# GROSS-ZAGIER REVISITED

BRIAN CONRAD

Contents


1. Introduction 1
2. Some properties of abelian schemes and modular curves 3
3. The Serre-Tate theorem and the Grothendieck existence theorem 7
4. Computing naive intersection numbers 10
5. Intersection formula via Hom groups 12
6. Supersingular cases with $r_{\mathscr{A}}(m) = 0$ 15
7. Application of a construction of Serre 21
8. Intersection theory via meromorphic tensors 29
9. Self-intersection formula and application to global height pairings 34
10. Quaternionic explications 43
Appendix A. Elimination of quaternionic sums (by W.R. Mann) 49
References 64


## 1. Introduction

The aim of this paper is to rework [GZ, Ch III] based on more systematic deformation-theoretic methods so as to treat all imaginary quadratic fields, all residue characteristics, and all $j$-invariants on an equal footing. This leads to more conceptual arguments in several places and interpretations for some quantities which appear to otherwise arise out of thin air in [GZ, Ch III]. For example, the sum in [GZ, Ch III, Lemma 8.2] arises for us in (9.6), where it is given a deformation-theoretic meaning. Provided the analytic results in [GZ] are proven for even discriminants, the main results in [GZ] would be valid without parity restriction on the discriminant of the imaginary quadratic field. Our order of development of the basic results follows [GZ, Ch III], but the methods of proof are usually quite different, making much less use of the "numerology" of modular curves.

Here is a summary of the contents. In §2 we consider some background issues related to maps among elliptic curves over various bases and horizontal divisors on relative curves over a discrete valuation ring. In §3 we provide a brief survey of the Serre-Tate theorem and the Grothendieck existence theorem, since these form the backbone of the deformation-theoretic methods which underlie all subsequent arguments. Due to reasons of space, some topics (such as intersection theory on arithmetic curves, Gross' paper [Gr1] on quasi-canonical liftings, and $p$-divisible groups) are not reviewed but are used freely where needed.

In §4 we compute an elementary intersection number on a modular curve in terms of cardinalities of isomorphism groups between infinitesimal deformations. This serves as both a warm-up to and key ingredient in §5–§6, where we use cardinalities of Hom groups between infinitesimal deformations to give a formula (in Theorem 5.1) for a local intersection number $(x.T_m(x^\sigma))_v$, where $x \in X_0(N)(H)$ is a Heegner point with CM by the ring of integers of an imaginary quadratic field $K$ (with Hilbert class field $H$), $\sigma \in \mathrm{Gal}(H/K)$ is an element which corresponds to an ideal class $\mathscr{A}$ in $K$ under the Artin isomorphism, and $T_m$ is a Hecke

*Date*: February 11, 2003.

This work was partially supported by the NSF and the Alfred P. Sloan Foundation. I would like to thank B. Gross and J-P. Serre for helpful suggestions, and Mike Roth for helping to resolve some serious confusion.

correspondence with $m \geq 1$ relatively prime to $N$. An essential hypothesis in Theorem 5.1 is the vanishing of the number $r_{\mathscr{A}}(m)$ of integral ideals of norm $m$ in the ideal class $\mathscr{A}$. This corresponds to the requirement that the divisors $x$ and $T_m(x^\sigma)$ on $X_0(N)$ are disjoint. Retaining the assumption $r_{\mathscr{A}}(m) = 0$, in §7 we develop and apply a construction of Serre in order to translate the formula in Theorem 5.1 into the language of quaternion algebras. The resulting quaternionic formulas in Corollary 7.15 are a model for the local intersection number calculation which is required for the computation of global height pairings in [GZ], except the condition $r_{\mathscr{A}}(m) = 0$ in Corollary 7.15 has to be dropped.

To avoid assuming $r_{\mathscr{A}}(m) = 0$, we have to confront the case of divisors which may contain components in common. Recall that global height pairings of degree 0 divisors are defined in terms of local pairings, using a "moving lemma" to reduce to the case in which horizontal divisors intersect properly. We want an explicit formula for the global pairing $\langle c, T_m(d^\sigma)\rangle_H$ where $c = x - \infty$ and $d = x - 0$, so we must avoid the abstract moving lemma and must work directly with improper intersections. Whereas [GZ] deal with this issue by using a technique from [Gr2, §5] which might be called "intersection theory with a tangent vector", in §8 we develop a more systematic method which we call "intersection theory with a meromorphic tensor". This theory is applied in §9 to give a formula in Theorem 9.6 which expresses a global height pairing in terms of local intersection numbers whose definition does not require proper intersections.

In §10 we use deformation theory to generalize Corollary 7.15 to include the case $r_{\mathscr{A}}(m) > 0$, and Appendix A (by W.R. Mann) recovers all three main formulas in [GZ, Ch III, §9] as consequences of the quaternionic formulas we obtain via intersection theory with meromorphic tensors. The appendix follows the argument of Gross-Zagier quite closely, but explains some background, elaborates on some points in more detail than in [GZ, Ch III, §9], and works uniformly across all negative fundamental discriminants without parity restrictions. This parity issue is the main technical contribution of the appendix, and simply requires being a bit careful.

Some conventions. As in [GZ], we normalize the Artin map of class field theory to associate uniformizers to arithmetic Frobenius elements. Thus, if $K$ is an imaginary quadratic field with Hilbert class field $H$, then for a prime ideal $\mathfrak{p}$ of $K$ the isomorphism between $\mathrm{Gal}(H/K)$ and the class group $\mathrm{Cl}_K$ of $K$ associates the ideal class $[\mathfrak{p}]$ to an arithmetic Frobenius element.

Following [GZ], we only consider Heegner points with CM by the maximal order $\mathscr{O}_K$ in an imaginary quadratic field $K$. A *Heegner diagram* (for $\mathscr{O}_K$) over an $\mathscr{O}_K$-scheme $S$ is an $\mathscr{O}_K$-linear isogeny $\phi : E \to E'$ between elliptic curves over $S$ which are equipped with $\mathscr{O}_K$-action which is "normalized" in the sense that the induced action on the tangent space at the identity is the same as that obtained through $\mathscr{O}_S$ being a sheaf of $\mathscr{O}_K$-algebras via the $\mathscr{O}_K$-scheme structure on $S$. For example, when we speak of Heegner diagrams (or Heegner points on modular curves) over $\mathbf{C}$, it is implicitly understood that an embedding $K \hookrightarrow \mathbf{C}$ has been fixed for all time.

Under the action of $\mathrm{Cl}_K \simeq \mathrm{Gal}(H/K)$ on Heegner points $X_0(N)(H) \subseteq X_0(N)(\mathbf{C})$, with $N > 1$ having all prime factors split in $K$, the action of $[\mathfrak{a}] \in \mathrm{Cl}_K$ sends the Heegner point $([\mathfrak{b}], \mathfrak{n}) \stackrel{\mathrm{def}}{=} (\mathbf{C}/\mathfrak{b} \to \mathbf{C}/\mathfrak{n}^{-1}\mathfrak{b})$ to $([\mathfrak{b}][\mathfrak{a}]^{-1}, \mathfrak{n})$. The appearance of inversion is due to our decision to send uniformizers to arithmetic (rather than geometric) Frobenius elements. This analytic description of the Galois action on Heegner points will play a crucial role in the proof of Corollary 7.11.

If $x$ is a Heegner point with associated CM field $K$, we will write $u_x$ to denote the cardinality of the group of roots of unity in $K$ (so $u_x = 2$ unless $K = \mathbf{Q}(\sqrt{-1})$ or $K = \mathbf{Q}(\sqrt{-3})$).

If $S$ is a finite set, we will write either $|S|$ or $\#S$ to denote the cardinality of $S$.

If $(R, \mathfrak{m})$ is a local ring, we write $R_n$ to denote $R/\mathfrak{m}^{n+1}$. If $X$ and $Y$ are $R$-schemes, we write $\mathrm{Hom}_{R_n}(X, Y)$ to denote the set of morphisms of mod $\mathfrak{m}^{n+1}$ fibers.

If $B$ is a central simple algebra of finite dimension over a field $F$, we write $\mathrm{N} : B \to F$ and $\mathrm{T} : B \to F$ to denote the reduced norm and reduced trace on $B$.

It is recommended (but not necessary) that anyone reading these notes should have a copy of [GZ] at hand. One piece of notation we adopt, following [GZ], is that $r_{\mathscr{A}}(m)$ denotes the number of integral ideals of norm $m > 0$ in an ideal class $\mathscr{A}$ of an imaginary quadratic field $K$ which is fixed throughout the discussion. Since $r_{\mathscr{A}} = r_{\mathscr{A}^{-1}}$ due to complex conjugation inducing inversion on the class group without chaging norms, if we change the Artin isomorphism $\mathrm{Gal}(H/K) \simeq \mathrm{Cl}_K$ by a sign then this has no impact on a formula for $\langle c, T_m d^\sigma \rangle_v$ in terms of $r_{\mathscr{A}}$, where $\mathscr{A} \in \mathrm{Cl}_K$ "corresponds" to $\sigma \in \mathrm{Gal}(H/K)$.

## 2. Some properties of abelian schemes and modular curves

We begin by discussing some general facts about elliptic curves. Since the proofs for elliptic curves are the same as for abelian varieties, and more specifically it is not enough for us to work with passage between a number field and $\mathbf{C}$ (e.g., we need to work over discrete valuation rings with positive characteristic residue field, etc.) we state the basic theorem in the more natural setting of abelian varieties and abelian schemes.

**Theorem 2.1.** *Let $A, B$ be abelian varieties over a field $F$.*

(1) *If $F$ is separably closed and $F'$ is an extension field, then $\mathrm{Hom}_F(A, B) \to \mathrm{Hom}_{F'}(A_{/F'}, B_{/F'})$ is an isomorphism. In other words, abelian varieties over a separably closed field never acquire any "new" morphisms over an extension field.*

(2) *If $F = \mathrm{Frac}(R)$ for a discrete valuation ring $R$ and $A$, $B$ have Néron models $\mathscr{A}$, $\mathscr{B}$ over $R$ which are proper (i.e., $A$ and $B$ have good reduction relative to $R$), then*

$$\mathrm{Hom}_F(A, B) = \mathrm{Hom}_R(\mathscr{A}, \mathscr{B}) \to \mathrm{Hom}_k(\mathscr{A}_0, \mathscr{B}_0)$$

*is injective, where $k$ is the residue field of $R$ and $(\cdot)_0$ denotes the "closed fiber" functor on $R$-schemes. In other words, $(\cdot)_0$ is a faithful functor from abelian schemes over $R$ to abelian schemes over $k$. In fact, this latter faithfulness statement holds for abelian schemes over* any *local ring $R$ whatsoever.*

*Proof.* The technical details are a bit of a digression from the main aims of this paper, so although we do not know a reference we do not give the details. Instead, we mention the basic idea: for $\ell$ invertible on the base, $\ell$-power torsion is "relatively schematically dense" in an abelian scheme (in the sense of [EGA, IV$_3$, 11.10]); this ultimately comes down to the classical fact that such torsion is dense on an abelian variety over an algebraically closed field. Such denseness, together with the fact that $\ell^n$-torsion is finite étale over the base, provides enough rigidity to descend morphisms for the first part of the theorem, and enough restrictiveness to force injectivity in the second part of the theorem. ■

We will later need the faithfulness of $(\cdot)_0$ in Theorem 2.1(2) for cases in which $R$ is an artin local ring, so it is not adequate to work over discrete valuation rings and fields.

Let's now recall the basic setup in [GZ]. We have fixed an imaginary quadratic field $K \subseteq \mathbf{C}$ with discriminant $D < 0$ (and we take $\overline{\mathbf{Q}}$ to be the algebraic closure of $\mathbf{Q}$ inside of $\mathbf{C}$). We write $H \subseteq \overline{\mathbf{Q}}$ for the Hilbert class field of $K$, and we choose a positive integer $N > 1$ which is relatively prime to $D$ and for which all prime factors of $N$ are split in $K$. Let $X = X_0(N)_{/\mathbf{Z}}$ be the coarse moduli scheme as in [KM], so $X$ is a proper flat curve over $\mathbf{Z}$ which is smooth over $\mathbf{Z}[1/N]$ but has some rather complicated fibers modulo prime factors of $N$. We have no need for the assumption that $D < 0$ is odd (i.e, $D \equiv 1 \bmod 4$), whose main purpose in [GZ] is to simplify certain aspects of calculations on the analytic side of the [GZ] paper, so we avoid such conditions (and hence include cases with even discriminant).

Let $x \in X(H) \subseteq X(\mathbf{C})$ be a Heegner point with CM by the maximal order $\mathscr{O}_K$. Because of our explicit knowledge of the action of $\mathrm{Gal}(H/\mathbf{Q})$ on Heegner points, we see that the action of $\mathrm{Gal}(H/\mathbf{Q})$ on $x \in X(H)$ is "free" (i.e., non-trivial elements in $\mathrm{Gal}(H/\mathbf{Q})$ do not fix our Heegner point $x \in X(H)$), so it follows that the map $x : \mathrm{Spec}(H) \to X_{/\mathbf{Q}}$ is a closed immersion. For each closed point $v$ of $\mathrm{Spec}\, \mathscr{O}_H$, by viewing $H$ as the fraction field of the algebraic localization $\mathscr{O}_{H,v}$ we may use the valuative criterion for properness to uniquely extend $x$ to a point in $X(\mathscr{O}_{H,v})$. A simple "smearing out" argument involving denominator-chasing shows that all of these maps arise from a unique morphism $\underline{x} : \mathrm{Spec}(\mathscr{O}_H) \to X$.

It is *not* a priori clear if the map $\underline{x}$ is a closed immersion (in general it isn't). More specifically, if $Z_x \hookrightarrow X$ is the scheme-theoretic closure of $x$ (i.e., the scheme-theoretic image of $\underline{x}$), then $Z_x \to \mathrm{Spec}(\mathbf{Z})$ is proper, flat, and quasi-finite, hence finite flat, so it has the form $Z_x = \mathrm{Spec}(A_x)$ with $\mathbf{Q} \otimes_{\mathbf{Z}} A_x = H$. Thus, $A_x$ is an order in $\mathscr{O}_H$, but it isn't obvious if $A_x = \mathscr{O}_H$. This is an important issue in subsequent intersection theory calculations because the intersection theory will involve closed subschemes of (various base chages on) $X$. For this reason, we must distinguish $\mathrm{Spec}\, A_x \hookrightarrow X$ and $\underline{x} : \mathrm{Spec}\, \mathscr{O}_H \to X$.

To illustrate what can go wrong, consider the map $\mathbf{Z}_p[T] \to \mathbf{Z}_p[\zeta_{p^2}]$ defined by $T \mapsto p\zeta_{p^2}$. Over $\mathbf{Q}_p$ this defines an immersion $\phi_\eta : \mathrm{Spec}(\mathbf{Q}_p(\zeta_{p^2})) \hookrightarrow \mathbf{A}^1_{\mathbf{Z}_p} \subseteq \mathbf{P}^1_{\mathbf{Z}_p}$ which is a closed immersion since $\mathbf{Q}_p(\zeta_{p^2}) = \mathbf{Q}_p[p\zeta_{p^2}]$, and $\phi_\eta$ comes from a map $\phi : \mathrm{Spec}(\mathbf{Z}_p[\zeta_{p^2}]) \to \mathbf{P}^1_{\mathbf{Z}_p}$ which is necessarily the one we would get from applying the valuative criterion for properness to $\phi_\eta$.

The "integral model" map $\phi$ is *not* a closed immersion because $\mathbf{Z}_p[T] \to \mathbf{Z}_p[\zeta_{p^2}]$ defined by $T \mapsto p\zeta_{p^2}$ is not a surjection. In fact, the generic fiber closed subscheme of $\mathbf{P}^1_{\mathbf{Q}_p}$ defined by $\phi_\eta$ has scheme-theoretic closure in $\mathbf{P}^1_{\mathbf{Z}_p}$ given by $\mathrm{Spec}(A)$ where $A = \mathrm{image}(\phi) = \mathbf{Z}_p + p\mathbf{Z}_p[\zeta_{p^2}]$ is a non-maximal order in $\mathbf{Z}_p[\zeta_{p^2}]$. The moral is that even if we can compute the "field of definition" of a point on a generic fiber smooth curve, it is not true that the closure of this in a particular proper integral model of the curve is cut out by a Dedekind subscheme or that it lies in the relative smooth locus. It is a very fortunate fact in the Gross-Zagier situation that this difficulty usually does not arise for the "horizontal divisors" they need to consider.

Here is the basic result we need concerning Heegner divisors in the relative smooth locus.

**Lemma 2.2.** *Let $x \in X(H)$ be a Heegner point and $\Lambda_v$ denote the valuation ring of the completion $H_v$ at a place $v$ over a prime $p$. The map $\mathrm{Spec}(\Lambda_v) \to X_{/\mathbf{Z}_p}$ corresponding to the pullback $x_v \in X(H_v)$ of $x$ factors through the relative smooth locus, and the induced natural map $\underline{x}_v : \mathrm{Spec}(\Lambda_v) \to X \times_{\mathbf{Z}} \Lambda_v$ arising from $x_v$ is a closed immersion into the relative smooth locus over $\Lambda_v$ (and hence lies in the regular locus).*

*Proof.* Using smoothness of $X$ over $\mathbf{Z}[1/N]$ when $p$ doesn't divide $N$ and [GZ, Ch III, Prop 3.1] when $p|N$, we see that the image of the closed point under $\underline{x}_v$ lies in the smooth locus over $\Lambda_v$. Since $\underline{x}_v$ is a section to a separated map, it is a closed immersion and necessarily lands inside of the relative smooth locus over $\Lambda_v$. ∎

*Remark* 2.3. One reason for the importance of Lemma 2.2 is that the curve $X_v = X \times_{\mathbf{Z}} \Lambda_v$ is usually *not* regular, so to do intersection theory one must resolve singularities. A minimal regular resolution $X^{\mathrm{reg}} = X^{\mathrm{reg}}_{/\mathbf{Z}_{(p)}}$ can be obtained by means of successive normalizations and blow-ups over the non-regular locus (thanks to a deep theorem of Lipman [L]), so the resolution process doesn't do anything over the relative smooth locus.

Thus, even though we can't expect to do intersection theory on $X_v$, we are assured by Lemma 2.2 that the intersection numbers among properly intersecting Heegner divisors will be computable directly on $X_v$. We need to keep track of smoothness because regularity is often destroyed by ramified base change, such as $\mathbf{Z}_{(p)} \to \mathscr{O}_{H,v}$ when $p$ is ramified in $K$ (in such cases the relative curve $X^{\mathrm{reg}} \times_{\mathbf{Z}_{(p)}} \Lambda_v$ can be non-regular, but some Heegner divisor intersection numbers may still be computed on this curves, such as happens in [GZ, Ch III, Prop 3.3]).

With $v$ fixed as above, let $W$ denote the completion of a maximal unramified extension of $\Lambda_v$ (if $p$ ramifies in $K$, this is not a Witt ring). The section $\underline{x}_v \in X_{/W}(W)$ lies in the relative smooth locus. Although $X_{/W}$ is not a fine moduli scheme, the point $\underline{x}_v$ does arise from a $\Gamma_0(N)$-structure over $W$. This is an important point. Before we prove it, as a preliminary step recall that from the classical CM theory we can find an algebraic model $\phi : E \to E'$ over $H$ which represents our Heegner point in $X(H)$. It is *not* true in general that $E$ and $E'$ have to admit everywhere good reduction over $\mathscr{O}_H$. But what is true, and suffices for us, is that we can always find such good models over $W$. This is given by Theorem 2.5, but before proving this we recall a general lemma concerning good reduction for CM elliptic curves.

**Lemma 2.4.** *Let $R$ be a complete discrete valuation ring with fraction field $F$ of characteristic $0$, and let $E_{/F}$ be an CM elliptic curve with CM field $K \subseteq F$ (so the CM-action is automatically defined over $F$). If $K = \mathbf{Q}(\sqrt{-1})$ or $K = \mathbf{Q}(\sqrt{-3})$, assume moreover that $E_{/F}$ actually has CM by the full ring of integers of $K$. Then there exists a "twist" $E'$ of $E$ over an unramified extension of $F$ (so $E'$ is CM with the same CM ring) such that $E'$ has good reduction over $R$.*

When $E$ begins life over a number field (which is the only case we really need for treating Heegner points), this lemma follows from Corollary 2 to Theorem 9 in [ST]. Since the argument in [ST] rests on class field theory, we prefer the argument below which uses only general principles concerning abelian varieties. The argument in [ST] has the merit of treating more cases when $K = \mathbf{Q}(\sqrt{-1})$ and $K = \mathbf{Q}(\sqrt{-3})$.

*Proof.* Since $\mathrm{End}(E_{/F}) = \mathscr{O}_K$, so $\mathrm{End}(E_{/F_s}) = \mathscr{O}_K$ and hence $\mathrm{Aut}(E_{/F_s}) = \mathscr{O}_K^\times = \mathrm{Aut}(E_{/F})$, twisted forms are described in terms of Galois cohomology $\mathrm{H}^1(\mathrm{Gal}(F_s/F), \mathscr{O}_K^\times)$ with $\mathscr{O}_K^\times$ given a *trivial* action by $\mathrm{Gal}(F_s/F)$. Passing to completions on a henselian discrete valuation ring doesn't affect the Galois group in characteristic 0, so without loss of generality we can assume that $R$ has a separably closed residue field $k$ (i.e., pass to the completion of the maximal unramified extension). Now we seek to find a twist $E'$ of $E$ over $F$ with good reduction over $R$.

A priori $E$ has potentially good reduction (as does any CM type abelian variety over $F$), so there exists a finite Galois extension $F'/F$ such that $E_{/F'}$ has good reduction. Let $R'$ be the integral closure of $R$ in $F'$ and let $\mathscr{E}_{/R'}$ denote the Néron model of $E_{/F'}$. Note in particular that the residue field extension $k'/k$ is purely inseparable (since $k$ is separably closed), so $\mathrm{Aut}(k'/k) = \{1\}$. Let $\Gamma = \mathrm{Gal}(F'/F)$, so for each $\sigma \in \Gamma$ there is (by the Néronian property of abelian schemes over $R$) a unique isomorphism $\phi_\sigma : \mathscr{E} \simeq \sigma^*(\mathscr{E})$ of elliptic curves over $R'$ which extends the evident descent data isomorphism $E_{/F'} \simeq \sigma^*(E_{/F'})$ of elliptic curves over $F'$. In particular, we have the cocycle property $\sigma^*(\phi_\tau) \circ \phi_\sigma = \phi_{\sigma\tau}$ and also $\phi_\sigma$ commutes with the CM actions by $K$ on source and target, as this can all be checked on the generic fibers.

Since $\sigma$ induces the identity on $k'$ (as $\mathrm{Aut}(k'/k) = \{1\}$!), on the closed fiber we therefore get an *automorphism* $\overline{\phi}_\sigma$ of $\mathscr{E}_{/k'}$ which defines an anti-homomorphism $\rho : \Gamma \to \mathrm{Aut}_{k'}(\mathscr{E}_{/k'})$ landing in the commutator of the $K$-action. But an imaginary quadratic field acting (in the isogeny category) on an elliptic curve is necessarily its own commutator in the endomorphism ring of the elliptic curve (even in the supersingular case!), so we deduce that $\rho(\Gamma) \subseteq \mathscr{O}_K^\times$ and in particular $\rho$ is actually a homomorphism since $\mathscr{O}_K^\times$ is abelian. Let $\chi(\sigma) \in \mathscr{O}_K^\times$ be the *unique* (cf. Theorem 2.1(2)) unit inducing the same action as $\rho(\sigma)$ on the closed fiber of $\mathscr{E}_{/R'}$ (recall our initial assumption that when $\mathscr{O}_K^\times$ is larger than $\{\pm 1\}$, then the whole ring $\mathscr{O}_K$ acts on $E$ and hence on the Néron model of $E_{/F'}$). Clearly $\chi$ is a homomorphism.

For each $\sigma \in \Gamma$ we see that $\psi(\sigma) \stackrel{\mathrm{def}}{=} \phi_\sigma \circ \chi(\sigma)^{-1} : \mathscr{E}_{/R'} \simeq \sigma^*(\mathscr{E}_{/R'})$ makes sense and induces the identity on the closed fiber. Since passage to the closed fiber is a faithful functor for abelian schemes (see Theorem 2.1(2)), we conclude that $\sigma \mapsto \psi(\sigma)$ satisfies the Galois cocycle condition on the generic fiber over $F'$ (indeed, to check this condition we may work over $R'$, and even on closed fibers over $R'$ by faithfulness). Thus, by Galois descent we obtain a twist $E'$ of $E$ over $F$ and an isomorphism $E'_{/F'} \simeq E_{/F'}$ of elliptic curves which carries the canonical descent data on $E'_{/F'}$ over to the "twisted" descent data $\psi$.

I claim that $E'$ has good reduction over $F$. To prove this, by Néron-Ogg-Shafarevich it suffices to pick a prime $\ell$ distinct from the residue characteristic of $R$ and to show that the $\Gamma$-action on the $\ell$-adic Tate module of $E'_{/F}$ is trivial (note that the Galois action on this Tate module certainly factors through $\Gamma$, since at least over $F'$ we know we have good reduction for $E'_{/F'} \simeq E_{/F'}$). But this triviality is equivalent to the natural action of $\Gamma$ on the closed fiber of the Néron model of $E'$ over $R'$ being the identity. This in turn follows from how $E'$ was constructed. ∎

Recall that $W$ denotes the completion of the maximal unramified extension of the algebraic localization $\mathscr{O}_{H,v}$ of $\mathscr{O}_H$ at the place $v$ over $p$.

**Theorem 2.5.** *In the above notation, the point in $X(W)$ arising from a Heegner point in $X(H)$ is necessarily induced by a Heegner diagram over $W$.*

It seems likely that Theorem 2.5 is false if we try to replace $W$ with $\mathscr{O}_{H,v}$ (or its completion $\Lambda_v$); we almost surely have to pass to some extension of $H$ unramified over $v$.

Due to the properness (even finiteness) of the fine moduli scheme of Drinfeld $\Gamma_0(N)$-structures on a given elliptic curve over a base, it follows that a diagram over $W$ as in Theorem 2.5 is automatically a $\Gamma_0(N)$-structure (i.e., a cyclic $N$-isogeny) since this is true on the generic fiber of characteristic 0.

*Proof.* Let's begin with a Heegner diagram $\phi : E \to E'$ over $H$ which induces the given point in $X(H)$. It is generally *not* true that this necessarily admits good reduction over $W$ (e.g., make a ramified quadratic twist), but Lemma 2.4 (whose additional hypothesis for $K = \mathbf{Q}(\sqrt{-1})$ or $K = \mathbf{Q}(\sqrt{-3})$ is satisfied because of our general assumption of CM by the maximal order for our Heegner points) shows that over the fraction field $F$ of $W$ there exists a twist of $E$ with good reduction. More specifically, such a twist is given by an element $\chi \in \mathrm{H}^1(\mathrm{Gal}(F_s/F), \mathscr{O}_K^\times) = \mathrm{Hom}_{\mathrm{cont}}(\mathrm{Gal}(F_s/F), \mathscr{O}_K^\times)$. Since $\phi$ is $\mathscr{O}_K$-equivariant, we can use the same Galois character $\chi$ to twist $E'$, and the resulting $\mathscr{O}_K^\times$-valued twisted descent data automatically has to commute with the $\mathscr{O}_K$-compatible $\phi$.

Consequently, we can twist the entire given Heegner diagram relative to $\chi$ and thereby obtain a Heegner diagram over $F$ for which the two elliptic curves have good reduction over $W$ and the induced geometric point in $X(F_s)$ coincides with the original Heegner point. Using Néron functoriality we can extend the twisted $\phi$ over $W$ and thereby get a Heegner diagram over $W$ inducing the point in $X(W)$ arising from our given Heegner point in $X(H)$ (recall that $X(W) \to X(F)$ is bijective).

■

In all that follows, we will fix a Heegner diagram $\underline{x}$ over $W$ inducing a given Heegner point in $X(W)$ (or rather, inducing the $W$-section of $X_{/W}$ arising from an initial choice of Heegner point in $X(H)$). The existence of such a diagram follows from Theorem 2.5. It is important to note that the data of $\underline{x}$ is *determined* up to non-unique isomorphism by the corresponding section in $X(W)$, so this section will also be denoted $\underline{x}$. This uniqueness up to isomorphism follows from the following result, applied to $R = W$:

**Theorem 2.6.** *Let $R$ be a complete discrete valuation ring with separably closed residue field $k$ and fraction field $L$. Let $A_1$ and $A_2$ be two abelian schemes over $R$. Then for any field extension $L'$ of $L$, the natural map $\mathrm{Hom}_R(A_1, A_2) = \mathrm{Hom}_L(A_{1/L}, A_{2/L}) \to \mathrm{Hom}_{L'}(A_{1/L'}, A_{2/L'})$ is an isomorphism.*

*In other words, there are no "new" geometric morphisms between $A_1$ and $A_2$ over an extension of $L$ which don't already show up over $L$ (or equivalently, over $R$).*

*Proof.* By the first part of Theorem 2.1, we are reduced to the case in which $L'$ is a separable closure of $L$, and by direct limit considerations we can even assume $L'$ is a finite separable extension of $L$ which we may moreover suppose to be Galois. Let $\Gamma = \mathrm{Gal}(L'/L)$. By Galois descent, it suffices to show that $\sigma^*(f) = f$ for any $f : A_{1/L'} \to A_{2/L'}$ and any $\sigma \in \Gamma$. Let $A'_j$ denote the Néron model of $A_{j/L'}$ over the integral closure $R'$ of $R$ in $L'$ (so $R'$ is a Dedekind domain which is finite over $R$), so $A'_j = A_j \times_R R'$. Let $F : A'_1 \to A'_2$ denote the map induced by $f$ on Néron models, so it suffices to prove $\sigma^*(F) = F$ for any $\sigma \in \Gamma$. The crucial role of the hypothesis that $R$ is complete is that it ensures $R'$ is again *local* (i.e., a discrete valuation ring).

By the second part of Theorem 2.1, it suffices to check equality on the closed fiber over $R'$. But since $R'$ has residue field which is purely inseparable over $R$ (so $\sigma$ reduces to the idenity!), it is immediate that $\sigma^*(F)$ and $F$ coincide on closed fibers over $R'$.

■

Throughout all that follows, the prime $p$ will be fixed and we will write $F$ to denote the fraction field of $W$. We will also write $\underline{X}$ to denote $X_{/W}$, and $\pi$ to denote a uniformizer of $W$.

We now record an important immediate consequence of the above considerations (upon recalling how correspondences act on rational points, via pushforward and pullback).

**Corollary 2.7.** *Fix a positive integer $m$ relatively prime to $Np$. Consider the Hecke divisor $T_m(\underline{x}_{/F})$ in $X_{/F}$. Its scheme-theoretic closure in $X_{/W}$ is a sum of sections lying inside of the relative smooth locus.*

By "sum of sections" we mean in the sense of relative effective Cartier divisors, which is to say that we take products of ideal sheaves (all of which are invertible, thanks to the sections lying in the relative smooth locus). It is immediate from the corollary, that the closed fiber of this closure is computed exactly by the habitual "Hecke correspondence" formula on the level of geometric points. Thus, this scheme-theoretic closure may be denoted $T_m(\underline{x})$ without risk of confusion.

*Proof.* We have to prove that all points in $T_m(\underline{x}_{/F})$ are $F$-rational with corresponding $W$-point in the *smooth* locus of $\underline{X}$ over $W$. It is here that one uses that $m$ is prime to $Np$, the crux of the matter being that all prime-to-$p$ torsion on elliptic curves over $W$ is finite étale and hence constant (since $W$ has separably closed residue field) and all prime-to-$N$ isogenies induce isomorphisms on level $N$-structures over any base. Thus, all level structures of interest can be defined over $W$ and we just have to show that if $\underline{z}$ is a $\Gamma_0(N)$-structure over $W$ which is $m$-isogenous to $\underline{x}$, then the resulting section $\mathrm{Spec}(W) \hookrightarrow \underline{X}$ lands in the smooth locus.

If $p \nmid N$ then $\underline{X}$ is $W$-smooth, so we're done. If $p|N$, then by [GZ, Ch III, Prop 3.1] the closed point $\underline{x}_0 \in \underline{X}(W/\pi)$ is an ordinary point in either the $(0,n)$ or $(n,0)$ component of the closed fiber $\underline{X}_0$ (where $n = \mathrm{ord}_p(N) \geq 1$). Since $\underline{z}_0$ is $m$-isogenous to $\underline{x}_0$ with $\gcd(m,N) = 1$, we see that the $\Gamma_0(N)$-structure $\underline{z}_0$ is another ordinary point on the same component as $\underline{x}_0$ and in particular is a smooth point. Thus, $\underline{z}$ lies in the relative smooth locus.

■

The analogue of this corollary when $p|m$ is much more subtle, and will be treated in §6.

## 3. The Serre-Tate theorem and the Grothendieck existence theorem

The main technical problem in the arithmetic intersection component of [GZ] is to compute various intersection numbers by means of counting morphisms between deformations of elliptic curves. We will have two elliptic curves $E$ and $E'$ over a complete local noetherian ring $(R, \mathfrak{m})$ (often artin local or a discrete valuation ring) and will want to count the size of the finite set $H_n = \mathrm{Hom}_{R_n}(E_n, E'_n)$, where $(\cdot)_n$ denotes reduction modulo $\mathfrak{m}^{n+1}$. By the last line of Theorem 2.1(2), the natural "reduction" map $H_{n+1} \to H_n$ (corresponding to base change by $R_{n+1} \twoheadrightarrow R_n$) is injective for every $n \geq 0$, so we can view the $H_n$'s as a decreasing sequence of subgroups of $H_0 = \mathrm{Hom}_k(E_0, E'_0)$ (where $k = R/\mathfrak{m}$). When $k$ has positive characteristic $p$, the Serre-Tate theorem (to be stated in Theorem 3.3) will "compute" these Hom groups $H_n$ between elliptic curves in terms of Hom groups between $p$-divisible groups. In fact, the Serre-Tate theorem does much more: it essentially identifies the deformation theory of an elliptic curve with that of its $p$-divisible group. It will turn out that $p$-divisible groups are more tractable for our counting purposes, by means of the theory of formal groups (a theory which is closely connected to that of $p$-divisible groups in the situations of interest).

Once we have a way to work with the $H_n$'s by means of $p$-divisible groups via the Serre-Tate theorem, it will still be important to understand one further issue: what is the intersection of the $H_n$'s inside of $H_0$? For example, the faithfulness at the end of Theorem 2.1(2) yields a natural inclusion

$$\mathrm{Hom}_R(E, E') \subseteq \bigcap H_n \simeq \varprojlim H_n$$

inside of $H_0$. Using a special case of the Grothendieck existence theorem, to be recorded in Theorem 3.4, this inclusion is an equality.

Let us now set forth the general context in which the Serre-Tate theorem takes place. Since the specificity of elliptic curves leads to no essential simplifications on the argument which is used for abelian schemes, we will work in the more general setting of abelian schemes (although only the case of elliptic curves intervenes in [GZ]). We recommend [Mum] as a basic reference for abelian varieties and [GIT, Ch 6] as a basic reference for abelian schemes. Let $S$ be a scheme, and $A, A'$ two abelian schemes over $S$. Assume that a prime number $p$ is locally nilpotent on $S$ (i.e., every point of $S$ has residue field of characteristic $p$). The case of most interest will be $S = \mathrm{Spec}(W/\pi^{n+1})$ where $W$ is either a ring of Witt vectors or some finite discrete valuation ring extension thereof (with uniformizer $\pi$), though more general base rings are certainly needed for deformation theory arguments.

Let $\Gamma = A[p^\infty]$ and $\Gamma' = A'[p^\infty]$ denote the $p$-divisible groups arising from the $p$-power torsion schemes on $A$ and $A'$ over $S$. The problem considered by Serre and Tate was that of lifting morphisms of abelian schemes through an infinitesimal thickening of the base. More specifically, let $S_0 \hookrightarrow S$ be a closed immersion with defining ideal sheaf $\mathscr{I}$ satisfying $\mathscr{I}^N = 0$ for some positive integer $N$. Consider the problem of lifting an element in $\mathrm{Hom}_{S_0}(A_0, A'_0)$ to an element in $\mathrm{Hom}_S(A, A')$. Let's first record that this lifting problem has a unique solution if it has any at all, and that a similar uniqueness holds for $p$-divisible groups (not necessarily coming from abelian schemes):

**Lemma 3.1.** *Let $S_0 \hookrightarrow S$ be as above. Let $A$ and $A'$ be arbitrary abelian schemes over $S$, and let $\Gamma$ and $\Gamma'$ be arbitrary $p$-divisible groups over $S$. Then the natural maps $\mathrm{Hom}_S(A, A') \to \mathrm{Hom}_{S_0}(A_0, A'_0)$ and $\mathrm{Hom}_S(\Gamma, \Gamma') \to \mathrm{Hom}_{S_0}(\Gamma_0, \Gamma'_0)$ are injective.*

*Proof.* In both cases we may assume $S$ is local. By the last line of Theorem 2.1(2), base change to the residue field is a faithful functor on abelian schemes over a local ring, so the case of abelian schemes is settled by means of base change to the residue field. For $p$-divisible groups, use [K, Lemma 1.1.3(2)] (which also handles the abelian scheme case). ∎

As was mentioned above, one aspect of the Serre-Tate theorem is that it identifies solutions to the infinitesimal lifting problem for morphisms of abelian schemes with solutions to the analogous problem for their $p$-divisible groups. In order to put ourselves in the right frame of mind, we first record the fact that abelian scheme morphisms can be safely viewed as morphisms between $p$-divisible groups:

**Lemma 3.2.** *If $A$ and $A'$ are abelian schemes over a base $S$, and $\Gamma$ and $\Gamma'$ are the associated $p$-divisible groups for a prime number $p$, then the natural map $\mathrm{Hom}_S(A, A') \to \mathrm{Hom}_S(\Gamma, \Gamma')$ is injective.*

Note that this lemma does not impose any requirements on residue characteristics at points in $S$. We call a map $\Gamma \to \Gamma'$ *algebraic* if it arises from a (necessarily unique) map $A \to A'$.

*Proof.* We can assume $S$ is local, and by the last line in Theorem 2.1(2) we can even assume $S = \mathrm{Spec}(k)$ for a field $k$ which we may assume is algebraically closed. If $p$ is distinct from the characteristic of $k$, then the injectivity can be proven by a relative schematic density argument (this is needed in the proof of Theorem 2.1, and uses results from [EGA, $\mathrm{IV}_3$, 11.10]). If $p$ coincides with the characteristic of $k$, then under the equivalence of categories between connected $p$-divisible groups over $k$ and commutative formal groups of finite height over $k$, the connected component of the $p$-divisible group of an abelian variety over $k$ is identified with the formal group of the abelian variety. Thus, if $f, g : A \to A'$ satisfy $f[p^\infty] = g[p^\infty]$ then $f$ and $g$ induce the same maps $\widehat{\mathscr{O}}_{A',0} \to \widehat{\mathscr{O}}_{A,0}$ on formal groups at the origin. Thus, $f = g$. ∎

As an example, if $E$ is an elliptic curve over $\mathbf{Q}_p$ with good reduction over $\mathbf{Z}_p$ and $\mathscr{E}$ is the Néron model of $E$, then $\mathrm{End}_{\mathbf{Q}_p}(E) = \mathrm{End}_{\mathbf{Z}_p}(\mathscr{E}) \subseteq \mathrm{End}_{\mathbf{Z}_p}(\mathscr{E}[p^\infty])$. We want to describe the image of this inclusion. More generally, suppose we are given a map $f : \Gamma \to \Gamma'$ between the $p$-divisible groups of two abelian schemes $A$ and $A'$ over $S$, and assume its reduction $f_0$ over $S_0$ is algebraic (in the sense that it comes from a map $A_0 \to A'_0$ which is then necessarily unique). The theorem of Serre and Tate asserts that $f$ is also algebraic (and more):

**Theorem 3.3.** (Serre-Tate) *The natural commutative square* (*with injective arrows*)

$$\begin{array}{ccc} \mathrm{Hom}_S(A, A') & \longrightarrow & \mathrm{Hom}_S(\Gamma, \Gamma') \\ \downarrow & & \downarrow \\ \mathrm{Hom}_{S_0}(A_0, A'_0) & \longrightarrow & \mathrm{Hom}_{S_0}(\Gamma_0, \Gamma'_0) \end{array}$$

*is cartesian. In other words, a map $f_0 : A_0 \to A'_0$ lifts to a map $f : A \to A'$ if and only if the induced map $f_0[p^\infty] : \Gamma_0 \to \Gamma'_0$ lifts to a map $h : \Gamma \to \Gamma'$, in which case $f$ and $h$ are unique and $f[p^\infty] = h$.*

*Moreover, if the ideal sheaf $\mathscr{I}$ of $S_0$ in $S$ satisfies $\mathscr{I}^N = 0$ for some $N$ and we are just given an abelian scheme $A_0$ over $S_0$ and a $p$-divisible group $\Gamma$ over $S$ equipped with an isomorphism $\iota_0 : \Gamma_0 \simeq A_0[p^\infty]$, then there exists an abelian scheme $A$ over $S$ and an isomorphism $\iota : \Gamma \simeq A[p^\infty]$ lifting $\iota_0$. The pair $(A, \iota)$ is unique up to unique isomorphism.*

*Proof.* For an exposition of Drinfeld's elegant proof of this theorem, see [K, §1]. This proof uses the point of view of fppf abelian sheaves. ∎

In words, Theorem 3.3 says that the infinitesimal deformation theory of an abelian scheme coincides with the infinitesimal deformation theory of its $p$-divisible group when all points of the base have residue characteristic $p$. We stress that it is crucial for the Serre-Tate theorem that all points of $S$ have residue characteristic $p$ and that we work with *infinitesimal* deformations. For example, $p$-divisible groups are étale away from points of residue characteristic $p$, so there are no obstructions to infinitesimal deformations of $p$-divisible group maps at such points, whereas there can certainly be obstructions to infinitesimally deforming abelian scheme maps at such points (e.g., supersingular elliptic curves over $k = \overline{\mathbf{F}}_p$ always lift to $W(k)$, and it follows from Corollary 3.5 below that some of the endomorphisms in characteristic $p$ cannot lift through all infinitesimal levels of the deformation to $W(k)$).

One reason for the significance of Theorem 3.3 for our purposes is that if $A$ and $A'$ are abelian schemes over a complete local noetherian ring $R$ of residue characteristic $p > 0$ (the case of elliptic curves being the only one we'll need), then $\mathrm{Hom}_{R_n}(A_n, A'_n)$ is naturally identified with the group of "algebraic" elements in $\mathrm{Hom}_k(\Gamma_0, \Gamma'_0)$ which lift (necessarily uniquely) to $\mathrm{Hom}_{R_n}(\Gamma_n, \Gamma'_n)$. This is particularly interesting when $k$ is algebraically closed and we are working with $A = A'$ equal to a supersingular elliptic curve $E_0$ over $k$. In such a case, $\Gamma_0$ "is" the unique commutative formal group over $k$ of dimension 1 and height 2, so $\mathrm{End}_k(\Gamma_0)$ is identified with the maximal order in the unique quaternion division algebra over $\mathbf{Q}_p$. Thus, problems in the deformation theory of endomorphisms are transformed into problems in quaternion algebras (and some such problems are solved in [Gr1], as we shall use later on).

The Grothendieck existence theorem addresses the problem of realizing a compatible family of infinitesimal deformations of a proper variety as the reductions of a common non-infinitesimal deformation.

**Theorem 3.4.** (Grothendieck) *Let $R$ be a noetherian ring which is separated and complete with respect to the $I$-adic topology for an ideal $I$. Let $X$ and $Y$ be* proper *$R$-schemes, and $R_n, X_n, Y_n$ the reductions modulo $I^{n+1}$. The natural map of sets $\mathrm{Hom}_R(X,Y) \to \varprojlim \mathrm{Hom}_{R_n}(X_n, Y_n)$ is bijective.*

*Moreover, if $\{X_n\}$ is a compatible system of proper schemes over the $R_n$'s and $\mathscr{L}_0$ is an ample line bundle on $X_0$ which lifts compatibly to a line bundle $\mathscr{L}_n$ on each $X_n$, then there exists a pair $(X, \mathscr{L})$ consisting of a proper $R$-scheme and ample line bundle which compatibly reduces to each $(X_n, \mathscr{L}_n)$, and this data over $R$ is unique up to unique isomorphism.*

*Proof.* See [EGA, $\mathrm{III}_1$, §5] for the proof and related theory. ∎

The first part of Grothendieck's theorem identifies the category of proper $R$-schemes with a full subcategory of the category of compatible systems $\{X_n\}$ of proper schemes over the $R_n$'s, and it is really the second part of the theorem which is usually called the existence theorem (as it asserts the existence of an $R$-scheme giving rise to specified infinitesimal data over the $R_n$'s).

It is not true that all compatible systems $\{X_n\}$ of proper schemes over the $R_n$'s actually arise from a proper scheme $X$ over $R$, and those which do are usually described as being *algebraizable.* Since the setup in Grothendieck's theorem is compatible with fiber products, to give compatible group scheme structures on the $X_n$'s is equivalent to giving a group scheme structure on $X$. Also, various fundamental openness results from [EGA, $\mathrm{IV}_3$, §11–12] ensure that for many important properties $\mathbf{P}$ of morphisms, a map $X \to Y$ of *proper* $R$-schemes has property $\mathbf{P}$ if and only if each $X_n \to Y_n$ does. For example, if each $X_n$ is $R_n$-smooth of pure relative dimension $d$, then the same holds for $X$ over $R$.

Since elliptic curves possess canonical ample line bundles (namely, the inverse ideal sheaf of the identity section), we see that Grothendieck's theorem provides the conceptual explanation (i.e., independent of Weierstrass equations) for why a compatible system of elliptic curve deformations over the $R_n$'s uniquely lifts to an elliptic curve over $R$. In higher relative dimension, formal abelian schemes need not be algebraizable. However, when given two abelian schemes $A$ and $A'$ over $R$ we conclude from Grothendieck's theorem that

$$\mathrm{Hom}_R(A, A') = \bigcap \mathrm{Hom}_{R_n}(A_n, A'_n) \subseteq \mathrm{Hom}_{R/\mathfrak{m}}(A_0, A'_0).$$

Combining the Grothendieck existence theorem with the Serre-Tate theorem, we arrive at the following crucial result (to be applied in the case of elliptic curves over a complete discrete valuation ring $W$).

**Corollary 3.5.** *Let $(R, \mathfrak{m})$ be a complete local noetherian ring with residue field $k = R/\mathfrak{m}$ of characteristic $p > 0$, and let $A, A'$ be abelian schemes over $R$ with associated $p$-divisible groups $\Gamma$ and $\Gamma'$. Then*

$$\mathrm{Hom}_R(A, A') = \mathrm{Hom}_k(A_0, A'_0) \bigcap \left( \bigcap_n \mathrm{Hom}_{R_n}(\Gamma_n, \Gamma'_n) \right)$$

*inside of* $\mathrm{Hom}_k(\Gamma_0, \Gamma_0')$.

Let $(R, \mathfrak{m})$ be as above, with perfect residue field $k$ of characteristic $p > 0$. Let $A_0$ be an ordinary abelian variety over $k$. Since $k$ is perfect, the connected-étale sequence $0 \to \Gamma_0^0 \to \Gamma_0 \to \Gamma_0^{\mathrm{et}} \to 0$ of $\Gamma_0 = E_0[p^\infty]$ is uniquely split. One lifting of $\Gamma_0$ to a $p$-divisible group over $R$ is the split form, namely a product of the unique (up to unique isomorphism) liftings of $\Gamma_0^0$ and $\Gamma_0^{\mathrm{et}}$ to multiplicative and étale $p$-divisible groups over $R$. Such a split lifting involves no ambiguity in the splitting data. More precisely, if the connected-étale sequence of a $p$-divisible group $\Gamma$ over $R$ is a split sequence, then the splitting is unique. Indeed, any two splittings differ by a morphism from the étale part to the connected part, so it suffices to show $\mathrm{Hom}_R(\Gamma_1, \Gamma_2) = 0$ for an étale $p$-divisible group $\Gamma_1$ and a connected $p$-divisible group $\Gamma_2$ over $R$. We may make a local faithfully flat base change on $R$ to get to the case where $R$ has algebraically closed residue field, in which case $\Gamma_1$ is a product of $\mathbf{Q}_p/\mathbf{Z}_p$'s, so it suffices to show that $\cap p^n \cdot \Gamma_2(R) = 0$. Since $\Gamma_2$ is a formal Lie group in finitely many variables, with multiplication by $p$ given by $T_i \mapsto T_i^p + p(\cdot)$ in terms of formal parameters $T_i$, $\cap p^n \cdot \Gamma_2(R)$ vanishes because $\cap \mathfrak{m}^n = 0$ (as follows from Krull's intersection theorem for noetherian local rings). With this uniqueness of splittings in hand, we can make a definition:

**Definition 3.6.** A *Serre-Tate canonical lift* of $A_0$ to $R$ is a pair $(A, \iota)$ where $A$ is an abelian scheme over $R$, $\iota$ is an isomorphism between $A_{/k}$ and $A_0$, and the connected-étale sequence of $A[p^\infty]$ is split.

It is immediate from the above theorems of Serre-Tate and Grothendieck that a Serre-Tate canonical lift is characterized by the condition that it is a lift of $A_0$ whose $p$-divisible group has a split connected-étale sequence over $R$, and that it is unique up to unique isomorphism (as a lift of $A_0$). Moreover, if $A_0'$ is another ordinary abelian variety with Serre-Tate canonical lift $(A', \iota')$, then $\mathrm{Hom}_R(A, A') \to \mathrm{Hom}_k(A_0, A_0')$ is a *bijection*. In general, when $R$ is not artinian one cannot expect a Serre-Tate canonical lift to exist; the best one can do is get a formal abelian scheme. However, in the case of elliptic curves the existence is always satisfied.

## 4. Computing naive intersection numbers

We are interested in computing the local intersection pairing $\langle \underline{x} - (\underline{\infty}), T_m(\underline{x^\sigma}) - T_m(\underline{0}) \rangle_v$, with $x \in X(H) \subseteq \underline{X}(W)$ a Heegner point and $\sigma \in \mathrm{Gal}(H/K)$. Serious complications will be caused by the possibility that $\underline{x}$ is a component of $T_m(\underline{x^\sigma})$. By [GZ, Ch III, Prop 4.3], the multiplicity of $x$ as a component in $T_m(x^\sigma)$ is $r_{\mathscr{A}}(m)$, the number of integral ideals of norm $m$ in the ideal class $\mathscr{A}$ associated to $\sigma \in \mathrm{Gal}(H/K) \simeq \mathrm{Cl}_K$. Thus, we will first concentrate on the case $r_{\mathscr{A}}(m) = 0$, so all divisors of interest on $\underline{X}$ intersect properly. The formula we wish to establish will relate local intersection numbers with Iso groups and Hom groups for infinitesimal deformations of Heegner points. This section is devoted to some general preliminaries concerning the simplest situation: intersecting two sections $\underline{y}, \underline{y}'$ in $\underline{X}(W)$ whose generic fibers are distinct and whose closed points $\underline{y}_0$ and $\underline{y}_0'$ are assumed to be disjoint from the cuspidal divisor and supported in the smooth locus. Heegner points will play no role here.

Recall that a section through the regular locus on a proper flat curve over a field or discrete valuation ring is automatically supported in the smooth locus, so we could equivalently be assuming that the closed points of our two sections lie in the (non-cuspidal part of the) regular locus of the closed fiber or that the sections lie in the (non-cuspidal part of the) regular locus on $\underline{X}$. The most important instance of this setup is given by sections arising from Heegner points or components of prime-to-$p$ Hecke divisors obtained from such points (as in Corollary 2.7).

For such sections $\underline{y}$ and $\underline{y}'$ we need to make an *additional assumption* that there exist $\Gamma_0(N)$-structure diagrams over $W$ which actually induce $\underline{y}$ and $\underline{y}$'. This is an additional assumption because $\underline{X}$ is not a fine moduli scheme (but we'll prove in a moment that in many cases this additional assumption is satisfied). Note in particular (thanks to Theorem 2.6) that such models over $W$ are unique up to non-unique isomorphism, so it is easily checked that subsequent considerations which use such models (e.g., Theorem 4.1 below) are intrinsic to the given sections $\underline{y}$ and $\underline{y}'$ in $\underline{X}(W)$.

In this setting it makes sense to consider the intersection number

$$(\underline{y}.\underline{y}') \stackrel{\mathrm{def}}{=} \mathrm{length}(\underline{y} \cap \underline{y}') \tag{4.1}$$

where $\underline{y} \cap \underline{y}'$ denotes a scheme-theoretic intersection inside of $\underline{X}$. Although general intersection theory on arithmetic curves usually works with degree 0 divisors and requires passing to a regular resolution of $\underline{X}$ and using a moving lemma there, the regular resolution process can be carried out without affecting the relative smooth locus. Thus, the geometric input needed for an intersection calculation on a regular resolution is the length as in (4.1). Taking the point of view of the right side of (4.1) avoids general intersection theory and hence makes sense with *no* assumption of regularity/smoothness. Thus, we could contemplate trying to establish a formula for this length with *no* regularity assumptions at all. However, we'll see that a regularity assumption is crucial if we want to avoid restrictions on automorphism groups at closed points. Here is the formula we wish to prove:

**Theorem 4.1.** *Let $\underline{y}, \underline{y}' \in \underline{X}(W)$ be sections which intersect properly and reduce to regular non-cuspidal points in the special fiber (and hence are supported in the relative smooth locus over $W$). Assume moreover that these sections are induced by (necessarily unique) respective $\Gamma_0(N)$-structures denoted $\underline{y}$ and $\underline{y}'$ over $W$. Then* $\mathrm{length}(\underline{y} \cap \underline{y}') = \frac{1}{2}\sum_{n \geq 0} \#\mathrm{Isom}_{W_n}(\underline{y}', \underline{y})$, *where $W_n = W/\pi^{n+1}$.*

Before we prove Theorem 4.1, we should record the following result which explains why the condition of existence of "$W$-models" for $\underline{y}, \underline{y}' \in \underline{X}(W)$ is usually automatically satisfied.

**Lemma 4.2.** *If $\underline{y} \in \underline{X}(W)$ is a section which is disjoint from the cuspidal locus and has $\mathrm{Aut}(\underline{y}_0) = \{\pm 1\}$, then there exists a $\Gamma_0(N)$-structure over $W$ which induces $\underline{y}$. Moreover, with no assumptions on $\mathrm{Aut}(\underline{y}_0)$, the complete local ring of $\underline{X}$ at the closed point $\underline{y}_0$ of $\underline{y}$ is the subring of $\mathrm{Aut}(\underline{y}_0)$-invariants in the universal deformation ring of $\underline{y}_0$ and $\{\pm 1\}$ acts trivially on this ring.*

Since $W/\pi$ is algebraically closed, every non-cuspidal point in $\underline{X}(W/\pi)$ is actually represented by a $\Gamma_0(N)$-structure over $W/\pi$ (unique up to non-unique isomorphism). Thus, the concept of $\mathrm{Aut}(\underline{y}_0)$ makes sense prior to the proof of the lemma (and the existence of a universal deformation ring follows from the theory of Drinfeld structures on elliptic curves).

*Proof.* (of Lemma 4.2) Since formal deformations of $\Gamma_0(N)$-diagrams are algebraizable (thanks to Theorem 3.4), it suffices to prove the assertion concerning universal deformation rings.

If one looks at how the coarse moduli scheme $\underline{X}$ is constructed from a fine moduli scheme $\underline{X}(\iota)$ (away from the cusps) upon adjoining enough prime-to-$p$ level structure $\iota$, and one notes that formation of coarse moduli schemes in our context commutes with flat base change (such as $\mathbf{Z} \to W$), it follows that the complete local ring at $\underline{y}_0$ on $\underline{X}$ is exactly the subring of $\Gamma = \mathrm{Aut}(\underline{y}_0)$-invariants in the complete local ring on $\underline{X}(\iota)$ at a point corresponding to $\underline{y}_0$ with supplementary $\iota$-structure added. The justification of this assertion is a standard argument in deformation theory which we omit.

Since $\underline{X}(\iota)$ is a fine moduli scheme away from the cusps, its complete local rings are formal deformation rings. Combining this with the fact that étale level structure $\iota$ is "invisible" when considering formal deformations, we conclude that the complete local ring in question on $\underline{X}$ really is naturally identified with the $\Gamma$-invariants in the universal deformation ring of $\underline{y}_0$ (with the *natural* action of $\Gamma$!). It therefore remains to check that $-1 \in \Gamma$ acts trivially on this deformation ring. But since inversion uniquely lifts to all deformations of a $\Gamma_0(N)$-structure (other closed fiber automorphisms usually don't lift!), the triviality of the action of $-1$ on universal deformation rings drops out.

■

*Proof.* (of Theorem 4.1) If $\underline{y}_0 \neq \underline{y}'_0$, then both sides of the formula are 0 and we're done. If these closed points coincide, so $\underline{y}_0 \simeq \underline{y}'_0$ as $\Gamma_0(N)$-structures over $W/\pi$, we at least see that the automorphism groups of $\underline{y}_0$ and $\underline{y}'_0$ (by which we mean the automorphism groups of representative $\Gamma_0(N)$-structures over $W/\pi$) are abstractly isomorphic.

Let $z \in \underline{X}(W/\pi)$ be the common non-cuspidal point arising from $\underline{y}_0$ and $\underline{y}'_0$. For later purposes let's also *fix* choices of isomorphisms of $\Gamma_0(N)$-structures $\underline{y}_0 \simeq z$ and $\underline{y}'_0 \simeq z$ in order to view $\underline{y}$ and $\underline{y}'$ (or more specifically *fixed* choices of $W$-models for these sections) as $W$-deformations of the $\Gamma_0(N)$-structure $z$.

Let $A_z$ denote the universal deformation ring for $z$. This ring is regular of dimension 2 (as is proven in [KM]) and has a unique structure of $W$-algebra (compatibly with its residue field identification with $W/\pi$), so by commutative algebra the existence of a $W$-section (such as coming from $\underline{y}$ or $\underline{y}'$) forces $A_z$ to be formally $W$-smooth, which is to say of the form $W[\![T]\!]$ as an abstract local $W$-algebra.

By Lemma 4.2, the complete local ring $\widehat{\mathscr{O}}_{\underline{X},z}$ is exactly the subring $A_z^{\overline{\Gamma}}$, where $\overline{\Gamma} = \mathrm{Aut}(z)/\{\pm 1\}$ acts naturally. Let $d = |\overline{\Gamma}|$ denote the size of $\overline{\Gamma}$. A crucial point is that $\overline{\Gamma}$ acts *faithfully* on $A_z$ (i.e., the only elements of $\mathrm{Aut}(z)$ acting trivially on the deformation ring are the elements $\pm 1$). This amounts to the assertion that $\pm 1$ are the only automorphisms of $z$ which lift to all deformations of $z$. That is, the generic deformation cannot have automorphisms other than $\pm 1$. This follows from the 1-dimensionality of modular curves and the existence of elliptic curves with automorphism group $\{\pm 1\}$.

The finite group $\overline{\Gamma}$ of order $d$ now acts faithfully on the formally smooth $W$-algebra $A_z \simeq W[\![T]\!]$, and if we define the "norm" $t = \mathrm{Norm}_{\overline{\Gamma}}(T) = \prod_{\gamma \in \overline{\Gamma}} \overline{\gamma}(T)$ then by [KM, p. 508] the subring $\widehat{\mathscr{O}}_{\underline{X},z}$ of $\overline{\Gamma}$-invariants is exactly the formal power series ring $W[\![t]\!]$ and [Mat, Thm 23.1] ensures that the resulting finite map $W[\![t]\!] \to W[\![T]\!]$ between 2-dimensional regular local rings is necessarily *flat*. Moreover, using Artin's theorem on the subfield of invariants under a faithful action of a finite group on a field (such as $\overline{\Gamma}$ acting on the fraction field of $W[\![T]\!]$) we see that $\widehat{\mathscr{O}}_{\underline{X},z} \to A_z$ is finite flat of degree $d = |\overline{\Gamma}|$. XC The two sections $\underline{y}, \underline{y}' : \widehat{\mathscr{O}}_{\underline{X},z} \twoheadrightarrow W$ both arise from specified deformations and hence (uniquely) lift to $W$-sections of $A_z$. We can choose $T$ without loss of generality to cut out the section corresponding to the deformation $\underline{y}$, and we let $T'$ correspond likewise to the deformation $\underline{y}'$. Define $t'$ to be the $\overline{\Gamma}$-norm of $T'$ in $\widehat{\mathscr{O}}_{\underline{X},z}$. By *definition*, we have $(\underline{y}.\underline{y}') = \mathrm{length}(\widehat{\mathscr{O}}_{\underline{X},z}/(t,t')) = (1/d)\mathrm{length}(A_z/(t,t'))$ since $\widehat{\mathscr{O}}_{\underline{X},z} \to A_z$ is finite *flat* of degree $d$.

By induction and short exact sequence arguments with lengths, one shows that if $A$ is any 2-dimensional regular local ring and $a, a' \in A$ are two non-zero elements of the maximal ideal with $A/(a,a')$ of finite length and prime factorization $a = \prod p_i$, $a' = \prod q_j$ (recall $A$ is a UFD), then $A/(p_i, q_j)$ has *finite* length for all $i, j$ and $\mathrm{length}(A/(a,a')) = \sum_{i,j} \mathrm{length}(A/(p_i,q_j))$. Thus, in our setup we have

$$\mathrm{length}(A_z/(t,t')) = \sum_{\gamma,\gamma' \in \overline{\Gamma}} \mathrm{length}(A_z/(\gamma(T), \gamma'(T'))) = d \sum_{\gamma' \in \overline{\Gamma}} \mathrm{length}(A_z/(T, \gamma'(T'))).$$

Consequently, we get

$$(\underline{y}.\underline{y}') = \sum_{\gamma' \in \overline{\Gamma}} \mathrm{length}(A_z/(T, \gamma'(T'))) \tag{4.2}$$

We have $A_z/T = W$ corresponding to the deformation $\underline{y}$, and $A_z/(T, \gamma'(T'))$ is of finite length and hence of the form $W/\pi^{k_{\gamma'}}$ for a unique positive integer $k_{\gamma'}$. If we let $\gamma'(\underline{y}')$ denote the deformation obtained from the $\Gamma_0(N)$-structure $\underline{y}'$ by composing its residual isomorphism with $z$ with a representative automorphism $\widetilde{\gamma}' \in \mathrm{Aut}(z)$ for $\gamma' \in \overline{\Gamma} = \mathrm{Aut}(z)/\{\pm 1\}$, then $\gamma'(\underline{y}')$ corresponds to $A_z/\gamma'(T')$ and hence there exists a (unique!) isomorphism of *deformations* $\underline{y} \bmod \pi^k \simeq \gamma'(\underline{y}') \bmod \pi^k$ if and only if $k \le k_{\gamma'}$. Equivalently, there exists a (unique) isomorphism $\underline{y} \bmod \pi^k \simeq \underline{y}' \bmod \pi^k$ lifting $\widetilde{\gamma}' \in \mathrm{Aut}(z)$ if and only if $k \le k_{\gamma'}$.

Any isomorphism $\underline{y} \bmod \pi^k \simeq \underline{y}' \bmod \pi^k$ merely as *abstract* $\Gamma_0(N)$*-structures* (ignoring the deformation structure with respect to $z$) certainly either lifts a unique $\widetilde{\gamma}'$ or else its negative does (but not both!). Since passage to the mod $\pi$ fiber is *faithful* (Theorem 2.1(2)), so any $\Gamma_0(N)$-structure isomorphism between $\underline{y}$ and $\underline{y}'$ modulo $\pi^k$ is uniquely detected as a compatible system of isomorphisms modulo $\pi^n$'s for $n \le k$, we conclude from (4.2) that $(\underline{y}.\underline{y}')$ counts exactly the sum over all $n \ge 1$ of the sizes of the sets $\mathrm{Isom}_{W_n}(\underline{y}, \underline{y}')$ up to identifying isomorphisms with their negatives. This is essentially just a jazzed-up way of interchanging the order of a double summation. Since no isomorphism can have the same reduction as its negative, we have completed the proof of Theorem 4.1 by purely deformation-theoretic means.

■

## 5. Intersection formula via Hom groups

The link between intersection theory and deformation theory is given by:

**Theorem 5.1.** *Let $\sigma \in \mathrm{Gal}(H/K)$ correspond to the ideal class $\mathscr{A}$ of $K$, and assume that the number $r_{\mathscr{A}}(m)$ of integral ideals in $\mathscr{A}$ of norm $m$ vanishes, where $m \geq 1$ is relatively prime to $N$. Then*

$$(\underline{x}.T_m(\underline{x^\sigma})) = \frac{1}{2}\sum_{n\geq 0} |\operatorname{Hom}_{W_n}(\underline{x^\sigma}, \underline{x})_{\deg m}|. \tag{5.1}$$

*Remark* 5.2. Recall that the hypothesis $r_{\mathscr{A}}(m) = 0$ says that $\underline{x}$ and $T_m(\underline{x^\sigma})$ intersect properly, so Theorem 4.1 is applicable (thanks to the description of $T_m(\underline{x^\sigma})$ as a sum of Cartier divisors in Corollary 2.7).

In order to prove Theorem 5.1, we have to treat three essentially different cases: $p$ not dividing $m$, $p|m$ with $p$ split in $K$, and $p|m$ with $p$ inert or ramified in $K$. The first case will follow almost immediately from what we have already established, while the second case amounts to a careful study of ordinary elliptic curves and Serre-Tate canonical liftings, and the third case is a subtle variant on the second case where we have to deal with supersingular reduction and must replace Serre-Tate theory with Gross' variant as discussed in [Gr1]. In particular, the phrases "canonical" and "quasi-canonical" liftings will have different meanings depending on whether we are in the ordinary or supersingular cases (for $p|m$).

In this section we take care of the easy case $p \nmid m$ and the less involved case where $p|m$ but $p$ splits in $K$ (i.e., the ordinary case). Our proof will essentially be the one in [GZ], except we alter the reasoning a little bit to avoid needing to use explicit Hecke formulas on divisors. The supersingular case (i.e., $p$ not split in $K$) will roughly follow the same pattern as the ordinary case, except things are a bit more technical (e.g., it seems unavoidable to make explicit Hecke computations on divisors) and hence we postpone such considerations until the next section.

Let's now take care of the easy case $p \nmid m$. In this case, all $m$-isogenies are finite étale, so by Corollary 2.7 we see that if $\{C\}$ denotes the (finite) set of order $m$ subgroup schemes of $\underline{x^\sigma}_0$, each of which uniquely lifts to any deformation of $\underline{x^\sigma}_0$ (thanks to the invariance of the étale site with respect to infinitesimal thickenings). As relative effective Cartier divisors in the relative smooth locus of $\underline{X}$ over $W$ we get the equality

$$T_m\underline{x^\sigma} = \sum_C \underline{x^\sigma}_C$$

where $\underline{x^\sigma}_C$ denotes the quotient $\Gamma_0(N)$-structure on $\underline{x^\sigma}$ by the unique order $m$ subgroup scheme over $W$ lifting $C$ modulo $\pi$ (this makes sense since $m$ is relatively prime to $N$).

By Theorem 4.1 and the symmetry of intersection products, we obtain the formula

$$(\underline{x}.T_m\underline{x^\sigma}) = \sum_C (\underline{x}.\underline{x^\sigma}_C) = \sum_{n\geq 0}\sum_C \frac{1}{2}|\operatorname{Isom}_{W_n}(\underline{x^\sigma}_C, \underline{x})| \tag{5.2}$$

with all sums implicitly finite (i.e., all but finitely many terms vanish). Since any non-zero map between elliptic curves over $W_n$ is automatically finite flat (thanks to the fiber-by-fiber criterion for flatness), we conclude that the inner sum over $C$'s is equal to $\frac{1}{2}|\operatorname{Hom}_{W_n}(\underline{x^\sigma}, \underline{x})_{\deg m}|$. This gives Theorem 5.1 in case $p \nmid m$.

Now assume $p|m$, so we have $m = p^t r$ with $t, r \geq 1$ and $p \nmid r$. In particular, this forces $p \nmid N$, so the $\Gamma_0(N)$-structures are *étale* and $\underline{X}$ is a *proper smooth curve* over $W$. Hence, we have a good theory of Cartier divisors and intersection theory on $\underline{X}$ without needing to do any resolutions at all, and we can keep track of compatibilities with $\Gamma_0(N)$-structures merely by checking such compatibility modulo $\pi$. Also, we can work with Hecke correspondences directly on the level of "integral model" Cartier divisors (rather than more indirectly in terms of closures from characteristic 0 on abstract regular resolutions). These explications are important because much of our analysis will take place on the level of $p$-divisible groups and deformations thereof (for which level $N$ structure is totally invisible) and we will have to do very direct analysis of Hecke correspondences on relative effective Cartier divisors in characteristic 0, characteristic $p$, and at the infinitesimal level. The a priori knowledge that a deformation of a $\Gamma_0(N)$-compatible map is *automatically* $\Gamma_0(N)$-compatible will be the reason that we can focus most of our attention on $p$-power torsion and not worry about losing track of morphisms within the category of $\Gamma_0(N)$-structures.

In $\underline{X}$ we have the equality of relative effective Cartier divisors $T_m(\underline{x^\sigma}) = T_{p^t}(T_r(\underline{x^\sigma}))$ with $T_r(\underline{x^\sigma})$ a sum of various $W$-sections $\underline{z}$, thanks to Corollary 2.7. Thus,

$$(\underline{x}.T_m(\underline{x^\sigma})) = \sum_{\underline{z}} (\underline{x}.T_{p^t}(\underline{z})). \tag{5.3}$$

Before we focus our attention on the proof of Theorem 5.1 when $p|m$, we make one final observation. Using the connected-étale sequence for finite flat group schemes over $W_n$ (such as the kernels of dual isogenies $\underline{x} \to \underline{x^\sigma}$ of degree $m$) and the fact that finite flat group schemes of prime-to-$p$ order in residue characteristic $p$ are automatically étale (and therefore constant when the residue field is algebraically closed), we see that both sides of Theorem 5.1 naturally break up into into sums over all $\underline{z}$'s. It suffices to establish the equality

$$(\underline{x}.T_{p^t}(\underline{z})) = \frac{1}{2}\sum_{n\geq 0} |\operatorname{Hom}_{W_n}(\underline{z},\underline{x})_{\deg p^t}| \tag{5.4}$$

for each irreducible component $\underline{z} \in \underline{X}(W)$ of $T_r(\underline{x^\sigma})$.

We now prove (5.4) when $p$ is split in $K$, which is to say that all the Heegner points (such as $\underline{x}$ and $\underline{x^\sigma}$) have ordinary reduction. The key observation is that the $W$-models corresponding to all such Heegner points must be (Serre-Tate) *canonical lifts* of their closed fibers, which is to say that the connected-étale sequences of their $p$-divisible groups over $W$ are *split* (and non-canonically isomorphic to $\mathbf{Q}_p/\mathbf{Z}_p \times \mathbf{G}_m[p^\infty]$). To prove this, we first note that by Tate's isogeny theorem for $p$-divisible groups over $W$ it suffices to check that on the generic fiber, the $p$-adic Tate modules underlying Heegner points over $W$ are isomorphic to $\mathbf{Z}_p \times \mathbf{Z}_p(1)$ as Galois modules.

What we know from ordinary reduction and the fact that $W$ has algebraically closed residue field is that these $p$-adic Tate modules must be extensions of $\mathbf{Z}_p$ by $\mathbf{Z}_p(1)$. In order to get the splitting, we use the fact that for any elliptic curve $E$ over $F$ with CM by $K$, the CM-action by $K$ is defined over $K$ and hence also over $F$, so the Galois action commutes with the *faithful* action of

$$\mathbf{Z}_p \otimes_{\mathbf{Z}} \mathscr{O}_K = \mathbf{Z}_p \otimes_{\mathbf{Z}} \operatorname{Hom}_F(E,E) \hookrightarrow \operatorname{Hom}_F(T_p(E), T_p(E)).$$

Thus, when $p$ is split in $K$ and $E$ has good ordinary reduction then the action of the $\mathbf{Z}_p$-algebra $\mathbf{Z}_p \otimes_{\mathbf{Z}} \mathscr{O}_K \simeq \mathbf{Z}_p \times \mathbf{Z}_p$ gives rise to a decomposition of $T_p(E)$ into a direct sum of two Galois characters which moreover *must* be the trivial and $p$-adic cyclotomic characters, as desired. This proves that our Heegner points are in fact automatically the Serre-Tate canonical lifts of their ordinary closed fibers. Note that each section $\underline{z} \in \underline{X}(W)$ of $T_r(\underline{x^\sigma})$ has $p$-adic Tate module *isomorphic* to that of the Heegner point $\underline{x^\sigma}$ and hence all such $\underline{z}$'s are Serre-Tate canonical lifts of their closed fibers (even though such $\underline{z}$'s generally are not Heegner points).

By the Serre-Tate theorem, we conclude

$$\operatorname{Hom}_W(\underline{x^\sigma},\underline{x}) \hookrightarrow \operatorname{Hom}_{W/\pi}(\underline{x^\sigma},\underline{x}) \tag{5.5}$$

is an isomorphism because on the $p$-divisible groups side it is obvious the endomorphisms of $\mathbf{Q}_p/\mathbf{Z}_p \times \mathbf{G}_m[p^\infty]$ over $W/\pi$ uniquely lift to $W$ (and recall that for the level $N$-structure we don't have to check anything once the residual level respects this data).

The assumption $r_{\mathscr{A}}(m) = 0$ says that there are no degree $m$ isogenies on *generic geometric fibers* between $\underline{x^\sigma}$ and $\underline{x}$, so there are no such isogenies over $W$ thanks to Theorem 2.6. Since (5.5) is an isomorphism, there are no such isogenies modulo $\pi^{n+1}$ for each $n \geq 0$. This renders the right side of Theorem 5.1 equal to 0. Thus, to prove the theorem we must prove that $(\underline{x}.T_{p^t}(\underline{z})) = 0$ for any point $\underline{z} \in \underline{X}(W)$ which is prime-to-$p$ isogenous to $\underline{x^\sigma}$. In more explicit terms, we need to show that a $\Gamma_0(N)$-structure (over some finite extension $F'$ of $F$) which is $p$-power isogenous to $\underline{z}_{/F'}$ (and hence has to have good reduction) must have corresponding closed subscheme in $\underline{X}$ which is *disjoint* from $\underline{x}$.

Assuming the contrary, suppose that $\underline{x}$ actually meets $T_{p^t}(\underline{z})$. Thus, some point in $T_{p^t}(\underline{z})$ (viewed as a $\Gamma_0(N)$-structure over a finite extension $W'$ of $W$) would have to admit $\underline{x}_{/W'}$ as the canonical lift of its closed fiber. We have just seen that $\underline{x}$ is *not* one of the component points of $T_{p^t}(\underline{z})$, yet $\underline{x}$ as a $\Gamma_0(N)$-structure is the Serre-Tate canonical lift of its (ordinary) closed fiber. To get a contradiction, it therefore suffices to check that for any non-cuspidal generic geometric point $\underline{z}$ on $\underline{X}$ which has good reduction and is the

Serre-Tate canonical lift of its closed fiber, every point in the geometric generic fiber of of $T_{p^t}(\underline{z})$ has the Serre-Tate canonical lift of its closed fiber (with lifted $\Gamma_0(N)$-structure) as one of the (geometric generic) points of $T_{p^t}(\underline{z})$. Due to the level of generality of this claim (with respect to the hypotheses on $\underline{z}$), we can use induction on $t$ and the recursive formula $T_{p^{t+1}} = T_p T_{p^t} - pT_{p^{t-1}}$ to reduce to the case $t = 1$.

By the very definition of $T_p$, we break up $T_p(\underline{z})_{/\overline{F}}$ (viewed on the level of Néron models of "elliptic curve moduli" over a sufficiently large extension of $W$) into two parts: the quotient by the order $p$ connected subgroup and the $p$ quotients by the étale order $p$ subgroups. Since $\underline{z}$ is a Serre-Tate canonical lift, it has two canonical quotients (one with connected kernel and one with étale kernel) which are visibly Serre-Tate canonical lifts of their closed fibers. Moreover, these two quotients yield the two distinct geometric points which we see in $T_p(\underline{z})$ on the closed fiber divisor level. This completes the case when $p$ is split in $K$.

## 6. Supersingular cases with $r_{\mathscr{A}}(m) = 0$

We now treat the hardest case of Theorem 5.1: $p|m$ with $p$ either inert or ramified in $K$. All elliptic curves in question will have (potentially) supersingular reduction. For as long as possible, we shall simultaneously treat the cases of $p$ inert in $K$ and $p$ ramified in $K$. We will prove (5.4) for $\underline{z}$ in $T_r(\underline{x^\sigma})$.

Let $K_p = K \otimes_{\mathbf{Q}} \mathbf{Q}_p$ be the completion of $K$ at the unique place over $p$, so $K_p$ is a quadratic extension *field* of $\mathbf{Q}_p$, with valuation ring $\mathscr{O} = \mathscr{O}_K \otimes_{\mathbf{Z}} \mathbf{Z}_p$. The $p$-divisible groups of our Heegner points over $W$ are formal $\mathscr{O}$-modules of height 1 over $W$. Instead of the Serre-Tate theory of canonical lifts, we will have to use Gross' theory of canonical and quasi-canonical lifts for height 1 formal $\mathscr{O}$-modules. In this theory, developed in [Gr1], the role of canonical lifts is played by the unique (up to non-canonical isomorphism) height 1 formal $\mathscr{O}$-module over $W$, namely a Lubin-Tate formal group. The formal groups of $\underline{x}$, $\underline{x^\sigma}$, and $\underline{z}$ all admit this additional "module" structure (replacing the property of being Serre-Tate canonical lifts in the ordinary case treated above). The points in the $T_{p^t}(\underline{z})$'s will correspond to various quasi-canonical lifts in the sense of [Gr1], and working out the field/ring of definition of such $\Gamma_0(N)$-structures will play an essential role in the intersection theory calculations in the supersingular case.

Our first order of business is to give a precise description of $T_{p^t}(\underline{z})$ for *any* section $\underline{z} \in \underline{X}(W)$ which is supported away from the cuspidal locus and is represented by a $\Gamma_0(N)$-structure $\phi : E \to E'$ over $W$ for which the "common" $p$-divisible group of $E$ and $E'$ is endowed with a structure of Lubin-Tate formal $\mathscr{O}$-module over $W$ (note $\phi$ is an $N$-isogeny and $p \nmid N$, so $\phi$ induces an isomorphism on $p$-divisible groups). The $\underline{z}$'s we care about will be known to have such properties due to their construction from Heegner points, but in order to keep straight what really matters we avoid assuming here that $\underline{z}$ is a CM point or even that $\underline{z}$ has any connection to Heegner points at all.

Since $W$ is the completion of the maximal unramified extension of $K$, by Lubin-Tate theory (or more specifically local class field theory for $K$ at the unique place over the ramified/inert $p$) we know that the Galois representation $\chi : G_F \to \mathscr{O}^\times$ on the generic fiber $p$-divisible group associated to a Lubin-Tate formal $\mathscr{O}$-module of height 1 is surjective and (due to the uniqueness of such formal $\mathscr{O}$-modules up to non-unique isomorphism) this character is independent of the specific choice of such formal group. That is, by thinking in terms of this character $\chi$ we are dealing with a canonical concept. One may object that the data of the $\mathscr{O}$-action on the $p$-divisible group of $\underline{z}$ is extra data which we have imposed, but we'll see that the "counting" conclusions we reach will not depend on this choice. Note also that in local class field theory it is shown that $\chi$ is the reciprocal of the reciprocity map for $K_p$ (if one associates local uniformizers to arithmetic Frobenius elements; if one adopts the geometric Frobenius convention of Deligne then $\chi$ is the reciprocity map). This will lead us to the connection with local ring class fields over $K_p$, a connection of paramount importance in the proof of Theorem 6.4 below.

In order to describe $T_{p^t}(\underline{z})$ where $\underline{z} = (\phi : E \to E')_{/W}$, we first determine the geometric generic points of $T_{p^t}(\underline{z})$, which is to say that we consider the situation over $\overline{F}$. By definition as a divisor on $\underline{X}_{/\overline{F}}$,

$$T_{p^t}(\underline{z})_{/\overline{F}} = T_{p^t}(\underline{z}_{/\overline{F}}) = \sum_{C} (\phi_C : E/C \to E'/\phi(C))$$

where $C$ runs over all order $p^t$ subgroups of $E$ and $\phi_C$ is the naturally induced map (still an $N$-isogeny since $p \nmid N$). We break up the collection of $C$'s into collections based on the largest $u \geq 0$ for which $C$ contains

$E[p^u]$. We have the natural isomorphism $E/E[p^u] \simeq E$ which carries $C$ over to a cyclic subgroup of order $p^{t-2u}$, and we note that $\phi$ carries $E[p^u]$ isomorphically over to $E'[p^u]$. Letting $s = t - 2u \geq 0$, we therefore get

$$T_{p^t}(\underline{z})_{/\overline{F}} = \sum_{s \equiv t \bmod 2, 0 \leq s \leq t} \sum_{|C|=p^s} (\phi_C : E/C \to E'/\phi(C))$$

with the inner sum now taken over *cyclic* subgroups of order $p^s$.

We wish to determine which of the $\overline{F}$-points on $\underline{X}$ in this sum correspond to a common closed point on $\underline{X}_{/F}$. By Galois theory, this amounts to working out the $\mathrm{Gal}(\overline{F}/F)$-orbits on these points, where $\sigma \in \mathrm{Gal}(\overline{F}/F)$ acting through functoriality on $\underline{X}(\overline{F})$ carries a point $\phi_C$ to the point $\phi_{\sigma(C)}$ where $\sigma$ acts naturally on $C \subseteq E(\overline{F})$ (and $\sigma(\phi(C)) = \phi(\sigma(C))$ since $\phi$ is a morphism over $F$). Fortunately, $\sigma(C)$ is something we can compute because $C \subseteq E[p^\infty](\overline{F})$ and $G_F$ acts on $E[p^\infty](\overline{F})$ through the character $\chi$! Since $\chi$ is a surjective map onto $\mathscr{O}^\times$, the problem of working out Galois orbits (for which we may fix $s$) is exactly the problem of computing orbits for the natural $\mathscr{O}^\times$-action on the $p^{s-1}(p+1)$ cyclic subgroups of order $p^s$ in $\mathscr{O}/p^s \simeq (\mathbf{Z}/p^s)^{\oplus 2}$ when $s \geq 1$ (the case $s = 0$ is trivial: in this case $t$ is even and there is a single orbit corresponding to the subgroup $E[p^{t/2}]$, which is to say the point $\underline{z}$).

First suppose $p$ is inert in $K$, so $\mathscr{O}/p$ is a field (of order $p^2$). Every element of additive order $p^s$ in $\mathscr{O}/p^s$ is a multiplicative unit, so there is a unique orbit. This orbit is a closed point whose degree over $F$ is 1 when $s = 0$ and $p^{s-1}(p+1)$ when $s > 0$ (since for $s > 0$ we just saw that the orbit contains $p^{s-1}(p+1)$ distinct geometric points). This yields

$$T_{p^t}(\underline{z}) = \sum_{0 \leq s \leq t, s \equiv t \bmod 2} \underline{y}(s) \tag{6.1}$$

with each "horizontal divisor" $\underline{y}(s)$ having its generic fiber equal to a single closed point in $\underline{X}_{/F}$ of degree $p^s + p^{s-1}$ over $F$ for $s > 0$ and of degree 1 over $F$ for $s = 0$.

From the local class field theory interpretation of the preceding calculation, we see that the generic point of $\underline{y}(s)$ is defined over an abelian extension of $F$ which is obtained from the abelian extension of $K_p$ whose norm group has image in $(\mathscr{O}/p^s)^\times$ which is spanned by the stablizer $(\mathbf{Z}/p^s)^\times$ of a "line". This corresponds to the subgroup $\mathbf{Z}_p^\times \cdot (1 + p^s\mathscr{O}) = (\mathbf{Z}_p + p^s\mathscr{O})^\times$ of index $(p+1)p^{s-1}$ in $\mathscr{O}^\times$. If $F(s)/F$ denotes the corresponding finite extension (abelian over $K_p$), we see see that the field of definition of $\underline{y}(s)_{/F}$ is exactly $\mathrm{Spec}(F(s))$. We let $W(s)$ denote the corresponding valuation ring and let $\pi_s$ be a uniformizer, so $\underline{y}(s)$ corresponds to an element in $\underline{X}(W(s))$. It is clear from the construction that the $p$-divisible group underlying $\underline{y}(s)$ is *exactly* a level $s$ quasi-canonical deformation (in the sense of [Gr1]) of the formal $\mathscr{O}$-module underlying $\underline{z}_0^{(p^t)} \simeq \underline{x}_0$. Moreover, by taking scheme-theoretic closure from the generic fiber (in conjunction with the generic fiber description of quasi-canonical deformations in [Gr1]) we see that *the point* $\underline{y}(s) \in \underline{X}(W(s))$ *arises from a* $\Gamma_0(N)$*-structure defined over* $W(s)$. We stress that from the point of view of coarse moduli schemes this is a slightly surprising fact (for which we will be very grateful in the proof of Theorem 6.4). By Theorem 2.6 such a $W(s)$-model is unique up to non-unique isomorphism.

Meanwhile, when $p$ is ramified in $K$ then if $\pi \in \mathscr{O}$ is a uniformizer (recall $W$ is unramified over $\mathscr{O}$, so this is not really an abuse of notation), we see that $\mathscr{O}/p$ is a non-split $\mathscr{O}$-module extension of $\mathscr{O}/\pi = \mathbf{F}_p$ by $\pi\mathscr{O}/\pi^2\mathscr{O} \simeq \mathscr{O}/\pi$. For $s > 0$, among the $p^s + p^{s-1}$ cyclic subgroups of order $p^s$ in $\mathscr{O}/p^s$ we have a natural decomposition into the $p^{s-1}$ such subgroups whose generators map to non-zero elements in $\pi\mathscr{O}/\pi^2\mathscr{O} \subseteq \mathscr{O}/p$ and the $p^s$ others whose generators have non-zero image in $\mathscr{O}/\pi$. In the first case the generators are unit multiples of $\pi$ and in the second case the generators are units. Thus, both such clumps are orbits, so for each $s > 0$ with $s \equiv t \bmod 2$ we get an orbit of size $p^s$ and an orbit of size $p^{s-1}$. When $t$ is even, the case $s = 0$ also gives rise to another singleton orbit corresponding again to the point $\underline{z}$. For both options of the parity of $t$, we arrive at $t + 1$ distinct closed points $\underline{y}(s)_{/F}$ on $\underline{X}_{/F}$ of degrees $p^s$ over $F$ for $0 \leq s \leq t$. In other words, we get

$$T_{p^t}(\underline{z}) = \sum_{0 \leq s \leq t} \underline{y}(s) \tag{6.2}$$

with the generic fiber of $\underline{y}(s)$ of degree $p^s$ over $F$. These generic fiber closed points of $\underline{X}_{/F}$ have residue field identical to the finite abelian extension $F(s)$ of $F$ corresponding (via local class field theory for $K_p$) to the subgroup $(\mathbf{Z}_p + \pi^s\mathscr{O})^\times$ in $\mathscr{O}^\times$. Once again we recover level $s$ quasi-canonical deformations in the sense of [Gr1], and the points $\underline{y}(s) \in \underline{X}(W(s))$ are represented by $\Gamma_0(N)$-structures over $W(s)$.

The identities (6.1) and (6.2) are precisely the inert and ramified cases in [GZ, Ch III, (5.2)]. In both the inert and ramified cases, for $s > 0$ the point $\underline{y}(s) \in \underline{X}(W(s))$ is represented by a $\Gamma_0(N)$-structure over the valuation ring $W(s)$ of the local ring class field of level $s$, with its $p$-divisible group endowed with a natural structure of formal module of height 1 over a suitable *non-maximal* order in $\mathscr{O}$ (depending on $s$ and the $\mathscr{O}$-structure specified on $\underline{z}$). More specifically, the explicit description above shows that for $s > 0$ these formal modules are *quasi-canonical* deformations (in the sense of [Gr1]) which are *not* "canonical" (i.e., not formal modules over $\mathscr{O}$).

When $t$ is even, the point $\underline{y}(0)$ is $\underline{z}$ (use the isogeny $p^{t/2}$), corresponding to a "canonical" deformation of the closed fiber. When $s = 1$ in the inert case (so $t$ is odd), the $\Gamma_0(N)$-structure $\underline{y}(1)_0$ over $W/\pi$ is the quotient of $\underline{z}_0$ by its unique order $p$ subgroup scheme, which is to say that it is the Frobenius base change $\underline{z}_0^{(p)}$ of the $\Gamma_0(N)$-structure $\underline{z}_0$ over $W/\pi$. Finally, when $s = 0$ with odd $t$ in the ramified case, then $\underline{y}(0) \simeq \underline{z}^{\sigma_\mathfrak{p}}$ as $\Gamma_0(N)$-structures, where $\sigma_\mathfrak{p} \in \mathrm{Gal}(H/K)$ is the arithmetic Frobenius element at the unique prime $\mathfrak{p}$ of $K$ over $p$. To see this assertion in the ramified case, we just have to construct a $p$-isogeny between $\underline{z}$ and $\underline{z}^{\sigma_\mathfrak{p}}$ (and then compose it with $p^{(t-1)/2}$ to realize $\underline{z}^{\sigma_\mathfrak{p}}$ as the unique $W$-section $\underline{y}(0)$ in $T_{p^t}(\underline{z})$. Since $\sigma_\mathfrak{p}$ lies in the decomposition group at $\mathfrak{p}$ and is the restriction to $H$ of the arithmetic Frobenius automorphism of the completed maximal unramified extension $F$ of $K_\mathfrak{p}$, the existence of a $p$-isogeny $\underline{z}^{\sigma_\mathfrak{p}} \to \underline{z}$ as $\Gamma_0(N)$-structures follows from Theorem 6.2 below (which does not depend on the preceding discussion).

The next two basic results will help us to construct elements of Hom groups. This is essential in the proof of (5.4) when $p|m$.

**Theorem 6.1.** *Let $E$ and $E'$ be elliptic curves over $W$ with supersingular reduction and CM by $\mathscr{O}_K$, so $p$ is inert or ramified in $K$. Let $f : E'_{/W_n} \to E_{/W_n}$ be an isogeny, with $n \geq 0$.*

- *If $n \geq 1$ in the inert case and $n \geq 2$ in the ramified case, then the map $f$ has degree divisible by $p^2$ if and only if $f \bmod \pi^{n+1}/p = [p] \circ g$ for some other isogeny $g : E'_{/W/(\pi^{n+1}/p)} \to E_{/W/(\pi^{n+1}/p)}$.*
- *For any $n \geq 0$ the map $[p] \circ f$ always lifts (uniquely) to an isogeny over $W/p\pi^{n+1}$.*
- *If $p$ is inert in $K$, $n = 0$, and $f$ has degree $pr$ for $r \geq 1$ not divisible by $p$, then $f$ does* not *lift to an isogeny over $W/\pi^2$.*

Note that the uniqueness of liftings follows from the second part of Theorem 2.1 over the artinian quotients of $W$. The last part of the theorem is crucial for success in the inert case with $t$ odd, and it is not needed for any other cases. It should also be noted that Theorem 6.1 rests on the fact that $W$ is the completion of a maximal *unramified* extension of $\mathscr{O} = \mathbf{Z}_p \otimes_\mathbf{Z} \mathscr{O}_K$ (i.e., has algebraically closed residue field and *no* ramification over $\mathscr{O}$), as otherwise the theorem is generally false.

*Proof.* Since $W$ is the completion of the maximal unramified extension of $\mathscr{O}$, we may (and do) take our uniformizer $\pi$ of $W$ to come from $\mathscr{O}$.

For the first part of the theorem, we first want to prove that if $f$ has degree divisible by $p^2$, then $f \bmod \pi^{n+1}/p$ factors through the isogeny $[p]$ on $E' \bmod \pi^{n+1}/p$. Let $\Gamma_{/W}$ denote a fixed Lubin-Tate formal $\mathscr{O}$-module of height 1 and let $R = \mathrm{End}_{W_0}(\Gamma_{/W_0})$, a maximal order in a quaternion algebra over $\mathbf{Q}_p$. We can (and do) fix isomorphisms of formal $\mathscr{O}$-modules $E[p^\infty] \simeq \Gamma$ and $E'[p^\infty] \simeq \Gamma$. By means of these isomorphisms, the map $f[p^\infty]$ induced by $f$ on $p$-divisible groups is converted into an *endomorphism* of $\Gamma_{/W_n}$ as a formal group (*not* necessarily respecting the $\mathscr{O}$-structure). It is proven in [Gr1] that the endomorphism ring of $\Gamma_{/W_n}$ is exactly the subring $\mathscr{O} + \pi^n R$ inside of the endomorphism ring $R$ of the closed fiber of $\Gamma$. In this way, $f$ gives rise to an element $\alpha \in \mathscr{O} + \pi^n R$ with reduced norm $\mathrm{N}(\alpha)$ divisible by $p^2$. By the Serre-Tate theorem relating the deformation theory of elliptic curves with that of their $p$-divisible groups, as well as the fact that infinitesimal deformations are automatically compatible with $\Gamma_0(N)$-structures when $p \nmid N$ (provided such compatibility holds over the residue field), the condition that $f \bmod \pi^{n+1}/p$ factor through $[p]$ is exactly the statement that $\alpha$ as an element of $\mathscr{O} + (\pi^n/p)R$ be divisible by $p$ (note that $\pi^n/p \in \mathscr{O}$ in the ramified case since we require $n \geq 2$ in this case). Now we are faced with a problem in quaternion algebras.

For the factorization aspect of the first part of the theorem, we aim to show for any $n \geq 1$ in the inert case and any $n \geq 2$ in the ramified case, $p(\mathscr{O} + (\pi^n/p)R) = p\mathscr{O} + \pi^n R$ contains the elements $\alpha \in \mathscr{O} + \pi^n R$ of reduced norm divisible by $p^2$. Any element of $R$ lies inside of the valuation ring $\mathscr{O}_L$ of a quadratic extension $L$ of $\mathbf{Q}_p$ which lies inside of $R \otimes_{\mathbf{Z}_p} \mathbf{Q}_p$. Moreover, the reduced norm of such an element coincides with its relative norm from $L$ down to $\mathbf{Q}_p$. An element of $\mathscr{O}_L$ with norm down to $\mathbf{Z}_p$ divisible by $p^2$ is certainly divisible by $p$ in $\mathscr{O}_L$, so we can definitely write $\alpha = p\beta$ for some $\beta \in R$. Thus, for $\beta \in R$ with $p\beta \in \mathscr{O} + \pi^n R$ we must prove $\beta \in \mathscr{O} + (\pi^n/p)R$. Choose an element $r \in R$ such that $R = \mathscr{O} \oplus \mathscr{O}r$, so if $\beta = a + br$ with $a, b \in \mathscr{O}$ then $pb \in \pi^n \mathscr{O}$ and we want $b \in (\pi^n/p)\mathscr{O}$. This is obvious.

Now consider the assertion that $[p] \circ f$ lifts to $W/p\pi^{n+1}$ for any $n \geq 0$. Using the Serre-Tate theorem and the description in [Gr1] of endomorphism rings of $\Gamma$ over $W_r$'s as recalled above, we wish to prove that any element $\alpha \in \mathscr{O} + \pi^n R$ has the property that $p\alpha \in \mathscr{O} + p\pi^n R$. This is obvious.

Finally, consider the last part of the theorem. Once again using Serre-Tate and [Gr1], we want to show that if $\alpha \in R$ has reduced norm in $\mathbf{Z}_p$ equal to a unit multiple of $p$, then $\alpha$ does *not* lie in the subring $\mathscr{O} + \pi R$. Suppose that $\alpha \in \mathscr{O} + \pi R$. Since $\alpha$ is not a unit in $R$ (as its norm down to $\mathbf{Z}_p$ is not a unit), its image in $R/pR$ is not a unit. When $p$ is inert in $K$, so we may take $\pi = p$, the non-unit image of $\alpha \in \mathscr{O} + pR$ in $R/pR$ lies inside of the field $\mathscr{O}/p$ and hence $\alpha$ has vanishing image in $R/pR$. We conclude that $\alpha \in pR$, so $\alpha$ has reduced norm divisible by $p^2$, contrary to hypothesis. ■

The case of ramified $p$ and odd $t$ will also require the following theorem.

**Theorem 6.2.** *Let $E$ and $E'$ be as in Theorem 6.1, but assume $p$ is ramified in $K$. Let $\varphi$ denote the Frobenius endomorphism on the completion $W$ of the maximal unramified extension of $\mathscr{O}$. Let $(\cdot)^\varphi$ denote the operation of base change by $\varphi$ on $W$-schemes. For any positive integer $d$ and any non-negative integer $n$ there is a natural isomorphism of groups*

$$\mathrm{Hom}_{W_n}(E', E)_{\deg d} \simeq \mathrm{Hom}_{W_{n+1}}(E'^{\varphi}, E)_{\deg pd}. \tag{6.3}$$

*Moreover, if we endow $E$ and $E'$ with $\Gamma_0(N)$-structures and give $E'^{\varphi}$ the base change $\Gamma_0(N)$-structure, then this bijection carries $\Gamma_0(N)$-compatible maps to $\Gamma_0(N)$-compatible maps.*

*Proof.* As usual, we may (and do) choose a uniformizer of $W$ to be taken from a uniformizer of $\mathscr{O} = \mathscr{O}_K \otimes_{\mathbf{Z}} \mathbf{Z}_p$. Consider the $p$-torsion $E'[p]$, a finite flat group scheme over $W$ with order $p^2$. The action of $\mathscr{O}_K$ on this factors through an action of $\mathscr{O}_K/p = \mathscr{O}_K/\mathfrak{p}^2 \simeq \mathscr{O}/\pi^2$, where $\mathfrak{p}$ is the unique prime of $\mathscr{O}_K$ over $p$. Thus, the torsion subscheme $E'[\mathfrak{p}]$ coincides with the kernel of the multiplication map by $\pi$ on the formal $\mathscr{O}$-module $E'[p^\infty]$. In particular, this torsion subscheme is finite flat of order $p$ over $W$. Let $E'' = E'/E'[\mathfrak{p}]$, so there is a natural degree $p$ isogeny $E' \to E''$ whose reduction is uniquely isomorphic to the relative Frobenius map on the closed fiber of $E'$. Since $p \nmid N$ there is a unique $\Gamma_0(N)$-structure on $E''$ compatible via $E' \to E''$ with any given $\Gamma_0(N)$-structure on $E'$. Consider the degree $p$ dual isogeny $\psi : E'' \to E'$. Let $\psi_e = \psi \bmod \pi^{e+1}$ for any $e \geq 0$.

We claim that if $f : E'_{/W_n} \to E_{/W_n}$ is an isogeny then $f \circ \psi_n$ lifts to $W_{n+1}$ (with degree clearly divisible by $p$), and conversely every isogeny over $W_{n+1}$ of degree divisible by $p$ arises via this construction (from a necessarily unique $f$). This will establish the theorem with $E''$ in the role of $E'^{\varphi}$ (with the $\Gamma_0(N)$-aspect easily checked by unwinding the construction process), and we will then just have to construct an isomorphism of elliptic curves $E'' \simeq E'^{\varphi}$ over $W$ compatible with the recipe for $\Gamma_0(N)$-structures induced from $E'$. Using Serre-Tate, our problem for relating $E''$ and $E$ is reduced to one on the level of $p$-divisible groups.

If $\Gamma = E'[p^\infty]$, so $\Gamma$ is a formal $\mathscr{O}$-module of height 1 over $W$ and we write $R$ for $\mathrm{End}_{W_0}(\Gamma)$, then multiplication by $\pi$ makes sense on $\Gamma$ (whereas it does *not* make sense on the level of elliptic curves in case $\mathfrak{p}$ is not principal). Since all formal $\mathscr{O}$-modules of height 1 over $W$ are (non-canonically) isomorphic, so there exists a $W$-isomorphism $E[p^\infty] \simeq \Gamma$, by using [Gr1] as in the proof of Theorem 6.1 we can reduce our lifting problem to showing that any element $\alpha \in \mathscr{O} + \pi^n R$ has the property that $\alpha\pi^\vee \in \mathscr{O} + \pi^{n+1} R$ (with $\pi^\vee$ corresponding to the dual isogeny to $\pi$ with respect to the Cartier-Nishi self-duality of the $p$-divisible group of any elliptic curve, such as $E'_{/W}$), and conversely that any element of $\mathscr{O} + \pi^{n+1}R$ with reduced norm divisible by $p$ necessarily has the form $\alpha\pi^\vee$ for some $\alpha \in \mathscr{O} + \pi^n R$.

The self-duality of $E'$ induces complex conjugation on $\mathscr{O}_K$, and hence induces the unique non-trivial automorphism on $\mathscr{O}$ as a $\mathbf{Z}_p$-algebra. We conclude that $\pi^\vee$ is a unit multiple of $\pi$ in $\mathscr{O}$. Thus, we are reduced to the algebra problem of proving that $(\mathscr{O} + \pi^n R)\pi \subseteq R$ is exactly the set of elements in $\mathscr{O} + \pi^{n+1}R$ with reduced norm divisible by $p$. Since $\pi R = R\pi$ (due to the uniqueness of maximal orders in finite-dimensional division algebras over local fields), we need to prove that $\pi\mathscr{O} + \pi^{n+1}R$ is the set of elements in $\mathscr{O} + \pi^{n+1}R$ with reduced norm divisible by $p$. Since $\pi$ has reduced norm $p$ and an element of $\mathscr{O} + \pi^{n+1}R$ not in $\pi\mathscr{O} + \pi^{n+1}R$ is a unit in $R$ and hence has unit reduced norm, we're done.

It remains to construct a $W$-isomorphism $E'' \simeq E'^{\varphi}$ compatible with $\Gamma_0(N)$-recipes from $E'$. Over $W/\pi$ we have a canonical isomorphism $\overline{E}'' \simeq \overline{E}'/\ker(\mathrm{Frob}) \simeq \overline{E}'^{(p)} \simeq \overline{(E'^{\varphi})}$ which is visibly "$\Gamma_0(N)$-compatible" (with respect to $\overline{E'}$). Thus, we just have to lift *this* isomorphism to $W$. By Serre-Tate, it suffices to make a lift on the level of $p$-divisible groups. For $\Gamma = E'[p^\infty]$ we have $E''[p^\infty] \simeq \Gamma/\Gamma[\pi]$ and $E'^{\varphi}[p^\infty] \simeq \Gamma^\varphi$, so we seek a $W$-isomorphism $\Gamma/\Gamma[\pi] \simeq \Gamma^\varphi$ lifting the canonical isomorphism $\overline{\Gamma}/\overline{\Gamma}[\pi] \simeq \overline{\Gamma}^{(p)}$ over $W/\pi$. In other words, we seek to construct a $p$-isogeny $\Gamma \to \Gamma^\varphi$ over $W$ which lifts the relative Frobenius over $W/\pi$. This assertion is intrinsic to $\Gamma$. Since $\Gamma$ is, up to non-canonical isomorphism over $W$, the unique formal $\mathscr{O}$-modules of height 1 over $W$, it suffices to make such a $p$-isogeny lifting the relative Frobenius for *one* formal $\mathscr{O}$-module of height 1 over $W$. By base change compatibility, it suffices to do the construction for a single formal $\mathscr{O}$-module over $\mathscr{O}$. Since $\mathscr{O}$ residue field is $\mathbf{F}_p$, as $p$ is ramified in $\mathscr{O}_K$, over $\mathscr{O}$ we seek a suitable *endomorphism* of a formal $\mathscr{O}$-module of height 1. Using a Lubin-Tate formal group $\mathscr{F}_{\pi,f}$ for a uniformizer $\pi$ of $\mathscr{O}$ and the polynomial $f(X) = \pi X + X^p$, the endomorphism $[\pi]$ does the job. ∎

The nonzero contributions to (5.4) require some control on closed fibers, and this is provided by:

**Lemma 6.3.** *For $\underline{z}$ in $T_r(\underline{x^\sigma})$, the closed fiber of $T_{p^t}(\underline{z})$ is supported at a single point in $\underline{X}(W/\pi)$, corresponding to the $\Gamma_0(N)$-structure $\underline{z}_0$ when $t$ is even and $\underline{z}_0^{(p)}$ when $t$ is odd.*

*Proof.* First consider the inert case. When $t$ is even then $\underline{y}(0)$ makes sense and we have seen that its closed fiber is $\underline{z}_0$. When $t$ is odd then $\underline{y}(1)$ makes sense and we have seen that its closed fiber is $\underline{z}_0^{(p)}$. Thus, for the inert case we just have to check that $\underline{y}(s)_0$ in $\underline{X}(W/\pi)$ only depends on $s \bmod 2$. From the explicit construction, we get $\underline{y}(s)_0$ from $\underline{z}$ as $\Gamma_0(N)$-structures over $W/\pi$ by $s$-fold iteration of the process of passing to quotients by the unique order $p$ subgroup scheme (i.e., the kernel of the relative Frobenius) in our elliptic curves. But over a field (such as $W/\pi$) of characteristic $p$, going through two steps of this process is exactly the same as passing to the quotient by $p$-torsion, so it brings us back to where we began! Thus, the construction only depends on $s \bmod 2$.

Now consider the ramified case. This goes essentially as in the preceding paragraph (where we barely used the property of $p$ being inert). The only difference is that our divisor has contributions from both even and odd values of $s$. Looking back at where $\underline{y}(s)$ came from, we see that the contribution for $s \equiv t \bmod 2$ is $\underline{z}_0$ for $t$ even and $\underline{z}_0^{(p)}$ for $t$ odd. Meanwhile, for $s \not\equiv t \bmod 2$ our setup is obtained from an order $p^{s+1}$ quotient on the $\Gamma_0(N)$-structure $\underline{z}_0$, so this again corresponds to $\underline{z}_0$ for $t$ even and to $\underline{z}_0^{(p)}$ for $t$ odd. ∎

One important consequence of Lemma 6.3 is that the intersection number $(\underline{x}.T_{p^t}(\underline{z}))$ in (5.4) is non-zero if and only if $\underline{x}_0 \simeq \underline{z}_0$ for even $t$ and if and only if $\underline{x}_0 \simeq \underline{z}_0^{(p)}$ for odd $t$ (or in perhaps more uniform style, $\underline{x}_0 \simeq \underline{z}_0^{(p^t)}$). Here, isomorphisms are as $\Gamma_0(N)$-structures.

We need one more result before we can prove (5.4). The following theorem rests on the fact that the points $\underline{y}(s) \in \underline{X}(W(s))$ in the support of $T_{p^t}(\underline{z})$ are quasi-canonical liftings (in the sense of [Gr1]).

**Theorem 6.4.** *For a point $\underline{z}$ in $T_r(\underline{x^\sigma})$ and a point $\underline{y}(s)$ in the divisor $T_{p^t}(\underline{z})$ with $s > 0$, we have*

$$(\underline{x}.\underline{y}(s)) = (1/2)|\mathrm{Isom}_{W/\pi}(\underline{y}(s)_0, \underline{x}_0)|.$$

This result is [GZ, Ch III, Prop 6.1]. Our (very different) proof uses deformation theory.

*Proof.* We can assume that there exists an isomorphism $\iota : \underline{x}_0 \simeq \underline{y}(s)_0$ as $\Gamma_0(N)$-structures over $W/\pi$, as otherwise both sides of the desired equation vanish. Now *fix such an isomorphism* $\iota$.

Since the $p$-divisible group underlying $\underline{x}$ is a "canonical" lifting of its closed fiber (i.e., its endomorphism ring is a maximal order in a quadratic extension of $\mathbf{Z}_p$) while the $p$-divisible group underlying $\underline{y}(s)$ for $s > 0$ is merely a level $s$ quasi-canonical lifting in the sense of [Gr1] (so its endomorphism ring is a non-maximal order), by [Gr1, Prop 5.3(3)] we see that for $s > 0$ the $p$-divisible group of $\underline{y}(s)$ over $W(s)$ is *not* isomorphic modulo $\pi_s^2$ to the $p$-divisible group of $\underline{x}_{/W(s)}$ *as deformations of* $\underline{x}_0$ (i.e., in a manner respecting the map induced by $\iota$ on $p$-divisible groups). Consider the universal formal deformation ring $A$ for $\underline{x}_0$ on the category of complete local noetherian $W$-algebras with residue field $W/\pi$.

Abstractly $A \simeq W[\![t]\!]$, and we get $\mathrm{Spec}(W) \to \mathrm{Spec}(A)$ and $\mathrm{Spec}(W(s)) \to \mathrm{Spec}(A)$ corresponding to $\underline{x}$ and $\underline{y}(s)$ respectively. A key technical point is that both of these maps are *closed immersions* (i.e., the ring maps are surjective). In fact, the corresponding ring maps $A \to W$ and $A \to W(s)$ send $t$ to a uniformizer (using Lubin-Tate theory for $\underline{x}$ and [Gr1, Prop 5.3(3)] for $\underline{y}(s)$). Since the subscheme of $\mathrm{Spec}(A)$ cut out by $\underline{x}$ corresponds to the principal ideal generated by $t - \pi u$ for some unit $u \in W^\times$ and the map $A \twoheadrightarrow W(s)$ corresponding to $\underline{y}(s)$ sends $t$ to $\pi_s u_s$ for a unit $u_s \in W(s)^\times$, we compute that the scheme-theoretic intersection of our two closed subschemes in $\mathrm{Spec}(A)$ is $\mathrm{Spec}(W(s)/(\pi_s u_s - \pi u)) \simeq \mathrm{Spec}(W(s)/\pi_s)$ since $\pi/\pi_s$ lies in the maximal ideal of $W(s)$. In geometric terms, the closed subschemes $\underline{x}$ and $\underline{y}(s)$ in $\underline{X}$ are tranverse.

Let $\overline{\Gamma} = \mathrm{Aut}(\underline{x}_0)/\{\pm 1\}$, so we want $d \stackrel{\mathrm{def}}{=} |\overline{\Gamma}|$ to equal $(\underline{x}.\underline{y}(s))$ (assuming this intersection number is non-zero). From Lemma 4.2 and the proof of Theorem 4.1, we see that $\overline{\Gamma}$ acts faithfully on $A \simeq W[\![T]\!]$ and that $\widehat{\mathscr{O}}_{\underline{X},\underline{x}_0}$ is naturally identified with the subring of invariants $A^{\overline{\Gamma}} \simeq W[\![t]\!]$, where $t = \mathrm{Norm}_{\overline{\Gamma}}(T)$. Consider the map $A \to W(s)$. This map depends on $\iota$, and we showed above that it is surjective. The action of $\overline{\Gamma}$ on $A$ is compatible with making permutations on the choices of $\iota$, so the given map $A \to W(s)$ sends $\gamma(T)$ to a uniformizer of $W(s)$ for all $\gamma \in \overline{\Gamma}$. In particular, the map $W[\![t]\!] \simeq \widehat{\mathscr{O}}_{\underline{X},\underline{x}_0} \to W(s)$ corresponding to $\underline{y}(s)$ sends $t$ to an element of normalized order in $W(s)$ equal to $|\overline{\Gamma}| = d$, whence we see that the closed subscheme $\underline{y}(s) \hookrightarrow \underline{X}$ is *not* generally $\mathrm{Spec}(W(s))$ but rather corresponds to the order of index $d$ in $W(s)$. Likewise, the surjective map $W[\![t]\!] \to W$ corresponding to $\underline{x}$ sends $t$ to an element of normalized order $d$ in $W$. Thus, the scheme-theoretic intersection $\underline{x} \cap \underline{y}(s)$ in $\underline{X}$ is the spectrum of the artinian quotient

$$W(s)/(\pi_s^d(\mathrm{unit}) - \pi^d u). \tag{6.4}$$

Since $W(s)$ is totally ramified over $W$ of degree $p^s + p^{s-1} > 1$ in the inert case and of degree $p^s > 1$ in the ramified case, we see that $\pi_s^d$ has strictly smaller order than $\pi^d$ in $W(s)$, so the artinian quotient (6.4) has length $d$, as desired. ∎

We are now in position to prove Theorem 5.1 when $p|m$.

When $s = 0$ occurs, we have $\underline{z} \neq \underline{x}$ when $t$ is even. Indeed, since $\underline{z}$ is $p^t$-isogenous to itself (via $p^{t/2}$) and hence occurs as a point in $T_{p^t}(\underline{z}) \subseteq T_m(\underline{x^\sigma})$, the vanishing of $r_{\mathscr{A}}(m)$ forces $\underline{z} \neq \underline{x}$. Thus, $\underline{y}(0) = \underline{z}$ in $\underline{X}(W)$ is distinct from $\underline{x}$ for even $t$. Likewise, when $t$ is odd and the $s = 0$ case occurs (i.e., $p$ is ramified in $K$), then $\underline{y}(0) = \underline{z}^{\sigma_{\mathfrak{p}}}$ is again distinct from $\underline{x}$. The reason is that $\underline{z}^{\sigma_{\mathfrak{p}}}$ is $p$-isogenous to $\underline{z}$, whence (by multiplying against $p^{(t-1)/2}$) is $p^t$-isogenous to $\underline{z}$, so if $\underline{y}(0) = \underline{x}$ then $\underline{x}$ would occur in $T_{p^t}(\underline{z}) \subseteq T_m(\underline{x^\sigma})$, contradicting that $r_{\mathscr{A}}(m) = 0$. Theorem 4.1 therefore gives us a formula for $(\underline{x}.\underline{y}(0))$ in all cases when $s = 0$ occurs.

Combining this with Lemma 6.3 and Theorem 6.4, we can compute $(\underline{x}.T_{p^t}(\underline{z}))$ as a sum of various terms, with the contribution from $s = 0$ being treated separately. We get: for inert $p$,

$$(\underline{x}.T_m(\underline{z})) = \begin{cases} \frac{1}{2}\sum_{n \geq 0} \#\,\mathrm{Hom}_{W_n}(\underline{z},\underline{x})_{\deg 1} + \frac{t}{2} \cdot \frac{1}{2}\#\,\mathrm{Hom}_{W/\pi}(\underline{z},\underline{x})_{\deg 1}, & t \text{ even} \\ \frac{t+1}{2} \cdot \frac{1}{2}\#\,\mathrm{Hom}_{W/\pi}(\underline{z},\underline{x})_{\deg p}, & t \text{ odd} \end{cases} \tag{6.5}$$

and for ramified $p$,

$$(\underline{x}.T_m(\underline{z})) = \begin{cases} \frac{1}{2}\sum_{n\geq 0}\#\operatorname{Hom}_{W_n}(\underline{z},\underline{x})_{\deg 1} + t\cdot\frac{1}{2}\#\operatorname{Hom}_{W/\pi}(\underline{z},\underline{x})_{\deg 1}, & t \text{ even} \\ \frac{1}{2}\sum_{n\geq 0}\#\operatorname{Hom}_{W_n}(\underline{z}^{\sigma}_{\mathfrak{p}},\underline{x})_{\deg 1} + t\cdot\frac{1}{2}\#\operatorname{Hom}_{W/\pi}(\underline{z}^{\sigma_{\mathfrak{p}}},\underline{x})_{\deg 1}, & t \text{ odd}\end{cases} \tag{6.6}$$

with the $\sum_{n\geq 0}(\dots)$ term corresponding to $s=0$ contributions. For odd $t$ in the ramified case, note also that $\underline{z}^{\sigma_{\mathfrak{p}}}$ has closed fiber $\underline{z}_0^{(p)}$.

One aspect of (6.5) and (6.6) which perhaps requires some further explanation is in the case of odd $t$ and inert $p$, for which we need to explain why a degree $p$ map $\underline{z}_0 \to \underline{x}_0$ as $\Gamma_0(N)$-structures is the "same" as an isomorphism $\underline{z}_0^{(p)} \simeq \underline{x}_0$ of $\Gamma_0(N)$-structures; this latter isomorphism data is what one naturally gets when applying Lemma 6.3 and Theorem 6.4 with $\underline{y}(s)_0 \simeq \underline{z}_0^{(p)}$ for odd $s$. The point is that there is a *unique* order $p$ subgroup scheme in a supersingular elliptic curve over a field of characteristic $p$, the quotient by which is the Frobenius base change. Thus, there is only one possible kernel for a degree $p$ map $\underline{z}_0 \to \underline{x}_0$, so the passage between degree $p$ maps $\underline{z}_0 \to \underline{x}_0$ and isomorphisms $\underline{z}_0^{(p)} \simeq \underline{x}_0$ (respecting $\Gamma_0(N)$-structures) is immediate.

It remains to identify (6.5) and (6.6) with the right side of (5.4) in all four cases (depending on the parity of $t$ and whether $p$ is inert or ramified in $K$). For even $t$ and inert $p$, use multiplication by $p^{t/2}$ to lift isomorphisms to $p^t$-isogenies over thicker artinian bases by repeated application of the second part of Theorem 6.1, and use the first part of Theorem 6.1 to ensure that iteration of this construction gives the right side of (5.4). A similar argument using multiplication by $p^{(t-1)/2}$ and the third part of Theorem 6.1 takes care of odd $t$ and inert $p$; the role of the final part of Theorem 6.1 is to ensure that the sum on the right side of (5.4) in such cases has vanishing terms for $n > (t-1)/2$. The case of even $t$ and ramified $p$ goes by an argument as in the case of even $t$ and inert $p$, upon noting that the second part of Theorem 6.1 causes $[p]\circ f$ to lift from $W_n$ to $W_{n+2}$ since $\operatorname{ord}_K(p) = 2$ in the ramified case. The most subtle case of all is odd $t$ and ramified $p$, for which Theorem 6.2 (and Theorem 6.1) provides the ability to translate this case of (6.6) into the form on the right side of (5.4).

## 7. Application of a construction of Serre

With Theorem 5.1 settled, the next task is to explicate the right side of (5.1) in terms of quaternion algebras (still maintaining the assumption $r_{\mathscr{A}}(m) = 0$; that is, the ideal class $\mathscr{A}$ corresponding to $\sigma$ under class field theory contains no integral ideals of norm $m$). The simplest case is when $p$ is split in $K$, for then $\underline{x}$ and $\underline{x^{\sigma}}$ are Serre-Tate canonical lifts of their closed fibers, so

$$\operatorname{Hom}_{W_n}(\underline{x^{\sigma}},\underline{x}) = \operatorname{Hom}_W(\underline{x^{\sigma}},\underline{x}) = \operatorname{Hom}_F(x^{\sigma},x)$$

for all $n \geq 0$ (see (5.5)). Thus, the right side of (5.1) vanishes since $r_{\mathscr{A}}(m) = 0$. Thus, for the purpose of explicitly computing the right side of Theorem 5.1 in terms of quaternion algebras, we lose no generality in immediately restricting to the case in which $p$ is not split in $K$, so $p \nmid N$, $\mathscr{O} = \mathbf{Z}_p \otimes_{\mathbf{Z}} \mathscr{O}_K$ is the ring of integers of a quadratic extension of $\mathbf{Q}_p$, and $\underline{x}$ has supersingular reduction (as does $\underline{x^{\sigma}}$).

We need the following classical result (which is noted below [GZ, Ch III, Prop 7.1] and for which we give a non-classical proof).

**Lemma 7.1.** *The ring $R = \operatorname{End}_{W/\pi}(\underline{x})$ is an order in a quaternion division algebra $B$ over $\mathbf{Q}$ which is non-split at exactly $p$ and $\infty$. More specifically, $\mathbf{Z}_p \otimes_{\mathbf{Z}} R$ is the maximal order in the division algebra $\mathbf{Q}_p \otimes_{\mathbf{Q}} B$ and for $\ell \neq p$ the order $\mathbf{Z}_\ell \otimes_{\mathbf{Z}} R$ is conjugate to the order $\left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix} \in M_2(\mathbf{Z}_\ell) \,|\, c \equiv 0 \bmod N\right\}$ in $\mathbf{Q}_\ell \otimes_{\mathbf{Q}} B \simeq M_2(\mathbf{Q}_\ell)$*

*Proof.* By the classical theory of supersingular elliptic curves, we know the first part. Thus, the whole point is to work out the local structure, and for this we use Tate's isogeny theorem for abelian varieties over finite fields. Let $k_0$ be a sufficiently large finite field over which there is a model $\underline{x}_0$ of the Heegner diagram $\underline{x}$ mod $\pi$ for which all "geometric" endomorphisms of the underlying elliptic curve (ignoring the étale level structure) are defined, so in particular $R = \mathrm{End}_{k_0}(\underline{x}_0)$. By Tate's isogeny theorem (with the trivial adaptation to keep track of the level structure), for $\ell \neq p$ the natural map

$$\mathbf{Z}_\ell \otimes_{\mathbf{Z}} R \to \mathrm{End}_{\mathbf{Z}_\ell[G_{k_0}]}(\underline{x}_0[\ell^\infty])) \tag{7.1}$$

is an isomorphism, where the right side denotes the ring of $\ell$-adic Tate module endomorphisms which are Galois-equivariant and respect the $\ell$-part of the level structure. Since the left side has $\mathbf{Z}_\ell$-rank equal to 4 and the right side is an order inside of $M_2(\mathbf{Z}_\ell)$, we conclude that the Galois group acts through scalars (this only requires injectivity of the Tate map, and this is completely elementary, so we haven't yet used the full force of Tate's theorem). Thus, we can replace $\mathbf{Z}_\ell[G_{k_0}]$ with $\mathbf{Z}_\ell$ on the right side of (7.1), and this yields the desired matrix description for all $\ell \neq p$.

Now consider the situation at $p$. Since there is no level structure involved, the claim is that if $E$ is a supersingular elliptic curve over a finite field $k_0$ such that all "geometric" endomorphisms of $E$ are defined over $k_0$, then $\mathbf{Z}_p \otimes_{\mathbf{Z}} \mathrm{End}_{k_0}(E)$ is the *maximal* order in a quaternion division algebra over $\mathbf{Q}_p$. The crucial point is that the action of a Frobenius element $\phi_{k_0}$ on $E$ is through an *integer*. Indeed, since $\mathbf{Z}$ is a direct summand of the endomorphism algebra of $E$ it follows that $\phi_{k_0}$ acts as an integer on $E$ as long as it acts as an $\ell$-adic integer on an $\ell$-adic Tate module of $E$. This integrality property follows from the fact that (7.1) is an isomorphism, since $\mathbf{Q}_\ell$ is the center of $\mathbf{Q}_\ell \otimes_{\mathbf{Q}} B$.

It follows that on the height 2 formal group $\Gamma = E[p^\infty]_{/\overline{k}_0}$ of $E$ over an algebraic closure $\overline{k}_0$, every endomorphism commutes with the $k_0$-Frobenius and hence is actually *defined* over $k_0$. By the theory of formal groups over *separably closed fields*, the endomorphism ring of $\Gamma$ is a maximal order in a quaternion division algebra over $\mathbf{Q}_p$. But we have just seen that this endomorphism ring coincides with that of the $p$-divisible group of $E$ over $k_0$. Thus, it suffices to prove that the natural map

$$\mathbf{Z}_p \otimes_{\mathbf{Z}} \mathrm{End}_{k_0}(E) \to \mathrm{End}_{k_0}(E[p^\infty]) \tag{7.2}$$

is an isomorphism. Notice that if we replace $p$ with $\ell \neq p$ then this is exactly the usual statement of Tate's isogeny theorem (up to the identification of the category of $\ell$-divisible groups over $k_0$ with a certain category of $\mathbf{Z}_\ell[G_{k_0}]$-modules).

This "$\ell = p$" case of Tate's theorem is proven by essentially the same exact method as Tate's theorem: the only change in the proof is that one has to use Dieudonné modules to replace the use of Tate modules. Nearly every theorem in [Mum] concerning Tate modules also works (usually with the same proof) for Dieudonné modules, and this enables one to extend various results (such as computing the characteristic polynomial of Frobenius over a finite field) to the $p$-part in characteristic $p$.

■

Using the "closed fiber" functor, we have a natural injection $\mathscr{O}_K = \mathrm{End}_W(\underline{x}) \to \mathrm{End}_{W/\pi}(\underline{x}) = R$ which gives rise to a $\mathbf{Q}$-linear injection $K \to B$. By Skolem-Noether, there exists $j \in B$ such that $jaj^{-1} = \overline{a}$ for all $a \in K$, where $a \mapsto \overline{a}$ is the non-trivial automorphism of $K$ over $\mathbf{Q}$. In particular, $j^2 \in K^\times$ and any other such element $j'$ of $B$ lies in $K^\times j$ (as $K$ is its own centralizer in $B$). Thus, $B_- = Kj$ is intrinsic to $K \hookrightarrow B$, so the decomposition $B = K \oplus B_-$ is intrinsic. Observe that $B_- = Kj$ can also be intrinsically described as the set of elements $b \in B$ such that $ba = \overline{a}b$ for all $a \in K \hookrightarrow B$.

If $b \in B$ is written as $b = b_+ + b_-$ according to this decomposition then $b_- = cj$ for some $c \in K$. Since $(cj)^2 = c\sigma(c) \in \mathbf{Q}$, when $b_- \neq 0$ it generates a quadratic field over $\mathbf{Q}$ in which its conjugate is $-b_-$. Thus, we can compute the reduced norm $\mathrm{N}(b)$ as

$$\mathrm{N}(b) = \mathrm{N}(b_+)\mathrm{N}(1 + (c/b_+)j) = \mathrm{N}(b_+)(1 + (c/b_+)j)(1 - (c/b_+)j) = \mathrm{N}(b_+) + \mathrm{N}(b_-)$$

when $b_+ \neq 0$, and the case $b_+ = 0$ is trivial. In other words, the reduced norm is additive with respect to the decomposition $B = K \oplus B_-$.

Just as the decomposition $B = K \oplus B_-$ is intrinsic, for our $p$ which is non-split in $K$ (so $\mathrm{Gal}(K/\mathbf{Q}) \simeq \mathrm{Gal}(K_p/\mathbf{Q}_p)$) it follows that tensoring with $\mathbf{Q}_p$ gives the analogous decomposition for the non-split $\mathbf{Q}_p \otimes_{\mathbf{Q}} B$. If we define $R_p \stackrel{\text{def}}{=} \mathbf{Z}_p \otimes_{\mathbf{Z}} R \simeq \mathrm{End}_{W/\pi}(\underline{\widehat{x}})$ (the isomorphism being Tate's isomorphism (7.2)), then by [Gr1, Prop 4.3] we see that the subring of $p$-divisible group endomorphisms lifting to $W_n$ is given by those

$b = b_+ + b_- \in R_p$ satisfying $D\,\mathrm{N}(b_-) \equiv 0 \bmod p\mathrm{N}(v)^n$, where $\mathrm{N}(v)$ denotes the ideal-theoretic norm of the unique prime of $\mathscr{O}_K$ over $p$. Thus,

$$\mathrm{End}_{W_n}(\widehat{\underline{x}}) = \{b \in R_p \mid D\mathrm{N}(b_-) \equiv 0 \bmod p\mathrm{N}(v)^n\}.$$

The Serre-Tate lifting theorem ensures that $\mathrm{End}_{W_n}(\underline{x})$ consists of those elements of $R = \mathrm{End}_{W_0}(\underline{x})$ lifting to a $W_n$-endomorphism of the *p-divisible* group $\widehat{\underline{x}}$ of $\underline{x}$. Thus,

$$\mathrm{End}_{W_n}(\underline{x}) = \{b \in R \mid D\mathrm{N}(b_-) \equiv 0 \bmod p\mathrm{N}(v)^n\}. \tag{7.3}$$

Describing $\mathrm{Hom}_{W_n}(\underline{x^\sigma}, \underline{x})$ is more subtle, since it rests on an interesting tensor construction of Serre's which unfortunately appears to not be explained adequately in the literature outside of the context of abelian varieties over a field (which is inadequate for applications such as our present situation where we have to work over artin local rings). There is a more general discussion of Serre's construction in [Gi], but that is also not adequate for our needs. We now develop Serre's tensor construction in a general setting, essentially to make functorial sense of an isomorphism $\underline{x^\sigma} \simeq \mathrm{Hom}(\mathfrak{a}, \underline{x}) \simeq \mathfrak{a}^{-1} \otimes_{\mathscr{O}_K} \underline{x}$ in a relative situation, where $\mathfrak{a}$ represents the ideal class $\mathscr{A}$ associated to $\sigma$ under the Artin isomorphism. From a more algebraic point of view, the problem is that we do not yet have a recipe *over* $W$ for constructing $\underline{x^\sigma}$ in terms of $\underline{x}$ (unless $\sigma \in \mathrm{Gal}(H/K)$ lies in the decomposition group at $v$). It is this recipe that we must intrinsically construct; see Corollary 7.11 for the answer. The basic idea is to construct $\underline{x^\sigma}$ as an "$\mathscr{O}_K$-tensor product" of $\underline{x}$ against a suitable fractional ideal of $\mathscr{O}_K$.

To motivate things, if $S = \mathrm{Spec}(\mathbf{C})$ and $E^{\mathrm{an}} \simeq \mathbf{C}/\mathfrak{b}$ for a fractional ideal $\mathfrak{b}$ of $\mathscr{O}_K$, then we expect to have an $\mathscr{O}_K$-linear analytic isomorphism

$$(\mathfrak{a} \otimes_{\mathscr{O}_K} E)^{\mathrm{an}} \overset{?}{\simeq} \mathfrak{a} \otimes_{\mathscr{O}_K} (\mathbf{C}/\mathfrak{b}) \simeq \mathbf{C}/\mathfrak{a}\mathfrak{b}. \tag{7.4}$$

This sort of construction on the analytic side *does* describe Galois actions $x \mapsto x^\sigma$ on Heegner points when viewed as $\mathbf{C}$-points of a modular curve (if $\mathfrak{a}$ is a prime ideal and $\sigma$ is a *geometric* Frobenius at this prime), but if we are to have any hope of working with such Galois twisting on the level of $W$-points (or *anything* other than $\mathbf{C}$-points), we have to find a non-analytic mechanism for constructing $\mathbf{C}/\mathfrak{a}\mathfrak{b}$ from $\mathbf{C}/\mathfrak{b}$.

The device for algebraically constructing "$\mathfrak{a} \otimes_{\mathscr{O}_K} E$" is due to Serre, and involves representing a functor. Rather than focus only on $\mathscr{O}_K$-module objects in the construction, we prefer to give a construction valid for more general coefficient rings, since abelian varieties of dimension $> 1$ (or in positive characteristic) tend to have non-commutative endomorphism algebras and Drinfeld modules are naturally "module schemes" over rings such as $\mathbf{F}_p[t]$. To include such a diversity of phenomena, we will construct a scheme $M \otimes_A \mathfrak{M}$ for an arbitrary associative ring $A$, a finite projective right $A$-module $M$, and an arbitrary left $A$-module scheme $\mathfrak{M}$ over any base scheme $S$ (i.e., $\mathfrak{M}$ is a commutative $S$-group scheme endowed with a left $A$-action).

In the interest of saving space, we will omit most of the proofs of the assertions we make below concerning Serre's tensor construction; we hope to provide a more detailed development elsewhere. Many proofs are mechanical, though one must carry them out in the correct order to avoid complications, and a few arguments require input from Dieudonné theory. Here is the setup for the basic existence result.

Let $S$ be a scheme and let $A$ be an associative ring with identity. Let $M$ be a finite projective right $A$-module with dual left $A$-module $M^\vee = \mathrm{Hom}_A(M, A)$ of right $A$-linear homomorphisms (with $(a.\phi)(m) \overset{\mathrm{def}}{=} a\phi(m)$ for $\phi \in M^\vee$, $a \in A$, $m \in M$). By using the analogous duality construction for left modules, there is a natural map $M \to M^{\vee\vee}$ which is an isomorphism of right $A$-modules (use projectivity of $M$).

**Theorem 7.2.** *Let $\mathfrak{M}$ be a left $A$-module scheme. The functor $T \rightsquigarrow M \otimes_A \mathfrak{M}(T) \simeq \mathrm{Hom}_A(M^\vee, \mathfrak{M}(T))$ on $S$-schemes is represented by a commutative group scheme over $S$, denoted $M \otimes_A \mathfrak{M}$ or $\mathrm{Hom}_A(M^\vee, \mathfrak{M})$, and $M \otimes_A (\cdot)$ carries closed immersions to closed immersions, surjections to surjections, and commutes with formation of fiber products.*

*For each of the following properties $\mathbf{P}$ of $S$-schemes, if $\mathfrak{M}$ satisfies property $\mathbf{P}$ then so does $M \otimes_A \mathfrak{M}$: quasi-compact and quasi-separated, locally finitely presented, finitely presented, locally finite type, separated, proper, finite, locally quasi-finite, quasi-finite and quasi-separated, smooth, étale, flat. In particular, if $\mathfrak{M}$ is finite locally free over $S$ then so is $M \otimes_A \mathfrak{M}$.*

In Theorem 7.7, we'll see some nice applications of this result in the context of finite flat group schemes (beware that without an additional constancy condition on the "$A$-rank" of $M$, appropriately defined, the functor $M \otimes_A (\cdot)$ may have a slightly funny effect on the order of a finite locally free group scheme).

By considering left $A$-modules as right $A^{\mathrm{opp}}$-modules and vice-versa, the property of being a finite projective module is preserved and everything we say below carries over with trivial modifications upon switching the words "left" and "right".

*Proof.* First we establish representability, and then will consider flatness; the rest is left as an exercise in functorial criteria, etc. When $M$ is a finite free right $A$-module, it is clear that an $r$-fold fiber product of $\mathfrak{M}$ does the job. In general, choose a finite presentation of the finite projective left $A$-module $M^\vee$:

$$A^{\oplus r} \to A^{\oplus s} \to M^\vee \to 0. \tag{7.5}$$

Applying $\mathrm{Hom}_A(\cdot, \mathfrak{M}(T))$ yields a left exact sequence, so by representability of scheme-theoretic kernels we deduce representability in general.

I am grateful to Serre for suggesting the following argument for the flatness property; this is much simpler than my original argument. Since $M$ is projective, it is a direct summand of some $A^{\oplus n}$ as a right $A$-module. Thus, $M \otimes_A \mathfrak{M}$ is a direct factor of the $S$-flat $\mathfrak{M}^n$ as an $S$-scheme. It therefore suffices to check that if $X$ and $Y$ are $S$-schemes with $Y(S)$ non-empty and $X \times_S Y$ is $S$-flat, then $X$ is $S$-flat. We may assume $S$ is local and $X$ is affine, and can replace $Y$ by an affine open neighborhood of the closed point of the identity section. We thereby get two commutative ring extensions $R \to B, C$ with $B \otimes_R C$ flat over $R$ and $R \to C$ having a section, so $C = R \oplus I$ as $A$-modules. Thus, $B$ is a direct summand of the $R$-flat $B \otimes_R C$ as an $R$-module and hence is $R$-flat. ∎

Now we wish to study fibers and exact sequences. This requires us to first define what it means to say that a finite projective right $A$-module $M$ has constant rank $r$ over $A$. When $A$ is commutative, the meaning is clear (i.e., the vector bundle $\widetilde{M}$ on $\mathrm{Spec}(A)$ has constant rank) and can be formulated by saying that for any ring map $\phi : A \to k$ to a (commutative) field $k$, the base change $M \otimes_{A,\phi} k$ is $r$-dimensional as a $k$-vector space. Moreover, it suffices to demand this condition for algebraically closed fields $k$. The correct notion in the non-commutative case is unclear in general. However, rings acting on (reasonable) schemes are special. For example, if $\mathfrak{M}$ is locally finite type over $S$ then $A$ acts on the tangent spaces at geometric points of $\mathfrak{M}$, and hence we get many maps from $A$ to matrix algebras (or more conceptually, central simple algebras of finite dimension) over residue fields at geometric points on $S$. This example suggests a definition which, while surely "wrong" for finite projective modules over general associative rings, works well for the situations we really care about (i.e., module schemes locally of finite type over a base).

**Definition 7.3.** We say that a finite projective right $A$-module $M$ has *constant rank* $r$ if, for every map $\phi : A \to C$ to a finite-dimensional central simple algebra over an algebraically closed field $k = Z(C)$, with $\dim_k C = d_C^2$, the finite right $C$-module $M \otimes_{A,\phi} C$ has length $d_C r$. In other words, as a right $C$-module $M \otimes_{A,\phi} C$ is isomorphic to an $r$-fold direct sum of copies of $C$.

*Remark* 7.4. When $A$ is commutative, this is equivalent to the more familiar notion from commutative algebra (i.e. a rank $r$ vector bundle on $\mathrm{Spec}(A)$) since a finite projective module over a local ring is free and $C$ in Definition 7.3 is isomorphic to a direct sum of $d_C$ copies of its unique simple module $I_C$. Dropping the commutativity assumption, since $\dim_k I_C = d_C$ in the notation of Definition 7.3, the $k$-length of a finite $C$-module is equal to $d_C$ times its $C$-length. Thus, in Definition 7.3 one could equivalently require $\dim_k M \otimes_{A,\phi} C = r \dim_k(C)$ where $k$ is the center of $C$. This latter formula provides a definition that works without requiring $k$ to be algebraically closed, shows that the quantifiers in Definition 7.3 involve no set theory or universe issues, and makes it clear that the finite projective dual left $A$-module $M^\vee$ is of constant rank $r$ (defined in terms of $C \otimes_{\phi,A} M^\vee$) if and only if $M$ is of constant rank $r$ (in the sense of Definition 7.3). One sees this latter compatibility with the help of the natural isomorphism of left $C$-modules

$$C \otimes_{\phi,A} M^\vee \simeq (M \otimes_{A.\phi} C)^\vee$$

(with the right side a $C$-dual) defined by $c \otimes \ell \mapsto (m \otimes c' \mapsto c\phi(\ell(m))c')$. This is particularly useful when considering (Picard or Cartier) duality for left and right $A$-module schemes which are abelian schemes or finite locally free.

In order to apply Serre's tensor construction to abelian schemes and their torsion subschemes, we need to record some more properties.

**Theorem 7.5.** *Let $\mathfrak{M} \to S$ be a locally finite type left $A$-module scheme, and let $M$ be a finite projective right $A$-module.*

- *If the $S$-fibers of $\mathfrak{M}$ are connected, then so are the $S$-fibers of $M \otimes_A \mathfrak{M}$. In particular, if $\mathfrak{M}$ is an abelian scheme over $S$ then so is $M \otimes_A \mathfrak{M}$.*
- *If $\mathfrak{M}$ has fibers over $S$ of dimension $d$ and $M$ has constant rank $r$ over $A$ then $M \otimes_A \mathfrak{M}$ has fibers of dimension $dr$.*
- *Let $0 \to \mathfrak{M}' \to \mathfrak{M} \to \mathfrak{M}'' \to 0$ be a short exact sequence of locally finitely presented $S$-flat left $A$-module schemes. Then applying $M \otimes_A (\cdot)$ yields another short exact sequence of such $A$-module schemes, and in particular the map $M \otimes_A \mathfrak{M} \to M \otimes_A \mathfrak{M}''$ is faithfully flat.*

We will be particularly interested in the case that $A$ is the ring of integers of a number field and $M$ is a fractional ideal (corresponding to the case of rank 1), so $M \otimes_A E$ is an elliptic curve for $E \to S$ an elliptic curve (with $S = \mathrm{Spec}(W)$, $\mathrm{Spec}(W/\pi^{n+1})$, etc.).

Before we address the special case of the behavior of Serre's construction on finite locally free commutative group schemes, we digress to study an important example which recovers (7.4).

Suppose that $\mathfrak{M}$ is an abelian variety over $\mathbf{C}$ and $A$ is an associative ring acting on $\mathfrak{M}$ on the left. By the analytic theory, we functorially have $\mathfrak{M}^{\mathrm{an}} \simeq V/\Lambda$ where $V = \mathrm{Tan}_0(\mathfrak{M}^{\mathrm{an}})$ is the universal covering and $\Lambda = \mathrm{H}_1(\mathfrak{M}^{\mathrm{an}}, \mathbf{Z})$. In particular, there is a nature left $A$-action on $V$ commuting with the $\mathbf{C}$-action and with respect to which the lattice $\Lambda$ is stable. It then makes sense to form $M \otimes_A V$ and $M \otimes_A \Lambda$, and it follows by using a right $A$-linear isomorphism $M \oplus N \simeq A^{\oplus r}$ for suitable $N$ that $M \otimes_A V$ is a finite-dimensional $\mathbf{C}$-vector space and $M \otimes_A \Lambda$ is a finitely generated *closed* subgroup which is co-compact and hence a lattice. Thus, the quotient $(M \otimes_A V)/(M \otimes_A \Lambda)$ makes sense as a complex torus, and it is only natural to guess that this must be $(M \otimes_A \mathfrak{M})^{\mathrm{an}}$ (which we know to be a complex torus, since $M \otimes_A \mathfrak{M}$ is already known to be an abelian variety). This guess is of course correct, and is crucial for the usefulness of Serre's construction (see the *proof* of Corollary 7.11).

**Theorem 7.6.** *With notation as above, there is an natural $A$-linear isomorphism of $\mathbf{C}$-analytic Lie groups $(M \otimes_A V)/(M \otimes_A \Lambda) \simeq (M \otimes_A \mathfrak{M})^{\mathrm{an}}$ which is functorial in both $\mathfrak{M}$ and $M$.*

We will use this theorem in the case $\dim \mathfrak{M} = 1$ (i.e., elliptic curves).

If $\underline{x}$ is a "level $N$" Heegner diagram over an $\mathscr{O}_K[1/N]$-scheme $S$ (such as $W$ or $\mathbf{C}$), it makes sense to apply Serre's tensor construction $\mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, \cdot)$ to this diagram, provided that this construction carries cyclic isogenies of degree $N$ to cyclic isogenies of degree $N$. While such a property is easily checked over $\mathbf{C}$ by means of the preceding theorem, in order to work over a base which is not a subfield of $\mathbf{C}$ (such as a non-field ring like $W$ or an artinian quotient of $W$) we need to do some more work.

First we record a general result, the proof of which requires Dieudonné theory.

**Theorem 7.7.** *Let $\mathfrak{M}$ be a left $A$-module scheme which is finite locally free over $S$, so its Cartier dual $\mathfrak{M}^\vee$ is a right $A$-module scheme. Then $M \otimes_A \mathfrak{M}$ is also finite locally free over $S$, of rank $d^r$ if $\mathfrak{M}$ has constant rank $d$ over $S$ and $M$ has constant rank $r$ over $A$.*

*Quite generally, in terms of Cartier duality there is a natural isomorphism $(M \otimes_A \mathfrak{M})^\vee \simeq \mathfrak{M}^\vee \otimes_A M^\vee$.*

Using the above theorems and some additional arguments, one can deduce the following two consequences.

**Corollary 7.8.** *Let $S$ be a henselian local scheme. The functor $M \otimes_A (\cdot)$ respects formation of the connected-étale sequence of a finite locally free commutative group scheme over $S$.*

*Moreover, if $S$ is an arbitrary scheme with all points of positive characteristic $p$ and $\mathfrak{M} \to S$ is a left $A$-module scheme which is an abelian scheme having ordinary (resp. supersingular) fibers, then $M \otimes_A \mathfrak{M} \to S$ has the same property.*

**Corollary 7.9.** *Let $f : \mathfrak{M} \to \mathfrak{N}$ be an isogeny (i.e., finite locally free map) of constant rank $d$ between left $A$-module schemes which are flat and locally finitely presented schemes over $S$. Then*

$$1_M \otimes f : M \otimes_A \mathfrak{M} \to M \otimes_A \mathfrak{N}$$

*is an isogeny, and the natural map $M \otimes_A \ker(f) \to \ker(1_M \otimes f)$ is an isomorphism.*

*When $f$ has constant degree $d$ and $M$ has constant rank $r$ over $A$, then $1 \otimes f$ has constant degree $d^r$. Likewise, when $f$ is étale then so is $1_M \otimes f$.*

The applications to Heegner points require that we keep track of *cyclicity* of kernels of isogenies. This amounts to:

**Corollary 7.10.** *Let $M$ be a finite projective right $A$-module of constant rank 1 and let $\mathfrak{M}$ be a left $A$-module scheme whose underlying group scheme over $S$ is finite étale and "cyclic" of constant order $d$. Then the same holds for $M \otimes_A \mathfrak{M}$.*

An important corollary of these considerations is that for a Heegner diagram $\underline{x}$ of level $N$ over any $\mathscr{O}_K[1/N]$-scheme $S$ and any fractional ideal $\mathfrak{a}$ of $K$, applying $\mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, \cdot)$ to this diagram yields another Heegner diagram. Now we can finally prove the main result:

**Corollary 7.11.** *Let $\underline{x}$ be a Heegner diagram over $W$ arising from a Heegner diagram $x$ over $H$, and let $\sigma \in \mathrm{Gal}(H/K)$ correspond to the ideal class represented by a fractional ideal $\mathfrak{a}$ of $K$. There exists an isomorphism $\underline{x^\sigma} \simeq \mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, \underline{x}) \simeq \mathfrak{a}^{-1} \otimes_{\mathscr{O}_K} \underline{x}$ as Heegner diagrams over $W$.*

*Proof.* Note that the assertion only depends on the isomorphism classes of the Heegner diagrams, and by Theorem 2.6 two Heegner diagrams over $W$ are $W$-isomorphic if and only if they become isomorphic over some extension field of the fraction field of $W$. Since Serre's tensor construction is of formation compatible with *arbitrary* base change, by using an embedding $W \hookrightarrow \mathbf{C}$ as $\mathscr{O}_H$-algebras (which exists by cardinality considerations on transcendence degrees over $H$) it suffices to prove the result with $W$ replaced by $\mathbf{C}$. From the analytic description of Galois action on Heegner points, if $x$ over $\mathbf{C}$ is analytically isomorphic to

$$\mathbf{C}/\mathfrak{b} \to \mathbf{C}/\mathfrak{b}\mathfrak{n}^{-1}$$

then $x^\sigma$ over $\mathbf{C}$ is analytically isomorphic to the analogous diagram with $\mathfrak{b}$ replaced by $\mathfrak{a}^{-1}\mathfrak{b}$. Since $\mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, \mathfrak{M}) \simeq \mathfrak{a}^{-1} \otimes_{\mathscr{O}_K} \mathfrak{M}$ for an $A$-module scheme $\mathfrak{M}$, if we use functoriality then Theorem 7.6 ensures that applying $\mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, \cdot)$ to $x$ over $\mathbf{C}$ yields a Heegner diagram which is analytically and hence (by GAGA) algebraically isomorphic to the diagram of $x^\sigma$ over $\mathbf{C}$, using $(\mathfrak{a} \otimes_{\mathscr{O}_K} \mathbf{C})/(\mathfrak{a} \otimes_{\mathscr{O}_K} \Lambda) \simeq \mathbf{C}/\mathfrak{a}\Lambda$ for any $\mathscr{O}_K$-lattice $\Lambda$ in $\mathbf{C}$ (with this isomorphism functorial in $\Lambda \hookrightarrow \mathbf{C}$). ∎

Having explored the properties of Serre's tensor construction, we're ready to apply it:

**Theorem 7.12.** *Let $\mathfrak{a}$ be an ideal in the ideal class $\mathscr{A}$ and let $\sigma \in \mathrm{Gal}(H/K)$ correspond to this ideal class. Then there is an isomorphism of groups*

$$\mathrm{Hom}_{W_n}(\underline{x^\sigma}, \underline{x}) \simeq \mathrm{End}_{W_n}(\underline{x}) \cdot \mathfrak{a} \subseteq B \tag{7.6}$$

*under which an isogeny $\phi : \underline{x^\sigma} \to \underline{x}$ corresponds to an element $b \in B$ with $\mathrm{N}(b) = \deg(\phi)\mathrm{N}(\mathfrak{a})$ as ideals in $\mathbf{Z}$.*

In addition to needing the preceding theory to justify the isomorphism in (7.6), particularly over an artin local base such as $W_n$, we will need to work a little more to keep track of degrees. Postponing the proof of Theorem 7.12 for a short while, note that by using an abstract isomorphism $\underline{x^\sigma} \simeq \mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, \underline{x})$ over $W$ (and hence over any $W$-scheme, such as $W_n$) as follows from Corollary 7.11, we obtain for any $W$-scheme $S$ that

$$\mathrm{Hom}_S(\underline{x^\sigma}, \underline{x}) \simeq \mathrm{Hom}_S(\mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, \underline{x}), \underline{x}) \simeq \mathfrak{a} \otimes_{\mathscr{O}_K} \mathrm{End}_S(\underline{x}) \tag{7.7}$$

where $\mathscr{O}_K$ acts on $\mathrm{End}_S(\underline{x})$ through "inner composition" and the second isomorphism arises from:

**Lemma 7.13.** *Let $A$ be an associative ring and $M$ a finite projective left $A$-module. For any left $A$-module scheme $\mathfrak{M}$ over $S$ and any commutative $S$-group scheme $G$, view the group $\mathrm{Hom}_S(\mathfrak{M}, G)$ as a right $A$-module via the $A$-action on $\mathfrak{M}$. Then the natural map $\xi_M : \mathrm{Hom}_S(\mathfrak{M}, G) \otimes_A M \to \mathrm{Hom}_S(\mathrm{Hom}_A(M, \mathfrak{M}), G)$ defined functorially by $\xi_M(\phi \otimes m) : f \mapsto \phi(f(m))$ (on the level of points in $S$-schemes) is well-defined and an isomorphism.*

This lemma is a simple exercise in definition-chasing (keep in mind how we *defined* the $A$-action on $\mathrm{Hom}_S(\mathfrak{M},\mathfrak{N})$), using functoriality with respect to product decompositions in $M$ and the existence of an isomorphism $M\oplus N\simeq A^{\oplus r}$ to reduce to the trivial case $M=A$.

Strictly speaking, this lemma does not include the case of Hom groups between the Heegner diagrams $\underline{x}$ and $\underline{x^\sigma}$, since these are data of elliptic curves with some additional level structure. Since $\mathfrak{a}$ is typically not principal, we need to be a bit careful to extract what we wish from Lemma 7.13. The simple trick to achieve this is to observe that the $\mathscr{O}_K$-module Hom groups of $\Gamma_0(N)$-structures which we really want to work with are described as kernels of certain $\mathscr{O}_K$-linear maps of Hom modules to which Lemma 7.13 *does* apply. By exactness of tensoring against a flat module (commuting with formation of kernels), we get the second isomorphism in (7.7) as a canonical isomorphism.

Let's now record an easy mild strengthening of a special case of Lemma 7.13, from which we'll be able to compute $\mathrm{End}_{W_n}(\underline{x^\sigma})$ in terms of $\mathrm{End}_{W_n}(\underline{x})$.

**Lemma 7.14.** *Let $A$ be an associative ring, $M$ and $M'$ two finite projective left $A$-modules. Let ${M'}^\vee$ denote the right module of left-linear maps from $M'$ to $A$. For any two left $A$-module schemes $\mathfrak{M}$ and $\mathfrak{M}'$ over a base $S$, view the group $\mathrm{Hom}_S(\mathfrak{M},\mathfrak{M}')$ as a right $A$-module via the $A$-action on $\mathfrak{M}$ and as a left $A$-module via the action on $\mathfrak{M}'$. Then the natural map*

$$\xi_{M',M}: {M'}^\vee\otimes_A \mathrm{Hom}_S(\mathfrak{M},\mathfrak{M}')\otimes_A M\to \mathrm{Hom}_S(\mathrm{Hom}_A(M,\mathfrak{M}),\mathrm{Hom}_A(M',\mathfrak{M}'))$$

*defined functorially by $(\xi_{M',M}(\ell'\otimes\phi\otimes m))(f): m'\mapsto \ell'(m')\phi(f(m))$ (on the level of points in $S$-schemes) is well-defined and an isomorphism.*

*In particular, there is a natural isomorphism $M^\vee\otimes_A \mathrm{End}_S(\mathfrak{M})\otimes_A M\simeq \mathrm{End}_S(\mathrm{Hom}_A(M,\mathfrak{M}))$ given by $\ell\otimes\phi\otimes m\mapsto (f\mapsto \ell'(\cdot)\cdot\phi(f(m)))$, and this is an isomorphism of associative rings.*

As a consequence of this lemma, we get a natural ring isomorphism

$$\mathrm{End}_S(\underline{x^\sigma})\simeq \mathfrak{a}^{-1}\otimes_{\mathscr{O}_K}\mathrm{End}_S(\underline{x})\otimes_{\mathscr{O}_K}\mathfrak{a} \tag{7.8}$$

for any $W$-scheme $S$, where the ring structure on the right is the obvious one obtained via the pairing $\mathfrak{a}^{-1}\times\mathfrak{a}\to\mathscr{O}_K$. Now we have enough machinery to prove Theorem 7.12.

*Proof.* (of Theorem 7.12). Specializing the preceding preliminary considerations to the case $S=W_n$ and identifying $\mathrm{End}_{W_n}(\underline{x})$ with a subring of $\mathrm{End}_{W_0}(\underline{x})=R$, if we view $\mathscr{O}_K=\mathrm{End}_W(\underline{x})$ as embedded in $R$ via reduction it follows that making $\mathscr{O}_K$ act on $\mathrm{End}_{W_n}(\underline{x})$ through "inner composition" as in Lemma 7.13 corresponds to making it act by right multiplication in $R$ (since multiplication in $R$ is defined as composition of morphisms and all modern algebraists define composition of morphisms to begin on the right). This therefore yields an isomorphism of groups $\mathrm{Hom}_{W_n}(\underline{x^\sigma},\underline{x})\simeq \mathrm{End}_{W_n}(\underline{x})\otimes_{\mathscr{O}_K}\mathfrak{a}$ where $\mathscr{O}_K\subseteq R$ acts through right multiplication on $\mathrm{End}_{W_n}(\underline{x})\subseteq R$. Since $\mathfrak{a}$ is an invertible $\mathscr{O}_K$-module and $B$ is torsion-free, the natural multiplication map $\mathrm{End}_{W_n}(\underline{x})\otimes_{\mathscr{O}_K}\mathfrak{a}\to B$ is an isomorphism onto $\mathrm{End}_{W_n}(\underline{x})\cdot\mathfrak{a}$.

It remains to chase degrees of isogenies. Since the degree of an isogeny over $W_n$ can be computed on the $W/\pi$-fiber, it suffices to work over the *field* $W/\pi$ and to show the following: given the data consisting of

- a supersingular elliptic curve $E$ over an algebraically closed field $k$ whose positive characteristic $p$ is a prime which is non-split in $K/\mathbf{Q}$,
- an action of $\mathscr{O}_K$ on $E$ (i.e., an injection of $\mathscr{O}_K$ into the quaternion division algebra $B=\mathbf{Q}\otimes_{\mathbf{Z}}\mathrm{End}_k(E)$),
- a fractional ideal $\mathfrak{a}$ of $K$,

under the isomorphism

$$\mathrm{End}_k(E)\cdot\mathfrak{a}\simeq \mathrm{Hom}_k(\mathrm{Hom}(\mathfrak{a},E),E) \tag{7.9}$$

defined by $\psi\cdot a\mapsto (f\mapsto\psi(f(a)))$ we claim that an isogeny $\phi:\mathrm{Hom}(\mathfrak{a},E)\to E$ corresponds to an element $b\in B$ with reduced norm $\deg(\phi)\mathrm{N}(\mathfrak{a})$. The delicate point here is that $\mathfrak{a}$ might not be a principal ideal in $\mathscr{O}_K$. The trick by means of which we will "reduce" to the principal ideal situation is to investigate what happens on the level of $\ell$-divisible groups for *every* prime $\ell$, since $\mathscr{O}_{K,\ell}\overset{\mathrm{def}}{=}\mathbf{Z}_\ell\otimes_{\mathbf{Z}}\mathscr{O}_K$ is semi-local (either a discrete valuation ring or a product of two discrete valuation rings) and hence every invertible ideal in this latter ring is principal.

To make things precise, fix an arbitrary prime $\ell$ of $\mathbf{Z}$. The $\ell$-part of $\deg(\phi)$ is exactly the degree of the isogeny $\phi_\ell$ induced by $\phi$ on $\ell$-divisible groups. To keep matters simple, note that everything depends only on the isomorphism class of $\mathfrak{a}$ as an $\mathscr{O}_K$-module, or in other words everything is compatible with replacing this ideal by another representative of the same ideal class. More specifically, if $\phi \mapsto b$ under (7.9) and $\phi_c$ is the composite of $\phi$ with the isomorphism $c^{-1} : \mathrm{Hom}(c\mathfrak{a}, E) \simeq \mathrm{Hom}(\mathfrak{a}, E)$ then $\deg(\phi) = \deg(\phi_c)$ and $\phi_c \mapsto cb$, with $\mathrm{N}b = \deg(\phi)\mathrm{N}\mathfrak{a} \Leftrightarrow \mathrm{N}(cb) = \deg(\phi_c)\mathrm{N}(c\cdot\mathfrak{a})$. Thus, we may assume without loss of generality that $\mathfrak{a}$ is an integral ideal and hence elements of this ideal act on $E$. If we replace $k$ with a sufficiently big finite subfield $k_0$ down to which $E$ descends and over which all "geometric" morphisms among $E$ and $\mathrm{Hom}_{\mathscr{O}_K}(\mathfrak{a}, E)$ are defined, we can consider the isomorphism $\mathrm{End}_{k_0}(E)\cdot\mathfrak{a} \simeq \mathrm{Hom}_{k_0}(\mathrm{Hom}(\mathfrak{a}, E), E)$. We now apply Tate's isogeny theorem *at any prime* $\ell$ to both sides. Using the language of $\ell$-divisible groups so as to treat all primes on an equal footing, tensoring both sides with $\mathbf{Z}_\ell$ yields an isomorphism

$$\mathrm{End}_{k_0}(E[\ell^\infty])\cdot\mathfrak{a}_\ell \simeq \mathrm{Hom}_{k_0}(\mathrm{Hom}(\mathfrak{a}, E)[\ell^\infty], E[\ell^\infty])$$

with $\mathfrak{a}_\ell = \mathfrak{a}\mathscr{O}_{K,\ell}$ an invertible ideal of $\mathscr{O}_{K,\ell} = \mathbf{Z}_\ell \otimes_{\mathbf{Z}} \mathscr{O}_K$.

By functoriality, it is clear that

$$\mathrm{Hom}(\mathfrak{a}, E)[\ell^\infty] = \varinjlim(\mathfrak{a}^{-1}\otimes_{\mathscr{O}_K} E[\ell^n]) = \varinjlim(\mathfrak{a}_\ell^{-1}\otimes_{\mathscr{O}_{K,\ell}} E[\ell^n]).$$

Corollary 7.9 ensures these tensored group schemes really do form an $\ell$-divisible group (of height 2). If we write $\mathfrak{a}_\ell^{-1}\otimes E[\ell^\infty]$ to denote this $\ell$-divisible group, then we have an isomorphism

$$\mathrm{End}_{k_0}(E[\ell^\infty])\cdot\mathfrak{a}_\ell \simeq \mathrm{Hom}_{k_0}(\mathfrak{a}_\ell^{-1}\otimes E[\ell^\infty], E[\ell^\infty]) \tag{7.10}$$

defined in the obvious manner. If $\pi \in \mathscr{O}_{K,\ell}$ is a generator of the invertible (principal!) ideal $\mathfrak{a}_\ell$ then every element $b_\ell$ of the left side of (7.10) can be written in the form $b_\ell = \psi_\ell\cdot\pi$ for a unique endomorphism $\psi_\ell$ of $E[\ell^\infty]$ and $b_\ell$ corresponds under (7.10) to the composite morphism $\phi_\ell : \mathfrak{a}_\ell^{-1}\otimes E[\ell^\infty] \simeq E[\ell^\infty] \to E[\ell^\infty]$, where the first step is the isomorphism of multiplication by $\pi$ and the second step is $\psi_\ell$. Thus, $\phi_\ell$ has degree equal to that of $\psi_\ell$, which in turn is just the reduced norm of $\psi_\ell$ (in the $\ell$-adic quaternion algebra $\mathbf{Q}_\ell\otimes_{\mathbf{Q}} B$). But since $b_\ell = \psi_\ell\pi$, this reduced norm is equal to $\mathrm{N}(b_\ell)/\mathrm{N}(\pi) = \mathrm{N}(b_\ell)/\mathrm{N}(\mathfrak{a}_\ell)$, thereby giving the $\ell$-part $\deg(\phi_\ell) = \mathrm{N}(b_\ell)/\mathrm{N}(\mathfrak{a}_\ell)$ of the desired "global" equality of $\mathbf{Z}$-ideals $\mathrm{N}b = \deg(\phi)\mathrm{N}\mathfrak{a}$. ∎

**Corollary 7.15.** *Assume $r_{\mathscr{A}}(m) = 0$ and $\gcd(m, N) = 1$. Choose a representative ideal $\mathfrak{a}$ for $\mathscr{A}$ which is prime to $p$.*

(1) *If $p$ is inert in $K$ and $v$ is a place of $H$ over $p$, then $q_v = p^2$ and*

$$(\underline{x}.T_m(\underline{x^\sigma})) = \sum_{b\in R\mathfrak{a}/\pm1,\,\mathrm{N}b=m\mathrm{N}\mathfrak{a}} \frac{1}{2}(1+\mathrm{ord}_p(\mathrm{N}b_-)). \tag{7.11}$$

(2) *If $p\mathscr{O}_K = \mathfrak{p}^2$ and $v$ is a place of $H$ over $p$, then $q_v = p^k$ where $k\in\{1,2\}$ is the order of $[\mathfrak{p}]$ in the class group of $K$ and*

$$(\underline{x}.T_m(\underline{x^\sigma})) = \sum_{b\in R\mathfrak{a}/\pm1,\,\mathrm{N}b=m\mathrm{N}\mathfrak{a}} \mathrm{ord}_p(D\mathrm{N}b_-)). \tag{7.12}$$

*Proof.* This result is [GZ, Ch III, Cor 7.4], for which the main inputs in the proof are Theorem 5.1, (7.3), and Theorem 7.12. We refer the reader to [GZ] for the short argument to put it all together (which requires that $\mathfrak{a}$ be prime to $p$). The only aspect on which we offer some further clarification is to explain conceptually why $b_-$ is always non-zero for the $b$'s under consideration and why $\mathrm{ord}_p(\mathrm{N}b_-)$ is odd when $p\nmid D$.

The non-vanishing of $b_-$ is simply because an element $b\in R\mathfrak{a}$ with $b_- = 0$ is an element of which lifts to $\mathrm{Hom}_{W_n}(\underline{x^\sigma}, \underline{x})$ for all $n\ge 0$, and hence corresponds to an element in $\mathrm{Hom}_W(\underline{x^\sigma}, \underline{x})$ by Serre-Tate and Grothendieck's existence theorem. But this latter Hom module vanishes due to the assumption $r_{\mathscr{A}}(m) = 0$. Thus, since $b\ne 0$ (as its reduced norm is $m\mathrm{N}\mathfrak{a}\ne 0$) we have $b_-\ne 0$. Meanwhile, when our non-split $p$ satisfies $p\nmid D$, so $p$ is inert $K$, the choice of $j$ (as in the unramified case of [Gr1]) may be made so that

$j^2 \in K^\times$ is a uniformizer at the place over $p$. Thus, any $b_- = cj \in Kj$ satisfies $\mathrm{N}(b_-) = \mathrm{N}_{K/\mathbf{Q}}(c)\mathrm{N}(j^2)$, where $\mathrm{N}_{K/\mathbf{Q}}(c)$ has even order at $p$ and $\mathrm{N}(j^2)$ is a uniformizer at $p$ (as $p$ is inert in $K$). Hence, $\mathrm{ord}_p(\mathrm{N}(b_-))$ is odd.

■

This completes our consideration of cases with $r_{\mathscr{A}}(m) = 0$.

## 8. Intersection theory via meromorphic tensors

In order to go beyond the local cases with $r_{\mathscr{A}}(m) = 0$, we need a stronger geometric technique for computing intersection pairings without using a moving lemma. This section is devoted to developing the necessary generalities in this direction, so modular curves and Heegner points play no role in this section. Unfortunately, the letters $K$, $D$, $F$ have already been assigned meaning in our earlier analysis, with $F$ being the fraction field of $W$. Since $W$ and the discriminant of $K/\mathbf{Q}$ will make no appearance in this section, *for this section only we reserve the letter $F$ to denote a global field and $D$ to denote a divisor on a curve.* We fix a proper smooth geometrically connected curve $X_{/F}$ (having nothing to do with $X_0(N)$). Let $D, D'$ be degree 0 divisors on $X$ whose common support $|D| \cap |D'|$, if non-empty, consists entirely of $F$-rational points. The example of interest to us is $X_0(N)_{/H}$, $D = (x) - (0)$ (resp. $D = (x) - (\infty)$) for some Heegner point $x \in X_0(N)(H)$ (note the two cusps are $H$-rational too), and $D' = T_m((x^\sigma) - (\infty))$ (resp. $D' = T_m((x^\sigma) - (0))$) for $\sigma \in \mathrm{Gal}(H/K)$, with $\gcd(m, N) = 1$. In this case, we see that any overlap of supports between $D$ and $D'$ certainly consists entirely of non-cuspidal $H$-rational points, and moreover such non-empty overlap occurs exactly when $r_{\mathscr{A}}(m) > 0$.

In our general setup, we wish to define a local symbol $\langle D, D'\rangle_v$ for each place $v$ of $F$ with the properties:

- it is bilinear in such pairs $\{D, D'\}$ (note the hypothesis $|D| \cap |D'| \subseteq X(F)$ is preserved under formation of linear combinations),
- it agrees with the canonical local height pairing of $D_v$ and $D'_v$ when $|D| \cap |D'| = \emptyset$,
- the sum $\sum_v \langle D, D'\rangle_v$ is equal to the canonical global height pairing.

We recall that in the case of disjoint support, the local height pairing was uniquely characterized by abstract properties (including functorial behavior with respect to change of the base field) and for non-archimedean $v$ is explicitly constructed via intersection theory using any regular proper model over the local ring $\mathscr{O}_{F,v}$ at $v$. For the generalization allowing $|D| \cap |D'|$ to consist of some $F$-rational points, the local terms $\langle D, D'\rangle_v$ will *not* be canonical, but rather will depend on a certain non-canonical *global* choice. Happily, the product formula will ensure that this global choice only affects local terms by amounts whose total sum is 0. This will retain the connection to the global height pairings (which is what we really care about).

In order to clarify the local nature of the construction, we will carry it out entirely in the context of a local field, only returning to the global case after the basic local construction has been worked out. Thus, we now write $F_v$ to denote a local (perhaps archimedean) field with normalized absolute value $|\cdot|_v$. Feel free to think of this as arising from completing a global field $F$ at a place $v$, but such a "global" model will play no role in our local construction. Let $\mathscr{O}_{F_v}$ denote the valuation ring of $v$ in the non-archimedean case. We fix $X_{v/F_v}$ a proper smooth and geometrically connected curve. The following definition is considered in [Gr2, §5].

**Definition 8.1.** If $x \in X_v(F_v)$ is a point, $f \in F_v(X_v)^\times$ is a non-zero rational function, and $t_x$ is a *fixed* uniformizer at $x$, we define

$$f[x] = f_{t_x}[x] \stackrel{\text{def}}{=} \frac{f}{t_x^{\mathrm{ord}_x(f)}}(x) \in F_v^\times.$$

This definition clearly depends on $t_x$ only through its non-zero image $\omega_x$ in the cotangent line $\mathrm{Cot}_x(X_v)$ at $x$, so we may denote it $f_{\omega_x}[x]$ instead. For non-archimedean $v$ the absolute value $|f[x]|_v$ only depends on this non-zero cotangent vector up to $\mathscr{O}_{F_v}^\times$-multiple. For technical reasons (e.g., the fact that various Hecke divisors do not have all support consisting of rational points), it is convenient to extend the definition in the absolute value aspect to the case in which $x \in X_v$ is merely a closed point and not necessarily $F_v$-rational (this mild generalization does not seem to appear in [Gr2]).

For any such $x$, the residue field $\kappa(x)$ at $x$ is a finite extension of $F_v$ and hence for a cotangent vector $\omega_x$ represented by a uniformizer $t_x$ at $x$, the value

$$f[x] = f_{\omega_x}[x] \stackrel{\text{def}}{=} (f/t_x^{\mathrm{ord}_x(f)})(x) \in \kappa(x)^\times$$

makes sense (depending again only on $\omega_x$), and when $\mathrm{ord}_x(f) = 0$ this is just $f(x)$. Thus, the absolute value $|f_{\omega_x}[x]|_v$ make sense, where we write $|\cdot|_v$ to also denote the unique extension of $|\cdot|_v$ to an absolute value on $\kappa(x)$, and it is just $|f(x)|_v$ when $\mathrm{ord}_x(f) = 0$. Extending by $\mathbf{Z}$-linearity and multiplicativity, we can define the symbol

$$|f_{\underline{\omega}}[D]|_v = \prod_{x \in |D|} |f_{\omega_x}[x]|_v^{\mathrm{ord}_x(D)}$$

for any divisor $D$ on $X_v$ for which we have *fixed* a set $\underline{\omega}$ of non-zero cotangent vectors at each $x \in |D| \cap |\mathrm{div}(f)|$ (keep in mind that $|f_\omega[x]|_v$ does not depend on $\omega \in \mathrm{Cot}_x(X_v)$ when $x \notin |\mathrm{div}(f)|$, so the choice of cotangent vector at such $x$'s really should be made when defining $|f_{\underline{\omega}}[D]|_v$ but actually doesn't matter). We will write $|f[D]|_v$ instead of $|f_{\underline{\omega}}[D]|_v$ when the choice of cotangent vectors is understood from context. We do *not* claim to define $f_{\underline{\omega}}[D]$, as this would make no sense if the support $|D|$ contains several non-rational points (whose residue fields are not canonically identified with each other).

We trivially have $|(fg)[D]|_v = |f[D]|_v \cdot |g[D]|_v$ with both sides depending on choices of cotangent vectors at points of $|D|$ where either $f$ or $g$ has a zero or pole, and we also have $|f[D + D']|_v = |f[D]|_v \cdot |f[D']|_v$ (assuming the same cotangent vector has been chosen relative to $D$ and $D'$ at any common point in their support). Now we are ready for the main definition, which we will justify shortly.

**Definition 8.2.** Let $D, D'$ be degree 0 divisors on $X_{v/F_v}$ with $|D| \cap |D'| = \{x_1, \ldots, x_n\} \subseteq X_v(F_v)$. *Choose* a set of cotangent vectors $\underline{\omega} = \{\omega_1, \ldots, \omega_n\}$ at each such point. For $f \in F_v(X_v)^\times$ such that $|D + \mathrm{div}(f)| \cap |D'| = \emptyset$, define

$$\langle D, D' \rangle_{v,\underline{\omega}} = \langle D + \mathrm{div}_{X_v}(f), D' \rangle_v - \log |f_{\underline{\omega}}[D']|_v, \tag{8.1}$$

where the first term on the right side is the canonical local height pairing for disjoint divisors of degree 0.

It is clear that this definition does not depend on the auxiliary choice of $f$, thanks to properties of the canonical local height pairing, so when $D$ and $D'$ have disjoint supports we see via the case $f = 1$ that this definition coincides with the canonical local height pairing for disjoint divisors of degree 0. The bilinearity in $D$ and $D'$ (relative to consistent choices of cotangent vectors) is clear.

A crucial observation is that *if* we begin in a global situation $X_{/F}$ with degree 0 divisors $D, D'$ on $X$ whose supports overlap precisely in global rational points $\{x_1, \ldots, x_n\} \subseteq X(F)$, then by choosing a set of *global* cotangent vectors $\underline{\omega}$ at the $x_j$'s and choosing a *global* $f \in F(X)^\times$ for which $D + \mathrm{div}(f)$ and $D'$ have disjoint support, we get

$$\begin{aligned}
\sum_v \langle D_v, D'_v \rangle_{v,\underline{\omega}} &= \sum_v \langle D + \mathrm{div}(f), D' \rangle_v - \sum_v \log |f_{\underline{\omega}}[D']|_v \\
&= \langle D + \mathrm{div}(f), D' \rangle - \sum_{x' \in |D'|} \mathrm{ord}_{x'}(D') \sum_v \sum_j \log |f_{\omega_{x'}}[x'_j]|_{v,j},
\end{aligned}$$

where the inner sum is over all factor fields $\kappa(x'_j)$ of the reduced ring $\kappa(x) \otimes_F F_v$ (on which the unique multiplicative norm extending $|\cdot|_v$ corresponds to an absolute value $|\cdot|_{v,j}$ on $\kappa(x'_j)$) and $x'_j$ runs over all points on $X_v$ over $x' \in X$. If $t_{x'}$ is a uniformizer representing $\omega_{x'}$ then

$$\sum_j \log |f_{\omega_{x'}}[x'_j]|_{v,j} = \log |\mathrm{N}_{\kappa(x')/F}(f_{\omega_{x'}}[x'])|_v,$$

so by the product formula for $\mathrm{N}_{\kappa(x')/F}(f_{\omega_{x'}}[x]) \in F^\times$ we get

$$\sum_v \langle D_v, D'_v \rangle_{v,\underline{\omega}} = \langle D + \mathrm{div}(f), D' \rangle = \langle D, D' \rangle,$$

recovering the canonical global height pairing.

It is very important for later purposes to observe that the local construction of $\langle\cdot,\cdot\rangle_{v,\underline{\omega}}$ is unaffected by replacing the choices of the $\omega_x$'s in the local theory with $\mathscr{O}_{F_v}^\times$-multiples. Somewhat more generally, if two degree 0 divisors $D, D'$ on $X_{v/F_v}$ satisfy $|D| \cap |D'| = \{x_1, \dots, x_n\} \subseteq X_v(F_v)$, and we define $\omega'_{x_j} = \alpha_j \omega_{x_j}$, then since $\mathrm{ord}_x(f) = -\mathrm{ord}_x(D)$ for all $x \in |D| \cap |D'|$ in (8.1) we get

$$\langle D, D'\rangle_{v,\underline{\omega}'} = \langle D, D'\rangle_{v,\underline{\omega}} - \sum_j \mathrm{ord}_{x_j}(D)\mathrm{ord}_{x_j}(D')\log|\alpha_j|_v. \tag{8.2}$$

This follows immediately from the definitions, using the fact that uniformizers $t_x$ and $t'_x$ representing $\omega_x$ and $\omega'_x = \alpha\omega_x$ are related by $t'_x = \alpha t_x + (\dots)$. Note that (8.2) is exactly in accordance with [GZ, (5.2), p. 250] (which treates the case of a one-point overlap), since our $\alpha$ is the reciprocal of that in [GZ].

An explication of (8.1) which plays an essential role in calculations at complex places is briefly described in [GZ, Ch II, §5.1] (in which the *normalized* absolute value $|\cdot|_v$ on $\mathbf{C}$ is denoted $|\cdot|^2$ for the usual reasons). Due to its importance, we want to justify that this alternative description really matches (8.1), and in particular really works for any local field at all (not just $\mathbf{C}$). The starting point for this alternative description is the observation that since $X_v$ is $F_v$-smooth, for any $x \in X_v(F_v)$ there is a small open neighborhood of $x$ in $X_v^{\mathrm{an}}$ which is analytically isomorphic to an open unit disc (so there is a plentiful supply of rational points near $x$ in the sense of the topology on $F_v$), and the ordinary local height pairing has nice continuity properties with respect to slightly modifying a divisor variable.

**Theorem 8.3.** *Let $D, D'$ be degree 0 divisors on $X_v$ with $|D| \cap |D'| = \{x_1, \dots, x_n\} \subseteq X_v(F_v)$. Choose $y_j \in X_v(F_v)$ near (but not equal to) $x_j$ in the topological space $X_v(F_v)$. Define $D_{\underline{y}}$ to be the degree 0 divisor obtained from $D$ by replacing all appearances of $x_j$ with $y_j$. Then*

$$\langle D, D'\rangle_{v,\underline{\omega}} = \lim_{\underline{y}\to\underline{x}} \left( \langle D_{\underline{y}}, D'\rangle_v - \sum_j \mathrm{ord}_{x_j}(D)\mathrm{ord}_{x_j}(D')\log|t_{x_j}(y_j)|_v \right),$$

*where $t_{x_j}$ is a uniformizer representing $\omega_{x_j}$. In particular, this limit actually exists.*

It is clear that the term in the limit is independent of the choice of uniformizer lifting each $\omega_{x_j}$.

*Proof.* Subtracting the left side (as defined in (8.1)) from the right side, we get the limit as $\underline{y} \to \underline{x}$ of

$$\langle \sum_j \mathrm{ord}_{x_j}(D) y_j - \sum_{x \in |\mathrm{div}(f)| - \underline{x}} \mathrm{ord}_x(f)x, D'\rangle_v - \sum_j \mathrm{ord}_{x_j}(D)\mathrm{ord}_{x_j}(D')\log|t_{x_j}(y_j)|_v$$
$$+ \left( \sum_{x' \in |D'| - \underline{x}} \mathrm{ord}_{x'}(D')\log|f(x')|_v + \sum_j \mathrm{ord}_{x_j}(D')\log|\frac{f}{t_{x_j}^{\mathrm{ord}_{x_j}(f)}}(x_j)|_v \right),$$

where the sum of the last two terms is $\log|f_{\underline{\omega}}[D']|_v$. Since $\mathrm{ord}_{x_j}(D) = -\mathrm{ord}_{x_j}(f)$, after some cancellation and noting that

$$\frac{f}{t_{x_j}^{\mathrm{ord}_{x_j}(f)}}(x_j) = \lim_{y_j \to x_j} \frac{f(y_j)}{t_{x_j}(y_j)^{\mathrm{ord}_{x_j}(f)}},$$

we are left with the limiting value of $-\langle \mathrm{div}_{X_v}(f)_{\underline{y}}, D'\rangle_v + \langle D'_{\underline{y}}, \mathrm{div}_{X_v}(f)\rangle_v$ as $\underline{y} \to \underline{x}$. Thus, it suffices to prove the general claim that if $D, D'$ are two degree 0 divisors on $X_v$ with overlap at the set of rational points $\underline{x}$, and $\underline{y}$ is a slight deformation of this set of points, then $\langle D_{\underline{y}}, D'\rangle_v - \langle D, D'_{\underline{y}}\rangle_v \to 0$ as $\underline{y} \to \underline{x}$ (with both terms ordinary canonical local height pairings).

By choosing a rational point $P$ away from these (which we avoid in the limiting process) and expressing $D$ and $D'$ as $\mathbf{Z}$-linear combinations of differences $Q - [\kappa(Q) : F_v]P$ for closed points $Q \in X_v$, we reduce to the following general situation. Let $x_1, x_2 \in X_v$ be two closed points distinct from $P \in X_v(F_v)$, with $\kappa_j = \kappa(x_j)$ the residue field at $x_j$. Define $\widetilde{x}_j = x_j$ if $x_j$ is not a rational point, and otherwise define $\widetilde{x}_j$ to be a rational point near $x_j$, with the requirement that $\widetilde{x}_j \neq x_j$ for both $j$'s when $x_1 = x_2$. Note in particular that $\widetilde{x}_j$ has residue field degree over $F_v$ equal to $[\kappa_j : F_v]$ in all cases. The main claim is that

$$\langle \widetilde{x}_1 - [\kappa_1 : F_v]P, x_2 - [\kappa_2 : F_v]P\rangle_v - \langle x_1 - [\kappa_1 : F_v]P, \widetilde{x}_2 - [\kappa_2 : F_v]P\rangle_v \to 0$$

as $\underline{\widetilde{x}} \to \underline{x}$ (the condition $\widetilde{x}_j \to x_j$ being a tautology if $x_j$ is not a rational point).

If neither $x_1$ nor $x_2$ is a rational point, there's nothing to say. If $x_1$ is a rational point but $x_2$ is not (or vice-versa), then by bilinearity we reduce to the continuity of the local height pairing when rational points within a fixed divisor are moved around in $X_v(F_v)$. Thus, we may now assume $x_1, x_2 \in X_v(F_v)$. By treating separately the cases in which we have $\widetilde{x}_j = x_j$ for some $j$ and when this does not hold for either $j$ (and in the latter case, distinguishing $x_1 = x_2$ from the case $x_1 \neq x_2$), one can use related continuity/bilinearity arguments.

■

Now we turn to explicating our modified local intersection pairing in the non-archimedean case. Two natural questions arise for non-archimedean $F_v$:

(1) Can we carry out the local construction using a pairing on the level of (suitable) divisors *not* necessarily of degree 0, somewhat generalizing the use of intersection theory for regular proper models in the case of *disjoint* effective divisors?

(2) Are there ways to package the $\mathscr{O}_{F_v}^{\times}$-multiple class of a non-zero cotangent vector by using higher (meromorphic) tensors?

Thanks to bilinearity, in order to address the first question, the central issue is to properly define $(x.x)_v$ for $x \in X_v(F_v)$ when a non-zero cotangent vector $\omega_x$ is chosen at $x$. More specifically, suppose that $\underline{X}_{v/\mathscr{O}_{F_v}}$ is a regular proper model for $X_v$, and let $\underline{x} \in \underline{X}_v(\mathscr{O}_{F_v})$ be the section arising from $x$ via scheme-theoretic closure. Since $\underline{X}_v$ is regular, this section $\underline{x}$ lies in the relative smooth locus over $\mathscr{O}_{F_v}$. We have the natural linear isomorphism among cotangent spaces $\mathrm{Cot}_x(X_v) \simeq F_v \otimes_{\mathscr{O}_{F_v}} \mathrm{Cot}_{\underline{x}}(\underline{X}_v)$, so given a non-zero $\omega_x \in \mathrm{Cot}_x(X_v)$ there exists an $\alpha_x \in F_v^{\times}$ unique up to $\mathscr{O}_{F_v}^{\times}$-multiple such that $\alpha_x \omega_x \in \mathrm{Cot}_x(X_v)$ is a basis of the rank 1 lattice $\mathrm{Cot}_{\underline{x}}(\underline{X}_v)$. In particular, we see that $\mathrm{ord}_v(\alpha_x)$ depends only on the data of $\omega_x$ (not $t_x$) and $\underline{X}_v$. We'll now see that the choice of $\underline{X}_v$ doesn't matter.

**Theorem 8.4.** *Fix a cotangent vector $\omega_x$ at every $x \in X_v(F_v)$, and let $\underline{\omega} = \{\omega_x\}_{x \in X_v(F_v)}$ denote the corresponding indexed collection. Define $(x.x)_{v,\underline{\omega}} = \mathrm{ord}_v(\alpha_x)$ for each $x \in X_v(F_v)$ as above. For distinct closed points $x', x'' \in X_v$ define $(x'.x'')_{v,\underline{\omega}} = (x'.x'')_v$ as in* (4.1) *applied to the regular proper model $\underline{X}_v$. Extending these definitions to define $(D.D')_{v,\underline{\omega}}$ via $\mathbf{Z}$-bilinearity for divisors $D, D'$ on $X_v$ with $|D| \cap |D'| \subseteq X_v(F_v)$, we get*

$$\langle D, D' \rangle_{v,\underline{\omega}} = -(D.D')_{v,\underline{\omega}} \log q_v \tag{8.3}$$

*whenever $D$ and $D'$ have degree* 0 (*defining the left side as in* (8.1) *via the same choices $\underline{\omega}$*) *and the "closure" of either $D$ or $D'$ in $\underline{X}_v$ has intersection* 0 *with every irreducible component of the closed fiber of $\underline{X}_v$.*

*Remark* 8.5. Our $\alpha_x$ is the reciprocal of the $\alpha$ in [GZ], so the lack of a minus sign in [GZ, (8.1), p. 263] is consistent with our formula with a minus sign. Beware that $\mathrm{ord}_v(\alpha_x)$ is very dependent on the particular choice of regular model $\underline{X}_v$, so $(x.x)_{v,\underline{\omega}}$ depends on $\underline{X}_v$ (but we omit such dependence from the notation so as to avoid tediousness). Such dependence is reasonable to expect in view of the fact that the ability to *compute* $\langle D, D' \rangle_{v,\underline{\omega}}$ in terms of intersection theory on $\underline{X}_v$ is also heavily dependent on a hypothesis involving $\underline{X}_v$ (namely, that the closure of either $D$ or $D'$ in $\underline{X}_v$ has intersection 0 with all irreducible components of the closed fiber of $\underline{X}_v$).

*Remark* 8.6. It is immediate from the definitions that if $c \in F_v^{\times}$ then

$$(x.x)_{v,c\underline{\omega}} = (x.x)_{v,\underline{\omega}} - \mathrm{ord}_v(c). \tag{8.4}$$

*Proof.* By symmetry, we may assume the closure of $D$ in $\underline{X}_v$ has intersection 0 with all closed fiber irreducible components. We wish to use the limit formula from Theorem 8.3. For $\underline{y}$ "near" $\underline{x}$ as in that theorem, we first claim that $D_{\underline{y}}$ satisfies the same property which we just imposed on $D$. This will enable us to compute the term $\langle D_{\underline{y}}, D' \rangle_v$ via intersection theory on $\underline{X}_v$, since $D_{\underline{y}}$ and $D'$ are certainly disjoint divisors on $X_v$.

When taking the closure of $D_{\underline{y}}$ in $\underline{X}_v$ we get essentially the same divisor as the closure of $D$ in $\underline{X}_v$ except for possibly the contributions of the sections $y_j \in X_v(F_v) = \underline{X}_v(\mathscr{O}_{F_v})$. It suffices to check that if $x \in X_v(F_v)$ is a given point and $y \in X_v(F_v)$ is sufficiently close to $x$ in the topological space $X_v(F_v)$, then for any irreducible component $\Delta$ of the closed fiber of $\underline{X}_v$ we have $(\overline{x}.\Delta) = (\overline{y}.\Delta)$, where $\overline{x}$ and $\overline{y}$ are the unique sections of $\underline{X}_v \to \operatorname{Spec}(\mathscr{O}_{F_v})$ extending $x$ and $y$ respectively (which, when viewed as closed subschemes of $\underline{X}_v$, are the scheme-theoretic closures of the respective closed points $x, y$ in the open subscheme $X_v \subseteq \underline{X}_v$). Since $\underline{X}_v$ is regular with smooth generic fiber, the section $\overline{x}$ must factor through the relative smooth locus. Moreover, if $x_0$ is the closed point of $\overline{x}$ then the preimage $B(x_0)$ of $x_0$ under the reduction map

$$X_v(F_v) = \underline{X}_v(\mathscr{O}_{F_v}) \to \underline{X}_v(\mathscr{O}_{F_v}/\mathfrak{m}_v)$$

is an open neighborhood of $x$. Thus, we consider only $y$ in this neighborhood, so $\overline{y}$ has closed point $x_0$.

If $T$ is a generator of the ideal sheaf of $\overline{x}$ in a neighborhood of $x_0$, then $\widehat{\mathscr{O}}_{\underline{X}_v, x_0} \simeq \mathscr{O}_{F_v}[\![T]\!]$ and $y \mapsto T(y)$ sets up a topological isomorphism between $B(x_0)$ and the open unit disc in $F_v$ around the origin. Thus, if the ideal sheaf of $\Delta$ is generated by the regular element $f \in \mathscr{O}_{F_v}[\![T]\!]$, then $f$ has non-zero constant term (as otherwise the Cartier divisor $\Delta$ would contain $\overline{x}$ and so would not be a closed fiber component) and $(\overline{y}.\Delta)_v = \operatorname{ord}(f(T(y)))$. We have to show that $\operatorname{ord}(f(t)) = \operatorname{ord}(f(0))$ for $t \in F_v$ sufficiently near 0. Since $f(0) \neq 0$ and $f$ has integral coefficients, it suffices to take $|t|_v < |f(0)|_v$. This completes the justification that we may compute $\langle D_{\underline{y}}, D' \rangle_v$ using intersection theory on $\underline{X}_v$ when $\underline{y}$ is sufficiently close to $\underline{x}$ and the closure of $D$ in $\underline{X}_v$ has intersection 0 with all closed fiber irreducible components of $\underline{X}_v$.

We conclude from the limit formula in Theorem 8.3 that it suffices to prove

$$-(x.x)_{v,\omega_x} \log q_v = \lim_{y \to x} (-(\overline{y}.\overline{x})_v \log q_v - \log |t_x(y)|_v)$$

where $y \in X_v(F_v) - \{x\}$. In other words, we must prove that this limit not only exists, but is equal to

$$-\operatorname{ord}_v(\alpha_x) \log q_v = \log |\alpha_x|_v$$

(and so in particular only depends on $\omega_x$ and not the specific representative uniformizer $t_x$, which is a priori clear). Since $t_x$ is a regular function in a Zariski neighborhood of $x$, it can be viewed as an "analytic" function $t_x^{\mathrm{an}}$ on a small disc centered at $x$ with parameter $T$ as in the preceding paragraph (the domain of this analytic function might not be the entire open unit disc: $t_x$ might have a pole near $x$). If $\tau = T(y)$ then $-(\overline{y}.\overline{x})_v \log q_v = -\operatorname{ord}_v(\tau) \log q_v = \log |\tau|_v$ and $t_x(y) = t_x^{\mathrm{an}}(\tau)$. We therefore have to compute

$$\lim_{\tau \to 0} (\log |\tau|_v - \log |t_x^{\mathrm{an}}(\tau)|_v)$$

where $t_x = (1/\alpha_x)T + \dots$ in $F[\![T]\!]$ with $T$ as defined above serving as a parameter on an open unit disc (so this also gives the expansion of the analytic function $t_x^{\mathrm{an}}$).

Since $t_x^{\mathrm{an}} = (1/\alpha_x)T \cdot h$ where $h$ is an analytic function near 0 with $h(0) = 1$, we conclude that $t_x^{\mathrm{an}}(\tau) = (1/\alpha_x)\tau h(\tau)$, so $\log |\tau|_v - \log |t_x^{\mathrm{an}}(\tau)| = \log |\alpha_x|_v - \log |h(\tau)|_v$, and as $\tau \to 0$ we have $\log |h(\tau)|_v \to \log |1|_v = 0$. Thus, we conclude that $\lim_{y \to x}(-(\overline{y}.\overline{x})_v \log q_v - \log |t_x(y)|_v) = \log |\alpha_x|_v$, as desired.

∎

As the above proof shows, in order to *compute* $(D.D')_{v,\underline{\omega}}$ we may make a base change to the completion of the maximal unramified extension of $F_v$, and we note that $\underline{X}_v \times_{\mathscr{O}_{F_v}} \widehat{\mathscr{O}_{F_v}^{\mathrm{sh}}}$ is still regular.

With these preliminary constructions settled, we turn to the question of computing self-intersection numbers with the help of an auxiliary tensor. Fix an integer $k$ (we allow $k \leq 0$) and fix a non-zero rational section $\theta$ of $(\Omega^1_{X_v/F_v})^{\otimes k}$ with $r_x \stackrel{\mathrm{def}}{=} \operatorname{ord}_x(\theta) \neq -k$. When $k = 0$ we are requiring the non-zero rational function $\theta$ to have a zero or pole at $x$, so the situation is always "abstract". We may write

$$\theta = (C_x t_x^{r_x} + \dots)(\mathrm{d}t_x)^{\otimes k}, \tag{8.5}$$

where $C_x \in F_v^\times$ and $t_x$ is a uniformizer lifting $\omega_x$ (so $C_x$ only depends on $\omega_x$, not $t_x$). If we use $\omega'_x = c\omega_x$ with $c \in F_v^\times$, then upon using the representative uniformizer $t'_x = ct_x$ we obtain the transformation formula

$$C'_x = \frac{C_x}{c^{r_x+k}}. \tag{8.6}$$

Note that when $\theta$ is *given* then $\omega_x$ determines $C_x$ up to $\mathscr{O}_{F_v}^\times$-scaling and vice-versa, but when $k + r_x \neq \pm 1$ then because of (8.6) we certainly cannot expect to find $\omega_x$ realizing an arbitrary desired $C_x \in F_v^\times$ for a given $\theta$. In general, the best we can do is determine $\omega_x$ up to $\mathscr{O}_{F_v}^\times$-multiple in terms of $\theta_x$ by the requirement that $0 \leq \operatorname{ord}_v(C_x) < |r_x + k|$. When $k = 1$ and $r_x = 0$ this recovers the normalization used for intersection theory via tangent vectors, but in general we cannot expect to make $C_x$ a unit at $v$. Although this milder normalization does give us a way to use $\theta$ to single out a "preferred" $\omega_x$ up to $\mathscr{O}_{F_v}^\times$-multiple (since $\mathrm{ord}_x(\theta) \neq -k$), and hence the choice of $\theta$ is "all" the input we need to make sense of $C_x$, it will actually be convenient to not minimize the choices in this way. Rather, we prefer to think of fixing $\theta$ in advance and still retaining the freedom to pick whatever $\omega_x$ we like at the point $x$.

Observe that if $\alpha_x\omega_x$ induces a basis of $\mathrm{Cot}_{\underline{x}}(\underline{X}_v)$, then there is a generator $T_x$ of the height 1 prime ideal $\mathfrak{p}_{\underline{x}}$ of the section $\underline{x}$ in $\mathscr{O}_{\underline{X}_v,\underline{x}_0}$ such that $\alpha_x t_x \equiv T_x$ in $\mathrm{Cot}_x(X_v)$. Thus,

$$\theta = \left(\frac{C_x}{\alpha_x^{r_x+k}}T_x^{r_x} + \dots\right)(\mathrm{d}T_x)^{\otimes k}, \tag{8.7}$$

so $(\alpha_x^{r_x+k}/C_x)\theta$ is a basis of the invertible $\mathscr{O}_{F_v}$-module $(\Omega^1_{\underline{X}_v/\mathscr{O}_{F_v}})^{\otimes k}(-r_x\cdot\underline{x})_{\underline{x}_0}/\mathfrak{p}_{\underline{x}}$. Consequently, we can now recast the entire discussion in terms of a fixed choice of $\theta$.

More specifically, let us now *fix* a choice of non-zero rational section $\theta$ of $(\Omega^1_{X_v/F_v})^{\otimes k}$ (in practice this will arise from a global rational tensor). For $v \nmid \infty$ and any $x \in X_v(F_v)$ with $r_x \stackrel{\mathrm{def}}{=} \mathrm{ord}_x(\theta) \neq -k$, we choose a uniformizer $t_x$ at $x$ (in practice arising from a global uniformizer at a global rational point), and let $C_x \in F_v^\times$ be the leading coefficient of the $t_x$-adic expansion of $\theta$ (this coefficient transforms by the reciprocal multiplier when the non-zero cotangent vector $\omega_x \in \mathrm{Cot}_x(X_v)$ attached to $t_x$ is replaced with an $F_v^\times$-multiple). If $\beta_x \in F_v^\times$ is chosen so that $\beta_x\theta$ is a basis of $(\Omega^1_{\underline{X}_v/\mathscr{O}_{F_v}})^{\otimes k}(-r_x\cdot\underline{x})_{\underline{x}_0}/\mathfrak{m}_{\underline{x}}$, then $\beta_x$ is unique up to $\mathscr{O}_{F_v}^\times$-multiple, $\beta_x$ *does not depend on* $\omega_x$, and via (8.7) we see

$$\frac{\mathrm{ord}_v(C_x\beta_x)}{r_x+k} = \mathrm{ord}_v(\alpha_x) = (x.x)_{v,\omega_x} \in \mathbf{Z}. \tag{8.8}$$

In this way, we can compute $(x.x)_{v,\omega_x}$ at any $x$ for which $r_x \stackrel{\mathrm{def}}{=} \mathrm{ord}_x(\theta) \neq -k$.

## 9. Self-intersection formula and application to global height pairings

We wish to apply the theory in §8 to generalize Corollary 7.15 to cases with $r_{\mathscr{A}}(m) > 0$. This amounts to finding systematic (non-canonical!) local definitions of self-intersection numbers for rational points in such a way that one still recovers the canonical global height pairing as a sum of local terms. In [GZ] this is carried out with the help of a tangent vector. Unfortunately, the application of this method in the proof of [GZ, Ch III, Lemma 8.2] uses the 1-form $\eta^4(q)\mathrm{d}q/q$ which doesn't actually live on $X_0(N)$, but only on a degree 6 covering $X' \to X_0(N)$. At points of ramification for this covering (i.e., elliptic points) one gets the zero map on tangent spaces, and elsewhere at points away from the branch locus there is no reason why global $H$-rational points on $X_0(N)$ have to lift to $H$-rational points on $X'$. Without the ability to work with $H$-rational points, and hence with $H$-rational tangent vectors, one encounters complications when formulating a global formula over $H$ in terms of local tangent vector calculations: there has got to be *some* link between all of the local tangent vectors (e.g., they all come from a global one) in order for the sum of non-canonical local terms to recover canonical global height pairings in the case of degree 0 divisors with non-disjoint supports. Our alternative approach will also have to confront the issue of what to do at elliptic points, but the merit of using deformation theory is that we will be able to treat all points in a uniform manner without needing to use specialized arguments for the elliptic case.

Our method might be called "intersection theory with meromorphic tensors". The motivation is that although $\eta^4(q)\mathrm{d}q/q$ only lives on a cyclic covering of $X_0(N)$, $\Delta(q)(\mathrm{d}q/q)^{\otimes 6}$ makes sense as a meromorphic tensor on $X_0(N)_{/\mathbf{Q}}$. More importantly, $\Delta$ has a functorial interpretation and so makes sense within the context of deformation theory. For our purposes, the special role of $\Delta$ is that (via the relative Kodaira-Spencer isomorphism) on universal deformation rings for elliptic curves (*without* level structure), the leading coefficient of $\Delta$ as a 6-tensor is always a *unit*. After we have carried out our approach, we will revisit the method used in [GZ] and see how it can be understood as a special case of our approach, at least if one avoids the quadratic fields $K = \mathbf{Q}(\sqrt{-1})$ and $K = \mathbf{Q}(\sqrt{-3})$. An added bonus of our approach via $\Delta$ and meromorphic tensors is that we will be able to argue with abstract local deformation theory instead of relying on the global geometry of modular curves. This allows us to avoid complications traditionally caused by small primes and special $j$-invariants.

The main result of this section is really Theorem 9.6, which gives a rather general formula (9.16) for global height pairings. The proof of this global formula will require a formula for certain self-intersection numbers. In the global formula, the local junk terms in Theorem 9.2 essentially all cancel out, and we are left with something that is computable. In §10 we will take up the problem of computing the non-archimedean terms (denoted $\langle x, T_m(x^\sigma)\rangle_v^{\mathrm{GZ}}$ in (9.16)) in terms of quaternionic data.

We will have to do some hard work in the special case of coarse moduli schemes, whereas the general intersection theory discussion in §8 took place on rather general smooth curves over local fields. Let us return to our earlier standard notation, with $\underline{X} = X_0(N)_{/W}$, where $W$ is the completion of a maximal unramified extension of the local ring $\mathscr{O}_{H,v}$ at a place $v$ of $H$ over a prime $p$ of $\mathbf{Q}$. Also, $F$ denotes the fraction field of $W$ and $X$ denotes the generic fiber of $\underline{X}$. We do *not* make any assumptions on the behavior of $p$ in $K$: it may be inert, split, or ramified.

Let $\underline{x} \in \underline{X}(W)$ be a section disjoint from the cuspidal locus, and assume there exists a $\Gamma_0(N)$-diagram over $W$ which represents $\underline{x}$. We emphasize that $\underline{x}$ need *not* come from a Heegner point, though by Theorem 2.5 our hypotheses are satisfied for sections arising from Heegner points over $H$, and by Theorem 2.6 such a $W$-diagram realizing a *specified* $F$-diagram is unique up to canonical $W$-isomorphism (such uniqueness having nothing to do with Heegner points). When the $F$-diagram is not specified (i.e., only the $F$-point of the coarse moduli scheme is specified) then we lose the canonicalness of the $W$-diagram. We write $\underline{x}$ to denote a Heegner diagram over $W$ (unique up to non-canonical isomorphism) representing a given section $\underline{x} \in \underline{X}(W)$ for which some such diagram exists. For our present purposes, the connection with Heegner points is not relevant; the isomorphism class of the diagram over $W$ is all that matters. We first focus on a purely local assertion concerning diagrams over $W$ (viewed as certain sections of $\underline{X}_{/W}$).

Let $x \in \underline{X}(F)$ denote the generic fiber rational point corresponding to $\underline{x}$. Fix an arbitrary integer $k$ and a rational section $\theta$ of $(\Omega^1_{X_{/F}})^{\otimes k}$ such that $r_x \stackrel{\mathrm{def}}{=} \mathrm{ord}_x(\theta) \neq -k$. Taking $k = 0$ and $\theta$ a non-zero rational function with divisor having a zero or pole at $x$ is one option, for example. Another option of interest is the case $k = 6$ and $\theta = \Delta$, but it doesn't matter what choice we make. In practice the choice we make will have to arise from the global model over $H$. For conceptual clarity, we avoid specifying a particular choice of $k$ or $\theta$ at the outset.

*Choose* a cotangent vector $\omega_x$ at $x$ on $\underline{X}_{/F}$, with $t_x$ a uniformizer representing $\omega_x$. We let $C_x$ denote the leading coefficient of the $t_x$-adic expansion of the rational section $\theta$ (relative to the basis $(\mathrm{d}t_x)^{\otimes k}$), so $C_x$ depends only on $\omega_x$. This data allows us to define $(x.x)_{v,\omega_x}$ in accordance with the recipe of the previous section (see (8.8))

The group $\mathrm{Aut}_F(x)$ concides with the "geometric" automorphism group of $x$ (by Theorem 2.6), and hence has order $2u_x$ with $u_x \in \{1, 2, 3\}$ (as $F$ has characteristic 0). For example, if $\underline{x}$ came from a global point over $H$, then by Theorem 2.1(1) the value of $u_x$ would actually be of global nature (i.e., it would depend only on the resulting $\overline{\mathbf{Q}}$-point or $\mathbf{C}$-point) and not at all depend on $v$. We now introduce an auxiliary quantity which intuitively measures the fact that we are trying to do intersection theory on a (regular model of a) coarse moduli scheme rather than on a Deligne-Mumford stack. To get started, we require a preliminary lemma in deformation theory.

We state the lemma in the specific context we need (i.e., $\Gamma_0(N)$-structures), but the reader will see that the method of proof works much more generally. The basic point is to provide a mechanism for passing between deformation rings of objects in characteristic 0 and objects in characteristic $p$ (in cases for which the universal deformations are algebraizable, so it actually makes sense to base change a formal deformation from residue characteristic $p$ over to residue characteristic 0; e.g., there is a map $\mathrm{Spec}(\mathbf{Q}_p) \to \mathrm{Spec}(\mathbf{Z}_p)$ but there is no map $\mathrm{Spf}(\mathbf{Q}_p) \to \mathrm{Spf}(\mathbf{Z}_p)$).

**Lemma 9.1.** *Let $\underline{x}$ be a $\Gamma_0(N)$-structure over $W$, with generic fiber $x$ over $F$. Let $\mathscr{R}_0$ be the universal deformation ring of $\underline{x}_0$ on the category of complete local noetherian $W$-algebras with residue field $W/\pi$. Let $\mathscr{I}_{\underline{x}} \subseteq \mathscr{R}_0$ be the ideal of the section corresponding to the $W$-deformation $\underline{x}$. The $F \otimes_W \mathscr{I}_{\underline{x}}$-adic completion* *of the algebraized universal deformation over $F \otimes_W \mathscr{R}_0$ is the universal deformation of $x$ on the category of complete local noetherian $F$-algebras with residue field $F$.*

Let us explicate the meaning of the lemma in the case $N = 1$ (which, strictly speaking, is not a case we have ever been considering. but for which the lemma is true). Let $E$ be an elliptic curve over $W$ with closed fiber $E_0$ over $W/\pi$, and let $\mathscr{R}$ be the universal deformation ring of $E_0$. Let $\mathscr{E} \to \mathrm{Spec}(\mathscr{R})$ be an elliptic curve lifting $E_0$ which algebraizes the universal formal deformation. By abstract deformation theory $\mathscr{R}$ is formally smooth over $W$ of relative dimension 1, and the ideal cutting out the $W$-section of $\mathrm{Spec}(\mathscr{R})$ arising from the special deformation $E_{/W}$ is a height 1 prime. By choosing a generator $T$ of this we can identify $\mathscr{R} \simeq W[\![T]\!]$ such that reduction modulo $T$ recovers $E_{/W}$. The content of the lemma is that the elliptic curve $\mathscr{E}_{/F[\![T]\!]}$ is an (algebraized) universal deformation of the *generic* fiber $E_{/F}$ of our original elliptic curve $E$ over $W$. In this way, we have a precise link between universal deformation rings of the generic and closed fibers of $E_{/W}$.

To see the general meaning of the lemma (with $N > 1$ allowed), suppose $\underline{x}$ actually occurs in a universal algebraic family over an affine finite type $W$-scheme $\mathrm{Spec}(R)$. Let $\mathfrak{m}$ denote the maximal ideal associated to $\underline{x}_0$ and let $\mathfrak{p}_{\underline{x}}$ denote the prime associated to $\underline{x}$. Note that the $\mathfrak{p}_{\underline{x}}$-adic completion of $R$ is naturally isomorphic to $\widehat{R}_{\mathfrak{m}}$ with a slightly weaker topology, since $R/\mathfrak{p}_{\underline{x}} \simeq W$ is max-adically complete. For example, if $\widehat{R}_{\mathfrak{m}} \simeq W[\![T]\!]$ with its maximal-adic topology (and $T$ corresponding to the section $\underline{x}$), then the $\mathfrak{p}_{\underline{x}}$-adic completion would be isomorphic to $W[\![T]\!]$ with just the $T$-adic (rather than $(\pi, T)$-adic) topology.

If we let $R_\eta = F \otimes_W R$, then the above lemma is just the obvious commutative algebra assertion that the completion of $R_\eta$ along the section $x$ is naturally isomorphic to the completed tensor product $F \widehat{\otimes}_W \widehat{R}_{\mathfrak{m}}$ where $F$ and $W$ are given the discrete topology and $\widehat{R}_{\mathfrak{m}}$ is given its topology as $\mathfrak{p}_{\underline{x}}$-adic completion of $R$. In even more concrete terms, this is a jazzed-up version of the assertion that if we form the $T$-adic completion of $F \otimes_W (W[\![T]\!])$ then we naturally get $F[\![T]\!]$ (beware that we're not actually assuming formal smoothness over $W$ for the deformation ring $\mathscr{R}_0$ in the lemma, though such smoothness does hold in the cases to which we'll be applying this lemma later on).

*Proof.* Although we could give a proof using universal algebraic families, for aesthetic reasons we prefer to give a proof entirely within the framework of deformation theory, as this clarifies the general nature of the argument (i.e., the role of elliptic curves and level structures is rather inessential to the argument). Essentially, the same method will apply "whenever" one is studying a moduli problem which is finite over one whose deformation rings are formally smooth, whose Isom-schemes are finite unramified, and whose universal deformations are algebraizable over the base. First, we remove the appearance of $\Gamma_0(N)$-structures (which we view as pairs consisting of an elliptic curve and an auxiliary subgroup scheme with $\Gamma_0(N)$-structure). Let the complete local noetherian $W$-algebra $R_0$ denote the universal deformation ring for the elliptic curve underlying $\underline{x}_0$, and let $I_{\underline{x}}$ denote the ideal corresponding to the $W$-section given by the elliptic curve underlying $\underline{x}$.

The functor of $\Gamma_0(N)$-structures on an elliptic curve is finite, so if $R_0'$ denotes the finite $R_0$-algebra classifying such structures on deformations of the elliptic curve underlying $\underline{x}_0$, then $R_0'$ is the product of finitely many local rings (as $R_0$ is local henselian), and the ring $\mathscr{R}_0$ is the unique local factor ring of $R_0'$ corresponding to the $\Gamma_0(N)$-structure $\underline{x}_0$, or equivalently is the unique *local* factor ring of $R_0'$ supporting the $W$-section arising from $\underline{x}$. It follows that the $F \otimes_W \mathscr{I}_{\underline{x}}$-adic completion of $F \otimes_W \mathscr{R}_0$ is a factor ring of the ring of $\Gamma_0(N)$-structures on the "universal" elliptic curve over the $F \otimes_W I_{\underline{x}}$-adic completion of $F \otimes_W R$. Moreover, this completion of $F \otimes_W \mathscr{R}_0$ also supports the $F$-section corresponding to $x$. Since $F \otimes_W (R/I_{\underline{x}}) = F$ and $F \otimes_W (\mathscr{R}_0/\mathscr{I}_{\underline{x}}) = F$ are *local*, so the indicated completions of $F \otimes_W \mathscr{R}_0$ and $F \otimes_W R$ are also local, it follows that the $F \otimes_W \mathscr{I}_{\underline{x}}$-adic completion of $F \otimes_W \mathscr{R}_0$ (along with the $\Gamma_0(N)$-structure over it!) is the desired universal deformation of $x$ *provided* that the $F \otimes_W I_{\underline{x}}$-adic completion of $F \otimes_W R$ (along with the elliptic curve over it!) is the universal deformation of the elliptic curve underlying $x$. This latter statement is exactly the original problem for elliptic curves without the interference of level structures.

Throwing away the level structures, we begin with an elliptic curve $E_0$ over $W/\pi$ and a deformation $E$ of $E_0$ to $W$. Let $R$ denote the universal deformation ring of $E_0$ and let $J \subseteq R$ be the ideal of the $W$-section corresponding to $E$. Define $R_\eta = F \otimes_W R$ and $J_\eta = F \otimes_W J$, so $J_\eta$ is a maximal ideal in $R_\eta$ (with residue field $F$). We need to prove that the $J_\eta$-adic completion of $R_\eta$ (along with the elliptic curve over it) is the universal deformation of $E_{/F}$ (on the category of complete local noetherian $F$-algebras with residue field $F$). Let $\widetilde{R}_\eta$ be the universal deformation ring of $E_{/F}$, so we have a natural map

$$\widetilde{R}_\eta \to \widehat{(R_\eta)_{J_\eta}} \tag{9.1}$$

which we want to prove to be an isomorphism.

From the deformation theory of elliptic curves, we know that both source and target in (9.1) are formal power series rings over $F$ of the same dimension (which happens to be 1, but we ignore this fact to maintain conceptual understanding of the situation). Thus, it suffices to prove that the map is surjective. More specifically, it suffices to prove that the map

$$\widetilde{R}_\eta \to R_\eta/J_\eta^2 = F \otimes_W (R/J^2) \tag{9.2}$$

is surjective. The whole point is to prove surjectivity onto the maximal ideal $F \otimes_W (J/J^2)$, with $J/J^2$ a *finite free* $W$-module. Assuming failure of surjectivity, we can choose an appropriate codimension 1 lattice in $J/J^2$ which eats up the entire image of the maximal ideal of $\widetilde{R}_\eta$ under (9.2) (after tensoring with $F$). Thus, we can find a $W$-algebra quotient of $R = W \oplus J$ of the form $W[\varepsilon]$ (with $(\varepsilon)$ the image of $J$) so that the resulting elliptic curve deformation over $F[\varepsilon]$ is trivial. In other words, we will have a non-trivial deformation $\mathscr{E}$ of $E_{/W}$ to $W[\varepsilon]$ which induces a trivial deformation of $E_{/F}$ over $F[\varepsilon]$. We now show that such a deformation cannot exist.

Consider the Isom-scheme $I = \underline{\mathrm{Isom}}(\mathscr{E}, E_{/W[\varepsilon]})$ over $W[\varepsilon]$ which classifies elliptic curve isomorphisms over variable $W[\varepsilon]$-schemes (*not* necessarily respecting structures as deformations of $E_{/W}$!). This is a finite unramified $W[\varepsilon]$-scheme (by the deformation theory of elliptic curves), and we just have to prove that $I(W[\varepsilon]) \to I(F[\varepsilon])$ is surjective, since an isomorphism $\mathscr{E} \simeq E_{/W[\varepsilon]}$ over $W[\varepsilon]$ which induces an isomorphism of $F[\varepsilon]$-*deformations* of $E_{/F}$ is automatically an isomorphism of *deformations* of $E_{/W}$ (as the "generic fiber" functor from flat separated $W$-schemes to $F$-schemes is faithful). Since $W[\varepsilon]$ is a henselian local ring, any finite unramified $W[\varepsilon]$-scheme can be realized as a closed subscheme of a finite étale $W[\varepsilon]$-scheme. But since $W[\varepsilon]$ is even strictly henselian, it follows that the only finite étale $W[\varepsilon]$-algebras are finite products of copies of $W[\varepsilon]$. Thus, $I$ is a finite disjoint union of spectra of quotients of $W[\varepsilon]$. But the ideal theory of $W[\varepsilon]$ is sufficiently easy that we see by inspection that the only quotient of $W[\varepsilon]$ which admits a $W[\varepsilon]$-algebra map to $F[\varepsilon]$ is $W[\varepsilon]$ itself. Thus, $I(W[\varepsilon]) \to I(F[\varepsilon])$ is surjective (and even bijective). ■

Now we return to our original situation with the universal deformation ring $\mathscr{R}_0$ of $\underline{x}_0$. We have a natural map $\widehat{\mathscr{O}}_{\underline{X},\underline{x}} = \widehat{\mathscr{O}}_{\underline{X},\underline{x}_0} \to \mathscr{R}_0$ which computes the subring of invariants in $\mathscr{R}_0$ under $\mathrm{Aut}_{W/\pi}(\underline{x}_0)$. After inverting $\pi$ and passing to completions along the sections defining $x$, the source ring becomes $\widehat{\mathscr{O}}_{X,x}$ while the target ring becomes the universal deformation ring $\mathscr{R}_\eta$ of $x$, thanks to Lemma 9.1. Moreover, by chasing universal (algebraized) objects in Lemma 9.1 (and projecting to artinian quotients of $\mathscr{R}_\eta$), we see that this induced natural map

$$\widehat{\mathscr{O}}_{X,x} \to \mathscr{R}_\eta \tag{9.3}$$

is indeed the canonical isomorphism of $\widehat{\mathscr{O}}_{X,x}$ onto the subring of $\mathrm{Aut}_F(x)$-invariants in the universal deformation ring $\mathscr{R}_\eta$ of $x$. Since the group $\mathrm{Aut}_F(x)/\{\pm 1\}$ of order $u_x$ acts faithfully on the universal deformation ring $\mathscr{R}_\eta \simeq F[\![T]\!]$, so the totally (tamely) ramified extension (9.3) has ramification degree $u_x$, we conclude that the generator $\theta_x$ of $(\widehat{\Omega}^1_{\widehat{\mathscr{O}}_{X,x}/F})^{\otimes k}(-r_x)$ maps to a generator of $(\widehat{\Omega}^1_{\mathscr{R}_\eta/F})^{\otimes k}(-(r_x u_x + k(u_x - 1)))$, where the twisting notation "$(n)$" denotes tensoring with the $n$th power of the inverse of the maximal ideal; this amounts to nothing more than the calculation that if $T' = \mathrm{unit} \cdot T^u$ then

$$T'^r (\mathrm{d}T')^{\otimes k} = \mathrm{unit}' \cdot T^{ru + k(u-1)} (\mathrm{d}T)^{\otimes k} \tag{9.4}$$

in $\widehat{\Omega}_{F[\![T]\!]/F}$ (since $u \in F^\times$).

If we let $\mathscr{J}$ denote the (invertible!) height 1 prime ideal of the $W$-section $\underline{x}$ of the formally smooth deformation ring $\mathscr{R}_0$ and let $\mathscr{J}_\eta$ denote the ideal of the $F$-section $x$ of $\mathscr{R}_\eta$, then for any integer $m$ we have a natural isomorphism of 1-dimensional $F$-vector spaces

$$(\widehat{\Omega}^1_{\mathscr{R}_\eta/F})^{\otimes k}(m)/\mathscr{J}_\eta \simeq F \otimes_W \left( (\widehat{\Omega}^1_{\mathscr{R}_0/W})^{\otimes k}(m)/\mathscr{J} \right),$$

where on the right side we are twisting relative to the invertible ideal $\mathscr{J}$. The left side is a 1-dimensional $F$-vector space and the "integral" differentials on the right side provide a natural rank 1 lattice in this vector space. In this way, taking $m = -(r_x u_x + k(u_x - 1))$ enables us to define an integer $\mathrm{ord}_{v,x}(\theta)$ which measures the extent to which $\theta_x \bmod \mathscr{J}_\eta$ fails to arise from a generator of the lattice of "integral" differential tensors. In more explicit terms, if we compute the formal expansion of the tensor $\theta$ relative to a formal parameter along the section $\underline{x}$ of the "integral" deformation ring $\mathscr{R}_0$ of $\underline{x}_0$, then we get a leading coefficient in $F^\times$ which is well-defined up to unit, and $\mathrm{ord}_{v,x}(\theta)$ is just the $\mathrm{ord}_v$ of this coefficient.

With these preliminary considerations settled, we are now in position to state the main local formula:

**Theorem 9.2.** *With notation as defined above, including $C_x$ defined as in* (8.5)*, we have*

$$(x.x)_{v,\omega_x} = \frac{1}{2}\sum_{n\geq 0}(|\operatorname{Aut}_{W_n}(\underline{x})| - |\operatorname{Aut}_W(\underline{x})|) + \frac{\mathrm{ord}_v(C_x u_x^k) - \mathrm{ord}_{v,x}(\theta)}{r_x + k} \tag{9.5}$$

*Remark* 9.3. Despite possible appearances to the contrary, we will see in §10 that (9.5) with $\theta = \Delta$ recovers [GZ, Ch III, Lemma 8.2] at non-elliptic points.

*Remark* 9.4. If $x$ arises from a global point over $H$ and we choose a global $\theta$ and a global $\omega_x$, then $C_x$, $u_x$, and $r_x$ all have global meaning independent of $v$ (e.g., $2u_x$ is the order of the geometric automorphism group of $x$, which can be computed over $\overline{\mathbf{Q}}$, $\mathbf{C}$, or an algebraic closure of the fraction field of $W$). We regard the first term on the right side of (9.5) as the interesting part, and the rest as "junk". The only junk term which is somewhat subtle is $\mathrm{ord}_{v,x}(\theta)$, since it is defined via formal deformation theory at $v$ and hence is not obviously $\mathrm{ord}_v$ of some globally defined quantity when $x$, $\omega_x$, and $\theta_x$ are globally defined at the start. Thus, we will need to do a little work to explicate the "global" meaning of $\mathrm{ord}_{v,x}(\theta)$ in such situations. When we take $k = 6$ and $\theta = \Delta$, then typically $r_x = 0$ and $\mathrm{ord}_{v,x}(\theta) = 0$. However, at elliptic points we may have $r_x \neq 0$ and when $v|N$ we may have $\mathrm{ord}_{v,x}(\theta) \neq 0$. The choice $\theta = \Delta$ is the one we will use later on in order to explicate formulas for global height pairings.

For our purposes the extra local "junk" terms will not be problematic because when we form the sum over *all* places $v$ of $H$ (in the context of Heegner points), then by the product formula these terms will essentially add up to zero once the archimedean analogue is adapted to our method of computation and once we have found a more conceptual interpretation of $\mathrm{ord}_{v,x}(\theta)$ for a well-chosen $\theta$ (namely, $\theta = \Delta$). It is crucial for such an application of the product formula that the denominator $r_x + k$ in (9.5) does *not* depend on $v$ when $x \in \underline{X}(F)$ arises from a global point over $H$ and $\theta$, $\omega_x$ are globally chosen over $H$.

Now we prove Theorem 9.2.

*Proof.* Although $\underline{X}$ need not be regular, as usual it suffices to work with this model for our intersection theory calculations because $\underline{x}$ lies in the $W$-smooth locus on this model (cf. (8.8)). We also observe that both sides of (9.5) transform the same way under a change of $\omega_x$. Indeed, if we use $\omega'_x = c\omega_x$ for some $c \in F_v^\times$ and if $\alpha_x\omega_x$ is a basis of $\mathrm{Cot}_{\underline{x}}(\underline{X}_v)$, then $\alpha'_x = \alpha_x/c$ makes $\alpha'_x\omega'_x = \alpha_x\omega_x$ an $\mathscr{O}_{F_v}$-basis of the cotangent line along $\underline{x}$, so

$$(x.x)_{v,\omega'_x} = \mathrm{ord}_v(\alpha'_x) = -\mathrm{ord}_v(c) + \mathrm{ord}_v(\alpha_x) = -\mathrm{ord}_v(c) + (x.x)_{v,\omega_x},$$

while by (8.6) we have $\mathrm{ord}_v(C'_x)/(r_x + k) = -\mathrm{ord}_v(c) + \mathrm{ord}_v(C_x)/(r_x + k)$. Since $u_x$ and $\mathrm{ord}_{v,x}(\theta)$ and the Aut-terms on the right side of (9.5) do not depend on $\omega_x$, we conclude that indeed both sides of (9.5) transform in the same way under change in $\omega_x$. Thus, it suffices to prove the result for one choice of $\omega_x$. We will make the choice later in the proof.

Let $G = \operatorname{Aut}_{W/\pi}(\underline{x}_0)/\{\pm 1\}$, a quotient of the group of automorphisms of $\underline{x}_0$ as a $\Gamma_0(N)$-structure. Let $\mathscr{R}_0$ denote the universal deformation ring of $\underline{x}_0$, so there is a natural *injective* $W$-algebra action map

$$[\cdot] : G \to \operatorname{Aut}_W(\mathscr{R}_0),$$

an abstract isomorphism $\mathscr{R}_0 \simeq W[\![T_0]\!]$, and an isomorphism $\widehat{\mathscr{O}}_{\underline{X},\underline{x}_0} \simeq \mathscr{R}_0^G$, where $T_0$ cuts out the deformation $\underline{x}$ over $W$. The norm $T_{\underline{x}} \stackrel{\text{def}}{=} \mathrm{Norm}_G(T_0) = \prod_{g\in G}[g](T_0)$ is a formal parameter of $\mathscr{R}_0^G \simeq \widehat{\mathscr{O}}_{\underline{X},\underline{x}_0} \simeq \widehat{\mathscr{O}}_{\underline{X},\underline{x}}$ over $W$, so $T_{\underline{x}}$ serves as a formal parameter along $\underline{x}$ on $\underline{X}$.

We now give a deformation-theoretic interpretation of

$$\frac{1}{2}\sum_{n\geq 0}(|\operatorname{Aut}_{W_n}(\underline{x})| - |\operatorname{Aut}_W(\underline{x})|). \tag{9.6}$$

The method we use applies to a rather more general class of deformation problems with formally smooth deformation rings of relative dimension 1 (as the following abstract argument makes clear). Since a representative in $\operatorname{Aut}_{W/\pi}(\underline{x})$ of $g \in G$ lifts to $W_n$ (as an automorphism of $\Gamma_0(N)$-structures) if and only if its negative lifts, it makes to speak of an element of $G$ lifting to $W_n$. Thus, by including the factor of $1/2$ we see that the $n$th term in (9.6) counts the number of $g \in G$ which lift to $W_n$ but don't lift to $W$. But $g \in G$ lifts to $W_n$ if and only if $T_0$ and $[g](T_0)$ generate the same ideal in the universal *deformation* ring $\mathscr{R}_0/\pi^{n+1}$ of $\underline{x} \bmod \pi^{n+1}$ over $W_n$-algebras (with residue field $W/\pi$). Since the quotients $\mathscr{R}_0/(\pi^{n+1}, T_0)$ and $\mathscr{R}_0/(\pi^{n+1}, [g](T_0))$ are abstractly $W_n$-isomorphic (via $[g]$) and hence have the same *finite* length, a surjection of $\mathscr{R}_0$-algebras $\mathscr{R}_0/(\pi^{n+1}, [g](T_0)) \twoheadrightarrow \mathscr{R}_0/(\pi^{n+1}, T_0)$ is necessarily an isomorphism. Thus, we conclude that $g$ lifts to $W_n$ if and only if $[g](T_0)$ is a multiple of $T_0$ in $\mathscr{R}_0/\pi^{n+1}$. By viewing $[g](T_0) \in \mathscr{R}_0 \simeq W[\![T_0]\!]$ as a formal power series in $T_0$, to say that $[g](T_0)$ is a multiple of $T_0$ modulo $\pi^{n+1}$ amounts to saying that the constant term $[g](T_0)(0)$ (i.e., the image of $[g](T_0)$ in $\mathscr{R}_0/(T_0) \simeq W$) is divisible by $\pi^{n+1}$.

We conclude that $g \in G$ doesn't lift to $W$ if and only if $[g](T_0)$ has non-zero constant term $[g](T_0)(0)$, in which case $g$ lifts to $W_n$ if and only if $n+1 \leq \mathrm{ord}_v([g](T_0)(0))$. It follows that (9.6) is exactly the sum of the $\mathrm{ord}_v$'s of the non-zero constant terms among the $[g](T_0)$'s. Equivalently, when we consider the formal parameter $T_{\underline{x}} = \mathrm{Norm}_G(T_0) = \prod_{g\in G}[g](T_0)$ along $\underline{x}$ as an element in the universal deformation ring $\mathscr{R}_0 \simeq W[\![T_0]\!]$, the $\mathrm{ord}_v$ of its least degree non-zero coefficient is exactly equal to (9.6).

Since the $G$-norm $T_{\underline{x}}$ of $T_0$ formally cuts out $\underline{x} \in \underline{X}(W)$, it also induces a uniformizer in the complete local ring at $x \in X(F)$. Thus, when $T_{\underline{x}}$ is viewed as an element in $W[\![T_0]\!] \subseteq F[\![T_0]\!]$ (with the latter ring naturally identified with the universal deformation ring of $x$, by Lemma 9.1), it has the form

$$T_{\underline{x}} = b_x T_0^{u_x} + \dots \tag{9.7}$$

with $b_x \neq 0$ and $u_x = (1/2)|\operatorname{Aut}_F(x)|$ the ramification degree of the universal deformation ring $F[\![T_0]\!]$ over $\widehat{\mathscr{O}}_{X,x}$. We have seen already that when the $G$-norm $T_{\underline{x}}$ of $T_0$ is expanded as a product of $[g](T_0)$'s, the product of the non-zero constant terms among the $[g](T_0)$'s (taken with respect to $T_0$-adic expansions) has $\mathrm{ord}_v$ equal to (9.6). This product of constant terms is $b_x$ up to $W^\times$-multiple (arising from the $g$'s lifting to $W$, for which $[g](T_0)$ is a unit multiple of $T_0$), so we conclude that (9.6) is exactly $\mathrm{ord}_v(b_x)$.

We will take $\omega_x$ to be represented by the formal parameter $t_x = T_{\underline{x}}$ in $\widehat{\mathscr{O}}_{X,x}$ (an inspection of our definition of $(x.x)_{v,\omega_x}$ and the calculation (8.8) makes it clear that we may work with formal uniformizers that don't necessarily arise from rational functions in $\mathscr{O}_{X,x}$ on the algebraic curve $X_{/F}$). In $(\widehat{\Omega}^1_{\widehat{\mathscr{O}}_{X,x}/F})^{\otimes k}(-r_x\cdot x)$, by (9.7) we have

$$\theta = (C_x T_{\underline{x}}^{r_x} + \dots)(\mathrm{d}T_{\underline{x}})^{\otimes k} = (C_x b_x^{r_x+k} u_x^k T_0^{r_x u_x + k(u_x-1)} + \dots)(\mathrm{d}T_0)^{\otimes k}.$$

Thus, by definition we have

$$\mathrm{ord}_{v,x}(\theta) = \mathrm{ord}_v(C_x b_x^{r_x+k} u_x^k) = \mathrm{ord}_v(C_x u_x^k) + (r_x+k)\mathrm{ord}_v(b_x). \tag{9.8}$$

Now let's consider the main identity (9.5). Since $\theta$ has its $T_{\underline{x}}$-adic expansion with leading coefficient $C_x$, where $T_{\underline{x}}$ is a formal parameter along the $W$-section $\underline{x}$ on $\underline{X}$, we take $\beta_x = 1/C_x$ in (8.8), so $(x.x)_{v,\omega_x} = 0$. Meanwhile, by (9.8) the right side of (9.5) is equal to

$$\mathrm{ord}_v(b_x) + \frac{\mathrm{ord}_v(C_x u_x^k) - \mathrm{ord}_{v,x}(\theta)}{r_x + k} = 0 = (x.x)_{v,\omega_x}.$$

■

Before we move on to establish formulas for $(\underline{x}.T_m(\underline{x^\sigma}))_v$ – when properly defined – in cases with $r_{\mathscr{A}}(m) > 0$, it is convenient to introduce some convenient shorthand so as to avoid having to carry out the "junk" local terms from (9.5). For a $\Gamma_0(N)$-diagram $\underline{x}$ over $W$ whose associated section in $\underline{X}(W)$ lies in $\underline{X}^{\mathrm{sm}}(W)$ (and is *not* assumed to arise from a Heegner point, though Heegner points *do* satisfy these hypotheses), we define

$$(x.x)_v^{\mathrm{GZ}} = \frac{1}{2}\sum_{n\geq 0}(|\operatorname{Aut}_{W_n}(\underline{x})| - |\operatorname{Aut}_W(\underline{x})|). \tag{9.9}$$

We emphasize that (9.9) is in general not equal to our self-intersection pairing, but we now explain why it does coincide with the one used in [GZ, Ch III, §8] away from elliptic points and characteristics dividing $N$. This requires a definition.

**Definition 9.5.** For $\underline{x}$ as above which is non-elliptic (i.e., $u_x = 1$, so $\operatorname{ord}_x(\Delta) = (6/u_x)(u_x - 1) = 0$) and factors through the relative smooth locus, let $\omega$ denote a 1-form near $x$ lifting a non-zero cotangent vector denoted $\omega_x$ at $x$. Let $g_\omega = \omega^{\otimes 6}/\Delta$, a non-zero rational function with no zero or pole at $x$. Define

$$(x.x)^{\eta^4} = (x.x)^{\omega_x} + \frac{1}{6}\operatorname{ord}_v(g_\omega(x)).$$

This is independent of the choice of $\omega$ (and $\omega_x$).

To see the independence of the choice of $\omega$ (and $\omega_x$), we note that using $\omega' = h\omega$ with a rational function $h$ requires $h$ to not have a zero or pole at $x$ and hence the equalities $\omega'_x = h(x)\omega_x$ and $g_{\omega'} = h^6 g_\omega$ force

$$(x.x)^{\omega'_x} + \frac{1}{6}\operatorname{ord}_v(h(x)^6 g_\omega(x)) = (x.x)^{\omega_x} - \operatorname{ord}_v(h(x)) + \operatorname{ord}_v(h(x)) + \frac{1}{6}\operatorname{ord}_v(g_\omega(x)).$$

The reason for the notation $(x.x)^{\eta^4}$ is that if $\eta^4$ actually made sense as a meromorphic 1-form on $X_0(N)$ and were non-vanishing and regular at $x$ then we would recover the old definition of "intersection theory with a cotangent vector" (based on $\eta^4_x$). To see this, if we pretend $\eta^4$ lives on $X$ then we can write $\Delta = (\eta^4)^6$ and hence formally

$$\frac{1}{6}\operatorname{ord}_v(g_\omega(x)) = \operatorname{ord}_v(\frac{\omega}{\eta^4}(x)),$$

so

$$(x.x)^{\eta^4} = (x.x)^{\omega_x} + \operatorname{ord}_v((\omega/\eta^4)(x)) \tag{9.10}$$

and since $\omega_x = \eta^4_x \cdot (\omega/\eta^4)(x)$ we could use (8.4) to thereby "recover" the definition of self-intersection at $x$ using the hypothetical cotangent vector $\eta^4_x$.

In any case, to actually compute the intrinsic $(x.x)^{\eta^4}$ we may work formally at $x$ and so can choose $\omega = \mathrm{d}T$ with $T$ a formal parameter along the section $\underline{x} \in \underline{X}^{\mathrm{sm}}(W)$. In terms of earlier notation from (8.5) with $\theta = \Delta$ we have $g_\omega(x) = 1/C_x$. Combining this with the equalities $u_x = 1$ and $r_x + k = 6$, we conclude from Theorem 9.2 that

$$(x.x)^{\eta^4} = (x.x)^{\omega_x} - \frac{1}{6}\operatorname{ord}_v(C_x) = (x.x)^{\mathrm{GZ}} - \frac{1}{6}\operatorname{ord}_{v,x}(\Delta).$$

But we will prove in Lemma 10.1 that under the above hypotheses, $\operatorname{ord}_{v,x}(\Delta)/6 = \operatorname{ord}_v(\mathfrak{n})$. Thus, as long as $v \nmid N$, we have $(x.x)^{\eta^4} = (x.x)^{\mathrm{GZ}}$. It is the modified intersection pairing based on Definition 9.5 which is used in the intersection theory considerations in [GZ, Ch III, §8] (so Theorem 9.2 with $v \nmid N$ really does recover [GZ, Ch III, Lemma 8.2]), but now that we have seen how to recover the point of view in [GZ] at non-archimedean places via our perspective if one avoids elliptic points, we won't make any further reference to Definition 9.5 because when $v|N$ or $v|\infty$ our local terms seem different from the ones in [GZ] (but of course the *global* sums coincide).

We define $\langle x, x\rangle_v^{\mathrm{GZ}} \stackrel{\mathrm{def}}{=} -(x.x)_v^{\mathrm{GZ}}\log(q_v)$. Since $\operatorname{ord}_x(T_m(x^\sigma)) = r_{\mathscr{A}}(m)$, by Theorem 9.2 this definition then yields

$$(x.T_m(x^\sigma))_{v,\omega_x} = (x.T_m(x^\sigma))_v^{\mathrm{GZ}} + r_{\mathscr{A}}(m)\left(\frac{\operatorname{ord}_v(C_x u_x^k) - \operatorname{ord}_{v,x}(\theta)}{r_x + k}\right),$$

where we define $(.)_v^{\mathrm{GZ}}$ to be essentially the usual intersection pairing except with (9.9) replacing the self-intersection formula, and

$$\theta = (C_x t_x^{r_x} + \dots)(\mathrm{d}t_x)^{\otimes k}$$

with $t_x$ a uniformizer representing $\omega_x$ (and $C_x$ only depends on $\omega_x$, not $t_x$). We also define

$$\langle x, T_m(x^\sigma)\rangle_v^{\mathrm{GZ}} = -(x.T_m(x^\sigma))_v^{\mathrm{GZ}} \log(q_v) \tag{9.11}$$

using (9.9) and the canonical local height pairing.

Define $c = x - \infty$, $d = x - 0$. Using Theorem 8.4 and Theorem 9.2, for non-archimedean places $v$ of $H$

$$\langle c, T_m(d^\sigma)\rangle_{v,\omega_x} = \langle x, T_m(x^\sigma)\rangle_v^{\mathrm{GZ}} + r_{\mathscr{A}}(m)\frac{\log|C_x u_x^k|_v - \log|\theta|_{x,v}}{r_x + k}, \tag{9.12}$$

where $|\theta|_{x,v} \stackrel{\mathrm{def}}{=} q_v^{-\mathrm{ord}_{v,x}(\theta)}$ denotes the $v$-adic absolute value of the leading coefficient of $\theta$ when expanded relative to a formal parameter along $\underline{x}$ in the universal deformation ring at $\underline{x}_0$. Here we have used that the closure of $T_m(0)$ is $\sigma_1(m).\underline{0}$ and that $\underline{x}$ and $T_m(\underline{x^\sigma})$ are disjoint from the cuspidal locus since CM points are everywhere potentially good reduction.

We define

$$\mathfrak{a}_x = \prod_{v\nmid\infty} \mathfrak{p}_v^{\mathrm{ord}_{v,x}(\theta)}. \tag{9.13}$$

By Lemma 10.1, when $\theta = \Delta$ and $u_x = 1$ then $\mathfrak{a}_x = \overline{\mathfrak{n}}^6$. We conclude that for every non-archimedean place $v$ of $H$,

$$\langle c, T_m(d^\sigma)\rangle_{v,\omega_x} = \langle x, T_m(x^\sigma)\rangle_v^{\mathrm{GZ}} + r_{\mathscr{A}}(m)\frac{\log|C_x u_x^k \mathfrak{a}_x^{-1}|_v}{r_x + k}, \tag{9.14}$$

where the factor $C_x u_x^k$ on the right side is an element in $H^\times$ and $\mathfrak{a}_x$ is a fractional ideal. This puts us in excellent position to use the product formula to check that the junk terms will (almost) globally sum to 0 (and hence we can ignore them). What we need to do is work out the appropriate archimedean analogue of the calculation (9.14).

Fix a complex place $v$ of $H$ and let $\mathbf{C}_v$ denote $H_v$ (with $\mathbf{C}_v$ non-canonically isomorphic to $\mathbf{C}$). Fix a choice of $\sqrt{-1} \in \mathbf{C}_v$ to define an orientation on $\mathbf{C}_v$-manifolds, so we may realize the associated upper half-space $\mathfrak{h}_v$ as the base of a universal $\mathbf{C}_v$-analytic family of elliptic curves with trivialized (oriented) relative homology. We thereby get a canonical $\mathbf{C}_v$-analytic "uniformization" $\pi_v : \mathfrak{h}_v \to X_0(N)^{\mathrm{an}}_{/\mathbf{C}_v} = X_0(N)(\mathbf{C}_v)$ and we may (and do) choose $z_v \in \mathfrak{h}_v$ lifting $x_v \in X_0(N)(\mathbf{C}_v)$. The pullback tensor $\pi_v^*(\theta_v^{\mathrm{an}})$ has order $r_x u_x + k(u_x - 1)$ at $z_v$ because of a calculation such as in (9.4) and the fact that $\mathfrak{h}_v$ is the base of a universal family for an analytic moduli problem which is étale over the $\Gamma_0(N)$-moduli problem in the $\mathbf{C}_v$-analytic category (so $\pi_v$ computes the same ramification degrees $u_x$). Let $g_{z_v}$ be a local analytic uniformizer at $z_v$ on $\mathfrak{h}_v$ which enjoys the property that the $g_{z_v}$-adic analytic expansion

$$\pi_v^*(\theta_v) = (g_{z_v}^{r_x u_x + k(u_x-1)} + \dots)(\mathrm{d}g_{z_v})^{\otimes k} \tag{9.15}$$

has a leading coefficient of 1. Such a $g_{z_v}$ exists since $\mathbf{C}_v$ is algebraically closed. We will use this $g_{z_v}$ shortly (and we do not care if $g_{z_v}$ extends meromorphically to all of $\mathfrak{h}_v$).

By Theorem 8.3, we have

$$\langle c, T_m(d^\sigma)\rangle_{v,\omega_x} = \lim_{y\to x}(\langle c_y, T_m(d^\sigma)\rangle_v - r_{\mathscr{A}}(m)\log|t_x(y)|_v),$$

where $y \in X_0(N)(\mathbf{C}_v) - \{x\}$ converges to $x$ and $c_y$ is the divisor obtained from $c$ by replacing $x$ with $y$. Combining this identity with (9.14), we may sum over *all* places $v$ of $H$ and exploit the product formula for $C_x u_x^k \in H^\times$ and the identity

$$\sum_{v\nmid\infty} \log|\mathfrak{a}_x^{-1}|_v = \log|\mathrm{N}_{H/\mathbf{Q}}(\mathfrak{a}_x)|$$

(with this fractional ideal norm viewed as a positive rational number) to obtain a formula for the global height pairing: $\langle c, T_m(d^\sigma)\rangle$ is equal to

$$\sum_{v\nmid\infty}\langle x, T_m(x^\sigma)\rangle_v^{\rm GZ} + \frac{r_{\mathscr{A}}(m)\log|{\rm N}_{H/\mathbf{Q}}(\mathfrak{a}_x)|}{r_x+k}$$
$$+\sum_{v|\infty}\lim_{y\to x}\left(\langle c_y, T_m(d^\sigma)\rangle_v - \frac{r_{\mathscr{A}}(m)}{r_x+k}((r_x+k)\log|t_x(y)|_v + \log|C_x u_x^k|_v)\right).$$

In order to put this into a more useful form, we need to examine the term

$$(r_x+k)\log|t_x(y)|_v + \log|C_x u_x^k|_v$$

for $v|\infty$. We may write

$$\pi_v^*(t_x^{\rm an}) = \alpha_{v,x} h_{v,x} g_{z_v}^{u_x}$$

for some $\alpha_{v,x}\in\mathbf{C}_v^\times$ and $h_{v,x}$ analytic near $z_v$ with $h_{v,x}(z_v)=1$. Since $\theta_v = (C_x t_x^{r_x}+\dots)({\rm d}t_x^{\otimes k})$, we get

$$\pi_v^*(\theta_v^{\rm an}) = (C_x\alpha_{v,x}^{r_x+k}u_x^k g_{z_v}^{u_x r_x + k(u_x-1)}+\dots)({\rm d}g_{z_v})^{\otimes k}.$$

We conclude via (9.15) that $C_x u_x^k \alpha_{v,x}^{r_x+k}=1$. Thus, $(r_x+k)\log|\alpha_{v,x}|_v = -\log|C_x u_x^k|_v$. Taking $y_v$ in $\mathfrak{h}_v$ near $z_v$ and lying over $y\in X_v(\mathbf{C}_v)$, we obtain

$$\begin{aligned}(r_x+k)\log|t_x(y)|_v + \log|C_x u_x^k|_v &= (r_x+k)(\log|\alpha_{v,x}|_v + \log|h_{v,x}(y_v)|_v + u_x\log|g_{z_v}(y_v)|_v) + \log|C_x u_x^k|_v\\ &= (r_x+k)u_x\log|g_{z_v}(y_v)|_v + (r_x+k)\log|h_{v,x}(y_v)|_v\end{aligned}$$

where $h_{v,x}(y_v)\to 1$ as $y_v\to z_v$ in $\mathfrak{h}_v$. Thus, when forming the limit as $y\to x$ in $X_v(\mathbf{C}_v)$ we can drop the $h_{v,x}(y)$ term and hence arrive at the main result:

**Theorem 9.6.** *Let $x\in X_0(N)(H)$ be a Heegner point, and let $u_x$ be half the size of the geometric automorphism group of $x$ (so $u_x = (1/2)|\mathscr{O}_K^\times|$). Choose a non-zero rational section $\theta$ of $(\Omega^1_{X_0(N)/H})^{\otimes k}$ with $r_x \overset{\rm def}{=} {\rm ord}_x(\theta)\neq -k$. Let $\mathfrak{a}_x$ be the fractional ideal of $H$ constructed via deformation theory as in* (9.13). *Define $c = x-\infty$, $d = x-0$.*

*The global height pairing $\langle c, T_m(d^\sigma)\rangle$ is equal to*

$$\sum_{v\nmid\infty}\langle x, T_m(x^\sigma)\rangle_v^{\rm GZ} + \sum_{v|\infty}\lim_{y_v\to z_v}(\langle c_{y_v}, T_m(d^\sigma)\rangle_v - u_x r_{\mathscr{A}}(m)\log|g_{z_v}(y_v)|_v) + \frac{r_{\mathscr{A}}(m)\log|{\rm N}_{H/\mathbf{Q}}\mathfrak{a}_x|}{r_x+k}, \tag{9.16}$$

*where*

- *the local term $\langle x, T_m(x^\sigma)\rangle_v^{\rm GZ}$ for $v\nmid\infty$ is defined by* (9.9) *and* (9.11),
- *for $v|\infty$, $z_v\in\mathfrak{h}_v$ projects onto $x_v^{\rm an}$ under the analytic uniformization $\pi_v:\mathfrak{h}_v\to X_0(N)^{\rm an}_{/H_v}$,*
- *$g_{z_v}$ is an analytic uniformizer at $z_v$ with respect to which the local analytic expansion of the meromorphic $k$-tensor $\pi_v^*(\theta_v^{\rm an})$ has leading coefficient equal to* 1.
- *for $v|\infty$, the limit runs over $y_v\in\mathfrak{h}_v-\{z_v\}$ converging to $z_v$,*

As we have seen above, the local factors of $\mathfrak{a}_x$ are determined by the incarnation of $\theta$ in local deformation theory on good models for $x$. There is a particularly nice choice to be made in our situation, namely $k=6$ and $\theta=\Delta$. Since $\Delta$ *over* $\mathbf{C}$ is nowhere vanishing away from the cusps, this choice renders

$$r_x u_x + k(u_x-1) = 0 \tag{9.17}$$

for all $x$ (cf. (9.15)). In particular, $r_x+k\neq 0$ for all $x$ (so we may indeed use $\Delta$ as the basic tensor in the preceding considerations). Consequently, since $\Delta(q)(\frac{{\rm d}q}{q})^{\otimes 6} = ((2\pi i)\eta^4(z){\rm d}z)^{\otimes 6}$, with this choice it is trivial to check that at *any* (fixed) point $z_v\in\mathfrak{h}_v$ we may take the analytic uniformizer $g_{z_v}(z') = (2\pi i)\eta^4(z_v)(z'-z_v)$ for varying $z'\in\mathfrak{h}_v$. With this choice, the contribution from $v|\infty$ in (9.16) is an "explicit" limit involving archimedean local heights and certain analytic functions $g_{z_v}$ on the upper half-plane over $H_v\simeq\mathbf{C}$, and is exactly the archimedean height contribution which is computed in [GZ, Ch II, (5.3), (5.5)]. Since $\log p$'s for distinct rational primes $p$ are $\mathbf{Q}$-linearly independent, it follows that Theorem 9.6 in this case must agree place-by-place (over $\mathbf{Q}$) with the global formula used in [GZ], where the contribution in (9.16) over a prime $p$ comes from the first and third pieces of the formula (upon factoring out a $-\log q_v$, using Theorem 8.4):

$$\langle c, T_m d^\sigma\rangle_p \stackrel{\text{def}}{=} -\sum_{v|p}((x.T_m(x^\sigma))_v^{\mathrm{GZ}} - \frac{r_{\mathscr{A}}(m)}{r_x+k}\mathrm{ord}_{v,x}(\theta))\log q_v. \tag{9.18}$$

The aim of §10 is to make the $v$-term on the right side of (9.18) explicit when $\theta = \Delta$; see Theorem 10.5. The appendix uses this to give an explication of (9.18) depending on the splitting behavior of $p$ in $K$.

The method we have outlined is of fairly general nature for computing global height pairings on coarse moduli schemes. The real problem in any given situation is to make the archimedean height pairings and analytic functions $g_{z_v}$ explicit enough to carry out computations, and to compute the ideal $\mathfrak{a}_x$ (this latter data being the non-archimedean aspect which is sensitive to the choice of $\theta$). The trick is to find a single well-understood $\theta$ that will satisfy $\mathrm{ord}_x(\theta) \neq -k$ at *all* Heegner points. For our purposes, the choice $\theta = \Delta$ works best.

## 10. Quaternionic explications

We are now in position to carry out the computation of the local term

$$(x.T_m(x^\sigma))_v^{\mathrm{GZ}} - \frac{r_{\mathscr{A}}(m)}{r_x+k}\mathrm{ord}_{v,x}(\theta) \tag{10.1}$$

from (9.18) for *any* value of $r_{\mathscr{A}}(m)$ but with $\theta = \Delta$. The difference (10.1) is the main geometric local contribution to the non-archimedean contribution (9.18) to the global height pairing formula in Theorem 9.6, with the intersection pairing $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$ computed by means of usual local intersection pairings for disjoint divisors and the modified self-pairing $(x.x)_v^{\mathrm{GZ}}$ given by (9.9). In particular, we see that $(x.x)_v^{\mathrm{GZ}} = 0$ whenever $\mathrm{Aut}_W(\underline{x}) = \mathrm{Aut}_{W/\pi}(\underline{x}_0)$.

Since the set $\mathrm{Isom}_W(\underline{x},\underline{y})$ of $W$-isomorphisms is empty when the corresponding generic geometric points $x, y$ are distinct (thanks to Theorem 2.6), we conclude trivially that the identity

$$(x.y)_v^{\mathrm{GZ}} = \frac{1}{2}\sum_{n\geq 0}|\mathrm{Isom}_{W_n}^{\mathrm{new}}(\underline{x},\underline{y})| \tag{10.2}$$

holds for *all* (possibly equal) pairs of points $x, y \in X(F)$ whose associated $W$-points come from (necessarily unique up to isomorphism) $\Gamma_0(N)$-diagrams over $W$, where we define a "new" isomorphism as one not lifting to $W$. Keep in mind that when $x \neq y$ we are using Theorem 4.1, and in this formula the local intersection pairing on the left side is computed on the $W$-scheme $\underline{X} = X_0(N)_{/W}$ (which is often non-regular, but points arising from Heegner data and Hecke correspondences thereof lie in the relative smooth locus, so there is no ambiguity concerning these local intersection numbers).

Consequently, the *exact* same method which we used to prove Theorem 5.1 in the case $p \nmid m$ carries over to give the (finite) formula

$$(x.T_m(x^\sigma))_v^{\mathrm{GZ}} = \frac{1}{2}\sum_{n\geq 0}|\,\mathrm{Hom}_{W_n}^{\mathrm{new}}(\underline{x^\sigma},\underline{x})_{\deg(m)}| \tag{10.3}$$

whenever $v \nmid m$, where a "new" homomorphism is one not lifting to $W$. Once again, $\underline{x} \in \underline{X}(W)$ can be *any* non-cuspidal section which is represented by a $\Gamma_0(N)$-diagram over $W$ (and every $W$-point coming from a Heegner point over $H$ has been seen to have this property), with $x$ the generic fiber of $\underline{x}$ over $W$.

In order to make progress toward computing (10.1) explicitly when $\theta = \Delta$, we need to compute $\mathrm{ord}_{v,x}(\Delta)$. This is provided by:

**Lemma 10.1.** *Let $\underline{x} \in \underline{X}(W)$ be disjoint from the cuspidal locus and represented by a $\Gamma_0(N)$-diagram. When $v \nmid N$ or $v|\mathfrak{n}$, then $\mathrm{ord}_{v,x}(\Delta) = 0$. When $v|\overline{\mathfrak{n}}$ then $\mathrm{ord}_{v,x}(\Delta)/(r_x+6) = u_x\mathrm{ord}_v(N)$. In other words, for all $v \nmid \infty$ $\mathrm{ord}_{v,x}(\Delta) = u_x(r_x+k)\mathrm{ord}_v(\overline{\mathfrak{n}}) = k\cdot\mathrm{ord}_v(\overline{\mathfrak{n}})$, so $\mathfrak{a}_x = \overline{\mathfrak{n}}^k = \overline{\mathfrak{n}}^6$.*

*Remark* 10.2. The method of proof is fairly abstract, so the same method should give a formula for $\mathfrak{a}_x$ in terms of $x : \mathrm{Spec}(\mathscr{O}_H) \to \mathscr{M}_{1,1}$ *if* $\theta$ merely comes from level 1 (over $\mathscr{O}_H$). Since we have to choose a specific $\theta$ eventually, it seems simplest to just prove Lemma 10.1 for $\theta = \Delta$ and to leave more general considerations to the reader's imagination.

*Proof.* By (9.17), we have $(r_x + 6)u_x = 6$, so when $v|\mathfrak{n}$ we really want to prove $\mathrm{ord}_{v,x}(\Delta) = \mathrm{ord}_v(N^6)$. The exponent of 6 will arise from the fact that $\Delta$ is a 6-tensor. Let $R$ denote the universal deformation ring of the elliptic curve $E_0$ underlying $x_0$, equipped with universal (algebraized) elliptic curve $E_{/R}$, and let $R'$ denote the universal deformation ring of the $\Gamma_0(N)$-structure $x_0$, with $\iota$ the corresponding universal structure on the base change $E_{/R'}$ (with closed fiber $\iota_0$ recovering the level structure $x_0$ enhancing $E_0$). By the general theory, since the $\Gamma_0(N)$-moduli problem is finite flat over the moduli stack of (smooth) elliptic curves it follows that $R'$ is a finite flat $R$-algebra: it is simply the factor ring of the finite flat universal $\Gamma_0(N)$-structure algebra over $R$ corresponding to the residual structure $x_0$ on $E_0$.

The crucial point is that when $v \nmid N$ or $v|\mathfrak{n}$ then $R \to R'$ is an *isomorphism*. Indeed, in the case $v \nmid N$ the $N$-torsion on deformations of $E_0$ is *étale* and hence is invisible from the point of view of deformation theory. For the same reason, when $v|N$ then the prime-to-$p$ part of the $\Gamma_0(N)$-structure lifts uniquely to any deformation of $E_0$. All the action therefore happens in the $p$-part. If $v|\mathfrak{n}$ then the residual $p$-part of the level structure is multiplicative and more specifically is selected out functorially by the connected-étale sequence on the $p^{\mathrm{ord}_p(N)}$-torsion of deformations of $E_0$. Hence, we once again only have to deform $E_0$ and the level structure uniquely compatibly deforms for free, so again $R \to R'$ is an isomorphism. The case $v|\overline{\mathfrak{n}}$ is more subtle (the entire $\Gamma_0(N)$-structure is étale, but specifying the $p$-part on a deformation amounts to *splitting* the connected-étale sequence, so this imposes genuine extra data on deformations, leading to a bigger deformation ring than the one for $E_0$ alone).

Recall that in general $\mathrm{ord}_{v,x}(\theta)$ is defined as follows: we have an isomorphism $R' \simeq W[\![T]\!]$ with $T = 0$ corresponding to the $W$-structure $x$ (recall that $R'$ is regular and $W$-flat of dimension 2 by the general theory as in [KM], and it has a $W$-section arising from $x$ and hence really does have to have the form $W[\![T]\!]$), and $\mathrm{ord}_{v,x}(\theta)$ is defined to be $\mathrm{ord}_v$ of the leading coefficient of the $T$-adic expansion of the $k$-tensor $\theta$. It makes a big difference whether $R'$ coincides with the universal deformation ring of $E_0$ or if it is genuinely bigger.

We first dispose of the cases $v \nmid N$ and $v|\mathfrak{n}$, and then will settle $v|\overline{\mathfrak{n}}$. As long as $v \nmid \overline{\mathfrak{n}}$, we have just seen that $R' = R$. Thus, if we write $W[\![T]\!]$ for the universal deformation ring of $E_0$ (with $T = 0$ cutting out the elliptic curve over $W$ underlying $x$) then we must show that the 6-tensor $\Delta$ has $T$-adic expansion with unit leading coefficient. The crux of the matter is that $\Delta$ is a generator of the line bundle $\omega^{\otimes 12}$ on the (open) moduli stack $\mathscr{M}_{1,1}$, and the Kodaira-Spencer map $\mathrm{KS}_{1,1} : \omega \to \Omega^1_{\mathscr{M}_{1,1}/\mathbf{Z}}$ is what converts it into a 6-tensor on this stack. This is the usual recipe that converts even weight modular forms into tensors on modular curves. The general theory of the Kodaira-Spencer map ensures that $\mathrm{KS}_{1,1}$ is an *isomorphism* (we are working away from the cuspidal substack) precisely because the deformation theory of an elliptic curve (with the marked identity section!) coincides with the deformation theory of its underlying "bare" curve (as we can always lift the identity section, thanks to smoothness, and can then translate it via the group law to put it in the correct position). Consequently, the image of $\Delta$ in the invertible $R$-module $(\widehat{\Omega}^1_{R/W})^{\otimes 6}$ is a generator and hence it has unit leading coefficient.

There remains the case $v|\overline{\mathfrak{n}}$. Since $v|N$, so $p$ is split in $K$, the elliptic curve underlying $x$ is the Serre-Tate canonical lift of $E_0$. Moreover, the deformation theory of $E_0$ coincides with that of its $p$-divisible group. Thus, we may identify $R$ with $W[\![T]\!]$ where $q = 1 + T$ is the so-called Serre-Tate parameter. The deformation theory of the $\Gamma_0(N)$-structure only matters through its $p$-part (as the other primary components deform uniquely). Let $p^e$ be the $p$-part of $N$ (with $e > 0$). We claim

$$R' = R[T']/((1 + T')^{p^e} - (1 + T)) \simeq W[\![T']\!]. \tag{10.4}$$

Once this is shown, then $T = p^e(T' + \dots) + T'^{p^e}$, so $\Delta = \mathrm{unit}'(\mathrm{d}T)^{\otimes 6} = \mathrm{unit}'(p^e)^6(\mathrm{d}T')^{\otimes 6}$. More intrinsically, $\Delta$ in the invertible $R$-module $\widehat{\Omega}^1_{R'/W}$ is $(p^e)^6$ times a generator, whence $\mathrm{ord}_{v,x}(\Delta) = \mathrm{ord}_v((p^e)^6) = \mathrm{ord}_v(N^6) = \mathrm{ord}_p(N^6)$ as desired.

To establish (10.4) (and thereby complete the proof), recall that the deformation aspect of the level structure only matters through its $p$-part. More specifically, we are not just imposing an arbitrary $\Gamma_0(p^e)$-structure on deformations of $E_0$ but rather one which is *étale* (here is where the condition $v|\overline{\mathfrak{n}}$ is used). Hence, our task is really one of deforming *splittings* of the connected-étale sequence on $p^e$-torsion of deformations of the ordinary $E_0$. Upon choosing a generator of the étale part of $E_0[p^e]$, we uniquely identify $E_0[p^e]$ with $\mathbf{Z}/p^e \times \mu_{p^e}$ compatibly with the scheme-theoretic Weil pairing. For any infinitesimal deformation $E'$ of $E_0$ we can uniquely write its connected-étale sequence as $0 \to \mu_{p^e} \to E'[p^e] \to \mathbf{Z}/p^e \to 0$ in a manner lifting the corresponding canonical sequence for $E_0[p^e]$. Let $\pi$ denote the projection map from $E'[p^e]$ onto its maximal étale quotient. Thus, the deformation problem is equivalently one of choosing a section of the fiber $\pi^{-1}(1)$ on deformations $E'$ of $E_0$. In the universal situation over $W[\![T]\!]$ (with $q = 1+T$ the Serre-Tate canonical parameter and $T = 0$ cutting out the Serre-Tate canonical lift which is the elliptic curve underlying $x$), the torsion levels are described by rather explicit group schemes considered in [KM, §8.9ff]. This description rests crucially on the fact that $q = 1+T$ is the Serre-Tate canonical parameter, and in any case explicates the scheme $\pi^{-1}(1)$ by means of the equation $X^{p^e} = 1+T$. Thus, if we define $T' = X - 1$ then we obtain the desired description $R' = R[T']/((1+T')^{p^e} - (1+T))$ for the (local) ring $R'$ as a finite $R$-algebra. ∎

As an application of (10.3), since maps between *ordinary* elliptic curves always uniquely lift to maps between Serre-Tate canonical deformations, we see that if $p$ splits in $K$ and $p \nmid mN$ (so (10.3) can be applied and maps between level $N$-structures uniquely deform) then the $\mathrm{Hom}^{\mathrm{new}}$-terms on the right side of (10.3) vanish. This yields the following generalization of the vanishing observation which was noted at the beginning of §7:

**Lemma 10.3.** *If $p$ is split in $K$, $\gcd(m,N) = 1$, and $p \nmid mN$, then $(\underline{x}.T_m(\underline{x^\sigma}))_v^{\mathrm{GZ}} = 0$ for any place $v$ of $H$ over $p$, without restriction on $r_{\mathscr{A}}(m)$.*

How about the case in which $p|N$ (so $p$ splits in $K$ and $p \nmid m$, but the $N$-torsion schemes need not be étale)? This is answered by:

**Theorem 10.4.** *Assume that $v|N$ and $\gcd(m,N) = 1$. Let $\underline{x} \in \underline{X}(W)$ be a Heegner point. Then*

$$(\underline{x}.T_m(\underline{x^\sigma}))_v^{\mathrm{GZ}} - \frac{r_{\mathscr{A}}(m)}{r_x + k}\mathrm{ord}_{v,x}(\Delta) = \begin{cases} 0 & \text{if } v|\mathfrak{n} \\ -u_x r_{\mathscr{A}}(m)\mathrm{ord}_p(N) & \text{if } v|\overline{\mathfrak{n}} \end{cases}$$

*Proof.* There are two cases to consider, depending on the prime factor $\mathfrak{n}$ of $N$ over $p$ which kills the level structure on $\underline{x}$. If $v|\mathfrak{n}$, then the $p$-part of the level structure on $\underline{x}$ (and hence also on $\underline{x^\sigma}$) is connected and in fact multiplicative. Since infinitesimal deformations of ordinary elliptic curves possess *unique* connected multiplicative subgroups of a specified order and any map between such deformations respects the specification of such a subgroup (thanks to the functoriality of the connected-étale sequence), we can conclude by the same argument as for $p \nmid mN$ split in $K$ that the right side of (10.3) vanishes when $v|\mathfrak{n}$. This gives the $v|\mathfrak{n}$ case of Theorem 10.4, since $\mathrm{ord}_{v,x}(\Delta) = 0$ by Lemma 10.1 for such $v$.

The case $v|\overline{\mathfrak{n}}$, for which the $p$-part of the $\Gamma_0(N)$-structure is étale, has the property that the $p$-part of the $\Gamma_0(N)$-structure on $\underline{x}$ and $\underline{x^\sigma}$ amounts to giving (non-canonical) splittings of connected-étale sequences over $W$ so as to single out the $p$-part of the étale level structures. General elements in $\mathrm{Hom}_{W/\pi}(\underline{x^\sigma}, \underline{x})$ have no non-trivial compatibility condition imposed on the $p$-part of the level structure because connected-étale sequences uniquely and canonically split over $W/\pi$. However, when such a map is uniquely lifted to a map of the (Serre-Tate canonical) deformed elliptic curves underlying $\underline{x}$ and $\underline{x^\sigma}$ over $W$ it is generally not true that such a lifted map must respect chosen splittings of the connected-étale sequences over $W$ (and hence generally does not give a map of $\Gamma_0(N)$-structures over $W$). But our situation is special because the $p$-part of the level structure coincides with the piece of the $\overline{\mathfrak{n}}$-divisible group of order equal to the $p$-part of $N$. Since there are no non-zero maps from an étale $p$-divisible group to a connected one (over a local noetherian base), we conclude that the argument used for $v|\mathfrak{n}$ actually still works for $v|\overline{\mathfrak{n}}$. Thus, we get $(x.T_m(x^\sigma))_v^{\mathrm{GZ}} = 0$ even when $v|\overline{\mathfrak{n}}$, so we must prove $\mathrm{ord}_{v,x}(\Delta)/(r_x + k) = u_x \mathrm{ord}_p(N) = u_x \mathrm{ord}_v(N)$ (the latter equality holding since $p$ is unramified in $H$). This is provided by Lemma 10.1. ∎

Now we compute (10.1) when $v \nmid N$ and $\theta = \Delta$. By Lemma 10.1, the $\mathrm{ord}_{v,x}(\Delta)$ term vanishes for such $v$. Thus, our task comes down to computing $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$.

**Theorem 10.5.** *Assume $v \nmid N$, $\gcd(m, N) = 1$, and $\underline{x} \in \underline{X}(W)$ is a Heegner point as usual. Let $\mathfrak{a}$ represent the ideal class $\mathscr{A}$ and be prime to $p$, and let $\sigma \in \mathrm{Gal}(H/K)$ correspond to $\mathscr{A}$. Let $v$ be a place of $H$ over $p$.*

(1) *If $p$ is inert in $K$, then*

$$(x.T_m(x^\sigma))_v^{\mathrm{GZ}} = \sum_{b \in R\mathfrak{a}/\pm 1, \mathrm{N}b = m\mathrm{N}\mathfrak{a}, b_- \neq 0} \frac{1}{2}(1 + \mathrm{ord}_p(\mathrm{N}(b_-))) \ + \ \frac{1}{2} u_x r_{\mathscr{A}}(m)\mathrm{ord}_p(m).$$

(2) *If $p$ is ramified in $K$ then*

$$(x.T_m(x^\sigma))_v^{\mathrm{GZ}} = \sum_{b \in R\mathfrak{a}/\pm 1, \mathrm{N}b = m\mathrm{N}\mathfrak{a}, b_- \neq 0} \mathrm{ord}_p(D\mathrm{N}(b_-)) \ + \ u_x r_{\mathscr{A}}(m)\mathrm{ord}_p(m).$$

(3) *If $p = \mathfrak{p} \cdot \overline{\mathfrak{p}}$ is split in $K$ and $v$ lies over $\mathfrak{p}$ then*

$$(x.T_m(x^\sigma))_v^{\mathrm{GZ}} = u_x \kappa_{\mathfrak{p}}$$

*where $\kappa_{\mathfrak{p}}$ and $\kappa_{\overline{\mathfrak{p}}}$ are non-negative integers which are intrinsic to the prime ideals $\mathfrak{p}$ and $\overline{\mathfrak{p}}$ (e.g., they are defined without reference to $x$) and satisfy $\kappa_{\mathfrak{p}} + \kappa_{\overline{\mathfrak{p}}} = r_{\mathscr{A}}(m)\mathrm{ord}_p(m)$.*

This theorem is [GZ, Ch III, Prop 8.5]. Our proof is longer, but more conceptual.

*Proof.* The supersingular cases (i.e., $p$ not split in $K$) will be treated by essentially the exact same method which we used when $r_{\mathscr{A}}(m) = 0$, but the ordinary case (i.e., $p$ split in $K$) will require some new work.

It follows from (10.3) that the quaternionic formulas from Corollary 7.15 carry over to the general case (i.e., $r_{\mathscr{A}}(m) > 0$ is allowed) to compute $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$ *when we also assume $v \nmid m$*, provided one augments the condition on the summation to require $b_- \neq 0$ (as we recall that this is the quaternionic translation of the condition that a morphism not lift to $W$). This gives the first two cases of the theorem when $\mathrm{ord}_p(m) = 0$. Note that Corollary 7.15 only occurs when $p \nmid N$ (as $p|N$ puts us in the split cases). The more interesting case of Corollary 7.15 is $v|m$, or equivalently $p|m$.

Before considering the modifications needed to get the theorem when $p|m$ (so $p \nmid N$) and $r_{\mathscr{A}}(m) > 0$, we first note that if $p$ is inert in $K$ (so $p^2$ is the norm of the unique prime $p\mathscr{O}_K$ over $p$) then the positivity of $r_{\mathscr{A}}(m)$ forces $\mathrm{ord}_p(m)$ to be *even*, with $(1/2)\mathrm{ord}_v(m) = (1/2)\mathrm{ord}_p(m)$. Meanwhile, if $p$ is ramified in $K$ then $\mathrm{ord}_v(m)$ is even with $(1/2)\mathrm{ord}_v(m) = \mathrm{ord}_p(m)$. Thus, the "extra" term on the right side of the inert and ramified cases of the theorem can be uniformly described by the formula

$$(1/2)ur_{\mathscr{A}}(m)\mathrm{ord}_v(m), \tag{10.5}$$

where $u = u(K) = u_x$.

Let $m = p^t r$ with $p \nmid r$ and $t > 0$. The first two cases of the theorem require nothing beyond our earlier work on the analogues when $r_{\mathscr{A}}(m) = 0$. The condition $v \nmid N$ is crucial throughout, since the deformation theory analysis uses quite critically that $N$-torsion is étale (and hence deformations of elliptic curves maps over any $W_r$ are automatically $\Gamma_0(N)$-compatible when this is true over $W_0$). Also, the condition $r_{\mathscr{A}}(m) = 0$ played essentially *no* role in our earlier treatment of supersingular cases. The only relevance of this condition was to ensure that when cases with $s = 0$ arise then the various $W$-structures $\underline{z} = \underline{y}(0)$ which show up in $T_m(\underline{x^\sigma})$ for $\underline{z}$ in $T_r(\underline{x^\sigma})$ are never equal to $\underline{x}$. If we allow $r_{\mathscr{A}}(m)$ to perhaps be positive, then the analysis of the $\underline{y}(s)$'s for $s > 0$ goes through completely unchanged (since the value of $r_{\mathscr{A}}(m)$ was never relevant in that analysis) and the associated closed points $\underline{y}(s)_{/F}$ on $\underline{X}_{/F}$ were shown to have residue field $F(s)$ strictly bigger than $F$. In particular, such points cannot contribute to an appearance of the $F$-rational point $\underline{x}$ in the divisor $T_m(\underline{x^\sigma})$. Thus, for $r_{\mathscr{A}}(m) > 0$ we get almost the exact same formulas for $(\underline{x}.T_m(\underline{x^\sigma}))_v^{\mathrm{GZ}}$ as in (6.5) and (6.6) with the two modifications that the case of $p$ inert in $K$ only gives rise to cases with *even* $t = \mathrm{ord}_p(m)$ – as we have seen in the deduction of (10.5) – and the summation terms

$$\frac{1}{2} \cdot \sum_{n \geq 0} \# \operatorname{Hom}_{W_n}(\underline{z}, \underline{x})_{\deg 1}$$

which arise as formulas for $(\underline{z}.\underline{x})$ must be replaced with their "new" counterparts in accordance with (10.2). This latter modification causes the $\sum_{n\geq 0}(\dots)$ terms in the "$r_{\mathscr{A}}(m)=0$" formula for $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$ in (6.5) and (6.6) to be replaced with "new" counterparts as well.

If we look back at our argument which used Theorem 6.2 and Theorem 6.1 to translate (6.5) and (6.6) into the language of cardinalities of Hom-groups as in Theorem 5.1, we see that those arguments *never* used any hypothesis concerning the value of $r_{\mathscr{A}}(m)$. Thus, when converting the terms

$$\frac{t}{2}\cdot\frac{1}{2}\#\operatorname{Hom}_{W/\pi}(\underline{x^\sigma},\underline{x})_{\deg r},\ \ t\cdot\frac{1}{2}\#\operatorname{Hom}_{W/\pi}(\underline{x^\sigma},\underline{x})_{\deg r},\ \ t\cdot\frac{1}{2}\#\operatorname{Hom}_{W/\pi}((\underline{x^\sigma})^{\sigma_{\mathfrak{p}}},\underline{x})_{\deg r} \tag{10.6}$$

into various $(1/2)\#\operatorname{Hom}_{W_e}(\underline{x^\sigma},\underline{x})$'s for small $e$, we just have to account for that fact that we now want to break up such terms into a "new" part and a "non-new" part. As we see from (10.6), there are $t/2=(1/2)\mathrm{ord}_p(m)$ such terms when $p$ is inert in $K$ and there are $t=\mathrm{ord}_p(m)$ such terms when $p$ is ramified in $K$. We thereby pick up a "non-new" contribution of

$$(t/2)\cdot(1/2)|\operatorname{Hom}_W(\underline{x^\sigma},\underline{x})_{\deg m}|=(t/2)ur_{\mathscr{A}}(m)=\frac{1}{2}ur_{\mathscr{A}}(m)\mathrm{ord}_p(m)$$

when $p$ is inert in $K$ and

$$t\cdot(1/2)|\operatorname{Hom}_W(\underline{x^\sigma},\underline{x})_{\deg m}|=ur_{\mathscr{A}}(m)\mathrm{ord}_p(m)$$

when $p$ is ramified in $K$, with the remaining "new" part of $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$ given by the right side of (10.3). We compute the right side of (10.3) in our supersingular situation by the exact same quaternionic method as before, and this yields the summation terms given in the first two parts of the theorem, the extra condition $b_-\neq 0$ simply reflecting the fact that we're only counting the number of "new" homomorphisms (i.e., those not lifting to $W$) at each infinitesimal level. This completes the justification of the first two parts of the theorem.

Now we turn to the third part of the theorem, with $p=\mathfrak{p}\overline{\mathfrak{p}}$ split in $\mathscr{O}_K$ (and $p\nmid N$), so the Heegner data have underlying elliptic curves with ordinary reduction. We have already seen via (10.3) that when $p\nmid mN$ is split in $K$ then $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}=0$, exactly in accordance with the asserted formula in the split case. Thus, from now on we may (and do) assume $\mathrm{ord}_p(m)>0$ (which forces $p\nmid N$ anyway). We cannot use (10.3) in such cases, so we will give a more explicit analysis of the situation. The goal is to get a formula for $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$ which agrees with the one we want for $p|m$ split in $K$.

As a first step toward computing $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$ when $p$ is split in $K$ and $v\nmid N$ (but $p|m$), we need to determine which order $m$ subgroup schemes $C\subseteq\underline{x^\sigma}$ have the property that the quotient $\underline{x^\sigma}_C$ by $C$ has closed fiber isomorphic to the closed fiber of $\underline{x}$ (with the isomorphism respecting the level structures). That is, we want to know when $C_0$ is the kernel of an isogeny $\phi_0:\underline{x^\sigma}_0\to\underline{x}_0$ respecting the $\Gamma_0(N)$-structures. By the Serre-Tate theorem, we have $\operatorname{Hom}_{W/\pi}(\underline{x^\sigma}_0,\underline{x}_0)=\operatorname{Hom}_W(\underline{x^\sigma},\underline{x})$ since $v\nmid N$ (so compatibility with $\Gamma_0(N)$-structures does not impose any additional condition when deforming maps). In particular,

$$(x.x)_v^{\mathrm{GZ}}=\frac{1}{2}\sum_{n\geq 0}(|\operatorname{Aut}_{W_n}(\underline{x})|-|\operatorname{Aut}_W(\underline{x})|)=0. \tag{10.7}$$

Any map $\underline{x^\sigma}\to\underline{x}$ is automatically $\mathscr{O}_K$-compatible (as can be checked by showing that the $\mathscr{O}_H$-module tangent "lines" of $x$ and $x^\sigma$ have the same (canonical) $\mathscr{O}_K$-structure through action of $\mathscr{O}_K$ on the elliptic curve, and this is trivial to check since $\sigma\in\mathrm{Gal}(H/K)$ acts as the identity on $K\subseteq H$). It follows that $\phi_0$ as above is $\mathscr{O}_K$-compatible, so $C_0$ is an $\mathscr{O}_K$-submodule scheme of $\underline{x^\sigma}_0$.

Let $\mathfrak{a}$ be an integral ideal in the ideal class $\mathscr{A}$ with norm $m$. Consider the unique degree $m$ isogeny

$$\phi:\mathfrak{a}^{-1}\otimes\underline{x}=\underline{x^\sigma}\to\underline{x}$$

lifting $\phi_0$, so $\phi\in\mathfrak{a}$ and $\mathrm{N}(\phi)=\deg(\phi)\mathrm{N}(\mathfrak{a})=m^2$ (as one checks using the same arguments with $\ell$-divisible groups that we employed in our earlier degree calculations during the quaternionic considerations). In other words, in $\mathscr{O}_K$ we have an equality of ideals $\phi\mathscr{O}_K=\mathfrak{a}\cdot\mathfrak{b}$ with $\mathfrak{b}$ an integral ideal of norm $m$ in the ideal class $\mathscr{A}^{-1}$. Conversely, it is clear that when we are given an integral ideal in $\mathscr{A}^{-1}$ with norm $m$ then we get a map $\phi$ as considered above. Thus, up to composing $\phi$ (or $\phi_0$) with a unit of $\mathscr{O}_K$ we see that the set of order $m$ subgroups $C_0\subseteq\underline{x^\sigma}_0$ of interest to us is of cardinality equal to $r_{\mathscr{A}^{-1}}(m)=r_{\mathscr{A}}(m)$. We fix such a $\mathfrak{b}$ and a representative generator $\phi_{\mathfrak{b}}$ of $\mathfrak{ab}$, and we seek to determine all order $m$ subgroups $C \subseteq \underline{x^\sigma}_{/\overline{F}}$ for which $C_0 = \ker((\phi_{\mathfrak{b}})_0)$, written more loosely as "$C \equiv \ker((\phi_{\mathfrak{b}})_0)$" (changing $\phi_{\mathfrak{b}}$ by an $\mathscr{O}_K^\times$-multiple doesn't matter).

To find all order $m$ subgroups $C \subseteq x^\sigma_{/\overline{F}}$ such that $C \equiv \ker((\phi_{\mathfrak{b}})_0)$, note that on the prime-to-$p$ part everything is uniquely determined, so we focus our attention on the $p$-part. Let $t = \mathrm{ord}_p(m) > 0$. There is an evident inclusion $\overline{\mathfrak{b}} \cdot (\mathfrak{a}^{-1} \otimes \underline{x}[p^t]) \subseteq \ker(\phi_{\mathfrak{b}})_p$ inside of the $p$-part of $\ker(\phi_{\mathfrak{b}})$ and this is an equality for order reasons. Because of the $\mathscr{O}_K$-compatibility of $\phi_{\mathfrak{b}}$ and the canonical splitting of the $p$-divisible groups of $\underline{x}$ and $\underline{x^\sigma}$ into $\mathfrak{p}$-divisible and $\overline{\mathfrak{p}}$-divisible parts (where $p\mathscr{O}_K = \mathfrak{p}\overline{\mathfrak{p}}$), with the $\mathfrak{p}$ denoting the prime under $v$, we have a decomposition $\ker(\phi_{\mathfrak{b}})_p = (\overline{\mathfrak{b}} \cdot (\mathfrak{a}^{-1} \otimes \underline{x}[p^t]))_{\overline{\mathfrak{p}}} \times (\overline{\mathfrak{b}} \cdot (\mathfrak{a}^{-1} \otimes \underline{x}[p^t]))_{\mathfrak{p}}$ which expresses $\ker(\phi_{\mathfrak{b}})_p$ as the product of étale and connected pieces over $W$ (the first factor is étale and the second is connected because $\mathfrak{p}$ is the prime under $v$, and everything is really finite flat because of Raynaud's scheme-theoretic closure trick which is particularly well-behaved in the context of multiplicative and étale group schemes without ramification restrictions). The analogous decomposition of $\mathfrak{a}^{-1} \otimes \underline{x}[p^t] = \underline{x^\sigma}[p^t]$ into a $\overline{\mathfrak{p}}$-part and a $\mathfrak{p}$-part reflects the decomposition arising from the unique splitting of the $p$-divisible group of $\underline{x^\sigma}$ into étale and connected pieces.

By the Serre-Tate theorem, we conclude (by considering deformations of endomorphisms of the $p$-divisible groups $\mathbf{Q}_p/\mathbf{Z}_p$ and its dual $\mathbf{G}_m[p^\infty]$ separately) that the condition that a deformation of $(\phi_{\mathfrak{b}})_0$ to some $W'/\pi'^{n+1}$ *not* lift to $W'$ (with $W'$ a finite discrete valuation ring extension of $W$) is *exactly* that its kernel not split as a product of connected and étale pieces in accordance with the splitting of the $p$-divisible group of $\underline{x^\sigma}$ (reduced modulo $\pi^{n+1}$). Thus, contributions to the points in the divisor $T_m(x^\sigma)$ (aside from $r_{\mathscr{A}}(m) \cdot x$) which "reduce to $(\phi_{\mathfrak{b}})_0$" correspond on the level of *geometric* points to subgroups $C$ inside of the geometric generic fiber of the finite flat $W$-group scheme $\mathfrak{a}^{-1} \otimes \underline{x}[p^t]$ with $(\overline{\mathfrak{b}} \cdot (\mathfrak{a}^{-1} \otimes \underline{x}[p^t]))_{\mathfrak{p}} \subseteq C$ and the cokernel of this inclusion projecting isomorphically onto $(\overline{\mathfrak{b}} \cdot (\mathfrak{a}^{-1} \otimes \underline{x}[p^t]))_{\overline{\mathfrak{p}}}$, yet with this quotient map *not* being split compatibly with the natural (Serre-Tate "canonical") splitting of the connected-étale sequence of $\mathfrak{a}^{-1} \otimes \underline{x}[p^t]$. The analysis of such $C$'s (e.g., finding Galois orbits, etc) as $\overline{F}$-points takes places entirely in the $p$-divisible group of $\underline{x^\sigma} = \mathfrak{a}^{-1} \otimes \underline{x}$, so since this $p$-divisible group only depends on $\mathfrak{a}$ through the $\mathscr{O}$-module $\mathfrak{a} \otimes_{\mathscr{O}_K} \mathscr{O}$ which is free of rank 1, we can drop the appearance of the functor "$\mathfrak{a}^{-1} \otimes (\cdot)$" without any harm.

As we run over all $r_{\mathscr{A}^{-1}}(m) = r_{\mathscr{A}}(m)$ possibilities for $\mathfrak{b}$, only the $p$-part $\mathfrak{p}^i \overline{\mathfrak{p}}^{t-i}$ of $\mathfrak{b}$ matters for the analysis of possible $C$'s (with $\mathfrak{a}$ harmlessly removed). This $p$-part sequence is $0 \to \underline{x}[\mathfrak{p}^i] \to C \to \underline{x}[\overline{\mathfrak{p}}^{t-i}] \to 0$. Upon trivializing the étale part of the $p$-divisible group of $\underline{x}$, we may naturally identify $\underline{x}[p^e]$ with $\mu_{p^e} \times \mathbf{Z}/p^e$, with $\mu_{p^e} = \underline{x}[\mathfrak{p}^e]$ and $\mathbf{Z}/p^e = \underline{x}[\overline{\mathfrak{p}}^e]$, all this compatible with change in $e$. Thus, we really have

$$0 \to \mu_{p^i} \to C \to p^i\mathbf{Z}/p^t\mathbf{Z} \to 0$$

inside of $\mu_{p^t} \times \mathbf{Z}/p^t\mathbf{Z}$. Passing to the quotient by $\mu_{p^i}$ on $C$ and taking the part of $\underline{x}[p^t]$ lying over $p^i\mathbf{Z}/p^t\mathbf{Z} \subseteq \mathbf{Z}/p^t\mathbf{Z}$, upon using the canonical isomorphisms $\mu_{p^t}/\mu_{p^i} \simeq \mu_{p^{t-i}}$ and $p^i\mathbf{Z}/p^t\mathbf{Z} \simeq \mathbf{Z}/p^{t-i}\mathbf{Z}$ we see that the specification of $C$ (now collapsed to its étale quotient) amounts to giving a *non-trivial* splitting over $\overline{F}$ of the base change of the connected-étale sequence of $\mu_{p^{t-i}} \times \mathbf{Z}/p^{t-i}$ (the trivial splitting contributes to the point $\underline{x}$ in $T_m(\underline{x^\sigma})$, but $(x.x)_v^{\mathrm{GZ}} = 0$).

The possible $C$'s are given by subgroups generated by elements $(\zeta, 1) \in (\mu_{p^{t-i}} \times \mathbf{Z}/\mathfrak{p}^{t-i})(\overline{F})$. Up to $\mathrm{Gal}(\overline{F}/F)$-conjugacy such $C$'s are classified by the order $p^s$ of $\zeta$ (as $W \simeq W(\overline{\mathbf{F}}_p)$ is absolutely unramified, so it's cyclotomic theory is "as big" as possible), with $1 \le s \le t - i$ (the case $s = 0$ corresponds to the point $\underline{x}$ in the support of $T_m(\underline{x^\sigma})$ which we're not considering). The associated divisor point on $\underline{X}_{/F}$ has residue field $F(s) \overset{\mathrm{def}}{=} F(\zeta_{p^s})$, and it is represented by a $\Gamma_0(N)$-structure $\underline{y}(s)$ over the valuation ring $W(s) \overset{\mathrm{def}}{=} W[\zeta_{p^s}]$. Let $\pi_s = \zeta_{p^s} - 1$, a uniformizer of $W(s)$. Note that $\underline{y}(s) \bmod \pi_s^2$ is *not* the Serre-Tate lift of its closed fiber $\underline{x}_0$ as a "bare" elliptic curve, since $\zeta_{p^s} \not\equiv 1 \bmod \pi_s^2$ (in other words, $\underline{y}(s)$ is a quasi-canonical lifting in the Serre-Tate sense, as is alluded to at the very end of [Gr1]). Thus, under the canonical map $W[\![T_0]\!] \to W(s)$ from the universal deformation ring of the $\Gamma_0(N)$-structure $\underline{x}_0$ (with $T_0 = 0$ corresponding to $\underline{x}$) we have to have that $T_0$ maps to a uniformizer of $W(s)$. Beware that this does not imply in general that natural map $\mathrm{Spec}(W(s)) \to \underline{X}$ over $W$ (sending the closed point to $\underline{x}_0$) is a closed immersion, let alone that the corresponding closure of the closed point $\mathrm{Spec}(F(s))$ in $\underline{X}_{/F}$ is transverse to the closed subscheme

$\underline{x} = \mathrm{Spec}(W)$. The problem is that the complete local ring at $\underline{x}_0$ on $\underline{X}$ might not be the universal deformation ring (as a $\Gamma_0(N)$-structure).

To keep track of what is really happening on the scheme $\underline{X}$, recall from (9.7) that the natural map

$$W[\![T_{\underline{x}}]\!] \simeq \widehat{\mathscr{O}}_{\underline{X},\underline{x}_0} \to W[\![T_0]\!]$$

is $T_{\underline{x}} \mapsto b_x T_0^{u_x} + \dots$ with $\mathrm{ord}_v(b_x) = (x.x)_v^{\mathrm{GZ}} = 0$. Thus, $W[\![T_{\underline{x}}]\!] = \widehat{\mathscr{O}}_{\underline{X},\underline{x}_0} \to W(s)$ arising from $\underline{y}(s)$ sends $T_{\underline{x}}$ to the $u_x$th power of a uniformizer of $W(s)$. Hence, the closure of $y(s) : \mathrm{Spec}(F(s)) \hookrightarrow \underline{X}_{/F}$ in $\underline{X}$ is the order of level $u_x$ in $W(s)$ and $(x.y(s))_v = u_x$ for $1 \le s \le t-i$. We conclude that $\mathfrak{b}$ contributes $u_x(t-i) = u_x \mathrm{ord}_{\overline{\mathfrak{p}}}(\mathfrak{b})$ to $(x.T_m(x^\sigma))_v^{\mathrm{GZ}}$. Consequently, we get

$$(x.T_m(x^\sigma))_v^{\mathrm{GZ}} = u_x \cdot \sum_{\mathfrak{b}} \mathrm{ord}_{\overline{\mathfrak{p}}}(\mathfrak{b}) \tag{10.8}$$

when $v$ lies over $p|m$ which is split in $K$ (with $\mathfrak{p}|p$ the prime of $\mathscr{O}_K$ under $v$). The sum in (10.8) is taken over all integral ideals of norm $m$ in the ideal class of $\mathscr{A}^{-1}$, and will be called $\kappa_{\mathfrak{p}}$ (it clearly is intrinsic to $\mathfrak{p}$ and in general is not divisible by $r_{\mathscr{A}}(m)$). Since $\mathrm{ord}_{\overline{\mathfrak{p}}}(\mathfrak{b}) + \mathrm{ord}_{\mathfrak{p}}(\mathfrak{b}) = t$ for each of the $r_{\mathscr{A}}(m)$ different $\mathfrak{b}$'s, we conclude that $\kappa_{\mathfrak{p}} + \kappa_{\overline{\mathfrak{p}}} = r_{\mathscr{A}}(m)t = r_{\mathscr{A}}(m)\mathrm{ord}_p(m)$.

■

## Appendix A. Elimination of quaternionic sums (by W.R. Mann)

We wish to explicitly compute $\langle c, T_m(d^\sigma)\rangle_p$ as defined in (9.18) with $\theta = \Delta$ and $x \in X(H)$ a Heegner point (with CM by the ring of integers of $K$). Recall that $K$ is the imaginary quadratic field we have fixed from the outset, $D < 0$ is its discriminant, and $H$ is the Hilbert class field of $K$. Also, the level $N$ has prime factors which are split in $K$.

Our approach will build upon the results in §10 and use further arguments with quaternion algebras. The results will be expressed in terms of the arithmetic of $K$, with the answers separated according to whether $p$ is split, inert, or ramified in $K$ (and keep in mind that we allow the discriminant of $K$ to be even). Combining Theorem 8.4, Lemma 10.1, Theorem 10.4 (if $p|N$), and Theorem 10.5 (if $p \nmid N$), we obtain the split case [GZ, Ch III, Prop 9.2]:

**Theorem A.1.** *If $p$ is split in $K$, then $\langle c, T_m(d^\sigma)\rangle_p = -u_x r_{\mathscr{A}}(m) h_K \mathrm{ord}_p(m/N)\log(p)$.*

The appearance of $h_K \log p$ in Theorem A.1 is due to the identity $h_K \log p = \sum_{v|\mathfrak{p}} \log q_v$, where $q_v$ is the size of the residue field at the place $v$ of $H$ and $\mathfrak{p}$ is a choice of either prime over $p$ in $K$.

Now we may assume $p$ is non-split in $K$, so in particular $p \nmid N$. Let $\mathfrak{p}$ be the unique prime of $K$ over $p$. By Theorem 8.4 and Lemma 10.1, we have

$$\langle c, T_m(x^\sigma)\rangle_p = -\sum_{v|\mathfrak{p}} (x.T_m(x^\sigma))_v^{\mathrm{GZ}} \log q_v$$

where the intersection number for the $v$-term is given by Theorem 10.5(1) (resp. Theorem 10.5(2)) when $p$ is inert (resp. ramified) in $K$. Since $\sum_{v|\mathfrak{p}} \log q_v = h_K \log \mathrm{N}\mathfrak{p}$ in the non-split case, with $q_v = p^2$ in the inert case since the principal prime $\mathfrak{p} = p\mathscr{O}_K$ is totally split in $H$ (Principal Ideal Theorem), Theorem 10.5(1) implies

$$\langle c, T_m d^\sigma\rangle_p = -ur_{\mathscr{A}}(m)h_K \mathrm{ord}_p(m)\log p - \log p \cdot \sum_{v|\mathfrak{p}} \sum_{\substack{b \in R_v\mathfrak{a}/\pm 1 \\ \mathrm{N}b = m\mathrm{N}\mathfrak{a}, b_- \neq 0}} (1 + \mathrm{ord}_p(\mathrm{N}(b_-))) \tag{A.1}$$

in the inert case, with $R_v = \mathrm{End}_{W_v/\pi_v}(\underline{x}_v)$. Meanwhile, in the ramified case we get

$$\langle c, T_m d^\sigma\rangle_p = -ur_{\mathscr{A}}(m)h_K \mathrm{ord}_p(m)\log p - \log p \cdot \sum_{v|\mathfrak{p}} f_v \sum_{\substack{b \in R_v\mathfrak{a}/\pm 1 \\ \mathrm{N}b = m\mathrm{N}\mathfrak{a}, b_- \neq 0}} \mathrm{ord}_p(D\,\mathrm{N}(b_-)), \tag{A.2}$$

where $q_v = p^{f_v}$ and $R_v = \mathrm{End}_{W_v/\pi_v}(\underline{x}_v)$.

Our aim is to find an expression for the inner quaternionic sums in (A.1) and (A.2) depending solely on the arithmetic of $K$, and we will see that no small effort is required to combine and manipulate these quaternionic formulae. These local sums are *exactly* the local sums which are analyzed in [GZ, Ch III, §9], and are computed in terms of the arithmetic of $K$ in [GZ, Ch III, Prop 9.7, Prop 9.11]. Since the explanations there are sometimes a bit terse, in order to clarify what is happening and moreover to show that the results work for even $D$, the rest of this appendix is devoted to explaining the analysis of these quaternionic sums and deriving the formulas obtained by Gross-Zagier. The additional burden of treating even discriminants will be a slight nuisance, but will not require any essentially new ideas.

Before getting into the details, let us briefly outline the argument. The sums of interest from Theorem 10.5 involve sums which extend over certain elements inside a quaternionic order isomorphic to the order $\mathrm{End}_{W/\pi}(\underline{x})$. The first task is to find a convenient model for the quaternion algebra in question (this algebra is the unique one over $\mathbf{Q}$ ramified at precisely $p$ and $\infty$, as we saw in Lemma 7.1), paying careful attention to the specification of an embedding of $K$ into this model. We will then have to find a model for the order of interest within this algebra, *but* it turns out that Lemma 7.1 fails to identify this order up to $\mathscr{O}_K$-algebra isomorphism. Specifically, when one takes into account the embedding of $\mathscr{O}_K$ into the order, there are finitely many non-isomorphic orders which satisfy the conditions of Lemma 7.1. We will see that as $\mathrm{Gal}(H/K)$ acts on the places over $p$, it also serves to simply transitively permute the isomorphism classes of these orders (as $\mathscr{O}_K$-algebras). This is precisely what enables us to obtain a formula depending only on the arithmetic of $K$ when we sum over all places $v$ of $H$ over the unique prime $\mathfrak{p}$ of $K$ over $p$ in (A.1) and (A.2).

Though we retain the notation $K$ with the same meaning as throughout this paper, in this appendix we will use $F$ to denote an arbitrary field of characteristic 0 (though only characteristic 2 requires extra care). This will pose no risk of conflict with the use of $F$ to denote the fraction field of $W$ in the main paper, since that fraction field will never show up in this appendix.

To get started, we review some terminology in the theory of quaternion algebras. Recall that a *quaternion algebra* $B$ over a field $F$ is a 4-dimensional central simple $F$-algebra. If $E$ is an extension of $F$, then $B_E \stackrel{\mathrm{def}}{=} E \otimes_F B$ is a quaternion algebra over $E$. An important example of a quaternion algebra is the matrix algebra $M_2(F)$. Another family of examples is provided by the following, which we will use (keep in mind we assume $F$ has characteristic 0):

*Example* A.2. Pick $e, f \in F^\times$. There is a unique quaternion algebra $B$ with basis $1, i, j, ij$ satisfying the requirements that $i^2 = e$, $j^2 = f$, and $ij = -ji$ (so $(ij)^2 = -ef$). This algebra is denoted $\left(\frac{e,f}{F}\right)$.

It is a basic fact that a quaternion algebra $B$ over $F$ is either a division algebra or is a matrix algebra over $F$, and in the latter case the isomorphism $B \simeq M_2(F)$ is unique up to inner automorphism. This latter case is called *split*, and the division algebra case is called *non-split*. It is a subtle algebraic problem to determine whether or not the construction in Example A.2 is split (for a given pair $e, f \in F^\times$). An extension field $E$ of $F$ is called a *splitting field* of $B$ if $B_E$ is split. In general there exists a finite Galois extension $E$ of $F$ which splits $B$. Since the trace and determinant on $M_2(E)$ are invariant under conjugation by $M_2(E)^\times = \mathrm{GL}_2(E)$, if we choose a Galois splitting field $E/F$ for $B$ we can transport the trace and determinant to define $\mathrm{Tr}_\iota : B_E \to E$ and $\mathrm{N}_\iota : B_E \to E$ via an isomorphism $\iota : B_E \simeq M_2(E)$. The $\mathrm{GL}_2(E)$-conjugacy ambiguity in the choice of $\iota$ does not affect these constructions. If we extend scalars on $\iota$ through $\sigma \in \mathrm{Gal}(E/F)$ on source and target, we get a new isomorphism $\iota^\sigma$. Thus, $\mathrm{Tr}_{\iota^\sigma} = \mathrm{Tr}_\iota$ and $\mathrm{N}_{\iota^\sigma} = \mathrm{N}_\iota$, yet also $\mathrm{Tr}_{\iota^\sigma}$ is the extension of scalars on $\mathrm{Tr}_\iota$ by $\sigma$, and similarly for the comparison of $\mathrm{N}_{\iota^\sigma}$ and $\mathrm{N}_\iota$. It follows that $\mathrm{Tr}_\iota$ and $\mathrm{N}_\iota$ are $\mathrm{Gal}(E/F)$-equivariant and independent of $\iota$, so these descend to define the canonical *reduced trace* and *reduced norm*

$$\mathrm{Tr} : B \to F, \ \ \mathrm{N} : B \to F.$$

These are $F$-linear and multiplicative respectively, with $\mathrm{Tr}(bb') = \mathrm{Tr}(b'b)$ for any $b, b' \in B$. For $b \in B$, we define $\overline{b} = b - \mathrm{Tr}(b)$. Extending scalars to a splitting field, $b \mapsto \overline{b}$ becomes $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}$. Using this, one checks that $b\overline{b} = \overline{b}b = \mathrm{N}(b)$ and $b \mapsto \overline{b}$ is an anti-automorphism of $B$.

Note in particular (by Cayley-Hamilton over a splitting field) that $b \in B$ is a root of $X^2 - \mathrm{Tr}(b)X + \mathrm{N}(b)$, so the reduced trace and norm restrict to the usual trace and norm on any quadratic subfield of $B$ over $F$. Moreover, for $b \in B$ outside of $F$, this is the *unique* quadratic polynomial over $F$ satisfied by $b$. It follows that

$F[b]$ is 2-dimensional for any $b \in B - F$, and in Example A.2 we have $\overline{x + yi + zj + wij} = x - yi - zj - wij$ for $x, y, z, w \in F$.

Class field theory leads to an abstract classification:

**Lemma A.3.** *If $F$ is a local field other than $\mathbf{C}$, there is a unique quaternion division algebra over $F$ up to isomorphism. Any quadratic extension of $F$ can be embedded into $B$. If $F$ is a number field, a quaternion algebra $B$ over $F$ is split at all but finitely many places, with the set of non-split places of even size. Any even set of places arises as the set of non-split places of a unique quaternion algebra over $F$. In particular, if $B$ is split at all places of $F$ then $B$ is split.*

Although Lemma A.3 is useful for theoretical purposes, we need to build some concrete quaternion algebras. For example, we will need to describe the unique quaternion division algebra over $\mathbf{Q}_p$ in terms of the construction in Example A.2.

**Definition A.4.** When $F$ is a local field the *Hilbert symbol* $(\cdot, \cdot) : F^\times \times F^\times \to \mu_2(F)$ is defined by the requirement that $(e, f) = 1$ when the quaternion algebra $\left(\frac{e,f}{F}\right)$ is split, and $(e, f) = -1$ when this algebra is a division algebra.

In the case of a number field $F$, if we write $(\cdot, \cdot)_v$ to denote the associated local Hilbert symbol for $F_v$ at a place $v$, then the evenness in Lemma A.3 corresponds to the *product formula*: $\prod_v (e, f)_v = 1$ for any $e, f \in F^\times$, the product taken over all places $v$ of $F$. As one special case, if $(e, f)_v = 1$ for all but possibly one place $v_0$, then $(e, f)_{v_0} = 1$. We will be interested in the case $F = \mathbf{Q}$, and the first problem we will have to solve is that of building quaternion algebras over $\mathbf{Q}$ with a specified even set of non-split places, using the language in Example A.2. Often there will be a particularly problematic place, but if we can check that a quaternion algebra has the desired splitting/non-splitting away from one place, then the behavior at the missing place is forced by the product formula.

Returning to the original problem of interest, recall that our prime $p$ is either inert or ramified in $K$, and we want to construct a model for the quaternion division algebra $B$ in Lemma 7.1: this is the unique such algebra over $\mathbf{Q}$ which is non-split at precisely $p$ and $\infty$. Since the sums in Theorem 10.5 involve elements of (an order in) $B$ which are identified with respect to right multiplication by a nonzero ideal $\mathfrak{a}$ of $\mathscr{O}_K$, a model for $B$ needs to encode an embedding of $K$. We want to make such a model in the simplest manner possible. It makes sense to try to build a model $\left(\frac{e,f}{\mathbf{Q}}\right)$ with $e = i^2 = D$ since we want to keep track of how $K = \mathbf{Q}(\sqrt{D})$ sits in $B$. Thus, we want to find an $f \in \mathbf{Q}^\times$ for which the algebra $\left(\frac{D,f}{\mathbf{Q}}\right)$ is a division algebra non-split at exactly $p$ and $\infty$. That is, we want $(D, f)_v = -1$ exactly when $v = p, \infty$. To find a suitable $f$, we need to understand how to compute local Hilbert symbols over places of $\mathbf{Q}$. This is done via:

**Lemma A.5.** *Let $F$ be a local field. The Hilbert symbol $(\cdot, \cdot) : F^\times \times F^\times \to \mu_2(F)$ factors through $F^\times/(F^\times)^2 \times F^\times/(F^\times)^2$, yielding a nondegenerate symmetric bilinear form. If $e \in F^\times$ is not a square, then $(e, f) = 1$ if and only if $f \in \mathrm{N}_{F(\sqrt{e})/F}(F(\sqrt{e})^\times)$.*

Before we return to the problem of finding a model for $B$, for later convenience we recall how Lemma A.5 helps us to compute the local symbols for quaternion algebras over $\mathbf{Q}$.

*Example* A.6. We wish to compute $(a, b)_v$ for $a, b \in \mathbf{Q}^\times$ and $v$ a place of $\mathbf{Q}$. If $v = \infty$ then $\mathbf{Q}_v = \mathbf{R}$, so by Lemma A.5 we see that $(a, b)_\infty = -1$ exactly when $a$ and $b$ are both negative. If $v = \ell$ is an odd prime, the situation is only slightly more difficult. The classes in $\mathbf{Q}_\ell^\times/(\mathbf{Q}_\ell^\times)^2$ are represented by $\{1, \varepsilon, \ell, \varepsilon\ell\}$ where $\varepsilon \in \mathbf{Z}_\ell^\times$ is not a quadratic residue mod $\ell$ and hence is not a square in $\mathbf{Z}_\ell^\times$. Since $\mathbf{Q}_\ell(\sqrt{\varepsilon})$ is the quadratic unramified extension of $\mathbf{Q}_\ell$, it has norm group generated by $\mathbf{Z}_\ell^\times$ and $\ell^2$. Thus, $(\varepsilon, \varepsilon)_\ell = 1$ and $(\varepsilon, \ell)_\ell = -1$. Also, since $-\ell$ is a norm from $\mathbf{Q}_\ell(\sqrt{\ell})$, we have $(\ell, -\ell)_\ell = 1$. Thus, $(\ell, \ell)_\ell = (\ell, -1)_\ell$. By Lemma A.5, this is 1 if and only if $-1$ is a square mod $\ell$, so $(\ell, \ell)_\ell = (-1)^{(\ell-1)/2}$. This generates the other possible pairings by bilinearity (e.g., $(u, v)_\ell = 1$ for $u, v \in \mathbf{Z}_\ell^\times$).

For $v = 2$, there are more cases because $\mathbf{Q}_2^\times/(\mathbf{Q}_2^\times)^2$ has order 8 with generators $-1, 2, 3$. Rather than delve into tedious examniantion of cases, we use the product formula to compute these symbols. Consider the quaternion algebras over $\mathbf{Q}$ isomorphic to $\left(\frac{a,b}{\mathbf{Q}}\right)$ for some $a, b$ chosen from $\{-1, 2, 3\}$, and recall that an even number of places will be non-split. Since these choices of $a$ and $b$ are units for all finite primes $\ell > 3$, our observations above show that $(a,b)_\ell = 1$ for all such $\ell$. All that is left are the places $2, 3, \infty$, and each algebra must be non-split at either none or two of these three places. For the sake of completeness, we note that an examination of the local symbols at 3 and $\infty$ yields: $(-1,-1)_2 = (-1,3)_2 = (2,3)_2 = (3,3)_2 = -1$, $(-1,2)_2 = (2,2)_2 = 1$.

Now recall that we wish to find $f \in \mathbf{Q}^\times$ so that $\left(\frac{D,f}{\mathbf{Q}}\right)$ is non-split at exactly $p$ and $\infty$. There will be infinitely many $f$ that work, but we want the simplest possible choice. We must require $f < 0$ to force non-splitting at $\infty$. We will have to treat separately the cases when $p$ is inert in $K$ and when $p$ is ramified in $K$. First consider the inert case. It is necessary that $\mathrm{ord}_p(f)$ be odd. To see this, first note that the completion $K_{\mathfrak{p}} = \mathbf{Q}_p \otimes_{\mathbf{Q}} K$ is an unramified quadratic extension inside of $B_p$, so its norm group in $\mathbf{Q}_p^\times$ consists of elements of even order. Thus, if $\mathrm{ord}_p(f)$ is even then $-f = \mathrm{N}_{K_{\mathfrak{p}}/\mathbf{Q}_p}(c)$ for some $c \in K_{\mathfrak{p}}^\times$, so via the decomposition $B = K \oplus Kj$ with $j^2 = f$ (cf. the discussion following Lemma 7.1) we'd get

$$\mathrm{N}(c + j) = \mathrm{N}(c) + \mathrm{N}(j) = \mathrm{N}_{K_{\mathfrak{p}}/\mathbf{Q}_p}(c) + f = 0,$$

a contradiction since $B_p$ is non-split and $c + j \neq 0$. One might hope that $\left(\frac{D,-p}{\mathbf{Q}}\right)$ works, as it is clearly nonsplit at $p$ and $\infty$, but in general this algebra can be nonsplit at some primes dividing $D$ (and possibly 2). For example, if $\ell | D$ is an odd prime then

$$(D,-p)_\ell = (\ell,-p)_\ell (D/\ell,-p)_\ell = (\ell,-p)_\ell = \left(\frac{-p}{\ell}\right), \tag{A.3}$$

and this is generally not equal to 1.

Thus, the simplest possible choice for $f$ in the inert case is $f = -pq$ for some *auxiliary* prime $q \nmid Dp$. For odd primes $\ell \nmid Dq$, we already know $(D,-pq)_\ell = 1$ by Example A.6, since $D$ and $-pq$ are units modulo $\ell$. Thus, we're left to handle a finite number of primes, at which we will get congruence conditions on $q$ (for which there will be infinitely many solutions, by Dirichlet's theorem). For odd $\ell | D$, computing as in (A.3) gives the condition $\left(\frac{-pq}{\ell}\right) = 1$, which is a congruence condition (with $(\ell - 1)/2$ solutions) for $q$ modulo each such $\ell$. It remains to consider the places 2, $p$, and $q$ (possibly $p = 2$ or $q = 2$). If $p = 2$ then $D$ and $q$ are odd, so the requirement $(D,-pq)_p = -1$ can be satisfied by a mod 8 condition on $q$. This makes $\left(\frac{D,-pq}{\mathbf{Q}}\right)$ nonsplit at $p$ and $\infty$, and split at all odd primes except for possibly $q$, and splitting is then forced at $q$ by the product formula. Now suppose $p$ is odd. In this case, since $p$ is inert in $K$, $(D,-pq)_p = (D,-q)_p(D,p)_p = (D,p)_p = \left(\frac{D}{p}\right) = -1$, so $\left(\frac{D,-pq}{\mathbf{Q}}\right)$ is nonsplit at $p$ and $\infty$. By the product formula, it remains to verify splitting at either 2 or $q$. This renders the case $q = 2$ trivial, so we may assume $q$ is odd. With $pq$ now odd, we will force splitting at 2 instead, by means of the congruence $-pq \equiv 1 \bmod 8$, which forces $-pq \in \mathbf{Q}_2^\times$ to be a square (so $(D,-pq)_2 = 1$). This settles the construction of our quaternion algebra model in the inert case. Observe that $q$ is automatically split in $K$. Indeed, for odd $q$ we have $1 = (D,-pq)_q = (D,-p)_q(D,q)_q = (D,q)_q = \left(\frac{D}{q}\right)$, while for $q = 2$ we have $(D,-2p)_2 = 1$, so $D \equiv 1 \bmod 8$ (recall that odd fundamental discriminants are $\equiv 1 \bmod 4$), so 2 splits in $K$.

*Remark* A.7. In all inert cases, $q$ satisfies the condition that for all primes $\ell | D$, $-pq \in \mathbf{Z}_\ell^\times$ is a square. Also, requiring $-pq \equiv 1 \bmod D$ is consistent with the requirements on $q$ in inert cases. In ramified cases, the congruence $-q \equiv 1 \bmod D/p$ (which involves no condition at $p$, unless $p = 2$) may always be imposed without inconsistency.

We can similarly treat the case when $p$ is ramified in $K$; i.e., $p | D$. The simplest choice of $f$ we might try is $f = -q$ for some prime $q \nmid D$. If $p$ is odd, then as in (A.3) we get $(D,-q)_p = \left(\frac{-q}{p}\right)$, so requiring this to be $-1$ is a congruence condition on $q \bmod p$. If $p = 2$, we similarly find that $(D,-q)_2 = -1$ follows from a congruence condition on $q \bmod 8$. For example, if $8 \nmid D$ then $q \equiv 1 \bmod 4$ is necessary and sufficient, while if $8 | D$ then $q \equiv 3 \bmod 8$ is sufficient. For odd $\ell \nmid Dq$, just as above it is clear that $(D,-q)_\ell = 1$. For odd $\ell | D$, we have $(D,-q)_\ell = \left(\frac{-q}{\ell}\right)$, and we want this to be 1, imposing a congruence on $q \bmod \ell$. We're now left with the places 2 and $q$, which means that we are done if $p = 2$ (as we've considered it already) or if $q = 2$ (product formula). Otherwise $q$ is odd and all congruence conditions have been at odd primes, so we can also require $q \equiv -1 \bmod 8$ to force $(D,-q)_2 = 1$. The product formula now ensures $(D,-q)_q = 1$, so in fact $q$ again splits in $K$, just as in the inert case. Even if $q = 2$ (so $D$ is odd and $p|D$ is odd), from $(D,-2p)_2 = 1$ we get $D \equiv 1 \bmod 8$, so $q = 2$ is split in $K$. It is important later that our choice of $q$ is automatically split in $K$ (for $p$ of either inert or ramified type in $K$).

Having found a model for the quaternion algebra $B$ over $\mathbf{Q}$, we need to get a handle on certain orders $R$ in $B$, as this is what intervenes in Theorem 10.5. We will begin our discussion of this problem by reviewing some basic properties of orders in quaternion algebras which will be used quite a lot in what follows. Let $F$ be the fraction field of a Dedekind domain $A$ of characteristic 0; we are primarily interested in the case when $A$ is the ring of integers of a local field or number field. Let $B$ be a quaternion algebra over $F$.

**Definition A.8.** A *lattice* in $B$ is a finite $A$-submodule $L$ which spans $B$ over $F$ (so $L$ is locally free of rank 4 over $A$). An *order* in $B$ is an $A$-subalgebra $R$ which is also a lattice. We say that $R$ is a *maximal* order if it is not strictly contained in any larger order, and we say that $R$ is an *Eichler order* if it is the intersection of two maximal orders.

Note that if $R$ is an order and $r \in R$, then $A[r]$ is a finite $A$-module, so $r$ is integral over $A$. Since $A[r]$ is locally free of rank 2 when $r \notin R \cap F = A$, the minimal polynomial $X^2 - \mathrm{Tr}(r)X + \mathrm{N}(r)$ of $r$ over $F$ must have $A$ coefficients when $r \notin A$. Since Tr and N are $A$-valued on $A$, we see that Tr and N are $A$-valued on any order $R$ in $B$.

**Lemma A.9.** *Suppose $A$ is a complete discrete valuation ring, with maximal ideal $\mathfrak{m}$ and uniformizer $\pi$. Let $B$ be a quaternion algebra over $F$.*

(1) *If $B$ is nonsplit, then it contains a unique maximal order and this contains all other orders.*
(2) *If $B$ is split, then all maximal orders in $B$ are conjugate under $B^\times$ and every order lies in one. In $M_2(F)$ a maximal order is $M_2(A)$, and any Eichler order is conjugate to $\left(\begin{smallmatrix} A & A \\ \pi^n & A \end{smallmatrix}\right)$ for a unique $n \geq 0$. Any two Eichler orders which are abstractly isomorphic as rings are conjugate in $B$.*

*Remark* A.10. The order $\left(\begin{smallmatrix} A & A \\ \mathfrak{m}^n & A \end{smallmatrix}\right)$ in (2) is the intersection of the maximal orders $M_2(A)$ and $\gamma_n M_2(A)\gamma_n^{-1} = \left(\begin{smallmatrix} A & \mathfrak{m}^{-n} \\ \mathfrak{m}^n & A \end{smallmatrix}\right)$ (with $\gamma_n = \left(\begin{smallmatrix} 0 & 1 \\ \pi^n & 0 \end{smallmatrix}\right)$). This example is called a *standard* Eichler order.

*Proof.* For the first part, it suffices to show that the set $R$ of elements of $B$ integral over $R$ is an order. That is, we must show $R$ is finite as an $A$-module and is a subring of $B$. Note that $R$ is stable under the involution $b \mapsto \overline{b}$. The key to the subring property is that if $b \in B$ has $\mathrm{N}(b) \in A$, then $\mathrm{Tr}(b) \in A$. Indeed, $F[b]$ is a field on which the reduced norm and trace agree with the usual norm and trace (relative to $F$), and by completeness of $A$ we know that the valuation ring of $F[b]$ is characterized by having integral norm. Thus, to show that $R$ is stable under multiplication we just need that if $x, y \in R$ then $\mathrm{N}(xy) \in A$. But $\mathrm{N}(xy) = \mathrm{N}(x)\mathrm{N}(y)$. Meanwhile, for addition (an issue because non-commutativity does not make it evident that a sum of intergal elements is integral), we note that if $x, y \in R$ then

$$\mathrm{N}(x+y) = (x+y)(\overline{x}+\overline{y}) = \mathrm{N}(x) + \mathrm{N}(y) + \mathrm{Tr}(x\overline{y}).$$

But this final reduced trace term lies in $A$ because $x\overline{y} \in R$. Hence, $R$ is a subring of $B$. In particular, $R$ is an $A$-submodule since $A \subseteq R$.

To show that $R$ is $A$-finite, we may pick a model for $B$ as in Example A.2, and may assume $i^2 = e, j^2 = f \in A$. Thus, $i, j \in R$, so $ij \in R$. For any $x \in R$, we have $x, xi, xj, xij \in R$. Taking reduced traces of all four of these elements and writing $x = c + c_1 i + c_2 j + c_3 ij$ for $c, c_1, c_2, c_3 \in F$, we get

$$x \in \frac{1}{2}A + \frac{1}{2e}Ai + \frac{1}{2f}Aj + \frac{1}{2ef}Aij.$$

Thus, $R$ lies inside of a finite $A$-module and hence is $A$-finite.

Now we turn to the split case, so we may assume $B = M_2(F)$. To keep the picture clear, we suppose $B = \mathrm{End}_F(V)$ for a 2-dimensional $F$-vector space $V$ on which we do not impose a basis. If $R$ is any order in $B$, then for a lattice $L$ in $V$ clearly $RL$ is another finite $A$-submodule spanning $V$ over $F$, so $N = RL$ is an $R$-stable lattice. Hence, $R \subseteq \mathrm{End}_A(N)$. Since any two lattices are conjugate to each other, the assertions concerning maximal orders come down to the claim that if $\mathrm{End}_A(N_0) \subseteq \mathrm{End}_A(N_1)$ for two lattices $N_0, N_1$ in $V$, then this inclusion is an equality. We may scale $N_1$ so $N_1 \subseteq N_0$ and $N_1$ is not contained in $\mathfrak{m}N_0$, so

$N_0/N_1 \simeq A/\mathfrak{m}^n$ for some $n \geq 0$. We can then pick bases so that $N_0 = Ae \oplus Ae'$ and $N_1 = Ae \oplus A\mathfrak{m}^n e'$. If $n > 0$, then it is obvious that $\mathrm{End}_A(N_0)$ does not lie in $\mathrm{End}_A(N_1)$ inside of $\mathrm{End}_F(V) = B$.

Now consider Eichler orders. It suffices to focus attention on those Eichler orders $R$ which are not maximal orders in $B = F \otimes_A R$. Using a suitable conjugation by $B^\times$, given any two non-maximal Eichler orders we may suppose they have the form $R \cap R_1$ and $R \cap R_2$ for maximal orders $R, R_1, R_2$ which we may assume to be distinct. Arguing as above, $R = \mathrm{End}_A(N)$ for some lattice $N$ in $V$ and $R_j = \mathrm{End}_A(N_j)$ for a sublattice $N_j \subseteq N$ with $N/N_j \simeq A/\mathfrak{m}^{n_j}$ for some $n_j > 0$. Relative to a basis $\{e_j, e_j'\}$ of $N$ for which $\{e_j, \pi^{n_j} e_j'\}$ is a basis of $N_j$, we get an identification of $R \cap R_j$ with $R_j \stackrel{\mathrm{def}}{=} \left(\begin{smallmatrix} A & A \\ \mathfrak{m}^{n_j} & A \end{smallmatrix}\right)$ inside of $R \simeq M_2(A)$ (isomorphism using the basis $\{e_j, e_j'\}$), yielding the desired description of Eichler orders. In fact, $n_j$ is intrinsic to $R_j$ because on the quaternion algebra $B = F \otimes_A R_j$ we can apply the involution $b \mapsto \overline{b}$ to form $\overline{R}_j$, and by inspection $R/(R_j \cap \overline{R}_j) \simeq A/\mathfrak{m}^{n_j}$ as an $A$-module.

Thus, when comparing two abstractly isomorphic Eichler orders $R \cap R_j$ as considered above, necessarily $n_1 = n_2 = n$. But we can then use the two bases on $N$ adapted to the $N_j$'s to define an element in $R^\times = \mathrm{Aut}_A(N)$ which carries $N_1$ into $N_2$ and hence conjugates $R \cap R_1$ over to $R \cap R_2$.

■

*Remark* A.11. One important consequence of the preceding proof is that in the split case, for each non-maximal Eichler order $R$ in $B$, there is a *unique* pair of maximal orders $S$ and $S'$ with $S \cap S' = R$. Indeed, by conjugation we may consider the case $B = M_2(F)$ and $R = \begin{pmatrix} A & A \\ \mathfrak{m}^n & A \end{pmatrix}$ a standard Eichler order as in Remark A.10. Thus, a maximal order $S = \mathrm{End}_A(N)$ contains $R$ if and only if $R(N) \subseteq N$. Since we can scale to get $N \subseteq A^2$ with $A^2/N$ a cyclic module (necessarily $A/\mathfrak{m}^n$), it is easy to check (using $A$-module generators of the explicit standard Eichler order $R$) that $N$ must be the span of $e_1$ and $\pi^j e_2$ for some $0 \leq j \leq n$. Only the extreme options $j = 0, n$ provide a pair whose intersection is the standard $R$.

In order to build a model of the global order in Lemma 7.1, there is one further aspect of the local theory of orders in quaternion algebras which we need to address: discriminants. The trace form $(x, y) \mapsto \mathrm{Tr}(xy)$ is a symmetric bilinear form on $B$, and its restriction to an order $R$ is a symmetric $A$-valued bilinear form on $R$. Thus, we can define the *discriminant* $\mathrm{disc}(R)$ of $R$ to be the discriminant of this bilinear form (i.e., the ideal generated by the determinant of its values on pairs from an $A$-basis of $R$; note that $R$ is $A$-free since we are in the local case). Since Tr is invariant under $B^\times$-conjugation, conjugate orders in $B$ have the same discriminant ideal. Also, since the trace form is non-degenerate (as can be checked in the split case upon extending scalars), $\mathrm{disc}(R) \neq 0$. Formation of the discriminant ideal commutes with faithfully flat base change $A \to A'$ to another (complete) discrete valuation ring. If $R' \subseteq R$ is a suborder with the finite-length $A$-module $R/R'$ of length $n$, then $\mathrm{disc}(R') = \mathfrak{m}^{2n}\mathrm{disc}(R)$. Since every order lies inside of a maximal order (all of which are conjugate), we see that discriminants of orders in $B$ are off by a square factor from the common discriminant of the maximal orders in $B$.

*Example* A.12. If $B = M_2(F)$ and $R$ is the Eichler order $\left(\begin{smallmatrix} A & A \\ \mathfrak{m}^n & A \end{smallmatrix}\right)$, then $\mathrm{disc}(R) = \mathfrak{m}^{2n}$. In particular, in the split case $\mathrm{disc}(R) = A$ if and only if $R$ is a maximal order, and all orders have square discriminant.

In general, $B$ has an unramified splitting field $F'/F$, so if $R$ is a maximal order in $B$ and $A'$ is the valuation ring of $F'$ then $R' = A' \otimes_A R$ is an order in the split $F' \otimes_F B$. It follows that $\mathrm{disc}(R)A' = \mathrm{disc}(R')$ is a square, hence $\mathrm{disc}(R)$ is a square since $A \to A'$ is unramified. This motivates us to define the *reduced discriminant* $\mathrm{Disc}(R)$ of an order $R$ in $B$ to be the ideal of $A$ whose square is $\mathrm{disc}(R)$.

*Example* A.13. If $R' \subseteq R$ is an inclusion of orders in $B$ and $R/R'$ has $A$-length $n$, then $\mathrm{Disc}(R') = \mathfrak{m}^n\mathrm{Disc}(R)$. For example, if $R_n$ is the Eichler order $\left(\begin{smallmatrix} A & A \\ \mathfrak{m}^n & A \end{smallmatrix}\right)$ from Remark A.10, then $\mathrm{Disc}(R_n) = \mathfrak{m}^n$. This clarifies the uniqueness of $n$ in Lemma A.9(2).

We can also compute $\mathrm{Disc}(R)$ which $R$ is maximal and $B$ is non-split. Since there is only one such $B$ over $F$ up to isomorphism, and all $R$'s are conjugate in $B$, it suffices to compute in a single example (over $F$). Let $F'$ be the unramified quadratic extension of $F$, with valuation ring $A'$. Since $\pi$ is not a norm from $F'$, by Lemma A.5 it follows that $B \stackrel{\mathrm{def}}{=} F' \oplus F'j$ with $j^2 = \pi$ is a non-split quaternion algebra over $F$. One can then check that $R = A' \oplus A'j$ is the set of $A$-integral elements, so this is the maximal order. Writing down the matrix for the Tr pairing, we get $\mathrm{disc}(R) = \mathrm{disc}(A'/A)^2(j^2)^2 = \mathfrak{m}^2$, so $\mathrm{Disc}(R) = \mathfrak{m}$. Thus, in the non-split case an order $R$ is maximal if and only if $\mathrm{Disc}(R) = \mathfrak{m}$.

The non-triviality of reduced discriminants of maximal orders in the non-split case is a reason that non-split quaternion algebras in the local case are referred to as "ramified".

Now we are in position to globalize to the case of orders in quaternion algebras over the fraction field $F$ of a Dedekind domain $A$ (the case of a number field, especially $F = \mathbf{Q}$, is the one of most importance for us). Fix a quaternion algebra $B$ over $F$ and an order $R$ in $B$ (of which there are clearly many, for example by intersecting $B$ with an order in a split extension $B_E$). We can define $\mathrm{disc}(R)$ as in the local case to be the (nonzero) discriminant ideal of the symmetric bilinear nondegenerate Tr pairing. Since $R$ is merely locally free as an $A$-module, and probably not free, this discriminant ideal is constructed by localizing on $A$, just as in the definition of the discriminant of a symmetric bilinear form on any locally free module of finite rank over a commutative ring. Note that $\mathrm{disc}(R) = \prod_v(\mathrm{disc}(R_v) \cap A)$, with $\mathrm{disc}(R_v) = A_v$ for all but finitely many $v$. By passing to the local case, we see that $\mathrm{disc}(R)$ is a square, so we may define the *reduced discriminant* ideal $\mathrm{Disc}(R)$ in $A$. This ideal is the unit ideal at all but finitely many places, so the preceding discriminant calculations show that $R_v$ is a maximal order and $B_v$ is split for all but finitely many maximal ideals $v$ of $A$. In contrast with the local case, in the global case it is no longer true that maximal orders have to be conjugate.

We can uniquely construct orders by specifying local data. This will be an essential ingredient in our construction of models for global orders, so let us briefly summarize how this goes. As with lattices in any finite-dimensional $F$-vector space, if $R$ and $R'$ are two orders in a quaternion algebra $B$ over $F$, then $R_v = R'_v$ inside of $B_v = F_v \otimes_F B$ for all but finitely many maximal ideals $v$ of $A$. Conversely, if we pick an order $S_v$ in $B_v$ for all $v$ with $S_v = R_v$ for all but finitely many $v$, then there exists a unique order $S$ in $B$ with $A_v \otimes_A S = S_v$ inside of $B_v$ for all $v$. This is an easy application of weak approximation for Dedekind domains. In particular, an order $R$ in $B$ is maximal if and only if $R_v$ is maximal in $B_v$ for all $v$, and likewise $R$ is an Eichler order (i.e,, an intersection of two maximal orders) if and only if $R_v$ is an Eichler order for all $v$.

Note that when $R$ is maximal, $\mathrm{Disc}(R)$ is the product of the maximal ideals of $A$ at precisely those places where $B$ is non-split. If $R$ is merely an Eichler order, then $\mathrm{Disc}(R)$ is the product of the maximal ideals at the non-split places of $B$ (the ramified primes of $B$, for which $B_v$ has only one maximal order) and powers of maximal ideals at the split places where $R$ is a non-maximal Eichler order.

*Example* A.14. The order $R$ in Lemma 7.1 is an Eichler order with reduced discriminant $Np$. Indeed, the lemma assures us that $R$ is maximal at the non-split place $p$, and the local description at all $\ell \neq p$ (where $B$ is split) is exactly a standard Eichler order in $M_2(\mathbf{Z}_\ell)$ whose index is the $\ell$-part of $N$.

With Example A.14 in hand, specialize to the case $A = \mathbf{Z}$. We seek a *potential* model for the order $R = \mathrm{End}_{W/\pi}(\underline{x})$ inside of $B$ which arises in Theorem 10.5. Inside of our concrete model $\left(\frac{D,?}{\mathbf{Q}}\right)$ for $B$ (in the inert and ramified cases separately), we want to find an Eichler order containing $\mathscr{O}_K$ and having reduced discriminant $Np$. We will find such an order, but we will not know it to be ($\mathscr{O}_K$-)isomorphic, let alone conjugate, to $\mathrm{End}_{W/\pi}(\underline{x})$. This problem will be overcome by the fact that in (A.1) and (A.2) we are summing over all places of $H$ over $p$. We carry out the construction in two separate cases: $p$ inert in $K$ and $p$ ramified in $K$.

When $p$ is inert in $K$, recall that we found an isomorphism $B \simeq \left(\frac{D,-pq}{\mathbf{Q}}\right)$ where $q \nmid pD$ is a prime satisfying certain congruence conditions at the primes $\ell|D$ and possibly at 8 as well. Clearly $R' = \mathscr{O}_K \oplus \mathscr{O}_K j$ is an order, and we can compute its reduced discriminant to be $Dpq$. Thus, $R'$ is maximal at $p$ and at all primes not dividing $Dq$. Since a maximal order of $B$ has reduced discriminant $p$ (because $B$ is non-split at precisely $p$ and $\infty$), it follows that $R'$ has index $Dq$ inside of any maximal order containing $R'$. In particular, an order containing $R'$ with index $Dq$ *must* be a maximal order of $B$.

To examine how one might find a maximal order in $B$ containing $R'$, we first will find local models for maximal orders at all primes $\ell|Dq$ (exactly the places where $R'$ is not maximal), and then we will globalize as in the discussion preceding Example A.14. First let $\ell$ be a prime dividing $D$ (so $\ell$ ramifies in $K$), and consider the split algebra $B_\ell = M_2(\mathbf{Q}_\ell)$. By Remark A.7, $-pq$ is a unit square in $\mathbf{Z}_\ell$, so $-pq$ is a unit norm from the ramified quadratic extension $\mathscr{O}_{K,\ell}$. Pick $X_\ell \in \mathscr{O}_{K,\ell}$ with $\mathrm{N}(X_\ell) = -pq = j^2$. The element $X_\ell - j \in B_\ell$ has nonzero reduced trace but vanishing reduced norm, so it is a zero divisor in $B_\ell$ which generates a left $B_\ell$-module of dimension 2 over $\mathbf{Q}_\ell$. This module must have $X_\ell - j$ and $\omega(X_\ell - j)$ as a $\mathbf{Z}_\ell$-basis, where $\mathscr{O}_K = \mathbf{Z}[\omega]$ (so $\mathscr{O}_{K,\ell} = \mathbf{Z}_\ell[\omega]$). We may take $\omega = (1+i)/2$ when $D$ is odd and $\omega = i/2$ when $D$ is even, where $i^2 = D$. The natural map from $B_\ell$ to $\mathrm{End}_{\mathbf{Q}_\ell}(B_\ell(X_\ell - j)) \simeq M_2(\mathbf{Q}_\ell)$ gives a concrete splitting, where we use the basis $X_\ell - j$, $\omega(X_\ell - j)$. Under this isomorphism, $R'_\ell = \mathscr{O}_{K,\ell} \oplus \mathscr{O}_{K,\ell} j$ clearly maps into $M_2(\mathbf{Z}_\ell)$ (just compute action of $\mathbf{Z}_\ell$-generators of $R'_\ell$ on the $\mathbf{Q}_\ell$-basis $\{X_\ell - j, \omega(X_\ell - j)\}$ of $B_\ell(X_\ell - j)$).

Since $(X_\ell + j)(X_\ell - j) = 0$, we compute (using the identities $jX_\ell = \overline{X}_\ell j$ and $\mathrm{N}(X_\ell) = j^2$):

$$(X_\ell + j)\omega(X_\ell - j) = (X_\ell + j)\frac{i}{2}(X_\ell - j) = \frac{i}{2}(X_\ell - j)^2 = \frac{i}{2}(2X_\ell)(X_\ell - j) = (2X_\ell)(\frac{i}{2}(X_\ell - j)).$$

For odd $D$ it follows that $\frac{i}{D}(X_\ell + j)$ lies in $M_2(\mathbf{Z}_\ell)$ and has additive order $D_\ell$ in $B_\ell/R'_\ell$ (where $D_\ell$ is the $\ell$-part of $D$). Thus, the maximal order $R_\ell = M_2(\mathbf{Z}_\ell)$ is generated by $\frac{i}{D}(X_\ell + j)$ and $R'_\ell$ (since $R'_\ell$ has index $(Dq)_\ell = D_\ell$ in a maximal order). Similarly, if $2|D$ then $\frac{i}{D}(X_\ell + j)$ and $\frac{1}{2}(X_\ell + j)$ act as elements of $M_2(\mathbf{Z}_\ell)$ and additively generate a subgroup of order $D_\ell$ in $B_\ell/R'_\ell$. Thus, these two elements along with $R'_\ell$ generate $R_\ell$ when $D$ is even, so the maximal order $R_\ell$ is always spanned over $\mathbf{Z}_\ell$ by $1, \omega, (i/D)(X_\ell + j)$, and $e\cdot(X_\ell + j)$, where $e = 1$ or $e = 1/2$, according as whether $2 \nmid D$ or $2|D$.

A more uniform and succinct way to describe the maximal order $R_\ell \subseteq B_\ell$ at $\ell|D$ is:

$$\{\alpha + \beta j \mid \alpha, \beta \in \mathfrak{d}_\ell^{-1},\ \alpha - X_\ell\beta \in \mathscr{O}_{K,\ell}\}, \tag{A.4}$$

where $\mathfrak{d}_\ell^{-1}$ is the inverse different of $\mathscr{O}_{K,\ell}$. To see that the lattice (A.4) is $R_\ell$, one first checks directly that it contains the $\mathbf{Z}_\ell$-generators $1, \omega, (i/D)(X_\ell + j), e\cdot(X_\ell + j)$ of $R_\ell$, so the problem is to show the reverse containment (direct verification that (A.4) is an order seems quite painful). For any pair $(\alpha, \beta)$ satisfying the criteria in (A.4), we have $\alpha + \beta j = (\alpha - X_\ell\beta) + \beta(X_\ell + j) \in \mathscr{O}_{K,\ell} + \mathfrak{d}_\ell^{-1}(X_\ell + j)$, so we just need $\mathfrak{d}_\ell^{-1}(X_\ell + j)$ to lie in $R_\ell$. Since $R_\ell$ contains $\mathscr{O}_{K,\ell}$ and is an order, it suffices to pick a generator $\gamma$ of the fractional ideal $\mathfrak{d}_\ell^{-1}$ and to show $\gamma(X_\ell + j)$ lies in $R_\ell$. For odd $D$ or $\ell \neq 2$, we can take $\gamma = i/D$. For even $D$ and $\ell = 2$, we use the fact that $i/D$ and $1/2$ generate $\mathfrak{d}_2^{-1}$ as a $\mathbf{Z}_2$-module in such cases (ultimately because $\mathfrak{d}_2$ is generated by $i = \sqrt{D}$ and $D/2$ over $\mathbf{Z}_2$), by treating separately the cases $D \equiv 4 \bmod 8$ and $8|D$.

The importance of the description (A.4) is that it provides a description which only depends on $X_\ell \bmod \mathfrak{d}_\ell$. To make this point clearer, first observe (for $\ell|D$, being careful if $\ell = 2$) that changing an element of $\mathscr{O}_{K,\ell}$ modulo $\mathfrak{d}_\ell$ does not change its norm (to $\mathbf{Z}_\ell$) modulo $D_\ell$. Thus, even though we chose $X_\ell$ above to satisfy $\mathrm{N}(X_\ell) = -pq$ so as to carry out the preceding calculations, if we replace $X_\ell$ with some $X'_\ell \in \mathscr{O}_{K,\ell}$ which is the same mod $\mathfrak{d}_\ell$, then $\mathrm{N}(X'_\ell) \equiv -pq \bmod D_\ell$ and using $X'_\ell$ instead of $X_\ell$ in (A.4) yields the same lattice (i.e., the maximal order $R_\ell \subseteq B_\ell$ we constructed at $\ell$). In fact, the *only* thing which matters about $X_\ell \in \mathscr{O}_{K,\ell}$ is that it satisfies $\mathrm{N}(X_\ell) \equiv -pq \bmod D_\ell$, since we claim that any such $X_\ell$ is congruent mod $\mathfrak{d}_\ell$ to an element of $\mathscr{O}_{K,\ell}$ with norm equal to $-pq$. Indeed, since $-pq$ is a unit norm at $\ell$, we really just have to show that if $u \in \mathscr{O}_{K,\ell}^\times$ satisfies $\mathrm{N}(u) \equiv 1 \bmod D_\ell$ then $u \equiv u' \bmod \mathfrak{d}_\ell$ with $\mathrm{N}(u') = 1$. The case of odd $\ell$ (resp. $\ell = 2$ and $D_2 = 8$) is easy since $1 + \ell\mathbf{Z}_\ell$ (resp. $1 + 8\mathbf{Z}_2$) consists of squares, and the case $\ell = 2$ and $D_2 = 4$ (so $D \equiv 4 \bmod 8$) requires checking additionally that some (and thus every) element congruent to 5 mod 8 is a norm, and we see that $\mathrm{N}(1+i) = 1 - D \equiv 5 \bmod 8$. From now on, for each $\ell|D$, rather than require $X_\ell$ to have exact norm $-pq$, we merely require it to represent an element in $\mathscr{O}_{K,\ell}/\mathfrak{d}_\ell$ whose norm in $\mathbf{Z}_\ell/D_\ell$ is $-pq$.

With the construction of explicit maximal orders at all $\ell|D$ settled, we now turn to the easier task of finding a maximal order at $q$. This is easier because $q = \mathfrak{q}\overline{\mathfrak{q}}$ splits in $K$. It is simplest to check directly that if $\alpha \in \mathscr{O}_K$ and $\beta \in \mathfrak{q}^{-1}$, then the resulting elements $\alpha + \beta j$ form an order $R^{(q)}$ in $B$ which is maximal at $q$. This set is stable under multiplication because if $\beta, \beta' \in \mathfrak{q}^{-1}$ then $(\beta j)(\beta' j) = \beta\overline{\beta}' j^2 = \beta\overline{\beta}' \cdot (-pq) \in \mathscr{O}_K$ since $(q) = \mathfrak{q}\overline{\mathfrak{q}}$. The reduced discriminant of $R^{(q)}$ is a $q$-unit, so we get maximality of this order at the split place $q$ of $B$.

A global order that satisfies all of the above local conditions and contains $R' = \mathscr{O}_K \oplus \mathscr{O}_K j$ will have reduced discriminant $p$, as it would now be maximal at all primes dividing $Dq$. If we pick $X \in \mathscr{O}_K$ such that $\mathrm{N}(X) \equiv -pq \bmod D$, then $X$ is congruent modulo $\mathfrak{d}_\ell$ to a legitimate choice of $X_\ell$ for each $\ell|D$. If we let $\mathfrak{d}$ denote the different for $K$ over $\mathbf{Q}$, and $\mathscr{O}_\mathfrak{d}$ the semilocal subring of $K$ consisting of elements which are integral at the places dividing $\mathfrak{d}$, then the lattice $S' = S'_X = \{\alpha + \beta j \in B \,|\, \alpha \in \mathfrak{d}^{-1},\ \beta \in \mathfrak{d}^{-1}\mathfrak{q}^{-1}, \alpha - X\beta \in \mathscr{O}_{\mathfrak{d}}\}$ contains $R'$ and satisfies the above local conditions at each finite place to make it a maximal order of $B$ (for $\ell \nmid Dpq$, $S'$ has $\ell$-unit reduced discriminant). We are looking for an Eichler order containing $\mathscr{O}_K$ and having reduced discriminant $Np$. Recall that the primes of $N$ are split in $K$, so we can fix a factorization $N = \mathfrak{n}\overline{\mathfrak{n}}$ as a product of relatively prime conjugate ideals (in fact, $\mathfrak{n}$ is chosen in the main text to correspond to the kernel of an isogeny at a chosen Heegner point, but the choice won't matter here). In this case, $\mathfrak{n}S'\mathfrak{n}^{-1}$ is an order which must also be maximal, as its localization at any place is conjugate to that of $S'$ at the same place (since localizations of $\mathfrak{n}$ are principal). Since $\mathfrak{n}$ is relatively prime to $D$, this conjugation has no impact at the ramified places, so $\mathfrak{n}S'\mathfrak{n}^{-1}$ and $S'$ have the same congruence conditions at places dividing $D$. In fact,

$$\mathfrak{n}S'\mathfrak{n}^{-1} = \{\alpha + \beta j \,|\, \alpha \in \mathfrak{d}^{-1},\ \beta \in \mathfrak{d}^{-1}\mathfrak{q}^{-1}\mathfrak{n}\overline{\mathfrak{n}}^{-1},\ \alpha - X\beta \in \mathscr{O}_{\mathfrak{d}}\}.$$

The factor of $\overline{\mathfrak{n}}^{-1}$ is due to the relation $jx = \overline{x}j$ for all $x \in K$.

The intersection $S = S' \cap \mathfrak{n}S'\mathfrak{n}^{-1}$ is an Eichler order containing $\mathscr{O}_K$. Since $\mathfrak{n}$ and $\overline{\mathfrak{n}}$ are relatively prime, $S$ is explicitly described as

$$S = S_X = \{\alpha + \beta j \,|\, \alpha \in \mathfrak{d}^{-1},\ \beta \in \mathfrak{d}^{-1}\mathfrak{q}^{-1}\mathfrak{n},\ \alpha - X\beta \in \mathscr{O}_{\mathfrak{d}}\}. \tag{A.5}$$

The quotient $\mathscr{O}_K$-module $S'/S$ is $\mathscr{O}_K/\mathfrak{n}$, so $S$ has index $N$ in $S'$. Thus, $S$ has reduced discriminant $Np$ as desired. If we had also required $-pq \equiv 1 \bmod D$, then we could have taken $X = 1$ (see Remark A.7). This seems to implicitly be the choice made in [GZ]; certainly one does not get an order in (A.5) using $X = 1$ unless $-pq \equiv 1 \bmod D$.

When $p\mathscr{O}_K = \mathfrak{p}^2$ is ramified in $K$ and we take $\left(\frac{D,-q}{\mathbf{Q}}\right)$ as the explicit model for $B$ as discussed earlier, we can similarly compute the reduced discriminant of the order $R' = \mathscr{O}_K + \mathscr{O}_K j$ to be $Dq$. If $p$ is odd, we need only find $X \in \mathscr{O}_K$ such that $\mathrm{N}(X) \equiv -q \bmod D/p$; in exactly the same way as above, we find that the following is an Eichler order (containing $\mathscr{O}_K$) with reduced discriminant $Np$:

$$S = S_X = \{\alpha + \beta j \,|\, , \alpha \in \mathfrak{p}\mathfrak{d}^{-1},\ \beta \in \mathfrak{p}\mathfrak{d}^{-1}\mathfrak{q}^{-1}\mathfrak{n},\ \alpha - X\beta \in \mathscr{O}_{\mathfrak{d}}\}. \tag{A.6}$$

Note that $\mathfrak{p}\mathfrak{d}^{-1}$ has no factor at $\mathfrak{p}$ (since $p \neq 2$), and thus there is no local congruence condition at $\mathfrak{p}$, as $\alpha$ and $\beta$ are already integral at $\mathfrak{p}$. Of course, no alteration from $R'$ is required above $p$, as $\mathrm{ord}_p(Dq) = 1$, so $R'$ is maximal at the non-split place $p$ of $B$.

If $p = 2$ is ramified in $K$, the situation is somewhat different, but the explicit description of the order is the same as (A.6), where we again pick $X \in \mathscr{O}_K/\mathfrak{d}\mathfrak{p}^{-1}$ with $\mathrm{N}(X) \equiv -q \bmod D/p$. It is only necessary to explain the local conditions above 2, where $R'$ fails to be maximal (since $4|Dq$). Recall that the unique maximal order of a non-split quaternion algebra over a local field consists of precisely the elements of integral norm. If $\alpha$ and $\beta$ are non-integral at 2, the condition that $\mathrm{N}(\alpha) + q\mathrm{N}(\beta) = \mathrm{N}(\alpha + \beta j)$ be 2-integral is equivalent to $\mathrm{N}(\alpha/\beta) \equiv -q \bmod \mathrm{N}(\beta)^{-1}$. Such a congruence cannot happen $\bmod D_2$, since then $q$ would be a local norm and hence we could find a nonzero $\alpha + \beta j \in B$ with vanishing reduced norm, contrary to the fact that $B_2$ is a division algebra. There is no obstruction $\bmod D_2/2$, and indeed one can easily check that there is a unique $X_2 \in \mathscr{O}_{K,2}/\mathfrak{p}^{-1}\mathfrak{d}$ satisfying $\mathrm{N}(X_2) \equiv -q \bmod D_2/2$. Thus we conclude that $\alpha - X_2\beta \in \mathscr{O}_{K,2}$ (recall that the valuation ring of a local field is characterized by the property of having integral norm in a local subfield), and $\alpha, \beta \in \mathfrak{p}\mathfrak{d}^{-1}$. Likewise, (A.6) holds. Note that in the ramified case (for $p = 2$ and for $p \neq 2$), Remark A.7 allows us to take $X = 1$ (as in [GZ]) by also requiring $-q \equiv 1 \bmod D/p$ (this does not impose a congruence at $p$, unless $p = 2$).

What we have done so far is find an Eichler order containing $\mathscr{O}_K$ and having the correct reduced discriminant. This has not yet been shown to have a connection to the specific order $R = \mathrm{End}_{W/\pi}(\underline{x})$ in Theorem 10.5. Fortunately, the data of the embedding of $\mathscr{O}_K$ ensures some connection between such Eichler orders. The following theorem and its corollary are due to Eichler, and we include a proof for the reader's convenience. Corollary A.16 will be applied with $F = \mathbf{Q}$ and $E = K$.

**Theorem A.15.** *Let $F$ be a nonarchimedean local field with valuation ring $A$, maximal ideal $\mathfrak{m}$, and uniformizer $\pi$. Let $B$ be a split quaternion algebra over $F$ with a fixed embedding of the quadratic field $E$ over $F$. Let $S$ and $S'$ be Eichler orders in $B$ with the same isomorphism class* (*i.e.*, $\mathrm{Disc}S = \mathrm{Disc}S'$). *If $E \cap S = E \cap S' = \mathscr{O}_E$, then there exists a nonzero $x \in \mathscr{O}_E$ such that $xSx^{-1} = S'$.*

*Proof.* We will first prove the stament when $S$ and $S'$ are maximal orders, which we can assume are distinct. By Lemma A.9, we can assume $B = \mathrm{End}_F(V) = M_2(F)$ for a vector space $V = F^2$ over $F$, with $S = \mathrm{End}_A(N) = M_2(A)$ for a lattice $N = A^2$ in $V$ and $S' = \mathrm{End}_A(N')$ for some other lattice $N'$ in $V$ which is not a scalar multiple of $N$. By scaling $N'$ and changing the basis of $N$ if necessary, we can assume $N = Ae_1 \oplus Ae_2$ and $N' = Ae_1 \oplus \mathfrak{m}^n e_2$ for some $n > 0$, so

$$S' = \mathrm{End}_A(Ae_1 \oplus \mathfrak{m}^n e_2) = \begin{pmatrix} A & \mathfrak{m}^{-n} \\ \mathfrak{m}^n & A \end{pmatrix} = \gamma_n M_2(A)\gamma_n^{-1} = \gamma_n S \gamma_n^{-1},$$

where $\gamma_n = \left(\begin{smallmatrix} 0 & 1 \\ \pi^n & 0 \end{smallmatrix}\right)$. Note that the elements of $B^\times = \mathrm{GL}_2(F)$ which conjugate $S$ into $S'$ are exactly the elements of the coset $\gamma_n S^\times = \gamma_n \mathrm{GL}_2(A)$. Since $S \cap S' = \left(\begin{smallmatrix} A & A \\ \mathfrak{m}^n & A \end{smallmatrix}\right)$ is the standard Eichler order of reduced discriminant $\mathfrak{m}^n$ and $\mathscr{O}_E \subseteq S \cap S'$ by hypothesis, we can find $a, b, c \in A$ so that an $A$-basis of $\mathscr{O}_E$ is given by the identity 1 and the element $\rho = \left(\begin{smallmatrix} a & b \\ \pi^n c & 0 \end{smallmatrix}\right)$. Since $\rho/\pi \in E$ is not in $\mathscr{O}_E$, yet $S \cap E = S' \cap E = \mathscr{O}_E$, it follows that $\rho/\pi$ is not contained in either of $S$ or $S'$. Hence, by inspection of the matrix for $\rho$ we see that either $a$ or $bc$ must be a unit in $A$. If $bc \in A^\times$ then $\rho \in \gamma_n \mathrm{GL}_2(A)$, while if not then $a$ is a unit and hence we can instead replace $\rho$ with $\rho + \pi^n = \left(\begin{smallmatrix} a+\pi^n & b \\ \pi^n c & \pi^n \end{smallmatrix}\right)$ which does lie in $\gamma_n \mathrm{GL}_2(A)$. That is, we may assume $\rho \in \mathscr{O}_E$ conjugates $S$ into $S'$. This settles the case when $S$ and $S'$ are maximal.

Now consider the case in which the isomorphic $S$ and $S'$ are non-maximal, say with reduced discriminant $\mathfrak{m}^n$ for some $n > 0$. By conjugation, we may assume $S = \left(\begin{smallmatrix} A & A \\ \mathfrak{m}^n & A \end{smallmatrix}\right)$ is the standard Eichler order of reduced discriminant $\mathfrak{m}^n$ in $M_2(A)$. Thus, $S = S_1 \cap S_2$ where the $S_j$'s are the maximal orders considered above (i.e., $S_1 = M_2(A)$ and $S_2 = \mathrm{End}_A(Ae_1 \oplus \mathfrak{m}^n e_2)$). We may write $S' = S_1' \cap S_2'$ where the $S_j'$'s are maximal orders in $B$. The special case of maximal orders ensures that there exists a nonzero $x \in \mathscr{O}_E$ with $xS_1x^{-1} = S_1'$. If $S$ and $x^{-1}S'x$ are conjugate by a nonzero element of $\mathscr{O}_E$ (as in the statement of the theorem), then we'll be done. Thus, we may assume $S_1' = M_2(A) = S_1$.

Since the two Eichler orders $S$ and $S'$ are presented as intersections of a *common* maximal order $S_1$ with other maximal orders $S_2$ and $S_2'$, from the *proof* of Lemma A.9 there exists $y \in S_1^\times = \mathrm{GL}_2(A)$ such that $ySy^{-1} = S'$. The given condition $\mathscr{O}_E \subseteq S' = ySy^{-1}$ says exactly that $y^{-1}\rho y \in S_2 = \left(\begin{smallmatrix} A & \mathfrak{m}^{-n} \\ \mathfrak{m}^n & A \end{smallmatrix}\right)$, where $\{1, \rho\}$ is an $A$-basis of $\mathscr{O}_E$, say with $\rho$ chosen as above. Since $y^{-1}\rho y \in S_1 = M_2(A)$, the condition for membership in $S_2$ is just the condition that the lower left corner entry of $y^{-1}\rho y$ lie in $\mathfrak{m}^n$. Explicitly, if $y = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$, then this condition is equivalent to saying $c'(aa' + bc') \in \mathfrak{m}^n$. Conversely, any $y = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in \mathrm{GL}_2(A) = S_1^\times$ satisfying this latter property automatically satisfies $\mathscr{O}_E \subseteq ySy^{-1}$, with the Eichler order $ySy^{-1}$ given by intersecting $S_1 = M_2(A)$ with another maximal order.

Suppose $c'$ is not a unit in $A$. If $c' \in \mathfrak{m}^n$ (automatic if $n = 1$), then $y \in \Gamma_0(\mathfrak{m}^n) \stackrel{\mathrm{def}}{=} S_2 \cap S_1^\times = S_2 \cap \mathrm{GL}_2(A)$ conjugates $S_2$ into itself, so $S = S'$ and we are done. If, on the other hand, $n > 1$ and $c' \in \mathfrak{m} - \mathfrak{m}^n$, then since $c'(aa' + bc') \in \mathfrak{m}^n$ we get $aa' + bc' \in \mathfrak{m}$, or in other words $aa' \in \mathfrak{m}$. But $y \in \mathrm{GL}_2(A)$ and $c' \in \mathfrak{m}$, so $a' \notin \mathfrak{m}$ and hence $a \in \mathfrak{m}$. This combination of conditions implies $\rho/\pi \in E$ has integral trace and norm as a matrix, so $\rho/\pi \in E$ is integral over $A$ and therefore lies in $\mathscr{O}_E$. This is a contradiction, so we can assume $c' \in A^\times$, so $aa' + bc' \in \mathfrak{m}^n$.

We must have $a \notin \mathfrak{m}$, as otherwise $a, b \in \mathfrak{m}$, a contradiction (since $\rho/\pi \notin S_1 = M_2(A)$). There is a more useful way to express these conditions: for $t = \left(\begin{smallmatrix} 0 & 1 \\ a & b \end{smallmatrix}\right) \in \mathrm{GL}_2(A)$, the condition on $y$ is precisely that $ty \in \Gamma_0(\mathfrak{m}^n)$. We are now in position to find a nonzero $x \in \mathscr{O}_E$ such that $xSx^{-1} = S'$. Let $Z$ be the group generated by the central element $\pi \in S_1 = M_2(A) \subseteq B$. Note that conjugation by $\gamma_n$ is an involution of both $S$ and $S^\times = \Gamma_0(\mathfrak{m}^n)$, since it interchanges $S_1 = M_2(A)$ and $S_2 = \gamma_n M_2(A)\gamma_n^{-1}$ (with $S = S_1 \cap S_2$). The group $G$ generated by $\Gamma_0(\mathfrak{m}^n)$, $\gamma_n$, and $Z$ is exactly the subgroup of elements of $B^\times = \mathrm{GL}_2(F)$ which conjugate $S$ into itself. To prove this, first note that if $b \in B^\times$ conjugates $S$ into itself, then conjugation by $b$ either preserves or swaps the unique pair of maximal orders $S_1$ and $S_2$ whose intersection is the non-maximal Eichler order $S$ (see Remark A.11). Multiplying against $\gamma_n$ if necessary permits us to assume that conjugation by $b$ preserves both $S_j$'s. But for a maximal order $R$ in $B$, the elements of $B^\times$ which conjugate $R$ into $R$ are the elements of $F^\times R^\times = \pi^{\mathbf{Z}} R^\times$ (since $R = \mathrm{End}_A(N)$ forces $bRb^{-1} = \mathrm{End}_A(b(N))$, and $\mathrm{End}_A(b(N)) = \mathrm{End}_A(N)$ if and only if $b(N) = cN$ for some $c \in F^\times$). Thus,

$$b \in \pi^{\mathbf{Z}}(S_1^\times \cap (\pi^{\mathbf{Z}} \cdot S_2^\times)) = \pi^{\mathbf{Z}}(S_1^\times \cap S_2^\times) = \pi^{\mathbf{Z}} S^\times = Z \cdot \Gamma_0(\mathfrak{m}^n).$$

Here, the first equality uses that the reduced norm of a unit in an order is a unit in $A$.

It suffices to find a nonzero $x \in \mathscr{O}_E$ such that $y^{-1}x \in G$, since then $x = y(y^{-1}x)$ conjugates $S$ into $S'$. Since $y^{-1} \in \Gamma_0(\mathfrak{m}^n)t$, we need to find $x \in \mathscr{O}_E$ such that $tx \in G$. This is something we can accomplish by observation: letting $x = (-a + \pi^r) + \rho$ for an $r > 0$ to be determined, we compute

$$tx = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix} \begin{pmatrix} \pi^r & b \\ c\pi^n & -a + \pi^r \end{pmatrix} = \begin{pmatrix} c\pi^n & -a + \pi^r \\ a\pi^r + bc\pi^n & b\pi^r \end{pmatrix}.$$

If $bc \in \mathfrak{m}$ then $r = n$ makes $tx \in \Gamma_0(\mathfrak{m}^n)$, while if $bc \notin \mathfrak{m}$ then $r = n + 1$ does the same.

■

**Corollary A.16.** *Let $F$ be a number field, and $B$ a quaternion algebra over $F$ equipped with a fixed embedding of a quadratic extension $E$. If $S$ and $S'$ are Eichler orders of $B$ with the same reduced discriminant such that $S \cap E = S' \cap E = \mathscr{O}_E$, then there exists a nonzero ideal $I$ of $\mathscr{O}_E$ such that $ISI^{-1} = S'$.*

*Proof.* A priori $S_v = S'_v$ for all but finitely many $v$. For all such $v$ define $x_v = 1$. At the remaining places, use Theorem A.15 to construct $x_v \in \mathscr{O}_{E_v}$ so that $x_v S x_v^{-1} = S'$. Let $I$ be the ideal with $\mathrm{ord}_v(I) = \mathrm{ord}_v(x_v)$ for all non-archimedean places $v$. The equality $ISI^{-1} = S'$ can be checked place by place.

■

In the situation given in Corollary A.16, if we have $S' = ISI^{-1}$ with $I$ a *principal* ideal, then picking a generator gives an explicit isomorphism from $S$ to $S'$ which is $\mathscr{O}_E$-linear. Since $E$ is its own centralizer in $B$ and all automorphisms of $B$ are inner (Skolem-Noether), we conclude that the set of $\mathscr{O}_E$-linear isomorphism classes of Eichler orders in $B$ with a fixed reduced discriminant and containing $\mathscr{O}_E$ admits a simply transitive action by the class group of $\mathscr{O}_E$. In our situation of interest (with $F = \mathbf{Q}$ and $E = K$), upon fixing a place $v$ of $H$ over $p$ we have already established (in (7.8)) for $R_v = \mathrm{End}_{W/\pi}(\underline{x})$ that if $\sigma \in \mathrm{Gal}(H/K) \simeq \mathrm{Cl}_K$ corresponds to the ideal class of some nonzero integral ideal $\mathfrak{a}$, then there is an abstract isomorphism $\mathfrak{a}^{-1}R_v\mathfrak{a} \simeq \mathrm{End}_{W/\pi}(\underline{x^\sigma})$ as $\mathscr{O}_K$-algebras, where $W = W_v$ is the completion of a maximal unramified extension over $v$. Thus, if $B$ denotes the unique (up to isomorphism) quaternion algebra over $\mathbf{Q}$ which is ramified at exactly $p$ and $\infty$ and we fix an embedding of $K$ into $B$, then we have a bijection between two sets: the set of isomorphism classes of Eichler orders in $B$ containing $\mathscr{O}_K$ and having reduced discriminant $Np$, and the set of Heeger points $\{x^\sigma\}$ in $X_0(N)(H)$, via $\mathrm{End}_{W/\pi}(\underline{x^\sigma}) \leftrightarrow \sigma$.

Since $p$ is non-split in $K$, we have a transitive free action of $\mathrm{Gal}(H/K)$ on the $v$'s of $H$ over $p$. An important consequence is that rather than fixing $\sigma$ and considering the isomorphism classes $\mathfrak{a}^{-1}R_v\mathfrak{a} \simeq \mathrm{End}_{W_v/\pi}(\underline{x^\sigma})$ as the place $v$ varies over $p$, we can fix one place $v_0$ and study $\mathfrak{a}^{-1}R_{v_0}\mathfrak{a}$ as the ideal class $\mathscr{A}$ of $\mathfrak{a}$ varies. Both processes exhaust (without repetition) the set of $\mathscr{O}_K$-algebra isomorphism classes of Eichler orders in $B$ with reduced discriminant $Np$ and containing $\mathscr{O}_K$. Thus, we can fix the Eichler order $S = S_X$ from (A.5) and then for (A.1) we may contemplate a sum over the orders $R^{\mathfrak{b}} = \mathfrak{b}^{-1}S\mathfrak{b}$ as $\mathfrak{b}$ runs over integral ideals representing the ideal classes of $K$. It is this latter point of view (which avoids the language of supersingular elliptic curves) which will dominate all that follows.

We are now ready to prove [GZ, Ch III, Prop 9.7, 9.11], allowing $D$ to be even. We need to introduce some notation. We write $[\mathfrak{a}]$ to denote the ideal class of a fractional ideal $\mathfrak{a}$ of $K$, and for any ideal class $\mathscr{B}$ we define $R_{\mathscr{B}}(m) = \sum_c r_{\mathscr{B}c}(m)$ where $c$ runs over the *squares* in the ideal class group of $K$. Note that $R_{\mathscr{B}\mathscr{B}'} = R_{\mathscr{B}^{-1}\mathscr{B}'}$ for any two ideal classes $\mathscr{B}$ and $\mathscr{B}'$. For example, if a prime $\ell$ of $\mathbf{Q}$ is split in $K$ (such as our prime $q$ chosen above), say $\ell = \mathfrak{l}\bar{\mathfrak{l}}$, then $[\mathfrak{l}]$ and $[\bar{\mathfrak{l}}]$ are inverse to each other in $\mathrm{Cl}_K$. Taking $\ell = q$, we conclude that $R_{\mathscr{A}[\mathfrak{q}\mathfrak{n}]}$ is independent of the choice of prime $\mathfrak{q}$ of $K$ over $q$, and is likewise independent of the choice of factorization $N = \mathfrak{n}\bar{\mathfrak{n}}$. In particular, the formula in Theorem A.17 below does not depend on the choice of $\mathfrak{q}$ (but only through the entire identity do we see that the right side is independent of the choice of $q$ as above):

**Theorem A.17.** *Suppose $p$ is inert in $K$ and $\mathfrak{q}$ is chosen over $q$. For $n \geq 1$, define $\delta(n) = 2^{e(n)}$ where $e(n)$ is the number of prime factors of $\gcd(D, n)$. Then*

$$\langle c, T_m d^\sigma \rangle_p = -ur_{\mathscr{A}}(m)h_K \mathrm{ord}_p(m)\log p - \log p \cdot u^2 \cdot \sum_{\substack{0 < n < m|D|/N \\ p|n}} \mathrm{ord}_p(pn) r_{\mathscr{A}}(m|D| - nN)\delta(n) R_{\mathscr{A}[\mathfrak{q}\mathfrak{n}]}(n/p).$$

We specify a choice of $\mathfrak{q}$ in the statement of the theorem because of the appearance of $\mathfrak{q}$ as the final term on the product on the right side, but keep in mind that the choice does not affect the value of that term.

*Proof.* By (A.1) and the preceding discussion, we have to prove the identity

$$\sum_{\mathfrak{b}} \sum_{\substack{b\in R^{(\mathfrak{b})}\mathfrak{a}/\pm 1\\ \mathrm{N}(b)=m\mathrm{N}\mathfrak{a},\, b_-\neq 0}} (1+\mathrm{ord}_p(\mathrm{N}(b_-))) = u^2\cdot \sum_{\substack{0<n<m|D|/N\\ p|n}} \mathrm{ord}_p(pn) r_{\mathscr{A}}(m|D|-nN)\delta(n)R_{\mathscr{A}[\mathfrak{q}\mathfrak{n}]}(n/p), \tag{A.7}$$

where $\mathfrak{b}$ runs over representative of ideal classes $\mathscr{B}$ of $K$ and $R^{(\mathfrak{b})} \stackrel{\mathrm{def}}{=} \mathfrak{b}S\mathfrak{b}^{-1}$ ranges over the $\mathscr{O}_K$-algebra isomorphism classes of Eichler orders in $B$ which contain $\mathscr{O}_K$ and have reduced discriminant $Np$. Here, $S = S_X$ is chosen as in (A.5), with $\mathrm{N}(X) \equiv -pq \bmod D$. We need to analyze sums over all such $R^{(\mathfrak{b})}$'s, with the sum for each $R = R^{(\mathfrak{b})}$ ranging over $R\mathfrak{a}/\pm 1$, where $\mathfrak{a}$ is a prime-to-$p$ representative of the ideal class $\mathscr{A}$ corresponding to our fixed $\sigma \in \mathrm{Gal}(H/K)$. Multiplying $\mathfrak{a}$ by a principal ideal which is prime to $p$ has no impact on the left side of (A.7), so we may select our choice to satisfy the additional property that $\gcd(\mathfrak{a}, D) = 1$. This will be convenient later.

Taking $\mathfrak{b}$ fixed above (so we write $R$ rather than $R^{(\mathfrak{b})}$), using (A.5) and the fact that $j$ acts on $K$ through the involution in $\mathrm{Gal}(K/\mathbf{Q})$ yields

$$R\mathfrak{a} = \{\alpha+\beta j \mid \alpha\in\mathfrak{d}^{-1}\mathfrak{a},\ \beta\in\mathfrak{d}^{-1}\mathfrak{q}^{-1}\mathfrak{n}\mathfrak{b}\overline{\mathfrak{b}}^{-1}\overline{\mathfrak{a}},\ \alpha - X^{\mathfrak{a}\mathfrak{b}}\beta\in\mathscr{O}_{\mathfrak{d}}\}, \tag{A.8}$$

where $X^{\mathfrak{a}\mathfrak{b}} \in \mathscr{O}_K/\mathfrak{d}$ is an element we need to define in terms of $X \in \mathscr{O}_K/\mathfrak{d}$ from (A.5); recall that only $X \in \mathscr{O}_K/\mathfrak{d}$ (rather than $X \in \mathscr{O}_K$) matters for the construction of the model $S$ in (A.5). Note that the separate conditions on $\alpha$ and $\beta$ in (A.8) define a lattice and hence are properties which can be checked locally (where all fractional ideals of $\mathscr{O}_K$ such as $\mathfrak{a}$ and $\mathfrak{b}$ become principal).

To define $X^{\mathfrak{a}\mathfrak{b}}$, we define an element $X^I$ more generally for any nonzero integral ideal $I$ of $\mathscr{O}_K$. Picking an element $y \in \mathscr{O}_K$ for which $y\mathscr{O}_{\mathfrak{d}} = I_{\mathfrak{d}}$ (i.e., $(y)$ has the same factors as $I$ at places dividing $\mathfrak{d}$), note that $y\overline{y}^{-1}$ is a unit at $\mathfrak{d}$ since the primes of $\mathfrak{d}$ are ramified over $\mathbf{Q}$. Moreover, the residue fields at such primes coincide with the prime field, so changing $y$ by a $\mathfrak{d}$-unit has no impact on $y\overline{y}^{-1} \bmod \mathfrak{d} \in (\mathscr{O}_K/\mathfrak{d})^\times$. Thus, the quantity $y\overline{y}^{-1}X \bmod \mathfrak{d}$ only depends on $I_{\mathfrak{d}}$ and $X \bmod \mathfrak{d}$. Hence, if we define $X^I \in \mathscr{O}_K$ to be any solution to the congruence $X^I \equiv y\overline{y}^{-1}X \bmod \mathfrak{d}$, then $X^I$ is well-defined modulo $\mathfrak{d}$ and $\mathrm{N}(X^I) \equiv -pq \bmod D$ since $\mathrm{N}(X) \equiv -pq \bmod D$ and $\mathrm{N}(y\overline{y}^{-1}) = y\overline{y}^{-1}\overline{y}y^{-1} = 1$. Thus, $R = R^{(\mathfrak{b})} = \mathfrak{b}S_X\mathfrak{b}^{-1}$ is equal to $S_{X^{\mathfrak{b}}}$ and we see why $X^{\mathfrak{a}\mathfrak{b}}$ intervenes in the description (A.8).

The $X \mapsto X^I$ construction provides an "action" of the multiplicative monoid of nonzero integral ideals on the set of solutions in $\mathscr{O}_K/\mathfrak{d}$ to the congruence $\mathrm{N}(X) \equiv -pq \bmod D$. We need to get a handle on the set of such solutions since we are working with an Eichler order $S$ from (A.5) which depends on the specification of such a solution. For each $\ell|D$, we claim that the congruence $\mathrm{N}(X) \equiv -pq \bmod D_\ell$ in $\mathscr{O}_{K,\ell}/\mathfrak{d}_\ell$ has exactly two (necessarily unit) solutions, or equivalently (since the congruence *does* have solutions, due to how we chose $q$) that $\mathrm{N} : (\mathscr{O}_{K,\ell}/\mathfrak{d}_\ell)^\times \to (\mathbf{Z}_\ell/D_\ell)^\times$ has kernel of order 2. For odd $\ell$ this map is just squaring on $\mathbf{F}_\ell^\times$. If $\ell = 2$ and $D_2 = 8$ then $\mathscr{O}_{K,\ell} = \mathbf{Z}_2[\sqrt{\pm 2d}]$ with $d \in \mathbf{Z}_2^\times$ mattering mod 8, and one verifies the claim by direct calculation in all four cases. If $\ell = 2$ and $D_2 = 4$ then $\mathscr{O}_{K,\ell} = \mathbf{Z}_2[\sqrt{-1}]$ or $\mathbf{Z}_2[\sqrt{3}]$, and again a direct calculation does the job.

We conclude that if $\mathfrak{d}$ has $t$ prime factors then there are $2^t$ congruence classes $X \in \mathscr{O}_K/\mathfrak{d}$ that satisfy the norm congruence condition $\mathrm{N}(X) \equiv -pq \bmod D$. At each place, $X \mapsto X^I$ either fixes the two local solutions or swaps them, so we deduce $X^{I^2} \equiv X \bmod \mathfrak{d}$ for any $I$. On the other hand, the action of a prime factor $\mathfrak{m}$ of $\mathfrak{d}$ is non-trivial on the local congruence condition at $\mathfrak{m}$ and is trivial at the other factors. Indeed, if the residue characteristic $\ell$ at $\mathfrak{m}$ is odd then we may take $y$ very congruent to $i = \sqrt{D}$ at $\mathfrak{m}$ and very congruent to 1 at the other factors of $\mathfrak{d}$ (so $y\overline{y}^{-1} \equiv -1 \bmod \ell$ but $y\overline{y}^{-1}$ is congruent to 1 along the other primary factors of $D$). Meanwhile, if $\ell$ is 2 and $D_2 = 8$ then we can argue similarly using $i/2$. Finally, if $\ell = 2$ and $D_2 = 4$ (so $\mathfrak{d}_2 = (2)$) then we can pick $y$ highly congruent to $1 + i/2$ over 2, so $y\overline{y}^{-1} \equiv i/2 \not\equiv 1 \bmod \mathfrak{d}_2$.

At this point it is helpful to interject a few comments about the class group of quadratic imaginary fields. It is well-known, and easily checked directly, that 2-torsion subgroup of the class group is generated by the ideal classes above the ramified primes. For odd primes $\ell|D$ and $\ell = 2$ if $8|D$, these are of the form $(\ell, \sqrt{D})$, while for $D_2 = 4$ we have $(2, 1+\sqrt{D/4})$ above 2. The order of the group generated by these elements is $2^{t-1}$ if $t$ distinct primes divide $D$, since the only non-square ideal over the ramified places that becomes trivial in the ideal class group is $(\sqrt{D}) = (i)$ (unless $D_2 = 4$, which requires a separate argument). In this way, we see that the above construction really defines a transitive free action of $(\mathbf{Z}/2)^t$ on the set of possible $X$'s, and this $(\mathbf{Z}/2)^t$ may be naturally identified with an extension of $\mathrm{Cl}_K[2]$ by $\mathbf{Z}/2$. Our formula (A.8) is slightly more general than [GZ, Ch III, (9.3)] in that it allows $D$ to be even, in which case the action $X \mapsto X^I$ affects the local congruence above 2 in a slightly more complicated manner when $D_2 = 4$ (if we avoid even $D$ and take $X = 1$, then $X^{\mathfrak{ab}}$ collapses to a product of local signs, as in [GZ]).

Continuing with the method of Gross-Zagier, note that an element $b = \alpha + \beta j \in R\mathfrak{a}$ has reduced norm which is the sum of the reduced norms of $b_+ = \alpha$ and $b_- = \beta j$, so in the presence of the conditions on the sum on the left side of (A.7) we get $m\mathrm{N}(\mathfrak{a}) = \mathrm{N}(b) = \mathrm{N}(\alpha) + pq\mathrm{N}(\beta)$, where $pq\mathrm{N}(\beta) = \mathrm{N}(b_-) \neq 0$, so necessarily $\beta \neq 0$. If we introduce the ideals

$$\mathfrak{c} = (\alpha)\mathfrak{d}\mathfrak{a}^{-1}, \quad \mathfrak{c}' = (\beta)\mathfrak{d}\mathfrak{q}\mathfrak{n}^{-1}\mathfrak{b}^{-1}\overline{\mathfrak{b}}\overline{\mathfrak{a}}^{-1} \neq 0, \tag{A.9}$$

then

$$\mathrm{N}(\mathfrak{c}) + (Np)\mathrm{N}(\mathfrak{c}') = \mathrm{N}(\alpha)\frac{|D|}{\mathrm{N}(\mathfrak{a})} + \mathrm{N}(\beta)\cdot Np\frac{|D|q}{\mathrm{N}(\mathfrak{a})N} = (\mathrm{N}(\alpha) + pq\mathrm{N}(\beta))\frac{|D|}{\mathrm{N}(\mathfrak{a})} = m|D|.$$

In particular, since $\gcd(m, N) = \gcd(N, D) = 1$ and $N > 1$, so $Np \nmid m|D|$, we must have $\mathrm{N}(\mathfrak{c}) \neq 0$, so $\mathfrak{c} \neq 0$ (so $\alpha \neq 0$).

To summarize, the conditions on $\alpha$ and $\beta$ say that $\mathfrak{c}$ and $\mathfrak{c}'$ are nonzero integral ideals with $\mathrm{N}(\mathfrak{c}) + Np\mathrm{N}(\mathfrak{c}') = m|D|$. Since $\mathfrak{d} = (\sqrt{D})$ is a principal ideal, we see that $\mathfrak{c}$ is in the ideal class $\mathscr{A}^{-1}$ while $\mathfrak{c}'$ is in the ideal class $\mathscr{A}\mathscr{B}^{-2}[\mathfrak{q}\mathfrak{n}^{-1}]$, where $\mathscr{B}$ is the ideal class of $\mathfrak{b}$ (the same $\mathfrak{b}$ which is implicit in our model $R = R^{(\mathfrak{b})}$). Thus, for each $\alpha + \beta j \in R\mathfrak{a}$ we have constructed a pair of nonzero integral ideals $\mathfrak{c}$ and $\mathfrak{c}'$ of $\mathscr{O}_K$ which lie in specified ideal classes and satisfy a single norm relation. The key is to prove that this construction exhausts all such pairs of integral ideals, and to determine how badly this construction fails to be injective. Recall that on the left side of (A.7), we are varying $\mathscr{B}$ over all ideal classes of $K$, each of which we include once. However, there is a complication caused by the fact that the class of $\mathfrak{c}'$ only depends on $\mathscr{B}^2$, not $\mathscr{B}$, so a pair of integral ideals $(\mathfrak{c}, \mathfrak{c}')$ may arise from elements $\alpha + \beta j$ in several Eichler orders in $B$ which contain $\mathscr{O}_K$ and have reduced discriminant $Np$, but are not conjugate as $\mathscr{O}_K$-algebras.

To work out how much repetition we encounter, suppose that we start with a pair of nonzero integral ideals $\mathfrak{c} \in \mathscr{A}^{-1}$ and $\mathfrak{c}' \in \mathscr{A}\mathscr{B}^{-2}[\mathfrak{q}\mathfrak{n}^{-1}]$ with $\mathrm{N}(\mathfrak{c}) + Np\mathrm{N}(\mathfrak{c}') = m|D|$. Define $n = p\mathrm{N}(\mathfrak{c}')$, so $\mathrm{N}(\mathfrak{c}) = m|D| - nN$. We can reverse the original definitions (A.9) to define nonzero principal ideals

$$\mathfrak{c}\mathfrak{a}\mathfrak{d}^{-1}, \quad \mathfrak{c}'\mathfrak{d}^{-1}\mathfrak{q}^{-1}\mathfrak{n}\mathfrak{b}\overline{\mathfrak{b}}^{-1}\overline{\mathfrak{a}}. \tag{A.10}$$

Since $u = |\mathscr{O}_K^\times|/2$, there are $4u^2 = (2u)^2$ choices of respective generators $\alpha$ and $\beta$ of these principal ideals. Because of the integrality $\mathfrak{c}$ and $\mathfrak{c}'$, all local conditions in (A.8) necessary to make $\alpha + \beta j \in R\mathfrak{a}$ are automatically satisfied *except* for possibly the congruence conditions $\alpha - X^{\mathfrak{ab}}\beta \in \mathscr{O}_{K,\ell}$ at the primes $\ell | D$.

For $\ell | D$, all we have to work with is

$$\mathrm{N}(\alpha + \beta j) = \mathrm{N}(\alpha) + pq\mathrm{N}(\beta) = \frac{\mathrm{N}(\mathfrak{c}) + Np\cdot\mathrm{N}(\mathfrak{c}')}{|D|}\cdot\mathrm{N}(\mathfrak{a}) = m\mathrm{N}(\mathfrak{a}),$$

so $\mathrm{N}(\alpha) \equiv -pq\mathrm{N}(\beta) \pmod{\mathscr{O}_{K,\ell}}$. In particular, since $-pq$ is a local unit, $\alpha$ and $\beta$ are either both $\ell$-integral or else neither is $\ell$-integral and they have the same pole order (so $\alpha/\beta$ is an $\ell$-unit). We claim that there is a squarefree integral ideal $I | \mathfrak{d}$ (possibly more than one) such that $\alpha \equiv X^{\mathfrak{ab}I}\beta \pmod{\mathscr{O}_{\mathfrak{d}}}$. This follows from:

**Lemma A.18.** *For each prime $\ell | D$, we have $\alpha - X'_\ell\beta \in \mathscr{O}_{K,\ell}$ for at least one of the two units $X'_\ell \in \mathscr{O}_{K,\ell}/\mathfrak{d}_\ell$ with norm $-pq \in \mathbf{Z}_\ell/D_\ell$. Both such units work if and only if $\ell | \mathrm{N}(\mathfrak{c}')$.*

*Proof.* We treat separately the case of odd $\ell$ and $\ell = 2$. First assume $\ell \neq 2$, and we want to prove that at least one choice works. This is clear when $\alpha$ and $\beta$ are $\ell$-integral (both choices work). When neither is $\ell$-integral, $\alpha/beta \in \mathscr{O}_{K,\ell}^\times$ has norm congruent to $-pq$ mod $D_\ell$, so there is a unique $X' \in \mathscr{O}_K/\mathfrak{d}_\ell$ with norm $-pq$ mod $D_\ell$ such that $\alpha/\beta \equiv X' \pmod{\mathscr{O}_{K,\ell}}$. But look at the expression $(\beta) = \mathfrak{c}'\mathfrak{d}^{-1}\mathfrak{q}^{-1}\mathfrak{n}\mathfrak{b}\overline{\mathfrak{b}}^{-1}\overline{\mathfrak{a}}$. Since $\mathfrak{d}$ is squarefree at $\ell$, $\gcd(D, \mathfrak{a}qN) = 1$, and any ramified prime factor in $\mathfrak{b}$ is cancelled against its inverse appearing in $\overline{\mathfrak{b}}^{-1}$, we see that $\beta$ is $\ell$-integral if and only if $\mathfrak{c}'$ is divisible by the prime over $\ell$, and that otherwise $\beta$ (hence $\alpha$) has a simple pole at that prime.

When $\alpha$ and $\beta$ have simple poles over $\ell$, $\alpha \equiv X'_\ell \beta \pmod{\mathscr{O}_{K,\ell}}$ is equivalent to $\alpha/\beta \equiv X'_\ell \bmod \mathfrak{d}_\ell$. Thus, in general for odd $\ell|D$ we see that $\alpha \equiv X'_\ell \beta \pmod{\mathscr{O}_{K,\ell}}$ for at least one of the two possible values of $X'_\ell$, with both occurring if and only if $\alpha$ and $\beta$ are $\ell$-integral, which in turn is equivalent to $\ell|\mathrm{N}(\mathfrak{c}')$.

Now we turn to the more subtle case $\ell = 2|D$. For this case it will be convenient to recall from Remark A.7 that $-pq \equiv 1 \bmod 8$ when $D$ is even in the inert case. Once again, either $\alpha$ and $\beta$ are both $\ell$-integral or neither is, and when both are integral then clearly $2|\mathrm{N}(\mathfrak{c}')$. To keep track of the pole order and the condition of whether $2 \nmid \mathrm{N}(\mathfrak{c}')$ in non-integral cases, we separately treat the cases $D_2 = 4$ and $D_2 = 8$. First assume $D_2 = 4$, so $D = -4d$ with positive squarefree $d \equiv 1 \bmod 4$. Representatives for $(\mathfrak{d}_2^{-1}/\mathscr{O}_{K,2}) - \{0\}$ are $1/2$, $i/4$, and $1/2 + i/4$ (recall $i = \sqrt{D}$). The norms of these classes are well-defined in $D_2^{-1}\mathbf{Z}_2/\mathbf{Z}_2 = 4^{-1}\mathbf{Z}/\mathbf{Z}$, and are represented by $1/4$, $-D/16 \equiv 1/4$, and $(1+d)/4 \equiv 1/2$. Since $\mathrm{N}(X'_2) \equiv -pq \bmod D_2$ and $-pq \equiv 1 \bmod 8$, the two solutions $X'_2 \in \mathscr{O}_{K,2}/\mathfrak{d}_2$ are $1$ and $i/2$ (since $d \equiv 1 \bmod 4$). Since $\beta$ has at worst a double pole in $K_2$, so $\mathrm{N}(\beta)$ has at worst a double pole in $\mathbf{Q}_2$, the congruence $\mathrm{N}(\alpha) \equiv -pq\mathrm{N}(\beta) \pmod{\mathbf{Z}_2}$ is equivalent to $\mathrm{N}(\alpha) \equiv \mathrm{N}(\beta) \pmod{\mathbf{Z}_2}$.

Thus, $\alpha, \beta \in \mathfrak{d}_2^{-1}/\mathscr{O}_{K,2}$ are either both in the class of $1/2 + i/4$ or are each in one of the two classes $\{1/2, i/4\}$. Running through both options and recall that $X'_2 \in \{1, i/2\}$, we see that $\alpha \equiv X'_2\beta \pmod{\mathscr{O}_{K,2}}$ can be solved in the non-integral case, with both $X'_2$ options working if and only if we are in the case where $\alpha$ and $\beta$ are in the class of $1/2 + i/4$. This is also the only case in which $(\beta)\mathfrak{d}_2$ lies in the maximal ideal of $\mathscr{O}_{K,2}$ (i.e., $\beta$ has a simple pole, rather than a double pole), or equivalently $2|\mathrm{N}(\mathfrak{c}')$.

The case $D_2 = 8$ (i.e., $D = -8d$ with odd squarefree $d > 0$) goes similarly, but there are more cases to consider. Now $\mathfrak{d}_2^{-1}/\mathscr{O}_{K,2} \simeq \mathbf{Z}/4 \times \mathbf{Z}/2$ as an abelian group, with generators $i/8$ (of order 4) and $1/2$. Norms on here are well-defined in $D_2^{-1}\mathbf{Z}_2/\mathbf{Z}_2 = 8^{-1}\mathbf{Z}/\mathbf{Z}$. The two solutions in $\mathscr{O}_{K,2}/\mathfrak{d}_2$ to $\mathrm{N}(X'_2) \equiv -pq \bmod D_2 = 1 \bmod 8$ are $X'_2 = \pm 1$, so for non-zero classes $\alpha, \beta \in \mathfrak{d}_2^{-1}/\mathscr{O}_{K,2}$ we want $\mathrm{N}(\alpha) \equiv \mathrm{N}(\beta) \pmod{\mathbf{Z}_2}$ if and only if $\alpha = \pm\beta \pmod{\mathscr{O}_{K,2}}$, and that both such signs work precisely when $\beta$ is 2-torsion (i.e., does not have a triple pole, which is to say exactly that $(\beta)\mathfrak{d}_2$ lies in the maximal ideal, or in other words $2|\mathrm{N}(\mathfrak{c}')$). This is straightforward by inspection of the norm of each non-zero class in $\mathfrak{d}_2^{-1}/\mathscr{O}_{K,2}$.

■

We conclude that when given $\mathfrak{c}$ and $\mathfrak{c}'$ of the desired type and constructing pairs $\alpha$ and $\beta$ from these, necessarily $\alpha \equiv X^{\mathfrak{a}\mathfrak{b}I}\beta \pmod{\mathscr{O}_\mathfrak{d}}$ for *some* integral squarefree ideal $I|\mathfrak{d}$ (which is visibly 2-torsion in the class group). This corresponds to using $\mathfrak{b}I$ instead of $\mathfrak{b}$ (representing a multiple of $\mathscr{B}$ by a 2-torsion ideal class), so any such choice of $\alpha$ and $\beta$ satisfies conditions so that $\alpha + \beta j \in R'\mathfrak{a}$ with $R' = IRI^{-1} = R^{(\mathfrak{b}I)}$ for some squarefree integral ideal $I|\mathfrak{d}$. Note that $I\overline{I}^{-1} = (1)$, so replacing $\mathfrak{b}$ with $\mathfrak{b}I$ in the definition of $\mathfrak{c}'$ in terms of $(\beta)$ (or $(\beta)$ in terms of $\mathfrak{c}'$) has no impact.

Now we are ready to carry out the calculation of the left side of (A.7). Although the outer sum runs over $\mathscr{O}_K$-algebra isomorphism class of Eichler orders $R$ of reduced discriminant $Np$ (containing $\mathscr{O}_K$) inside of $B$, rather than considering contributions from suitable elements $b = \alpha + \beta j$ which lie in $R^{(\mathfrak{b})}\mathfrak{a}$ for each of a list of class group representatives $\mathfrak{b}$, it is more convenient to consider a coset $\rho$ in $\mathrm{Cl}_K/\mathrm{Cl}_K[2] \simeq \mathrm{Cl}_K^2$, and determine *only* the subsummation within (A.7) coming from those ideal classes $\mathscr{B}$ in the chosen coset. These classes are precisely those whose square $\mathscr{B}^2$ is some specified square class $\mathscr{C} \in \mathrm{Cl}_K^2$, and if we choose a "base point" representative $\mathfrak{b}$ for one such $\mathscr{B}$, then $\{\mathfrak{b}I\}$ provides a 2-to-1 set of representatives of elements in the coset $\rho$ as $I$ runs over squarefree divisors of the principal ideal $\mathfrak{d}$. Recall that $\mathrm{Cl}_K[2]$ is presented as the free $\mathbf{F}_2$-vector space on the prime factors of $\mathfrak{d}$, subject to one relation: when $D_2 \neq 4$ we require that the sum of these elements is zero, and when $D_2 = 4$ (so $D = -4d$ with positive $d \equiv 1 \bmod 4$) we require the vanishing of the sum over the factors in odd residue characteristic when $d > 1$ (and the class group is trivial for $d = 1$). Though we are interested in the subsummation of (A.7) running over the ideal classes lying in the 2-torsion coset $\rho$ (or having square $\mathscr{C}$), we shall consider instead the analogous sum $\Sigma'_{\mathscr{C}}$ carried out over all concrete orders $R^{(\mathfrak{b}I)}$ as $I$ runs through squarefree factors of $\mathfrak{d}$. This collection of orders indexed by $I$'s collapses 2-to-1 when considering $\mathscr{O}_K$-algebra isomorphism classes, so at the end we will need to divide by 2 (it would be inconvenient to explicitly avoid the double overcount caused by isomorphism class repetition).

Fix data consisting of nonzero integral ideals $\mathfrak{c}$ and $\mathfrak{c}'$ with $\mathfrak{c} \in \mathscr{A}^{-1}$ and $\mathfrak{c}' \in \mathscr{A}\mathscr{C}^{-1}[\mathfrak{q}\mathfrak{n}^{-1}]$ with $\mathrm{N}(\mathfrak{c}) + Np\mathrm{N}(\mathfrak{c}') = m|D|$. The ideals in (A.10) are unaffected by replacing $\mathfrak{b}$ with $\mathfrak{b}I$ for squarefree divisors $I$ of $\mathfrak{d}$. Thus, for $\Sigma'_{\mathscr{C}}$ viewed as a sum of sums labelled by orders $R^{(\mathfrak{b}I)}$, it makes sense to focus on the contribution of elements $b = \alpha + \beta j \in R^{(\mathfrak{b}I)}\mathfrak{a}/\pm 1$ for which $\alpha$ and $\beta$ are respective generators of the ideals in (A.10). Our collection of $2^t$ Eichler orders $R^{(\mathfrak{b}I)}$, which is 2-to-1 on the level of $\mathscr{O}_K$-algebra isomorphism classes, is naturally indexed by the $2^t$ elements $X \in \mathscr{O}_K/\mathfrak{d}$ satisfying $\mathrm{N}(X) \equiv -pq \bmod D$:

$$\{\alpha + \beta j \mid \alpha \in \mathfrak{d}^{-1},\ \beta \in \mathfrak{d}^{-1}\mathfrak{q}^{-1}\mathfrak{n}\mathfrak{b}\overline{\mathfrak{b}}^{-1},\ \alpha - X\beta \in \mathscr{O}_{\mathfrak{d}}\}.$$

Define $n = p\mathrm{N}(\mathfrak{c}')$. Each choice of generators $\alpha$ and $\beta$ for the principal ideals in (A.10) contributes to $\Sigma'_{\mathscr{C}}$, but may contribute multiple times, via membership of $b = \alpha + \beta j$ in perhaps more than one of the $2^t$ orders with which we are working. The issue comes down to the fact that for each $\ell | D$, the local congruence $\alpha \overset{?}{\equiv} X'_\ell \beta \bmod \mathscr{O}_{K,\ell}$ might be satisfied for both solutions $X'_\ell \in \mathscr{O}_K/\mathfrak{d}_\ell$ to $\mathrm{N}(X'_\ell) \equiv -pq \bmod D_\ell$. By Lemma A.18, we see that the $\ell | D$ for which both congruences hold are exactly those for which $\ell | \mathrm{N}(\mathfrak{c}') = n/p$, or equivalently $\ell | \gcd(n, D)$. If $e(n)$ denotes the number of prime factors of $\gcd(n, D)$, then we see that such terms $\alpha + \beta j$ contribute $1 + \mathrm{ord}_p(\mathrm{N}(b_-)) = \mathrm{ord}_p(p\mathrm{N}(\beta j))$ exactly $2^{e(n)} = \delta(n)$ times.

Thus, for $\mathfrak{c} \in \mathscr{A}^{-1}$ and $\mathfrak{c}' \in \mathscr{A}\mathscr{C}^{-1}[\mathfrak{q}\mathfrak{n}^{-1}]$ we have $4u^2$ choices of pairs $(\alpha, \beta)$ generating the principal ideals in (A.10), and each such pair gives rise to an element $b = \alpha + \beta j$ which lies in $R'\mathfrak{a}$ for $\delta(n)$ of the orders $R'$ which arise in $\Sigma'_{\mathscr{C}}$. Since we are really summing over $R'\mathfrak{a}/\pm 1$, where $\pm b$ are the same, we have to divide by 2. Thus, the amount contributed to $\Sigma'_{\mathscr{C}}$ by the $\delta(n)$ appearances of $b$ is $(1 + \mathrm{ord}_p(\mathrm{N}(\beta j)))/2$. We have $\mathrm{N}(\beta j) = \mathrm{N}(\beta) \cdot (-pq)$, and the $p$-part of $\mathrm{N}(\beta)$ agrees with that of $n/p = \mathrm{N}(\mathfrak{c}')$, due to (A.9) and the fact that $p$ does not divide $\mathrm{N}(\mathfrak{a}) \cdot qND$. Thus, for each pair $(\mathfrak{c}, \mathfrak{c}')$ we get a total contribution to $\Sigma'_{\mathscr{C}}$ given by $(4u^2)\delta(n) \cdot \mathrm{ord}_p(pn)/2$. Remembering now to divide by an additional factor of 2 from the initial double overcount on isomorphism classes of models (upon fixing $\mathscr{C}$), the pair $(\mathfrak{c}, \mathfrak{c}')$ contributes $u^2\delta(n)\mathrm{ord}_p(pn)$, where $n = p\mathrm{N}(\mathfrak{c}')$ is an integer divisible by $p$, and $Nn < m|D|$. This only depends on $(\mathfrak{c}, \mathfrak{c}')$ through the value of $n$. Since $\mathfrak{c}$ has norm $m|D| - nN$, the total number of such pairs $(\mathfrak{c}, \mathfrak{c}')$ is $r_{\mathscr{A}}(m|D| - nN)r_{\mathscr{A}\mathscr{C}^{-1}[\mathfrak{q}\mathfrak{n}^{-1}]}(n/p)$. For the first factor we have used that $r_{\mathscr{A}^{-1}} = r_{\mathscr{A}}$ since conjugation of an ideal inverts the ideal class but fixes the norm.

As we vary the square class $\mathscr{C}$, $\mathscr{A}\mathscr{C}^{-1}[\mathfrak{q}\mathfrak{n}^{-1}]$ varies over all products of $\mathscr{A}[\mathfrak{q}\mathfrak{n}]$ against a square ideal class. The number of integral ideals of norm $n/p$ which lie in such classes is precisely what is counted by $R_{\mathscr{A}[\mathfrak{q}\mathfrak{n}]}(n/p)$. This completes the proof. ∎

*Remark* A.19. Another way to interpret the double overcounting is that we should identify contributions from $\alpha + \beta j$ and $\alpha - \beta j$, since $\alpha - \beta j = i(\alpha + \beta j)i^{-1}$ and it is precisely conjugation by the generator $i = \sqrt{D}$ of $\mathfrak{d}$ which serves to identify repetitions within an $\mathscr{O}_K$-algebra isomorphism class in the collection of orders used in the preceding proof (relative to a chosen coset $\rho \in \mathrm{Cl}_K/\mathrm{Cl}_K[2]$).

In the ramified case, the final formula in [GZ, Ch III, §9] is valid without parity restriction on $D$, but some care is required to handle even $D$. The essential deviations from the proof of Theorem A.17 are detailed below.

**Theorem A.20.** *Suppose $p = \mathfrak{p}^2$ is ramified in $K$. Pick a prime $\mathfrak{q}$ over $q$ as above, and define $\delta(n) = 2^{e(n)}$ where $e(n)$ is the number of prime factors of* $\gcd(n, D)$. *Then*

$$\langle c, T_m d^\sigma \rangle_p = -r_{\mathscr{A}}(m)h_K u\mathrm{ord}_p(m)\log p - \log p \cdot u^2 \cdot \sum_{\substack{0 < n < m|D|/N \\ p|n}} \mathrm{ord}_p(n) r_{\mathscr{A}}(m|D| - nN)\delta(n)R_{\mathscr{A}[\mathfrak{q}\mathfrak{p}\mathfrak{n}]}(n/p).$$

*Proof.* The first essential difference from the inert case is that the places $v|\mathfrak{p}$ are generally no longer in a bijective correspondence with the elements of $\mathrm{Cl}_K$. Indeed, if $[\mathfrak{p}]$ has order 2 in $\mathrm{Cl}_K$ then $f_v = 2$ for all $v|\mathfrak{p}$. In this case, there are only $h/2$ isomorphism classes of Eichler orders in $B$ with reduced discriminant $Np$ containing $\mathscr{O}_K$: if we fix $S = S_X$ as in (A.6), then $\mathfrak{p}S\mathfrak{p}^{-1} = S$. This is a minor change, as summing over the places $v$ will instead correspond to summing over the orders $R^{(\mathfrak{b})}$ as defined previously, where $\mathfrak{b}$ ranges over a representative set of ideals for $\mathrm{Cl}_K/[\mathfrak{p}]$.

As in the inert case, we will perform the sum over ideal classes by summing on the outside over cosets in $\mathrm{Cl}_K^2$ and on the inside over those ideal classes in the coset *modulo* $[\mathfrak{p}]$ – these can be represented by the squarefree divisors $I$ of $\mathfrak{d}$ that are relatively prime to $\mathfrak{p}$, but not necessarily in a 2-to-1 fashion as before. Losing the order 2 action at $\mathfrak{p}$ has cut the cardinality of this representative set in half, but if $[\mathfrak{p}]$ is trivial, the cardinality of the $\mathrm{Cl}_K[2]$ coset is the same as in the inert case. Thus we can conclude that the representation is 2-to-1 exactly when $f_v = 2$ for all $v|\mathfrak{p}$, and otherwise 1-to-1.

Respecting these changes, the method of summation is the same, but there is a change (relative to what we saw in the inert case) when counting the number of representative models for a given coset of $\mathrm{Cl}_K[2]/[\mathfrak{p}]$ that will contain a choice of $\alpha + \beta j$ corresponding to ideals $(\mathfrak{c}, \mathfrak{c}')$. Previously this count was $\delta(n)$, but now we always have $p|n$, and the resulting local factor of 2 that $p$ contributes to $\delta(n)$ is superfluous, as there is not a corresponding order 2 action at $\mathfrak{p}$ on the representatives.

Along with the observation that $\mathrm{ord}_p(D\mathrm{N}(b_-)) = \mathrm{ord}_p(n)$, the formula in the theorem is assembled in the same manner as in the inert case.

■

## References

[BLR] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Springer-Verlag, Berlin, 1990.

[EGA] J. Dieudonné, A. Grothendieck, *Éléments de géométrie algébrique*, Publ. Math. IHES, **4, 8, 11, 17, 20, 24, 28, 32**, (1960-7).

[Gi] J. Giraud, *Remarque sur une formule de Shimura-Taniyama*, Inventiones Math. **5** (1968), pp. 231–236.

[Gr1] B. Gross, *On canonical and quasi-canonical liftings*, Inventiones Math. **84** (1986), pp. 321–326.

[Gr2] B. Gross, "Local heights on curves" in *Arithmetic geometry* (Cornell/Silverman, etc.), Springer-Verlag, 1986.

[GZ] B. Gross, D. Zagier, *Heegner points and derivatives of L-series*, Inventiones Math. **84** (1986), pp. 225–320.

[K] N. Katz, "Serre-Tate local moduli" in *Surfaces algébriques*, Lecutre Notes in Math 868, Spriger-Verlag, NY, 1981, pp. 138–202.

[KM] N. Katz, B. Mazur, *Arithmetic moduli of elliptic curves*, Princetion University Press, Princeton, New Jersey, 1985.

[L] J. Lipman, *Desingularization of two-dimensional schemes*, Annals of Math **107** (1978), pp. 151–207.

[Mat] H. Matsumura, *Commutative Ring Theory*, Cambridge Univ. Press, 1986.

[GIT] D. Mumford, *Geometric invariant theory*, Ergebnisse 34, Springer Verlag, New York (1965).

[Mum] D. Mumford, *Abelian varieties*, Published for the Tata Institute of Fundamental Research, Bombay, 1970, Tata Institute of Fundamental Research Studies in Mathematics, No. 5.

[ST] J-P. Serre, J. Tate, *Good reduction of abelian varieties*, Annals of Mathematics **88** (1968), pp. 492–517.

[Tate] J. Tate, *Endomorphisms of abelian varieties over finite fields*, Inventiones Math. **2** (1966), pp. 134–144.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
*E-mail address*: `bdconrad@umich.edu`

Department of Mathematics, Brown University, Providence, RI 02912, USA
*E-mail address*: `wrmann@math.brown.edu`