Any element of $\left\langle \begin{matrix} a & b \\ c & d \end{matrix} \right\rangle \in PGL(2,\mathbb{Q})^+$ acts on $\mathcal{H}$ by the fractional linear transformation $z \mapsto (az+b)/(cz+d)$. For any positive integer n define $A(n) = \{\left\langle \begin{matrix} a & b \\ 0 & d \end{matrix} \right\rangle \mid ad = n, \gcd(a,b,d) = 1, 0 \leq b < d\}$.*

Lemma 2.2. *There exists a $t \in \mathbb{R}$ such that if $\mathrm{Im}\,(z_1) > t$ and $\mathrm{Im}\,(z_2) > t$ and $f(z_1) = f(z_2)$ then $z_1 - z_2 \in \mathbb{Z}$.*

Proof: Let $\bar{f}$ be the function meromorphic on the interior of the unit disk defined by the relation $\bar{f}(\exp(2\pi iz)) = f(z)$ for all $z \in \mathcal{H}$. Then the coefficients of the Laurent expansion of $\bar{f}$ are the Fourier coefficients a_n of f. The Lemma follows immediately from the fact that $\bar{f}(q)$ has a simple pole at $q = 0$. ∎

Define $\mathcal{H}_t = \{z \in \mathcal{H} \mid \mathrm{Im}\,(z) > t\}$. We shall call $\mathcal{H}_t$ the injective region. Recall the definition of a local symmetry of f given in the introduction.

Proposition 2.3. *If $z_1, z_2 \in \mathcal{H}$ and $f(z_1) = f(z_2)$ then there exists a local symmetry $\alpha : D_1 \to D_2$ of f for which $\alpha(z_1) = z_2$.*

The proof of Proposition 2.3 is a generalization of that of Corollary 3.7 in [K], obtained by restricting attention to those $n \in \mathcal{N}_K$. The only place where this generalization is not immediate is in establishing the following:

Lemma 2.4. *If $z_1, z_2 \in \mathcal{H}$ are such that $f(z_1) = f(z_2)$ and $f'(z_1) = 0$, then also $f'(z_2) = 0$.*

The remainder of this section is devoted to the proof of Lemma 2.4. We start with three lemmas.

Lemma 2.5.

- **a** *Choose any $n \in \mathcal{N}_K$. If $\mathrm{Im}\,(z_0) > nt$, then all roots y of $F_n(f(z_0), y)$ will be simple roots.*
- **b** *Choose any $n \in \mathcal{N}_K$ and $\beta \in A(n)$. If $f'(z_0) = 0$ then either $f'(\beta(z_0)) = 0$, or $f(\beta(z_0))$ is a multiple root of $F_n(f(z_0), y)$.*
- **c** *Choose any $n \in \mathcal{N}_K$ and $\beta \in A(n)$. If $f(\beta(z_0))$ is a multiple root of $F_n(f(z_0), y)$, then either $f'(z_0) = 0$, or $f(z_0)$ is a multiple root of $F_n(f(\beta(z_0)), y)$.*

Proof: Part **a** is immediate from property **MI.3′** of F_n and Lemma 2.2. For any $n \in \mathcal{N}_K$, and any $\beta = \left\langle \begin{matrix} n/d & r \\ 0 & d \end{matrix} \right\rangle \in A(n)$, differentiating $F_n(f(z), f(\beta(z))) = 0$ gives:

$$\frac{\partial F_n}{\partial x}\big(f(z_0), f(\beta(z_0))\big)\, f'(z_0) + \frac{\partial F_n}{\partial y}\big(f(z_0), f(\beta(z_0))\big)\, \frac{n}{d^2} f'(\beta(z_0)) = 0. \qquad (2.1)$$

Part **b** follows from equation (2.1), while **c** follows from equation (2.1) and property **MI.1** of F_n. ∎

Lemma 2.6. *If $f'(z_0) \neq 0$, then for all sufficiently large primes $p \in \mathcal{N}_K$ and for all $\ell \in \mathbb{Z}^{>0}$, $f(p^\ell z_0)$ is a simple root of $F_{p^\ell}(f(z_0), y)$.*

Proof: Choose any prime $p \in \mathcal{N}_K$ and $\ell \in \mathbb{Z}^{>0}$ such that $f(p^\ell z_0)$ is a multiple root of $F_{p^\ell}(f(z_0), y)$: $\frac{\partial F_{p^\ell}}{\partial y}(f(z_0), f(p^\ell z_0)) = 0$. By Lemma 2.5 **c**, we see that $f(z_0)$ is a multiple root of $F_{p^\ell}(f(p^\ell z_0), y)$. This means that $f(z_0) = f(\beta(p^\ell z_0))$ for some $\beta \in A(p^\ell)$, $\beta(p^\ell z_0) \neq z_0$.

If there were infinitely many such primes p, then we would have a sequence $w_i = p_i^{2m_i} z_0 + s_i$, $0 \leq s_i < 1$, $i = 1, 2, \ldots$ of distinct numbers w_i at which $f(w_i) = f(z_0)$. By Lemma 2.2 all but finitely many of these w_i have $m_i = 0$, and so the w_i lie in a bounded region bounded away from the real axis. The analyticity of f would then force $f(z) = f(z_0)$ everywhere, a contradiction. ∎

Lemma 2.7. *Suppose $z_1, z_2 \in \mathcal{H}$, $f(z_1) = f(z_2)$, $f'(z_1) = 0$ and $f'(z_2) \neq 0$. Then there exists an NN such that for all primes $p > NN$, $p \in \mathcal{N}_K$, and any $\ell \in \mathbb{Z}^{>0}$, there is some $z_{p^\ell} = \beta(z_1)$, where $\beta \in A(p^m)$, for which both $f(z_{p^\ell}) = f(p^\ell z_2)$ and $f'(z_{p^\ell}) = 0$.*

Proof: Take NN sufficiently large so that for all primes $p > NN$, $f(p^\ell z_2)$ is a simple root of $F_{p^\ell}(f(z_2), y) = F_{p^\ell}(f(z_1), y)$. That such an NN exists follows from Lemma 2.6. Then $f(p^\ell z_2) = f(z_{p^\ell})$ for some $z_{p^\ell} = \beta(z_1)$ for some $\beta \in A(p^\ell)$ and hence $f(z_{p^\ell})$ must be a simple root of $F_{p^\ell}(f(z_1), y)$ and Lemma 2.5 **b** implies $f'(z_{p^\ell}) = 0$.

∎

Proof of Lemma 2.4: Assume that $f'(z_2) \neq 0$, we shall show that this leads to a contradiction.

First note that if $p \in \mathcal{N}_K$ is a prime such that $p > NN$ and $\mathrm{Im}\,(pz_2) > t$, then by Lemma 2.7 z_p and pz_2 satisfy $f(z_p) = f(pz_2)$, $f'(z_p) = 0$ and, since pz_2 is in the injective region, $f'(pz_2) \neq 0$. Thus, by redefining z_1 to be z_p and z_2 to be pz_2 if necessary, we may assume in our search for a contradiction that $\mathrm{Im}\,(z_2) > t$.

With notation as in Lemma 2.7, choose any primes $p, p' \in \mathcal{N}_K$, $p, p' > NN$. We may find increasing sequences $\ell_i, \ell_i' \in \mathbb{Z}^{>0}$ such that $1 < p^{\ell_i}/p'^{\ell_i'} < 2$ (see Lemma 4.2 below). Let $n_i = p^{\ell_i}$, $m_j = p'^{\ell_j'}$. By Lemma 2.5 **a** each root of $F_{m_i}(f(z_{n_i}), y) = F_{m_i}(f(n_i z_2), y)$ will be simple. Then by Lemma 2.5 **b**, $f'(m_i z_{n_i}) = 0$. Thus we obtain an infinite number of distinct points $z_i' = m_i z_{n_i}$, such that $f'(z_i') = 0$. Note that $\mathrm{Im}(z_1/2) < \mathrm{Im}(z_i')$ and the fact that $f'(z_i') = 0$ means that $\mathrm{Im}(z_i') \leq t$. Thus, translating by integers if necessary, all the z_i' lie in a compact region in which f is analytic. This forces f to be constant, which is impossible. ∎

§3. Existence of a unique maximal domain.

We know from Proposition 2.3 that if $f(z_1) = f(z_2)$, then there exists a local symmetry $\alpha : D_1 \to D_2$ with $z_i \in D_i$ and $\alpha(z_1) = z_2$. Our goal in this section is to show this α has a unique maximal domain to which it can be analytically continued. Throughout this section let $\sigma, \sigma' : [0,1] \to \mathcal{H}$ be any two continuous maps which are injective except possibly at finitely many points of $[0,1]$.

Definition 3.1. *Write $\sigma \approx \sigma'$ if for each $s \in [0,1]$ there exists a local symmetry α_s, defined in an open disc D_s about $\sigma(s)$, such that $\alpha_s \circ \sigma = \sigma'$ in some neighbourhood V_s of s.*

Note that $\approx$ is an equivalence relation. Also the equation $\alpha_s \circ \sigma = \sigma'$ implies that $\alpha_s(z) = \alpha_{s'}(z)$ for all $s' \in V_s$, and for all z in $D_s \cap D_{s'}$.

Lemma 3.2.

a *If $\sigma \approx \sigma'$, $n \in \mathcal{N}_K$ and $\beta \in A(n)$, then there exists a $\beta' \in A(n)$ such that $\beta \circ \sigma \approx \beta' \circ \sigma'$.*

b *If $\alpha : D_1 \to D_2$ is a local symmetry, $n \in \mathcal{N}_K$ and $\beta \in A(n)$, then there exists a $\beta' \in A(n)$ such that $\beta' \circ \alpha \circ \beta^{-1} : \beta(D_1) \to \beta'(D_2)$ is also a local symmetry.*

c *Let $\alpha : D_1 \to D_2$ be a local symmetry which is not a translation. Then for all sufficiently large primes $p \in \mathcal{N}_K$, there exists an $\beta = \left\langle \begin{matrix} 1 & r \\ 0 & p \end{matrix} \right\rangle \in A(p)$ such that $\beta \circ \alpha$ is a local symmetry on pD_1.*

Proof: First note that the points $z \in \mathcal{H}$ at which $F_n(f(z), y)$ has multiple roots – say $z \in Z_n$ – are isolated. To see this, note that $y = f(\beta(z))$ will be a multiple root if and only if

$$\frac{\partial F_n}{\partial y}\big(f(z), f(\beta(z))\big) = 0. \tag{3.1}$$

The left side of equation (3.1) is an analytic function of z. It cannot be identically zero by Lemma 2.2 (take $\mathrm{Im}\,(z) > nt$); therefore its zeros must be isolated. Taking the union over all $\beta \in A(n)$ of the solutions z of equation (3.1), we find that indeed the points of Z_n are isolated.

To prove **a**, choose any $s \in [0,1]$, and pick any $z \in D_s$, $z \notin Z_n$. Then the equation

$$f(\beta(z)) = f(\beta' \circ \alpha_s(z)) \tag{3.2}$$

uniquely determines some $\beta' \in A(n)$. By continuity, equation (3.2) continues to hold for the same β', for all other $z \in D_s$, $z \notin Z_n$; thus again by continuity and the fact the points in Z_n are isolated, it holds for the same β', for all $z \in D_s$. That equation (3.2) holds for all s now follows from uniqueness of β'. Part **a** is now immediate (the local symmetry is $\beta' \circ \alpha_s \circ \beta^{-1}$).

The proof of **b** is similar. Part **c** follows from **b** and Lemma 2.2. ∎

Lemma 3.3. *If $\sigma \approx \sigma'$ and σ is closed (i.e. $\sigma(0) = \sigma(1)$), then so is σ'.*

Proof: Choose any prime $p \in \mathcal{N}_K$ for which $p\sigma'(0)$ and $p\sigma'(1)$ both lie in $\mathcal{H}_t$. By Lemma 3.2 **a** there exists a $\beta \in A(p)$ such that $\beta \circ \sigma \approx p\sigma'$. Then $f(p\sigma'(0)) = f(\beta \circ \sigma(0)) = f(\beta \circ \sigma(1)) = f(p\sigma'(1))$; by Lemma 2.1 we must have $\sigma'(0) - \sigma'(1) \in \frac{1}{p}\mathbb{Z}$. Choosing a different prime p then forces $\sigma'(0) - \sigma'(1)$ to be an integer, n say.

Now choose any $\ell, m \in \mathcal{N}_K$ for which $\frac{\ell}{m}\sigma' \subset \mathcal{H}_t$. Again by Lemma 3.2 **a**, there exists some $\beta \in A(\ell m)$ such that $\beta \circ \sigma \approx \frac{\ell}{m}\sigma'$. This means, from the above paragraph, that $n\frac{\ell}{m} \in \mathbb{Z}$ which, varying ℓ and m, forces $n = 0$. ∎

Proposition 3.4. *Every local symmetry has a unique maximal domain D.*

Proof: Let $\alpha : D_1 \to D_2$ be a local symmetry of f. Choose a point $z_0 \in D_1$ and let D be the set of all points $z \in \mathcal{H}$ such that there exists a curve $\sigma([0,1])$ from z_0 to z along which α can be analytically continued. Clearly D is path connected and open and hence connected. By Lemma 3.3 α extends to a well-defined analytic function on D. If D' is any other domain to which α can be analytically continued, then by construction all the points of D' are contained in D and hence D is the unique maximal domain for α. ∎

§4. Local symmetries are global symmetries.

In this section we show that the local symmetry of Proposition 2.3 extends to an automorphism of $\mathcal{H}$.

Lemma 4.1. *Let D_1 and D_2 be the maximal domains for local symmetries α_1 and α_2 respectively. If there exists a nonempty open set $E \subset D_1 \cap D_2$ such that $\alpha_1(E) \subset \mathcal{H}_t$ and $\alpha_2(E) \subset \mathcal{H}_t$, then $D_1 = D_2$.*

Proof: We have a map $\alpha_2 \circ \alpha_1^{-1} : \alpha_1(E) \longrightarrow \alpha_2(E)$. This is a local symmetry of f and so, since both $\alpha_1(E)$ and $\alpha_2(E)$ are in the injective region, it must be a translation, θ say. Set $\alpha_3 = \theta \circ \alpha_1$. Then D_1 is also a maximal domain for α_3 and, moreover, the analytic functions α_2 and α_3 are equal on the open set E. Proposition 3.4 now forces $D_1 = D_2$. ∎

Lemma 4.2. *[HW, Theorem 438] If $s \in \mathbb{R}^{>0}$ is irrational, then the set $\{m - ns \mid m, n \in \mathbb{Z}^{>0}\}$ is dense in $\mathbb{R}$.*

Proposition 4.3. *Let $\alpha : B \longrightarrow A$ be a local symmetry which is not a translation, with B the maximal domain of α. Then B is invariant under two transformations $z \mapsto a_i z + b_i$, $i = 1, 2$ where a_1 and a_2 are positive and not equal to 1 and $a_1 \neq a_2^r$ for any rational r.*

Proof: Choose $B' \subseteq B$ to be an open disc of radius η and centre $z_0 = x_0 + iy_0$ and let $A' = \alpha(B')$. Choose any primes $p, q \in \mathcal{N}_K$. By Lemma 4.2 there exist strictly increasing sequences $\ell_i, m_i \in \mathbb{Z}^{>0}$, $i = 1, 2, \ldots$, for which $1 < p^{\ell_i}/q^{m_i} < 2$. Put $p_i = p^w p^{\ell_i}$ and $q_i = q^{m_i}$ where w is an integer chosen sufficiently large so that $p^w A' \subset \mathcal{H}_t$. By Lemma 3.2 **b** there are local symmetries between $p_i A'$ and $B_i' = \beta_i(B')$ for some $\beta_i \in A(p_i)$. Similarly there exist local symmetries between $B_i = q_i B_i'$ and suitable images A_i of $p_i A'$. By the above inequalities all the A_i are contained in $\mathcal{H}_t$. If $B_i \cap \mathcal{H}_t \neq \emptyset$, then the local symmetry between A_i and B_i must be a translation, which implies that α is also a translation: a contradiction. So $B_i \cap \mathcal{H}_t = \emptyset$. Let $z_i = x_i + iy_i$ and η_i be the centre and radius of B_i. Since $q_i/p_i > 1/(2p^w)$ we have $y_i > y_0/(2p^w)$ and, since $B_i \cap \mathcal{H}_t = \emptyset$, $y_i < t$. Thus, translating by integers if necessary, all the z_i will lie in a compact subset of $\mathcal{H}$. In addition $\eta_i > \eta/(2p^w)$ and so by computing areas, we see that intersections must occur: say $B_i \cap (B_j + s)$ is nonempty for some integer s. By Lemma 4.1, the maximal domains corresponding to B_i and $B_j + s$ must be equal. This gives us some transformation $a_1 z + b_1$ which must fix B. Since the sequence m_i is strictly increasing, a straightforward calculation shows that $a_1 = p^\sigma q^\tau$ with $\tau \neq 0$. To obtain the second transformation $a_2 z + b_2$ with $a_1 \neq a_2^r$ for any rational r, repeat the construction with two different primes. ∎

Lemma 4.4. *Let a_1, a_2, b_1, b_2 be rational numbers such that:*

1. *a_1 and a_2 are positive and not equal to 1.*
2. *$a_1 \neq a_2^r$ for any rational r.*
3. *$c = b_1/(1 - a_1) - b_2/(1 - a_2) \neq 0$.*

Let G be the subgroup of $GL(2, \mathbb{R})$ generated by $g_1 = \begin{pmatrix} a_1 & b_1 \\ 0 & 1 \end{pmatrix}$ and $g_2 = \begin{pmatrix} a_2 & b_2 \\ 0 & 1 \end{pmatrix}$, then the closure of G in $GL(2, \mathbb{R})$ is the subgroup $\{\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mid a, b \in \mathbb{R}, a > 0\}$.

Proof: A easy calculation shows that

$$(g_1^{-n} g_2^{-m} g_1^n g_2^m)^s = \begin{pmatrix} 1 & sc(1 - a_1^{-n})(1 - a_2^{-m}) \\ 0 & 1 \end{pmatrix}.$$

Taking the limit $m \to \pm\infty$ as appropriate, we have that the closure of G contains $\begin{pmatrix} 1 & sc(1-a_1^n) \\ 0 & 1 \end{pmatrix}$ and hence $\begin{pmatrix} 1 & sc(a_1^{n_1} - a_2^{n_2}) \\ 0 & 1 \end{pmatrix}$ for all $n_1, n_2, s \in \mathbb{Z}$. By a suitable choice of n_1 and n_2 we can make $c(a_1^{n_1} - a_2^{n_2})$ arbitrarily small and nonzero. Since s is any integer the closure of the set of all these translations gives all real translations.

Next, note that by conditions **1** and **2** we may apply Lemma 4.2 to $\log(a_1)$ and $\log(a_2)$ to conclude that the set $\{a_1^n a_2^m \mid n, m \in \mathbb{Z}\}$ is dense in the positive reals. Now $g_1^n g_2^m = \begin{pmatrix} a_1^n a_2^m & d \\ 0 & 1 \end{pmatrix}$ for some d. Thus the closure of G contains $\begin{pmatrix} 1 & -d \\ 0 & 1 \end{pmatrix} g_1^n g_2^m = \begin{pmatrix} a_1^n a_2^m & 0 \\ 0 & 1 \end{pmatrix}$ and so the closure of G contains all dilations by positive reals.

As $\{\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mid a, b \in \mathbb{R}, a > 0\}$ is generated by real translations and dilations by positive reals the result follows. ∎

Corollary 4.5. *The maximal domain B of a local symmetry α is either the upper half plane or the interior of a region bounded by two lines which meet at a common point on the real axis.*

Proof: Let $z \mapsto a_i z + b_i$, $i = 1, 2$ be the two transformations produced by Proposition 4.3. If $b_1/(1-a_1) \neq b_2/(1-a_2)$, then by Lemma 4.4 these two transformations generate a group that is dense in a group that acts transitively on the upper half plane. Hence B is the upper half plane and the result follows. If $b_1/(1-a_1) = b_2/(1-a_2)$, then the two transformations generate a subgroup of the group Θ of transformations $z \mapsto wz + (1-w)c$, $w \in \mathbb{R}^{>0}$, $c = b_1/(1-a_1)$. In fact this subgroup is dense by the same argument as used in Lemma 4.4. Thus if z_0 is a point in B, then B also contains the Θ-orbit of z_0. As the orbits of Θ are rays with tail at the point c on the real axis the result follows. ∎

Proposition 4.6. *Any local symmetry of f extends to a global symmetry.*

Proof: Let α be a local symmetry of f. If α is a translation then we are done, so assume otherwise. Clearly α^{-1} is also a local symmetry of f. Let B_1 be the maximal domain of α and B_2 be the maximal domain for α^{-1}. Applying Lemma 3.3 to α^{-1} shows that there are no points of B_1 at which α is not invertible. Thus $\alpha(B_1) \subset B_2$ and reversing the roles of α and α^{-1} gives $\alpha^{-1}(B_2) \subset B_1$ and hence $\alpha(B_1) = B_2$. By Corollary 4.5 there are four cases to consider. If neither B_1 nor B_2 is $\mathcal{H}$ then note first that we can construct an explicit isomorphism $\gamma_1 : B_1 \to \mathcal{H}$ given by $\gamma_1(z) = (a_1(z - a_2))^{a_3}$ for suitable choices of the real parameters a_2 and a_3 and a complex parameter a_1. Similarly there is an isomorphism $\gamma_2 : B_2 \to \mathcal{H}$. Thus $\alpha = \gamma_2^{-1} \circ M \circ \gamma_1$ for some $M \in SL(2, \mathbb{R})$. We can thus find a continuous curve $\sigma : [0, 1) \to B_1$ such that $\lim_{t \to 1} \sigma(t) = z_0 \in \partial B_1 \cap \mathcal{H}$ and $\sigma' = \alpha \circ \sigma$ such that $\sigma' : [0, 1) \to B_2 \cap \mathcal{H}$. Let $w_0 = \lim_{t \to 1} \sigma'(t)$. By continuity $f(w_0) = f(z_0)$ and so there is a local symmetry from a neighbourhood of z_0 to a neighbourhood of w_0. This however is a contradiction since it gives an analytic continuation of α across the boundary of B_1 which was assumed to be the maximal domain of α.

Consider next the case where $B_1 \neq \mathcal{H}$ and $B_2 = \mathcal{H}$. As above we know that $\alpha = M \circ \gamma_1$. Consider an open disc in $\mathcal{H}$, bounded away from the real axis such that its centre lies on one of the boundary lines of B_1. Let D_1 be the intersection of this disc with B_1. Let $\alpha(D_1) = D_2$. Under the map γ_1, D_1 maps to a simply connected region which has a

bounded interval of the real axis as part of its boundary. Under M this interval must remain bounded, since if not D_2 would contain points arbitrarily close to $i\infty$ and so f would be unbounded on D_2. This is a contradiction since f is bounded on D_1 and α is a local symmetry of f.

Thus we have $\alpha : D_1 \to D_2$ with D_1 bounded away from the real axis and D_2 such that part of the boundary of D_2 is an interval of the real line. Now by Lemma 3.2 **c** for all sufficiently large primes $p \in \mathcal{N}_K$ we have a local symmetry $\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \circ \alpha \circ \beta^{-1} : \beta(D_1) \to pD_2$, where $\beta = \begin{pmatrix} 1 & k \\ 0 & p \end{pmatrix}$ for some $0 \leq k < p$. Thus taking p sufficiently large and applying a translation if necessary, we can find a local symmetry $\alpha' : D_1' \to D_2'$ with $D_1' \subset D_2'$, $D_1' \neq D_2'$ and D_1' bounded and bounded away from the real axis. Let w_0 be a point of D_2' that is not in D_1'. The points $\alpha'^{-k}(w_0)$, $k \in \mathbb{Z}^{>0}$ are distinct, are in D_1' and f has the same value at each point: a contradiction as f is analytic on $\mathcal{H}$ and not constant. The case $B_1 = \mathcal{H}$ and $B_2 \neq \mathcal{H}$ is similar.

Thus the only possible case is $B = B' = \mathcal{H}$ and so α is an automorphism of $\mathcal{H}$. ∎

Recall that the group of analytic automorphisms of $\mathcal{H}$ is $PSL(2,\mathbb{R})$. Our goal in the next section is to show that the subgroup of $PSL(2,\mathbb{R})$ generated by the symmetries of f is a congruence group.

§5. Modular invariance.

In this section 1_2 will denote the identity 2×2 matrix and for any subgroup G of $SL(2,\mathbb{R})$ G_∞ will denote the subgroup of G that stabilizes $i\infty$. We shall obtain the following:

Proposition 5.1. *Let G be a subgroup of $SL(2,\mathbb{R})$ and K be a positive integer. Suppose*

1 *G is a discrete subgroup.*

2 $G_\infty = \langle -1_2, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$.

3 *For all $n \in \mathcal{N}_K$ and all $m = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$, there exist integers ℓ and k such that $\ell \mid n$, $0 \leq -k < n/\ell$ and such that*

$$\begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \ell & -k \\ 0 & n/\ell \end{pmatrix}^{-1} = \begin{pmatrix} na/\ell & ka + \ell b \\ c/\ell & (kc + \ell d)/n \end{pmatrix} \tag{5.1}$$

is in G.

Then either $G = \langle -1_2, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$ or G contains $\pm\Gamma_0(N)$ where $N \mid K^\infty$.

We start with some lemmas. For $m \in SL(2,\mathbb{R})$ define c_m to be the lower left entry of m. Note that for any discrete subgroup Γ of $SL(2,\mathbb{R})$, $|c_m|$ is an invariant of the double coset $\Gamma_\infty m \Gamma_\infty$.

Lemma 5.2. *[Sh, p.11] Let Γ be a discrete subgroup of $SL(2,\mathbb{R})$. Given $M > 0$, there are only finitely many double cosets $\Gamma_\infty m \Gamma_\infty$ with $m \in \Gamma$ and $|c_m| \leq M$.*

Lemma 5.3. *[Sh, p.11]*

a *Let Γ be a discrete subgroup of $SL(2,\mathbb{R})$. There exists an $r > 0$, depending only on Γ, such that if $m \in \Gamma - \Gamma_\infty$ then $|c_m| \geq r$.*

b *For any r satisfying the condition of* **a** *let $U = \{z \in \mathcal{H} \mid \mathrm{Im}(z) > 1/r\}$. Then $\Gamma_\infty = \{m \in \Gamma \mid m(U) \cap U \neq \emptyset\}$.*

We shall not require Lemma 5.3 **b** until Section 8.

Lemma 5.4. *If G is as in Proposition 5.1 and $m \in G$ then $m = \lambda m'$ where $m' \in GL(2,\mathbb{Q})^+$ (the superscript denotes positive determinants).*

Proof: Let $m = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $c = 0$ then the result follows from property **2** of G. Otherwise define $r_m = a/c \pmod 1$. As $G_\infty = \langle -1_2, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$, r_m is an invariant of the double coset $G_\infty m G_\infty$. Also, by property **3** of G and Lemma 5.3 **a**, for all but finitely many primes $p \in \mathcal{N}_K$, G contains $m(p) = \begin{pmatrix} pa & * \\ c & * \end{pmatrix}$, with $r_{m(p)} = p r_m$ and $c_{m(p)} = c_m$. If a/c is irrational, then for distinct primes $p_1, p_2 \in \mathcal{N}_K$, we have $p_1 a/c \neq p_2 a/c \pmod 1$. So we obtain infinitely many double cosets $G_\infty m(p) G_\infty$ with the same value of $c_{m(p)}$, which is a contradiction by property **1** of G and Lemma 5.2. Thus $m = \begin{pmatrix} \lambda a' & b \\ \lambda c' & d \end{pmatrix}$ with a' and c' rational. Applying the same argument to m^{-1} gives $m = \begin{pmatrix} \lambda a' & b \\ \lambda c' & \lambda d' \end{pmatrix}$ with d' rational. Finally applying the argument to m^2 (or $\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} m \right)^2$ if $a' + d' = 0$) we get $m = \lambda m'$ with $m' \in GL(2,\mathbb{Q})^+$ as required. ∎

Definition 5.5. *Call an integer matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ primitive if $\gcd(a,b,c,d) = 1$. Let M^* be the set of primitive integer 2×2 matrices. For $m \in G$, by Lemma 5.4 we can write $m = \lambda \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M^*$. Define $|m| = ad - bc$. This is well-defined, since the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is unique up to multiplication by -1.*

Lemma 5.6. *$|m|$ is an invariant of $G_\infty m G_\infty$.*

Proof: Clear. ∎

In the rest of this section we use the notation $a \| b$ if a and b are integers such that a divides b and a is coprime to b/a.

Lemma 5.7. *Let $m_1 \in G$ with $\lambda^{-1} m_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$ an integer matrix with $c_1 \neq 0$. Choose any $n \in \mathcal{N}_K$ and let $m_i = \lambda \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}$ $i \geq 2$ be a sequence in G obtained*

recursively by applying property 3 of Proposition 5.1 to m_{i-1}. *Let* ℓ_i, k_i *be the relevant parameters in (5.1). Let* $c_\infty = c_1/\prod_{i=1}^{\infty} \ell_i$, *then*

- **a** *If* $c_\infty \in \mathbb{Z}$ *then for each* $i \geq 1$ d_i *is an integer;*
- **b** *If* $c_\infty \in \mathbb{Z}$ *then* $c_\infty \neq 0$ *and if* p *divides* $gcd(n, c_\infty)$ *then* p *divides* d_i *for all* $i \geq 1$;
- **c** *There exists a* W, *depending only on* λ *and* c_1, *such that if all prime divisors of* n *are larger than* W, *then* $\ell_i = 1$ *and* $d_i \in \mathbb{Z}$ *for all* $i \geq 1$.

Proof: Part **a**: by Lemma 5.3 **a**, $\ell_i = 1$ for all but finitely i and so, since $c_\infty \in \mathbb{Z}$, ℓ_i will divide c_i for all $i \geq 1$. Suppose $d_{i_0} \notin \mathbb{Z}$ for some i_0. Then, since $d_{i+1} = (k_i c_i + \ell_i d_i)/n$ and ℓ_i divides both n and c_i, it follows that $d_i \notin \mathbb{Z}$ for all $i > i_0$. Thus we can choose i_0 so that $d_{i_0} \notin \mathbb{Z}$ and $\ell_i = 1$ for all $i \geq i_0$.

By construction $d_i = x_i/y_i$ where x_i and y_i are coprime integers and $y_i \,|\, n^\infty$. Since by hypothesis d_{i_0} is not an integer, there is a prime p such that $p \,|\, y_{i_0}$. Let $p^\eta \,\|\, n$ and $p^s \,\|\, y_{i_0}$. For $i \geq i_0$ we have $d_{i+1} = (k_i c_i + d_i)/n$ and so $p^{(i-i_0)\eta+s} \,\|\, y_i$. Let $p^{s'} \,\|\, \det(m_1)$. To find the exact power of p which divides $|m_i|$ note that $\det(m_i) = \det(m_1)$ and that a_i, b_i and c_i are integers for all $i \geq 1$ and so $p^{2((j-i)\eta+s)+s'} \,\|\, |m_i|$ for all $i \geq i_0$. It then follows from Lemma 5.6 that $G_\infty m_i G_\infty$, $i \geq i_0$ are infinitely many distinct double cosets, but $c_{m_i} = c_{m_{i_0}}$ which contradicts Lemma 5.2. This establishes **a**.

To prove **b**, as noted above $\ell_i = 1$ for all but finitely many i and as $c_1 \neq 0$ we must have $c_\infty \neq 0$ as required. If $p \,|\, \gcd(c_\infty, n)$ then $p \,|\, c_i$ for all i. If there is some i_0 for which $p \nmid d_{i_0}$ then since $d_{i+1}(n/\ell_i) = k_i c_{i+1} + d_i$ it follows that $p \nmid d_i$ for all $i > i_0$. However for i sufficiently large $d_{i+1} n = k_i c_\infty + d_i$ so that $p \,|\, d_i$: a contradiction and so $p \,|\, d_i$ for all $i \geq 1$.

For part **c** take $W > \lambda c_1/r$. If $\ell_i \neq 1$ for some i then there is some prime $p \,|\, \ell_i$ and the condition on W implies that $\lambda c_1/\ell_i \leq \lambda c_1/p < r$: a contradiction by Lemma 5.3 **a**. That d_i is an integer for all $i \geq 1$ now follows from part **a** since $c_\infty = c_1$ is an integer. ∎

Proof of Proposition 5.1: If all $m \in G$ have $c_m = 0$ then $G = \langle -1_2, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$. Otherwise G contains

$$m' \begin{pmatrix} 1 & -|m'| \\ 0 & 1 \end{pmatrix} m'^{-1} = \begin{pmatrix} 1 + a'c' & -a'^2 \\ c'^2 & 1 - a'c' \end{pmatrix} \in G \cap SL(2, \mathbb{Z}) \tag{5.2}$$

and $c' \neq 0$.

Now consider any prime p which divides c' but not K. Choose any $\eta \geq 1$ so that $p^\eta = 1 \pmod K$, and take $n = p^\eta$. Let notation be as in Lemma 5.7, then $\prod_i \ell_i = p^j$ for some j and let $p^{j'} \,\|\, c'^2$. If $j' > j$ then c_∞ is an integer divisible by p and as $d = 1 \pmod p$ Lemma 5.7 **b** yields a contradiction. Hence we have $j \geq j'$. Moreover if we replace λ by $\lambda p^{j'-j}$ and make the appropriate rescaling of the entries of $m' = m_1$, then the hypotheses of Lemma 5.7 **a** are satisfied and so $p^{j-j'} d_i \in \mathbb{Z}$ for all i. As noted in the proof of Lemma 5.7 there is some i_0 such that $c_{i_0} = c_\infty$. If we construct m_{i_0} and apply the transformation of equation (5.2) then from the above comments the result is an integer matrix for which p does not divide the lower left entry. Now iterate this process until all such primes are exhausted. Finally applying the transformation of equation (5.2) and take the Kth power of the resulting matrix. This is equivalent to applying the transformation of equation (5.2)

with $|m'|$ replaced by $K|m'|$ and yields a matrix $m = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G \cap SL(2,\mathbb{Z})$ such that $K \mid c$, $c \neq 0$ and $d = 1 \pmod K$.

Let $W(m)$ the constant of Lemma 5.7 **c**. As c and d are coprime, by considering $m \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ for a suitable choice of k we find that G contains $g = \begin{pmatrix} a & b'' \\ c & d'' \end{pmatrix}$ with $d'' = 1 \pmod K$, d'' positive and divisible only by primes larger than $W(m)$. As m and g both have $\lambda = 1$ and the same bottom left entry we have $W(g) = W(m)$ and so applying Lemma 5.7 **c** to g with $n = d''$ and property **3** of Proposition 5.1 we find that G contains $\begin{pmatrix} ad'' & b'' \\ c & 1 \end{pmatrix}$. All operations used preserve determinants, hence, premultiplying by a suitable translation, we see that G contains $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$.

Consider any nonzero $a, b \in \mathbb{Z}$ with $\gcd(a, bc) = 1$ and hence $\gcd(a, K) = 1$ since $K \mid c$. Choose any prime $q > bc/r$, with r as in Lemma 5.3 **a** such that $aq = 1 \pmod K$, and any prime $p > qbc/r$ such that $p = a \pmod{bc}$. G contains $\begin{pmatrix} 1 & 0 \\ qbc & 1 \end{pmatrix}$ so, postmultiplying by a suitable translation, we know G contains an element of the form $\begin{pmatrix} 1 & * \\ qbc & p* \end{pmatrix}$ with determinant equal to 1. By our choice of primes p, q, Lemma 5.7 **a** with $n = pq$ implies in equation (5.1) we must have $\ell = q$ and $k = 0$, so G contains an element of the form $\begin{pmatrix} p & * \\ bc & * \end{pmatrix}$ and hence also $\begin{pmatrix} a & * \\ bc & * \end{pmatrix}$, both with determinant equal to 1. As these matrices are a complete set of coset representatives for the subgroup $\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$ in $\Gamma_0(c)$ the result follows with $N = c$. ∎

§6. Generalized modular equations and Hauptmoduln.

The aim of this section is to prove Theorem 1.4. In outline the proof is similar to that for the j function, see for example [Knapp, p.335]. However, we have been unable to locate a discussion of this material in the literature and so we include a self-contained treatment here.

Let $\mathcal{F}_N$ be the field of automorphic functions of $\Gamma(N)$ with Fourier coefficients in $\mathbb{Q}(\zeta_N)$. For n coprime to N, $*n$ will denote the Galois automorphism of $\mathbb{Q}(\zeta_N)$ such that $\zeta_N * n = \zeta_N^n$. In Sections 6.1 to 6.5 of [Sh] it is shown that:

Theorem 6.0.

a *$\mathcal{F}_N = \mathbb{Q}(j, f_a \mid a \in (\mathbb{Z}/N\mathbb{Z})^2, a \neq (0,0))$, for functions $f_a(z)$, known as the Fricke functions, related to the the division points of order N of the elliptic curve $\mathbb{C}/(\mathbb{Z}+z\mathbb{Z})$.*

b *$\mathcal{F}_N$ is a Galois extension of $\mathbb{Q}(j) = \mathcal{F}_1$ with Galois group $GL(2,\mathbb{Z}/N\mathbb{Z})/\{\pm 1\}$. The action of $\alpha \in GL(2,\mathbb{Z}/N\mathbb{Z})$ is given by $f_a \mapsto f_{a\alpha}$. If $m \in SL(2,\mathbb{Z})$ then $f_a \circ m = f_{am}$.*

c *Let n be an integer coprime to N. Then $\begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \in GL(2,\mathbb{Z}/N\mathbb{Z})$ acts on $\mathcal{F}_N$ by $h \mapsto h * n$ where $h * n$ is obtained from h by applying $*n$ to its Fourier coefficients.*

d *Let $M_2(\mathbb{Z})$ be the set of 2×2 integer matrices. Let $U = \prod_p GL(2,\mathbb{Z}_p) \times GL(2,\mathbb{R})^+$, where the superscript denotes positive determinants. For every $u \in U$ and every N, there exists an element α of $M_2(\mathbb{Z}) \cap GL(2,\mathbb{Q})^+$ such that $u_p = \alpha \pmod{N \cdot \mathbb{Z}_p}$, where u_p is the pth component of u. Set $au = a\alpha$ for all $a \in (\mathbb{Z}/N\mathbb{Z})^2$. Then $f_a \mapsto f_{au}$ defines an element of $Gal(\mathcal{F}_N/\mathcal{F}_1)$, call it $\tau(u)$.*

Proof:

a See [Sh, Proposition 6.9 (1)] and [Sh, p.134].

b See [Sh, Proposition 6.6 (1) and (2)] and [Sh, equation (6.1.3)].

c This follows from the explicit form of the $q^{1/N}$-expansion of the f_a in [Sh, equation (6.2.1)] and parts **a** and **b**.

d See the proof of [Sh, Proposition 6.21] and the discussion in [Sh, section 6.4]. ∎

In the rest of this section f will be a nonconstant element of $\mathcal{F}_N$ and where no confusion can arise we shall write G for $G(f)$ and $G * n$ for $G(f * n)$. The initial results of this section hold without further restrictions on f. However later results require additional hypotheses. For example, in Proposition 6.8 we require that G contains $\Gamma_0(N)$ and that $G * n^2 = G$ for all n coprime to N. From Proposition 6.14 to the end of the section we shall require that f is a Hauptmodul of G and that G contains $\Gamma_0(N)$.

Lemma 6.1. *G is a discrete subgroup of $SL(2,\mathbb{R})$ which contains $\Gamma(N)$ with finite index.*

Proof: As $f \in \mathcal{F}_N$, $\Gamma(N)$ is a subgroup of G and the index must be finite, since if not f would give rise to a nonconstant function on $X(\Gamma(N))$ taking the same value at infinitely many distinct points: a contradiction. It then follows that G is discrete by [Sh, Proposition 1.11]. ∎

Lemma 6.2. *There is a group homomorphism $\phi : G \longrightarrow PGL(2,\mathbb{Q})^+$ with $\ker(\phi) = \{\pm 1_2\}$. $\overline{G} = \phi(G)$ is a discrete subgroup of $PGL(2,\mathbb{Q})^+$.*

Proof: As G is commensurable with $SL(2,\mathbb{Z})$, it follows ([Sh, Proposition 1.30]) that the set of cusps of G is the same as the set of cusps of $SL(2,\mathbb{Z})$, i.e. $\mathbb{Q} \cup \{i\infty\}$. Clearly if w is a cusp and m is an element of G then $m(w)$ is also a cusp. It is not difficult, for example by considering $m(i\infty), m(0), m^{-1}(i\infty)$ and $m^{-1}(0)$, to use this fact to show that any element of G is a multiple of a matrix with rational entries. We may define ϕ to be the corresponding map to $PGL(2,\mathbb{Q})^+$. The only elements of $SL(2,\mathbb{R})$ which are multiples of the identity are $\pm 1_2$, however by definition $-1_2 \in G(f)$ and so $\ker(\phi) = \{\pm 1_2\}$. As $\overline{G}$ contains $\overline{\Gamma_0(N)} = \phi(\Gamma_0(N))$ with finite index and $\Gamma_0(N)$ is a discrete subgroup of $PGL(2,\mathbb{Q})^+$, again by [Sh, Proposition 1.11] $\overline{G}$ is a discrete subgroup of $PGL(2,\mathbb{Q})^+$. ∎

Recall from section 2 that elements of $PGL(2,\mathbb{Q})^+$ by angular brackets and that if m is a nonsingular 2×2 integer matrix with positive determinant then $\langle m \rangle$ will denote the corresponding element of $PGL(2,\mathbb{Q})^+$. Note also that by Lemma 6.2 any $m \in G$ is a multiple of a primitive integer matrix. Recall the definition of $|m|$ given in 5.5.

Lemma 6.3. *[N2] If $\langle m \rangle \in \overline{G}$ then every prime p dividing $|m|$ also divides N.*

Lemma 6.4. *If $\gcd(n,N) = 1$ and $j = (\overline{G} : \overline{\Gamma(N)})$ then*

$$\overline{G} = \bigcup_{i=1}^{j} \left\langle \begin{matrix} a_i & nb_i \\ nc_i & d_i \end{matrix} \right\rangle \overline{\Gamma(N)} = \bigcup_{i=1}^{j} \overline{\Gamma(N)} \left\langle \begin{matrix} d_i & -nb_i \\ -nc_i & a_i \end{matrix} \right\rangle,$$

where $\begin{pmatrix} a_i & nb_i \\ nc_i & d_i \end{pmatrix}$, $1 \le i \le j$ *is a primitive integer matrix.*

Proof: Take $\langle B \rangle \in \overline{G}$ with B a primitive integer matrix, so that by Lemma 6.3 its determinant is coprime to n. Since reduction of $SL(2,\mathbb{Z})$ mod nN is surjective onto $SL(2,\mathbb{Z}/nN\mathbb{Z})$ we can find $m \in SL(2,\mathbb{Z})$ such that $m = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\mathrm{mod}\, N)$ and $m = B^{-1}\begin{pmatrix} \det(B) & 0 \\ 0 & 1 \end{pmatrix} (\mathrm{mod}\, n)$. Then $\langle Bm \rangle$ is in the same $\overline{\Gamma(N)}$ coset as $\langle B \rangle$ and Bm is diagonal mod n.

The second equality follows from the first and the observation that for any group G and subgroup H if $\{g_i\}$ is a complete set of left coset representatives of H in G, then $\{g_i^{-1}\}$ is complete set of right coset representatives. ∎

Theorem 6.5. *[Sh, p.147]*

a *For every* $\alpha \in GL(2,\mathbb{Q})^+$ *and every* $h \in \mathcal{F}_N$, $h \circ \alpha \in \mathcal{F}_{N'}$ *for some* N'.

b *If* $\alpha \in GL(2,\mathbb{Q})^+$, $\beta \in GL(2,\mathbb{Q})^+$, $u \in U$, $v \in U$ *and* $\alpha u = v\beta$ *then* $(j \circ \alpha)^{\tau(u)} = j \circ \beta$ *and* $(f_a \circ \alpha)^{\tau(u)} = f_{av} \circ \beta$.

Corollary 6.6.

a *If* $h \in \mathcal{F}_N$ *and* $\alpha \in GL(2,\mathbb{Q})^+ \cap M_2(\mathbb{Z})$ *then* $h \circ \alpha \in \mathcal{F}_{N\det(\alpha)}$.

b *If* G *contains* $\Gamma_0(N)$, *then for any* n *coprime to* N $G * n$ *contains* $\Gamma_0(N)$.

Proof: For part **a** consider first $f_a \in \mathcal{F}_N$. From Theorem 6.5 **a**, $f_a \circ \alpha$ is a modular function of some level. Moreover, since $\alpha\Gamma(N\det(\alpha))\alpha^{-1} \subset \Gamma(N)$, $f_a \circ \alpha$ is a modular function of level $N\det(\alpha)$. We may find $\gamma \in SL(2,\mathbb{Z})$ such that $\gamma^{-1}\alpha = \alpha'$ is upper triangular . Then by Theorem 6.0 **b**, $f_a \circ \alpha = f_a \circ (\gamma\alpha') = f_{a\gamma} \circ \alpha'$. Since $f_{a\gamma}$ is an element of $\mathcal{F}_N$, it has a Fourier expansion with respect to $\exp(2\pi iz/N)$ with coefficients in $\mathbb{Q}(\zeta_N)$, thus $f_{a\gamma} \circ \alpha'$ has a Fourier expansion with respect to $\exp(2\pi iz/N\det(\alpha))$ with coefficients in $\mathbb{Q}(\zeta_{N\det(\alpha)})$ and so it, and hence $f_a \circ \alpha$, is in $\mathcal{F}_{N\det(\alpha)}$. The corresponding result for $j(z)$ follows similarly and so the corollary follows by Theorem 6.0 **a**.

To show part **b** note first that for any $\gamma \in \Gamma_0(N)$ there exists a $\gamma' \in \Gamma_0(N)$ such that:

$$\begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \gamma = \gamma' \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \pmod N.$$

So for any $a \in (\mathbb{Z}/N\mathbb{Z})^2$, $a \neq (0,0)$, we have, by Theorem 6.0 **b** and **c**, $(f_a * n) \circ \gamma = (f_a \circ \gamma') * n$. As $\mathcal{F}_N$ is generated over $\mathbb{Q}(j)$ by these f_a, this relation holds for any element of $\mathcal{F}_N$, in particular $(f * n) \circ \gamma = (f \circ \gamma') * n = f * n$ since $\Gamma_0(N) \subset G$. ∎

The following result is given in [N2]. It is, however, a corollary of Shimura's more general result contained in Theorem 6.5 and we give a proof using this fact. Norton also give an outline for the proof of Proposition 6.14 which again we obtain from Shimura's Theorem.

Corollary 6.7. *Let* n *be coprime to* N *and let* $\alpha = \begin{pmatrix} a & b \\ nc & d \end{pmatrix}$ *and* $\beta = \begin{pmatrix} a & nb \\ c & d \end{pmatrix}$. *If* α *is a primitive integer matrix and* $\langle \alpha \rangle \in \overline{G}$, *then* $\langle \beta \rangle \in \overline{G * n}$. *Also if* β *is a primitive integer matrix and* $\langle \beta \rangle \in \overline{G * n}$, *then* $\langle \alpha \rangle \in \overline{G}$.

Proof: Define $u, v \in U$ by

$$u_p = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} & \text{for } p \mid N \\ \beta & \text{for } p \nmid N \text{ and } p = \infty \end{cases}$$

and

$$v_p = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} & \text{for } p \mid N \\ \alpha & \text{for } p \nmid N \text{ and } p = \infty \end{cases}$$

Here u and v are well-defined since $\det(\beta) = \det(\alpha)$ and by Lemma 6.3 $\det(\alpha)|N^\infty$. If $N'|N^\infty$ then since $u_p = v_p = \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \pmod{N' \cdot \mathbb{Z}_p}$ for all primes p we have, by Theorem 6.0 **c** and **d**, that if $h \in \mathcal{F}_{N'}$ then $h^{\tau(u)} = h^{\tau(v)} = h * n$. In particular for any $f_a \in \mathcal{F}_N$ we have $f_{av} \circ \beta = (f_a^{\tau(v)}) \circ \beta = (f_a * n) \circ \beta$ and by Lemma 6.3 and Corollary 6.6 **a** $(f_a \circ \alpha)^{\tau(u)} = (f_a \circ \alpha) * n$. Thus from Theorem 6.5 **b** we have $(f_a \circ \alpha) * n = (f_a * n) \circ \beta$. Similarly $(j \circ \alpha) * n = j \circ \beta$. So by Theorem 6.0 **a**, for any $h \in \mathcal{F}_N$ we have $(h \circ \alpha) * n = (h * n) \circ \beta$. Since $f \in \mathcal{F}_N$ and $\langle \alpha \rangle \in \overline{G(f)}$, we have $f * n = (f \circ \alpha) * n = (f * n) \circ \beta$ and so $\langle \beta \rangle \in \overline{G * n}$.

The proof of the second part of the Corollary is very similar to the first part. ∎

Let $A(n)$ be as in section 2. Note that $\operatorname{card}(A(n)) = \psi(n)$ where $\psi(n) = n \prod_{\substack{p \mid n \\ p \text{ prime}}} (1 + \frac{1}{p})$. Also fix $\alpha_n = \left\langle \begin{matrix} n & 0 \\ 0 & 1 \end{matrix} \right\rangle \in A(n)$.

Proposition 6.8. *Let* $T = \{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \in SL(2, \mathbb{R}) \mid t \in \mathbb{Z} \}$ *and* $\overline{T} = \phi(T)$. *Let* n *be any integer coprime to* N. *If* G *contains* $\Gamma_0(N)$, *then*

$$\bigcup_{\alpha \in A(n)} \overline{T} \alpha \overline{G * n^{-1}} = \bigcup_{\alpha \in A(n)} \overline{G} \alpha = \overline{\Gamma_0(N)} \alpha_n \overline{G * n^{-1}}.$$

Proof: We begin by showing

$$\bigcup_{\alpha \in A(n)} \overline{T} \alpha \overline{\Gamma_0(N)} = \bigcup_{\alpha \in A(n)} \overline{\Gamma_0(N)} \alpha = \overline{\Gamma_0(N)} \alpha_n \overline{\Gamma_0(N)}. \tag{6.1}$$

Define

$$M^*(n, N) = \{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M^* \mid c = 0 \pmod{N},\ ad - bc = n \}.$$

Let $\langle M^*(n, N) \rangle$ be the corresponding subset of $PGL(2, \mathbb{Q})^+$. By Lemma 9.14 of [Knapp], we get

$$\langle M^*(n, N) \rangle = \bigcup_{\alpha \in A(n)} \overline{\Gamma_0(N)} \alpha.$$

From this equality it is clear that $\bigcup_{\alpha\in A(n)} \overline{\Gamma_0(N)}\alpha^{-1}\overline{T} \subseteq \bigcup_{\alpha\in A(n)} \overline{\Gamma_0(N)}\alpha$. Moreover, the reverse inclusion also holds, since for each $\alpha \in A(n)$ we have $\alpha = \alpha'^{-1}t$ for some translation t and some $\alpha' \in A(n)$ and hence

$$\bigcup_{\alpha\in A(n)} \overline{\Gamma_0(N)}\alpha = \bigcup_{\alpha\in A(n)} \overline{\Gamma_0(N)}\alpha^{-1}\overline{T}.$$

The first equality of equation (6.1) now follows by taking inverses. The equality

$$\langle M^*(n,N)\rangle = \overline{\Gamma_0(N)}\alpha_n\overline{\Gamma_0(N)}$$

is established in [Sh, Proposition 3.32(1)], completing the proof of equation (6.1).

Now choose any $\alpha \in A(n)$ and $g \in \overline{G}$. By equation (6.1) we know $\alpha = h\alpha_n h'$ for some $h, h' \in \overline{\Gamma_0(N)}$. By Lemma 6.4 we may write $gh \in \overline{G}$ as $h''m_i$ where m_i equals some $\begin{pmatrix} a_i & nb_i \\ nc_i & d_i \end{pmatrix}$ as in Lemma 6.4 and $h'' \in \overline{\Gamma_0(N)}$. Now $m_i\alpha_n = \alpha_n m_i'$ where $m_i' = \begin{pmatrix} a_i & b_i \\ n^2c_i & d_i \end{pmatrix} \in \overline{G * n^{-1}}$ by Corollary 6.7. Thus we have shown that

$$\bigcup_{\alpha\in A(n)} \overline{G}\alpha \subset \overline{\Gamma_0(N)}\alpha_n\overline{G * n^{-1}},$$

where we absorbed the h' because by Corollary 6.6 **b** $\overline{\Gamma_0(N)} \subseteq \overline{G * n^{-1}}$. The reverse inclusion follows by identical arguments. Finally, multiplying equation (6.1) on the right by $\overline{G * n^{-1}}$, we get

$$\bigcup_{\alpha\in A(n)} \overline{T}\alpha\overline{G * n^{-1}} = \overline{\Gamma_0(N)}\alpha_n\overline{G * n^{-1}},$$

as required. ∎

The following result is contained in the proof of [N2, Theorem 3].

Corollary 6.9. *Let n be any integer coprime to N. Suppose G contains $\Gamma_0(N)$ and c is a cusp of $\overline{G}$. Then $\overline{G}(\alpha(c)) = \overline{G}(i\infty)$ for some $\alpha \in A(n)$ if and only if $\overline{G * n^{-1}}(c) = \overline{G * n^{-1}}(i\infty)$.*

Proof: If $\overline{G}(\alpha(c)) = \overline{G}(i\infty)$ then $g\alpha(c) = i\infty$ for some $g \in \overline{G}$ and so by Proposition 6.8 $t\alpha' g'(c) = i\infty$ for some $t \in \overline{T}$, for some $\alpha' \in A(n)$ and for some $g' \in \overline{G * n^{-1}}$. Since $t\alpha'$ fixes $i\infty$ this yields $g'(c) = i\infty$.

For the reverse implication, if $g'(c) = i\infty$ for some $g' \in \overline{G * n^{-1}}$ then $\alpha_n g'(c) = i\infty$ and so by Proposition 6.8 there is some $g \in \overline{G}$ and some $\alpha \in A(n)$ such that $g\alpha(c) = i\infty$ and so $\overline{G}(\alpha(c)) = \overline{G}(i\infty)$ as required. ∎

The following result is a small generalization of a result contained in the proof of [Sh, Proposition 6.9]. It will be used in the proof of Proposition 6.16.

Proposition 6.10. *Let G be a discrete subgroup of $SL(2,\mathbb{R})$ which contains $\Gamma(N)$ and let A_0 be the field of automorphic functions of G. If $A_0 = \mathbb{C}(f_1,\ldots,f_m)$ for some functions $f_1,\ldots,f_m \in \mathcal{F}_N$, then $A_0 \cap \mathcal{F}_N = \mathbb{Q}(\zeta_N, f_1,\ldots,f_m)$.*

Proof: The inclusion $\mathbb{Q}(\zeta_N, f_1, ..., f_m) \subset A_0 \cap \mathcal{F}_N$ is obvious. For the reverse inclusion, consider any $g \in A_0 \cap F_N$ and expand

$$g(q) = \sum_{i}^{\infty} a_i q^{i/N} = \sum_{i=1}^{m} c_i f_i(q) = \sum_{i=1}^{m}\sum_{j}^{\infty} c_i b_{i,j} q^{j/N},$$

where each $a_i, b_{i,j} \in \mathbb{Q}(\zeta_N)$, $c_i \in \mathbb{C}$. This equality yields a linear system for the c_i with coefficients in $\mathbb{Q}(\zeta_N)$ which has at least one solution in $\mathbb{C}$ and hence a solution in $\mathbb{Q}(\zeta_N)$.∎

Definition 6.11. *For any n coprime to N we define $w(n) \in GL(2,\mathbb{Z}/N\mathbb{Z})$ by $w(n) = \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix} \pmod N$. By Theorem 6.0* **b**, *$w(n)\{\pm 1\}$ is an element of the Galois group of $\mathcal{F}_N$ over $\mathbb{Q}(j)$ and we shall write its action on any $h \in \mathcal{F}_N$ as $h \mapsto h!n$.*

The next three results show that in the case that f is a Hauptmodul for G and n is coprime to N then $G * n^2 = G$.

Lemma 6.12. *If $h \in \mathcal{F}_N$, $m \in \overline{G(f)}$ and n is coprime to N then $(h \circ m)!n = (h!n) \circ m$.*

Proof: By Lemma 6.2, $m = \langle m' \rangle$, where m' is a primitive integer matrix. The proof is essentially identical to that of Corollary 6.7 with $\alpha = \beta = m'$ and $u = v$ where

$$v_p = \begin{cases} \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix} & \text{for } p|N, \\ \alpha & \text{for } p \not| N \text{ and } p = \infty. \end{cases}$$

∎

Lemma 6.13. *If $\overline{G}$ contains $\overline{\Gamma_0(N)}$ then $f!n = f * n^2$ for all n coprime to N.*

Proof: Any upper triangular matrix in $SL(2,\mathbb{Z}/N\mathbb{Z})$ is the image under reduction modulo N of some element of $\Gamma_0(N)$. So for any n coprime to N we can find $m(n) \in \Gamma_0(N)$ such that $m(n) = \begin{pmatrix} n^{-1} & 0 \\ 0 & n \end{pmatrix} \pmod N$. By Theorem 6.0 **b** and **c** we find $(f_a \circ m(n))!n = (f_{am(n)})!n = f_{am(n)w(n)} = f_a * n^2$. Hence, by Theorem 6.0 **a**, for any $h \in \mathcal{F}_N$, $(h \circ m(n))!n = h * n^2$. In particular, since $m(n) \in G$, $f!n = f * n^2$. ∎

In the remainder of this section will take $f \in \mathcal{F}_N$ to be a Hauptmodul of G which will contain $\Gamma_0(N)$. We shall also normalize f to have zero constant term.

Proposition 6.14. *$f * n^2 = f$ for any n coprime to N.*

Proof: Let n be an integer coprime to N. Since $*n^2$ is an automorphism, $f * n^2 \in \mathcal{F}_N$. Also, by Lemma 6.12 and Lemma 6.13, $f * n^2$ is fixed by G and so by [Sh, Proposition 2.6] it is an automorphic function for G. Since f is a Hauptmodul the field of automorphic functions of G is $\mathbb{C}(f)$ and hence by Proposition 6.10, $f * n^2 \in \mathbb{Q}(\zeta_N, f)$. Similarly $f * n^{-2}$ can be expressed as a rational function of f and so applying $*n^2$ to this expression we find that $f \in \mathbb{C}(f * n^2)$ and hence $\mathbb{C}(f) = \mathbb{C}(f * n^2)$. Thus $f * n^2$ is also a normalized Hauptmodul for G and since there is only one such function we have $f * n^2 = f$. ∎

Lemma 6.15. *For any n coprime to N, with $A(n)$ as in equation (6.1), the functions $f \circ \alpha, \alpha \in A(n)$ are distinct.*

Proof: Since f is a Hauptmodul its q-expansion is of the form $f = 1/q + \ldots$. Now use the same proof as for the j function, see for example [Knapp, Lemma 11.31]. ∎

We now follow the standard proof of the existence of modular polynomials for the j function. See for example [Knapp, p.334-6] or [Sh, p.108-110]. Note that the possibility that the coefficients of f are irrational introduces a slight complication into the proof of the following result, as we do not know *a priori* that $f * n$ is a Hauptmodul. In fact, once we have proven Theorem 1.3 we can use it to show that $f * n$ is a Hauptmodul for $G * n$; the condition that the coefficients of f are algebraic integers is not required since, for example by Corollary 6.6 **b**, $G * n$ is not trivial.

Proposition 6.16. *For any n coprime to N, any symmetric polynomial of $\{f \circ \alpha \mid \alpha \in A(n)\}$ is an element of $H[f * n]$ where H is the field generated over $\mathbb{Q}$ by the coefficients of f.*

Proof: Let s be any symmetric polynomial with rational coefficients of the $\{f \circ \alpha \mid \alpha \in A(n)\}$. By Corollary 6.6 **a** each $f \circ \alpha$, $\alpha \in A(n)$ is a modular function of level nN, and hence so is s. By Proposition 6.8 and Lemma 6.15, $s(z)$ is invariant under $\overline{G * n}$ and it follows from [Sh, Proposition 2.6] that $s(z)$ is an automorphic function for $\overline{G * n}$. In particular $s(z)$ has a q-expansion (rather than a $q^{1/n}$-expansion) as it is invariant under $z \mapsto z + 1$. Each coefficient in this q-expansion lies in $\mathbb{Q}(\exp(2\pi i/nN))$, the Galois group of which is $(\mathbb{Z}/nN\mathbb{Z})^* \cong (\mathbb{Z}/n\mathbb{Z})^* \times (\mathbb{Z}/N\mathbb{Z})^*$. The action of $r \times 1$ is to replace $f \circ \alpha$ by $f \circ \alpha'$ where if $\alpha = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ then $\alpha' = \begin{pmatrix} a & b' \\ 0 & d \end{pmatrix}$ where $b' = rb \pmod{d}$. The effect of $1 \times w$ is to replace $f \circ \alpha$ by $(f * w) \circ \alpha$. It follows that the q coefficients of $s(z)$ are fixed by $(\mathbb{Z}/n\mathbb{Z}) \times \mathrm{Fix}(H)$ and so lie in H. In particular we have $s(z) \in \mathcal{F}_N$.

By Corollary 6.6 **b**, if G contains $\Gamma_0(N)$ then so does $G * n$ and so $G * n$ satisfies the hypotheses of Proposition 6.8. The proof of Lemma 6.15 also works for $f * n$. Thus we can repeat the argument at the start of this proof to deduce $s * n$ is invariant under $\overline{G}$. By Proposition 6.10, $s * n = (P * n)(f)/(Q * n)(f)$ where $P(x), Q(x) \in \mathbb{Q}(\zeta_N)[x]$ and so $s = P(f * n^{-1})/Q(f * n^{-1}) = P(f * n)/Q(f * n)$.

It remains to show that Q is a constant. Suppose $s = P(f * n)/Q(f * n)$ and that Q is not a constant. Then Q has some nontrivial factor $f * n - c$. Since f is analytic in the upper half plane so is s and so there can be no point $z_0 \in \mathcal{H}$ with $(f * n)(z_0) = c$. Also since $f * n$ is a modular function of level N it must take the value c at some point in $\mathcal{H}^*$ (see, for example, [Mu, Corollary 1.2]). Thus c must be a cusp value of $f * n$. Let z_0 be one of the corresponding cusps, so $\lim_{z \to z_0}(f * n)(z) = c$ (limits at cusps are taken in the topology described, for example, in [Sh, Chapter 1]). We have $(f * n)(i\infty) = \infty$ and so there is no $\langle m \rangle \in \overline{G * n}$ such that $\langle m \rangle(z_0) = i\infty$ and so by Corollary 6.9 for each $\alpha \in A(n)$, there is no $\langle m' \rangle \in \overline{G}$ such that $\langle m' \rangle(\alpha(z_0)) = i\infty$. Since f is a Hauptmodul it is finite at any cusp not in $G(i\infty)$ and it follows that for each $\alpha \in A(n)$, $\lim_{z \to z_0} f(\alpha(z))$ is finite and hence $\lim_{z \to z_0} s(z)$ is finite. This is a contradiction and so Q must be a constant and s is a polynomial in $f * n$ as required.

Finally, if we write $s = \sum c_i (f*n)^i$, then equating the nonpositive q coefficients on each side yields a linear system for the c_i with coefficients in H. It follows that $s(z) \in H[f*n]$.∎

Let

$$F_n(y) = \prod_{\alpha \in A(n)} (y - f \circ \alpha)$$

Proposition 6.17. *For each n coprime to N, the polynomial $F_n(y)$ has coefficients in $H[f*n]$ and is irreducible over $\mathbb{C}(f*n)$.*

Proof: By Proposition 6.16 the coefficients of $F_n(y)$ are in $H[f*n] \subset \mathbb{C}(f*n)$. Also $F_n(y)$ splits in $\mathbb{C}(f \circ \alpha | \alpha \in A(n))$; $\overline{G*n}$ acts on this field, fixes $\mathbb{C}(f*n)$ and permutes the roots of $F_n(y)$. By Proposition 6.8, each $\alpha \in A(n)$ is in $\overline{\Gamma_0(N)\alpha_n G*n}$ so we have $\alpha = \gamma\alpha_n\gamma'$, $\gamma \in \overline{\Gamma_0(N)}$ and $\gamma' \in \overline{G*n}$. Hence $\gamma^{-1}\alpha = \alpha_n\gamma'$ and so $\overline{G*n}$ acts transitively on the roots. By Lemma 6.15 the roots are distinct and so $F_n(y)$ is irreducible over $\mathbb{C}(f*n)$. ∎

Since the ring $H[f*n]$ is isomorphic to $H[x]$ there is a unique polynomial $F_n(x,y) \in H[x,y]$ such that $F_n(f*n, y) = F_n(y)$.

Proposition 6.18. *For any $n > 1$ coprime to N, $F_n(x,y)$ is a generalized modular polynomial for f.*

Proof: By Proposition 6.17 $F_n(x,y) \in H[x,y]$ and by the definition of $F_n(x)$ properties **MI.2** and **MI.3'** are satisfied. Thus it remains to show that $F_n(x,y) = (F_n * n)(y,x)$. We follow the proof in [K, Proposition 3.2]. Let $a, b, d \in \mathbb{Z}$ with $ad = n$, $\gcd(a,b,d) = 1$ and $0 \le b < d$. Also let $b' = a - b$, $\beta_1 = \begin{pmatrix} d & b' \\ 0 & a \end{pmatrix}$, $\beta_2 = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ and $z' = \beta_2(z)$. We have that $F_n((f * n)(z'), f(\beta_1(z'))) = 0$. Since $\beta_1 \circ \beta_2(z) = z + 1$ it follows that $F_n((f*n)(\beta_2(z)), f(z)) = 0$. Let n' be such that $n' = n \pmod N$ and $n' = 1 \pmod n$. Then applying $*n'$ we find $(F_n * n)(f(\beta_2(z)), (f*n)(z)) = 0$. Thus the roots of $\chi_n(y) = (F_n * n)(y, (f*n)(z))$ include those of $F_n(y) = F_n((f*n)(z), y)$. So $c(y)F_n(y) = \chi_n(y)$ for some polynomial $c(y)$. We can compute c as follows: a straightforward calculation shows that the leading term in the q-expansion of $F_n(0)$ is q^t, $t = -\psi(n)$. Since $F_n(0)$ is a polynomial in $f*n$ we can deduce that its leading term is $(f*n)^t$. The other coefficients of $F_n(y)$ are lower degree polynomials in $f*n$. It follows that the leading term of $\chi_n(y)$ is y^t and as the leading term of $F_n(y)$ is y^t the result follows. ∎

Proof of Theorem 1.4 1: This follows from Proposition 6.14 and Proposition 6.18. ∎

§7. The main result.

In this section we obtain Theorem 1.3 and Theorem 1.4 **2**; before giving the proof we start with two Lemmas.

Lemma 7.1. *If f satisfies the hypotheses of Theorem 1.3 and if G is the group of symmetries of f, then G satisfies the hypotheses of Proposition 5.1.*

Proof: First, assume for contradiction that G is not discrete. Pick a sequence of distinct elements $\alpha_i \in G, i \in \mathbb{Z}^{>0}$ which converges to $\alpha \in SL(2,\mathbb{R})$, such that for all i and j, $\alpha_i \neq \pm\alpha_j$. For each i, j, $\alpha_i\alpha_j^{-1}$ has at most 1 fixed point in $\mathcal{H}$, and thus we can choose some $z_1 \in \mathcal{H}$ which avoids those countably many points. It will obey $\alpha_i(z_1) \neq \alpha_j(z_1)$ $\forall i, j$.

We have $\lim_{i\to\infty} \alpha_i(z_1) = \alpha(z_1)$ and for all i $f(\alpha_i(z_1)) = f(z_1)$. But the points $\alpha_i(z_1)$ are distinct, so f takes on the same value infinitely many times in a compact neighbourhood of $\alpha(z_1)$, which is impossible since f is not constant. Thus G must be discrete.

Next, suppose that the transformation $z \mapsto (az+b)/d$ is in G. Then $a \neq 0$ and for $\mathrm{Im}\,(z)$ sufficiently large we have that both $(az+b)/d \in \mathcal{H}_t$ and $z \in \mathcal{H}_t$. As $f((az+b)/d) = f(z)$ by Lemma 2.2 we must have $(az+b)/d - z \in \mathbb{Z}$. It follows that $a = d$ and d divides b so that $z \mapsto (az+b)/d$ is a translation by an integer as required.

Finally, note that the third condition of Proposition 5.1 is just Lemma 3.2 **b** in the case that α is a global symmetry. ∎

Lemma 7.2. *With f as in Lemma 7.1, suppose $\overline{G}(f)$ is trivial (i.e. consists only of translations by integers) and that all the coefficients a_n of f are algebraic integers. Then $f = q^{-1} + \zeta q$ where $\zeta^{\gcd(24,K)+1} = \zeta$.*

Proof: If $f = q^{-1}$ then we are done, otherwise let a_j be the first nonzero coefficient of f. Then $|a_j^\sigma| \geq 1$ for some $\sigma \in \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, as can be seen by looking at the constant term of the minimal polynomial of a_j. Let h be the formal q-series corresponding to f as in the introduction. It is straightforward to show that as a formal series h^σ satisfies a modular equation of order n for all $n = 1 \pmod K$. In fact this modular equation is just F_n^σ. Thus, as discussed in the introduction, results of Mahler and Kozlov imply that $(h^\sigma)(\exp(2\pi i z))$ extends to an analytic function on the upper half plane which we shall call f^σ. It follows that f^σ satisfies the hypotheses of Theorem 1.3. By Lemma 7.1 and Proposition 5.1 either $\overline{G(f^\sigma)}$ is trivial or f^σ is a modular function of level N for some $N \mid K^\infty$. The latter case, however, is impossible: By [Sh; Proposition 6.1] we have $f^\sigma = P/Q$, where P and Q are polynomials with complex coefficients in j and the Fricke functions (notation as in Theorem 6.0). Using a similar argument to that of Proposition 6.10 we can take the coefficients of P and Q to lie in $\bar{\mathbb{Q}}$. But then applying σ^{-1} to f^σ it is clear that f is also a modular function of level N so that $\overline{\Gamma(N)} \subset \overline{G}$: a contradiction since by hypothesis $\overline{G}$ is trivial.

Let $\overline{f^\sigma}$ be as in Lemma 2.2. We have shown that $\overline{f^\sigma}$ is injective on $|q| < 1$. By Theorem 1.4 of [Mark],

$$\sum_{i=j}^{\infty} i|a_i^\sigma|^2 \leq 1. \tag{7.1}$$

Since $|a_j^\sigma| \geq 1$, it follows that $|a_1^\sigma| = 1$ and $a_i = 0$ for $i > 1$. But then examining the sum of the roots of the modular equation of order $n = K+1$ for f shows that $1/q^{K+1} + a_1 q^{K+1} = Q_{K+1}(f(q))$ and comparing the terms in q^{K+1} yields $a_1^{K+1} - a_1 = 0$ and so a_1 is a primitive sth root of 1 for some $s \mid K$. It remains to show that $s \mid 24$.

Let $n, s \in \mathbb{Z}$ satisfy $n = 1 \pmod K$ and $s \mid K$. Also let $R = \mathbb{Q}(\zeta\zeta_n)[q^{-1/n}, q^{1/n}]$ where ζ is a primitive sth root of 1. Let W be the group of automorphisms of R generated by $g_1 : q^{1/n} \mapsto \zeta_n q^{1/n}$, and $g_2 : q^{1/n} \mapsto \zeta^{-1} q^{-1/n}$. Let $f = q^{-1} + \zeta q$ and $R^W = \{r \in R \mid g(r) = r, \forall g \in W\}$. Then it is not difficult to show that $R^W = \mathbb{Q}(\zeta\zeta_n)[f]$.

Thus if f satisfies a modular equation of order n, then the group W permutes the roots of $P(x) = F_n(f, x)$. In particular if $r_1 = \zeta_n^{-ab} q^{-a^2/n} + \zeta\zeta_n^{ab} q^{a^2/n}$ is one such root, then so is $g_2(r_1)$, this can only be the case, however, if $a^2 = 1 \pmod s$. Repeating this

argument for sufficiently many n we find that the exponent of the group $(\mathbb{Z}/s\mathbb{Z})^*$ is either 2 or 1 and so $s \mid 24$. ∎

Proof of Theorem 1.3: If f is such that for all $z_1, z_2 \in \mathcal{H}$ $f(z_1) = f(z_2)$ implies that $z_1 - z_2 \in \mathbb{Z}$ and the coefficients of f are algebraic integers, then by Lemma 7.2 f must be one of the trivial functions. If not, choose two points $z_1, z_2 \in \mathcal{H}$ such that $f(z_1) = f(z_2)$ and $z_1 - z_2 \notin \mathbb{Z}$. By Proposition 2.3 and Proposition 4.6 there exists an element $m \in SL(2, \mathbb{R})$ such that $m(z_1) = z_2$ and $f(m(z)) = f(z)$. Let $m(z) = (az + b)/(cz + d)$. If $c = 0$ then by Lemma 7.1 m is a translation by an integer: a contradiction. Thus we must have $c \neq 0$. From Proposition 5.1 and Lemma 7.1 the symmetry group G of f contains $\Gamma_0(N)$ for some N, with $N \mid K^\infty$.

Thus $f(z)$ gives rise to a function $\hat{f}$ on $X(G)$ that is analytic except possibly at the cusps. At the cusp corresponding to infinity $\hat{f}$ has a simple pole. At the other cusps, if any, $\hat{f}$ is bounded and hence has removable singularities. It follows that f is an automorphic function. By Proposition 2.3 and Proposition 4.6 $\hat{f}$ is injective on $X(G)$, which is hence of genus zero and f is a Hauptmodul for G, as required. ∎

Proof of Theorem 1.4 2: If $f(z) = q^{-1} + \zeta q$ with ζ a primitive sth root of 1 and $s \mid \gcd(24, K)$ then using the techniques in the proofs of Lemma 6.16 and Lemma 7.2 it can be shown that any symmetric polynomial in $\{f \circ \alpha \mid \alpha \in A(n)\}$ lies in $\mathbb{Q}(\zeta)[f]$. Defining $F_n(x, y)$ as in Proposition 6.18, the proof of Proposition 6.18 shows that F_n is a generalized modular polynomial for f.

In the case that $f(z) = q^{-1}$ the proof is identical, except that in this case the group W of Lemma 7.2 is generated by $g_1 : q^{1/n} \mapsto \zeta_n q^{1/n}$. ∎

§8. Comments and conjectures.

Replicable functions In [N1] Norton introduced the notion of a replicable function. A formal q-series

$$h(q) = q^{-1} + \sum_{n=1}^{\infty} h_n q^n$$

is said to be replicable if for each $a \in \mathbb{Z}^{>0}$ there exists a formal q-series

$$h^{(a)}(q) = q^{-1} + \sum_{n=1}^{\infty} h_n^{(a)} q^n$$

such that the analogue of equation (1.1) holds, namely:

$$\sum_{\substack{ad=n \\ 0 \le b < d}} h^{(a)}(\zeta_d^b q^{\frac{a}{d}}) = Q_n(h(q)),$$

where Q_n depends on h and is the unique polynomial such that

$$Q_n(h(q)) = q^{-n} + \text{terms of positive degree.}$$

One aim of this definition is to investigate to what extent the moonshine conjectures can be extended to include other groups and automorphic functions. In the case that h has *rational integer* coefficients Norton conjectured that h is replicable if and only if h is the q-expansion of a function f analytic on the upper half plane and satisfying condition **2** of Theorem 1.3. In [CuN] it was shown that any Hauptmodul with rational integer coefficients is replicable.

Our results do not provide a proof of Norton's conjecture. However Norton also introduced the notion of a completely replicable function of order K. These are replicable functions which satisfy the additional constraints * that for all $a, b \in \mathbb{Z}^{>0}$ $h^{(a)}$ is replicable, $h^{(a)^{(b)}} = h^{(ab)}$ and $h^{(a)} = h^{(\gcd(a,K))}$. These properties are satisfied by the moonshine functions. Completely replicable functions are discussed in more detail by Kozlov [K]. He proves in particular that if f is a completely replicable function of order K then f satisfies a modular equation of order n for all n coprime to K ([K, Proposition 4.1] and [K, Proposition 3.2]), see also [Mar]. Thus from Theorem 1.3 we deduce that every completely replicable function of order K with coefficients which are algebraic integers satisfies condition **2** of Theorem 1.3. The converse is not true even if we retrict to the case where $h(q)$ has rational coefficients: there are Hauptmoduln satisfying condition **2** of Theorem 1.3 which are not completely replicable [N1].

A generalization of the Mahler recurrence relations Mahler [Mah] used the modular equation of order 2 to give recurrence relations for the coefficients of the j function. More generally these recurrences determine the coefficients of any odd level Hauptmodul, $f = q^{-1} + \sum_{n=1}^{\infty} a_n q^n$, $a_n \in \mathbb{Q}$, $n = 1, 2, \ldots$, once a_1, a_2, a_3 and a_5 are given. Using Theorem 1.4 we can derive a similar set of recurrence relations that hold for any Hauptmodul of level N whose coefficients lie in $\mathbb{Q}(\zeta_N)$ when N is odd. These recurrences are, for $k \geq 1$:

$$a_{4k} = a_{2k+1} * 2 + \sum_{j=1}^{k-1} (a_j a_{2k-j}) * 2 + \frac{1}{2}((a_k * 2)^2 - a_k),$$

$$a_{4k+1} = a_{2k+3} * 2 + \sum_{j=1}^{k} (a_j a_{2k+2-j}) * 2 + \frac{1}{2}((a_{k+1} * 2)^2 - a_{k+1}) + \frac{1}{2}(a_{2k}^2 + a_{2k} * 2)$$
$$- (a_2 a_{2k}) * 2 + \sum_{j=1}^{k-1} a_j a_{4k-4j} + \sum_{j=1}^{2k-1} (-1)^j a_j a_{4k-j},$$

$$a_{4k+2} = a_{2k+2} * 2 + \sum_{j=1}^{k} (a_j a_{2k+1-j}) * 2,$$

* Note although these conditions can be imposed on any $h(q)$, Norton originally proposed them only for the case where the coefficients h_n are rational. A conjecture has been made as to how the definition of a replicable function should be modified when the h_n lie in a composite of quadratic fields [N2]. That such a modification is necessary if Hauptmoduln with irrational coefficients are to be replicable can be seen, for example, by considering the sum of the roots of a generalized modular equation $F_n((h * n)(q), y)$ when $n \neq 1 \pmod{N}$.

and

$$a_{4k+3} = a_{2k+4} * 2 + \sum_{j=1}^{k+1}(a_j a_{2k+3-j}) * 2 - \frac{1}{2}(a_{2k+1}^2 - a_{2k+1} * 2)$$

$$- (a_2 a_{2k+1}) * 2 + \sum_{j=1}^{k} a_j a_{4k+2-4j} + \sum_{j=1}^{2k}(-1)^j a_j a_{4k+2-j}.$$

Weakening the hypotheses of Theorem 1.3 It seems probable that the condition in part **2** of Theorem 1.3 that the coefficents of f are algebraic integers can be weakened. We make the following:

Conjecture 8.1. *Let f be an analytic function on $\mathcal{H}$ with Fourier expansion $f(z) = q^{-1} + \sum_{n=1}^{\infty} a_n q^n$, $q = \exp(2\pi i z)$, $a_n \in \mathbb{C}$, $n \in \mathbb{Z}^{>0}$ and let K be a positive integer. Then the following are equivalent:*

1 *f satisfies a modular equation of order n for all $n = 1 \pmod K$.*

2 *f is either $q^{-1} + \zeta q$ where $\zeta^{\gcd(24,K)+1} = \zeta$, or is a Hauptmodul for a subgroup G of $SL(2,\mathbb{R})$ satisfying:*

- **a** *G contains $\Gamma_0(N)$ with finite index for some $N \mid K^\infty$.*
- **b** *G contains $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ if and only if $k \in \mathbb{Z}$.*
- **c** *$X(G)$ has genus zero.*
- **d** *$a_i \in \mathbb{Z}[\zeta_N]$, $i \in \mathbb{Z}^{>0}$.*

Some evidence is provided for this conjecture by the following:

Proposition 8.2. *Let $K \in \mathbb{Z}^{>0}$ and $\mathcal{P} \subset \mathcal{N}_K$ be any infinite set of primes. Then there are only finitely many formal q-series h of the form $h = q^{-1} + \sum_{n=1}^{\infty} a_n q^n$, $a_n \in \mathbb{C}$ such that h satisfies a modular equation of order p for all $p \in \mathcal{P}$. Moreover the coefficients of any such h lie in a finite extension of $\mathbb{Q}$.*

Proof: In sections 13 and 29 of [Mah] it is shown that if h satisfies a modular equation of order ℓ for some prime ℓ and $m = \ell^2 + \ell - 1$, then the a_i, $i > m$ are polynomials in $a_1, \ldots, a_m$. The coefficients of these polynomials are in $\mathbb{Q}$ by a argument similar to that of Proposition 6.16. Consider $p \in \mathcal{P}$ and $F_p(x) = F_p(h(q), x) = c_0(q) + c_1(q)x + \ldots + c_t(q)x^t$ where $F_p(x,y)$ is a modular polynomial for h and $t = p+1$. Initially we only know that the coefficients of $F_p(x,y)$ are complex numbers. Once again, however, an argument similar to the one in the proof of Proposition 6.16 shows that the $c_i(q)$ are q-series with coefficients in $\mathbb{Q}[a_1, a_2, \ldots]$ and as noted above $\mathbb{Q}[a_1, a_2, \ldots] \subseteq \mathbb{Q}[a_1, \ldots, a_m]$. Since $F_p(x,y)$ is a modular polynomial each $c_i(q)$ is a polynomial P_i, say, in $h(q)$ which can be calculated by equating the coefficients and the nonpositive terms in the q-expansions of $c_i(q)$ and P_i. The fact that the leading term of $h(q)^i$ is q^{-i} means that P_i has coefficients in $\mathbb{Q}[a_1, \ldots, a_m]$. Hence each modular polynomial $F_p(x,y)$ for $h(q)$ has coefficients in $\mathbb{Q}[a_1, \ldots, a_m]$. It is not difficult to see that $h(q)$ will satisfy $M.3$ if and only if $F_p(h(q), h(q^p)) = 0$. Thus h will satisfy a modular equation of order p for all $p \in \mathcal{P}$ if and only if $(a_1, \ldots, a_m)$ is a zero of some ideal I in $\mathbb{Q}[x_1, \ldots, x_m]$.

As noted in the introduction, the results of Mahler and Kozlov imply that $f(z) = h(\exp(2\pi iz))$ extends to an analytic function on the upper half plane. Not all of our previous results apply. However the results of section 2 are still available by the following argument: For any analytic function $g(z)$ on $\mathcal{H}$ such that $g(z) = g(z+1)$ define the Hecke operators

$$T_n(g) = \sum_{ad=n,\, 0\le b<d} g((az+b)/d), \quad T_n^*(g) = \sum_{\substack{ad=n,\, 0\le b<d \\ \gcd(a,b,d)=1}} g((az+b)/d).$$

The standard algebra of Hecke operators gives

$$T_n(T_m(g)) = \sum_{d\mid \gcd(m,n)} d\, T_{mn/d^2}(g) \tag{8.1}$$

and we also have

$$T_n(g) = \sum_{d^2\mid n} T^*_{n/d^2}(g). \tag{8.2}$$

We observe that $f(z)$ obeys a modular equation of order n if and only if $T_n^*(f^j)$ is a polynomial in $f(z)$ for all $j \in \mathbb{Z}^{>0}$. If all the prime divisors of n are in $\mathcal{P}$, then equations (8.1) and (8.2) by induction show that $T_n^*(f^j)$ is a polynomial in $f(z)$ for all $j \in \mathbb{Z}^{>0}$ and so $f(z)$ satisfies a modular equation of order n.

The results of sections 3 and 4 still apply and together with the first part of Lemma 7.1 show that if $f(z_1) = f(z_2)$ then there exists $m \in SL(2,\mathbb{R})$ such that $m(z_1) = z_2$ and that $G(f)$ satisfies conditions **1** and **2** of Proposition 5.1. Let U be as in Lemma 5.3 **b**. If $z_1, z_2 \in U$, $f(z_1) = f(z_2)$ and $m(z_1) = z_2$ then by Lemma 5.3 **b** we must have $m \in G_\infty$. Thus $f(z)$ is injective modulo 1 on U. Moreover if $m \in G$ then $m' = m \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} m^{-1} \in G$ and $c_{m'} = c_m^2$. Iterating this procedure we see from Lemma 5.3 **a** that for all $m \in G$ we must have $|c_m| \ge 1$. Thus we may choose $r = 1$ in Lemma 5.3 so that $f(z)$ is injective modulo 1 on $\{z \in \mathcal{H} \mid \mathrm{Im}(z) > 1\}$. Making an appropriate rescaling and applying equation (7.1) we find that the q coefficients of $f(z)$ are bounded by

$$|a_j| \le e^{2\pi(j+1)}, \quad j = 1, 2, \ldots. \tag{8.3}$$

It can be shown, for example by constructing a Gröbner basis using a reverse lexicographic term ordering, that either there are only finitely many zeros of I all of which lie in some finite extension of $\mathbb{Q}$ or there are zeros $(a_1, \ldots, a_m)$ of I with $\max_j |a_j|$ arbitrarily large (see for example [BW, Propositions 7.42 and 7.52]). However the latter case is impossible by equation (8.3). The result now follows. ∎

References

[**B**] R.E. Borcherds, *Monstrous Moonshine and monstrous Lie superalgebras,* Invent. Math. **109**, 405-444 (1992).

[**BR**] R.E. Borcherds and A. J. E. Ryba, *Modular Moonshine II,* to appear in Duke J. Math. (1996).

[**BW**] T. Becker and V. Weispfenning, *Gröbner Bases,* Springer-Verlag, (1993).

[**CM**] H. Cohn and J. McKay, *Spontaneous generation of modular invariants,* to appear in Math of Computations **66** (1996).

[**CN**] J.H. Conway and S.P. Norton, *Monstrous Moonshine,* Bull. Lond. Math. Soc. **11**, 308-339 (1979).

[**CuN**] C.J. Cummins and S.P. Norton, *Rational hauptmoduls are replicable,* Can. J. Math.]bf 47, 1201-1218 (1995).

[**CY**] I. Chen and N. Yui, *Singular values of Thompson series,* in Groups, Difference sets and the Monster, eds. K.T. Arusu *et al*, de Gruyter (1995).

[**D**] P.L. Duren, *Univalent Functions,* Springer-Verlag (1983).

[**FLM**] I.B. Frenkel, J. Lepowsky and A. Meurman, *Vertex operators and the monster,* Academic Press: Boston (1988).

[**HW**] G.H. Hardy and E.M. Wright *An introduction to the theory of numbers,* Fifth Edition, Clarendon Press: Oxford, (1979).

[**Knapp**] A. W. Knapp, *Elliptic Curves,* Princeton University Press, (1992).

[**Koi**] M. Koike, *On replication formula and Hecke operators,* Nagoya University, (preprint).

[**K**] D. N. Kozlov, *On completely replicable functions and extremal poset theory,* Ph.D. thesis, Department of Mathematics, University of Lund, Sweden, (1994).

[**L**] S. Lang, *Elliptic functions,* 2nd edition, Addison-Wesley: Reading Mass (1987).

[**Mah**] K. Mahler, *On a class of non-linear functional equations connected with modular functions,* J. Austral. Math. Soc. **22A**, 65-118 (1976).

[**Mar**] Y. Martin, *On modular invariance of completely replicable functions,* in Moonshine, the Monster, and related Topics, eds. C Dong and G. Mason, Contemporary Mathematics **193**, 263–286, Amer. Math. Soc., Providence, RI, (1996).

[**Mark**] A. Markushevich, *Theory of functions of a complex variable, Vol. III,* Chelsea Publishing Company, (1977).

[**M**] G. Mason (with an appendix by S.P. Norton), *Finite groups and modular functions,* Proc. Symp. Pure Math. **47**, Part 1, *The Arcata Conference on Representations of Finite Groups (Arcata, Calif., 1986)*, 181–210, Amer. Math. Soc., Providence, RI, (1987).

[**Mu**] V. K. Murty, *Introduction to Abelian Varieties,* CRM Monograph Series, Amer. Math. Soc., Providence, RI, (1993).

[**N1**] S. P. Norton, *More on Moonshine,* in Computational Group Theory ed. M. D. Atkinson, Academic Press, 185-193 (1984).

[**N2**] S. P. Norton, *Non-monstrous Moonshine,* in Groups, Difference Sets, and the Monster, eds. K. T. Arasu *et al*, de Gruyter, 433-441 (1996).

[**Sh**] G. Shimura, *Introduction to the arithmetic theory of automorphic functions,* Princeton University Press, (1971).

[**T**] J.G. Thompson, *A finiteness theorem for subgroups of* $PSL(2,\mathbb{R})$ *which are commensurable with* $PSL(2,\mathbb{Z})$, Proc. Sym. Pure. Math., **37**, *Santa Cruz Conference on finite groups*, 533–555, Amer. Math. Soc., Providence RI, (1980).

Centre interuniversitaire en calcul mathématique algébrique (CICMA)
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd West
Montréal, H3G 1M8
Québec
CANADA

Max-Planck-Institut für Mathematik
Gottfried-Claren-Strasse 26
53225 Bonn
GERMANY