REDUCTIONS OF ABELIAN SURFACES OVER GLOBAL FUNCTION FIELDS

DAVESH MAULIK, ANANTH N. SHANKAR, AND YUNQING TANG

ABSTRACT. Let A be a non-isotrivial ordinary abelian surface over a global function field with good reduction everywhere. Suppose that A does not have real multiplication by any real quadratic field with discriminant a multiple of p. We prove that there are infinitely many places modulo which A is isogenous to the product of two elliptic curves.

CONTENTS

1. INTRODUCTION

1.1. **The main results.** Let p be an odd prime and let $\mathcal{A}_2$ denote the moduli stack of principally polarized abelian surfaces over $\mathbb{F}_p$. We view $\mathcal{A}_2$ as (the special fiber of the canonical integral model of) a GSpin Shimura variety and let $Z(m)$ denote the Heegner divisors in $\mathcal{A}_2$; more precisely, $Z(m)$ parametrizes abelian surfaces with a special endomorphism s such that $s \circ s$ is the endomorphism given by multiplication by m (see §2.2).

Theorem 1. *Assume $p \geq 5$. Let C be an irreducible smooth quasi-projective curve with a finite morphism $C \to \mathcal{A}_{2,\bar{\mathbb{F}}_p}$. Assume that the generic point of C corresponds to an ordinary abelian surface.*

(1) If the image of C is not contained in any Heegner divisor $Z(m)$, and if C is projective, then there exist infinitely many $\bar{\mathbb{F}}_p$-points on C which correspond to non-simple abelian surfaces.
(2) If the image of C is contained in some $Z(m)$ such that $p \nmid m$, then there exist infinitely many $\bar{\mathbb{F}}_p$-points on C which correspond to abelian surfaces isogenous to self-products of elliptic curves

In Theorem 1(2), note that the elliptic curve may vary for these points. An equivalent statement is that there exist infinitely many $\bar{\mathbb{F}}_p$-points on C which correspond to abelian surfaces whose Néron–Severi ranks are strictly larger than that of the generic point of C. Note that in the case (2), any irreducible component of $Z(m) \subset \mathcal{A}_2$ is an irreducible component of a Hecke translate of some

Hilbert modular surface associated to the real quadratic field $F = \mathbb{Q}(\sqrt{m})$ (if m is a square number, then we obtain a Hecke translate of the self-product of the modular curve).

Remark. The assumption that the generic point is ordinary is necessary (especially if we formulate the theorem in terms of the Néron–Severi rank). For instance, in the case (2), we may take C to be an irreducible component of the non-ordinary locus. If p is inert in F, then all the points on C are supersingular and the Néron–Severi rank does not jump. If p is split in F, then the only points where the Néron–Severi rank jumps are the finitely many supersingular points.

Remark. We make the (technical) assumption that C is projective in (1) because the Heegner divisors $Z(m)$ are all non-compact and we plan to remove this assumption in future work. On the other hand, the Hilbert modular surfaces considered in (2) do contain compact special divisors (see the second half of §2.2 for the definitions of special divisors in the Hilbert case, and §4.3.3 for a criterion of when these special divisors are compact) whose $\overline{\mathbb{F}}_p$ points parameterize abelian surfaces isogenous to a self-product of elliptic curves. By working exclusively with these compact special divisors, we no longer need assume that C is projective.

Remark 2. A modification of our argument shows that with the same assumption in (1), for a fixed real quadratic number field F, there are infinitely many ordinary $\bar{\mathbb{F}}_p$-points on C such that the corresponding abelian surfaces admit real multiplication by F. Here we need to assume $p \geq 7$ if p is ramified in F. Otherwise, $p \geq 5$ is enough.

To prove Theorem 1(1), we consider the intersection number of C and $Z(\ell^2)$, where ℓ is a varying prime number. If we consider $Z(\ell)$ with $\ell \equiv 3 \bmod 4$ instead, we prove

Theorem 3. *Suppose we have the same assumptions as in Theorem 1(1). Then there are infinitely many ordinary $\bar{\mathbb{F}}_p$-points on C such that, for each of these points, the corresponding abelian surface admits real multiplication by the ring of integers of some real quadratic field (note that the quadratic fields may vary for these points).*

It would be interesting to find $\bar{\mathbb{F}}_p$-points of complex multiplication by maximal orders, but our current method only asserts real multiplication by maximal orders.

1.2. **Previous work and heuristics.** Theorem 1 is a generalization of [CO06, Proposition 7.3], where Chai and Oort proved Theorem 1(2) with $\mathcal{A}_1 \times \mathcal{A}_1$ taking the place of a Hilbert modular surface. Their proof crucially uses the product structure of the Shimura variety, as well as the product structure of the Frobenius morphism. Following the discussion in §7 of [CO06], Theorem 1 is related to a bi-algebraicity conjecture. See §1.4 for more details.

We offer the following heuristic for Theorem 1(1). Using Honda and Tate's classification of $\mathbb{F}_q^n$-isogeny classes of abelian varieties in terms of Weil-q^n numbers, the number of $\mathbb{F}_q$-isogeny classes of abelian varieties is seen to equal $q^{n(3/2+o(1))}$. Similarly, the number of *split* $\mathbb{F}_{q^n}$-isogeny classes in $\mathcal{A}_2$ is seen to equal $q^{n(1+o(1))}$. If we treat the map from $C(\mathbb{F}_{q^n})$ to the set of $\mathbb{F}_{q^n}$-isogeny classes as a random map, we expect that the number of $\mathbb{F}_{q^n}$ points of C which are not simple is around $q^{n/2(1+o(1))}$. Letting n approach infinity, this heuristic suggests that infinitely many points of $C(\overline{\mathbb{F}}_q)$ that are split. There are analogous questions in other settings. For the case of equicharacteristic 0, these results are well known (for instance, the density of Noether–Lefschetz loci is discussed in [Voi02]). In mixed characteristic, the analogue of Theorem 1(2) is treated in [Cha18], [ST17]. The major difference between Theorem 1 and these other cases is that the ordinary generic point assumption is crucial since the result is simply false otherwise (as remarked in §1.1).

Indeed, this difference hints at the key difficulty in our setting, which is that the local intersection number at a supersingular point is of the same magnitude as the total intersection number, which makes the approach more complicated than that of [ST17]; we discuss this in more detail in §1.3.

1.3. **Proof of the main results.** We view both Hilbert modular surfaces and the Siegel three-fold as GSpin Shimura varieties attached to a quadratic space (V, Q). In each setting, we have a notion of special endomorphisms and special divisors and, for simplicity, we use the same notation $Z(m)$.

The main idea of the proof is to compare the global and local intersection numbers of $C.Z(m)$[1] for appropriate sequences of m and show it is not possible for finitely many points to account for the total global intersection as m increases.

More precisely,

(1) The global intersection number $I(m) := C.Z(m)$ is controlled by Borcherds theory [Bor98] (see also [Mau14] and [HMP]).
(2) We prove that as $m \to \infty$, the total local contribution from supersingular points is at most $\frac{11}{12}I(m)$ by studying special endomorphisms.[2]
(3) We prove that the local contribution from a non-supersingular point is $o(I(m))$ as $m \to \infty$.

This allows us to conclude that, as $m \to \infty$, more and more points of C contribute to the intersection $C.Z(m)$. In order to prove Theorem 1(1), the sequence of m will consist only of squares, and in order to prove Theorem 3, the sequence will consist only of primes. Note that in $\mathcal{A}_2$, the Heegner divisor $Z(m)$ for square m parametrizes abelian surfaces which are not geometrically simple, thereby allowing us to deduce Theorem 1(1). Similar arguments allow us to deduce part Theorem 1(2), and also Theorem 3.

Compared to the number field situation, the main difficulty of the positive characteristic function field case is that the local contributions at supersingular points are of the same magnitude as the global contribution. More precisely, taking the Hilbert case as an example, Borcherds theory implies that the generating series of $Z(m)$ is a non-cuspidal modular form of weight 2; on the other hand, the theta series attached to the special endomorphism lattice at a supersingular point is also a non-cuspidal weight 2 modular form since the lattice is of rank 4. Therefore, even without considering higher intersection multiplicities, the local intersection number of $C.Z(m)$ at a supersingular point is also of the same magnitude as the growth rate of Fourier coefficients of an Eisenstein series of weight 2.

Bounding the local contribution from a supersingular point. Let $A \to C$ denote the family of principally polarized abelian surfaces induced from a morphism $C \to \mathcal{A}_{2,\overline{\mathbb{F}}_p}$, and let $\operatorname{Spf} \overline{\mathbb{F}}_p[[t]] \to C$ denote the formal neighborhood of a supersingular point. For a special endomorphism s such that $s \circ s = m$, we say that s is of norm m.

The local contribution to $C.Z(m)$ from this supersingular point equals $\sum_{n=0}^{\infty} r_n(m)$, where $r_n(m)$ is the number of special endomorphisms of $A \mod t^{n+1}$ with norm m. Therefore, in order to bound the local contribution, it suffices to prove that, as $n \to \infty$, there are many special endomorphisms of $A \mod t^n$ which decay rapidly enough (see Definition 5.1.1, Theorem 5.1.2, Theorem A.0.1 for precise statements).

A similar decay result appears in the mixed characteristic setting (see [ST17]), by a straightforward application of Grothendieck–Messing theory. In the equicharacteristic case, however, proving our decay results is much more involved. In particular, we need to use Kisin's description [Kis10, §1.4, 1.5] of the F-crystal associated to a certain automorphic vector bundle $\mathbb{L}_{\mathrm{cris}}$, whose F-invariant part is the lattice of special endomorphisms, in order to prove the required decay. See §3.1.5 and the proof of Theorem 5.1.2 for more details.

We will focus on the Siegel case from now on. Let L_0 denote the lattice of special endomorphisms of $A \mod t$, and let $L_n \subset L_0$ be the lattice of special endomorphisms of $A \mod t^{n+1}$. These

[1] Although C is not a substack of $\mathcal{A}_2$, we may define $C.Z(m)$ as the degree of the pull back of $Z(m)$ via $C \to \mathcal{A}_{2,\bar{\mathbb{F}}_p}$ when C is projective.

[2] Indeed, the ratio depends on p and it goes to $1/2$ as $p \to \infty$.

lattices are of rank 5 and are equipped with natural quadratic forms such that $A \mod t^{n+1}$ admits a special endomorphism of norm m if and only if m is represented by L_n. Broadly speaking, we can bound the local contribution by using geometry-of-numbers techniques. To obtain the desired estimate, we choose the sequence m as follows. We first prove the existence of a rank 2 sublattice $P_n \subset L_n$ that has the following property: for all m bounded by an appropriate function of n, the abelian surface $A \mod t^{n+1}$ has a special endomorphism of norm m only if the quadratic form restricted to P_n represents m. This fact follows from the existence of a rank 3 submodule of special endomorphisms which decay rapidly (Theorems 5.1.2 and A.0.1). Furthermore, the discriminant of P_n goes to infinity as $n \to \infty$. Therefore, the density of numbers (or primes, or prime-squares) represented by the *binary* quadratic form P_n approaches zero, as $n \to \infty$. We now pick a sequence of prime-squares m none of which are represented by P_n defined by the finitely many supersingular points on C.

The non-ordinary locus is singular at superspecial points. This allows us to prove the existence of a special endomorphism that decays "more rapidly than expected" (see Definition 5.1.1(3)). Consequently, by the explicit formula of Eisenstein series in these cases by [BK01], we prove that the sum of local contributions at supersingular points is at most 11/12 of the global contribution.

We remark that our proof is more involved than the proof of [CO06, Proposition 7.3] because the intersection theory on Hilbert modular surfaces and Siegel three-folds is more complicated than that on the product of j-lines.

1.4. **Additional remarks.** The key difference between the number field and function field situation is the following. Let A be an abelian surface over $\mathcal{O}_K$, where K is a local field. The $\mathbb{Z}_p$-module of special endomorphisms of $A[p^\infty]$ has rank ≤ 3. This rank equals three if and only if A can be realized as the limit point (in the analytic topology) of a sequence of CM points. This can happen in the mixed characteristic case, but not in the equicharacteristic p case unless A is defined over a finite field.[3] Thus, we have a rank 3 decay in the Decay Lemmas (Theorems 5.1.2 and A.0.1).

In the setting of higher dimensional GSpin Shimura varieties, for the same reason, we expect that generalizations of the Decay Lemma will only yield a rank 3 $\mathbb{Z}_p$-module that decays rapidly. This has the consequence of the existence of formal curves, such that the module of special endomorphisms of the p-divisible group over these formal curves have large rank. An interesting bi-algebraicity question is whether such formal curves can be algebraic without being special. In the ordinary case, Chai has the following conjecture:

Conjecture 4 ([Cha03, Conj. 7.2, Remark 7.2.1, Prop. 5.3, Remark 5.3.1])**.** *Let X be a subvariety in a mod p Shimura variety passing through an ordinary point P. Assume that the formal germ of X at P is a formal torus in the Serre–Tate coordinates. Then X is a Shimura subvariety.*

1.5. **Organization of paper.** In §2, we recall the notion of special endomorphisms, special divisors and crystalline realization $\mathbb{L}_{\text{cris}}$ of the automorphic vector bundle of special endomorphisms. In §3, we recall the lattices of special endomorphisms of a supersingular point and compute $\mathbb{L}_{\text{cris}}$ on its deformation space. In §4, we recall Borcherds theory and the explicit formula for the Fourier coefficients of vector-valued Eisenstein series due to Bruinier–Kuss; we use them to compare the global intersection number and the mod t local intersection number at a supersingular point. Sections §5 and §6 are the key technical part of the paper. We prove the decay theorems for special endomorphisms, which we will use to bound the higher local intersection multiplicities at supersingular points. Section §7 provides the outline of the main proofs and by geometry-of-numbers arguments, we prove Theorem 1(2) in §8 and prove Theorem 1(1) and Theorem 3 in §9.

In order to get the main idea of the proof, the reader may focus on Theorem 1(2) and start from §§7,8 and refer back to §§3-5 when necessary.

[3]Ordinary abelian varieties which have CM are defined over finite fields.

1.6. **Notation.** We write $f \asymp g$ if $f = O(g), g = O(f)$. Throughout the paper, p is an odd prime.

Acknowledgement. We thank Johan de Jong, Keerthi Madapusi Pera, Arul Shankar, Salim Tayou, and Jacob Tsimerman for helpful discussions. D.M. is partially supported by NSF FRG grant DMS-1159265. Y.T. is partially supported by the NSF grant DMS-1801237. We would like to thank the referee for valuable suggestions on the exposition of the paper.

2. Special endomorphisms

In this section, we first introduce quadratic lattices (L, Q) such that the associated GSpin Shimura varieties will be $\mathcal{A}_2$ and certain Hilbert modular surfaces related to the Heegner divisors $Z(m)$. The definition of special endomorphisms and Heegner divisors are given in §2.2.

2.1. **The global lattice L.** For a quadratic $\mathbb{Z}$-lattice (L, Q), let $C(L)$ (resp. $C^+(L)$) denote the (resp. even) Clifford algebra of L. Let $(-)'$ denote the standard involution on $C(L)$ fixing all elements in L given by $(v_1 \cdots v_n)' = v_n \cdots v_1$ for $v_i \in L$. Let V denote $L \otimes \mathbb{Q}$ endowed with the quadratic form Q. There is a bilinear form $[-,-]$ on V given by $[x, y] := Q(x + y) - Q(x) - Q(y)$.

Let L_S be the rank 5 $\mathbb{Z}$-lattice endowed with the quadratic form $Q(x) = x_0^2 + x_1x_2 - x_3x_4$ for $x = (x_0, \cdots, x_4) \in \mathbb{Z}^5$. This quadratic form has signature $(3, 2)$ and L_S is an even lattice, maximal among $\mathbb{Z}$-valued sublattices in $L_S \otimes \mathbb{Q}$. For $p > 2$, L_S is self-dual at p. A direct computation shows that $C^+(L_S) \cong M_4(\mathbb{Z})$. Let

$$v_0 = (1, 0, 0, 0, 0), v_1 = (0, 1, 1, 0, 0), v_2 = (0, 1, -1, 0, 0), v_3 = (0, 0, 0, 1, 1), v_4 = (0, 0, 0, 1, -1).$$

Then $\delta := v_0 \cdots v_4 \in C(L_S)$ lies in the center of $C(L_S)$ and $\delta' = \delta, \delta^2 = 1$. Therefore, there is an isomorphism between quadratic spaces given by $L_S \xrightarrow{\simeq} \delta L_S \subset C^+(L_S)$. (See for instance [KR00, App. A].)

Given a vector $x \in L_S$ such that $Q(x) = m, m \in \mathbb{Z}_{>0}$, the orthogonal complement $x^\perp \subset L_S$ endowed with the restriction of Q on $x^\perp$ is a quadratic $\mathbb{Z}$-lattice of signature $(2, 2)$ and let $L_H \subset x^\perp \otimes \mathbb{Q}$ be a maximal lattice containing $x^\perp$. If m is not a perfect square, let F denote the real quadratic field $\mathbb{Q}(\sqrt{m})$. A direct computation shows that there is an isomorphism $L_H \otimes \mathbb{Q} \cong \mathbb{Q}^2 \oplus F$ such that $Q((a, b, \gamma)) = ab + \mathrm{Nm}_{F/\mathbb{Q}}\,\gamma$. The assumption $p \nmid m$ and $p > 2$ implies that $x^\perp$ and hence L_H are self-dual at p.

Now let (L, Q) have signature $(n, 2)$, and let p be a prime such (L, Q) is self-dual at p. As in [AGHMP18, §4.1 §4.2], [KR00, §1], there is a GSpin Shimura variety M attached to (L, Q) and this Shimura variety also admits a smooth integral model $\mathcal{M}$ over $\mathbb{Z}_{(p)}$ since L is self-dual at p; the Shimura variety (and its integral model) recovers the moduli space of principally polarized abelian surfaces when $L = L_S$ (see Remark 2.2.2 for details) and it is a Hilbert modular surface when $L = L_H$ (see for instance [HY12, §2.2, §3.1]). We may write M_L and $\mathcal{M}_L$ to emphasis on the dependence on L.

To prove Theorem 1(1) and Theorem 3, we will take $L = L_S$ and to prove Theorem 1(2), we will take $L = L_H$.

2.2. **Special endomorphisms and special divisors.** We first introduce the notion of special endomorphisms when $L = L_S$ and $\mathcal{M}$ is the moduli space of principally polarized abelian surfaces. Given an $\mathcal{M}$-scheme S, let A_S denote the pullback of the universal principally polarized abelian surface on $\mathcal{M}$ via $S \to \mathcal{M}$; let $\dagger$ denote the Rosati involution on A_S.

Definition 2.2.1. A *special endomorphism* of A_S is an element $s \in \mathrm{End}(A_S)$ such that $s^\dagger = s$ and $\mathrm{Tr}\, s = 0$, where Tr is the reduced trace on the semisimple algebra $\mathrm{End}(A_S) \otimes \mathbb{Q}$.

Remark 2.2.2. Our definition of special endomorphisms is essentially the same as the one given by Kudla–Rapoport ([KR00, Def. 2.1, Eqn. (2.21)]). Indeed, as in [KR00, §§1-2], the moduli problem indicates that every $\mathcal{M}$-scheme S gives rise to a principally polarized abelian scheme B_S over S with $\iota : C^+(L) \hookrightarrow \mathrm{End}(B_S)$ and a polarization such that the induced Rosati involution $\dagger$ satisfies $\iota(c)^\dagger = \iota(c^T)$, where $(-)^T$ is the transpose on $C^+(L) \simeq M_4(\mathbb{Z})$ (see condition (iii) and the first paragraph of [KR00, p.701]); moreover, for each $\ell \neq p$, there is an isomorphism $C^+(L) \otimes \mathbb{Z}_\ell \simeq T_\ell(B_S)$, where T_ℓ denotes the ℓ-adic Tate module, compatible with the $C^+(L)$-action (it acts on itself via left multiplication; see [KR00, p.703]).[4] Therefore, via ι, we have $B_S \cong A_S^4$, where A_S is an abelian surface and by the compatibility of the polarization with ι (see also [KR00, Eqn. (1.9), (1.10)]), and the polarization on B_S is induced by the self-product of a principal polarization on A_S. Hence $\mathcal{M}$ parameterizes principally polarized abelian surfaces. Moreover, an element s_B in $\mathrm{End}(B_S) \cong M_4(\mathrm{End}(A_S))$ commuting with $\iota(C^+(L))$ is of form $\mathrm{diag}(s, s, s, s)$, where an endomorphism s of A_S. In the sense of Kudla–Rapoport, such s_B is special if and only if it is traceless and fixed by the Rosati involution on B_S; this is equivalent to that s is traceless and fixed by the Rosati involution on A_S. Therefore, our definition is the same as that of Kudla–Rapoport.

Definition 2.2.3. Let $\mathbb{D}$ denote the Dieudonné crystal over $\mathcal{M}_{\mathbb{F}_p}$ (i.e., the first relative crystalline homology of the universal family of principally polarized abelian surface over $\mathcal{M}_{\mathbb{F}_p}$). Let $\mathbb{L}_{\mathrm{cris}} \subset \mathrm{End}(\mathbb{D})$ denote the sub-crystal of trace 0 elements fixed by the Rosati involution.[5]

By definition, when S is a $\mathcal{M}_{\mathbb{F}_p}$-scheme, an element $s \in \mathrm{End}(A_S)$ is a special endomorphism if and only if the crystalline realization of $s \in \mathrm{End}(\mathbb{D}_S)$ lies in $\mathbb{L}_{\mathrm{cris},S}$.

Definition 2.2.4. For the p-divisible group $A_S[p^\infty]$, we say $s \in \mathrm{End}(A_S[p^\infty])$ is a *special endomorphism* if the image of s in $\mathrm{End}(\mathbb{D}_S)$ lies in $\mathbb{L}_{\mathrm{cris},S}$.

Remark 2.2.5. In [MP16, §4.14], there is a definition of $\mathbb{L}_{\mathrm{cris}}$ as a direct summand of the endomorphism of the first relative crystalline cohomology of the Kuga–Satake abelian scheme over $\mathcal{M}_{\mathbb{F}_p}$. More precisely, the left multiplication of $\mathrm{GSpin}(V, Q) \subset C^+(V)^\times$ acting on $C(V)$ induces a variation of Hodge structures on $C(V)$ over M; this gives rise to the Kuga–Satake abelian scheme A^{KS} over M and the Kuga–Satake abelian scheme extends over $\mathcal{M}$. The 8-dimensional abelian scheme considered by Kudla–Rapoport is a sub abelian scheme of A^{KS} via the natural embedding $C^+(V) \subset C(V)$. (Note that in [KR00], $\gamma \in \mathrm{GSpin}(V, Q)$ acts on $C^+(V)$ by the right multiplication by γ and $C^+(V)$ acts on $C^+(V)$ by left multiplication, which is opposite to the convention in [MP16]. This difference is due to the different choices of the symplectic pairing on $C^+(V)$ and $C(V)$ in [KR00, (1.9)] and [MP16, §1.6]. If we use the symplectic pairing in [MP16] for the discussion in [KR00], then we obtain similar results as in [KR00] but with the convention consistent with that in [MP16].)

Let $\mathbb{D}^{\mathrm{KS}}$ denote the Dieudonné crystal of A^{KS} over $\mathcal{M}_{\mathbb{F}_p}$; Madapusi Pera defined $\mathbb{L}_{\mathrm{cris}} \subset \mathrm{End}(\mathbb{D}^{\mathrm{KS}})$ by the crystalline realization of the absolute Hodge cycle induced by the $\mathrm{GSpin}(V, Q)$-invariant idempotent which realizes $V \subset \mathrm{End}(C(V))$ as a direct summand. Since the element δ given in §2.1 lies in the center of $C(L)$, then it induces an isomorphism $\mathrm{End}(C(L)) \supset L \cong \delta L \subset \mathrm{End}(C^+(L))$ compatible with $\mathrm{GSpin}(V, Q)$-action. Therefore, δ induces an isomorphism between the crystals $\mathbb{L}_{\mathrm{cris}}$ in our sense and the one in the sense of Madapusi Pera; in particular, the notions of special endomorphisms coincide under the identification via δ. Also, for a special endomorphism s in both cases, $s \circ s$ is a scalar multiple $Q(s)$ on the suitable abelian scheme; since $\delta^2 = 1$, hence $Q(s)$ remains

[4] Although Kudla–Rapoport uses abelian schemes up to isogeny to give the moduli interpretation, one may translate it into abelian schemes up to isomorphism; see also [AGHMP17, §2.2].

[5] Note that Frobenius is not an endomorphism on $\mathrm{End}(\mathbb{D})$, due to the existence of negative slopes. However, we will abuse terminology, and still treat $\mathrm{End}(\mathbb{D})$ and $\mathbb{L}_{\mathrm{cris}}$ as F-crystals in the sense that we will view Frobenius as a map from $\mathrm{End}(\mathbb{D})$ to $\mathrm{End}(\mathbb{D})[1/p]$, while remembering the integral structure, and similarly for $\mathbb{L}_{\mathrm{cris}}$.

the same for images of s under various identification of special endomorphisms. By [MP16, Lem. 5.2], $Q(s) > 0$ for all nonzero special endomorphism s.

Definition 2.2.6. For $m \in \mathbb{Z}_{>0}$, the *special divisor* $\mathcal{Z}(m)$ is the Deligne–Mumford stack over $\mathcal{M}$ with functor of points $\mathcal{Z}(m)(S) = \{s \in \mathrm{End}(A_S) \text{ special } | Q(s) = m\}$ for any $\mathcal{M}$-scheme S. We use the same notation for the image of $\mathcal{Z}(m)$ in $\mathcal{M}$. By for instance [AGHMP18, Prop. 4.5.8], $\mathcal{Z}(m)$ is flat over $\mathbb{Z}_{(p)}$ and hence $\mathcal{Z}(m)_{\mathbb{F}_p}$ is still a divisor of $\mathcal{M}_{\mathbb{F}_p}$; we denote $\mathcal{Z}(m)_{\mathbb{F}_p}$ by $Z(m)$.

Lemma 2.2.7. *Every $\bar{\mathbb{F}}_p$-point of $Z(m^2)$ corresponds to a geometrically non-simple abelian surface.*

Proof. Let s be a special endomorphism of an abelian surface A such that $s \circ s = [m^2]$. Then $(s - [m]) \circ (s + [m]) = 0$. Since $\mathrm{Tr}\, s = 0$, then $s \pm [m] \neq 0$ and hence $s \pm [m]$ are not invertible. Then $\ker(s - [m])$ defines a non-trivial sub abelian scheme of A. □

We now discuss the case when $L = L_H$. We keep the same notation as in §2.1. For simplicity, we first discuss the case when $L_H = x^\perp$, where $x \in L_S$ and $Q(x) = m$ with $p \nmid m$; for the general case, the following discussion still holds true when replacing endomorphisms with suitable elements in $\mathrm{End} \otimes \mathbb{Q}$ (see the end of this subsection). When $L_H = x^\perp \subset L_S$, the Shimura variety (and its integral model) $\mathcal{M}_{L_H}$ defined by L_H is naturally a sub-Shimura variety of $\mathcal{M}_{L_S}$, the moduli space of principally polarized abelian surfaces, and hence a point on $\mathcal{M}_{L_H}$ corresponds to a polarized abelian surface with real multiplication by $\mathcal{O} := \mathbb{Z}[x]/(x^2 - m)$. Let σ denote the ring automorphism on $\mathcal{O}$ satisfying $x^\sigma = -x$. As before, let S be a $\mathcal{M}_{L_H}$-scheme, and let A_S denote the abelian surface over S with real multiplication by $\mathcal{O}$.

Definition 2.2.8 ([HY12, §3.1 p.26])**.** A *special endomorphism* (resp. *special quasi-endomorphism*) of A_S is an element $s \in \mathrm{End}(A_S)$ (resp. $s \in \mathrm{End}(A_S) \otimes \mathbb{Q}$) such that $s^\dagger = s$ and $s \circ f = f^\sigma \circ s$ for all $f \in \mathcal{O}$.

We still use $\mathbb{D}$ to denote the pullback to $\mathcal{M}_{L_H,\mathbb{F}_p}$ the Dieudonné crystal over $\mathcal{M}_{L_S,\mathbb{F}_p}$ in Definition 2.2.3; since the abelian surfaces over $\mathcal{M}_{L_H}$ admit an $\mathcal{O}$-action, the Dieudonné crystal $\mathbb{D}$ is also endowed with an $\mathcal{O}$-action.

Definition 2.2.9. Let $\mathbb{L}_{\mathrm{cris}} \subset \mathrm{End}(\mathbb{D})$ denote the sub-crystal of elements v fixed by Rosati involution and $s \circ f = f^\sigma \circ s$ for all $f \in \mathcal{O}$. For the p-divisible group $A_S[p^\infty]$, we say $s \in \mathrm{End}(A_S[p^\infty])$ is a *special endomorphism* if the image of s in $\mathrm{End}(\mathbb{D}_S)$ lies in $\mathbb{L}_{\mathrm{cris},S}$.

Remark 2.2.10. By Remark 2.2.5 and [AGHMP17, Prop. 2.5.1, Prop. 2.6.4], in order to show that the above definitions of special endomorphisms and $\mathbb{L}_{\mathrm{cris}}$ can be identified with those by Madapusi Pera, we only need to show that for an endomorphism s (of either the abelian surface or of its Dieudonné crystal $\mathbb{D}$) fixed by the Rosati involution is traceless and orthogonal to x if and only if $s \circ x = -x \circ s$. To see this, note that if $\mathrm{Tr}\, s = 0$, then $s \perp x$ if and only if $Q(s+x) - Q(s) - Q(x) = s \circ x + x \circ s = 0$; on the other hand, if $s \circ x = -x \circ s$, then $x^{-1} \circ s \circ x = -s$ and hence $\mathrm{Tr}\, s = 0$.

2.2.11. In general (i.e. when $x^\perp \subsetneq L_H$), we may still use the same definition for $\mathbb{L}_{\mathrm{cris}}$ and special endomorphisms of p-divisible groups, as $x^\perp$ is self-dual at p and hence $x^\perp \otimes \mathbb{Z}_p = L_H \otimes \mathbb{Z}_p$. On the other hand, we consider *special quasi-endomorphisms* $s \in \mathrm{End}(A_S) \otimes \mathbb{Q}$ which satisfy the following integrality condition: the ℓ-adic realizations of s lie in $L_H \otimes \mathbb{Z}_\ell \subset \mathrm{End}(T_\ell(A_S) \otimes \mathbb{Q}_\ell)$ for all $\ell \neq p$ and the crystalline realizations of s lie in $\mathbb{L}_{\mathrm{cris},S}$. As in Definition 2.2.6, the *special divisor* $\mathcal{Z}(m)$ is the Deligne–Mumford stack over $\mathcal{M}_{L_H}$ with $\mathcal{Z}(m)(S)$ given by

$$\{s \in \mathrm{End}(A_S) \otimes \mathbb{Q} \text{ special quasi-endomorphism satisfying the integrality condition above } \mid Q(s) = m\}$$

for any $\mathcal{M}$-scheme S. By the proof of [AGHMP18, Prop. 4.5.8], where they used [MP16, Prop. 5.21], $\mathcal{Z}(m)$ is flat over $\mathbb{Z}_p$. We use $Z(m)$ to denote the image of $\mathcal{Z}(m)_{\mathbb{F}_p}$ in $\mathcal{M}_{L_H,\mathbb{F}_p}$, which is a divisor in $\mathcal{M}_{L_H,\mathbb{F}_p}$.

2.3. **Lattices of special endomorphisms of supersingular points.** For a fixed supersingular point, let A denote the abelian surface attached to this point.

Definition 2.3.1. Let L'' denote the $\mathbb{Z}$-lattice of special endomorphisms of A (resp. special quasi-endomorphisms when $L = L_H$). Let $L'' \subset L' \subset L'' \otimes \mathbb{Q}$ be a $\mathbb{Z}$-lattice which is maximal at all $\ell \neq p$ and $L'' \otimes \mathbb{Z}_p = L' \otimes \mathbb{Z}_p$. Let Q' denote the natural quadratic form on L' given by $s \circ s = [Q'(s)] \in \mathrm{End}(A) \otimes \mathbb{Q}$. By the positivity of the Rosati involution, Q' is positive definite (see for instance [MP16, Lem. 5.12]).

Even though there seem to be choices involved here, we will see that for our computation, these choices do not matter and the result will only depend on the Ekedahl–Oort stratum where the supersingular point lies in. The information of $L' \otimes \mathbb{Z}_p$ will be provided in §3.

Lemma 2.3.2. *$(L' \otimes \mathbb{Z}_\ell, Q') \cong (L \otimes \mathbb{Z}_\ell, Q)$ for $\ell \neq p$.*

Proof. Both lattices shall be maximal at ℓ and by [HP17, Rem. 7.2.5], $(L' \otimes \mathbb{Q}_\ell, Q') \cong (L \otimes \mathbb{Q}_\ell, Q)$. Then we conclude by the fact that there is a unique isometry class of $\mathbb{Z}_\ell$-maximal sublattices of a given $\mathbb{Q}_\ell$-quadratic space (see for instance, [HP17, Thm. A.1.2]). □

Remark 2.3.3. Actually, for the case of Hilbert modular surfaces, the essential part of the above lemma is [HY12, Prop. 3.1.3]. For the $\mathcal{A}_2$ case, we can explicitly compute L'' as follows and it is maximal. By [Eke87, Prop. 5.2], for any $\ell \neq p$, there is a unique class (up to $\mathrm{GL}_4(\mathbb{Z}_\ell)$-conjugation) of principal polarizations on the Tate module $T_\ell(A)$. Therefore, to compute $L'' \otimes \mathbb{Z}_\ell$, we may assume that $A = E^2$ and endowed with the product principal polarization, where E is a supersingular elliptic curve. Hence the quadratic form on the lattice L'', which is the trace 0 part of $H^2(A)$, is given by $x_0^2 + \mathrm{Nm}$, where Nm is the quadratic form given by the reduced norm on the quaternion algebra $\mathrm{End}(E)$.

3. The F-crystals $\mathbb{L}_{\mathrm{cris}}$ on local deformation spaces of supersingular points

In this section, we compute the lattices ($L'' \otimes \mathbb{Z}_p$ in Definition 2.3.1) of special endomorphisms of supersingular points with the natural quadratic forms following Howard–Pappas [HP17, §§5-6].[6] In conjunction with [Kis10, §1], we then obtain $\mathbb{L}_{\mathrm{cris}}$ (see Definition 2.2.3 and Definition 2.2.9) on the formal neighborhoods of supersingular points in the Shimura variety $\mathcal{M}$. As a direct consequence, we obtain the local equation of the non-ordinary locus in §3.4. These are the key inputs to §§5-6; in particular, we use the explicit descriptions of this section to prove our decay results.

3.1. **A brief review of the work of Howard–Pappas and Kisin.** Since both [HP17] and [Kis10] apply to GSpin Shimura varieties of any dimension, we first recall their results in the general setting.

Let (V, Q) denote a quadratic $\mathbb{Q}$-vector space of signature $(n, 2)$ and let $L \subset V$ be a maximal even lattice which is self-dual at p. Let $\mathcal{M}$ denote the smooth canonical integral model over $\mathbb{Z}_p$ of the GSpin Shimura variety attached to (L, Q) in [Kis10].

Set $k = \overline{\mathbb{F}}_p, W = W(k), K = W[1/p]$. In this section, we consider a fixed supersingular point $P \in \mathcal{M}(k)$. In the case of abelian surfaces considered in §2 (with $L = L_S$ or L_H), P supersingular means the corresponding abelian surface over P is supersingular. This in turn is equivalent to the action of the crystalline Frobenius φ on $\mathbb{L}_{\mathrm{cris},P}(W)$ being pure, with slope 0. In the general setting, let $\mathbb{D}$ denote the Dieudonné crystal of the universal Kuga–Satake abelian variety over $\mathcal{M}_{\mathbb{F}_p}$ and let $\mathbb{L}_{\mathrm{cris}} \subset \mathrm{End}(\mathbb{D})$ denote the sub crystal corresponding to $L \subset C(L)$ defined in [MP16, §4.14].[7] Let φ denote the crystalline Frobenius on $\mathbb{D}_P(W)$ and $\mathbb{L}_{\mathrm{cris},P}(W)$. Then we say P is *supersingular* if φ acts on $\mathbb{L}_{\mathrm{cris},P}(W)$ with pure slope 0 (see for instance [HP17, Lem. 4.2.4, §7.2.1]).

[6] One may also carry out the computation following Ogus [Ogu79, §3].

[7] Note that in the cases $L = L_H, L_S$ in §2, we still take $\mathbb{D}$ to be the Dieudonné crystal of the universal abelian surfaces, not that of the Kuga–Satake abelian varieties.

By Dieudonné theory, we have $L'' \otimes \mathbb{Z}_p = \mathbb{L}_{\mathrm{cris},P}(W)^{\varphi=1}$. In order to compute $L'' \otimes \mathbb{Z}_p$ and the φ-action on $\mathbb{L}_{\mathrm{cris},P}(W)$, we introduce another free W-module $\mathbb{L}_P^{\#}(W)$ following [HP17, §6.2.1].[8]

Definition 3.1.1. The filtration on $\mathbb{D}_P(W)$ is given by $\mathrm{Fil}^1 \mathbb{D}_P(W) := \varphi^{-1}(p\mathbb{D}_P(W))$. We define $\mathbb{L}_P^{\#}(W) := \{v \in \mathbb{L}_{\mathrm{cris},P}(W) \otimes_W K \mid v \,\mathrm{Fil}^1 \mathbb{D}_P(W) \subset \mathrm{Fil}^1 \mathbb{D}_P(W)\}$.

3.1.2. By [HP17, Thm. 7.2.4], studying supersingular points and their formal neighborhood in $\mathcal{M}$ reduces to study the points and their formal neighborhood in the associated Rapoport–Zink spaces and hence we use results in [HP17, §§5-6].

By [HP17, Prop. 6.2.2], $\varphi(\mathbb{L}_P^{\#}(W)) = \mathbb{L}_{\mathrm{cris},P}(W)$. In particular,

$$L'' \otimes \mathbb{Z}_p = \mathbb{L}_{\mathrm{cris},P}(W)^{\varphi=1} = \mathbb{L}_P^{\#}(W)^{\varphi=1}.$$

Recall that in Definition 2.3.1, we endow $V' := L'' \otimes \mathbb{Q}_p$ with a quadratic form Q'; let $[-,-]'$ denote the bilinear form on V' given by $[x, y]' = Q'(x + y) - Q'(x) - Q'(y)$. Hence

$$V' = (\mathbb{L}_{\mathrm{cris},P}(W) \otimes_W K)^{\varphi=1}.$$

Since P is supersingular, we have $n = \mathrm{rk}_W \mathbb{L}_{\mathrm{cris},P}(W) = \mathrm{rk}_{\mathbb{Z}_p} L'' = \dim_{\mathbb{Q}_p} V'$.

Let $\Lambda_P \subset V'$ denote the dual of $L'' \otimes \mathbb{Z}_p$ with respect to $[-,-]'$. Then by [HP17, Prop. 6.2.2], Λ_P is a *vertex lattice*, i.e., Λ_P is a $\mathbb{Z}_p$-lattice in V' such that $p\Lambda_P \subset \Lambda_P^\vee \subset \Lambda_P$. The *type* t_P of Λ_P is defined to be $\dim_{\mathbb{F}_p}(\Lambda_P/\Lambda_P^\vee)$. By [HP17, Prop. 5.1.2, (1.2.3.1)], there is $t_{\max} \in 2\mathbb{Z}$ which only depends on n and $\det(V') = \det(V_{\mathbb{Q}_p})$[9] such that $t_P \in 2\mathbb{Z}$ and $2 \le t_P \le t_{\max}$. Moreover, there exists a vertex lattice $\Lambda \subset V'$ of type $t_{\max}$ such that $\Lambda_P \subset \Lambda$. Indeed, the proof of [HP17, Prop. 5.1.2] constructs all possible isometry classes of Λ (with the quadratic form) for all (V, Q) (note that in *loc. cit.*, they proved that for given (V, Q), the isometry class of Λ is unique).

Therefore, given (V, Q), we first obtain the isometry class of Λ of type $t_{\max}$ and then all isometry classes of the lattices of special endomorphisms $L'' \otimes \mathbb{Z}_p$ attached to all supersingular points are given by the duals of the vertex lattices contained in Λ.

From Λ, we may compute all possible isomorphism classes of $\mathbb{L}_{\mathrm{cris},P}(W)$ and $\mathbb{L}_P^{\#}(W)$ as rank n free W-modules endowed with a quadratic form/bilinear form and a σ-linear Frobenius φ (here we use σ to denote the Frobenius action on W) following [HP17, Prop. 6.2.2, §5.3.1]. Indeed, $\mathbb{L}_P^{\#}(W) \subset \Lambda \otimes_{\mathbb{Z}_p} W =: \Lambda_W$ is the preimage of a Lagrangian $\overline{L}_P^{\#} \subset \Lambda_W/\Lambda_W^\vee$ with respect to the quadratic form $pQ' \bmod p$ such that

$$\dim(\overline{L}_P^{\#} + \overline{\varphi}(\overline{L}_P^{\#})) = t_{\max}/2 + 1, \tag{3.1.1}$$

where we use φ to denote the σ-linear map on Λ_W given by $\mathrm{Id} \otimes \sigma$ and $\overline{\varphi}(\overline{v}) := \overline{\varphi(v)}$ is well-defined for $\overline{v} \in \Lambda_W/\Lambda_W^\vee$ with a lift $v \in \Lambda_W$. The quadratic form and φ-action on $\mathbb{L}_P^{\#}(W)$ are the restrictions of the quadratic forms and φ-action on Λ_W. We then obtain $\mathbb{L}_{\mathrm{cris},P}(W) = \varphi(\mathbb{L}_P^{\#}(W))$. Note that by [HP17, Prop. 5.1.2], the even dimensional $\mathbb{F}_p$-quadratic space $(\Lambda/\Lambda^\vee, pQ' \bmod p)$ does not have a Lagrangian defined over $\mathbb{F}_p$ and hence is nonsplit; see [HP14, §§3.2-3.3] for a discussion on how to find all such $\overline{L}_P^{\#}$.

Definition 3.1.3. For a supersingular point P, we say P is *superspecial* if $t_P = 2$;[10] we say P is *supergeneric* if $t_P = t_{\max} \neq 2$.

[8] Note that in [HP17], they use y to denote a point in $\mathcal{M}(k)$ and $\mathbb{L}_P^{\#}(W)$ is denoted by L_y while $\mathbb{L}_{\mathrm{cris},P}(W)$ is denoted by $L_y^{\#}$.

[9] The determinant $\det(V')$ is the determinant of the Gram matrix $([x_i, x_j]')_{i,j=1,\ldots,n+2}$, where $\{x_i\}_{i=1}^{n+2}$ is a $\mathbb{Q}_p$-basis of V'; we view $\det(V')$ as an element in $\mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2$.

[10] In the settings in §2, P is superspecial if and only if the corresponding abelian surface is isomorphic to the product of two supersingular elliptic curves, which is the usual definition for an abelian surface to be superspecial.

By [HP17, Prop. 5.2.2], P is superspecial if and only if

$$\varphi^2(\mathbb{L}_P^{\#}(W)) \subset \mathbb{L}_P^{\#}(W) + \varphi(\mathbb{L}_P^{\#}(W)). \tag{3.1.2}$$

By [HP17, (1.2.3.1)], in the setting of §2, we have $t_{\max} \leq 4$ and hence the supersingular points in question are either superspecial or supergeneric.

Remark 3.1.4. By [MP16, Prop. 4.7 (iii) (iv)], $\mathrm{GSpin}(L, Q)_W$ acts on $\mathbb{D}_P(W)$ and $\mathbb{L}_{\mathrm{cris},P}(W)$; moreover, as W-quadratic spaces, $\mathbb{L}_{\mathrm{cris},P}(W) \cong L \otimes W$ (we use Q_W to denote the quadratic form on $L'' \otimes \mathbb{Z}_p$) and for $x \in \mathbb{L}_{\mathrm{cris},P}(W), x \circ x = Q_W(x) \cdot \mathrm{Id} \in \mathrm{End}(\mathbb{D}_P(W))$. Therefore Q' on $L'' \otimes \mathbb{Z}_p$ is the restriction of Q on $\mathbb{L}_{\mathrm{cris},P}(W)$ to $L'' \otimes \mathbb{Z}_p$. We introduce the notation Q' to emphasize that Q' and Q (as $\mathbb{Z}_p$-quadratic forms) are restrictions of Q_W to $\mathbb{Z}_p$-lattices in different $\mathbb{Q}_p$-subspaces. Hence $\mathrm{GSpin}(L, Q)_W = \mathrm{GSpin}(\mathbb{L}_{\mathrm{cris},P}(W), Q')$.

3.1.5. We now describe the F-crystal $\mathbb{L}_{\mathrm{cris}}$ over the formal completion $\widehat{\mathcal{M}}_P$ along the supersingular point P following [Kis10, §§1.4-1.5] and [Moo98, §4.5]; see also [HP17, §§3.1.4, 3.1.6].

The Hodge filtration $\mathrm{Fil}^1 \mathbb{D}_P(W) \bmod p \subset \mathbb{D}_P(k)$ corresponds to a cocharacter $\overline{\mu} : \mathbb{G}_{m,k} \to \mathrm{GSpin}(L, Q)_k$ and we pick a cocharacter $\mu : \mathbb{G}_{m,W} \to \mathrm{GSpin}(L, Q)_W$ which lifts $\overline{\mu}$. Let $U_P \subset \mathrm{GSpin}(L, Q)_W$ denote the opposite unipotent of the parabolic subgroup defined by μ; and let $\widehat{U_P}$ denote the formal completion of U_P along the identity. Pick coordinates and write $\widehat{U_P} = \mathrm{Spf}\, W[[x_1, \dots, x_d]]$ such that $x_1 = \cdots = x_d = 0$ defines the identity element in U_P. Let σ denote the Frobenius action on $W[[x_1, \dots, x_d]]$ which lifts the σ-action on W and for which $\sigma(x_i) = x_i^p$.

Let R denote $\widehat{\mathcal{O}}_{\mathcal{M},P}$, the complete local ring of $\mathcal{M}$ at P. Then there exists an isomorphism from $\mathrm{Spf}\, R$ to $\widehat{U_P}$ (and we still use σ to denote the Frobenius action on R via the identification to $W[[x_1, \dots, x_d]]$) such that

(1) $\mathbb{D}(R) = \mathbb{D}_P(W) \otimes_W R$ and $\mathbb{L}_{\mathrm{cris}}(R) = \mathbb{L}_{\mathrm{cris},P}(W) \otimes_W R$ as R-modules;
(2) and under the above identifications, the σ-linear Frobenius action, denoted by Frob, on $\mathbb{D}(R)$ and $\mathbb{L}_{\mathrm{cris}}(R)$ is given by $u \cdot (\varphi \otimes \sigma)$, where u denotes the universal $W[[x_1, \dots, x_d]]$-point in $\widehat{U_P}$ and φ is the crystalline Frobenius on $\mathbb{D}_P(W)$ or $\mathbb{L}_{\mathrm{cris},P}(W)$ given in §3.1.2.

On $\mathbb{L}_{\mathrm{cris}}$, the $\mathrm{GSpin}(L, Q)_W$ action factors through the quotient $\mathrm{SO}(L, Q)_W$. So from now on, since we will only care about Frob on $\mathbb{L}_{\mathrm{cris}}$, then by Remark 3.1.4, we will work with $\mu : \mathbb{G}_{m,W} \to \mathrm{SO}(\mathbb{L}_{\mathrm{cris},P}(W), Q')$ and U_P the opposite unipotent of μ in $\mathrm{SO}(\mathbb{L}_{\mathrm{cris},P}(W), Q')$.

In the rest of this section, we will apply §§3.1.2, 3.1.5 to the setting in §2 and we will work with the coordinates on $\widehat{U_P}$. When $L = L_H$, we write $\widehat{U_P} = \mathrm{Spf}\, W[[x, y]]$ and when $L = L_S$, we write $\widehat{U_P} = \mathrm{Spf}\, W[[x, y, z]]$. We will use $\epsilon \in \mathbb{Z}_p^\times$ to denote an element which is not a perfect square in $\mathbb{Z}_p$. Let $\mathbb{Z}_{p^2}$ (resp. $\mathbb{Q}_{p^2}$) denote $W(\mathbb{F}_{p^2})$ (resp. $W(\mathbb{F}_{p^2})[1/p]$) and let $\lambda \in \mathbb{Z}_{p^2}^\times$ be an element such that $\sigma(\lambda) = -\lambda$ (for instance, we can take λ to be a root in $\mathbb{Z}_{p^2}$ of $x^2 - \epsilon = 0$). We will use $\{v_i\}_{i=1}^{n+2}$ to denote a W-basis of $\mathbb{L}_{\mathrm{cris},P}(W)$ and $\{w_i\}_{i=1}^{n+2}$ to denote a $\mathbb{Z}_p$-basis of $\Lambda_P^\vee = \mathbb{L}_{\mathrm{cris},P}(W)^{\varphi=1}$; note that $\mathrm{Span}_W\{w_i\}$ is a W-sublattice of $\mathbb{L}_{\mathrm{cris},P}(W)$.

3.2. **The Hilbert case $L = L_H$.** Recall that as in Theorem 1(2), we have $p \nmid m \in \mathbb{Z}_{>0}$.

3.2.1. Assume that p is inert in $\mathbb{Q}(\sqrt{m})$;[11] then we have $t_{\max} = 4$.

The vertex lattice with type $t_{\max}$ is $\Lambda = \mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1\} \oplus Z$, where

$$[Z, e_1]' = [Z, f_1]' = [e_1, e_1]' = [f_1, f_1]' = 0,\ [e_1, f_1]' = 1/p,\ Z \cong \mathbb{Z}_{p^2},\ Q'(x) = x\sigma(x)/p,\ \forall x \in Z.$$

[11] If $m \in \mathbb{Z}$ is a perfect square, then by convention, we view every prime p to be split in $\mathbb{Q}[x]/(x^2 - m)$ and the discussion of the split case still holds.

Hence $\Lambda^\vee = p\Lambda$. Set $e_2 = (1\otimes 1 + (1/\lambda)\otimes\lambda)/2, f_2 = (1\otimes 1 + (-1/\lambda)\otimes\lambda)/2 \in \mathbb{Z}_{p^2}\otimes_{\mathbb{Z}_p} Z$. Then, as elements in Λ_W,

$$\varphi(e_1) = e_1, \varphi(f_1) = f_1, \varphi(e_2) = f_2, \varphi(f_2) = e_2, [e_2, e_2]' = [f_2, f_2]' = 0, [e_2, f_2]' = 1/p.$$

All possible $\overline{L}_P^{\#}$ are given by two families of Lagrangians in k-quadratic space spanned by $\bar{e}_1, \bar{e}_2, \bar{f}_1, \bar{f}_2 \in \Lambda_W/\Lambda_W^\vee$ with quadratic form pQ satisfying (3.1.1):

$$\overline{L}_P^{\#} = \mathrm{Span}_k\{\bar{e}_1 + \sigma^{-1}(\bar{c})\bar{f}_2, \bar{e}_2 - \sigma^{-1}(\bar{c})\bar{f}_1\}, \text{ or } \overline{L}_P^{\#} = \mathrm{Span}_k\{\bar{e}_1 + \sigma^{-1}(\bar{c})\bar{e}_2, \sigma^{-1}(\bar{c})\bar{f}_1 - \bar{f}_2\},$$

where $\bar{c} \in k$.[12] Therefore, we have that

$$\mathbb{L}_{\mathrm{cris},P}(W) = \mathrm{Span}_W\{e_1 + ce_2, cf_1 - f_2, pe_2, pf_1\}, \text{ or } \mathbb{L}_{\mathrm{cris},P}(W) = \mathrm{Span}_W\{e_1 + cf_2, e_2 - cf_1, pf_1, pf_2\},$$

where $c \in W$.[13] Moreover, by (3.1.2), P is superspecial if and only if $\sigma^{-1}(c) - \sigma(c) \in pW$, which is equivalent to the Teichmuller lift of $\bar{c}$ lying in $\mathbb{Z}_{p^2}$. Note that if $c - c' \in pW$, then c, c' define the same $\mathbb{L}_{\mathrm{cris},P}(W)$. Therefore, without loss of generality, from now on, we will only work with $c \in W$ which is the Teichmuller lifting of $\bar{c} \in k$. Hence P is superspecial if and only if there exists $c \in \mathbb{Z}_{p^2}$ such that $\mathbb{L}_{\mathrm{cris},P}(W)$ is given by the above form.

In order to compute the F-crystal $\mathbb{L}_{\mathrm{cris}}$, we pick the following W-basis $\{v_1, \ldots, v_4\}$ of $\mathbb{L}_{\mathrm{cris},P}(W)$ such that the Gram matrix of $[-,-]'$ with respect to this basis is $\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$, where I denotes the 2×2 identity matrix. For the first family, take

$$v_1 = f_2 - cf_1, v_2 = e_1 + ce_2 - \sigma^{-1}(c)cf_1 + \sigma^{-1}(c)f_2, v_3 = pe_2 - p\sigma^{-1}(c)f_1, v_4 = pf_1;$$

for the second family, take

$$v_1 = e_2 - cf_1, v_2 = \sigma^{-1}(c)e_2 - \sigma^{-1}(c)cf_1 + e_1 + cf_2, v_3 = pf_2 - p\sigma^{-1}(c)f_1, v_4 = pf_1.$$

Then on $\mathbb{L}_{\mathrm{cris},P}(W)$, with respect to $\{v_1, \ldots, v_4\}$, we have

$$\varphi = b\sigma, \text{ with } b = \begin{bmatrix} 0 & \sigma(c) - \sigma^{-1}(c) & p & 0 \\ 0 & 1 & 0 & 0 \\ 1/p & 0 & 0 & 0 \\ (\sigma^{-1}(c) - \sigma(c))/p & 0 & 0 & 1 \end{bmatrix}.$$

The filtration on $\mathbb{L}_{\mathrm{cris},P}(k)$ is given by

$$\mathrm{Fil}^1 \mathbb{L}_{\mathrm{cris},P}(k) = \mathrm{Span}_k\{\bar{v}_3\}, \mathrm{Fil}^0 \mathbb{L}_{\mathrm{cris},P}(k) = \mathrm{Span}_k\{\bar{v}_2, \bar{v}_3, \bar{v}_4\}, \mathrm{Fil}^{-1} \mathbb{L}_{\mathrm{cris},P}(k) = \mathbb{L}_{\mathrm{cris},P}(k),$$

so we may choose $\mu : \mathbb{G}_{m,W} \to \mathrm{SO}(\mathbb{L}_{\mathrm{cris},P}(W), Q')$ to be $t \mapsto \mathrm{diag}(t^{-1}, 1, t, 1)$. Then $\widehat{U_P} = \mathrm{Spf}\, W[[x,y]]$ with the universal point

$$u = \begin{bmatrix} 1 & x & -xy & y \\ 0 & 1 & -y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -x & 1 \end{bmatrix} \text{ and } \mathrm{Frob} = ub\sigma, \text{ with } ub = \begin{bmatrix} -xy/p - ay/p & a + x & p & y \\ -y/p & 1 & 0 & 0 \\ 1/p & 0 & 0 & 0 \\ -x/p - a/p & 0 & 0 & 1 \end{bmatrix},$$

where $a = \sigma(c) - \sigma^{-1}(c)$; we have $a = 0$ if P is superspecial and $a \in W^\times$ if P is supergeneric.

---

[12]Indeed, as $\dim_k \Lambda_W/\Lambda_w^\vee$ is small, in this case, all Lagrangians satisfy (3.1.1). There are two families and each is parametrized by $\mathbb{P}^1(k)$ so more accurately, we shall view $\bar{c} \in \mathbb{P}^1(k)$, i.e., there are two more Lagrangians $\mathrm{Span}_k\{\bar{f}_1, \bar{f}_2\}$ and $\mathrm{Span}_k\{\bar{e}_2, \bar{f}_1\}$; however, since the role of e_i and f_i are symmetric, the computation for these two cases are equivalent to $\mathrm{Span}_k\{\bar{e}_1, \bar{e}_2\}$ and $\mathrm{Span}_k\{\bar{e}_1, \bar{f}_2\}$ so we may safely omit them and only take $\bar{c} \in k$. Moreover, we use $\sigma^{-1}(\bar{c})$ to be the parameter here because eventually we want to work with $\mathbb{L}_{\mathrm{cris},P}(W) = \varphi(\mathbb{L}_P^{\#}(W))$.

[13]Here we notice that φ swaps two families of $\mathbb{L}_P^{\#}(W)$; in particular, the general formula for $\mathbb{L}_{\mathrm{cris},P}(W)$ is the same as that for $\mathbb{L}_P^{\#}(W)$ (other than swapping between the two families). This observation holds true in general by [HP14, Rmk. 3.5].

When P is superspecial, $\{w_1 = pv_1 + v_3, w_2 = \lambda(pv_1 - v_3), w_3 = v_2, w_4 = v_4\}$ is a $\mathbb{Z}_p$-basis of $L'' \otimes \mathbb{Z}_p$. Using $\{w_1, \dots, w_4\}$ as a K-basis of $\mathbb{L}_{\mathrm{cris},P}(W)[1/p]$, we have

$$\text{(3.2.1)} \qquad \mathrm{Frob} = \left(I + \begin{bmatrix} -\frac{xy}{2p} & \frac{\lambda xy}{2p} & \frac{x}{2p} & \frac{y}{2p} \\ -\frac{xy}{2\lambda p} & \frac{xy}{2p} & \frac{x}{2\lambda p} & \frac{y}{2\lambda p} \\ -y & \lambda y & 0 & 0 \\ -x & \lambda x & 0 & 0 \end{bmatrix} \right) \circ \sigma.$$

When P is supergeneric, $\{w_1 = v_4, w_2 = pv_1 + v_3 + (c + \sigma^{-1}(c))v_4, w_3 = \lambda(pv_1 - v_3 + (c - \sigma^{-1}(c))v_4), w_4 = pv_2 - cv_3 - p\sigma^{-1}(c)v_1 - c\sigma^{-1}(c)v_4\}$ is a $\mathbb{Z}_p$-basis of $L'' \otimes \mathbb{Z}_p$ and with respect to this basis, $\mathrm{Frob} = (I + \frac{y}{p}A + xB) \circ \sigma$, where

$$A = \begin{bmatrix} -c & -c^2 & -\lambda c^2 & 0 \\ 1/2 & 0 & \lambda c & c^2/2 \\ 1/(2\lambda) & c/\lambda & 0 & -c^2/(2\lambda) \\ 0 & -1 & \lambda & c \end{bmatrix}, B = \begin{bmatrix} 0 & -1 + cy/p & \lambda cy/p + \lambda & -c^2y/p \\ 0 & -y/(2p) & \lambda y/(2p) & 1/2 + cy/(2p) \\ 0 & -y/(2\lambda p) & y/(2p) & 1/(2\lambda) + cy/(2p\lambda) \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

3.2.2. Assume that p is split in $\mathbb{Q}(\sqrt{m})$; then we have $t_{\max} = 2$ and hence every P is superspecial.

The vertex lattice with type $t_{\max}$ is $\Lambda = \{(x_1, x_2, x_3, x_4) \in \mathbb{Z}_p^4\}$ with

$$Q'((x_1, x_2, x_3, x_4)) = x_1^2 - \epsilon x_2^2 - p^{-1}x_3^2 + \epsilon p^{-1}x_4^2;$$

we have $\Lambda^\vee = \mathrm{Span}_{\mathbb{Z}_p}\{e_1, e_2, pe_3, pe_4\}$, where e_i is the vector with $x_i = 1$ and $x_j = 0$ for $j \neq i$. Recall that we take $\epsilon = \lambda^2$; we then have[14] that

$$\mathbb{L}_{\mathrm{cris},P}(W) = \mathrm{Span}_W\{v_1 = \frac{1}{2}(e_3+\lambda^{-1}e_4), v_2 = \frac{1}{2}(e_1+\lambda^{-1}e_2), v_3 = -\frac{1}{2}(pe_3-\lambda^{-1}pe_4), v_4 = \frac{1}{2}(e_1-\lambda^{-1}e_2)\}.$$

The Gram matrix is $\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$ and on $\mathbb{L}_{\mathrm{cris},P}(W)$, the Frobenius $\varphi = b\sigma$ with

$$b = \begin{bmatrix} 0 & 0 & -p & 0 \\ 0 & 0 & 0 & 1 \\ -1/p & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The filtration on $\mathbb{L}_{\mathrm{cris},P}(k)$ given by φ is the same as in §3.2.1 and hence we may use the same μ and u there. Therefore, on $\mathbb{L}_{\mathrm{cris}}(W[[x,y]])$, we have

$$\mathrm{Frob} = ub\sigma, \text{ with } ub = \begin{bmatrix} xy/p & y & -p & x \\ y/p & 0 & 0 & 1 \\ -1/p & 0 & 0 & 0 \\ x/p & 1 & 0 & 0 \end{bmatrix}.$$

Moreover, $\{w_1 = pv_1 - v_3, w_2 = \lambda(pv_1 + v_3), w_3 = v_2 + v_4, w_4 = \lambda(v_4 - v_2)\}$ is a $\mathbb{Z}_p$-basis of $L'' \otimes \mathbb{Z}_p$ and with respect to this basis,

$$\text{(3.2.2)} \qquad \mathrm{Frob} = \left(I + \begin{bmatrix} \frac{xy}{2p} & -\frac{\lambda xy}{2p} & \frac{x+y}{2p} & \frac{-\lambda(x-y)}{2p} \\ \frac{xy}{2\lambda p} & -\frac{xy}{2p} & \frac{x+y}{2\lambda p} & \frac{-(x-y)}{2p} \\ \frac{x+y}{2} & \frac{-\lambda(x+y)}{2} & 0 & 0 \\ \frac{x-y}{2\lambda} & \frac{-(x-y)}{2} & 0 & 0 \end{bmatrix} \right) \circ \sigma.$$

[14]There are exactly two Lagrangians and the other one is given by replacing λ by $-\lambda$. Since λ and $-\lambda$ play the same role in our later computation, there is no loss of generality here.

3.3. **The Siegel case $L = L_S$.** We now compute $\mathbb{L}_{\mathrm{cris}}$ for Theorem 1(1) and Theorem 3. In this case, we have $t_{\max} = 4$.

The vertex lattice with type $t_{\max}$ is $\Lambda = \mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1\} \oplus Z_S$, where $Z_S = \{(x_1, x_2, x_3) \in \mathbb{Z}_p^3\}$ $[Z_S, e_1]' = [Z_S, f_1]' = [e_1, e_1]' = [f_1, f_1]' = 0$, $[e_1, f_1]' = 1/p, Q'((x_1, x_2, x_3)) = c(-\epsilon x_1^2 - p^{-1}x_2^2 + \epsilon p^{-1} x_3^2)$, for some $c \in \mathbb{Z}_p^\times$. Since $\det \Lambda = \det L \in \mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2$ and $\det L = 2$, we have $c = -1$. Let $g = (1, 0, 0) \in Z_S$ and $Z = \mathrm{Span}_{\mathbb{Z}_p}\{(0, 1, 0), (0, 0, 1)\} \subset Z_S$. Then $\Lambda/\Lambda^\vee = \mathrm{Span}_{\mathbb{F}_p}\{\bar{e}_1, \bar{f}_1\} \oplus Z/Z^\vee$. Note that $\mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1\} \oplus Z$ is exactly the same quadratic $\mathbb{Z}_p$-lattice which is denoted by Λ in §3.2.1; hence the same computation there applies to find $\mathbb{L}_{\mathrm{cris},P}(W) \subset \Lambda \otimes W$. More precisely, there exist $v_1, \dots, v_4 \in \mathrm{Span}_W\{e_1, f_1\} \oplus Z \otimes W$ and $c \in W$ which is the Teichmuller lift of $\bar{c} \in k$ such that

(1) $\mathbb{L}_{\mathrm{cris},P}(W) = \mathrm{Span}_W\{v_1, \dots, v_4, v_5\}$, where $v_5 = g$;

(2) the Gram matrix of $[-, -]'$ with respect to $\{v_1, \dots, v_5\}$ is $\begin{bmatrix} 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & 2\epsilon \end{bmatrix}$, where I is the 2×2 identity matrix;

(3) The Frobenius φ on $\mathbb{L}_{\mathrm{cris},P}(W)$ with respect to the basis $\{v_i\}$ is

$$\varphi = b\sigma, \text{ with } b = \begin{bmatrix} 0 & \sigma(c) - \sigma^{-1}(c) & p & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1/p & 0 & 0 & 0 & 0 \\ (\sigma^{-1}(c) - \sigma(c))/p & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix};$$

(4) P is superspecial if and only if $\sigma^2(c) = c$.

We may choose $\mu : \mathbb{G}_{m,W} \to \mathrm{SO}(\mathbb{L}_{\mathrm{cris},P}(W), Q')$ to be $t \mapsto \mathrm{diag}(t^{-1}, 1, t, 1, 1)$. Then $\widehat{U_P} = \mathrm{Spf}\, W[[x, y, z]]$ with the universal point

$$u = \begin{bmatrix} 1 & x & -xy - \frac{z^2}{4\epsilon} & y & z \\ 0 & 1 & -y & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -x & 1 & 0 \\ 0 & 0 & -\frac{z}{2\epsilon} & 0 & 1 \end{bmatrix} \text{ and } \mathrm{Frob} = ub\sigma, \text{ with } ub = \begin{bmatrix} -\frac{1}{p}(xy + \frac{z^2}{4\epsilon}) - \frac{ay}{p} & a + x & p & y & z \\ -\frac{y}{p} & 1 & 0 & 0 & 0 \\ \frac{1}{p} & 0 & 0 & 0 & 0 \\ -\frac{x+a}{p} & 0 & 0 & 1 & 0 \\ -\frac{z}{2\epsilon p} & 0 & 0 & 0 & 1 \end{bmatrix}$$

acting on $\mathbb{L}_{\mathrm{cris}}(W[[x, y, z]])$, where $a = \sigma(c) - \sigma^{-1}(c)$; note that $a = 0$ if and only if P is superspecial.

For the proofs of Theorem 1(1) and Theorem 3, we only need to study superspecial points so we only give the matrix of Frob with respect to a basis of $\mathbb{L}_{\mathrm{cris}} \otimes_W K$ consisting of elements in $L'' \otimes \mathbb{Z}_p$ when P is superspecial; we refer the reader to the appendix for the discussion when P is supergeneric.

We now assume that P is superspecial. Let $w_1 = \lambda(pv_1 - v_3), w_2 = pv_1 + v_3, w_3 = v_2, w_4 = v_4, w_5 = v_5$. Then $L'' \otimes \mathbb{Z}_p = \mathrm{Span}_{\mathbb{Z}_p}\{w_1, \dots, w_5\}$. We view $\{w_i\}_{i=1}^5$ as a K-basis of $\mathbb{L}_{\mathrm{cris},P}(W) \otimes K$, then the Frobenius on $\mathbb{L}_{\mathrm{cris}}(W[[x, y, z]])$ is given by

$$\mathrm{Frob} = \left(I + \begin{bmatrix} \frac{1}{2p}(xy + \frac{z^2}{4\epsilon}) & -\frac{1}{2\lambda p}(xy + \frac{z^2}{4\epsilon}) & \frac{x}{2\lambda p} & \frac{y}{2\lambda p} & \frac{z}{2\lambda p} \\ \frac{\lambda}{2p}(xy + \frac{z^2}{4\epsilon}) & -\frac{1}{2p}(xy + \frac{z^2}{4\epsilon}) & \frac{x}{2p} & \frac{y}{2p} & \frac{z}{2p} \\ \lambda y & -y & 0 & 0 & 0 \\ \lambda x & -x & 0 & 0 & 0 \\ \frac{\lambda z}{2\epsilon} & -\frac{z}{2\epsilon} & 0 & 0 & 0 \end{bmatrix} \right) \circ \sigma. \tag{3.3.1}$$

3.4. **Equation of non-ordinary locus.** We now use the computation in §§3.2, 3.3 to obtain the local equation of the non-ordinary locus in a formal neighborhood of a supersingular point P using results in [Ogu01]. Although [Ogu01] only focuses on the case of K3 surfaces, the results that we recall here apply to any GSpin Shimura varieties. We follow the notation in §3.1. For a perfect field k' of characteristic p, for $P' \in \mathcal{M}(k')$, we say P is *ordinary* if the slopes of the crystalline Frobenius φ on $\mathbb{L}_{\mathrm{cris},P'}(W)$ are $-1, 1$ with multiplicity 1 and 0 with multiplicity n.[15]

The cocharacter $\overline{\mu}$ defines a filtration $\mathrm{Fil}^i, i = -1, 0, 1$ on $\mathbb{L}_{\mathrm{cris},P}(k)$, which is the Hodge filtration in [Ogu01] and in particular, $\dim \mathrm{Fil}^1 \mathbb{L}_{\mathrm{cris},P}(k) = 1, \dim \mathrm{Fil}^0 \mathbb{L}_{\mathrm{cris},P}(k) = n+1, \dim \mathrm{Fil}^{-1} \mathbb{L}_{\mathrm{cris},P}(k) = n+2$ and the annihilator of $\mathrm{Fil}^1 \mathbb{L}_{\mathrm{cris},P}(k)$ in $\mathbb{L}_{\mathrm{cris},P}(k)$ with respect to Q is $\mathrm{Fil}^0 \mathbb{L}_{\mathrm{cris},P}(k)$.[16] The Hodge filtration over the mod p complete local ring $R \otimes_W k$ at P is given by $\mathrm{Fil}^i \mathbb{L}_{\mathrm{cris}}(R \otimes_W k) := \mathrm{Fil}^i \mathbb{L}_{\mathrm{cris},P}(k) \otimes_k (R \otimes k)$. Note that $\mathrm{Frob}(\mathrm{Fil}^0 \mathbb{L}_{\mathrm{cris}}(R \otimes_W k)) \subset \mathrm{Fil}^0 \mathbb{L}_{\mathrm{cris}}(R \otimes_W k)$, so we have a well-defined map $p\,\mathrm{Frob} : \mathrm{gr}_{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes_W k) \to \mathrm{gr}_{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes_W k)$, where $\mathrm{gr}_{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes_W k) := \mathrm{Fil}^{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes_W k) / \mathrm{Fil}^0 \mathbb{L}_{\mathrm{cris}}(R \otimes_W k)$.

Lemma 3.4.1 (Ogus)**.** *For a supersingular point P, The non-ordinary locus (over k) in the formal neighborhood of P is given by the equation*

$$p\,\mathrm{Frob}\,|_{\mathrm{gr}_{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes_W k)} = 0.$$

Proof. By [Ogu01, Prop. 11], the discussion of the conjugate filtration on [Ogu01, p.333-334], and the fact that the annihilator of $\mathrm{Fil}^1 \mathbb{L}_{\mathrm{cris}}(R \otimes k)$ in $\mathbb{L}_{\mathrm{cris}}(R \otimes k)$ with respect to Q is $\mathrm{Fil}^0 \mathbb{L}_{\mathrm{cris}}(R \otimes k)$, we have that the equation defining the non-ordinary locus is the projection of the conjugate filtration (denoted by F^2_{con} in *loc. cit.*) to $\mathrm{gr}_{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes k)$. By definition, $F^2_{con} = p\,\mathrm{Frob}\,\mathbb{L}_{\mathrm{cris}}(R \otimes k)$ and then the lemma follows. ☐

Corollary 3.4.2. *When $L = L_H$, the local equation of the non-ordinary locus in a formal neighborhood of a supersingular point P is $xy = 0$ if P is superspecial and is $y = 0$ if P is supergeneric; when $L = L_S$, the local equation is $xy + z^2/(4\epsilon) = 0$ if P is a superspecial point and $(x+a)y + z^2/(4\epsilon) = 0$ if P supergeneric, where $a \in W(k)^\times$ depends on P.*

Proof. We will prove this corollary in the case $L = L_S$, since the other case is handled the same way. Recall we have the basis $v_1 \dots v_5$ of $\mathbb{L}_{\mathrm{cris}}$, with $\mathrm{Fil}^-1 = \mathbb{L}_{\mathrm{cris}}$ and Fil^0 being spanned by v_2, v_3, v_4 and v_5. Therefore, using the explicit formulas from the previous section, we see the map $p\,\mathrm{Frob} : \mathrm{gr}_{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes_W k) \to \mathrm{gr}_{-1} \mathbb{L}_{\mathrm{cris}}(R \otimes_W k)$ is given by $p\,\mathrm{Frob}(v_1) = -(xy + \frac{z^2}{4\epsilon} + ay)v_1$. Our result now follows from Ogus's description of the non-ordinary locus. ☐

4. Arithmetic Borcherds Theory, Siegel mass formula, and Eisenstein series

We use arithmetic Borcherds theory [HMP] to control the global intersection number of a curve C with special divisors. More precisely, we use the work of Bruinier and Kuss in [BK03] to study the Fourier coefficients of the Eisenstein part of the (vector-valued) modular form arising from Borcherds theory. In order to compare the global intersection number with the local contribution later in the paper, we also apply the computations in [BK03] and the Siegel mass formula to the Eisenstein part of the theta series attached to a supersingular point and reduce the question to a computation of local densities and determinants of the lattices L and L' introduced in §2.1 and Definition 2.3.1 (in §4.2, we will summarize the properties of L'). We use Hanke's method in [Han04] to compute the local densities. Throughout this section, p is an odd prime such that L is self-dual at p. For a prime ℓ, we use $v_\ell : \mathbb{Z}_\ell \backslash \{0\} \to \mathbb{Z}_{\geq 0}$ to denote the ℓ-adic valuation.

[15]When $L = L_H, L_S$, the point P' is ordinary if and only if the corresponding abelian surface over k' is ordinary by the definition of $\mathbb{L}_{\mathrm{cris}}$.

[16]See also [Ogu82, p.411] for the definition. Note that here we directly work on the crystalline cohomology without using the canonical isomorphism to the de Rham cohomology. Note that our filtration is shifted by 1 when comparing to the filtration in [Ogu01] because his Frobenius is p times our Frobenius.

4.1. **Arithmetic Borcherds theory and the explicit formula for the Eisenstein series.** Recall the special divisors $Z(m)$ from Definition 2.2.6 and §2.2.11. The following modularity result is the key input to the estimate of the intersection number $Z(m).C$.

In order to state the result using vector-valued modular forms, for $\mu \in L^\vee/L, m \in \mathbb{Q}_{>0}$, let $\mathcal{Z}(m,\mu)$ denote the special divisors over $\mathbb{Z}$ in $\mathcal{M}$ defined in [AGHMP18, §4.5, Def. 4.5.6]. By definition, $\mathcal{Z}(m,0)$ is the divisor $\mathcal{Z}(m)$ defined in §2.2; and roughly speaking, $\mathcal{Z}(m,\mu)$ parametrizes abelian surfaces A with a special quasi-endomorphism s such that $Q(s)=m$ and the ℓ-adic and crystalline realizations of s lie in the image of $(\mu+L)\otimes\mathbb{Z}_\ell$ and $(\mu+L)\otimes\mathbb{Z}_p$ in $\mathrm{End}(T_\ell(A)\otimes\mathbb{Q}_\ell)$ and $\mathrm{End}(\mathbb{D}\otimes_W W[1/p])$ respectively, where $\mathbb{D}$ is the Dieudonné module of A . By the proof of [AGHMP18, Prop. 4.5.8] and [MP16, Prop. 5.21], the assumption that L is self-dual at p implies that $\mathcal{Z}(m,\mu)$ is flat over $\mathbb{Z}_p$. Let $Z(m,\mu)$ denote $\mathcal{Z}(m,\mu)_{\mathbb{F}_p}$. Let $(\mathfrak{e}_\mu)_{\mu\in L^\vee/L}$ denote the standard basis of $\mathbb{C}[L^\vee/L]$. Let $\omega\in\mathrm{Pic}(\mathcal{M}_{\mathbb{F}_p})_\mathbb{Q}$ denote the Hodge line bundle in the $\mathbb{Q}$-Picard group of $\mathcal{M}_{\mathbb{F}_p}$; in other words, ω is the line bundle of weight 1 modular forms (see for instance [AGHMP18, Thm. 4.4.6] for a definition of ω).

Theorem 4.1.1 (Borcherds, Howard–Madapusi-Pera)**.** *Assume (L,Q) is a maximal quadratic lattice of signature $(n,2)$ such that L is self-dual at p. The generating series*

$$\omega^{-1}\mathfrak{e}_0+\sum_{m>0,\mu\in L^\vee/L} Z(m,\mu)q^m\mathfrak{e}_\mu,\ \ \textit{where } q=e^{2\pi i\tau},$$

lies in $M_{1+\frac{n}{2}}(\rho_L)\otimes\mathrm{Pic}(\mathcal{M}_{\mathbb{F}_p})_\mathbb{Q}$. Here, ρ_L denotes the Weil representation on $\mathbb{C}[L^\vee/L]$ and $M_{1+\frac{n}{2}}(\rho_L)$ denotes the space of vector-valued modular forms of $\mathrm{Mp}_2(\mathbb{Z})$ with respect to ρ_L of weight $1+\frac{n}{2}$.[17] *In particular, for any $\mathbb{Q}$-linear functional $\alpha:\mathrm{Pic}(\mathcal{M}_{\mathbb{F}_p})_\mathbb{Q}\to\mathbb{C}$, the vector-valued power series*

$$\alpha(\omega^{-1})\mathfrak{e}_0+\sum_{m>0,\mu\in L^\vee/L}\alpha(Z(m,\mu))q^m\mathfrak{e}_\mu$$

is the Fourier expansion of an element of $M_{1+\frac{n}{2}}(\rho_L)$.

Proof. By abuse of notation, we also use ω to denote the Hodge line bundle over $\mathcal{M}$. By [HMP, Thm. B], the generating series $\omega^{-1}\mathfrak{e}_0+\sum_{m>0,\mu\in L^\vee/L}\mathcal{Z}(m,\mu)q^m\mathfrak{e}_\mu\in M_{1+\frac{n}{2}}(\rho_L)\otimes\mathrm{Pic}(\mathcal{M})_\mathbb{Q}$. Since $\mathcal{Z}(m,\mu)$ are flat over $\mathbb{Z}_p$, then the desired assertion follows from intersecting with $\mathcal{M}_{\mathbb{F}_p}$.

Alternatively, Borcherds [Bor99, Thm. 4.5] proved the modularity assertion for $\mathcal{Z}(m,\mu)_\mathbb{C}$. By [Mau14, Lemma 5.12] and the flatness of $\mathcal{Z}(m,\mu)$, the proof of Borcherds implies the desired modularity for $Z(m,\mu)$. □

4.1.2. In the setting of Theorem 1(2) (i.e. the case when $L=L_H$), we work with curves C that are not necessarily proper. We therefore need a version of the above modularity result that holds for the special fiber a toroidal compactification of $\mathcal{M}$. To that end, let $\mathcal{M}^{\mathrm{tor}}$ denote a toroidal compactification of $\mathcal{M}$, and let $D_1,\dots,D_k$ denote irreducible components of the boundary $\mathcal{M}^{\mathrm{tor}}_{\mathbb{F}_p}\setminus\mathcal{M}_{\mathbb{F}_p}$. In [BBGK07, Theorem 6.2], the authors prove the modularity result for $\mathcal{M}^{\mathrm{tor}}$, which will directly imply the modularity result for $\mathcal{M}^{\mathrm{tor}}_{\mathbb{F}_p}$. The constant term is still given by the Hodge line bundle, still denoted by ω, on $\mathcal{M}^{\mathrm{tor}}_{\mathbb{F}_p}$ and the special divisors $Z(m,\mu)$ are replaced by[18] $Z'(m,\mu)+E(m,\mu)$, where $Z'(m,\mu)$ is the Zariski-closure of $Z(m,\mu)$ in $\mathcal{M}^{\mathrm{tor}}_{\mathbb{F}_p}$, and $E(m,\mu)$ is a "correction term", and has as its irreducible components the D_i with appropriate multiplicity. Crucially, when $Z(m,\mu)$ is proper (see §4.3.3 for when this happens), the multiplicities of the D_i in correction term

[17] In [Bor99], [BK01], [BK03], they work with $(L,-Q)$ and the modular form is with respect to the dual of the Weil representation of $(L,-Q)$, which is the Weil representation of (L,Q). Our convention is the same as the one in [HMP] and [Bru17].

[18] Our notation $Z'+E$ is different from the notation used in *loc. cit.*

$E(m,\mu)$ are all zero and hence $E(m,\mu)$ is trivial. Therefore, compact special divisors stay as they are in the modularity theorem for $\mathcal{M}^{\mathrm{tor}}_{\mathbb{F}_p}$.[19]

4.1.3. Recall that we have a finite morphism $\pi: C \to \mathcal{M}_{\bar{\mathbb{F}}_p}$. When C is proper, for $Z \in \mathrm{Pic}(\mathcal{M}_{\mathbb{F}_p})_{\mathbb{Q}}$, we define $C.Z$ as the degree of $\pi^* Z \in \mathrm{Pic}(C)_{\mathbb{Q}}$. For Theorem 1(2), we pick a toroidal compactification $\mathcal{M}^{\mathrm{tor}}$ of the Hilbert modular surface $\mathcal{M}$ and let C' denote the smooth compactification of C and the finite morphism π extends to a finite morphism $\pi': C' \to \mathcal{M}^{\mathrm{tor}}_{\bar{\mathbb{F}}_p}$. Then for a proper divisor Z in $\mathcal{M}_{\mathbb{F}_p}$, we use $C.Z$ to denote $\deg_{C'}(\pi^* Z)$; since Z is proper, $C' \cap Z = C \cap Z$ so we only need to consider points in $\mathcal{M}_{\bar{\mathbb{F}}_p}$.

4.1.4. We apply Theorem 4.1.1 and §4.1.2 to $\alpha(Z) := C.Z$ defined n §4.1.3 for $Z \in \mathrm{Pic}(\mathcal{M}_{\mathbb{F}_p})_{\mathbb{Q}}$ (and we further assume that Z is proper when $L = L_H$). We decompose the modular form $-(\omega.C)\mathfrak{e}_0 + \sum_{m>0,\mu\in L^\vee/L} Z(m,\mu).Cq^m\mathfrak{e}_\mu$ as $E(q) + G(q)$, where $E(q) \in M_{1+\frac{n}{2}}(\rho_L)$ is an Eisenstein series and $G(q) \in M_{1+\frac{n}{2}}(\rho_L)$ is a cusp form. Note that the constant term of $E(q)$ is $-(\omega.C)\mathfrak{e}_0$.

We now recall the vector-valued Eisenstein series $E_0(\tau) \in M_{1+\frac{n}{2}}(\rho_L)$ which has constant term $\mathfrak{e}_0$. This Eisenstein series has been studied in [Bru02, §1.2.3], [BK01, §4], and [BK03, §3]. Here we follow [Bru17, §2.1] as we use the same convention of quadratic forms. We denote an element in $\mathrm{Mp}_2(\mathbb{Z})$ by (g,σ), where $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ and σ is a choice of the square root of $c\tau + d$. Let $\Gamma'_\infty \subset \mathrm{Mp}_2(\mathbb{Z})$ denote the stabilizer of ∞. Then for $n \geq 3$, the following summation converges on the upper half plane and we define

$$E_0(\tau) = \sum_{(g,\sigma)\in\Gamma'_\infty\backslash \mathrm{Mp}_2(\mathbb{Z})} \sigma(\tau)^{-(2+n)}(\rho_L(g,\sigma)^{-1}\mathfrak{e}_0).$$

When $n = 2$, we define $E_0(\tau)$ use analytic continuation following [BK03, §3]. Write $\tau = x + iy$ and define for $s \in \mathbb{C}$,

$$E_0(\tau, s) = \sum_{(g,\sigma)\in\Gamma'_\infty\backslash \mathrm{Mp}_2(\mathbb{Z})} \sigma(\tau)^{-(2+n)}(\rho_L(g,\sigma)^{-1}(y^s\mathfrak{e}_0)),$$

which converges on the upper half plane for s with $\Re s > 0$ ($n = 2$ here). By [BK03, p. 1697], $E_0(\tau, s)$ has meromorphic continuation in s to the entire $\mathbb{C}$ and it is holomorphic at $s = 0$ and we define $E_0(\tau)$ to be the value at $s = 0$ of the meromorphic continuation of $E_0(\tau, s)$. Moreover, by *loc. cit.*, $E_0(\tau)$ is holomorphic and hence lies in $M_{1+\frac{n}{2}}(\rho_L)$ if ρ_L does not contain the trivial representation as a subquotient. In the proof of Theorem 1(2), we work with $L = L_H$ and this condition for ρ_L is always satisfied as far as the m in the statement of Theorem 1(2) is not a perfect square, i.e., $\mathcal{M}$ is not the product of modular curves.

We denote the q-expansion of $E_0(\tau)$ as $\sum_{m\geq 0, m\in\mathbb{Z}+Q(\mu)} q_L(m,\mu)q^m\mathfrak{e}_\mu$ and set $q_L(m) := q_L(m, 0)$ for $m \in \mathbb{Z}_{>0}$.

4.1.5. We fix some notations before we state the explicit formula of $q_L(m)$ given by Bruinier–Kuss. Given a quadratic lattice L (not necessarily the lattice L_H, L_S), we write $\det(L)$ for the determinant of its Gram matrix. We have $|L^\vee/L| = |\det(L)|$.

[19] We note that in [BBGK07], the authors work with Hilbert modular surfaces attached to real quadratic fields with prime discriminant D and state the modularity result using modular forms with level $\Gamma_0(D)$. However, their proof, which uses Borcherds product for the Fourier expansion and the flatness of $\mathcal{Z}(m,\mu)$, applies for all Hilbert modular surfaces in the setting of vector-valued modular forms by using the original work of Borcherds [Bor98]. We hence deduce modularity for $\mathcal{M}^{\mathrm{tor}}_{\mathbb{F}_p}$. Although the integral special divisors (denoted by $\mathcal{T}(n)$ in [BBGK07]) are defined by taking Zariski closure in $\mathcal{M}^{\mathrm{tor}}$ of the special divisors on the generic fiber $\mathcal{M}^{\mathrm{tor}}_{\mathbb{Q}}$, this notion coincides with our definition by the flatness of the integral special divisors in both definitions.

For a rational prime ℓ, we use $\delta(\ell, L, m)$ to denote the local density of L representing m over $\mathbb{Z}_\ell$. More precisely, $\delta(\ell, L, m) = \lim_{a\to\infty} \ell^{a(1-\operatorname{rk} L)}\#\{v \in L/\ell^a L \mid Q(v) \equiv m \bmod \ell^a\}$. [BK01, Lem. 5] asserts that the limit is stable once $a \geq 1 + 2v_\ell(2m)$. In particular, if m is representable by $(L \otimes \mathbb{Z}_\ell, Q)$, then $\delta(\ell, L, m) > 0$.

Given $0 \neq D \in \mathbb{Z}$ such that $D \equiv 0, 1 \bmod 4$, we use χ_D to denote the Dirichlet character $\chi_D(a) = \left(\frac{D}{a}\right)$, where $\left(\frac{\cdot}{\cdot}\right)$ is the Kronecker symbol. For a Dirichlet character χ, we set $\sigma_s(m, \chi) = \sum_{d|m} \chi(d)d^s$.

Theorem 4.1.6 (Bruinier–Kuss; see also [Bru17, Thms. 2.3, 2.4])**.** *Consider $L = L_H, L_S$ defined in §2.1 and $m \in \mathbb{Z}_{>0}$.*

(1) For $L = L_H$, the Fourier coefficient $q_L(m)$ is

$$-\frac{4\pi^2 m\sigma_{-1}(m, \chi_{4\det L})}{\sqrt{|L^\vee/L|}L(2, \chi_{4\det L})} \prod_{\ell | 2\det(L)} \delta(\ell, L, m).$$

(2) For $L = L_S$, write $m = m_0 f^2$, where $\gcd(f, 2\det L) = 1$ and $v_\ell(m_0) \in \{0, 1\}$ for all $\ell \nmid 2\det L$. Then the Fourier coefficient $q_L(m)$ is

$$-\frac{16\sqrt{2}\pi^2 m^{3/2} L(2, \chi_{\mathcal{D}})}{3\sqrt{|L^\vee/L|}\zeta(4)} \left(\sum_{d|f} \mu(d)\chi_{\mathcal{D}}(d)d^{-2}\sigma_{-3}(f/d)\right) \prod_{\ell|2\det L} \Big(\delta(\ell, L, m)/(1 - \ell^{-4})\Big),$$

where μ is the Mobius function and $\mathcal{D} = -2m_0 \det L$.[20]

Proof. When $L = L_S$, this is [BK01, Thm. 11]. When $L = L_H$, one modifies the proof of [BK01, Thm. 11] as follows. Using [BK03, Prop. 3.1] instead of [BK01, Prop. 2], we obtain [BK01, Prop. 3] since Shintani's formula works in general. To express the formula in [BK01, Prop. 3] as a product of local terms, we use [Iwa97, §11.5, p. 196]. The rest of the proof, which computes the local terms at $\ell \nmid 2\det L$, works in the same way (see also [Iwa97, eqns (11.71)–(11.74)]). □

If $Z(m) \neq \emptyset$, then m is representable by (L, Q) and in particular for every ℓ, m is representable by $(L \otimes \mathbb{Z}_\ell, Q)$ and hence $\delta(\ell, L, m) > 0$. By Theorem 4.1.6, we have $q_L(m) < 0$ when $Z(m) \neq \emptyset$.

4.2. The lattice L' and the Siegel mass formula.

4.2.1. For a supersingular point $P \in \mathcal{M}(k)$, we defined L'', the lattice of special endomorphisms, in Definition 2.3.1 and picked $L' \supset L''$ which is maximal at all $\ell \neq p$ and $L' \otimes \mathbb{Z}_p = L'' \otimes \mathbb{Z}_p$. Though there may be choices for L', the local lattices $L' \otimes \mathbb{Z}_\ell$ are well-defined up to isometry. More precisely, for $\ell \neq p$, $L' \otimes \mathbb{Z}_\ell$ is given by Lemma 2.3.2; and for $\ell = p$, $L' \otimes \mathbb{Z}_p = L'' \otimes \mathbb{Z}_p$ is computed in §§3.2-3.3. Note that given L, the isometry class of the quadratic lattice $L' \otimes \mathbb{Z}_p$ only depends on whether P is superspecial or supergeneric; indeed, following the notation in §3.1.2, if $t_P = t_{\max}$ (for instance, when P is supergeneric), then Λ_P is a maximal lattice with respect to pQ' and hence its isometry class (and thus the isometry class of $L' \otimes \mathbb{Z}_p = \Lambda_P^\vee$) is unique; if $t_P = 2$, i.e., P is superspecial, then $\Lambda_P^\vee$ is a maximal lattice with respect to Q' and hence is unique up to isometry.

In order to compute the local intersection number of $Z(m).C$ at P, we also need to consider sublattices L''' of L' such that $L''' \otimes \mathbb{Z}_\ell = L' \otimes \mathbb{Z}_\ell$ for all $\ell \neq p$ (more precisely, we will take L''' to be the lattices defined in §7.2.3). In particular, $\det L''' = p^{2a} \det L'$ for some $a \in \mathbb{Z}_{\geq 0}$.

Let $\theta_{L'''}(q)$ denote the theta series of the positive definite lattice L''', which is a modular form of weight $\operatorname{rk} L'/2$; we decompose $\theta_{L'''}(q) = E_{L'''}(q) + G_{L'''}(q)$, where $E_{L'''}$ is an Eisenstein series and $G_{L'''}$ is a cusp form. Let $q_{L'''}(m)$ denote the m-th Fourier coefficients of $E_{L'''}$ (at the cusp ∞). The following theorem asserts that $q_{L'''}(m)$ only depends on the genus of L''' and gives explicit formula for $q_{L'''}(m)$. In particular, when we consider the theta series for L', we have that $q_{L'}(m)$ is

[20] Since $\det L_S = 2$ and more generally for odd rank quadratic lattice L, we have $2 \mid \det L$, then $\mathcal{D} \equiv 0 \bmod 4$.

independent of the choice of L' above and it only depends on L and whether P is superspecial or supergeneric.

Theorem 4.2.2 (Siegel mass formula)**.** *Notation as in §4.2.1. The Eisenstein series $E_{L'''}$ only depends on the genus of L'''. Moreover, for $m \in \mathbb{Z}_{>0}$,*

(1) when $L = L_H$,

$$q_{L'''}(m) = \frac{4\pi^2 m \sigma_{-1}(m, \chi_{4\det L'})}{\sqrt{|L'''^\vee/L'''|}L(2, \chi_{4\det L'})} \prod_{\ell | 2 det L'} \delta(\ell, L''', m);$$

(2) when $L = L_S$,

$$q_{L'''}(m) = \frac{16\sqrt{2}\pi^2 m^{3/2} L(2, \chi_{\mathcal{D}'})}{3\sqrt{|L'''^\vee/L'''|}\zeta(4)} \left(\sum_{d|f} \mu(d)\chi_{\mathcal{D}'}(d) d^{-2}\sigma_{-3}(f/d)\right) \prod_{\ell | 2\det L'} \Big(\delta(\ell, L''', m)/(1-\ell^{-4})\Big),$$

where we write $m = m_0 f^2$ with $\gcd(f, 2\det L') = 1$ and $v_\ell(m_0) \in \{0, 1\}$ for all $\ell \nmid 2\det L'$ and $\mathcal{D}' = -2m_0 \det L'$

Proof. The first assertion follows from the Siegel mass formula; see for instance [IK04, Thm. 20.9, eqn. (20.121), and pp. 479-480]. In order to obtain the formula above, we note that the proof of [BK01, Thm. 11] using [BK01, Thm. 6] also applies to L''' and hence we conclude that the formula in [Bru17, Thms. 2.3, 2.4] also applies to L''' and obtain the formulae in the theorem with all L' replaced by L'''. Note that by the computations in §§3.2-3.3, we have $p \mid \det L'$, and hence $\ell \mid 2\det L'''$ if and only if $\ell \mid 2\det L'$; also $\chi_{4\det L'''} = \chi_{4\det L'}$ and $\chi_{\mathcal{D}'} = \chi_{-2m_0\det L'''}$. Hence using L' (instead of L''') for $\chi, \mathcal{D}'$ and the product $\ell \mid 2\det L'$ yields the same formulae. ☐

4.3. **The asymptotic of $q_L(m)$.** The discussion of this subsection also applies to $q_{L'''}(m)$ when m is representable by (L''', Q'), but we only focus on $q_L(m)$ here.

4.3.1. Assume that m is representable by $(L \otimes \mathbb{Z}_\ell, Q)$ for every prime ℓ. We will also assume that, as m varies within a specified set T, there exists an absolute constant $C > 0$ such that for all $\ell \mid 2\det L$, we have $v_\ell(m) \le C$. As we shall see in §4.3.3, we will always be in this situation.

For a given $\ell \mid 2\det L$, as in [Bru17, proof of Prop. 2.5], by [BK01, Lem. 5], we have $\delta(\ell, L, m) = \ell^{a(1-\mathrm{rk}\, L)}\#\{v \in L/\ell^a L \mid Q(v) \equiv m \bmod \ell^a\}$ with $a = 1 + 2C + 2v_\ell(2)$ and hence $\ell^{a(1-\mathrm{rk}\, L)} \le \delta(\ell, L, m) \le \ell^a$.[21]

Therefore, given (L, Q), by Theorem 4.1.6, we have that $|q_L(m)| \asymp m\sigma_{-1}(m, \chi_{4\det L})$ and hence $m^{1-\epsilon} \ll_\epsilon |q_L(m)| \ll_\epsilon m^{1+\epsilon}$ for $L = L_H$; and $|q_L(m)| \asymp m^{3/2} L(2, \chi_{\mathcal{D}}) \sum_{d|f} \mu(d)\chi_{\mathcal{D}}(d) d^{-2}\sigma_{-3}(f/d)$ for $L = L_S$. As in the proof of [Bru17, Prop. 2.5], we have $\sum_{d|f} \mu(d)\chi_{\mathcal{D}}(d)d^{-2}\sigma_{-3}(f/d) \ge 1/5$ and

$$\sum_{d|f} \mu(d)\chi_{\mathcal{D}}(d)d^{-2}\sigma_{-3}(f/d) \le \sum_{d|f} d^{-2}\sigma_{-3}(f/d) < \sum_{d|f} d^{-2}\zeta(3) < \zeta(2)\zeta(3);$$

moreover, by *loc. cit.*, $L(2, \chi_D) \ge \zeta(4)/\zeta(2)$ and $L(2, \chi_D) \le \prod_p (1-p^{-2})^{-1} = \zeta(2)$. Hence $|q_L(m)| \asymp m^{3/2}$ when $L = L_S$.

Lemma 4.3.2. *We fix the same assumptions as in §4.3.1. For $m \gg 1$, we have $Z(m) \neq \emptyset$ and the intersection number $Z(m).C = -q_L(m)(\omega.C) + o(|q_L(m)|)$. More precisely, when $L = L_H$, the error term can be bounded by $O_\epsilon(m^{1/2+\epsilon})$ and when $L = L_S$, the error term can be bounded by $O(m^{5/4})$.*

[21] When $\mathrm{rk}\, L \ge 5$, for a fixed ℓ, it is well known that $\delta(\ell, L, m) \asymp 1$ for all m representable by $(L \otimes \mathbb{Z}_\ell, Q)$ without imposing any bound on $v_\ell(m)$; see for instance [Iwa97, pp. 198-199].

Proof. We follow the discussion in §4.1.4. Let $g(m), m \in \mathbb{Z}_{>0}$ denote the m-th Fourier coefficients of $\mathfrak{e}_0$-component of $G(q)$, which is also a cusp form of weight $1 + \frac{n}{2}$ with respect to certain congruence subgroup of $\mathrm{Mp}_2(\mathbb{Z})$ depending on L. When $L = L_H$, by Deligne's bound ([Del73, Del74]), we have $|g(m)| \ll m^{1/2}\sigma_0(m) \ll_\epsilon m^{1/2+\epsilon} = o_\epsilon(m^{1-\epsilon}) = o(|q_L(m)|)$ for any $0 < \epsilon < 1/4$. When $L = L_S$, the trivial bound yields $|g(m)| \ll m^{5/4} = o(m^{3/2})$ (see [Sar90, Prop. 1.3.5]). Therefore by Theorem 4.1.1, $Z(m).C = -q_L(m)(\omega.C) + o(|q_L(m)|)$; in particular, for $m \gg 1$, $Z(m).C > 0$ and hence $Z(m) \neq \emptyset$. □

4.3.3. When $L = L_S$, recall from §2.1 that the quadratic form is $Q(x) = x_0^2 + x_1x_2 - x_3x_4$ and hence every $m \in \mathbb{Z}_{>0}$ is representable by (L, Q). In particular, $Z(m) \neq \emptyset$ and $\delta(\ell, L, m) > 0$ for all ℓ. Moreover, in order to prove Theorem 1(1) and Remark 2, we will work with $m \in T := \{Dq^2 \mid q \text{ prime and } q \neq p\}$, where we take $D = 1$ for Theorem 1(1) and D being the discriminant of the real quadratic field in Remark 2; and for Theorem 3, we work with $m \in T := \{q \mid q \text{ prime and } q \neq p, q \text{ is a quadratic residue mod } p, \text{ and } q \equiv 3 \bmod 4\}$. In particular, for all such m, we have $v_\ell(m) \leq 2 + v_\ell(D)$ and hence the assumptions in §4.3.1 are satisfied.

When $L = L_H$, since L is maximal and isotropic, we have that the quadratic form on $L \otimes \mathbb{Z}_\ell$ is given by $xy + Q_1(z)$, where $x, y \in \mathbb{Z}_\ell, z \in \mathbb{Z}_\ell^2$ and Q_1 is some quadratic form. Then $\delta(\ell, L, m) > 0$ for all ℓ; indeed, by [Han04, Def. 3.1, Lem. 3.2], $\delta(\ell, L, m) > 0$ if there exists $x, y \in \mathbb{Z}/\ell^{1+2v_\ell(2)}$ such that $xy \equiv m \bmod \ell^{1+2v_\ell(2)}$ and $x \not\equiv 0 \bmod \ell$ (by the terminology in [Han04], this construct a good type solution (taking $z = 0$) for $(L, Q) \bmod \ell^{1+2v_\ell(2)}$, which can be lifted to $\mathbb{Z}/\ell^k$ for any $k \geq 1 + 2v_\ell(2)$). Such x, y always exists and hence every $m \in \mathbb{Z}_{>0}$ is representable by $(L \otimes \mathbb{Z}_\ell, Q)$ for all ℓ and hence by Lemma 4.3.2, there exists $N \in \mathbb{Z}_{>0}$ such that for all $m > N$, m is representable by (L, Q). For the proof of Theorem 1(2), we work with m in

$$T := \{m \in \mathbb{Z} \mid m > N, p \nmid m, v_\ell(m) \leq C, \forall \ell \mid 2\det L, \text{ and } \exists\, q || m \text{ such that } q \text{ inert in } F\},$$

where F is the real quadratic field attached to the Hilbert modular surface and the constant C is chosen so that this set is non-empty. The existence of q implies that $m \neq \mathrm{Nm}_{F/\mathbb{Q}}\, \gamma$ for any $\gamma \in F$ and hence for any $v \in L_H \otimes \mathbb{Q}$ such that $Q(v) = m$, we have $v^\perp \subset L_H \otimes \mathbb{Q}$ is anisotropic. Note that if $Z(m)$ is non-compact in $\mathcal{M}_{\mathbb{F}_p}$, then $Z(m)$ parametrizes abelian surfaces which are isogenous to the self-product of elliptic curves and then $v^\perp$ is isotropic. Therefore, for any $m \in T$, we have that $Z(m)$ is compact in $\mathcal{M}_{\mathbb{F}_p}$. Note that $T \subset \mathbb{Z}_{>0}$ is of positive density.

Lemma 4.3.4. *For $L = L_H$ and $M > 0$, we have $\sum_{1 \leq m \leq M, m \in T} |q_L(m)| \asymp M^2$.*

Proof. By §§4.3.1,4.3.3, we have for $m \in T$, $|q_L(m)| \asymp m\sigma_{-1}(m, \chi)$, where $\chi = \chi_{4\det L}$. We write

$$\begin{aligned}
\sum_{1 \leq m \leq M, m \in T} m\sigma_{-1}(m, \chi) &= \sum_{1 \leq m \leq M, m \in T} \sum_{d \mid m} d \cdot \chi(m/d) = \sum_{1 \leq d \leq M, 1 \leq f \leq M, df \leq M, df \in T} d \cdot \chi(f) \\
&= \sum_{1 \leq d \leq M^{1/2}} d \sum_{1 \leq f \leq M/d, df \in T} \chi(f) + \sum_{1 \leq f \leq M^{1/2}} \chi(f) \sum_{1 \leq d \leq M/f, df \in T} d \\
&\quad -\Big(\sum_{1 \leq d \leq M^{1/2}} d\Big(\sum_{1 \leq f \leq M^{1/2}, df \in T} \chi(f)\Big)\Big).
\end{aligned}$$

Note that

$$\Big| \sum_{1 \leq d \leq M^{1/2}} d \sum_{1 \leq f \leq M/d, df \in T} \chi(f) \Big| \leq \sum_{1 \leq d \leq M^{1/2}} d \cdot (M/d) = O(M^{3/2}),$$

$$\Big|\Big(\sum_{1 \leq d \leq M^{1/2}} d\Big(\sum_{1 \leq f \leq M^{1/2}, df \in T} \chi(f)\Big)\Big)\Big| \leq \Big(\sum_{1 \leq d \leq M^{1/2}} d\Big) \cdot \Big(\sum_{1 \leq f \leq M^{1/2}} 1\Big) = O(M^{3/2}).$$

The second term is the main term. First let $T' := \{m \in \mathbb{Z} \mid m > N, p \nmid m, v_\ell(m) \leq C, \forall \ell \mid 2\det L\}$ then

$$\sum_{1\leq f\leq M^{1/2}} \chi(f) \sum_{1\leq d\leq M/f, df\in T'} d = \sum_{1\leq f\leq M^{1/2}, p\nmid f} \chi(f) \sum_{1\leq d\leq M/f, p\nmid d, v_\ell(d)\leq C, \forall \ell\mid 2\det L} d,$$

because $v_\ell(df) \leq C \iff v_\ell(d) \leq C, \forall \ell \mid 2\det L$ when $v_\ell(f) = 0, \forall \ell \mid 2\det L$ and if $v_\ell(f) > 0$ for some $\ell \mid 2\det L$, then $\chi(f) = 0$. Since $\sum_{1\leq d\leq M/f, p\nmid d, v_\ell(d)\leq C, \forall \ell\mid 2\det L} d = C_1 \frac{M^2}{f^2} + O(M)$, where C_1 and the implicit constant only depend on C, L, p. Hence

$$\sum_{1\leq f\leq M^{1/2}} \chi(f) \sum_{1\leq d\leq M/f, df\in T'} d = C_1 M^2 \sum_{1\leq f\leq M^{1/2}, p\nmid f} \chi(f)/f^2 + O(X^{3/2}) \asymp M^2.$$

To finish the proof, we only need to show that $|\sum_{1\leq f\leq M^{1/2}} \chi(f) \sum_{1\leq d\leq M/f, df\in T'-T} d| = o(M^2)$. Since $M/f \geq M^{1/2}$, by definition of T, $\#\{d \mid 1 \leq d \leq M/f, df \in T'-T\} = o(M/f)$ with implicit constant independent of f and hence we obtain the desired bound. $\square$

4.4. **Local densities at p and the ratios of Fourier coefficients.** We set the same notation as in §4.2.1. Theorem 4.1.6 and Theorem 4.2.2 reduce the comparison between $q_L(m)$ and $q_{L'''}(m)$ to the computation of the local density $\delta(p, L''', m)$, which we now compute following [Han04, §3]. Recall that p is an odd prime and $v_p(m) \leq 1$ for all $m \in T$ defined in §4.3.3. For an arbitrary quadratic lattice (L, Q), let $\alpha(p, L, m) := p^{1-\operatorname{rk} L}\#\{v \in L/pL \mid Q(v) \equiv m \bmod p\}$; if we diagonalize $L \otimes \mathbb{Z}_p$ such that Q is given by $\sum_{i=1}^{\operatorname{rk} L} a_i x_i^2$ with $a_i \in \mathbb{Z}_p$, then we define

$\alpha^*(p, L, m) := p^{1-\operatorname{rk} L}\#\{v = (x_1, \dots, x_{\operatorname{rk} L}) \in L/pL \mid Q(v) \equiv m, \exists i \text{ such that } v_p(a_i) = 0, x_i \not\equiv 0 \bmod p\}$.

Lemma 4.4.1 (Hanke). *If $p \nmid m$, we have*

$$\delta(p, L''', m) = \alpha(p, L''', m);$$

if $v_p(m) = 1$, we have

$$\delta(p, L''', m) = \alpha^*(p, L''', m) + p^{1-s_0}\alpha(p, L'''_I, m/p),$$

where if we write $(L''' \otimes \mathbb{Z}_p, Q')$ into diagonal form $\sum_{i=1}^{\operatorname{rk} L'''} a_i x_i^2$ with $a_i \in \mathbb{Z}_p$, we define $s_0 = \#\{a_i \mid v_p(a_i) = 0\}$ and L'''_I is the quadratic lattice with quadratic form $\sum_{i=1}^{\operatorname{rk} L'''} a'_i x_i^2$, where $a'_i = pa_i$ if $v_p(a_i) = 0$ and $a'_i = p^{-1}a_i$ if $v_p(a_i) \geq 1$.

Proof. If $p \nmid m$, the assertion follows from [Han04, Rmk. 3.4.1 (a), Lem. 3.2]; If $v_p(m) = 1$, then we only have good type and bad type I solutions in the sense of [Han04, Def. 3.1, p. 360] and the assertion follows from [Han04, Lem. 3.2, p. 360, Rmk. 3.4.1 (a)]. $\square$

We first compute $\delta(p, L', m)$ by Lemma 4.4.1. We always pick $\epsilon \in \mathbb{Z}_p^\times \backslash (\mathbb{Z}_p^\times)^2$ as in §3.1.2.

4.4.2. Consider $L = L_H$ and recall that $p \nmid m, \forall m \in T$. Let F denote the real quadratic field attached to the Hilbert modular surface defined by L_H.

(1) Assume that p is inert in F and P is supergeneric. By §3.2.1, $L' \otimes \mathbb{Z}_p = \Lambda^\vee = p\Lambda$ and hence $p \mid Q'(v), \forall v \in L'$; in particular, $\delta(p, L', m) = 0$.
(2) Assume that p is inert in F and P is superspecial. By §3.2.1, $Q'(v) = xy + p(z^2 - \epsilon w^2)$, where w_i are given right above (3.2.1) and $v = xw_3 + yw_4 + zw_1 + ww_2$ with $x, y, z, w \in \mathbb{Z}_p$. Hence $\delta(p, L', m) = \alpha(p, L', m) = 1 - 1/p$.
(3) Assume that p is split in F; hence P is superspecial. By §3.2.2, $L' \otimes \mathbb{Z}_p = \Lambda^\vee$ with $Q'(v) = x^2 - \epsilon y^2 - pz^2 + \epsilon p w^2$, where $v = xe_1 + ye_2 + z(pe_3) + w(pe_4)$ with $x, y, z, w \in \mathbb{Z}_p$. Hence $\delta(p, L', m) = \alpha(p, L', m) = 1 + 1/p$.

4.4.3. Consider $L = L_S$.

(1) Assume that P is superspecial. By §3.3, we have $Q'(v) = xy + \epsilon z^2 + pw^2 - p\epsilon u^2$, where $v = xw_3 + yw_4 + zw_5 + ww_2 + uw_1$ with $x, y, z, w, u \in \mathbb{Z}_p$ and w_i are given right above (3.3.1). Hence if $p \nmid m$, then $\delta(p, L', m) = \alpha(p, L', m) \leq 1 + 1/p$ by [Han04, Table 1]. If $v_p(m) = 1$, then the quadratic form of L'_I is $p(xy + \epsilon z^2) + w^2 - \epsilon u^2$ and hence $\delta(p, L', m) = \alpha^*(p, L', m) + p^{-2}\alpha(p, L'_I, m/p) = (1 - p^{-2}) + p^{-2}(1 + p^{-1}) = 1 + p^{-3}$.

(2) Assume that P is supergeneric. By §3.3, $L' \otimes \mathbb{Z}_p = \Lambda^\vee$ and hence the quadratic form is $pxy + \epsilon z^2 + pw^2 - p\epsilon u^2$. If $p \nmid m$, then $\delta(p, L', m) = \alpha(p, L', m) = 0$ or 2; if $v_p(m) = 1$, then the quadratic form of L'_I is $p\epsilon z^2 + xy + w^2 - \epsilon u^2$ and hence $\delta(p, L', m) = \alpha^*(p, L', m) + \alpha(p, L'_I, m/p) = 0 + 1 + p^{-2} = 1 + p^{-2}$ by [Han04, Table 1].

We now estimate $\delta(p, L''', m)$ for sublattices lattices L''' of L' defined in §4.2.1.

Lemma 4.4.4. *If $p \nmid m$, then $\delta(p, L''', m) \leq 2$.*

Proof. By Lemma 4.4.1, $\delta(p, L''', m) = \alpha(p, L''', m)$. Write the quadratic form Q' on L''' into the diagonal form $\sum_{i=1}^{\operatorname{rk} L'''} a_i x_i^2$ with $a_i \in \mathbb{Z}_p$ and we may assume that there exists a_i such that $p \nmid a_i$; otherwise $\delta(p, L''', m) = 0$ then we are done. Now let $\widetilde{L}'''$ denote the quadratic form $\sum_{1 \leq i \leq \operatorname{rk} L''', p \nmid a_i} a_i x_i^2$. Then by definition, $\alpha(p, L''', m) = \alpha(p, \widetilde{L}''', m)$. Since $p \mid \operatorname{disc} L'$, then $p \mid \operatorname{disc} L'''$ and $\operatorname{rk} \widetilde{L'''} \leq \operatorname{rk} L''' - 1 \leq 4$. Then by [Han04, Table 1], $\alpha(p, \widetilde{L}''', m) \leq 2$ and hence $\delta(p, L''', m) \leq 2$. $\square$

Lemma 4.4.5. *Assume that $L = L_S$ and $v_p(m) = 1$. We have $\delta(p, L''', m) \leq 2 + 2p$. Moreover, if P is superspecial and $[L' : L'''] = p$, then $\delta(p, L''', m) \leq 4$.*

Proof. By Lemma 4.4.1, $\delta(p, L''', m) = \alpha^*(p, L''', m) + p^{1-s_0}\alpha(p, L'''_I, m/p) \leq \alpha(p, L''', m) + p\alpha(p, L'''_I, m/p)$. By the proof of Lemma 4.4.4, we have $\alpha(p, L''', m) = \alpha(p, \widetilde{L}''', m) \leq 2$. The same argument implies that $\alpha(p, L'''_I, m/p) \leq 2$ if $\operatorname{rk}(\widetilde{L}''') \leq 4$. If $\operatorname{rk}(\widetilde{L}''') = 5$, then it is isotropic and we write the quadratic form as $xy + Q_1(z)$. The equation $xy + Q_1(z) \equiv (m/p) \bmod p$ has $(p-1)p^3$ solutions in $\mathbb{F}_p^5$ with $x \neq 0$ and has at most p^4 solutions with $x = 0$. Hence $\alpha(p, L''', m/p) = \alpha(p, \widetilde{L}''', m/p) < 2$. Therefore, $\delta(p, L''', m) \leq 2 + 2p$.

If P is superspecial and $[L' : L'''] = p$, then $s_0 \geq 1$ and hence $\delta(p, L''', m) \leq \alpha^*(p, L''', m) + \alpha(p, L'''_I, m/p) \leq 4$. $\square$

The following lemma, which is the main goal of this subsection, will be used to compare the local intersection number at a supersingular point P with the global intersection number.

Lemma 4.4.6. *Notation as in §4.2.1 and consider $m \in T$ (defined in §4.3.3).*

(1) If P is superspecial or $L = L_H$, then $\dfrac{q(m)_{L'}}{-q(m)_L} \leq \dfrac{1}{p-1}$.

(2) If $L = L_S$ and P is supergeneric, then $\dfrac{q(m)_{L'}}{-q(m)_L} \leq \dfrac{2}{p^2-1}$.

(3) If $p \nmid m$, then $\dfrac{q(m)_{L'''}}{-q(m)_L} \leq \dfrac{2}{\sqrt{|(L''' \otimes \mathbb{Z}_p)^\vee/(L''' \otimes \mathbb{Z}_p)|}(1-p^{-2})}$.

(4) Assumption as in Lemma 4.4.5, then $\dfrac{q(m)_{L'''}}{-q(m)_L} \leq \dfrac{2p}{\sqrt{|(L''' \otimes \mathbb{Z}_p)^\vee/(L''' \otimes \mathbb{Z}_p)|}(1-p^{-1})}$*; moreover, if P is superspecial and $[L' : L'''] = p$, then* $\dfrac{q(m)_{L'''}}{-q(m)_L} \leq \dfrac{4}{p^2-1}$.

Proof. Recall from §4.2.1 that $L''' \otimes \mathbb{Z}_\ell \cong L \otimes \mathbb{Z}_\ell, \forall \ell \neq p$; hence for $\ell \neq p$, we have $\delta(p, L''', m) = \delta(p, L, m)$ and $\det L''' = p^k \det L$ for some $k \in \mathbb{Z}_{\geq 0}$. Since L is self-dual at p, then $p \nmid \det L$; by §3.1.2,

$\det L' = p^{2b}\det L$ for some $b \in \mathbb{Z}_{>0}$ (concretely, one may deduce this fact by the explicit formula of Q' in §§4.4.2-4.4.3) and hence $k \in 2\mathbb{Z}_{>0}$. Thus $\chi_{4\det L}(d) = \chi_{4\det L'}(d)$ and $\chi_{-2m_0\det L}(d) = \chi_{-2m_0\det L'}(d)$ if $p \nmid d$.

Therefore, by Theorem 4.1.6 and Theorem 4.2.2, we have that for $L = L_H$, $p \nmid m$

$$\frac{q(m)_{L'''}}{-q(m)_L} = \frac{\delta(p, L''', m)}{\sqrt{|(L''' \otimes \mathbb{Z}_p)^\vee/(L''' \otimes \mathbb{Z}_p)|}(1-\chi_{4\det L}(p)p^{-2})} \leq \frac{\delta(p, L''', m)}{\sqrt{|(L''' \otimes \mathbb{Z}_p)^\vee/(L''' \otimes \mathbb{Z}_p)|}(1-p^{-2})};$$

for $L = L_S$, $v_p(m) \leq 1$, we observe that m_0 remains the same for L and L''' and $p \nmid f$ and hence

$$\frac{q(m)_{L'''}}{-q(m)_L} = \frac{\delta(p, L''', m)(1-\chi_{\mathcal{D}}(p)p^{-2})}{\sqrt{|(L''' \otimes \mathbb{Z}_p)^\vee/(L''' \otimes \mathbb{Z}_p)|}(1-p^{-4})} \leq \frac{\delta(p, L''', m)}{\sqrt{|(L''' \otimes \mathbb{Z}_p)^\vee/(L''' \otimes \mathbb{Z}_p)|}(1-p^{-2})}.$$

Therefore, (1)(2) follow from §§4.4.2-4.4.3; (3) follows from Lemma 4.4.4; (4) follows from Lemma 4.4.5. □

5. The decay lemma for supersingular points and its proof in the Hilbert case

The goal of this section is to prove that special endomorphisms "decay rapidly". More precisely, consider a generically ordinary two-dimensional abelian scheme over $\bar{\mathbb{F}}_p[[t]]$ whose special fiber is supersingular. We consider the lattice of special endomorphisms of the abelian scheme mod t^N as N varies, and establish bounds for the covolume of these lattices. These bounds are exactly what we need to bound the local intersection multiplicity $\operatorname{Spf}\bar{\mathbb{F}}_p[[t]] \cdot Z(m)$ – see Lemma 7.2.1. The precise definitions and results are in Definition 5.1.1 and Theorem 5.1.2.

Throughout this section, as in §3, $k = \bar{\mathbb{F}}_p$, $W = W(k)$, $K = W[1/p]$. We focus on the behavior of the curve C in Theorems 1 and 3 in a formal neighborhood of a supersingular point P, so we may let $C = \operatorname{Spf} k[[t]]$ denote a generically ordinary formal curve in $\mathcal{M}_k$ which specializes to P. As in §3.1.5, σ denote both the Frobenius on K and the Frobenius on the coordinate rings $W[[x,y]], W[[x,y,z]]$ of $\widehat{\mathcal{M}}_P$, which is the unique extension of the Frobenius action on W for which $\sigma(x) = x^p, \sigma(y) = y^p, \sigma(z) = z^p$. For a matrix M with entries in $K[[x,y]]$ or $K[[x,y,z]]$, we use $M^{(n)}$ to denote $\sigma^n(M)$. Also recall we set $\lambda \in \mathbb{Z}_{p^2}^\times$ such that $\sigma(\lambda) = -\lambda$. We use σ_t to denote the Frobenius on $K[[t]]$ which extends σ on K and sends t to t^p.

5.1. Statement of the Decay Lemma and the first reduction step.

The map $C \to \mathcal{M}_k$ gives rise to a local ring homomorphism from $k[[x,y]] \to k[[t]]$ (in the Hilbert case) or $k[[x,y,z]] \to k[[t]]$ (in the Siegel case), and we denote by $x(t)$, $y(t)$, and $z(t)$ the images of x, y, and z respectively. Let v_t denote the t-adic valuation map on $k[[t]]$. Let A denote the t-adic valuation of the local equation defining the non-ordinary locus in Corollary 3.4.2. More precisely, if P superspecial, then $A = v_t(xy)$ in the Hilbert case and $A = v_t(xy + \frac{z^2}{4\epsilon})$ in the Siegel case.

Definition 5.1.1. Let w denote a special endomorphism of the p-divisible group at P (i.e., w is an element in $L' \otimes \mathbb{Z}_p$; see Definition 2.2.4 and Definition 2.2.9).

(1) We say that w *decays rapidly* if $p^n w$ does not lift to an endomorphism modulo t^{A_n+1} for all $n \in \mathbb{Z}_{\geq 0}$, where $A_n := [A(p^n + p^{n-1} + \cdots + 1 + \frac{1}{p})]$; here $[x]$ denote the maximal integer y such that $y \leq x$.
(2) We say that a $\mathbb{Z}_p$-submodule of $L' \otimes \mathbb{Z}_p$ *decays rapidly* if every primitive vector in the submodule decays rapidly.
(3) We say that w decays *very* rapidly if $p^n w$ does not lift to an endomorphism modulo $t^{A_{n-1}+ap^n+1}$ for some constant $a \leq A/2$, for all $n \in \mathbb{Z}_{\geq 0}$, where A_n is defined in (1) and we define $A_{-1} = [A/p]$.

Theorem 5.1.2 (Decay Lemma)**.** *Assume P is superspecial. There exists a rank* 3 *$\mathbb{Z}_p$-submodule of $L' \otimes \mathbb{Z}_p$ which decays rapidly and furthermore, there is a primitive vector in this submodule which decays* very *rapidly.*

Here we only state the decay lemma for a superspecial point since we do not need to work with supergeneric points to prove Theorems 1 and 3. We refer the reader to the appendix for a decay lemma when P is supergeneric.

Proof assuming Proposition 5.1.3. For $m \in \mathbb{Z}_{\geq 0}$, let S_m denote $\operatorname{Spec} k[t]/(t^m)$ and let D_m denote the p-adic completion of the PD enveloping algebra of the ideal (t^m, p) in $W[[t]]$. Let ι_m denote the composite map $S_m \to \operatorname{Spf} k[[t]] \to \operatorname{Spf} k[[x, y]]$ or $\operatorname{Spf} k[[x, y, z]]$. Then by [dJ95, §2.3], there exists a functor from the category of p-divisible groups over S_m to the category Dieudonné modules over D_m. More precisely, a special endomorphism $\tilde{w}_m$ of the p-divisible group over S_m which specializes to $w \in L' \otimes \mathbb{Z}_p$ gives rise to an endomorphism of the Dieudonné module which specializes to w. By functoriality of Dieudonné modules, images of special endomorphisms are horizontal sections of $\iota_m^* \mathbb{L}_{\mathrm{cris}}(D_m)$ stable under the Frobenius action; here the connection on $\iota_m^* \mathbb{L}_{\mathrm{cris}}(D_m)$ is the pull-back of the connection on $\mathbb{L}_{\mathrm{cris}}(W[[x, y]]), \mathbb{L}_{\mathrm{cris}}(W[[x, y, z]])$ by a ring homomorphism $W[[x, y]] \to W[[t]]$ or $W[[x, y, z]] \to W[[t]]$ which lifts $k[[x, y]] \to k[[t]]$ or $k[[x, y, z]] \to k[[t]]$ given by C[22] and the σ_t-linear Frobenius is given in [Moo98, §4.3.3].[23]

The connection on $\mathbb{L}_{\mathrm{cris}}(W[[x, y]])$ or $\mathbb{L}_{\mathrm{cris}}(W[[x, y, z]])$ gives rise to a connection on $\mathbb{L}_{\mathrm{cris},P}(W) \otimes_W K[[x, y]] \supset \mathbb{L}_{\mathrm{cris}}(W[[x, y]]$ or $\mathbb{L}_{\mathrm{cris},P}(W) \otimes_W K[[x, y, z]] \supset \mathbb{L}_{\mathrm{cris}}(W[[x, y, z]]$. Let $\tilde{w}$ denote the horizontal section in $\mathbb{L}_{\mathrm{cris},P}(W) \otimes_W K[[x, y]]$ or $\mathbb{L}_{\mathrm{cris},P}(W) \otimes_W K[[x, y, z]]$ extending $w \in L' \otimes \mathbb{Z}_p \subset \mathbb{L}_{\mathrm{cris},P}(W)$. Since the image of $\tilde{w}_m$ in $\iota_m^* \mathbb{L}_{\mathrm{cris}}(D_m)$ is horizontal and the connection on $\iota_m^* \mathbb{L}_{\mathrm{cris}}(D_m)$ is the pull-back connection, then $\tilde{w}_m = \iota_m^* \tilde{w}$. Therefore, if w lift to a special endomorphism in S_m, then $\iota_m^* \tilde{w} \in \iota_m^* \mathbb{L}_{\mathrm{cris}}(D_m) \subset \mathbb{L}_{\mathrm{cris},P}(W) \otimes_W K[[t]]$.

The section $\tilde{w}$ is constructed in [Kis10, §1.5.5] as follows. Recall from §§3.2-3.3, the Frobenius on $\mathbb{L}_{\mathrm{cris}}(W[[x, y]]), \mathbb{L}_{\mathrm{cris}}(W[[x, y, z]])$, with respect to a φ-invariant basis $\{w_i\}$, is given by $(I + F) \circ \sigma$ for some matrix F with entries in $(x, y)K[x, y]$ or $(x, y, z)K[x, y, z]$. We define F_∞ to be the infinite product $\prod_{i=0}^{\infty}(1 + F^{(i)})$, where $F^{(i)}$ is the i-th σ-twist of F (recall $\sigma(x) = x^p, \sigma(y) = y^p, \sigma(z) = z^p$). Since $v_t(y), v_t(x), v_t(z) \geq 1$, the product is well-defined and the entries of F_∞ are power series valued in $K[[t]]$. The $\mathbb{Q}_p$-span of the columns of F_∞ are vectors of $\mathbb{L}_{\mathrm{cris},P}(W) \otimes K[[x, y]], \mathbb{L}_{\mathrm{cris},P}(W) \otimes K[[x, y, z]]$ which are Frobenius stable and horizontal.

Now we are ready to reduce to the proof of the decay lemma to the following proposition. Indeed, by Proposition 5.1.3, with respect to $\{w_i\}$, there exists a rank 3 $\mathbb{Z}_p$-submodule of $L' \otimes \mathbb{Z}_p$ such that for every primitive w in this submodule, the coefficient of t^{k_n} for some $k_n \leq A(1 + p + \cdots + p^{n+1})$ in $p^n \tilde{w}$ does not lie in $(p^{-1}W)^4$; since $p\mathbb{L}_{\mathrm{cris},P}(W) \subset L' \otimes W$, with respect to a W-basis of $\mathbb{L}_{\mathrm{cris},P}(W)$, the coefficient of t^{k_n} in $p^n \tilde{w}$ does not lie in W^4. On the other hand, for any $N < p(A_n + 1)$, we have $p^{-1} t^N \notin D_{A_n+1}$. Note that $p(A_n + 1) > pA(p^n + \cdots + 1/p) = A(p^{n+1} + \cdots + 1) \geq k_n$. Hence $p^n \tilde{w}$ does not extend to a special endomorphism over S_{A_n+1}. Thus, this rank 3 submodule decays rapidly. Moreover, the existence of a vector decaying very rapidly follows by the second assertion of Proposition 5.1.3 via the same argument and the fact that $p(A_{n-1} + ap^n + 1) > p(A(p^{n-1} + \cdots + 1/p) + ap^n) = A(p^n + \cdots + 1) + ap^{n+1}$. □

[22]We may pick a lift $k \to W$, for instance, the Teichmüller lift and hence view $x(t), y(t), z(t)$ as power series in $W[[t]]$.

[23]Here we refer to [Moo98] for the existence of an explicit formula of the σ_t-linear Frobenius, but we do not need this explicit formula for our purpose. We will always carry out our computation using the σ-linear Frobenius; see the rest of the proof for the details.

Proposition 5.1.3. *Assume P is superspecial. With respect to the w_i-basis in §§3.2-3.3, there exists a rank 3 $\mathbb{Z}_p$-submodule of $L' \otimes \mathbb{Z}_p$ such that for every primitive w in this submodule, the coefficients of $1 = t^0, \dots, t^{A(1+p+\dots+p^n)}$ in the power series $p^n\tilde{w} \in (K[[t]])^4$ do not all lie in W^4 for all $n \in \mathbb{Z}_{\geq 0}$ (property DR); moreover, there exist $a \leq A/2$ and a primitive w in the rank 3 submodule such that the coefficients of $1, \dots, t^{A(1+p+\dots+p^{n-1})+ap^n}$ in $p^n\tilde{w} \in (K[[t]])^4$ do not all lie in W^4 for all $n \in \mathbb{Z}_{\geq 0}$ (property DvR).*

By a slight abuse of terminology, if a submodule of $L' \otimes \mathbb{Z}_p$ satisfies the property DR (with respect to basis $\{w_i\}$), we also say that this submodule *decays rapidly*; if a primitive vector satisfies property DvR, we also say that this vector *decays very rapidly.* By the proof of Theorem 5.1.2 above, property DR (resp. DvR) implies decaying (resp. very) rapidly in the sense of Definition 5.1.1.

The rest of this section is devoted to prove Proposition 5.1.3 for the Hilbert case and its proof for the Siegel case is given in §6. In the following, the split/inert case means that p is split/inert in the real quadratic field attached to the Hilbert modular surface.

In the Hilbert case, by Corollary 3.4.2, the non-ordinary locus is cut out by the equation $xy = 0$. As in the proof of reducing Theorem 5.1.2 to Proposition 5.1.3, we pick a lift $W[[x, y]] \to W[[t]]$ of of the local ring homomorphism $k[[x, y]] \to k[[t]]$ defined by C. Since C is generically ordinary, we have that both x and y map to non-zero power series in $W[[t]]$. Without loss of generality, we assume that $v_t(x) \leq v_t(y)$, and that $x(t) = t^a + \dots$ and $y(t) = \alpha t^b + \dots$, where $\alpha \in W^\times$.

5.2. **Decay in the split case.** Notation as in the proof of Theorem 5.1.2. We first compute $F_\infty = \prod_{i=0}^{\infty}(1 + F^{(i)})$, where by (3.2.2),

$$F = \begin{bmatrix} \frac{xy}{2p} & -\frac{\lambda xy}{2p} & \frac{x+y}{2p} & \frac{-\lambda(x-y)}{2p} \\ \frac{xy}{2\lambda p} & -\frac{xy}{2p} & \frac{x+y}{2\lambda p} & \frac{-(x-y)}{2p} \\ \frac{x+y}{2} & \frac{-\lambda(x+y)}{2} & 0 & 0 \\ \frac{x-y}{2\lambda} & \frac{-(x-y)}{2} & 0 & 0 \end{bmatrix}.$$

Let $F_\infty(1)$ and $F_\infty(2)$ denote the top-left and top-right 2×2 blocks of F_∞ respectively. To simplify the notation, define[24]

$$G = \begin{bmatrix} \frac{1}{2} & \frac{-\lambda}{2} \\ \frac{1}{2\lambda} & \frac{-1}{2} \end{bmatrix}, H_u = \begin{bmatrix} \frac{1}{2} & \frac{-\lambda}{2} \\ \frac{1}{2\lambda} & \frac{-1}{2} \end{bmatrix}, H_l = \begin{bmatrix} \frac{1}{2} & \frac{-\lambda}{2} \\ \frac{1}{2\lambda} & \frac{-1}{2} \end{bmatrix},$$

and let F_t, F_u and F_l denote the top-left, top-right, and bottom-left 2×2 blocks of F. The following elementary lemma picks out the terms in $F_\infty(1), F_\infty(2)$ with the desired p-power on the denominators.

Lemma 5.2.1. *(1) The part of $F_\infty(1)$ with p-adic valuation $-(n+1)$ consists of sums of products of the form $\prod_{i=0}^{m_1+2m_2} X_i^{(n_i)}$. Here X_i is either F_t, F_u or F_l,[25] $m_1 + 1$ is the number of occurrences of F_t, and m_2 is the number of occurrences of the pair F_u, F_l, $m_1 + m_2 = n$, and n_i is a strictly increasing sequence of non-negative integers. The analogous statement holds for $F_\infty(2)$ as well.*

[24] These three matrices are the same; however, we use different notations to be consistent with the proof for the Siegel case in §6.

[25] The terms X_i are chosen so that the product makes sense, and has the right size. Note that this would imply that F_u, F_l must occur in consecutive pairs.

(2) Fix values of m_1, m_2 as above. Among all the terms in the above sum, the ones with minimal t-adic valuation only occur when $n_i = i$, and either when $X_0 = X_1 = \ldots = X_{m_1} = F_t$, or $X_0 = X_2 = \ldots = X_{2m_2-2} = F_u$. The analogous statement holds for $F_\infty(2)$ as well.

(3) (for $F_\infty(1)$) The product $\prod_{i=0}^{m_1} F_t^{(i)} \prod_{i=0}^{m_2-1} F_u^{(m_1+2i+1)} F_l^{(m_1+2i+2)}$ (modulo terms with smaller p-power in denominators[26]) equals

$$\frac{1}{p^{n+1}} \prod_{i=0}^{m_1} G^{(i)} (xy)^{(i)} \prod_{i=0}^{m_2-1} H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)} (x^{1+p} + y^{1+p})^{(m_1+2i+1)}.$$

(4) (for $F_\infty(2)$) The product $\prod_{i=0}^{m_1} F_t^{(i)} \prod_{i=0}^{m_2-1} F_u^{(m_1+2i+1)} F_l^{(m_1+2i+2)} \cdot F_u^{(m_1+2m_2+1)}$ (modulo terms with smaller p-power in denominators) equals

$$\frac{1}{p^{n+2}} \prod_{i=0}^{m_1} G^{(i)} (xy)^{(i)} \prod_{i=0}^{m_2-1} H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)} (x^{1+p} + y^{1+p})^{(m_1+2i+1)} \cdot F_u^{(m_1+2m_2+1)}$$

5.2.2. **Notations.** We make the following definition to further lighten the notation.

Let $P(1)_{m_2,n}$ denote the product

$$\prod_{i=0}^{m_1} G^{(i)} \prod_{i=0}^{m_2-1} H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)}.$$

Recall that $A = a + b$ denotes the t-adic valuation $v_t(xy)$ of xy and let B denote $v_t(x^{p+1} + y^{p+1})$. Note that $B \geq a(p+1)$ and the equality holds unless $a = b$.

In order to prove Proposition 5.1.3, we will consider the following case-by-case analysis depending on the relation between a and b. The following elementary lemmas will be used in the case-by-case analysis.

Lemma 5.2.3. *Let n, e, f be in $\mathbb{Z}_{\geq 0}$.*

(1) The kernel of the 2×2 matrix $P(1)_{e,n}$ modulo p is defined over $\mathbb{F}_{p^2}$ but not over $\mathbb{F}_p$.

(2) The reductions of $P(1)_{e,n}$ and $P(1)_{f,n}$ modulo p are not scalar multiples (over k) of each other if $e \not\equiv f \mod 2$. In particular, these reductions are not scalar multiples of each other if $f = e \pm 1$.

Proof. As the entries of G, H_u and H_l are all in $W(\mathbb{F}_{p^2})[1/p]$, it follows that $G^{(2m)} = G$ and $G^{(2m+1)} = G^{(1)}$ (and the analogous statements hold for H_u and H_l). A direct computation shows that $GG^{(1)}G = G$, $H_u H_l^{(1)} H_u H_l^{(1)} = H_u H_l^{(1)}$, and $H_u^{(1)} H_l H_u^{(1)} H_l = H_u^{(1)} H_l$. Therefore, if $n - e$ is odd, then $P(1)_{e,n}$ simplifies to either $GG^{(1)} H_u H_l^{(1)}$, $GG^{(1)}$ or $H_u H_l^{(1)}$; if $n - e$ is even, $P(1)_{e,n}$ simplifies to G or $GH_u^{(1)} H_l$. A direct computation shows that the matrices $GG^{(1)}$, $H_u H_l^{(1)}$ and $GG^{(1)} H_u H_l^{(1)}$ (resp. G and $GH_u^{(1)} H_l$) are equal to

$$\begin{bmatrix} \frac{1}{2} & \frac{\lambda}{2} \\ \frac{1}{2\lambda} & \frac{1}{2} \end{bmatrix} \left(\text{resp. } \begin{bmatrix} \frac{1}{2} & \frac{-\lambda}{2} \\ \frac{1}{2\lambda} & \frac{-1}{2} \end{bmatrix} \right).$$

In either case, since $\lambda \in W(\mathbb{F}_{p^2}) \backslash \mathbb{Z}_p$, there is no non-trivial $\mathbb{F}_p$-linear combination of the columns modulo p which equals zero; this implies part (1). Furthermore, the above matrices are clearly not scalar multiples of each other, whence part (2) follows. □

[26] We use here that $x^p \pm y^p \equiv (x \pm y)^p \mod p$.

Lemma 5.2.4. *Let n, e, f be in $\mathbb{Z}_{\geq 0}$.*

(1) *The kernel of the 2×2 matrix $P(1)_{e,n-1} \cdot H_u^{(n+e)}$ modulo p is defined over $\mathbb{F}_{p^2}$ but not $\mathbb{F}_p$.*

(2) *The reductions of $P(1)_{e,n-1} \cdot H_u^{(n+e)}$ and $P(1)_{f,n-1} \cdot H_u^{(n+f)}$ modulo p are not scalar multiples of each other if $e \not\equiv f \mod 2$. In particular, these reductions are not scalar multiples of each other if $f = e \pm 1$.*

Proof. We argue along the lines of the proof of Lemma 5.2.3. Indeed, if $n - e$ is odd (resp. even), we are reduced to the cases of $GG^{(1)}H_uH_l^{(1)}H_u$, $GG^{(1)}H_u$, $H_uH_l^{(1)}H_u$, and H_u (resp. $GH_u^{(1)}H_lH_u^{(1)}$ and $GH_u^{(1)}$). The rest of the argument is similar. $\square$

We now prove Proposition 5.1.3 when p is split in the real quadratic field defining the Hilbert modular surface. The proof is a case-by-case study in the following four cases based on the relation of $a = v_t(x)$ and $b = v_t(y)$. The idea is to pick out the term(s) with minimal t-adic valuation among all the terms with the same p-power denominators given in Lemma 5.2.1. Case 4 is the generic case and it is easy to pick out such terms so we give the proof directly. In Cases 1-3, we first state the lemmas on the terms with minimal t-adic valuation and then prove the decay lemma. For the convenience of the reader, we summarize the desired vectors which decay rapidly enough at the beginning of each case.

Case 1: $a = b$. Recall that $A = v_t(xy) = a + b = 2a$.

We will prove that every vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_i\}$ decays rapidly, where $w_i = w_4$ if the t-adic valuation of $x - y$ is $> a$, and $w_i = w_3$ otherwise. Moreover, w_i, $i = 3, 4$ respectively, decays very rapidly.

Lemma 5.2.5. (1) *Among the terms appearing in $F_\infty(1)$ described in Lemma 5.2.1 with denominator p^{n+1}, the unique term with minimal t-adic valuation is*

$$P(1)_{0,n}(xy)^{1+p+\ldots+p^n}.$$

(2) *Among the terms appearing in $F_\infty(2)$ described in Lemma 5.2.1 with denominator p^{n+1}, the unique term with minimal t-adic valuation is*

$$P(1)_{0,n-1} \cdot F_u^{(n)}(xy)^{1+p+\ldots+p^{n-1}}.$$

This lemma follows directly from Lemma 5.2.1 and the assumption that $a = b$.

Proof of Proposition 5.1.3 in this case. We first prove that every primitive vector $w \in \mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ decays rapidly. Indeed, write $w = cw_1 + dw_2$, by Lemma 5.2.3(1) and Lemma 5.2.5(1), there is a unique (non-vanishing) term in $F_\infty(1)w$ with denominator $1/p^{n+1}$ and minimal t-adic valuation $A(1 + p + \cdots + p^n)$ given by $P(1)_{0,n}[c \quad d]^T(xy)^{1+p+\cdots+p^n}$. Hence, modulo $t^{A(1+p+\cdots+p^n)+1}$, the horizontal section $p^n\tilde{w} = F_\infty(p^n w)$ does not lie in $W[[t]]$ and hence w decays rapidly.

Secondly, let $i \in \{3, 4\}$ be defined as above and we show that w_i decays *very* rapidly. Note that our definition of w_i implies that the first two entries of the i^{th} row of F have t-adic valuation equalling a. Furthermore, by Lemma 5.2.3(1), $P(1)_{0,n-1} \cdot v \neq 0 \bmod p$, where v is the n^{th} Frobenius twist of either column of H_u. Therefore, among the terms in the i^{th} column of F_∞ with denominator p^{n+1}, the term with minimal t-adic valuation has t-adic valuation $2a(1+p+\ldots+p^{n-1})+ap^n$. Hence w_i decays very rapidly since $a \leq (2a)/2 = A/2$.

Finally, we show that every vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_i\}$ decays rapidly. Let w_u denote a primitive vector in the span of w_1, w_2. It suffices to show that every vector which either has the form $p^m w_u + w_i$ or $w_u + p^m w_i$ decays rapidly, where $m \geq 0$. We first prove that every vector which has the form $p^m w_u + w_i$ decays rapidly where $m \geq 0$. Indeed, consider the two-dimensional vector whose entries are the first two entries of $F_\infty \cdot p^m w_u$. The t-adic valuation of the coefficient of $1/p^{n+1}$ equals $2a(1+p+\ldots+p^{n+m})$. Similarly, consider the two-dimensional vector whose entries are the first two entries of $F_\infty \cdot w_i$. The t-adic valuation of the coefficient of $1/p^{n+1}$ equals $2a(1+p+\ldots+p^{n-1})+ap^n$. Regardless of the value of m, the latter quantity is always smaller than the former quantity, whence it follows that $p^m w_u + w$ decays rapidly. Now, consider a vector of the form $w_u + p^m w_i$, where $m > 0$. Analogous to the previous case, consider the two-dimensional vector whose entries are the first two entries of $F_\infty \cdot w_u$. The t-adic valuation of the sum of all terms with denominator p^{n+1} equals $2a(1 + p + \ldots + p^n)$. Similarly, consider the two-dimensional vector whose entries are the first two entries of $F_\infty \cdot p^m w_i$. The t-adic valuation of the coefficient of $1/p^{n+1}$ equals $2a(1+p+\ldots+p^{n+m-1})+ap^{n+m}$. Regardless of the value of m (recall that $m > 0$), the latter quantity is always greater than the former quantity, whence it follows that $p^m w_u + w$ decays rapidly. $\square$

Case 2: $b = p^{2e}a$ **for some** $e \in \mathbb{Z}_{\geq 1}$**.** We will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w\}$ decays rapidly where w is some primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$. We will further prove that w decays *very* rapidly.

Lemma 5.2.6. *(1) Among the terms appearing in $F_\infty(1)$ described in Lemma 5.2.1 with denominator p^{n+1}, the unique term with minimal t-adic valuation is*

$$P(1)_{e,n}(xy)^{1+p+\ldots+p^{n-e}}x^{p^{n-e+1}+p^{n-e+2}+\ldots+p^{n+e}}.$$

(2) Among the terms appearing in $F_\infty(2)$ described in Lemma 5.2.1 with denominator p^{n+1}, there are exactly two terms with minimal t-adic valuation, and they are

$$P(1)_{e,n-1} \cdot F_u^{(n+e-1)}(xy)^{1+p+\ldots+p^{n-e-1}}x^{p^{n-e}+p^{n-e+1}+\ldots+p^{n+e-2}}, \text{ and}$$

$$P(1)_{e+1,n-1} \cdot F_u^{(n+e)}(xy)^{1+p+\ldots+p^{n-e-2}}x^{p^{n-e-1}+p^{n-e}+\ldots+p^{n+e-1}}.$$

Proof. In the following, we will prove part (1); part (2) will follow by an identical argument.

Note that the t-adic valuation of all the entries of $F(1)$ is $a + b$, and the t-adic valuation of the entries of F_u and F_l is a . Let k, l be in $\mathbb{Z}_{\geq 0}$ such that $k + l = n + 1$. Consider the following terms of $F_\infty(1)$ with denominator exactly p^{n+1}:

$$X_{k,l} := F(1) \cdot F(1)^{(1)} \ldots \cdot F(1)^{(k-1)} \cdot F_u^{(k)} F_l^{(k+1)} \ldots F_u^{(k+2l-2)} F_l^{(k+2l-1)}.$$

Similar to Lemma 5.2.1(2), we observe that among all the terms of $F_\infty(1)$ with denominator exactly p^{n+1} given in Lemma 5.2.1(1), for each other term X not listed above, there exists at least one $X_{k,l}$ (as k and l vary over all non-negative integers constrained by $k + l = n + 1$) such that $v_t(X_{k,l}) < v_t(X)$. Therefore, to prove (1), it suffices to show that $v_t(X_{k,l})$ with $k = n - e + 1$ and $l = e$ is less than $v_t(X_{k,l})$ with any other choice of k, l.

Since $b = ap^{2e}$ and $k + l = n + 1$, then $f(k) := v_t(X_{k,n}) = a\left((1 + p^{2e})\frac{p^k-1}{p-1} + \frac{p^{2(n-k+1)}-1}{p-1}p^k\right)$, and we need to prove that $k = n - e + 1$ minimizes this expression as k ranges over $\mathbb{Z} \cap [0, n + 1]$. Note that if we allow k to take all real values in the interval $[0, n + 1]$, a direct computation shows that f is convex (i.e., $f''(k) > 0$). Therefore, it suffices to show that $f(n - e + 1) < f(n - e)$ and $f(n - e + 1) < f(n - e + 2)$. These claims can be verified directly and hence we prove (1). $\square$

Proof of Proposition 5.1.3 in this case. We first prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ decays rapidly. Indeed, let w' be a primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$. Lemma 5.2.3(1) implies that $P(1)_{e,n} \cdot w' \bmod p$ is non-zero. This fact taken in conjunction with Lemma 5.2.6(1) yields that w' decays rapidly.

Secondly, we prove that there exists a primitive vector $w \in \mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ (independent of n) which decays very rapidly. Set $Y_{e,n} := P(1)_{e,n-1} \cdot F_u^{(n+e-1)}(xy)^{1+p+\ldots+p^{n-e-1}}x^{p^{n-e}+p^{n-e+1}+\ldots+p^{n+e-2}} + P(1)_{e+1,n-1} \cdot F_u^{(n+e)}(xy)^{1+p+\ldots+p^{n-e-2}}x^{p^{n-e-1}+p^{n-e}+\ldots+p^{n+e-1}}$, which is the sum of the two terms with minimal t-adic valuation listed in Lemma 5.2.6(2). The sum $Y_{e,n}$ is non-zero modulo p by Lemma 5.2.3(2). Furthermore, up to Frobenius twists and multiplication by scalars, the matrix $Y_{e,n} \bmod p$ is independent of n. Therefore, there exists a vector $w \in \mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ which is independent of n and does not lie in the kernel of $Y_{e,n} \mod p$. The very rapid decay of w follows from this observation and Lemma 5.2.6(2).

Finally, a valuation-theoretic argument analogous to Case 1 shows that every primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w\}$ decays rapidly, thereby establishing Proposition 5.1.3 in this case. □

Case 3: $b = p^{2e+1}a$ for some $e \in \mathbb{Z}_{\geq 0}$. We will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4, w\}$ decays rapidly where w is some primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ and that $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ decays *very* rapidly.

Lemma 5.2.7. *(1) Among the terms appearing in $F_\infty(2)$ described in Lemma 5.2.1 with denominator p^{n+1}, the unique term with minimal t-adic valuation is*

$$P(1)_{e,n-1} \cdot H_u^{(n+e)}(xy)^{1+p+\ldots+p^{n-e-1}} x^{p^{n-e}+p^{n-e+1}+\ldots+p^{n+e}}.$$

(2) Among the terms appearing in $F_\infty(1)$ described in Lemma 5.2.1 with denominator p^{n+1}, there are exactly two terms with minimal t-adic valuation, and they are

$$P(1)_{e,n}(xy)^{1+p+\ldots+p^{n-e-1}} x^{p^{n-e}+p^{n-e+1}+\ldots+p^{n+e-1}}, \text{ and}$$

$$P(1)_{e+1,n}(xy)^{1+p+\ldots+p^{n-e-2}} x^{p^{n-e-1}+p^{n-e}+\ldots+p^{n+e}}.$$

Proof. The proof of this lemma is identical to that of Lemma 5.2.6, so we omit the details. □

Proof of Proposition 5.1.3 in this case. Analogous to Case 2, Lemma 5.2.4 and Lemma 5.2.7(2) imply the existence of a primitive $w \in \mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ that decays rapidly; and by Lemma 5.2.4(1) and Lemma 5.2.7(1), $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ decays very rapidly. Finally, a valuation-theoretic argument shows that every primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w, w_3, w_4\}$ decays rapidly. □

Case 4: $b \neq ap^e$ for any value of e.

Proof of Proposition 5.1.3. As this is the easiest case, we will be content with merely sketching a proof. Analogous to Lemmas 5.2.6 and 5.2.7, it is easy to see that in this case there are unique terms with minimal t-adic valuations with denominator p^{n+1} occurring in both $F_\infty(1)$ and $F_\infty(2)$. It follows that every primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ decays rapidly and every vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ decays very rapidly. Finally, a valuation theoretic argument similar to Case 1 shows that every vector in the span of w_1, w_2, w_3, w_4 does decay rapidly, finishing the proof of Proposition 5.1.3. □

5.3. **Decay in the inert case.** Notation as in the proof of Theorem 5.1.2 and §3.2.1. Recall that P is superspecial and we will show that The $\mathbb{Z}_p$-span of w_1, w_2, w_3 decays rapidly, and the vector w_3 decays very rapidly.

Proof of Proposition 5.1.3. The proof goes along the same lines as the proof of the decay lemma for split Hilbert modular varieties, so we will be content with just outlining the salient points.

We first compute $F_\infty = \prod_{i=0}^{\infty}(1 + F^{(i)})$, where by (3.2.1), with respect to the basis $\{w_1, w_2, w_3, w_4\}$, $F = \begin{pmatrix} F_t & F_u \\ F_l & 0 \end{pmatrix}$, where

$$F_t = \frac{xy}{2p}\begin{pmatrix} -1 & \lambda \\ -1/\lambda & 1 \end{pmatrix}, F_u = \frac{1}{2p}\begin{pmatrix} x & y \\ x/\lambda & y/\lambda \end{pmatrix}, F_l = \begin{pmatrix} -y & \lambda y \\ -x & \lambda x \end{pmatrix}.$$

Recall that the non-ordinary locus is cut out by the equation $xy = 0$ and $a = v_t(x), b = v_t(y) \in \mathbb{Z}_{>0}$.

Similar to Lemma 5.2.1, it is easy to see that the top-left 2×2 block of F_∞ with p-adic valuation $-(n+1)$ has a term of the form $F_t F_t^{(1)} \ldots F_t^{(n)}$, and this term is the unique term with minimal t-adic valuation (equalling $(a+b)(1+p+\ldots p^n)$). Similarly, the top-right 2×2 block of F_∞ with

p-adic valuation $-(n+1)$ has a term of the form $F_tF_t^{(1)}\dots F_t^{(n-1)}F_u^{(n)}$, and this term is the unique term with minimal t-adic valuation (equaling $(a+b)(1+p+\dots p^{n-1})+ap^n$).

Arguments identical to Lemma 5.2.3 and Lemma 5.2.4 yield that every primitive vector in the $\mathbb{Z}_p$ span of w_1, w_2 (and in the span of w_3) decays rapidly (very rapidly, in the case of w_3). Further, as the t-adic valuation of $F_tF_t^{(1)}\dots F_t^{(m)}$ is different from the t-adic valuation of $F_tF_t^{(1)}\dots F_t^{(n-1)}F_u^{(n)}$ for every pair of integers n, m, it follows that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$ also decays rapidly. The argument is elaborated on in the last paragraph of the proof for Case 1 in §5.2. $\square$

6. Proof of the decay lemma in the Siegel case

In this section, we prove Proposition 5.1.3 and hence Theorem 5.1.2 (for superspecial points) in the Siegel case. We refer the reader to the appendix for a decay lemma for supergeneric points. The main idea of the proof is similar to that of the Hilbert case in §5.

6.1. **Preparation of the proof.** We follow the notation in §5, $k=\bar{\mathbb{F}}_p$, $W=W(k)$, $K=W[1/p]$, $\lambda\in\mathbb{Z}_{p^2}^\times$ such that $\sigma(\lambda)=-\lambda$, and $C=\mathrm{Spf}\, k[[t]]$ a generically ordinary formal curve in $\mathcal{M}_k$ which specializes to a superspecial point P. This gives rise to a local ring homomorphism $k[[x,y,z]]\to k[[t]]$ and we pick a lift $W[[x,y,z]]\to W[[t]]$ (still a ring homomorphism), and we denote by $x(t), y(t)$ and $z(t)$ the images of x, y, z respectively.

Let a, b, c denote the t-adic valuations of $x(t), y(t)$ and $z(t)$ respectively. We adopt the convention that a, b, c may take on the value ∞ if the corresponding power series is 0. As before, v_t denotes the t-adic valuation map on $K[[t]]$ or $k[[t]]$.

Also recall that σ denotes both the Frobenius on K and the Frobenius on the coordinate rings $W[[x,y,z]]$ with $\sigma(x)=x^p, \sigma(y)=y^p, \sigma(z)=z^p$; and for a matrix M with entries in $K[[x,y,z]]$, $M^{(n)}$ denotes $\sigma^n(M)$.

The preparation lemmas of the Siegel case are very similar to that of the split Hilbert case in the beginning of §5.2.

6.1.1. **Notations.** Recall that $F_\infty=\prod_{i=0}^{\infty}(1+F^{(i)})$, where by (3.3.1), with respect to the basis $\{w_1,\cdots,w_5\}$,

$$F=\begin{bmatrix} \frac{1}{2p}(xy+\frac{z^2}{4\epsilon}) & -\frac{1}{2\lambda p}(xy+\frac{z^2}{4\epsilon}) & \frac{x}{2\lambda p} & \frac{y}{2\lambda p} & \frac{z}{2\lambda p} \\ \frac{\lambda}{2p}(xy+\frac{z^2}{4\epsilon}) & -\frac{1}{2p}(xy+\frac{z^2}{4\epsilon}) & \frac{x}{2p} & \frac{y}{2p} & \frac{z}{2p} \\ \lambda y & -y & 0 & 0 & 0 \\ \lambda x & -x & 0 & 0 & 0 \\ \frac{\lambda z}{2\epsilon} & -\frac{z}{2\epsilon} & 0 & 0 & 0 \end{bmatrix},$$

where $\epsilon=\lambda^2\in\mathbb{Z}_p^\times$. We denote by F_t, F_u, and F_l the top-left 2×2 block, the top-right 2×3 block, and the bottom-left 3×2 block of F respectively. Define

$$G=\begin{bmatrix} \frac{1}{2} & \frac{-1}{2\lambda} \\ \frac{\lambda}{2} & \frac{-1}{2} \end{bmatrix}, H_u=\begin{bmatrix} \frac{1}{2\lambda} & \frac{1}{2\lambda} & \frac{1}{2\lambda} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \text{ and } H_l=\begin{bmatrix} \lambda & -1 \\ \lambda & -1 \\ \lambda & -1 \end{bmatrix}.$$

Let $F_\infty(1)$ and $F_\infty(2)$ denote the top-left 2×2 block and top-right 2×3 of F_∞ respectively.

By Corollary 3.4.2, the non-ordinary locus is cut out by the equation $xy+z^2/(4\epsilon)=0$. Let ηt^A and μt^B denote the leading terms of $xy+z^2/(4\epsilon)$ and $xy^p+x^py+z^{1+p}/(2\epsilon)$ respectively. In particular, $A=v_t(xy+z^2/(4\epsilon))$, and $B=v_t(xy^p+x^py+z^{1+p}/(2\epsilon))$.

The following is analogous to Lemma 5.2.1.

Lemma 6.1.2. *(1) The part of $F_\infty(1)$ with p-adic valuation $-(n+1)$ consists of sums of products of the form $\prod_{i=0}^{m_1+2m_2} X_i^{(n_i)}$. Here, X_i is either F_t, F_u or F_l,*[27] *m_1+1 is the number of occurrences of F_t, and m_2 is the number of occurrences of the pair F_u, F_l, $m_1+m_2=n$, and $\{n_i\}_{i=0}^{m_1+2m_2}$ is a strictly increasing sequence of non-negative integers. The analogous statement holds for $F_\infty(2)$ as well.*

(2) Fix values of m_1, m_2 as above. Among all the terms in the above sum, the ones with minimal t-adic valuation only occur when $n_i = i$ for all i, and either when $X_0 = X_1 = \ldots X_{m_1} = F_t$, or $X_0 = X_2 = \ldots = X_{2m_2-2} = F_u$, depending on whether $A \geq B$. The analogous statement holds for $F_\infty(2)$ as well.

(3) (for $F_\infty(1)$) The product $\prod_{i=0}^{m_1} F_t^{(i)} \prod_{i=0}^{m_2-1} F_u^{(m_1+1+2i)} F_l^{(m_1+2i+2)}$ equals

$$\frac{1}{p^{n+1}} \prod_{i=0}^{m_1} G^{(i)}(xy+z^2/2)^{(i)} \prod_{i=0}^{m_2-1} \frac{1}{3} H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)} (xy^p + x^p y + z^{p+1})^{(m_1+2i+1)}.$$

(4) (for $F_\infty(2)$) The product $\prod_{i=0}^{m_1} F_t^{(i)} \prod_{i=0}^{m_2-1} F_u^{(m_1+2i+1)} F_l^{(m_1+2i+2)} \cdot F_u^{(m_1+2m_2+1)}$ equals

$$\frac{1}{p^{n+2}} \prod_{i=0}^{m_1} G^{(i)}(xy+z^2/2)^{(i)} \prod_{i=0}^{m_2-1} \frac{1}{3} H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)} (xy^p + x^p y + z^{p+1})^{(m_1+2i+1)} \cdot F_u^{(m_1+2m_2+1)}$$

6.1.3. **Notation.** Let $P(1)_{m_2,n}$ denote the product $\prod_{i=0}^{m_1} G^{(i)} \prod_{i=0}^{m_2-1} \frac{1}{3} H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)}$.

The following will play a similar role as Lemma 5.2.3.

Lemma 6.1.4. *The kernel of $P(1)_{g,f+g} \bmod p$ does not contain any non-zero vector defined over $\mathbb{F}_p$. Moreover, if f is odd (resp. even), the kernel of $P(1)_{g,f+g} \bmod p$ does not contain the vector $\begin{bmatrix} \lambda^{-1} \\ 1 \end{bmatrix}$ (resp. $\begin{bmatrix} -\lambda^{-1} \\ 1 \end{bmatrix}$).*

Proof. We prove the assertions by explicit computation as in Lemmas 5.2.3 and 5.2.4. Note that

$$\frac{1}{3} H_u^{(2m)} H_l^{(2m+1)} = \frac{-1}{2} \begin{bmatrix} 1 & \lambda^{-1} \\ \lambda & 1 \end{bmatrix}, \frac{1}{3} H_u^{(2m-1)} H_l^{(2m)} = \frac{1}{2} \begin{bmatrix} -1 & \lambda^{-1} \\ \lambda & -1 \end{bmatrix}$$

Both these matrices satisfy the relation $X^2 = -X$ and hence $\prod_{i=0}^{m_2-1} H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)}$ equals, up to a multiple of ± 1, one of these matrices depending on the parity of m_1. Similarly, we have

$$G \cdots G^{(2m)} = \frac{1}{2} \begin{bmatrix} 1 & -\lambda^{-1} \\ \lambda & -1 \end{bmatrix}, \; G \cdots G^{(2m+1)} = \frac{1}{2} \begin{bmatrix} 1 & \lambda^{-1} \\ \lambda & 1 \end{bmatrix}.$$

Therefore, $P(1)_{g,f+g}$ equals $\pm\frac{1}{2} \begin{bmatrix} 1 & \lambda^{-1} \\ \lambda & 1 \end{bmatrix}$ if f is odd, and equals $\pm\frac{1}{2} \begin{bmatrix} 1 & -\lambda^{-1} \\ \lambda & -1 \end{bmatrix}$ if f is even. The lemma then follows immediately. □

[27] The terms X_i are chosen so that the product makes sense, and has the right size. Note that this would imply that F_u, F_l must occur in consecutive pairs.

For fixed n, among the terms listed in Lemma 6.1.2 with denominator p^{n+1}, the number of terms with equal minimal t-adic valuation depends on certain numerical relation between A and B. We then perform the following case-by-case analysis in §§6.2-6.4 to prove the Decay Lemma. The first case, while technically the easiest, holds the main ideas in general.

6.2. **Case 1:** $A < B$**.**
Note that if $a + b \neq 2c$, or more generally, if the leading terms of xy and $z^2/(4\epsilon)$ do not cancel, then $A < B$.

Proof of Proposition 5.1.3 in this case. For the ease of exposition, we assume that $a \leq b \leq c$. Note that this forces $2a \leq A$. Even though the statement of Proposition 5.1.3 is not symmetric in a, b, c, an identical argument as the one below suffices to deal with all the other cases.

We will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$ decays rapidly. For a primitive vector $w \in \mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$, write $w = \alpha_u w_u + \alpha_l w_3$, where w_u is a primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$, and $\alpha_u, \alpha_l \in \mathbb{Z}_p$. Since w is primitive, then either α_u or α_l is a p-adic unit. We may assume that α_u is a unit – the other case is entirely analogous to this one. Suppose that the p-adic valuation of α_l is $m \geq 0$.

Consider the terms appearing in $F_\infty(1)$ described in Lemma 6.1.2 with denominator p^{n+1}. As $A < B$, the one with minimal t-adic valuation is $P(1)_{0,n}(xy + z^2/(4\epsilon))^{1+p+\ldots+p^n}$, and this is the unique term with this property. Similarly, consider the terms appearing in $F_\infty(2)$ with denominator p^{n+1+m}. As $A < B$, the unique term whose first column has minimal t-adic valuation is $P(1)_{0,n+m-1} \cdot F_u^{(n+m)}(xy + z^2/(4\epsilon))^{1+p+\ldots+p^{n+m-1}}$.

Let P denote the 2×3 matrix whose first two columns equal $P(1)_{0,n}(xy+z^2/(4\epsilon))^{1+p+\ldots+p^n}$ (part of $F_\infty(1)$), and whose last column is the first column of $P(1)_{0,n+m-1} \cdot F_u^{(n+m)}(xy+z^2/(4\epsilon))^{1+p+\ldots+p^{n+m-1}}$ (part of $F_\infty(2)$). Since $1 \leq a < A$, then for any $m \in \mathbb{Z}_{\geq 0}$, we have $A(1 + \ldots + p^n) \neq A(1 + \ldots + p^{n+m-1}) + ap^{m+n}$. Therefore, regardless of the value of m, the t-adic valuation of entries of the first two columns of P are different from the t-adic valuation of the last column of P.

To prove that w decays rapidly, it suffices to prove that among the monomials in Pw with p-adic valuation equalling $-(n+1)$, there exists a monomial with t-adic valuation $\leq A(1 + \ldots p^n)$. By the proof of Proposition 5.1.3 in Case 1 in §5.2, this in turn reduces to proving the following statement: if $m \geq 1$, then $w_u \bmod p$ is not in the kernel of $P(1)_{0,n} \bmod p$; and if $m = 0$, the vector $\begin{bmatrix} (\lambda^{-1})^{(n)} \\ 1 \end{bmatrix} \bmod p$ is not in the kernel of $P(1)_{0,n-1} \bmod p$. Both statements follow from Lemma 6.1.4, establishing the decay of the rank 3 submodule $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$.

Proposition 5.1.3 in this case follows from the observation that since $2a \leq A$, then w_3 decays very rapidly. □

6.3. **Case 2:** $A \geq B, a \neq b$**.**
Note that if $A \geq B$, then $a + b = 2c$ (as the only way this can happen is if xy has the same t-adic valuation as $z^2/(4\epsilon)$). We may therefore assume without loss of generality that $a < b$. It follows then that $a < c < b$. Within this case, we will need to consider the following two subcases.

Subcase $(2.1)_e$*:* $B(1 + p^{2e-1}) < A(1 + p) < B(1 + p^{2e+1})$ *for some* $e \in \mathbb{Z}_{\geq 1}$. In this subcase, we will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_i\}$ decays rapidly, where $i \in \{3, 4, 5\}$ will be chosen depending on the values of a, b and c.

The following lemma, in conjunction with Lemma 6.1.4, implies (as in Case 1) that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ decays rapidly. It can be proved by the same argument as in the proof of Lemma 5.2.6(1), so we omit its proof.

Lemma 6.3.1. *Among the terms appearing in $F_\infty(1)$ described in Lemma 6.1.2 with denominator p^{n+1}, the unique term with minimal t-adic valuation is*

$$P(1)_{e,n}(xy+z^2/(4\epsilon))^{(1+\ldots+p^{n-e})}(xy^p+x^py+z^{1+p}/(2\epsilon))^{p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-1}}.$$

The t-adic valuation of this term is $A(1+\ldots+p^{n-e})+B(p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-1})$.

The following lemmas will be used to show that one of w_3, w_4, w_5 also decays rapidly. These lemmas imply that among the terms appearing in $F_\infty(2)$ with denominator p^{n+1}, for at least one of the columns of this matrix, there exists a unique term with minimum t-adic valuation.

Lemma 6.3.2. *Given $g\in\mathbb{Z}_{\geq 1}, n\in\mathbb{Z}_{\geq 0}$, consider the multiset consisting of numbers of the form $A(1+\ldots+p^{n-f-1})+B(p^{n-f}+p^{n-f+2}+\ldots+p^{n+f-2})+gp^{n+f}$, as f varies over $\mathbb{Z}\cap[0,n]$. If the minimal number in this multiset occurs more than once, then it must occur for consecutive values of f.*

Proof. For any choice of f, let us denote the expression by $v(f)$. It suffices to prove the following statement: for $f_1<f_2-1$, if $v(f_1)=v(f_2)$, then $v(f_2)>v(f_2-1)$. To that end, suppose that $v(f_1)=v(f_2)$. Then $A(1+p+\ldots p^{f_2-f_1-1})=B(p^{f_2-f_1}-1)(p^{f_2+f_1}+1)/(p^2-1)+gp^{f_2}(p^{f_2}-p^{f_1})$.

To prove $v(f_2)>v(f_2-1)$, note that $p^{-(n-f_2)}(v(f_2)-v(f_2-1))=B(p^{2f_2-1}+1)/(p+1)+gp^{2f_2-1}(p-1)-A$. Multiplying this by $(1+p+\ldots+p^{f_2-f_1})$ and applying the relation of A and B above, we have

$$\frac{1+p+\ldots+p^{f_2-f_1-1}}{p^{n-f_2}}(v(f_2)-v(f_2-1))=\frac{B(p^{f_2-f_1}-1)(p^{2f_2-1}-p^{f_1+f_2})}{p^2-1}+g(p^{f_2-f_1}-1)(p^{2f_2-1}-p^{f_1+f_2}),$$

which is positive since $f_2>f_1+1$. The lemma follows. $\square$

Lemma 6.3.3. *There are at most two numbers g in the set $\{a,b,c\}$ such that there exists an integer f (f is allowed to depend on the choice of g) with $A(1+\ldots p^{n-f-1})+B(p^{n-f}+p^{n-f+2}+\ldots p^{n+f-2})+gp^{n+f}=A(1+\ldots p^{n-f})+B(p^{n-f+1}+p^{n-f+1}+\ldots p^{n+f-3})+gp^{n+f-1}$.*[28]

Proof. Suppose there existed choices of $f\in\mathbb{Z}_{\geq 0}$ for all three choices of g. Let f_1,f_2,f_3 be the choices for f. Then, by the proof of Lemma 6.3.2, we have that $ap^{2f_1-1}(p-1)=A-B(1+p^{2f_1-1})/(1+p)$, and similarly $bp^{2f_2-1}(p-1)=A-B(1+p^{2f_2-1})/(1+p)$, $cp^{2f_3-1}(p-1)=A-B(1+p^{2f_3-1})/(1+p)$. Substituting these expressions in the equality $a+b=2c$ yields the equation

$$(p^{1-2f_1}+p^{1-2f_2}-2p^{1-2f_3})A=\frac{B}{p+1}(p^{1-2f_1}+p^{1-2f_2}-2p^{1-2f_3}).$$

Since $A\geq B\geq p+1$, we have $A\neq B/(p+1)$ and hence $p^{1-2f_1}+p^{1-2f_2}-2p^{1-2f_3}=0$. Since $f_1,f_2,f_3\in\mathbb{Z}_{\geq 1}$, we must have $f_1=f_2=f_3$ and hence $a=b=c$, which is a contradiction. $\square$

Proof of Proposition 5.1.3 in this case. Let $h\in\{a,b,c\}$ be such that there is no f which satisfies the hypothesis of Lemma 6.3.3 (indeed, the lemma guarantees the existence of such an h).

We first show the existence of a rank 3 submodule which decays rapidly. Without loss of generality, we may assume that $h=a$ and we will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2,w_3\}$ decays rapidly (if $h=b$ or c, the identical proof will show sufficient decay, with w_4 or w_5 taking the place of w_3).

As in Case 1, Lemmas 6.1.4 and 6.3.1 to 6.3.3 imply that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2\}$ and $\mathrm{Span}_{\mathbb{Z}_p}\{w_3\}$ both decay rapidly. Therefore, it suffices to show that $\alpha_u w_u+\alpha_3 w_3$ decays rapidly, where w_u is a primitive vector in the span of w_1,w_2, and either α_u or α_3 in $\mathbb{Z}_p$ is a p-adic unit.

By Lemma 6.3.1, the t-adic valuation of the coefficient of $1/p^{n+1}$ of $F_\infty w_u$ is $d(n)=A(1+\ldots+p^{n-e})+B(p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-1})$. Similarly, the t-adic valuation of the coefficient of

[28]Note that if the equation holds, then f is independent of n, since the equation is actually independent of n; see the proof of Lemma 6.3.2.

$1/p^{m+1}$ of $F_\infty \cdot w_3$ is $c(m) = A(1+\ldots+p^{m-f-1}) + B(p^{m-f}+p^{m-f+2}+\ldots+p^{m+f-2}) + ap^{m+f}$ for some $f \in \mathbb{Z} \cap [0,n]$. As in Case 1, it suffices to prove that $d(n)$ is never equal to $c(m)$, regardless of the values of n and m.

Let $c(f', m) = A(1+\ldots+p^{m-f-1}) + B(p^{m-f'}+p^{m-f'+2}+\ldots+p^{m+f'-2}) + ap^{m+f'}$, for any value of $f' \leq m$. By the definition of f, $c(m) = c(f, m)$, and $f' = f$ minimizes the value of $c(f', m)$.

If $n \geq m$, since $a < A$, then $d(n) > c(e, m) \geq c(f, m) = c(m)$, as required. On the other hand, if $m > n$, we have $c(m) > A(1+\ldots+p^{m-f-1}) + B(p^{m-f}+p^{m-f+2}+\ldots+p^{m+f-2}) \geq d(n)$, where the last inequality follows from Lemma 6.3.1.

Finally, we treat the question of very rapid decay. If we may take $h = a$ or $h = c$, the very rapid decay of w_3 or w_5 is established by the inequality $2a < 2c \leq A$. Otherwise, h must be b and for both a, c, there exist f_1, f_3 satisfying the equation in Lemma 6.3.3. Since $a \neq c$, then $f_1 \neq f_3$ and at least one $f_i \geq 2$. By the proof of Lemma 6.3.3, we have $A - B(1+p^{2f_i-1})/(p+1) > 0$ and hence $A \geq 7B > 2b$. Thus, w_4 decays very rapidly. $\square$

Subcase $(2.2)_e$: *$A(1+p) = B(1+p^{2e-1})$ for some $e \in \mathbb{Z}_{\geq 1}$.* In this subcase, we will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4, w_5\}$ decays rapidly. We first need the following lemma.

Lemma 6.3.4. *Among the terms appearing in $F_\infty(2)$ described in Lemma 6.1.2 with denominator p^{n+1}, the unique term with minimal t-adic valuation is*

$$P(1)_{e-1,n-1}F_u^{(n+e-1)}(xy+z^2/(4\epsilon))^{(1+\ldots+p^{n-e})}(xy^p+x^py+z^{1+p}/(2\epsilon))^{p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-3}}.$$

The t-adic valuation of the i^th column term is $A(1+\ldots+p^{n-e}) + B(p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-3}) + gp^{n+e-1}$, where g is either a, b or c depending on whether i is $1, 2$ or 3.

Proof. It suffices to prove that choice of $f = e$ minimizes the expression $A(1+p+\ldots+p^{n-f}) + B(p^{n-f+1}+p^{n-f+3}\ldots+p^{n+f-3}) + gp^{n+f-1}$, where f is allowed to range between 0 and n. This can be verified by direct calculation. $\square$

Proof of Proposition 5.1.3 in this case. It follows from Lemmas 6.1.4 and 6.3.4 that w_3, w_4 and w_5 individually decay rapidly, and that w_3 decays very rapidly. In order to show that $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4, w_5\}$ decays rapidly, it suffices to show that the t-adic valuations of the coefficients $1/p^{l+1}, 1/p^{m+1}, 1/p^{n+1}$ of $F_\infty(w_3), F_\infty(w_4), F_\infty(w_5)$ are always distinct, regardless of the values of l, m, n. By Lemma 6.3.4, these quantities equal $A(1+p+\ldots+p^{l-e}) + B(p^{l-e+1}+p^{l-e+3}+\ldots+p^{l+e-3}) + ap^{l+e-1}$, $A(1+p+\ldots+p^{m-e}) + B(p^{m-e+1}+p^{m-e+3}+\ldots+p^{m+e-3}) + bp^{m+e-1}$ and $A(1+p+\ldots+p^{n-e}) + B(p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-3}) + cp^{n+e-1}$.

As a, b, c are all strictly less than B, these quantities will all be different unless two of l, m, n are equal. In this case, the quantities still differ, because a, b, c are all distinct integers by assumption. Therefore, $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4, w_5\}$ decays rapidly. $\square$

6.4. **Case 3: $A \geq B$ and $a = b$.** In this case, $a = b = c$. We may assume that $x(t) = t^a$, $y(t) = \beta t^a + \sum_{i=a+1}^\infty \beta_i t^i$, and $z(t) = \gamma t^a + \sum_{i=a+1}^\infty \gamma_i t^i$. Since $A \geq B$, we have $\beta + \gamma^2/(4\epsilon) = 0$. We will break the proof of the Decay Lemma into two subcases and the following lemma will be used in both cases.

Lemma 6.4.1. *Suppose that $\gamma \in \mathbb{F}_p$. Let $a' > a$ denote the smallest integer such that either $\beta_{a'} \neq 0$ or $\gamma_{a'} \neq 0$. Then both $\beta_{a'}$ and $\gamma_{a'}$ are non-zero and moreover, $B \geq (p-1)a + 2a'$.*

Proof. Since $\gamma \in \mathbb{F}_p$ and $\beta + \gamma^2/(4\epsilon) = 0$, then $\beta \in \mathbb{F}_p$. Therefore, in $k[[t]]$,

$$xy + z^2/(4\epsilon) = \sum_{i \geq a'} (\beta_i + \gamma\gamma_i/(2\epsilon))t^{i+a} + (4\epsilon)^{-1} \sum_{i,j \geq a'} \gamma_i\gamma_j t^{i+j},$$

$$xy^p + x^py + z^{1+p}/(2\epsilon) = \sum_{i \geq a'} (\beta_i + \gamma\gamma_i/(2\epsilon))t^{i+pa} + \sum_{i \geq a'} (\beta_i^p + \gamma\gamma_i^p/(2\epsilon))t^{pi+a} + (2\epsilon)^{-1} \sum_{i,j \geq a'} \gamma_i\gamma_j^p t^{i+jp}.$$

If one of $\beta_{a'}$ and $\gamma_{a'}$ were zero, then $A = a' + a$, whereas $B \geq a' + pa$; this contradicts with the assumption that $A \geq B$. Hence, we obtain the first assertion of the lemma.

Let $a'' \geq a'$ denote the smallest integer such that $\beta_i + \gamma\gamma_i/(2\epsilon) \neq 0$. Then by applying the Frobenius action, we have $\beta_{a''}^p + \gamma\gamma_{a''}^p/(2\epsilon) \neq 0$, and $B \geq \min\{(p+1)a', a'' + pa\}$. If $B \geq (p+1)a'$, then the second assertion of the lemma follows.

Therefore, we assume that $B = a'' + pa < (p+1)a'$. The expansion of $xy + z^2/(4\epsilon)$ above has a non-zero term of the form $(\beta_{a''} + \gamma\gamma_{a''}/(2\epsilon))t^{a+a''}$. As $A \geq B$, the term $(\beta_{a''} + \gamma\gamma_{a''}/(2\epsilon))t^{a+a''}$ has to be cancelled out by a term of the form $(4\epsilon)^{-1}\sum_{i+j=a+a'',i,j\geq a'}\gamma_i\gamma_j t^{i+j}$. Therefore, it follows that $2a' \leq a + a''$ and hence $B = a'' + pa \geq (p-1)a + 2a'$. $\square$

Case $(3.1)_e$*:* $B(1+p^{2e-1}) < (p+1)A < B(1+p^{2e+1})$ *for some* $e \in \mathbb{Z}_{\geq 1}$*.*

The same argument as in Case 2.1 suffices to prove Proposition 5.1.3, unless $A = B\frac{1+p^{2e-1}}{1+p} + a(p^{2e} - p^{2e-1})$. Therefore, we will assume that this is the case.

Lemma 6.4.2. *Among the terms appearing in* $F_\infty(2)$ *described in Lemma 6.1.2 with denominator* p^{n+1}*, there are exactly two with minimal* t*-adic valuation. They are:*

$$P(1)_{e-1,n-1}F_u^{(n+e-1)}(xy + z^2/(4\epsilon))^{(1+\ldots+p^{n-e})}(xy^p + x^py + z^{1+p}/(2\epsilon))^{p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-3}},$$

$$P(1)_{e,n-1}F_u^{(n+e)}(xy + z^2/(4\epsilon))^{(1+\ldots+p^{n-e-1})}(xy^p + x^py + z^{1+p}/(2\epsilon))^{p^{n-e}+p^{n-e+2}+\ldots+p^{n+e-2}}.$$

Both the terms have t*-adic valuation* $A(1+\ldots+p^{n-e})+B(p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-3})+ap^{n+e-1}$*.*

Proof. This lemma follows from a similar argument as Lemma 5.2.6(2) and the proofs of Lemmas 6.3.2 and 6.3.3, so we omit the details. $\square$

Proof of Proposition 5.1.3 in this case. We will show that either w_3 or w_5 decays very rapidly. There are two terms with minimal t-adic valuation as in Lemma 6.4.2, appearing in the coefficient of $1/p^{n+1}$ of $F_\infty(w_3)$ and $F_\infty(w_5)$. A direct computation yields that the sum of these two terms equals by

$$\frac{1}{2p^{n+1}}P(1)_{0,n-e-1}(xy + z^2/(4\epsilon))^{1+p+\ldots+p^{n-e-1}}(X(t)u(t)^{p^{2e}} + Y(t)u(t)^{p^{2e-1}})^{(n-e)},$$

where

- $u(t)$ stands for either $x(t)$ or $z(t)$, according to whether we work with w_3 or w_5,
- $X(t) = pF_u \cdot F_l^{(1)} \cdot pF_u^{(2)} \cdots F_l^{(2e-1)} \cdot [(\lambda^{-1})^{(2e)}, 1]^{\mathrm{T}}$, and
- $Y(t) = pF_t \cdot pF_u^{(1)} \cdot F_l^{(2)} \cdots pF_u^{(2e-3)} \cdot F_l^{(2e-2)} \cdot [(\lambda^{-1})^{(2e-1)}, 1]^{\mathrm{T}}$. The superscript T stands for transpose.

The decay of w_3 and w_5 is determined by the t-adic valuation of the entries of $X(t)u(t)^{p^{2e}} + Y(t)u(t)^{p^{2e-1}}$. For the rest of the proof, it suffices to focus on the second row of $X(t), Y(t)$ and hence we view them as functions. We prove the very rapid decay of w_3 or w_5 in two cases.

(1) Both $\beta, \gamma \in \mathbb{F}_p$.

In this case, we claim that the t-adic valuation of $X(t)u(t)^{p^{2e}} + Y(t)u(t)^{p^{2e-1}}$ is at most $A+B(p+p^3+\ldots+p^{2e-3})+a'p^{2e-1}$ for at least one choice of $u(t)$ between $x(t)$ and $z(t)$, where a' is defined in Lemma 6.4.1. This claim implies that the t-adic valuation of the coefficient of $1/p^{n+1}$ of $F_\infty(w_3)$ or $F_\infty(w_5)$ is at most $A(1+\ldots+p^{n-e})+B(p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-3})+a'p^{n+e-1}$. This is sufficient to prove the rapid decay of w_3 or w_5. Indeed, this quantity is strictly less than $A(1+\ldots+p^{n-f}) + B(p^{n-f+1} + p^{n-f+3} + \ldots + p^{n+f-3}) + ap^{n+f-1}$ for all values of $f \neq e, e+1$ by Lemma 6.4.1 and hence the sum of the two terms in Lemma 6.4.2 gives the minimal t-adic valuation term of the coefficient of $1/p^{n+1}$ in $F_\infty(w_3)$ or $F_\infty(w_5)$. Moreover, the bounds on a' in Lemma 6.4.1 proves that w_3 or w_5 decays very rapidly.

We now prove the claim by contradiction. Suppose that $X(t)x(t)^{p^{2e}}+Y(t)x(t)^{p^{2e-1}}$ has t-adic valuation greater than $A+B(p+p^3+\ldots+p^{2e-3})+a'p^{2e-1}$. Since $z(t)=\gamma x(t)+\gamma_{a'}t^{a'}+\ldots$ with $\gamma\in\mathbb{F}_p,\gamma_{a'}\neq 0$ and we have assumed that $A=B\frac{1+p^{2e-1}}{1+p}+a(p^{2e}-p^{2e-1})$, it follows that there is a unique monomial in $X(t)z(t)^{p^{2e}}+Y(t)z(t)^{p^{2e-1}}$ with t-adic valuation $A+B(p+p^3+\ldots+p^{2e-3})+a'p^{2e-1}$, thereby establishing the claim for $u(t)=z(t)$.

(2) Either β or γ is not in $\mathbb{F}_p$.

In this case, as $\beta+\gamma^2/(4\epsilon)=0$, we may assume that $\gamma\notin\mathbb{F}_p$. We again consider the function $X(t)u(t)^{p^{2e}}+Y(t)u(t)^{p^{2e-1}}$. Suppose that the leading coefficient of $X(t)$ is μ_X and that of $Y(t)$ is μ_Y. Then, the terms of minimal equal t-adic valuations cancel out in the case when $u(t)=x(t)$ only if $\mu_X+\mu_Y=0$, otherwise by the same idea as in (1), w_3 decays very rapidly. Therefore, we may assume that $\mu_X+\mu_Y=0$. However in this case, if we pick $u(t)=z(t)$, then the terms terms with minimal equal t-adic valuations cancel out only if $\mu_X\gamma^{p^{2e}}+\mu_Y\gamma^{p^{2e-1}}=0$, which is not possible as $\gamma^{p^{2e}}\neq\gamma^{p^{2e-1}}$. In other words, we show that in this case, w_5 decays very rapidly.

As in Case 2.1, $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2\}$ decays rapidly, and also every vector that can be written as $\alpha_u w_u+\alpha_i w_i$ with $\alpha_i\in\mathbb{Z}_p^\times$ ($i=3,5$ depending on whether w_3 or w_5 decays) decays very rapidly. The latter statement follows by the same valuation-theoretic argument as in the proof of Case 2.1, which also proves that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2,w_i\}$ decays rapidly. ☐

Case $(3.2)_e: A(1+p)=B(1+p^{2e-1})$ *for some* $e\in\mathbb{Z}_{\geq 1}$.

Lemma 6.4.3. *Among the terms appearing in* $F_\infty(1)$ *described in Lemma 6.1.2 with denominator* p^{n+1}*, there are exactly two with minimal* t*-adic valuation. They are:*

$$P(1)_{e,n}(xy+z^2/(4\epsilon))^{1+\ldots+p^{n-e}}(xy^p+x^py+z^{1+p}/(2\epsilon))^{p^{n-e+1}+p^{n-e+3}+\ldots+p^{n+e-1}},$$

$$P(1)_{e-1,n}(xy+z^2/(4\epsilon))^{1+\ldots+p^{n-e+1}}(xy^p+x^py+z^{1+p}/(2\epsilon))^{p^{n-e+2}+p^{n-e+4}+\ldots+p^{n+e-2}}.$$

Both these terms have t*-adic valuation* $A(1+\ldots+p^{n-e})+B(p^{n-e+1}+p^{n-e+3}\ldots+p^{n+e-1})$.

As we have seen many lemmas of this flavor, we omit the proof.

This lemma shows that there are two terms with the same t-adic valuation, which could therefore lead to cancellation, and such phenomenon prevents us from proving that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2\}$ decays rapidly. Nevertheless, the following lemma shows that there is at least a saturated rank one submodule of $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2\}$ which decays rapidly.

Lemma 6.4.4. *There is a vector* w_0 *in* $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2\}$ *which decays rapidly.*

Proof. By Lemma 6.4.3 and the proof of Lemma 6.1.4, the coefficient (viewed as a power series in t) of the sum of the two terms with minimal t-adic valuation among the terms with denominator p^{n+1} is of the form $\mu_1M_1+\mu_2M_2$, for some p-adic units μ_i, where $\{M_1,M_2\}=\left\{\begin{bmatrix}1&\lambda^{-1}\\\lambda&1\end{bmatrix},\begin{bmatrix}1&-\lambda^{-1}\\\lambda&-1\end{bmatrix}\right\}$.

As M_1 mod p and M_2 mod p are not scalar multiples of each other, the linear combination $\mu_1M_1+\mu_2M_2$ mod p is non-zero. Therefore, there exists a vector $\bar{w}_0$ defined over $\mathbb{F}_p$ which does not lie in $\ker(\mu_1M_1+\mu_2M_2 \bmod p)$. Choosing $w_0\in\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2\}$ which lifts $\bar{w}_0$ finishes the proof of this lemma. ☐

We are now ready to prove the last remaining case of Proposition 5.1.3 (and also the Decay Lemma Theorem 5.1.2).

Proof of Proposition 5.1.3. We will first prove that there is a rank 2 submodule of $\mathrm{Span}_{\mathbb{Z}_p}\{w_3,w_4,w_5\}$ which decays rapidly. For ease of notation, let $\bar{F}_u$ denote the matrix $\frac{1}{t^a}F_u$ evaluated at $t=0$.

Let K denote $\ker(P(1)_{n-1,e-1}\bar{F_u}^{(n+e-1)} \bmod p) \cap \mathrm{Span}_{\mathbb{F}_p}\{w_3, w_4, w_5\}$. If $\dim_{\mathbb{F}_p} K \leq 1$, then lifting two linearly independent $\mathbb{F}_p$-vectors $\notin K$ gives the desired rank 2 submodule. Therefore, we assume that $\dim_{\mathbb{F}_p} K = 2$ (note that since $P(1)_{n-1,e-1}\bar{F_u}^{(n+e-1)} \bmod p$ is not the zero matrix, so $\dim_{\mathbb{F}_p} K \neq 3$). It follows that $\beta, \gamma \in \mathbb{F}_p$.

We will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ decays rapidly. First, since $K \cap \mathrm{Span}_{\mathbb{F}_p}\{w_3, w_4\} = \mathrm{Span}_{\mathbb{F}_p}\{\beta w_3 - w_4\}$, then any primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ which modulo p is not a multiple of $\beta w_3 - w_4$ must decay rapidly. Now we consider $\beta w_3 - w_4$. Up to constants, the coefficient of the $1/p^{n+1}$ part of the first entry of $F_\infty(\beta w_3 - w_4)$ equals $\beta_{a'} t^{A(1+\dots+p^{n-e})+B(p^{n-e+1}+p^{n-e+2}\dots+p^{n+e-3})+a'p^{n+e-1}}$. Lemma 6.4.1 establishes the required decay as follows: firstly, as $a' \leq B \leq A$, we have that the vector $\beta w_3 - w_4$ decays rapidly. Secondly, the exact bound for a' in Lemma 6.4.1 implies (as in the proof in Case 2.1) that $\mathrm{Span}_{\mathbb{Z}_p}\{w_3, w_4\}$ decays rapidly. Finally, the *very* rapid decay of w_3,w_4 follows from the bound $2a' \leq B \leq A$.

Then, Proposition 5.1.3 follows by an argument analogous to that in Case 2.1 with Lemma 6.4.4. □

7. The setup of the main proofs

In this section, we provide the general setup of the proofs of Theorems 1 and 3. As mentioned in §1.3, the proofs consist of the following parts:

(1) The sum of the local contributions at supersingular points is at most 11/12 of the global contribution; and
(2) the local contribution from non-supersingular points is of smaller magnitude.

Proposition 7.2.5 makes (1) precise, and is stated in §7.2. We will prove Proposition 7.2.5 and (2) in §8 for the Hilbert case and in §9 for the Siegel case. The idea involved in the statement of Proposition 7.2.5 is that we break the global intersection number $C.Z(m)$ into pieces, one for each non-ordinary point on C, by using the relation between the Hasse invariant and the Hodge line bundle in §7.1. We also relate the local intersection multiplicity at a point to a lattice-point count.

7.1. **The global contribution and its decomposition.** Recall that in §4.3.3, we list the set T of $m \in \mathbb{Z}_{>0}$ for which we will study $C.Z(m)$ to prove our main theorems. In order to study the asymptotic behavior, we define $T_M = \{m \in T \mid m \leq M\}$ for $M \in \mathbb{Z}_{>0}$. Moreover, in §§8-9, we will construct a subset $S_M \subset T_M$ which consists of bad values of m that we want to rule out. The total global intersection number that we will consider is $\sum_{m\in T_M - S_M} C.Z(m)$. We sum over m instead of working with individual m because geometry-of-numbers techniques which we use to bound the local intersection multiplicity (for cumulative m) do not work for individual m. The following lemma gives the asymptotics of the global term using results in §4.

Lemma 7.1.1. *Assume that $\#S_M = O(M^{1-\epsilon}) = O(\#T_M^{1-\epsilon})$ for some $\epsilon > 0$ if $L = L_H$ and that $\#S_M = o(\#T_M)$ if $L = L_S$. Then*

$$\sum_{m\in T_M - S_M} C.Z(m) = (\omega.C) \sum_{m\in T_M - S_M} |q_L(m)| + o(\sum_{m\in T_M - S_M} |q_L(m)|).$$

Moreover, we have, for Theorem 1(2), $\sum_{m\in T_M - S_M} C.Z(m) \asymp M^2$; for Theorem 1(1) and Remark 2, $\sum_{m\in T_M - S_M} C.Z(m) \asymp M^2/\log M$; for Theorem 3, $\sum_{m\in T_M - S_M} C.Z(m) \asymp M^{5/2}/\log M$.

Proof. By §4.3.1 and the assumption on S_M, we have $\sum_{m\in S_M} |q_L(m)| = o(\sum_{m\in T_M} |q_L(m)|)$. Then the assertions follow from §4.3.1, Lemma 4.3.2, Lemma 4.3.4, and the prime number theorem. □

For each non-ordinary point P on $C \cap Z(m)$, we introduce the notion of *global intersection number $g_P(m)$ at P* using the following (well-known) relation between the non-ordinary locus and the divisor class of the Hodge bundle. Note that in the proof, we will only use the notion $g_P(m)$ for a supersingular point.

Lemma 7.1.2. *The non-ordinary locus in $\mathcal{M}_k$ and $\mathcal{M}_k^{\mathrm{tor}}$ is cut out by a Hasse-invariant H, which is a section of ω^{p-1}, and hence the number of non-ordinary points (counted with multiplicity) on C is given by $(p-1)(C.\omega)$.*

See for instance [Box15, §§1.4-1.5, Theorem 6.2.3] for an explanation of this fact (and we use the fact that the ordinary Newton stratum coincides with the ordinary Ekedahl–Oort stratum). For the last assertion in the lemma, we remark that when $L = L_H$, the boundary $\mathcal{M}_k^{\mathrm{tor}} \backslash \mathcal{M}_k$ is ordinary and hence the intersection of C' (in §4.1.3) with the non-ordinary locus is the same as the intersection of C with the non-ordinary locus.

Definition 7.1.3. Let t be the local coordinate at P (i.e., $\widehat{C}_P = \operatorname{Spf} k[[t]]$) and let $A = v_t(H)$. We define $g_P(m) = \frac{A}{p-1}|q_L(m)|$.

Note that by the above lemmas, we have the following decomposition

$$\sum_{P \in C \text{ non-ord}} \sum_{m \in T_M - S_M} g_P(m) = \sum_{m \in T_M - S_M} |q_L(m)|(\omega.C) = \sum_{m \in T_M - S_M} C.Z(m) + o\Big(\sum_{m \in T_M - S_M} |q_L(m)|(\omega.C)\Big).$$

7.2. **The lattices and the outline of the proof.** Let $B \to \operatorname{Spf} k[[t]]$ denote the generically ordinary abelian surface given by fulling back the universal family over $\mathcal{M}_k$ to $\widehat{C}_P = \operatorname{Spf} k[[t]]$ for some point $P \in C$. Recall the notion of special endomorphisms from §2.2 and by a slight abuse of terminology, when $L = L_H$, we will also refer to a special quasi-endomorphism with certain integrality condition in §2.2.11 as a special endomorphism. For any $n \in \mathbb{Z}_{>0}$, the lattice is special endomorphisms of $B \bmod t^n$ is a sublattice of $B \bmod t$, which is equipped with a positive definite quadratic form Q' (see Definition 2.3.1).

Lemma 7.2.1. *The local intersection multiplicity of $C.Z(m)$ at P, denoted by $l_P(m)$, equals*

$$\sum_{n=1}^{\infty} \#\{\textit{Special endomorphisms } s \textit{ of } B \bmod t^n \textit{ with } Q'(s) = m\}.$$

The lemma follows directly from the moduli interpretation of $Z(m)$. Note that as B generically has no special endomorphisms, this infinite sum can actually be be truncated at some finite stage (which will depend on m).

Remark 7.2.2. Given B, the lattices of special endomorphisms of $B \bmod t^n$ have the same rank for all $n \in \mathbb{Z}_{>0}$. Indeed, the work of de Jong, Moonen and Kisin cited in the proof of Theorem 5.1.2 applies to any P and for any special endomorphism w of $B \bmod t$, we have the parallel extension $\tilde{w} \in K[[t]]$, which is invariant under the Frobenius on $\mathbb{L}_{\mathrm{cris}}(W[[t]])$. By de Jong's theory (here we need the fully faithfulness of the Dieudonné functor, see [dJ95, Cor. 2.4.9]), whether w extends over modt^n depends on the p-powers in the denominators of the coefficients of $\tilde{w}$. Therefore, given n, there exists N such that $p^N w$ extends over modt^n and hence these lattices tensor $\mathbb{Z}_\ell, \ell \neq p$ are all isomorphic and in particular, the rank of the lattices is independent of n.

Motivated by the Decay Lemmas Theorems 5.1.2 and A.0.1, we define the following lattices for supersingular points (note that the notation is slightly different from that in the introduction and we will use the notation in this section for the rest of the paper).

7.2.3. Assume P is supersingular and recall that $A = v_t(H)$, where H is the Hasse invariant and in Definition 5.1.1, $A_n = [A(p^n + p^{n-1} + \cdots + 1 + \frac{1}{p})]$.

- If B mod t is superspecial, define $L_{0,1}$, $L_{n,1}, n \in \mathbb{Z}_{>0}$, and $L_{n,2}, n \in \mathbb{Z}_{\geq 0}$ to be the lattices of special endomorphisms of B mod t, mod $t^{A_{n-1}+1}$, and mod $t^{A_{n-1}+ap^n+1}$ respectively. As in Definition 2.3.1, we pick a lattice $L'_{n,i} \subset L'$ such that $L_{n,i} \subset L'_{n,i}$ and for $\ell \neq p$, $L'_{n,i} \otimes \mathbb{Z}_\ell = L' \otimes \mathbb{Z}_\ell$ and $L'_{n,i} \otimes \mathbb{Z}_p = L_{n,i} \otimes \mathbb{Z}_p$. In particular, $L'_{0,1} = L'$ and by Theorem 5.1.2, we have $[L'_{n,1} : L'_{n,2}] \geq p$ and $[L' : L'_{n,1}] \geq p^{3n}$.
- If B mod t is supergeneric, define L_0 and $L_n, n \in \mathbb{Z}_{>0}$ to be the lattice of special endomorphisms of B mod t and mod $t^{A_{n-1}+1}$. Again, as in Definition 2.3.1, we pick a lattice $L'_n \subset L'$ such that $L_n \subset L'_n$ and for $\ell \neq p$, $L'_n \otimes \mathbb{Z}_\ell = L' \otimes \mathbb{Z}_\ell$ and $L'_n \otimes \mathbb{Z}_p = L_n \otimes \mathbb{Z}_p$; Theorem A.0.1 implies that $[L' : L'_n] \geq p^{3n}$.

Since we assume that C does not admit any global special endomorphisms, we have $\cap_{n=0}^{\infty} L_{n,i} = \{0\} = \cap_{n=0}^{\infty} L_n$. By Remark 7.2.2, the difference between $L'_n, L'_{n,i}$ and $L_n, L_{n,i}$ is the same as that between $L_0, L_{0,i}$ and L', we also have $\cap_{n=0}^{\infty} L'_{n,i} = \{0\} = \cap_{n=0}^{\infty} L'_n$.

Corollary 7.2.4. *(1) If P is superspecial, then*

$$l_P(m) \leq \frac{A(p+2)}{2p} r_{0,1}(m) + \frac{A}{2} r_{0,2}(m) + \sum_{n=1}^{\infty} \frac{Ap^n}{2} (r_{n,1}(m) + r_{n,2}(m)),$$

where $r_{n,i}(m) = \#\{s \in L'_{n,i} \mid Q'(s) = m\}$.

(2) If P is supergeneric, then

$$l_P(m) \leq \frac{A(p+1)}{p} r_0(m) + \sum_{n=1}^{\infty} Ap^n r_n(m),$$

where $r_n(m) = \#\{s \in L'_n \mid Q'(s) = m\}$.

Proof. By Lemma 7.2.1 and §7.2.3, we have that for P superspecial,

$$\begin{aligned} l_P(m) &\leq (A_{-1} + a) r_{0,1}(m) + (A_0 - A_{-1} - a) r_{0,2}(m) + \sum_{n=1}^{\infty} (ap^n r_{n,1}(m) + (A_n - A_{n-1} - ap^n) r_{n,2}(m)) \\ &= A_{-1}(r_{0,1}(m) - r_{0,2}(m)) + a \sum_{n=0}^{\infty} p^n (r_{n,1}(m) - r_{n,2}(m)) + \sum_{n=0}^{\infty} A_n (r_{n,2}(m) - r_{n+1,2}(m)) \\ &\leq \frac{A}{p}(r_{0,1}(m) - r_{0,2}(m)) + \frac{A}{2} \sum_{n=0}^{\infty} p^n (r_{n,1}(m) - r_{n,2}(m)) + \sum_{n=0}^{\infty} (A(p^n + \cdots + 1 + p^{-1})(r_{n,2}(m) - r_{n+1,2}(m)), \end{aligned}$$

where the last equality follows from the facts that $r_{n,1}(m) \geq r_{n,2}(m), r_{n,2}(m) \geq r_{n+1,2}(m)$ and $a \leq A/2, A_n \leq A(p^n + \cdots + p^{-1})$. We then obtain the assertion in (1) by rearranging the summations. A similar argument gives (2). □

The main task of the next two sections is to prove that

Proposition 7.2.5. *Given C, there exists S_M satisfying the assumption in Lemma 7.1.1 such that for every supersingular point P on C, we have*

$$\sum_{m \in T_M - S_M} l_P(m) \leq \frac{11}{12} \sum_{m \in T_M - S_M} g_P(m) + o(\sum_{m \in T_M - S_M} g_P(m)).$$

Once we have this proposition, we will prove that the local contribution from non-supersingular points have smaller order of magnitude, whence we conclude that there are infinitely many non-supersingular points on C which lie in the desired special divisors.

7.3. **Ordinary points.** In order to bound $l_P(m)$, we need the following decay lemma for ordinary points, which follows directly from Serre–Tate theory. We thank Keerthi Madapusi Pera for pointing this out to us. Let $B \to \operatorname{Spf} k[[t]]$ denote the abelian surface with ordinary reduction given by pulling back the universal family over $\mathcal{M}_k$ to $\widehat{C}_P = \operatorname{Spf} k[[t]]$ for an ordinary point P.

Lemma 7.3.1. *If w is not a special endomorphism for the p-divisible group $B[p^\infty] \bmod t^{A+1}$, then pw is not a special endomorphism for $B[p^\infty] \bmod t^{pA+1}$.*

Proof. Note that an endomorphism of $B[p^\infty] \bmod t^n$ is special if and only if its reduction on $B[p^\infty] \bmod t$ is special. Hence we only need to consider the deformation of endomorphisms. Let $\{q_i\}$ ($i = 1, 2$ for the Hilbert case and $i = 1, 2, 3$ for the Siegel case) denote q-coordinates of the formal group in the Serre–Tate deformation theory. Then, the formal neighborhood of the Shimura variety is given by $\operatorname{Spf} W[[t_i]]$, where $t_i = q_i - 1$. Note that over $W[[t_i]]$, s is an endomorphism of a point $x = (x_1, x_2, x_3)$ [29] if and only if, all points $y = (y_1, y_2, y_3)$ satisfying the condition $(y_i + 1)^p = x_i + 1$ (equivalently $(y_i + 1)^p - 1 = x_i$) have the property that ps is an endomorphism of y. In other words, if $f(t_i) = f(q_i - 1) = 0$ is the local equation of the locus on $W[[t_i]]$ such that the endomorphism w deforms, then $f(q_i^p - 1) = f((t_i + 1)^p - 1) = 0$ is the local equation for pw to deform. Hence over $k[[t]]$, if $f = 0$ is the local equation that deforms w, then $f^p = 0$ is the local equation for pw, and this gives the desired assertion. $\square$

Lemma 7.3.2. *Let $L_0, L_n, n \in \mathbb{Z}_{>0}$ be the lattices of special endomorphisms of $B \bmod t$ and $B \bmod t^{Ap^{n-1}+1}$ respectively where $A \in \mathbb{Z}_{>0}$. Then*

(1) for any A, we have $\operatorname{rk}_{\mathbb{Z}} L_n \leq 2$ if $L = L_H$ and $\operatorname{rk}_{\mathbb{Z}} L_n \leq 3$ if $L = L_S$;
(2) there exist a constant A and a $\mathbb{Z}_p$-lattice Λ (depending on P) with $\operatorname{rk}_{\mathbb{Z}_p} \Lambda \leq 1$ when $L = L_H$ and $\operatorname{rk}_{\mathbb{Z}_p} \Lambda \leq 2$ when $L = L_S$ such that $L_n \subset (\Lambda + p^{n-1} L_1 \otimes \mathbb{Z}_p) \cap L_0$.

In particular, if $\operatorname{rk}_{\mathbb{Z}} L_n = 3$ when $L = L_S$ or $\operatorname{rk}_{\mathbb{Z}} L_n = 2$ when $L = L_H$, then $(\operatorname{disc} L_n)^{1/2} \geq p^{n-1}$.

Proof. Note that $L_n \subset L_n \otimes \mathbb{Z}_p \subset L_0 \otimes \mathbb{Z}_p = \mathbb{L}_{\mathrm{cris},P}(W)^{\varphi=1}$, where $\mathbb{L}_{\mathrm{cris},P}$ is the fiber of the F-crystal $\mathbb{L}_{\mathrm{cris}}$ defined in Definition 2.2.3 and Definition 2.2.9 and φ is the Frobenius action. Since P is ordinary, then φ acts on $\mathbb{L}_{\mathrm{cris},P}(W)$ with slope $-1, 1, 0, 0$ (Hilbert case) or $-1, 1, 0, 0, 0$ (Siegel case) and hence (1) follows.

Let Λ' be the $\mathbb{Z}_p$-lattice of special endomorphisms of $B[p^\infty]$. Since $\widehat{C}_P$ is not contained in any special divisor,[30] $B[p^\infty]$ admits at most a rank 2 (resp. rank 1) module of special endomorphisms when $L = L_S$ (resp. $L = L_H$); indeed, if $\operatorname{rk}_{\mathbb{Z}_p} \Lambda' = 3$ (resp. 2), then $\Lambda' \otimes \mathbb{Q}_p = L_0 \otimes \mathbb{Q}_p$, and thus the B admits special endomorphisms.

We now mimic the proof of [ST17, Thm. 4.1.1] using Lemma 7.3.1 instead of [ST17, Lem. 4.1.2(2)]. Let $\Lambda \subset L_0 \otimes \mathbb{Z}_p$ be the saturation of Λ' in $L_0 \otimes \mathbb{Z}_p$; then there exists $\Lambda_0 \subset L_0 \otimes \mathbb{Z}_p$ such that $L_0 \otimes \mathbb{Z}_p = \Lambda \oplus \Lambda_0$. Let Λ_n denote $(L_n \otimes \mathbb{Z}_p + \Lambda) \cap \Lambda_0$; then $L_n \otimes \mathbb{Z}_p + \Lambda = \Lambda \oplus \Lambda_n$. It suffices to show that there exists A such that $\Lambda_n \subset p\Lambda_{n-1}$ (and this implies that $\Lambda_n \subset p^{n-1}\Lambda_1$).

By definition, none of the elements in Λ_0 extend to $\operatorname{Spf} k[[t]]$, then there exists A such that $\Lambda_1 \subset p\Lambda_0$. For $n \geq 2$, assume for contradiction that there exists $\alpha \in \Lambda_n \backslash p\Lambda_{n-1}$. If $\alpha \in p\Lambda_{n-2}$, then write $\alpha = p\beta$ with $\beta \in \Lambda_{n-2}$. Since $p\beta = \alpha \in \Lambda_n$, then by Lemma 7.3.1, $\beta \in \Lambda_{n-1}$, which contradicts with the assumption that $\alpha \notin p\Lambda_{n-1}$. Thus we have $\alpha \notin p\Lambda_{n-2}$; by iterating the argument, we have $\alpha \notin p\Lambda_0$. This is a contradiction since $\alpha \in \Lambda_n \subset \Lambda_1 \subset p\Lambda_0$. $\square$

[29] i.e., x is defined by setting $t_1 = x_1,\ t_2 = x_2$ and $t_3 = x_3$

[30] This is the assumption of Theorem 1(1) and Theorem 3; and for Theorem 1(2), we may assume this as otherwise, the conclusion is automatic.

8. Proof of Theorem 1(2)

In this section, we use the results proved in §§4-5 to prove Proposition 7.2.5 in the case of Hilbert modular surfaces. This, in conjunction with Lemma 8.1.2, yields Theorem 1(2).

8.1. The bad set S_M and the local intersection multiplicities at non-supersingular points.

We first construct the set S_M; the following lemma only concerns ordinary and superspecial points because we only need to consider such P for the proof of Theorem 1(2). Indeed, if $P \in Z(m)$, then P is either ordinary or supersingular and if $P \in Z(m), p \nmid m$, then by §4.4.2(1), P is not supergeneric. Therefore for $P \in Z(m), m \in T$, P is either superspecial or ordinary.

Lemma 8.1.1. *Notation as in §7.1,7.2.3 and Lemma 7.3.2. Given a finite set* $\{P_i\} \subset (C \cap (\cup_{m\in\mathbb{Z}_{>0}} Z(m)))(k)$, *there exists* $S_M \subset T_M$ *with* $\#S_M = O(M^{1-\epsilon})$ *for some* $0 < \epsilon < 1/6$ *such that for all* i,

(1) if P_i is superspecial, then $\{s \in L'_{N,1} \mid 0 \neq Q'(s) \leq M, Q'(s) \notin S_M\} = \emptyset$ *where* $N = \frac{1+\epsilon}{3}\log_p M$;

(2) if P_i is ordinary, then $\{s \in L_N \mid 0 \neq Q'(s) \leq M, Q'(s) \notin S_M\} = \emptyset$ *where* $N = \epsilon \log_p M$.

Proof. Since the union of finitely many sets with cardinality $O(M^{1-\epsilon})$ still has cardinality to be $O(M^{1-\epsilon})$, it suffices to prove the assertion for each P_i separately. We follow the idea of the proof of [ST17, Thm. 4.3.3].

If P_i is superspecial, we take $S_M = \{m \in T_M \mid \exists s \in L'_{N,1} \text{ with } Q'(s) = m\}$ and then it satisfies (1) by definition. Note that $\#S_M \leq \#\{s \in L'_{N,1} \mid Q'(s) \leq M\}$. Then by a geometry-of-numbers argument (see for instance [ST17, Lem. 4.2.1]) and Theorem 5.1.2, we have

$$\#\{s \in L'_{N,1} \mid Q'(s) \leq M\} = O(M^2/p^{3N} + M^{3/2}/p^{2N} + M/p^N + M^{1/2}/d_N),$$

where d_N is the first successive minimum of $L'_{N,1}$ and $d_N \to \infty$ as $N \to \infty$ because $\cap L'_{N,1} = \{0\}$. Then $\#S_M = O(M^{1-\epsilon})$ by the definition of N.

If P_i is ordinary, then $\operatorname{rk} L_N = 2$ by Lemma 7.3.2 and the fact that $\operatorname{rk} L_N = \operatorname{rk} L_0$ is even by the Tate conjecture. Similar to the superspecial case, we take $S_M = \{m \in T_M \mid \exists s \in L_N \text{ with } Q'(s) = m\}$ and then by Lemma 7.3.2, $\#S_M = O(M/p^N + M^{1/2}/d_N) = O(M^{1-\epsilon})$. □

Lemma 8.1.2. *Notation as in Lemma 8.1.1. For an ordinary point* $P = P_i \in C(k)$, *we have*

$$\sum_{m\in T_M - S_M} l_P(m) = O(M^{1+\epsilon}) = o(M^2).$$

Proof. By Lemmas 7.2.1, 7.3.2 and 8.1.1,

$$\sum_{m\in T_M-S_M} l_P(m) = \sum_{m\in T_M-S_M} A(r_0(m) + \sum_{n=1}^{N}(p^n - p^{n-1})r_n(m)) \leq A\sum_{n=0}^{N} p^n \sum_{m=1}^{M} r_n(m),$$

where $r_n(m) = \#\{s \in L_n \mid Q'(s) = m\}$. By a geometry-of-numbers argument and Lemma 7.3.2, we have $\sum_{m=1}^{M} r_n(m) = O(M/p^n + M^{1/2}/d_n)$, where d_n is the first successive minimum of L_n and the implicit constant here only depends on p. Thus $\sum_{m\in T_M-S_M} l_P(m) = O(NM + p^N M^{1/2}) = O(M^{1+\epsilon})$. □

8.2. Proof of Proposition 7.2.5 in the Hilbert case.

We follow the notation in Lemma 8.1.1 and $P = P_i$ superspecial. We break $\sum_{m\in T_M-S_M} l_P(m)$ into two parts and are treated in the following lemmas.

Lemma 8.2.1. *Notation as in Corollary 7.2.4. For any $\epsilon > 0$, there exists $c \in \mathbb{Z}_{>0}$ which only depends on P and ϵ such that*

$$\sum_{m \in T_M - S_M} \sum_{n=c}^{\infty} \frac{Ap^n}{2}(r_{n,1}(m) + r_{n,2}(m)) \leq \epsilon M^2 + o(M^2).$$

Proof. By Lemma 8.1.1, $r_{n,i}(m) = 0$ for $n > N = \frac{1+\epsilon}{3}\log_p M$ and hence

$$\sum_{m \in T_M - S_M} \sum_{n=c}^{\infty} \frac{Ap^n}{2}(r_{n,1}(m)+r_{n,2}(m)) = \sum_{m \in T_M - S_M} \sum_{n=c}^{N} \frac{Ap^n}{2}(r_{n,1}(m)+r_{n,2}(m)) \leq \sum_{n=c}^{N}\sum_{m=1}^{M} Ap^n r_{n,1}(m)$$

since $r_{n,1}(m) \geq r_{n,2}(m)$.

By a geometry-of-numbers argument, $\sum_{m=1}^{M} r_{n,1}(m) \leq c_2(M^2/p^{3n}+M^{3/2}/p^{2n}+M/p^n+M^{1/2}/d_n)$, where c_2 is an absolute constant and d_n is the first successive minimum of $L'_{n,1}$. Hence

$$\sum_{n=c}^{N} Ap^n \sum_{m=1}^{M} r_{n,1}(m) \leq Ac_2M^2\sum_{n=c}^{N} 1/p^{2n} + \sum_{n=c}^{N} Ap^n c_2(M^{3/2}/p^{2n} + M/p^n + M^{1/2}/d_n).$$

Note that $Ac_2\sum_{n=c}^{N} \leq Ac_2(p^{2c}(1-p^{-2}))^{-1}$, which goes to 0 as $c \to \infty$ and the second term is

$$O(M^{3/2}) + O((\log M)M) + O(M^{1/2})\sum_{n=c}^{N} p^n = O(M^{3/2}).$$

Thus we obtain the desired estimate. $\square$

Lemma 8.2.2. *Notation as in Corollary 7.2.4. For any $c \in \mathbb{Z}_{>0}$, we have*

$$\sum_{m \in T_M - S_M} (\frac{A(p+2)}{2p} r_{0,1}(m) + \frac{A}{2} r_{0,2}(m) + \sum_{n=1}^{c} \frac{Ap^n}{2}(r_{n,1}(m)+r_{n,2}(m))) \leq \alpha \sum_{m \in T_M - S_M} g_P(m) + o(M^2),$$

where $\alpha < 11/12$ is an absolute constant.

Proof. Let $\theta_{n,i}$ denote the theta series attached to the lattice $L'_{n,i}$. We decompose $\theta_{n,i} = E_{n,i} + G_{n,i}$, where $G_{n,i}$ is a cusp form and $E_{n,i}$ is an Eisenstein series as in §4.2 and follow the proof of Lemma 4.3.2.

Let $E = \frac{A(p+2)}{2p}E_{0,1} + \frac{A}{2}E_{0,2} + \sum_{n=1}^{c}\frac{Ap^n}{2}(E_{n,1}+E_{n,2})$, $G = \frac{A(p+2)}{2p}G_{0,1} + \frac{A}{2}G_{0,2} + \sum_{n=1}^{c}\frac{Ap^n}{2}(G_{n,1}+G_{n,2})$.

Note that G is a weight 2 cusp form and by Deligne's Weil bound, we have that its m-th Fourier coefficient $q_G(m) = O(m^{1/2+\epsilon})$. Hence the total contribution from the cusp form G is $\sum_{m \in T_M - S_M} q_G(m) = O(M^{3/2+\epsilon})$.

Let $q_{n,i}(m)$ and $q(m)$ denote the m-th Fourier coefficient of $E_{n,i}$ and E. Recall that for $p \nmid m$ for $m \in T_M$; by Lemma 4.4.6 and the fact that $|L'^{\vee}/L'| = p^2$, we have for any n, i

$$\frac{q_{n,i}(m)}{|q(m)_L|} \leq \frac{2p}{(p^2-1)[L' : L'_{n,i}]}, \text{ and } \frac{q_{0,1}(m)}{|q(m)_L|} \leq \frac{1}{p-1}.$$

Recall from §7.2.3 that $[L' : L'_{n,1}] \geq p^{3n}$ and $[L' : L'_{n,1}] \geq p^{3n+1}$; therefore,

$$\frac{q(m)}{|q_L(m)|} \leq \frac{A(p+2)}{2p}\cdot\frac{1}{p-1} + \frac{A}{2}\frac{2p}{(p^2-1)p} + \sum_{n=1}^{c}\frac{Ap^n}{2}\cdot\frac{2p}{p^2-1}(p^{-3n}+p^{-3n-1}) \leq \frac{A}{p-1}\left(\frac{p+2}{2p} + \frac{p}{p^2-1}\right).$$

Take $\alpha = \frac{p+2}{2p} + \frac{p}{p^2-1}$, which is $< 11/12$ when $p \geq 5$. We have the left hand side equals

$$\sum_{m \in T_M - S_M} (q(m) + q_G(m)) \leq \sum_{m \in T_M - S_M} \frac{\alpha A}{p-1} |q_L(m)| + O(M^{3/2+\epsilon}),$$

which gives the desired estimate by the definition of $g_P(m)$. □

Proof of Proposition 7.2.5 when $L = L_H$. The set S_M is constructed by Lemma 8.1.1 and taking $\{P_i\}$ to contain all of (the finitely many) supersingular points in $C \cap (\cup_{p \nmid m} Z(m))$. Then the desired estimate follows from Lemma 8.2.1 and Lemma 8.2.2 by taking c such that $\epsilon < \frac{11}{12} - \alpha$. □

Proof of Theorem 1(2). If C is contained in $Z(m)$ with m being a perfect square, then by applying suitable Hecke translates, we may assume that C is contained in the product of modular curves and then the assertion is a special case of [CO06, Proposition 7.3]. Now for the rest of the proof, we may assume that C is contained in some Hilbert modular surface and we will use $Z(m)$ to denote special divisors on the Hilbert modular surface. Note that any point on $Z(m)$ corresponds to an abelian surface isogenous to the self-product of an elliptic curve. Thus we assume for contradiction that there are only finitely many points on $C \cap (\cup_{m \in T} Z(m))$ and take $\{P_i\}$ to be this finite set and apply Lemma 8.1.1 to construct S_M. Since all $Z(m)$ are compact, it makes sense to consider $C.Z(m)$. We deduce a contradiction by Lemma 7.1.1, Proposition 7.2.5, and Lemma 8.1.2. □

9. Proofs of Theorem 1(1) and Theorem 3

In this section, we prove all of Theorem 1 and Theorem 3. §9.1 consists of results pertaining to squares represented by positive definite quadratic forms.[31] In §9.2, we prove Proposition 7.2.5 by combining results proved in §§4, 6, and 9.1. Finally, we deal with the intersection multiplicities at non-supersingular points in §9.3 to finish the proof of the main theorem.

We now set up notation that we will use for §9. For supersingular points P, recall that we defined $L'_{n,i}$ for superspecial points and L'_n for supergeneric point in §7.2.3. Let $l(n)_i, i = 1, \ldots, 5$ denote the i^{th} successive minimum of the quadratic form Q' restricted to $L'_{n,1}$ or L'_n. Let P_n denote a rank two sublattice of $L'_{n,1}$ or L'_n with minimal discriminant. Note that $l(n)_1 l(n)_2 \asymp d_n$, where d_n denotes the root discriminant of P_n. Moreover, since $\cap_{n=0}^{\infty} L'_{n,i} = \{0\} = \cap_{n=0}^{\infty} L'_n$, we have $l(n)_1 \to \infty$ as $n \to \infty$.

9.1. **Preparation.**

We need the following results to prove Proposition 7.2.5. Although Lemma 9.1.2 is stated for the rank 5 lattices $L'_{n,1}, L'_n$, the proof does not use the assumption on rank and hence it holds for the lattices L_n for ordinary points (notation as in Lemma 7.3.2) when $\mathrm{rk}_{\mathbb{Z}} L_0 = 3$; see §9.3 for details.

Lemma 9.1.1. *We have $l(n)_1 l(n)_2 \cdots l(n)_i \gg p^{(i-2)n}$ for $i \geq 3$.*

Proof. Note that if we have two lattices $L_1 \supset L_2$, then the successive minima of L_2 give upper bounds of that of L_1. Thus we may enlarge $L'_{n,i}, L'_n$ and prove the assertion for the enlarged lattices.

We enlarge $L'_{n,i}, L'_n$ as follows. For $\ell \neq p$, we still require $L'_{n,i} \otimes \mathbb{Z}_\ell = L' \otimes \mathbb{Z}_\ell$ and $L'_n \otimes \mathbb{Z}_\ell = L' \otimes \mathbb{Z}_\ell$; at p, let Λ_0 denote the rank 3 submodule of $L' \otimes \mathbb{Z}_p$ which decays rapidly in the Decay Lemmas (Theorems 5.1.2 and A.0.1), then we enlarge $L'_{n,1}, L'_n$ such that $L'_{n,1} \otimes \mathbb{Z}_p = p^n \Lambda_0 + L' \otimes \mathbb{Z}_p$ and $L'_{n,1} \otimes \mathbb{Z}_p = p^n \Lambda_0 + L' \otimes \mathbb{Z}_p$.

[31] Recall that we must prove our curve intersects special divisors of the form $Z(D\ell^2)$ at infinitely many points. This involved dealing with squares represented by quadratic forms, and hence the Geometry-of-numbers arguments are more involved than in the Hilbert case.

For the enlarged $L'_{n,1}, L'_n$, we have

$$l(n)_j \ll p^n, j = 1, \dots, 5, \quad l(n)_1 l(n)_2 \cdots l(n)_5 \asymp p^{3n},$$

where the implied constants only depend on the lattice L'. Thus the assertion follows. $\square$

Lemma 9.1.2. *Suppose that $d_n^2 M = o(p^{2n})$ as $n \to \infty$. Then for any vector $v \in L'_{n,1}$ or L'_n such that $Q(v) \le M$, we have that $v \in P_n$ for $n \gg 1$. In particular, if $d_n \le p^{n/2}$, then for any vector $v \in L'_{n,1}$ or L'_n such that $Q'(v) < p^{n-\epsilon}$ for some absolute constant $\epsilon > 0$, we have that $v \in P_n$ for $n \gg 1$. (All the implicit constants here are independent of n, M.)*

Proof. Recall that $l(n)_1 \cdot l(n)_2 \asymp d_n$. Thus, by Lemma 9.1.1, we have

$$l(n)_1 l(n)_2 l(n)_3 \gg p^n, \quad l(n)_3 \gg p^n/d_n.$$

In other words, for any vector v linearly independent to P_n, we have $Q'(v) \ge l(n)_3^2 \gg p^{2n}/d_n^2$. Then the first assertion follows. The second assertion follows directly from the first assertion by taking $M = p^{n-\epsilon}$. $\square$

Proposition 9.1.3. *Fix $D \in \mathbb{Z}_{>0}$. Recall $r_{n,i}(m), r_n(m)$ from Corollary 7.2.4. Then we have the following two bounds: (here we state the results for $r_n(m)$ and the same results also hold for $r_{n,1}(m)$)*

(1) $$\sum_{m=D\ell^2, m\le M, \ell \text{ prime}} r_n(m) = O_\epsilon\Big(\frac{M^{2+\epsilon}}{p^{2n}} + \frac{M^{3/2+\epsilon}}{p^n} + M^{1+\epsilon}\Big).$$

(2) $$\sum_{m=D\ell^2, m\le M, \ell \text{ prime}} r_n(m) \text{ and } \sum_{\ell \le M, \ell \text{ prime}} r_n(\ell) \text{ are both } O\Big(\tfrac{M^{5/2}}{p^{3n}} + \tfrac{M^2}{p^{2n}} + \tfrac{M^{3/2}}{p^n} + \tfrac{M}{d_n} + \tfrac{M^{1/2}}{l(n)_1}\Big).$$

Proof. In the proof, we work with L'_n and everything holds true for $L'_{n,1}$, too.

We note that (2) is a trivial upper bound from a geometry-of-numbers argument. Indeed, both $\sum_{m=D\ell^2, m\le M, \ell \text{ prime}} r_n(m)$ and $\sum_{\ell\le M, \ell \text{ prime}} r_n(\ell)$ are no greater than $\sum_{m=1}^{M} r_n(m)$; we then obtain the desired bound by [ST17, Lem. 4.2.1] and Lemma 9.1.1.

Now we prove (1). We may assume that there exists a vector $v_0 \in L'_0$ such that $Q'(v_0) = D\ell_0^2$ for some prime ℓ_0. Otherwise $r_n(m) = 0$ for all $m = D\ell^2$ for any prime ℓ. Let e_1 denote a primitive vector in L'_n such that $e_1 = p^k v_0$ for some $k \in \mathbb{Z}_{\ge 0}$. By definition, $p^n L'_0 \subset L'_n$ and thus $p^n v_0 \in L'_n$. Therefore $k \le n$. Since e_1 is primitive in L'_n, we extend it into a basis $\{e_1, e_2, \dots e_5\}$ of L'_n. Let $\widetilde{L'}_n$ denote the sublattice of L'_0 spanned by $f_1 := v_0 = e_1/p^k, e_2, \dots e_5$; since $\widetilde{L'}_n$ is a sublattice of L'_0, then $Q'|_{\widetilde{L'}_n}$ is still $\mathbb{Z}$-valued. Since $Q'(v_0) = D\ell_0^2 =: N$, then there exist vectors $f_2 \dots f_5 \in (2N)^{-1}\widetilde{L'}_n$ such that $[f_i, f_1]' = 0$ for $i \ge 2$, and $\mathrm{Span}_{\mathbb{Z}}\{f_1, f_2, f_3, f_4, f_5\} \supset \widetilde{L'}_n$.

Let $\widetilde{Q'}$ denote the restriction of $Q' \otimes \mathbb{Q}$ to $\mathrm{Span}_{\mathbb{Z}}\{f_2, f_3, f_4, f_5\} \subset L'_0 \otimes \mathbb{Q}$. By the definition of f_i, we have $\widetilde{Q'}$ is a $(2N)^{-2}\mathbb{Z}$-valued quadratic form. Let $\widetilde{l(n)}_1, \dots, \widetilde{l(n)}_4$ denote the successive minima of $\mathrm{Span}_{\mathbb{Z}}\{f_2, f_3, f_4, f_5\}$. Since $(2N)^{-1}\widetilde{L'}_n = (2N)^{-1}L'_n + (2N)^{-1}p^{-k}\mathbb{Z}e_1$, then $\widetilde{l(n)}_1 \cdots \widetilde{l(n)}_i \gg p^{(i-1)n-k} \ge p^{(i-2)n}$ for $i \ge 2$ (note that $k \le n$ and N is absorbed in the implicit constant as N is independent of n, k). Then the standard geometry-of-numbers argument gives

$$Y_n := \#\{y \in \mathrm{Span}_{\mathbb{Z}}\{f_2, f_3, f_4, f_5\} \mid \widetilde{Q'}(y) \le M\} = O\Big(\frac{M^2}{p^{2n}} + \frac{M^{3/2}}{p^n} + M\Big).$$

On the other hand, on $\mathrm{Span}_{\mathbb{Z}}\{f_1, f_2, f_3, f_4, f_5\}$, for $v = xf_1 + y_2 f_2 + \cdots + y_5 f_5$, we have $Q'(v) = D\ell_0^2 x^2 + \widetilde{Q'}(v_y)$, where $v_y = y_2 f_2 + \cdots + y_5 f_5$. If $Q'(v) = D\ell^2 \le M$, then $Q'(v_y) \le Q'(v) \le M$ and $Q'(v_y) = D(\ell - \ell_0 x)(\ell + \ell_0 x)$. For a given v_y with $Q'(v_y) \le M$, there are at most $O_\epsilon(M^\epsilon)$ ways to factor $Q'(v_y)/D$ into two factors and thus there are at most $O_\epsilon(M^\epsilon)$ possible x such that for $v = xf_1 + v_y$, we have $Q'(v) = D\ell^2 \leq M$ for some prime ℓ. Since $L'_n \subset \mathrm{Span}_{\mathbb{Z}}\{f_1, f_2, f_3, f_4, f_5\}$, then $\sum_{m=D\ell^2, m\leq M, \ell \text{ prime}} r_n(m) = O_\epsilon(M^\epsilon Y_n)$, which gives the (1) by the above bound for Y_n. $\square$

Proposition 9.1.4. *Fix $D \in \mathbb{Z}_{>0}$. The proportion of primes $\ell \leq (M/D)^{1/2}$ such that $D\ell^2$ is represented by the quadratic form restricted to P_n goes to zero as $n \to \infty$.*

Proof. Let R_n denote the imaginary quadratic ring with discriminant $-d_n^2$. The class group of R_n is in bijection with equivalence classes of binary quadratic forms of discriminant $-d_n^2$. Let $\mathfrak{a}$ denote the ideal corresponding to Q' restricted to P_n. Recall that $l(n)_1 \to \infty$ as $n \to \infty$. Thus for $n \gg 1$, we have that $\mathfrak{a}$ is not equivalent to any ideal whose norm is D, i.e., (P_n, Q') does not represent D. Note that it suffices to deal with primes ℓ which are relatively prime to Dd_n^2.

The correspondence between ideal classes and binary quadratic forms yields that the integer $D\ell^2$ is represented by (P_n, Q') if and only if there exists an invertible ideal $\mathfrak{b}$ equivalent to $\mathfrak{a}$ with $\mathrm{Nm}\,\mathfrak{b} = D\ell^2$. This implies that $\ell = \mathfrak{c}_1\mathfrak{c}_2$ (i.e. the prime ℓ splits in R_n), and that $\mathfrak{b} = \mathfrak{d}\mathfrak{c}_1^2$ or $\mathfrak{b} = \mathfrak{d}\mathfrak{c}_2^2$, where $\mathfrak{d}$ is some ideal such that $\mathrm{Nm}\,\mathfrak{d} = D$ (the case $\mathfrak{b} = \mathfrak{c}_1\mathfrak{c}_2$ is ruled out by the above discussion that $\mathfrak{a}$ and therefore $\mathfrak{b}$ is not equivalent to any ideal whose norm is D). In other words, Q' restricted to P_n represents $D\ell^2$ if and only if there exist some ideals $\mathfrak{c}, \mathfrak{d}$ such that $\mathrm{Nm}\,\mathfrak{c} = \ell, \mathrm{Nm}\,\mathfrak{d} = D$ and $\mathfrak{c}^2\mathfrak{d}$ is equivalent to $\mathfrak{a}$.

Let C denote the equivalence classes of ideals $\mathfrak{c}$ such that $\mathfrak{c}^2$ is equivalent to $\mathfrak{a}\mathfrak{d}^{-1}$ for some $\mathfrak{d}$ with $\mathrm{Nm}\,\mathfrak{d} = D$. Since D is fixed, then C is a finite (independent of n) union of torsors for the 2-torsion of the class group of R_n, when C is nonempty. By the Genus theory, the cardinality of the two-torsions of the class group of R_n is bounded by the number of prime divisors of d_n^2; thus, $\#C = O_\epsilon(d_n^\epsilon)$.

We finish the proof in two cases.

(1) If $d_n \leq (\log M)^2$, it follows by [TZ18] that the proportion of primes represented by the quadratic form associated to any ideal class $\mathfrak{c}$ is $1/d_n$ because $d_n \asymp$ the class number of R_n. Thus the total proportion of ℓ such that $D\ell^2$ is representable is $\#C/d_n = O_\epsilon(d_n^{\epsilon-1})$, which goes to 0 as $d_n \to \infty$.

(2) If $d_n \geq (\log M)^2$, let $f_\mathfrak{c}$ denote the binary quadratic form associated to $\mathfrak{c}$. Then as in the proof of [ST17, Claim 3.1.9], we have

$$\#\{\ell \mid \ell < (M/D)^{1/2} \text{ representable by } f_\mathfrak{c}\} \leq \#\{m \mid m < (M/D)^{1/2} \text{ representable by } f_\mathfrak{c}\} = O(M^{1/2}/d_n + M^{1/4}).$$

Thus by the above discussion,

$$\#\{\ell \mid \exists v \in P_n, Q'(v) = D\ell^2 \leq M\} = (\#C)O(M^{1/2}/d_n + M^{1/4}) = o(M^{1/2}/\log M),$$

which finishes the proof. $\square$

The following result gives a bound of Fourier coefficients of the cuspidal part of our theta series in terms of the discriminant of the quadratic lattice.

Proposition 9.1.5 (Duke, Waibel)**.** *Let S be a fixed finite set of primes. Let θ be the theta series attached to a positive definite quadratic lattice of rank 5 with discriminant D_θ such that all prime factors of D_θ lie in S. Write $\theta = E + G$, where E is an Eisenstein series and G is a cusp form. Then, there exist absolutely bounded positive constants N_0 and C such that for all $m \in T$ (the set T defined in §4.3.3), the m-th Fourier coefficient $q_G(m)$ of G satisfies that $q_G(m) \leq CD_\theta^{N_0} m^{1+1/4}$.*

By Remark 7.2.2, we have that $\mathrm{disc}\, L'_{n,i}, \mathrm{disc}\, L'_n$ are independent of n, i away from p and hence all the theta series attached to these lattices satisfy the assumption on D_θ.

An analogous result of Proposition 9.1.5 was proved by Duke in the case of ternary quadratic forms. The main steps of his proof carry through in this case too, so we will be content with just sketching his proof.

Proof. The proof of [Duk05, Lem. 1] and the discussion on [Duk05, p.40] apply to rank 5 quadratic forms (with suitable modification of the power of D_θ) and we have that the Petersson norm of G satisfies $||G|| = O(D_\theta^{N_1})$ for some absolute constant N_1 (here we use the fact that the level N_θ of G is $O(D_\theta)$.

Thus to obtain a bound for $q_G(m)$, we only need to bound the Fourier coefficients $a_j(m)$ for an orthonormal basis of the space of cusp forms of weight $5/2$ and level N_θ (with respect to certain quadratic character determined by θ). Now we apply [Wai18, Theorem 1]. Using the notation there, we have that if $m = \ell$, then $t = \ell, v = 1, w = 1, (m, N_\theta) = O(1)$; if $m = D\ell^2$, then $t = D, v \asymp 1, w \asymp \ell, (m, N_\theta) = O(1)$. Thus $|a_j(m)| \ll_\epsilon m^{\frac{27}{28}+\epsilon} D_\theta^\epsilon$ for $m = \ell$ and $|a_j(m)| \ll_\epsilon m^{\frac{3}{4}+\epsilon} D_\theta^\epsilon$, which gives the desired bound once we combine with the above estimate of $||G||$. $\square$

9.2. **Proof of Proposition 7.2.5 in the Siegel case.** Notation as in §7.2.3 and Corollary 7.2.4. For a supersingular point P, we estimate $\sum_{m \in T_M - S_M} r_{n,i}(m), \sum_{m \in T_M - S_M} r_n(m)$ with respect to different ranges of n.

Definition 9.2.1. Given absolute constants $\epsilon_0, \epsilon_1 > 0$ (we will choose ϵ_0, ϵ_1 in the proof of Proposition 7.2.5), the ranges of n are defined as follows:

- n is *small* if $n \leq \epsilon_0 \log_p M$, where ϵ_0 is a constant independent of M.
- n is in the *lower medium* range if $\epsilon_0 \log_p M < n \leq \frac{3}{4} \log_p M$
- n is in the *upper medium* range if $\frac{3}{4} \log_p M < n \leq (1 + \epsilon_1) \log_p M$.
- n is *large* if $n > (1 + \epsilon_1) \log_p M$.

Proof Proposition 7.2.5 for Theorem 1(1) and Remark 2 with p split in F. For $m \in T_M$, we have $m = D\ell^2$, where D is a non-zero quadratic residue mod p. Then by §4.4.3(2), any supergeneric point does not lie on $Z(m)$. Hence we will only consider P superspecial.

Recall from Lemma 7.1.1 that for any S_M such that $\#S_M = o(\#T_M)$, we have

$$\sum_{m \in T_M - S_M} C.Z(m) \asymp \sum_{m \in T_M - S_M} g_P(m) \asymp M^2 (\log M)^{-1}.$$

We will first prove that there exists S_M such that $\#S_M = o(\#T_M)$ and the contribution from $n \geq \epsilon_0 \log_p M$ is $o(M^2 / \log M)$.

The lower medium range. By Proposition 9.1.3(1),

$$\begin{aligned}
\sum_{m \in T_M - S_M} \sum_{n = \lceil \epsilon_0 \log_p M \rceil}^{[\frac{3}{4} \log_p M]} \frac{Ap^n}{2} (r_{n,1}(m) + r_{n,2}(m)) &\leq \sum_{m = D\ell^2, m \leq M} \sum_{n = \lceil \epsilon_0 \log_p M \rceil}^{[\frac{3}{4} \log_p M]} Ap^n r_{n,1}(m) \\
&= A \sum_{n = \lceil \epsilon_0 \log_p M \rceil}^{[\frac{3}{4} \log_p M]} p^n O_\epsilon(M^{2+\epsilon}/p^{2n} + M^{3/2+\epsilon}/p^n + M^{1+\epsilon}) \\
&= O_\epsilon(M^{2+\epsilon-\epsilon_0} + M^{3/2+\epsilon} \log M + M^{7/4+\epsilon}),
\end{aligned}$$

which is $o(M^2 / \log M)$ once we take $\epsilon < \min\{\epsilon_0, 1/4\}$.

The upper medium range. We treat this part in two ways according to whether $d_{n_0} \leq M^{1/8}$, where $n_0 = \lceil \frac{3}{4} \log_p M \rceil$.

(1) If $d_{n_0} \geq M^{1/8}$, then we bound this part using geometry-of-numbers.
Since $L'_{n,1} \subset L'_{n_0,1}$ for all $n \geq n_0$, then by definition, $d_n \geq d_{n_0} \geq M^{1/8}$. By Proposition 9.1.3(2), we have that

$$\sum_{m \in T_M - S_M} \sum_{n=\lceil \frac{3}{4} \log_p M \rceil}^{[(1+\epsilon_1)\log_p M]} \frac{Ap^n}{2}(r_{n,1}(m) + r_{n,2}(m)) \leq \sum_{n=\lceil \frac{3}{4} \log_p M \rceil}^{[(1+\epsilon_1)\log_p M]} \sum_{m=D\ell^2, m \leq M} Ap^n r_{n,1}(m)$$
$$= O(M + M^{5/4} + M^{3/2}\log M + M^{15/8+\epsilon_1} + M^{3/2+\epsilon_1}),$$

which is $o(M^2/\log M)$ once we take $\epsilon_1 < 1/8$.

(2) If $d_{n_0} < M^{1/8}$, we control this part by putting m's in this range into S_M.
More precisely, consider $R_M := \{m \in T_M \mid \exists v \in L'_{n_0,1}, Q'(v) = m\}$. By our assumption, $d_{n_0}^2 M < M^{5/4} = o(p^{2n_0})$ and by Lemma 9.1.2, for $M \gg 1$, if $m \in R_M$, then m is represented by $Q' \mid_{P_{n_0}}$, which is a binary quadratic form. Then by Proposition 9.1.4 (note that $n_0 \to \infty$ as $M \to \infty$), $\# R_M = o(M^{1/2}/\log M) = o(\# T_M)$. Thus we may choose S_M such that $S_M \supset R_M$ and then

$$\sum_{m \in T_M - S_M} \sum_{n=\lceil \frac{3}{4} \log_p M \rceil}^{[(1+\epsilon_1)\log_p M]} \frac{Ap^n}{2}(r_{n,1}(m) + r_{n,2}(m)) = 0.$$

The large n's. Let $n_0 = \lceil (1+\epsilon_1)\log_p M \rceil$ and let $R'_M := \{m \in T_M \mid \exists v \in L'_{n_0,1}, Q'(v) = m\}$. We will show that $\# R'_M = o(M^{1/2}/\log M)$ and thus we may choose S_M such that $S_M \supset R'_M$ and then

$$\sum_{m \in T_M - S_M} \sum_{n=\lceil (1+\epsilon_1)\log_p M \rceil}^{\infty} \frac{Ap^n}{2}(r_{n,1}(m) + r_{n,2}(m)) = 0.$$

We bound the size of R'_M case by case depending on the size of $d_{n_0}, l(n_0)_1$.

- Case (1): $d_{n_0} \leq M^{1/2+\epsilon_2}$ for some absolute constant $\epsilon_2 < \epsilon_1/2$.
Then $d_{n_0} \leq M^{1/2+\epsilon_2} < p^{n_0/2}$ and $M < p^{n_0 - \epsilon_1}$. By Lemma 9.1.2, for $M \gg 1$, if $m \in R'_M$, then m is represented by $Q' \mid_{P_{n_0}}$. By Proposition 9.1.4, $\# R'_M = o(M^{1/2}/\log M)$.
- Case (2): $d_{n_0} > M^{1/2+\epsilon_2}$ for all $\epsilon_2 < \epsilon_1/2$ and $l(n_0)_1 > M^{\epsilon_3}$ for some absolute constant $\epsilon_3 > 0$.
We have $\# R'_M \leq \#\{v \in L'_{n_0,1} \mid Q'(v) \in T_M\}$, which is $O(M^{1/2-\epsilon_1} + M^{1/2-\epsilon_2} + M^{1/2}/l(n_0)_1) = o(M^{1/2}/\log M)$ by Proposition 9.1.3(2).
- Case (3) $d_{n_0} > M^{1/2+\epsilon_2}$ for some $\epsilon_2 < \epsilon_1/2$ and $l(n_0)_1 \leq M^{\epsilon_3}$ for some $\epsilon_3 < \epsilon_2$.
Then $l(n_0)_2 = d_{n_0}/l(n_0)_1 > M^{1/2}$. In other words, any vector $v \in L'_{n_0,1}$ which is not a scalar multiple of the chosen vector v_0 of the minimum length has $Q'_{n_0}(v) \leq l(n_0)_2^2 > M$. Therefore, any $m \in R'_M$ has to be represented by the rank 1 quadratic form spanned by v_0. As $M \to \infty$, we have $l(n_0)_1 \to \infty$. Thus once M is large enough such that $l(n_0)_1^2 > D$, then this rank 1 quadratic form would represent at most one element in T_M and hence $\# R'_M = o(\# T_M)$.

In conclusion, taking $S_M = R_M \cup R'_M$, we have $\# S_M = o(\# T_M)$ and

$$\sum_{m \in T_M - S_M} \sum_{n=\lceil \epsilon_0 \log_p M \rceil}^{\infty} \frac{Ap^n}{2}(r_{n,1}(m) + r_{n,2}(m)) = o(M^2/\log M).$$

The small n's. We follow the notation and the idea of the proof in Lemma 8.2.2.

We enlarge $L'_{n,1}$ as in the proof of Lemma 9.1.1; also let w be the vector which decays very rapidly in the Decay Lemma for superspecial points, then we enlarge $L'_{n,2}$ such that $L'_{n,2} \otimes \mathbb{Z}_p = L'_{n,1} \otimes \mathbb{Z}_p + p^{n+1}\mathbb{Z}_p w$. Then $\operatorname{disc} L'_{n,i} \asymp p^{6n}$ with the implicit constant only depending on P. Note that Corollary 7.2.4 still holds with the new definitions of $L'_{n,i}$.

Let

$$E = \frac{A(p+2)}{2p} E_{0,1} + \frac{A}{2} E_{0,2} + \sum_{n=1}^{[\epsilon_0 \log_p M]} \frac{Ap^n}{2}(E_{n,1}+E_{n,2}), G = \frac{A(p+2)}{2p} G_{0,1} + \frac{A}{2} G_{0,2} + \sum_{n=1}^{[\epsilon_0 \log_p M]} \frac{Ap^n}{2}(G_{n,1}+G_{n,2}).$$

Note that here the Eisenstein series E and the cusp form G depend on M.

Since $\operatorname{disc} L'_{n,i} = O(p^{6\epsilon_0 \log_p M}) = O(M^{6\epsilon_0})$ for $n \leq \epsilon_0 \log_p M$, then by Proposition 9.1.5, the m-th Fourier coefficient

$$q_G(m) \ll (M^{6\epsilon_0})^{N_0} m^{5/4} \sum_{n=0}^{[\epsilon_0 \log M]} p^n \ll M^{(6N_0+1)\epsilon_0} m^{5/4}$$

and $\sum_{m \in T_M - S_M} q_G(m) = O(M^{(6N_0+1)\epsilon_0 + 7/4}) = o(M^2/\log M)$ once we take $\epsilon_0 < (24N_0+4)^{-1}$.

The computation for the Eisenstein part is the same as in the proof of Lemma 8.2.2. More precisely, since $p \nmid m$, by Lemma 4.4.6(1)(3), we have

$$\frac{q(m)}{|q_L(m)|} \leq \frac{A(p+2)}{2p} \cdot \frac{1}{p-1} + \frac{A}{2} \frac{2p}{(p^2-1)p} + \sum_{n=1}^{[\epsilon_0 \log M]} \frac{Ap^n}{2} \cdot \frac{2p}{p^2-1}(p^{-3n} + p^{-3n-1}) \leq \frac{11}{12} \cdot \frac{A}{p-1}.$$

Thus we finish the proof by putting all parts together and using Corollary 7.2.4. □

Proof of Proposition 7.2.5 for Remark 2 when p inert or ramified in F. Comparing to the previous case when p is split in F, the only differenced are (1) $Z(m)$ may contain supergeneric points; and (2) $p \mid m$ when p is ramified and the estimate for $\frac{q(m)}{|q_L(m)|}$ may change. Nevertheless, except the computation of the Fourier coefficients of the Eisenstein series, all the bounds for n in the medium and large range and for the cuspidal contribution for small n remain valid for supergeneric points if we replace $L'_{n,1}$ by L'_n for supergeneric points. Note that we only enlarge $L'_n, L'_{n,i}$ as in the proof of Lemma 9.1.1 when we treat small n. Thus, to finish the proof, we compute the contribution from the Eisenstein part (for the enlarged lattices).

If P supergeneric, let θ_n denote the theta series attached to the lattice L'_n. We decompose $\theta_n = E_n + G_n$, where G_n is a cusp form and E_n is an Eisenstein series. Let $E = \frac{A(p+1)}{p} E_0 + \sum_{n=1}^{[\epsilon_0 \log_p M]} Ap^n E_n$.

If p is inert, i.e., $p \nmid D$, $p \nmid m$, then by Lemma 4.4.6(2)(3) and the fact that $|(L'_0)^\vee / L'_0| = p^2$ for P supergeneric, we have

$$\frac{q(m)}{|q_L(m)|} \leq \frac{A(p+1)}{p} \cdot \frac{2}{p^2-1} + \sum_{n=1}^{[\epsilon_0 \log_p M]} Ap^n \frac{2}{(p^2-1)[L'_0 : L'_n]} < \frac{A}{p-1}\left(\frac{2}{p} + \frac{2}{(p+1)(p^2-1)}\right) < \frac{11}{12} \cdot \frac{A}{p-1}.$$

If p is ramified, we have $v_p(m) = 1$. For P superspecial, by Lemma 4.4.6(1)(4), we have

$$\frac{q(m)}{|q_L(m)|} \leq \frac{A(p+2)}{2p} \cdot \frac{1}{p-1} + \frac{A}{2} \cdot \frac{4}{p^2-1} + \sum_{n=1}^{[\epsilon_0 \log M]} \frac{Ap^n}{2} \cdot \frac{2p}{p-1}\left(\frac{1}{p^{3n}} + \frac{1}{p^{3n+1}}\right) < \frac{A}{p-1}\left(\frac{p+2}{2p} + \frac{p+3}{p^2-1}\right),$$

which is $< \frac{11}{12} \cdot \frac{A}{p-1}$ for $p \geq 7$. Similarly, for P supergeneric, by Lemma 4.4.6(2)(4), we have

$$\frac{q(m)}{|q_L(m)|} \leq \frac{A(p+1)}{p} \cdot \frac{2}{p^2-1} + \sum_{n=1}^{[\epsilon_0 \log_p M]} Ap^n \frac{2}{(p-1)[L_0' : L_n']} < \frac{A}{p-1}\left(\frac{2}{p} + \frac{2}{p^2-1}\right) < \frac{11}{12} \cdot \frac{A}{p-1}.$$

Thus we finish the proof. □

Proof of Proposition 7.2.5 for Theorem 3. Since every $m \in T_M$ in this case is a non-zero quadratic residue mod$p$, hence by §4.4.3(2), all supersingular points on $Z(m)$ are superspecial. The idea of the proof is similar to the case of Theorem 1 (1).

By Lemma 7.1.1, we have $\sum_{m \in T_M - S_M} g_P(m) \asymp M^{5/2}(\log M)^{-1}$. We construct S_M by large n. More precisely, we set $S_M = \{m \in T_M \mid \exists v \in L'_{n_0,1}, Q'(v) = m\}$, where $n_0 = \lceil (1+\epsilon_1)\log_p M \rceil$. Then

$$\#S_M \leq \#\{v \in L'_{n_0,1} \mid Q'(v) \leq M\} = O(M^{5/2}/p^{3n_0} + M^2/p^{2n_0} + M^{3/2}/p^{n_0} + M/d_{n_0} + M^{1/2}) = O(M^{1/2} + M/d_{n_0}),$$

which is $o(M/\log M) = o(\#T_M)$ if there exists an absolute constant $\epsilon > 0$ such that $d_{n_0} \gg M^{\epsilon}$. If not, then by Lemma 9.1.2, we have that for $M \gg 1$, all $m \in S_M$ representable by the binary quadratic form $Q'|_{P_{n_0}}$. Since $d_{n_0} \to \infty$, the density of primes representable by $Q'|_{P_{n_0}}$ goes to zero, i.e., we still have $\#S_M = o(\#T_M)$. With this choice of S_M, we have

$$\sum_{m \in T_M - S_M} \sum_{n = \lceil (1+\epsilon_1)\log_p M \rceil}^{\infty} \frac{Ap^n}{2}(r_{n,1}(m) + r_{n,2}(m)) = 0.$$

For n in the medium range,

$$\sum_{m \in T_M - S_M} \sum_{n = \lceil \epsilon_0 \log_p M \rceil}^{\lceil (1+\epsilon_1)\log_p M \rceil} \frac{Ap^n}{2}(r_{n,1}(m) + r_{n,2}(m)) \ll \sum_{n = \lceil \epsilon_0 \log_p M \rceil}^{\lceil (1+\epsilon_1)\log_p M \rceil} p^n \sum_{m \leq M} r_{n,1}(m) = o(M^{5/2}/\log M)$$

since $\sum_{m \leq M} r_{n,1}(m) = O(M^{5/2}/p^{3n} + M^2/p^{2n} + M^{3/2}/p^n + M)$. The estimate for small n's is exactly as in the case for Theorem 1(1) above and thus we finish the proof. □

9.3. **Contribution from non-supersingular points and conclusions.** To finish the proof, we only need to show that $\sum_{m \in T_M - S_M} l_P(m)$ for non-supersingular points P are $o(\sum_{m \in T_M - S_M} C.Z(m))$, which is $o(M^2/\log M)$ for Theorem 1(1) and Remark 2 and is $o(M^{5/2}/\log M)$ for Theorem 3. We still use the notation in Lemma 7.3.2 for ordinary points.

Recall that an abelian surface is ordinary, almost ordinary (i.e., its Newton polygon has slopes $0, 1/2, 1$), or supersingular.

Lemma 9.3.1. *If P is almost ordinary or if P is ordinary with $\mathrm{rk}_{\mathbb{Z}} L_0 \neq 3$, then*

$$\sum_{m \in T_M - S_M} l_P(m) = o(\sum_{m \in T_M - S_M} C.Z(m)).$$

Proof. By the classification of endomorphism rings of char p abelian surfaces (see for instance [Tat71, Thm. 1]), we see that if the abelian surface corresponding to P has almost ordinary reduction, then its lattice of special endomorphisms has rank at most 1. On the other hand, if P is ordinary, then $\mathrm{rk}_{\mathbb{Z}} L_0$ is odd and hence $\mathrm{rk}_{\mathbb{Z}} L_0 = 1$. In both cases, let $a_n x^2$ to denote the quadratic form with one variable given by Q' restricted to the lattice of special endomorphisms of the abelian surface mod t^n. Since the lattice mod t^{n+1} is a sublattice of the one mod t^n, we have $a_n \mid a_{n+1}$.

Since C does not have any global special endomorphisms, we have $a_n \to \infty$ and hence $a_n x^2$ does not represent any element in $T_M \subset \{D\ell^2 \mid \ell \text{ prime }\}$ or $T_M \subset \{\ell \mid \ell \text{ prime }\}$ once $n \gg 1$ (with then implicit constant only depending on P).

Thus $\sum_{m\in T_M-S_M} l_P(m) = \sum_{m\in T_M-S_M} O(M^{1/2}) = o(\sum_{m\in T_M-S_M} C.Z(m))$. $\square$

Now it only remains to treat the case when P is ordinary and $\mathrm{rk}_{\mathbb{Z}} L_0 = 3$. We first construct S_M for such P.

Lemma 9.3.2. *Given M, set $n_0 = \lceil (1+\epsilon_0)\log_p M\rceil$ and $S_M = \{m \in T_M \mid \exists v \in L_{n_0}$ with $Q'(v) = m\}$. Then $\#S_M = o(\#T_M)$.*

Proof. By a geometry-of-numbers argument and Lemma 7.3.2, we have

$$\#S_M \le \{v \in L_{n_0} \mid Q'(v) \le M\} = O(M^{3/2}/p^{n_0} + M/b_{n_0} + M^{1/2}/a_{n_0}),$$

where a_{n_0} is the minimal length of a non-zero vector in L_{n_0} and b_{n_0} is the minimal root discriminant of a rank 2 sublattice in L_{n_0}. Since C does not have any global special endomorphisms, we have $a_{n_0}, b_{n_0} \to \infty$ as $M \to \infty$. Fix $0 < \epsilon_1 < \epsilon_0/4$. We prove the desired estimate by a case-by-case discussion based on the size of a_{n_0}, b_{n_0}.

(1) $a_{n_0} < M^{\epsilon_1}$ and $b_{n_0} > M^{1/2+2\epsilon_1}$. Then we conclude as in the proof Proposition 7.2.5 for Theorem 1(1) for large n case (3). More precisely all $v \in L_{n_0}$ with $Q'(v) \le M$ lie in a rank 1 sublattice of L_{n_0} and thus the total number of such v is $o(\#T_M)$.

(2) $a_{n_0} \ge M^{\epsilon_1}$ and $b_{n_0} > M^{1/2+2\epsilon_1}$. Then

$$\#S_M = O(M^{3/2}/p^{n_0} + M/b_{n_0} + M^{1/2}/a_{n_0}) = O(M^{1/2-\epsilon_1}) = o(M^{1/2}/\log M).$$

(3) $b_{n_0} \le M^{1/2+2\epsilon_1}$. Then $p^{n_0/2} = M^{1/2+\epsilon_0/2} \ge b_{n_0}$ and by Lemma 9.1.2 (note the proof of this lemma applies to this case), for $M \gg 1$, if $m \in S_M$, then m is represented by the binary quadratic form given by restricting Q' to the rank 2 sublattice in L_{n_0} with minimal discriminant($=b_{n_0}^2$). Since $b_{n_0} \to \infty$, then we conclude by Proposition 9.1.4 for Theorem 1(1) and Remark 2 and by the fact that the density of primes represented by such quadratic forms goes to 0 for Theorem 3. $\square$

Now we estimate the total local contribution at an ordinary point with $\mathrm{rk}_{\mathbb{Z}} L_0 = 3$.

Proposition 9.3.3. *Assume P is ordinary with $\mathrm{rk}_{\mathbb{Z}} L_0 = 3$. After possible enlarging S_M in Lemma 9.3.2 (still with $\#S_M = o(\#T_M)$), we have $\sum_{m\in T_M-S_M} l_P(m) = o(\sum_{m\in T_M-S_M} C.Z(m))$.*

Proof. Notation as in Lemma 7.3.2. By Lemmas Lemma 7.2.1, Lemma 7.3.2, and Lemma 9.3.2, we have

$$\sum_{m\in T_M-S_M} l_P(m) = \sum_{m\in T_M-S_M} A(r_0(m) + \sum_{n=1}^{[(1+\epsilon_0)\log_p M]} (p^n - p^{n-1}) r_n(m)) \ll \sum_{n=0}^{[(1+\epsilon_0)\log_p M]} p^n \sum_{m\in T_M-S_M} r_n(m).$$

Notation as in Lemma 9.3.2. We have $\sum_{m=1}^{M} r_n(m) = O(M^{3/2}/p^n + M/b_n + M^{1/2}/a_n)$.

For Theorem 3, we have

$$\sum_{m\in T_M-S_M} l_P(m) \ll \sum_{n=0}^{[(1+\epsilon_0)\log_p M]} p^n (M^{3/2}/p^n + M) = O(M^{2+\epsilon_0}) = o(M^{5/2}/\log M),$$

when we take $\epsilon_0 < 1/2$.

For Theorem 1(1) and Remark 2, set $n_1 = \lceil \frac{3}{4}\log_p M\rceil$. First,

$$\sum_{n=0}^{n_1} p^n \sum_{m\in T_M-S_M} r_n(m) \ll \sum_{n=0}^{n_1} p^n (M^{3/2}/p^n + M) = O(M^{7/4}) = o(M^2/\log M).$$

Second, for $\sum_{n=n_1}^{[(1+\epsilon_0)\log_p M]} p^n \sum_{m\in T_M-S_M} r_n(m)$, we bound it by studying the following two cases separately.

(1) $b_{n_1} \geq M^{1/8}$. As in the first part, we have

$$\sum_{n=n_1}^{[(1+\epsilon_0)\log_p M]} p^n \sum_{m\in T_M - S_M} r_n(m) \ll \sum_{n=n_1}^{[(1+\epsilon_0)\log_p M]} p^n(M^{3/2}/p^n + M/b_n + M^{1/2}),$$

which is $O(M^{3/2}\log M + M^{2+\epsilon_0-1/8} + M^{3/2+\epsilon_0}) = o(M^2/\log M)$ once we take $\epsilon_0 < 1/8$.

(2) $b_{n_1} < M^{1/8}$. We are going to enlarge S_M to be $\{m \in T_M \mid \exists v \in L_{n_1} \text{ with } Q'(v) = m\}$. Since $b_{n_1}^2 M < M^{5/4} = o(p^{2n_1})$, then we conclude, as in the upper medium range Case (2) in the proof of Proposition 7.2.5 for Theorem 1(1), by Lemma 9.1.2 and Proposition 9.1.4 that $\#S_M = o(\#T_M)$ and $\sum_{n=n_1}^{[(1+\epsilon_0)\log_p M]} p^n \sum_{m\in T_M - S_M} r_n(m) = 0$. $\square$

Proof of Theorem 1(1), Remark 2, and Theorem 3. Assume for contradiction that there are only finitely many points on $C \cap (\cup_{m\in T} Z(m))$. Then we construct S_M by taking the union of the S_M in Proposition 7.2.5 for supersingular points and that in Lemma 9.3.2 and Proposition 9.3.3 for ordinary points with $\mathrm{rk}_{\mathbb{Z}} L_0 = 3$. Since it is a finite union, we still have $\#S_M = o(\#T_M)$. We deduce a contradiction by Lemma 7.1.1, Proposition 7.2.5, Lemma 9.3.1, and Proposition 9.3.3. $\square$

Appendix A. The decay lemma in the supergeneric case

Notation as in §5, especially §5.1. The goal of this section is to prove the following theorem:

Theorem A.0.1 (Decay Lemma)**.** *At a supersingular point P on C, there exists a rank 3 submodule of the lattice $L' \otimes \mathbb{Z}_p$ of special endomorphisms which decays rapidly (see Definition 5.1.1).*

We will first prove the Siegel case, i.e., $L = L_S$ and then we deduce the Hilbert case ($L = L_H$) from the proof of the Siegel case. Since we have proved the above theorem when P is superspecial in §§5-6, we will only treat the case when P is supergeneric in this appendix and hence we only need to consider the case when p is inert for the Hilbert case.

A.1. **Siegel case.** First, we follow §3.3 and let $w_1 := pv_1 + v_3 + (c+\sigma^{-1}(c))v_4$, $w_2 := \lambda p v_1 - \lambda v_3 + \lambda(c - \sigma^{-1}(c))v_4$, $w_3 := v_4$, $w_4 := -\sigma^{-1}(c)pv_1 + pv_2 - cv_3 - \sigma^{-1}(c)cv_4$, and $w_5 := v_5$, where recall that $\lambda \in \mathbb{Z}_{p^2}^\times$ such that $\sigma(\lambda) = -\lambda$. Then a direct computation implies that

$$\mathrm{Span}_{\mathbb{Z}_p}\{w_1, \ldots, w_5\} = \mathbb{L}_{\mathrm{cris},P}(W)^{\varphi=1} = L' \otimes \mathbb{Z}_p.$$

By §3.1.5 and using the universal unipotent u in §3.3, we have that the Frobenius on $\mathbb{L}_{\mathrm{cris}}(W[[x,y,z]])$, with respect to the basis $\{w_i\}_{i=1}^5$, is $I + \frac{y}{p}\mathbf{A} + \frac{Q}{p}\mathbf{B} + x\mathbf{C} + z\mathbf{D}$, where $Q = xy + \frac{z^2}{4\epsilon}$,

$$\mathbf{A} = \begin{bmatrix} 0 & c\lambda & \frac{1}{2} & \frac{c^2}{2} & 0 \\ \frac{c}{\lambda} & 0 & \frac{1}{2\lambda} & \frac{-c^2}{2\lambda} & 0 \\ -c^2 & -\lambda c^2 & -c & 0 & 0 \\ -1 & \lambda & 0 & c & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \frac{-1}{2} & \frac{\lambda}{2} & 0 & \frac{c}{2} & 0 \\ \frac{-1}{2\lambda} & \frac{1}{2} & 0 & \frac{c}{2\lambda} & 0 \\ c & -c\lambda & 0 & -c^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2\lambda} & 0 \\ -1 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{2p} \\ 0 & 0 & 0 & 0 & \frac{1}{2\lambda p} \\ 0 & 0 & 0 & 0 & \frac{-c}{p} \\ 0 & 0 & 0 & 0 & 0 \\ \frac{-1}{2\epsilon} & \frac{\lambda}{2\epsilon} & 0 & \frac{c}{2\epsilon} & 0 \end{bmatrix}.$$

To lighten notation, let $\mathbf{G}$ denote the matrix $p\mathbf{D}\mathbf{D}^{(1)}$. We then have:

$$\mathbf{G} = \begin{bmatrix} \frac{-1}{4\epsilon} & \frac{-\lambda}{4\epsilon} & 0 & \frac{c^{(1)}}{4\epsilon} & 0 \\ \frac{-1}{4\lambda\epsilon} & \frac{-1}{4\epsilon} & 0 & \frac{c^{(1)}}{4\lambda\epsilon} & 0 \\ \frac{c}{2\epsilon} & \frac{c\lambda}{2\epsilon} & 0 & \frac{-cc^{(1)}}{2\epsilon} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-1}{2\epsilon} \end{bmatrix}.$$

Here, as in §3.3, we assume that $c \in W$ is the Teichmuller lift of its reduction $\bar{c}$ in k; since P is supergeneric, we have that $\bar{c} \neq \sigma^2(\bar{c})$ and hence $c \pm \sigma(c) \in W^\times$.

As in §5.1, we consider the formal curve $C = \operatorname{Spf} k[[t]]$ which is generically ordinary and specializes to P; the map $C \to \mathcal{M}_k$ gives rise to a map $k[[x,y,z]] \to k[[t]]$, and we denote by $x(t), y(t), z(t)$ the images of x, y, z. The proof of Theorem A.0.1 in the case $y(t) = 0$ is much simpler than that of the case $y(t) \neq 0$, so in the rest of this appendix, we will present the proof of Theorem A.0.1 assuming $y(t) \neq 0$; similar ideas (with simpler case-by-case discussion) yield the proof when $y(t) = 0$. Without loss of generality, we assume that $y(t) = t^{v_y}$ for some $v_y \in \mathbb{Z}_{>0}$ and $z(t)$ has t-adic valuation v_z, and that the leading coefficient of $z(t)^2/4\epsilon$ (here we also use ϵ to denote the reduction of $\epsilon \in \mathbb{Z}_p$ mod p) is α (set $v_z = \infty$ and $\alpha = 0$ if $z(t) = 0$). Let $\tilde{\alpha} \in W$ be the Teichmuller lift of α, and define $\mathbf{E} = \mathbf{A} + \tilde{\alpha}\mathbf{B}$.

Definition A.1.1. For $n \in \mathbb{Z}_{>0}$, define X_n to be the product $\prod_{i=0}^{n} X^{(i)}$, where X stands for either $\mathbf{A}, \mathbf{B}$ or $\mathbf{E}$.

We record the following lemma which we will make crucial use of in the proof of the Decay Lemma.

Lemma A.1.2. *(1) The matrices $\mathbf{A}_n$ and $\mathbf{A}_n \bmod p$ have rank 1. Further, the first four rows of $\mathbf{A}_n$ are each scalar multiples of $R^{(n-1)}$ by p-adic units, where*

$$R = \begin{bmatrix} c + c^{(1)} & \lambda(c - c^{(1)}) & 1 & -cc^{(1)} & 0 \end{bmatrix}.$$

(2) $\mathbf{B}_n$ and $\mathbf{B}_n \bmod p$ have rank 1. Further, the first three rows of $\mathbf{B}_n$ are each scalar multiples of $S^{(n-1)}$ by p-adic units, where

$$S = \begin{bmatrix} 1 & \lambda & 0 & -c^{(1)} & 0 \end{bmatrix}.$$

(3) $\mathbf{E}_n$ has rank ≤ 1. Further, the rows of $\mathbf{E}_n$ are contained in the span of $R^{(n-1)} - \alpha^{(n)} S^{(n-1)}$. If $n \geq 2$ and $\mathbf{E}_n \bmod p$ has rank 1 if and only if $c - c^{(2)} - \alpha^{(1)}$ is a p-adic unit.

Proof. As each of the above cases follow from similar straightforward inductive calculations, we will content ourselves with solely proving part (2). That the rows are spanned by $S^{(n-1)}$ follows directly from the fact that the row span of $\mathbf{B}_n$ is contained in the row span of $\mathbf{B}^{(n)}$, which is indeed equal to the span of $S^{(n-1)}$.

We will prove that the first three rows of $\mathbf{B}_n$ are scalar multiples of $S^{(n-1)}$ by p-adic units by an inductive argument. The case of $n = 1$ can be checked by a direct calculation. Observing that $\mathbf{B}_n = \mathbf{B}_{n-1}\mathbf{B}^{(n)}$, it suffices to prove that the 1×5 matrix $S^{(n-1)} \cdot \mathbf{B}^{(n)}$ has its first, second and fourth entries equalling p-adic units. This is also seen by a direct calculation, and the lemma follows. $\square$

For the remainder of our paper, we will replace $c, \bar{c}$ by $\tilde{c}, c$ respectively. We will also abuse notation by allowing λ to denote both λ and its mod p reduction – it will be clear from context what we refer to.

As in the proof of Theorem 5.1.2 in §5.1, we consider the matrix F_∞, which has a product expansion $\prod_{m=0}^{\infty}(I + \frac{y(t)}{p}\mathbf{A} + \frac{Q(t)}{p}\mathbf{B} + x(t)\mathbf{C} + z(t)\mathbf{D})^{(m)}$ (note that as in §5.1, the twist is with respect to $\sigma(x) = x^p, \sigma(y) = y^p, \sigma(z) = z^p$ and then we specialize to $x = x(t), y = y(t), z = z(t)$). Given any positive integer n, the coefficient of $1/p^n$ in F_∞ having minimal t-adic valuation arise from products of (Frobenius-twisted) powers of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$, with $\mathbf{C}$ contributing only larger order terms. We will therefore ignore $\mathbf{C}$ while proving our decay results.

We will need to consider the following cases:

A.2. Case 1: $v_y > 2v_z$.

Proof of Theorem A.0.1. We will use part (2) of Lemma A.1.2 to prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_4\}$ decays rapidly. As $v_y > 2v_z$, the t-adic valuation of $Q = xy + z^2/(4\epsilon)$ equals $2v_z$, which is strictly smaller than that of y. Moreover, by Corollary 3.4.2, the equation of the non-ordinary locus is $(x + a)y + \frac{z^2}{4\epsilon}$ and hence $A = 2v_z$, where A in Definition 5.1.1 is the t-adic valuation of the equation. Further, the t-adic valuation of z^{1+p} equals $(p + 1)v_z$, which is strictly greater than $2v_z = v_t(Q)$. Therefore, the term in the coefficient of $1/p^{n+1}$ with minimum t-adic valuation equals $(Q\mathbf{B}) \cdots (Q\mathbf{B})^{(n)}$. In order to prove that the property DR (in the statement of Proposition 5.1.3) holds, by Lemma A.1.2(2), it suffices to prove that the kernel of the matrix $S^{(n-1)} \bmod p$ does not contain any non-zero $\mathbb{F}_p$-linear combinations of $w_1, w_2, w_4 \bmod p$. Indeed, this follows directly from the fact that c and therefore all its Frobenius twists do not lie in $\mathbb{F}_{p^2}$. Thus Theorem A.0.1 follows from the property DR by the same argument that Proposition 5.1.3 implies Theorem 5.1.2 in §5.1. $\square$

A.3. Case 2: $v_y < 2v_z$.

Proof of Theorem A.0.1. It suffices to prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$ decays rapidly. As $v_y < 2v_z$, the term in the coefficient of $1/p^{n+1}$ with minimum t-adic valuation equals $(y\mathbf{A}) \dots (y\mathbf{A})^{(n)}$. Note also that A in Definition 5.1.1 is $v_t((x + a)y + \frac{z^2}{4\epsilon}) = v_y$. As in Case 1, it suffices to prove that the kernel of $\mathbf{A} \cdots \mathbf{A}^{(n)} \bmod p$ contains no non-zero $\mathbb{F}_p$-linear combinations of $w_1, w_2, w_3 \bmod p$.

To that end, let w be some $\mathbb{F}_p$-linear combination of these three vectors, which were in the kernel of $\mathbf{A} \cdots \mathbf{A}^{(n)}$. By Lemma A.1.2(1), the (first four) rows of $\mathbf{A} \cdots \mathbf{A}^{(n)}$ are unit-scalar multiples of $R^{(n-1)}$. Therefore, $R^{(n-1)} \cdot w = 0$. As w is Frobenius-invariant, this is equivalent to $R \cdot w = 0$.

The existence of w implies that there exist $s_1, s_2, s_3 \in \mathbb{F}_p$ such that $s_1(c+c^{(1)})+s_2\lambda(c-c^{(1)})+s_3 = 0$. If either $s_1 = 0$ or $s_2 = 0$, it follows that either $c + c^{(1)} \in \mathbb{F}_p$ or $\lambda(c - c^{(1)}) \in \mathbb{F}_p$. Either case would imply that $c = c^{(2)}$, which is a contradiction as we have assumed that $c \notin \mathbb{F}_{p^2}$. Therefore, we assume that $s_1 = -1$. We now have

$$c + c^{(1)} = \lambda s_2(c - c^{(1)}) + s_3,$$

and therefore applying σ to the above equation, we also have

$$c^{(1)} + c^{(2)} = \lambda s_2(c^{(2)} - c^{(1)}) + s_3.$$

Subtracting the two equations yields

$$c - c^{(2)} = \lambda s_2(c - c^{(2)}).$$

This is a contradiction, as $\lambda \notin \mathbb{F}_p$. Therefore, such w could not have been in the kernel of $\mathbf{A} \cdots \mathbf{A}^{(n)} \bmod p$, whence it follows that every primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$ satisfies the condition DR in the statement of Proposition 5.1.3, as required. Therefore, Theorem A.0.1 follow in this case. $\square$

A.4. **Case 3:** $v_y = 2v_z$**.** In this case, it follows that $z(t) \neq 0$, and thus $\alpha \neq 0$.

Subcase 1: $\alpha^{(1)} - c + c^{(2)} \neq 0$**.**

Proof of Theorem A.0.1. We will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_i, w_3, w_5\}$ decays rapidly where i is either 1 or 2. In this case, we have $A \geq 2v_z$. The conditions imposed on v_y, v_z imply that the term with minimal t-adic valuation in the coefficient of $1/p^{n+1}$ is

$$\mathbf{E}\mathbf{E}^{(1)} \cdots \mathbf{E}^{(n)} t^{2v_z(1+p+\cdots+p^n)} + \mathbf{E}\mathbf{E}^{(1)} \cdots \mathbf{E}^{(n-1)}(p\mathbf{D})^{(n)} t^{2v_z(1+p+\cdots+p^{n-1})+v_z p^n}.$$

Note that the first term in the sum has its last column equalling zero, and the second term has its first four columns equalling zero.

For brevity, we will henceforth denote a Frobenius twist with a superscript of $'$. We claim that at least one of $c + c' - \alpha'$ and $\lambda(c - c' - \alpha')$ does not lie in $\mathbb{F}_p$. Indeed, the first element being in $\mathbb{F}_p$ implies that $c' + c'' - \alpha'' = c + c' - \alpha'$, consequently $-(\alpha' - \alpha'') = c'' - c$. Similarly, $\lambda(c - c' - \alpha')$ being an element of $\mathbb{F}_p$ implies that $\alpha' + \alpha'' = c - c''$. Therefore, both elements being in $\mathbb{F}_p$ implies that $\alpha' + \alpha'' = \alpha' - \alpha''$, which in turn implies that $\alpha = 0$, a contradiction.

Without loss of generality, we will assume that $c + c' - \alpha' \notin \mathbb{F}_p$, and prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_3, w_5\}$ decays rapidly (if $\lambda(c - c' - \alpha') \notin \mathbb{F}_p$, then an identical argument would yield that $\mathrm{Span}_{\mathbb{Z}_p}\{w_2, w_3, w_5\}$ decays rapidly). We first prove that every primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_3\}$ decays rapidly. Indeed, let w be any primitive vector. In order to prove that w decays rapidly, it suffices to prove that $w \bmod p$ is not in the kernel of $\mathbf{E}\mathbf{E}^{(1)} \cdots \mathbf{E}^{(n)} \bmod p$. However, as the 4th row of this matrix is a unit-multiple of $R^{(n-1)} - \alpha^{(n)} S^{(n-1)}$ by Lemma A.1.2 (3), it suffices to prove that $w \bmod p$ is not in the kernel of the 1×5 matrix $R^{(n-1)} - \alpha^{(n)} S^{(n-1)}$. As $w \bmod p$ is $\mathbb{F}_p$-rational, this is equivalent to asking that $w \bmod p$ is not in the kernel of $R - \alpha' S \bmod p$. This follows from the fact that $c + c' - \alpha' \notin \mathbb{F}_p$.

We now show that w_5 decays rapidly. Indeed, the last column of $\mathbf{E}\mathbf{E}^{(1)} \cdots \mathbf{E}^{(n-1)}(p\mathbf{D})^{(n)}$ has $(R^{(n-2)} - \alpha^{(n-1)} S^{(n-2)}) \cdot v^{(n)}$ as its fourth entry (up to a unit), where v is the last column of $p\mathbf{D}$. It suffices to prove that $(R^{(n-2)} - \alpha^{(n-1)} S^{(n-2)}) \cdot v^{(n)} \neq 0 \bmod p$, equivalently $(R - \alpha' S) \cdot v^{(2)} \neq 0 \bmod p$. A direct computation shows that this element equals $-(\alpha' - c + c'')$, which we have assumed to be not zero. Therefore, it follows that w_5 decays rapidly.

Let w denote a primitive vector in the span of w_1, w_3. Consider a $\mathbb{Z}_p$-linear combination $aw + bw_5$, where either a or b is a p-adic unit. The only way for $aw + bw_5$ to not decay rapidly is if the t-adic valuation of the coefficient of $1/p^{n+1}$ in $F_\infty w$ equalled the t-adic valuation of the coefficient of $1/p^{m+1}$ in $F_\infty w_5$. However, the former equals $t^{2v_z(1+p+\ldots+p^n)}$ and the latter equals $t^{2v_z(1+p+\ldots+p^{m-1})+v_z p^m}$. These two quantities are clearly never equal, thereby establishing the required decay as in Case 1 in §5.2. $\square$

Subcase 2: $\alpha^{(1)} - c + c^{(2)} = 0$**.** In this case, we see that $\alpha = c^{(-1)} - c^{(1)}$, and thus by Corollary 3.4.2, the local equation of the non ordinary locus is given by the equation $Q(t) - \alpha y(t)$. In particular, we see that $H(t) := Q(t) - \alpha y(t) = Q(t) - \alpha t^{v_y} \neq 0$. We will now express

$$\mathrm{Frob} = (I + \frac{y(t)}{p}\mathbf{E} + \frac{H(t)}{p}\mathbf{B} + x(t)\mathbf{C} + z(t)\mathbf{D}) \circ \sigma,$$

where $\mathbf{E}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are defined in §A.1. We will establish the required decay by considering the fourth row of F_∞. As mentioned above, we omit $\mathbf{C}$ and powers of $x(t)$ in this analysis, as there are no negative powers of p in the entries of $\mathbf{C}$.

We need the following lemma:

Lemma A.4.1. *Consider all products of the form $W_0W_1W_2 \dots W_n$, where W_i is the i^{th} Frobenius twist of $\mathbf{E}, \mathbf{B}$ or $\mathbf{D}$. The only products which have a non-zero fourth row are those with the following properties:*

(1) $W_0 = \mathbf{E}$.

(2) Suppose that $W_i, W_j \neq$ twists of $\mathbf{D}$ but $W_{i+1}, \dots, W_{j-1} =$ twists of $\mathbf{D}$ for $1 \leq i < j \leq n$. Then, $j - i$ has to be odd. Equivalently, any maximal consecutive subsequence consisting exclusively of Frobenius twists of $\mathbf{D}$ has to have even length, unless the subsequence is terminates with W_n.

(3) Apart from $i = 0$, the only possible $j \leq n$ such that $W_j = \mathbf{E}^{(j)}$ is $j = n$.

Further, a product that satisfies the above properties does indeed have a non-zero fourth row.

Finally, consider the length four vector given by the first four rows of the last column of the product. This vector is nonzero if and only if $W_n = \mathbf{D}^{(n)}$ and the number of occurrences of twists of $\mathbf{D}$ is odd. If this is the case, the first four columns of the product are all zero.

Proof. (1)(2) are clear. Part (3) follows from a direct computation. We will illustrate this computation in the particular case

$$\mathbf{E}\prod_{i=1}^{m}\mathbf{B}^{(i)}\prod_{j=m+1}^{m+2e}\mathbf{D}^{(j)}\mathbf{E}^{(m+2e+1)} = \frac{1}{p^e}\mathbf{E}\prod_{i=1}^{m}\mathbf{B}^{(i)}\prod_{j=m+1}^{m+e}\mathbf{G}^{(2j-m-1)}\mathbf{E}^{(m+2e+1)}. \tag{A.4.1}$$

It will follow from explicitly computing the product that multiplying it by either $\mathbf{E}^{(m+2e+2)}$, $\mathbf{D}^{(m+2e+2)}$ or $\mathbf{B}^{(m+2e+2)}$ yields the zero matrix. The other cases (where the W_i are other choices of $\mathbf{B}, \mathbf{D}$) are entirely analogous, and the same computation goes through.

An easy inductive argument shows that, the product $\prod_{i=1}^{m}\mathbf{B}^{(i)}\prod_{j=m+1}^{m+2e}\mathbf{D}^{(j)} = \frac{1}{p^e}\prod_{i=1}^{m}\mathbf{B}^{(i)}\prod_{j=m+1}^{m+e}\mathbf{G}^{(2j-m-1)}$ equals

$$\frac{(-1)^{m+e}}{(2p\epsilon)^e}\begin{bmatrix} \frac{1}{2} & \frac{(-1)^{m+1}\lambda}{2} & 0 & \frac{-\tilde{c}^{(2e+m)}}{2} & 0 \\ \frac{-1}{2\lambda} & \frac{(-1)^m}{2} & 0 & \frac{\tilde{c}^{(2e+m)}}{2\lambda} & 0 \\ -\tilde{c}^{(1)} & (-1)^m\lambda\tilde{c}^{(1)} & 0 & \tilde{c}^{(1)}\tilde{c}^{(2e+m)} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * \end{bmatrix}.$$

Multiplying this matrix on the left by the fourth row of $\mathbf{E}$ and on the right by $\mathbf{E}^{(m+2e+1)}$ and using the relation $\alpha^{(1)} = c - c^{(2)}$ yields

$$\frac{(-1)^{m+e}}{(2\epsilon p)^e}\begin{bmatrix} -\tilde{c}^{(2e+m+1)} - \tilde{c}^{(2e+m+2)} & (-1)^m\lambda(\tilde{c}^{(m+2e+1)} - \tilde{c}^{(m+2e+2)}) & -1 & \tilde{c}^{(m+2e+1)}\tilde{c}^{(m+2e+2)} & 0 \end{bmatrix}.$$

Note that the product (not just the fourth row) matrix has rank one, and so every other row must be some multiple of this row. In order to show that the product in (A.4.1) multiplied by W_{m+2e+2}

(where W_{m+2e+2} is either $\mathbf{B}^{(m+2e+2)}, \mathbf{D}^{(m+2e+2)}$ or $\mathbf{E}^{(m+2e+2)}$) is zero, it suffices to prove that the fourth row of this product is zero. This can be checked by direct computation. $\square$

We record the fourth row of the product in (A.4.1) (scaled by p^e) for future use:

Lemma A.4.2. *The fourth row of the product* $p^e\mathbf{E}\prod_{i=1}^{m}\mathbf{B}^{(i)}\prod_{j=m+1}^{m+2e}\mathbf{D}^{(j)}\mathbf{E}^{(m+2e+1)}$ *equals*

$$\frac{(-1)^{m+e}}{(2\epsilon)^e}\begin{bmatrix}-\tilde{c}^{(2e+m+1)}-\tilde{c}^{(2e+m+2)} & (-1)^m\lambda(\tilde{c}^{(m+2e+1)}-\tilde{c}^{(m+2e+2)}) & -1 & \tilde{c}^{(m+2e+1)}\tilde{c}^{(m+2e+2)} & 0\end{bmatrix}.$$

Let $H(t)=\eta t^A+t^{A+1}(z_2(t))$, where $\eta\in k^\times$. Let $\tilde{\eta}$ denote the Teichmuller lift of η.

We break this final case into two cases.

Subsubcase 1: $v_z(2p^{2e}-p^{2e-1}+1)<A<v_z(2p^{2e+2}-p^{2e+1}+1)$**.** In this case, we will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2,w_3\}$ decays rapidly.

Lemma A.4.3. *The term with minimal t-adic valuation in the coefficient of $1/p^{n+1}$ in the fourth row of (the top-left 4×4 block of) F_∞ is*

$$p^e\mathbf{E}\mathbf{B}^{(1)}\cdots\mathbf{B}^{(n-e-1)}\mathbf{D}^{(n-e)}\cdots\mathbf{D}^{(n+e-1)}\mathbf{E}^{(n+e)}.$$

Proof. Note that F_∞ is composed of sums of products as in Lemma A.4.1, where each W_i is multiplied by:

- $y(t)^{(i)}=t^{2v_zp^i}$ if $W_i=\mathbf{E}^{(i)}$;
- $h(t)^{(i)}=\tilde{\eta}^{(i)}t^{Ap^i}+\cdots$ if $W_i=\mathbf{B}^{(i)}$;
- $z(t)^{(i)}=\beta^{(i)}t^{v_zp^i}$ if $W_i=\mathbf{D}^{(i)}$, where β is the leading coefficient in $z=\beta t^{v_z}+\cdots$.

Consider products as in Lemma A.4.1. As we are looking for matrices where the first four columns are not all zero, it follows that the number of occurrences of twists of $\mathbf{D}$ must be even. Therefore, consider a product with n_1 occurrences of twists of either $\mathbf{E}$ or $\mathbf{B}$, and $2n_2$ occurrences of twists of $\mathbf{D}$. The fourth row of such a product would have a p-adic valuation of $-(n_1+n_2)$, and hence we assume that $n+1=n_1+n_2$.

It is clear that the t-adic valuation of the expression is minimized if the first and last matrices in the product are both twists of $\mathbf{E}$. Indeed, the t-adic valuation is minimized when $W_i=\mathbf{E}^{(i)}$ for as many i as possible, and Lemma A.4.1 implies that this happens when the first and last matrices are both twists of $\mathbf{E}$. As the t-adic valuation of $H(t)$ is strictly greater than that of $z(t)$, it follows that products of the form $\mathbf{E}\mathbf{B}^{(1)}\cdots\mathbf{B}^{(n_1-2)}\mathbf{D}^{n_1-1}\cdots\mathbf{D}^{(n_1+2n_2-2)}\mathbf{E}^{(n_1+2n_2-1)}$ contain the term with minimal t-adic valuation.

As in Case 4 in §5.2, a convexity argument yields that the t-adic valuation is minimized exactly for the product listed in the statement of this result, thereby concluding the proof. $\square$

Proof of Theorem A.0.1 in this case. We will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1,w_2,w_3\}$ decays rapidly. Let R_e denote the mod p reduction of the row detailed in Lemma A.4.2. It suffices to show that there is no $\mathbb{F}_p$-linear combination of the first three entries of R_e which evaluates to zero. This is equivalent to prove that the elements $c+c^{(1)}$, $\lambda(c-c^{(1)})$ and 1 are $\mathbb{F}_p$-linearly independent. This fact has already been established in Case 2 in §A.3. The result follows. $\square$

Subsubcase 2: $v_z(2p^{2e}-p^{2e-1}+1)=A$**.**

Lemma A.4.4. *(1) There are two terms with minimal t-adic valuation in the coefficient of $1/p^{n+1}$ in the fourth row of (the top-left 4×4 block of) F_∞. They are:*

- $p^e\mathbf{E}\mathbf{B}^{(1)}\cdots\mathbf{B}^{(n-e-1)}\mathbf{D}^{(n-e)}\cdots\mathbf{D}^{(n+e-1)}\mathbf{E}^{(n+e)}\eta^{(1+p+\cdots+p^{n-e-1})}\beta^{p^{n-e}+\cdots+p^{n+e-1}}$*; and*
- $p^{e-1}\mathbf{E}\mathbf{B}^{(1)}\cdots\mathbf{B}^{(n-e)}\mathbf{D}^{(n-e+1)}\cdots\mathbf{D}^{(n+e-2)}\mathbf{E}^{(n+e-1)}\eta^{(1+p+\cdots+p^{n-e})}\beta^{p^{n-e+1}+\cdots+p^{n+e-2}}$.

(2) *The term with minimal t-adic valuation in the coefficient of $1/p^{n+1}$ in the fourth row of the last column of F_∞ is*

$$\mathbf{E}\mathbf{B}^{(1)}\cdots\mathbf{B}^{(n-e-1)}\mathbf{D}^{(n-e)}\cdots\mathbf{D}^{(n+e)}.$$

Proof. The proof goes along the same lines as that of Lemma A.4.3, and so we will not spell out the details. □

Proof of the final case of Theorem A.0.1. We will show that there exists a rank 2 submodule $W \subset \mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$ such that $W + \mathbb{Z}_p w_5$ decays rapidly.

The term with minimal t-adic valuation in the coefficient of $1/p^{n+1}$ of the fourth row of the top-left 4×4 block of F_∞ is the sum of the corresponding parts in the two matrices in Lemma A.4.4(1). The t-adic valuation of this sum is $v_z(2+p^{n-e}+\cdots+p^{n+e-1}+2p^{n+e})+A(p+\cdots+p^{n-e-1})$, which is $\leq A(1+\cdots+p^n)$.

In order to prove that W decays rapidly with the above t-adic valuation, it suffices to show that the kernel of the sum of the fourth rows of the two matrices in Lemma A.4.4(1) mod $p \cap W$ mod p is $\{0\}$ for a suitable choice of W. To that end, let $\mu = \frac{\beta^{p^{n-e}+p^{n+e-1}}}{2\epsilon}$ and $\nu = \eta^{p^{n-e}}$. Then, the sum of the two rows with equal minimal t-adic valuation mod p is a scalar multiple[32] of

$$\begin{bmatrix}-\alpha_1 & (-1)^{n-e-1}\alpha_2 & -\alpha_3 & \alpha_4 & 0\end{bmatrix}.$$

Lemma A.4.2 yields that,

- $\alpha_1 = \mu(c^{(n+e)}+c^{(n+e+1)})+\nu(c^{(n+e-1)}+c^{(n+e)})$;
- $\alpha_2 = \lambda\big(\mu(c^{(n+e)}-c^{(n+e+1)})-\nu(c^{(n+e-1)}-c^{(n+e)})\big)$;
- $\alpha_3 = \mu+\nu$.

We now prove that depending on the values of μ and ν, we may take W to be either $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$, $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_3\}$ or $\mathrm{Span}_{\mathbb{Z}_p}\{w_2, w_3\}$. To further lighten notation, let $\delta_1 = (\mu+\nu)c^{(n+e)}$ and $\delta_2 = \mu c^{(n+e+1)}+\nu c^{(n+e-1)}$.

(1) Suppose $\nu+\mu = 0$. We will show that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ decays rapidly. It suffices to prove that there is no non-trivial $\mathbb{F}_p$-linear relation between $\alpha_1 = \mu(c^{(n+e-1)}-c^{(n+e+1)})$ and $\alpha_2 = \lambda\mu(c^{(n+e-1)}-c^{(n+e+1)})$. However, this follows directly from the facts that $\lambda \notin \mathbb{F}_p$ and that $c^{(n+e+1)} \neq c^{(n+e-1)}$.

(2) Suppose that $\delta_1 \neq \pm\delta_2$. We will show that either $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ or $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_3\}$ decays rapidly. Indeed, the former happens when $\alpha_1 = \delta_1+\delta_2$ and $\alpha_2 = \lambda(\delta_1-\delta_2)$ are not $\mathbb{F}_p$ multiples of each other. Therefore, suppose that they were. Then we have that for some $l \in \mathbb{F}_p$, $\delta_1+\delta_2 = l\lambda(\delta_1-\delta_2)$. Note that this yields that $\delta_2/\delta_1 \in \mathbb{F}_{p^2}$. We will prove that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_3\}$ decays rapidly. If not, then $\alpha_1 = s\alpha_3$, where $s \in \mathbb{F}_p$. That is, $\delta_1+\delta_2 = s\delta_1/c^{(n+e)}$. Equivalently, $\delta_2/\delta_1 = (s-c^{(n+e)})/c^{(n+e)}$. As $c \notin \mathbb{F}_{p^2}$, it follows that $\delta_2/\delta_1 \notin \mathbb{F}_{p^2}$, which is a contradiction.

(3) Suppose that $\delta_1 = \delta_2$. We will show that $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_3\}$ decays rapidly, by showing that α_1, α_3 are $\mathbb{F}_p$-linearly independent. Indeed, $\alpha_1 = 2(\nu+\mu)c^{(n+e)}$, and $\alpha_3 = \nu+\mu$. As $c \notin \mathbb{F}_{p^2}$, the two quantities are $\mathbb{F}_p$-linearly independent as required.

(4) Suppose that $\delta_1 = -\delta_2$. The same argument as above works to show that $\mathrm{Span}_{\mathbb{Z}_p}\{w_2, w_3\}$ decays rapidly.

Thus we have proved that there exists a rank 2 submodule W of $\mathrm{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$ which decays rapidly and the minimal t-valuation in the coefficient of $1/p^{n+1}$ is $v_z(2+p^{n-e}+\cdots+p^{n+e-1}+2p^{n+e})+A(p+\cdots+p^{n-e-1})$.

---

[32] The scalar multiple factor is $(-1)^{n-1}\frac{\eta^{1+p+\dots p^{n-e-1}}\beta^{p^{n-e+1}+\dots p^{n+e-2}}}{(2\epsilon)^{e-1}}$.

On the other hand, the term with minimal t-adic valuation in the coefficient of $1/p^{m+1}$ of the last column of F_∞ is the last column of the matrix in Lemma A.4.4(2) - it is easy to see that this last column is non-zero. Hence w_5 decays rapidly. The t-adic valuation of this term is $v_z(2+p^{m-e}+\cdots+p^{m+e-1}+p^{m+e})+A(p+\dots p^{m-e-1})$, and we notice that this is always different from $v_z(2+p^{n-e}+\cdots+p^{n+e-1}+2p^{n+e})+A(p+\cdots+p^{n-e-1})$ regardless of the values of m,n. Therefore, $W+\mathbb{Z}_p w_5$ decays rapidly by the same argument at the end of Case 3 Subcase 1 above. $\square$

A.5. **The case of inert Hilbert modular surfaces.** That we have a rank 3 submodule of special endomorphisms follows directly from the Siegel case. Indeed, by §3.2.1 and §3.3, the F-crystal $\mathbb{L}_{\mathrm{cris}}$ in the setting of inert Hilbert modular surfaces is obtained by setting $z=0$ in the Siegel case. The required decay follows directly from §A.3.

References

[AGHMP17] Fabrizio Andreatta, Eyal Z. Goren, Benjamin Howard, and Keerthi Madapusi Pera, *Height pairings on orthogonal Shimura varieties*, Compos. Math. **153** (2017), no. 3, 474–534.

[AGHMP18] ———, *Faltings heights of abelian varieties with complex multiplication*, Ann. of Math. (2) **187** (2018), no. 2, 391–531.

[Bor98] Richard E. Borcherds, *Automorphic forms with singularities on Grassmannians*, Invent. Math. **132** (1998), no. 3, 491–562.

[Bor99] ———, *The Gross-Kohnen-Zagier theorem in higher dimensions*, Duke Math. J. **97** (1999), no. 2, 219–233.

[Box15] George Andrew Boxer, *Torsion in the Coherent Cohomology of Shimura Varieties and Galois Representations*, ProQuest LLC, Ann Arbor, MI, 2015. Thesis (Ph.D.)–Harvard University.

[Bru02] Jan H. Bruinier, *Borcherds products on O(2, l) and Chern classes of Heegner divisors*, Lecture Notes in Mathematics, vol. 1780, Springer-Verlag, Berlin, 2002.

[Bru17] Jan Hendrik Bruinier, *Borcherds products with prescribed divisor*, Bull. Lond. Math. Soc. **49** (2017), no. 6, 979–987.

[BBGK07] Jan H. Bruinier, José I. Burgos Gil, and Ulf Kühn, *Borcherds products and arithmetic intersection theory on Hilbert modular surfaces*, Duke Math. J. **139** (2007), no. 1, 1–88.

[BK03] Jan Hendrik Bruinier and Ulf Kühn, *Integrals of automorphic Green's functions associated to Heegner divisors*, Int. Math. Res. Not. **31** (2003), 1687–1729.

[BK01] Jan Hendrik Bruinier and Michael Kuss, *Eisenstein series attached to lattices and modular forms on orthogonal groups*, Manuscripta Math. **106** (2001), no. 4, 443–459.

[Cha03] Ching-Li Chai, *Families of ordinary abelian varieties: canonical coordinates, p-adic monodromy, Tate-linear subvarieties and Hecke orbits* (2003). available on https://www.math.upenn.edu/ chai.

[CO06] Ching-Li Chai and Frans Oort, *Hypersymmetric abelian varieties*, Pure Appl. Math. Q. **2** (2006), no. 1, Special Issue: In honor of John H. Coates., 1–27.

[Cha18] François Charles, *Exceptional isogenies between reductions of pairs of elliptic curves*, Duke Math. J. **167** (2018), no. 11, 2039–2072.

[dJ95] A. J. de Jong, *Crystalline Dieudonné module theory via formal and rigid geometry*, Inst. Hautes Études Sci. Publ. Math. **82** (1995), 5–96 (1996).

[Del73] P. Deligne, *Formes modulaires et représentations de* GL(2), Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), Springer, Berlin, 1973, pp. 55–105. Lecture Notes in Math., Vol. 349 (French).

[Del74] Pierre Deligne, *La conjecture de Weil. I*, Inst. Hautes Études Sci. Publ. Math. **43** (1974), 273–307 (French).

[Duk05] W. Duke, *On ternary quadratic forms*, J. Number Theory **110** (2005), no. 1, 37–43.

[Eke87] Torsten Ekedahl, *On supersingular curves and abelian varieties*, Math. Scand. **60** (1987), no. 2, 151–178.

[Han04] Jonathan Hanke, *Local densities and explicit bounds for representability by a quadratic form*, Duke Math. J. **124** (2004), no. 2, 351–388.

[HMP] Benjamin Howard and Keerthi Madapusi Pera, *Arithmetic of Borcherds products*. arXiv:1710.00347.

[HP14] Benjamin Howard and Georgios Pappas, *On the supersingular locus of the* $\mathrm{GU}(2,2)$ *Shimura variety*, Algebra Number Theory **8** (2014), no. 7, 1659–1699.

[HP17] ———, *Rapoport-Zink spaces for spinor groups*, Compos. Math. **153** (2017), no. 5, 1050–1118.

[HY12] Benjamin Howard and Tonghai Yang, *Intersections of Hirzebruch-Zagier divisors and CM cycles*, Lecture Notes in Mathematics, vol. 2041, Springer, Heidelberg, 2012.

[Iwa97] Henryk Iwaniec, *Topics in classical automorphic forms*, Graduate Studies in Mathematics, vol. 17, American Mathematical Society, Providence, RI, 1997.

[IK04] Henryk Iwaniec and Emmanuel Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004.

[Kis10] Mark Kisin, *Integral models for Shimura varieties of abelian type*, J. Amer. Math. Soc. **23** (2010), no. 4, 967–1012.

[KR00] Stephen S. Kudla and Michael Rapoport, *Cycles on Siegel threefolds and derivatives of Eisenstein series*, Ann. Sci. École Norm. Sup. (4) **33** (2000), no. 5, 695–756 (English, with English and French summaries).

[MP16] Keerthi Madapusi Pera, *Integral canonical models for spin Shimura varieties*, Compos. Math. **152** (2016), no. 4, 769–824.

[Mau14] Davesh Maulik, *Supersingular K3 surfaces for large primes*, Duke Math. J. **163** (2014), no. 13, 2357–2425. With an appendix by Andrew Snowden.

[Moo98] Ben Moonen, *Models of Shimura varieties in mixed characteristics*, Galois representations in arithmetic algebraic geometry (Durham, 1996), London Math. Soc. Lecture Note Ser., vol. 254, Cambridge Univ. Press, Cambridge, 1998, pp. 267–350.

[Ogu79] Arthur Ogus, *Supersingular $K3$ crystals*, Journées de Géométrie Algébrique de Rennes (Rennes, 1978), Astérisque, vol. 64, Soc. Math. France, Paris, 1979, pp. 3–86.

[Ogu82] ———, *Hodge cycles and crystalline cohomology*, Hodge cycles, motives, and Shimura varieties, Lecture Notes in Mathematics, vol. 900, Springer-Verlag, Berlin-New York, 1982, pp. 357–414.

[Ogu01] ———, *Singularities of the height strata in the moduli of $K3$ surfaces*, Moduli of abelian varieties (Texel Island, 1999), Progr. Math., vol. 195, Birkhäuser, Basel, 2001, pp. 325–343.

[Sar90] Peter Sarnak, *Some applications of modular forms*, Cambridge Tracts in Mathematics, vol. 99, Cambridge University Press, Cambridge, 1990.

[ST17] Ananth N. Shankar and Yunqing Tang, *Exceptional splitting of reductions of abelian surfaces over number fields* (2017), 1–18. Available on arXiv: 1706.08154.

[Tat71] John Tate, *Classes d'isogénie des variétés abéliennes sur un corps fini (d'après T. Honda)*, Séminaire Bourbaki. Vol. 1968/69: Exposés 347–363, Lecture Notes in Math., vol. 175, Springer, Berlin, 1971, pp. Exp. No. 352, 95–110 (French).

[TZ18] Jesse Thorner and Asif Zaman, *A Chebotarev Variant of the Brun–Titchmarsh Theorem and Bounds for the Lang-Trotter conjectures*, Int. Math. Res. Not. IMRN **16** (2018), 4991–5027.

[Voi02] Claire Voisin, *Théorie de Hodge et géométrie algébrique complexe*, Cours Spécialisés [Specialized Courses], vol. 10, Société Mathématique de France, Paris, 2002 (French). viii+595.

[Wai18] Fabian Waibel, *Fourier coefficients of half-integral weight cusp forms and Waring's problem*, Ramanujan J. **47** (2018), no. 1, 185–200.