REDUCTIONS OF ABELIAN SURFACES OVER GLOBAL FUNCTION FIELDS

DAVESHL 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.

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 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 $m$. The notion of Heegner divisors remains the same for any GSpin Shimura varieties and we will also use $Z(m)$ to denote these Heegner divisors in Hilbert modular surfaces.

Theorem 1. Let $C$ be an irreducible quasi-projective curve in $\mathcal{A}_{2,\mathbb{F}_p}$. Assume that the generic point of $C$ corresponds to an ordinary abelian surface.

(1) If $C$ is not contained in any Heegner divisor $Z(m)$, and if $C$ is projective, then there exist infinitely many $\mathbb{F}_p$-points on $C$ which correspond to non-simple abelian surfaces.

(2) If $C$ is contained in some $Z(m)$ such that $p \nmid m$, then there exist infinitely many $\mathbb{F}_p$-points on $C$ which correspond to abelian surfaces isogenous to self-products of elliptic curves (note that the elliptic curve may vary for these points).

An equivalent statement to Theorem 1(2) is that there exist infinitely many $\mathbb{F}_p$-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 some Hilbert modular surface associated to the real quadratic field $F = \mathbb{Q}(\sqrt{m})$ (if $m$ is a square number, then we obtain the self-product of the modular curve).

Remark. The ordinary generic point assumption 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.

On the other hand, the assumption that $C$ is projective in (1) seems to be a technical assumption and we plan to remove it in future work.

Remark. 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 $\mathbb{F}_p$-points on $C$ such that the corresponding abelian surfaces admit real multiplication by $F$.

To prove Theorem 1(1), we consider the intersection number of $C$ and $Z(\ell^2)$, where $\ell$ is a varying prime number. If we consider $Z(\ell)$ instead, we prove...
Theorem 2. Suppose we have the same assumptions as in Theorem 1(1). Then there are infinitely many ordinary $\overline{F}_p$-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 $\overline{F}_p$-points of complex multiplication by maximal orders, but our current method only asserts maximality for the totally real parts of the endomorphism rings.

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 $A_1 \times A_1$ 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 mod $p$ analogue of the Andre–Oort conjecture. See §1.4 for more details.

One can give a heuristic justification for our theorem as follows. Following results of Achter and Howe in [AH17], the number of non-simple principally polarized abelian surfaces over $F_{q^n}$ is roughly $q^{n(5/2+o(1))}$, and the number of non-simple isogeny classes is roughly $q^{n(1+o(1))}$. Similarly, the total number of isogeny classes in $A_2$ is roughly $q^{n(3/2+o(1))}$. If we treat the map from $C(F_{q^n})$ to the set of $F_{q^n}$ isogeny classes as a random map, the number of $F_{q^n}$ points of $C$ “should” be around $q^{n/2(1+o(1))}$.

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 (see §1.1 remark).

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; we discuss this in more details in §1.3. Moreover, the mixed characteristic analogue of Theorem 1(1) is not yet known (see §1.4 for details).

1.3. Proof of the main results. We view a Hilbert modular surface/a Siegel three-fold as a GSpin Shimura variety attached to a quadratic space $(V,Q)$. The main idea of the proof is to compare the global and local intersection numbers of $C.Z(m)$ for appropriate sequences of $m$. 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 $\frac{3}{4}I(m)$ by studying special endomorphisms.

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 will consist only of squares, and in order to prove Theorem 2 the sequence will consist only of primes. Note that in $A_2$, 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 2.

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

\footnote{The ratio $3/4$ is not a sharp upper bound.}
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 the inclusion $C \subset A_{2\mathbb{F}_p}$, and let $\text{Spf} \mathbb{F}[[t]] \to C$ 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 5.1.2). A similar result appears in the mixed characteristic setting (see [ST17]), but in this case, proving our decay results is much more involved. In particular, we need to use Kisin’s description of the $F$-crystal of special endomorphisms (see [Kis10] §1.4, 1.5) in order to prove the required decay, whereas a straightforward application of Grothendieck–Messing theory suffices in the mixed characteristic setting.

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 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 bound the local contribution by using geometry-of-numbers techniques. To obtain the desired estimate, it requires to 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 (Theorem 5.1.3). 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 $P_n$.

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 5.1.2(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 three-fourths 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 mixed characteristic analogue of Theorem 1 is not yet known. The key difference 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 $\leq 3$. 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 $\mathbb{F}_p$.

Towards the higher dimensional generalization of Theorem 1 there are extra difficulties related to conjectures along the lines of a mod $p$ André–Oort conjecture. More precisely, in analogy with the bi-algebraicity theorem of special subvarieties used in the proof of the André–Oort conjecture, Chai conjectured the following for the mod $p$ case.

---

2Ordinary abelian varieties which have CM are defined over finite fields.
Conjecture 3 ([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.

This conjecture pertains to potential higher dimensional generalizations of Theorem 1 as follows. Let $X$ be an algebraic family of abelian varieties inside a GSpin Shimura variety, and let $x \in X$ be a closed ordinary point. Consider the $p$-divisible group associated to the abelian scheme in a formal neighborhood of $x$. The conjecture provides a non-trivial upper bound of the $\mathbb{Z}_p$-rank of the module of special endomorphisms of this $p$-divisible group.

1.5. Organization of paper. In §2 we recall the Dieudonné module of special endomorphisms of a supersingular point and the $F$-crystal on its deformation space. In §3 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 §4 and §5 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 §6 provides the outline of the main proofs and by geometry-of-numbers arguments, we prove Theorem 1(2) in §7 with inputs from §§3,4 and prove Theorem 1(1) and Theorem 2 in §8 with inputs from §§3,5.

In order to get the main idea of the proof, the reader may focus on Theorem 1(2) and start from §§6,7 and refer back to §§2-4 when necessary.

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.

2. Dieudonné modules of special endomorphisms of supersingular points

In this section, we compute the Dieudonné modules of special endomorphisms of supersingular points with the natural quadratic forms following [HP17] and [LO98]. By [Kis10], we then obtain the $F$-crystals of special endomorphisms on the deformation space of supersingular points. As a direct consequence, we obtain the local equation of the non-ordinary locus in §2.5. These are the key inputs to §4-5. As the preparation for §3 to compare the global and local modular forms, we compute the $\mathbb{Z}$-lattices of special endomorphisms at supersingular points. The Hilbert case is discussed in §§2.1-2.2 and the Siegel case is discussed in §§2.3-2.4.

2.1. The Hilbert case via Howard–Pappas [HP17]. To prove Theorem 1(2), in this subsection and §2.2, we consider the Hilbert modular surface associated to a real quadratic number field $F$.

We first describe how to compute (a certain variant of) the Dieudonné module of special endomorphisms at a supersingular point as a quadratic space following [HP17] §§5-6 and then we apply Kisin’s result [Kis10 §1] to obtain the $F$-crystal.

As the method of Howard–Pappas works for GSpin Shimura varieties of any dimension, we first recall the algorithm in the general setting.

2.1.1. Let $S$ be the special fiber of the canonical integral model of an orthogonal type Shimura variety defined by GSpin$(n, 2)$ with hyperspecial level at $p$ and let $(V, Q)$ denote the quadratic space of signature $(n, 2)$. (Note that some of the notation in this subsection is local and we use different notation later.)

3Unfortunately, this non-trivial upper bound is still not enough for us to apply our strategy to prove Theorem 1 for all orthogonal Shimura varieties.
(1) In the proof of [HPT17, Prop. 5.1.2], they construct all possible maximal vertex lattices \( \Lambda \) (with the quadratic form). Vertex lattices are those lattices such that \( p\Lambda \subset \Lambda^\vee \subset \Lambda \) and its type \( t \) is defined to be the dimension of the \( \mathbb{F}_p \)-vector space \( \Lambda/\Lambda^\vee \). Frobenius acts trivially on these \( \mathbb{Z}_p \)-lattices and the type \( t_{\text{max}} \) of a maximal vertex lattice depends on \( n \) and \( \det(V_{\mathbb{Q}_p}) \). In particular, for our setting, when \( n = 2 \), we consider the exceptional isomorphism \( \text{SO}(2, 2) \cong \text{Res}_{\mathbb{Q}}^\mathbb{R} \text{SL}_2 \), then \( t_{\text{max}} = 4 \) when \( p \) is inert in \( F \) and \( t_{\text{max}} = 2 \) when \( p \) is split in \( F \). When \( n = 3 \), we have \( t_{\text{max}} = 4 \).

(2) Let \( k \) be an algebraic closure of \( \mathbb{F}_p \) and let \( W = W(k) \) and \( K = W[1/p] \). For a \( k \)-point \( y \) on the supersingular locus of \( S \) or on the corresponding Rapoport–Zink space, we consider the corresponding lattice \( L = \{ x \in V_K : x(\text{Fil}^1 D) \subset \text{Fil}^1 D \} \), where \( D \) is the contravariant Dieudonné module of \( y \) over \( W \), we use \( \varphi \) to denote the Frobenius on \( D \), and \( \text{Fil}^1 D = \varphi^{-1}(pD) \). By [HPT17, Prop. 6.2.2, §5.3.1], such lattice \( L \) is given by preimage of Lagrangians \( L \in \Lambda_{W/\Lambda_{W}^\vee} \) such that \( \dim(L + \varphi(L)) = t_{\text{max}}/2 + 1 \), where \( \varphi \) also denotes the induced Frobenius on \( V \).

(3) We can work with the lattice \( L \) since the \( \varphi \)-invariant sublattice in \( L \) preserves both the filtration (this by definition) and the \( W \)-Dieudonné module (by [HPT17, Prop. 6.2.2]). When the supersingular locus is 1-dimensional (from (1), for our applications, the supersingular locus is of dimension 0 or 1), then the superspecial points corresponding to those Lagrangians contained in a smaller vertex lattice. Indeed, by [HPT17] Thm. D(ii),(iii), Prop. 5.3.2, Thm. 6.3.1], the dimension of the Rapoport–Zink space \( R\Lambda \) is \( t/2 - 1 \); for supergeneric points, \( R\Lambda \) is (some irreducible components of) the entire Rapoport–Zink space \( RZ \) of the basic locus and hence it is 1-dimensional. For superspecial points, they lie on the intersection distinct irreducible components of \( RZ \) and the corresponding \( R\Lambda \) is 0-dimensional and hence \( t = 2 \) and this vertex lattice is contained in some maximal vertex lattice.

2.1.2. Notation. We define \([x, y] = Q(x + y) - Q(x) - Q(y)\) to be the bilinear form induced by \( Q \). In \( \mathbb{Q}_p^2 \), we use \( x' \) to denote \( \sigma(x) \). Let \( \lambda \in W(\mathbb{F}_p^\times) \) be chosen such that \( \lambda' = -\lambda \) (to relate to non-square \( u \in \mathbb{Q}_p \), we can take \( \lambda \) to be a root in \( \mathbb{Q}_p^2 \) of \( x^2 - u = 0 \)). We use \( \varphi \) to denote the \( \sigma \)-linear map on Dieudonné modules. Let \( \bar{L} \) denote the mod \( p \) reduction of a \( W \)- or \( \mathbb{Z}_p \)-lattice \( L \). We use \( L^\# \) to denote \( \text{Span}_W(\varphi(L)) \). Recall that \( p \) is an odd prime and unramified in \( F \).

We now explicitly compute the Dieudonné module for our setting and use [Kis10, §§1.4-1.5] to compute the Frobenius on the \( F \)-crystal on the versal deformation space.

2.1.3. Assume that \( n = 2 \) and \( p \) is inert in \( F \).

(1) The maximal vertex lattice (with trivial \( \varphi \)-action) is \( \Lambda = \text{Span}_{\mathbb{Z}_p}\{e_1, f_1\} \oplus \mathbb{Z} \), where
\[
[Z, e_1] = [Z, f_1] = [e_1, e_1] = [f_1, f_1] = 0, [e_1, f_1] = 1/p, Z \cong \mathbb{Z}_p^2, Q(x) = xx'/p, \forall x \in Z.
\]
Hence \( \Lambda^\vee = p\Lambda \). Set \( e_2 = (1 \otimes 1 + (1/\lambda) \otimes \lambda)/2, f_2 = (1 \otimes 1 + (-1/\lambda) \otimes \lambda)/2 \in \mathbb{Z}_p^2 \otimes_{\mathbb{Z}_p} \mathbb{Z} \). Then
\[
\varphi(e_2) = f_2, \varphi(f_2) = e_2, [e_2, e_2] = [f_2, f_2] = 0, [e_2, f_2] = 1/p.
\]

(2) There are two families of Lagrangians in \( k \)-quadratic space spanned by \( e_1, e_2, f_1, f_2 \) with quadratic form \( pq \) satisfying the dimension restriction. One is \( \text{Span}\{e_1, e_2\} \) under the action of \( e_1 \mapsto e_1 + cf_2, e_2 \mapsto e_2 - cf_1 \) with \( c \in W^\times \cup \{0\} \), completed with \( \text{Span}\{f_1, f_2\} \) under similar action \( f_1 \mapsto f_1 + ce_2, f_2 \mapsto f_2 - ce_1 \). The second family is \( \text{Span}\{e_1 + ce_2, cf_1 - f_2\} \), completed with \( \text{Span}\{ce_1 + e_2, f_1 - cf_2\} \).

In the first case, the special lattice \( L \) is \( \text{Span}_W\{e_1 + cf_2, e_2 - cf_1, pf_1, pf_2\} \). The second case, \( L = \text{Span}_W\{e_1 + ce_2, cf_1 - f_2, pe_2, pf_1\} \).

(3) By [HPT17, Prop. 5.2.2], since the type \( t \) of superspecial point is 2, the above lattice \( L \) is superspecial if and only if \( L + \varphi(L) \) is \( \varphi \)-invariant. In both families, this is equivalent to that \( c \in \mathbb{Z}_p \), i.e., \( \sigma^2(c) = c \). For any special lattice \( L \) in (2), the lattice \( L^\# \) is the \( W \)-Dieudonné.
module of special endomorphisms. Note that \( L \mapsto L^\# \) just switches the two families with \( c \) replaced by \( \sigma(c) \), hence from now on, we work with \( L^\# \) by using the same notation as \( L \).

Now we treat the first family, and work with the supergeneric \( \sigma \), where

\[
x_1 = e_2 - cf_1, x_2 = \sigma^{-1}(c)e_2 - \sigma^{-1}(c)cf_1 + e_1 + cf_2, x_3 = pf_2 - p\sigma^{-1}(c)f_1, x_4 = pf_1.
\]

such that \( Q(x_i) = 0, [x_i, x_j] = 1 \) only when \( \{i, j\} = \{1, 3\}, \{2, 4\} \) and otherwise \( [x_i, x_j] = 0 \).

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

Following \cite{Kis10} §§1.4-1.5, let \( \text{Frob} \) denote the Frobenius on the \( F \)-crystal over the versal deformation space; we can choose the Hodge cocharacter to be \( \text{Frob} \) is split in the \( w \)-basis given by

\[
\begin{pmatrix} 1 & x & -xy & y \\ 0 & 1 & -y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -x & 1 \end{pmatrix}
\]

and \( \text{Frob} = b\sigma, \) with \( b = \begin{pmatrix} -xy/p - ay/p & a + x & p & y \\ -y/p & 1 & 0 & 0 \\ 1/p & 0 & 0 & y \\ -y/p - a/p & 0 & 0 & 1 \end{pmatrix} \),

where \( a = \sigma(c) - \sigma^{-1}(c) \) and \( a = 0 \) if and only if \( L \) corresponds to a superspecial point. Later, by a supergeneric point, we mean a non-superspecial supersingular point.

For the second family, we consider the \( W \)-basis given by

\[
x_1 = f_2 - cf_1, x_2 = e_1 + ce_2 - \sigma^{-1}(c)cf_1 + \sigma^{-1}(c)f_2, x_3 = pe_2 - p\sigma^{-1}(c)f_1, x_4 = pf_1.
\]

The rest of the computation yields the same result.

2.1.4. Assume that \( n = 2 \) and \( p \) is split in \( F \).

There is a basis \( \{v_1, v_2, v_3, v_4\} \) of \( V \) such that the Gram matrix of the bilinear form is

\[
\begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
\]

and a \( \varphi \)-invariant basis of \( V \) is \( \{w_1 = pv_1 + v_2, w_2 = \lambda pv_1 - \lambda v_2, w_3 = v_3 + v_4, w_4 = \lambda v_3 - \lambda v_4\} \).

The bilinear form restricted to the \( \mathbb{Q}_p \)-span of the \( w_i \) has no isotropic vectors. Then by the same argument, the Frobenius on the \( F \)-crystal associated to \( V \) in the above basis has the form

\[
\begin{bmatrix} -xy/p & p & x & -y \\ 1/p & 0 & 0 & 0 \\ x/p & 0 & 0 & 1 \\ -y/p & 0 & 1 & 0 \end{bmatrix}
\]

The Frobenius on the \( F \)-crystal with respect to the the \( w_i \) is:

\[
\begin{bmatrix} 1 - \frac{xy}{2p} & \frac{\lambda xy}{2p} & \frac{x-y}{2p} & -\frac{\lambda (x+y)}{2p} \\ -\frac{xy}{2p} & 1 + \frac{xy}{2p} & \frac{x-y}{2p} & \frac{-(x+y)}{2p} \\ \frac{x-y}{2p} & \frac{\lambda (x-y)}{2p} & 1 & 0 \\ \frac{x+y}{2p} & -\frac{x-y}{2p} & 0 & 1 \end{bmatrix}
\]
2.2. **Lattices for Hilbert modular surface.** In this subsection, based on the local information in §2.1 we construct \( \mathbb{Z} \)-lattices \( L' \) with a quadratic form for supersingular points and compare these lattices with the \( \mathbb{Z} \)-lattice \( L \) used in Borcherds theory in §3.

2.2.1. As in [BBGK07, §2.3], [HY12 §2], [KR99], the pair \((V,Q)\) attached to the Hilbert modular surface is is given by \( V = F \oplus \mathbb{Q}^2 \) with \( Q = ab - \gamma \gamma' \) for \( \gamma, \gamma' \in F, a, b \in \mathbb{Q} \). As in [HMP], we take \( L \) to be a maximal lattice of a given quadratic space over \( \mathbb{Q} \). When the discriminant of the totally real field \( F \) is a prime, see [BB03 §4].

By definition of \( L \), we have the following.

**Lemma 2.2.2.** \( \det L = \text{disc } F. \)

Now we consider the lattice \( L'' \) of special endomorphisms attached to a superspecial point equipped with the natural quadratic form \( Q \) given by \( s \circ s = Q(s) \cdot \text{Id} \). Note, it may be an abuse of notation since we will not show that the lattices (with quadratic form) attached to any superspecial points are isomorphic. However, this problem will be solved by working with a lattice \( L' \) containing \( L'' \) and we will show that for every \( \ell \), the isomorphism class of \((L' \otimes \mathbb{Z}_\ell, Q')\) is independent of the choice of a superspecial point. Moreover, in [HY12], it is shown that the isomorphism class of \((L'' \otimes \mathbb{Q}, Q)\) is independent of the choice of a supersingular point.

**Lemma 2.2.3.** We have \((\text{disc } F)p^2 \mid \det L''. \)

**Proof.** Note that \( L'' \otimes \mathbb{Q}_\ell \cong L \otimes \mathbb{Q}_\ell \) for \( \ell \neq p \). Since \( L \) has discriminant \( \text{disc } F \) and \( L \) is maximal, so \( \text{disc } F \mid \det L' \). On the other hand, the behavior of \( L'' \otimes \mathbb{Z}_p \) is studied in [HP17] and by the classification in the proof of [HP17, Prop. 5.1.2], we have \( p^2 = |(L''/L') \otimes \mathbb{Z}_p| \). \( \square \)

2.2.4. For the rest of the text, we will consider a lattice \( L' \supset L'' \) such that it maximal at all \( \ell \neq p \) (i.e. maximal among all lattices such that \( Q \) evaluated on the lattice has value in \( \mathbb{Z}_\ell \)) and \( L'' \otimes \mathbb{Z}_p = L' \otimes \mathbb{Z}_p \). The same argument as above shows that \((\text{disc } F)p^2 \mid \det L' \). Also, the Fourier coefficients of the theta series of \( L' \) is certainly no less than that of \( L'' \).

**Lemma 2.2.5.** \((L' \otimes \mathbb{Z}_\ell, Q') \cong (L \otimes \mathbb{Z}_\ell, Q)\) for \( \ell \neq p \). Also, \(|\det L'| = (\text{disc } F)p^2\).

**Proof.** Both lattices shall be maximal at \( \ell \) and the quadratic spaces are isomorphic over \( \mathbb{Q}_\ell \). 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. \( \square \)

2.2.6. Similarly, we also define \( L' \) for supegeneric point to be the lattice such that (1) it contains the lattice of special endomorphisms at a supegeneric point; (2) it is maximal at all \( \ell \neq p \), and (3) over \( \mathbb{Z}_p \), it equals to the lattice computed in §2.1.

2.3. **Siegel case via Li–Oort** [LO98]. In this subsection, we explicitly describe the crystal of special endomorphisms as in the case when \( n = 2 \) via the identification of \( \text{GSpin}_{3,2} \) with \( \text{GSp}_4 \).

2.3.1. Let \( G \) denote the derived group of \( \text{GSpin}_{3,2} \). Note that the adjoint group is the split form of \( \text{SO}_5 \). The center of \( \text{GSpin}_{3,2} \) acts trivially on \( V \), and so the action factors through \( \text{SO}_5 \). Further, \( V \) is the unique non-trivial subrepresentation of \( \bigwedge^2 W \), where \( W \) is the standard 4-dimensional representation of \( G \) (where we now treat \( G \) as a symplectic group). Therefore, in order to determine the possible values of \( b \in \text{GSpin}_{3,2}(\mathbb{Q}_p) \) (note that Frobenius is given by \( b\sigma \)), it suffices to explicitly describe the Dieudonné modules that arise from principally polarized supersingular abelian surfaces. In this subsection, we use \( \mathbb{F} \) to denote an algebraically closed field of characteristic \( p \), let \( \mathbb{D} \) denote the (contravariant) Dieudonné module of abelian surfaces, and let \( \varphi \) denote the Frobenius (on the supersingular abelian surfaces or on its Dieudonné module \( \mathbb{D} \)).
2.3.2. By [LO98 §4.1], every principally polarized supersingular abelian surface $A$ over $\mathbb{F}$ can be expressed as a (polarized) quotient of a polarized superspecial abelian surface $A_0$. The kernel of the polarization on $A_0$ is defined to be $A_0[\varphi]$, the kernel of Frobenius, and the kernel of the isogeny $A_0 \to A$ is an order $p$ subgroup of $A_0[\varphi]$ (see [LO98 §3.6]). In general, although there are several (but only finitely many) polarizations on $A$ equivalent on the level of the polarization on $A_0$ as follows.

(1) The Dieudonné module of $A_0$ has a basis $f_1, f_2, f_3, f_4$ with respect to which the matrix of Frobenius $\varphi$ is as below:

$$
\begin{bmatrix}
0 & 0 & p & 0 \\
0 & 0 & 0 & p \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
$$

Let $A_0^*\!$ be the dual abelian surface of $A_0$. By definition, the polarization induces a map $\mathbb{D}(A_0^*) \to \mathbb{D}(A_0)$ whose image is $\varphi\mathbb{D}(A_0)$. Therefore, in the basis $\{f_1, f_2, f_3, f_4\}$, the $W(\mathbb{F})[1/p]$-valued symplectic bilinear form on $\mathbb{D}(A_0)$ induced by the polarization is given by the matrix

$$
\begin{bmatrix}
0 & p^{-1} & 0 & 0 \\
-p^{-1} & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & -1 & 0
\end{bmatrix}.
$$

(2) In order to obtain $A$ as a polarized quotient of $A_0$, we choose an order $p$ subgroup of $A_0[\varphi]$. On the level of Dieudonné modules, this corresponds to choosing a sublattice $\mathbb{D}'$ such that $\varphi\mathbb{D}(A_0) \subseteq \mathbb{D}' \subseteq \mathbb{D}(A_0)$. Therefore, a basis of the module $\mathbb{D}'$ is $\{f'_1, f'_2, f_3, f_4\}$, where either $f'_1 = f_1 + cf_2$ and $f'_2 = pf_1$ (where $c \in W(\mathbb{F})$), or $f'_1 = pf_1$ and $f'_2 = f_2$. Note that $c$ is not unique. Indeed, different choices of $c$ which agree mod $p$ yield the same $\mathbb{D}'$, and therefore the same $A$. Furthermore, if $f'_1 = pf_1$ or if $c = 0$, the abelian surface is then superspecial (and therefore, the polarized Dieudonné modules will be isomorphic). Thus, we may assume from now on that $f'_1 = f_1 + cf_1, f'_2 = pf_2$.

By construction, the symplectic form restricted to $\mathbb{D}'$ is a perfect pairing and the matrix of the pairing with respect to $\{f'_1, f'_2, f_3, f_4\}$ is as below:

$$
\begin{bmatrix}
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & -1 & 0
\end{bmatrix}.
$$

The matrix of $\varphi$ in the basis $\{f'_1, f'_2, f_3, f_4\}$ is

$$
\begin{bmatrix}
0 & 0 & p & 0 \\
0 & 0 & -c & 1 \\
1 & 0 & 0 & 0 \\
\sigma(c) & p & 0 & 0
\end{bmatrix}.
$$

The above discussion classifies all possible $b \in \text{GSp}_4(W(\mathbb{F})[1/p])$ which occur as Frobenii for a principally polarized supersingular abelian surface over $\mathbb{F}$ (up to $\sigma$-conjugacy by an element of $\text{GSp}_4(W(\mathbb{F}))$). Note that $b = z \cdot b'$, where $z$ is a central element of $\text{GSp}_4$, and $b' \in G$. The element
$b' \in G$, view as an action on $W$, has the following matrix:

$$
\begin{bmatrix}
0 & 0 & p^{1/2} & 0 \\
0 & 0 & -cp^{-1/2} & p^{-1/2} \\
p^{-1/2} & 0 & 0 & 0 \\
\sigma(c)p^{-1/2} & 0 & 0 & p^{1/2}
\end{bmatrix}.
$$

Although this is not a $W(\mathbb{F})[1/p]$-rational point of $G$, its image $b_0 \in SO_5$ is $W(\mathbb{F})[1/p]$-rational.

2.3.3. We now consider the representation $V$ of $GSp_{3,2}$ and $SO_5$. For ease of notation, we set $f_3 = f_3, f_4 = f_4$ and as above, we work with the basis $\{f'_1, f'_2, f'_3, f'_4\}$. We realize $V$ as a direct summand of $\wedge^2 W$, with a one-dimensional complement arising from the symplectic form $f'_1 \wedge f'_2 + f'_3 \wedge f'_4$ (here, we view the symplectic form as an element of $\wedge^2 W$). An analysis of the weight spaces of the standard maximal torus of $G$ yields that $f'_1 \wedge f'_2, f'_1 \wedge f'_4, f'_2 \wedge f'_3, f'_2 \wedge f'_4 \subset V$. Finally, the only non-trivial subspaces of $\text{Span}\{f'_1 \wedge f'_2, f'_3 \wedge f'_4\}$ are $\text{Span}\{f'_1 \wedge f'_2 + f'_3 \wedge f'_4\}$ and $\text{Span}\{f'_1 \wedge f'_2 - f'_3 \wedge f'_4\}$. Therefore $V$ is $\text{Span}\{e_1, \ldots, e_5\}$, where $e_1 = f'_1 \wedge f'_2, e_2 = f'_1 \wedge f'_3, e_3 = f'_2 \wedge f'_4, e_4 = f'_2 \wedge f'_3, e_5 = f'_1 \wedge f'_2 - f'_3 \wedge f'_4$. Then, $2e_1 e_3 + 2e_2 e_4 + e_5^2/2$ is the unique quadratic form (up to scaling) whose stabilizer is $SO_5$. By replacing $e_5$ with $\sqrt{2}e_5$, the symmetric bilinear form associated to the quadratic form is

$$
\begin{bmatrix}
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}.
$$

Set $v_1 = e_1 - \sigma^{-1}(c)e_2, v_2 = e_2, v_3 = e_3, v_4 = e_4 + \sigma^{-1}(c)e_3, v_5 = e_5$ and the matrix of the symmetric form remains as above. In this basis, the Frobenius is $b\sigma$, where

$$
b = \begin{bmatrix}
-1 & 0 & 0 & 0 & 0 \\
\sigma(c) - \sigma^{-1}(c) & 0 & 0 & 0 & p \\
0 & \sigma(c) - \sigma^{-1}(c) & -1 & 0 & 0 \\
0 & 0 & -\frac{p}{2} & 0 & 0 \\
0 & 0 & 0 & 0 & -1
\end{bmatrix}.
$$

2.3.4. The Hodge cocharacter $\mu$ can be chosen to be as follows:

$$
\begin{bmatrix}
1 \\
1 \tau \\
1 \\
1
\end{bmatrix}
$$

Kisin in [Kis10, §1.5.4] explicitly describes the filtered $F$-crystal associated to $V$ over the deformation space of the principally polarized surface $A$. The filtration on $V$ is the one induced by the cocharacter $\mu$ (although we will not need the filtration for our purpose). The Frobenius is given by composing a generic unipotent element associated to the opposite unipotent of $\mu$ with (the trivial extension of) $\varphi$ at the supersingular point. With respect to the basis $\{v_1, \ldots, v_5\}$, the opposite unipotent is given by

$$
\begin{bmatrix}
1 & 0 & 0 & -y & 0 \\
x & 1 & y & -xy - z^2/2 & z \\
0 & 0 & 1 & -x & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & -z & 1
\end{bmatrix}.
$$
and the Frobenius $Frob$ on the $F$-crystal is
\[
\begin{bmatrix}
-1 & -y/p & 0 & 0 & 0 \\
-x + a & (ay - z^2/2 - xy)/p & -y & p & -z \\
0 & (a - x)/p & -1 & 0 & 0 \\
0 & 1/p & 0 & 0 & 0 \\
0 & -z/p & 0 & 0 & -1
\end{bmatrix},
\]
where $a = \sigma(c) - \sigma^{-1}(c)$.

The key input to the proof of Theorem 1 is the Frobenius at a superspecial point (i.e., $a = 0$) with respect to a $\varphi$-invariant basis (note that as in §2.1, by $\varphi$-invariant, we mean a basis of $V$ which is invariant under the Frobenius $\varphi$ at the supersingular point) given by the following:
\[
I + \begin{bmatrix}
z^2/2+xy & -z^2/2-xy & -x & -y & -z \\
\lambda(z^2/2+xy) & \lambda(z^2/2-xy) & -\lambda x & -\lambda y & -\lambda z \\
-\lambda y & y/\lambda & 0 & 0 & 0 \\
-x & x/\lambda & 0 & 0 & 0 \\
-z & z/\lambda & 0 & 0 & 0
\end{bmatrix}.
\]

We refer the reader to the appendix for the Frobenius at a supergeneric point with respect to an $F$-invariant basis.

2.4. **Lattices for Siegel case.** We may adopt the same idea as in Section 2.2 (i.e. [HP17] for the $p$-adic lattice and the uniqueness of maximal lattice at all other places) to construct the lattices for global and local contribution in the Siegel case. However, we give an explicit computation based on the fact that the abelian surfaces in question are principally polarized. Since the local information at $p$ is given by §2.3, we focus on comparing the $\mathbb{Z}_\ell$-lattices.

2.4.1. Let $L'$ denote the quadratic lattice of special endomorphisms of a principally polarized superspecial abelian surface $A$. By [Eke87] Prop. 5.2, for any $\ell$, there is a unique class (up to $\text{GL}_4(\mathbb{Z}_\ell)$-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 $x_0^2 + Nm$, where $Nm$ is the quadratic form given by the reduced norm on the quaternion algebra $\text{End}(E)$.

Note that the difference of quadratic lattices of special endomorphisms at superspecial and supergeneric points is only at $p$. The above discussion also recovers the $\mathbb{Z}_\ell ll$-lattices for supergeneric points.

2.4.2. The lattice $L$ for global contribution is a maximal lattice of $V$ (for our application, different choices of $L$ lead to the same Eisenstein series in §3). To obtain $L$, we replace $\text{End}(E)$ in 2.4.1 by the split quaternion algebra $M_2$ since $L$ can be computed as the $\mathbb{Z}$-Betti cohomology of a point on the Shimura variety (see for instance [AGHMP18]). Therefore, the quadratic form on $L$ is $x_0^2 + x_1 x_4 - x_2 x_3$, where $\{x_0, \ldots, x_4\}$ is a basis of $L$. The discriminant of $L$ is 2.

2.5. **Equation of non-ordinary locus.** We now use the computation in §§2.1 2.3 to obtain the local equation of the non-ordinary locus in a formal neighborhood of a supersingular point. The results here are well known, see for instance [Ogu01] Prop. 3, Prop. 11, but we record the results here for completeness in the coordinates that we will use later.

**Lemma 2.5.1.** The non-ordinary locus is given by the equation $pFrob|_{\text{gr}, -1} = 0$ over $\bar{F}_p$.

**Proof.** For all GSpin Shimura varieties attached to $(V, Q)$, a point is ordinary if and only if the Newton polygon on the $F$-crystal $V$ at this point has a slope $-1$. Indeed, the latter statement
implies that the Newton polygon is the same as the Hodge polygon. Then the assertion follows from the weak admissibility of the Newton polygon by considering the trace of \( \text{Frob} \) on \( V \).

\( \square \)

**Corollary 2.5.2.** When \( n = 2 \), 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 \( n = 3 \), the local equation is \( xy + z^2/2 = 0 \) if \( P \) is a superspecial point and \( (x - a)y + z^2/2 = 0 \) if \( P \) supergeneric, where \( 0 \neq a \in \mathcal{W}(\overline{\mathbb{F}}_p) \) depends on \( P \).

### 3. Borcherds Theory and Eisenstein series

We use Borcherds theory to control the global intersection number of a curve \( C \) with special divisors. More precisely, we use the work of Bruinier and Kuss in \( \textbf{BK03} \) 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 \( \textbf{BK03} \) to the Eisenstein part of the theta series attached to supersingular points and reduce the question to a computation of local densities and determinants of the lattices \( L \) and \( L' \) introduced in \( \S \textbf{2.2} \). We use Hanke’s method in \( \textbf{Han04} \) to compute the local densities.

In the proofs in \( \S \textbf{2.7,8} \) we only work with Heegner divisors \( Z(m) \) with \( p \nmid m \), for which the computation of local density is relatively easier. Nevertheless, such computations of local density can be carried out for general \( m \) and we refer the interested reader to Appendix A.

#### 3.1. Borcherds theory.

We start from the setting of integral models of orthogonal Shimura varieties. Let \( L \) be an even lattice of signature \((n,2)\) and we use \( Q \) to denote the quadratic form. For simplicity, we assume that \( L \) is self-dual at \( p \). We always assume that \( p \) is an odd prime. Let \( \mu \in L' / L \) and \( m \in \mathbb{Q}_{>0} \); we consider the special divisor (which is defined over \( \mathbb{Z} \)) \( Z(m,\mu) \). We define \( Z(0,\mu) = 0 \) for \( \mu \neq 0 \) and set \( Z(0,0) = \omega^{-1} \). Here, \( \omega \) is the Hodge line bundle, i.e. the line bundle of weight one modular forms. By Howard–Madapusi-Pera, the generating series \( \sum_{m \geq 0, \mu \in L'/L} Z(m,\mu)q^m \) is a vector-valued (in the Chow group of the integral model of Shimura variety over \( \mathbb{Z} \)) modular form of \( \text{Mp}_2(\mathbb{Z}) \) with respect to the Weil representation \( \rho_L \) on \( \mathbb{C}(L'/L) \).

We may restrict to the mod \( p \) special fiber and still obtain the same modular result. We remark that the same conclusion follows by using \( \textbf{Han04} \) Lemma 5.12, the definition of integral divisors and the characteristic 0 Borcherds theory.

We consider the (vector-valued) Eisenstein series \( E_0(\tau) \) (notation as in \( \textbf{BK01} \); see \( \S 4 \), eqn(8)). More precisely, \( E_0(\tau) \) has constant term 1 at 0 in \( L'/L \), and 0 at the non-zero \( \mu \in L'/L \). This Eisenstein series is a vector-valued modular form with respect to \( \rho_L \). Note that when \( n = 2 \), the definition in \( \textbf{BK01} \) is not convergent, so we need to use analytic continuation, see for instance \( \textbf{BK03} \) Prop. 3.1] to replace \( \textbf{BK01} \) Prop 2. As discussed in \( \textbf{BK03} \), if the Weil representation does not have a copy of the trivial representation as a subquotient, then the weight 2 Eisenstein series is also holomorphic.

Note that in \( \textbf{HMP} \), we work with signature \((n,2)\) and in \( \textbf{BK01} \) and etc., they work with \((2,n)\). The difference is just to work with the negative of the quadratic form. See for instance \( \textbf{Bru17} \) for such discussion (there, Bruinier works with \((n,2)\) and translated all results into this form).

Let \( q(m,\mu) \) be the Fourier coefficient of \( E_0(\tau) \) (with \( m \in \mathbb{Z} - \mathcal{Q}(\mu) \)). Note that by \( \textbf{BK01} \) Prop. 14], for \( n > 0 \), we have \( q(m,\mu) \leq 0 \) and it is non-zero when \( Z(m,\mu) \neq \emptyset \) (this claim will be clear once we see the explicit formula in the next section). Let \( C \) be a projective curve in the mod \( p \) fiber of the Shimura variety.

**Lemma 3.1.1.** The intersection number \( Z(m,\mu).C \sim -q(m,\mu)(\omega.C) \).

\footnote{In classical Borcherds theory, people 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 \((L,Q)\).}

\footnote{We use the convention in \( \textbf{Bru17} \).}
3.2. General theory on Fourier coefficients. For the purpose of this paper, we only work with the component of $E_0(\tau)$ at $0 \in L^\vee/L$ and in this case, $m \in \mathbb{Z}$. We remark that similar formulæ hold for other choices of $\mu \in L^\vee$.

Given a lattice $L$, we write $\det(L)$ for the determinant of its Gram matrix. We have $|L^\vee/L| = |\det(L)|$. Given a number $D$, we use $\chi_D$ to denote the Dirichlet character $\chi_D(a) = (D/a)$ (the Kronecker symbol). We set $\sigma_s(m, \chi) = \sum d | m \chi(d)d^s$.

We use $\delta(\ell, L, m)$ to denote the local density of $L$ representing $m$ over $\mathbb{Z}_\ell$. More precisely, $\delta(\ell, L, m) = \lim_{a \to \infty} \varepsilon(a)\ell^{\#\{ r \in L/\ell L \mid Q(r) \equiv m \mod \ell^a\}$ [BK01, Lem. 5] asserts that the limit is stable once $a \geq 1 + 2v_\ell(2m)$. Iwaniec [Iwa97, p. 196] gives another bound on $\delta$ which only depends on $\det L$. It follows that the product of local densities only depends on congruence classes of $m \mod 64(\det L)^2$.

**Theorem 3.2.1** (BK01 Thm 11 and Iwa97 Thm 11.2, §11.5; for complete formula, see [Bru17 Thm. 2.3]).

In the following, we use $L$ to denote the global lattice (and $q_L(m)$ to denote the Fourier coefficient $q(m,0)$) and $L'$ to denote the positive definite lattice from supersingular points defined in Section 2.2 (and $q_{L'}(m)$ to denote the Fourier coefficient of the Eisenstein series associated to the theta-series of $L'$).

1. For $n = 2$, the Fourier coefficient $q_L(m)$ is

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

2. In the signature $(4,0)$ case, the Fourier coefficient $q_{L'}(m)$ of $q^m$ is

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

3. For $n = 3$, write $m = m_0f^2$, 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

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

where $\mu$ is the Mobius function and $D = -2m_0\det L$. (Note that $2 \mid \det L$ since $\text{rk} L$ is odd.)

4. In the signature $(5,0)$ case, the analogous formula holds without a negative sign, after replacing $L$ with $L'$. (Note that the definition of $m_0$ and $f$ may change accordingly).

**Proof.** For the $n = 3$ case, this is in BK01. For $n = 2$, by using Hecke’s trick (analytic continuation of Eisenstein series) and BK03 Prop. 3.1 we obtain BK01 Prop. 2. This yields BK01 Prop. 3, as 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. (Also compare BK01 Thm. 11 and Iwa97 eqn (11.71)−(11.74)).

3.3. Comparing Fourier coefficients: $n = 2$. The following lemma can be viewed as a refinement of the statement $\sum_{m=1}^M q_L(m) = O(M^2)$. In what follows, the error term here may not be sharp.

---

6In the terminology in Han13, we work with the Hessian matrix. More precisely, $2Q(x) = x^t Sx$, where $S$ is the matrix whose determinant we take.
Lemma 3.3.1. Fix $a, D \in \mathbb{N}$ and $\chi$ a character such that $\chi(\ell) = 0$ for all $1 < \ell \mid D$. (This assumption always holds for our application.) Then

$$\sum_{m \leq X, m \equiv a \mod D} m \sigma_{-1}(m, \chi) = \frac{L(2, \chi)}{2D} X^{2} + O(X^{3/2}).$$

Proof. We use the standard hyperbola method used to sum the convolution of two functions.

$$LHS = \sum_{m \leq X, m \equiv a \mod D} \sum_{d \mid m} d \cdot \chi(m/d) = \sum_{d \leq X, f \leq X, df \equiv a \mod D} d \cdot \chi(f) = \sum_{d \leq X^{1/2}} d \cdot \chi(f) + \sum_{f \leq X^{1/2}} \chi(f) \sum_{d \leq X/f, df \equiv a \mod D} d - (\sum_{d \leq X^{1/2}} d)(\sum_{f \leq X^{1/2}} \chi(f))$$

Note that the absolute value of the first term is bounded above by $\sum_{d \leq X^{1/2}} d \cdot (X/d) = X^{3/2}$ and the absolute value of the third term is bounded above by $(\sum_{d \leq X^{1/2}} d) \cdot (\sum_{d \leq X^{1/2}} 1) \leq X^{3/2}$. The second term is the main term. We may assume that $f$ is always invertible mod $D$ (otherwise $\chi(f) = 0$). Therefore, $\sum_{d \leq X/f, df \equiv a \mod D} d = \frac{X^{2}}{2Df^{2}} + O(X)$. Hence

$$LHS = \frac{X^{2}}{2D} \sum_{f \leq X^{1/2}} \chi(f)/f^{2} + O(X^{3/2}) = \frac{L(2, \chi)}{2D} X^{2} + O(X^{3/2}).$$

□

Now we compare $q_{L}(m)$ and $q_{L'}(m)$ in the following lemma, which is a direct consequence of Theorem 3.2.1.

Lemma 3.3.2. Suppose $L'$ is the lattice corresponding to a superspecial point. For $m \not\equiv 0 \mod p$,

$$\frac{q(m)_{L'}}{-q(m)_{L}} = \frac{\delta(p, L', m)}{p(1 - \chi_{4_{\det}L}(p)p^{-2})}.$$

If $L'$ is corresponds to a supergeneric point the same formula holds except with an extra factor of $p$ in the denominator.

We compute this constant in two cases. We follow [HP17, §6.1] and note that $\Lambda^{\vee}$ there is the lattice of special endomorphism over $\mathbb{Z}_{p}$ (more precisely, from the discussion in [HP17, §6.1.1], the lattice of special endomorphism contains $\Lambda^{\vee}$ and by [HP17, Prop. 5.2.2], we see that these two $\mathbb{Z}_{p}$-lattices coincide).

1. $p$ is inert in $F$. Then $\chi_{4_{\det}L}(p) = \chi_{\det}L(p) = -1$. At a superspecial point, the $\mathbb{Z}_{p}$-quadratic form is $xy + p(z^{2} - Dw^{2})$ and to compute $\delta$, we only need to work mod $p$. In this case, the density is $(1 - 1/p)$. Hence the ratio is $(p - 1)/(p^{2} + 1)$.

At a supergeneric point, the $\mathbb{Z}_{p}$-quadratic form is $p(xy + z^{2} - Dw^{2})$, and hence the density is zero when $p \nmid m$.

2. $p$ is split in $F$. Then $\chi_{4_{\det}L}(p) = \chi_{\det}L(p) = 1$. The quadratic form over $\mathbb{Z}_{p}$ is $x^{2} - uy^{2} - pz^{2} + upw^{2}$, where $u \in \mathbb{Z}_{p}^{\times}$ is not a square. We again need to work mod $p$ to compute $\delta$, and it equals $(1 + 1/p)$. Hence the ratio is $1/(p - 1)$. 


We first give a rough estimate on the asymptotic of \( q_L(m) \). Assume \( m_0 \) is fixed (this is the case in our later application). Modulo constants, the main varying term is\(^7\)
\[
f^3 \sum_{d | f} \mu(d) \chi_D(d) d^{-2} \sigma_3(f/d) < f^3 \sum_{d | f} d^{-2} \zeta(3) < f^3 \zeta(2) \zeta(3).
\]

and hence the sum of all such \( q(m, \nu) \) for \( m = r^2 < X \) is \( CX^2 + o(X^2) \), where \( C \) is a constant.

Now we compare the local and global Eisenstein series similar to the \( n = 2 \) case. We first treat the superspecial points.

**Lemma 3.4.1.** Assume that \( p^2 \nmid m \). Then
\[
\frac{q(m)_L'}{-q(m)_L} = \frac{\delta(p, L', m)(1 - \chi_D(p)p^{-2})}{p(1 - p^{-4})}.
\]
(Note that in this case, \( m_0, f \) remain the same and \( p \nmid f \) and hence for all \( d \mid f \), we have \( \chi_D(d) = \chi_D'(d) \). Moreover, for the principally polarized Siegel case, \( L = 2 \) and hence \( D = -4m_0 \).

Now we compute this ratio case by case and the local density computation still follows from [Han04 §3].

1. \( p \| m \). In this case, \( p \mid m_0 \) and hence \( \chi_D(p) = 0 \); the local density is given by \( \delta = 1 + p^{-2}(p+1) \). Therefore the ratio is \( (1 + p^{-1} + p^{-2})/(p - p^{-3}) \), which is less than \( (p-1)^{-1} \).
2. \( p \nmid m \). A naive estimate (i.e. without computing whether \( \chi(p) \) is 1 or −1) indicates that \( \delta \leq 1 + p^{-1} \) and the ratio is \( \leq (p-1)^{-1} \).

**3.4.2.** Now we treat supergeneric points. The difference is that now the total ratio is further divided by \( p \) and we need to compute \( \delta \) in this case. Note that the quadratic form is \( (u)x^2 + pQ' \), where \( Q' \) is a four variable quadratic form with discriminant prime to \( p \).

1. \( p \nmid m \). Then \( \delta \) is either 0 or 2 and we have
\[
\frac{q(m)_L'}{-q(m)_L} \leq \frac{2(1 + p^{-2})}{p^2(1 - p^{-4})} = \frac{2}{p^2 - 1}.
\]
2. \( v_p(m) = 1 \). Then \( \delta = 1 + \delta_{bad} \leq 1 + p^{-2} \) and we have
\[
\frac{q(m)_L'}{-q(m)_L} \leq \frac{(2 + p^{-2})(1 + p^{-2})}{p^2(1 - p^{-4})} = \frac{2 + p^{-2}}{p^2 - 1}.
\]

In case (2) above, we use Hanke’s notion of lattices points of good/zero/bad type (see [Han04 Def. 3.1]) and we refer the reader to [Han04 §3] for the algorithm of computing local densities inductively via these notions.

**3.5. Smaller lattices.** We will work with sublattices lattices of \( L' \) with larger power of \( p \) in discriminant coming from Theorem 5.1.3.

**Lemma 3.5.1.** Let \( L'' \subset L' \) be a sub-lattice such that \( L'' \otimes \mathbb{Z}_\ell = L' \otimes \mathbb{Z}_\ell \) for all \( \ell \neq p \). Then for any \( p \nmid m \), the local density \( \delta(m) \) of \( L'' \) is at most 2.

**Proof.** Since \( p \nmid m \), the density \( \delta(m) \) only contains the good density and hence the density equals the density of the mod \( p \) quadratic form on \( L''/pL'' \). Since \( p \mid \text{disc } L' \), when we write the quadratic form on \( L'' \) in diagonal form, at least one coefficient is divisible by \( p \) and hence the mod \( p \) quadratic form is at most of rank 4. On the other hand, the mod \( p \) quadratic form has rank at least 1 (otherwise, the local density representing \( m \) will be 0). Then we check [Han04 Table 1], to see that all the local densities are bounded by 2. \(\square\)

\(^7\)Here we give an upper bound. Lower bounds are given in [Bru17 proof of Prop. 2.5] via similar argument.
4. Decay lemma for supersingular points 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 with real multiplication over $\mathbb{F}_p[[t]]$, whose special fiber is a supersingular abelian surface. 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 (this is exactly what we need in order to bound the local intersection multiplicity $\text{Spf } \mathbb{F}_p[[t]] \cdot Z(m)$ – see Lemma 6.2.1). The precise definitions and results are in Definition 4.1.1 and Theorem 4.1.2.

Throughout this section, $k$ denotes $\overline{\mathbb{F}}_p$ (or more generally, any subfield of $\overline{\mathbb{F}}_p$ over which $C$ is defined) and $W$ denotes $W(k)$. Let $\sigma$ be a lift of Frobenius on $W(k)[1/p]$ and we will also use $\sigma$ to denote the extension of Frobenius on $W(k)[1/p][[t]]$ which sends $t$ to $t^p$. For a matrix $M$ with entries in $W(k)[1/p][[t]]$, we use $M^{(n)}$ to denote $\sigma^n(M)$. Further, we define the Hilbert modular surface to be “inert” or “split” if the associated real quadratic field is inert or split at $p$ respectively (note that we avoid the case when the real quadratic field is ramified). Finally, let $\lambda$ be as in §2.1.2 (i.e. $\sigma(\lambda) = -\lambda$).

4.1. Statement of the Decay Lemma. In this section, we focus on the behavior of $C$ in a formal neighborhood of a supersingular point, and so let $C = \text{Spf } k[[t]]$ denote a generically ordinary formal curve in the Hilbert modular surface which specializes to a supersingular point. This is equivalent to a local ring homomorphism from $k[[x, y]] \to k[[t]]$, and we denote by $x(t)$ and $y(t)$ the images of $x$ and $y$ 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 2.5.2.

Definition 4.1.1. Let $w$ denote a special endomorphism. We say that $w$ decays rapidly if $p^aw$ does not lift to an endomorphism modulo $t^{A(1+p+\cdots+p^n)}$ for all $n$. We say that $w$ decays very rapidly if $p^aw$ does not lift to an endomorphism modulo $t^{A(1+p+\cdots+p^{n-1})+ap^n}$ for some $a \leq A/2$, for all $n$. We say that a submodule decays rapidly if every primitive vector in the submodule which is not a multiple of $p$ decays rapidly.

In particular, if a vector decays very rapidly, then it decays rapidly.

The main theorem of this section is the decay lemma:

Theorem 4.1.2 (Decay Lemma in the Hilbert case). There exists a rank 3 submodule of special endomorphisms which decays rapidly. If the point is superspecial, there is a primitive vector in this rank 3 submodule which decays very rapidly.

4.2. Decay in the split case. Every supersingular point is superspecial in this case. As the non-ordinary locus is cut out by the equation $xy = 0$, both $x$ and $y$ map to non-zero power series in $t$ in the map $k[[x, y]] \to k[[t]]$. Without loss of generality, we assume that $v_t(x) \leq v_t(y)$, and that $x(t) = t^a + \cdots$ and $y(t) = \alpha t^b + \cdots$, where $\alpha \in k$.

4.2.1. Recall from §2.1 the Frobenius on the crystal, with respect to a $\varphi$-invariant basis $\{w_1, \cdots, w_4\}$ of $V$, is given by $I + F$, where $F$ denote the matrix

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


We define $F_\infty$ to be the infinite product $\prod_{i=0}^{\infty} (1 + F^{(i)})$. Since $v_t(y) \geq v_t(x) \geq 1$, the entries of $F_\infty$ consist of power series valued in $W[1/p][[t]]$. The $\mathbb{Z}_p$-span of the columns of $F_\infty$ are vectors of the isocrystal of special endomorphisms which are Frobenius stable (see for instance [Kis10] §1.4.1.5); moreover, these vectors are also horizontal with respect to the connection on the isocrystal. Let $w$ be any vector in this span; the coordinates of $w$ consist of power series with entries in $W[1/p][[t]]$. Let $m$ be some integer, and consider $w \mod t^m$. Then $p^n w$ extends to a special endomorphism of the abelian surface modulo $t^m$, but $p^n w$ does not. Therefore, in order to prove the Decay Lemma, we need to carefully analyze $F_\infty$.

Let $F_\infty(1)$ and $F_\infty(2)$ denote the top left and top right $2 \times 2$ blocks of $F_\infty$ respectively. To simplify the notation, define

$$G = \begin{bmatrix} -\frac{1}{7} & \frac{\lambda}{7} \\ \frac{7}{2} & \frac{7}{2} \end{bmatrix}, H_u = \begin{bmatrix} \frac{1}{7} & -\frac{\lambda}{7} \\ \frac{7}{2} & -\frac{1}{2} \end{bmatrix}, H_l = \begin{bmatrix} 1 & -\lambda \\ \frac{1}{2} & 1 \end{bmatrix},$$

and let $F_t$, $F_u$ and $F_l$ denote the top-left, top-right, and bottom-left $2 \times 2$ 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 4.2.2.**  
(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 $X_0$ is either $F_t$, $F_u$ or $F_l$\footnote{Here, to view $x(i), y(i)$ as power series in $W[[t]]$, we pick a lift $k \to W$, for instance, the Teichmüller lift.} $m_1 + 1$ 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 $n_0$ 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 $n_i = 0$, and either when $X_0 = X_1 = \ldots X_{m_1-1} = F_t$, or $X_0 = X_1 = \ldots X_{m_2-1} = F_u$. 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\footnote{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.}) equals

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

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

$$\frac{1}{p^{n+2}} \prod_{i=0}^{m_1} G^{(i)}(x) H_u^{(m_1+2i+1)} H_l^{(m_1+2i+2)} (x^1+p + y^1+p)^{(m_1+2i+1)} . F_u^{(m_1+2m_2+1)}.$$
4.2.3. 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 that there exists a rank-3 submodule which decays rapidly, 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 4.2.4. Let \( n, e, f \) be in \( \mathbb{Z}_{\geq 0} \).

1. The kernel of the \( 2 \times 2 \) 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 of each other if \( e \neq f \) 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 \( GG^{(1)}G = G, H_u H_l^{(1)} H_u H_l^{(1)} = 2H_u H_l^{(1)} \), and \( H_u^{(1)} H_l H_u^{(1)} H_l = 2H_u^{(1)} H_l \). Therefore, if \( n - e \) is odd, then \( P(1)_{e,n} \) simplifies to either \( GG^{(1)} H_u H_l^{(1)} \), \( GG^{(1)} H_u H_l^{(1)} \), 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 \( GG^{(1)} H_u H_l^{(1)} \) are scalar multiples of
\[
\begin{bmatrix}
\frac{1}{p} & \frac{1}{p} \\
\frac{1}{p} & \frac{1}{p}
\end{bmatrix}
\]
(resp. \( G \) and \( GH_u^{(1)} H_l \) are scalar multiples of
\[
\begin{bmatrix}
\frac{1}{p} & \frac{1}{p} \\
\frac{1}{p} & \frac{1}{p}
\end{bmatrix}
\].

In either case, since \( \lambda \in W(\mathbb{F}_{p^2}) \setminus \mathbb{F}_p \), 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.

Lemma 4.2.5. Let \( n, e, f \) be in \( \mathbb{Z}_{\geq 0} \).

1. The kernel of the \( 2 \times 2 \) 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 \( e \neq f \) 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 4.2.4. Indeed, if \( n - e \) is odd (resp. even), we are reduced to the cases of \( GG^{(1)} H_u H_l^{(1)} H_u \), \( GG^{(1)} H_u \), \( H_u^{(1)} H_l H_u \), and \( H_u \) (resp. \( GH_u^{(1)} H_l H_u \) and \( GH_u^{(1)} \)). The rest of the argument is similar.

We now prove the Decay Lemma, Theorem 4.1.2 when \( p \) is inert 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 4.2.2.

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 $\text{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_i\}$ decays rapidly, where $w_i = w_4$ if the $t$-adic valuation of $x - y$ is $> a$, and $w_i = w_3$ otherwise. Moreover, $w_i$ decays very rapidly.

**Lemma 4.2.6.**  
(1) Among the terms appearing in $F(x)(1)$ described in Lemma 4.2.2 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 $F(x)(2)$ described in Lemma 4.2.2 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 4.2.2 and the assumption that $a = b$.

**Proof of the Decay Lemma in this case.** We first prove that every primitive vector $w \in \text{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$ decays rapidly. Indeed, write $w = cw_1 + dw_2$, by Lemma 4.3.1 and Lemma 4.2.6, there is a unique (non-vanishing) term in $F(x)(1)w$ with denominator $1/p^{n+1}$ and minimal $t$-adic valuation $A(1 + p + \ldots + p^n)$ given by $P(1)_{0,n} c dT(xy)^{1+p+\ldots+p^{n}}$. Hence, modulo $tA(1+p+\ldots+p^n)$, the horizontal section $F(x)(p^n w)$ 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 equaling $a$. Furthermore, by Lemma 4.2.6, $P(1)_{0,n-1} \cdot v \neq 0 \mod p$, where $v$ is the $n^{th}$ Frobenius twist of either column of $H_w$. Therefore, among the terms in the $i^{th}$ column of $F(x)$ 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 $\text{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_i\}$ 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 $F(x) \cdot p^m w_u$. 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 $F(x) \cdot w_i$. 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 $p^m w_u + w_i$ decays rapidly. Now, consider a vector of the form $w_u + p^m w_i$, where $m \geq 0$. Analogous to the previous case, consider the two-dimensional vector whose entries are the first two entries of $F(x) \cdot w_i$. The $t$-adic valuation of the sum of all terms with denominator $p^{n+1}$ equals $2a(1+p+\ldots+p^{n-1})$. Similarly, consider the two-dimensional vector whose entries are the first two entries of $F(x) \cdot p^m w_i$. The $t$-adic valuation of the coefficient of $1/p^{n+1}$ equals $2a(1+p+\ldots+p^{n+m-1}) + ap^n + p^m$. Regardless of the value of $m$ (recall that $m \geq 0$), the latter quantity is always greater than the former quantity, whence it follows that $p^m w_u + w_i$ decays rapidly.

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

**Lemma 4.2.7.**  
(1) Among the terms appearing in $F(x)(1)$ described in Lemma 4.2.2 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 $F(x)(2)$ described in Lemma 4.2.2 with denominator $p^{n+1}$, there are exactly two terms with minimal $t$-adic valuation, and they are

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

and
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 $p^{n+1}$:

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

Similar to Lemma 4.2.2(2), we observe that among all the terms of $F_\infty(1)$ with denominator exactly $p^{n+1}$ given in Lemma 4.2.2(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 $f(k) := v_t(X_{k,l}) = a \left( (1 + p^{2e}) \frac{p^{k-1}}{p-1} + \frac{p^{n-k+1} - 1}{p-1} \right)$, 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).

Proof of the Decay Lemma in this case. We first prove that $\text{Span}_{\mathbb{Z}_p}(w_1, w_2)$ decays rapidly. Indeed, let $w'$ be a primitive vector in $\text{Span}_{\mathbb{Z}_p}(w_1, w_2)$. Lemma 4.2.4(1) implies that $P(1)_{e,n} \cdot w'$ mod $p$ is non-zero. This fact taken in conjunction with Lemma 4.2.7(1) yields that $w'$ decays rapidly.

Secondly, we prove that there exists a primitive vector $w \in \text{Span}_{\mathbb{Z}_p}(w_3, w_4)$ (independent of $n$) which decays very rapidly. Set $Y_{e,n} := P(1)_{e,n-1} F_u(n+e-1) (xy)^{1+p+\ldots+p^{n-e-1}} x^{p^{n-e}+p^{n-e+1}+\ldots+p^{n+e}} + P(1)_{e+1,n-1} F_u(n+e) (xy)^{1+p+\ldots+p^{n-e-2}} x^{p^{n-e}+p^{n-e+1}+\ldots+p^{n+e+1}}$, which is the sum of the two terms with minimal $t$-adic valuation listed in Lemma 4.2.7(2). The sum $Y_{e,n}$ is non-zero modulo $p$ by Lemma 4.2.4(2). Furthermore, up to Frobenius twists and multiplication by scalars, the matrix $Y_{e,n}$ mod $p$ is independent of $n$. Therefore, there exists a vector $w \in \text{Span}_{\mathbb{Z}_p}(w_3, w_4)$ 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 4.2.7(2).

Finally, a valuation-theoretic argument analogous to Case 1 shows that every primitive vector in $\text{Span}_{\mathbb{Z}_p}(w_1, w_2, w)$ decays rapidly, thereby establishing the Decay Lemma in this case.

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

Lemma 4.2.8. (1) Among the terms appearing in $F_\infty(2)$ described in Lemma 4.2.2 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 4.2.2 with denominator $p^{n+1}$, there are exactly two terms with minimal $t$-adic valuation, and they are

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

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

Proof. The proof of this lemma is identical to that of Lemma 4.2.7(2) so we omit the details.

Proof of the Decay Lemma in this case. Analogous to Case 2, Lemma 4.2.5 and Lemma 4.2.8(2) imply the existence of a primitive $w \in \text{Span}_{\mathbb{Z}_p}(w_1, w_2)$ that decays rapidly; and by Lemma 4.2.5(1)
and Lemma [4.2.8(1), Span\(_{\mathbb{Z}_p}\{w_3, w_4\}\) decays very rapidly. Finally, a valuation-theoretic argument shows that every primitive vector in Span\(_{\mathbb{Z}_p}\{w, w_3, w_4\}\) decays rapidly. 

**Case 4:** \(b \neq ap^e\) for any value of \(e\).

**Proof of the Decay Lemma.** As this is the easiest case, we will be content with merely sketching a proof. Analogous to Lemmas [4.2.7 and 4.2.8] 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 \(\text{Span}_{\mathbb{Z}_p}\{w_1, w_2\}\) decays rapidly and every vector in \(\text{Span}_{\mathbb{Z}_p}\{w_3, w_4\}\) 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 the Decay Lemma.

4.3. Decay in the inert case.

4.3.1. **Superspecial case.** The Frobenius on the crystal with respect to a \(\varphi\)-invariant basis \(\{w_1, w_2, w_3, w_4\}\) is \(I + \begin{pmatrix} A_1 & A_2 \\ A_3 & 0 \end{pmatrix}\), where

\[
A_1 = \frac{xy}{2p} \begin{pmatrix} -1 & \lambda x \\ -1/\lambda & 1 \end{pmatrix}, \quad A_2 = \frac{1}{2p} \begin{pmatrix} x & y \\ x & y \end{pmatrix}, \quad A_3 = \begin{pmatrix} -y & \lambda y \\ -x & \lambda x \end{pmatrix}.
\]

As in the split case, the non-ordinary locus is cut out by the equation \(xy = 0\). Therefore, we may assume that both \(x\) and \(y\) map to non-zero power series in the map \(k[[x, y]] \to k[[d]]\). We will again assume that \(a = v_t(x) \leq b = v_t(y)\).

**Proposition 4.3.1.** The \(\mathbb{Z}_p\)-span of \(w_1, w_2, w_3\) decays rapidly, and the vector \(w_3\) decays very rapidly.

**Proof.** 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.

Similar to Lemma [4.2.2], it is easy to see that the top-left \(2 \times 2\) block of \(F_\infty\) with \(p\)-adic valuation \(-n+1\) has a term of the form \(A_n := A_1A_1(1) \ldots A_1(n)\), and this term is the unique term with minimal \(t\)-adic valuation (equalling \((a+b)(1+p + \ldots p^n))\). Similarly, the top-right \(2 \times 2\) block of \(F_\infty\) with \(p\)-adic valuation \(-(n+1)\) has a term of the form \(B_n := A_1A_1(1) \ldots A_1(n-1)A_2(n)\), and this term is the unique term with minimal \(t\)-adic valuation (equalling \((a+b)(1+p + \ldots p^{n-1}) + ap^n)\).

Arguments identical to Lemma [4.2.4 and 4.2.5] 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 \(A_m\) is different from the \(t\)-adic valuation of \(B_n\) for every pair of integers \(n, m\), it follows that \(\text{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}\) also decays rapidly. The argument is elaborated on in the last paragraph of Case 1 of the decay lemma for split Hilbert modular surfaces.

4.3.2. **Supergeneric case.** At a supergeneric point, set \(w_1 = x_1, w_2 = px_1 + x_3 + (c+\sigma^{-1}(c))x_4, w_3 = c(px_1-x_3+(c-\sigma^{-1}(c))x_4, w_4 = px_2-cx_3-p\sigma^{-1}(c)x_1-\sigma^{-1}(c)x_4,\) then \(L^{e=1} = \text{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3, w_4\}\).

**Proposition 4.3.2.** The \(\mathbb{Z}_p\)-span of \(v_1, v_2, v_3\) decays rapidly.

**Proof.** As before, we write the Frobenius matrix with respect to \(w_i\) as \(I + \frac{y}{p}A + xB\), where

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

Note that every denominator \(p\) must be accompanied by a factor of \(y\). Hence, the leading term arises from a twisted product of \(A\) with itself (if \(B\) is involved, then it must lie in higher power of \(m\).
Hence the leading term is of the form \( y^{1+p+\cdots+p^n-1}/p^n \) and the last row \( R_n \) of the twisted product \( AA^\sigma \cdots A^\sigma_{n-1} \) is \( R_n = \pm (c - \sigma^2(c)) \sigma(R_{n-1}), \) \( n \geq 3 \) and \( R_2 = (\pm 1 - \sigma(c) - c \lambda(c - c) - \sigma(c)) \) (this can easily be proved by induction). The rest of the proof is the same as the proof in §B.2. \( \Box \)

5. Decay lemma at supersingular points in the Siegel case

In this section, we prove the analogue of Theorem 4.1.2 in the Siegel case. We deal only with superspecial points in \( \mathcal{A}_2 \mathbb{F}_p \) (see Theorem 5.1.3 for the precise statement), as the proof of Theorem 1.1 does not require an analysis of supergeneric points. However, we still prove the analogous result at supergeneric points in the appendix, and we refer the interested reader to Appendix B.

Throughout this section, \( k \) denotes \( \mathbb{F}_p \) (or more generally, any subfield of \( \mathbb{F}_p \) over which \( C \) is defined) and \( W \) denotes \( W(k) \). Let \( \sigma \) be a lift of Frobenius on \( W(k)[1/p] \) and we will also use \( \sigma \) to denote the extension of Frobenius on \( W(k)[1/p][[t]] \) which sends \( t \) to \( t^p \). For a matrix \( M \) with entries in \( W(k)[1/p][[t]] \), we use \( M^{(n)} \) to denote \( \sigma^n(M) \). Let \( \lambda \) be as in §4.

5.1. Statement of the Decay Lemma. Let \( C = \text{Spf} \ k[[t]] \) denote a generically ordinary formal curve which specializes to a superspecial point. This is equivalent to a local ring homomorphism \( k[[x, y, z]] \to k[[t]] \), and we denote by \( x(t), y(t) \) and \( z(t) \) the images of \( x, y, z \) respectively. By Corollary 2.5.2 the non-ordinary locus is cut out by the equation \( xy + z^2/2 = 0 \).

5.1.1. Notations. 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 \( \infty \) if the corresponding power series is 0. As before, \( v_t \) denotes the t-adic valuation map on \( W(1/p)[[t]] \) or \( k[[t]] \).

Let \( \eta^A \) and \( \mu^B \) denote the leading terms of \( xy + z^2/2 \) and \( xy^p + x^p y + z^{1+p} \) respectively. In particular, \( A = v_t(xy + z^2/2) \), and \( B = v_t(xy^p + x^p y + z^{1+p}) \).

With the new definition of \( A \) in the Siegel case, the same definition of decay (very) rapidly in §4 works and we recall it here.

Definition 5.1.2. Let \( w \) denote a special endomorphism.

(1) We say that \( w \) decays rapidly if \( p^n w \) does not lift to an endomorphism modulo \( t^{A(1+p+\cdots+p^n)} \) for all \( n \).

(2) We say that a submodule 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 \( t^{A(1+p+\cdots+p^n-1) + a p^n} \) for some constant \( a \leq A/2 \), for all \( n \).

The following is the main result of this section.

Theorem 5.1.3 (Decay lemma in the Siegel case). At a superspecial point, there exists a rank 3 submodule of special endomorphisms which decays rapidly and furthermore, there is a primitive vector in this rank 3 submodule which decays very rapidly.

5.2. Preparation of the proof. The preparation lemmas of the Siegel case are very similar to that of the split Hilbert case in the beginning of §4.2.

5.2.1. Recall from §2.3 the Frobenius on the crystal, with respect to a \( \varphi \)-invariant basis \( \{w_1, \cdots, w_5\} \) of \( V \), is given by \( I + F \), where \( F \) is the matrix

\[
\begin{bmatrix}
\frac{x^2/2 + xy}{2p} & \frac{-x^{2}/2-xy}{2p} & \frac{-x}{2p} & \frac{-y}{2p} & \frac{-z}{2p} \\
\frac{\lambda(x^2/2+xy)}{2p} & \frac{-\lambda x^{2}/2-xy}{2p} & \frac{-\lambda x}{2p} & \frac{-\lambda y}{2p} & \frac{-\lambda z}{2p} \\
-\frac{y}{2p} & \frac{y}{\lambda} & 0 & 0 & 0 \\
-\frac{x}{2p} & \frac{x}{\lambda} & 0 & 0 & 0 \\
-\frac{z}{2p} & \frac{z}{\lambda} & 0 & 0 & 0
\end{bmatrix}
\]
We define $F_{\infty}$ to be the infinite product $\prod_{i=0}^{\infty}(1 + F^{(i)})$. As in the discussion in [4.2.1], the entries of $F_{\infty}$ consist of power series valued in $W[1/p][[t]]$ and the $\mathbb{Z}_p$-span of the columns of $F_{\infty}$ are vectors of the iso-crystal of special endomorphisms which are Frobenius stable. For a vector $w$ in this span, $m \in \mathbb{Z}_{\geq 0}$, let $n$ denote the smallest positive integer such that $p^nw$ has integral entries modulo $t^m$. Then $p^nw$ extends to a special endomorphism of the abelian surface modulo $t^m$, but $p^{n-1}w$ does not. Therefore, in order to prove the Decay Lemma, we need to analyze $F_{\infty}$.

5.2.2. **Notations.** We denote by $F_t$, $F_u$, and $F_l$ the top-left $2 \times 2$ block, the top-right $2 \times 3$ block, and the bottom-left $3 \times 2$ block of $F$ respectively. Define

$$G = \left[ \frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{3}{2} \right], \quad H_u = \left[ \frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{3}{2} \right], \quad \text{and} \quad H_l = \left[ -1, -1, -\frac{1}{2}, -\frac{1}{2} \right].$$

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

The following is analogous to Lemma 4.2.2.

**Lemma 5.2.3.**

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 $F_l$\[11\] 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 $X_0 = X_1 = \ldots X_{m_1-1} = F_t$, 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 $\prod_{i=0}^{m_1} F_t^{(i)} \prod_{i=0}^{m_2-1} F_u^{(m_1+2i+2)} F_l^{(m_1+2i+2)}$ equals

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

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

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

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

The following will play a similar role as Lemma 4.2.4.
Lemma 5.2.5. The kernel of $P(1)_{g,f+g}$ mod $p$ does not contain any non-zero vector defined over $\mathbb{F}_p$. Moreover, if $f$ is odd (resp. even), the kernel of $P(1)_{g,f+g}$ mod $p$ does not contain the vector $\left[\begin{array}{c} 1 \\ \lambda \end{array}\right]$ (resp. $\left[\begin{array}{c} 1 \\ -\lambda \end{array}\right]$).

Proof. We prove the assertions by explicit computation as in Lemmas 4.2.4 and 4.2.5. Note that

$$H_u^{(2m)}H_l^{(2m+1)} = \frac{1}{2} \left[\begin{array}{cc} 1 & \lambda^{-1} \\ \lambda & 1 \end{array}\right], \quad H_u^{(2m+1)}H_l^{(2m)} = \frac{1}{2} \left[\begin{array}{cc} 1 & -\lambda^{-1} \\ -\lambda & 1 \end{array}\right].$$

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 one of these matrices depending on the parity of $m_1$. Similarly, we have

$$G \cdots G^{(2m)} = \frac{1}{2} \left[\begin{array}{cc} 1 & -\lambda^{-1} \\ -\lambda & 1 \end{array}\right], \quad G \cdots G^{(2m+1)} = \frac{1}{2} \left[\begin{array}{cc} 1 & \lambda^{-1} \\ \lambda & 1 \end{array}\right].$$

Therefore, $P(1)_{g,f+g}$ equals $\left[\begin{array}{c} 1 \\ \lambda \end{array}\right]$ if $f$ is odd, and equals $\left[\begin{array}{c} 1 \\ -\lambda \end{array}\right]$ if $f$ is even. The lemma then follows immediately. $\square$

For fixed $n$, among the terms listed in Lemma 5.2.3 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 §§5.3,5.5 to prove the Decay Lemma. The first case, while technically the easiest, holds the main ideas in general.

5.3. Case 1: $A < B$.

Note that if $a + b \neq 2c$, or more generally, if the leading terms of $xy$ and $z^2/2$ do not cancel, then $A < B$.

Proof of the Decay Lemma 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 Decay Lemma 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 $\text{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$ decays rapidly. For a primitive vector $v \in \text{Span}_{\mathbb{Z}_p}\{w_1, w_2, w_3\}$, write $w = \alpha_u w_u + \alpha_v w_v$, where $w_u$ is a primitive vector in $\text{Span}_{\mathbb{Z}_p}\{w_1, w_2\}$, and $\alpha_u, \alpha_v \in \mathbb{Z}_p$. Since $w$ is primitive, then either $\alpha_u$ or $\alpha_v$ is a $p$-adic unit. We may assume that $\alpha_u$ is a unit -- the other case is entirely analogous to this. Suppose that the $p$-adic valuation of $\alpha_v$ is $m \geq 0$.

Consider the terms appearing in $F_\infty(1)$ described in Lemma 5.2.3 with denominator $p^{n+1}$. As $A < B$, the one with minimal $t$-adic valuation is $P(1)_{0,n}(xy + z^2/2)^{1+p+\ldots+p^n}$, 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 $P(1)_{0,n+m-1} \cdot F_u^{(n+m)}(xy + z^2/2)^{1+p+\ldots+p^n+m-1}$.

Let $P$ denote the $2 \times 3$ matrix whose first two columns equal $P(1)_{0,n}(xy + z^2/2)^{1+p+\ldots+p^n}$ (part of $F_\infty(1)$), and whose last column is the first column of $P(1)_{0,n+m-1} \cdot F_u^{(n+m)}(xy + z^2/2)^{1+p+\ldots+p^n+m-1}$ (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 $\leq A(1 + \ldots + p^n)$. This in turn reduces to proving the following statement: if $m \geq 1$, then $w_u$ mod $p$ is not in the kernel of $P(1)_{0,n}$ mod $p$; and if $m = 0$, the vector $\left[\begin{array}{c} 1 \\ \lambda(\alpha) \end{array}\right]$ mod $p$ is not in the kernel of $P(1)_{0,n-1}$ mod $p$. For the ease of exposition, we assume that $a \leq b \leq c$. Note that this forces $2a \leq A$. Even though the Decay Lemma is not symmetric in $a,b,c$, an identical argument as the one below suffices to deal with all the other cases.
p. Both statements follow from Lemma 5.2.5 establishing the decay of the rank 3 submodule \(\text{Span}_p \{w_1, w_2, w_3\}\).

The decay lemma in this case follows from the observation that since \(2a \leq A\), then \(w_3\) decays very rapidly. \(\square\)

5.4. 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 \(x(t)y(t)\) has the same \(t\)-adic valuation as \(z^2/2\)). 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): \(B(1 + p^{2c-1}) < A(1 + p) < B(1 + p^{2e+1})\) for some \(e \in \mathbb{Z}_{\geq 1}\). In this subcase, we will prove that \(\text{Span}_p \{w_1, w_2, w_3\}\) decays rapidly, where \(e \in \{3, 4, 5\}\) will be chosen depending on the values of \(a, b\) and \(c\).

The following lemma, in conjunction with Lemma 5.2.5, implies (as in Case 1) that \(\text{Span}_p \{w_1, w_2\}\) decays rapidly. It can be proved by the same argument as in the proof of Lemma 4.2.7(1), so we omit its proof.

**Lemma 5.4.1.** Among the terms appearing in \(F_{\infty}(1)\) described in Lemma 5.2.5 with denominator \(p^{n+1}\), the unique term with minimal \(t\)-adic valuation is

\[
P(1)_{e,n} (xy + z^2/2)(1 + \ldots + p^{-e}) (xy^p + xy + \ldots + p^{e+1}).
\]

The \(t\)-adic valuation of this term is \(\max(1 + \ldots + p^{-e}) + \min(B, p^{e+1} + \ldots + p^{e+1}). \)

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 5.4.2.** Given \(g \in \mathbb{Z}_{\geq 1}, n \in \mathbb{Z}_{\geq 0}\), consider the multiset consisting of numbers of the form \(A(1 + \ldots + p^{n-f-1}) + B(p^{n-f+1} + p^{n-f+2} + \ldots + p^{n+1}) + gp^{n+f}, \) 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 \(f_1 < f_2\), 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 \(A(1 + p + \ldots + p^{f_2-f_1-1}) = B(p^{f_2-f_1-1})(p^{f_2-f_1+1} + 1)/(p^2 - 1) + gp^{f_2}(p^{f_2-f_1})\).

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

\[
1 + p + \ldots + p^{f_2-f_1-1} \frac{v(f_2) - v(f_2 - 1)}{p^{n-f_2}} = \frac{B(p^{f_2-f_1-1})(p^{f_2-f_1+1})}{p^2 - 1} + gp^{f_2-f_1} = \frac{A(1 + \ldots + p^{n-f_1})}{p^{n-f_1+1}} + B(p^{n-f_1+1} + p^{n-f_1+2} + \ldots + p^{n+1}) + gp^{n+f_1}.
\]

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

**Lemma 5.4.3.** There are at most two numbers \(g\) in the set \(\{a, b, c\}\) such that there exists an integer \(f\) (if \(f\) is allowed to depend on the choice of \(g\)) with \(A(1 + \ldots + p^{n-f-1}) + B(p^{n-f} + p^{n-f+2} + \ldots + p^{n+1}) + gp^{n+f} = A(1 + \ldots + p^{n-f}) + B(p^{n-f+1} + p^{n-f+2} + \ldots + p^{n+1}) + gp^{n+f-1}\).

**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 5.4.2, we have that \(ap^{f_1-1}(p-1) = A - B(1 + p^{f_1-1})/(1 + p)\),

\[\text{Note that if the equation holds, then } f \text{ is independent of } n, \text{ since the equation is actually independent of } n; \text{ see the proof of Lemma 5.4.2.}\]
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 the Decay Lemma in this case.** Let \( h \in \{a, b, c\} \) be such that there is no \( f \) which satisfies the hypothesis of Lemma 5.4.3 (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 \( \text{Span}_{\mathbb{Z}_p} \{w_1, w_2, w_3\} \) 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 5.2.5, 5.4.1 and 5.4.3 imply that \( \text{Span}_{\mathbb{Z}_p} \{w_1, w_2\} \) and \( \text{Span}_{\mathbb{Z}_p} \{w_3\} \) 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 \( \alpha_u \) or \( \alpha_3 \) in \( \mathbb{Z}_p \) is a \( p \)-adic unit.

By Lemma 5.4.1, 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 \( 1/p^{m+1} \) of \( F_\infty w_3 \) is \( c(m) = A (1 + \ldots + p^{m-f}) + B (p^{m-f} + p^{m-f+2} + \ldots + p^{m+2}) + ap^{m+f} \) 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 \( c(f', m) = A (1 + \ldots + p^{m-f}) + B (p^{m-f} + p^{m-f+2} + \ldots + p^{m+2}) + ap^{m+f} \), 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 \( d(n) > c(f, m) \geq c(f, m) = c(m) \), as required. On the other hand, if \( m > n \), we have \( c(m) > A (1 + \ldots + p^{m-f}) + B (p^{m-f} + p^{m-f+2} + \ldots + p^{m+2}) \geq d(n) \), where the last inequality follows from Lemma 5.4.1.

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 \( f_1, f_2 \) satisfying the equation in Lemma 5.4.3. Since \( a \neq c \), then \( f_1 \neq f_2 \) and at least one \( f_i \geq 2 \). By the proof of Lemma 5.4.3, we have \( A - B (1+p^{2f_1-1})/(p+1) > 0 \) and hence \( A > 7B > 2b \). Thus, \( w_4 \) decays very rapidly. \(\square\)

**Subcase (2.2.2):** \( A (1 + p) = B (1 + p^{2e-1}) \) for some \( e \in \mathbb{Z}_{\geq 1} \). In this subcase, we will prove that \( \text{Span}_{\mathbb{Z}_p} \{w_3, w_4, w_5\} \) decays rapidly. We first need the following lemma.

**Lemma 5.4.4.** Among the terms appearing in \( F_\infty(2) \) described in Lemma 5.2.3 with denominator \( p^{n+1} \), the unique term with minimal \( t \)-adic valuation is

\[
P(1) e_{-1, n-1} F_u^{(n+e-1)} (xy + z^2/2) (1 + \ldots + p^{n-e}) (x + p^y + z^{1+p} + p^{n-e+1} + p^{n-e+3} + \ldots + p^{n+e-3}).
\]

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 \( A (1 + p + \ldots + p^{n-f}) + B (p^{n-f+1} + p^{n-f+3} + \ldots + p^{n+1}) + gp^{n+f-1} \), where \( f \) is allowed to range between 0 and \( n \). This can be verified by direct calculation. \(\square\)

**Proof of the Decay Lemma in this case.** It follows from Lemmas 5.2.5 and 5.4.4 that \( w_3, w_4 \) and \( w_5 \) individually decay rapidly, and that \( w_3 \) decays very rapidly. In order to show that \( \text{Span}_{\mathbb{Z}_p} \{w_3, w_4, w_5\} \) 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 5.4.4, these quantities equal \( A (1 + p + \ldots + p^{l-e}) + B (p^{l-e+1} + p^{l-e+3} + \ldots + p^{l+e-3}) + ap^{l+e-1} \), \( A (1 + p + \ldots + p^{l+e-1}) + B (p^{l+e+1} + p^{l+e+3} + \ldots + p^{l+1}) + ap^{l+1} \), and \( A (1 + p + \ldots + p^{l+e-1}) + B (p^{l+e+1} + p^{l+e+3} + \ldots + p^{l+1}) + ap^{l+1} \). These are easily shown to be distinct for different values of \( l, m, n \). \(\square\)
Lemma 5.5.2. As in Lemma 5.4.2 and 5.4.3, so we omit the details. This lemma follows from a similar argument as Lemma 4.2.7(2) and the proofs of Lemmas 5.2.3.

Proof. \[1 \geq \frac{1}{p^{n+1}} P(1)_{a,n−1}(xy + z^2/2)^{(1+...+p^{n−e})} (xy^p + x^p y + z^1 p^{p−e+1} + p^{n−e+3} + ... + p^{n+e−3}) \]

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, \(\text{Span}_{\mathbb{Z}_p} \{w_3, w_4, w_5\}\) decays rapidly. □

5.5. 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 \(z(t) = \gamma t^a + \sum_{i=a}^{\infty} \gamma_i t^i\). Since \(A \geq B\), we have \(\beta + \gamma^2/2 = 0\). We will break the proof of the Decay Lemma into two subcases and the following lemma will be used.

Lemma 5.5.1. 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 \(\beta + \gamma^2/2 = 0\), then \(\beta \in \mathbb{F}_p\). Therefore, in \(k[[t]]\),
\[
xy + z^2/2 = \sum_{i ≥ a'} (\beta_i + \gamma_{i} t^i + a + 1/2 \sum_{i,j ≥ a'} \gamma_i j t^{i+j},
\]
\[
xy + x^p y + z^{1+p} = \sum_{i ≥ a'} (\beta_i + \gamma_{i} t^i + a + 1/2 \sum_{i,j ≥ a'} \gamma_i j t^{i+j} + 1/2 \sum_{i,j ≥ a'} \gamma_i j t^{i+j},
\]

If one of \(\beta_{a'}\) and \(\gamma_{a'}\) were zero, then \(A = a' + a\), whereas \(B ≥ a' + pa\). Hence, we obtain the first assertion of the lemma.

Let \(a'' ≥ a'\) denote the smallest integer such that \(\beta_i + \gamma_i \neq 0\). Then \(\beta_{a''} + \gamma_{a''} \neq 0\), and \(B ≥ \min\{(p + 1)a', a'' + pa\}. If \(B ≥ (p + 1)a'\), then the second assertion of the lemma follows.

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

Case (3.1): \(B(1 + p^{2e−1}) < (p+1)A < B(1 + p^{2e+1})\) for some \(e \in \mathbb{Z}_{≥1}\).

The same argument as in Case 2.1 suffices to prove the Decay Lemma, unless \(A = B1+p^{2e−1}+1/p^{1+p}\). Therefore, we will assume that this is the case.

Lemma 5.5.2. Among the terms appearing in \(F_{∞}(2)\) described in Lemma 5.2.3 with denominator \(p^{n+1}\), there are exactly two with minimal \(t\)-adic valuation. They are:
\[
P(1)_{e−1,n−1}F_u^{(n+e−1)}(xy + z^2/2)^{(1+...+p^{n−e})} (xy^p + x^p y + z^{1+p} p^{p−e+1} + p^{n−e+3} + ... + p^{n+e−3}),
\]
\[
P(1)_{e,n−1}F_u^{(n+e)}(xy + z^2/2)^{(1+...+p^{n−e−1})} (xy^p + x^p y + z^{1+p} p^{p−e} + p^{n−e+2} + ... + p^{n+e−2}).
\]

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

Proof. This lemma follows from a similar argument as Lemma 4.2.7(2) and the proofs of Lemmas 5.4.2 and 5.4.3 so we omit the details. □

Proof of the Decay Lemma 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 5.5.2 appearing in the coefficient of \(1/p^{n+1}\) of \(F_{∞}(w_3)\) and \(F_{∞}(w_5)\). A direct computation yields that the sum of these two terms equals by
\[
\frac{1}{p^{n+1}} P(1)_{a,n−e−1}(xy + z^2/2)^{1+p+...+p^{n−e−1}} (X(t)u(t)^{p^{2e}} + Y(t)u(t)^{p^{2e−1}})^{(n−e)},
\]

- \(u(t)\) stands for either \(x(t)\) or \(z(t)\), according to whether we work with \(w_3\) or \(w_5\),
The decay of \( w_3 \) and \( w_5 \) is determined by the \( t \)-adic valuation of the entries of \( X(t)u(t)p^{2e} + Y(t)u(t)p^{2e-1} \). Note that \( X(t) \) and \( Y(t) \) are multiples of the same vector \([1, \lambda]^{T}\) by some functions in \( W[[t]] \), and hence we will abuse notation by treating \( X(t) \) and \( Y(t) \) as functions which multiply the constant vector \([1, \lambda]^{T}\). 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 \( A + B(p + p^3 + \ldots + p^{2e-3}) + \alpha'p^{2e-1} \) for at least one choice of \( u(t) \) between \( x(t) \) and \( z(t) \), where \( \alpha' \) is defined in Lemma \( 5.5.1 \) 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 \( A(1 + \ldots + p^{2e-} \alpha) + B(p^{n+e-1} + p^{n+e+1} + \ldots + p^{n+e-1}) + \alpha'p^{n+e-1} \). This is sufficient to prove the rapid decay of \( w_3 \) or \( w_5 \). Indeed, this quantity is strictly less than \( A(1 + \ldots + p^{n-1}) + B(p^{n-f+1} + p^{n-f+1} + \ldots + p^{n+f-3}) + ap^{n-1} \) for all values of \( f \neq e, e + 1 \) by Lemma \( 5.5.1 \) and hence the sum of the two terms in Lemma \( 5.5.2 \) 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 \( \alpha' \) in Lemma \( 5.5.1 \) proves that \( w_3 \) or \( w_5 \) decays very rapidly.

We now prove the claim by contradiction. Suppose that \( X(t)z(t)p^{2e} + Y(t)x(t)p^{2e-1} \) has \( t \)-adic valuation greater than \( A + B(p + p^3 + \ldots + p^{2e-3}) + \alpha'p^{2e-1} \). Since \( z(t) = \gamma x(t) + \gamma a't^\gamma + \ldots \) 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 \( X(t)z(t)p^{2e} + Y(t)x(t)p^{2e-1} \) with \( t \)-adic valuation \( A + B(p + p^3 + \ldots + p^{2e-3}) + \alpha'p^{2e-1} \), thereby establishing the claim for \( u(t) = z(t) \).

2. Either \( \beta \) or \( \gamma \) is not in \( \mathbb{F}_p \).

In this case, as \( \beta + \gamma^2 / 2 = 0 \), 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 \( \mu_X \) and that of \( Y(t) \) is \( \mu_Y \). 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), \( w_5 \) 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 \( \mu_X \gamma p^{2e} + \mu_Y \gamma^3 p^{2e-1} = 0 \), 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, \( \text{Span}_{\mathbb{Z}_p} \{w_1, w_2\} \) decays rapidly, and also every vector that can be written as \( \alpha, w_1 + \alpha w_2 \) with \( \alpha \in \mathbb{Z}_p^{\alpha} \) (\( i = 3, 5 \) depending on whether \( w_3 \) or \( w_5 \) 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 \( \text{Span}_{\mathbb{Z}_p} \{w_1, w_2\} \) decays rapidly.

\[ (3.2) : A(1 + p) = B(1 + p^{2e-1}) \text{ for some } e \in \mathbb{Z}_{\geq 1}. \]

\textbf{Lemma 5.5.3.} Among the terms appearing in \( F_{\infty}(1) \) described in Lemma \( 5.2.3 \) with denominator \( p^{n+1} \), there are exactly two with minimal \( t \)-adic valuation. They are:

\[ P(1)_{e, n}(x^2 + y^2 + \ldots + p^{n-e}(x^2 + y^2 + \ldots + p^{n-e} + \ldots + p^{n+e-1}), \]

\[ P(1)_{e-1, n}(x^2 + y^2 + \ldots + p^{n-e+1}(x + y + \ldots + p^{n-e+1} + \ldots + p^{n+e-2}). \]

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

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 \( \text{Span}_{\mathbb{Z}_p} \{w_1, w_2\} \)
decays rapidly. Nevertheless, the following lemma shows that there is at least a saturated rank one submodule of $\text{Span}_{F_p} \{w_1, w_2\}$ which decays rapidly.

**Lemma 5.5.4.** There is a vector $w_0$ in $\text{Span}_{F_p} \{w_1, w_2\}$ which decays rapidly.

**Proof.** By Lemmas 5.2.5 and 5.5.3, the coefficient (viewed as a power series in $t$) of the sum of the two terms with minimal $p$-adic valuation among the terms with denominator $p^{n+1}$ is of the form

$$\mu_1 M_1 + \mu_2 M_2,$$

for some $p$-adic units $\mu_i$, 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 $M_1 \text{ mod } p$ and $M_2 \text{ mod } p$ are not scalar multiples of each other, the linear combination $\mu_1 M_1 + \mu_2 M_2 \text{ mod } p$ is non-zero. Therefore, there exists a vector $\tilde{w}_0$ defined over $\mathbb{F}_p$ which does not lie in $\ker(\mu_1 M_1 + \mu_2 M_2 \text{ mod } p)$. Choosing $w_0 \in \text{Span}_{F_p} \{w_1, w_2\}$ which lifts $\tilde{w}_0$ finishes the proof of this lemma. \hfill \square

We are now ready to prove the last remaining case of the Decay Lemma.

**Proof of the Decay Lemma.** We will first prove that there is a rank 2 submodule of $\text{Span}_{F_p} \{w_3, w_4, w_5\}$ which decays rapidly. For ease of notation, let $F_u$ denote the matrix $\frac{1}{t} F_u$ evaluated at $t = 0$.

Let $K$ denote $\ker(P(1)_{u-1, e-1} F_u \text{ mod } p)$. If $\dim_{F_p} K \leq 1$, then lifting two linearly independent $\mathbb{F}_p$-vectors $\not\in K$ gives the desired rank 2 submodule. Therefore, we assume that $\dim_{F_p} K = 2$ (not that since the matrix mod $p$ is non-zero, so $\dim_{F_p} K \neq 3$). It follows that $\beta, \gamma \in \mathbb{F}_p$.

We will prove that $\text{Span}_{F_p} \{w_3, w_4\}$ decays rapidly. First, since $K \cap \text{Span}_{F_p} \{w_3, w_4\} = \text{Span}_{F_p} \{\beta w_3 - w_4\}$, then any primitive vector in $\text{Span}_{F_p} \{w_3, w_4\}$ 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 $\beta_{\alpha} t^A(1+\ldots+t^p-\ldots) + B(p^n+1+p^n+2+\ldots+p^n\ldots) + a'p^n\ldots$. Lemma 5.5.1 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 5.5.1 implies (as in the proof in Case 2.1) that $\text{Span}_{F_p} \{w_3, w_4\}$ decays rapidly. Finally, the very rapid decay of $w_3, w_4$ follows from the bound $2a' \leq B \leq A$.

Then, the Decay Lemma follows by an argument analogous to that in Case 2.1 with Lemma 5.5.4. \hfill \square

6. THE SETUP OF THE MAIN PROOFS

In this section, we provide the general setup of the proofs of Theorems 1 and 2. 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 $3/4$ of the global contribution; and

(2) the local contribution from non-supersingular points is of smaller magnitude.

Proposition 6.2.2 makes (1) precise, and is stated in §6.2. We will prove Proposition 6.2.2 and (2) in §7 for the Hilbert case and in §8 for the Siegel case. The idea involved in the statement of Proposition 6.2.2 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 §6.1. We also relate the local intersection multiplicity at a point to a lattice-point count.

6.1. Decomposition of the global contribution. 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 lemmas. Note that in the proof, we will only use this notion for a supersingular point. We recall the statement of the following lemma from §3.

**Lemma 6.1.1.** The intersection number $Z(m).C \sim -q_L(m)(\omega.C)$.  

28
The relation between the non-ordinary locus and the vanishing of the Hodge bundle is well known. The precise statement we need is:

**Lemma 6.1.2.** The non-ordinary locus is cut out by a Hasse-invariant $H$, which is a section of $\omega^{p-1}$, and hence the number of non-ordinary points (counted with multiplicity) on $C$ is given by $(p-1)(C.\omega)$. Note that this claim extends to the boundary of the Shimura variety.

See for instance [Box15 §1.4] for an explanation of this fact (and we use the fact that the ordinary Newton stratum coincides with the ordinary Ekedahl–Oort stratum).

**Definition 6.1.3.** Let $t$ be the local coordinate at $P$ and let $A = v_t(H)$. We define $g_P(m) = \frac{A}{p-1}|q_L(m)|$.

Note that by the above lemmas, $\sum_{P \in C \text{ non-ord}} g_P(m) = |q_L(m)|(\omega.C) \sim Z(m).C$.

### 6.2. The lattices and the outline of the proof.

Motivated by the Decay Lemmas in previous sections, we define the following lattices (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). Let $B \rightarrow \text{Spf} \mathbb{F}[[t]]$ denote a generically ordinary abelian surface with supersingular reduction. We will think of $B$ as an abelian surface with no extra endomorphisms or as a surface with real multiplication depending on whether the focus is on part (1) or (2) of Theorem [1]. We first make explicit what the local intersection multiplicity $\text{Spf} \mathbb{F}[[t]] . Z(m)$ is:

**Lemma 6.2.1.** The local intersection multiplicity $\text{Spf} \mathbb{F}[[t]] . Z(m)$ equals

$$\sum_{k=1}^{\infty} \#\{\text{Special endomorphisms of } B \text{ mod } t^k \text{ with norm } m\}.$$ \[\]

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 $m$). For brevity, we denote this quantity by $l_P(m)$. Recall that $A$ is the integer such that the Hasse invariant has defining equation $t^A$.

- If $B \mod t$ is superspecial, define $L_{n, 1}$ to be the lattice of special endomorphisms of $B \mod t^{A(1+p+...p^n)}$ and $L_{n, 2}$ to be the lattice of special endomorphisms of $B \mod t^{A(1+p+...p^n+1/2p^n+1)}$. Note that the decay lemmas imply that $L_{n, 2}$ has index at least $p$ inside $L_{n, 1}$, $L_{n, 1}$ has index at least $p^{3n}$, and $p^n L_{0, 1} \subset L_{n, 1}$.
- If $B \mod t$ is suporganerically, define $L_n$ to be the lattice of special endomorphisms of $B \mod t^{A(1+p+...p^n)}$. Again, the decay lemmas imply that $L_n$ has index at least $p^{3n}$ inside $L_0$ and $p^n L_0 \subset L_n$.

In particular, $L_{0, 1} = L'$ in Section 2.2 and $L_{0, 2}$ is the lattice after the first time a vector decays very rapidly.

To prove the main theorems, we consider the asymptotic of $\sum_{m \in T_M - S_M} C.Z(m)$ and the sum of the corresponding local contributions $\sum_{m \in T_M - S_M} l_P(m)$ as $M \rightarrow \infty$. Here $T_M = \{1, \ldots, M\}$ for Theorem [1](2), $T_M = \{\ell^2 \mid \ell \text{ prime, } \ell^2 \leq M\}$ for Theorem [1](1) and $T_M = \{\ell \mid \ell \text{ prime, } \ell \leq M\}$ for Theorem [2] and $S_M$ is a subset of $T_M$ with density $0$ as $M \rightarrow \infty$. 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 main task of the next two sections is to prove that

**Proposition 6.2.2.** Given $C$, there exists $S_M$ such that for every supersingular point $P$ on $C$, we have $\sum_{m \in T_M - S_M} l_P(m) \leq \frac{3}{4} \sum_{m \in T_M - S_M} g_P(m) + o(\sum_{m \in T_M - S_M} g_P(m))$.

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. Proof of Theorem 1(2)

In this section, we use the results proved in §§3-4 to prove Proposition 6.2.2 in the case of Hilbert modular surfaces. This, in conjunction with Lemma 7.1.1, yields Theorem 1(2).

7.1. Non-supersingular points.

Lemma 7.1.1. The local intersection at non-supersingular point is of smaller magnitude of the global intersection.\(^\text{13}\)

Proof. In the Hilbert case, the only points on special divisors are ordinary and supersingular points. For ordinary points, the lattice of special endomorphism is a rank 2 quadratic form. Since we have assumed that the curve \(C\) does not admit any global special endomorphisms, the discriminant of the rank 2 quadratic form goes to \(\infty\). Therefore, as \(M \to \infty\), the density of integers/primes represented by these quadratic goes to 0. We enlarge \(S_M\) in Section 7.2 to contain these bad numbers. Hence, there exists a constant \(c\) (which only depends on the ordinary point, the curve \(C\), \(p\), and \(\epsilon\), the small density of bad numbers; it is independent of \(m, M\)) such that for any \(m \notin S_M\), the local multiplicity is less than \(c\). Therefore, the local contribution of the ordinary point (intersecting with all \(m, M\) will be bounded by \(c \sum_{m=1}^{M} r_0(m) = O(M)\). It is easy to see that the global contribution is, up to multiplying by a constant bounded away from zero, \(M^2\). \(\square\)

7.2. Proof of Proposition 6.2.2 in the Hilbert case. The proof consists of three steps:

1. We construct a set \(S_M \subset \{1, \ldots, M\}\) of density bounded away from 1 (our methods are robust enough to deal with just prime numbers, if need be) such that the lattice (after the \(N^{th}\) decay) does not represent any \(m \notin S_M\). Here, \(N := (1/2 + \epsilon) \log_p M\). We choose a (necessarily positive and bounded away from 0) constant \(C_1\) such that the leading term of the global contribution is \(C_1 M^2\).

2. We observe that from (1), the contribution from beyond \(c\)-th decay is bounded by

\[
\sum_{n=c}^{N} \sum_{m=1, m \notin S_M}^{M} Cp^n r_n(m),
\]

where \(C\) is an absolute constant (depending on the curve and \(p\), but independent of \(m, n\)) and \(r_n(m)\) is the number of points in \(L_{n,1}\) (superspecial) or \(L_{n}\) (supergeneric) with norm \(m\). In other words, \(r_n\) is the theta-series associated to the lattice \(L_{n,1}\) (superspecial) or \(L_{n}\) (supergeneric). For any fixed \(\epsilon > 0\), we choose an absolute constant \(c\) only depending on the given \(\epsilon\) such that the above sum is less than \(c M^2 + o(M^2)\).

3. We bound the main term of the local contribution \(\sum_{m=1, m \notin S_M}^{M} \sum_{n=0}^{\alpha(n)} \alpha(n) s_n(m) + \beta(n) r_n(m)\), where \(\alpha(n) + \beta(n) = C p^n\) and \(\alpha(n) < \beta(n)\). Here \(r_n(m)\) is as above and \(s_n(m)\) is the number of points in \(L_{n,2}\) with norm \(m\) (and zero for supergeneric points). This main term will be controlled by breaking the theta-series into a sum of an Eisenstein series and a cusp form.

Step 1. As in [ST17], let \(n = N\), we see that the number of bad numbers (i.e. with local multiplicity larger than \(n\)) is bounded by \(O(M^2/p^{3n} + M^{3/2}/p^{2n} + M/p^n + M^{1/2}/d_n)\), which is \(O(M^{1-\epsilon})\) if \(p^{2n} = M^{1+\epsilon}\).

Global contribution. Since \(\{1, \ldots, M\} - S_M\) is a positive density set of numbers, we have the global contribution with leading term \(C_1 M^2\).

\(^{13}\)Note that we can achieve a sharper estimate by using Serre–Tate theory. See the proof of the ordinary points in the \((3, 2)\) case.
Step 2. In this part, we will show that \( \sum_{n=c}^{N} \sum_{m=1}^{M} Cp^nr_n(m) = \epsilon(c)M^2 + o(M^2) \), where \( \epsilon(c) \to 0 \) as \( c \to \infty \). In next part, we will see that the main contribution in Step 3 is strictly smaller than the global contribution (no matter which finite number \( c \) we choose) and hence we will choose \( c \) here so that \( \epsilon(c) \) is small enough.

Standard geometry-of-numbers arguments (see \[EK95\] Equations (5),(6) and Lemma 2.4\[14\]) yield the estimate \( \sum_{m=1}^{M} r_n(m) \leq C_2 M^2/p^3n + O(M^{3/2}/p^2n + M/p^n + M^{1/2}/dn) \), where \( C_2 \) is an absolute constant coming from the volume of unit ball. Hence we will take \( \epsilon(c) = C_2(p^{2c}(1 - p^{-2}))^{-1} \). On the other hand, the tail term is

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

Step 3. For each \( n \in \{0, \ldots, c\} \), we consider sublattices \( L_{n,1}, L_{n,2} \subset L' \) as in §6. (In particular, \( L_{0,1} = L' \), \( L_{0,2} \) admits one decay vector, \( L_{1,1} \) admits all three rapidly decay vector for the first time, etc.) For each lattice, we break the theta series \( \theta_{n,i} = E_{n,i} + f_{n,i} \), where \( f_{n,i} \) is a cusp form and \( E_{n,i} \) is an Eisenstein series whose Fourier coefficients are given in section 3. Let \( E = \sum_{n=0}^{\infty} \alpha(n)E_{n,1} + \beta(n)E_{n,2} \) and \( f = \sum_{m=0}^{c} \alpha(n)f_{n,1} + \beta(n)f_{n,2} \). Note that \( f \) is a weight 2 cusp form and we apply Deligne’s Weil bound, we have \( |f(m)| = O(m^{1/2+\epsilon}) \) (here the \( \epsilon \) is just to be safe, maybe unnecessary). Hence the total contribution from the cusp form \( f \) is \( \sum_{m=1}^{M} q_{n,i}(m) = O(M^{3/2+\epsilon}) \). Hence the main contribution come from the Eisenstein series \( E \). We compare the Fourier coefficients \( q_{n,i}(m), q(m)_{L'} \) of each \( E_{n,i} \) with \( E_{L'} \) (note that in §3 we’ve compared \( E_{L'} \) and \( E_L \)).

**Lemma 7.2.1.** For \( (m,p) = 1 \), we have \( \frac{q_{n,i}(m)}{q(m)_{L'}} \leq \frac{2p}{(p^2 + 1)|[L':L_{n,i}]|} \), where we have + for split case and − for inert case.

Recall that for \( p \nmid m \), there is no contribution from supergeneric points. Recall from §3 that if \( p \) is inert, \( q(m)_{L'}/(-q(m)_{L}) = (p-1)/(p^2 + 1); \) and if \( p \) is split, \( q(m)_{L'}/(-q(m)_{L}) = 1/(p-1) \). In other words, we have

\[
\frac{q_{n,i}(m)}{-q(m)_{L}} \leq \frac{2p}{(p^2 + 1)|[L':L_{n,i}]|},
\]

where we have + for inert and − for split. We now compute \( \frac{q_{n,i}(m)}{-q(m)_{L}} = \sum_n \gamma_{n,i} \). We want this ratio to be smaller than \( A/(p-1) \), where \( A = v_t(xy) \). Let \( B = v_t(x) \) and we may assume \( v_t(x) \leq v_t(y) \). By abuse of notation, we will use \( \gamma_{n,i} \) to denote the bound we get, may be larger than the actual ratio.

1. Inert case. \( \gamma_{0,1} = B/(p-1)^{p+1}, \gamma_{0,2} = 2(A-B)/(p^2+1), \gamma_{1,1} = 2(pA-pB)/(p^2+1)p^2, \gamma_{1,2} = 2pA/(p^2+1)p^2, \gamma_{2,1} = 2pB/(p^2+1)p^2, \gamma_{2,2} = 2B/(p^2+1)p^2 \), the rest terms are similar. We have

\[
\sum_{n=0}^{\infty} \gamma_{n,i} \leq \frac{B(p-1)}{p^2 + 1} + \frac{2B}{(p^2 + 1)(1-p^{-2})} + \frac{2(A-B)}{(p^2 + 1)(1-p^{-2})}.
\]

2. Split case. \( \gamma_{0,1} = B/(p-1)^{p+1}, \gamma_{0,2} = 2(pA-pB)/(p^2-1)p^2, \gamma_{1,1} = 2pB/(p^2-1)p^2, \gamma_{1,2} = 2B/(p^2-1)(1-p^{-2}), \gamma_{2,1} = 2pA/(p^2-1)p^2, \gamma_{2,2} = 2A/B/(p^2-1)(1-p^{-2}) \), the rest are similar. We have

\[
\sum_{n=0}^{\infty} \gamma_{n,i} \leq \frac{B}{p-1} + \frac{2B}{p(p^2-1)(1-p^{-2})} + \frac{2(A-B)}{(p^2-1)(1-p^{-2})} \leq \frac{3A}{4(p-1)}.
\]

\[\text{Sch68} \] for a proof of \([EK95\] Lemma 2.4\]}
Therefore, the local contribution of this supersingular point is at most $\frac{3}{2}$ of the global contribution (associated to this point) up to $O(M^{3/2+\epsilon})$.

To finish the proof for Theorem 1(2), we note that the set of $m$ such that $Z(m)$ is compact has density 1 (more precisely, for such $m$, the power of inert primes must be even) and hence we only consider intersection of $C$ with compact $Z(m)$.

8. PROOF OF THEOREM 1(1) AND THEOREM 2

In this section, we prove all of Theorem 1 and Theorem 2. §8.1 consists of results pertaining to squares represented by positive definite quadratic forms. In §8.2, we prove Proposition 6.2.2 by combining results proved in §6.2 and §8.1. Finally, we deal with the intersection multiplicities at non-supersingular points in §8.3 to finish the proof of the main theorem.

We now set up notation that we will use for §8 and in the proof of Proposition 6.2.2. Let the lattices $L_n,i$ be as in §6.2. Let $D_n$ denote the root discriminant of the lattice $L_{n,1}$. Let $P_n$ denote a rank two sublattice with minimal discriminant. Let $l(n), i = 1, \ldots, 5$ denote the $i$th successive minimum of the quadratic form restricted to $L_{n,1}$. Note that $l(n)_1l(n)_2$ (up to an absolutely bounded constant) equals $d_n$, the root discriminant of $P_n$.

8.1. Preparation. We need the following results to prove Proposition 6.2.2.

**Lemma 8.1.1.** Notation as above. Suppose that $d_n^2M \leq p^{2n}$. If there exists a vector $v \in L_{n,1}$ such that $Q(v) < M$, then $v \in P_n$.

**Proof.** Consider the vector $w = d_n^2v$. It is easy to see that $w \in P_n \oplus P_n^\perp$. Further, $Q(w) < d_n^2M$. Write $w = w_1 + w_2$ with $w_1 \in P_n$ and $w_2 \in P_n^\perp$. However, for every $v' \in P_n^\perp$ we have $p^{2n} | Q(v')$. If $w_2 \neq 0$, then $p^{2n} \leq Q(w_2) \leq Q(w)$. This is a contradiction, as we assumed that $Q(w) < d_n^2M \leq p^{2n}$.

**Lemma 8.1.2.** Suppose that $d_n \leq p^{n/2}$. If there exists a vector $v$ such that $Q_n(v) < p^{n-\epsilon}$, then $v \in P_n$.

**Proof.** Let $\ell_i$ denote the $i$th successive minima of $Q_n$. Each $\ell_i \leq p^n$, $\ell_1\ell_2 \sim d_n$ and $\ell_1\ell_2\cdots\ell_5 \sim p^{3n}$, where the implied constants are absolute. Therefore, it follows that $\ell_1\ell_2\ell_3 \geq p^n$, and therefore $\ell_3 \geq p^{n/2}$. In other words, any vector $v$ linearly independent to $v_1, v_2$ (a basis of $P_n$) has length $\geq p^{n/2}$, i.e. $Q(v) \geq p^n$. The lemma follows.

**Proposition 8.1.3.** Let $X_n$ denote the number of $v \in L_{n,1}$ such that $Q(v) = \ell^2$, where $\ell$ is a prime bounded above by $M^{1/2}$. Then we have the following two bounds: (this $\epsilon$ can be taken as an absolute constant; it just gives a lower bound of $M$. Once we fix $\epsilon$, the implied constant is absolute)

1. $X_n = O(M^{2+\epsilon \over p^m} + M^{3+2\epsilon \over p^m} + M^{1+\epsilon \over e_n}).$
2. $X_n = O(M^{5/2 \over p^m} + M^2 p^{-n} + M^{3/2 \over p^m} + M \over d_n + M^{1/2 \over e_n}).$

**Proof.** We note that (2) is a trivial upper bound. Indeed, the RHS of (2) bounds the number of $v \in L_n$ with $Q(v) \leq M$. Therefore, it remains to prove (1). There exists a primitive vector $e_1 \in L_n$ with the following properties:

- $Q(e_1) = p^m$ for some $m \leq n$.
- The bilinear form evaluated on $e_1, v$ is a multiple of $p^m$ for every $v \in L_n$.

\[\text{[15]}\] Recall that we must prove our curve intersects special divisors of the form $Z(m^2)$ 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.
Let \( e_1, e_2, \ldots, e_5 \) be a basis for \( L_n \). Then, by the above two properties, the lattice \( L'_n \) spanned by \( f_1 = e_1/p^m, e_2, \ldots, e_5 \) has the property that \( Q \otimes Q \) restricted to \( L'_n \subset L_n \otimes Q \) is an integral binary quadratic form. For brevity, we denote this form on \( L'_n \) by \( Q \). Note that \( Q(e_1/p^m) = 1 \). Therefore, there exist vectors \( f_2 \ldots f_5 \in L'_n \) such that the \( f_i \) are orthogonal to \( f_1 \) (\( i \geq 2 \), and \( f_1, f_2, f_3, f_4, f_5 \) is a basis for \( L'_n \).

Consider \( Q \) restricted to the span of \( f_2 \ldots f_5 \). Denote this form by \( Q' \). It is positive definite and has cumulative products of successive minima bounded below by \( 1, p^{n-m}, p^{2n-m}, p^{3n-m} \). Therefore, \( X_n \) is bounded by the number of solutions to \( x^2 + Q'(y_2 \ldots y_5) = z^2 \), with \( z^2 \leq M \).

We therefore have \((z + x)(z - x) = Q'(y_i)\), where \( M^{1/2} \geq z \geq x \). For each fixed value of \((z + x)(z - x)\), there are at most \( M' \) ways of factoring it as a product. Therefore, \( X_n \) is bounded by \( M'y_n \), where \( Y_n \) is the number of \( v \) such that \( Q'(v) \leq M \). As the cumulative products of successive minima of \( Q' \) are bounded below by \( 1, p^{n-m}, p^{2n-m}, p^{3n-m} \), we have \( Y_n = O\left(\frac{Y^2}{p^m} + \frac{Y^{3/2}}{p^{n/2}} + Y\right) \). The proposition follows.

\[ \square \]

**Proposition 8.1.4.** The proportion of primes \( \ell \leq M^{1/2} \) such that \( \ell^2 \) is represented by the quadratic form restricted to \( P_n \) goes to zero as \( n \) grows to infinity.

**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 \( D_n \). Let \( \mathfrak{a} \) denote the ideal corresponding to \( Q \) restricted to \( P_n \). We may assume that \( \mathfrak{a} \) is not equivalent to the unit idea because \( Q \) does not represent 1. Note that it suffices to deal with primes \( \ell \) which are relatively prime to \( d_n^2 \).

The correspondence between ideal classes and binary quadratic forms yields that a prime \( \ell^2 \) is represented by \( Q \) if and only if there exists an invertible ideal \( \mathfrak{b} \) equivalent to \( \mathfrak{a} \) with \( Nm \mathfrak{b} = \ell^2 \). This implies that \( \ell = c_1c_2 \) (i.e. the prime \( \ell \) splits in \( R_n \)), and that \( \mathfrak{b} = c_1^2 \) or \( \mathfrak{b} = c_2^2 \) (the case \( \mathfrak{b} = c_1c_2 \) is ruled out as we observed that \( \mathfrak{a} \) and therefore \( \mathfrak{b} \) is not equivalent to the unit ideal). In other words, \( Q \) restricted to \( P_n \) represents \( \ell^2 \) if and only if there exists some ideal \( \mathfrak{c} \) with norm \( \ell \) whose square is equivalent to \( \mathfrak{a} \).

Let \( C \) denote the equivalence class of ideals \( \mathfrak{c} \) such that \( \mathfrak{c}^2 \) is equivalent to \( \mathfrak{a} \) – note that \( C \) is a torsor for the 2-torsion of the class group when \( C \) is nonempty. We deal with two cases: if \( d_n \leq (\log M)^2 \), it follows by [TZ18] that the proportion of primes represented by any one form \( \mathfrak{c} \) is \( 1/d_n \). On the other hand, if \( d_n \geq (\log M)^2 \), then the proportion of integers \( \leq M^{1/2} \) represented by \( \mathfrak{c} \) is \( 1/d_n \). Further, \( \#C \leq d_n^6 \). The proposition follows. \[ \square \]

The following result gives a bound of Fourier coefficients of the cuspidal part of our theta series in terms of discriminant of quadratic lattice. Let \( \theta_n \) denote the modular form associated to the lattice \( L_n \). Let \( \theta = E + f \), where \( E \) is the associated Eisenstein series and \( f \) is the associated cusp form. Let \( r_m, q_m \) and \( a_m \) denote the Fourier coefficients of \( \theta \), \( E \) and \( f \) respectively. We have the following result:

**Proposition 8.1.5** (Duke). Suppose that the quadratic form associated to \( \theta \) has discriminant \( D \). Then, there exist absolutely bounded positive constants \( N_0 \) and \( C \) such that \( a_m \leq CD^{N_0}m^{1+1/4} \).

In the above result, the exponent of \( 1 + 1/4 \) can be improved to \( 1 + \epsilon \) for any \( \epsilon > 0 \), but then the constants would depend on the choice of \( \epsilon \). Further, this result 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.

\[ ^{16} \text{Genus theory yields that the two-torsion of the class group of } R_n \text{ is bounded by the number of prime divisors of } d_n^2 \]
Proof. As in Lemma 1 (and the discussion following the statement of Lemma 1) [Duk88], the Peterson norm of the cusp form \( f \) can be bounded polynomially in terms of \( D \). [Wai18, Theorem 1] can now be used to obtain the required bounds on \( a_m \).

8.2. Proof of the main theorem in the Siegel case. Instead of summing over squares, we sum over prime squares, and hence the global contribution is \( \epsilon_0 M^2 (\log M)^{-1} \). We treat the supersingular contribution with respect to different ranges of the discriminant.

**Definition 8.2.1.** The ranges of the discriminant are defined as follows: (here, by \( \log \), we mean \( \log_q \))

- \( n \) is defined to be small if \( n \leq \epsilon_0 \log M \), where \( \epsilon_0 \) is a constant independent of \( M \).
- \( n \) is defined to be in the lower medium range if \( \epsilon_0 \log M < n \leq \frac{3}{4} \log M \)
- \( n \) is defined to be in the upper medium range if \( \frac{3}{4} \log M < n \leq (1 + \epsilon_1) \log M \) where \( \epsilon_1 \) is a constant independent of \( M \), which can be chosen to be arbitrarily small.
- \( n \) is defined to be large if \( n > (1 + \epsilon_1) \log M \).

**Bounding the contribution from \( n \) in the lower medium range.** We need to bound the quantity \( \sum_{n=\epsilon_0 \log M}^{\frac{3}{4} \log M} p^n X_n \). Using the bounds on \( X_n \) from Proposition 8.1.3(1), we see that \( \sum_{n=\epsilon_0 \log M}^{\frac{3}{4} \log M} p^n X_n = O(M^{2+\epsilon-\epsilon_0} + M^{3/2+\epsilon} \log M + M^{7/4+\epsilon}) \) which is \( o(M^{2-\delta}) \) for \( \delta = \epsilon/2 \), as long as \( \epsilon > 2\epsilon_0 \).

**Bounding the contribution from \( n \) in the upper medium range.** Let \( n_0 = \frac{3}{4} \log M \). We will deal with two cases according to whether \( d_{n_0} \leq M^{1/8} \) or not.

**Case 1: \( d_{n_0} \geq M^{1/8} \).** We need to bound the quantity \( \sum_{n=\frac{3}{4} \log M}^{(1+\epsilon_1) \log M} p^n X_n \). Using the bounds on \( X_n \) from Proposition 8.1.3(2), we see that \( \sum_{n=\frac{3}{4} \log M}^{(1+\epsilon_1) \log M} p^n X_n = O(M^{1/4} + M^{5/4} + M^{3/2} \log M + M^{15/8+\epsilon_1} + \frac{M^{3/2+\epsilon_1}}{\epsilon_0}) \) which is again \( o(M^{2-\delta}) \) for appropriately chosen \( \delta \). (note that \( \epsilon_1 \) is fixed and can be chosen to be arbitrarily small.)

**Case 2: \( d_{n_0} < M^{1/8} \).** Then \( d_{n_0} M < M^{3/2} \) and by Lemma 8.1.1, if \( m \leq M \) is represented by \( Q_{n_0} \), then it is represented by \( P_{n_0} \). Then by Proposition 8.1.4, the size of the set of bad numbers (whose square is represented by \( P_{n_0} \)) is \( o(M^{1/2}/\log M) \) and we put (the squares of) these numbers into \( S_M \).

**Bounding the contribution from large \( n \).** Let \( n_0 = (1 + \epsilon_1) \log M \). Let \( \epsilon_2 \) be a small constant so that \( \epsilon_2 < \epsilon_1/2 \).

**Case 1: \( d_{n_0} \leq M^{1/2+\epsilon_2} \).** We will produce a density 0 set \( S_M \) such that for any prime square \( m \notin S_M \), it will not be represented by \( L_{n_0} \). Since \( d_{n_0} \leq M^{1/2+\epsilon_2} \) and \( M < p^{n_0-\epsilon_1} \), by Lemma 8.1.2, if \( m \) is represented by \( Q_{n_0} \), then it is represented by \( P_{n_0} \). By Proposition 8.1.4, the number of such \( m \) is \( o(M^{1/2}/\log M) \) (note that \( n_0 \to \infty \) as \( M \to \infty \) and hence \( d_n \to \infty \)).

\(^{17}\)Note that \( \epsilon_0 \) is an absolute constant chosen independent of \( M \), and \( \epsilon \) can be chosen to be as small as we want, as long as \( M \) is large enough.
Case 2: $d_{n_0} > M^{1/2+\epsilon_2}$. The number of bad squares is bounded by $X_{n_0}$. By Proposition 8.1.3(2), we have $X_{n_0} = O(M^{1/2-\epsilon_1} + M^{1/2-\epsilon_2} + M^{1/2}/\epsilon_{n_0})$. This will be $o(M^{1/2}/\log M)$ if $\epsilon_{n_0} > M^{\epsilon_3}$ for any fixed $\epsilon_3 > 0$.

If not, i.e. $\epsilon_{n_0} \leq M^{\epsilon_3}$, we take $\epsilon_3 < \epsilon_2$, then $l(n_0) > M^{1/2}$. In other words, any vector $v$ which is not a scalar multiple of the chosen vector of the smallest length has length $> M^{1/2}$, i.e. $Q_{n_0}(v) > M$. Therefore, any $m < M$ is represented by our form has to be represented by the rank 1 quadratic form. Since we only consider prime squares, we are done (indeed, as long as $\epsilon_{n_0} > 1$, the rank one quadratic form will represent at most one prime square. In the proof, we will choose $M$ large enough so that $\epsilon_{n_0} > 1$).

**Estimating the contribution from small $n$.** We consider $\sum_{\ell=1, m=\ell^2 \in S_M} \sum_{n=0}^{\epsilon_{n_0} M} \alpha(n) s_n(m) + \beta(n) r_n(m)$, where $\alpha(n) + \beta(n) = C_1 p^n$ and the definition of $s_n, r_n$ is the same as in \S 7.2. As in previous section, we decompose the above sum into $\sum_m q_E(m) + q_f(m)$, where $q_E(m)$ is the Eisenstein contribution and $q_f(m)$ is the cuspidal contribution. The difference is that in the Hilbert case, we have a finite sum $\sum_{n=0}^{\epsilon_{n_0} M}$ which is fixed for all $m, M$ while here the Eisenstein series $E$ and the cuspidal form $f$ depend on $M$.

For the cuspidal form $f$, by Proposition 8.1.5 we have $q_f(m) \leq C_0 (p^{6n_0 log M})^{N_0} m^{5/4} \sum_{n=0}^{\epsilon_{n_0} M} s_n(m)$.

As in the Hilbert case, we give an estimate of the Eisenstein series part independent of the choice of $\epsilon_0$. We write $\frac{q_E(m)}{\ell L} = \sum_{n,i} \gamma_{n,i}$ and we want this ratio to be $< A/(p-1)$, where $A = v_t(xy+z^2/2)$ if superspecial and $A = v_t(xy-(\sigma(c)-\sigma^{-1}(c))) + z^2/2$ if suprergeneric.

(1) Superspecial case. As notation in Theorem 5.1.3. By Theorem 5.1.3 and the computation of local density, we have $\gamma_{n,1} = \frac{a}{p-1}$. For smaller lattices $L_{n,i}$, we have $\frac{q_{n,i}(m)}{\ell L} \leq \frac{2a}{(p-1)^2}$.

Hence $\gamma_{n,2} = \frac{2(A-a)}{p-1}$, $\gamma_{1,1} = \frac{2pa}{p(p-1)} = \frac{2a}{p-1}$. The estimate is the same as the split Hilbert case and we conclude that $\sum_{n=0}^{\infty} \gamma_{n,i} \leq \frac{2A}{p-1}$.

(2) Supergeneric case. In this case, we only use the decay theorem which asserts that there is a rank 3 module which decays rapidly (i.e. without using any knowledge of some vector decays very rapidly hence we do not distinguish $L_{n,1}$ and $L_{n,2}$). Then $\gamma_0 = \frac{2A}{p^2-1}$, $\gamma_1 = \frac{2pA}{p^3(p^2-1)}$, $\gamma_2 = \frac{2p^2A}{p^3(p-1)}$ and etc. We have $\sum_{n=0}^{\infty} \gamma_n \leq \frac{A}{p^2-1}(1-\frac{1}{p-1}) \leq \frac{9A}{16(p-1)}$.

Note that for the proofs of Theorem 1 and Theorem 2 we only work with $Z(m)$ with $p \nmid m$ so every supersingular point on $Z(m)$ is superspecial and we do not need case (2) above. We include the computation nevertheless, which uses results from Appendix 3.

To finish the proof of Theorem 1, we only need to show that away from a small density set, the local contribution at a non-supersingular point has smaller order of magnitude than the global intersection number. This is proved in the next subsection.

In order to prove Theorem 2 we consider intersection $C.Z(\ell)$ and the rest of argument goes along similar lines.

### 8.3. Contribution from non-supersingular points.

To finish the proof, we only need to show that the contribution from non-supersingular points are $o(M^2/\log M)$. By the classification of endomorphism rings of char $p$ abelian surfaces (and Tate’s Honda–Tate theory paper), we see that if the abelian surface has almost ordinary reduction (i.e. slopes 0, 1/2 and 1), then its module of special endomorphism has rank at most 1. Since $C$ does not have global endomorphism, the rank 1 quadratic forms stops representing a prime square after any decay. In other words, the local contribution at a middle case point is $c_0 M/\log M = o(M^2/\log M)$. 

35
Now we work with ordinary points. The following lemma follows directly from Serre–Tate theory. We thank Keerthi Madapusi Pera for pointing this out to us.

**Lemma 8.3.1.** Suppose that $f$ is a local equation at an ordinary point (with respect to $q$-coordinates) for the special divisor $Z(m)$. Then $f^p$ is a local equation for the special divisor $Z(p^2m)$.

Now we can control the local contribution as in the supersingular case. Note that mod$p$, at the ordinary point, the lattice $L_0$ is at most rank 3. If it is of rank 1, the same argument as in the almost ordinary reduction case as above works. Since $C$ does not admit global endomorphism, the $p$-divisible group over $C$ admits at most a rank 2-module of special endomorphisms (if it is rank 3, then it must coincides with $L_0$, which is rational and thus the abelian surface admits extra endomorphisms). Therefore, there exists a vector $w \in L_0$ such that $w$ is not an endomorphism of the $p$-divisible group over $C$ and by Lemma 8.3.1 if $w$ is not an endomorphism for the abelian surface over mod $t^A$, then $pw$ is not an endomorphism over mod $t^{pA}$. Fix $A$ and let $L_n$ be the lattice of special endomorphisms over mod $t^{p^{n-1}A}$, i.e. $L_n \subset (\Lambda + p^nL_0) \cap L_0$, where $\Lambda$ is a module over $\mathbb{Z}_p$ of rank at most 2. In particular, the discriminant of $L_n$ is at least $p^n$.

First, we need the following lemma. From now on, by log we mean $\log_p$.

**Lemma 8.3.2.** Given $M$, set $n_0 = (1 + \epsilon_0)\log M$. The density of primes squares in $[1, \ldots, M]$ represented by $L_{n_0}$ goes to 0 as $M \to \infty$.

**Proof.** The number of positive integers in $[1, \ldots, M]$ represented by $L_n$ is at most $X_n = O(M^{3/2} + \frac{M}{b_n} + \frac{M^{1/2}}{a_n})$, where $a_n$ is the minimal length of a non-zero vector in $L_n$ and $b_n$ is the minimal root discriminant of a rank 2 sublattice in $L_n$. Fix $\epsilon_1 < \epsilon_0/4$.

(1) $a_{n_0} < M^{\epsilon_1}$ and $b_{n_0} > M^{1/2 + 2\epsilon_1}$. Then any vector in $L_{n_0}$ which is not a scalar multiple of the given vector $v_0$ with length $a_{n_0}$ has length at least $b_{n_0}/a_{n_0} > M^{1/2 + \epsilon_1}$. Hence for primes squares $\leq M$ represented by $L_{n_0}$, it must be represented by the rank 1 quadratic form (spanned by $v_0$). Note that for $M$ large enough, $a_{n_0} > 1$ and hence no prime square is represented in this case.

(2) $a_{n_0} \geq M^{\epsilon_1}$ and $b_{n_0} > M^{1/2 + 2\epsilon_1}$. Then the number of bad prime squares is $X_{n_0} = O(M^{1/2+\epsilon_1}) = o(M^{1/2}/\log M)$.

(3) $b_{n_0} \leq M^{1/2 + 2\epsilon_1}$. Then $p^{n_0/2} = M^{1/2 + \epsilon_0/2} \geq b_{n_0}$ and by Lemma 8.1.2 (note the proof of this lemma applies to this case), if an integer $\leq M = p^{n_0-\epsilon}$ is represented by $L_{n_0}$, then it is represented by the rank 2 lattice $P_{n_0}$ (i.e. the rank 2 sublattice in $L_{n_0}$ with minimal discriminant). By Proposition 8.1.4 the density of such prime goes to 0.

Now we prove the total local contribution at an ordinary point.

**Proposition 8.3.3.** At an ordinary point, the sum of the local intersection number of $C$ with $Z(\ell^2)$ for all good primes $\ell \leq M^{1/2}$ is $o(M^{2}/\log M)$ (here by good, we mean from a density zero set–the density zero set contains the bad primes above)

**Proof.** Notation as before, let $A \in \mathbb{N}$ such that the first decay happens at $t^A$. We write the total local contribution as $\sum_{A=1}^{A=M^{1/2}} \sum_{\ell \leq x} A(p^n - p^{n-1})r_n(\ell^2)$. Set $n_1 = \frac{3}{4} \log M$.

First, $\sum_{n=0}^{n_1} A(p^n - p^{n-1})r_n(\ell^2) \leq \sum_{n=0}^{n_1} C_0 A(p^n - p^{n-1})(\frac{M^{3/2}}{p^n} + O(M)) = O(M^{7/4})$.

For $\sum_{n=n_1}^{(1+\epsilon_0)\log M} A(p^n - p^{n-1})r_n(\ell^2)$, we bound it by studying the following two cases separately.

---

\[18\] This is not necessarily unique since we only talk about the local equation up to unit.
Appendix A. Local density computation

To prove our main theorems, we only work with $Z(m)$ when $p \nmid m$. By the same method in Section 3, we can still compare the Fourier coefficients of Eisenstein series from global and local contribution for any $m$.

A.0.1. We include the results here for future references. Following [Han04, §3, Def. 3.1], for vectors in a quadratic lattice over a local field, we separate them into good, zero, and bad types. We also define (representability) density of good, zero, or bad type to be the density of good, zero, or bad type vectors representing a given number.

A.1. The Hilbert case. We give the analogous result of Lemma 3.3.2 when $p \mid m$. Notice that

$$\sigma_s(m, \chi) = \prod_l \sigma_s(l^{r_l(m)}, \chi)$$

and $\chi_{4 \det L'}(p) = 0$. We have the following lemma.

**Lemma A.1.1.** Suppose that $L'$ is superspecial. For any $m \in \mathbb{N}$, we have

$$\frac{q(m)_{L'}}{-q(m)_L} = \frac{\delta(p, L', m)\sigma_1(m, \chi_{4 \det L'})}{p(1 - \chi_{4 \det L}(p)p^{-2})\sigma_1(m, \chi_{4 \det L})} = \frac{\delta(p, L', m)}{p(1 - \chi_{4 \det L}(p)p^{-2})\sigma_1(p^{v_p(m)}, \chi_{4 \det L})}.$$

If $L'$ is supergeneric, then divide the above ratio by an extra $p$. Note that $\sigma_1(p^{v_p(m)}, \chi_{4 \det L}) \geq \min\{1, 1 + \chi_{4 \det L}(p)/p\}$.

More explicitly, using the $\mathbb{Z}_p$-lattices given above, we can compute the local density by the inductive method in [Han04, §3].

1. $p$ inert in $F$. At superspecial points, we claim that $\delta < 2$. Indeed, if $p \mid m$, then we have both good and bad I types so $\delta = 2 - 2/p$ (good) $+ p^{-1}(1+1/p) = 2 - 1/p + 1/p^2$, where the second term $1+1/p$ is the good density of representing $m/p$ by $Q' = pxy + z^2 - Dw^2$. When $p^2 \mid m$, then we only have good and zero types, therefore $\delta(m) = 2 - 2/p + p^{-2}\delta(m/p^2)$ and by induction, we see $\delta(m) < 2$. Actually, the proof yields the inequality $\delta \leq 2 - 1/p + 1/p^2$.

We now deal with the case of superspecial points. If $p \mid m$, then $\delta = p(1 + p^{-2})$. If $p^2 \mid m$, we have $\delta_k(m) = p\delta_{k-1}(Q/p(m/p))$ and the good density for $Q/p$ at $m/p$ is $1 - 1/p + 1/p^2$. We only have good and zero types and by arguing inductively, we have $\delta < p$ in this case. Actually, we have $\delta < (1 - p^{-1} + 2p^{-2} - p^{-3})$.

2. $p$ split in $F$. As far as $p \mid m$, there is no good type. If $p \mid m$, then $\delta = p^{-1} + p^{-2}$. For $p^2 \mid m$, then $\delta_k(m) = p^{-2}\delta_{k-2}(m/p^2)$ and hence $\delta(m) \leq p^{-2} + p^{-3}$. Actually, we have $\delta(m) = p^{-v_p(m)} + p^{-v_p(m)+1}$.

A.2. the Siegel case. For $n = 3$, superspecial points, we have the following.

**Proposition A.2.1.** For $L'$ attached to superspecial points and for $m$ such that $p^2 \mid m$, the ratio $q(m)_{L'}$ is always strictly smaller than $(p - 1)^{-1}$. With the above computation, we see that for any $m$, the ratio is no greater than $(p - 1)^{-1}$.

**Proof.** We separate the two cases by $v_p(m)$ being even or odd. We give the details of the even case and the odd case is similar.
Lemma A.3.1. \( v_p(m) \) even. Then \( p \nmid \mathcal{D}, p \mid \mathcal{D}' \), for any \( s \) such that \( p \nmid s \), we have \( \chi_D(s) = \chi_{D'}(s) \); also \( f = p^{v_p(f)} f' \). Set \( A' = \sum_{d|f'} \mu(d) \chi_{D'}(d)d^{-2}\sigma_3(f'/d) \) and set \( A \) similarly. By definition of \( \mu \), we have \( A = I + II \), where \( I = \sum_{d|f'} \mu(d) \chi_{D}(d)d^{-2}\sigma_3(f/d) \) and \( II = \sum_{d=pq', d'|f} \mu(d) \chi_{D}(d)d^{-2}\sigma_3(f/d) \). By the property of divisor functions, we have \( I = A' \cdot \sigma_3(p^{v_p(f)}) \) and \( II = -\chi_D(p)p^{-2}\sigma_3(p^{v_p(f)-1})A' \). By our assumption on \( m \), we have \( \sigma_3(p^{v_p(f)}) \geq \max(1+p^{-3}, \sigma_3(p^{v_p(f)-1})) \) and \( \sigma_3(p^{v_p(f)-1}) \geq 1 \).

Similar to the simple case, we have \( \frac{q(m)_{L'}}{-q(m)_{L}} = \frac{\delta(p,L',m)(1-\chi_D(p)p^{-2})A'}{p(1-p^{-4})A} \). Discuss separately for \( \chi_D(p) = 1, -1 \), we have \( \frac{q(m)_{L'}}{-q(m)_{L}} \leq \frac{\delta(p,L',m)}{p(1-p^{-4})}. \) To prove the desired ratio bound, we only need to prove that \( \delta \leq (1+1/p-2+p^{-3}) \).

We use the same idea as before to compute \( \delta \) inductively: \( \delta_{\text{good}} = 1, \delta_{\text{bad}} = 0 \) because the \( p \)-multiple part is \( x_2^3 - ux_4^2 \), and \( \delta_{\text{zero}}(m) = p^{-3}\delta(m/p^2) \). Therefore, \( \delta = 1 + p^{-3} + \ldots + p^{-3v_p(f)} + p^{-3v_p(f)-1} < 1 + p^{-4} + p^{-3}(1-p^{-3})^{-1}, \) which is smaller than the desired bound.

(2) \( v_p(m) \) odd. Then \( \chi_D(s) = \chi_{D'}(s) \) for any \( s \) since \( p \nmid \mathcal{D} \) and \( \chi_D(s) = 0 \) if \( p \mid s \). Therefore, notation as above, \( A = A' \sigma_3(p^{v_p(f)}) \) and hence \( A'/A \leq (1+p^{-3})^{-1} \). To have the desired ratio for \( \frac{q(m)_{L'}}{-q(m)_{L}} \), it is enough to show that \( \delta \leq (1+1/p-2+p^{-3}) \).

Similar as above, we have \( \delta_{\text{good}} = 1, \delta_{\text{bad}} = 0, \delta_{\text{zero}}(m) = p^{-3}\delta(m/p^2) \) and inductively, we have \( \delta = 1 + p^{-3} + \ldots + p^{-3v_p(f)} + p^{-3v_p(f)-1} + p^{-3v_p(f)}-2, \) which is smaller than the desired bound.

\( \square \)

For supergeneric point, in addition to \( v_p(m) = 1 \), 0 cases, we have

(1) \( v_p(m) \geq 2 \) even. Then \( \delta(m) \leq 1 + 1 + p^{-1} - p^{-2} + p^{-3}\delta(m/p^2) \). Hence \( \delta < \frac{2+2p^{-1}p^{-2}}{1-p^{-3}} \) and

\[
\frac{q(m)_{L'}}{-q(m)_{L}} \leq \frac{\delta(p,L',m)}{(p^2-1)(1+p^{-2}+p^{-3})} < \frac{2+p^{-1}}{p^2-1}.
\]

(2) \( v_p(m) \geq 2 \) odd. Then \( \delta(m) \leq 1 + 1 + p^{-1} - p^{-2} + p^{-3}\delta(m/p^2) \). Hence \( \delta < \frac{2+2p^{-1}p^{-2}}{1-p^{-3}} \)

(different initial data, but we get the same bound) and

\[
\frac{q(m)_{L'}}{-q(m)_{L}} \leq \frac{\delta(p,L',m)}{(p^2-1)(1+p^{-2}+p^{-3})} < \frac{2+p^{-1}}{p^2-1}.
\]

A.3. Smaller lattices for the Siegel case. For general \( m \), we have the following estimate, which will not be needed in the proof of main theorems. Notation as in §6.

**Lemma A.3.1.** Fix \( m \) and consider a superspecial point. We have \( \frac{q(m)_{L'}}{q(m)_{L}} \leq C 2^np^{-9m/5} \).

Note that the character \( \chi \) will remain the same, hence the above lemma is a direct consequence of the local density \( \delta_n(m) \) of the quadratic form on \( L'_n \). More precisely, we will prove, by induction on \( n \), that

\[
\delta_n(m) \leq C 2^np^{6n/5}.
\]

**Remark A.3.2.** We may also obtain a bound by induction on \( v_p(m) \). In the base case when \( p \nmid n \), we have \( \delta_n(m) = \delta_{n, \text{good}}(m) \), which is the density computed on \( L'_n/pL'_n \) and \( \delta_{\text{good}} \leq 2 \) by [Han04, Table 1].

**Proof of the bound of local density.** Since \( p > 2 \), we may always diagonalize the quadratic form on \( L'_n \) as \( \sum_{i=1}^5 u_i^2 \), we have \( \sum_{i=1}^5 v_p(u_i) = 2 + 6n \). Since \( p^n L' \subset L'_n \subset L' \subset L'' \subset p^{-1}L' \), we have...
\[ L_n^\prime \subset p^{-2n-1}L_n. \] Therefore, \( v_p(u_i) \leq 2n + 1. \) Set \( s_j, j = 0,1 \) to be the number of \( u_i \) such that \( v_p(u_i) = j. \) Then \( s_0 + s_1 \leq 2. \) We have the following induction formula:

\[
\delta_n(m) = \delta_{n,\text{good}}(m) + p^{-3}\delta_n(m/p^2) + \min\{s_1,1\}p^{1-s_0}\delta_{\text{good}}(m/p) + \min\{s_2,1\}p^{2-s_0-s_1}\delta''(m/p^2),
\]

where the good densities \( \delta_{n,\text{good}}, \delta_{\text{good}} \) are both \( < 2 \) and \( \delta'' \) is the local density associated to the quadratic form by replacing all \( w_i, v_p(u_i) \geq 2 \) by \( u_i'' = u_i - 2. \) Note that after the change, \( v_p(u_i'') \leq 2(n-1) + 1 \) and \( \sum v_p(u_i'') \leq 2 + 6(n-1) \) (equality holds when \( s_0 + s_1 = 2 \)).

1. Assume \( s_0 + s_1 = 2. \) In this case, the quadratic form giving \( \delta'' \) satisfies the same condition as the quadratic form of \( L_{n-1}^\prime, \) so we will denote it by \( \delta_{n-1} \). Then

\[
\delta_n(m) \leq 2 + p^{-3}\delta_n(m/p^2) + 2p + \delta_{n-1}(m/p^2).
\]

Recall for \( n = 0, \) we have the base case \( \delta_0(m) \leq 1 + p^{-1}. \)

- \( v_p(m) \geq 2n. \) Induction shows

\[
\delta_n(m) \leq C2^n,
\]

where \( C \) is a fixed constant only depend on \( p. \) (e.g. \( C = 2 + 2p; \) by modifying \( C, \) we may replace 2 by any number > \( (1 - p^{-3})^{-1}. \)

- \( v_p(m) < 2n. \) The induction process stops after \( v_p(m)/2 \) steps and we have

\[
\delta_n(m) < (2 + 2p)p^3((1 + p^{-3})^{v_p(m)/2} - 1) + 2(1 + p^{-3})^{v_p(m)/2} < C2^{v_p(m)/2}.
\]

2. Assume \( s_0 + s_1 = 1. \) We have

\[
\delta_n(m) \leq 2 + p^{-3}\delta_n(m/p^2) + 2p + p\delta''(m/p^2).
\]

After at most \( k = (1 + 3n)/4 \) steps, we reduce to the \( s_0 + s_1 \geq 2 \) case and the induction argument gives

\[
\delta_n(m) \leq Cp^{3n/4}2^n.
\]

3. Assume \( s_0 = s_1 = 0. \) We have

\[
\delta_n(m) \leq 2 + p^{-3}\delta_n(m/p^2) + 2p + p^2\delta''(m/p^2).
\]

After at most \( k = (1 + 3n)/5 \) steps, we reduce to the \( s_0 + s_1 \geq 1 \) case (and then we further reduce to \( s_0 + s_1 \geq 2 \) case) and in conclusion we have

\[
\delta_n(m) \leq Cp^{6n/5}2^n. \quad \square
\]

**Appendix B. Decay in the n = 3 superspecial case**

In this section, we treat supersingular points which are not superspecial. The notion of decaying rapidly is the same as in Definition 5.1.2 with \( A \) being the multiplicity of Hasse invariant (although \( A \) is used for some other matrices).

**Theorem B.0.1** (Decay lemma in the Siegel case). At a supersingular point, there exists a rank 3 submodule of special endomorphisms which decays rapidly.

We first compute \( V^{F=1}. \) This is a free \( \mathbb{Z}_p \)-module of rank 5. For convenience, let \( \lambda \) denote a unit in \( W(\mathbb{F}_p) \) such that \( \sigma(\lambda) = -\lambda. \) Clearly, \( w_3 := \lambda e_3 \) and \( w_5 := \lambda e_5 \) generate a saturated submodule of \( V^{F=1}. \) A computation yields that \( w_1 := p\lambda e_4 + d\lambda e_3 - e_4, w_2 := \lambda pe_2 + \lambda de_3 - \lambda e_4 \) and \( w_4 := p\lambda e_1 - d\lambda e_4 \) are also fixed by \( F. \) For a non-superspecial point, we have \( d \notin \mathbb{F}_p^2. \) Following Kisin's constructions, Frobenius on the crystal (in the \( F \)-invariant basis described just above) has the form

\[
Id + Ax/p + B(xy + z^2/2)/p + Cy + Dz,
\]

where
To lighten notation, let

\[ A = \begin{bmatrix} 0 & \frac{-d}{\lambda} & \frac{-1}{\lambda} & \frac{d^2}{\lambda^2} & 0 \\ -d\lambda & 0 & \frac{-1}{\lambda} & \frac{-d^2\lambda}{2} & 0 \\ d^2 & \frac{d^2}{\lambda} & d & 0 & 0 \\ -1 & \frac{1}{\lambda} & 0 & -d & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \]

\[ B = \begin{bmatrix} \frac{1}{\lambda} & \frac{-1}{\lambda} & 0 & \frac{d}{\lambda} & 0 \\ \frac{d\lambda}{2} & \frac{d\lambda}{2} & 0 & \frac{-d^2 \lambda}{2} & 0 \\ -d & \frac{d}{\lambda} & 0 & -d^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \]

\[ C = \begin{bmatrix} 0 & 0 & 0 & \frac{-1}{\lambda} & 0 \\ 0 & 0 & 0 & \frac{d}{2\lambda} & 0 \\ -1 & \frac{1}{\lambda} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \]

and

\[ D = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{-1}{\lambda} \\ 0 & 0 & 0 & 0 & \frac{d}{2\lambda} \\ 0 & 0 & 0 & 0 & 0 \\ -1 & \frac{1}{\lambda} & 0 & -d & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \]

To lighten notation, let \( G \) denote the matrix \( pDD^{(1)} \).

\[ G = \begin{bmatrix} \frac{1}{\lambda} & \frac{1}{\lambda} & 0 & \frac{d^{(1)}}{\lambda} & 0 \\ \frac{1}{\lambda} & \frac{1}{\lambda} & 0 & \frac{d^{(1)}}{\lambda} & 0 \\ -d & \frac{d}{\lambda} & 0 & -dd^{(1)} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \]

As in the case of inert Hilbert modular surfaces, given any positive \( n \), the coefficient of \( 1/p^n \) which has the smallest powers of \( (x, y, z) \) arise from powers of \( A, B \) and \( D \) with \( C \) contributing only larger order terms. We will therefore ignore \( C \) while proving our decay results.

Again, consider a formal curve \( \text{Spf} \mathbb{F}[[t]] \) and a principally polarized abelian surface over \( \mathbb{F}[[t]] \) which is generically ordinary and has supersingular (but not superspecial) reduction. This corresponds to a map \( \mathbb{F}[[x, y, z]] \rightarrow \mathbb{F}[[t]] \), and we denote by \( x(t), y(t), z(t) \) the images of \( x, y, z \). Without loss of generality, we assume that \( x(t) = \ell^a \) and \( z(t) \) has \( t \)-adic valuation \( c \), and that the leading coefficient of \( z(t)^2/2 \) is \( \alpha \). We will need the following Lemmas.

**Lemma B.0.2.** Let \( E = A + \alpha B \) and suppose that \( n \geq 0 \).

1. Let \( w \) be any vector in the \( \mathbb{F}_p \)-span of \( w_1, w_2, w_4 \) mod \( p \). Then \( w \) is not in the kernel of \( p^{n+1}BB^{(1)} \ldots B^{(n)} \mod p \).

2. Suppose that \( n \geq 1 \). If \( \alpha^{(1)} - \frac{d}{\lambda} + \frac{d^2}{\lambda^2} = 0 \), then \( E \ldots E^{(n)} = 0 \) when \( n \geq 2 \). Otherwise, the 4th row of \( E \ldots E^{(n)} \) is a multiple of \( R^{(n-1)} \) by a \( p \)-adic unit. Here \( R \) is the 4th row of \( EE^{(1)} \) and

\[ R = \begin{bmatrix} d + d^{(1)} - \alpha^{(1)} & \frac{d - d^{(1)} - \alpha^{(1)}}{\lambda} & 1 & -\alpha^{(1)}d^{(1)} & dd^{(1)} \end{bmatrix}. \]
Proof. (1) An easy inductive argument shows that $p^{n+1}B \ldots B^{(n)}$ equals

$$
\begin{pmatrix}
\frac{1}{2} & \frac{(-1)^{n+1}}{2} & 0 & d^{(n)} & 0 \\
\frac{1}{2} & \frac{(-1)^{n+1}}{2} & 0 & \frac{d^{(n)}}{2} & 0 \\
-d & \frac{-d}{\lambda} & 0 & -dd^{(n)} & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}.
$$

The kernel of the above matrix intersected with the $\mathbb{F}$-span of $w_1, w_2, w_4 \mod p$ is clearly seen to equal the span of $w_1 + (-1)^n\lambda w_2, -d^{(n)}w_1 + w_4$. As $d \not\in \mathbb{F}_p$, there is no non-trivial $\mathbb{F}_p$ linear relation between $1, \lambda^{-1}$ and $d^{(n)}$. It follows that the above span contains no non-zero $\mathbb{F}_p$-combinations of $w_1, w_2, w_4$ as required.

(2) The proof goes by induction. If the claim were true for the last row of $A \ldots A^{(n)}$, it suffices to prove that $R^{(n-1)}A^{(n+1)}$ is a unit multiple of $R^{(n)}$. As $R^{(n-1)}A^{(n+1)} = [RA^{(2)}](n-1)$, it suffices to prove that $RA^{(2)}$ is a $p$-adic unit multiple of $R^{(1)}$. This can be checked by direct calculation.

For $E$ is similar.....Also, say that when $\alpha^{(1)} - d + d^{(2)} = 0$, the assertion can be checked by computing $EE^{(1)}E^{(2)} = 0$ directly.

We will need to consider the following cases:

B.1. Case 1: $a > 2c$.

Proof of decay lemma. We will use part (1) of Lemma B.0.2 to prove that every special endomorphism in the span of $w_1, w_2, w_4$ decays rapidly. As $a > 2c$, the $t$-adic valuation of $xy + z^2/2$ equals $2c$, which is strictly smaller than that of $x$. Further, the $t$-adic valuation of $z^{1+p}$ equals $(p+1)c$, which is strictly greater than the valuation of $xy + z^2/2$. Therefore, the term in the coefficient of $1/p^{n+1}$ with minimum $t$-adic valuation equals $B \ldots B^{(n)}$. That the $\mathbb{Z}_p$-span span of $w_1, w_2, w_4$ decays rapidly follows from (1) of Lemma B.0.2.

B.2. Case 2: $a < 2c$.

Proof of decay lemma. It suffices to prove that every vector in the span of $w_1, w_2, w_3$ decays rapidly. As $a < 2c$, the term in the coefficient of $1/p^{n+1}$ with minimum $t$-adic valuation equals $A \ldots A^{(n)}$. It suffices to prove that the kernel of $p^{n+1}A \ldots A^{(n)} \mod p$ contains no non-zero $\mathbb{F}_p$ linear combinations of $w_1, w_2, w_3 \mod p$. Let $w$ be some $\mathbb{F}_p$-linear combination of these three vectors, which were in the kernel of $p^{n+1}A \ldots A^{(n)}$.

By Lemma B.0.2 the 4th row of $p^{n+1}A \ldots A^{(n)}$ is $R^{(n-1)}$. Therefore, $R^{(n-1)} \cdot w = 0$. As $w$ is Frobenius-invariant, this is equivalent to $R \cdot w = 0$. A direct computation shows that

$$
R = [d + d^{(1)} \frac{d - d^{(1)}}{\lambda} 1 dd^{(1)}].
$$

The existence of $w$ implies that there exist $a, b, c \in \mathbb{F}_p$ such that $a(d + d^{(1)}) + b(d - d^{(1)})/\lambda + c = 0$. If either $a = 0$ or $b = 0$, it follows that either $d + d^{(1)} \in \mathbb{F}_p$ or $(d - d^{(1)})/\lambda \in \mathbb{F}_p$. Either case would imply that $d = d^{(2)}$, which is a contradiction as we have assumed that $d \not\in \mathbb{F}_p$. Therefore, we assume that $a = -1$. Therefore, we have

$$
d + d^{(1)} = b(d - d^{(1)})/\lambda + c,
$$

and therefore

$$
d^{(1)} + d^{(2)} = b(d^{(2)} - d^{(1)})/\lambda + b.
$$

Subtracting the two equations yields

$$
\text{[Equation]}.
$$
\[ d - d^{(2)} = b(d - d^{(2)})/\lambda. \]

This is a contradiction, as \( \lambda \notin \mathbb{F}_p \). Therefore, such \( w \) could not have been in the kernel of \( p^{n+1}A \ldots A^{(n)} \mod p \), whence it follows that every vector in the \( \mathbb{Z}_p \) span of \( w_1, w_2, w_3 \) decays rapidly.

\[ \square \]

B.3. Case 3: \( a = 2c \).

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

**Proof.** We will prove that every vector in the \( \mathbb{Z}_p \)-span of \( w_1, w_3, w_5 \) decays rapidly where \( i \) is either 1 or 2. The conditions imposed on \( a, c \) imply that the term with minimal \( t \)-adic valuation in the coefficient of \( 1/p^{n+1} \) of \( F_\infty \) is \( EE^{(1)} \ldots E^{(n)}/2c(1+p^{n+1}) + EE^{(1)} \ldots E^{(n-1)}D^{(n)}2c(1+p^{n+1}+cp^n) \).

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 denote by \( u' \) the Frobenius twist of \( u \). We claim that both \( d + d' - \alpha' \) and \( d - d' - \alpha' \) cannot be elements of \( \mathbb{F}_p \). Indeed, the first element being in \( \mathbb{F}_p \) implies that \( d' + d'' - \alpha'' = d + d' - \alpha' \), consequently \( \alpha' - \alpha'' = d - d'' \). Similarly, \( d - d' - \alpha' \) being an element of \( \mathbb{F}_p \) implies that \( \alpha' + \alpha'' = d - d'' \). Therefore, both elements being in \( \mathbb{F}_p \) implies that \( \alpha = 0 \), which is a contradiction.

Without loss of generality, we will assume that \( d + d' - \alpha' \notin \mathbb{F}_p \), and prove that every vector in the span of \( w_1, w_3, w_5 \) decays rapidly (if \( d - d' - \alpha' \notin \mathbb{F}_p \), then an identical argument would yield that every vector in the span of \( w_2, w_3, w_5 \) decays rapidly).

We first prove that every vector in the \( \mathbb{Z}_p \) span of \( w_1, w_3 \) decays rapidly. Indeed, let \( w \) be any vector (which is not a multiple of \( p \)). In order to prove that \( w \) decays rapidly, it suffices to prove that \( w \mod p \) is not in the kernel of \( EE^{(1)} \ldots E^{(n)} \mod p \). However, as the 4th row of this matrix is a unit-multiple of \( R^{(n-1)} \) (\( R \) is as in (3) of Lemma [B.0.2]), it suffices to prove that \( w \mod p \) is not in the kernel of the \( 1 \times 5 \) matrix \( R^{(n-1)} \mod p \). As \( w \mod p \) is \( \mathbb{F}_p \)-rational, this is equivalent to asking that \( w \mod p \) is not in the kernel of \( R \mod p \).

This follows from the fact that \( d + d' - \alpha' \notin \mathbb{F}_p \).

We now show that \( w_5 \) decays rapidly. Indeed, the last column of \( EE^{(1)} \ldots E^{(n-1)}D^{(n)} \) has \( R^{(n-2)} \cdot v^{(n)} \) as its fourth entry, where \( v \) is the last column of \( D \). It suffices to prove that \( R^{(n-2)} \cdot v^{(n)} \neq 0 \mod p \), equivalently \( R \cdot v^{(2)} \neq 0 \mod p \). A direct computation shows that this element equals \( \alpha' - d + d'' \), which we have assumed is 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 \( 1/p^{n+1} \) in \( F_\infty w_5 \). However, the former equals \( 2c(1+p^{n+1}) \) and the latter equals \( 2c(1+p^{n-1}+cp^n) \). These two quantities are clearly never equal, thereby establishing the required decay.

\[ \square \]

**Subcase 2:** \( \alpha^{(1)} - d + d^{(2)} = 0. \) We will establish the required decay by considering the fourth row of \( F_\infty \). As mentioned above, we omit \( C \) and powers of \( y \) in this analysis, as there are no negative powers of \( p \) in the entries of \( C \).

We need the following lemma:

**Lemma B.3.1.** Consider all products of the form \( W_0W_1W_2 \ldots W_n \), where \( W_i \) is the \( i \)th Frobenius twist of \( E, B \) or \( D \). The only products which have a non-zero fourth row are those with the following properties:

(1) \( W_0 = E \).
(2) Suppose that \( W_i, W_j \neq D \) but \( W_{i+1} \ldots W_{j-1} = D \) 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 \( D \) 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 \( W_j = E \) is \( j = n \).

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

Finally, this product has nonzero last column if and only if \( W_n = D \) and the number of occurrences of \( D \) 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

\[
E \prod_{i=1}^m B^{(i)} \prod_{j=k+1}^{m+2e} D^{(j)} E^{(m+2e+1)}
\]

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

An easy inductive argument shows that the product \( \prod_{i=1}^m B^{(i)} \prod_{j=k+1}^{m+2e} D^{(j)} \) equals

\[
\begin{pmatrix}
\frac{1}{2} & -(\frac{-1}{2})^{m+1} & 0 & \frac{d(2e+m)}{2} \\
\frac{-1}{2} & -(\frac{-1}{2})^m & 0 & \frac{-d(2e+m)}{2} \\
-d^{(1)} & -(\frac{-1}{2})^{m+1}d^{(1)} & 0 & -d(1)d^{(2e+m)} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

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

\[
\begin{pmatrix}
d(2e+m+1) + d^{(2e+m+2)} & -(\frac{-1}{2})^{m+1}d(2e+m+1)d^{(2e+m+2)} - d^{(2e+m+2)} \\
\frac{d(2e+m+1)}{\chi} & 1 & d(2e+m+1)d^{(2e+m+2)} & 0
\end{pmatrix}.
\]

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

We record the fourth row of the product in (B.3.1) for future use:

Lemma B.3.2. The fourth row of the product \( E \prod_{i=1}^m B^{(i)} \prod_{j=k+1}^{m+2e} D^{(j)} E^{(m+2e+1)} \) equals

\[
\begin{pmatrix}
d(2e+m+1) + d^{(2e+m+2)} & -(\frac{-1}{2})^{m+1}d(2e+m+1)d^{(2e+m+2)} - d^{(2e+m+2)} \\
\frac{d(2e+m+1)}{\chi} & 1 & d(2e+m+1)d^{(2e+m+2)} & 0
\end{pmatrix}.
\]

Define \( z_1 = z^2/2 + xy - \alpha x \). As this function is the local equation cutting out the non-ordinary locus at our point (note that \( \alpha = d^{(-1)} - d^{(1)} \)), \( z_1(t) \neq 0 \). Let \( z_1(t) = \eta t^M + t^{M+1}(z_2(t)) \) where \( \eta \in \overline{\mathbb{F}_p} \).

We break this final case into two cases.

Subsubcase 1: \( c(2p^{2e} - p^{2e-1} + 1) < M < c(2p^{2e+2} - p^{2e+1} + 1) \). In this case, we will prove that every vector in the span of \( w_1, w_2, w_3 \) decays rapidly. Therefore, it suffices for us to work with the top-left \( 4 \times 4 \) block of \( F_\infty \).
Lemma B.3.3. The term with minimal $t$-adic valuation in the coefficient of $1/p^n$ in the fourth row of (the top-left block of) $F_\infty$ is

$$EB^{(1)} \ldots B^{(n-e-1)}D^{(n-e)} \ldots D^{(n+e-1)}E^{(n+e)}$$

Proof. Note that $F_\infty$ is composed of sums of products as in Lemma [B.3.1] where each $W_i$ is multiplied by:

- $x(i)^{(i)} = i^{2ip^d}$ if $W_i = E$.
- $z(i)^{(i)} = \eta(i)^{(i)}t^{Mp^i} + \ldots$ if $W_i = B$.
- $z(i)^{(i)} = \beta(t)^{(i)}t^{e}$ if $W_i = D$, where $\beta$ is the leading coefficient in $z = \beta t^e + \cdots$.

Consider products as in Lemma [B.3.1]. As we are looking for matrices where the first four columns are not all zero, it follows that the number of occurrences of $D$ must be even. Therefore, consider a product with $n_1$ occurrences of either $E$ or $B$, and $2n_2$ occurrences of $D$. 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 $E$. Indeed, the $t$-adic valuation is minimized when $W_i = E$ for as many $i$ as possible, and Lemma [B.3.1] implies that this happens when the first and last matrices are both $E$. As the $t$-adic valuation of $z_1(t)$ is strictly greater than that of $z(t)$, it follows that products of the form $EB^{(1)} \ldots B^{(n_2-1)}D^{n_1-1} \ldots D^{(n_1+2n_2-2)}E^{(n_1+2n_2-1)}$ contain the term with minimal $t$-adic valuation.

As in the case of split Hilbert modular surfaces, 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 the Decay lemma in this case. We will prove that the span of $w_1, w_2, w_3$ decays rapidly. Let $R$ denote the mod $p$ reduction of the row detailed in Lemma [B.3.2]. It suffices to show that there is no $\mathbb{F}_p$-linear combination of the first three entries of $R$ which evaluates to zero. This is tantamount to proving that the elements $d + d^{(1)}, \frac{d-d^{(1)}}{x}$ and $1$ are $\mathbb{F}_p$-linearly independent. But this has already been established in Case 2 of the decay lemma. The result follows. $\square$

Subsubcase 2: $c(2p^{2e} - p^{2e-1} + 1) = M$.

Lemma B.3.4. (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 block of) $F_\infty$. They are:

- $EB^{(1)} \ldots B^{(n-e-1)}D^{(n-e)} \ldots D^{(n+e)}\eta^{(1+p+\ldots p^{n-e-1})}\beta p^{n-e+\ldots+p^{n+e-1}}$.
- $EB^{(1)} \ldots B^{(n-e)}D^{(n-e+1)} \ldots D^{(n+e-2)}\eta^{(1+p+\ldots p^{n})}\beta p^{n-e+1+\ldots+p^{n+e-2}}$.

(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 $F_\infty$ is

$$EB^{(1)} \ldots B^{(n-e-1)}D^{(n-e)} \ldots D^{(n+e)}$$

Proof. The proof goes along the same lines as that of Lemma [B.3.3] and so we will not spell out the details. $\square$

Proof of the final case of the Decay lemma. We will show that there exists a dimension 2 vector space $W \subset \mathbb{Z}_p(w_1, w_2, w_3)$ such that $W \oplus \mathbb{Z}_p w_5$ decays rapidly.

The term with minimal $t$-adic valuation in the coefficient of $1/p^{n+1}$ of the top-left block of $F_\infty$ is the sum of the two matrices detailed in (1) of Lemma [B.3.4]. The $t$-adic valuation of this sum is $c(2 + p^{n-e} + \ldots p^{n-e-1} + 2p^{n+e}) + M(p + \ldots p^{n-e-1})$.

On the other hand, the term with minimal $t$-adic valuation in the coefficient of $1/p^{n+1}$ of the last column of $F_\infty$ is the last column of the matrix detailed in (2) of Lemma [B.3.4]. It is easy to
We now prove that depending on the values of \( \mu \) and \( \nu \), either the span of \( w_1, w_2, w_3 \) or \( w_2, w_3 \) decays rapidly. To further lighten notation, let \( \delta_1 = (\mu + \nu) d(n+e) \) and \( \delta_2 = \mu d(n+e+1) + \nu d(n+e-1) \).

(1) Suppose \( \nu + \mu = 0 \). We will show that the span of \( w_1, w_2 \) decays rapidly. It suffices to prove that there is no non-trivial \( \mathbb{F}_p \) relation between \( \alpha_2 = \mu (d(n+e-1) + d(n+e+1)) \) and \( \alpha_3 = \frac{\lambda}{\mu} (d(n+e-1) - d(n+e+1)) \). However, this follows directly from the facts that \( \lambda \notin \mathbb{F}_p \) and that \( d(n+e+1) \neq d(n+e-1) \).

(2) Suppose that \( \delta_1 \neq \pm \delta_2 \). We will show that either the span of \( w_1, w_2 \) or \( w_1, w_3 \) decays rapidly. Indeed, the former happens when \( \alpha_1 = \delta_1 + \delta_2 \) and \( \alpha_2 = \frac{\lambda}{\mu} (\delta_1 - \delta_2) \) are not \( \mathbb{F}_p \) multiples of each other. Therefore, suppose that they were. Then we have the equations (where \( l \in \mathbb{F}_p \))

\[
\delta_1 + \delta_2 = l \left( \frac{-1}{\lambda} (\delta_1 - \delta_2) \right)
\]

Note that this yields that \( \delta_2 / \delta_1 \in \mathbb{F}_p^\times \). We will prove \( w_1, w_3 \) decays rapidly. If not, then \( \alpha_1 = s \alpha_3 \), where \( s \in \mathbb{F}_p \). That is, \( \delta_1 + \delta_2 = s \delta_1 / d(n+e) \). Equivalently, \( \delta_2 / \delta_1 = (s - d(n+e)) / d(n+e) \). As \( d \notin \mathbb{F}_p^\times \), it follows that \( \delta_2 / \delta_1 \notin \mathbb{F}_p^\times \), which is a contradiction.

(3) Suppose that \( \delta_1 = \delta_2 \). We will show that the span \( w_1, w_3 \) decays rapidly, by showing that \( \alpha_1, \alpha_3 \) are \( \mathbb{F}_p \) linearly independent. Indeed, \( \alpha_1 = 2(\nu + \mu) d(n+e) \), and \( \alpha_3 = \nu + \mu \). As \( d \notin \mathbb{F}_p^\times \), 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 the span of \( w_2, w_3 \) decays rapidly.

\[ \square \]

References


