PICARD RANKS OF K3 SURFACES OVER FUNCTION FIELDS AND THE HECKE ORBIT CONJECTURE

DAVESH MAULIK, ANANTH N. SHANKAR, AND YUNQING TANG

Abstract. Let $\mathscr{X} \to C$ be a non-isotrivial and generically ordinary family of K3 surfaces over a proper curve C in characteristic $p \geq 5$. We prove that the geometric Picard rank jumps at infinitely many closed points of C. More generally, suppose that we are given the canonical model of a Shimura variety $\mathcal{S}$ of orthogonal type, associated to a lattice of signature $(b, 2)$ that is self-dual at p. We prove that any generically ordinary proper curve C in $\mathcal{S}_{\overline{\mathbb{F}}_p}$ intersects *special divisors* of $\mathcal{S}_{\overline{\mathbb{F}}_p}$ at infinitely many points. As an application, we prove the ordinary Hecke orbit conjecture of Chai–Oort in this setting; that is, we show that ordinary points in $\mathcal{S}_{\overline{\mathbb{F}}_p}$ have Zariski-dense Hecke orbits. We also deduce the ordinary Hecke orbit conjecture for certain families of unitary Shimura varieties.

Contents

1. Introduction

1.1. **Families of K3 surfaces.** Given a family of complex K3 surfaces, it is a well-known fact that the Picard ranks of the fibers will jump at infinitely many special points, so long as the family is not isotrivial. More precisely, let us recall the following Hodge-theoretic result, due to Green [Voi02] and Oguiso [Ogu03]. Let Δ be the unit disc in $\mathbb{C}$, and let $\mathscr{X} \to \Delta$ be a non-isotrivial family of (compact) K3 surfaces. If ρ denotes the minimal Picard rank of $\mathscr{X}_s$, $s \in \Delta$, then the set of points $t \in \Delta$ for which the Picard rank of $\mathscr{X}_t$ is greater than ρ is a countable, dense subset of Δ. In particular, there are infinitely many such points.

In positive characteristic, this question is more subtle. Suppose we are given $\mathscr{X} \to C$, where $C/\overline{\mathbb{F}}_p$ is a curve and $\mathscr{X}$ is a non-isotrivial family of K3 surfaces. It is now no longer the case that the Picard rank has to jump at infinitely many points of C. For example, there exist families where every fiber X_t is a supersingular K3 surface, in which case the rank is always 22. By studying families of non-ordinary Kummer surfaces, one can produce families where the Picard rank jumps, but only at finitely many points (see 1.4 for an example).

In both of these examples, the generic fiber is not *ordinary*. The first main result of this paper shows that, under additional hypotheses, if the generic fiber is ordinary, then there will be infinitely many points where the Picard rank jumps. That is, we show the following:

Theorem 1.1. *Let $C/\overline{\mathbb{F}}_p$ denote a smooth proper curve where $p \geq 5$ is a prime number, and let $\mathscr{X} \to C$ denote a generically ordinary non-isotrivial family of K3 surfaces. Suppose that the discriminant*[1] *of the generic Picard lattice is prime to p. Then there exist infinitely many points $c \in C(\overline{\mathbb{F}}_p)$ such that the Picard rank of $\mathscr{X}_c$ is greater than the generic Picard rank of $\mathscr{X}$.*

Broadly speaking, Theorem 1.1 is proved by studying moduli spaces of K3 surfaces, viewed as (canonical integral models of) GSpin Shimura varieties $\mathcal{S}$ associated to quadratic $\mathbb{Z}$-lattices (L, Q) having signature $(b, 2)$ with $b \leq 19$. These Shimura varieties admit families of "special divisors", which are themselves GSpin Shimura varieties associated to sublattices of (L, Q) having signature $(b-1, 2)$, whose points parameterize K3 surfaces with Picard rank greater than those parameterized by "generic points" of the ambient Shimura variety.

The notion of special divisors makes sense in the more general setting of GSpin Shimura varieties $\mathcal{S}$ associated to quadratic lattices (L, Q) having signature $(b, 2)$ for all positive integers b. For every positive integer m, there exists a divisor $\mathcal{Z}(m) \subset \mathcal{S}$ which, if not empty, is also (the integral model of) a GSpin Shimura variety. We prove the following theorem which directly implies Theorem 1.1.

Theorem 1.2. *Let $\mathcal{S}$ denote the canonical integral model over $\mathbb{Z}_p$ of the GSpin Shimura variety associated to a quadratic $\mathbb{Z}$-lattice (L, Q) of signature $(b, 2)$, such that p does not divide the discriminant of (L, Q). Assume that $b \geq 3, p \geq 5$. Let C be an irreducible smooth proper curve with a finite morphism $C \to \mathcal{S}_{\bar{\mathbb{F}}_p}$ such that the generic point of C is ordinary and that the image of C does not lie in any special divisors $Z(m) := \mathcal{Z}(m)_{\bar{\mathbb{F}}_p}, m \in \mathbb{Z}_{>0}$. Then there exist infinitely many $\bar{\mathbb{F}}_p$-points on C which lie in $\cup_{m \in \mathbb{N}, p \nmid m} Z(m)$.*

In the case of $\mathcal{S}$ being a Hilbert modular surface or a Siegel modular threefold (with $b = 2, 3$ respectively), Theorem 1.2 follows from our earlier paper [MST], but the general setting considered here requires additional techniques.

1.2. **The Hecke orbit conjecture.** The second goal of this paper is to apply Theorem 1.2 to the study of Hecke orbits in characteristic p. In general, Shimura varieties are naturally equipped with a set of correspondences, known as Hecke correspondences. Roughly speaking, these Hecke correspondences permute[2] the set of special divisors. In characteristic zero, the dynamics of Hecke correspondences are well-understood. For example, work of Clozel–Oh–Ullmo [COU01] proves that the Hecke orbit of a point equidistributes in the analytic topology. However, in characteristic p, the behavior of the Hecke orbit of a point is still far from understood.

The first result along these lines is due to Chai [Cha95], who proved that the prime-to-p Hecke orbit of an ordinary point is Zariski dense in $\mathcal{A}_{g,\mathbb{F}_p}$, the moduli space of principally polarized abelian varieties over $\mathbb{F}_p$. Guided by this, Chai and Oort have the following more general conjecture for arbitrary Shimura varieties.

Conjecture 1.3 (Chai–Oort)**.** *Let $\mathcal{S}$ denote the canonical integral model of a Shimura variety of Hodge type (with hyperspecial level) and let $\mathcal{S}_{\overline{\mathbb{F}}_p}$ denotes its special fiber.*[3] *Then the prime-to-p Hecke orbit of a μ-ordinary point is Zariski dense in $\mathcal{S}_{\overline{\mathbb{F}}_p}$.*[4]

There is a further generalization to non-ordinary points, which we do not discuss here. For more about the Hecke orbit conjecture and generalizations, see [Cha03], [Cha05],[Cha06], [CO06], [CO09]

[1] Note that the Picard lattice of a K3 surface is equipped with a non-degenerate quadratic form arising from the intersection pairing. The discriminant of the Picard lattice is defined to be the discriminant of this quadratic form.

[2] The special divisors $Z(d)$ and $Z(m^2d)$ are in the same Hecke orbit.

[3] The conjecture was made in the PEL case, but is expect to hold for Hodge type Shimura varieties too, which includes the case of PEL Shimura varieties.

[4] Being μ-ordinary means that this point lies in the open Newton stratum of $\mathcal{S}_{\bar{\mathbb{F}}_p}$ and it means ordinary if the ordinary locus in $\mathcal{S}_{\bar{\mathbb{F}}_p}$ is nonempty, which will be the case for us in the rest of the paper.

and [CO19]. Using Theorem 1.2 as our main input, we establish the ordinary Hecke orbit conjecture for GSpin Shimura varieties, as well as certain unitary Shimura varieties.

Theorem 1.4. *Let $\mathcal{S}_{\mathbb{F}_p}$ denote the mod p (where $p \geq 5$) fiber of the canonical integral Shimura variety associated to one of the following data:*

The orthogonal case. *A quadratic $\mathbb{Z}$-lattice with signature $(b, 2)$ having discriminant prime to p with the associated Shimura variety defined in §2.1.*

The unitary case. *An imaginary quadratic field K split at p, and an $\mathcal{O}_K$-Hermitian lattice having signature $(n, 1)$, with discriminant prime to p with the associated Shimura variety defined in* [RSZ20, §3, §4.1].[5]

Then the prime-to-p Hecke orbit of an ordinary point is dense in $\mathcal{S}_{\mathbb{F}_p}$.

As far as we know, this result is the first of its kind towards settling the Hecke orbit conjecture in the setting of orthogonal Shimura varieties.

1.3. Outline of the proof of Theorem 1.2.

There are two broad steps in our proof:

(1) We use Borcherds theory to compute the asymptotic of the intersection numbers $(C.Z(m))$ as $m \to \infty$;

(2) We then prove that given finitely many points $P_1, \cdots, P_n \in C(\overline{\mathbb{F}}_p)$, the local contributions have the property $\sum_{p \nmid m,\ 1 \leq m \leq X} \sum_{i=1}^{n} i_{P_i}(C.Z(m)) < \sum_{p \nmid m, 1 \leq m \leq X} C.Z(m)$ for large enough $X \in \mathbb{Z}$.

These two steps together prove that as $X \to \infty$, more and more points $P \in C(\overline{\mathbb{F}}_p)$ must contribute to the intersection $C.(\sum_{p \nmid m, 1 \leq m \leq X} Z(m))$, thereby yielding Theorem 1.2.

The second step involves both local and global techniques. We use the moduli interpretation of the special divisors $Z(m)$ to express the local contribution $i_P(C.Z(m))$ in terms of a lattice point count in an *infinite* nested sequence of lattices. This is another way in which the characteristic-p nature of this work complicates matters – the analogous expression in the characteristic 0 setting involved a lattice point count in a *finite* nested sequence of lattices, which makes matters far more tractable. The most technical part of the paper deals with controlling the main term of $i_P(C.Z(m))$ for supersingular points, which we do over Sections 4,5 and 6. This requires using Ogus's and Kisin's work to explicitly understand the equicharacteristic deformation theory of special endomorphisms at supersingular points in terms of crystalline theory.

One of the difficulties of this result compared to [MST] is that, for $b \leq 3$, it is relatively easy to bound the error terms of $i_P(C.Z(m))$. However, in general, high levels of tangency between C and special divisors could in principle cause this term to grow uncontrollably. In order to control this, we use a global argument which first appeared in [SSTT]. Note that $i_P(C.Z(m))$ is necessarily bounded above by the global intersection number $(C.Z(m))$. We use the fact that the global bound holds for *every positive integer m* (representable by (L, Q)) in order to obtain a sufficient control on the error term of the local contribution $i_P(C.Z(m))$ *on average*, as we average over all positive integers m. More precisely, we prove that if the error term of $i_P(C.Z(m))$ is too close to the global intersection number for several values of m, then there must exist a positive integer m_0 for which the local intersection number $i_P(C.Z(m_0))$ is greater than the global intersection number $(C.Z(m_0))$, which is a contradiction.

It is crucial to our proof of the local bound that the curve C is an algebraic curve. Given a formal curve having the form $\operatorname{Spf} \overline{\mathbb{F}}_p[[t]] \subset \mathcal{S}_{\mathbb{F}_p}$ with closed point P, the term $i_P(\operatorname{Spf} \overline{\mathbb{F}}_p[[t]].Z(m))$ is well defined. It is easy to construct examples of formal curves that have the property that $i_P(\operatorname{Spf} \overline{\mathbb{F}}_p[[t]].Z(m_i))$ grows exponentially fast for appropriate sequences of integers m_i, as we discuss

[5] Here we only work with the special case when the CM field is an imaginary quadratic field; in the special case when the polarization is principle, see also [KR14, §2] or [SSTT, §9.3].

in Section 3.5. Indeed, the growth rate of $i_P(\operatorname{Spf} \overline{\mathbb{F}}_p[[t]].Z(m))$ can be used as a necessary criterion to determine whether or not a formal curve contained in $\mathcal{S}$ is algebraizable.

1.4. **Contributions from supersingular points.** As stated in the outline, the most technical part of our paper is dealing with supersingular points. The main difficulty is caused by the fact that the local contribution $i_P(C.Z(m))$ from a supersingular point $P \in C(\overline{\mathbb{F}}_p)$ has the same order of magnitude as the global intersection number $(C.Z(m))$ as $m \to \infty$; Indeed, the global intersection number $(C.Z(m))$ can be expressed in terms of the m-th Fourier coefficient of a non-cuspidal modular form of weight $1 + b/2$, whose Eisenstein part is well understood (Lemma 7.5 and Theorem 7.4). The main term of the local contribution $i_P(C.Z(m))$ at a supersingular point P is controlled by the m-th Fourier coefficients of the theta series associated to a nested sequence of positive definite lattices each having rank $b + 2$, and is therefore also asymptotic to the m-th Fourier coefficient of an Eisenstein series of weight $1 + b/2$ (see §7.14).

Therefore, a more refined understanding of the constants involved in the global intersection number and the supersingular contribution is needed to prove our theorem. In fact, this is precisely what goes wrong when C is no longer generically ordinary. There are examples when finitely many supersingular points can indeed conspire to fully make up the entire global intersection number. We illustrate this with the following example.

Consider the setting of $X \to C$, where $C/\overline{\mathbb{F}}_p$ is a curve and X is a non-isotrivial family of Kummer surfaces. Indeed, let $E \to C$ denote a non-isotrivial family of elliptic curves, and let $E_0 \to C$ denote a supersingular elliptic curve pulled back to C. Consider the family $K(E \times_C E_0) \to C$, where $K(E \times_C E_0)$ denotes the Kummer surface associated to the abelian surface $E \times_C E_0$. The set of points $c \in C(\overline{\mathbb{F}}_p)$ such that the Picard rank of $K(E \times E_0)_c$ is greater than the generic Picard rank of $K(E \times E_0)$ is precisely the set of $c \in C(\overline{\mathbb{F}}_p)$ such that the fiber of E at c is supersingular, and therefore the total global intersection number is made up from the local contributions from these finitely many supersingular points.

1.5. **Outline of proof of Theorem 1.4.** We now survey the proof of the Hecke orbit conjecture, using Theorem 1.2 as input. We will focus on the orthogonal case, since the unitary case follows by a similar argument.

Let us first observe that Chai's approach for $\mathcal{A}_g$ does not easily extend to this case. Chai's argument involves several steps, some of which generalize to the case of all Shimura varieties of Hodge type, but there are many ideas in Chai's work which don't generalize to our setting. Indeed, a key step in his paper is the so-called Hilbert trick which states that every $\overline{\mathbb{F}}_p$-valued point of $\mathcal{A}_g$ is contained in a positive-dimensional Shimura subvariety of $\mathcal{A}_g$, namely a Hilbert modular variety. Unfortunately, this fact does not hold for most Shimura varieties. The case of Hilbert modular varieties is more tractable than $\mathcal{A}_g$ because the geometrically simple factors of the associated reductive groups have rank 1.

Instead, our idea is to use an inductive argument on the dimension of $\mathcal{S}_{\overline{\mathbb{F}}_p}$. Our argument consists of the following steps:

(1) The setting of Shimura varieties associated to quadratic lattices having signature $(1, 2)$ follows because these Shimura varieties are one-dimensional, and it is well known that the Hecke orbit of an ordinary point is an infinite set. Now, inductively assume that the Hecke orbit conjecture has been proved for all Shimura varieties associated to quadratic lattices having signature $(b - 1, 2)$, with discriminant relatively prime to p, where $b > 1$.

(2) Let $Z \subset \mathcal{S}_{\overline{\mathbb{F}}_p}$ denote a generically ordinary Hecke stable subvariety, where $\mathcal{S}$ is the canonical integral model of a Shimura variety associated to a quadratic lattice having signature $(b, 2)$ with discriminant relatively prime to p. Such a subvariety Z necessarily has to be positive dimensional, as the Hecke orbit of an ordinary point is necessarily infinite.

(3) Suppose that Z contains a proper curve C that is generically ordinary. Then, Theorem 1.2 implies that C intersects the union of special divisors $\bigcup_{p\nmid m} Z(m)$ at infinitely many points, and therefore at an ordinary point $x \in Z(m_0)$. The special divisor $Z(m_0)$ is the special fiber of a Shimura variety in its own right, associated to a quadratic lattice having signature $(b-1, 2)$ and prime-to-p discriminant (because $p \nmid m_0$), and so the prime-to-p Hecke orbit of x contains a Zariski-dense subset of $Z(m_0)$ by the inductive hypothesis. Therefore, $Z(m_0) \subset Z$, and the result follows from the fact that the Hecke orbit of any special divisor is Zariski dense in $\mathcal{S}_{\mathbb{F}_p}$.

(4) To deal with the case when Z might not contain a proper curve, we directly prove that any generically ordinary Hecke stable subvariety that intersects the boundary of the Baily–Borel compactification of $\mathcal{S}_{\mathbb{F}_p}$ (constructed in [MP19]) must be all of $\mathcal{S}_{\mathbb{F}_p}$.

In other words, even though a "generic" $\overline{\mathbb{F}}_p$-valued point of $\mathcal{S}$ may not lie in a smaller positive-dimensional Shimura variety, we are able to reduce to the case of a smaller Shimura variety using the intersection-theoretic input of Theorem 1.2.

1.6. **Previous work.** In addition to the ones mentioned above, we discuss here other related work in the literature.

Chai and Oort [CO06] proved Theorem 1.1 for Kummer surfaces associated to the product of two elliptic curves and Theorem 1.2 for $\mathcal{S} = \mathcal{A}_1 \times \mathcal{A}_1$ without the assumption that C is proper. The number field analogs of Theorem 1.1 and Theorem 1.2 have been proved in [SSTT], based on the previous work by Charles [Cha18] and [ST20] for $\mathcal{A}_1 \times \mathcal{A}_1$ and Hilbert modular surfaces respectively. For characteristic zero families, [Tay20] proved an equidistribution result on the the Noether–Lefschetz locus, which is a refinement of the theorem of Green.

For the results on Hecke orbits, Chai has also proved Conjecture 1.3 in the setting of Hilbert modular varieties, as well as for some PEL type C Shimura varieties. Building on work of Chai, the second named author [Sha] proved Conjecture 1.3 for the ordinary locus in Deligne's modèles étranges.

There is also a generalization of Conjecture 1.3 to $\bar{\mathbb{F}}_p$-points in other Newton strata (see [Cha06, Conj. 3.2]). In the case of $\mathcal{A}_g$, there is extensive work of Chai and Oort studying the properties of Newton strata (see their survey paper [CO19] and the references there); in combination with work of Yu, this gives the full Hecke orbit conjecture for $\mathcal{A}_g$ and Hilbert modular varieties (see for instance [Cha05] for the proofs). More recently, Zhou proved Conjecture 1.3 for (the μ-ordinary loci of) quaternionic Shimura varieties associated to quaternion algebras over some totally real fields ([Zho, Thm. 3.1.3, Rmk. 3.1.4]); and Xiao proved the generalized version for certain PEL Shimura varieties of type A and C and the points in those Newton strata which contain certain hypersymmetric points ([Xia, Thm. 7.1, Cors. 7.5, 7.6]).

1.7. **Organization of paper.** In §2, we recall the definitions of GSpin Shimura varieties, special endomorphisms, and special divisors. In §3, we formulate Theorems 3.2 and 3.3 which describe the decay of lattices of special endomorphisms at supersingular points. The proof of these statements occupies the next three sections, which may be skipped on a first reading. In §4, we recall from Ogus's work [Ogu79] the explicit description of the lattices of special endomorphisms at supersingular points and we use Kisin's work [Kis10] to compute an F-crystal $\mathbf{L}_{\mathrm{cris}}$, which controls the deformation of special endomorphisms. In sections §§5-6 we use this explicit description to prove the decay results. In §7, we prove Theorem 1.2 following the outline given above. In §8, we prove Theorem 1.4 using Theorem 1.2; we only use the statement (not the proof) of Theorem 1.2 and the reader who is interested in the Hecke orbit conjecture may directly proceed to §8 after §2.

Notation. Throughout the paper, $p \geq 5$ is a prime. We write $f \asymp g$ if $f = O(g)$ and $g = O(f)$.

Acknowledgement. We thank George Boxer, Ching-Li Chai, Johan de Jong, Kai-Wen Lan, Keerthi Madapusi Pera, Frans Oort, Arul Shankar, Andrew Snowden, Salim Tayou, and Tonghai Yang for helpful discussions, as well as Arthur and D.W. Read for additional assistance. A.S. has been partially supported by the NSF grant DMS-2100436 and Y.T. has been partially supported by the NSF grant DMS-1801237. Y.T. was a chargée de recherche at CNRS and Université Paris-Saclay from February 2020 to June 2021. We would like to thank the referee for a careful and thorough reading, and for valuable suggestions which vastly improved the exposition of the paper.

2. GSpin Shimura varieties and special divisors

In this section, we review basic definitions, terminology, and notation for GSpin Shimura varieties, special endomorphisms, and special divisors that we need in the rest of the paper.

Let (L, Q) be a quadratic $\mathbb{Z}$-lattice of signature $(b, 2)$, $b \geq 1$. We assume that (L, Q) is self-dual at p. We recall the canonical integral model of the GSpin Shimura variety associated to (L, Q) and the definition of special divisors. The main references are [MP16, §§3-5] and [AGHMP18, §§4.1-4.3];[6] see also [SSTT, §2] for a brief summary.

2.1. Let $V := L \otimes_{\mathbb{Z}} \mathbb{Q}$ and let $[-,-]$ denote the bilinear form on V given by $[x, y] = Q(x+y) - Q(x) - Q(y)$. Let $G := \mathrm{GSpin}(L \otimes \mathbb{Z}_{(p)}, Q)$ be the group of spinor similitudes of $L \otimes \mathbb{Z}_{(p)}$, which is a reductive group over $\mathbb{Z}_{(p)}$ and naturally a subgroup of $C(L \otimes \mathbb{Z}_{(p)})^\times$, where $C(-)$ denotes the Clifford algebra. The group $G(\mathbb{R})$ acts on the Hermitian symmetric domain $D_L = \{z \in V_{\mathbb{C}} \mid [z, z] = 0, [z, \bar{z}] < 0\}/\mathbb{C}^\times$ via $G_{\mathbb{Q}} \to \mathrm{SO}(V)$. For $[z] \in D_L$ with $z \in V_{\mathbb{C}}$, let $h_{[z]} : \mathrm{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m \to G_{\mathbb{R}}$ denote the unique homomorphism which induces the Hodge decomposition on $V_{\mathbb{C}}$ given by $V_{\mathbb{C}}^{1,-1} = \mathbb{C}z, V_{\mathbb{C}}^{0,0} = (\mathbb{C}z \oplus \mathbb{C}\bar{z})^\perp, V_{\mathbb{C}}^{-1,1} = \mathbb{C}\bar{z}$. Thus $(G_{\mathbb{Q}}, D_L)$ is a Shimura datum with reflex field $\mathbb{Q}$.

Let $\mathbb{K} \subset G(\mathbb{A}_f)$ be a compact open subgroup contained in $G(\mathbb{A}_f) \cap C(L \otimes \widehat{\mathbb{Z}})^\times$, where $C(L \otimes \widehat{\mathbb{Z}})$ is the Clifford algebra of $(L \otimes \widehat{\mathbb{Z}}, Q)$ and we assume that $\mathbb{K}_p = G(\mathbb{Z}_p)$. Then we have the GSpin Shimura variety $Sh := Sh(G_{\mathbb{Q}}, D_L)_{\mathbb{K}}$ over $\mathbb{Q}$ with $Sh(G_{\mathbb{Q}}, D_L)_{\mathbb{K}}(\mathbb{C}) = G(\mathbb{Q}) \backslash D_L \times G(A_f)/\mathbb{K}$ and by [Kis10, Theorem 2.3.8], Sh admits a canonical smooth integral model $\mathcal{S} := \mathcal{S}_{\mathbb{K}}$ over $\mathbb{Z}_{(p)}$.

2.2. Let H denote the Clifford algebra $C(L)$ equipped with the right action by itself via right multiplication, and equip $H \otimes \mathbb{Z}_{(p)}$ with the action of G by left multiplication. By picking a suitable symplectic form on H, we have $G_{\mathbb{Q}} \to \mathrm{GSp}(H \otimes \mathbb{Q})$, which induces a morphism from $(G_{\mathbb{Q}}, D_L)$ to a Siegel Shimura datum. Thus there is a Kuga–Satake abelian scheme $A^{\mathrm{univ}} \to Sh$ whose first $\mathbb{Z}$-coefficient Betti cohomology $\mathbf{H}_B$ is the local system induced by H (and its $G_{\mathbb{Q}}$-action). This Kuga–Satake abelian scheme $A^{\mathrm{univ}} \to Sh$ extends to an abelian scheme $\mathcal{A}^{\mathrm{univ}} \to \mathcal{S}$ equipped with a left $C(L)$-action. Let $\mathbf{H}_{\mathrm{dR}}, \mathbf{H}_{\ell,\text{ét}}$ denote the first relative de Rham cohomology and ℓ-adic étale cohomology with $\mathbb{Z}_\ell$-coefficient of $\mathcal{A}^{\mathrm{univ}} \to \mathcal{S}$ for $\ell \neq p$, and let $\mathbf{H}_{\mathrm{cris}}$ denote the first relative crystalline cohomology of $\mathcal{A}^{\mathrm{univ}}_{\mathbb{F}_p} \to \mathcal{S}_{\mathbb{F}_p}$.

The action of L on H via left multiplication induces a $G_{\mathbb{Q}}$ equivariant map $L \otimes \mathbb{Q} \to \mathrm{End}_{C(L)}(H \otimes \mathbb{Q})$, and thus we have a $\mathbb{Z}$-local system $\mathbf{L}_B$ over Sh with a natural embedding $\mathbf{L}_B \to \mathrm{End}_{C(L)}(\mathbf{H}_B)$. There are a filtered vector bundle with connection $\mathbf{L}_{\mathrm{dR}} \subset \mathrm{End}_{C(L)}(\mathbf{H}_{dR})$, a $\mathbb{Z}_\ell$-lisse sheaf $\mathbf{L}_{\ell,\text{ét}} \subset \mathrm{End}_{C(L)}(\mathbf{H}_{\ell,\text{ét}})$ and an F-crystal $\mathbf{L}_{\mathrm{cris}} \subset \mathrm{End}_{C(L)}(\mathbf{H}_{\mathrm{cris}})$ such that these embeddings along with $\mathbf{L}_B \to \mathrm{End}_{C(L)}(\mathbf{H}_B)$ are compatible under Betti-de Rham, Betti-étale, de Rham-crystalline comparison maps (see [MP16, Prop. 3.11, 3.12, Prop. 4.7]). By [AGHMP18, §4.3], $\mathbf{L}_?, ? = B, \mathrm{dR}, (\ell, \text{ét}), \mathrm{cris}$ are equipped with a natural quadratic form $\mathbf{Q}$ given by $f \circ f = \mathbf{Q}(f) \cdot \mathrm{Id}$ for a section f of $\mathbf{L}_?$.

Definition 2.3 ([AGHMP18, Def. 4.3.1]). Let T denote an $\mathcal{S}$-scheme.

[6] Since we work with the hyperspecial case, all the results listed here are in [MP16] and we follow the convention of using cohomology as in [MP16].

(1) An endomorphism $v \in \mathrm{End}_{C(L)}(\mathcal{A}_T^{\mathrm{univ}})$ is *special* if all cohomological realizations of v lie in the image of $\mathbf{L}_? \to \mathrm{End}_{C(L)}(\mathbf{H}_?)$, where $? = B, \mathrm{dR}, \mathrm{cris}, (\ell, \text{ét})$, for all $\ell \neq p$.[7]

(2) Assume that $T \otimes \mathbb{F}_p \neq \emptyset$. Let $\mathcal{A}_T^{\mathrm{univ}}[p^\infty]$ denote the p-divisible group associated to $\mathcal{A}_T^{\mathrm{univ}}$. An endomorphism $v \in \mathrm{End}_{C(L)}(\mathcal{A}_T^{\mathrm{univ}}[p^\infty])$ is *special* if its crystalline realization lies in $\mathbf{L}_{\mathrm{cris}}$.

Remark 2.4. For connected T, an endomorphism $v \in \mathrm{End}_{C(L)}(\mathcal{A}_T^{\mathrm{univ}})$ or $\mathrm{End}_{C(L)}(\mathcal{A}_T^{\mathrm{univ}}[p^\infty])$ is special if and only if there exists a geometric point $t \in T$ such that $v_t \in \mathrm{End}_{C(L)}(\mathcal{A}_t^{\mathrm{univ}})$ or $\mathrm{End}_{C(L)}(\mathcal{A}_t^{\mathrm{univ}}[p^\infty])$ is special (see [AGHMP18, Prop. 4.3.4, Lem. 4.3.5] and their proofs). Moreover, if $T_{\mathbb{F}_p} \neq \emptyset$, then we may pick a geometric point $t \in T_{\mathbb{F}_p}$ and for such t, $v_t \in \mathrm{End}_{C(L)}(\mathcal{A}_t^{\mathrm{univ}})$ is special if and only if the crystalline realization of v_t lies in $\mathbf{L}_{\mathrm{cris}}$ (see [MP16, Cor. 5.22, §5.24]). In this paper, we will mainly work with T which is an $\mathcal{S}_{\mathbb{F}_p}$-scheme and thus we will only use $\mathbf{L}_{\mathrm{cris}}$ to verify special endomorphisms.

Remark 2.5. By [MP16, Lem. 5.2], for $v \in \mathrm{End}_{C(L)}(\mathcal{A}_T^{\mathrm{univ}})$ special, we have $v \circ v = [Q(v)]$ for some $Q(v) \in \mathbb{Z}_{\geq 0}$ and $v \mapsto Q(v)$ is a positive definite quadratic form on the $\mathbb{Z}$-lattice of special endomorphisms of $\mathcal{A}_T^{\mathrm{univ}}$.

Definition 2.6. For $m \in \mathbb{Z}_{>0}$, the *special divisor* $\mathcal{Z}(m)$ is the Deligne–Mumford stack over $\mathcal{S}$ with functor of points $\mathcal{Z}(m)(T) = \{v \in \mathrm{End}(\mathcal{A}_T^{\mathrm{univ}}) \text{ special } | Q(v) = m\}$ for any $\mathcal{S}$-scheme T. We use the same notation for the image of $\mathcal{Z}(m)$ in $\mathcal{S}$. By for instance [AGHMP18, Prop. 4.5.8], $\mathcal{Z}(m)$ is an effective Cartier divisor and is flat over $\mathbb{Z}_{(p)}$ and hence $\mathcal{Z}(m)_{\mathbb{F}_p}$ is still an effective Cartier divisor of $\mathcal{S}_{\mathbb{F}_p}$; we denote $\mathcal{Z}(m)_{\mathbb{F}_p}$ by $Z(m)$.

3. Lattice decay statements and heuristics

In this section, we formulate local intersection multiplicities in terms of counting points from a nested sequence of lattices. In the supersingular case, we then state decay estimates for this nested sequence that will be crucial for controlling the local contributions. Proving these estimates will occupy §§4, 5, and 6. We give a heuristic explanation for why these decay estimates suffice. Finally, at the end of the section, we construct a formal family where the local multiplicities behave wildly; as a consequence, in our argument, it is necessary to use the global geometry to control the local error terms.

Preliminaries and main statements. Let k denote $\overline{\mathbb{F}}_p$ and recall from Theorem 1.2 that $C \to \mathcal{S}_k$ is a smooth proper curve whose generic point maps to the ordinary locus of $\mathcal{S}_k$. Given a perfect field k' of characteristic p, and a point $P' \in \mathcal{S}(k')$, we say P' is *ordinary* if the slopes of the crystalline Frobenius φ on $\mathbf{L}_{\mathrm{cris},P'}(W(k'))$ are $-1, 1$ with multiplicity 1 and 0 with multiplicity b. For $P' \in \mathcal{S}(k)$, we say P' is *supersingular* if the crystalline Frobenius φ on $\mathbf{L}_{\mathrm{cris},P'}(W(k))$ is isoclinic of slope 0.[8] Let $P \in C(k)$, and let t be a local coordinate at P (i.e., $\widehat{C}_P = \mathrm{Spf}\, k[[t]]$). Let $\mathcal{A}/k[[t]]$ denote the pullback of the universal abelian scheme $\mathcal{A}^{\mathrm{univ}}/\mathcal{S}$. Finally, let L_n denote the $\mathbb{Z}$-module of special endomorphisms of $\mathcal{A} \bmod t^n$. The moduli-theoretic description of the special divisors yields the following expression:

$$i_P(C.Z(m)) = \sum_{n=1}^{\infty} \#\{v \in L_n \mid Q(v) = m\}. \tag{3.1}$$

[7] We drop the ones which do not make sense. For instance, if p is invertible in T, we drop cris; if $T_{\mathbb{Q}} = \emptyset$, we drop B.

[8] By [HP17, Lem. 4.2.4], this definition of being supersingular is equivalent to that the corresponding Kuga–Satake abelian variety is supersingular.

As discussed in the introduction, one of the main difficulties in comparing local and global intersections is the contribution of supersingular and especially superspecial points; these are the supersingular points for which the lattice of special endomorphisms is as large as possible (see §4.1 for a precise definition). We will therefore assume that the image of P in $\mathcal{S}_k$ is contained in the supersingular locus of $\mathcal{S}_k$.

We now define the Cartier divisor associated to the reduced locus of non-ordinary points of $\mathcal{S}_k$, as well as the Hasse invariant on $\mathcal{S}_k$ which defines this locus. This is implicitly done in [Ogu01] in the case when $\mathcal{S}_k$ parameterizes K3 surfaces – the more general case can also be deduced from *loc. cit.*, because for the most part, Ogus' constructions and proofs only require K3 crystals. However, for completeness, we include a more thorough definition.

Consider the filtered vector bundle with connection $\mathbf{L}_{\mathrm{dR}}$ on $\mathcal{S}_k$. Define $E_1 \subset \mathbf{L}_{\mathrm{dR}}$ to be the conjugate filtration $F^1_{\mathrm{con}}\mathbf{L}_{\mathrm{dR}}$, which in turn is defined as follows[9]. At any closed point $s \in \mathcal{S}_k$, $E_{1,s}$ can be defined as in [Ogu01, page 327] – Ogus' definition only requires a K3 crystal (as defined in [Ogu79, Section 3]). To see that this definition globalizes to all of $\mathcal{S}_k$, we may use the conjugate filtration on $\mathbf{H}_{\mathrm{dR}}$ (which is well defined as $\mathbf{H}_{\mathrm{dR}}$ is the relative deRham cohomology of $\mathcal{A}/\mathcal{S}_k$) to induce a filtration (which we define to be F^1_{con}) on $\mathbf{L}_{\mathrm{dR}}$, which agrees with $E_{1,s}$ at any closed point s of $\mathcal{S}_k$. Having defined $F^1_{\mathrm{con}} \subset \mathbf{L}_{\mathrm{dR}}$ and therefore E_1, we consider the locus $\mathcal{F}_2 \subset \mathcal{S}_k$ defined in [Ogu01, Proposition 11], with $h = 1$. By [Ogu01, Proposition 3], $\mathcal{F}_2$ is the same as the non-ordinary locus as defined by the slopes of $\mathbf{L}_{\mathrm{cris}}$. Finally, by [Ogu01, Theorem 15], the equation defining the locus $\mathcal{F}_2$ is a section of $\mathrm{Fil}^1(\mathbf{L}_{\mathrm{dR}})^{\otimes(p-1)}$, and hence is a weight $p-1$ modular form which we define to be the *Hasse invariant* H on $\mathcal{S}_k$.

We now return to the setting of $P \in C \to \mathcal{S}_k$, with P supersingular. See Lemma 4.9 for the description of H in a neighborhood of $P \in \mathcal{S}_k$. We have that $\mathcal{A}$ is generically ordinary and specializes to a supersingular point. Therefore, after restriction to $k[[t]]$ the Hasse invariant H vanishes to some finite order at P.

In order to control the number of points in the nested family of lattices L_n, as n grows, we will prove that the covolumes of these lattices grow rapidly; note that the covolume of a lattice determines – to first order – the number of lattice points with bounded norm.

We define h_P to be $v_t(H)$, namely the t-adic valuation of H restricted to $\widehat{C}_P$. Our bounds will be in terms of the quantity h_P, and so we make the following definitions.

Definition 3.1. Let $r \geq 0$ denote an integer, and let $a = \frac{h_P}{2}$. Define $h_r = [h_P(p^r + \dots p + 1 + 1/p)]$, $h'_r = [h_P(p^r + \dots + 1) + a/p]$ and $h'_{-1} = [a/p]$.

Suppose that the point P is supersingular, but not superspecial. Then we have:

Theorem 3.2. *The index $|L_1/L_n|$ of L_n inside L_1 satisfies the inequality*

$$|L_1/L_n| \geq p^{2+2r}$$

if $h_r + 1 \leq n \leq h_{r+1}$.

We remind the reader that L_1 contains the lattices L_n with index a power of p (see [MST, Rmk. 7.2.2]). The content of the above result is that the for any n that is larger than $h_P(1 + 1/p)$, the abelian scheme $\mathcal{A} \bmod t^n$ has fewer special endomorphisms than $\mathcal{A} \bmod t$, and that the index of L_n in L_1 is at least p^2. For n greater than $h_P(p + 1 + 1/p)$, $\mathcal{A} \bmod t^n$ has still fewer special endomorphisms than $\mathcal{A} \bmod t$, and in fact the index of L_n in L_1 is at least p^4, etc.

As the lattice of special endomorphisms at P is maximal when P is superspecial, we need better bounds in this case. In §6, we establish the following result:

Theorem 3.3. *When P is superspecial, the index $|L_1/L_n|$ of L_n inside L_1 satisfies one of the following two inequalities:*

[9] We note that our notation differs from Ogus' by a Tate twist, so our F^1_{con} corresponds to Ogus' F^2_{con}

(1) $|L_1/L_n| \geq p^{1+2r}$ *if* $h'_{r-1} + ap^r + 1 \leq n \leq h'_r$ *and* $|L_1/L_n| \geq p^{2+2r}$ *if* $h'_r + 1 \leq n \leq h'_r + ap^{r+1}$.

(2) $|L_1/L_n| \geq p$ *if* $h'_{-1} + a + 1 \leq n \leq h'_0$ *and* $|L_1/L_n| \geq p^{3+2r}$ *if* $h'_r + 1 \leq n \leq h'_{r+1}$.

The above results show that there is a dichotomy between the local behavior at superspecial points, and supersingular points that are not superspecial. This is because the vanishing locus of the Hasse invariant on $\mathcal{S}_k$ is singular precisely at superspecial points (see for instance [Ogu01, the proof of Cor. 16]). This singularity forces the covolume of L_n to increase faster than it otherwise would.

3.4. A heuristic.

To motivate our approach, we give a heuristic argument here for the expectation that for $p \gg_\epsilon 1$, the sum of local intersection multiplicities $i_P(C.Z(m))$ at supersingular points on C with $Z(m)$ is at most $(\frac{1}{2} + \epsilon)C.Z(m)$ as $m \to \infty$ using Theorems 3.2 and 3.3. The proof of Theorem 7.18 verifies this expectation when we average over m. In particular, this heuristic explains why we need a stronger decay estimate for superspecial points and why such decay should exist. In order to just convey the basic idea, we will keep the argument presented here brief, even a little vague; more precise statements and proofs will come later in §7 and the reader may consult there for the precise statements and proofs.

Theorems 3.2 and 3.3 imply that for $p \gg 1$, the major contribution in $i_P(C.Z(m))$ comes from the intersection of $k[t]/t^{h_P}$ and $Z(m)$ (as the covolumes of L_n increase). The intersection multiplicity of $k[t]/t^{h_P}$ and $Z(m)$ is at most h_P times the number $b(m, P)$ of branches of the formal completion $\widehat{Z(m)}_P \subset \widehat{\mathcal{S}}_{k,P}$.

Indeed, $b(m, P) = \#\{v \in L_1 \mid Q(v) = m\}$. By studying the theta series associated to L_1, we have that $b(m, P)$ is roughly $|q_L(m)|/p^{t_P/2}$, where $q_L(m)$ denotes the m-th Fourier coefficient of the vector-valued Eisenstein series E_0 of weight $1 + b/2$ defined in §7.2 and t_P (which is an even positive integer) is the type of P defined in §4.2. In particular, this bound is the worst when $t_P = 2$, which by definition means that P is superspecial.

We now consider the extreme case when all non-ordinary points on C are superspecial. Since the Hasse invariant is a weight $p - 1$ modular form on $\mathcal{S}_k$, then we have

$$\sum_{P \in C(k) \text{ superspecial}} h_P = (p-1)C.\omega,$$

where ω is the line bundle of modular forms of weight one. Then without considering the first inequalities in each of (1) and (2) of Theorem 3.3, an initial estimate of $\sum_{P \text{ supersingular}} i_P(C.Z(m))$ is

$$\sum_{P \in C(k) \text{ superspecial}} h_P |q_L(m)|/p = \frac{p-1}{p} |q_L(m)|(C.\omega),$$

and a priori this should be a lower bound as we have ignored tangencies of order greater than h_P.

On the other hand, as we discuss in Lemma 7.5, based on Borcherds theory, this is roughly the same size as the global intersection $C.Z(m)$. Thus we need some extra input, which is exactly given by Theorem 3.3; this result lets us replace h_P by $h_P/2$ for the major term in $i_P(C.Z(m))$ and then obtain our expectation.

We can see how this works in the simplest situation, when C intersects all local formal branches of $\widehat{Z(m)}_P$ transversely. Then $i_P(C.Z(m)) = b(m, P)$. On the other hand, since the singular locus of the non-ordinary locus in $\mathcal{S}_k$ consists of the superspecial points, then $h_P \geq 2$ for all superspecial points and thus the total number of superspecial points on C is at most $\frac{p-1}{2} C.\omega$ and thus by the above estimate of $b(m, P)$, we see that $\sum_{P \text{ supersingular}} i_P(C.Z(m)) \leq (\frac{1}{2} + \epsilon)(C.Z(m))$.

3.5. An example of a formal curve. We will now construct a formal curve $\operatorname{Spf} k[[t]] \subset \mathcal{S}_k$ with closed point P where the local multiplicities $i_P(C.Z(m_i))$ grows exponentially fast for appropriate sequences of integers m_i. Our example will in fact be a case in which P is ordinary. For ease of exposition, we assume that the quadratic lattice has even rank, and consequently let $b = 2c$. There exist $\mathcal{S}_k, P$ such that L_1, the $\mathbb{Z}$-module of special endomorphisms at P, has rank $2c = b$, with the condition that the discriminant of L_1 is prime to p (for instance, we may always choose a GSpin Shimura variety which admits a map from the modular curve in characteristic 0, and then choose a CM point $\tilde{P}$ in the modular curve, and finally choose p to be a large enough ordinary prime for $\tilde{P}$, and P to be $\tilde{P}$ mod p). Suppose that $e_1, f_1, e_2, f_2, \dots e_c, f_c$ is an orthogonal $\mathbb{Z}$-basis of L_1. Let $\mathcal{L}_n = L_n \otimes \mathbb{Z}_p$ denote the module of (formal) special endomorphisms $\mathcal{A}[p^\infty]$ mod t^n.

Serre–Tate theory yields the existence of coordinates $\{q_i, q_i' : 1 \leq i \leq c\}$ such that the formal neighborhood of $\mathcal{S}_k$ at P is given by $\operatorname{Spf} k[[q_i - 1, q_i' - 1]]$ where the coordinates q_i, q_i' canonically endow $\operatorname{Spf} k[[q_i - 1, q_i' - 1]]$ with the structure of a formal torus – this follows by work of Noot [Noo96, Thm. 2.8, §3]. The locus in the formal deformation space of an ordinary abelian variety to which a specific endomorphism extends is well known to be a (possibly non-smooth) formal subgroup. Work of Madapusi Pera [MP16, Cor. 5.17] implies that the deformation space of a formal special endomorphism is a divisor. Furthermore, Madapusi Pera [MP16, Cor. 5.17, Cor. 5.19] proves that if a $\mathbb{Z}_p$-linearly independent set of special endomorphisms of P spans an n-dimensional self-dual $\mathbb{Z}_p$-submodule of the special endomorphisms of the ordinary point $\mathcal{A}[p^\infty]$ mod t, then the sub-locus of the formal neighborhood of $\mathcal{S}_k$ at P to which all endomorphisms in this module extend is a smooth formal subscheme of codimension n, and therefore a formal subtorus of codimension n. Therefore, we may assume that the coordinates q_i (respectively q_i') satisfy the property that the local equation defining the locus where the formal special endomorphisms e_i (respectively f_i) deform is just $q_i - 1$ (respectively $q_i' - 1$). It now follows that the local equation defining the locus where the formal special endomorphism $\sum_{i=1}^c (\lambda_i e_i + \mu_i f_i) \in \mathcal{L}_1,\ \lambda_i, \mu_i \in \mathbb{Z}_p$ deforms is just $\displaystyle\prod_{i=1}^c q_i^{\lambda_i} q_i'^{\mu_i} - 1$. Note that this has following consequence: if f is the local equation defining the locus where some special endomorphism w deforms, then f^p is the equation for pw.

We now choose μ_i to be irrational p-adic integers which are "very well approximated" by actual integers. Specifically, choose $\mu_i = \mu = \sum a_n p^n$ where $0 \leq a_n \leq p - 1$ and $a_0 = 1$. We will choose the precise values for a_n, $n \geq 1$ below.

We will now construct our formal curve to satisfy the property that $\mathcal{A}/\operatorname{Spf} k[[t]]$ admits no non-zero special endomorphisms, but $\mathcal{A}[p^\infty]/\operatorname{Spf} k[[t]]$ admits special endomorphisms by $\operatorname{Span}_{\mathbb{Z}_p}\{e_i + \mu f_i\}_{i=1}^c \subset \mathcal{L}_1$. Choosing $\operatorname{Spf} k[[t]] \subset \mathcal{S}_k$ to be defined by the quotient map $\rho : k[[q_i - 1, q_i' - 1]] \to k[[t]]$, with $\rho(q_i) = (1 + \alpha_i t)^{-\mu}$ and $\rho(q_i') = (1 + \alpha_i t)$, where $\alpha_i \in k$ are linearly independent over $\mathbb{F}_p$, is one such example and we will treat this example.

With this setup, we are now prepared to compute the lattices $\mathcal{L}_N$, and therefore also L_N and $i_P(\operatorname{Spf} k[[t]].Z(m_i))$. The assumption that the elements $\alpha_i \in k$ are $\mathbb{F}_p$-linearly independent and $\mu \in \mathbb{Z}_p^\times$ implies that the local equation defining the locus in $\operatorname{Spf} k[[t]]$ such that any primitive $w \in \operatorname{Span}\{e_1 \dots e_c\}$ deforms is just t. As the endomorphisms $e_i + \mu f_i, 1 \leq i \leq c$ extend to the whole of $\operatorname{Spf} k[[t]]$, we have $\mathcal{L}_2, \dots, \mathcal{L}_p = \operatorname{Span}\{pe_1, \dots, pe_c, e_1 + \mu f_1, \dots, e_c + \mu f_c\}$; $\mathcal{L}_{p+1}, \dots, \mathcal{L}_{p^2} = \operatorname{Span}\{p^2 e_1, \dots, p^2 e_c, e_1 + \mu f_1, \dots, e_c + \mu f_c\}$; and we finally have $\mathcal{L}_{p^{n-1}+a} = \operatorname{Span}\{p^n e_1, \dots, p^n e_c, e_1 + \mu f_1, \dots, e_c + \mu f_c\}$, where $a \geq 1$ and $p^{n-1} + a \leq p^n$. Finally, we have that $L_N = L_1 \cap \mathcal{L}_N$ (with the intersection in $\mathcal{L}_1$).

The fact that $\mu = \displaystyle\sum_{n \geq 0} a_n p^n$ implies that $v_{i,0} = e_i + a_0 f_i \in L_p$, $v_{i,1} = e_i + (a_0 + a_1 p) f_i \in L_{p^2}, \dots$, $v_{i,n} = e_i + (a_0 + a_1 p + \dots a_n p^n) f_i \in L_{p^{n+1}}$, etc. We finally choose our sequence of a_n – recall that we have already chosen $a_0 = 1$. To that end, define $n_0 = 0$, and recursively define $n_{j+1} = p^{2n_j}$. We define $a_{n_j} = 1$ and $a_n = 0$ if $n \neq n_j$, $\forall j \in \mathbb{Z}_{\geq 0}$. For any positive integer j_0, we see that $v_{i,n_{j_0+1}-1} = e_i + (\sum_{j=0}^{j_0} p^{n_j}) f_j \in L_{p^{n_{j+1}}}$. It is easy to see that $m_j := Q(v_{i,n_{j+1}-1}) \asymp p^{2n_j}$. Therefore, we have that $i_P(\operatorname{Spf} k[[t]].Z(m_j)) \geq p^{n_{j+1}}$, whose size is clearly exponential in m_j! We have therefore constructed an example of a formal curve, as well as a sequence of special divisors $Z(m_j)$, such that $i_P(\operatorname{Spf} k[[t]].Z(m_j))$ is exponential in m_j. In fact, $L_{p^{n_{j+1}}}$ contains a rank-c sublattice with discriminant $\asymp p^{2cn_j}$ (spanned by $\{v_{i,n_{j+1}-1}\}_{i=1}^c$). Therefore, when $c > 2$, by choosing our initial values $Q(e_i), Q(f_i)$ carefully, we may even arrange for $i_P(\operatorname{Spf} k[[t]].Z(m))$ growing exponentially in m_j (and therefore growing faster than any polynomial in m) for *a density one* set of $m \in [m_j, m_j^N]$.

In [MST], we are able to get around this difficulty because $c \leq 2$, and hence our lattices all have relatively small rank. Indeed, in that setting, the lattices $L_{p^{n_{j+1}}}$ may contain sublattices with discriminant logarithmic in $p^{n_{j+1}}$, but these sublattices necessarily have rank bounded above by 2, and the set of integers represented by rank two positive definite lattices has density zero.

4. The F-crystal $\mathbf{L}_{\text{cris}}$ on local deformation spaces of supersingular points

The goal of this section and §§5 and 6 is to prove Theorems 3.2 and 3.3 by analyzing the deformation behavior of special endomorphisms at supersingular points.

To set up this analysis, in this section, we compute $\mathbf{L}_{\text{cris}}$ over the formal neighborhoods of supersingular points in $\mathcal{S}_{\bar{\mathbb{F}}_p}$. As in [MST, §3], we first compute $\mathbf{L}_{\text{cris},P}(W)$ at a supersingular point P, which is a quadratic space over $W := W(\bar{\mathbb{F}}_p)$ with a σ-linear Frobenius action φ, and then we use Kisin's work [Kis10] to obtain $\mathbf{L}_{\text{cris}}$ over the formal neighborhood of P. Here we use the work of Ogus [Ogu79, §3] to compute $\mathbf{L}_{\text{cris},P}(W)$ while we follow [HP17] in [MST]; the extra input is [Ogu79, Thm. 3.21].

In [Ogu79], he uses the notion of supersingular K3 crystals [Ogu79, Def. 3.1], which are isoclinic of slope 1; these crystals differ by a Tate twist applied to our $\mathbf{L}_{\text{cris},P}(W)$ (which is isoclinic of slope 0). Our convention is the same as that in [HP17]. In particular, our Frobenius φ differs from the Frobenius in [Ogu79] by a factor of $1/p$. For the convenience of the reader, we give references to [Ogu79] whenever possible in this paper and the reader may check [MST] for the references to [HP17].

The F-crystal $\mathbf{L}_{\text{cris}}$ at a supersingular point.

4.1. Set $k = \bar{\mathbb{F}}_p, W = W(k), K = W[1/p]$ and let σ denote the usual Frobenius action on K. Given a supersingular point P, $\mathbb{L} := \mathbf{L}_{\text{cris},P}(W)$ is equipped with a quadratic form $\mathbf{Q}$ (see §2.2) and a σ-linear Frobenius action φ. We note that φ is not a endomorphism of $\mathbb{L}$, but is a σ-linear map $\mathbf{L}_{\text{cris},P}(W) \to \frac{1}{p}\mathbf{L}_{\text{cris},P}(W)$. Let $\langle -, - \rangle$ denote the bilinear form on $\mathbb{L}$ given by $\langle x, y \rangle = \mathbf{Q}(x + y) - \mathbf{Q}(x) - \mathbf{Q}(y)$. By the definition of $\mathbf{Q}$, we have $\langle \varphi(x), \varphi(y) \rangle = \sigma(\langle x, y \rangle)$.

Let $\mathcal{L}$ denote the $\mathbb{Z}_p$-lattice of special endomorphisms the p-divisible group $A_P[p^\infty]$, where $A_P := \mathcal{A}_P^{\text{univ}}$. By Dieudonné theory, we have that $\mathcal{L} = \mathbb{L}^{\varphi=1}$. Since P is supersingular, we have that $\operatorname{rk}_{\mathbb{Z}_p} \mathcal{L} = \operatorname{rk}_W \mathbb{L} = \operatorname{rk}_{\mathbb{Z}} L$ and $\mathbb{L} \subset \mathcal{L} \otimes_{\mathbb{Z}_p} K$.

By [Ogu82, Thm. 3.4], there is a decomposition of $\mathbb{Z}_p$-quadratic lattices $(\mathcal{L}, \langle , \rangle) = (\mathcal{L}_0, \langle , \rangle_0) \oplus (\mathcal{L}_1, \langle , \rangle_1)$, where $p \mid \langle , \rangle_0$, both $\frac{1}{p}\langle , \rangle_0$ and $\langle , \rangle_1$ are perfect, and $2 \mid \operatorname{rk}_{\mathbb{Z}_p} \mathcal{L}_0$. Thus $p\langle , \rangle_0$ induces a perfect $\mathbb{F}_p$-valued quadratic form on the $\mathbb{F}_p$-vector space $\frac{1}{p}\mathcal{L}_0/\mathcal{L}_0$; we also denote this quadratic form by $p\langle , \rangle_0$. The type of P, denoted by t_P, is defined to be $\operatorname{rk}_{\mathbb{Z}_p} \mathcal{L}_0$;[10] by [Ogu79, Cor. 3.11], $2 \mid t_P, 2 \leq t_P \leq \operatorname{rk} L$. We say P is *superspecial* if $t_P = 2$; otherwise, we say P is non-superspecial.

4.2. The above decomposition of $\mathcal{L}$ induces a decomposition of $\mathbb{L}$, which allows us to compute $\mathbb{L}$ explicitly. More precisely, by [Ogu79, Thms. 3.5,3.20], the W-quadratic lattice $\mathbb{L}$ with Frobenius

[10] By [Ogu01, p. 327], $t_P/2$ is the Artin invariant if A_P is the Kuga–Satake abelian variety associated to a $K3$ surface.

action decomposes as $\mathbb{L}_0 \oplus \mathbb{L}_1$, where

$$\mathcal{L}_0 \otimes W \subset \mathbb{L}_0 \subset \frac{1}{p}\mathcal{L}_0 \otimes W$$

and $\mathbb{L}_0 \bmod \mathcal{L}_0 \otimes W \subset (\frac{1}{p}\mathcal{L}_0/\mathcal{L}_0) \otimes k$ is a totally isotropic subspace with respect to $\langle,\rangle_0$ of dimension $t_P/2$ satisfying certain conditions and $\mathbb{L}_1 = \mathcal{L}_1 \otimes W$.

We first provide explicit descriptions of $\mathcal{L}_0$ and $\mathbb{L}_0$.

Lemma 4.3 (Ogus). *Set $n = t_P/2$ and $\lambda \in \mathbb{Z}_{p^2}^\times$ such that $\lambda^2 \bmod p \in \mathbb{F}_p$ is a quadratic non-residue. There exists a $\mathbb{Z}_p$-basis $\{e_1, \dots, e_n, f_1, \dots, f_n\}$ of $\mathcal{L}_0$ and the quadratic form $\langle,\rangle_0$ is given by $\langle e_i, f_i\rangle_0 = p$ for $i > 1$, $\langle e_1, e_1\rangle_0 = 2p$, $\langle f_1, f_1\rangle_0 = -2\lambda^2 p$, and $\langle v, w\rangle_0 = 0$ for all $(v, w) \in \{e_1, \dots, e_n, f_1, \dots, f_n\}^2$ such that $(v, w) \neq (e_i, f_i), (f_i, e_i), i > 1$ or $(e_1, e_1), (f_1, f_1)$.*

Proof. The assertion follows from Theorem 3.4 and the proof of Lemma 3.15 in [Ogu79]. □

Lemma 4.4 (Ogus). *Fix the $\mathbb{Z}_p$-quadratic space $\mathcal{L}_0$ as in Lemma 4.3. For each supersingular point P with $(\mathbb{L}_0)^{\varphi=1} = \mathcal{L}_0$, there exists a vector $v \in \frac{1}{p}(\mathcal{L}_0 \otimes_{\mathbb{Z}_p} W)$ such that*

$$\mathbb{L}_0 = \mathrm{Span}_W\{v, \sigma(v), \dots, \sigma^{n-1}(v)\} + \mathcal{L}_0 \otimes W$$

and v satisfies the following conditions:

(1) $\mathrm{Span}_W\{v, \sigma(v), \dots, \sigma^{2n-1}(v)\} = \frac{1}{p}\mathcal{L}_0 \otimes_{\mathbb{Z}_p} W$.

(2) $\mathrm{Span}_W\{v, \sigma(v), \dots, \sigma^{n-1}(v)\}$ is isotropic for $\langle,\rangle_0$.

(3) $\langle v, \sigma^n(v)\rangle_0 = 1/p$,

where we use σ to denote the action $1 \otimes \sigma$ on $\mathcal{L}_0 \otimes_{\mathbb{Z}_p} K$. The quadratic form and φ action on $\mathbb{L}_0$ are induced by those on $\mathcal{L}_0 \otimes_{\mathbb{Z}_p} K$ via $\mathbb{L}_0 \subset \mathcal{L}_0 \otimes_{\mathbb{Z}_p} K$, where φ on $\mathcal{L}_0 \otimes_{\mathbb{Z}_p} K$ is given by $1 \otimes \sigma$. Finally, the set of vectors $\{v, \sigma(v), \dots, \sigma^{n-1}(v), p\sigma^n(v), \dots, p\sigma^{2n-1}(v)\}$ forms a W-basis for $\mathbb{L}_0$.

Proof. Consider the inclusion $\mathcal{L}_0 \otimes W \subset \mathbb{L}_0 \subset \frac{1}{p}\mathcal{L}_0 \otimes W$. Recall from §4.1 that the quadratic form $p^{-1}\langle,\rangle_0$ yields a perfect bilinear form on $\mathcal{L}_0 \otimes W$. By [Ogu79, Theorem 3.5], the data of $\mathbb{L}_0$ is in bijection with the data of an n-dimensional subspace $\overline{H} \subset \mathcal{L}_0 \otimes k$ which is isotropic for $p^{-1}\langle,\rangle_0$, where $\overline{H}$ satisfies conditions 3.5.2 and 3.5.3 of *loc. cit.*.[11] Let $H \subset \mathcal{L}_0 \otimes W$ denote any lift of $\overline{H}$ and then the crystal $\mathbb{L}_0$ corresponding to $\overline{H}$ is defined to be $\frac{1}{p}H + \mathcal{L}_0 \otimes W$. Note that $\mathbb{L}_0$ only depends on $\overline{H}$ and not on H itself, and that $\overline{H}$ is indeed the kernel of the natural map $\mathcal{L}_0 \otimes k \to \mathbb{L}_0 \otimes k$.

The discussion in the paragraph above Theorem 3.21 in [Ogu79] implies that there exists a vector $e' \in \overline{H}$ such that $\{e', \dots, \sigma^{n-1}(e')\}$ yields a basis of $\overline{H}$, and the set $\{e', \dots, \sigma^{2n-1}(e')\}$ is a basis of $\mathcal{L}_0 \otimes k$. Note that although the discussion in *loc. cit.* is in the context of $\varphi^{-1}(\overline{H}) \subset \mathcal{L}_0 \otimes W$ and not $\overline{H}$, everything applies to our setting too, by defining e' to be $\sigma(e)$, where e is as in [Ogu79, p. 33], and note that $\varphi(\mathcal{L}_0) = \mathcal{L}_0$.

A straightforward application of Hensel's lemma yields a specific choice of an *isotropic* n-dimensional $H_0 \subset \mathcal{L}_0 \otimes W$ along with a vector $\tilde{e}'$, with the property that H_0 and $\tilde{e}'$ reduce to $\overline{H}$ and e' mod p such that H_0 is the W-span of $\tilde{e}', \sigma(\tilde{e}'), \dots, \sigma^{n-1}(\tilde{e}')$. It then follows that the W-span of $\tilde{e}', \sigma(\tilde{e}'), \dots, \sigma^{2n-1}(\tilde{e}')$ equals $\mathcal{L}_0 \otimes W$. By replacing $\tilde{e}'$ by an appropriate $W^\times$-multiple, we may also assume that $\frac{1}{p}\langle \tilde{e}', \sigma^n(\tilde{e}')\rangle_0 = 1$. The the lemma follows by defining $v = \frac{1}{p}\tilde{e}'$. □

Lemma 4.5. *Set $v_i = \sigma^{i-1}(v), i = 1, \dots, 2n$ for the vector v in Lemma 4.4. Then there exist vectors $w_1, \dots, w_n \in \mathbb{L}_0$ such that*

(1) $v_1, \dots, v_n, w_1 \dots, w_n$ form a W-basis of $\mathbb{L}_0$;

[11] Ogus proved that the isomorphism classes of so-called K3 crystals ([Ogu79, Def. 3.1]) are in bijection with the data in [Ogu79, Thm. 3.5] described here; indeed, the isomorphism classes of K3 crystals in Ogus's sense are isomorphism classes of $\mathbb{L}$ for supersingular points by [HP17].

(2) The Gram matrix of $\langle , \rangle_0$ *with respect to this basis is* $\left[\begin{array}{c|c} 0 & I \\ \hline I & 0 \end{array}\right]$*;*

(3) The Frobenius φ *on* $\mathbb{L}_0$ *with respect to this basis is of form* $B_0\sigma$*, where*[12]

$$B_0 = \left[\begin{array}{cccc|cccc} 0 & & & & & & & p \\ 1 & & & & & & & pb_1 \\ & \ddots & & & & & & \vdots \\ & & 1 & & & & & pb_{n-1} \\ \hline & & & p^{-1} & -b_1 & \dots & -b_{n-1} & 0 \\ & & & & 1 & & & 0 \\ & & & & & \ddots & & \vdots \\ & & & & & & 1 & 0 \end{array}\right] \text{ with } b_i \in W.$$

Proof. By the final claim of Lemma 4.4, $\{v_1, \dots, v_n, pv_{n+1}, \dots, pv_{2n}\}$ is a basis of $\mathbb{L}_0$ over W. By Lemma 4.4(2)(3) and the fact that $\langle \varphi(x), \varphi(y)\rangle_0 = \sigma(\langle x, y\rangle_0)$, we have that $\langle v_i, pv_j\rangle_0 = 0$ for $j \leq i+n-1$ and $\langle v_i, pv_{i+n}\rangle_0 = 1$; thus by modifying $pv_{n+1}, \dots, pv_{2n}$ by an upper-unipotent matrix, we obtain $w_1, \dots, w_n$ satisfying condition (2). Moreover, the left half of B_0 in condition (3) also follows from the definition of v_i and that $w_1 = pv_{n+1}$.

We now consider the top-right block of B. To deduce that the first $n-1$ columns of this block vanish, (2) shows that it suffices to prove $\langle \varphi(w_i), w_j\rangle_0 = 0$ for $1 \leq i \leq n-1$ and $1 \leq j \leq n$. By definition, the $w_i, 1 \leq i \leq n-1$ are W-linear combinations of $pv_{n+1} = p\varphi^n(v), \dots, pv_{2n-1} = p\varphi^{n-2}(v)$ and thus $\varphi(w_i)$ is contained in $\mathrm{Span}_W\{pv_{n+1}, \dots, pv_{2n}\} = \mathrm{Span}_W\{w_1, \dots, w_n\}$. Since $\mathrm{Span}_W\{w_1, \dots, w_n\}$ is isotropic by Lemma 4.4(2), then $\langle \varphi(w_i), w_j\rangle_0$ for $1 \leq i \leq n-1$ and $1 \leq j \leq n$ as required. In order to prove that the last column of this block is as claimed in the lemma, it suffices to prove that $\langle \varphi(w_n), w_1\rangle_0 = \sigma(\langle w_n, \varphi^{-1}(w_1)\rangle_0) = p$ and $p \mid \langle \varphi(w_n), w_j\rangle_0 = \sigma(\langle w_n, \varphi^{-1}(w_j)\rangle_0)$ for $j \leq n$. Note that $\varphi^{-1}(w_1) = pv_n$ and then the first equality follows. For the rest, note that φ gives a σ-linear endomorphism of $\mathcal{L}_0 \otimes W$ and $p \mid \langle , \rangle_0$ on $\mathcal{L}_0$, thus $w_1 \dots w_n, \varphi(w_n) \in \mathcal{L}_0 \otimes W$ and $p \mid \langle \varphi(w_n), w_j\rangle_0$ for all $j \leq n$.

Similarly, for the bottom-right part of B, it suffices to show that $\langle \varphi(w_i) - w_{i+1}, v_j\rangle_0 = 0$, $p\langle \varphi(w_i), v_1\rangle_0 = -\langle \varphi(w_n), w_{i+1}\rangle_0$, $\forall 1 \leq i \leq n-1, 2 \leq j \leq n$ and $\langle \varphi(w_n), v_j\rangle_0 = 0$, $\forall 1 \leq j \leq n$. Note that $\langle \varphi(w_i) - w_{i+1}, v_j\rangle_0 = \langle \varphi(w_i), v_j\rangle_0 - \langle w_{i+1}, v_j\rangle_0 = \langle w_i, v_{j-1}\rangle_0 - \langle w_{i+1}, v_j\rangle_0 = 0$ by condition (2). Then $\varphi(w_i) = w_{i+1} + a_i w_1$ for some $a_i \in W$. Thus $w_{i+1} = \varphi(w_i - \sigma^{-1}(a_i)pv_n)$ and then $\langle \varphi(w_n), w_{i+1}\rangle_0 = \sigma(\langle w_n, w_i - \sigma^{-1}(a_i)pv_n\rangle_0) = -pa_i$; in other words, $p\langle \varphi(w_i), v_1\rangle_0 = -\langle \varphi(w_n), w_{i+1}\rangle_0$. Moreover, $\langle \varphi(w_n), v_j\rangle_0 = \sigma(\langle w_n, v_{j-1}\rangle_0) = 0$ for $j \geq 2$. For $\langle \varphi(w_n), v_1\rangle_0 =: c$, by the above discussion, $\varphi(w_n) = pv_1 + pb_1v_2 + \dots + pb_{n-1}v_n + cw_1$ and thus $\langle \varphi(w_n), \varphi(w_n)\rangle_0 = 2pc$; on the other hand, $\langle \varphi(w_n), \varphi(w_n)\rangle_0 = \sigma(\langle w_n, w_n\rangle_0) = 0$ and then $c = 0$, which finishes the proof of the lemma. □

4.6. Let S_0 denote the change-of-basis matrix from $\{e_i, f_i\}_{i=1}^n$ to $\{v_i, w_i\}_{i=1}^n$ in Lemma 4.5. More precisely, $S_0 \in M_{2n}(K)$ whose first (resp. last) n columns are the coordinates of v_i (resp. w_i) in terms of the basis $\{e_i, f_i\}_{i=1}^n$. For the simplicity of computations in §5, let S_0' the change-of-basis matrix from $\{e_i, f_i\}_{i=1}^n$ to $\{pv_i, w_i\}_{i=1}^n$. From the proof of Lemma 4.5, $\mathrm{Span}_W\{e_i, f_i\}_{i=1}^n = \mathrm{Span}_W\{pv_i, w_i\}_{i=1}^n$; thus $S_0' \in \mathrm{GL}_{2n}(W)$. Moreover, by definition, $S_0 = S_0' \left[\begin{array}{c|c} p^{-1}I & 0 \\ \hline 0 & I \end{array}\right]$.

4.7. We now describe $\mathbb{L}_1$ and $\mathcal{L}_1$ defined in §§4.1,4.2. Recall that $\mathbb{L}_1 = \mathcal{L}_1 \otimes W$, the Frobenius φ on $\mathbb{L}_1$ is given by $1 \otimes \sigma$ and the quadratic form on $\mathbb{L}_1$ is also induced by the one on $\mathcal{L}_1$, so we only need to classify $\mathcal{L}_1$. Unlike $\mathcal{L}_0$, which is completely determined by t_P (see Lemma 4.3), the $\mathbb{Z}_p$-quadratic

[12] All empty entries in the matrix are 0.

lattice $\mathcal{L}_1$ depends on $\dim L$ and $\operatorname{disc} L$ (see [Ogu79, Thm. 3.4] and [HP17, §4.3.1]). Since $\mathcal{L}_1$ is self-dual, we have the following three cases:[13]

(1) $\dim_{\mathbb{Z}_p} \mathcal{L}_1 = 2m$ and there is an m-dimensional isotropic subspace of $\mathcal{L}_1$ over $\mathbb{Z}_p$. We call this the *split* case.
(2) $\dim_{\mathbb{Z}_p} \mathcal{L}_1 = 2m$ and there does *not* exist an m-dimensional isotropic subspace of $\mathcal{L}_1$ over $\mathbb{Z}_p$.
(3) $\dim_{\mathbb{Z}_p} \mathcal{L}_1$ is odd.

Note that for cases (2)(3), one may always embed $\mathcal{L}_1$ into a split $\mathbb{Z}_p$-quadratic lattice of larger dimension. Therefore, we deal exclusively with the *split* case and we will remark in the proofs of the decay lemmas in §§5-6 that by the above embedding trick, the computation in the split case will also prove the decay lemmas in all other cases. We use $\{e'_i, f'_i\}_{i=1}^m$ to denote a $\mathbb{Z}_p$-basis of $\mathcal{L}_1$ in the split case such that the Gram matrix with respect to this basis is $\left[\begin{array}{c|c} 0 & I \\ \hline I & 0 \end{array}\right]$.

Description of $\mathbf{L}_{\text{cris}}$ at the formal neighborhood.

4.8. Following [Kis10, §§1.4-1.5], we will describe the formal neighborhood of the Shimura variety at the supersingular point P, and also compute the F-crystal $\mathbf{L}_{\text{cris}}$ over this formal neighborhood (see also [MST, §3.1.5, §3.2.1]). We first summarize Kisin's description in abstract terms, before providing an explicit description of the F-crystal in terms of the coordinates provided earlier in this section.

Recall from §4.1 that the quadratic form $\mathbf{Q}$ on $\mathbb{L}$ is compatible with the Frobenius φ on $\mathbb{L}$; moreover, $\mathbf{L}_{\text{dR}}$ defined in §2.2 carries the Hodge filtration so, by the canonical de Rham-crystalline comparison, $\mathbb{L} \otimes_W k$ also carries a filtration which we call *the* $\bmod p$ *Hodge filtration.* Let $\mu : \mathbb{G}_{m,W} \to \mathrm{SO}(\mathbb{L}, \mathbf{Q})$ denote any co-character (which we shall refer to as "the Hodge co-character") whose mod p reduction induces the above filtration. Let U denote the opposite unipotent in $\mathrm{SO}(\mathbb{L}, \mathbf{Q})$ with respect to μ, and let $\operatorname{Spf} R = \widehat{U}$ denote the completion of U at the identity section. Pick $\sigma : R \to R$ to be a lift of the Frobenius endomorphism on $R \bmod p$. Let u be the tautological R-point of U.

Then, by [Kis10, §§1.4, 1.5], there exists an isomorphism between the complete local ring of the Shimura variety at P and $\operatorname{Spf} R$ such that the F-crystal $\mathbf{L}_{\text{cris}}(R)$ is isomorphic to $\mathbb{L} \otimes_W R$ as an R-module, and the Frobenius action on $\mathbf{L}_{\text{cris}}(R)$, denoted by Frob, is given by $\text{Frob} = u \circ (\varphi \otimes \sigma)$ on $\mathbb{L} \otimes_W R$ via the isomorphism $\mathbf{L}_{\text{cris}}(R) \cong \mathbb{L} \otimes_W R$. For the simplicity of notation, we will fix the above mentioned isomorphisms and write $\mathbf{L}_{\text{cris}}(R) = \mathbb{L} \otimes_W R$. By the canonical de Rham-crystalline comparison, the Hodge filtration on $\mathbf{L}_{\text{dR}}$ induces a filtration on $\mathbf{L}_{\text{cris}}(R \otimes_W k)$ which we also call *the Hodge filtration.*

We will now provide an explicit description in terms of coordinates of the above objects. By Lemma 4.5(3) and Mazur's theorem on determining the $\bmod p$ Hodge filtration using φ (see for instance [Ogu82, p. 411]), the $\bmod p$ Hodge filtration on $\mathbb{L} \otimes_W k$ is given by

$$\operatorname{Fil}^1 \mathbb{L} \otimes_W k = \operatorname{Span}_k\{\bar{w}_n\}, \operatorname{Fil}^0 \mathbb{L} \otimes_W k = \operatorname{Span}_k\{\bar{v}_i, \bar{w}_j, \bar{e}'_l, \bar{f}'_l\}_{i=1,\ldots,n-1, j=1,\ldots,n, l=1,\ldots,m}, \operatorname{Fil}^{-1} \mathbb{L} \otimes_W k = \mathbb{L} \otimes_W k,$$

where $\bar{v}_i, \bar{w}_j, \bar{e}'_l, \bar{f}'_l$ denote the reduction of $v_i, w_j, e'_l, f'_l \bmod p$. Thus, with respect to the basis $\{v_i, w_i, e'_j, f'_j\}_{i=1,\ldots,n, j=1,\ldots,m}$, we choose the Hodge cocharacter $\mu : \mathbb{G}_{m,W} \to \mathrm{SO}(\mathbb{L}, \mathbf{Q})$ in the local Shimura datum to be

[13] Comparing to [MST, §3], §3.2.1 in *loc. cit.* is a special case of the split even dimensional case, §3.2.2 in *loc. cit.* is a special case of the non-split even dimensional case, and §3.3 in *loc. cit.* is a special case of the odd dimensional case.

$$\mu(t) = \left[\begin{array}{cccc|cccc|ccc}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & t^{-1} & & & & & & & \\
\hline
 & & & & 1 & & & & & & \\
 & & & & & \ddots & & & & & \\
 & & & & & & 1 & & & & \\
 & & & & & & & t & & & \\
\hline
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{array}\right],$$

where the diagonal blocks have sizes n, n, and $2m$.

Moreover, there exist local coordinates $\{x_i, y_i, x'_j, y'_j\}_{i=1,\dots,n-1, j=1,\dots,m}$ such that the complete local ring $\widehat{\mathcal{O}}_{\mathcal{S},P}$ of $\mathcal{S}$ at P is isomorphic to $\operatorname{Spf} R$, where $R = W[[x_i, y_i, x'_j, y'_j]]_{i=1,\dots,n-1,j=1,\dots,m}$. We define $\sigma : R \to R$, the operator that restricts to the usual Frobenius element on W and which lifts the Frobenius endomorphism on $R \mod p$, to be $\sigma(x_i) = x_i^p, \sigma(y_i) = y_i^p, \sigma(x'_j) = (x'_j)^p, \sigma(y'_j) = (y'_j)^p$. The tautological point of the opposite unipotent in $\mathrm{SO}(\mathbb{L}, \mathbf{Q})$ with respect to μ has the following description in terms of our coordinates:

$$u = I + \left[\begin{array}{cccc|cccc|cccccc}
 & & & & & & & -y_1 & & & & & & \\
 & & & & & & & \vdots & & & & & & \\
 & & & & & & & -y_{n-1} & & & & & & \\
x_1 & \dots & x_{n-1} & 0 & y_1 & \dots & y_{n-1} & Q & x'_1 & \dots & x'_m & y'_1 & \dots & y'_m \\
\hline
 & & & & & & & -x_1 & & & & & & \\
 & & & & & & & \vdots & & & & & & \\
 & & & & & & & -x_{n-1} & & & & & & \\
 & & & & & & & 0 & & & & & & \\
\hline
 & & & & & & & -y'_1 & & & & & & \\
 & & & & & & & \vdots & & & & & & \\
 & & & & & & & -y'_m & & & & & & \\
 & & & & & & & -x'_1 & & & & & & \\
 & & & & & & & \vdots & & & & & & \\
 & & & & & & & -x'_m & & & & & &
\end{array}\right],$$

where $Q = -\sum_{i=1}^{n-1} x_i y_i - \sum_{j=1}^{m} x'_j y'_j$.

The Frobenius action Frob on $\mathbf{L}_{\mathrm{cris}}(R) = \mathbb{L} \otimes_W R$ is given by $\mathrm{Frob} = u \circ (\varphi \otimes \sigma)$. Thus, with respect to the R-basis $\{v_i \otimes 1, w_i \otimes 1, e'_j \otimes 1, f'_j \otimes 1\}_{i=1,\dots,n,j=1,\dots,m}$, we have that $\mathrm{Frob} = (uB) \circ \sigma$, where $B = \left[\begin{array}{c|c} B_0 & 0 \\ \hline 0 & I \end{array}\right]$, and B_0 is given in Lemma 4.5.

Equation of the non-ordinary locus. We now compute the local equation of the non-ordinary locus in a formal neighborhood of a supersingular point. Recall that we have the Hodge cocharacter μ, whose mod p reduction induces the mod p Hodge filtration on $\mathbf{L}_{\mathrm{cris},P}(k) = \mathbb{L} \otimes_W k$. By [Moo98,

§4.5], the Hodge filtration on $\mathbf{L}_{\mathrm{cris}}(R \otimes_W k) = \mathbb{L} \otimes_W (R \otimes_W k)$ is given by

$$\mathrm{Fil}^i(\mathbf{L}_{\mathrm{cris}}(R \otimes_W k)) = \mathrm{Fil}^i(\mathbb{L} \otimes_W k) \otimes_k (R \otimes_W k), i = -1, 0, 1.$$

As in [MST, §3.4], we note that $p\,\mathrm{Frob}$ induces a map $\mathrm{gr}_{-1}\,\mathbf{L}_{\mathrm{cris}}(R \otimes_W k) \to \mathrm{gr}_{-1}\,\mathbf{L}_{\mathrm{cris}}(R \otimes_W k)$. Ogus proved the following result.

Lemma 4.9 (Ogus)**.** *For a supersingular point P, the non-ordinary locus (over k) in the formal neighborhood of P is given by the equation*

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

See [Ogu01][Prop. 11 and p 333-334] (or [MST][Lemma 3.4.1] which elaborates on [Ogu01]).

Corollary 4.10. *For a supersingular point P, the non-ordinary locus (over k) in the formal neighborhood of P is given by the equation $Q = 0$ if P is superspecial; otherwise, the equation is given by $y_1 = 0$.*

Proof. In what follows, the number n is as in Lemma 4.5, i.e., $2n = t_P = \dim_W \mathbb{L}_0$ and we follow the notation in §4.8. The space $\mathrm{gr}_{-1}\,\mathbf{L}_{\mathrm{cris}}(R \otimes_W k)$ is spanned by $\bar{v}_n$. We use the description of $\mathrm{Frob} = (uB) \circ \sigma$ (from Lemma 4.5 and the explicit description of u in §4.8) to see that the map $p\,\mathrm{Frob} : \mathrm{gr}_{-1}\,\mathbf{L}_{\mathrm{cris}}(R \otimes_W k) \to \mathrm{gr}_{-1}\,\mathbf{L}_{\mathrm{cris}}(R \otimes_W k)$ has the explicit description

$$p\,\mathrm{Frob}(\bar{v}_n) = Q\bar{v}_n \text{ if } n = 1;\ p\,\mathrm{Frob}(\bar{v}_n) = y_1\bar{v}_n \text{ if } n > 1.$$

The result now follows from the fact that P is superspecial if and only if $t_P = 2$ if and only if $n = 1$. □

4.11. In order to compute the powers of Frob in the proofs of the decay lemmas later, we describe Frob with respect to the K-basis $\{e_i, f_i, e'_j, f'_j\}_{i=1,\dots,n, j=1,\dots,m}$ of $\mathbb{L} \otimes_W K$. Let $S = \left[\begin{array}{c|c} S_0 & 0 \\ \hline 0 & I \end{array}\right], S' = \left[\begin{array}{c|c} S'_0 & 0 \\ \hline 0 & I \end{array}\right]$, where S_0, S'_0 are defined in §4.6 and thus $S' \in \mathrm{GL}_{2n+2m}(W)$. Then by definition, $B = S^{-1}\sigma(S)$.

We view $\{e_i, f_i, e'_j, f'_j\}_{i=1,\dots,n, j=1,\dots,m}$ as an $R[1/p]$-basis of $\mathbf{L}_{\mathrm{cris}}(R) \otimes_R R[1/p]$ and then Frob is given by $S(uB)\sigma(S^{-1}) \circ \sigma = SuS^{-1} \circ \sigma = S'u'(S')^{-1} \circ \sigma$, where

$$u' = I + \left[\begin{array}{cccc|cccc|cccccc}
 &&&& &&& -y_1/p &&&&&& \\
 &&&& &&& \vdots &&&&&& \\
 &&&& &&& -y_{n-1}/p &&&&&& \\
x_1 & \dots & x_{n-1} & 0 & y_1/p & \dots & y_{n-1}/p & Q/p & x'_1/p & \dots & x'_m/p & y'_1/p & \dots & y'_m/p \\ \hline
 &&&& &&& -x_1 &&&&&& \\
 &&&& &&& \vdots &&&&&& \\
 &&&& &&& -x_{n-1} &&&&&& \\
 &&&& &&& 0 &&&&&& \\ \hline
 &&&& &&& -y'_1 &&&&&& \\
 &&&& &&& \vdots &&&&&& \\
 &&&& &&& -y'_m &&&&&& \\
 &&&& &&& -x'_1 &&&&&& \\
 &&&& &&& \vdots &&&&&& \\
 &&&& &&& -x'_m &&&&&&
\end{array}\right].$$

5. Decay for non-superspecial supersingular points

The goal of this and the next section is to prove that, at supersingular points, special endomorphisms "decay rapidly" in the sense of [MST, Def. 5.1.1], which we will recall below.

Throughout these sections, $k = \bar{\mathbb{F}}_p$, $W = W(k)$, $K = W[1/p]$. We focus on the behavior of the curve C in Theorem 1.2 in a formal neighborhood of a supersingular point P, so we may let $C = \operatorname{Spf} k[[t]]$ denote a generically ordinary formal curve in $\mathcal{S}_k$ which specializes to P. In this section, we will focus on the case when P is non-superspecial and we treat the superspecial case in §6.

Let $\mathcal{A}/k[[t]]$ denote the pullback of the universal abelian scheme $\mathcal{A}^{\text{univ}}$ over the integral model $\mathcal{S}$ of the GSpin Shimura variety via $\operatorname{Spf} k[[t]] \to \mathcal{S}_k$ and let A denote $\mathcal{A} \bmod t$, and we will consider the p-divisible groups $\mathcal{A}[p^\infty], A[p^\infty]$ associated to $\mathcal{A}, A$. Let h denote the t-adic valuation of the local equation defining the non-ordinary locus given in Corollary 4.10. Recall from §4.1, $\mathcal{L}$ is the lattice of special endomorphisms of $A[p^\infty]$.

Definition 5.1 ([MST, Def. 5.1.1])**.** We say that $w \in \mathcal{L}$ *decays rapidly* if for every $r \in \mathbb{Z}_{\geq 0}$, the special endomorphism $p^r w$ does not lift to an endomorphism of $\mathcal{A}[p^\infty]$ modulo t^{h_r+1}, where $h_r := [h(p^r + \cdots + 1 + 1/p)]$. We say that a $\mathbb{Z}_p$-submodule of $\mathcal{L}$ decays rapidly if every primitive vector in this submodule decays rapidly.

The main theorem of this section is the following:

Theorem 5.2 (The Decay Lemma)**.** *There exists a rank* 2 *saturated* $\mathbb{Z}_p$*-submodule of* $\mathcal{L}$ *which decays rapidly.*

Theorem 3.2 follows directly from the Decay Lemma:

Proof of Theorem 3.2. We first note that $\mathcal{L}_n$, the lattice of special endomorphisms of $\mathcal{A}[p^\infty] \bmod t^n$, is precisely $L_n \otimes \mathbb{Z}_p$. Upon choosing a basis of $\mathcal{L}$ that extends a basis of the submodule that decays rapidly (which we may do, as this submodule is saturated in $\mathcal{L}$), we see that the index $|\mathcal{L}/\mathcal{L}_n|$ of $\mathcal{L}_n$ in $\mathcal{L}$ is at least p^{2r+2} if $n \geq h_r + 1$. The corresponding statements for L_n now follow directly. □

5.3. We first give an indication why the reader should expect a statement along these lines to hold. Note that in the mixed characteristic setting, namely while deforming from k to $W(k)$, applying Grothendieck–Messing theory yields that if a special endomorphism α lifts mod p^n but not mod p^{n+1}, then the special endomorphism $p\alpha$ would lift mod p^{n+1} but *not* p^{n+2} (see [ST20, Lemma 4.1.2]). However, Grothendieck–Messing theory is inherently limited in the equicharacteristic p setting, and the bounds it yields are worse than the bounds it yields in the mixed characteristic setting.

We illustrate this with the following example. Let $\mathscr{G}$ denote a p-divisible group over $\operatorname{Spec} k[t]/(t^a)$, suppose that α is any endomorphism of $\mathscr{G}$, and suppose that $\mathscr{G}'$ over $\operatorname{Spec} k[t]/(t^{pa})$ is a p-divisible group that deforms $\mathscr{G}$. We claim that the endomorphism $p\alpha$ deforms to $\mathscr{G}'$ regardless of how α behaves. Indeed, let $\mathbb{D}$ denote the Dieudonné crystal of $\mathscr{G}/\operatorname{Spec} k[t]/(t^a)$. As the map $\operatorname{Spec} k[t]/(t^a) \to \operatorname{Spec} k[t]/(t^{pa})$ is naturally equipped with a divided powers structure (and this is the key point in this observation), we may evaluate the Dieudonné crystal $\mathbb{D}$ at $\operatorname{Spec} k[t]/(t^{pa})$, and Grothendieck–Messing theory implies that the choice of deformation $\mathscr{G}'$ is equivalent to the choice of a filtration of $\mathbb{D}(\operatorname{Spec} k[t]/(t^{pa}))$ which is compatible with the filtration on $\mathbb{D}(\operatorname{Spec} k[t]/(t^a))$ given by $\mathscr{G}$. This corresponds to $\mathrm{Fil} \subset \mathbb{D}(\operatorname{Spec} k[t]/(t^{pa}))$, which is a free $k[t]/(t^{pa})$ sub-module of $\mathbb{D}(\operatorname{Spec} k[t]/(t^{pa}))$, which itself is a free $k[t]/(t^{pa})$-module. Moreover, any endomorphism β of $\mathscr{G}$ induces an endomorphism of the crystal $\mathbb{D}$, and therefore induces an endomorphism of $\mathbb{D}(\operatorname{Spec} k[t]/(t^{pa}))$. Finally, β deforms to an endomorphism of $\mathscr{G}'$ if and only if $\beta(\mathrm{Fil}) \subset \mathrm{Fil}$. Given that α induces an endomorphism of $\mathbb{D}(\operatorname{Spec} k[t]/(t^{pa}))$ (which need not preserve Fil), it follows that $p\alpha$ induces the zero map on $\mathbb{D}(\operatorname{Spec} k[t]/(t^{pa}))$, and thus tautologically preserves Fil whether or not α does. Therefore, it follows that if α is an endomorphism of $\mathscr{G}$ over $k[t]/(t^a)$, then $p\alpha$ lifts to any deformation of $\mathscr{G}$ to $\operatorname{Spec} k[t]/(t^{pa})$, which suggests that it is not possible to expect a much faster rate of decay than defined in Definition 5.1.

We now work in the setting of a p-divisible group $\mathcal{A}[p^\infty]/k[[t]]$. Let α denote an endomorphism of $\mathcal{A}[p^\infty]$ mod t, that extends to an endomorphism of $\mathcal{A}[p^\infty]$ mod t^a, but not t^{a+1}. The example considered above implies that $p\alpha$ extends to an endomorphism mod t^{pa}. However, Grothendieck–Messing theory cannot be naively applied to find an effective integer b (in terms of a) which has the property that $p\alpha$ does not extend to an endomorphism of $\mathcal{A}[p^\infty]$ mod t^b. Therefore, in order to prove Theorem 5.2, we use Kisin's description of the F-crystal $\mathbf{L}_{\mathrm{cris}}$, which controls the t-adic deformation of the special endomorphisms of $\mathcal{A}[p^\infty]$ mod t – see §5.4 for a sketch of how we proceed.

On the other hand, we remark that in some special cases when the generic point has extra endomorphisms (e.g., when the Kuga–Satake family over C is isomorphic to self-product of a family of elliptic curves), the work of Keating, based on the formal cohomology theory of Lubin–Tate and Drinfeld, yields the desired decay (see [Kea88, Thm. 1.1])[14], which gives another justification on the shape of h_r.

5.4. We give a rough idea of the proof of Theorem 5.2 here (see §§5.5-5.6 for details and references); we also provide a toy example of the explicit computation in this section. The reader should feel free to read this description and skip the details of our proof on a first reading. Consider a special endomorphism $w \in \mathcal{L}$. We give a criterion for w to not extend to an endomorphism of $\mathcal{A}[p^\infty]$ mod t^r in terms of the crystal $\mathbf{L}_{\mathrm{cris}}$. To that end, recall that R denotes the ring $W[[x_i, y_i, x'_j, y'_j]]$ (notation as in §4.8), and consider $w \in \mathcal{L} \subset \mathbb{L}$ as an element of $\mathbb{L} \otimes_W R = \mathbf{L}_{\mathrm{cris}}(R)$. Let $\widetilde{w} \in \mathbb{L} \otimes K[[x_i, y_i, x'_j, y'_j]]$ denote the unique horizontal continuation of w with respect to the connection on $\mathbf{L}_{\mathrm{cris}}$. The entries of the vector $\widetilde{w}$ are power series valued in $K[[x_i, y_i, x'_j, y'_j]]$, and the p-adic valuation of the coefficients of these power series go to $-\infty$. Loosely speaking, in order to understand whether w extends to a $k[t]/(t^r)$-point of R arising from a $k[[t]]$-point of R, we just need to restrict $\widetilde{w}$ to $\mathbb{L} \otimes_W K[[t]]$. This yields a vector with entries in the power series ring $K[[t]]$, and it suffices to understand the p-adic valuations of the coefficients of these power series. An explicit expression for $\widetilde{w}$ is given by $\lim_{n\to\infty} \mathrm{Frob}^n(w)$, where as in §4.8, Frob is the σ-linear Frobenius action on $\mathbf{L}_{\mathrm{cris}}$ (see [Kis10, §1.5.5]).

We illustrate this computation (which we carry out in full detail in the following pages) with a toy model. Consider $\mathrm{Frob} = (I+F)\circ\sigma$ with respect to a φ-invariant basis of $\mathbb{L}$, where $F = \begin{bmatrix} xy/p & x/p \\ y & 0 \end{bmatrix}$, $R = W[[x,y]]$, $\sigma(x) = x^p, \sigma(y) = y^p$ and when we restrict ourselves to C, we plug in x, y by certain power series $x(t), y(t) \in W[[t]]$, which are chosen based on $C \to \operatorname{Spf} k[[x,y]]$. Thus $\prod_{n=1}^{\infty}(I+F)\circ\sigma$ is an infinite summation of products of $F^{(i)} := \sigma^i(F)$. Consider $w = [1,0]^T$ (with respect to the chosen basis of $\mathbb{L}$), then a direct computation of matrix products implies that for the first coordinate of $\widetilde{w}$, among all the terms with p-adic valuation $-(r+1)$, the unique term with minimal t-adic valuation is $\prod_{i=1}^{r+1} \sigma^{i-1}(xy/p) = (xy)^{1+p+\cdots+p^r}/p^{r+1}$. This observation allows us to prove the Decay Lemma.

The setup. The setup and the first reduction steps in the proof of Theorem 5.2 is the same as that in the proof of [MST, Thm. 5.1.2] in [MST, §5.1]. We briefly introduce the notation for the proof of Theorem 5.2 here and the reader may see [MST] for more details.

5.5. Recall from §4.8 that $\widehat{\mathcal{O}}_{\mathcal{S},P} = \operatorname{Spf} W[[x_1, \ldots, x_{n-1}, y_1, \ldots, y_{n-1}, x'_1, \ldots, x'_m, y'_1, \ldots, y'_m]]$ when $\mathcal{L}_1$ is split. Since P is non-superspecial, we have $n \geq 2$ through out this section.

The formal curve C gives rise to the tautological ring homomorphism

$$W[[x_1, \ldots, x_{n-1}, y_1, \ldots, y_{n-1}, x'_1, \ldots, x'_m, y'_1, \ldots, y'_m]] \to k[[t]],$$

[14] We would like to thank the anonymous referee for pointing out this reference to us.

and we denote by $x_i(t)$ (respectively $x'_i(t), y_i(t), y'_i(t)$) the images of x_i (respectively x'_i, y_i, y'_i) in $k[[t]]$. For each $x_i(t)$ (respectively $x'_i(t), y_i(t), y'_i(t)$), let $X_i(t)$ (respectively $X'_i(t), Y_i(t), Y'_i(t)$) denote the power series in $W[[t]]$ whose coefficients are the Teichmuller lifts of the coefficients of $x_i(t)$ (respectively $x'_i(t), y_i(t), y'_i(t)$). We define $Y_n(t) = -\sum_{i=1}^{n-1} X_i(t)Y_i(t) - \sum_{j=1}^{m} X'_j(t)Y'_j(t)$, and define $Y_{n+1}(t) = -\sum_{i=1}^{m}(X'_i(t)(Y'_i(t))^p + (X'_i(t))^p Y'_i(t))$. Let $a_i = v_t(Y_i(t))$ for $1 \leq i \leq n+1$, where v_t denotes the function of taking t-adic valuation. By Corollary 4.10, since P is a non-superspecial supersingular point, the local equation of the non-ordinary locus is given by $y_1(t) = 0$ and hence $h = v_t(y_1(t)) = v_t(Y_1(t)) = a_1$.

5.6. We now relate the lift of special endomorphisms to explicit computations of powers of the Frobenius matrix given in §4.11. For $s \in \mathbb{Z}_{\geq 0}$, let D_s denote the p-adic completion of the PD enveloping algebra of the ideal (t^s, p) in $W[[t]]$ and let ι_s denote the map $\operatorname{Spec} k[t]/(t^s) \to \operatorname{Spf} k[[t]] \to \operatorname{Spf} R \otimes_W k$. By de Jong's theory [dJ95, §2.3], if $w \in \mathcal{L}$ lifts to a special endomorphism of $\mathcal{A}$ mod t^s, then it gives rise to a horizontal section in the Dieudonné module $(\iota_s^* \mathbf{L}_{\text{cris}})(D_s)$. Thus, in order to find the largest s such that w lifts to $k[t]/(t^s)$, we first compute the horizontal section $\tilde{w}_s \in (\iota_s^* \mathbf{L}_{\text{cris}})(D_s) \otimes_W K$ whose restriction to the fiber $t = 0$ equals w and then study the p-adic integrality of $\tilde{w}_s$.

Here we recall the construction of $\tilde{w}_s$ following [Kis10, §1.5.5] and the rest of this section is devoted to the study of the p-adic integrality. More precisely, Kisin constructed a Frobenius stable section $\tilde{w}$ in $\mathbb{L} \otimes_W K[[x_i, y_i, x'_j, y'_j]]$ whose restriction to $\mathbb{L}$ (the fiber at P) is a given φ-invariant vector $w \in \mathbb{L}$; moreover, since Frob is a horizontal morphism, he concluded that $\tilde{w}$ is horizontal. In our setting, since the connection on $\iota_s^* \mathbf{L}_{\text{cris}}$ is the pullback of the connection on $\mathbf{L}_{\text{cris}}$, the horizontal section $\tilde{w}_s$ is the pullback via $R \to W[[t]] \to D_s$ of the horizontal section $\tilde{w} \in \mathbb{L} \otimes_W K[[x_i, y_i, x'_j, y'_j]]$.

Let F' denote the matrix $S'(u' - I)(S')^{-1}$ in §4.11; with respect to the basis $\{e_i, f_i, e'_j, f'_j\}$, $\text{Frob} = (I + F) \circ \sigma$, where σ is defined in §4.8.. Let $F'^{(i)}$ denote the i-th σ-twist of F; more precisely, $F'^{(i)}$ is given by $\sigma^i(S'(u' - I)(S')^{-1})$. Let

$$F'_\infty = \prod_{i=0}^{\infty}(I + F'^{(i)}) \in M_{2n+2m}(K[[x_i, y_i, x'j, y'_j]]).$$

We define F_∞ by substituting $X_i(t)$ (resp. $X'_j(t), Y_i(t), Y'_j(t)$) for x_i (resp. x'_j, y_i, y'_j) in F'_∞. More explicitly, let $F^{(i)}$ be the matrix obtained by substituting $X_i(t)$ (resp. $X'_j(t), Y_i(t), Y'_j(t)$) for x_i (resp. x'_j, y_i, y'_j) in $F'^{(i)}$. In other words, we first compute the Frobenius twist of F', and *only then* substitute the power series in t for the variables x_i, y_i, x'_j, y'_j.[15] So we obtain

$$F_\infty = \prod_{i=0}^{\infty}(I + F^{(i)}) \in M_{2n+2m}(K[[t]]).$$

This product is well-defined and the $\mathbb{Q}_p$-span of the columns of F_∞ are vectors in $\mathbb{L} \otimes_W K[[t]]$ which are Frob-invariant and horizontal. Let ι denote the map $\operatorname{Spf} k[[t]] \to \operatorname{Spf} R \otimes_W k$; thus $\iota^* \tilde{w} \in \mathbb{L} \otimes_W K[[t]] \cong K[[t]]^{2m+2n}$ is given by $F_\infty w$, where we write w as a column vector with respect to the basis $\{e_i, f_i, e'_j, f'_j\}$. The horizontal section $\tilde{w}_s$ is $F_\infty w$, which indeed lies in $(\iota_s^* \mathbf{L}_{\text{cris}})(D_s) \otimes_W K \subset \mathbb{L} \otimes_W K[[t]]$. One way to see this claim is that by §5.3, for each $w \in \mathcal{L}$, there exists $N \gg 1$ (depending on s, w) such that $p^N w$ extends to an endomorphism of $\mathcal{A}[p^\infty]$ mod t^s and thus by de Jong's theory, $p^N \tilde{w}_s \in (\iota_s^* \mathbf{L}_{\text{cris}})(D_s)$ and thus $\tilde{w}_s \in (\iota_s^* \mathbf{L}_{\text{cris}})(D_s) \otimes_W K$.

[15] For more details, see [MST, Proof of Thm. 5.1.2 assuming Prop. 5.1.3].

In order to show that w decays rapidly, it suffices to show that for every r, we have that $p^r\tilde{w}_{h_r+1}$ does not lie in $\mathbf{L}_{\mathrm{cris}}(D_{h_r+1})$. More precisely, since for every $N < p(h_r+1)$, we have $t^N/p \notin D_{h_r+1}$, it suffices to show that there exists an entry of $\iota^*\tilde{w}$ (viewed as a power series in $K[[t]]$) which has a term of form ct^N/p^{r+2} with $N < p(h_r+1)$ and $c \in W^\times$.[16] Thus in what follows, we use the definition of F_∞ to expand $F_\infty w$ into an infinite sum of vectors in $K[[t]]^{2m+2n}$ whose entries are monomials in $X_i(t), Y_i(t), X'_j(t), Y'_j(t)$ and will find the vector with minimal t-adic valuation in the expansion of $F_\infty w$ among all vectors with p-adic valuation $-r$. Note that by the t-adic and p-adic valuations of a vector, we mean the minimal t-adic/p-adic valuation of all entries.

The terms in F_∞ with minimal t-adic valuation among ones with fixed p-adic valuation. In order to prove Theorem 5.2, it suffices to work with $F_\infty(1)$, the top-left $2n\times 2n$ block of F_∞ (see the first paragraph of the proof of Theorem 5.2 right after Lemma 5.13 for details) and in what follows, we compute $F_\infty(1)$ explicitly.

5.7. Let A (resp. C, D) denote the top-left $2n\times 2n$ (resp. top-right $2n\times m$ and bottom-left $m\times 2n$) block of $u'-I$ in §4.11 with $X_i(t)$ (resp. $X'_j(t), Y_i(t), Y'_j(t)$) substituted in place of x_i (resp. x'_j, y_i, y'_j). Thus

$$F = \begin{bmatrix} S'_0 A S'^{-1}_0 & S'_0 C \\ D S'^{-1}_0 & 0 \end{bmatrix}.$$

For $i < n$, we let A_i denote the $2n\times 2n$ matrix with the $n, n+i$ entry equal to $p^{-1}Y_i(t)$, the $i, 2n$ entry equalling $-p^{-1}Y_i(t)$ and all other entries equal to 0. We let A_n denote the matrix with zeros everywhere except for the $n, 2n$ entry, which equals $p^{-1}Y_n(t)$. Let K_i equal $S'_0 A_i (S'_0)^{-1}$. Let $A_{n+1} = CD^{(1)}$ and $K_{n+1} = S'_0 A_{n+1}((S'_0)^{-1})^{(1)}$. Note that A_{n+1} is a $2n\times 2n$ matrix with zeros everywhere except for the $n, 2n$-entry which equals $p^{-1}Y_{n+1}(t)$.

For brevity, let $F(1)$ denote the top-left $2n\times 2n$ block of F. We have $F(1) = \sum_{i=1}^{n} K_i + B(x)$, where $B(x)$ involves only the X_i and has p-adically integral coefficients since $S'_0 \in \mathrm{GL}_{2n}(W)$. We further break $B(x)$ into $B(x) = B_1(x) + B_2(x)$, where the (i,j) entry of $S'^{-1}_0 B_1(x) S'_0$ is zero unless $i = n, j < n$, in which case the entry is x_j, and the (i,j) entry of $S'^{-1}_0 B_2(x) S'_0$ is zero unless $j = 2n, n < i < 2n-1$, in which case the entry is $-x_{n-i}$. Moreover, we observe that $F_\infty(1)$ is made up of sums of finite products of σ-twists of $B_1(x)$, $B_2(x)$ and $K_i, i = 1, \ldots, n+1$.

The following two lemmas identify the nonzero products of σ-twists of $K_i, i = 1, \ldots, n+1$. For brevity, let $R_1 \ldots R_{2n}$ denote the rows of the matrix $(S'_0)^{-1}$. Since $(S'_0)^{-1} \in \mathrm{GL}_{2n}(W)$, we have that $\{\overline{R}_i\}_{i=1}^{2n}$ is a basis of k^{2n}, where we use $\overline{R}_i$ to denote $R_i \bmod p$.

Lemma 5.8. *We have $\sigma(R_i) = R_{i+1}$ for $n+1 \le i \le 2n-1$; $\sigma(R_{2n}) = R_1$, $\sigma(R_i) = R_{i+1} - b_i R_1$ for $1 \le i \le n-1$, and $\sigma(R_n) = R_{n+1} + \sum_{i=1}^{n-1} b_i R_{n+i+1}$, where $b_i \in W$ in Lemma 4.5.*

Proof. Recall from §4.11 that $B_0 = S_0^{-1}\sigma(S_0)$ and thus by §4.6, we have $\sigma((S'_0)^{-1}) = (B'_0)^{-1}(S'_0)^{-1}$, where $B'_0 = \left[\begin{array}{c|c} p^{-1}I & 0 \\ \hline 0 & I \end{array}\right] B_0 \left[\begin{array}{c|c} pI & 0 \\ \hline 0 & I \end{array}\right]$. We then obtain the assertions by a direct computation using Lemma 4.5(3). □

Lemma 5.9. *Notation as in §§5.5,5.7. All Frobenius twists below are defined in the same way as $F^{(i)}$: first applying σ^i to x_i, y_i, x'_j, y'_j and then substitute $X_i(t), Y_i(t), X'_j(t), Y'(t)_j$ for x_i, y_i, x'_j, y'_j. We also view v_i, w_i in Lemma 4.5 as vectors $v_i \otimes 1, w_i \otimes 1$ in $K[[t]]^{2n} = \mathbb{L}_0 \otimes_W K[[t]]$, where this identification uses the φ-invariant basis $\{e_i, f_i\}_{i=1}^n$ of $\mathbb{L}_0 \otimes_W K$ in Lemma 4.3.*

[16] Here we have p^{r+2} instead of p^{r+1} is due to the fact that $\mathrm{Span}_W\{e_i, f_i, e'_j, f'_j\} \neq \mathbb{L}$ but we still have $\mathrm{Span}_W\{e_i, f_i, e'_j, f'_j\} \supset p\mathbb{L}$.

(1) For $i, j, l \in \mathbb{Z}_{>0}$ such that $i, j, l \leq n+1$ and $l \leq i$, the matrix $K_iK_j^{(l)} = 0$ unless $i = l$. Moreover, the image of $K_iK_j^{(i)}$ is $\mathrm{Span}_{K[[t]]}\{v_n\}$ and $K_iK_j^{(i)} = S_0'M$ where $M \in M_{2n}(K[[t]])$ is the matrix with its n^{th} row being $Y_iY_j^{(i)}p^{-2}\sigma^{i+j-1}(R_{n+1})$ and all other rows being 0.

(2) For $i_1, i_2, \ldots, i_l \in \mathbb{Z}_{\geq 1}$ such that $i_1, i_2, \ldots, i_l \leq n+1$, the image of the matrix $\prod_{j=1}^{l} K_{i_j}^{(i_1+\cdots+i_{j-1})}$ is $\mathrm{Span}_{K[[t]]}\{v_n\}$. Moreover, $\prod_{j=1}^{l} K_{i_j}^{(i_1+\cdots+i_{j-1})} = S_0'M$ where $M \in M_{2n}(K[[t]])$ is the matrix with its n^{th} row being $(\prod_{j=1}^{l} Y_{i_1}^{(i_1+\cdots i_{j-1})})p^{-l}\sigma^{i_1+i_2+\cdots+i_l-1}(R_{n+1})$ and all other rows being 0.

Proof. (1) By §5.7, $\ker K_i = \mathrm{Span}_{K[[t]]}\{v_1, \ldots, v_n, w_1, \ldots, w_{i-1}, w_{i+1}, \cdots, w_{n-1}\}$ for $1 \leq i < n$, and $\ker K_n = \mathrm{Span}_{K[[t]]}\{v_1, \ldots, v_n, w_1, \ldots, w_{n-1}\}$, $\ker K_{n+1} = \mathrm{Span}_{K[[t]]}\{\varphi(v_1), \ldots, \varphi(v_n), \varphi(w_1) \ldots \varphi(w_{n-1})\}$. Note that if we view v_i, w_i as vectors in K^{2n} by using the basis $\{e_i, f_i\}_{i=1}^{n}$, then $\varphi(v_i), \varphi(w_i)$ are just applying σ to all coordinates. On the other hand, $\mathrm{im}\, K_j = \mathrm{Span}_{K[[t]]}\{v_j, v_n\}$ for $1 \leq j \leq n$ and $\mathrm{im}\, K_{n+1} = \mathrm{Span}_{K[[t]]}\{v_n\}$; hence by Lemma 4.5(3), for $1 \leq l \leq n$,

$$\mathrm{im}\, K_j^{(l)} = \varphi^l(\mathrm{im}\, K_j) \subset \mathrm{Span}_{K[[t]]}\{v_1, \ldots, v_n, w_1, \ldots, w_l\},$$

and $\mathrm{im}\, K_j^{(n+1)} \subset \mathrm{Span}_{K[[t]]}\{\varphi(v_1), \ldots, \varphi(v_n), \varphi(w_1), \ldots, \varphi(w_n)\}$. Therefore, $K_iK_j^{(l)} = 0$ if $l < i$.

Suppose now that $l = i$. Then $\mathrm{im}\, K_iK_j^{(i)} = \mathrm{Span}_{K[[t]]}\{K_iw_i\} = \mathrm{Span}_{K[[t]]}\{v_n\}$ for $i \leq n$ and $\mathrm{im}\, K_{n+1}K_j^{(n+1)} = \mathrm{Span}_{K[[t]]}\{K_{n+1}\varphi(w_n)\} = \mathrm{Span}_{K[[t]]}\{v_n\}$. Thus the matrix M has only its n^{th} row non-zero. We now compute the n^{th} row of M. For $i, j \leq n$, note that $M = A_i(S_0')^{-1}(S_0')^{(i)}A_j^{(i)}((S_0')^{-1})^{(i)}$; if $i = n+1$ or $j = n+1$, the matrix M is given by the same formula once we replace the $(S_0')^{-1}$ after A_i or A_j by $((S_0')^{-1})^{(1)}$. For $j \leq n$, the product $A_j^{(i)}((S_0')^{-1})^{(i)}$ has only its j^{th} and n^{th} rows non-zero (if $j = n$, then only the n^{th} row is non-zero), and these rows equal $-Y_j^{(i)}p^{-1}\sigma^i(R_{2n})$ and $Y_j^{(i)}p^{-1}\sigma^i(R_{n+j})$ respectively (if $j = n$, the row $Y_n^{(i)}p^{-1}\sigma^i(R_{2n})$); and $A_{n+1}^{(i)}((S_0')^{-1})^{(i+1)}$ only has its n^{th} row non-zero and its n^{th} row is $Y_{n+1}^{(i)}p^{-1}\sigma^{i+1}(R_{2n})$. Similarly, for $i \leq n$, the n^{th} row in the product $A_i(S_0')^{-1}$ equals $Y_ip^{-1}R_{n+i}$; the n^{th} row in $A_{n+1}((S_0')^{-1})^{(1)})$ equals $Y_{n+1}p^{-1}\sigma(R_{2n}) = Y_{n+1}p^{-1}R_1$ by Lemma 5.8. For $j < n$, by the above computation, we write $A_j^{(i)}((S_0')^{-1})^{(i)}$ as $(-Y_j^{(i)}p^{-1})e_j\sigma^i(R_{2n}) + (Y_j^{(i)}p^{-1})e_n\sigma^i(R_{n+j})$, where e_j (resp. e_n) is the column vector with all coordinates 0 expect the j^{th} (resp. n^{th}) coordinates being 1. Then

$$(S_0')^{(i)}A_j^{(i)}((S_0')^{-1})^{(i)} = (-Y_j^{(i)}p^{-1})\varphi^i(pv_j)\sigma^i(R_{2n}) + (Y_j^{(i)}p^{-1})\varphi^i(pv_n)\sigma^i(R_{n+j}).$$

By definition of R_{n+i}, R_1, we have $R_{n+i}v_j = 0 = R_1w_j$, $R_{n+i}w_j = 0$ for $j \neq i$, $R_{n+i}w_i = 1$, $R_1v_j = 0$ for $j > 1$, and $R_1(pv_1) = 1$. For $i \leq n$, the coefficient of w_i in $\varphi^i(pv_j)$ (resp. $\varphi^i(pv_n)$) is 0 (resp. 1) by Lemma 4.5(3) and thus the n^{th} row of M is $Y_iY_j^{(i)}p^{-2}\sigma^i(R_{n+j}) = Y_iY_j^{(i)}p^{-2}\sigma^{i+j-1}(R_{n+1})$ by Lemma 5.8. The cases when $i = n+1$ or $j = n, n+1$ also follow from Lemma 4.5 and Lemma 5.8 by direct computations as above.

(2) We prove by induction on l. Indeed, we only need to verify the expression of M. The base case is just (1). We assume that $i_1 \leq n$ and the case $i_1 = n+1$ follows by a similar computation. Note that

$$M = A_{i_1}(S_0')^{-1}(\prod_{j=2}^{l} K_{i_j}^{(i_2+\cdots+i_{j-1})})^{(i_1)} = A_{i_1}(S_0')^{-1}(S_0')^{(i_1)}(Y_{i_2}^{(i_1)} \cdots Y_{i_l}^{(i_1+\cdots+i_{l-1})}p^{1-l})e_n\sigma^{i_1+\cdots+i_l-1}(R_{n+1})$$

by the induction hypothesis. By the computation in (1), we have $A_{i_1}(S_0')^{-1}$ has only its i_1^{th} and n^{th} rows non-zero (if $i_1 = n$, then only the n^{th} row is non-zero), and these rows equal $-Y_{i_1}p^{-1}R_{2n}$ and

$Y_{i_1}p^{-1}R_{n+i_1}$ respectively; moreover, by Lemma 4.5 and the definition of S'_0, we have $(S'_0)^{(i_1)}e_n = pv_{n+i_1}$, which is a linear combination of $w_1, \dots, w_{i_1}$ with the coefficient of w_{n+i_1} being 1; then by the defintion of R_{n+i_1} and R_{2n}, we have $R_{n+i_1}(pv_{n+i_1}) = R_{n+i_1}(w_{i_1}) = 1$ and $R_{2n}(pv_{n+i_1}) = 0$. Therefore we have $A_{i_1}(S'_0)^{-1}(S'_0)^{(i_1)}e_n = (Y_{i_1}p^{-1})e_n$ and the assertion follows. □

Recall from §5.7 that $F_\infty(1)$ is an infinite sum of finite products of σ-twists of $B(x)$ and $K_i, i = 1, \dots, n+1$.[17] The following lemma picks out which of these finite products have the minimal t-adic valuation among those with a fixed p-adic valuation. We observe that σ-twists of $B_1(x), B_2(x)$ have non-negative p-adic valuation and positive t-adic valuation; therefore if we have a finite product of σ-twists of $B_1(x), B_2(x)$ and K_i which shows up in $F_\infty(1)$, then by removing all the σ-twists of $B(x)$ and lowering the twist degrees for K_i, we still obtain a product which shows up in $F_\infty(1)$ and have smaller t-adic valuation.

We illustrate this with two examples. Consider the term $K_1B_1(x)^{(1)}K_2^{(2)}K_5^{(4)}$, and suppose that it is non-zero. Then, the term $K_1K_2^{(1)}K_5^{(3)}$ is non-zero (this follows as the image of $B_1(x)$ is contained in the image of K_2, and in fact equals the image of $K_2K_5^{(2)}$), and visibly has smaller t-adic valuation, while the p-adic valuation is the same. Indeed, this argument works for any sub-product that starts with a twist of $B_1(x)$ and has exactly one occurrence of a twist of K_j, which appears at the end of this sub-product – the term that replaces this sub-product will be an appropriate twist of K_j.

Similarly, by considering row-spans instead of images, it follows that if a term that looks like $K_2B_2(x)^{(i)}K_3^{(j)}$ is non-zero, then so is $K_2^{(i)}K_3^{(j)}$ as the row span of K_2 (in fact, of every K_j for every j) contains the row span of B_2. And so, in this case, we replace $K_2B(x)^{(i)}K_3^{(j)}$ with $K_2K_3^{(j-i)}$. As above, this argument works for any sub-product that ends with a twist of $B_2(x)$, and has exactly one occurrence of some twist of some K_i, which appears at the beginning of this sub-product. Thus, it suffices to exclusively consider products of σ-twists of the K_i.

5.10. We introduce some notation for the lemmas. For $r \in \mathbb{Z}_{>0}$, define $\mathbb{I}_r = \{1, 2, \dots, n+1\}^r$. For $I = (i_1, \dots, i_r) \in \mathbb{I}_r$, define $P_I = K_{i_1} \cdot K_{i_2}^{(i_1)} \cdot K_{i_3}^{(i_1+i_2)} \cdots K_{i_r}^{(i_1+i_2+\cdots+i_{r-1})}$ and define the *weight* of I, denoted by μ_I, to be $\sum_{j=1}^r i_j$. By Lemma 5.9, we write $P_I = S'_0 M_I$ and note that all nonzero entries in P_I have the same t-adic valuation $\sum_{j=1}^r p^{i_1+\cdots+i_{j-1}}a_{i_j} =: \nu_I$ (recall that $a_i = v_t(Y_i)$).

Note that in the expansion of $F_\infty(1)$ into an infinite sum of finite products of σ-twists of $B(x)$ and $K_i, i = 1, \dots, n+1$, among all such finite products with a p-adic valuation $-r$, the ones with the minimal t-adic valuation have to be of the form P_I with $I \in \mathbb{I}_r$. Let ν_r denote this minimal t-adic valuation. Define $\mathbb{I}_r^{\min} = \{I \in \mathbb{I}_r : v_t(P_I) = \nu_r\}$. In other words, among all finite products of σ-twists of $B(x)$ and K_i with p-adic valuation $-r$ in the expansion of $F_\infty(1)$, the ones with minimal t-adic valuations are $P_I, I \in \mathbb{I}_r^{\min}$. The following lemma provides some information of the set $\mathbb{I}_r^{\min}$.

Lemma 5.11. *Notation as in §5.10 and let $I = (i_1, \dots, i_r) \in \mathbb{I}_r^{\min}$. Then:*

(1) $(i_2, \dots, i_r) \in \mathbb{I}_{r-1}^{\min}$. Conversely, if $(j_2, \dots j_r) \in \mathbb{I}_{r-1}^{\min}$, then there exists $j_1 \in \{1, \dots, n+1\}$ such that $(j_1, j_2, \dots, j_r) \in \mathbb{I}_r^{\min}$.

(2) $i_1 \le i_2 \le \cdots \le i_r$ and $a_{i_1} \ge a_{i_2} \ge \cdots \ge a_{i_r}$.

(3) Let $J = (j_1, \dots j_r) \in \mathbb{I}_r^{\min}$. Then all $(l_1, \dots, l_r) \in \mathbb{I}_r^{\min}$ must have each l_α to be either i_α or j_α for $1 \le \alpha \le r$.

(4) Let J be as in (3). Then $|\mu_I - \mu_J| < n+1$.

(5) Suppose that $|\mathbb{I}_r^{\min}| > 1$. Then there exist two elements in $\mathbb{I}_r^{\min}$ with different weights. Further, there is a unique $I \in \mathbb{I}_r^{\min}$ with maximal weight, and a unique $J \in \mathbb{I}_r^{\min}$ with minimal weight.

[17] Note that we also consider $B(x)$ and K_i as the zeroth σ-twist of themselves.

Proof. (1) By §5.10, we have $\nu_I = a_{i_1} + p^{i_1}\nu_{(i_2,\dots,i_r)}$, and thus $\nu_{(i_2,\dots,i_r)}$ has to be minimized in order for ν_I to be minimal. On the other hand, take $j_1 = i_1$ and then we have $\nu_{(j_1,\dots,j_r)} = a_{i_1} + p^{i_1}\nu_{(j_2,\dots,j_r)} = a_{i_1} + p^{i_1}\nu_{(i_2,\dots,i_r)} = \nu_r$.

(2) We prove the assertion by induction on r. By the inductive hypothesis and (1), we may assume that $i_2 \le i_3 \le \cdots i_r$ and $a_{i_2} \ge a_{i_3} \ge \cdots \ge a_{i_r}$.

Assume for contradiction that $i_1 > i_2$. If $a_{i_2} \le a_{i_1}$, then $\nu_{(i_2,i_2,i_3,\dots i_r)} < \nu_I$, which contradicts with $I \in \mathbb{I}_r^{\min}$. Therefore $a_{i_1} < a_{i_2}$. Let $I' = (i_2, i_1, i_3, \dots, i_r)$. We have $\nu_{I'} - \nu_I = \nu_{(i_2,i_1)} - \nu_{(i_1,i_2)} = a_{i_1}(p^{i_2} - 1) - a_{i_2}(p^{i_1} - 1) < 0$. Thus we must have $i_1 \le i_2$.

Now assume for contradiction that $a_{i_1} < a_{i_2}$. Then $\nu_{(i_1,i_1,i_3,\dots i_r)} < \nu_I$. Thus $a_{i_1} \ge a_{i_2}$ as required.

(3) Suppose that $J = (j_1, \dots, j_r)$. By (1), it follows that $\nu_{(j_2,\dots,j_r)} = \nu_{(i_2,\dots,i_r)} = \nu_{r-1}$, whence $a_{i_1} + p^{i_1}\nu_{r-1} = a_{j_1} + p^{j_1}\nu_{r-1}$. It follows that $(i_1, j_2, j_3, \dots, j_r), (j_1, i_2, \dots, i_r) \in \mathbb{I}_r^{\min}$. (3) now follows by induction on r.

(4) $\mu_I - \mu_J = i_1 - (\sum_{\alpha=1}^{r-1}(j_\alpha - i_{\alpha+1})) - j_r$. By (2) and (3), $j_\alpha \le i_{\alpha+1}$; since $j_r > 0$, then $\mu_I - \mu_J < i_1 \le n+1$. Similarly, $\mu_J - \mu_I < n+1$; thus the result follows.

(5) For $1 \le \alpha \le r$, set $M_\alpha = \max_{J\in\mathbb{I}_r^{\min}}\{j_\alpha\}$ and $m_\alpha = \min_{J\in\mathbb{I}_r^{\min}}\{j_\alpha\}$. By applying (3) repeatedly the set of all $J \in \mathbb{I}_r^{\min}$, it follows that $(M_1, \dots, M_r), (m_1, \dots, m_r) \in \mathbb{I}_r^{\min}$. By definition, $(M_1, \dots, M_r)$ is the unique element of $\mathbb{I}_r^{\min}$ with maximal weight, and $(m_1, \dots, m_r) \in \mathbb{I}_r^{\min}$ is the unique element with minimal weight. □

Other preparation lemmas. Recall that $\overline{R}_i$ denote R_i mod p.

Lemma 5.12. *For any $0 \neq v \in \mathbb{F}_p^{2n}$, we have $\overline{R}_{n+1}v \neq 0$. Consequently, if $\{z_1, \dots, z_{2n}\}$ is a basis of $\mathbb{F}_p^{2n}$, then $\overline{R}_{n+1}z_1, \dots, \overline{R}_{n+1}z_{2n} \in k$ are linearly independent over $\mathbb{F}_p$.*

Proof. If $\overline{R}_{n+1}v = 0$, then $\sigma^i(\overline{R}_{n+1})v = 0$ for all $i \ge 0$. By Lemma 5.8, $\mathrm{Span}_W\{\sigma^i(R_{n+1})\}_{i=0}^{2n-1} = W^{2n}$; thus $\overline{R}_{n+1}, \sigma(S_{n+1}), \dots, \sigma^{2n-1}(\overline{R}_{n+1})$ form a basis of k^{2n}. Therefore, if $\overline{R}_{n+1}v = 0$, then $v = 0$; this proves the first assertion.

For the second assertion, suppose there exists a non-trivial linear relation $\sum_{i=1}^{2n} a_i(\overline{R}_{n+1}z_i) = 0$ with $a_i \in \mathbb{F}_p$. Then, $\overline{R}_{n+1}(\sum_{i=1}^{2n} a_i z_i) = 0$, which contradicts the first assertion. □

Lemma 5.13. *Let $\alpha_0, \dots, \alpha_n \in k$ such that $(\alpha_0, \dots, \alpha_n) \neq (0, \dots, 0)$. Consider the linear combination $\overline{R} = \sum_{i=0}^{n} \alpha_i\sigma^i(\overline{R}_{n+1})$. Then $\dim_{\mathbb{F}_p}\{v \in \mathbb{F}_p^{2n} \mid \overline{R}v = 0\} \le n$.*

Proof. For any $z \in \mathbb{F}_p^{2n}$, note that $\overline{R}z = \vec{\alpha}\vec{\beta}(z)$, where

$$\vec{\alpha} = (\alpha_0, \dots, \alpha_n),\ \vec{\beta}(z) = \begin{bmatrix} \overline{R}_{n+1}z \\ (\overline{R}_{n+1}z)^{(1)} \\ \vdots \\ (\overline{R}_{n+1}z)^{(n)} \end{bmatrix}.$$

Now assume for contradiction that there exist linearly independent vectors $z_1, z_2, \dots, z_{n+1} \in \mathbb{F}_p^{2n}$ such that $\overline{R}z_j = 0$. This implies that $\vec{\alpha}\vec{\beta}(z_j) = 0$ for every $1 \le j \le n+1$. In particular, this implies that the row vector $\vec{\alpha}$ is in the (left) kernel of the $(n+1)\times(n+1)$ matrix $M(z)$ whose j^{th} column is $\vec{\beta}(z_j)$. This contradicts the assumption $\vec{\alpha} \neq 0$ once we show that $M(z)$ is invertible. Indeed, note that the $(i+1)$th row of $M(z)$ is the Frobenius twist of the i^{th} row, and hence $M(z)$ is a Moore matrix. The determinant of a Moore-matrix vanishes if and only if the entries of the first row are linearly independent over $\mathbb{F}_p$. However, the first row of $M(z)$ consists of the elements $\{\overline{R}_{n+1}z_i\}_{i=1}^{n+1}$, and by Lemma 5.12 these elements are linearly independent over $\mathbb{F}_p$. The lemma follows. □

Decay in the non-superspecial case.

Proof of Theorem 5.2. We follow the argument in §5.6. Let w be a primitive vector (that is, not a multiple of p) in $\mathrm{Span}_{\mathbb{Z}_p}\{e_1,\dots,e_n,f_1,\dots,f_n\}\subset\mathcal{L}$. With respect to the basis $\{e_i,f_i,e'_j,f'_j\}$, we view w as a vector in W^{2n+2m}, which has the last $2m$ coordinates being 0. Let w_0 denote the vector in W^{2n} whose coordinates are the first $2n$ coordinates of w (indeed, as vectors in $\mathcal{L}$, $w=w_0$). Then for any $r,s\in\mathbb{Z}_{\geq 0}$, if $p^rF_\infty(1)w_0$ is not integral in $\mathbb{L}_0\otimes_W D_s$, then neither is $p^rF_\infty w$ in $\mathbb{L}_0\otimes_W D_s\oplus\mathbb{L}_1\otimes_W D_s=\mathbb{L}\otimes_W D_s=\iota_s^*(\mathbf{L}_{\mathrm{cris}})(D_s)$. Thus to prove the Decay Lemma, it suffices to work with $F_\infty(1)$. Thus for a general $\mathbb{L}_1$, we may embed it into the split case as described in §4.7 to reduce the proof to the split case because such an embedding only changes $\mathbb{L}_1$ part and the $\mathbb{L}_0$ part remains the same; more precisely, the decay vectors that we choose lie in $\mathbb{L}_0$ and the computation only considers the projection of $F_\infty w$ into $\mathbb{L}_0\otimes D_s$ and thus the same argument applies.

Notation as in §5.10; let M_{r+1} denote the matrix in $M_{2n}(K)$ such that

$$\sum_{I\in\mathbb{I}^{\min}_{r+1}}P_I=S'_0\sum_{I\in\mathbb{I}^{\min}_{r+1}}M_I=t^{\nu_{r+1}}S'_0M_{r+1}+t^{\nu_{r+1}+1}N_{r+1},\ \text{for some } N_{r+1}\in M_{2n}(K[[t]]).$$

By definition, $p^{r+1}M_{r+1}\in M_{2n}(W)$. First we follow the reduction step as in the last paragraph of the proof of Thm. 5.1.2 assuming Prop. 5.1.3 in [MST]. Note that, because $S'_0\in\mathrm{GL}_{2n}(W)$, we have that $\ker(p^{r+2}M_{r+2} \bmod p)=\ker(p^{r+2}S'_0M_{r+2} \bmod p)$. If $w_0 \bmod p\notin\ker(p^{r+2}M_{r+2} \bmod p)=\ker(p^{r+2}S'_0M_{r+2} \bmod p)$, then the coefficient of $t^{\nu_{r+2}}$ in $p^{r+2}F_\infty(1)w_0$ modulo $p\mathcal{L}_0\otimes_{\mathbb{Z}_p}W$ is given by $p^{r+2}S'_0M_{r+2}w_0 \bmod p\not\equiv 0$. In other words, the coefficient of $t^{\nu_{r+2}}$ in $p^rF_\infty(1)w_0$ does not lie in $p^{-1}\mathcal{L}_0\otimes_{\mathbb{Z}_p}W$. Since $\mathbb{L}_0\subset p^{-1}\mathcal{L}_0\otimes_{\mathbb{Z}_p}W$, then the coefficient of $t^{\nu_{r+2}}$ in $p^rF_\infty(1)w_0$ does not lie in $\mathbb{L}_0$. For $s>\nu_{r+2}/p$, by the definition of D_s, we have that $p^{-1}t^{\nu_{r+2}}\notin D_s$ and then $p^rF_\infty w_0\notin\mathbf{L}_{\mathrm{cris}}(D_s)$. Thus the following claim implies the Decay lemma.

Claim. (1) There exists a saturated $\mathbb{Z}_p$-submodule $\Lambda\subset\mathcal{L}$ of rank at least n such that if $v\in\Lambda$ is a primitive vector, then $v \bmod p\notin\ker(p^{r+2}M_{r+2} \bmod p)$ for all $r\in\mathbb{Z}_{\geq 0}$.

(2) For all $r\in\mathbb{Z}_{\geq 0}$, we have $\nu_{r+2}/p<h_r+1$, where h_r is defined in Definition 5.1.

For (1), by Lemma 5.9(2), each $p^{r+2}M_I\in M_{2n}(K[[t]])$ with $I\in\mathbb{I}^{\min}_{r+2}$ has only its n^{th} row non-zero and its n^{th} row has the form

$$c_It^{\nu_{r+2}}\sigma^{\mu_I-1}(R_{n+1})+(\text{higher powers of } t)\in K[[t]]^{2n},$$

where $c_I\in W^\times$ as it is the product of twists of leading coefficients of Y_{i_j} (these leading coeffcients are all Teichmuller lifts of non-zero elements in k); therefore, summing over all $I\in\mathbb{I}^{\min}_{r+2}$, we have that $p^{r+2}M_{r+2}$ only has its n^{th} row non-zero and its n^{th} row is a W-linear combination of $\{\sigma^{\mu_I-1}(R_{n+1})\}_{I\in\mathbb{I}^{\min}_{r+2}}$. By Lemma 5.11(5), let $J,J'\in\mathbb{I}^{\min}_{r+2}$ denote the unique elements such that $\mu_J=\max_{I\in\mathbb{I}^{\min}_{r+2}}\mu_I$ and $\mu_{J'}=\min_{I\in\mathbb{I}^{\min}_{r+2}}\mu_I$. Thus the n^{th} row of $p^{r+2}M_{r+2} \bmod p$ is a k-linear combination of $\{\sigma^{\mu-1}(\overline{R}_{n+1})\}_{\mu_{J'}\leq\mu\leq\mu_J}$ and by Lemma 5.11(4), this set consists at most $n+1$ elements. Moreover since J,J' are unique, the coefficients of $\sigma^{\mu_{J'}-1}(\overline{R}_{n+1})$ and $\sigma^{\mu_J-1}(\overline{R}_{n+1})$ are $\sum_{I\in\mathbb{I}^{\min}_{r+2},\mu_I=\mu_{J'}}\overline{c}_I=\overline{c}_{J'}$ and $\sum_{I\in\mathbb{I}^{\min}_{r+2},\mu_I=\mu_J}\overline{c}_I=\overline{c}_J$ respectively; and both are non-zero in k as $c_J,c_{J'}\in W^\times$. Then by Lemma 5.13, $\dim_{\mathbb{F}_p}S_{r+1}\leq n$, where $S_{r+1}:=\{\bar v\in\mathbb{F}_p^{2n}\mid\bar v\in\ker(p^{r+1}M_{r+1} \bmod p)\}$. By Lemma 5.11(1)(3), $\mathbb{I}^{\min}_{r+2}=\mathbb{I}'\times\mathbb{I}^{\min}_{r+1}$, where

$$\mathbb{I}'=\{i\mid 1\leq i\leq n+1,\exists J(i)\in\mathbb{I}^{\min}_{r+1}\text{ such that }(i,J(i))\in\mathbb{I}^{\min}_{r+2}\}.$$

Then $p^{r+2}M_{r+2}=\sum_{i\in\mathbb{I}'}A_i(p^{r+1}M_{r+1})^{(i)}$, where $A_i\in M_{2n}(W)$. Therefore, $S_{r+1}\subset S_{r+2}$. Thus $S_\infty:=\bigcup_{j=1}^\infty S_j$ is a subspace of $\mathbb{F}_p^{2n}$ of dimension at most n. Thus there exists a saturated $\mathbb{Z}_p$-submodule $\Lambda\subset\mathcal{L}$ or rank at least n such that $\Lambda \bmod p\cap S_\infty=\{0\}$ and any primitive vector $v\in\Lambda$ satisfies the desired condition.

For (2), note that $\nu_{r+2} \leq \nu_{(1,\ldots,1)} = h(1+\cdots+p^{r+1})$ since $h = a_1$; thus $h_r + 1 > h(p^r + \cdots + p^{-1}) \geq \nu_{r+2}/p$. $\square$

5.14. Ogus [Ogu01, Lem. 2, Prop. 11] gives an explicit description of the local equation of Newton strata of $\mathcal{S}_k$ and by using the explicit coordinates in §5.5, in the formal neighborhood of a supersingular point P of type $2n$, the Newton stratum of codimension s is cut out by the single equation $y_s = 0$ in the Newton stratum of codimension $s-1$ for $s \leq n$.[18] Let $C \to \mathcal{S}_k$ be a formal curve which specializes to P. Assume that the generic point of C lies in the open Newton stratum of codimension $s - 1 \leq n - 1$. We say a special endomorphism w of $A[p^\infty]$ *decays rapidly* if it satisfies the condition in Definition 5.1 with $h = v_t(y_s(t)) = a_s$. The only place in the proof of Theorem 5.2 where we used the generic ordinary assumption of C is Claim (2). Once we replace the computation there by $\nu_{r+2} \leq \nu_{(s,\ldots,s)} = a_s(1 + \cdots + p^{r+1})$, we obtain the following general version of the Decay Lemma for non-superspecial supersingular points.

Theorem 5.15 (Generalized decay lemma in the generic case)**.** *Suppose that $C \to \mathcal{S}_k$ is a formal curve which specializes to a non-superspecial supersingular point of type $2n$ (i.e., Artin invariant n) and is generically in an open Newton stratum of codimension $\leq n-1$. Then there exists a saturated rank n submodule of special endomorphisms which decays rapidly.*

6. Decay for superspecial points

The goal of this section is to prove a Decay Lemma for superspecial points (Theorem 6.2). The computations in the proof of Theorem 6.2 go along very similar lines to the calculations carried out in [MST, §5]. We will therefore be brief and will refer to [MST] whenever appropriate.

Throughout this section, we work in the setting of a formal curve $C = \operatorname{Spf} k[[t]] \to \mathcal{S}_k$ which is generically ordinary, and specializes to a superspecial point P. Recall that $\mathcal{A}/k[[t]]$ denotes the pullback of $\mathcal{A}^{\mathrm{univ}}$, h denotes the t-adic valuation of the local equation of the non-ordinary locus given in Corollary 4.10, and $\mathcal{L}$ is the lattice of special endomorphisms of the p-divisible group at P.

In order to obtain sufficiently strong bounds to prove Theorem 1.2, we require a Decay Lemma which is slightly stronger than Theorem 5.2. In order to do this, we introduce the notion of very rapid decay; also the following definition for rapid decay in the superspecial case is stronger than Definition 5.1.

Definition 6.1. For a superspecial point P,

1. We say that $w \in \mathcal{L}$ *decays rapidly* (resp. *very* rapidly) if for every $r \in \mathbb{Z}_{\geq 0}$, the special endomorphism $p^r w$ does not lift to an endomorphism of $\mathcal{A}[p^\infty]$ modulo $t^{h'_r+1}$ (resp.$t^{h'_{r-1}+ap^r+1}$),for some $a \leq \frac{h}{2}$ independent of r; here $h'_r := [h(p^r + \cdots + 1) + a/p)]$ and $h'_{-1} := [a/p]$.
2. We say that $w \in \mathcal{L}$ *decays rapidly (resp. very rapidly) to first order* if w does not extend to an endomorphism modulo $t^{[h+a/p]+1}$ (resp. $t^{[a+a/p]+1}$) for some $a \leq \frac{h}{2}$.
3. A $\mathbb{Z}_p$-submodule of $\mathcal{L}$ *decays rapidly* if every primitive vector in this submodule decays rapidly. Given a submodule $\Lambda \subset \mathcal{L}$ which decays rapidly and a vector $w \in \mathcal{L}$ such that $w \notin \Lambda$, we say that *the pair* (L, w) *decays very rapidly to first order* if w decays very rapidly to first order, and every primitive vector in $\operatorname{Span}_{\mathbb{Z}_p}\{L, w\}$ decays rapidly to first order.

The main theorem of this section is the following:

[18] Here and also in the statement of Theorem 5.15, by the Newton stratum associated to ν in Kottwitz's set, we mean the closed subscheme in the Shimura variety parametrizing points whose Newton points/polygons $\nu' \leq \nu$ with respect to the partial order in Kottwitz's set; equivalently, this closed subscheme is the Zariski closure of the locally closed subscheme parametrizing points whose Newton points/polygons are exactly ν. We refer to this locally closed subscheme as the open Newton stratum.

Theorem 6.2 (Decay Lemma in the superspecial case)**.** *There exists a saturated rank 2 $\mathbb{Z}_p$-submodule $\Lambda \subset \mathcal{L}$ which decays rapidly. Moreover, the submodule Λ may be chosen so that at least one of the following statements holds:*

(1) there exists a primitive $w \in \Lambda$ which decays very rapidly;

(2) there exists a primitive vector $w \in \mathcal{L}$ such that $w \notin \Lambda$ and the pair (L, w) decays very rapidly to first order.

We expect that an analogous statement of [MST, Thm. 5.1.2] holds; more precisely, we expect that there is a rank 3 submodule of $\mathcal{L}$ which decays rapidly, and moreover, there exists a vector in this rank 3 submodule which decays very rapidly. In order to prove Theorem 1.2, the weaker statement Theorem 6.2 suffices.

Theorem 3.3 follows directly from Theorem 6.2:

Proof of Theorem 3.3. The argument used to deduce Theorem 3.2 from Theorem 5.2 works in this setting, with Theorem 3.3(1) following from Theorem 6.2(1), and Theorem 3.3(2) following from Theorem 6.2(2). □

The setup. Here we carry out all the computation for the split case described in §4.7 and we will explain in the proof how to deduce the general case from the split case.

6.3. Recall from §4.8 that $\widehat{\mathcal{O}}_{\mathcal{S},P} = \operatorname{Spf} W[[x_1, \ldots, x_m, y_1, \ldots, y_m]]$ (note that for P superspecial, we have $n = t_P/2 = 1$; and x_j, y_j here were denoted by x'_j, y'_j in §4.8); the formal curve C gives rise to the tautological map of local rings

$$W[[x_1, \ldots, x_m, y_1, \ldots, y_m]] \to k[[t]]$$

and we let $x_i(t)$ (respectively $y_i(t)$) denote the images of the x_i (respectively y_i)) in $k[[t]]$. For each of the $x_i(t)$ (respectively $y_i(t)$), let $X_i(t) \in W[[t]]$ (respectively $Y_i(t) \in W[[t]]$) denote the power series whose coefficients are the Teichmuller lifts of those of $x_i(t)$ (respectively $y_i(t)$). Let $Q(t) = -\sum_{i=1}^{m} X_i(t)Y_i(t)$ (compare to §4.8, here we use Q to denote the lift of itself), and let $R(t) = -\sum_{i=1}^{m} \Big(X_i(t)(Y_i(t))^p + (X_i(t))^p Y_i(t)\Big)$. By Corollary 4.10, $Q(t) \bmod p = 0$ is the local equation for the non-ordinary locus, so $h = v_t(Q(t))$. Let $h' = v_t(R(t))$. Without loss of generality, we may assume that $v_t(X_1(t)) \leq v_t(X_i(t))$ and $v_t(X_1(t)) \leq v_t(Y_i(t))$ as everything is symmetric in the x_i, y_i, and let a denote $v_t(X_1(t))$. By definition, we have that $2a \leq h$ and $(p+1)a \leq h'$.

6.4. Following the notation of Lemma 4.3 for $n = 1$, the vector v_1 in Lemma 4.5 must be $\frac{1}{2p}(e_1 + f_1/\lambda)$ and $w_1 = \frac{1}{2}(e_1 - f_1/\lambda)$.[19] By §§4.8,4.11, we have that with respect to the basis $\{e_1, f_1, e'_i, f'_i\}_{i=1}^m$, the Frobenius on $\mathbf{L}_{\mathrm{cris}}(W[[x_i, y_i]])$ is given by

$$\operatorname{Frob} = (I + F) \circ \sigma, \text{ where } F = \left[\begin{array}{cc|cccccc} \frac{Q}{2p} & \frac{-\lambda Q}{2p} & \frac{x_1}{2p} & \cdots & \frac{x_m}{2p} & \frac{y_1}{2p} & \cdots & \frac{y_m}{2p} \\ \frac{Q}{2p\lambda} & \frac{-Q}{2p} & \frac{x_1}{2p\lambda} & \cdots & \frac{x_m}{2p\lambda} & \frac{y_1}{2p\lambda} & \cdots & \frac{y_m}{2p\lambda} \\ \hline -y_1 & \lambda y_1 & & & & & & \\ \vdots & \vdots & & & & & & \\ -y_m & \lambda y_m & & & & & & \\ -x_1 & \lambda x_1 & & & & & & \\ \vdots & \vdots & & & & & & \\ -x_m & \lambda x_m & & & & & & \end{array}\right].$$

[19] There is another possible choice with λ replaced by $-\lambda$; given the computation is the same for both cases, we will just work with the first case.

Let F_t, F_r and F_l denote the top-left 2×2 block, the top-right $2 \times 2m$ block and the bottom-left $2m \times 2$ block of F respectively. Let $F_{r,i}, F_{l,i}$ denote the i^{th} column of F_r and i^{th} row of F_l respectively.

As in §5.6, in order to prove the Decay Lemma, we study the expansion of $F_\infty = \prod_{i=0}^{\infty}(I + F^{(i)})$. Let $F_\infty(1)$, $F_\infty(2)$ and $F_\infty(1,3)$ denote the top-left 2×2 block, the top-right $2 \times 2m$ block, and the $(m+1)^{\text{th}}$ row of the bottom-left $2m \times 2$ block of F_∞ respectively. We denote by $F_\infty(2)_i$ the i^{th} column of $F_\infty(2)$ and $F_\infty(2,3)_i$ the i^{th} entry of the $(m+1)^{\text{th}}$ row of the bottom-right $2m \times 2m$ block of F_∞.

Let

$$M = \frac{1}{2}\begin{bmatrix} 1 & -\lambda \\ \frac{1}{\lambda} & -1 \end{bmatrix}, N = \frac{1}{2}\begin{bmatrix} 1 & \lambda \\ \frac{1}{\lambda} & 1 \end{bmatrix}.$$

Preliminary lemmas. As in [MST, §5.2], we expand the blocks $F_\infty(1,3), F_\infty(2,3)$ in F_∞ as an infinite sum of finite products of σ-twists of F_t, F_r and F_l. The following lemma follows directly from the shape of F and an elementary analysis on t-adic valuations.

Lemma 6.5. *Fix $r \in \mathbb{Z}_{\geq 0}$*

(1) Among all finite products of σ-twists of F_t, F_r and F_l in the expansion of $F_\infty(1,3)$ with p-adic valuation $-(r+1)$, the terms which have the smallest t-adic valuation are contained in the set

$$\mathbb{S}_{1,r+1} = \{F_{l,m+1}\prod_{i=1}^{\alpha} F_t^{(i)} \prod_{j=1}^{\beta} F_r^{(\alpha+2j-1)} F_l^{(\alpha+2j)} \mid \alpha, \beta \in \mathbb{Z}_{\geq 0}, \alpha + \beta = r+1\}.$$

(2) Among all finite products of σ-twists of F_t, F_r and F_l in the expansion of $F_\infty(2,3)_s$ with p-adic valuation $-(r+1)$, the terms with the smallest t-adic valuation are contained in the set

$$\mathbb{S}_{2,s,r+1} = \{F_{l,m+1}\prod_{i=1}^{\alpha} F_t^{(i)} \prod_{j=1}^{\beta} F_r^{(\alpha+2j-1)} F_l^{(\alpha+2j)} F_{r,s}^{(\alpha+2\beta+1)} \mid \alpha, \beta \in \mathbb{Z}_{\geq 0}, \alpha + \beta = r\}.$$

The following lemma follows from a direct computation similar to [MST, Lem. 5.2.1, Lem. 5.2.3].

Lemma 6.6. *Consider the product* $P_{\alpha,\beta} = \prod_{i=1}^{\alpha} F_t^{(i)} \cdot \prod_{j=1}^{\beta} F_r^{(\alpha+2j-1)} F_l^{(\alpha+2j)}$.

(1) If α is odd, the product equals $p^{-(\alpha+\beta)} \prod_{i=1}^{\alpha} Q^{(i)} \cdot \prod_{j=1}^{\beta} R^{(\alpha+2j-1)} M^{(1)}$.

(2) If α is even, the product equals $p^{-(\alpha+\beta)} \prod_{i=1}^{\alpha} Q^{(i)} \cdot \prod_{j=1}^{\beta} R^{(\alpha+2j-1)} N^{(1)}$.

In either case, the kernel of $F_{l,m+1} p^{\alpha+\beta} P_{\alpha+\beta} \bmod p$ does not contain any non-zero $\mathbb{F}_p$-rational vectors or any $k^\times$-multiple of $([1, \lambda^{-1}]^{(\alpha+1)})^T$.

Decay in the superspecial case.

Proof of Theorem 6.2. We first prove the theorem in the split case. We continue the argument in §5.6.

Case 1: $h < h'$. It follows from Lemmas 6.5 and 6.6 that there is a unique element of $\mathbb{S}_{1,r+1}$ (respectively $\mathbb{S}_{2,1,r+1}$) with minimal t-adic valuation. This term is $F_{l,m+1}P_{r+1,0}$ (respectively $F_{l,m+1}P_{k,0}F_{r,1}^{(r+1)}$), and has t-adic valuation $a + h(p + \cdots + p^{r+1})$ (respectively $a + h(p + \cdots + p^r) + ap^{r+1}$). By the last assertion of Lemma 6.6, we conclude that for any primitive vector $w \in \operatorname{Span}_{\mathbb{Z}_p}\{e_1, f_1\}$, in the entry of the vector $p^r F_\infty w$ corresponding to its f_1'-coordinate in $K[[t]]$, the coefficient of $t^{a+h(p+\cdots+p^{r+1})}$ does not lie in W. Since $\gamma_0 e_1 + \delta_0 f_1 + \sum_{i=1}^m \gamma_i e_i' + f_1' + \sum_{i=2}^m \delta_i f_i'$ is primitive in $\mathbb{L} = \mathbf{L}_{\text{cris},P}(W)$ for any $\gamma_i, \delta_i \in W$, we conclude that the horizontal section $p^r F_\infty w$ is not integral in $\iota_s^* \mathbf{L}_{\text{cris}}(D_s)$ if

$s > (a + h(p + \cdots + p^{r+1}))/p$ as for any $N < sp$, we have $t^N/p \notin D_s$. Thus $\mathrm{Span}_W\{e_1, f_1\}$ decays rapidly. Similarly by Lemma 6.6, the special endomorphism e_1' decays very rapidly. The fact that $\mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1, e_1'\}$ decays rapidly follows from an argument identical to that outlined in the last paragraph of [MST, Proof of Prop. 5.1.3 Case 1 in §5.2]. For the convenience of the reader, we give a brief sketch of this argument. Let w be a primitive vector in $\mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1\}$. As in *loc. cit.*, it suffices to prove that the t-adic valuation of the term of $F_\infty w$ with denominator p^{r_1} is different from the t-adic valuation of the term of $F_\infty e_1'$ with denominator p^{r_2} for any $r_1, r_2 \in \mathbb{Z}_{>0}$. The former quantity equals $a + h(p + \cdots + p^{r_1})$, and the latter quantity equals $a + h(p + \ldots p^{r_2-1}) + ap^{r_2}$. As $1 \leq a \leq h/2$, it follows that these quantities can never be the same and the result follows.

Note that in this case, we have proved that a rank 3 submodule of $\mathcal{L}$ must decay.

Case 2: $h'(1+p^{2e-1}) < h(1+p) < h'(1+p^{2e+1})$**, for some** $e \in \mathbb{Z}_{\geq 1}$**.** As in Case 1, by Lemma 6.5, $F_{l,m+1}P_{r-e+1,e}$ is the unique element of $\mathbb{S}_{1,r+1}$ with minimal t-adic valuation (the argument is similar to that of [MST, Lem. 5.2.6]); moreover $v_t(F_{l,m+1}P_{r-e+1,e}) < v_t(F_{l,m+1}P_{r+1,0}) < p(h_r' + 1)$. Thus, by Lemma 6.6 and the same argument as in Case 1, $\mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1\}$ decays rapidly. On the other hand, $|S_{2,1,1}| = 1$ and the unique element has t-adic valuation $a + pa$ and thus e_1' decays very rapidly to first order. The fact that the pair $(\mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1\}, e_1')$ decays very rapidly to first order follows from an argument identical to the one outlined at the end of Case 1.

Case 3: $h'(1+p^{2e-1}) = h(1+p)$ **for some** $e \in \mathbb{Z}_{\geq 1}$**.** In this case, by Lemma 6.5 and a computation similar to [MST, Lem. 5.2.7], we have that $F_{l,m+1}P_{r-e+1,e-1} \cdot F_{r,1}^{(r+e)}$ is a unique element of $\mathbb{S}_{2,1,r+1}$ with the smallest t-adic valuation and

$$v_t(F_{l,m+1}P_{r-e+1,e-1} \cdot F_{r,1}^{(r-e)}) \leq v_t(F_{l,m+1}P_{r,0} \cdot F_{r,1}^{(r+1)}) < p(h_{r-1}' + ap^r + 1).$$

Then the last assertion of Lemma 6.6 implies that e_1' decays very rapidly.

Claim. At least one of e_1, f_1 decays rapidly.

Proof. When $r < e - 1$, there is a unique element of $\mathbb{S}_{1,r+1}$ with minimal t-adic valuation, and thus the argument as in Case 1 shows that for any primitive vector $w \in \mathrm{Span}_{\mathbb{Z}_p}\{e_1, f_1\}$, we have that $p^r w$ does not lift to an endomorphism $\mathrm{mod} t^{h_r'+1}$.

When $r \geq e - 1$, there are exactly *two* distinct elements of $\mathbb{S}_{1,r+1}$ with minimal t-adic valuation, and they are $P_{r-e+1,e}$ and $P_{r-e+2,e-1}$. We first prove that at least one of $p^{e-1}e_1, p^{e-1}f_1$ does not extend to an endomorphism modulo $t^{h_{e-1}'+1}$. Indeed, by Lemma 6.6, we have that $F_{l,m+1}(P_{0,e} + P_{1,e-1}))$ equals $p^{-e}(AF_{l,m+1}M^{(1)} + BF_{l,m+1}N^{(1)})$, where $A = \prod_{i=1}^{e} R^{(2i-1)}, B = Q^{(1)} \cdot \prod_{i=1}^{(e-1)} R^{(2i)}$. Let $\gamma, \delta \in W^\times)$ denote the leading coefficients of $A, B \in W[[t]]$. As $p > 2$, we have that at most one of $\gamma - \delta$ and $\gamma + \delta$ lie in pW. Suppose that $\gamma + \delta \notin pW$, then $[1, 0]^T$ does not lie in $\ker(\gamma[-1, \lambda]N^{(1)} + \delta[-1, \lambda]M^{(1)})$ and thus the e_1'-coordinate of the horizontal section $\tilde{e}_1$, up to a scalar multiple in $W^\times$, is $p^{-e}t^{a+h'(p+p^3+\cdots+p^{2e-1})}+$ other powers of t. Therefore, $p^{e-1}e_1$ does not lift to $\mathrm{mod} t^{h_{e-1}'+1}$ because $h_{e-1}' + 1 > h(p^{e-1} + p^{e-2} + \cdots 1) + a/p \geq (a + h'(p + p^3 + \cdots + p^{2e-1}))/p$. On the other hand, if $\gamma - \delta \notin pW$, then $[0, 1]^T$ does not lie in $\ker \gamma[-1, \lambda]M^{(1)} + \delta[-1, \lambda]N^{(1)}$ and the same argument as above implies that $p^{e-1}f_1$ does not lift to $\mathrm{mod} t^{h_{e-1}'+1}$.

Now we show that if $\gamma + \delta \notin pW$, then e_1 decays rapidly. A similar computation as above shows that for $r \geq e$, we have that $F_{l,m+1}(P_{r-e+1,e} + P_{r-e+2,e-1}) = p^{-(r+1)}X_1 \prod_{i=1}^{r-e+1} Q^{(i)}(A^{(r-e+1)}M^{(1)} + B^{(r-e+1)}N^{(1)})$ when $r-e$ is even and $F_{l,m+1}(P_{r-e+1,e} + P_{r-e+2,e-1}) = p^{-(r+1)}X_1 \prod_{i=1}^{r-e+1} Q^{(i)}(A^{(r-e+1)}N^{(1)} + B^{(r-e+1)}M^{(1)})$ when $r - e$ is odd. Since $\gamma + \delta \notin pW$, then $\gamma^{(r-e+1)} + \delta^{(r-e+1)} \notin pW$. Thus $[1, 0]^T$ does not lie in $\ker(\gamma^{(r-e+1)}[-1, \lambda]N^{(1)} + \delta^{(r-e+1)}[-1, \lambda]M^{(1)})$ and $\ker(\gamma^{(r-e+1)}[-1, \lambda]M^{(1)} + \delta^{(r-e+1)}[-1, \lambda]N^{(1)})$. Therefore, we conclude that e_1 decays rapidly by a direct computation of the t-adic valuation of $F_{l,m+1}P_{r-e+1,e}$ as the $r = e - 1$ case.

If $\gamma - \delta \notin pW$, then an identical argument as above implies that f_1 decays rapidly. □

To finish the proof of Case 3, we notice that an argument identical to the one outlined at the end of Case 1 goes through to show that if e_1 (resp. f_1) decays rapidly, then $\mathrm{Span}_{\mathbb{Z}_p}\{e_1, e_1'\}$ (resp. $\mathrm{Span}_{\mathbb{Z}_p}\{f_1, e_1'\}$) decays rapidly.

□

6.7. It follows from [HP17, Prop. 4.2.5] that the lattice $\mathcal{L}_1$ is split if and only if the global lattice defining the Shimura variety (with signature $(b,2)$, and which is self-dual at p by assumption) is *not* split mod p. Therefore, we have established the required decay whenever the Shimura variety is defined by an even-dimensional orthogonal $\mathbb{Z}$-lattice (with signature $(b,2)$, and which is self-dual at p by assumption) which is *not* split mod p. Therefore, suppose that the global lattice is either odd-dimensional, or is even-dimensional but *is split* mod p. We embed this lattice inside a higher-rank lattice, which is even-dimensional (and has signature $(b',2)$, and is self-dual at p), and which is *not* split mod p – after choosing level structure away from p appropriately, this induces an embedding of Shimura varieties, and we abuse notation by letting P (resp. $\mathrm{Spf}\, k[[t]]$) denote the image of our original supersingular point (resp. of the formal curve) in the larger Shimura variety. Let $\mathcal{L}'$ denote the $\mathbb{Z}_p$-module of special endomorphisms of the p-divisible group of P when we view it as in the bigger variety. That $\mathrm{Spf}\, k[[t]]$ actually lies in our original orthogonal Shimura variety implies that $\mathcal{L}' = \mathcal{L} \oplus \mathcal{L}''$, with every special endomorphism in $\mathcal{L}''$ extending to $\mathrm{Spf}\, k[[t]]$ (in fact, $\mathcal{L}''$ necessarily arises from actual endomorphisms of the abelian variety at P, not just endomorphisms of the p-divisible group at P). Our computation applies to yield a $\mathbb{Z}_p$-sublattice of $\mathcal{L}'$ that decays rapidly. Adding any special endomorphism contained in $\mathcal{L}''$ to a special endomorphism that decays rapidly doesn't affect the rapidity of decay, because of the above remark. Therefore, we may assume that the lattice of special endomorphisms that decays rapidly is contained in $\mathcal{L}$, thereby establishing the necessary decay even when $\mathcal{L}_1$ is odd-dimensional, or non-split.

7. Proof of Theorem 1.2

In this section, we prove Theorem 1.2. As sketched in the introduction, our approach is to combine global bounds from Borcherds theory with bounds on the average local intersection multiplicities. At supersingular points, these are obtained using section 3 (Theorem 3.2 and Theorem 3.3).

Note that Theorem 1.2 is independent of the choice of level structure of $\mathcal{S}$ and is equivalent for different $\mathcal{S}$ with the same quadratic space $(L \otimes \mathbb{Q}, Q)$ over $\mathbb{Q}$; thus without loss of generality, we may assume that $L \subset V = L \otimes \mathbb{Q}$ is maximal among all lattices over which Q is $\mathbb{Z}$-valued. Recall from Theorem 1.2 that we assume that C is not contained in any special divisors $Z(m)$.

The global intersection number and its decomposition.

7.1. Let $S \subset \mathbb{Z}_{>0}$ be a set of positive density (i.e., $\lim_{X\to\infty} \frac{1}{X}|\{m \in S \mid m \leq X\}|$ exists and > 0) and we also assume that each $m \in S$ is representable by (L, Q) and for any $m \in S$, we have $p \nmid m$. By the theory of quadratic forms, such S exists.[20] For $X \in \mathbb{Z}_{>0}$, we use S_X to denote $\{m \in S \mid X \leq m \leq 2X\}$.

7.2. We use vector-valued modular forms to control the asymptotic of $C.Z(m)$ as $m \to \infty$. Let $L^\vee$ denote the dual of L in V with respect to the bilinear form $[-,-]$ induced by Q and let $\{\mathfrak{e}_\mu\}_{\mu \in L^\vee/L}$ denote the standard basis of $\mathbb{C}[L^\vee/L]$. Let ρ_L denote the Weil representation on $\mathbb{C}[L^\vee/L]$ of the metaplectic group $\mathrm{Mp}_2(\mathbb{Z})$. As in [MST, §4.1.4], we consider the Eisenstein series $E_0(\tau), \tau \in \mathbb{H}$ defined by $E_0(\tau) = \displaystyle\sum_{(g,\sigma)\in \Gamma'_\infty \backslash \mathrm{Mp}_2(\mathbb{Z})} \sigma(\tau)^{-(2+b)}(\rho_L(g,\sigma)^{-1}\mathfrak{e}_0)$, where $\Gamma'_\infty \subset \mathrm{Mp}_2(\mathbb{Z})$ is the stabilizer

[20] Indeed, by [SSTT, Lem. 4.7], every $m \gg 1$ is representable since L is maximal.

of ∞. Note that the constant term of E_0 is $\mathfrak{e}_0$ and $E_0(\tau)$ is a weight $1+\frac{b}{2}$ modular form with respect to ρ_L.

The following theorem of Bruinier–Kuss [BK01] gives explicit formulae of the Fourier coefficients of E_0. As we are using different convention of the signature of (V, Q) as in [BK01], we refer the reader to the formulae in [Bru17, Thms. 2.3, 2.4].

7.3. We first introduce some notation for an arbitrary quadratic lattice (L, Q) over $\mathbb{Z}$. We write $\det(L)$ for the determinant of its Gram matrix. For a rational prime ℓ, we use $\delta(\ell, L, m)$ to denote the local density of L representing m over $\mathbb{Z}_\ell$. More precisely, $\delta(\ell, L, m) = \lim_{a\to\infty} \ell^{a(1-\operatorname{rk} L)}\#\{v \in L/\ell^a L \mid Q(v) \equiv m \bmod \ell^a\}$. If m is representable by $(L\otimes\mathbb{Z}_\ell, Q)$, then $\delta(\ell, L, m) > 0$; moreover, when $\operatorname{rk} L \geq 5$ (this is the case for our application), by for instance [Iwa97, pp. 198-199], for a fixed ℓ, we have that $\delta(\ell, L, m) \asymp 1$ for all m representable by $(L \otimes \mathbb{Z}_\ell, Q)$.

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

Theorem 7.4 ([BK01, Thm. 11])**.** *Recall that (L, Q) is a quadratic lattice over $\mathbb{Z}$ of signature $(b, 2)$ with $b \geq 3$. Let $q_L(m)$ denote the coefficient of $q^m\mathfrak{e}_0$ in the q-expansion of E_0.*

(1) For b even, the Fourier coefficient $q_L(m)$ is
$$-\frac{2^{1+b/2}\pi^{1+b/2}m^{b/2}\sigma_{-b/2}(m, \chi_{(-1)^{1+b/2}4\det L})}{\sqrt{|L^\vee/L|}\Gamma(1+b/2)L(1+b/2, \chi_{(-1)^{1+b/2}4\det L})}\prod_{\ell|2\det(L)}\delta(\ell, L, m).$$

(2) For b odd, 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
$$-\frac{2^{1+b/2}\pi^{1+b/2}m^{b/2}L((b+1)/2, \chi_{\mathcal{D}})}{\Gamma(1+b/2)\sqrt{|L^\vee/L|}\zeta(b+1)}\left(\sum_{d|f}\mu(d)\chi_{\mathcal{D}}(d)d^{-(b+1)/2}\sigma_{-b}(f/d)\right)\prod_{\ell|2\det L}\Big(\delta(\ell, L, m)/(1-\ell^{-1-b})\Big),$$
where μ is the Mobius function and $\mathcal{D} = (-1)^{(b-1)/2}2m_0\det L$.

In particular, $|q_L(m)| \asymp m^{b/2}$ for all m representable by (L, Q).

Here the last assertion is a direct consequence of the above explicit formulae and the fact $\delta(\ell, L, m) \asymp 1$ (see also [MST, §4.3.1]).

Recall that $C \to \mathcal{S}_k$ is a smooth proper curve such that the generic point of C maps into the ordinary locus of $\mathcal{S}_k$.

Lemma 7.5. *Let ω denote the tautological line bundle on $\mathcal{S}_k$ corresponding to $\operatorname{Fil}^1 V \subset V$ (i.e., ω is the line bundle of modular forms on $\mathcal{S}_k$ of weight 1). Then the intersection number $Z(m).C = |q_L(m)|(\omega.C) + O(m^{(b+2)/4})$. In particular, $\sum_{m\in S_X} Z(m).C \asymp (\omega.C)\sum_{m\in S_X}|q_L(m)| \asymp X^{1+b/2}$ for S_X defined in §7.1.*

Proof. By the modularity theorem of Borcherds [Bor99] or its arithmetic version by Howard and Madapusi Pera [HP20], we have that $-(\omega.C)+\sum_{m=1}^\infty Z(m).C$ is the $\mathfrak{e}_0$-component of a vector-valued modular form with respect to ρ_L of weight $(1+b/2)$ and its Eisenstein part is given by the $\mathfrak{e}_0$-component of $-(\omega.C)E_0$ (see [MST, Thm. 4.1.1, §4.1.4]). The difference of $-(\omega.C)+\sum_{m=1}^\infty Z(m).C$ and the $\mathfrak{e}_0$-component of $-(\omega.C)E_0$ is a cusp form, and thus the first assertion follows from the trivial bound on Fourier coefficients of cusp forms (see [Sar90, Prop. 1.3.5]). We then obtain the last assertion by Theorem 7.4. □

In order to compare $C.Z(m)$ with the local intersection number $i_P(C.Z(m))$ for a point $P \in (C\cap Z(m))(k)$, we introduce the notion of *global intersection number* $g_P(m)$ as follows.

Definition 7.6 ([MST, Def. 7.1.3])**.** Let H denote the Hasse-invariant on $\mathcal{S}_k$ (i.e., $H = 0$ cuts out the non-ordinary locus). Let t be the local coordinate at P (i.e., $\widehat{C}_P = \operatorname{Spf} k[[t]]$) and let $h_P = v_t(H)$, the t-adic valuation of H restricted to $\widehat{C}_P$. We define

$$g_P(m) = \frac{h_P}{p-1}|q_L(m)|.$$

In particular, $g_P(m) = 0$ for P ordinary and

$$\sum_{P \in (C \cap Z(m))(k)} g_P(m) = |q_L(m)|(\omega.C)$$

since H is a section of ω^{p-1}.

Local intersection number: preparation and non-supersingular points. We first introduce some notation and reformulate the calculation of local intersection number as a lattice counting problem.

7.7. Recall that $P \in (C \cap Z(m))(k)$ for some m. Let $\mathcal{A}/k[[t]]$ denote the pullback of the universal abelian scheme $\mathcal{A}^{\mathrm{univ}}$ via $\operatorname{Spf} k[[t]] = \widehat{C}_P \to \mathcal{S}_k$. Let L_n denote the $\mathbb{Z}$-lattice of special endomorphisms of $\mathcal{A} \bmod t^n$. By definition, $L_{n+1} \subset L_n$ for every $n \geq 1$, and our assumption that C is not contained in any special divisor yields that $\cap_n L_n = \{0\}$. By [MST, Rmk. 7.2.2], all L_n have the same rank. Moreover, by [HP17, Lem. 4.2.4], P is supersingular if and only if $\operatorname{rk}_{\mathbb{Z}} L_1 = b+2$. Since the quadratic form $\langle -, - \rangle$ on $\mathbf{L}_{\mathrm{cris},P}(W)$ satisfies that $\langle \varphi(x), \varphi(y) \rangle = \sigma(\langle x, y \rangle)$, then the slope s part of $\mathbf{L}_{\mathrm{cris},P}(W)$ has the same dimension as the slope $-s$ part and hence the slope non-zero part cannot have rank $b+1$; thus if P is not supersingular, then $\operatorname{rk}_{\mathbb{Z}} L_1 \leq b$.

On the other hand, by Remark 2.4, we have a positive definite quadratic form Q on L_n given by $v \circ v = [Q(v)]$ for $v \in L_n$. By the moduli-theoretic description of the special divisors and the fact that C intersects $Z(m)$ properly (due to the assumption that the image of C does not lie in any $Z(m)$), we have

$$i_P(C.Z(m)) = \sum_{n=1}^{\infty} \#\{v \in L_n \mid Q(v) = m\}. \tag{7.1}$$

Note that although for a fixed m, the set $\{v \in L_n \mid Q(v) = m\}$ is empty for $n \gg 1$, but this bound on n is in general dependent on m. In the work of Chai and Oort [CO06], they use the canonical product structure in the setting $\mathcal{S} = \mathcal{A}_1 \times \mathcal{A}_1$ and work with a sequence of divisors for which the local contributions from any one fixed point is *absolutely bounded*, independent of the special divisor.

By Lemma 7.5 and the last assertion of Theorem 7.4, there exists an absolute constant c_1 (which depends only on the curve C) such that

$$i_P(C.Z(m)) \leq (C.Z(m)) < c_1 m^{b/2}. \tag{7.2}$$

We now recall the definitions of the successive minima of the L_n from [EK95].

Definition 7.8.

(1) For $i \in \{1, \ldots, r = \operatorname{rk}_{\mathbb{Z}} L_n\}$, the successive minima $\mu_i(n)$ of L_n is defined as $\inf\{y \in \mathbb{R}_{>0} \mid \exists v_1, \ldots, v_i \in L_n$ linearly independent, and $Q(v_j) \leq y^2, 1 \leq j \leq i\}$.

(2) For $n \in \mathbb{Z}_{\geq 1}, 1 \leq i \leq r$, define $a_i(n) = \prod_{j=1}^{i} \mu_j(n)$; define $a_0(n) = 1$.

The determinant of a quadratic lattice (which is approximately the product of all the successive minima) gives first order control on the number of lattice points with bounded norm – however, the error term does depend on the lattice in question. In our setting, we must count lattice points of bounded norm in an infinite family of lattices, and so considering the determinants alone doesn't allow us sufficient control across this family of lattices. Indeed, in the example of a formal curve constructed in §3.5, the error terms involved can get very large, even on average. As seen in [EK95], the data of each individual successive-minima controls the error term in a way that is uniform across all lattices of a fixed rank, and hence we keep track of this more refined data in our setting of a nested family of lattices.

We have the following result establishing lower bounds for the $a_i(n)$, which is similar to [SSTT, Lem. 7.6].

Lemma 7.9. *We have that* $a_i(n) \gg n^{i/b}$ *for* $1 \leq i \leq \mathrm{rk}_{\mathbb{Z}} L_n$*.*

Proof. Let $0 \neq v \in L_n$ be a vector that minimizes the quantity $Q(v)$, and thus $a_1(n) = Q(v)^{1/2}$. Note that $v \in L_n$ implies $v \in L_i$ for every $i \leq n$. Take $m = Q(v)$. Eqn. (7.1) yields that $i_P(C \cdot Z(m)) \geq n$, and then by Eqn. (7.2), $n < c_1 m^{b/2}$. As $a_1(n)^2 = m$, it follows that $c_1 a_1(n)^b > n$, whence $a_1(n) \gg n^{1/b}$. The bounds for the other $a_i(n)$ follow from the observation that $\mu_i(n) \geq a_1(n)$, and hence $a_i(n) \geq a_1(n)^i$. □

Corollary 7.10. *For S_X defined in §7.1, there exists a constant c_2 depending only on C such that*

$$\sum_{m \in S_X} i_P(C.Z(m)) = \sum_{n=1}^{c_2 X^{b/2}} \sum_{m \in S_X} \#\{v \in L_n \mid Q(v) = m\}.$$

Proof. Lemma 7.9 implies that there exists a constant c_2 only depending on C such that for $n > c_2 X^{b/2}$, we have $a_1(n) > (2X)^{1/2}$. In other words, $\min_{0 \neq v \in L_n} Q(v) > 2X$. Then the corollary follows from Eqn. (7.1). □

We are now ready to bound the local intersection number $i_P(C.Z(m))$ on average over m for P not supersingular, which is the analogue of [SSTT, Prop. 7.7].

Proposition 7.11. *For P not supersingular, we have that*

$$\sum_{m=1}^{2X} i_P(C.Z(m)) = O(X^{b/2} \log X),$$

where the implicit constant only depends on C. In particular, $\sum_{m \in S_X} i_P(C.Z(m)) = O(X^{b/2} \log X)$.

Proof. By §7.7, we have that $r := \mathrm{rk}_{\mathbb{Z}} L_n \leq b$. By [EK95, Lem. 2.4, Eqns (5)(6)] and Lemma 7.9, we have

$$\sum_{n=1}^{c_2 X^{b/2}} \sum_{m=1}^{2X} \#\{v \in L_n \mid Q(v) = m\} \ll \sum_{n=1}^{c_2 X^{b/2}} \sum_{i=0}^{r} \frac{(2X)^{i/2}}{a_i(n)} \ll \sum_{n=1}^{c_2 X^{b/2}} \sum_{i=1}^{r} \frac{(2X)^{i/2}}{n^{i/b}},$$

where the implicit constant in the first inequality is absolute and the implicit constant in the second inequality only depends on C. For any $1 \leq i < b$, we see that

$$\sum_{n=1}^{c_2 X^{b/2}} \frac{(2X)^{i/2}}{n^{i/b}} = (2X)^{i/2} \sum_{n=1}^{c_2 X^{b/2}} \frac{1}{n^{i/b}} = O(X^{b/2}),$$

as required. If $i = b$, the identical calculation yields a bound of $O(X^{b/2} \log X)$. The result then follows directly by Corollary 7.10. □

Local intersection number at supersingular points.

7.12. For a supersingular point $P \in C$, we break the local intersection number into two parts for a fixed $T \in \mathbb{Z}_{>0}$, to be chosen later, as follows: $\sum_{m\in S_X} i_P(C.Z(m)) = i_P(X,T)_{\mathrm{err}} + i_P(X,T)_{\mathrm{mt}}$, where

$$i_P(X,T)_{\mathrm{err}} = \sum_{n=T}^{c_2X^{b/2}} \sum_{m\in S_X} \#\{v \in L_n \mid Q(v) = m\}, \quad i_P(X,T)_{\mathrm{mt}} = \sum_{n=1}^{T-1} \sum_{m\in S_X} \#\{v \in L_n \mid Q(v) = m\}$$

and the equality holds due to Corollary 7.10.

We first bound the error term $i_P(X,T)_{\mathrm{err}}$.

Proposition 7.13. *There exists an absolute constant $c_3 > 0$ (independent of X, T) such that*

$$i_P(X,T)_{\mathrm{err}} \leq \frac{c_3}{T^{2/b}} X^{\frac{b+2}{2}} + O(X^{(b+1)/2}).$$

Proof. As in the proof of Proposition 7.11, by Lemma 7.9, we have

$$i_P(X,T)_{\mathrm{err}} \leq \sum_{n=T}^{c_2X^{b/2}} \sum_{m=1}^{2X} \#\{v \in L_n \mid Q(v) = m\} \ll \sum_{n=T}^{c_2X^{b/2}} \sum_{i=0}^{b+2} \frac{(2X)^{i/2}}{n^{i/b}}.$$

As in the proof of Proposition 7.11, we have that $\sum_{n=T}^{c_2X^{b/2}} \frac{(2X)^{i/2}}{n^{i/b}} = O(X^{\frac{b+1}{2}})$ for all $1 \leq i \leq b+1$. For $i = b+2$, we have $\sum_{n=T}^{c_2X^{b/2}} \frac{(2X)^{\frac{b+2}{2}}}{n^{(b+2)/b}} < \sum_{n=T}^{\infty} \frac{(2X)^{\frac{b+2}{2}}}{n^{(b+2)/b}} \leq \frac{c_3}{T^{2/(b+2)}} X^{\frac{b+2}{2}}$ for some absolute constant $c_3 > 0$ by a direct computation. □

In order to bound $i_P(X,T)_{\mathrm{mt}}$, we study the theta series attached to (certain lattices containing) L'_n.

7.14. Let $L'_n \subset L_n \otimes \mathbb{Q}$ be a $\mathbb{Z}$-lattice such that $L'_n \supset L_n$, L'_n is maximal at all primes $\ell \neq p$, and $L'_n \otimes \mathbb{Z}_p = L_n \otimes \mathbb{Z}_p$; we may choose $L'_n \subset L'_1$ and we will assume this for the rest of this section; the quadratic form Q restricts to a positive definite quadratic form on L'_n. Let θ_n denote the theta series attached to L'_n and we write its q-expansion as $\theta_n(q) = \sum_{m=0}^{\infty} r_n(m)q^m$. By definition, $r_n(m) \geq \#\{v \in L_n \mid Q(v) = m\}$ and hence $i_P(X,T)_{\mathrm{mt}} \leq \sum_{n=1}^{T} \sum_{m\in S_X} r_n(m)$.

The theta series θ_n is a weight $1 + b/2$ modular form and we decompose $\theta_n(q) = E_{L'_n}(q) + G_n(q)$, where $E_{L'_n}$ is an Eisenstein series and G_n is a cusp form. Let $q_{L'_n}(m)$ and $a(m)$ denote the m-th Fourier coefficients of $E_{L'_n}$ and $\sum_{n=1}^{T} G_n$ respectively. By [Sar90, Prop. 1.3.5], we have $a(m) = O_T(m^{(b+2)/4})$ and thus

$$i_P(X,T)_{\mathrm{mt}} \leq \sum_{n=1}^{T} \sum_{m\in S_X} q_{L'_n}(m) + \sum_{m\in S_X} a(m) = \sum_{n=1}^{T} \sum_{m\in S_X} q_{L'_n}(m) + O_T(X^{1+(b+2)/4}).$$

The following theorem gives explicit formulae of $q_{L'_n}(m)$.

Theorem 7.15 (Siegel mass formula)**.** *Given any $L' \subset L'_1$ sublattice such that $L' \otimes \mathbb{Z}_\ell = L'_1 \otimes \mathbb{Z}_\ell$ for all $\ell \neq p$, let $q_{L'}(m)$ be the m-th Fourier coefficient of the Eisenstein part of the theta series attached to L'.*

(1) *For b even,*

$$q_{L'}(m) = \frac{2^{1+b/2}\pi^{1+b/2}m^{b/2}\sigma_{-b/2}(m, \chi_{(-1)^{1+b/2}4\det L'_1})}{\sqrt{|L'^\vee/L'|}\Gamma(1+b/2)L(1+b/2, \chi_{(-1)^{1+b/2}4\det L'_1})} \prod_{\ell | 2 det L'_1} \delta(\ell, L', m).$$

(2) *For b odd, $q_{L'}(m)$ equals*

$$\frac{2^{1+b/2}\pi^{1+b/2}m^{b/2}L((b+1)/2, \chi_{\mathcal{D}'})}{\Gamma(1+b/2)\sqrt{|L'^\vee/L'|}\zeta(b+1)} \left(\sum_{d|f} \mu(d)\chi_{\mathcal{D}}(d)d^{-(b+1)/2}\sigma_{-b}(f/d)\right) \prod_{\ell | 2\det L'_1} \Big(\delta(\ell, L', m)/(1-\ell^{-1-b})\Big),$$

where we write $m = m_0 f^2$, where $\gcd(f, 2\det L'_1) = 1$ and $v_\ell(m_0) \in \{0,1\}$ for all $\ell \nmid 2\det L'_1$, μ is the Mobius function, and $\mathcal{D}' = (-1)^{(b-1)/2}2m_0\det L'_1$.

Proof. This theorem is a direct consequence of the Siegel mass formula by the same proof in [MST, Thm. 4.2.2]. □

We may apply this theorem to $L' = L'_n$ in §7.14 because all L'_n are maximal at $\ell \neq p$ and thus $L'_n \otimes \mathbb{Z}_\ell = L'_1 \otimes \mathbb{Z}_\ell$.

Lemma 7.16. *For $p \nmid m$, we have that*

$$\frac{q_{L'_n}(m)}{|q_L(m)|} \leq \frac{2}{\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|}(1 - p^{-[(b+2)/2]})}.$$

Moreover, if P is superspecial, then

$$\frac{q_{L'_1}(m)}{|q_L(m)|} \leq \frac{1+p^{-1}}{p(1-p^{-[(b+2)/2]})}.$$

Proof. By [HP17, Rmk. 7.2.5], $L \otimes \mathbb{Q}_\ell \cong L'_n \otimes \mathbb{Q}_\ell$ as quadratic spaces for all $\ell \neq p$; since L, L'_n are both maximal at $\ell \neq p$, then $L \otimes \mathbb{Z}_\ell \cong L'_n \otimes \mathbb{Z}_\ell$ as $\mathbb{Z}_\ell$-quadratic lattices for all $\ell \neq p$ (see for instance [HP17, Thm. A.1.2]). Moreover, since $p \nmid m$, then by Theorems 7.4 and 7.15, we have that

$$\frac{q_{L'_n}(m)}{|q_L(m)|} = \frac{\delta(p, L'_n, m)}{\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|}(1 - \chi_{(-1)^{1+b/2}4\det L}(p)p^{-1-b/2})} \text{ if } 2 \mid b;$$

$$\frac{q_{L'_n}(m)}{|q_L(m)|} = \frac{\delta(p, L', m)(1-\chi_{\mathcal{D}}(p)p^{-(b+1)/2})}{\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|}(1-p^{-1-b})} \text{ if } 2 \nmid b.$$

Therefore,

$$\frac{q_{L'_n}(m)}{|q_L(m)|} \leq \frac{\delta(p, L'_n, m)}{\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|}(1 - p^{-[(b+2)/2]})}.$$

For the first assertion, it remains to show that $\delta(p, L'_n, m) \leq 2$. Write the quadratic form Q on $L'_n \otimes \mathbb{Z}_p$ into the diagonal form $\sum_{i=1}^{b+2} a_i x_i^2$ with $a_i \in \mathbb{Z}_p$ and we may assume that there exists a_i such that $p \nmid a_i$; otherwise $\delta(p, L'_n, m) = 0$ then we are done. Now let $\widetilde{L}'_n$ denote the quadratic lattice over $\mathbb{Z}_p$ with the quadratic form $\widetilde{Q}$ given by $\sum_{1 \leq i \leq b+2, p \nmid a_i} a_i x_i^2$. By [Han04, Rmk. 3.4.1(a), Lem. 3.2], we have that

$$\delta(p, L'_n, m) = p^{-b-1}\#\{v \in L'_n/pL'_n \mid Q(v) \equiv m \bmod p\} = p^{1-\operatorname{rk}\widetilde{L}'_n}\#\{v \in \widetilde{L}'_n/p\widetilde{L}'_n \mid \widetilde{Q}(v) \equiv m \bmod p\},$$

where the last equality follows from definition. If $\operatorname{rk}\widetilde{L}'_n \geq 3$, the $\mathbb{F}_p$-quadratic form $\widetilde{Q} \bmod p$ is isotropic, then we may write $\widetilde{Q} \bmod p = xy + Q'(z)$. For $x \in \mathbb{F}_p^\times$, for any value of z, there is at most one $y \in \mathbb{F}_p$ such that $\widetilde{Q} \equiv m \bmod p$, this yields $(p-1)p^{\operatorname{rk}\widetilde{L}'_n - 2}$ solutions; for $x = 0$, there are at most $p^{\operatorname{rk}\widetilde{L}'_n - 1}$ solutions. Therefore $p^{1-\operatorname{rk}\widetilde{L}'_n}\#\{v \in \widetilde{L}'_n/p\widetilde{L}'_n \mid Q(v) \equiv m \bmod p\} < 2$. If $\operatorname{rk}\widetilde{L}'_n = 1, 2$, [Han04, Table 1] implies that $p^{1-\mathrm{rk}\,\widetilde{L}'_n}\#\{v \in \widetilde{L}'_n/p\widetilde{L}'_n \mid Q(v) \equiv m \bmod p\} \leq 2$. Thus we conclude that $\delta(p, L'_n, m) \leq 2$.

For the second assertion, by definition, for a superspecial point, we have $\sqrt{|(L'_1 \otimes \mathbb{Z}_p)^\vee / L'_1 \otimes \mathbb{Z}_p|} = p^{t_P/2} = p$ and thus it remains to show that $\delta(p, L'_1, m) \leq 1 + p^{-1}$. As in the discussion for the first assertion, we have $\delta(p, L'_1, m) = p^{1-\mathrm{rk}\,\widetilde{L}'_1}\#\{v \in \widetilde{L}'_1/p\widetilde{L}'_1 \mid \widetilde{Q}(v) \equiv m \bmod p\}$, where $\widetilde{L}'_1/p\widetilde{L}'_n$ is a $\mathbb{F}_p$-vector space equipped with a non-degenerate quadratic form. We will prove that for any non-degenerate $\mathbb{F}_p$-quadratic space (M, Q_M) with $\dim M \geq 3$, we have

$$p^{1-\dim M}\#\{v \in M \mid Q_M(v) \equiv m \bmod p\} \leq 1 + p^{-1}$$

by induction on $\dim M$. This is enough to prove the second assertion because $p^2 || \operatorname{disc} L'_1$ and $\mathrm{rk}\, L'_1 = b + 2 \geq 5$, which implies $\mathrm{rk}\, \widetilde{L}'_1 \geq 3$. If $\dim M = 3, 4$, then the desired bound follows from [Han04, Table 1]. For $\dim M \geq 5$, we note that M is isotropic and thus can be decomposed into an orthogonal direct sum of a hyperbolic plane and a non-degenerate $\mathbb{F}_p$-quadratic space $(M', Q_{M'})$; as in the discussion for the first assertion in the previous paragraph, we have

$$p^{1-\dim M}\#\{v \in M \mid Q_M(v) \equiv m \bmod p\} = (1 - p^{-1}) + p^{-\dim M'}\#\{v \in M' \mid Q_{M'}(v) \equiv m \bmod p\}.$$

Since $\dim M' = \dim M - 2$, then by the inductive hypothesis, we have

$$p^{1-\dim M'}\#\{v \in M' \mid Q_{M'}(v) \equiv m \bmod p\} \leq 1 + p^{-1}$$

and $p^{1-\dim M}\#\{v \in M \mid Q_M(v) \equiv m \bmod p\} = (1 - p^{-1}) + p^{-\dim M'}(1 + p^{-1}) \leq 1 + p^{-1}$. □

Proposition 7.17. *There exists an absolute constant* $0 < \alpha < 1$ *such that*

$$i_P(X, T)_{\mathrm{mt}} = \alpha \sum_{m \in S_X} g_P(m) + O_T(X^{1+(b+2)/4}).$$

Proof. For brevity, we set $h = h_P$ in Definition 7.6; by §7.14, it suffices to show that

$$\sum_{n=1}^{T} \frac{q_{L'_n}(m)}{|q_L(m)|} \leq \alpha \frac{h}{p-1}$$

for some constant $0 < \alpha < 1$ (recall from Definition 7.6 that $g_P(m) = \frac{h}{p-1}|q_L(m)|$). We will prove this claim using the decay statements from Section 3 by a similar computation as in [MST, Cor. 7.2.4, Lem. 8.2.2]. We will apply these here using the fact that $L_n \otimes \mathbb{Z}_p = L'_n \otimes \mathbb{Z}_p$ and the identity

$$\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|} = \sqrt{|(L'_1 \otimes \mathbb{Z}_p)^\vee / L'_1 \otimes \mathbb{Z}_p|} \cdot |L'_1/L'_n|.$$

If P is a nonsuperspecial supersingular point, then by definition, $\sqrt{|(L'_1 \otimes \mathbb{Z}_p)^\vee / L'_1 \otimes \mathbb{Z}_p|} \geq p^2$. Moreover, by the above identity and Theorem 3.2, for $h_r + 1 \leq n \leq h_{r+1}, r \in \mathbb{Z}_{\geq 0}$, we have $\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|} \geq p^{4+2r}$. Thus by Lemma 7.16 (recall from §7.1 that $p \nmid m$ for all $m \in S_X$),

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{q_{L'_n}(m)}{|q_L(m)|} &= \sum_{n=1}^{h_0} \frac{q_{L'_n}(m)}{|q_L(m)|} + \sum_{r=0}^{\infty} \sum_{n=h_r+1}^{h_{r+1}} \frac{q_{L'_n}(m)}{|q_L(m)|} \leq \sum_{n=1}^{h_0} \frac{2}{p^2(1 - p^{-[(b+2)/2]})} + \sum_{r=0}^{\infty} \sum_{n=h_r+1}^{h_{r+1}} \frac{2}{p^{4+2r}(1 - p^{-[(b+2)/2]})} \\ &\leq \frac{2}{1 - p^{-2}} \left(\frac{h(p^{-1}+1)}{p^2} + \frac{hp}{p^4} + \frac{hp^2}{p^6} + \cdots \right) \leq \frac{h}{p-1} \cdot \frac{2(p^2 - p + 1)}{p(p^2 - 1)} \leq \frac{11}{12} \cdot \frac{h}{p-1}\end{aligned}$$

for all $p \geq 3$.

If P is superspecial and statement (1) in Theorem 3.3 holds for P, then for $h'_{r-1} + ap^r + 1 \leq n \leq h'_r, r \in \mathbb{Z}_{\geq 0}$ where $a = h/2$, we have $\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|} \geq p^{2+2r}$, and for $h'_r + 1 \leq$

$n \leq h'_r + ap^{r+1}, r \in \mathbb{Z}_{\geq 0}$, we have $\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|} \geq p^{3+2r}$. Thus for $b \geq 4$, we have $1 - p^{-[(b+2)/2]} \geq 1 - p^{-3}$ and by Lemma 7.16,

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{q_{L'_n}(m)}{|q_L(m)|} &\leq \frac{1+p^{-1}}{p(1-p^{-3})}(a(1+p^{-1})) + \frac{2(h-a)}{p^2(1-p^{-3})} + \frac{2ap}{p^3(1-p^{-3})} + \frac{2(h-a)p}{p^4(1-p^{-3})} + \cdots \\ &\leq \frac{h}{p-1}\left(\frac{(p+1)^2}{2(p^2+p+1)} + \frac{2p}{p^2+p+1}(1+p^{-1}+p^{-2}+\cdots)\right) \leq \frac{61}{62}\frac{h}{p-1}\end{aligned}$$

for all $p \geq 5$. For $b = 3$, we remark that the proof of [MST, Thm. 5.1.2] applies to all (L, Q) with $b = 3$ and L self-dual at p, not just the one associated to principally polarized abelian surfaces. Thus in this case, there is a rank 3 submodule which decays rapidly in the sense of Definition 5.1. Thus the computation in [MST, §9.2 small n's] proves that $\displaystyle\sum_{n=1}^{\infty} \frac{q_{L'_n}(m)}{|q_L(m)|} \leq \frac{11}{12}\frac{h}{p-1}$ for all $p \geq 5$.

If P is superspecial and statement (2) in Theorem 3.3 holds for P, then there exists a constant $a \leq h/2$ such that for $ap^{-1} + a + 1 \leq n \leq ap^{-1} + h$, we have $\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|} \geq p^2$, and for $h'_r + 1 \leq n \leq h'_{r+1}, r \in \mathbb{Z}_{\geq 0}$, we have $\sqrt{|(L'_n \otimes \mathbb{Z}_p)^\vee / L'_n \otimes \mathbb{Z}_p|} \geq p^{4+2r}$. Thus by Lemma 7.16

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{q_{L'_n}(m)}{|q_L(m)|} &\leq \frac{1+p^{-1}}{p(1-p^{-2})}(a(1+p^{-1})) + \frac{2(h-a)}{p^2(1-p^{-2})} + \frac{2hp}{p^4(1-p^{-2})} + \frac{2hp^2}{p^6(1-p^{-2})} + \cdots \\ &\leq \frac{h}{p-1}\left(\frac{1+p^{-1}}{2} + (p+1)^{-1} + \frac{2}{p+1}(p^{-1}+p^{-2}+p^{-3}+\cdots)\right) \leq \frac{17}{20}\frac{h}{p-1}\end{aligned}$$

for all $p \geq 5$. □

Theorem 7.18. *There is an absolute constant $0 < \alpha' < 1$ such that for S_X defined in §7.1 and for any $P \in C(k)$ supersingular, we have that*

$$\sum_{m \in S_X} i_P(C.Z(m)) = \alpha' \sum_{m \in S_X} g_P(m) + O(X^{(b+1)/2}).$$

Indeed, we may state this theorem without assuming P is supersingular since the statement here for non-supersingular P is a weaker version of Proposition 7.11.

Proof. We may take α' to be any absolute constant such that $1 > \alpha' > \alpha$, where α is given in Proposition 7.17. Then we choose $T \in \mathbb{Z}_{>0}$ such that $\frac{c_3}{T^{2/b}} X^{1+b/2} \leq (\alpha' - \alpha)\sum_{m \in S_X} g_P(m)$; such T exists since $\sum_{m \in S_X} g_P(m) \asymp X^{1+b/2}$ by Lemma 7.5. Once we fix such a T, which may be chosen only depending on α, α', S (not S_X), the desired bound follows from Propositions 7.13 and 7.17. □

Now we combine the previous results in this section to prove Theorem 1.2.

Proof of Theorem 1.2. If there were only finitely many points P in $C \cap (\cup_{p \nmid m} Z(m))(k)$, then by Proposition 7.11, Theorem 7.18, and Definition 7.6, we have that

$$\sum_{m \in S_X} C.Z(m) = \sum_{m \in S_X} \sum_{P \in C \cap (\cup_{m \in S_X} Z(m))(k)} i_P(C.Z(m)) = \alpha'(\omega.C) \sum_{m \in S_X} |q_L(m)| + O(X^{(b+1)/2}),$$

which contradicts Lemma 7.5. □

8. Application to the Hecke orbit problem

We prove Theorem 1.4 using Theorem 1.2 in this section. For $x \in \mathcal{S}_{\mathbb{F}_p}(k)$, where $k = \overline{\mathbb{F}}_p$, we use T_x to denote the set of all prime-to-p Hecke translates of x and let $\overline{T_x}$ denote the Zariski closure of T_x in $\mathcal{S}_k$. We will prove that for x ordinary, $\overline{T_x} = \mathcal{S}_k$. Our proof will be by induction on b, the dimension of $\mathcal{S}_{\mathbb{F}_p}$ – we will use Theorem 1.2 to reduce to the case of a smaller dimensional Shimura variety in the case that T_x contains a proper generically ordinary curve, with the observation that the theorem is trivial when $b = 1$. In the case where we do not have this *a priori* knowledge, we will deduce our result by studying compactifications of $\mathcal{S}_k$. We will prove the GSpin case first and in the end of this section, we will remark on how to adapt the same line of ideas to the unitary case (see Remark 8.12).

8.1. Recall from §2.1 that the quadratic lattice (L, Q) is self-dual at p and the level we pick is hyperspecial at p. By [MP19, Thm 3], the canonical integral model $\mathcal{S}$ of the Hodge type Shimura variety Sh admits a projective normal compactification $\mathcal{S}^{\mathrm{BB}}$ over $\mathbb{Z}_{(p)}$ such that $\mathcal{S}^{\mathrm{BB}}_{\mathbb{Q}}$ is the Bailey–Borel/minimal compactification Sh^{BB} of Sh; moreover, the classical stratification of Sh^{BB} by quotients by finite groups of Shimura varieties of Hodge type extends to a stratification of $\mathcal{S}^{\mathrm{BB}}$ by quotients by finite groups of integral models of these Shimura varieties; in particular, the stratification on $\mathcal{S}^{\mathrm{BB}}$ is flat. We use $\overline{T_x}^{\mathrm{BB}}$ to denote the Zariski closure of $\overline{T_x}$ in $\mathcal{S}^{\mathrm{BB}}_k$. In addition, the Hecke correspondences on $\mathcal{S}$ associated to $G(\mathbb{A}^p_f)$ extend naturally to correspondences on $\mathcal{S}^{\mathrm{BB}}$. Since these are algebraic correspondences, we have that $\overline{T_x}$ and $\overline{T_x}^{\mathrm{BB}}$ are stable under the Hecke correspondences associated to $G(\mathbb{A}^p_f)$.

Once we choose an admissible complete smooth cone decomposition, by [MP19, Thms 1, 2, 4.1.5], the canonical integral model $\mathcal{S}$ admits a smooth toroidal compactification $\mathcal{S}^{\mathrm{tor}}$ such that $\mathcal{S}^{\mathrm{tor}}_{\mathbb{Q}}$ is the toroidal compactification of Sh constructed in [AMRT10, Pin90]. Moreover, the stratification of $\mathcal{S}^{\mathrm{tor}}_{\mathbb{Q}}$ by quotients by finite groups of mixed Shimura varieties extends to a stratification of $\mathcal{S}^{\mathrm{tor}}$ with all boundary components being flat divisors and the formal completions of $\mathcal{S}^{\mathrm{tor}}$ along the boundary components of the same shape as that of $\mathcal{S}^{\mathrm{tor}}_{\mathbb{Q}}$. There is also a natural map $\pi : \mathcal{S}^{\mathrm{tor}} \to \mathcal{S}^{\mathrm{BB}}$ which extends the identity map on $\mathcal{S}$ and this map is compatible with the stratifications.

Thus for the rest of this section, we follow [BZ21, §§3.2, 3.3] and [Zem20, §4] for the explicit descriptions of $\mathcal{S}^{\mathrm{tor}}_{\mathbb{C}}, \mathcal{S}^{\mathrm{BB}}_{\mathbb{C}}$ and use it for $\mathcal{S}^{\mathrm{tor}}_{\mathbb{F}_p}$ and $\mathcal{S}^{\mathrm{BB}}_{\mathbb{F}_p}$ by the work of Madapusi Pera summarized above. In particular, the boundary components (cusps) in $\mathcal{S}^{\mathrm{BB}}_{\mathbb{F}_p}$ are either 0-dimensional or 1-dimensional.

0-dimensional cusps. We first prove Theorem 1.4 assuming that $\overline{T_x}^{\mathrm{BB}}$ contains a 0-dimensional cusp in $\mathcal{S}^{\mathrm{BB}}_{\mathbb{F}_p}$. The argument for this is essentially the same as in [Cha95, §2], and we will follow the approach there closely, indicating the places where modifications are necessary. The idea of the argument in [Cha95] is as follows. Given a 0-dimensional cusp, we study the Hecke-stabilizer of the cusp and its action on the formal neighborhood to argue that any invariant subscheme which is not $\mathcal{S}^{\mathrm{BB}}_{\mathbb{F}_p}$ is contained in the boundary.

8.2. *Coordinates.* To describe the action in coordinates, we follow the notation in [MP19] and refer to section 2 there for more details. Fix a prime $\ell \neq p$. We will work with level structure K_n given by embedding into GSp and restricting the full level ℓ^n structure there;[21] let $\mathcal{S}_{n,k}$ denote the corresponding special fiber over k of the canonical model of the Shimura variety. Given a zero-dimensional cusp x_n, we fix a cusp label representative Φ describing the cusp, which includes the data of an admissible parabolic subgroup $P \subset G_{\mathbb{Q}}$ which in this case is the stabilizer of an isotropic line in $L_{\mathbb{Q}}$. As n varies, Φ defines a compatible system of cusps $\{x_n\}$ in the inverse system $\{\mathcal{S}^{\mathrm{BB}}_{n,k}\}$ and a point $x \in \lim_{\leftarrow} \mathcal{S}^{\mathrm{BB}}_{n,k}$.

Let U_P denote the unipotent radical of P and $W \subset U_P$ denote the center of U_P. By [MP19, §2.1.11, §2.1.16], we can associate to K_n a lattice $\mathbf{B}_{K_n} \subset W(\mathbb{Q})$ with dual lattice $\mathbf{S}_{K_n} \subset W(\mathbb{Q})^\vee$ and an arithmetic group Δ_{K_n} acting on $\mathbf{B}_{K_n}$. We also have an open self-adjoint convex cone $\mathbf{H} \subset W(\mathbb{R})$

[21] Here we follow the convention in [MP19, §3.1] that we use an embedding into the group of symplectic similitudes of a symplectic space over $\mathbb{Q}$ which admits a self-dual $\mathbb{Z}$-lattice; this embedding may be different from the one in §2.2, but can be constructed from the one in §2.2 using Zarhin's trick as explained in [MP19, p. 442].

preserved by Δ_K by [MP19, §2.1.6, §2.1.16].[22] In terms of this data, by [MP19, Cor. 5.1.8, Cor. 5.2.8], the complete local ring of $\mathcal{S}_{n,k}^{\mathrm{BB}}$ at x_n is given by the ring of invariants

$$R_{\ell^n} = k[[q^\lambda]]_{\lambda\geq 0}^{\Delta_{K_n}}$$

where $\lambda \geq 0$ denotes elements of $\mathbf{S}_{K_n}$ which have non-negative pairing with $\mathbf{H}$. If we pass to the inverse limit, we get the ring

$$R_\ell = \cup_n R_{\ell^n}.$$

The Hecke correspondences at finite level are induced by an action of the group $G(\mathbf{Q}_\ell)$ on the inverse limit $\lim_{\leftarrow} \mathcal{S}_{n,k}^{\mathrm{BB}}$. In order to study Hecke-stable subvarieties, rather than study the full $G(\mathbf{Q}_\ell)$-action, it suffices to study the action of $\mathbf{B}_\ell := \mathbf{B}_{K_n} \otimes \mathbb{Z}[1/\ell] \subset W(\mathbb{Q}_\ell)$ which fixes the point x in the inverse limit $\lim_{\leftarrow} \mathcal{S}_{n,k}^{\mathrm{BB}}$ and therefore acts on the ring R_ℓ.[23] Given $T \in \mathbf{B}_\ell$, its action on $f \in R_\ell$ is given by the formula

$$f = \sum_\lambda a_\lambda q^\lambda \mapsto T(f) = \sum_\lambda \mathbf{e}((T,\lambda)) a_\lambda q^\lambda.$$

Here, $(T,\lambda) \in \mathbb{Z}[1/\ell]$ is the pairing of $T \in W(\mathbb{Q})$ and $\lambda \in W(\mathbb{Q})^\vee$ and $\mathbf{e} : \mathbb{Z}[1/\ell] \to \mu_{\ell^\infty}(k)$ is the group homomorphism defined in [Cha95, p. 452] as follows. The choice of cusp and the full level structure determines a compatible system (ζ_{ℓ^n}) of primitive ℓ^n-th roots of unity; given this, the map $\mathbf{e}(\frac{a}{\ell^n}) = (\zeta_{\ell^n})^a$ is a well-defined homomorphism.

Invariant ideals of the complete local ring. In terms of the above coordinates, the main proposition is the following, based on Proposition 2 of [Cha95].

Proposition 8.3. *Let $I_{\ell^n} \subset R_{\ell^n}$ be a nonzero ideal such that $I = I_{\ell^n} R_\ell$ is stable under the action of $\mathbf{B}_\ell$. Then* Spf R_{ℓ^n}/I_{ℓ^n} *is contained in the formal completion of the boundary of $\mathcal{S}_{n,k}^{\mathrm{BB}}$.*

Again, we merely summarize the argument from [Cha95]. Rather than work directly with R_{ℓ^n}, it is more convenient to pass to a toroidal compactification $\mathcal{S}_{n,k}^{\mathrm{tor}}$. The choice of compactification in particular specifies a smooth cone decomposition of the rational closure of the cone $\mathbf{H}$.[24] By [MP19, §§5.1.5, 2.1.17, 2.1.18], the formal completion of $\mathcal{S}_{n,k}^{\mathrm{tor}}$ along the preimage of x_n is covered by affine formal subschemes S_α parametrized by cones $\sigma_\alpha \subset \mathbf{H}$. For each such cone σ, the corresponding formal scheme has coordinate ring given by the completion R_{σ,ℓ^n} of the algebra

$$\oplus_{\lambda\in\mathbf{S}_K\cap\sigma^\vee} k \cdot q^\lambda$$

along the ideal generated by the monomials q^λ with λ a non-invertible element in the monoid $\mathbf{S}_K \cap \sigma^\vee$. Let I_σ denote the ideal of R_{σ,ℓ^n} generated by monomials q^λ where λ is strictly positive on σ. This ideal is principal, and corresponds to the reduced formal subscheme associated to the toroidal boundary.

Following Chai, let $J_\sigma \subset I_\sigma$ denote the ideal generated by q^λ where $\lambda > 0$ on $\sigma \cap \mathbf{H}$. Given $f \in R_{\sigma,\ell^n}$, we say that f has a leading term with respect to J_σ if there exists $\lambda \in \mathbf{S}_K \cap \sigma^\vee$ and $a \in k^\times$ such that $f \in aq^\lambda(1 + J_\sigma)$. This implies that the ideal generated by f is a monomial ideal, and contains a power of I_σ, so the subscheme cut out by f is contained in the toroidal boundary. The main claim to be proven is that, given I as in Proposition 8.3, for each cone σ in the decomposition of $\mathbf{H}$, there exists $f_\sigma \in I$ which has a leading term with respect to J_σ. The proof of this in [Cha95, pp. 455-456] is purely cone-theoretic, so applies identically in our setting. The key step ([Cha95, Lem. 1]) is a cancellation algorithm: given $f \in I$, and a finite collection $S = \{\lambda_0, \dots, \lambda_r\}$ for which f has nonzero coefficients, there exists $g \in I$ given by a finite linear

[22] In [MP19] there is a twist by $2\pi i$ which we are suppressing.

[23] Note that by definition in [MP19, §2.1.11], $\mathbf{B}_{K_n} \otimes \mathbb{Z}[1/\ell]$ is independent of n for our K_n.

[24] Here we call $\mathbf{H}^*$ in [MP19, §2.1.22] the rational closure of the cone $\mathbf{H}$.

combination of translates $T(f)$ for which the corresponding coefficients are all zero except for λ_0. This is proven using the explicit formula for $T(f)$.

1-dimensional cusps. We now treat the case when $\overline{T_x}^{\mathrm{BB}}$ contains at least one k-point in a 1-dimensional cusp. We choose an admissible complete smooth cone decomposition and let $\overline{T_x}^{\mathrm{tor}}$ denote the Zariski closure of T_x in $\mathcal{S}_k^{\mathrm{tor}}$. We will show that either $\overline{T_x}^{\mathrm{tor}} = \mathcal{S}_k^{\mathrm{tor}}$ or $\dim_k \overline{T_x}^{\mathrm{BB}} \setminus \overline{T_x} = 0$ and $\dim \overline{T_x}^{\mathrm{BB}} \geq 2$.

8.4. By the first paragraph in [BZ21, §3.3], there is a unique cone decomposition for a given 1-dimensional cusp and the boundary strata in $\mathcal{S}^{\mathrm{tor}}$ over 1-dimensional cusps in $\mathcal{S}^{\mathrm{BB}}$ are canonical. Thus by [MP19, Prop. 2.1.19, §4.1.12, Prop. 4.1.13], the Hecke correspondences associated to $G(\mathbb{A}_f^p)$ on Sh extend uniquely to $\pi^{-1}(\mathcal{S}^{\mathrm{BB}} \setminus \{\text{0-dim cusps}\})$ satisfying certain explicit description of these correspondences on formal completion along boundary components given in [MP19, §4.1.12]. Set $\overline{T_x}^{\mathrm{tor},1} := \overline{T_x}^{\mathrm{tor}} \cap \pi^{-1}(\mathcal{S}^{\mathrm{BB}} \setminus \{\text{0-dim cusps}\})$. Then for any $g \in G(\mathbb{A}_f^p)$, we have $g.\overline{T_x}^{\mathrm{tor},1} \supset g.\overline{T_x} = \overline{T_x}$ and thus $g.\overline{T_x}^{\mathrm{tor},1} = \overline{T_x}^{\mathrm{tor},1}$. In particular, for any $y \in \overline{T_x}^{\mathrm{tor},1}(k)$, the Zariski closure of all prime-to-p Hecke orbits of y in $\mathcal{S}_k^{\mathrm{tor}}$ is contained in $\overline{T_x}^{\mathrm{tor}}$. In particular, we will study the Hecke correspondences on a boundary point $y \in (\overline{T_x}^{\mathrm{tor},1} \setminus \overline{T_x})(k)$ in order to deduce certain properties for $\overline{T_x}$.

8.5. Let Υ be a 1-dimensional cusp in $\mathcal{S}^{\mathrm{BB}}$. We first follow [Zem20, §4] to give an explicit description of $\pi^{-1}(\Upsilon(\mathbb{C}))$. By [Zem20, Prop. 4.3, Thm. 4.5] (see also [BZ21, Lem. 3.18, Prop. 3.19]), up to quotient by a finite group, $\pi^{-1}(\Upsilon(\mathbb{C}))$ is a torsor under an abelian scheme over the modular curve (with suitable level); moreover, let $I \subset L$ be a (saturated) isotropic plane corresponding to Υ (thus the admissible parabolic in this case is the stabilizer of I) and set $\Lambda = I^{\perp}/I$, then the above mentioned abelian scheme is given by $\mathcal{E} \otimes_{\mathbb{Z}} \Lambda$, where $\mathcal{E}$ is the universal family of elliptic curves over the modular curve. Therefore, by [MP19, Thm. 4.1.5], $\pi^{-1}(\Upsilon)$ is a quotient by a finite group of a $\mathcal{E} \otimes \Lambda$-torsor over the modular curve.

Since the prime-to-p Hecke correspondences on $\pi^{-1}(\Upsilon)$ are the extensions of the Hecke correspondences on $\pi^{-1}(\Upsilon(\mathbb{C}))$ obtained by taking the normalizations of the Zariski closures of the graphs of these correspondences in characteristic 0, we first study the Hecke orbits of $y \in \pi^{-1}(\Upsilon(\mathbb{C}))$.

Proposition 8.6. *Notation as in §8.5. For $y \in \pi^{-1}(\Upsilon(\mathbb{C}))$, let $T_{y,\ell}$ denote the set of all ℓ-power Hecke translates of y. Then $T_{y,\ell}$ contains all the translates of y by ℓ-power torsion points in $\mathcal{E}_{\pi(y)} \otimes \Lambda$, where $\mathcal{E}_{\pi(y)}$ denotes the fiber of $\mathcal{E}$ at $\pi(y)$ (in the modular curve) and recall that $\pi^{-1}(\pi(y))$ is an $\mathcal{E}_{\pi(y)} \otimes \Lambda$-torsor.*

Proof. Recall that $I \subset L$ denotes the (saturated) isotropic subspace corresponding to Υ; let $P \subset G_{\mathbb{Q}} = \mathrm{GSpin}(L \otimes \mathbb{Q})$ denote the maximal parabolic which is the stabilizer of I, let U denote the unipotent radical of P, and let W denote the center of U; set $\mathcal{V} := U/W$. By [MP19, §2.1.10], $\mathcal{V}(\mathbb{Q})$ acts on the on the $\mathcal{E} \otimes \Lambda$-torsor $\pi^{-1}(\Upsilon(\mathbb{C}))$ over $\Upsilon(\mathbb{C})$ and the explicit form of this action is given by [BZ21, Lem. 3.11].

More precisely, following [Zem20, §4], we pick a $\mathbb{Z}$-basis $\{z, w\}$ of I; Using the bilinear form $[-,-]$ induced by the quadratic form Q, we naturally identify the dual $L^\vee \subset V = L \otimes \mathbb{Q}$. Let $\zeta, \omega \in L^\vee$ be a basis dual to (z, w).[25] Recall that Υ is the modular curve with suitable level and let τ be a lift of $\pi(y) \in \Upsilon(\mathbb{C})$ to the upper half plane. Then by [Zem20, Thm. 4.5, proof of Prop. 4.3, Eqns (25)(26)], $\pi^{-1}(\pi(y))$ is isomorphic to the quotient of $W_{\mathbb{C}}^{1,\tau} := \{\zeta' + \tau\omega' + e \mid e \in \Lambda \otimes_{\mathbb{Z}} \mathbb{C}\} \subset V_{\mathbb{C}}/I \otimes_{\mathbb{Z}} \mathbb{C}$ by the translation action of $(\Lambda \oplus \tau\Lambda)$. By [BZ21, Lem. 3.11], $a + b\tau \in \mathcal{V}(\mathbb{Z}[1/\ell]) \cong \Lambda \otimes \mathbb{Z}[1/\ell] \oplus \tau\Lambda \otimes \mathbb{Z}[1/\ell]$ acts by sending $\zeta + \tau\omega + e$ to $\zeta + \tau\omega + (e + a + b\tau)$. Since U is the Heisenberg group described in

[25] This means that $[-,-]$ induces an isomorphism between $\mathrm{Span}_{\mathbb{Z}}\{\zeta, \omega\}$ and $\mathrm{Hom}(I, \mathbb{Z})$ with ζ, ω mapping to the basis dual to $\{z, w\}$; the existence of such a basis is given by [Zem20, Def. 2.1, Lem. 2.2].

[Zem20, §1, Prop. 1.6, Cor. 1.9], then all elements in $\mathcal{V}(\mathbb{Z}[1/\ell])$ lift to elements in $U(\mathbb{Z}[1/\ell])$; thus the Hecke translates of y by elements in $U(\mathbb{Z}[1/\ell])$ contains all translates of y by ℓ-power torsion points in $\mathcal{E}_{\pi(y)} \otimes \Lambda$. □

Corollary 8.7. *Let $y \in \pi^{-1}(\Upsilon(k))$, where Υ is a 1-dimensional cusp of $\mathcal{S}^{\mathrm{BB}}$, and let $T_{y,\ell}$ denote the set of all ℓ-power Hecke translates of y. Then $T_{y,\ell} \cap \pi^{-1}(\pi(y))$ is Zariski dense in $\pi^{-1}(\pi(y))$.*

Proof. We first argue the Hecke action of $U(\mathbb{Z}[1/\ell])$ on y is given by translates of y by ℓ-power torsion points $\mathcal{E}_{\pi(y)} \otimes \Lambda$. Following the description in [MP19, §4.1.12, 4.1.1 - 4.1.3], the extension of Hecke correspondences from characteristic 0 to characteristic p is obtained by taking the normalizations in the sense of footnote (2) in [MP19]. In particular, if one picks a lift $\tilde{y}$ of y to characteristic 0, then the Hecke translates of y are the mod p reductions of Hecke translates of $\tilde{y}$. By Proposition 8.6 and its proof, the action on $\tilde{y}$ is given by torsion translates, so the same holds for the reduction mod p.

Note that $\pi^{-1}(\pi(y)) \simeq \mathcal{E}_{\pi(y)} \otimes \Lambda$ as varieties over k (this isomorphism is non-canonical) and thus the union of the translates of ℓ-power torsion points is Zariski dense in $\pi^{-1}(\pi(y))$ since the set of ℓ-power torsion points of an abelian variety over k is Zariski dense. □

Corollary 8.8. *Recall that $x \in \mathcal{S}_{\mathbb{F}_p}(k)$ ordinary and assume that $b \geq 3$. If $\overline{T_x}^{\mathrm{BB}}$ contains a k-point which lies on a 1-dimensional cusp of $\mathcal{S}_{\mathbb{F}_p}^{\mathrm{BB}}$. Then either (1) $\overline{T_x} = \mathcal{S}_k$ or (2) $\dim_k \overline{T_x}^{\mathrm{BB}} \setminus \overline{T_x} = 0$ and $\dim \overline{T_x}^{\mathrm{BB}} \geq 2$.*

Proof. Since $\overline{T_x}^{\mathrm{BB}}$ is stable under Hecke correspondences, then for any 1-dimensional cusp Υ, we have that $\overline{T_x}^{\mathrm{BB}} \cap \Upsilon_k$ is stable under the Hecke correspondences associated to $\mathrm{GL}_2(\mathbb{A}_f^p)$ on Υ_k. Thus $\overline{T_x}^{\mathrm{BB}} \cap \Upsilon_k = \Upsilon_k$ or $\dim_k \overline{T_x}^{\mathrm{BB}} \cap \Upsilon_k = 0$.

If there exists an Υ such that $\overline{T_x}^{\mathrm{BB}} \cap \Upsilon_k = \Upsilon_k$, then $\pi(\overline{T_x}^{\mathrm{tor}}) = \overline{T_x}^{\mathrm{BB}} \supset \Upsilon_k$. By Corollary 8.7 and §8.4, we have that $\overline{T_x}^{\mathrm{tor}} \supset \pi^{-1}(\Upsilon_k)$ and thus $\dim_k \overline{T_x} \geq \dim_k \pi^{-1}(\Upsilon_k) + 1 = \dim \mathcal{S}_k$. Moreover, since $G(\mathbb{A}_f^p)$ acts transitively on the π_0 of the inverse limit of the canonical models of Sh with varying levels away from p by [Kis10, Lem. 2.2.5], then by the definition of canonical integral models, the only Hecke-stable subvariety of $\mathcal{S}_k$ of dimension $\dim \mathcal{S}_k$ must be the entire $\mathcal{S}_k$ and thus $\overline{T_x} = \mathcal{S}_k$.

If for any 1-dimensional cusp Υ, we have $\dim_k \overline{T_x}^{\mathrm{BB}} \cap \Upsilon_k = 0$, then $\dim_k \overline{T_x}^{\mathrm{BB}} \setminus \overline{T_x} = 0$. On the other hand, by the assumption, there exists $y' \in \Upsilon(k)$ for some Υ such that $y' \in \overline{T_x}^{\mathrm{BB}}$; then there exists $y \in \pi^{-1}(\Upsilon)(k)$ such that $y \in \overline{T_x}^{\mathrm{tor}}$ and $\pi(y) = y'$. By Corollary 8.7, we have $\dim_k \overline{T_x} \geq 1 + \dim \overline{T_{y,\ell}} = b - 1 \geq 2$. Thus we conclude that (2) holds. □

Proposition 8.9 (Existence of a proper curve)**.** *Let $x \in \mathcal{S}_{\mathbb{F}_p}(k)$ be ordinary. Then, either $\overline{T_x} = \mathcal{S}_k$ or $\overline{T_x}$ contains a proper curve which is generically ordinary.*

Proof. The prime-to-p Hecke orbit of an ordinary point is always infinite, and so $\overline{T_x}$ has dimension at least 1. The case where $\mathcal{S}_k$ is one-dimensional follows.

Suppose that $b = 2$, i.e., $\mathcal{S}_k$ is two-dimensional. Then, $\mathcal{S}_k^{\mathrm{BB}}$ contains 1-dimensional cusps if and only if the reductive group defining $\mathcal{S}$ is $\mathbb{Q}$-split if and only if $\mathcal{S}$ is a product of two modular curves. In this case, the density of ordinary Hecke orbits follows from the product structure of $\mathcal{S}$ and thus $\overline{T_x} = \mathcal{S}_k$. Therefore, suppose that $\mathcal{S}^{\mathrm{BB}}$ does not contain any 1-dimensional cusps. Then, $\overline{T_x}$ either equals $\mathcal{S}_k$ or $\overline{T_x}$ is a proper curve (in which cases the lemma follows), or $\overline{T_x}^{\mathrm{BB}}$ intersects the boundary of $\mathcal{S}_k^{\mathrm{BB}}$ non-trivially. We now have $\overline{T_x}^{\mathrm{BB}} = \mathcal{S}_k^{\mathrm{BB}}$ by Proposition 8.3 and the lemma follows.

Therefore, we may assume that $b \geq 3$. If $\overline{T_x}$ is proper, the lemma follows. Otherwise, $\overline{T_x}^{\mathrm{BB}}$ intersects the boundary non-trivially. Suppose that $\overline{T_x}^{\mathrm{BB}}$ contains a zero-dimensional cusp. Then, the lemma follows from Proposition 8.3. Therefore, suppose that $\overline{T_x}^{\mathrm{BB}}$ in $\mathcal{S}_k^{\mathrm{BB}}$ contains a point in a 1-dimensional cusp. Then either (1) or (2) of Corollary 8.8 must hold. The lemma follows directly if Case (1) holds so we assume that we are in Case (2). But in this case, we have that $\dim_k \overline{T_x}^{\mathrm{BB}} \setminus \overline{T_x} = 0$ and $\dim \overline{T_x}^{\mathrm{BB}} \geq 2$ – the lemma follows in this case, because it is always possible to find a proper curve in the two dimensional *projective*[26] variety $\overline{T_x}^{\mathrm{BB}}$ passing through an ordinary point which avoids the finitely many boundary points $\overline{T_x}^{\mathrm{BB}} \setminus \overline{T_x}$. □

Proof of the Hecke orbit conjecture. We first recall some results on Hecke orbits which we will need. As the results and their proofs are standard, we will content ourselves with only a sketch of their proofs.

Lemma 8.10. *Let $f : Sh_1 \to Sh_2$ be a morphism of Shimura varieties of Hodge type with hyperspecial level at p and let $G_i, i = 1, 2$ denote the reductive group of Sh_i. Let $\mathcal{S}_i$ denote the canonical integral model of Sh_i and we still use f to denote the unique extension $f : \mathcal{S}_1 \to \mathcal{S}_2$ (such an extension exists by the theory of canonical models; see* [Kis10, Thm. 2.3.8]*). Let $X \subset \mathcal{S}_{2,k}$ be a subvariety that intersects the ordinary locus (here we assume that the ordinary locus in $\mathcal{S}_{2,k}$ is not empty), and let $\overline{T_X}$ denote the Zariski closure of the Hecke orbit T_X of X with respect to the Hecke correspondences associated to $G_2(\mathbb{A}_f^p)$. Then*

(1) for any Shimura subvariety $Z \subset Sh_2$, we have that $\overline{T_X} \subset \mathcal{S}_{2,k}$ is not contained in $\mathcal{Z}_k$, where $\mathcal{Z}$ denotes the Zariski closure of Z in $\mathcal{S}_2$;

(2) $f^{-1}(\overline{T_X})$ is stable under the Hecke correspondences associated to $G_1(\mathbb{A}_f^p)$ on $\mathcal{S}_{1,k}$.

Proof. (1) In order to discuss ℓ-adic monodromy, we fix a geometric point $x \in \overline{T_X}$ and let A_x denote the fiber of the universal abelian variety over $\mathcal{S}_2$ at x; although in the statement of the lemma, we consider $\overline{T_X}$ over k, in the proof here, we choose a finite extension of $\mathbb{F}_p$ over which $\overline{T_X}$ is defined as we will use the full arithmetic étale fundamental group (not just the geometric part). By [Kis10, §2.2] (see also [MP16, Prop. 3.11]) and [AGHMP18, Remark 4.2.3], the ℓ-adic lisse sheaf given by the relative $H^1_{\ell,\text{ét}}$ of the universal abelian variety is endowed with tensors (as global sections of the suitable tensor products of the sheaf and its dual) and these tensors restricted to the fiber $H^1_{\ell,\text{ét}}(A_x)$ cut out a subgroup $G_2(\mathbb{Q}_\ell) \subset \mathrm{GL}(H^1_{\ell,\text{ét}}(A_x))$. Moreover, at each point in $\overline{T_X}$, the Galois group at the point fixes these tensors, by combining [Kis10, Lemma 2.2.1] with the comparison map of étale cohomology groups between characteristic 0 and characteristic p. Therefore, the ℓ-adic monodromy representation $\rho_{\ell,x} : \pi_1^{\text{ét}}(\overline{T_X}, x) \to \mathrm{GL}(H^1_{\ell,\text{ét}}(A_x))$ factors through $G_2(\mathbb{Q}_\ell) \subset \mathrm{GL}(H^1_{\ell,\text{ét}}(A_x))$. (The same conclusion holds for any subvariety of $\mathcal{S}_{2,k}$ in place of $\overline{T_X}$.)

The image of the ℓ-adic monodromy of the ℓ-adic lisse sheaf given by the relative $H^1_{\ell,\text{ét}}$ of the universal abelian variety restricted to any Hecke-stable subvariety in $\mathcal{S}_{2,k}$ must be Zariski dense in $G^{\mathrm{ad}}_{2,\mathbb{Q}_\ell}$ via the quotient map $G_2 \to G_2^{\mathrm{ad}}$. Indeed, the Hecke correspondences associated to $G_2(\mathbb{Q}_\ell)$ on the Hecke-stable subvariety induce the conjugation action of $G_2(\mathbb{Q}_\ell)$ on $\mathrm{End}(H^1_{\ell,\text{ét}})$ and thus the ℓ-adic monodromy must be stable under the conjugation action and thus a normal subgroup of $G_2(\mathbb{Q}_\ell)$.

Moreover, if we pick any ordinary point in the Hecke-stable subvariety, the Frobenius at this point must have non-identity component in each $\mathbb{Q}_\ell$-simple factor of $G^{\mathrm{ad}}_{2,\mathbb{Q}_\ell}$. Indeed, since the Hodge cocharacter is non-trivial on each $\mathbb{Q}$-simple factor of G_2^{ad}, the Frobenius is non-trivial on each $\mathbb{Q}$-simple factor of G_2^{ad}; the claim on $\mathbb{Q}_\ell$-simple factor then follows from the fact [Kis17, Cor. 2.3.1] that the Frobenius is conjugate to an element in $G_2^{\mathrm{ad}}(\mathbb{Q})$. Consequently, the image of the ℓ-adic monodromy in $G_2^{\mathrm{ad}}(\mathbb{Q}_\ell)$ is a normal subgroup which

[26] because $\mathcal{S}^{\mathrm{BB}}_{\mathbb{F}_p}$ is projective by [MP19, Thm. 3]

has nontrivial projection onto every $\mathbb{Q}_\ell$-simple factor of $G^{\mathrm{ad}}_{2,\mathbb{Q}_\ell}$, and so must be Zariski dense in $G^{\mathrm{ad}}_{2,\mathbb{Q}_\ell}$.

Note that since T_X is Hecke-stable, then $\overline{T_X}$ is Hecke-stable as all Hecke correspondences are algebraic. It then follows that $\overline{T_X}$ is not contained in any $\mathcal{Z}_k$ since the ℓ-adic monodromy of the family of abelian varieties over Z is contained in the reductive group associated to Z, whose image is a proper algebraic subgroup of $G^{\mathrm{ad}}_{2,\mathbb{Q}_\ell}$.

(2) Note that $\overline{T_X}$ is stable under $G_2(\mathbb{A}^p_f)$, then it suffices to prove that for any $x \in \mathcal{S}_1(k)$ and for any $g \in G_1(\mathbb{A}^p_f)$, if $x' \in g.x$, then there exists $g' \in G_2(\mathbb{A}^p_f)$ such that $f(x') \in g'.f(x)$. Indeed, we may take $g' = f(g)$, where we view $f : G_1 \to G_2$, and the desired property follows from the definition of Hecke correspondences via the extension property of canonical integral models given in [Kis10, Thm. 2.3.8]. □

Lemma 8.11. *Notation as in Proposition 8.9; the proper curve in $\overline{T_x}$ can be chosen such that it is not contained in any special divisor $Z(m)$.*

Proof. By the proof of Lemma 8.10(1), there exists at least one irreducible component of $\overline{T_x}$ whose ℓ-adic monodromy group is Zariski dense in $G^{\mathrm{ad}}_{\mathbb{Q}_\ell}$ and we only need to show that there exists a curve $C' \subset \overline{T_x}$ which has the same ℓ-adic monodromy as this irreducible component of $\overline{T_x}$. (As we will not use any other property of $\overline{T_x}$ other than having large monodromy group, by abuse of notation, we will still use $\overline{T_x}$ to denote this irreducible component with large monodromy group so for the rest of the proof, $\overline{T_x}$ is irreducible.) More precisely, fix a geometric point $y \in C'$ and let A_y denote the Kuga–Satake abelian variety at y, then the image of the ℓ-adic monodromy representation $\rho_{\ell,C',y} : \pi_1^{\text{ét}}(C', y) \to \pi_1^{\text{ét}}(\overline{T_x}, y) \to \mathrm{GL}(H^1_{\ell,\text{ét}}(A_y))$ coincides with the image of $\pi_1^{\text{ét}}(\overline{T_x}, y)$ (and indeed, we will see from the proof below that this restriction can be combined with the proof of Proposition 8.9 to construct a generically ordinary and proper curve C').

Since $\overline{T_x}$ is positive dimensional, we may pick a Zariski local coordinate and view an open part of $\overline{T_x}$ as a variety T of dimension $\dim_k \overline{T_x} - 1$ over $k(t)$. To find a curve C' in $\overline{T_x}$, it suffices to find a k'-point in T, where k' is some finite extension of $k(t)$. If we replace $k(t)$ by a number field, then the desired assertion (the existence of points in T over number fields with the same ℓ-adic monodromy group as the generic point) is exactly [And96, Thm. 5.2 (3)], which not only proves the existence, but also shows that the set of points not having the largest possible monodromy is thin (mince in French) in the sense of [Ser89, §9.1 Definition]; see for instance [Ser89, p. 149] for a proof which reduces the claim to the Hilbert irreducibility theorem. Serre's proof holds word-by-word in the global function field case once we replace the classical Hilbert irreducibility theorem by the analogous statement for global function fields, which is [BSE21, Thm. 1.1]. Since [BSE21, Cor. 3.5] gives a quantitative version of the comparison of Galois group which is more directly related to what we need, we will finish the proof using this corollary following Serre's argument.

More precisely, the ℓ-adic monodromy map over T defines a Galois extension over the function field of T, i.e., the Galois group is isomorphic to the ℓ-adic monodromy group G_ℓ of T_x. We start from the simplified case: assume that this extension were finite and that there exists a non-empty Zariski open subset $U \subset T$ contained in $\mathbb{A}^{\dim_{k(t)} T}_{k(t)}$. In this case, we refer the reader to [Ser89, p. 123, first paragraph] or [BSE21, paragraph after Rmk. 3.1] for the concrete definition of how to specialize the Galois extension from the generic point to k'-points. In our case, with the ℓ-adic Galois representation associated to T, the Galois group G_y of the specialization of the extension at a k'-point y is the image of $\mathrm{Gal}(\overline{k'}/k')$ in the monodromy representation (well-defined up to conjugation). As the Galois extension of the function field of U, which is also the function field of T, is assumed to be finite, we may find a monic, irreducible, and separable polynomial with coefficients in regular functions on $\mathbb{A}^{\dim_{k(t)} T}_{k(t)}$ such that this Galois extension is the splitting field of this polynomial. By [BSE21, Cor. 3.5] and the proof of [BSE21, Thm. 3.6], the set $\{y \in U(k(t)) \mid G_y \not\simeq G_\ell\}$ is of density 0.

In order to treat the profinite group G_ℓ, Serre proved that the Frattini subgroup $\Phi(G_\ell)$ of G_ℓ, which is defined to be the intersection of all maximal subgroups of G_ℓ, is open in G_ℓ (see [Ser89, pp. 148-149]) and thus we reduce the question to study the finite Galois extension associated to $G_\ell/\Phi(G_\ell)$. More precisely, we conclude as above that there exists a density 1 set of y such that $G_y\Phi(G_\ell)/\Phi(G_\ell) = G_\ell/\Phi(G_\ell)$ and hence $G_y = G_\ell$ by the definition of the Frattini group. In general, there exists a non-empty Zariski open subset $U \subset T$ which admits a finite étale map to a Zariski open subset of $\mathbb{A}_{k(t)}^{\dim_{k(t)} T}$; we may apply the above argument to the monodromy group of the pushforward of the local system $H^1_{\ell,\text{ét}}$ on U and the preimages in U of any $k(t)$-point not in the bad density 0 set in $\mathbb{A}_{k(t)}^{\dim_{k(t)} T}$ give rise to k'-points with maximal possible monodromy group for some finite extension k' of $k(t)$. □

Proof of Theorem 1.4 orthogonal case. We will induct on $\dim_k \mathcal{S}_k = b$. When $b = 1$, $\mathcal{S}_k$ is a curve; since the prime-to-p Hecke orbit of an ordinary point is infinite, its Zariski closure must be positive dimensional and thus the base case is verified.

Now assume that $b \geq 2$ and that Theorem 1.4 holds for all ordinary points in the special fiber of the canonical integral model of GSpin Shimura varieties of dimension $b-1$ with hyperspecial level. Consider $x \in \mathcal{S}(k)$ ordinary and $\overline{T_x}$, the Zariski closure of the prime-to-p Hecke orbit of x.

By Proposition 8.9, $\overline{T_x}$ is either equal to $\mathcal{S}_k$ (in which case we are done), or $\overline{T_x}$ contains a proper curve C' that is generically ordinary and we may assume that C' is not contained in any special divisor $Z(m)$ by Lemma 8.11. We now apply Theorem 1.2 to the normalization C of C' with the natural map $C \to C' \to \mathcal{S}_k$; in the case of $b = 2$, we apply the proof of [MST, Thm. 1(2)] instead. As a result, there exists an ordinary point x' on $C' \subset \overline{T_x}$ such that $x' \in Z(m)(k)$ for some $p \nmid m$ representable by (L, Q), as there are only finitely many non-ordinary points on C'.

Let $\mathcal{S}' \subset \mathcal{Z}(m)$ denote the canonical integral model of the Shimura subvariety of $\mathcal{S}$ which consists some irreducible components of $\mathcal{Z}(m)$ and $x' \in \mathcal{S}'(k)$. Note that since $p \nmid m$, $\mathcal{S}'$ has hyperspecial level at p and $\dim_k \mathcal{S}'_k = b - 1$.[27] By Lemma 8.10(2), $\overline{T_x} \cap \mathcal{S}'_k$ is a generically ordinary Hecke-stable subvariety of $\mathcal{S}'_k$. Then by the inductive hypothesis, we have that $\overline{T_x} \cap \mathcal{S}'_k = \mathcal{S}'_k$, and thus $\mathcal{S}'_k \subset \overline{T_x}$. In fact, an identical argument yields that $Z'(m) \subset \overline{T_x}$ for infinitely many m, where $Z'(m)$ is some irreducible component of $Z(m)$; indeed, if there were only finitely many such $Z'(m)$, they only intersect C at finitely many k-points and we may always pick x' different from these finitely many points when we apply Theorem 1.2. Since the Zariski closure of infinitely many distinct subvarieties of dimension $b - 1$ must be at least b-dimensional, we conclude that $\overline{T_x}$ must contain at least one irreducible component of $\mathcal{S}_k$. Moreover, since the Hecke action $G(\mathbb{A}_f^p)$ on the inverse limit of $\mathcal{S}_k$ with varying levels away from p permutes all its irreducible/connected components, we conclude that $\overline{T_x} = \mathcal{S}_k$. □

Remark 8.12. Let $\mathcal{S}_k$ denote the mod $\mathfrak{p}$ special fiber of the canonical integral model $\mathcal{S}$ over $\operatorname{Spec} \mathcal{O}_{K,(\mathfrak{p})}$ of the PEL type unitary Shimura variety considered in [RSZ20, §3, §4.1] and [RSZ21, §3.4, §4.1] where $\mathfrak{p} \mid p$ and p splits in $K/\mathbb{Q}$ and p does not divide the discriminant of the Hermitian form (we work with the special case of [RSZ20, RSZ21] that the CM field is imaginary quadratic; if we further

[27] More precisely, as explained on [AGHMP18, p. 434] that $\mathcal{Z}(m)_{\mathbb{Q}}$ is a finite disjoint union of GSpin Shimura varieties associated to quadratic spaces isomorphic to $(v^\perp, Q|_{v^\perp}) \subset (V, Q)$, where $v \in L$ with $Q(v) = m$; since $p \nmid m$, then $(v^\perp, Q|_{v^\perp})$ is self-dual at p and the embedding $v^\perp \subset V$ also induces a hyperspecial level for the Shimura variety associated to $v^\perp$. By [AGHMP18, Prop. 4.4.2] (or [MP16, Cor. 6.23]), $\mathcal{Z}(m)$ is normal and flat and thus is the disjoint union of canonical integral models of the GSpin Shimura varieties associated to isomorphism classes of $v^\perp$ with $Q(v) = m$.

restrict ourselves to the principally polarized case, see also [BHK$^+$20, §2.1] and [KR14, §2.1, Notation 2.6]). The assumption that p is not ramified in $K/\mathbb{Q}$ and p does not divide the discriminant implies that we can work with the hyperspecial level and the PEL Shimura variety has good reduction at $\mathfrak{p}$ and the assumption that p splits in $K/\mathbb{Q}$ implies that the ordinary locus in $\mathcal{S}_k$ is nonempty. More concretely, following [RSZ21, Def. 3.9, Def. 4.1], given a Hermitian lattice L self-dual at p, $\mathcal{S}$ is the moduli space of the moduli problem which associates to a locally noetherian $\mathcal{O}_{K,(\mathfrak{p})}$-scheme T the groupoid of $(A_0, A, \iota_0, \iota, \lambda)$, where[28]

- A_0 is an elliptic curve over T,
- $\iota_0 : \mathcal{O}_K \to \mathrm{End}(A_0)$ such that $\mathcal{O}_K$ acts on $\mathrm{Lie}\, A_0$ via the structure map $\mathcal{O}_K \to \mathcal{O}_{K,(\mathfrak{p})} \to \mathcal{O}_T$;
- A is an abelian scheme over T of relative dimension $(n+1)$,
- $\iota : \mathcal{O}_K \to \mathrm{End}(A)$ satisfies the Kottwitz determinant condition $\det(t - \iota(\alpha)|\,\mathrm{Lie}\, A) = (t-\alpha)^n(t-\overline{\alpha}) \in \mathcal{O}_T[t]$,
- $\lambda : A \to A^\vee$ is a quasi-polarization such that
  - its Rosati involution satisfies $\iota(\alpha)^\dagger = \iota(\overline{\alpha})$ for all $\alpha \in \mathcal{O}_K$,
  - λ induces a principal polarization on the p-divisible group $A[p^\infty]$,
  - $\mathrm{Hom}_{\widehat{\mathcal{O}_K}^p}(\widehat{T}^p(A_0), \widehat{T}^p(A))$ is isomorphic to $L \otimes_{\mathcal{O}_K} \widehat{\mathcal{O}_K}^p$ as Hermitian spaces, where $\widehat{T}^p(A_0), \widehat{T}^p(A))$ denote the product of prime-to-p Tate modules and $\widehat{\mathcal{O}_K}^p$ denotes the product of completions $\mathcal{O}_{K,\ell}$ away from p and the Hermitian form on $\mathrm{Hom}_{\widehat{\mathcal{O}_K}^p}(\widehat{T}^p(A_0), \widehat{T}^p(A))$ is given by $x \mapsto x^\vee \circ x \in \mathrm{Hom}_{\widehat{\mathcal{O}_K}^p}(\widehat{T}^p(A_0), \widehat{T}^p(A_0)) \otimes \mathbb{A}^p_{K,f} \cong \mathbb{A}^p_{K,f}$, where $(-)^\vee$ is the dual map induced by the polarizations on A_0 and A.

There are special divisors in $\mathcal{S}$ described in [RSZ20, §3.5] (see also [BHK$^+$20, §2.5] and[KR14, §2.2, Def. 2.8]), parametrizing $(A_0, A, \iota_0, \iota, \lambda, x)$ with $(A_0, A, \iota_0, \iota, \lambda)$ as above and $x \in \mathrm{Hom}_{\mathcal{O}_K}(A_0, A)$.

By [SSTT, §9.3], given $C \to \mathcal{S}_k$ such that the image of the generic point of C is ordinary, we can construct a morphism from (a finite étale cover of) C to the special fiber of the canonical integral model of a GSpin Shimura variety associated to a quadratic space of signature $(2n, 2)$ such that the image of generic point of C is ordinary; moreover, the construction also has the property that if $P \in C(k)$ maps to a point in a special divisor in the GSpin special fiber, then P also maps to a point in a special divisor in the unitary special fiber. Thus, as a direct consequence of Theorem 1.2, we prove that there are infinitely many k-points on C which lie in the union of special divisors in $\mathcal{S}_k$; moreover, we may further assume that the special morphisms $x \in \mathrm{Hom}_{\mathcal{O}_K}(A_0, A)$ corresponding to points on the special divisors have the property that $x^\vee \circ x \in \mathrm{Hom}_{\mathcal{O}_K}(A_0, A_0) \cong \mathcal{O}_K$ is coprime to p. This is the analogue of Theorem 1.2 in the unitary case. Note that similar to the orthogonal case, the smaller unitary Shimura varieties associated to the prime-to-p special divisors still have discriminants of the Hermitian spaces prime to p.

In order to prove Theorem 1.4 unitary case, we use the above analogue of Theorem 1.2 in the unitary case and adapt the above inductive proof for the orthogonal case to the unitary case if $\overline{T_x}$ is proper; thus to finish the proof, it remains to treat the case when $\overline{T_x}^{\mathrm{BB}}$ hits the boundary of $\mathcal{S}_k^{\mathrm{BB}}$.

The arithmetic compactifications of $\mathcal{S}$ are described in [BHK$^+$20, §3].[29] More precisely, by [BHK$^+$20, Thm. 3.7.1, Prop. 3.4.4], the boundary components of $\mathcal{S}^{\mathrm{BB}}$ are 0-dimensional (relative to $\mathrm{Spec}\,\mathcal{O}_{K,\mathfrak{p}}$); the toroidal compactification $\mathcal{S}^{\mathrm{tor}}$ is canonical and the fibers over the cusps of $\mathcal{S}^{\mathrm{tor}} \to \mathcal{S}^{\mathrm{BB}}$ are abelian schemes and each of these abelian schemes, up to quotient by a finite group, is isomorphic (over some finite extension of $\mathcal{O}_{K,(\mathfrak{p})}$) to $E \otimes_{\mathcal{O}_K} \Lambda_0$, where E is an elliptic curve CM

[28] Here we work with isomorphism classes of abelian varieties; one may also describe the moduli problem in the prime-to-p isogeny category of abelian varieties; see for instance [LZ21, §11.2] for a short summary of the discussion in [RSZ20, RSZ21].

[29] Even though [BHK$^+$20] works with principal polarization case, since we work with hyperspecial level at p, the description also applies to our case here.

by $\mathcal{O}_K$ and Λ_0 is an $\mathcal{O}_K$-lattice of rank $n-1$. Since $\mathcal{S}^{\mathrm{tor}}$ is canonical, the Hecke correspondences associated to $G(\mathbb{A}_f^p)$ (here G denotes the reductive group associated to $\mathcal{S}$) extend to $\mathcal{S}^{\mathrm{tor}}$. For each cusp in $\mathcal{S}^{\mathrm{BB}}$, we may choose an isotropic line $J \subset W$, where W is the Hermitian space over K of signature $(n,1)$ used to define $\mathcal{S}$. The admissible parabolic associated to the cusp is the stabilizer of J and by [How15, §3.3, p. 673], the $\mathbb{Z}[1/\ell]$-points of the unipotent part of this parabolic acts on $E \otimes_{\mathcal{O}_K} \Lambda_0$ by translations of ℓ-power torsion points and thus we prove the analogous statement of Proposition 8.6 for the unitary case. Therefore we prove the unitary case of Theorem 1.4 by the proof of Corollary 8.8.

References

[And96] Yves André, *Pour une théorie inconditionnelle des motifs*, Inst. Hautes Études Sci. Publ. Math. **83** (1996), 5–49 (French).

[AGHMP18] Fabrizio Andreatta, Eyal Z. Goren, Benjamin Howard, and Keerthi Madapusi Pera, *Faltings heights of abelian varieties with complex multiplication*, Ann. of Math. (2) **187** (2018), no. 2, 391–531.

[AMRT10] Avner Ash, David Mumford, Michael Rapoport, and Yung-Sheng Tai, *Smooth compactifications of locally symmetric varieties*, 2nd ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2010. With the collaboration of Peter Scholze.

[BSE21] Lior Bary-Soroker and Alexei Entin, *Explicit Hilbert's irreducibility theorem in function fields*, Abelian varieties and number theory, Contemp. Math., vol. 767, Amer. Math. Soc., [Providence], RI, 2021, pp. 125–134.

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

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

[BHK+20] Jan H. Bruinier, Benjamin Howard, Stephen S. Kudla, Michael Rapoport, and Tonghai Yang, *Modularity of generating series of divisors on unitary Shimura varieties*, Astérisque **421, Diviseurs arithmétiques sur les variétés orthogonales et unitaires de Shimura** (2020), 7–125.

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

[BZ21] Jan Hendrik Bruinier and Shaul Zemel, *Special cycles on toroidal compactifications of orthogonal Shimura varieties*, Math. Ann. (2021). published online.

[Cha95] Ching-Li Chai, *Every ordinary symplectic isogeny class in positive characteristic is dense in the moduli*, Invent. Math. **121** (1995), no. 3, 439–479.

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

[Cha05] ______, *Hecke orbits on Siegel modular varieties*, Geometric methods in algebra and number theory, Progr. Math., vol. 235, Birkhäuser Boston, Boston, MA, 2005, pp. 71–107.

[Cha06] ______, *Hecke orbits as Shimura varieties in positive characteristic*, International Congress of Mathematicians. Vol. II, Eur. Math. Soc., Zürich, 2006, pp. 295–312.

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

[CO09] ______, *Moduli of abelian varieties and p-divisible groups*, Arithmetic geometry, Clay Math. Proc., vol. 8, Amer. Math. Soc., Providence, RI, 2009, pp. 441–536. MR2498069

[CO19] ______, *The Hecke orbit conjecture: a survey and outlook*, Open problems in arithmetic algebraic geometry, Adv. Lect. Math. (ALM), vol. 46, Int. Press, Somerville, MA, 2019, pp. 235–262.

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

[COU01] Laurent Clozel, Hee Oh, and Emmanuel Ullmo, *Hecke operators and equidistribution of Hecke points*, Invent. Math. **144** (2001), no. 2, 327–351.

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

[EK95] Alex Eskin and Yonatan R. Katznelson, *Singular symmetric matrices*, Duke Math. J. **79** (1995), no. 2, 515–547, DOI 10.1215/S0012-7094-95-07913-7.

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

[How15] Benjamin Howard, *Complex multiplication cycles and Kudla-Rapoport divisors, II*, Amer. J. Math. **137** (2015), no. 3, 639–698.

[HP20] Benjamin Howard and Keerthi Madapusi Pera, *Arithmetic of Borcherds products*, Astérisque **421, Diviseurs arithmétiques sur les variétés orthogonales et unitaires de Shimura** (2020), 187–297.

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

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

[Kea88] Kevin Keating, *Lifting endomorphisms of formal A-modules*, Compositio Math. **67** (1988), no. 2, 211–239.

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

[Kis17] ———, mod p *points on Shimura varieties of abelian type*, J. Amer. Math. Soc. **30** (2017), no. 3, 819–914.

[KR14] Stephen Kudla and Michael Rapoport, *Special cycles on unitary Shimura varieties II: Global theory*, J. Reine Angew. Math. **697** (2014), 91–157.

[LZ21] Chao Li and Wei Zhang, *Kudla–Rapoport cycles and derivatives of local densities*, J. Amer. Math. Soc. (2021). published online.

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

[MP19] ———, *Toroidal compactifications of integral models of Shimura varieties of Hodge type*, Ann. Sci. Éc. Norm. Supér. (4) **52** (2019), no. 2, 393–514 (English, with English and French summaries).

[MST] Davesh Maulik, Ananth N. Shankar, and Yunqing Tang, *Reductions of abelian surfaces over global function fields*, Compos. Math. to appear.

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

[Noo96] Rutger Noot, *Models of Shimura varieties in mixed characteristic*, J. Algebraic Geom. **5** (1996), no. 1, 187–207.

[Ogu03] Keiji Oguiso, *Local families of $K3$ surfaces and applications*, J. Algebraic Geom. **12** (2003), no. 3, 405–433.

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

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

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

[Pin90] Richard Pink, *Arithmetical compactification of mixed Shimura varieties*, Bonner Mathematische Schriften [Bonn Mathematical Publications], vol. 209, Universität Bonn, Mathematisches Institut, Bonn, 1990. Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, 1989.

[RSZ20] M. Rapoport, B. Smithling, and W. Zhang, *Arithmetic diagonal cycles on unitary Shimura varieties*, Compos. Math. **156** (2020), no. 9, 1745–1824.

[RSZ21] ———, *On Shimura varieties for unitary groups*, Pure Appl. Math. Q. **17** (2021), no. 2, 773–837.

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

[Ser89] Jean-Pierre Serre, *Lectures on the Mordell-Weil theorem*, Aspects of Mathematics, E15, Friedr. Vieweg & Sohn, Braunschweig, 1989. Translated from the French and edited by Martin Brown from notes by Michel Waldschmidt.

[Sha] Ananth N. Shankar, *The Hecke orbit conjecture for "modèles étranges"*. preprint.

[SSTT] Ananth N. Shankar, Arul Shankar, Yunqing Tang, and Salim Tayou, *Exceptional jumps of Picard ranks of reductions of K3 surfaces over number fields*. available on arXiv: 1909.07473.

[ST20] Ananth N. Shankar and Yunqing Tang, *Exceptional splitting of reductions of abelian surfaces*, Duke Math. J. **169** (2020), no. 3, 397–434.

[Tay20] Salim Tayou, *On the equidistribution of some Hodge loci*, J. Reine Angew. Math. **762** (2020), 167–194.

[Voi02] Claire Voisin, *Théorie de Hodge et géométrie algébrique complexe*, Cours Spécialisés [Specialized Courses], vol. 10, Société Mathématique de France, Paris, 2002 (French). viii+595.

[Xia] Luciena X. Xiao, *On The Hecke Orbit Conjecture for PEL Type Shimura Varieties*. arXiv:2006.06859.

[Zem20] Shaul Zemel, *The structure of integral parabolic subgroups of orthogonal groups*, J. Algebra **559** (2020), 95–128.

[Zho] Rong Zhou, *Motivic cohomology of quaternionic Shimura varieties and level raising*. arXiv:1901.01954.

Department of mathematics, Massachusetts Institute of Technology
E-mail address: maulik@mit.edu

Department of mathematics, University of Wisconsin, Madison
E-mail address: ashankar@math.wisc.edu

Department of mathematics, Princeton University
E-mail address: yunqingt@math.princeton.edu