arXiv:1503.05690v1 [math.NT] 19 Mar 2015

MODULAR EMBEDDINGS OF TEICHMÜLLER CURVES

MARTIN MÖLLER AND DON ZAGIER

Contents

Abstract. Fuchsian groups with a modular embedding have the richest arithmetic properties among non-arithmetic Fuchsian groups. But they are very rare, all known examples being related either to triangle groups or to Teichmüller curves.

In Part I of this paper we study the arithmetic properties of the modular embedding and develop from scratch a theory of twisted modular forms for Fuchsian groups with a modular embedding, proving dimension formulas, coefficient growth estimates and differential equations.

In Part II we provide a modular proof for an Apéry-like integrality statement for solutions of Picard-Fuchs equations. We illustrate the theory on a worked example, giving explicit Fourier expansions of twisted modular forms and the equation of a Teichmüller curve in a Hilbert modular surface.

In Part III we show that genus two Teichmüller curves are cut out in Hilbert modular surfaces by a product of theta derivatives. We rederive most of the known properties of those Teichmüller curves from this viewpoint, without using the theory of flat surfaces. As a consequence we give the modular embeddings for all genus two Teichmüller curves and prove that the Fourier developments of their twisted modular forms are algebraic up to one transcendental scaling constant. Moreover, we prove that Bainbridge's compactification of Hilbert modular surfaces is toroidal. The strategy to compactify can be expressed using continued fractions and resembles Hirzebruch's in form, but every detail is different.

Introduction

Modular forms are certainly best understood for the full modular group $\mathrm{SL}(2,\mathbb{Z})$, closely followed by those for its congruence subgroups and other arithmetic groups. Among the non-arithmetic Fuchsian groups, the groups having the best arithmetic properties are those admitting a modular embedding. Here, *modular embedding* refers to the existence of a map $\varphi : \mathbb{H} \to \mathbb{H}$ intertwining the action of a Fuchsian group Γ and its Galois conjugate. The notion of modular embedding (in this sense) appears for the first time in work of Cohen and Wolfart ([9]). They show that triangle groups admit modular embeddings, and for more than a decade these remained the only examples. An infinite collection of new examples were found with the discovery of new Teichmüller curves by Calta [7] and McMullen [20]. To find the modular embeddings for them is one of the motivations for this paper.

For a reader whose main focus is modular forms this paper wants to advertise an interesting new class of modular forms. For example, we explain an integrality phenomenon for the coefficients of a solution of a Picard-Fuchs differential equation, like Beukers's modular proof of the corresponding phenomenon for Apéry's famous differential equations, except that this time the explanation requires using a *pair* of "q-coordinates." For a reader with main focus on Teichmüller curves, we show how to rediscover many of their properties without referring to the theory of flat surfaces.

The paper has three parts, linked by the aim to describe modular embeddings. In Part I we set up a general theory of modular forms for Fuchsian groups admitting a modular embedding. We call these *twisted modular forms* and we prove the twisted analogs of the properties that appear in most textbooks about modular forms in the untwisted case. In Part II we start from an example of a Fuchsian group with modular embedding where the Picard-Fuchs differential equations can be explicitly computed. We invite the reader to discover the properties of Fourier coefficients of the modular embeddings and of twisted modular forms via this worked example. In Part III we give explicitly the Fourier expansions of the modular embedding for the genus two Teichmüller curves found by Calta [7] and McMullen [20]. In the rest of the introduction we describe the results in more detail, highlighting the main theorems (not necessarily in the same order) with bullet points.

Modular embeddings and twisted modular forms. Suppose that the Fuchsian group $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$ has coefficients in a quadratic number field K with Galois group generated by σ. The essential ingredient of a modular embedding for Γ is a holomorphic function $\varphi : \mathbb{H} \to \mathbb{H}$ with the Γ-equivariance property

$$\varphi\Big(\frac{az+b}{cz+d}\Big) \;=\; \frac{a^\sigma \varphi(z) + b^\sigma}{c^\sigma \varphi(z) + d^\sigma}\,.$$

We show that such a modular embedding always has a "Fourier expansion," and an old theorem of Carathéodory gives us a good estimate for its Fourier coefficients. Analyzing φ leads us to the definition of *φ-twisted modular form* of bi-weight (k,ℓ). This is a holomorphic function $f : \mathbb{H} \to \mathbb{C}$ with the transformation property

$$f\Big(\frac{az+b}{cz+d}\Big) \;= (cd+d)^k (c^\sigma \varphi(z) + d^\sigma)^\ell f(z)\,.$$

For example, direct calculation shows that $\varphi'(z)$ is a twisted modular form of bi-weight $(2,-2)$.

We develop a theory of twisted modular forms from scratch, analyzing to which extent classical topics of modular forms generalize to this new notion. Our first topic is the coefficient growth.

- For $\ell > 0$ and $k+\ell > 2$ the Fourier coefficients of a twisted modular form $f(z) = \sum_{n \geq 0} a_n q^n$ of (k,ℓ) satisfy the estimate $a_n = \mathrm{O}(n^{k+\ell-1})$.

Similar estimates are given for other bi-weights as well; see Theorem 2.1 for the complete statement. The proofs combine the well-known Hecke argument in the untwisted case and the mechanism underlying the equidistribution of long horocycles.

The next classical topic is the dimension of the space of modular forms. A modular embedding comes with one basic invariant λ_2, that one can view is several ways: as an integral over a conformal density (10), as a ratio of the degrees of the natural line bundles whose sections are twisted modular forms, or as a Lyapunov exponent for the Teichmüller geodesic flow in the case of Teichmüller curves.

- For $k+\ell$ even and $k+\lambda_2\ell > 2$ the dimension of the space of twisted modular forms of bi-weight (k,ℓ) is the sum of $(k+\lambda_2\ell-1)(g-1)$ and contributions from the cusps and elliptic fixed points. Here g denotes the genus of $\mathbb{H}/\Gamma$.

For a torsion-free Fuchsian group this is of course a classical Riemann-Roch calculation. Hence the main point is to determine the elliptic fixed point contributions in the twisted case. See (17) for the definition of the characteristic quantities of elliptic fixed points and Theorem 3.2 for the complete statement.

Finally, the statement that modular forms expressed in terms of a modular function satisfy differential equations also carries over to the twisted case.

- If $f(z)$ is a twisted modular form of bi-weight (k, ℓ) and t a modular function, then the function $y(t)$ defined locally by $y(t(z)) = f(z)$ satisfies a linear differential equation of order $(k+1)(\ell+1)$ with algebraic coefficients (Theorem 4.1).

Modular embeddings via differential equations. The starting point of the whole project was a worked example, the Teichmüller curves for $D = 17$, that we present in Section 6–8. (The definition of Teichmüller curves along with a summary of the classification results for Teichmüller curves in genus two is given in §5.3 and §5.4.) Starting from the flat geometry definition we briefly explain the derivation of the equation of the Teichmüller curve as family of hyperelliptic curves and computation of the Picard-Fuchs differential equations, following [4].

In this way, we present in §7.1 the Fourier expansion of twisted modular forms explicitly. The corresponding group Γ is neither arithmetic nor commensurable to a triangle group, and the modular forms cannot be expressed in terms of hypergeometric functions. The twisted modular forms have two curious properties.

- The twisted modular forms do not have a power series expansion in $K[[q]]$ for the standard modular parameter $q = e^{2\pi i z/\alpha}$, where α is the width of the cusp, but lie instead in $K[[Aq]]$, where A is a transcendental number of Gelfond-Schneider type (i.e., a number of the form $\beta_1 \beta_2^{\beta_3}$ with all β_i algebraic).
- If f is a twisted modular form and t a suitably scaled modular function, then the function $y(t)$ locally defined by $y(t(z)) = f(z)$ has $\mathcal{O}_K$-integral coefficients.

The second of these observations was already proved in [4] using p-adic differential equations. In Section 8 we will give a "modular" proof of both statements. The surprising feature here is that, while the classical proof by Beukers of the integrality of the Apéry coefficients using modularity relies on the integrality of the Fourier coefficients of the q-expansions of modular forms on arithmetic groups, here the expansions of both $f(z)$ and $t(z)$ with respect to Aq have coefficients with infinitely many prime factors in their denominators, and yet the integrality of f with respect to t still holds. To give a modular argument for this integrality, we have to use the relationship between twisted modular forms on the Teichmüller curve and Hilbert modular forms on the ambient surface.

For $D = 17$ the full ring of symmetric Hilbert modular forms has been determined by Hermann ([14]). We recall and use his construction to write down explicitly the equation of the Teichmüller curves for $D = 17$ on the (rational) symmetric Hilbert modular surface in Theorem (8.3).

- The exist coordinates U, V on the Hilbert modular surface X_{17}, explicitly given in terms of theta functions, such that the two Teichmüller curves on X_{17} are cut out by the quadratic equations (70) and (71).

Modular embeddings via derivatives of theta functions. The concrete example $D = 17$ led us to the discovery of a general construction of the modular form cutting out Teichmüller curves.

- The vanishing locus of the Hilbert modular form $\mathcal{D}\theta$ of weight $(3, 9)$, given as a product of derivatives of odd theta functions, is precisely the union of one or two Teichmüller curves on the Hilbert modular surface X_D (Theorem 9.1).

Starting from the theta function viewpoint we prove the characterizing properties of genus two Teichmüller curves without relying either on the geodesic definition or on any flat surface properties. Maybe these ideas can be used to construct new Teichmüller curves. Given the length of the paper, we simplify our task and prove the following results only for fundamental discriminants D. With appropriate care, the proofs can certainly be adapted to the general case.

- The vanishing locus of $\mathcal{D}\theta$ is transversal to one of the two natural foliations of the Hilbert modular surface X_D (Theorem 12.1).
- The vanishing locus of $\mathcal{D}\theta$ is disjoint from the reducible locus (Theorem 12.2).

On the compactified Hilbert modular surface, the reducible locus and the vanishing locus of a Hilbert modular form always intersect and the number of intersection points is proportional to the volume of the Hilbert modular surface. So the claim is that all these intersection points lie on the boundary of the Hilbert modular surface, hence at cusps of the vanishing locus of $\mathcal{D}\theta$. While for the second statement we have also have a proof using theta functions, we give proofs of both these statements relying on the following description of the cusps.

- The cusps of the vanishing locus of $\mathcal{D}\theta$ are in bijection with pairs consisting of a standard quadratic form $[a,b,c]$ of discriminant D and a class $r \in \mathbb{Z}/\gcd(a,c)\mathbb{Z}$ (Theorem 10.1).

Here an indefinite quadratic form is called *standard* if $a > 0 > c$ and $a+b+c < 0$. As a statement about cusps of Teichmüller curves, this result already appears in [22] and [1]. Our proof, however, is completely different. We now explain the main idea. Suppose that a power series $f = \sum c_\nu q_1^\nu q_2^{\sigma(\nu)}$ has to vanish on a branch of a curve parameterized by $q_1 = q^{\alpha_1} e^{\varepsilon_1(q)}$ and $q_2 = q^{\alpha_2} e^{\varepsilon_2(q)}$. After these parametrizations are plugged into f, the lowest order exponent (in q) has to appear twice, since otherwise the terms cannot cancel. In the concrete case of $f = \mathcal{D}\theta$ we are led to the following notion. Given a fractional $\mathcal{O}_D$ ideal $\mathfrak{a}$, we say that a primitive element $\alpha \in (\mathfrak{a}^2)^\vee$ is a *multiminimizer* for $\mathfrak{a}$ if the quadratic form $x \mapsto \mathrm{tr}(\alpha x^2)$ on $\mathfrak{a}$ takes its minimum on one of the three non-trivial cosets $\frac{1}{2}\mathfrak{a}/\mathfrak{a}$ at least (and then, as we show, exactly) twice (with x and $-x$ not distinguished). We show on the one hand that multiminimizers for $\mathfrak{a}$ are in bijection with standard quadratic forms in the wide ideal class. On the other hand, on any branch of the vanishing locus the local parameter can be chosen such that $\alpha = \alpha_1 = \sigma(\alpha_2)$ is a multiminimizer and that the multiminimizers (up to multiplication by the square of a unit) determine the branch uniquely up to an element of $\mathbb{Z}/\gcd(a,c)\mathbb{Z}$.

We have given the definition of multiminimizers and the description of branches of the vanishing locus in detail since this notion and construction reappears twice in the rest of the paper. First, multiminimizers appear prominently in the discussion of Bainbridge's compactification below and, secondly, this description of the branches immediately gives the Fourier expansion of the modular embedding of the uniformizing group of the genus two Teichmüller curves for any D (see Theorem 13.1). Moreover, both "curious properties" mentioned in the case $D = 17$ hold in general. In particular, we have:

- For any D, any cusp of the vanishing locus of $\mathcal{D}\theta$ with corresponding Fuchsian group Γ and modular embedding φ, the φ-twisted modular forms of bi-weight

(k,ℓ) have a basis with Fourier expansions of the form $\sum_{n\geq 0} a_n (Aq)^n$ with a_n algebraic and A transcendental of Gelfond-Schneider type (Theorem 13.2).

As another application of the description of Teichmüller curves via theta derivatives, we give in Theorem 13.3 a description of the quadratic differentials on the leaves of the natural foliation of a Hilbert modular surface whose integral measures the flat distance between the two zeros of the eigenform for real multiplication. These quadratic differentials can be packaged together to a meromorphic modular form of weight $(-2,4)$ that we give as the quotient of theta series and their derivatives. Our result has been used by McMullen [26] to describe the beautiful and complicated flat structure on the leaves visually ("snow falling on cedars").

Hirzebruch's compactification and Bainbridge's compactification. Hirzebruch constructed a minimal smooth compactification of Hilbert modular surfaces. His compactification is remarkable in many ways. First, it is the prototype of what is nowadays a called a toroidal compactification, i.e., it is given by a *fan* of decreasing slopes, periodic under the action of the squares of units. The fan is given for each cusp $\mathfrak{a}$ of the Hilbert modular surface by the corners of the lower convex hull of $(\mathfrak{a}^2)^\vee$ in $\mathbb{R}^2_+$.

Bainbridge ([1]) observed that the closure in the Deligne-Mumford compactification of the Torelli preimage of Hilbert modular surfaces provides another compactification. This compactification is in general neither smooth nor minimal, but it was useful in his calculation of Euler characteristics of Teichmüller curves. It is amusing to compare the two types of compactifications and to discover that they are parallel in spirit, but different in every concrete detail.

- Bainbridge's compactification is a toroidal compactification, given for each cusp $\mathfrak{a}$ by the fan of multiminimizers (lying in $(\mathfrak{a}^2)^\vee$) for $\mathfrak{a}$. (Theorem 11.5).

The second remarkable property of Hirzebruch's compactification is that it can easily be computed using a continued fraction algorithm.

- Hirzebruch's compactification is driven by the "fast minus" continued fraction algorithm, while Bainbridge's compactification is driven by a "slow plus" continued fraction algorithm.

The reader will find the precise description of the algorithms in §11.2 and §10.4 respectively. The bijection between standard and reduced indefinite quadratic forms induces a subtle relationship between the number of boundary components of Hirzebruch's compactification and Bainbridge's compactification. In particular, the number of curves in the Bainbridge compactification of any cusp is always the same as the number for the Hirzebruch compactification of some cusp, but not necessarily the same one! The details and several examples are given in Section 11.

Acknowledgements. The first named author is partially supported by the ERC starting grant 257137 "Flat surfaces." He would also like to thank the Max Planck Institute for Mathematics in Bonn, where much of this work was done.

Part I: Modular embeddings and twisted modular forms

The notion modular embedding in the sense used here appears for the first time in a paper by Cohen and Wolfart ([9]). They study holomorphic maps $\mathbb{H} \to \mathbb{H}$ equivariant with respect to a Fuchsian group and its Galois conjugate.

In particular Cohen and Wolfart show that all triangle groups admit modular embeddings. Subsequent work of Schmutz-Schaller and Wolfart ([34]) gave some necessary conditions for a group to admit a modular embeddding. Some Fuchsian quadrangle groups were shown in [32] not to admit modular embeddings, but it took more than a decade until new examples of modular embeddings were discovered.

The first new examples arose from the Teichmüller curves discovered by Calta and McMullen (see [7] and [20], and [27] for the modular viewpoint). All Teichmüller curves give rise to modular embeddings. We summarize the known results of Teichmüller curves (and thus the known groups admitting a modular embedding) briefly at the end of Section 1.

Here, in Part I, we think of the group Fuchsian group Γ as given (e.g. in terms of a presentation) and study properties of the modular embeddings as holomorphic maps. We give an extension of the notion of modular forms to this context that we call twisted modular forms. The aim of the first part is to study this new object and to derive the analogues of the standard results on modular forms (Fourier coefficients, dimension, differential equations) from scratch in the context of twisted modular forms.

1. Hilbert modular embeddings

The term *modular embedding* is used in the literature both for equivariant maps from $\mathbb{H} \to \mathbb{H}^g$ (starting with [9]) and from $\mathbb{H}^g \to \mathbb{H}_g$ (already in [13]) . To distinguish, we call them "Hilbert modular embeddings" and "Siegel modular embeddings," respectively, according to the range of the corresponding map. We will be interested mostly in the quadratic case $g = 2$ and refer to [34] for basic notions is the general case.

Throughout this paper we denote by K a real quadratic field, with a fixed embedding $K \subset \mathbb{R}$. We use the letter σ to denote the Galois conjugation of K or the second embeddding of K into $\mathbb{R}$, writing $\sigma(x)$ or x^σ interchangeably for $x \in K$. By a *Hilbert modular group* for K we will mean any subgroup Γ_K of $\mathrm{SL}(2, K)$ commensurable with $\mathrm{SL}(2, \mathcal{O})$ for some order $\mathcal{O} \subset K$. (Later we will make specific choices.) Such a group acts discretely and cofinitely on $\mathbb{H}^2$ by $(z_1, z_2) \mapsto \left(\dfrac{az_1 + b}{cz_1 + d}, \dfrac{a^\sigma z_2 + b^\sigma}{c^\sigma z_2 + d^\sigma}\right)$. Here $\mathbb{H}$ denotes the upper half-plane.

We will be interested only in *Hilbert modular embeddings* for which the first coordinate in $\mathbb{H}^2$ is a local coordinate everywhere. A modular embedding of this type is described by the data (Γ, φ), where

- Γ is a subgroup of some Hilbert modular group $\Gamma_K \subset \mathrm{SL}(2, K)$ which, viewed as a subgroup of $\mathrm{SL}(2, \mathbb{R})$, is Fuchsian (i.e., discrete and cofinite).
- $\varphi : \mathbb{H} \to \mathbb{H}$ is a holomorphic map satisfying $\varphi \circ \gamma = \gamma^\sigma \circ \varphi$ for all $\gamma \in \Gamma$.

For such a pair (Γ, φ), the map $z \mapsto (z, \varphi(z))$ defines a map from the curve $\mathbb{H}/\Gamma$ to the Hilbert modular surface $\mathbb{H}^2/\Gamma_K$.

Written out explicitly, the condition on φ means that we have

$$\varphi\Big(\frac{az+b}{cz+d}\Big) \;=\; \frac{a^\sigma\varphi(z)+b^\sigma}{c^\sigma\varphi(z)+d^\sigma} \tag{1}$$

for all $\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma$ and $z \in \mathbb{H}$. We will use this transformation property in §2 to define a 1-cocycle on Γ and hence a new type of modular form ("twisted" form) on Γ. Just as for usual modular forms, these must satisfy suitable growth conditions at the cusps of Γ, and to formulate these we need to know how φ behaves near the cusps. Assume first that one of these cusps is at ∞, with the stabilizer of ∞ in the image $\bar\Gamma$ of Γ in $\mathrm{PSL}(2,\mathbb{R})$ being generated by $\pm\left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1\end{smallmatrix}\right)$ with $\alpha \in K \cap \mathbb{R}_+$. Then we have:

Proposition 1.1. *Suppose that the stabilizer of ∞ in Γ is generated by $z \mapsto z+\alpha$ with $a \in K$, $\alpha > 0$. Then α^σ is also positive and $\varphi(z)$ has an expansion of the form*

$$\varphi(z) \;=\; \frac{\alpha^\sigma}{\alpha}z \,+\, \sum_{n=0}^{\infty} B_n\,\mathbf{e}\big(\frac{nz}{\alpha}\big) \qquad \big(\forall z\in\mathbb{H}\big)\,, \tag{2}$$

where $\mathbf{e}(x) := e^{2\pi i x}$ and where the coefficients B_n satisfy the inequalities

$$|B_n| \;\le\; 2\,\Im(B_0) \qquad \text{for all } n\ge 1. \tag{3}$$

Proof. From (1) we have $\varphi(z+\alpha) = \varphi(z) + \alpha^\sigma$, so the function $\varphi(z) - \alpha^\sigma z/\alpha$ is invariant under $z \mapsto z+\alpha$ and hence equals $f(\mathbf{e}(z/\alpha))$ for some holomorphic function $f(q)$ in the punctured disc $\mathbb{D}^* = \{q \,:\, 0<|q|<1\}$. Define a second holomorphic function F in $\mathbb{D}^*$ by $F(q) = \mathbf{e}\big(f(q)/|\alpha^\sigma|\big)$. From $\varphi(\mathbb{H}) \subseteq \mathbb{H}$ we deduce that $|q^{\pm1}F(q)| < 1$ in $\mathbb{D}^*$, where the sign is chosen so that $\pm\alpha^\sigma > 0$. It follows that $F(q)$ extends to a meromorphic function in $\mathbb{D} = \{q \,:\, |q|<1\}$ with at most a simple pole at $q=0$. But then the fact that F has a single-valued logarithm in $\mathbb{D}^*$ implies that its order of vanishing at 0 must be zero, so α^σ must be positive and f extends holomorphically to the full disc and hence has a convergent Taylor expansion $\sum_{n=0}^\infty B_n q^n$, proving the first claim. For the second, we note first that the estimate $|F(q)| \le |1/q|$ and the holomorphy of F at 0 imply by the maximum principle that $|F(q)|\le 1$ in the disk $\mathbb{D}$ (this is just the Schwarz lemma, applied to the function $qF(q)$), so B_0 has positive imaginary part and the function $f(q)/B_0$ takes on values in the right half-plane. An elementary argument then gives the estimate $|B_n| \le 2n\Im(B_0)$. (Write $f(q) = (B_0 - \bar B_0\lambda(q))/(1-\lambda(q))$ where λ sends $\mathbb{D}$ to $\mathbb{D}$ and 0 to 0; then $B_n = 2i\Im(B_0)\sum_{m=1}^n[\lambda^m]_n$, where $[\lambda^m]_n$ denotes the coefficient of q^n in $\lambda(q)^m$, which is bounded in absolute value by 1 because λ is.) The stronger estimate $|B_n| \le 2\Im(B_0)$ follows from a theorem of Carathéodory [8], which says precisely that a holomorphic function mapping $\mathbb{D}$ to the right half-plane and sending 0 to 1 has all its Taylor coefficients at 0 bounded by 2 in absolute value. □

Corollary 1.2. *The imaginary part of $\varphi(z)$ satisfies the inequalities*

$$\frac{\alpha^\sigma}{\alpha}y \;\le\; \Im(\varphi(z)) \;\le\; \frac{\alpha^\sigma}{\alpha}y \,+\, \frac{C}{y} \tag{4}$$

for all $z = x+iy \in \mathbb{H}$, where C is a constant independent of z.

Proof. The first statement is just the inequality $|F(q)|\le 1$ established in the course of the above proof, and the second follows from (3) because $\sum_{n>0}|q|^n \ll 1/y$. □

Exactly similar statements hold for all of the other cusps of Γ. Recall that by definition, such a cusp is an element $\nu \in \mathbb{P}^1(K)$ whose stabilizer $\bar{\Gamma}_\nu$ in $\bar{\Gamma}$ is infinite cyclic, say $\bar{\Gamma}_\nu = \langle \pm\gamma_\nu \rangle$. Choose $g \in \mathrm{SL}(2,K)$ with $g(\infty) = \nu$. Then $g^{-1}\gamma_\nu g = \pm\left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right)$ for some positive element α of K. Equation (1) implies that the function $\varphi_g = g^{\sigma\,-1} \circ \phi \circ g$ satisfies $\varphi_g(z+\alpha) = \varphi_g(z) + \alpha^\sigma$, because $\varphi_g \circ \left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right) = \varphi_g g^{-1}\gamma_\nu g = g^{\sigma\,-1}\varphi\,\gamma_\nu g = g^{\sigma\,-1}\gamma_\nu^\sigma \varphi\, g = g^{\sigma\,-1}\gamma_\nu^\sigma g^\sigma \varphi_g = (g^{-1}\gamma_\nu g)^\sigma \varphi_g = \left(\begin{smallmatrix} 1 & \alpha^\sigma \\ 0 & 1 \end{smallmatrix}\right) \circ \varphi_g$. Then the same proof as for the case $g = \mathrm{Id}$ shows that α^σ is positive and that

$$\varphi_g(z) = \frac{\alpha^\sigma}{\alpha} z + \sum_{n=0}^{\infty} B_n\, \mathbf{e}\Bigl(\frac{nz}{\alpha}\Bigr) \tag{5}$$

for some constants $B_n \in \mathbb{C}$, satisfying the same estimate (3) as before, and for all $z \in \mathbb{H}$. Of course α and B_n depend on ν, and also on the choice of g, but the expansions for Γ-equivalent cusps are the same up to trivial rescalings, because $\varphi_{\gamma g} = \varphi_g$ for $\gamma \in \Gamma$, so that there are only finitely many essentially distinct expansions.

Another basic property of φ, obtained by applying the Schwarz lemma to the map $I_{\varphi(z')} \circ \varphi \circ I_{z'}^{-1} : \mathbb{D} \to \mathbb{D}$, where I_a for $a \in \mathbb{H}$ denotes the standard isomorphism $(\mathbb{H}, a) \to (\mathbb{D}, 0)$ given by $z \mapsto (z-a)/(z-\bar{a})$, is that one has the inequalities

$$\left|\frac{\varphi(z) - \varphi(z')}{\varphi(z) - \overline{\varphi(z')}}\right| \;\le\; \left|\frac{z - z'}{z - \overline{z'}}\right|, \qquad \frac{|\varphi(z) - \varphi(z')|^2}{\Im(\varphi(z))\,\Im(\varphi(z'))} \;\le\; \frac{|z - z'|^2}{\Im(z)\,\Im(z')} \tag{6}$$

or equivalently $d(\varphi(z), \varphi(z')) \le d(z, z')$ for any $z,\, z' \in \mathbb{H}$, where $d : \mathbb{H} \times \mathbb{H} \to \mathbb{R}_{\ge 0}$ denotes the Poincaré metric. (This is of course a standard property of any holomorphic map from the upper half-plane to itself.) Fixing z and letting $\Im(z')$ tend to infinity, or fixing z' and letting $\Im(z)$ tend to 0, we obtain second proofs of the two inequalities in (4), while letting z' tend to z we obtain the estimate

$$|\varphi'(z)| \;\le\; \frac{\Im(\varphi(z))}{\Im(z)} \qquad (z \in \mathbb{H})\,. \tag{7}$$

or equivalently $|\kappa(z)| \le 1$, where $\kappa : \mathbb{H} \to \mathbb{C}$ is the map defined by

$$\kappa(z) \;=\; \frac{\Im(z)}{\Im(\varphi(z))}\, \varphi'(z)\,. \tag{8}$$

From the equivariance property (1) we obtain the formulas

$$\Im(\varphi(\gamma z)) \;=\; \frac{\Im(\varphi(z))}{|c^\sigma \varphi(z) + d^\sigma|^2}\,, \quad \varphi'\Bigl(\frac{az+b}{cz+d}\Bigr) \;=\; \frac{(cz+d)^2}{(c^\sigma \varphi(z) + d^\sigma)^2}\, \varphi'(z)\,, \tag{9}$$

and these together with the standard formula $\Im(\gamma z) = \Im(z)/|cz+d|^2$ imply that the function κ is Γ-invariant. We can therefore introduce a basic invariant $\lambda_2 = \lambda_2(\Gamma, \varphi)$ of the pair (Γ, φ) by

$$\lambda_2 \;=\; \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} \iint_{\Gamma\backslash\mathbb{H}} \frac{|\varphi'(z)|^2}{\Im(\varphi(z))^2}\, dx\, dy \;=\; \frac{\iint_{\mathbb{H}} |\kappa|^2\, d\mu}{\iint_{\mathbb{H}} d\mu} \;\in\; (0,1]\,, \tag{10}$$

where $d\mu = y^{-2} dx\, dy$ (with $z = x + iy$ as usual) is the standard $\mathrm{SL}(2,\mathbb{R})$-invariant measure on $\mathbb{H}$, and where the integral can be taken over any fundamental domain for the action of Γ on $\mathbb{H}$ and is absolutely convergent because $|\kappa| \le 1$. The invariant λ_2, whose values are always rational numbers, can be interpreted either as a ratio of intersection numbers on the Hilbert modular surface or as the second Lyapunov exponent ([30]), which explains the notation.

We close this section by describing briefly the four known types of Hilbert modular embeddings. The first two will play a role in this paper and will be discussed in more detail in Section 5. The other two are mentioned only for the sake of completeness.

Type 1: Modular curves. Here $\Gamma = \Gamma_A = \{\gamma \in \Gamma_K \mid A\gamma = \gamma^\sigma A\}$ where A is a suitable element of $\mathrm{GL}^+(2,K)$, and the map φ is defined by $\varphi(z) = Az$. The corresponding curves $(1,\varphi)(\mathbb{H}/\Gamma) \subset \mathbb{H}^2/\Gamma_K$ in this case are the irreducible components of the curves T_N studied in [16] and [17] and reviewed in §5.2. In particular, there are infinitely many curves of this type on each Hilbert modular surface, and conversely each of these curves lies on infinitely many Hilbert modular surfaces in $\mathbb{H}_2/\mathrm{Sp}(4,\mathbb{Z})$.

Type 2: Teichmüller curves. These are defined abstractly as the algebraic curves in the moduli space $\mathcal{M}_g$ of curves of genus g that are totally geodesic submanifolds for the Teichmüller metric. In genus 2, they always lie on Hilbert modular surfaces and (apart from one exception for the field $\mathbb{Q}(\sqrt{5})$) have a modular interpretation as the components of the moduli space of genus 2 curves whose Jacobian has real multiplication by an order in a real quadratic field and such that the unique (up to a scalar) holomorphic form on the curve that is equivariant with respect to this action[1] has a double zero. There are at most two curves of this type on each Hilbert modular surface X_K, and conversely each Teichmüller curve lies on exactly one X_K. The proof that these curves have a modular embedding comes from [27]. (See also the proof of Proposition 5.6.)

There exists a variant of these curves, not used in this paper but studied in detail by Weiß in [36], called "twisted Teichmüller curves," obtained as the images of Teichmüller curves under the action of elements of $\mathrm{GL}(2,K)^+$. They are still geodesic for the Kobayashi metric in X_K, but no longer for the Kobayashi (= Teichmüller) metric in $\mathcal{M}_2$. There are in general infinitely many of these curves on any Hilbert modular surface.

Types 3 and 4: Curves related to Prym varieties. Recall that a *Prym variety* is the kernel of the map $\mathrm{Jac}(C) \to \mathrm{Jac}(C_0)$ induced by a double cover $C \to C_0$ of curves. By the Riemann-Hurwitz formula, it is 2-dimensional if and only if the genus g of C lies between 2 and 5. For the cases $g = 3$ or 4, there is a construction of Teichmüller curves in the moduli space $\mathcal{M}_g$ corresponding to certain Prym varieties having real multiplication by an order in a real quadratic field [23]. Our cases 3. and 4. are these two cases, in the order $g = 4$, $g = 3$.

The four types 1.–4. are distinguished by the invariant λ_2, which takes on the values 1 for Type 1, $\frac{1}{3}$ for Type 2, $\frac{1}{5}$ for Type 3, and $\frac{1}{7}$ for Type 4 ([1] and [30], Proposition 5.1). For each of these values, there is an infinite number of commensurability classes of Fuchsian groups with this invariant.

The exceptional Teichmüller curve over $\mathbb{Q}(\sqrt{5})$ corresponds to $\lambda_2 = 1/2$, and at the time of writing this is the only known commensurability class of a Fuchsian group with this value of the invariant λ_2.

---

[1] i.e. $(\mu\circ)^*\varphi = \mu\varphi$ for all μ in the order; cf. §5.1 ("first eigendifferential") for details

2. Twisted modular forms

For any function $\varphi:\mathbb{H}\to\mathbb{H}$ satisfying (1) we can define two factors $J(g,z)$ and $\widetilde{J}(g,z)$ for $g\in\mathrm{SL}(2,K)$ and $z\in\mathbb{H}$ by

$$J(g,z) \;=\; cz+d\,, \qquad \widetilde{J}(g,z) \;=\; c^\sigma\varphi(z)+d^\sigma \qquad \text{if } g=\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right). \tag{11}$$

The calculation

$$\begin{aligned}\widetilde{J}(\gamma_1\gamma_2,z) \;&=\; (c_1a_2+d_1c_2)^\sigma\,\varphi(z)\;+\;(c_1b_2+d_1d_2)^\sigma\\ &=\; \Bigl(c_1^\sigma\,\frac{a_2^\sigma\varphi(z)+b_2^\sigma}{c_2^\sigma\varphi(z)+d_2^\sigma}\;+\;d_1^\sigma\Bigr)\bigl(c_2^\sigma\varphi(z)+d_2^\sigma\bigr)\\ &=\; \Bigl(c_1^\sigma\,\varphi\bigl(\frac{a_2z+b_2}{c_2z+d_2}\bigr)\;+\;d_1^\sigma\Bigr)\bigl(c_2^\sigma\varphi(z)+d_2^\sigma\bigr) \qquad\qquad \text{(by eq. (1))}\\ &=\; \widetilde{J}(\gamma_1,\gamma_2z)\,\widetilde{J}(\gamma_2,z) \qquad \text{for } \gamma_1,\,\gamma_2\in\Gamma\end{aligned}$$

shows that $\widetilde{J}$ is a cocycle for Γ. (The corresponding statement for J, which follows from the same calculation with $\varphi\equiv\mathrm{Id}$, is, of course, standard.) It follows that the map $f\mapsto f|_{(k,\ell)}g$ of the space of holomorphic functions in $\mathbb{H}$ to itself defined for $k,\,\ell\in\mathbb{Z}$ and $g\in\mathrm{GL}^+(2,K)$ by

$$\bigl(f|_{(k,\ell)}g\bigr)(z) \;=\; J(g,z)^{-k}\,\widetilde{J}(g,z)^{-\ell}\,f(gz) \tag{12}$$

is a group action when restricted to Γ.

We now define a *φ-twisted modular form of bi-weight (k,ℓ)* on Γ to be a holomorphic function $f:\mathbb{H}\to\mathbb{C}$ satisfying $f|_{(k,\ell)}\gamma=f$ for all $\gamma\in\Gamma$ together with the growth requirement that the function f_g is bounded as $\Im(z)\to\infty$ for every $g\in\mathrm{SL}(2,K)$, where $f_g(z)=(cz+d)^{-k}(c^\sigma\varphi_g(z)+d^\sigma)^{-\ell}f(gz)$. The function f_g depends only on the coset Γg.

To describe its Fourier expansion, recall that the cusp $g\infty$ is called *irregular* if Γ does not contain -1 and if $\bigl(g^{-1}\Gamma g\bigr)_\infty=\bigl\langle-\bigl(\begin{smallmatrix}1&\alpha\\0&1\end{smallmatrix}\bigr)\bigr\rangle$. If $-1\in\Gamma$, then $k+\ell$ is even if $M_{k,\ell}(\Gamma)$ is to be non-zero. In this case $f_g(z+\alpha)=f_g(z)$ where α as before is the totally positive element of K with $\bigl(g^{-1}\Gamma g\bigr)_\infty=\bigl\langle\pm\bigl(\begin{smallmatrix}1&\alpha\\0&1\end{smallmatrix}\bigr)\bigr\rangle$. So in this case f_g has a Fourier expansion of the form

$$f_g(z) \;=\sum_{n\ge0}a_n\,\mathbf{e}(nz/\alpha) \quad\text{as}\quad \Im(z)\to\infty \tag{13}$$

where n ranges over integers. If $-1\notin\Gamma$ and the cusp is regular, or if $k+\ell$ is even, then the Fourier expansion has the same form. Only if the cusp is irregular and $k+\ell$ is odd, then $f_g(z+\alpha)=-f_g(z)$ and the Fourier expansion is as in (13), but now with n ranging over $\mathbb{Z}+\frac12$.

If the Fourier coefficient a_0 is 0 for all cusps (a condition that is automatically satisfied at regular cusps if $k+\ell$ is odd), we call f a *cusp form.* The spaces of φ-twisted modular forms and cusp forms of bi-weight (k,ℓ) will be denoted by $M_{k,\ell}(\Gamma,\varphi)$ and $S_{k,\ell}(\Gamma,\varphi)$, respectively. We will often omit the "φ" when no confusion can result.

Obviously, ordinary modular forms of weight k on Γ are φ-twisted modular forms of bi-weight $(k,0)$ for any φ, and in fact $M_{k,0}(\Gamma,\varphi)=M_k(\Gamma)$. We give three examples of twisted modular forms with $\ell\neq0$.

i) We always have $\varphi' \in M_{2,-2}(\Gamma,\varphi)$, by virtue of the second equation in (9) and the expansions of φ at the cusps given in Section 1. This example shows that the weights k and ℓ of a holomorphic twisted modular form in do not both have to be positive.

ii) In the case of modular curves, when the map $\varphi : \mathbb{H} \to \mathbb{H}$ is given by a fractional linear transformation $A \in \mathrm{GL}^+(2,K)$, the calculation

$$\widetilde{J}(\gamma,z) = c^\sigma Az + d^\sigma = J(\gamma^\sigma, Az) = \frac{J(\gamma^\sigma A,z)}{J(A,z)} = \frac{J(A\gamma,z)}{J(A,z)} = \frac{J(A,\gamma z)}{J(A,z)}\, J(\gamma,z)$$

for $\gamma \in \Gamma$ shows that if f belongs to $M_{(k,\ell)}(\Gamma,\varphi)$, then the function $f_A(z) = J(A,z)^{-\ell} f(z)$ belongs to $M_{k+\ell}(\Gamma)$ in the usual sense, so that here we do not get a new kind of modular forms. For Teichmüller curves, on the other hand, the automorphy factor $\widetilde{J}$ cannot be reduced to an automorphy factor of the classical sort, and the twisted forms are a genuinely new type of modular form.

iii) If Γ_K is a Hilbert modular group containing Γ and F is a Hilbert modular form of weight (k,ℓ) on Γ_K (i.e., $F : \mathbb{H}^2 \to \mathbb{C}$ is a holomorphic map satisfying $F(\gamma z_1, \gamma^\sigma z_2) = (cz_1+d)^k (c^\sigma z_2 + d^\sigma)^\ell F(z_1,z_2)$ for all $\gamma \in \Gamma_K$ and all z_1, $z_2 \in \mathbb{H}$), then the restriction of F to $\mathbb{H}$ under the embedding $(1,\varphi) : \mathbb{H} \to \mathbb{H}^2$ is an element of $M_{k,\ell}(\Gamma,\varphi)$.

The last example provides many twisted modular forms for any (Γ,φ). But not all twisted modular forms arise in this way, and it makes sense to study the twisted forms independently of the two-variable theory. In particular, one can ask for the dimensions of the spaces $M_{k,\ell}(\Gamma,\varphi)$ and $S_{k,\ell}(\Gamma,\varphi)$ and for the structure of the bigraded ring $M_{**}(\Gamma,\varphi) = \oplus_{k,\ell} M_{k,\ell}(\Gamma,\varphi)$, just as is usually done for classical modular forms when $\ell = 0$, and we can also study the classical topic of growth of Fourier coefficients. We shall give a general formula for the dimensions in the next section and a description of the ring of twisted forms in a special example in Section 7, while the rest of this section is devoted to the study of the coefficient growth.

Theorem 2.1. *Let $f(z) = \sum a_n q^n$ be a twisted modular form of bi-weight (k,ℓ), and set $K = k + |\ell|$. Then the Fourier coefficients of f satisfy the estimates*

$$a_n \;=\; \begin{cases} \mathrm{O}(n^{K/2}) & \text{if } k+\ell < 2 \text{ or } f \text{ is a cusp form,} \\ \mathrm{O}(n^{K/2}\log n) & \text{if } k+\ell = 2, \\ \mathrm{O}(n^{k-1+\max(0,\ell)}) & \text{if } k+\ell > 2\,. \end{cases}$$

Proof. Suppose first that f is cuspidal. Here we use a modification of the well-known argument given by Hecke in the untwisted case. We construct the real-valued continuous function

$$F(z) = |f(z)|\, y^{k/2}\, \widetilde{y}^{\ell/2},$$

where $y = y(z) = \Im(z)$ and $\widetilde{y} = \widetilde{y}(z) = \Im(\varphi(z))$. This function is Γ-invariant by the definition of a twisted modular form. Since f is a cusp form, F decays rapidly at cusps and hence is bounded, so $f(z) = \mathrm{O}(y^{-k/2}\widetilde{y}^{-\ell/2})$. On the other hand, $a_n = \frac{1}{\alpha}\int_0^\alpha f(x+iy)\mathbf{e}(-n(x+iy)/\alpha)\,dx$ for any y. Specializing to $y = 1/n$ and using the first or the second inequality in (4) depending on the sign of ℓ, we obtain the estimate stated.

In the remaining cases, still

$$|a_n| \le \frac{1}{\alpha}\int_0^\alpha F(x+i/n)\, n^{k/2}\, \Im(\varphi(x+1/n))^{-\ell/2}dx \ll n^{K/2}\int_0^\alpha F(x+i/n)\,dx,$$

by (4) (where the constant implied by $\ll$ depends only on Γ), but now $F(z) = F(\gamma(z)) = \mathrm{O}(H(z)^{(k+\ell)/2})$ instead of $\mathrm{O}(1)$, where $H(z)$ is defined in the lemma below. Since the exponent $k-1+\max(0,\ell)$ is equal to $K/2+(k+\ell)/2-1$, the remaining statement is precisely the content of the lemma. □

Lemma 2.2. *Let Γ be a non-cocompact Fuchsian group, with the width of the cusp at ∞ equal to 1, and define the height function $H(z) = \max_{\gamma\in\Gamma}\Im(\gamma z)$. Then for $n>1$ and $\lambda>0$ one has the estimates*

$$\int_0^1 H(x+\tfrac{i}{n})^\lambda\,dx \;=\; \begin{cases} \mathrm{O}(1) & \text{if } 0<\lambda<1,\\ \mathrm{O}(\log n) & \text{if } \lambda=1,\\ \mathrm{O}(n^{\lambda-1}) & \text{if } \lambda>1, \end{cases}$$

where the implied constant does not depend on n.

Proof. The case $\lambda=\frac12$ is [33], Proposition 2.2. Essentially the same method can be used to give all cases. We provide the details only for $\lambda>\frac12$. We let $T=\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and choose for each of the h cusps η_j of Γ a matrix $N_j\in\mathrm{SL}(2,\mathbb{R})$ such that $N_j\eta_j=\infty$, and such that the stabilizer of ∞ in $\Gamma_j = N_j\Gamma N_j^{-1}$ is always $\langle T\rangle$. Let $\mathcal{F}$ be a closed fundamental domain for Γ, which we may choose so that the cusp neighborhoods have the shape

$$N_j(\mathcal{F})\cap\{\zeta\in\mathbb{H} : \Im(z)>B\} = [0,1]\times[B,\infty) \qquad (j=1,\dots,h)$$

for some $B>1$, and are disjoint. We define the truncation function $\lfloor x\rfloor_B$ to be x if $x>B$ and 0 otherwise. Since the complement in the fundamental domain of the cusp neighborhoods is compact, it suffices to prove the statement of the lemma with $\lfloor H\rfloor_B$ in the place of H. Note that

$$\int_0^1 \lfloor H(x+\tfrac{i}{n})\rfloor_B^\lambda\,dx \;=\; \sum_{j=1}^h\sum_{\gamma\in\langle T\rangle\backslash\Gamma_j}\int_0^1\lfloor\Im(\gamma(x+\tfrac{i}{n}))\rfloor_B^\lambda\,dx\,. \tag{14}$$

Suppose that $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$ gives a non-zero contribution to the right hand side. Recall that Shimizu's Lemma states that in a Fuchsian group, normalized so that the cusp ∞ has width 1, either $c=0$ or $1\le|c|$. Here the truncation implies that $c\ne0$ and that

$$1\le\frac{1/n}{|(c(x+\frac{i}{n})+d|^2}\,,$$

from which

$$1\le|c|\le\sqrt{n}\quad\text{and}\quad -\frac{d}{c}\in[-1,2].$$

On the other hand, with the substitution $x=-\frac{d}{c}+\frac{t}{n}$ we get

$$\begin{aligned}\int_0^1\lfloor\Im(\gamma(x+\tfrac{i}{n}))\rfloor_B^\lambda\,dx &\le \int_{-\infty}^\infty\left(\frac{1/n}{(cx+d)^2+c^2/n^2}\right)^\lambda dx\\ &\le \frac{n^{\lambda-1}}{|c|^{2\lambda}}\int_{-\infty}^\infty\frac{dt}{(t^2+1)^\lambda} \;=\; \mathrm{O}\Big(\frac{n^{\lambda-1}}{|c|^{2\lambda}}\Big)\,.\end{aligned}$$

(It is this estimate which has to be changed, taking into account $\lfloor\,\cdot\,\rfloor_B$, when $\lambda \leq \frac{1}{2}$.) We define

$$C^j_{\mu,\nu}(x,X) \;=\; \left\{\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \langle T\rangle\backslash\Gamma_j \,:\, x < |c| \leq X,\; -\frac{d}{c} \in [\mu,\nu]\right\}$$

and $C^j_{\mu,\nu}(X) = C^j_{\mu,\nu}(0,X)$. The crucial observation now is that the cardinality of this set is bounded for $X > 1$, any j and $\mu < \nu$ by

$$\#C^j_{\mu,\nu}(X) \leq (\nu-\mu)X^2 + 1.$$

This again follows from Shimizu's Lemma, which implies that for two matrices $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $\gamma' = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$ in $C^j_{a,b}(X)$ we have $|\frac{d'}{c'} - \frac{d}{c}| \geq |cc'|^{-1} \geq X^{-2}$ (see [19], Proposition 2.8, for details).

If $\lambda = 1$, then the contribution of the j-th cusp to the right hand side of (14) is bounded above by a constant times

$$\begin{aligned}\sum_{\gamma\in C^j_{-1,2}(\sqrt{n})} \frac{1}{|c|^2} \;&\leq\; \sum_{k=0}^{\frac{1}{2}\log_2 n} \;\sum_{\gamma\in C^j_{-1,2}(2^{k-1},2^k)} \frac{1}{|c|^2} \\ &\leq\; \sum_{k=0}^{\frac{1}{2}\log_2 n} \frac{3\cdot 2^{2k}}{2^{2k-2}} \;=\; \mathrm{O}(\log n).\end{aligned}$$

The other cases with $\lambda > 1/2$ are calculated the same way, the estimate for the left-hand side of (14) now being $\mathrm{O}\bigl(n^{\lambda-1}\sum_{k\leq\frac{1}{2}\log_2 n} 2^{2k(1-\lambda)}\bigr)$, which is $\mathrm{O}(n^{\lambda-1})$ for $\lambda > 1$ and $\mathrm{O}(1)$ for $\lambda < 1$. $\square$

3. Dimensions and degrees

Twisted modular forms of bi-weight (k,ℓ) can be thought of as sections of the bundle $\mathcal{L}^{\otimes k}\otimes\widetilde{\mathcal{L}}^{\otimes \ell}$ with appropriate growth conditions at the cusps, where $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ are the line bundles over $\mathbb{H}/\Gamma$ defined as the quotients of $\mathbb{H}\times\mathbb{C}$ by the equivalence relations $(z,u)\sim(\gamma z, J(\gamma,z)u)$ and $(z,u)\sim(\gamma z,\widetilde{J}(\gamma,z)u)$ for $\gamma\in\Gamma$. The dimension of the space of such forms for a given group can therefore be computed by the Riemann-Roch theorem for curves, just as in the case of classical modular forms, if we know the degrees of the two bundles $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ and the numbers of cusps and elliptic fixed points of various orders of the group. In particular, if Γ had no cusps and no fixed points (a situation of which, so far as we know, there is no example), then the Riemann-Roch theorem would give $\dim M_{k,\ell}(\Gamma) = (k\deg(\mathcal{L}) + \ell\,\deg(\widetilde{\mathcal{L}}))(g-1)$, where g is the genus of $\mathbb{H}/\Gamma$. The presence of cusps (including possibly irregular ones) and elliptic fixed points will make the actual formula more complicated.

Let Π be a torsion-free subgroup of finite index in Γ. Such a group always exists, since the level three subgroup of the Hilbert modular group is torsion-free and Γ is a subgroup of the Hilbert modular group. We define Π_0 to be a subgroup of finite index such that the eigenvalues of all parabolic elements are one (i.e. all cusps are regular). Such a subgroup exists, since Π is free if it is has a cusp. By passing to a smaller subgroup if necessary, we may suppose $\Pi_0\subset\Gamma$ to be normal. We let $\mathcal{L}_0$ and $\widetilde{\mathcal{L}}_0$ be the line bundles over $\mathbb{H}/\Pi_0$ defined by the automorphy factors J and $\widetilde{J}$ respectively.

The basic invariant we attach to a Hilbert modular embedding is the ratio

$$\lambda_2 \;=\; \deg(\widetilde{\mathcal{L}}_0)\,/\,\deg(\mathcal{L}_0). \tag{15}$$

As a consequence of the proof of Theorem 3.2 below, we see that this number does not depend on the choice of Π_0 among torsion-free subgroups of Γ with regular cusps. The value of the invariant λ_2 is given by 1, 1/3, 1/5, 1/7 on the four classes of Hilbert modular embeddings described in Section 1.

The definition of λ_2 and the classical result that ordinary modular forms of weight 2 are differential forms on $\mathbb{H}/\Pi_0$ imply the following Proposition.

Proposition 3.1. *Let g_0 and s_0 denote the genus and the number of cusps of $\mathbb{H}/\Pi_0$, respectively. Then*

$$\deg(\mathcal{L}_0) = g_0 - 1 + s_0/2 \quad \textit{and} \quad \deg(\widetilde{\mathcal{L}}_0) = \lambda_2\,(g_0 - 1 + s_0/2)\,. \tag{16}$$

Now we have to pass from Π_0 to Γ, which may have both elliptic fixed points and irregular cusps. If Γ contains $-I$, then we will assume that $k+\ell$ is even, since otherwise the equation $f|_{(k,\ell)}(-I) = -f$ implies that the space of twisted modular forms of bi-weight (k,ℓ) is 0.

We define *characteristic numbers* at elliptic fixed points and cusps for the bundle of twisted modular forms of bi-weight (k,ℓ) in the following way. Suppose that x is an elliptic fixed point and that the isotropy group Γ_x is of order n_x in $\mathrm{SL}(2,\mathbb{R})$. We take a generator $\gamma = \left(\begin{smallmatrix} * & * \\ c & * \end{smallmatrix}\right) \in \Gamma_x$ that acts on the tangent space at x by a rotation by $2\pi/n_x$ in the positive direction, i.e. such that $\arccos(\mathrm{tr}(\gamma)/2) = 2\pi/n_x$ and $c\,\sin(2\pi/n_x) \le 0$. We let $r_x = \frac{1}{n_x}$ and define $r_x^\sigma \in \frac{1}{n_x}\mathbb{Z}$ by

$$\cos(2\pi r_x^\sigma) \;=\; \mathrm{tr}(\gamma^\sigma)/2 \quad \text{and} \quad c^\sigma \sin(2\pi r_x^\sigma) \;\le\; 0\,.$$

Then the characteristic number at x is defined as

$$b_x(k,\ell) = \left\{-k r_x - \ell r_x^\sigma\right\}, \tag{17}$$

where the curly braces denote the fractional part (in $[0,1)$) of the rational number. If x is a cusp, we define the characteristic number $b_x(k,\ell)$ to be 1/2 if the cusp is irregular, $-I \notin \Gamma$, and $k+\ell$ is odd, and we let $b_x(k,\ell) = 0$ in all other cases.

We remark that characteristic numbers are a finer information than the usual *type* of the elliptic fixed point ([35], Section I.5 and V.7), since there are two possibilities even for fixed points of order two (in $\mathrm{PSL}(2,\mathbb{R})$). We let $\Delta = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \sqrt{D} \end{smallmatrix}\right)$ and $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathcal{O}_D)$. Then the contribution of $\Delta S \Delta^{-1} \in \mathrm{SL}(\mathcal{O}_D^\vee \oplus \mathcal{O}_D)$ is $b_x(k,\ell) = \left\{\frac{-k+\ell}{4}\right\}$, whereas the contribution of S is $b_x(k,\ell) = \left\{\frac{-k-\ell}{4}\right\}$. Note that in all these calculation we consider modular embeddings to $\mathbb{H}^2$. If we consider a modular embedding to $\mathbb{H}\times\mathbb{H}^-$, the fixed point of S is $-i$ in the second factor, so its contribution is $b_x(k,\ell) = \left\{\frac{-k+\ell}{4}\right\}$.

We can now give the dimension of the space of modular forms in terms of the topology of $\mathbb{H}/\Gamma$ and those characteristic numbers. We let $\overline{n_x}$ be the order of the isotropy group Γ_x in $\mathrm{PSL}(2,\mathbb{R})$.

Theorem 3.2. *Let k and ℓ be integers. If $k+\ell$ is odd and Γ contains $-I$, then $\dim M_{k,\ell}(\Gamma,\varphi) = 0$. If $-I \notin \Gamma$ or $k+\ell$ is even, then twisted modular forms of* *bi-weight (k,ℓ) are precisely the global sections of a line bundle $\mathcal{L}_{k,\ell}$ of degree*

$$\deg(\mathcal{L}_{k,\ell}) \;=\; (k+\lambda_2\ell)\Big(g-1+\frac{s}{2}+\frac{1}{2}\sum_{x\in\mathbb{H}/\Gamma}\big(1-\frac{1}{\overline{n_x}}\big)\Big) \;-\; \sum_{x\in\overline{\mathbb{H}/\Gamma}} b_x(k,\ell)\,,$$

where g and s denote the genus and the number of cusps of Γ. If also $k+\lambda_2\ell\geq 2$, then

$$\dim M_{k,\ell}(\Gamma,\varphi) \;=\; \deg(\mathcal{L}_{k,\ell}) \;+\; (1-g)\,.$$

Proof. We mimic the standard argument for ordinary modular forms and describe $\mathcal{L}_{k,\ell}$ as a subsheaf of $\mathcal{L}^{\otimes k}\otimes\widetilde{\mathcal{L}}^{\otimes \ell}$. If t is a local parameter at x, for x both in $\mathbb{H}/\Gamma$ or being a cusp, the stalk $(\mathcal{L}_{k,\ell})_x$ at consists of all germs of holomorphic functions f with $f(\gamma t) = J(\gamma,t)^k\widetilde{J}(\gamma,\varphi(t))^\ell f(t)$ for all $\gamma\in\Gamma_x$ in the stabilizer of x. With this definition, twisted modular forms of bi-weight (k,ℓ) are obviously the global sections of $\mathcal{L}_{k,\ell}$.

In order to compute the degree of $\mathcal{L}_{k,\ell}$ we use the map $\pi:\mathbb{H}/\Pi_0\to\mathbb{H}/\Gamma$. This induces an inclusion $\pi^*(\mathcal{L}_{k,\ell})\to\mathcal{L}_0^k\mathcal{L}_0^\ell$. Since we know the degree of the image in terms of g, s and λ_2, it suffices to compute the degree of its cokernel $\mathcal{K}$. This cokernel is supported at the elliptic fixed points and at the cusps.

Suppose first that $z_\gamma\in\mathbb{H}$ is the fixed point of $\gamma\in\Gamma$. The point $\varphi(z_\gamma)$ is fixed by γ^σ, so that by the cocycle condition both $J(\gamma,z_\gamma)$ and $\widetilde{J}(\gamma,\varphi(z_\gamma))$ are roots of unity of some order that divides the order of the isotropy group n_x.

Now let y be one of the preimages of x and let u be a local parameter at y, so that $t=u^{\overline{n_x}}$. Then

$$(\mathcal{L}_0^k\widetilde{\mathcal{L}}_0^\ell)_y\cong\mathbb{C}[[u]] \quad\text{and}\quad (\pi^*(\mathcal{L}_{k,\ell}))_y = u^{\overline{n_x}B_x(k,\ell)}\,\mathbb{C}[[u]],$$

where $B_x(k,\ell)\in[0,1)$ is the unique rational number such that

$$J(\gamma,z_\gamma)^k\widetilde{J}(\gamma,\varphi(z_\gamma))^\ell \;=\; \mathbf{e}(\,B_x(k,\ell))\,.$$

Consequently, $\dim\mathcal{K}_y=\overline{n_x}\,B_x(k,\ell)$ for each of the $\deg(\pi)/\overline{n_x}$ points y above x.

Next we want to show that $B_x(k,\ell)=b_x(k,\ell)$. If $x=i$, then the generator of Γ_x specified above is $\gamma=\left(\begin{smallmatrix}\cos(2\pi/n_x) & \sin(2\pi/n_x)\\ -\sin(2\pi/n_x) & \cos(2\pi/n_x)\end{smallmatrix}\right)$ and hence $J(\gamma_1,i)^k=\mathbf{e}(-kr_x)$ and $\widetilde{J}(\gamma_1,i)^\ell=\mathbf{e}(-\ell r_x^\sigma)$ by definition of r_x and r_x^σ. This proves the claim in the special case $x=i$. For the general case note first that for any $\alpha\in\mathrm{SL}(2,\mathbb{R})$ the cocycle property implies $J(\gamma,z_\gamma)=J(\alpha\gamma\alpha^{-1},z_{\alpha\gamma\alpha^{-1}})$, where $z_{\alpha\gamma\alpha^{-1}}=\alpha z_\gamma$ is the fixed point of $\alpha\gamma\alpha^{-1}$. The equivariance property (1) implies that $\varphi(z_\gamma)$ is the fixed point of γ^σ and hence $\widetilde{J}(\gamma,z_\gamma)=J(\gamma^\sigma,z_{\gamma^\sigma})$. If α takes the fixed point of γ to i, then α^σ takes the fixed point of γ^σ to i and so

$$\begin{aligned} &J(\gamma,z_\gamma)^k\,\widetilde{J}(\gamma,\varphi(z_\gamma))^\ell = J(\gamma,z_\gamma)^k\,J(\gamma^\sigma,z_{\gamma^\sigma})^\ell\\ &= J(\alpha\gamma\alpha,i)^k\,J((\alpha^\sigma\gamma^\sigma\alpha^\sigma)^{-1},i)^\ell = J(\alpha\gamma\alpha,i)^k\,\widetilde{J}((\alpha\gamma\alpha)^\sigma,i)^\ell \end{aligned}\tag{18}$$

reduces to the case already considered.

Now suppose that x is a cusp and let y be one of the cusps of Π_0 above x. If $\overline{n_x}$ denotes the degree of the covering π at y, then there are $\deg(\pi)/\overline{n_x}$ cusps above x since π is Galois.

We start with the case $-I\notin\Gamma$. Then the stabilizer Γ_x is infinite cyclic. Let γ be a generator. The same argument as for (18) allows us to assume that the fixed point $z_\gamma = \infty$. Note that γ^σ also fixes ∞, so that $J(\gamma,\infty) = \widetilde{J}(\gamma,\varphi(\infty)) = \gamma_{2,2}$, the lower right entry of γ. Since the cusp is irregular if and only if the generator γ has $\gamma_{2,2} = -1$, we deduce

$$J(\gamma,\infty)^k\,\widetilde{J}(\gamma,\varphi(\infty))^\ell = \mathbf{e}(b_x(k,\ell))$$

for $b_x(k,\ell)$ defined above. On the other hand, let a be the width of the cusp ∞ of Γ_0, so that $q = \mathbf{e}(z/a)$ is a local parameter at y. Then

$$(\mathcal{L}_0^k\widetilde{\mathcal{L}}_0^\ell)_y \cong \mathbb{C}[[q]] \quad \text{and} \quad (\pi^*(\mathcal{L}_{k,\ell})_y = q^{\overline{n_x}b_x(k,\ell)}\,\mathbb{C}[[q]],$$

so that in total $\dim \mathcal{K}_y = \overline{n_x}\,b_x(k,\ell)$.

With the same local calculation one checks that if $-I \in \Gamma$ always $(\pi^*(\mathcal{L}_{k,\ell})_y = (\mathcal{L}_0^k\widetilde{\mathcal{L}}_0^\ell)_y$. Hence in this case, too, $\dim \mathcal{K}_y = \overline{n_x}\,b_x(k,\ell) = 0$ holds by definition.

Altogether, this implies

$$\deg(\mathcal{L}_{k,\ell}) \;=\; \frac{1}{\deg(\pi)}\Big((k+\lambda_2\ell)\deg(\mathcal{L}_0)\Big) - \sum_{x\in\overline{\mathbb{H}/\Gamma}} b_x(k,\ell).$$

The number of cusps of Γ is $s_0 = \deg(\pi)\,\sum_{x\in\partial(\mathbb{H}/\Gamma)}\frac{1}{\overline{n_x}}$. Together with (16) and the Riemann-Hurwitz formula

$$\frac{g(\mathbb{H}/\Pi_0)-1}{\deg(\pi)} = g(\mathbb{H}/\Gamma) - 1 + \frac{1}{2}\sum_{x\in\overline{\mathbb{H}/\Gamma}}\Big(1-\frac{1}{\overline{n_x}}\Big)$$

this implies the degree claim. The dimension statement then follows from Riemann-Roch. □

4. Differential equations coming from twisted modular forms

A basic fact about modular forms, whose proof will be recalled below, is that for any Fuchsian group $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$, any modular function t on Γ and any modular form f of integral weight $k \geq 1$ on Γ, the function $y(t)$ defined locally by $f(z) = y(t(z))$ satisfies a linear differential equation of order $k+1$ with algebraic coefficients (and even with polynomial coefficients if $\mathbb{H}/\Gamma$ has genus 0 and t is a hauptmodule[2]). In this subsection we prove the corresponding statement for twisted modular forms. This statement will give one of the two approaches used in this paper to describe Teichmüller curves explicitly on Hilbert modular surfaces, by comparing the differential equations coming from their geometric definition (Picard-Fuchs differential equations) with the differential equations satisfied by suitable twisted modular forms on them.

Theorem 4.1. *Let $f(z)$ be a twisted modular form on (Γ,φ) of bi-weight (k,ℓ), with $k,\ell \geq 0$, and $t(z)$ a modular function with respect to the same group Γ. Then the function $y(t)$ defined locally by $f(z) = y(t(z))$ satisfies a linear differential equation of order $(k+1)(\ell+1)$ with algebraic coefficients.*

Proof. It suffices to prove this for the two cases $(k,\ell) = (1,0)$ and $(0,1)$, since the general case follows from these. (The number $(k+1)(\ell+1)$ arises as the dimension of $\mathrm{Sym}^k(\mathrm{V}_1)\otimes\mathrm{Sym}^k(\mathrm{V}_2)$ where $\dim V_1 = \dim V_2 = 2$.) The first case is the classical theorem mentioned above, of which several proofs are known (see e.g. §5.3 of [6]).

[2]Recall that a "hauptmodule" (or "Hauptmodul" if one retains the German spelling) is a modular function t giving an isomorphism between $\overline{\mathbb{H}/\Gamma}$ and $\mathbb{P}^1(\mathbb{C})$ if the former has genus 0.

We reproduce one of them here since it generalizes directly to the more complicated case of bi-weight $(0,1)$.

Let, then, $f(z)$ be an ordinary modular form of weight 1 and $t(z)$ a modular function on Γ. By definition we have the two transformation equations

$$t\Big(\frac{az+b}{cz+d}\Big) \;=\; t(z)\,, \qquad f\Big(\frac{az+b}{cz+d}\Big) \;=\; (cz+d)\,f(z)$$

for all matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. Differentiating these equations gives the further transformation equations

$$\begin{aligned}
t'\Big(\frac{az+b}{cz+d}\Big) &= (cz+d)^2 t'(z)\,,\\
f'\Big(\frac{az+b}{cz+d}\Big) &= (cz+d)^3 f'(z) \,+\, c(cz+d)^2\, f(z)\,,\\
f''\Big(\frac{az+b}{cz+d}\Big) &= (cz+d)^5 f''(z) \,+\, 4c(cz+d)^4 f'(z) \,+\, 2c^2(cz+d)^3 f(z)\,.
\end{aligned}$$

The first of these equations says that t' is a (meromorphic) modular form of weight 2, and by combining the others we find that the expression $2f'^2 - ff''$ is a modular form of weight 6. It follows that

$$\frac{t'(z)}{f(z)^2} \;=\; A\big(t(z)\big), \qquad \frac{2f'(z)^2 - f(z)f''(z)}{t'(z)f(z)^4} \;=\; B\big(t(z)\big) \tag{19}$$

for some rational (or, if t is not a hauptmodule, algebraic) functions $A(t)$ and $B(t)$. A direct calculation shows that

$$\frac{1}{t'}\bigg(\frac{t'}{f^2}\,\frac{1}{t'}\,f'\bigg)' \;+\; \frac{2f'^2 - ff''}{t'f^4}\,f \;=\; 0\,,$$

It follows that the function $y(t)$ defined parametrically by the equation $y(t(z)) = f(z)$ (which of course can only hold locally, since $t(z)$ is Γ-invariant and $f(z)$ isn't) satisfies the second order linear differential equations

$$\big(A(t)\,y'(t)\big)' \;+\; B(t)\,y(t) \;=\; 0\,, \tag{20}$$

or $Ay'' + A'y' + By = 0$. This proves the theorem in the case $(k,\ell) = (1,0)$.

Now suppose that f is a twisted modular form of bi-weight (0,1), i.e., f satisfies the transformation equation $f\big(\frac{az+b}{cz+d}\big) = (c^\sigma\varphi(z) + d^\sigma)\,f(z)$ for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. From this equation, and from equation (1) and its derivative (= second equation in (9)), we find by further differentiating the transformation equations

$$\begin{aligned}
\varphi''\Big(\frac{az+b}{cz+d}\Big) &= \frac{(cz+d)^4}{(c^\sigma\varphi(z)+d^\sigma)^2}\varphi''(z) \,+\, \frac{2c(cz+d)^3}{(c^\sigma\varphi(z)+d^\sigma)^2}\,\varphi'(z)\\
&\quad -\, \frac{2c^\sigma(cz+d)^4}{(c^\sigma\varphi(z)+d^\sigma)^3}\,\varphi'(z)^2\,,\\
f'\Big(\frac{az+b}{cz+d}\Big) &= (cz+d)^2(c^\sigma\varphi(z)+d^\sigma)\,f'(z) \,+\, c^\sigma(cz+d)^2\varphi'(z)f(z)\,,\\
f''\Big(\frac{az+b}{cz+d}\Big) &= (cz+d)^4(c^\sigma\varphi(z)+d^\sigma)\,f''(z)\\
&\quad +\, \big[2c(cz+d)^3(c^\sigma\varphi(z)+d^\sigma) \,+\, 2c^\sigma(cz+d)^4\varphi'(z)\big]\,f'(z)\\
&\quad +\, \big[c^\sigma(cz+d)^4\varphi''(z) \,+\, 2cc^\sigma(cz+d)^3\varphi'(z)\big]\,f(z)\,.
\end{aligned}$$

From these equations it follows that the combination $(2f'^2 - ff'')\varphi' + ff'\varphi''$ is a modular form of weight 6. But we have already seen that t' and φ' are twisted modular of bi-weights $(2,0)$ and $(2,-2)$, respectively. It follows that

$$\begin{aligned} \frac{t'(z)}{\varphi'(z)f(z)^2} &= A\bigl(t(z)\bigr), \\ \frac{(2f'(z)^2 - f(z)f''(z))\,\varphi'(z) + f(z)f'(z)\varphi''(z)}{t'(z)\varphi'(z)^2 f(z)^4} &= B\bigl(t(z)\bigr) \end{aligned} \tag{21}$$

for some algebraic (resp. rational if t is a hauptmodule) functions $A(t)$ and $B(t)$, and since by direct calculation we have

$$\frac{1}{t'}\Bigl(\frac{t'}{\varphi' f^2}\,\frac{1}{t'}\,f'\Bigr)' + \frac{(2f'^2 - ff'')\varphi' - ff'\varphi''}{t'\varphi'^2 f^4}\,f = 0$$

in this case, we deduce that f satisfies a second order linear differential equation of the same form (20) as before. □

Remark. The two weight 6 modular forms $2f'^2 - ff''$ (for $f \in M_{1,0}(\Gamma)$) and $(2f'^2 - ff'')\varphi' + ff'\varphi''$ (for $f \in M_{0,1}(\Gamma)$) used above, which are easily checked to be holomorphic at the cusps, are special cases of the classical Rankin-Cohen bracket and of a twisted version of it, respectively. Without going into details, we mention that the twisted Rankin-Cohen brackets of two twisted modular forms $f_i \in M_{k_i,\ell_i}(\Gamma,\varphi)$ $(i = 1, 2)$ can be defined as the product of the usual Rankin-Cohen brackets of ${\varphi'}^{\ell_1/2} f_1$ and ${\varphi'}^{\ell_2/2} f_2$ (which by example i) of Section 2 are ordinary modular forms of weight $k_1 + \ell_1$ and $k_2 + \ell_2$ on Γ) with a suitable power of φ'.

Part II: Modular embeddings via differential equations

In Section 4 we have seen abstractly how classical or twisted modular forms give rise to differential equations. In Part II we show conversely, in a specific example, how to obtain from these differential equations the Hilbert modular embedding φ. The example that we will consider in detail is $D = 17$, for which the differential equations needed were computed in [4]. In Section 6 we will sketch how these were obtained, referring to that paper for the full details. In Section 7 we discuss the arithmetical properties of the solutions of these differential equations and compute the Fourier expansions of the corresponding modular forms at all cusps. We turn Theorem 3.2 into a concrete description of the ring of modular forms (Theorem 5.5 for the general result and Propositions 7.1 and 7.2 for the special case $D = 17$), since the corresponding local invariants can be computed for Teichmüller curves. In Section 8 we will show how to use these solutions to obtain an explicit embedding of the Teichmüller curve in the Hilbert modular surface. The introductory Section 5 provides the necessary background on Hilbert modular surfaces and Teichmüller curves in genus 2.

5. Curves on Hilbert modular surfaces

As we have already said, there are two basic examples of the situation described in Section 1: modular curves and Teichmüller curves. In this section we describe both of these, the first relatively briefly since it is well known and the second in more detail. We begin with a preliminary subsection specifying more precisely the Hilbert modular surfaces that will be used in this paper. The main new result in this section is the dimension formula Theorem 5.5.

5.1. **Hilbert modular groups and Hilbert modular surfaces.** As before, we denote by K be a real quadratic field, together with a fixed embedding $K \subset \mathbb{R}$, and denote by σ both the Galois conjugation and the second embedding of K into $\mathbb{R}$. In §1 we briefly defined Hilbert modular groups and Hilbert modular surfaces, denoting them generically by Γ_K and $\mathbb{H}^2/\Gamma_K$. Now we want to be more specific. Our general reference are Hirzebruch's seminal paper [15] and the book [35] by van der Geer.

Usually when one speaks of "the" Hilbert modular group for K one means the group $\mathrm{SL}(2, \mathcal{O}_D)$, where D is the discriminant of an order $\mathcal{O} = \mathcal{O}_D \subset K$. However, since we want principally polarized abelian surfaces, we need to work instead with the modified Hilbert modular group

$$\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O}) \;=\; \begin{pmatrix} \mathcal{O} & \mathcal{O}^\vee \\ (\mathcal{O}^\vee)^{-1} & \mathcal{O} \end{pmatrix} \;\cap\; \mathrm{SL}(2, K)\,,$$

where $\mathcal{O}^\vee$ denotes the set of $x \in K$ for which xy has integral trace for all $y \in \mathcal{O}$. One has $\mathcal{O}^\vee = \mathfrak{d}^{-1}$, where in the case of quadratic fields the ideal $\mathfrak{d}$, called the *different* of K, is simply the principal ideal $(\sqrt{D}) = \sqrt{D}\,\mathcal{O}$. Note that the two groups $\mathrm{SL}(2, \mathcal{O})$ and $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$ are conjugate in $\mathrm{GL}(2, K)$ by the action of the diagonal matrix $\Delta = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \sqrt{D} \end{smallmatrix}\right)$, and in particular are isomorphic as abstract groups. But the action of Δ interchanges the upper and lower half-planes in the second factor (since the Hilbert modular group acts on the second factor via its Galois conjugates and the Galois conjugate of the determinant of Δ is negative), so the quotient is the Hilbert modular surface

$$X_D \;=\mathbb{H}^2/\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O}),$$

which is isomorphic to $X_{\mathcal{O}}^- = \mathbb{H}\times\mathbb{H}^-/\mathrm{SL}(2,\mathcal{O})$ and not in general isomorphic to the "standard" Hilbert modular surface $X_{\mathcal{O}} = \mathbb{H}^2/\mathrm{SL}(2,\mathcal{O})$. (They do not even necessarily have the same Euler characteristic.) If $\mathcal{O}$ contains a unit ε of negative norm, which happens, for instance, when D is prime, then $\mathfrak{d}$ is principal in the narrow sense and the varieties $X_{\mathcal{O}}$ and $X_{\mathcal{O}}^-$ are isomorphic via $(z_1,z_2)\mapsto(\varepsilon z_1,\,\varepsilon' z_2)$.

To a point $\mathbf{z}=(z_1,\,z_2)\in\mathbb{H}^2$ we associate the polarized abelian surface $A_{\mathbf{z}} = \mathbb{C}^2/\mathfrak{L}_{\mathbf{z}}$, where $\mathfrak{L}_{\mathbf{z}}\subset\mathbb{C}^2$ is the lattice

$$\mathfrak{L}_{\mathbf{z}} \;=\; \bigl\{(az_1\,+\,b,\,a^\sigma z_2+b^\sigma)\mid a\in\mathcal{O}^\vee,\;\, b\in\mathcal{O})\bigr\}\,, \tag{22}$$

with the action of $\mathcal{O}$ on $A_{\mathbf{z}}$ induced from the action $\lambda(v_1,v_2)=(\lambda v_1,\lambda^\sigma v_2)$ of $\mathcal{O}$ on $\mathbb{C}^2$ and with the polarization induced from the antisymmetric pairing

$$\langle\,(a,b),\,(a',b')\,\rangle = \mathrm{tr}_{K/\mathbb{Q}}\bigl(ab'-a'b\bigr) \qquad (a,\,a'\in(\mathcal{O}^\vee)^{-1},\;\,b,\,b'\in\mathcal{O})\,. \tag{23}$$

This pairing is unimodular and the polarization is principal, which is why that case is of special interest.

We observe that the action of $\mathcal{O}$ on $A_{\mathbf{z}}$ gives a canonical splitting of the 2-dimensional space of holomorphic 1-forms on A into two 1-dimensional eigenspaces, generated by the differential forms $\omega = dv_1$ and $\widetilde{\omega} = dv_2$, which we will call the *first* and *second eigendifferential*, respectively. If $A_{\mathbf{z}}$ is the Jacobian of a curve C of genus 2, then by the canonical identification of the spaces of holomorphic 1-forms on C and on $A_{\mathbf{z}}$ we obtain corresponding eigendifferentials on C. These will be used in the definition of Teichmüller curves in §5.3.

Since the isomorphism class of $A_{\mathbf{z}}$ depends only on the image of $\mathbf{z}$ in X_D, and since polarized abelian surfaces are parametrized by points in the quotient of the Siegel upper half-space $\mathbb{H}_2$ by $\mathrm{Sp}(4,\mathbb{Z})$ we get an embedding of the Hilbert modular surface X_D into the quotient $\mathbb{H}_2/\mathrm{Sp}(4,\mathbb{Z})$, a *Siegel modular embedding.* Explicitly, let ψ be the map from $\mathbb{H}^2$ to $\mathbb{H}_2$ given by

$$\psi:\quad \mathbf{z}\;=\;(z_1,z_2)\;\mapsto\; B\begin{pmatrix} z_1 & 0\\ 0 & z_2\end{pmatrix}B^T \tag{24}$$

where, for some $\mathbb{Z}$-basis (ω_1,ω_2) of $\mathcal{O}$ we let

$$B=\begin{pmatrix}\omega_1 & \omega_1^\sigma\\ \omega_2 & \omega_2^\sigma\end{pmatrix}\qquad\text{and}\qquad A=B^{-1}\,. \tag{25}$$

We define a homomorphism $\Psi:\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O})\to\mathrm{Sp}(4,\mathbb{Z})$ by

$$\Psi:\quad \begin{pmatrix} a & b\\ c & d\end{pmatrix}\mapsto\begin{pmatrix} B^T & 0\\ 0 & A\end{pmatrix}\begin{pmatrix}\widehat{a} & \widehat{b}\\ \widehat{c} & \widehat{d}\end{pmatrix}\begin{pmatrix} A^T & 0\\ 0 & B\end{pmatrix}, \tag{26}$$

where $\widehat{a}$ for $a\in K$ denotes the diagonal matrix $\mathrm{diag}(a,a^\sigma)$. Then the map ψ is equivariant with respect to the actions of $\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O})$ on $\mathbb{H}^2$ and of $\Psi(\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O}))$ on $\mathbb{H}_2$, so it induces a map, also denoted by ψ, on the level of quotient spaces.

Remark on notation: We will use the letters ψ and Ψ for Siegel modular embeddings and φ and Φ for Hilbert modular embeddings. The capital letter will denote the map on the level of modular groups and the small letter the map on the level of symmetric spaces or quotient spaces.

More generally, for any invertible $\mathcal{O}$-ideal $\mathfrak{a}$ the trace pairing (23) on the lattice $\mathfrak{a}^\vee \oplus \mathfrak{a}$ is unimodular and consequently, the abelian surface $\mathfrak{L}_{\mathbf{z},\mathfrak{a}}$, defined as in (22) with $\mathcal{O}$ replaced by $\mathfrak{a}$, is principally polarized. This implies that the Hilbert modular surfaces

$$X_{D,\mathfrak{a}} \;=\; \mathbb{H}/\mathrm{SL}(\mathfrak{a}^\vee \oplus \mathfrak{a}),$$

where

$$\mathrm{SL}(\mathfrak{a}^\vee \oplus \mathfrak{a}) \;=\; \begin{pmatrix} \mathcal{O} & \mathfrak{a}^\vee(\mathfrak{a})^{-1} \\ \mathfrak{a}(\mathfrak{a}^\vee)^{-1} & \mathcal{O} \end{pmatrix} \,\cap\, \mathrm{SL}(2,K)\,, \tag{27}$$

also parametrize principally polarized abelian varieties with real multiplication by $\mathcal{O}$. The only difference is that now cusp at ∞ is in general a different one than for X_D. We will use these variants $X_{D,\mathfrak{a}}$ when we discuss cusps of Hilbert modular surface in Section 10. If we construct B using some $\mathbb{Z}$-basis (ω_1,ω_2) of $\mathfrak{a}$, then (24) defines a map ψ that is equivariant with respect to a homomorphism

$$\Psi : \mathrm{SL}(\mathfrak{a}^\vee \oplus \mathfrak{a}) \to \mathrm{Sp}(4,\mathbb{Z}) \tag{28}$$

given by the same definition (25). Hence the pair (ψ,Ψ) defines a Siegel modular embedding of $X_{D,\mathfrak{a}}$.

5.2. **Modular curves.** We already defined the modular curves in Section 1 as the quotients of $\mathbb{H}$ by subgroups of $\mathrm{SL}(2,\mathbb{R})$ of the form $\Gamma_A = \{\gamma \in \Gamma_K \mid A\gamma = \gamma^\sigma A\}$ where A is "a suitable element" of $\mathrm{GL}(2,K)$, embedded into appropriate Hilbert modular surfaces via $z \mapsto (z, Az)$. Here "suitable" means that the adjoint of A equals its Galois conjugate, so that $A = \left(\begin{smallmatrix} \lambda^\sigma & -b\sqrt{D} \\ a\sqrt{D} & \lambda \end{smallmatrix}\right)$ for some $(a,b,\lambda) \in \mathbb{Q}\times\mathbb{Q}\times K$, which after multiplying A by a suitable scalar in $\mathbb{Q}^\times$ (which does not change the definition of Γ_A) we can assume belongs to $\mathbb{Z}\times\mathbb{Z}\times\mathcal{O}_K$. The corresponding embedded curve in $\mathbb{H}^2$ is defined as $az_1z_2 + \nu^\sigma z_1 + \nu z_2 + b = 0$, where $\nu = \lambda/\sqrt{D} \in \mathcal{O}^\vee = \mathfrak{d}^{-1}$, and the union of these curves (or rather, of their images in the Hilbert modular surface) when A ranges over all matrices as above with given determinant $N \in \mathbb{N}$ is denoted by T_N. These curves T_N were studied in detail (for the Hilbert modular surface $X_\mathcal{O}$) in [16] and [17]. We recall a few results that we will use. The curve T_N is non-empty if and only if N is congruent modulo D to the norm of an element of $\mathcal{O}$, and is non-compact (i.e., passes through the cusps) if and only if N is the norm of an integral ideal $\mathfrak{a}$ of K, in which case each of its components is non-compact. It is not in general irreducible, for three reasons. First, we have $T_N = \bigcup_{d^2|N} F_{N/d^2}$, where F_N is defined like T_N but with the additional requirement that (a,b,ν) is primitive in the lattice $\mathbb{Z}\times\mathbb{Z}\times\mathfrak{d}^{-1}$. Secondly, the F_N are in general not irreducible either, but decompose as $\bigcup_\alpha F_N(\alpha)$, where α ranges over the elements of $\mathfrak{d}^{-1}/\mathcal{O}$ with $N(\alpha) \equiv N/D \pmod 1$ and $F_N(\alpha)$ is defined by requiring $\nu \equiv \alpha \pmod{\mathcal{O}}$ ([39], p. 4, Remark 1). Finally, even the $F_N(\alpha)$ need not be irreducible. (For instance, if $D = p$ is prime and $p^2|N$, then the two Legendre symbols (a/p) and (b/p) cannot both vanish or have opposite values, so $F_N = F_N(0)$ has two components distinguished by the invariant $\varepsilon \in \{\pm 1\}$ defined by $(a/p) = \varepsilon$ or $(b/p) = \varepsilon$; cf. [11].) However, if N is "admissible" in the sense of the proposition on p. 57 of [17] (i.e., N is the norm of a primitive ideal in the principal genus), then that proposition says that each $F_N(\alpha)$ is irreducible, as one shows by counting the number of cusps of each component separately and of the whole curve F_N.

The same results apply to the curves in $X_\mathcal{O}^-$ defined (and denoted) in the same way but with the determinant of A now being $-N$ rather than N. Using the identification $(z_1, z_2) \mapsto (\sqrt{D}z_1, -\sqrt{D}z_2)$ of $X_{\mathcal{O}}^-$ with X_D, we can consider these as curves in X_D, the defining equation now being

$$ADz_1z_2 + \lambda^\sigma z_1 + \lambda z_2 + B \;=\; 0 \qquad (A,\, B \in \mathbb{Z},\ \lambda \in \mathcal{O},\ ABD - \lambda\lambda^\sigma = N). \quad (29)$$

A special union of these curves will be play a role a in characterization of Teichmüller curves below. In the moduli space of principally polarized abelian surfaces $\mathcal{A}_2$ we denote by P the *product locus* (also called *reducible locus*), i.e. the locus of abelian varieties that split, as polarized abelian varieties, into a product of elliptic curves. The Torelli map gives an isomorphism

$$t : \mathcal{M}_2 \to \mathcal{A}_2 \smallsetminus P\,.$$

The intersection of P with the Hilbert modular surface X_D will be denoted by P_D. It is a union of modular curves, as described in the following proposition.

Proposition 5.1. *The decomposition into irreducible components of P_D is given by*

$$P_D \;=\; \bigcup_{\nu\in\mathfrak{d}^{-1},\ \nu\gg0,\ \mathrm{Tr}(\nu)=1} P_{D,\nu} \;=\; \bigcup_{r\equiv D \ (\mathrm{mod}\ 2),\ |r|<2\sqrt{D}} P_{D,\frac{r+\sqrt{D}}{2\sqrt{D}}}\,,$$

where $P_{D,\nu}$ is the image in X_D of the curve $(\nu, \nu^\sigma)\mathbb{H} \subset \mathbb{H}^2$.

Proof. This is essentially Corollary 3.5 of [24], which states that

$$P_D = \bigcup_{N,\,r\in\mathbb{Z},\ N>0,\ D=r^2+4N} T_N\Big(\frac{r+\sqrt{D}}{2\sqrt{D}}\Big)\,. \quad (30)$$

Since each N occurring is admissible (it is the norm of the primitive principal ideal generated by $\frac{r+\sqrt{D}}{2}$), we have that $T_N(\nu) = F_N(\nu)$ is irreducible for each $\nu = \frac{r+\sqrt{D}}{2\sqrt{D}}$ and hence coincides with its subcurve $P_{D,\nu}$. $\square$

We would like to say a few words to explain where the equation $D = r^2 + 4N$ in (30) comes from. A point of P_D corresponds to a product $E \times E'$ of elliptic curves having real multiplication by $\mathcal{O}$, i.e., for which there is an endomorphism $\Phi = \left(\begin{smallmatrix} a & \lambda' \\ \lambda & b \end{smallmatrix}\right)$ of $E\times E'$ satisfying a quadratic equation of discriminant D over $\mathbb{Z}$. Since for generic points the elliptic curves E and E' do not have complex multiplication, we have $a \in \mathrm{End}(E) = \mathbb{Z}$, $b \in \mathrm{End}(E') = \mathbb{Z}$, and $\lambda\lambda' = \lambda'\lambda = N$ with $N = \deg\lambda \in \mathbb{N}$ and hence $\Phi^2 - (a+b)\Phi + (ab-N) = 0$, $D = (a+b)^2 - 4(ab-N) = (a-b)^2 + 4N$. We should also mention that the statement $P_D \subseteq \bigcup_{r^2+4N=D} T_N$ is just the special case $D' = 1$ of the general statement that the intersection of two Humbert surfaces $\mathcal{H}_D$ and $\mathcal{H}_{D'}$ in the moduli space $\mathcal{A}_2 = \mathbb{H}_2/\mathrm{Sp}(4,\mathbb{Z})$ is contained in the union of T_N with $DD' = r^2 + 4N$ for some $r \in \mathbb{Z}$, $N \in \mathbb{N}$. This statement is well known, and is given implicitly in the proof of Prop. XI.2.8, p. 215, of [35], but since we could not find a convenient reference and since the proof is easy, we give it here. We recall (cf. [35], Chapter IX) that the Humbert surface $\mathcal{H}_D$ is defined as the image in $\mathcal{A}_2$ of the union of the curves

$$\Big\{\left(\begin{smallmatrix} \tau_1 & \tau_2 \\ \tau_2 & \tau_3 \end{smallmatrix}\right) \in \mathbb{H}_2 \;\Big|\; a\tau_1 + b\tau_2 + c\tau_3 + d(\tau_2^2 - \tau_1\tau_3) + e \;=\; 0\Big\} \quad (31)$$

with $(a, b, c, d, e) \in \mathbb{Z}^5$, $b^2 - 4ac - 4de = D$. If D is a fundamental discriminant, then $\mathcal{H}_D$ is irreducible and hence can be given by any one of the equations in (31). The locus of products of elliptic curves in $\mathcal{A}_2$ is $\mathcal{H}_1$, because the standard embedding $(\mathbb{H}/\mathrm{SL}_2(\mathbb{Z}))^2 \to \mathbb{H}_2/\mathrm{Sp}_4(\mathbb{Z})$ is given by the equation $\tau_2 = 0$ in $\mathbb{H}_2$, which has the form (31) with $(a,b,c,d,e) = (0,1,0,0,0)$, $b^2-4ac-4de = 1$. The Hilbert modular surface X_D can be identified with $\mathcal{H}_D$, because if we write $N(x\omega_1+y\omega_2) = Ax^2+Bxy+Cy^2$ (A, B, $C \in \mathbb{Z}$, $B^2-4AC = D$), then the map (24) is given by

$$\begin{pmatrix} \tau_1 & \tau_2 \\ \tau_2 & \tau_3 \end{pmatrix} = \begin{pmatrix} \omega_1^2 z_1 + {\omega_1^\sigma}^2 z_2 & \omega_1\omega_2 z_1 + \omega_1^\sigma \omega_2^\sigma z_2 \\ \omega_1\omega_2 z_1 + \omega_1^\sigma\omega_2^\sigma z_2 & \omega_2^2 z_1 + {\omega_2^\sigma}^2 z_2 \end{pmatrix}, \tag{32}$$

which satisfies an equation of the form (31) with $(a,b,c,d,e) = (C,-B,A,0,0)$, $b^2-4ac-4de = D$. In general, to compute the intersection $\mathcal{H}_D \cap \mathcal{H}_{D'}$ we substitute the expression in (32) into the an equation of the form (31) with $b^2-4ac-4de = D'$. This gives the equation

$$-dDz_1z_2 \; + \; (a\omega_1^2 + b\omega_1\omega_2 + c\omega_2^2)z_1 \; + \; (a\omega_1^2 + b\omega_1\omega_2 + c\omega_2^2)^\sigma z_2 \; + \; e \; = \; 0\,,$$

which has the form (29) with

$$N = -deD - N(a\omega_1^2 + b\omega_1\omega_2 + c\omega_2^2) \; = \; \frac{DD' \; - \; (2Aa + Bb + 2Cc)^2}{4}$$

as asserted. In the special case $D' = 1$, we recover the equation $D = r^2 + 4N$ and also see that we are on the component $F_N\big(\frac{r+\sqrt{D}}{2\sqrt{D}}\big)$ of F_N, as claimed in (30), since it is easily seen that $a\omega_1^2 + b\omega_1\omega_2 + c\omega_2^2 \equiv \frac{r+\sqrt{D}}{2} \pmod{\mathfrak{d}}$.

5.3. **Teichmüller curves and Veech groups.** A *Teichmüller curve* is an irreducible algebraic curve W in the moduli space $\mathcal{M}_g$ of curves of genus g which is a totally geodesic submanifold for the Teichmüller metric. Teichmüller curves are generated by a pair consisting of a curve C and a non-zero holomorphic one-form $\omega \in H^1(C, \Omega^1_C)$. Such pairs are called *flat surfaces*. On the set of flat surfaces there is an action of $\mathrm{GL}(2,\mathbb{R})$ and Teichmüller curves are the projection to $\mathcal{M}_g$ of the orbit $\mathrm{GL}(2,\mathbb{R}) \cdot (C,\omega)$. The uniformizing group Γ such that $W = \mathbb{H}/\Gamma$, called a *Veech group*, can be read off from the flat geometry of the pair (C,ω). Let K_Γ be the trace field of Γ and $r = [K_\Gamma : \mathbb{Q}]$. Teichmüller curves with $r = g$ are called *algebraically primitive*. Under the Torelli map, algebraically primitive Teichmüller curves map to the locus of abelian varieties with real multiplication by K ([27] Theorem 2.6). In particular for $g = 2$ the universal covering of an algebraically primitive Teichmüller curve defines a map

$$(\varphi_0, \varphi) : \mathbb{H} \to \mathbb{H}^2$$

equivariant with respect to the action of the Veech group (acting on the left in the obvious way and on the right via its embedding into $\mathrm{SL}(2,K) \hookrightarrow \mathrm{SL}(2,\mathbb{R})^2$). The geodesic definition of Teichmüller curves implies that φ_0 is a Möbius transformation. Moreover we may suppose $\varphi_0 = \mathrm{id}$ using appropriate choices in the universal covering map. Consequently, Teichmüller curves define Hilbert modular embeddings in the above sense. ([27], Section 3.)

The space of flat surfaces (C,ω) is naturally stratified by the number and multiplicities of the zeros of ω. In particular, for $g = 2$ we have two strata $\Omega\mathcal{M}_2(1,1)$ and $\Omega\mathcal{M}_2(2)$, corresponding to ω having two distinct zeros or one double zero, respectively. For $g = 2$ we have the following classification for algebraically primitive Teichmüller curves ([20], [21], [28], [22]).

Theorem 5.2. *There is only one Teichmüller curve in the stratum $\Omega\mathcal{M}_2(1,1)$. It lies in the Hilbert modular surface X_5.*

The stratum $\Omega\mathcal{M}_2(2)$ contains infinitely many algebraically primitive Teichmüller curves, each lying in a unique Hilbert modular surface. For each non-square discriminant $D \geq 5$ the Hilbert modular surface X_D contains exactly one Teichmüller curve if $D \not\equiv 1 \pmod 8$ and exactly two if $D \equiv 1 \pmod 8$.

The union W_D of the Teichmüller curves in X_D of the second type is the locus in $\mathcal{M}_2$ of curves whose Jacobians have real multiplication by $\mathcal{O}_D$ and such that the eigendifferential on which $\mathcal{O}_D$ acts via the embedding $K \subset \mathbb{R}$ has a double zero.

The two components in the case $D \equiv 1 \pmod 8$ are distinguished by a spin invariant $\delta \in \{0, 1\}$ and will be denoted by $W_D^\delta = \mathbb{H}/\Gamma_D^\delta$, so that $W_D = W_D^0 \cup W_D^1$ in this case. The definition of the spin invariant is given in [21] and will not be repeated here, but in §9.2 we will be able to give a new and equivalent definition in terms of our description of Teichmüller curves via theta functions.

The Teichmüller curves in $\Omega\mathcal{M}_2(2)$ admit the following characterization, which is an adaptation of the criterion in [27], Theorem 5.3. Let $\mathcal{F}_i$ ($i = 1$ or 2) be the two natural foliations of a Hilbert modular surface X_D for which the i-th cooordinate is constant in the uniformization.

Theorem 5.3. *An algebraic curve $W \subset X_D$ is a union of Teichmüller curves if and only if*
(i) *W is disjoint from the reducible locus and*
(ii) *W is everywhere transversal to $\mathcal{F}_1$.*

Sketch of proof. If W is a Teichmüller curve, then (i) and (ii) hold by definition and by the fact that we can use the first coordinate as a parameter, respectively.

For the converse recall that over a Hilbert modular surface the relative first cohomology with coefficients in K splits into two eigenspaces, two local systems over K that we denote by $\mathbb{L}$ and $\widetilde{\mathbb{L}}$ and that are interchanged by the Galois group of K. Consequently, over any curve in a Hilbert modular surface the cohomology splits in the same way.

Condition (i) is equivalent to W being in the image of the locus of Jacobians with real multiplication under the Torelli map. To apply the criterion of [27], Theorem 5.3, we need to show that the Kodaira-Spencer map for $\mathbb{L}$ or $\widetilde{\mathbb{L}}$ vanishes nowhere on $\overline{W}$. Condition (ii) implies the non-vanishing of the Kodaira-Spencer map for the corresponding $\mathbb{L}$ in the interior of X_D, while at the cusps non-vanishing is automatic, by a local calculation as in [5], Proposition 2.2. □

One can generalize this setup using algebraic curves in A_g that are *totally geodesic for the Kobayashi metric.* See [29] for a characterization of these Kobayashi geodesics.

5.4. **Twisted modular forms for W_D.** The topology of W_D and the ratio λ_2 are completely determined, combining the work of several authors. We summarize the results and combine them with Theorem 3.2 to determine the dimension of the space of twisted modular forms.

Theorem 5.4. *For any non-square discriminant D, the fundamental invariants of the curves W_D are as follows.*

(i) The orbifold Euler characteristic of W_D equals

$$\chi(W_D) = -\tfrac{9}{2}\chi(X_D), \tag{33}$$

where X_D is the Hilbert modular surface $\mathbb{H}^2/\mathrm{SL}(\mathcal{O}_D^\vee \oplus \mathcal{O}_D)$.

(ii) The cusps of W_D are in bijection with standard quadratic forms of discriminant D (see Section 10).

(iii) For $D = 5$, the curve W_D has two fixed points, one of order two and one of order five. For $D \neq 5$, there are $e_2(D)$ elliptic fixed points of order two on W_D and no other fixed points, where $e_2(D)$ is a sum of class numbers of imaginary quadratic orders ([31], Table 1). In particular for $D \equiv 1 \mod 8$, there are $e_2(D) = \frac{1}{2}h(-4D)$ elliptic fixed points of order two.

(iv) The curves W_D^0 and W_D^1 are defined over $\mathbb{Q}(\sqrt{D})$ and are Galois conjugate.

(v) The curves W_D^0 and W_D^1 are homeomorphic.

(vi) For a torsion-free subgroup of the Veech group of any component W_D^i of W_D the ratio λ_2 of the degrees of the line bundles $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ equals $1/3$.

Proof. Statement (i) is the main result of [1]. Statement (ii) is implicit in [21] and explicit in [1]. Statement (iii) is the main result of [31]. Statement (iv) is Theorem 3.3(b) of [4] and (v) follows directly. Statement (vi) was shown in [1], Corollary 12.4, and with a different proof in [4], Corollary 2.4. □

We recall that the value of $\chi(X_D)$ is known, and is given for a fundamental discriminant D by

$$\chi(X_D) \;=\; 2\,\zeta_K(-1) \;=\; \frac{1}{30}\sum_{D=b^2+4ac} a \tag{34}$$

(see [15]), and in general by a similar explicit formula.

Every curve of genus two is hyperelliptic and consequently, $-I$ is in the Veech group for every Teichmüller curve in genus two. The dimension of the space of twisted modular forms can now be deduced from Theorem 3.2.

Theorem 5.5. *For $D > 5$ the space of twisted modular forms $M_{k,\ell}$ on W_D is zero for $k+\ell$ odd. For $k+\ell$ even and $D \not\equiv 1 \mod 8$*

$$\dim M_{k,\ell}(\Gamma) \;=\; -\frac{1}{2}\Big(k+\frac{\ell}{3}\Big)\,\chi(W_D) - \Big\{\frac{-k+\ell}{4}\Big\}\,e_2(D),$$

where $\{x\}$ is the fractional part of x, and for each of the two components for $D \equiv 1 \mod 8$

$$\dim M_{k,\ell}(\Gamma) \;=\; -\frac{1}{4}\Big(k+\frac{\ell}{3}\Big)\,\chi(W_D) - \frac{1}{4}\Big\{\frac{-k+\ell}{4}\Big\}\,e_2(D),$$

where $\chi(W_D)$ and $e_2(D)$ is given in Theorem 5.4.

Proof. The first statement holds because the Veech group contains $-I$. Given the general dimension calculation in Theorem 3.2 and the Euler characteristic in (33) it remains to show that for all the fixed points x of order two the local contribution $b_x(k,\ell)$ is $\{\frac{-k+\ell}{4}\}$, not $\{\frac{-k-\ell}{4}\}$ for some of them.

Suppose that $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(\mathcal{O}_D^\vee \oplus \mathcal{O}_D)$ is of order 4 and stabilizes $\mathbf{z} = (z_1, z_2) \in \mathbb{H}^2$. Then multiplication by the diagonal matrix with diagonal entries $((cz_1+d)^{-1}, (c^\sigma z_2 + d^\sigma)^{-1})$ defines a linear map J of $\mathbb{C}^2$ that stabilizes the lattice $\mathcal{L}_{\mathbf{z}}$ from (22) (i.e. the corresponding abelian surface has complex multiplication by the ring generated by $\mathcal{O}_D$ and J). To show that $b_{\mathbf{z}}(k,\ell) = \{\frac{-k+\ell}{4}\}$ is hence equivalent to showing that J^{-1} (or J) has two eigenspaces of dimension one, rather than a two-dimensional eigenspace.

Mukamel ([31]) studies, along with his classification of fixed points of Teichmüller curves, the locus $\mathcal{M}_2(D_8)$ of genus two surfaces with automorphism group containing the dihedral group of order 8. He shows that all the fixed points of order two on the Teichmüller curves W_D lie on the intersection (in A_2) of the Hilbert modular surface X_D with $\mathcal{M}_2(D_8)$. The family of curves over $M_2(D_8)$ is given by the hyperelliptic equation

$$Y^2 = (X^2-1)(X^2+aX+1), \quad a \in \mathbb{C} \setminus \{\pm 2\}.$$

The automorphism of order four is $J(X,Y) = (\frac{1}{X}, \frac{iY}{X^3})$ and the eigendifferentials are $dX/Y + X\,dX/Y$ and $dX/Y - X\,dX/Y$, which lie in the eigenspace for $+1$ and for -1 respectively. This proves the claim on the J-eigenspaces. □

5.5. **Gauss-Manin connection and Picard-Fuchs equation.** Here we explain why Teichmüller curves give rise to twisted modular forms and how to obtain the differential equations we attached to them in Section 4 geometrically. For the moment, let W be any curve in $\mathcal{M}_2$ such that the corresponding family of Jacobians has real multiplication by an order in K. Then the vector bundle with fiber $H^1(C,\mathbb{C})$ over the point $[C] \in M_2$ splits (over $\mathbb{R}$, and in fact over K), as in the proof of Theorem 5.3, into rank two subbundles $\mathbb{L}$ and $\widetilde{\mathbb{L}}$. This vector bundle also comes with a flat (Gauss-Manin) connection ∇. The bundles $\mathbb{L}$ and $\widetilde{\mathbb{L}}$ come with holomorphic subbundles $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ respectively, whose fibers over X are the holomorphic one-forms on C that are eigenforms for the real multiplication. The bundles $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ naturally extend over the cusps $\overline{W} \setminus W$, where the fibers are stable forms. We denote them by the same letters. (We recall that a form is called *stable* if in the limit as $t \to t_0$, where the genus 2 curve parametrized by $t \neq t_0$ degenerates to a curve of genus 0 with double points, the corresponding differential on the normalization of this curve has simple poles with opposite residues at the points that get identified.)

Suppose for simplicity that W is a rational curve with parameter t. If we choose sections $\omega(t)$ of $\mathcal{L}$ and $\widetilde{\omega}(t)$ of $\widetilde{\mathcal{L}}$, then $\{\omega(t), \nabla(\partial/\partial t)\omega(t), \nabla(\partial/\partial t)^2\omega(t)\}$ are linearly dependent in cohomology. Concretely, this means that if L is the corresponding second order differential linear operator, a quadratic polynomial in $\partial/\partial t$, then the image of $\omega(t)$ under L is exact. Similarly, $\{\widetilde{\omega}(t), \nabla(\partial/\partial t)\widetilde{\omega}(t), \nabla(\partial/\partial t)^2\widetilde{\omega}(t)\}$ are linearly dependent and give a second order differential operator that makes $\widetilde{\omega}(t)$ exact. It follows that the periods, defined as the integral of $\omega(t)$ and $\widetilde{\omega}(t)$ over any fixed element of $H_1(C,\mathbb{C})$ are annihilated by L and $\widetilde{L}$ respectively. (Here "fixed" means that we use the property of being a local system to identify the homology groups $H_1(C_t,\mathbb{C}) = H_1(C_t,\mathbb{Z}) \otimes \mathbb{C}$ with each other locally.) These are the well-known *Picard-Fuchs differential equations* satisfied by periods.

Now assume that W is a Teichmüller curve. We show that the periods just described are (twisted) modular forms of weight $(1,0)$ and $(0,1)$ respectively, with respect to a modular embedding φ as defined in Section 1. More precisely we have the following correspondence.

Proposition 5.6. *Suppose that W is a Teichmüller curve with uniformization $\mathbb{H}/\Gamma$ as above, and let L and $\widetilde{L}$ be the rank two differential operators associated with sections $\omega(t)$ and $\widetilde{\omega}(t)$ of $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ as above. Then there is a rank-one submodule (in the rank-two $\mathcal{O}_{\overline{W}}(W)$-module of solutions of L) consisting of holomorphic modular forms of weight $(1,0)$, and a rank-one submodule (in the rank-two $\mathcal{O}_{\overline{W}}(W)$-module* *of solutions of $\widetilde{L}$) consisting of twisted holomorphic modular forms of weight* $(0,1)$. *If $\omega(t)$ (resp. $\widetilde{\omega}(t)$) extends to a stable form over a cusp of W, then the corresponding (twisted) modular form is holomorphic at this cusp.*

This defines a $1{:}1$ *correspondence between holomorphic sections of $\mathcal{L}$ over $\overline{W}$ and holomorphic twisted modular forms on Γ of weight* $(1,0)$, *and a* $1{:}1$ *correspondence between holomorphic sections of $\widetilde{\mathcal{L}}$ over $\overline{W}$ and holomorphic twisted modular forms on Γ of weight* $(0,1)$.

Proof. In [27] it was shown that there exists an oriented basis β, α of the kernel of $\widetilde{\omega}$ in $H_1(C,\mathbb{R})$ such that the monodromy representation of $\pi_1(W)=\Gamma$ on that subspace is the identity, and similarly a basis β^σ, α^σ of the kernel of ω with respect to which the monodromy representation is given by the Galois conjugate group Γ^σ.

Consequently, the period map $z\mapsto\int_\beta\omega(z)/\int_\alpha\omega(z)$ is equivariant with Γ acting on domain and range, hence the identity after an appropriate conjugation by a Möbius transformation. Moreover, the period map $z\mapsto\int_{\beta^\sigma}\widetilde{\omega}(z)/\int_{\alpha^\sigma}\widetilde{\omega}(z)$ is equivariant with Γ acting on the domain and Γ^σ on the range. Hence this map agrees with φ in the definition of the modular embedding by the uniqueness of modular embeddings.

As said above, the periods $f(z)=\int_\alpha\omega(z)$ and $f_1(z)=\int_\beta\omega(z)=zf(z)$ span the space of solutions of L (pulled back to $\mathbb{H}$ via t). The statement above about the monodromy implies that for all $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\Gamma$

$$\begin{pmatrix}f_1(\gamma z)\\ f(\gamma z)\end{pmatrix}=\gamma\begin{pmatrix}f_1(z)\\ f(z)\end{pmatrix}=\begin{pmatrix}af_1(z)+bf(z)\\ cf_1(z)+df(z)\end{pmatrix}=\begin{pmatrix}(az+b)f(z)\\ (cz+d)f(z)\end{pmatrix}.$$

The second row implies that f is a twisted modular form for Γ of weight $(1,0)$. Similarly, the periods $\widetilde{f}(z)=\int_{\alpha^\sigma}\widetilde{\omega}(z)$ and $f_1(z)=\int_{\beta^\sigma}\widetilde{\omega}(z)=\varphi(z)\widetilde{f}(z)$ span the space of solutions of $\widetilde{L}$ and we have

$$\begin{pmatrix}\widetilde{f}_1(\gamma z)\\ \widetilde{f}(\gamma z)\end{pmatrix}=\gamma^\sigma\begin{pmatrix}\widetilde{f}_1(z)\\ \widetilde{f}(z)\end{pmatrix}=\begin{pmatrix}a^\sigma\widetilde{f}_1(z)+b^\sigma\widetilde{f}(z)\\ c^\sigma\widetilde{f}_1(z)+d^\sigma\widetilde{f}(z)\end{pmatrix}=\begin{pmatrix}(a^\sigma\varphi(z)+b^\sigma)\widetilde{f}(z)\\ (c^\sigma\varphi(z)+d^\sigma)\widetilde{f}(z)\end{pmatrix}.$$

Again, the second row implies that $\widetilde{f}$ is a twisted modular form for Γ of weight $(0,1)$.

Holomorphicity of f and $\widetilde{f}$ in the interior of $\mathbb{H}$ is obvious by the definition of a period. To show that they are holomorphic at the cusps, we may assume without loss of generality that $z_0=\infty$ and $t_0=0$. There, it follows from the definition of the monodromy representation that α (resp. α^σ) is characterized in $\langle\widetilde{\omega}\rangle^\perp$ (resp. in $\langle\omega\rangle^\perp$) as the elements invariant under the local monodromy group. The period of a stable form along such a cycle is finite.

To establish the last statement of the proposition, we just need to assign to every holomorphic twisted modular form of weight $(1,0)$ (resp. weight $(0,1)$) a section of $\mathcal{L}$ (resp. of $\widetilde{\mathcal{L}}$). This is well-known in the untwisted case and was done in both the untwisted and twisted cases in Section 4 of Part I. ☐

We end with a remark on Galois conjugation and spin. We defined Γ to be monodromy group of the local system $\mathbb{L}$. Then, of course, the monodromy group of the Galois conjugate $\widetilde{\mathbb{L}}$ is Γ^σ. We will see in the next section in an example, and at the end of the paper in general, that the solutions y and $\widetilde{y}$ also have coefficients in the field K. However, the Galois conjugate solution y^σ is neither equal to $\widetilde{y}$ nor to any other solution of $\widetilde{\mathbb{L}}$. In fact, y^σ is naturally a solution of a differential operator associated with the Galois conjugate Teichmüller curve W^σ. For $D \equiv 1 \mod 8$ this is the curve with the other spin invariant (see Theorem 5.4 iv)). For $D \not\equiv 1 \mod 8$ this Galois conjugate curve is isomorphic to the original curve by McMullen's classification recalled in Theorem 5.2. The equation of this curve for $D = 13$ is given explicitly in [4].

6. Example: the curve W_{17}^1 and its associated differential equations

Our running example, from now until the end of Part II, will be the Teichmüller curve W_{17}^1 on the Hilbert modular surface X_{17}. In this section we gather the known results for this curve, applying the algorithm of McMullen to compute the Veech group and summarizing the construction from [4] to compute the equation of the universal family and the corresponding Picard-Fuchs differential equations.

6.1. **The Veech group for $D = 17$ and spin 1.** For small values of D the Veech groups $\Gamma_D^\delta = \Gamma(W_D^\delta)$ can be calculated using the algorithm in [20]. We describe this in detail for the case $D = 17$, $\delta = 1$, i.e., for the Teichmüller curve of non-trivial spin. McMullen's algorithm gives a subgroup of $\mathrm{SL}(2, \mathcal{O}_D)$, so the group that we will get (which we will denote simply by Γ, or by Γ_{17} when needed for clarity, with quotient $W = \mathbb{H}/\Gamma$) is actually the conjugate $\Delta\Gamma_{17}^1\Delta^{-1}$ of Γ_{17}^1 by $\Delta = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \sqrt{D} \end{smallmatrix}\right)$, and for the same reason the function φ used to make the modular embedding will go from $\mathbb{H}$ to $\mathbb{H}^-$ rather than from $\mathbb{H}$ to $\mathbb{H}$. Later, when we use this modular embedding explicitly to compare the twisted modular forms on W with standard Hilbert modular forms for $\mathbb{Q}(\sqrt{17})$, we will conjugate back to make the comparison easier.

We denote by $\alpha = (1+\sqrt{17})/2$ the standard generator of $\mathcal{O} = \mathcal{O}_{17}$ over $\mathbb{Z}$, and for ease of reading will sometimes use the abbreviated notation $[m, n]$ for $m + n\alpha \in \mathcal{O}$. The group Γ can be embedded into $\mathrm{SL}(2, \mathbb{R})$ by the standard embedding of $\mathcal{O}$ into $\mathbb{R}$ and then acts discretely. (Note that the other embedding of $\mathcal{O}$ into $\mathbb{R}$ would lead to a non-discrete subgroup of $\mathrm{SL}(2, \mathbb{R})$!) A fundamental domain for this action is shown in Figure 1 (b), while Figure 1 (a) shows the explicit "L-shaped region" needed to apply the algorithm and obtain the fundamental domain. The group Γ has three cusps, at $z = \infty$, 1, and $\alpha/2$, and an elliptic fixed point of order 2 at $z = i$, where z is the coordinate in $\mathbb{H}$. The stabilizers of the cusps are the infinite cyclic groups generated by the parabolic elements

$$M_\infty = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}, \ M_1 = \begin{pmatrix} -2\alpha-2 & 2\alpha+3 \\ -2\alpha-3 & 2\alpha+4 \end{pmatrix}, \ M_{\alpha/2} = \begin{pmatrix} -2\alpha-3 & 3\alpha+4 \\ -2\alpha-2 & 2\alpha+5 \end{pmatrix},$$

and the stabilizer of i is generated by the element $M_i = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. The presentation of the group Γ given by McMullen's algorithm is then

$$\Gamma = \left\langle M_\infty, M_i, M_1, M_{\alpha/2} \,\middle|\, M_i^2 = -1, \ M_\infty M_i M_1 M_{\alpha/2} = 1 \right\rangle.$$

It will be useful in the sequel to deal not only with the Fuchsian group Γ but also with a certain index 4 subgroup Π of it, already mentioned in §3. This group is more convenient for purposes of calculation because it is free and also because the universal genus 2 curve over $\overline{\mathbb{H}/\Pi}$ has a stable model. We want that $\widetilde{\Pi} = \Pi \times \{\pm 1\}$

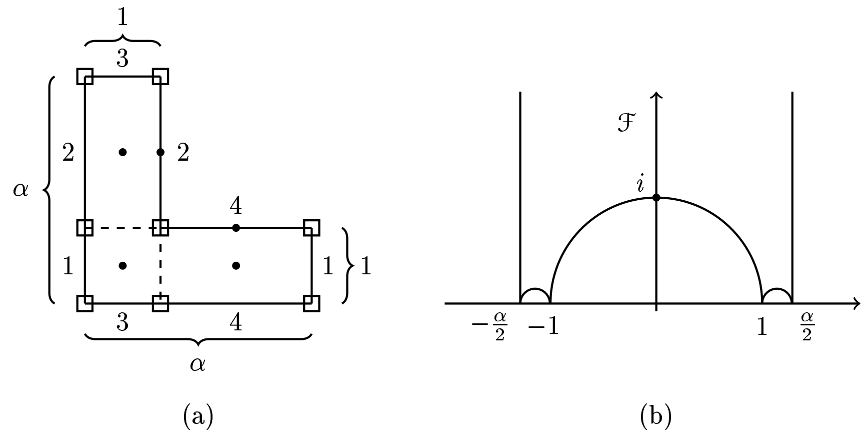

FIGURE 1. (a) Flat surface generating the Teichmüller curve W_{17} when $\alpha = (1+\sqrt{17})/2$. The square is the double zero of ω, the black points are the remaining 5 Weierstrass points.
(b) A fundamental domain for Γ.

has index two in Γ. (This already implies that Π has no torsion and is thus free.) This group is not unique. We fix the choice

$$\Pi \;=\; \langle M_\infty,\; M_{\alpha/2},\; M_1^2,\; -M_1^{-1}M_\infty M_1\rangle\,.$$

The group $\widetilde{\Pi} = \Pi \times \{\pm 1\}$ is the kernel of the homomorphism from $\Gamma \to \{\pm1\}$ sending M_∞ and $M_{\alpha/2}$ to 1 and M_i and M_1 to -1.

Both curves $\mathbb{H}/\Gamma$ and $\mathbb{H}/\Pi$ have genus 0, so there are modular functions $u(z)$ and $t(z)$ on Γ and Π giving isomorphism between their compactifications and $\mathbb{P}^1(\mathbb{C})$ (hauptmodules). We can normalize them so that the involution induced by M_i : $z \mapsto -1/z$ corresponds to $t \mapsto 1/t$ and that the three cusps of $\mathbb{H}/\Pi$ are at 0, 1 and ∞. Then the quotient map $t \mapsto s$ is given by

$$s \;=\; -\frac{4\,\kappa_0\,t}{(t-1)^2}\,, \qquad \text{where} \quad \kappa_0 = \frac{-895+217\sqrt{17}}{256} \tag{35}$$

and the values of s and t at the cusps and elliptic fixed points are given, according to the calculations in [4] (where a different parameter on $\mathbb{H}/\Pi$ was used), by the table

	z	∞	0	1	$\alpha/2$	$-2/\alpha$	i
Π	$t = t(z)$	0	∞	1	λ^{-1}	λ	-1
Γ	$s = s(z)$	0	0	∞	1	1	κ_0

(36)

where

$$\lambda \;=\; t\Bigl(-\frac{2}{\alpha}\Bigr) \;=\; \frac{31-7\sqrt{17}}{2}\,, \qquad \kappa_0 \;=\; s(i) \;=\; -\frac{(\lambda-1)^2}{4\lambda}\,. \tag{37}$$

For later use we emphasize that both t and s are local parameters of the Teichmüller curve at the cusp $z=\infty$. The whole situation is summarized by the following diagram.

$$\begin{tikzcd}
\overline{W}_\Pi = \overline{\mathbb{H}/\Pi} \arrow[r, "t"] \arrow[d, "2:1"'] & \mathbb{P}^1(\mathbb{C}) \arrow[d, "s = -\frac{4\kappa_0 t}{(1-t)^2}"] \\
\overline{W} = \overline{\mathbb{H}/\Gamma} \arrow[r, "s"] & \mathbb{P}^1(\mathbb{C})
\end{tikzcd}$$

6.2. **The universal family over W.** The modular curve $W = \mathbb{H}/\Gamma$ parametrizes a family of genus 2 curves with real multiplication by $\mathcal{O} = \mathbb{Z}+\mathbb{Z}\alpha$ on their Jacobians. This family, and its associated Picard-Fuchs differential equations, was determined explicitly in [4]. In this subsection we review these results, and also give somewhat simpler equations by making suitable changes of coordinates.

The explicit equation in [4] was actually given for the family over the double cover $\mathbb{H}/\Pi$ of W, with the parameter t, and has the form

$$\begin{aligned} Y^2 \;=\; P_5(X,t) \;=\; \bigl(X \,+\, (At+B)\bigr) \,\times\, \bigl(X \,+\, (Bt+A)\bigr) \times \\ \times\; \bigl(X^3 \,+\, C(t+1)X^2 \,+\, (D(t+1)^2+Et)X \,+\, F(t+1)^3+Gt(t+1)\bigr)\,, \end{aligned} \tag{38}$$

with coefficients $A,\dots,G$ given (with our above notation $[m,n] = m+n\alpha$) by

$$\begin{aligned} &A = 5\,[2,1]\,, \quad B = -2\,[5,3]\,, \quad C = [3,1]\,, \quad D = -\frac{1}{4}\,[827,529]\,, \\ &E = 2^4\cdot 17\,[3,2]\,, \quad F = -\frac{1}{2}\,[4597,2943]\,, \quad G = 2\cdot 17\cdot [271,173]\,. \end{aligned} \tag{39}$$

(These coefficients are not quite as bad as they look since they all factor into small prime factors, e.g. $D = -\pi_2^{11}/4\varepsilon^3$ and $F = -\pi_2^{14}\varepsilon^4/2$ where $\varepsilon = 4+\sqrt{17}$ is the fundamental unit of $\mathbb{Q}(\sqrt{17})$ and $\pi_2 = (3+\sqrt{17})/2$ one of the prime factors of 2.) We explain briefly how this equation is derived.

We can represent the fiber C_t over t as a hyperelliptic curve $Y^2 = P_6(X,t)$, where P_6 is a polynomial of degree 6 in X whose roots correspond to the Weierstrass points of C_t. From the action of the group $\Pi \subset \Gamma$ on the Veech groups on these points (which can be analyzed by looking at the Figure 1A, in which the Weierstrass points are indicated by black points), we see that they break up into three orbits of size 1 and one of size 3, with two of the 1-element orbits being interchanged by the symmetry $t \mapsto t^{-1}$. Placing the other 1-element orbit, the singularity of ω, at infinity, we get a new equation of the form $Y^2 = P_5(X,t)$ where P_5 factors into two linear and one cubic polynomial, and by degree computations together with the symmetry under $t \mapsto t^{-1}$ we find that these factors must have the form given in (38) for some constants $A,\dots,G$. They are not unique, since we can make a change of variables $X \mapsto \alpha X + \beta(t+1)$, but become unique up to scaling if we assume that the two eigendifferentials $\omega = dv_1$ and $\widetilde{\omega} = dv_2$ are dX/Y and $X\,dX/Y$, respectively. To determine them, we note that at each cusp $t = c \in \{0,\,1,\,\lambda\}$ this polynomial must acquire two double roots (the degeneracy of the genus 2 curve at infinity cannot consist of just two roots coming together, because the real multiplication forces the subspace of $H_1(C;\mathbb{Z})$ that collapses to be an $\mathcal{O}$-module and hence to have even rank over $\mathbb{Z}$), so we have $P_5(X,c) = (X-X_0^c)(X-X_1^c)^2(X-X_2^c)^2$. (The corresponding factorizations at the two other cusps $t=\infty$ and λ^{-1} are then automatic because of the $t \mapsto t^{-1}$ symmetry.) These conditions do not yet suffice to determine the equation, but there is one further condition at each cusp. This comes from the fact that a singular curve of the form $Y^2 = (X-X_0)(X-X_1)^2(X-X_2)^2$ has genus 0.

A parametrization with parameter T is given by $X = (X_1 - X_0)T^2 + X_0$ and $Y = (X_1 - X_0)^{5/2}T(T^2-1)(T^2-\rho^2)$, where $\rho^2 = (X_2 - X_0)/(X_1 - X_0)$ denotes the cross-ratio of X_0, X_1, X_2 and ∞. The differential form dX/Y corresponds under this map to a multiple of the differential form $\left(\frac{1}{T-1} - \frac{1}{T+1} + \frac{\rho^{-1}}{T-\rho} - \frac{\rho^{-1}}{T+\rho}\right) dT$ on $\mathbb{P}^1$ having four simple poles with residues summing to 0 in pairs, and with the ratio of the non-paired residues being $\pm\rho$. But for the cusps of the Teichmüller curves we know *a priori* that these ratios of residues must equal the ratio of the top and bottom sides of the L-shaped region as it degenerates. From the horizontal sides Figure 1A we read off the value $\rho_0 = \alpha$. Redrawing this figure decomposed into cylinders in the direction of slope one (corresponding to the cusp $t = 1$) and slope $\alpha/2$ (corresponding to the cusp $t = \lambda$ by the table in the preceding subsection) we find $\rho_1 = \alpha/2$, $\rho_\lambda = (1+\alpha)/2$. This information now suffices to determine all of the unknown coefficients, up to the ambiguity already mentioned (in particular the second eigendifferential form $\widetilde{\omega} = X\,dX/Y$ automatically has the correct ratio of residues, namely, the Galois conjugates of the ones for ω, so that there are no extra restrictions on the coefficients coming from this condition), and carrying out the calculation we find the values given in (39).

We remark that equation (38) can be simplified considerably by substituting $(1+t)(1+X\sqrt{17})/4$ for X, in which case $P_5(X,t)$, up to a factor $(\sqrt{17}(1+t)/4)^5$, takes on the much simpler form

$$F_5(X,u) \;=\; \big((X-1)^2 \,-\, [4,5]\,u\big)\,\big((X+\sqrt{17})(X+1)^2 \,-\, 8u(2X+[9,5])\big)$$

with $u = 4\varepsilon\left(\frac{1-t}{1+t}\right)^2 = \frac{4\varepsilon}{1-s/\kappa_0}$. This gives an explicit and relatively simple equation for the family of genus 2 curves over the Teichmüller curve $\mathbb{H}/\Gamma$.

6.3. **The Picard-Fuchs equations for W and their solutions.** As already discussed in Section 3, even though we are considering only the single curve $W = \mathbb{H}/\Gamma$, there are *two* Picard-Fuchs differential equations, corresponding to the variation of the periods of the two eigendifferentials ω and $\widetilde{\omega}$ for the action of $\mathcal{O}$ on the space of holomorphic differentials of the fibers. It will be crucial for our calculations to have both of them, since together they will tell us explicitly how the Teichmüller curve W is embedded in the Hilbert modular surface X_{17}.

Obtaining the Picard-Fuchs differential equations satisfied by the periods of the two eigendifferentials ω and $\widetilde{\omega}$ is straightforward once the equation of the family of curves has been obtained. One has to find differential operators L and $\widetilde{L}$ mapping the one-forms ω and $\widetilde{\omega}$ to exact forms. The result, given in [4], is a pair of differential operators of the same form as in (20), namely

$$L \;=\; \frac{d}{dt}\,A(t)\,\frac{d}{dt} \,+\, B(t)\,, \quad \widetilde{L} \;=\; \frac{d}{dt}\,\widetilde{A}(t)\,\frac{d}{dt} \,+\, \widetilde{B}(t)\,, \tag{40}$$

where $A(t)$ and $B(t)$ are the polynomials given by

$$\begin{aligned} A(t) \;&=\; t\,(t-1)\,(t-\lambda)(t-\lambda^{-1}) \;=\; t^4 \,-\, \beta t^3 \,+\, \beta t^2 \,-t\,, \\ B(t) \;&=\; \frac{3}{4}\left(3t^2 \,-\, (\beta+\gamma)\,t \,+\, \gamma\right), \end{aligned} \tag{41}$$

with $\lambda = (31 - 7\sqrt{17})/2$ as in (37) and β and γ defined by

$$\beta = \lambda \,+\, \lambda^{-1} + 1 = \frac{1087 - 217\sqrt{17}}{64}\,, \quad \gamma = \frac{27 - 5\sqrt{17}}{4}\,, \tag{42}$$

and where $\widetilde{A}(t)$ and $\widetilde{B}(t)$ are the rational functions

$$\begin{aligned}\widetilde{A}(t) &= A(t)\Big/\Big(t^2 + \frac{137-95\sqrt{17}}{128}\,t + 1\Big)\,,\\ \widetilde{B}(t) &= \Big(\frac{1}{4}\,t^4 + \frac{1113-399\sqrt{17}}{512}\,t^3 - \frac{260375-69633\sqrt{17}}{16384}\,t^2 \\ &\quad - \frac{1387-301\sqrt{17}}{128}\,t + \frac{23-5\sqrt{17}}{8}\Big)\Big/\Big(t^2 + \frac{137-95\sqrt{17}}{128}\,t + 1\Big)^2\,.\end{aligned} \tag{43}$$

The differential operator L has five singularities, at infinity and at the roots of $A(t)$. The differential operator $\widetilde{L}$ has seven singularities, these five and two more at the poles of $\widetilde{A}$, but these last two are only apparent singularities of the differential equation, i.e., all solutions of the equation are holomorphic at these points.

The unique solutions in $1 + t\,\mathbb{C}[[t]]$ of the differential equations $Ly = 0$ and $\widetilde{L}\widetilde{y} = 0$ can easily be calculated recursively. The first few terms are given by

$$\begin{aligned} y &= 1 + \tfrac{81-15\sqrt{17}}{16}\,t + \tfrac{4845-1155\sqrt{17}}{64}\,t^2 + \tfrac{3200225-775495\sqrt{17}}{2048}\,t^3 + \cdots \\ &\approx 1 + 1.197\,t + 1.294\,t^2 + 1.356\,t^3 + 1.402\,t^4 + 1.439\,t^5 + \cdots \\ \widetilde{y} &= 1 + \tfrac{23-5\sqrt{17}}{8}\,t + \tfrac{5561-1343\sqrt{17}}{128}\,t^2 + \tfrac{452759-109793\sqrt{17}}{512}\,t^3 + \cdots \\ &\approx 1 + 0.2981\,t + 0.1849\,t^2 + 0.1384\,t^3 + 0.1131\,t^4 + 0.0973\,t^5 + \cdots \end{aligned}$$

There are also unique power series y_1 and $\widetilde{y}_1$ without constant term such that $y\,\log(t) + y_1$ and $\widetilde{y}\,\log(t) + \widetilde{y}_1$ are solutions of the same differential equations as y and $\widetilde{y}$, respectively. These series begin

$$\begin{aligned} y_1 &= \tfrac{439-97\sqrt{17}}{64}\,t + \tfrac{563089-135575\sqrt{17}}{4096}\,t^2 + \tfrac{200641639-48642353\sqrt{17}}{65536}\,t^3 + \cdots \\ &\approx 0.6103\,t + 1.001\,t^2 + 1.283\,t^3 + 1.504\,t^4 + 1.687\,t^5 + \cdots \\ \widetilde{y}_1 &= \tfrac{1575-369\sqrt{17}}{128}\,t + \tfrac{1749337-423695\sqrt{17}}{8192}\,t^2 + \tfrac{1764480419-427927381\sqrt{17}}{393216}\,t^3 + \cdots \\ &\approx 0.4185\,t + 0.2927\,t^2 + 0.2305\,t^3 + 0.1958\,t^4 + 0.1748\,t^5 + \cdots \end{aligned}$$

We have given the numerical values of the first coefficients of each of these four power series to emphasize that they are quite small (and the same is true of the first few hundred, which we have computed). In fact, the coefficients in each case grow like λ^n, where $\lambda = 1.06913\cdots$ is the number defined by (37), since the radius of convergence is the absolute value of the nearest singularity $t \neq 0$, and the singularities are at $t = 0$, λ^{-1}, 1, λ and ∞. The growth in each case is quite regular, with the coefficient of t^n being asymptotic to a constant times λ^n/n. It is perhaps worth mentioning that if we took the Galois conjugates L^σ and $\widetilde{L}^\sigma$ of the differential operators L and $\widetilde{L}$, which give the Picard-Fuchs equations for the other Teichmüller curve W_{17}^0 (cf. [4] or §9), then the power series $y, \ldots, \widetilde{y}_1$ would also be replaced by their Galois conjugates and would look algebraically very similar to those above, but would have completely different real coefficients and growth, e.g., the expansion of y^σ begins

$$\begin{aligned} y^\sigma &= 1 + \tfrac{81+15\sqrt{17}}{16}\,t + \tfrac{4845+1155\sqrt{17}}{64}\,t^2 + \tfrac{3200225+775495\sqrt{17}}{2048}\,t^3 + \cdots \\ &\approx 1 + 8.928\,t + 150.11\,t^2 + 3123.9\,t^3 + 71667\,t^4 + 1738907\,t^5 + \cdots\,, \end{aligned}$$

now with coefficients growing like $(\lambda^\sigma)^n/n$ with $\lambda^\sigma = 29.93086\cdots$. The corresponding Fuchsian group, although isomorphic to Π as an abstract group, is not conjugate to it in $\mathrm{SL}(2,\mathbb{R})$, and the quotients of the upper half-plane by these two groups, which are the curves W_D^0 and W_D^1, represent different points of the moduli space $\mathcal{M}_{0,5}$.

Another very striking property of the expansions of y and $\widetilde{y}$ given above (and then of course also of their conjugates y^σ and $\widetilde{y}^\sigma$) is that the only denominators one sees are powers of 2, i.e., the first few coefficients of these power series all belong to the ring $\mathcal{O}[\frac{1}{2}]$. A calculation to higher accuracy shows that the same holds for the first few hundred coefficients, and in fact it is a theorem, proved in [4], that it holds for all coefficients. We will return to this question at the end of the next section because it is has a very interesting aspect that was in fact the point of departure for our whole investigation.

7. Arithmetic properties of modular forms for W_{17}^1

With the preparations in the preceding sections we can now compute in §7.1 the modular embedding φ in the example $D = 17$. In the process we compute the Fourier expansions of some modular forms and later, in §7.3 we completely determine the ring of twisted modular forms in this specific example. This arithmetic of the coefficients reveals two surprising phenomena, a transcendental constant needed for the correct choice of the q-parameter and the integrality statement mentioned above and proved in [4], which cannot be explained using one modular q-variable. We discuss these in §7.2, and provide the explanations in Section 8.

7.1. **Modular parametrization of the differential equations.** We have already mentioned that the differential equations $Ly = 0$ and $\widetilde{L}\widetilde{y} = 0$ have the same form as the differential equation (20) satisfied by ordinary or twisted modular forms with respect to a hauptmodule. This is of course not a coincidence: we have

$$y(t(z)) \;=\; f(z)\,, \quad \widetilde{y}(t(z)) \;=\; \widetilde{f}(z) \qquad \text{for } \Im(z) \text{ large, } |t(z)| \text{ small}$$

where $t : \mathbb{H}/\Pi \to \mathbb{C}$ is the map defined in §6.1 and $f(z)$ and $\widetilde{f}(z)$ are a modular form of weight 1 and a twisted modular form of weight (0,1), respectively, on the same group Π. In this subsection we will work out this statement in more detail, obtaining in particular a way to calculate the expansion (5) of the function $\varphi : \mathbb{H} \to \mathbb{H}^-$ whose graph gives the embedding of $\mathbb{H}/\Pi$ into $\mathbb{H} \times \mathbb{H}^-/\mathrm{SL}(2,\mathcal{O}_{17})$ as discussed in Section 1. (Here we need $\mathbb{H}^-$, rather than $\mathbb{H}$ as in Section 1, because we have conjugated the original Veech group by $\Delta = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \sqrt{17} \end{smallmatrix}\right)$ to embed it into $\mathrm{SL}(2,\mathcal{O}_D)$ and the Galois conjugate of Δ has negative determinant.) We will also calculate the q-expansion of $\widetilde{f}(z)$, obtaining our first explicit example of a twisted modular form. In Section 8 we will use this information to determine completely the rings of twisted modular forms for Γ and Π and the algebraic description of the Teichmüller curve W inside the Hilbert modular surface X_{17}.

We begin with the functions y and f. As stated in §6.1, the cusp at infinity for either Γ or Π has width α, i.e., its stabilizer is generated by the transformation $z \mapsto z+\alpha$, where $\alpha = (1+\sqrt{17})/2$ is our standard generator of $\mathcal{O}_{17}$, so any modular function or modular form on either group can be written as a power series in the variable $q = e^{2\pi i z/\alpha}$. On the other hand, we know that the space of solutions of the differential equation satisfied by any weight 1 modular form $f(z)$ with respect to any modular function on the same group is spanned by $f(z)$ and $zf(z)$. Our first thought is thus that q coincides with the "mirror parameter"

$$Q \;=\; Q(t) \;=\; t\exp(y_1/y)\,, \tag{44}$$

where $y = y(t)$ and $y_1 = y_1(t)$ are the two power series in t defined in §6.3. This is indeed what happens in the case of the Apéry or Apéry-like differential equations (see [41]), at least if one normalizes the hauptmodule correctly. Here, however, it is not quite true. We can see this numerically as follows. The function $Q(t)$ has a Taylor expansion beginning

$$Q(t) \;=\; t \,+\, \tfrac{439-97\sqrt{17}}{64}\,t^2 \,+\, \tfrac{249125-60195\sqrt{17}}{2048}\,t^3 \,+\, \cdots\,. \tag{45}$$

We can invert this power series to obtain

$$t \;=\; t(Q) \;=\; Q \,-\, \tfrac{439-97\sqrt{17}}{64}\,Q^2 \,+\, \tfrac{103549-24971\sqrt{17}}{2048}\,Q^3 \,+\, \cdots \tag{46}$$

and then substitute this into the expansion of $f(z) = y(t)$ to express $f(z)$ as a power series

$$\begin{aligned} f(z) \;&=\; 1 \,+\, \tfrac{81-15\sqrt{17}}{16}\,Q \,+\, \tfrac{8613-2019\sqrt{17}}{512}\,Q^2 \,+\, \tfrac{726937-175823\sqrt{17}}{16384}\,Q^3 \,+\, \cdots \\ &\approx\; 1 + 1.197\,Q + 0.563\,Q^2 + 0.122\,Q^3 + 0.0082\,Q^4 - 0.0011\,Q^5 - \cdots \end{aligned} \tag{47}$$

in the new local parameter Q at infinity. Looking at the first few numerical coefficients in this expansion, we see that they seem to be tending to 0 rapidly, suggesting that the radius of convergence of this power series is larger than 1, which is the value it would have to have if we were expanding with respect to q. The point is that, although the function $\log Q = \log t \,+\, y_1/y$ has the same behavior at infinity as $2\pi i z/\alpha$, namely, that it is well-defined up to an integer multiple of $2\pi i$, this property determines it only up to an additive constant. Therefore q and Q are related by

$$Q \;=\; A\,q \;=\; A\,e^{2\pi i z/\alpha} \tag{48}$$

for some constant $A \neq 0$ that has no reason to be equal to 1. The radius of convergence of the series in (47) is then equal to the absolute value of this constant.

We can use this idea, or a modification of it, to calculate A numerically. First, by computing a few hundred coefficients of the series in (47) and calculating its radius of convergence by the standard formula $R = \liminf |a_n|^{-1/n}$, where a_n denotes the nth coefficient, we find that $|A|$ is roughly equal to 7.5. However, this direct approach has very poor convergence (because the coefficients of the expansion of $f(z)$ in Q, unlike those of the same function when written as a power series $y(t)$ in t, do not behave in a regular way), and anyway gives only the absolute value of the scaling constant A. To find the actual value to high precision, we apply a simple trick. From the data in the table (36), we know that the value of $t(z)$ at $z = i$ equals -1 and that this value is taken on with multiplicity 1 (because t is a hauptmodule for a group with no elliptic fixed points) and is not taken on at any point in the upper half-plane with imaginary part bigger than 1 (because i and its translates by multiples of α are the highest points in the Π-orbit of i). It follows that the function $1/(t(z)+1)$ has a simple pole at $z = i$ and that if we express this function as a power series in q (resp. Q), then its singularity nearest the origin is a simple pole at $q_0 = e^{-2\pi/\alpha}$ (resp. $Q_0 = Aq_0$). In other words, $1/(1+t)$ is the sum of $c/(1 - Q/Q_0)$ for some non-zero constant c and a function holomorphic in a disc of radius strictly larger than $|Q_0|$. This implies that if we expand $1/(1+t) = \sum b_n Q^n$, then the coefficients b_n are given by $b_n = cQ_0^{-n}(1 + \mathrm{O}(a^{-n}))$ for some $a > 1$, and hence that the quotients b_{n+1}/b_n tend to Q_0 with exponential rapidity. Calculating a few hundred of the coefficients b_n numerically, we find from this the value

$$A \;\approx\; -7.48370822991173536914114556623211$$

to very high precision. After some trial and error we can recognize this number "in closed form" as

$$A \;\stackrel{?}{=}\; -2\left(3+\sqrt{17}\right)\left(\frac{5-\sqrt{17}}{2}\right)^{(\sqrt{17}-1)/4}, \tag{49}$$

and we will see later that this guessed value is indeed the correct one.

Equations (46)–(49) now give as many terms as desired of the q-expansions of the modular function $t(z)$ and modular form $f(z)$. We can (and of course did) then use this to check the correctness of these equations numerically to high accuracy by verifying the invariance of $t(z)$, and the invariance of $f(z)$ up to an automorphy factor $cz+d$, under modular transformations $z \mapsto (az+b)/(cz+d)$ in the group Π. Similarly, by inverting (45) we can also give the inverse of the uniformizing map $\mathbb{H} \to W_\Pi$ explicitly as

$$\begin{aligned} z(t) \;&=\; \frac{\alpha}{2\pi i}\,\log\frac{Q}{A} \;=\; \frac{\alpha}{2\pi i}\Bigl(\log t \,+\, \frac{y_1(t)}{y(t)} \,-\, \log A\Bigr) \\ &=\; \frac{\alpha}{2\pi i}\bigl(\log\frac{t}{A} \,+\, \tfrac{439-97\sqrt{17}}{64}\,t \,+\, \tfrac{321913-77807\sqrt{17}}{4096}\,t^2 \,+\,\cdots\bigr)\,. \end{aligned} \tag{50}$$

Exactly the same considerations apply to the second differential operator $\widetilde{L}$, with the difference that here the mirror parameter

$$\widetilde{Q} \;=\; t\,e^{\widetilde{y}_1/\widetilde{y}} \;=\; t + \tfrac{1575-369\sqrt{17}}{128}\,t^2 \,+\, \tfrac{4814915-1166773\sqrt{17}}{16384}\,t^3 \,+\,\cdots$$

is related to the variable z in the upper half-plane by

$$\widetilde{Q} \;=\; \widetilde{A}\,\widetilde{q} \qquad\text{with}\qquad \widetilde{q} \;=\; e^{2\pi i\varphi(z)/\alpha^\sigma}\,,$$

where $\varphi : \mathbb{H} \to \mathbb{H}^-$ is the twisting map and $\widetilde{A}$ is some constant. A calculation like the one for A gives the numerical value

$$\widetilde{A} \;\approx\; -40.95654078902989227160445729576 85\,,$$

which we can recognize as the "conjugate-in-the-exponent" of the value in (49):

$$\widetilde{A} \;\stackrel{?}{=}\; -2\,(3+\sqrt{17})\left(\frac{5-\sqrt{17}}{2}\right)^{(-\sqrt{17}-1)/4}. \tag{51}$$

We will show later that also this formula is indeed correct.

We can now calculate the Fourier expansions of both the twisted modular form $\widetilde{f}(z) = y_1(t)$ and the twisting map $\varphi : \mathbb{H} \to \mathbb{H}^-$ as

$$\widetilde{f}(z) \;=\; 1 \,+\, \tfrac{23-\sqrt{17}}{8}\,A\,q \,+\, \tfrac{1951-473\sqrt{17}}{256}\,A^2\,q^2 \,+\, \tfrac{184453-44739\sqrt{17}}{8192}\,A^3\,q^3 \,+\,\cdots$$

and

$$\begin{aligned}
\varphi(z) \;&=\; \frac{\alpha^\sigma}{2\pi i}\,\log\frac{\widetilde{Q}}{\widetilde{A}} \;=\; \frac{\alpha^\sigma}{2\pi i}\Bigl(\log t \,+\, \frac{\widetilde{y}_1(t)}{\widetilde{y}(t)\,-\,\log\widetilde{A}}\Bigr)\\
&=\; \frac{\alpha^\sigma}{2\pi i}\bigl(\log\frac{t}{\widetilde{A}} \,+\, \tfrac{1575-369\sqrt{17}}{128}\,t \,+\, \tfrac{1208617-292799\sqrt{17}}{8192}\,t^2 \,+\cdots\bigr)\\
&=\; \tfrac{-9+\sqrt{17}}{8}z + \tfrac{1-\sqrt{17}}{4\pi i}\Bigl(\tfrac{-\sqrt{17}}{2}\log(\tfrac{5-\sqrt{17}}{2}) \,+\, \tfrac{697-175\sqrt{17}}{128}\,A\,q\\
&\qquad +\, \tfrac{-29767+7249\sqrt{17}}{8192}\,A^2\,q^2 \,+\, \tfrac{3091637-749587\sqrt{17}}{393216}\,A^3\,q^3 \,+\,\cdots\Bigr)\,.
\end{aligned} \tag{52}$$

Again these Fourier expansions, unlike the expansions of the same functions as power series in $t(z)$, converge exponentially rapidly for all z in the upper half-plane and can be used to compute the functions $\widetilde{f}(z)$ and $\varphi(z)$ numerically and to verify the modular transformation properties (1) and $\widetilde{f}\bigl(\frac{az+b}{cz+d}\bigr) = (c^\sigma\varphi(z)+d^\sigma)\,\widetilde{f}(z)$ numerically to high accuracy, giving us the first explicit example of a non-classical twisted modular form on a Teichmüller curve.

7.2. **Modularity and integrality.** At the end of §6.3 we stated that all the coefficients of the expansions of y and $\widetilde{y}$ as power series in t belong to the ring $\mathcal{O}[\frac{1}{2}]$. This integrality has a rather puzzling aspect, which we discuss here and resolve in §8.3.

If we write y as $\sum c_n t^n$, then the differential equation $Ly=0$ translates into the recursion

$$(n+1)^2 c_{n+1} \;=\; \bigl(\beta(n^2+n)+\tfrac{3}{4}\gamma\bigr)\,c_n \;-\; \bigl(\beta(n^2-\tfrac{1}{4})+\tfrac{3}{4}\gamma\bigr)\,c_{n-1} \;+\; (n-\tfrac{1}{2})^2 c_{n-2}$$

for the coefficients c_n, where β and γ are given by (42). The integrality (away from 2) of the c_n is far from automatic from this recursion, because at each stage one has to divide a linear combination of previous coefficients by $(n+1)^2$, so that *a priori* one would only expect $n!^2 c_n$ to be 2-integral. Divisibility properties of this type are familiar from well-known recursions like the recursion

$$(n+1)^2\,A_{n+1} \;=\; (11n^2+11n+3)\,A_n - n^2\,A_{n-1} \tag{53}$$

used by Apéry in his famous proof of the irrationality of $\zeta(2)$, or the similar one he used in his even more famous proof of the irrationality of $\zeta(3)$. However, they are extremely rare. For instance, in [41] it was found that of the first 100,000,000 members of the three-parameter family of recursions obtained by varying the coefficients "11," "3" and "-1" in (53), only 7 (if one excluded certain degenerate families, and up to scaling) had integral solutions.

Apéry proved the integrality of the solution of his recursion (53) by giving the explicit closed formula $A_n = \sum_{k=0}^n \binom{n}{k}^2\binom{n+k}{n}$. We do not know a corresponding expression in our case. However, soon after Apéry's original proof, a more conceptual explanation was found by Beukers [3], who saw that the differential equation corresponding to Apéry's recursion has a modular parametrization $y=f(z)\in M_1(\Gamma)$, $t=t(z)\in M_0^{\mathrm{mer}}(\Gamma)$ of the type discussed in §4, the group Γ in this case being $\Gamma_1(5)$, and this implies the integrality because we have $f(z)\in\mathbb{Z}[[q]]$, $t(z)\in q+q^2\mathbb{Z}[[q]]$ and hence $y\in\mathbb{Z}[[t]]$. Similar statements hold for all seven of the "Apéry-like" equations mentioned above, leading to the conjecture (which was made explicitly in [41]) that the integrality property for recursions of this type occurs *precisely* when the corresponding differential equation has a modular parametrization.

The surprise is now this. In our case, just as in the seven "Apéry-like" ones, the differential equation (at least for y) *is* modular, and *does* have power series solutions with integral coefficients (away from 2), but now the modularity does not explain the integrality in the same way as above, because here the relevant Fuchsian group is not arithmetic and the q-expansions of t and y are not integral. Indeed, as we saw in §7.1, the coefficients of these q-expansions are not even algebraic numbers, since they involve powers of the scaling constant A, which according to (49) and Gelfond's theorem is a transcendental number. But even if we rescale by replacing q by $Q = Aq$, then, although the first few coefficients as listed in equations (46) and (47) have denominators that are powers of 2, this property fails if we compute more coefficients. For example, the coefficient of Q^{11} in $f(z)$ equals

$$\frac{16063132006911958155776129 - 3895881761337356780171815\sqrt{17}}{2^{53}\cdot 3^3\cdot 5\cdot 7},$$

and calculating further we find that the first 100 coefficients contain in their denominators all primes less than 100 that do not split in $\mathbb{Q}(\sqrt{17})$, and similarly for $t(z)$. Thus, although our differential equation (40) does not actually contradict the hypothetical statement

"integrality occurs only when the differential equation is modular"

mentioned above, the mechanism

$$\begin{aligned} y(t) \text{ modular } &\Rightarrow\ y \text{ and } t \text{ both have integral } q\text{-expansions}\\ &\Rightarrow\ y \text{ has an integral } t\text{-expansion}\end{aligned}$$

which previously explained that statement now breaks down completely. This puzzle, which was in fact the original motivation for the investigation described in this paper, will be solved in Section 8, where we will provide a purely modular explanation of the integrality property by expanding y and $\widetilde{y}$ with respect to *both* q and $\widetilde{q}$, using Hilbert modular forms rather than modular forms in one variable.

7.3. **The ring of twisted modular forms for W and W_Π.** We can now calculate the rings $M_{*,*}(\Gamma,\varphi)$ and $M_{*,*}(\Pi,\varphi)$ of twisted modular forms on the Teichmüller curve $W = \mathbb{H}/\Gamma$ and its double cover $W_\Pi = \mathbb{H}/\Pi$. This information will be used in the following section to embed the curve W^1_{17} into the Hilbert modular surface X_{17}.

We already know two twisted modular forms on Π, namely $f(z) = y(t(z))$ in $M_{1,0}(\Pi)$ and $\widetilde{f}(z) = \widetilde{y}(t(z))$ in $M_{0,1}(\Pi)$. (From now on we omit the "φ".) Any holomorphic or meromorphic twisted modular form of weight (k,ℓ) on Π is then equal to $f^k\widetilde{f}^\ell$ times a rational function of $t = t(z)$. The next proposition tells us which ones are holomorphic.

Proposition 7.1. *For $k,\ \ell \geq 0$ the vector space of twisted modular forms of weight (k,ℓ) on Π is given by*

$$M_{(k,\ell)}(\Pi) \;=\; \big\langle\, f(z)^k \widetilde{f}(z)^\ell\, t(z)^c \mid 3k+\ell \geq 2c \geq 0\,\big\rangle_{\mathbb{C}}\,. \tag{54}$$

Proof. The group Π has no elliptic fixed points and five cusps, one of which is irregular, so Theorem 5.5 implies the dimension formula

$$\dim M_{(k,\ell)}(\Pi) = 1 + \left\lfloor \frac{3k+\ell}{2} \right\rfloor\,. \tag{55}$$

Since the right-hand side of this equals the number of monomials $f^k \widetilde{f}^l t^c$ in (54), it suffices to prove that each of these monomials is holomorphic or equivalently, that $f(z)$ and $\widetilde{f}(z)$ are holomorphic everywhere (including at the cusps) and vanish to orders $1/2$ and $3/2$, respectively, at the cusp $t = \infty$, where the order is measured with respect to the local parameter $1/t$. The holomorphy at the cusps is a special case of Proposition 5.6, since the construction of the defining equation (38) of W given in §6.2 was based on choosing the coefficients in such a way as to make the two differential forms $\omega = dX/Y$ and $\widetilde{\omega} = X\,dX/Y$ stable at all of the cusps of W. We therefore only need to check that the order of vanishing of f and $\widetilde{f}$ at $z = 0$ (corresponding to $t = \infty$) are at most, and hence exactly, equal to $3/2$ and $1/2$, respectively. We will give two arguments to see this.

The first way is to use the action of the element $S = M_i$ of Γ, which corresponds to the involution $t \mapsto 1/t$ on the base of the family (38). This involution extends via $X \mapsto X/t$ and $Y \mapsto Y/t^{5/2}$ to an involution ι of the whole family, with $\iota^*\omega(t) = t^{3/2}\omega(t)$ and $\iota^*\widetilde{\omega}(t) = t^{1/2}\widetilde{\omega}(t)$. Near $t = 0$, the section f is the period of ω along the unique cycle (up to scale) β_0 that is orthogonal to $\widetilde{\omega}(t)$ and extends across $t = 0$. Near $t = \infty$, it is the period of ω along the unique cycle (up to scale) β_∞ that is orthogonal to $\widetilde{\omega}(t)$ and extends across $t = \infty$. From these defining properties it follows that $i^*\beta_0$ is proportional to β_∞. The same argument applies for $\widetilde{f}$. In modular terms, this translates into the statement that the function $f(z)$ transforms via

$$\frac{1}{z}\, f\Big(-\frac{1}{z}\Big) \;=\; f(z)\, t(z)^{3/2} \tag{56}$$

for some appropriate choice of the square-root of $t(z)^{3/2}$ (which we need only make at one point since this function has no zeros or poles in $\mathbb{H}$), and similarly

$$\frac{1}{\varphi(z)}\, \widetilde{f}\Big(-\frac{1}{z}\Big) \;=\; \widetilde{f}(z)\, t(z)^{1/2}\,. \tag{57}$$

Equations (56) and (57) clearly imply the statement that the modular forms f^2t^3 and $\widetilde{f}^2 t$ at ∞ are holomorphic everywhere, as claimed.

The other approach, not using the accidental fact that the cusps 0 and ∞ of $\mathbb{H}/\Pi$ happen to be interchanged by an element in the normalizer of Π and therefore applicable in other situations, is based on the equation

$$\frac{\alpha}{2\pi i}\, t'(z) \;=\; f(z)^2\, A(t(z))\,, \tag{58}$$

where $A(t)$ is the 4th degree polynomial given in (41), see (19). Since the polynomial $A(t)$ is divisible by t, we find that $f(z)^2$ multiplied by a cubic polynomial in $t(z)$ is equal to the logarithmic derivative of the modular function $t(z)$, and hence is holomorphic at the cusp $t = \infty$, so the order of f at $t = \infty$ is $\le 3/2$, as desired. A similar argument, this time using (21), applies also to $\widetilde{f}$, with the polynomial $A(t)$ replaced by the rational function $\widetilde{A}(t)$, which grows like t^2 as $t \to \infty$. $\square$

Remarks. i) We make some comments about the half-integer order of f and $\widetilde{f}$ at ∞ and about the appearance of the function $\sqrt{t(z)}$ in (56) and (57). By the discussion in §6.1, the image $\bar{\Pi}$ of Π in $\bar{\Gamma} = \Gamma/\{\pm 1\}$ has index 2 and is hence normal, so $S\bar{\Pi}S^{-1} = \bar{\Pi}$ and hence $F|_k S$ must be a modular form on Π for any modular form F (twisted or not) of *even* weight k on Π. But the subgroup Π of Γ is not normal and is not normalized by S, so the intersection of Π with $\Pi' = S\bar{\Pi}S^{-1}$

is a proper subgroup (of index 2) in Γ. The space of holomorphic modular forms of weight 1 on this group is 4-dimensional, spanned by the functions $ft^{j/2}$ with $0 \le j \le 3$, with the spaces of modular forms of weight 1 on Π and Π' separately being spanned by (f, ft) and by $(ft^{1/2}, ft^{3/2})$, respectively. Notice that the group $\Pi \cap \Pi'$ does not contain the stabilizer M_∞ of ∞ in Γ (or Π), but only its square, so that it has width 2 and hence a local uniformizer $q^{1/2}$ at ∞. This group has genus 0, with $t(z)^{1/2}$ as a Hauptmodul.

ii) In the above proof we gave an implicit estimate of the period integrals that define f and $\widetilde{f}$ in the neighborhood of any cusp. These periods are given by integrating $\omega = dX/Y$ and $\widetilde{\omega} = X\,dX/Y$ over a linear combination of paths that are invariant under the local monodromy around the given cusp. It is perhaps worthwhile giving a more explicit proof in the special case at hand, since this makes the argument clearer and also shows how to find the full expansion, and not just the order of vanishing, of f and $\widetilde{f}$ at every cusp. We will just give the main formulas, without complete details. At each cusp we choose a local parameter $\varepsilon = \varepsilon_j$. Then we can find the expansions of $y(t) = f(z)$ and $\widetilde{y}(t) = \widetilde{f}(z)$ near $t = t_j$ by looking at the explicit form of the degenerations of the differentials ω and $\widetilde{\omega}$ there, as explained at the end of the discussion in §6.2. We consider the cusp $t = \infty$ here and the other cusps even more briefly in iii) below. Near $t = \infty$ we make the substitutions $(X, t) = \bigl(\frac{cT^2 - A}{\varepsilon}, \frac{1}{\varepsilon}\bigr)$, with $c = \frac{17 - 3\sqrt{17}}{2}$ and $A = 5\,\frac{5+\sqrt{17}}{2}$ as in (39), and where ε tends to 0. Then by direct computation we find

$$P_5(X,t) \;=\; (c/\varepsilon)^5 \left[T\left(T^2 - \lambda_1^2\right)\left(T^2 - \lambda_2^2\right)\right]^2 \;+\; \mathrm{O}\left(\varepsilon^{-4}\right),$$

with $\lambda_1 = \frac{5+\sqrt{17}}{2}$ and $\lambda_2 = \frac{3+\sqrt{17}}{4}$. The fact that the leading coefficient of the right-hand side as a Laurent series in ε is a square corresponds to the degeneration of the fiber over the cusp to a rational curve, and lets us compute the differential form $\omega = dX/\sqrt{P_5(X,t)}$ as

$$\omega \;=\; \frac{2\,\varepsilon^{3/2}}{c^{3/2}} \left[\frac{1}{\left(T^2 - \lambda_1^2\right)\left(T^2 - \lambda_2^2\right)} \;+\; \cdots\right] dT\,,$$

where the omitted terms contain higher powers of ε with coefficients that are rational functions of T having poles only at $\pm\lambda_1$, $\pm\lambda_2$ that can easily be found explicitly with a suitable mathematical software program. The factor $\varepsilon^{3/2}$ gives the vanishing order we claimed, and the rest of the expansion gives us the complete expansion of $f(z)$ near $z = 0$. Specifically, the homology of $\mathbb{P}^1 \smallsetminus \{\pm\lambda_1, \pm\lambda_2, \infty\}$ is spanned by the four small loops γ_i around the four poles $\pm\lambda_1$ and $\pm\lambda_2$, and the integral of the above form around each such loop is given simply by the residue of the form at that pole, so that we can easily calculate the periods around each γ_i to any order in ε. When one does this calculation, one finds that these integrals are given, up to a constant, by

$$\int_{\gamma_1,\gamma_2,\gamma_3,\gamma_4} \omega(\varepsilon) \;\doteq\; \left(1, -1, \tfrac{1+\sqrt{17}}{2}, -\tfrac{1+\sqrt{17}}{2}\right)\left(1 + \tfrac{81 - 15\sqrt{17}}{16}\varepsilon + \tfrac{4845 - 1155\sqrt{17}}{64}\varepsilon^2 + \cdots\right).$$

The surprising observation that they are all proportional is explained by the fact the loops γ_i correspond to the core curves (in both directions) of parallel cylinders in the generating flat surface, here concretely the vertical cylinders in Figure 1 (a). More precisely, these curves stay parallel in a neighborhood of the cusp of the Teichmüller curve by definition of these curves, and hence the periods remain proportional.

j	t	z	w_j	A_j
1	0	∞	$\alpha = \frac{1+\sqrt{17}}{2}$	$-2\,(3+\sqrt{17})\left(\frac{5-\sqrt{17}}{2}\right)^{\frac{\sqrt{17}-1}{4}}$
2	1	1	$8+2\sqrt{17}$	$4\,(3+\sqrt{17})\,(4+\sqrt{17})^{\frac{5+\sqrt{17}}{4}}$
3	λ^{-1}	$\alpha/2$	$\frac{-3+\sqrt{17}}{2}$	$2^{11}\,(4+\sqrt{17})^9\left(\frac{3+\sqrt{17}}{2}\right)^{-\frac{39+\sqrt{17}}{2}}$

TABLE 1. Cusps

Another striking property of the above expansion, namely the integrality (up to powers of two) of its coefficients as a power series in the local parameter $\varepsilon = 1/t$ at the cusp $t = \infty$, is obvious both from the proof in [4] and from the one that we will give in Section 8.

iii) We now also indicate briefly how to find the Fourier expansions of f and $\widetilde{f}$ at a cusp t_j other than $t = 0$. To define them, we must first choose a matrix M_j mapping the point z_j with $t(z_j) = t_j$ to ∞. For definiteness's sake we choose $M_1 =$Id and $M_j = \left(\begin{smallmatrix} 0 & -1 \\ 1 & -z_j \end{smallmatrix}\right)$ for $j \neq 1$. The *width* w_j of the cusp is defined as the smallest positive number w with $M_j^{-1}\left(\begin{smallmatrix} 1 & w \\ 0 & 1 \end{smallmatrix}\right)M_j \in \Gamma$, and we define $q_j = e^{2\pi i M_j(z)/w_j}$ and $Q_j = A_j q_j$, where A_j is chosen so that $Q_j = \varepsilon + \mathrm{O}(\varepsilon^2)$ as $\varepsilon \to 0$. Notice that both the width w_j and the value of the scaling constant A_j depend on our choices of ε_j and M_j; with the choices given above, they are given as in Table 1 for representatives of Γ-equivalence classes of cusps. At each cusp we expand the integral as a power series in the local parameter, observing that each coefficient is the integral of a rational function on a punctured Riemann sphere, and proceed just as we did above for $t \to \infty$. We omit the calculations and give only the results (for f; those for $\widetilde{f}$ can be obtained in the same way):

$$\begin{aligned} f|_1 M_2 &= \frac{1}{8}(3+\sqrt{17})^{5/2}(4+\sqrt{17})^{1/2}\left(1 - \tfrac{3}{4}Q_3 + \tfrac{3807+915\sqrt{17}}{128}Q_3^2 + \cdots\right), \\ f|_1 M_3 &= \frac{-i}{2^{29/2}}(3+\sqrt{17})^{11}(4+\sqrt{17})^{-4}\left(1 + \tfrac{-255+1959\sqrt{17}}{1024}Q_4 + \cdots\right). \end{aligned}$$

This concludes our discussion of the modular forms and twisted modular forms on the group Π. Proposition 7.1 also allows us to describe the twisted modular forms on the Teichmüller curve itself, i.e. for the group Γ. Since this group contains $-I$, there are no twisted modular forms of weight (k,ℓ) for $k+\ell$ odd. Theorem 3.2 and the arguments in the preceding proof immediately imply the following statement.

Proposition 7.2. *The ring of weight modular forms of even total weight $k+\ell$ for Γ has the graded dimensions*

$$\dim M_{(k,\ell)}(\Gamma) \;=\; 1 + \left\lfloor \frac{3k+\ell}{4} \right\rfloor .$$

It consists of the invariants of $M_{(k,\ell)}(\Pi)$ under the involution

$$f \mapsto f t^{3/2}, \quad \widetilde{f} \mapsto \widetilde{f} t^{1/2}.$$

As a corollary, we see that that the ring of *parallel weight* twisted modular forms on Γ is freely generated by the two forms

$$\xi \;=\; (1-t)^2\, f\widetilde{f}\,, \qquad \eta \;=\; \frac{45-11\sqrt{17}}{8}\; t\, f\widetilde{f} \;=\; -\frac{19+5\sqrt{17}}{4}\, s\,\xi\,, \tag{59}$$

with s as in (35). (The numerical factors in the definition of η have been included for later convenience.) In particular, when we embed the Teichmüller curve into the Hilbert modular surface, then the restriction of a Hilbert modular form of weight (k,k) is a homogeneous polynomial of degree k in ξ and η, or equivalently equals ξ^k times a polynomial in s of degree at most k. We will use this in the next section to find an explicit description of this embedding.

8. The Hilbert modular embedding of W_{17}^1

We continue to study the example of the special Teichmüller curve $W = W_{17}^1$, using the information given in the previous section to give a complete description of the embedding of W into the Hilbert modular surface X_{17}.

8.1. **Modular embedding via Eisenstein series.** We can use our explicit knowledge of the twisting map $\varphi:\mathbb{H}\to\mathbb{H}^-$ and of the twisted modular forms on Π and on Γ to embed $\mathbb{H}/\Pi$ and $\mathbb{H}/\Gamma$ into X_{17} by expressing the restrictions of Hilbert modular forms as polynomials in f, $\widetilde{f}$, and t (or simply in ξ and η if we restrict our attention to forms of parallel weight). We begin by recalling the main points of the theory of Hilbert modular forms, using $\mathrm{SL}(2,\mathcal{O})$ rather than $\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O})$ since it is more familiar, and then later transform our results back to $\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O})$ for $D=17$, using the isomorphism of the two groups in this case.

A *Hilbert modular form* of weight (k,ℓ) on the full modular group $\mathrm{SL}(2,\mathcal{O}_D)$ is a holomorphic function $F:\mathbb{H}^2\to\mathbb{C}$ satisfying the transformation law $F(\gamma z_1,\,\gamma^\sigma z_2) = (cz_1+d)^k(c^\sigma z_2+d^\sigma)^\ell F(z_1,z_2)$ for $g = \left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \mathrm{SL}(2,\mathcal{O}_D)$. If $k=\ell$ ("parallel weight"), we call the weight simply k. We denote the space of such forms by $M_{k,\ell}(\mathrm{SL}(2,\mathcal{O}_D))$, or simply $M_k(\mathrm{SL}(2,\mathcal{O}_D))$ if $k=\ell$, and more generally write $M_{k,\ell}(\Gamma,\chi)$ for the corresponding space for forms with respect to a subgroup Γ of $\mathrm{SL}(2,\mathcal{O}_D)$ and character χ. On $M_k(\mathrm{SL}(2,\mathcal{O}_D))$ we have the involution induced by the symmetry $\iota:(z_1,z_2)\mapsto(z_2,z_1)$ of $\mathbb{H}^2$, so we can split this space into the direct sum of the subspaces $M_k^\pm(\mathrm{SL}(2,\mathcal{O}_D))$ of symmetric and antisymmetric Hilbert modular forms.

A Hilbert modular form F on $\mathrm{SL}(2,\mathcal{O}_D)$ has a Fourier expansion of the form

$$F(\mathbf{z}) \;=\; F(z_1,z_2) \;=\; b_0 \;+ \sum_{\nu\in\mathcal{O}_D^\vee,\ \nu\gg 0} b_\nu\; \mathbf{e}\big(\mathrm{tr}(\nu\mathbf{z})\big) \tag{60}$$

where $\mathcal{O}_D^\vee$ is the inverse discriminant and $\mathrm{tr}(\nu\mathbf{z})$ for $\mathbf{z}=(z_1,z_2)\in\mathbb{H}^2$ means $\nu z_1+\nu^\sigma z_2$. If we choose a $\mathbb{Z}$-basis for $\mathcal{O}_D^\vee$ and write X and Y for the corresponding exponential functions $\mathbf{e}(\mathrm{tr}(\nu\mathbf{z}))$, then the right-hand side of (60) becomes a Laurent series in X and Y (or even a power series if the basis is chosen appropriately). In practice it is sometimes more convenient to choose only a $\mathbb{Q}$-basis for $\mathcal{O}_D^\vee$, in which case we work with power series with congruence conditions on the exponents of X and Y. The simplest choice is

$$X \;=\; X(\mathbf{z}) \;=\; \mathbf{e}\Big(\frac{z_1+z_2}{2}\Big)\,, \qquad Y \;=\; Y(\mathbf{z}) \;=\; \mathbf{e}\Big(\frac{z_1-z_2}{2\sqrt{D}}\Big)\,, \tag{61}$$

in which case $\{\mathbf{e}(\mathrm{tr}(\nu\mathbf{z}))\} = \{X^m Y^n \mid m \equiv n \pmod 2\}$ and (60) becomes

$$F(\mathbf{z}) \;=\; \sum_{m=0}^{\infty} \Biggl(\sum_{\substack{|n| \le |m|\sqrt{D} \\ n \equiv m \pmod 2}} b_{m,n}\, Y^n \Biggr) X^m \quad \in\; \mathbb{C}[Y, Y^{-1}][[X]]\,. \tag{62}$$

For Hilbert modular forms the same remarks as in §5.1 for Hilbert modular surfaces apply concerning the (less standard) Hilbert modular groups $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$ or functions on $\mathbb{H}\times\mathbb{H}^-$, the only difference in the latter case being that the condition $\nu \gg 0$ in (60) must be replaced by $\nu > 0 > \nu^\sigma$.

The simplest Hilbert modular forms to construct are the Eisenstein series of weight (k,k) ($k = 2, 4, 6, \dots$), with Fourier expansion given by[3]

$$\mathcal{E}_k^D(\mathbf{z}) \;=\; \frac{\zeta_K(1-k)}{4} \;+\; \sum_{\nu \in \mathcal{O}_D^\vee,\ \nu \gg 0} \sigma_{k-1}^K\bigl(\nu\sqrt{D}\bigr)\, \mathbf{e}(\mathrm{tr}(\nu\mathbf{z}))\,. \tag{63}$$

Here $\zeta_K(s)$ denotes the Dedekind zeta function of $K = \mathbb{Q}(\sqrt{D})$ and $\sigma_{k-1}^K(\nu\sqrt{D})$ for $\nu \in \mathcal{O}_D^\vee$ with $\nu \gg 0$ is given by

$$\sigma_{k-1}^K\bigl(\nu\sqrt{D}\bigr) \;:=\; \sum_{\mathfrak{b} \mid \nu\sqrt{D}} N(\mathfrak{b})^{k-1} \;=\; \sum_{d \mid \nu\sqrt{D}} d^{k-1}\, \sigma_{k-1}\Bigl(\frac{D\nu\nu'}{d^2}\Bigr)\,,$$

where the first sum runs over integral ideals $\mathfrak{b}$ of K dividing the integral ideal $\nu\sqrt{D}$ and the second sum (whose equality with the first is proved in [40], Lemma on p. 66) runs over positive integers d such that $d^{-1}\nu \in \mathcal{O}_D^\vee$, and where $\sigma_{k-1}(m)$ for $m \in \mathbb{N}$ has its usual meaning as the sum of the $(k-1)$st powers of the (positive) divisors of m. In particular, for $D = 17$ the first three Eisenstein series $\mathcal{E}_k = \mathcal{E}_k^{17}$ have Fourier expansions beginning

$$\begin{aligned}
\mathcal{E}_2(\mathbf{z}) \;&=\; \frac{1}{12} + \bigl(3Y^3 + 7Y + 7Y^{-1} + 3Y^{-3}\bigr)\, X \;+\; \bigl(Y^8 + 21Y^6 + 14Y^4 \\
&\qquad + 45Y^2 + 18 + 45Y^{-2} + 14Y^{-4} + 21Y^{-6} + Y^{-8}\bigr)\, X^2 \;+\; \cdots\,, \\
\mathcal{E}_4(\mathbf{z}) \;&=\; \frac{41}{120} + \bigl(9Y^3 + 73Y + 73Y^{-1} + 9Y^{-3}\bigr)\, X \;+\; \bigl(Y^8 + 657Y^6 \\
&\qquad + 2198Y^4 + 5265Y^2 + 4914 + 5265Y^{-2} + \cdots + Y^{-8}\bigr)\, X^2 \;+\; \cdots\,, \\
\mathcal{E}_6(\mathbf{z}) \;&=\; \frac{5791}{252} + \bigl(33Y^3 + 1057Y + 1057Y^{-1} + 33Y^{-3}\bigr)\, X \;+\; \bigl(Y^8 + 34881Y^6 \\
&\qquad + 371294Y^4 + 1116225Y^2 + 1419858 + \cdots + Y^{-8}\bigr)\, X^2 \;+\; \cdots\,,
\end{aligned}$$

with X and Y as in (61). As a check, we can verify that if set $X = q$ and $Y = 1$ then these Fourier expansions agree to the accuracy computed (several hundred terms) with those of the classical $\mathrm{SL}(2,\mathbb{Z})$ modular forms $\frac{1}{12}E_4$, $\frac{41}{120}E_4^2$, and $\frac{44}{3}E_4^3 + \frac{2095}{252}E_6^2$, respectively.

We now compute the restrictions of these Eisenstein series to the Teichmüller curve $W = \mathbb{H}/\Gamma$ (or rather, of its double covering $\mathbb{H}/\Pi$) that we studied in §6. As explained there, the algorithm used produced Γ as a subgroup of $\mathrm{SL}(2,\mathcal{O})$, rather

[3]For general discriminants there are several Hilbert-Eisenstein series for each value of k, with sum $\mathcal{E}_k^D$. If the class number of D is one, as is the case for $D = 17$, there is only one.

than $\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O})$, so that the twisting function $\varphi(z)$ maps $\mathbb{H}$ to $\mathbb{H}^-$ (cf. eq. (52)). Hence we must use the embedding of W into $\mathbb{H}^2/\mathrm{SL}(2,\mathcal{O})$ by

$$z \quad\mapsto\quad (z_1,z_2) \;=\; \bigl(\varepsilon z,\,\varepsilon^\sigma\varphi(z)\bigr) \qquad (z\in\mathbb{H})\,, \tag{64}$$

where $\varepsilon=4+\sqrt{17}$ is the fundamental unit of $\mathbb{Q}(\sqrt{17})$. Using the expansions of Q and $\widetilde{Q}$ as power series in t that were given in §7.1, we find that the monomial $X^mY^n=\mathbf{e}(\mathrm{tr}(\nu\mathbf{z}))$ for $\nu=\frac{m}{2}+\frac{n}{2\sqrt{17}}\in\mathcal{O}^\vee$ has the t-expansion

$$\begin{aligned} X^mY^n \;&=\; \mathbf{e}\bigl(\varepsilon\nu z\,+\,\varepsilon^\sigma\nu^\sigma\varphi(z)\bigr) \;=\; (Q/A)^{\varepsilon\alpha\nu}\,(\widetilde{Q}/\widetilde{A})^{(\varepsilon\alpha\nu)^\sigma} \\ &=\; \bigl(-2(3+\sqrt{17})\bigr)^{-\frac{21m+5n}{2}}\left(\frac{5-\sqrt{17}}{2}\right)^{-8m-2n} t^{\frac{21m+5n}{2}} \\ &\quad\times\left(1\,+\,\frac{16597m+3985n-(3827m+919n)\sqrt{17}}{128}\,t\,+\,\cdots\right). \end{aligned} \tag{65}$$

(More precisely, this is true under the assumption that the constants A and $\widetilde{A}$ relating $q=\mathbf{e}(z/\alpha)$ and $\widetilde{q}=\mathbf{e}(\varphi(z)/\alpha^\sigma)$ to $Q=t+\cdots$ and $\widetilde{Q}=t+\cdots$ indeed have the values given in (49) and (51); we will return to this point below.) Inserting the expansion (65) into the Fourier development of $\mathcal{E}_k(\mathbf{z})$ or any other Hilbert modular form, we can compute its restriction to W as a power series in t. On the other hand, as we saw above, the restriction of any $F\in M_k(SL(2,\mathcal{O}))$ to W is a linear combination of monomials $\xi^i\eta^j$ with $i+j=k$, where ξ and η are the functions defined in (59), whose expansions in t are known, so we can find the coefficients of this linear combination by linear algebra. The result of the computation for the first three Eisenstein series is

$$\begin{aligned} 12\,\mathcal{E}_2\big|_W \;&=\; \xi^2\,-\,\frac{11+\sqrt{17}}{4}\,\xi\eta\,+\,\eta^2\,, \\ 120\,\mathcal{E}_4\big|_W \;&=\; 41\,\xi^4\,-\,\frac{1855+365\sqrt{17}}{2}\,\xi^3\eta\,+\,\frac{18245+3979\sqrt{17}}{8}\,\xi^2\eta^2 \\ &\qquad+\,\frac{151+35\sqrt{17}}{4}\,\xi\eta^3\,+\,41\,\eta^4\,, \\ 252\,\mathcal{E}_6\big|_W \;&=\; 5791\,\xi^6\,-\,\frac{867831+173541\sqrt{17}}{4}\,\xi^5\eta\,+\,\frac{11350461+2429643\sqrt{17}}{16}\,\xi^4\eta^2 \\ &\qquad+\,\frac{1883335+652933\sqrt{17}}{16}\,\xi^3\eta^3\,+\,\frac{69270195+16881483\sqrt{17}}{64}\xi^2\eta^4 \\ &\qquad+\,\frac{1983525+452397\sqrt{17}}{32}\xi\eta^5\,+\,5791\,\eta^6\,. \end{aligned}$$

(It was to simplify the coefficients in these polynomials that we introduced the factor c in (59). The coefficient of ξ^k here is just the constant term of $\mathcal{E}_k$, and in particular rational, and the coefficient of η^k has the same value because the matrix $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ belongs to $\mathrm{SL}(2,\mathcal{O})$, though not to Γ, so that the constant terms at the two cusps $s(0)=0$ and $s(1)=\infty$ of the restriction of any Hilbert modular form to W are the same up to scaling.) Now by elimination we find a polynomial

$$\begin{aligned} P(\mathcal{E}_2,\,\mathcal{E}_4,\,\mathcal{E}_6) \;&=\; \bigl(3465994203567-840620808790\,\sqrt{17}\bigr)\,\mathcal{E}_6^4 \\ &\qquad+\,\cdots\,+\,\bigl(7395484320944244318526129490625 \\ &\qquad-\,1711627845603248913114298550625\,\sqrt{17}\bigr)\,\mathcal{E}_2^{12} \end{aligned}$$

(in which we have omitted seventeen equally gigantic intermediate terms) whose restriction to W vanishes. We have thus obtained an explicit algebraic equation cutting out the Teichmüller curve W on the Hilbert modular surface X_{17}, but it is too big in the sense that its vanishing locus is reducible and W is only one of its components. Indeed, we know from the results of Bainbridge [1] that there must be a symmetric Hilbert modular form of weight 12 which vanishes precisely on the Teichmüller curve and its image under the involution $\iota : (z_1, z_2) \mapsto (z_2, z_1)$, whereas the above equation has weight 24. The reason for this is twofold:

- the Eisenstein series generate only a subring of the full ring of symmetric Hilbert modular forms, and there is no reason that the minimal defining equation of W should belong to this subring; and
- we did not even use all the Eisenstein series, but only $\mathcal{E}_2$, $\mathcal{E}_4$ and $\mathcal{E}_6$.

The second point can be dealt with by expressing the restriction of each $\mathcal{E}_k$ to W as a polynomial in ξ and η and looking for the first weight in which some linear combination of monomials in these restricted Eisenstein series vanishes. This weight, however, still turns out to be 14 rather than 12, and the answer is not even unique: there is a 2-dimensional space of linear combinations of $\mathcal{E}_2^7, \ldots, \mathcal{E}_{14}$ that vanish on W, all having huge coefficients so that we do not reproduce them here. (As a side remark, it is actually surprising that there should be even one relation in such a low weight, let alone two, since there are 15 monomials of weight 14 in $\Gamma_2, \ldots, \Gamma_{14}$ and also 15 monomials of weight 14 in ξ and η, so that one would not expect the former to lie in a non-trivial subspace, let alone a subspace of codimension 2. The first weight in which there has to be a relation for dimensional reasons is 16. Similarly, it is surprising that we found a relation among $\mathcal{E}_2$, $\mathcal{E}_4$ and $\mathcal{E}_6$ in as low a weight as 24, since *a priori* the first weight in which the number of monomials in these three forms is larger than the number of monomials in ξ and η is 38. This suggests that the restrictions of the Eisenstein series to W have some non-generic property, but we do not know what it is.) To address the first point, we need to have a full set of generators of the ring of Hilbert modular forms. Finding such a set of generators in general is a comparatively difficult (though always algorithmically solvable) problem, but in the case $D = 17$ the result has been obtained by Hermann [14]. We will describe his results in the next subsection and use them to determine the symmetric Hilbert modular form of minimal weight 12 vanishing on W.

Another pertinent remark is that, although we have so far only used Hilbert modular forms of "parallel" weight (k, k), whose restriction to W is a polynomial in ξ and η, one can equally well consider Hilbert modular forms of mixed weight (k, ℓ) with $k \neq \ell$, in which case the restrictions become polynomials in f, $\widetilde{f}$ and $t^{\pm 1}$. For example, the Rankin-Cohen bracket $2E_2'(\mathbf{z})E_4(\mathbf{z}) - E_2(\mathbf{z})E_4'(\mathbf{z})$, where the prime denotes differentiation with respect to the first variable z_1, is a Hilbert modular form of weight (8,6), and its restriction to W could be computed explicitly as $f(z)^8 \widetilde{f}(z)^6$ times a Laurent polynomial in $t(z)$. Such mixed weight forms will play a role in Part III of this paper, e.g. in Section 9, where we will use a different construction to find explicitly a non-symmetric Hilbert modular form of non-parallel weight (3, 9) which vanishes precisely on W, again in accordance with the general results of Bainbridge.

8.2. **Hilbert modular forms for the discriminant 17.** We continue to work with the Hilbert modular group $\Gamma_D = \mathrm{SL}(2,\mathcal{O})$ for the case $D = 17$, i.e., for $\mathcal{O} = \mathbb{Z}[\alpha]$ with $\alpha = (1+\sqrt{17})/2$. We will describe the structure of the ring of symmetric Hilbert modular forms of parallel even weight, following Hermann [14], who obtains these modular forms by restriction of Siegel modular forms of genus 2. Later we will look also at non-symmetric Hilbert modular forms and Hilbert modular forms of odd or non-parallel weight on Γ_{17}.

Hermann begins with the sixteen genus 2 Siegel theta series

$$\Theta_{m,m'}(Z,v) \;=\; \sum_{x\in\mathbb{Z}^2+m} \mathbf{e}\Big(\tfrac{1}{2}xZx^t + x(v+m')^T\Big)\,. \tag{66}$$

Here $m,\,m' \in \{0,\frac{1}{2}\}^2$ and the independent variables Z and v are in the Siegel half-space $\mathbb{H}_2$ and in $\mathbb{C}^2$, respectively. Ten of these (those for which the theta characteristic (m,m') is even, i.e. $4m\cdot m' \equiv 0 \pmod 2$) are even functions of v and hence give Siegel modular forms of weight $\frac{1}{2}$ after restricting to $v=0$. (The other six are odd and hence give 0 on restriction, but their derivatives with respect to v give non-trivial restrictions that will play a crucial rule in the constructions of Part III of this paper.) Using a modular embedding from $\mathbb{H}^2$ to $\mathbb{H}_2$ like the one described in the previous sub-section, we get ten Hilbert theta series, all of weight $\frac{1}{2}$ with respect to a suitable subgroup of Γ_D.

It is convenient to re-index the sixteen theta characteristics in a way that makes the action of Γ_D more transparent. This part works for any $D\equiv 1\ (8)$, i.e., for D such that the prime 2 splits as $\pi_2\pi_2^\sigma$ in $K=\mathbb{Q}(\sqrt{D})$ for some prime ideal $\pi_2 \neq \pi_2^\sigma$ (in our case, the principal ideal generated by $1+\alpha$). Define sets S and $\widehat{S}$ by

$$S \;=\; \{0,1,\infty\}\,, \qquad \widehat{S} \;=\; S\cup\{X\}$$

where X is a symbol, and let Γ_D act on $\widehat{S}\times\widehat{S}$ by fixing X and identifying $S\times S$ with $\mathbb{P}^1(\mathcal{O}/p_2)\times\mathbb{P}^1(\mathcal{O}/\sigma(p_2))$. We match the usual indexing by characteristic with these symbols by Table 2, in which the even characteristics correspond to the pairs

$m\setminus m'$	X	0	1	∞
X	1111	1011	1010	1110
0	0111	0011	0010	0110
1	0101	0001	0000	0100
∞	1101	1001	1000	1100

TABLE 2. Reindexing of theta characteristics

$(a,b)\in\widehat{S}^2$ in which a and b are either both equal to or both different from X, and the odd characteristics to the pairs (a,b) in which exactly one of a and b is equal to X. This gives us by restriction 10 Hilbert modular forms $\theta_{X,X}$ and $\theta_{a,b}$ $(a,\,b\in S)$ of weight 1/2.

Set $\Theta = \theta_{X,X}^2$ and for each permutation π of the set S set $\eta_\pi = \pm\prod_{s\in S}\theta_{s,\pi(s)}$, where the sign is $+1$ if π is the identity and -1 otherwise. Up to powers of i, the Hilbert modular group preserves Θ and permutes the η_π, preserving the parity of π, so that if we set

$$\eta_\pm \;=\; \sum_{\pi\text{ even}}\eta_\pi \;\pm\; \sum_{\pi\text{ odd}}\eta_\pi$$

then Θ, η_+ and η_- are Hilbert modular forms, with multiplier systems, for the full Hilbert modular group. More precisely, we have

$$\Theta \in M_1(\Gamma_D, v_0), \qquad \eta_\pm^2 \in M_3(\Gamma_D, v_0), \qquad \eta_+\eta_- \in M_3(\Gamma_D, v_0^{-1})\,, \tag{67}$$

where

$$v_0 : \Gamma_D \to \mu_4, \quad \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \mapsto -1, \quad \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) \mapsto i^{\operatorname{tr}(x)} \text{ for } x \in \mathcal{O} \tag{68}$$

is a character of order 4. The form $\theta_{X,X}$ is antisymmetric with respect to the involution ι and hence vanishes on the diagonal $\mathbb{H}/\mathrm{SL}(2,\mathbb{Z}) \subset \mathbb{H}^2/\mathrm{SL}(2,\mathcal{O}_D)$. Moreover, in the case $D = 17$ this is its full vanishing locus, so that any Hilbert modular form vanishing on the diagonal is divisible by $\theta_{X,X}$ and any symmetric Hilbert modular form vanishing on the diagonal is divisible by Θ. For example, since the restrictions of both η_+^2 and η_-^2 are proportional to $\sqrt{\Delta}$ (where we use E_4, E_6 and $\Delta = (E_4^3 - E_6^2)/1728$ to denote the standard generators of $M_*(\mathrm{SL}(2,\mathbb{Z}))$), some linear combination of them, which turns out to be $\eta_-^2 - 4\eta_+^2$, vanishes on the diagonal and hence is divisible by Θ. This gives us the following five symmetric Hilbert modular forms of even weight and trivial character:

$$G_2 = \frac{\eta_-^2 - 4\eta_+^2}{\Theta}, \quad G_4 = \eta_+\eta_-\Theta, \quad H_4 = \Theta^4, \quad G_6 = \eta_-^2\Theta^3, \quad H_6 = \eta_-^3\eta_+\,,$$

where the index of each form indicates its weight.

Theorem 8.1 (Hermann, [14]). *The ring $M_{\mathrm{ev}}^+(\Gamma_{17}) = \bigoplus_{k\ge 0} M_{2k}^+(\Gamma_{17})$ of symmetric Hilbert modular forms of even weight for $D = 17$ is generated by the five Hilbert modular forms G_2, G_4, H_4, G_6 and H_6, with the relations*

$$G_4G_6 \;=\; H_4H_6\,, \quad G_4^3 \;=\; \frac{1}{4}G_6(H_6 + G_2G_4)\,, \quad G_6^2 \;=\; H_4\,(4G_4^2 - G_2G_6)\,.$$

In particular, $M_{\mathrm{ev}}^+(\Gamma_{17})$ is a free module of rank 4 over the algebra $\mathbb{C}[G_2, H_4, H_6]$, with basis $\{1, G_4, G_6, G_4^2\}$.

Sketch of proof (following Hermann). The relations among Hermann's five forms, like any relations among modular forms, can be verified algorithmically by looking at a finite part of the Fourier expansions of the functions involved, so we only have to show that these forms generate the whole ring. Let $F = F_0 \in M_k(\Gamma_{17})$ be a symmetric Hilbert modular form of even weight k with trivial character. Then the restriction of F to the diagonal has weight $2k$ divisible by 4, and since the ring of modular forms on $\mathrm{SL}(2,\mathbb{Z})$ of weight divisible by 4 is generated by the forms E_4 and Δ, which are multiples of the restrictions to the diagonal of G_2 and H_6, there is a weighted homogeneous polynomial $P_0(G_2, H_6)$ in G_2 and H_6 whose restriction to the diagonal coincides with that of F_0. By what we said above, we then have $F_0 = P_0(G_2, H_6) + \Theta F_1$ for some Hilbert modular form $F_1 \in M_{k-1}(\Gamma_{17}, v_0^{-1})$. The restriction of F_1 to the diagonal has the character of $\sqrt{\Delta}$ and weight congruent to 2 modulo 4, so by the same argument as before coincides with the restriction of $\eta_+\eta_-P_1(G_2, H_6)$ for some weighted homogeneous polynomial $P_1(G_2, H_6)$. This implies in turn $F_1 = \eta_+\eta_-P_1(G_2, H_6) + \Theta F_2$ for some $F_2 \in M_{k-2}(\Gamma_{17}, v_0^2)$. A similar argument shows that F_2 has the same restriction to the diagonal as $\eta_+^2\eta_-^2P_2(G_2, H_6)$ for yet a third polynomial P_2, so $F_2 = \eta_+^2\eta_-^2P_2(G_2, H_6) + \Theta F_3$ for some $F_3 \in M_{k-3}(\Gamma_{17}, v_0)$, and a final iteration gives a fourth polynomial P_3 such that $F_3 = \eta_-^2P_3(G_2, H_6) + \Theta F_4$ for some Hilbert modular form F_4 of weight $k-4$, now again with trivial character. Combining these successive identities we have written F

as $P_0 + G_4P_1 + G_4^2P_2 + G_6P_3 + H_4F_4$ where each P_i belongs to $\mathbb{C}[G_2, H_6]$ and $F_4 \in M^+_{k-4}(\Gamma_{17})$, and now iterating the whole argument we see that F is a linear combination of 1, G_4, G_4^2 and G_6 with coefficients in $\mathbb{C}[G_2, H_4, H_6]$ as claimed. $\square$

Example. The Fourier expansions of η_+, η_- and Θ begin

$$\begin{aligned}
\eta_+ &= -4\,X^{1/4}\bigl((Y+Y^{-1}) + (13Y^4 - 19Y^2 - 19Y^{-2} + 13Y^{-4})\,X + \cdots\bigr)\,,\\
\eta_- &= 16\,X^{1/4}\bigl(1 - (Y^5 + 3Y^3 - Y - Y^{-1} + 3Y^{-3} + Y^{-5})\,X + \cdots\bigr)\,,\\
\Theta &= 4\,X^{1/2}\bigl((Y^2 - 2 + Y^{-2}) - 2(Y^5 - Y - Y^{-1} + Y^{-5})\,X + \cdots\bigr)\,.
\end{aligned}$$

(As a check, if we set $Y = 1$ and $X = q$ then these reduce to $-8\eta^6$, $16\eta^6$, and 0.) Comparing with the expansions of the first three Eisenstein series $\mathcal{E}_k$ given above, we find that these forms are given in terms of Hermann's generators of $M^+_{\rm ev}$ by

$$-192\,\mathcal{E}_2 = G_2\,, \qquad 640\,\mathcal{E}_4 = \frac{41}{48}\,G_2^2 - 39\,G_4 - 57\,H_4\,,$$
$$14336\,\mathcal{E}_6 = -\frac{5791}{72}\,G_2^3 + 8571\,G_4G_2 + 11463\,H_4G_2 - \frac{32865}{4}\,G_6 - 6285\,H_6\,.$$

Corollary 8.2. *The function field field of the symmetric Hilbert modular surface with $D = 17$ is the rational function field is generated by the two functions*

$$U = \frac{H_4}{G_4} \quad \Bigl(= \frac{\Theta^3}{\eta_-\eta_+}\Bigr)\,, \qquad V = \frac{H_6}{H_6 - G_2G_4} \quad \Bigl(= \frac{\eta_-^2}{4\eta_+^2}\Bigr)\,. \tag{69}$$

Proof. The relations among Hermann's generators imply

$$\Bigl(\frac{G_4}{G_2^2}\,,\ \frac{H_4}{G_2^2}\,,\ \frac{G_6}{G_2^3}\,,\ \frac{H_6}{G_2^3}\Bigr) = \Bigl(\frac{UV}{4(V-1)^2}\,,\ \frac{U^2V}{4(V-1)^2}\,,\ \frac{U^2V^2}{4(V-1)^3}\,,\ \frac{UV^2}{4(V-1)^3}\Bigr)\,.$$

so the corollary follows immediately from the theorem. $\square$

8.3. **The equation of the Teichmüller curve.** The corollary just given tells us that the Hilbert modular surface for $D = 17$ is rational, with coordinates U and V. In particular, the image of the Teichmüller curves $W_{17}^{\pm}$ on this surface must be given by polynomial equations in these coordinates. In this subsection we will give these equations, which turn out to be several orders of magnitude simpler than the previously obtained equation $P(\mathcal{E}_2, \mathcal{E}_4, \mathcal{E}_6) = 0$. We will also describe better systems of Fourier coordinates and will resolve two questions that we raised earlier by showing that the values of A and $\widetilde{A}$ given in (49) and (51) on the basis of numerical computations are indeed correct and by giving a purely modular proof of the integrality (away from the prime 2) of the Taylor expansions of y and $\widetilde{y}$ as power series in t.

Theorem 8.3. *On the (rational) symmetric Hilbert modular surface with coordinates U and V, the Teichmüller curves W_{17}^1 and W_{17}^0 are given by the equations*

$$W_D^1:\quad V + \frac{5+\sqrt{17}}{2}\,U^2 + 3\,\frac{7+\sqrt{17}}{8}\,U + \frac{1-\sqrt{17}}{8} = 0 \tag{70}$$

and

$$W_D^0:\quad V + \frac{5-\sqrt{17}}{2}\,U^2 + 3\,\frac{7-\sqrt{17}}{8}\,U + \frac{1+\sqrt{17}}{8} = 0\,. \tag{71}$$

Proof. Since we gave the expressions for the Hilbert-Eisenstein series $\mathcal{E}_2$, $\mathcal{E}_4$ and $\mathcal{E}_6$ in the Hermann generators in the last subsection, we could derive (70) from the results of §8.1 giving the restrictions to the curve W_{17}^1 of these Eisenstein series. However, it is much simpler to work directly with Hermann's generators, obtaining their Fourier expansions from those of $\eta_\pm$ and Θ as given above and then using (65) to obtain the t-expansions of their restrictions to W. The results of the computations are given in the following table, in which we have used the results from the end of §7.3 to write the expansion of the restriction of each of G_4, H_4, G_6 and H_6 to W (that of $G_2 = -192\mathcal{E}_2$ was already given above) as a power of ξ times a polynomial in s:

F	$\varepsilon G_4/2^4\pi_2^8$	$\varepsilon H_4/2^4\pi_2^6$	$\varepsilon^2 G_6/2^6\pi_2^{12}$	$\varepsilon^2 H_6/2^6\pi_2^{14}$
$\xi^{-k}\,F\|W$	$s(s-1)(s-\kappa_1)$	$s(s-1)$	$s^2(s-1)^2$	$s^2(s-1)^2(s-\kappa_1)$

Here $\kappa_1 = \pi_2^{-2} = \frac{13-3\sqrt{17}}{8}$. From the definitions of the Hilbert modular functions U and V it then follows that their restrictions to W are given by

$$U\big|_W = \frac{\kappa_1}{s-\kappa_1}\,, \qquad V\big|_W = \frac{-1+\sqrt{17}}{8}\,\frac{s(s-1)}{(s-\kappa_1)^2}\,, \tag{72}$$

and equation (70) follows immediately. Equation (71) is proved in a similar way using the Galois conjugate differential equation, as discussed in the remark at the end of §5.5; the resulting expansions are the Galois conjugates of those for W_{17}^1 and since the Hermann generators have rational Fourier coefficients the final equation is necessarily also the Galois conjugate of that of W_{17}^1. In fact, this Galois conjugation property holds for all D, as was already recalled in Theorem 5.4 (iv) of §5.4. □

We end this section by discussing four points related to the equations given in Theorem 8.3.

1. Equation (70) describes a Hilbert modular function that vanishes precisely on the curve W_{17}^1. We can also ask for the holomorphic Hilbert modular form of smallest weight with the same property. If we multiply the left-hand side of (70) through by $\eta_+^2\eta_-^2$, then by equation (69) the result is

$$F_{17}^1 \;=\; \frac{1}{4}\,\eta_-^4 \;+\; \frac{5+\sqrt{17}}{2}\,\Theta^6 \;+\; 3\,\frac{7+\sqrt{17}}{8}\,\Theta^3\eta_+\eta_- \;+\; \frac{1-\sqrt{17}}{8}\,\eta_+^2\eta_-^2\,, \tag{73}$$

and by Theorem 8.3 this vanishes precisely on W_{17}^1. According to (67), F_{17}^1 is a holomorphic Hilbert modular form of weight 6 on the full Hilbert modular group, but with quadratic character v_0^2, where v_0 is defined as in (68). If we further multiply it by Θ^2, then we get a Hilbert modular form of weight 8 on the full Hilbert modular group and with trivial character, given in terms of the Hermann generators by

$$\Theta^2 F_{17}^1 \;=\; \frac{1}{4}\,G_2G_6 \;+\; \frac{5+\sqrt{17}}{2}\,H_4^2 \;+\; 3\,\frac{7+\sqrt{17}}{8}\,G_4H_4 \;+\; \frac{9-\sqrt{17}}{8}\,G_4^2\,,$$

but this form now vanishes not only on W_{17}^1, but also (doubly) on the diagonally embedded modular curve $\mathbb{H}/\mathrm{SL}(2,\mathbb{Z}) \subset \mathbb{H}^2/\mathrm{SL}(2,\mathcal{O})$. On the other hand, if we multiply F_{17}^1 by its Galois conjugate $F_{17}^0 = (F_{17}^1)^\sigma$, then the product F_{17} vanishes precisely on the full Teichmüller locus $W_{17} = W_{17}^0 \cup W_{17}^1$, and this is now a holomorphic Hilbert modular form on the full modular group $\mathrm{SL}(2,\mathcal{O})$ and with trivial character, given in terms of the basis of $M_{12}(\mathrm{SL}(2,\mathcal{O}))$ from Theorem 8.1 by

$$F_{17} \;=\; (\text{explicit polynomial in } G_2,\, G_4,\, H_4,\, G_6,\, H_6 \text{ with rational coefficients})\,.$$

We do not write out the polynomial, since it is a bit complicated, but observe that it involves only 11 of the 16 generators of $M_{12}(\mathrm{SL}(2,\mathcal{O}))$. The fact that here there is a single Hilbert modular form of weight 12 whose vanishing locus is precisely the union of the Teichmüller curves on X_D is a special case of the theorem of Bainbridge, already mentioned in §8.1, stating that such a form F_D exists for every D. We will give a different proof of this in Part III by constructing $F_D \in M_{12}(\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O}))$ in general as a product of twelve derivatives of theta series of weight $(\frac12,\frac32)$ or $(\frac32,\frac12)$.

2. The next point concerns the choice of coordinates for our Fourier expansions. We replace the previously used coordinates X and Y from (61) by the new Fourier variables

$$X_1 \;=\; X^{-3}Y^{13}, \quad Y_1 \;=\; X^5Y^{-21}\,.$$

This has several advantages. First of all they form a $\mathbb{Z}$-basis of the group of Fourier monomials $\mathbf{e}(\mathrm{tr}(x\mathbf{z}))$, whereas X and Y generated a subgroup of index 2. Secondly, symmetric Hilbert modular forms of even weight are symmetric in X_1 and Y_1, as one sees using the action of ε. For instance the (X_1,Y_1)-expansions of the first two Hermann generators begin

$$\begin{aligned} -\tfrac{1}{192}\,G_2 \;&=\; \tfrac{1}{12} \,+\, Y_1X_1 \,+\, (9Y_1^2+3Y_1^3)\,X_1^2 \,+\, (3Y_1^2+10Y_1^3+15Y_1^4)\,X_1^3 \,+\, \cdots\,,\\ \tfrac{1}{256}\,G_4 \;&=\; \qquad -\, Y_1X_1 \,+\, (-9Y_1^2+Y_1^3)\,X_1^2 \,+\, (Y_1^2-14Y_1^3+Y_1^4)\,X_1^3 \,+\, \cdots\,, \end{aligned}$$

in which the coefficients of $Y_1^3X_1^2$ and $Y_1^2X_1^3$ are the same. Thirdly, and most important, both are holomorphic near the cusp of W and hence have power series expansions in t, with valuations 1 and 0 there rather than $21/2$ and $5/2$ as for X and Y. This has to do with the Hirzebruch resolution of the cusp singularities, according the which different $\mathbb{Z}$-bases of the group just mentioned are good coordinates in different parts of the resolution cycle. (We will discuss this in much more detail in Section 11.) Here we have to choose the coordinates that are adapted to the point of the cusp resolution through which W passes. Explicitly, these expansions are

$$\begin{aligned} X_1 \;&=\; X^{-3}Y^{13} \;=\; -\tfrac{11+3\sqrt{17}}{64}\,t \,+\, \tfrac{403-229\sqrt{17}}{2048}\,t^2 \,+\, \cdots\,,\\ Y_1 \;&=\; X^5Y^{-21} \;=\; \tfrac{21-5\sqrt{17}}{2} \,-\, \tfrac{895-217\sqrt{17}}{8}\,t \,+\, \cdots\,. \end{aligned} \tag{74}$$

(For comparison, the leading terms of X and Y are $\sqrt{\frac{524445-220267\sqrt{17}}{2^{80}}t^{21/2}}$ and $\sqrt{-\frac{3+11\sqrt{17}}{2^{20}}t^{5/2}}$, respectively.) It is then very easy to restrict a Hilbert modular form F given in its (X_1,Y_1)-expansion to W: to get the expansion up to order t^n it suffices to expand F up to X_1^n in $\mathbb{C}[Y_1][[X_1]]$.

3. The third point concerns the correctness of the values of A and $\widetilde{A}$ in (49) and (51). We have been assuming throughout (for instance in (74)) that these guesses were correct, and now finally prove it. We write

$$\begin{aligned} X_1 \;&=\; -\tfrac{11+3\sqrt{17}}{64}\,c_1\,t\,\exp\Bigl(\varepsilon\alpha\lambda\frac{y_1}{y} + (\varepsilon\alpha\lambda)^\sigma\,\frac{\widetilde{y}_1}{\widetilde{y}}\Bigr)\,,\\ Y_1 \;&=\; \quad\tfrac{21-5\sqrt{17}}{2}\,c_2\,t\,\exp\Bigl(\varepsilon\alpha\mu\,\frac{y_1}{y} + \bigl(\varepsilon\alpha\mu\bigr)^\sigma\,\frac{\widetilde{y}_1}{\widetilde{y}}\Bigr)\,, \end{aligned} \tag{75}$$

where $\lambda = \frac{13-3\sqrt{17}}{2\sqrt{17}}$ and $\mu = \frac{-21+5\sqrt{17}}{2\sqrt{17}}$ (the factor ε in the exponent takes care of the identification of $\mathbb{H}\times\mathbb{H}$ and $\mathbb{H}\times\mathbb{H}^-$, the factor α is the cusp width, and the factors λ and μ come from the passage from (X,Y) to (X_1,Y_1)) and where c_1 and c_2 are numerical factors that are both equal to one if (and only if) the formulas for A and $\widetilde{A}$ in (49) and (51) are correct. (One could—and we originally did—also do the whole calculation without including the prefactors $-\frac{11+3\sqrt{17}}{64}$ and $\frac{21-5\sqrt{17}}{2}$ in (75), but then the numbers in the calculation would be even worse.) We substitute these expressions into G_2, divide by ξ^2, with ξ as in (59), and write the quotient as a power series $\sum_i C_i s^i$, with $s = \frac{4t}{(1-t)^2}$ as usual. By Proposition 7.2, we know that this power has to reduce to a quadratic polynomial if c_1 and c_2 have the correct values. If we instead treat c_1 and c_2 as unknowns, then the coefficients of this power series are polynomials in c_1 and c_2 (with huge coefficients), and we have to show that the infinite system of polynomial equations $C_i(c_1,c_2)=0$ for all $i>2$ has the unique solution $c_1=c_2=1$. By computer calculation we find that the g.c.d. of the resultants $\mathrm{Res}_{c_1}(C_3,C_4)$ and $\mathrm{Res}_{c_1}(C_4,C_5)$ is equal to $-2^{-111}c_2^8(c_2-1)$, and since c_2 cannot be 0, it must be 1. Then substituting $c_2=1$ into the g.c.d. of C_3 and C_4 gives $c_1=1$.

4. The final point concerns the integrality (away from the prime 2) of our Fourier expansions. The coefficients of the expansion of any Hilbert modular form in X_1 and Y_1 with rational coefficients always has bounded denominators, like in the examples for G_2 and G_4 above. The same is true also for Hilbert modular functions, e.g.

$$\begin{aligned} U+1 \;&=\; (Y_1-Y_1^2)\,X_1 \;+\; (-Y_1-15Y_1^2+17Y_1^3-Y_1^4)\,X_1^2 \;+\;\cdots\,,\\ \tfrac14 V \;&=\; Y_1\,X_1 \;-\; (22Y_1^2+2Y_1^3)\,X_2 \;+\; (-2Y_1^2+289Y_1^3+12Y_1^4+Y_1^5)\,X_1^3 \;+\;\cdots \end{aligned}$$

for the generators of the field of symmetric Hilbert modular functions given in Corollary 8.2 above. We introduce new generators of this function field, namely $U_1 = 1-\frac{4(U+1)}{V}$ and $V_1=\frac14 V$. (Other choices would be equally good.)

Proposition 8.4. *In the new set of generators the equality*

$$\mathbb{Z}[[X_1,Y_1]]^{\mathrm{sym}} \;=\; \mathbb{Z}[[U_1,V_1]]$$

of power series rings holds.

Proof. We have $\mathbb{Z}[[X_1,Y_1]]^{\mathrm{sym}} = \mathbb{Z}[[S,P]]$, where $S=X_1+Y_1$ and $P=X_1Y_1$. We have already seen that U and V, and hence also U_1 and V_1, are symmetric in X_1 and Y_1. We can thus express U_1 and V_1 as a power series in S and P. Concretely, these expansions start

$$\begin{aligned} U_1 \;&=\; S \;+\; (-7+3S+3S^2)\,P \;+\; (13-65S+37S^2+6S^3)P^2 \;+\;\cdots\,,\\ V_1 \;&=\; P \;+\; (-22-2S)\,P^2 \;+\; (289+12S+S^2)\,P^3 \;+\;\cdots\,. \end{aligned}$$

They have integral expansions, so $\mathbb{Z}[[U_1,V_1]] \subseteq \mathbb{Z}[[S,P]]$. Conversely, since the expansions begin $S+\mathrm{O}(P)$ and $P+\mathrm{O}(P^2)$, we can recursively compute S and P as power series in U_1 and V_1, and these again have integral coefficients:

$$\begin{aligned} S \;&=\; U_1 \;+\; (7+3U_1+3U_1^2)\,V_1 \;+\; (120-20U_1-82U_1^2+6U_1^3)\,V_1^2 \;+\;\cdots\,,\\ P \;&=\; V_1 \;+\; (22+2U_1)\,V_1^2 \;+\; (693+158U_1+U_1^2)\,V_1^3 \;+\;\cdots\,. \end{aligned}$$

It follows that $\mathbb{Z}[[U_1,V_1]] \supseteq \mathbb{Z}[[S,P]]$, as desired. $\square$

Corollary 8.5. *The restriction to W_Π of any symmetric Hilbert modular form of even weight with integral Fourier coefficients belongs to $R[[t]]$, where $R = \mathcal{O}_{17}[\frac{1}{2}]$.*

Proof. Since $\kappa_0 \in R^\times$, eq. (35) implies that $R[[t]] = R[[s]]$. Since $\kappa_1 \in R^\times$, eq. (72) implies that the restrictions of U_1 and W_1 to W belong to this ring. Since any Hilbert modular form has a Fourier expansion with exponents in a cone strictly contained in the positive quadrant (explicitly, if $F = \sum_{r,s} a_{r,s} X_1^r Y_1^s$, then $a_{r,s} = 0$ unless $\frac{9-\sqrt{17}}{8} r \le s \le \frac{9+\sqrt{17}}{8} r\big)$, it contains only finitely many monomials contributing to any fixed power of t in the t-expansion of its restriction to W. (Explicitly, $X_1^r Y_1^s + X_1^s Y_1^r$ is divisible by $P^{\min(r,s)}$, and $P\big|_W = \mathrm{O}(t)$.) The corollary follows. □

Proposition 8.6. *The power series $y(t)$ and $\widetilde{y}(t)$ have $\mathcal{O}_{17}$-integral expansions up to denominator 2, i.e. $y(t), \widetilde{y}(t) \in R[t]$, where $R = \mathcal{O}_{17}[\frac{1}{2}]$.*

Proof. The differential operators $D_1 = \frac{\sqrt{17}}{2\pi i}\frac{\partial}{\partial z_1}$ and $D_2 = \frac{\sqrt{17}}{2\pi i}\frac{\partial}{\partial z_2}$ can be written as

$$D_1 \;=\; \lambda\, X_1 \frac{\partial}{\partial X_1} \;+\; \mu\, Y_1 \frac{\partial}{\partial Y_1}\,, \quad D_2 \;=\; \lambda^\sigma\, X_1 \frac{\partial}{\partial X_1} \;+\; \mu^\sigma\, Y_1 \frac{\partial}{\partial Y_1}\,,$$

where $\lambda = \frac{13-3\sqrt{17}}{2}$ and $\mu = \frac{-21+5\sqrt{17}}{2}$, and hence map $R[[U_1, V_1]]$ to itself. On the other hand, they send Hilbert modular functions to meromorphic Hilbert modular forms of weight $(2,0)$ and $(0,2)$ respectively. By Proposition 7.2, the quotients of the restrictions to W_Π of these derivatives by y^2 (resp. $\widetilde{y}^2$) must be rational functions of t. We have to make the right choices of these functions in order not to introduce unwanted denominators. Since $1/U$ and V/U^2 restrict to polynomials in s (cf. eq. (72)), we choose these as the Hilbert modular functions to be differentiated, finding

$$y^2 \;=\; c\,\frac{(1-t)^4}{P(t)} \cdot D_1\Big(\frac{V}{U^2}\Big)\Big|_{W_\Pi}\,, \qquad \widetilde{y}^2 = \tilde{c}\,\frac{(1-t)^2}{t(1+t)} \cdot D_2\Big(\frac{1}{U}\Big)\Big|_{W_\Pi}\,,$$

where $c = \frac{29+7\sqrt{17}}{2}$, $\tilde{c} = \frac{7+\sqrt{17}}{2}$ and $P(t)$ is the polynomial

$$P(t) \;=\; t\,(1+t)\,\big(1 - \tfrac{31-7\sqrt{17}}{2}t\big)\,\big(1 - \tfrac{31+7\sqrt{17}}{64}t\big)\,\big(1 - \tfrac{647-153\sqrt{17}}{8}t + t^2\big)\,.$$

Since c and $\tilde{c}$ belong to R and the polynomials in the denominators are in $1 + tR[t]$, it follows that y^2 and $\widetilde{y}^2$, and hence also y and $\widetilde{y}$, are in $R[[t]]$. □

Although we have been working with $D = 17$ all the time and using concrete generators of the field of Hilbert modular functions, it is clear from the proof that the basic principle—the use of integral coefficients for Hilbert modular forms and the base change in two variables—can be applied for any D, giving the $\mathcal{O}_D$-integrality (up to finitely many primes in the denominator) of solutions $y(t)$ and $\widetilde{y}(t)$ of the corresponding Picard-Fuchs equations for any D.

Part III: Modular embeddings via derivatives of theta functions

In Part II we deduced the equation of the Teichmüller curve in the Hilbert modular surface for $D = 17$ and of a Hilbert modular form cutting out the Teichmüller curve starting from the differential operators L and $\widetilde{L}$, which in turn were deduced from the explicit algebraic model of the family as given in (38). In Part III we will show that there is a general construction of a Hilbert modular form of mixed weight that cuts out the Teichmüller curve. The construction, given in Section 9, uses derivatives of theta functions. The short proof depends on the description of Teichmüller curves using eigenforms for real multiplication with a double zero (see Theorem 5.2). We also verify that in the case $D = 17$ we get the same equation for the Teichmüller curve as the one already obtained in Section 8.

Next, in Sections 10 and 12, we develop the theory of Teichmüller curves in genus two "from scratch" starting from the new definition as vanishing loci of theta derivatives. In particular we give new proofs from this point of view of the cusp classification and of the facts that these curves are Kobayashi geodesics and are disjoint from the reducible locus and hence are Teichmüller curves. We do not know how to reprove the irreducibility from the viewpoint of theta functions.

Along the way, in Section 11 we show that Bainbridge's compactification of Hilbert modular surfaces using the moduli space of curves is indeed a toroidal compactification. Recall that this property of Hirzebruch's compactification was the model on which the notion of toroidal compactifications was developed.

9. Teichmüller curves are given by theta derivatives

Bainbridge has shown in [1] that the Teichmüller curves W_D defined in Section 5.3 is given as the vanishing locus of a modular form $(3, 9)$ for all D. We determine this form explicitly. It turns out to be a product of derivatives of theta series restricted from the Siegel half space to Hilbert modular varieties.

9.1. Theta functions and their restrictions to Hilbert modular varieties.

We recall the definition of the classical theta-functions and properties of their derivatives. Although we are ultimately interested in $g = 2$ only, we can keep g general without effort when setting up the definitions.

For $m, m' \in (\frac{1}{2}\mathbb{Z})^g$ (considered as row vectors) we define the (Siegel) theta function

$$\Theta_{(m,m')} : \left\{ \begin{array}{lll} \mathbb{C}^g \times \mathbb{H}_g & \to & \mathbb{C} \\ (v, Z) & \mapsto & \displaystyle\sum_{x \in \mathbb{Z}^g + m} \mathbf{e}\left(\tfrac{1}{2} x Z x^T + x(v + m')^T\right). \end{array} \right.$$

with characteristic (m, m'). The evaluation of a theta-function at $v = 0$ is called a *theta constant*. The theta-function (and the characteristic (m, m')) is called *odd* if $4m(m')^T$ is odd and *even* otherwise. Odd theta-constants vanish identically as functions of Z. Up to sign, $\Theta_{(m,m')}$ depends only on m and m' modulo $\mathbb{Z}^g$.

The theta constants $\Theta_{(m,m')}(0, Z)$ are modular forms of weight $\frac{1}{2}$ for some subgroup (in fact $\Gamma(4, 8)$, see e.g. [18]) of $\mathrm{Sp}(2g, \mathbb{Z})$. The partial derivatives with respect to any v_i are not modular, but if we restrict to $v = 0$ and consider the gradient (as column vector)

$$\nabla\Big(\Theta_{(m,m')}(0,Z)\Big) \;=\; \left(\frac{\partial}{\partial v_i}\Theta_{(m,m')}(v,Z)|_{v=0}\right)_{i=1,\dots,g},$$

then we get a vector-valued modular form. That is, if $\Theta_{(m,m')}(0,Z) = 0$, one calculates that for any $M = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in \Gamma(4,8)$ the gradient transforms as

$$\nabla\Big(\Theta_{(m,m')}\Big)(0, M\cdot Z) \;=\; \zeta_8\, \det(CZ+D)^{1/2}\,(CZ+D)\,\nabla\Big(\Theta_{(m,m')}\Big)(0,Z),$$

where ζ_8 is an 8-th root of unity depending on M.

If K is a totally real number field of degree g over $\mathbb{Q}$ with ring of integers $\mathcal{O}$, then just as in the special case $g=2$ we can define a g-dimensional Hilbert modular variety $X_K = \mathbb{H}^g/\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$ and a Siegel modular embedding (Ψ,ψ) of X_K into $\mathbb{H}_g/\mathrm{Sp}(2g,\mathbb{Z})$, given by a matrix $B \in \mathrm{GL}(g,\mathbb{R})$ as in (25). Recall that this means that $\Psi : \mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O}) \to \mathrm{Sp}(2g,\mathbb{Z})$ is a homomorphism and that $\psi(\mathbf{z}) = B^T \mathrm{diag}(z_1,\dots,z_g)B$ is a map that is equivariant with respect to Ψ. We then denote by $\theta_{(m,m')}(\mathbf{z}) = \Theta_{(m,m')}(0,\psi(\mathbf{z}))$ and $\nabla\theta_{(m,m')}(\mathbf{z}) = \nabla\Theta_{(m,m')}(0,\psi(\mathbf{z}))$ the restriction of the theta functions and their gradients to $\mathbb{H}^g$. We also write $\theta_{(m,m'),\Psi}$ if we want to emphasize the dependence on the modular embedding. The modularity of the Siegel theta functions imply that the (Hilbert) theta constants $\theta_{(m,m')}(\mathbf{z})$ are modular forms with a character of order 8 of weight $(\frac{1}{2},\dots,\frac{1}{2})$ for a subgroup of finite index of the Hilbert modular group $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$. The theta constants are non-zero if and only if (m,m') is even, while the theta gradients are modular if and only if (m,m') is odd.

The modular transformation of the derivative of theta constants for (m,m') odd now reads

$$\nabla\theta_{(m,m')}(\gamma\cdot\mathbf{z}) \;=\; \zeta_8\, \det(\hat{c}\psi(\mathbf{z})+\hat{d})^{1/2}\, B^{-1}J(\mathbf{z},\gamma)B\,\nabla\theta_{(m,m')}(\mathbf{z})$$

for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in the subgroup $\Psi^{-1}(\Gamma(4,8)) \subset \mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O})$, where $\hat{e}$ for $e \in K$ denotes the diagonal matrix with entries $\sigma_j(e)$ given by the different real embeddings σ_j of K and $J = J(\mathbf{z},\gamma)$ is the diagonal matrix with entries $\sigma_j(c)z_j + \sigma_j(d)$. Consequently, the vector-valued modular form $B\,\nabla\theta_{(m,m')}$ transforms with the automorphy factor $\zeta_8 \det(\hat{c}\psi(\mathbf{z})+\hat{d})^{1/2}J(\mathbf{z},\gamma)$, which is also a diagonal matrix. We will calculate the root of unity ζ_8 for $K = \mathbb{Q}(\sqrt{D})$ with $D \equiv 1 \pmod 8$ in detail below.

To summarize, the i-th entry $D_i\theta_{(m,m')}(\mathbf{z})$ of the column vector $B\,\nabla\theta_\psi$ is a Hilbert modular form of multi-weight $(\frac{1}{2},\dots,\frac{1}{2},\frac{3}{2},\frac{1}{2},\dots,\frac{1}{2})$ and a character stemming from the 8-th root of unity.

We can also express the functions $D_i\theta_{(m,m')}$ as derivatives in certain eigendirections. For this purpose replace the original coordinates $\mathbf{v} = (v_1,\dots,v_g) \in \mathbb{C}^g$ by the "eigendirection coordinates" $\mathbf{u} = B\mathbf{v} = (u_1,\dots,u_g)$. Then we may write

$$D_i\theta_{(m,m')}(\mathbf{z}) = \frac{\partial}{\partial u_i}\Theta_{(m,m')}(\mathbf{v},\psi(\mathbf{z}))|_{\mathbf{u}=0}.$$

9.2. **W_D is the vanishing locus of the theta-derivatives.** With this preparation we can now determine Bainbridge's modular form for general D.

Theorem 9.1. *The function*

$$\mathcal{D}\theta(\mathbf{z}) \;=\; \prod_{(m,m')\ \mathrm{odd}} D_2\theta_{(m,m')}(\mathbf{z})$$

is a modular form of weight $(3,9)$ *(with character) for the full Hilbert modular group* $\mathrm{SL}(\mathcal{O}_D^\vee \oplus \mathcal{O}_D)$. *Its vanishing locus is precisely the Teichmüller curve* W_D.

We will give the details if the prime (2) splits, i.e., if $D \equiv 1 \pmod 8$. In that case, as already mentioned in §8.2, the discussion of the numbering of theta characteristics given there for $D = 17$ always holds. In particular, one has a quartic character v_0 given by (68). Here, $\mathcal{D}\theta$ is a product of two modular forms and its character is v_0^2. Indeed, using the shorthand notation introduced in §8.2 we calculate the action of generators of $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$ on the functions $D_2\theta_{(m,m')}$. From Table 2 in §8.2, we deduce that the products of theta derivatives

$$D_L \;=\; D_2\theta_{0X}\cdot D_2\theta_{1X}\cdot D_2\theta_{\infty X} \quad\text{and}\quad D_R \;=\; D_2\theta_{X0}\cdot D_2\theta_{X1}\cdot D_2\theta_{X\infty}. \tag{76}$$

are modular forms for the full group $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$ of weight $(\frac{3}{2},\frac{9}{2})$ with a character of order 8 according to Table 3 below, in which $\alpha = \frac{1+\sqrt{D}}{2}$ as usual. The function $\mathcal{D}\theta$ then equals $D_L D_R$ and has character v_0^2 of order 2. This computation also shows

$f(z_1,z_2)$	$f(z_1+1,z_2+1)$	$f(z_1+\alpha,z_2+\sigma(\alpha))$	$z_1^{-3/2}z_2^{-9/2}f(1/z_1,1/z_2)$
D_L	$-D_L$	$-\zeta_8 D_L$	$-D_L$
D_R	$-D_R$	$\zeta_8^{-1} D_R$	$-D_R$

TABLE 3. The action of $\mathrm{SL}(2,\mathcal{O})$ on the $D_2\theta_L$ and $D_2\theta_R$

that if $D \equiv 1 \pmod 8$, then the Teichmüller curve W_D has two components, the vanishing loci of D_L and of D_R. This is our equivalent of McMullen's spin invariant that we referred to after Theorem 5.2.

Before proceeding to the proof of Theorem 9.1, we write out the Fourier expansions of the theta derivatives explicitly. Fix a basis (ω_1,ω_2) of $\mathcal{O}_D$ as we did in (25) for the construction of a Siegel modular embedding, say $(\omega_1,\omega_2) = (1,\gamma_0 = \frac{D+\sqrt{D}}{2})$. Then the first row the matrix B is the basis of $\mathcal{O}$ and the second row consists of the Galois conjugates. Given $m \in \frac{1}{2}\mathbb{Z}^2$, we let

$$\rho(x) \;=\; (x+m)\,\cdot\,(1,\gamma_0)^T$$

for $x \in \mathbb{Z}^2$. Then for (m,m') odd the theta derivative is given by

$$D_2\theta_{(m,m')}(z_1,z_2) \;=\; i^{4m(m')^T}\sum_{x\in\mathbb{Z}^2}(-1)^{2x\cdot(m')^T}\sigma(\rho(x))q_1^{\rho(x)^2/2}q_2^{\sigma(\rho(x))^2/2}\,, \tag{77}$$

where $q_j = \mathbf{e}(z_j)$ as usual.

Proof of Theorem 9.1. The modularity of $\mathcal{D}\theta$ for some level subgroup follows from the corresponding property of the Siegel theta functions, as shown in §9.1. It follows from the calculations above for $D \equiv 1 \pmod 8$ that the factor group permutes the set of theta characteristics, preserving their parity. Moreover we read off from the table that $\mathcal{D}\theta$ has character v_0^2 in this case. In other cases, $\mathcal{D}\theta$ is still a modular form, with some character, for the full Hilbert modular group.

For the second statement, let C be a curve of genus two with period matrix Z. It is cut out in its Jacobian as the vanishing locus of $\Theta_{(0,0)}(v,Z)$. The two-torsion points in $\mathrm{Jac}(C)$ are $Zm^T + (m')^T$ for $m,m' \in \left(\frac{1}{2}\mathbb{Z}\right)^2$. The Weierstrass points of C are precisely the 6 points $Zm^T + (m')^T$ for (m,m') odd. Moreover, a holomorphic differential form ω has a double zero on C if and only if ω vanishes (automatically doubly) at a Weierstrass point.

Suppose the point $[C] \in \mathcal{M}_2$ lies on W_D. Then C has real multiplication, so $Z = \psi(\mathbf{z})$ for some fixed Siegel modular embedding ψ. We identify the universal covering V of the Jacobian $J(C)$ with the dual to $H^0(C, \Omega^1_C)$. The characterizing condition on the existence of an eigenform with a double zero (see Theorem 5.2) is equivalent to the vanishing of the derivative of the Riemann theta-function in the second eigendirection u_2 at a Weierstrass point. Since we defined $\mathbf{u} = B\mathbf{v}$ this just means that

$$\sum_{j=1}^{2} \omega_j^\sigma \left(\frac{\partial}{\partial v_j} \Theta_{(0,0),D}(\mathbf{v}, \psi(\mathbf{z})) \right) |_{\mathbf{v}=\psi(\mathbf{z})m^T + (m')^T} \;=\; 0$$

for some even (m, m').

Differentiating the defining equation

$$\Theta_{(0,0)}(\mathbf{v} + Zm^T + (m')^T, Z)) \;=\; \mathbf{e}(-\pi i m Z m^T - 2\pi i m(\mathbf{v} + m')^T) \cdot \Theta_{(m,m')}(\mathbf{v}, Z)$$

for $Z = \psi(\mathbf{z})$ with respect to v_i, we see that any point $\mathbf{z} \in \mathbb{H}^2$ corresponding to $\mathrm{Jac}(C)$ lies in the vanishing locus of $\mathcal{D}\theta$.

To show that the modular form M_{W_D} vanishes nowhere else there are several options. The first is to remark that the above argument can be inverted for Jacobians of smooth curves. So one has just to show that the vanishing locus of $\mathcal{D}\theta$ is disjoint from the reducible locus. We give two proofs of this fact that do not rely on any Teichmüller theory in Section 12. Yet another way to conclude is to compare the weight of $\mathcal{D}\theta$ with the modular form that cuts out W_D in Bainbridge's theorem from [1]. They are both of weight $(3, 9)$. □

The example $D = 17$ revisited. If we calculate the theta series and their derivatives for $D = 17$ as was done in Section 8.2, then we can verify that the product of $D_R(z_1, z_2)$ with $D_R(z_2, z_1)$ is indeed proportional to the function F^1_{17} given in (73), and similarly that the product of $D_L(z_1, z_2)$ with $D_L(z_2, z_1)$ is proportional to the Galois conjugate function F^0_{17}.

10. Cusps and multiminimizers

Fix an invertible $\mathcal{O}_D$-ideal $\mathfrak{a}$. Our aim in this section is to list the branches of the theta-derivative vanishing locus W_D through the cusp of the Hilbert modular surface X_D determined by the class of $\mathfrak{a}$. Our Ansatz is to describe the branch of W_D defined by a Hilbert modular form in (q_1, q_2) by $q_1 = q^\alpha$ and $q_2 = q^{\sigma(\alpha)}(1+P)$, where q a suitable local parameter and P a power series with positive valuation in q. This will lead us to consider a minimization problem on quadratic forms. The solutions are given by so-called "multiminimizers."

The main result of this section will be the following characterization of cusps of W_D, derived from Fourier expansions only. We call an indefinite quadratic form $Q = [a, b, c]$ *standard*[4] if $a > 0 > c$ and $a + b + c < 0$. To a quadratic form Q we

[4] There seems to be no standard terminology for these quadratic forms and quadratic irrationalities. They appear implicitly in [10] and [1].

associate the quadratic irrationality $\lambda_Q = \frac{-b+\sqrt{D}}{2a}$. The quadratic form is standard if and only if $\lambda = \lambda_Q$ satisfies

$$\lambda > 1 > 0 > \lambda^\sigma . \tag{78}$$

A quadratic irrationality λ satisfying (78) will also be called *standard*.

Theorem 10.1. *Suppose that D is a fundamental discriminant or, more generally, that $\mathfrak{a}$ is an invertible $\mathcal{O}_D$-ideal. Then there is a bijection between the cusps of W_D mapping to the cusp $\mathfrak{a}$ of X_D and the set of pairs $(Q,\overline{r})$ consisting of a standard quadratic form $Q = [a,b,c]$ of discriminant D such that $[\langle 1, \lambda_Q\rangle] = [\mathfrak{a}]$ together with a residue class $\overline{r} \in \mathbb{Z}/(a,c)\mathbb{Z}$.*

This will be proved in §10.2. We will also comment on the case of non-invertible ideals after a discussion of cusps of Hilbert modular surfaces in the same subsection.

Given that W_D is a Teichmüller curve, as we showed in Theorem 9.1 using the eigenform definition and will show again in the next section using the theta viewpoint only, this characterization reproves the list of cusps of [21]. (See also [1].) For comparison, we will briefly sketch the approach based on flat surfaces in §10.3.

The proof of Theorem 10.1 can be applied verbatim to prove the following result on the reducible locus P_D (see §5.2) from the theta function viewpoint. This result was also proven by Bainbridge using the flat surface viewpoint on cusps.

Theorem 10.2. *Let D and $\mathfrak{a}$ be as above. Then there is a bijection between the cusps of P_D mapping to the cusp $\mathfrak{a}$ of X_D and the set of pairs $(Q,\overline{r})$ as in Theorem 10.1. In particular, for every given $\mathfrak{a}$ the numbers of cusps of P_D and W_D mapping to the cusp $\mathfrak{a}$ of X_D coincide.*

We will end the section by giving an algorithm in §10.4 to compute multiminimizers using continued fractions.

10.1. **Multiminimizers.** Let $F(x,y) = Ax^2 + Bxy + Cy^2$ be a positive definite binary quadratic form with real coefficients. On each of the cosets of $2\mathbb{Z}^2$ in $\mathbb{Z}^2$ this form assumes its minimum a finite number of times. Of course on the trivial coset the minimum is 0 and is assumed exactly once, while on each of the other cosets the minimum is generically attained exactly twice, by some non-zero vector and its negative. We call the form $[A,B,C]$ *multiminimizing* if on at least one of the three non-trivial cosets the minimum is attained more than twice. These forms are classified by the following proposition.

Proposition 10.3. *A positive definite binary quadratic form is multiminimizing if and only if it is diagonalizable over $\mathbb{Z}$. In this case, there is exactly one coset of $2\mathbb{Z}^2$ in $\mathbb{Z}^2$ on which the form has a multiple minimum; this minimum is assumed exactly twice (up to sign) and is the sum of the minima in the other two cosets.*

Proof. This is proved using reduction theory. Suppose that the form F is multiminimizing. Since this property is obviously $\mathrm{SL}(2,\mathbb{Z})$-invariant, we can assume that F is reduced, i.e. $F = [A,B,C]$ with $C \geq A \geq |B|$. By Cauchy's inequality we have $|Bxy| \leq A(x^2+y^2)/2$, and hence

$$F(x,y) \;\geq\; \frac{A}{2}\,x^2 \,+\, \Big(C - \frac{A}{2}\Big)\,y^2\,,$$

for all $(x,y)\in\mathbb{R}^2$. In particular,

$$\begin{array}{llll}
|x|\ge 1,\ |y|\ge 2 & \Rightarrow & F(x,y)\ \ge\ 4C\,-\,3A/2 & >\ A\,,\\
|x|\ge 2,\ |y|\ge 1 & \Rightarrow & F(x,y)\ \ge\ C\,+\,3A/2 & >\ C\,,\\
|x|\ge 3,\ |y|\ge 1 & \Rightarrow & F(x,y)\ \ge\ C\,+\,4A & >\ A+|B|+C\,,\\
|x|\ge 1,\ |y|\ge 3 & \Rightarrow & F(x,y)\ \ge\ 9C\,-\,4A & >\ A+|B|+C\,.
\end{array}$$

These equations show that the smallest value of F on the coset "(odd, even)" is attained only at $(\pm1,0)$ and equals A, that the smallest value of F on the coset "(even, odd)" is attained only at $(0,\pm1)$ and equals C, and that the two smallest values of F on the coset "(odd, odd)" are attained only at $(\pm1,\pm1)$ and are equal to $A-B+C$ and $A+B+C$. In particular, F is multi-minimizing if and only if B vanishes, in which case the only coset on which it attains its minimum more than once is "(odd, odd)" and the minimum there is attained exactly twice (up to sign) and is the sum of the unique minima in the other two cosets. $\square$

Let $\mathfrak{a}$ be a fractional $\mathcal{O}_D$-ideal in a real quadratic field K and ξ a non-zero coset of $\mathfrak{a}$ in $\frac{1}{2}\mathfrak{a}$. (Thus there are three possibilities for ξ given $\mathfrak{a}$.) We denote by $\widetilde{\mathrm{MM}}(\mathfrak{a},\xi)$ the set of non-zero $\alpha\in K$ such that the quadratic form $F_\alpha(x)=\mathrm{tr}(\alpha x^2)$ is positive definite and assumes its minimum value on the coset ξ more than twice (up to sign), and call the elements of this set *multiminimizers* for ξ. We denote by $\widetilde{\mathrm{MM}}(\mathfrak{a})$ the set of all *multiminimizers for* $\mathfrak{a}$, i.e. the union of the sets $\widetilde{\mathrm{MM}}(\mathfrak{a},\xi)$ for all three cosets ξ. Clearly, $\widetilde{\mathrm{MM}}(\mathfrak{a})$ is invariant under multiplication by positive rational numbers and by the squares of elements of the group U_D of units ε of $\mathcal{O}_D$. We set

$$\mathrm{MM}(\mathfrak{a})\ =\ \widetilde{\mathrm{MM}}(\mathfrak{a})/\big(\mathbb{Q}_+^\times\cdot U_D^2\big)\,. \tag{79}$$

Later we will often use the representatives of $\alpha\in\mathrm{MM}(\mathfrak{a})$ that are primitive in $(\mathfrak{a}^2)^\vee$. They are unique up to multiplication by U_D^2.

Proposition 10.4. *Let $\mathfrak{a}$ be a fixed fractional $\mathcal{O}_D$ ideal of K. Then there is a bijection between $\mathrm{MM}(\mathfrak{a})$ and the set of standard quadratic forms $Q=[a,b,c]$ with $b^2-4ac=D$ in the wide ideal class of $\mathfrak{a}$.*

Moreover, given $\alpha\in\widetilde{\mathrm{MM}}(\mathfrak{a})$, there is a unique basis (ω_1,ω_2) such that $\omega_2>\omega_1>0$ of $\mathfrak{a}$ with respect to which the form F_α is diagonal, and the coset ξ is then $\frac{1}{2}(\omega_1+\omega_2)+\mathfrak{a}$.

Proof. Given $Q=[a,b,c]$ choose $\mu\in K^\times$ positive such that $\mu\langle1,\lambda_Q\rangle=\mathfrak{a}$ and take

$$\alpha\ =\ \frac{-ac}{\mu^2\lambda_Q\sqrt{D}}\ =\ \frac{a}{\mu^2}\cdot\frac{b+\sqrt{D}}{2\sqrt{D}}\,. \tag{80}$$

This is positive definite if Q is standard. Since Q is simple the basis $\omega_1=\mu$, $\omega_2=\mu\lambda_Q$ satisfies the conditions stated. Moreover, in this basis the quadratic form F_α is

$$F_\alpha=[A,0,C],\qquad A=\mathrm{tr}(\alpha\omega_1^2)=a,\quad C=\mathrm{tr}(\alpha\omega_2^2)=-c$$

since $\mathrm{tr}(\alpha\omega_1\omega_2)=\mathrm{tr}(-ac/\sqrt{D})=0$. Thus α is a multiminimizer.

Conversely, if α is a representative of a class in $\mathrm{MM}(\mathfrak{a})$ then it follows from Proposition 10.3 that there exists a unique basis (ω_1,ω_2) of $\mathfrak{a}$ (up to interchanging the ω_i and changing their signs) of $\mathfrak{a}$ with respect to which $F_a(x)$ is diagonal, and the coset ξ is then $\frac{1}{2}(\omega_1+\omega_2)+\mathfrak{a}$. This also proves the last statement. In this basis

$F_\alpha = [A, 0, C]$ with $A = \mathrm{tr}(\alpha\omega_1^2)$ and $C = \mathrm{tr}(\alpha\omega_2^2)$. We choose signs and order the basis such that $\lambda = \omega_2/\omega_1 > 1$ and claim that then $\sigma(\lambda) < 0$, so that λ is the root of a standard form. In fact, we have $\alpha = q/(\omega_1\omega_2\sqrt{D})$ for some $q \in \mathbb{Q}_+^\times$. Thus $C = \mathrm{tr}(q\lambda/\sqrt{D})$ and

$$A = \mathrm{tr}(q/(\lambda\sqrt{D})) = -N(\lambda)C.$$

Since F_α is positive definite (by definition of a multiminimizer), the numbers A and C are positive, so this implies that $N(\lambda) < 0$ as claimed. One checks immediately that λ does not depend on the representative of the multiminimizer in $\mathrm{MM}(\mathfrak{a}, \xi)$ we have chosen. We take $Q = [a, b, c]$ so that λ satisfies $a\lambda^2 + b\lambda + c = 0$ with $a > 0$ and a, b, c coprime integers, and since $\langle 1, \lambda\rangle$ is an invertible $\mathcal{O}_D$ module, we then have $b^2 - 4ac = D$.

Obviously, the composition $\lambda_Q \mapsto \alpha(\lambda_Q) \mapsto \lambda(\alpha)$ is the identity. In the other direction, note that α is determined by (ω_1, ω_2) up to a positive rational number and that μ with $\mu\langle 1, \lambda_Q\rangle = \mathfrak{a}$ is determined up to a unit. Consequently, each of the distinguished basis elements ω_i is determined up to a unit and α is determined up to a square of this unit. Since multiminimizers were defined in (79) by these two equivalence relations, this shows the bijection we claimed. □

10.2. **Cusps of Hilbert modular surfaces.** Classically, cusps of the Hilbert modular surface X_D are defined to be the equivalence classes of points in $\mathbb{P}^1(K)$ under the action of $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$. Equivalently, we may define a cusp as an exact sequence

$$0 \to \mathfrak{a}^\vee \to \mathcal{O}_D^\vee \oplus \mathcal{O}_D \to \mathfrak{a} \to 0$$

of torsion-free $\mathcal{O}$-modules up to the action of $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$ on $\mathcal{O}^\vee \oplus \mathcal{O}$ and its sub $\mathcal{O}$-modules. The modules $\mathfrak{a}$ arising in this way are *quasi-invertible*, i.e. invertible $\mathcal{O}_E$-module for some order $\mathcal{O}_E \supseteq \mathcal{O}_D = \mathcal{O}$. Yet another equivalent viewpoint to define a cusp is by the class of an invertible $\mathcal{O}_E$-ideal $\mathfrak{a}$ together with an element $r \in \mathbb{Z}/\sqrt{\frac{D}{E}}\mathbb{Z}$.

We briefly recall how to see the equivalence of these definitions. For the equivalence of the first two definitions, intersect the line $L \subset K^2$ determined by a point in $\mathbb{P}^1(K)$ with a fixed embedding of $\mathcal{O}^\vee \oplus \mathcal{O}$ in K^2 to get the exact sequence and conversely tensor the exact sequence with K over $\mathcal{O}$. That such an extension class is determined by $r \in \mathbb{Z}/\sqrt{\frac{D}{E}}\mathbb{Z}$ can be deduced from the calculation in [1], Proposition 7.20 or [2], Theorem 2.1.

As a preparation for the proof of Theorem 10.1 we determine the Fourier series of $\mathcal{D}\theta$ at a given cusp $\mathfrak{a}$ of X_D. For a basis $\boldsymbol{\omega} = (\omega_1, \omega_2)$ of the ideal $\mathfrak{a}$ let $\rho_{\boldsymbol{\omega}}(x) = (x + m) \cdot \boldsymbol{\omega}^T$ for $x \in \mathbb{Z}^2$, with the dependence on m suppressed in the notation. Then for $(m, m') \in (\frac{1}{2}\mathbb{Z})^2$ odd we define

$$D_2\theta_{(m,m'),\boldsymbol{\omega}}(z_1, z_2) \;=\; \sum_{x \in \mathbb{Z}^2} (-1)^{2x\cdot(m')^T} \sigma(\rho_{\boldsymbol{\omega}}(x)) q_1^{\rho_{\boldsymbol{\omega}}(x)^2/2} q_2^{\sigma(\rho_{\boldsymbol{\omega}}(x))^2/2}, \qquad (81)$$

where $q_i = \mathbf{e}(z_i)$. Note that $D_2\theta_{(m,m'),\boldsymbol{\omega}}$ depends on the chosen basis $\boldsymbol{\omega}$, but a base change can be compensated for by letting the base change matrix also act an the characteristic (m, m'). Consequently, the product

$$\mathcal{D}\theta_{\mathfrak{a}} = \prod_{(m,m')\ \mathrm{odd}} D_2\theta_{(m,m'),\boldsymbol{\omega}} \qquad (82)$$

is of the form $\sum_{y\in\mathfrak{a}} c(y)\exp(\mathrm{tr}(y^2z))$ and thus invariant under an upper triangular matrix in $\mathrm{SL}(\mathfrak{a}^\vee\oplus\mathfrak{a})$.

Lemma 10.5. *The Fourier expansion of $\mathcal{D}\theta$ at the cusp $\mathfrak{a}$ is proportional to $\mathcal{D}\theta_\mathfrak{a}$.*

Proof. To avoid the generally hard problem of finding the Fourier development of a modular form at a different cusp we use the fact that the vanishing locus of $\mathcal{D}\theta$ has an intrinsic formulation in terms of eigenforms. In Theorem 9.1 we proved that this vanishing locus corresponds to the set of principally polarized abelian varieties with real multiplication such that the first eigenform has a double zero. This proof works for any Siegel modular embedding, for example the one given at the end of §5.1, where the locus of real multiplication is $X_{D,\mathfrak{a}}$ and ψ is constructed with the help of the matrix B as in (25) having $\boldsymbol{\omega}=(\omega_1,\omega_2)$ as its first column. The restriction of the Siegel theta function with characteristic (m,m') via this modular embedding is just $D_2\theta_{(m,m'),\boldsymbol{\omega}}$. To complete the proof, we note that the cusp at ∞ of $X_{D,\mathfrak{a}}$ is just the cusp $\mathfrak{a}$ of X_D. To see this, take a matrix $\left(\begin{smallmatrix}\mathfrak{a}^{-1} & \mathfrak{a}^\vee\\ (\mathfrak{a}^\vee)^{-1} & \mathfrak{a}\end{smallmatrix}\right)\cap\mathrm{SL}(2,K)$. It conjugates $\mathrm{SL}(\mathcal{O}^\vee\oplus\mathcal{O})$ into $\mathrm{SL}(\mathfrak{a}^\vee\oplus\mathfrak{a})$, since $\mathfrak{a}^\vee=\mathfrak{a}^{-1}\mathcal{O}^\vee$ and the line at infinity in $\mathbb{P}^1_K$ intersects $\mathfrak{a}^\vee\oplus\mathfrak{a}$ in the submodule $\mathfrak{a}^\vee$, as required. □

Proof of Theorem 10.1. We first determine the cusps of the vanishing locus of $\mathcal{D}\theta_\mathfrak{a}$, as defined in (82), that map to the cusp ∞ of the Hilbert modular surface $X_{D,\mathfrak{a}}$. Because functions in a neighborhood of the cusp $\mathfrak{a}$ have the form $\sum_\nu c_\nu\mathbf{e}(\mathrm{tr}(\nu z))$ with $\nu\in(\mathfrak{a}^\vee\mathfrak{a}^{-1})^\vee=\mathfrak{a}^2$, we can choose a local parameter q at a branch of this locus, with $q=0$ at the cusp, of the form $q=\mathbf{e}(z/\alpha)$ with $\alpha\in(\mathfrak{a}^2)^\vee$ primitive. A lift of the vanishing locus to $\mathbb{H}\times\mathbb{H}$ looks locally like $z_2=\varphi(z_1)$, where $\varphi(z)=\frac{\alpha^\sigma}{\alpha}z+C+\frac{\varepsilon(q)}{2\pi i}$ as $\Im(z)\to\infty$ for some $C\in\mathbb{C}$ and some power series ε in q with no constant term. Then we have

$$\mathbf{e}(\nu z_1+\nu^\sigma z_2)|_{\mathrm{locus}} \;=\; \mathbf{e}(C\nu^\sigma)\,q^{\mathrm{tr}(\alpha\nu)}e^{\sigma(\nu)\varepsilon(q)}\qquad\text{for all}\quad \nu\in\mathfrak{a}^2\,,\tag{83}$$

so that the restriction of any Hilbert modular form becomes a power series in q. Making a different choice of the lifting would change C by an integral multiple of $N(\mathfrak{a})^2\sqrt{D}/\alpha$, so that the quantity

$$S \;=\; \mathbf{e}\Big(\frac{C\alpha}{N(\mathfrak{a})^2\sqrt{D}}\Big)\;\in\;\mathbb{C}^*\tag{84}$$

is independent of the choice of the lifting. We will show below that S is in fact a rational power of an element of K^*.

The resulting q-exponents after plugging (83) into (81) are of the form $\mathrm{tr}(\alpha\rho_\omega(x)^2)$. In order for the theta derivative to vanish, the smallest exponent of q must occur twice, so this quadratic form has to take its minimum twice (with x and $-x$ not distinguished). Hence α is a multiminimizer for $\mathfrak{a}$. Recall that this specifies α only up to $\mathbb{Q}_+^\times\cdot U_D^2$, but here the fact that $q=\mathbf{e}(z/\alpha)$ is a local parameter (or equivalently, that α is primitive) eliminates the $\mathbb{Q}_+^\times$-ambiguity. The U_D^2-ambiguity corresponds to the fact that this group (considered as diagonal matrices in the Hilbert modular group) stabilizes the given cusp α of the Hilbert modular surface. Since U_D^2 acts transitively on the three non-trivial cosets ζ of $\frac12\mathfrak{a}/\mathfrak{a}$ and since a multiminimizer has the multiple minimum property on exactly one of the three cosets by Proposition 10.3, we may suppose from now on that the branch is chosen such that $D_2\theta_{(m,m'),\boldsymbol{\omega}}$ for $m=(\frac12,\frac12)$ vanishes on that branch.

We next determine α and exhibit for this purpose a convenient basis of $\mathfrak{a}^2$. We suppose that ω_1, ω_2 was from the beginning of the discussion the distinguished basis of $\mathfrak{a}$ associated with a multiminimizer in Proposition 10.4. We let $g = \gcd(a, c)$, where $ax^2 + bx + c = 0$ is the minimal polynomial of $\lambda = \omega_2/\omega_1$. Choose $s, t \in \mathbb{Z}$ such that $sa + tc = g$. It is easily verified that

$$\mathfrak{a}^2 \;=\; \langle \omega_1^2, \omega_1\omega_2, \omega_2^2 \rangle \;=\; \langle \alpha^*, \beta^* \rangle,$$

where

$$\alpha^* \;=\; \frac{g}{a}\omega_1^2 + \frac{tb}{a}\omega_1\omega_2 \quad \text{and} \quad \beta^* \;=\; \frac{1}{g}\omega_1\omega_2$$

since $\gcd(a, b, c) = 1$. (In fact, $\mathfrak{a}^2$ contains $\frac{c}{a}\omega_1^2 + \frac{b}{a}\omega_1\omega_2$ using the minimal polynomial, so $\frac{b}{g}\omega_1\omega_2$ and finally $\frac{1}{g}\omega_1\omega_2$ by the gcd condition. The ideal also contains $\frac{ct}{a}\omega_1^2 + \frac{tb}{a}\omega_1\omega_2$ and, since $\gcd(a, tc) = g$, also α^*. The converse inclusion follows from the line below (87).) This basis is chosen such that the dual basis is $\{\alpha, \beta\}$ where

$$\alpha \;=\; \frac{-ac}{g\omega_1^2\lambda\sqrt{D}} \;=\; \frac{a}{g\omega_1^2} \cdot \frac{b + \sqrt{D}}{2\sqrt{D}}\,. \tag{85}$$

As local coordinates on the Hilbert modular surface we now pick $X = q_1^{\alpha^*} q_2^{\sigma(\alpha^*)}$ and $Y = q_1^{\beta^*} q_2^{\sigma(\beta^*)}$, which are power series in the local coordinate q beginning with q and S, respectively. In these coordinates the factors of $\mathcal{D}\theta_{\mathfrak{a}}$ have the expansion

$$D_2\theta_{(m,m'),\boldsymbol{\omega}}(z_1, z_2) \;=\; \sum_{x \in \mathbb{Z}^2} (-1)^{2x\cdot(m')^T} \rho_{\boldsymbol{\omega}}(x)^{\sigma}\, X^{F(\widetilde{x}_1,\widetilde{x}_2)/2} Y^{G(\widetilde{x}_1,\widetilde{x}_2)/2}\,, \tag{86}$$

where $\widetilde{x}_i = x_i + \frac{1}{2}$ and where

$$F \;=\; \Big[\frac{a}{g}, 0, \frac{-c}{g}\Big], \quad G \;=\; \Big[-bt, 2g, bs\Big]\,. \tag{87}$$

(This follows from $\frac{a}{g}\alpha^* - bt\beta^* = \omega_1^2$, $2g\beta^* = 2\omega_1\omega_2$ and $-\frac{c}{g}\alpha^* - bs\beta^* = \omega_2^2$.) Then $D_2\theta_{(m,m'),\boldsymbol{\omega}}/X^{1/8}Y^{1/8}$ has an expansion in integral powers of X and $Y^{1/2}$, but we actually need the product $\mathcal{F} = D_2\theta_{((\frac{1}{2},\frac{1}{2}),(\frac{1}{2},0)),\boldsymbol{\omega}} D_2\theta_{((\frac{1}{2},\frac{1}{2}),(0,\frac{1}{2})),\boldsymbol{\omega}}$, and this product, divided by $X^{1/4}Y^{1/4}$, has integral powers of Y as well.

Finally we need to show that map from cusps to multiminimizers is onto and that the fibers have cardinality g. Analyzing the lowest order coefficient in q of $\mathcal{F}$, to which precisely the summands $\widetilde{x}_1 = \pm\frac{1}{2}$ and $\widetilde{x}_2 = \pm\frac{1}{2}$ contribute, and noting that $G(\frac{1}{2}, \frac{1}{2}) - G(\frac{1}{2}, -\frac{1}{2}) = g$, we find that the S of (84) has to be a solution of

$$S^g \;=\; \left(\frac{\rho_{\boldsymbol{\omega}}(\frac{1}{2}, -\frac{1}{2})}{\rho_{\boldsymbol{\omega}}(\frac{1}{2}, \frac{1}{2})}\right)^{\sigma}. \tag{88}$$

This equation has precisely g solutions, differing by gth roots of unity. For each such solution of the lowest order term in q there is a unique power series $\varepsilon(q)$ such that (83) is in the vanishing locus of $\mathcal{D}\theta_{\mathfrak{a}}$, since the coefficients of $\varepsilon(q)$ are recursively determined by a triangular system of equations. (We will discuss the arithmetic properties of $\varepsilon(q)$ in §13.1.)

This completes the proof for the vanishing locus of $\mathcal{D}\theta_{\mathfrak{a}}$ at the cusp ∞ and by Lemma 10.5 also the proof of Theorem 10.1. $\square$

A statement like Theorem 10.1 can certainly be proven along the same lines also if $\mathfrak{a}$ is not an invertible $\mathcal{O}_D$-ideal (which can of course only happen for non-fundamental D). This has an effect in Lemma 10.5, and the fact that $\gcd(a,b,c) > 1$ changes the computation of the basis $\{\alpha^*, \beta^*\}$ that was used in the proof. Since our aim is just to demonstrate the method of reproving the properties of genus two Teichmüller curves using theta functions, we do not carry this out in detail.

Proof of Theorem 10.2. The reducible locus is the vanishing locus of the product of all 10 even theta functions. Thus branches of the vanishing locus have to be parametrized as in (83) with α a multiminimizer. For $m = (0,0)$ the forms are never multiminimizing and in all other cases the proof proceeds as the proof of Theorem 10.1. Only the coefficients of the equation (88) change, but not the exponents. This does not affect the number of cusps for each multiminimizer. $\square$

10.3. **Cusps of W_D via flat surfaces and prototypes.** The cusps of the Teichmüller curve W_D were first determined in [21] based on the following observations. It was discovered by Veech along with the definition of Teichmüller curves that cusps correspond to directions (considered as elements of $\mathbb{R}^2/\mathbb{R}^*$) of *saddle connections* (i.e. geodesics for the flat metric $|\omega|$ starting and ending at a zero of ω). Applying a rotation to the flat surface, we may suppose that the direction is horizontal, so that the subgroup of $\mathrm{GL}(2,\mathbb{R})$ stabilizing the direction consists of upper triangular matrices. McMullen discovered that flat surfaces parametrized by W_D always decompose in saddle connection directions into two cylinders as indicated in Figure 2. The action of the upper triangular group allows us to assume the upper cylinder to be a square, while still having the freedom to normalized by the action of $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. McMullen further deduces from real multiplication that the two cylinders must be isogenous with a homothety in a real quadratic field. We thus may suppose that (a, c, q) are integral and λ is real quadratic. A more careful analysis of the isogeny (see [21], Equation (2.1), or [1], Proposition 7.20), imposing moreover that $\mathcal{O}_D$ is the exact endomorphism ring of a generic abelian surface parametrized by the Teichmüller curve, implies that $\lambda = [a, b, c]$ with a and c as in Figure 2 and such that $(a, b, c, r) = 1$.

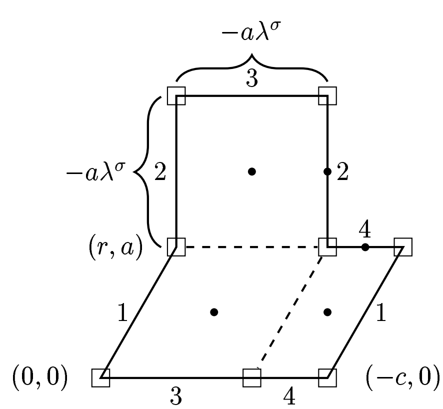

FIGURE 2. Prototype of a flat surface parametrizing cusps of the Teichmüller curves W_D. Sides with the same label are identified.

Obviously $a > 0$, $c < 0$ and the side length inequality $-a\lambda^\sigma < -c$ is equivalent to $a + b + c < 0$. By the action of T we may reduce mod $\gcd(a, c)$. Altogether this implies that cusps of W_D correspond bijectively to pairs $(\lambda = [a, b, c], q)$ with λ standard and $q \in \mathbb{Z}/(\mathrm{a}, \mathrm{c})$, as in Theorem 10.1.

10.4. **Computing multiminimizers.** There are many variants of how to expand real numbers into continued fractions. One may subtract (slowly) one at each step or group these steps together, one may subtract one while being greater than one, greater than zero, or until becoming negative. We will need two of them, one here and one in §11.2.

The *slow plus greater-than-one continued fraction expansion* of $x \in \mathbb{R}_{>1}$ is defined by $x_0 = x$ and then inductively by

$$x_{n+1} = \begin{cases} x_n - 1 & \text{if} \quad x_n > 2 \\ 1/(x_n - 1) & \text{if} \quad 2 > x_n > 1 \end{cases},$$

so that

$$x = 1 + \cdots + 1 + \left(1 + \cfrac{1}{1 + \cdots + 1 + \left(1 + \frac{1}{\ddots}\right)}\right).$$

The class of the ideal $[\mathfrak{a}] = [\langle 1, x_n \rangle]$ is unchanged under this continued fraction algorithm, since $\langle 1, x_n \rangle = \langle 1, x_n - 1 \rangle = (x_n - 1)\langle 1, \frac{1}{x_n - 1} \rangle$.

Lemma 10.6. *The slow plus greater-than-one continued fraction expansion of any quadratic irrational λ is periodic, and is pure periodic if and only if λ is standard.*

Proof. The classical proof of Lagrange (for the more usual fast continued fraction algorithm with plus sign, where λ is sent to $\lambda - \lfloor \lambda \rfloor$ if $\lambda > 1$ and to $1/\lambda$ if $0 < \lambda < 1$) applies here as well. The slow algorithm here just introduces intermediate steps and combines two steps when $1 < \lambda < 2$. The second statement is proven in [10]. $\square$

This observation gives us an algorithm to compute all multiminimizers for a given ideal $\mathfrak{a}$. Write $\mathfrak{a} = \mu_0 \langle \lambda_0, 1 \rangle$ and apply the continued fraction algorithm to λ_0 until it becomes pure periodic, say at a standard quadratic irrational λ_1 with $\mathfrak{a} = \mu_1 \langle \lambda_1, 1 \rangle$. The first multiminimizer is then the $U_D^2 \cdot \mathbb{Q}^*$-class of $\alpha_1 = \frac{-1}{\mu_1^2 \lambda_1 \sqrt{D}}$ as in (80). Proceeding in this way with $\alpha_k = \frac{-1}{\mu_k^2 \lambda_k \sqrt{D}}$ where $\mathfrak{a} = \mu_k \langle \lambda_k, 1 \rangle$ for $\lambda_2, \ldots, \lambda_n$ along the period of λ_1 gives all the multiminimizers for $\mathfrak{a}$.

In each of the steps we may choose α in its $\mathbb{Q}^*$-class to be primitive in some fixed $\mathcal{O}_D$-module. We will use this normalization for the Bainbridge compactification in the next section.

11. The Hirzebruch and the Bainbridge compactification

Hirzebruch's minimal smooth compactification is a toroidal compactification. In this section we reinterpret the Bainbridge compactification of Hilbert modular surfaces as a toroidal compactification, see Theorem 11.5. This compactification was defined originally via the Deligne-Mumford compactification of the moduli space of one-forms on Riemann surfaces, We will see that both Hirzebruch's and Bainbridge's compactification can be computed using continued fraction algorithms.

We start with a review of toroidal compactifications from the viewpoint of curve degenerations and we recall Hirzebruch's continued fraction algorithm. The procedure presented in §10.4 to compute multiminimizers is very similar to Hirzebruch's, but the continued fraction algorithms are different. The geometry of Bainbridge's compactification has been described in [1]. We recall this material in §11.4, since it is part of the main Theorem 11.5. Once this is proven, all geometric properties will follow from general facts about toroidal compactifications. (See Proposition 11.2.) The proof of the main theorem will be given in §11.5.

The description of toroidal compactifications will make it easy to determine (in §11.6) where the cusps of modular curves F_N or Teichmüller curves W_D intersect the cusp resolution cycle both of the Hirzebruch compactification and the Bainbridge compactification. Finally, in §11.8 we compare the two continued fraction algorithms governing the two compactifications and give formulas for the total lengths or the cusp resolution cycles in both cases.

11.1. **Toroidal compactifications.** Locally near $(\infty,\infty) \in \mathbb{H}^2$ a Hilbert modular surfaces is $\mathbb{H}^2/G(M,V)$, where M is a complete submodule of K (i.e., an additive subgroup of K that is free abelian of rank two) and $G(M,V) \subset \mathrm{SL}(\mathcal{O}_D^\vee \oplus \mathcal{O}_D)$ is the semidirect product of M and some subgroup V of totally positive units (see e.g. [35]).

We study the limiting behavior of a complex curve C in $\mathbb{H}^2$ parametrized by $\tau \in \mathbb{H}$ and having the asymptotic form

$$\begin{aligned} z_1 &= \gamma\tau + A_0 + A_1 q + A_2 q^2 + \cdots \\ z_2 &= \gamma^\sigma\tau + B_0 + B_1 q + B_2 q^2 + \cdots \end{aligned} \tag{89}$$

for $\Im(\tau) \to \infty$, where $\gamma \in K$ is totally positive, $q = \mathbf{e}(\tau)$ and $A_i, B_i \in \mathbb{C}$. If this curve descends to an algebraic curve in the Hilbert modular surface X_D, so that the intersection with the subgroup M of the cusp stabilizer is not trivial, we can always assume that the leading coefficients are of that form.

To each totally positive $\alpha \in M$ we associate a copy $\mathbb{P}^1_\alpha$ of $\mathbb{P}^1(\mathbb{C})$. We have a map

$$X_\alpha : \left\{ \begin{array}{ccl} \mathbb{H}^2/M & \to & \mathbb{C}^*_\alpha \subset \mathbb{P}^1_\alpha \\ (z_1,z_2) & \mapsto & \mathbf{e}\left(\mathrm{tr}(\alpha^\vee \underline{z})\right) = \mathbf{e}\left(\frac{\alpha^\sigma z_1 - \alpha z_2}{N(M)\sqrt{D}}\right) \end{array}\right. ,$$

where $N(M)$ denotes the norm of M. Using this identification, we glue $\mathbb{P}^1_\alpha$ at the cusp (∞,∞) of $\mathbb{H}^2$ topologically as follows. A sequence of points $\underline{z} = (z_1,z_2) \in \mathbb{H}^2$ with imaginary parts y_1, y_2 both tending to ∞ with limiting value of y_1/y_2 being equal to $t \in \mathbb{R}_+$ converges to the point $0 \in \mathbb{P}^1_\alpha$ if $t < \alpha/\alpha^\sigma$, to the point $\infty \in \mathbb{P}^1_\alpha$ if $t > \alpha/\alpha^\sigma$ and to a finite point $X = \mathbf{e}(\theta/N(M)\sqrt{D})$ for some $\theta \in \mathbb{C}$ if $t = \alpha/\alpha^\sigma$ and if $\alpha^\sigma z_1 - \alpha z_2 = \theta + o(1)$.

Consequently, the curve C meets $\mathbb{P}^1_\alpha$ at ∞ if $\mathrm{tr}(\alpha^\vee\gamma) < 0$, at 0 if $\mathrm{tr}(\alpha^\vee\gamma) > 0$, and at a finite, non-zero, point if γ is a rational multiple of α.

To any oriented $\mathbb{Q}$-basis (α,β), i.e. with $\frac{\alpha}{\alpha^\sigma} < \frac{\beta}{\beta^\sigma}$, corresponds a pair of projective lines $\mathbb{P}^1_\alpha$ and $\mathbb{P}^1_\beta$ meeting at one point $(\infty, 0) \in \mathbb{P}^1_\alpha \times \mathbb{P}^1_\beta$. Suppose that $\gamma \in K$ has the property

$$\frac{\alpha}{\alpha^\sigma} < \frac{\gamma}{\gamma^\sigma} < \frac{\beta}{\beta^\sigma}$$

holds, so that the curve C passes through the point $(\infty, 0)$. If we write $\gamma = p\alpha + q\beta$ with $p, q \in \mathbb{Q}_+$ then, near the point $(\infty, 0)$ the curve (89) looks like

$$(X_\alpha^{-1})^p = (X_\beta)^q.$$

We will, of course, apply this in particular to the curves F_N and to Teichmüller curves.

A V-invariant partial compactification $\overline{\mathbb{H}^2/M}$ of a Hilbert modular surface is defined by adding not just one $\mathbb{P}^1_\alpha$, but an appropriate sequence of $\mathbb{P}^1_\alpha$'s.

Definition 11.1. *A sequence of numbers $(\alpha_n)_{n\in\mathbb{N}}$ in M forms a* fan *if i) the α_n are all totally positive, ii) $V \cdot \{\alpha_n\}_{n\in\mathbb{N}} = \{\alpha_n\}_{n\in\mathbb{N}}$ and iii) the ratios α_n^σ/α_n are a strictly decreasing sequence.*

Our definition includes that the fan is V-invariant. If $V = \langle\varepsilon\rangle$, then ii) is equivalent to the existence of some k such that $\varepsilon\alpha_n = \alpha_{n+k}$. The minimal positive k with that property will be called the *length* of the fan.

Proposition 11.2. *A fan $(\alpha_n)_{n\in\mathbb{N}}$ determines a partial compactification $\overline{\mathbb{H}^2/M}$ of $\mathbb{H}^2/M$ with the following properties. For each n there is an irreducible curve $\mathbb{P}^1_{\alpha_n}$ in $\overline{\mathbb{H}^2/M} \smallsetminus \mathbb{H}^2/M$. The curves $\mathbb{P}^1_{\alpha_n}$ and $\mathbb{P}^1_{\alpha_{n+1}}$ intersect in one point. For $k > 1$ the curves $\mathbb{P}^1_{\alpha_n}$ and $\mathbb{P}^1_{\alpha_{n+k}}$ are disjoint in $\overline{\mathbb{H}^2/M}$.*

The action of V on $\mathbb{H}^2/M$ extends to an action on $\overline{\mathbb{H}^2/M}$. Hence a fan determines a partial compactification $\overline{\mathbb{H}^2/G(M,V)}$ of $\mathbb{H}^2/G(M,V)$.

The partial compactification is always smooth along $\mathbb{P}^1_{\alpha_n}$ minus the intersection points with the $\mathbb{P}^1_{\alpha_{n\pm1}}$. At the intersection point of $\mathbb{P}^1_{\alpha_n}$ and $\mathbb{P}^1_{\alpha_{n+1}}$ the compactification is smooth if and only $M = \mathbb{Z}\alpha_n + \mathbb{Z}\alpha_{n+1}$. More generally, if $\mathbb{Z}\alpha_n + \mathbb{Z}\alpha_{n+1}$ has index k in M, then $\overline{\mathbb{H}^2/M}$ has a cyclic quotient singularity of order k at that point. In particular, the compactification is normal.

Proof. We first recall the general setup of two-dimensional toric varieties. For each two-dimensional cone σ given, say, as the span of α_n and α_{n+1} we let U_σ be the variety with coordinate ring $\mathbb{C}[\sigma^\vee \cap M^\vee]$. For each one-dimensional cone τ at the boundary of σ_2, say spanned by α_n we let $V_\tau \subset U_\sigma$ be the variety with coordinate ring $\mathbb{C}[\sigma^\vee \cap M^\vee]$. If τ is contained both in σ_1 and σ_2 we may glue U_{σ_1} and U_{σ_2} along the open set V_τ. In particular a_n (or $\tau = \langle a_n\rangle$) determines a rational curve $\mathbb{P}^1_{\alpha_n} = (U_{\sigma_1} \cup U_{\sigma_2}) \smallsetminus V_\tau$. The zero-dimensional cone 0 corresponds to the variety W with coordinate ring $\mathbb{C}[M^\vee]$ and sits as open part in all the V_τ and U_σ. Here $M^\vee$ is the dual $\mathbb{Z}$-module, which we will identify from now on as a submodule of K using the trace pairing.

In our situation we want to identify W with $\mathbb{C}^2/M$. On the level of local coordinate rings this is done by assigning to $b \in M^\vee$ the coordinate function

$$X_b^\vee(z_1, z_2) = \mathbf{e}(bz_1 + b^\sigma z_2).$$

(This is the same coordinate as $X_{\alpha^\vee}$ associated above with $\alpha \in M$, using the identification of M with $M^\vee$ via $\alpha \mapsto \alpha^\vee = \frac{\alpha^\sigma}{N(M)\sqrt{D}}$. We refer to $\alpha^\vee$ as the *trace* *dual* of α.) Since $\mathbb{H}^2/M$ sits inside $\mathbb{C}^2/M$, the partial compactification of W by the U_σ defines the desired partial compactification of $\mathbb{H}^2/M$.

Given a cone τ generated by α_n, the element $\alpha_\vee$ is the unique (up to sign) primitive element of $M^\vee$ with $\mathrm{tr}(\alpha_n\alpha_n^\vee) = 0$. If we complete $\alpha_n^\vee$ to a basis of $M^\vee$ using β_n, whose sign we may choose such that $\mathrm{tr}(\alpha_n^\vee\beta_n) > 0$, then the curve $\{X_{\beta_n}^\vee = 0\}$ is independent of the choice of β_n and coincides with the curve $\mathbb{P}^1_{\alpha_n}$ defined in the text preceding the proposition.

Given two consecutive elements α_n and α_{n+1} of the fan, the monotonicity of the ratios implies that

$$\mathrm{tr}(\alpha_n^\vee\alpha_{n+1}) < 0 \quad \text{and} \quad \mathrm{tr}(\alpha_{n+1}^\vee\alpha_n) > 0.$$

Hence $\alpha_n^\vee$ and $-\alpha_{n+1}^\vee$ can play the role of β_n above. Consequently, the intersection point of $\mathbb{P}^1_{\alpha_n}$ and $\mathbb{P}^1_{\alpha_{n+1}}$ is the point $\{X_{\alpha_n^\vee}^\vee = 0, (X_{\alpha_{n+1}}^\vee)^{-1} = 0\}$. Since X_{α_j} is a coordinate on $\mathbb{P}^1_{\alpha_j}$, we retrieve that the intersection point of $\mathbb{P}^1_{\alpha_n}$ and $\mathbb{P}^1_{\alpha_{n+1}}$ is $(\infty, 0)$ in that coordinate system.

The disjointness of $\mathbb{P}^1_{\alpha_n}$ and $\mathbb{P}^1_{\alpha_{n+k}}$ is obvious from the gluing procedure. The statement on the V-action is an obvious consequence of the V-invariance of a fan.

The singularity statements are described in detail in [12], Section 2.2. □

Replacing the $\mathbb{Q}$-basis (α, β) by $(\alpha+\beta, \beta)$ resp. by $(\alpha, \alpha+\beta)$ corresponds to performing a sigma-transformation (or blowup) defined by the coordinate changes

$$(X_\alpha^{-1}, X_\beta) \mapsto (X_\alpha^{-1}X_\beta^{-1}, X_\beta) \quad \text{resp.} \quad (X_\alpha^{-1}, X_\beta) \mapsto (X_\alpha^{-1}, X_\alpha X_\beta)$$

at that point. It follows that the partial compactifications defined by any two fans are related by repeated blowing up and blowing down.

11.2. **Hirzebruch's compactification.** We describe the fan of Hirzebruch's compactification. There are many detailed expositions of this, in particular [15] and Chapter II of [35]. The aim is to compare the fans of Hirzebruch and Bainbridge below. We will use the letters A_k for Hirzebruch's fan (which is consistent with [35]) and subscripts k for the indexing that produces an increasing sequences of slopes. (Thus if we replace A_n by A_{k_0-n} for some k_0 we fit exactly the Definition 11.1.)

Suppose we want to compactify the Hilbert modular surface X_D at the cusp $\mathfrak{a}$ or equivalently the cusp at ∞ in $X_{D,\mathfrak{a}}$. Then, in the notation of §11.1 the module $M = \mathfrak{a}^\vee(\mathfrak{a}^{-1}) = (\mathfrak{a}^2)^\vee$. We choose the A_k to be the set of extreme points in $M^+ = M \cap (\mathbb{R}_+)^2$, i.e. the points lying on the convex hull $\mathrm{conv}(M^+)$ of M^+ in $(\mathbb{R}_+)^2$, indexed by increasing slope then then A_k form a fan since $\mathrm{conv}(M^+)$ is V-invariant. This compactification is smooth, because any two adjacent points on the boundary of $\mathrm{conv}(M^+)$ form a $\mathbb{Z}$-basis of M ([15] or [35], II) Lemma 2.1). Since the A_k lie on the boundary of the convex hull, we can write

$$A_{k-1} + A_{k+1} = p_k A_k \quad \text{with} \quad p_k \in \mathbb{Z},\, p \geq 2\,. \tag{90}$$

On this smooth compactification the self intersection number of $\mathbb{P}^1_{A_k}$ is $-p_k$ ([35], § II.2). Consequently, the compactification using at a cusp with stabilizer $G(M, V)$ the fan given by the boundary points of $\mathrm{conv}(M^+)$ is the minimal smooth compactification.

We call x *reduced* if x is real quadratic and

$$x > 1 > x' > 0.$$

The *fast-minus continued fraction algorithm* of $x \in \mathbb{R} > 1$ is defined by $x_0 = x$ and then inductively by

$$x_{k+1} = 1/(p_k - x_k) \tag{91}$$

so that

$$x_k \;=\; p_k - \cfrac{1}{p_{k+1} - \cfrac{1}{\ddots}} \quad .$$

Note that the narrow class of the ideal $[M] = [\langle 1, x\rangle]$ is unchanged under this continued fraction algorithm, since $\langle 1, x\rangle = (p-x)\langle 1, \frac{1}{p-x}\rangle$. The following lemma is the analogue of Lemma 10.6 above.

Lemma 11.3. *The fast minus continued fraction expansion of any quadratic irrational x is periodic, and is pure periodic if and only if x is reduced.*

Proof. This is well known and is stated, for instance, in §2.5 of [15], where the relation of the fast plus continued fraction expansion and the fast minus continued fraction expansion is also given. $\square$

This observation gives us an algorithm to compute the convex hull. Write $M = \mu_0\langle 1, x_0\rangle$ and apply the continued fraction algorithm until it becomes pure periodic, say at a reduced quadratic irrational x_1 with $M = \mu_1\langle 1, x_1\rangle$. Take $A_0 = \mu_1$ and $A_1 = \mu_1 x_1$ and then $A_k = A_{k-1}/x_k$ for $k \geq 2$, where $x_1, x_2, \ldots, x_n$ form the cycle of the continued fraction algorithm. The recursive definition of the x_k in (91) then is equivalent to (90).

Conclusion: The Hirzebruch compactification is given by the fan stemming from the lower convex hull or, equivalently, triggered by the fast minus continued fraction algorithm.

11.3. **Bainbridge's compactification.** Bainbridge's compactification of a Hilbert modular surface is defined using the Deligne-Mumford compactification of the moduli space of curves of genus two. For the details we recall facts about various bundles of one-forms over moduli spaces.

In Section 5.3 we introduced together with Teichmüller curves the vector bundle $\Omega\mathcal{M}_g$ of holomorphic one-forms over the moduli space $\mathcal{M}_g$ of curves of genus g. The moduli space of curves comes with the Deligne-Mumford compactification by stable curves $\overline{\mathcal{M}_g}$ and the vector bundle $\Omega\mathcal{M}_g$ extends to a vector bundle $\overline{\Omega\mathcal{M}_g}$, whose sections are *stable forms*. A stable form is a differential form on the normalization of the stable curves, holomorphic except for at most simple poles at the preimages of nodes and such that the residues at the two branches of a node add up to zero. We refer to $\overline{\Omega\mathcal{M}_g}$ and to the corresponding projective bundle $\mathbb{P}\overline{\Omega\mathcal{M}_g}$ as the *Deligne-Mumford compactification* of $\Omega\mathcal{M}_g$ (resp. of $\mathbb{P}\Omega\mathcal{M}_g$).

Strictly contained between $\mathcal{M}_g$ and $\overline{\mathcal{M}_g}$ is the partial compactification $\widetilde{\mathcal{M}_g}$ of stable curves of compact type, i.e. stable curves whose Jacobian is compact or equivalently of arithmetic genus g.

For $g = 2$ the Torelli map $t : \widetilde{\mathcal{M}_2} \to \mathcal{A}_2$ is an isomorphism. It extends to an isomorphism of the bundles of stable one-forms $t : \widetilde{\Omega\mathcal{M}_2} \to \Omega\mathcal{A}_2$ and also to the projectivized bundles $t : \mathbb{P}\widetilde{\Omega\mathcal{M}_2} \to \mathbb{P}\Omega\mathcal{A}_2$.

On the other hand, we have seen in Section 5.1 that a Hilbert modular surface X_D parametrizes principally polarized abelian varieties with real multiplication. There is a unique holomorphic one-form on such an abelian variety that is an eigenform for the action of real multiplication (with the embedding $K \to \mathbb{R}$ that we fixed throughout). The quotient map of a Siegel modular embedding by the action of $\mathrm{SL}(2, \mathcal{O}_D)$ defines a map $X_D \to \mathcal{A}_2$ and the choice of an eigenform lifts this map to an injection

$$\psi : X_D \to \mathbb{P}\Omega\mathcal{A}_2.$$

We thus use the same letter for this map as for the modular embedding. The image $t^{-1}(\psi(X_D))$ is called the *eigenform locus* (maybe the *projectivized eigenform locus* would be more precise). It parametrizes stable curves of genus two of compact type with real multiplication by $\mathcal{O}_D$. We denote by $\overline{X_D}^{DM}$ the closure of the eigenform locus in the Deligne-Mumford compactification $\mathbb{P}\overline{\Omega\mathcal{M}_2}$.

Definition 11.4. *The* Bainbridge compactification (called the *geometric compactification* in [1]) $\overline{X_D}^{B}$ *of a Hilbert modular surface* X_D *is the normalization of* $\overline{X_D}^{DM}$.

We now identify this compactification as a toroidal compactification. We restrict to the case D fundamental for simplicity and since the preparations in Section 10 have been carried out for this case only. In §11.5 we will prove:

Theorem 11.5. *Suppose that D is a fundamental discriminant. For each cusp given by the ideal class $\mathfrak{a}$ the sequence of multiminimizers for $\mathfrak{a}$, as derived from the continued fraction algorithm in § 10.4, forms a fan. The Bainbridge compactification is the toroidal compactification of the Hilbert modular surface obtained by using this sequence of multiminimizers.*

Comparing with §11.2, we can summarize this theorem and the algorithm in §10.4 as follows.

Conclusion: The Bainbridge compactification is given by the fan stemming from the multiminimizers or, equivalently, triggered by the slow plus greater-than-one continued fraction algorithm.

In §11.4 we review the properties of the compactification by Bainbridge. For the proof of Theorem 11.5 we will need only part of these properties. The local structure at the intersection points of the compactification curves is forced by normality and could also be derived from Proposition 11.2.

11.4. **Period coordinates and properties of the Bainbridge compactification.** In order to identify the Bainbridge compactification and to prove Theorem 11.5 we use the part of the work of Bainbridge, where he gives a coordinate system of $\overline{\Omega\mathcal{M}_2}$ and describes $\overline{X_D}^{DM}$ in there. He uses a lift by choosing a scaling of the one-form ω by fixing locally a loop α_1 and by imposing that $\int_{\alpha_1} \omega = r_1$ for some $r_1 \in K$. We will then compare the coordinates introduced for toroidal compactifications to this coordinate system and thereby prove the claimed isomorphism.

There are two relevant types of coordinate systems, both called *period coordinates* depending on the type of the stable curves. We follow [1], Section 6.6. If (X, ω) is in the boundary of $\overline{X_D}^{DM}$, then $g(X)$ is zero. The first type of coordinate system is around a stable curve X with two non-separating nodes. We moreover suppose that ω has two simple zeros. We let $\pm r_1$ and $\pm r_2$ be the residues of ω. Choose loops α_1, α_2 around the punctures such that $\int_{\alpha_i} \omega = r_i$. For a smooth surfaces in a neighborhood choose loops β_1, β_2 that complete the α_i-curves to a symplectic basis. Finally, choose a path I joining the two zeros of ω. Then on a neighborhood of (X, ω) in $\Omega\mathcal{M}_2$ the functions

$$\begin{aligned} v^{(E)} &= \int_{\alpha_1} \omega, \qquad w^{(E)} = \int_{\alpha_2} \omega, \\ y^{(E)} &= \mathbf{e}\left(\int_{\beta_1} \omega/v\right), \quad z^{(E)} = \mathbf{e}\left(\int_{\beta_2} \omega/w\right), \quad x^{(E)} = \int_I \omega \end{aligned} \tag{92}$$

are well-defined, i.e. independent of ambiguity in the choice of β_i given by Dehn twists around the corresponding α_i. These five functions form a system of coordinates on $\Omega\mathcal{M}_2$. We provide them with a superscript E (edge) to distinguish them. They will correspond to edges of the boundary of $\overline{X_D}^{DM}$. Note that $x = x^{(E)}$ is only well-defined up to an additive constant depending on the path of integration and that its sign depends on the choice of ordering of the two zeros.

From the geometry of the Deligne-Mumford compactification Bainbridge derives the following proposition that we use to prove Theorem 11.5.

Proposition 11.6 ([1], Proposition 7.18, Theorem 7.17 and Theorem 7.22)**.** *A pair (r_1, r_2) appears as residues of an eigenform of X_D if and only if $\lambda = r_2/r_1 \in \mathbb{Q}(\sqrt{D})$ and $N(\lambda) < 0$. More precisely, the irreducible components of the boundary of $\overline{X_D}^{DM}$ are in bijection with the unordered projective tuples $(r_1 : r_2)$, or equivalently to standard quadratic irrationals λ in $\mathcal{O}_D$.*

Near a boundary component labeled by λ, the Hilbert modular surface $\overline{X_D}^{DM}$ is cut out in $\Omega\mathcal{M}_g$ by the equations

$$v^{(E)} = r_1, \quad w^{(E)} = r_2 \quad \textit{and} \quad (y^{(E)})^a = (z^{(E)})^{-c}. \tag{93}$$

and the boundary curve is given in these coordinates as $\{y^{(E)} = z^{(E)} = 0\}$.

The boundary of $\overline{X_D}^B$ is a union of rational curves C_λ where $\lambda = r_1/r_2$.

We explain the last statement. If $\gcd(a, c) > 1$, then (93) shows that the compactification is not normal near $\{y^{(E)} = z^{(E)} = 0\}$: the normalization has $\gcd(a, c)$ local branches. Nevertheless, the preimage of $\{y^{(E)} = z^{(E)} = 0\}$ in the normalization $\overline{X_D}^B$ is a connected curve C_λ. The normalization map is a cyclic covering of order $\gcd(a, c)$ ramified precisely over the intersection points of C_λ with its two adjacent curves in the cusp resolution.

We include the following proposition for a complete description of the Bainbridge compactification. It is not needed for the proof of Theorem 11.5.

The second type of coordinate system is around a stable curve X consisting of two irreducible components joined at three non-separating nodes. Here we choose $\alpha_1, \alpha_2, \alpha_3$ to be the loops around the punctures and let $r_i = \int_{\alpha_i} \omega$ be the residue. We may orient the α_i so that $r_1 - r_2 + r_3 = 0$. The one-form ω necessarily has one zero on each of the irreducible components. We let γ_i be a path joining these two zeros that crosses α_i once with positive intersection and no other α_j. Then the five functions

$$\begin{aligned} v^{(V)} &= \int_{\alpha_1} \omega, \qquad w^{(V)} = \int_{\alpha_2} \omega, \quad x^{(V)} = \mathbf{e}\Big(\int_{\gamma_3} \omega/(w^{(V)} - v^{(V)})\Big), \\ y^{(V)} &= \mathbf{e}\Big(\int_{\gamma_1} \omega/v^{(V)}\Big), \quad z^{(V)} = \mathbf{e}\Big(\int_{\gamma_2} \omega/w^{(V)}\Big) \end{aligned} \tag{94}$$

form a system of coordinates on ΩM_2 near (X, ω). We provide them with a superscript V (vertex of the compactification cycle) to distinguish them. We also have to say how the coordinates $v^{(V)}, \ldots, z^{(V)}$ and $v^{(E)}, \ldots, z^{(E)}$ are related near a vertex. Suppose we unpinch the node corresponding to α_3. Then on this nearby surface

$$\beta_1 = \gamma_1 - \gamma_3, \quad \beta_2 = \gamma_2 + \gamma_3, \quad I = \gamma_3.$$

Consequently, we have

$$v^{(E)} = v^{(V)} =: v\,, \qquad w^{(E)} = w^{(V)} =: w$$

and

$$x^{(V)} = \mathbf{e}\big(x^{(E)}/(w-v)\big), \quad y^{(V)} = y^{(E)} \cdot \mathbf{e}\big(x^{(E)}/v\big), \quad z^{(V)} = z^{(E)} \cdot \mathbf{e}\big(x^{(E)}/w\big)\,.$$

The following proposition describes the intersection points of boundary curves of $\overline{X_D}^B$. It is proven in [1], Theorem 7.27 but it is also an immediate consequence of Theorem 11.5 and Proposition 11.2, in particular its last statement.

Proposition 11.7. *If λ^+ denotes the successor of λ for the slow plus greater-than-one continued fraction, then the curves C_λ and C_{λ^+} have exactly one point c_λ in common. Near this intersection point the Hilbert modular surface $\overline{X_D}^B$ is cut out in $\overline{\Omega M_g}$ by the equations*

$$v^{(V)} = r_1, \quad w^{(V)} = r_2 \quad \text{and} \quad \big(y^{(V)}\big)^a = \big(z^{(V)}\big)^{-c} \big(x^{(V)}\big)^{-a-b-c}.$$

The point c_λ is a cyclic quotient singularity of order

$$m_\lambda = \frac{a}{\gcd(a,c)\gcd(a,a+b+c)}\,.$$

11.5. **The proof of Theorem 11.5.** The first step is to show that the multiminimizers form a fan. For the cusp $\mathfrak{a}$ we may choose the normalization as in (85) and suppose that the multiminimizers are primitive elements in $(\mathfrak{a}^2)^\vee$. We next examine the slopes. Suppose that at some step of the slow plus greater-than-one continued fraction we have the representation $\mathfrak{a} = \mu_k\langle\lambda_k, 1\rangle$. Then

$$\frac{\sigma(\alpha_k)}{\alpha_k} = \frac{\lambda_k\mu_k^2}{\sigma(\lambda_k\mu_k^2)} \quad \text{and} \quad \frac{\sigma(\alpha_{k+1})}{\alpha_{k+1}} = \frac{(\lambda_k-1)\mu_k^2}{\sigma((\lambda_k-1)\mu_k^2)}$$

in both cases of the continued fraction algorithm. The ratio of these two fractions is

$$\frac{\sigma(\alpha_k)}{\alpha_k}\Big/\frac{\sigma(\alpha_{k+1})}{\alpha_{k+1}} = \Big(1+\frac{1}{\lambda_k-1}\Big)\Big(1-\frac{1}{\sigma(\lambda_k)}\Big) > 1$$

since λ_k is standard. Consequently the slopes of multiminimizers are decreasing. Since the continued fraction algorithm is periodic this shows that the sequence multiminimizers in continued fraction order forms a fan. We denote the compactification of the Hilbert modular surface using these fans of multiminimizers by $\overline{X_D}^{MM}$.

We let $\overline{X_D}^{MM,*}$, $\overline{X_D}^{B,*}$, resp. $\overline{X_D}^{DM,*}$ be the complement of the intersection points of the cusps resolution curves in the three compactifications, i.e. obtained by removing the codimension two boundary strata. Our aim to show that there is a map $m : \overline{X_D}^{MM,*} \to \overline{X_D}^{DM,*}$, which is an isomorphism over X_D and which maps the boundary components labeled by a standard quadratic irrational λ onto the component with the same label by an unramified cyclic covering of degree $g = \gcd(a,c)$. Since the Bainbridge compactification is normal, the map m factors through a map $\widehat{m} : \overline{X_D}^{MM,*} \to \overline{X_D}^{B,*}$, which is an isomorphism by the local description of the map $\overline{X_D}^{B,*} \to \overline{X_D}^{DM,*}$. Since the multiminimizer compactification and the Bainbridge compactification are normal, the codimension two indeterminacy of $\widehat{m}$ (on domain an range) can be resolved to a global isomorphism.

We want to define m in a neighborhood of a point on the component of the boundary of $\overline{X_D}^{MM,*}$ given by the multiminimizer α with corresponding standard quadratic irrational λ. Since $\alpha^\vee$ is the unique (up to sign) primitive element in $(\mathfrak{a}^2)^\vee$ with $\mathrm{tr}(\alpha\alpha^\vee) = 0$, local coordinates near this point as defined in the proof of Proposition 11.2 are just the coordinates $X = q_1^{\alpha^*} q_2^{\sigma(\alpha^*)}$ and $Y = q_1^{\beta^*} q_2^{\sigma(\beta^*)}$ used in the proof of Theorem 10.1.

The map m is given by assigning to a point in X_D the curve given as the vanishing locus of the Siegel theta function restricted to X_D together with the first eigenform. In order to understand the local behavior of this map near the boundary we may choose any convenient translate of the theta function. As in the previous proofs we take the characteristic $((\frac{1}{2},\frac{1}{2}),(\frac{1}{2},0)$ and the basis $\boldsymbol{\omega} = (\omega_1,\omega_2)$ of $\mathfrak{a}$ that is distinguished by the multiminimizer. Moreover, we define elliptic coordinates

$$S = \mathbf{e}(\mathrm{tr}(\omega_1^\vee \mathbf{u})), \quad T = \mathbf{e}(\mathrm{tr}(\omega_2^\vee \mathbf{u}))\,.$$

In the coordinates X, Y, S, T, the theta function is

$$\begin{aligned}\theta_{(m,m'),\boldsymbol{\omega}}(z_1,z_2) \;&=\; \sum_{x\in\mathbb{Z}^2} (-1)^{2x\cdot(m')^T} X^{F(\widetilde{x}_1,\widetilde{x}_2)/2} Y^{G(\widetilde{x}_1,\widetilde{x}_2)/2} S^{x_1} T^{x_2} \qquad (95)\\ &= 2X^{\frac{a+|c|}{2}} S^{-1/2}T^{-1/2}\left(Y^{G(\frac{1}{2},\frac{1}{2})/2}(ST-1) + Y^{G(\frac{1}{2},-\frac{1}{2})/2}(S-T) + O(X)\right),\end{aligned}$$

where $\widetilde{x}_i = x_i + \frac{1}{2}$ and where

$$F \;=\; \big[a/g, 0, -c/g\big], \quad G \;=\; \big[-bt, 2g, -bs\big]\,. \qquad (96)$$

Consequently, since the boundary is given by $X = 0$ the vanishing locus of θ degenerates there. As in §6.2 the limiting curve to a rational curve. We abbreviate $Z = Y^{G(\frac{1}{2},\frac{1}{2})/2}/Y^{G(\frac{1}{2},-\frac{1}{2})/2} = Y^{g/2}$, and obtain (for Z fixed) the equation

$$Z^{-1}(ST-1) + (S-T) \;=\; 0 \quad \text{or, equivalently} \quad (S-Z)(T+Z) \;=\; 1 - Z^2$$

for the rational curve. A parametrization with coordinate t is given by $S = t + Z$, $T = (1-Z^2)/t - Z$.

On a boundary component of the genus two Deligne-Mumford compactification parametrizing rational curves with two nodes, the cross-ratio of the four points is a coordinate. This is also the relative period x_E that was used by Bainbridge as coordinate (cf. Proposition 11.6), up to a Möbius transformation depending on r_1 and r_2, as one checks easily be integrating the limiting stable form.

Our last task is therefore to express the cross-ratio in terms of the coordinate X. The first eigenform is

$$\omega \;=\; du_1 \;=\; (\omega_2^\vee)^\sigma \frac{dS}{S} - (\omega_1^\vee)^\sigma \frac{dT}{T} \;=\; \left(\frac{(\omega_2^\vee)^\sigma}{t+Z} + \frac{(\omega_1^\vee)^\sigma}{t(1-\frac{Z}{1-Z^2}t)} \right) dt. \tag{97}$$

in the given rational parametrization. The cross-ratio of the four poles $0, \infty, -Z$, and $\frac{1-Z^2}{Z}$ is given by a Möbius transformation of $Z^{-2} = Y^{\gcd(a,c)}$. This completes the proof of the properties claimed about the map m, and of Theorem 11.5.

11.6. **Intersection of F_N with Bainbridge's and Hirzebruch's boundary.** We can now answer the question about the intersection of the modular and Teichmüller curves with the two compactifications. For the Hirzebruch compactification the statement in Theorem 11.8 is of course already in the literature.

Fix the cusp to be infinity and let F_N be one of the modular curves, as defined in §5.2, passing through the cusp at infinity. Then F_N can be given by an equation of the form

$$\lambda z_1 + \lambda^\sigma z_2 + B \;=\; 0 \qquad (B \in \mathbb{Z},\ \lambda \in \mathcal{O},\ \lambda\lambda^\sigma \;=\; -N). \tag{98}$$

The following result now follows immediately from Proposition 11.2 and the descriptions of the Hirzebruch and Bainbridge compactifications given in this section.

Theorem 11.8. *The curve F_N passes through an interior point of the cusp resolution cycle of the Bainbridge compactification (resp. of the Hirzebruch compactification) if there is a multiminimizer α_n for this cusp such that $\alpha_n/\mu \in \mathbb{Q}$ (resp. if there is an element A_n of Hirzebruch's lower convex hull fan such that $A_n/\mu \in \mathbb{Q}$).*

Otherwise, if

$$\frac{\alpha_n}{\alpha_n^\sigma} > \frac{\mu}{\mu^\sigma} > \frac{\alpha_{n+1}}{\alpha_{n+1}^\sigma}, \quad \textit{resp.} \quad \frac{A_n}{A_n^\sigma} > \frac{\mu}{\mu^\sigma} > \frac{A_{n+1}}{A_{n+1}^\sigma},$$

then the curve F_{N_μ} passes through the node corresponding to the intersection of the curves associated with α_n and α_{n+1} (resp. with A_n and A_{n+1}).

This result together with Theorem 10.1 and Theorem 10.2 reproves from the theta viewpoint another result of Bainbridge.

Corollary 11.9. *The curves W_D and the components of the reducible locus P_D intersect the boundary of the Bainbridge compactification only in interior points of the boundary curves.*

Note that the component $P_{D,\nu}$ of P_D as defined in Proposition 5.1 is given by equation (98) with $\lambda = \nu\sqrt{D}$, $B = 0$.

11.7. **Examples.** *Case* $D = 17$. We consider the cusp at ∞ for $\mathrm{SL}(\mathcal{O}^\vee \oplus \mathcal{O})$. In this case $M = \mathcal{O}_{17}^\vee$ and the standard quadratic forms are $[2,-3,-1]$, $[2,-1,-2]$, $[1,-3,-2]$,$[1,-1,-4]$ and $[1,1,-4]$. The following table contains the corresponding multiminimizers $\alpha \in \mathcal{O}_D^\vee$, scaled by the factor $\sqrt{17}$.

standard	$[1,-3,-2]$	$[1,-1,-4]$	$[1,1,-4]$	$[2,-3,-1]$	$[2,-1,-2]$
n	1	2	3	4	5
λ	$\frac{3+\sqrt{17}}{2}$	$\frac{1+\sqrt{17}}{2}$	$\frac{-1+\sqrt{17}}{2}$	$\frac{3+\sqrt{17}}{4}$	$\frac{1+\sqrt{17}}{4}$
$\sqrt{17}\alpha$	$\frac{-3+\sqrt{17}}{2}$	$\frac{-1+\sqrt{17}}{2}$	$\frac{1+\sqrt{17}}{2}$	$\frac{3+\sqrt{17}}{2}$	$4+\sqrt{17}$
$\sqrt{17}A_k$	$\frac{-3+\sqrt{17}}{2}$	$\frac{-1+\sqrt{17}}{2}$	$\frac{1+\sqrt{17}}{2}$	$\frac{3+\sqrt{17}}{2}$	$4+\sqrt{17}$
x_k	$\frac{7+\sqrt{17}}{4}$	$\frac{9+\sqrt{17}}{8}$	$\frac{7+\sqrt{17}}{8}$	$\frac{5+\sqrt{17}}{4}$	$\frac{5+\sqrt{17}}{2}$
k	5	4	3	2	1

By the singularity criterion Proposition 11.2 this compactification is smooth.

A reduced quadratic form is $x_1 = (5+\sqrt{17})/2$. Its fast minus continued fraction is listed in the lower part of that table. The point $A_1 = \frac{1}{\sqrt{17}}(4+\sqrt{17})$ lies on the lower convex hull of $\mathcal{O}^\vee_{\sqrt{17}} \cap (\mathbb{R}_+)^2$. By the algorithm for Hirzebruch's compactification the subsequent points on the lower convex hull are defined by $A_{k+1} = A_k/x_k$, as listed in the table.

The Bainbridge compactification has no singularities at the points c_λ and the number of reduced quadratic forms equals the number of standard quadratic forms. Hence the Hirzebruch and the Bainbridge compactification coincide in this case. This is consistent with the table listing the same values for $\sqrt{17}\alpha_n$ and for $\sqrt{17}A_k$.

Case $D = 41$. There are 11 standard and also 11 reduced quadratic forms. But here the Hirzebruch and the Bainbridge compactification do not coincide.

standard	$[1,-5,-4]$	$\cdots$	$[4,-5,-1]$	$[2,-3,-4]$	$[2,1,-5]$	$\cdots$	$[4,-3,-2]$
n	1	$\cdots$	6	7	8	$\cdots$	11
λ_n	$\frac{5+\sqrt{41}}{2}$	$\cdots$	$\frac{5+\sqrt{41}}{8}$	$\frac{3+\sqrt{41}}{4}$	$\frac{-1+\sqrt{41}}{4}$	$\cdots$	$\frac{3+\sqrt{41}}{8}$
$\sqrt{41}\alpha_n$	$\frac{-5+\sqrt{41}}{2}$	$\cdots$	$\frac{5+\sqrt{41}}{2}$	$\frac{19+3\sqrt{41}}{2}$	$\frac{83+13\sqrt{41}}{2}$	$\cdots$	$\frac{429+67\sqrt{41}}{2}$
$\sqrt{41}A_k$	$\frac{-5+\sqrt{41}}{2}$	$\cdots$	$\frac{5+\sqrt{41}}{2}$	$6+\sqrt{41}$	$\frac{19+3\sqrt{41}}{2}$	$\cdots$	$826+129\sqrt{41}$
x_k	$\frac{13+\sqrt{41}}{8}$	$\cdots$	$\frac{11+\sqrt{41}}{8}$	$\frac{9+\sqrt{41}}{10}$	$\frac{7+\sqrt{41}}{4}$	$\cdots$	$\frac{11+\sqrt{41}}{10}$
k	11	$\cdots$	6	5	4	$\cdots$	1

At the intersection points of the curves C_{λ_7} and C_{λ_8} and also at the intersection points of C_{λ_8} and C_{λ_9} the Bainbridge compactification is smooth, but

$$\alpha_7 + \alpha_9 = \alpha_8 \quad \text{and also} \quad \alpha_9 + \alpha_{11} = \alpha_{10}$$

and hence this compactification is not minimal. In fact, C_{λ_8} (and also $C_{\lambda_{10}}$) is a (-1)-curve and the corresponding values α_8 and α_{10} do not show up in the list of A_k. On the other hand, at the intersection point of C_{λ_6} and C_{λ_7} the Bainbridge compactification has a quotient singularity of order two, since α_6 and α_7 generate an index two subgroup of $\mathcal{O}^\vee_{41}$. It can be resolved by blowing up, adding a (-2)-curve, corresponding to the value $\sqrt{41}A_5 = 6 + \sqrt{41}$ that does not show up in the list of $\sqrt{41}\alpha_n$. The singularity can also be read off from the quadratic form $[4, 5, -1]$ and Proposition 11.6. In terms of the convex hull of $\mathcal{O}^\vee_{41}$, the multiminimizer fan has two interior points and skips two boundary points, as is shown in Figure 3 below.

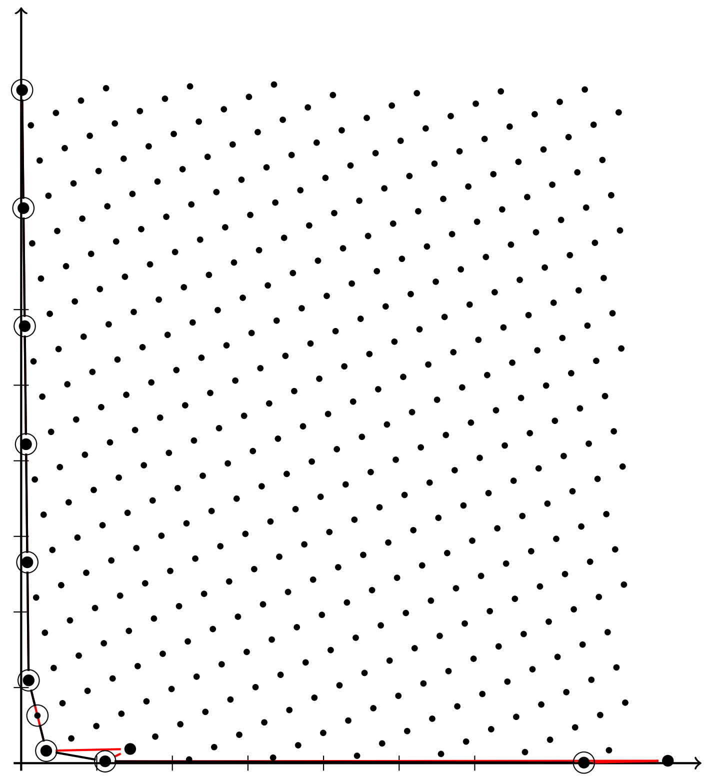

FIGURE 3. The multiminimizer fan (thick black points) and the Hirzebruch fan (circles) for $D = 41$. The rightmost thick black point is an interior point of the convex hull. Another point of the Hirzebruch fan skipped by the multiminimizer fan is not drawn (far to the right, close to the horizontal axis)

11.8. **Relating the two compactifications and two continued fraction algorithms.** The preceding examples show that the Bainbridge and the Hirzebruch compactification may sometimes agree, but that they are different in general. They also illustrate the general algorithm how to go from Bainbridge's compactification to Hirzebruch's: blow down curves corresponding to interior points and blow up points where boundary points of the lower convex hull have been omitted by the multiminimizers.

At first glance the previous examples suggest that at least the length of the boundary cycles in the Bainbridge and the Hirzebruch compactification agree. This is true for class number one, but otherwise the truth is more subtle, as we now explain.

To determine the geometry of the cusp $\mathfrak{a}$ in the Bainbridge compactification, we need to run the multiminimizer algorithm given in §10.4 for λ such that $\mathfrak{a} = \langle 1, \lambda \rangle$ and get as an output $\alpha \in (\mathfrak{a}^2)^\vee$. The geometry of Hirzebruch's minimal smooth resolution, however, depends only on the square of $\mathfrak{a}^2$, since we need to determine by the algorithm in §11.2 the lower convex hull of $(\mathfrak{a}^2)^\vee \cap \mathbb{R}^2_+$.

Consequently, whenever the squaring map is not an isomorphism on the ideal class group of $K = \mathbb{Q}(\sqrt{D})$, only the reduced quadratic irrationalities x such that $(\langle 1, x\rangle)^\vee$ is a square appear as label in the cusp resolution of the Hilbert modular surface X_D in the focus of our interest. However all the curves labeled by standard quadratic forms appear on X_D. An example to illustrate this is $D = 65$, where the class group (both in the narrow and wide sense) is of order two, so the multiminimizers for both cusps lie in an ideal equivalent to $\mathcal{O}_{65}^\vee$. Representatives of the multiminimizers are

$$\begin{aligned}
\mathrm{MM}(\mathcal{O}_{65}) &= \Big\{\tfrac{5+\sqrt{65}}{2\sqrt{65}}, \tfrac{7+\sqrt{65}}{2\sqrt{65}}, \tfrac{23+3\sqrt{65}}{2\sqrt{65}}, \tfrac{8+\sqrt{65}}{\sqrt{65}}, \tfrac{153+19\sqrt{65}}{2\sqrt{65}}, \\
&\qquad \tfrac{137+17\sqrt{65}}{2\sqrt{65}}, \tfrac{395+49\sqrt{65}}{2\sqrt{65}}, \tfrac{1685+209\sqrt{65}}{2\sqrt{65}}, \tfrac{1943+241\sqrt{65}}{2\sqrt{65}}\Big\} \\
\tfrac{7-\sqrt{65}}{2}\,\mathrm{MM}(\langle 1, \tfrac{-3+\sqrt{65}}{4}\rangle) &= \Big\{\tfrac{5+\sqrt{65}}{2\sqrt{65}}, \tfrac{7+\sqrt{65}}{2\sqrt{65}}, \tfrac{8+\sqrt{65}}{\sqrt{65}}, \tfrac{137+17\sqrt{65}}{2\sqrt{65}} \\
&\qquad \tfrac{395+49\sqrt{65}}{2\sqrt{65}}, \tfrac{653+81\sqrt{65}}{2\sqrt{65}}, \tfrac{911+113\sqrt{65}}{2\sqrt{65}}\Big\}
\end{aligned}$$

while the elements on the convex hull, up to multiplication by U_{65}^2, are given by

$$\begin{aligned}
\Big\{&\tfrac{5+\sqrt{65}}{2\sqrt{65}}, \tfrac{7+\sqrt{65}}{2\sqrt{65}}, \tfrac{8+\sqrt{65}}{\sqrt{65}}, \tfrac{137+17\sqrt{65}}{2\sqrt{65}}, \tfrac{395+49\sqrt{65}}{2\sqrt{65}}, \tfrac{653+81\sqrt{65}}{2\sqrt{65}}, \\
&\tfrac{911+113\sqrt{65}}{2\sqrt{65}}, \tfrac{1169+145\sqrt{65}}{2}, \tfrac{1427+177\sqrt{65}}{2}\Big\}
\end{aligned}$$

There is however an equality of total lengths of cusp resolution cycles, if one takes all Hilbert modular surfaces $\mathbb{H}^2/\mathrm{SL}(\mathfrak{b} \oplus \mathcal{O}_D)$ into account. We start our considerations on the level of quadratic irrationalities.

Reduced quadratic irrationalities as defined in (91) are well-known and have been used to label the boundary curves of Hirzebruch's compactification. Standard quadratic irrationalities as defined in (78) have been used to label the boundary curves of Bainbridge's compactification. There is an obvious bijection between these two classes, given by

$$\boxed{\begin{array}{l}\text{reduced}\\ x > 1 > x^\sigma > 0\end{array}} \quad \underset{\lambda = \frac{x}{x-1}}{\overset{x = \frac{\lambda}{\lambda-1}}{\longleftrightarrow}} \quad \boxed{\begin{array}{l}\text{standard}\\ \lambda > 1 > 0 > \lambda^\sigma\end{array}}$$

In order to pass from this correspondence of quadratic irrationalities to cusp resolutions, we define

$$C^{\mathrm{std}}(\mathfrak{a}) \;=\; \{(\mu, \lambda) \in K^+/U_D^2 \times K \,:\, \mathfrak{a} = \mu\langle 1, \lambda\rangle,\ \lambda \text{ standard}\}$$

and

$$C^{\mathrm{red}}(\mathfrak{a}) \;=\; \{(\rho, x) \in K^+/U_D^2 \times K \,:\, \mathfrak{a} = \rho\langle 1, x\rangle,\ x \text{ reduced}\}.$$

Then the map

$$\phi : C^{\mathrm{std}}(\mathfrak{a}) \to C^{\mathrm{red}}(\mathfrak{a}), \quad (\mu, \lambda) \mapsto (\rho, x) = (\mu(\lambda - 1), \tfrac{\lambda}{\lambda-1})$$

is obviously a bijection.

Proposition 11.10. *For any $\mathfrak{a}$, the length of the cycle of the Bainbridge compactification for the cusp corresponding to $\mathfrak{a}^2$ is equal to the length of the cycle of the Hirzebruch compactification for the cusp corresponding to $\mathfrak{a}$.*

Proof. Curves in the Bainbridge compactification for $\mathfrak{a}^2$ are in bijection to $C^{\mathrm{std}}(\mathfrak{a}^2)$ by Theorem 11.5 together with Proposition 10.4. Curves in the Hirzebruch compactification for $\mathfrak{a}$ are in bijection to $C^{\mathrm{red}}(\mathfrak{a}^2)$ by the algorithm in §11.2. We now can use the identification ϕ. □

This explains the initial observations for $D = 17$ and $D = 41$ and in general for $\mathfrak{a} = \mathfrak{a}^2 = \mathcal{O}_D$. For a more symmetric formulation we recall that the isomorphism class of the Hilbert modular surface $\mathbb{H}^2/\mathrm{SL}(\mathfrak{b} \oplus \mathcal{O}_D)$ depends on $\mathfrak{b}$ only up to squares of ideals and multiplication by a totally positive element in K. Consequently, there are $2^{t-1} = |\mathrm{Ker}(\mathrm{Sq} : \mathcal{C}l^+(\mathrm{D}) \to \mathcal{C}l^+(\mathrm{D}))|$ Hilbert modular surfaces for a given D, where t is the number of distinct prime factors of D. This also implies that every ideal class $\mathfrak{a}$ (in the wide sense) appears 2^{t-1} times as the module M associated with a cusp $\mathbb{H}^2/G(M, V)$ on the total collection of Hilbert modular surfaces for a given D. Altogether, this implies that the number

$$\ell(D) \;=\; \sum_{\substack{[a,b,c] \text{ reduced} \\ b^2-4ac=D}} 1 \;=\; \sum_{\substack{[a,b,c] \text{ standard} \\ b^2-4ac=D}} 1$$

appears in two incarnations.

Proposition 11.11. *The total number of boundary curves of the Bainbridge compactification of X_D is equal to $\ell(D)$.*

The total number of boundary curves of the Hirzebruch compactifications of the Hilbert modular surfaces $\mathbb{H}^2/\mathrm{SL}(\mathfrak{b}_i \oplus \mathcal{O}_D)$ for $\mathfrak{b}_i$ in a set of representatives of $\mathcal{C}l^+(D)/\mathrm{Sq}(\mathcal{C}l^+(D))$ is equal to $2^{t-1}\ell(D)$.

We could presumably make this statement even more symmetrical if we compactified all Hilbert modular surfaces $\mathbb{H}^2/\mathrm{SL}(\mathfrak{b} \oplus \mathcal{O}_D)$ using multiminimizers defined by theta functions, but we leave it to the reader to explore this.

12. Uniformization and disjointness from the reducible locus

The aim of this section is to give an independent proof of "W_D is a Teichmüller curve" using the definition via theta functions and applying Theorem 5.3. We emphasize that we derive all properties of W_D *ab ovo*, i.e., using just the definition as the vanishing locus of $\mathcal{D}\theta$ and without using anything that follows from the "geodesic" definition. Here again we simplify our counting task by restricting to fundamental discriminants.

12.1. **Transversality of W_D to the foliation $\mathcal{F}_1$.** Recall that in §5.3 we defined the "first Hilbert modular foliation" $\mathcal{F}_1$ of the Hilbert modular surface X_D to be the foliation defined by the constancy of the first cooordinate in the uniformization (i.e., by the equation $dz_1 = 0$, which is invariant under the action of the Hilbert modular group). Let $\varphi : \mathbb{H} \to \mathbb{H}$ be a holomorphic map such that $z \mapsto (z, \varphi(z))$ defines a branch of the vanishing locus of $\mathcal{D}\theta$, and suppose that the corresponding component W_D of the vanishing locus has the uniformization $W_D = \mathbb{H}/\Gamma$.

Theorem 12.1. *Suppose that D is a fundamental discriminant. Then the restriction to W_D of the derivative $\mathcal{D}^3\theta = \frac{\partial}{\partial z_2}\mathcal{D}\theta$ is a φ-twisted modular form of weight $(3, 11)$ for Γ that vanishes only at the cusps of W_D.*

Since the vanishing locus of a holomorphic function F is transversal to $\mathcal{F}_1$ at a point p if and only if $\frac{\partial}{\partial z_2}F(p) \neq 0$, the theorem immediately implies the transversality we want to prove.

For all $(m, m') \in (\frac{1}{2}\mathbb{Z}^2)^2$ and for any basis $\boldsymbol{\omega} = (\omega_1, \omega_2)$ of an $\mathcal{O}$-ideal $\mathfrak{a}$, define

$$D_2^3\theta_{(m,m'),\boldsymbol{\omega}}(\mathbf{z}) \;=\; \frac{\partial^3}{\partial u_2^3}\Big(\Theta_{(m,m')}(u, \psi_{\boldsymbol{\omega}}(\mathbf{z}))\Big)\Big|_{u=0},$$

where the Siegel modular embedding $\psi_{\boldsymbol{\omega}}$ is defined using $\boldsymbol{\omega}$. If we drop the index $\boldsymbol{\omega}$ we tacitly assume that we have chosen some basis of $\mathcal{O}$, as we did in (77). It follows that

$$D_2^3\theta_{(m,m'),(}(z_1, z_2) \;=\; \frac{\partial}{\partial z_2}\Big(D_2\theta_{(m,m')}(z_1, z_2)\Big)$$

(by the heat equation or by direct computation). Consequently,

$$\mathcal{D}^3\theta(z_1, z_2) \;=\; \sum_{(m_0,m_0')\text{ odd}} D_2^3\theta_{(m_0,m_0')}(z_1, z_2) \prod_{\substack{(m,m')\text{ odd}\\(m,m')\neq(m_0,m_0')}} D_2\theta_{(m,m')}(z_1, z_2),$$

explaining also the name given to this twisted modular form.

Proof. By applying the chain rule one sees that the restriction of the z_i-derivative of a Hilbert modular form to the vanishing locus of the form satisfies a modular transformation property with respect to the subgroup stabilizing this vanishing locus. This calculation also shows that the the z_i-derivative increases the weight in the ith component by two, proving the first claim.

To prove the second claim we will show that the vanishing orders of $\mathcal{D}^3\theta$ at the cusps sum up to the total vanishing order of a twisted modular form of this bi-weight. Since $\mathcal{D}^3\theta$ is holomorphic, it cannot then have any zeros at finite points. On a minimal compactification of the Hilbert modular surface X_D the number of intersection points of two modular forms of bi-weights (k_1, ℓ_1) and (k_2, ℓ_2) is $\frac{1}{4}(k_1\ell_2 + k_2\ell_1)|\chi(X_D)|$. This follows e.g. from [35], Section IV.2. This calculation is still valid when intersecting the vanishing locus W_D of a modular form with a section of the bundle of modular forms of bi-weight (k_2, ℓ_2) to W_D. Since here $(k_1, \ell_1) = (3, 9)$ and $(k_2, \ell_2) = (3, 11)$, the function $\mathcal{D}^3\theta_a$ has $15|\chi(X_D)|$ zeros on the closure of W_D. We have to show that they all lie at the cusps.

Note that precisely one of the two quadratic forms $[a, b, c]$ and $[-c, -b, -a]$ of discriminant D with $a > 0$ and $c < 0$ satisfies the additional condition $a + b + c < 0$ required to make it a standard quadratic form. From (34) we consequently deduce that

$$15\chi(X_D)| \;=\; \frac{1}{2}\sum_{\substack{D=b^2-4ac\\a>0,\;c<0}} a \;=\; \frac{1}{2}\sum_{\substack{[a,b,c]\text{ standard}\\D=b^2-4ac}} \big(a + |c|\big)\,.$$

To complete the proof it thus suffices to show that at each of the g cusps of W_D corresponding to λ, the vanishing order of $\mathcal{D}^3\theta$ is at least (and hence precisely) equal to $(a + |c|)/2g$. Here λ is a zero of the standard form $[a, b, c]$ and $g = \gcd(a, c)$. (The order of zero may indeed be half-integral, in accordance with the fact that $\mathcal{D}\theta$ is a modular form with a quadratic character.)

By Lemma 10.5 the vanishing orders of $\mathcal{D}^3\theta$ at the cusps of W_D mapping to α can be computed as the vanishing orders of

$$\mathcal{D}^3\theta_{\mathfrak{a}}(z_1,z_2) = \sum_{(m_0,m_0')\,\mathrm{odd}} D_2^3\theta_{(m_0,m_0'),\boldsymbol{\omega}}(z_1,z_2) \prod_{\substack{(m,m')\,\mathrm{odd}\\(m,m')\neq(m_0,m_0')}} D_2\theta_{(m,m'),\boldsymbol{\omega}}(z_1,z_2),$$

on the Hilbert modular surface $X_{D,\mathfrak{a}}$ at the cusps mapping to infinity. By Theorem 10.1 we associate a multiminimizer α to such a cusp and we may assume that $\boldsymbol{\omega} = (\omega_1,\omega_2)$ has been chosen to be the distinguished bases for this multiminimizer. If t is a local parameter of a cusps of W_D, as in the proof of Theorem 10.1 then the terms appearing in the expansion of $D_2\theta_{(m_0,m_0'),\boldsymbol{\omega}}$ and $D_2^3\theta_{(m_0,m_0'),\boldsymbol{\omega}}$ are $t^{F(\widetilde{x}_1,\widetilde{x}_2)/2}$, where $\widetilde{x}_i = x_i + m_i$ with $x_i \in \mathbb{Z}$ and where

$$F = \left[\frac{a}{g}, 0, \frac{-c}{g}\right]$$

as in (87). The minimal t-exponents are greater or equal to $(a+|c|)/8g$, $a/8g$, $|c|/8g$ in case m is equal to $(\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ respectively, both for $D_2\theta$ and $D_2^3\theta$. Since for each of these m there are precisely two m' such that (m,m') is odd, the total vanishing order is at least $(a+|c|)/2g$, which is what we wanted to show. $\square$

12.2. **Disjointness of W_D from the reducible locus (by counting zeros).** In this and the following subsection we give two completely different proofs of the following result, which is the second half of what we need to apply the criteria of Theorem 5.3 and show that the vanishing locus of $\mathcal{D}\theta$ is a Teichmüller curve. The first proof is similar to the one used for Theorem 12.1, by comparing the number of known zeros of a twisted modular form with its total number of zeros.

Theorem 12.2. *The vanishing locus W_D of $\mathcal{D}\theta$ is disjoint in X_D from the reducible locus P_D.*

Proof. Recall that the reducible locus is the vanishing locus of the product of all 10 even theta functions. This product is a Hilbert modular form of weight $(5,5)$, so its restriction to W_D is a modular form for W_D of bi-weight $(5,5)$. As in the preceding proof we deduce that the degree of its divisor (on a compactification of W_D) is $\frac{1}{4}(5\cdot 3 + 5\cdot 9)\,|\chi(X_D)| \;=\; 15\,|\chi(X_D)|$.

As in the preceding proof it suffices to show that the restriction of the product $\prod_{(m,m')\,\mathrm{even}}\theta_{(m,m')}$ to W_D vanishes at each of the g cusps of W_D corresponding to λ to the order at least $(a+|c|)/2g$. Here again we can work at the cusp ∞ of $X_{D,\mathfrak{a}}$. There, by the same argument as in Lemma 10.5, the product of the ten even theta functions is given by $\prod_{(m,m')\,\mathrm{even}}\theta_{(m,m'),\boldsymbol{\omega}}$, where $\boldsymbol{\omega}$ is some basis of $\mathfrak{a}$ and where $\theta_{(m,m'),\boldsymbol{\omega}}(\mathbf{z}) = \Theta_{(m,m'),\boldsymbol{\omega}}(0,\psi_{\boldsymbol{\omega}}(\mathbf{z}))$ with the modular embedding $\psi_{\boldsymbol{\omega}}$ defined using $\boldsymbol{\omega}$. The rest of the proof proceeds as above. To each cusp we associate its multiminimizer and take $\boldsymbol{\omega}$ to be the distinguished basis. The terms appearing in the expansion of $\theta_{(m,m'),\boldsymbol{\omega}}$ at such a cusp are $t^{F(\widetilde{x}_1,\widetilde{x}_2)/2}$, where $\widetilde{x}_i = x_i + m_i$ with $x_i \in \mathbb{Z}$ and where $F = \left[\frac{a}{g}, 0, \frac{-c}{g}\right]$. The minimal t-exponents are greater or equal to $(a+|c|)/8g$, $a/8g$, $|c|/8g$ and 0 in case m is equal to $(\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0)$, $(0,\frac{1}{2})$ and $(0,0)$ respectively. Each of the first three cases occurs twice among the ten even theta characteristics (and the irrelevant last case four times). Summing up these contributions gives again the vanishing order at least $(a+|c|)/2g$ that we claimed. $\square$

12.3. **Disjointness of W_D from the reducible locus (via theta products).** In this subsection we give a proof of Theorem 12.2 based on a completely different idea, by establishing a formula for the restriction of theta derivatives to the reducible locus. Let

$$\theta_{00}(z) = \sum_{n\in\mathbb{Z}} q^{n^2/2}\,, \quad \theta_{\frac{1}{2}0}(z) = \sum_{n\in\mathbb{Z}+1/2} q^{n^2/2}\,,$$
$$\theta_{0\frac{1}{2}}(z) = \sum_{n\in\mathbb{Z}} (-1)^n q^{n^2/2}\,, \quad \theta_{\frac{1}{2}\frac{1}{2}}(z) = \sum_{n\in\mathbb{Z}+1/2} (-1)^{n-1/2} n q^{n^2/2}\,.$$

(Here $\theta_{\frac{1}{2}\frac{1}{2}}$ should perhaps be called $\theta'_{\frac{1}{2}\frac{1}{2}}$, since the corresponding Jacobi Thetanullwert vanishes identically, but with these notations it will be easier to write a closed expression.) The product formulas found by Jacobi for these four functions, which in a modern notation say that they are equal to $\eta(2z)^5/\eta(z)^2\eta(4z)^2$, $2\eta(4z)^2/\eta(2z)$, $\eta(z)^2/\eta(2z)$ and $\eta(z)^3$, respectively, show that none of these functions vanish anywhere in the upper half-plane. Recall from Proposition 5.1 that the reducible locus is the union of irreducible curves $P_{D,\nu} = F_N(\nu)$ for $\nu = \frac{r+\sqrt{D}}{2\sqrt{D}}$, where $D = r^2 + 4N$ with $N \in \mathbb{N}$. Therefore Theorem 12.2 follows from the following theorem, which for simplicity we formulate only for D odd, the case of even D being similar.

Theorem 12.3. *Let $D \equiv 1 \pmod 4$ be a fundamental discriminant. Then for any odd theta characteristic (m,m') the restriction of the modular form $D_2\theta_{(m,m')}$ to the curve $P_{D,\nu}$ for ν as above has the factorization*

$$D_2\theta_{(m,m')}(\nu z, \nu^\sigma z) = \begin{cases} -\ \theta_{\hat m_1 m'_1}(z)\theta_{m_2\hat m'_2}(Nz) & \text{if } m_1 = m'_1 = 1/2 \\ -\sqrt{D}\nu^\sigma\, \theta_{\hat m_1 m'_1}(z)\theta_{m_2\hat m'_2}(Nz), & \text{if } m_2 = m'_2 = 1/2 \end{cases}$$

as a product of Jacobi theta functions, where $\hat m_1$ and $\hat m'_2$ are defined by

$$\begin{array}{llll} \hat m_1 = m_1 + m_2 \mod (1), & \hat m'_2 = m'_1 + m'_2 \mod (1) & \text{if } r \equiv 1 \mod 4 \\ \hat m_1 = m_1, & \hat m_2 = m_2, & \text{if } r \equiv 3 \mod 4. \end{array}$$

for $m = (m_1, m_2)$, $m' = (m'_1, m'_2)$. In particular, this restriction vanishes only at cusps.

Proof. As above we use a tilde to denote elements of the shifted lattice, i.e., $\widetilde{x}_i = x_i + m_i$. The restriction of the theta derivative to $P_{D,\nu}$ is

$$D_2\theta_{(m,m')}(\nu z, \nu^\sigma z) \ = \sum_{(x_1,x_2)\in\mathbb{Z}^2} (-1)^{2(x_1,x_2)(m')^T} \rho(\widetilde{x}_1,\widetilde{x}_2)^\sigma q^{\mathrm{tr}(\nu\rho(\widetilde{x}_1,\widetilde{x}_2)^2)/2}\,,$$

where $\rho(x_1,x_2) = x_1 + \frac{1+\sqrt{D}}{2}x_2$. With the $\mathbb{Z}$-linear transformation $\widetilde{y}_1 = \widetilde{x}_1 + \frac{r+1}{2}\widetilde{x}_2$ $\widetilde{y}_2 = -\widetilde{x}_2$ we obtain

$$\rho(\widetilde{x}_1,\widetilde{x}_2) \ = \ \widetilde{y}_1 + \frac{r-\sqrt{D}}{2}\widetilde{y}_2 \ = \ \widetilde{y}_1 + \sqrt{D}\nu^\sigma\widetilde{y}_2 \ =: \tau(\widetilde{y}_1,\widetilde{y}_2)$$

and

$$D_2\theta_{(m,m')}(\nu z, \nu^\sigma z) = \sum_{(\widetilde{y}_1,\widetilde{y}_2)\in\mathbb{Z}^2+(\hat m_1,m_2)} \epsilon(\widetilde{y}_1,\widetilde{y}_2)\tau(\widetilde{y}_1,\widetilde{y}_2)q^{(\widetilde{y}_1^2+N\widetilde{y}_2^2)/2},$$

where

$$\epsilon(\widetilde{y}_1,\widetilde{y}_2) = (-1)^{2(\widetilde{y}_1-\hat m_1)m'_1+2(\widetilde{y}_2-m_2)\hat m'_2}$$

and where we have used $\mathrm{tr}(\nu\tau(y_1,y_2)^2) = y_1^2 + Ny_2^2$. The q-exponent is invariant under both $\widetilde{y}_1 \mapsto -\widetilde{y}_1$ and $\widetilde{y}_2 \mapsto -\widetilde{y}_2$. Under, say, $\widetilde{y}_2 \mapsto -\widetilde{y}_2$ the sign of $\epsilon(\widetilde{y}_1,\widetilde{y}_2)$ is unchanged unless $m_2 = \hat{m}'_2 = 1/2$. Hence, unless $m_2 = \hat{m}'_2 = 1/2$, the $\sqrt{D}\nu^\sigma \widetilde{y}_2$-contribution of $\tau(\widetilde{y}_1, \widetilde{y}_2)$ cancels and one checks the formula by multiplying the unary theta functions. For an odd theta constant (m, m') precisely one of the cases $m_1 = \hat{m}'_1 = 1/2$ or $\hat{m}_2 = m'_2 = 1/2$ happens and a similar cancellation gives the formula in the second case, too. Notice that in each case of the theorem, one of the theta series in the decomposition is the function $\theta_{\frac{1}{2}\frac{1}{2}}$ of weight $3/2$, so that the total weight is always 2. □

Open problem: Can one reprove the irreducibility, stated in Theorem 5.2 and proved by McMullen using combinatorial number theory of the set of cusps, exclusively with techniques of (Hilbert) modular forms?

13. Applications

13.1. **The modular embedding via theta functions.** Using the description of W_D as vanishing locus of $\mathcal{D}\theta$, we can now give the modular embedding φ as in the "Fourier expansion" as defined in (2) for any cusp of W_D. In fact most of this was already achieved in the proof of Theorem 10.1. There we identified a cusp of W_D with a standard quadratic form $[a, b, c]$ (and hence a multiminimizer α) together with an element in $r \in \mathbb{Z}/g\mathbb{Z}$ where $g = \gcd(a, c)$ (equivalently, a solution S of (88)). Recall from this proof that for each such solution there is a unique branch, given by (83), of the locus $D_2\theta_{(m,m'),\mathfrak{a}}(z_1, z_2) = 0$. The map $z \mapsto (z, \varphi(z))$ describing this branch was given by

$$\varphi(z) = \frac{\alpha^\sigma}{\alpha} z + C + \frac{\varepsilon(q)}{2\pi i}. \tag{99}$$

We can now describe the arithmetic properties of this expansion.

Theorem 13.1. *The coefficients of the modular embedding describing the branch determined by the quadratic form $[a, b, c]$ and number S as in (88) through the cusp $\mathfrak{a}$ of X_D have the following properties.*

i) The constant C in (99) belongs to $\frac{1}{2\pi i} K^\times \log K^\times$. In fact, $\mathbf{e}\Big(\frac{g\alpha C}{N(\mathfrak{a})^2\sqrt{D}}\Big) \in K \smallsetminus \mathbb{Q}$.

ii) The number S lies in $K^{1/g}$.

iii) For each β completing the multiminimizer α determined by $[a, b, c]$ to a basis of $(\mathfrak{a}^2)^\vee$ there exists $A \in \mathbb{C}^*$ such that in the local parameter $Q = Aq$ of the cusp

$$\mathbf{e}(\nu z + \nu^\sigma \varphi(z)) \;=\; S^{\mathrm{tr}(\beta\nu)}\, Q^{\mathrm{tr}(\alpha\nu)} e^{\sigma(\nu)\varepsilon(q)} \qquad \text{for all} \quad \nu \in \mathfrak{a}^2\,.$$

The scalar A is transcendental of Gelfond-Schneider type, more precisely of the form x^y with x and y in $K \smallsetminus \mathbb{Q}$.

iv) The coefficients a_n of the power series $\varepsilon(q) = \sum_{n\ge1} a_n Q^n$ expanded in a local parameter Q as in ii) lie in the number field $K(S)$.

Proof. Statement i) obviously follows from (88) and (84). For ii) it suffices to test for ν the dual basis $\{\alpha^*, \beta^*\} \subset \mathfrak{a}^2$ of $\{\alpha, \beta\}$. Plugging in β^* confirms that S here is the same as in (84), since $\beta^* = \alpha^\sigma/N(\mathfrak{a})^2\sqrt{D}$. Plugging in $\alpha^* = \beta^\sigma/N(\mathfrak{a})^2\sqrt{D}$ implies that

$$A \;=\; \mathbf{e}\Big(\frac{\beta\, C}{N(\mathfrak{a})^2\sqrt{D}}\Big) \;=\; S^{\beta/\alpha} \;\in\; K^{\beta/g\alpha}\,.$$

The last statement iii) follows since the coefficients of Fourier expansion of the theta function lie in $K(S)$ and solving recursively for the a_n involves only these coefficients and integral powers of S. For concreteness, we perform the first step in this procedure in the case $|c| < a$. The summands $(\widetilde{x}_1, \widetilde{x}_2) = (\pm\frac{1}{2}, \pm\frac{1}{2})$ contribute to the lowest order term (in Q) of the restriction of the theta derivative to the branch $(z, \varphi(z)$. The next lowest order term is determined by $(\widetilde{x}_1, \widetilde{x}_2) = (\pm\frac{1}{2}, \pm\frac{3}{2})$ and we can solve for

$$a_1 \;= S^{g+2bs} \frac{\rho_\omega(\frac{1}{2},\frac{3}{2})\big(\sigma(\alpha^*)F(\frac{1}{2},\frac{3}{2}) + \sigma(\beta^*)G(\frac{1}{2},\frac{3}{2})\big)}{\rho_\omega(\frac{1}{2},\frac{1}{2})\big(\sigma(\alpha^*)F(\frac{1}{2},\frac{1}{2}) + \sigma(\beta^*)G(\frac{1}{2},\frac{1}{2})\big)}$$

using the notation ρ_ω introduced in the lines before (81), where F and G given in (87) and $g = sa + tc$. □

We remark that the "A" of §7.1 is not quite the same as the one above, and that the numerical value given in (49) is not of the form x^y with x, $y \in K$, but an algebraic multiple of this. This is because we normalized our Q in §7.1 so that the expansion of $t = Q + \cdots$ had leading coefficient 1. If we changed Q by the algebraic factor, then t and y would still have Q-expansions with coefficients in K, so that this would be an equally good choice in the special case $D = 17$, but for the general statement it seems best to normalize Q as above.

13.2. **Fourier coefficients of twisted modular forms.** Fix D, a fractional $\mathcal{O}_D$-ideal $\mathfrak{a}$, and a branch of the vanishing locus of $\mathcal{D}\theta$ through the cusp at ∞ of the Hilbert modular surface $X_{D,\mathfrak{a}}$ given by the quadratic form $[a, b, c]$ and S as above. Let $\Gamma \subset \mathrm{SL}(\mathfrak{a}^\vee \oplus \mathfrak{a})$ be the subgroup stabilizing this branch. It is the Fuchsian group uniformizing the curve W_D, normalized so that the cusp labeled with $([a, b, c], S)$ is the cusp at ∞.

The first twisted modular form for Γ we encountered was the form $\varphi'(z)$ of bi-weight $(2, -2)$ in the case $D = \sqrt{17}$. By the preceding theorem we know that this modular form has a Fourier expansion of the form

$$\varphi'(z) = \sum_{n \geq 0} b_n (Aq)^n$$

with A transcendental of Gelfond-Schneider type and b_n in the field $K(S)$ $(= K$ in this case, since $g = 1)$. We will show that such a statement holds for all twisted modular forms.

This section is inspired by work of Wolfart in the case of non-compact Fuchsian triangle groups. Let $\Delta(\infty, q, r)$ be such a group. We normalize it so that $\Delta(\infty, q, r) \subset \mathrm{SL}(2, \bar{\mathbb{Q}})$ and that ∞ is a cusp of $\Delta(\infty, q, r)$. In this situation Wolfart shows in [37]) that there exists some $A \in \mathbb{C}$ such that the space of (ordinary, i.e. of bi-weight $(k, 0)$) modular forms for $\Delta(\infty, q, r)$ admits a basis given by forms f_m with Fourier expansions

$$f_m(z) = \sum_{n \geq 0} r_{m,n} (Aq)^n \quad \text{with } r_{m,n} \text{ rational},$$

where $q = \mathbf{e}(z/a_0)$. The constant A is transcendental of Gelfond-Schneider type if the Fuchsian group is non-arithmetic and is algebraic in the finitely many other cases.

The uniformizing group of the curves W_D is non-arithmetic and the following result extends Wolfart's non-algebraicity result to this class of curves.

Theorem 13.2. *The space of twisted modular form of all bi-weights for the group* Γ *has a basis of forms with Fourier expansions* $\sum_{n\geq 0} a_n Q^n$ *with* $a_n \in K(S)$ *with* $Q = A\mathbf{e}(z/\alpha)$ *and* S *as in Theorem 13.1 ii). The number* A*, and also the radius of convergence* $1/|A|$ *of this series, is transcendental of Gelfond-Schneider type.*

Wolfart's proof of A being of Gelfond-Schneider type relies on properties of Γ-functions and trigonometric calculations. It overlaps with our result in the few cases ($D = 5, 8, 12$) where the uniformizing group of W_D is a triangle group. The rationality of r_n in Wolfart's result follows easily from the rationality of the coefficients of the Picard-Fuchs differential operators. This rationality does not hold for general D, as can be seen from our example (41). The statement of the theorem can indeed not be strengthened to rationality of the coefficients, even in the case $k = \ell = 0$, since the modular function t in (46) admits no rescaling that has rational coefficients.

Proof. For each (k, ℓ) the space of Hilbert modular forms of weight (k, ℓ) has a basis of forms whose Fourier expansions have rational coefficients. Suppose that $f = \sum_{\nu\in\mathfrak{a}^2} c_\nu \mathbf{e}(\nu z_1 + \nu^\sigma z_2)$ is such a basis element. Then the restriction to $(z_1, z_2) = (z, \varphi(z))$ with φ as in (99) is

$$f(q) = \sum_{\nu\in\mathfrak{a}^2} c_\nu S^{\mathrm{tr}(\beta\nu)}\, Q^{\mathrm{tr}(\alpha\nu)} e^{\sigma(\nu)\varepsilon(q)}$$

and so for this twisted modular form the claim directly follows from Theorem 13.1.

For k and ℓ sufficiently large the restriction map is surjective since $\mathcal{L}_1 \otimes \mathcal{L}_2$ is ample on X_D. Here $\mathcal{L}_1$ and $\mathcal{L}_2$ are the natural line bundles on X_D such that Hilbert modular forms of weight (k, ℓ) are sections of $\mathcal{L}_1^k \otimes \mathcal{L}_2^\ell$. (The surjectivity in fact holds already for $k \geq 4$ and $\ell \geq 10$. To see this, we tensor the structure sequence for $W_D \subset X_D$ with $\mathcal{L}_1^k \otimes \mathcal{L}_2^\ell$ and note that the cokernel of the restriction map lies in $H^1(X_D, \mathcal{I}_{W_D} \otimes \mathcal{L}_1^k \otimes \mathcal{L}_2^\ell)$. Since $\mathcal{I}_{W_D} \cong \mathcal{O}_{X_D}(-W_D) \cong \mathcal{L}_1^{-3}\mathcal{L}_2^{-9}$ and since $\mathcal{L}_1 \otimes \mathcal{L}_2$ is ample on X_D, Kodaira's vanishing theorem implies the claim.)

For the remaining cases, note that with f and f also $1/f$ and fg have Fourier expansions as claimed in the theorem. It thus suffices to consider the products of modular forms of bi-weight (k, ℓ) with a given modular form of large enough weight and then to apply the restriction argument. □

13.3. **The foliation by constant absolute periods.** As quotients of $\mathbb{H} \times \mathbb{H}$, Hilbert modular surfaces come with two natural holomorphic foliations, which we called $\mathcal{F}_1$ and $\mathcal{F}_2$. They also admit an interesting foliation defined using the $\mathrm{SL}(2, \mathbb{R})$-action on the space $\Omega\mathcal{M}_2$. The leaves of this foliation are upper half-planes, but the foliation is not holomorphic. It is studied in detail in [24]. In the context of studying the $\mathrm{SL}(2, \mathbb{R})$-action a natural local coordinate system on $\Omega\mathcal{M}_2$ is given by period coordinates, i.e. by integrating the holomorphic one-form ω along a chosen basis of the first homology relative to the zeros of ω. In the special case of genus two, we have used this coordinate system in (92).

We may identify the (Torelli preimage of a) Hilbert modular surface as a subset of $\mathbb{P}\Omega\mathcal{M}_2 = (\Omega\mathcal{M}_2)/\mathbb{C}^*$ by mapping X to the class of (X, ω), where ω is the eigenform for the first embedding of K (in the order that we have chosen once and for all). It follows from [1] or [24] that the foliation by constant absolute periods, where only x_E defined in (92) is allowed to vary, is the *first* foliation $\mathcal{F}_1$. The function x_E is not globally well-defined: its sign depends on the choice of an orientation and it may also be altered by a constant by adding the period of a closed loop. However $q = (dx_E)^2$ is a well-defined quadratic differential on each leaf of $\mathcal{F}_1$, independent of these choices.

The horizontal trajectories of this quadratic differential exhibit beautiful structures on the leaves of $\mathcal{F}_1$. They have been determined by McMullen in [26], using the following theorem.

Theorem 13.3. *The quadratic differential q is proportional to the restriction of the meromorphic modular form*

$$Q(z_1,z_2) \;=\; \Bigg(\prod_{(m,m')\,\mathrm{odd}} D_2\theta_{(m,m')}(z_1,z_2)\Bigg)\Bigg/\Bigg(\prod_{(m,m')\,\mathrm{even}} \theta_{(m,m')}(z_1,z_2)\Bigg)$$

of weight $(-2,4)$ to the leaf where z_1 is constant.

Proof. This follows directly from [1] and Theorem 9.1. More precisely, Bainbridge has determined the dependence on z_1. Since the quadratic differential depends on the choice of a holomorphic one-form on each Riemann surface, there exists a linear map Q_1 from the first eigenform bundle to quadratic differentials on the leaves of $\mathcal{F}_1$ that is locally defined by $q = (dx_E)^2$. Such a map is the same object as a meromorphic modular form of weight $(-2,4)$. This modular form Q_1 vanishes at W_D, where the zeros collide, and acquires a pole at the reducible locus P_D, where the zeros are infinitely far apart. It is shown moreover in [1] that both the vanishing order and the pole order are equal to one. Consequently, by Theorem 9.1 and the fact that the even theta characteristics vanish precisely at the reducible locus, Q/Q_1 is a holomorphic function on X_D. It then extends holomorphically to the Baily-Borel compactification by Hartog's theorem, since the boundary has codimension two. We deduce that Q/Q_1 has to be constant, which proves the claim. □

References

[1] Bainbridge, M., *Euler characteristics of Teichmüller curves in genus two*, Geom. Topol. **11** (2007), 1887–2073

[2] Bainbridge, M., Möller, M., *The Deligne-Mumford compactification of the real multiplication locus and Teichmüller curves in genus 3.* Acta Math. 208 no. 1 (2012), 1-92

[3] Beukers, F., *Irrationality proofs using modular forms*, Astérisque **147–148** (1987), 271–283

[4] Bouw, I., Möller, M., *Differential equations associated with non-arithmetic Fuchsian groups*, Journal London Math. Soc. **81** No. 1 (2010), 65–90

[5] Bouw, I., Möller, M., *Triangle groups, Teichmüller curves and Lyapunov exponents*, Annals of Math. **172** (2010), 139–185

[6] Bruinier, J., van der Geer, G. and Zagier, D., *The 1-2-3 of Modular Forms*, Springer, 2009

[7] Calta, K., *Veech surfaces and complete periodicity in genus two*, J. Amer. Math. Soc. **17** (2004), 871–908

[8] Carathéodory, C., *Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen*, Math. Ann. 64 (1907), 95–115

[9] Cohen, P., Wolfart, J., *Modular embeddings for some non-arithmetic Fuchsian groups*, Acta Arithm. **56** (1990), 93–110

[10] Choie, Y.J., Zagier, D., *Rational period functions for* $\mathrm{PSL}(2,\mathbb{Z})$, Contemp. Math. **143**, Amer. Math. Soc., Providence, RI (1993), 89–107

[11] Franke, H.-G., *Kurven in Hilbertschen Modulflächen und Humbertsche Flächen im Siegel-Raum.* Dissertation, Universität Bonn, 1977

[12] Fulton, W., *Introduction to toric varieties.* Annals of Mathematics Studies 131. Princeton University Press, Princeton, NJ (1993)

[13] Hammond, W., *The modular groups of Hilbert and Siegel*, Amer. J. Math. **88** (1966), 497–516
[14] Hermann, C.F., *Symmetrische Hilbertsche Modulformen und Modulfunktionen zu* $\mathbb{Q}(\sqrt{17})$, Math. Ann. **256** (1981), 191–197
[15] Hirzebruch, F., *Hilbert modular surfaces.* Enseignement Math. (2) **19** (1973), 183–281
[16] Hirzebruch, F., Zagier, D., *Intersection numbers of curves on Hilbert modular surfaces and modular forms of Nebentypus*, Invent. Math. **36** (1976), 57-113
[17] Hirzebruch, F., Zagier, D., *Classification of Hilbert modular surfaces*, in *Complex Analysis and Algebraic Geometry*, Iwanami Shoten & Cambridge Univ. Press, Tokyo (1977), 43–77
[18] Igusa, J.-I., *Theta functions.* Grundlehren der math. Wiss. **194**. Springer, New York-Heidelberg, (1972)
[19] Iwaniec, H., *Introduction to the Spectral Theory of Automorphic forms*, Bibl. Rev. Math. Iberoamericana, Madrid (1995)
[20] McMullen, C., *Billiards and Teichmüller curves on Hilbert modular surfaces*, J. Amer. Math. Soc. **16** (2003), 857–885
[21] McMullen, C., *Teichmüller curves in genus two: discriminant and spin*, Math. Ann. **333** (2005), 87–130
[22] McMullen, C., *Teichmüller curves in genus two: torsion divisors and ratios of sines*, Invent. Math. **165**(2006), 651–672
[23] McMullen, C., *Prym varieties and Teichmüller curves*, Duke Math. J. **133** (2006), 569–590
[24] McMullen, C., *Foliations of Hilbert modular surfaces*, Amer. J. Math. **129** (2007), 365–425
[25] McMullen, C., *Rigidity of Teichmüller curves*, Math. Res. Letters **16** (2009), 647–650
[26] McMullen, C., *Moduli spaces of isoperiodic forms on Riemann surfaces*, preprint (2012), Duke Math. J. **163** (2012), 2271–2323
[27] Möller, M., *Variations of Hodge structures of Teichmüller curves*, J. Amer. Math. Soc. **19** (2006), 327–344
[28] Möller, M., *Periodic points on Veech surfaces and the Mordell–Weil group over a Teichmüller curve*, Invent. Math. **165** (2006), 633–649
[29] Möller, M., Viehweg, E., *Kobayashi geodesics in* A_g, J. Differential Geom. **86** (2010), 355–379
[30] Möller, M., *Prym covers, theta functions and Kobayashi geodesics in Hilbert modular surfaces*, preprint (2011), to appear in Amer. J. Math.
[31] Mukamel, R., *Orbifold points on Teichmüller curves and Jacobians with complex multiplication*, Geom. Topol. **18** (2014), 779–829
[32] Ricker, S., *Symmetric Fuchsian quadrilateral groups and modular embeddings*, Quart. J. Math. 53 (2002), 75–86
[33] Stömbergsson, A., *On the uniform equidistribution of long closed horocycles*, Duke Math. J. **123** (2004), 507–547
[34] Schmutz-Schaller, P., Wolfart, J., *Semi-arithmetic Fuchsian groups and modular embeddings*, J. London Math. Soc. **61** (2000), 13–24
[35] van der Geer, G., *Hilbert Modular Sufaces*, Springer-Verlag (1987)
[36] Weiss, C., *Twisted Teichmüller curves*, Lecture Notes in Mathematics 2041 (2014), 166 pages
[37] Wolfart, J., *Eine arithmetische Eigenschaft automorpher Formen zu gewissen nicht-arithmetischen Gruppen*, Math. Ann. **262** (1983), 1–21
[38] Wolfart, J., *Der Überlagerungsradius gewisser algebraischer Kurven und die Werte der Betafunktion an rationalen Stellen*, Math. Ann. **273** (1985), 1–15
[39] Zagier, D., *Modular forms associated to real quadratic fields*, Inv. math. **30** (1975), 1–46
[40] Zagier, D., *On the values at negative integers of the zeta-function of a real quadratic field*, Enseignement Math. **22** (1976), 55–95
[41] Zagier, D., *Integral solutions of Apéry-like recurrence equations*, in *Groups and Symmetries: From the Neolithic Scots to John McKay*, CRM Proceedings and Lecture Notes **47** (2009), Amer. Math. Society, 349–366