Algebra & Number Theory

Volume 5
2011
No. 4

mathematical sciences publishers

Algebra & Number Theory

www.jant.org

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

The subscription price for 2011 is US $150/year for the electronic version, and $210/year (+$35 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory (ISSN 1937-0652) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
http://msp.org/

A NON-PROFIT CORPORATION
Typeset in LATEX

ALGEBRA AND NUMBER THEORY 5:4(2011)

Global descent obstructions for varieties

Jean-Marc Couveignes and Emmanuel Hallouin

We show how to transport descent obstructions from the category of covers to the category of varieties. We deduce examples of curves having **Q** as field of moduli, that admit models over every completion of **Q**, but have no model over **Q**.

1. Introduction

If k is a field, a k-variety is by definition a separated scheme of finite type over $\mathrm{Spec}(k)$. A k-curve is a variety of dimension 1 over k. A k-surface is a variety of dimension 2 over k.

1a. ***Statement of the main results.*** This work constructs descent obstructions in the category of varieties. For example, we prove the following:

Theorem 1.1. *There exists a projective, integral and smooth curve over* $\overline{\mathbf{Q}}$, *having* **Q** *as field of moduli, which has models over all the completions of* **Q** *but not over* **Q** *itself.*

The main idea is to start from a descent obstruction in the category of finite morphisms (also called covers) of curves, and to transport it into various other categories: the category of quasiprojective surfaces, the category of proper surfaces, and finally the category of smooth curves.

Theorem 1.2. *Let k be a field of characteristic zero, and k^a be an algebraic closure of k. Let X_k be a smooth, projective, geometrically integral curve over k and let X denote the base change to k^a of X_k. Let Y be a smooth, projective, integral curve over k^a and let $\varphi : Y \to X$ be a (possibly ramified) cover over k^a, having k as field of moduli. There exists a smooth, projective, integral curve over k^a having k as field of moduli and having exactly the same fields of definition as the initial cover φ.*

This work is supported by the French Agence Nationale pour la Recherche through project ALGOL (ANR-07-BLAN-0248).

MSC2000: primary 11R34; secondary 12G05, 11G35, 14D22.

Keywords: algebraic curve, global field, descent, field of definition, field of moduli, stack.

Descent obstructions have been constructed in the categories of G-covers and covers [Coombes and Harbater 1985; Dèbes and Fried 1994; Couveignes and Granboulan 1994] and in the category of dynamical systems [Silverman 1995]. As far as we know, no example of purely global descent obstruction was known for varieties. Mestre [1991] gave some examples of local obstructions for hyperelliptic curves. Dèbes and Emsalem [1999] give a criterion in order to have a model over its field of moduli for a curve. This criterion involves a particular model for the quotient of the curve by its automorphism group. Dèbes and Emsalem also prove that the local-global principle applies to the descent problem for a curve *together with its automorphisms*. However they leave open the question of the local-global principle for a curve (or a variety in general). Purely global descent obstructions for covers have been constructed by Ros and Couveignes:

Theorem 1.3 [Couveignes and Ros 2004, Corollaire 2]. *There exists a connected ramified $\overline{\mathbf{Q}}$-cover of $\mathbf{P}^1_{\mathbf{Q}}$ having $\mathbf{Q}$ as field of moduli, having models over all the completions of $\mathbf{Q}$, but having no model over $\mathbf{Q}$.*

Applying Theorem 1.2 to these obstructions, one proves Theorem 1.1.

1b. ***Overview of the paper.*** Let k be a field with characteristic zero. Let X_k be a smooth, projective and geometrically integral curve over k and set $X = X_k \times_k k^a$ to be the pullback of X_k along the inclusion $k \hookrightarrow k^a$. Our starting point is a smooth projective and integral curve Y over k^a and a nonconstant morphism $\varphi : Y \to X$. We look for a variety having the same field of moduli and the same fields of definition as φ. A natural candidate is the complement $X \times Y - G(\varphi)$ of the graph $G(\varphi)$ of φ in the product $X \times Y$. We call it the *mark* of φ. We expect this surface to have the same field of moduli and the same fields of definition as φ. In order to prove it, we construct a morphism from the stack of models of φ onto the stack of models of the mark of φ. Then, we try to prove that this morphism is an equivalence of categories.

In Section 2, we recall the definition of the stack of "models" of an algebraic variety over k^a (or of a cover of curves over k^a). Then, we explain how a morphism between the two gerbes associated with two objects relates the definition and moduli properties of either objects. It turns out that the key point is to control the group of automorphisms of the two objects. We first prove, in Section 3, that the base curve X of our starting cover φ can be assumed to have no nontrivial k^a-automorphism. In other words, we construct another k-curve X'_k without any nontrivial k^a-automorphism and a k^a-cover $Y' \to X'_k \times_k k^a$ having the same field of moduli and the same fields of definition as φ. In Section 4, we take it for granted that X has no nontrivial automorphism and we prove that the mark of φ has the same field of moduli and the same fields of definition as φ. In Section 5, we assume that the field of moduli of the cover φ is k and we construct a proper normal k^a-surface having k as field of moduli and the same fields of definition as φ. This proper surface is a cover of $X \times Y$, strongly ramified along the graph of φ. Finally, in Section 6, we construct a projective k^a-curve, having k as field of moduli, and having the same fields of definition as the initial cover φ. This curve lies on the previous surface, and is obtained by deformation of a stable curve having the same automorphism group as the surface.

Notation. If k is a field, we denote by k^a its algebraic closure. Let l be a k-extension and let X_l be an l-variety. We denote by $\mathrm{Aut}_l(X_l)$ or simply $\mathrm{Aut}(X_l)$ the group of automorphisms of the l-variety X_l (that is, automorphisms over $\mathrm{Spec}(l)$). We denote by $\mathrm{Aut}_k(X_l)$ the group of automorphisms of the *k-scheme* X_l (that is, automorphisms over $\mathrm{Spec}(k)$). For $f \in l(X_l)$ we write $(f)_0$ for the divisor of zeros of f, and $(f)_\infty$ for the divisor of poles.

2. Stack of "models"

In this section k is a field of characteristic zero and k^a is an algebraic closure of k.

2a. ***The conjugate of a variety.*** Let X be a k^a-variety. We denote by $\pi : X \to \mathrm{Spec}(k^a)$ the structural morphism. Let $\sigma : k^a \to k^a$ be a k-isomorphism. We denote by ${}^\sigma X$ the k^a-variety defined to be X itself with the structural morphism ${}^\sigma\pi = \mathrm{Spec}(\sigma)^{(-1)} \circ \pi$. The square below is cartesian and ${}^\sigma\pi$ is the pullback of π along $\mathrm{Spec}(\sigma)$.

$$\begin{tikzcd}
{}^\sigma X \arrow[r, "\pi"] \arrow[d, "\mathrm{Id}"] & \mathrm{Spec}(k^a) \arrow[r, "\mathrm{Spec}(\sigma)^{(-1)}"] & \mathrm{Spec}(k^a) \arrow[d, "\mathrm{Spec}(\sigma)"] \\
X \arrow[rr, "\pi"] & & \mathrm{Spec}(k^a)
\end{tikzcd}$$

With this (slightly abusive) notation one has ${}^\tau({}^\sigma(\pi)) = {}^{\tau\sigma}\pi$ and ${}^\tau({}^\sigma(X)) = {}^{\tau\sigma}X$. If X is an affine variety, then ${}^\sigma X$ is obtained from X by letting σ act on the coefficients in the defining equations of X.

2b. ***The field of moduli.*** It is natural to ask whether X and ${}^\sigma X$ are isomorphic. They are certainly isomorphic as schemes (and even equal by definition), but as varieties over k^a, they are isomorphic if and only if there exists an isomorphism ϕ_σ that makes the following diagram commute:

$$\begin{tikzcd}
X \arrow[r, "\phi_\sigma"] \arrow[d, "\pi"'] & X \arrow[d, "\pi"] \\
\mathrm{Spec}(k^a) \arrow[r, "\mathrm{Spec}(\sigma)"] & \mathrm{Spec}(k^a)
\end{tikzcd} \tag{1}$$

The existence of such a square means that the isomorphism $\mathrm{Spec}(\sigma)$ of $\mathrm{Spec}(k^a)$ lifts to an isomorphism ϕ_σ of X. The set of such σ is a subgroup of $\mathrm{Gal}(k^a/k)$. The fixed field of this subgroup is a finite extension of k called the *field of moduli* of X.

2c. ***Fields of definition.*** Another natural question: given $l \subset k^a$ an algebraic extension of k, do an l-variety $\pi_l : X_l \to \mathrm{Spec}(l)$ and a cartesian square

$$\begin{tikzcd}
X_l \arrow[d, "\pi_l"'] & X \arrow[l] \arrow[d, "\pi"] \\
\mathrm{Spec}(l) & \mathrm{Spec}(k^a) \arrow[l, "\mathrm{Spec}(\subset)"']
\end{tikzcd} \tag{2}$$

(where the line below is the spectrum of the inclusion) exist or not? If such a square exists we say that l is a *field of definition* of X. We say that $\pi_l : X_l \to \mathrm{Spec}(l)$ is a *model* of $\pi : X \to \mathrm{Spec}(k^a)$ over l. One may wonder whether it is important to impose the arrow below in the definition above. The answer is yes, it is, in general. The existence of such a cartesian square may depend on the chosen arrow below. However, if k is the field of moduli of X, then we may compose the cartesian squares in (1) and (2) as

$$\begin{tikzcd}
X_l \arrow[d, "\pi_l"'] & X \arrow[l] \arrow[d, "\pi"] & X \arrow[l, "\phi_\sigma"'] \arrow[d, "\pi"] \\
\mathrm{Spec}(l) & \mathrm{Spec}(k^a) \arrow[l, "\mathrm{Spec}(\subset)"'] & \mathrm{Spec}(k^a) \arrow[l, "\mathrm{Spec}(\sigma)"']
\end{tikzcd}$$

and choose any of the arrows below.

Another simple observation: if X has a model $\pi_k : X_k \to \mathrm{Spec}(k)$ over k, then k is the field of moduli of X. Indeed, we write X as a fiber product $X = X_k \times_k k^a$ and we take for ϕ_σ the fiber product $\mathrm{Id}_{X_k} \times_{\mathrm{Spec}(k)} \mathrm{Spec}(\sigma)$ where $\mathrm{Id}_{X_k} : X_k \to X_k$ is the identity on X_k. One may wonder whether the converse is true or not.

2d. ***Descent obstructions.*** Let's assume that k is the field of moduli of the k^a-variety X. Does there exist a model over k? If the answer is no, we say that there is a *descent obstruction.* In case k is a number field, we say that the obstruction is *purely global* if k is the field of moduli and if there is no model over k, but for every place v of k there exists a model of X over some extension $l \subset k^a$ such that l can be embedded in the completion k_v of k at v.

2e. ***The fibered category of "models" of a variety.*** We denote by $\mathrm{Et} / \mathrm{Spec}(k)$ the category of finite étale morphisms over $\mathrm{Spec}(k)$. An object U in this category is a structural morphism $\mathrm{Spec}(l) \to \mathrm{Spec}(k)$, where $k \hookrightarrow l$ is a finite étale k-algebra. We define a *covering* of U to be a surjective family $(U_i \to U)_i$ of morphisms in

Et / Spec(k). This turns Et / Spec(k) into a site called the *étale site* on Spec(k). It satisfies the three axioms of a site: the pullback of a covering exists and is a covering, a covering of a covering is a covering, and the identity is a covering.[1]

Given a k^a-variety X, we define the fibered category over Et / Spec(k) of its "models". For any k-algebra l, we must say what we mean by a "model" of X over Spec(l). If l is a field, we say that an l-variety $\pi_l : X_l \to \mathrm{Spec}(l)$ is a "model" of X over Spec(l) if and only if there exist an embedding $e : l \hookrightarrow k^a$ over k and a cartesian square

$$\begin{tikzcd} X_l \arrow[d, "\pi_l"'] & X \arrow[l] \arrow[d, "\pi"] \\ \mathrm{Spec}(l) & \mathrm{Spec}(k^a). \arrow[l, "\mathrm{Spec}(e)"'] \end{tikzcd}$$

This time, we insist on the fact that we do not fix an embedding of l into k^a. In particular, if l is a subfield of k^a containing k, we accept models of X but also models of all its conjugates. So the word model here is less restrictive than in Section 2c. That is why we write the word model between quotation marks in that case. As we have already noticed, the two notions coincide when k is the field of moduli of X. If l is any finite étale algebra over k, then it is a direct product of finitely many finite field extensions of k. We define a "model" of X over Spec(l) to be a disjoint union of "models" of X over every connected component of Spec(l).

Definition 2.1 (the category $\mathbb{M}_X$ of "models" of X). Let X be a k^a-variety. The category of "models" of X, denoted $\mathbb{M}_X$, is the category whose objects are all "models" of X over all finite étale k-algebras, and whose morphisms are the cartesian squares

$$\begin{tikzcd} X_l \arrow[rr] \arrow[d, "\pi_l"'] & & X_m \arrow[d, "\pi_m"] \\ \mathrm{Spec}(l) \arrow[rr] \arrow[dr] & & \mathrm{Spec}(m) \arrow[dl] \\ & \mathrm{Spec}(k). & \end{tikzcd}$$

The functor that associates Spec(l) to every "model" over Spec(l) turns $\mathbb{M}_X$ into a fibered category over Et / Spec(k); we denote by $\mathbb{M}_X(l)$ or $\mathbb{M}_X(\mathrm{Spec}(l))$ the fiber over Spec(l).

[1] Note that in this paper, we use the word *covering* in the context of sites. We keep the word *cover* for a finite morphism between two varieties, for example, two smooth projective and geometrically integral curves.

In particular, we can pull back a "model" $X_l \to \mathrm{Spec}(l)$ along any morphism $\mathrm{Spec}(m) \to \mathrm{Spec}(l)$ over $\mathrm{Spec}(k)$. Let's note that pulling back is not quite innocent since it can turn a model into its conjugates, so to say.

2f. ***Descent data.*** We recall that under mild conditions, the fibered category $\mathbb{M}_X$ is a stack. We need a few definitions and elementary results about descent data (see Giraud [1964] or Vistoli's notes [2005]). Let $\mathbb{S}$ be a site and let $\mathbb{X}$ be a fibered category over $\mathbb{S}$. Let U be an object in $\mathbb{S}$ and let $\mathcal{U} = (U_i \to U)_i$ be a covering of U. A *descent datum* from $\mathcal{U}$ to U is a collection of objects $X_i \to U_i$. For every i and every j, we also want an isomorphism $\phi_{ij} : \pi_2^*(X_j) \to \pi_1^*(X_i)$, where π_1 and π_2 are the two "projections" in the cartesian diagram

$$\begin{tikzcd}
& U_i \times_U U_j \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\
U_i \arrow[dr] & & U_j. \arrow[dl] \\
& U &
\end{tikzcd}$$

We also require the compatibility relation

$$\pi_{12}^*(\phi_{ij}) \circ \pi_{23}^*(\phi_{jk}) = \pi_{13}^*(\phi_{ik}) \quad \text{for any } i,\ j, \text{ and } k. \tag{3}$$

Here the π_{12}, π_{23}, π_{31} are the partial "projections" in the cube

$$\begin{tikzcd}
& U_{ijk} \arrow[rr, "\pi_{23}"] \arrow[dl, "\pi_{12}"'] \arrow[dd, "\pi_{13}"'] & & U_{jk} \arrow[dl] \arrow[dd] \\
U_{ij} \arrow[rr] \arrow[dd] & & U_j \arrow[dd] & \\
& U_{ik} \arrow[rr] \arrow[dl] & & U_k \arrow[dl] \\
U_i \arrow[rr] & & U &
\end{tikzcd}$$

and $U_{ij} = U_i \times_U U_j$ and $U_{ijk} = U_i \times_U U_j \times_U U_k$.

A morphism of descent data is a collection of local morphisms that are compatible with the glueing morphisms on either sides. Thus we obtain a category $\mathrm{Desc}_{\mathbb{X}}(\mathcal{U}, U)$ for every covering $\mathcal{U}$ of U. We denote by $\mathbb{X}(U)$ the fiber of $\mathbb{X}$ above U. There is a functor $\mathbb{X}(U) \to \mathrm{Desc}_{\mathbb{X}}(\mathcal{U}, U)$ that associates to any object over U the collection of its restrictions over the U_i. These constructions are functorial. For example, if $\mathbb{Y}$ is another fibered category and $\mathbb{F} : \mathbb{X} \to \mathbb{Y}$ a cartesian functor, then $\mathbb{F}$ induces a functor from $\mathbb{X}(U)$ to $\mathbb{Y}(U)$ and a functor from $\mathrm{Desc}_{\mathbb{X}}(\mathcal{U}, U)$ to $\mathrm{Desc}_{\mathbb{Y}}(\mathcal{U}, U)$. Also, the composite functors $\mathbb{X}(U) \to \mathbb{Y}(U) \to \mathrm{Desc}_{\mathbb{Y}}(\mathcal{U}, U)$ and $\mathbb{X}(U) \to \mathrm{Desc}_{\mathbb{X}}(\mathcal{U}, U) \to \mathrm{Desc}_{\mathbb{Y}}(\mathcal{U}, U)$ are isomorphic. A fibered category

$\mathbb{X}$ over a site $\mathbb{S}$ is a *stack* if and only if all the functors $\mathbb{X}(U) \to \mathrm{Desc}_{\mathbb{X}}(\mathcal{U}, U)$ are equivalences of categories.

2g. ***When $\mathbb{M}_X$ is a stack, respectively a gerbe.*** If X is a k^a-variety, then $\mathbb{M}_X$ is a fibered category over Et / Spec(k) and it makes sense to ask if it is a stack or not. We first notice that if $l \subset m$ are two finite field extensions of k, then $\mathrm{Spec}(m) \to \mathrm{Spec}(l)$ is a covering of Spec(l). If further m/l is Galois, then a descent datum from Spec(m) to Spec(l) is a model $\pi_m : X_m \to \mathrm{Spec}(m)$ of X over Spec(m) and, for every σ in $\mathrm{Gal}(m/l)$, an automorphism $\phi_\sigma : X_m \to X_m$ of *l-scheme*, such that the diagram

$$\begin{tikzcd}
X_m \arrow[r, "\phi_\sigma"] \arrow[d, "\pi_m"'] & X_m \arrow[d, "\pi_m"] \\
\mathrm{Spec}(m) \arrow[r, "\mathrm{Spec}(\sigma)"] & \mathrm{Spec}(m)
\end{tikzcd}$$

commutes. We stress that the ϕ_σ need not be automorphisms of the m-variety X_m but only automorphisms of the l-scheme X_m. Let $\mathrm{Aut}_l(X_m)$ denote the set of automorphisms of the l-scheme X_m. The meaning of condition (3) is that the map $\mathrm{Spec}(\sigma) \mapsto \phi_\sigma$ must be a group homomorphism from $\mathrm{Aut}_{\mathrm{Spec}(l)}(\mathrm{Spec}(m))$ into $\mathrm{Aut}_l(X_m)$.

Proposition 2.2. *Let X be a variety over k^a. If X is affine or projective or if every finite subset of $X(k^a)$ is contained in an affine subvariety, then the fibered category $\mathbb{M}_X$ is a stack over* Et / Spec(k).

Proof. This is a consequence of Weil's descent theory. See the initial article of Weil [1956] or Serre's book [1959, Chapter V, Section 4]. □

Let us recall that a locally nonempty and locally connected stack is called a *gerbe*. More precisely, a stack $\mathbb{X}$ over a site $\mathbb{S}$ is a gerbe if and only if

(1) for every object U in $\mathbb{S}$, there exists a covering $(U_i \to U)_i$ of U such that the fibers over the U_i are nonempty,

(2) given two objects $X \mapsto U$ and $Y \mapsto U$ above U, there exists a covering $(U_i \to U)_i$ such that for every i the pullbacks $X \times_U U_i$ and $Y \times_U U_i$ are isomorphic over U_i, and

(3) for every object U in $\mathbb{S}$, the fiber $\mathbb{X}(U)$ is a groupoid.

The stack $\mathbb{M}_X$ of "models" of a variety X always satisfies conditions one and three, whereas the second holds true if and only if k is the field of moduli of X.

2h. ***The stack, next the gerbe of "models" of a cover of curves.*** Since the starting point of our construction is a cover of curves, we now define the stack of "models" of a cover of curves. Let X_k be a smooth, projective, geometrically integral curve over k. We set $X = X_k \times_k k^a$. Let Y be a smooth projective and integral curve over k^a and let $\varphi : Y \to X$ be a nonconstant morphism. We say that φ is a *cover* of X. An isomorphism between two covers $\varphi : Y \to X$ and $\psi : Z \to X$ is an isomorphism of k^a-varieties $i : Y \to Z$ such that $\psi \circ i = \varphi$.

The conjugate of a cover. If σ is a k-automorphism of k^a, the conjugate variety ${}^{\sigma}X$ is obtained from X by composing the structural morphism on the left with $\mathrm{Spec}(\sigma)^{(-1)}$. The same is true for Y. So any k^a-morphism φ from Y to X can be seen as a k^a-morphism ${}^{\sigma}\varphi$ from ${}^{\sigma}Y$ to ${}^{\sigma}X$. Since X is the fiber product of X_k and $\mathrm{Spec}(k^a)$ over $\mathrm{Spec}(k)$, we have a canonical isomorphism $\phi_\sigma = \mathrm{Id}_{X_k} \times_{\mathrm{Spec}(k)} \mathrm{Spec}(\sigma)$ between X and ${}^{\sigma}X$. The composite map $\phi_\sigma^{(-1)} \circ {}^{\sigma}\varphi$ is a morphism of k^a-varieties from ${}^{\sigma}Y$ to X. We call it the conjugate of φ by σ. We may denote it ${}^{\sigma}\varphi$ also by abuse of notation.

The field of moduli. This is the fixed field by the subgroup of $\mathrm{Gal}(k^a/k)$ consisting of all σ such that ${}^{\sigma}\varphi$ is isomorphic to φ.

Fields of definition, models. If $l \subset k^a$ is an algebraic extension of k, we set $X_l = X_k \times_k l$. Let Y_l be a smooth projective and geometrically connected l-curve. Let $\varphi_l : Y_l \to X_l$ be a cover. If we lift φ_l along the spectrum of the inclusion $l \subset k^a$, we obtain a morphism from $Y_l \times_l k^a$ onto $X = X_l \times_l k^a$. If this cover is isomorphic to $\varphi : Y \to X$, we say that φ_l is a model of φ over l. So it makes sense to ask whether (purely global) obstructions to descent for covers of curves exist or not. It is proved in [Couveignes and Ros 2004] that such obstructions do exist.

The fibered category of "models" of a cover. Given a finite étale k-algebra l, we explain the meaning of a "model" of φ over l. If l is a finite field extension of k, we set $X_l = X_k \times_k l$. Let Y_l be a smooth projective and geometrically integral curve over $\mathrm{Spec}(l)$, and let $\varphi_l : Y_l \to X_l$ be a cover defined over $\mathrm{Spec}(l)$. We pick now any embedding $e : l \to k^a$ over k. The pullback of X_l along $\mathrm{Spec}(e)$ is X (up to unique isomorphism) and we have the following diagram:

$$\begin{tikzcd}
 & & Y_l \arrow[dl, "\varphi_l"'] \arrow[ddl] & & Y_l \times_l k^a \arrow[ll] \arrow[dl] \arrow[ddl] \\
X_k \arrow[d] & X_l \arrow[l] \arrow[d] & & X \arrow[ll] \arrow[d] & \\
\mathrm{Spec}(k) & \mathrm{Spec}(l) \arrow[l] & & \mathrm{Spec}(k^a) \arrow[ll, "\mathrm{Spec}(e)"'] &
\end{tikzcd}$$

We say that φ_l is a "model" of φ if the cover $\varphi_l \times_{\mathrm{Spec}(l)} \mathrm{Spec}(k^a) : Y_l \times_l k^a \to X$ is isomorphic to φ. Again the choice of the embedding e does not matter. We just ask such an embedding to exist. If l is any finite étale algebra over k, we define a "model" of φ over $\mathrm{Spec}(l)$ to be a disjoint union of "models" of φ over every connected component of Spec(l). We write $\mathbb{M}_\varphi$ for the category of all models of φ. This is a fibered category over Et / Spec(k). The following proposition is a consequence of Weil's descent theorem.

Proposition 2.3. *Let X_k be a smooth, projective, geometrically integral curve over k and set $X = X_k \times_k k^a$. Let Y be a smooth projective and integral curve over k^a and let $\varphi : Y \to X$ be a nonconstant morphism of k^a curves. Then the fibered category $\mathbb{M}_\varphi$ is a stack over* Et / Spec(k).

As in the case of varieties, the stack $\mathbb{M}_f$ is a gerbe if and only if k is the field of moduli of f.

2i. ***Transporting obstructions.*** The next result summarizes the previous sections.

Proposition 2.4. *Let X be a k^a-variety* (*or a cover of curves*). *Then k is the field of moduli of X if and only if the stack $\mathbb{M}_X$ is a gerbe. The field l is a field of definition of X if and only if the fiber $\mathbb{M}_X(l)$ is not empty.*

Let X and Y be two k^a-varieties. A cartesian morphism of stacks $\mathbb{F} : \mathbb{M}_X \to \mathbb{M}_Y$ is a functor of fibered categories (it transforms cartesian squares into cartesian squares). Such an $\mathbb{F}$ associates an l-model $\mathbb{F}(X_l)$ of Y to every l-model X_l of X, and commutes with base change. One can use it to relate descent obstruction for X and for Y.

Proposition 2.5. *Let X and Y be either k^a-varieties or covers of curves. Suppose that there exists a morphism $\mathbb{F} : \mathbb{M}_X \to \mathbb{M}_Y$ of stacks.*

(1) *If k is the field of moduli of X, then k is the field of moduli of Y.*

(2) *If l is a field of definition of X, then l is also a field of definition of Y.*

(3) *Assume further that k is the field of moduli of X and that $\mathbb{F}$ is fully faithful. Then l is a field of definition of X if and only if l is a field of definition of Y.*

Proof. Assertions 1 and 2 result from Proposition 2.4 and Section 2g. Assertion 3 is a consequence of the following lemma. □

Lemma 2.6. *Let $\mathbb{X}$ and $\mathbb{Y}$ be two gerbes over a site $\mathbb{S}$ and let $\mathbb{F} : \mathbb{X} \to \mathbb{Y}$ be a cartesian morphism. If $\mathbb{F}$ is fully faithful, then $\mathbb{F}$ is essentially surjective.*

Proof. Let U be an object in $\mathbb{S}$ and let $Y \to U$ be an object in the fiber $\mathbb{Y}(U)$. Locally $\mathbb{X}(U)$ is not empty: there exist a covering $(U_i \to U)_i$ of U and objects $X_i \in \mathbb{X}(U_i)$ for every i. Set $Y_i = Y \times_U U_i$. Locally, Y_i and $\mathbb{F}(X_i)$ are isomorphic: there exists a covering $(U_{ij} \to U_i)_j$ such that $Y_i \times_{U_i} U_{ij}$ and $\mathbb{F}(X_i \times_{U_i} U_{ij})$ are isomorphic. Set $X_{ij} = X_i \times_{U_i} U_{ij}$ and $Y_{ij} = Y_i \times_{U_i} U_{ij}$.

The set of objects $(Y_{ij} \to U_{ij})_{ij}$ defines a descent datum from $(U_{ij} \to U)_{ij}$ to U; indeed for every i, j, i', j', pulling back identity gives rise to isomorphisms

$$\Phi_{iji'j'} : Y_{i'j'} \times_{U_{i'j'}} U_{iji'j'} \to Y_{ij} \times_{U_{ij}} U_{iji'j'}$$

which clearly satisfy the compatibility conditions (3) of Section 2f. Since $\mathbb{F}$ is fully faithful, there exist isomorphisms

$$\Psi_{iji'j'} : X_{i'j'} \times_{U_{i'j'}} U_{iji'j'} \to X_{ij} \times_{U_{ij}} U_{iji'j'}$$

which, in turn, satisfy the compatibility conditions (3) of Section 2f. We deduce that there exists $X \to U$ in $\mathbb{X}(U)$ such that $\mathbb{F}(X) = Y$. □

We now give an example of morphism between two stacks of "models" of a variety.

Proposition 2.7. *Let X be a integral variety over k^a having k as field of moduli and let G be a finite subgroup of* $\mathrm{Aut}_{k^a}(X)$ *which is normal in the group* $\mathrm{Aut}_k(X)$. *Assume that every orbit of G is contained in an affine open subset of X. Then there is a morphism from $\mathbb{M}_X$ to $\mathbb{M}_{X/G}$, where X/G denotes the quotient variety of X by G.*

Let $Y \subset X/G$ be the complement of the branch locus of $X \to X/G$. This is an open subvariety of X/G and there is a morphism of stacks from $\mathbb{M}_X$ to $\mathbb{M}_Y$.

Proof. We first need to define the image of an object. Let l be an extension of k and let X_l be a "model" of X in $\mathbb{M}_X(l)$. The elements of G may not be all defined over l, but there exists a finite Galois extension m of l over which they are. Set $X_m = X_l \times_l m$. Let $p_m : X_m \to X_m/G$ be the canonical projection onto the quotient of X_m by the group G. According to Section 2g, there exists a group homomorphism $\sigma \mapsto \phi_\sigma$ from $\mathrm{Gal}(m/l)$ to $\mathrm{Aut}_l(X_m)$. Since G is a normal subgroup of $\mathrm{Aut}_k(X)$, it is normal in $\mathrm{Aut}_l(X_m)$ also. Thus for every $g \in G$ and every $\sigma \in \mathrm{Gal}(m/l)$, one has $\phi_\sigma \circ g \circ \phi_\sigma^{-1} \in G$. We deduce that $p_m \circ \phi_\sigma \circ g = p_m \circ \phi_\sigma$ for every $g \in G$. This implies that ϕ_σ factorizes into $\psi_\sigma : X_m/G \to X_m/G$. By uniqueness of this factorization, the correspondence $\sigma \mapsto \psi_\sigma$ is a group homomorphism from $\mathrm{Gal}(m/l)$ to $\mathrm{Aut}_l(X_m/G)$; therefore the quotient X_m/G descends to l.

Next, we define the image of a morphism. Let $X_i \to \mathrm{Spec}(l_i)$ for $i = 1, 2$ be two "models" of X. One can complete a cartesian square involving the X_i as

$$\begin{tikzcd}
X_1 \arrow[d] & X_2 \arrow[l] \arrow[d] & X_m \arrow[l] \arrow[d] \\
\mathrm{Spec}(l_1) & \mathrm{Spec}(l_2) \arrow[l] & \mathrm{Spec}(m), \arrow[l]
\end{tikzcd}$$

where m is a finite Galois extension of k such that all the elements of G are defined over m. We know that there exist isomorphisms Φ_1 and Φ_2 making the following diagrams commute:

$$\begin{tikzcd}
Y_1 \arrow[d] & X_m/G \arrow[l, "\Phi_1"'] \arrow[d] & Y_2 \arrow[d] & X_m/G \arrow[l, "\Phi_2"'] \arrow[d] \\
\mathrm{Spec}(l_1) & \mathrm{Spec}(m) \arrow[l] & \mathrm{Spec}(l_2) & \mathrm{Spec}(m) \arrow[l]
\end{tikzcd}$$

The image of the starting cartesian square is nothing but

$$\begin{tikzcd}
Y_1 \arrow[d] & Y_2 \arrow[l, "\Phi_1 \circ \Phi_2^{-1}"'] \arrow[d] \\
\mathrm{Spec}(l_1) & \mathrm{Spec}(l_2). \arrow[l]
\end{tikzcd}$$

This completes the proof of the first statement. The second is true since taking the branch locus commutes with base changes. □

3. Cancellation of the automorphism group of the base curve

In this section, k is a field of characteristic zero, k^a is an algebraic closure of it, and $l \subset k^a$ is an algebraic extension of k. Let X_k be a projective, smooth, geometrically integral curve over k and set $X = X_k \times_k k^a$. We assume that we are given a smooth projective and integral curve Y over k^a and a cover $\varphi : Y \to X$ having k as field of moduli. We want to construct covers having the same field of moduli and the same fields of definition as φ but satisfying additional properties. For example, we would like the base curve X to have no nontrivial k^a-automorphism. We first prove that the degree of the cover can be multiplied by any prime integer not dividing the initial degree.

Proposition 3.1. *Let X_k be a smooth, projective, geometrically integral curve over k and set $X = X_k \times_k k^a$. Let Y be a smooth projective and integral curve over k^a and let $\varphi : Y \to X$ be a degree d cover. For every prime p not dividing d, there exist a smooth projective curve Y' over k^a and a degree pd cover $\psi : Y' \to X$ having the same field of moduli and the same fields of definition as φ.*

Proof. Let $f \in k(X_k)$ be a nonconstant function whose divisor is simple and does not meet the ramification locus of φ. The equation $h^p = f$ defines a degree p extension of $k(X_k)$. We denote by X'_k the smooth, projective, geometrically integral curve corresponding to this function field and we set $X' = X'_k \times_k k^a$. The morphism $\nu : X' \to X$ is a degree p cyclic cover. We fix an algebraic closure Ω of $k^a(X)$ and embeddings of $k^a(X')$ and $k^a(Y)$ in Ω. Let Y' be the smooth projective k^a-curve corresponding to the compositum of $k^a(Y)$ and $k^a(X')$. Since the field extensions $k^a(Y)$ and $k^a(X')$ are linearly disjoint over $k^a(X)$, the cover $\psi : Y' \to X$ has degree pd:

$$\begin{tikzcd}
& Y' \arrow[dl, "\varphi'"'] \arrow[dr, "\nu'"] \arrow[dd, "\psi"] & \\
X' \arrow[dr, "\nu"'] & & Y \arrow[dl, "\varphi"] \\
& X &
\end{tikzcd}$$

We prove this construction yields a morphism of stacks $\mathbb{F} : \mathbb{M}_\varphi \to \mathbb{M}_\psi$. Let $l \subset k^a$ be a finite extension of k. Set $X_l = X_k \times_k l$ and $X'_l = X'_k \times_k l$ and consider $\varphi_l : Y_l \to X_l$ an l-model of φ. In the construction above, one can replace X, X', Y by X_l, X'_l, Y_l. The l-curve Y'_l corresponding to the compositum of the two function fields $l(X'_l)$ and $l(Y_l)$ is smooth, projective, geometrically integral (because l is algebraically closed in the compositum) curve, and the l-cover $\psi_l : Y'_l \to X_l$ is an l-model of ψ. We define the morphism $\mathbb{F}$ by setting $\mathbb{F}(\varphi_l) = \psi_l$. Since the function f has been chosen in $k(X)$, the functor $\mathbb{F}$ maps cartesian squares to cartesian squares. Thus $\mathbb{F}$ is a morphism of stacks. According to Proposition 2.5, if l is a field of definition of φ then l is a field of definition of ψ and if φ has k as field of moduli then ψ has k as field of moduli.

To prove the converse, we construct a morphism the other way around using Proposition 2.7. Let ν' denote the Galois cover $Y' \to Y$. We need to show that the group $\mathrm{Aut}(\nu')$ is normal in $\mathrm{Aut}_k(\psi)$. Let $\Phi' \in \mathrm{Aut}_k(\psi)$. It induces maps $\Phi : Y \to Y$ and $\Psi : X \to X$ such that the following diagram commutes:

$$\begin{tikzcd}
Y' \arrow[rr, "\Phi'"] \arrow[d, "\nu'"'] \arrow[dd, bend right=60, "\psi"'] & & Y' \arrow[d, "\nu'"] \arrow[dd, bend left=60, "\psi"] \\
Y \arrow[rr, "\Phi"] \arrow[d, "\varphi"'] & & Y \arrow[d, "\varphi"] \\
X \arrow[rr, "\Psi"] \arrow[dr] & & X \arrow[dl] \\
& \mathrm{Spec}(k) &
\end{tikzcd} \tag{4}$$

(horizontal arrows are morphisms of k-schemes). The existence of Ψ results form X being defined over k. The morphism Φ exists because $Y \xrightarrow{\varphi} X$ is the maximal subcover of $Y' \xrightarrow{\psi} X$ unramified at the support of f. And f is k-rational. Now if $\Lambda \in \mathrm{Aut}_{k^a}(\nu')$, that is, $\nu' \circ \Lambda = \nu'$, then

$$\nu' \circ \Phi' \circ \Lambda = \Phi \circ \nu' \circ \Lambda = \Phi \circ \nu' = \nu' \circ \Phi',$$

so $\Phi' \circ \Lambda \circ \Phi'^{-1} \in \mathrm{Aut}_{k^a}(\nu')$, which was to be proved. So we have a morphism $\mathbb{G} : \mathbb{M}_\psi \to \mathbb{M}_\varphi$ of stacks and the lemma follows. □

Remark. The functor $\mathbb{F} : \mathbb{M}_\varphi \to \mathbb{M}_\psi$ is not fully faithful because ψ has more automorphisms than φ. This is why we cannot apply point (3) of Proposition 2.5 here. We construct another functor $\mathbb{G} : \mathbb{M}_\psi \to \mathbb{M}_\varphi$ instead and we apply points (1) and (2) of Proposition 2.5 to either functors $\mathbb{F}$ and $\mathbb{G}$ successively. We notice that $\mathbb{G}$ is a left inverse of $\mathbb{F}$.

Next, we show that the base curve can be assumed to have genus greater than 2.

Proposition 3.2. *Let X_k be a smooth, projective, geometrically integral curve over k and set $X = X_k \times_k k^a$. Let Y be a smooth projective and integral curve over k^a and let $\varphi : Y \to X$ be a degree d cover. There exist a smooth, projective, geometrically integral curve X'_k over k of genus greater than 2 and a cover $\varphi' : Y' \to X'_k \times_k k^a$ having the same field of moduli and the same fields of definition as φ.*

Proof. We use the construction and notation of diagram (4) above. We further assume that the chosen function f has degree at least 3. By the Hurwitz genus formula, the curve X' has a genus greater than or equal to 2. This construction yields a morphism of stacks $\mathbb{F} : \mathbb{M}_\varphi \to \mathbb{M}_{\varphi'}$. The cover $\varphi : Y \to X$ is the maximal subcover of $\psi : Y' \to X$ unramified at the support of f. Therefore, there exists a morphism from $\mathrm{Aut}_{k^a}(\varphi') \to \mathrm{Aut}_{k^a}(\varphi)$. This morphism is bijective because $k^a(X')$ and $k^a(Y)$ are linearly disjoint over $k^a(X)$. So the morphism $\mathbb{F}$ is fully faithful. We conclude by invoking Proposition 2.5. □

We now prove that one can assume that the base curve has no nontrivial k^a-automorphism.

Proposition 3.3. *Let X_k be a smooth, projective, geometrically integral curve over k and set $X = X_k \times_k k^a$. Let Y be a smooth projective and integral curve over k^a and let $\varphi : Y \to X$ be a cover. There exists a smooth, projective, geometrically integral curve X'_k over k, of genus greater that 2, such that $X' = X'_k \times_k k^a$ has no nontrivial automorphism and there exists a cover $\varphi' : Y' \to X'$ having the same field of moduli and the same fields of definition as φ.*

Proof. Thanks to Proposition 3.2, one can assume that the genus $g(X)$ of X is greater than 2. So the group $\mathrm{Aut}(X)$ of k^a-automorphisms is finite. Let $p \geq 3$ be a prime integer.

We first show that there exists a nonconstant function $f \in k(X)$ that is nonsingular above 2, -2 and ∞, of degree greater than $2 + 4p(g(X) - 1) + 2p^2$, such that the set $f^{-1}(\{-2, 2\})$ is not invariant by any nontrivial automorphism of X, and such that the set of singular values of φ does not meet the set $f^{-1}(\{2, -2, \infty\})$. Indeed, let D be a simple effective divisor on X with degree greater than $2 +$

$4p(g(X)-1)+2p^2$. We also assume that D is disjoint from the set of singular values of φ and the linear space $L(D)$ associated with D generates $k^a(X)$ over k^a. In particular, for every $\theta \in \mathrm{Aut}(X)$, this linear space is not contained in the Kernel of $\theta - \mathrm{Id}$. It is not contained in the kernel of $\theta + \mathrm{Id}$ either because it contains k^a. If D has been chosen of large enough degree, the functions in $L(D)$ having degree less than the degree of D are contained in a finite union of strict vector subspaces. Therefore there exists a nonconstant function $f \in L(D)$ such that $\deg(f) = \deg(D)$ and $\theta(f) \neq \pm f$ for all $\theta \in \mathrm{Aut}(X) \setminus \{\mathrm{Id}\}$. By construction, this function is not singular above ∞ and $f^{-1}(\infty)$ does not meet the singular values of φ. We can also assume that $f \in k(X_k)$.

By construction, the function f^2 has no nontrivial automorphism (in short, $\mathrm{Aut}_{k^a(f^2)}(k^a(X)) = \{\mathrm{Id}\}$). Using Lemma 7.3, we deduce that almost all the fibers of f^2 are nonsingular and not fixed by any nontrivial automorphism of $\mathrm{Aut}(X)$. In particular, there exists $\lambda \in k^*$ such that the fiber of f^2 above λ^2 is nonsingular, not fixed by any nontrivial automorphism in $\mathrm{Aut}(X)$ and does not meet the singular values of φ. The function $2f/\lambda$ satisfies all the required properties. Denote it by f.

Now the equation $h^p + h^{-p} - f = 0$ defines a regular extension of $k(X_k)$. Let X''_k be the smooth, projective, geometrically integral curve associated with this function field. We denote by w the automorphism of X''_k given by $w(h) = h^{-1}$ and by X'_k the quotient $X''_k/\langle w\rangle$; this is a smooth, projective, geometrically integral k-curve, covering X_k by a degree p k-cover $\nu_k : X'_k \to X_k$. Extending scalars to k^a, we obtain a Galois cover $X'' \to X$ of k^a-curves, with Galois group D_p, and whose singular values are exactly $f^{-1}(\{2,-2,\infty\})$. Since the subgroup $\langle w\rangle$ is self-normalized in D_p, the quotient by this subgroup is a subcover $\nu : X' \to X$ of k^a-curves of degree p having no nontrivial automorphism.

Because the ramification loci do not meet, the function fields $k^a(X'')$ and $k^a(Y)$ are linearly disjoint over $k^a(X)$. Let Y' and Y'' be the smooth, projective, integral curves corresponding to the compositum of $k^a(Y)$ with $k^a(X')$ and $k^a(X'')$, respectively. We have the diagram

$$\begin{tikzcd}
\mathbf{P}^1 \arrow[d] & X'' \arrow[l, "h"'] \arrow[d] & Y'' \arrow[l] \arrow[d] \\
\mathbf{P}^1 \arrow[d] & X' \arrow[l, "h+\frac{1}{h}"'] \arrow[d, "\nu"'] & Y' \arrow[l, "\varphi'"] \arrow[d] \\
\mathbf{P}^1 & X \arrow[l, "f"'] & Y \arrow[l, "\varphi"] \\
\underbrace{\hphantom{\mathbf{P}^1 \longleftarrow X}}_{k\text{-rational towers}} & &
\end{tikzcd}$$

The cover $Y'' \to Y$ is again a D_p-Galois cover and the cover $Y'' \to Y'$ has degree 2. Let us show that the cover $\varphi' : Y' \to X'$ has the expected properties.

First of all, it is clear that the construction above yields a morphism of stacks $\mathbb{F} : \mathbb{M}_\varphi \to \mathbb{M}_{\varphi'}$. The Galois equivariance is a direct consequence of the fact that the middle tower is defined over k. This morphism is in fact fully faithful because the subcover $\varphi : Y \to X$ of $\nu \circ \varphi' : Y' \to X$ is the maximal subcover unramified at $f^{-1}(\{2, -2, \infty\})$.

It remains to prove that the curve X' has no nontrivial automorphism. Let θ' be an automorphism of X'. Call Z the image of $\nu \times (\nu \circ \theta') : X' \to (X \times X)$. Let $\pi_1 : X \times X \to X$ be the projection onto the first factor. The map ν factors as $\nu : X' \to Z \xrightarrow{\pi_1} X$ and it has prime degree p. So Z is either isomorphic to X or birationaly equivalent to X'. In the latter case, the geometric genus of Z would be $> \frac{1}{4}\deg(f)p \geq 1 + 2p(g(X) - 1) + p^2$ by the Hurwitz genus formula. But the bidegree of Z is $\leq (p, p)$. So, by Lemma 7.1, its virtual arithmetic genus is less than $1 + 2p(g(X) - 1) + p^2$, a contradiction. Thus Z is a correspondence of bidegree $(1, 1)$ that defines an automorphism θ of X such that $\theta \circ \nu = \nu \circ \theta'$. Such an automorphism preserves the ramification data of ν, that of its Galois closure $X'' \to X$, and that of the unique subcover of degree 2 of the cover $X'' \to X$. Since this last cover is exactly ramified above $f^{-1}(\{-2, 2\})$, we deduce that $\theta = \mathrm{Id}$ and then that θ' is a k^a-automorphism of the cover ν. Since ν has no nontrivial automorphism, $\theta' = \mathrm{Id}$. □

4. Quasiprojective surfaces

In this section we prove the following theorem:

Theorem 4.1. *Let k be a field of characteristic zero. Let X_k be a smooth, projective, geometrically integral curve over k and set $X = X_k \times_k k^a$. Let Y be a smooth, projective, integral curve over k^a and let $\varphi : Y \to X$ be a nonconstant morphism of k^a curves. Then there exists a smooth quasiprojective integral k^a-surface having the same field of moduli and the same fields of definition as φ.*

First of all, by Propositions 3.2 and 3.3, one can assume the base curve X to have genus greater than 2 and no nontrivial k^a-automorphism. Let $G(\varphi) \subset X \times Y$ be the graph of φ. Let U be the open complementary set of $G(\varphi)$ in $X \times Y$. We show that the surface U has all the expected properties. We call it the *mark* of the cover $\varphi : Y \to X$. We first prove two lemmas.

Lemma 4.2. *Let l/k be a finite extension of k inside k^a. Let X_k be a smooth, projective, geometrically integral k-curve. Set $X = X_k \times_k k^a$ and assume that the genus of X is greater than 2 and that X has no nontrivial k^a-automorphism. Let U_l and V_l be the marks of two nontrivial geometrically integral l-covers $\varphi_l : Y_l \to X_l$ and $\psi_l : Z_l \to X_l$, where $X_l = X_k \times_k l$.*

Then every morphism of covers between $\varphi_l : Y_l \to X_l$ and $\psi_l : Z_l \to X_l$ induces a morphism between the corresponding marks U_l and V_l. Conversely, every *surjective l-morphism from U_l to V_l is equal to* $\mathrm{Id} \times \gamma_l$, *where $\gamma_l : Y_l \to Z_l$ is an l-morphism between the covers $\varphi_l : Y_l \to X_l$ and $\psi_l : Z_l \to X_l$.*

Proof. An l-morphism between the covers $Y_l \xrightarrow{\varphi_l} X_l$ and $Z_l \xrightarrow{\psi_l} X_l$ is a morphism of l-curves $\gamma_l : Y_l \to Z_l$ such that $\psi_l \circ \gamma_l = \varphi_l$. The product $\mathrm{Id} \times \gamma_l : X_l \times Y_l \to X_l \times Z_l$ maps the graph of φ_l to the graph of ψ_l and also the mark U_l to the mark V_l.

Conversely, let υ_l be a surjective l-morphism form U_l to V_l. We denote by $\upsilon : U \to V$, $\varphi : Y \to X$, $\psi : Z \to X$ the base change to k^a of υ_l, φ_l, ψ_l, respectively. Let y be a closed k^a-point of Y. Let $\pi_2 : X \times Z \to Z$ be the projection onto the second factor. The restriction of $\pi_2 \circ \upsilon$ to $(X \times \{y\}) \cap U$ is a constant because the genus of X is less than that of Z. We denote by $\gamma(y)$ this constant; this defines a morphism $\gamma : Y \to Z$ which cannot be constant since υ is surjective. Let $\pi_1 : X \times Z \to X$ be the projection onto the first factor. The restriction of $\pi_1 \circ \upsilon$ to $(X \times \{y\}) \cap U$ is a morphism β_y with values in X. Let $F \subset Y$ the set of closed k^a-points of Y such that the morphism β_y is constant. This is a closed set, and a finite one because υ is surjective. For a closed k^a-point $y \notin F$, the morphism β_y induces an automorphism of X, which is trivial because X has no nontrivial automorphism. Thus $\upsilon(x, y) = (x, \gamma(y))$ for every closed k^a-point x on X and y on Y with $y \notin F$ and $(x, y) \in U$. Let x be a closed k^a-point of X. The restriction of $\pi_1 \circ \upsilon$ to $(\{x\} \times Y) \cap U$ is constant and equal to x on the nonempty open set $(\{x\} \times (Y - F)) \cap U$. So it is a constant function. So F is empty and υ is the restriction of $\mathrm{Id} \times \gamma$ to U. Thus $\mathrm{Id} \times \gamma$ maps U to V and therefore $\psi \circ \gamma = \varphi$. Moreover γ must be defined over l since υ, U, V are defined over l. □

Lemma 4.3. *Suppose X_k is a smooth, projective, geometrically integral k-curve. Assume that $X = X_k \times_k k^a$ has genus greater than* 2 *and no nontrivial k^a-automorphism. Let U be the mark of a nonconstant k^a-cover $\varphi : Y \to X$, where Y is a smooth, projective, integral k^a-curve. Then*

(1) *k is the field of moduli of U (in the category of quasiprojective varieties) if and only if it is the field of moduli of the cover $\varphi : Y \to X$, and*

(2) *an algebraic extension of k is a field of definition of U if and only if it is a field of definition of the cover $\varphi : Y \to X$.*

Proof. It is clear that the construction of the mark from the cover commutes with base change. This yields a morphism of stacks $\mathbb{F} : \mathbb{M}_\varphi \to \mathbb{M}_U$, which is fully faithful according to Lemma 4.2. The result follows by Proposition 2.5. In particular, $\mathbb{F}$ has an inverse functor $\mathbb{G} : \mathbb{M}_U \to \mathbb{M}_\varphi$. □

5. Proper normal surfaces

In this section we prove the following theorem:

Theorem 5.1. *Let k be a field of characteristic zero. Let X_k be a smooth, projective, geometrically integral curve over k and set $X = X_k \times_k k^a$. Let Y be a smooth projective, integral curve over k^a and let $\varphi : Y \to X$ be a cover. Assume that k is the field of moduli of φ. Then, there exists a proper, normal and integral surface S over k^a, having k as field of moduli, and having the same fields of definition as φ.*

We construct the surface in question as a cover of the product $X \times Z$, strongly ramified along the graph of ψ, where $\psi : Z \to X$ is a well chosen cover derived from φ.

5a. ***Construction of the surface S.*** The construction of the surface is divided into several steps.

Step 1 (starting point). We keep notation and assumptions of Theorem 5.1. We denote by $g(X)$ the genus of X and by d the degree of the cover φ. According to Proposition 3.3, we may assume that $g(X)$ is at least 2 and that X has no nontrivial automorphism over k^a.

Step 2 (a system of generators $f_1, \dots, f_I$ of the function field $k(X_k)$). We need many *k-rational* functions on X.

Lemma 5.2 (the functions f_i on X and the primes p_i). *There exist $I \in \mathbf{N}^*$, some prime integers $p_1, \dots, p_I > d$, and functions $f_1, \dots, f_I \in k(X_k)$ satisfying the following conditions*:

(1) *The functions $(f_i)_{1 \le i \le I}$ generate the field $k(X_k)$ over k.*

(2) *For every $1 \le i \le I$ and every $\lambda \in k^a$, none of the functions $f_i - \lambda$ is a p_i-th power in $k^a(X)$.*

(3) *Let $\Pi = \prod_{i=1}^{I} p_i$ and let M and m be the maximum and minimum, respectively, among the degrees of the f_i. Then*

$$1 + 2(g(X) - 1)\Pi + \Pi^2 < m \le \deg(f_i) \le M$$

for all i such that $1 \le i \le I$.

Proof. We first choose a finite generating system $(h_j)_{1 \le j \le J}$ of $k(X_k)$ over k. We assume that none of the h_j is a power in $k^a(X)$. We set $I = 2J$ and let $\Pi = \prod_{i=1}^{I} p_i$ be the product of the first I prime integers greater than the degree d of φ. We choose two distinct prime integers a and b, both greater than $1 + 2(g(X) - 1)\Pi + \Pi^2$. For every $1 \le j \le J$, we set

$$f_j = h_j^a \quad \text{and} \quad f_{j+J} = h_j^b.$$

We can choose a and b in so that none of the functions $f_i - \lambda$ is a p_i-th power in $k^a(X)$ for $\lambda \in k^a$ and $1 \le i \le I$: this is evident for $\lambda = 0$. If $\lambda \ne 0$ and if $h_i^a - \lambda = \prod_{0 \le k \le a-1}(h_i - \zeta_a^k \lambda^{1/a})$ is a power, then h_i has at least a distinct singular values. This is impossible if we choose an a greater than the number of singular values of h_i.

We also note that the $(f_i)_{1\leq i\leq I}$ generate $k(X_k)$ over k and that they all have a degree greater than $1+2(g(X)-1)\Pi+\Pi^2$, as expected. □

Step 3 (a cover $\psi : Z \to X$ of large enough degree). Let p be a prime integer greater than $(g(X)+IM)\Pi$. We call Z the curve and $\psi : Z \to X$ the degree pd cover given by Proposition 3.1. The genus of Z is greater than $dp > (g(X)+IM)\Pi$ and the covers φ and ψ have the same field of moduli and the same fields of definition.

Step 4 (a system of functions $g_1, \dots, g_I$ on $X \times Z$). Using the previous functions f_i, we define functions on $X \times Z$.

Lemma 5.3 (the functions g_i on $X \times Z$). *For every* $1 \leq i \leq I$, *let* g_i *be the function on* $X \times Z$ *defined by*

$$g_i(P, Q) = f_i(\psi(Q)) - f_i(P).$$

Then

(1) *the negative part* $(g_i)_\infty$ *of the divisor of* g_i *is* $(f_i)_\infty \times Z + X \times (f_i \circ \psi)_\infty$,

(2) *the positive parts* $(g_i)_0$ *are such that* $\gcd_i((g_i)_0) = G(\psi)$, *where* $G(\psi)$ *is the graph of* ψ, *and*

(3) *for every point* $P \in X$ *the function* $Q \mapsto g_i(P, Q)$ *on* $P \times Z$ *is not a* p_i*-th power.*

Proof. The first two points are easy. To prove the third one, we notice that each function f_i is such that none of the $f_i - \lambda$ for $\lambda \in k^a$ is a p_i-th power (Lemma 5.2). Since the degree pd of ψ is prime to p_i, none of the function $f_i \circ \psi - \lambda$ is a p_i-th power in $k^a(Z)$. Condition 3 follows. □

We stress that if ψ is defined over a field l, then so are the functions g_i.

Step 5 (the surface S). Let $k^a(X \times Z)$ be the field of functions of $X \times Z$. We define a regular radicial extension of $k^a(X \times Z)$ by setting

$$y_i^{p_i} = g_i \quad \text{for every } 1 \leq i \leq I.$$

We denote by S the normalization of $X \times Z$ in the latter radicial extension. It is a normal surface by construction and there is a ramified cover

$$\chi : S \to X \times Z$$

that is Galois over k^a with Galois group $\prod_{i=1}^{I} \mathbf{Z}/p_i\mathbf{Z}$.

5b. ***The group of automorphisms of S.*** We denote by A the group of k^a-automorphisms of ψ. An element in A induces a k^a-automorphism of $X \times Z$, and this latter automorphism can be lifted uniquely to an automorphism of $k^a(S)/k^a$ that fixes all y_i and stabilizes $k^a(X \times Z)$. In the sequel we shall use the same notation for an automorphism of ψ, the induced automorphism of $X \times Z$ and its lift to S. In other words, A can be identified with a subgroup of $\mathrm{Aut}_{k^a}(S)$, the group of k^a-automorphisms of S.

Another subgroup of $\mathrm{Aut}_{k^a}(S)$ is the Galois group $B = \prod_{i=1}^{I} \mathbf{Z}/p_i\mathbf{Z}$ of the extension $k^a(S)/k^a(X \times Z)$. The groups A and B are the sets of α and β, respectively, such that the following diagrams commute:

$$\begin{tikzcd} S \arrow[r, "\alpha"] \arrow[d, "\chi"'] & S \arrow[d, "\chi"] \\ X \times Z \arrow[r, "\alpha"] & X \times Z \end{tikzcd} \qquad \begin{tikzcd} S \arrow[rr, "\beta"] \arrow[dr, "\chi"'] & & S \arrow[dl, "\chi"] \\ & X \times Z & \end{tikzcd} \tag{5}$$

It is clear that $A \times B \subset \mathrm{Aut}_{k^a}(S)$. We now prove that this inclusion is an equality. To this end, we introduce a family of curves on S.

Lemma 5.4 (the curves E_Q). *For any point Q on Z, we call E_Q the inverse image of $X \times Q$ by χ and we denote by $\chi_Q : E_Q \to X \times Q$ the restriction of χ to E_Q. The geometric genus of E_Q can be bounded from above*:

$$g(E_Q) \leq (g(X) + IM)\Pi < g(Z), \tag{6}$$

and the genus of any nontrivial subcover of χ_Q can be bounded from below:

$$1 + 2(g(X) - 1)\Pi + \Pi^2 < m \leq g(\textit{nontrivial subcover of } \chi_Q : E_Q \to X). \tag{7}$$

Proof. If Q is the generic point on Z, then E_Q is a geometrically integral curve and χ_Q is a degree Π, geometrically connected cover. The degree of the ramification divisor of this cover is bounded from above by the product $2IM$ (where I is the number of functions in the family $(f_i)_i$ and M is the maximum of the degrees of these functions). The upper bound follows.

As for the lower bound, let us consider a nontrivial subcover of χ_Q. Such a cover has degree at least $p_1 \geq 3$ and its ramification divisor has degree at least m (where m is the minimum among the degrees of the functions f_i). So its genus is greater than m and the lower bound follows. □

Lemma 5.5. *The group* $\mathrm{Aut}(S)$ *of k^a-automorphisms of S is $A \times B$.*

Proof. Let θ be a k^a-automorphism of S. Let Q be the generic point of Z. We know from inequality (6) of Lemma 5.4 that $g(E_Q) < g(Z)$. We deduce that $\theta(E_Q) = E_{\alpha(Q)}$, where α is a k^a-automorphism of Z.

We now prove that the isomorphism between E_Q and $E_{\alpha(Q)}$ induced by θ makes the following diagram commute:

$$\begin{tikzcd} E_Q \arrow[r, "\theta"] \arrow[d, "\chi_Q"'] & E_{\alpha(Q)} \arrow[d, "\chi_{\alpha(Q)}"] \\ X \times Q \arrow[r, "\mathrm{Id}\times\alpha"] & X \times \alpha(Q) \end{tikzcd} \tag{8}$$

Indeed, the cartesian product of the maps χ_Q and $\chi_{\alpha(Q)} \circ \theta$ defines a morphism

$$E_Q \xrightarrow{\chi_Q \times (\chi_{\alpha(Q)} \circ \theta)} X \times X,$$

whose image W is a divisor with bidegree $\leq (\Pi, \Pi)$. Using Lemma 7.1 we deduce that the arithmetic genus of W is smaller than or equal to $1 + 2(g(X) - 1)\Pi + \Pi^2$. Let $\pi_1 : X \times X \to X$ be the projection on the first factor. The morphism χ_Q factors as

$$\chi_Q : E_Q \to W \xrightarrow{\pi_1} X.$$

The map $W \xrightarrow{\pi_1} X$ is a birational isomorphism, else it would define a nontrivial subcover of $\chi_Q : E_Q \to X$. But we know from inequality (7) of Lemma 5.4 that such a subcover has geometric genus greater than or equal to $m > 1+2(g(X)-1)\Pi+\Pi^2$. A contradiction. We deduce that W is a correspondence of bidegree $(1, 1)$. Since X has no nontrivial k^a-automorphism we deduce that diagram (8) commutes.

We now prove that $\alpha \in A$. We have just showed that θ induces an isomorphism between the covers $\chi_Q : E_Q \to X$ and $\chi_{\alpha(Q)} : E_{\alpha(Q)} \to X$. Therefore these two covers have the same ramification data: For every $1 \leq i \leq I$, the points P such that $f_i(P) = f_i(\psi(Q))$ and those such that $f_i(P) = f_i(\psi(\alpha(Q)))$ are the same. Thus

$$f_i(\psi(Q)) = f_i(\psi(\alpha(Q))) \quad \text{for all } i;$$

therefore $\psi(Q) = \psi(\alpha(Q))$, because the f_i generate $k^a(X)$ over k^a (Lemma 5.2). So $\psi = \psi \circ \alpha$, and $\alpha \in A$.

Diagram (8) implies that the map $\chi_{\alpha(Q)} \circ \theta : E_Q \to E_{\alpha(Q)}$ is equal to $(\mathrm{Id} \times \alpha) \circ \chi_Q$. And this is $\chi_{\alpha(Q)} \circ \alpha$ according to diagram (5). We set $\beta = \theta \circ \alpha^{-1}$ and we check that $\chi_{\alpha(Q)} \circ \beta = \chi_{\alpha(Q)}$. Since Q is generic and α surjective, we deduce that $\chi \circ \beta = \chi$ so $\beta \in B$. We conclude that $\theta = \beta\alpha \in A \times B$ as was to be shown. □

Remark. We have proved something slightly stronger than Lemma 5.5: the group of birational k^a-automorphisms of S is $A \times B$. We won't need this stronger result.

5c. ***Field of moduli and fields of definition of S.*** To prove Theorem 5.1, we have to show that the cover φ and the surface S share the same field of moduli and the same fields of definition. One can replace the cover φ by the cover ψ, since those two covers have the same field of moduli and fields of definition.

The construction of Section 5a yields a morphism of stacks $\mathbb{F} : \mathbb{M}_\psi \to \mathbb{M}_S$. To see that, let us consider an extension $l \subset k^a$ of k and let $\psi_l : Z_l \to X_l$ be an l-model of ψ. We follow the lines of the construction above, replacing ψ by ψ_l. Since the functions f_i are k-rational, the functions g_i lie in $l(X_l \times Z_l)$. Then the radical extension defined by the equations $y_i^{p_i} = g_i$ is a regular extension of $l(X_l \times Z_l)$. The normalization of $X_l \times Z_l$ in this extension is a surface S_l which is defined over l. This surface S_l is an l-model of S and the morphism $\mathbb{F}$ is defined on objects by $\mathbb{F}(\psi_l) = S_l$. Because functions f_i are k-rational, $\mathbb{F}$ is a morphism of stacks. According to Proposition 2.5, k is the field of moduli of S and every field of definition of ψ (or φ) is a field of definition of S.

Unfortunately, $\mathbb{F}$ is not fully faithful. As we did in Proposition 3.1, we use Proposition 2.7 to construct a morphism the other way around. The group $\mathrm{Aut}_{k^a}(S)$ is a normal subgroup of $\mathrm{Aut}_k(S)$. The action of $\mathrm{Aut}_k(S)$ on $\mathrm{Aut}_{k^a}(S)$ stabilizes the unique subgroup of order Π, which is nothing but $B = \mathrm{Aut}_{k^a}(\chi)$. Let U be the mark of the cover ψ. This is the complementary set of the branch locus of the quotient map $\chi : S \to X \times Z$. According to Proposition 2.7, taking the complementary set of the branch locus of a quotient map defines a morphism of stacks $\mathbb{G} : \mathbb{M}_S \to \mathbb{M}_U$. Therefore, every field of definition of S is a field of definition of the mark of ψ and then also a field of definition of ψ by Lemma 4.3. Indeed the proof of this lemma provides a morphism from $\mathbb{M}_U$ to $\mathbb{M}_\psi$ and the proof of Proposition 3.1 provides a morphism from $\mathbb{M}_\psi$ to $\mathbb{M}_\varphi$.

6. Curves

In this section we prove Theorem 1.2. We shall make use of the surface S constructed in Section 5, so we keep the notation there. We know that S has field of moduli k and the same fields of definition as the initial cover $\varphi : Y \to X$ (or equivalently $\psi : Z \to X$). The main idea is to draw on S a singular (but stable) curve inheriting the field of moduli and fields of definition of S, and then to deform it to obtain a smooth projective curve.

6a. ***Two stable curves.*** In Section 5a, we have constructed a cover $\chi : S \to X \times Z$ strongly ramified along the graph of $\psi : Z \to X$. For any point P on X, we call F_P the inverse image of $P \times Z$ by χ and $\chi_P : F_P \to P \times Z$ the corestriction of χ to $P \times Z$. We call Γ the union of the supports of all divisors of the functions g_i of Lemma 5.3. It contains the ramification locus of the cover χ.

Lemma 6.1. *There exist two nonconstant k-rational functions $f, g \in k^a(X)$ such that*

(1) *the divisor $((f)_0 + (f)_\infty) \times Z$ crosses Γ transversally*;

(2) *the divisor* $X \times ((g \circ \psi)_0 + (g \circ \psi)_\infty)$ *crosses* $\Gamma \cup [((f)_0 + (f)_\infty) \times Z]$ *transversally*;

(3) *any* k^a*-automorphism of* Z *that stabilizes the fiber* $(g \circ \psi)_0$ *is an automorphism of the cover* ψ (*note that the preceding condition implies that this fiber is simple*);

(4) *for any zero* P *of* f, *the cover* $g \circ \psi \circ \chi_P : F_P \to \mathbf{P}^1$ *has no automorphism other than the elements of* $A \times B$. *So* $\mathrm{Aut}_{k^a}(g \circ \psi \circ \chi_P) = \mathrm{Aut}_{k^a}(\psi \circ \chi_P) = A \times B$.

Proof. Let $f \in k^a(X)$ be a k-rational nonconstant function. We apply Lemma 7.2 to k, X, Z, Γ and f. We deduce that there exist two distinct scalars x and y in k^a such that $(f)_x \times Z$ and $(f)_y \times Z$ cross transversally Γ. We can even choose x and y in k and such that for every point P in $f^{-1}(x)$ or $f^{-1}(y)$, the fiber of every function $f_i \in k(X)$ above $f_i(P)$ does not meet the singular values of ψ, that is,

$$f_i^{-1}(f_i(P)) \cap \{\text{singular values of } \psi\} = \varnothing \quad \text{for all } P \in f^{-1}(x) \cup f^{-1}(y). \tag{9}$$

We replace f by $(f-x)/(f-y)$ and the first condition is fulfilled.

Now, for every zero P of f, we see that F_P is smooth and geometrically integral, because $(f)_0 \times Z$ crosses transversally the ramification locus Γ of χ. We now prove that

$$\mathrm{Aut}_{k^a}(\psi \circ \chi_P) = A \times B.$$

Indeed the function field $k^a(F_P)$ is the compositum

$$\begin{tikzcd}
 & k^a(F_P) \arrow[dl, dash] \arrow[dd, dash] \arrow[dr, dash] & \\
k^a(Z) \arrow[dr, dash, "pd"'] & & k^a(X') \overset{\text{def.}}{=} k^a\big((f_i - f_i(P))^{1/p_i}, 1 \le i \le I\big) \\
 & k^a(X) \arrow[ur, dash, "\prod_i p_i"'] &
\end{tikzcd}$$

where $X' \to X$ is an abelian cover with Galois group $B = \prod_{i=1}^{I} \mathbf{Z}/p_i\mathbf{Z}$. The $k^a(X)$-extensions $k^a(Z)$ and $k^a(X')$ are linearly disjoint (their degrees are coprime and one of them is Galois) and condition (9) implies that the extension $k^a(Z)/k^a(X)$ is not ramified above the zeros of the functions $f_i - f_i(P)$.

Now, any subcover of $X' \to X$ is ramified above the zeros of at least one of the functions $f_i - f_i(P)$. The same is true for any subcover of $F_P \to Z$. We deduce that $Z \to X$ is the maximal subcover of $F_P \to X$ that is not ramified above the zeros of the functions $f_i - f_i(P)$. Therefore any $k^a(X)$-automorphism of $k^a(F_P)$ stabilizes $k^a(Z)$. Thus

$$\mathrm{Aut}_{k^a(X)}(k^a(F_P)) = \mathrm{Aut}_{k^a(X)}(k^a(Z)) \times \mathrm{Aut}_{k^a(X)}(k^a(X')),$$

as was to be shown.

Next we look for a function g in $k(X)$ such that $g \circ \psi$ has no k^a-automorphism but elements of A and, for every zero P of f, the cover $g \circ \psi \circ \chi_P$ has no k^a-automorphism but elements of $\mathrm{Aut}_{k^a}(\psi \circ \chi_P) = A \times B$. According to Lemma 7.4, the functions in $k(X)$ that do not fulfill all these conditions lie in a finite union of strict sub-k-algebras. Therefore there exists such a function g.

According to Lemma 7.2, the scalars x in k such that $(g \circ \psi)_x$ does not cross $\Gamma \cup [((f)_0 + (f)_\infty) \times Z]$ transversally are finitely many. According to Lemma 7.3, the x in k such that $(g \circ \psi)_x$ has a k^a-automorphism not in A are finitely many. Therefore there exist two distinct scalars x and y in k such that $(g \circ \psi)_x$ and $(g \circ \psi)_y$ cross $\Gamma \cup [((f)_0 + (f)_\infty) \times Z]$ transversally and $(g \circ \psi)_x$ has no automorphism but those in A. We replace g by $(g - x)/(g - y)$ and the last three conditions are satisfied. □

The curves C_0 and D_0. Let C_0 be the curve on $X \times Z$ with equation

$$f(P) \times g \circ \psi(Q) = 0.$$

Let D_0 be the inverse image of C_0 by χ. These are singular curves over k^a. The two following lemmas are concerned with the stability and the automorphism groups of these two curves.

Lemma 6.2. *The curve C_0 is stable and* $\mathrm{Aut}_{k^a}(C_0) \simeq A$.

Proof. The curve C_0 is geometrically reduced because the zeros of f and $g \circ \psi$ are simple. The singular points on C_0 are the couples (P, Q) on $X \times Z$ such that $f(P) = g \circ \psi(Q) = 0$. These are ordinary double points. Therefore C_0 is semistable. It is geometrically connected also. Its irreducible components are isomorphic to X or Z. So they all have genus ≥ 2. Therefore C_0 is a stable curve.

We now prove that the group of k^a-automorphisms of C_0 is the group A of k^a-automorphisms of ψ. It is clear that A is included in $\mathrm{Aut}_{k^a}(C_0)$. Conversely, let θ be a k^a-automorphism of C_0. Then θ permutes the irreducible components of C_0. Some of these components are isomorphic to X, and the others are isomorphic to Z. Since X and Z are not k^a-isomorphic, θ stabilizes the two subsets of components. If we restrict θ to a component isomorphic to X and then compose with the projection on X, we obtain a nonconstant k^a-morphism from X to itself. This morphism must be the identity because X has no nontrivial k^a-automorphism. Therefore θ stabilizes each component isomorphic to Z. The singular points on such a component are the zeros of $g \circ \psi$. The set of these zeros is stabilized by no k^a-automorphism of Z but those of ψ by (3) of Lemma 6.1. So the restriction of θ to any component isomorphic to Z is in A. Composing θ with a well chosen element in A, we may assume that θ is trivial on one component isomorphic to Z. Therefore θ stabilizes every component isomorphic to X. Since these components have no nontrivial automorphism, θ acts trivially on them. Now let $P \times Z$ be a component of C_0 isomorphic to Z. The restriction of θ to it is an automorphism that fixes the singular points. These points are the zeros of $g \circ \psi$. So the restriction of θ to $P \times Z$ is in A. Since A acts faithfully on the set of zeros of $g \circ \psi$, we deduce that θ acts trivially on every component isomorphic to Z. □

Examining the full group of k^a-automorphisms of D_0 seems difficult to us. We restrict our attention to the subgroup $\mathrm{Aut}_{k^a}^{\mathrm{adm.}}(D_0)$ of *admissible* automorphisms. We first explain what we mean by an admissible automorphism.

We first note that the components of D_0 are of two different kinds. Some of them are covers of some $X \times Q$ where Q is a zero of $g \circ \psi$. We denote such a component by E_Q. The other components are covers of some $P \times Z$ where P is a k^a-zero of f. Such a component is denoted by F_P. We call $\chi_P : F_P \to P \times Z$ and $\chi_Q : E_Q \to X \times Q$ the restrictions of χ to components of D_0. Now let T be a singular point on D_0 such that $\chi(T) = (P, Q)$. So T lies in the intersection of E_Q and F_P. The point on E_Q corresponding to T is denoted U. The point on F_P corresponding to T is denoted V. So $\chi_Q(U) = P$ and $\chi_P(V) = Q$. Condition (2) in Lemma 6.1 implies that $f \circ \chi_Q$ is a uniformizing parameter for E_Q at U; likewise, condition (1) implies that $g \circ \psi \circ \chi_P$ is a uniformizing parameter for E_Q at F_P at V. Let θ be an automorphism of D_0 and let $T' = (U', V')$ be the image of $T = (U, V)$ by θ. We write $\chi(T') = (P', Q')$. We observe that $f \circ \chi_{Q'} \circ \theta$ is a uniformizing parameter for E_Q at U and $g \circ \psi \circ \chi_{P'} \circ \theta$ is a uniformizing parameter for F_P at V.

We say that θ is an *admissible automorphism* of D_0 if for every singular point T of D_0, we have

$$\frac{f \circ \chi_{Q'} \circ \theta}{f \circ \chi_Q}(U) \times \frac{g \circ \psi \circ \chi_{P'} \circ \theta}{g \circ \psi \circ \chi_P}(V) = 1, \tag{10}$$

where $\chi(T) = (P, Q)$ and $\chi(\theta(T)) = (P', Q')$. The justification for this definition is given in Section 6b. Admissible automorphisms form a subgroup of the group of k^a-automorphisms of D_0.

Lemma 6.3. *The curve D_0 is stable and* $\mathrm{Aut}_{k^a}^{\mathrm{adm.}}(D_0) \simeq A \times B$.

Proof. It is clear that $A \times B$ acts faithfully on D_0, and the corresponding automorphisms are admissible.

We can now prove that the curve $D_0 \subset S$ is a stable curve. Points (1) and (2) of Lemma 6.1 imply that the ramification locus Γ of χ does not contain any singular points of C_0. Therefore every singular point on C_0 gives rise to $\deg(\chi)$ singular points on D_0. These are all ordinary double points. To prove that D_0 is connected, we observe that the function g_i restricted to any irreducible component of C_0 is not a p_i-th power because none of the functions $f_i - \lambda$, $\lambda \in k^a$ is a p_i-th power (and the $f_i \circ \psi - \lambda$ are not either) as shown in Lemma 5.2. Also the irreducible components of D_0 correspond bijectively to those of C_0.

Now let us prove that $\mathrm{Aut}_{k^a}^{\mathrm{adm.}}(D_0) \simeq A \times B$. The components F_P and E_Q have different genera. Therefore any k^a-automorphism θ of D_0 stabilizes the set of all components F_P (and also the set of all E_Q).

Let Q and Q' be two k^a-zeros of $g \circ \psi$ such that $\theta(E_Q) = E_{Q'}$. As in the proof of Lemma 5.5, we notice that the image of E_Q in the product $X \times X$ by the morphism $\chi_Q \times \chi_{Q'} \circ \theta$ has an arithmetic genus no more than $1 + 2(g(X) - 1)\Pi + \Pi^2$. Again, this implies that this image is k^a-isomorphic to X (otherwise, this image would have geometric genus bigger than $1 + 2(g(X) - 1)\Pi + \Pi^2$ by the Hurwitz formula). Since X has no k^a-automorphism, we deduce that θ induces an isomorphism of covers between the restrictions $\chi_Q : E_Q \to X$ and $\chi_{Q'} : E_{Q'} \to X$ of χ. Thus

$$\chi_Q = \chi_{Q'} \circ \theta. \tag{11}$$

This implies that θ stabilizes every component F_P, where P is any k^a-zero of f. Indeed, let $T = (U, V) \in E_Q \cap F_P$ be a singular point, where P is a k^a-zero of f and Q is a k^a-zero of $g \circ \psi$. We have $\chi(T) = (P, Q) \in X \times Z$ and there exist $P' \in X(k^a)$ and $Q' \in Z(k^a)$ such that $\theta(T) \in F_{P'} \cap E_{Q'}$. We deduce from Equation (11) that

$$P' = \chi_{Q'} \circ \theta(T) = \chi_Q(T) = P.$$

We conclude that $P = P'$ and $\theta(F_P) = F_P$.

Now, we deduce from formulas (10) and (11) that

$$\frac{g \circ \psi \circ \chi_P \circ \theta}{g \circ \psi \circ \chi_P}(V) = 1. \tag{12}$$

Denote by θ_P the restriction of θ to F_P. This is an automorphism of F_P. We prove θ_P is the restriction to F_P of an element of $A \times B$. To this end, we introduce the function $h_P = g \circ \psi \circ \chi_P \in k^a(F_P)$. The degree of h_P is $\deg(g) \times pd \times \Pi$ and its zeros are all simple. These zeros are the intersection points between F_P and the other components of D_0. Since θ_P permutes these zeros, the functions $h_P \circ \theta_P$ and h_P have the same divisor of zeros. Therefore the only possible poles of the function $h_P/(h_P \circ \theta_P) - 1$ are the poles of h_P, and its degree is $\leq h_P$. But according to (12), the zeros of h_P are also zeros of $h_P/(h_P \circ \theta_P) - 1$. So if the function $h_P/(h_P \circ \theta_P) - 1$ is nonzero, it has the same divisor as h_P. Therefore there exists a constant $c \in k^a$ such that

$$\frac{h_P}{h_P \circ \theta_P} - 1 = ch_P \qquad \text{or equivalently} \qquad \frac{1}{h_P \circ \theta_P} = \frac{1}{h_P} + c.$$

Since θ_P has finite order e and k^a has characteristic zero, we deduce that $ce = 0$, and then $c = 0$. Then $h_P \circ \theta_P = h_P$ and θ_P is an automorphism of the cover

$h_P = g \circ \psi \circ \chi_P : F_P \to \mathbf{P}^1$. According to point (4) of Lemma 6.1, we deduce that θ_P is the restriction to F_P of an element in $A \times B$. We replace θ by θ composed with the inverse of this element. So we can now assume that θ acts trivially on F_P for some P. In particular θ fixes every singular point on F_P. So θ stabilizes every component E_Q. The restriction θ_Q of θ to E_Q is an automorphism of χ_Q according to (11). Further θ_Q fixes a point (and every point) in the unramified fiber above P of the Galois cover $\chi_Q : E_Q \to X$. Therefore θ_Q is the identity. We have proved that θ is trivial on every component E_Q.

To conclude, we prove that θ is also trivial on the components $F_{P'}$ for every zero P' of f. Let us recall that we have already assumed this to be true for one of these zeros. We call $\theta_{P'}$ the restriction of θ to $F_{P'}$. We have already proved that $\theta_{P'}$ is the restriction of an element in $A \times B$. Further $\theta_{P'}$ fixes all the singular points of D_0 lying on $F_{P'}$. These points are the zeros of $g \circ \psi \circ \chi_{P'}$. Therefore, it only remains to be proved that the action of $A \times B$ on the set of zeros of $g \circ \psi \circ \chi_{P'}$ is free. This is the case for elements in B because the zeros of $g \circ \psi$ are, by hypothesis, unramified in the Galois cover $\chi_{P'} : F_{P'} \to Z$. This is also true for elements in $A \times B$ because the action of A on the set of zeros of $g \circ \psi$ is free. □

6b. *Deformations.* We now deform the two stable curves C_0 and D_0. If $t \in k^a$ is a scalar, we call $C_t \subset W = X \times Z$ the curve with equation $f(P) \times g(\psi(Q)) = t$. We call D_t the inverse image of C_t by χ. In this and the next paragraph, we prove that for almost all scalars t in k, the curve D_t is smooth, geometrically integral, with k^a-automorphism group equal to $A \times B$, and with the same field of moduli and the same fields of definition as the original cover φ. To this end, we would like to consider the families $(C_t)_t$ and $(D_t)_t$ as fibrations above $\mathbf{P}^1$. We should be careful however: the family $(C_t)_t$ has base points. So we blow up $W = X \times Z$ along

$$\Delta = ((f)_\infty \times (g \circ \psi)_0) \cup ((f)_0 \times (g \circ \psi)_\infty).$$

Note that Δ is the union of $2 \times \deg(f) \times \deg(g \circ \psi)$ simple geometric points. We call $W_{\infty,\infty} \subset W = X \times Z$ the complementary open set of

$$((f)_\infty \times Z) \cup (X \times (g \circ \psi)_\infty) \quad \text{in } X \times Z.$$

We similarly define $W_{0,0}$, $W_{0,\infty}$, $W_{\infty,0}$. These four open sets cover $X \times Z$.

Let $\mathbf{P}^1 = \mathrm{Proj}(k^a[T_0, T_1])$ be the projective line over k^a. We set $F = 1/f$ and $G = 1/g$. Let $C_{\infty,0} \subset W_{\infty,0} \times \mathbf{P}^1$ be the set of $(P, Q, [T_0 : T_1])$ such that $f(P)T_0 = G(\psi(Q))T_1$. Let $C_{0,\infty} \subset W_{0,\infty} \times \mathbf{P}^1$ be the set of $(P, Q, [T_0 : T_1])$ such that $g(\psi(Q))T_0 = F(P)T_1$. Let $C_{\infty,\infty} \subset W_{\infty,\infty} \times \mathbf{P}^1$ be the set of $(P, Q, [T_0 : T_1])$ such that $f(P)g(\psi(Q))T_0 = T_1$. Let $C_{0,0} \subset W_{0,0} \times \mathbf{P}^1$ be the set of $(P, Q, [T_0 : T_1])$ such that $T_0 = F(P)G(\psi(Q))T_1$. We glue together these four algebraic varieties and obtain a variety $C \subset W \times \mathbf{P}^1$. Let $\pi_W : C \to W$ be the projection on the first factor and let $\pi_C : C \to \mathbf{P}^1$ be the projection on $\mathbf{P}^1$. This is a flat, projective, surjective morphism.

Let $D \subset S \times \mathbf{P}^1$ be the inverse image of C by $\chi \times \mathrm{Id}$ where $\mathrm{Id} : \mathbf{P}^1 \to \mathbf{P}^1$ is the identity. This is the blow up of S along $\chi^{-1}(\Delta)$. Note that $\chi^{-1}(\Delta)$ is the union of $\deg(\chi) \times \deg(f) \times \deg(g \circ \psi)$ simple geometrical points because χ is unramified above Δ. Actually, D is the normalization of C in $k^a(S \times \mathbf{P}^1)$. We denote by $\chi : D \to C$ the corresponding morphism. We call $\pi_S : D \to S$ the projection on the first factor. We call $\pi_D : D \to \mathbf{P}^1$ the projection on the second factor. This is the composed morphism $\pi_D = \pi_C \circ \chi$. This is a flat, proper and surjective morphism.

Let $\mathbf{A}^1 \subset \mathbf{P}^1$ be the spectrum of $k^a[T]$, where $T = T_1/T_0$. Using the function T we identify $\mathbf{P}^1(k^a)$ and $k^a \cup \{\infty\}$. If t is a point on $\mathbf{P}^1(k^a)$ we denote by C_t the fiber of π_C above t and by D_t the fiber of π_D above t. The restriction of π_W to C_t is a closed immersion. So we can see C_t as a curve on $W = X \times Z$. Similarly, the restriction of π_S to D_t is a closed immersion. So we can see D_t as a curve drawn on S. In particular, the fiber of π_C at 0 is isomorphic by π_W to the stable curve C_0 introduced in Section 6a. Similarly, the fiber of π_D at 0 is isomorphic by π_S to the stable curve D_0 introduced in Section 6a.

Let us call C_η the generic fiber of π_C and D_η the generic fiber of π_D.

We prove that the curve C_η over $k^a(\mathbf{P}^1)$ is geometrically connected and that for almost every $t \in \mathbf{P}^1(k^a)$, the curve C_t over k^a is connected. According to Stein's factorization theorem [Liu 2002, Chapter 5, Exercise 3.11], we can factor $\pi_C : C \to \mathbf{P}^1$ as $\pi_f \circ \pi_c$, where π_c has geometrically connected fibers and π_f is finite and dominant. The fiber of π_f above 0 is trivial because C_0 is connected and reduced. Therefore the degree of π_f is 1 according to [Liu 2002, Chapter 5, Exercise 1.25]. Therefore π_f is an isomorphism above a nonempty open set of $\mathbf{P}^1$. The generic fiber C_η is geometrically connected over $k^a(\mathbf{P}^1)$ and for almost all $t \in \mathbf{P}^1(k^a)$ the curve C_t over k^a is connected.

We now prove that C_η is smooth (and therefore geometrically integral). Indeed, it is smooth outside the points $(P, Q) \in C_\eta \subset X \times Z$, where $df(P) = 0$ and $d(g \circ \psi)(Q) = 0$. Such points are defined over k^a. Therefore $f(P) \times g(\psi(Q))$ cannot take the transcendental value T at these points.

The ramification locus $\Gamma \subset W$ of χ cuts the fiber C_0 transversally. Therefore it cuts the generic fiber C_η transversally. So D_η is smooth and geometrically integral. Thus for almost every $t \in k^a$ the fibers C_t and D_t are smooth and integral.

We now compute $\mathrm{Aut}_{k(\mathbf{P}^1)^s}(D_\eta)$. Let $R = k^a[[T]]$ be the completed local ring at the point $T = 0$ of $\mathbf{P}^1$. The curve $\hat{D} = D \times_{\mathbf{P}^1} \mathrm{Spec}(R)$ is stable over $\mathrm{Spec}(R)$. According to [Liu 2002, Chapter 10, Proposition 3.38, Remark 3.39] the functor "automorphism group" $t \mapsto \mathrm{Aut}_t(\hat{D}_t)$ is representable by a finite unramified scheme over Spec R and the specialization morphism $\mathrm{Aut}_{k^a((T))}(\hat{D}_\eta) \to \mathrm{Aut}_{k^a}(D_0)$ is injective. According to Lemma 7.6, the image of this morphism is included in the subgroup of admissible k^a-automorphisms of D_0. Since Spec R has no unramified cover, we can deduce

$$A \times B \subset \mathrm{Aut}_{k^a(\mathbf{P}^1)^s}(D_\eta) \subset \mathrm{Aut}_{k^a((T))}(\hat{D}_\eta) \subset \mathrm{Aut}_{k^a}^{\mathrm{adm.}}(D_0).$$

We know that the rightmost group is equal to $A \times B$. So

$$\mathrm{Aut}_{k^a(\mathbf{P}^1)^s}(D_\eta) = A \times B$$

as was intended to be proved.

6c. ***Fields of moduli and fields of definition of fibers.*** We just have proved that for almost all $t \in \mathbf{A}^1(k)$, the fiber D_t is smooth and geometrically integral. Using Lemma 7.7 on the specialization of the automorphism group we deduce that for almost all $t \in \mathbf{A}^1(k)$, the group of k^a-automorphisms of the fiber D_t is isomorphic to the group of $k(\mathbf{A}^1)^s$-automorphisms of the generic fiber. Since the latter group is isomorphic to the automorphism group $A \times B$ of the surface S, we deduce that, for almost all t, the restriction map is an isomorphism:

$$\mathrm{Aut}_{k^a}(S) \xrightarrow{\simeq} \mathrm{Aut}_{k^a}(D_t). \tag{13}$$

Now let $t \in k$ be such that D_t is smooth and geometrically integral and such that $\mathrm{Aut}_{k^a}(D_t) = A \times B$. We call $\pi_t : D_t \to S$ the corresponding embedding. We construct a functor $\mathbb{F}_t : \mathbb{M}_S \to \mathbb{M}_{\pi_t}$. We first define the image of an object by $\mathbb{F}_t$. Let $l \subset k^a$ be a finite extension of k and S_l an l-model of S. Using the functor $\mathbb{M}_S \to \mathbb{M}_U$ given in Section 5c followed by the functor $\mathbb{M}_U \to \mathbb{M}_\psi$ in the proof of Lemma 4.3, we obtain an l model $\psi_l : Z_l \to X_l$ of the cover ψ, where $X_l = X_k \times_k l$ and Z_l is an l-model of Z. There is also an abelian cover $\chi_l : S_l \to X_l \times Z_l$ which is well defined up to an automorphism of S_l. We denote by $C_{t,l}$ the curve on $X_l \times Z_l$ with the equation $f \cdot g \circ \psi_l - t = 0$. Let $D_{t,l}$ be the inverse image of $C_{t,l}$ by χ_l. Let $\pi_{t,l} : D_{t,l} \hookrightarrow S_l$ be the inclusion map. The image of the object S_l by the functor $\mathbb{F}_t$ is defined to be $\pi_{t,l}$. We still need to define the image of a morphism by the functor $\mathbb{F}_t$. Let l' be another finite extension of k and let $\sigma : l \to l'$ be a k-homomorphism. Let $S'_{l'}$ be an l'-model of S and let $\alpha : S_l \to S'_{l'}$ be a morphism above Spec(σ). We call $\pi'_{t,l'} : D'_{t,l'} \hookrightarrow S'_{l'}$ the image by $\mathbb{F}_t$ of $S'_{l'}$. Then α maps $D_{t,l}$ to $D'_{t,l'}$. We denote by β the restriction of α to $D_{t,l}$. The image of α by $\mathbb{F}_t$ is defined to be the morphism (α, β) from $\pi_{t,l}$ to $\pi'_{t,l'}$. If we compose $\mathbb{F}_t : \mathbb{M}_S \to \mathbb{M}_{\pi_t}$ with the forgetful functor $\mathbb{M}_{\pi_t} \to \mathbb{M}_{D_t}$, we obtain a cartesian functor $\mathbb{G}_t : \mathbb{M}_S \to \mathbb{M}_{D_t}$. Further, identity (13) implies that the functor $\mathbb{G}_t$ is fully faithful. So, by Proposition 2.5, both S and D_t have k as field of moduli and they have the same fields of definition. In view of Section 5c, D_t, ψ and φ also share the same fields of definition. Theorem 1.2 is proved.

7. Six lemmas about curves and surfaces

In this section we state and prove six lemmas that are needed in the proof of Theorem 1.2.

7a. *About curves and products of two curves.*

Lemma 7.1. *Let k be a algebraically closed field. Let X and Y be two projective, smooth and integral curves over k. Let β be the genus of X and let γ be the genus of Y. We fix a geometric point P on X and a geometric point Q on Y. We identify the curves X and $X \times Q$ and the curves Y and $P \times Y$. Let D be a divisor on $X \times Y$ of bidegree (b, c), that is, $b = X \cdot D$ and $c = Y \cdot D$. The virtual arithmetic genus π of D is at most $1 + bc + c(\beta - 1) + b(\gamma - 1)$. When $b = c$ this bound reads $1 + 2b(\beta - 1) + b^2$.*

Proof. The algebraic equivalence class of the canonical divisor on $X \times Y$ is $K = 2(\beta - 1)Y + 2(\gamma - 1)X$. The virtual arithmetic genus π of D, as defined in [Hartshorne 1977, Exercise V-1.3], is $\pi = \frac{1}{2}D \cdot (D + K) + 1$. Thus $\pi = D \cdot (D + 2(\beta - 1)Y + 2(\gamma - 1)X)/2 + 1$. We deduce from Castelnuovo's and Severi's inequality (see [Hartshorne 1977, Exercise V-1.9]) that $D \cdot D \le 2bc$. This finishes the proof of the lemma. □

Lemma 7.2. *Let k be an algebraically closed field. Let X and Y be two projective, smooth, integral curves over k. Let D be an effective divisor without multiplicity on the surface $X \times Y$. Let $f \in k(X)$ be a nonconstant function. For all but finitely many scalars x in k, the divisor $(f)_x \times Y$ crosses D transversally, where $(f)_x$ is the positive part of the divisor of $f - x$.*

Proof. We call $p_X : X \times Y \to X$ the projection on the first factor. Let E be the set of points in $X(k)$ such that at least one of the following condition holds: $p_X^{-1}(P)$ contains a singular point on D, or $p_X^{-1}(P)$ contains a ramified point of the morphism $p_X : D \to X$, or the fiber $p_X^{-1}(P)$ is contained in D. The set E is finite. For all $x \in k$ but finitely many, the fiber $f^{-1}(x)$ avoids E and is simple. □

Lemma 7.3. *Let k be an algebraically closed field. Let X be a projective, smooth, integral curve over k with genus at least 2. Let $f \in k(X)$ be a nonconstant function. We denote by G the group of k-automorphisms of f. This is the set of all k-automorphisms θ of X such that $f \circ \theta = f$. For any $x \in \mathbf{P}^1(k)$, we write $(f)_x = f^{-1}(x)$ for the fiber above x, and G_x for the group of k-automorphisms of X that stabilize the set of k-points of $(f)_x$. For all x in $\mathbf{P}^1(k)$ but finitely many, we have $G_x = G$.*

Proof. The group $H = \mathrm{Aut}_k(X)$ of k-automorphisms of X is finite. Let θ be an automorphism in $H \setminus G$ and let $x \in \mathbf{P}^1(k)$. Assume that the k-points in $(f)_x$ are permuted by θ. Let P be one of them. Then $f \circ \theta(P) = f(P) = x$. So P is a zero of the nonzero function $f \circ \theta - f$. For each θ there are finitely many such zeros. And the θ are finitely many. So the images by f of such P are finitely many also. □

Lemma 7.4. *Let k be a field. Let X_k be a projective, smooth, geometrically integral curve over k. Set $X = X_k \times_k k^a$ and assume that X has genus at least 2. Let Y be a projective, smooth, integral curve over k^a and let $\varphi : Y \to X$ be a nonconstant k^a-cover. If f is any nonconstant function in $k^a(X)$, then* $\mathrm{Aut}(\varphi) \subset \mathrm{Aut}(f \circ \varphi)$. *Let $V \subset k(X_k)$ be the set of functions $f \in k(X_k)$ such that* $\mathrm{Aut}(\varphi) \neq \mathrm{Aut}(f \circ \varphi)$. *This set V is contained in a finite union of strict k-subalgebras of $k(X_k)$.*

Proof. The statement concerns the three function fields $k^a(f) \subset k^a(X) \subset k^a(Y)$, and the groups involved are the following ones:

$$\begin{cases} \mathrm{Aut}(\varphi) = \mathrm{Aut}_{k^a(X)}(k^a(Y)), \\ \mathrm{Aut}(f \circ \varphi) = \mathrm{Aut}_{k^a(f)}(k^a(Y)), \\ \mathrm{Aut}(Y) = \mathrm{Aut}_{k^a}(k^a(Y)), \end{cases} \quad \Rightarrow \quad \mathrm{Aut}(\varphi) \subset \mathrm{Aut}(f \circ \varphi) \subset \mathrm{Aut}(Y).$$

Now, the set V can be described as

$$V = \Big(\bigcup_{\theta \in \mathrm{Aut}(Y) \setminus \mathrm{Aut}(\varphi)} k^a(Y)^\theta \cap k^a(X) \Big) \cap k(X_k) = \bigcup_{\theta \in \mathrm{Aut}(Y) \setminus \mathrm{Aut}(\varphi)} k^a(Y)^\theta \cap k(X_k).$$

This is a union of sets indexed by elements in the finite set $\mathrm{Aut}(Y) \setminus \mathrm{Aut}(\varphi)$ (recall that $\mathrm{Aut}(Y)$ is finite because the genus of Y is at least 2). Since $\theta \notin \mathrm{Aut}(\varphi)$, each $k^a(Y)^\theta \cap k^a(X)$ is a strict subfield of $k^a(X)$ containing k^a. Therefore $k^a(Y)^\theta \cap k(X_k) \subsetneq k(X_k)$. □

7b. ***Deformation of an automorphism of a nodal curve.*** In this subsection we give a *necessary* condition for an automorphism of a nodal curve to extend to a given deformation of this curve. Let R be a complete discrete valuation ring. Let π be a uniformizing parameter and let k be the residue field. Assume that k is algebraically closed. Let D be a semistable curve over $\mathrm{Spec}(R)$. Let D_η be the generic fiber and D_0 the special fiber. Assume D_η is smooth over the fraction field of R. Let T be a singular point of D_0. According to [Liu 2002, Chapter 10, Corollary 3.22], the completion of the local ring of D at T takes the form

$$\hat{\mathcal{O}}_{D,T} = R[\![f, g]\!]/\langle fg - \pi^e \rangle,$$

where e is a positive integer. This integer is called the *thickness* of D at T. We also say that f and g form a coordinate system for D at T. If we reduce modulo π, we obtain the completion of the local ring of D_0 at T:

$$\hat{\mathcal{O}}_{D_0,T} = \hat{\mathcal{O}}_{D,T}/\langle \pi \rangle = k[\![\bar{f}, \bar{g}]\!]/\langle \bar{f}\bar{g} \rangle,$$

where $\overline{f} = f \bmod \pi$ and $\bar{g} = g \bmod \pi$. Because T is an ordinary double point, D_0 has two branches F and G at T. These correspond to the two irreducible components of the completion at T. Be careful that these two branches may lie on the same irreducible component of D_0. Anyway, the functions $\overline{f}$ and $\bar{g}$ are the uniformizing parameters of either branches. We call P and Q the points of F and G above T.

Now let T' be another singular point of D_0, and let f', g', e', F', and G' be the corresponding data. Let θ be an automorphism of D over R such that $\theta(T) = T'$ and $\theta(F) = F'$, $\theta(G) = G'$. The functions $f' \circ \theta$ and $g' \circ \theta$ form a coordinate system for D at T. So $e' = e$ and both $f' \circ \theta / f$ and $g' \circ \theta / g$ are units in $\hat{\mathbb{O}}_{D,T}$ (indeed, in either fraction, the numerator and denominator have the same Weil divisor). Since $f \times g = \pi^e = f' \circ \theta \times g' \circ \theta$, we have $\frac{f' \circ \theta}{f}(T) \times \frac{g' \circ \theta}{g}(T) = 1$. We reduce this identity modulo π and obtain the following identity, where the first factor is a function on F evaluated at P and the second is a function on G evaluated at Q:

$$\frac{\overline{f}' \circ \overline{\theta}}{\overline{f}}(P) \times \frac{\bar{g}' \circ \overline{\theta}}{\bar{g}}(Q) = 1. \tag{14}$$

Definition 7.5. Let R be a complete discrete valuation ring. Assume that the residue field k is algebraically closed. Let D be a semistable curve over $\mathrm{Spec}(R)$. The generic fiber of D is assumed to be smooth. Assume we are given a coordinate system at each singular point of the special fiber D_0. Let $\bar{\theta}$ be an automorphism of the special fiber D_0. We say that $\bar{\theta}$ is *admissible* in $D/\mathrm{Spec}(R)$ if for every singular point T of D_0, the image $\bar{\theta}(T)$ has the same thickness as T in D, and if equality (14) holds true.

We have just proved this:

Lemma 7.6. *With the notation of Definition 7.5, the set of automorphisms of D_0 that are admissible in $D/\mathrm{Spec}(R)$ is a subgroup of $\mathrm{Aut}_k(D_0)$. If θ is an automorphism of D over $\mathrm{Spec}(R)$, its reduction $\bar{\theta} = \theta \bmod \pi$ is an automorphism of D_0 and is admissible in $D/\mathrm{Spec}(R)$.*

One may compare this statement with [Wewers 1999, Theorem 3.1.1], where the deformation of morphisms between two distinct curves is studied.

Remark. The converse of Lemma 7.6 is not true. For example, consider the elliptic curve E with modular invariant $j = 0$ (or 1728). Every automorphism of E is admissible because there are no singular points on the curve (the condition in Definition 7.5 is empty). However, the only automorphisms that can be extended to the generic elliptic curve are the identity and the involution.

7c. ***Automorphisms of curves in a family.*** We state and prove a standard lemma about specialization of automorphism groups.

Lemma 7.7. *Let k be a field and let U be a smooth, geometrically integral curve over k. Let X be a quasiprojective, smooth, geometrically integral surface over k. Let $\pi : X \to U$ be a surjective, projective, smooth morphism of relative dimension* 1. *Assume that for any point x of U, the fiber X_x at x is geometrically integral. We call η the generic point of U and call $\overline{X}_\eta = X_\eta \times_{\mathrm{Spec}(k(U))} \mathrm{Spec}(k(U)^a)$ the generic fiber, seen as a curve over the algebraic closure of the function field of the basis U. We assume the genus of X_η is at least* 2.

There exists a nonempty open subset V of U over k such that for any geometric point $x \in V(k^a)$ the group of k^a-automorphisms of the fiber at x is equal to the group $\mathrm{Aut}_{k(U)^a}(\overline{X}_\eta)$ of automorphisms of $\overline{X}_\eta$.

The following proof was communicated to us by Qing Liu.

Proof. This is a consequence of a general result by Deligne and Mumford. Let $X \to S$ be a flat projective morphism over a noetherian scheme S. The functor $T \to \mathrm{Aut}_T(X_T)$ from the category of S-schemes to the category of groups is representable by a group scheme $\mathrm{Aut}_{X/S}$ over S. See [Kollár 1996, Exercise 1.10.2] for example. When $X \to S$ is a stable curve with genus at least 2, Deligne and Mumford [1969, Theorem 1.11] prove that the scheme $\mathrm{Aut}_{X/S}$ is finite and unramified over S. In our lemma, S is a smooth, geometrically integral curve U over k. Replacing S by a nonempty open subset, we may assume that $\mathrm{Aut}_{X/S}$ is finite étale over S. At the expense of a finite surjective base change $T \to S$, we may assume that the generic fiber of $\mathrm{Aut}_{X/S} \to S$ consists of rational points. So $\mathrm{Aut}_{X/S} \to S$ is now a disjoint union of étale sections and the fibers have constant degree. In particular, the fibers are constant and the specialization maps $\mathrm{Aut}_S(X) = \mathrm{Aut}_{X/S}(S) \to \mathrm{Aut}_s(X_s) = \mathrm{Aut}_{X/S}(k(s))$ are isomorphisms. □

References

[Coombes and Harbater 1985] K. Coombes and D. Harbater, "Hurwitz families and arithmetic Galois groups", *Duke Math. J.* **52**:4 (1985), 821–839. MR 87g:14012 Zbl 0601.14023

[Couveignes and Granboulan 1994] J.-M. Couveignes and L. Granboulan, "Dessins from a geometric point of view", pp. 79–113 in *The Grothendieck theory of dessins d'enfants* (Luminy, 1993), edited by L. Schneps, London Math. Soc. Lecture Note Ser. **200**, Cambridge Univ. Press, 1994. MR 96b:14015 Zbl 0835.14010

[Couveignes and Ros 2004] J.-M. Couveignes and N. Ros, "Des obstructions globales à la descente des revêtements", *Acta Arith.* **114**:4 (2004), 331–348. MR 2006h:11075 Zbl 1056.12002

[Dèbes and Emsalem 1999] P. Dèbes and M. Emsalem, "On fields of moduli of curves", *J. Algebra* **211**:1 (1999), 42–56. MR 99k:14044 Zbl 0934.14019

[Dèbes and Fried 1994] P. Dèbes and M. D. Fried, "Nonrigid constructions in Galois theory", *Pacific J. Math.* **163**:1 (1994), 81–122. MR 95c:12008 Zbl 0788.12001

[Deligne and Mumford 1969] P. Deligne and D. Mumford, "The irreducibility of the space of curves of given genus", *Inst. Hautes Études Sci. Publ. Math.* 36 (1969), 75–109. MR 41 #6850 Zbl 0181.48803

[Giraud 1964] J. Giraud, *Méthode de la descente*, Bull. Soc. Math. France Mém. **2**, 1964. MR 32 #7556

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

[Kollár 1996] J. Kollár, *Rational curves on algebraic varieties*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **32**, Springer, Berlin, 1996. MR 98c:14001 Zbl 0877.14012

[Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics **6**, Oxford University Press, 2002. MR 2003g:14001 Zbl 0996.14005

[Mestre 1991] J.-F. Mestre, "Construction de courbes de genre 2 à partir de leurs modules", pp. 313–334 in *Effective methods in algebraic geometry* (Castiglioncello, 1990), edited by T. Mora and C. Traverso, Progr. Math. **94**, Birkhäuser, Boston, MA, 1991. MR 92g:14022 Zbl 0752.14027

[Serre 1959] J.-P. Serre, *Groupes algébriques et corps de classes*, 2nd ed., Publications de l'institut de mathématique de l'université de Nancago **7**, Hermann, Paris, 1959. MR 21 #1973 Zbl 0097. 35604

[Silverman 1995] J. H. Silverman, "The field of definition for dynamical systems on $\mathbf{P}^1$", *Compositio Math.* **98**:3 (1995), 269–304. MR 96j:11090 Zbl 0849.11090

[Vistoli 2005] A. Vistoli, "Grothendieck topologies, fibered categories and descent theory", pp. 1–104 in *Fundamental algebraic geometry*, Math. Surveys Monogr. **123**, Amer. Math. Soc., Providence, RI, 2005. MR 2223406

[Weil 1956] A. Weil, "The field of definition of a variety", *Amer. J. Math.* **78** (1956), 509–524. MR 18,601a

[Wewers 1999] S. Wewers, "Deformation of tame admissible covers of curves", pp. 239–282 in *Aspects of Galois theory* (Gainesville, FL, 1996), edited by H. Völklein et al., London Math. Soc. Lecture Note Ser. **256**, Cambridge Univ. Press, 1999. MR 2001b:14048 Zbl 0995.14008

Communicated by Bjorn Poonen
Received 2009-09-11 Revised 2010-11-22 Accepted 2011-02-19

jean-marc.couveignes@math.u-bordeaux1.fr
Institut de Mathématiques de Bordeaux,
Université Bordeaux I et CNRS, 161, cours de la Libération,
F-33405 Talence, France
http://www.math.u-bordeaux1.fr/~couveign/

hallouin@univ-tlse2.fr
Département de mathématiques et informatique,
Université de Toulouse le Mirail, 5, allées Antonio Machado,
F-31058 Toulouse, France
http://www.math.univ-toulouse.fr/~hallouin/

ALGEBRA AND NUMBER THEORY 5:4(2011)

Specializations of elliptic surfaces, and divisibility in the Mordell–Weil group

Patrick Ingram

Let $\mathscr{E} \to C$ be an elliptic surface defined over a number field k, let $P : C \to \mathscr{E}$ be a section, and let ℓ be a rational prime. We bound the number of points of low algebraic degree in the ℓ-division hull of P at the fibre $\mathscr{E}_t$. Specifically, for $t \in C(\bar{k})$ with $[k(t) : k] \leq B_1$ such that $\mathscr{E}_t$ is nonsingular, we obtain a bound on the number of $Q \in \mathscr{E}_t(\bar{k})$ such that $[k(Q) : k] \leq B_2$, and such that $\ell^n Q = P_t$ for some $n \geq 1$. This bound depends on $\mathscr{E}$, P, ℓ, B_1, and B_2, but is independent of t.

1. Introduction

One of the central problems in the study of elliptic surfaces is to determine the extent to which the geometry of the surface determines the arithmetic of its fibres. Let $\mathscr{E} \to C$ be an elliptic surface defined over a number field k. Then if the fibre $\mathscr{E}_t$ above $t \in C(k)$ is nonsingular, there is a homomorphism

$$\sigma_t : \mathscr{E}(C) \to \mathscr{E}_t(k),$$

where $\mathscr{E}(C)$ is the group of sections $P : C \to \mathscr{E}$ (we include the existence of a section in our definition of an elliptic surface). By a theorem of Silverman [1994, Chapter III, Theorem 11.4], this map is injective for all but finitely many $t \in C(k)$. The map σ_t is not, in general, surjective, since the rank of $\mathscr{E}_t(k)$ may exceed that of $\mathscr{E}(C)$ (see [Salgado 2009]), but another result of Silverman shows that for elliptic surfaces over $\mathbb{P}^1_{\mathbb{Q}}$, there are infinitely many fibres for which the image of σ_t is, at least, not divisible in $\mathscr{E}_t(\mathbb{Q})$.

Theorem [Silverman 1985]. *Let $\mathscr{E} \to \mathbb{P}^1$ be an elliptic surface defined over $\mathbb{Q}$, with nonconstant j-invariant. Then there exist infinitely many $t \in \mathbb{P}^1(\mathbb{Q})$ such that the quotient $\mathscr{E}_t(\mathbb{Q})/\sigma_t(\mathscr{E}(\mathbb{P}^1))$ is torsion-free.*

We cannot replace ‘there exist infinitely many’ with ‘for all but finitely many’ in Silverman’s result, since the elliptic surface defined over $\mathbb{P}^1_{\mathbb{Q}}$ by $\mathscr{E} : y^2 = x^3 - tx + t$

MSC2000: primary 11G05; secondary 14J27, 14G05.

Keywords: elliptic surface, specialization theorem.

has no section of order two, but $(\eta, 0) \in \mathscr{E}_t(\mathbb{Q})$ is a point of order two whenever $t = \eta^3/(\eta - 1)$.

Saying that $\mathscr{E}_t(\mathbb{Q})/\sigma_t(\mathscr{E}(\mathbb{P}^1))$ is torsion-free amounts to saying that for any prime ℓ and any section $P : \mathbb{P}^1 \to \mathscr{E}$, the specialization $P_t = \sigma_t(P)$ is divisible by ℓ in the Mordell–Weil group $\mathscr{E}_t(\mathbb{Q})$ only if P is already divisible by ℓ in the group $\mathscr{E}(\mathbb{P}^1)$ of sections. Since Silverman's result only treats infinitely many of the fibres, however, it is still conceivable that one could construct a section $P : \mathbb{P}^1 \to \mathscr{E}$ that is not divisible by a prime ℓ, but whose specializations $P_t \in \mathscr{E}_t(\mathbb{Q})$ are divisible by arbitrarily large powers of ℓ (as t varies).

Our main result is that this sort of construction is not possible, and we prove this for elliptic surfaces $\mathscr{E} \to C$ over arbitrary base curves, defined over a number field k. In fact, even if we are allowed to consider fibres and points of bounded algebraic degree over k, the extent to which the specializations of a given section $P : \mathscr{E} \to C$ might be ℓ-divisible is limited.

Theorem 1. *Let $k/\mathbb{Q}$ be a number field, let $\mathscr{E} \to C$ be an elliptic surface, with nonconstant j-invariant, over the smooth projective curve C, and let $P : C \to \mathscr{E}$ be a section (all defined over k). Then for any $B_1, B_2 \geq 1$, there is a value $M(B_1, B_2)$ such that*

$$\#\bigl\{Q \in \mathscr{E}_t(\bar{k}) : [k(Q) : k] \leq B_1 \text{ and } \ell^n Q = P_t \text{ for some } n \geq 1\bigr\} \leq M,$$

as $t \in C(\bar{k})$ varies over the places of good reduction for $\mathscr{E}$ with $[k(t) : k] \leq B_2$.

Theorem 1 is a weak form of a natural conjecture. For a given subgroup $H \subseteq \mathscr{E}(C)$ and a given extension k'/k, define the *k'-rational division hull of H on the fibre above $t \in C(k')$* by

$$H_t^{\text{div}}(k') = \bigl\{Q \in \mathscr{E}_t(k') : NQ \in H_t \text{ for some } N \geq 1\bigr\},$$

where $H_t \subseteq \mathscr{E}_t(k')$ denotes the specialization of H at t. It is plausible to conjecture that if $\mathscr{E} \to C$ is a nonisotrivial elliptic surface and $H \subseteq \mathscr{E}(C)$ is any subgroup, then the index $(H_t^{\text{div}}(k') : H_t)$ is bounded for $t \in C(k')$, and that the bound depends only on $D = [k' : k]$, $\mathscr{E}$, and H. In the case where H has rank zero, this conjecture follows immediately from Merel's theorem [1996], since $H_t^{\text{div}}(k')$ will simply be the k'-rational torsion on $\mathscr{E}_t$. Indeed, Merel's theorem plays a role in the results in this paper, since we have not assumed that P is a section of infinite order. The main result of this paper is equivalent to the boundedness of the ℓ-power part of $(H_t^{\text{div}}(k') : H_t)$, in terms of $D = [k' : k]$, in the rank one case. In the case $D = 1$, Silverman [1985] has shown that the conjecture follows from a conjecture of Lang, which posits a certain lower bound for the Néron–Tate height on an elliptic curve. Lang's conjecture, in turn, has been shown by Hindry and Silverman [1988] to follow from Szpiro's conjecture, and hence the *abc* conjecture of Masser and

Oesterlé. If one supposes that the constants in Lang's conjecture depend only on the degree of the number field involved, then the full version of the conjecture above follows by Silverman's argument.

In general, the finiteness of Theorem 1 is the best we can do in this setting, since one is free to choose $P = \ell^N P_0$ for some section $P_0 : C \to \mathscr{E}$, and N arbitrarily large. This ensures that the sets in the statement of Theorem 1 have size at least N. If, however, we cast out finitely many primes and finitely many fibres, and consider only k-rational points, we obtain something much more explicit. We call the prime ℓ a *special prime* for the elliptic surface $\mathscr{E}$ if it is one of the finitely many primes such that either $\ell = 2$ or the j-invariant $j_{\mathscr{E}} : C \to \mathbb{P}^1$ has a pole of order divisible by ℓ.

Theorem 2. *Let $\mathscr{E}$ and P be as above, suppose that ℓ is not a special prime for $\mathscr{E}$, and suppose further that P is not of the form ℓP_0 for any section $P_0 : C \to \mathscr{E}$. Then*

$$\#\big\{Q \in \mathscr{E}_t(k) : \ell^n Q = P_t \text{ for some } n \geq 1\big\} \leq \ell^2$$

for all but finitely many $t \in C(k)$. The upper bound can be replaced with 0 *if C has genus at least* 1 *or if the j-invariant $j_{\mathscr{E}}$ has at least* 5 *distinct poles in $C(\bar{k})$;* 4 *poles suffice if $\ell = 5$, and* 3 *suffice if $\ell \geq 7$.*

Of course, this theorem is trivially true if C has genus 2 or greater.

Remark. The proof of Theorem 2 is a modification of the proof of Theorem 1, and with slightly more work, one can obtain a version for points of bounded degree over k. Specifically, in proving the first claim of Theorem 2, we actually prove (under the conditions of the theorem) that $\ell^n Q = P_t$ implies $n \leq 1$, except on fibres corresponding to finitely many $t \in C(k)$. If the argument is extended, one can show that for all but finitely many $t \in C(k)$, if $Q \in \mathscr{E}_t(\bar{k})$ with $[k(Q) : k] \leq D$, then we have $\ell^n Q = P_t$ only if $n < \log_2 D + 5$. Note that, if we are allowed to consider points with $[k(Q) : k] \leq D$, then we can find examples with $\ell^n Q = P_t$ for any $n \leq \log D/(2 \log \ell)$.

There is one important case in which Theorem 2 does not apply: if we'd like to discuss torsion on specializations of an elliptic surface, then we should like to apply Theorem 2 with $P = \mathbb{O}$, the identity section. However, it is always true that $\mathbb{O} = \ell\mathbb{O}$. As it happens, this is not a fundamental obstacle. For any finite set S of rational primes, and any elliptic curve E, let $E^{\mathrm{Tors},S}$ denote the S-primary torsion on E, that is, the union of $E[N]$ as $N \in \mathbb{N}$ ranges over S-units.

Theorem 3. *Let $\mathscr{E}$ be as above, and suppose that $j_{\mathscr{E}}$ has at least* 5 *distinct poles in $C(\bar{k})$. Then for all but finitely many $t \in C(k)$, the torsion subgroup of $\mathscr{E}_t(k)$ is exactly $\mathscr{E}_t^{\mathrm{Tors},S}(k)$, where S is the set of special primes for $\mathscr{E}$.*

As mentioned above, it follows from Merel's theorem (which plays a role in the proof of Theorems 1 and 3) that size of the torsion subgroup of $\mathscr{E}_t(k)$ is bounded independent of $t \in C(k)$. The significance of Theorem 3, then, is that it more precisely describes the structure of $\mathscr{E}_t(k)^{\mathrm{Tors}}$ on all but finitely many fibres.

If E/k is an elliptic curve over a number field, then one might consider E as an arithmetic surface $E \to \operatorname{Spec}(R)$, where R is the ring of integers of k, and ask if the result analogous to Theorem 1 holds. Indeed, results of this general type, that is, local-to-global results about divisibility in the Mordell–Weil group, have already been considered over number fields, for example, the work of Banaszak, Gajda, and Krasoń [Banaszak et al. 2005]. However, for elliptic curves over number fields, it is quite easy to show that something rather different from Theorem 1 is true.

Theorem 4. *Let k be a number field and let E/k be an elliptic curve. If $P \in E(k)$ is a point of infinite order and ℓ is any rational prime, then for any $M \geq 1$, we may choose an infinite set of primes S_M such that*

$$\#\big\{Q \in E_{\mathfrak{p}}(k_{\mathfrak{p}}) : \ell^n Q = P_{\mathfrak{p}} \textit{ for some } n \geq 1\big\} \geq M$$

for all $\mathfrak{p} \in S_M$. Moreover, we can choose this set of primes to have density $M^{-2/5+o(1)}$, where $o(1) \to 0$ as $M \to \infty$.

The results of Banaszak, Gajda, and Krasoń are largely Galois-theoretic, while those of Silverman [1985] are obtained by studying the variation of the Néron–Tate height across fibres of $\mathscr{E}$. By contrast, Theorems 1 and 2 require a mixture of Galois theory and some deep results in diophantine geometry. Since it may be of independent interest, we mention here the main diophantine lemma used, which is a slight adjustment of [Faber et al. 2009, Lemma 4.5], which in turn derives from work of Vojta [1992] and Song and Tucker [2001].

Lemma 5. *Let*

$$C_0 \xleftarrow{\phi_1} C_1 \xleftarrow{\phi_2} C_2 \xleftarrow{\phi_3} \cdots$$

be a tower of (smooth, projective) curves connected by nonconstant morphisms, defined over the number field k. Let R_{ϕ_n} denote the ramification divisor of ϕ_n, and suppose that there are constants $c_1 > 0$ and c_2 such that

$$\frac{\deg R_{\phi_n}}{2 \deg \phi_n} \geq c_1 2^n - c_2 \quad \textit{for all } n.$$

Then for each $B \geq 1$, there exists an $N(B)$ such that $C_{N(B)}(\bar{k})$ contains at most finitely many points Q with $[k(Q) : k] \leq B$. (In fact, we see in Section 5 below that we may take $N(B)$ to be at most $\log_2 B + O(1)$ as $B \to \infty$.)

Before proceeding with the outline of the main argument, we remark on the requirement that $\mathscr{E}$ have nonconstant j-invariant. While Silverman's result above has been considered in the case of elliptic surfaces $\mathscr{E} \to \mathbb{P}^1$ with constant j-invariant, by Gupta and Ramsay [1997], it is clear that our main result cannot hold for all split elliptic surfaces. If E/k is an elliptic curve with rank at least 1, let $\mathscr{E} \to E$ be an elliptic surface birational to $E \times E$, with projection onto the second coordinate. Let $Q \in E(k)$ be a point of infinite order, and let P be the diagonal section $P : E \to \mathscr{E}$ defined by $t \mapsto (t, t)$. Then for any N, if we set $t = \ell^N Q$, we clearly have

$$\#\big\{Q' \in E_t(k) = E(k) : \ell^n Q' = P_t = \ell^N Q\big\} \geq N.$$

2. Notation and outline of the argument

Although much of the proof of Theorem 1 will take place in function fields, it is useful to keep in mind the geometric picture. The strategy of the proof is as follows: let $\mathscr{E} \to C$ be our elliptic surface, and let $\Gamma_0 \subseteq \mathscr{E}$ be the image of our section $P : C \to \mathscr{E}$. For each n, let Γ_{n+1} be the pullback of Γ_n by the rational function $[\ell] : \mathscr{E} \to \mathscr{E}$. For any extension F/k, points in $\Gamma_n(F)$ parametrize fibres $\mathscr{E}_t$ of $\mathscr{E}$ with $t \in C(F)$, with a marked point $Q \in \mathscr{E}_t(F)$ such that $\ell^n Q = P_t$. In general, these curves might be singular and/or reducible, but we will imagine for the moment that their normalizations $\widetilde{\Gamma}_n$ are (geometrically) irreducible. In other words, we have a tower

$$\widetilde{\Gamma}_0 \leftarrow \widetilde{\Gamma}_1 \leftarrow \widetilde{\Gamma}_2 \leftarrow \widetilde{\Gamma}_3 \leftarrow \cdots$$

of smooth projective curves, connected by dominant morphisms (namely, those induced by multiplication-by-ℓ), all defined over the number field k. Lemma 5 gives us control over points of low algebraic degree on the curves $\widetilde{\Gamma}_n$, provided that the morphisms above ramify enough (equivalently, the genera of the curves increase quickly enough). Obtaining the appropriate lower bound on ramification provides for some tricky geometry, since the only possibility for ramification is where the curves Γ_n intersect singular fibres of $\mathscr{E}$, and these are precisely the points at which one might need to blow up in order the resolve the singularities of Γ_n. We avoid this by moving the entire problem into the function field setting, and applying Tate's nonarchimedean uniformization of elliptic curves. It is here that we use the hypothesis that $j_{\mathscr{E}}$ be nonconstant, since poles of $j_{\mathscr{E}}$ correspond to fibres at which $\mathscr{E}$, after suitable base extension, has semistable reduction. The resulting estimates on ramification, combined with Lemma 5, suffice to prove the results in this special case.

Generally, we can't hope for the curves $\widetilde{\Gamma}_n$ to actually be irreducible (in particular, if P is a multiple by ℓ of another section, then $\widetilde{\Gamma}_1$ has a component birational to C), but each is the disjoint union of finitely many components, and the rational map $[\ell] : \mathscr{E} \to \mathscr{E}$ induces a map from each component of $\widetilde{\Gamma}_{n+1}$ to some component of $\widetilde{\Gamma}_n$. Denoting the components of the normalized curves by $\widetilde{\Gamma}_i^{(j)}$, we have a tree of curves with dominant morphisms which looks something like this:

$$\begin{tikzcd}
 & \widetilde{\Gamma}_1^{(1)} \arrow[dl] & \widetilde{\Gamma}_2^{(1)} \arrow[l] \\
\widetilde{\Gamma}_0 & & \widetilde{\Gamma}_2^{(2)} \cdots \arrow[dl] \\
 & \widetilde{\Gamma}_1^{(2)} \arrow[ul] & \widetilde{\Gamma}_2^{(3)} \cdots \arrow[l] \\
 & & \widetilde{\Gamma}_2^{(4)} \cdots \arrow[ul]
\end{tikzcd}$$

The key is to show that this tree is eventually nonbranching. In other words, we want to show that the tree depicted above contains only finitely many infinite paths, so that we may apply Lemma 5 to each of these paths. Looking at the surface as an elliptic curve E over $K = k(C)$, this amounts to showing there is some N such that the sets $[\ell]^{-n}P \subseteq E(\overline{K})$ contain at most N Galois orbits for any $n \geq 1$. For elliptic curves over number fields, this follows from Kummer theory, but it seems that these results have not previously been extended to elliptic curves over complex function fields. In Section 3, we prove the appropriate Galois-theoretic results to show that the number of components of the curves $\widetilde{\Gamma}_n$ eventually stabilizes. In Section 4, we will employ Tate's uniformization of elliptic curves over local fields to study the ramification of the maps $\widetilde{\Gamma}_{n+1} \to \widetilde{\Gamma}_n$ induced by $[\ell]$. In both of these sections, we consider E over the extension $K \otimes_k \mathbb{C}$ of K, in order to obtain geometric results. In Section 5, we assemble the proof of Theorem 1, and in Sections 6 and 7, the proofs of Theorems 2 and 4.

Throughout the paper, $\mathscr{E}$ is a smooth elliptic surface (with some chosen 'identity section') defined over the number field k. We denote by E/K the generic fibre of $\mathscr{E}$, an elliptic curve over the function field $K = k(C)$. To obtain geometric results, in Sections 3 and 4, we will frequently work over the extension $K_{\mathbb{C}} = K \otimes_k \mathbb{C}$, determined by some fixed embedding of k into $\mathbb{C}$. The curves Γ_n are as defined above, and $\widetilde{\Gamma}_n$ are their normalizations. Since $\widetilde{\Gamma}_0 \cong C$, we often identify these curves tacitly. Also, in a slight abuse of notation, P will stand both for the section $P : C \to \mathscr{E}$, as well as the corresponding point in $E(K)$.

3. Galois orbits

For any prime ℓ, the action of the absolute Galois group $\mathrm{Gal}(\overline{K}_{\mathbb{C}}/K_{\mathbb{C}})$ on $E(K_{\mathbb{C}})$ partitions

$$[\ell]^{-n}P = \{Q \in E(\overline{K}_{\mathbb{C}}) : \ell^n Q = P\}$$

into a certain number of orbits for each n, and we wish to show that this number is bounded as $n \to \infty$. In other words, we wish to show that Galois acts nearly as freely on $[\ell]^{-n}P$ as the group structure allows. If ℓ is not a special prime and

$P \notin \ell E(K)$, we will show that $[\ell]^{-n}P$ is in fact Galois-irreducible for all n (in the sense that all of its elements are conjugate under the action of Galois).

For each n we set $K_n = K_{\mathbb{C}}(E[\ell^n])$, the ℓ^n-division field of $E/K_{\mathbb{C}}$, and we set K_∞ to be the union of the $K_n/K_{\mathbb{C}}$. Let

$$T_\ell(E) = \varprojlim E[\ell^n]$$

be the ℓ-adic Tate module of E. Fixing a basis for $T_\ell(E)$ allows us to define a representation

$$\rho_\ell : \mathrm{Gal}(K_\infty/K_{\mathbb{C}}) \to \mathrm{GL}_2(\mathbb{Z}_\ell).$$

In the number field case, a deep theorem of Serre [1972] states that such representation is surjective for all but finitely many primes ℓ (unless the elliptic curve has complex multiplication). For elliptic curves over $K_{\mathbb{C}}$, we cannot expect this to be true. It is easy to show, using the Weil pairing on $T_\ell(E)$, that the image of the representation ρ_ℓ must be contained in $\mathrm{SL}_2(\mathbb{Z}_\ell)$, but the following lemma tells us that the representation (almost) surjects onto this subgroup.

Lemma 6. *Let*

$$\rho_\ell : \mathrm{Gal}(K_\infty/K_{\mathbb{C}}) \to \mathrm{SL}_2(\mathbb{Z}_\ell)$$

be the ℓ-adic Galois representation associated to $E/K_{\mathbb{C}}$. Then the image of ρ_ℓ has finite index in $\mathrm{SL}_2(\mathbb{Z}_\ell)$. *Moreover, if ℓ is not a special prime for $\mathscr{E}$, then ρ_ℓ is surjective.*

Proof. The first claim is a theorem of Igusa [1959]. More precisely, Cox and Parry [1984] show that the image of Galois in

$$\mathrm{SL}_2(\widehat{\mathbb{Z}}) = \prod_\ell \mathrm{SL}_2(\mathbb{Z}_\ell)$$

contains the congruence subgroup

$$\Gamma(N) = \left\{ M \in \mathrm{SL}_2(\mathbb{Z}_\ell) : M \equiv \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \pmod{N} \right\},$$

for N twice the least common multiple of the orders of the poles of the j-invariant $j_{\mathscr{E}} : C \to \mathbb{P}^1$. In other words, if ℓ is not a special prime for $\mathscr{E}$ (that is, if ℓ is odd and prime to the orders of the poles of $j_{\mathscr{E}}$), then the image of Galois is all of $\mathrm{SL}_2(\mathbb{Z}_\ell)$. □

We will need some facts about the Galois cohomology of elliptic curves over complex function fields. In the number field case, the theory is reasonably well understood due to work of Bashmakov [1972]; see also Ribet [1979]. The techniques rely, however, on the fact that in the number field setting, Galois acts on the ℓ-primary torsion as an open subgroup of $\mathrm{GL}_2(\mathbb{Z}_\ell)$. In particular, Bashmakov exploits elements of the centre of GL_2, while the centre of SL_2 is decidedly less interesting. Nonetheless, the result we need is still true in this setting.

Lemma 7. *The first cohomology group $H^1(\mathrm{Gal}(K_\infty/K_\mathbb{C}), T_\ell(E))$ has finite exponent. Furthermore, if ℓ is an odd prime and if the representation ρ_ℓ is surjective, then the group is trivial.*

Proof. In the case where ρ_ℓ is surjective, the traditional proof works: the group $G = \mathrm{Gal}(K_\infty/K_\mathbb{C}) \cong \mathrm{SL}_2(\mathbb{Z}_\ell)$ contains an element that acts as -1 on $T_\ell(E)$. Since this element is in the centre of G, we know (by a lemma of Sah) that multiplication by -2 annihilates the first cohomology group. Since the group $H^1(G, T_\ell(E))$ is ℓ-power torsion (multiplication by ℓ^m kills $E[\ell^m]$, and $H^1(G, T_\ell(E))$ is the projective limit of $H^1(\mathrm{Gal}(K_n/K_\mathbb{C}), E[\ell^n])$), it has exponent $\gcd(2, \ell)$.

We now treat the more general case, modifying an argument of Tate presented by Coates [1970]. For simplicity, choose a basis for $T_\ell(E)$ and identify G with its image in $\mathrm{SL}_2(\mathbb{Z}_\ell)$. By Lemma 6, the image of the map

$$\mathrm{Gal}(K_\infty/K_\mathbb{C}) \to \mathrm{SL}_2(\mathbb{Z}_\ell)$$

contains the congruence subgroup $\Gamma(\ell^N)$, the kernel of reduction modulo ℓ^N, for some $N \geq 0$. Now, let $H \subseteq \Gamma(\ell^N) \subseteq G$ be the subgroup generated by the set of matrices

$$\left\{ \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix}, \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} : \beta \in \ell^N \mathbb{Z}_\ell \right\}.$$

We will show that $H^1(H, T_\ell(E))$ has finite exponent. As noted at the beginning of the proof of [Bandini et al. 2009, Lemma 3.10], H contains $\Gamma(\ell^{2N})$, and so

$$(G : H) \leq (G : \Gamma(\ell^{2N})) \leq (\mathrm{SL}_2(\mathbb{Z}_\ell) : \Gamma(\ell^{2N})) = \ell^{1+6N}(\ell^2 - 1).$$

Since the composition of the restriction and corestriction maps

$$H^1(G, T_\ell(E)) \xrightarrow{\mathrm{Res}} H^1(H, T_\ell(E)) \xrightarrow{\mathrm{Cor}} H^1(G, T_\ell(E))$$

is simply multiplication by $(G : H)$, proving that $H^1(H, T_\ell(E))$ has finite exponent would be enough to prove the same of $H^1(G, T_\ell(E))$.

Let $f : H \to T_\ell(E)$ be a continuous 1-cocyle. For ease of reference, we will write

$$f\begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} f_1(\beta) \\ f_2(\beta) \end{pmatrix}.$$

Note that, by the cocycle relation,

$$\begin{aligned} \begin{pmatrix} f_1(\beta_1+\beta_2) \\ f_2(\beta_1+\beta_2) \end{pmatrix} &= \begin{pmatrix} f_1(\beta_1) \\ f_2(\beta_1) \end{pmatrix} + \begin{pmatrix} 1 & \beta_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} f_1(\beta_2) \\ f_2(\beta_2) \end{pmatrix} \\ &= \begin{pmatrix} f_1(\beta_1) + f_2(\beta_2) + \beta_1 f_2(\beta_2) \\ f_2(\beta_1) + f_2(\beta_2) \end{pmatrix}, \end{aligned} \tag{1}$$

and so, in particular,

$$f_2(\beta_1+\beta_2)=f_2(\beta_1)+f_2(\beta_2)\quad\text{for all } \beta_1,\beta_2\in\ell^N\mathbb{Z}_\ell.$$

Since cocycles send the identity to the identity, f_2 is actually a homomorphism from $\ell^N\mathbb{Z}_\ell$ to $\mathbb{Z}_\ell$.

Now, for any $\alpha\in 1+\ell^{2N}\mathbb{Z}_\ell$, we have $\operatorname{diag}(\alpha,\alpha^{-1})\in H$. This follows from the aforementioned comment in [Bandini et al. 2009], or more directly from observing that for any $\gamma\in\mathbb{Z}_\ell$,

$$\begin{pmatrix}1 & 0\\ -\ell^N/(1+\ell^N\gamma) & 1\end{pmatrix}\begin{pmatrix}1 & \ell^N\gamma\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ \ell^N & 1\end{pmatrix}\begin{pmatrix}1 & -\ell^N\gamma/(1+\ell^N\gamma)\\ 0 & 1\end{pmatrix}\\ =\begin{pmatrix}1+\ell^{2N}\gamma & 0\\ 0 & (1+\ell^{2N}\gamma)^{-1}\end{pmatrix}.$$

At this point, to simplify notation, we will write

$$\sigma_{\alpha,\beta}=\begin{pmatrix}\alpha & \beta\\ 0 & \alpha^{-1}\end{pmatrix}.$$

Now, on the one hand, we have the relation

$$\sigma_{\alpha,0}\sigma_{1,\beta}\sigma_{\alpha,0}^{-1}=\sigma_{1,\alpha^2\beta},$$

by simply multiplying the matrices. On the other hand, since f is a cocyle we have $f(\sigma^{-1})=-\sigma^{-1}f(\sigma)$ for all σ. Thus, if $\alpha\in\mathbb{Z}\cap(1+\ell^{2N}\mathbb{Z}_\ell)$, (suppressing the first coordinate for convenience)

$$\begin{aligned}\begin{pmatrix}* \\ \alpha^2 f_2(\beta)\end{pmatrix}=\begin{pmatrix}* \\ f_2(\alpha^2\beta)\end{pmatrix}&=f(\sigma_{1,\alpha^2\beta})=f(\sigma_{\alpha,0}\sigma_{1,\beta}\sigma_{\alpha,0}^{-1})\\ &=f(\sigma_{\alpha,0})+\sigma_{\alpha,0}f(\sigma_{1,\beta})+\sigma_{\alpha,0}\sigma_{1,\beta}f(\sigma_{\alpha,0}^{-1})\\ &=f(\sigma_{\alpha,0})+\sigma_{\alpha,0}f(\sigma_{1,\beta})-\sigma_{\alpha,0}\sigma_{1,\beta}\sigma_{\alpha,0}^{-1}f(\sigma_{\alpha,0})\\ &=\begin{pmatrix}0 & -\alpha^2\beta\\ 0 & 0\end{pmatrix}f(\sigma_{\alpha,0})+\sigma_{\alpha,0}f(\sigma_{1,\beta})=\begin{pmatrix}* \\ \alpha^{-1}f_2(\beta)\end{pmatrix}.\end{aligned}$$

Thus, for a given β, we have $\alpha^2 f_2(\beta)=\alpha^{-1}f_2(\beta)$ for any integer $\alpha\equiv 1 \pmod{\ell^{2N}}$. Clearly then, $f_2(\beta)=0$ for all β.

We now have f_2 vanishing identically, and from (1) above, f_1 must be a homomorphism. Again, we have $f_1(\alpha\beta)=\alpha f_1(\beta)$ for all $\alpha\in\mathbb{Z}\cap(1+\ell^{2N}\mathbb{Z}_\ell)$ and $\beta\in\ell^N\mathbb{Z}_\ell$. Write

$$f\begin{pmatrix}\alpha & 0\\ 0 & \alpha^{-1}\end{pmatrix}=\begin{pmatrix}g_1(\alpha)\\ g_2(\alpha)\end{pmatrix}\quad\text{for all } \alpha\in 1+\ell^{2N}\mathbb{Z}_\ell.$$

We compute again

$$\begin{aligned}\begin{pmatrix}\alpha^2 f_1(\beta)\\ 0\end{pmatrix} &= f(\sigma_{1,\alpha^2\beta}) = f(\sigma_{\alpha,0}\sigma_{1,\beta}\sigma_{\alpha,0}^{-1})\\ &\ \vdots\\ &= \begin{pmatrix}0 & -\alpha^2\beta\\ 0 & 0\end{pmatrix} f(\sigma_{\alpha,0}) + \sigma_{\alpha,0} f(\sigma_{1,\beta})\\ &= \begin{pmatrix}-\alpha^2\beta g_2(\alpha)\\ 0\end{pmatrix} + \begin{pmatrix}\alpha f_1(\beta)\\ 0\end{pmatrix}.\end{aligned}$$

Thus, taking $\alpha = 1 + \ell^{2N}$, we obtain

$$\left(\alpha^2 - \alpha\right) f_1(\beta) = -\alpha^2\beta g_2(\alpha),$$

and therefore

$$\ell^{2N} f\begin{pmatrix}1 & \beta\\ 0 & 1\end{pmatrix} = \begin{pmatrix}-\alpha g_2(\alpha)\beta\\ 0\end{pmatrix} \quad \text{for all } \beta.$$

Using essentially the same argument, we can also show that

$$\ell^{2N} f\begin{pmatrix}1 & 0\\ \beta & 1\end{pmatrix} = \begin{pmatrix}0\\ g_1(\alpha)\beta\end{pmatrix}.$$

Thus, for

$$\sigma \in \left\{\begin{pmatrix}1 & \beta\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\ \beta & 1\end{pmatrix} : \beta \in \ell^N \mathbb{Z}_\ell\right\},$$

we obtain

$$\ell^{2N} f(\sigma) = \sigma(\xi) - \xi \quad \text{for } \xi = \begin{pmatrix}g_1(\alpha)\\ -\alpha g_2(\alpha)\end{pmatrix}.$$

The same relation must hold on all of H, since matrices of this form generate H, and so $\ell^{2N} f$ is a 1-coboundary. Since f was arbitrary, $\ell^{2N} H^1(H, T_\ell(E))$ is trivial, proving the result. □

Lemma 8. *For sufficiently large m, if $Q \in E(\overline{K}_{\mathbb{C}})$ and $\ell^m Q = P$, then $Q \notin E(K_\infty)$. If ℓ is not a special prime and $P \notin \ell E(K_{\mathbb{C}})$, then $\ell Q = P$ implies $Q \notin E(K_\infty)$.*

Proof. Suppose that $Q \in K_n$, and consider the map

$$f : G_n = \mathrm{Gal}(K_n/K_{\mathbb{C}}) \to E[\ell^m]$$

defined by $f(\sigma) = \sigma(Q) - Q$, which is a 1-cocyle. By Lemma 7, there is some s, not depending on n and m, such that $\ell^s H^1(G_n, E[\ell^m])$ is trivial. It follows that $\ell^s f$ is a 1-coboundary, so that $\ell^s f(\sigma) = \sigma(\xi) - \xi$ for some fixed $\xi \in E[\ell^m]$ and all $\sigma \in G_n$. One checks that this implies $\sigma(\ell^s Q - \xi) = \ell^s Q - \xi$ for all $\sigma \in G_n$, and so $\ell^s Q - \xi \in E(K_{\mathbb{C}})$. This implies

$$\ell^s P = \ell^m(\ell^s Q - \xi) \in \ell^m E(K_{\mathbb{C}}),$$

which cannot be true if m is large enough, since s is independent of m, and $E(K_{\mathbb{C}})$ is finitely generated.

If ℓ is not a special prime, let $\ell Q = P$ for some $Q \in E(K_n)$. The triviality of $H^1(G_n, E[\ell])$, by the argument above with $s = 0$, leads to $Q + \xi \in E(K_{\mathbb{C}})$ for some $\xi \in E[\ell]$. This implies $P \in \ell E(K_{\mathbb{C}})$, which we have assumed is not true. □

Now choose a consistent family of preimages of P, that is, a sequence Q_s in $E(\overline{K}_{\mathbb{C}})$ such that $Q_0 = P$ and $\ell Q_{s+1} = Q_s$. We consider the maps

$$f_s : \mathrm{Gal}(K_\infty(Q_s)/K_\infty) \to E[\ell^s], \quad \sigma \mapsto \sigma(Q_s) - Q_s.$$

Then f_s is a homomorphism, since it is a 1-cocyle and the domain acts trivially on the image. It is also clear that f_s is injective, since any $\sigma \in \mathrm{Gal}(K_\infty(Q_s)/K_\infty)$ fixing Q_s must fix everything. What's more, the diagram

$$\begin{tikzcd}
\mathrm{Gal}(K_\infty(Q_{s+1})/K_\infty) \arrow[r, "\text{restriction}"] \arrow[d, "f_{s+1}"'] & \mathrm{Gal}(K_\infty(Q_s)/K_\infty) \arrow[d, "f_s"'] \\
\mathrm{im}(f_{s+1}) \arrow[r, "{[\ell]}"'] & \mathrm{im}(f_s)
\end{tikzcd}$$

commutes (since $[\ell]$ is defined over the ground field), and so we may take (compatible) projective limits of the top and bottom. If we let K' denote the union of all fields $K_\infty(Q_s)$, we thereby obtain a map f_∞ from $\mathrm{Gal}(K'/K_\infty)$ to the Tate module $T_\ell(E)$. The group

$$H_\infty = \varprojlim \mathrm{im}(f_s) \subseteq \varprojlim E[\ell^s] = T_\ell(E)$$

is a submodule of $T_\ell(E)$ (both in terms of the Galois module structure, and the $\mathbb{Z}_\ell$ module structure). To sum up, projective limits give us the following diagram, with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Gal}(K'/K_\infty) \arrow[r] \arrow[d, "f_\infty"] & \mathrm{Gal}(K'/K_{\mathbb{C}}) \arrow[r] \arrow[d] & \mathrm{Gal}(K_\infty/K_{\mathbb{C}}) \arrow[r] \arrow[d, "\rho_\ell"] & 0 \\
0 \arrow[r] & T_\ell(E) \arrow[r] & T_\ell(E) \rtimes \mathrm{SL}_2(\mathbb{Z}_\ell) \arrow[r] & \mathrm{SL}_2(\mathbb{Z}_\ell) \arrow[r] & 0.
\end{tikzcd}$$

The next lemma shows that $\mathrm{im}(f)_\infty$ has finite index in $T_\ell(E)$. Note that in light of Lemma 6 and the exactness of the sequences above, this is the same as showing that the image of $\mathrm{Gal}(K'/K_{\mathbb{C}})$ has finite index in $T_\ell(E) \rtimes \mathrm{SL}_2(\mathbb{Z}_\ell)$,

Lemma 9. *The group H_∞ has finite index in $T_\ell(E)$. If ℓ is not a special prime, then in fact $H_\infty = T_\ell(E)$.*

Proof. This simply follows from the fact that H_∞ is a submodule of $T_\ell(E)$ and that the image of the action of Galois on $T_\ell(E)$ is an open subgroup of $\mathrm{SL}_2(\mathbb{Z}_\ell)$. First suppose that f_1 is surjective, so that $\mathrm{im}(f_1) = E[\ell]$. Then H_∞ is a submodule of $T_\ell(E)$, with the property that $T_\ell(E) = H_\infty + \ell T_\ell(E)$. It follows from Nakayama's lemma [Lang 2002, Lemma 4.2, page 425] that $H_\infty = T_\ell(E)$.

We now treat the general case. First of all, it is clear that H_∞ is not cyclic. If it were, then, as H_∞ is a Galois submodule of $T_\ell(E)$, the image of the representation

$$\rho_\ell : \mathrm{Gal}(K_\infty/K_{\mathbb{C}}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}_\ell)$$

would be contained in a Borel subgroup, which clearly violates Lemma 6. So we may choose two linearly independent elements in H_∞. Let

$$\begin{pmatrix} u_1\ell^{a_1} \\ u_2\ell^{a_2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} v_1\ell^{b_1} \\ v_2\ell^{b_2} \end{pmatrix}$$

be these two elements, with the u_i and v_i units in $\mathbb{Z}_\ell$. By standard linear algebra, we may rewrite this basis and multiply by a power of ℓ, to obtain $\ell^s e_1, \ell^s e_2 \in H_\infty$, where e_1 and e_2 are the standard basis vectors. But then $\ell^s T_\ell(E) \subseteq H_\infty$, and so

$$(T_\ell(E) : H_\infty) \le (T_\ell(E) : \ell^s T_\ell(E)) = \ell^{2s}. \qquad \square$$

We now state the main claim of this section, namely that Galois acts nearly as freely on $[\ell]^{-n}P$ as the group structure allows.

Lemma 10. *The number of distinct Galois orbits in $[\ell]^{-s}P$ over K_∞ is bounded by*

$$(E[\ell^s] : \mathrm{im}(f_s)) \le (T_\ell(E) : H_\infty).$$

Proof. Recall the point Q_s such that $f_s(\sigma) = \sigma(Q_s) - Q_s$. By definition, the Galois orbit of Q_s is simply $Q_s + \mathrm{im}(f_s)$. Any other Galois orbit is of the form $Q' + \mathrm{im}(f_s)$ for some $\ell^s Q' = P$. For each such Q', there is a $\xi' \in E[\ell^s]$ such that $Q' = Q_s + \xi'$, and so the Galois orbit of Q' is $Q_s + \xi' + \mathrm{im}(f_s)$. This gives an explicit bijection between the Galois orbits of $\ell^{-s}P$, and the cosets of $\mathrm{im}(f_s)$ by $E[\ell^s]$. That this number is bounded by the index $(T_\ell(E) : H_\infty)$ follows from the fact that $T_\ell(E)/H_\infty$ is the projective limit of the groups $E[\ell^s]/\mathrm{im}(f_s)$. $\square$

4. Tate uniformization and ramification

The aim of this section is to show that the tower of preimage curves described in Section 2 is sufficiently ramified. One can see geometrically why this must be true: if v is a place of C over which $E/K_{\mathbb{C}}$ has split multiplicative reduction, then the fibre of the Néron model over v is the union of $v(j_E)$ lines (with intersection points removed). The restriction of $[\ell]$ to any of the components of this Néron polygon is an ℓ-to-one map to some other component. In particular, if the group of sections on $\mathscr{E}$ contains all of the ℓ^n-th preimages of P, then $\ell^n \mid v(j_E)$. This argument applies to extensions $\mathscr{E} \times_C C'$ (given a Néron model over the extension) as well, and so if w is a prolongation of v to a field over which $[\ell]^{-n}P$ is rational for large n, we have $w(j_E) > v(j_E)$, so v is ramified in this extension.

Although the geometric argument above can be turned into a proof, it is not entirely straightforward, because the relationship between the Néron model of $\mathscr{E}$ and the Néron model of the base extension of $\mathscr{E} \times_C C'$ for some curve $C' \to C$ is somewhat subtle when the covering is ramified. We obtain a simpler proof by considering the function field version of the problem. The main tool is Tate's v-adic uniformization of elliptic curves, which is described over number fields in [Silverman 1994, V.3–V.6]. The results over function fields are identical, and may be found in [Roquette 1970].

Throughout this section, we fix a prime v at which $E/K_{\mathbb{C}}$ has split multiplicative reduction (we assume that one exists), and we suppose that P does not reduce to the singular point modulo this prime. In the proof of Theorem 1, we will reduce the problem to the case where these assumptions hold.

Theorem (Tate). *Let F be a field complete with respect to the nonarchimedean valuation v, and suppose that E/F is an elliptic curve with split multiplicative reduction at v. Then there is a unique $q \in F^*$ with $|q|_v < 1$, and maps such that*

$$0 \to q^{\mathbb{Z}} \to F^* \to E(F) \to 0$$

is an exact sequence. Also, if F'/F is a Galois extension, then the corresponding sequence is an exact sequence of Galois modules.

In essence, completing with respect to a prime of split multiplicative reduction, then allows us to glean a lot of information about the elliptic curve E by considering the multiplicative group of the completion of the field.

Lemma 11. *Let F be field complete with respect to the normalized discrete valuation v, with ring of integers R, and an algebraically closed residue field R/v of characteristic not dividing n. Then for any $\alpha \in F^*$, we have $\alpha \in (F^*)^n$ if and only if $n \mid v(\alpha)$.*

Proof. In one direction, note that if $\alpha \in (F^*)^n$, then $v(\alpha) = nv(\beta)$ for some $\beta \in F^*$. It follows at once that $n \mid v(\alpha)$.

Let $r_v : R \to R/v$ be the reduction-modulo-v map. First suppose that $\alpha \in R^*$. Then $r_v(\alpha) \neq 0$, and since the residue field is algebraically closed, $x^n - r_v(\alpha)$ has a simple root in R/v. By Hensel's Lemma [Serre 1979, p. 34], there is a root of $x^n - \alpha$ in R. If u is such a root, then $nv(u) = v(\alpha) = 0$, and so $u \in R^*$, whence $\alpha = u^n \in (R^*)^n$.

Now suppose that $\alpha \in F^* \setminus R^*$. By taking reciprocals if necessary, suppose that $\alpha \in R$. If $v(\alpha) = m$, write $\alpha = \gamma\pi^m$, where π is a uniformizer for v, and $\gamma \in R^*$. By the previous argument, $\gamma = u^m$ for some $u \in R^*$, and so $\alpha = (u\pi)^m \in (F^*)^m$. □

To state our next result, we will define the *ramification tree* of the point $Q_0 = P_v$ on $\Gamma_0 \subseteq \mathscr{E}$. The nodes of the tree are the points on the curves $\widetilde{\Gamma}_n$ which map down to Q_0, with a point Q_{n+1} on $\widetilde{\Gamma}_{n+1}$ linked to a point Q_n on $\widetilde{\Gamma}_n$ if Q_{n+1} maps to Q_n by the map induced by $[\ell]$. We will weight these edges with the ramification index of this map at Q_{n+1}, so that the weights of the edges above any given point sum to ℓ^2. For convenience, we will refer to the points on $\widetilde{\Gamma}_n$ as nodes *at level n* in our tree. When we speak of nodes *above Q*, we mean nodes at the level immediately above that of Q, which are connected to Q by an edge.

We may give an equivalent definition of the ramification tree in terms of function fields. By identifying points of $C(\mathbb{C})$ with the corresponding valuations on $K_{\mathbb{C}}$, points above Q_0, on the components of the curves $\widetilde{\Gamma}_n$, correspond to valuations on the function fields of the corresponding components which extend v, and so we may take the valuations to be the nodes of our tree. The valuation v_{n+1} of $\mathbb{C}(\widetilde{\Gamma}^{(i)}_{n+1})$ is linked to the valuation v_n of $\mathbb{C}(\widetilde{\Gamma}^{(j)}_n)$ just if the former field extends the latter (that is, $\widetilde{\Gamma}^{(i)}_{n+1}$ maps onto $\widetilde{\Gamma}^{(j)}_n$), and $v_{n+1} \,|\, v_n$. Again, the weights on the edges are simply the ramification indices $e(v_{n+1}/v_n)$. Note that the Galois orbits in $[\ell]^{-n}P$ correspond to the components of $\mathbb{C}(\widetilde{\Gamma}_n)$, with

$$\mathbb{C}(\widetilde{\Gamma}_n) \cong K_{\mathbb{C}}(Q_n) \quad \text{for } Q_n \in [\ell]^{-n}P$$

any representative of the appropriate Galois orbit.

It turns out that there are only three possible types of branching above a node in our tree: there might be

(1) ℓ^2 edges above a given node, each necessarily of weight 1;

(2) ℓ edges of weight 1 and $\ell - 1$ of weight ℓ; or

(3) ℓ edges, each of weight ℓ.

The remainder of the section is devoted to proving this, and establishing the exact structure of the tree.

Let $\widehat{K}_{\mathbb{C}}$ be the completion of $K_{\mathbb{C}}$ with respect to v. We recall some basic facts about extensions of local fields; see [Serre 1979, Section II.3]. If $L = K_{\mathbb{C}}(Q, E[\ell^n])$ for some $Q \in E(\overline{K}_{\mathbb{C}})$ with $[\ell]^n Q = P$, then L is a Galois extension of $K_{\mathbb{C}}$. If we fix a prolongation w of v to L, then the decomposition group

$$D_w = \{\sigma \in \mathrm{Gal}(L/K_{\mathbb{C}}) : w \circ \sigma = w\}.$$

of w/v is precisely the Galois group of $\widehat{L}/\widehat{K}_{\mathbb{C}}$. If $Q_1, \ldots, Q_g$ are a complete set of representatives of the $\mathrm{Gal}(L/K_{\mathbb{C}})$-orbits in $[\ell]^{-n}P$, then the prolongations of

v to the (distinct) fields $L(Q_i)$ are simply the valuations of the form $w \circ \sigma$ for $\sigma \in \mathrm{Gal}(L/K_{\mathbb{C}})$. Two automorphisms generate the same valuation if and only if they are in the same coset of D_w in $\mathrm{Gal}(L/K_{\mathbb{C}})$. Thus, the prolongations are exactly determined by the D_w-orbits (that is, the $\mathrm{Gal}(\widehat{L}/\widehat{K}_{\mathbb{C}})$-orbits) in $[\ell]^{-n}P$. Additionally, since all residual degrees are 1 (the residue field is always $\mathbb{C}$), the ramification index $e(w/v)$ of the prolongation associated to the decomposition orbit containing Q is exactly $[\widehat{K}_{\mathbb{C}}(Q) : \widehat{K}_{\mathbb{C}}]$ (this is true because, as we will see below, $\widehat{K}_{\mathbb{C}}(Q)/\widehat{K}_{\mathbb{C}}$ is a Galois extension even though $K_{\mathbb{C}}(Q)/K_{\mathbb{C}}$ may not be). Thus, the nodes at level n in our tree correspond to distinct orbits in $[\ell]^{-n}P$ under the decomposition group of some fixed valuation of L extending v. We will suppose throughout that we have extended v in some way to $\overline{K}_{\mathbb{C}}$, and the *decomposition group* of a field L will always refer to the decomposition group of the restriction of this valuation to L.

More generally, for w a prolongation of v corresponding to the point $Q \in [\ell]^{-n}P$, the nodes above w in the ramification tree correspond to the decomposition orbits of points $Q' \in [\ell]^{-1}Q$. Given a prolongation w' corresponding to (the decomposition orbit of) Q', the ramification index $e(w'/w)$ is simply

$$[\widehat{K}_{\mathbb{C}}(Q') : \widehat{K}_{\mathbb{C}}(Q)] = [\widehat{K}_{\mathbb{C}}(Q') : \widehat{K}_{\mathbb{C}}]/[\widehat{K}_{\mathbb{C}}(Q) : \widehat{K}_{\mathbb{C}}].$$

By Tate's v-adic uniformization, there is a unique $q \in \widehat{K}_{\mathbb{C}}^*$ such that for any Galois extension $F/\widehat{K}_{\mathbb{C}}$, we have an exact sequence

$$0 \to q^{\mathbb{Z}} \to F^* \xrightarrow{\phi} E(F) \to 0$$

which respects the action of the Galois group (which acts trivially on q). We point out that the units R^* map, by ϕ, onto the connected component $E_0(\widehat{K}_{\mathbb{C}})$ containing the identity; see [Silverman 1994, p. 431]. Also, $E[\ell^n]$ is generated by $\phi(q^{1/\ell^n})$ and $\phi(\zeta_{\ell^n})$, where q^{1/ℓ^n} is some ℓ^n-th root of q, and ζ_{ℓ_n} is a primitive ℓ^n-th root of unity. Note that, since $\zeta_{\ell^n} \in \mathbb{C} \subseteq F$, the group $E(\widehat{K}_{\mathbb{C}})$ contains at least cyclic ℓ^n-torsion for all n.

We now consider the points in $[\ell]^{-n}P$. Recall that we are assuming $P \in E_0(\widehat{K}_{\mathbb{C}})$, and thus we may fix, once and for all, a value $\beta \in R^*$ with $\phi(\beta) = P$. By Lemma 11, $\beta \in (R^*)^{\ell^n}$ for all n, and so we will fix a compatible system of roots $\beta^{1/\ell^n} \in R^*$ (compatible in the sense that $(\beta^{1/\ell^{n+1}})^{\ell} = \beta^{1/\ell^n}$). The elements of $[\ell]^{-n}P$ are precisely the images under ϕ of the points

$$\{\beta^{1/\ell^n} q^{a/\ell^n} \zeta_{\ell^n}^b : 0 \le a, b < \ell^n\}.$$

Suppose that

$$Q = \phi(\beta^{1/\ell^n} q^{a/\ell^n} \zeta_{\ell^n}^b).$$

Since $\zeta_{\ell^n} \in \mathbb{C} \subseteq K_{\mathbb{C}}$ for all n and since $\beta^{1/\ell^n} \in R^* \subseteq \widehat{K}_{\mathbb{C}}$ for all n, we have

$$\widehat{K}_{\mathbb{C}}(Q) = \widehat{K}_{\mathbb{C}}(\beta^{1/\ell^n} q^{a/\ell^n} \zeta_{\ell^n}^b) = \widehat{K}_{\mathbb{C}}(q^{a/\ell^n}).$$

In particular, if q^a has order ℓ^m in $\widehat{K}_{\mathbb{C}}^*/(\widehat{K}_{\mathbb{C}}^*)^{\ell^n}$, then the conjugates of Q by the decomposition group are simply the elements of the form $\phi(\beta^{1/\ell^n} q^{a/\ell^n} \zeta_{\ell^n}^b \zeta_{\ell^m}^c)$, for $c \in \mathbb{Z}/\ell^m\mathbb{Z}$. In other words, the extension $\widehat{K}_{\mathbb{C}}(Q)/\widehat{K}_{\mathbb{C}}$ is a cyclic Galois extension of order ℓ^s, where $0 \le s \le n - \operatorname{ord}_\ell(v(q))$ is the greatest value such that $a \equiv 0 \pmod{\ell^s}$.

In particular, the quantity $a \in \mathbb{Z}/\ell^n\mathbb{Z}$ is an invariant of the decomposition orbit of $Q = \phi(\beta^{1/\ell^n} q^{a/\ell^n} \zeta_{\ell^n}^b)$ (although there may be more than one orbit with the same value a), and hence an invariant of the corresponding node at level n in the ramification tree. Furthermore, if $a' \in \mathbb{Z}/\ell^{n+1}\mathbb{Z}$ is the corresponding quantity for a node corresponding to the decomposition orbit of $Q' \in [\ell]^{-1}Q$, then $a' \equiv a \pmod{\ell^n}$. To describe the structure of the tree, we will set

$$m = \operatorname{ord}_\ell(v(q)) = \operatorname{ord}_\ell(v(j_E))$$

and say that a node has

(1) Type A if $n < m$;

(2) Type B_r if $n \ge m$ and $a \equiv 0 \pmod{\ell^{n-m}}$, and $0 \le r \le m$ is the greatest value with $a \equiv 0 \pmod{\ell^{n+r-m}}$; and

(3) Type C if $n \ge m$ and $a \not\equiv 0 \pmod{\ell^{n-m}}$.

Lemma 12. *The ramification tree observes the following rules*:

(1) *All nodes at level* 0 *through* $m-1$ *have Type* A (*if* $m = 0$, *then there are no such nodes*), *and each of these nodes have* ℓ^2 *distinct edges above them* (*necessarily each of weight* 1).

(2) *All nodes at level m are Type* B_r *nodes for some* $0 \le r \le m$, *and at any level of the tree*,

   (a) *if* $m = 0$, *then a Type* B_0 *node has* ℓ *Type* B_0 *nodes above it*, *with weight* 1 *each*, *and* $\ell - 1$ *Type* C *nodes above it*, *with weight* ℓ *each*;

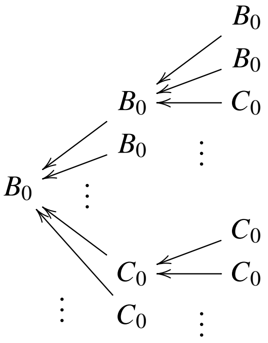
   (b) *if* $m \ge 1$, *then a Type* B_0 *node has* ℓ *Type* C *nodes above it*, *each with weight* ℓ;

   (c) *a Type* B_r *node*, *for* $1 \le r < m$, *has* ℓ^2 *Type* B_{r-1} *nodes above it*, *each with weight* 1;

   (d) *if* $m \ge 1$, *a Type* B_m *node has* ℓ *nodes of Type* B_m *above it*, *each of weight* 1, *and* $\ell - 1$ *nodes of Type* B_{m-1} *above it*, *each of weight* ℓ.

(3) *Each Type* C *node has* ℓ *Type* C *nodes above it*, *each with weight* ℓ.

Thus, when $\mathrm{ord}_\ell(j_E) = 0$, the ramification tree looks something like this:



We should point out that the lemma above does not uniquely define the structure of the tree, but only gives the information that we will need in the proof of Theorem 1. In the proof below, however, it is pointed out that there are precisely $\ell^m\varphi(\ell^{m-r})$ nodes of Type B_r at level m for each $0 \le r \le m$, where φ is the Euler totient function. This fact (combined with Lemma 12) uniquely determines the full structure of the tree.

Proof. The claim that all nodes at levels 0 through $m-1$ have Type A is clear from the definition, as is the claim that all nodes at level m have Type B_r for some $0 \le r \le m$ (since the condition $a \equiv 0 \pmod{\ell^0}$ is satisfied trivially). We now conduct a case-by-case examination of the nodes immediately above nodes of any given type, proving the various claims in the lemma. Throughout, we take $Q = \phi(\beta^{1/\ell^n} q^{a/\ell^n} \zeta_{\ell^n}^b)$ as a representative of the appropriate decomposition orbit, and consider the decomposition orbits of possible

$$Q' = \phi(\beta^{1/\ell^{n+1}} q^{a'/\ell^{n+1}} \zeta_{\ell^{n+1}}^{b'}) \in [\ell]^{-1}Q.$$

Case: Type A. First, suppose that $n < m$. Then $q^{a'/\ell^{n+1}}$ is $\widehat{K}_{\mathbb{C}}$-rational for all $a' \in \mathbb{Z}/\ell^{n+1}\mathbb{Z}$, since $q \in (\widehat{K}_{\mathbb{C}}^*)^{\ell^m}$ and $n+1 \le m$. In particular, every extension of the form $\widehat{K}_{\mathbb{C}}(q^{a'/\ell^{n+1}})/\widehat{K}_{\mathbb{C}}(q^{a/\ell^n})$ is trivial, and so there are ℓ^2 distinct decomposition orbits above Q. This gives ℓ^2 distinct nodes above the node corresponding to Q, and each must be unramified, since their ramification indices sum to ℓ^2 (or, simply because the inertia groups are trivial). As noted above, if $n+1 < m$, these points are all Type A, while if $n+1 = m$, the points above Q have Type B_r for some r. It is, in fact, quite easy to compute the number of nodes of Type B_r at level m for each $0 \le r \le m$, thereby completely describing the structure of the tree. The nodes of Type B_r are in one-to-one correspondence with pairs (a, b) with $b \in \mathbb{Z}/\ell^m\mathbb{Z}$, and $a \in \mathbb{Z}/\ell^m\mathbb{Z}$ of the form $\ell^r u$ for $u \in \mathbb{Z}/\ell^m\mathbb{Z}$ a unit. In other words, there are precisely $\ell^m\varphi(\ell^{m-r})$ nodes of Type B_r at level m for each $0 \le r \le m$.

Case: Type B_r. Now, suppose that $n \geq m$ and that $a \equiv 0 \pmod{\ell^{n-m}}$, so that Q has Type B_r for some $0 \leq r \leq m$. Since $q \in (\widehat{K}_{\mathbb{C}}^*)^{\ell^m}$, say $q = q_0^{\ell^m}$, we see that the field

$$\widehat{K}_{\mathbb{C}}(Q) = \widehat{K}_{\mathbb{C}}(q^{a/\ell^n}) = \widehat{K}_{\mathbb{C}}(q_0^{a/\ell^{n-m}})$$

is simply $\widehat{K}_{\mathbb{C}}$. Now let $Q' = \phi(\beta^{1/\ell^{n+1}} q^{a'/\ell^{n+1}} \zeta_{\ell^{n+1}}^{b'})$ be an element of $\ell^{-1}Q$, so that $a' \equiv a \pmod{\ell^n}$ and $b' \equiv b \pmod{\ell^n}$. Consider the extension

$$\widehat{K}_{\mathbb{C}}(Q')/\widehat{K}_{\mathbb{C}}(Q) = \widehat{K}_{\mathbb{C}}(q^{a'/\ell^{n+1}})/\widehat{K}_{\mathbb{C}}(q^{a/\ell^n}).$$

If $a' \equiv 0 \pmod{\ell^{n+1-m}}$, then this extension is trivial, and each Q' of this form is fixed by the decomposition group. To each of these points (if there are any), corresponds a prolongation of w to $K_{\mathbb{C}}(Q')$ which is unramified. If $a' \not\equiv 0 \pmod{\ell^{n+1-m}}$, then this extension is cyclic of degree ℓ. In this case, the decomposition orbit of Q' corresponds to a prolongation of w to $K_{\mathbb{C}}(Q')$ which is ramified of index ℓ.

Now, if Q has Type B_r for some $r \geq 1$, then $m \geq 1$ and $a \equiv 0 \pmod{\ell^{n+1-m}}$. So for all points Q' above Q, we have $a' \equiv a \equiv 0 \pmod{\ell^{n+1-m}}$ since $a' \equiv a \pmod{\ell^n}$ and $n+1-m \leq n$. In this case, all points Q' will be fixed by the decomposition group, and so correspond to unramified prolongations of w to $K_{\mathbb{C}}(Q')$. If $r < m$, we have $a' \equiv a \not\equiv 0 \pmod{\ell^{n+r+1-m}}$, and so each $Q' \in \ell^{-1}Q$ is a point of Type B_{r-1}. If, on the other hand, $r = m$, then have $a \equiv 0 \pmod{\ell^n}$. There is one choice $a' \in \mathbb{Z}/\ell^{n+1}\mathbb{Z}$ with $a' \equiv a \pmod{\ell^n}$ and $a' \equiv 0 \pmod{\ell^{n+1}}$, and $\ell - 1$ choices with $a' \not\equiv 0 \pmod{\ell^{n+1}}$. By the same argument, the choice $a' = 0$ yields ℓ decomposition-invariant points, and so ℓ distinct, unramified prolongations of w. Each of those points will again have Type B_m. The $\ell - 1$ choices with $a' \neq 0$ each yield a single decomposition orbit, and so a single prolongation of w with ramification index ℓ. Since $a' \equiv 0 \pmod{\ell^n}$ but $a' \not\equiv 0 \pmod{\ell^n}$, these $\ell - 1$ points have Type B_{m-1}.

Now suppose that Q has Type B_0, which we will subdivide into two cases, depending on whether or not $m = 0$. If $m \geq 1$, then $a \equiv 0 \pmod{\ell^{n-m}}$, but $a \not\equiv 0 \pmod{\ell^{n+1-m}}$. It follows that any point $Q' \in \ell^{-1}Q$, which must satisfy $a' \equiv a \pmod{\ell^n}$, satisfies $a' \not\equiv 0 \pmod{\ell^{n-m}}$. These points, then, are all Type C. For each of these values of $a' \in \mathbb{Z}/\ell^{n+1}/\mathbb{Z}$, we find that Q' is in a decomposition orbit of size ℓ. Thus, there are ℓ prolongations of w, each with ramification index ℓ. On the other hand, if $m = 0$, then the condition on a is simply that $a \equiv 0 \pmod{\ell^n}$. Consider the different values $a' \in \mathbb{Z}/\ell^{n+1}\mathbb{Z}$ with $a' \equiv a \pmod{\ell^n}$. If $a' = 0$, then the decomposition group acts trivially on Q'. This gives ℓ distinct prolongations of w, each with ramification index ℓ. Each of the points corresponding to these prolongations has Type B_0. The $\ell - 1$ choices of $a' \neq 0$ each yield one prolongation, with Type C.

Case: Type C. In this case, we have $a \not\equiv 0 \pmod{\ell^{n-m}}$, and so if $a' \in \mathbb{Z}/\ell^{n+1}\mathbb{Z}$, with $a' \equiv a \pmod{\ell^n}$, we have $a' \not\equiv 0 \pmod{\ell^{n+1-m}}$. Thus, any point $Q' \in \ell^{-1}Q$ is a Type C point. Now, any of the extensions $\widehat{K}_{\mathbb{C}}(q^{a'/\ell^{n+1}})/\widehat{K}_{\mathbb{C}}(q^{a/\ell^n})$ must be cyclic of order ℓ, since $q^{a'}$ generates a cyclic subgroup of $\widehat{K}_{\mathbb{C}}^*/(\widehat{K}_{\mathbb{C}}^*)^{\ell^{n+1}}$ of order ℓ^{n+1-m}, while q^a generates a cyclic subgroup of $\widehat{K}_{\mathbb{C}}^*/(\widehat{K}_{\mathbb{C}}^*)^{\ell^n}$ of order ℓ^n. Thus, there are ℓ primes above w, each with ramification index ℓ, and corresponding to a point of Type C. □

5. Proof of Theorem 1

We proceed with the proof of the main result, first making several reductions to simplify the argument. For simplicity, we introduce the following notation

$$X(k, B_1, \mathscr{E}, P, t) = \left\{Q \in \mathscr{E}_t(\bar{k}) : [k(Q) : k] \leq B_1 \text{ and } \ell^n Q = P_t \text{ for some } n \geq 1\right\}.$$

Our first lemma shows that we are free to replace $\mathscr{E} \to C$ by some base extension defined over an algebraic extension of k. In other words, by thinking in terms of the generic fibre E/K, it suffices to prove Theorem 1 for E over an algebraic extension K'/K.

Lemma 13. *Let k'/k be an algebraic extension, and let $\phi : C' \to C$ be a dominant morphism of curves defined over k'. If Theorem 1 is true for an elliptic surface $\mathscr{E}' \to C'$ birational to $\mathscr{E} \times_C C'$, then it is true for $\mathscr{E}$.*

Proof. In proving Theorem 1, it is clear that we are proving something stronger if we pass to an algebraic extension of k. So we will, replacing k by k' if necessary, suppose that $\phi : C' \to C$ is defined over k. Now let $\mathscr{E}'$ be birational to $\mathscr{E} \times_C C'$, and let $P : C \to \mathscr{E}$ be a section. Then P lifts uniquely to a section $P' : C' \to \mathscr{E}'$, and since $\mathscr{E}'_t \cong \mathscr{E}_{\phi(t)}$ over k for all but finitely many $t \in C'(\bar{k})$, we have immediately

$$X(k, B_1, \mathscr{E}, P, \phi(t)) = X(k, B_1, \mathscr{E}', P', t).$$

Now, for any $s \in C(\bar{k})$, there is some $t \in C'(\bar{k})$ with $\phi(t) = s$, and we have $[k(t) : k] \leq \deg(\phi)[k(s) : k]$. Thus, we may conclude Theorem 1 for $\mathscr{E}$ and P, with $B_1 = D_1$ and $B_2 = D_2$ by applying Theorem 1 to $\mathscr{E}'$ and P', with $B_1 = D_1$ and $B_2 = \deg(\phi)D_2$. □

From this point forward, we will assume that $\mathscr{E}$ has been replaced with an appropriate base extension $\mathscr{E}'$, so that there is at least one place $v \in C(\bar{k})$ such that the fibre $\mathscr{E}_v$ is multiplicative.

Lemma 14. *Let k, $\mathscr{E}$, and ℓ be as in Theorem 1, and let $N \geq 1$. If Theorem 1 is true for $P = NP'$, then it is true for $P = P'$.*

Proof. Suppose there is some point $Q \in \mathscr{E}_t(\bar{k})$ with $\ell^n Q = P'_t$. Then $\ell^n NQ = NP'_t$, and clearly $k(NQ) \subseteq k(Q)$ (since multiplication by N is given by rational maps defined over k). In particular, in the notation above, the image of $X(k, B_1, \mathscr{E}, P', t)$ by the morphism $[N] : \mathscr{E}_t \to \mathscr{E}_t$ is wholly contained in the set $X(k, B_1, \mathscr{E}, NP', t)$. But this means that the cardinality of the former set is at most N^2 times the cardinality of the latter. □

Lemma 15. *Theorem 1 is true for P a point of finite order on $E(K)$.*

Proof. In light of Lemma 14, we may replace P by NP for any $N \geq 1$, and so it suffices to consider the case where $P = \mathbb{O}$ is the identity section. But

$$\left\{Q \in \mathscr{E}_t(\bar{k}) : \ell^n Q = \mathbb{O}_t \text{ for some } n \geq 1\right\}$$

is just $\mathscr{E}_t[\ell^\infty] \setminus \{\mathbb{O}_t\}$, and so in particular is a subset of $\mathscr{E}_t(\bar{k})_{\mathrm{Tors}}$.

A simple modification of a theorem of Merel (for the appropriate variant, see [Silverman 2007, Corollary 6.64]) implies that for all elliptic curves E/F, where $F/\mathbb{Q}$ is a number field, the set

$$\bigcup_{[L:F] \leq n} E(L)_{\mathrm{Tors}}$$

is bounded in cardinality by some quantity that depends only on n and $[F : \mathbb{Q}]$. Since the (smooth) fibres $\mathscr{E}_t$ for $[k(t) : k] \leq B_2$ are all elliptic curves defined over number fields of degree at most $B_2[k : \mathbb{Q}]$, invoking the claim above with $n = B_1 B_2 [k : \mathbb{Q}]$ gives us a uniform bound on the sizes of the sets $X(k, B_1, \mathscr{E}, \mathbb{O}, t)$, for $t \in C(\bar{k})$ with $[k(t) : k] \leq B_2$. □

We now prove Lemma 5, relying heavily on [Faber et al. 2009, Lemma 4.5].

Proof of Lemma 5. For any morphism of curves $\phi : X \to Y$ with ramification divisor R_ϕ, we set

$$\rho(\phi) = \frac{\deg R_\phi}{2 \deg \phi}$$

and recall that we have assumed that there exist constants $c_1 > 0$ and c_2 such that

$$\rho(\phi_n) \geq c_1 2^n - c_2.$$

Lemma 4.5 of [Faber et al. 2009] states that if

$$X_0 \xleftarrow{\phi_1} X_1 \xleftarrow{\phi_2} \cdots \xleftarrow{\phi_N} X_N$$

is a tower of (smooth projective) curves equipped with nonconstant morphisms all defined over a number field k, then, with

$$B_N = \min_{1 \leq m \leq N} 2^{N-m} \rho(\phi_m) \quad \text{and} \quad b_N = \min_{1 \leq m \leq N} \rho(\phi_m), \tag{2}$$

the set

$$\{P \in X_N(\bar{k}) : [k(P):k] < B_N \quad \text{and} \quad [k(\phi_1 \circ \cdots \circ \phi_N(P)):k] \geq b_N\}$$

is finite. Note that if any of the maps ϕ_n is unramified, the result is trivial, since in this case $B_N = b_N = 0$. Thus, we fix n_0 such that $\rho(\phi_n) > \frac{1}{2}c_1 2^n$ for $n > n_0$, (for example, we could take n_0 to be anything larger than $\log_2(2c_2/c_1)$, assuming $c_2 > 0$, or $n_0 = 0$ otherwise). We then apply the lemma to the tower of curves

$$\mathbb{P}^1 \xleftarrow{\psi_1} \mathbb{P}^1 \xleftarrow{\psi_2} C_{n_0} \xleftarrow{\psi_3=\phi_{n_0+1}} C_{n_0+1} \xleftarrow{\psi_4=\phi_{n_0+2}} \cdots ,$$

where the two leftmost maps are any morphisms of degree at least 2 (defined over k). On the one hand, the conditions on ϕ_n and the fact that ψ_1 and ψ_2 are not unramified ensure that $\rho(\psi_n) \geq \epsilon 2^n$ for some $\epsilon > 0$. Thus,

$$B_N = \min_{1 \leq m \leq N} 2^{N-m}\rho(\psi_m) \geq \min_{1 \leq m \leq N} 2^{N-m}\epsilon 2^m = \epsilon 2^N \quad \text{for any } N \geq 1.$$

On the other hand, $\psi_1 : \mathbb{P}^1 \to \mathbb{P}^1$, and so the Hurwitz formula guarantees that

$$b_N \leq \rho(\psi_1) = \frac{2d-2}{2d} < 1.$$

By the lemma from [Faber et al. 2009], we know that for any $N > n_0$, the set

$$\{P \in C_N(\bar{k}) : [k(P):k] < \delta 2^N \quad \text{and} \quad [k(\psi_1 \circ \cdots \circ \psi_N(P)):k] \geq 1\}$$

is finite, with $\delta = \epsilon 2^{2-n_0}$. Since the condition $[k(\psi_1 \circ \cdots \circ \psi_N(P)):k] \geq 1$ is trivial, we have proven the lemma (with the more explicit information that it suffices to take N larger than $\log_2(B/\delta)$). □

We now proceed with the body of the proof of the main result. Let k, $\mathscr{E}$, P, ℓ, B_1, and B_2 be as in the statement of Theorem 1. In light of Lemma 15, we will suppose that P is a section of infinite order. The j-invariant $j_{\mathscr{E}} : C \to \mathbb{P}^1$ was assumed nonconstant, and hence is dominant. In particular, over $\mathbb{C}$, it must have a pole. If necessary, we may replace $\mathscr{E}$ with an elliptic surface $\mathscr{E}'$, birational to a base extension of $\mathscr{E}$, such that this pole is a place of multiplicative reduction for $\mathscr{E}'$. After replacement of $\mathscr{E}$ with this base extension (employing Lemma 13), the generic fibre E of $\mathscr{E}$ over $K = k(C)$ has split multiplicative reduction at v over $K_{\mathbb{C}} = K \otimes_k \mathbb{C}$. Finally, in light of Lemma 14, we will also replace P by $v(j_E)P$ in order to ensure that $P \in E_0(\widehat{K_{\mathbb{C}}})$, where $\widehat{K_{\mathbb{C}}}$ is the completion of $K_{\mathbb{C}}$ at v.

Let $\Gamma_0 \subseteq \mathscr{E}$ be the image of P and, as in the introduction, let $\Gamma_{n+1} = [\ell]^{-1}\Gamma_n$. If $t \in C(\bar{k})$ has $[k(t):k] \leq B_1$, and $Q \in \mathscr{E}_t(\bar{k})$ has $[k(Q):k] \leq B_2$, and $\ell^n Q = P_t$, then the pair (Q, t) corresponds to a point on Γ_n of degree at most $B = B_1 B_2$ over k, and hence a point of degree at most B on the normalization $\widetilde{\Gamma}_n$. We will use Lemma 5 to bound n.

First, note that since the tree of components of the curves $\widetilde{\Gamma}_n$ contains only finitely many paths (by Lemma 10), it suffices to prove the result for any of the finitely many distinct towers

$$C_0 \xleftarrow{\phi_0} C_1 \xleftarrow{\phi_1} \cdots,$$

where $C_n \subseteq \widetilde{\Gamma}_n$ is irreducible and ϕ_n is the map induced by $[\ell] : \mathscr{E} \to \mathscr{E}$. In particular, we may choose n_0 large enough that for all $n \geq n_0$, C_{n+1} is the *only* component of $\widetilde{\Gamma}_{n+1}$ which maps to C_n by the map $\widetilde{\Gamma}_{n+1} \to \widetilde{\Gamma}_n$ induced by $[\ell] : \mathscr{E} \to \mathscr{E}$. Now suppose that w is a place above v corresponding to a point on C_{n_0}. We claim that there is a place $w' \,|\, w$, corresponding to a point of a curve of C_n for some $n \geq n_0$, that has Type C (in the sense of Section 4). Of course, if w itself has Type C, then we're done. Suppose that the w is a node in the ramification tree of Type B_r for some $0 \leq r \leq m$. By Lemma 12, then, there is a node of Type C above w if w has Type B_0, and a node of Type B_{r-1} above w otherwise. By induction, then, we eventually have a node $w' \,|\, w$ of Type C in the ramification tree. Finally, if w has Type A, then there is some node of Type B_0 above w, and we apply the previous case.

So, increasing n_0 if necessary, and replacing w with the node of Type C above it, we may simply assume that w has Type C. Now, every Type C node in the ramification tree splits into ℓ Type C nodes with ramification index ℓ, at the next level of the tree. Thus, the curve C_{n_0+m} contains ℓ^m distinct points above w, all of Type C. In other words, the ramification divisor of the morphism

$$\phi_{n_0+m} : C_{n_0+m} \to C_{n_0+m+1}$$

has degree at least $\ell^m(\ell - 1)$. Since each of the maps ϕ_n has degree at most ℓ^2, it follows that for $n > n_0$,

$$\rho(\phi_n) \geq \frac{\ell^{n-n_0}(\ell-1)}{\ell^2} \geq \ell^n \left(\frac{\ell - 1}{\ell^{n_0+2}} \right).$$

Since there are only finitely many $n \leq n_0$, this shows that there are constants $c_1 > 0$ and c_2 such that

$$\rho(\phi_n) \geq c_1 2^n - c_2$$

(in fact, we could replace the 2^n with ℓ^n, but this provides no gains after the application of Lemma 5). We are now in a position to apply Lemma 5. The lemma tells us that for any $B \geq 1$, there is an $N(B)$ such that $C_{N(B)}(\bar{k})$ contains only finitely many points z with $[k(z) : k] \leq B$. Applying the argument to each of the finitely many towers of components of the curves $\widetilde{\Gamma}_n$ proves the same thing for those curves.

Given $B \geq 1$, let $Y_B \subseteq C(\bar{k})$ be the (finite) set of t corresponding to points in $\Gamma_{N(B)}(\bar{k})$ of degree at most B. Now suppose that $t \in C(\bar{k})$ and $Q \in \mathscr{E}_t(\bar{k})$, with

$\ell^n Q = P_t$ and $[k(t, Q) : k] \leq B$. If $t \notin Y_B$, then $n \leq n_0$. The number of points Q is at most

$$\begin{aligned}\#\ell^{-1}P_t + \#\ell^{-2}P_t + \cdots + \#\ell^{-n_0}P &= \ell^2 + \ell^4 + \cdots + \ell^{2n_0(B)} \\ &= \frac{\ell^{2n_0(B)+1} - \ell^2}{\ell^2 - 1},\end{aligned}$$

which does not depend on t.

Now suppose that $t \in Y_B$ and that $\mathscr{E}_t$ is nonsingular. If $\hat{h}_t : \mathscr{E}_t(\bar{k}) \to \mathbb{R}^+$ is the Néron–Tate height on the elliptic curve $\mathscr{E}_t$, then for any $Q \in \mathscr{E}_t(\bar{k})$ with $\ell^n Q = P_t$ for some $n \geq 1$, we have

$$\hat{h}_t(Q) = \ell^{-2n}\hat{h}_t(P_t) \leq \hat{h}_t(P_t).$$

Thus, there are a finite number of points in $\mathscr{E}_t(\bar{k})$ such that $\ell^n Q = P_t$ for some $n \geq 1$, and $[k(Q) : k] \leq B_1$. So, since Y_B is finite, we have an upper bound on the size of the set

$$\left\{Q \in \mathscr{E}_t(\bar{k}) : [k(Q) : k] \leq B_1 \text{ and } \ell^n Q = P_t \text{ for some } n \geq 1\right\}$$

for $t \in C(\bar{k})$ with $[k(t) : k] \leq B_2$, whether $t \in Y_B$ or not. This proves Theorem 1.

Remark. It is natural to ask how far one might extend this argument. In particular, if we fix a finite set of primes and let Z be the set of positive integers divisible only by these primes, then it is not hard to see that a slight modification of the proof of Theorem 1 shows that the sets

$$\left\{Q :\in \mathscr{E}_t(\bar{k}) : nQ = P_t \text{ for some } n \in Z \text{ and } [k(Q) : k] \leq B_1\right\}$$

are uniformly bounded in size for $t \in C(\bar{k})$ of degree at most B_2. Indeed, if $\Omega(n)$ denotes the number of prime divisors of n (counted with multiplicity), then it should be possible to modify proofs above to show that the curves

$$\Gamma(n) = [n]^{-1}C \subseteq \mathscr{E}$$

satisfy $\rho(\phi) \gg 2^{\Omega(n)}$ for $\phi : \Gamma(\ell n) \to \Gamma(n)$ the morphism induced by $[\ell]$. In other words, it should follow by a simple modification of Lemma 5 (which turns out to depend only very minimally on the particular tower of curves) that $\Gamma(n)$ contains only finitely many points of degree at most $B = \max\{B_1, B_2\}$, so long as $\Omega(n)$ is large enough. The problem is that since there are infinitely many integers n with $\Omega(n) = M$, say, the set Y_B of points on C above which there is a point on some $\Gamma(n)$ with $\Omega(n) = M$ might still be infinite. Since we cannot reduce the problem to one on finitely many fibres, the final step of the proof of Theorem 1will not apply.

6. The proof of Theorems 2 and 3

As mentioned in the introduction, the set of points $Q \in \mathscr{E}_t(\bar{k})$ of bounded degree (indeed, of degree 1), such that $\ell^N Q = P_t$ for some N, can be made arbitrarily large by simply choosing P to be divisible in $\mathscr{E}(C)$ be a large power of ℓ. Requiring that ℓ be nonspecial, we can prove the stronger claim of Theorem 2, namely that if P is not of the form ℓP_0 for any section $P_0 : C \to \mathscr{E}$, then

$$\#\{Q \in \mathscr{E}_t(k) : \ell^n Q = P_t \text{ for some } n \geq 1\} \leq \ell^2,$$

for all but finitely many places $t \in C(k)$, provided that $\mathscr{E}$ has at least one multiplicative fibre. Taking a cue from work of Baragar and McKinnon [2010], we note that we may replace the upper bound with 0 if $\mathscr{E}$ has at least 5 distinct multiplicative fibres (4 multiplicative fibres suffice if $\ell = 5$, or 3 if $\ell \geq 7$). We should note that since the group of sections on $\mathscr{E}$ is finitely generated, P is an ℓ-th multiple of another section only for finitely many primes ℓ.

The result follows from the Mordell conjecture (now a theorem of Faltings), once one shows that the curve $\widetilde{\Gamma}_2$ (or, with the additional hypotheses, $\widetilde{\Gamma}_1$) is irreducible, and has genus at least 2. For if this is the case, then there are only finitely many fibres on which $[\ell]^{-2}P_t$ (respectively, $[\ell]^{-1}P_t$) contains any k-rational points at all. The result follows since $[\ell]^{-1}P_t$ contains at most ℓ^2 points. Thus, Theorem 2 is proven once we establish this:

Lemma 16. *Let ℓ be a nonspecial prime for $\mathscr{E}$, and let $P : C \to \mathscr{E}$ be a section which is* not *an ℓ-th multiple, and suppose that $j_{\mathscr{E}}$ is nonconstant. Then $\widetilde{\Gamma}_2$ is an irreducible curve of genus at least* 4. *If we suppose, additionally, that $j_{\mathscr{E}} : C \to \mathbb{P}^1$ has at least* 5 *distinct poles over $\mathbb{C}$ (at least* 4 *poles if $\ell = 5$, or at least* 3 *poles if $\ell \geq 7$), then $\widetilde{\Gamma}_1$ is an irreducible curve of genus at least* 2.

Proof. The irreducibility follows from the results in Lemma 10. In particular, since ℓ is not a special prime, Lemma 6 tells us that Galois group of the covering $\widetilde{\Gamma}_n \to \Gamma_0$ is isomorphic (in the natural way) to a semidirect product $(\mathbb{Z}/\ell^n\mathbb{Z})^2 \rtimes \mathrm{SL}_2(\mathbb{Z}/\ell^n\mathbb{Z})$. In particular, the action is transitive, and the curve $\widetilde{\Gamma}_n$ is irreducible.

For convenience, we will work in the function field setting, considering the generic fibre $E/K_{\mathbb{C}}$. We will denote the function fields (over $\mathbb{C}$) of $\widetilde{\Gamma}_1$ and $\widetilde{\Gamma}_2$ by F_1 and F_2, respectively.

Since $j_{\mathscr{E}} : C \to \mathbb{P}^1$ is not constant, it is dominant, and so must have a pole. Let v be a pole of $j_{\mathscr{E}} : C \to \mathbb{P}^1$, and suppose for the time being that E has multiplicative reduction at v. If we have $P \in E_0(\widehat{K}_{\mathbb{C}})$, then by Lemma 12 (since $v(j_{\mathscr{E}})$ is prime to ℓ), the place v is a point of Type B_0 in the terminology of Section 4. A quick examination of the structure of the tree (referring to Lemma 12) shows that there are ℓ places of Type B_0 of F_1 above v, and above each of these there are $\ell - 1$ places w of F_2 with $e_w(F_2/F_1) = \ell$. Similarly, there are $\ell - 1$ places of Type C for F_1 above v, and above each of these, ℓ places w of F_2 with $e_w(F_2/F_1) = \ell$. Thus the map (of degree ℓ^2) $\widetilde{\Gamma}_2 \to \widetilde{\Gamma}_1$ has ramification divisor of degree at least $2\ell(\ell-1)^2$. By the Riemann–Hurwitz formula,

$$2\left(g(\widetilde{\Gamma}_2)-1\right) \geq \ell^2 2\left(g(\widetilde{\Gamma}_1)-1\right)+2\ell(\ell-1)^2,$$

or, using the trivial bound $g(\widetilde{\Gamma}_1) \geq 0$,

$$g(\widetilde{\Gamma}_2) \geq \ell^3 - 3\ell^2 + \ell + 1 \geq 4$$

(recall that $\ell \geq 3$).

If $P \notin E_0(\widehat{K}_{\mathbb{C}})$, then we cannot apply Lemma 12. However, the general approach of Section 4 still applies. If

$$0 \to q^{\mathbb{Z}} \to \widehat{K}_{\mathbb{C}} \xrightarrow{\phi} E(\widehat{K}_{\mathbb{C}}) \to 0$$

is the Tate uniformization of E at v, as in Section 4, then set $m = v(q) = -v(j_{\mathscr{E}})$ and write $q = q_0^m$ for some $q_0 \in \widehat{K}_{\mathbb{C}}^*$. Since q_0 is a uniformizer for v, we may write $P = \phi(uq_0^p)$ for some $0 < p < m$ and some v-unit u (we may take $0 < p < m$ because $q^{\mathbb{Z}} = \ker(\phi)$). The places of F_2 above v correspond to decomposition orbits (relative to a fixed prolongation of v) of points of the form

$$Q' = \phi(u^{1/\ell^2} q_0^{(p+am)/\ell^2} \zeta_{1/\ell^2}^b) \quad \text{for } a \text{ and } b \in \mathbb{Z}/\ell^2\mathbb{Z}.$$

By hypothesis, m is prime to ℓ, and so the function $q \mapsto p + am$ simply permutes $\mathbb{Z}/\ell^2\mathbb{Z}$. In other words, the places of F_2 above v simply correspond to the decomposition orbits of points of the form

$$Q' = \phi(u^{1/\ell^2} q_0^{a/\ell^2} \zeta_{1/\ell^2}^b) \quad \text{for } a \text{ and } b \in \mathbb{Z}/\ell^2\mathbb{Z}.$$

Exactly as in Section 4, the $\ell(\ell-1)$ choices of a such that $a \not\equiv 0 \pmod{\ell^2}$ each yield a place w of F_2 for which $e_w(F_2/F_1) = \ell$. The $\ell - 1$ choices of $a \equiv 0 \pmod{\ell}$ but $a \not\equiv 0 \pmod{\ell^2}$ give $\ell(\ell-1)$ places w of F_2 for which $e_w(F_2/F_1) = \ell$. Just as in the previous case, we obtain

$$g(\widetilde{\Gamma}_2) \geq \ell^3 - 3\ell^2 + \ell + 1 \geq 4.$$

Now we suppose that v is a pole of $j_{\mathscr{E}} : C \to \mathbb{P}^1$ but that v is a place of additive reduction. Then there is some quadratic extension $K'/K_{\mathbb{C}}$ and an elliptic curve E'/K' which is K'-isomorphic to E, such that E' has multiplicative reduction at v; see [Silverman 1994, p. 442]. Moreover, $v = (v')^2$ ramifies in this extension. Let $Q_1, Q_2 \in E'(\overline{K'})$ with $\ell^2 Q_2 = \ell Q_1 = P$. We have shown that there are ℓ places w of $K'(Q_1)$ above v' with $e_w(K'(Q_1)/K') = 1$, and $\ell - 1$ with $e_w(K'(Q_1)/K') = \ell$. In other words, there are ℓ places w of $K'(Q_1)$ above v with $e_w(K'(Q_1)/K) = 2$, and $\ell - 1$ with $e_w(K'(Q_1)/K) = 2\ell$. But $K'(Q_1) = K'F_1$, and so a prime of $K'(Q_1)$ is totally ramified in the extension K'/K only if it is in the extension $K'(Q_1)/F_1$. In particular, every prime w of $K'(Q_1)$ above v satisfies $e_w(K'(Q_1)/F_1) = 2$. Thus each place w of F_1 above v extends uniquely to a place w' of $K'(Q_1)$, and we have $e_w(F_1/K) = \frac{1}{2}e_{w'}(K'(Q_1)/K)$. It follows that there are ℓ places w of F_1 with $e_w(F_1/K)$, and $\ell - 1$ with $e_w(F_1/K) = \ell$. After conducting the same analysis for the extension F_2/F_1, we see that the factorizations of v in these extensions are identical to the previous case. The Hurwitz formula again gives $g(\widetilde{\Gamma}_2) \geq 4$.

Now we consider the genus of $\widetilde{\Gamma}_1$ in terms of the number of poles of $j_{\mathscr{E}}$. By the arguments above, if v is a pole of $j_{\mathscr{E}}$, then there are at least $\ell - 1$ places of $\widetilde{\Gamma}_1$ above v at which the map $\widetilde{\Gamma}_1 \to \widetilde{\Gamma}_0$ is ramified with index ℓ. Thus, the ramification divisor of the map $\widetilde{\Gamma}_1 \to \widetilde{\Gamma}_0$ has degree at least $N(\ell - 1)^2$, where N is the number of distinct poles of $j_{\mathscr{E}}$ (over $\mathbb{C}$). By The Riemann–Hurwitz formula (since the morphism $\widetilde{\Gamma}_1 \to \widetilde{\Gamma}_0$ has degree ℓ^2), we have

$$g(\widetilde{\Gamma}_1) \geq 1 - \ell^2 + \tfrac{1}{2}N(\ell - 1)^2.$$

This is of no use to us if $N \leq 2$, but one can check that if $N \geq 5$, then $g(\ell) \geq 2$ for all $\ell \geq 3$. Furthermore, if $N \geq 4$, then $g(\ell) \geq 8$ for $\ell \geq 5$, and if $N \geq 3$, then $g(\ell) \geq 6$ for all $\ell \geq 7$.

If $\Gamma_0 \cong C$ has genus at least 1, then the estimate on the ramification of the map $\widetilde{\Gamma}_1 \to \widetilde{\Gamma}_0$ and the Hurwitz formula, gives

$$g(\widetilde{\Gamma}_1) \geq 1 + \tfrac{1}{2}(\ell - 1)^2 \geq 3 \quad \text{for } \ell \geq 2,$$

regardless of the number of poles of $j_{\mathscr{E}}$ (provided that $j_{\mathscr{E}}$ is nonconstant). □

The remark after the statement of Theorem 2 is proved by a similar argument. In particular, in this simplified case one knows that all of the curves $\widetilde{\Gamma}_n$ are irreducible, and an examination of the ramification tree shows that the map

$$\phi_{n+1} : \widetilde{\Gamma}_{n+1} \to \widetilde{\Gamma}_n$$

has ramification of degree at least $(n + 1)\ell^n(\ell - 1)^2$. Lemma 4.5 of [Faber et al. 2009] now implies an upper bound on n such that $[\ell]^{-n}P_t$ contains points of degree at most D on infinitely many fibres. The lower bound comes from observing that the points in $[\ell]^{-n}P_t$ each have degree at most ℓ^{2n} over k.

Proof of Theorem 3. Let $\mathscr{E}$ and S be as in the statement of the theorem, and suppose that $\mathscr{E}_t(k)$ contains a point of order N for N not an S-unit. Then $\mathscr{E}_t(k)$ contains a point of order ℓ for some prime $\ell \notin S$. Note that, by Merel's theorem ℓ is bounded in terms of $[k : \mathbb{Q}]$. Thus, it suffices to show the finiteness of the set of $t \in C(k)$ such that $\mathscr{E}_t(k)$ contains a point of order ℓ for any given $\ell \notin S$.

Now, fix $\ell \notin S$. Since $E[\ell] \setminus \{\mathbb{O}\}$ is Galois-irreducible, the curve $[\ell]^{-1}\mathbb{O}$ has two components, one of which is birational to C (this is the curve on $\mathscr{E}$ corresponding to $\mathbb{O}$). Let C' be the normalization of the component birational to C, and let Γ be the normalization of the other component. If $\ell \geq 13$, then using the obvious map $\Gamma \to X_1(\ell)$, we see that $g(\Gamma) \geq 2$, and hence $\Gamma(k)$ is finite. Thus we only have something to prove for $\ell = 3, 5, 7$.

If $j_{\mathscr{E}}$ has a pole at the place v on C, then (by the same argument as in the proof of Theorem 2) there are a total of ℓ places above v, on the union of these two curves, at which the map induced by $[\ell]$ is unramified, and $\ell - 1$ at which the map has ramification index ℓ. Since $C' \cong C$, all of the ramified places must lie on Γ, and so the places above v contribute $(\ell - 1)^2$ to the degree of the ramification divisor of the map $\Gamma \to C$ (this map has degree $\ell^2 - 1$).

If C has genus 1 (or greater), then the existence of a single pole of $j_{\mathscr{E}}$ gives the lower bound

$$g(\Gamma) \geq 1 + \tfrac{1}{2}(\ell - 1)^2 \geq 3 \quad \text{for } \ell \geq 3.$$

Otherwise, the existence of N distinct poles of $j_{\mathscr{E}} : C \to \mathbb{P}^1$ gives a lower bound of

$$g(\Gamma) \geq 1 - (\ell^2 - 1) + \tfrac{N}{2}(\ell - 1)^2 \geq 3 \quad \text{for } N \geq 5 \text{ and } \ell \geq 3. \qquad \square$$

7. The proof of Theorem 4

Let E/k be an elliptic curve defined over a number field, and fix a rational prime ℓ and a value $M \geq 0$. Let $n \geq 1$ be the least integer such that $M < \ell^{2n}$, and let $F = k(E[\ell^n], [\ell]^{-n}P)$. Then F/k is a Galois extension, and if $\mathfrak{p}$ is a prime of k whose Frobenius element in $\mathrm{Gal}(F/k)$ is trivial, then $E_{\mathfrak{p}}(k_{\mathfrak{p}})$ contains $E_{\mathfrak{p}}[\ell^n]$ and contains a point Q such that $\ell^n Q = P_{\mathfrak{p}}$. In particular, for each $m \leq n$ there are at least ℓ^{2m} values $R \in E_{\mathfrak{p}}(k_{\mathfrak{p}})$ such that $\ell^m R = P_{\mathfrak{p}}$ for some $m \geq 1$. Thus, there are at least

$$\ell^{2n} + \ell^{2(n-1)} + \cdots + \ell^2 = \frac{\ell^{2(n+1)} - 1}{\ell^2 - 1} - 1$$

points $R \in E_{\mathfrak{p}}(k_{\mathfrak{p}})$ such that $\ell^m R = P_{\mathfrak{p}}$ for some $m \geq 1$. On the other hand, the density of this set of primes (removing the finitely many bad primes) is at least $[F : k]^{-1}$ by the Chebotarev density theorem. Since $\mathrm{Gal}(F/k)$ embeds in $E[\ell^n] \rtimes \mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z})$, we have

$$[F : k] \leq \#\big(E[\ell^n] \rtimes \mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z})\big) \leq \ell^{2n} \times \ell^{3n-2}(\ell^2 - 1) \leq \ell^{5n}.$$

Thus, since $M \geq \ell^{2n-2}$, our chosen set of primes has density at least

$$M^{-(2/5)(1-1/n)} \geq M^{-(2/5)(1-2\log\ell/\log M)}.$$

It is worth noting that, since an analogue of the Chebotarev density theorem is true for function fields in positive characteristic [Kumar Murty and Scherk 1994], the same is true of Theorem 4.

Acknowledgements

I would like to thank David McKinnon and Xander Faber for useful conversations during the writing up of these results, and Joseph Silverman and Soroosh Yazdani for helpful comments on an earlier draft.

References

[Banaszak et al. 2005] G. Banaszak, W. Gajda, and P. Krasoń, "Detecting linear dependence by reduction maps", *J. Number Theory* **115**:2 (2005), 322–342. MR 2006g:11113 Zbl 1089.11030

[Bandini et al. 2009] A. Bandini, I. Longhi, and S. Vigni, "Torsion points on elliptic curves over function fields and a theorem of Igusa", *Expo. Math.* **27**:3 (2009), 175–209. MR 2010j:11087 Zbl 05572047

[Baragar and McKinnon 2010] A. Baragar and D. McKinnon, "$K3$ surfaces, rational curves, and rational points", *J. Number Theory* **130**:7 (2010), 1470–1479. MR 2645231 Zbl 1195.14029

[Bašmakov 1972] M. I. Bašmakov, "Cohomology of Abelian varieties over a number field", *Uspehi Mat. Nauk* **27**:6(168) (1972), 25–66. In Russian. MR 53 #2961 Zbl 0256.14016

[Coates 1970] J. Coates, "An application of the division theory of elliptic functions to diophantine approximation", *Invent. Math.* **11** (1970), 167–182. MR 44 #3963 Zbl 0216.04403

[Cox and Parry 1984] D. A. Cox and W. R. Parry, "Representations associated with elliptic surfaces", *Pacific J. Math.* **114**:2 (1984), 309–323. MR 86a:14042 Zbl 0582.14013

[Faber et al. 2009] X. Faber, B. Hutz, P. Ingram, R. Jones, M. Manes, T. J. Tucker, and M. E. Zieve, "Uniform bounds on pre-images under quadratic dynamical systems", *Math. Res. Lett.* **16**:1 (2009), 87–101. MR 2009m:11095 Zbl 05548667

[Gupta and Ramsay 1997] R. Gupta and K. Ramsay, "Indivisible points on families of elliptic curves", *J. Number Theory* **63**:2 (1997), 357–372. MR 98c:11050 Zbl 0885.11038

[Hindry and Silverman 1988] M. Hindry and J. H. Silverman, "The canonical height and integral points on elliptic curves", *Invent. Math.* **93**:2 (1988), 419–450. MR 89k:11044 Zbl 0657.14018

[Igusa 1959] J.-i. Igusa, "Fibre systems of Jacobian varieties, III: Fibre systems of elliptic curves", *Amer. J. Math.* **81** (1959), 453–476. MR 21 #3422 Zbl 0115.38904

[Kumar Murty and Scherk 1994] V. Kumar Murty and J. Scherk, "Effective versions of the Chebotarev density theorem for function fields", *C. R. Acad. Sci. Paris Sér. I Math.* **319**:6 (1994), 523–528. MR 95j:11104 Zbl 0822.11077

[Lang 2002] S. Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics **211**, Springer, New York, 2002. MR 2003e:00003 Zbl 0984.00001

[Merel 1996] L. Merel, "Bornes pour la torsion des courbes elliptiques sur les corps de nombres", *Invent. Math.* **124**:1-3 (1996), 437–449. MR 96i:11057 Zbl 0936.11037

[Ribet 1979] K. A. Ribet, "Kummer theory on extensions of abelian varieties by tori", *Duke Math. J.* **46**:4 (1979), 745–761. MR 81g:14019 Zbl 0428.14018

[Roquette 1970] P. Roquette, *Analytic theory of elliptic functions over local fields*, Hamburger Mathematische Einzelschriften (N.F.) **1**, Vandenhoeck & Ruprecht, Göttingen, 1970. MR 41 #5376 Zbl 0194.52002

[Salgado 2009] C. Salgado, "Rank of elliptic surfaces and base change", *C. R. Math. Acad. Sci. Paris* **347**:3-4 (2009), 129–132. MR 2010g:11098 Zbl 05527932

[Serre 1972] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR 52 #8126 Zbl 0235.14012

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York, 1979. MR 82e:12016 Zbl 0423.12016

[Silverman 1985] J. H. Silverman, "Divisibility of the specialization map for families of elliptic curves", *Amer. J. Math.* **107**:3 (1985), 555–565. MR 87f:14027 Zbl 0613.14029

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 96b:11074 Zbl 0911.14015

[Silverman 2007] J. H. Silverman, *The arithmetic of dynamical systems*, Graduate Texts in Mathematics **241**, Springer, New York, 2007. MR 2008c:11002 Zbl 1130.37001

[Song and Tucker 2001] X. Song and T. J. Tucker, "Arithmetic discriminants and morphisms of curves", *Trans. Amer. Math. Soc.* **353**:5 (2001), 1921–1936. MR 2002a:11063 Zbl 0982.11035

[Vojta 1992] P. Vojta, "A generalization of theorems of Faltings and Thue–Siegel–Roth–Wirsing", *J. Amer. Math. Soc.* **5**:4 (1992), 763–804. MR 94a:11093 Zbl 0778.11037

Communicated by Barry Mazur
Received 2009-10-01 Revised 2010-03-10 Accepted 2010-08-21

pingram@math.uwaterloo.ca *Department of Pure Mathematics, University of Waterloo,*
Waterloo, ON N2L 3G1, Canada

Current address: *Department of Mathematics, Colorado State University,*
Fort Collins, CO 80523-1874, United States

pingram@math.colostate.edu

ALGEBRA AND NUMBER THEORY 5:4(2011)

Explicit CM theory for level 2-structures on abelian surfaces

Reinier Bröker, David Gruenewald and Kristin Lauter

For a complex abelian surface A with endomorphism ring isomorphic to the maximal order in a quartic CM field K, the Igusa invariants $j_1(A)$, $j_2(A)$, $j_3(A)$ generate an unramified abelian extension of the reflex field of K. In this paper we give an explicit geometric description of the Galois action of the class group of this reflex field on $j_1(A)$, $j_2(A)$, $j_3(A)$. Our description can be expressed by maps between various Siegel modular varieties, and we can explicitly compute the action for ideals of small norm. We use the Galois action to modify the CRT method for computing Igusa class polynomials, and our run time analysis shows that this yields a significant improvement. Furthermore, we find cycles in isogeny graphs for abelian surfaces, thereby implying that the 'isogeny volcano' algorithm to compute endomorphism rings of ordinary elliptic curves over finite fields does not have a straightforward generalization to computing endomorphism rings of abelian surfaces over finite fields.

1. Introduction

Class field theory describes the abelian extensions of a given number field K. For $K = \mathbb{Q}$, the Kronecker–Weber theorem tells us that every abelian extension of K is contained in a *cyclotomic extension*. In 1900, Hilbert asked for a similar 'explicit description' for higher degree number fields. This is known as Hilbert's twelvth problem, and it is still largely unsolved.

Besides $K = \mathbb{Q}$, the answer is only completely known for imaginary quadratic fields. In this case, the solution is provided by *complex multiplication* theory; see for example [Silverman 1994, Chapter 2]. The techniques used can be generalized to *CM fields*, that is, imaginary quadratic extensions of totally real fields. However, for general CM fields we do not always get an explicit description of the *maximal* abelian extension. From a computational perspective, the case of general CM fields is far less developed than the imaginary quadratic case.

Gruenewald thanks Microsoft Research, where this research was undertaken, for its hospitality.
MSC2000: 11G15.
Keywords: abelian surface, isogeny, level structure.

In this article, we solely focus on degree 4 primitive CM fields K. For such fields, invariants of principally polarized abelian surfaces (p.p.a.s.) with endomorphism ring isomorphic to the maximal order $\mathbb{O}_K$ of K generate a subfield of the Hilbert class field of the *reflex field* of K (a degree 4 subfield of the normal closure of K). To explicitly compute the resulting extension, we compute an *Igusa class polynomial*

$$P_K = \prod_{\{A \text{ p.p.a.s} \mid \mathrm{End}(A)=\mathbb{O}_K\}/\cong} (X - j_1(A)) \in \mathbb{Q}[X].$$

Here, j_1 is one of the *three* Igusa invariants of A. A contrast with the case of imaginary quadratic fields — where we compute the *Hilbert class polynomial* — is that the polynomial P_K has rational coefficients that are not integers in general, and it need not be irreducible over $\mathbb{Q}$.

There are three methods to explicitly compute the polynomial P_K: complex analytic evaluation of the invariants [Spallek 1994; van Wamelen 1999; Weng 2003], the CRT method using finite field arithmetic [Eisenträger and Lauter 2009] and the computation of a canonical lift [Gaudry et al. 2006; Carls et al. 2008] using p-adic arithmetic for $p = 2, 3$. However, none of these three approaches exploit the Galois action of the maximal abelian extension of the reflex field on the set of principally polarized abelian surfaces with endomorphism ring $\mathbb{O}_K$. The goal of this article is to make this Galois action explicit and give a method to compute it.

Our algorithm to compute the Galois action significantly speeds up the CRT-approach described in [Eisenträger and Lauter 2009] to compute Igusa class polynomials and it can be used to improve the 3-adic approach [Carls et al. 2008] as well. The improvement in computing Igusa class polynomials parallels the improvements given in [Belding et al. 2008] for computing Hilbert class polynomials. Our run time analysis is similar to the analysis in [Belding et al. 2008]. Contrary to the genus 1 algorithm however, the genus 2 algorithm is not quasilinear in the size of the output. We suggest further refinements that might yield a quasilinear algorithm as area of further study in Section 6.

Besides speeding up the computation of Igusa class polynomials, our algorithm gives a method of computing *isogenous* abelian surfaces over finite fields. Computing an isogeny is a basic computational problem in arithmetic geometry, and we expect that our algorithm can be used in a variety of contexts, ranging from point counting on Jacobians of curves to cryptographic protocols.

Our computations naturally lead us to study the (l, l)-isogeny graph of abelian surfaces over finite fields. For ordinary elliptic curves, the l-isogeny graph looks like a volcano and this observation forms the heart of the algorithm [Kohel 1996] to compute the endomorphism ring of an ordinary elliptic curve over a finite field. We show that for abelian surfaces, the (l, l)-isogeny graph does *not* have a volcano shape. This shows that a straightforward generalization of the elliptic curve algorithm to abelian surfaces does not work.

The structure of this paper is as follows. In Section 2 we recall the basic facts of complex multiplication theory and background on CM abelian surfaces and their invariants. In Section 3 we describe the Galois action on the set of isomorphism classes of abelian surfaces with CM by $\mathbb{O}_K$ in a geometric way. Our algorithm to compute this action is intrinsically linked to Siegel modular functions of higher level. Section 4 gives the definitions and properties of the four Siegel modular functions that we use. The algorithm to compute the Galois action is detailed in Section 5 and we apply it in Section 6 to improving the method to compute an Igusa class polynomial modulo a prime p. We give a detailed run time analysis of our algorithm in Section 6 as well. We illustrate our approach with various detailed examples in Section 7. A final Section 8 contains the obstruction to the volcano picture for abelian surfaces.

2. CM abelian surfaces

2.1. ***CM theory.*** In this section we recall the basic facts of CM theory for higher dimensional abelian varieties. Most of the material presented in this section is an adaptation to our needs of the definitions and proofs of Shimura's [1998] and Lang's [1983] textbooks.

We fix an embedding of $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$. By a real number field, we mean a field that is fixed by complex conjugation. With this convention, a *CM field* K is a totally imaginary quadratic extension of a totally real number field. Let K^+ denote the real quadratic subfield of K, and let n be the degree of K^+ over $\mathbb{Q}$. The $2n$ embeddings $K \hookrightarrow \overline{\mathbb{Q}}$ naturally come in pairs. Indeed, we can choose n embeddings $\Phi = \{\varphi_1, \dots, \varphi_n\}$ such that we have $\mathrm{Hom}(K, \overline{\mathbb{Q}}) = \Phi \cup \overline{\Phi}$. We call such a set Φ a *CM type* for K, and we interpret a CM type in the natural way as a map $K \hookrightarrow \mathbb{C}^n$.

If Φ *cannot* be obtained as a lift of a CM type of a CM subfield of K, then we call Φ *primitive*. For instance, in the simplest case $K^+ = \mathbb{Q}$, CM fields K are imaginary quadratic and every choice for $K \hookrightarrow \overline{\mathbb{Q}}$ determines a primitive CM type. If K has degree, four then every choice of a CM type is primitive when K does not contain an imaginary quadratic field. It is not hard to show [Shimura 1998, Section 8.4] that this occurs exactly for $\mathrm{Gal}(L/\mathbb{Q}) = D_4, C_4$, where L denotes the normal closure of K. We say that the field K is primitive in this case.

In this article, we will only consider primitive quartic CM fields K. For the remainder of this section, we fix such a field K. We say that a principally polarized abelian surface $A/\mathbb{C}$ has *CM* by the maximal order $\mathbb{O}_K$ if there exists an isomorphism $\mathbb{O}_K \xrightarrow{\sim} \mathrm{End}(A)$. The CM type distinguishes these surfaces. More precisely, a surface A that has CM by $\mathbb{O}_K$ has *type* $\Phi = \{\varphi_1, \varphi_2\}$ if the complex representation $R_{\mathbb{C}}$ of the endomorphism algebra $\mathrm{End}(A) \otimes_{\mathbb{Z}} \mathbb{Q}$ satisfies

$$R_{\mathbb{C}} \cong \varphi_1 \oplus \varphi_2.$$

One shows [Lang 1983, Theorem 1.3.6] that a principally polarized abelian surface that has CM by $\mathbb{O}_K$ of type Φ is *simple*, that is, is not isogenous to the product of elliptic curves.

Let Φ be a CM type for K. For an $\mathbb{O}_K$-ideal I, the quotient $A_I = \mathbb{C}^2/\Phi(I)$ is an abelian surface of type Φ by [Lang 1983, Theorem 4.1]. This surface need not admit a principal polarization. The dual variety of A_I is given by $\hat{A}_I = \mathbb{C}^2/\Phi(\bar{I}^{-1}\mathfrak{D}_K^{-1})$, where

$$\mathfrak{D}_K^{-1} = \{x \in K \mid \mathrm{Tr}_{K/\mathbb{Q}}(x\mathbb{O}_K) \subseteq \mathbb{Z}\}$$

is the inverse different and $\bar{I}$ denotes the complex conjugate of I. If $\pi \in K$ satisfies $\Phi(\pi) \in (i\mathbb{R}_{>0})^2$ and $\pi\mathfrak{D}_K = (I\bar{I})^{-1}$, then the map $A_I \to \hat{A}_I$ given by

$$(z_1, z_2) \mapsto (\varphi_1(\pi)z_1, \varphi_2(\pi)z_2)$$

is an isomorphism [Shimura 1998, pages 102–104] and A_I is principally polarizable. All principally polarized abelian surfaces with CM by $\mathbb{O}_K$ of type Φ arise via this construction.

Let L be the normal closure of K. We extend Φ to a CM type Φ' of L, and we define the *reflex field*

$$K_\Phi = \mathbb{Q}\big(\big\{\textstyle\sum_{\phi \in \Phi'} \phi(x) \mid x \in K\big\}\big).$$

The CM type on K induces a CM type $f_\Phi = \{\sigma^{-1}|_{K_\Phi} : \sigma \in \Phi'\}$ of the reflex field K_Φ. The field K_Φ is a subfield of L of degree 4. In particular, it equals K in the case K is Galois. If $L/\mathbb{Q}$ is dihedral, then K_Φ and K are not isomorphic. However, the two different CM types yield isomorphic reflex fields in this case. Furthermore, we have

$$(K_\Phi)_{f_\Phi} = K$$

and the induced CM type on $(K_\Phi)_{f_\Phi}$ equals Φ.

An automorphism σ of K induces an isomorphism $(A, \Phi) \xrightarrow{\sim} (A^\sigma, \Phi^\sigma)$ of CM abelian surfaces, where $\Phi^\sigma = \{\varphi_1\sigma, \varphi_2\sigma\}$. Thus two CM types that are complex conjugates of each other produce the same sets of isomorphism classes of abelian surfaces. In the Galois case there is only one CM type up to isomorphism and in the dihedral case there are two distinct CM types.

2.2. ***Igusa invariants.*** Any principally polarized abelian surface over $\mathbb{C}$ is of the form $A_\tau = \mathbb{C}^2/(\mathbb{Z}^2 + \mathbb{Z}^2\tau)$, where τ is an element of the *Siegel upper half plane*

$$\mathbb{H}_2 = \{\tau \in \mathrm{Mat}_2(\mathbb{C}) \mid \tau \text{ symmetric}, \Im(\tau) \text{ positive definite}\}.$$

The moduli space $\mathcal{A}_2$ of principally polarized abelian surfaces is 3-dimensional. We are mostly interested in the subspace $\mathcal{M}_2 \subset \mathcal{A}_2$ of Jacobians of curves. The structure of $\mathcal{M}_2$ is well known; we recall it for convenience. Let

$$Y^2 = a_6X^6 + \cdots + a_0 = f(X)$$

be a genus 2 curve and write $\alpha_1, \ldots, \alpha_6$ for the roots of f. For simplifying notation, let (ij) denote the quantity $(\alpha_{k_i} - \alpha_{k_j})$ for a given ordering of the roots. The *Igusa–Clebsch invariants* I_2, I_4, I_6, I_{10} (denoted by A, B, C, D in [Igusa 1960, Section 3]) are defined by

$$\begin{aligned}
I_2 &= a_6^2 \sum_{15} (12)^2(34)^2(56)^2,\\
I_4 &= a_6^4 \sum_{10} (12)^2(23)^2(31)^2(45)^2(56)^2(64)^2,\\
I_6 &= a_6^6 \sum_{60} (12)^2(23)^2(31)^2(45)^2(56)^2(64)^2(14)^2(25)^2(36)^2,\\
I_{10} &= a_6^{10} \sum_{i<j} (ij) = a_6^{10}\operatorname{disc}(f),
\end{aligned}$$

where we sum over all root orderings $\{\alpha_{k_i}\}$ that give distinct summands; the subscript indicates the number of terms we sum over.

Theorem 2.1. *The moduli space $\mathcal{M}_2$ is isomorphic to*

$$\{[I_2 : I_4 : I_6 : I_{10}] \in \mathbb{P}^3_w(\mathbb{C}) \mid I_{10} \neq 0\},$$

where $\mathbb{P}^3_w$ denotes weighted projective space with weights 2, 4, 6 *and* 10.

Proof. See [Igusa 1960]. □

We note that the condition $I_{10} \neq 0$ ensures that the polynomial f defining the genus 2 curve is separable.

Instead of working with a subset of weighted projective space, many people work with an affine subspace of $\mathcal{M}_2$. This nonweighted subspace is given by

$$(j_1, j_2, j_3) = \left(\frac{I_2^5}{I_{10}}, \frac{I_4I_2^3}{I_{10}}, \frac{I_6I_2^2}{I_{10}}\right).$$

The functions j_i are commonly called the *Igusa functions*. We remark that there are various definitions of these functions and there are different opinions for which choice is the best. Our functions are the same as those in [van Wamelen 1999], for example. They have the property that for τ, τ' corresponding to Jacobians of curves, the equality $j_i(\tau) = j_i(\tau') \neq 0$ for $i = 1, 2, 3$ implies that C and C' are isomorphic. A detailed description on computing $j_i(\tau)$ for a point $\tau \in \mathbb{H}_2$ can be found in [Dupont 2006; Weng 2003].

A weak version of the main theorem of complex multiplication theory is that, for a primitive quartic CM field K, the Igusa invariants of an abelian variety with CM by $\mathcal{O}_K$ generate an unramified abelian extension of a reflex field of K. More precisely, we have the following result.

Theorem 2.2 [Spallek 1994, Theorem 5.8]. *Let (K, Φ) be a primitive quartic CM type. Let I be an $\mathcal{O}_K$-ideal with the property that there exists a principal polarization on $A_I = \mathbb{C}^2/\Phi(I)$. Then the field $K_\Phi(j_1(A_I), j_2(A_I), j_3(A_I))$ is a subfield of the Hilbert class field of K_Φ. The polynomial*

$$P_K = \prod_A (X - j_1(A)),$$

with A ranging over the isomorphism classes of principally polarized abelian surfaces with endomorphism ring $\mathcal{O}_K$, has rational coefficients. The same is true for the polynomials Q_K and R_K for the j_2 and j_3-invariants.

We will see in Corollary 3.3 that, for any primitive CM type Φ, there always exists an $\mathcal{O}_K$-ideal I such that A_I is principally polarizable.

3. CM action

Throughout this section, we let K be a fixed primitive quartic CM field. We also fix a CM type $\Phi : K \hookrightarrow \mathbb{C}^2$ and let $A/\mathbb{C}$ be a principally polarized abelian surface that has complex multiplication by $\mathcal{O}_K$ of CM type Φ.

3.1. ***Galois action of the class group.*** We define a group $\mathfrak{C}(K)$ as

$$\{(\mathfrak{a}, \alpha) \mid \mathfrak{a} \text{ a fractional } \mathcal{O}_K\text{-ideal with } \mathfrak{a}\bar{\mathfrak{a}} = (\alpha) \text{ and } \alpha \in K^+ \text{ totally positive}\}/\sim,$$

where two pairs $(\mathfrak{a}, \alpha)$ and $(\mathfrak{b}, \beta)$ are equivalent if and only if there exists a unit $u \in K^*$ with $\mathfrak{b} = u\mathfrak{a}$ and $\beta = u\bar{u}\alpha$. The multiplication is defined componentwise, and $(\mathcal{O}_K, 1)$ is the neutral element of $\mathfrak{C}(K)$.

The group $\mathfrak{C}(K)$ naturally acts on the finite set $S(K, \Phi)$ of isomorphism classes of principally polarized abelian surfaces that have CM by $\mathcal{O}_K$ of a given type Φ. Indeed, any such surface is given by an ideal I determining the variety and a 'Φ-positive' element $\pi \in K$ giving the principal polarization. We now put

$$(\mathfrak{a}, \alpha) \cdot (I, \pi) = (\mathfrak{a}I, \alpha\pi) \quad \text{for } (\mathfrak{a}, \alpha) \in \mathfrak{C}(K).$$

By [Shimura 1998, Section 14.6], the action of $\mathfrak{C}(K)$ on $S(K, \Phi)$ is transitive and free. In particular, we have $|\mathfrak{C}(K)| = |S(K, \Phi)|$.

The structure of the group $\mathfrak{C}(K)$ is best described by the following theorem. Denote by $\mathrm{Cl}^+(\mathcal{O}_{K^+})$ the narrow class group of $\mathcal{O}_{K^+}$ and write $(\mathcal{O}_{K^+}^*)^+$ for the group of totally positive units of $\mathcal{O}_{K^+}$.

Theorem 3.1. *Let K be a primitive quartic CM field. Then the sequence*

$$1 \to (\mathbb{O}^*_{K^+})^+/N_{K/K^+}(\mathbb{O}^*_K) \xrightarrow{u \mapsto (\mathbb{O}_K, u)} \mathfrak{C}(K) \xrightarrow{(\mathfrak{a},\alpha)\mapsto \mathfrak{a}} \mathrm{Cl}(\mathbb{O}_K) \xrightarrow{N_{K/K^+}} \mathrm{Cl}^+(\mathbb{O}_{K^+}) \to 1$$

is exact.

Proof. The exactness at $(\mathbb{O}^*_{K^+})^+/N_{K/K^+}(\mathbb{O}^*_K)$ is [Shimura 1998, Section 14.5]. It remains to show that the sequence is exact at $\mathrm{Cl}^+(\mathbb{O}_{K^+})$. To prove this, we first prove[1] that there is a *finite* prime that is ramified in K/K^+.

Suppose that K/K^+ is unramified at all finite primes. By genus theory, we then have $K = K^+(\sqrt{n})$ with $n \in \mathbb{Z}$. However, K then has $\mathbb{Q}(\sqrt{n})$ as quadratic subfield and K is a biquadratic field. This contradicts our assumption that K is primitive.

Because there is a finite prime of K^+ that ramifies in K, the extensions K/K^+ and $H^+(K^+)/K^+$ are linearly disjoint. Here, H^+ denotes the narrow Hilbert class field. By Galois theory, we then have

$$\mathrm{Gal}(H(K)/K) \twoheadrightarrow \mathrm{Gal}(KH^+(K^+)/K) \xrightarrow{\sim} \mathrm{Gal}(H^+(K^+)/K^+). \qquad \square$$

Remark 3.2. The surjectivity of the last arrow was also proved in [Kohel 2008, Lemma 2.1] under the assumption that there exists a finite prime that ramifies in K/K^+. Our proof shows in fact that such a prime always exists.

Corollary 3.3. *Let K be a primitive quartic CM field. The set $S(K)$ of isomorphism classes of principally polarized abelian surfaces with CM by $\mathbb{O}_K$ has cardinality*

$$|S(K)| = \begin{cases} |\mathfrak{C}(K)| & \textit{if } \mathrm{Gal}(K/\mathbb{Q}) \cong C_4, \\ 2|\mathfrak{C}(K)| & \textit{if } \mathrm{Gal}(K/\mathbb{Q}) \cong D_4. \end{cases}$$

Proof. By Theorem 3.1, the cardinality $|S(K, \Phi)| = |\mathfrak{C}(K)|$ is independent of the choice of a CM type Φ. If we let n denote the number of CM types up to conjugacy, then the theorem follows immediately from the equality

$$|S(K)| = n|S(K, \Phi)|. \qquad \square$$

The Galois group $\mathrm{Gal}(K_\Phi(j_1(A))/K_\Phi)$ acts in the following way on the set $S(K, \Phi)$. With f_Φ the CM type on K_Φ induced by Φ, we define $N_\Phi : K_\Phi \to K$ by

$$N_\Phi(x) = \prod_{\varphi \in f_\Phi} \varphi(x).$$

For an $\mathbb{O}_{K_\Phi}$-ideal I, the $\mathbb{O}_K$-ideal $N_\Phi(I)$ is called the *typenorm* of I. We get a natural map $m : \mathrm{Cl}(\mathbb{O}_{K_\Phi}) \to \mathfrak{C}(K)$ defined by

$$m(\mathfrak{p}) = (N_\Phi(\mathfrak{p}), N_{K_\Phi/\mathbb{Q}}(\mathfrak{p})) \quad \text{for degree 1 prime representatives } \mathfrak{p}.$$

[1] We thank Everett Howe for suggesting this argument.

The Galois group of $K_\Phi(j_1(A))/K_\Phi$ is a quotient of $\mathrm{Gal}(H(K_\Phi)/K_\Phi) \cong \mathrm{Cl}(\mathbb{O}_{K_\Phi})$, and by [Shimura 1998, Section 15.2], the induced map

$$m : \mathrm{Gal}(K_\Phi(j_1(A))/K_\Phi) \to \mathfrak{C}(K)$$

is *injective*. This describes the Galois action. Indeed, the group $\mathfrak{C}(K)$ acts on the set of all principally polarized abelian surfaces that have CM by $\mathbb{O}_K$, and m maps the Galois group injectively into $\mathfrak{C}(K)$. In Example 7.2 we will see that the natural map $\mathrm{Cl}(\mathbb{O}_{K_\Phi}) \to \mathfrak{C}(K)$ need not be injective.

The typenorm can be defined in a slightly different way as well. If $K/\mathbb{Q}$ is Galois with $\mathrm{Gal}(K/\mathbb{Q}) = \langle\sigma\rangle$, then for $\Phi = \{1, \sigma\}$ we have $N_\Phi(\mathfrak{p}) = \mathfrak{p}^{1+\sigma^3}$. If K is not Galois, then we have $N_\Phi(\mathfrak{p}) = N_{L/K}(\mathfrak{p}\mathbb{O}_L)$. We will use this description both for actual computations and in the proof of Lemma 6.5.

In the remainder of Section 3.1, we provide the theoretical framework that will allow us to explicitly compute the CM action. Let I be a $\mathbb{O}_{K_\Phi}$-ideal of norm l. We assume for simplicity that l is prime. We have $m(I) = (N_\Phi(I), l) = (J, l) \in \mathfrak{C}(K)$, where J is an $\mathbb{O}_K$-ideal of norm l^2.

Lemma 3.4. *Let I be an $\mathbb{O}_{K_\Phi}$-ideal of prime norm l with typenorm $N_\Phi(I) = J \subset \mathbb{O}_K$. Then J divides $(l) \subset \mathbb{O}_K$.*

Proof. This follows from the relation $N_\Phi(I)\overline{N_\Phi(I)} = N_{K_\Phi/\mathbb{Q}}(I) = l$. □

For an $\mathbb{O}_K$-ideal M, we define the M-torsion of the abelian surface A by

$$A[M] = \{P \in A(\mathbb{C}) \mid \text{for all } \alpha \in M : \alpha(P) = 0\}.$$

We assume here that we have *fixed* an isomorphism $\mathrm{End}(A) \xrightarrow{\sim} \mathbb{O}_K$, meaning that M is an $\mathrm{End}(A)$-ideal as well. If M is generated by an integer n, then $A[M]$ equals the n-torsion $A[n]$.

Lemma 3.4 implies that $A[J]$ is a 2-dimensional subspace of the l-torsion $A[l]$ of A. The polarization of A induces a symplectic form on $A[l]$, and $A[l]$ is a *symplectic* vector space of dimension 4 over the finite field $\mathbb{F}_l$. By CM theory we know that the quotient $A/A[J]$ is again a *principally polarized* abelian surface. By [Mumford 1970, Section 23], this implies that $A[J]$ is an *isotropic* 2-dimensional subspace of $A[l]$, that is, the symplectic form vanishes on $A[J]$. We recall that an isogeny $A \to B$ between principally polarized abelian surfaces whose kernel is a 2-dimensional isotropic subspace of $A[l]$ is called an *(l, l)-isogeny*, and $A \to A/A[J]$ is an example of an (l, l)-isogeny.

The moduli space of all pairs (A, G), with A a principally polarized abelian surface over $\mathbb{C}$ and G a 2-dimensional isotropic subspace of $A[l]$, can be described by an ideal $V(l) \subset \mathbb{Q}[X_1, Y_1, Z_1, X_2, Y_2, Z_2]$. More precisely, the variety corresponding to $V(l)$ equals the Siegel modular variety $Y_0^{(2)}(l)$ studied, for example, in [Bröker and Lauter 2009]. As a complex Riemann surface, we have

$$Y_0^{(2)}(l) = \Gamma_0^{(2)}(l) \setminus \mathbb{H}_2,$$

with

$$\Gamma_0^{(2)}(l) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \in \mathrm{Sp}_4(\mathbb{Z}) \mid c \equiv 0_2 \bmod l \right\}.$$

If we specialize $V(l)$ at a point $(X_1, Y_1, Z_1) = (j_1(A), j_2(A), j_3(A))$, the resulting ideal $V'(l)$ is 0-dimensional. The corresponding variety is a union of points corresponding to the '(l, l)-isogenous abelian surfaces'. Since there are $[\mathrm{Sp}_4(\mathbb{Z}) : \Gamma_0^{(2)}(l)] = (l^4 - 1)/(l - 1)$ isotropic subspaces of dimension 2 in $A[l]$ by [Bröker and Lauter 2009, Lemma 6.1], there are exactly $(l^4 - 1)/(l - 1)$ solutions to the system of equations given by V'. By construction, the triple

$$(j_1(A/J), j_2(A/J), j_3(A/J))$$

is one of the solutions. There are $l^3 + l^2 + l$ other solutions, and we will see in Section 6 that for CM computations it is relatively easy to determine which of the solutions come from the typenorm of an $\mathcal{O}_{K_\Phi}$-ideal.

Unfortunately, the ideal $V(l)$ can only be computed for very small l. Indeed, the only case that has been done is $l = 2$ [Dupont 2006, Section 10.4.2] and it takes roughly 50 megabytes to store the 3 generators of V. By [Bröker and Lauter 2009], knowing the ideal $V(l)$ for some prime l implies that we have an equation for the *Humbert surface* of discriminant l^2. Since computing Humbert surfaces is a traditionally hard problem, we do not expect that much progress can be made in computing $V(l)$ for primes $l > 2$.

3.2. ***Richelot isogeny.*** Although one could use the ideal $V(2)$ from [Dupont 2006, Section 10.4.2] to compute $(2, 2)$-isogenies, there is a more efficient way. This alternative, known as the *Richelot isogeny*, is classical and we recall it here for convenience. Let K be a field of characteristic different from 2, and let C/K be a nonsingular genus 2 curve. We can choose an equation $Y^2 = f(X)$ for C, with $f \in K[X]$ a monic polynomial of degree 6. Any factorization $f = ABC$ into three monic degree 2 polynomials defines a genus 2 curve C' given by

$$\Delta Y^2 = [A, B][A, C][B, C],$$

where Δ is the determinant of A, B, C with respect to the basis $1, X, X^2$, and $[A, B] = A'B - AB'$ with A' the derivative of A. This new curve is nonsingular precisely when Δ is nonzero.

One proves [Bost and Mestre 1988] that C and C' are $(2, 2)$-isogenous. It is not hard to see that there are exactly $15 = (2^4 - 1)/(2 - 1)$ different curves C' that can be obtained this way. It follows that this construction gives all $(2, 2)$-isogenous Jacobians $\mathrm{Jac}(C')$.

4. Smaller functions

The Igusa functions introduced in Section 2 are 'too large' to be practical in our computation of the CM action: currently we cannot compute an ideal describing the variety $Y_0^2(l)$ for primes $l > 2$. In this section we introduce smaller functions $f_1, \dots, f_4$ that are more convenient from a computational perspective. For $N > 1$, we define the *congruence subgroup* of level N as the kernel of the reduction map $\mathrm{Sp}_4(\mathbb{Z}) \to \mathrm{Sp}_4(\mathbb{Z}/N\mathbb{Z})$, denoted by $\Gamma(N)$.

For $x, y \in \{0, 1\}^2$, define the functions $\theta_{x,y} : \mathbb{H}_2 \to \mathbb{C}$ by

$$\theta_{x,y}(\tau) = \sum_{n\in\mathbb{Z}^2} \exp \pi i \big((n + \tfrac{1}{2}x)^T \tau (n + \tfrac{1}{2}x) + (n + \tfrac{1}{2}x)^T y\big). \tag{4-1}$$

The functions $\theta_{x,y}$ are known as the *theta constants* and arise naturally from the construction of theta functions [Igusa 1964]. The equality $\theta_{x,y}(\tau) = (-1)^{x^T y}\theta_{x,y}(\tau)$ shows that only 10 of the 16 theta constants are nonzero.

The fourth powers of the functions $\theta_{x,y}$ are Siegel modular forms of weight 2 for the congruence subgroup $\Gamma(2) \subset \mathrm{Sp}_4(\mathbb{Z})$. The Satake compactification $X(2)$ of the quotient $\Gamma(2) \setminus \mathbb{H}_2$ has a natural structure of a projective variety, and the fourth powers $\theta_{x,y}^4$ define an embedding of $X(2)$ into projective space.

Theorem 4.1 [van der Geer 1982, Theorem 5.2]. *Let $M_2(\Gamma(2))$ denote the $\mathbb{C}$-vector space of all Siegel modular forms of weight 2 for the congruence subgroup $\Gamma(2)$. Then the space $M_2(\Gamma(2))$ is 5-dimensional and is spanned by the ten modular forms $\theta_{x,y}^4$. Furthermore, the map $X(2) \to \mathbb{P}^4 \subset \mathbb{P}^9$ defined by the functions $\theta_{x,y}^4$ is an embedding. The image is the quartic threefold in $\mathbb{P}^4$ defined by*

$$u_2^2 - 4u_4 = 0 \quad \textit{with} \;\; u_k = \sum_{x,y} \theta_{x,y}^{4k}.$$

The Igusa functions j_1, j_2, j_3 can be readily expressed in terms of $\theta_{x,y}^4$; see for example [Igusa 1967, page 848]. Thus we have an inclusion

$$\mathbb{C}(j_1, j_2, j_3) \subseteq \mathbb{C}(\theta_{x,y}^4/\theta_{x',y'}^4)$$

where we include *all* quotients of theta fourth powers. The functions $\theta_{x,y}^4/\theta_{x',y'}^4$ are rational Siegel modular *functions* of level 2. Whereas $(j_1(\tau), j_2(\tau), j_3(\tau))$ depends only on the $\mathrm{Sp}_4(\mathbb{Z})$-equivalence class of $\tau \in \mathbb{H}_2$, a value $(\theta_{x,y}^4(\tau)/\theta_{x',y'}^4(\tau))_{x,x',y,y'}$ depends on the $\Gamma(2)$-equivalence class of τ. Since the affine points of $\Gamma(2) \setminus \mathbb{H}_2 \subset X(2)$ correspond to isomorphism classes of pairs $(A, \{P_1, P_2, P_3, P_4\})$ consisting of a principally polarized 2-dimensional abelian variety A together with a basis $\{P_1, P_2, P_3, P_4\}$ of the 2-torsion, the functions $\theta_{x,y}^4/\theta_{x',y'}^4$ not only depend on the abelian variety in question but also on an ordering of its 2-torsion. For every isomorphism class $\mathrm{Sp}_4(\mathbb{Z})\tau$ of abelian varieties, there are $[\mathrm{Sp}_4(\mathbb{Z}) : \Gamma(2)] = 720$

values for the tuple $(\theta^4_{x,y}(\tau)/\theta^4_{x',y'}(\tau))_{x,x',y,y'}$. The functions $\theta^4_{x,y}/\theta^4_{x',y'}$ are smaller than the Igusa functions in the sense that their Fourier coefficients are smaller. A natural idea is to get even smaller functions by considering the quotients $\theta_{x,y}/\theta_{x',y'}$ themselves instead of their fourth powers.

We define the four functions $f_1, f_2, f_3, f_4 : \mathbb{H}_2 \to \mathbb{C}$ by

$$f_1 = \theta_{(0,0),(0,0)} \quad f_2 = \theta_{(0,0),(1,1)} \quad f_3 = \theta_{(0,0),(1,0)} \quad f_4 = \theta_{(0,0),(0,1)},$$

with $\theta_{(x,y),(x',y')} = \theta_{x,y}/\theta_{x',y'}$. We stress that the particular choice of the 'theta constants' is rather arbitrary; our only requirement is that we define 4 different functions. The three quotients f_1/f_4, f_2/f_4, f_3/f_4 are rational Siegel modular functions.

Theorem 4.2. *If $\tau, \tau' \in \mathbb{H}_2$ satisfy $(f_1(\tau), \dots, f_4(\tau)) = (f_1(\tau'), \dots, f_4(\tau'))$, then we have $(j_1(\tau), j_2(\tau), j_3(\tau)) = (j_1(\tau'), j_2(\tau'), j_3(\tau'))$. Furthermore, the quotients f_1/f_4, f_2/f_4, f_3/f_4 are invariant under the subgroup $\Gamma(8)$.*

Proof. The vector space $M_2(\Gamma(2))$ is spanned by $\{f_1^4, \dots f_4^4, g^4\}$, where $g = \theta_{(0,1),(0,0)}$. The relation in Theorem 4.1, together with the five linear relations between the $\theta^4_{x,y}$ from Riemann's theta formula [Igusa 1964, page 232], yield that g^4 satisfies a degree 4 polynomial P over $L = \mathbb{C}(f_1, f_2, f_3, f_4)$. The polynomial P factors over L as a product of the 2 irreducible quadratic polynomials

$$P_-, P_+ = T^2 - (f_1^4 - f_2^4 + f_3^4 - f_4^4)T + (f_1^2 f_3^2 \pm f_2^2 f_4^2)^2.$$

By looking at the Fourier expansions of $f_1, \dots, f_4$ and g, we see that g^4 is a root of P_-. Hence, the extension $L(g^4)/L$ is quadratic and generated by a root of P_-.

For each of the 2 choices of a root of P_-, the other 5 fourth powers of theta functions will be uniquely determined. Indeed, the fourth powers are functions on the space $M_2(\Gamma(2))$ and this space is 5-dimensional by Theorem 4.1. This means that we get a priori *two* Igusa triples (j_1, j_2, j_3) for every tuple (f_1, f_2, f_3, f_4). However, a close inspection of the formulas expressing the Igusa functions in terms of theta fourth powers yields that these Igusa triples coincide. Hence, the triple (j_1, j_2, j_3) does not depend on the choice of a root of P_-. This proves the first statement in the theorem.

The second statement follows immediately from a result of Igusa, who proves in [Igusa 1964, page 242] that the field M generated by *all* theta quotients is invariant under a group that contains $\Gamma(8)$. Since the field $\mathbb{C}(f_1/f_4, f_2/f_4, f_3/f_4)$ is a subfield of M, Theorem 4.2 follows. □

Since the functions f_1/f_4, f_2/f_4, f_3/f_4 are invariant under $\Gamma(8)$, the moduli interpretation is that they depend on an abelian variety together with a level 8-structure. Let $\mathrm{Stab}(f)$ be the stabilizer of f_1/f_4, f_2/f_4, f_3/f_4 inside the symplectic group $\mathrm{Sp}_4(\mathbb{Z})$. We have inclusions

$$\Gamma(8) \subset \mathrm{Stab}(f) \subset \mathrm{Sp}_4(\mathbb{Z})$$

and the quotient $Y(f) = \mathrm{Stab}(f) \setminus \mathbb{H}_2$ has a natural structure of a quasiprojective variety by the Baily–Borel theorem [1966]. However, this variety is not smooth.

We let

$$\mathbb{H}_2^* = \{\tau \in \mathbb{H}_2 \mid \tau \text{ is not } \mathrm{Sp}_4(\mathbb{Z})\text{-equivalent to a diagonal matrix}\}$$

be the subset of $\mathbb{H}_2$ of those τ that do not correspond to a product of elliptic curves with the product polarization. The argument in [Runge 1993, Section 5] shows that $G = \Gamma(8)/\,\mathrm{Stab}(f)$ acts freely on $Y(8)$. By [Mumford 1970, Chapter 2, Section 7], the quotient

$$Y(f)^* = \mathrm{Stab}(f) \setminus \mathbb{H}_2^*$$

is a *smooth* variety.

Lemma 4.3. *The map $Y(f)^* \to Y(1)$ induced by the inclusion* $\mathrm{Stab}(f) \to \mathrm{Sp}_4(\mathbb{Z})$ *has degree* $23040 = 32 \cdot 720$.

Proof. The map factors as $Y(f)^* \to Y(2) \to Y(1)$; thus it suffices to determine the degrees of each part. The degree of the map $Y(f)^* \to Y(2)$ can be seen from the proof of Theorem 4.2: given a projective tuple $(f_4^4, f_2^4, f_3^4, f_4, g^4)$ representing a point Q of $Y(2)$, over a splitting field there are $4^3 = 64$ projective tuples (f_1, f_2, f_3, f_4) and exactly half of these satisfy $P_- = 0$ and hence are valid preimages of Q. Thus $Y(f)^* \to Y(2)$ has degree 32. The degree of $Y(2) \to Y(1)$ equals $[\mathrm{Sp}_4(\mathbb{Z}) : \Gamma(2)] = 720$. This completes the proof. □

From a tuple $(f_1(\tau), \dots, f_4(\tau))$, the proof of Theorem 4.2 shows how to compute an Igusa triple $(j_1(\tau), j_2(\tau), j_3(\tau))$. For convenience, we make this explicit in the next subsection.

4.1. ***Transformation formulas.*** As in the proof of Theorem 4.2, let $g = \theta_{(0,1),(0,0)}$. Now g^4 is a root of the quadratic polynomial P_-. From values (f_1, f_2, f_3, f_4), we can pick any root of P_- as a value for g^4. The functions $\{f_1^4, \dots f_4^4, g^4\}$ form a basis of $M_2(\Gamma(2))$. Define new functions x_i by

$$\begin{aligned}
x_1 &= -f_1^4 + 2f_2^4 - f_3^4 + 2f_4^4 + 3g^4,\\
x_2 &= -f_1^4 + 2f_2^4 - f_3^4 - f_4^4,\\
x_3 &= -f_1^4 - f_2^4 - f_3^4 + 2f_4^4,\\
x_4 &= 2f_1^4 - f_2^4 - f_3^4 - f_4^4,\\
x_5 &= -f_1^4 - f_2^4 + 2f_3^4 - f_4^4,\\
x_6 &= 2f_1^4 - f_2^4 + 2f_3^4 - f_4^4 - 3g^4.
\end{aligned}$$

The x_i are called level 2 *Satake coordinate functions*. In terms of these functions we obtain a model for $X[2]$ embedded in $\mathbb{P}^5$ given by

$$s_1 = 0 \quad \text{and} \quad s_2^2 - 4s_4 = 0,$$

where $s_k = \sum_{i=1}^6 x_i^k$ are the k-th power sums.

The action of $\mathrm{Sp}_4(\mathbb{Z})/\Gamma(2)$ on $x_i(\tau)$ is equivalent to that of $\mathrm{Sym}(\{x_1, \dots, x_6\})$ permuting the coordinates. Thus we can write level 1 modular functions as symmetric functions of the x_i, and the *Igusa–Clebsch invariants* from Section 2.2 are given by

$$\begin{aligned} I_2 &= \frac{5(48s_6 - 3s_2^3 - 8s_3^2)}{3(12s_5 - 5s_2s_3)}, \\ I_4 &= 3^{-1}s_2^2, \\ I_6 &= 3^{-2}(3I_2I_4 - 2s_3), \\ I_{10} &= 2^{-2}3^{-6}5^{-1}(12s_5 - 5s_2s_3), \end{aligned}$$

from which we can compute absolute Igusa invariants (j_1, j_2, j_3).

Conversely, if $(j_i(\tau))$ corresponds to the Jacobian of a curve, then we can compute a value for $(f_1(\tau), \dots, f_4(\tau))$ as follows. First we compute the Igusa–Clebsch invariants, then we apply the transformation

$$\begin{aligned} s_2 &= 3I_4, \\ s_3 &= 3/2(I_2I_4 - 3I_6), \\ s_5 &= 5/12s_2s_3 + 3^5 \cdot 5I_{10}, \\ s_6 &= 27/16I_4^3 + 1/6s_3^2 + 3^6/2^2I_2I_{10}, \end{aligned}$$

after which we can compute the level 2 Satake coordinate functions as the roots $x_1, \dots, x_6$ of the *Satake sextic polynomial*

$$X^6 - \tfrac{1}{2}s_2X^4 - \tfrac{1}{3}s_3X^3 + \tfrac{1}{16}s_2^2X^2 + (\tfrac{1}{6}s_2s_3 - \tfrac{1}{5}s_5)X + (\tfrac{1}{96}s_2^3 + \tfrac{1}{18}s_3^2 - \tfrac{1}{6}s_6)$$

with coefficients in $\mathbb{Q}(s_2, s_3, s_5, s_6)$. One choice for $f_1^4, f_2^4, f_3^4, f_4^4$ is given by

$$\begin{aligned} f_1^4 &= (-x_2 - x_3 - x_5)/3, \\ f_2^4 &= (-x_3 - x_4 - x_5)/3, \\ f_3^4 &= (-x_2 - x_3 - x_4)/3, \\ f_4^4 &= (-x_2 - x_4 - x_5)/3. \end{aligned}$$

Finally, we extract fourth roots to find values for $(f_1(\tau), \dots, f_4(\tau))$ satisfying $P_- = 0$. It is easy to find a solution to $P_- = 0$: if $(f_1, \dots, f_4)$ is not a solution, then $(\sqrt{-1}f_1, \dots, f_4)$ is a solution.

The coefficients of the Satake sextic polynomial are in $\mathbb{Z}[\frac{1}{2}, \frac{1}{3}, I_2, I_4, I_6, I_{10}]$. In particular, this means that our transformation formulas are also valid over finite fields of characteristic greater than 3.

5. The CM action and level structure

We let $\mathrm{Stab}(f)$ be the stabilizer of the three quotients f_1/f_4, f_2/f_4, f_3/f_4 defined in Section 4. By Theorem 4.2, we have $\Gamma(8) \subseteq \mathrm{Stab}(f)$. For a prime $l > 2$, we now define

$$Y(f; l)^* = (\mathrm{Stab}(f) \cap \Gamma_0^{(2)}(l)) \setminus \mathbb{H}_2^*$$

which we view as an equality of Riemann surfaces. By the Baily–Borel theorem, the space $Y(f; l)^*$ has a natural structure of a variety. Since we restricted to $\mathbb{H}_2^*$, the variety is affine. Just like in the case $l = 1$ from Section 4, $Y(f; l)^*$ is smooth.

The moduli interpretation of $Y(f; l)^*$ is the following. Points are isomorphism classes of triples (A, G, L), where A is a principally polarized complex abelian surface, G is a 2-dimensional isotropic subspace of $A[l]$, and L is a level 8-structure. The notion of isomorphism is that (A, G, L) and (A', G', L') are isomorphic if and only if there is an isomorphism $\varphi : A \to A'$ of principally polarized abelian surfaces that satisfies $\varphi(G) = G'$ and $\varphi(L) = L'$.

Lemma 5.1. *The map $Y(f; l)^* \to Y(f)^*$ that is induced by the inclusion map $(\mathrm{Stab}(f) \cap \Gamma_0^{(2)}(l)) \to \mathrm{Stab}(f)$ has degree $(l^4 - 1)/(l - 1)$ for primes $l > 2$.*

Proof. This is clear: the choice of a level 8-structure L is independent of the choice of a subspace of the l-torsion for $l > 2$. □

Besides the map $Y(f; l)^* \to Y(f)^*$ from the lemma, we also have a map $Y(f; l)^* \to Y(f)^*$ given by $(A, G, L) \mapsto (A/G, L')$. Indeed, the isogeny $\varphi : A \to A/G$ induces an isomorphism $A[8] \to (A/G)[8]$ and we have $L' = \varphi(L)$. As was explained in Section 3.2, this map also has degree $(l^4 - 1)/(l - 1)$. Putting all the varieties together, the picture is as follows.

$$\begin{tikzcd}
 & Y(f;l)^* \arrow[dl, "s"'] \arrow[dr, "t"] \arrow[ddl, bend left] \arrow[ddr, bend right] & \\
Y(f)^* \arrow[d, "f"] & & Y(f)^* \arrow[d, "f"'] \\
Y(1) & & Y(1) \arrow[d, dotted] \\
\mathbb{A}^3 \arrow[u] & & \mathbb{A}^3
\end{tikzcd}$$

The map s sends $(A, G, L) \in Y(f; l)^*$ to $(A, L) \in Y(f)^*$ and t is the map induced by the isogeny $A \to A/G$. This diagram allows us to find all the abelian surfaces that are (l, l)-isogenous to a given surface A, where we assume that A is the Jacobian of a genus 2 curve. Indeed, we first map the Igusa invariants $(j_1(A), j_2(A), j_3(A))$ to a point in $Y(1)$, say given by the Igusa–Clebsch invariants. We then *choose* (A, L) on $Y(f)^*$ lying over this point. Although there are 23040 choices for L, it does not matter which one we choose. Above (A, L), there are $(l^4-1)/(l-1)$ points in $Y(f; l)^*$ and via the map $t : Y(f; l)^* \to Y(f)^*$ we map all of these down to $Y(f)^*$. Forgetting the level 8-structure now yields $(l^4-1)/(l-1)$ points in $Y(1)$. If A is simple, that is, not isogenous to a product of elliptic curves, then we can transform these into absolute Igusa invariants.

Assuming we can compute an ideal

$$V(f; l) \subset \mathbb{Q}[W_1, X_1, Y_1, Z_1, W_2, X_2, Y_2, Z_2]$$

defining the quasiprojective variety $Y(f; l)^*$, we derive the following algorithm to compute all (l, l)-isogenous abelian surfaces.

Algorithm 5.2. Input: A Jacobian $A/\mathbb{C}$ of a genus 2 curve given by its Igusa invariants, and the ideal $V(f; l)$ defining $Y(f; l)^*$.

Output: The Igusa invariants of all principally polarized abelian surfaces that are (l, l)-isogenous to A.

(1) Compute Igusa–Clebsch invariants $(I_2, I_4, I_6, I_{10}) \in \mathbb{C}^4$ corresponding to A.

(2) Choose an element $(f_1, f_2, f_3, f_4) \in Y(f)^*$ that maps to (I_2, I_4, I_6, I_{10}) using the method described in Section 4.1.

(3) Specialize the ideal $V(f; l)$ in $(W_1, X_1, Y_1, Z_1) = (f_1, f_2, f_3, f_4)$ and solve the remaining system of equations.

(4) For each solution found in the previous step, compute the corresponding point in $Y(1)$ using the method given in Section 4.1.

5.1. ***Computing $V(f; l)$.*** In this subsection, we use an algorithm of Gruenewald [2008] to compute the ideal $V(f; l)$ needed in Algorithm 5.2. Our approach only terminates in a reasonable amount of time in the simplest case $l = 3$.

The expression for the theta constants in (4-1) can be written in terms of the individual matrix entries, and with some minor modifications we can represent it as a power series with integer coefficients. Writing $\tau = \left(\begin{smallmatrix}\tau_1 & \tau_2\\ \tau_2 & \tau_3\end{smallmatrix}\right) \in \mathbb{H}_2$, we have

$$\begin{aligned}\theta_{(a,b),(c,d)}(\tau) = (-1)^{\frac{ac+bd}{2}} &\sum_{(x_1,x_2)\in\mathbb{Z}^2} (-1)^{x_1c+x_2d} p^{(2x_1+a)^2} q^{(2x_1+a+2x_2+b)^2} r^{(2x_2+b)^2} \\ &\in \mathbb{Z}[[p, q, r]],\end{aligned}$$

where $p = e^{2\pi i(\tau_1-\tau_2)/8}$, $q = e^{2\pi i\tau_2/8}$ and $r = e^{2\pi i(\tau_3-\tau_2)/8}$. We see that it is easy to compute Fourier expansions for the Siegel modular forms f_i.

One of the surfaces (l, l)-isogenous to $\mathbb{C}^2/(\mathbb{Z}^2 + \mathbb{Z}^2 \cdot \tau)$ is $\mathbb{C}^2/(\mathbb{Z}^2 + \mathbb{Z}^2 \cdot l\tau)$, and we want to find a relation between the f_i and the functions $f_i(l\tau)$. The expansion for $f_i(l\tau)$ can be constructed easily from the Fourier expansion of $f_i(\tau)$ by replacing p, q, r with p^l, q^l, r^l.

Starting with $n = 2$, we compute all homogeneous monomials of degree n in $f_i(\tau)$, $f_i(l\tau)$ represented as truncated power series and then use exact linear algebra to find linear dependencies between them. In this manner we obtain a basis for the degree n homogeneous component of the relation ideal. We then check experimentally whether our list of relations generate $V(f; l)$ or not by computing the degree of the projection maps. If one of the projection maps has degree larger than $l^3 + l^2 + l + 1$, then more relations are required, in which case we increment n by 1 and repeat the procedure. We stop once we have found sufficiently many relations to generate $V(f; l)$.

Using this method we computed the ideal $V(f; 3)$. The (3, 3)-isogeny relations in $V(f; 3)$ are given by 85 homogeneous polynomials of degree six. The whole ideal takes 35 kilobytes to store in a text file; see the online supplement. The individual relations are fairly small, having at most 40 terms. Furthermore, the coefficients are 7-smooth and bounded by 200 in absolute value, which makes them amenable for computations.

However, we have not rigorously proven that the ideal $V(f; 3)$ is correct. To do this we would need to show that our 85 polynomials define relations between Siegel modular forms rather than just truncated Fourier expansions. From the work of Poor and Yuen [2000] there is a computable bound for which a truncated Fourier expansion uniquely determines the underlying Siegel modular form. Thus with high enough precision our relations are able to be proven. A Gröbner basis computation in Magma [Bosma et al. 1997] informs us that the projection maps have the expected degree 40; hence we have obtained enough relations. Under the assumption that these relations hold, the ideal $V(f; 3)$ is correct.

Our (3, 3)-isogeny relations hold for all Jacobians of curves. We remark that if we restrict ourselves to CM-abelian surfaces defined over unramified extensions of $\mathbb{Z}_3$, then there are smaller (3, 3)-isogeny relations; see [Carls et al. 2008]. These smaller relations cannot be used however to improve the ‘CRT-algorithm’ as in Section 6.3.

6. The CM action over finite fields

6.1. ***Reduction theory.*** The theory we developed in Sections 3 through 5 uses the *complex analytic* definition of abelian surfaces and the Riemann surfaces $Y_0^{(2)}(l)$

and $Y(f;l)^*$. We now explain why we can use the results in *positive characteristic* as well. Firstly, if we take a prime p that splits completely in K, then by [Goren 1997, Theorems 1 and 2] the reduction modulo p of an abelian surface $A/H(K_\Phi)$ with endomorphism ring $\mathbb{O}_K$ is *ordinary*. The reduced surface again has endomorphism ring $\mathbb{O}_K$.

Furthermore, one can naturally associate an algebraic stack $\mathfrak{A}_{\Gamma_0(p)}$ to $Y_0^{(2)}(l)$ and prove that the structural morphism $\mathfrak{A}_{\Gamma_0(p)} \to \mathrm{Spec}(\mathbb{Z})$ is smooth outside l; see [Chai and Norman 1990, Corollary 6.1.1]. In more down-to-earth computational terminology, this means the moduli interpretation of the ideal $V \subset \mathbb{Q}[X_1, \dots, Z_2]$ remains valid when we reduce the elements of V modulo a prime $p \neq l$.

The reduction of $Y(f;l)^*$ is slightly more complicated. The map $Y(8l) \to Y(f;l)^*$ is finite étale by [Katz and Mazur 1985, Theorem A.7.1.1], where we now view the affine varieties $Y(f;l)^*$ and $Y(8l)$ as schemes. It is well known that $Y(N)$ is smooth over $\mathrm{Spec}(\mathbb{Z}[1/N])$ for $N \geq 3$, so in particular, the scheme $Y(f;l)^*$ is smooth over $\mathrm{Spec}(\mathbb{Z}[1/(2l)])$. Again, this means that the moduli interpretation for the ideal $V(f;l) \subset \mathbb{Q}[W_1, \dots, Z_2]$ remains valid when we reduce the elements of $V(f;l)$ modulo a prime $p \nmid 2l$.

We saw at the end of Section 4 that our transformation formulas are valid modulo p for primes $p > 3$. Putting this all together, we obtain the following result:

Lemma 6.1. *Let l be prime, and let $p \nmid 6l$ be a prime that splits completely in a primitive CM field K. On input of the Igusa invariants of a principally polarized abelian surface $A/\overline{\mathbb{F}}_p$ with* $\mathrm{End}(A) = \mathbb{O}_K$ *and the ideal $V(f;l) \subset \overline{\mathbb{F}}_p[W_1, \dots, Z_2]$, Algorithm 5.2 computes the Igusa invariants of all (l,l)-isogenous abelian surfaces.*

6.2. ***Finding (l,l)-isogenous abelian surfaces.*** Now fix a primitive quartic CM field K, and let $p \nmid 6l$ be a prime that splits completely in the subfield

$$K_\Phi(j_1(A), j_2(A), j_3(A))$$

of the Hilbert class field of K_Φ. By the choice of p, the Igusa invariants of an abelian surface $A/\overline{\mathbb{F}}_p$ with $\mathrm{End}(A) = \mathbb{O}_K$ are defined over the prime field $\mathbb{F}_p$. Moreover, p splits in K_Φ and as it splits in its normal closure L it will split completely in K; hence Lemma 6.1 applies.

Algorithm 5.2 applied to the point $(j_1(A), j_2(A), j_3(A))$ and the ideal $V(f;l)$ yields $(l^4-1)/(l-1)$ triples of Igusa invariants. All these triples are Igusa invariants of principally polarized abelian surfaces with endomorphism *algebra* K; some are defined over the prime field $\mathbb{F}_p$ and some are not. However, since p splits completely in the field of moduli $K_\Phi(j_1(A), j_2(A), j_3(A))$, the Igusa invariants of the surfaces that have endomorphism ring $\mathbb{O}_K$ *are defined over the field* $\mathbb{F}_p$.

Algorithm 6.2. Input: The Igusa invariants of a simple principally polarized abelian surface $A/\mathbb{F}_p$ with $\mathrm{End}(A) = \mathbb{O}_K$, and the ideal $V(f; l) \subset \mathbb{F}_p[W_1, \dots, Z_2]$. Here, l is a prime such that there exists a prime ideal in K_Φ of norm l. Furthermore, we assume that $p \nmid 6l$.

Output: The Igusa invariants of all principally polarized abelian surfaces $A'/\mathbb{F}_p$ with $\mathrm{End}(A') = \mathbb{O}_K$ that are (l, l)-isogenous to A.

(1) Apply Algorithm 5.2 to A and $V(f; l)$. Let S be the set of all Igusa invariants that are defined over $\mathbb{F}_p$.

(2) For each $(j_1(A'), j_2(A'), j_3(A')) \in S$, construct a genus 2 curve C having these invariants using Mestre's algorithm; see [Mestre 1991; Cardona and Quer 2005].

(3) Apply the Freeman–Lauter algorithm [2008] to test whether $\mathrm{Jac}(C)$ has endomorphism ring $\mathbb{O}_K$. Return the Igusa invariants of all the curves that pass this test.

We can predict beforehand *how many* triples will be returned by Algorithm 6.2. We compute the prime factorization

$$(l) = \mathfrak{p}_1^{e_1} \dots \mathfrak{p}_k^{e_k}$$

of (l) in K_Φ. Say that we have $n \le 4$ prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ of norm l in this factorization, disregarding multiplicity. For each of these ideals $\mathfrak{p}_i$ we compute the typenorm map $m(\mathfrak{p}_i) \in \mathfrak{C}(K)$. The size of

$$\{m(\mathfrak{p}_1), \dots, m(\mathfrak{p}_n)\} \subset \mathfrak{C}(K).$$

equals the number of triples computed by Algorithm 6.2.

Remark 6.3. Step 1 of the algorithm requires working in an extension of $\mathbb{F}_p$. The degree of this extension depends on the splitting behavior of 2 in $\mathbb{O}_K$. An upper bound is given by $4[\mathbb{F}_p(A[2]) : \mathbb{F}_p] \le 24$, where $\mathbb{F}_p(A[2])$ denotes the field obtained by adjoining the coordinates of all 2-torsion points of A.

6.3. ***Igusa class polynomials.*** The CRT algorithm [Eisenträger and Lauter 2009] for computing the Igusa class polynomials $P_K, Q_K, R_K \in \mathbb{Q}[X]$ of a primitive quartic CM field K also computes the reductions of these 3 polynomials modulo primes p which split completely in the Hilbert class field of K_Φ. The method suggested in [Eisenträger and Lauter 2009] loops over all p^3 possible Igusa invariants and runs an endomorphism ring test for each triple $(j_1(A'), j_2(A'), j_3(A'))$, to see if A' has endomorphism ring $\mathbb{O}_K$.

We propose two main modifications to this algorithm. Firstly, we only demand that the primes p split completely in the subfield $K_\Phi(j_1(A), j_2(A), j_3(A))$ of the Hilbert class field of K_Φ that we obtain by adjoining the Igusa invariants of an abelian surface A that has CM by $\mathcal{O}_K$. To find such primes, we simply loop over $p = 5, 7, 11\dots$, and for the primes $(p) = \mathfrak{P}_1\mathfrak{P}_2\mathfrak{P}_3\mathfrak{P}_4 \subset \mathcal{O}_{K_\Phi}$ that split completely in K_Φ we test if

$$m(\mathfrak{P}_1) = (\mu) \subset \mathcal{O}_K \quad \text{and} \quad N(\mathfrak{P}_1) = \mu\bar{\mu},$$

with $\bar{\mu}$ the complex conjugate of μ. By [Shimura 1998, Section 15.3, Theorem 1], a prime p satisfying these conditions splits completely in $K_\Phi(j_1(A), j_2(A), j_3(A))$. It includes the primes that split completely in the Hilbert class field of K_Φ.

Our second modification is a big improvement in computing the Igusa class polynomial modulo p. Instead of looping over all $O(p^3)$ curves, we exploit the Galois action in a similar vein as in [Belding et al. 2008]. Below we give the complete algorithm to compute P_K, R_K, Q_K modulo a prime p meeting our conditions.

Step 1. Compute the class group

$$\mathrm{Cl}(\mathcal{O}_{K_\Phi}) = \langle \mathfrak{p}_1, \dots, \mathfrak{p}_k \rangle \tag{6-1}$$

of the reflex field, where we take degree 1 prime ideals $\mathfrak{p}_i$. For each of the ideals $\mathfrak{p}_i$ of *odd* norm $N_{K_\Phi/\mathbb{Q}}(\mathfrak{p}_i) = l_i$, compute the ideal $V(f; l_i)$ describing the Siegel modular variety $Y(f; l_i)^*$.

Step 2. Find an abelian surface $A/\mathbb{F}_p$ that has endomorphism ring isomorphic to $\mathcal{O}_K$, as follows. We factor $(p) \subseteq \mathcal{O}_K$ into primes $\mathfrak{P}_1, \overline{\mathfrak{P}}_1, \mathfrak{P}_2, \overline{\mathfrak{P}}_2$ and compute a generator π for the principal ideal $\mathfrak{P}_1\mathfrak{P}_2$. We compute the minimal polynomial f_π of π over $\mathbb{Q}$. We try *random* curves $C/\mathbb{F}_p$ until we find a curve with

$$\#C(\mathbb{F}_p) \in \{p + 1 \pm \mathrm{Tr}_{K/\mathbb{Q}}(\pi)\} \quad \text{and} \quad \#\,\mathrm{Jac}(C) \in \{f_\pi(1), f_\pi(-1)\}. \tag{6-2}$$

By construction, such a curve C has endomorphism algebra K. We test whether $\mathrm{Jac}(C)$ has endomorphism ring $\mathcal{O}_K$ using the algorithm of Freeman and Lauter [2008]. If it does, continue with Step 3, otherwise try more random curves C until we find one for which its Jacobian has endomorphism ring $\mathcal{O}_K$.

Step 3. Let $A/\mathbb{F}_p$ be the surface found in Step 2. The group $G = m(\mathrm{Cl}(\mathcal{O}_{K_\Phi}))$ acts in a natural way on A and we compute the set

$$G \cdot (j_1(A), j_2(A), j_3(A)) \subseteq S(K)$$

as follows. For $x = m(I)$ we write $I = \prod_i \mathfrak{p}_i^{a_i}$. The action of $\mathfrak{p}_1$ is computed using Algorithm 6.2 in case the norm of $\mathfrak{p}_1$ is odd and by applying a Richelot isogeny (see Section 3.2) if $\mathfrak{p}_1$ has norm 2. By successively applying the action of $\mathfrak{p}_1$, we compute the action of $\mathfrak{p}_1^{a_1}$. We then continue with the action of $\mathfrak{p}_2$, and so on. This allows us to compute the action of x on the surface A, and doing this for all x we compute the set $G \cdot (j_1(A), j_2(A), j_3(A))$. This part of the algorithm is analogous the one in [Belding et al. 2008].

Step 4. In contrast to genus 1 and the algorithm in [Belding et al. 2008], it is unlikely that *all* surfaces with endomorphism ring $\mathbb{O}_K$ are found. This is partly because we only find surfaces having the *same* CM type as the initial surface A, so in the dihedral case we are missing surfaces with the second CM type. Even in the cyclic case where there are $|\mathfrak{C}(K)|$ isomorphism classes, it is possible that the map

$$m : \mathrm{Cl}(\mathbb{O}_{K_\Phi}) \to \mathfrak{C}(K)$$

is not surjective, meaning that we do not find all surfaces of a given CM type. The solution is simple: compute the cardinality of $S(K)$ using Corollary 3.3, and if the number of surfaces found is less than $|S(K)|$, go back to Step 2.

Step 5. Once we have found all surfaces with endomorphism ring $\mathbb{O}_K$, expand

$$P_K \bmod p = \prod_{\{A \text{ p.p.a.s} \mid \mathrm{End}(A) = \mathbb{O}_K\}/\cong} (X - j_1(A)) \in \mathbb{F}_p[X]$$

and likewise for Q_K and R_K. The main difference with the method of [Eisenträger and Lauter 2009] is that we do *not* find all roots of P_K by a random search: we exploit the Galois action.

6.4. ***Run time analysis.*** We proceed with the run time analysis of the algorithm to compute the Igusa class polynomials using the 'modified CRT-approach' from Section 6.3. The input of the algorithm is a degree four CM field K. The discriminant D of K can be written as $D_1 D_0^2$, with D_0 the discriminant of the real quadratic subfield K^+ of K. We will give the run time in terms of D_1 and D_0.

First we analyze the size of the primes p used in the algorithm. The primes we use split completely in a subfield $S = K_\Phi(j_i(A))$ of the Hilbert class field of the reflex field K_Φ of K. If GRH holds, then there exists [Lagarias and Odlyzko 1977] an effectively computable constant $c > 0$, independent of K, such that the smallest such prime p satisfies

$$p \le c \cdot (\log|\mathrm{disc}(S/\mathbb{Q})|)^2,$$

where $\mathrm{disc}(S/\mathbb{Q})$ denotes the discriminant of the extension $S/\mathbb{Q}$. Since S is a totally unramified extension of K_Φ, we have

$$\mathrm{disc}(S/\mathbb{Q})^{1/[S:\mathbb{Q}]} = \mathrm{disc}(K_\Phi/\mathbb{Q})^{1/[K_\Phi:\mathbb{Q}]}$$

and we derive $\mathrm{disc}(S/\mathbb{Q}) = \mathrm{disc}(K_\Phi/\mathbb{Q})^{[S:K_\Phi]}$. Theorem 3.1 yields the bound $[S : K_\Phi] \le 4h^-(K)$, where $h^-(K) = |\mathrm{Cl}(\mathbb{O}_K)|/|\mathrm{Cl}(\mathbb{O}_{K^+})|$ denotes the *relative* class number of K. Using the bound (see [Louboutin 2003])

$$h^-(K) = \widetilde{O}(\sqrt{D_1 D_0}), \tag{6-3}$$

we derive that the smallest prime p is of size $\widetilde{O}(D_1 D_0)$. Here, the $\widetilde{O}$-notation indicates that factors that are of logarithmic order in the main term have been disregarded.

The Igusa class polynomials have rational coefficients, and at the moment the best known bound for the logarithmic height of the denominator of a coefficient is $\widetilde{O}(D_1^{3/2} D_0^{5/2})$. This bound is proven in [Streng 2010, Section 2.9] and is based on the denominator bounds in [Goren and Lauter 2010]. A careful analysis [Streng 2010, Section 2.11] yields that each coefficient of P_K, R_K, Q_K has logarithmic height $\widetilde{O}(D_1^{3/2} D_0^{5/2})$ as well. A standard argument as in [Belding et al. 2008, Lemma 5.3] shows that the $\widetilde{O}(D_1^{3/2} D_0^{5/2})$ primes that we need can be taken to be of size $\widetilde{O}(D_1^2 D_0^3)$ if GRH holds true. We find these primes in time $\widetilde{O}(D_1^2 D_0^3)$. We remark that better bounds on the denominators of the coefficients translate into better bounds on the size of the primes we need.

If GRH holds true, then the ideals $\mathfrak{p}_i$ in Step 1 can be chosen to have norm at most $12(\log D_1 D_0^2)^2$ by [Bach 1990]. Since the method from Section 5 for computing the ideal $V(f; N(\mathfrak{p}_i))$ is heuristic, we will rely on the following heuristic for our analysis.

Heuristic 6.4. Given a prime $l > 2$, we can compute generators for the ideal $V(f; l)$ in time polynomial in l.

At the moment, our computation of $V(f; l)$ only terminates in a reasonable amount of time for $l = 3$. However, *in theory* we only spend heuristic time $(\log D_1 D_0^2)^n$ in Step 1 for some $n \geq 2$ that is independent of D_1 and D_0. This is negligible compared to other parts of the algorithm.

We continue with the analysis of computing $P_K \bmod p$. As we think that the bound $p = \widetilde{O}(D_1^2 D_0^3)$ is too pessimistic, we will do the analysis in terms of both p and D_1, D_0. First we analyze the time spent on the random searches to find abelian surfaces with endomorphism ring $\mathcal{O}_K$. Every time we leave Step 2, we compute a factor $F \mid P_K \bmod p$ of the (first) Igusa class polynomial. Let $k \leq 2[\mathfrak{C}(K) : m(\mathrm{Cl}(\mathcal{O}_{K_\Phi}))]$ be the number of factors F we need to compute. The first time we invoke Step 2, we will with probability 1 compute a new factor F_1 of P_K. The second time we call Step 2 we need to ensure that we compute a *different* factor $F_2 \mid P_K$. Hence, we expect that we need to call Step 2 $k/(k-1)$ times to compute F_2. We see that we expect that we have to do Step 2

$$k(1 + 1/2 + \cdots + 1/k) = \widetilde{O}(k)$$

times to compute all factors $F_1, \ldots, F_k$.

Lemma 6.5. *We have* $[\mathfrak{C}(K) : m(\mathrm{Cl}(\mathcal{O}_{K_\Phi}))] \leq 2 \cdot 2^{6\omega(D)}$ *for any primitive quartic CM field* K, *where* $\omega(D)$ *denotes the number of prime divisors of* D.

Proof. We will bound the index of the image of the map $\tilde{m} : \mathrm{Cl}(\mathcal{O}_{K_\Phi}) \to \mathrm{Cl}(\mathcal{O}_K)$ inducing m. By Theorem 3.1, this index differs by at most a factor

$$|(\mathcal{O}_{K^+}^*)^+ / N_{K/K^+}(\mathcal{O}_K^*)| \leq 2$$

from $[\mathfrak{C}(K) : m(\mathrm{Cl}(\mathcal{O}_{K_\Phi}))]$.

If $K/\mathbb{Q}$ is dihedral with normal closure L, then the image of the norm map $N_{L/K} : \mathrm{Cl}(\mathcal{O}_L) \to \mathrm{Cl}(\mathcal{O}_K)$ has index at most 2 by class field theory. In the cyclic case, it is not hard to check that $\Im(\tilde{m})$ contains the squares. It suffices to bound the 2-torsion $\mathrm{Cl}(\mathcal{O}_K)$ in this case. The 2-rank of $\mathrm{Cl}(\mathcal{O}_K)$ is determined by *genus theory*. Using a combination of group cohomology and Nakayama's lemma, one can show [Rosen 2011] that the 2-rank is at most $6t$, with t the number of primes that ramify in the cyclic CM extension $K/\mathbb{Q}$. The lemma follows. □

We remark that outside a zero-density subset of very smooth integers, we have $\omega(n) < 2 \log\log n$ and we can then absorb the factor $\widetilde{O}(2^{6\omega(D)}) = 2^{6\omega(D)}\widetilde{O}(\log(D))$ into the $\widetilde{O}$-notation.

The probability that one of the random searches performed in this step will yield an abelian surface $\mathrm{Jac}(C)$ with endomorphism ring $\mathcal{O}_K$ is bounded from below by

$$h^-(K)/p^3 = \tilde{\Omega}(\sqrt{D_1 D_0}/p^3)$$

where we have used the effective lower bound $h^-(K) = \tilde{\Omega}(\sqrt{D_1 D_0})$ proved in [Louboutin 2003]. We therefore expect that we have to compute the number of points on C and on $\mathrm{Jac}(C)$ for

$$\widetilde{O}(p^3/\sqrt{D_1 D_0})$$

curves $C/\mathbb{F}_p$. Since point counting on genus 2 curves is polynomial time by [Pila 1990], this takes time $\widetilde{O}(p^3/\sqrt{D_1 D_0})$.

For all the curves $C/\mathbb{F}_p$ that satisfy equation (6-2), we have to check whether we have $\mathrm{End}(\mathrm{Jac}(C)) \cong \mathcal{O}_K$ or not. The probability that $\mathrm{End}(\mathrm{Jac}(C))$ is isomorphic to $\mathcal{O}_K$ is bounded from below by

$$\frac{h^-(K)}{\sum_{\mathcal{O}} h(\mathcal{O})}, \tag{6-4}$$

where the sum ranges over all orders $\mathcal{O} \subseteq \mathcal{O}_K$ that contain $\mathbb{Z}[\pi, \bar{\pi}]$. Assuming mild ramification conditions on the prime 2, there are only $O(\log n)$ orders $\mathcal{O} \subseteq \mathcal{O}_K$ of index n; see [Nakagawa 1996, Corollary 1]. We assume the following heuristic.

Heuristic 6.6. For any quartic CM field K, there are $O(\log n)$ orders $\mathcal{O} \subseteq \mathcal{O}_K$ of index n.

Justification of heuristic. As indicated in [Nakagawa 1996], the splitting condition on 2 is purely technical and should not affect the result. □

We can bound the class number $h(\mathbb{O})$ by $2[\mathbb{O}_K : \mathbb{Z}[\pi,\bar{\pi}]]h(\mathbb{O}_K)$ by [Stevenhagen 2008, Theorem 6.7]. It follows that we can bound the probability in (6-4) by

$$\Omega\left(\frac{1}{[\mathbb{O}_K : \mathbb{Z}[\pi,\bar{\pi}]]^{1+\varepsilon}h(\mathbb{O}_{K^+})}\right),$$

where we have used the bound n^ε for the number of divisors of n. Using the index bound

$$[\mathbb{O}_K : \mathbb{Z}[\pi,\bar{\pi}]] \le \frac{16p^2}{D_0\sqrt{D_1}}$$

from [Freeman and Lauter 2008, Proposition 6.1], we expect that we have to do

$$O\left(\frac{2^{6\omega(D)}p^{2+2\varepsilon}\sqrt{D_0}}{(\sqrt{D_1}D_0)^{1+\varepsilon}}\right)$$

endomorphism ring computations.

At the moment, the only known algorithm [Freeman and Lauter 2008] to test whether $\mathrm{End}(\mathrm{Jac}(C)) \cong \mathbb{O}_K$ holds has a run time $\widetilde{O}(p^{18})$, and one application of this algorithm dominates the computation of $P_K \in \mathbb{Q}[X]$. To make the run time analysis of our algorithm easier once a better algorithm to compute $\mathrm{End}(\mathrm{Jac}(C))$ has been found, we will use the bound $O(X)$ for the run time to compute $\mathrm{End}(\mathrm{Jac}(C))$. In total, we see that we spend

$$\widetilde{O}(2^{6\omega(D)}(p^3/\sqrt{D_1D_0} + (p^{2+2\varepsilon}\sqrt{D_0}/(\sqrt{D_1}D_0)^{1+\varepsilon})X))$$

time in all calls of Step 2.

The action of $\mathfrak{p}_i$ on $A/\mathbb{F}_p$ in Step 3 is computed in polynomial time in the norm l_i of $\mathfrak{p}_i$. As l_i is, under GRH, of polynomial size in $\log(D_1D_0^2)$, we spend time

$$\widetilde{O}(\sqrt{D_1D_0})$$

for every time we call Step 3. We call Step 3 as often as Step 2, so in total we spend time $\widetilde{O}(2^{6\omega(D)}\sqrt{D_1D_0})$ in Step 3.

The time spent in Step 4 is negligible, and the time spent in Step 5 is $\widetilde{O}(\sqrt{D_1D_0})$. Combining all five steps, we see that we compute $P_K \bmod p$ in time

$$\widetilde{O}(2^{6\omega(D)}(p^3/\sqrt{D_1D_0} + (p^{2+2\varepsilon}\sqrt{D_0}/(\sqrt{D_1}D_0)^{1+\varepsilon})X + \sqrt{D_1D_0})). \tag{6-5}$$

Theorem 6.7. *If GRH and Heuristic assumptions 6.4, 6.6 hold true, then we can compute the polynomials P_K, Q_K, R_K in probabilistic time*

$$\widetilde{O}(2^{6\omega(D)}(D_1^7D_0^{11} + XD_1^{5+\varepsilon}D_0^{8+2\varepsilon})).$$

Here, X denotes the run time of an algorithm that, given $A/\mathbb{F}_p$, decides whether $\mathrm{End}(A)$ *is isomorphic to $\mathbb{O}_K$ or not.*

Proof. Substitute $p = \widetilde{O}(D_1^2 D_0^3)$ in equation (6-5) to get the time per prime. The result follows from the fact that we need to compute P_K, R_K, Q_K modulo p for $\widetilde{O}(D_1^{3/2} D_0^{5/2})$ primes. □

We conclude this section with some remarks on the run time of our algorithm. At the moment, the main bottleneck is checking whether $\mathrm{End}(A) \cong \mathbb{O}_K$ holds or not. In Section 8 we show that a straightforward generalization of Kohel's algorithm [2008] is impossible and that a new approach is needed.

From a practical point of view, we are limited by the fact that we can only compute the ideal $V(f; 3)$ in a reasonable amount of time. By only using the primes lying over 2 and 3, we only use a subgroup of the group $\mathfrak{C}(K)$ giving the Galois action.

Even when these two problems are solved, there is a bottleneck not present in the genus 1 algorithm from [Belding et al. 2008]. The random searches take time $\widetilde{O}(p^3/\sqrt{D_1 D_0})$; even for the smallest prime p this is of size $\widetilde{O}(D_1^{5/2} D_0^{5/2})$.

Doing only the random searches for this prime already takes more time than it takes to write down the output $P_K, Q_K, R_K \in \mathbb{Q}[X]$. Hence, our algorithm is at the moment *not* quasilinear in the size of the output.

As noted in [Gruenewald 2010, Section 6], we can speed up this step of the algorithm by first computing a model for the Humbert surface describing all principally polarized abelian surfaces that have real multiplication by the quadratic subfield K^+ of K. We then perform our random search on this two-dimensional subspace of the three-dimensional moduli space. The time for the random searches would, for the smallest prime p, drop to

$$\widetilde{O}(D_1^{3/2} D_0^{3/2}).$$

Although this is less than the size of the output, our algorithm is not quasilinear once all primes p are taken into account.

To get a quasilinear algorithm, we think one should do the random searches on a one-dimensional subspace of the moduli space. This approach is an object of further study.

7. Examples and applications

In this section we illustrate our algorithm by computing the Igusa class polynomials modulo primes p for various CM fields. We point out the differences with the analogous genus 1 computations.

Example 7.1. In the first example we let $K = \mathbb{Q}[X]/(X^4 + 185X^2 + 8325)$ be a *cyclic* CM field of degree 4. All CM types are equivalent in this case, and the reflex field of K is K itself. The discriminant of K equals $5^2 \cdot 37^3$, and the real quadratic subfield of K is $K^+ = \mathbb{Q}(\sqrt{37})$. An easy computation shows that the narrow class group of K^+ is trivial. In particular, all ideal classes of K are principally polarizable, and we have

$$\mathfrak{C}(K) \cong \mathrm{Cl}(\mathbb{O}_K).$$

We compute $\mathrm{Cl}(\mathbb{O}_K) = \mathbb{Z}/10\mathbb{Z} = \langle \mathfrak{p}_3 \rangle$, where $\mathfrak{p}_3$ is a prime lying over 3. The prime ideal $\mathfrak{p}_3$ has norm 3, and its typenorm $N_\Phi(\mathfrak{p}_3)$ generates a subgroup of order 5 in $\mathrm{Cl}(\mathbb{O}_K)$.

The smallest prime that splits in the Hilbert class field of K is $p = 271$. We illustrate our algorithm by computing the Igusa class polynomials for K modulo this prime. First we do a 'random search' to find a principally polarized abelian surface over $\mathbb{F}_p$ with endomorphism ring $\mathbb{O}_K$ in the following way. We factor $(p) \subset \mathbb{O}_K$ into primes $\mathfrak{P}_1, \overline{\mathfrak{P}}_1, \mathfrak{P}_2, \overline{\mathfrak{P}}_2$ and compute a generator π of the principal $\mathbb{O}_K$-ideal $\mathfrak{P}_1\mathfrak{P}_2$. The element π has minimal polynomial

$$f = X^4 + 9X^3 + 331X^2 + 2439X + 73441 \in \mathbb{Z}[X].$$

If the Jacobian $\mathrm{Jac}(C)$ of a hyperelliptic curve C has endomorphism ring $\mathbb{O}_K$, then the Frobenius morphism of $\mathrm{Jac}(C)$ is a root of either $f(X)$ or $f(-X)$. With the factorization

$$f = (X - \tau_1)(X - \tau_2)(X - \tau_3)(X - \tau_4) \in K[X],$$

a *necessary* condition for $\mathrm{Jac}(C)$ to have endomorphism ring $\mathbb{O}_K$ is

$$\#C(\mathbb{F}_p) \in \{p + 1 \pm (\tau_1 + \tau_2 + \tau_3 + \tau_4)\} = \{261, 283\}$$

and

$$\#\mathrm{Jac}(C)(\mathbb{F}_p) \in \{f(1), f(-1)\} = \{71325, 76221\}.$$

We try random values $(j_1, j_2, j_3) \in \mathbb{F}_p^3$ and write down a hyperelliptic curve C with those Igusa invariants using Mestre's algorithm [Mestre 1991; Cardona and Quer 2005]. If C satisfies the 2 conditions above, then we check whether $\mathrm{Jac}(C)$ has endomorphism ring $\mathbb{O}_K$ using the algorithm in [Freeman and Lauter 2008]. If it passes this test, we are done. Otherwise, we select a new random value (j_1, j_2, j_3).

We find that $w_0 = (133, 141, 89)$ is a set of invariants for a surface $A/\mathbb{F}_p$ with endomorphism ring $\mathbb{O}_K$. We apply Algorithm 6.2 to w_0. The Igusa–Clebsch invariants corresponding to w_0 are $[133, 54, 82, 56]$. With the notation from Section 4, we have $s_2 = 162$, $s_3 = 106$, $s_5 = 128$, $s_6 = 30$. The Satake sextic polynomial

$$\mathcal{S} = X^6 + 190X^4 + 55X^3 + 82X^2 + 18X + 63 \in \mathbb{F}_p[X]$$

factors over $\mathbb{F}_{p^5}$ and we write $\mathbb{F}_{p^5} = \mathbb{F}_p(\alpha)$, where α satisfies $\alpha^5 + 2\alpha + 265 = 0$. We express the 6 roots of $\mathcal{S}$ in terms of α and pick

$$\begin{aligned}
f_1^4 &= 147\alpha^4 + 147\alpha^3 + 259\alpha^2 + 34\alpha + 110,\\
f_2^4 &= 176\alpha^4 + 211\alpha^3 + 14\alpha^2 + 134\alpha + 190,\\
f_3^4 &= 163\alpha^4 + 93\alpha^3 + 134\alpha^2 + 196\alpha + 115,\\
f_4^4 &= 226\alpha^4 + 261\alpha^3 + 99\alpha^2 + 9\alpha + 27
\end{aligned}$$

as values for the fourth powers of our Siegel modular functions. The fourth roots of $(f_1^4, f_2^4, f_3^4, f_4^4)$ are all defined over $\mathbb{F}_{p^{10}}$, but the proof of Theorem 4.2 shows that not every choice corresponds to the Igusa invariants of A. We pick fourth roots (r_1, r_2, r_3, r_4) such that the polynomial P_- from Section 4 vanishes when evaluated at $(T, f_1, f_2, f_3, f_4) = (\theta^4_{(0,1),(0,0)}, r_1, r_2, r_3, r_4)$. Here, $\theta^4_{(0,1),(0,0)}$ is computed from the Igusa–Clebsch invariants. For an arbitrary choice of fourth roots for r_1, r_2, r_3, there are two solutions $\pm r_4$ to $P_- = 0$. Indeed, if we take $\mathbb{F}_{p^{10}} = \mathbb{F}_p(\beta)$ with $\beta^{10}+\beta^6+133\beta^5+10\beta^4+256\beta^3+74\beta^2+126\beta+6=0$, then the tuple (r_1, r_2, r_3, r_4) given by

$$\begin{aligned}
r_1 &= 179\beta^9 + 69\beta^8 + 203\beta^7 + 150\beta^6 + 29\beta^5 + 258\beta^4 + 183\beta^3 + 240\beta^2 + 255\beta + 226,\\
r_2 &= 142\beta^9 + 105\beta^8 + 227\beta^7 + 244\beta^6 + 72\beta^5 + 155\beta^4 + 2\beta^3 + 129\beta^2 + 137\beta + 23,\\
r_3 &= 63\beta^9 + 112\beta^8 + 132\beta^7 + 244\beta^6 + 94\beta^5 + 40\beta^4 + 191\beta^3 + 263\beta^2 + 85\beta + 70,\\
r_4 &= 190\beta^9 + 41\beta^8 + 62\beta^7 + 170\beta^6 + 151\beta^5 + 240\beta^4 + 270\beta^3 + 56\beta^2 + 16\beta + 257
\end{aligned}$$

is a set of invariants for A together with some level 8-structure.

Next we specialize our ideal $V(f; 3)$ at $(W_1, X_1, Y_1, Z_1) = (r_1, r_2, r_3, r_4)$ and we solve the remaining system of 85 equations in 4 unknowns. Let (r_1', r_2', r_3', r_4') be the solution where

$$r_1' = 184\beta^9 + 48\beta^8 + 99\beta^7 + 83\beta^6 + 20\beta^5 + 232\beta^4 + 16\beta^3 + 223\beta^2 + 85\beta + 108.$$

The quadruple (r_1', r_2', r_3', r_4') are invariants of an abelian surface A' together with level 8-structure that is $(3, 3)$-isogenous to A. To map this quadruple to the Igusa invariants of A', we compute a root of the quadratic polynomial

$$P_-(T, r_1', r_2', r_3', r_4').$$

This root is a value for $\theta^4_{(0,1),(0,0)}$. Since we now know *all* theta fourth powers, we can apply the formulas relating theta functions and Igusa functions in Section 4.1 to find the Igusa triple (238, 10, 158).

In total, we find 16 Igusa triples defined over $\mathbb{F}_p$. All these triples are Igusa invariants of surfaces that have endomorphism *algebra* K. To check which ones have endomorphism *ring* $\mathcal{O}_K$, we apply the algorithm of Freeman and Lauter [2008]. We find that only the four triples

$$(253, 138, 96),\ (257, 248, 58),\ (238, 10, 158),\ (140, 159, 219)$$

are invariants of surfaces with endomorphism ring $\mathcal{O}_K$. The fact that we find 4 new sets of invariants should come as no surprise. Indeed, there are 4 ideals of norm 3 lying over 3 in $\mathcal{O}_K$ and each ideal gives us an isogenous surface.

Since the typenorm map $m : \mathrm{Cl}(\mathcal{O}_K) \to \mathfrak{C}(K)$ is not surjective, we are forced to do a *second* random search to find a 'new' abelian surface with endomorphism ring $\mathcal{O}_K$. We apply our isogeny algorithm to $w_1 = (74, 125, 180)$ as before, and we again find 4 new sets of invariants:

$$(174, 240, 246), (193, 85, 15), (268, 256, 143), (75, 263, 182).$$

In the end we expand the Igusa polynomials

$$\begin{aligned} P_K &= X^{10} + 92X^9 + 72X^8 + 217X^7 + 98X^6 \\ &\qquad + 195X^5 + 233X^4 + 140X^3 + 45X^2 + 123X + 171, \\ Q_K &= X^{10} + 232X^9 + 195X^8 + 45X^7 + 7X^6 \\ &\qquad + 195X^5 + 173X^4 + 16X^3 + 33X^2 + 247X + 237, \\ R_K &= X^{10} + 240X^9 + 57X^8 + 213X^7 + 145X^6 \\ &\qquad + 130X^5 + 243X^4 + 249X^3 + 181X^2 + 134X + 81 \end{aligned}$$

modulo $p = 271$.

Example 7.2. In the previous example, all the prime ideals of K lying over 3 gave rise to an isogenous abelian surface. This phenomenon does not always occur. Indeed, let K be a primitive quartic CM field and let $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ be the prime ideals of norm 3. If we have a principally polarized abelian surface $A/\mathbb{F}_p$ with endomorphism ring $\mathcal{O}_K$, then the number of (3, 3)-isogenous abelian surfaces with the same endomorphism ring equals the cardinality of

$$\{m(\mathfrak{p}_1), \dots, m(\mathfrak{p}_n)\}.$$

There are examples where this set has *less* than n elements.

Take the cyclic field $K = \mathbb{Q}[X]/(X^4 + 219X^2 + 10512)$. The class group of K is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. The prime 3 ramifies in K, and we have $(3) = \mathfrak{p}_1^2\mathfrak{p}_2^2$. The primes $\mathfrak{p}_1$ and $\mathfrak{p}_2$ in fact generate $\mathrm{Cl}(\mathcal{O}_K)$. It is easy to see that for this field we have

$$m(\mathfrak{p}_1) = m(\mathfrak{p}_2) \in \mathfrak{C}(K),$$

so we only find *one* isogenous surface.

Example 7.3. Our algorithm is not restricted to cyclic CM fields. In this example we let $K = \mathbb{Q}[X]/(X^4 + 22X^2 + 73)$ be a CM field with Galois group D_4. There are 2 equivalence classes of CM types. We fix a CM type $\Phi : K \to \mathbb{C}^2$ and let K_Φ be the reflex field for Φ. We have $K_\Phi = \mathbb{Q}[X]/(X^4 + 11X^2 + 12)$, and K and K_Φ have the same Galois closure L.

Since the real quadratic subfield $K^+ = \mathbb{Q}(\sqrt{3})$ has narrow class group $\mathbb{Z}/2\mathbb{Z}$, the group $\mathfrak{C}(K)$ fits in an exact sequence

$$1 \to \mathbb{Z}/2\mathbb{Z} \to \mathfrak{C}(K) \to \mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 1$$

and a close inspection yields $\mathfrak{C}(K) \cong \mathbb{Z}/4\mathbb{Z}$. The prime 3 factors as $(3) = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{p}_3^2$ in the reflex field, and we have $\mathrm{Cl}(\mathcal{O}_{K_\Phi}) = \mathbb{Z}/4\mathbb{Z} = \langle[\mathfrak{p}_1]\rangle$. The element $m(\mathfrak{p}_1) \in \mathfrak{C}(K)$ has order 4, and under the map

$$\mathfrak{C}(K) \xrightarrow{f} \mathrm{Cl}(\mathcal{O}_K) = \mathbb{Z}/4\mathbb{Z},$$

the element $f(m(\mathfrak{p}_1))$ has order 2. We see that even though the ideal $N_{L/K}(\mathfrak{p}_1\mathcal{O}_L)$ has order 2 in the class group, the typenorm of $\mathfrak{p}_1$ has order 4.

Of the 4 ideal classes of K, only 2 ideal classes are principally polarizable for Φ. The other 2 ideal classes are principally polarizable for ‘the other’ CM type. Furthermore, the two principally polarizable ideal classes each have *two* principal polarizations.

The prime $p = 1609$ splits completely in the Hilbert class field of K_Φ. As in Example 7.1, we do a random search to find that a surface $A/\mathbb{F}_p$ with Igusa invariants $w_0 = (1563, 789, 704) \in \mathbb{F}_p^3$ has endomorphism ring $\mathcal{O}_K$. We apply Algorithm 6.2 to this point. As output, we get w_0 again and two new points

$$w_1 = (1396, 1200, 1520) \quad \text{and} \quad w_2 = (1350, 1316, 1483).$$

The fact that we find w_0 again should come as no surprise since $m(\mathfrak{p}_3) \in \mathfrak{C}(K)$ is the trivial element. The points w_1 and w_2 correspond to $\mathfrak{p}_1$ and $\mathfrak{p}_2$.

As expected we compute that the cycle

$$w_0 = (1563, 789, 704) \xrightarrow{\mathfrak{p}_1} (1396, 1200, 1520) \xrightarrow{\mathfrak{p}_1} (1276, 1484, 7) \xrightarrow{\mathfrak{p}_1} (1350, 1316, 1483) \xrightarrow{\mathfrak{p}_1} w_0$$

has length 4. To find the full Igusa class polynomials modulo p, we do a second random search. The remaining 4 points are $(782, 1220, 257)$, $(1101, 490, 1321)$, $(577, 35, 471)$, $(1154, 723, 1456)$.

8. Obstruction to isogeny volcanoes

For an ordinary elliptic curve $E/\mathbb{F}_p$ over a finite field, Kohel [1996] introduced an algorithm to compute the endomorphism ring $\mathrm{End}(E)$, which has recently been improved in [Bisson and Sutherland 2011]. One first computes the endomorphism *algebra* K by computing the trace of the Frobenius morphism π of E. If the index $[\mathbb{O}_K : \mathbb{Z}[\pi]]$ is only divisible by small primes l, then Kohel's algorithm uses the l-isogeny graph to determine the endomorphism ring. The algorithm depends on the fact that the graph of l-isogenies looks like a volcano, and one can quotient by subgroups of order l to move down the volcano until one hits the bottom. We refer to [Fouquet and Morain 2002; Kohel 1996] for the details of this algorithm. This approach succeeds because of the following fact.

Lemma 8.1. *Let E, $E'/\mathbb{F}_p$ be two ordinary elliptic curves whose endomorphism rings are isomorphic to the same order $\mathbb{O}$ in an imaginary quadratic field K. Let $l \neq p$ be a prime such that the index $[\mathbb{O}_K : \mathbb{O}]$ is divisible by l. Then there are no isogenies of degree l between E and E'.*

Proof. This result is well known. Since the proof helps us understand what goes wrong in dimension 2, we give the short proof. Suppose that there does exist an isogeny $\varphi : E \to E'$ of degree l. By the Deuring lifting theorem [Lang 1987, Theorem 13.14], we can lift φ to an isogeny $\widetilde{\varphi} : \widetilde{E} \to \widetilde{E}'$ defined over the ring class field for $\mathbb{O}$. By CM theory, we can write $\widetilde{E}' = \mathbb{C}/I$ with I an *invertible* $\mathbb{O}$-ideal of norm l. But since l divides the index $[\mathbb{O}_K : \mathbb{O}]$, there are no invertible ideals of norm l. □

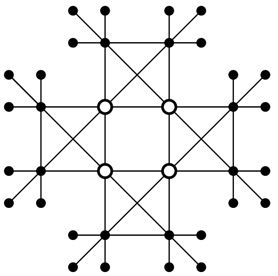
Unlike the elliptic curve case, there are a greater number of possibilities for the endomorphism ring of an (l, l)-isogenous abelian surface $A/\mathbb{F}_p$. Necessarily, the order must contain $\mathbb{Z}[\pi, \bar{\pi}]$, where π corresponds to the Frobenius endomorphism of A. Let $\varphi : A \to A'$ be an (l, l)-isogeny of principally polarized abelian surfaces where $\mathbb{O} = \mathrm{End}(A)$ contains $\mathbb{O}' = \mathrm{End}(A')$. Since φ splits multiplication by l, it follows that $\mathbb{Z} + l\mathbb{O} \subseteq \mathbb{O}' \subseteq \mathbb{O}$ and hence $\mathbb{O}'$ has index dividing l^3 in $\mathbb{O}$. In addition, since the $\mathbb{Z}$-rank is greater than two, it is possible to have several nonisomorphic suborders of $\mathbb{O}$ having the same index.

A natural question is whether the 'volcano approach' for elliptic curves can be generalized to ordinary principally polarized abelian surfaces $A/\mathbb{F}_p$. The extension of Schoof's algorithm [Pila 1990] enables us to compute the endomorphism algebra $K = \mathrm{End}(A) \otimes_{\mathbb{Z}} \mathbb{Q}$, and the problem is to compute the subring $\mathrm{End}(A) = \mathbb{O} \subseteq \mathbb{O}_K$. By working with explicit l-torsion points for primes $l \mid [\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$, one can determine this subring [Eisenträger and Lauter 2009; Freeman and Lauter 2008]. This approach requires working over large extension field of $\mathbb{F}_p$ and a natural question is whether we can generalize the volcano algorithm directly by using (l, l)-isogenies between abelian surfaces. However, the statement analogous to Lemma 8.1 — *there are no (l, l)-isogenies between A and A'* if $\mathrm{End}(A)$ and $\mathrm{End}(A')$ have isomorphic endomorphisms rings whose conductor in $\mathbb{O}_K$ divides l —

does *not* hold in general. This is a theoretical obstruction to a straightforward generalization of the algorithm for elliptic curves.

The next example shows that the analogue of Lemma 8.1 for abelian surfaces fails.

Example 8.2. In Example 7.3 we found the point $(782, 1220, 257) \in \mathbb{F}_{1609}^3$. Below we depict the connected component of the $(3, 3)$-isogeny graph. The white dots represent surfaces with endomorphism ring $\mathbb{O}_K$, and the black dots correspond to surfaces whose endomorphism ring is nonmaximal. The lattice of suborders of $\mathbb{O}_K$ of 3-power index that contain $\mathbb{Z}[\pi, \bar{\pi}]$ is completely described by the indices of the suborders in this case. We have $\mathbb{Z}[\pi, \bar{\pi}] \subset \mathbb{O}_{27} \subset \mathbb{O}_9 \subset \mathbb{O}_3 \subset \mathbb{O}_K$, where the subscript denotes the index in $\mathbb{O}_K$.



The leaf nodes all have endomorphism ring $\mathbb{O}_{27}$ and the remaining eight black vertices have endomorphism ring $\mathbb{O}_3$. We observe that there are cycles in this graph other than at the 'surface' of the volcano. ◇

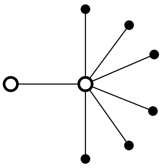
The reason that cycles can occur is the following. Just like in Lemma 8.1, we can lift an isogeny $\varphi : A \to A'$ to characteristic zero. By CM theory, we can now write $\widetilde{A} = \mathbb{C}^2/\Phi(I)$ for some invertible $\mathbb{O}$-ideal I. The isogenous surface $\widetilde{A}'$ equals $\mathbb{C}^2/\Phi(\mathfrak{a}^{-1}I)$ for an invertible $\mathbb{O}$-ideal $\mathfrak{a}$ of norm l^2. The difference from the elliptic curve case is that there *do* exist invertible $\mathbb{O}$-ideals of norm l^2. Hence, the isogeny graph for abelian surfaces need not look like a 'volcano'.

Another ingredient of the endomorphism ring algorithm for elliptic curves can fail. In the elliptic curve case, the following property of the l-isogeny graph is essential. Suppose that $E/\mathbb{F}_p$ has endomorphism ring $\mathbb{O}$ and let $\varphi : E \to E'$ be an isogeny from E to an elliptic curve with endomorphism ring of index l. If φ is defined over $\mathbb{F}_p$, then *all* $l + 1$ isogenies of degree l are defined over $\mathbb{F}_p$.

The analogous statement for dimension 2 is that *all* (l, l)-isogenies are defined over $\mathbb{F}_p$ as soon as there is one (l, l)-isogeny $\varphi : A \to A'$ that is defined over $\mathbb{F}_p$. Here, A' is an abelian surface with endomorphism ring of index dividing l^3. This statement is *not* true, as the following example shows.

Example 8.3. Consider the cyclic quartic CM field $K = \mathbb{Q}[X]/(X^4 + 12X^2 + 18)$, which has class number 2. The Igusa class polynomials have degree 2 and over $\mathbb{F}_{127}$

we find the corresponding moduli points $w_0 = (118, 71, 63)$ and $w_1 = (98, 82, 56)$. The isogeny graph is not regular:



The white dots represent the points having maximal endomorphism ring. There are 7 points isogenous to w_0, which includes w_1. One cannot identify w_1 from the graph structure alone. This demonstrates that the isogeny graph is insufficient to determine the endomorphism rings; the polarized CM lattices are also required. ◇

The shape of this graph can be explained as follows. Let $\pi \in \mathbb{O}_K$ correspond to the Frobenius morphism of a surface A belonging to the vertex w_1. If A' is (l, l)-isogenous to A, then A' is defined over $\mathbb{F}_p$ if and only if its endomorphism ring contains π. Since there are several orders of index dividing l^3 in $\mathbb{O}_K$, it can happen that π is contained in some of them, and not in others. In our example, the black points all have the same endomorphism ring $\mathbb{O}'$ with $\pi \in \mathbb{O}'$. The 33 other isogenous surfaces have an endomorphism ring that contains π^3, but not π.

Acknowledgments

The authors thank Nils Bruin, Everett Howe, David Kohel, and John Voight for helpful discussions and improvements to an early version the paper, the anonymous referee both for various detailed comments and for motivating us to include a run time analysis, and Mike Rosen [2011] for proving a genus theory result that allowed us to prove Lemma 6.5.

References

[Bach 1990] E. Bach, “Explicit bounds for primality testing and related problems”, *Math. Comp.* **55**:191 (1990), 355–380. MR 91m:11096 Zbl 0701.11075

[Baily and Borel 1966] W. L. Baily, Jr. and A. Borel, “Compactification of arithmetic quotients of bounded symmetric domains”, *Ann. of Math.* (2) **84** (1966), 442–528. MR 35 #6870 Zbl 0154.08602

[Belding et al. 2008] J. Belding, R. Bröker, A. Enge, and K. Lauter, “Computing Hilbert class polynomials”, pp. 282–295 in *Algorithmic number theory*, edited by A. J. van der Poorten and A. Stein, Lecture Notes in Comput. Sci. **5011**, Springer, Berlin, 2008. MR 2009j:11200 Zbl 1205.11139

[Bisson and Sutherland 2011] G. Bisson and A. V. Sutherland, “Computing the endomorphism ring of an ordinary elliptic curve over a finite field”, *J. Number Theory* **131**:5 (2011), 815–831. MR 2772473 Zbl 05876849

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, “The Magma algebra system, I: The user language”, *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR 1484478 Zbl 0898.68039

[Bost and Mestre 1988] J.-B. Bost and J.-F. Mestre, "Moyenne arithmético-géométrique et périodes des courbes de genre 1 et 2", *Gaz. Math.* **38** (1988), 36–64. MR 89k:14072 Zbl 0682.14031

[Bröker and Lauter 2009] R. Bröker and K. Lauter, "Modular polynomials for genus 2", *LMS J. Comput. Math.* **12** (2009), 326–339. MR 2010k:11096 Zbl 05947706

[Cardona and Quer 2005] G. Cardona and J. Quer, "Field of moduli and field of definition for curves of genus 2", pp. 71–83 in *Computational aspects of algebraic curves*, edited by T. Shaska, Lecture Notes Ser. Comput. **13**, World Sci. Publ., Hackensack, NJ, 2005. MR 2006h:14036 Zbl 1126.14031

[Carls et al. 2008] R. Carls, D. Kohel, and D. Lubicz, "Higher-dimensional 3-adic CM construction", *J. Algebra* **319**:3 (2008), 971–1006. MR 2010e:14042 Zbl 1140.14042

[Chai and Norman 1990] C.-L. Chai and P. Norman, "Bad reduction of the Siegel moduli scheme of genus two with $\Gamma_0(p)$-level structure", *Amer. J. Math.* **112**:6 (1990), 1003–1071. MR 91i:14033 Zbl 0734.14010

[Dupont 2006] R. Dupont, *Moyenne arithmético-géométrique, suites de Borchardt et applications*, thesis, École polytechnique, 2006.

[Eisenträger and Lauter 2009] K. Eisenträger and K. Lauter, "A CRT algorithm for constructing genus 2 curves over finite fields", pp. 161–176 in *Proceedings of Arithmetic, Geometry, and Coding Theory (AGCT-10)*, Séminaires & Congrès **21**, Société de Mathématique Francaise, Paris, 2009.

[Fouquet and Morain 2002] M. Fouquet and F. Morain, "Isogeny volcanoes and the SEA algorithm", pp. 276–291 in *Algorithmic number theory*, edited by C. Fieker and D. R. Kohel, Lecture Notes in Comput. Sci. **2369**, Springer, Berlin, 2002. MR 2005c:11077 Zbl 1058.11041

[Freeman and Lauter 2008] D. Freeman and K. Lauter, "Computing endomorphism rings of Jacobians of genus 2 curves over finite fields", pp. 29–66 in *Algebraic geometry and its applications*, Ser. Number Theory Appl. **5**, World Sci. Publ., Hackensack, NJ, 2008. MR 2010a:14042 Zbl 1151.14314

[Gaudry et al. 2006] P. Gaudry, T. Houtmann, D. Kohel, C. Ritzenthaler, and A. Weng, "The 2-adic CM method for genus 2 curves with application to cryptography", pp. 114–129 in *Advances in cryptology—ASIACRYPT 2006*, edited by X. Lai and K. Chen, Lecture Notes in Comput. Sci. **4284**, Springer, Berlin, 2006. MR 2009j:94110 Zbl 1172.94576

[van der Geer 1982] G. van der Geer, "On the geometry of a Siegel modular threefold", *Math. Ann.* **260**:3 (1982), 317–350. MR 84a:10028 Zbl 0473.14017

[Goren 1997] E. Z. Goren, "On certain reduction problems concerning abelian surfaces", *Manuscripta Math.* **94**:1 (1997), 33–43. MR 98m:14048 Zbl 0924.14023

[Goren and Lauter 2010] E. Z. Goren and K. Lauter, "Genus 2 curves with complex multiplication", preprint, 2010. arXiv 1003.4759

[Gruenewald 2008] D. Gruenewald, *Explicit Algorithms for Humbert Surfaces*, thesis, University of Sydney, 2008.

[Gruenewald 2010] D. Gruenewald, "Computing Humbert surfaces and applications", pp. 59–69 in *Arithmetic, geometry, cryptography and coding theory 2009* (Marseille, 2009), edited by D. Kohel and R. Rolland, Contemp. Math. **521**, Amer. Math. Soc., Providence, RI, 2010. MR 2744034 Zbl 05831703

[Igusa 1960] J.-i. Igusa, "Arithmetic variety of moduli for genus two", *Ann. of Math.* (2) **72** (1960), 612–649. MR 22 #5637 Zbl 0122.39002

[Igusa 1964] J.-i. Igusa, "On the graded ring of theta-constants", *Amer. J. Math.* **86** (1964), 219–246. MR 29 #2258 Zbl 0146.31703

[Igusa 1967] J.-i. Igusa, "Modular forms and projective invariants", *Amer. J. Math.* **89** (1967), 817–855. MR 37 #5217 Zbl 0159.50401

[Katz and Mazur 1985] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies **108**, Princeton University Press, 1985. MR 86i:11024 Zbl 0576.14026

[Kohel 1996] D. R. Kohel, *Endomorphism rings of elliptic curves over finite fields*, thesis, University of California, Berkeley, 1996, available at http://tinyurl.com/42lezaq. MR 2695524

[Kohel 2008] D. R. Kohel, "Complex multiplication and canonical lifts", pp. 67–83 in *Algebraic geometry and its applications*, edited by J. Chaumine et al., Ser. Number Theory Appl. **5**, World Sci. Publ., Hackensack, NJ, 2008. MR 2010d:14064 Zbl 1151.14329

[Lagarias and Odlyzko 1977] J. C. Lagarias and A. M. Odlyzko, "Effective versions of the Chebotarev density theorem", pp. 409–464 in *Algebraic number fields: L-functions and Galois properties* (Durham, NC, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR 56 #5506 Zbl 0362.12011

[Lang 1983] S. Lang, *Complex multiplication*, Grundlehren der Mathematischen Wissenschaften **255**, Springer, New York, 1983. MR 85f:11042 Zbl 0536.14029

[Lang 1987] S. Lang, *Elliptic functions*, 2nd ed., Graduate Texts in Mathematics **112**, Springer, New York, 1987. MR 88c:11028 Zbl 0615.14018

[Louboutin 2003] S. Louboutin, "Explicit lower bounds for residues at $s = 1$ of Dedekind zeta functions and relative class numbers of CM-fields", *Trans. Amer. Math. Soc.* **355**:8 (2003), 3079–3098. MR 2004f:11134 Zbl 1026.11085

[Mestre 1991] J.-F. Mestre, "Construction de courbes de genre 2 à partir de leurs modules", pp. 313–334 in *Effective methods in algebraic geometry* (Castiglioncello, 1990), edited by T. Mora and C. Traverso, Progr. Math. **94**, Birkhäuser, Boston, MA, 1991. MR 92g:14022 Zbl 0752.14027

[Mumford 1970] D. Mumford, *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics **5**, Tata Institute, Bombay, 1970. MR 44 #219 Zbl 0223.14022

[Nakagawa 1996] J. Nakagawa, *Orders of a quartic field*, Mem. Amer. Math. Soc. **583**, American Mathematical Society, Providence, RI, 1996. MR 96k:11125 Zbl 0865.11069

[Pila 1990] J. Pila, "Frobenius maps of abelian varieties and finding roots of unity in finite fields", *Math. Comp.* **55**:192 (1990), 745–763. MR 91a:11071 Zbl 0724.11070

[Poor and Yuen 2000] C. Poor and D. S. Yuen, "Linear dependence among Siegel modular forms", *Math. Ann.* **318**:2 (2000), 205–234. MR 2001j:11024 Zbl 0972.11035

[Rosen 2011] M. Rosen, "The p-rank of the class group in cyclic p-power extensions", in preparation, 2011.

[Runge 1993] B. Runge, "On Siegel modular forms: I", *J. Reine Angew. Math.* **436** (1993), 57–85. MR 94c:11041 Zbl 0772.11015

[Shimura 1998] G. Shimura, *Abelian varieties with complex multiplication and modular functions*, Princeton Mathematical Series **46**, Princeton University Press, 1998. MR 99e:11076 Zbl 0908.11023

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 96b:11074 Zbl 0911.14015

[Spallek 1994] A.-M. Spallek, *Kurven vom Geschlecht 2 und ihre Anwendung in Public-Key-Kryptosystemen*, thesis, Universität Gesamthochschule Essen, 1994.

[Stevenhagen 2008] P. Stevenhagen, "The arithmetic of number rings", pp. 209–266 in *Algorithmic number theory: lattices, number fields, curves and cryptography*, edited by J. P. Buhler and P. Stevenhagen, Math. Sci. Res. Inst. Publ. **44**, Cambridge Univ. Press, 2008. MR 2009k:11213 Zbl 05532104

[Streng 2010] M. Streng, *Complex multiplication of abelian surfaces*, thesis, Universiteit Leiden, 2010.

[van Wamelen 1999] P. van Wamelen, "Examples of genus two CM curves defined over the rationals", *Math. Comp.* **68**:225 (1999), 307–320. MR 99c:11079 Zbl 0906.14025

[Weng 2003] A. Weng, "Constructing hyperelliptic curves of genus 2 suitable for cryptography", *Math. Comp.* **72**:241 (2003), 435–458. MR 2003i:14029 Zbl 1013.11023

Communicated by Hendrik W. Lenstra
Received 2009-10-23 Revised 2011-04-18 Accepted 2011-07-12

reinier@math.brown.edu *Department of Mathematics, Brown University, Box 1917,*
151 Thayer Street, Providence, RI 02912, United States

davidg@maths.usyd.edu.au *Laboratoire de Mathématiques Nicolas Oresme, CNRS UMR*
5139, UFR Sciences, Campus 2, Boulevard Maréchal Juin,
Université de Caen Basse-Normandie, 14032 Caen cedex,
France

klauter@microsoft.com *Microsoft Research, One Microsoft Way,*
Redmond, WA 98052, United States

ALGEBRA AND NUMBER THEORY 5:4(2011)

On the cluster category of a marked surface without punctures

Thomas Brüstle and Jie Zhang

We study the cluster category $\mathscr{C}_{(S,M)}$ of a marked surface (S, M) without punctures. We explicitly describe the objects in $\mathscr{C}_{(S,M)}$ as direct sums of homotopy classes of curves in (S, M) and one-parameter families related to noncontractible closed curves in (S, M). Moreover, we describe the Auslander–Reiten structure of the category $\mathscr{C}_{(S,M)}$ in geometric terms and show that the objects without self-extensions in $\mathscr{C}_{(S,M)}$ correspond to curves in (S, M) without self-intersections. As a consequence, we establish that every rigid indecomposable object is reachable from an initial triangulation.

1. Introduction

We study the cluster category of a marked surface without punctures: Consider a compact connected oriented 2-dimensional bordered Riemann surface S and a finite set of marked points M lying on the boundary ∂S of S with at least one marked point on each boundary component. The condition $M \subset \partial S$ means that we do not allow the marked surface (S, M) to have punctures. (While some of the background results we need are valid in the more general context of punctured surfaces, the main result we need from [Assem et al. 2010] works only without punctures.) In [Fomin et al. 2008] a cluster algebra $\mathscr{A}(S, M)$ is associated to the marked surface (S, M). The initial seed of $\mathscr{A}(S, M)$ corresponds to a triangulation Γ of (S, M), and the mutation of a cluster variable corresponds to the flip of an arc in the triangulation (see [Fomin and Zelevinsky 2002] for the definition of a cluster algebra). The cluster algebra, defined by iterated mutations, is thus independent of the chosen triangulation Γ of (S, M).

The cluster category $\mathscr{C}_{(S,M)}$ providing a categorification of the algebra $\mathscr{A}(S, M)$ was defined in [Amiot 2009]. In fact, in [Labardini-Fragoso 2009a; Assem et al. 2010] a quiver with potential (Q_Γ, W_Γ) has been defined for each triangulation Γ

Thomas Brüstle is supported by NSERC. Jie Zhang is partially supported by China Scholarship Council and Bishop's University. Both authors stayed at INRIA Sophia Antipolis when writing this paper.

MSC2000: primary 16G99; secondary 16G20, 57N05, 57M50, 16G70.

Keywords: cluster category, marked surface.

of (S, M). Since the Jacobian algebra $J(Q_\Gamma, W_\Gamma)$ is finite-dimensional, one can use [Amiot 2009] to define the cluster category $\mathscr{C}_\Gamma$ associated to Γ. By [Keller and Yang 2009; Fomin et al. 2008; Labardini-Fragoso 2009a], this category $\mathscr{C}_\Gamma$ does not depend on the triangulation Γ of (S, M) and is just denoted by $\mathscr{C}_{(S,M)}$ (more precisely, $\mathscr{C}_\Gamma$ is triangle equivalent to $\mathscr{C}_{\Gamma'}$ if Γ' is another triangulation of (S, M)).

Cluster categories associated to quivers are object of intense investigation (see for example [Geiss et al. 2005; Caldero et al. 2006; Buan et al. 2006; 2008; Caldero and Chapoton 2006; Geiß et al. 2006; Assem et al. 2008]) and are understood explicitly in terms of representations of the quiver. The category $\mathscr{C}_{(S,M)}$, defined by a quiver with potential, is however quite difficult to compute and only a few cases of surfaces allow direct calculations. The aim of this paper is to provide an explicit description of the objects and the irreducible morphisms in the category $\mathscr{C}_{(S,M)}$ in terms of the surface (S, M), independent of the choice of a triangulation. In fact, the category $\mathscr{C}_{(S,M)}$ is closely related to module categories of string algebras, and a well-know classification theorem [Butler and Ringel 1987] describes all modules over a string algebra by listing the indecomposable modules in a list of string modules and a collection of band modules. Similarly, the indecomposable objects of $\mathscr{C}_{(S,M)}$ can be described as belonging to two classes which we call string objects and band objects. The cluster categories $\mathscr{C}_{(S,M)}$ are k-categories, where k denotes a fixed algebraically closed field. Moreover they are Krull–Schmidt; thus it is sufficient to classify indecomposable objects up to isomorphism.

Theorem 1.1. *A parametrization of the isoclasses of indecomposable objects in $\mathscr{C}_{(S,M)}$ is given by "string objects" and "band objects", where*

(1) *the string objects are indexed by the homotopy classes of noncontractible curves in (S, M) which are not homotopic to a boundary segment of (S, M), subject to the equivalence relation $\gamma \sim \gamma^{-1}$;*

(2) *the band objects are indexed by $k^* \times \Pi_1^*(S, M)/\sim$, where $k^* = k \setminus \{0\}$ and $\Pi_1^*(S, M)/\sim$ is given by the nonzero elements of the fundamental group of (S, M) subject to the equivalence relation generated by $a \sim a^{-1}$ and cyclic permutation.*

Remark 1.2. The curves we consider in Theorem 1.1 are allowed to have self-intersections.

The boundary of the surface S consists of a collection of disjoint circles, each one inheriting an orientation from the orientation of S. Given any curve γ in (S, M), we denote by ${}_s\gamma$ the curve obtained from γ by moving its starting point clockwise to the next marked point on the boundary. Likewise, moving the ending point of γ clockwise along the boundary to the next marked point, one obtains a curve γ_e from γ. These moves establish the irreducible morphisms between string objects, as explained in the following theorem. Note that in the special case where S is a disc, this description of irreducible morphisms has been given in [Caldero et al. 2006], establishing one of the first categorifications of a cluster algebra (of type A).

In the next few statements, we use the abbreviation AR to refer to the Auslander–Reiten structure.

Theorem 1.3. *Let γ be a noncontractible curve in (S, M) that is not homotopic to a boundary segment of (S, M). Then there is an AR-triangle in $\mathscr{C}_{(S,M)}$ as follows, where ${}_s\gamma$ or γ_e are zero objects in $\mathscr{C}_{(S,M)}$ if they are boundary segments*:

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1].$$

Moreover, all AR-triangles between string objects in $\mathscr{C}_{(S,M)}$ are of this form.

We obtain also the AR-translation in the cluster category $\mathscr{C}_{(S,M)}$:

Proposition 1.4. *The AR-translation in $\mathscr{C}_{(S,M)}$ is given by the simultaneous counterclockwise rotation of the starting and ending points of a curve to the next marked points at the boundary.*

Moreover, from Theorem 1.3 one can deduce the shape of the AR-components in the category $\mathscr{C}_{(S,M)}$. In fact, each boundary component of S with t marked points gives rise to a tube of rank t in $\mathscr{C}_{(S,M)}$. All other components formed by string objects are composed of meshes with exactly two middle terms.

We also study the effect of a change of the triangulation. It is well-known that any two triangulations of (S, M) can be transformed into each other by a sequence of flips $f_i(\Gamma)$ which locally change one arc in Γ. On the other hand, each triangulation yields a cluster-tilting object T in $\mathscr{C}_{(S,M)}$, and there one can apply mutations $\mu_i(T)$ locally changing one summand. We show that these two operations are compatible:

Theorem 1.5. *Each triangulation Γ of (S, M) yields a cluster-tilting object T_Γ in $\mathscr{C}_{(S,M)}$, and*

$$\mu_i(T_\Gamma) = T_{f_i(\Gamma)}.$$

We then study the relation between extensions in the category $\mathscr{C}_{(S,M)}$ and intersections of the corresponding curves. Given any two curves in (S, M) with intersections, we explicitly construct one or two new curves, sometimes resolving the intersection, and sometimes increasing the winding number. These new curves serve as middle term of certain nonsplit short exact sequences which allow to prove the following theorem and corollary:

Theorem 1.6. *Curves in (S, M) without self-intersections correspond to the indecomposable objects without self-extensions in $\mathscr{C}_{(S,M)}$.*

Corollary 1.7. *There is a bijection between triangulations of (S, M) and cluster-tilting objects of $\mathscr{C}_{(S,M)}$. In particular, each indecomposable object without self-extension is reachable from the cluster-tilting object T_Γ (the initial cluster-tilting object).*

This paper is organized as follows: we first describe in Section 2 the indecomposable objects in $\mathscr{C}_{(S,M)}$, using the description of modules over a string algebra. In Section 3 we describe the irreducible morphisms in $\mathscr{C}_{(S,M)}$ and study the shape of its AR-components. Section 4 is devoted to compare the effect of a flip of an arc in the triangulation and a mutation of the corresponding cluster-tilting object. We also compare it to results obtained for decorated representations of quivers with potential. Finally, in Section 5 we prove that curves without self-intersections correspond to objects without self-extensions in $\mathscr{C}_{(S,M)}$ and we establish the bijection between triangulations and cluster-tilting objects.

2. Indecomposable objects in $\mathscr{C}_{(S,M)}$

In this section, we choose a triangulation Γ of (S, M), and by studying the corresponding Jacobian algebra, we give a geometric characterization of the indecomposable objects in $\mathscr{C}_{(S,M)}$. By a curve in (S, M), we mean a continuous function $\gamma : [0, 1] \to S$ with $\gamma(0), \gamma(1) \in M$. A closed curve is one with $\gamma(0) = \gamma(1)$, and a simple curve is one where γ is injective, except possibly at the endpoints. For a curve γ, we denote by γ^{-1} the inverse curve $\gamma^{-1} : [0, 1] \to S,\ t \mapsto \gamma(1-t)$. We always consider curves up to homotopy, and for any collection of curves we implicitly assume that their mutual intersections are minimal possible in their respective homotopy classes. We recall from [Fomin et al. 2008] the definition of a triangulation:

Definition 2.1. An arc δ in (S, M) is a simple noncontractible curve in (S, M). The boundary of S is a disjoint union of circles, which are subdivided by the points in M into boundary segments. We call an arc δ a *boundary arc* if it is homotopic to such a boundary segment. Otherwise, δ is said to be an *internal arc*. A *triangulation* of (S, M) is a maximal collection Γ of arcs that do not intersect except at their endpoints. We call a triangle $\triangle$ in Γ an *internal triangle* if all edges of $\triangle$ are internal arcs.

Proposition 2.2 [Fomin et al. 2008]. *In each triangulation of (S, M), the number of internal arcs is*

$$n = 6g + 3b + c - 6$$

where g is the genus of S, b is the number of boundary components, and $c = |M|$ is the number of marked points.

***The quiver with potential* (Q_Γ, W_Γ).** We recall from [Assem et al. 2010; Labardini-Fragoso 2009a] that each triangulation Γ of (S, M) yields a quiver Q_Γ with potential W_Γ:

(1) $Q_\Gamma = (Q_0, Q_1)$, where the set of vertices Q_0 is given by the internal arcs of Γ, and the set of arrows Q_1 is defined as follows: whenever there is a triangle $\triangle$ in Γ containing two internal arcs a and b, there is an arrow $\rho : a \to b$ in Q_1 if a is a predecessor of b with respect to clockwise orientation at the joint vertex of a and b in $\triangle$.

(2) Every internal triangle $\triangle$ in Γ gives rise to an oriented cycle $\alpha_\triangle \beta_\triangle \gamma_\triangle$ in Q, unique up to cyclic permutation of the factors $\alpha_\triangle$, $\beta_\triangle$, $\gamma_\triangle$. We define

$$W_\Gamma = \sum_\triangle \alpha_\triangle \beta_\triangle \gamma_\triangle$$

where the sum runs over all internal triangles $\triangle$ of Γ.

Unless we compare two different triangulations, we omit the subscript and denote the quiver with potential defined by Γ just with (Q, W). We refer to [Derksen et al. 2008] for more details on quivers with potentials, such as the definition of the Jacobian algebra $J(Q, W)$ associated to (Q, W). From [Assem et al. 2010] we know that $J(Q, W)$ is a finite-dimensional string algebra provided (Q, W) is defined by a triangulation of a marked surface as above.

Moreover, we denote by $\Lambda_{(Q,W)}$ the corresponding Ginzburg dg-algebra (see [Ginzburg 2006] for more details), and denote by $D(\Lambda_{(Q,W)})$ its derived category [Keller 2006]. The *cluster category* $\mathscr{C}_\Gamma = \mathscr{C}_{(Q,W)}$ associated to (Q, W) is defined in [Amiot 2009] as the quotient of triangulated categories $\operatorname{Per}\Lambda_{(Q,W)}/D^b(\Lambda_{(Q,W)})$ where $\operatorname{Per}\Lambda_{(Q,W)}$ is the thick subcategory of $D(\Lambda_{(Q,W)})$ generated by $\Lambda_{(Q,W)}$ and $D^b(\Lambda_{(Q,W)})$ is the full subcategory of $D(\Lambda_{(Q,W)})$ of the dg-modules whose homology is of finite total dimension.

Theorem 2.3 [Amiot 2009; Koenig and Zhu 2008]. *Suppose (Q, W) is a quiver with potential whose Jacobian algebra $J(Q, W)$ is finite-dimensional.*

(1) *$\mathscr{C}_\Gamma$ is 2-Calabi–Yau, Hom-finite and the image T_Γ of the free module $\Lambda_{(Q,W)}$ in the quotient* $\operatorname{Per}\Lambda_{(Q,W)}/D^b(\Lambda_{(Q,W)})$ *is a cluster-tilting object.*

(2) *$\mathscr{C}_\Gamma/T_\Gamma$ is equivalent to* $\operatorname{mod} J(Q, W)$, *the category of finite dimensional modules over $J(Q, W)$. Moreover, the projection functor $\mathscr{C}_\Gamma \to$* $\operatorname{mod} J(Q, W)$ *is given by* $\operatorname{Ext}^1_{\mathscr{C}_\Gamma}(T_\Gamma, -)$.

As explained in the introduction, the category $\mathscr{C}_\Gamma$ is (up to triangle equivalence) independent of the choice of a triangulation of (S, M), and is therefore denoted by $\mathscr{C}_{(S,M)}$. Moreover, it is a Krull–Schmidt category, so it is sufficient to describe all indecomposable objects up to isomorphism in order to describe its objects. From

Theorem 2.3(2), the indecomposable objects in $\mathscr{C}_{(S,M)}$ are either indecomposable modules over the Jacobian algebra $J(Q, W)$ or one of the $|Q_0| = 6g+3b+|M|-6$ summands of T_Γ. In order to give a description of all indecomposable objects in $\mathscr{C}_{(S.M)}$, we first study the Jacobian algebra $J(Q, W)$.

The string algebra $J(Q, W)$. We now recall some basic definitions related to string algebras and prove Theorem 1.1. Recall from [Butler and Ringel 1987] that a finite-dimensional algebra A is a *string algebra* if there is a quiver Q and an admissible ideal I such that $A = kQ/I$ and the following conditions hold:

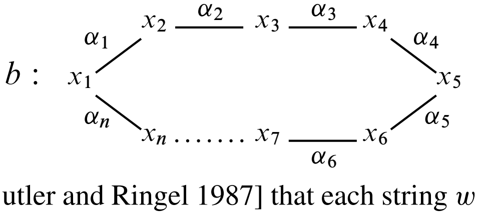
(S1) At each vertex of Q start at most two arrows and stop at most two arrows.

(S2) For each arrow α there is at most one arrow β and at most one arrow δ such that $\alpha\beta \notin I$ and $\delta\alpha \notin I$.

Given an arrow $\beta \in Q_1$, let $s(\beta)$ be its starting point and $e(\beta)$ its ending point. We denote β^{-1} the formal inverse of β with $s(\beta^{-1}) = e(\beta)$ and $e(\beta^{-1}) = s(\beta)$. A word $w = \alpha_n\alpha_{n-1}\cdots\alpha_1$ of arrows and their formal inverses is called a *string* if $\alpha_{i+1} \neq \alpha_i^{-1}$, $e(\alpha_i) = s(\alpha_{i+1})$ for all $1 \leq i \leq n-1$, and no subword nor its inverse belongs to I. Thus a string w can be viewed as a walk in the quiver Q avoiding the zero relations defining the ideal I:

$$w:\quad x_1 \overset{\alpha_1}{\text{———}} x_2 \overset{\alpha_2}{\text{———}} \cdots x_{n-1} \overset{\alpha_{n-1}}{\text{———}} x_n \overset{\alpha_n}{\text{———}} x_{n+1}$$

where x_i are vertices of Q and α_i are arrows in either direction. We denote by $s(w) = s(\alpha_1)$ and $e(w) = e(\alpha_n)$ the starting point and the ending point of w, respectively. For technical reasons, we also consider the empty string which we also call zero string. A string w is called *cyclic* if the first vertex x_1 and the last vertex x_{n+1} coincide. A *band* $b = \alpha_n\alpha_{n-1}\cdots\alpha_2\alpha_1$ is defined to be a cyclic string b such that each power b^m is a string, but b itself is not a proper power of any string. Thus b can be viewed as a cyclic walk:



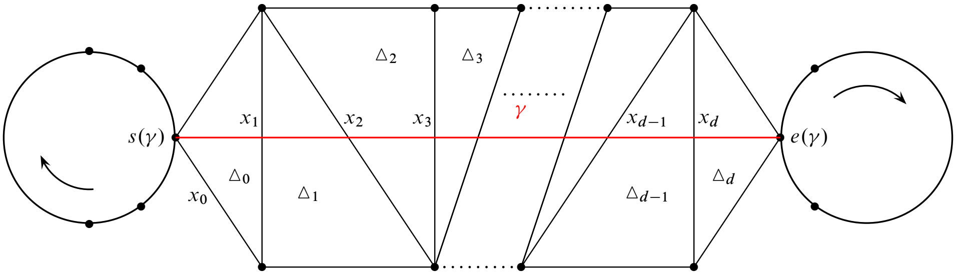
We recall from [Butler and Ringel 1987] that each string w in A defines a *string module* $M(w)$ in $\operatorname{mod} A$. The underlying vector space of $M(w)$ is obtained by replacing each x_i in w by a copy of the field k. The action of an arrow α of Q on $M(w)$ is induced by the relevant identity morphisms if α lies on w, and is zero otherwise. For the zero string 0, we let $M(0)$ be the zero module. Each band b defines a family of band modules $M(b, n, \phi)$ with $n \in \mathbb{N}$ and $\phi \in \operatorname{Aut}(k^n)$ by replacing each x_i in b a copy of the vector space k^n, and the action of an arrow α

on $M(b, n, \phi)$ is induced by identity morphisms if $\alpha = \alpha_j$ for $j = 1, 2 \dots n-1$ and by ϕ if $\alpha = \alpha_n$ (see [Butler and Ringel 1987]).

Let Γ be a triangulation of the marked surface (S, M), and denote by (Q, W) the corresponding quiver with potential. In [Assem et al. 2010] the strings and bands of $J(Q, W)$ are related to the noncontractible curves and simple noncontractible closed curves respectively in (S, M): For two curves γ', γ in (S, M) we denote by $I(\gamma', \gamma)$ the minimal intersection number of two representatives of the homotopic classes of γ' and γ. For each curve γ in (S, M) with

$$d = \sum_{\gamma' \in \Gamma} I(\gamma', \gamma)$$

we fix an orientation of γ, and let $x_1, x_2, \dots, x_d$ be the internal arcs of Γ that intersect γ in the fixed orientation of γ, as in the figure:



Here we denote by $s(\gamma) = \gamma(0)$ the starting point of γ and by $e(\gamma) = \gamma(1)$ its endpoint. Both points lie on the boundary, indicated by the circles in the preceding figure. Along its way, the curve γ is passing through (not necessarily distinct) triangles $\triangle_0, \triangle_1, \dots, \triangle_d$. Thus we obtain a string $w(\gamma)$ in $J(Q, W)$:

$$w(\gamma) : \quad x_1 \overset{\alpha_1}{\text{———}} x_2 \overset{\alpha_2}{\text{———}} \cdots x_{d-2} \overset{\alpha_{d-2}}{\text{———}} x_{d-1} \overset{\alpha_{d-1}}{\text{———}} x_d .$$

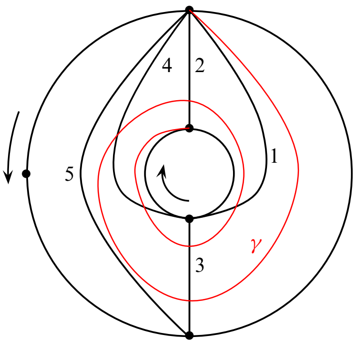
We recall the following result concerning the map $\gamma \mapsto w(\gamma)$:

Theorem 2.4 [Assem et al. 2010]. *Given a triangulation Γ of a marked surface (S, M), the map $\gamma \mapsto w(\gamma)$ is a bijection between the homotopy classes of noncontractible curves in (S, M) not homotopic to an arc in Γ and the strings of $J(Q, W)$.*

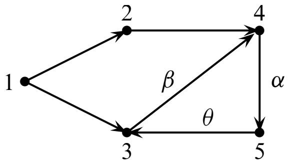
Similarly, each noncontractible closed curve b in S defines a cyclic walk in $J(Q, W)$. If b is not a proper power of any element in $\Pi_1^*(S, M)$, then it defines a band $w(b)$.

Remark 2.5. It will be convenient later to assume that each element $b \in \Pi_1^*(S, M)$ is given by a noncontractible closed curve b with endpoints $b(0) = b(1)$ lying on a marked point. Remark that as an element in $\Pi_1^*(S, M)$, b does not depend on the choice of the marked point.

Example 2.6. We consider an annulus with two marked points on each boundary and a triangulation Γ with internal arcs 1, 2, 3, 4, 5 as follows:



The associated quiver Q_Γ is cluster-tilted of type $\tilde{A}_4$ with potential $W_\Gamma = \alpha\beta\theta$:



From the diagram at the top of the page we read off

$$w(\gamma): \quad 3 \longrightarrow 4 \longleftarrow 2 \longleftarrow 1 \longrightarrow 3 \longrightarrow 4$$

where γ is the curve in red, and

$$w(b): \quad \begin{tikzcd} & 2 \arrow[dr] & \\ 1 \arrow[ur] \arrow[dr] & & 4 \\ & 3 \arrow[ur] & \end{tikzcd}$$

where b is a simple closed curve representing a generator of the fundamental group of the annulus.

Proof of Theorem 1.1. Let Γ be a triangulation of (S, M). As explained on page 533, the indecomposable objects in $\mathscr{C}_{(S,M)}$ are either given by indecomposable modules over the Jacobian algebra $J(Q, W)$, or they correspond to the indecomposable summands of T_Γ, thus to the internal arcs in Γ. The (finite-dimensional) indecomposable modules over a string algebra A are classified in [Butler and Ringel 1987]: Each indecomposable A-module is (isomorphic to) a string or a band module. The string module $M(w)$ is isomorphic to the string module $M(w^{-1})$ defined by the inverse string w^{-1}, and the band module $M(b, n, \phi)$ is isomorphic to $M(b', n, \phi)$ whenever b' is obtained from b by inversion or cyclic permutation. Apart from that, there are no isomorphisms between string or band modules.

Thus each noncontractible curve γ that is not homotopic to a boundary segment of (S, M) corresponds to an indecomposable object in $\mathscr{C}_{(S,M)}$. If γ is not an internal arc in Γ, it corresponds to the string module $M(w(\gamma))$. From Theorem 2.4 and what we have discussed above we conclude that two such curves γ, δ are isomorphic as objects in $\mathscr{C}_{(S,M)}$ precisely when γ is homotopic to δ or to its inverse δ^{-1}. We refer to these objects as the string objects or curves in (S, M), as described in part (1) of Theorem 1.1.

The remaining indecomposable objects in $\mathscr{C}_{(S,M)}$ correspond to the band modules $M(b, n, \phi)$ over $J(Q, W)$; we refer to them as band objects. They are parametrized by a positive integer n, an automorphism ϕ of k^n which is given by an element of k^* since k is algebraically closed, and a band b of $J(Q, W)$. The fundamental group $\Pi_1(S, M)$ is a free group with a finite number of generators which are given by simple noncontractible closed curves in S (see [Küçük 2005]). In order to avoid counting curves with opposite orientation twice we consider the elements in $\Pi_1(S, M)$ up to the equivalence relation $a \sim a^{-1}$.

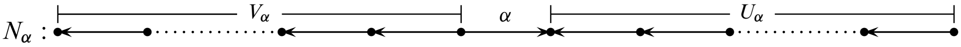
Moreover, to comply with the definition of a band module, we write each element a in $\Pi_1(S, M)$ as $a = b^n$ (multiplicatively written) for some $b \in \Pi_1(S, M)$ which itself is not a proper power of an element in $\Pi_1(S, M)$. Furthermore, we consider the elements of $\Pi_1(S, M)$ up to cyclic permutation of their factors. Then it is clear that the band modules $M(b, n, \phi)$ over $J(Q, W)$ correspond bijectively to $k^* \times \Pi_1^*(S, M)/\sim$, where $\Pi_1^*(S, M)/\sim$ is given by the nonzero elements of the fundamental group of (S, M) subject to the equivalence relation generated by $a \sim a^{-1}$ and cyclic permutation. This is the description of the band objects in $\mathscr{C}_{(S,M)}$ given in part (2) of Theorem 1.1. □

3. Irreducible morphisms in $\mathscr{C}_{(S,M)}$

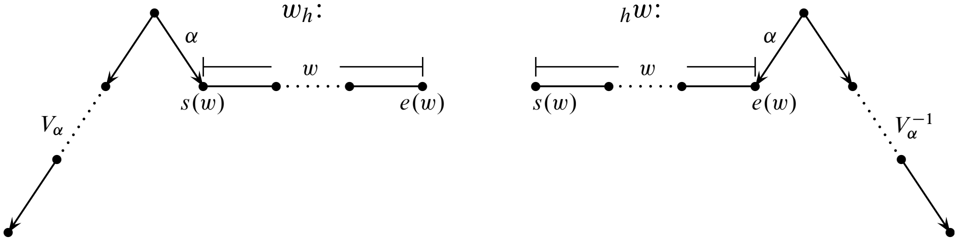
Based on the geometric characterization of the indecomposable objects in $\mathscr{C}_{(S,M)}$ in the previous section, we study in this section the irreducible morphisms in $\mathscr{C}_{(S,M)}$.

The AR-quiver of a string algebra. We first recall some basic definitions from [Butler and Ringel 1987]. Let $A = kQ/I$ be a finite-dimensional string algebra with $Q = (Q_0, Q_1)$ and $\mathscr{S}$ the set of all strings in A. A string w *starts* (*or ends*) *on a peak* if there is no arrow $\alpha \in Q_1$ with $w\alpha \in \mathscr{S}$ (or $\alpha^{-1}w \in \mathscr{S}$); likewise, a string w *starts* (*or ends*) *in a deap* if there is no arrow $\beta \in Q_1$ with $w\beta^{-1} \in \mathscr{S}$ (or $\beta w \in \mathscr{S}$).

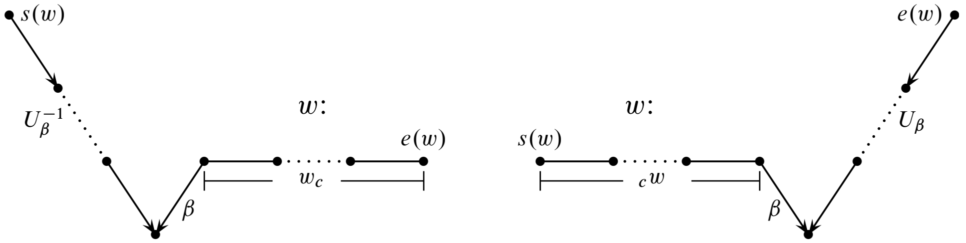
A string $w = \alpha_1\alpha_2 \cdots \alpha_n$ with all $\alpha_i \in Q_1$ is called *direct string*, and a string of the form w^{-1} where w is a direct string is called *inverse string*. Strings of length zero are both direct and inverse. For each arrow $\alpha \in Q_1$, let $N_\alpha = U_\alpha \alpha V_\alpha$ be the unique string such that U_α and V_α are inverse strings and N_α starts in a deep and ends on a peak. (See figure.)



If the string w does not start on a peak, we define $w_h = w\alpha V_\alpha$ and say that w_h is obtained from w by adding a hook on the starting point $s(w)$. Dually, if w does not end on a peak, we define ${}_hw = {V_\alpha}^{-1}\alpha^{-1}w$ and say that ${}_hw$ is obtained from w by *adding a hook* on the ending point $e(w)$. (See figure.)



Suppose now that the string w starts on a peak. If w is not a direct string, we can write $w = w_c\beta^{-1}\gamma_1\gamma_2\cdots\gamma_r = w_c\beta^{-1}U_\beta^{-1}$ for some $\beta \in Q_1$ and $r \geq 0$. We say in this case that w_c is obtained from w by *deleting a cohook* on $s(w)$. If w is a direct string, we define $w_c = 0$. Dually, assume that w ends on a peak. Then, if w is not an inverse string, we can write $w = {\gamma_r}^{-1}\cdots{\gamma_2}^{-1}{\gamma_1}^{-1}\beta_c w = U_\beta\beta_c w$ for some $\beta \in Q_1$ and $r \geq 0$. We say that ${}_cw$ is obtained from w by deleting a cohook on $e(w)$, and if w is an inverse string, we define ${}_cw = 0$. (See figure.)



Theorem 3.1 [Butler and Ringel 1987]. *For a string algebra A, let w be a string such that $M(w)$ is not an injective A-module. Then the AR-sequence starting in $M(w)$ is given,*

(1) *if w neither starts nor ends on a peak, by*

$$0 \longrightarrow M(w) \longrightarrow M(w_h) \oplus M({}_hw) \longrightarrow M({}_hw_h) \longrightarrow 0;$$

(2) *if w does not start but ends on a peak, by*

$$0 \longrightarrow M(w) \longrightarrow M(w_h) \oplus M({}_cw) \longrightarrow M({}_cw_h) \longrightarrow 0;$$

(3) *if w starts but does not end on a peak, by*

$$0 \longrightarrow M(w) \longrightarrow M(w_c) \oplus M({}_hw) \longrightarrow M({}_hw_c) \longrightarrow 0;$$

(4) *if w both starts and ends on a peak, by*

$$0 \longrightarrow M(w) \longrightarrow M(w_c) \oplus M({}_cw) \longrightarrow M({}_cw_c) \longrightarrow 0.$$

Remarks 3.2. (1) Each arrow $\alpha \in Q_1$ defines a string $N_\alpha = U_\alpha \alpha V_\alpha$ that starts in a deep and ends on a peak. Then $w = U_\alpha$ does not start but ends on a peak, hence we have an AR-sequence by Theorem 3.1:

$$0 \longrightarrow M(w) \longrightarrow M(w_h) \oplus M({}_c w) \longrightarrow M({}_c w_h) \longrightarrow 0.$$

Here $M({}_c w) = 0$ since w is an inverse string, and $w_h = N_\alpha$ and ${}_c w_h = {}_c(N_\alpha) = V_\alpha$ by definition. Hence, for each $\alpha \in Q_1$, there is an AR-sequence with an indecomposable middle term in mod A:

$$0 \longrightarrow M(U_\alpha) \longrightarrow M(N_\alpha) \longrightarrow M(V_\alpha) \longrightarrow 0.$$

In fact, the AR-sequences between string modules which admit only one direct summand in the middle term are indexed by the arrows in Q (see [Butler and Ringel 1987]).

(2) If the string module $M(w)$ is injective, then w both starts and ends on a peak. Thus we can write

$$w = \gamma_r^{-1}\gamma_{r-1}^{-1}\cdots\gamma_1^{-1}\beta_1\beta_2\cdots\beta_{s-1}\beta_s \quad \text{or}$$
$$w = U_{\beta_1}\beta_1\beta_2\cdots\beta_{s-1}\beta_s = \gamma_r^{-1}\gamma_{r-1}^{-1}\cdots\gamma_1^{-1}U_{\gamma_1}^{-1},$$

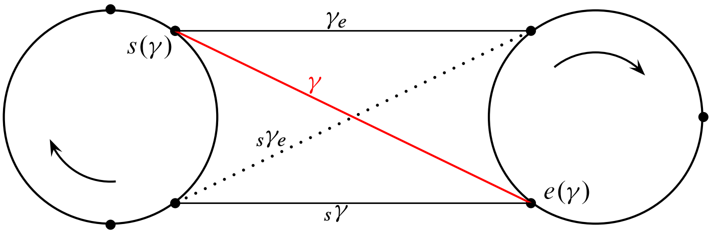
where $r + s \geq 1$; hence ${}_c w = \gamma_r^{-1}\gamma_{r-1}^{-1}\cdots\gamma_2^{-1}$, $w_c = \beta_2\cdots\beta_{s-1}\beta_s$, and $M(w)/\mathrm{Soc}(M(w)) = M({}_c w) \oplus M(w_c)$ which yields two irreducible morphisms $M(w) \to M({}_c w)$ and $M(w) \to M(w_c)$. Moreover, ${}_c w_c = {}_c(w_c) = ({}_c w)_c$ is the empty string, hence $M({}_c w_c)$ is the zero module. Thus one might say that the case where M(w) is injective is contained in the case (4) of Theorem 3.1:

(3) The number of indecomposable summands in the middle term of an AR-sequence between string modules is at most two.

Irreducible morphisms in $J(Q, W)$. We now fix a triangulation $\Gamma = \{\tau_1, \tau_2, \ldots \tau_n, \tau_{n+1}, \ldots, \tau_{n+m}\}$ of the marked surface (S, M), where $\tau_1, \ldots, \tau_n$ are internal arcs and $\tau_{n+1}, \ldots, \tau_{n+m}$ are boundary arcs. The aim of this subsection is to describe the AR-quiver of the string algebra $J(Q, W)$ in terms of objects in $\mathscr{C}_{(S,M)}$, which we identified with noncontractible curves and noncontractible closed curves in (S, M). A curve γ in (S, M) defines a string $w(\gamma)$ in $J(Q, W)$ (which is empty if γ is contractible or homotopic to an arc in Γ), and consequently a string module $M(w(\gamma))$ in mod $J(Q, W)$ (which is zero if γ is contractible or homotopic to an arc in Γ). We use the notation $M(\gamma)$ for $M(w(\gamma))$ and from the discussion in Section 2 we know that the string module $M(\gamma)$ is the image of γ under the projection functor $\mathrm{Ext}^1_{\mathscr{C}_\Gamma}(T_\Gamma, -) : \mathscr{C}_{(S,M)} \to \mathrm{mod}\, J(Q, W)$.

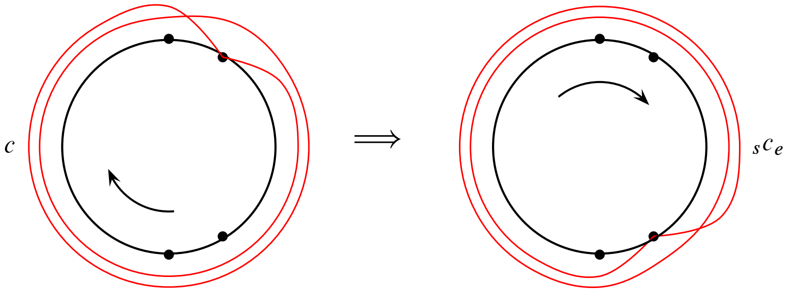
To define elementary moves on curves in (S, M) we use the fact that the orientation of S induces an orientation on each boundary component of S: For any curve

γ in (S, M) we denote by ${}_s\gamma$ the *pivot elementary move* of γ on its starting point, meaning that the curve ${}_s\gamma$ is obtained from γ by moving the starting point $s(\gamma)$ clockwise to the next marked point b on the same boundary (note that $b = s(\gamma)$ if there is only one marked point lying on the same boundary). Similarly, we denote by γ_e the pivot elementary move of γ on its ending point. (See figure.)



Iterated pivot elementary moves are denoted ${}_s\gamma_e = {}_s(\gamma_e) = ({}_s\gamma)_e$, ${}_{s^2}\gamma = {}_s({}_s\gamma)$ and $\gamma_{e^2} = (\gamma_e)_e$, respectively.

Remark 3.3. If $c = d^n$ is a noncontractible closed curve, where $n \geq 1$ and d is a simple closed curve, we view c as a closed curve in (S, M) with endpoints on the boundary as in Remark 2.5. By the definition of pivot elementary moves we obtain that ${}_sc_e$ is again a closed curve whose endpoints are moved to the next point on the boundary; thus ${}_sc_e = d^n = c$ as elements in $\Pi_1^*(S, M)$. See the picture for an example.

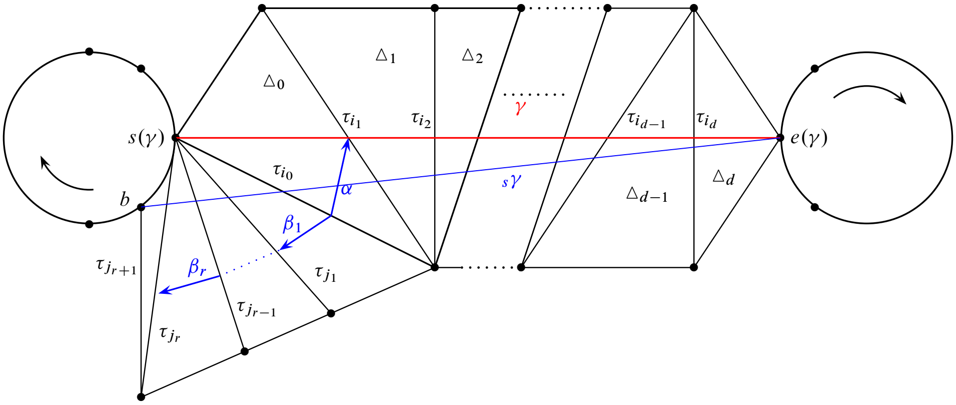


Lemma 3.4. *Let Γ be a triangulation of a marked surface (S, M), and let γ be a curve in (S, M) such that the string $w(\gamma)$ of $J(Q, W)$ is nonempty.*

(1) *If $w(\gamma)$ does not start on a peak, then $w({}_s\gamma)$ is obtained by adding a hook on $s(w(\gamma))$. Thus $w({}_s\gamma) = w(\gamma)_h$ and there is an irreducible morphism in* $\operatorname{mod} J(Q, W) : M(\gamma) \longrightarrow M({}_s\gamma)$.

(2) *If $w(\gamma)$ does not end on a peak, then $w(\gamma_e)$ is obtained by adding a hook on $e(w(\gamma))$. Thus $w(\gamma_e) = {}_hw(\gamma)$ and there is an irreducible morphism in* $\operatorname{mod} J(Q, W) : M(\gamma) \longrightarrow M(\gamma_e)$.

Proof. We only prove (1), since (2) is obtained dually.



As shown in the figure above, we denote by $\tau_{i_1}, \tau_{i_2}, \dots, \tau_{i_d}$ the internal arcs of Γ that intersect γ. Let b be the marked point lying clockwise next to $s(\gamma)$ on the same boundary component. Since γ does not start on a peak, there exits an arrow $\alpha : \tau_{i_0} \to \tau_{i_1}$ in Q such that $w(\gamma)\alpha$ is a string in $J(Q, W)$, for some internal arc $\tau_{i_0} \in \Gamma$. Let $\tau_{j_1}, \tau_{j_2}, \dots, \tau_{j_r}$ be all internal arcs in Γ which intersect τ_{i_0} in the vertex $s(\gamma)$ and which are successors of τ_{i_0} with respect to clockwise orientation at the common vertex $s(\gamma)$. We denote by $\beta_1, \beta_2, \dots, \beta_r$ the arrows of Q induced by the internal arcs $\tau_{i_0}, \tau_{j_1}, \dots, \tau_{j_r}$. Then

$$w(_s\gamma) = w(\gamma)\alpha\beta_1^{-1}\beta_2^{-1}\cdots\beta_r^{-1} = w(\gamma)\alpha V_\alpha,$$

where

$$V_\alpha = \beta_1^{-1}\beta_2^{-1}\cdots\beta_r^{-1}.$$

This means that $w(_s\gamma)$ is obtained from $w(\gamma)$ by adding a hook on $s(w(\gamma))$. By Theorem 3.1 there is an irreducible morphism $M(\gamma) \to M(_s\gamma)$ in mod $J(Q, W)$. □

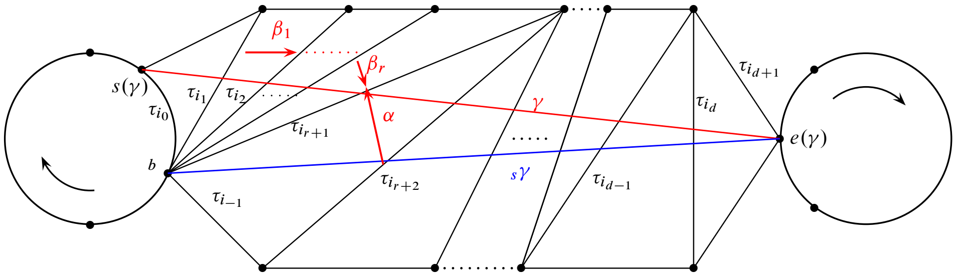
Lemma 3.5. *Let Γ be a triangulation of a marked surface (S, M), and let γ be a curve in (S, M) such that the string $w(\gamma)$ of $J(Q, W)$ is nonempty.*

(1) *If $w(\gamma)$ starts on a peak, then $w(_s\gamma)$ is obtained by deleting a cohook on $s(w(\gamma))$. Thus $w(_s\gamma) = w(\gamma)_c$ and if $w(\gamma)_c$ is nonempty there is an irreducible morphism in* mod $J(Q, W) : M(\gamma) \longrightarrow M(_s\gamma)$.

(2) *If $w(\gamma)$ ends on a peak, then $w(\gamma_e)$ is obtained by deleting a cohook on $e(w(\gamma))$. Thus $w(\gamma_e) = {_c}w(\gamma)$ and if ${_c}w(\gamma)$ is nonempty there is an irreducible morphism in* mod $J(Q, W) : M(\gamma) \longrightarrow M(\gamma_e)$.

Proof. As before we only prove part (1) of the lemma.

We denote by b the marked point lying clockwise next to $s(\gamma)$ on the same boundary. We further denote by $\tau_{i_1}, \tau_{i_2}, \dots, \tau_{i_d}$ the internal arcs that intersect γ



in the order indicated in the figure above. Since $w(\gamma)$ starts on a peak, there is an $r \geq 1$ such that the arcs $\tau_{i_1}, \tau_{i_2}, \ldots, \tau_{i_{r+1}}$ intersect in the vertex b and induce arrows $\beta_1, \beta_2, \ldots, \beta_r$ in Q_1 as shown in the figure. We choose r to be maximal and distinguish two cases:

(i) If $r + 1 = d$ then $w(\gamma)$ is a direct string and $w({}_s\gamma) = w(\gamma)_c$ is the empty string.

(ii) If $r+1 < d$ then by maximality of r there is an arrow α in Q_1 from $\tau_{i_{r+2}}$ to $\tau_{i_{r+1}}$. Since γ starts on a deep, τ_{i_0} is a boundary arc which implies $U_\alpha^{-1} = \beta_r \beta_{r-1} \cdots \beta_1$. Moreover, we know $w(\gamma) = w_0 \alpha^{-1} \beta_r \beta_{r-1} \cdots \beta_1 = w_0 \alpha^{-1} U_\alpha^{-1}$. Thus $w({}_s\gamma) = w_0$ which means that $w({}_s\gamma)$ is obtained by deleting a cohook on $s(w(\gamma))$. If $w({}_s\gamma)$ is nonempty, we obtain by Theorem 3.1 that there is an irreducible morphism in $\operatorname{mod} J(Q, W) : M(\gamma) \longrightarrow M({}_s\gamma)$. □

Theorem 3.6. *Let Γ be a triangulation of a marked surface (S, M), and let γ be a curve in (S, M). Then each irreducible morphism in* mod $J(Q, W)$ *starting in $M(\gamma)$ is obtained by pivot elementary moves on endpoints of γ. Moreover, all AR-sequences between string modules in* mod $J(Q, W)$ *are of the form*

$$0 \longrightarrow M(\gamma) \longrightarrow M({}_s\gamma) \oplus M(\gamma_e) \longrightarrow M({}_s\gamma_e) \longrightarrow 0$$

for some curve γ in (S, M).

Proof. The irreducible morphisms and AR-sequences between string modules are described in Theorem 3.1. Among the four cases listed there, we only consider the second case here, the others being similar: Suppose $w(\gamma)$ does not start but ends on a peak. Then we get from Lemma 3.4 and Lemma 3.5 two irreducible morphisms, $M(\gamma) \to M({}_s\gamma)$ and $M(\gamma) \to M(\gamma_e)$. Moreover, the construction of ${}_s\gamma$ does not change any information of γ on $e(\gamma)$, that is $w({}_s\gamma)$ also ends on a peak if it is nonempty. By Lemma 3.5 again we get an irreducible morphism $M({}_s\gamma) \to M({}_s\gamma_e)$ in case $w({}_s\gamma_e)$ is nonempty, and similarly, there is an irreducible morphism $M(\gamma_e) \to M({}_s\gamma_e)$ induced by γ_e. This completes the proof since the number of summands in the middle term of an AR-sequence between string modules is at most two. □

To complete the description of irreducible morphisms in $\operatorname{mod} J(Q, W)$, we recall from [Butler and Ringel 1987] that each band in $J(Q, W)$ yields a k^*-family of homogeneous tubes in the AR-quiver of $J(Q, W)$, given by an embedding $\operatorname{mod} k[t, t^{-1}] \to \operatorname{mod} J(Q, W)$.

Corollary 3.7. *The AR-translation in* $\operatorname{mod} J(Q, W)$ *is given by simultaneous counterclockwise rotation of starting and ending point of a curve to the next marked points at the boundary, that is*

$$\tau^{-1}(M(\gamma)) = M({}_s\gamma_e)$$

if $M(\gamma)$ is not an injective $J(Q, W)$-module.

Proof. For string modules this is shown in Theorem 3.6 above. For any band b in $J(Q, W)$ we have $b = {}_sb_e$ as elements in $\Pi_1(S, M)$. Moreover, the corresponding band modules $M(b, n, \phi)$ lie on homogeneous tubes and thus satisfy $\tau^{-1}(M(b, n, \phi)) = M(b, n, \phi)$. □

Example 3.8. We reconsider the curve γ from Example 2.6. From the diagram defining Γ it is easy to obtain the following descriptions:

$$M_\gamma = \begin{array}{cccccc} & & & 1 & & \\ 3 & & 2 & & 3 & \\ & 4 & & & & 4 \end{array} \qquad M({}_s\gamma) = \begin{array}{ccccccccc} & & 1 & & & & 1 & & \\ & 2 & & 3 & & 2 & & 3 & \\ 4 & & & & 4 & & & & 4 \end{array}$$

$$M(\gamma_e) = \begin{array}{ccccccc} & & & 1 & & & \\ 3 & & 2 & & 3 & & 2 \\ & 4 & & & & 4 & \end{array} \qquad M({}_s\gamma_e) = \begin{array}{cccccccccc} & & 1 & & & & 1 & & & \\ & 2 & & 3 & & 2 & & 3 & & 2 \\ 4 & & & & 4 & & & & 4 & \end{array}$$

Hence, we have an AR-sequence:

$$0 \longrightarrow M(\gamma) \longrightarrow M({}_s\gamma) \oplus M(\gamma_e) \longrightarrow M({}_s\gamma_e) \longrightarrow 0.$$

Remark 3.9. If $M(\gamma)$ is an injective $J(Q, W)$-module, it follows from the description in Remark 3.2(2) that ${}_s\gamma_e$ is an internal arc in Γ which implies $M({}_s\gamma_e) = 0$. Dually, if $\gamma \in \Gamma$ is an internal arc, $M({}_s\gamma_e)$ is a projective module in $\operatorname{mod} J(Q, W)$.

Irreducible morphisms in $\mathscr{C}_{(S,M)}$. Recall that $\mathscr{C} := \mathscr{C}_\Gamma$ is a triangulated Hom-finite k-category which is 2-Calabi–Yau. We know that $\mathscr{C}$ has a cluster-tilting object $T_\Gamma = \tau_1 \oplus \tau_2 \oplus \cdots \oplus \tau_n$, where $\tau_1, \ldots \tau_n$ are the internal arcs in Γ. Denote by [1] the suspension functor of the triangulated category $\mathscr{C}$. Moreover, since $\mathscr{C}$ is 2-Calabi–Yau, $\mathscr{C}$ has AR-triangles and the AR-translation is given by $\tau = [1]$ (see [Reiten and Van den Bergh 2002]).

Curves in (S, M) which are noncontractible and not homotopic to a boundary arc give a parametrization of the isoclasses of string objects in $\mathscr{C}$, we refer to them as non-boundary curves. We identify contractible curves and boundary arcs with the zero object in $\mathscr{C}$. We further say that a curve γ is not in Γ and write $\gamma \notin \Gamma$ if it is noncontractible and not homotopic to an arc in the triangulation Γ. Since the non-boundary arcs in Γ yield the cluster-tilting object T_Γ, the string module $M(\gamma)$ is nonzero if $\gamma \notin \Gamma$. The following lemma will be used frequently:

Lemma 3.10 [Keller and Reiten 2007; Koenig and Zhu 2008]. *Let γ be a curve in (S, M) such that $\gamma \notin \Gamma$.*

(1) *$M(\gamma)$ is projective in* $\operatorname{mod} J(Q, W)$ *if and only if $\gamma[1] \in \Gamma$.*

(2) *$M(\gamma)$ is injective in* $\operatorname{mod} J(Q, W)$ *if and only if $\gamma[-1] \in \Gamma$.*

From [Koenig and Zhu 2008], we know that any sink (or source) map in $\mathscr{C}$ is again a sink (or source) map in $\operatorname{mod} J(Q, W)$, therefore each AR-triangle in $\mathscr{C}$

$$\gamma' \longrightarrow \bigoplus \gamma_i \longrightarrow \gamma'' \longrightarrow \gamma'[1]$$

with $\gamma' \notin \Gamma$ and $\gamma'[-1] \notin \Gamma$ yields an AR-sequence in $\operatorname{mod} J(Q, W)$:

$$0 \longrightarrow M(\gamma') \longrightarrow \bigoplus M(\gamma_i) \longrightarrow M(\gamma'') \longrightarrow 0.$$

Moreover, all AR-sequences in $\operatorname{mod} J(Q, W)$ are obtained in this way and the AR-translation in $\operatorname{mod} J(Q, W)$ is induced by that in $\mathscr{C}$. Combining this with Corollary 3.7 and Remark 3.9, we can easily get this:

Proposition 3.11. *The AR-translation in $\mathscr{C}_{(S,M)}$ is given by simultaneous counterclockwise rotation of starting and ending point of a curve to the next marked points at the boundary. That is $\gamma[-1] = {}_s\gamma_e$ if γ is a string object and $(\lambda, b^n)[-1] = (\lambda, {}_sb^n_e)$ if (λ, b^n) is a band object in $\mathscr{C}_{(S,M)}$.*

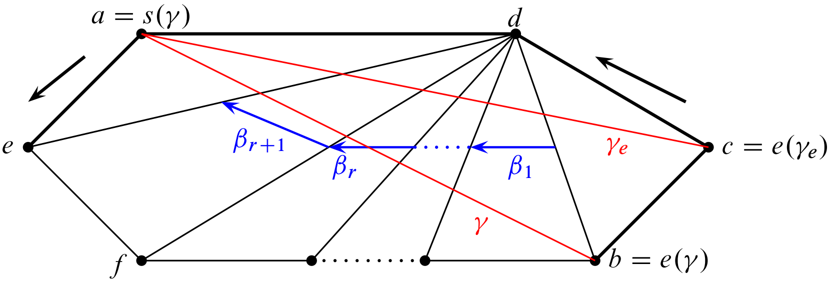
Proof. Assume γ is a string object. If γ is a non-boundary curve in (S, M) such that $M(\gamma)$ is not injective, then $M(\gamma[-1]) = M(\tau^{-1}(\gamma)) = \tau^{-1}(M(\gamma)) = M({}_s\gamma_e)$, which implies $\gamma[-1] = {}_s\gamma_e$. If $M(\gamma)$ is injective, then Remark 3.9 shows that ${}_s\gamma_e \in \Gamma$ is an internal arc, and by Lemma 3.10 and the discussion in Section 2 we obtain $\gamma[-1] = {}_s\gamma_e$. Similarly, if $\gamma \in \Gamma$ is an internal arc, then $\gamma[-1] = {}_s\gamma_e$.

Assume (λ, b^n) is a band object, where $\lambda \in k^*$, $b^n \in \Pi_1^*(S, M)$ with $n \geq 1$ and b a noncontractible closed curve. Then $M((\lambda, b^n)[-1]) = M(\tau^{-1}(\lambda, b^n)) = \tau^{-1}(M((\lambda, b^n))) = \tau^{-1}(M(w(b), n, \phi_\lambda)) = M(w(b), n, \phi_\lambda) = M(w({}_sb_e), n, \phi_\lambda) = M((\lambda, {}_sb^n_e))$, where $\phi_\lambda \in \operatorname{Aut}(k^n)$ is induced by λ. Hence $(\lambda, b^n)[-1] = (\lambda, {}_sb^n_e)$ as a band object in $\mathscr{C}_{(S,M)}$. □

Lemma 3.12. *Let $\gamma \notin \Gamma$ be a curve in (S, M) such that $M(\gamma_e)$ is a nonzero projective-injective module in* $\operatorname{mod} J(Q,W)$. *There is a source map in* $\operatorname{mod} J(Q,W)$:

$$M({}_s\gamma_e) \longrightarrow M({}_{s^2}\gamma_e).$$

Proof. Assume $M(\gamma_e)$ is a projective-injective module. Then $w(\gamma_e)$ must be an inverse (or a direct) string which both starts and ends on a peak and also both starts and ends in a deep. Without loss of generality, assume $w(\gamma_e) = \beta_1^{-1} \cdots \beta_r^{-1} \beta_{r+1}^{-1}$, where $r \geq 0$.



Let $a = s(\gamma) = s(\gamma_e)$, $b = e(\gamma)$, $c = e(\gamma_e)$, and let d, e, f be marked points as in the figure above. Since $w(\gamma_e)$ both starts and ends in a deep, bc and da are boundary arcs. Similarly, ae and dc are boundary arcs since γ_e both starts and ends on a peak, therefore b, c, d, a, e lie consecutively on the same boundary.

(1) If $M({}_s\gamma_e)$ is an injective module, that is $w({}_s\gamma_e) = \beta_1^{-1} \cdots \beta_r^{-1}$ both starts and ends on a peak, then ef is a boundary arc. Then the definition of pivot elementary move implies that

$$M({}_{s^2}\gamma_e) = M(\beta_1^{-1} \cdots \beta_{r-1}^{-1}) = M({}_s\gamma_e)/\operatorname{Soc}(M({}_s\gamma_e))$$

and ${}_s\gamma_{e^2} = de \in \Gamma$ is an internal arc. Hence by Remark 3.2(2) there is a source map: $M({}_s\gamma_e) \to M({}_{s^2}\gamma_e)$.

(2) If $M({}_s\gamma_e)$ is not an injective module, that is $w({}_s\gamma_e) = \beta_1^{-1} \cdots \beta_r^{-1}$ does not start on a peak, then there exists $\alpha \in Q_1$ such that $w({}_s\gamma_e)\alpha$ is a string. By Theorem 3.6, there is an AR-sequence:

$$0 \longrightarrow M({}_s\gamma_e) \longrightarrow M({}_{s^2}\gamma_e) \oplus M({}_s\gamma_{e^2}) \longrightarrow M({}_{s^2}\gamma_{e^2}) \longrightarrow 0,$$

where $w({}_s\gamma_e) = \beta_1^{-1} \cdots \beta_r^{-1} = U_\alpha$, $w({}_{s^2}\gamma_e) = w({}_s\gamma_e)_h = N_\alpha$ by Lemma 3.4, and ${}_s\gamma_{e^2} = de \in \Gamma$ is an internal arc in Γ. Thus there is a source map: $M({}_s\gamma_e) \to M({}_{s^2}\gamma_e)$ in mod $J(Q, W)$. □

Proposition 3.13. *Let γ be a non-boundary curve and*

$$0 \longrightarrow M(\gamma) \longrightarrow M({}_s\gamma) \oplus M(\gamma_e) \longrightarrow M({}_s\gamma_e) \longrightarrow 0 \qquad (3\text{-}1)$$

an AR-sequence in mod $J(Q, W)$ *with $M({}_s\gamma) \neq 0 \neq M(\gamma_e)$. Then there is an AR-triangle with two middle terms in $\mathscr{C}$ as follows*:

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1].$$

Proof. Since AR-sequences in $\operatorname{mod} J(Q, W)$ are induced by AR-triangles in $\mathscr{C}$, we assume that

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \oplus \sigma \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1] \tag{3-2}$$

is an AR-triangle in $\mathscr{C}$ with $M(\sigma) = 0$. It suffices to prove that $\sigma = 0$.

Otherwise, σ contains a direct summand ρ which is an internal arc in Γ. Then ${}_s\gamma[1] \in \Gamma$ or $\gamma_e[1] \in \Gamma$ since $M(\rho[1]) \neq 0$ in $\operatorname{mod} J(Q, W)$ by Lemma 3.10 and since the number of summands in the middle term of an AR-sequence in $J(Q, W)$ is at most two. Similarly, we conclude ${}_s\gamma[-1] \in \Gamma$ or $\gamma_e[-1] \in \Gamma$. Hence there are four cases; we consider only two of them, since the others are similar.

(1) Assume ${}_s\gamma[1] \in \Gamma$ and $\gamma_e[-1] \in \Gamma$. Then Lemma 3.10 implies that $M({}_s\gamma)$ is projective and $M(\gamma_e)$ is injective. Hence $M(\gamma)$ must be a direct summand of $\operatorname{rad} M({}_s\gamma)$ and $M({}_s\gamma_e)$ must be a direct summand of $M(\gamma_e)/\operatorname{Soc} M(\gamma_e)$. This yields a contradiction by comparing the dimensions in (3-1).

(2) Assume $\gamma_e[1] \in \Gamma$ and $\gamma_e[-1] \in \Gamma$. Then Lemma 3.10 implies that $M(\gamma_e)$ is a projective-injective module, and so by Lemma 3.12 there is a source map

$$M({}_s\gamma_e) \longrightarrow M({}_{s^2}\gamma_e) \tag{3-3}$$

in $\operatorname{mod} J(Q, W)$. But after shifting (3-2) by $[-1]$, there is an AR-sequence in $\operatorname{mod} J(Q, W)$:

$$0 \longrightarrow M({}_s\gamma_e) \longrightarrow M({}_{s^2}\gamma_e) \oplus M({}_s\rho_e) \longrightarrow M({}_{s^2}\gamma_{e^2}) \longrightarrow 0.$$

with $M({}_s\rho_e) = M(\rho[-1]) \neq 0$ a projective module. This contradicts the fact that the source map (3-3) ends with just one indecomposable module. □

Since neither injective modules nor projective modules occur in homogeneous tubes, the following corollary related to band objects can be obtained similarly.

Corollary 3.14. *Let (λ, b^n) be a band object in $\mathscr{C}$, where $\lambda \in k^*$, $b^n \in \Pi_1^*(S, M)$ with $n \geq 1$ and b itself is not a proper power of an element in $\Pi_1^*(S, M)$. Then there is an AR-triangle in $\mathscr{C}$:*

$$(\lambda, b^n) \longrightarrow (\lambda, b^{n+1}) \oplus (\lambda, b^{n-1}) \longrightarrow (\lambda, b^n) \longrightarrow (\lambda, b^n)[1],$$

where (λ, b^0) is the zero object in $\mathscr{C}$.

Therefore, the band objects are closed under irreducible morphisms and the corresponding AR-components in $\mathscr{C}$ are homogeneous tubes. It remains to consider the string objects in $\mathscr{C}$.

Lemma 3.15. *If γ is an internal arc in Γ, the AR-triangle in $\mathscr{C}$ starting in γ is of the form*

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1].$$

Proof. Suppose γ is an internal arc in Γ. Then $M({}_s\gamma_e)$ is projective in $\operatorname{mod} J(Q,W)$ by Lemma 3.10. We consider $\gamma[-1] = {}_s\gamma_e$.

(1) If $M({}_s\gamma_e)$ is not injective, then Theorem 3.6 induces an AR-sequence in $\operatorname{mod} J(Q, W)$:

$$0 \longrightarrow M({}_s\gamma_e) \longrightarrow M({}_{s^2}\gamma_e) \oplus M({}_s\gamma_{e^2}) \longrightarrow M({}_{s^2}\gamma_{e^2}) \longrightarrow 0. \tag{3-4}$$

The definition of a cluster-tilting object guarantees that ${}_{s^2}\gamma_e$ and ${}_s\gamma_{e^2}$ cannot be internal arcs in Γ. If neither ${}_{s^2}\gamma_e$ nor ${}_s\gamma_{e^2}$ is boundary arc (thus $M({}_{s^2}\gamma_e) \neq 0 \neq M({}_s\gamma_{e^2})$), then Proposition 3.13 induces an AR-triangle

$${}_s\gamma_e \longrightarrow {}_{s^2}\gamma_e \oplus {}_s\gamma_{e^2} \longrightarrow {}_{s^2}\gamma_{e^2} \longrightarrow {}_s\gamma_e[1]$$

which yields the AR-triangle

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1].$$

Assume now that one of ${}_{s^2}\gamma_e$ and ${}_s\gamma_{e^2}$ is a boundary arc. Without loss of generality, let ${}_{s^2}\gamma_e$ be a boundary arc, then by definition of a cluster-tilting object, $\operatorname{Hom}_{\mathscr{C}}({}_s\gamma_e, \sigma) = 0$ for any $\sigma \in \Gamma$, hence (3-4) induces an AR-triangle of the form

$${}_s\gamma_e \longrightarrow {}_{s^2}\gamma_e \oplus {}_s\gamma_{e^2} \longrightarrow {}_{s^2}\gamma_{e^2} \longrightarrow {}_s\gamma_e[1]$$

with ${}_{s^2}\gamma_e = 0$. The definition of a pivot elementary move implies that γ_e is also a boundary arc, hence there is an AR-triangle of the form

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1]$$

where $\gamma_e = 0$ is a boundary arc in $\mathscr{C}$.

(2) If $M({}_s\gamma_e) = M(\gamma[-1])$ is injective, then $M({}_s\gamma_e)$ is a projective-injective module by Lemma 3.10. By definition of a cluster-tilting object and Remark 3.2(3) we can assume that the AR-triangle starting in γ is of the form

$$\gamma \longrightarrow \delta_1 \oplus \delta_2 \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1]$$

where neither δ_1 nor δ_2 are internal arcs in Γ. Since $M({}_s\gamma_e)$ is a projective-injective module in $\operatorname{mod} J(Q, W)$, one of δ_1 and δ_2 must be a boundary arc. Without loss of generality, assume δ_2 is a boundary arc; then Lemma 3.12 implies

$$M(\delta_1) = \operatorname{rad} M({}_s\gamma_e) = M({}_s\gamma);$$

that is, $\delta_1 = {}_s\gamma$ and $\delta_2 = \gamma_e$ is a boundary arc. Therefore the AR-triangle is of the form

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1]$$

where $\gamma_e = 0$ is a boundary arc. □

As shown in Remark 3.2(3), the middle terms of AR-sequences in mod $J(Q, W)$ contain at most two indecomposable summands. The following proposition establishes the same result for AR-triangles in $\mathscr{C}$.

Proposition 3.16. *The number of indecomposable summands in the middle term of an AR-triangle in $\mathscr{C}_{(S,M)}$ is at most two.*

Proof. We only need to consider string objects in $\mathscr{C}$. Let therefore γ be a non-boundary curve, and let

$$\gamma \longrightarrow \bigoplus_{i=1}^{r} \delta_i \oplus \bigoplus_{j=1}^{s} \tau_j \longrightarrow \gamma[-1] \longrightarrow \gamma[1] \tag{3-5}$$

be an AR-triangle in $\mathscr{C}$ starting in γ with $\delta_i \notin \Gamma$ and where τ_j are internal arcs in Γ. Remark 3.2(3) implies $r \leq 2$ and $s \leq 2$. If $\gamma \in \Gamma$ or $\gamma[-1] \in \Gamma$, then by Lemma 3.15, there is nothing to prove. We suppose now $\gamma \notin \Gamma$ and $\gamma[-1] \notin \Gamma$. If $r = 2$, Proposition 3.13 implies $s = 0$. If $r = 1$ and $s = 2$, then (3-5) induces an AR-sequence with one middle term in mod $J(Q, W)$:

$$0 \longrightarrow M(\gamma) \longrightarrow M(\delta_1) \longrightarrow M(\gamma[-1]) \longrightarrow 0,$$

hence there exists $\alpha \in Q_1$ such that $w(\delta_1) = N_\alpha$ by Remark 3.2(1). On the other hand, if we shift (3-5) by [1] and [−1], then Remark 3.2(3) implies that $\delta_1[1]$ and $\delta_1[-1]$ must be the internal arcs in Γ, since

$$M(\tau_1[1]) \neq 0 \neq M(\tau_2[1]) \quad \text{and} \quad M(\tau_1[-1]) \neq 0 \neq M(\tau_2[-1])$$

in mod $J(Q, W)$ by Lemma 3.10. Thus $M(\delta_1) = M(N_\alpha)$ is a projective-injective module in mod $J(Q, W)$. The definition of N_α implies that the quiver Q is of type A_2. But this is impossible, since the cluster category of type A_2 is well-known, and there is no AR-triangle with more than two summands in the middle term. □

Now we consider the AR-sequences with just one middle term in mod $J(Q, W)$.

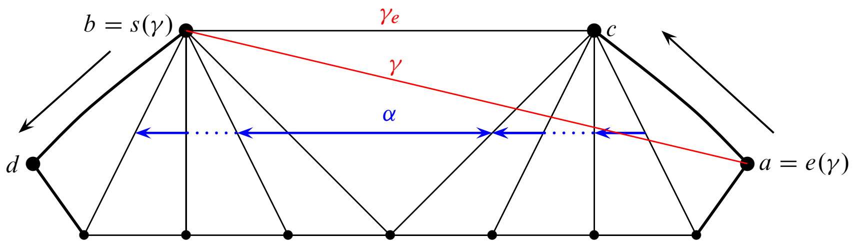
Lemma 3.17. *Let γ be a non-boundary curve in (S, M) such that $w(\gamma) = U_\alpha$, for some $\alpha \in Q$. Then $w({}_s\gamma) = N_\alpha$ and $w(\gamma_e) = 0$ is a zero string, and the AR-triangle in $\mathscr{C}$ starting in γ is of the form*

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1].$$

Proof. Since $w(\gamma) = U_\alpha$ does not start but ends on a peak, we know by Lemma 3.4 that $w({}_s\gamma) = w(\gamma)_h = N_\alpha$ and by Lemma 3.5 that

$$w(\gamma_e) = {}_c w(\gamma) = {}_c U_\alpha = 0$$

is a zero string. Hence γ_e might be an internal arc in Γ or a boundary arc in (S, M), as in the figure:



Let $b = s(\gamma)$ and $a = e(\gamma)$ be the endpoints of γ and let d be the marked point lying clockwise next to b on the same boundary. Since U_α ends on a peak, ac is a boundary arc. The definition of pivot elementary move implies $c = e(\gamma_e)$, hence there is an AR-sequence with one middle term in mod $J(Q, W)$ (see Remark 3.2(1)):

$$0 \longrightarrow M(\gamma) \longrightarrow M({}_s\gamma) \longrightarrow M({}_s\gamma_e) \longrightarrow 0. \tag{3-6}$$

(1) If $\gamma_e \in \Gamma$, Lemma 3.10 implies that $M(\gamma_e[1])$ is an injective module. Remark 3.2(2) induces an irreducible morphism $M(\gamma_e[1]) \to M(\gamma)$ in mod $J(Q, W)$; hence there are two irreducible morphisms in $\mathscr{C}$:

$$\gamma_e[1] \to \gamma \quad \text{and} \quad \gamma \to \gamma_e[1][-1] = \gamma_e.$$

Combining this with (3-6) and Proposition 3.16, we get the AR-triangle in $\mathscr{C}$:

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1].$$

(2) If γ_e is a boundary arc, then $w(\gamma)$ starts in a deep and ends on a peak which implies $M(\gamma) = M(U_\alpha)$ is projective in mod $J(Q, W)$; thus $\gamma[1] \in \Gamma$ by Lemma 3.10. By (3-6) and the definition of a cluster-tilting object, the AR-triangle starting in γ is of the form

$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1]$$

where $\gamma_e = 0$ in $\mathscr{C}$. □

By the following theorem we finally obtain the results formulated in Theorem 1.3 in the introduction:

Theorem 3.18. *Let (S, M) be a marked surface without punctures, and let γ be a non-boundary curve in (S, M). Then each irreducible morphism in $\mathscr{C}_{(S,M)}$ starting in γ is obtained by pivot elementary moves on endpoints of γ. The AR-triangle starting in γ is of the form*

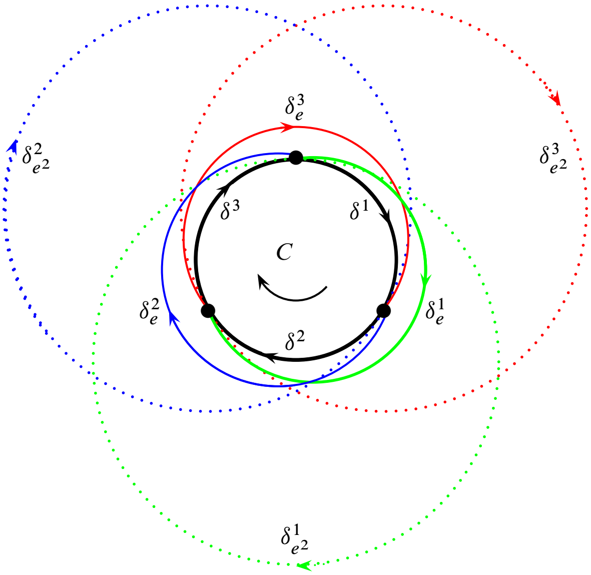
$$\gamma \longrightarrow {}_s\gamma \oplus \gamma_e \longrightarrow {}_s\gamma_e \longrightarrow \gamma[1].$$

Proof. Lemma 3.15 implies the case when γ or $\gamma[-1] = {}_s\gamma_e$ is an internal arc in Γ. Proposition 3.13 and Lemma 3.17 yield the remaining case. □

AR-components in $\mathscr{C}_{(S,M)}$. The aim of this subsection is to describe the AR-components in $\mathscr{C}_{(S,M)}$. From Corollary 3.14 we know that all band objects of $\mathscr{C}_{(S,M)}$ lie in homogeneous tubes, so we focus on string objects from now on. It follows from Theorem 3.18 that each string object γ in $\mathscr{C}_{(S,M)}$ is starting point of a mesh with two middle terms in the AR-quiver of $\mathscr{C}_{(S,M)}$ except when one of ${}_s\gamma$ or γ_e is a boundary arc. The situation where one of ${}_s\gamma$ or γ_e is a boundary arc is explicitly described in the following corollary.

Corollary 3.19. *Let C be a boundary component of S with t marked points. If we choose a numbering and an orientation of the boundary arcs $\delta^1, \dots, \delta^t$ such that all δ^i are clockwise oriented and $e(\delta^i) = s(\delta^{i+1})$ for all i, then the objects $\delta^i_{e^j}$ with $i = 1, \dots, t$ and $j \geq 1$ form a tube of rank t in $\mathscr{C}_{(S,M)}$. Moreover, there is a bijection between the boundary components of S and the tubes in $\mathscr{C}_{(S,M)}$ which are not formed by band objects.*

Proof. The proof follows easily from the description of the AR-triangles given in Theorem 3.18 once the orientation of the boundary arcs is chosen. We illustrate the situation where $t = 3$:



Note that $\delta^i_{e^{t-1}}$ is a noncontractible closed curve for all surfaces except a disc, where all closed curves are contractible. Thus, for a disc the objects $\delta^i_{e^{t-1}}$ are zero, and the remaining objects $\delta^i_{e^j}$ with $j > t-1$ have to be identified accordingly, see the description in [Caldero et al. 2006]. □

The preceding corollary yields a description of all AR-components (formed by string objects) which contain some meshes with only one middle term. The remaining AR-components in $\mathscr{C}_{(S,M)}$ are formed by meshes with exactly two middle terms. If Λ is one such component, one can choose one object γ in Λ and then the component is formed by all ${}_{s^i}\gamma_{e^j}$ with $i, j \geq 1$. In case there are no identifications of the objects, one obtains thus components of the form $\mathbb{Z}\mathbb{A}_\infty^\infty$. If S is neither a disc nor an annulus, then the Jacobian algebra associated to a triangulation will in general be of nonpolynomial growth (see [Assem et al. 2010]), and there are plenty of components of the form $\mathbb{Z}\mathbb{A}_\infty^\infty$ in the AR-quiver of $\mathscr{C}_{(S,M)}$. If S is an annulus, the cluster category $\mathscr{C}_{(S,M)}$ is of type $\widetilde{\mathbb{A}}$ and there is no component of the form $\mathbb{Z}\mathbb{A}_\infty^\infty$.

4. Flips and mutations

In the previous sections we fixed one triangulation Γ of (S, M) and studied the irreducible morphisms and AR-components of the cluster category $\mathscr{C}_\Gamma$ defined by the quiver with potential (Q_Γ, W_Γ). The aim of this section is to study the effect of a change in the triangulation on the cluster category $\mathscr{C}_\Gamma$.

If τ_i is an internal arc in Γ, then there exists exactly one internal arc $\tau_i' \neq \tau_i$ in (S, M) such that $f_{\tau_i}(\Gamma) := (\Gamma \backslash \{\tau_i\}) \cup \{\tau_i'\}$ is also a triangulation of (S, M). In fact, the internal arc τ_i is a diagonal in the quadrilateral formed by the two triangles of Γ containing τ_i, and τ_i' is the other diagonal in that quadrilateral, see [Fomin et al. 2008]. We denote τ_i' by $f_\Gamma(\tau_i)$ and say that $f_{\tau_i}(\Gamma)$ is obtained from Γ by applying a flip along τ_i. In fact, by applying iterated flips one can obtain all triangulations of (S, M):

Theorem 4.1 [Fomin et al. 2008]. *For any two triangulations of (S, M) there is a sequence of flips which transforms one triangulation into the other.*

As shown in Theorem 1.1, we can view the non-boundary curves in (S, M) as objects in $\mathscr{C}_\Gamma$. If we denote all internal arcs of Γ by $\tau_1, \ldots, \tau_n$, their direct sum $T_\Gamma = \tau_1 \oplus \tau_2 \oplus \cdots \oplus \tau_n$ is a cluster-tilting object in $\mathscr{C}_\Gamma$. The following theorem is adapted from its source to our setup.

Theorem 4.2 [Iyama and Yoshino 2008]. *If $\tau_i \in \Gamma$ is an internal arc, there is a curve τ_i^* in (S, M) (unique up to homotopy) which is not homotopic to τ_i such that the object $\mu_{\tau_i}(T_\Gamma)$ obtained from T_Γ by replacing τ_i with τ_i^* is also a cluster-tilting object in $\mathscr{C}_\Gamma$.*

The object $\mu_{\tau_i}(T_\Gamma)$ is called the *mutation* of T_Γ in τ_i, and (τ_i, τ_i^*) is called an *exchange pair* in $\mathscr{C}_\Gamma$. As shown in [Iyama and Yoshino 2008], any exchange pair (τ_i, τ_i^*) induces the following nonsplit triangles (unique up to isomorphism) which are referred to as *exchange triangles*:

$$\tau_i \xrightarrow{f} \tau \longrightarrow \tau_i^* \longrightarrow \tau_i[1] \quad \text{and} \quad \tau_i^* \longrightarrow \tau \xrightarrow{g} \tau_i \longrightarrow \tau_i^*[1]$$

Here f is a minimal left $\operatorname{add}(T_\Gamma \setminus \{\tau_i\})$-approximation and g is a minimal right $\operatorname{add}(T_\Gamma \setminus \{\tau_i\})$-approximation. Since

$$\operatorname{End}(T_\Gamma, T_\Gamma) \cong \operatorname{End}(T_\Gamma[1], T_\Gamma[1]) \cong J(Q_\Gamma, W_\Gamma),$$

we know that the quiver of the endomorphism algebra of T_Γ in $\mathscr{C}_\Gamma$ does not contain loops at any vertex. Hence we obtain the following lemma which is a special case of Lemma 7.5 in [Keller 2010].

Lemma 4.3. *If τ_i is an internal arc of Γ, the exchange triangles are given by*

$$\tau_i \xrightarrow{f} \bigoplus_{\tau_j \to \tau_i} \tau_j \longrightarrow \tau_i^* \longrightarrow \tau_i[1] \quad \textit{and} \quad \tau_i^* \longrightarrow \bigoplus_{\tau_i \to \tau_k} \tau_k \xrightarrow{g} \tau_i \longrightarrow \tau_i^*[1],$$

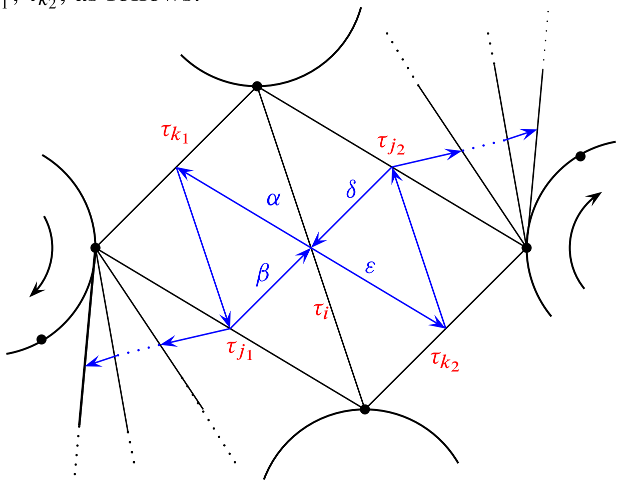
where f is a minimal left $\operatorname{add}(T_\Gamma \setminus \tau_i)$-approximation and g is a minimal right $\operatorname{add}(T_\Gamma \setminus \tau_i)$-approximation.

For an internal arc $\tau_i \in \Gamma$, we discussed the definition of the flip $f_\Gamma(\tau_i)$ above. On the other hand, if we view τ_i as an indecomposable rigid object in $\mathscr{C}_\Gamma$, also the mutation $\mu_\Gamma(\tau_i) := \tau_i^*$ of τ_i is defined. The following theorem shows that flip and mutation of an internal arc are compatible (viewed as objects in the cluster category $\mathscr{C}_\Gamma$):

Theorem 4.4. *If Γ is a triangulation of (S, M) and $\tau_i \in \Gamma$ is an internal arc, then*

$$\mu_\Gamma(\tau_i) = f_\Gamma(\tau_i).$$

Proof. The internal arc τ_i is a diagonal of a quadrilateral formed by the internal arcs $\tau_{j_1}, \tau_{j_2}, \tau_{k_1}, \tau_{k_2}$, as follows:



The triangles of Γ containing τ_i induce arrows $\alpha, \beta, \delta, \varepsilon$ in Q_Γ as indicated above. By Lemma 4.3 there is a nonsplit triangle in $\mathscr{C}_\Gamma$:

$$\tau_i \xrightarrow{f} \tau_{j_1} \oplus \tau_{j_2} \longrightarrow \tau_i^* \longrightarrow \tau_i[1]$$

where f is a minimal left add$(T_\Gamma \setminus \tau_i)$-approximation. We obtain the following right exact sequence in mod $J(Q_\Gamma, W_\Gamma)$ by applying $M(-) \cong \mathrm{Ext}^1_{\mathscr{C}_\Gamma}(T_\Gamma, -)$ to this triangle:

$$M(\tau_i[-1]) \longrightarrow M(\tau_{j_1}[-1]) \oplus M(\tau_{j_2}[-1]) \longrightarrow M(\tau_i^*[-1]) \longrightarrow 0.$$

This sequence is in fact a minimal projective resolution in mod $J(Q_\Gamma, W_\Gamma)$ whose projective modules can be described by strings as follows:

$$\begin{aligned} M(\tau_i[-1]) &= M({V_\alpha}^{-1} V_\varepsilon), \\ M(\tau_{j_1}[-1]) &= M({V_\alpha}^{-1} \beta V_\beta), \\ M(\tau_{j_2}[-1]) &= M({V_\delta}^{-1} \delta^{-1} V_\varepsilon). \end{aligned}$$

Hence $M(\tau_i^*[-1]) \cong M(V_\delta^{-1}\delta^{-1}\beta V_\beta)$, which implies $w({}_s(\tau_i^*)_e) = V_\delta^{-1}\delta^{-1}\beta V_\beta$ and thus $\mu_\Gamma(\tau_i) = \tau_i^* = f_\Gamma(\tau_i)$. □

For any curve γ in (S, M), the definition of the string $w(\gamma)$ depends on the Jacobian algebra $J(Q_\Gamma, W_\Gamma)$. In order to compare string modules in two Jacobian algebras arising from different triangulations of (S, M), we denote in the following by $w(\Gamma, \gamma)$ and $M(\Gamma, \gamma)$ the string and the string module in the Jacobian algebra $J(Q_\Gamma, W_\Gamma)$. Similarly, the band in $J(Q_\Gamma, W_\Gamma)$ given by a noncontractible closed curve b is denoted by $w(\Gamma, b)$. If $\gamma = \tau_i \in \Gamma$, we denote by $M(\Gamma, \tau_i)$ the associated simple decorated representation of (Q_Γ, W_Γ); see more details in [Derksen et al. 2008; Labardini-Fragoso 2009b]. It is shown in this latter reference that the flips of triangulations are compatible with the mutations of decorated representations.

On the other hand, let $\Gamma' = f_{\tau_i}(\Gamma)$ be the triangulation of (S, M) obtained by a flip along τ_i; then [Keller and Yang 2009] establishes an equivalence $\bar{\mu}_i : \mathscr{C}_\Gamma \to \mathscr{C}_{\Gamma'}$ for each $1 \leq i \leq n$. Viewing the indecomposable objects in $\mathscr{C}_\Gamma$ as noncontractible curves and noncontractible closed curves in (S, M), or as indecomposable decorated representation of (Q_Γ, W_Γ), then [Plamondon 2010] shows the compatibility between the equivalence $\bar{\mu}_i$ and the mutation of decorated representations of (Q_Γ, W_Γ). Therefore, we get following lemma.

Lemma 4.5. *Let $\Gamma' = f_{\tau_i}(\Gamma)$ be the triangulation of (S, M) obtained by a flip along τ_i and let $T_{\Gamma'}$ be the corresponding cluster-tilting object in $\mathscr{C}_{\Gamma'}$. If $\bar{\mu}_i : \mathscr{C}_\Gamma \longrightarrow \mathscr{C}_{\Gamma'}$ denotes the equivalence just discussed, then*

$$\mathrm{Ext}^1_{\mathscr{C}_{\Gamma'}}(T_{\Gamma'}, \bar{\mu}_i(\gamma)) = M(\Gamma', \gamma)$$

for any non-boundary curve γ in (S, M).

Proof. Let $\mathrm{Ind}_s\mathscr{C}_\Gamma$ be the set of all indecomposable string objects in $\mathscr{C}_\Gamma$ which is indexed by non-boundary curves in (S, M), and

$$D_\Gamma = \{M(\Gamma, \gamma) |\ \gamma \text{ curve in } (S, M)\}.$$

Then by Proposition 4.1 in [Plamondon 2010], we have the commutative diagram

$$\begin{tikzcd}
\mathrm{Ind}_s\mathscr{C}_\Gamma \arrow[r, "\mathrm{Ext}^1_{\mathscr{C}_\Gamma}(T_\Gamma, -)"] \arrow[d, "\mu_i"'] & D_\Gamma \arrow[d, "\bar{\mu}_i"] \\
\mathrm{Ind}_s\mathscr{C}_{\Gamma'} \arrow[r, "\mathrm{Ext}^1_{\mathscr{C}_{\Gamma'}}(T_{\Gamma'}, -)"] & D_{\Gamma'}
\end{tikzcd}$$

where μ_i is the mutation of decorated representations. Therefore, for each curve γ in (S, M) we have

$$\begin{aligned}\mathrm{Ext}^1_{\mathscr{C}_{\Gamma'}}(T_{\Gamma'}, \bar{\mu}_i(\gamma)) &= \mu_i(\mathrm{Ext}^1_{\mathscr{C}_\Gamma}(T_\Gamma, \gamma)) \\ &\cong \mu_i(M(\Gamma, \gamma)) \cong M(f_{\tau_i}(\Gamma), \gamma) = M(\Gamma', \gamma),\end{aligned}$$

where the last isomorphism is given by the main result in [Labardini-Fragoso 2009b]. This completes the proof. □

Since $\bar{\mu}_i : \mathscr{C}_\Gamma \to \mathscr{C}_{\Gamma'}$ is an equivalence, the objects $\bar{\mu}_i(\gamma)$ in $\mathscr{C}_{\Gamma'}$ can also be described by curves in (S, M) and we denote $\bar{\mu}_i(\gamma)$ again by γ in $\mathscr{C}_{\Gamma'}$.

Corollary 4.6. *Each triangulation of (S, M) yields a cluster-tilting object in $\mathscr{C}_{(S,M)}$ and $\mathrm{Ext}^1_{\mathscr{C}}(\gamma, \gamma) = 0$ if γ is an internal arc in (S, M).*

Proof. Let Γ be the triangulation of (S, M) that we studied before, and Γ' be another triangulation. By Theorem 4.1, there exists a sequence of flips which transform Γ to Γ':

$$\Gamma = \Gamma_0 \overset{f_{i_1}}{\text{---}} \Gamma_1 \overset{f_{i_2}}{\text{---}} \Gamma_2 \overset{f_{i_3}}{\text{---}} \cdots \overset{f_{i_n}}{\text{---}} \Gamma_n = \Gamma',$$

where Γ_j is a triangulation of (S, M) and f_{i_j} is a flip for $1 \le j \le n$. We know that Γ induces a cluster-tilting object T_Γ which is the direct sum of all internal arcs in Γ, and Theorem 4.4 implies that Γ_1 also induces a cluster-tilting object T_{Γ_1} given by all internal arcs in Γ_1. Lemma 4.5 allows us to keep doing this until $\Gamma_n = \Gamma'$, which completes the proof. □

Let Γ be a triangulation of (S, M) and let γ and δ be two non-boundary curves in (S, M). From Theorem 2.3 we know that the morphism space $\mathrm{Hom}_{\mathscr{C}}(\gamma, \delta)$ in $\mathscr{C} = \mathscr{C}_\Gamma$ can be decomposed as k-vector space as follows:

$$\mathrm{Hom}_{J(Q_\Gamma, W_\Gamma)}\big(M(\Gamma, \gamma), M(\Gamma, \delta)\big) \oplus \big\{f \in \mathrm{Hom}_{\mathscr{C}}(\gamma, \delta) \mid f \text{ factors through } T_\Gamma\big\}.$$

Moreover, by Lemma 3.3 in [Palu 2008], the morphisms factoring through T_Γ can be replaced by the morphisms of two other related objects in mod $J(Q_\Gamma, W_\Gamma)$, so $\mathrm{Hom}_{\mathscr{C}}(\gamma, \delta)$ is given by

$$\mathrm{Hom}_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \gamma), M(\Gamma, \delta)) \oplus D\,\mathrm{Hom}_{J(Q_\Gamma, W_\Gamma)}(\tau^{-1}M(\Gamma, \delta), \tau M(\Gamma, \gamma)).$$

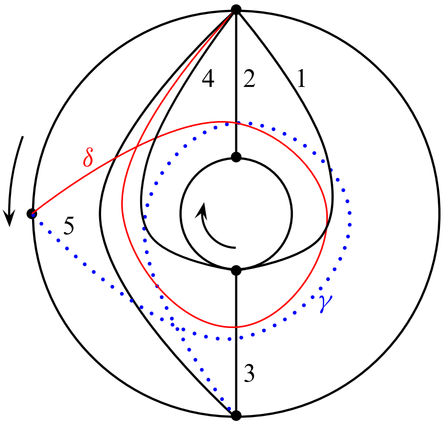
In [Crawley-Boevey 1989] a basis of the Hom-space between string modules is explicitly described using factor strings and substrings. Using Theorem 2.4, one can translate this into a description using curves and the triangulation Γ of (S, M); thus one could give a basis of $\operatorname{Hom}_{\mathscr{C}}(\gamma, \delta)$ combinatorially in terms of Γ.

Another possibility is to choose a triangulation Γ' such that a given morphism in $\mathscr{C}$ factoring through T_Γ can be described by a morphism in

$$\operatorname{Hom}_{J(Q_{\Gamma'}, W_{\Gamma'})}\big(M(\Gamma', \gamma), M(\Gamma', \delta)\big).$$

This is illustrated in the following example.

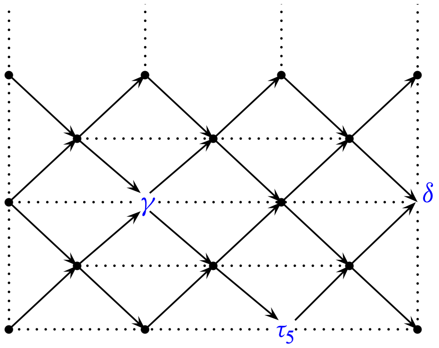
Example 4.7. We consider the following two curves γ, δ from Example 2.6:



(the solid red line represents δ, and the dotted line γ). It is easy to see that

$$M(\Gamma, \delta) = \begin{smallmatrix} & 1 & & & \\ 3 & & 2 & & \\ & & & 4 & \\ & & & & 5 \end{smallmatrix} \quad \text{and} \quad M(\Gamma, \gamma) = \begin{smallmatrix} 5 & & 1 & & \\ & 3 & & 2 & \\ & & & & 4 \end{smallmatrix}$$

Hence $\operatorname{Hom}_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \gamma), M(\Gamma, \delta))$ is one-dimensional with a basis element f mapping the factor string 5 of $w(\Gamma, \gamma)$ to the substring 5 of $w(\Gamma, \delta)$. From Corollary 3.19 one obtains that γ and δ lie in a tube of rank 3 in $\mathscr{C}$ as follows:



Thus there is, besides the morphism f induced from $\operatorname{mod} J(Q_\Gamma, W_\Gamma)$, another nonzero morphism $g \in \operatorname{Hom}_{\mathscr{C}}(\gamma, \delta)$ factoring through τ_5. Obviously, g cannot be described by morphisms in $\operatorname{mod} J(Q_\Gamma, W_\Gamma)$ since $M(\Gamma, \tau_5) = 0$, but we can realize g in another module category: Applying a flip along τ_5 we obtain a new triangulation Γ' whose associated quiver with potential is $Q_{\Gamma'} = \tilde{A}_4$, $W_{\Gamma'} = 0$. Denote by $5^* = M(\Gamma', \tau_5)$ the simple $\tilde{A}_4$-module concentrated in 5; then

$$M(\Gamma', \delta) = \begin{array}{ccccc} & & 1 & & \\ & 3 & & 2 & \\ 5^* & & & & 4 \end{array} \qquad M(\Gamma', \gamma) = \begin{array}{ccccc} & 1 & & & \\ 3 & & 2 & & 5^* \\ & & & 4 & \end{array} \qquad M(\Gamma', \tau_5) = \; 5^*$$

Hence $\operatorname{Hom}_{\tilde{A}_4}(M(\Gamma', \gamma), M(\Gamma', \delta))$ is also one-dimensional, given by a basis element g' mapping the factor string 5^* of $w(\Gamma', \gamma)$ to the substring 5^* of $w(\Gamma', \delta)$, which factors through $M(\Gamma', \tau_5)$. Therefore $g \in \operatorname{Hom}_{\mathscr{C}}(\gamma, \delta)$ can be described by $g' \in \operatorname{Hom}_{\tilde{A}_4}(M(\Gamma', \gamma), M(\Gamma', \delta))$. Moreover, since $\tilde{A}_4$ is hereditary, we can conclude that

$$\begin{aligned} \operatorname{Hom}_{\mathscr{C}}(\gamma, \delta) &\cong \operatorname{Hom}_{\mathscr{C}(\tilde{A}_4)}(M(\Gamma', \gamma), M(\Gamma', \delta)) \\ &= \operatorname{Hom}_{\tilde{A}_4}(M(\Gamma', \gamma), M(\Gamma', \delta)) \oplus \operatorname{Ext}^1_{\tilde{A}_4}(M(\Gamma', \gamma), M(\Gamma', \gamma)) \end{aligned}$$

is two-dimensional. Hence

$$\operatorname{Hom}_{\mathscr{C}}(\gamma, \delta) \cong \operatorname{Hom}_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \gamma), M(\Gamma, \delta)) \oplus \operatorname{Hom}_{\tilde{A}_4}(M(\Gamma', \gamma), M(\Gamma', \delta))$$

as k-vector spaces.

Using Lemma 3.3 in [Palu 2008], we have

$$\tau M(\Gamma, \gamma) = M(\Gamma, \delta) = \begin{array}{ccccc} & 1 & & & \\ 3 & & 2 & & \\ & & & 4 & \\ & & & & 5 \end{array} \quad \text{and} \quad \tau^{-1} M(\Gamma, \delta) = M(\Gamma, \gamma) = \begin{array}{ccccc} 5 & & 1 & & \\ & 3 & & 2 & \\ & & & & 4 \end{array}$$

Hence, $\operatorname{Hom}_{\mathscr{C}}(\gamma, \delta)$ is given by

$$\operatorname{Hom}_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \gamma), M(\Gamma, \delta)) \oplus D \operatorname{Hom}_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \gamma), M(\Gamma, \delta))$$

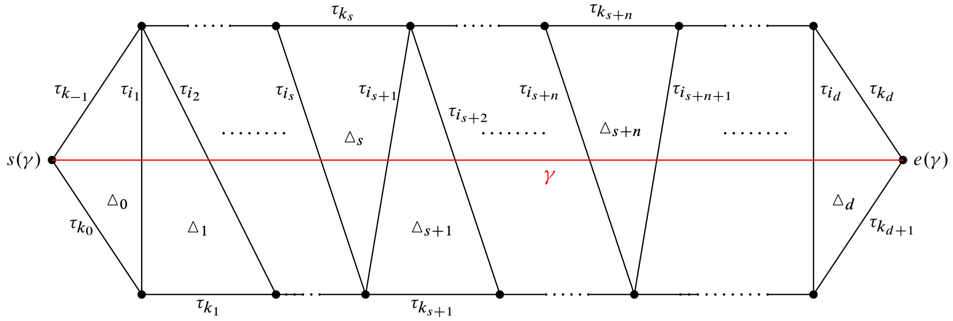
as k-vector space.

5. Extensions and intersections

Next we study the relation between extensions in the category $\mathscr{C} = \mathscr{C}_{(S,M)}$ and intersections of curves in (S, M). Given any curve in (S, M) with self-intersections, we explicitly construct one or two new curves, sometimes resolving the self-intersection, and sometimes increasing the winding number. These curves serve as middle term of certain nonsplit short exact sequences which allow to prove the following theorem:

Theorem 5.1. $\operatorname{Ext}^1_{\mathscr{C}}(\gamma, \gamma) = 0$ *if and only if* γ *is an internal arc in* (S, M).

Proof. Let γ be a curve in (S, M) and fix a triangulation Γ of (S, M). We denote by $\tau_{i_1}, \tau_{i_2}, \dots, \tau_{i_d}$ the internal arcs of Γ that intersect γ, as indicated here:



Along its way, the curve γ passes through (not necessarily distinct) triangles $\triangle_0, \triangle_1, \dots, \triangle_d$. For $1 \le l \le d-1$, the triangle $\triangle_l$ is formed by the internal arcs τ_{i_l} and $\tau_{i_{l+1}}$ and a third arc which is denoted by τ_{k_l}. In the first triangle $\triangle_0$, we denote the arcs clockwise by τ_{i_1}, τ_{k_0} and $\tau_{k_{-1}}$. Similarly, $\triangle_d$ is formed by τ_{i_d}, τ_{k_d} and $\tau_{k_{d+1}}$, as shown in the figure.

By Corollary 4.6 we only need to prove one direction of the theorem, namely that $\operatorname{Ext}^1_{\mathscr{C}}(\gamma, \gamma) \neq 0$ in case the curve γ has self-intersections. To do so, it is enough to prove $\operatorname{Ext}^1_{J(Q_{\Gamma'}, W_{\Gamma'})}(M(\Gamma', \gamma), M(\Gamma', \gamma)) \neq 0$ for some triangulation Γ'. In some cases one can choose $\Gamma' = \Gamma$, but sometimes it will be necessary to change Γ to Γ' by a sequence of flips in order to realize the extension over some algebra $J(Q_{\Gamma'}, W_{\Gamma'})$.

We therefore assume that the curve γ has self-intersections. Thus there are $0 < r < r' < 1$ such that $\gamma(r) = \gamma(r')$. We choose r and r' in such a way that the restriction $b = \gamma|_{[r,r']}$ is a simple noncontractible closed curve. Hence the subword $w(\Gamma, \gamma|_{[r,r']})$ of $w(\Gamma, \gamma)$ defines a band $w(\Gamma, b)$ in mod $J(Q_\Gamma, W_\Gamma)$ as follows:

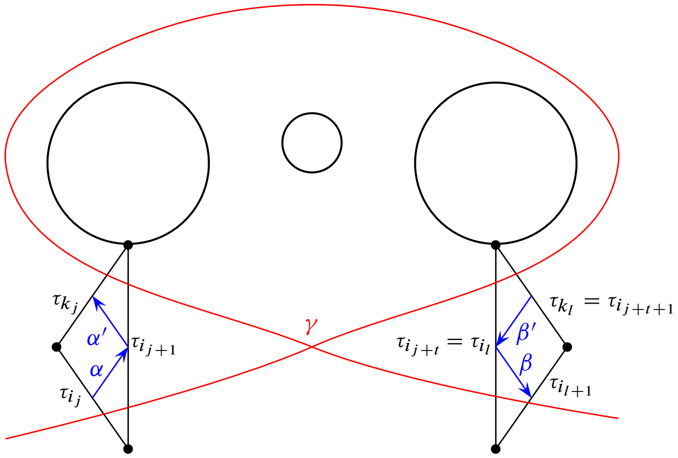
$$\begin{tikzcd}
 & & \tau_{i_{s+2}} \arrow[rr, dash] & & \tau_{i_{s+3}} \arrow[dr, dash] & \\
w(\Gamma, b): & \tau_{i_{s+1}} \arrow[ur, dash] \arrow[dr, "\theta"'] & & & & \tau_{i_{s+4}} \arrow[dl, dash] \\
 & & \tau_{i_{s+n}} & \cdots \arrow[r, dash] & \tau_{i_{s+5}} &
\end{tikzcd}$$

The difference between the string $w(\Gamma, \gamma|_{[r,r']})$ and the band $w(\Gamma, b)$ is the arrow between the endpoints. Up to duality, we may assume that this arrow θ (which is induced by the triangle $\triangle_{i_s}$), is oriented from $\tau_{i_{s+1}}$ towards $\tau_{i_{s+n}}$. We distinguish several cases of how the band $w(\Gamma, b)$ is embedded in the string $w(\Gamma, \gamma)$:

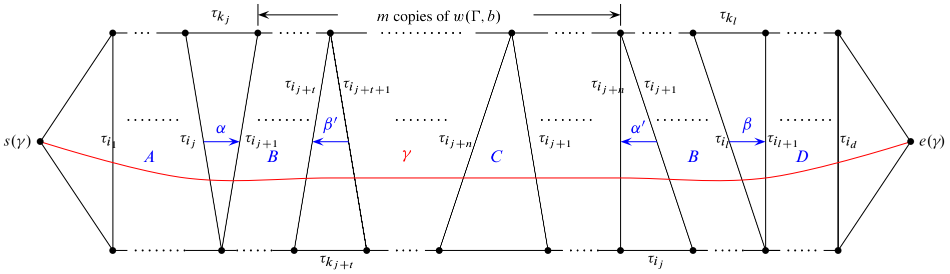
Case I: $w(\Gamma, \gamma)$ contains the band $w(\Gamma, b)$.

In this case, $w(\Gamma, \gamma)$ contains the string $w(\Gamma, \gamma|_{[r,r']})$ and the arrow θ and thus $m \ge 1$ consecutive copies of $w(\Gamma, b)$. We extend the subword of $w(\Gamma, \gamma)$ which is induced by b to maximal length, that is, we choose $j \le s$ and $l \ge s+n$ such that all arrows between $\tau_{i_{j+1}}$ and τ_{i_l} belong to $w(\Gamma, b)$, but the arrows $\tau_{i_j} \overset{\alpha}{\text{—}} \tau_{i_{j+1}}$ induced by $\triangle_{i_j}$ and $\tau_{i_l} \overset{\beta}{\text{—}} \tau_{i_{l+1}}$ induced by $\triangle_{i_l}$ are not induced by b.

Without loss of generality, assume α is oriented as $\tau_{i_j} \to \tau_{i_{j+1}}$. Then β has orientation $\tau_{i_l} \to \tau_{i_{l+1}}$ since γ intersects itself. Therefore the arrows $\alpha' : \tau_{i_{j+1}} \to \tau_{k_j}$ induced by $\triangle_{i_j}$ and the arrow $\beta' : \tau_{k_l} \to \tau_{i_l}$ induced by $\triangle_{i_l}$ belong to $w(\Gamma, b)$. We assume $\tau_{i_{j+t}} = \tau_{i_l}$, $\tau_{i_{j+t+1}} = \tau_{k_l}$ with $1 \leq t \leq n$. If $m = 1$, the situation is depicted as follows:



Here, the boundaries are indicated by the circles, and $\tau_{i_j} = \tau_{k_0}$ when $j = 0$ (also note that τ_{i_j} might be a boundary arc). We rewrite the figure from the preceding page accordingly:



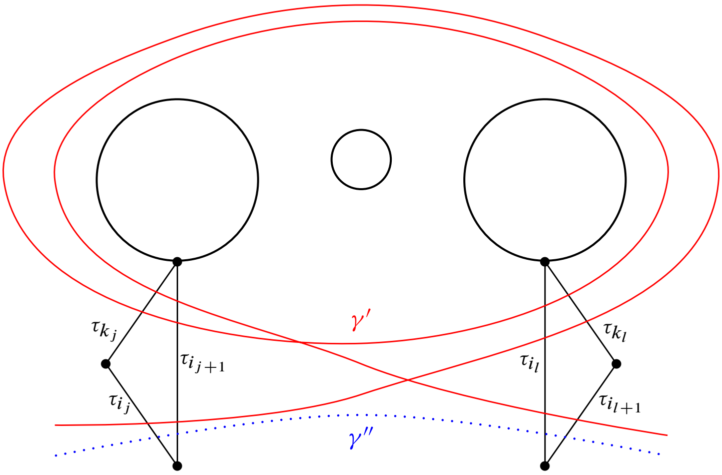
Thus $w(\Gamma, \gamma)$ is given by

$$w(\Gamma, \gamma) = A \text{—} \tau_{i_j} \xrightarrow{\alpha} \tau_{i_{j+1}} \text{—} B \text{—} \tau_{i_l} \xleftarrow{\beta'} \tau_{k_l} \text{—} C \text{—} \tau_{i_{j+n}} \xleftarrow{\alpha'} \tau_{i_{j+1}} \text{—} B \text{—} \tau_{i_l} \xrightarrow{\beta} \tau_{i_{l+1}} \text{—} D.$$

Consider the following two strings in mod $J(Q_\Gamma, Q_\Gamma)$:

$$\begin{aligned} w(\Gamma, \gamma') = A \text{—} \tau_{i_j} \xrightarrow{\alpha} \tau_{i_{j+1}} \text{—} B \text{—} \tau_{i_l} \xleftarrow{\beta'} \tau_{k_l} \text{—} C \text{—} \tau_{i_{j+n}} \xleftarrow{\alpha'} \tau_{i_{j+1}} \text{—} B \text{—} \tau_{i_l} \xleftarrow{\beta'} \tau_{k_l} \\ \text{—} C \text{—} \tau_{i_{j+n}} \xleftarrow{\alpha'} \tau_{i_{j+1}} \text{—} B \text{—} \tau_{i_l} \xrightarrow{\beta} \tau_{i_{l+1}} \text{—} D, \\ w(\Gamma, \gamma'') = A \text{—} \tau_{i_j} \xrightarrow{\alpha} \tau_{i_{j+1}} \text{—} B \text{—} \tau_{i_l} \xrightarrow{\beta} \tau_{i_{l+1}} \text{—} D, \end{aligned}$$

where γ' and γ'' are the corresponding curves in (S, M). If $m = 1$, then γ', γ'' can be visualized as follows, where γ' is given by the full curve and γ'' is described by the dotted curve:



We easily construct a nonsplit exact sequence in mod $J(Q_\Gamma, W_\Gamma)$:

$$0 \longrightarrow M(\Gamma, \gamma) \xrightarrow{\left[\begin{smallmatrix} f \\ g \end{smallmatrix}\right]} M(\Gamma, \gamma') \oplus M(\Gamma, \gamma'') \longrightarrow M(\Gamma, \gamma) \longrightarrow 0,$$

where f identifies the factor string

$$A - \tau_{i_j} \xrightarrow{\alpha} \tau_{i_{j+1}} - B - \tau_{i_l} \xleftarrow{\beta'} \tau_{k_l} - C - \tau_{i_{j+n}} \xleftarrow{\alpha'} \tau_{i_{j+1}} - B - \tau_{i_l}$$

of $w(\Gamma, \gamma)$ with a substring of $w(\Gamma, \gamma')$ and g sends the factor string

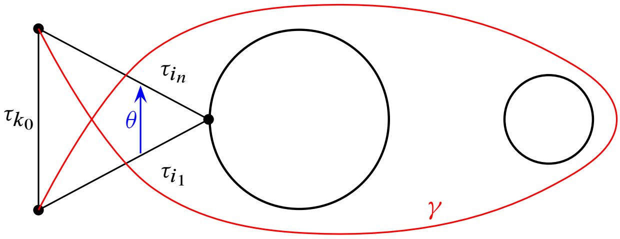
$$\tau_{i_{j+1}} - B - \tau_{i_l} \xrightarrow{\beta} \tau_{i_{l+1}} - D$$

of $w(\Gamma, \gamma)$ to the same substring of $w(\Gamma, \gamma'')$.

Case II: $w(\Gamma, \gamma)$ does not contain the band $w(\Gamma, b)$.

In this case, $w(\Gamma, \gamma)$ does not contain θ but it contains $w(\Gamma, \gamma|_{[r,r']})$ as a subword. We distinguish a number of subcases:

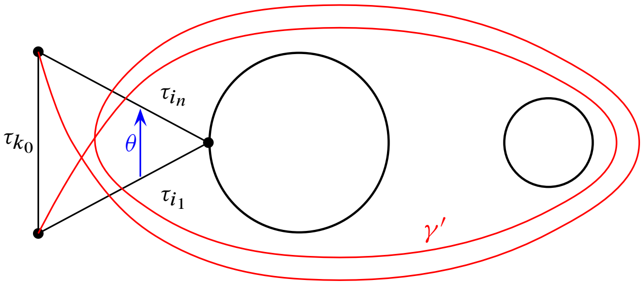
(II.1) $\underline{0 = s < s + n = d}$. Here τ_{i_s} equals τ_{k_0} and the subword $w(\Gamma, \gamma|_{[r,r']})$ equals $w(\Gamma, \gamma)$ (see figure):



Since $w(\Gamma, b)$ is a band we know that there exists a string of the form

$$w(\Gamma, \gamma') = \tau_{i_1} \leftrightsquigarrow \tau_{i_n} \xleftarrow{\theta} \tau_{i_1} \leftrightsquigarrow \tau_{i_n}$$

in mod $J(Q_\Gamma, W_\Gamma)$, where γ' can be visualized as follows:

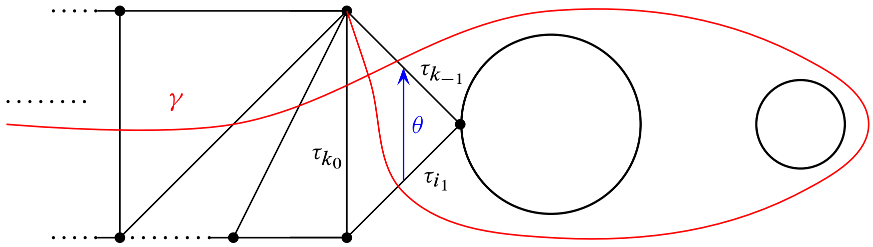


It is easy to see that there is a nonsplit exact sequence in mod $J(Q_\Gamma, W_\Gamma)$:

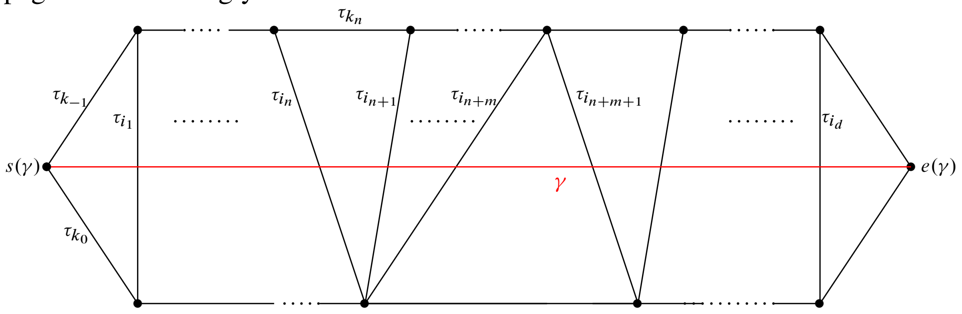
$$0 \longrightarrow M(\Gamma, \gamma) \xrightarrow{f} M(\Gamma, \gamma') \longrightarrow M(\Gamma, \gamma) \longrightarrow 0$$

where f is induced by the embedding of the string $w(\Gamma, \gamma)$ as a substring of $w(\Gamma, \gamma')$.

(II.2) $\underline{0 = s < s + n < d}$. Here $\tau_{i_s} = \tau_{i_0} = \tau_{k_0}$ and the subword $w(\Gamma, \gamma|_{[r,r']})$ is a proper subword of $w(\Gamma, \gamma)$ which ends before reaching the endpoint $e(\gamma)$ of γ:



Since γ intersects itself, $\tau_{k_{-1}} = \tau_{i_n}$ and $\tau_{i_{n+1}} = \tau_{k_0}$. We rewrite the diagram from page 557 accordingly:



where $\tau_{i_{n+2}}, \tau_{i_{n+3}} \ldots \tau_{i_{n+m}}$ with $m \geq 1$ are all internal arcs lying clockwise before $\tau_{i_{n+1}}$. After applying flips along $\tau_{i_{n+1}}, \tau_{i_{n+2}} \ldots \tau_{i_{n+m}}$, we get a new triangulation

Γ' and a new band $w(\Gamma', b)$ in mod $J(Q_{\Gamma'}, W_{\Gamma'})$ related to same noncontractible closed curve b:

$$\begin{tikzcd}[arrows=dash]
 & & \tau^*_{i_{n+m-1}} \arrow[r, dotted] & \tau^*_{i_{n+1}} \arrow[dr] & \\
w(\Gamma', b): & \tau^*_{i_{n+m}} \arrow[ur] \arrow[dr] & & & \tau_{i_1} \\
 & & \tau_{i_n} \arrow[r, dotted] & {} \arrow[ur] &
\end{tikzcd}$$

If $d = n + m$, the subword $w(\Gamma', b)$ is equal to

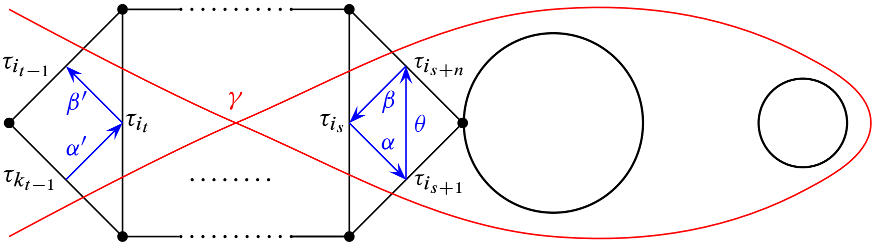
$$w(\Gamma', \gamma) = \tau^*_{i_{n+m}} - \tau^*_{i_{n+m-1}} \leftrightsquigarrow \tau_{i_n}$$

in mod $J(Q_{\Gamma'}, W_{\Gamma'})$, and this is the same case as in II.1 when mod $J(Q_{\Gamma'}, W_{\Gamma'})$ is considered. If $d > n + m$, then

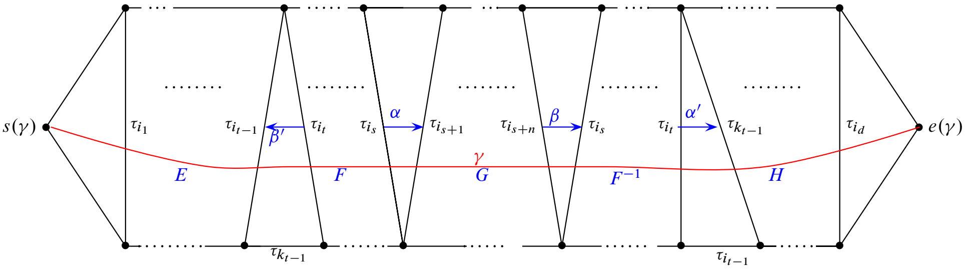
$$w(\Gamma', \gamma) = \tau^*_{i_{n+m}} - \tau^*_{i_{n+m-1}} \leftrightsquigarrow \tau_{i_n} - \tau^*_{i_{n+m}} - \cdots ;$$

hence $w(\Gamma', \gamma)$ contains the band $w(\Gamma', b)$ as subword — the same case as in I.

(II.3) $\underline{0 < s < s + n < d}$. Here the subword $w(\Gamma, \gamma|_{[r,r']})$ is a proper subword of $w(\Gamma, \gamma)$ that starts after τ_{i_1} and ends before reaching the endpoint $e(\gamma)$ of γ. Since γ intersects itself, there must exist $1 \le t \le s$ such that $\tau_{i_{s+n+l}} = \tau_{i_{s+1-l}}$ for $1 \le l \le s - t + 1$ and $\tau_{i_{s+n+t+1}} = \tau_{k_{t-1}}$. The situation is this:



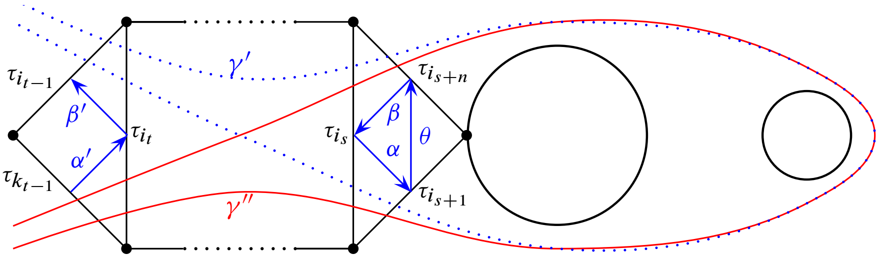
where α and β are arrows induced by the triangle $\triangle_{i_s}$, and α' and β' are induced by the triangle $\triangle_{i_{t-1}}$. We rewrite the diagram from page 557 accordingly:



Hence $w(\Gamma, \gamma)$ can be given as

$$E - \tau_{i_{t-1}} \xleftarrow{\beta'} \tau_{i_t} - F - \tau_{i_s} \xrightarrow{\alpha} \tau_{i_{s+1}} - G - \tau_{i_{s+n}} \xrightarrow{\beta} \tau_{i_s} - F^{-1} - \tau_{i_t} \xleftarrow{\alpha'} \tau_{k_{t-1}} - H.$$

We consider the two curves in (S, M) marked γ' (solid line) and γ'' (dotted line) in the figure:



Then $w(\Gamma, \gamma')$ is given by

$$E - \tau_{i_{t-1}} \xleftarrow{\beta'} \tau_{i_t} - F - \tau_{i_s} \xrightarrow{\alpha} \tau_{i_{s+1}} - G - \tau_{i_{s+n}} \xrightarrow{\beta} \tau_{i_s} - F^{-1} - \tau_{i_t} \xrightarrow{\beta'} \tau_{i_{t-1}} - E^{-1}$$

and $w(\Gamma, \gamma'')$ by

$$H^{-1} - \tau_{k_{t-1}} \xrightarrow{\alpha'} \tau_{i_t} - F - \tau_{i_s} \xrightarrow{\alpha} \tau_{i_{s+1}} - G^{-1} - \tau_{i_{s+n}} \xrightarrow{\beta} \tau_{i_s} - F^{-1} - \tau_{i_t} \xleftarrow{\alpha'} \tau_{k_{t-1}} - H.$$

Hence there is a nonsplit exact sequence in mod $J(Q_\Gamma, W_\Gamma)$

$$0 \longrightarrow M(\Gamma, \gamma) \xrightarrow{\begin{bmatrix} f \\ g \end{bmatrix}} M(\Gamma, \gamma') \oplus M(\Gamma, \gamma'') \longrightarrow M(\Gamma, \gamma) \longrightarrow 0,$$

where f sends the factor string $w_0 = E - \tau_{i_{t-1}} \xleftarrow{\beta'} \tau_{i_t} - F - \tau_{i_s}$ of $w(\Gamma, \gamma)$ to w_0^{-1} as a substring of $w(\Gamma, \gamma')$ and g identifies the factor string

$$w_1 = \tau_{i_t} - F - \tau_{i_s} \xrightarrow{\alpha} \tau_{i_{s+1}} - G - \tau_{i_{s+n}} \xrightarrow{\beta} \tau_{i_s} - F^{-1} - \tau_{i_t} \xleftarrow{\alpha'} \tau_{k_{t-1}} - H$$

with w_1^{-1} as a substring of $w(\Gamma, \gamma'')$.

(II.4) $\underline{0 < s < s + n = d}$. Here the subword $w(\Gamma, \gamma|_{[r,r']})$ is a proper subword of $w(\Gamma, \gamma)$ which ends at the endpoint $e(\gamma)$ of γ. This subcase is dual to II.2. □

Remark 5.2. $\mathrm{Ext}^1_{\mathscr{C}}(\gamma, \gamma) = 0$ implies $\mathrm{Ext}^1_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \gamma), M(\Gamma, \gamma)) = 0$ for any triangulation Γ of (S, M). But $\mathrm{Ext}^1_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \gamma), M(\Gamma, \gamma)) = 0$ for one triangulation Γ of (S, M) does not imply $\mathrm{Ext}^1_{\mathscr{C}}(\gamma, \gamma) = 0$ in general. We reconsider δ in Example 4.7. It is easy to see that

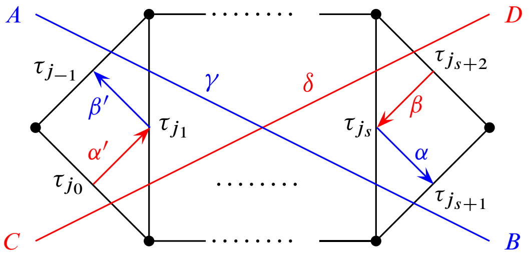
$$M(\Gamma, \delta) = \begin{array}{ccccc} & & 1 & & \\ & 3 & & 2 & \\ & & & & 4 \\ & & & & & 5 \end{array} \quad \text{and} \quad M(\Gamma, \delta[1]) = \begin{array}{ccc} & 1 & \\ 2 & & 3 \\ & 4 & \end{array}$$

Hence $\mathrm{Ext}^1_{J(Q_\Gamma, W_\Gamma)}(M(\Gamma, \delta), M(\Gamma, \delta)) = 0$ by the Auslander–Reiten formula in mod $J(Q_\Gamma, W_\Gamma)$, since the nonzero morphism from $M(\Gamma, \delta)$ to $\tau(M(\Gamma, \delta)) = M(\Gamma, \delta[1])$ factors through the injective module I_5. But after applying a flip along 5, we get $\mathrm{Ext}^1_{\tilde{A}_4}(M(\Gamma', \delta), M(\Gamma', \delta)) \neq 0$, which implies that $\mathrm{Ext}^1_{\mathscr{C}}(\delta, \delta) \neq 0$.

Similarly as in Theorem 5.1, we can study the extensions of two different curves by their intersections in (S, M).

Proposition 5.3. *Let γ, δ be distinct curves in (S, M) and suppose $I(\gamma, \delta) \neq 0$. Then* $\operatorname{Ext}^1_{\mathscr{C}}(\gamma, \delta) \neq 0 \neq \operatorname{Ext}^1_{\mathscr{C}}(\delta, \gamma)$.

Proof. We only consider one of the intersections of γ and δ. After some flips (if needed), there is a triangulation Γ' such that $w(\Gamma', \gamma)$ and $w(\Gamma', \delta)$ share a common subword (associated to the intersection) $w = \tau_{j_1} \cdots \tau_{j_s}$ with $s \geq 1$, as shown:



Here either τ_{j_0} or $\tau_{j_{-1}}$ might be a boundary arc, and one of $\tau_{j_{s+1}}$ and $\tau_{j_{s+2}}$ can also be a boundary arc. Hence $w(\Gamma', \gamma)$ and $w(\Gamma', \delta)$ are of the form

$$w(\Gamma', \gamma) = A - \tau_{j_{-1}} \xleftarrow{\beta'} \tau_{j_1} \leftrightsquigarrow \tau_{j_s} \xrightarrow{\alpha} \tau_{j_{s+1}} - B,$$

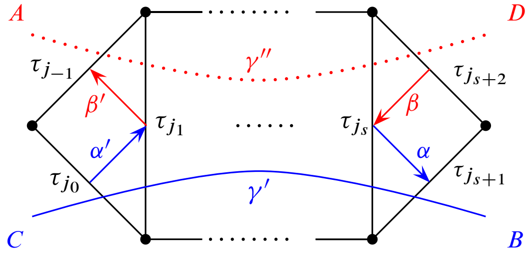
$$w(\Gamma', \delta) = C - \tau_{j_0} \xrightarrow{\alpha'} \tau_{j_1} \leftrightsquigarrow \tau_{j_s} \xleftarrow{\beta} \tau_{j_{s+2}} - D.$$

Consider the strings

$$(\Gamma', \gamma') = C - \tau_{j_0} \xrightarrow{\alpha'} \tau_{j_1} \leftrightsquigarrow \tau_{j_s} \xrightarrow{\alpha} \tau_{j_{s+1}} - B,$$

$$w(\Gamma', \gamma'') = A - \tau_{j_{-1}} \xleftarrow{\beta'} \tau_{j_1} \leftrightsquigarrow \tau_{j_s} \xleftarrow{\beta} \tau_{j_{s+2}} - D,$$

where γ' and γ'' are given in the picture:



Then there is a nonsplit exact sequence in mod $J(Q_{\Gamma'}, W_{\Gamma'})$

$$0 \longrightarrow M(\Gamma', \gamma) \xrightarrow{\begin{bmatrix} f \\ g \end{bmatrix}} M(\Gamma', \gamma') \oplus M(\Gamma', \gamma'') \longrightarrow M(\Gamma', \delta) \longrightarrow 0,$$

where f sends the factor string

$$w_0 = \tau_{j_1} \leftrightsquigarrow \tau_{j_s} \xrightarrow{\alpha} \tau_{j_{s+1}} - B$$

of $w(\Gamma', \gamma)$ to w_0 as a substring of $w(\Gamma', \gamma')$ and g identifies the factor string $w_1 = A \text{—} \tau_{j_{-1}} \xleftarrow{\beta'} \tau_{j_1} \leftrightsquigarrow \tau_{j_s}$ as a substring of $w(\Gamma', \gamma'')$. Therefore

$$\operatorname{Ext}^1_{J(Q_{\Gamma'}, W_{\Gamma'})}(M(\Gamma', \delta), M(\Gamma', \gamma)) \neq 0,$$

which implies that $\operatorname{Ext}^1_{\mathscr{C}}(\delta, \gamma) \neq 0$. Then, since $\mathscr{C}$ is 2-Calabi–Yau, we obtain

$$\operatorname{Ext}^1_{\mathscr{C}}(\gamma, \delta) \cong D \operatorname{Ext}^1_{\mathscr{C}}(\delta, \gamma) \neq 0. \qquad \square$$

Proposition 5.3 and Corollary 4.6 imply:

Corollary 5.4. *Let γ, δ be distinct internal arcs in (S, M). Then $\operatorname{Ext}^1_{\mathscr{C}}(\gamma, \delta) = 0$ if and only if $I(\gamma, \delta) = 0$.*

Corollary 5.5. *There is a bijection between triangulations of (S, M) and cluster-tilting objects of $\mathscr{C}_{(S,M)}$. In particular, each indecomposable object without self-extensions is reachable from the cluster-tilting object T_Γ (the initial cluster-tilting object).*

Proof. Corollary 4.6 implies that each triangulation of (S, M) yields a cluster-tilting object in $C_{(S,M)}$. Hence it suffices to prove that each cluster tilting object $T = T_1 \oplus \cdots \oplus T_n$ gives a triangulation of (S, M). By Theorem 5.1, we can assume $T = \tau_1 \oplus \cdots \oplus \tau_n$ where each τ_i is an internal arc in (S, M) corresponding to T_i. The definition of a triangulation and the above corollary yield a unique triangulation $\Gamma_T = \{\tau_1, \cdots, \tau_n, \tau_{n+1}, \cdots, \tau_{n+m}\}$ where $\tau_{n+1}, \cdots, \tau_{n+m}$ are boundary arcs in (S, M). $\square$

Acknowledgement

Jie Zhang thanks Claire Amiot and Yann Palu for interesting discussions on this problem.

References

[Amiot 2009] C. Amiot, "Cluster categories for algebras of global dimension 2 and quivers with potential", *Ann. Inst. Fourier* (*Grenoble*) **59**:6 (2009), 2525–2590. MR 2011c:16026 Zbl 05673905

[Assem et al. 2008] I. Assem, T. Brüstle, and R. Schiffler, "Cluster-tilted algebras as trivial extensions", *Bull. Lond. Math. Soc.* **40**:1 (2008), 151–162. MR 2009c:16086

[Assem et al. 2010] I. Assem, T. Brüstle, G. Charbonneau-Jodoin, and P.-G. Plamondon, "Gentle algebras arising from surface triangulations", *Algebra Number Theory* **4**:2 (2010), 201–229. MR 2011h:16045 Zbl 05704468

[Buan et al. 2006] A. B. Buan, R. Marsh, M. Reineke, I. Reiten, and G. Todorov, "Tilting theory and cluster combinatorics", *Adv. Math.* **204**:2 (2006), 572–618. MR 2007f:16033 Zbl 1127.16011

[Buan et al. 2008] A. B. Buan, R. J. Marsh, and I. Reiten, "Cluster mutation via quiver representations", *Comment. Math. Helv.* **83**:1 (2008), 143–177. MR 2008k:16026 Zbl 1193.16016

[Butler and Ringel 1987] M. C. R. Butler and C. M. Ringel, "Auslander–Reiten sequences with few middle terms and applications to string algebras", *Comm. Algebra* **15**:1-2 (1987), 145–179. MR 88a:16055 Zbl 0612.16013

[Caldero and Chapoton 2006] P. Caldero and F. Chapoton, "Cluster algebras as Hall algebras of quiver representations", *Comment. Math. Helv.* **81**:3 (2006), 595–616. MR 2008b:16015 Zbl 1119.16013

[Caldero et al. 2006] P. Caldero, F. Chapoton, and R. Schiffler, "Quivers with relations arising from clusters (A_n case)", *Trans. Amer. Math. Soc.* **358**:3 (2006), 1347–1364. MR 2007a:16025 Zbl 1137.16020

[Crawley-Boevey 1989] W. W. Crawley-Boevey, "Maps between representations of zero-relation algebras", *J. Algebra* **126**:2 (1989), 259–263. MR 90k:16035 Zbl 0685.16018

[Derksen et al. 2008] H. Derksen, J. Weyman, and A. Zelevinsky, "Quivers with potentials and their representations. I. Mutations", *Selecta Math.* (*N.S.*) **14**:1 (2008), 59–119. MR 2010b:16021

[Fomin and Zelevinsky 2002] S. Fomin and A. Zelevinsky, "Cluster algebras. I. Foundations", *J. Amer. Math. Soc.* **15**:2 (2002), 497–529. MR 2003f:16050

[Fomin et al. 2008] S. Fomin, M. Shapiro, and D. Thurston, "Cluster algebras and triangulated surfaces. I. Cluster complexes", *Acta Math.* **201**:1 (2008), 83–146. MR 2010b:57032

[Geiss et al. 2005] C. Geiss, B. Leclerc, and J. Schröer, "Semicanonical bases and preprojective algebras", *Ann. Sci. École Norm. Sup.* (4) **38**:2 (2005), 193–253. MR 2007h:17018 Zbl 1131.17006

[Geiß et al. 2006] C. Geiß, B. Leclerc, and J. Schröer, "Rigid modules over preprojective algebras", *Invent. Math.* **165**:3 (2006), 589–632. MR 2007g:16023 Zbl 1167.16009

[Ginzburg 2006] V. Ginzburg, "Calabi–Yau algebras", preprint, 2006. arXiv math/0612139

[Iyama and Yoshino 2008] O. Iyama and Y. Yoshino, "Mutation in triangulated categories and rigid Cohen–Macaulay modules", *Invent. Math.* **172**:1 (2008), 117–168. MR 2008k:16028 Zbl 1140.18007

[Keller 2006] B. Keller, "On differential graded categories", pp. 151–190 in *International Congress of Mathematicians*, vol. 2, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2006. MR 2008g:18015 Zbl 1140.18008

[Keller 2010] B. Keller, "Cluster algebras, quiver representations and triangulated categories", pp. 76–160 in *Triangulated categories*, London Math. Soc. Lecture Note Ser. **375**, Cambridge Univ. Press, Cambridge, 2010. MR 2011h:13033 Zbl 1215.16012

[Keller and Reiten 2007] B. Keller and I. Reiten, "Cluster-tilted algebras are Gorenstein and stably Calabi-Yau", *Adv. Math.* **211**:1 (2007), 123–151. MR 2008b:18018 Zbl 1128.18007

[Keller and Yang 2009] B. Keller and D. Yang, "Quiver mutation and derived equivalences", preprint, 2009. arXiv 0906.0761

[Koenig and Zhu 2008] S. Koenig and B. Zhu, "From triangulated categories to abelian categories: cluster tilting in a general framework", *Math. Z.* **258**:1 (2008), 143–160. MR 2008m:18021 Zbl 1133.18005

[Küçük 2005] A. Küçük, "The fundamental group of an oriented surface of genus *n* with *k* boundary surfaces", *Appl. Math. Comput.* **160**:1 (2005), 141–145. MR 2100181

[Labardini-Fragoso 2009a] D. Labardini-Fragoso, "Quivers with potentials associated to triangulated surfaces", *Proc. Lond. Math. Soc.* (3) **98**:3 (2009), 797–839. MR 2010b:16033 Zbl 05551833

[Labardini-Fragoso 2009b] D. Labardini-Fragoso, "Quivers with potentials associated to triangulated surfaces, II: arc representations", preprint, 2009. arXiv 0909.4100

[Palu 2008] Y. Palu, "Cluster characters for 2-Calabi-Yau triangulated categories", *Ann. Inst. Fourier (Grenoble)* **58**:6 (2008), 2221–2248. MR 2009k:18013 Zbl 1154.16008

[Plamondon 2010] P.-G. Plamondon, "Cluster algebras via cluster categories with infinite-dimensional morphism spaces", preprint, 2010. arXiv 1004.0830

[Reiten and Van den Bergh 2002] I. Reiten and M. Van den Bergh, "Noetherian hereditary abelian categories satisfying Serre duality", *J. Amer. Math. Soc.* **15**:2 (2002), 295–366. MR 2003a:18011 Zbl 0991.18009

Communicated by Andrei Zelevinsky
Received 2010-05-14 Revised 2010-08-13 Accepted 2010-09-12

thomas.brustle@usherbrooke.ca *Département de Mathématiques, Université de Sherbrooke, Sherbrooke J1K 2R1, Canada*
Department of Mathematics, Bishop's University, Sherbrooke J1M 1Z7, Canada

Jie.Zhang@usherbrooke.ca *Département de Mathématiques, Université de Sherbrooke, Sherbrooke J1K 2R1, Canada*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in ANT are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 5 No. 4 2011

Global descent obstructions for varieties 431
JEAN-MARC COUVEIGNES and EMMANUEL HALLOUIN

Specializations of elliptic surfaces, and divisibility in the Mordell–Weil group 465
PATRICK INGRAM

Explicit CM theory for level 2-structures on abelian surfaces 495
REINIER BRÖKER, DAVID GRUENEWALD and KRISTIN LAUTER

On the cluster category of a marked surface without punctures 529
THOMAS BRÜSTLE and JIE ZHANG

1937-0652(2011)5:4;1-D