Algebra & Number Theory

Volume 19
2025
No. 1

msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2025 is US $565/year for the electronic version, and $820/year (+$70, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

ALGEBRA AND NUMBER THEORY 19:1 (2025)
https://doi.org/10.2140/ant.2025.19.1

A modification of the linear sieve, and the count of twin primes

Jared Duker Lichtman

We introduce a modification of the linear sieve whose weights satisfy strong factorization properties, and consequently equidistribute primes up to size x in arithmetic progressions to moduli up to $x^{10/17}$. This surpasses the level of distribution $x^{4/7}$ with the linear sieve weights from well-known work of Bombieri, Friedlander, and Iwaniec, and which was recently extended to $x^{7/12}$ by Maynard. As an application, we obtain a new upper bound on the count of twin primes. Our method simplifies the 2004 argument of Wu, and gives the largest percentage improvement since the 1986 bound of Bombieri, Friedlander, and Iwaniec.

1. Introduction

Given a finite set $\mathcal{A}$ of positive integers, sieve methods offer a broad framework for estimating the number of elements in $\mathcal{A}$ all whose prime factors exceed z, denoted by $S(\mathcal{A}, z)$, in terms of the approximate density $g_{\mathcal{A}}(d) = g(d)$ of multiples of d in $\mathcal{A}$, denoted by $\mathcal{A}_d$. Note one often expects

$$S(\mathcal{A}, z) \approx |\mathcal{A}| \prod_{p<z} (1 - g(p)).$$

Combinatorial sieves may be viewed as refinements of the basic inclusion-exclusion principle, and are described by a sequence of weights $\lambda(d) \in \{-1, 0, 1\}$ supported on integers up to some level $D \geqslant 1$. We refer the reader to Opera de Cribro [Friedlander and Iwaniec 2010] for a more thorough introduction to the subject.

In particular, the upper bound weights $\lambda^+(d)$ for the linear sieve satisfy

$$S(\mathcal{A}, z) \leqslant |\mathcal{A}| \prod_{p<z} (1 - g(p)) \left(F\left(\frac{\log D}{\log z}\right) + o(1) \right) + \sum_{\substack{d \leqslant D \\ p \mid d \Rightarrow p < z}} \lambda^+(d)(|\mathcal{A}_d| - |\mathcal{A}| g(d)) \tag{1-1}$$

as $D \to \infty$, provided $g = g_{\mathcal{A}}$ satisfies some mild conditions. Here the function $F : \mathbb{R}_{\geqslant 1} \to \mathbb{R}_{\geqslant 1}$ is defined by a delay-differential equation, as in (2-5). For sets $\mathcal{A}$ sufficiently equidistributed in arithmetic progressions the second sum over $d \leqslant D$ in (1-1) contributes negligibly, in which case the main term is

$$S(\mathcal{A}, z) \lesssim |\mathcal{A}| \prod_{p<z} (1 - g(p)) F(s),$$

MSC2020: primary 11N35, 11N36; secondary 11N05.
Keywords: linear sieve, well-factorable weights, level of distribution, switching principle, Buchstab identity.

where $z = D^{1/s}$. In fact, $F(s) \to 1$ as $s \to \infty$ so the main term confirms the naïve expectation in this case. Moreover, F is optimal in the sense that the bound (1-1) is attained sharply for a particular set $\mathcal{A}$.

We introduce this sieve theory setup more formally in Section 2.2 below, and define the sieve weights λ^+ explicitly in Section 3; see [Friedlander and Iwaniec 2010, Section 12] for further details on the linear sieve ($\beta = 2$), as well as [loc. cit., Section 11] for its generalization to the β-sieve.

The linear sieve is powerful when combined with equidistribution estimates which make the final sum in (1-1) small. For example, the Bombieri–Vinogradov theorem shows that for every ε, $A > 0$, letting $Q = x^{1/2-\varepsilon}$ we have

$$\sum_{q \leqslant Q} \sup_{(a,q)=1} \left| \pi(x; q, a) - \frac{\pi(x)}{\varphi(q)} \right| \ll_{\varepsilon, A} \frac{x}{(\log x)^A}. \tag{1-2}$$

So by taking $D = Q$, (1-1) can give a good upper bound when the set $\mathcal{A}$ is related to the primes, such as when $\mathcal{A} = \{p + 2 : p \leqslant x\}$, in which case (1-1) gives an upper bound for the count of twin primes.

The estimate (1-2) may be viewed as an assertion of the generalized Riemann hypothesis on average over moduli up to $Q = x^{1/2-\varepsilon}$. It remains an important open problem to extend the range to $Q = x^{1/2+\delta}$ for some fixed $\delta > 0$. Indeed, Elliott and Halberstam [1970] conjectured such an extension up to $Q = x^{1-\varepsilon}$ for any $\varepsilon > 0$.

In some contexts it suffices to relax the setup in (1-2) in order to raise the level of distribution. In particular, in the case of a fixed residue class $a \in \mathbb{Z}$, and the absolute values replaced by well-factorable weights $\lambda(q)$ (see Definition 2.1), the celebrated result of Bombieri, Friedlander and Iwaniec [Bombieri et al. 1986] raised the level up to $Q = x^{4/7-\varepsilon}$,

$$\sum_{\substack{q \leqslant Q \\ (q,a)=1}} \lambda(q) \left(\pi(x; q, a) - \frac{\pi(x)}{\varphi(q)} \right) \ll_{a, A, \varepsilon} \frac{x}{(\log x)^A}. \tag{1-3}$$

While the linear sieve weights are not themselves well-factorable, Iwaniec [1980] constructed a well-factorable variant $\tilde{\lambda}^+$ of the weights λ^+ (and so (1-3) holds with $\lambda = \tilde{\lambda}^+$), which are only slightly altered from λ^+ so that $\tilde{\lambda}^+$ enjoys an analogous linear sieve bound as in (1-1), notably with an identical form of the main term,

$$S(\mathcal{A}, z) \leqslant |\mathcal{A}| \prod_{p<z} (1 - g(p)) \left(F\left(\frac{\log D}{\log z} \right) + o(1) \right) + \sum_{\substack{d \leqslant D \\ p \mid d \Rightarrow p < z}} \tilde{\lambda}^+(d)(|\mathcal{A}_d| - |\mathcal{A}| g(d)). \tag{1-4}$$

The bound (1-3) stood for several decades, but quite recently Maynard [2020] managed to extend the level in (1-3) further to $Q = x^{7/12-\varepsilon}$ in the case of the weights $\lambda = \tilde{\lambda}^+$. Given the currently available equidistribution estimates for primes, we note the level $x^{7/12}$ is a natural barrier for these weights.

In this article, we modify the technical construction of the linear sieve weights to avoid this barrier, and thereby produce new sieve weights that induce stronger equidistribution estimates for primes.

Theorem 1.1. *Let $D = x^{10/17-\varepsilon}$. There exists a sequence $\tilde{\lambda}^*(d) \in \{-1, 0, 1\}$ satisfying*:

(1) *Equidistribution for primes*: *For any fixed $a \in \mathbb{Z}$, $A, \varepsilon > 0$, we have*

$$\sum_{\substack{d \leqslant D \\ (d,a)=1}} \tilde{\lambda}^*(d) \left(\pi(x; d, a) - \frac{\pi(x)}{\varphi(d)} \right) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}.$$

(2) *Sieve upper bound*: *For $s \geqslant 1$, $z = D^{1/s}$, we have*

$$S(\mathcal{A}, z) \leqslant |\mathcal{A}| \prod_{p<z} (1 - g(p))(F^*(s) + o(1)) + \sum_{\substack{d \leqslant D \\ p \mid d \Rightarrow p<z}} \tilde{\lambda}^*(d)(|\mathcal{A}_d| - |\mathcal{A}| g(d)),$$

where $F^*(s) \leqslant 1.000081 F(s)$ when $1 \leqslant s \leqslant 3$, for the linear sieve function F as in (2-5).

The key feature of Theorem 1.1 is to obtain equidistribution up to level $x^{10/17}$ at the cost of only a tiny loss in the main term. See Theorem 2.12 and Proposition 5.4 for full technical statements and additional variations that may be of independent interest.

1.1. ***Application to twin primes.*** We expect that Theorem 1.1 should give numerous improvements to sieve bounds related to the primes. As proof of concept in this direction, we give a new upper bound for the count of twin primes up to x, denoted by $\pi_2(x)$. Recall Hardy and Littlewood [1923] conjectured the asymptotic formula

$$\pi_2(x) \sim \frac{2x}{(\log x)^2} \prod_{p>2} \frac{1 - 2/p}{(1 - 1/p)^2} =: \Pi(x). \tag{1-5}$$

Theorem 1.2. *As x tends to infinity, we have*

$$\pi_2(x) \lesssim 3.29956 \Pi(x).$$

Theorem 1.2 gives a 2.94% refinement from the previous record bound of Wu [2004]. For reference, this gives the largest percentage improvement since the work of Bombieri, Friedlander, and Iwaniec [1986]. See Table 1 for a chronology of the known upper bounds on $\pi_2(x)/\Pi(x)$. Also see Siebert [1976], Riesel and Vaughan [1983, Lemma 5] for numerically explicit forms of Selberg's bound [1952].

The main ingredients for these results come from applying sieve bounds to the set $\mathcal{A} = \{p + 2 : p \leqslant x\}$, and using equidistribution of primes in arithmetic progressions to handle remainder terms. Bombieri and Davenport obtained $\pi_2(x)/\Pi(x) \lesssim 4$ as a consequence of the Bombieri–Vinogradov theorem (1-2) and a standard sieve upper bound of level $x^{1/2-\varepsilon}$. More generally, if one proves level of distribution $x^{\theta-\varepsilon}$ then one immediately obtains $\pi_2(x)/\Pi(x) \lesssim 2/\theta$. Bombieri, Friedlander and Iwaniec proved $\pi_2(x)/\Pi(x) \lesssim \frac{7}{2}$ by the well-factorable variant (1-3) level of distribution $x^{4/7-\varepsilon}$, together with the linear sieve with well-factorable remainder (1-4).

Year	Author(s)	$\pi_2(x)/\Pi(x) \lesssim$
1919	Brun [1919]	O(1)
1947	Selberg [1952]	8
1964	Pan [1964]	6
1966	Bombieri and Davenport [1966]	4
1978	Chen [1978]	3.9171
1983	Fouvry and Iwaniec [1983]	$3.7777\cdots = 34/9$
1984	Fouvry [1984]	$3.7647\cdots = 64/17$
1986	Bombieri, Friedlander and Iwaniec [1986]	3.5
1986	Fouvry and Grupp [1986]	3.454
1990	Wu [1990]	3.418
2003	Cai and Lu [2003]	3.406
2004	Wu [2004]	3.39951

Table 1. Upper bounds for $\pi_2(x)/\Pi(x)$.

The other key ingredient to subsequent improvements is the *switching principle*, introduced in Chen's celebrated result [1973] that there are infinitely many primes p such that $p+2$ has at most two prime factors. The basic insight is to use a weighted sieve inequality to split the problem into multiple cases, apply sieve bounds to $\mathcal{A} = \{p+2 : p \leqslant x\}$ in certain cases, and then reinterpret the remaining cases as new sieving problems for switched sets $\mathcal{B} = \{m-2 \leqslant x\}$ where the numbers m are constructed from $\mathcal{A}$ (as prescribed depending on the case).

1.2. ***Outline of main ideas in Theorem 1.1.*** Maynard's new equidistribution results show equidistribution of the primes with sieve weights $\tilde{\lambda}^+(d)$, provided $d = p_1 \cdots p_r$ is restricted to suitably well-factorable integers. Unfortunately, the original linear sieve weights only partially satisfy these well-factorable conditions. In particular for $\eta > 0$, when looking at the linear sieve of level $x^{7/12+\eta}$, some integers d in its support do not satisfy the conditions, which means that $x^{7/12}$ is the limit for the linear sieve given our current equidistribution technology. Nevertheless, the key observation here is that only a few exceptional d fail to satisfy these conditions. Moreover up to level $x^{10/17}$, i.e., $\eta < \frac{1}{204}$, the anatomy of exceptional d may be precisely characterized in terms of η (given specifically as $\mathcal{P}_4$, $\mathcal{P}_6$ in (3-6)). In particular, as $\eta > 0$ grows the family of exceptional integers contribute $O(\eta^5)$ to the sieve bound. However, we note this characterization breaks down when $\eta \geqslant \frac{1}{204}$, and the contribution becomes considerably larger and more complicated.

As such we carefully revise the construction of the linear sieve, altering a few particular inclusion-exclusion steps in order to avoid the exceptional integers d with bad factorizations. Once these terms no longer contribute to the sieve, this produces a worse and more complicated main term, but since there are only a very small number of such terms the resulting loss is small. And since these modified weights now satisfy stronger factorization properties in their support, we can now leverage the full strength of Maynard's equidistribution results.

Notation

We use the Vinogradov $\ll$ and $\gg$ asymptotic notation, and the big oh $O(\,\cdot\,)$ and $o(\,\cdot\,)$ asymptotic notation. We use $f \sim g$, $f \lesssim g$, and $f \gtrsim g$ to denote $f = (1+o(1))g$, $f \leqslant (1+o(1))g$, and $f \geqslant (1+o(1))g$, respectively. Dependence on a parameter will be denoted by a subscript.

The letter p will always be reserved to denote a prime number, $\pi(x)$ is the prime counting function, and $\pi(x; d, a)$ is the count of primes up to x congruent to a (mod d). We use φ to denote the Euler totient function, μ the Möbius function, and $e(x) := e^{2\pi i x}$ the complex exponential. We use $\mathbf{1}$ to denote the indicator function of a statement. For example, for a set A denote

$$\mathbf{1}_{a\in A} = \begin{cases} 1, & \text{if } a \in A, \\ 0, & \text{else,} \end{cases} \qquad \mathbf{1}^{a_0 \notin A}_{a_1,\dots,a_i \in A} = \begin{cases} 1, & \text{if } a_1, \dots, a_i \in A \text{ and } a_0 \notin A, \\ 0, & \text{else.} \end{cases}$$

Finally, we refer to various sieve weights referred places throughout the article, so we take a moment to list them here:

We generically write λ to denote a sequence of weights in $\{0, \pm 1\}$. In particular, λ^+ and λ^- refer to the (upper and lower bound) weights of the linear sieve, given by restrictions of the Möbius function, $\lambda^{\pm}(d) = \mu(d)\mathbf{1}_{d\in\mathcal{D}^{\pm}}$. Analogously, the modified (upper bound) linear sieve weights λ^* are given by $\lambda^*(d) = \mu(d)\mathbf{1}_{d\in\mathcal{D}^*}$. Here the support sets $\mathcal{D}^{\pm}$ and $\mathcal{D}^*$ are defined in (3-1) and (3-5). We also write $\lambda^{(r)}$ to refer to the weights λ^+ or λ^- (and $\mathcal{D}^{(r)}$ to refer to $\mathcal{D}^+$ or $\mathcal{D}^-$), depending on whether r is odd or even.

The well-factorable weights $\tilde{\lambda}^{\pm}$ are defined in (5-17). Following Iwaniec, this construction involves certain auxiliary weights at intermediate steps, namely, $\lambda_{(D_1,\dots,D_r)}$ defined in (5-14), and $\lambda^{(r)}_{(D_1,\dots,D_r)} = \lambda_{(D_1,\dots,D_r)} * \lambda^{(r)}$ defined in (5-15). The analogous construction starting from λ^* gives modified weights $\tilde{\lambda}^*$, defined in (5-16).

2. Technical setup and results

2.1. *Factorization of weights and their level of distribution.*

Definition 2.1 (well-factorable). Let $Q \in \mathbb{R}_{\geqslant 1}$. A sequence $\lambda(q)$ is *well-factorable of level* Q, if for every factorization $Q = Q_1 Q_2$ into $Q_1, Q_2 \in \mathbb{R}_{\geqslant 1}$, there exist sequences γ_1, γ_2 such that:

(1) $|\gamma_1(q_1)|, |\gamma_2(q_2)| \leqslant 1$ for all $q_1, q_2 \in \mathbb{N}$.

(2) $\gamma_i(q) = 0$ if $q \notin [1, Q_i]$ for $i = 1, 2$.

(3) We have $\lambda = \gamma_1 * \gamma_2$, i.e.,

$$\lambda(q) = \sum_{q = q_1 q_2} \gamma_1(q_1)\gamma_2(q_2).$$

Bombieri, Friedlander and Iwaniec [1986, Theorem 10] established level of distribution $x^{4/7-\varepsilon}$ with well-factorable weights.

Theorem 2.2 [Bombieri et al. 1986]. *Fix any $a \in \mathbb{Z}$ and let $A, \varepsilon > 0$. For any well-factorable sequence λ of level $Q \leqslant x^{4/7-\varepsilon}$, we have*

$$\sum_{\substack{q \leqslant Q \\ (q,a)=1}} \lambda(q)\left(\pi(x; q, a) - \frac{\pi(x)}{\varphi(q)}\right) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}.$$

Maynard [2020] considered a natural strengthening of well-factorable sequences.

Definition 2.3 (triply well-factorable). Let $Q \in \mathbb{R}_{\geqslant 1}$. A sequence $\lambda(q)$ is *triply well-factorable of level Q*, if for every factorization $Q = Q_1Q_2Q_3$ into $Q_1, Q_2, Q_3 \in \mathbb{R}_{\geqslant 1}$, there exist sequences $\gamma_1, \gamma_2, \gamma_3$ such that:

(1) $|\gamma_1(q_1)|, |\gamma_2(q_2)|, |\gamma_3(q_3)| \leqslant 1$ for all $q_1, q_2, q_3 \in \mathbb{N}$.

(2) $\gamma_i(q) = 0$ if $q \notin [1, Q_i]$ for $i = 1, 2, 3$.

(3) We have $\lambda = \gamma_1 * \gamma_2 * \gamma_3$, i.e.,

$$\lambda(q) = \sum_{q=q_1q_2q_3} \gamma_1(q_1)\gamma_2(q_2)\gamma_3(q_3).$$

The definitions of well-factorable and triply well-factorable sequences are quite natural and relatively simple from a conceptual standpoint. Maynard [2020, Theorem 1.1] obtains powerful equidistribution results for triply well-factorability that are beyond the scope of well-factorability. Unfortunately, triply well-factorability is too restrictive a condition for us to produce Theorem 1.1. As such we are forced to identify the precise mechanism that enables Maynard's equidistribution results, and extract the following technical definition that is implicit in [Maynard 2020].[1]

Definition 2.4 (programmably factorable). Let $0 < \delta < 10^{-5}$. For $x \in \mathbb{R}_{>1}$, a sequence $\lambda(q)$ is *programmably factorable of level Q (relative to x, δ)*, if for every $N \in [x^{2\delta}, x^{1/3+\delta/2}]$ there exists a factorization $Q = Q_1Q_2Q_3$ with $Q_1, Q_2, Q_3 \in \mathbb{R}_{\geqslant 1}$, satisfying the system

$$\begin{aligned} Q_1 &\leqslant Nx^{-\delta}, \\ N^2Q_2Q_3^2 &\leqslant x^{1-\delta}, \\ N^2Q_1Q_2^4Q_3^3 &\leqslant x^{2-\delta}, \\ NQ_1Q_2^5Q_3^2 &\leqslant x^{2-\delta}. \end{aligned} \tag{2-1}$$

And for every such factorization $Q = Q_1Q_2Q_3$ there exist sequences $\gamma_1, \gamma_2, \gamma_3$ such that:

(1) $|\gamma_1(q_1)|, |\gamma_2(q_2)|, |\gamma_3(q_3)| \leqslant 1$ for all $q_1, q_2, q_3 \in \mathbb{N}$.

(2) $\gamma_i(q) = 0$ if $q \notin [1, Q_i]$ for $i = 1, 2, 3$.

[1] Indeed, the definition of programmably factorable in the special case $Q_3 = 1$ gives the implicit condition (which is implied by well-factorable) that enables Bombieri, Friedlander and Iwaniec to get equidistribution (1-3); also see Lemma 5 in [Fouvry and Grupp 1986].

(3) We have $\lambda = \gamma_1 * \gamma_2 * \gamma_3$, i.e.,

$$\lambda(q) = \sum_{q = q_1 q_2 q_3} \gamma_1(q_1)\gamma_2(q_2)\gamma_3(q_3).$$

Programmable factorability is the key technical definition in this article. It is named in allusion to the linear programming-type system of inequalities (2-1) that the factors satisfy. The diagram below displays the various implications among the definitions:

$$\begin{tikzcd}
\lambda \text{ is \textit{triply well-factorable} of level } Q \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \lambda \text{ is \textit{well-factorable} of level } Q \\
\lambda \text{ is \textit{programmably factorable} of level } Q \text{ (relative to } x,\ \delta) &
\end{tikzcd}$$

In the key result [Maynard 2020, Theorem 1.1], Maynard extended the level of distribution up to $Q < x^{3/5}$ for programmably factorable weights. Note that level $x^{3/5}$ is the natural barrier for (2-1) to admit a solution.

Theorem 2.5 [Maynard 2020]. *Fix any $a \in \mathbb{Z}$ and let A, $\varepsilon > 0$. For any programmably factorable sequence λ of level $Q \leqslant x^{3/5-\varepsilon}$ (relative to x, $\varepsilon/50$), we have*

$$\sum_{\substack{q \leqslant Q \\ (q,a)=1}} \lambda(q)\left(\pi(x; q, a) - \frac{\pi(x)}{\varphi(q)}\right) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}.$$

Remark 2.6. Theorem 1.1 in [Maynard 2020] was stated for triply factorable sequences, but its proof in fact gives the result for programmably factorable sequences.

Note the weights $\tilde{\lambda}^+$ are composed of well-factorable — but not necessarily programmably factorable — sequences of given level D. Nevertheless, Maynard showed the upper bound weights $\tilde{\lambda}^+$ of sieve level $D = x^{7/12-\varepsilon}$ are programmably factorable of level $D \leqslant Q = x^{3/5-\varepsilon}$ (relative to x, $\varepsilon/50$). By Theorem 2.5 this gives [Maynard 2020, Theorem 1.2] below.

Corollary 2.7 [Maynard 2020]. *For any fixed $a \in \mathbb{Z}$ and A, $\varepsilon > 0$, the weights $\tilde{\lambda}^+$ from (2-4) of sieve level $D = x^{7/12-\varepsilon}$ satisfy*

$$\sum_{\substack{d \leqslant D \\ (d,a)=1}} \tilde{\lambda}^+(d)\left(\pi(x; d, a) - \frac{\pi(x)}{\varphi(d)}\right) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}.$$

Later, in Proposition 5.4, we shall obtain technical improvements of Corollary 2.7 for Iwaniec's weights $\tilde{\lambda}^\pm$ (both upper and lower), in special cases where equidistribution is restricted to moduli which are smooth, or otherwise amenable to programmable factorization.

We may summarize the definitions and results of the section up to this point as follows:

λ is *triply well-factorable* of level Q:

- Equidistributed for $Q < x^{3/5}$.
- Can take $\lambda = \tilde{\lambda}^+$ for $D \leqslant Q^{2/3}$.

λ is *well-factorable* of level Q:

- Equidistributed for $Q < x^{4/7}$.
- Can take $\lambda = \tilde{\lambda}^+$ for $D \leqslant Q$.

λ is *programmably factorable* of level Q (relative to x, δ):

- Equidistributed for $Q < x^{3/5}$.
- Can take $\lambda = \tilde{\lambda}^+$ for $D \leqslant Q < x^{7/12}$.

For each type of sequence, we have outlined their corresponding levels of distribution, and the levels at which the type is satisfied by the upper bound weights for the linear sieve. Observe well-factorability is flexible enough to accommodate the linear sieve to any level, but has weaker equidistribution. On the other hand, triple well-factorability has stronger equidistribution, but is too rigid to accommodate the linear sieve (at nontrivial levels). Finally, programmable factorability also has strong equidistribution in addition to (nontrivially) accommodating the linear sieve, though at the cost of conceptual technicality.

Remark 2.8. In general, λ well-factorable of level Q directly implies λ triply well-factorable of level $Q^{2/3}$.[2] In particular, for $\lambda = \tilde{\lambda}^+$ the triply well-factorable level $Q^{2/3} < x^{2/5}$ is sharp.[3]

2.2. ***Sieve theory setup and bounds.*** We recall the standard sieve-theoretic notation. Given a finite set $\mathcal{A} \subset \mathbb{N}$, set of primes $\mathcal{P}$, and a threshold $z > 0$, we define $\mathcal{A}_d = \{n \in \mathcal{A} : d \mid n\}$ and remainder $r_{\mathcal{A}}$ via

$$|\mathcal{A}_d| = g(d)|\mathcal{A}| + r_{\mathcal{A}}(d),$$

where g is a multiplicative function, with $0 \leqslant g(p) < 1$ for $p \in \mathcal{P}$ (we assume $g(p) = 0$ if $p \notin \mathcal{P}$). Also define $P(z) = \prod_{p<z,\, p\in\mathcal{P}} p$ and $V(z) = \prod_{p \mid P(z)}(1 - g(p))$. The central object of interest is the *sifted sum*

$$S(\mathcal{A}, z) = S(\mathcal{A}, \mathcal{P}, z) = \sum_{n\in\mathcal{A}} \mathbf{1}_{(n,P(z))=1}. \tag{2-2}$$

Later for our application of interest, we will set $g(d) = 1/\varphi(d)$. For now, it suffices for us to assume for all $2 \leqslant w \leqslant z$,

$$\frac{V(w)}{V(z)} = \prod_{\substack{w\leqslant p<z\\ p\in\mathcal{P}}} (1 - g(p)) = \frac{\log z}{\log w}\left(1 + O\left(\frac{1}{\log w}\right)\right). \tag{2-3}$$

Remark 2.9. The proof of the upper bound for the standard linear sieve only requires a one-sided inequality for $V(w)/V(z)$, whereas our modification requires the above two-sided condition (2-3).

[2] Indeed, take any factorization $Q = Q_1Q_2Q_3$, with (say) $Q_1 \geqslant Q_2 \geqslant Q_3 \geqslant 1$. Note $Q_3 \leqslant Q^{1/3}$. If λ is well-factorability of level $Q/Q_3 = Q_1Q_2$, there are sequences γ_1, γ_2 supported on $[1, Q_1]$, $[1, Q_2]$ with $\lambda = \gamma_1 * \gamma_2 = \gamma_1 * \gamma_2 * \delta$. Here $\delta(q) = \mathbf{1}_{q=1}$. Hence λ is triply well-factorable of level $\inf_{Q=Q_1Q_2Q_3} Q/Q_3 \geqslant Q^{2/3}$.

[3] Indeed, consider the factorization $Q = Q_1Q_2Q_3$ with $(Q_1, Q_2, Q_3) = (Q^{1/3-\varepsilon}, Q^{1/3-\varepsilon}, Q^{1/3+2\varepsilon})$. Then for $q = p_1p_2p_3$ of size $p_1, p_2 \sim Q^{1/3}$, $p_3 \sim Q^{1/9}$, we see $p_1, p_2 > Q_1 = Q_2$. Thus all sequences γ_i supported on Q_i satisfy $\gamma_1 * \gamma_2 * \gamma_3(q) = 0$. In particular $\gamma_1 * \gamma_2 * \gamma_3 \neq \tilde{\lambda}^+$.

The basic result which we shall adapt is the linear sieve with well-factorable remainder, as in [Friedlander and Iwaniec 2010, Theorem 12.20].

Theorem 2.10 [Friedlander and Iwaniec 2010]. *Let $\varepsilon > 0$ and $D > 1$ be sufficiently small and large, respectively. Then for $s \geqslant 1$ and $z = D^{1/s}$, we have*

$$S(\mathcal{A}, z) \leqslant |\mathcal{A}| V(z)(F(s) + O(\varepsilon)) + \sum_{d \mid P(z)} \tilde{\lambda}^{+}(d) r_{\mathcal{A}}(d),$$

$$S(\mathcal{A}, z) \geqslant |\mathcal{A}| V(z)(f(s) + O(\varepsilon)) - \sum_{d \mid P(z)} \tilde{\lambda}^{-}(d) r_{\mathcal{A}}(d),$$

where the implied constant only depends that of (2-3). *Here the weights $\tilde{\lambda}^{\pm}$ are*

$$\tilde{\lambda}^{\pm}(d) = \sum_{j \leqslant \exp(\varepsilon^{-3})} \lambda_j^{\pm}(d) \tag{2-4}$$

for some well-factorable sequences $\lambda_j^{\pm}$ of level D. The functions $F, f : \mathbb{R}^{+} \to \mathbb{R}$ satisfy the system of delay-differential equations

$$\begin{aligned} sF(s) &= 2e^{\gamma} \quad [s \leqslant 3] \; (sF(s))' = f(s-1), \\ sf(s) &= 0 \qquad [s \leqslant 2] \; (sf(s))' = F(s-1). \end{aligned} \tag{2-5}$$

Remark 2.11. See [Iwaniec 1980, Theorem 1] for an alternate formulation and proof, which gives sharper quantitative bounds than $O(\varepsilon)$. However, it is more technical than necessary for our purposes.

The main result of this article is the following modification of the linear sieve with programmably factorable remainder.

Theorem 2.12. *Let $\mathcal{A}$ be a finite set of positive integers with density function $g(d)$ satisfying* (2-3), *and $F(s)$ the function defined by the system* (2-5). *Let $\varepsilon > 0$ and $x > 1$ be sufficiently small and large, respectively. Then for $\eta \geqslant 0$, $D = x^{7/12+\eta}$, $s \geqslant 1$, and $z = D^{1/s}$, we have*

$$S(\mathcal{A}, z) \leqslant |\mathcal{A}| V(z)(F^{*}(s) + O(\varepsilon)) + \sum_{d \mid P(z)} \tilde{\lambda}^{*}(d) r_{\mathcal{A}}(d),$$

where the implied constant only depends that of (2-3). *Here the weights $\tilde{\lambda}^{*}$ are*

$$\tilde{\lambda}^{*}(d) = \sum_{j \leqslant \exp(\varepsilon^{-3})} \lambda_j^{*}(d) \tag{2-6}$$

for some programmably factorable sequences λ_j^{*} of level D (relative to x, $\varepsilon/50$). For $\eta < \frac{1}{204}$ we have $F^{*}(s) = F(s) + O(\eta^5)$, and $F^{*}(s) \leqslant 1.000081 F(s)$ for $1 \leqslant s \leqslant 3$, $\eta = \frac{1}{204}$.

Note Theorem 2.5, applied to each $\lambda = \lambda_j^{*}$ above, immediately implies the following.

Corollary 2.13. *Given any fixed $a \in \mathbb{Z}$ and A. For $\eta < \frac{1}{204}$, the weights $\tilde{\lambda}^*$ as in* (2-6) *of level $D = x^{7/12+\eta}$ satisfy*

$$\sum_{\substack{d \leqslant D \\ (d,a)=1}} \tilde{\lambda}^*(d) \left(\pi(x; d, a) - \frac{\pi(x)}{\varphi(d)} \right) \ll_{a,A,\eta} \frac{x}{(\log x)^A}.$$

3. Programmably factorable support

The upper and lower bound weights $\lambda^{\pm}$ for the linear sieve of level D are defined by

$$\lambda^{\pm}(d) = \mu(d) \mathbf{1}_{d \in \mathcal{D}^{\pm}},$$

where $\mathcal{D}^{\pm} = \mathcal{D}^{\pm}(D)$ are the standard support sets

$$\begin{aligned} \mathcal{D}^{+} &= \{p_1 \cdots p_r : D^{1/2} \geqslant p_1 \geqslant \cdots \geqslant p_r, \text{ and } p_1 \cdots p_{l-1} p_l^3 \leqslant D \text{ for each odd } l \leqslant r\}, \\ \mathcal{D}^{-} &= \{p_1 \cdots p_r : D^{1/2} \geqslant p_1 \geqslant \ldots \geqslant p_r, \text{ and } p_1 \cdots p_{l-1} p_l^3 \leqslant D \text{ for each even } l \leqslant r\}. \end{aligned} \tag{3-1}$$

We may also write $\mathcal{D}^{(r)}$ to denote $\mathcal{D}^{+}$ or $\mathcal{D}^{-}$, when r is even or odd, respectively.

Observe that both sets satisfy the containment $\mathcal{D}^{\pm}(D) \subset \mathcal{D}^{\text{well}}(D)$, where

$$\mathcal{D}^{\text{well}} = \{p_1 \cdots p_r : D^{1/2} \geqslant p_1 \geqslant \ldots \geqslant p_r \text{ and } p_1 \cdots p_{l-1} p_l^2 \leqslant D \text{ for each } l \leqslant r\}. \tag{3-2}$$

We shall return to this observation later in the section.

Maynard [2020] deduces Corollary 2.7 for $\tilde{\lambda}^{+}$ from the general Theorem 2.5 by means of the following key result [loc. cit., Proposition 9.1] (along with a construction of Iwaniec we shall address in later sections), which programmably factorizes elements of the support $\mathcal{D}^{+}$.

Proposition 3.1 [Maynard 2020]. *Let $0 < \delta < 10^{-3}$ and let $D = x^{7/12-50\delta}$, $N \in [x^{2\delta}, x^{1/3+\delta/2}]$. Then every $d \in \mathcal{D}^{+}(D)$ has a factorization $d = d_1 d_2 d_3$ such that $d_1 \leqslant N x^{-\delta}$ and*

$$N^2 d_2 d_3^2 \leqslant x^{1-\delta}, \quad N^2 d_1 d_2^4 d_3^3 \leqslant x^{2-\delta}, \quad N d_1 d_2^5 d_3^2 \leqslant x^{2-\delta}. \tag{3-3}$$

Remark 3.2. The level $x^{7/12}$ is sharp in this construction. Indeed, *heuristically speaking*, the linear sieve weights are not programmably factorable of level $D = x^{7/12+\eta}$ for any $\eta > 0$, because the support set contains obstructing (families of) elements $d \in \mathcal{D}^{+}(D)$ of the form $d = p_1 \cdots p_r$ where $p_1 \approx \cdots \approx p_6 \approx D^{1/7}$, or where $p_1 \approx p_2 \approx D^{2/7}$ and $p_3 \approx p_4 \approx D^{1/7}$. This heuristic description of the obstructions is made precise by the families $\mathcal{P}_4$, $\mathcal{P}_6$ (in (3-6) below), and thereby tells us how we should restrict the support set in order to increase the level (namely, to $\mathcal{D}^*$ in (3-5) below).

For $\eta > 0$, level $D = x^{7/12+\eta}$, we define the modified weights $\lambda^* = \lambda^*_{\eta}$,

$$\lambda^*(d) = \mu(d) \mathbf{1}_{d \in \mathcal{D}^*}, \tag{3-4}$$

for the support set $\mathcal{D}^*$,

$$\mathcal{D}^* = \mathcal{D}^{+}(x^{7/12}) \cup \{p_1 \cdots p_r \in \mathcal{D}^{+}(x^{7/12+\eta}) : p_1 \cdots p_i \notin \mathcal{P}_i,\ i \leqslant r,\ i \in \{4, 6\}\}. \tag{3-5}$$

Here $\mathcal{P}_4$ and $\mathcal{P}_6 = \mathcal{P}_{6,1} \cup \mathcal{P}_{6,2}$ are exceptional subsets of $\mathcal{D}^+(x^{7/12+\eta})$, given by

$$\begin{aligned}\mathcal{P}_4 &= \{p_1 \cdots p_4 : p_1 < x^{1/6+2\eta} \text{ and } p_2 p_4 > x^{1/4-3\eta}\},\\ \mathcal{P}_{6,1} &= \{p_1 \cdots p_6 : p_1 p_2 < x^{1/6+2\eta} \text{ and } p_2 p_3 p_4 > x^{1/4-3\eta} \text{ and } p_6 > x^{1/12-5\eta}\},\\ \mathcal{P}_{6,2} &= \{p_1 \cdots p_6 : p_1,\, p_2 p_3 < x^{1/6+2\eta} \text{ and } p_1 p_4,\, p_2 p_3 p_4 > x^{1/4-3\eta} \text{ and } p_6 > x^{1/12-5\eta}\}.\end{aligned} \tag{3-6}$$

The modified support set $\mathcal{D}^* = \mathcal{D}^*_\eta$ is understood to depend on $\eta > 0$ (as do $\mathcal{P}_4$, $\mathcal{P}_6$), but we will suppress this for notational convenience.

In this section, we establish a programmable factorization of the elements of the support $\mathcal{D}^*$ provided $D < x^{10/17}$, i.e., $\eta < \frac{1}{204}$. This will serve as the key technical input for the proof of Theorem 2.12.

Proposition 3.3 (factorization of elements of $\mathcal{D}^*$). *Let $0 < \delta < 10^{-5}$, and take $0 < \eta < \frac{1}{204} - 3\delta$ and $N \in [x^\delta, x^{1/3-\delta/2}]$. If $d \in \mathcal{D}^*$ for $D = x^{7/12+\eta-50\delta}$, then we may factor $d = d_1 d_2 d_3$ such that $d_1 \leqslant N x^{-\delta}$ and*

$$N^2 d_2 d_3^2 \leqslant x^{1-\delta}, \quad N^2 d_1 d_2^4 d_3^3 \leqslant x^{2-\delta}, \quad N d_1 d_2^5 d_3^2 \leqslant x^{2-\delta}. \tag{3-7}$$

On the first attempt working through technicalities, we encourage the reader to set $\delta = 0$ in order to better view the key features.

Before proving the proposition, we need some lemmas. The first gives a general-purpose criterion to factor an integer d.

Lemma 3.4. *Let $D = x^{7/12+\eta-50\delta}$ for $-\frac{1}{84} < \eta < \frac{1}{60}$. A factorization $d = d_1 d_2 d_3$ satisfies* (3-7), *provided $d_1, d_2, d_3 \geqslant 1$ satisfy*

$$d_2 \in [x^{1/6+2\eta}, x^{1/4-3\eta}], \quad d_1 \leqslant N x^{-\delta} \text{ and } d_3 \leqslant D/N d_2. \tag{3-8}$$

Proof. By (3-8), $N d_3 \leqslant D/d_2$ and so

$$\begin{aligned}N^2 d_2 d_3^2 &\leqslant D^2/d_2 \leqslant x^{2(7/12+\eta-50\delta)-(1/6+2\eta)} = x^{1-\delta},\\ N^2 d_1 d_2^4 d_3^3 &\leqslant D^3 d_2 \leqslant x^{3(7/12+\eta-50\delta)+(1/4-3\eta)} = x^{2-\delta},\\ N d_1 d_2^5 d_3^2 &\leqslant D^2 d_2^3 \leqslant x^{2(7/12+\eta-50\delta)+3(1/4-3\eta)} = x^{23/12-7\eta-100\delta} < x^{2-\delta},\end{aligned}$$

using $\eta \in \left(-\frac{1}{84}, \frac{1}{60}\right)$. This gives (3-7). □

The above criterion implies factorizations in the following special cases.

Lemma 3.5. *Let $D = x^{7/12+\eta-50\delta}$ for $\eta < \frac{1}{60}$. For $r \geqslant 4$, let $x^{1/6+2\eta} > p_1 \geqslant \ldots \geqslant p_r$ be primes for which $d = p_1 \cdots p_r \in \mathcal{D}^+(D)$. Suppose d_2 is one of the subproducts in $\{p_1 p_4, p_2 p_3, p_2 p_4, p_2 p_3 p_4\}$. Then d has a factorization $d = d_1 d_2 d_3$ satisfying* (3-7), *provided*

$$d_2 \in [x^{1/6+2\eta}, x^{1/4-3\eta}].$$

Proof. Let $C = D/N d_2$ and note either $p_1 \leqslant N$ or $p_1 \leqslant C$, since

$$p_1^2 \leqslant D^{2/3} = x^{2/3(7/12+\eta-50\delta)} < x^{1/3+4\eta-50\delta} \leqslant D/d_2 = NC.$$

We proceed by induction on $r \geqslant 4$. As the base case $r = 4$, by Lemma 3.4 it suffices for each b to factor $p_1 \cdots p_4/d_2 = d_1 d_3$ for $d_1 \leqslant N$, $d_3 \leqslant C$. Indeed, this holds when $d_2 = p_2 p_3 p_4$ since $p_1^2 \leqslant AC$, and similarly:

- If $d_2 = p_2 p_4$ then $p_3^2 \leqslant D/p_1 p_2 p_3 \leqslant NC/p_1$ implies $p_1 p_3 = d_1 d_3$ for some $d_1 \leqslant N$, $d_3 \leqslant C$.
- If $d_2 = p_1 p_4$ then $p_3^2 \leqslant D/p_1 p_2 p_3 \leqslant NC/p_2$ implies $p_2 p_3 = d_1 d_3$ for some $d_1 \leqslant N$, $d_3 \leqslant C$.
- If $d_2 = p_2 p_3$ then $p_4^2 \leqslant D/p_1 p_2 p_3 \leqslant NC/p_1$ implies $p_1 p_4 = d_1 d_3$ for some $d_1 \leqslant N$, $d_3 \leqslant C$.

Now for $r \geqslant 5$, we inductively assume a factorization $p_1 \cdots p_{r-1} = d_1 d_2 d_3$ with $d_1 \leqslant N$, $d_3 \leqslant C$. Then $p_r^2 \leqslant D/p_1 \cdots p_{r-1} = NC/(ac)$ so either $d_1 p_r \leqslant N$ or $d_3 p_r \leqslant C$, extending the factorization. Hence Lemma 3.4 applies again, and completes the proof. □

Finally, if the primes dividing d are small enough, we may use the greedy algorithm to factor d as follows.

Lemma 3.6. *Let $D = x^{7/12+\eta-50\delta}$ for $\eta < \frac{1}{60}$. For $r \geqslant 4$, let $x^{1/6+2\eta} > p_1 \geqslant \ldots \geqslant p_r$ be primes for which $d = p_1 \cdots p_r \in \mathcal{D}^+(D)$, and $p_6 < x^{1/12-5\eta}$ if $r \geqslant 6$. Then d has a factorization $d = abc$ satisfying* (3-7), *provided there is a factorization $p_1 p_2 p_3 p_4 = d_1 d_2 d_3$ satisfying*

$$d_1 \leqslant N x^{-\delta}, \quad d_3 \leqslant x^{1-2\delta}/DN, \quad d_2 \leqslant D^2/x^{1-3\delta} = x^{1/6+2\eta+3\delta}. \tag{3-9}$$

Proof. Let $D_1 = N x^{-\delta}$, $D_2 = D^2/x^{1-3\delta}$, $D_3 = x^{1-2\delta}/(DN)$, so that $d_i \leqslant D_i$ by assumption.

We now greedily append primes to d_i while preserving $d_i \leqslant D_i$ for all i, i.e., where at the j-th step we replace $d_i \mapsto d_i p_j$ (for one of $i = 1, 2, 3$) provided $d_i p_j \leqslant D_i$. Starting from $j = 5$, we stop either when we have exhausted all primes (i.e., $j = r$), or $d_i p_j > D_i$ for each $i = 1, 2, 3$. In the former case, we have the desired $d_1 d_2 d_3 = d = p_1 \cdots p_r$ and $d_i \leqslant D_i$ so we easily get

$$\begin{aligned} D_1 &= N x^{-\delta}, \\ N^2 D_2 D_3^2 &= x^{1-\delta}, \\ N^2 D_1 D_2^4 D_3^3 &= D^5 x^{-1+5\delta} \leqslant x^{5\cdot(3/5)-1-245\delta} < x^{2-\delta}, \\ N D_1 D_2^5 D_3^2 &= D^8 x^{-3+10\delta} \leqslant x^{8\cdot(3/5)-3-390\delta} < x^{2-\delta}. \end{aligned}$$

Thus $d_1 d_2 d_3 = d = p_1 \cdots p_r$ gives the desired factorization.

In the latter case, there exists a terminal index $j < r$ for which $d_i p_j > D_i$ for all $i = 1, 2, 3$. Note if j is odd, then $d_i p_j \leqslant D_i$ for some i, since

$$p_j^3 \leqslant \frac{D}{p_1 \cdots p_{j-1}} = \frac{D_1 D_2 D_3}{d_1 d_2 d_3}.$$

So the terminal j is even with $j \geqslant 6$. By assumption $p_j \leqslant p_6 \leqslant x^{1/12-5\eta}$ is smaller than the width of the interval $[x^{1/6+2\eta}, x^{1/4-3\eta}]$. And since $d_2 \leqslant D_2 = x^{1/6+2\eta} < d_2 p_j$, we deduce $e_2 := d_2 p_j$ lies in the interval $e_2 \in [x^{1/6+2\eta}, x^{1/4-3\eta}]$.

Thus letting $E_3 := D_2 D_3 / e_2$, for each $l > j$ in turn we shall greedily append the prime p_l onto either d_1 or d_3 while preserving $d_1 < D_1$ and $d_3 < E_3$. Indeed, for all $l > j$,

$$p_l^2 \leqslant \frac{D}{p_1 \cdots p_{l-1}} = \frac{D_1 D_2 D_3}{d_1 d_2 d_3 p_j \cdots p_{l-1}} \leqslant \frac{D_1 E_3}{d_1 d_3 p_{j+1} \cdots p_{l-1}},$$

so there is a factorization $e_1 e_3 = d_1 d_3 p_{j+1} \cdots p_l$ with $e_1 \leqslant D_1 = N$ and $e_3 \leqslant E_3 = D/(A e_2 x^{2\delta})$. Hence the result now follows by Lemma 3.4 for the factorization $e_1 e_2 e_3 = d_1 d_2 d_3 p_j \cdots p_l = p_1 \cdots p_l$. □

Proof of Proposition 3.3. We shall consider 3 cases, depending on the sizes of p_1 and $p_2 p_3$ compared to the endpoints of the key interval $[x^{1/6+2\eta}, x^{1/4-3\eta}]$.

Case 1 ($p_1 \geqslant D^2/x = x^{1/6+2\eta}$). Let $d_2 := p_1$, $C := D/N d_2$. Note $C = D/N d_2 \geqslant D^{2/3}/N \geqslant 1$.

Next $D \geqslant p_1^3 \geqslant p_1 p_2^2$ implies $p_2^2 \leqslant D/p_1 = NC$, so either $p_2 \leqslant N$ or $p_2 \leqslant C$. Similarly, since $p_1 \cdots p_{j-1} p_j^2 \leqslant D$ for all $j \leqslant r$, we get $p_j^2 \leqslant AC/(p_2 \cdots p_{j-1})$ for $3 \leqslant j \leqslant r$. As such, we may factor $p_2 \cdots p_r = d_1 d_3$ for $d_1 \leqslant N, d_3 \leqslant C$. Hence by Lemma 3.4 $p_1 \cdots p_r = d_1 d_2 d_3$ satisfies (3-7).

In the remaining cases, we assume $p_1 < x^{1/6+2\eta}$. By Lemma 3.5, it remains to consider $p_2 p_3 > x^{1/4-3\eta}$ or $p_2 p_3 < x^{1/6+2\eta}$. Note

$$p_2 p_3 \leqslant p_1^{1/3} (p_1 p_2 p_3^3)^{1/3} \leqslant (x^{1/6+2\eta})^{1/3} D^{1/3} = x^{1/3(1/6+7/12+3\eta-50\eta)} < x^{1/4+\eta-16\delta}. \tag{3-10}$$

Case 2 ($p_2 p_3 > x^{1/4-3\eta}$ and $p_1 < x^{1/6+2\eta}$). The proof follows by Lemma 3.5 if $p_2 p_4 \in [x^{1/6+2\eta}, x^{1/4-3\eta}]$. Thus by definition of $\mathcal{P}_4$, in this case we may assume

$$p_2 p_4 < x^{1/6+2\eta}. \tag{3-11}$$

Hence we have $p_4 < x^{12\eta+50\delta}$, since

$$p_2 > p_2 (p_1 p_2 p_3^3)/D > (p_2 p_3)^3/D > x^{3((1/4)-3\eta)}/D = x^{1/6-10\eta+50\delta}. \tag{3-12}$$

If $p_1 p_4 > x^{1/6+2\eta}$, then the proof follows by Lemma 3.5 where $d_2 = p_1 p_4$ is $< x^{(1/6+2\eta)+12\eta+50\delta} < x^{1/4-3\eta}$, since $\eta < \frac{1}{204} - 3\delta$.

Else $p_1 p_4 < x^{1/6+2\eta}$. We shall apply Lemma 3.6 with $d_2 = p_1 p_4$.

If either $N x^{-\delta}$ or $x^{1-2\delta}/DN$ is greater than $x^{1/4+\eta-16\delta} \geqslant p_2 p_3$, by (3-10), then Lemma 3.6 completes the proof with $(d_1, d_3) = (p_2 p_3, 1)$ or $(1, p_2 p_3)$, respectively. Otherwise, $N x^{-\delta}, x^{1-2\delta}/DN \in [x^{1/6-2\eta-64\delta}, x^{1/4+\eta-16\delta}]$, since $x/D = x^{5/12-\eta+50\delta}$. But then, using $\eta < \frac{1}{108}$,

$$\begin{aligned} \max(N x^{-\delta}, x^{1-2\delta}/DN) &\geqslant (x^{1-2\delta}/D)^{1/2} = x^{(5/2)/12-\eta/2+24\delta} > x^{1/6+2\eta} > p_2, \\ \min(N x^{-\delta}, x^{1-2\delta}/DN) &\geqslant x^{1/6-2\eta-64\delta} > x^{1/8+\eta/2-8\delta} \geqslant (p_2 p_3)^{1/2} \geqslant p_3, \end{aligned} \tag{3-13}$$

by (3-10), which suffices again for Lemma 3.6. Note $p_6 < x^{12\eta+50\delta} < x^{1/12-5\eta}$ when $r \geqslant 6$, using $\eta < \frac{1}{204} - 3\delta$.

Case 3 ($p_2 p_3 < x^{1/6+2\eta}$ and $p_1 < x^{1/6+2\eta}$). By Lemma 3.5, it suffices to consider either $p_1 p_4 < x^{1/6+2\eta}$ or $p_1 p_4 > x^{1/4-3\eta}$.

Subcase 3.1 ($p_1 p_4 < x^{1/6+2\eta}$). Suppose we can show $p_6 < x^{1/12-5\eta}$ (when $r \geqslant 6$). Then since $x^{1-3\delta}/D > D^4/x^{2-6\delta}$, either $Nx^{-\delta}$ or $x^{1-2\delta}/DN$ is greater than $D^2/x^{1-3\delta}$. Thus Lemma 3.6 will complete the proof, with $(d_1, d_2, d_3) = (p_1 p_4, p_2 p_3, 1)$ or $(1, p_2 p_3, p_1 p_4)$.

If $p_1 p_3 > x^{1/4-3\eta}$ then in this subcase

$$x^{1/12+\eta} > (p_2 p_3)^{1/2} \geqslant p_3 = p_4 \frac{p_1 p_3}{p_1 p_4} > p_4 x^{1/12-5\eta},$$

so $p_4 < x^{6\eta}$. Hence $p_6 \leqslant p_4 < x^{1/12-5\eta}$ since $\eta < \frac{1}{108}$, which completes the proof.

Else $p_1 p_3 < x^{1/4-3\eta}$. By Lemma 3.5, it suffices $p_1 p_3 < x^{1/6+2\eta}$. Then we see $p_3 > x^{1/12-5\eta}$ implies $p_1 < x^{1/12+7\eta}$. If further $p_1 p_2 > x^{1/4-3\eta}$, then similarly

$$x^{1/12+7\eta} > p_1 \geqslant p_2 = p_3 \frac{p_1 p_2}{p_1 p_3} > p_3 x^{1/12-5\eta},$$

so $p_3 < x^{12\eta}$. Hence $p_6 \leqslant p_3 < x^{1/12-5\eta}$ since $\eta < \frac{1}{204}$, which completes the proof.

Else $p_1 p_2 < x^{1/4-3\eta}$. By Lemma 3.5, we may assume $p_1 p_2 < x^{1/6+2\eta}$.

Similarly, suppose $p_2 p_3 p_4 < x^{1/4-3\eta}$. By Lemma 3.5 we may assume $p_2 p_3 p_4 < x^{1/6+2\eta}$, and so

$$p_6 \leqslant (p_2 p_3 p_4)^{1/3} \leqslant x^{(2/3)/12+(2/3)\eta} \leqslant x^{1/12-5\eta},$$

using $\eta < \frac{1}{204}$, which completes the proof.

Thus we may assume $p_2 p_3 p_4 > x^{1/4-3\eta}$. But unless $p_6 < x^{1/12-5\eta}$, this subcase will contradict the definition of $\mathcal{P}_{6,1}$ in (3-6), hence completing the proof.

Subcase 3.2 ($p_1 p_4 > x^{1/4-3\eta}$). If $d_2 = p_2 p_3 p_4 < x^{1/6+2\eta}$, then Lemma 3.6 completes the proof with $(d_1, d_3) = (p_1, 1)$ or $(1, p_1)$, since

$$p_6 \leqslant (p_2 p_3 p_4)^{1/3} \leqslant x^{(1/3)(1/6+2\eta)} \leqslant x^{1/12-5\eta}$$

for $\eta < \frac{1}{204}$. And if $p_2 p_3 p_4 \in [x^{1/6+2\eta}, x^{1/4-3\eta}]$ the proof follows by Lemma 3.5.

Else $p_2 p_3 p_4 > x^{1/4-3\eta}$. Note $p_4 < x^{1/12+\eta}$ and $p_1 = p_1 p_4/p_4 > x^{1/6-4\eta}$ and $p_2 p_3 p_4 < x^{1/4+3\eta}$. Also note we may factor $p_1 p_4 = d_1 d_3$ for $d_1 \leqslant N$, $d_3 \leqslant x^{1-2\delta}/DN$ (Indeed this follows if N or $x^{1-2\delta}/DN$ exceeds $x^{1/4+3\eta} \geqslant p_1 p_4$. Else $N, x^{1-2\delta}/DN \in [x^{1/6-4\eta-2\delta}, x^{1/4+3\eta}]$, which also works similarly as with (3-13), since $p_4 < x^{1/12+\eta} < x^{1/6-4\eta-2\delta}$ and $p_1 < x^{1/6+2\eta} < x^{1/2(5/12-\eta-\delta)}$ by $\eta < \frac{1}{60}$.)

If further $p_6 > x^{1/12-5\eta}$, then this subcase contradicts the definition of $\mathcal{P}_{6,2}$ in (3-6). Hence we have $p_6 \leqslant x^{1/12-5\eta}$, and so by the above paragraph Lemma 3.6 completes the proof with $d_2 = p_2 p_3$.

Combining all cases completes the proof of Proposition 3.3. □

3.1. ***Refined factorization of $\mathcal{D}^{\text{well}}$.*** Proposition 3.3 (programmably) factorizes each $d \in \mathcal{D}^* \subset \mathcal{D}^+(x^{7/12+\eta})$, and forms the key step to prove the weights λ^* are programmably factorable. With applications in mind to twin primes, we shall similarly (programmably) factorize certain subsets of the well-factorable support $\mathcal{D}^{\text{well}}$, as in (3-2).

In the following result, we factorize $d \in \mathcal{D}^{\text{well}}(D)$ for variable level $D \in (x^{4/7}, x^{3/5})$, depending on the anatomy of d. As $\mathcal{D}^{\pm} \subset \mathcal{D}^{\text{well}}$, this has implications to both upper and lower bounds for the standard linear sieve.

Proposition 3.7. *Let $\mathcal{D}^{\text{well}}(D)$ as in* (3-2) *for $D = x^{7/12+\eta-50\delta}$ and $-\frac{1}{84} < \eta < \frac{1}{60} - 30\delta$. Let $x^{1/4-3\eta} \geqslant p_1 \geqslant \ldots \geqslant p_r$ be primes for which $d = p_1 \cdots p_r \in \mathcal{D}^{\text{well}}(D)$. Then d has factorization $d = abc$ satisfying* (3-7) *if $p_3 \leqslant x^{1/12-5\eta}$, or if*

$$d_2 \in [x^{1/6+2\eta}, x^{1/4-3\eta}] \quad \textit{with } d_2 \mid p_1 p_2 p_3,\ d_2 \neq p_3.$$

Proof. For $i = 1, 2, 3$, suppose $d_2 = p_1 \cdots p_i$ lies $[x^{1/6+2\eta}, x^{1/4-3\eta}]$, and let $A = Nx^{-\delta}$, $C = x^{\delta} D/Nd_2$. Since $p_1 \cdots p_{j-1} p_j^2 \leqslant D$ for all $i < j \leqslant r$, we get $p_j^2 \leqslant AC/(p_{i+1} \cdots p_{j-1})$ for $i < j \leqslant r$. As such, we may factor $p_{i+1} \cdots p_r = d_1 d_3$ for $d_1 \leqslant A$, $d_3 \leqslant C$. Hence by Lemma 3.4 $p_1 \cdots p_r = d_1 d_2 d_3$ satisfies (3-7).

Else, by assumption $p_1 < x^{1/(4)-3\eta}$ so we may assume further $p_1 < x^{1/(6)+2\eta}$. In particular this gives $p_1^2 \leqslant D/d_2$. For the remaining $d_2 \mid p_1 p_2 p_3$:

- If $d_2 = p_2 p_3$ then $p_1^2 \leqslant D/d_2 = AC$ implies $p_1 \leqslant A$ or $p_1 \leqslant C$.
- If $d_2 = p_1 p_3$ then $p_2^2 \leqslant D/d_2 = AC$ implies $p_2 \leqslant A$ or $p_2 \leqslant C$.
- If $d_2 = p_2$ then $p_3^2 \leqslant D/p_1 d_2$ implies a factorization $p_1 p_3 = d_1 d_3$ for $d_1 \leqslant A$, $d_3 \leqslant C$.

For each d_2 above, we factored $p_1 p_2 p_3 = d_1 d_2 d_3$ for $d_1 \leqslant A$, $d_3 \leqslant C$. Since $p_1 \cdots p_{j-1} p_j^2 \leqslant D$ for all $j \leqslant r$, by induction we may factor $p_1 \cdots p_r = d_1 d_2 d_3$ for $d_1 \leqslant A$, $d_3 \leqslant C$. By Lemma 3.4 $p_1 \cdots p_r = d_1 d_2 d_3$ satisfies (3-7).

Finally, suppose $p_3 \leqslant x^{1/12-5\eta}$ is less than the width of the interval $[x^{1/6+2\eta}, x^{1/4-3\eta}]$. Since $p_1 < x^{1/6+2\eta}$, we have $p_1 p_3 < x^{1/4-3\eta}$ so by the above argument we may assume $d_2 := p_1 p_3 < x^{1/6+2\eta}$. Then $p_2^3 \leqslant p_1 p_2^2 \leqslant D$ implies

$$p_2^2 \leqslant x^{(2/3)(7/12+\eta)} < x^{5/12-\eta+47\delta} = \frac{x^{1-3\delta}}{D},$$

since $\eta < \frac{1}{60} - 30\delta$. Thus $p_2 \leqslant Nx^{-\delta}$ or $p_2 \leqslant x^{1-2\delta}/DN$, so there is a factorization $p_1 p_2 p_3 = d_1 d_2 d_3$ satisfying (3-9). Hence the same greedy argument as in Lemma 3.6 completes the proof, with p_3 playing the role of p_6. □

Taking the maximum valid η as above, we may reexpress the above factorization of level x^{θ}, $\theta = \frac{7}{12} + \eta$, as follows. Note the maximum θ for which $t \in \left[\frac{1}{6} + 2\eta, \frac{1}{4} - 3\eta\right]$ is given by

$$\theta(t) = \begin{cases} \frac{2-t}{3} & \text{if } t > \frac{1}{5}, \\ \frac{1+t}{2} & \text{if } t \leqslant \frac{1}{5}. \end{cases} \tag{3-14}$$

Similarly the maximum $\theta = \frac{7}{12} + \eta$ for which $t \leqslant \frac{1}{12} - 5\eta$ is $\frac{1}{5}(3 - t)$.

Corollary 3.8. *Let $p_1 \geqslant \ldots \geqslant p_r$ be primes and write $p_i = x^{t_i}$. If $d = p_1 \cdots p_r \in \mathcal{D}^{\mathrm{well}}(x^{\theta - 50\delta})$, then there is a factorization $d = d_1 d_2 d_3$ satisfying* (3-7) *provided*

$$\theta \leqslant \theta(t_1),$$

for $\theta(t)$ as in (3-14). *Moreover if $t_1 \leqslant \frac{1}{5}$, then it suffices that*

$$\theta \leqslant \theta(t_1, t_2, t_3) := \max\left\{\frac{3 - t_3}{5}, \theta(t_1), \theta(t_2), \theta(t_1 + t_2 + t_3), \theta(t_1 + t_2), \theta(t_1 + t_3), \theta(t_2 + t_3)\right\}. \quad (3\text{-}15)$$

4. Modification of the linear sieve

In this section we shall bound the modified linear sieve, analogous to the bounds for the linear sieve (sometimes called the Jurkat–Richert theorem). This bound will form the basis for our final result in the next section, in which we modify the construction of Iwaniec's weights.

Proposition 4.1. *Let $\varepsilon > 0$ be sufficiently small. For $\eta \leqslant \frac{1}{204}$, the modified weights λ^* as in* (3-4) *of level $D = x^{7/12 + \eta - \varepsilon}$ satisfy*

$$S(\mathcal{A}, z) \leqslant |\mathcal{A}| V(z) \left(F^*\left(\frac{\log D}{\log z}\right) + o(1)\right) + \sum_{d \mid P(z)} \lambda^*(d) r_{\mathcal{A}}(d),$$

where $F^* = F^*_\eta$ is a function satisfying $F^*(s) = F(s) + O(\eta^5)$ for F as in (2-5).

Remark 4.2. It suffices for our purposes to obtain qualitative error $o(1)$ in the factor accompanying F^*. Though as with the Jurkat–Richert theorem, with greater care one should obtain a quantitative refinement, e.g., $O((\log D)^{-1/6})$; see (12.4)–(12.8) in [Friedlander and Iwaniec 2010].

We now adapt the proof. Let $D = x^{7/12+\eta}$ and $D_0 = x^{7/12}$. For $n \geqslant 1$, primes $p_1 \geqslant \ldots \geqslant p_n$, if $p_1 \cdots p_n \notin \mathcal{D}^+(D)$ then there exists a minimal index $l \leqslant n$ such that $p_1 \cdots p_l \notin \mathcal{D}^+(D)$. By definition such minimal l is odd. (Explicitly, this occurs when $p_1 \cdots p_{l-1} p_l^3 > D$ but $p_1 \cdots p_{m-1} p_m^3 \leqslant D$ for all odd $m < l$.) Similarly, if $p_1 \cdots p_n \notin \mathcal{D}^*$ there exists a minimal index $l \leqslant n$ such that $p_1 \cdots p_l \notin \mathcal{D}^*$, which is also odd.

Indeed, to show this let $l \leqslant n$ be the minimal index such that $p_1 \cdots p_l \notin \mathcal{D}^*$. If $(p_1, \ldots, p_j) \notin \mathcal{P}_j$ for all $j \leqslant l$, $j \in \{4, 6\}$, then clearly $l > j$ must be odd, as with $\mathcal{D}^+(D)$. On the other hand, if $(p_1, \ldots, p_j) \in \mathcal{P}_j$ for some $j \leqslant l$, $j \in \{4, 6\}$, a priori one might expect l could be even. However, the key point in this case is that $p_1 \cdots p_j \in \mathcal{D}^+(D_0) \subset \mathcal{D}^*$ (since $p_1 \cdots p_j \approx D^{\frac{6}{7}}$ by definition of $\mathcal{P}_j$). Thus $l > j$ is the minimal index such that $p_1 \cdots p_l \notin \mathcal{D}^+(D_0)$, and hence must be odd as claimed.

Using this minimal index, we show the following lemma.

Lemma 4.3. *Let h be a multiplicative function with $0 \leqslant h(p) \leqslant 1$ for all primes p. Then we have*

$$\prod_{p \mid n} (1 - h(p)) \leqslant \sum_{d \mid n} \lambda^*(d) h(d).$$

Proof. Note if $h = 1$ identically, we interpret the product as $\mathbf{1}_{n=1}$. Now by definition,

$$\sum_{d\,|\,n} \lambda^*(d)h(d) - \prod_{p\,|\,n}(1-h(p)) = \sum_{\substack{d\,|\,n\\ d\in\mathcal{D}^*}} \mu(d)h(d) - \sum_{d\,|\,n} \mu(d)h(d) = -\sum_{\substack{d\,|\,n\\ d\notin\mathcal{D}^*}} \mu(d)h(d).$$

Then splitting up $d \notin \mathcal{D}^*$ by its minimal index,

$$-\sum_{\substack{d\,|\,n\\ d\notin\mathcal{D}^*}} \mu(d)h(d) = \sum_{\text{odd } l} \sum_{\substack{p_l<\cdots<p_1<z\\ p_1\cdots p_{l-1}\in\mathcal{D}^*\\ p_1\cdots p_l\notin\mathcal{D}^*}} h(p_1\cdots p_l) \sum_{\substack{p_1\cdots p_l b\,|\,n\\ b\,|\,P(p_l)}} \mu(b)h(b) \geqslant 0,$$

since $h \geqslant 0$ and the inner sum over b factors as $\prod_{p\,|\,(P(p_l),n)}(1-h(p)) \geqslant 0$, since $h(p) \leqslant 1$. □

By Lemma 4.3 with $h(d) = 1$, we have

$$\mathbf{1}_{n=1} \leqslant \sum_{d\,|\,n} \lambda^*(d), \tag{4-1}$$

in which case we obtain

$$\begin{aligned} S(\mathcal{A},\mathcal{P},z) &= \sum_{n\in A} \mathbf{1}_{(n,P(z))=1} \leqslant \sum_{n\in A}\sum_{d\,|\,(n,P(z))} \lambda^*(d) = \sum_{d\,|\,P(z)} \lambda^*(d)|\mathcal{A}_d| \\ &= X\sum_{d\,|\,P(z)} \lambda^*(d)g(d) + \sum_{d\,|\,P(z)} \lambda^*(d)r_{\mathcal{A}}(d) =: XV^*(D,z) + R^*_{\mathcal{A}}(D,z). \end{aligned} \tag{4-2}$$

Following Lemma 4.3 with $h = g$, we have the identity

$$V^*(D,z) := \sum_{\substack{d\,|\,P(z)\\ d\in\mathcal{D}^*}} \mu(d)g(d) = V(z) + \sum_{\text{odd } n} \sum_{\substack{p_n<\cdots<p_1<z\\ p_1\cdots p_{n-1}\in\mathcal{D}^*\\ p_1\cdots p_n\notin\mathcal{D}^*}} g(p_1\cdots p_n)V(p_n), \tag{4-3}$$

and similarly

$$V^+(D,z) = V(z) + \sum_{\text{odd } n} \sum_{\substack{p_n<\cdots<p_1<z\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)\\ p_1\cdots p_n\notin\mathcal{D}^+(D)}} g(p_1\cdots p_n)V(p_n) =: V(z) + \sum_{\text{odd } n} V_n(D,z). \tag{4-4}$$

Then the difference of V^* and V^+ is

$$V^*(D,z) - V^+(D,z) = \sum_{\text{odd } n} \sum_{p_n<\cdots<p_1<z} g(p_1\cdots p_n)V(p_n)\mathbf{\Delta}, \tag{4-5}$$

where $\mathbf{\Delta}$ is the difference of indicator functions,

$$\mathbf{\Delta} := \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^*\\ p_1\cdots p_{n-1}\in\mathcal{D}^*}} - \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)}} = \mathbf{1}_{\substack{p_1\cdots p_n\in\mathcal{D}^+(D)\setminus\mathcal{D}^*\\ p_1\cdots p_{n-1}\in\mathcal{D}^*}} - \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)\setminus\mathcal{D}^*}},$$

recalling $\mathcal{D}^* \subset \mathcal{D}^+(D)$. Note if a point is $(p_1, \dots, p_6) \in \mathcal{P}_6$ then its projection is $(p_1, \dots, p_4) \notin \mathcal{P}_4$. So by definitions of $\mathcal{D}^*$, $\mathcal{D}^+(D)$ from (3-5), (3-1), for odd n we have the identities

$$\mathbf{1}_{\substack{p_1\cdots p_n\in\mathcal{D}^+(D)\setminus\mathcal{D}^*\\ p_1\cdots p_{n-1}\in\mathcal{D}^*}} = \sum_{\substack{j\in\{4,6\}\\ j<n}} \mathbf{1}_{(p_1,\dots,p_j)\in\mathcal{P}_j}\cdot \mathbf{1}_{\substack{p_1\cdots p_n\in\mathcal{D}^+(D)\setminus\mathcal{D}^+(D_0)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D_0)}}, \tag{4-6}$$

$$\mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)\setminus\mathcal{D}^*}} = \sum_{\substack{j\in\{4,6\}\\ j<n}} \mathbf{1}_{(p_1,\dots,p_j)\in\mathcal{P}_j}\cdot \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)\setminus\mathcal{D}^+(D_0)}}. \tag{4-7}$$

We may plug (4-6) and (4-7) into $\boldsymbol{\Delta}$. In addition, we strategically add and subtract the indicator function of $\{p_1 \cdots p_n \notin \mathcal{D}^+(D),\ p_1 \cdots p_{n-1} \in \mathcal{D}^+(D_0)\}$, which together give

$$\begin{aligned}
\boldsymbol{\Delta} &= \sum_{\substack{j\in\{4,6\}\\ j<n}} \mathbf{1}_{(p_1,\dots,p_j)\in\mathcal{P}_j}\cdot \Big(\mathbf{1}_{\substack{p_1\cdots p_n\in\mathcal{D}^+(D)\setminus\mathcal{D}^+(D_0)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D_0)}} - \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)\setminus\mathcal{D}^+(D_0)}}\Big)\\
&= \sum_{\substack{j\in\{4,6\}\\ j<n}} \mathbf{1}_{(p_1,\dots,p_j)\in\mathcal{P}_j}\cdot \Big(\mathbf{1}_{\substack{p_1\cdots p_n\in\mathcal{D}^+(D)\setminus\mathcal{D}^+(D_0)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D_0)}} + \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D_0)}}\\
&\qquad\qquad - \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)\setminus\mathcal{D}^+(D_0)}} - \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D_0)}}\Big)\\
&= \sum_{\substack{j\in\{4,6\}\\ j<n}} \mathbf{1}_{(p_1,\dots,p_j)\in\mathcal{P}_j}\cdot \Big(\mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D_0)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D_0)}} - \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)}}\Big).
\end{aligned}$$

Thus plugging $\boldsymbol{\Delta}$ back into (4-5) gives

$$\begin{aligned}
&V^*(D,z) - V^+(D,z)\\
&\quad = \sum_{j\in\{4,6\}} \sum_{\substack{p_j<\cdots<p_1<z\\ (p_1,\dots,p_j)\in\mathcal{P}_j}} \sum_{\text{odd } n>j} \sum_{p_n<\cdots<p_{j+1}<p_j} g(p_1\cdots p_n)V(p_n)\Big(\mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D_0)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D_0)}} - \mathbf{1}_{\substack{p_1\cdots p_n\notin\mathcal{D}^+(D)\\ p_1\cdots p_{n-1}\in\mathcal{D}^+(D)}}\Big).
\end{aligned}$$

Recalling the definition of $V_n(D, z)$ in (4-4), since g is multiplicative we have

$$\begin{aligned}
&V^*(D,z) - V^+(D,z)\\
&\quad = \sum_{j\in\{4,6\}} \sum_{\substack{p_j<\cdots<p_1<z\\ (p_1,\dots,p_j)\in\mathcal{P}_j}} g(p_1\cdots p_j) \sum_{\text{odd } n>j} \bigg(V_{n-j}\Big(\frac{D_0}{p_1\cdots p_j}, p_j\Big) - V_{n-j}\Big(\frac{D}{p_1\cdots p_j}, p_j\Big)\bigg)\\
&\quad = \sum_{j\in\{4,6\}} \sum_{\substack{p_j<\cdots<p_1<z\\ (p_1,\dots,p_j)\in\mathcal{P}_j}} g(p_1\cdots p_j)\bigg(V^+\Big(\frac{D_0}{p_1\cdots p_j}, p_j\Big) - V^+\Big(\frac{D}{p_1\cdots p_j}, p_j\Big)\bigg).
\end{aligned} \tag{4-8}$$

as $V^+(D, z) - V^+(D', z) = \sum_{\text{odd } n}[V_n(D, z) - V_n(D', z)]$.

Now assuming the two-sided condition (2-3) for g, the proof of [Friedlander and Iwaniec 2010, Theorem 11.12] (see (12.4)–(12.8)) gives asymptotic equality,

$$V^+(D,z) = V(z)\bigg\{F\Big(\frac{\log D}{\log z}\Big) + O((\log D)^{-1/6})\bigg\} \qquad (z \leqslant D), \tag{4-9}$$

so that (4-8) becomes

$$\begin{aligned}V^*(D,z)=V(z)&\Big\{F\Big(\frac{\log D}{\log z}\Big)+O((\log D)^{-1/6})\Big\}\\&+\sum_{j\in\{4,6\}}\sum_{\substack{p_j<\cdots<p_1<z\\(p_1,\ldots,p_j)\in\mathcal{P}_j}}g(p_1\cdots p_j)V(p_j)\\&\times\Big\{F\Big(\frac{\log D_0/p_1\cdots p_j}{\log p_j}\Big)-F\Big(\frac{\log D/p_1\cdots p_j}{\log p_j}\Big)+O\Big(\log\Big(\frac{D}{p_1\cdots p_j}\Big)^{-1/6}\Big)\Big\}.\end{aligned}\tag{4-10}$$

By partial summation and the prime number theorem, for each j we have,

$$\begin{aligned}&\sum_{\substack{p_j<\cdots<p_1<z\\(p_1,\ldots,p_j)\in\mathcal{P}_j}}g(p_1\cdots p_j)V(p_j)F\Big(\frac{\log D_0/p_1\cdots p_j}{\log p_j}\Big)\\&\qquad=\big(\tfrac{7}{12}+\eta\big)\int_{(x_1,\ldots,x_j)\in P_j}\frac{\mathrm{d}x_1\cdots\mathrm{d}x_j}{x_1\cdots x_{j-1}x_j^2}F\Big(\frac{7/12-x_1-\cdots x_j}{x_j}\Big)+O((\log D)^{-1/6}).\end{aligned}$$

Here P_j is the polytope in Euclidean space $\mathbb{R}^j$ corresponding to $\mathcal{P}_j$, as below.

Hence from (4-10), we obtain

$$V^*(D,z)=V(z)\Big\{F^*\Big(\frac{\log D}{\log z}\Big)+O((\log D)^{-1/6})\Big\}\qquad(z\leqslant D),\tag{4-11}$$

where the function F^* satisfies

$$\begin{aligned}&sF^*(s)-sF(s)\\&=\big(\tfrac{7}{12}+\eta\big)\cdot\sum_{j\in\{4,6\}}\int_{(x_1,\ldots,x_j)\in P_j}\frac{\mathrm{d}x_1\cdots\mathrm{d}x_j}{x_1\cdots x_{j-1}x_j^2}\Big[F\Big(\frac{7/12-x_1-\cdots x_j}{x_j}\Big)-F\Big(\frac{7/12+\eta-x_1-\cdots x_j}{x_j}\Big)\Big].\end{aligned}\tag{4-12}$$

Namely, $P_4\subset\mathbb{R}^4$ is given by

$$P_4=\big\{(x_1,\ldots,x_4)\in\mathrm{D}^+\big(\tfrac{7}{12}+\eta\big):x_1<\tfrac{1}{6}+2\eta\text{ and }x_2+x_4>\tfrac{1}{4}-3\eta\big\},$$

and $P_6=P_{6,1}\cup P_{6,2}\subset\mathbb{R}^6$ is given by

$$\begin{aligned}P_{6,1}&=\big\{(x_1,\ldots,x_6)\in\mathrm{D}^+\big(\tfrac{7}{12}+\eta\big):x_1+x_2<\tfrac{1}{6}+2\eta\text{ and }x_6>\tfrac{1}{12}-5\eta\text{ and }x_2+x_3+x_4>\tfrac{1}{4}-3\eta\big\},\\P_{6,2}&=\big\{(x_1,\ldots,x_6)\in\mathrm{D}^+\big(\tfrac{7}{12}+\eta\big):x_1,x_2+x_3<\tfrac{1}{6}+2\eta\\&\qquad\qquad\text{and }x_6>\tfrac{1}{12}-5\eta\text{ and }x_1+x_4,x_2+x_3+x_4>\tfrac{1}{4}-3\eta\big\}.\end{aligned}\tag{4-13}$$

Similarly, D^+ is the set in Euclidean space corresponding to $\mathcal{D}^+$, namely,

$$\mathrm{D}^+(\tau)=\{(x_1,\ldots,x_r):x_1>\cdots>x_r>0\text{ and }x_1+\cdots x_{l-1}+3x_l<\tau\text{ for each odd }1\leqslant l\leqslant r\}.$$

Hence Proposition 4.1 follows.

4.1. ***Sieve function computation.*** We now compute F^* in terms of F.

Proposition 4.4. *Let* $\eta = \frac{1}{204}$. *Then for* $1 \leqslant s \leqslant 3$, *we have*

$$F^*(s) \leqslant 1.000081 F(s). \tag{4-14}$$

Proof. From (4-12) we have

$$sF^*(s) = sF(s) + \left(\tfrac{7}{12} + \eta\right) \cdot 2e^{\gamma}\eta(J_4 + J_6), \tag{4-15}$$

for integrals J_j, $j \in \{4, 6\}$,

$$\begin{aligned} J_j &:= \frac{1}{2e^{\gamma}\eta} \int_{(x_1,\dots,x_j)\in P_j} \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_j}{x_1 \cdots x_{j-1}x_j^2} \left[F\left(\frac{7/12 - x_1 - \cdots x_j}{x_j}\right) - F\left(\frac{7/12 + \eta - x_1 - \cdots x_j}{x_j}\right)\right] \\ &= \int_{(x_1,\dots,x_j)\in P_j} \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_j}{x_1 \cdots x_j} \left[\left(\tfrac{7}{12} - x_1 - \cdots - x_j\right)\left(\tfrac{7}{12} + \eta - x_1 - \cdots - x_j\right)\right]^{-1}, \end{aligned}$$

since $sF(s) = 2e^{\gamma}$ for $s \in [1, 3]$. In particular $|P_j| = O(\eta^j)$ implies $J_j = O(\eta^j)$, and so from (4-15) we obtain $F^*(s) = F(s) + O(\eta^5)$.

For $\eta = \frac{1}{204}$, we use Mathematica[4] to compute that

$$J_4 \leqslant 0.016896. \tag{4-16}$$

Next we bound J_6. For $(x_1, \dots, x_6) \in P_6$ we have $x_4 < \frac{1}{2}(x_2+x_3) < \frac{1}{12}+\eta$ and $\frac{7}{12}+\eta-x_1-\cdots-x_6 > x_5$ so

$$J_6 \leqslant \int_{\overline{P_6}} \frac{\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3}{x_1x_2x_3} \int_{1/12-5\eta<x_6<x_5<x_4<1/12+\eta} \frac{\mathrm{d}x_4\,\mathrm{d}x_5\,\mathrm{d}x_6}{x_4x_5^2x_6(x_5-\eta)},$$

where $\overline{P_6} = \overline{P_{6,1}} \cup \overline{P_{6,2}}$ is the (3-dimensional) projection of P_6, given explicitly by

$$\begin{aligned} \overline{P_{6,1}} &= \left\{(x_1, x_2, x_3) \in \mathrm{D}^+\left(\tfrac{7}{12}+\eta\right) : x_1+x_2 < \tfrac{1}{6}+2\eta \text{ and } x_3 > \tfrac{1}{12}-5\eta \text{ and } x_2+2x_3 > \tfrac{1}{4}-3\eta\right\}, \\ \overline{P_{6,2}} &= \left\{(x_1, x_2, x_3) \in \mathrm{D}^+\left(\tfrac{7}{12}+\eta\right) : x_1, x_2+x_3 < \tfrac{1}{6}+2\eta \text{ and } x_3 > \tfrac{1}{12}-5\eta \text{ and } x_1+x_3, x_2+2x_3 > \tfrac{1}{4}-3\eta\right\}. \end{aligned}$$

For $\eta = \frac{1}{204}$, we compute $J_6 \leqslant (J_{6,1} + J_{6,2})J_{6,0}$ where

$$\begin{aligned} J_{6,0} &= \int_{1/12-5\eta<x_6<x_5<x_4<1/12+\eta} \frac{\mathrm{d}x_4\,\mathrm{d}x_5\,\mathrm{d}x_6}{x_4x_5^2x_6(x_5-\eta)} \leqslant 2.33838, \\ J_{6,1} &= \int_{\overline{P_{6,1}}} \frac{\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3}{x_1x_2x_3} \leqslant 0.000806853, \\ J_{6,2} &= \int_{\overline{P_{6,2}}} \frac{\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3}{x_1x_2x_3} \leqslant 0.00397946. \end{aligned}$$

Hence combining with (4-16), for $s \in [1, 3]$ we conclude

$$\frac{F^*(s)}{F(s)} \leqslant 1 + \left(\tfrac{7}{12} + \eta\right) \cdot \eta(J_4 + (J_{6,1} + J_{6,2})J_{6,0}) \leqslant 1.000081. \qquad \square$$

[4] The Mathematica package and code are available at `arxiv.org/abs/2109.02851`.

5. Factorable remainder, after Iwaniec

In Theorem 2.10, Iwaniec constructed a well-factorable variant $\tilde{\lambda}^{\pm}$ of the weights $\lambda^{\pm}$ from the (Jurkat–Richert) linear sieve. In this section, we prove Theorem 2.12 for the programmably factorable variant $\tilde{\lambda}^{*}$ by adapting Iwaniec's construction, similarly building on the Jurkat–Richert type Proposition 4.1 that we obtained in the previous section. We shall also prove a technical variation on this result, with a variable level depending on anatomy of the moduli.

To set up the construction, we first adapt [Friedlander and Iwaniec 2010, Proposition 12.18]. Denote $P(z,u) = P(z)/P(u) = \prod_{u<p\leqslant z} p$.

Proposition 5.1. *Let* $\eta > 0$, *and* $D = x^{(7/12+\eta)/(1+\varepsilon+\tau)}$ *for* $\varepsilon > 0$ *sufficiently small. Let* $\boldsymbol{D}_r^*$ *be defined by* (5-8). *Let* $\lambda^{(r)}$ *be the standard* (*upper and lower*, *for r odd and even*, *resp.*) *weights for the linear sieve of level* D^{ε}. *Then for* $u = D^{\varepsilon^2}$, $\tau = \varepsilon^9$,

$$S(\mathcal{A}, z) \leqslant |\mathcal{A}| V(z)\{F^*(s) + O(\varepsilon^5)\} + \sum_{0\leqslant r\leqslant\varepsilon^{-2}} \sum_{(D_1,\dots,D_r)\in\boldsymbol{D}_r^*} \frac{(-1)^r}{\gamma(D_1,\dots,D_r)} \sum_{\substack{p_1\cdots p_r \mid P(z,u)\\ D_j<p_j\leqslant D_j^{1+\tau}}} \sum_{\substack{b\mid P(u)\\ b\leqslant D^{\varepsilon}}} \lambda^{(r)}(b) r_{\mathcal{A}}(bp_1\cdots p_r). \tag{5-1}$$

Proof. First we write

$$S(\mathcal{A}, z) \leqslant S^*(\mathcal{A}, z) - \sum_{\text{odd } n\leqslant N} S_n(\mathcal{A}, z) \tag{5-2}$$

for any $N \geqslant 1$, where

$$S^*(\mathcal{A}, z) := \sum_{0\leqslant r\leqslant\varepsilon^{-2}} (-1)^r \sum_{\substack{u\leqslant p_r<\dots p_1<z\\ p_1\cdots p_r\in\mathcal{D}^*}} |\mathcal{A}_{p_1\cdots p_r}|, \quad S_n(\mathcal{A}, z) := \sum_{\substack{p_n<\cdots<p_1<z\\ p_1\cdots p_{n-1}\in\mathcal{D}^*\\ p_1\cdots p_n\notin\mathcal{D}^*}} S(\mathcal{A}_{p_1\cdots p_n}, p_n).$$

We apply the inequality (5-2), not for $\mathcal{A} = (a_n)$ itself but rather for the subsequence $\tilde{\mathcal{A}} = (a_n \mathbf{1}_{(n,P(u))=1})$. Here we take $u = D^{\varepsilon^2}$, and then return to $\mathcal{A}$ by means of the fundamental lemma.

Let $z = D^{1/s}$ with $2 \leqslant s \leqslant \varepsilon^{-1}$. Since $z > u$, the only change to the above bound (5-2) when passing to $\tilde{\mathcal{A}}$ is the term $S^*(\tilde{\mathcal{A}}, z)$, provided that N is not too large in terms of ε. Specifically, we require the lower bound for p_n (by induction, the linear sieve conditions imply $p_1 \cdots p_m < D^{1-2^{-m}}$)

$$p_n \geqslant (D/p_1\cdots p_{n-1})^{1/3} \geqslant D^{2^{1-N}/3} \tag{5-3}$$

to be larger than $u = D^{\varepsilon^2}$, which certainly holds provided

$$N \leqslant \frac{1}{2}\log\frac{1}{\varepsilon}. \tag{5-4}$$

Now it remains to evaluate $S^*(\tilde{\mathcal{A}}, z)$,

$$S^*(\tilde{\mathcal{A}}, z) = \sum_{0\leqslant r\leqslant\varepsilon^{-2}} (-1)^r \sum_{\substack{u\leqslant p_r<\dots p_1<z\\ p_1\cdots p_r\in\mathcal{D}^*}} |\tilde{\mathcal{A}}_{p_1\cdots p_r}|. \tag{5-5}$$

For each r, we break the range in the inner sum into boxes. Namely, we let $D_1, \dots, D_r$ run over numbers of form

$$D^{\varepsilon^2(1+\tau)^j}, \quad j = 0, 1, 2, \dots, \tag{5-6}$$

with $\tau = \varepsilon^9$. We denote by $\boldsymbol{D}_r^+ = \boldsymbol{D}_r^+(D)$ the set of r-tuples $(D_1, \dots, D_r)$ with $D_r \leqslant \dots \leqslant D_1 \leqslant \sqrt{D}$ such that

$$\boldsymbol{D}_r^+ = \begin{cases} \{(D_1, \dots, D_r) : D_1 \cdots D_{m-1} D_m^3 < D \text{ for all odd } m \leqslant r\} & \text{if } r \text{ even,} \\ \{(D_1, \dots, D_r) : D_1 \cdots D_{m-1} D_m^3 < D^{1/(1+\tau)} \text{ for all odd } m \leqslant r\} & \text{if } r \text{ odd.} \end{cases}$$

We note, for $\varepsilon > 0$ sufficiently small, the cardinalities of the $\boldsymbol{D}_r^+$ are bounded by

$$\sum_{0 \leqslant r \leqslant \varepsilon^{-2}} |\boldsymbol{D}_r^+| \leqslant \exp(\varepsilon^{-3}). \tag{5-7}$$

Hereafter let $D = x^{(7/12+\eta)/(1+\tau+\varepsilon)}$, and define

$$\boldsymbol{D}_r^* = \{(D_1, \dots, D_r) \in \boldsymbol{D}_r^+(D) : (D_1, \dots, D_i) \notin \boldsymbol{P}_{i,r} \text{ for } i \leqslant r,\ i \in \{4, 6\}\}. \tag{5-8}$$

where $\boldsymbol{P}_{4,r}$, $\boldsymbol{P}_{6,r}$ are (τ-enlarged, for even r) analogues of the polytopes $\mathcal{P}_4$, $\mathcal{P}_6$ in (3-6), e.g.,

$$\boldsymbol{P}_{4,r} = \begin{cases} \{(D_1, \dots, D_4) : D_1^{1+\tau} < x^{1/6+2\eta} \text{ and } (D_2 D_4)^{1/(1+\tau)} > x^{1/4-3\eta}\} & \text{if } r \text{ even,} \\ \{(D_1, \dots, D_4) : D_1 < x^{1/6+2\eta} \text{ and } D_2 D_4 > x^{1/4-3\eta}\} & \text{if } r \text{ odd.} \end{cases}$$

Observe each integer $p_1 \cdots p_r$ has a unique vector $(D_1, \dots, D_r)$ such that $p_i \in (D_i, D_i^{1+\tau}]$ for all $i \leqslant r$, inducing a map $\nu : \mathbb{N} \to \bigcup_r \boldsymbol{D}_r^+$. As a convention $\nu(1) = (\,)$ is the empty vector. By construction, for even r if $p_1 \cdots p_4 \notin \mathcal{P}_4$ then $\nu(p_1 \cdots p_4) \notin \boldsymbol{P}_{4,r}$, and if $(D_1, \dots, D_4) \notin \boldsymbol{P}_{4,r}$ then $\nu^{-1}(D_1, \dots, D_4) \cap \mathcal{P}_4 = \varnothing$ for odd r. Continuing this argument, we deduce

$$\begin{aligned} p_1 \cdots p_r \in \mathcal{D}^* &\implies \nu(p_1 \cdots p_r) \in \boldsymbol{D}_r^* && \text{if } r \text{ even,} \\ (D_1, \dots, D_r) \in \boldsymbol{D}_r^* &\implies \nu^{-1}(D_1, \dots, D_r) \subset \mathcal{D}^* && \text{if } r \text{ odd.} \end{aligned} \tag{5-9}$$

Without loss, we may restrict $\boldsymbol{D}_r^*$ to vectors with nonempty preimage in $\mathcal{D}^*$. Hence by construction, (5-5) becomes[5]

$$S^*(\tilde{\mathcal{A}}, z) \leqslant \sum_{0 \leqslant r \leqslant \varepsilon^{-2}} \sum_{(D_1, \dots, D_r) \in \boldsymbol{D}_r^*} \frac{(-1)^r}{\gamma(D_1, \dots, D_r)} \sum_{\substack{p_1 \cdots p_r \mid P(z) \\ D_j < p_j \leqslant D_j^{1+\tau}}} |\tilde{\mathcal{A}}_{p_1 \cdots p_r}|, \tag{5-10}$$

where $\gamma(D_1, \dots, D_r) = k_1! \cdots k_\ell!$ for the corresponding multiplicities $k_i \geqslant 1$ (i.e., we have $r = k_1 + \cdots + k_\ell$ and $D_1 = \cdots = D_{k_1} < D_{k_1+1} = \cdots = D_{k_2} < \cdots = D_r$.). Note the term $r = 0$ corresponds to $|\tilde{\mathcal{A}}|$ with $p_1 = \cdots = p_r = 1$.

[5] Indeed, we have reverse engineered the definition of $\boldsymbol{D}_r^*$ just so that (5-10) holds.

Now by the fundamental lemma [Friedlander and Iwaniec 2010, Theorem 6.9], we have upper (and lower) bounds

$$|\tilde{\mathcal{A}}_{p_1\cdots p_r}| = S(\mathcal{A}_{p_1\cdots p_r}, u) \leqslant g(p_1\cdots p_r)|\mathcal{A}|V(u)\{1+O(e^{-1/\varepsilon})\} + \sum_{b\leqslant D^\varepsilon} \lambda^{(r)}(b) r_{\mathcal{A}}(bp_1\cdots p_r),$$

(with $\leqslant$ replaced by $\geqslant$ for the lower bound) where $\lambda^{(r)}$ is the upper (lower) bound β-sieve of level D^ε when r is even (odd). For further details on the fundamental lemma and β-sieve, we refer the reader to [Friedlander and Iwaniec 2010, Sections 6 and 11].

Plugging back into (5-10), we get

$$\begin{aligned} S(\mathcal{A}, z) &\leqslant S^*(\tilde{\mathcal{A}}, z) \\ &\leqslant \sum_{0\leqslant r\leqslant \varepsilon^{-2}} \sum_{(D_1,\dots,D_r)\in \boldsymbol{D}_r^*} \frac{(-1)^r}{\gamma(D_1,\dots,D_r)} \sum_{\substack{p_1\cdots p_r \mid P(z)\\ D_j<p_j\leqslant D_j^{1+\tau}}} \\ &\qquad \times \Bigl\{ g(p_1\cdots p_r)|\mathcal{A}|V(u)\{1+O(e^{-1/\varepsilon})\} + \sum_{b\leqslant D^\varepsilon} \lambda^{(r)}(b) r_{\mathcal{A}}(bp_1\cdots p_r)\Bigr\}. \qquad (5\text{-}11) \end{aligned}$$

We now compare the main term above to that of the modified linear sieve, as in (4-3)–(4-11) from the proof of Proposition 4.1, namely,

$$V^*(D, z) := \sum_{d \mid P(z)/P(u)} \lambda^*(d) g(d) = V(z)\{F^*(s)+o(1)\}.$$

The difference between these main terms is accounted for by those d with two close prime factors, within a ratio D^τ, and those d near the boundary. The former contribution is

$$\sum_{\substack{d \mid P(z)/P(u)\\ p\leqslant p'<pD^\tau\\ pp' \mid d}} g(d) \leqslant \sum_{\substack{u\leqslant p<z\\ p\leqslant p'<pD^\tau}} g(pp')\cdot \prod_{u\leqslant p<z} (1+g(p)),$$

and the latter contribution is

$$\sum_r \sum_{\substack{u<p_r<\dots p_1<z\\ D^{1/(1+\tau)}<p_1\cdots p_m^3<D}} g(p_1\cdots p_r), \qquad (5\text{-}12)$$

where m is the first index ($m \leqslant r$) for which this occurs. Both contributions may be shown to be $O(\varepsilon^5)$, see [Friedlander and Iwaniec 2010, pages 254–255]. Hence the Proposition follows. □

Remark 5.2. We make a minor technical point. Namely, at an admissible cost $O(\varepsilon^5)$ we may assume Equation (5-1) holds, where $\boldsymbol{D}_r^*$ is further restricted to vectors $(D_1,\dots,D_r)$ satisfying

$$\nu^{-1}(D_1^{1/(1+\tau)},\dots,D_r^{1/(1+\tau)}) \subset \mathcal{D}^*, \qquad (5\text{-}13)$$

regardless of parity of r. To show this, note by definitions of $\boldsymbol{P}_{4,r}$, $\boldsymbol{P}_{6,r}$ the integers $p_1 \cdots p_r$ with $p_j \in [D_j^{1/(1+\tau)}, D_j]$, $j \leqslant r$, that lie outside $\mathcal{D}^*$ must satisfy

$$x^{1/6+2\eta}/p_1 \quad \text{or} \quad x^{1/12-5\eta}/p_6 \in [x^{-2\tau}, x^{2\tau}].$$

Then for $B_1 = x^{1/6+2\eta+2\tau}$, $B_6 = x^{1/12-5\eta+2\tau}$, we have

$$\sum_{B_i \geqslant p_i \geqslant \max(B/x^{4\tau}, u)} g(p_i) \ll \log \frac{\log B_i}{\log \max(B_i/x^{4\tau}, u)} \ll \frac{\log x^{4\tau}}{\log u} \ll \frac{\tau}{\varepsilon^2},$$

and so the contribution of such integers to the main term of (5-11) is

$$\ll \sum_r \sum_{\substack{u < p_r < \cdots < p_1 < z \\ B_i \geqslant p_i > \max(B_i/x^{4\tau}, u), i \in \{1,6\}}} g(p_1 \cdots p_r) \ll \frac{\tau}{\varepsilon^2} \prod_{u<p<z} (1 + g(p)) \ll \frac{\tau}{\varepsilon^2} \frac{\log z}{\log u} \ll \varepsilon^5.$$

Hence (5-13) follows.

We now proceed to Theorem 2.12. For each vector $(D_1, \ldots, D_r) \in \boldsymbol{D}_r$ we define the weight $\lambda_{(D_1,\ldots,D_r)}$ supported on d in $\nu^{-1}(D_1, \ldots, D_r)$, namely,

$$\lambda_{(D_1,\ldots,D_r)}(d) := \mathbf{1}^{d = p_1 \cdots p_r}_{D_j < p_j \leqslant D_j^{1+\tau} \forall j}. \tag{5-14}$$

Next, we may decompose an integer d into its u-smooth and rough components, $d = b(p_1 \cdots p_r)$. Recall $\mathcal{D}^{(r)} = \mathcal{D}^{\pm}$, $\lambda^{(r)} = \lambda^{\pm}$ (depending on the parity of r), and for $b \mid P(u)$ we have $\lambda^{(r)}(b) = \mu(b)$ if $b \in \mathcal{D}^{(r)}$, and $\lambda^{(r)}(b) = 0$ else. Thus we may define the convolution $\lambda^{(r)}_{(D_1,\ldots,D_r)} := \lambda_{(D_1,\ldots,D_r)} * \lambda^{(r)}$, i.e.,

$$\lambda^{(r)}_{(D_1,\ldots,D_r)}(d) = \begin{cases} \mu(b) & \text{if } d = bp_1 \cdots p_r \text{ for } b \in \mathcal{D}^{(r)}(D^{\varepsilon}),\, b \mid P(D^{\varepsilon^2}), \text{ and } D_j < p_j \leqslant D_j^{1+\tau} \forall j \leqslant r, \\ 0 & \text{else.} \end{cases} \tag{5-15}$$

Hence the remainder in (5-1) equals

$$\sum_{0 \leqslant r \leqslant \varepsilon^{-2}} \sum_{(D_1,\ldots,D_r) \in \boldsymbol{D}_r^*} \frac{(-1)^r}{\gamma(D_1, \ldots, D_r)} \sum_{d \mid P(z)} \lambda^{(r)}_{(D_1,\ldots,D_r)}(d) r_{\mathcal{A}}(d) = \sum_{d \mid P(z)} \tilde{\lambda}^*(d) r_{\mathcal{A}}(d)$$

for the weights

$$\tilde{\lambda}^* = \sum_{0 \leqslant r \leqslant \varepsilon^{-2}} \sum_{(D_1,\ldots,D_r) \in \boldsymbol{D}_r^*} \frac{(-1)^r}{\gamma(D_1, \ldots, D_r)} \lambda^{(r)}_{(D_1,\ldots,D_r)}. \tag{5-16}$$

Recalling the cardinality of $\boldsymbol{D}_r^* \subset \boldsymbol{D}_r^+$ from (5-7), it suffices to show the weights $\lambda^{(r)}_{(D_1,\ldots,D_r)}$ are programmably factorable for each vector in $\boldsymbol{D}_r^*$. To this we have the following.

Lemma 5.3. *For an integer d denote the vector $\nu(d) = (D_1, \ldots, D_r)$ from* (5-9). *If d has a factorization as in* (3-7) *at level D, then the corresponding weights $\lambda_{\nu(d)}$ and $\lambda^{(r)}_{\nu(d)}$, as in* (5-14) *and* (5-15), *resp., are programmably factorable sequences of levels $D^{1+\tau}$ and $D^{1+\tau+\varepsilon} = x^{7/12+\eta}$, resp.* (*relative to x, $\varepsilon/50$*).

Proof. By assumption for each $N \in [1, x^{1/3}]$, there is a factorization $d = d_1d_2d_3$ satisfying the system (3-7). For $j = 1, 2, 3$, write $d_j = \prod_{i \in I_j} p_i$ for the induced partition of indices $\{1, \dots, r\} = I_1 \cup I_2 \cup I_3$. Thus letting $Q_j = \prod_{i \in I_j} D_i$, the factorization $D_1 \cdots D_r = Q_1Q_2Q_3$ satisfies (2-1), since $D_i < p_i$.

Further, writing the corresponding subvectors $(D_i)_{i \in I_j}$ for $j = 1, 2, 3$, the weights $\lambda_{(D_i)_{i \in I_j}}$ are 1-bounded, supported on $[1, Q_j^{1+\tau}]$, and give the desired triple convolution,

$$\lambda_{(D_1,\dots,D_r)} = \lambda_{(D_i)_{i \in I_1}} * \lambda_{(D_i)_{i \in I_2}} * \lambda_{(D_i)_{i \in I_3}}.$$

Hence $\lambda_{(D_1,\dots,D_r)}$ is programmably factorable of level $D^{1+\tau}$ as claimed. Similarly $\lambda^{(r)}_{(D_1,\dots,D_r)} = \lambda_{(D_i)_{i \in I_1}} * \lambda_{(D_i)_{i \in I_2}} * \lambda^{(r)}_{(D_i)_{i \in I_3}}$ is programmably factorable of level $D^{1+\tau+\varepsilon}$. □

Now for each vector $(D_1, \dots, D_r) \in \boldsymbol{D}_r^*$, by (5-13) there exists $d = p_1 \cdots p_r \in \mathcal{D}^*$ for some primes $p_i \in (D_i^{1/(1+\tau)}, D_i]$. Then for all $N \in [1, x^{1/3}]$ Proposition 3.3 gives a factorization of d as in (3-7), and so by Lemma 5.3 $\lambda^{(r)}_{(D_1,\dots,D_r)}$ is programmably factorable of level $x^{7/12+\eta}$.

This completes the proof of Theorem 2.12.

5.1. ***Variable level of distribution for the linear sieve weights.*** We now return to Iwaniec's well-factorable weights $\tilde{\lambda}^{\pm}$ for the (upper and lower) linear sieve, given explicitly from (5-15) as the weighted sum,

$$\tilde{\lambda}^{\pm} = \sum_{0 \leqslant r \leqslant \varepsilon^{-2}} \sum_{(D_1,\dots,D_r) \in \boldsymbol{D}_r^{\pm}} \frac{(-1)^r}{\gamma(D_1, \dots, D_r)} \lambda^{(r)}_{(D_1,\dots,D_r)}. \tag{5-17}$$

We introduce the analogous set of well-factorable vectors $\boldsymbol{D}_r^{\text{well}} = \boldsymbol{D}_r^{\text{well}}(D)$,

$$\boldsymbol{D}_r^{\text{well}} = \{(D_1, \dots, D_r) : D_1 \cdots D_{m-1} D_m^2 < D \quad \text{for all } m \leqslant r\}. \tag{5-18}$$

Note $\boldsymbol{D}_r^{\pm} \subset \boldsymbol{D}_r^{\text{well}}$, having dropped parity conditions on the indices $m \leqslant r$.

We have the following technical variation on Theorem 2.12 for the original linear sieve.

Proposition 5.4. *Let $(D_1, \dots, D_r) \in \boldsymbol{D}_r^{\text{well}}(D)$ and write $D = x^{\theta}$, $D_i = x^{t_i}$ for $i \leqslant r$. If $\theta \leqslant \theta(t_1) - \varepsilon$ as in* (3-14), *then*

$$\sum_{\substack{b = p_1 \cdots p_r \\ D_i < p_i \leqslant D_i^{1+\tau}}} \sum_{\substack{d = bc \leq x^{\theta} \\ c \mid P(p_r) \\ (d,a)=1}} \tilde{\lambda}^{\pm}(d) \left(\pi(x; d, a) - \frac{\pi(x)}{\varphi(d)} \right) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}. \tag{5-19}$$

Moreover if $t_1 \leqslant \frac{1}{5}$ and $r \geqslant 3$, then (5-19) *holds provided that $\theta \leqslant \theta(t_1, t_2, t_3) - \varepsilon$ as in* (3-15).

If $t_1 \leqslant \frac{1}{5}$ and $r \leqslant 2$, then provided $\theta \leqslant \frac{1}{5}(3 - u) - \varepsilon$,

$$\sum_{\substack{b = p_1 \cdots p_r \\ D_i < p_i \leqslant D_i^{1+\tau}}} \sum_{\substack{d = bc \leqslant x^{\theta} \\ c \mid P(x^u) \\ (d,a)=1}} \tilde{\lambda}^{\pm}(d) \left(\pi(x; d, a) - \frac{\pi(x)}{\varphi(d)} \right) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}.$$

In particular for $r = 0$ (i.e., the empty vector), $\theta \leqslant \frac{1}{5}(3-u) - \varepsilon$, this simplifies as

$$\sum_{\substack{d \leqslant x^\theta \\ d \mid P(x^u) \\ (d,a)=1}} \tilde{\lambda}^{\pm}(d) \biggl(\pi(x; d, a) - \frac{\pi(x)}{\varphi(d)} \biggr) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}.$$

Proof. Given $(D_1, \dots, D_r)$, take an integer $b = p_1 \cdots p_r$ with $D_i < p_i \leqslant D_i^{1+\tau}$. Then for all multiples d of b with $d/b \mid P(p_r)$, the weight $\lambda^{(s)}_{(D'_1,\dots,D'_s)}(d)$ vanishes unless the vector $(D'_1, \dots, D'_s)$ extends $(D_1, \dots, D_r)$. That is, $D'_i = D_i$ for all $i \leqslant r$. Conversely, given such a vector $(D'_1, \dots, D'_s)$ we have $\lambda^{(s)}_{(D'_1,\dots,D'_s)}(d') = 0$ unless the first s primes of d' are $p_1 \cdots p_s$ with $D_i < p_i \leqslant D_i^{1+\tau}$, $i \leqslant r$. So by the definition of $\tilde{\lambda}^{\pm}$ as in (5-17), we have

$$\sum_{\substack{b = p_1 \cdots p_r \\ D_i < p_i \leqslant D_i^{1+\tau}}} \sum_{\substack{d = bc \leqslant x^\theta \\ c \mid P(p_r) \\ (d,a)=1}} \tilde{\lambda}^{\pm}(d) = \sum_{r \leqslant s \leqslant \varepsilon^{-2}} \sum_{\substack{(D'_1,\dots,D'_s) \in \boldsymbol{D}_s^{\pm} \\ D'_i = D_i, i \leqslant r}} \frac{(-1)^s}{\gamma(D'_1, \dots, D'_s)} \sum_{\substack{d \leqslant x^\theta \\ (d,a)=1}} \lambda^{(s)}_{(D'_1,\dots,D'_s)}(d). \tag{5-20}$$

Here we have extended (by zero) the inner sum to all $d \leqslant x^\theta$, $(d,a) = 1$.

Next, take such a vector $(D'_1, \dots, D'_s) \in \boldsymbol{D}_s^{\pm}(x^\theta)$ with $D'_i = D_i$ for $i \leqslant r$. Each integer $d = p_1 \cdots p_s$ with $D'_i < p_i \leqslant (D'_i)^{1+\tau}$ lies in $d \in \mathcal{D}^{\mathrm{well}}(x^{\theta+\tau})$. In particular $p_1 \leqslant D_1^{1+\tau} \leqslant x^{t_1+\tau}$ so by Corollary 3.8, d has a factorization as in (3-7) at level $x^{\theta(t_1+\tau)}$. Since $\theta(t)$ is continuous (in fact, piecewise linear), for $\tau > 0$ sufficiently small $\theta(t_1+\tau) \geqslant \theta(t_1) - \varepsilon \geqslant \theta$. Thus by Lemma 5.3 the weights $\lambda^{(s)}_{(D_1,\dots,D_s)}$ are programmably factorable sequences of level x^θ. Hence for each such vector, by Theorem 2.5 we have

$$\sum_{\substack{d \leqslant x^\theta \\ (d,a)=1}} \lambda^{(s)}_{(D'_1,\dots,D'_s)}(d) \ll_{a,A,\varepsilon} \frac{x}{(\log x)^A}. \tag{5-21}$$

Plugging (5-21) back into (5-20) gives the bound (5-19), as claimed.

Moreover, if $t_1 \leqslant \frac{1}{5}$ and $r \geqslant 3$ then proceeding as in the above paragraph, by Corollary 3.8 d will factor as in (3-7) to level $x^{\theta(t_1+\tau, t_2+\tau, \tau_3+\tau)}$. Again $\theta(t, u, v)$ is continuous, so for $\tau > 0$ sufficiently small

$$\theta(t_1+\tau, t_2+\tau, \tau_3+\tau) \geqslant \theta(t_1, t_2, t_3) - \varepsilon \geqslant \theta.$$

Hence (5-19) also follows in this case.

Similarly if $t_1 \leqslant \frac{1}{5}$ and $r \leqslant 2$, proceeding as above with $d = p_1 \cdots p_s$, the assumption $d/b \mid P(x^u)$ implies $s \leqslant 2$ or $p_3 \leqslant D_3^{1+\tau} \leqslant x^{u+\tau}$. Thus by Corollary 3.8 d will factor as in (3-7) to level $x^{(3-u-\tau)/5} \geqslant x^\theta$. Hence (5-19) follows in this case as well. □

5.2. *Equidistribution for products of primes.* We use an extension of Theorem 2.5 to products of k primes. This is the analogue in the programmably factorable setting of [Wu 1990, Lemma 7] extending Theorem 2.2 of Bombieri, Friedlander and Iwaniec.

Proposition 5.5. *Let $\varepsilon > 0$ and λ be a programmably factorable sequence of level $D \leqslant x^{3/5-\varepsilon}$ (relative to x, $\varepsilon/50$). Take real numbers $\varepsilon_1, \dots, \varepsilon_k \geqslant \varepsilon$ such that $\sum_{i\leqslant k} \varepsilon_i = 1$. Then for any fixed integer $a \in \mathbb{Z}$, $A, B > 0$, letting $\Delta = 1 + (\log x)^{-B}$ we have*

$$\sum_{\substack{d\leqslant D\\(d,a)=1}} \lambda(d)\Biggl(\sum_{\substack{p_1\cdots p_k\equiv a \pmod d\\ x^{\varepsilon_i}/\Delta<p_i\leqslant x^{\varepsilon_i}\forall i\leqslant k}} 1-\frac{1}{\varphi(d)}\sum_{\substack{(p_1\cdots p_k,d)=1\\ x^{\varepsilon_i}/\Delta<p_i\leqslant x^{\varepsilon_i}\forall i\leqslant k}} 1\Biggr) \ll_{a,\varepsilon,A,B} \frac{x}{(\log x)^A}. \tag{5-22}$$

Proof. This follows by the same proof method as in Theorem 2.5 (i.e., [Maynard 2020, Theorem 1.1]). Indeed, Maynard just uses the Heath-Brown identity to decompose the indicator function of primes into Type I/II sums. A similar decomposition holds for products of k primes, after which we may apply the same Type I/II estimates in Propositions 5.1 and 5.2 of [Maynard 2020]. □

In addition, by replacing Theorem 2.5 with Proposition 5.5 in the proof, we obtain analogues of Proposition 5.4 for the linear sieve weights $\lambda = \tilde{\lambda}^{\pm}$ in the case of products of k primes.

Corollary 5.6. *Let $(D_1, \dots, D_r) \in \boldsymbol{D}_r^{\text{well}}(D)$ and write $D = x^{\theta}$, $D_i = x^{t_i}$ for $i \leqslant r$. Let $\varepsilon > 0$ and real numbers $\varepsilon_1, \dots, \varepsilon_k \geqslant \varepsilon$ such that $\sum_{i\leqslant k} \varepsilon_i = 1$. Fix an integer $a \in \mathbb{Z}$, $A, B > 0$, and let $\Delta = 1+(\log x)^{-B}$. If $\theta \leqslant \theta(t_1) - \varepsilon$ as in* (3-14),

$$\sum_{\substack{b=p'_1\cdots p'_r\\ D_i<p'_i\leqslant D_i^{1+\tau}}}\ \sum_{\substack{d=bc\leqslant x^{\theta}\\ c\,|\,P(p'_r)\\(d,a)=1}} \tilde{\lambda}^{\pm}(d)\Biggl(\sum_{\substack{p_1\cdots p_k\equiv a \pmod d\\ x^{\varepsilon_i}/\Delta<p_i\leqslant x^{\varepsilon_i}\forall i\leqslant k}} 1-\frac{1}{\varphi(d)}\sum_{\substack{(p_1\cdots p_k,d)=1\\ x^{\varepsilon_i}/\Delta<p_i\leqslant x^{\varepsilon_i}\forall i\leqslant k}} 1\Biggr) \ll_{a,\varepsilon,A,B} \frac{x}{(\log x)^A}. \tag{5-23}$$

Moreover if $t_1 \leqslant \frac{1}{5}$, $r \geqslant 3$, then (5-23) *holds provided that $\theta \leqslant \theta(t_1, t_2, t_3) - \varepsilon$ as in* (3-15).

In addition, if $r \leqslant 2$, $u \leqslant t_r$, $t_1 \leqslant \frac{1}{5}$, and $\theta \leqslant \frac{3-u}{5} - \varepsilon$, then

$$\sum_{\substack{b=p'_1\cdots p'_r\\ D_i<p'_i\leqslant D_i^{1+\tau}}}\ \sum_{\substack{d=bc\leqslant x^{\theta}\\ c\,|\,P(x^u)\\(d,a)=1}} \tilde{\lambda}^{\pm}(d)\Biggl(\sum_{\substack{p_1\cdots p_k\equiv a \pmod d\\ x^{\varepsilon_i}/\Delta<p_i\leqslant x^{\varepsilon_i}\forall i\leqslant k}} 1-\frac{1}{\varphi(d)}\sum_{(p_1\cdots p_k,d)=1 x^{\varepsilon_i}/\Delta<p_i\leqslant x^{\varepsilon_i}\forall i\leqslant k} 1\Biggr) \ll_{a,\varepsilon,A,B} \frac{x}{(\log x)^A}. \tag{5-24}$$

6. Upper bound for twin primes

Now we shall apply the modified sieve to the set of twin primes

$$\mathcal{A} := \{p+2 : p \leqslant x\}.$$

In this case the sieve notation specializes as $\mathcal{P} = \{p > 2\}$ and $g(d) = 1/\varphi(d)$ for odd d, so that $V(z) = \prod_{2<p<z}(1 - 1/\varphi(p))$. Recall $V(z) \sim \mathfrak{S}_2/e^{\gamma} \log z$ by Mertens theorem, for the Hardy–Littlewood constant $\mathfrak{S}_2 = 2\prod_{p>2}(1-2/p)/(1-1/p)^2$ appearing in $\Pi(x) = \mathfrak{S}_2 x/(\log x)^2$.

We begin in the spirit of Fouvry and Grupp [1986], and apply a weighted sieve inequality. To each nonswitched term, we apply the Buchstab identity in order to lower the sieve threshold down to $z = x^{\epsilon}$ for some tiny $\epsilon > 0$. By Proposition 5.4, such smooth sums will satisfy level of distribution $\frac{1}{5}(3-\epsilon)$.

Combined with variations on a theme, which identify programmably factorable weights in certain cases, the consequent increase in level will be sufficient to obtain the bound in Theorem 1.2. For a final bit of savings, we use refinements obtained by Wu's iteration method [2004].

Before moving on, we note that sieve methods and the switching principle have also yielded progress on the Goldbach problem. Indeed, the upper bound in [Wu 2004, Theorem 3] for twin primes is obtained by using the same formulae as in [loc. cit., Theorem 1] for the Goldbach problem, except for altering the level from $\frac{1}{2}$ to $\frac{4}{7}$ (this amounts to replacing factors of 4 with $\frac{7}{2}$ in a few instances). Importantly, the quantitative upper bounds for twin primes are much stronger than those of the Goldbach problem. This is because the latter relies on level $\frac{1}{2}$ from Bombieri and Vinogradov for a growing residue $a = N$, while the former may appeal to level of distribution $\frac{4}{7}$ from Bombieri, Friedlander and Iwaniec for the fixed residue $a = 2$ (and now the subsequent improvements of Maynard). As such our methods have nothing new to say for the Goldbach problem.

6.1. ***Lemmas.*** We begin with a standard lemma for $x^{1/u}$-rough numbers in terms of the Buchstab function $\omega(u) = (f(u) + F(u))/(2e^{\gamma})$ for linear sieve functions f, F as in Theorem 2.10. Alternatively, ω is directly defined via

$$\begin{aligned} \omega(u) &= \frac{1}{u} && \text{for } 1 \leqslant u \leqslant 2, \\ (u\omega(u))' &= \omega(u-1) && \text{for } 2 \leqslant u. \end{aligned}$$

Lemma 6.1. *Let $x \geqslant 2$ and $y = x^{1/u}$. Then we have*

$$\sum_{\substack{n\leqslant x \\ p\,|\,n \Rightarrow p \geqslant y}} 1 = \omega(u)\frac{x}{\log y} + O\biggl(\frac{x}{(\log y)^2}\biggr).$$

Proof. This is [Wu 1990, Lemma 12]. □

The argument of Wu makes essential use of weighted sieve inequalities, as in [Wu 2004, Lemmas 4.1 and 4.2]. We shall employ the latter inequality in the special case $d = 1$, $\sigma = 1$.

Lemma 6.2. *For $\frac{3}{10} \geqslant \rho \geqslant \tau_3 > \tau_2 > \tau_1 \geqslant \rho' \geqslant \frac{1}{20}$, we have*

$$5S(\mathcal{A}, x^{\rho}) \lesssim \sum_{1\leqslant n\leqslant 21} \Gamma_n,$$

where

$$\Gamma_1 := 4S(\mathcal{A}, x^{\rho'}) + S(\mathcal{A}, x^{\tau_1}),$$

$$\Gamma_2 := -\sum_{x^{\rho'}\leqslant p<x^{\rho}} S(\mathcal{A}_p, x^{\rho'}),$$

$$\Gamma_3 := -\sum_{x^{\rho'}\leqslant p<x^{\tau_2}} S(\mathcal{A}_p, x^{\rho'}),$$

$$\Gamma_4 := -\sum_{x^{\rho'}\leqslant p<x^{\tau_3}} S(\mathcal{A}_p, x^{\rho'}),$$

$$\Gamma_{12} := \sum\sum\sum_{\substack{x^{\rho'}\leqslant p_1<p_2<x^{\tau_1} \\ x^{\tau_3}\leqslant p_3<x^{\rho}}} S(\mathcal{A}_{p_1p_2p_3}, p_2),$$

$$\Gamma_{13} := \sum\sum\sum_{x^{\rho'}\leqslant p_1<x^{\tau_1}\leqslant p_2<x^{\tau_2}\leqslant p_3<x^{\rho}} S(\mathcal{A}_{p_1p_2p_3}, p_2),$$

$$\Gamma_{14} := \sum\sum\sum_{\substack{x^{\rho'}\leqslant p_1<x^{\tau_1} \\ x^{\tau_2}\leqslant p_2<p_3<x^{\rho}}} S(\mathcal{A}_{p_1p_2p_3}, p_2),$$

$$\Gamma_5 := \sum\sum_{x^{\rho'}\leqslant p_1<p_2<x^{\tau_2}} S(\mathcal{A}_{p_1p_2}, x^{\rho'}),$$

$$\Gamma_6 := \sum\sum_{\substack{x^{\rho'}\leqslant p_1<x^{\tau_1}\\ x^{\tau_2}\leqslant p_2<x^{\tau_3}}} S(\mathcal{A}_{p_1p_2}, x^{\rho'}),$$

$$\Gamma_7 := \sum\sum_{x^{\rho'}\leqslant p_1<p_2<x^{\tau_1}} S(\mathcal{A}_{p_1p_2}, p_1),$$

$$\Gamma_8 := \sum\sum_{x^{\rho'}\leqslant p_1<x^{\tau_1}\leqslant p_2<x^{\tau_2}} S(\mathcal{A}_{p_1p_2}, p_1),$$

$$\Gamma_9 := \sum\sum\sum_{x^{\tau_1}\leqslant p_1<p_2<p_3<x^{\tau_3}} S(\mathcal{A}_{p_1p_2p_3}, p_2),$$

$$\Gamma_{10} := \sum\sum\sum_{x^{\tau_1}\leqslant p_1<p_2<x^{\tau_2}\leqslant p_3<x^{\rho}} S(\mathcal{A}_{p_1p_2p_3}, p_2),$$

$$\Gamma_{11} := \sum\sum\sum_{x^{\tau_1}\leqslant p_1<x^{\tau_2}\leqslant p_2<p_3<x^{\tau_3}} S(\mathcal{A}_{p_1p_2p_3}, p_2),$$

$$\Gamma_{15} := \sum\sum\sum_{x^{\tau_1}\leqslant p_1<x^{\tau_2}\leqslant p_2<x^{\tau_3}\leqslant p_3<x^{\rho}} S(\mathcal{A}_{p_1p_2p_3}, p_2),$$

$$\Gamma_{16} := \sum\sum\sum\sum_{x^{\tau_2}\leqslant p_1<p_2<p_3<p_4<x^{\tau_3}} S(\mathcal{A}_{p_1p_2p_3p_4}, p_3),$$

$$\Gamma_{17} := \sum\sum\sum\sum_{x^{\tau_2}\leqslant p_1<p_2<p_3<x^{\tau_3}\leqslant p_4<x^{\rho}} S(\mathcal{A}_{p_1p_2p_3p_4}, p_3),$$

$$\Gamma_{18} := \sum\sum\sum\sum_{x^{\tau_2}\leqslant p_1<p_2<x^{\tau_3}\leqslant p_3<p_4<x^{\rho}} S(\mathcal{A}_{p_1p_2p_3p_4}, p_3),$$

$$\Gamma_{19} := \sum\sum\sum\sum_{\substack{x^{\tau_1}\leqslant p_1 x^{\tau_2}\\ x^{\tau_3}\leqslant p_2<p_3<p_4<x^{\rho}}} S(\mathcal{A}_{p_1p_2p_3p_4}, p_3),$$

$$\Gamma_{20} := \sum\sum\sum\sum\sum_{x^{\tau_2}\leqslant p_1<x^{\tau_3}\leqslant p_2<p_3<p_4<p_5<x^{\rho}} S(\mathcal{A}_{p_1p_2p_3p_4p_5}, p_4),$$

$$\Gamma_{21} := \sum\sum\sum\sum\sum\sum_{x^{\tau_3}\leqslant p_1<p_2<p_3<p_4<p_5<p_6<x^{\rho}} S(\mathcal{A}_{p_1p_2p_3p_4p_5p_6}, p_5).$$

Proof. This is [Wu 2004, Lemma 4.2] with $d = 1$, $\sigma = 1$. Here we simplify Wu's notation slightly, using $(\underline{d}^{1/s}, \underline{d}^{1/\kappa_3}, \underline{d}^{1/\kappa_2}, \underline{d}^{1/\kappa_1}, \underline{d}^{1/s'}) = (x^{\rho}, x^{\tau_3}, x^{\tau_2}, x^{\tau_1}, x^{\rho'})$. The basic proof idea is to iterate the Buchstab identity and to strategically neglect some terms by positivity. □

6.2. *Computations.* Given $0.1 \leqslant \rho' \leqslant \tau_1 < 0.2 \leqslant \tau_2 < \tau_3 \leqslant \rho \leqslant 0.3.$, we define integrals $I_n = I_n(\rho, \rho', \tau_1, \tau_2, \tau_3)$ by

$$\begin{aligned}
I_n &= \int_{\mathbb{D}_n} \omega\Big(\frac{1-t-u-v}{u}\Big)\frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v}{tu^2v} && (9 \leqslant n \leqslant 15),\\
I_n &= \int_{\mathbb{D}_n} \omega\Big(\frac{1-t-u-v-w}{v}\Big)\frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w}{tuv^2w} && (16 \leqslant n \leqslant 19),\\
I_{20} &= \int_{\mathbb{D}_{20}} \omega\Big(\frac{1-t-u-v-w-x}{w}\Big)\frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,\mathrm{d}x}{tuvw^2x},\\
I_{21} &= \int_{\mathbb{D}_{21}} \omega\Big(\frac{1-t-u-v-w-x-y}{x}\Big)\frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,\mathrm{d}x\,\mathrm{d}y}{tuvwx^2y},
\end{aligned} \tag{6-1}$$

where ω is the Buchstab function, and where the domains $\mathbb{D}_n$ are

$$\begin{aligned}
\mathbb{D}_9 &= \{(t, u, v) : \tau_1 < t < u < v < \tau_3\},\\
\mathbb{D}_{10} &= \{(t, u, v) : \tau_1 < t < u < \tau_2 < v < \rho\},\\
\mathbb{D}_{11} &= \{(t, u, v) : \tau_1 < t < \tau_2 < u < v < \tau_3\},\\
\mathbb{D}_{12} &= \{(t, u, v) : \rho' < t < u < \tau_1, \tau_3 < v < \rho\},\\
\mathbb{D}_{13} &= \{(t, u, v) : \rho' < t < \tau_1 < u < \tau_2 < v < \rho\},\\
\mathbb{D}_{14} &= \{(t, u, v) : \rho' < t < \tau_1, \tau_2 < u < v < \rho\},
\end{aligned}$$

$$\begin{aligned}
\mathbb{D}_{15} &= \{(t,u,v) : \tau_1 < t < \tau_2 < u < \tau_3 < v < \rho\},\\
\mathbb{D}_{16} &= \{(t,u,v,w) : \tau_2 < t < u < v < w < \tau_3\},\\
\mathbb{D}_{17} &= \{(t,u,v,w) : \tau_2 < t < u < v < \tau_3 < w < \rho\},\\
\mathbb{D}_{18} &= \{(t,u,v,w) : \tau_2 < t < u < \tau_3 < v < w < \rho\},\\
\mathbb{D}_{19} &= \{(t,u,v,w) : \tau_1 < t < \tau_2, \tau_3 < u < v < w < \rho\},\\
\mathbb{D}_{20} &= \{(t,u,v,w,x) : \tau_2 < t < \tau_3 < u < v < w < x < \rho\},\\
\mathbb{D}_{21} &= \{(t,u,v,w,x,y) : \tau_3 < t < u < v < w < x < y < \rho\}.
\end{aligned}$$

Recall the definitions (3-14) and (3-15),

$$\theta(t) = \begin{cases} \frac{2-t}{3} & \text{if } t > \frac{1}{5},\\ \frac{1+t}{2} & \text{if } t \leqslant \frac{1}{5}. \end{cases}$$

We let $\theta_\epsilon = \frac{3-\epsilon}{5}$ and

$$\theta(t,u,v) := \max\Big\{\frac{3-v}{5}, \theta(t), \theta(u), \theta(t+u+v), \theta(t+u), \theta(t+v), \theta(u+v)\Big\}.$$

We also define

$$\begin{aligned} G_1 &= 4G(\rho') + G(\tau_1), & G_3 &= G_0 + \overline{G}(\tau_2),\\ G_2 &= G_0 + \overline{G}(\rho), & G_4 &= G_0 + \overline{G}(\tau_3), \end{aligned} \tag{6-2}$$

where for $c \leqslant \frac{1}{5}$,

$$\begin{aligned} G(c) = \frac{1}{\epsilon}F(\theta_\epsilon/\epsilon) - \frac{1}{\epsilon}\int_\epsilon^c \frac{\mathrm{d}t}{t} f((\theta_\epsilon - t)/\epsilon) + \frac{1}{\epsilon}\int_\epsilon^c\int_\epsilon^t \frac{\mathrm{d}t\,\mathrm{d}u}{tu} F((\theta_\epsilon - t - u)/\epsilon)\\ - \int_\epsilon^c\int_\epsilon^t\int_\epsilon^u \frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v}{tuv^2} f((\theta(t,u,v) - t - u - v)/v), \end{aligned} \tag{6-3}$$

and for $c > \frac{1}{5}$,

$$\overline{G}(c) = -\frac{1}{\epsilon}\int_{1/5}^c \frac{\mathrm{d}t}{t} f((\theta(t) - t)/\epsilon) + \int_{1/5}^c\int_\epsilon^{\rho'} \frac{\mathrm{d}t\,\mathrm{d}u}{tu^2} F((\theta(t) - t - u)/u) \tag{6-4}$$

as well as

$$\begin{aligned} G_0 = -\frac{1}{\epsilon}\int_{\rho'}^{1/5} \frac{\mathrm{d}t}{t} f((\theta_\epsilon - t)/\epsilon) + \frac{1}{\epsilon}\int_{\rho'}^{1/5}\int_\epsilon^{\rho'} \frac{\mathrm{d}t\,\mathrm{d}u}{tu} F((\theta_\epsilon - t - u)/\epsilon)\\ - \int_{\rho'}^{1/5}\int_\epsilon^{\rho'}\int_\epsilon^u \frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v}{tuv^2} f((\theta(t,u,v) - t - u - v)/v). \end{aligned} \tag{6-5}$$

We similarly let

$$
\begin{aligned}
G_5 &= \frac{1}{\epsilon}\int_{\rho'}^{1/5}\int_{\rho'}^{t}\frac{\mathrm{d}t\,\mathrm{d}u}{tu}F((\theta_\epsilon - t - u)/\epsilon) + \frac{1}{\rho'}\int_{1/5}^{\tau_2}\int_{\rho'}^{t}\frac{\mathrm{d}t\,\mathrm{d}u}{tu}F((\theta(t)-t-u)/\rho') \\
&\qquad - \int_{\rho'}^{1/5}\int_{\rho'}^{t}\int_{\epsilon}^{\rho'}\frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v}{tuv^2}f((\theta(t,u,v)-t-u-v)/v), \\
G_6 &= \frac{1}{\rho'}\int_{\tau_2}^{\tau_3}\int_{\rho'}^{\tau_1}\frac{\mathrm{d}t\,\mathrm{d}u}{tu}F((\theta(t)-t-u)/\rho'), \\
G_7 &= \frac{1}{\epsilon}\int_{\rho'}^{\tau_1}\int_{\rho'}^{t}\frac{\mathrm{d}t\,\mathrm{d}u}{tu}F((\theta_\epsilon - t - u)/\epsilon) - \int_{\rho'}^{\tau_1}\int_{\rho'}^{t}\int_{\epsilon}^{u}\frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v}{tuv^2}f((\theta(t,u,v)-t-u-v)/v), \\
G_8 &= \frac{1}{\epsilon}\int_{\tau_1}^{1/5}\int_{\rho'}^{\tau_1}\frac{\mathrm{d}t\,\mathrm{d}u}{tu}F((\theta_\epsilon - t - u)/\epsilon) + \int_{1/5}^{\tau_2}\int_{\rho'}^{\tau_1}\frac{\mathrm{d}t\,\mathrm{d}u}{tu^2}F((\theta(t)-t-u)/u) \\
&\qquad - \int_{\tau_1}^{1/5}\int_{\rho'}^{\tau_1}\int_{\epsilon}^{u}\frac{\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v}{tuv^2}f((\theta(t,u,v)-t-u-v)/v).
\end{aligned}
\tag{6-6}
$$

Recall the sieve functions F, f satisfy $F(s) = 2e^\gamma/s$ for $s \in [1, 3]$, $f(s) = 2e^\gamma \log(s-1)/s$ for $s \in [2, 4]$ and $F(s) = 2e^\gamma/s \cdot \left[1 + \int_2^{s-1} f(t)\,\mathrm{d}t\right]$ for all $s \geqslant 1$.

The main bound is the following.

Proposition 6.3. *Let* $0 < \epsilon \leqslant 0.1 \leqslant \rho' \leqslant \tau_1 < 0.2 \leqslant \tau_2 < \tau_3 \leqslant \rho \leqslant 0.3$. *Then for* I_n, G_n, *and* $G(c)$ *as in* (6-1), (6-2), (6-6), *and* (6-3), *we have*

$$
S(\mathcal{A}, x^\rho) \lesssim \frac{\Pi(x)}{5e^\gamma}\biggl(\sum_{n=1}^{8} G_n + G\bigl(\tfrac{1}{5}\bigr)\sum_{n=9}^{21} I_n\biggr). \tag{6-7}
$$

Proof. We first bound $S(\mathcal{A}, x^c)$ for $c \in \left[\epsilon, \frac{1}{5}\right]$. By the Buchstab identity,

$$
S(\mathcal{A}, x^c) = S(\mathcal{A}, x^\epsilon) - \sum_{x^\epsilon \leqslant p < x^c} S(\mathcal{A}_p, p).
$$

Iterating twice more, we obtain

$$
S(\mathcal{A}, x^c) = S(\mathcal{A}, x^\epsilon) - \sum_{x^\epsilon \leqslant p_1 < x^c} S(\mathcal{A}_{p_1}, x^\epsilon) + \sum_{x^\epsilon \leqslant p_2 < p_1 < x^c} S(\mathcal{A}_{p_1 p_2}, x^\epsilon) - \sum_{x^\epsilon \leqslant p_3 < p_2 < p_1 < x^c} S(\mathcal{A}_{p_1 p_2 p_3}, p_3). \tag{6-8}
$$

To each term $S(\mathcal{A}_d, x^\epsilon)$ above, we apply the linear sieve of level x^{θ_ϵ} for $\theta_\epsilon = \frac{1}{5}(3-\epsilon)$, as in Theorem 2.10. And to each term $S(\mathcal{A}_{p_1 p_2 p_3}, p_3)$, we apply the linear sieve of level x^θ for $\theta = \theta(t_1, t_2, t_3)$, where $p_i = x^{t_i}$.

To handle the corresponding error terms, note for primes $x^\epsilon \leqslant p_2 < p_1 < x^c \leqslant x^{1/5}$ and $d \in \{1, p_1, p_1 p_2\}$, the prime factors of q/d above are bounded by x^ϵ so that the sets $\mathcal{A}_q$ are equidistributed to level θ_ϵ. Hence for each $x^\epsilon \leqslant p_2 < p_1 < x^{1/5}$, by Proposition 5.4 with $u = \epsilon$, $b = 1$, p_1, $p_1 p_2$, we have

$$
\sum_{\substack{q \leqslant x^{\theta_\epsilon} \\ q \mid P(x^\epsilon)}} \tilde{\lambda}^+(q)\biggl(|\mathcal{A}_q| - \frac{|\mathcal{A}|}{\varphi(q)}\biggr) = \sum_{\substack{q \leqslant x^{\theta_\epsilon} \\ q \mid P(x^\epsilon)}} \tilde{\lambda}^+(q)\biggl(\pi(x; q, -2) - \frac{\pi(x)}{\varphi(q)}\biggr) \ll_A \frac{x}{(\log x)^A}
$$

and

$$\sum_{p_1}\sum_{\substack{q=p_1m\leqslant x^{\theta_\epsilon}\\ m\,|\,P(x^\epsilon)}}\tilde{\lambda}^{-}(q)\biggl(|\mathcal{A}_q|-\frac{|\mathcal{A}|}{\varphi(q)}\biggr)\ll_A\frac{x}{(\log x)^A}$$

and

$$\sum_{p_2,p_1}\sum_{\substack{q=p_1p_2m\leqslant x^{\theta_\epsilon}\\ m\,|\,P(x^\epsilon)}}\tilde{\lambda}^{+}(q)\biggl(|\mathcal{A}_q|-\frac{|\mathcal{A}|}{\varphi(q)}\biggr)\ll_A\frac{x}{(\log x)^A}.$$

In addition for each $p_3<p_2<p_1<x^{1/5}$, $p_i=x^{t_i}$, letting $\theta=\theta(t_1,t_2,t_3)$, by Proposition 5.4 with $b=p_1p_2p_3$,

$$\sum_{p_3,p_2,p_1}\sum_{\substack{q=p_1p_2p_3m\leqslant x^{\theta}\\ m\,|\,P(p_3)}}\tilde{\lambda}^{-}(q)\biggl(|\mathcal{A}_q|-\frac{|\mathcal{A}|}{\varphi(q)}\biggr)\ll_A\frac{x}{(\log x)^A}.$$

Thus for $c\leqslant\frac{1}{5}$, the linear sieve bounds give

$$S(\mathcal{A},x^\epsilon)\lesssim|\mathcal{A}|V(x^\epsilon)F\Bigl(\frac{\theta_\epsilon}{\epsilon}\Bigr) \tag{6-9}$$

and

$$\sum_{x^\epsilon\leqslant p_1<x^c}S(\mathcal{A}_{p_1},x^\epsilon)\gtrsim\sum_{x^\epsilon\leqslant p_1<x^c}|\mathcal{A}|g(p_1)V(x^\epsilon)f\Bigl(\frac{\theta_\epsilon-t_1}{\epsilon}\Bigr) \tag{6-10}$$

and

$$\sum_{x^\epsilon\leqslant p_2<p_1<x^c}S(\mathcal{A}_{p_1p_2},x^\epsilon)\lesssim\sum_{x^\epsilon\leqslant p_2<p_1<x^c}|\mathcal{A}|g(p_1p_2)V(x^\epsilon)F\Bigl(\frac{\theta_\epsilon-t_1-t_2}{\epsilon}\Bigr) \tag{6-11}$$

and

$$\sum_{x^\epsilon\leqslant p_3<p_2<p_1<x^c}S(\mathcal{A}_{p_1p_2p_3},p_3)\gtrsim\sum_{x^\epsilon\leqslant p_3<p_2<p_1<x^c}|\mathcal{A}|g(p_1p_2p_3)V(p_3)f\Bigl(\frac{\theta-t_1-t_2-t_3}{t_3}\Bigr). \tag{6-12}$$

Hence by (6-9), (6-10), (6-11), (6-12), we observe that (6-8) becomes

$$\begin{aligned}S(\mathcal{A},x^c)\lesssim-|\mathcal{A}|&\sum_{x^\epsilon\leqslant p_3<p_2<p_1<x^c}g(p_1p_2p_3)V(p_3)f\Bigl(\frac{\theta-t_1-t_2-t_3}{\epsilon}\Bigr)\\ &+|\mathcal{A}|V(x^\epsilon)\biggl(F\Bigl(\frac{\theta_\epsilon}{\epsilon}\Bigr)-\sum_{x^\epsilon\leqslant p_1<x^c}g(p_1)f\Bigl(\frac{\theta_\epsilon-t_1}{\epsilon}\Bigr)+\sum_{x^\epsilon\leqslant p_2<p_1<x^c}g(p_1p_2)F\Bigl(\frac{\theta_\epsilon-t_1-t_2}{\epsilon}\Bigr)\biggr).\end{aligned} \tag{6-13}$$

Recall $V(z)\sim\mathfrak{S}_2/e^\gamma\log z$ by Merten's theorem. Thus by partial summation and the prime number theorem, we obtain

$$S(\mathcal{A},x^c)\lesssim\frac{\Pi(x)}{e^\gamma}G(c), \tag{6-14}$$

for $G(c)$ as in (6-3). Hence for $c=\rho',\tau_1$ we have $c\in\bigl[\epsilon,\frac{1}{5}\bigr]$, so we bound Γ_1 as

$$\Gamma_1=4S(\mathcal{A},x^{\rho'})+S(\mathcal{A},x^{\tau_1})\lesssim\frac{\Pi(x)}{e^\gamma}(4G(\rho')+G(\tau_1))=\frac{\Pi(x)}{e^\gamma}G_1.$$

Now consider $c, c' \in \left[\frac{1}{5}, \frac{2}{7}\right]$. We shall apply the linear sieve of level $\theta(t_1)$, as in (3-14). In general, for $p_1 = x^{t_1}$ and $\theta(t_1)$ as in (3-14), Proposition 5.4 gives

$$\sum_{p_1} \sum_{\substack{p_1 \mid q, q \leqslant x^{\theta(t_1)} \\ q/p_1 \mid P(p_1)}} \tilde{\lambda}^{-}(q)\left(|\mathcal{A}_q| - \frac{|\mathcal{A}|}{\varphi(q)}\right) \ll_A \frac{x}{(\log x)^A},$$

so that for $c, c' \in \left[\frac{1}{5}, \frac{2}{7}\right]$, the linear sieve of level $\theta(t_1)$ gives

$$\sum_{x^{c'} \leqslant p_1 < x^c} S(\mathcal{A}_{p_1}, p_1) \gtrsim \sum_{x^{c'} \leqslant p_1 < x^c} |\mathcal{A}| g(p_1) V(p_1) f\left(\frac{\theta(t_1) - t_1}{t_1}\right). \tag{6-15}$$

Thus by partial summation and the prime number theorem,

$$\Gamma_2 = \sum_{x^{\rho'} \leqslant p < x^{\rho}} S(\mathcal{A}_p, x^{\rho'}) \lesssim \frac{\Pi(x)}{e^{\gamma}}(G + \bar{G}(\rho)) = \frac{\Pi(x)}{e^{\gamma}} G_2.$$

Similarly, we obtain

$$\Gamma_n \lesssim \frac{\Pi(x)}{e^{\gamma}} G_n \quad \text{for } 1 \leqslant n \leqslant 8. \tag{6-16}$$

Finally, for the remaining Γ_n, we apply the switching principle. Namely, for Γ_9 we have

$$\Gamma_9 = \sum\sum\sum_{x^{\tau_1} \leqslant p_1 < p_2 < p_3 < x^{\tau_3}} S(\mathcal{A}_{p_1 p_2 p_3}, p_2) = S(\mathcal{B}, x^{1/2}) + O(x^{1/2}) \tag{6-17}$$

for the set

$$\mathcal{B} = \{p_1 p_2 p_3 m - 2 \leqslant x : x^{\tau_1} \leqslant p_1 < p_2 < p_3 < x^{\tau_3},\ p' \mid m \Rightarrow p' \geqslant p_2\}. \tag{6-18}$$

Note since $p_2, p_1 > x^{\tau_1} > x^{0.1}$, each m above has at most 7 prime factors.

Now by a standard subdivision argument, $\mathcal{B}$ is similarly equidistributed in arithmetic progressions as is $\mathcal{A}$. Indeed, the basic idea is to partition $\mathcal{B} = \bigcup_{r \leqslant 7} \mathcal{B}^{(r)}$, where $\mathcal{B}^{(r)}$ is the subset corresponding to integers m with r prime factors. Then we cover the prime tuples $(p_1, \ldots, p_r)$ into hypercubes of the form

$$[\Delta^{l_1}, \Delta^{l_1+1}) \times \cdots \times [\Delta^{l_r}, \Delta^{l_r+1})$$

for $\Delta = 1 + (\log x)^{-B}$ with $B > 0$ sufficiently large, and apply Corollary 5.6 with $u = \epsilon$, $b = 1$. This gives

$$\sum_{\substack{q \leqslant x^{\theta_\epsilon} \\ q \mid P(x^{\epsilon})}} \tilde{\lambda}^{+}(q)\left(|\mathcal{B}_q| - \frac{|\mathcal{B}|}{\varphi(q)}\right) \ll_A \frac{x}{(\log x)^A}.$$

Similarly for each $x^{\varepsilon} < p_3' < p_2' < p_1' < x^{1/5}$, by Corollary 5.6 with $u = \epsilon$ and $b = p_1', p_1' p_2'$, we have

$$\sum_{p_1'} \sum_{\substack{q = p_1' m' \leqslant x^{\theta_\epsilon} \\ m' \mid P(x^{\epsilon})}} \tilde{\lambda}^{-}(q)\left(|\mathcal{B}_q| - \frac{|\mathcal{B}|}{\varphi(q)}\right) \ll_A \frac{x}{(\log x)^A}$$

and

$$\sum_{p_2',p_1'}\sum_{\substack{q=p_1'p_2'm'\leqslant x^{\theta_\epsilon}\\ m'\mid P(x^\epsilon)}}\tilde{\lambda}^{+}(q)\biggl(|\mathcal{B}_q|-\frac{|\mathcal{B}|}{\varphi(q)}\biggr)\ll_A\frac{x}{(\log x)^A}.$$

In addition for each $p_3' < p_2' < p_1' < x^{1/5}$, $p_i' = x^{t_i}$, letting $\theta=\theta(t_1,t_2,t_3)$, by Corollary 5.6 with $b=p_1'p_2'p_3'$, we have

$$\sum_{p_3',p_2',p_1'}\sum_{\substack{q=p_1'p_2'p_3'm'\leqslant x^{\theta}\\ m'\mid P(p_3')}}\tilde{\lambda}^{-}(q)\biggl(|\mathcal{B}_q|-\frac{|\mathcal{B}|}{\varphi(q)}\biggr)\ll_A\frac{x}{(\log x)^A}.$$

Iterating the Buchstab identity, we have

$$\begin{aligned}
&S(\mathcal{B},x^{1/2})\\
&\quad\leqslant S(\mathcal{B},x^{1/5})\\
&\quad=S(\mathcal{B},x^{\epsilon})-\sum_{x^\epsilon\leqslant p_1'<x^{1/5}}S(\mathcal{B}_{p_1'},x^{\epsilon})+\sum_{x^\epsilon\leqslant p_2'<p_1'<x^{1/5}}S(\mathcal{B}_{p_1'p_2'},x^{\epsilon})-\sum_{x^\epsilon\leqslant p_3'<p_2'<p_1'<x^{1/5}}S(\mathcal{B}_{p_1'p_2'p_3'},p_3'),
\end{aligned}$$

and hence by the linear sieve bounds we obtain

$$\Gamma_9\lesssim S(\mathcal{B},x^{1/2})\leqslant S(\mathcal{B},x^{1/5})\lesssim e^{-\gamma}\frac{G\bigl(\frac{1}{5}\bigr)}{\log x}\mathfrak{S}_2|\mathcal{B}|. \tag{6-19}$$

Now to compute $|\mathcal{B}|$ in (6-18), by Lemma 6.1 we have

$$|\mathcal{B}|\sim\frac{x}{\log x}\int_{\tau_1<t_1<t_2<t_3<\tau_2}\frac{\mathrm{d}t_1\,\mathrm{d}t_2\,\mathrm{d}t_3}{t_1t_2^2t_3}\,\omega\biggl(\frac{1-t_1-t_2-t_3}{t_2}\biggr)=\frac{x}{\log x}\cdot I_9.$$

Thus we have $\Gamma_9\lesssim e^{-\gamma}G\bigl(\frac{1}{5}\bigr)\Pi(x)I_9$. Similarly, we obtain

$$\Gamma_n\lesssim G\bigl(\tfrac{1}{5}\bigr)\frac{\Pi(x)}{e^\gamma}I_n\quad(\text{for } 9\leqslant n\leqslant 21). \tag{6-20}$$

Hence plugging (6-16), (6-20) into Lemma 6.2 completes the proof. □

Let

$$\begin{aligned}
\rho&=0.27195, & \tau_3&=0.24589,\\
\rho'&=0.12313, & \tau_2&=0.20867,\\
\epsilon&=0.002, & \tau_1&=0.16288.
\end{aligned}\tag{6-21}$$

For such choices of parameters, we compute the following integrals from Proposition 6.3,

$$\sum_{9\leqslant n\leqslant 21}I_n\leqslant 0.174404,\qquad\sum_{1\leqslant n\leqslant 8}G_n\leqslant 28.34581,\qquad G\bigl(\tfrac{1}{5}\bigr)\leqslant 5.99237.$$

Thus by Proposition 6.3, we obtain the bound

$$\pi_2(x) \lesssim S(\mathcal{A}, x^\rho) \lesssim \frac{\Pi(x)}{5e^\gamma}\biggl(\sum_{n=1}^{8} G_n + G\bigl(\tfrac{1}{5}\bigr)\sum_{n=9}^{21} I_n\biggr) \lesssim 3.30042\Pi(x). \tag{6-22}$$

We also record the individual bounds (see table at end of paper).

6.3. ***Completing the proof of Theorem 1.2.*** We shall refine our argument in certain cases for which Wu's iteration method [2004] applies directly without modification. As such we have chosen simplicity over full optimization.

In the lemma below, we consider the cases of sets $\mathcal{A}_{p_1 p_2}$, where p_1, p_2 lie in a prescribed range, and such that for all multiples $bp_1 p_2$, the sets $\mathcal{A}_{bp_1 p_2}$ are equidistributed to level x^θ (we also need level x^θ for corresponding switched sets $\mathcal{B}$ of integers $mbp_1 p_2 - 2$).

Lemma 6.4. *Let $\theta \in \bigl[\frac{1}{2}, 1\bigr]$, $s \in [2, 3]$, and $\mathcal{A} = \{p + 2 : p \leqslant x\}$. There is a function $H_\theta(s)$, monotonically increasing in θ for fixed s and decreasing in s for fixed θ, such that the following holds: For each $(D_1, D_2) \in \boldsymbol{D}_2^{\mathrm{well}}(x^\theta)$, we have*

$$\sum_{\substack{D_1 < p_1 < D_1^{1+\tau} \\ D_2 < p_2 < D_2^{1+\tau}}} S(\mathcal{A}_{p_1 p_2}, z) \lesssim \frac{\Pi(x)}{e^\gamma} \sum_{\substack{D_1 < p_1 < D_1^{1+\tau} \\ D_2 < p_2 < D_2^{1+\tau}}} \frac{\log x}{\varphi(p_1 p_2)\log z}\biggl(F(s) - \frac{2e^\gamma}{s} H_\theta(s)\biggr), \tag{6-23}$$

where $z^s = x^\theta / p_1 p_2$, provided (5-22) holds for $\lambda = \tilde{\lambda}^+$ at level $D = x^\theta$ with $(x^{\varepsilon_1}, x^{\varepsilon_2}) = (D_1, D_2)$, and provided for all vectors $(D_1, \ldots, D_r) \in \boldsymbol{D}_r^{\mathrm{well}}(x^\theta)$ extending (D_1, D_2),

$$\sum_{\substack{b = p_1 \cdots p_r \\ D_i < p_i \leqslant D_i^{1+\tau}}} \sum_{\substack{b \mid d, d \leqslant x^\theta \\ (d,a)=1}} \tilde{\lambda}^\pm(d)\biggl(\pi(x; d, a) - \frac{\pi(x)}{\varphi(d)}\biggr) \ll_{a,\varepsilon,A} \frac{x}{(\log x)^A}.$$

Proof. Wu iterates the weighted sieve inequality [Wu 2004, Lemmas 4.1 and 4.2] on the subset of (nonswitched) terms whose sieving parameter s lies in the interval $s \in [2, 3]$. This yields a recurrence relation for a function $H_\theta(s)$, which encodes the percent savings over the (normalized) linear sieve $sF(s)/(2e^\gamma)$.

Starting from a term $S(\mathcal{A}_{p_1 p_2}, z)$, each successive iteration of the weighted sieve inequality is composed of terms of the form $S(\mathcal{A}_{bp_1 p_2}, z')$ for some multiple $bp_1 p_2$ corresponding to some vector extending (D_1, D_2). By assumption, all such sets are equidistributed to level x^θ, when weighted by the upper/lower linear sieve. Similarly, the switched sets are also equidistributed to level x^θ (here we only need the upper bound weights $\tilde{\lambda}^+$). Finally, the savings function $H_\theta(s)$ inherits the stated monotonicity properties by construction of the iteration. □

The function H_θ depends on the known level of distribution x^θ (i.e., Wu used $\theta = \frac{1}{2}$ for Goldbach, and $\theta = \frac{4}{7}$ for twin primes). For parameters as in Tables 1 and 2 of [Wu 2004, pages 30–32],

$$H_{1/2}(t) \geqslant \begin{cases} 0.0223939 & \text{if } 2.0 < t \leqslant 2.2, \\ 0.0217196 & \text{if } 2.2 < t \leqslant 2.3, \\ 0.0202876 & \text{if } 2.3 < t \leqslant 2.4, \\ 0.0181433 & \text{if } 2.4 < t \leqslant 2.5, \\ 0.0158644 & \text{if } 2.5 < t \leqslant 2.6, \\ 0.0129923 & \text{if } 2.6 < t \leqslant 2.7, \\ 0.0100686 & \text{if } 2.7 < t \leqslant 2.8, \\ 0.0078162 & \text{if } 2.8 < t \leqslant 2.9, \\ 0.0072943 & \text{if } 2.9 < t \leqslant 3.0, \\ 0 & \text{else,} \end{cases}$$

and

$$H_{4/7}(t) \geqslant \begin{cases} 0.0287118 & \text{if } 2.0 \leqslant t \leqslant 2.1, \\ 0.0280509 & \text{if } 2.1 < t \leqslant 2.2, \\ 0.0264697 & \text{if } 2.2 < t \leqslant 2.3, \\ 0.0241936 & \text{if } 2.3 < t \leqslant 2.4, \\ 0.0214619 & \text{if } 2.4 < t \leqslant 2.5, \\ 0.0183875 & \text{if } 2.5 < t \leqslant 2.6, \\ 0.0149960 & \text{if } 2.6 < t \leqslant 2.7, \\ 0.0117724 & \text{if } 2.7 < t \leqslant 2.8, \\ 0.0094724 & \text{if } 2.8 < t \leqslant 2.9, \\ 0.0090024 & \text{if } 2.9 < t \leqslant 3.0, \\ 0 & \text{else.} \end{cases}$$

As such, Wu [2004, Theorem 3] obtained $\pi_2(x)/\Pi(x) \lesssim \frac{7}{2}(1 - H_{4/7}(2.1)) \leqslant 3.39951$.

To complete our proof of Theorem 1.2 we apply Lemma 6.4, now valid up to level $x^{7/12}$ by Corollary 2.7 and Proposition 5.5 for $\tilde{\lambda}^+$. Note when the largest integration variable is $t \geqslant \frac{1}{5}$, we have $\theta(t) = \frac{1}{3}(2-t) \leqslant \frac{7}{12}$ if and only if $t \geqslant \frac{1}{4}$. A key feature we use to satisfy the conditions of Lemma 6.4 is that the level $x^{\theta(t_1)}$ persists, since the largest prime $p_1 = x^{t_1}$ is preserved through successive iterations.

Thus, in practice, Lemma 6.4 simply amounts to modifying the integral in G_2 by substituting $F(s) - \frac{2e^\gamma}{s} H_\theta(s)$ in for $F(s)$, $s = (\theta(t) - t - u)/u$, when $t \geqslant \frac{1}{4}$ (the only parameter $\geqslant \frac{1}{4}$ is ρ, so we only refine G_2). Denote this as G_2^{Wu}. For ease we also use $H_\theta(s) \geqslant H_{4/7}(s)$, by monotonicity in θ. Doing so, with the same parameter choices (6-21), we obtain $G_2^{\mathrm{Wu}} \leqslant -5.598667$ and hence

$$\pi_2(x) \lesssim \frac{\Pi(x)}{5e^\gamma}\biggl(G_2^{\mathrm{Wu}} + \sum_{1 \leqslant n \neq 2 \leqslant 8} G_n + G\bigl(\tfrac{1}{5}\bigr) \sum_{9 \leqslant n \leqslant 21} I_n\biggr) \lesssim 3.299552\Pi(x). \tag{6-24}$$

This completes the proof of Theorem 1.2.

n	G_n	n	I_n	n	I_n
1	39.00163	9	0.0332157	17	0.000315
2	-5.591009	10	0.0228322	18	0.000269
3	-3.986553	11	0.0092564	19	0.000164
4	-5.060499	12	0.0150101	20	$\leqslant 2.70 \cdot 10^{-6}$
5	1.864133	13	0.0547244	21	$\leqslant 5.50 \cdot 10^{-9}$
6	0.741181	14	0.0260202		
7	0.453663	15	0.0124636		
8	0.923736	16	0.0001314		

Table 2. Values of $G_1, \dots, G_8$ and $I_9, \dots, I_{21}$.

For slight numerical gains, one may compute $H_\theta(s)$ when $\theta \in \left[\frac{4}{7}, \frac{7}{12}\right]$, by tweaking the formulae in [Wu 2004]. More substantially, Wu defined a lower bound savings $h_\theta(s)$, for a substitution of $f(s)$ by $f(s) + \frac{2e^\gamma}{s} h_\theta(s)$. But in practice, to compute h would require derivations (analogous to H) of as yet undetermined formulae. We leave these to the reader.

Acknowledgments

The author is grateful to James Maynard for suggesting the problem and for many valuable discussions. The author also thanks Carl Pomerance, Kyle Pratt, and the anonymous referee for helpful feedback. The author was supported by a Clarendon Scholarship at the University of Oxford.

References

[Bombieri and Davenport 1966] E. Bombieri and H. Davenport, "Small differences between prime numbers", *Proc. Roy. Soc. Lond. Ser. A* **293** (1966), 1–18. MR Zbl

[Bombieri et al. 1986] E. Bombieri, J. B. Friedlander, and H. Iwaniec, "Primes in arithmetic progressions to large moduli", *Acta Math.* **156**:3-4 (1986), 203–251. MR Zbl

[Brun 1919] V. Brun, "La série $1/5+1/7+1/11+1/13+1/17+1/19+1/29+1/31+1/41+1/43+1/59+1/61+\cdots$ où les dénominateurs sont nombres premiers jumeaux est convergente ou finie", *Bull. Sci. Math.* **43** (1919), 100–104, 124–128. Zbl

[Cai and Lu 2003] Y. Cai and M. Lu, "On the upper bound for $\pi_2(x)$", *Acta Arith.* **110**:3 (2003), 275–298. MR Zbl

[Chen 1973] J. R. Chen, "On the representation of a larger even integer as the sum of a prime and the product of at most two primes", *Sci. Sinica* **16** (1973), 157–176. MR Zbl

[Chen 1978] J. R. Chen, "On the Goldbach's problem and the sieve methods", *Sci. Sinica* **21**:6 (1978), 701–739. MR Zbl

[Elliott and Halberstam 1970] P. D. T. A. Elliott and H. Halberstam, "A conjecture in prime number theory", pp. 59–72 in *Symposia Mathematica*, *IV* (Rome, 1968/1969), Academic Press, London, 1970. MR Zbl

[Fouvry 1984] É. Fouvry, "Autour du théorème de Bombieri–Vinogradov", *Acta Math.* **152**:3-4 (1984), 219–244. MR Zbl

[Fouvry and Grupp 1986] É. Fouvry and F. Grupp, "On the switching principle in sieve theory", *J. Reine Angew. Math.* **370** (1986), 101–126. MR Zbl

[Fouvry and Iwaniec 1983] É. Fouvry and H. Iwaniec, "Primes in arithmetic progressions", *Acta Arith.* **42**:2 (1983), 197–218. MR Zbl

[Friedlander and Iwaniec 2010] J. Friedlander and H. Iwaniec, *Opera de cribro*, Amer. Math. Soc. Colloq. Publ. **57**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[Hardy and Littlewood 1923] G. H. Hardy and J. E. Littlewood, “Some problems of ‘partitio numerorum’, III: On the expression of a number as a sum of primes”, *Acta Math.* **44**:1 (1923), 1–70. MR Zbl

[Iwaniec 1980] H. Iwaniec, “A new form of the error term in the linear sieve”, *Acta Arith.* **37** (1980), 307–320. MR Zbl

[Maynard 2020] J. Maynard, “Primes in arithmetic progressions to large moduli, II: Well-factorable estimates”, preprint, 2020. arXiv 2006.07088

[Pan 1964] C.-d. Pan, “A new application of the Yu. V. Linnik large sieve method”, *Acta Math. Sinica* **14** (1964), 597–606. In Chinese, translated in *Chinese Math. Acta* **5** (1964), 642–652. MR Zbl

[Riesel and Vaughan 1983] H. Riesel and R. C. Vaughan, “On sums of primes”, *Ark. Mat.* **21**:1 (1983), 46–74. MR Zbl

[Selberg 1952] A. Selberg, “On elementary methods in prime number theory and their limitations”, pp. 13–22 in *Den* 11*te Skandinaviske Matematikerkongress* (Trondheim, Norway, 1949), Tanum, Oslo, 1952. MR Zbl

[Siebert 1976] H. Siebert, “Montgomery’s weighted sieve for dimension two”, *Monatsh. Math.* **82**:4 (1976), 327–336. MR Zbl

[Wu 1990] J. Wu, “Sur la suite des nombres premiers jumeaux”, *Acta Arith.* **55**:4 (1990), 365–394. MR Zbl

[Wu 2004] J. Wu, “Chen’s double sieve, Goldbach’s conjecture and the twin prime problem”, *Acta Arith.* **114**:3 (2004), 215–273. MR Zbl

Communicated by Andrew Granville
Received 2022-05-08 Revised 2024-01-09 Accepted 2024-02-13

jared.d.lichtman@gmail.com *Mathematical Institute, University of Oxford, Oxford, United Kingdom*

ALGEBRA AND NUMBER THEORY 19:1 (2025)
https://doi.org/10.2140/ant.2025.19.39

Ranks of abelian varieties in cyclotomic twist families

Ari Shnidman and Ariel Weiss

Let A be an abelian variety over a number field F, and suppose that $\mathbb{Z}[\zeta_n]$ embeds in $\operatorname{End}_{\bar{F}} A$, for some root of unity ζ_n of order $n = 3^m$. Assuming that the Galois action on the finite group $A[1-\zeta_n]$ is sufficiently reducible, we bound the average rank of the Mordell–Weil groups $A_d(F)$, as A_d varies through the family of μ_{2n}-twists of A. Combining this result with the recently proved uniform Mordell–Lang conjecture, we prove near-uniform bounds for the number of rational points in twist families of bicyclic trigonal curves $y^3 = f(x^2)$, as well as in twist families of theta divisors of cyclic trigonal curves $y^3 = f(x)$. Our main technical result is the determination of the average size of a 3-isogeny Selmer group in a family of μ_{2n}-twists.

1. Introduction

Let A be an abelian variety over a number field F and let $G_F = \operatorname{Gal}(\bar{F}/F)$. Any G_F-stable subgroup $H \subset \operatorname{Aut}_{\bar{F}} A$ gives rise to a twist family of abelian varieties A_ξ over F, indexed by the elements ξ of the Galois cohomology set $H^1(G_F, H)$. The base change of A_ξ to $\bar{F}$ is isomorphic to $A_{\bar{F}}$, but with G_F-action twisted by ξ. Our goal is to study the distributions of the ranks of the Mordell–Weil groups $A_\xi(F)$ in such twist families, and to give some applications.

Every abelian variety A has the automorphism -1, and since $H^1(G_F, \{\pm 1\}) \simeq F^\times/F^{\times 2}$, we obtain the *quadratic twist family* of A. The average rank of $A_\xi(F)$ in quadratic twist families has been extensively studied in the case of elliptic curves [Brumer 1992; Heath-Brown 1994; Katz and Sarnak 1999; Smith 2017; Bhargava et al. 2019], with [Bhargava et al. 2019] addressing many cases in higher dimension as well.

In this paper, we consider the case $H = \mu_{2n}$, the group of $2n$-th roots of unity, where $n = 3^m$ for some $m \geq 1$. More precisely, we assume that there is a G_F-equivariant ring embedding $\mathbb{Z}[\zeta] \hookrightarrow \operatorname{End}_{\bar{F}} A$, where $\zeta = \zeta_n \in \bar{F}$ is a root of unity of order n. We say that such an A has *ζ-multiplication*. For example, the Jacobian J of a curve of the form $y^3 = xf(x^{3^{m-1}})$ has ζ-multiplication induced from the order n automorphism $(x, y) \mapsto (\zeta^3 x, \zeta y)$.

Since $\mu_{2n} = \langle -\zeta \rangle \subset \operatorname{Aut}_{\bar{F}} A$, there is a twist A_d for each $d \in F^\times/F^{\times 2n} \simeq H^1(G_F, \mu_{2n})$. In the example above, J_d is the d-th quadratic twist of the Jacobian of $y^3 = xf\left(\frac{1}{d}x^{3^{m-1}}\right)$. In Section 5B, we recall a height function $h : F^\times/F^{\times 2n} \to \mathbb{R}$ with the property that the sets $\Sigma_X := \{d \in F^\times/F^{\times 2n} : h(d) < X\}$ are finite. When $F = \mathbb{Q}$, the height $h(d)$ is the absolute value of the smallest integer representing d. The average rank of $A_d(F)$ is then, by definition,

$$\operatorname{avg}_d \operatorname{rk} A_d(F) = \lim_{X\to\infty} \operatorname*{avg}_{d\in\Sigma_X} \operatorname{rk} A_d(F).$$

MSC2020: primary 11G10; secondary 11E76, 11S25, 14G05.
Keywords: arithmetic statistics, twist families, rational points on curves, ranks of abelian varieties.

In general, it is not known whether this limit exists or even if the limsup is finite. In the latter case, we say that the *average rank of $A_d(F)$ is bounded.*

1A. ***Mordell–Weil ranks.*** If A has ζ-multiplication, the endomorphism $1-\zeta \in \operatorname{End}_{\bar{F}} A$ descends to an isogeny $\pi : A \to A'$ over F (see Section 2). The kernel $A[\pi]$ is a G_F-stable subgroup of the 3-torsion group $A[3]$, and hence is a finite-dimensional $\mathbb{F}_3$-vector space. We show that the average rank of $A_d(F)$ is bounded, assuming that the G_F-action on $A[\pi]$ is sufficiently reducible.

Theorem 1.1. *Let A be an abelian variety with ζ_{3^m}-multiplication over F.*

(i) *If $A[\pi]$ is a direct sum of characters, then* $\operatorname{avg}_d \operatorname{rk} A_d(F)$ *is bounded.*

(ii) *If $A[\pi]$ has a full flag, then* $\operatorname{avg}_d \operatorname{rk} A_d(F)$, *over squarefree $d \in F^\times/F^{\times 2n}$, is bounded.*

Here, we say that $A[\pi]$ *has a full flag* if there are G_F-modules $0 = H_0 \subset H_1 \subset \cdots \subset H_k = A[\pi]$ such that $\dim_{\mathbb{F}_3} H_{i+1}/H_i = 1$. We say that $d \in F^\times/F^{\times 2n}$ is *squarefree* if $v(d) \equiv 0$ or $1 \pmod{2n}$, for all finite places v of F. Theorem 1.1 is a simultaneous generalization of [Bhargava et al. 2019, Theorem 2.2] and [Bhargava et al. 2020, Theorem 5] to a larger class of twist families of abelian varieties.

If J is the Jacobian of $y^3 = xf(x^{3^{m-1}})$, then the representation theoretic conditions on $J[\pi]$ translate into conditions on the Galois group $\operatorname{Gal}(f)$ of the splitting field of $f(x)$ over F. More generally, we deduce the following result from Theorem 1.1:

Corollary 1.2. *Let $f(x) \in F[x]$ be separable and nonconstant, and let J be the Jacobian of either $y^{3^m} = f(x)$ or $y^3 = xf(x^{3^{m-1}})$.*

(i) *If* $\operatorname{Gal}(f) \simeq (\mathbb{Z}/2\mathbb{Z})^k$, *for some $k \geq 0$, then* $\operatorname{avg}_d \operatorname{rk} J_d(F)$ *is bounded.*

(ii) *If* $\operatorname{Gal}(f)$ *is an extension of* $(\mathbb{Z}/2\mathbb{Z})^k$ *by a* 3*-group, then* $\operatorname{avg}_d \operatorname{rk} J_d(F)$, *over squarefree $d \in F^\times/F^{\times 2n}$, is bounded.*

Our proof of Theorem 1.1 gives an explicit upper bound on the average rank, however, the bound depends on subtle arithmetic properties of A. The following crude upper bound has the virtue that it applies to a large class of abelian varieties and depends only mildly on A.

Theorem 1.3. *Suppose that $A[\pi]$ has a full flag and that A admits a ζ_n-stable principal polarization. Let S be the set of places of F dividing $3\mathfrak{f}_A\infty$, where $\mathfrak{f}_A$ is the conductor of A. Then the average rank of $A_d(F)$, for squarefree $d \in F^\times/F^{\times 2n}$, is at most* $\dim A \cdot (\#S + 3^{-\#S})$.

For most A, this bound is significantly weaker than what our method actually gives. An interesting case is when A has complex multiplication (CM), i.e., $\dim A = 3^{m-1}$, in which case $\dim_{\mathbb{F}_3} A[\pi] = 1$ and the reducibility hypotheses are automatically satisfied. When the complex multiplication is defined over F, we obtain especially strong results:

Theorem 1.4. *Suppose that* $\dim A = 3^{m-1}$, *so that A has complex multiplication by $\mathbb{Z}[\zeta_{3^m}]$. Assume moreover that $\zeta_{3^m} \in F$, so that the complex multiplication is defined over F. Then the average $\mathbb{Z}[\zeta_{3^m}]$-module rank of $A_d(F)$ is at most $\frac{1}{2}$, and at least* 50% *of twists A_d have rank* 0.

In the CM case, we expect that 100% of twists A_d have rank 0, in which case our result is halfway towards the analogue of Goldfeld's conjecture in this context.

Such A arise as factors of the Jacobians of the curves $y^{3^m} = x^a(1-x)^b$, which have good reduction away from 3. Even when the CM is not defined over F, we obtain average rank bounds that depend only on $\dim A$. Over $\mathbb{Q}$, for example, the average rank of $A_d(\mathbb{Q})$ is at most $\frac{19}{9}\dim A$ by Theorem 1.3, a bound which can be improved to $\frac{13}{9}\dim A$ with a more refined analysis.

1B. ***Rational points on curves.*** Theorem 1.1 has concrete consequences for the arithmetic of curves C/F of genus $g \geq 2$. It was nearly 40 years ago that Faltings proved that $C(F)$ is a finite set, but very recently, there has been significant progress towards a uniform upper bound on $\#C(F)$. Building on work of Dimitrov, Gao, and Habegger [Dimitrov et al. 2021], Kühne [2021] has shown that

$$\#C(F) \leq c_g^{1+\operatorname{rk}\operatorname{Jac}(F)},$$

where c_g is a constant depending only g. Building on this work, Gao, Ge, and Kühne [Gao et al. 2021] proved the more general uniform Mordell–Lang conjecture for closed subvarieties of abelian varieties. These results reduce the question of uniform bounds for rational points on a large class of varieties to a question about ranks of abelian varieties.

By combining these results with Theorem 1.1, we show that "near-uniformity" holds for twists of bicyclic trigonal curves:

Theorem 1.5. *Let $f(x) \in F[x]$ be separable, of degree at least two, and with all of its roots nonzero elements of F. Consider the bicyclic trigonal curve $C : y^3 = f(x^2)$, and let $C_d : dy^3 = f(dx^2)$ be the corresponding sextic twist family. Then for every $\varepsilon > 0$, there is a constant N_ε such that the lower density of classes $d \in F^\times/F^{\times 6}$ for which*

$$\#C_d(F) \leq N_\varepsilon$$

is at least $1-\varepsilon$.

To prove Theorem 1.5, we apply Theorem 1.1 not to the Jacobian of C, but to the Prym variety for the double cover $C \to C'$, where $C' : y^3 = f(x)$; see Section 9. We remark that the constant N_ε depends only on ε, $\deg(f)$, and $\#S$ (using the notation of Theorem 1.3).

Unlike the curves in Theorem 1.5, a general cyclic trigonal curve $C : y^3 = f(x)$ has no sextic twists, so it may seem that Theorem 1.1 says nothing about rational points on the twists of C itself. However, for these curves, we can consider sextic twists of a theta divisor $\Theta \subset J = \operatorname{Jac}(C)$. Recall that Θ is birational to the symmetric power $C^{(g-1)}$, so its rational points parametrize low-degree points on C. We can choose Θ so that it is preserved by the μ_{2n}-action (see Section 9), which allows us to consider the twist $\Theta_d \subset J_d$, for each $d \in F^\times$.

Theorem 1.6. *Let $f(x) \in F[x]$ be separable and suppose that* $\operatorname{Gal}(f) \simeq (\mathbb{Z}/2\mathbb{Z})^k$ *for some $k \geq 0$. Let $C : y^3 = f(x)$, and suppose that* $\operatorname{Jac}(C)$ *is geometrically simple. Let $\Theta \subset$* $\operatorname{Jac}(C)$ *be a symmetric theta* *divisor. Then for every $\varepsilon > 0$, there is a constant N_ε such that the lower density of classes $d \in F^\times/F^{\times 6}$ for which*

$$\#\Theta_d(F) \leq N_\varepsilon$$

is at least $1 - \varepsilon$.

This result again follows from Theorem 1.1 and [Gao et al. 2021], and N_ε depends only on ε, $\deg(f)$, and $\#S$. Since the results of [Gao et al. 2021] are ineffective, we cannot say anything explicit about the constant N_ε in general. However, one can prove explicit results in this direction by instead combining our work with the Chabauty method. We illustrate this by way of an example.

Theorem 1.7. *Consider the sextic twist family $C_d : y^3 = (x^2 - d)(x^2 - 4d)$ of genus-3 curves. For at least $\frac{1}{3}$ of squarefree $d \in \mathbb{Z}$ such that $d \equiv 2$ or $11 \pmod{36}$, we have $\#C_d(\mathbb{Q}) \leq 5$.*

The curve C_d admits a double cover $p : C_d \to E_d$ to the elliptic curve $E_d : y^3 = (x - d)(x - 4d)$. Moreover, C_d embeds in the abelian surface $P_d = \operatorname{Jac}(C_d)/p^* \operatorname{Pic}^0(E_d)$. By making the rank bound in Theorem 1.1 explicit, we show that $\operatorname{rk} P_d(\mathbb{Q}) \leq 1$ for at least $\frac{1}{3}$ of twists d. Then we invoke, and generalize slightly, Stoll's uniform Chabauty result for twist families [2006].

The same method works for sextic twist families of the form $C_{a,d} : y^3 = (x^2 - d)(x^2 - ad)$. To prove the existence of twists with $\operatorname{rk} P_{a,d}(\mathbb{Q}) \leq 1$, we must check that a certain local 3-adic root number takes the value -1 for some twist d. We can verify this condition in Magma for seemingly any given curve $C_{a,d}$, but it would be nice to have a proof for all or most values of a.[1] It would also be interesting to prove explicit results for symmetric squares of trigonal plane quartics, as in Theorem 1.6, by using [Caro and Pasten 2023].

1C. *3-isogeny Selmer groups.*

Having discussed some applications of Theorem 1.1, let us discuss its proof. Theorem 1.1 follows from a more precise result about Selmer groups. Let A/F be an abelian variety with ζ-multiplication and admitting a 3-isogeny $\phi : A \to B$. If $A[\phi] \subset A[\pi]$, or equivalently, if ϕ is ζ-linear, then each twist A_d is endowed with its own 3-isogeny $\phi_d : A_d \to B_d$. For each d, we consider the ϕ_d-Selmer group $\operatorname{Sel}_{\phi_d}(A_d)$, which sits in the exact sequence

$$0 \to B_d(F)/\phi_d A_d(F) \to \operatorname{Sel}_{\phi_d}(A_d) \to Ш(A_d)[\phi_d] \to 0.$$

The main technical result of this paper is the exact computation of $\operatorname{avg}_d \# \operatorname{Sel}_{\phi_d}(A_d)$.

To state the precise result, we recall the global Selmer ratio $c(\phi_d) = \prod_v c_v(\phi_d)$, where for each place v of F, we define

$$c_v(\phi_d) = \frac{\#\operatorname{coker}(A_d(F_v) \to B_d(F_v))}{\#\ker(A_d(F_v) \to B_d(F_v))}.$$

For $v \nmid 3\infty$, we have $c_v(\phi_d) = c_v(B_d)/c_v(A_d)$, where $c_v(A)$ is the Tamagawa number of A over F_v. Thus, up to some subtle factors at places v above 3 and ∞, the number $c(\phi_d)$ is the ratio of the global Tamagawa

[1] In [Shnidman and Weiss 2023, Theorem 1.4], we prove that a positive proportion of $P_{a,d}$ have rank at most 1 in the case that a is a square, using a different argument which sidesteps the root number question.

numbers $c(B_d)/c(A_d)$. In particular, we have $c_v(\phi_d) \in 3^{\mathbb{Z}}$, and $c_v(\phi_d) = 1$ for all but finitely many v (having fixed d).

We say $A[\phi]$ is *almost everywhere locally a direct summand* of $A[\pi]$ if, for almost all places v of F, the G_{F_v}-module $A[\phi]$ is a direct summand of $A[\pi]$.

Theorem 1.8. *Assume that $A[\phi]$ is almost everywhere locally a direct summand of $A[\pi]$. Then $\mathrm{avg}_d \, \# \mathrm{Sel}_{\phi_d}(A_d) = 1 + \mathrm{avg}_d \, c(\phi_d)$, where both averages are finite and taken over $d \in F^\times / F^{\times 2n}$, ordered by height.*

This result is a simultaneous generalization of [Bhargava et al. 2019, Theorem 2.1] and [Bhargava et al. 2020, Theorem 1]. Interestingly, the condition of being everywhere locally a direct summand, which is automatically satisfied for the families considered in [Bhargava et al. 2019; 2020], seems to be an obstruction to computing the average size of $\# \mathrm{Sel}_{\phi_d}(A_d)$ in the entire family of twists, at least using our methods. In any case, if we only consider squarefree twists, then this obstruction goes away:

Theorem 1.9. *Let $\phi : A \to B$ be a ζ-linear 3-isogeny. Then the average size of $\# \mathrm{Sel}_{\phi_d}(A_d)$ over squarefree $d \in F^\times / F^{\times 2n}$ is finite and equal to $1 + \mathrm{avg}_d \, c(\phi_d)$.*

The quantity $\mathrm{avg}_d \, c(\phi_d)$ is governed by local arithmetic data which can be made explicit in certain cases. For example, in Theorem 1.4 we have $c(\pi_d) = c(\phi_d) = 1$ for all d. However, in general, computing the exact value of $\mathrm{avg}_d \, c(\phi_d)$ is hard. Nonetheless, one can give an explicit upper bound on $\mathrm{avg}_d \, \# \mathrm{Sel}_{\phi_d}(A_d)$ depending only on F, $\dim A$, and the number of primes dividing the conductor of A (Proposition 5.5).

In Section 6 we show how to deduce Theorem 1.1 from Theorems 1.8 and 1.9. In the remainder of the introduction, we discuss the proofs of the latter two results.

1D. ***Methods.*** We prove Theorems 1.8 and 1.9 using geometry-of-numbers methods. As in the previous works [Bhargava et al. 2019; 2020] of the first author and his collaborators, we first identify the elements of $\mathrm{Sel}_{\phi_d}(A_d)$ with $\mathrm{SL}_2(F)$-orbits of binary cubic forms of discriminant d. We then wish to use lattice-point counting techniques, which have been extended to global fields in [Bhargava et al. 2015], to count the number of such orbits of bounded discriminant. However, it is not at all clear (and indeed, it is not always true) that the $\mathrm{SL}_2(F)$-orbits corresponding to Selmer elements are *integral*, i.e., that they contain cubic forms whose coefficients are algebraic integers. This integrality is of course essential for lattice-point counting.

For the quadratic twist families considered in [Bhargava et al. 2019], integrality follows quickly once it is realized that the local Selmer conditions are very mild outside the finitely many primes dividing the conductor of A. One needs only to "clear denominators" at those finitely many primes, and then the Selmer orbits become integral. For the families considered in this paper, the question of integrality is more subtle. The first interesting case is the family of sextic twists $E : y^2 = x^3 + d$ considered in [Bhargava et al. 2020], which is the unique twist family of elliptic curves with 3-power ζ-multiplication. In that special family, the authors give an *explicit* bijection between Selmer elements and binary cubic forms. Integrality is then proven using a direct connection with Bhargava's higher composition laws.

In the more general setup of this paper, we cannot rely on such an explicit parametrization, nor do we expect one to exist. Instead, our method is more abstract and involves two independent steps. First, we give a complete analysis of the *integral* arithmetic invariant theory for the representation $\mathrm{Sym}^3 \mathbb{Z}^2$ of SL_2, in the sense that we identify precisely which $\mathrm{SL}_2(F_v)$-orbits of binary cubic forms, over a local field F_v, have integral representatives. We show that once the valuation of the discriminant $v(d)$ is at least 3, all orbits have integral representatives, and we determine what happens for $v(d) \leq 2$ as well. Our strategy is to translate the question into one about cubic rings, whose local structure we understand well. Second, we study the Selmer groups $\mathrm{Sel}_{\phi_d}(A_d)$ from a purely cohomological point of view, in the spirit of Mazur and Rubin [2007]; see also [Klagsbrun et al. 2013]. Upon comparing the results, we find that for all but finitely many primes v, the Selmer orbits of discriminant d are v-integral, except possibly when $v(d) = 2$. In particular, the Selmer orbits are integral when d is squarefree. When $v(d) = 2$, we find that the local direct summand condition on $A[\phi] \subset A[\pi]$ exactly matches up with the local integrality condition.

1E. ***Future directions.*** In many situations, the integral orbits of a reductive group G acting on a representation V have been shown to parametrize Selmer elements in a certain explicit family of abelian varieties [Bhargava and Shankar 2015a; 2015b; Bhargava and Ho 2016; Bhargava and Gross 2014; Thorne 2013; Laga 2023]. The results of this paper show that there is a tremendous amount of flexibility in these constructions, in the sense that (G, V) can be used to parametrize Selmer elements in very different looking families over the same space of invariants $V /\!\!/ G$ (which is $\mathbb{A}^1$ in our case). Our analysis of the integral arithmetic invariant theory of $(\mathrm{SL}_2, \mathrm{Sym}^3 \mathbb{Z}^2)$ can be adapted to some of these other representations (G, V), so it would be interesting to understand the following (vaguely formulated) question. For which families $\phi : \mathcal{A} \to \mathcal{B}$ of isogenies of abelian varieties over $S = V /\!\!/ G$ can the elements of $\mathrm{Sel}_{\phi_s}(\mathcal{A}_s)$, for $s \in S(F)$, be parametrized by orbits of $G(F)$ on $V(F)$?

1F. ***Outline.*** We begin in Section 2 with basics on abelian varieties with ζ-multiplication. Sections 3–6 are the technical heart of the paper. In Section 3, we give a complete analysis of the integral arithmetic theory for the representation $\mathrm{Sym}^3 \mathbb{Z}^2$ of SL_2. In Section 4, we give a parallel, but independent analysis of the Selmer groups $\mathrm{Sel}_{\phi_d}(A_d)$ of the twists ϕ_d of a general ζ-linear 3-isogeny. In Section 5, we combine these two sections and prove Theorems 1.8 and 1.9. In Section 6, we apply the results of Section 5 to prove Theorems 1.1 and 1.3.

The remainder of the paper is devoted to applications of our main results. In Section 7, we study the average ranks of the Jacobians of the curves $y^3 = xf(x^{3^{m-1}})$ and $y^{3^m} = f(x)$ and prove Corollary 1.2. In Section 8, we give explicit results for abelian varieties with CM and prove Theorem 1.4. In Section 9, we study rational points in twist families of curves as in Section 1B, and prove Theorems 1.5 and 1.6. Finally, in Section 10, we study twist families of genus-3 curves and prove Theorem 1.7.

2. Abelian varieties with ζ-multiplication

Let F be a field of characteristic 0. Fix an odd prime p, an integer $n = p^m$, and a primitive n-th root of unity $\zeta = \zeta_n$. In this section, φ denotes Euler's totient function.

Definition 2.1. An abelian variety with ζ-multiplication is a pair (A, ι_A), where A is an abelian variety over F and $\iota_A : \mathbb{Z}[\zeta] \hookrightarrow \operatorname{End}_{\overline{F}} A$ is a G_F-equivariant injective ring homomorphism.

We usually suppress any mention of ι_A and view $\mathbb{Z}[\zeta]$ as a subring of $\operatorname{End}_{\overline{F}} A$. In this section, we collect some basic facts and constructions relating to abelian varieties with ζ-multiplication.

2A. ***The isogeny π.*** If A is an abelian variety with ζ-multiplication, then since $1-\zeta$ divides p in $\mathbb{Z}[\zeta]$, the map $1-\zeta \in \operatorname{End}_{\overline{F}} A$ is an isogeny whose degree is a power of p.

Lemma 2.2. *The kernel of $1-\zeta$ is G_F-stable, and hence is an F-subgroup of $A[p]$. In particular, there is an abelian variety $A^{(1)}$ over F, such that the endomorphism $1-\zeta$ of A over $\overline{F}$ descends to an isogeny $\pi : A \to A^{(1)}$ over F.*

Proof. If $P \in A_{\overline{F}}[1-\zeta]$ and $\sigma \in G_F$, then $\zeta^{\sigma^{-1}} = \zeta^i$ for some $i \in (\mathbb{Z}/n\mathbb{Z})^\times$ and

$$\zeta(P^\sigma) = (\zeta^{\sigma^{-1}}(P))^\sigma = (\zeta^i P)^\sigma = P^\sigma,$$

which shows that $P^\sigma \in A_{\overline{F}}[1-\zeta]$. Hence, $A_{\overline{F}}[1-\zeta]$ descends to an F-subgroup H of A. Thus, we obtain a an isogeny $\pi : A \to A/H =: A^{(1)}$ over F.

The equality of ideals $(1-\zeta)^{\varphi(n)} = (p)$ in $\mathbb{Z}[\zeta]$ shows that $A_{\overline{F}}[1-\zeta] \subset A_{\overline{F}}[p]$. □

If $\zeta \in F$, then $A^{(1)} = A/A[1-\zeta] \simeq A$, and $\pi : A \to A^{(1)}$ can be identified with the endomorphism $1-\zeta$. If $\zeta \notin F$, then $A^{(1)} = A/A[\pi]$ is a twist of A which we now identify:

Lemma 2.3. *$A^{(1)}$ is the twist of A corresponding to the cocycle $\sigma \mapsto \frac{1-\zeta^\sigma}{1-\zeta} \in H^1(F, \mathbb{Z}[\zeta]^\times)$.*

Proof. Over $F(\zeta)$, the map $\eta : A/A[1-\zeta] \to A$ given by $\bar{x} \mapsto (1-\zeta)x$ defines an isomorphism. Hence, $A^{(1)}$ is the twist corresponding to the cocycle $\sigma \mapsto \eta^\sigma \eta^{-1}$. □

The abelian variety $A^{(1)}$ also has ζ-multiplication. Iterating Lemma 2.2, for each integer s, we obtain an abelian variety $A^{(s)} = A/A[(1-\zeta)^s]$. As in Lemma 2.3, $A^{(s)}$ is isomorphic to the twist of A corresponding to the cocycle

$$\sigma \mapsto \frac{(1-\zeta^\sigma)^s}{(1-\zeta)^s} \in H^1(F, \mathbb{Z}[\zeta]^\times).$$

We define π^s to be the corresponding isogeny $A = A^{(0)} \to A^{(s)}$.

Note that $A^{(\varphi(n))} \simeq A$ and that $\pi^{\varphi(n)} : A \to A$ is multiplication by pu, for some unit $u \in \operatorname{Aut} A$. In particular, when writing $A^{(s)}$, we can always consider s modulo $\varphi(n) = p^{m-1}(p-1)$, and we have inclusions

$$A[\pi] \subset A[\pi^2] \subset \cdots \subset A[\pi^{\varphi(n)-1}] \subset A[\pi^{\varphi(n)}] = A[p].$$

In general, the Galois action on $A^{(s)}$ is related to the Galois action on A in a convoluted way. However, on a subset of the torsion of A, the action is especially simple.

Lemma 2.4. *Let $s = p^r$ for some $0 \le r < m$, and let $i \in \mathbb{Z}$. Then, as G_F-modules,*

$$A^{(is)}[\pi^s] \simeq A[\pi^s] \otimes \chi_p^i,$$

where χ_p is the mod p cyclotomic character.

Proof. By Lemma 2.3, the abelian variety $A^{(is)}$ is the twist of A corresponding to the cocycle

$$\sigma \mapsto \frac{(1-\zeta^\sigma)^{is}}{(1-\zeta)^{is}} \in H^1(F, \mathbb{Z}[\zeta]^\times).$$

Equivalently, there is an isomorphism $\phi : A \to A^{(is)}$ over $F(\zeta)$ such that for all $\sigma \in G_F$ and $P \in A(\bar{F})$, we have

$$(\phi^{-1})^\sigma \circ \phi(P) = \frac{(1-\zeta^\sigma)^{is}}{(1-\zeta)^{is}} P.$$

Hence, if $P \in A[\pi^s]$ and $\sigma \in G_F$, then in $A^{(is)}[\pi^s]$, we have

$$(\phi(P))^\sigma = \frac{(1-\zeta^\sigma)^{is}}{(1-\zeta)^{is}} \phi(P^\sigma).$$

Suppose that $\sigma : \zeta \mapsto \zeta^j$ for some $j \in (\mathbb{Z}/n\mathbb{Z})^\times$. Then, since $(1-\zeta^{p^r})\phi(P^\sigma) = (1-\zeta)^{p^r}\phi(P^\sigma) = 0$, we have

$$\begin{aligned} \frac{(1-\zeta^\sigma)^{ip^r}}{(1-\zeta)^{ip^r}} \phi(P^\sigma) &= (1+\zeta+\zeta^2+\cdots+\zeta^{s-1})^{ip^r}\phi(P^\sigma) \\ &= (1+\zeta^{p^r}+\zeta^{2p^r}+\cdots+\zeta^{p^r(j-1)})^i\phi(P^\sigma) \\ &= j^i\phi(P^\sigma). \end{aligned}$$

Since, by definition, $j \pmod p = \chi_p(\sigma)$, we see that $(\phi(P))^\sigma = \chi_p(\sigma)^i\phi(P^\sigma)$, as claimed. □

2B. *ζ-linear isogenies.* We keep the notation $n = p^m$ and $\zeta = \zeta_n$.

Definition 2.5. Let (A, ι_A) and (B, ι_B) be abelian varieties over F with ζ-multiplication, and let $\phi : A \to B$ be an isogeny. We say that ϕ is *ζ-linear* if $\iota_B(\alpha) \circ \phi = \phi \circ \iota_A(\alpha)$ for all $\alpha \in \mathbb{Z}[\zeta]$.

Lemma 2.6. *If $\phi : A \to B$ is a ζ-linear p-isogeny, then $A[\phi] \subset A[\pi]$. Conversely, if $H \subset A[\pi]$ is a G_F-stable subgroup of order p, then the quotient $B := A/H$ inherits a ζ-multiplication from A, and the canonical p-isogeny $\phi : A \to B$ is ζ-linear.*

Proof. Since ϕ is ζ-linear, if $P \in A[\phi]$, then so is ζP. Hence, the action of ζ is given by a homomorphism $\mu_n = \mu_{p^m} \to \operatorname{Aut}_{\bar{F}} A[\phi] \simeq (\mathbb{Z}/p\mathbb{Z})^\times$, which must of course be trivial. Thus, ζ acts as the identity on $A[\phi]$, so $A[\phi] \subset A[1-\zeta] = A[\pi]$.

For the converse, since $\iota_A(\zeta)$ fixes H, we have $\ker(\phi) = \ker(\phi \circ \iota_A(\zeta))$, so that $\phi \circ \iota_A(\zeta)$ factors through ϕ. That is, there exists an automorphism $\zeta_B : B \to B$ such that $\phi \circ \iota_A(\zeta) = \zeta_B \circ \phi$. This automorphism ζ_B has order n and has the same minimal polynomial as $\iota_A(\zeta)$. Thus, the map $\iota_B : \mathbb{Z}[\zeta] \to \operatorname{End}_{\bar{F}} B$ given by $\zeta \mapsto \zeta_B$ is a ζ-multiplication on B, and the p-isogeny ϕ is ζ-linear by construction. □

2C. *Twists.* If (A, ι_A) has ζ_n-multiplication, then ι_A induces an inclusion $\mathbb{Z}[\zeta_{2n}]^\times \subset \operatorname{Aut}_{\bar{F}} A$ of G_F-modules. For each $d \in F^\times$, let A_d be the twist of A corresponding to the image of d under

$$F^\times \to F^\times/F^{\times 2n} \simeq H^1(F, \mu_{2n}) \to H^1(F, \operatorname{Aut}_{\bar{F}} A).$$

Then A_d is an abelian variety over F that becomes isomorphic to A over $F(d^{1/2n})$. Moreover, A_d also has ζ_n-multiplication.

Remark 2.7. If $\operatorname{Aut}_{\bar{F}} A = \mu_{2n}$, then distinct $2n$-th power classes d give nonisomorphic A_d. However, if $\operatorname{Aut}_{\bar{F}} A \supsetneq \mu_{2n}$, then the map $H^1(F, \mu_{2n}) \to H^1(F, \operatorname{Aut}_{\bar{F}} A)$ need not be injective, and hence the twists A_d need not be distinct.

Now let $\phi : A \to B$ be a ζ-linear p-isogeny over F. By Lemma 2.6, the automorphisms $\pm\zeta \in \operatorname{Aut}_{\bar{F}} A$ preserve the subgroup $A[\phi]$, giving an inclusion of G_F-modules $\mu_{2n} \hookrightarrow \operatorname{Aut}_{\bar{F}}(\phi)$, where $\operatorname{Aut}_{\bar{F}}(\phi)$ is the subgroup of $\operatorname{Aut}_{\bar{F}} A$ stabilizing $A[\phi]$. For $d \in F^\times$, let $\phi_d : A_d \to B_d$ be the twist of ϕ corresponding to the image of d under

$$F^\times \to F^\times/F^{\times 2n} \simeq H^1(F, \mu_{2n}) \to H^1(F, \operatorname{Aut}_{\bar{F}}(\phi)).$$

Then ϕ_d is a ζ-linear p-isogeny over F. Similarly, we may twist the isogeny $\pi : A \to A^{(1)}$ to obtain $\pi_d : A_d \to A_d^{(1)}$, and this is the canonical isogeny "π" associated to A_d (and its ζ-multiplication).

Remark 2.8. By Lemma 2.3, the abelian variety $A^{(1)}$ is the twist of A corresponding to the cocycle $\xi : \sigma \mapsto \frac{1-\zeta^\sigma}{1-\zeta}$ in $H^1(F, \mathbb{Z}[\zeta_n]^\times)$. When $n = 3$, we have $\mathbb{Z}[\zeta_{2n}]^\times = \mu_6$, and since $\frac{1-\zeta}{\sqrt[6]{-27}} \in \mu_6$, the cocycle ξ is in the same cohomology class as $-27 \in F^\times/F^{\times 6}$. It follows that $A^{(1)} = A_{-27}$, which is the quadratic twist of A by the mod 3 cyclotomic character. More generally, we have $A^{(\frac{1}{2}\varphi(n))} = A_{(p^*)^n}$, where $p^* = (-1)^{(p-1)/2}p$ is such that $\mathbb{Q}(\sqrt{p^*}) \subset \mathbb{Q}(\zeta_p)$. However, for general s and n, the twist $A^{(s)}$ need not be isomorphic to A_d, for any $d \in F^\times/F^{\times 2n}$.

3. Integral orbits of binary cubic forms

In this section, we classify the integral $\mathrm{SL}_2(F)$-orbits on the space of binary cubic forms over a local field F. We first recall some facts from [Bhargava et al. 2020, §2] and [Bhargava 2004, Theorem 13].

Let $V = \operatorname{Sym}^3 \mathbb{Z}^2$ be the space of binary cubic forms. The group SL_2 acts on V, and the ring of invariant functions is generated by the usual polynomial discriminant $\operatorname{Disc} : V \to \mathbb{Z}$. Let F be any field of characteristic not 2 or 3, and for any $d \neq 0$ in F, define $V(F)_d := \{f \in V(F) : \operatorname{Disc}(f) = d\}$. There is a unique *reducible* $\mathrm{SL}_2(F)$-orbit of cubic forms $f \in V(F)_d$. The stabilizer of such an f is a commutative F-group scheme C_d of order 3. The Galois action on $C_d(\bar{F})$ is by the quadratic character $\chi_d : \operatorname{Gal}(F(\sqrt{d})/F) \to \{\pm 1\}$.

Proposition 3.1. *The group $H^1(F, C_d)$ is in bijection with the* $\mathrm{SL}_2(F)$*-orbits on $V(F)_d$.*

Now let F be a local field of residue characteristic neither 2 nor 3, with surjective discrete valuation $v : F^\times \to \mathbb{Z}$, ring of integers $\mathcal{O}_F$, maximal ideal $\mathfrak{m}$ and residue field $\mathbb{F}$ of cardinality q.

We wish to determine which $\mathrm{SL}_2(F)$-orbits in $V(F)_d$ have representatives in $V(\mathcal{O}_F)_d$. We call these orbits the *integral orbits*, and we let $H^1_{\mathrm{int}}(C_d)$ be the subset of $H^1(F, C_d)$ that they correspond to under the bijection of Proposition 3.1. Of course, a necessary condition for there to be any integral orbits at all is that $d \in \mathcal{O}_F$. We see that even though the abstract group $H^1(F, C_d)$ depends only on the square-class of d, the notion of integrality depends on the actual value of d, and in particular its valuation.

We recall some facts about cubic rings over F and over $\mathcal{O}_F$ [Bhargava et al. 2013]. The action of SL_2 on V extends to the following action of GL_2: if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then

$$(\gamma \cdot f)(x, y) = \frac{1}{\det \gamma} f(ax + cy, bx + dy).$$

If R is a principal ideal domain, then a *cubic ring over* R is an R-algebra S that is free of rank 3 as an R-module. The discriminant of S is a well-defined element of $R^{\times}/R^{\times 2}$.

Proposition 3.2 (Levi, Delone–Faddeev, Gan–Gross–Savin). *For any principal ideal domain R, there is a discriminant preserving bijection between $\mathrm{GL}_2(R)$-orbits on $V(R)$ and isomorphism classes of cubic rings over R. Moreover, this bijection is functorial in R.*

Proof. Building on [Levi 1914; Delone and Faddeev 1940; Gan et al. 2002], it is shown in [Bhargava et al. 2013] that the bijection sends a cubic R-ring S to any binary cubic form representing the cubic map $S/R \to \wedge^2_R(S/R)$, $s \mapsto s \wedge s^2$, which is functorial in R. □

If $\gamma \in \mathrm{GL}_2(F)$ and $f \in V(F)$, then $\mathrm{Disc}(\gamma f) = \det(\gamma)^2 \,\mathrm{Disc}(f)$. It follows that isomorphism classes of cubic F-algebras L of discriminant d are in bijection with $\mathrm{GL}_2(F)_{\pm 1}$-orbits on $V(F)_d$. Here, $\mathrm{GL}_2(F)_{\pm 1}$ is the subgroup of $\mathrm{GL}_2(F)$ consisting of elements with determinant ± 1. Since $\mathrm{SL}_2(F)$ has index 2 in $\mathrm{GL}_2(F)_{\pm 1}$, the $\mathrm{GL}_2(F)_{\pm 1}$-orbits break up into at most two $\mathrm{SL}_2(F)$-orbits. It is a fun exercise to show that there are exactly two orbits if and only if L is a field; the orbits are represented by $f(x, y)$ and $f(y, x)$.

Remark 3.3. The trivial class in $H^1(F, C_d)$ corresponds to the unique orbit of *reducible* forms of discriminant d. Hence, $\alpha \in H^1(F, C_d)$ is nontrivial if and only if the corresponding cubic algebra L is a field (if and only if L is generated over F by a root of $f(x, 1)$). The trivial class corresponds to $F \times E_d$, where $E_d = F[x]/(x^2 - d)$ is the quadratic F-algebra of discriminant d. Note that the trivial class is represented by $\frac{d}{4}x^3 + xy^2$, which is integral as long as d is.

From the functoriality in Proposition 3.2 applied to the base change $\mathcal{O}_F \hookrightarrow F$, we deduce:

Proposition 3.4. *Let $\alpha \in H^1(F, C_d)$, and let L be the corresponding cubic F-algebra. Then α is integral if and only if there is an $\mathcal{O}_F$-order $S \subset \mathcal{O}_L$ with $v(\mathrm{Disc}\ S) = v(d)$.*

Proof. The "only if" direction is clear. For the "if" direction, observe that any $\mathcal{O}_F$-order $S \subset \mathcal{O}_L$ has discriminant congruent to d modulo $F^{\times 2}$. So if $v(\mathrm{Disc}\ S) = v(d)$, then we see that $\mathrm{Disc}(S)$ is congruent to d modulo $\mathcal{O}_F^{\times 2}$. Thus we may choose bases so that S corresponds to a binary cubic form with coefficients in $\mathcal{O}_F$ of exact discriminant d. □

The following two facts about cubic orders will be useful [Bhargava et al. 2013, Propositions 15–16].

Proposition 3.5. *Let L be an étale cubic F-algebra. Suppose $f(x, y)$ corresponds to the maximal order $\mathcal{O}_L$ under the bijection of Proposition 3.2. Then the factorization type of $f(x, y)$ over the residue field $\mathbb{F}$ is the factorization type of the maximal ideal $\mathfrak{m}$ of $\mathcal{O}_F$ in the ring $\mathcal{O}_L$.*

Proposition 3.6. *Let $f(x, y) \in V(\mathcal{O}_F)_d$ correspond to a cubic ring S over $\mathcal{O}_F$. Then the sub-$\mathcal{O}_F$-rings $S' \subset S$ of index q correspond bijectively with the zeros of $f \pmod{\mathfrak{m}}$ in $\mathbb{P}^1(\mathbb{F})$.*

	$d \in F^{\times 2}$	$-3d \in F^{\times 2}$
$\zeta_3 \in F$	$\dim H^1(F, C_d) = 2$ $\dim H^1_{\mathrm{un}}(F, C_d) = 1$	
$\zeta_3 \notin F$	$\dim H^1(F, C_d) = 1$ $\dim H^1_{\mathrm{un}}(F, C_d) = 1$	$\dim H^1(F, C_d) = 1$ $\dim H^1_{\mathrm{un}}(F, C_d) = 0$

Table 1. Dimensions of $H^1(F, C_d)$ and $H^1_{\mathrm{un}}(F, C_d)$.

We also need the following result, which requires $\operatorname{char} \mathbb{F} \neq 3$, and which describes the subgroup $H^1_{\mathrm{un}}(F, C_d) \subset H^1(F, C_d)$ of unramified classes.

Proposition 3.7. *Suppose $0 \neq \alpha \in H^1(F, C_d)$ corresponds to the cubic extension L/F. Then $\alpha \in H^1_{\mathrm{un}}(F, C_d)$ if and only if L is unramified.*

Proof. Let $f_L \in V(F)$ be the corresponding binary cubic form. If L is unramified, then since f_L becomes reducible over L, the restriction of α to $H^1(L, C_d)$ is trivial. Thus, α is an unramified class. If L is ramified, then f_L remains irreducible over every unramified extension of F, and hence α is ramified. □

Lemma 3.8. *Assume the residue characteristic of F is not* 3. *Then*:

(i) $\dim H^1(F, C_d) = \dim H^0(F, C_d) + \dim H^0(F, C_{-3d})$.

(ii) $\dim H^1_{\mathrm{un}}(F, C_d) = \dim H^0(F, C_d)$.

When d has even valuation, these dimensions are computed in Table 1.

Proof. First note that C_d is Cartier dual to $C_{-3d} \simeq C_d \otimes \mu_3$. Since the residue characteristic is not 3, the Euler–Poincaré characteristic formula [Milne 1986, I.2.8] immediately gives (i). Let $I_F \subset G_F$ be the inertia group and let g be the Frobenius element of G_F/I_F. Then the groups $H^1_{\mathrm{un}}(F, C_d) \simeq H^0(I_F, C_d)/(g-1)H^0(I_F, C_d)$ and $H^0(I_F, C_d)[g-1] = H^0(F, C_d)$ have the same cardinality, which proves (ii). The table is computed using the fact that $\dim H^0(F, C_d) = 1$ if and only if $d \in F^{\times 2}$ and the dimension is 0 otherwise. □

The main result of this section is the following classification of the integral orbits in $V(F)_d$.

Theorem 3.9. *Let $\mathcal{O}_F$ be the ring of integers of a local field F with* $\operatorname{char} \mathcal{O}_F/\mathfrak{m} > 3$, *and let $d \in \mathcal{O}_F$ be nonzero.*

(a) *If $v(d) = 0$, then $H^1_{\mathrm{int}}(F, C_d) = H^1_{\mathrm{un}}(F, C_d)$.*

(b) *If $v(d)$ is odd, then $H^1_{\mathrm{int}}(F, C_d) = H^1(F, C_d) = 0$.*

(c) *If $v(d) = 2$, then the only nonintegral classes are the nontrivial unramified classes.*

(d) *If $v(d) > 2$, then all classes are integral.*

Proof. (a) This case follows from Propositions 3.4 and 3.7.

(b) We have $H^1(F, C_d) = 0$ by Lemma 3.8 (since $H^0(F, C_d) = 0$ whenever d has odd valuation). By Remark 3.3, the trivial class is integral.

(c) The ramified classes α correspond to totally ramified cubic extensions L/F. For such L we have $v(\operatorname{Disc} \mathcal{O}_L) = v(d)$, and hence these α are integral by Proposition 3.4. If $H^1(F, C_d)$ has a nontrivial unramified class α, then it corresponds to the unique unramified cubic extension L/F, which has unit discriminant. By Proposition 3.4, α is integral if and only if $\mathcal{O}_L$ has an order of index q. By Proposition 3.5, the binary form corresponding to $\mathcal{O}_L$ has no root over $\mathbb{F}$. So by Proposition 3.6, $\mathcal{O}_L$ has no order of index q. Hence the nontrivial unramified classes are indeed nonintegral.

(d) If $\alpha \in H^1(F, C_d)$ corresponds to a ramified cubic extension L/F, then $\mathcal{O}_L$ has discriminant of valuation 2. By Propositions 3.5 and 3.6, $\mathcal{O}_L$ has a unique order S of index q, and hence $v(\operatorname{Disc}\ S) = 4$. Note that if S_0 is a cubic $\mathcal{O}_F$-ring, then $S_0' = \mathcal{O}_F + \mathfrak{m}S_0$ is a subring of S of index q^2 and $\operatorname{Disc}(S_0') = q^4 \operatorname{Disc}(S_0)$. Thus, by considering the orders $\mathcal{O}_F + \mathfrak{m}^k\mathcal{O}_L$ and $\mathcal{O}_F + \mathfrak{m}^k S$, we can find an order $S' \subset \mathcal{O}_L$ with $v(\operatorname{Disc}\ S') = 2k$, for any $k \geq 1$.

Next let $\alpha \in H^1(F, C_d)$ be a nontrivial unramified class corresponding to the unramified cubic extension L/F. Then $v(\operatorname{Disc}(\mathcal{O}_F + \mathfrak{m}\mathcal{O}_L)) = 4$. By Proposition 3.6, there are $q+1$ suborders $S' \subset \mathcal{O}_F + \mathfrak{m}\mathcal{O}_L$ of index q, so that $v(\operatorname{Disc}\ S') = 6$. As before, we deduce that there exists an order $S'' \subset \mathcal{O}_L$ with $v(\operatorname{Disc}\ S'') = 2k$, for any $k \geq 2$. □

4. Local Selmer conditions for ζ-linear isogenies

Let F be a finite extension of $\mathbb{Q}_p$ with surjective discrete valuation v, ring of integers $\mathcal{O}_F$, uniformizer ϖ, and residue field $\mathbb{F}$. Let $m \geq 1$, $n = 3^m$, and let $\zeta = \zeta_n$ be a primitive n-th root of unity. Let A, B be abelian varieties over F that admit ζ-multiplication.

Let $\phi : A \to B$ be a ζ-linear 3-isogeny over F, as defined in Section 2B. For each $d \in F^\times$, we consider the 3-isogeny $\phi_d : A_d \to B_d$, as in Section 2C. We will assume in this section that $A[\phi](F) \neq 0$, that is, that $A[\phi]$ is generated by a rational point. This can always be achieved by replacing ϕ with an appropriate twist, so there is no loss in generality. We also assume that $0 \leq v(d) < 2n$, again with no loss in generality.

The group $H^1(F, A_d[\phi_d])$ is a finite-dimensional $\mathbb{F}_3$-vector space. In fact, if $\chi_d : G_F \to \mathbb{F}_3^\times$ is the quadratic character cutting out $F(\sqrt{d})$, then $A_d[\phi_d] \simeq A[\phi] \otimes \chi_d$ is isomorphic to C_d from Section 3. Thus, $H^1(F, A_d[\phi_d]) \simeq H^1(F, C_d)$. The exact sequence

$$0 \to A_d[\phi_d] \to A_d \to B_d \to 0$$

induces a Kummer map

$$\partial_d : B_d(F) \to H^1(F, A_d[\phi_d]).$$

We call its image $\operatorname{im}(\partial_d) \subset H^1(F, A_d[\phi_d])$ the subgroup of *soluble classes*. The goal of this section is to prove the following theorem describing the soluble classes and to compute the local Selmer ratios of ϕ_d.

Theorem 4.1. *Assume that* $\operatorname{char}\mathbb{F} \neq 3$, *that A has good reduction, and that* $A[\phi](F) \neq 0$. *Suppose that* $0 \leq v(d) < 2n$.

(a) *If* $v(d) = 0$, *then* $\operatorname{im}(\partial_d) = H^1_{\mathrm{un}}(F, A_d[\phi_d])$.

(b) *If $v(d)$ is odd, or more generally, if $F(\sqrt{d})/F$ is ramified, then* $\operatorname{im}(\partial_d) = H^1(F, A_d[\phi_d]) = 0$.

(c) *If $v(d) > 0$ is even and $d \notin F^{\times 2}$, then* $H^1_{\mathrm{un}}(F, A_d[\phi_d]) = 0$.

(d) *If $v(d) > 0$ is even and $d \in F^{\times 2}$, then write $d = \varpi^{v(d)}u$ with $u \in \mathcal{O}_F^\times$ and let* $s = \gcd(3^m, v(d))$. *Then* $\operatorname{im}(\partial_d) \cap H^1_{\mathrm{un}}(F, A_d[\phi_d]) = 0$ *if and only if $A_u[\phi]$ is a direct summand of $A_u[\pi^s]$.*

We retain these assumptions on A and $\mathbb{F}$ for the remainder of this section.

4A. ***Nonsquare and unramified twists.*** We first prove parts (a)–(c).

Proof of Theorem 4.1(a)–(c). For (a), assume at first that $\operatorname{char}\mathbb{F} > 3$. Then $F(d^{1/2n})$ is unramified over F, since $v(d) = 0$ and $(\operatorname{char}\mathbb{F}, 2n) = 1$. Since A_d is isomorphic to A over $F(d^{1/2n})$, it has good reduction over an unramified extension, and hence has good reduction already over F. Since A_d has good reduction and $\operatorname{char}\mathbb{F} \nmid \deg(\phi_d)$, the image of the Kummer map ∂_d is exactly the unramified classes [Česnavičius 2016, Proposition 2.7(d)]. The proof just given works even when $\operatorname{char}\mathbb{F} = 2$, as long as $F(\sqrt{d})/F$ is unramified. When this extension is ramified, the result follows from (b). Part (b) itself follows from Theorem 3.9(b). The case $\operatorname{char}\mathbb{F} = 2$ was not dealt with there, but the proof is identical.

For (c), we have $H^1_{\mathrm{un}}(F, A_d[\phi_d]) = H^1_{\mathrm{un}}(F, C_d) = 0$ by Lemma 3.8, since $H^0(F, C_d) = 0$ whenever $d \notin F^{\times 2}$. □

4B. ***Twists of positive valuation.*** The proof of Theorem 4.1(d) will take more work. Indeed, we will compute more generally the size of $\operatorname{im}(\partial_d)$ for all d (including $d \notin F^{\times 2}$) such that $v(d)$ is even and positive. Let $s = \gcd(v(d), 3^m)$, and write $d = \varpi^{v(d)}u$ for $u \in \mathcal{O}_F^\times$. Recall the map $\pi^s : A \to A^{(s)}$ defined in Section 2, and let $\psi^s : B \to A^{(s)}$ be the isogeny such that $\psi^s \circ \phi = \pi^s$.

4B1. *Extension classes.* For each $t \in F^\times$, let κ_t^s be the extension class corresponding to the short exact sequence

$$0 \to A_t[\phi_t] \to A_t[\pi_t^s] \xrightarrow{\phi_t} B_t[\psi_t^s] \to 0. \tag{4-1}$$

Thus $\kappa_t^s = 0$ if and only if $A_t[\phi_t]$ is a direct summand of $A_t[\pi_t^s]$ as a G_F-module. Similarly, let $\hat{\kappa}_t^s$ be the class of the extension

$$0 \to B_t[\psi_t^s] \to B_t[\pi_t^s] \xrightarrow{\psi_t^s} A_t^{(s)}[\phi_t^{(s)}] \to 0. \tag{4-2}$$

Thus, $\hat{\kappa}_t^s = 0$ if and only if $B_t[\psi_t^s]$ is a direct summand of $B_t[\pi_t^s]$.

Remark 4.2. By Lemma 2.4, the cocycle $\hat{\kappa}_t^s$ is equal to the class of the extension

$$0 \to B_t^{(-s)}[\psi_t^{-s}] \to B_t^{(-s)}[\pi_t^s] \to A_t[\phi_t] \to 0.$$

	$d\in F^{\times 2}$	$-3d\in F^{\times 2}$	$d,-3d\notin F^{\times 2}$
$\zeta_3\in F$	$\dim H^1(F,A_d[\phi_d])=2$ $\dim H^1_{\mathrm{un}}(F,A_d[\phi_d])=1$		$\dim H^1(F,A_d[\phi_d])=0$ $\dim H^1_{\mathrm{un}}(F,A_d[\phi_d])=0$
$\zeta_3\notin F$	$\dim H^1(F,A_d[\phi_d])=1$ $\dim H^1_{\mathrm{un}}(F,A_d[\phi_d])=1$	$\dim H^1(F,A_d[\phi_d])=1$ $\dim H^1_{\mathrm{un}}(F,A_d[\phi_d])=0$	$\dim H^1(F,A_d[\phi_d])=0$ $\dim H^1_{\mathrm{un}}(F,A_d[\phi_d])=0$

Table 2. Dimensions of $H^1(F,A_d[\phi_d])$ and $H^1_{\mathrm{un}}(F,A_d[\phi_d])$.

Here, we have $\phi\circ\psi_t^{-s}=\pi^s : B_t^{(-s)}\to B_t$. By duality, $\hat{\kappa}_t^s$ is the class of the extension

$$0\to \widehat{B}_t[\hat{\phi}_t]\to \widehat{B}_t[\hat{\pi}_t^s]\to \hat{A}_t[\hat{\psi}_t^{-s}]\to 0.$$

Thus $\hat{\kappa}_t^s=0$ if and only if $\widehat{B}_t[\hat{\phi}_t]$ is a direct summand of $\widehat{B}_t[\hat{\pi}_t^s]$, which explains the notation.

Let $|\kappa_u^s|$ and $|\hat{\kappa}_u^s|$ denote the orders of the classes κ_u^s and $\hat{\kappa}_u^s$ in their respective Ext-groups.

4B2. *The image of* ∂_d. The following theorem relates the size of $\operatorname{im}\partial_d$ to $|\kappa_u^s|$ and $|\hat{\kappa}_u^s|$ and will finish the proof of Theorem 4.1(d).

Theorem 4.3. *Assume that $v(d)$ is even and positive, and write $d=\varpi^{v(d)}u$ with $u\in\mathcal{O}_F^\times$.*

(i)
$$\#\operatorname{im}\partial_d\cap H^1_{\mathrm{un}}(F,A_d[\phi_d])=\begin{cases}|\kappa_u^s| & \textit{if } d\in F^{\times 2},\\ 1 & \textit{otherwise.}\end{cases}$$

(ii)
$$\#\left(\frac{\operatorname{im}\partial_d}{\operatorname{im}\partial_d\cap H^1_{\mathrm{un}}(F,A_d[\phi_d])}\right)=\begin{cases}\dfrac{3}{|\hat{\kappa}_u^s|} & \textit{if } -3d\in F^{\times 2},\\ 1 & \textit{otherwise.}\end{cases}$$

In particular, if $d\in F^{\times 2}$, then $\operatorname{im}\partial_d\cap H^1_{\mathrm{un}}(F,A_d[\phi_d])=0$ *if and only if $A_u[\phi_u]$ is a direct summand of $A_u[\pi_u^s]$.*

The proof of Theorem 4.3 requires several preliminary results.

Lemma 4.4. *The dimensions of $H^1(F,A_d[\phi_d])$ and $H^1_{\mathrm{un}}(F,A_d[\phi_d])$ are as in Table 2.*

Proof. Since $H^1(F,A_d[\phi_d])\simeq H^1(F,C_d)$ and $H^1_{\mathrm{un}}(F,A_d[\phi_d])\simeq H^1_{\mathrm{un}}(F,C_d)$, the dimensions in Table 2 are identical to those in Table 1. The bottom right cell is only relevant when $\operatorname{char}\mathbb{F}=2$, and follows from Lemma 3.8(i). □

Lemma 4.5. *If $t\in F^{\times 2s}$, then $A_t[\pi_t^s]\simeq A[\pi^s]$ and $B_t[\pi_t^s]\simeq B[\pi^s]$ as G_F-modules.*

Proof. By the definition of A_t, there is an isomorphism $\phi : A\to A_t$ over $F(t^{1/2n})$ such that if $P\in A[\pi^s]$ and $\sigma\in G_F$, then in $A_t[\pi_t^s]$, we have

$$(\phi(P))^\sigma=\frac{\sigma(t^{1/2n})}{t^{1/2n}}\phi(P^\sigma).$$

If $t\in F^{\times 2s}$, then $\frac{\sigma(t^{1/2n})}{t^{1/2n}}\in\langle\zeta^s\rangle$. Since ζ^s acts as the identity on $A[\pi^s]$, we see that $A_t[\pi_t^s]\simeq A[\pi^s]$. The proof that $B_t[\pi_t^s]\simeq B[\pi^s]$ is identical. □

Remark 4.6. In particular, if $s = 1$, the condition in Theorem 4.1(d) is simply that $A[\phi]$ is a direct summand of $A[\pi]$, which is independent of u.

Corollary 4.7. (i) *There is an isomorphism between*

$$0 \to A_u[\phi_u] \to A_u[\pi_u^s] \xrightarrow{\phi_u} B_u[\psi_u^s] \to 0$$

and

$$0 \to A_d[\phi_d] \to A_d[\pi_d^s] \xrightarrow{\phi_d} B_d[\psi_d^s] \to 0$$

as short exact sequences of G_F-modules.

(ii) *There is an isomorphism between*

$$0 \to B_u[\psi_u^s] \to B_u[\pi_u^s] \xrightarrow{\psi_u^s} A_u^{(s)}[\phi_u^{(s)}] \to 0$$

and

$$0 \to B_d[\psi_d^s] \to B_d[\pi_d^s] \xrightarrow{\psi_d^s} A_d^{(s)}[\phi_d^{(s)}] \to 0$$

as short exact sequences of G_F-modules.

Proof. The first claim follows from Lemma 4.5 together with the observation that the isomorphism $A_u[\pi_u^s] \to A_d[\pi_d^s]$ restricts to an isomorphism $A_u[\phi_u] \to A_d[\phi_d]$. The second claim follows similarly. □

Lemma 4.8. *Suppose that $v(d) = 2a \cdot 3^r$ is even and positive, and let k be an unramified extension of F. For $X \in \{A, B\}$, we have $X_d[3](k) = X_d[\pi_d^{3^r}](k)$.*

Proof. Let F_{un} be the maximal unramified extension of F, let $L = F_{\text{un}}(\sqrt{d})$, and let $M = L(d^{1/n}) = F_{\text{un}}(d^{1/2n})$. Since $k \subset L$, it suffices to show that $X_d[3](L) = X_d[\pi_d^{3^r}](L)$.

The extension M/L is tamely ramified of order 3^{n-r}, and we can choose a generator τ of $\operatorname{Gal}(M/L)$ so that $\tau(d^{1/2n})/d^{1/2n} = \zeta^{3^r}$. Since X has good reduction, we have $X[3] \subset X(L)$. Since X and X_d are isomorphic over $L(d^{1/n})$, it follows that $X_d[3] \subset X(M)$.

Now, if $\phi : X_d \to X$ is an isomorphism over M and $P \in X_d(M)$, then by definition,

$$\phi(P)^\tau = \tau(d^{1/2n})/d^{1/2n}\phi(P^\tau) = \zeta^{3^r}\phi(P^\tau).$$

Hence, if $P \in X_d[3](L)$, then $\phi(P) \in X[3] \subset X[3](L)$, so $\phi(P) = \zeta^{3^r}\phi(P)$. It follows that

$$0 = (1-\zeta^{3^r})\phi(P) = (1-\zeta)^{3^r}\phi(P) = \pi^{3^r}\phi(P) = \phi(\pi_d^{3^r}P),$$

where the second equality uses the fact that P is a 3-torsion point. It follows that $P \in X_d[\pi_d^{3^r}]$, so $X_d[\pi_d^{3^r}](L) = X_d[3](L)$. □

From (4-1), we obtain a long exact sequence

$$0 \to A_d[\phi_d](F) \to A_d[\pi_d^s](F) \xrightarrow{\phi_d} B_d[\psi_d^s](F) \xrightarrow{\delta_d} H^1(F, A_d[\phi_d]).$$

Similarly, from (4-2), there is a long exact sequence

$$0 \to B_d[\psi_d^s](F) \to B_d[\pi_d^s](F) \xrightarrow{\psi_d^s} A_d^{(s)}[\phi_d^{(s)}](F) \xrightarrow{\hat{\delta}_d} H^1(F, B_d[\psi_d^s]). \tag{4-3}$$

Clearly $\operatorname{im}\delta_d \subset \operatorname{im}\partial_d$, and ∂_d induces an injective map

$$\frac{B_d[\pi_d^s](F)}{B_d[\psi_d^s](F)} \hookrightarrow \frac{\operatorname{im}\partial_d}{\operatorname{im}\delta_d}.$$

Proposition 4.9. *We have*

$$\operatorname{im}\delta_d = \operatorname{im}\partial_d \cap H^1_{\mathrm{un}}(F, A_d[\phi_d]).$$

Proof. Consider the Kummer exact sequence

$$0 \to A_d[\phi_d](F) \to A_d(F) \xrightarrow{\phi_d} B_d(F) \xrightarrow{\partial_d} H^1(F, A_d[\phi_d]).$$

Since $\operatorname{char}\mathbb{F} \neq 3$, we have $B_d(F)/\phi_d A_d(F) = B_d[3^\infty](F)/\phi_d A_d[3^\infty](F)$, so ∂_d depends only on the exact sequence

$$0 \to A_d[\phi_d](F) \to A_d[3^\infty](F) \xrightarrow{\phi_d} B_d[3^\infty](F) \xrightarrow{\partial_d} H^1(F, A_d[\phi_d]).$$

By Lemma 4.8, this exact sequence is the same as

$$0 \to A_d[\phi_d](F) \to A_d[\pi_d^s](F) \xrightarrow{\phi_d} B_d[\pi_d^s](F) \xrightarrow{\partial_d} H^1(F, A_d[\phi_d]).$$

By Corollary 4.7, the long exact sequence

$$0 \to A_d[\phi_d](F) \to A_d[\pi_d^s](F) \xrightarrow{\phi_d} B_d[\psi_d^s](F) \xrightarrow{\delta_d} H^1(F, A_d[\phi_d])$$

is isomorphic to the long exact sequence

$$0 \to A_u[\phi_u](F) \to A_u[\pi_u^s](F) \xrightarrow{\phi_u} B_u[\psi_u^s](F) \xrightarrow{\delta_u} H^1(F, A_u[\phi_u]).$$

Since A_u has good reduction, the image of δ_u is contained in $H^1_{\mathrm{un}}(F, A_u[\phi_u])$ [Česnavičius 2016, Proposition 2.7(d)]. Thus the image of δ_d is contained in both $H^1_{\mathrm{un}}(F, A_d[\phi_d])$ and $\operatorname{im}\partial_d$.

Conversely, if $y \in B_d[\pi_d^s](F) \setminus B_d[\psi_d^s](F)$, then $\partial_d(y)$ is the cocycle $\sigma \mapsto x^\sigma - x$, where $\phi_d(x) = y$. But $x \in A_d[3] \setminus A_d[\pi_d^s]$, so by Lemma 4.8, x cannot be defined over an unramified extension of F. It follows that $\partial_d(y) \notin H^1_{\mathrm{un}}(F, A_d[\phi_d])$. Hence $\operatorname{im}\delta_d \supset \operatorname{im}\partial_d \cap H^1_{\mathrm{un}}(F, A_d[\phi_d])$. □

We next relate the sizes of the images of δ_d and $\hat{\delta}_d$ to the sizes of $|\kappa_u^s|$ and $|\hat{\kappa}_u^s|$.

Lemma 4.10. *If $d \in F^{\times 2}$, then $\#\operatorname{im}\delta_d = |\kappa_u^s|$. Similarly, if $-3d \in F^{\times 2}$, then $\#\operatorname{im}\hat{\delta}_d = |\hat{\kappa}_u^s|$.*

Proof. By duality, we have $A_u[\phi_u] \simeq \widehat{B}_u[\hat{\phi}_u] \otimes \chi_3$. Thus, the two claims of the lemma are in fact equivalent to each other; see Remark 4.2. We prove the second one. Since $-3d \in F^{\times 2}$, we have $A_u^{(s)}[\phi_u^{(s)}] = \mathbb{F}_3$ as a G_F-module. Moreover, by definition we have $\hat{\delta}_d(1) = \hat{\kappa}_d^s$. By Corollary 4.7, $|\hat{\kappa}_d^s| = |\hat{\kappa}_u^s|$. It follows that $|\hat{\kappa}_u^s| = \#\operatorname{im}\hat{\delta}_d$. □

Proof of Theorem 4.3. If $d \in F^{\times 2}$, then part (i) follows immediately from Proposition 4.9 and Lemma 4.10. If $d \notin F^{\times 2}$, then $H^1_{\mathrm{un}}(F, A_d[\phi_d]) = 0$ by Table 2.

	$d \in F^{\times 2}$	$-3d \in F^{\times 2}$	$d, -3d \notin F^{\times 2}$
$\zeta_3 \in F$	$\frac{\lvert\kappa_u^s\rvert}{\lvert\hat{\kappa}_u^s\rvert}$		1
$\zeta_3 \notin F$	$\frac{\lvert\kappa_u^s\rvert}{3}$	$\frac{3}{\lvert\hat{\kappa}_u^s\rvert}$	1

Table 3. Values of $c(\phi_d)$ over local fields F, when $v(d)$ is even and positive.

By Lemma 4.8 and Proposition 4.9,

$$\frac{\operatorname{im}\partial_d}{\operatorname{im}\partial_d \cap H^1_{\mathrm{un}}(F, A_d[\phi_d])}$$

is isomorphic to the image of the injective map

$$\frac{B_d[\pi_d^s](F)}{B_d[\psi_d^s](F)} \to \frac{\operatorname{im}\partial_d}{\operatorname{im}\delta_d}$$

induced by ∂_d. From (4-3), we have

$$\#\left(\frac{B_d[\pi_d^s](F)}{B_d[\psi_d^s](F)}\right) = \frac{\#A_d^{(s)}[\phi_d^{(s)}](F)}{\#\operatorname{im}\hat{\delta}_d}.$$

If $-3d \in F^{\times 2}$, this is $\frac{3}{|\hat{\kappa}_u^s|}$ by Lemma 4.10. If $-3d \notin F^{\times 2}$, then $H^1(F, A_d[\phi_d]) = H^1_{\mathrm{un}}(F, A_d[\phi_d])$ by Table 2, so the result follows from Proposition 4.9. □

4C. *Local Selmer ratios.* For applications, we record the local Selmer ratios

$$c(\phi_d) = \frac{\#\operatorname{coker}(A_d(F) \to B_d(F))}{\#\ker(A_d(F) \to B_d(F))} = \frac{\#\operatorname{im}(\partial_d)}{\#A_d[\phi_d](F)},$$

which have implicitly been computed in the previous subsection.

Theorem 4.11. *Assume that* $\operatorname{char}\mathbb{F} \neq 3$, *that A has good reduction, and that $A[\phi](F) = \mathbb{Z}/3\mathbb{Z}$. Then $c(\phi_d) = 1$ unless $v(d)$ is even and positive. If $v(d)$ is even and positive, write $d = \varpi^{v(d)}u$ with $u \in \mathcal{O}_F^\times$, and let $s = \gcd(3^m, v(d))$. Let κ_u^s and $\hat{\kappa}_u^s$ be the classes defined in the previous section, and write $|\kappa_u^s|$ and $|\hat{\kappa}_u^s|$ for their orders. Then $c(\phi_d)$ is as in Table 3.*

Proof. If neither d nor $-3d$ is a square in F, then $\#A_d[\phi_d](F) = 1$, and by Table 2, $\#\operatorname{im}(\partial_d) = 1$, so $c(\phi_d) = 1$ as claimed. Henceforth, we assume that either d or $-3d$ is a square in F. When $v(d) = 0$, then A_d has good reduction, so by [Shnidman 2021, Proposition 3.1]

$$c(\phi_d) = c(B_d)/c(A_d) = 1,$$

where $c(A_d)$ and $C(B_d)$ are the Tamagawa numbers of A_d and B_d. When $v(d)$ is odd, then $\operatorname{im}(\partial_d) = 0$ (Theorem 4.1) and $A_d[\phi_d](F) = 0$, so again $c(\phi_d) = 1$. So it remains to compute $c(\phi_d)$ when $v(d)$ is even and positive. This is done by combining the formula for $\#\operatorname{im}\partial_d$ in Theorem 4.3 with the fact that $\#A_d[\phi_d](F) = 3$ if d is a square and 1 otherwise. The result of this computation is Table 3. □

5. Selmer groups and integrality

We return to the global setting of the introduction, so that F is a number field and $n = 3^m$ for some $m \geq 1$. Recall that $\zeta \in \bar{F}$ is a primitive n-th root of unity. Let $\phi : A \to B$ be a ζ-linear 3-isogeny over F. Recall from Section 2 that there is a twist $A^{(1)}$ of A, and an isogeny $\pi : A \to A^{(1)}$, which becomes isomorphic to the endomorphism $1 - \zeta$ over $F(\zeta)$.

For any F-algebra K, define $\mathbb{B}(K) = K^\times / K^{\times 2n}$. The notation is meant to suggest that $\mathbb{B}$ is the classifying stack $B\mu_{2n}$. For $d \in \mathbb{B}(F)$, let $\phi_d : A_d \to B_d$ be a twist of ϕ corresponding to

$$d \in F^\times / F^{\times 2n} \simeq H^1(F, \mu_{2n}) \to H^1(F, \mathrm{Aut}_{\bar{F}}(\phi)),$$

as in Section 2C. The Selmer group $\mathrm{Sel}_{\phi_d}(A_d)$ is the subgroup of $H^1(F, A_d[\phi_d])$ consisting of classes whose restriction lies in the image of the Kummer map

$$\partial_{d,v} : B_d(F_v)/\phi_d(A_d(F_v)) \hookrightarrow H^1(F_v, A_d[\phi_d])$$

for all places v of F. Sometimes we use the notation $\mathrm{Sel}(\phi_d)$ instead of the more clunky $\mathrm{Sel}_{\phi_d}(A_d)$.

The goal of this section is to compute the average size of $\mathrm{Sel}_{\phi_d}(A_d)$ as d varies. The idea is to view Selmer elements as $\mathrm{SL}_2(F)$-orbits of binary cubic forms and then apply geometry-of-numbers counting techniques. To carry this out, we must show that the orbits corresponding to Selmer elements have representatives with bounded denominator. In fact, we will show that this boundedness only holds if $A[\phi]$ is almost everywhere locally a direct summand of $A[\pi]$.

5A. *Integrality of Selmer elements.* We assume for simplicity that $A[\phi] \simeq \mathbb{Z}/3\mathbb{Z}$ as group schemes. This is not really a constraint, since there is always a quadratic twist of A with this property. This assumption implies that $A_d[\phi_d] \simeq C_d$, where C_d is the order 3 group scheme cut out by the quadratic field of discriminant d, from Section 3.

Recall the space V of binary cubic forms from Section 3. Recall also the set $V(F)_d$ of cubic forms of discriminant d, whose $\mathrm{SL}_2(F)$-orbits are in bijection with $H^1(F, C_d) \simeq H^1(F, A_d[\phi_d])$. Similarly, for each place v of F, there is a bijection between $\mathrm{SL}_2(F_v)$-orbits on the set $V(F_v)_d$ and $H^1(F_v, A_d[\phi_d])$. Let $V(F_v)_d^{\mathrm{sol}}$ denote the subset of $V(F_v)_d$ corresponding to classes $\alpha \in H^1(F_v, A_d[\phi_d])$ in the image of $\partial_{d,v}$. Similarly, let $V(F)_d^{\mathrm{sel}}$ denote the subset of $V(F)_d$ corresponding to classes in $\mathrm{Sel}_{\phi_d}(A_d)$. Define

$$V(F)^{\mathrm{sel}} = \bigcup_{0 \neq d \in \mathcal{O}_F} V(F)_d^{\mathrm{sel}} \quad \text{and} \quad V(F_v)^{\mathrm{sol}} = \bigcup_{d \in \mathcal{O}_v(2n)} V(F_v)_d^{\mathrm{sol}},$$

where $\mathcal{O}_v$ is the ring of integers in F_v, and $\mathcal{O}_v(2n) = \{d \in \mathcal{O}_v : v(d) < 2n\}$. Similarly, define

$$V(F)^{\mathrm{sel}}_{\mathrm{sq.free}} = \bigcup_{0 \neq d \in \mathcal{O}_F \ \mathrm{sq.free}} V(F)_d^{\mathrm{sel}} \quad \text{and} \quad V(F_v)^{\mathrm{sol}}_{\mathrm{sq.free}} = \bigcup_{d \in \mathcal{O}_v(2)} V(F_v)_d^{\mathrm{sol}},$$

where $\mathcal{O}_v(2) = \{d \in \mathcal{O}_v : v(d) < 2\}$ and the union on the left runs over elements $d \in \mathcal{O}_F$ that are squarefree. Of course, these sets depend on the initial choices of A and ϕ.

In order to count the $\mathrm{SL}_2(F)$-orbits on $V(F)^{\mathrm{sel}}$ and $V(F)^{\mathrm{sel}}_{\mathrm{sq.free}}$ of bounded discriminant, we wish to prove that these orbits have representatives in $V(\mathcal{O}_F)$, or at least representatives with denominators that are uniformly bounded. Since SL_2 has class number 1, this is ultimately a local question, and it is enough to prove that for almost all v, each $\mathrm{SL}_2(F_v)$-orbit in $V(F_v)^{\mathrm{sol}}$ and $V(F_v)^{\mathrm{sol}}_{\mathrm{sq.free}}$ has a representative in $V(\mathcal{O}_v)$. For $V(F)^{\mathrm{sel}}_{\mathrm{sq.free}}$, this integrality holds without any conditions. However, for $V(F)^{\mathrm{sel}}$, we are forced to assume that $A[\phi]$ is almost everywhere locally a direct summand of $A[\pi]$, as defined in the introduction.

The following integrality result is crucial for the proofs of Theorems 5.2 and 5.3 below.

Theorem 5.1. *Let $v \nmid 6\infty$ be a place of F at which A has good reduction.*

(i) *Every element of $V(F_v)^{\mathrm{sol}}_{\mathrm{sq.free}}$ is $\mathrm{SL}_2(F_v)$-equivalent to an element of $V(\mathcal{O}_v)$.*

(ii) *If $A[\phi]$ is a direct summand of $A[\pi]$ as a G_{F_v}-module, then every element of $V(F_v)^{\mathrm{sol}}$ is $\mathrm{SL}_2(F_v)$-equivalent to an element of $V(\mathcal{O}_v)$.*

Proof. For each $d \in \mathcal{O}_v$, we must show that each class of $H^1(F_v, A_d[\phi_d]) \simeq H^1(F_v, C_d)$ that lies in the image of $\partial_{d,v}$ corresponds to an integral orbit of discriminant d. This follows from a comparison of Theorem 3.9 with Theorem 4.1 and Remark 4.6. □

5B. ***Average size of the Selmer group.*** There is a natural height function on $\mathbb{B}(F)$ defined as follows. Let M_∞ be the set of archimedean places of F. If $d \in \mathbb{B}(F)$ with lift $d_0 \in F^\times$, then define the ideal $I = \{a \in F : a^{2n} d_0 \in \mathcal{O}_F\}$. The height of d is then

$$h(d) = \mathrm{Nm}(I)^{2n} \prod_{v \in M_\infty} |d_0|_v.$$

This is independent of the lift d_0, by the product formula. If $F = \mathbb{Q}$, then $h(d) = |d_0|$, where d_0 is the unique $2n$-th power free integer representing d. For any $X > 0$, there are finitely many $d \in \mathbb{B}(F)$ with $h(d) < X$.

In order to state a robust version of Theorems 1.8 and 1.9, we recall from [Bhargava et al. 2020] the notion of functions on F that are defined by local conditions. Let $F_\infty = \prod_{v \in M_\infty} F_v$. We say a function $\psi : F \to [0, 1]$ is *defined by local congruence conditions* if there exist local functions $\psi_v : F_v \to [0, 1]$ for every finite place v of F, and a function $\psi_\infty : F_\infty \to [0, 1]$, such that the following two conditions hold:

(1) For all $w \in F$, the product $\psi_\infty(w) \prod_{v \notin M_\infty} \psi_v(w)$ converges to $\psi(w)$.

(2) For each finite place v, and for $v = \infty$, the function ψ_v is nonzero on some open set and locally constant outside some closed subset of F_v of measure 0.

A subset of F is said to be *defined by local congruence conditions* if its characteristic function is defined by local congruence conditions.

Let Σ_0 be a fundamental domain for the action of $F^\times$ on F defined by $\alpha \cdot \beta = \alpha^{2n}\beta$. We may take Σ_0 so that it is defined by local congruence conditions. For any $X > 0$, let F_X denote the set of $d \in F^\times$ such that $h(d) < X$. Then $\Sigma_0 \cap F_X$ is finite and we think of the abelian varieties A_d as elements of Σ_0, so that the set of all A_d, with $d \in \mathbb{B}(F)$ and $h(d) < X$, is naturally in bijection with the finite set $\Sigma_0 \cap F_X$.

A *family of twists* $\{A_d\}$ *defined by local congruence conditions* is then a subset $\Sigma \subset \Sigma_0$ defined by local congruence conditions. In that case, the characteristic function χ_Σ of Σ factors as

$$\chi_\Sigma = \chi_{\Sigma,\infty} \prod_{v \notin M_\infty} \chi_{\Sigma_v}.$$

For each finite place v of F, let Σ_v be the subset of F_v whose characteristic function is χ_{Σ_v}, and let Σ_∞ be the subset of F_∞ whose characteristic function is $\chi_{\Sigma,\infty}$.

We say that Σ is *large* if Σ_v contains the set $\mathcal{O}_v(2) := \{d \in \mathcal{O}_v : v(d) < 2\}$ for all but finitely many finite places v, and if Σ_∞ is a nonempty union of cosets in $\mathbb{B}(F_\infty)$. By construction, we have $\Sigma_{0,v} = \mathcal{O}_v(2n) \supset \mathcal{O}_v(2)$ for all finite v, so Σ_0 is itself large.

If f is a positive function on $\mathbb{B}(F)$ and $\Sigma \subset \Sigma_0$, we write $\Sigma(X) = \Sigma \cap F_X$ and define

$$\operatorname{avg}_\Sigma f(d) = \lim_{X\to\infty} \frac{1}{\#\Sigma(X)} \sum_{d\in\Sigma(X)} f(d).$$

Our formula for $\operatorname{avg}_\Sigma \mathrm{Sel}_{\phi_d}(A_d)$ is most neatly formulated in terms of the global Selmer ratios $c(\phi_d)$ defined in the introduction, and first defined in [Bhargava et al. 2020].

Theorem 5.2. *Let Σ be a large family of twists A_d. For each $k \in \mathbb{Z}$, let $T_k \subset \Sigma$ be the subset of $d \in \Sigma$ such that $c(\phi_d) = 3^k$. Assume either that $A[\phi]$ is almost everywhere locally a direct summand of $A[\pi]$ or that $\Sigma_v = \mathcal{O}_v(2)$ for all but finitely many v. If T_k is nonempty, then*

$$\operatorname*{avg}_{d\in T_k} \# \mathrm{Sel}_{\phi_d}(A_d) = 1 + 3^k.$$

When they are nonempty, the sets T_k are countable disjoint unions of large sets. Using the uniformity estimate [Bhargava et al. 2015, Theorem 17] and copying the argument from [Bhargava et al. 2020, pp. 319–320], we deduce Theorem 5.2 from the following result. To state it, we define for any $d = (d_v)_{v\in M_\infty} \in \mathbb{B}(F_\infty)$,

$$c_\infty(\phi_d) := \prod_{v\in M_\infty} c_v(\phi_{d_v}).$$

We also let $\overline{\Sigma}_\infty$ denote the image of Σ_∞ in the finite group $\mathbb{B}(F_\infty)$.

Theorem 5.3. *Let Σ be a large family of twists A_d. Assume either that $A[\phi]$ is almost everywhere locally a direct summand of $A[\pi]$ or that $\Sigma_v = \mathcal{O}_v(2)$ for all but finitely many v. Then*

$$\operatorname{avg}_\Sigma \# \mathrm{Sel}_{\phi_d}(A_d) = 1 + \frac{\displaystyle\int_{d\in\overline{\Sigma}_\infty} c_\infty(\phi_d)\mu_\infty(d)}{\displaystyle\int_{d\in\overline{\Sigma}_\infty} \mu_\infty(d)} \prod_{v\notin M_\infty} \frac{\displaystyle\int_{d\in\Sigma_v} c_v(\phi_d)\mu_v(d)}{\displaystyle\int_{d\in\Sigma_v} \mu_v(d)},$$

where μ_v is any Haar measure on $\mathcal{O}_v$ and μ_∞ is the uniform measure on $\mathbb{B}(F_\infty)$.

Proof. Assume at first that $A[\phi](F) \simeq \mathbb{Z}/3\mathbb{Z}$, as in Section 5A. Then the left-hand side is equal to the limit as $X \to \infty$ of the average number of $\mathrm{SL}_2(F)$-orbits of discriminant d in $V(F)^{\mathrm{sel}}$, for d in the finite set $\Sigma \cap F_X$. Given our key result Theorem 5.1, Theorem 5.3 now follows exactly as in the proof of [Bhargava et al. 2020, Theorem 11], and we refer the reader there for the details. Note that two proofs were given in that paper, one in the case $F = \mathbb{Q}$ and one for general number fields F. The proof over general fields F relies on the geometry-of-numbers machinery developed in the preprint [Bhargava et al. $\geq$ 2025], which has still not appeared. One can alternatively make use of the techniques in [Bhargava et al. 2015], again using Theorem 5.1 as a key input, to deduce the formula for general F. Note that in [Bhargava et al. 2015], the authors count cubic extensions of F with prescribed local conditions (ordered by discriminant), exactly by counting *integral* $\mathrm{SL}_2(F)$-orbits of binary cubic forms with prescribed local conditions.

The proof in the general case where $A[\phi](F) \neq \mathbb{Z}/3\mathbb{Z}$ is exactly the same except we identify elements of $\mathrm{Sel}_{\phi_d}(A_d)$ with orbits of binary cubic forms of discriminant dk, where $k \in \mathcal{O}_F$ is chosen so that $F(\sqrt{k}) = F(A[\phi])$, as is done in [Bhargava et al. 2019, §4]. □

Taking $\Sigma = \Sigma_0$ in Theorem 5.3, we deduce Theorem 1.8, and taking $\Sigma = \{d : d \in \mathcal{O}_v(2)\ \forall v\}$, we deduce Theorem 1.9.

5C. *Explicit Selmer rank bounds.* The following consequence of Theorem 5.2 will be helpful in giving explicit average rank bounds.

Proposition 5.4. *Let $\phi : A \to B$, Σ, and T_k be as in Theorem 5.2. Then*:

(i) *For each $d \in F^\times$, we have*

$$c(\phi_d) = \frac{\#\,\mathrm{Sel}(\phi_d)}{\#\,\mathrm{Sel}(\hat{\phi}_d)} \cdot \frac{\#\widehat{B}_d[\hat{\phi}_d](F)}{\#A_d[\phi_d](F)}.$$

(ii) *If T_k is nonempty, then it has positive density and*

$$\operatorname*{avg}_{d \in T_k} \dim_{\mathbb{F}_3} \mathrm{Sel}(\phi_d) \oplus \mathrm{Sel}(\hat{\phi}_d) \leq |k| + 3^{-|k|}.$$

(iii) *For a proportion of at least $1 - \frac{1}{2 \cdot 3^{|k|}}$ of $d \in T_k$, we have $\dim_{\mathbb{F}_3} \mathrm{Sel}(\phi_d) \oplus \mathrm{Sel}(\hat{\phi}_d) = |k|$.*

Proof. Since ϕ and $\hat{\phi}$ determine dual local conditions in their respective Selmer groups [Česnavičius 2017, B.1], the Greenberg–Wiles formula [Neukirch et al. 2000, 8.7.9] applies and gives (i). The argument for (ii) is then exactly the same as [Bhargava et al. 2020, Theorem 50].

For (iii), we may assume $k \geq 0$, by switching ϕ and $\hat{\phi}$ if necessary. By (i), we have

$$\dim \mathrm{Sel}(\phi_d) + \dim \mathrm{Sel}(\hat{\phi}_d) = k \quad \text{if and only if} \quad \dim \mathrm{Sel}(\hat{\phi}_d) = 0,$$

at least for the 100% of d such that $\#\widehat{B}_d[\hat{\phi}_d](F) = \#A_d[\phi_d](F) = 1$. By Theorem 5.2, the average size of $\mathrm{Sel}(\hat{\phi}_d)$ for $d \in T_k$ is $1 + 3^{-k}$. Hence, if s_0 is the lim inf of the natural density of $d \in T_k$ with $\#\,\mathrm{Sel}(\hat{\phi}_d) = 1$, then

$$s_0 + 3(1 - s_0) \leq 1 + 3^{-k},$$

and hence $s_0 \geq 1 - \frac{1}{2 \cdot 3^{|k|}}$. □

Proposition 5.5. *Let $\phi : A \to B$ and T_k be as in Theorem 5.2, with Σ the set of squarefree twists. Let S be the set of places of F dividing $3\mathfrak{f}_A\infty$, where $\mathfrak{f}_A$ is the conductor of A. Then $T_k = \varnothing$ if $|k| > \#S$. As a consequence, we have* $\operatorname{avg}_\Sigma \dim_{\mathbb{F}_3} \operatorname{Sel}(\phi_d) \oplus \operatorname{Sel}(\hat{\phi}_d) \le \#S + 3^{-\#S}$.

Proof. If $v \nmid 3$ is a prime of good reduction, then by Theorem 4.11 and the assumption that d is squarefree, $c_v(\phi_d) = 1$. On the other hand, directly from the definition, we see that $c_v(\phi_d) \ge \frac{1}{3}$ for any v. Hence,

$$c(\phi_d) = \prod_{v \mid 3\mathfrak{f}_A\infty} c_v(\phi_d) \ge \prod_{v \mid 3 f_A \infty} \frac{1}{3} = 3^{-\#S}.$$

For almost all $d \in \Sigma$, we have $A_d[\phi_d](F) = 0 = \widehat{B}_d[\hat{\phi}_d](F)$, in which case Proposition 5.4(i) gives

$$c(\phi_d)c(\hat{\phi}_d) = \frac{\#\operatorname{Sel}(\hat{\phi}_d)}{\#\operatorname{Sel}(\phi_d)} \cdot \frac{\#\operatorname{Sel}(\phi_d)}{\#\operatorname{Sel}(\hat{\phi}_d)} = 1.$$

By the above, we also have $c(\hat{\phi}_d) \ge 3^{-\#S}$, and hence $c(\phi_d) \le 3^{\#S}$. Since $3^{-\#S} \le c(\phi_d) \le 3^{\#S}$, we have $T_k = \varnothing$ if $|k| > \#S$. The second claim now follows from Proposition 5.4(ii). □

6. The average rank is bounded in cyclotomic twist families

In this section, we apply Theorems 1.8 and 1.9 to prove Theorems 1.1 and 1.3, i.e., to bound the average Mordell–Weil rank of $A_d(F)$.

Proof of Theorem 1.1(ii). Since $A[\pi]$ admits a full flag, by Lemma 2.4, so does $A^{(i)}[\pi]$ for each $i = 1, \dots, 2 \cdot 3^{m-1} - 1$. Hence, for each i, there is a sequence of ζ-linear 3-isogenies

$$A^{(i)} = B_0^{(i)} \xrightarrow{\phi_1^{(i)}} B_1^{(i)} \to \cdots \to B_{k-1}^{(i)} \xrightarrow{\phi_k^{(i)}} B_k^{(i)} = A^{(i+1)}.$$

By Theorem 1.9, for each i, j, the average rank of $\operatorname{Sel}(\phi_{j,d}^{(i)})$, for squarefree $d \in F^\times/F^{\times 2n}$ is bounded.

Recall [Bhargava et al. 2019, Lemma 9.1], that if $\psi_1 : A_1 \to A_2$ and $\psi_2 : A_2 \to A_3$ are isogenies of abelian varieties, then there is an exact sequence

$$\operatorname{Sel}_{\psi_1}(A_1) \to \operatorname{Sel}_{\psi_2\circ\psi_1}(A_1) \to \operatorname{Sel}_{\psi_2}(A_2). \tag{6-1}$$

Hence, for each d, we have

$$\begin{aligned} \operatorname{rk} A_d(F) \le \dim_{\mathbb{F}_3} \frac{A_d(F)}{3A_d(F)} &\le \dim_{\mathbb{F}_3} \operatorname{Sel}_3(A_d) \\ &\le \sum_{i=0}^{2\cdot 3^{m-1}-1} \sum_{j=1}^{k} \dim_{\mathbb{F}_3} \operatorname{Sel}(\phi_{j,d}^{(i)}) \\ &\le \sum_{i=0}^{2\cdot 3^{m-1}-1} \sum_{j=1}^{k} \#\operatorname{Sel}(\phi_{j,d}^{(i)}). \end{aligned}$$

Here, the second inequality follows inductively from (6-1), and the third inequality follows from the trivial inequality $k \leq 3^k$ for all $k \in \mathbb{Z}$. Taking the average over d, the result follows from Theorem 1.9. □

In the remainder of this section, we prove Theorem 1.1(i). The proof just given does not apply: by assumption, $B_0^{(i)}[\phi_1^{(i)}]$ is a direct summand of $A^{(i)}[\pi]$, however, there is no reason that $B_{j-1}^{(i)}[\phi_j^{(i)}]$ should be a direct summand of $B_{j-1}^{(i)}[\pi]$. Hence, we cannot directly apply Theorem 1.8 to get the result. Instead, we exploit the fact that we can take the isogenies $\phi_j^{(i)}$ in any order, together with the following dual version of Theorem 1.8:

Corollary 6.1. *Let B, C be abelian varieties over F with ζ-multiplication and a ζ-linear* 3*-isogeny $\phi : B \to C$ defined over F. Suppose that $\widehat{C}[\hat{\phi}]$ is almost everywhere locally a direct summand of $\widehat{C}[\pi]$. Then the average size of the Selmer group* $\mathrm{Sel}(\phi_d)$ *is bounded.*

Proof. Let $\hat{\phi} : \widehat{C} \to \widehat{B}$ be the dual isogeny, which is itself a ζ-linear 3-isogeny, with respect to the natural ζ-multiplication structure on $\widehat{C}$ and $\widehat{B}$. Let $\mathbb{B} = \mathbb{B}_{2n}$ as in Section 5. For each $d \in \mathbb{B}(F)$, Proposition 5.4(i) gives

$$\frac{\#\,\mathrm{Sel}(\phi_d)}{\#\,\mathrm{Sel}(\hat{\phi}_d)} = \frac{\# B_d[\phi_d](F)}{\#\widehat{C}_d[\hat{\phi}_d](F)} c(\phi_d).$$

The classes d for which $B_d[\phi_d](F) = \mathbb{Z}/3\mathbb{Z}$ or $\widehat{C}_d[\hat{\phi}_d](F) = \mathbb{Z}/3\mathbb{Z}$, form a union of at most two square-classes in $\mathbb{B}(F)$, so we can ignore these values of d when trying to bound the average size of $\mathrm{Sel}(\phi_d)$. In any case, up to a harmless factor of 3, the formula reads

$$\frac{\#\,\mathrm{Sel}(\phi_d)}{\#\,\mathrm{Sel}(\hat{\phi}_d)} = c(\phi_d).$$

By Theorem 1.8 applied to $\hat{\phi}$, the average size of $\#\,\mathrm{Sel}(\hat{\phi}_d)$ is bounded. In fact, if we restrict to the set $T_k = T_k(\hat{\phi}) \subset \mathbb{B}(F)$ where $c(\hat{\phi}_d) = 3^k$, then the average size is equal to $1 + 3^k$ by Theorem 5.2. Since $c(\phi_d) = 1/c(\hat{\phi}_d)$, the average size of $\mathrm{Sel}(\phi_d)$ is equal to $3^{-k} + 1$, and in particular converges, at least on this set T_k.

To see that the average size of $\#\,\mathrm{Sel}(\phi_d)$ converges (to the expected number) on all of $\mathbb{B}(F)$, we can argue as in [Bhargava et al. 2020, §6.4], using the uniformity estimate [Bhargava et al. 2013, Proposition 29] to give a tail bound for those binary cubic forms with large square part in their discriminant. The uniformity estimate applies to elements of $\#\,\mathrm{Sel}(\hat{\phi}_d)$ since, under the hypotheses of the corollary, they are represented by integral binary cubic forms. However, we need to bound the average size of $c(\phi_d)\#\,\mathrm{Sel}(\hat{\phi}_d)$ and not $\#\,\mathrm{Sel}(\hat{\phi}_d)$, so we also need to control the size of $c(\phi_d)$. By Table 3, we have $c(\phi_d) = O(3^m)$, where m is the number of primes v of F with $v(d)$ even and positive. Moreover, each element of $\mathrm{Sel}(\hat{\phi}_d)$ corresponds to a binary cubic form whose discriminant has norm divisible by q_v^2, for all such v. Since

$$3^m \prod_{v(d)=2} q_v^{-2} \gg 3q_{v_1}^{-2} \gg q_{v_1}^{-2},$$

the same argument as in [loc. cit.] goes through. We refer there for the details. □

Lemma 6.2. *Let F be any field and suppose there is a commutative diagram of isogenies of abelian varieties over F,*

$$\begin{tikzcd}
A \arrow[r, "\phi_1"] \arrow[d, "\phi_2"] & B_1 \arrow[d, "\psi_2"] \\
B_2 \arrow[r, "\psi_1"] & C
\end{tikzcd}$$

such that ϕ_2 maps $A[\phi_1]$ isomorphically onto $B_2[\psi_1]$. Then ψ_2 induces an injection

$$\frac{B_1(F)}{\phi_1(A(F))} \hookrightarrow \frac{C(F)}{\psi_1(B_2(F))}.$$

In particular, there is an embedding $\mathrm{Sel}_{\phi_1}(A) \hookrightarrow \mathrm{Sel}_{\psi_1}(B_2)$.

Proof. Taking cohomology, we obtain the following commutative diagram, with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & A[\phi_1](F) \arrow[r] \arrow[d, equal, "\phi_2"] & A(F) \arrow[r, "\phi_1"] \arrow[d, "\phi_2"] & B_1(F) \arrow[r, "\delta_1"] \arrow[d, "\psi_2"] & H^1(F, A[\phi_1]) \arrow[d, equal, "\phi_2"] \\
0 \arrow[r] & B_2[\psi_1](F) \arrow[r] & B_2(F) \arrow[r, "\psi_1"] & C(F) \arrow[r, "\partial_1"] & H^1(F, B_2[\psi_1])
\end{tikzcd}$$

Consider the composite of maps

$$B_1(F) \xrightarrow{\psi_2} C(F) \to \frac{C(F)}{\psi_1(B_2(F))}.$$

We show that the kernel is exactly $\phi_1(A(F))$ and, therefore, that there is an injection

$$\psi_2 : \frac{B_1(F)}{\phi_1(A(F))} \hookrightarrow \frac{C(F)}{\psi_1(B_2(F))},$$

from which the result follows. If $x \in A(F)$, then $\psi_2(\phi_1(x)) = \psi_1(\phi_2(x)) = 0$. Conversely, if $y \in B_1(F)$ and $\psi_2(y) = \psi_1(z)$ for some $z \in B_2(F)$, then

$$\phi_2(\delta_1(y)) = \partial_1(\psi_2(y)) = \partial_1(\psi_1(z)) = 0.$$

Since ϕ_2 is an isomorphism on cohomology, it follows that $\delta_1(y) = 0$ and, hence, that y is in the image of ϕ_1. □

Recall that $A[\pi]$ decomposes as a direct sum of characters, so that $A[\pi](\bar{F}) = \langle P_1, \dots, P_k\rangle$, and $\mathrm{Gal}(\bar{F}/F)$ stabilizes each of the subgroups $\langle P_i\rangle$. Thus, π factors as a product of ζ-linear 3-isogenies:

$$\begin{tikzcd}
A = B_0 \arrow[rrrr, "\pi"] \arrow[dr, "\phi_1"'] & & & & A^{(1)} = B_k \\
& B_1 = A/\langle P_1\rangle \arrow[r, "\phi_2"] & \cdots \arrow[r, "\phi_{k-1}"] & B_{k-1} = A/\langle P_1, \dots, P_{k-1}\rangle \arrow[ur, "\phi_k"'] &
\end{tikzcd}$$

Moreover, each of the maps $\phi_i : B_{i-1} \to B_i$ can be twisted to a map $\phi_{i,d} : B_{i-1,d} \to B_{i,d}$.

Corollary 6.3. *For each $i = 1, \dots, k$, the average size of* $\mathrm{Sel}_{\phi_{i,d}}(B_{i-1,d})$, *for $d \in \mathbb{B}(F)$, is bounded.*

Proof. The assumption that $A[\pi]$ decomposes as a sum of characters means that we can take the P_i's in any order. Hence, for each i, d, there is a commutative diagram

$$\begin{tikzcd}
B_{i-1,d} \arrow[r, "\phi_{i,d}"] \arrow[d] & B_{i,d} \arrow[d] \\
B'_{k-1,d} \arrow[r, "\psi_{i,d}"] & A_d^{(1)}
\end{tikzcd}$$

where

$$B'_{k-1,d} := \frac{A_d}{\langle P_1, \dots, P_{i-1}, P_{i+1} \dots, P_k \rangle}$$

and $\ker(\psi_{i,d}) = \ker(\phi_{i,d}) = \langle P_i \rangle$.

By Lemma 2.4, since $A[\pi]$ is completely reducible, so is $A^{(1)}[\pi]$ and hence so is $\hat{A}^{(1)}[\pi]$. Thus $\hat{A}^{(1)}[\hat{\psi}_{i,d}]$ is a direct summand of $\hat{A}^{(1)}[\pi]$. It follows from Corollary 6.1 that the average size of $\mathrm{Sel}_{\psi_{i,d}}(B'_{k-1,d})$ is bounded. By Lemma 6.2, we have embeddings $\mathrm{Sel}_{\phi_{i,d}}(B_{i-1,d}) \hookrightarrow \mathrm{Sel}_{\psi_{i,d}}(B'_{k-1,d})$ for each $d \in \mathbb{B}(F)$, so the average size of $\mathrm{Sel}_{\phi_{i,d}}(B_{i-1,d})$ is bounded as well. □

Proof of Theorem 1.1(i). By Lemma 2.4, since $A[\pi]$ is completely reducible, so is $A^{(i)}[\pi]$ for all i. Hence, we can factor $A^{(i)} \to A^{(i+1)}$ as

$$A^{(i)} = B_0^{(i)} \xrightarrow{\phi_1^{(i)}} B_1^{(i)} \to \dots \to B_{k-1}^{(i)} \xrightarrow{\phi_k^{(i)}} B_k^{(i)} = A^{(i+1)}.$$

By Corollary 6.3, for each i, j, the average rank of $\mathrm{Sel}(\phi_{j,d}^{(i)})$ is bounded. The result now follows exactly as in the proof of Theorem 1.1(ii). □

Proof of Theorem 1.3. The hypotheses imply that the Rosati involution associated to the polarization restricts to complex conjugation on the subring $\mathbb{Z}[\zeta]$. Thus, after identifying $A \simeq \hat{A}$ and $A_{-3^n} \simeq \hat{A}_{-3^n}$, we can factor multiplication by -3 on A_d as the composition

$$A_d \xrightarrow{\pi_d^{3^{m-1}}} A_d^{(3^{m-1})} = A_{-3^n d} \xrightarrow{\hat{\pi}_d^{3^{m-1}}} \hat{A}_d,$$

where the middle equality is Remark 2.8. As in the proof of Theorem 1.1(ii), we can factor $\pi_d^{3^{m-1}}$ as a product of $\dim A$ 3-isogenies $\phi_{j,d}$. By duality, $\hat{\pi}_d^{3^{m-1}}$ factors as the product of the dual isogenies $\hat{\phi}_{j,d}$. Thus, for each d, we have

$$\begin{aligned}
\operatorname{rk} A_d(F) &\le \dim_{\mathbb{F}_3} \frac{A_d(F)}{3A_d(F)} \le \dim_{\mathbb{F}_3} \mathrm{Sel}_3(A_d) \\
&\le \sum_{j=1}^{\dim A} \dim_{\mathbb{F}_3} \mathrm{Sel}(\phi_{j,d}) \oplus \mathrm{Sel}(\hat{\phi}_{j,d})
\end{aligned}$$

and hence, by Proposition 5.5,

$$\mathrm{avg}_d \operatorname{rk} A_d(F) \le \dim A \cdot (\#S + 3^{-\#S}). \qquad \square$$

7. The average rank in twist families of trigonal Jacobians

Next we use Theorem 1.1 to prove Corollary 1.2. In this section, F is a number field and $\zeta \in \bar{F}$ is a primitive 3^m-th of unity, for some $m \geq 1$. Let $n = 3^m$, as always.

7A. ***Trigonal Jacobians.*** Let $f(x) \in F[x]$ be a monic separable polynomial such that $f(0) \neq 0$, and let C be the smooth projective curve with affine model

$$C : y^3 = xf(x^{3^{m-1}}).$$

If $m > 1$, then C has a unique rational point ∞ at infinity, and has genus $g = 3^{m-1}\deg(f)$. In fact, we will assume $m > 1$, since the case $m = 1$ will be subsumed by the results of Section 7B.

Let $J = \mathrm{Jac}(C)$ be the Jacobian of C, a g-dimensional principally polarized abelian variety over F. The automorphism $(x, y) \mapsto (\zeta^3 x, \zeta y)$ induces an automorphism of $J_{\bar{F}}$ of order 3^m, which we again call ζ, and which endows J with ζ_n-multiplication. As in Section 2, the endomorphism $1 - \zeta \in \mathrm{End}_{\bar{F}}(J)$ descends to an isogeny $\pi : J \to J^{(1)}$ over F.

Lemma 7.1. *Let G be an extension of $(\mathbb{Z}/2\mathbb{Z})^k$ by a 3-group H. Then every irreducible representation $\rho : G \to \mathrm{GL}_N(\mathbb{F}_3)$ is one-dimensional. Consequently, any G-representation V over $\mathbb{F}_3$ admits a full flag.*

Proof. Since $H \lhd G$ is normal and ρ is semisimple, $\rho|_H$ is also semisimple. Now, any nontrivial representation V over $\mathbb{F}_3$ of a 3-group contains a nonzero fixed vector [Serre 1977, Proposition 26]. Thus, by semisimplicity and induction on $\dim V$, we see that $\rho|_H$ is trivial. Thus, ρ factors through a representation of $(\mathbb{Z}/2\mathbb{Z})^k$. For any $g \in (\mathbb{Z}/2\mathbb{Z})^k$, $\rho(g) \in \mathrm{GL}_N(\mathbb{F}_3)$ has order at most 2, so its minimal polynomial, either $X \pm 1$ or $X^2 - 1$, has distinct $\mathbb{F}_3$-rational roots. Hence, $\rho(g)$ is diagonalizable, and since $(\mathbb{Z}/2\mathbb{Z})^k$ is abelian, the operators $\rho(g)$ are simultaneously diagonalizable. In other words, ρ is a direct sum of characters. Since ρ is irreducible, it follows that ρ is one-dimensional. □

We first prove Corollary 1.2 for Jacobians of the curves $C : y^3 = xf(x^{3^{m-1}})$. In Theorem 7.4, we will address the curves $C : y^{3^m} = f(x)$.

Proof of Corollary 1.2. By assumption $\mathrm{Gal}(f)$ is an extension of $(\mathbb{Z}/2\mathbb{Z})^k$ by a 3-group H. By Theorem 1.1, it is enough to show that the Galois representation $J[\pi]$ has a full flag, and splits as a direct sum of characters if $H = 1$.

Lemma 7.2. *$J[\pi^{3^{m-1}}] = J[1 - \zeta_3]$ is a maximal isotropic $\mathbb{F}_3$-subspace of $J[3]$ of dimension g.*

Proof. Degree considerations show that $\dim J[1 - \zeta_3] = g$, so we need only show that $J[1 - \zeta_3]$ is isotropic with respect to the Weil pairing $J[3] \times J[3] \to \mu_3$. Now, the Rosati involution † sends the ideal $(1 - \zeta_3) = (\sqrt{-3})$ to itself (by Lemma 10.4 below), and $\langle \alpha P, Q\rangle = \langle P, \alpha^\dagger Q\rangle$ for all $\alpha \in \mathrm{End}(J_{\bar{F}})$ and $P, Q \in J[3]$. If $P, Q \in J[(\sqrt{-3})]$, we may write $Q = \sqrt{-3}(R)$, for some $R \in J[3]$. We then compute

$$\langle P, Q\rangle = \langle P, \sqrt{-3}(R)\rangle = \langle -\sqrt{-3}(P), R\rangle = 1,$$

showing that $J[1 - \zeta_3] = J[\sqrt{-3}]$ is isotropic. □

It follows that $\dim_{\mathbb{F}_3} J[\pi] = \deg(f)$. Explicitly, if α is a root of f and $\beta^{3^{m-1}} = \alpha$, then the divisor

$$(\beta, 0) + (\zeta^3\beta, 0) + (\zeta^{3\cdot 2}\beta, 0) + \cdots + (\zeta^{3^m-3}\beta, 0) - 3^{m-1}\infty$$

is fixed by ζ, and $J[\pi]$ is generated by the above divisor classes. Moreover, if K is the splitting field of f over F, then each of these divisors defines an element of $J(K)$.

The action of G_F on $J[\pi]$ induces a homomorphism $\rho : G_F \to \mathrm{GL}_N(\mathbb{F}_3)$, where $N = \deg(f)$, whose kernel is exactly G_K. The image is therefore isomorphic to $\mathrm{Gal}(K/F)$, which is an extension of $(\mathbb{Z}/2\mathbb{Z})^k$ by a 3-group H. Hence, by Lemma 7.1, $J[\pi]$ admits a full flag. If, moreover, $H = 1$, then $J[\pi]$ is completely reducible, as explained in the proof of Lemma 7.1. □

7B. ***Iterated triple covers and Pryms.*** For our second class of Jacobians, let $f(x) \in F[x]$ be a monic separable polynomial of degree $N > 1$. Let C be the smooth projective curve with affine model $y^{3^m} = f(x)$. The degree $n = 3^m$ map $C \to \mathbb{P}^1$, sending $(x, y) \mapsto x$, has Galois group μ_n, at least over $\bar{F}$. If $3 \nmid N$, then the fiber above infinity is a unique F-rational point and C has genus $g = \frac{1}{2}(N-1)(3^m-1)$. If $3 \mid N$, then the fiber above infinity may have more than one point and they may not be F-rational.

Let $J = \mathrm{Jac}(C)$ be the Jacobian. The order 3^m automorphism $(x, y) \mapsto (x, \zeta y)$ of C induces an automorphism ζ on J. When $m = 1$, this endows J with ζ_3-multiplication, and we are in a trigonal situation similar to Section 7A. However, if $m \geq 2$, the automorphism $\zeta \in \mathrm{Aut}_{\bar{F}}(J)$ does not give rise to a ζ_n-multiplication on J, as we have defined it in this paper.

Example 7.3. Consider the curve $C : y^9 = x^2 - 1$ of genus 4. This admits a degree three map to the elliptic curve $E : y^3 = x^2 - 1$, and the Jacobian $J = \mathrm{Jac}(C)$ is isogenous to $A \times E$, for some abelian 3-fold $A \subset J$. The order 9 automorphism ζ induces an order 9 automorphism on A and an order 3 automorphism on E. It thereby endows A with ζ_9-multiplication and E with ζ_3-multiplication, but it does *not* give a ζ_9-multiplication on J. Indeed, the minimal polynomial for $\zeta \in \mathrm{End}_{\bar{F}} J$ is $\Phi_9(x)\Phi_3(x)$ and not $\Phi_9(x) = x^6 + x^3 + 1$.

While J does not admit ζ_n-multiplication, it is nonetheless isogenous to a product of abelian varieties $P_1 \times P_2 \times \cdots \times P_m$ where each P_i has ζ_{3^i}-multiplication (see Lemma 7.5). In any case, we have $\mu_{2n} \subset \mathrm{Aut}_{\bar{F}} J$, and we may speak of the twists J_d, for each $d \in F^\times/F^{\times 2n}$.

Theorem 7.4. *Assume that* $\mathrm{Gal}(f)$ *is an extension of* $(\mathbb{Z}/2\mathbb{Z})^k$ *by a* 3*-group* H. *Then the average rank of the twists* J_d *for squarefree* $d \in F^\times/F^{\times 2n}$ *is bounded. If* $H = 1$, *then the average rank of the twists* J_d, *for all* $d \in F^\times/F^{\times 2n}$ *is bounded.*

Proof. Let $C' : y^{3^{m-1}} = f(x)$, and let J' be its Jacobian. Note that when $m = 1$, we have $C' \simeq \mathbb{P}^1$ and $J' = 0$. The map $q : C \to C'$ sending $(x, y) \mapsto (x, y^3)$ induces a surjection $q_* : J \to J'$, and we let P be the identity component of the kernel, i.e., P is the (generalized) Prym variety for the cover q. Since q is a ramified triple cover, the map $q^* : J' \to J$ is injective. We may identify $q^* = \hat{q}_*$, and it follows that the kernel of q_* is already connected, and hence equal to P.

Lemma 7.5. *P admits ζ_{3^m}-multiplication.*

Proof. It is enough to show that the minimal polynomial for ζ_{3^m}, as an endomorphism of P, is $\Phi_{3^m}(x) = x^{2\cdot 3^{m-1}} + x^{3^{m-1}} + 1$. For this, it is enough to show that the characteristic polynomial of ζ_{3^m}, acting on the homology lattice $H_1(P_{\mathbb{C}}, \mathbb{Z})$ is $\Phi_{3^m}(x)^{N-1}$. By [Arul 2021, Lemma 3.16], the characteristic polynomial of ζ_{3^m} acting on $H_1(C_{\mathbb{C}}, \mathbb{Z}) \simeq H_1(J_{\mathbb{C}}, \mathbb{Z})$ is $(1 + x + x^2 + \cdots + x^{3^m-1})^{N-1}$. The claim now follows, by induction on m. □

Let $\pi : P \to P^{(1)}$ be the isogeny over F descending $1-\zeta$ over $F(\zeta)$, as usual. Note that $P[\pi] \subset P[3]$ since P has ζ_{3^m}-multiplication, whereas $J[1-\zeta]$ is not in general contained in $J[3]$.

Lemma 7.6. *The representation $G_F \to \mathrm{Aut}_{\mathbb{F}_3} P[\pi]$ factors through* $\mathrm{Gal}(f)$.

Proof. If $\alpha_1, \dots, \alpha_N$ are the roots of $f(x)$, then the divisor classes $(\alpha_i, 0) - (\alpha_j, 0)$ generate the group $J[1-\zeta]$ [Schaefer 2018, §3] and so the G_F-action on $P[\pi] \subset J[1-\zeta]$ factors through $\mathrm{Gal}(f)$. □

Now we finish the proof of Theorem 7.4. By Lemmas 7.1 and 7.6, the Galois module $P[\pi]$ has a full flag, and is completely reducible if $H = 1$. Thus Theorem 1.1 says that the average rank of P_d, for squarefree $d \in F^\times/F^{\times 2\cdot 3^m}$, is bounded (and without the squarefree condition if $H = 1$). Up to isogeny, we have $J_d \simeq P_d \times J'_d$, where J'_d is the twist of J' (which has $\mu_{2\cdot 3^{m-1}}$ twists) by the image of d under $F^\times/F^{\times 2\cdot 3^m} \to F^\times/F^{\times 2\cdot 3^{m-1}}$. By induction, the average rank of J'_d, for $d \in F^\times/F^{\times 2\cdot 3^m}$ is bounded. (We view the family J'_d over $\mathbb{B}_{2\cdot 3^m}$, instead of the more natural $\mathbb{B}_{2\cdot 3^{m-1}}$, but the same proof works for this slightly "redundant" family as well.) Thus the average rank of J_d is bounded. □

Remark 7.7. Combining the two families considered in this section, we obtain similar results for the curves $C_{k,j} : y^{3^k} = xf(x^{3^j})$. The Jacobian $\mathrm{Jac}(C_{k,j})$ is isogenous to $\prod_{r=1}^{k} P_{r,j}$, where each $P_{r,j} = \mathrm{Prym}(C_{r,j} \to C_{r-1,j})$ is a generalized Prym variety with $\zeta_{3^{r+j}}$-multiplication.

8. CM abelian varieties

Next, we prove Theorem 1.4. Let $\zeta = \zeta_{3^m}$ be a primitive 3^m-th root of unity. We recall our definition of complex multiplication:

Definition 8.1. An abelian variety A over a number field F *has complex multiplication by* $\mathbb{Z}[\zeta]$ if A has dimension 3^{m-1} and a G_F-equivariant ring embedding $\mathbb{Z}[\zeta] \hookrightarrow \mathrm{End}_{\bar{F}} A$.

Proof of Theorem 1.4. Set $g = 3^{m-1} = \dim A$. The assumption $\mathbb{Q}(\zeta) \subset F$ ensures that $A \simeq A^{(1)}$ by Lemma 2.3. Hence, we can view the 3-isogeny $\pi : A \to A^{(1)}$ as an endomorphism of A, and we have $\pi^{2g} = 3u$ for some automorphism u of A. By the multiplicativity of the global Selmer ratio [Shnidman 2021, Corollary 3.5], we have

$$c(\pi_d)^{2g} = c(\pi_d^{2g}) = c([3]u) = c([3]). \tag{8-1}$$

We claim that $c([3]) = 1$. If v is an infinite prime, then since $F \supset \mathbb{Q}(\zeta)$, we have $F_v \simeq \mathbb{C}$ and $c_v([3]) = \#A[3](\mathbb{C})^{-1} = 3^{-2g}$. If $v \nmid 3$ is a finite prime then $c_v([3]) = c_v(A_d)/c_v(A_d) = 1$. Finally, $\prod_{v|3} c_v([3]) = 3^{g[F:\mathbb{Q}]}$, by [Shnidman 2021, Proposition 3.1]. Let $[F : \mathbb{Q}] = 2N$. Then F has N complex places, so $c([3]) = 3^{-2gN} \cdot 3^{g\cdot 2N} = 1$, as claimed. By (8-1), we also have $c(\pi_d) = 1$ for all $d \in F^\times/F^{\times 6g}$.

It follows from Theorem 5.2 that the average size of $\mathrm{Sel}_{\pi_d}(A_d)$ is $1+1=2$. Since $2r \leq 3^r - 1$ for all integers r, we have for all d:

$$\mathrm{rk}_{\mathbb{Z}[\zeta]} A_d(F) \leq \dim_{\mathbb{F}_3} \mathrm{Sel}_{\pi_d}(A_d) \leq \frac{1}{2}(3^{\mathrm{rk}\,\mathrm{Sel}_{\pi_d}(A_d)} - 1) = \frac{1}{2}(\#\,\mathrm{Sel}_{\pi_d}(A_d) - 1).$$

Thus, the average $\mathbb{Z}[\zeta]$-rank of $A_d(F)$ is at most $\frac{1}{2}$. The second part of the theorem follows from Proposition 5.4(ii). □

Over general number fields F, it is no longer true that $c(\pi_d) = 1$ for all d, even for abelian varieties with complex multiplication. Moreover, one must consider more than one 3-isogeny to bound the average rank of $A_d(F)$, in general. However, one can still prove upper bounds which are significantly stronger than Theorem 1.3. For example, for CM abelian varieties A of dimension $g = 3^{m-1}$ over $\mathbb{Q}$ with ζ_{3^m}-multiplication, we can show that the average rank of $A_d(F)$ is at most $\frac{13}{9}g$. Over the totally real field $\mathbb{Q}(\zeta_{3^m} + \bar{\zeta}_{3^m})$, we can also prove that an explicit positive proportion of twists have $\mathrm{rk}\, A_d(F) = 0$. We omit these proofs, since they are somewhat technical and could probably be optimized further. Finally, we remark that the only other result in the literature in this direction that we are aware of is [Diaconu and Tian 2005], which proves that an infinite but density zero set of twists of the degree p Fermat Jacobian over $\mathbb{Q}(\zeta_p + \bar{\zeta}_p)$ have rank 0.

9. Rational points on hyperbolic varieties

Proof of Theorem 1.5. It is not enough to simply invoke Theorem 1.1 and [Kühne 2021, Theorem 4], since $\mathrm{Jac}(C_d)$, is not the d-th sextic twist of $J = \mathrm{Jac}(C)$ in our sense. Indeed, the twists C_d come from the involution $\tau(x, y) = (-x, y)$, which does not induce -1 on $\mathrm{Jac}(C)$. Instead, we consider the Prym variety $P = \ker(\mathrm{Jac}(C) \to \mathrm{Jac}(C/\tau))$, and its dual $\widehat{P}$. Note that ζ preserves P, and τ restricts to -1 on P. Thus, Theorem 1.1 applies to the twist family P_d, for $d \in F^\times/F^{\times 6}$, and it follows that $\mathrm{avg}_d\, \mathrm{rk}\, \widehat{P_d}(F) = \mathrm{avg}_d\, \mathrm{rk}\, P_d(F)$ is bounded.

The inclusion $P_d \hookrightarrow \mathrm{Jac}(C_d)$ induces a surjection $q : \mathrm{Jac}(C_d) \twoheadrightarrow \widehat{P_d}$. Suppose that $C_d(F)$ is nonempty and that $z_0 \in C_d(F)$. Then composing with the Abel–Jacobi map $C_d \hookrightarrow \mathrm{Jac}(C_d)$, using z_0 as base point, we obtain a map $j : C_d \to \widehat{P_d}$. We prove that j is injective on points, except possibly at the fixed points of τ (the points where $x = 0$ and the point(s) at infinity). If $j(w) = j(z)$, then $w - z$ is the pullback of a divisor on C_d/τ. Thus $\tau(w - z) \equiv w - z$, and so $\tau(w) + z \equiv w + \tau(z)$. But if D is a divisor of degree 2 on a nonhyperelliptic curve C, then by Riemann–Roch, we have $h^0(D) = 2 + 1 - g + h^0(K - D) = 3 - g + g - 2 = 1$. Thus, we must have $\tau(w) + z = w + \tau(z)$. If $\tau(w) = \tau(z)$, then $w = z$, as desired. The only other possibility is that $\tau(w) = w$ and $\tau(z) = z$. This proves the claimed injectivity.

Thus, to prove Theorem 1.5, we may replace C_d with the image of j, which is a closed irreducible curve in $\widehat{P_d}$. By [Gao et al. 2021, Theorem 1.1], there is a constant c such that $\#C_d(F) \leq c^{1+\mathrm{rk}\,\widehat{P_d}(F)}$, for all d. But since $\mathrm{avg}_d\, \mathrm{rk}\, \widehat{P_d}(F)$ is bounded, this implies that for all $\varepsilon > 0$, there exists N_ε such that $C_d(F) \leq N_\varepsilon$ for at least $1 - \varepsilon$ of twists d. □

In order to set up the proof (and statement) of Theorem 1.6, we need to give a precise description of theta divisors for the curves C with affine model $y^3 = f(x)$. Recall that for any smooth projective curve C/F of genus $g \geq 2$, the symmetric power $C^{(g-1)}$ parametrizes effective divisors D on C of degree $g-1$. Given $\kappa \in \mathrm{Div}(C)$ of degree $g-1$, there is a morphism $C^{(g-1)} \to \mathrm{Jac}(C)$ sending $D \mapsto D - \kappa$. Its image is the theta divisor, denoted $\Theta = \Theta_\kappa$. The divisor itself depends on κ, though its class in the Néron–Severi group of $\mathrm{Jac}(C)$ does not. If 2κ is canonical, then Θ is preserved by the involution -1 on $\mathrm{Jac}(C)$, by Riemann–Roch. Such a κ exists over $\bar{F}$, but need not exist over F, in general. If in addition there exists $\mu_n \subset \mathrm{Aut}_{\bar{F}}(C)$ which fixes κ, then for each $d \in F^\times/F^{\times 2n}$, we may consider the twist Θ_d of Θ, which is a divisor in $\mathrm{Jac}(C)_d$.

In our case, let $f : C \to \mathbb{P}^1$ be the degree three map $(x, y) \mapsto x$. The ramification divisor has the form $2D$, and satisfies $K_C = f^*K_{\mathbb{P}^1} + 2D$. Hence $\kappa = D - f^*0$ is a half-canonical divisor (over F) which is fixed by the μ_3-action. We therefore obtain sextic twists $\Theta_d \subset \mathrm{Jac}(C)_d$, for each $d \in F^\times/F^{\times 6}$, as in the statement of Theorem 1.6.

Proof of Theorem 1.6. This now follows from Theorem 1.1 and [Gao et al. 2021, Theorem 1.1]. □

10. Abelian surfaces with ζ_3-multiplication

For our final application, we study a family of abelian surfaces with ζ_3-multiplication, arising as Prym varieties. We prove results on the Mordell–Weil groups in sextic twist families of such surfaces, and give applications to explicit uniform bounds on rational points in sextic twist families of bielliptic trigonal curves of genus 3 (Theorem 1.7). For some recent results on rank statistics in larger families of Prym surfaces, see [Laga 2023].

Let F be a number field and let $f(x) = x^2 + ax + b \in F[x]$ be a quadratic polynomial with nonzero discriminant and $b \neq 0$. Then $y^3 = f(x^2) = x^4 + ax^2 + b$ is an affine model of a smooth projective plane quartic curve C. Note the double cover $\pi : (x, y) \mapsto (x^2, y)$ to the elliptic curve $E : y^3 = f(x)$. We refer to these genus-3 curves as *bielliptic Picard curves*; see [Laga and Shnidman 2023]. As in Section 9, we consider the Prym variety $P = \mathrm{Prym}_{C/E}$, i.e., the kernel of the map $J = \mathrm{Jac}(C) \to E$ induced by Albanese functoriality. The Prym P need not be principally polarized over $\mathbb{Q}$, but it admits a polarization $\lambda : P \to \widehat{P}$ whose kernel is order 4 [Mumford 1974]. The ζ_3-multiplication on J induces ζ_3-multiplication on P, and hence we may speak of the sextic twists P_d. In fact, P_d is itself the Prym variety of $C_d : y^3 = x^4 + adx^2 + bd^2$, which covers the elliptic curve $E_d : y^3 = x^2 + adx + bd^2$.

Lemma 10.1. *Let $\pi : P \to P_{-27}$ denote the descent of $1 - \zeta$ to F. Then $P[\pi](\bar{F}) \simeq (\mathbb{Z}/3\mathbb{Z})^2$ is spanned by $(s, 0) - (-s, 0)$ and $(t, 0) - (-t, 0)$, where $\pm s$, $\pm t$ are the four roots of $f(x^2)$.*

In order to apply our result to P, we assume that $f(x)$ has linear factors over F, so that $P[\pi]$ decomposes as a direct sum of two 1-dimensional Galois modules, corresponding to the quadratic characters $G_F \to \mathbb{F}_3^\times$ cut out by the fields $F(s)$ and $F(t)$. Then Theorem 1.1 says that the average rank of $P_d(F)$ is bounded, and it is interesting to ask whether there is some positive proportion of d with $\mathrm{rk}\, P_d(\mathbb{Q}) \leq 1$, so that we may apply the Chabauty method. We do not quite prove that such a positive proportion of d exists for general Pryms of this type, but we can prove it in seemingly any given example with the help of explicit computations. We demonstrate the idea by proving the following result stated in the introduction:

Theorem 1.7. *Consider the sextic twist family $C_d : y^3 = (x^2 - d)(x^2 - 4d)$ of genus-3 curves. For at least $\frac{1}{3}$ of squarefree $d \in \mathbb{Z}$ such that $d \equiv 2$ or $11 \pmod{36}$, we have $\#C_d(\mathbb{Q}) \leq 5$.*

We will use the following variant of Chabauty's method:

Theorem 10.2 (Stoll). *Let C be a smooth projective curve of genus $g \geq 2$ over a number field F, and let H be a G_F-stable subgroup of $\operatorname{Aut}_{\overline{F}} C$. Embed $C \hookrightarrow \operatorname{Jac}(C)$ using any positive degree H-invariant divisor as basepoint. Suppose there is a quotient B of $\operatorname{Jac}(C)$ such that the composition $\iota : C \hookrightarrow \operatorname{Jac}(C) \to B$ is an embedding, and suppose there exists $H \hookrightarrow \operatorname{Aut}_{\overline{F}} B$, compatible with the H-action on C, via ι. Then for all but finitely many H-twists C_ξ with $\operatorname{rk} B_\xi(F) < \dim B$, we have*

$$\#C_\xi(F) \leq f_C(\operatorname{rk} B_\xi(F) + g - \dim B) + \#C_\xi^{\text{triv}}(F) + \#C_\xi^{\text{triv,non-tors}}(\overline{F}) \setminus C_\xi^{\text{triv}}(F).$$

Here, f_C is the explicit function defined in [Stoll 2006, §3] *and C_ξ^{triv} is the subscheme of points fixed by some nontrivial automorphism in H, and $C_\xi^{\text{triv,non-tors}}$ is the subscheme of trivial points which map to nontorsion points of B.*

Proof. This is a straightforward generalization of [Stoll 2006, Theorem 5.1], which is the special case where $B = \operatorname{Jac}(C)$. (We have stated an ineffective version of the result, which is sufficient for our purposes. This is what allows us to use the function f_C as opposed to Stoll's $\tilde{f}_C$.) The proof is the same, except that instead of the nondegenerate pairing

$$\Omega(C/F) \times \operatorname{Jac}(C)(F) \otimes \mathbb{Q} \to F$$

used in [Stoll 2006, §6], we use the nondegenerate pairing $\Omega(C/F)^B \times B(F) \otimes \mathbb{Q} \to F$, where $\Omega(C/F)^B$ is the image of $\iota^* : \Omega(B/F) \to \Omega(C/F)$. □

We deduce Theorem 1.7 from Theorem 10.2 and the following theorem, whose proof will occupy the remainder of this section:

Theorem 10.3. *Let $C : y^3 = (x^2 - 1)(x^2 - 4)$, and let P be the corresponding Prym variety. Let Σ be the set of squarefree $d \in \mathbb{Z}$ such that $d \equiv 2$ or $11 \pmod{36}$.*[2] *Then the average rank of P_d, for $d \in \Sigma$, is at most $\frac{7}{3} \approx 2.33$. Moreover, for at least $\frac{1}{3}$ of $d \in \Sigma$, we have $\operatorname{rk} P_d \leq 1$.*

Proof of Theorem 1.7. The curve C embeds in $B = \widehat{P} = J/\pi^*E$, the dual of the Prym P; see [Barth 1987, 1.12]. We apply Theorem 10.2 to the cyclic group H of order six generated by $(x, y) \mapsto (-x, \zeta_3 y)$. We embed C in its Jacobian using the point $\infty = [0:1:0]$. Consulting [Stoll 2006, Lemma 3.1], we have $f_C(r) \leq 4$ when C is a plane quartic and $r \leq 2$. We have $C_d^{\text{triv}}(\mathbb{Q}) = \{\infty\}$ for all $d \in \Sigma$, so the second term in Theorem 10.2 is 1. The third term is 0 since all eight of the trivial points on C map to torsion points of B. Indeed, the points with $y = 0$ map to 3-torsion points on $\operatorname{Jac}(C_d)$, and if $P = (0, y_0) \in C_d$,

[2] Σ is the set of d such that $d, -3d \notin \mathbb{Q}_2^{\times 2} \cup \mathbb{Q}_3^{\times 2}$.

then $2P - 2\infty \in \pi^* E_d$; hence P is sent to a 2-torsion point on $B = J_d/\pi^* E_d$. Altogether we get a bound of $C_\xi(F) \leq 5$ in Stoll's theorem, which combined with Theorem 10.3 proves Theorem 1.7. □

10A. ***Bielliptic Picard curves.*** Before proving Theorem 10.3, we prove some preliminary lemmas.

Lemma 10.4. *Let (J, λ) be the Jacobian of a curve C with a nontrivial automorphism ζ inducing ζ-multiplication on J. Then the Rosati involution $\alpha \mapsto \lambda^{-1}\hat{\alpha}\lambda$ on* $\mathrm{End}(J)$ *restricts to complex conjugation on* $\mathbb{Z}[\zeta] \subset \mathrm{End}(J)$.

Proof. Let D_0 be a degree $g-1$ divisor fixed by ζ. Consider the theta divisor

$$\Theta = \{D - D_0 : \deg(D) = g-1,\ D \text{ effective}\} \subset J$$

and set $\mathcal{L} = \mathcal{O}_J(\Theta)$. We have $\lambda = \varphi_{\mathcal{L}} : J \to \hat{J}$. Since Θ is fixed by ζ, we have $\zeta^*\mathcal{L} \simeq \mathcal{L}$ and hence $\varphi_{\mathcal{L}} = \varphi_{\zeta^*\mathcal{L}} = \hat{\zeta}\varphi_{\mathcal{L}}\zeta$. Rearranging, we see that the Rosati involution sends ζ to $\zeta^{-1} = \bar{\zeta}$. □

Remark 10.5. The proof shows, more generally, that if α is an automorphism of a curve C, and α^* is the induced automorphism of $\mathrm{Jac}(C)$, then the Rosati involution sends α^* to its inverse.

Now let $C : y^3 = x^4 + ax^2 + b$ be a bielliptic Picard curve defined over $\mathbb{Q}$. Let P be the Prym surface for the covering $C \to E$ where $E : y^3 = x^2 + ax + b$. Since P has ζ_3-multiplication, the endomorphism $[-3] : P \to P$ factors as $[-3] = \pi_{-27} \circ \pi$, where $\pi : P \to P_{-27}$ is the canonical $(3, 3)$-isogeny coming from Lemma 2.2 (see also Remark 2.8). Let $\pi_d : P_d \to P_{-27d}$ be the sextic twist family of $(3, 3)$-isogenies, and let $\hat{\pi}_d : \widehat{P}_{-27d} \to \widehat{P}_d$ denote the dual isogeny.

Lemma 10.6. $\mathrm{Sel}(\pi_{-27d}) \simeq \mathrm{Sel}(\hat{\pi}_d)$.

Proof. Let $C_d : y^3 = x^4 + adx^2 + bd^2$, let $E_d : y^3 = x^2 + adx + bd^2$, and let $J_d = \mathrm{Jac}(C_d)$.[3] The abelian variety P_d is, by definition, $\ker(J_d \to E_d)$, where the map $J_d \to E_d$ is induced by the double cover $C_d \to E_d$. Let λ_J denote the principal polarization of J_d, and let ζ_J be the automorphism of J_d induced by the map $(x, y) \mapsto (x, \zeta_3 y)$ on C_d. By Lemma 10.4 we have a commutative diagram

$$\begin{tikzcd}
P_d \arrow[r] \arrow[rrr, bend left, "\lambda_d"] \arrow[d, "\bar{\zeta}"] & J_d \arrow[r, "\lambda_J"] \arrow[d, "\bar{\zeta}"] & \hat{J}_d \arrow[r] \arrow[d, "\hat{\zeta}"] & \widehat{P}_d \arrow[d, "\hat{\zeta}"] \\
P_d \arrow[r] \arrow[rrr, bend right, "\lambda_d"'] & J_d \arrow[r, "\lambda_J"'] & \hat{J}_d \arrow[r] & \widehat{P}_d
\end{tikzcd}$$

[3] Note that this is not the same as the d-th sextic twist coming from the ζ-multiplication on J. The latter is isomorphic to the d-th quadratic twist of the Jacobian of $dy^3 = f(x^2)$, and is in general not a Jacobian.

It follows that $\zeta^{-1} = \lambda_d^{-1}\hat{\zeta}\lambda_d$ in $\mathrm{End}(P_d)$, and hence

$$(1-\hat{\zeta}^{-1})\circ\lambda_d = \lambda_d\circ(1-\zeta)$$

over $\bar{F}$. Over F we must therefore have $\hat{\pi}_{-27d}\circ\lambda_d = \lambda_{-27d}\circ\pi_d$, and since λ_d is prime-to-3, we deduce $\mathrm{Sel}(\pi_d)\simeq\mathrm{Sel}(\hat{\pi}_{-27d})$. □

Since $[-3] = \pi_{-27}\circ\pi$, it follows that

$$\mathrm{rk}(P_d)\le\dim_{\mathbb{F}_3}\mathrm{Sel}_3(P_d)\le\dim_{\mathbb{F}_3}(\mathrm{Sel}(\pi_d)\oplus\mathrm{Sel}(\pi_{-27d})) = \dim_{\mathbb{F}_3}\mathrm{Sel}(\pi_d)+\dim_{\mathbb{F}_3}\mathrm{Sel}(\hat{\pi}_d).$$

The following result relates the parity of $\dim_{\mathbb{F}_3}\mathrm{Sel}_3(P_d)$ to the global Selmer ration $c(\pi_d)$.

Proposition 10.7. *Let $d\in\mathbb{Q}^\times$ be such that $P_{-27d}[\pi_{-27d}](\mathbb{Q}) = 0$, and write $c(\pi_d) = 3^m$. Then we have the congruence* $\dim_{\mathbb{F}_3}\mathrm{Sel}_3(P_d)\equiv m \pmod 2$.

Proof. By the Greenberg–Wiles formula, we have $\#\mathrm{Sel}(\pi_d)/\#\mathrm{Sel}(\hat{\pi}_d) = c(\pi_d) = 3^m$. Since $[-3] = \pi_d\circ\pi_{-27d}$, we have an exact sequence

$$0\to\mathrm{Sel}(\pi_d)\to\mathrm{Sel}_3(P_d)\to\mathrm{Sel}(\pi_{-27d})\to\frac{\text{Ш}(P_{-27d})[\pi_{-27d}]}{\pi_d(\text{Ш}(P_d)[3])}\to 0.$$

Exactness on the left is because $P_{-27d}[\pi_{-27d}](\mathbb{Q}) = 0$. By Lemma 10.6, there is an isomorphism $\mathrm{Sel}(\pi_{-27d})\simeq\mathrm{Sel}(\hat{\pi}_d)$, so we see that

$$m\equiv\dim_{\mathbb{F}_3}\mathrm{Sel}_3(P_d)+\dim_{\mathbb{F}_3}\frac{\text{Ш}(P_{-27d})[\pi_{-27d}]}{\pi_d(\text{Ш}(P_d)[3])}\pmod 2. \tag{10-1}$$

Let

$$\langle\cdot,\cdot\rangle:\text{Ш}(P_{-27d})\times\text{Ш}(\widehat{P}_{-27d})\to\mathbb{Q}/\mathbb{Z}$$

be the Cassels–Tate pairing. Using the polarization $\lambda_{-27d}:P_{-27d}\to\widehat{P}_{-27d}$, define

$$\langle\cdot,\cdot\rangle_\lambda:\text{Ш}(P_{-27d})\times\text{Ш}(P_{-27d})\to\mathbb{Q}/\mathbb{Z}$$

by $\langle x,y\rangle_\lambda = \langle x,\lambda_{-27d}(y)\rangle$. As in [Shnidman 2021, Theorems 4.3, 4.4], if $x\in\text{Ш}(P_{-27d})[\pi_{-27d}]$, then y is in the image of $\pi_d:\text{Ш}(P_d)\to\text{Ш}(P_{-27d})$ if and only if $\langle x,\lambda_{-27d}(y)\rangle_\lambda = 0$ for all $y\in\text{Ш}(P_{-27d})[\pi_{-27d}]$. Thus, the Cassels–Tate pairing $\langle\cdot,\cdot\rangle_\lambda$ restricts to a nondegenerate paring on the finite group

$$\frac{\text{Ш}(P_{-27d})[\pi_{-27d}]}{\pi(\text{Ш}(P_d)[3])}.$$

Moreover, since both P_d and P_{-27d} are prime-to-3 polarized, this pairing is antisymmetric, and therefore alternating. The nondegeneracy implies that it has even $\mathbb{F}_3$-rank. Combining with (10-1), we deduce the desired congruence modulo 2. □

The following general lemma will be used to compute local Selmer ratios below.

Lemma 10.8. *Let $\alpha:A\to B$ be an isogeny of abelian varieties over a nonarchimedean characteristic* 0 *local field F. Then $c_\ell(\alpha)c_\ell(\hat{\alpha}) = \#(\mathcal{O}_F/\deg(\alpha)\mathcal{O}_F)$.*

Proof. By [Česnavičius 2017, B.1], the groups $B(F)/\alpha A(F)$ and $\hat{A}(F)/\hat{\alpha}\widehat{B}(F)$ are orthogonal complements under Tate–Shatz local duality

$$H^1(F, A[\alpha]) \times H^1(F, \widehat{B}[\hat{\alpha}]) \to \mathbb{Q}/\mathbb{Z}.$$

Thus

$$c_\ell(\alpha)c_\ell(\hat{\alpha}) = \frac{\#B(F)/\alpha A(F)}{\#A(F)[\alpha]} \cdot \frac{\#\hat{A}(F)/\hat{\alpha}\widehat{B}(F)}{\#\widehat{B}(F)[\hat{\alpha}]} = \frac{\#H^1(F, A[\alpha])}{\#A(F)[\alpha] \cdot \#\widehat{B}(F)[\hat{\alpha}]} = \#(\mathcal{O}_F/\deg(\alpha)\mathcal{O}_F),$$

where the final equality follows from the Euler–Poincaré characteristic formula. □

10B. ***The example.*** Now specialize to the context of Theorem 10.3 and the specific curves $C_d : y^3 = (x^2-d)(x^2-4d)$.

The isogeny $\pi : P \to P_{-27}$ factors as

$$P \xrightarrow{\phi} B \xrightarrow{\psi} P_{-27},$$

where $B = P/\langle(1,0)-(-1,0)\rangle$. Since $(1,0)-(-1,0) \in P[\pi]$, we obtain twists $\phi_d : P_d \to B_d$ and $\psi_d : B_d \to P_{-27d}$ by Lemma 2.6.

Let us compute the local Selmer ratios for ϕ_d and ψ_d, for all $d \in \Sigma$ (where Σ is as in Theorem 10.3). For any $d \in \mathbb{Q}^\times$, we have $c_\infty(\phi_d) = c_\infty(\psi_d)$, since the kernels of ϕ and ψ are both $\mathbb{Z}/3\mathbb{Z}$. Note that P has good reduction at all $p > 3$, since C does. Thus, for $d \in \Sigma$ and for all $p \nmid 6\infty$, by Theorem 4.11, we have $c_p(\phi_d) = 1 = c_p(\psi_d)$. If $p = 2$, then since $d \equiv 2, 3 \pmod 4$, neither d nor $-3d$ is a square in $\mathbb{Q}_2$, so by Table 2, $c_2(\phi_d) = 1 = c_2(\psi_d)$ as well. To compute the ratios $c_3(\phi_d)$ and $c_3(\psi_d)$, we use Lemma 10.8. By multiplicativity, we have

$$c_3(\phi_d)c_3(\psi_d)c_3(\phi_{-27d})c_3(\psi_{-27d}) = c_3([3]) = 9.$$

Since $d, -27d \notin \mathbb{Q}_3^{\times 2}$, all four of these local Selmer ratios are integers, and the same is true for the dual isogenies. Thus, by Lemma 10.8, each ratio must be either 1 or 3. Hence, of the four Selmer ratios $c_3(\phi_d), c_3(\psi_d), c_3(\phi_{-27d}), c_3(\psi_{-27d})$, exactly two are 3 and the other two are 1.

Lemma 10.9. *If $d \in \Sigma$, then $c_3(\phi_d) \neq c_3(\psi_d)$.*

Proof. By Proposition 10.7, the parity of $\dim_{\mathbb{F}_3} \mathrm{Sel}_3(P_d)$ is odd if and only if $c(\pi_d) = c(\phi_d)c(\psi_d)$ is an odd power of 3. Hence, by our local computations at all other primes $p \neq 3$ given above, the parity of $\dim_{\mathbb{F}_3} \mathrm{Sel}_3(P_d)$ is odd if and only if $c_3(\phi_d) \neq c_3(\psi_d)$. Since $\mathrm{Jac}(C_d)$ is prime-to-3-isogenous to $P_d \times E_d$, we have $\mathrm{Sel}_3(J_d) = \mathrm{Sel}_3(P_d) \oplus \mathrm{Sel}_3(E_d)$. Also, letting $K = \mathbb{Q}(\zeta_3)$ and $X \in \{J_d, P_d, E_d\}$, we have

$$\dim \mathrm{Sel}_3(X) \equiv \dim \mathrm{Sel}_\pi(X) + \dim \mathrm{Sel}_{\pi_{-27}}(X_{-27}) = \dim \mathrm{Sel}_\pi(X_K) \pmod 2,$$

where π is the map induced by $1-\zeta$ on divisors, and X_K is the base change to K. It follows that $c(\phi_d) \neq c(\psi_d)$ if and only if $\dim \mathrm{Sel}_\pi(J_{d,K}) - \dim \mathrm{Sel}_\pi(E_{d,K})$ is odd. The latter two π-Selmer groups can be computed in Magma [Bosma et al. 1997] for any choice of d, using the command `PhiSelmerGroup`.

In fact, it is enough to take $d = 2$, since the isomorphism class of A_d over $\mathbb{Q}_3$ depends only on the image of d in $\mathbb{Q}_3^\times/\mathbb{Q}_3^{\times 6}$, and all elements of Σ map to the sixth-power class of $d = 2$.[4] For $d = 2$, we find that $\dim \mathrm{Sel}_\pi(J_{d,K}) - \dim \mathrm{Sel}_\pi(E_{d,K}) = 1$. □

To recap, for $d \in \Sigma$, we have $c(\phi_d) = c_3(\phi_d)c_\infty(\phi_d)$ and $c(\psi_d) = c_3(\psi_d)c_\infty(\psi_d)$, and we know $c_\infty(\phi_d) = c_\infty(\psi_d)$ (which is equal to 1 or 1/3, depending on the sign of d) and $\{c_3(\phi_d), c_3(\psi_d)\} \subset \{1, 3\}$. Combining this with Lemma 10.9, we conclude that exactly one of $c(\phi_d)$ and $c(\psi_d)$ is 1 and the other is $3^{\pm 1}$.

Proof of Theorem 10.3. For all d, we have

$$\mathrm{rk}\, P_d(\mathbb{Q}) \le \mathrm{rk}(\mathrm{Sel}(\pi_d) \oplus \mathrm{Sel}(\hat{\pi}_d)) \le \mathrm{rk}(\mathrm{Sel}(\phi_d) \oplus \mathrm{Sel}(\hat{\phi}_d)) + \mathrm{rk}(\mathrm{Sel}(\psi_d) \oplus \mathrm{rk}\, \mathrm{Sel}(\hat{\psi}_d)).$$

For $d \in \Sigma$, we have seen that exactly one of $c(\phi_d)$ and $c(\psi_d)$ is 1 and the other is $3^{\pm 1}$. Thus, the average rank of $P_d(\mathbb{Q})$ for $d \in \Sigma$ is at most $\left(1 + \frac{4}{3}\right) = \frac{7}{3}$, by Proposition 5.4. This proves the first claim of Theorem 10.3.

Next we show that $\dim_{\mathbb{F}_3} \mathrm{Sel}_3(P_d) = 1$ for at least $\frac{1}{3}$ of $d \in \Sigma$. Without loss of generality we may assume that $c(\phi_d) = 1$ and $c(\psi_d) = 3^{\pm 1}$. By Proposition 5.4(iii), for at least $\frac{1}{2}$ of $d \in \Sigma$, we have $\mathrm{Sel}(\phi_d) = 0 = \mathrm{Sel}(\hat{\phi}_d)$, and for at least $\frac{5}{6}$ of $d \in \Sigma$, we have $\dim_{\mathbb{F}_3} \mathrm{Sel}(\psi_d) \oplus \mathrm{Sel}(\hat{\psi}_d) = 1$. Thus, for at least $\frac{5}{6} - \frac{1}{2} = \frac{1}{3}$ of $d \in \Sigma$, we have

$$\dim_{\mathbb{F}_3} \mathrm{Sel}_3(P_d) \le \dim \mathrm{Sel}(\phi_d) + \dim \mathrm{Sel}(\hat{\phi}_d) + \dim \mathrm{Sel}(\psi_d) + \dim \mathrm{Sel}(\hat{\psi}_d) \le 1.$$

This implies that $\mathrm{rk}\, P_d(\mathbb{Q}) \le 1$ for at least $\frac{1}{3}$ of $d \in \Sigma$. □

10C. ***More general curves.*** It is plausible that for *every* Prym P associated to some curve $C_{a,b} : y^3 = (x^2 - a)(x^2 - b)$, our method shows that $\mathrm{rk}\, P_d \le 1$ for a positive proportion of d. This holds if one can check a certain 3-adic condition on the numbers $c_3(\phi_d)$ and $c_3(\psi_d)$, exactly as in the proof of Lemma 10.9. This condition is satisfied in all examples we checked, but we do not have a proof in general. Since the local Selmer ratios $c_3(\phi_{a,b,d})$ and $c_3(\psi_{a,b,d})$ are locally constant as functions on $\mathbb{Q}_3^3 = \{(a, b, d)\}$, we can at least say that this condition holds for a large class of bielliptic Picard curves $C_{a,b}$, with a and b satisfying certain congruence conditions modulo a power of 3.

In [Shnidman and Weiss 2023], we prove that a positive proportion of P_d have rank at most 1, in the case where a/b is a square, using an extra argument which avoids the local 3-adic computation. In general, we have the following result, whose proof is an easier version of the argument given above, so we omit it.

Theorem 10.10. *Fix $a \in \mathbb{Q} \setminus \{0, \pm 1\}$. For $d \in \mathbb{Q}^\times/\mathbb{Q}^{\times 6}$, let $P_{a,d}$ be the Prym surface for the genus-*3 *curve $y^3 = (x^2 - d)(x^2 - ad)$. Then* $\mathrm{rk}\, P_{a,d}(\mathbb{Q}) \le 2$ *for a positive proportion of d.*

[4] To check this, use the fact that $\mathbb{Z}_3^{\times 6} = 1 + 9\mathbb{Z}_3$.

Acknowledgments

We are grateful to Nils Bruin, Jef Laga, Max Lieblich, Dino Lorenzini, Beth Malmskog, Zach Scherr, and Michael Stoll for helpful conversations and correspondences. We thank the referees for their helpful and detailed comments and corrections, and for suggesting to us cleaner proofs of Lemmas 3.8, 4.8, and 10.8. Shnidman was supported by the Israel Science Foundation (grant 2301/20). Weiss was supported by an Emily Erskine Endowment Fund postdoctoral fellowship at the Hebrew University of Jerusalem, by the Israel Science Foundation (grant 1963/20), and by the US-Israel Binational Science Foundation (grant 2018250).

References

[Arul 2021] V. Arul, "Division by $1-\zeta$ on superelliptic curves and Jacobians", *Int. Math. Res. Not.* **2021**:4 (2021), 3143–3185. MR Zbl

[Barth 1987] W. Barth, "Abelian surfaces with $(1, 2)$-polarization", pp. 41–84 in *Algebraic geometry* (Sendai, Japan, 1985), edited by T. Oda, Adv. Stud. Pure Math. **10**, North-Holland, Amsterdam, 1987. MR Zbl

[Bhargava 2004] M. Bhargava, "Higher composition laws, I: A new view on Gauss composition, and quadratic generalizations", *Ann. of Math.* (2) **159**:1 (2004), 217–250. MR Zbl

[Bhargava and Gross 2014] M. Bhargava and B. H. Gross, "Arithmetic invariant theory", pp. 33–54 in *Symmetry: representation theory and its applications*, edited by R. Howe et al., Progr. Math. **257**, Birkhäuser, New York, 2014. MR Zbl

[Bhargava and Ho 2016] M. Bhargava and W. Ho, "Coregular spaces and genus one curves", *Camb. J. Math.* **4**:1 (2016), 1–119. MR Zbl

[Bhargava and Shankar 2015a] M. Bhargava and A. Shankar, "Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves", *Ann. of Math.* (2) **181**:1 (2015), 191–242. MR Zbl

[Bhargava and Shankar 2015b] M. Bhargava and A. Shankar, "Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0", *Ann. of Math.* (2) **181**:2 (2015), 587–621. MR Zbl

[Bhargava et al. 2013] M. Bhargava, A. Shankar, and J. Tsimerman, "On the Davenport–Heilbronn theorems and second order terms", *Invent. Math.* **193**:2 (2013), 439–499. MR Zbl

[Bhargava et al. 2015] M. Bhargava, A. Shankar, and X. Wang, "Geometry-of-numbers methods over global fields, I: Prehomogeneous vector spaces", preprint, 2015. arXiv 1512.03035

[Bhargava et al. 2019] M. Bhargava, Z. Klagsbrun, R. J. Lemke Oliver, and A. Shnidman, "3-isogeny Selmer groups and ranks of abelian varieties in quadratic twist families over a number field", *Duke Math. J.* **168**:15 (2019), 2951–2989. MR Zbl

[Bhargava et al. 2020] M. Bhargava, N. Elkies, and A. Shnidman, "The average size of the 3-isogeny Selmer groups of elliptic curves $y^2 = x^3 + k$", *J. Lond. Math. Soc.* (2) **101**:1 (2020), 299–327. MR Zbl

[Bhargava et al. $\geq$ 2025] M. Bhargava, A. Shankar, and X. Wang, "Geometry-of-numbers methods over global fields, II: Coregular representations", in preparation.

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system, I: The user language", *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR Zbl

[Brumer 1992] A. Brumer, "The average rank of elliptic curves, I", *Invent. Math.* **109**:3 (1992), 445–472. MR Zbl

[Caro and Pasten 2023] J. Caro and H. Pasten, "A Chabauty–Coleman bound for surfaces", *Invent. Math.* **234**:3 (2023), 1197–1250. MR Zbl

[Česnavičius 2016] K. Česnavičius, "Selmer groups as flat cohomology groups", *J. Ramanujan Math. Soc.* **31**:1 (2016), 31–61. MR Zbl

[Česnavičius 2017] K. Česnavičius, "p-Selmer growth in extensions of degree p", *J. Lond. Math. Soc.* (2) **95**:3 (2017), 833–852. MR Zbl

[Delone and Faddeev 1940] B. N. Delone and D. K. Faddeev, "Theory of irrationalities of third degree", *Acad. Sci. URSS. Trav. Inst. Math. Stekloff*, **11** (1940), 3–340. In Russian. MR Zbl

[Diaconu and Tian 2005] A. Diaconu and Y. Tian, "Twisted Fermat curves over totally real fields", *Ann. of Math.* (2) **162**:3 (2005), 1353–1376. MR Zbl

[Dimitrov et al. 2021] V. Dimitrov, Z. Gao, and P. Habegger, "Uniformity in Mordell–Lang for curves", *Ann. of Math.* (2) **194**:1 (2021), 237–298. MR Zbl

[Gan et al. 2002] W. T. Gan, B. Gross, and G. Savin, "Fourier coefficients of modular forms on G_2", *Duke Math. J.* **115**:1 (2002), 105–169. MR Zbl

[Gao et al. 2021] Z. Gao, T. Ge, and L. Kühne, "The uniform Mordell–Lang conjecture", preprint, 2021. arXiv 2105.15085

[Heath-Brown 1994] D. R. Heath-Brown, "The size of Selmer groups for the congruent number problem, II", *Invent. Math.* **118**:2 (1994), 331–370. MR Zbl

[Katz and Sarnak 1999] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, Amer. Math. Soc. Colloq. Publ. **45**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

[Klagsbrun et al. 2013] Z. Klagsbrun, B. Mazur, and K. Rubin, "Disparity in Selmer ranks of quadratic twists of elliptic curves", *Ann. of Math.* (2) **178**:1 (2013), 287–320. MR Zbl

[Kühne 2021] L. Kühne, "Equidistribution in families of abelian varieties and uniformity", preprint, 2021. arXiv 2101.10272

[Laga 2023] J. Laga, "Arithmetic statistics of Prym surfaces", *Math. Ann.* **386**:1-2 (2023), 247–327. MR Zbl

[Laga and Shnidman 2023] J. Laga and A. Shnidman, "The geometry and arithmetic of bielliptic Picard curves", preprint, 2023. arXiv 2308.15297

[Levi 1914] F. Levi, "Kubische Zahlkörper und binäre kubische Formenklassen", *Sitz.ber. Sächs. Akad. Wiss. Leipz. Math.-Nat.wiss. Kl.* **66** (1914), 26–37. Zbl

[Mazur and Rubin 2007] B. Mazur and K. Rubin, "Finding large Selmer rank via an arithmetic theory of local constants", *Ann. of Math.* (2) **166**:2 (2007), 579–612. MR Zbl

[Milne 1986] J. S. Milne, *Arithmetic duality theorems*, Persp. Math. **1**, Academic Press, Boston, MA, 1986. MR Zbl

[Mumford 1974] D. Mumford, "Prym varieties, I", pp. 325–350 in *Contributions to analysis*, edited by L. V. Ahlfors et al., Academic Press, New York, 1974. MR Zbl

[Neukirch et al. 2000] J. Neukirch, A. Schmidt, and K. Wingberg, *Cohomology of number fields*, Grundl. Math. Wissen. **323**, Springer, 2000. MR Zbl

[Schaefer 2018] E. F. Schaefer, "Explicit descent for Jacobians of prime power cyclic covers of the projective line", *Trans. Amer. Math. Soc.* **370**:5 (2018), 3487–3505. MR Zbl

[Serre 1977] J.-P. Serre, *Linear representations of finite groups*, Grad. Texts in Math. **42**, Springer, 1977. MR Zbl

[Shnidman 2021] A. Shnidman, "Quadratic twists of abelian varieties with real multiplication", *Int. Math. Res. Not.* **2021**:5 (2021), 3267–3298. MR Zbl

[Shnidman and Weiss 2023] A. Shnidman and A. Weiss, "Rank growth of elliptic curves over n-th root extensions", *Trans. Amer. Math. Soc. Ser. B* **10** (2023), 482–506. MR Zbl

[Smith 2017] A. Smith, "2^{∞}-Selmer groups, 2^{∞}-class groups, and Goldfeld's conjecture", preprint, 2017. arXiv 1702.02325

[Stoll 2006] M. Stoll, "Independence of rational points on twists of a given curve", *Compos. Math.* **142**:5 (2006), 1201–1214. MR Zbl

[Thorne 2013] J. A. Thorne, "Vinberg's representations and arithmetic invariant theory", *Algebra Number Theory* **7**:9 (2013), 2331–2368. MR Zbl

Communicated by Melanie Matchett Wood
Received 2022-05-27 Revised 2023-12-12 Accepted 2024-01-22

ariel.shnidman@mail.huji.ac.il
Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel

weiss.742@osu.edu
Department of Mathematics, The Ohio State University, Columbus, OH, United States

Department of Mathematics, Ben-Gurion University of the Negev, Be'er Sheva, Israel

ALGEBRA AND NUMBER THEORY 19:1 (2025)
https://doi.org/10.2140/ant.2025.19.77

Picard rank jumps for K3 surfaces with bad reduction

Salim Tayou

Let X be a K3 surface over a number field. We prove that X has infinitely many specializations where its Picard rank jumps, hence extending our previous work with Shankar, Shankar and Tang to the case where X has bad reduction. We prove a similar result for generically ordinary nonisotrivial families of K3 surfaces over curves over $\overline{\mathbb{F}}_p$ which extends previous work of Maulik, Shankar and Tang. As a consequence, we give a new proof of the ordinary Hecke orbit conjecture for orthogonal and unitary Shimura varieties.

1. Introduction

Let X be a K3 surface over a number field K. Let $\mathscr{X} \to \mathscr{S}$ be a smooth projective model, where $\mathscr{S} \hookrightarrow \operatorname{Spec}(\mathcal{O}_K)$ is an open subset of the spectrum of the ring of integers $\mathcal{O}_K$. For every place $\mathfrak{P}$ of $\mathcal{O}_K$ with finite residual field $k(\mathfrak{P})$, we have an injective specialization map

$$\operatorname{Pic}(X_{\overline{K}}) \hookrightarrow \operatorname{Pic}(\mathscr{X}_{\overline{k(\mathfrak{P})}}),$$

and both groups have finite rank, *the Picard rank*, denoted $\rho(X_{\overline{K}})$ and $\rho(\mathscr{X}_{\overline{k(\mathfrak{P})}})$ respectively.

Inspired by the classical density result of Noether–Lefschetz loci for weight-2 polarized variations of Hodge structures, see [Voisin 2002; Oguiso 2003], Charles [2014] asked what can be said about the *arithmetic Noether–Lefschetz locus*

$$\mathrm{NL} = \{\mathfrak{P} \in \mathscr{S} \mid \rho(X_{\overline{K}}) < \rho(\mathscr{X}_{\overline{k(\mathfrak{P})}})\}.$$

In a prior work [Shankar et al. 2022, Theorem 1.1], we proved that the set NL is infinite under the additional assumption that X has potentially everywhere good reduction, i.e., up to taking a finite extension

MSC2020: 11G18, 14G40, 14J28.
Keywords: K3 surfaces, GSpin Shimura varieties, Arakelov theory.

of K, we assumed that $\mathscr{S} = \operatorname{Spec}(\mathcal{O}_K)$. The first main result of this paper is the following unconditional result.

Theorem 1.1. *Let X be a K3 surface over a number field K. Then the set* NL *is infinite.*

This theorem is a particular instance of Theorem 4.1 which is formulated for GSpin Shimura varieties and which has many other applications. As a consequence, Theorems 1.4, 1.6, and Corollary 1.7 in [Shankar et al. 2022] hold with no assumptions on the reduction type. In particular, we have the following theorem.

Theorem 1.2. *Let K be a number field and A an abelian surface over K. Then there exist infinitely many places where A has good reduction, and the reduction is geometrically nonsimple.*

1A. ***Picard rank jumps over function fields.*** Let $p \geq 5$ be a prime number. Let $\mathscr{X} \to \mathscr{S}$ be a family of K3 surfaces over a curve $\mathscr{S}$ over $\overline{\mathbb{F}}_p$. Let η be the generic point of $\mathscr{S}$. For every $s \in \mathscr{S}(\overline{\mathbb{F}}_p)$, we have similarly an inequality of Picard ranks

$$\rho(\mathscr{X}_{\bar{\eta}}) \leq \rho(\mathscr{X}_s),$$

and one can introduce similarly the Noether–Lefschetz locus as the subset of $\mathscr{S}$ where the above inequality is strict:

$$\mathrm{NL} = \{s \in \mathscr{S}(\overline{\mathbb{F}}_p) \mid \rho(\mathscr{X}_{\bar{\eta}}) < \rho(\mathscr{X}_s)\}.$$

Maulik, Shankar and Tang [Maulik et al. 2022a, Theorem 1.1] proved that if $\mathscr{S}$ is proper and the family $\mathscr{X} \to \mathscr{S}$ is generically ordinary and not isotrivial then the set NL is infinite. Our second main theorem in this paper is to remove the properness assumption in their result.

Theorem 1.3. *Let $\mathscr{X} \to \mathscr{S}$ be a generically ordinary nonisotrivial family of K3 surfaces over a smooth curve $\mathscr{S}$ over $\overline{\mathbb{F}}_p$ with $p \geq 5$. Suppose that the discriminant of the generic geometric Picard lattice is prime to p. Then the locus* NL *is infinite.*

The theorem is also a particular instance of Theorem 4.8 for GSpin Shimura varieties, which has several other applications and also has an analogue for unitary Shimura varieties, see Theorem 6.1. In particular, we have the following theorem which extends [Maulik et al. 2022b, Theorem 1(1)] to the quasiprojective case.

Theorem 1.4. *Let A be a nonisotrivial ordinary abelian surface over the function field of a curve over $\overline{\mathbb{F}}_p$. Then A has infinitely many smooth and nonsimple specializations.*

Both Theorem 1.1 and Theorem 1.3 are motivated by the density of Hodge loci in polarized variations of Hodge structure of weight 2 of K3 type; see for example [Voisin 2002; Oguiso 2003; Tayou 2020]. Recent density results for general polarized variations of Hodge structures of level less than 2 as in [Tayou and Tholozan 2023; Baldi et al. 2024] suggest that density of Hodge loci in arithmetic and function field settings are natural problems to investigate, and we hope to address these questions in future work.

1B. ***Hecke orbit conjecture.*** As an application of Theorem 1.3, we give a new proof of the Hecke orbit conjecture for orthogonal and certain unitary Shimura varieties. We refer to [Maulik et al. 2022a, Section 1.2] for the context and prior results on this conjecture.

Theorem 1.5. *Let $\mathcal{M}_{\mathbb{F}_p}$ be the reduction at $p \geq 5$ of the integral model of a Shimura variety of either*:

(1) *Orthogonal type associated to a lattice of signature $(b, 2)$ having discriminant prime to p.*

(2) *Unitary type associated to an imaginary quadratic field K split at p and to a Hermitian lattice over $\mathcal{O}_K$ of signature $(n, 1)$ with discriminant prime to p.*

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

The density of Hecke orbits in characteristic zero is a consequence of the work of Clozel and Ullmo [2005], see also [Eskin and Oh 2006] for a dynamical approach using Ratner theory. Chai [1995] first proved the Hecke orbit conjecture for the ordinary locus of the moduli space of principally polarized abelian varieties. For orthogonal and some unitary Shimura varieties, a first proof of the Hecke orbit conjecture in the ordinary case has been obtained by Maulik, Shankar and Tang [2022a] and our approach is inspired from theirs. Very recently, Pol Van Hoften [2024] proved this conjecture for the ordinary locus of Shimura varieties of abelian type under certain conditions on the reflex field and using completely different methods.

1C. ***Strategy of the proof.*** Theorem 1.1 and Theorem 1.3 are proved using a strategy initiated by Chai and Oort [2006] and Charles [2018] for the product of two modular curves and subsequently used in [Maulik et al. 2022b; Shankar and Tang 2020] for Hilbert modular surfaces over number fields and Siegel threefolds over $\overline{\mathbb{F}}_p$. Here we follow the set-up in [Shankar et al. 2022] and [Maulik et al. 2022a] to which we refer for more details. For Theorem 1.1, we first translate it into an intersection-theory-type statement between a curve and a sequence of divisors in the integral model of a toroidal compactification of a Shimura variety of GSpin type. For this matter, we use the Arakelov intersection theory with prelog forms developed in [Bruinier et al. 2007]. We follow a similar approach for Theorem 1.3, using the usual intersection theory on the reduction modulo p of the aforementioned compactification of the integral model of a GSpin Shimura variety. The new ingredients which were missing in both [Shankar et al. 2022] and [Maulik et al. 2022a] are the local estimates on multiplicities of intersection with special divisors at points of bad reduction and the estimates of extra terms coming from the boundary divisors in the global intersection numbers coming from the work of [Bruinier and Zemel 2022]; see also [Engel et al. 2023] for a recent approach. These are the main contributions of this paper. To obtain the first estimates, we use an explicit description of the special divisors in the formal completions along toroidal boundary components. This allows us to define in each case a decreasing sequence of positive definite lattices (L_n, Q) which computes the local intersection number. We give an estimate on the growth of the successive minima of these lattices, then a geometry-of-numbers-type argument allows us to derive the desired estimates. To obtain the bounds on the extra terms in the global intersection number, we use the explicit expressions from [Bruinier and Zemel 2022] and [Bruinier 2002] combined with an equidistribution result from [Duke 1988; Eskin and Oh 2006].

1D. ***Organization of the paper.*** The key input of this paper is the description of the special divisors in terms of local coordinates of integral models of toroidal compactifications of Shimura varieties of GSpin type. In Section 2, we explain these constructions following [Howard and Madapusi Pera 2020] and [Madapusi Pera 2016], and the section culminates with a description of the special divisors in formal completions along locally closed boundary divisors. In Section 3, we recall briefly Arakelov arithmetic intersection theory with prelog forms following [Bruinier et al. 2007], and we assemble different ingredients from the literature [Bruinier and Zemel 2022; Howard and Madapusi Pera 2020; Borcherds 1999] to state the modularity of the generating series of special divisors in the integral models of toroidal compactifications of Shimura varieties of GSpin type. In Section 4, we state the archimedean and finite place estimates needed to prove our main theorems, and then we prove the archimedean estimates. Section 5 is devoted to estimating contributions from bad reduction places. Finally, we prove the application to Hecke orbit conjecture in Section 6.

2. GSpin Shimura varieties: integral models and their compactifications

This section summarizes the construction of the GSpin Shimura variety, its toroidal compactifications and their integral models following [Bruinier and Zemel 2022; Howard and Madapusi Pera 2020; Madapusi Pera 2019; Andreatta et al. 2018], see also [Kisin 2010; Madapusi Pera 2016; Pink 1989] for earlier work. The ultimate goal is to describe the special divisors in formal completions along the toroidal boundary strata. The familiar reader may wish to skip directly to Section 2D for these results.

2A. ***The GSpin Shimura variety.*** Let (L, Q) be an even quadratic lattice of signature $(b, 2)$ with $b \geq 1$ and with associated even bilinear form

$$(\,.\,) : L \times L \to \mathbb{Z}$$

such that $Q(x) = (x.x)/2 \in \mathbb{Z}$ for all $x \in L$.

Let $G = \mathrm{GSpin}(L_{\mathbb{Q}})$ be the algebraic group over $\mathbb{Q}$ of spinor similitudes defined as in [Madapusi Pera 2016, Section 1.2]. The group $G(\mathbb{R})$ acts on the Hermitian symmetric space

$$\mathcal{D} =: \{z \in \mathbb{P}(L_{\mathbb{C}}) \mid (z.z) = 0, (z.\bar{z}) < 0\}.$$

The pair $(G, \mathcal{D})$ is the *GSpin Shimura datum*. Its reflex field is $\mathbb{Q}$ by [Madapusi Pera 2016, Section 3.1].

For $K \subset G(\mathbb{A}_f)$ a compact open subgroup, the *GSpin Shimura variety*

$$M(\mathbb{C}) = G(\mathbb{Q})\backslash \mathcal{D} \times G(\mathbb{A}_f)/K$$

is the set of complex points of a Deligne–Mumford stack M defined over $\mathbb{Q}$. In what follows, we choose the compact open group $K \subset G(\mathbb{A}_f)$ as in [Andreatta et al. 2018, Equation (4.1.2)]. Its image in

$\mathrm{SO}(L_{\mathbb{Q}})(\mathbb{A}_f)$ stabilizes $L \otimes \widehat{\mathbb{Z}} \subseteq L \otimes \mathbb{A}_f$ and is equal to the subgroup that acts trivially on the quotient $\widehat{L}^{\vee}/\widehat{L} = L^{\vee}/L$, where the dual lattice $L^{\vee}$ is defined as

$$L^{\vee} = \{x \in L_{\mathbb{Q}} \mid \forall y \in L, (x.y) \in \mathbb{Z}\}.$$

The Shimura variety M carries a line bundle of weight-1 modular forms that we denote by $\mathcal{L}_{\mathbb{Q}}$ and we refer to [Andreatta et al. 2018, Section 4.1] for a definition. The Shimura datum $(G, \mathcal{D})$ is of Hodge type by [Andreatta et al. 2017, Section 2.2]: there exists a Shimura datum of Siegel type $(G^{\mathrm{Sg}}, \mathcal{D}^{\mathrm{Sg}})$ and a compact open subgroup $K^{\mathrm{sg}} \subset G^{\mathrm{sg}}(\mathbb{A}_f)$ such that we have an embedding of Shimura varieties over $\mathbb{Q}$

$$M \hookrightarrow M^{\mathrm{Sg}}.$$

This is the *Kuga–Satake embedding*. The pull-back of the universal abelian scheme on M^{Sg} yields the *Kuga–Satake* abelian scheme $A \to M$.

2B. ***Toroidal compactifications over*** $\mathbb{C}$. In this section, we describe the toroidal compactifications of M as well as the structure of the boundary components following [Bruinier and Zemel 2022] and [Howard and Madapusi Pera 2020]. See also [Ash et al. 1975] for the general theory of toroidal compactifications over $\mathbb{C}$.

Recall from [Howard and Madapusi Pera 2020, Section 2.2] that an admissible parabolic subgroup $P \subseteq G$ is either a maximal proper parabolic subgroup of G or G itself.[1] A *cusp label representative* $\Phi = (P, \mathcal{D}^{\circ}, h)$ is a triple constituted from an admissible parabolic subgroup P, a connected component $\mathcal{D}^{\circ} \subset \mathcal{D}$ and an element $h \in G(\mathbb{A}_f)$.

Attached to a cusp label representative $\Phi = (P, \mathcal{D}^{\circ}, h)$, there exists a mixed Shimura variety that we now describe. Let U_{Φ} be the unipotent radical of P and let W_{Φ} be the center of U_{Φ}.[2] Let Q_{Φ} be the normal subgroup of P defined as in [Pink 1989, Sectoin 4.7], see also [Howard and Madapusi Pera 2020, 2.2]. Define as in [loc. cit.] $\mathcal{D}_{\Phi} = Q_{\Phi}(\mathbb{R})W_{\Phi}(\mathbb{C})\mathcal{D}^{\circ}$ and let $K_{\Phi} = hKh^{-1} \cap Q_{\Phi}(\mathbb{A}_f)$. We define then the mixed Shimura variety

$$M_{\Phi}(\mathbb{C}) = Q_{\Phi}(\mathbb{Q})\backslash\mathcal{D}_{\Phi} \times Q_{\Phi}(\mathbb{A}_f)/K_{\Phi}. \tag{2B.1}$$

By [Pink 1989, Proposition 12.1], $M^{\Phi}(\mathbb{C})$ has a canonical model M^{Φ} also defined over $\mathbb{Q}$. Let $\overline{Q}_{\Phi} = Q_{\Phi}/W_{\Phi}$ and $\overline{\mathcal{D}}_{\Phi} = W_{\Phi}(\mathbb{C})\backslash\mathcal{D}_{\Phi}$. Let $\overline{K}_{\Phi}$ be the image of K_{Φ} under the quotient map $Q_{\Phi}(\mathbb{A}_f) \to \overline{Q}_{\Phi}(\mathbb{A}_f)$. Then from the data $(\overline{Q}_{\Phi}, \overline{\mathcal{D}}_{\Phi}, \overline{K}_{\Phi})$ we define similarly to (2B.1) a mixed Shimura variety $\overline{M}_{\phi}$ and we have a canonical morphism

$$M_{\Phi} \to \overline{M}_{\Phi}. \tag{2B.2}$$

This map has a torsor structure that we now describe. Let $\Gamma_{\Phi} = K_{\Phi} \cap W_{\Phi}(\mathbb{Q})$. It is a $\mathbb{Z}$-lattice in $W_{\Phi}(\mathbb{Q})$. By [Howard and Madapusi Pera 2020, Proposition 2.3.1], the map (2B.2) is canonically a torsor under the torus $T_{\Phi,\mathbb{Q}}$ whose cocharacter group is Γ_{Φ}.

[1] G^{ad} is simple in our case.

[2] We follow the notation of [Madapusi Pera 2019] which differs from other references.

The mixed Shimura variety $\overline{M}_\Phi$ has itself a fibration structure over a pure Shimura variety constructed as follows; see [Madapusi Pera 2019, 2.1.7] for more details.

Let $G^h_\Phi = Q_\Phi/U_\Phi$ be the Levi quotient of Q_Φ, $V_\Phi = U_\Phi/W_\Phi$ the unipotent radical of $\overline{Q}_\Phi$ and let $\mathcal{D}^h_\Phi = V_\Phi(\mathbb{R})\backslash\overline{D}_\Phi$. Then the pair $(G^h_\Phi, \mathcal{D}^h_\Phi)$ is a pure Shimura datum with reflex field equal to $\mathbb{Q}$. Let $K^h_\Phi \subset G^h_\Phi(\mathbb{A}_f)$ be the image of K_Φ. Then the quotient

$$M^h_\Phi(\mathbb{C}) = G^h_\Phi(\mathbb{Q})\backslash(\mathcal{D}^h_\Phi \times G^h_\Phi(\mathbb{A}_f))/K^h_\Phi$$

is the set of complex points of a Shimura variety which admits a canonical model M^h_Φ defined over $\mathbb{Q}$ and we have a canonical map

$$\overline{M}_\Phi \to M^h_\Phi. \tag{2B.3}$$

By [Madapusi Pera 2019, 2.1.12], there exists a natural abelian scheme $A_K(\Phi) \to M^h_\Phi$ such that the map (2B.3) is a torsor under $A_K(\Phi)$.

In what follows, we will describe the above data for the GSpin Shimura variety introduced in Section 2A following [Howard and Madapusi Pera 2020, Section 4] and [Bruinier and Zemel 2022, Section 3]. Let Φ be a cusp label representative. The admissible parabolic subgroup P is the stabilizer of a totally isotropic subspace I_Φ of $L_\mathbb{Q}$ of dimension at most 2. The dimension-0 case corresponds to $P = G$. If P is the stabilizer of a primitive isotropic line $I_\mathbb{Q} \subset L_\mathbb{Q}$, then the cusp label representative is said to be of type III. If P is the stabilizer of a primitive isotropic plane $J_\mathbb{Q} \subset L_\mathbb{Q}$, then Φ is said to be of type III. We will follow the notation of [Bruinier and Zemel 2022] and denote by Υ, resp. Ξ, a cusp label representative of type II, resp. of type III.

Given two cusp label representatives Φ_1 and Φ_2, there is a notion of a K-morphism $\Phi_1 \xrightarrow{(\gamma,q_2)_K} \Phi_2$ given by $\gamma \in G(\mathbb{Q})$ and $q_2 \in Q_{\Phi_2}(\mathbb{A}_f)$ which we don't define here and refer to [Madapusi Pera 2019, 2.1.14] for the definition, see also [Howard and Madapusi Pera 2020, Definition 2.4.1].

Let Φ be a cusp label representative. By the general theory of toroidal compactifications, see [Pink 1989, 4.15] or [Ash et al. 1975, Chapter II, Section 1.1] for the definitions, there exists a canonical open nondegenerate self-adjoint convex cone $C_\Phi \subset W_\Phi(\mathbb{R})$ homogeneous under $P(\mathbb{R})$ and which allows to realize $\mathcal{D}^\circ$ as a tube domain inside an affine space, see [Madapusi Pera 2019, 2.1.5]. We define the *extended cone* C^*_Φ as in [Madapusi Pera 2019, 2.1.22]: for any map $\Phi' \xrightarrow{(\gamma,q)_K} \Phi$, the conjugation by γ^{-1} induces an embedding

$$\mathrm{int}(\gamma^{-1}) : W_{\Phi'}(\mathbb{R}) \hookrightarrow W_\Phi(\mathbb{R})$$

and we define then

$$C^*_\Phi = \bigcup_{\Phi' \to \Phi} \mathrm{int}(C_{\Phi'}).$$

This cone lies between C_Φ and its topological closure in $W_\Phi(\mathbb{R})$ but in general, it is neither open nor closed. See also [Pink 1989, Definition-Proposition 4.22] for more details.

Recall from [Howard and Madapusi Pera 2020, Definition 2.4.3] that a *rational polyhedral cone decomposition* (rpcd for short) of C^*_Φ is a collection $\Sigma_\Phi = \{\sigma\}$ of rational polyhedral cones $\sigma \subset W_\Phi(\mathbb{R})$

satisfying natural compatibility conditions (we don't recall these conditions here and invite the reader to consult the reference above for more information). The rpcd Σ_Φ is said to be smooth if it is smooth in the sense of [Pink 1989, Section 5.2] with respect to the lattice Γ_Φ. It is complete if

$$C_\Phi^* = \bigcup_{\sigma \in \Sigma_\Phi} \sigma.$$

2B1. *Boundary components of type II.* Let Υ be a cusp label representative of type II. Then P is the stabilizer of a primitive isotropic plane $J_\mathbb{Q} \subseteq L_\mathbb{Q}$ and let $J = J_\mathbb{Q} \cap h.L$, where $h \in G(\mathbb{A}_f)$ acts on L via the map $G(\mathbb{A}_f) \to \mathrm{SO}(L_\mathbb{Q})(\mathbb{A}_f)$. Then by [Howard and Madapusi Pera 2020, page 31], the group W_Υ is identified with $\bigwedge^2 J_\mathbb{Q}$; hence it is one-dimensional. The lattice $\Gamma_\Upsilon \subset W_\Upsilon$ is also of rank 1. The open convex cone C_Υ is given by a half line $\mathbb{R}^+ \backslash \{0\}$ and the extended cone is $C_\Upsilon^* = \{0\} \cup C_\Upsilon$.

Let M_Υ and $\overline{M}_\Upsilon$ be the mixed Shimura varieties associated to Υ. Then $M_\Upsilon \to \overline{M}_\Upsilon$ is a torsor under the one-dimensional torus T_Υ with cocharacter group Γ_Υ. The group G_Υ^h is equal to SL_2 and $\mathcal{D}_\Upsilon^h$ is equal to the Poincaré upper half-plane. The Shimura variety M_Υ^h is a modular curve and the abelian scheme A_Υ is equal to the Kuga–Sato variety $D \otimes E$ where $E \to M_\Upsilon^h$ is the universal elliptic curve over M_Υ^h and D is the positive definite plane $J^\perp/J$; see [Bruinier and Zemel 2022, Corollary 3.17] and [Zemel 2020, Proposition 4.3] for details and proofs. Notice that our choice for the compact open subgroup K gives exactly the stable orthogonal group used in [Bruinier and Zemel 2022] and [Zemel 2020].

The only possible cone decomposition of C_Υ^* in this situation is $\Sigma_\Upsilon = \{\{0\}, C_\Upsilon \cup \{0\}\}$ and this determines a partial compactification $M_\Upsilon \hookrightarrow M_{\Upsilon,\Sigma}$ which is a fibration by $\mathbb{A}^1_\mathbb{C}$ over $\overline{M}_\Upsilon$. Finally, there is only one boundary divisor denoted by B_Υ associated to the ray C_Υ.

2B2. *Boundary components of type III.* Let Ξ be a cusp label representative of type III. Then P is the stabilizer of a primitive isotropic line $I_\mathbb{Q} \subset L_\mathbb{Q}$ and let $I = I_\mathbb{Q} \cap h.L$. Set $K_I = I^\perp/I$. Then by [Howard and Madapusi Pera 2020, Equation (4.4.2)], we have $U_\Xi = W_\Xi$ and we have an isomorphism of vector spaces

$$K_{I,\mathbb{Q}} \otimes I_\mathbb{Q} \simeq W_\Xi(\mathbb{Q}).$$

The lattice (K_I, Q) is a Lorentzian lattice of signature $(b-1, 1)$. Under the above isomorphism, and assuming we have chosen a primitive generator of I, the open convex cone $C_\Xi \subset W_\Xi(\mathbb{R})$, see [Howard and Madapusi Pera 2020, Section 2.4] is identified with a connected component of the light cone

$$\{x \in K_{I,\mathbb{R}},\ Q(x) < 0\}.$$

The spaces M_Ξ^h and $\overline{M}_\Xi$ are equal and are Shimura varieties of dimension zero that we can describe as follows. Let $(\mathbb{G}_m, \mathcal{H}_0)$ be the Shimura data given by

$$\mathcal{H}_0 := \{2\pi\epsilon : \epsilon^2 = -1\},$$

on which $\mathbb{R}^\times$ acts naturally through the quotient $\mathbb{R}^\times/\mathbb{R}_+^\times$. There is a morphism of mixed Shimura data $(Q_\Xi, \mathcal{D}_\Xi) \to (\mathbb{G}_m, \mathcal{H}_0)$ given by a canonical character $\nu_\Xi : Q_\Xi \to \mathbb{G}_m$ defined as in [Howard and Madapusi Pera 2020, Equation (4.4.1)] and a map $\mathcal{D}_\Xi \to \mathcal{H}_0$ given as in [loc. cit., Equation (4.6.3)]. Then the

Shimura variety $\mathrm{Sh}_{\nu_\Xi(K_\Xi)}(\mathbb{G}_m, \mathcal{H}_0)$ is zero-dimensional and the canonical map $M_\Xi \to \mathrm{Sh}_{\nu_\Xi(K_\Xi)}(\mathbb{G}_m, \mathcal{H}_0)$ is a torsor under the torus $T_\Xi = \mathrm{Spec}(\mathbb{Q}[q_\alpha]_{\alpha \in \Gamma_\Xi^\vee})$ with cocharacter group $\Gamma_\Xi = K_I$ by [Bruinier and Zemel 2022, Proposition 3.7].

The intermediate cone C_Φ^* can be described explicitly as follows, see also [Bruinier and Zemel 2022, page 23]: for any type-II boundary component Υ with corresponding isotropic plane J containing I, the quotient J/I has a generator $\omega_{\Xi,\Upsilon}$ lying on the boundary of the C_Ξ. Hence

$$C_\Xi^* = C_\Xi \cup \bigcup_\Upsilon \mathbb{R}\omega_{\Xi,\Upsilon}.$$

The rays $\mathbb{R}\omega_{\Xi,\Upsilon}$ will be referred to as the external rays and the rays in C_Ξ are the *inner rays*.

2B3. *Toroidal compactifications.* Recall from [Howard and Madapusi Pera 2020, Definition 2.4.4] that a K-admissible rational polyhedral cone decomposition for $(G, \mathcal{D})$ is a collection $\Sigma = \{\Sigma_\Xi, \Sigma_\Upsilon\}$ such that Σ_Ξ and Σ_Υ are rpcd for any cusp label representative Ξ and Υ respectively satisfying the compatibility conditions of [loc. cit., Definitions 2.4.3, 2.4.4]. It is said smooth (resp. complete) if every Σ_Φ is smooth (resp. complete).

A toroidal stratum representative is a pair (Φ, σ) where Φ is a cusp label representative and $\sigma \subset C_\Phi^*$ is a rational polyhedral cone whose interior is contained in C_Φ. There is similarly a notion of K-morphism between stratum representatives, see [loc. cit., Definition 2.4.6] and the set of K-isomorphism classes of toroidal stratum representatives will be denoted $\mathrm{Start}_K(G, \mathcal{D}, \Sigma)$. We say that Σ is finite if

$$|\mathrm{Start}_K(G, \mathcal{D}, \Sigma)| < \infty.$$

Let Σ be a finite K-admissible complete cone decomposition. The main result of [Pink 1989, Section 12], see also [Madapusi Pera 2019, Theorem 2.1.27], ensures that there exists a proper toroidal compactification

$$M \hookrightarrow M^\Sigma$$

in the category of Deligne–Mumford stacks over $\mathbb{Q}$ such that M^Σ is proper over $\mathbb{Q}$ and has a stratification

$$M^\Sigma = \bigsqcup_{(\Phi,\sigma) \in \mathrm{Start}_K(G,\mathcal{D},\Sigma)} B^{\Phi,\sigma} \tag{2B.4}$$

by locally closed subspaces indexed by the finite set of strata $\mathrm{Start}_K(G, \mathcal{D}, \Sigma)$. The stratum indexed by (Φ, σ) lies in the closure of the stratum index by (Φ', σ') if and only if there is a K-morphism of strata representatives $(\Phi, \sigma) \to (\Phi', \sigma')$. Then the closure of the stratum $B_K^{\Phi,\sigma}$ is given by

$$\overline{B^{\Phi,\sigma}} = \bigcup_{(\Phi',\sigma') \to (\Phi,\sigma)} B^{\Phi',\sigma'}.$$

Moreover, by [Howard and Madapusi Pera 2020, Theorem 3.4.1] following the work of Harris and Zucker [2001], the line bundle of weight-1 modular forms $\mathcal{L}$ extends to a line bundle on M^Σ which we still denote $\mathcal{L}$ by abuse of notation.

Let (Φ, σ) be a toroidal stratum representative. Then (Φ, σ) determines a partial compactification of the mixed Shimura variety $M_\Phi \hookrightarrow M_\Phi(\sigma)$ with boundary component index by σ denoted by $Z^\Phi(\sigma)$. Pink proved that there is a canonical isomorphism [Pink 1989, Corollary 7.17, Theorem 12.4], see also [Madapusi Pera 2019, Theorem 2.1.27], of Deligne–Mumford stacks

$$\Delta_K(\Phi, \sigma)\backslash Z^\sigma(\sigma) \simeq B^{\Phi,\sigma},$$

where $\Delta_K(\Phi, \sigma)$ is the finite group defined in [Madapusi Pera 2019, 2.1.19]. The latter induces an isomorphism of formal Deligne–Mumford stacks

$$\Delta_K(\Phi, \sigma)\backslash \widehat{M}_\Phi(\sigma) \simeq \widehat{M}^\Sigma, \tag{2B.5}$$

where $\widehat{M}_\Phi(\sigma)$ is the completion of $M_\Phi(\sigma)$ along the locally closed subspace $Z^\Phi(\sigma)$ and $\widehat{M}^\Sigma$ is the formal completion of M^Σ along the locally closed stratum $B^{\Phi,\sigma}$.

Our goal in the next two sections is to make the above isomorphisms explicit for type-II and type-III boundary strata.

2B4. *Formal completion along type-II boundary strata.* Let Υ be a cusp label representative of type II. By the discussion in Section 2B1, there is a unique choice of a one-dimensional ray σ and hence a unique choice of boundary stratum representative (Υ, σ) which corresponds to a locally closed divisor $B^{\Upsilon,\sigma}$.

The morphism $M_\Upsilon \to \overline{M}_\Upsilon$ is then a torsor under a one-dimensional torus T_Υ with cocharacter group $\Gamma_\Upsilon \simeq \mathbb{Z}$, i.e., $T_\Upsilon \simeq \mathrm{Spec}(\mathbb{Q}[q, q^{-1}])$. The partial compactification $T_\Upsilon(\sigma)$ is then isomorphic to $\mathrm{Spec}(\mathbb{Q}[q])$ and the partial toroidal compactification of M_Υ is given as a twisted torus embedding over $\overline{M}_\Upsilon$ with fiber $\mathrm{Spec}(\mathbb{Q}[q])$. Hence we have the following description of $\widehat{M}_\Upsilon(\sigma)$

$$\widehat{M}_\Upsilon(\Sigma) \xrightarrow{\mathrm{Spf}(\mathbb{Q}[\![X]\!])} \overline{M}_\Upsilon \xrightarrow{D\otimes E} M^h_\Upsilon.$$

2B5. *Formal completion along type-III boundary strata.* Let (Ξ, σ) be a toroidal stratum representative of type III such that σ is a one-dimensional inner ray. The corresponding boundary component is denoted by $B^{\Phi,\sigma}$ and is a locally closed divisor. Write $\sigma = \mathbb{R}\omega$, where $\omega \in C_\Xi \cap K$ is an integral primitive generator that satisfies $(\omega.\omega) < 0$.

The morphism $M_\Xi \to \mathrm{Sh}_{\nu_\Xi(K_\Xi)}(\mathbb{G}_m, \mathcal{H}_0)$ is a torsor under the torus

$$T_\Xi = \mathrm{Spec}(\mathbb{Q}[q_\alpha]_{\alpha\in\Gamma_\Xi^\vee}).$$

The partial compactification $T_\Xi(\sigma)$ is equal to

$$T_\Xi(\sigma) = \mathrm{Spec}(\mathbb{Q}[q_\alpha]_{(\alpha.\omega)\geq 0, \alpha\in\Gamma_\Xi^\vee})$$

and the ideal defining the boundary divisor is given by $I_\sigma = (q_\alpha, (\alpha, \omega) > 0)$. It is generated by $q_{\omega'}$ for any $\omega' \in \Gamma_\Xi^\vee$ for which $(\omega, \omega') = 1$. We fix such ω'.

The formal completion along the boundary divisor is then given by

$$\widehat{T}_\Xi(\sigma) = \mathrm{Spec}(\mathbb{Q}[q_\alpha, \alpha \in \Gamma_\Xi^\vee \cap \omega^\perp][\![q_{\omega'}]\!]),$$

and the map $M_\Xi(\sigma) \to \mathrm{Sh}_{\nu_\Xi(K_\Xi)}(\mathbb{G}_m, \mathcal{H}_0)$ is a twisted torus embedding with fibers $\widehat{T}_\Xi(\sigma)$. We will trivialize this fibration following an approach similar to [Howard and Madapusi Pera 2020, page 34].

First choose an auxiliary isotropic line $I_* \subset L_\mathbb{Q}$ such that $(I.I_*) \neq 0$. Then by [Howard and Madapusi Pera 2020, Equation (4.6.6)] and the discussion that follows, this determines a section

$$(\mathbb{G}_m, \mathcal{H}_0) \xrightarrow{s} (Q_\Xi, \mathcal{D}_\Xi).$$

The section s determines a Levi decomposition $Q_\Xi = \mathbb{G}_m \ltimes U_\Xi$. Let $K_0 \subset \mathbb{G}_m(\mathbb{A}_f)$ be a compact open subgroup small enough such that the image under the section s is contained in K_Ξ and let

$$K_{\Xi,0} = K_0 \ltimes (U_\Xi(\mathbb{A}_f \cap K_\Xi)) \subset K_\Xi.$$

Then by reasoning similarly to [Howard and Madapusi Pera 2020, Proposition 4.6.2], we have the following.

Proposition 2.1. *We have an isomorphism of formal algebraic spaces*

$$\bigsqcup_{a \in \mathbb{Q}^\times_{>0} \backslash \mathbb{A}^\times_f / K_0} \widehat{T}_\Xi(\sigma)_{/\mathbb{C}} \xrightarrow{\simeq} \widehat{M}_{K_{\Xi,0}}(\sigma)_{/\mathbb{C}},$$

and the map

$$\widehat{M}_{K_{\Xi,0}}(\sigma)_{/\mathbb{C}} \to \widehat{M}_{K_\Xi}(\sigma)_{/\mathbb{C}}$$

is a formally étale map of formal Deligne–Mumford stacks given by the quotient by $K_\Xi/K_{\Xi,0}$. In particular, if K is neat, then the above map is a formally étale surjection of algebraic spaces.

Proof. The same proof as in [Howard and Madapusi Pera 2020, Proposition 4.6.2] works with no change in our setting. □

2C. ***Integral models.*** We recall in this section the construction of integral models of GSpin Shimura varieties and their compactifications following [Howard and Madapusi Pera 2020; Andreatta et al. 2018; Madapusi Pera 2019]. We assume henceforth that the lattice (L, Q) is a maximal lattice, i.e., there is no strict superlattice in $L_\mathbb{Q}$ containing L over which Q is $\mathbb{Z}$-valued.

By [Andreatta et al. 2018, Section 4.4], there exists a flat and normal integral model $\mathcal{M} \to \mathrm{Spec}(\mathbb{Z})$ which is a Deligne–Mumford stack of finite type over $\mathbb{Z}$. It enjoys the following properties:

(1) If the lattice (L, Q) is almost self dual at a prime p then the restriction of the integral model to $\mathrm{Spec}(\mathbb{Z}_{(p)})$ is smooth.[3]

(2) If p is odd and p^2 does not divide the discriminant of (L, Q), the restriction of $\mathcal{M}$ to $\mathrm{Spec}(\mathbb{Z}_{(p)})$ is regular.

(3) If $n \geq 6$, the reduction mod p is geometrically normal.

(4) The line bundle of modular forms of weight 1 extends to a line bundle on $\mathcal{M}$ that we denote by $\mathcal{L}$.

[3] See [Howard and Madapusi Pera 2020, Definition 6.1.1].

Furthermore, given a K-admissible polyhedral complete cone decomposition, $\mathcal{M}$ admits by [Madapusi Pera 2019, Theorem 4.1.5] a toroidal compactification $\mathcal{M}^{\Sigma}$ proper over $\operatorname{Spec}(\mathbb{Z})$ and which extends the compactification M^{Σ} previously defined over $\mathbb{Q}$. Moreover, it has a stratification

$$\mathcal{M}^{\Sigma} = \bigsqcup_{(\Phi,\sigma)\in \operatorname{Start}_K(G,\mathcal{D},\Sigma)} \mathcal{B}^{\Phi,\sigma} \tag{2C.1}$$

which extends the stratification in (2B.4) and such every stratum is flat over $\mathbb{Z}$. The unique open stratum is $\mathcal{M}$ and its complement is a Cartier divisor. Moreover, for any cusp label representative (Φ, σ), the tower of maps

$$M_{\Phi}(\sigma) \to \overline{M}_{\phi} \to M^{h}_{\Phi}$$

has an integral model

$$\mathcal{M}_{\Phi}(\sigma) \to \overline{\mathcal{M}}_{\phi} \to \mathcal{M}^{h}_{\Phi}$$

which satisfies the following: the abelian scheme A_{Φ} has an extension $\mathcal{A}_{\Phi} \to \mathcal{M}^{h}_{\Phi}$ such that the map $\overline{\mathcal{M}}_{\phi} \to \mathcal{M}^{h}_{\Phi}$ is a torsor under $\mathcal{A}_{\Phi}$ and the map $\mathcal{M}_{\Phi}(\sigma) \to \overline{\mathcal{M}}_{\phi}$ is a twisted torus embedding with structure group the torsor $\mathcal{T}_{\Xi}$ extending T_{Ξ}. Finally, the boundary component $Z_{\Phi}(\sigma)$ has a flat extension $\mathcal{Z}_{\Phi}(\sigma)$ such that we have an isomorphism of completions:

$$\Delta_K(\Phi,\sigma)\backslash \widehat{\mathcal{M}}_{\Phi}(\sigma) \simeq \widehat{\mathcal{M}}^{\Sigma} \tag{2C.2}$$

extending the isomorphism in (2B.5). See [Madapusi Pera 2019, Theorem 4.1.5] and [Howard and Madapusi Pera 2020, Section 8.1] for more details.

Fix a prime p. The goal of the next two subsections is to describe the formal completions of $\mathcal{M}^{\Sigma}$ along the boundary divisors of these compactifications explicitly over $\mathbb{Z}_{(p)}$ in the type-II and the type-III case.

2C1. *Type II.* Let (Υ, σ) be a toroidal stratum representative of type II where σ is the unique one-dimensional ray.

Let $\mathcal{T}_{\Upsilon} = \operatorname{Spec}(\mathbb{Z}_{(p)}[q, q^{-1}])$ with partial compactification $\mathcal{T}_{\Upsilon}(\sigma) = \operatorname{Spec}(\mathbb{Z}_{(p)}[q])$. By (2C.1) and [Madapusi Pera 2019, Theorem 4.1.5(2–4)], the morphism $\mathcal{M}_{\Upsilon} \to \overline{\mathcal{M}}_{\Upsilon}$ is a torsor under $\mathcal{T}_{\Upsilon}$ and the morphism $\overline{\mathcal{M}}_{\Upsilon} \to \mathcal{M}^{h}_{\Xi}$ is a torsor under $D \otimes \mathcal{E}$, where $\mathcal{E} \to \mathcal{M}^{h}_{\Xi}$ is the universal elliptic curve. Moreover, the partial toroidal compactification of $\mathcal{M}_{\Xi}$ is given as a twisted torus embedding over $\overline{\mathcal{M}}_{\Upsilon}$ with fibers isomorphic to $\mathcal{T}_{\Upsilon}(\sigma)$. In particular, the formal completion of $\mathcal{M}_{\Upsilon}$ along the boundary component is describe by the following diagram:

$$\widehat{\mathcal{M}}_{\Upsilon}(\sigma) \xrightarrow{\widehat{\mathcal{T}}_{\Upsilon}(\sigma)} \overline{\mathcal{M}}_{\Upsilon} \xrightarrow{D\otimes\mathcal{E}} \mathcal{M}^{h}_{\Upsilon}, \tag{2C.3}$$

where $\widehat{\mathcal{T}}_{\Upsilon}(\sigma) = \operatorname{Spf}(\mathbb{Z}_{(p)}[\![q]\!])$.

2C2. *Type III.* Let (Ξ, σ) be a toroidal stratum representative of type III such that σ is one-dimensional and generated by a primitive integral element $\omega \in C_{\Xi}$ with $(\omega.\omega) = -2N$. Let $\mathcal{T}_{\Xi} = \operatorname{Spec}(\mathbb{Z}_{(p)}[q_{\alpha}]_{\alpha\in\Gamma^{\vee}_{\Xi}})$ and recall that we have a T_{Ξ} torsor structure

$$M_{\Xi} \to \operatorname{Sh}_{\nu_{\Xi}(K_{\Xi})}(\mathbb{G}_m, \mathcal{H}_0).$$

The cone σ determines a partial compactification $\mathcal{T}_\Xi(\sigma)=\mathrm{Spec}(\mathbb{Z}_{(p)}[q_\alpha]_{(\alpha,\omega)\geq 0,\alpha\in\Gamma_\Xi^\vee})$ and also a partial compactification $\mathcal{M}_\Xi\hookrightarrow\mathcal{M}_\Xi(\sigma)$ which is a twisted torus embedding with fibers $\mathcal{T}_\Xi(\sigma)$.

The boundary divisor in $\mathcal{T}_\Xi(\sigma)$ is defined by the ideal $I_\sigma=(q_\alpha,(\alpha,\omega)>0)$. If $\omega'\in\Gamma_\Xi^\vee$ is as before an element such that $(\omega'.\omega)=1$, then $I_\sigma=(q_{\omega'})$. The formal completion of $\mathcal{T}_\Xi(\sigma)$ along I_σ is then given by

$$\widehat{\mathcal{T}}_\Xi=\mathrm{Spf}(\mathbb{Z}_{(p)}[q_\alpha,\alpha\in\Gamma_\Xi^\vee\cap\omega^\perp][\![q_{\omega'}]\!]).$$

Recall that we have a morphism of Shimura data

$$(Q_\Xi,\mathcal{D}_\Xi)\xrightarrow{\nu_\Xi}(\mathbb{G}_m,\mathcal{H}_0),$$

and let s be the section of ν_Ξ defined in Section 2B5. Let $K_0\subset\mathbb{A}_f^\times$ be a compact open subgroup such that $s(K_0)\subset K_\Xi$. We can furthermore assume that K_0 factors as

$$K_0=\mathbb{Z}_p^\times.K_0^p.$$

Let F be the abelian extension of $\mathbb{Q}$ determined by the reciprocity morphism in global class field theory:

$$\mathrm{rec}:\mathbb{Q}_{>0}^\times\backslash\mathbb{A}_f^\times/K_0\simeq\mathrm{Gal}(F/\mathbb{Q}).$$

Fix a prime $\mathfrak{P}\subset\mathcal{O}_F$ above p and let R be the localization of $\mathcal{O}_F$ at $\mathfrak{P}$. Then using similar arguments as in [Howard and Madapusi Pera 2020, Proposition 8.2.3], we have the following proposition.

Proposition 2.2. *There is an isomorphism*

$$\coprod_{\mathbb{Q}_{>0}^\times\backslash\mathbb{A}_f^\times/K_0}\widehat{\mathcal{T}}_\Xi(\sigma)_{/R}\to\widehat{\mathcal{M}}_{\Xi,0}(\sigma)/R$$

of formal Deligne–Mumford stacks over R whose base change to $\mathbb{C}$ agrees with Proposition 2.1. Moreover, the map

$$\widehat{\mathcal{M}}_{\Xi,0}(\sigma)/R\to\widehat{\mathcal{M}}_\Xi(\sigma)/R$$

is an étale map of Deligne–Mumford stacks given as the quotient by $K_\Xi/K_{\Xi,0}$.

The proof follows from the description given over $\mathbb{C}$ Proposition 2.1, the flatness of both sides over $\mathbb{Z}_{(p)}$ and the fact the normalization of $\mathrm{Spec}(\mathbb{Z}_{(p)})$ in $\mathrm{Sh}_{K_0}(\mathbb{G}_m,\mathcal{H}_0)$ is isomorphic to $\coprod_{a\in\mathbb{Q}_{>0}^\times\backslash\mathbb{A}_f^\times/K_0}\mathrm{Spec}(R)$, see [Howard and Madapusi Pera 2020, Proposition 8.2.3] for a proof and more details.

2D. ***Special divisors.*** We continue to assume in this section that the lattice (L,Q) is maximal and let Σ be a smooth K-admissible cone decomposition.

For every $\beta\in L^\vee/L$, $m\in Q(\beta)+\mathbb{Z}$ such that $m>0$, one can define a *special divisor $\mathcal{Z}(\beta,m)\to\mathcal{M}$* following [Andreatta et al. 2018, Definition 4.5.6]. We recall briefly the definition and refer to [loc. cit.] for more details.

The Shimura variety $\mathcal{M}$ carries the family of Kuga–Satake abelian varieties $\mathcal{A}\to\mathcal{M}$. For any scheme $S\to\mathcal{M}$, a group of special quasiendomorphisms $V_\beta(\mathcal{A}_S)$ is defined in [Andreatta et al. 2018, Section 4.5].

Then the functor sending a scheme S to

$$\mathcal{Z}(\beta, m)(S) = \{x \in V_\beta(\mathcal{A}_S) \mid Q(x) = m\}$$

is representable by a Deligne–Mumford stack which is étale locally an effective Cartier divisor on $\mathcal{M}$. We will rather consider its image in $\mathcal{M}$ by a procedure described in [Howard and Madapusi Pera 2020] after Proposition 6.5.2. By abuse of notation, we also denote by $\mathcal{Z}(\beta, m)$ its closure in $\mathcal{M}^\Sigma$, which is again a Cartier divisor.

In what follows, we will give an explicit description of $\mathcal{Z}(\beta, m)$ in the formal completions of $\mathcal{M}^\Sigma$ along its boundary components. Since for our purposes we only need $\beta = 0$ and m coprime to p, we will only describe what happens in this situation and we abbreviate for short $\mathcal{Z}(\beta, m) = \mathcal{Z}(m)$. We assume that $m \geq 1$ is coprime to p for the rest of this section.

By [Andreatta et al. 2018, page 434], $\mathcal{Z}(m)(\mathbb{C})$ has a complex uniformization as follows: for any $g \in G(\mathbb{A}_f)$, let $L_g = g.\widehat{L} \cap L_{\mathbb{Q}}$ and consider the sub-Hermitian domain of $\mathcal{D}$

$$\mathcal{D}^\circ(\lambda) = \{x \in \mathcal{D}^\circ \mid (x, \lambda) = 0\},$$

where $\lambda \in L_g$, $Q(\lambda) = m$. Then $\mathcal{Z}(m)(\mathbb{C})$ is equal to the union of $\mathcal{D}^\circ(\lambda)$ for $g \in G(\mathbb{A}_f)$ and $\lambda \in L_g$ with $Q(\lambda) = m$.

For any $\lambda \in L_g$ with $Q(\lambda) = m$, let G_λ be the fixator of λ, L_λ the orthogonal lattice to λ in $L_{\mathbb{Q}}$, and let $\mathcal{D}_\lambda \subset \mathcal{D}$ be the orthogonal to λ. Notice that $\mathcal{D}_\lambda$ does not depend on g but only on $\lambda \in L_{\mathbb{C}}$. Notice that since m is coprime to p, the lattice L_λ is also maximal at p. Then (G_λ, D_λ) is again a Shimura datum of GSpin associated to the lattice (L_λ, Q) which is of signature $(b-1, 2)$ and has reflex field equal to $\mathbb{Q}$. If we choose $K_\lambda \subset G_\lambda(\mathbb{A}_f)$ a compact open subgroup as in [Andreatta et al. 2018, Equation (4.1.2)], then $K_\lambda \subset K \cap G_\lambda(\mathbb{A}_f)$ and we obtain a morphism of complex Shimura varieties

$$M_\lambda(\mathbb{C}) \to M(\mathbb{C}).$$

By the description [loc. cit., Equation (2.4)], the union over $g \in G(\mathbb{A}_f)$, $\lambda \in L_g$ with $Q(\lambda) = m$ of the images of $M_\lambda(\mathbb{C})$ is equal to $\mathcal{Z}(m)(\mathbb{C})$.

Now since (G_λ, D_λ) is again a Shimura variety of GSpin type associated to a lattice maximal at p, the discussion in the previous sections applies verbatim to the Shimura variety M_λ and yields similar description for the compactification and the integral model over $\mathbb{Z}_{(p)}$. In particular, we have a map between integral models $\mathcal{M}_\lambda \to \mathcal{M}$ over $\mathbb{Z}_{(p)}$ which factors through $\mathcal{Z}(m)$ by [Howard and Madapusi Pera 2020, page 82].

$$\mathcal{M}_\lambda \to \mathcal{Z}(m) \hookrightarrow \mathcal{M}$$

and the union over of images of such maps for $g \in G(\mathbb{A}_f)$ and $\lambda \in L_g$ with $Q(\lambda) = m$ is equal to $\mathcal{Z}(m)$.[4]

Let (Φ, σ) be a toroidal stratum representative for $\mathcal{M}$. From the description of the parabolic subgroups of GSpin$(b, 2)$, we have the following lemma.

[4] This union is in fact finite.

Lemma 2.3. *The group $P \cap G_\lambda$ is an admissible parabolic subgroup of G_λ if and only if $\lambda \in I_\Phi^\perp$.*

Notice also that if $\lambda \notin I_\Phi^\perp$, then the image of $\mathcal{D}_\lambda$ in $M^\Sigma(\mathbb{C})$ will not intersect the boundary components parametrized by Φ, as its projection to the Baily–Borel compactification will not do so. Hence they will not appear in the formal completions of $\mathcal{M}^\Sigma$ along these boundary components.

We can write $\Phi = (P, \mathcal{D}^\circ, h)$ and let $\lambda \in L_g$ with $Q(\lambda) = m$ such that $\lambda \in I_\Phi^\perp$. Lemma 2.3 shows that (Φ, σ) can also be seen as a toroidal stratum representative with respect to $(G_\lambda, \mathcal{D}_\lambda)$ by considering $P \cap G_\lambda$; see [Madapusi Pera 2016, Section 2.1.28] for more details. Let $\mathcal{M}_{\lambda,\Upsilon}$ be the integral model over $\mathbb{Z}_{(p)}$ of the mixed Shimura variety associated to Φ. We get then a morphism of mixed Shimura varieties

$$\mathcal{M}_{\lambda,\Phi} \to \mathcal{M}_\Phi,$$

as well as a morphism of partial compactifications respecting the strata

$$\mathcal{M}_{\lambda,\Phi}(\sigma) \to \mathcal{M}_\Phi(\sigma).$$

By [Madapusi Pera 2019, Proposition 2.1.29], the morphism induced at the level of formal completions along the boundary strata given by σ is compatible with the toroidal compactifications of $\mathcal{M}_\lambda$ and $\mathcal{M}$. In particular, we get a commutative diagram

$$\begin{tikzcd}
\widehat{\mathcal{M}}_{\lambda,\Phi}(\sigma) \arrow[r] \arrow[d] & \widehat{\mathcal{M}}_\Phi(\sigma) \arrow[d] \\
\widehat{\mathcal{Z}}(m) \arrow[r] & \widehat{\mathcal{M}}^\Sigma
\end{tikzcd}$$

where the right vertical map is an étale cover of Deligne–Mumford stacks, the left vertical map is an étale cover of an open and closed subset by [Howard and Madapusi Pera 2020, page 82]. Finally, the union over $g \in G(\mathbb{A}_f)$, $\lambda \in L_g$ with $Q(\lambda) = m$ of the images of the left map covers the whole $\widehat{\mathcal{Z}}(m)$.

2D1. *Special divisors along type-II boundary components.* Let (Υ, σ) be a toroidal stratum representative of type II.

Let $\lambda \in L$ with $Q(\lambda) = m$ such that $\lambda \in I_\Upsilon^\perp$ and m is coprime to p. We have a morphism of formal completions of the partial compactifications of mixed Shimura varieties

$$\widehat{\mathcal{M}}_{\lambda,\Upsilon}(\sigma) \to \widehat{\mathcal{M}}_\Upsilon(\sigma).$$

Let $x \in \mathcal{B}^{\Upsilon,\sigma}(\overline{\mathbb{F}}_p) \subset \mathcal{M}_\Upsilon(\sigma)(\overline{\mathbb{F}}_p)$ and let $\mathcal{O}_{\mathcal{M}_\Upsilon(\sigma),x}$ be the local ring at x. Let $\bar{x}$ be the image of x in $\overline{\mathcal{M}}_\Upsilon(\overline{\mathbb{F}}_p)$ and let z the image in $\mathcal{M}_\Upsilon^h(\overline{\mathbb{F}}_p)$. If follows from (2C.3) that the formal completion $\widehat{\mathcal{O}}_{\mathcal{M}_\Upsilon(\sigma),x}$ is isomorphic to

$$\widehat{\mathcal{O}}_{\mathcal{M}_\Upsilon(\sigma),x} \simeq \mathbb{Z}_p[\![X]\!] \mathbin{\widehat{\otimes}} \widehat{\mathcal{O}}_{\overline{\mathcal{M}}_\Phi,\bar{x}}.$$

Moreover, the pull-back of the torsor $\overline{\mathcal{M}}_\Upsilon \to \mathcal{M}_\Upsilon^h$ to $\operatorname{Spf}(\widehat{\mathcal{O}}_{\mathcal{M}_\Upsilon^h,z})$ is trivial, as it is trivial by reduction to $\overline{\mathbb{F}}_p$ and we can lift formally any section. Hence

$$\widehat{\mathcal{O}}_{\overline{\mathcal{M}}_\Phi,\bar{x}} \simeq \widehat{\mathcal{O}}_{D\otimes\mathcal{E},\bar{x}}.$$

For $\lambda \in D$, consider the map over $\mathcal{M}^h_\Upsilon$

$$D \otimes \mathcal{E} \xrightarrow{(\lambda.)\otimes \mathrm{Id}} \mathcal{E}. \tag{2D.1}$$

Its kernel is flat over $\mathcal{M}^h_\Upsilon$. Let $I_\lambda \subset \lambda$ be the ideal defining it. Then $\hat{I}_\lambda \hookrightarrow \widehat{\mathcal{O}}_{D\otimes\mathcal{E},\bar{x}}$ is flat over $\widehat{\mathcal{O}}_{\mathcal{M}^h_{\Upsilon},z}$.

Proposition 2.4. *The formal completion $\widehat{\mathcal{Z}}(m)$ along x is the union over $\lambda \in D$ with $Q(\lambda) = m$ of the vanishing loci inside $\widehat{\mathcal{O}}_{\mathcal{M}_\Upsilon(\sigma),x}$ of the ideals $\mathbb{Z}_p[\![X]\!] \mathbin{\widehat{\otimes}} \hat{I}_\lambda$.*

Proof. Let $\lambda \in L$ such that $\lambda \in J^\perp$ and $Q(\lambda) = m$. Then we have a description of the mixed Shimura variety $M_{\Upsilon,\lambda}$ similar to (2C.3), namely, it has a fibration structure which fits into the following diagram:

$$\begin{tikzcd}
\widehat{\mathcal{M}}_{\lambda,\Upsilon}(\sigma) \arrow[r, "\widehat{\mathcal{T}}_\Upsilon(\sigma)"] \arrow[d] & \overline{\mathcal{M}}_{\lambda,\Upsilon} \arrow[r, "D_\lambda\otimes\mathcal{E}"] \arrow[d] & \mathcal{M}^h_{\lambda,\Upsilon} \arrow[d] \\
\widehat{\mathcal{M}}_\Upsilon(\sigma) \arrow[r, "\widehat{\mathcal{T}}_\Upsilon(\sigma)"] & \overline{\mathcal{M}}_\Upsilon \arrow[r, "D\otimes\mathcal{E}"] & \mathcal{M}^h_\Upsilon
\end{tikzcd}$$

One can check that $D_\lambda = \bar{\lambda}^\perp$, where $\bar{\lambda}$ is the image of λ in $D = J^\perp/J$. Moreover, the right vertical map in the above diagram is an étale cover and the vertical middle map is equivariant with respect to the inclusion

$$\bar{\lambda}^\perp \otimes \mathcal{E} \hookrightarrow D \otimes \mathcal{E},$$

and the left vertical map has image given by an open and closed subset of $\widehat{\mathcal{Z}}(m)$.

Let $z' \in \mathcal{M}^h_{\lambda,\Upsilon}(\overline{\mathbb{F}}_p)$ be a point mapping to z, then $\widehat{\mathcal{O}}_{\mathcal{M}^h_{\lambda,\Upsilon,z'}} \simeq \widehat{\mathcal{O}}_{\mathcal{M}^h_{\Upsilon,z}}$. Hence the above diagram becomes at the level of completed local rings

$$\begin{tikzcd}
\mathrm{Spf}(\mathbb{Z}_p[\![X]\!] \mathbin{\widehat{\otimes}} \widehat{\mathcal{O}}_{\bar{\lambda}^\perp\otimes\mathcal{E},\overline{x'}}) \arrow[r] \arrow[d] & \mathrm{Spf}(\widehat{\mathcal{O}}_{\bar{\lambda}^\perp\otimes\mathcal{E},\overline{x'}}) \arrow[r] \arrow[d] & \mathrm{Spf}(\widehat{\mathcal{O}}_{\mathcal{M}^h_{\lambda,\Upsilon,z'}}) \arrow[d, "\simeq"] \\
\mathrm{Spf}(\mathbb{Z}_p[\![X]\!] \mathbin{\widehat{\otimes}} \widehat{\mathcal{O}}_{D\otimes\mathcal{E},\bar{x}}) \arrow[r] & \mathrm{Spf}(\widehat{\mathcal{O}}_{D\otimes\mathcal{E},\bar{x}}) \arrow[r] & \mathrm{Spf}(\widehat{\mathcal{O}}_{\mathcal{M}^h_{\Upsilon,z}})
\end{tikzcd}$$

where the vertical map is contained in the kernel of the map (2D.1). By considering all the $\lambda \in J^\perp$ that map to a given class $\bar{\lambda} \in D$, we get that the image is exactly the kernel of the map (2D.1) and hence the image of left vertical map is defined by the ideal $\mathbb{Z}_p[\![X]\!] \mathbin{\widehat{\otimes}} \hat{I}_\lambda$, see [Zemel 2020, Equation (26)] for a description over $\mathbb{C}$. Finally, since $\widehat{\mathcal{Z}}(m)$ is equal to the union of such images, the conclusion follows. □

2D2. *Special divisors along type-III boundary components.* Let (Ξ, σ) be a stratum representative of type III. Let $K_I = I^\perp/I$ be the Lorentzian lattice as introduced in Section 2B2 and we continue to assume that σ is a one-dimensional inner ray. Let $\omega \in K_I \cap C_\Xi$ be a generator of σ with $(\omega.\omega) = -2N$, $N \geq 1$. Let $\omega' \in K_I^\vee$ be an element such that $(\omega.\omega') = 1$.

Let $\lambda \in L$ with $Q(\lambda) = m$ and such that $\lambda \in I^\perp$. The projection $\bar{\lambda} \in K_I$ defines a divisor in the torus $\mathcal{T}_\Xi = \mathrm{Spec}(\mathbb{Z}_{(p)}[q^\alpha]_{\alpha\in\Gamma_\Xi^\vee})$ given by the equation $q^{\bar{\lambda}} = 1$.

In the partial compactification $\mathcal{T}_\Xi \hookrightarrow \mathcal{T}_\Xi(\sigma)$, the equation of this divisor becomes $q^{\bar{\lambda}}-1=0$ if $(\bar{\lambda}.\omega)\geq 0$ or $q^{-\bar{\lambda}}-1=0$ otherwise. Notice also that this divisor intersects the toric boundary divisor defined by σ if and only if $(\omega.\bar{\lambda})=0$. We will hence restrict ourselves to this latter situation and we denote by

$$\mathcal{T}_\Xi(\lambda,\sigma)\hookrightarrow \mathcal{T}_\Xi(\sigma)$$

the divisor defined by λ. By construction, it only depends on the class of λ in K_I.

Proposition 2.5. *Let $\widehat{\mathcal{Z}}(m)$ be the formal completion of $\mathcal{Z}(m)$ along the boundary component of $\mathcal{M}^\Xi$ index by (Ξ,σ). Then the following diagram is commutative and compatible with Proposition 2.1.*

$$\begin{tikzcd}
\bigsqcup_{a\in\mathbb{Q}^\times_{>0}\backslash\mathbb{A}^\times_f/K_0}\bigsqcup_{\lambda\in K,\,Q(\lambda)=m,\,(\lambda.\omega)=0}\widehat{\mathcal{T}}_{\Xi,0}(\lambda,\sigma)_{/\mathbb{C}} \arrow[r] \arrow[d] & \bigsqcup_{a\in\mathbb{Q}^\times_{>0}\backslash\mathbb{A}^\times_f/K_0}\widehat{\mathcal{T}}_{\Xi,0}(\sigma) \arrow[d] \\
\widehat{\mathcal{Z}}(m) \arrow[r] & \widehat{\mathcal{M}}^\Sigma
\end{tikzcd}$$

The vertical maps are étale coverings of formal Deligne–Mumford stacks and the union over $\lambda\in I^\perp$ covers $\widehat{\mathcal{Z}}(m)$.

Proof. Let $\lambda\in L\cap I^\perp$ with $Q(\lambda)=m$ and such that projection $\bar{\lambda}\in I^\perp/I$ is orthogonal to ω. Then we have similarly a description of the mixed Shimura variety $\mathcal{M}_{\lambda,\Xi}$ associated to the Shimura datum $(G_\lambda,\mathcal{D}_\lambda)$ as a torus fibration and such that the following diagram is commutative:

$$\begin{tikzcd}
\widehat{\mathcal{M}}_{\lambda,\Xi}(\sigma) \arrow[r, "\widehat{\mathcal{T}}_{\Xi,0}(\lambda,\sigma)"] \arrow[d] & \mathcal{S}(\mathbb{G}_m,\mathcal{H}_0)_{/R} \arrow[d] \\
\widehat{\mathcal{M}}_\Xi(\sigma) \arrow[r, "\widehat{\mathcal{T}}_{\Xi,0}(\sigma)"] & \mathcal{S}(\mathbb{G}_m,\mathcal{H}_0)_{/R}
\end{tikzcd}$$

The left vertical map is equivariant with respect to the inclusion $\widehat{\mathcal{T}}_\Xi(\lambda,\sigma)\hookrightarrow\widehat{\mathcal{T}}_\Xi(\sigma)$ and its image only depends on $\bar{\lambda}\in I^\perp/I$. Since the formal completion $\widehat{\mathcal{Z}}(m)$ is the union over $\lambda\in L$ of the images of the left vertical maps, we get the desired result. □

3. Arithmetic intersection theory and modularity

We recall in this section the Arakelov arithmetic intersection theory on $\mathcal{M}^\Sigma$ following [Bruinier et al. 2007], the modularity results of the special divisors from [Howard and Madapusi Pera 2020; Borcherds 1999] and its extension to complex toroidal compactification by [Bruinier and Zemel 2022]. Then we derive a further extension to the integral model of the toroidal compactifications of GSpin Shimura varieties.

3A. ***Modularity of special divisors.*** Let (L,Q) be a maximal quadratic lattice with signature $(b,3)$ and assume that $b\geq 3$.

Let $K \subset G(\mathbb{A}_f)$ be the compact open subgroup from Section 2A and let Σ be a K-admissible smooth polyhedral cone decomposition. Denote by $\mathcal{M}^{\Sigma}$ the toroidal compactification of the integral model of the GSpin Shimura variety constructed in Section 2C. Let $\widehat{\mathrm{CH}}^1(\mathcal{M}^{\Sigma}, \mathcal{D}_{\mathrm{pre}})_{\mathbb{Q}}$ be the first Chow group of prelog forms as defined in [Bruinier et al. 2007, Definition 1.15].

Let Υ be a cusp label representative of type II. Then there is a unique one-dimensional ray in the cone decomposition associated to Υ and we denote by abuse of notation $\mathcal{B}^{\Upsilon}$ the closure of the boundary divisor associated to Υ.

Consider now (Ξ, σ) a toroidal stratum representative of type III such that σ is a one-dimensional inner ray in the cone decomposition Σ. Then we denote by $\mathcal{B}^{\Xi,\sigma}$ the closed boundary divisor in $\mathcal{M}^{\Sigma}$ associated to (Ξ, σ).

Let $\beta \in L^{\vee}/L$ and $m \in Q(\beta) + \mathbb{Z}$ with $m > 0$. For every toroidal stratum representative Υ and (Ξ, σ), let $\mu_{\Upsilon}(\beta, m)$ and $\mu_{\Xi,\omega}(\beta, m)$ be the real numbers defined by (4E.1) and (4F.1), see also [Bruinier and Zemel 2022]. Consider then the following divisor on $\mathcal{M}^{\Sigma}$:

$$\mathcal{Z}^{\mathrm{tor}}(\beta, m) = \mathcal{Z}(\beta, m) + \sum_{\Upsilon} \mu_{\Upsilon}(\mu, m) \cdot \mathcal{B}^{\Upsilon} + \sum_{(\Xi,\omega)} \mu_{\Xi,\omega}(\mu, m) \cdot \mathcal{B}^{\Xi,\omega}, \tag{3A.1}$$

where the two last sums are over toroidal stratum representatives of type II and type III respectively. Then by [Bruinier and Zemel 2022], the Cartier divisor $\mathcal{Z}^{\mathrm{tor}}(\beta, m)$ can be endowed with a Green function $\Phi_{\beta,m}$ such that the resulting pair

$$\widehat{\mathcal{Z}}^{\mathrm{tor}}(\beta, m) = (\mathcal{Z}^{\mathrm{tor}}(\beta, m), \Phi_{\beta,m})$$

is an element of the first Chow group of prelog forms $\widehat{\mathrm{CH}}^1(\mathcal{M}^{\Sigma}, \mathcal{D}_{\mathrm{pre}})_{\mathbb{Q}}$. For $m = 0$ and $\beta = 0$, we define $\widehat{\mathcal{Z}}(0, 0)$ to be any arithmetic divisor whose is class is the dual of the hermitian line bundle $\widehat{\mathcal{L}} = (\mathcal{L}, \|\cdot\|_{\mathrm{pet}})$ endowed with the Petersson metric $\|z\|^2 = [z, \bar{z}]$.

Consider then the following generating series

$$\Phi_L := \sum_{\beta \in L^{\vee}/L} \sum_{m \in Q(\beta)+\mathbb{Z}} \widehat{\mathcal{Z}}^{\mathrm{tor}}(\beta, m) q^m e_\beta \in \mathbb{C}[L^{\vee}/L][\![q^{1/D_L}]\!] \otimes \widehat{\mathrm{CH}}^1(\mathcal{M}^{\Sigma}, \mathcal{D}_{\mathrm{pre}})_{\mathbb{Q}},$$

where $(e_\beta)_{\beta \in L^{\vee}/L}$ is a basis of the $\mathbb{C}$-vector space $\mathbb{C}[L^{\vee}/L]$, D_L is the discriminant of L, and $q = e^{2i\pi\tau}$, where $\tau \in \mathbb{H}$ is in the upper-half plane.

Let

$$\rho_L : \mathrm{Mp}_2(\mathbb{Z}) \to \mathrm{Aut}_{\mathbb{C}}(\mathbb{C}[L^{\vee}/L])$$

be the Weil representation associated to the quadratic lattice (L, Q), where $\mathrm{Mp}_2(\mathbb{R})$ is the metaplectic double cover if $\mathrm{Mp}_2(\mathbb{R})$. For $k \in \frac{1}{2}\mathbb{Z}$, let $\mathrm{Mod}_k(\rho_L)$ denote the vector space of vector valued modular forms of weight k with respect to ρ_L. We then have the following theorem.

Theorem 3.1. *The generating series Φ_L is the Fourier development of a vector-valued modular forms of weight $1 + \frac{b}{2}$ and representation ρ_L, i.e.,*

$$\Phi_L \in \mathrm{Mod}_{1+b/2}(\rho_L) \otimes \widehat{\mathrm{CH}}^1(\mathcal{M}^{\Sigma}, \mathcal{D}_{\mathrm{pre}})_{\mathbb{Q}}.$$

Proof. Let $F \in M^{!}_{1-b/2}(\overline{\rho_L})$ be a weakly holomorphic modular form of weight $1-\frac{b}{2}$ with respect to the complex conjugate Weil representation of ρ_L such that F has integral principal part, and let Ψ be the associated Borcherds product. Then by [Bruinier and Zemel 2022, Theorem 5.5], the divisor in $\mathcal{M}^{\Sigma}(\mathbb{C})$ of $\Psi(F)_{\mathbb{C}}$ is equal to

$$\sum_{\beta\in L^{\vee}/L}\ \sum_{m\in Q(\beta)+\mathbb{Z}} c_{\beta}(-m)\mathcal{Z}^{\mathrm{tor}}(\beta, m)(\mathbb{C}).$$

Since Borcherds products are defined rationally by [Howard and Madapusi Pera 2020, Theorem A], we only need to check that the divisor of the Borcherds products has the expected form over $\mathbb{Z}$ and this will be automatic if all the special divisors and the boundary divisors are flat. By [Madapusi Pera 2019, Theorem 4.1.5], the boundary divisors are flat and by [Howard and Madapusi Pera 2020, Proposition 7.2.2], the special divisors are flat over $\mathbb{Z}\left[\frac{1}{2}\right]$ and over $\mathbb{Z}$ if $b \geq 4$. For $b = 3$, one can use the algebraic version of the Borcherds embedding trick as in [Howard and Madapusi Pera 2020, Section 9.2] to prove that no further components appear at 2 and hence the divisor of the Borcherds product has the correct form. Hence we conclude by the criterion in [Bruinier and Zemel 2022, Proposition 5.4]. □

4. The main estimates and proof of the main theorems

We state in this section the local and global estimates that will allow us to prove Theorem 1.1 and Theorem 1.3. Then we will prove the global estimates and we postpone the proof of local estimates to the next section.

4A. ***Number field setting.*** Let X be K3 surface over a number field K. Given an embedding $\tau : K \hookrightarrow \mathbb{C}$, let (L, Q) be a maximal lattice containing the transcendental lattice of $X^{\tau}(\mathbb{C})$. It is an even lattice of signature $(b, 2)$ whose genus is independent from τ. We can assume furthermore that $b \geq 3$, as the case $b \leq 2$ has already been treated, see [Charles 2018; Shankar and Tang 2020].

Let $\mathcal{M}$ be the integral model of the GSpin Shimura variety associated to the lattice (L, Q) and, given an admissible polyhedral cone decomposition Σ, let $\mathcal{M}^{\Sigma}$ be its toroidal compactifications as in Section 2. By [Madapusi Pera 2015], the K3 surface has an associated Kuga–Satake abelian variety which we can also assume to be defined over the number field K, up to taking a finite extension. Hence it defines a K-point of $\mathcal{M}^{\Sigma}$. By the extension property of the integral model, there exists $N \geq 1$ such that, up to taking a finite extension of K, we have a flat morphism over $\mathbb{Z}$:

$$\mathrm{Spec}\big(\mathcal{O}_K\left[\tfrac{1}{N}\right]\big) \to \mathcal{M},$$

and by properness, this map extends to

$$\rho : \mathcal{Y} = \mathrm{Spec}(\mathcal{O}_K) \to \mathcal{M}^{\Sigma}.$$

By construction, the image of this map is not contained in any special divisor. A prime over N is said to be a prime of bad reduction and otherwise of good reduction.

As in [Shankar et al. 2022, Theorem 2.4], we will rather prove the following more general version, which is easily seen to imply Theorem 1.1.

Theorem 4.1. *Let $\mathcal{Y} \in \mathcal{M}^{\Sigma}(\mathcal{O}_K)$ with smooth reduction outside N. Let $D \in \mathbb{Z}_{>0}$ be a fixed integer represented by (L, Q) and coprime to N. Assume that $\mathcal{Y}_K \in M(K)$ is not contained in any special divisor $\mathcal{Z}(m)(K)$. Then there are infinitely many places $\mathfrak{P}$ of K of good reduction such that $\mathcal{Y}_{\overline{\mathfrak{P}}}$ lies in the image of $\mathcal{Z}(Dm^2) \to \mathcal{M}$ for some $m \in \mathbb{Z}_{>0}$ coprime to N.*

Let $\rho : \mathcal{Y} \to \mathcal{M}^{\Sigma}$ be as in the previous theorem. We first begin by the following proposition.

Proposition 4.2. *There exists a refinement of the cone decomposition Σ, such that the map $\rho : \mathcal{Y} \to \mathcal{M}^{\Sigma}$ satisfies the following property: for any prime $\mathfrak{P}$ of bad reduction, the image of the closed point $\{\mathfrak{P}\}$ under ρ is contained in a stratum which is a locally closed divisor of $\mathcal{M}^{\Sigma}$.*

Proof. Let $s_{\mathfrak{P}} \in \mathcal{Y}$ be the closed point $\mathfrak{P}$ of $\mathcal{Y}$. By (2C.1), the image of $s_{\mathfrak{P}}$ lies in a stratum indexed either by either a type-II boundary component Υ or a type-III (Ξ, σ) toroidal stratum representative. In the type-II case, the boundary is already a divisor and there is nothing to prove. In the type-III case, let r be the dimension of the cone σ. Then we get a morphism

$$\operatorname{Spf}(W(\overline{\mathbb{F}}_p)) \to \widehat{\mathcal{M}}^{\Sigma}, \tag{4A.1}$$

where $\widehat{\mathcal{M}}^{\Sigma}$ is the formal completion along the boundary component defined by (Ξ, σ). By a similar analysis to Section 2C2, we have an étale cover of formal Deligne–Mumford stacks

$$\widehat{\mathcal{T}}_{\Xi}(\sigma) \to \widehat{\mathcal{M}}^{\Sigma}.$$

Hence the map (4A.1) lifts to a morphism

$$\operatorname{Spf}(W(\overline{\mathbb{F}}_p)) \to \widehat{\mathcal{T}}_{\Xi}(\sigma), \tag{4A.2}$$

where

$$\widehat{\mathcal{T}}_{\Xi}(\sigma) = \operatorname{Spf}(\mathbb{Z}_p[q^{\alpha} \mid (\alpha, \sigma) = 0] \otimes_{\mathbb{Z}_p} \mathbb{Z}_p[\![q^{\alpha} \mid (\alpha, \sigma) > 0]\!]).$$

Hence this corresponds to a map

$$\mathbb{Z}_p[\![q^{\alpha} \mid (\alpha, C) > 0]\!] \otimes \mathbb{Z}_p[q^{\alpha} \mid (\alpha, C) = 0] \to W(\overline{\mathbb{F}}_p).$$

The linear form on $\Gamma_{\Xi}^{\vee}$ given by sending an element α to the p-adic valuation of the image of q^{α} under the above map is represented by an element $\omega \in \Gamma_{\Xi}$ which satisfies $(\omega.\alpha) > 0$ whenever $(\alpha.\sigma) > 0$; hence ω is in σ. The cocharacter defined by ω is in fact tangent to the map given in (4A.2). Let σ' in σ be the ray defined by ω and let Σ' be the new cone decomposition obtained by refining Σ and which contains σ' as a one-dimensional ray. Then $\mathcal{M}^{\Sigma'}$ is a blow-up of $\mathcal{M}^{\Sigma}$ and by the preceding discussion, the point $s_{\mathfrak{P}}$ belongs to the boundary divisor parametrized by (Ξ, σ'). Since there are only finitely many primes of bad reduction, then by repeating this procedure finitely many times, we get the desired cone decomposition. □

We will work from now on with the toroidal compactification given by the above proposition. For $m \geq 1$ an integer, let $\mathcal{Z}(m)$ be the closed special divisor $\mathcal{Z}(0, m) \hookrightarrow \mathcal{M}^{\Sigma}$ and $\widehat{\mathcal{Z}}^{\mathrm{tor}}(m)$ the arithmetic divisor associated to $\mathcal{Z}^{\mathrm{tor}}(m)$ by (3A.1). The pullback via the period map $\rho : \mathcal{Y} \to \mathcal{M}^{\Sigma}$ allows us to define the height $h_{\widehat{\mathcal{Z}}^{\mathrm{tor}}(m)}(\mathcal{Y})$ of $\mathcal{Y}$ with respect to the arithmetic divisor $\widehat{\mathcal{Z}}^{\mathrm{tor}}(m)$ as its image under the composition

$$\widehat{\mathrm{CH}}^1(\mathcal{M}^{\Sigma}, \mathcal{D}_{\mathrm{pre}})_{\mathbb{Q}} \to \widehat{\mathrm{CH}}^1(\mathcal{Y}, \mathcal{D}_{\mathrm{pre}})_{\mathbb{Q}} \xrightarrow{\widehat{\deg}} \mathbb{R}, \quad \widehat{\mathcal{Z}}^{\mathrm{tor}}(m) \to h_{\widehat{\mathcal{Z}}^{\mathrm{tor}}(m)}(\mathcal{Y}).$$

By choice of the lattice (L, Q), the arithmetic curve $\mathcal{Y}$ intersects properly the divisors $\mathcal{Z}(m)$, $\mathcal{B}^{\Xi,\omega}$ and $\mathcal{B}^{\Upsilon}$ for every Υ and (Ξ, ω). Hence we have

$$h_{\widehat{\mathcal{Z}}^{\mathrm{tor}}(m)}(\mathcal{Y}) = \sum_{\tau : K \hookrightarrow \mathbb{C}} \Phi_m(\mathcal{Y}^{\tau}) + \sum_{\mathfrak{P}} (\mathcal{Y}.\mathcal{Z}^{\mathrm{tor}}(m))_{\mathfrak{P}} \log|\mathcal{O}_K/\mathfrak{P}|, \tag{4A.3}$$

where for $\tau : K \hookrightarrow \mathbb{C}$, we use $\mathcal{Y}^{\tau}$ to denote the point in $M(\mathbb{C})$ induced by

$$\mathrm{Spec}(\mathbb{C}) \xrightarrow{\tau} \mathrm{Spec}(\mathcal{O}_K) = \mathcal{Y} \to \mathcal{M}^{\Sigma}.$$

We have

$$(\mathcal{Y}.\mathcal{Z}^{\mathrm{tor}}(m))_{\mathfrak{P}} = (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} + \sum_{\Upsilon} \mu_{\Upsilon}(m)(\mathcal{Y}.\mathcal{B}^{\Upsilon})_{\mathfrak{P}} + \sum_{(\Xi,\omega)} \mu_{\Xi,\omega}(m) \cdot (\mathcal{Y}.\mathcal{B}^{\Xi,\omega})_{\mathfrak{P}}. \tag{4A.4}$$

Let us denote by $\mathcal{O}_{\mathcal{Y}\times_{\mathcal{M}^{\Sigma}}\mathcal{Z}(m),v}$ the étale local ring of $\mathcal{Y} \times_{\mathcal{M}} \mathcal{Z}(m)$ at v. Then

$$(\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = \sum_{v \in \mathcal{Y}\times_{\mathcal{M}}\mathcal{Z}(m)(\overline{\mathbb{F}}_{\mathfrak{P}})} \mathrm{length}(\mathcal{O}_{\mathcal{Y}\times_{\mathcal{M}^{\Sigma}}\mathcal{Z}(m),v}), \tag{4A.5}$$

where $\mathbb{F}_{\mathfrak{P}}$ denotes the residue field of $\mathfrak{P}$.

Let

$$(\mathcal{Y}.\mathcal{Z}(m)) = \sum_{\mathfrak{P}} (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} \log|\mathcal{O}_K/\mathfrak{P}|.$$

The first new contribution of this paper is to prove the following estimate which results from Borcherds modularity and ad hoc bounds on the multiplicities $\mu_{\Upsilon}(m)$ and $\mu_{\Xi,\omega}(m)$.

Proposition 4.3. *As $m \to \infty$, we have*

$$(\mathcal{Y}.\mathcal{Z}(m)) + \sum_{\tau : K \hookrightarrow \mathbb{C}} \Phi_m(\mathcal{Y}^{\tau}) = O(m^b/2).$$

As a corollary, we get the following bound, which is referred to as the diophantine bound in [Shankar et al. 2022, Equation (5.2)].

Corollary 4.4. *For any finite place $\mathfrak{P}$, we have the following bound*:

$$(\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = O(m^{b/2} \log m), \quad \Phi_m(\mathcal{Y}^{\tau}) = O(m^{b/2} \log m).$$

For our next estimate, we recall the notion of asymptotic density from [Shankar et al. 2022]: for a subset $S \subset \mathbb{Z}_{>0}$, the *logarithmic asymptotic density* of S is defined to be

$$\limsup_{X\to\infty} \frac{\log|S_X|}{\log X},$$

where $S_X := \{a \in S \mid X \leq a < 2X\}$.

Recall from Theorems 5.7 and 6.1 in [loc. cit.] that we have the following estimate:

Proposition 4.5. *There exists a subset $S_{\mathrm{bad}} \subset \mathbb{Z}_{>0}$ of zero logarithmic asymptotic density such that*

$$\sum_{\tau:K\hookrightarrow\mathbb{C}} \Phi_m(\mathcal{Y}^\tau) = c(m)\log(m) + o(m^{b/2}\log(m)),$$

where $-c(m) \asymp m^{b/2}$ and is defined in [loc. cit., Section 3.3].

For a prime $\mathfrak{P}$ of good reduction, i.e., where the intersection of $\mathcal{Y}$ and $\mathcal{Z}(m)$ above $\mathfrak{P}$ is supported in $\mathcal{M}$, we have the following estimate which follows easily from [loc. cit., Theorem 7.1].

Proposition 4.6. *Let $\mathfrak{P}$ be a finite place of good reduction. Let $D \in \mathbb{Z}_{\geq 1}$ coprime to N. For $X \in \mathbb{Z}_{>0}$, let $S_{D,X}$ denote the set*

$$\Big\{m \in \mathbb{Z}_{>0} \;\Big|\; X \leq m < 2X,\ \frac{m}{D} \in \mathbb{Z} \cap (\mathbb{Q}^\times)^2,\ (m, N) = 1\Big\}.$$

Then we have

$$\sum_{m\in S_{D,X}} (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = o(X^{(b+1)/2}\log X).$$

Finally, for a prime $\mathfrak{P}$ of bad reduction, we prove the following proposition which is the second new contribution of this paper.

Proposition 4.7. *Let $\mathfrak{P}$ a finite place of bad reduction. Let $D \in \mathbb{Z}_{\geq 1}$ coprime to N. For $X \in \mathbb{Z}_{>0}$, let $S_{D,X}$ be the set defined in the previous proposition. Then we have*

$$\sum_{m\in S_{D,X}} (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = o(X^{(b+1)/2}\log X).$$

4B. ***Function field setting.*** We assume in this section that the lattice (L, Q) is self-dual at p. Then the Shimura variety $\mathcal{M}$ has smooth reduction at p and we denote its reduction by $\mathcal{M}_{\mathbb{F}_p}$. Given an admissible cone decomposition Σ, we denote by $\mathcal{M}^{\Sigma}_{\mathbb{F}_p}$ the reduction of the toroidal compactification $\mathcal{M}^{\Sigma}$. We first give a new formulation of Theorem 1.3, see Theorem 4.8, then we will give the main estimates that will allow us to prove the latter.

Let $\mathcal{X} \to \mathscr{S}$ be a generically ordinary nonisotrivial family of K3 surfaces over a smooth curve $\mathscr{S}$ over $\overline{\mathbb{F}}_p$. The quadratic lattice (L, Q) in this case corresponds to a maximal quadratic lattice orthogonal to the generic geometric Picard group in the K3 lattice. Hence (L, Q) has discriminant coprime to p by assumption and we get a period map by [Madapusi Pera 2015, section 4]

$$\rho : \mathscr{S} \to \mathcal{M}_{\mathbb{F}_p},$$

which is a finite map and the image of the generic point is in the ordinary locus. The locus in $\mathscr{S}$ where the Picard rank jumps corresponds then exactly to the union over $m \geq 1$ of the intersections $\mathscr{S} \cap \mathcal{Z}(m)_{\mathbb{F}_p}$. Hence Theorem 1.3 follows from the following theorem.

Theorem 4.8. *Let $\mathscr{S} \to \mathcal{M}_{\mathbb{F}_p}$ be a finite map with generically ordinary image and not contained in any special divisor. Then there exists infinitely many closed points in $\mathscr{S}$ in the union of $\mathcal{Z}(m)_{\mathbb{F}_p}$ for integers m coprime with p.*

Let $\mathscr{S}$ be a smooth curve as in the theorem above. By properness, we can extend the map

$$\rho : \overline{\mathscr{S}} \to \mathcal{M}^{\Sigma}_{\mathbb{F}_p},$$

where $\overline{\mathscr{S}}$ is the smooth compactification of $\mathscr{S}$. We have the following proposition whose proof is similar to Proposition 4.2 and hence we omit it.

Proposition 4.9. *There exists a refinement of the cone decomposition Σ such that the image of $\overline{\mathscr{S}}$ in $\mathcal{M}_{\overline{\mathbb{F}}_p}$ intersects the boundary only in strata corresponding to locally closed divisors.*

Let Σ be a polyhedral cone decomposition which satisfies the conditions of the previous proposition. By abuse of notation, if $D \subset \mathcal{M}^{\Sigma}_{\mathbb{F}_p}$ is a Cartier divisor, we write

$$(D.\overline{\mathscr{S}}) = \deg_{\overline{\mathscr{S}}} \rho^* D.$$

We have then the following global estimate.

Proposition 4.10. *As $m \to \infty$, we have*

$$(\mathcal{Z}(m)_{\overline{\mathbb{F}}_p}.\overline{\mathscr{S}}) = |c(m)|(\overline{\mathscr{S}}.\mathcal{L}_{\mathbb{F}_p}) + o(m^{b/2}).$$

For any integer m, we have the decomposition

$$(\mathcal{Z}(m)_{\overline{\mathbb{F}}_p}.\overline{\mathscr{S}}) = \sum_{P \in \overline{\mathbb{F}}_p} m_P(\mathcal{Z}(m)_{\mathbb{F}_p}, \overline{\mathscr{S}}),$$

where $m_P(\mathcal{Z}(m)_{\mathbb{F}_p}, \overline{\mathscr{S}})$ is the multiplicity of intersection at P. Our next goal is to estimate in average these local multiplicities and we start by the good reduction case already treated in [Maulik et al. 2022a, Proposition 7.11, Theorem 7.18].

Let S be as in [loc. cit., Section 7.1], i.e., a set of integers of positive density such that every $m \in S$ is coprime to p and is representable by the quadratic lattice (L, Q).

For $P \in (\mathscr{S} \cap \mathcal{M})(\mathbb{F}_p)$, we define as in [loc. cit., Definition 7.6]

$$g_P(m) = \frac{h_P}{p-1}|c(m)|,$$

where h_P is the order of vanishing of the Hasse invariant at P, see [loc. cit.] The following proposition is the combination of Proposition 7.11 and Theorem 7.18 from [loc. cit.].

Proposition 4.11. *Let $P \in \mathscr{S}(\overline{\mathbb{F}}_p)$. Then*:

(1) *If P is not supersingular then*

$$\sum_{m \in S_X} m_P(\mathcal{Z}(m)_{\mathbb{F}_p}.\mathscr{S}) = O(X^{b/2} \log X).$$

(2) *There exists an absolute constant $0 < \alpha < 1$ such that for any supersingular point P we have*

$$\sum_{m \in S_X} m_P(\mathcal{Z}(m)_{\mathbb{F}_p}.\mathscr{S}) = \alpha \sum_{m \in S_X} g_p(m) + O(X^{(b+1)/2}).$$

Our new contribution in this setting is the following theorem which gives an estimate on intersection multiplicities at points where $\bar{S}$ intersects the boundary of $\mathcal{M}_p^{\Sigma}$.

Proposition 4.12. *Let $P \in \overline{\mathscr{S}}(\mathbb{F}_p)$ a point mapping to the boundary of $\mathcal{M}_{\mathbb{F}_p}^{\Sigma}$. Then we have the following estimate*:

$$\sum_{m \in S_X} m_P(\mathcal{Z}(m)_{\mathbb{F}_p}.\overline{\mathscr{S}}) = O(X^{b/2} \log X).$$

4C. *Proof of the main theorems.* Assuming the estimates in the previous section we now indicate how to prove Theorem 1.1 and Theorem 1.3.

Proof of Theorem 1.1. It is enough to prove Theorem 4.1 in a similar way to [Shankar et al. 2022, Section 8]. For convenience of the reader, we will sketch the proof. Assume for the sake of contradiction that there are only finitely many primes of good reduction such that $\mathcal{Y}$ intersects a special divisor of the form $\mathcal{Z}(Dm^2)$ where Dm^2 is coprime with N and is represented by (L, Q). By Proposition 4.3 and Proposition 4.5, there exists a subset $S_{\text{bad}} \subset \mathbb{Z}_{>0}$ of logarithmic asymptotic density zero such that

$$(\mathcal{Y}.\mathcal{Z}(m)) = -c(m)\log(m) + o(m^{b/2}\log(m)) \asymp m^{b/2}\log(m).$$

Let $S_{D,X}^{\text{good}} = \{m \in S_{D,X}, m \notin S_{\text{bad}}, (m, N) = 1\}$, then one can easily check that $|S_{D,X}^{\text{good}}| \asymp X^{1/2}$ and $c(m) \gg X^{b/2} \log X$ for $m \in S_{D,X}^{\text{good}}$. Hence we get

$$\sum_{m \in S_{D,X}^{\text{good}}} (\mathcal{Y}.\mathcal{Z}(m)) \asymp X^{(b+1)/2} \log X. \tag{4C.1}$$

On the other hand, by Propositions 4.5 and 4.7, we get by summing over the finitely many places where either $\mathcal{Y}$ intersects a $\mathcal{Z}(Dm^2)$ or which are of bad reduction

$$\sum_{m \in S_{D,X}^{\text{good}}} (\mathcal{Y}.\mathcal{Z}(m)) = o(X^{(b+1)/2} \log X),$$

which contradicts (4C.1). □

Proof Theorem 1.3. The proof is similar: assume that there are only finitely many points in the union $\big(\bigcup_{m, m \wedge p = 1} \mathcal{Z}(m) \cap \mathscr{S}\big)(\overline{\mathbb{F}}_p)$ and let S be a set as in Section 4B. Then by Proposition 4.10, we have

$$\sum_{m \in S_X} (\mathcal{Z}(m)_{\mathbb{F}_p}.\overline{\mathscr{S}}) = \sum_{m \in S_X} |c(m)|(\overline{\mathscr{S}}.\mathcal{L}_{\mathbb{F}_p}) + o(X^{b/2+1}).$$

On the other hand, by Propositions 4.11 and 4.12 we have

$$\begin{aligned}\sum_{m\in S_X}(\mathcal{Z}(m)_{\mathbb{F}_p}.\overline{\mathscr{S}}) &= \sum_{m\in S_X}\ \sum_{P\in\left(\bigcup_{m,m\wedge p=1}\mathcal{Z}(m)\cap\mathscr{S}\right)(\bar{\mathbb{F}}_p)} m_P(\mathcal{Z}(m)_{\mathbb{F}_p}.\overline{\mathscr{S}})\\ &=\alpha\sum_{m\in S_X}|g_P(m)|+O(X^{(b+1)/2}).\\ &\le\alpha\sum_{m\in S_X}|c(m)|(\overline{\mathscr{S}}.\mathcal{L}_{\mathbb{F}_p})+O(X^{(b+1)/2}),\end{aligned}$$

where the last equality results from the fact that the Hasse invariant is a section of $\mathcal{L}_{\mathbb{F}_p}^{\otimes p-1}$. These two estimates contradict each other, hence the result. □

4D. *Global estimate.* We prove in this section simultaneously Propositions 4.3 and 4.10.

By Theorem 3.1, the generating series

$$\sum_{\beta\in L^\vee/L}\ \sum_{m\in Q(\beta)+\mathbb{Z}} h_{\widehat{\mathcal{Z}}^{\mathrm{tor}}(\beta,m)}(\mathcal{Y})q^m e_\beta$$

and

$$\sum_{\beta\in L^\vee/L}\ \sum_{m\in Q(\beta)+\mathbb{Z}} (\mathcal{Z}^{\mathrm{tor}}(\beta,m)_{\mathbb{F}_p}.\overline{\mathscr{S}})q^m e_\beta$$

are elements of $\mathrm{Mod}_{1+b/2}(\rho_L)$. Classical estimates on the growth of coefficients of modular forms imply that (see [Tayou 2020, Example 2.3] for more details)

$$h_{\widehat{\mathcal{Z}}^{\mathrm{tor}}(m)}(\mathcal{Y})=O(m^{b/2})$$

and

$$(\mathcal{Z}(m)_{\mathbb{F}_p}^{\mathrm{tor}}.\overline{\mathscr{S}})=|c(m)|(\overline{\mathscr{S}}.\mathcal{L}_{\mathbb{F}_p})+o(m^{b/2}).$$

By (4A.3) and (4A.4), we can write

$$\begin{aligned}&(\mathcal{Y}.\mathcal{Z}(m))+\sum_{\tau:K\hookrightarrow\mathbb{C}}\Phi_m(\mathcal{Y}^\tau)\\ &\quad=h_{\widehat{\mathcal{Z}}^{\mathrm{tor}}(m)}(\mathcal{Y})-\sum_{\Upsilon}\mu_\Upsilon(m)(\mathcal{Y}.\mathcal{B}^\Upsilon)_{\mathfrak{P}}\log|\mathcal{O}_K/\mathfrak{P}|-\sum_{\Xi}\mu_{\Xi,\sigma}(m)\cdot(\mathcal{Y}.\mathcal{B}^{\Xi,\sigma})_{\mathfrak{P}}\log|\mathcal{O}_K/\mathfrak{P}|\end{aligned}\tag{4D.1}$$

and similarly, we can write

$$(\overline{\mathscr{S}}.\mathcal{Z}(m)_{\mathbb{F}_p})=(\mathcal{Z}^{\mathrm{tor}}(m)_{\mathbb{F}_p}.\overline{\mathscr{S}})-\sum_{\Upsilon}\mu_\Upsilon(m)(\overline{\mathscr{S}}.\mathcal{B}^{\Upsilon,\mathbb{F}_p})-\sum_{\Xi,\sigma}\mu_{\Xi,\omega}(m)\cdot(\overline{\mathscr{S}}.\mathcal{B}_{\mathbb{F}_p}^{\Xi,\sigma}).\tag{4D.2}$$

Hence we only have to bound the growth of the multiplicities $\mu_\Upsilon(m)$ and $\mu_{\Xi,\omega}(m)$.[5] This is given by the following lemma.

Proposition 4.13. *As $m\to\infty$, we have the following estimates*:

(1) *For any type-II cusp label representative Υ, we have*

$$\mu_\Upsilon(m)\ll_\epsilon m^{b/2-1+\epsilon}.$$

[5] ω is the unique integral generator of σ.

(2) *For any type-III toroidal stratum representative* (Ξ, σ) *such that* σ *is a ray*, *we have*

$$\mu_{\Xi,\omega}(m) \ll_\epsilon m^{(b-1)/2+\epsilon}.$$

This proposition will be proved in the following two sections.

4E. ***Estimates on type-II multiplicities.*** The goal of this section is to prove the type-II estimate in Proposition 4.13. First we recall some notation associated to isotropic planes introduced in [Bruinier and Zemel 2022, Section 3.2].

Let $\Upsilon = (P, \mathcal{D}^\circ, h)$ be a cusp label representative corresponding to a boundary component of type II. Recall from Section 2B1 that P is the stabilizer of an isotropic plane $J_\mathbb{Q}$ and $J = J_\mathbb{Q} \cap h.L$ is a primitive isotropic plane of $h.L \cap L_\mathbb{Q}$.

To simplify the notation, assume that $h.L \cap L_\mathbb{Q} = L$, the reader may otherwise replace L by $L_h = h.L \cap L_\mathbb{Q}$ in what follows. Define then

$$J_{L^\vee} = J_\mathbb{R} \cap L^\vee, \quad J^\perp_L = J^\perp \cap L, \quad J^\perp_{L^\vee} = J^\perp \cap L^\vee, \quad \text{and} \quad D = J^\perp_L / J.$$

The lattice D is positive definite lattice of rank $b-2$. Its dual lattice can be described as

$$D^\vee = J^\perp_{L^\vee} / J_{L^\vee}.$$

and the discriminant lattice is given by

$$\Delta_D = D^\vee / D = J^\perp_{L^\vee} / (J^\perp_L + J_{L^\vee} = L^\vee_J / (L + J_{L^\vee}),$$

where $L^\vee_J$ is the subgroup of $L^\vee$

$$L + J^\perp_{L^\vee} = \{\mu \in L^\vee \mid \exists \nu \in L \text{ such that } (\mu, \lambda) = (\nu, \lambda)\ \forall \lambda \in J\}.$$

Let Θ_D denote the vector-valued Theta function associated to D defined by

$$\Theta_D(\tau) = \sum_{\beta \in D^\vee} q^{Q(\beta)} e_{\beta + D} \in \mathbb{C}[\Delta_D][\![q^{1/|\Delta_D|}]\!].$$

It is an element of $M_{b/2-1}(\rho_D)$, which is the space of vector-valued modular forms of weight $\frac{b}{2}-1$ with respect to the Weil representation ρ_D associated to the positive definite lattice (D, Q). We can also write

$$\Theta_D(\tau) = \sum_{\beta \in D^\vee / D} \sum_{m \geq 0} c(D, \beta, m) q^m e_\beta,$$

where for $\beta \in D^\vee / D$, $m \in Q(\beta) + \mathbb{Z}$, $m \geq 0$, we have

$$c(D, \beta, m) = |\{\lambda \in \beta + D,\ Q(\lambda) = m\}|.$$

Following Bruinier and Zemel's notation [2022, Section 4.4], define

$$\uparrow^L_D (\Theta_D)(\tau) = \sum_{\beta \in J^\perp_{L^\vee} / J} q^{Q(\beta)/2} e_{\beta + L} = \sum_{\beta \in L^\vee / L} \sum_{m \in Q(\beta) + L} c(D, \beta, m) q^m e_\beta \in M_{b/2-1}(\rho_L),$$

where $c(D,\beta,m)=0$ if $\beta\notin J_{L^\vee}^{\perp}/J^{\perp}$ or $m\notin Q(\beta)+\mathbb{Z}$, and otherwise $c(D,\beta,m)=c(D,\bar{\beta},m)$ where $\bar{\beta}$ is the image of β under the reduction map $J_{L^\vee}^{\perp}\to D^\vee/D$.

In particular, we have

$$q\frac{d}{dq}\uparrow_D^L(\Theta_D)(\tau)=\sum_{\beta\in L^\vee/L}\ \sum_{m\in Q(\beta)+L} mc(D,\beta,m)q^m e_\beta,$$

which is a quasimodular form in the sense of [Imamoğlu et al. 2014, Definition 1].

Then by [Bruinier and Zemel 2022, Definition 4.18, Proposition 4.21 4.15], we can define

$$\mu_\Upsilon(m)=\frac{1}{b-2}\,\mathrm{CT}\left(\left\langle q\frac{d}{dq}\uparrow_D^L(\Theta_D),\,F_m^+\right\rangle_L\right),\tag{4E.1}$$

where F_m^+ is the holomorphic part of the harmonic mass form $F_{m,0}$ from [Bruinier and Zemel 2022, Proposition 4.2]. A direct computation shows then (see also the second formula in [Bruinier 2002, Theorem 2.14])

$$\mu_\Upsilon(m)=\frac{2}{b-2}mc(D,0,m).$$

Classical estimates on coefficients of modular forms, see for example [Sarnak 1990, Proposition 1.5.5], show that

$$|c(D,\beta,m)|\ll_\epsilon m^{b/2-2+\epsilon}\tag{4E.2}$$

for all $\epsilon>0$. Hence we get that

$$|\mu_\Upsilon(m)|\ll_\epsilon m^{b/2-1+\epsilon},$$

which proves the first part of Proposition 4.13.

4F. ***Estimates on type-III multiplicities.*** In this section, we prove the estimates on the type-III multiplicities in Proposition 4.13.

Let (Ξ,σ) be a toroidal stratum representative of type III such that σ is a ray. Keeping the notation from Section 2B2, let $I_{\mathbb{Q}}$ be the isotropic line of $L_{\mathbb{Q}}$ whose stabilizer is the parabolic subgroup attached to Ξ and let $I=I_{\mathbb{Q}}\cap h.L$. To simplify notation, we assume that $h.L=L$, the reader may notice that this is harmless, up to replacing L by $h.L$ in what follows.

The line I is an isotropic line of L and the lattice $K_I=I^\perp/I$ is Lorentzian. Let $C_{\mathbb{R}}$ be the cone of negative elements of the Lorentzian space $K_{I,\mathbb{R}}$ and let $C=C_{\mathbb{R}}\cap K$. As is explained in Section 2B2, the ray ω is generated by an element $\omega\in K_I\cap C$ which is primitive and such that $Q(\omega)=-N$. Following [Bruinier and Zemel 2022, Definition 4.18], we define

$$\mu_{\Xi,\omega}(m)=\frac{\sqrt{N}}{8\sqrt{2}\pi}\Phi_m^K\left(\frac{\omega}{\sqrt{N}}\right).\tag{4F.1}$$

Let $v=\frac{\omega}{\sqrt{N}}$. By [Bruinier 2002, Proposition 2.11 and Theorem 2.14], we have

$$\Phi_m^K(v)=\Phi_m^K\bigl(v,\tfrac{1}{2}+\tfrac{b}{4}\bigr)=\frac{2\Gamma\bigl(\frac{b-1}{2}\bigr)(4\pi m)^{b/2}}{1+\frac{b}{2}}\sum_{\lambda\in K_I,\,Q(\lambda)=m}\frac{F\bigl(\frac{b-1}{2},1,1+\frac{b}{2};\frac{m}{(Q(\lambda_{v^\perp}))}\bigr)}{(4\pi|Q(\lambda_{v^\perp})^{(b-1)/2}|)},$$

where $F(a,b,c;z)$ is the usual Gauss hypergeometric function given by

$$F(a,b,c;z)=\sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!},$$

and $(a)_n=\Gamma(a+n)/\Gamma(a)$. Recall that the above series has 1 as a radius of convergence and converges absolutely in the unit circle $|z|=1$ if $\mathcal{R}(c-a-b)>0$. In our situation, the latter quantity is equal to $1+b/2-1-(b-1)/2=\frac{1}{2}>0$. Hence the series $F((b-1)/2,1,1+b/2;z)$ is globally bounded over the unit disc. For $\lambda\in K$ such that $Q(\lambda)=m$, we have $m=Q(\lambda_v)+Q(\lambda_{v^\perp})$ and $Q(\lambda_v)\le 0$, hence $0<m\le Q(\lambda_{v^\perp})$. Hence we get

$$|\Phi_m^K(v)|\ll\sqrt{m}.\sum_{\sqrt{m}\lambda\in K_I,Q(\lambda)=1}\frac{1}{Q(\lambda_{v^\perp})^{(b-1)/2}}\ll\sqrt{m}\sum_{N\ge1}\sum_{\substack{Q(\lambda_v^\perp)\in[N,N+1[\\ \sqrt{m}\lambda\in K_I\\ Q(\lambda)=1}}\frac{1}{N^{(b-1)/2}}.$$

By Proposition 4.14 below, we have

$$|\{\lambda\in K_{I,\mathbb{R}},\,Q(\lambda)=1,\,\sqrt{m}\lambda\in K,\,Q(\lambda_v^\perp)\in[N,N+1[\}|\ll_\epsilon m^{b/2-1+\epsilon}N^{b/2-2}.$$

Hence

$$|\Phi_m^K(v)|\ll_\epsilon m^{(b-1)/2+\epsilon}\sum_{N\ge1}\frac{N^{b/2-2}}{N^{(b-1)/2}}\ll_\epsilon m^{(b-1)/2+\epsilon}\sum_{N\ge1}\frac{1}{N^{3/2}}\ll_\epsilon m^{(b-1)/2+\epsilon},$$

which proves the second part of Proposition 4.13.

Proposition 4.14. *Let $m\ge1$ be an integer and $X>0$ a positive real number. Then*

$$|\{\lambda\in K_{I,\mathbb{R}},\,Q(\lambda)=1,\,\sqrt{m}\lambda\in K_I,\,Q(\lambda_v^\perp)\in[N,N+1[\}|\ll_\epsilon m^{b/2-1+\epsilon}N^{b/2-2}.$$

Proof. Recall that (K_I,Q) is a quadratic lattice of signature $(b-1,1)$ and we have a canonical measure μ_∞ on the quadric $K_1:=\{x\in K_{I,\mathbb{R}}\mid Q(x)=1\}$ defined as follows: for W an open subset of $K_\mathbb{R}$, let

$$\mu_\infty(W\cap K_1)=\lim_{\epsilon\to0}\frac{\operatorname{Leb}(\{x\in W,\,|Q(x)-1|<\epsilon\})}{2\epsilon}.$$

Here Leb is the Lebesgue measure on $K_\mathbb{R}$ for which the lattice K is of covolume 1. One can then prove that (see for example the proof of [Shankar et al. 2022, Corollary 4.12]):

$$\mu_\infty(\{\lambda\in K_1,\,Q(\lambda_{v^\perp})\in[X,X+1[\})\ll X^{b/2-2}.$$

On the other hand, by the equidistribution of integral points in quadrics, see [Eskin and Oh 2006; Duke 1988],[6] we have

$$|\{\lambda\in K_1,\,\sqrt{m}\lambda\in K,\,Q(\lambda_v^\perp)\in[N,N+1[\}|\ll_\epsilon m^{b/2-1+\epsilon}\mu_\infty(\{\lambda\in K_1,\,Q(\lambda_{v^\perp})\in[X,X+1[\}),$$

which yields the desired result. □

[6] Or the circle method.

5. Bounding the contribution from bad reduction places

In this section we prove Propositions 4.7 and 4.12. Let $\mathcal{M}^{\Sigma}$ be as before the toroidal compactification of the GSpin Shimura variety associated to a quadratic lattice (L, Q) and a K-admissible polyhedral cone decomposition Σ. The lattice (L, Q) is assumed to be maximal in the number field case and moreover self-dual at p in the function field case.

5A. ***Bad reduction in the number field setting.*** In this section, we prove Proposition 4.7. We assume hence that the lattice (L, Q) is maximal and that the polyhedral cone decomposition Σ is chosen in such way that Proposition 4.2 is satisfied.

By the choice of the cone decomposition Σ, the intersection points of $\mathcal{Y}$ and $\mathcal{M}^{\Sigma}$ lie either in a boundary divisor of type II or a boundary divisor of type III associated to a toroidal stratum representative (Ξ, σ) of type III where σ is a ray.

Let $\mathfrak{P}$ be a prime of bad reduction, i.e., where $\mathcal{Y}$ intersects the boundary of $\mathcal{M}^{\Sigma}$. Let $K_{\mathfrak{P}}$ be the completion at $\mathfrak{P}$ of the number field K and $v_{\mathfrak{P}}$ its normalized valuation. Let $k_{\mathfrak{P}}$ be the residue field of $\mathfrak{P}$ and $\bar{k}_{\mathfrak{P}}$ an algebraic closure.

5A1. *Type-II degeneration.* Assume in this section that the boundary point lies in $\mathcal{B}^{\Upsilon}_{\mathbb{F}_p}$ where Υ is a cusp label representative of type II.

Let J be the primitive isotropic plane associated to Υ and let $D = J_L^{\perp}/J$; see Section 4E for notation.

Recall from (2B.5) and (2C.3) that the completion of $\mathcal{M}^{\Sigma}$ along the boundary divisor $\mathcal{B}^{\Upsilon}$ fits into the following commutative diagram:

$$\begin{tikzcd}
\widehat{\mathcal{M}}_{\Upsilon} \arrow[r, "\pi"] \arrow[d, "="] & \widehat{\mathcal{M}^{\Sigma}} & \\
\widehat{\mathcal{M}}_{\Upsilon} \arrow[r, "\mathrm{Spf}(\mathbb{Z}_p[\![X]\!])"] & \overline{\mathcal{M}}_{\Upsilon} \arrow[r, "D\otimes\mathcal{E}"] & \mathcal{M}^{h}_{\Upsilon}
\end{tikzcd}$$

where the map π is an étale map of formal Deligne–Mumford stacks.

The formal completion of $\mathcal{Y}$ along $\mathfrak{P}$ induces a map

$$\mathrm{Spf}(\mathcal{O}_{K_{\mathfrak{P}}}) \to \widehat{\mathcal{M}^{\Sigma}},$$

which lifts by étaleness of π to a map

$$\mathrm{Spf}(\mathcal{O}_{K_{\mathfrak{P}}}) \to \widehat{\mathcal{M}}_{\Upsilon}.$$

Denoting by x the image of the closed point $s_{\mathfrak{P}}$, then we get a map of local rings

$$\Psi : \widehat{\mathcal{O}}_{\widehat{\mathcal{M}}_{\Upsilon},x} \to \mathcal{O}_{K_{\mathfrak{P}}}.$$

Let $m \geq 1$ be an integer coprime to N. By Proposition 2.4, the formal completion of the divisor $\mathcal{Z}(m)$ is described as the union over $\lambda \in D$ with $Q(\lambda) = m$, of the vanishing set of the ideals $\mathbb{Z}_p[X]] \otimes \hat{I}_{\lambda}$. If f_{λ}

is a generator of $\hat{I}_\lambda$,[7] then the multiplicity of intersection of the branch parametrized by λ at $\mathfrak{P}$ is equal to

$$v(\lambda) = v_{\mathfrak{P}}(\Psi(f_\lambda)).$$

Hence the multiplicity of intersection of $\mathcal{Y}$ and $\mathcal{Z}(m)$ at $\mathfrak{P}$ is given by

$$(\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = \frac{1}{d} \sum_{\lambda \in D,\, Q(\lambda)=m} v(\lambda),$$

where d is the degree of π at $\rho(s_{\mathfrak{P}})$.

For an integer n, define the set

$$L_n = \{\lambda \in D \mid v(\lambda) \geq n\}, \tag{5A.1}$$

and notice that (L_n) is a decreasing chain of sets. It follows then

$$(\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} \leq \sum_{\lambda \in D,\, Q(\lambda)=m} v(\lambda) \leq \sum_{n \geq 1} |\{\lambda \in L_n \mid Q(\lambda) = m\}|. \tag{5A.2}$$

The proposition below should be compared to what happens in the good reduction case in [Shankar et al. 2022, Section 7]. For a definition of the successive minima used, we refer to [Eskin and Katznelson 1995, Definition 2.2].

Proposition 5.1. *The sequence $(L_n, Q)_n$ is a decreasing sequence of positive definite lattices which all have the same rank $r \leq b-2$. Moreover, the following holds*:

(1) $\bigcap_n L_n = \{0\}$.

(2) *For every $n \geq 1$, $pL_n \subseteq L_{n+1}$.*

(3) *For $1 \leq r \leq b-2$, let $\mu_i(L_n)$ be the i-th successive minima of L_n and let $a_i(L_n) = \prod_{1 \leq k \leq i} \mu_i(L_n)$. Then we have*

$$a_i(L_n) \gg_\epsilon n^{i/(b+\epsilon)}.$$

Proof. Let $\lambda, \lambda' \in L_n$. From (2D.1), we see that $\ker(p_\lambda) \cap \ker(p_{\lambda'})$ and thus

$$\hat{I}_{\lambda+\lambda} \subset \widehat{I_\lambda} + \widehat{I_{\lambda'}}.$$

It follows that

$$v(\lambda + \lambda') \geq \min\{v(\lambda), v(\lambda')\} \geq n.$$

We conclude that $L_n \subseteq D$ is a subgroup and (L_n, Q) is obviously positive definite. Moreover, since the curve $\mathcal{Y}$ is not contained in any special divisor, (1) follows immediately.

For (2), let $\lambda \in L_n$ with $v(\lambda) \geq n \geq 1$. Then $\hat{I}_{p\lambda}$ is the ideal defining the kernel of the composition

$$D \otimes \widehat{\mathcal{E}} \to \widehat{\mathcal{E}} \to \widehat{\mathcal{E}},$$

over $\mathrm{Spf}(\widehat{\mathcal{O}}_{\mathcal{M}^h,z})$.

[7] Recall that $\mathcal{Z}(m)$ is Cartier.

Since the multiplication by p map is ramified at 0 with ramification degree equal to p, we conclude that

$$v(p\lambda) \geq pv(\lambda) \geq n+1.$$

This also proves that the lattices L_n have the same rank.

For (3), let $n \geq 1$ and let w_0 be a vector in L_n such that $Q(w_0) = \mu_1(L_n)^2$. By choosing $m_0 = \mu_1(L_n)^2$, the height bound Corollary 4.4 implies

$$n \leq (\mathcal{Y}.\mathcal{Z}(m_0))_{\mathfrak{P}} \ll_\epsilon m^{(b+\epsilon)/2}.$$

Hence $\mu_1(L_n) \gg_\epsilon n^{1/(b+\epsilon)}$. Since $a_i(n) \geq \mu_1(n)^r$, this concludes the proof. □

Proposition 5.2. *Let $D \in \mathbb{Z}_{\geq 1}$. For $X \in \mathbb{Z}_{>0}$, let $S_{D,X}$ denote the set*

$$\left\{m \in \mathbb{Z}_{>0} \mid X \leq m < 2X, \frac{m}{D} \in \mathbb{Z} \cap (\mathbb{Q}^\times)^2, (m, N) = 1\right\}.$$

Then we have

$$\sum_{m \in S_{D,X}} (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = o(X^{(b+1)/2} \log X).$$

Proof. We have

$$\sum_{m \in S_{D,X}} (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} \leq \sum_{m \in S_{D,X}} \sum_{n \geq 1} |\{\lambda \in L_n \mid Q(\lambda) = m\}| = \sum_{n \geq 1} \sum_{m \in S_{D,X}} |\{\lambda \in L_n \mid Q(\lambda) = m\}|.$$

By [Eskin and Katznelson 1995, Lemma 2.4], we have the following estimate which only depends on the rank r of the lattices L_n and hence not on n

$$\sum_{m \in S_{D,X}} |\{\lambda \in L_n \mid Q(\lambda) = m\}| \ll \sum_{j=0}^{r} \frac{X^j}{a_j(L_n)}.$$

On the other hand, if $\lambda \in L_n$ with $Q(\lambda) = m \in S_{D,X}$, then $\mu_1(L_n)^2 \leq m \leq X$; hence $n \ll X^{(b+\epsilon)/2}$ and

$$\begin{aligned} \sum_{m \in S_{D,X}} (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} &\ll \sum_{m \in S_{D,X}} \sum_{n \geq 1}^{O_\epsilon(X^{(b+\epsilon)/2})} |\{\lambda \in L_n \mid Q(\lambda) = m\}| \\ &\ll \sum_{j=0}^{r} \sum_{n \geq 1}^{X^{(b+\epsilon)/2}} \frac{X^{j/2}}{n^{j/(b+\epsilon)}} \\ &\ll \sum_{j=0}^{r} X^{j/2 + (1 - j/(b+\epsilon))(b+\epsilon)/2} = O(X^{(b+\epsilon)/2}). \end{aligned}$$

Hence the result. □

5A2. *Type-III degeneration.* Let (Ξ, σ) be a toroidal stratum representative of type III such that σ is a ray. We use notation from Section 2B2.

By our choice of Σ, the curve $\mathcal{Y}$ touches the boundary of $\mathcal{M}^{\Sigma}$ at a locally closed boundary divisor $\mathcal{B}^{\Xi,\sigma}$. Let $\widehat{M}^{\Sigma}$ be the formal completion of $\mathcal{M}^{\Sigma}$ along $\mathcal{B}^{\Xi,\sigma}$ and hence we get a map

$$\widehat{\mathcal{Y}} \to \widehat{M}^{\Sigma}. \tag{5A.3}$$

By Section 2B5, the following maps of formal Deligne–Mumford stacks are finite étale:

$$\bigsqcup_{\mathbb{Q}^{\times}_{>0}\backslash\mathbb{A}^{\times}_{f}/K_0} \widehat{\mathcal{T}}_{\Xi/R} \to \widehat{\mathcal{M}}_{\Xi,\sigma} \to \widehat{\mathcal{M}^{\Sigma}}.$$

Hence map (5A.3) lifts to map

$$\widehat{\mathcal{Y}} \to \operatorname{Spf}(\mathbb{Z}_p[q_\alpha \mid \alpha \in \Gamma_{\Xi}^{\vee} \cap \omega^{\perp}][\![q_{\omega'}]\!]).$$

This corresponds to a morphism

$$\mathbb{Z}_{(p)}[\![q_{\omega'}]\!][q_\alpha]_{\alpha \in \Gamma_{\Xi}^{\vee}, (\alpha.\omega)=0} \to \mathcal{O}_{K_{\mathfrak{P}}}. \tag{5A.4}$$

Let $\lambda \in K_I = \Gamma_{\Xi}$ with $Q(\lambda) = m$. By Section 2D2 the branch of the special divisor $\mathcal{Z}(m)$ parametrized by λ intersects the boundary only if $(\lambda.\omega) = 0$. In the latter case, by Proposition 2.5, its equation is given by $q^{\lambda} - 1$ and the multiplicity of intersection of $\mathcal{Y}$ with the branch given by λ is the p-adic valuation of the element $q^{\lambda} - 1$ under the map (5A.4).

Let $x \in \mathcal{B}^{\Xi,\sigma}(\overline{\mathbb{F}}_p)$ be the image of $\mathfrak{P}$. Then by the previous discussion, we conclude that

$$(\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = \frac{1}{d} \sum_{\lambda \in K_I \cap \omega^{\perp}, Q(\lambda)=m} v_p(q^{\lambda} - 1),$$

where d is the degree of the map (5A.3) at x.

For $n \geq 1$, let

$$L_n = \{\lambda \in K_I \cap \omega^{\perp} \mid v_p(q^{\lambda} - 1) \geq n\}.$$

Then we can rewrite the multiplicity intersection at $\mathfrak{P}$ as

$$(\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = \frac{1}{d} \sum_{n \geq 1} \{\lambda \in L_n \mid Q(\lambda) = m\}.$$

Proposition 5.3. *The lattices (L_n, Q) are positive definite lattices of rank $r \leq b - 1$ independent from n and they satisfy the following properties*:

(1) $\bigcap_n L_n = \{0\}$.

(2) *For every $n \geq 1$, $pL_n \subseteq L_{n+1}$.*

(3) *For $1 \leq r \leq b - 1$, let $\mu_i(L_n)$ be the i-th successive minima and let $a_i(L_n) = \prod_{1 \leq k \leq i} \mu_i(L_n)$. Then we have*

$$a_i(L_n) \gg_{\epsilon} n^{i/(b+\epsilon)}.$$

Proof. The proof is similar to the proof of Proposition 5.1. Let $\lambda, \lambda' \in K \cap \omega^{\perp}$. By writing

$$q^{\lambda+\lambda'} - 1 = q^{\lambda}(q^{\lambda'} - 1) + q^{\lambda} - 1,$$

we get that L_n is a lattice and it is obviously positive definite as K_I is Lorentzian and ω is a negative normed vector.

Let π be a uniformizer of $\mathcal{O}_{K_{\mathfrak{P}}}$ and let $\lambda \in L_n$. Then $q^{\lambda} = 1 + \pi^n.u$ for some $u \in \mathcal{O}_{K_{\mathfrak{P}}}$. Hence

$$q^{p\lambda} - 1 = (1 + \pi^n.u)^p - 1 = \sum_{i \geq 1} \binom{p}{i} \pi^{ni} u^i = \pi^{n+1} u'.$$

Hence (2). The rest of the proof is similar to Proposition 5.1. □

As a consequence, we get the following proposition, whose proof is identical to that of Proposition 5.2 and we omit it.

Proposition 5.4. *Let $D \in \mathbb{Z}_{\geq 1}$ be coprime to N. For $X \in \mathbb{Z}_{>0}$, let $S_{D,X}$ denote the set*

$$\left\{ m \in \mathbb{Z}_{>0} \mid X \leq m < 2X,\, \frac{m}{D} \in \mathbb{Z} \cap (\mathbb{Q}^{\times})^2,\, (m, N) = 1 \right\}.$$

Then we have

$$\sum_{m \in S_{D,X}} (\mathcal{Y}.\mathcal{Z}(m))_{\mathfrak{P}} = o(X^{(b+1)/2} \log X).$$

5B. ***Function field setting.*** In this section, we prove Proposition 4.11. We assume here that the lattice (L, Q) is self-dual at p and we let $\mathcal{M}_{\mathbb{F}_p}$ be the mod p GSpin Shimura variety associated to (L, Q). Let Σ be a polyhedral cone decomposition which satisfies Proposition 4.9.

Let $\overline{\mathscr{S}} \to \mathcal{M}^{\Sigma}_{\mathbb{F}_p}$ be a finite map as before and let $P \in \overline{\mathscr{S}}(\overline{\mathbb{F}}_p)$ be a point mapping to the boundary of $\mathcal{M}^{\Sigma}_{\mathbb{F}_p}$. Let denote $k = \overline{\mathbb{F}}_p$. The point P lies either in a boundary stratum of type II or type III. We treat each case separately.

5B1. *Type-II degeneration.* Assume that the image of P is in $\mathcal{B}^{\Upsilon}_{\mathbb{F}_p}(k)$ where Υ is a cusp label representative of type II.

Let $\widehat{\overline{\mathscr{S}}} \simeq \mathrm{Spf}(k[\![t]\!])$ be the formal completion of $\overline{\mathscr{S}}$ along s. Then by reasoning similarly to Section 5A1, specifically using the reduction mod p of (2C.3), we get for every $\lambda \in D$ with $Q(\lambda) = m \geq 1$, m coprime to N a map

$$\Phi_p : \widehat{\mathcal{O}}_{\mathcal{M}_{\Upsilon,\mathbb{F}_p,x}} \to k[\![t]\!],$$

Let $v(\lambda)$ denote the t-adic valuation of the generator f_{λ} of $I_{\lambda,p}$. Then similarly to the number field case, we have:

Lemma 5.5. *The multiplicity of intersection of $\overline{\mathscr{S}}$ and $\mathcal{Z}(m)_{\mathbb{F}_p}$ at P satisfies*

$$m_P(\overline{\mathscr{S}}, \mathcal{Z}(m)_{\mathbb{F}_p}) \ll \sum_{n \geq 1} |\{\lambda \in L_n \mid Q(\lambda) = m\}|.$$

Now we are ready to prove Proposition 4.12.

Proposition 5.6. *The sequence of lattices* (L_n, Q) *satisfy the same properties as in Proposition 5.1 and letting* S *be as in Section 4B*, *we have the following estimate for* $X > 0$:

$$\sum_{m \in S_X} m_P(\overline{\mathscr{S}}, \mathcal{Z}(m)_{\mathbb{F}_p}) = O_\epsilon(X^{(b+\epsilon)/2})$$

Proof. The same proof as in Proposition 5.1 shows that the lattices (L_n, Q) enjoy the same properties of the aforementioned proposition. For the second part, we have

$$\begin{aligned}\sum_{m \in S_X} m_P(\overline{\mathscr{S}}, \mathcal{Z}(m)_{\mathbb{F}_p}) &\ll \sum_{m \in S_X} \sum_{n \geq 1} |\{\lambda \in L_n \mid Q(\lambda) = m\}| \\ &\ll \sum_{n=1}^{O(X^{(b+\epsilon)/2})} |\{\lambda \in L_n \mid Q(\lambda) \leq m\}| \\ &\ll \sum_{n=1}^{O(X^{(b+\epsilon)/2})} \sum_{j=0}^{r} \frac{X^{j/2}}{a_j(L_n)} \\ &\ll \sum_{j=0}^{r} \sum_{n=1}^{O(X^{(b+\epsilon)/2})} \frac{X^{j/2}}{n^{j/(b+\epsilon)}} = O(X^{(b+\epsilon)/2}). \qquad \square\end{aligned}$$

5B2. *Type-III degeneration.* Assume now that there exists a toroidal stratum representative (Ξ, σ) such that σ is a ray and such that P lies in $\mathcal{B}^{\Xi,\sigma}_{\mathbb{F}_p}(k)$. Using a similar approach to Section 5A2 by taking reduction mod p, we get a map

$$k[q_\alpha \mid \alpha \in \Gamma_\Xi^\vee \cap \omega^\perp][\![q_{\omega'}]\!] \to k[\![t]\!],$$

sending $q_{\omega'}$ to an element of the ideal (t). Let v denote the t-adic valuation on $k[\![t]\!]$. Then, for m coprime to N, the multiplicity of intersection of $\overline{\mathscr{S}}$ and $\mathcal{Z}(m)_{\mathbb{F}_p}$ at P satisfies

$$m_P(\overline{\mathscr{S}}, \mathcal{Z}(m)_{\mathbb{F}_p}) \leq \sum_{\lambda \in K_I \cap \omega^\perp, Q(\lambda) = m} v(q^\lambda - 1).$$

If we define the sequence lattices L_n as

$$L_n = \{\lambda \in K \cap \omega^\perp \mid v(q^\lambda - 1) \geq n\},$$

then

$$m_P(\overline{\mathscr{S}}, \mathcal{Z}(m)_{\mathbb{F}_p}) \leq \sum_{n \geq 1} |\{\lambda \in L_n, Q(\lambda) = m\}|.$$

Now the rest of the proof is similar to Section 5A2. This proves Proposition 4.12 in the remaining type-III case.

6. Applications

In this section, we present a proof of Theorem 1.5. This approach is inspired from [Maulik et al. 2022a].

6A. ***Hecke orbit conjecture.***

6A1. *The orthogonal case.* Since GSpin Shimura varieties are finite covers of orthogonal ones, it is enough to prove the result for GSpin Shimura varieties.

Let $\mathcal{M}_{\mathbb{F}_p}$ be the reduction mod $p \geq 5$ of a GSpin-type Shimura variety with hyperspecial level at p associated to a lattice (L, Q), which is assumed to be self-dual at p and of signature $(b, 2)$. We will prove Theorem 1.5 by induction on b, which is also the dimension of $\mathcal{M}_{\mathbb{F}_p}$.

The case $b = 1$ is immediate: the prime-to-p Hecke orbit of x is infinite, hence Zariski dense.

Assume now that $n \geq 2$ and the result of Theorem 1.5 holds for all ordinary points in GSpin Shimura varieties of dimension less than $b - 1$ with hyperspecial level at p. Let x be an ordinary point in $\mathcal{M}(\overline{\mathbb{F}}_p)$ and let $\overline{T_x}$ be the Zariski closure of its prime-to-p Hecke orbit. Then $\overline{T_x}$ has positive dimension and intersects the ordinary locus nontrivially. Hence we can find a smooth quasiprojective curve $\mathscr{S}$ and a finite map

$$\mathscr{S} \to \mathcal{M}_{\mathbb{F}_p}$$

whose image is contained in $\overline{T_x}$ and which is contained in the ordinary locus. Moreover, we can assume that this image is not contained in any special divisor. Indeed, the same argument used for proper curves in [Maulik et al. 2022a, Lemma 8.11] works in our setting with no change. By Theorem 1.3, the curve $\mathscr{S}$ intersects infinitely many divisors $\mathcal{Z}(m)_{\mathbb{F}_p}$ with $(m, p) = 1$. The special divisors $\mathcal{Z}(m)_{\mathbb{F}_p}$ are themselves the union of GSpin Shimura varieties of dimension $b - 1$ with hyperspecial level at p since m is coprime to p. Let $y \in \mathscr{S}(\overline{\mathbb{F}}_p) \cap \mathcal{Z}'(m)(\overline{\mathbb{F}}_p)$ for some irreducible component $\mathcal{Z}'(m)$ of $\mathcal{Z}(m)$. Then y is ordinary and the prime-to-p Hecke orbit of y in $\mathcal{Z}'(m)_{\mathbb{F}_p}$ is Zariski dense by the induction hypothesis. Since this orbit is a suborbit of the Hecke orbit in $\mathcal{M}_{\mathbb{F}_p}$, we conclude that $\mathcal{Z}'(m)_{\mathbb{F}_p} \subset \overline{T_x}$. Furthermore, it is straightforward to check that the collection of the divisors $\mathcal{Z}'(m)_{\mathbb{F}_p}$ must be infinite by Theorem 1.3. Hence we conclude that $\overline{T_x} = \mathcal{M}_{\mathbb{F}_p}$ which is the desired result.

6A2. *The unitary case.* We prove in this section the Hecke orbit conjecture in the unitary case using the reduction to the orthogonal case already used in [Maulik et al. 2022a, Remark 8.12] and in [Shankar et al. 2022, Section 9.3].

Let $\mathcal{M}_{\mathbb{F}_p}$ be the mod p points of the canonical model of a unitary Shimura variety associated to an imaginary quadratic field $\boldsymbol{k}$, a unitary group of signature $(r, 1)$ with hyperspecial level at p as described in [Bruinier et al. 2020, Section 2.1] such that p is split in $\boldsymbol{k}$. Consider the family of special divisors $\mathcal{Z}_{Kra}(m)$ as described in [loc. cit., Section 2.5] which are themselves unitary Shimura varieties associated to unitary groups of signature $(r - 1, 1)$ and hyperspecial at p when p does not divide m. Then using a similar argument to [Shankar et al. 2022, Section 9.3] and further explained in [Maulik et al. 2022a, Remark 8.12], we have the following theorem which is a consequence of Theorem 1.3.

Theorem 6.1. *Assume that $p \geq 5$ and let $\mathscr{S} \to \mathcal{M}_{\mathbb{F}_p}$ be a finite map from a smooth quasiprojective curve $\mathscr{S}$ over $\overline{\mathbb{F}}_p$ and with generically ordinary image. Then the union over m prime to p of the intersections $\mathscr{S} \cap \mathcal{Z}_{Kra}(m)$ is infinite.*

Now the Hecke orbit conjecture in the unitary case is an easy consequence of the above theorem and the induction method explained in the previous paragraph.

Acknowledgments

I am very grateful to François Charles, François Greer, Keerthi Madapusi Pera, Ananth Shankar and Yunqing Tang for many helpful discussions and insights. This project started while I was invited by Yunqing Tang to Princeton University, and I also benefited from support from the Institute for Advanced Study in Princeton. I thank both institutions for their hospitality. I thank the referee for the very valuable comments and suggestions.

References

[Andreatta et al. 2017] F. Andreatta, E. Z. Goren, B. Howard, and K. Madapusi Pera, "Height pairings on orthogonal Shimura varieties", *Compos. Math.* **153**:3 (2017), 474–534. MR Zbl

[Andreatta et al. 2018] F. Andreatta, E. Z. Goren, B. Howard, and K. Madapusi Pera, "Faltings heights of abelian varieties with complex multiplication", *Ann. of Math.* (2) **187**:2 (2018), 391–531. MR Zbl

[Ash et al. 1975] A. Ash, D. Mumford, M. Rapoport, and Y. Tai, *Smooth compactification of locally symmetric varieties*, Lie Groups History Frontiers Appl. **4**, Math. Sci. Press, Brookline, MA, 1975. MR Zbl

[Baldi et al. 2024] G. Baldi, B. Klingler, and E. Ullmo, "On the distribution of the Hodge locus", *Invent. Math.* **235**:2 (2024), 441–487. MR Zbl

[Borcherds 1999] R. E. Borcherds, "The Gross–Kohnen–Zagier theorem in higher dimensions", *Duke Math. J.* **97**:2 (1999), 219–233. MR Zbl

[Bruinier 2002] J. H. Bruinier, *Borcherds products on* $\mathrm{O}(2, l)$ *and Chern classes of Heegner divisors*, Lecture Notes in Math. **1780**, Springer, 2002. MR Zbl

[Bruinier and Zemel 2022] J. H. Bruinier and S. Zemel, "Special cycles on toroidal compactifications of orthogonal Shimura varieties", *Math. Ann.* **384**:1-2 (2022), 309–371. MR Zbl

[Bruinier et al. 2007] J. H. Bruinier, J. I. Burgos Gil, and U. Kühn, "Borcherds products and arithmetic intersection theory on Hilbert modular surfaces", *Duke Math. J.* **139**:1 (2007), 1–88. MR Zbl

[Bruinier et al. 2020] J. H. Bruinier, B. Howard, S. S. Kudla, M. Rapoport, and T. Yang, "Modularity of generating series of divisors on unitary Shimura varieties", pp. 7–125 in *Arithmetic divisors on orthogonal and unitary Shimura varieties*, Astérisque **421**, Soc. Math. France, Paris, 2020. MR Zbl

[Chai 1995] C.-L. Chai, "Every ordinary symplectic isogeny class in positive characteristic is dense in the moduli", *Invent. Math.* **121**:3 (1995), 439–479. MR Zbl

[Chai and Oort 2006] C.-L. Chai and F. Oort, "Hypersymmetric abelian varieties", *Pure Appl. Math. Q.* **2**:1 (2006), 1–27. MR Zbl

[Charles 2014] F. Charles, "On the Picard number of K3 surfaces over number fields", *Algebra Number Theory* **8**:1 (2014), 1–17. MR Zbl

[Charles 2018] F. Charles, "Exceptional isogenies between reductions of pairs of elliptic curves", *Duke Math. J.* **167**:11 (2018), 2039–2072. MR Zbl

[Clozel and Ullmo 2005] L. Clozel and E. Ullmo, "Équidistribution de sous-variétés spéciales", *Ann. of Math.* (2) **161**:3 (2005), 1571–1588. MR Zbl

[Duke 1988] W. Duke, "Hyperbolic distribution problems and half-integral weight Maass forms", *Invent. Math.* **92**:1 (1988), 73–90. MR Zbl

[Engel et al. 2023] P. Engel, F. Greer, and S. Tayou, "Mixed mock modularity of special divisors", preprint, 2023. arXiv 2301.05982

[Eskin and Katznelson 1995] A. Eskin and Y. R. Katznelson, “Singular symmetric matrices”, *Duke Math. J.* **79**:2 (1995), 515–547. MR Zbl

[Eskin and Oh 2006] A. Eskin and H. Oh, “Representations of integers by an invariant polynomial and unipotent flows”, *Duke Math. J.* **135**:3 (2006), 481–506. MR Zbl

[Harris and Zucker 2001] M. Harris and S. Zucker, *Boundary cohomology of Shimura varieties, III: Coherent cohomology on higher-rank boundary strata and applications to Hodge theory*, Mém. Soc. Math. France (N.S.) **85**, Soc. Math. France, Paris, 2001. MR Zbl

[van Hoften 2024] P. van Hoften, “On the ordinary Hecke orbit conjecture”, *Algebra Number Theory* **18**:5 (2024), 847–898. MR Zbl

[Howard and Madapusi Pera 2020] B. Howard and K. Madapusi Pera, “Arithmetic of Borcherds products”, pp. 187–297 in *Arithmetic divisors on orthogonal and unitary Shimura varieties*, Astérisque **421**, Soc. Math. France, Paris, 2020. MR Zbl

[Imamoḡlu et al. 2014] Ö. Imamoḡlu, M. Raum, and O. K. Richter, “Holomorphic projections and Ramanujan’s mock theta functions”, *Proc. Natl. Acad. Sci. USA* **111**:11 (2014), 3961–3967. MR Zbl

[Kisin 2010] M. Kisin, “Integral models for Shimura varieties of abelian type”, *J. Amer. Math. Soc.* **23**:4 (2010), 967–1012. MR Zbl

[Madapusi Pera 2015] K. Madapusi Pera, “The Tate conjecture for K3 surfaces in odd characteristic”, *Invent. Math.* **201**:2 (2015), 625–668. MR Zbl

[Madapusi Pera 2016] K. Madapusi Pera, “Integral canonical models for spin Shimura varieties”, *Compos. Math.* **152**:4 (2016), 769–824. MR Zbl

[Madapusi Pera 2019] K. Madapusi Pera, “Toroidal compactifications of integral models of Shimura varieties of Hodge type”, *Ann. Sci. École Norm. Sup.* (4) **52**:2 (2019), 393–514. MR Zbl

[Maulik et al. 2022a] D. Maulik, A. N. Shankar, and Y. Tang, “Picard ranks of K3 surfaces over function fields and the Hecke orbit conjecture”, *Invent. Math.* **228**:3 (2022), 1075–1143. MR Zbl

[Maulik et al. 2022b] D. Maulik, A. N. Shankar, and Y. Tang, “Reductions of abelian surfaces over global function fields”, *Compos. Math.* **158**:4 (2022), 893–950. MR Zbl

[Oguiso 2003] K. Oguiso, “Local families of K3 surfaces and applications”, *J. Algebraic Geom.* **12**:3 (2003), 405–433. MR Zbl

[Pink 1989] R. Pink, *Arithmetical compactification of mixed Shimura varieties*, Ph.D. thesis, Universität Bonn, 1989, available at https://people.math.ethz.ch/∼pink/ftp/phd/PinkDissertation.pdf.

[Sarnak 1990] P. Sarnak, *Some applications of modular forms*, Cambridge Tracts in Math. **99**, Cambridge Univ. Press, 1990. MR Zbl

[Shankar and Tang 2020] A. N. Shankar and Y. Tang, “Exceptional splitting of reductions of abelian surfaces”, *Duke Math. J.* **169**:3 (2020), 397–434. MR Zbl

[Shankar et al. 2022] A. N. Shankar, A. Shankar, Y. Tang, and S. Tayou, “Exceptional jumps of Picard ranks of reductions of K3 surfaces over number fields”, *Forum Math. Pi* **10** (2022), art. id. e21. MR Zbl

[Tayou 2020] S. Tayou, “On the equidistribution of some Hodge loci”, *J. Reine Angew. Math.* **762** (2020), 167–194. MR Zbl

[Tayou and Tholozan 2023] S. Tayou and N. Tholozan, “Equidistribution of Hodge loci, II”, *Compos. Math.* **159**:1 (2023), 1–52. MR Zbl

[Voisin 2002] C. Voisin, *Théorie de Hodge et géométrie algébrique complexe*, Cours Spécialisés **10**, Soc. Math. France, Paris, 2002. MR Zbl

[Zemel 2020] S. Zemel, “The structure of integral parabolic subgroups of orthogonal groups”, *J. Algebra* **559** (2020), 95–128. MR Zbl

Communicated by Martin Olsson
Received 2022-06-27 Revised 2023-07-31 Accepted 2024-01-09

tayou@math.harvard.edu *Department of Mathematics, Harvard University, Cambridge, MA, United States*

ALGEBRA AND NUMBER THEORY 19:1 (2025)
https://doi.org/10.2140/ant.2025.19.113

Curves with few bad primes over cyclotomic $\mathbb{Z}_\ell$-extensions

Samir Siksek and Robin Visser

Let K be a number field, and S a finite set of nonarchimedean places of K, and write $\mathcal{O}^\times$ for the group of S-units of K. A famous theorem of Siegel asserts that the S-unit equation $\varepsilon + \delta = 1$, with $\varepsilon, \delta \in \mathcal{O}^\times$, has only finitely many solutions. A famous theorem of Shafarevich asserts that there are only finitely many isomorphism classes of elliptic curves over K with good reduction outside S. Now instead of a number field, let $K = \mathbb{Q}_{\infty,\ell}$ which denotes the $\mathbb{Z}_\ell$-cyclotomic extension of $\mathbb{Q}$. We show that the S-unit equation $\varepsilon + \delta = 1$, with $\varepsilon, \delta \in \mathcal{O}^\times$, has infinitely many solutions for $\ell \in \{2, 3, 5, 7\}$, where S consists only of the totally ramified prime above ℓ. Moreover, for every prime ℓ, we construct infinitely many elliptic or hyperelliptic curves defined over K with good reduction away from 2 and ℓ. For certain primes ℓ we show that the Jacobians of these curves in fact belong to infinitely many distinct isogeny classes.

1. Introduction

Let ℓ be a rational prime and r a positive integer. Write $\mathbb{Q}_{r,\ell}$ for the unique degree ℓ^r totally real subfield of $\bigcup_{n=1}^{\infty} \mathbb{Q}(\mu_n)$, where μ_n denotes the set of ℓ^n-th roots of 1. We let $\mathbb{Q}_{\infty,\ell} = \bigcup_r \mathbb{Q}_{r,\ell}$; this is the $\mathbb{Z}_\ell$-cyclotomic extension of $\mathbb{Q}$, and $\mathbb{Q}_{r,\ell}$ is called the r-th layer of $\mathbb{Q}_{\infty,\ell}$. Now let K be a number field, and write $K_{\infty,\ell} = K \cdot \mathbb{Q}_{\infty,\ell}$ and $K_{r,\ell} = K \cdot \mathbb{Q}_{r,\ell}$. To ease notation we shall sometimes write K_∞ for $K_{\infty,\ell}$. We write $\mathcal{O}_\infty$ (or $\mathcal{O}_{\infty,\ell}$) for the integers in K_∞ (i.e., the integral closure of $\mathbb{Z}$ in K_∞), and write $\mathcal{O}_r$ (or $\mathcal{O}_{r,\ell}$) for the integers of $K_{r,\ell}$. Clearly $\mathcal{O}_{\infty,\ell} = \bigcup_r \mathcal{O}_{r,\ell}$. The motivation for the present paper is a series of conjectures and theorems that suggest that the arithmetic of curves (respectively abelian varieties) over K_∞ is similar to the arithmetic of curves (respectively abelian varieties) over K. One of these is the following conjecture of Mazur [1972], which in essence says that the Mordell–Weil theorem continues to hold over K_∞.

Conjecture (Mazur). *Let A/K_∞ be an abelian variety. Then $A(K_\infty)$ is finitely generated.*

Another is a conjecture of Parshin and Zarhin [1989, page 91] which is the analogue of Faltings' theorem (Mordell conjecture) over K_∞.

Conjecture (Parshin and Zarhin). *Let X/K_∞ be a curve of genus ≥ 2. Then $X(K_\infty)$ is finite.*

Siksek is supported by the EPSRC grant *Moduli of Elliptic curves and Classical Diophantine Problems* (EP/S031537/1). Visser is supported by an EPSRC studentship (EP/V520226/1).

MSC2020: primary 11G10; secondary 11G05.

Keywords: Shafarevich conjecture, Abelian varieties, cyclic fields, cyclotomic fields, integral points.

A third is the following theorem of Zarhin [2010, Corollary 4.2], which asserts that the Tate homomorphism conjecture (also a theorem of Faltings [1983] over number fields) continues to hold over K_∞.

Theorem (Zarhin). *Let A, B be abelian varieties defined over $K_{\infty,\ell}$, and denote their respective ℓ-adic Tate modules by $T_\ell(A)$, $T_\ell(B)$. Then the natural embedding*

$$\operatorname{Hom}_{K_\infty}(A, B) \otimes \mathbb{Z}_\ell \hookrightarrow \operatorname{Hom}_{\operatorname{Gal}(\overline{K_\infty}/K_\infty)}(T_\ell(A), T_\ell(B))$$

is a bijection.

Mazur's conjecture is now known to hold for certain elliptic curves. For example, if E is an elliptic curve defined over $\mathbb{Q}$ then $E(\mathbb{Q}_\infty)$ is finitely generated thanks to theorems of Kato, Ribet and Rohrlich [Greenberg 2001, Theorem 1.5]. From this one can deduce [Greenberg 2001, Theorem 1.24] that $X(\mathbb{Q}_\infty)$ is finite for curves $X/\mathbb{Q}$ of genus ≥ 2 equipped with a nonconstant morphism to an elliptic curve $X \to E$ defined over $\mathbb{Q}$. We also note that the conjecture of Parshin and Zarhin follows easily from Mazur's conjecture and Faltings' theorem. Indeed, using the Abel–Jacobi map we can deduce from Mazur's conjecture that $X(K_\infty) = X(K_r)$ for suitably large r, and we know that $X(K_r)$ is finite by Faltings' theorem.

It is natural to wonder whether other standard conjectures and theorems concerning the arithmetic of curves and abelian varieties over number fields continue to hold over K_∞. The purpose of this paper is to give counterexamples to potential generalizations of certain theorems of Siegel and Shafarevich to K_∞. A theorem of Siegel (e.g., [Abramovich 2009, Theorem 0.2.8]) asserts that $(\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}_{K,S})$ is finite for any number field K and any finite set of primes S; modern proofs can be found in [Kim 2005; Lawrence and Venkatesh 2020; Poonen 2021]. We show that the corresponding statement over $\mathbb{Q}_{\infty,\ell}$ is false, at least for $\ell = 2, 3, 5, 7$. We denote by υ_ℓ the totally ramified prime of $\mathbb{Q}_{\infty,\ell}$ above ℓ (the precise meaning of primes in infinite extensions of $\mathbb{Q}$ is clarified in Section 2).

Theorem 1. *Let $\ell = 2, 3, 5$ or 7. Let*

$$S = \begin{cases} \{\upsilon_\ell\} & \text{if } \ell = 2, 5, 7, \\ \varnothing & \text{if } \ell = 3. \end{cases} \tag{1}$$

Let $\mathcal{O}_S$ denote the S-integers of $\mathbb{Q}_{\infty,\ell}$. Then $(\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}_S)$ is infinite.

Remarks. • There have been several recent papers showing that $\mathbb{P}^1 - \{0, 1, \infty\}$ and other punctured curves have no or few integral points over various infinite families of number fields e.g., [Freitas et al. 2020; 2021a; 2021b; 2022; Triantafillou 2021]. In particular, it is shown in [Freitas et al. 2020] that $(\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}_\infty) = \varnothing$ for $\ell \neq 3$. The obstruction given in [loc. cit.] for $\ell \neq 3$ is local in nature. In essence, Theorem 1 complements this result, showing that we can obtain infinitely many integral or S-integral points in the absence of the local obstruction. The proof of Theorem 1 is constructive.

• Theorem 1 strongly suggests that the conjecture of Parshin and Zarhin does not admit a straightforward generalization to the broader context of integral points on hyperbolic curves. We also remark that there is a critical difference over K_∞ between complete curves X of genus ≥ 2 and $\mathbb{P}^1 - \{0, 1, \infty\}$. For the former, the group of K_∞-points of the Jacobian is expected to be finitely generated by Mazur's conjecture. For the latter, the analogue of the Jacobian is the generalized Jacobian which is $\mathbb{G}_m \times \mathbb{G}_m$, and its group of K_∞-points is $(\mathbb{G}_m \times \mathbb{G}_m)(K_\infty) = \mathcal{O}_\infty^\times \times \mathcal{O}_\infty^\times$, which is infinitely generated.

Variants of the proof of Theorem 1 give the following.

Theorem 2. *Let $\ell = 2, 3$ or 5. Let $S = \{\upsilon_\ell\}$ and write $\mathcal{O}_S$ for the S-integers of $\mathbb{Q}_{\infty,\ell}$. Let*

$$k \in \begin{cases} \{1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 24\} & \text{if } \ell = 2, 3, \\ \{1, 2, 4\} & \text{if } \ell = 5. \end{cases}$$

Then $(\mathbb{P}^1 - \{0, k, \infty\})(\mathcal{O}_S)$ is infinite.

Let ζ_{ℓ^n} denote a primitive ℓ^n-th root of 1, and write $\Omega_{n,\ell} = \mathbb{Q}(\zeta_{\ell^n})$, and $\Omega_{n,\ell}^+ = \mathbb{Q}(\zeta_{\ell^n} + \zeta_{\ell^n}^{-1})$. Let

$$\Omega_{\infty,\ell} = \bigcup_{n=1}^{\infty} \Omega_{n,\ell}, \quad \Omega_{\infty,\ell}^+ = \bigcup_{n=1}^{\infty} \Omega_{n,\ell}^+.$$

We note the inclusions $\Omega_{\infty,\ell} \supset \Omega_{\infty,\ell}^+ \supset \mathbb{Q}_{\infty,\ell}$. Nagell [1969, page 181] points out that $1 + \zeta_{\ell^n}$ is a unit for ℓ odd, and that therefore the equation $\varepsilon + \delta = 1$ has the solution $\varepsilon = -\zeta_{\ell^n}$, $\delta = 1 + \zeta_{\ell^n}$ in units belonging to $\Omega_{n,\ell}$. It follows straightforwardly from this (see the beginning of Section 3) that $\mathbb{P}^1 - \{0, 1, \infty\}$ has infinitely many integral points defined over $\Omega_{\infty,\ell}$. Many of our constructions of S-integral points on $\mathbb{P}^1 - \{0, 1, \infty\}$ apply in greater generality to the fields $\Omega_{\infty,\ell}$ and $\Omega_{\infty,\ell}^+$, where the statements are in fact much cleaner. For example, we prove the following theorem.

Theorem 3. *Let ℓ be an odd prime. Then $(\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}(\Omega_{\infty,\ell}^+))$ is infinite.*

Here $\mathcal{O}(\Omega_{\infty,\ell}^+)$ denotes the integers of $\Omega_{\infty,\ell}^+$.

Shafarevich's conjecture asserts that for a number field K, a dimension n, a degree d, and a finite set of places S, there are only finitely many isomorphism classes of polarized abelian varieties defined over K of dimension n with degree-d polarization and with good reduction away from S. This conjecture was proved by Shafarevich for elliptic curves (i.e., $n = 1$) and by Faltings [1983] in complete generality. If we replace K by $\mathbb{Q}_{\infty,\ell}$ then the Shafarevich conjecture no longer holds. For example, consider

$$E_\varepsilon : \varepsilon Y^2 = X^3 - X,$$

where $\varepsilon \in \mathcal{O}_\infty^\times$. This elliptic curve has good reduction away from the primes above 2. Moreover, E_ε, E_δ are isomorphic over $\mathbb{Q}_\infty$ if and only if ε/δ is a square in $\mathcal{O}_\infty^\times$. As $\mathcal{O}_\infty^\times/(\mathcal{O}_\infty^\times)^2$ is infinite, we deduce that there are infinitely many isomorphism classes of elliptic curves over $\mathbb{Q}_\infty$ with good reduction away from the primes above 2. It is however natural to wonder if a sufficiently weakened version of the Shafarevich conjecture continues to hold over $\mathbb{Q}_\infty$. Indeed, the curves E_ε in the above construction form a single $\overline{\mathbb{Q}}$-isomorphism class. This it is natural to ask if, for suitable ℓ and finite set of primes S, does the set of elliptic curves over $\mathbb{Q}_\infty$ with good reduction outside S form infinitely many $\overline{\mathbb{Q}}$-isomorphism classes?

Theorem 4. *Let $\ell = 2, 3, 5,$ or 7. Let S be given by* (1) *and let $S' = S \cup \{\upsilon_2\}$ where υ_2 is the unique prime of $\mathbb{Q}_{\infty,\ell}$ above 2. Then, there are infinitely many $\overline{\mathbb{Q}}$-isomorphism classes of elliptic curves defined over $\mathbb{Q}_{\infty,\ell}$ with good reduction away from S' and with full 2-torsion in $\mathbb{Q}_{\infty,\ell}$. Moreover, these elliptic curves form infinitely many distinct $\mathbb{Q}_{\infty,\ell}$-isogeny classes.*

Remarks. • By [Freitas et al. 2020, Lemma 2.1], a rational prime $p \neq \ell$ is inert in $\mathbb{Q}_{\infty,\ell}$ if and only if $p^{\ell-1} \not\equiv 1 \pmod{\ell^2}$. It follows from this that 2 is inert in $\mathbb{Q}_{\infty,\ell}$ for $\ell = 3, 5, 7$ and 11.

• Faltings' proof [1983] of the Mordell conjecture can be considered to have three major steps. In the first step, Faltings proves the Tate homomorphism conjecture. In the second step, Faltings derives the Shafarevich conjecture from the Tate homomorphism conjecture, and in the final step Faltings uses the "Parshin trick" to deduce the Mordell conjecture from the Shafarevich conjecture. Although Zarhin has extended the Tate homomorphism conjecture to K_∞, Theorem 4 suggests that there is no plausible strategy for proving the conjecture of Parshin and Zarhin by mimicking Faltings' proof of the Mordell conjecture.

It is natural to wonder if the isogeny classes appearing in the proof of Theorem 4 are finite or infinite. Rather reassuringly they turn out to be finite.

Theorem 5. *Let E be an elliptic curve over $\mathbb{Q}_{\infty,\ell}$ without potential complex multiplication. Then the $\mathbb{Q}_{\infty,\ell}$-isogeny class of E is finite.*

The original version of Shafarevich's conjecture [1963] (also proved by Faltings [1983, Korollar 1]) states that for a given number field K, a genus g and a finite set of places S, there are only finitely many isomorphism classes of genus-g curves C/K with good reduction away from S. Again this statement becomes false if we replace K by $\mathbb{Q}_{\infty,\ell}$ for any prime ℓ.

Theorem 6. *Let $g \geq 2$ and let $\ell = 3, 5, 7, 11$ or 13. There are infinitely many $\overline{\mathbb{Q}}$-isomorphism classes of genus-g hyperelliptic curves over $\mathbb{Q}_{\infty,\ell}$ with good reduction away from $\{\upsilon_2, \upsilon_\ell\}$.*

Theorem 7. *Let $\ell \geq 11$ be an odd prime and let $g = \left\lfloor \frac{\ell-3}{4} \right\rfloor$. There are infinitely many $\overline{\mathbb{Q}}$-isomorphism classes of genus-g hyperelliptic curves over $\mathbb{Q}_{\infty,\ell}$ with good reduction away from $\{\upsilon_2, \upsilon_\ell\}$. Moreover, if*

$$\ell \in \{11, 23, 59, 107, 167, 263, 347, 359\},$$

then the Jacobians of these curves form infinitely many distinct $\mathbb{Q}_{\infty,\ell}$-isogeny classes.

The paper is structured as follows. In Section 2 we recall basic results on units and S-units of the cyclotomic field $\mathbb{Q}(\zeta_{\ell^n})$. In Sections 3–6 we employ identities between cyclotomic polynomials to give constructive proofs of Theorems 1, 2 and 3. Section 7 gives a proof of Theorem 5, making use of a deep theorem of Kato to control the $\mathbb{Q}_{\infty,\ell}$-points on certain modular curves. Section 8 uses the integral and S-integral points on $\mathbb{P}^1 - \{0, 1, \infty\}$ furnished by Theorem 1 to construct infinite families of elliptic curves over $\mathbb{Q}_{\infty,\ell}$ for $\ell = 2, 3, 5, 7$, with good reduction away from $\{\upsilon_2, \upsilon_\ell\}$, which are used to give a proof of Theorem 4. Sections 9 and 10 give proofs of Theorems 6 and 7, making use of the relation, due to Kummer, between the class number of $\mathbb{Q}(\zeta_{\ell^n})^+$, and the index of cyclotomic units in the full group of units.

2. Units and S-units of $\mathbb{Q}(\zeta)$

Let K be a subfield of $\overline{\mathbb{Q}}$. We denote the integers of K (i.e., the integral closure of $\mathbb{Z}$ in K) by $\mathcal{O}(K)$. Let p be a rational prime. By a *prime of K above p* we mean a map $\upsilon : K \to \mathbb{Q} \cup \{\infty\}$ satisfying the following:

- $\upsilon(p) = 1$, $\upsilon(0) = \infty$.
- $\upsilon|_{K^\times} : K^\times \to \mathbb{Q}$ is a homomorphism.
- $\upsilon(1 + b) = 0$ whenever $\upsilon(b) > 0$.

Suppose $K = \bigcup K_n$ where $K_0 \subset K_1 \subset K_2 \subset \cdots$ is a tower of number fields (i.e., finite extensions of $\mathbb{Q}$), with $K_0 = \mathbb{Q}$. One sees that the primes of K above p are in one-to-one correspondence with sequences $\{\mathfrak{p}_n\}$ where:

- $\mathfrak{p}_n$ is a prime ideal of $\mathcal{O}(K_n)$.
- $\mathfrak{p}_{n+1} \mid \mathfrak{p}_n \mathcal{O}(K_{n+1})$.
- $\mathfrak{p}_0 = p\mathbb{Z}$.

Indeed, from υ one obtains the corresponding sequence $\{\mathfrak{p}_n\}$ via the formula $\mathfrak{p}_n = \{\alpha \in \mathcal{O}(K_n) : \upsilon(\alpha) > 0\}$. Given a sequence $\{\mathfrak{p}_n\}$, we can recover the corresponding υ by letting

$$\upsilon(\alpha) = \operatorname{ord}_{\mathfrak{p}_n}(\alpha) / \operatorname{ord}_{\mathfrak{p}_n}(p)$$

whenever $\alpha \in K_n^\times$. Given a finite set of primes S of K, we define the S-integers of K to be the set $\mathcal{O}(K, S)$ of all $\alpha \in K$ such that $\upsilon(\alpha) \geq 0$ for every prime $\upsilon \notin S$. We let $\mathcal{O}(K, S)^\times$ be the unit group of $\mathcal{O}(K, S)$; this is precisely the set of $\alpha \in K^\times$ such that $\upsilon(\alpha) = 0$ for every prime $\upsilon \notin S$. If $S = \varnothing$ then $\mathcal{O}(K, S) = \mathcal{O}(K)$ are the integers of K and $\mathcal{O}(K, S)^\times = \mathcal{O}(K)^\times$ are the units of K.

Fix a rational prime ℓ. For a positive integer n, let ζ_{ℓ^n} denote a primitive ℓ^n-th root of 1 which is chosen so that

$$\zeta_{\ell^{n+1}}^\ell = \zeta_{\ell^n}.$$

Let $\Omega_{n,\ell} = \mathbb{Q}(\zeta_{\ell^n})$; this has degree $\varphi(\ell^n)$, where φ is Euler totient function. Let

$$\Omega_{\infty,\ell} = \bigcup_{n=1}^{\infty} \Omega_{n,\ell}.$$

The prime ℓ is totally ramified in each $\Omega_{n,\ell}$, and we denote by λ_n the unique prime ideal of $\mathcal{O}(\Omega_{n,\ell})$ above ℓ. Thus

$$\ell \cdot \mathcal{O}(\Omega_{n,\ell}) = \lambda_n^{\varphi(\ell^n)}. \tag{2}$$

We write υ_ℓ for the unique prime of $\Omega_{\infty,\ell}$ above ℓ. For now fix $n \geq 1$ if $\ell \neq 2$ and $n \geq 2$ if $\ell = 2$. We recall that $\lambda_n = (1 - \zeta_{\ell^n}) \cdot \mathcal{O}(\Omega_{n,\ell})$. If $\ell \nmid s$ then $(1 - \zeta_{\ell^n}^s) \cdot \mathcal{O}(\Omega_{n,\ell}) = \lambda_n$; we can see this by applying the automorphism $\zeta_{\ell^n} \mapsto \zeta_{\ell^n}^s$ to (2).

Lemma 8. *Let s be an integer and let $t = \mathrm{ord}_\ell(s)$. Suppose $t < n$. Then*

$$(1-\zeta_{\ell^n}^s)\cdot \mathcal{O}(\Omega_{n,\ell}) = \lambda_n^{\ell^t}.$$

Moreover,

$$\upsilon_\ell(1-\zeta_{\ell^n}^s) = \frac{1}{\ell^{n-1-t}(\ell-1)}.$$

Proof. Write $\zeta = \zeta_{\ell^n}$. Note that ζ^s is a primitive ℓ^{n-t}-th root of 1. Thus

$$(1-\zeta^s)\cdot \mathcal{O}(\Omega_{n-t,\ell}) = \lambda_{n-t}.$$

As ℓ is totally ramified in $\Omega_{n,\ell}$, we have

$$(1-\zeta^s)\cdot \mathcal{O}(\Omega_{n,\ell}) = \lambda_n^{[\Omega_{n,\ell}:\Omega_{n-t,\ell}]} = \lambda_n^{\ell^t}.$$

For the final part of the lemma,

$$\upsilon_\ell(1-\zeta^s) = \frac{\mathrm{ord}_{\lambda_n}(1-\zeta^s)}{\mathrm{ord}_{\lambda_n}(\ell)} = \frac{\ell^t}{\varphi(\ell^n)} = \frac{1}{\ell^{n-1-t}(\ell-1)}.$$ □

Cyclotomic units and S-units. Write V_n for the subgroup of $\mathcal{O}(\Omega_n, \{\upsilon_\ell\})^\times$ generated by

$$\{\pm\zeta_{\ell^n}, 1-\zeta_{\ell^n}^k : 1 \le k < \ell^n\}$$

and let

$$C_n = V_n \cap \mathcal{O}(\Omega_n)^\times.$$

The group C_n is called [Washington 1997, Chapter 8] the group of *cyclotomic units* in Ω_n. We will often find it more convenient to work with the group V_n.

Lemma 9. *The abelian group $V_n/\langle\pm\zeta_{\ell^n}\rangle$ is free with basis*

$$\{1-\zeta_{\ell^n}^k : 1 \le k < \ell^n/2, \ \ell \nmid k\}. \tag{3}$$

Proof. The torsion subgroup of V_n is the torsion subgroup of $\Omega_n^\times$ which is $\langle\pm\zeta_{\ell^n}\rangle$. Thus $V_n/\langle\pm\zeta_{\ell^n}\rangle$ is torsion free. By definition of V_n, the group $V_n/\langle\pm\zeta_{\ell^n}\rangle$ is generated by $1-\zeta_{\ell^n}^k$ with $\ell^n \nmid k$. Write $k = \ell^r d$ with $\ell \nmid d$; thus $r < n$. Suppose $r \ge 1$. Then,

$$\begin{aligned}
1-\zeta_{\ell^n}^k &= 1-\zeta_{\ell^n}^{\ell^r d} \\
&= \prod_{i=0}^{\ell^r-1}(1-\zeta_{\ell^n}^d\zeta_{\ell^r}^i) \quad \text{using} \quad 1-X^{\ell^r} = \prod_{i=0}^{\ell^r-1}(1-\zeta_{\ell^r}^i X) \\
&= \prod_{i=0}^{\ell-1}(1-\zeta_{\ell^n}^{d+i\ell^{n-r}}).
\end{aligned}$$

It follows that $V_n/\langle\pm\zeta_{\ell^n}\rangle$ is generated by $1-\zeta_{\ell^n}^k$ with $\ell \nmid k$. If $\ell^n/2 < k < \ell^n$ and $\ell \nmid k$ then

$$1-\zeta_{\ell^n}^k = -\zeta_{\ell^n}^k(1-\zeta_{\ell^n}^{\ell^n-k}). \tag{4}$$

Thus (3) certainly generates $V_n/\langle \pm\zeta_\ell^n \rangle$. Note that (3) has cardinality $\varphi(\ell^n)/2$ where φ is the Euler totient function. It therefore suffices to show that V_n has rank $\varphi(\ell^n)/2$. A well-known theorem [Washington 1997, Theorem 8.3] states that C_n has finite index in $\mathcal{O}(\Omega_n)^\times$ and thus, by Dirichlet's unit theorem, C_n has rank $-1+\varphi(\ell^n)/2$. We note that C_n is the kernel of the surjective homomorphism $V_n \to \mathbb{Z}$, sending μ to $\mathrm{ord}_{\lambda_n}(\mu)$. Thus V_n has rank $\varphi(\ell^n)/2$ completing the proof. □

Lemma 10. *Let $n \ge 2$ if $\ell \ne 2$ and $n \ge 3$ if $\ell = 2$. Then $V_{n-1} \subset V_n$. Moreover,*

$$\prod_{\substack{1 \le k < \ell^n/2 \\ \ell \nmid k}} (1-\zeta_{\ell^n}^k)^{c_k} \in \langle \pm\zeta_{\ell^n}, V_{n-1} \rangle$$

if and only if $c_k = c_m$ whenever $k \equiv m \pmod{\ell^{n-1}}$.

Proof. The group V_{n-1} is generated, modulo roots of unity, by $1-\zeta_{\ell^{n-1}}^d$ with $\ell \nmid d$. By the proof of Lemma 9,

$$1-\zeta_{\ell^{n-1}}^d = 1-\zeta_{\ell^n}^{\ell d} = \prod_{i=0}^{\ell-1}(1-\zeta_{\ell^n}^{d+i\ell^{n-1}}).$$

The lemma follows from Lemma 9. □

Given $a \in \mathbb{Z}_\ell$, it makes sense to reduce a modulo ℓ^n and therefore it makes sense to write $\zeta_{\ell^n}^a$. We write $\{a\}_n$ for the unique integer satisfying

$$0 \le \{a\}_n < \ell^n/2, \quad \{a\}_n \equiv \pm a \pmod{\ell^n}.$$

Lemma 11. *Let $a_1, \dots, a_r \in \mathbb{Z}_\ell$ and $c_1, \dots, c_r \in \mathbb{Z}$. Suppose:*

(i) $c_1 \ne 0$.

(ii) $a_1 \not\equiv 0 \pmod{\ell}$.

(iii) $a_1 \ne \pm a_2, \pm a_3, \dots, \pm a_r \pmod{\ell^n}$.

Write

$$\varepsilon_n = \prod_{1 \le i \le r} (1-\zeta_{\ell^n}^{a_i})^{c_i}. \tag{5}$$

Then, $\varepsilon_n \notin \langle \pm\zeta_{\ell^n}, V_{n-1} \rangle$ for all sufficiently large n.

Proof. If $a_j \equiv 0 \pmod{\ell}$ then $(1-\zeta_{\ell^n}^{a_j}) \in V_{n-1}$. We may therefore suppose $a_j \not\equiv 0 \pmod{\ell}$ for all j. Write

$$\delta_n = \prod_{1 \le i \le r} (1-\zeta_{\ell^n}^{\{a_i\}_n})^{c_i}.$$

In view of the identity (4) it will be sufficient to show that $\delta_n \notin \langle \pm\zeta_{\ell^n}, V_{n-1} \rangle$ for n sufficiently large. Also, in view of Lemma 10, it is sufficient to show for sufficiently large n that $\{a_1\}_n \not\equiv \{a_j\}_n \pmod{\ell^n}$ for all $2 \le j \le n$. This is equivalent to $a_1 \ne \pm a_j$ for $2 \le j \le n$ which is hypothesis (iii). This completes the proof. □

The following corollary easily follows from Lemma 11.

Corollary 12. *Let $a_1, \dots, a_r \in \mathbb{Z}_\ell$ and $c_1, \dots, c_r \in \mathbb{Z}$. Suppose*:

(i) $c_1 \equiv 1 \pmod 2$.

(ii) $a_1 \not\equiv 0 \pmod \ell$.

(iii) $a_1 \neq \pm a_2, \pm a_3, \dots, \pm a_r \pmod{\ell^n}$.

Let ε_n be as in (5). *Then, $\varepsilon_n \notin \langle \pm\zeta_{\ell^n}, V_{n-1}, V_n^2 \rangle$ for all sufficiently large n.*

Units and S-units from cyclotomic polynomials. For $m \geq 1$, let $\Phi_m(X) \in \mathbb{Z}[X]$ be the *m-th cyclotomic polynomial* defined by

$$\Phi_m(X) = \prod_{\substack{1 \leq i \leq m \\ (i,m)=1}} (X - \zeta_m^i).$$

These satisfy the identity [Washington 1997, Chapter 2]

$$X^m - 1 = \prod_{d \mid m} \Phi_d(X). \tag{6}$$

It follows from the Möbius inversion formula that

$$\Phi_m(X) = \prod_{d \mid m} (X^d - 1)^{\mu(m/d)}, \tag{7}$$

where μ denotes the Möbius function.

Lemma 13. *Let ℓ be a prime and $n \geq 1$. Let $m \geq 1$, and suppose $\ell^n \nmid m$*:

(a) $\Phi_m(\zeta_{\ell^n}) \in V_n \subseteq \mathcal{O}(\Omega_{n,\ell}, S)^\times$, *where $S = \{\upsilon_\ell\}$.*

(b) *If $m \neq \ell^u$ for all $u \geq 0$, then $\Phi_m(\zeta_{\ell^n}) \in C_n \subseteq \mathcal{O}(\Omega_{n,\ell})^\times$.*

Moreover,

$$\upsilon_\ell(\Phi_{\ell^t}(\zeta_{\ell^n})) = \begin{cases} 1/(\ell^{n-1}(\ell-1)), & t = 0, \\ 1/\ell^{n-t}, & 1 \leq t \leq n-1. \end{cases}$$

Proof. Let $t = \operatorname{ord}_\ell(m) < n$. Observe that $\Phi_m(X) \mid (X^m - 1)$. Hence $\Phi_m(\zeta_{\ell^n}) \cdot \mathcal{O}(\Omega_{n,\ell})$ divides $(1 - \zeta_{\ell^n}^m) \cdot \mathcal{O}(\Omega_{n,\ell})$. By Lemma 8 we have $(1 - \zeta_{\ell^n}^m) \cdot \mathcal{O}(\Omega_{n,\ell}) = \lambda_n^{\ell^t}$, giving (a).

For (b), write $m = \ell^t k$ where $k > 1$. Then $\Phi_m(X)$ divides the polynomial $(X^m - 1)/(X^{\ell^t} - 1)$. Therefore $\Phi_m(\zeta_{\ell^n}) \cdot \mathcal{O}(\Omega_{n,\ell})$ divides

$$\frac{(1 - \zeta_{\ell^n}^m)}{(1 - \zeta_{\ell^n}^{\ell^t})} \cdot \mathcal{O}(\Omega_{n,\ell}) = \frac{\lambda_n^{\ell^t}}{\lambda_n^{\ell^t}} = 1 \cdot \mathcal{O}(\Omega_{n,\ell}).$$

Thus $\Phi_m(\zeta_{\ell^n})$ is a unit, giving (b).

The final part of the lemma follows from Lemma 8, and the formulae

$$\Phi_{\ell^t}(X) = \begin{cases} X - 1, & t = 0, \\ (X^{\ell^t} - 1)/(X^{\ell^{t-1}} - 1), & t \geq 1. \end{cases}$$

□

Lemma 14. *Let $n \geq 2$ if $\ell \neq 2$ and $n \geq 3$ if $\ell = 2$. Then $V_n/\langle \pm\zeta_{\ell^n}\rangle$ is free with basis*

$$\{\Phi_m(\zeta_{\ell^n}) : 1 \leq m < \ell^n/2, \ell \nmid m\}.$$

Proof. This follows from Lemma 9 thanks to identities (6) and (7). □

3. The S-unit equation over $\mathbb{Q}(\zeta_{\ell^n})^+$

We continue with the notation of the previous section. In particular, let K be a subfield of $\overline{\mathbb{Q}}$ and S be a finite set of primes of K. Let k be a nonzero rational integer. We shall make frequent use of the correspondence between elements of $(\mathbb{P}^1 - \{0, k, \infty\})(\mathcal{O}(K, S))$ and the set of solutions to the S-unit equation

$$\varepsilon + \delta = k, \quad \varepsilon, \delta \in \mathcal{O}(K, S)^\times,$$

sending $\varepsilon \in (\mathbb{P}^1 - \{0, k, \infty\})(\mathcal{O}(K, S))$ to $(\varepsilon, \delta) = (\varepsilon, k - \varepsilon)$.

Now, as before, let ℓ be a rational prime and n is a positive integer. If $\ell = 2$ suppose $n \geq 2$. Let $\zeta = \zeta_{\ell^n}$, and write $\Omega^+_{n,\ell} = \mathbb{Q}(\zeta + 1/\zeta)$ for the index-2 totally real subfield of $\Omega_{n,\ell}$. Let

$$\Omega^+_{\infty,\ell} = \bigcup_{n=1}^{\infty} \Omega^+_{n,\ell}.$$

In this section, for suitable S, we produce solutions to S-unit equations over $\Omega^+_{\infty,\ell}$.

As before, Φ_m denotes the m-cyclotomic polynomial. It is convenient to record the first few Φ_m:

$$\Phi_1 = X - 1, \quad \Phi_2 = X + 1, \quad \Phi_3 = X^2 + X + 1,$$
$$\Phi_4 = X^2 + 1, \quad \Phi_5 = X^4 + X^3 + X^2 + X + 1, \quad \Phi_6 = X^2 - X + 1,$$
$$\Phi_7 = X^6 + X^5 + X^4 + X^3 + X^2 + X + 1, \quad \Phi_8 = X^4 + 1,$$
$$\Phi_9 = X^6 + X^3 + 1, \quad \Phi_{10} = X^4 - X^3 + X^2 - X + 1.$$

We shall call a polynomial $F \in \mathbb{Z}[X]$ *supercyclotomic* if it is of the form $X^m f_1 f_2 \cdots f_k$ where each $f_i(X)$ is a cyclotomic polynomial. We know, thanks to Lemma 13, that if F is supercyclotomic and ℓ is a prime, then $F(\zeta_{\ell^n}) \in \mathcal{O}(\Omega_n, \{\upsilon_\ell\})^\times$ for n sufficiently large. We wrote a short computer program that lists all supercyclotomic polynomials of degree at most 20 and searches for ternary relations of the form $F - G = kH$ with F, G, H supercyclotomic, $\gcd(F, G, H) = 1$ and k is a positive integer. Note that any such relation $F - G = kH$ gives points

$$\varepsilon_n = F(\zeta_{\ell^n})/H(\zeta_{\ell^n}) \in (\mathbb{P}^1 - \{0, k, \infty\})(\mathcal{O}(\Omega_n, \{\upsilon_\ell\})),$$

for n sufficiently large. We found the following ternary relations between supercyclotomic polynomials:

$$\Phi_2(X)^2 - \Phi_3(X) = X, \tag{8}$$

$$\Phi_2(X)^2 - \Phi_4(X) = 2X, \tag{9}$$

$$\Phi_2(X)^2 - \Phi_6(X) = 3X, \tag{10}$$

$$\Phi_2(X)^2 - \Phi_1(X)^2 = 4X, \tag{11}$$

$$\Phi_2(X)^4 - \Phi_{10}(X) = 5X\Phi_3(X), \tag{12}$$

$$\Phi_2^2(X)\Phi_3(X) - \Phi_1(X)^2\Phi_6(X) = 6X\Phi_4(X), \tag{13}$$

$$\Phi_7(X) - \Phi_1(X)^6 = 7X\Phi_6(X)^2, \tag{14}$$

$$\Phi_2(X)^4 - \Phi_1(X)^4 = 8X\Phi_4(X), \tag{15}$$

$$\Phi_2(X)^4\Phi_5(X) - \Phi_1(X)^4\Phi_{10}(X) = 10X\Phi_4(X)^3. \tag{16}$$

From the identities (6) and (7) one easily sees that $F(X^k)$ is supercyclotomic for any supercyclotomic polynomial F and any positive integer k; thus each of the nine identities above in fact yields an infinite family of identities. We pose the following open problems:

- Are there ternary linear relations between supercyclotomic polynomials that are outside these nine families?
- Classify all ternary linear relations between supercyclotomic polynomials.

Lemma 15. *Let $c : \Omega_\ell \to \Omega_\ell$ denote complex conjugation. Let $n \geq 1$ and let $\zeta = \zeta_{\ell^n}$ be an ℓ^n-th root of 1. Let $m \geq 1$ and suppose $\ell^n \nmid m$. Then*

$$\frac{\Phi_m(\zeta)^c}{\Phi_m(\zeta)} = \begin{cases} \zeta^{-\varphi(m)}, & m \geq 2, \\ -\zeta^{-1}, & m = 1. \end{cases}$$

Proof. Note that $\zeta^c = \zeta^{-1}$. So

$$\frac{\Phi_1(\zeta)^c}{\Phi_1(\zeta)} = \frac{\zeta^{-1}-1}{\zeta-1} = -\zeta^{-1}, \qquad \frac{\Phi_2(\zeta)^c}{\Phi_2(\zeta)} = \frac{\zeta^{-1}+1}{\zeta+1} = \zeta^{-1}.$$

Let $m \geq 3$. The polynomial Φ_m is monic of degree $\varphi(m)$, and its roots are the primitive m-th roots of 1 which come in distinct pairs η, η^{-1}. Thus the trailing coefficient is 1. It follows that $X^{\varphi(m)}\Phi_m(X^{-1})$ is monic and has the same roots as Φ_m, therefore

$$\Phi_m(X) = X^{\varphi(m)}\Phi_m(X^{-1}).$$

Hence

$$\frac{\Phi_m(\zeta)^c}{\Phi_m(\zeta)} = \frac{\Phi_m(\zeta^{-1})}{\Phi_m(\zeta)} = \zeta^{-\varphi(m)}. \qquad \square$$

Lemma 16. *Let ℓ be a prime. Let $F \in \mathbb{Z}[X]$ be a product of powers of cyclotomic polynomials. Suppose that the exponents of $\Phi_1(X)$ and $\Phi_2(X)$ in the factorization of F are both even. Then F has even degree and, for suitably large n, we have*

$$\zeta^{-\deg(F)/2}F(\zeta) \in \mathcal{O}(\Omega_{n,\ell}^{+}, S)^\times,$$

where $\zeta = \zeta_{\ell^n}$ and $S = \{\upsilon_\ell\}$.

Proof. We note that Φ_m has degree $\varphi(m)$ which is even for $m \geq 3$. It follows from this that F has even degree. From Lemma 13 we have $\zeta^{-\deg(F)/2}F(\zeta) \in \mathcal{O}(\Omega_{n,\ell}, S)^\times$ for suitably large n. To prove the lemma we need to show that $\zeta^{-\deg(F)/2}F(\zeta)$ is fixed by complex conjugation. Let G be either Φ_1^2, or Φ_2^2, or Φ_m with $m \geq 3$. We claim that $\zeta^{-\deg(G)/2}G(\zeta)$ is fixed by complex conjugation. Since F is a product of such G, the lemma follows from our claim. The claim is trivially true for $G = \Phi_1^2$ and $G = \Phi_2^2$, and follows immediately from Lemma 15 for $G = \Phi_m$ with $m \geq 3$. □

Lemma 17. *Let $S = \{\upsilon_\ell\}$. Let*

$$k \in \{1, 2, 3, 4, 5, 6, 7, 8, 10\}.$$

Then $(\mathbb{P}^1 - \{0, k, \infty\})(\mathcal{O}(\Omega^+_{\infty,\ell}, S))$ is infinite.

Proof. The proof makes use of identities (8)–(16). Each identity has the form $P - Q = kXR$, where P, Q, and R are supercyclotomic polynomials. Let n be sufficiently large so that ζ_{ℓ^n} is not a root of PQR, and write

$$\varepsilon_n = \frac{P(\zeta_{\ell^n})}{\zeta_{\ell^n} R(\zeta_{\ell^n})}, \qquad \delta_n = \frac{-Q(\zeta_{\ell^n})}{\zeta_{\ell^n} R(\zeta_{\ell^n})}.$$

From the identity $P - Q = kXR$ we see that $\varepsilon_n + \delta_n = k$. We note the following features of the triples (P, Q, R) common to all the identities (8)–(16):

- In every case, P, Q, R are products of powers of cyclotomic polynomials where the exponents of Φ_1 and Φ_2 are both even.
- Write $d = \deg(P)$. Then $\deg(Q) = d$ and $\deg(R) = d - 2$. Indeed as supercyclotomic polynomials are monic, the relation $P - Q = kXR$ forces P and Q to have the same degree as soon as $k \geq 2$.

We may rewrite ε_n as

$$\varepsilon_n = \frac{\zeta_{\ell^n}^{-d/2} P(\zeta_{\ell^n})}{\zeta_{\ell^n}^{-(d-2)/2} R(\zeta_{\ell^n})}, \qquad \delta_n = \frac{-\zeta_{\ell^n}^{-d/2} Q(\zeta_{\ell^n})}{\zeta_{\ell^n}^{-(d-2)/2} R(\zeta_{\ell^n})}.$$

By Lemma 16, we have $\varepsilon_n, \delta_n \in \mathcal{O}(\Omega^+_{n,\ell}, S)^\times$ for n suitably large, and therefore ε_n is an $\mathcal{O}(\Omega^+_{\infty,}, S)$-point on $\mathbb{P}^1 - \{0, k, \infty\}$. To complete the proof we need to show that we obtain infinitely many distinct points as we vary n. We will do this for $k = 10$. The other cases are similar. Note that

$$\varepsilon_n = \frac{\Phi_2(\zeta_{\ell^n})^4 \Phi_5(\zeta_{\ell^n})}{\zeta_{\ell^n} \Phi_4(\zeta_{\ell^n})^3} = \frac{(1 - \zeta_{\ell^n}^2)^7 (1 - \zeta_{\ell^n}^5)}{\zeta_{\ell^n} (1 - \zeta_{\ell^n})^5 (1 - \zeta_{\ell^n}^4)^3} \in V_n.$$

To show that we obtain infinitely many distinct ε_n it is enough to show that $\varepsilon_n \notin V_{n-1}$ for n sufficiently large. This follows by an easy application of Lemma 10; to illustrate this let $\ell = 5$ and suppose $\varepsilon_n \in V_{n-1}$. Note that $1 - \zeta_{5^n}^5 \in V_{n-1}$. It follows that

$$(1 - \zeta_{5^n})^{-5} (1 - \zeta_{5^n}^2)^7 (1 - \zeta_{5^n}^4)^{-3} \in \langle \pm \zeta_{\ell^n}, V_{n-1} \rangle.$$

Now in the product on the left the exponent of $1 - \zeta_{5^n}$ is -5 whereas the exponent of $1 - \zeta_{5^n}^{1+5^{n-1}}$ is 0, contradicting Lemma 10. The proof is similar for $\ell = 2$, and for $\ell \neq 2, 5$. It follows that we have infinitely many $\mathcal{O}(\Omega^+_{\infty,\ell}, S)$-points on $\mathbb{P}^1 - \{0, 10, \infty\}$. □

Proof of Theorem 2 for $\ell = 2$ and 3. For $\ell = 2, 3$, we have $\Omega^+_{\infty,\ell} = \mathbb{Q}_{\infty,\ell}$. Indeed, if $\ell = 2$ then $\mathbb{Q}_{n,2} = \Omega^+_{n+2,2}$ and if $\ell = 3$ then $\mathbb{Q}_{n,3} = \Omega^+_{n+1,3}$. Therefore Theorem 2 with $\ell = 2$ and 3 follows immediately from Lemma 17 for $k \in \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$.

Also, if $\ell = 2$, then the infinitely many solutions $\varepsilon + \delta = 6$ yield infinitely many solutions for $2\varepsilon + 2\delta = 12$ and $4\varepsilon + 4\delta = 24$. And if $\ell = 3$, then the infinitely many solutions $\varepsilon + \delta = 4$ yield infinitely many solutions $3\varepsilon + 3\delta = 12$, and similarly the infinitely many solutions $\varepsilon + \delta = 8$ yield infinitely many solutions $3\varepsilon + 3\delta = 24$. This proves Theorem 2 for $\ell = 2, 3$ and $k \in \{12, 24\}$. □

Proof of Theorem 1 for $\ell = 2$. Theorem 1 for $\ell = 2$ is simply a special case of Theorem 2. □

4. The unit equation over $\mathbb{Q}(\zeta_{\ell^n})^+$

For roots of unity α, β, we let

$$E(\alpha, \beta) = \frac{\alpha^2 + \alpha^{-2}}{(\alpha\beta^{-1} + \alpha^{-1}\beta)(\alpha\beta + \alpha^{-1}\beta^{-1})} = \frac{\Phi_8(\alpha)}{\Phi_4(\alpha\beta)\Phi_4(\alpha/\beta)},$$

$$F(\alpha, \beta) = \frac{\beta^2 + \beta^{-2}}{(\alpha\beta^{-1} + \alpha^{-1}\beta)(\alpha\beta + \alpha^{-1}\beta^{-1})} = \frac{\Phi_8(\beta)}{\Phi_4(\alpha\beta)\Phi_4(\beta/\alpha)}.$$

We easily check that

$$E(\alpha, \beta) + F(\alpha, \beta) = 1. \tag{17}$$

Lemma 18. *Suppose ℓ is odd and $n \geq 1$. Let $\zeta = \zeta_{\ell^n}$. Let i, j be integers satisfying i, j, $i + j$, $i - j \not\equiv 0 \pmod{\ell^n}$. Then $E(\zeta^i, \zeta^j)$, $F(\zeta^i, \zeta^j) \in \mathcal{O}(\Omega^+_{n,\ell})^\times$, and satisfy the unit equation*

$$\varepsilon + \delta = 1, \quad \varepsilon, \delta \in \mathcal{O}(\Omega^+_{n,\ell})^\times. \tag{18}$$

Moreover,

$$\upsilon_\ell(E(\zeta^i, \zeta^j) - F(\zeta^i, \zeta^j)) = \frac{\ell^{\operatorname{ord}_\ell(i+j)} + \ell^{\operatorname{ord}_\ell(i-j)}}{\ell^{n-1}(\ell - 1)}. \tag{19}$$

Proof. It is clear that $E(\zeta^i, \zeta^j)$, $F(\zeta^i, \zeta^j)$ are fixed by complex conjugation $\zeta \mapsto \zeta^{-1}$ and so belong to $\Omega^+_{n,\ell}$. By Lemma 13, $E(\zeta^i, \zeta^j)$ and $F(\zeta^i, \zeta^j)$ are units. It remains to check (19). We observe

$$E(\zeta^i, \zeta^j) - F(\zeta^i, \zeta^j) = \frac{(\zeta^{i-j} - \zeta^{j-i})(\zeta^{i+j} - \zeta^{-i-j})}{(\zeta^{i-j} + \zeta^{j-i})(\zeta^{i+j} + \zeta^{-i-j})} = \frac{(\zeta^{2(i-j)} - 1)(\zeta^{2(i+j)} - 1)}{\Phi_4(\zeta^{i-j})\Phi_4(\zeta^{i+j})}.$$

The denominator is a unit by Lemma 13. Now (19) follows from Lemma 8. □

Proof of Theorem 3. We deduce this from Lemma 18. Let us take for example $i = 2$ and $j = 1$. Let $n \geq 2$ and let

$$\varepsilon_n = E(\zeta^2_{\ell^n}, \zeta_{\ell^n}), \quad \delta_n = F(\zeta^2_{\ell^n}, \zeta_{\ell^n}).$$

By Lemma 18, ε_n, $\delta_n \in \mathcal{O}(\Omega^+_{\infty,\ell})^\times$ and satisfy $\varepsilon_n + \delta_n = 1$. Thus $\varepsilon_n \in (\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}(\Omega^+_{\infty,\ell}))$. Moreover,

$$\upsilon_\ell(2\varepsilon_n - 1) = \upsilon_\ell(\varepsilon_n - \delta_n) = \begin{cases} 2/(\ell^{n-1}(\ell - 1)), & \ell > 3, \\ 2/3^{n-1}, & \ell = 3, \end{cases}$$

by (19). Thus $\varepsilon_n \neq \varepsilon_m$ whenever $n \neq m$. Hence $(\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}(\Omega^+_{\infty,\ell}))$ is infinite. □

Remark. Theorem 3 applies only for ℓ odd; for $\ell = 2$ it is easy to show that the statement is false. Indeed, let η_n be the prime ideal of $\mathcal{O}(\Omega_{n,2}^+)$ above 2. Then $\mathcal{O}(\Omega_{n,2}^+)/\eta_n \cong \mathbb{F}_2$, and a solution to $\varepsilon + \delta = 1$ with ε, $\delta \in \mathcal{O}(\Omega_{n,2}^+)^\times$ reduced modulo η_n gives $1 + 1 \equiv 1 \pmod 2$ which is impossible.

Proof of Theorem 1 for $\ell = 3$. We recall that $\mathbb{Q}_{\infty,3} = \Omega_{\infty,3}^+$. Therefore Theorem 1 for $\ell = 3$ follows immediately from Theorem 3. □

5. The S-unit equation over $\mathbb{Q}_{\infty,5}$

The purpose of the is section is to prove Theorems 1 and 2 for $\ell = 5$. These in fact follow immediately from the following lemma.

Lemma 19. *Let υ_5 be the unique prime of $\mathbb{Q}_{\infty,5}$ above 5, and write $S = \{\upsilon_5\}$. Then*:

(i) $(\mathbb{P}^1 - \{0, k, \infty\})(\mathcal{O}(\mathbb{Q}_{\infty,5}, S))$ *is infinite for* $k = 1, 4$.

(ii) $(\mathbb{P}^1 - \{0, 2, \infty\})(\mathcal{O}(\mathbb{Q}_{\infty,5}))$ *is infinite.*

Proof. Let $a \in \mathbb{Z}_5^\times$ be the element satisfying

$$a^2 = -1, \quad a \equiv 2 \pmod 5;$$

such an element exists and is unique by Hensel's lemma. Let $\sigma : \Omega_{\infty,5} \to \Omega_{\infty,5}$ be the field automorphism satisfying

$$\sigma(\zeta_{5^n}) = \zeta_{5^n}^a$$

for $n \geq 1$. Note that σ is an automorphism of order 4, and fixes a subfield of $\Omega_{\infty,5}$ of index 4. This subfield is precisely $\mathbb{Q}_{\infty,5}$.

Let

$$\begin{aligned} F &= (x_1x_2^2 + x_3x_4^2)(x_1^2x_4 + x_2x_3^2), \\ G &= (x_1^2x_2 + x_3^2x_4)(x_1x_4^2 + x_2^2x_3), \\ H &= (x_1 - x_3)(x_2 - x_4)(x_1x_2 - x_3x_4)(x_1x_4 - x_2x_3). \end{aligned}$$

Observe F, G, H are invariant under the 4-cycle (x_1, x_2, x_3, x_4). One can check that $F - G = H$. Let $n \geq 2$ and write $\zeta = \zeta_{5^n}$. Let

$$\varepsilon_n = \frac{F(\zeta, \zeta^a, \zeta^{a^2}, \zeta^{a^3})}{H(\zeta, \zeta^a, \zeta^{a^2}, \zeta^{a^3})}, \quad \delta_n = -\frac{G(\zeta, \zeta^a, \zeta^{a^2}, \zeta^{a^3})}{H(\zeta, \zeta^a, \zeta^{a^2}, \zeta^{a^3})}.$$

From the identity $F - G = H$ we have $\varepsilon_n + \delta_n = 1$. We shall show that $\varepsilon_n, \delta_n \in \mathcal{O}(\mathbb{Q}_{\infty,5}, S)^\times$.

Since σ cyclically permutes $\zeta, \zeta^a, \zeta^{-1}, \zeta^{-a}$ we conclude that $f(\zeta, \zeta^a, \zeta^{-1}, \zeta^{-a}) \in \mathbb{Q}_{\infty,5}$ for $f = F$, G, H. Thus $\varepsilon_n, \delta_n \in \mathbb{Q}_{\infty,5}$. Moreover,

$$\begin{aligned} F &= x_2x_3^3x_4^2 \cdot \Phi_2(x_1x_2^2/x_3x_4^2)\Phi_2(x_1^2x_4/x_2x_3^2), \\ G &= x_2^2x_3^3x_4 \cdot \Phi_2(x_1^2x_2/x_3^2x_4)\Phi_2(x_1x_4^2/x_2^2x_3), \\ H &= x_2x_3^3x_4^2 \cdot \Phi_1(x_1/x_3) \cdot \Phi_1(x_2/x_4) \cdot \Phi_1(x_1x_2/x_3x_4) \cdot \Phi_1(x_1x_4/x_2x_3). \end{aligned}$$

Hence

$$\begin{aligned}\varepsilon_n &= \frac{\Phi_2(\zeta^{2+4a})\Phi_2(\zeta^{4-2a})}{\Phi_1(\zeta^2)\Phi_1(\zeta^{2a})\Phi_1(\zeta^{2+2a})\Phi_1(\zeta^{2-2a})}\\ &= \frac{(1-\zeta^{4+8a})(1-\zeta^{8-4a})}{(1-\zeta^2)(1-\zeta^{2a})(1-\zeta^{2+2a})(1-\zeta^{2-2a})(1-\zeta^{2+4a})(1-\zeta^{4-2a})}.\end{aligned}$$

and

$$\begin{aligned}\delta_n &= \frac{-\zeta^{2a}\Phi_2(\zeta^{4+2a})\Phi_2(\zeta^{2-4a})}{\Phi_1(\zeta^2)\Phi_1(\zeta^{2a})\Phi_1(\zeta^{2+2a})\Phi_1(\zeta^{2-2a})}\\ &= \frac{-\zeta^{2a}(1-\zeta^{8+4a})(1-\zeta^{4-8a})}{(1-\zeta^2)(1-\zeta^{2a})(1-\zeta^{2+2a})(1-\zeta^{2-2a})(1-\zeta^{4+2a})(1-\zeta^{2-4a})}.\end{aligned}$$

We checked, using the fact that $a \equiv 7 \pmod{25}$, that the exponents of ζ in the above expressions for ε_n and δ_n all have 5-adic valuation 0 or 1. It follows from this that $\varepsilon_n, \delta_n \in V_n \subseteq \mathcal{O}(\Omega_n, S)^\times$ for $n \geq 2$. Hence $\varepsilon_n, \delta_n \in \mathbb{Q}_{\infty,5} \cap \mathcal{O}(\Omega_n, S)^\times = \mathcal{O}(\mathbb{Q}_{\infty,5}, S)^\times$ for $n \geq 2$. To complete the proof of the lemma for $k = 1$ it is enough to show that $\varepsilon_n \neq \varepsilon_m$ for $n > m$, and for this it is enough to show that $\varepsilon_n \notin \langle \pm\zeta_{5^n}, V_{n-1}\rangle$ for $n \geq 2$. Since $a \equiv 7 \pmod{25}$ we see that

$$4+8a \equiv 10, \quad 8-4a \equiv 5, \quad 2+4a \equiv 5, \quad 4-2a \equiv 15 \pmod{25}.$$

Thus the factors

$$1-\zeta^{4+8a}, \quad 1-\zeta^{8-4a}, \quad 1-\zeta^{2+4a}, \quad 1-\zeta^{4-2a}$$

all belong to V_{n-1}. Hence it is enough to show that

$$(1-\zeta^2)(1-\zeta^{2a})(1-\zeta^{2+2a})(1-\zeta^{2-2a}) \tag{20}$$

does not belong to $\langle \pm\zeta_{5^n}, V_{n-1}\rangle$. However, the exponents 2, $2a$, $2+2a$, $2-2a$ are respectively 2, 4, 1, 3 modulo 5, and hence certainly distinct modulo 5^{n-1}. It follows from Lemma 10 that the product (20) does not belong to $\langle \pm\zeta_{5^n}, V_{n-1}\rangle$ completing the proof for $k = 1$.

The proof for $k = 2$ is similar, and is based on the identity $F - G = 2H$, where

$$\begin{aligned}F &= (x_1^2 + x_1x_3 + x_3^2)(x_2^2 + x_2x_4 + x_4^2) = x_3^2x_4^2 \cdot \Phi_3(x_1/x_3) \cdot \Phi_3(x_2/x_4),\\ G &= (x_1^2 - x_1x_3 + x_3^2)(x_2^2 - x_2x_4 + x_4^2) = x_3^2x_4^2 \cdot \Phi_6(x_1/x_3) \cdot \Phi_6(x_2/x_4),\\ H &= (x_1x_4 + x_2x_3)(x_1x_2 + x_3x_4) = x_2x_3^2x_4 \cdot \Phi_2(x_1x_4/x_2x_3) \cdot \Phi_2(x_1x_2/x_3x_4),\end{aligned}$$

and likewise the proof for $k = 4$ is based on the identity $F - G = 4H$, where

$$\begin{aligned}F &= (x_1 + x_3)^2(x_2 + x_4)^2 = x_3^2x_4^2 \cdot \Phi_2(x_1/x_3)^2\Phi_2(x_2/x_4)^2,\\ G &= (x_1 - x_3)^2(x_2 - x_4)^2 = x_3^2x_4^2 \cdot \Phi_1(x_1/x_3)^2\Phi_1(x_2/x_4)^2,\\ H &= (x_1x_2 + x_3x_4)(x_1x_4 + x_2x_3) = x_2x_3^2x_4 \cdot \Phi_2(x_1x_2/x_3x_4)\Phi_2(x_1x_4/x_2x_3).\end{aligned}$$ □

Remark. It is appropriate to remark on how the identities in the above proof were found. Write

$$\Psi_m(X, Y) = Y^{\varphi(m)}\Phi_m(X/Y)$$

for the homogenization of the m-th cyclotomic polynomial. Now consider

$$f(x_1, x_2, x_3, x_4) = \Psi_m(u, v),$$

where u, v are monomials in variables x_1, x_2, x_3, x_4. Let ℓ be a prime. We see that evaluating any such f at $(\zeta^\alpha, \zeta^\beta, \zeta^\gamma, \zeta^\delta)$ gives an element of V_n (provided that it does not vanish). We considered products of such f of total degree up to 20 and picked out ones that are invariant under the 4-cycle (x_1, x_2, x_3, x_4), and searched for ternary relations between them. This yielded the identities used in the above proof.

Proof of Theorems 1 and 2 for $\ell = 5$. Theorems 1 and 2 for $\ell = 5$ follow immediately from Lemma 19. □

6. The S-unit equation over $\mathbb{Q}_{\infty,7}$

Lemma 20. *Let υ_7 be the unique prime of $\mathbb{Q}_{\infty,7}$ above 7, and write $S = \{\upsilon_7\}$. Then*

$$(\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}(\mathbb{Q}_{\infty,7}, S))$$

is infinite.

Proof. In view of the proof of Lemma 19, it would be natural to seek polynomials F, G, H in variables $x_1, \dots, x_6$ satisfying the following properties:

- $F \pm G = H$.
- F, G, H are invariant under the 6-cycle $(x_1, x_2, \dots, x_6)$.
- Each is a product of polynomials

  $$f(x_1, x_2, \dots, x_6) = \Psi_m(u, v),$$

  with u, v monomials in $x_1, \dots, x_6$.

Unfortunately, an extensive search has failed to produce any such triple of polynomials. We therefore need to proceed a little differently.

Let $a \in \mathbb{Z}_7$ be the element satisfying

$$a^2 + a + 1 = 0, \qquad a \equiv 2 \pmod 7;$$

such an element exists and is unique by Hensel's lemma. Let $\sigma, c : \Omega_{\infty,7} \to \Omega_{\infty,7}$ be the field automorphisms satisfying

$$\sigma(\zeta_{7^n}) = \zeta_{7^n}^a, \qquad c(\zeta_{7^n}) = \zeta_{7^n}^{-1}$$

for $n \geq 1$. Then $\mathbb{Q}_{\infty,7}$ is the field fixed by the subgroup of $\mathrm{Gal}(\Omega_{\infty,7}/\mathbb{Q})$ generated by σ and c. We work with polynomials in variables x_1, x_2, x_3. Let

$$\begin{aligned}
F &= (x_1x_2^2 + x_3^3)(x_2x_3^2 + x_1^3)(x_3x_1^2 + x_2^3),\\
G &= (x_1 - x_2)(x_2 - x_3)(x_3 - x_1)(x_1x_2 - x_3^2)(x_2x_3 - x_1^2)(x_3x_1 - x_2^2),\\
H &= (x_1^2x_2 + x_3^3)(x_2^2x_3 + x_1^3)(x_3^2x_1 + x_2^3).
\end{aligned}$$

These satisfy the identity $F - G = H$. Moreover, they are invariant under the 3-cycle (x_1, x_2, x_3) and all the factors are of the form $\Psi_m(u, v)$ where $m = 1$ or 2, and where u, v are suitable monomials in x_1, x_2, x_3. Evaluating any of F, G, H at $(\zeta, \zeta^a, \zeta^{a^2})$ yields an S-unit belonging to $\Omega_{n,7}^{\langle\sigma\rangle}$. Now we let

$$F' = \frac{F(x_1^2, x_2^2, x_3^2)}{x_1^6 x_2^6 x_3^6}, \quad G' = \frac{G(x_1^2, x_2^2, x_3^2)}{x_1^6 x_2^6 x_3^6}, \quad H' = \frac{H(x_1^2, x_2^2, x_3^2)}{x_1^6 x_2^6 x_3^6}.$$

Observe that the rational functions F', G', H' satisfy $F' - G' = H'$ and are moreover invariant under the 3-cycle (x_1, x_2, x_3). Moreover, F', G', H' evaluated at $(\zeta, \zeta^a, \zeta^{a^2})$ yield S-units belonging to $\Omega_{n,7}^{\langle\sigma\rangle}$. We need to check that these in fact belong to $\mathbb{Q}_{n-1,7} = \Omega_{n,7}^{\langle\sigma,c\rangle}$ and so we need to check that these expressions are invariant under c. This follows immediately on observing that F', G', H' may be rewritten as

$$\begin{aligned}
F' &= \left(\frac{x_1 x_2^2}{x_3^3} + \frac{x_3^3}{x_1 x_2^2}\right)\left(\frac{x_2 x_3^2}{x_1^3} + \frac{x_1^3}{x_2 x_3^2}\right)\left(\frac{x_3 x_1^2}{x_2^3} + \frac{x_2^3}{x_3 x_1^2}\right),\\
G' &= \left(\frac{x_1}{x_2} - \frac{x_2}{x_1}\right)\left(\frac{x_2}{x_3} - \frac{x_3}{x_2}\right)\left(\frac{x_3}{x_1} - \frac{x_1}{x_3}\right)\left(\frac{x_1 x_2}{x_3^2} - \frac{x_3^2}{x_1 x_2}\right)\left(\frac{x_2 x_3}{x_1^2} - \frac{x_1^2}{x_2 x_3}\right)\left(\frac{x_3 x_1}{x_2^2} - \frac{x_2^2}{x_3 x_1}\right),\\
H' &= \left(\frac{x_1^2 x_2}{x_3^3} + \frac{x_3^3}{x_1^2 x_2}\right)\left(\frac{x_2^2 x_3}{x_1^3} + \frac{x_1^3}{x_2^2 x_3}\right)\left(\frac{x_3^2 x_1}{x_2^3} + \frac{x_2^3}{x_3^3 x_1}\right).
\end{aligned}$$

Thus F', G', H' evaluated at $(\zeta, \zeta^a, \zeta^{a^2})$ yield elements of $\mathcal{O}(\mathbb{Q}_{\infty,7}, S)^\times$. We write

$$\varepsilon_n = \frac{F'(\zeta, \zeta^a, \zeta^{a^2})}{H'(\zeta, \zeta^a, \zeta^{a^2})}, \quad \delta_n = -\frac{G'(\zeta, \zeta^a, \zeta^{a^2})}{H'(\zeta, \zeta^a, \zeta^{a^2})}.$$

Then ε_n, δ_n belong to $\mathcal{O}(\mathbb{Q}_{\infty,7}, S)^\times$ and satisfy $\varepsilon_n + \delta_n = 1$. In fact it is straightforward to check that $\varepsilon_n \notin \langle \pm\zeta_{7^n}, V_{n-1}\rangle$, from which it follows that $\varepsilon_n \neq \varepsilon_m$ for $n > m$. The details are similar to those of the proof of Lemma 19 and we omit them. □

7. Isogeny classes of elliptic curves over $\mathbb{Q}_{\infty,\ell}$

The purpose of this section is to prove Theorem 5. Since isogenous elliptic curves share the same set of bad primes, the corresponding theorem over number fields is an immediate consequence of Shafarevich's theorem. However, as we intend to show in the following section, Shafarevich's theorem does not generalize to elliptic curves over $\mathbb{Q}_{\infty,\ell}$. We shall instead rely on a theorem of Kato to control $\mathbb{Q}_{\infty,\ell}$-points on certain modular Jacobians.

Our first lemma shows that there are only finitely many primes that can divide the degree of a cyclic isogeny of E.

Lemma 21. *Let ℓ be a prime and let $E/\mathbb{Q}_{\infty,\ell}$ be an elliptic curve without potential complex multiplication. Then there is a constant B, depending on E, such that for primes $p \geq B$, the elliptic curve E has no p-isogenies defined over $\mathbb{Q}_{\infty,\ell}$.*

Proof. Let n be the least positive integer such that E admits a model defined over $\mathbb{Q}_{n,\ell}$. By a famous theorem of Serre [1972], there is a constant B, depending on E, such that for $p \geq B$ the mod p

representation

$$\bar{\rho}_{E,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}_{n,\ell}) \to \mathrm{GL}_2(\mathbb{F}_p)$$

is surjective. We may suppose that $B \geq 5$. Thus, for $p \geq B$, the Galois group $\mathrm{Gal}(\mathbb{Q}_{n,\ell}(E[p])/\mathbb{Q}_{n,\ell})$ is isomorphic to $\mathrm{GL}_2(\mathbb{F}_p)$ which is nonsolvable. We will show that E has no p-isogeny defined over $\mathbb{Q}_{\infty,\ell}$. Suppose otherwise. Then such an isogeny is in fact defined over $\mathbb{Q}_{m,\ell}$ for some $m \geq n$. It follows that the extension $\mathbb{Q}_{m,\ell}(E[p])/\mathbb{Q}_{m,\ell}$ has Galois group isomorphic to a subgroup of a Borel subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$, with is solvable. As the extension $\mathbb{Q}_{m,\ell}/\mathbb{Q}_{n,\ell}$ is cyclic, we conclude that $\mathbb{Q}_{m,\ell}(E[p])/\mathbb{Q}_{n,\ell}$ is solvable. However, this contains the nonsolvable subextension $\mathbb{Q}_{n,\ell}(E[p])/\mathbb{Q}_{n,\ell}$, giving a contradiction. □

We shall make use of the following theorem of Kato [2004, Theorem 14.4] building on work of Rohrlich [1984].

Theorem 22 (Kato). *Let ℓ be a prime. Let A be an abelian variety defined over $\mathbb{Q}$ and admitting a surjective map $J_1(N) \to A$ for some $N \geq 1$. Then $A(\mathbb{Q}_{\infty,\ell})$ is finitely generated.*

Lemma 23. *Let p, ℓ be primes. Let E be an elliptic curve defined over $\mathbb{Q}_{\infty,\ell}$ without potential complex multiplication. Then, for m sufficiently large, E has no p^m-isogenies defined over $\mathbb{Q}_{\infty,\ell}$.*

Proof. Let r be the least positive integer such that the modular curve $X = X_0(p^r)$ has genus at least 2, and write $J = J_0(p^r)$ for the corresponding modular Jacobian. It follows from Kato's theorem that $J(\mathbb{Q}_{\infty,\ell})$ is finitely generated, and therefore that $J(\mathbb{Q}_{\infty,\ell}) = J(\mathbb{Q}_{n,\ell})$ for some $n \geq 1$. Consider the Abel–Jacobi map

$$X \hookrightarrow J, \quad P \mapsto [P - \infty]$$

where $\infty \in X(\mathbb{Q})$ denotes the infinity cusp. It follows from this embedding that $X(\mathbb{Q}_{\infty,\ell}) = X(\mathbb{Q}_{n,\ell})$. By Faltings' theorem, this set is finite.

Let $k = \#X(\mathbb{Q}_{\infty,\ell})$ and let $s = kr$. To prove the lemma we in fact show that E has no cyclic isogenies of degree p^s defined over $\mathbb{Q}_{\infty,\ell}$. Suppose otherwise, and let $\psi : E \to E'$ be a cyclic isogeny of degree p^s defined over $\mathbb{Q}_{\infty,\ell}$. Then, we may factor ψ into a sequence of cyclic isogenies defined over $\mathbb{Q}_{\infty,\ell}$

$$E = E_0 \xrightarrow{\psi_1} E_1 \xrightarrow{\psi_2} E_2 \cdots \xrightarrow{\psi_k} E_k = E',$$

where ψ_i is of degree p^r. Note that E_i and E_j are nonisomorphic over $\overline{\mathbb{Q}}$ for $i \neq j$; indeed they are related by a cyclic isogeny and E does not have potential complex multiplication. Thus the elliptic curves $E_0, E_1, \ldots, E_k$ support distinct $\mathbb{Q}_{\infty,\ell}$-points on $X = X_0(p^r)$. This contradicts the fact that $\#X(\mathbb{Q}_{\infty,\ell}) = k$. □

Remark. A famous theorem of Serre [1968, Section 2.1] asserts that the p-adic Tate module of a non-CM elliptic curve defined over a number field is irreducible. It is in fact possible to deduce Lemma 23 from Serre's theorem for $\ell \neq p$, but we have been unable to do this for $\ell = p$.

Proof of Theorem 5. Let E' belong to the $\mathbb{Q}_{\infty,\ell}$-isogeny class of E. Let $\psi : E \to E'$ be an isogeny defined over $\mathbb{Q}_{\infty,\ell}$. This has kernel of the form $\mathbb{Z}/a \times \mathbb{Z}/ab$ where a, b are positive integers, and so it can be factored into a composition

$$E \to E/E[a] \cong E \to E',$$

where the final morphism is cyclic of degree b. Thus to prove the proposition, it is enough to show that E has finitely many cyclic isogenies defined over $\mathbb{Q}_{\infty,\ell}$. The degree of any such isogeny is divisible by primes $p < B$ where B is as in Lemma 21. Also, for any $p < B$, we know the exponent of p in the degree of a cyclic isogeny $E \to E'$ is bounded by Lemma 23. Thus there are finitely many cyclic isogenies of E defined over $\mathbb{Q}_{\infty,\ell}$. □

8. From S-unit equations to elliptic curves

The aim of this section is to prove Theorem 4. We start by recalling a few facts about Legendre elliptic curves; see Proposition III.1.7 of [Silverman 1986] and its proof. Let K be a field of characteristic $\neq 2$ and let $\lambda \in (\mathbb{P}^1 - \{0, 1, \infty\})(K)$. Associated to λ is the Legendre elliptic curve

$$E_\lambda : Y^2 = X(X-1)(X-\lambda).$$

This model respectively has discriminant and j-invariant

$$\Delta = 16\lambda^2(1-\lambda)^2, \quad j = \frac{64(\lambda^2-\lambda+1)^3}{\lambda^2(1-\lambda)^2}. \tag{21}$$

Moreover, for $\lambda, \mu \in (\mathbb{P}^1 - \{0, 1, \infty\})(K)$, the Legendre elliptic curves E_λ and E_μ are isomorphic over K (or over $\overline{K}$) if and only if

$$\mu \in \left\{\lambda, \frac{1}{\lambda}, 1-\lambda, \frac{1}{1-\lambda}, \frac{\lambda}{\lambda-1}, \frac{\lambda-1}{\lambda}\right\}.$$

Now let K be a number field and S a finite set of nonarchimedean places. We let S' be the set of nonarchimedean places which are either in S or above 2. We let $\lambda \in (\mathbb{P}^1 - \{0, 1, \infty\})(\mathcal{O}(K, S))$. Then λ, $1-\lambda \in \mathcal{O}(K, S)^\times$. It follows from the expression for the discriminant that E_λ has good reduction away from S'.

Proof of Theorem 4. Let $\ell = 2, 3, 5$ or 7. Let S be given by (1) and let $S' = S \cup \{\upsilon_2\}$ as in the statement of Theorem 4. In proving Theorem 1 we constructed, for each positive integer n, elements ε_n, $\delta_n = 1 - \varepsilon_n$, belonging $\mathbb{Q}_{\infty,\ell} \cap V_n \subseteq \mathcal{O}(\mathbb{Q}_{\infty,\ell}, S)^\times$, and moreover verified, for $n \geq 2$, that $\varepsilon_n \notin \langle \zeta_{\ell^n}, V_{n-1} \rangle$. We let

$$E_n : Y^2 = X(X-1)(X-\varepsilon_n).$$

Then E_n is defined over $\mathbb{Q}_{\infty,\ell}$ and has good reduction away from S'. We claim, for $n > m$, that E_n and E_m are not isomorphic, even over $\overline{\mathbb{Q}}$. To see this, suppose E_n and E_m are isomorphic. Then ε_n equals one of $\varepsilon_m^{\pm 1}$, $\delta_m^{\pm 1}$, $(-\varepsilon_m \delta_m)^{\pm 1}$. This gives a contradiction as all of these belong to $\langle \pm\zeta_{\ell^n}, V_{n-1} \rangle$. This proves the claim.

It remains to show that the E_n form infinitely many isogeny classes over $\mathbb{Q}_{\infty,\ell}$. However, this immediately follows from Theorem 5 and the following lemma. □

Lemma 24. *For n sufficiently large, E_n does not have potential complex multiplication.*

Proof. Suppose E_n has potential complex multiplication by an order R in an imaginary quadratic field K. Write $j = j(E_n)$. We claim that $\mathbb{Q}(j)/\mathbb{Q}$ is a cyclic Galois extension of order ℓ^n for some n. Note that $\mathbb{Q}(j)$ is a subextension of $\mathbb{Q}_{\infty,\ell}$ of finite degree, and is thus contained in $\mathbb{Q}_{m,\ell}$ for some m. Hence $\mathbb{Q}(j)$ is the fixed field of some subgroup H (say) of $G = \mathrm{Gal}(\mathbb{Q}_{m,\ell}/\mathbb{Q})$. As G is cyclic, the group H is a normal subgroup, and therefore $\mathbb{Q}(j)/\mathbb{Q}$ is a Galois extension. Moreover, $\mathrm{Gal}(\mathbb{Q}(j)/\mathbb{Q}) \cong G/H$ which is cyclic of order ℓ^n for some n, proving our claim.

By standard CM theory [Shimura 1971, Theorem 5.7], we know that $\mathrm{Gal}(K(j)/K) \cong \mathrm{Pic}(R)$ and $[\mathbb{Q}(j) : \mathbb{Q}] = [K(j) : K]$. Since in our case $\mathbb{Q}(j)/\mathbb{Q}$ is Galois, $\mathrm{Gal}(\mathbb{Q}(j)/\mathbb{Q}) \cong \mathrm{Gal}(K(j)/K) \cong \mathrm{Pic}(R)$. However, $\mathbb{Q}(j) \subset \mathbb{Q}_{\infty,\ell}$ is totally real. It follows [Shimura 1971, page 124] that $\mathrm{Pic}(R)$ is an elementary abelian 2-group. However $\mathbb{Q}(j)/\mathbb{Q}$ is cyclic of order ℓ^n. Thus, $j \in \mathbb{Q}$ if $\ell \neq 2$, and $j \in \mathbb{Q}_{1,2} = \mathbb{Q}(\sqrt{2})$ if $\ell = 2$. However, from the expression for j in (21) we know that $[\mathbb{Q}(\varepsilon_n) : \mathbb{Q}(j)] \leq 6$. Thus ε_n belongs to a subfield of $\mathbb{Q}_{\infty,\ell}$ of degree at most 12. The lemma follows since, by Siegel's theorem, the S-unit equation has only finitely many solutions in any number field. □

9. Hyperelliptic curves over $\mathbb{Q}_{\infty,\ell}$ with few bad primes

Let ℓ be an odd prime. Let $g \geq 2$ be an integer satisfying

$$\begin{cases} g \equiv (\ell-3)/4 \text{ or } -1 \pmod{(\ell-1)/2} & \text{if } \ell \equiv 3 \pmod 4, \\ g \equiv -1 \pmod{(\ell-1)/4} & \text{if } \ell \equiv 1 \pmod 4. \end{cases} \tag{22}$$

Then there is a positive integer k such that

$$k \cdot \left(\frac{\ell-1}{2}\right) = \begin{cases} 2g+1 \text{ or } 2g+2 & \text{if } \ell \equiv 3 \pmod 4, \\ 2g+2 & \text{if } \ell \equiv 1 \pmod 4. \end{cases} \tag{23}$$

Let $n \geq 2$ be a positive integer satisfying

$$\ell^{n-1} \geq k. \tag{24}$$

In this section we construct a hyperelliptic D_n curve of genus g defined over $\mathbb{Q}_{n-1,\ell}$ with good reduction away from the primes above 2, ℓ.

Write

$$\mathcal{Z}_n = \{\zeta \in \Omega_{n,\ell} : \zeta^{\ell^n} = 1, \zeta^{\ell^i} \neq 1 \text{ if } i < n\}$$

for the set of primitive ℓ^n-th roots of 1. Write

$$\mathcal{Z}_n^+ = \{\zeta + \zeta^{-1} : \zeta \in \mathcal{Z}_n\} \subset \Omega_{n,\ell}^+.$$

We note that any element of $\mathcal{Z}_n^+$ generates $\Omega_{n,\ell}^+$.

Lemma 25. $$\#\mathcal{Z}_n^+ = \tfrac{1}{2}\ell^{n-1}(\ell-1).$$

Proof. We note that $\#\mathcal{Z}_n = \varphi(\ell^n) = \ell^{n-1}(\ell-1)$. Suppose $\alpha, \beta \in \mathcal{Z}_n$. Then

$$(\alpha+\alpha^{-1}) - (\beta+\beta^{-1}) = \alpha^{-1}\cdot(1-\alpha\beta)\cdot(1-\alpha\beta^{-1}). \tag{25}$$

Thus $\alpha+\alpha^{-1} = \beta+\beta^{-1}$ if and only if $\alpha=\beta$ or $\alpha=\beta^{-1}$. The lemma follows. □

Write

$$G_n = \mathrm{Gal}(\Omega^+_{n,\ell}/\mathbb{Q}_{n-1,\ell}), \qquad H_n = \mathrm{Gal}(\Omega^+_{n,\ell}/\Omega^+_{n-1,\ell}).$$

We note that these are both cyclic subgroups of $\mathrm{Gal}(\Omega^+_{n,\ell}/\mathbb{Q})$ having orders

$$\#G_n = (\ell-1)/2, \quad \#H_n = \ell.$$

Lemma 26. *Fix $\zeta \in \mathcal{Z}_n$. Let*

$$\eta_i = \zeta^{1+\ell^{n-1}(i-1)} + \zeta^{-1-\ell^{n-1}(i-1)}, \quad 1 \le i \le \ell. \tag{26}$$

Then $\eta_1,\dots,\eta_\ell \in \mathcal{Z}_n^+$ form a single orbit under the action of H_n, but have pairwise disjoint orbits under the action of G_n.

Proof. Let $\kappa \in \mathrm{Gal}(\Omega_{n,\ell}/\mathbb{Q})$ be given by $\kappa(\zeta) = \zeta^{1+\ell^{n-1}}$. We note that κ has order ℓ and fixes $\Omega_{n-1,\ell}$. We denote the restriction of κ to $\Omega^+_{n,\ell}$ by τ; this is a cyclic generator of H_n. Note that

$$\eta_i = \tau^{i-1}(\zeta+\zeta^{-1}), \quad 1 \le i \le \ell.$$

Let $\sigma_1, \sigma_2 \in G_n$. Let $1 \le i < j \le \ell$ and suppose $\sigma_1(\eta_i) = \sigma_2(\eta_j)$. Thus $\sigma_1\tau^{i-1}(\eta_1) = \sigma_2\tau^{j-1}(\eta_1)$, so $\tau^{1-j}\sigma_2^{-1}\sigma_1\tau^{i-1}$ fixes η_1. As η_1 generates $\Omega^+_{n,\ell}$, we have $\tau^{1-j}\sigma_2^{-1}\sigma_1\tau^{i-1} = 1$ is the identity element in $\mathrm{Gal}(\Omega^+_{n,\ell}/\mathbb{Q})$. However, $\mathrm{Gal}(\Omega^+_{n,\ell}/\mathbb{Q})$ is abelian, so

$$\tau^{i-j} = \sigma_1^{-1}\sigma_2 \in G_n \cap H_n = \{1\}.$$

Since $1 \le i \le j \le \ell$ and τ has order ℓ we have $i = j$. □

The Galois group G_n acts faithfully on $\mathcal{Z}_n^+$. This action has ℓ^{n-1} orbits. Assumption (24) ensures that the number of orbits is at least k. If $k > \ell$, then we *extend* the list $\eta_1,\dots,\eta_\ell \in \mathcal{Z}_n^+$ to $\eta_1,\dots,\eta_k \in \mathcal{Z}_n^+$, so that the η_i continue to have disjoint orbits under the action of G_n; if $\ell = 3$ the choice of η_4 will be important later, and we choose $\eta_4 = \zeta^2+\zeta^{-2}$. Consider the curve

$$D_n : Y^2 = \prod_{j=1}^{k} \prod_{\sigma\in G_n} (X - \eta_j^\sigma). \tag{27}$$

Lemma 27. *The curve D_n is hyperelliptic of genus g, is defined over $\mathbb{Q}_{n-1,\ell}$, and has good reduction away from the primes above 2 and ℓ.*

Proof. Our assumption on the orbits ensures that the polynomial on the right hand-side of (27) is separable. By (23), the degree of the polynomial is either $2g+1$ or $2g+2$. Thus D_n is a hyperelliptic curve of genus g. A priori, D_n is defined over $\Omega^+_{n,\ell}$. However, the roots of the hyperelliptic polynomial are permuted by the action of $G_n = \mathrm{Gal}(\Omega^+_{n,\ell}/\mathbb{Q}_{n-1,\ell})$ and so the polynomial belongs to $\mathbb{Q}_{n-1,\ell}[X]$. Hence D_n is defined over $\mathbb{Q}_{n-1,\ell}$.

Let $u_1, \dots, u_d$ be the roots of the hyperelliptic polynomial. Then the discriminant of hyperelliptic polynomial is

$$\prod_{1 \le i < j \le d} (u_i - u_j)^2.$$

However, u_i, u_j are distinct elements of $\mathcal{Z}_n^+$. Thus there are α, $\beta \in \mathcal{Z}_n$ with $\alpha \neq \beta$, β^{-1} such that $u_i = \alpha + \alpha^{-1}$, $u_j = \beta + \beta^{-1}$. From the identity (25),

$$u_i - u_j = \alpha^{-1}(1 - \alpha\beta^{-1})(1 - \alpha\beta).$$

Since $\alpha\beta$ and $\alpha\beta^{-1}$ are nontrivial ℓ-power roots of 1, we see that $u_i - u_j$ is a $\{\upsilon_\ell\}$-unit, and hence the discriminant of the hyperelliptic polynomial of D_n is a $\{\upsilon_\ell\}$-unit. □

Given four pairwise distinct elements z_1, z_2, z_3, z_4 of a field K, we shall employ the notation $(z_1, z_2; z_3, z_4)$ to denote the *cross ratio*

$$(z_1, z_2; z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}.$$

We extend the cross ratio to four distinct elements z_1, z_2, z_3, z_4 of $\mathbb{P}^1(K)$ in the usual way. We let $\mathrm{GL}_2(K)$ act on $\mathbb{P}^1(K)$ via fractional linear transformations

$$\gamma(z) = \frac{az + b}{cz + d}, \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It is well-known and easy to check that these fractional linear transformations leave the cross ratio unchanged:

$$(\gamma(z_1), \gamma(z_2); \gamma(z_3), \gamma(z_4)) = (z_1, z_2; z_3, z_4).$$

Lemma 28. *Let $\overline{K}$ be an algebraically closed field of characteristic* 0. *Let*

$$D : Y^2 = \prod_{i=1}^{d} (X - a_i), \qquad D' : Y^2 = \prod_{i=1}^{d} (X - b_i)$$

be genus-g curves defined over $\overline{K}$ where the polynomials on the right are separable. If D, D' are isomorphic then there is some permutation $\mu \in S_d$ such that for all quadruples of pairwise distinct indices $1 \le r, s, t, u \le d$

$$(a_r, a_s; a_t, a_u) = (b_{\mu(r)}, b_{\mu(s)}; b_{\mu(t)}, b_{\mu(u)}).$$

Proof. We shall make use of the following standard description (e.g., [Baker et al. 2005, Proposition 6.11]) of isomorphisms of hyperelliptic curves: every isomorphism $\pi : D \to D'$ is of the form

$$\pi(X, Y) = \left(\frac{aX + b}{cX + d}, \frac{eY}{(cX + d)^{g+1}} \right)$$

for some

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\bar{K}), \quad e \in \bar{K}^\times.$$

Observe that $\pi(a_i, 0)$ has Y-coordinate 0; thus

$$\{\gamma(a_1), \dots, \gamma(a_d)\} = \{b_1, \dots, b_d\}.$$

Hence there is a permutation $\mu \in S_d$ such that $\gamma(a_i) = b_{\mu(i)}$. The lemma follows from the invariance of the cross ratio under the action of $\mathrm{GL}_2(\bar{K})$. □

Lemma 29. *Let $\ell \geq 11$ be prime. Then there is some $a \in \mathbb{Z}_\ell^\times$ of order $\ell - 1$ such that*

$$1+a^2 \not\equiv 0, \pm(1-a^2), \pm(a+a^3), \pm(a-a^3), \pm(1+a^3), \pm(1-a^3), \pm(a+a^2), \pm(a-a^2) \pmod{\ell}. \quad (28)$$

Proof. Making use of the fact that a polynomial of degree n has at most n roots, we see that the number of $a \in \mathbb{F}_\ell$ that *do not satisfy* (28) is (very crudely) bounded by 37. An element $a \in \mathbb{Z}_\ell^\times$ of order $\ell - 1$ is the unique Hensel lift of an element $a \in \mathbb{F}_\ell^\times$ of order $\ell - 1$. There are precisely $\varphi(\ell - 1)$ elements of order $\ell - 1$ in $\mathbb{F}_\ell^\times$. A theorem of Shapiro [1943, page 23], asserts that $\varphi(n) > n^{\log 2/\log 3}$ for $n \geq 30$. We note that if $\ell \geq 317$ then $\varphi(\ell - 1) \geq 316^{\log 2/\log 3} \approx 37.8$, and so the lemma holds for $\ell \geq 317$. For the range $11 \leq \ell \leq 317$ we checked the lemma by brute force computer enumeration. □

Lemma 30. *Let $n > m$ be sufficiently large. Then D_n and D_m are nonisomorphic, even over $\bar{\mathbb{Q}}$.*

Proof. Note that all roots of the hyperelliptic polynomial for D_n in (27) belong to $\mathcal{Z}_n^+$. It follows from (25) that the cross ratio of any four of them belongs to V_n. Suppose D_n and D_m are isomorphic. Let u_1, u_2, u_3, u_4 be any distinct roots of the hyperelliptic polynomial for D_n given in (27). Then, by Lemma 28,

$$(u_1, u_2; u_3, u_4) \in V_m \subseteq V_{n-1}.$$

We shall obtain a contradiction through a careful choice of the four roots $u_1, \dots, u_4$.

We first suppose that $k \geq 2$ and $\ell \geq 5$. Let $\zeta = \zeta_{\ell^n}$ and $b = 1 + \ell^{n-1}$. Then, by Lemma 26, $\eta_1 = \zeta + \zeta^{-1}$ and $\eta_2 = \zeta^b + \zeta^{-b}$. Let $a \in \mathbb{Z}_\ell^\times$ have order $\ell - 1$. Let $\kappa \in \mathrm{Gal}(\Omega_{n,\ell}/\mathbb{Q}_{n-1,\ell})$ be given by $\kappa(\zeta) = \zeta^a$. Then κ is a cyclic generator for $\mathrm{Gal}(\Omega_{n,\ell}/\mathbb{Q}_{n-1,\ell})$. We shall denote the restriction of κ to $\Omega_{n,\ell}^+$ by μ. Then μ is a cyclic generator for $G_n = \mathrm{Gal}(\Omega_{n,\ell}^+/\mathbb{Q}_{n-1,\ell})$ having order $(\ell-1)/2$. We shall take

$$u_1 = \eta_1 = \zeta + \zeta^{-1}, \quad u_2 = \mu(\eta_1) = \zeta^a + \zeta^{-a}, \quad u_3 = \eta_2 = \zeta^b + \zeta^{-b}, \quad u_4 = \mu(\eta_2) = \zeta^{ab} + \zeta^{-ab}.$$

We compute the cross ratio with the help of identity (25), finding

$$(u_1, u_2; u_3, u_4) = \frac{(1-\zeta^{1+b})(1-\zeta^{1-b})(1-\zeta^{a+ab})(1-\zeta^{a-ab})}{(1-\zeta^{1+ab})(1-\zeta^{1-ab})(1-\zeta^{a+b})(1-\zeta^{a-b})}.$$

As $b \equiv 1 \pmod{\ell}$, and clearly $a \not\equiv \pm 1 \pmod{\ell}$, it is easy to check that $1 + b$ is the only one out of the eight exponents of ζ above that is $\pm 2 \pmod{\ell}$. Therefore by Lemma 11, the cross ratio is not an element of $\langle \pm\zeta_{\ell^n}, V_{n-1}\rangle$ for n sufficiently large, giving a contradiction for the case $k \geq 2$ and $\ell \geq 5$.

Next we suppose that $k = 1$. It follows from (23) that $\ell \geq 11$. We choose $a \in \mathbb{Z}_\ell^\times$ as in Lemma 29, and, as above, take μ to be the corresponding generator of G_n of order $(\ell - 1)/2 \geq 5$. We take

$$u_i = \mu^{i-1}(\eta_1) = \zeta^{a^{i-1}} + \zeta^{-a^{i-1}}, \quad 1 \leq i \leq 4;$$

observe that these are four roots of the hyperelliptic polynomial of D_n given in (27). The assumption that $\ell \geq 11$ ensures that a has order ≥ 10 and so u_1, u_2, u_3, u_4 are indeed pairwise distinct. We compute the cross ratio with the help of identity (25), finding

$$(u_1, u_2; u_3, u_4) = \frac{(1 - \zeta^{1+a^2})(1 - \zeta^{1-a^2})(1 - \zeta^{a+a^3})(1 - \zeta^{a-a^3})}{(1 - \zeta^{1+a^3})(1 - \zeta^{1-a^3})(1 - \zeta^{a+a^2})(1 - \zeta^{a-a^2})}.$$

Using Lemma 10 and our choice of a given by Lemma 29 we conclude that this cross ratio does not belong to $\langle \pm\zeta_{\ell^n}, V_{n-1} \rangle$ for n sufficiently large. This gives a contradiction for the case $k = 1$.

Finally, we consider $\ell = 3$. It follows from (23) that $k \geq 5$. Recall our choices of η_1, η_2, η_3 in Lemma 26, and our choice of $\eta_4 = \zeta^2 + \zeta^{-2}$ in the particular case $\ell = 3$. We choose the four roots $u_i = \eta_i$ for $i = 1, \dots, 4$, and obtain

$$(u_1, u_2; u_3, u_4) = \frac{(1 - \zeta^{2+2\times 3^{n-1}})(1 - \zeta^{-2\times 3^{n-1}})(1 - \zeta^{3+3^{n-1}})(1 - \zeta^{-1+3^{n-1}})}{(1 - \zeta^3)(1 - \zeta^{-1})(1 - \zeta^2)(1 - \zeta^{-3^{n-1}})}.$$

As before, with the help of Lemma 11, we easily verify that the cross ratio is not an element of $\langle \pm\zeta_{\ell^n}, V_{n-1} \rangle$ for n sufficiently large. This completes the proof. □

Proof of Theorem 6. If $\ell = 3$ or 5 then (22) does not impose any restriction on the genus. Therefore we obtain, as above, for every genus $g \geq 2$, infinitely many $\overline{\mathbb{Q}}$-isomorphism classes of genus-g hyperelliptic curves, defined over $\mathbb{Q}_{\infty,\ell}$, with good reduction away from $\{\upsilon_2, \upsilon_\ell\}$.

It remains to deal with $\ell = 7$, 11 and 13. Here, (22) imposes the restriction

$$g \equiv \begin{cases} 1 \text{ or } 2 \bmod 3 & \text{if } \ell = 7, \\ 2 \text{ or } 4 \bmod 5 & \text{if } \ell = 11, \\ 2 \bmod 3 & \text{if } \ell = 13. \end{cases}$$

We very briefly sketch how to remove the restriction. Instead of D_n defined as in (27), we consider the more general

$$D_n : Y^2 = h(X) \cdot \prod_{j=1}^{k} \prod_{\sigma \in G_n} (X - \eta_j^\sigma)$$

where

- h is a monic divisor of $X(X - 1)(X + 1)$;
- k and h are chosen to obtain the desired genus;
- $\eta_j \in \mathcal{Z}_n^+$ are chosen as before.

These D_n are clearly defined over $\mathbb{Q}_{n-1,\ell}$. To check that they have good reduction away from $S' = \{\upsilon_2, \upsilon_\ell\}$, we need to verify that the difference of any two distinct roots u, v of the hyperelliptic polynomial belongs to $\mathcal{O}(\Omega_n, S')^\times$. The proof of Lemma 27 shows this if $u, v \in \mathcal{Z}_n^+$. For the remaining possible differences it is enough to note that

$$\alpha + \alpha^{-1} = \alpha^{-1}\Phi_4(\alpha), \quad \alpha + \alpha^{-1} + 1 = \alpha^{-1}\Phi_3(\alpha), \quad \alpha + \alpha^{-1} - 1 = \alpha^{-1}\Phi_6(\alpha),$$

which are all units by Lemma 13. We omit the remaining details. □

10. Isogeny classes of hyperelliptic curves over $\mathbb{Q}_{\infty,\ell}$

A beautiful theorem of Kummer asserts that the index of the cyclotomic units C_n in the full unit group $\mathcal{O}(\Omega_{n,\ell})^\times$ equals the class number h_n^+ of $\Omega_{n,\ell}^+$. In this section, with the help of Kummer's theorem, we prove for certain primes ℓ the existence of infinitely many isogeny classes of hyperelliptic Jacobians over $\mathbb{Q}_{\infty,\ell}$ with good reduction away from ℓ. We first prove a few elementary lemmas.

Lemma 31. *Let K be a field of characteristic not 2, and let $L = K(\sqrt{\alpha_1}, \dots, \sqrt{\alpha_r})$, where $\alpha_i \in K^\times$. Then for any $x \in K$ such that $\sqrt{x} \in L$, we have*

$$x = \alpha_1^{e_1} \cdots \alpha_r^{e_r} q^2$$

for some integers $e_i \in \mathbb{Z}$ and $q \in K$.

Proof. Let M be a field of characteristic not 2, and let $d \in M$ be a nonsquare. Let $x \in M$ and suppose $\sqrt{x} \in M(\sqrt{d})$. Then $\sqrt{x} = y + z\sqrt{d}$ for some $y, z \in M$. Squaring, we deduce that $yz = 0$. Thus $x = y^2$ or $x = dz^2$.

We now prove the lemma by induction on r. The above establishes the case $r = 1$. Let $r \geq 2$, and let $x \in K$ satisfy $\sqrt{x} \in L$. Letting $M = K(\sqrt{\alpha_1}, \dots, \sqrt{\alpha_{r-1}})$ we see that $x \in M$ and $\sqrt{x} \in M(\sqrt{\alpha_r})$. Thus, by the above, $\sqrt{x} \in M$ or $\sqrt{x\alpha_r} \in M$. In other words,

$$\sqrt{x \cdot \alpha_r^e} \in M = K(\sqrt{\alpha_1}, \dots, \sqrt{\alpha_{r-1}})$$

for some $e \in \{0, 1\}$. By the inductive hypothesis, there are $e_1, \dots, e_{r-1} \in \mathbb{Z}$ and $q \in K$ such that

$$x \cdot \alpha_r^e = \alpha_1^{e_1} \cdots \alpha_{r-1}^{e_{r-1}} q^2.$$

The proof is complete on taking $e_r = -e$. □

Lemma 32. *Let ℓ be an odd prime. Let $q \in \Omega_{\infty,\ell}$ satisfy $q^2 \in V_n$. If the class number h_n^+ of $\Omega_{n,\ell}^+$ is odd, then $q \in V_n$.*

Proof. Let $q \in \Omega_{\infty,\ell}$ satisfy $q^2 \in V_n \subset \Omega_{n,\ell}$. As the extension $\Omega_{\infty,\ell}/\Omega_{n,\ell}$ is pro-ℓ, we conclude that $q \in \Omega_{n,\ell}$. However, $V_n \subseteq \mathcal{O}(\Omega_{n,\ell}, \{\upsilon_\ell\})^\times$, where, as usual, υ_ℓ denotes the prime above ℓ. Thus $q \in \mathcal{O}(\Omega_{n,\ell}, \{\upsilon_\ell\})^\times$. We claim that

$$[\mathcal{O}(\Omega_{n,\ell}, \{\upsilon_\ell\})^\times : V_n] = h_n^+.$$

The lemma follows immediately from the claim. To prove the claim, consider the commutative diagram with exact rows

$$\begin{tikzcd}
1 \arrow[r] & C_n \arrow[r] \arrow[d, hook] & V_n \arrow[r, "\kappa"] \arrow[d, hook] & \mathbb{Z} \arrow[r] \arrow[d] & 1 \\
1 \arrow[r] & \mathcal{O}(\Omega_{n,\ell})^\times \arrow[r] & \mathcal{O}(\Omega_{n,\ell}, \{\upsilon_\ell\})^\times \arrow[r, "\kappa"] & \mathbb{Z} \arrow[r] & 1
\end{tikzcd}$$

where $\kappa(\alpha) = \mathrm{ord}_{(1-\zeta)}(\alpha)$. By the snake lemma,

$$\mathcal{O}(\Omega_{n,\ell}, \{\upsilon_\ell\})^\times / V_n \cong \mathcal{O}(\Omega_{n,\ell})^\times / C_n.$$

Write $C_n^+ = C_n \cap \Omega_{n,\ell}^+$. The aforementioned theorem of Kummer asserts that

$$[\mathcal{O}(\Omega_{n,\ell})^\times : C_n] = [\mathcal{O}(\Omega_{n,\ell}^+)^\times : C_n^+] = h_n^+;$$

see, for example, [Washington 1997, Exercise 8.5] for the first equality, and [loc. cit., Theorem 8.2] for the second. This proves the claim. □

Lemma 33. *Let K be a field of characteristic $\neq 2$. Let $f \in K[X]$ be a monic separable polynomial of odd degree $d \geq 5$. Write $f = \prod_{i=1}^d (X - \alpha_i)$ with $\alpha_i \in \overline{K}$. Let C/K be a hyperelliptic curve given by $Y^2 = f(X)$ with Jacobian J. Then*

$$K(J[2]) = K(\alpha_1, \ldots, \alpha_d), \qquad K(J[4]) = K(J[2])(\{\sqrt{\alpha_i - \alpha_j}\}_{1 \leq i,j \leq d}).$$

Proof. Write ∞ for the point at infinity on the given model for C. The expression given for $K(J[2])$ is well-known; it may be seen by observing (see, for example [Schaefer 1995]) that the classes of the classes of degree 0 divisors $[(\alpha_i, 0) - \infty]$ with $i = 1, \ldots, d$ generate $J[2]$.

Yelton [2015, Theorem 1.2.2] gives a high-powered proof of the given expression for $K(J[4])$. For the convenience of the reader we give a more elementary argument. Let $L = K(J[2])$. The theory of 2-descent on hyperelliptic Jacobians furnishes, for any field $M \supseteq L$, an injective homomorphism [Schaefer 1995; Stoll 2001]

$$J(M)/2J(M) \hookrightarrow \prod_{i=1}^{d} M^*/(M^*)^2$$

known as the $X - \Theta$-map. This in particular sends the 2-torsion point $[(\alpha_i, 0) - \infty]$ to

$$\left((\alpha_i - \alpha_1), \ldots, (\alpha_i - \alpha_{i-1}), \prod_{j \neq i} (\alpha_i - \alpha_j), (\alpha_i - \alpha_{i+1}), \ldots, (\alpha_i - \alpha_d)\right).$$

The field $K(J[4])$ is the smallest extension of M of L such that all the images of the 2-torsion generators $[(\alpha_i, 0) - \infty]$ are trivial in $\prod_{i=1}^d M^*/(M^*)^2$. This is plainly the extension

$$M = L(\{\sqrt{\alpha_i - \alpha_j}\}_{1 \leq i,j \leq d}). \qquad □$$

Lemma 34. *Let p be a prime for which 2 is a primitive root (i.e., 2 is a generator for $\mathbb{F}_p^\times$). Let G be a cyclic group of order p, and let V be an $\mathbb{F}_2[G]$-module with $\dim_{\mathbb{F}_2}(V) = p-1$. Suppose that the action of G on $V - \{0\}$ is free. Then V is irreducible.*

Proof. Let W be a $\mathbb{F}_2[G]$-submodule of V, and write $d = \dim_{\mathbb{F}_2}(W)$. Since the action of G on $V - \{0\}$ is free, the set $W - \{0\}$ consists of G-orbits, all having size p. However, $\#(W - \{0\}) = 2^d - 1$, and so $p \mid (2^d - 1)$. By assumption, 2 is a primitive root modulo p, therefore $(p-1) \mid d$. Since W is an $\mathbb{F}_2$-subspace of V, which has dimension $p-1$, we see that $W = 0$ or $W = V$. □

Lemma 35. *Let $\ell = 2p+1$, where ℓ and p are odd primes. Suppose 2 is a primitive root modulo p. Let $g = (\ell-3)/4$. Let $n \ge 2$ and let $D_n/\mathbb{Q}_{n-1,\ell}$ be the hyperelliptic curve defined in Section 9. Let $A/\mathbb{Q}_{\infty,\ell}$ be an abelian variety and let $\phi : J(D_n) \to A$ be an isogeny defined over $\mathbb{Q}_{\infty,\ell}$. Then $\phi = 2^r \phi_{\text{odd}}$ where $\phi_{\text{odd}} : J(D_n) \to A$ is an isogeny of odd degree.*

We remark if ℓ and p are primes with $\ell = 2p+1$ then p is called a Sophie Germain prime, and ℓ is called as safe prime.

Proof of Lemma 35. Note that, in the notation of Section 9, $k = 1$, and the hyperelliptic polynomial for D_n has odd degree $2g+1 = (\ell-1)/2 = p$, and consists of a single orbit under action of $G_n = \mathrm{Gal}(\Omega_n^+/\mathbb{Q}_{n-1,\ell})$:

$$D_n : y^2 = \prod_{\sigma \in G_n} (X - \eta_1^\sigma), \qquad \eta_1 = \zeta_{\ell^n} + \zeta_{\ell^n}^{-1}.$$

In particular, the hyperelliptic polynomial is irreducible over $\mathbb{Q}_{\infty,\ell}$. It follows from this (e.g., [Stoll 2001, Lemma 4.3]) that $J(\mathbb{Q}_{\infty,\ell})[2] = 0$, where J denotes $J(D_n)$ for convenience. We note, by Lemma 33, that $\mathbb{Q}_{\infty,\ell}(J[2]) = \mathbb{Q}_{\infty,\ell}(\eta_1) = \Omega_{\infty,\ell}^+$. We consider the action of $G_\infty := \mathrm{Gal}(\Omega_{\infty,\ell}^+/\mathbb{Q}_{\infty,\ell})$ on $J[2]$. The group G_∞ is cyclic of order $(\ell-1)/2 = p$. Any element fixed by this action belongs to $J(\mathbb{Q}_{\infty,\ell})[2] = 0$. Thus G_∞ acts freely on $V - \{0\}$, where $V := J[2]$.

Now let $\phi : J \to A$ be an isogeny defined over $\mathbb{Q}_{\infty,\ell}$. Then $W := \ker(\phi) \cap J[2]$ is a subgroup of V stable under the action of G_∞, and therefore an $\mathbb{F}_2[G_\infty]$-submodule of the $\mathbb{F}_2[G_\infty]$-module V. Observe that $\dim_{\mathbb{F}_2}(V) = 2g = p-1$. By hypothesis, 2 is a primitive root modulo p. We apply Lemma 34 to deduce that $W = 0$ or $W = V$. Therefore, either ϕ already has odd degree, or $J[2] \subseteq \ker(\phi)$. In the latter case, observe that $\phi = 2\phi'$ where $\phi' : J \to A$ is an isogeny defined over $\mathbb{Q}_{\infty,\ell}$ of degree $\deg(\phi)/2^{2g}$. As ϕ has finite degree, by repeating the argument we eventually arrive at $\phi = 2^r \phi_{\text{odd}}$. □

Lemma 36. *Let $\ell = 2p+1$, where ℓ and p are odd primes. Suppose 2 is a primitive root modulo p. Suppose that the class number h_n^+ of $\Omega_{n,\ell}^+$ is odd for all n. Let $g = (\ell-3)/4$. For $n \ge 2$ let $D_n/\mathbb{Q}_{n-1,\ell}$ be the genus-g hyperelliptic curve defined in Section 9. Let $n > m$ be sufficiently large. Then there are no isogenies $J(D_n) \to J(D_m)$ defined over $\mathbb{Q}_{\infty,\ell}$.*

The assumption that h_n^+ is odd for all n may seem at first sight very restrictive. However, it is conjectured [Buhler et al. 2004] that $h_{n+1}^+ = h_n^+$ for all but finitely many pairs (ℓ, n). Moreover, Washington [1978] has shown that $\mathrm{ord}_p(h_n)$ remains bounded as $n \to \infty$, for any fixed prime p.

Proof of Lemma 36. Write J_n for $J(D_n)$. Suppose there is an isogeny $\phi : J_n \to J_m$ defined over $\mathbb{Q}_{\infty,\ell}$. By Lemma 35 we may suppose that ϕ has odd degree, and so $\ker(\phi) \cap J_n[4] = 0$. Thus ϕ restricted to $J_n[4]$ induces an isomorphism of $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}_{\infty,\ell})$-modules $J_n[4] \cong J_m[4]$. In particular, $\mathbb{Q}_{\infty,\ell}(J_n[4]) = \mathbb{Q}_{\infty,\ell}(J_m[4])$. As in the proof of Lemma 35 we have $\mathbb{Q}_{\infty,\ell}(J_n[2]) = \mathbb{Q}_{\infty,\ell}(J_m[2]) = \Omega^{+}_{\infty,\ell}$. Thus, by Lemma 33, the equality $\mathbb{Q}_{\infty,\ell}(J_n[4]) = \mathbb{Q}_{\infty,\ell}(J_m[4])$ may be rewritten as

$$\Omega^{+}_{\infty,\ell}(\{\sqrt{\vartheta_{n,i} - \vartheta_{n,j}}\}_{1 \le i,j \le (\ell-1)/2}) = \Omega^{+}_{\infty,\ell}(\{\sqrt{\vartheta_{m,i} - \vartheta_{m,j}}\}_{1 \le i,j \le (\ell-1)/2}),$$

where $\vartheta_{r,i} := \mu_r^{i-1}(\zeta_{\ell^r} + \zeta_{\ell^r}^{-1})$ where μ_r is a cyclic generator of G_r. This, in particular, implies that

$$\sqrt{\vartheta_{n,2} - \vartheta_{n,1}} \in \Omega^{+}_{\infty,\ell}(\{\sqrt{\vartheta_{m,i} - \vartheta_{m,j}}\}_{1 \le i,j \le (\ell-1)/2}).$$

We apply Lemma 31 to obtain

$$\vartheta_{n,2} - \vartheta_{n,1} = \pm \prod_{1 \le i < j \le (\ell-1)/2} (\vartheta_{m,i} - \vartheta_{m,j})^{e_{i,j}} \cdot q^2$$

for some integers $e_{i,j} \in \mathbb{Z}$ and $q \in \Omega^{+}_{\infty,\ell}$. By Lemma 32, we have $q \in V_n$. The generator μ_n of G_n is given by $\mu_n(\zeta_{\ell^n} + \zeta_{\ell^n}^{-1}) = \zeta_{\ell^n}^{a} + \zeta_{\ell^n}^{-a}$ where $a \in \mathbb{Z}_\ell^\times$ has order $(\ell - 1)$. Note

$$\vartheta_{n,2} - \vartheta_{n,1} = \zeta_{\ell^n}^{a} + \zeta_{\ell^n}^{-a} - \zeta_{\ell^n} - \zeta_{\ell^n}^{-1} = \zeta_{\ell^n}^{-a}(1 - \zeta_{\ell^n}^{a+1})(1 - \zeta_{\ell^n}^{a-1}).$$

Thus,

$$(1 - \zeta_{\ell^n}^{a+1})(1 - \zeta_{\ell^n}^{a-1}) \in \langle \pm\zeta_{\ell^n}, V_m, V_n^2 \rangle.$$

However, $(a+1) \not\equiv \pm(a-1) \pmod{\ell}$. Now Corollary 12 gives a contradiction. □

Proof of Theorem 7. Let $\ell \ge 11$. Let

$$g = \left\lfloor \frac{\ell - 3}{4} \right\rfloor = \begin{cases} (\ell-3)/4, & \ell \equiv 3 \pmod{4}, \\ (\ell-5)/4, & \ell \equiv 1 \pmod{4}. \end{cases}$$

Thus g satisfies (22). Let D_n be as in Section 9. By Lemma 27, the hyperelliptic curve $D_n/\mathbb{Q}_{n-1,\ell}$ has genus g, and good reduction away from $\{\upsilon_2, \upsilon_\ell\}$. Moreover, by Lemma 30, we have D_n and D_m are nonisomorphic, even over $\overline{\mathbb{Q}}$, for $n > m$ sufficiently large.

Now suppose

(i) $\ell = 2p + 1$ where p is also an odd prime;

(ii) 2 as a primitive root modulo p.

It then follows from Lemma 36 that $J(D_n)$ and $J(D_m)$ are nonisogenous over $\mathbb{Q}_{\infty,\ell}$ provided h_n^+ is odd for all n, where h_n^+ denotes the class number of $\Omega^{+}_{n,\ell}$. Write h_n for the class number of $\Omega_{n,\ell}$. It is known thanks to the work of Estes [1989] that h_1 is odd for all primes ℓ satisfying (i) and (ii); a simplified proof of this result is given Stevenhagen [1994, Corollary 2.3]. Moreover, Ichimura and Nakajima [2012] show, for primes $\ell \le 509$, that the ratio h_n/h_1 is odd for all n. The primes $11 \le \ell \le 509$ satisfying both (i) and (ii) are 11, 23, 59, 107, 167, 263, 347, 359. Thus for these primes h_n is odd for all n. As $h_n^+ \mid h_n$ (see for example [Washington 1997, Theorem 4.10]), we know for these primes that h_n^+ is odd for all n. This completes the proof. □

Remarks. • A key step in our proof of Theorem 7 is showing that $J(D_n)[2]$ is irreducible as an $\mathbb{F}_2[G_\infty]$-module whenever $\ell = 2p+1$ where p is a prime having 2 as a primitive root. It can be shown for all other ℓ that the $\mathbb{F}_2[G_\infty]$-module $J(D_n)[2]$ is in fact reducible.

• Another key step is the argument in the proof of Lemma 36 showing that for $n > m$ sufficiently large, the Jacobians $J(D_n)$ and $J(D_m)$ are not related via odd degree isogenies defined over $\mathbb{Q}_{\infty,\ell}$. This step can be made to work, with very minor modifications to the argument, for all $\ell \geq 11$, and all choices of genus g given in (22).

Acknowledgements

We are grateful to Minhyong Kim for drawing our attention to the conjecture of Parshin and Zarhin, and to Alain Kraus and David Loeffler for useful discussions. We thank the referee for many useful comments.

References

[Abramovich 2009] D. Abramovich, “Birational geometry for number theorists”, pp. 335–373 in *Arithmetic geometry* (Göttingen, Germany, 2006), edited by H. Darmon et al., Clay Math. Proc. **8**, Amer. Math. Soc., Providence, RI, 2009. MR Zbl

[Baker et al. 2005] M. H. Baker, E. González-Jiménez, J. González, and B. Poonen, “Finiteness results for modular curves of genus at least 2”, *Amer. J. Math.* **127**:6 (2005), 1325–1387. MR Zbl

[Buhler et al. 2004] J. Buhler, C. Pomerance, and L. Robertson, “Heuristics for class numbers of prime-power real cyclotomic fields”, pp. 149–157 in *High primes and misdemeanours*, edited by A. van der Poorten and A. Stein, Fields Inst. Commun. **41**, Amer. Math. Soc., Providence, RI, 2004. MR Zbl

[Estes 1989] D. R. Estes, “On the parity of the class number of the field of qth roots of unity”, *Rocky Mountain J. Math.* **19**:3 (1989), 675–682. MR Zbl

[Faltings 1983] G. Faltings, “Endlichkeitssätze für abelsche Varietäten über Zahlkörpern”, *Invent. Math.* **73**:3 (1983), 349–366. MR Zbl

[Freitas et al. 2020] N. Freitas, A. Kraus, and S. Siksek, “On asymptotic Fermat over $\mathbb{Z}_p$-extensions of $\mathbb{Q}$”, *Algebra Number Theory* **14**:9 (2020), 2571–2574. MR Zbl

[Freitas et al. 2021a] N. Freitas, A. Kraus, and S. Siksek, “Local criteria for the unit equation and the asymptotic Fermat’s last theorem”, *Proc. Natl. Acad. Sci. USA* **118**:12 (2021), art. id. 2026449118. MR Zbl

[Freitas et al. 2021b] N. Freitas, A. Kraus, and S. Siksek, “The unit equation over cyclic number fields of prime degree”, *Algebra Number Theory* **15**:10 (2021), 2647–2653. MR Zbl

[Greenberg 2001] R. Greenberg, “Introduction to Iwasawa theory for elliptic curves”, pp. 407–464 in *Arithmetic algebraic geometry* (Park City, UT, 1999), edited by B. Conrad and K. Rubin, IAS/Park City Math. Ser. **9**, Amer. Math. Soc., Providence, RI, 2001. MR Zbl

[Ichimura and Nakajima 2012] H. Ichimura and S. Nakajima, “On the 2-part of the class numbers of cyclotomic fields of prime power conductors”, *J. Math. Soc. Japan* **64**:1 (2012), 317–342. MR Zbl

[Kato 2004] K. Kato, “p-adic Hodge theory and values of zeta functions of modular forms”, pp. 117–290 in *Cohomologies p-adiques et applications arithmétiques, III*, edited by P. Berthelot et al., Astérisque **295**, Soc. Math. France, Paris, 2004. MR Zbl

[Kim 2005] M. Kim, “The motivic fundamental group of $\mathbb{P}^1 - \{0, 1, \infty\}$ and the theorem of Siegel”, *Invent. Math.* **161**:3 (2005), 629–656. MR Zbl

[Lawrence and Venkatesh 2020] B. Lawrence and A. Venkatesh, "Diophantine problems and p-adic period mappings", *Invent. Math.* **221**:3 (2020), 893–999. MR Zbl

[Mazur 1972] B. Mazur, "Rational points of abelian varieties with values in towers of number fields", *Invent. Math.* **18** (1972), 183–266. MR Zbl

[Nagell 1969] T. Nagell, "Sur un type particulier d'unités algébriques", *Ark. Mat.* **8** (1969), 163–184. MR Zbl

[Poonen 2021] B. Poonen, "The S-integral points on the projective line minus three points via finite covers and Skolem's method", pp. 583–587 in *Arithmetic geometry*, *number theory*, *and computation*, edited by J. S. Balakrishnan et al., Springer, 2021. MR Zbl

[Rohrlich 1984] D. E. Rohrlich, "On L-functions of elliptic curves and cyclotomic towers", *Invent. Math.* **75**:3 (1984), 409–423. MR Zbl

[Šafarevič 1963] I. R. Šafarevič, "Algebraic number fields", pp. 163–176 in *Proceedings of the International Congress of Mathematicians* (Stockholm, 1962), edited by V. Stenström, Mittag-Leffler, Djursholm, Sweden, 1963. In Russian; translated as pp. 25–39 in Amer. Math. Soc. Transl. (2) **31**, Amer. Math. Soc., Providence, RI, 1962. MR Zbl

[Schaefer 1995] E. F. Schaefer, "2-descent on the Jacobians of hyperelliptic curves", *J. Number Theory* **51**:2 (1995), 219–232. MR Zbl

[Serre 1968] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, Benjamin, New York, 1968. MR Zbl

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

[Shapiro 1943] H. Shapiro, "An arithmetic function arising from the ϕ function", *Amer. Math. Monthly* **50** (1943), 18–30. MR Zbl

[Shimura 1971] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Kanô Memorial Lect. **1**, Iwanami Shoten, Tokyo, 1971. MR Zbl

[Siksek 2022] S. Siksek, "Integral points on punctured abelian varieties", *Eur. J. Math.* **8**:suppl. 2 (2022), 687–703. MR Zbl

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Grad. Texts in Math. **106**, Springer, 1986. MR Zbl

[Stevenhagen 1994] P. Stevenhagen, "Class number parity for the pth cyclotomic field", *Math. Comp.* **63**:208 (1994), 773–784. MR Zbl

[Stoll 2001] M. Stoll, "Implementing 2-descent for Jacobians of hyperelliptic curves", *Acta Arith.* **98**:3 (2001), 245–277. MR Zbl

[Triantafillou 2021] N. Triantafillou, "There are no exceptional units in number fields of degree prime to 3 where 3 splits completely", *Proc. Amer. Math. Soc. Ser. B* **8** (2021), 371–376. MR

[Washington 1978] L. C. Washington, "The non-p-part of the class number in a cyclotomic $\mathbb{Z}_p$-extension", *Invent. Math.* **49**:1 (1978), 87–97. MR Zbl

[Washington 1997] L. C. Washington, *Introduction to cyclotomic fields*, 2nd ed., Grad. Texts in Math. **83**, Springer, 1997. MR Zbl

[Yelton 2015] J. S. Yelton, *Hyperelliptic Jacobians and their associated l-adic Galois representations*, Ph.D. thesis, Pennsylvania State University, 2015, available at https://www.proquest.com/docview/1748051112.

[Zarhin 2010] Y. G. Zarhin, "Endomorphisms of abelian varieties, cyclotomic extensions, and Lie algebras", *Mat. Sb.* **201**:12 (2010), 93–102. In Russian; translated in *Sb. Math.* **201**:12 (2010), 1801–1810. MR Zbl

[Zarhin and Parshin 1989] Y. G. Zarhin and A. N. Parshin, "Finiteness problems in Diophantine geometry", pp. 35–102 in *Eight papers translated from the Russian*, edited by B. Silver, Amer. Math. Soc. Transl. (2) **143**, Amer. Math. Soc., Providence, RI, 1989. Zbl arXiv 0912.4325

Communicated by Akshay Venkatesh
Received 2023-02-05 Revised 2023-11-29 Accepted 2024-02-13

s.siksek@warwick.ac.uk Mathematics Institute, University of Warwick, Coventry, United Kingdom

robin.visser@warwick.ac.uk Mathematics Institute, University of Warwick, Coventry, United Kingdom

ALGEBRA AND NUMBER THEORY 19:1 (2025)
https://doi.org/10.2140/ant.2025.19.143

Vanishing results for the coherent cohomology of automorphic vector bundles over the Siegel variety in positive characteristic

Thibault Alexandre

We prove vanishing results for the coherent cohomology of the good reduction modulo p of the Siegel modular variety with coefficients in some automorphic bundles. We show that for an automorphic bundle with highest weight λ near the walls of the antidominant Weyl chamber, there is an integer $e \geq 0$ such that the cohomology is concentrated in degrees $[0, e]$. The accessible weights with our method are not necessarily regular and not necessarily p-small. Since our method is technical, we also provide an algorithm written in SageMath that computes explicitly the vanishing results.

1. Introduction

1.1. ***History and motivation.*** The cohomology of automorphic vector bundles on Shimura varieties has played important roles in the study of arithmetic properties of automorphic representations as explained in [Harris 1985]. Let p be a prime number and $N \geq 3$ be an integer such that $p \nmid N$. Consider the Siegel modular variety Sh of level N and genus $g \geq 1$ over $\mathbb{F}_p$. It is defined as the fine moduli space of abelian schemes of dimension g over $\mathbb{F}_p$ with a principal polarization and a basis of their N-torsion. This scheme is smooth and not proper over $\mathbb{F}_p$ but we can consider a smooth toroidal compactification $\mathrm{Sh}^{\mathrm{tor}}$ as defined in [Faltings and Chai 1990]. This article is concerned with the coherent cohomology of $\mathrm{Sh}^{\mathrm{tor}}$. We consider automorphic vector bundles that are defined as the contracted product of the Hodge vector

MSC2020: 14G35.
Keywords: coherent cohomology, automorphic vector bundles, vanishing results, Siegel variety, Siegel modular variety, positive characteristic, highest weight, Weyl chamber, SageMath, Shimura varieties, automorphic representations, abelian schemes, g-zip theory.

bundle Ω with an algebraic representation of GL_g. Over a field of characteristic $p > 0$, there are two important indecomposable (but not necessarily irreducible) algebraic representations of highest weight λ: the standard module $\Delta(\lambda)$ and the costandard module $\nabla(\lambda)$.[1] Note that these two algebraic representations are isomorphic and irreducible when the weight λ is p-small for GL_g.[2] We use the same notation to denote the corresponding automorphic vector bundles on the Siegel modular variety. To better understand the coherent cohomology of $\mathrm{Sh}^{\mathrm{tor}}$, it is convenient to know that all but certain cohomological degrees must be zero. Lan and Suh [2012; 2013] prove many vanishing results for the coherent cohomology of PEL Shimura varieties. Let W denote the Weyl group of Sp_{2g} and I denote the type of the parabolic subgroup $P \subset \mathrm{Sp}_{2g}$ that stabilizes the Hodge filtration on the Siegel upper half-plane $\mathbb{H}_g$. In the Siegel case, Lan and Suh were able to access automorphic bundles $\Delta(\lambda)^\vee$ in all Weyl chambers as long as the weight λ can be written

$$\lambda = w \cdot \mu + \underline{k}, \tag{1}$$

where w is an element of the minimal left coset representatives ${}^I W$ of type I, μ is a sufficiently regular weight [Lan and Suh 2012, Definition 7.18] which is p-small for Sp_{2g} and such that $|\mu|_{\mathrm{re},+} < p$ [Lan and Suh 2012, 7.22] and $\underline{k}$ is a positive parallel weight.[3] They use results from [Polo and Tilouine 2002] on dual Bernstein–Gelfand–Gelfand complexes and a geometric plethysm that imposes many restrictions on the size of the weight compared to p. We note that for such a weight λ, we have $\Delta(\lambda)^\vee = \nabla(-w_0\lambda) = \Delta(-w_0\lambda)$. As there is only a finite number of p-small characters for Sp_{2g}, their method accesses only a finite number of weights up to positive parallel weights.

1.2. ***Main results.*** Let D_{red} denote the boundary of the toroidal compactification and let $\nabla^{\mathrm{sub}}(\lambda)$ denote the subcanonical extension $\nabla(\lambda)(-D_{\mathrm{red}})$ of the costandard automorphic vector bundle of highest weight λ. Our main result is a general recipe to produce new vanishing results from old ones. Namely, we define a nondecreasing function $g_{I_0,e}$ on the power set of characters

$$g_{I_0,e} : \mathcal{P}(X^*) \to \mathcal{P}(X^*)$$

that depends on a subset $I_0 \subset I$ where I is the type of the parabolic subgroup of Sp_{2g} that stabilizes the Hodge filtration and an integer $0 \le e \le d-1$ where $d = g(g+1)/2$ is the dimension of $\mathrm{Sh}^{\mathrm{tor}}$. Note that the definition of $g_{I_0,e}$ is technical and not very helpful because it is a byproduct of our method which relies on the partial degeneration of multiple spectral sequences. We describe these spectral sequences and give the exact definition of $g_{I_0,e}$ in the overview of the strategy.

Theorem (Theorem 6.16). *Assume that $p > g^2$. Let $\mathcal{C}$ be a set of characters λ for which the cohomology $H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\lambda))$ is concentrated in degrees* $[0, e+1]$. *Then, the image of $\mathcal{C}$ by the function $g_{I_0,e}$ is a set of characters λ for which the cohomology $H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\lambda))$ is concentrated in degrees* $[0, e]$.

[1] In the context of a highest weight category (see [Riche 2016, 3.7] for a general introduction to this notion), one is also concerned with the simple module $L(\lambda)$ and the tilting module $T(\lambda)$.

[2] It means that, for all $\alpha \in \phi^+$, $\langle \lambda + \rho, \alpha^\vee \rangle \le p$, where ρ is the half-sum of positive roots.

[3] It means of the form $(k, k, \dots, k)$ for some $k > 0$.

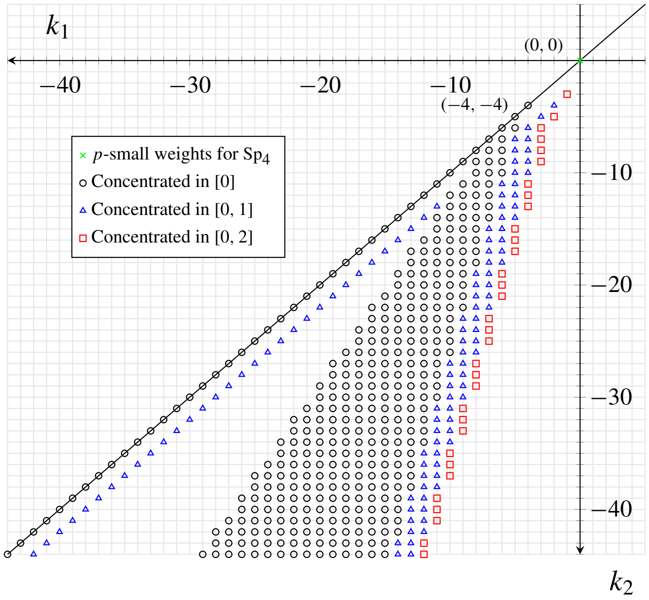

Figure 1. $g = 2$, $p = 5$.

Moreover, in the extreme cases $e = 0$ and $e = d - 1$, our method produces new vanishing results without any prior knowledge. These results can then be used in the other cases $0 < e < d - 1$. We illustrate our results in the special case $g = 2$, $p = 5$ in Figure 1.

The accessible weights with this method are not necessarily regular and not necessarily p-small (even up to a positive parallel weight) but they belong to the antidominant Weyl chamber.[4] Since the definition of the function $g_{I_0,e}$ is hard to grasp, we have implemented on SageMath an algorithm that compute the vanishing results with our method. Our method produces vanishing results for automorphic bundles $\nabla(\lambda)$ where λ is not necessarily of the form $w \cdot \mu + \underline{k}$ as in (1). In particular, we have no reason to expect that $\Delta(\lambda) = \nabla(\lambda)$. The p-smallness restriction is replaced with a much weaker restriction coming from the theory of G-Zip called orbitally p-closeness. We note that in the special case $g = 2$, $p = 5$, the only p-small character for Sp_4 is $(0, 0)$, which means that the method of Lan and Suh is only able to access weights of the form $w\rho - \rho + \underline{k}$.

1.3. ***Overview of the strategy.*** The first step is to consider the flag bundle $\pi : Y_{I_0}^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$ over the Siegel modular variety that parametrizes refinements of type $I_0 \subset I$ of the Hodge filtration of the universal semiabelian scheme on $\mathrm{Sh}^{\mathrm{tor}}$. Let d_0 denote the dimension of $Y_{I_0}^{\mathrm{tor}}$ over $\mathbb{F}_p$. Let P_0 denote the parabolic subgroup of Sp_{2g} of type I_0. For each character $\lambda \in X^*(P_0)$, we have a line bundle $\mathcal{L}_\lambda$ on $Y_{I_0}^{\mathrm{tor}}$ such that

$$\pi_* \mathcal{L}_\lambda = \nabla(\lambda).$$

Following the result from [Brunebarbe et al.], the second step is to use a result of [Goldring and Koskivirta 2019a] about the existence of generalized Hasse invariants on the stack Sp_{2g}-$\mathrm{ZipFlag}^{\mu,I_0}$ to prove that

[4] It corresponds to the dominant Weyl chamber in the work of Lan and Suh.

certain line bundles $\mathcal{L}_\lambda$ are D-ample (see Definition 5.2) on $Y_{I_0}^{\mathrm{tor}}$ where D is a certain effective Cartier divisor whose associated reduced divisor is the boundary D_{red}. The third step is to use a logarithmic version of the Kodaira–Nakano vanishing in positive characteristic due to Esnault and Viehweg to see that under the hypothesis $p > d_0 := \dim Y_{I_0}^{\mathrm{tor}}$, we have

$$H^i(Y_{I_0}^{\mathrm{tor}}, \Omega^{d_0-e}_{Y_{I_0}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}_\lambda^{\mathrm{sub}}) = 0$$

for all $i > e$ and all λ that admits generalized Hasse invariants. The fourth step is to filter the bundle

$$\Omega^{d_0-e}_{Y_{I_0}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}_\lambda^{\mathrm{sub}}$$

with an increasing filtration $F_\bullet$,

$$F_k = \pi^*\Omega^{d_0-e-k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \wedge \Omega^k_{Y_{I_0}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}_\lambda^{\mathrm{sub}},$$

and then consider the corresponding spectral sequence. It is a spectral sequence starting at the second page $E_2^{i,j}$ whose limit is zero when $i + j > e$ by the logarithmic Kodaira–Nakano vanishing considered above. In general, it is impossible to extract information on the second page of a spectral sequence whose limit is zero. However, if we can show that the second page degenerates (at least partially), then we can deduce that some terms $E_2^{i,j}$ must be zero.[5] The aim is to determine the vanishing results needed to ensure the partial degeneration of this spectral sequence. From the partial degeneration, we can deduce new vanishing results. Our method is technical as it involves recursively an unknown number of spectral sequences. Moreover, in the course of the argument, we are forced to contemplate tensor products of automorphic bundles $\nabla(\lambda) \otimes \nabla(\mu)$. To relate the cohomology of this tensor product to the cohomology of other automorphic bundles, we consider the spectral sequence associated to a ∇-filtration (see Definition 2.6) whose existence is ensured by [Mathieu 1990] and, like before, we determine the vanishing results needed to ensure its partial degeneration. The definition of the function $g_{I_0,e}$ on the power set of characters is a byproduct of our method that relies on the partial degeneration of relevant spectral sequences. More precisely, let $\mathcal{C}_{\mathrm{ample},I_0}$ denote the set of characters such that $\mathcal{L}_\lambda$ is D-ample on $Y_{I_0}^{\mathrm{tor}}$, r_0 denote the relative dimension of $\pi : Y_{I_0}^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$, $(\mu_j^k)_j$ denote the set of weights of the GL_g-module $\Lambda^k \operatorname{Sym}^2 \mathrm{std}$,[6] $s_M = \sum_{\alpha \in M} \alpha$ for any $M \subset \phi^+$ and $\rho_{I_0} = \frac{1}{2}\sum_{\alpha \in \phi_L^+ \setminus \phi_{I_0}^+} \alpha$. For any set $\mathcal{C}$ of characters, we define

$$g_{I_0,e}(\mathcal{C}) := \mu^{d-e}_{\binom{d}{d-e}} + X^*(P_0)^+ \cap (-2\rho_{I_0} + \mathcal{C}_{\mathrm{ample},I_0}) \cap \bigcap_{k,j,M} (s_M - 2\rho_{I_0} - \mu_j^{d-e+k} + \mathcal{C}),$$

where the last intersection is taken over the set of k, j, M where $0 \le k \le e$, $1 \le j \le \binom{d}{d-e} + k$ and $M \subset \phi_L^+ - \phi_{I_0}^+$ such that $|M| = r_0 - k$ with the exception of $j = \binom{d}{d-e}$ when $k = 0$.

[5] In the case $e = 0$, the spectral sequence is concentrated on one row which explains why we do not need any prior vanishing results.

[6] Actually, to follow the convention in Definition 3.10, we need to twist these weights by $w_0 w_{0,\mathrm{GL}_g}$ and assume they are ordered in a way that $w_0 w_{0,\mathrm{GL}_g}(\mu^n_{\binom{d}{n}})$is the highest weight.

1.4. ***Organization of the paper.*** In Section 2, we recall some results of algebraic representation of reductive groups in positive characteristic. In Section 3, we recall the definition of the Siegel modular variety and the different automorphic vector bundles. In Section 4, we recall how the theory of G-Zip can be used to study the EO stratification of the Siegel modular variety. In particular, we recall the main result of [Goldring and Koskivirta 2019a] about generalized Hasse invariants. In Section 5, we prove that line bundles of weight λ on the flag bundle over the Siegel modular variety that admit generalized Hasse invariants are D-ample. We also recall a logarithmic version of the Kodaira–Nakano vanishing in positive characteristic due to Esnault and Viehweg. In Section 6, we present our general method for producing new vanishing results and we give more details in the case $g = 2$. In Section 7, we explain how to compute new vanishing results with an algorithm written in SageMath and we plot the results we have obtained in some special cases with $g = 2$ and $g = 3$. See github.com/ThibaultAlexandre/vanishing-results-over-the-siegel-variety to download the algorithm.

2. Recollection on group theory

In this section, we follow mostly [Jantzen 2003] for generalities about algebraic representations of reductive groups over a field of positive characteristic. Let k be a field of positive characteristic $p > 0$ and G a geometrically connected split reductive algebraic group over k. The weights of the adjoint representation of G on its Lie algebra $\mathfrak{g}$ define a set of roots ϕ. We fix a Borel pair (B, T) defined over k where B is a Borel subgroup and $T \subset B$ is a maximal torus. This choice of Borel pair determines a subset of simple roots Δ and positive roots ϕ^+. We use a nonstandard convention for the positive roots as we declare α to be positive if the root group $U_{-\alpha}$ is contained in B.[7] Let ρ denote the half-sum of positive roots as $\mathbb{Q}$-character of T. If $I \subset \Delta$, we write ϕ_I (resp. ϕ_I^+) for the set of roots (resp. positive roots) generated from I. We write W for the Weyl group of G, $l : W \to \mathbb{N}$ for its length function and w_0 for its longest element. If $I \subset \Delta$, let $W_I \subset W$ denote the subgroup generated by the reflections s_α where $\alpha \in I$ and let ${}^I W \subset W$ denote the set of minimal length representatives of $W_I \backslash W$. We write $\langle \cdot , \cdot \rangle : X^*(T) \times X_*(T) \to \mathbb{Z}$ for the perfect pairing between the characters $X^*(T)$ of T and the cocharacters $X_*(T)$ of T. Since the characteristic of k is assumed to be positive, G is endowed with a relative Frobenius morphism $\varphi : G \to G^{(p)}$ where $G^{(p)} := G \times_{k,\sigma} k$ (with $\sigma : k \to k$ the Frobenius morphism of k) is again a reductive group over k. Unlike when the characteristic of k is 0, the category of algebraic representations of G on finite-dimensional vector spaces is no longer semisimple. The simple objects $L(\lambda)$ are still indexed by their highest weight λ but not every representation can be split into a direct sum of simple objects. This category $\mathrm{Rep}_k(G)$ has the structure of a highest weight category; see [Riche 2016, 3.7] for a general introduction to this notion.

Definition 2.1 [Jantzen 2003, Part I, Section 5.8]. Let $\lambda : T \to \mathbb{G}_m$ be a character of T. We define a line bundle $\mathcal{L}_\lambda$ on the flag variety G/B as the B-quotient of the vector bundle $G \times_k \mathbb{A}^1 \to G$, where B acts

[7] It simplifies the statement of the Proposition 3.20.

on $G \times_k \mathbb{A}^1$ by

$$(g, x)b = (gb^{-1}, \lambda(b^{-1})x),$$

and where λ is naturally extended by 0 on the unipotent part of B. The global section group $H^0(G/B, \mathcal{L}_\lambda)$ is given the structure of a G-module through left translation. As a consequence we get an algebraic representation of G, and we will denote it simply $\nabla(\lambda)$.

Proposition 2.2 [Jantzen 2003, Part II, Section 2.6]. *The G-module $\nabla(\lambda)$ is nonzero exactly when λ is dominant. Moreover, its highest T-weight is λ and we call $\nabla(\lambda)$ the induced module or costandard module of highest weight λ.*

Remark 2.3. A different convention can be found in the literature where we set the dominance to be relative to B.

Definition 2.4 [Jantzen 2003, Part II, Section 2.13]. Let $\lambda \in X^*(T)$ be a character. The standard module of highest weight λ can be defined

$$\Delta(\lambda) := \nabla(-w_0\lambda)^\vee,$$

where w_0 is the longest element of the Weyl group W of G and $\vee$ denotes the linear dual in $\mathrm{Rep}_k(G)$.

As a consequence from the definitions, $\nabla(\lambda)$ and $\Delta(\lambda)$ must have the same characters but they are usually not simple and not isomorphic. However $L(\lambda)$ is the socle of $\nabla(\lambda)$ and the head of $\Delta(\lambda)$; see [Jantzen 2003, Part II, Chapter 2]). We recall Kempf's vanishing theorem.

Proposition 2.5 [Jantzen 2003, Part II, Section 4.5]. *Let λ be a dominant character. For each $i > 0$, we have*

$$H^i(G/B, \mathcal{L}_\lambda) = 0.$$

More generally, let P be a standard parabolic of type $I \subset \Delta$ and λ a I-dominant character of P (i.e., $\langle \lambda, \alpha^\vee \rangle \geq 0$ for all $\alpha \in \Delta \backslash I$ and $\langle \lambda, \alpha^\vee \rangle = 0$ for all $\alpha \in I$). There is an associated line bundle $\mathcal{L}_\lambda$ on G/P and we have

$$H^i(G/P, \mathcal{L}_\lambda) = 0$$

for all $i > 0$.

Proof. We give a sketch of the argument. The first step is to show that $\mathcal{L}_\lambda$ is ample over the flag variety G/B exactly when λ is strictly dominant by reducing to the case $G = \mathrm{SL}_2$ and $G/B = \mathbb{P}^1_k$. Then, in characteristic 0, we can conclude with the Kodaira–Nakano vanishing theorem since the canonical bundle $\omega_{G/B}$ of G/B is antiample. Indeed, we have an isomorphism

$$\omega_{G/B} = \mathcal{L}_{-2\rho}$$

and if we consider a dominant character λ, the line bundle

$$\omega_{G/B}^{-1} \otimes_{\mathcal{O}_{G/B}} \mathcal{L}_\lambda = \mathcal{L}_{2\rho+\lambda}$$

is ample since $2\rho+\lambda$ is strictly dominant. The Kodaira–Nakano vanishing theorem applied to $\mathcal{L}_{2\rho+\lambda}$ says that

$$H^i(G/B, \underbrace{\omega_{G/B}\otimes\mathcal{L}_{2\rho+\lambda}}_{=\mathcal{L}_\lambda}) = 0$$

for all $i>0$. In positive characteristic, we can conclude with Serre's cohomological criterion for ampleness and the formula in [Jantzen 2003, Part II, Section 3.19] with the Steinberg module $\nabla((p^r-1)\rho)$. □

We insist on the fact that the proof in [loc. cit., Part II, Section 5.3] of the more general Borel–Weil–Bott theorem which gives information on the higher cohomology groups of $\mathcal{L}_\lambda$ when λ is no longer dominant requires to divide by binomial numbers $\binom{n}{k}$ with $n \geq p$, which is impossible in characteristic p. Actually, one can find counterexamples to the Borel–Weil–Bott theorem in positive characteristic; see [loc. cit., Part II, Section 15.8]. In characteristic 0, it is easier to understand tensor product of highest weight representations: we know that $L(\lambda)\otimes L(\mu)$ is a direct sum of $L(\lambda')$ where λ' can be expressed as $\lambda+\mu'$ where $\mu'\leq\mu$ is a weight of $L(\mu)$. Going back to our positive characteristic case, we would like to have a weaker but similar kind of result for $\nabla(\lambda)$'s.

Definition 2.6. Let V be an algebraic representation of G. We say that:

(1) V admits a ∇-filtration if there is a finite filtration

$$0 = V^n \subsetneq V^{n-1} \subsetneq \cdots \subsetneq V^0 = V$$

with graded pieces

$$V^i/V^{i+1} \simeq \nabla(\nu_i)$$

for some dominant characters ν_i.

(2) V admits a Δ-filtration if there is a finite filtration

$$0 = V^n \subsetneq V^{n-1} \subsetneq \cdots \subsetneq V^0 = V$$

with graded pieces

$$V^i/V^{i+1} \simeq \Delta(\nu_i)$$

for some dominant characters ν_i.

Remark 2.7. In the setting of a highest weight category, tilting modules are defined as modules that admit both a ∇- and a Δ-filtration.

The following proposition states the existence of a ∇-filtration for a tensor product $\nabla(\lambda)\otimes\nabla(\mu)$ and gives some details about its graded pieces. This result is due to Donkyn [1985] when G does not contain any components of type E_7, E_8 or that $p\neq 2$. His approach relies on a case by case analysis of each Dynkin diagram and requires long and difficult calculations. A more general proof, without the technical restrictions, was given later by Mathieu. We first need a lemma.

Lemma 2.8. *Let λ, μ denote T-characters such that* $\mathrm{Ext}^1_G(\nabla(\lambda),\nabla(\mu))\neq 0$. *Then*, $\lambda\geq\mu$.

Proof. We have

$$\begin{aligned}\operatorname{Ext}^1_G(\nabla(\lambda), \nabla(\mu)) &= H^1(G, \Delta(-w_0\lambda) \otimes \nabla(\mu)) \\ &= H^1(P, \Delta(-w_0\lambda) \otimes \mu) \qquad \text{by [Jantzen 2003, Part II, Section 4.7]}\end{aligned}$$

and by [loc. cit., Part II, Section 4.10 b)], there exists a weight ν of $\Delta(-w_0\lambda)$ such that $-(\nu+\mu)$ is a $\mathbb{N}$-linear combination of positive roots ϕ^+. In particular, we have $-\nu \geq \mu$. Since $w_0(-w_0\lambda) = -\lambda$ is the lowest weight of $\Delta(-w_0\lambda)$, we deduce that $\lambda \geq -\nu \geq \mu$. □

Proposition 2.9 [Mathieu 1990]. *Let λ, μ be two dominant characters in $X^*(T)$. Then $\nabla(\lambda) \otimes \nabla(\mu)$ admits a ∇-filtration $(V^i)_{i\geq 0}$ with graded pieces*

$$V^i/V^{i+1} \simeq \nabla(\lambda + \mu_i),$$

where $(\mu_i)_i$ is a collection of weights of $\nabla(\mu)$ with $\mu_0 = \mu$. In particular, the first graded piece is given by $V^0/V^1 = \nabla(\lambda+\mu)$.

Proof. We add some details to the result of Mathieu to explain how to get a filtration with the desired properties. The result of Mathieu assumes that G is a connected, simply connected, semisimple algebraic group over an algebraically closed field k of characteristic $p > 0$ and it is not hard to reduce to this case. By [Mathieu 1990, Theorem 1], there exists a filtration

$$0 = V^n \subset V^{n-1} \subset \cdots \subset V^1 \subset \cdots V^0 = \nabla(\lambda) \otimes \nabla(\mu),$$

where for each i the graded piece V^i/V^{i+1} is a costandard module $\nabla(\nu_i)$ for some dominant character ν_i. The character class of $\nabla(\lambda) \otimes \nabla(\mu)$ is

$$\operatorname{ch}(\nabla(\lambda) \otimes \nabla(\mu)) = \sum_i \operatorname{ch} \nabla(\lambda + \mu_i),$$

where the sum is taken over some weights $(\mu_i)_i$ of $\nabla(\mu)$. As the highest weight of this module, $\lambda+\mu$ contributes to the sum. Note that the nonzero terms are those such that $\lambda+\mu_i$ is dominant. We choose an ordering of the $(\mu_i)_i$ such that whenever $\mu_i < \mu_j$ for some i, j then $i > j$. It implies that there exists a permutation σ on $0, 1, \ldots, n-1$ such that

$$V^i/V^{i+1} = \nabla(\lambda + \mu_{\sigma(i)})$$

for all i between 0 and $n-1$. We remake the argument in [Jantzen 2003, Part 11, Section 4.16, Remark 4] to explain how to reorganize the terms. If $\sigma(i) < \sigma(i+1)$ for some i, $0 \leq i \leq n-2$, then $\lambda+\mu_{\sigma(i)} \not< \lambda+\mu_{\sigma(i+1)}$ and the exact sequence

$$0 \to \nabla(\lambda + \mu_{\sigma(i+1)}) \to V^i/V^{i+2} \to \nabla(\lambda + \mu_{\sigma(i)}) \to 0$$

is split because $\operatorname{Ext}^1_G(\nabla(\lambda + \mu_{\sigma(i+1)}), \nabla(\lambda + \mu_{\sigma(i)})) = 0$ by Lemma 2.8. It shows that

$$V^i/V^{i+2} = \nabla(\lambda + \mu_{\sigma(i+1)}) \oplus \nabla(\lambda + \mu_{\sigma(i)})$$

and we can replace V^{i+1} by a submodule $\tilde{V}^{i+1}$ between V^{i+2} and V^i such that $\tilde{V}^{i+1}/V^{i+2} = \nabla(\lambda+\mu_{\sigma(i)})$ and $V^i/\tilde{V}^{i+1} = \nabla(\lambda+\mu_{\sigma(i+1)})$. We iterate this process to produce the desired filtration. □

Remark 2.10. (1) Not all the weights $\mu' \leq \mu$ of $\nabla(\mu)$ such that $\lambda+\mu'$ is dominant will contribute to the filtration.

(2) Even if we will not need it, we note that the dual statement says that tensor products of standard modules $\Delta(\lambda)\otimes\Delta(\mu)$ admit a Δ-filtration.

Corollary 2.11. *Let V and W be two algebraic representations of G that admit a ∇-filtration. Then, $V\otimes W$ admits a ∇-filtration.*

We recall the Donkyn criterion.

Proposition 2.12. *Let V be an algebraic representation of G. The following proposition are equivalent*:

(1) *V admits a ∇-filtration.*

(2) *For all dominant characters λ and $i > 0$, $\mathrm{Ext}^i_G(\Delta(\lambda), V) = 0$.*

(3) *For all dominant characters λ, $\mathrm{Ext}^1_G(\Delta(\lambda), V) = 0$.*

Proof. See [Jantzen 2003, Part II, Section 4.16]. □

Corollary 2.13. *Let V and W be two algebraic representations of G. If V admits a ∇-filtration and W is a direct factor of V, then W admits a ∇-filtration.*

3. Recollection on Siegel varieties

In this section, we follow [Faltings and Chai 1990] for generalities about Siegel varieties. We denote by $\mathrm{Sch}_{\mathbb{F}_p}$ the category of schemes over $\mathbb{F}_p$ and $\mathbb{A}_f$ the finite adeles of $\mathbb{Q}$.

Definition 3.1. Let A and A' be abelian schemes of relative dimension g over a scheme S. A quasiisogeny $A \to A'$ is an equivalence class of pairs (α, N) where $\alpha : A \to A'$ is an isogeny over S and N is a positive integer with the relation

$$(\alpha, N) \sim (\alpha', N') \quad \text{if and only if} \quad N'\alpha = N\alpha'.$$

Definition 3.2. Let V be the $\mathbb{Z}$-module $\mathbb{Z}^{2g}$ with the standard nondegenerate symplectic pairing

$$\psi : V \times V \to \mathbb{Z}$$

such that $\psi(x, y) = {}^t x J x$ where

$$J = \begin{pmatrix} 0 & I_g \\ -I_g & 0 \end{pmatrix}.$$

We denote by Sp_{2g} the algebraic group over $\mathbb{Z}$ of $2g \times 2g$ matrices M that preserves the symplectic pairing ψ, i.e., such that

$${}^t M J M = J.$$

In the following proposition, we define the Siegel modular variety of level K as a scheme over $\mathbb{F}_p$ when the level is small enough and p is a prime such that $K_p = \mathrm{Sp}_{2g}(\mathbb{Z}_p)$.

Proposition 3.3 [Faltings and Chai 1990]. *Let $K \subset \mathrm{Sp}_{2g}(\mathbb{A}_f)$ be a subgroup that can be written as $K = K_p K^p$ for $K_p \subset \mathrm{Sp}_{2g}(\mathbb{Q}_p)$ hyperspecial and $K^p \subset \mathrm{Sp}_{2g}(\mathbb{A}_f^p)$ compact open. Consider the fibered category in groupoids $\mathcal{A}_{g,K}$ on $\mathrm{Sch}_{\mathbb{F}_p}$ whose S-points are groupoids with*:

- *Objects*: *$(A/S, \lambda, \psi)$ where A/S is abelian scheme over S of relative dimension g, $\lambda : A \to A^\vee$ is a $\mathbb{Z}_{(p)}$-multiple of a principal polarization and for all primes $l \neq p$ and all geometric points $s \in S$, ψ_l is a K_l-orbit of symplectic* ***isomorphisms*** *from $H_1(A_s, \mathbb{Q}_l)$ to $V \otimes \mathbb{Q}_l$ which is invariant under $\pi_1(S, s)$. The structure of symplectic $\mathbb{Q}_l$-vector space on the l-adic étale homology group $H_1(A_s, \mathbb{Q}_l)$ (it is also the rational Tate module of A_s) is the one induced by the polarization (which is an isomorphism since we tensor by $\mathbb{Q}_l$) and the Weil paring.*

- *Morphisms*: *A morphism $(A/S, \lambda, \psi) \to (A'/S, \lambda', \psi')$ is a quasiisogeny $\alpha : A \to A'$ over S such that the diagram*

$$\begin{tikzcd}
A \arrow[r, "\alpha"] \arrow[d, "\lambda"] & A' \arrow[d, "\lambda'"] \\
A^\vee & A'^\vee \arrow[l, "\alpha^\vee"']
\end{tikzcd}$$

is commutative up to a constant in $\mathbb{Z}_{(p)}$ and the pullback of ψ_l by the quasiisogeny α is ψ'_l.

If the level away from p, K^p, is small enough,[8] then $\mathcal{A}_{g,K}$ is representable by a smooth integral quasiprojective scheme over $\mathbb{F}_p$.

Remark 3.4. Without the hypothesis on the smallness of K, $\mathcal{A}_{g,K}$ is only a Deligne–Mumford stack over $\mathbb{F}_p$.

Notation 3.5. We fix some notation for the rest of this section. Let G denote the algebraic group Sp_{2g} over $\mathbb{F}_p$ where $g \geq 1$. We fix a neat finite level K that can be written as $K = K_p K^p$ for $K_p \subset G(\mathbb{Q}_p)$ hyperspecial and $K^p \subset G(\mathbb{A}_f^p)$ compact open. Let Sh denote the smooth quasiprojective variety $\mathcal{A}_{g,K}$ over $\mathbb{F}_p$. Let $\mu : \mathbb{G}_m \to G$ denote the minuscule cocharacter that stabilizes the Hodge filtration[9] and let $P^+ := P_\mu$, $P := P_{-\mu}$ denote the associated opposite parabolic subgroups with common Levi subgroup $L = \mathrm{GL}_g$ over $\mathbb{F}_p$. We consider the Borel B of upper triangular matrices in $G = \mathrm{Sp}_{2g}$, so that $B \subset P$. We write ϕ_L (resp. ϕ_L^+) for the roots of L (resp. positive roots of L).

Denote by $\pi : A \to \mathrm{Sh}$ the universal abelian scheme and $e : \mathrm{Sh} \to A$ its neutral section. The universal polarization of A gives to the algebraic de Rham cohomology $\mathcal{H}^1_{\mathrm{dR}}$ of A over Sh the structure of a Sp_{2g}-torsor over Sh.

[8] It is the case in particular when K is the kernel of the reduction map $\mathrm{Sp}_{2g}(\mathbb{Z}) \to \mathrm{Sp}_{2g}(\mathbb{Z}/N\mathbb{Z})$ with $N \geq 3$ such that $p \nmid N$.

[9] It maps z to $\left(\begin{smallmatrix} zI_g & 0 \\ 0 & z^{-1}I_g \end{smallmatrix}\right)$ with our choice of symplectic pairing in Definition 3.2.

Proposition 3.6. *The de Rham cohomology $\mathcal{H}^1_{\mathrm{dR}}$ is equipped with a Hodge filtration*

$$0 \to \Omega \to \mathcal{H}^1_{\mathrm{dR}} \to \Omega^\vee \to 0,$$

where

$$\Omega = e^*\Omega^1_{A/\,\mathrm{Sh}}$$

and

$$\Omega^\vee = R^1\pi_*\mathcal{O}_A.$$

We call Ω the Hodge vector bundle, it is a locally free sheaf of rank g over Sh. *Moreover, the Hodge bundle Ω is totally isotropic for the symplectic pairing on $\mathcal{H}^1_{\mathrm{dR}}$ which allows us to identify the Hodge filtration with a P-torsor on* Sh.

Proof. The Hodge filtration comes from the degeneration at the second page of the Hodge-de Rham spectral sequence which is a result of Deligne and Illusie [1987] in the case of abelian schemes. The vector bundle Ω is locally free of rank g because $\pi : A \to \mathrm{Sh}$ is smooth. Actually, we also have an isomorphism

$$\Omega \simeq \pi_*\Omega^1_{A/\,\mathrm{Sh}}.$$

Indeed, as a group scheme π satisfies

$$\Omega^1_{A/\,\mathrm{Sh}} = \pi^*e^*\Omega^1_{A/\,\mathrm{Sh}}$$

and for any proper morphism $f : X \to Y$ with geometrically connected fibers, we have

$$f_*\mathcal{O}_X = \mathcal{O}_Y.$$

From the projection formula, we deduce

$$\pi_*\Omega^1_{A/\,\mathrm{Sh}} = \pi_*(\pi^*e^*\Omega^1_{A/\,\mathrm{Sh}} \otimes \mathcal{O}_A) = e^*\Omega^1_{A/\,\mathrm{Sh}} \otimes \pi_*\mathcal{O}_A = \Omega^1_{A/\,\mathrm{Sh}}. \qquad \square$$

The Siegel modular variety Sh is not proper but we can consider a toroidal compactification.

Definition 3.7 [Faltings and Chai 1990, Chapter 4; Lan 2012, Theorem 2.15]. Let C be the cone of all positive semidefinite symmetric bilinear forms on $X^*(T) \otimes_{\mathbb{Z}} \mathbb{R}$ whose radicals are defined over $\mathbb{Q}$. Let $\Sigma = \{\sigma_\alpha\}_\alpha$ be a smooth $\mathrm{GL}(X^*(T))$-admissible decomposition in polyhedral cones of C as defined in [Faltings and Chai 1990, Chapter 4, Definition 2.2/2.3]. We assume that Σ admits a $\mathrm{GL}(X^*(T))$-equivariant polarization function as defined in [Faltings and Chai 1990, Chapter 4, Definition 2.4]. See [Ash et al. 1975] or [Kempf et al. 1973] for a proof of the existence of such polyhedral cone decompositions. We consider the corresponding toroidal compactification $\mathrm{Sh}^{\mathrm{tor},\Sigma}$ of the Siegel modular variety Sh. It is a smooth and projective scheme over $\mathbb{F}_p$ which satisfies the following properties:

(1) The complementary $D_{\mathrm{red}} = \mathrm{Sh}^{\mathrm{tor},\Sigma} - \mathrm{Sh}$, when endowed with its reduced structure, is a Cartier divisor with normal crossings.

(2) The universal abelian scheme $f : A \to \mathrm{Sh}$ extends to a semiabelian scheme $f^{\mathrm{tor}} : A^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$ such that $\Omega^{\mathrm{tor}} := e^*\Omega^1_{A^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor},\Sigma}}$ is a vector bundle that extends the Hodge bundle to $\mathrm{Sh}^{\mathrm{tor},\Sigma}$.

(3) By [Faltings and Chai 1990, Chapter 4] or [Lan 2012, Theorem 2.15, (2)] the semiabelian scheme $f^{\mathrm{tor}} : A^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$ can be compactified into a proper and log-smooth scheme $\bar{f}^{\mathrm{tor}} : \bar{A}^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$ which is projective and smooth over $\mathbb{F}_p$

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d] & A^{\mathrm{tor}} \arrow[r] \arrow[d, "f^{\mathrm{tor}}"] & \bar{A}^{\mathrm{tor}} \arrow[d] \arrow[dl, "\bar{f}^{\mathrm{tor}}"] \\
\mathrm{Sh} \arrow[r] & \mathrm{Sh}^{\mathrm{tor}} \arrow[r] & \operatorname{Spec} \mathbb{F}_p
\end{tikzcd}$$

and we denote again D_{red} the normal crossing divisor $\bar{A}^{\mathrm{tor}} - A$.

(4) Following [Faltings and Chai 1990, Chapter 4] or [Lan 2012, Theorem 2.15, (3)], the log-de Rham complex $\bar{\Omega}^{\bullet}_{\bar{A}^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}$ is the complex of log-differentials $\bar{\Omega}^{i}_{\bar{A}^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} := \Lambda^i \bar{\Omega}^{1}_{\bar{A}^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}$ where

$$\bar{\Omega}^{1}_{\bar{A}^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} = \Omega^1_{\bar{A}^{\mathrm{tor}}}(\log D_{\mathrm{red}})/(\bar{f}^{\mathrm{tor}})^{*}\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}).$$

(5) The log-de Rham cohomology

$$\mathcal{H}^1_{\log-\mathrm{dR}} := R^1(\bar{f}^{\mathrm{tor}})_* \bar{\Omega}^{\bullet}_{\bar{A}^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}$$

is a Sp_{2g}-torsor that extends $\mathcal{H}^1_{\mathrm{dR}}$ on Sh.

(6) The logarithmic Hodge-de Rham spectral sequence

$$E_1^{i,j} = R^j(\bar{f}^{\mathrm{tor}})_* \bar{\Omega}^{i}_{\bar{A}^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} \quad \Longrightarrow \quad \mathcal{H}^i_{\log-\mathrm{dR}} := R^i(\bar{f}^{\mathrm{tor}})_* \bar{\Omega}^{\bullet}_{\bar{A}^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}$$

degenerates at page 1. It defines a P-reduction of the Sp_{2g}-torsor $\mathcal{H}^1_{\log-\mathrm{dR}}$ on $\mathrm{Sh}^{\mathrm{tor},\Sigma}$ that extends the Hodge filtration on Sh.

From now on, we drop the superscript Σ to denote $\mathrm{Sh}^{\mathrm{tor},\Sigma}$ since the coherent cohomology does not depend on this choice.

Definition 3.8 [Faltings and Chai 1990]. We define the minimal compactification as the projective scheme

$$\mathrm{Sh}^{\mathrm{min}} := \operatorname{Proj}\left(\bigoplus_{n\geq 0} H^0(\mathrm{Sh}^{\mathrm{tor}}, \omega^{\otimes n})\right),$$

where $\omega = \det \Omega^{\mathrm{tor}}$ is the Hodge line bundle.

The minimal compactification $\mathrm{Sh}^{\mathrm{min}}$ is a normal and projective variety (independent of the choice of Σ) but it is not smooth in general. Moreover, the Hodge line bundle ω descends to an ample line bundle on $\mathrm{Sh}^{\mathrm{min}}$. From this construction, one can see that $\mathrm{Sh}^{\mathrm{tor}}$ is the normalization of the blow-up of $\mathrm{Sh}^{\mathrm{min}}$ along a coherent sheaf of ideals $\mathcal{J}$ of $\mathcal{O}_{\mathrm{Sh}^{\mathrm{min}}}$ and we write

$$\varphi : \mathrm{Sh}^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{min}}$$

for the induced morphism. The pullback $\varphi^*\mathcal{J}$ is of the form $\mathcal{O}_{\mathrm{Sh}^{\mathrm{tor}}}(-D)$ where D is an effective Cartier divisor whose associated reduced Cartier divisor is D_{red}. In particular, we deduce that there exists $\eta_0 > 0$ such that $\omega^{\otimes \eta}(-D)$ is ample on $\mathrm{Sh}^{\mathrm{tor}}$ for every $\eta \geq \eta_0$. In general, ω fails to be ample on $\mathrm{Sh}^{\mathrm{tor}}$.

Remark 3.9. The effective Cartier divisor D depends on the choice of the $\mathrm{GL}(X^*(T))$-equivariant polarization function on the decomposition in polyhedral cones Σ.

We are now able to define automorphic vector bundles with contracted products.

Definition 3.10. Let V be a finite-dimensional algebraic representation of $L = \mathrm{GL}_g$. We define the associated vector bundle $\mathcal{W}(V)$ on Sh (resp. its canonical extension to $\mathrm{Sh}^{\mathrm{tor}}$) to be the contracted product of V with the GL_g-torsor Ω (resp. Ω^{tor}). If $\lambda \in X^*(T)$ is an L-dominant character, we write simply $\nabla(\lambda)$ for the vector bundle corresponding to the induced representation $H^0(L/B_L, \mathcal{L}_\lambda)$ of L. It corresponds to the costandard representation of highest weight $w_0 w_{0,L}\lambda$.

Remark 3.11. We apologize for the weird convention in Definition 3.10. The advantage of this convention is to keep an easy formula in Proposition 3.20.

We recall the Kodaira–Spencer isomorphism.

Proposition 3.12 [Faltings and Chai 1990, Chapter 3, Section 9]. *The Kodaira–Spencer map*

$$\rho_{\mathrm{KS}} : \mathrm{Sym}^2 \Omega \to \Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})$$

is an isomorphism. This allows us to identify the logarithmic differentials $\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})$ with the automorphic vector bundle $\mathcal{W}(\mathrm{Sym}^2 \mathrm{std}_L) = \nabla(0, \dots, 0, -2)$. In particular, we have an isomorphism of line bundles

$$\Omega^d_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \simeq \nabla(-2\rho^L),$$

where d is the dimension of $\mathrm{Sh}^{\mathrm{tor}}$ and

$$\rho^L = \frac{1}{2} \sum_{\alpha \in \phi^+ \setminus \phi_L^+} \alpha.$$

Recall that the Hodge filtration on Sh is canonically identified with a P-torsor that extends to the toroidal compactification $\mathrm{Sh}^{\mathrm{tor}}$. From now on, I_0 denotes a subset of I and P_0 denotes its associated intermediate parabolic subgroup $B \subset P_0 \subset P$.

Definition 3.13. Let $S \to \mathrm{Sh}$ be a S-valued point of Sh and denote by A/S the corresponding abelian scheme. The flag bundle Y_{I_0} is the scheme over Sh whose S-valued points are P_0-reductions of the P-torsor corresponding to the Hodge filtration of A.

From the definition of Y_{I_0}, we get a smooth proper morphism $\pi : Y_{I_0} \to \mathrm{Sh}$ where each fiber is isomorphic to the flag variety P/P_0.

Remark 3.14. In the special case $I_0 = I$, the flag bundle Y_{I_0} coincide with the Siegel modular variety Sh.

The flag bundle Y_{I_0} extends to a flag bundle $Y_{I_0}^{\mathrm{tor}}$ over the toroidal compactification $\mathrm{Sh}^{\mathrm{tor}}$ because we have seen that the Hodge filtration over Sh extends to $\mathrm{Sh}^{\mathrm{tor}}$ in Definition 3.7(7). It implies by base change that the universal P_0-torsor on Y_{I_0} extends to $Y_{I_0}^{\mathrm{tor}}$. This allows us to define automorphic vector bundles on Y_{I_0} from algebraic representations of P_0.

Definition 3.15. Let V be a finite-dimensional algebraic representation of P_0. We define the associated vector bundle $\mathcal{L}(V)$ on Y_{I_0} (resp. its canonical extension on $Y_{I_0}^{\text{tor}}$) as the contracted product of V with the universal P_0-torsor on Y_{I_0} (resp. $Y_{I_0}^{\text{tor}}$). If $\lambda \in X^*(P_0)$ is a character, we write simply $\mathcal{L}_\lambda$ for the corresponding line bundle.

Remark 3.16. In the special case where $I_0 = \varnothing$, note that we have $X^*(P_0) = X^*(T)$.

There is an easy relation between $\mathcal{L}_\lambda$ and $\nabla(\lambda)$ that we want to explain. We first recall the proper base change theorem for coherent cohomology.

Proposition 3.17 (proper base change, nonreduced case). *Let $f : X \to S$ be a proper morphism between locally noetherian schemes. Let $\mathcal{F}$ be a coherent sheaf over X which is flat over S. Let $p \geq 0$ and $s \in S$. If $\theta_s^p : (R^p f_*\mathcal{F})_s \otimes_{\mathcal{O}_{S,s}} k(s) \to H^p(X_s, \mathcal{F}_{|X_s})$ is surjective, then there is an open neighborhood U of s such that for all $s' \in U$, $\theta_{s'}^p$ is an isomorphism and the following conditions are equivalent:*

(1) *θ_s^{p-1} is surjective.*

(2) *$R^p f_*\mathcal{F}$ is free on U.*

Under these conditions, the formation of $R^p f_*\mathcal{F}$ commutes under base change. This means that for any $g : S' \to S$, we have $g^* R^p f_*\mathcal{F} \simeq R^p f'_* g'^* \mathcal{F}$ where the maps are defined in the following cartesian diagram:

$$\begin{tikzcd}
X' \arrow[r, "g'"] \arrow[d, "f'"] & X \arrow[d, "f"] \\
S' \arrow[r, "g"] & S
\end{tikzcd}$$

Proof. See [Hartshorne 1977, Part III, Theorem 12.11]. □

Remark 3.18. (1) We assume θ_s^{-1} to be the zero morphism.

(2) The reference in Proposition 3.17 states a coherent base change theorem only for geometric points of S. To see how it implies the base change for any morphism $S' \to S$, see [Conrad, Proposition 2.1].

Lemma 3.19. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Artin stacks and $\pi : \mathcal{Y} \to \mathcal{X}$ a proper representable morphism. Let $\mathcal{L}$ be a coherent sheaf over $\mathcal{Y}$, flat over $\mathcal{X}$, such that for all geometric points $x :$ Spec $\bar{k} \to \mathcal{X}$ fitting in the cartesian diagram*

$$\begin{tikzcd}
\mathcal{Y}_x := \mathcal{Y} \times_{\mathcal{X},x} \operatorname{Spec} \bar{k} \arrow[r, "i"] \arrow[d, "\pi_x"] & \mathcal{Y} \arrow[d, "\pi"] \\
\operatorname{Spec} \bar{k} \arrow[r, "x"] & \mathcal{X}
\end{tikzcd}$$

the complex $R(\pi_x)_*\mathcal{L}_{|\mathcal{Y}_x}$ is concentrated in degree 0. Then, the complex $R\pi_*\mathcal{L}$ is also concentrated in degree 0.

Proof. Consider a presentation $f : X \to \mathcal{X}$ of the Artin stack $\mathcal{X}$ where X is a scheme and f is a surjective and smooth morphism. Consider the double cartesian diagram

$$\begin{tikzcd}
Y_x := Y \times_{X,x} \operatorname{Spec} \bar{k} \arrow[r, "i"] \arrow[d, "\pi_x"] & Y := X \times_{\mathcal{X}} \mathcal{Y} \arrow[r, "f'"] \arrow[d, "\tilde{\pi}"] & \mathcal{Y} \arrow[d, "\pi"] \\
\operatorname{Spec} \bar{k} \arrow[r, "x"] & X \arrow[r, "f"] & \mathcal{X}
\end{tikzcd}$$

where x is a geometric point of X. For any $i > 0$, we have $H^i(Y_x, \mathcal{L}_{|Y_x}) = 0$ by hypothesis. As a consequence, the base change morphism for the first cartesian diagram

$$\theta_x^i : R^i\tilde{\pi}_*\mathcal{L}_{|Y} \otimes_{\mathcal{O}_{X,x}} k(x) \to H^i(Y_x, \mathcal{L}_{|Y_x})$$

is surjective. By Proposition 3.17, we deduce that $\theta_{x'}^i$ is an isomorphism for all x' in a neighborhood of x. We deduce that $R^i\tilde{\pi}_*\mathcal{L}_{|Y}$ is zero for all $i > 0$. Since f is flat, the base change theorem for the second cartesian diagram says that there is an isomorphism

$$f^* \circ R\pi_*\mathcal{L} \to R\tilde{\pi}_* \circ (f')^*\mathcal{L}.$$

Since f is faithfully flat, it implies that $R\pi_*\mathcal{L}$ is concentrated in degree 0. □

Proposition 3.20. *Let λ be a character of P_0. Denote by $\pi : Y_{I_0} \to \mathrm{Sh}$ the flag bundle defined before. We have a canonical isomorphism*

$$\pi_*\mathcal{L}_\lambda \simeq \nabla(\lambda),$$

where we see λ as a character of T to construct $\nabla(\lambda)$. This isomorphism extends to the toroidal compactifications $Y_{I_0}^{\mathrm{tor}}$ and $\mathrm{Sh}^{\mathrm{tor}}$.

Proof. This isomorphism is a formal consequence of the definition of automorphic vector bundles in Definitions 3.10 and 3.15 and standard base change theorem combined with Kempf's theorem. We have a cartesian diagram

$$\begin{tikzcd}
Y_{I_0} \arrow[r, "\tilde{\zeta}"] \arrow[d, "\pi"] & \lfloor P_0 \backslash * \rfloor \arrow[d, "\tilde{\pi}"] \\
\mathrm{Sh} \arrow[r, "\zeta"] & \lfloor P \backslash * \rfloor
\end{tikzcd}$$

where the horizontal arrows corresponds to the universal P-torsor on Sh and the universal P_0-torsor on Y_{I_0} and where the vertical arrow $\tilde{\pi}$ between the classifying stacks is induced by the inclusion $P_0 \subset P$. For every $\lambda \in X^*(P_0)$, we have a line bundle $\mathcal{L}_\lambda$ on the classifying stack of P_0. We denote by $\nabla(\lambda)$ the vector bundle on the classifying stack of P associated to the P-module $H^0(P/P_0, \mathcal{L}_\lambda)$. By definition, we have isomorphisms

$$\tilde{\pi}_*\mathcal{L}_\lambda = \nabla(\lambda), \quad \tilde{\zeta}^*\mathcal{L}_\lambda = \mathcal{L}_\lambda, \quad \zeta^*\nabla(\lambda) = \nabla(\lambda)$$

on $\lfloor P\backslash * \rfloor$. Since ζ is flat, we have a base change theorem in the derived category of quasicoherent sheaves over Sh which says that the natural map

$$\zeta^* \circ R\tilde{\pi}_* \mathcal{L}_\lambda \to R\pi_* \circ \tilde{\zeta}^* \mathcal{L}_\lambda$$

is an isomorphism. If λ is I_0-dominant, Kempf's vanishing theorem from Proposition 2.5 combined with Lemma 3.19 implies that

$$R\pi_* \mathcal{L}_\lambda = \pi_* \mathcal{L}_\lambda, \quad R\tilde{\pi}_* \mathcal{L}_\lambda = \tilde{\pi}_* \mathcal{L}_\lambda,$$

and we get

$$\pi_* \mathcal{L}_\lambda \simeq \nabla(\lambda).$$

For the toroidal compactifications, the proof is exactly the same. □

4. G-Zips and stratifications

Let Sh denote the Siegel modular variety over $\mathbb{F}_p$ of genus $g \geq 1$ and neat level $K \subset \mathrm{Sp}_{2g}(\mathbb{A}_f)$ such that K_p is hyperspecial. The Siegel modular variety Sh has the Ekedahl–Oort stratification (EO stratification) which is a genuine new structure which does not exist in characteristic 0. For the modular curve defined over $\mathbb{F}_p$, there are two strata: the ordinary locus and the supersingular locus. The ordinary locus is an open subscheme corresponding to ordinary elliptic curves over $\mathbb{F}_p$ and the supersingular locus is a reduced closed subscheme corresponding to supersingular elliptic curves over $\mathbb{F}_p$. Hence, the closure of the ordinary locus is the whole modular curve. In the series of papers [Wedhorn 1999; Moonen and Wedhorn 2004; Pink et al. 2011; 2015], Moonen, Wedhorn, Pink and Ziegler define an Artin stack $G\text{-Zip}^\mu$ which depends on the reductive group G over $\mathbb{F}_p$ and a cocharacter μ of G. The underlying topological space of this stack is finite and its topology captures the closure relations of the EO stratification. Furthermore, one can construct a morphism

$$\zeta : \mathrm{Sh} \to G\text{-Zip}^\mu$$

from the G-torsor $\mathcal{H}^1_{\mathrm{dR}}$ corresponding to de Rham cohomology of the universal abelian scheme. In his thesis, Zhang [2018] has proven that ζ is a smooth morphism. One can recover the EO stratification on Sh through a pullback of some substack w of $G\text{-Zip}^\mu$. Recall that P is the parabolic associated to $-\mu$ defined in Notation 3.5 and $I \subset \Delta$ is its type. The EO stratification on the Shimura variety can be further generalized on the flag bundle Y_{I_0} corresponding to a standard parabolic subgroup $P_0 \subset P$ of type $I_0 \subset I$. Goldring and Koskivirta [2019b] define a stack $G\text{-ZipFlag}^{\mu, I_0}$, a smooth morphism

$$\zeta_{I_0} : Y_{I_0} \to G\text{-ZipFlag}^{\mu, I_0}$$

and one can define a stratification of the flag bundle Y_{I_0} through the pullback of a collection of some substacks $[w]$ of $G\text{-ZipFlag}^{\mu, I_0}$. For the convenience of the reader, we recall how [Goldring and Koskivirta 2019a; 2019b] use the formalism of G-Zips and G-ZipFlags to define and study the stratifications on the

Siegel modular variety Sh and its flag bundle Y_{I_0}. In particular, we recall their result on the existence of generalized Hasse invariants.

4.1. ***General theory.*** In order to consider the stratification of the stack G-ZipFlag$^{\mu, I_0}$ for all $I_0 \subset I$, it is convenient to use a general zip datum $\mathcal{Z}$ and to define a stack G-Zip$^{\mathcal{Z}}$ for a general reductive group G over k, a field of positive characteristic.

Definition 4.1. A zip datum of exposant $n \geq 1$ is a tuple

$$\mathcal{Z} = (G, P, L, Q, M, \varphi^n),$$

where G is a reductive group over $\mathbb{F}_p$, $\varphi : G \to G$ is the relative Frobenius map and $P, Q \subset G \times_{\mathbb{F}_p} \overline{\mathbb{F}}_p$ are parabolics over $\overline{\mathbb{F}}_p$ with Levi subgroups $L \subset P$, $M \subset Q$ such that $\varphi^n(L) = M$. We write U and V for the unipotent radical of P and Q.

Definition 4.2. A morphism of zip data of exposant $n \geq 1$

$$\mathcal{Z} = (G, P, L, Q, M, \varphi^n) \to \mathcal{Z}' = (G', P', L', Q', M', \varphi'^n)$$

is the data of a group morphism $f : G \to G'$ such that $f(\Diamond) \subset \Diamond'$ for $\Diamond = G, P, L, Q, M, U, V$.

Recall that for each $g \geq 1$, we have a minuscule cocharacter $\mu : \mathbb{G}_m \to \mathrm{Sp}_{2g}$ defined over $\mathbb{F}_p$. The couple (Sp_{2g}, μ) (Sp_{2g} is defined over $\mathbb{F}_p$) is a cocharacter datum according to the following definition.

Definition 4.3. A cocharacter datum is a couple (G, μ) where G is a reductive group over $\mathbb{F}_p$ and $\mu : \mathbb{G}_m \to G$ is a cocharacter defined over $\overline{\mathbb{F}}_p$. A morphism of cocharacter data

$$(G, \mu) \to (G', \mu')$$

is a group morphism $f : G \to G'$ such that $\mu = f \circ \mu'$. A cocharacter data (G, μ) determines a opposite parabolic subgroup P_μ, $P_{-\mu}$ with common Levi subgroup $L = P_{-\mu} \cap P_\mu$.

From a cocharacter datum (G, μ) we can construct a zip datum of exposant n

$$\mathcal{Z}_\mu = (G, P, L, Q, M, \varphi^n)$$

by setting $P = P_{-\mu}$, $Q = \varphi^n(P_\mu)$, $L = P_{-\mu} \cap P_\mu$, $M = \varphi^n(L)$. We explain how to define a stack G-Zip$^{\mathcal{Z}}$ from a zip datum $\mathcal{Z}$.

Definition 4.4. Let $\mathcal{Z}$ be a zip datum and S be a scheme over $\mathbb{F}_p$. A zip of type $\mathcal{Z}$ over S is a tuple

$$\underline{I} = (\mathcal{I}, \mathcal{I}_P, \mathcal{I}_Q, \psi),$$

where $\mathcal{I}$ is a G-torsor over S, $\mathcal{I}_P \subset \mathcal{I}$ is a P-reduction of $\mathcal{I}$, $\mathcal{I}_Q \subset \mathcal{I}$ is a Q-reduction of $\mathcal{I}$ and

$$\psi : (\varphi^n)^*(\mathcal{I}_P/U) \to \mathcal{I}_Q/V$$

is an isomorphism of M-torsors over S. A morphism of zips of type $\mathcal{Z}$ over S

$$\underline{I} = (\mathcal{I}, \mathcal{I}_P, \mathcal{I}_Q, \psi) \to \underline{I}' = (\mathcal{I}', \mathcal{I}'_P, \mathcal{I}'_Q, \psi')$$

is a morphism of G-torsors $f : \mathcal{I} \to \mathcal{I}'$ over S such that $f(\Diamond) \subset \Diamond'$ for $\Diamond = \mathcal{I}_P, \mathcal{I}_Q$ and such that the following diagram commutes

$$\begin{tikzcd}
(\varphi^n)^*(\mathcal{I}_P/U) \arrow[r, "\psi"] \arrow[d] & \mathcal{I}_Q/V \arrow[d] \\
((\varphi')^n)^*(\mathcal{I}'_P/U') \arrow[r, "\psi'"] & \mathcal{I}'_Q/V'
\end{tikzcd}$$

where the vertical arrows are induced by f.

Proposition 4.5. *Let $\mathcal{Z}$ be a zip datum and S be a scheme over $\mathbb{F}_p$. The category $G\text{-Zip}^{\mathcal{Z}}(S)$ of zips of type $\mathcal{Z}$ over S is a groupoid. The association $S \to G\text{-Zip}^{\mathcal{Z}}(S)$ defines an algebraic stack over $\mathbb{F}_p$ that we simply denote $G\text{-Zip}^{\mathcal{Z}}$.*

Proof. See [Pink et al. 2015, Propositions 3.2 and 3.11]. □

Note that the association $\mathcal{Z} \to G\text{-Zip}^{\mathcal{Z}}$ defines a functor from the category of zip data to the category of algebraic stacks over $\mathbb{F}_p$. We simply write $G\text{-Zip}^{\mu}$ instead of $G\text{-Zip}^{\mathcal{Z}_\mu}$ when the zip datum comes from a cocharacter datum (G, μ). Most of the interesting properties of $G\text{-Zip}^{\mathcal{Z}}$ can be deduced from its presentation as a quotient stack. From now on, we fix a zip datum of exposant n

$$\mathcal{Z} = (G, P, L, Q, M, \varphi^n).$$

Proposition 4.6. *$G\text{-Zip}^{\mathcal{Z}}$ is a smooth stack of dimension 0 over $\mathbb{F}_p$ and it is presented as a quotient stack*

$$\lfloor E_{\mathcal{Z}} \backslash G \rfloor,$$

where $E_{\mathcal{Z}} = \{(x, y) \in P \times Q \mid \varphi^n(\bar{x}) = \bar{y}\}$, $x \to \bar{x}$ denotes the natural projection $P \to L$, $Q \to M$ and $(x, y) \in E_{\mathcal{Z}}$ acts on $g \in G$ by

$$(x, y)g = xgy^{-1}.$$

Proof. See [Pink et al. 2015, Propositions 3.2 and 3.11]. □

Denote by W the Weyl group of G, $I \subset \Delta$ the type of the parabolic P, $J \subset \Delta$ the type of the parabolic Q, $W_I \subset W$ the subgroup generated by the reflexions in I, ${}^I W$ the set of elements w that are of minimal length in $W_I w$, $W_J \subset W$ the subgroup generated by the reflexions in J and W^J the set of elements w that are of minimal length in wW_J. The element of maximal length in W (resp. W_I and W_J) is denoted w_0 (resp. $w_{0,I}$ and $w_{0,J}$). Denote by z the element $w_0 w_{0,J}$.

Proposition 4.7. *If there exists a Borel pair (B, T) of G defined over $\mathbb{F}_p$, then there exists an element $z \in W$ such that the triple (B, T, z) is a W-frame for $\mathcal{Z}$. It means that the following conditions are satisfied*:

(1) $B \subset P$.

(2) $zBz^{-1} \subset Q$.

(3) $\varphi(B \cap L) = zBz^{-1} \cap M$.

Proof. See the proof of [Pink et al. 2011, Proposition 3.7]. □

For each $w \in W$, we choose a lift $\dot{w}$ in $N_G(T)$. The following proposition explains how G decomposes in $E_{\mathcal{Z}}$-orbits.

Proposition 4.8. *The map $w \mapsto G_w := E_{\mathcal{Z}} \dot{w} \dot{z}^{-1}$ restricts to bijections between:*[10]

(1) *${}^I W$ and the $E_{\mathcal{Z}}$-orbits of G.*

(2) *W^J and the $E_{\mathcal{Z}}$-orbits of G.*

Moreover, we have the following dimension formula for all $w \in {}^I W \cup W^J$

$$\dim G_w = l(w) + \dim(P).$$

Proof. See [Pink et al. 2011, Theorems 7.5 and 11.2]. □

Corollary 4.9. *The stack $G\text{-Zip}^{\mathcal{Z}}$ can be decomposed as*

$$G\text{-Zip}^{\mathcal{Z}} = \bigsqcup_{w \in {}^I W} \lfloor E_{\mathcal{Z}} \backslash G_w \rfloor.$$

The stack $G\text{-Zip}^{\mathcal{Z}}$ has a topology which describes the closure relations between the G_w.

Proposition 4.10. *The underlying topological space of $G\text{-Zip}^{\mathcal{Z}}$ is homeomorphic to the finite topological space ${}^I W$ where the topology is given by the partial order*

$$w \preccurlyeq w' \quad \textit{if and only if there is} \quad v \in W_I \textit{ such that } vwxv^{-1}x^{-1} \leq w',$$

where x is the unique element of minimal length in $W_J w_0 W_I$ and $\leq$ is the Bruhat order.

Proof. The result follows from the isomorphism

$$\overline{G_w} = \bigsqcup_{w' \preccurlyeq w,\ w' \in {}^I W} G_{w'}$$

for all $w \in {}^I W$ which is proven in [Pink et al. 2011, Theorem 6.2]. □

We simply write $[w]$ for the locally closed substack $\lfloor E_{\mathcal{Z}} \backslash G_w \rfloor$ of $G\text{-Zip}^{\mathcal{Z}}$. Now, we describe how to define a more general stack $G\text{-ZipFlag}^{\mathcal{Z}, P_0}$ which depends on the zip datum $\mathcal{Z}$ and an auxiliary parabolic subgroup $B \subset P_0 \subset P$.

Definition 4.11. Let $B \subset P_0 \subset P$ be a parabolic subgroup of P and S be a scheme over $\mathbb{F}_p$. A zip flag of type $(\mathcal{Z}, P_0)$ over S is a tuple

$$\underline{J} = (\underline{I}, \mathcal{J}),$$

where $\underline{I} = (\mathcal{I}, \mathcal{I}_P, \mathcal{I}_Q, \psi)$ is a zip of type $\mathcal{Z}$ over S and $\mathcal{J} \subset \mathcal{I}_P$ is a P_0-reduction of the P-torsor $\mathcal{I}_P$. A morphism of zip flags of type $(\mathcal{Z}, P_0)$ over S

$$\underline{J} = (\underline{I}, \mathcal{J}) \to \underline{J}' = (\underline{I}', \mathcal{J}')$$

[10] In the case $G = \mathrm{Sp}_{2g}$, these two bijections coincide.

is a morphism of zip $\underline{I} \to \underline{I}'$ of type $\mathcal{Z}$ over S such that the underlying morphism of G-torsor $\mathcal{I} \to \mathcal{I}'$ restricts to a morphism of P_0-torsor $\mathcal{J} \to \mathcal{J}'$ over S.

Proposition 4.12. *Let $B \subset P_0 \subset P$ be a parabolic subgroup of P and S be a scheme over $\mathbb{F}_p$. The category G-*ZipFlag$^{\mathcal{Z},P_0}(S)$ *of zip flags of type $(\mathcal{Z}, P_0)$ over S is a groupoid. The association $S \to$ G-*ZipFlag$^{\mathcal{Z},P_0}(S)$ *defines an algebraic stack over $\mathbb{F}_p$ that we simply denote G-*ZipFlag$^{\mathcal{Z},P_0}$.

From now on, we fix an auxiliary parabolic subgroup $B \subset P_0 \subset P$.

Proposition 4.13. *The stack G-*ZipFlag$^{\mathcal{Z},P_0}$ *is a smooth stack of dimension* $\dim(P/P_0)$ *over $\mathbb{F}_p$ and it can be presented as the quotient stack*

$$\lfloor E_{\mathcal{Z},P_0} \backslash G \rfloor,$$

where $E_{\mathcal{Z},P_0} := E_{\mathcal{Z}} \cap (P_0 \times G) \subset P_0 \times Q$ acts on G by restriction of the $E_{\mathcal{Z}}$-action on G. It can also be presented as the quotient stack

$$\lfloor E_{\mathcal{Z}} \times P_0 \backslash G \times P \rfloor,$$

where $((x, y), p_0) \in E_{\mathcal{Z}} \times P_0$ acts on $(g, p) \in G \times P$ through the formula

$$((x, y), p_0).(g, p) = (xgy^{-1}, xpp_0^{-1}).$$

Definition 4.14. The map sending a zip flag $\underline{J} = (\underline{I}, \mathcal{J})$ of type $(\mathcal{Z}, P_0)$ over S to the zip $\underline{I}$ of type $\mathcal{Z}$ over S defines a morphism of algebraic stacks over $\mathbb{F}_p$

$$\pi : G\text{-ZipFlag}^{\mathcal{Z},P_0} \to G\text{-Zip}^{\mathcal{Z}}.$$

Proposition 4.15. *The inclusion $E_{\mathcal{Z},P_0} \subset E_{\mathcal{Z}}$ induces a morphism*

$$\lfloor E_{\mathcal{Z},P_0} \backslash G \rfloor \to \lfloor E_{\mathcal{Z}} \backslash G \rfloor$$

which corresponds to π through the isomorphisms in Proposition 4.13.

Proposition 4.16. *The morphism $\pi : G$-*ZipFlag$^{\mathcal{Z},P_0} \to G$-Zip$^{\mathcal{Z}}$ *is proper and smooth with fibers isomorphic to the flag variety P_0/P.*

Proof of Propositions 4.12, 4.13, 4.15 and 4.16. See [Goldring and Koskivirta 2019b, Theorem 2.1.2]. □

It is natural to hope for a stratification of G-ZipFlag$^{\mathcal{Z},P_0}$ that generalizes the stratification on G-Zip$^{\mathcal{Z}}$ however the $E_{\mathcal{Z},P_0}$-orbits of G are not as easy to understand as the $E_{\mathcal{Z}}$-orbits. Instead, we define a smooth surjective map

$$G\text{-ZipFlag}^{\mathcal{Z},P_0} \to G\text{-Zip}^{\mathcal{Z}_0},$$

where $\mathcal{Z}_0$ is a zip datum constructed from $\mathcal{Z}$ and P_0 and then pullback the stratification of G-Zip$^{\mathcal{Z}_0}$.

Definition 4.17. We denote by $\mathcal{Z}_0$ the zip datum

$$\mathcal{Z}_0 = (G, P_0, L_0, Q_0, M_0, \varphi^n)$$

where Q_0 is a parabolic subgroup of Q defined by

$$Q_0 = \varphi^n(P_0 \cap L)R_u(Q) \subset Q,$$

with $R_u(Q)$ the unipotent radical of Q and where L_0, M_0 are the Levi subgroups of P_0, Q_0.

Proposition 4.18. *We have inclusions*

$$E_{\mathcal{Z},P_0} \subset E_{\mathcal{Z}_0} \subset P_0 \times Q_0$$

and the induced maps

$$G\text{-ZipFlag}^{\mathcal{Z},P_0} \xrightarrow{\psi_1} G\text{-Zip}^{\mathcal{Z}_0} \xrightarrow{\psi_2} \lfloor P_0 \backslash G/Q_0 \rfloor$$

are smooth and surjective.

Proof. See [Goldring and Koskivirta 2019b, Section 3.1]. □

Definition 4.19. The fine stratification of G-ZipFlag$^{\mathcal{Z},P_0}$ is the stratification of G-Zip$^{\mathcal{Z}_0}$ pulled back by ψ_1 and the coarse stratification of G-ZipFlag$^{\mathcal{Z},P_0}$ is the stratification of the Bruhat stack $\lfloor P_0 \backslash G/Q_0 \rfloor$ pulled back by $\psi_2 \circ \psi_1$. If $w \in {}^{I_0}W \cup W^{J_0}$, then we write G-ZipFlag$_w^{\mathcal{Z},P_0}$ for the corresponding fine strata.

In the special case where $P_0 = B$ is the Borel subgroup, the map ψ_2 is an isomorphism, so the coarse and the fine stratifications of G-ZipFlag$^{\mathcal{Z},P_0}$ coincide. Note that if we have an inclusion of auxiliary parabolics $B \subset P_0 \subset P_1 \subset P$, then there exists natural maps making the following diagram 2-cartesian:

$$\begin{tikzcd}
G\text{-ZipFlag}^{\mathcal{Z},P_0} \arrow[r] \arrow[d] & \lfloor P_0 \backslash G/Q_0 \rfloor \arrow[d] \\
G\text{-ZipFlag}^{\mathcal{Z},P_1} \arrow[r] & \lfloor P_1 \backslash G/Q_1 \rfloor
\end{tikzcd}$$

However, we don't know if a similar statement holds if we replace the Bruhat stacks with G-Zip$^{\mathcal{Z}_0}$ and G-Zip$^{\mathcal{Z}_1}$.

Corollary 4.20. *The stack G-ZipFlag$^{\mathcal{Z},P_0}$ can be decomposed as*

$$G\text{-ZipFlag}^{\mathcal{Z},P_0} = \bigsqcup_{w \in {}^{I_0}W} G\text{-ZipFlag}_w^{\mathcal{Z},P_0}$$

and for all $w \in {}^{I_0}W$, we have the closure relation

$$\overline{G\text{-ZipFlag}_w^{\mathcal{Z},P_0}} = \bigsqcup_{w' \preccurlyeq w,\ w' \in {}^{I_0}W} G\text{-ZipFlag}_{w'}^{\mathcal{Z},P_0},$$

where the order on ${}^{I_0}W$ is the one introduced in Proposition 4.10 with I replaced by I_0.

Corollary 4.21. *Let $w \in {}^{I_0}W \cup W^{J_0}$ and G-ZipFlag$_w^{\mathcal{Z},P_0}$ be the corresponding fine strata. Then G-ZipFlag$_w^{\mathcal{Z},P_0}$ is a smooth stack over $\mathbb{F}_p$ of pure dimension $l(w) + \dim P - \dim G$.*

Now we want to construct some sections of vector bundles on G-Zip$^{\mathcal{Z}}$, G-ZipFlag$^{\mathcal{Z},P_0}$ and relate their nonvanishing loci to the stratification we have introduced. We start by introducing vector bundles on G-Zip$^{\mathcal{Z}}$.

Definition 4.22. Let $\rho : L \to \mathrm{GL}(V)$ be a finite-dimensional algebraic representation of the Levi L. Consider the map $f : E_{\mathcal{Z}} \to L$ which is the composition of the first projection $E_{\mathcal{Z}} \to P$ with the quotient map $P \to L$. The composition $\rho \circ f$ is an algebraic representation of $E_{\mathcal{Z}}$. It induces a locally free sheaf $\mathcal{W}(V)$ of rank $\dim_{\mathbb{F}_p} V$ on $\lfloor E_{\mathcal{Z}} \backslash G \rfloor$. If $\lambda \in X^*(T)$ is a I-dominant character of T, we simply denote $\nabla(\lambda)$ the locally free sheaf $\mathcal{W}(H^0(L/B_L, \mathcal{L}_\lambda))$. Note that $H^0(L/B_L, \mathcal{L}_\lambda)$ is the costandard L-representation of highest weight $w_0 w_{0,L} \lambda$.

More generally, we can define vector bundles on $G\text{-ZipFlag}^{\mathcal{Z},P_0}$.

Definition 4.23. Let $\rho : P_0 \to \mathrm{GL}(V)$ be a finite-dimensional algebraic representation of the parabolic P_0. Consider the first projection map $f : E_{\mathcal{Z},P_0} \to P_0$. The composition $\rho \circ f$ is an algebraic representation of $E_{\mathcal{Z},P_0}$ and it induces a locally free sheaf $\mathcal{L}(V)$ of rank $\dim_{\overline{\mathbb{F}}_p} V$ on $\lfloor E_{\mathcal{Z},P_0} \backslash G \rfloor$. If $\lambda \in X^*(L_0) \subset X^*(T)$ is a character of L_0, we also denote by $\mathcal{L}_\lambda$ the line bundle $\mathcal{L}(\lambda)$ where we see λ as a one-dimensional representation of P_0.

We have defined vector bundles $\nabla(\lambda)$ on $G\text{-Zip}^{\mathcal{Z}}$ and line bundles $\mathcal{L}_\lambda$ on $G\text{-ZipFlag}^{\mathcal{Z},P_0}$ for certain characters $\lambda \in X^*(T)$. The next proposition gives a direct relation between them.

Proposition 4.24. *Recall that $\pi : G\text{-ZipFlag}^{\mathcal{Z},P_0} \to G\text{-Zip}^{\mathcal{Z}}$ is the proper and smooth map that forgets the P_0-torsor from a zip flag of type $(\mathcal{Z}, P_0)$. Let $\lambda \in X^*(L_0)$ be an I_0-dominant character of L_0. We have a canonical isomorphism*

$$\pi_* \mathcal{L}_\lambda \simeq \nabla(\lambda).$$

Proof. Consider the cartesian diagram

$$\begin{tikzcd} G\text{-ZipFlag}^{\mathcal{Z},P_0} \arrow[r, "\tilde{\zeta}"] \arrow[d, "\pi"] & \lfloor */P_0 \rfloor \arrow[d, "\tilde{\pi}"] \\ G\text{-Zip}^{\mathcal{Z}} \arrow[r, "\zeta"] & \lfloor */P \rfloor \end{tikzcd}$$

where the horizontal maps are given by the universal P_0-torsor on $G\text{-ZipFlag}^{\mathcal{Z},P_0}$ and the universal P-torsor on $G\text{-Zip}^{\mathcal{Z}}$. For each character $\lambda \in X^*(P_0)$, we have a line bundle $\mathcal{L}_\lambda$ on $\lfloor */P_0 \rfloor$ and a vector bundle $\nabla(\lambda)$ on $\lfloor */P \rfloor$ (corresponding to the induced P-representation $H^0(P/P_0, \mathcal{L}_\lambda)$) that satisfies

$$\tilde{\zeta}^* \mathcal{L}_\lambda = \mathcal{L}_\lambda, \quad \zeta^* \nabla(\lambda) = \nabla(\lambda).$$

It is straightforward from the definitions that

$$\tilde{\pi}_* \mathcal{L}_\lambda = \nabla(\lambda).$$

As the map is fibered in P/P_0 by Proposition 4.16, we know by Proposition 2.5 and Lemma 3.19 that for a I_0-dominant character λ, we have

$$R\tilde{\pi}_* \mathcal{L}_\lambda = \tilde{\pi}_* \mathcal{L}_\lambda, \quad R\pi_* \mathcal{L}_\lambda = \pi_* \mathcal{L}_\lambda.$$

Since ζ is flat, we conclude as in the end of the proof of Proposition 3.20 with the base change theorem in the derived category that says that the natural map

$$\zeta^* \circ R\tilde{\pi}_* \mathcal{L}_\lambda \to R\pi_* \circ \zeta^* \mathcal{L}_\lambda$$

is an isomorphism. □

On the Bruhat stack $\mathrm{Brh} = \lfloor B \backslash G/B \rfloor$, we have the Bruhat stratification

$$\mathrm{Brh} = \bigsqcup_{w \in W} \mathrm{Brh}_w,$$

where $\mathrm{Brh}_w = \lfloor B \backslash BwB/B \rfloor$ and for all $w \in W$ we have the closure relation

$$\overline{\mathrm{Brh}_w} = \bigsqcup_{w' \leq w,\, w' \in W} \mathrm{Brh}_{w'},$$

where $\leq$ is the Bruhat order. We consider the morphism

$$\psi : G\text{-ZipFlag}^{\mathcal{Z},B} \to \mathrm{Brh}$$

defined as the composition of the morphism induced by the inclusion

$$E_{\mathcal{Z},B} \subset B \times zBz^{-1}$$

with the isomorphism

$$\alpha_z : \lfloor B \backslash G/zBz^{-1} \rfloor \to \lfloor B \backslash G/B \rfloor,$$

that sends x to xz. We use this stack to construct some sections on $G\text{-ZipFlag}^{\mathcal{Z},P_0}$.

Proposition 4.25. *Given two characters $(\lambda, \eta) \in X^*(T) \times X^*(T)$, the associated line bundle $\mathcal{L}_{\lambda,\eta}$ on* Brh *has the following properties*:

(1) *We have a canonical isomorphism $\psi^* \mathcal{L}_{\lambda,\eta} = \mathcal{L}_{\lambda + p^\sigma(z\eta)}$, where $\sigma : \overline{\mathbb{F}}_p \to \overline{\mathbb{F}}_p$ is the inverse of the Frobenius.*

(2) *For all $w \in W$, we have $H^0(\mathrm{Brh}_w, \mathcal{L}_{\lambda,\eta}) \neq 0 \Leftrightarrow \eta = -w^{-1}\lambda$.*

(3) $\dim_{\overline{\mathbb{F}}_p} H^0(\mathrm{Brh}_w, \mathcal{L}_{\lambda,-w^{-1}\lambda}) = 1$.

(4) *For any nonzero $s \in H^0(\mathrm{Brh}_w, \mathcal{L}_{\lambda,-w^{-1}\lambda})$ viewed as a rational function on $\overline{\mathrm{Brh}}_w$, one has*

$$\mathrm{div}(s) = -\sum_{\alpha \in E_w} \langle \lambda, w\alpha^\vee \rangle \overline{\mathrm{Brh}}_{ws_\alpha},$$

where $E_w = \{\alpha \in \phi^+ \mid ws_\alpha < w \text{ and } l(ws_\alpha) = l(w) - 1\}$. The set of ws_α for $\alpha \in E_w$ is called the set of lower neighbors of w.

Proof. For (i), see [Goldring and Koskivirta 2019a, Lemma 3.1.1]. For (ii) to (iv), see [loc. cit., Theorem 2.2.1]. □

Definition 4.26. Let $w \in W$ and $n \geq 0$. We define by induction on n, the element $w^{(n)}$ by setting

(1) $w^{(0)} = e$,

(2) $w^{(n)} = {}^\sigma(w^{(n-1)}w)$ if $n \geq 1$.

Proposition 4.27. *The function*

$$\begin{aligned} D_w : X^*(T) &\to X^*(T), \\ \lambda &\mapsto \lambda - p\,{}^{\sigma}(zw^{-1}\lambda), \end{aligned}$$

induces a $\mathbb{Q}$-linear automorphism of $X^*(T) \otimes_{\mathbb{Z}} \mathbb{Q}$. If χ is a character, its inverse by D_w is given by the $\mathbb{Q}$-character

$$\lambda = \frac{-1}{p^{rn} - 1} \sum_{i=0}^{rn-1} p^i (zw^{-1})^{(i)}\,{}^{\sigma^i}\chi,$$

where r is an integer such that $(zw^{-1})^{(r)} = e$ and n is an integer such that χ is defined over $\mathbb{F}_{p^n}$.

Proof. See [Goldring and Koskivirta 2019a, Lemma 3.1.3]. □

Having defined sections on stacks G-$\text{Zip}^{\mathcal{Z}}$, we can study their vanishing locus.

Definition 4.28. Let $\lambda \in X^*(P_0)$ be a L_0-dominant character of P_0 and

$$s \in H^0(G\text{-ZipFlag}_w^{\mathcal{Z},P_0}, \mathcal{L}_\lambda),$$

a nonzero section. We say that s is a generalized Hasse invariant for G-$\text{ZipFlag}_w^{\mathcal{Z},P_0}$ if there exists some $d \geq 1$ such that s^d extends to $\overline{G\text{-ZipFlag}_w^{\mathcal{Z},P_0}}$ with nonvanishing locus G-$\text{ZipFlag}_w^{\mathcal{Z},P_0}$. We define the sets

$$\mathcal{C}_{\mathrm{Ha},I_0,w} = \{\lambda \in X^*(P_0) \mid \mathcal{L}_\lambda \text{ has a generalized Hasse invariant for } G\text{-ZipFlag}_w^{\mathcal{Z},P_0}\}$$

and

$$\mathcal{C}_{\mathrm{Ha},I_0} = \bigcap_{w \in W} \mathcal{C}_{\mathrm{Ha},I_0,w}.$$

Now, we give a strong result for the existence of generalized Hasse invariants on the stack G-$\text{ZipFlag}^{\mathcal{Z},P_0}$.

Proposition 4.29. *Let $\lambda \in X^*(P_0)$ be a L_0-dominant character, w be an element of ${}^{I_0}W$ and s be a nonzero section of $H^0(G\text{-ZipFlag}_w^{\mathcal{Z},P_0}, \mathcal{L}_\lambda)$. Then, the following statements are equivalent*:

(1) *s is a generalized Hasse invariant for G-$\text{ZipFlag}_w^{\mathcal{Z},P_0}$.*

(2) *For all $\alpha \in E_w$, we have*

$$\sum_{i=0}^{rn-1} \langle (zw^{-1})^{(i)}({}^{\sigma^i}\lambda), w\alpha^\vee \rangle p^i > 0,$$

where r is an integer such that $(zw^{-1})^{(r)} = e$ and n is an integer such that λ is defined over $\mathbb{F}_{p^n}$.

Proof. See [Goldring and Koskivirta 2019a, Proposition 3.2.1]. □

Example 4.30. We give more details in the case $G = \mathrm{Sp}_4$ and $\mathcal{Z} = \mathcal{Z}_\mu$ with μ the cocharacter that stabilizes the Hodge filtration of the Siegel datum. The Levi L is GL_2. We denote by s_1 and s_2 the simple reflections associated to the simple roots $(1,-1)$ and $(0,2)$. We represent the elements of W in the diagram

$$\begin{tikzcd}
 & w_0 = s_2s_1s_2s_1 \arrow[dl] \arrow[dr] & \\
w_1 = s_2s_1s_2 \arrow[d] \arrow[drr] & & w'_1 = s_1s_2s_1 \arrow[d] \arrow[dll] \\
w_2 = s_2s_1 \arrow[d] \arrow[drr] & & w'_2 = s_1s_2 \arrow[d] \arrow[dll] \\
w_3 = s_2 \arrow[dr] & & w'_3 = s_1 \arrow[dl] \\
 & e &
\end{tikzcd}$$

where an arrow is drawn from w to w' if $w' \leq w$ and $l(w') = l(w) - 1$. For each $w \in W$ and $\lambda \in X^*$, we denote by $s_{\lambda,w}$ the quasisection in $H^0(G\text{-ZipFlag}_w^{\mathcal{Z},B}, \mathcal{L}_\lambda)$ obtained via pullback from a nonzero quasisection of $H^0(\mathrm{Brh}_w, \mathcal{L}_{\chi,-w^{-1}\chi})$ where χ is a $\mathbb{Q}$-character such that $D_w(\chi) = \lambda$.[11] We write $\lambda = (k_1, k_2)$ and we compute $\operatorname{div}(s_{\lambda,w})$ for each w:

$$\begin{aligned}
\operatorname{div}(s_{\lambda,w_0}) &= \frac{1}{p^2-1}((p-1)(k_1-k_2)\overline{[w_1]} - (k_2+pk_1)\overline{[w'_1]}),\\
\operatorname{div}(s_{\lambda,w_1}) &= \frac{1}{p-1}(-k_1\overline{[w_2]} - k_2\overline{[w'_2]}),\\
\operatorname{div}(s_{\lambda,w'_1}) &= \frac{1}{p^2+1}(-((p-1)k_1+(p+1)k_2)\overline{[w_2]} + ((p+1)k_1-(p-1)k_2)\overline{[w'_2]}),\\
\operatorname{div}(s_{\lambda,w_2}) &= \left(\frac{k_1}{p+1} - \frac{k_2}{p-1}\right)\overline{[w_3]} + \frac{-k_2}{p-1}\overline{[w'_3]},\\
\operatorname{div}(s_{\lambda,w'_2}) &= -\left(\frac{k_1}{p-1} + \frac{k_2}{p+1}\right)\overline{[w_3]} + \frac{-k_1}{p-1}\overline{[w'_3]},\\
\operatorname{div}(s_{\lambda,w_3}) &= \frac{1}{p^2+1}(k_2-pk_1)\overline{[e]},\\
\operatorname{div}(s_{\lambda,w'_3}) &= \frac{1}{p+1}(k_1-k_2)\overline{[e]}.
\end{aligned}$$

We deduce that the set of characters

$$\mathcal{C}_1 := \left\{\lambda = (k_1,k_2) \;\middle|\; 0 > k_1 > \frac{p-1}{p+1}k_2 \text{ and } k_2 > pk_1\right\}$$

[11] Quasisection means a section of a certain positive tensorial power of $\mathcal{L}_\lambda$.

has generalized Hasse invariants for all strata of G-$\mathrm{ZipFlag}_w^{\mathcal{Z},B}$. Since the character of the Levi are $X^*(L) = \mathbb{Z}(1, 1)$ and the minimal length left coset representatives are

$$^I W = \{e, w_3, w_2, w_1\},$$

we deduce that the following set of characters

$$\mathcal{C}_2 := \{\lambda = (k_1, k_1) \mid k_1 < 0\}$$

has generalized Hasse invariant for all strata of G-$\mathrm{Zip}^{\mathcal{Z}}$.

Example 4.31. We give some details in the case $G = \mathrm{Sp}_6$. The Levi L is GL_3 and we denote by s_1, s_2 and s_3 the simple reflections associated to the simple roots $(1, -1, 0)$, $(0, 1, -1)$ and $(0, 0, 2)$. The Weyl group W is isomorphic to $S_3 \ltimes (\mathbb{Z}/2\mathbb{Z})^3$ (48 elements). The computations can be painful without a computer, so we have implemented an algorithm in SageMath that computes the divisor of all the $s_{w,\lambda}$ for any $g \geq 2$. See github.com/ThibaultAlexandre/generalized-hasse-invariants to download the algorithm. Take $p = 7$, $w = s_1 s_2 s_3$, and $\lambda = (-1, -3, -5)$. We get

$$\mathrm{div}(s_{w,\lambda}) = \tfrac{5}{6}[\overline{s_2 s_3}] + \tfrac{1}{2}[\overline{s_3 s_1}] + \tfrac{1}{6}[\overline{s_1 s_2}].$$

Koskivirta and Goldring introduced a notion called orbitally p-closeness that guarantees a character to have generalized Hasse invariants for all strata without having to compute all the $\mathrm{div}(s_{w,\lambda})$. However, this notion is not necessary for a character to have Hasse invariants.

Definition 4.32. Let λ be a character of T. For every coroot such that $\langle \lambda, \alpha^\vee \rangle \neq 0$, we set

$$\mathrm{Orb}(\lambda, \alpha^\vee) = \left\{ \frac{|\langle \lambda, w\alpha^\vee \rangle|}{|\langle \lambda, \alpha^\vee \rangle|} \;\middle|\; w \in W \rtimes \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p) \right\}$$

and we say that λ is

(1) orbitally p-close if $\max_{\alpha \in \phi} \mathrm{Orb}(\lambda, \alpha^\vee) \leq p - 1$,

(2) $\mathcal{Z}_0$-ample if $\langle \lambda, \alpha^\vee \rangle > 0$ for all $\alpha \in I \backslash I_0$ and $\langle \lambda, \alpha^\vee \rangle < 0$ for all $\alpha \in \phi^+ \backslash \phi_L^+$.

Proposition 4.33. *Let λ be a character of P_0. If λ is orbitally p-close and $\mathcal{Z}_0$-ample then, there exists $d \geq 1$ such that for all $w \in {}^{I_0}W$ and all nonzero section s in*

$$H^0(G\text{-}\mathrm{ZipFlag}_w^{\mathcal{Z},P_0}, \mathcal{L}_\lambda),$$

the d-th power s^d extends to $\overline{G\text{-}\mathrm{ZipFlag}_w^{\mathcal{Z},P_0}}$ with nonvanishing locus G-$\mathrm{ZipFlag}_w^{\mathcal{Z},P_0}$.

Proof. See [Goldring and Koskivirta 2019a, Proposition 3.2.3]. □

4.2. *G-Zip associated to the universal abelian scheme.*

In this subsection, we specialize our discussion to the Siegel case. Recall that the Siegel modular variety Sh is a smooth scheme over $k = \mathbb{F}_p$. We denote by $\pi : Y_{I_0} \to \mathrm{Sh}$ the Siegel flag bundle of type $I_0 \subset I$. Recall that π extends to the toroidal compactifications $\pi : Y_{I_0}^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$. We also have a minimal compactification $\mathrm{Sh}^{\mathrm{min}}$ for the Shimura variety but not for the flag bundle. The goal of this subsection is to define the maps ζ and ζ_{I_0}. We need some results on the Hodge and conjugate filtrations of abelian schemes. We recall a result due to Deligne and Illusie.

Proposition 4.34. *Let S be a scheme of characteristic p. Let $f : A \to S$ be an abelian scheme over S. Consider the Hodge to de Rham spectral sequence*

$$E_2^{i,j} = R^j f_* \Omega^i_{A/S} \quad \Longrightarrow \quad H^{i+j}_{\mathrm{dR}}(A/S).$$

and the conjugate spectral sequence

$$E'^{i,j}_1 = R^i f_*(\mathcal{H}^j(\Omega^\bullet_{A/S})) \quad \Longrightarrow \quad H^{i+j}_{\mathrm{dR}}(A/S).$$

Then

(1) *$E_2^{i,j}$ degenerates at page* 2,

(2) *$E'^{i,j}_1$ degenerates at page* 1.

Proof. See [Deligne and Illusie 1987, Corollaire 2.4 and Remarques 2.6(iv)]. □

Definition 4.35. Let S be a scheme of characteristic p. Let $f : A \to S$ be an abelian scheme over S. The Hodge filtration of A over S is the two-step underlying filtration on $H^1_{\mathrm{dR}}(A/S)$ coming from the degeneration of the Hodge to de Rham spectral sequence

$$0 \to \pi_*\Omega^1_{A/S} \to H^1_{\mathrm{dR}}(A/S) \to R^1\pi_*\mathcal{O}_A \to 0.$$

Definition 4.36. The conjugate filtration is the two-step underlying filtration on $H^1_{\mathrm{dR}}(A/S)$ coming from the degeneration of the conjugate spectral sequence

$$0 \to R^1\pi_*\mathcal{H}^0(\Omega^\bullet_{A/S}) \to H^1_{\mathrm{dR}}(A/S) \to \pi_*\mathcal{H}^1(\Omega^\bullet_{A/S}) \to 0.$$

The Hodge and the conjugate filtration are related on their graded pieces by the Cartier isomorphism which of we recall the definition.

Definition 4.37. Let S be a scheme of characteristic p and $f : A \to S$ be a smooth morphism.[12] The Cartier morphism is a map of graded algebras

$$C^{-1} : \bigoplus_i \Omega^i_{A^{(p)}/S} \to \bigoplus_i \mathcal{H}^i(F_*\Omega^\bullet_{A/S}),$$

where $F : A \to A^{(p)}$ is the relative geometric Frobenius of A over S. It is enough to define it in degree 0 and 1 and then use the graded algebra structure to extend it. In degree 0, it is the map $F^* : \mathcal{O}_{A^{(p)}} \to F_*\mathcal{O}_A$. In degree 1, it is a map

$$\Omega^1_{A^{(p)}/S} \to \mathcal{H}^1(F_*\Omega^\bullet_{A/S})$$

coming from the S-derivation $\delta : \mathcal{O}_{A^{(p)}} \to \mathcal{H}^1(F_*\Omega^\bullet_{A/S})$ satisfying (we use the isomorphism $\mathcal{O}_{A^{(p)}} = \mathcal{O}_A \otimes_{\mathcal{O}_S, F^*} \mathcal{O}_S$)

(1) $\delta(fs \otimes s') = \delta(f \otimes s^p s')$,

(2) $\delta(fg \otimes s') = f^p\delta(g \otimes s) + g^p\delta(f \otimes s)$

[12] Such as an abelian scheme A over S.

for all $f, g \in \mathcal{O}_A$ and $s, s' \in \mathcal{O}_S$. If $f \in \mathcal{O}_A$ and $s \in \mathcal{O}_S$, we define $\delta(f \otimes s)$ to be the cohomology class of $sf^{p-1}df$.

Proposition 4.38 [Cartier 1957]. *The Cartier morphism C^{-1} is an isomorphism and it satisfies*:

(1) $C^{-1}(1) = 1$.

(2) $C^{-1}(w \wedge w') = C^{-1}(w) \wedge C^{-1}(w')$ *for all* $w \in \Omega^i_{A^{(p)}/S}$, $w' \in \Omega^{i'}_{A^{(p)}/S}$.

(3) $C^{-1}(d(f \otimes 1)) = [f^{p-1}df]$.

The Hodge filtration and the conjugate filtration of an abelian scheme $f : A \to S$ of relative dimension g can be seen as a P-reduction $\mathcal{I}$ and a Q-reduction $\mathcal{I}$ of the $G = \mathrm{Sp}_{2g}$-torsor $H^1_{\mathrm{dR}}(A/S)$ where P and Q are the maximal parabolic subgroups associated to the cocharacter datum (Sp_{2g}, μ).[13] With the Cartier isomorphism, we can construct an isomorphism

$$\psi : (\varphi)^*(\mathcal{I}_P/U) \to \mathcal{I}_Q/V$$

of M-torsors. In other words, we can construct a zip $(H^1_{\mathrm{dR}}(A/S), \mathcal{I}_P, \mathcal{I}_Q, \psi)$ over S of type $\mathcal{Z} = (G, P, L, Q, M, \varphi)$.

Definition 4.39. The morphism

$$\zeta : \mathrm{Sh} \to G\text{-}\mathrm{Zip}^{\mathcal{Z}}$$

is the classifying map of the universal zip $\underline{I} = (\mathcal{H}^1_{\mathrm{dR}}, \mathcal{I}_P, \mathcal{I}_Q, \psi)$ associated to the universal abelian scheme $f : A \to \mathrm{Sh}$ over Sh. For all $w \in {}^I W$, we define the locally closed subscheme $\mathrm{Sh}_w := \zeta^{-1}(G\text{-}\mathrm{Zip}^{\mathcal{Z}}_w)$.

Over Y_{I_0}, we have a universal P_{I_0}-reduction $\mathcal{J}$ of the P-torsor $\mathcal{I}_P$ corresponding to the Hodge filtration. The pair $(\underline{I}, \mathcal{J})$ is a zip flag of type $(\mathcal{Z}, I_0)$.

Definition 4.40. The morphism

$$\zeta_{I_0} : Y_{I_0} \to G\text{-}\mathrm{ZipFlag}^{\mathcal{Z}, I_0}$$

is the classifying map of the universal zip flag $(\underline{I}, \mathcal{J})$ of type $(\mathcal{Z}, I_0)$. For all $w \in {}^{I_0}W$, we define the locally closed subscheme $(Y_{I_0})_w := \zeta_{I_0}^{-1}(G\text{-}\mathrm{ZipFlag}^{\mathcal{Z}}_w)$.

Proposition 4.41. *The morphisms ζ and ζ_{I_0} are smooth and surjective.*

Proof. See [Zhang 2018, Theorem 3.1.2] for the smoothness. See [Oort 2001] for the surjectivity. □

As generalizations lift along flat morphisms [Stacks 2005–, Tag 03HV], we deduce in particular that we have the following closure relations:

(1) For all $w \in {}^I W$, $\overline{\mathrm{Sh}}_w = \displaystyle\bigsqcup_{w' \preccurlyeq w,\, w' \in {}^I W} \mathrm{Sh}_{w'}$.

(2) For all $w \in {}^{I_0} W$, $\overline{(Y_{I_0})_w} = \displaystyle\bigsqcup_{w' \preccurlyeq w,\, w' \in {}^{I_0} W} (Y_{I_0})_{w'}$.

We give a statement about the extension of these results on the toroidal compactifications.

[13] See the paragraph after Definition 4.3.

Proposition 4.42. *The universal zip $\underline{I}$ extends to a zip of type $\mathcal{Z}$ over the toroidal compactification* $\mathrm{Sh}^{\mathrm{tor}}$ *of* Sh. *The corresponding classifying morphism ζ^{tor} extends the morphism ζ:*

$$\begin{tikzcd}
\mathrm{Sh}^{\mathrm{tor}} \arrow[r, "\zeta^{\mathrm{tor}}"] & G\text{-}\mathrm{Zip}^{\mathcal{Z}} \\
\mathrm{Sh} \arrow[u] \arrow[ur, "\zeta"'] &
\end{tikzcd}$$

Proof. See [Goldring and Koskivirta 2019a, Theorem 6.2.1]. □

Corollary 4.43. *The universal zip flag $(\underline{I}, \mathcal{J})$ extends to a zip flag of type $(\mathcal{Z}, I_0)$ over the toroidal compactification* $\mathrm{Sh}^{\mathrm{tor}}$ *of* Sh. *The corresponding classifying morphism $\zeta_{I_0}^{\mathrm{tor}}$ extends the morphism ζ_{I_0}:*

$$\begin{tikzcd}
Y_{I_0}^{\mathrm{tor}} \arrow[r, "\zeta_{I_0}^{\mathrm{tor}}"] & G\text{-}\mathrm{ZipFlag}^{\mathcal{Z}, I_0} \\
Y_{I_0} \arrow[u] \arrow[ur, "\zeta_{I_0}"'] &
\end{tikzcd}$$

Proposition 4.44. *The morphisms ζ^{tor} and $\zeta_{I_0}^{\mathrm{tor}}$ are smooth.*

Proof. See [Andreatta 2023, Theorem 1.2.]. □

5. Positive automorphic line bundles and Kodaira vanishing

Let $\mathrm{Sh}^{\mathrm{tor}}$ be a smooth and projective toroidal compactification of the special fiber of the Siegel modular variety as in Definition 3.7. Let $I_0 \subset I$ be a subset and $\pi : Y_{I_0}^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$ be the associated flag bundle that parametrizes P_0-reduction of the Hodge filtration over $\mathrm{Sh}^{\mathrm{tor}}$. In the last section, we have defined smooth morphisms

$$\zeta^{\mathrm{tor}} : \mathrm{Sh}^{\mathrm{tor}} \to G\text{-}\mathrm{Zip}^{\mathcal{Z}_\mu}$$

and

$$\zeta_{I_0}^{\mathrm{tor}} : Y_{I_0}^{\mathrm{tor}} \to G\text{-}\mathrm{ZipFlag}^{\mathcal{Z}_\mu, P_0},$$

which allowed us to construct generalized Hasse invariants on the stratification of $\mathrm{Sh}^{\mathrm{tor}}$ and $Y_{I_0}^{\mathrm{tor}}$. We denote by D_{red} the normal crossing Cartier divisors supported on the boundary of $\mathrm{Sh}^{\mathrm{tor}}$. Recall that there exists an effective Cartier divisor D whose associated reduced divisor is D_{red} and an integer $\eta_0 > 0$ such that $\omega^{\otimes \eta}(-D)$ is ample on $\mathrm{Sh}^{\mathrm{tor}}$ for every $\eta \geq \eta_0$.[14] To lighten our notation, we write D, D_{red} instead of $\pi^{-1}D, \pi^{-1}D_{\mathrm{red}}$ when no confusion is possible. Following the result of [Brunebarbe et al.], we use the generalized Hasse invariants to prove that certain line bundles $\mathcal{L}_\lambda$ are D-ample on the flag bundle $Y_{I_0}^{\mathrm{tor}}$.[15] The motivation for this notion comes from the determinant ω of the Hodge bundle Ω^{tor} which is not ample on $\mathrm{Sh}^{\mathrm{tor}}$ but only D-ample. Finally, we state a Kodaira–Nakano-like vanishing theorem for D-ample line bundles in positive characteristic from [Esnault and Viehweg 1992].

[14] See the paragraph after Definition 3.8.

[15] See Definition 5.2.

5.1. ***Positive line bundles.*** We recall the main positivity notion we will need for our automorphic line bundles. In this subsection X is a projective variety over k, a field of any characteristic.

Definition 5.1 [Lazarsfeld 2004, Chapter 1]. Let $\mathcal{L}$ a line bundle over X:

(1) $\mathcal{L}$ is ample if for any coherent module $\mathcal{F}$ over X, there is an integer $n_0 \geq 1$ such that for all $n \geq n_0$, the sheaf $\mathcal{F} \otimes \mathcal{L}^{\otimes n}$ is globally generated.

(2) Equivalently, $\mathcal{L}$ is ample if for any subvariety $V \subset X$, we have
$$c_1(\mathcal{L})^{\dim V} \cdot [V] > 0,$$
where $c_1(\mathcal{L})$ denotes the first Chern class of $\mathcal{L}$ and $\cdot$ the intersection product in the Chow ring of X.

(3) Equivalently, $\mathcal{L}$ is ample if for any subvariety $V \subset X$, there is an integer $d \geq 1$, a nonzero section s of $\mathcal{L}_{|V}^{\otimes d}$ and a point $x \in V$ such that $s(x) = 0$.

(4) $\mathcal{L}$ is nef if for any subvariety $V \subset X$, we have $c_1(\mathcal{L})^{\dim V} \cdot [V] \geq 0$.

(5) Equivalently, $\mathcal{L}$ is nef if for any subvariety $V \subset X$, there is an integer $d \geq 1$ and a nonzero section s of $\mathcal{L}_{|V}^{\otimes d}$.

(6) $\mathcal{L}$ is big if there is an integer $n \geq 1$ and an ample line bundle $\mathcal{A}$ such that $\mathcal{L}^{\otimes n} \otimes \mathcal{A}^{\otimes -1}$ is globally generated.

We now define the nonstandard notion of D-ample line bundle on a pair (X, D). It is a notion that appears in [Esnault and Viehweg 1992] without being explicitly named.

Definition 5.2. Let D be an effective Cartier divisor of X and $\mathcal{L}$ a line bundle over X. We say that $\mathcal{L}$ is D-ample if
$$\exists \eta_0 > 0, \forall \eta \geq \eta_0, \quad \mathcal{L}^{\otimes \eta}(-D) \text{ is ample.}$$

We recall some known facts about D-ample line bundles. Since we have not find a reference, we reprove them.

Proposition 5.3. *Le D be an effective Cartier divisor of X and $\mathcal{L}$ a line bundle over X. We have the following implication.*
$$\mathcal{L} \textit{ is ample} \quad \Longrightarrow \quad \mathcal{L} \textit{ is } D\textit{-ample.}$$

Proof. If $\mathcal{L}$ is ample, then $\mathcal{L}^{\otimes \eta}(-D) = \mathcal{L}^{\otimes \eta} \otimes \mathcal{O}_X(-D)$ must be ample for all $\eta \geq 1$ large enough. □

Proposition 5.4. *Let $\mathcal{L}$ be a line bundle over X. The following assertions are equivalent*:

(1) *$\mathcal{L}$ is nef and big.*

(2) *There exists an effective Cartier divisor D on X such that $\mathcal{L}$ is D-ample.*

Proof. Assume that there exists an effective Cartier divisor D on X such that $\mathcal{L}$ is D-ample. If $\mathcal{L}$ is not nef, then there is a curve $C \subset X$ such that the intersection product
$$c_1(\mathcal{L}) \cdot [C]$$

is negative. It implies that the intersection product

$$c_1(\mathcal{L}^{\otimes \eta}(-D)) \cdot [C] = \eta\, \underbrace{(c_1(\mathcal{L}) \cdot [C])}_{<0} - D \cdot [C]$$

must be negative when η is large enough, which contradicts the D-ampleness of $\mathcal{L}$. Moreover, since we can write $\mathcal{L}^{\otimes \eta_0}$ as a tensor product

$$\mathcal{L}^{\otimes \eta_0} = \mathcal{L}^{\otimes \eta_0}(-D) \otimes \mathcal{O}_X(D)$$

of an ample line bundle with an effective line bundle, $\mathcal{L}$ is big. We are left to show the implication (1) ⇒ (2). Since $\mathcal{L}$ is big, there exists an integer $n_0 \geq 1$ and an ample line bundle $\mathcal{A}$ such that $\mathcal{L}^{\otimes n_0} \otimes \mathcal{A}^{\otimes -1} = \mathcal{O}_X(D)$ with D an effective divisor. In particular, the line bundle $\mathcal{L}^{\otimes n_0}(-D)$ is ample. Since $\mathcal{L}$ is nef and the tensor product of an ample line bundle with a nef line bundle is ample, the line bundle $\mathcal{L}^{\otimes n}(-D)$ is ample for all integer $n \geq n_0$. □

Proposition 5.5. *Let D be an effective Cartier divisor and $\mathcal{L}$ a line bundle over X. If $\mathcal{L}$ and $\mathcal{L}'$ are D-ample line bundles on X, then $\mathcal{L} \otimes \mathcal{L}'$ is D-ample. If $\mathcal{L}^{\otimes n}$ is D-ample for a positive integer n, then $\mathcal{L}$ is D-ample.*

Proof. Assume that $\mathcal{L}$ and $\mathcal{L}'$ are D-ample line bundles. In particular, $\mathcal{L}'$ is nef by Proposition 5.4. For $n \geq 1$ large enough, the bundle

$$(\mathcal{L} \otimes \mathcal{L}')^{\otimes n}(-D) = \mathcal{L}^{\otimes n}(-D) \otimes (\mathcal{L}')^{\otimes n}$$

is ample as the tensor product of an ample line bundle with a nef line bundle. If $\mathcal{L}^{\otimes n}$ is D-ample for some $n \geq 1$, it implies that $\mathcal{L}^{\otimes n}$, hence $\mathcal{L}$, is nef. It also means that there is an integer $\eta_0 \geq 1$ such that $\mathcal{L}^{\otimes n\eta_0}(-D)$ is ample. Thus, the bundle

$$\mathcal{L}^{\otimes \eta - n\eta_0} \otimes \mathcal{L}^{\otimes n\eta_0}(-D) = \mathcal{L}^{\otimes \eta}(-D)$$

is ample for all $\eta \geq n\eta_0$. □

Proposition 5.6. *Let D be an effective Cartier divisor, $n \geq 1$ an integer and $\mathcal{L}$ a line bundle over X. The following assertions are equivalent*:

1. *$\mathcal{L}$ is D-ample.*
2. *$\mathcal{L}$ is nD-ample.*

Proof. Assume that $\mathcal{L}$ is D-ample. In particular $\mathcal{L}$ is nef and consider $\eta_0 \geq 1$ such that $\mathcal{L}^{\otimes \eta}(-D)$ is ample for all $\eta \geq \eta_0$. The bundle

$$\mathcal{L}^{\otimes \eta}(-nD) = \mathcal{L}^{\otimes \eta - n\eta_0} \otimes (\mathcal{L}^{\otimes \eta_0}(-D))^{\otimes n}$$

is ample for all $\eta \geq n\eta_0$ as a tensor product of a nef line bundle with an ample line bundle. Assume that $\mathcal{L}$ is nD-ample. It implies that the bundle

$$(\mathcal{L}^{\otimes \eta}(-D))^{\otimes n} = \mathcal{L}^{\otimes n\eta}(-nD)$$

is ample for all η large enough. □

5.2. ***D-ample automorphic line bundles.*** It is now convenient to introduce a relevant subset of characters of P_0.

Definition 5.7. We set

$$\mathcal{C}_{\mathrm{ample},I_0} = \{\lambda \in X^*(P_0)\mathcal{L}_\lambda \text{ is } D\text{-ample on } Y^{\mathrm{tor}}_{I_0}\}.$$

Proposition 5.8. *The subset $\mathcal{C}_{\mathrm{ample},I_0}$ is a saturated cone of $X^*(P_0)$ by Proposition 5.5 and we call it the D-ample cone of $Y^{\mathrm{tor}}_{I_0}$.*

Definition 5.9. Let $\lambda \in X^*(P_0)$ and $w \in {}^{I_0}W$. We call a generalized Hasse invariant for $\mathcal{L}_\lambda$ any section s of $\mathcal{L}_\lambda^{\otimes d}$ (for some $d \geq 1$) over $\overline{Y}^{\mathrm{tor}}_{I_0,w}$ that vanishes exactly on the boundary $\overline{Y}^{\mathrm{tor}}_{I_0,w} - Y^{\mathrm{tor}}_{I_0,w}$. Any $\mathcal{L}_\lambda$ with $\lambda \in \mathcal{C}_{\mathrm{Ha},I_0,w}$ admits a generalized Hasse invariant obtained as a pullback by $\zeta^{\mathrm{tor}}_{I_0}$.

We can now state and give a proof of the main result of this section.

Theorem 5.10. *If $\lambda \in X^*(P_0)$ is a character in $\mathcal{C}_{\mathrm{Ha},I_0}$, then $\mathcal{L}_\lambda$ is D-ample on $Y^{\mathrm{tor}}_{I_0}$.*

Proof. We start by proving that $\mathcal{L}_\lambda$ is nef on $Y^{\mathrm{tor}}_{I_0}$ for any $\lambda \in \mathcal{C}_{\mathrm{Ha},I_0}$. Let V be a subvariety of $Y^{\mathrm{tor}}_{I_0}$ and consider the minimal element w of ${}^{I_0}W$ such that $V \subset \overline{Y}^{\mathrm{tor}}_{I_0,w}$. Such an element always exists because $\overline{Y}^{\mathrm{tor}}_{I_0,w_0} = Y^{\mathrm{tor}}_{I_0}$. We consider a generalized Hasse invariant $s \in H^0(\overline{Y}^{\mathrm{tor}}_{I_0,w}, \mathcal{L}_\lambda^{\otimes d})$ (for some $d \geq 1$) and we claim that the restriction $s_{|V}$ is not identically zero. If it were, we would have

$$V \subset \overline{Y}^{\mathrm{tor}}_{I_0,w} - Y^{\mathrm{tor}}_{I_0,w} = \bigsqcup_{w' \preccurlyeq w,\, w' \neq w} Y^{\mathrm{tor}}_{I_0,w'},$$

which would contradict the minimality of w (V is irreducible). In particular, we have shown that $\mathcal{L}_\lambda$ is nef. Let λ be a character in $\mathcal{C}_{\mathrm{Ha},I_0}$. Recall that $\eta_0 \geq 1$ is an integer such that $\omega^{\otimes \eta}(-D)$ is ample for all $\eta \geq \eta_0$. Since $\mathcal{L}_\lambda$ is π-ample, we deduce that

$$\mathcal{L}_\lambda \otimes (\pi^*\omega^{\otimes \eta}(-D))^{\otimes m}$$

is ample on $Y^{\mathrm{tor}}_{I_0}$ for m large enough. Since λ belongs to $\mathcal{C}_{\mathrm{Ha},I_0}$ and the inequalities that define $\mathcal{C}_{\mathrm{Ha},I_0}$ are strict, we know that for all n large enough,

$$\mathcal{L}_\lambda^{\otimes n} \otimes \pi^*\omega^{\otimes -\eta}$$

has generalized Hasse invariants for all strata $Y^{\mathrm{tor}}_{I_0,w}$, so it is nef. Hence, we know

$$\mathcal{L}_\lambda \otimes (\pi^*\omega^{\otimes \eta}(-D))^{\otimes m} \otimes (\mathcal{L}_\lambda^{\otimes n} \otimes \pi^*\omega^{\otimes -\eta})^{\otimes m} = \mathcal{L}_\lambda^{\otimes nm+1}(-mD)$$

is ample on $Y^{\mathrm{tor}}_{I_0}$ for n, m large enough. We consider some integer $n_0, m_0 \geq 1$ such that $\mathcal{L}_\lambda^{\otimes n_0 m_0+1}(-m_0 D)$ is ample. Since $\mathcal{L}_\lambda$ is nef, we must have $\mathcal{L}_\lambda^{\otimes \eta}(-m_0 D)$ ample for all $\eta \geq n_0 m_0 + 1$. In particular, $\mathcal{L}_\lambda$ is $m_0 D$-ample, hence D-ample by Proposition 5.6. □

Remark 5.11. The theorem can be rephrased as an inclusion

$$\mathcal{C}_{\mathrm{Ha},I_0} \subset \mathcal{C}_{\mathrm{ample},I_0}.$$

Using Proposition 4.33 for the existence of generalized Hasse invariants on the stack G-ZipFlag$^{\mathcal{Z}_\mu, P_0}$, we get:

Theorem 5.12. *Let λ be a character of P_0. If λ is orbitally p-close and $\mathcal{Z}_0$-ample, then $\mathcal{L}_\lambda$ is D-ample on $Y_{I_0}^{\mathrm{tor}}$.*

5.3. ***A logarithmic Kodaira–Nakano vanishing theorem in positive characteristic.*** In this subsection, we review the Kodaira–Nakano vanishing theorem in positive characteristic due to Deligne and Illusie [1987] and a logarithmic version due to Esnault and Viehweg [1992]. Let X be a smooth projective variety of dimension n over a perfect field k of characteristic $p > 0$. Let D_{red} be a normal crossing divisor of X. We have an open immersion $\tau : U := X - D_{\mathrm{red}} \to X$.

Proposition 5.13. *Recall that X is a smooth projective variety over k and let $\mathcal{L}$ be an ample line bundle over X. Denote by d the dimension of X. Assume that $(X, \mathcal{L})$ lifts to $W_2(k)$ and $p \geq d$, then*

$$\forall i + j > d, \quad H^i(X, \Omega_X^j \otimes \mathcal{L}) = 0.$$

Proof. The detailed proof can be found in [Deligne and Illusie 1987]. □

Over the flag bundle $Y_{I_0}^{\mathrm{tor}}$ of the toroidal compactification of the Siegel modular variety, we have seen that certain line bundles $\mathcal{L}_\lambda$ are D-ample for some effective divisor D supported on the boundary. This motivates this refined version of the result of Deligne and Illusie due to Esnault and Viehweg.[16]

Proposition 5.14. *Recall that X is a smooth projective variety over k. Recall that D_{red} denotes a normal crossing divisor on X. Let D be an effective Cartier divisor whose associated reduced divisor is D_{red} and let $\mathcal{L}$ be a D-ample line bundle on X. Denote by d the dimension of X. Assume that the triple $(X, D_{\mathrm{red}}, \mathcal{L})$ lifts to $W_2(k)$ and $p \geq d$, then*

$$\forall i + j > d, \quad H^i(X, \Omega_X^j(\log D_{\mathrm{red}}) \otimes \mathcal{L}(-D_{\mathrm{red}})) = 0.$$

Proof. The proof of [Esnault and Viehweg 1992, Proposition 11.5] shows that

$$\forall i + j < \min(d, p), \quad H^i(X, \Omega_X^j(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{-1}) = 0,$$

which is equivalent to

$$\forall i + j > \max(2d - p, d), \quad H^i(X, \Omega_X^j(\log D_{\mathrm{red}}) \otimes \mathcal{L}(-D_{\mathrm{red}})) = 0$$

by Serre duality. We use that for all $i + j = n$, the pairing $\Omega_X^i(\log D_{\mathrm{red}}) \otimes \Omega_X^j(\log D_{\mathrm{red}}) \to \Omega_X^n(\log D_{\mathrm{red}})$ mapping $\alpha \otimes \beta$ to $\alpha \wedge \beta$ is perfect. □

Remark 5.15. Motivated by Proposition 5.4, one might be tempted to replace the assumption D-ample by nef and big. However, the Proposition 5.14 requires a normal crossing divisor.

[16] This result already appears in [Lan and Suh 2013].

6. Vanishing for automorphic vector bundles

In this section, we prove our vanishing results announced in Section 1.2. We start with some preliminary results concerning the spectral sequence associated to the cohomology of a filtered sheaf. Next, we construct the function $g_{I_0,e}$ on the power set of characters and prove that it produces new vanishing results from old ones. Finally, we give more details in the special case $g = 2$ as it is easier than the general case.

6.1. ***Spectral sequence associated to a filtered sheaf.*** We consider a scheme morphism $f : X \to S$ and a sheaf $\mathcal{F}$ on X endowed with an increasing filtration $\mathcal{F}_\bullet$ with graded pieces

$$\forall k \in \mathbb{Z}, \quad \mathrm{gr}_k = \mathcal{F}_k/\mathcal{F}_{k-1}.$$

Proposition 6.1. *There is a spectral sequence starting at page* 2

$$E_2^{t,k} = R^{t+k} f_*(\mathrm{gr}_k) \quad \Longrightarrow \quad R^{t+k} f_*(\mathcal{F}).$$

Proof. This result is well-known: see the appendix of [Esnault and Viehweg 1992] for example. We just recall how the differentials of the second page are defined. For all $k \in \mathbb{Z}$, there is an exact sequence

$$0 \to \mathcal{F}_{k-1}/\mathcal{F}_{k-2} \to \mathcal{F}_k/\mathcal{F}_{k-2} \to \mathcal{F}_k/\mathcal{F}_{k-1} \to 0$$

and the differentials are the connecting morphisms

$$\forall i \geq 0, \quad R^i f_*(\mathrm{gr}_k) \to R^{i+1} f_*(\mathrm{gr}_{k-1}). \qquad \square$$

From the study of this spectral sequence, we deduce several results.

Lemma 6.2. *Let* $i_0 \geq 0$ *and assume*

$$\forall k \in \mathbb{Z}, \quad R^{i_0} f_*(\mathrm{gr}_k) = 0.$$

Then,

$$R^{i_0} f_*(\mathcal{F}) = 0.$$

Proof. We pass from a page of a spectral sequence to the next one by taking cohomology and since $E_2^{i_0-k,k} = R^{i_0} f_*(\mathrm{gr}_k) = 0$ for all $k \in \mathbb{Z}$, we have

$$\forall a \geq 2, \forall k \in \mathbb{Z}, \quad E_a^{i_0-k,k} = 0.$$

Thus,

$$\forall a \geq 2, \forall k \in \mathbb{Z}, \quad E_\infty^{i_0-k,k} = 0$$

and

$$R^{i_0} f_*(\mathcal{F}) = 0. \qquad \square$$

Lemma 6.3. *Let* $i_0 \geq 0$ *and assume that there exists* $n \in \mathbb{Z}$ *such that for all* $k > n$, $\mathrm{gr}_k = 0$. *If*

$$R^{i_0} f_*(\mathcal{F}) = 0, \forall k \leq n-1, \quad R^{i_0+1} f_*(\mathrm{gr}_k) = 0,$$

then

$$R^{i_0} f_*(\mathrm{gr}_n) = 0.$$

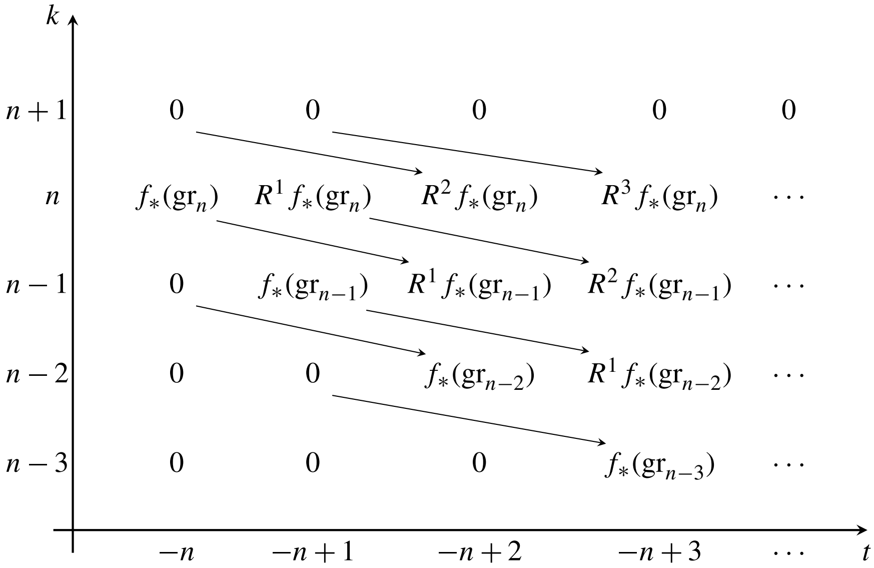
k
$n+1$
n
$n-1$
$n-2$
$n-3$
0 0 0 0 0
$f_*(\mathrm{gr}_n)$ $R^1f_*(\mathrm{gr}_n)$ $R^2f_*(\mathrm{gr}_n)$ $R^3f_*(\mathrm{gr}_n)$ $\cdots$
0 $f_*(\mathrm{gr}_{n-1})$ $R^1f_*(\mathrm{gr}_{n-1})$ $R^2f_*(\mathrm{gr}_{n-1})$ $\cdots$
0 0 $f_*(\mathrm{gr}_{n-2})$ $R^1f_*(\mathrm{gr}_{n-2})$ $\cdots$
0 0 0 $f_*(\mathrm{gr}_{n-3})$ $\cdots$
$-n$ $-n+1$ $-n+2$ $-n+3$ $\cdots$ t

Figure 2. E_2-page of the spectral sequence.

Proof. For a visual support, see the Figure 2. From the hypothesis on the graded pieces, we know that for all $a \geq 2$, the differential with target $E_a^{-n+i_0,n}$ vanishes. Since for all $k \leq n-1$ we have $R^{i_0+1}f_*(\mathrm{gr}_k) = 0$, then for all $a \geq 2$ the differential with source $E_a^{-n+i_0,n}$ must vanish. Thus, we get $R^{i_0}f_*(\mathrm{gr}_n) = E_2^{-n+i_0,n} = E_\infty^{-n+i_0,n}$ and $E_\infty^{-n+i_0,n} = 0$ as a graded piece of $R^{i_0}f_*(\mathcal{F})$. □

6.2. ***The general case.*** The goal of this subsection is to explain how to deduce new vanishing results for the coherent cohomology from known ones. Other Shimura varieties could be considered but we have restricted ourselves to the Siegel case for simplicity. We recall the notation. Let $\mathrm{Sh}^{\mathrm{tor}}$ be the special fiber over $\mathbb{F}_p$ of the Siegel modular variety of genus $g \geq 2$ and $\pi : Y_{I_0}^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$ the flag bundle in P/P_0 where $P_0 \subset \mathrm{Sp}_{2g}$ is a parabolic subgroup of type $I_0 \subset I \subset \Delta$ which is contained in the parabolic $P \subset \mathrm{Sp}_{2g}$ of type I. We denote by D_{red} the normal crossing divisor supported on the boundary of $\mathrm{Sh}^{\mathrm{tor}}$. We use the same notation D_{red} for the normal crossing divisor $\pi^{-1}D_{\mathrm{red}}$ of $Y_{I_0}^{\mathrm{tor}}$ when no confusion is possible. We denote by d, d_0 the dimension of $\mathrm{Sh}^{\mathrm{tor}}$, $Y_{I_0}^{\mathrm{tor}}$ and $r_0 = d_0 - d$ the relative dimension of π. We choose a system of positive roots in a way to obtain

$$I = \{e_i - e_{i+1} \mid i = 1, \dots, g-1\} \subset \Delta = \{e_i - e_{i+1} \mid i = 1, \dots, g-1\} \cup \{2e_g\}.$$

The Levi subgroup L of $P \subset \mathrm{Sp}_{2g}$ is GL_g and to each representation V of L, we have an associated vector bundle $\mathcal{W}(V)$ on $\mathrm{Sh}^{\mathrm{tor}}$. With our conventions, the Hodge bundle Ω is the vector bundle of rank g associated to the standard representation std_L of L. To each character λ of P_0, we have an associated line bundle $\mathcal{L}_\lambda$ on $Y_{I_0}^{\mathrm{tor}}$. Assuming that $p \geq d_0$, the basic idea is to use the logarithmic Kodaira–Nakano vanishing theorem (see Proposition 5.14) on the flag bundle $Y_{I_0}^{\mathrm{tor}}$ with D-ample line bundle $\mathcal{L}_\lambda$. Since the determinant of

$$\Omega^1_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}})$$

is a line bundle over $Y^{\mathrm{tor}}_{I_0}$, it is not hard to express it as an automorphic bundle and it provides vanishing results for cohomology groups H^i with $i > 0$. The accessible weights with this method are regular. To access less regular weights, a natural idea is to use the logarithmic Kodaira–Nakano vanishing theorem for

$$\Omega^m_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}})$$

with $m < d_0$. However, this bundle is not a line bundle and doesn't seem related to automorphic bundles (see Remark 6.9). A solution is to filter it by automorphic vector bundles and then use the associated spectral sequence. The following result is well-known but since we haven't found a reference, we give a proof.

Lemma 6.4. *We have an exact sequence of vector bundles*

$$0 \to \pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \to \Omega^1_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}}) \to \Omega^1_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \to 0.$$

Proof. By [Deligne 1970, II, Section 3], we have a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}} \arrow[r] \arrow[d, "a"] & \pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \arrow[r] \arrow[d, "b"] & \pi^*\mathcal{O}_{D_{\mathrm{red}}} \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \Omega^1_{Y^{\mathrm{tor}}_{I_0}} \arrow[r] & \Omega^1_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}}) \arrow[r] & \mathcal{O}_{\pi^{-1}D_{\mathrm{red}}} \arrow[r] & 0
\end{tikzcd}$$

where the rows are exact (use also that π is flat for the first row). Since π is smooth, $\ker a = 0$ and by the snake lemma, the sequence

$$0 \to \ker b \to 0 \to \Omega^1_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \to \Omega^1_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}})/\pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \to 0$$

is exact. The desired exact sequence is obtained from b. □

Definition 6.5. Let $e \geq 0$ be an integer. We define an increasing filtration $F_\bullet$ of $\Omega^{d_0-e}_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}})$ by

$$F_k = \pi^*\Omega^{d_0-e-k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \wedge \Omega^k_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}}),$$

with graded pieces

$$\mathrm{gr}_k = \pi^*\Omega^{d_0-e-k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^k_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}}.$$

From the Proposition 6.1, we get an associated spectral sequence starting at page 2 for each $\lambda \in X^*(P_0)$

$$E^{t,k}_{2,e,\lambda} = H^{t+k}(Y^{\mathrm{tor}}_{I_0}, \mathrm{gr}_k \otimes \mathcal{L}_\lambda(-D_{\mathrm{red}})) \quad \Longrightarrow \quad H^{t+k}(Y^{\mathrm{tor}}_{I_0}, \Omega^{d_0-e}_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}_\lambda(-D_{\mathrm{red}})). \tag{2}$$

This spectral sequence doesn't degenerate in general, so we need to consider weights λ that ensure partial degeneration results. This will allow us to deduce vanishing results for tensor product of the form

$$\Omega^k_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla(\lambda).$$

Another difficulty arises because, in positive characteristic, algebraic representations of reductive groups are not semisimple, so we can't easily deduce vanishing results for automorphic bundles from vanishing results for such tensor products. However, from Proposition 6.7, $\Omega^k_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})$ admits a ∇-filtration if $p > d$ and we can use Corollary 2.11 to see that the tensor product $\Omega^k_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla(\lambda)$ admits also a ∇-filtration: this will allow us to deduce new vanishing results for automorphic bundles. Since our method relies heavily on partial degeneration results that requires vanishing results, we can think of it as a way to deduce new vanishing results from known ones. This is why we present them in two steps:

- *Degeneration*: We determine the vanishing results we need to ensure the degeneration of relevant spectral sequences.
- *Propagation*: Given a set of known vanishing results, we determine the new vanishing results we can deduce from them.

To lighten our notation, we will denote the subcanonical automorphic bundle by $\nabla^{\mathrm{sub}}(\lambda)$ instead of $\nabla(\lambda)(-D_{\mathrm{red}})$ and $\mathcal{L}^{\mathrm{sub}}(V)$ instead of $\mathcal{L}(V)(-D_{\mathrm{red}})$. We introduce some notation for the weights of our automorphic bundles.

Definition 6.6. For all $n \geq 0$, we set

$$(\mu^n_j)_{1\leq j\leq \binom{d}{n}} = (w_0 w_{0,L} \nu^n_j)_{1\leq j\leq \binom{d}{n}},$$

where the ν^n_j are the characters of the L-representation

$$\wedge^n \operatorname{Sym}^2 \operatorname{std}_L .$$

We assume that $\nu^n_{\binom{d}{n}}$ is the highest weight.

Proposition 6.7. *If $p > d = g(g+1)/2$, then for any $n \geq 1$ the vector bundle $\Omega^n_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})$ admits a filtration*

$$0 = \mathcal{V}^s \subsetneq \mathcal{V}^{s-1} \subsetneq \cdots \subsetneq \mathcal{V}^0 = \Omega^n_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}),$$

where the graded pieces are automorphic vector bundles of the form $\nabla(\mu^n_j)$ with μ^n_j dominant.

Proof. Recall from Proposition 3.12 that the Kodaira–Spencer map induces an isomorphism

$$\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) = \mathcal{W}(\operatorname{Sym}^2 \operatorname{std}_L) = \nabla(0, \dots, 0, -2).$$

We only need to see that for any $1 \leq n \leq g(g+1)/2$, the GL_g-module $\Lambda^n \operatorname{Sym}^2 \operatorname{std}_L$ admits a ∇-filtration. The module

$$\operatorname{Sym}^2 \operatorname{std}_L = \nabla(2, 0, \dots, 0)$$

is already a costandard module. From Proposition 2.9, the module

$$(\operatorname{Sym}^2 \operatorname{std}_L)^{\otimes n}$$

admits also a ∇-filtration. Since $p > g(g+1)/2 \geq n$, p does not divide $n!$ and the surjection of G-modules

$$(\operatorname{Sym}^2 \mathrm{std}_L)^{\otimes n} \to \Lambda^n \operatorname{Sym}^2 \mathrm{std}_L$$

admits a GL_g-equivariant section s defined by the formula

$$s(v_1 \wedge \cdots \wedge v_n) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)}.$$

As a direct factor of $(\operatorname{Sym}^2 \mathrm{std}_L)^{\otimes n}$, the Corollary 2.13 implies that $\Lambda^n \operatorname{Sym}^2 \mathrm{std}_L$ admits a ∇-filtration. □

Proposition 6.8. *We have an isomorphism*

$$\Omega^1_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} = \mathcal{L}(\operatorname{Lie} L/\operatorname{Lie}(P_0 \cap L))^\vee,$$

and for all $i \geq 0$, the vector bundle $\Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}}$ is filtered by line bundles

$$\mathcal{L}_{-s_M}, \quad \textit{where } s_M = \sum_{\alpha \in M} \alpha \textit{ for all } M \subset \phi^+_L - \phi^+_{I_0} \textit{ such that } |M| = i.$$

In particular, $\Omega^{r_0}_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \simeq \mathcal{L}_{-2\rho_{I_0}}$ with

$$\rho_{I_0} = \frac{1}{2} \sum_{\alpha \in \phi^+_L \setminus \phi^+_{I_0}} \alpha.$$

Proof. Consider the cartesian diagram

$$\begin{tikzcd} Y^{\mathrm{tor}}_{I_0} \arrow[r, "\tilde{\zeta}"] \arrow[d, "\pi"] & \lfloor P_0 \backslash * \rfloor \arrow[d, "\tilde{\pi}"] \\ \mathrm{Sh}^{\mathrm{tor}} \arrow[r, "\zeta"] & \lfloor P \backslash * \rfloor \end{tikzcd}$$

where the horizontal arrows correspond to the universal P-torsor on $\mathrm{Sh}^{\mathrm{tor}}$ and the universal P_0-torsor on $Y^{\mathrm{tor}}_{I_0}$ and where the vertical arrow $\tilde{\pi}$ between the classifying stacks is induced by the inclusion $P_0 \subset P$. Coherent sheaves on the classifying stack $\lfloor P_0 \backslash * \rfloor$ are algebraic representations of P_0 and clearly, we have

$$\Omega^1_{\tilde{\pi}} = \operatorname{Lie}(P)/\operatorname{Lie}(P_0)^\vee,$$

where the action of P_0 on $\operatorname{Lie}(P)$ is induced by the restriction of the adjoint action of P. From the isomorphism

$$\tilde{\zeta}^* \Omega^1_{\tilde{\pi}} = \Omega^1_{\pi},$$

we deduce that

$$\Omega^1_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} = \mathcal{L}(\operatorname{Lie}(P)/\operatorname{Lie}(P_0)^\vee).$$

Since the T-weights on $\operatorname{Lie}(P)/\operatorname{Lie}(P_0)^\vee$ are the $-\alpha$ with $\alpha \in \phi^+_L - \phi^+_{I_0}$, the result follows. □

Remark 6.9. The exact sequence

$$0 \to \pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \to \Omega^1_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}}) \to \Omega^1_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \to 0$$

doesn't seem to split and we cannot prove the vanishing of the abelian group

$$\mathrm{Ext}^1_{\mathcal{O}_{Y^{\mathrm{tor}}_{I_0}}}(\Omega^1_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}}, \pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}))$$

using known vanishing results because the vector bundle

$$\pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^{1\vee}_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}}$$

is filtered by the $\mathcal{L}_\lambda$ with $\lambda \in X^*(P_0)$ outside the antidominant Weyl chamber for which the first cohomology is nonzero in general. Outside the case $I_0 = I$, we don't even know if $\Omega^1_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}})$ is automorphic. In other words, we don't know if $\Omega^1_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}})$ is of the form $\mathcal{L}(V)$ for an algebraic representation V of P_0.

6.2.1. *Degeneration.* Multiple subsets of $X^*(P_0)$ will occur in the formulations of our degeneration results, we gather them in the following definition.

Definition 6.10. Consider an integer $0 \leq e \leq d-1$. We denote $\mathcal{C}^0_{\mathrm{deg}}$, $\mathcal{C}^1_{\mathrm{deg},e}$ and $\mathcal{C}^2_{\mathrm{deg},e}$ the following sets of characters:

$$\mathcal{C}^0_{\mathrm{deg}} := \{\lambda \in X^*(P_0) \mid \lambda - 2\rho_{I_0} \in X^*(P_0)^+\},$$

$$\mathcal{C}^1_{\mathrm{deg},e} := \big\{\lambda \in X^*(P_0) \mid \forall i > e+1, \forall j, \forall 1 \leq k \leq e, \forall M \subset \phi_L^+ - \phi_{I_0}^+ \text{ such that } |M| = r_0 - k,\ H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\mu_j^{d-e+k} + \lambda - s_M)) = 0\big\},$$

$$\mathcal{C}^2_{\mathrm{deg},e} := \Big\{\lambda \in X^*(P_0) \mid \forall i > e+1, \forall j \neq \binom{d}{d-e} H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\mu_j^{d-e} + \lambda - 2\rho_{I_0})) = 0\Big\}.$$

Lemma 6.11. *Let* $\lambda \in \mathcal{C}^0_{\mathrm{deg}}$ *and* $\mathcal{F}$ *be a coherent sheaf on* $\mathrm{Sh}^{\mathrm{tor}}$. *For all* $0 \leq i \leq r_0$ *and* $n \geq 0$, *we have the following isomorphism*

$$H^n(Y^{\mathrm{tor}}_{I_0}, \pi^*\mathcal{F} \otimes \Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) = H^n(\mathrm{Sh}^{\mathrm{tor}}, \mathcal{F} \otimes \pi_*(\Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda)).$$

Proof. Let $i \geq 0$. We know by Proposition 6.8 that the vector bundle $\Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}}$ is filtered by line bundles

$$\mathcal{L}_{-s_M}, \quad \text{where } s_M = \sum_{\alpha \in M} \alpha \text{ for all } M \subset \phi_L^+ - \phi_{I_0}^+ \text{ such that } |M| = i.$$

From the definition of $\mathcal{C}^0_{\mathrm{deg}}$ and the fact that the roots in $\phi_L^+ - \phi_{I_0}^+$ are I_0-dominant, we know that all $\lambda - s_M$ are I_0-dominant characters. From Kempf's vanishing theorem (see Proposition 2.5 and Lemma 3.19), we get

$$\forall M \quad \forall k > 0, \quad R^k\pi_*(\mathcal{L}_{\lambda - s_M}) = 0$$

and by Lemma 6.2 we deduce

$$\forall k > 0, \quad R^k\pi_*(\Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}_\lambda) = 0.$$

Since $\pi^*\mathcal{O}_{\mathrm{Sh}^{\mathrm{tor}}}(-D_{\mathrm{red}}) = \mathcal{O}_{Y^{\mathrm{tor}}_{I_0}}(-D_{\mathrm{red}})$, the projection formula implies

$$\forall k > 0, \quad R^k\pi_*(\Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}_\lambda(-D_{\mathrm{red}})) = 0.$$

Using again the projection formula, it implies that the Leray spectral sequence

$$E_2^{t,k} = H^t(\mathrm{Sh}^{\mathrm{tor}}, R^k\pi_*(\pi^*\mathcal{F} \otimes \Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda)) \quad \Longrightarrow \quad H^{t+k}(Y^{\mathrm{tor}}_{I_0}, \pi^*\mathcal{F} \otimes \Omega^i_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda)$$

is concentrated on one row and we get the desired isomorphisms. □

Proposition 6.12. *Assume that $p > d = g(g+1)/2$. Let $e \geq 0$ and $i > e$ be integers. For any character*

$$\lambda \in \mathcal{C}^0_{\mathrm{deg}} \cap \mathcal{C}^1_{\mathrm{deg},e} \cap \mathcal{C}^2_{\mathrm{deg},e},$$

the vanishing

$$H^i(Y^{\mathrm{tor}}_{I_0}, \Omega^{d_0-e}_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) = 0$$

implies the vanishing

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\mu^{d-e}_{\binom{d}{d-e}} + \lambda - 2\rho_{I_0})) = 0.$$

Proof. We use Lemma 6.2 for the filtration of

$$\pi^*\Omega^{d-e+k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^{r_0-k}_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda$$

obtained from the one defined in Proposition 6.8 to see that the vanishing

$$\forall 1 \leq k \leq e, \quad H^{i+1}(Y^{\mathrm{tor}}_{I_0}, \pi^*\Omega^{d-e+k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^{r_0-k}_{Y^{\mathrm{tor}}_{I_0}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) = 0,$$

follows from the vanishing

$$H^{i+1}(Y^{\mathrm{tor}}_{I_0}, \pi^*\Omega^{d-e+k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_{\lambda-s_M}) = 0 \tag{3}$$

for all $1 \leq k \leq e$ and all $M \subset \phi^+_L - \phi^+_{I_0}$ such that $|M| = r_0 - k$. Since $\lambda \in \mathcal{C}^0_{\mathrm{deg}}$, we know by Lemma 6.11 that

$$H^{i+1}(Y^{\mathrm{tor}}_{I_0}, \pi^*\Omega^{d-e+k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_{\lambda-s_M}) = H^{i+1}(\mathrm{Sh}^{\mathrm{tor}}, \Omega^{d-e+k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla^{\mathrm{sub}}(\lambda - s_M)).$$

We use Propositions 2.9 and 6.7 to see that the bundle

$$\Omega^{d-e+k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla^{\mathrm{sub}}(\lambda - s_M)$$

admits a filtration where the graded pieces are isomorphic to

$$\nabla(\mu^{d-e+k}_j + \lambda - s_M).$$

By Lemma 6.2, we deduce that the vanishing in equality (3) follows from $\lambda \in \mathcal{C}^1_{\mathrm{deg},e}$. Since

$$H^i(Y^{\mathrm{tor}}_{I_0}, \Omega^{d_0-e}_{Y^{\mathrm{tor}}_{I_0}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) = 0$$

$$\begin{tikzcd}
k & & & & & & \\
2 & 0 \arrow[rrd] & 0 \arrow[rrd] & 0 & 0 & 0 & \\
1 & H^0(\mathrm{gr}_1 \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) \arrow[rrd] & H^1(\mathrm{gr}_1 \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) \arrow[rrd] & H^2(\mathrm{gr}_1 \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) & H^3(\mathrm{gr}_1 \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) & 0 & \\
0 & 0 \arrow[rrd] & 0 \arrow[rrd] & 0 & 0 & 0 & \\
-1 & 0 & 0 & 0 & 0 & 0 & \\
 & -1 & 0 & 1 & 2 & 3 & t
\end{tikzcd}$$

Figure 3. E_2-page of the spectral sequence when $e = 0$.

by hypothesis, we can apply Lemma 6.3 to $E_{2,d_0-e,\lambda}$ to deduce that

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \Omega^{d-e}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla^{\mathrm{sub}}(\lambda - 2\rho_{I_0})) = 0.$$

Combining again the Propositions 2.9 and 6.7, we know that

$$\Omega^{d-e}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla^{\mathrm{sub}}(\lambda - 2\rho_{I_0})$$

admits a ∇-filtration. Since $\lambda \in \mathcal{C}^2_{\mathrm{deg},e}$, we use again Lemma 6.3 for the ∇-filtration of $\Omega^{d-e}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla^{\mathrm{sub}}(\lambda - 2\rho_{I_0})$ to see that

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\mu^{d-e}_{\binom{d}{d-e}} + \lambda - 2\rho_{I_0})) = 0. \qquad \square$$

6.2.2. *Propagation.* In this section, we construct a nondecreasing function on the power set of characters that gives new vanishing results from known ones. Our main result is Theorem 6.16.

Definition 6.13. For all $k \geq 0$, we define a subset $\mathcal{C}^k_{\mathrm{van}}$ of X^* as

$$\mathcal{C}^k_{\mathrm{van}} = \{\lambda \in X^* \mid \forall i > k,\ H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\lambda)) = 0\}.$$

Remark 6.14. $\mathcal{C}^k_{\mathrm{van}}$ always contains the nondominant characters.

Definition 6.15. We define a function $g_{I_0,e} : \mathcal{P}(X^*) \to \mathcal{P}(X^*)$ by

$$g_{I_0,e}(\mathcal{C}) = \mu^{d-e}_{\binom{d}{d-e}} + X^*(P_0)^+ \cap (-2\rho_{I_0} + \mathcal{C}_{\mathrm{ample},I_0}) \cap \bigcap_{k,j,M} (s_M - 2\rho_{I_0} - \mu^{d-e+k}_j + \mathcal{C})$$

for all $\mathcal{C} \subset X^*$ where the last intersection is taken over the set of k, j, M where $0 \leq k \leq e$, $1 \leq j \leq \binom{d}{d-e} + k$ and $M \subset \phi^+_L - \phi^+_{I_0}$ such that $|M| = r_0 - k$ with the exception of $j = \binom{d}{d-e}$ when $k = 0$.

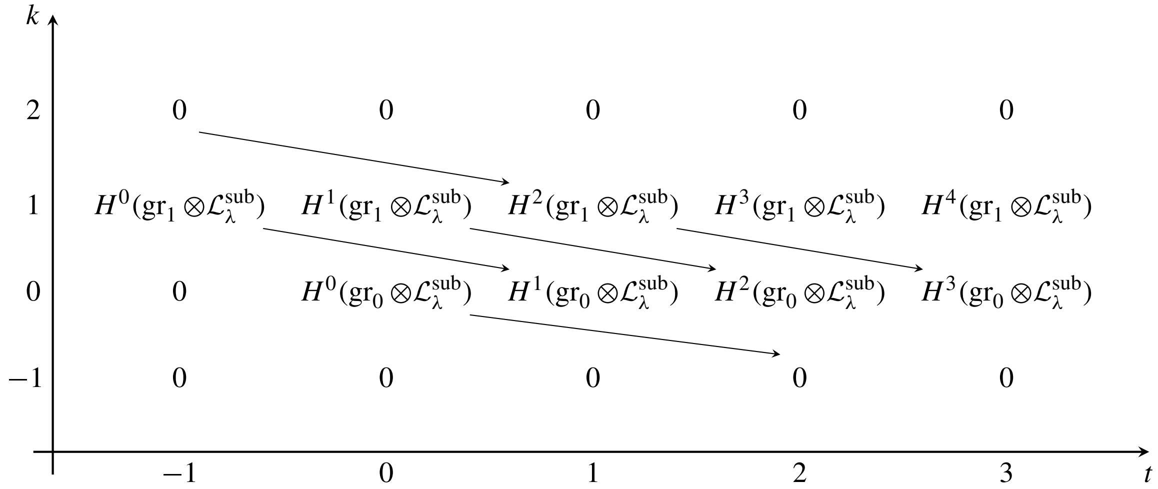
k
2
1
0
-1
0 0 0 0 0
$H^0(\mathrm{gr}_1 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$ $H^1(\mathrm{gr}_1 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$ $H^2(\mathrm{gr}_1 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$ $H^3(\mathrm{gr}_1 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$ $H^4(\mathrm{gr}_1 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$
0 $H^0(\mathrm{gr}_0 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$ $H^1(\mathrm{gr}_0 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$ $H^2(\mathrm{gr}_0 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$ $H^3(\mathrm{gr}_0 \otimes \mathcal{L}_\lambda^{\mathrm{sub}})$
0 0 0 0 0
-1 0 1 2 3
t

Figure 4. E_2-page of the spectral sequence when $e = 1$.

Theorem 6.16. *Assume that $p > d_0$. Let $\mathcal{C}$ be a subset of $\mathcal{C}_{\mathrm{van}}^{e+1}$. Then, we have*

$$g_{I_0,e}(\mathcal{C}) \subset \mathcal{C}_{\mathrm{van}}^{e}.$$

In other words, if we have a set $\mathcal{C}$ of characters λ for which the cohomology

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\lambda))$$

is concentrated in degrees $[0, e+1]$, then the image of $\mathcal{C}$ by the function $g_{I_0,e}$ is a set of characters λ for which the cohomology $H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\lambda))$ is concentrated in degrees $[0, e]$.

Proof. Since $g_{I_0,e}$ is nondecreasing, it suffices to show $g_{I_0,e}(\mathcal{C}_{\mathrm{van}}^{e+1}) \subset \mathcal{C}_{\mathrm{van}}^{e}$. Let $\lambda \in g_{I_0,e}(\mathcal{C}_{\mathrm{van}}^{e+1})$ be a character and define $\lambda' := \lambda + 2\rho_{I_0} - \mu^{d-e}_{\binom{d}{d-e}}$. From the definition of $g_{I_0,e}$, we first deduce that

$$\lambda' \in \mathcal{C}_{\mathrm{deg}}^{0} \cap \mathcal{C}_{\mathrm{deg}}^{1} \cap \mathcal{C}_{\mathrm{deg}}^{2}$$

and

$$\lambda' \in \mathcal{C}_{\mathrm{ample}, I_0}.$$

Since the triple $(Y_{I_0}^{\mathrm{tor}}, D_{\mathrm{red}}, \mathcal{L}_{\lambda'})$ lifts to $\mathbb{Z}/p^2\mathbb{Z}$ and $p \geq d_0$, we apply Proposition 5.14 to see that

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \Omega^{d_0-e}_{Y_{I_0}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}_{\lambda'}^{\mathrm{sub}}) = 0$$

for all $i + d_0 - e > d_0$ (i.e., $i > e$) and we use Proposition 6.12 (as $p > d_0 \geq d$) to see that

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\mu^{d-e}_{\binom{d}{d-e}} + \lambda' - 2\rho_{I_0})) = H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\lambda)) = 0$$

for all $i > e$. □

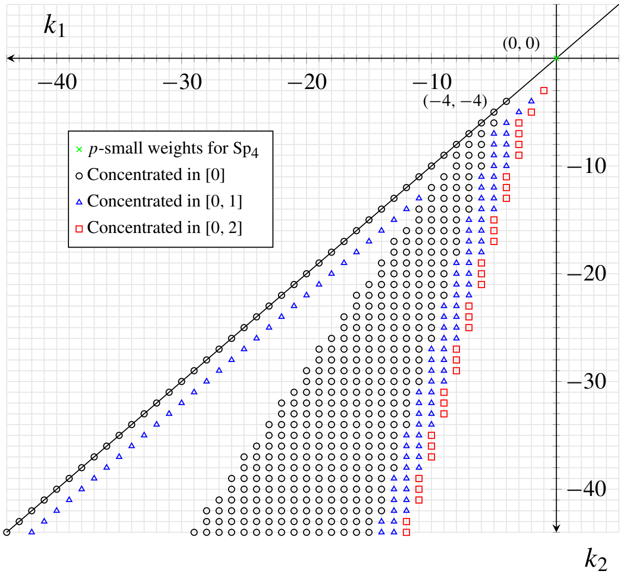

Figure 5. $g = 2$, $p = 5$. The weights $\lambda = (k_1 \geq k_2)$ such that the cohomology is concentrated in degree 0 contains in particular the positive parallel weights (k, k) below $(-4, -4)$. The vanishing results in the region located near the positive parallel line comes from the degeneration with $I = \{(1, -1)\}$ and the rest corresponds to the degeneration with $I = \varnothing$.

Remark 6.17. The theorem is still valid if we use a subset $\mathcal{C}'_{\mathrm{ample}, I_0}$ of $\mathcal{C}_{\mathrm{ample}, I_0}$ instead of $\mathcal{C}_{\mathrm{ample}, I_0}$ in the definition of $g_{I_0,e}$. In particular, by Theorem 5.12, we can use it with the subset of orbitally p-close and $\mathcal{Z}_0$-ample characters.

6.3. ***The Siegel threefold case.*** In this subsection, we give more details in the case $g = 2$ because we believe it already contains some of the idea of the general method and it requires less notation. Assume that p is a prime larger than $g^2 = 4$. The Siegel threefold $\mathrm{Sh}^{\mathrm{tor}}$ is projective variety of dimension $d = 3$ over $\mathbb{F}_p$. From the Kodaira–Spencer isomorphism, we have an identification

$$\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) = \nabla(0, -2).$$

From Proposition 6.7, we know that any exterior power of $\mathrm{Sym}^2 \mathrm{std}_{\mathrm{GL}_2}$ admits a ∇-filtration. It directly implies that we have

$$\Omega^2_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) = \nabla(-1, -3),$$
$$\Omega^3_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) = \nabla(-3, -3),$$

and the weights of these three automorphic vector bundles are

$$(\mu^1_j)_j = \{(-2, 0), (-1, -1), (0, -2)\},$$
$$(\mu^2_j)_j = \{(-3, -1), (-2, -2), (-1, -3)\},$$
$$(\mu^3_j)_j = \{(-3, -3)\}.$$

We start with the case $I_0 = \varnothing$. The associated complete flag bundle $\pi : Y^{\mathrm{tor}} \to \mathrm{Sh}^{\mathrm{tor}}$ parametrizes quotient line bundles of the rank 2 Hodge bundle Ω^{tor}. It is a $\mathbb{P}^1$-fibration and we have an identification:

$$\Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} = \mathcal{L}_{-2\rho} = \mathcal{L}_{(-1,1)}.$$

For any integer $0 \le e \le d-1 = 2$, we have an increasing filtration on the bundle $\Omega^{4-e}_{Y^{\mathrm{tor}}}(\log D_{\mathrm{red}})$ given by

$$F_k = \pi^*\Omega^{4-e-k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \wedge \Omega^k_{Y^{\mathrm{tor}}}(\log D_{\mathrm{red}})$$

with graded pieces

$$\mathrm{gr}_k = \pi^*\Omega^{4-e-k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^k_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}.$$

For any character $\lambda = (k_1 \ge k_2)$, we have an associated spectral sequence

$$E^{t,k}_{2,e,\lambda} = H^{t+k}(Y^{\mathrm{tor}}, \pi^*\Omega^{4-e-k}_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\Omega^k_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}\otimes\mathcal{L}^{\mathrm{sub}}_\lambda) \quad\Longrightarrow\quad H^{t+k}(Y^{\mathrm{tor}}, \Omega^{4-e}_{Y^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\mathcal{L}^{\mathrm{sub}}_\lambda)$$

starting at page 2. We will study this spectral sequence for each e, starting with $e = 0$. In this case (see the corresponding figure), the second page of the spectral sequence is concentrated in one row as the only graded piece is $\mathrm{gr}_1 = \pi^*\Omega^3_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}$. In particular, the spectral sequence degenerates at page 2 and we get

$$H^i(Y^{\mathrm{tor}}, \pi^*\Omega^3_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) = H^i(Y^{\mathrm{tor}}, \Omega^4_{Y^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_\lambda)$$

for all $i \ge 0$. Moreover, if we assume that $\lambda - 2\rho$ is dominant (which is equivalent to $k_1 \ge k_2 + 2$), we get

$$\begin{aligned} H^i(Y^{\mathrm{tor}}, \pi^*\Omega^3_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) &= H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla(-3,-3) \otimes \nabla^{\mathrm{sub}}(k_1-1, k_2+1)) \\ &= H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(k_1-4, k_2-2)) \end{aligned}$$

for all $i \ge 0$. We assume that $\mathcal{L}_{(k_1,k_2)}$ is D-ample on Y^{tor} and we use the logarithmic Kodaira–Nakano vanishing theorem to see that

$$H^i(Y^{\mathrm{tor}}, \Omega^4_{Y^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) = 0$$

for all $i > 0$. We summarize this discussion by saying that we have

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(k_1-4, k_2-2)) = 0$$

for all $i > 0$ and (k_1, k_2) such that

- $k_1 \ge k_2 + 2$,
- $(k_1, k_2) \in \mathcal{C}_{\mathrm{ample},\varnothing}$.

Now, we consider the spectral sequence in the case $e = 1$ (see the corresponding figure).

The graded pieces are $\mathrm{gr}_1 = \pi^*\Omega^2_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}$ and $\mathrm{gr}_0 = \pi^*\Omega^3_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})$. The limit is $H^i(Y^{\mathrm{tor}}, \Omega^3_{Y^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_\lambda)$ and by the logarithmic Kodaira–Nakano theorem, it vanishes for all $i > 1$ when $(k_1, k_2) \in \mathcal{C}_{\mathrm{ample},\varnothing}$. The critical differential is

$$d^{1,1} : H^2(Y^{\mathrm{tor}}, \pi^*\Omega^2_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) \to H^3(Y^{\mathrm{tor}}, \pi^*\Omega^3_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \mathcal{L}^{\mathrm{sub}}_\lambda),$$

because when $d^{1,1}=0$, we have $E_2^{1,1}=E_\infty^{1,1}$ and $E_2^{2,1}=E_\infty^{2,1}$. Under the additional hypothesis $k_1\geq k_2+2$, we deduce that

$$H^i(Y^{\mathrm{tor}},\pi^*\Omega^2_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}\otimes\mathcal{L}^{\mathrm{sub}}_\lambda)=H^i(\mathrm{Sh}^{\mathrm{tor}},\nabla(-1,-3)\otimes\nabla^{\mathrm{sub}}(k_1-1,k_2+1))=0$$

for all $i>1$ and $(k_1,k_2)\in\mathcal{C}_{\mathrm{ample},\varnothing}$. Consider an integer $i=2$ or 3. The tensor product of automorphic vector bundles

$$\nabla(-1,-3)\otimes\nabla^{\mathrm{sub}}(k_1-1,k_2+1)$$

is filtered by the automorphic bundles

$$\nabla^{\mathrm{sub}}(\mu_j^2+(k_1-1,k_2+1))$$

where $j=1,2,3$ and if we ask for the vanishing

$$H^{i+1}(\mathrm{Sh}^{\mathrm{tor}},\nabla^{\mathrm{sub}}(\mu_j^2+(k_1-1,k_2+1)))=0$$

for $j=1,2$, it will imply

$$H^i(\mathrm{Sh}^{\mathrm{tor}},\nabla^{\mathrm{sub}}((-1,-3)+(k_1-1,k_2+1)))=H^i(\mathrm{Sh}^{\mathrm{tor}},\nabla^{\mathrm{sub}}(k_1-2,k_2-2))=0.$$

To see that the critical differential $d^{1,1}$ is zero, it is sufficient to have

$$H^3(Y^{\mathrm{tor}},\pi^*\Omega^3_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\mathcal{L}^{\mathrm{sub}}_\lambda)=H^3(\mathrm{Sh}^{\mathrm{tor}},\nabla^{\mathrm{sub}}(k_1-3,k_2-3))=0.$$

We summarize this discussion by saying that we have

$$H^i(\mathrm{Sh}^{\mathrm{tor}},\nabla^{\mathrm{sub}}(k_1-2,k_2-2))=0$$

for all $i>1$ and (k_1,k_2) such that

- $k_1\geq k_2+2$,
- $(k_1,k_2)\in\mathcal{C}_{\mathrm{ample},\varnothing}$,
- $H^3(\mathrm{Sh}^{\mathrm{tor}},\nabla^{\mathrm{sub}}(\mu_j^2+(k_1-1,k_2+1)))=0$ for $j=1,2$,
- $H^3(\mathrm{Sh}^{\mathrm{tor}},\nabla^{\mathrm{sub}}(k_1-3,k_2-3))=0$.

Now, we consider the spectral sequence in the case $e=2$. The graded pieces are

$$\mathrm{gr}_1=\pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}\quad\text{and}\quad\mathrm{gr}_0=\pi^*\Omega^2_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}).$$

The limit is

$$H^i(Y^{\mathrm{tor}},\Omega^2_{Y^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\mathcal{L}^{\mathrm{sub}}_\lambda)$$

and by the logarithmic Kodaira–Nakano theorem, it vanishes for all $i>2$ when $(k_1,k_2)\in\mathcal{C}_{\mathrm{ample},\varnothing}$. The differential

$$d^{2,1}:H^3(Y^{\mathrm{tor}},\pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}}\otimes\mathcal{L}^{\mathrm{sub}}_\lambda)\to\underbrace{H^4(Y^{\mathrm{tor}},\pi^*\Omega^2_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}})\otimes\mathcal{L}^{\mathrm{sub}}_\lambda)}_{=0}$$

is already 0 since the Siegel threefold has dimension 3. Under the additional hypothesis $k_1 \geq k_2 + 2$, we deduce that

$$H^i(Y^{\mathrm{tor}}, \pi^*\Omega^1_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \Omega^1_{Y^{\mathrm{tor}}/\,\mathrm{Sh}^{\mathrm{tor}}} \otimes \mathcal{L}^{\mathrm{sub}}_\lambda) = H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla(0,-2) \otimes \nabla^{\mathrm{sub}}(k_1-1, k_2+1)) = 0$$

for all $i > 2$ and $(k_1, k_2) \in \mathcal{C}_{\mathrm{ample},\varnothing}$. The tensor product of automorphic vector bundles

$$\nabla(0,-2) \otimes \nabla^{\mathrm{sub}}(k_1-1, k_2+1)$$

is filtered by the automorphic bundles

$$\nabla^{\mathrm{sub}}(\mu^1_j + (k_1-1, k_2+1)),$$

where $j = 1, 2, 3$ and since the vanishing

$$H^{i+1}(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(\mu^2_j + (k_1-1, k_2+1))) = 0$$

for $j = 1, 2$ is automatic, it implies the vanishing of

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}((0,-2) + (k_1-1, k_2+1))) = H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(k_1-1, k_2-1)).$$

We summarize this discussion by saying that we have

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(k_1-1, k_2-1)) = 0$$

for all $i > 2$ and (k_1, k_2) such that

- $k_1 \geq k_2 + 2$,
- $(k_1, k_2) \in \mathcal{C}_{\mathrm{ample},\varnothing}$.

We consider the case $I_0 = I = \{(1,-1)\}$. This case corresponds to $Y^{\mathrm{tor}}_{I_0} = \mathrm{Sh}^{\mathrm{tor}}$. In this degenerate case, the spectral sequence is trivial and the D-ample automorphic line bundle are powers of the determinant of the Hodge bundle : $\nabla(k,k)$ for all $k < 0$. By the logarithmic Kodaira–Nakano vanishing theorem, we have

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \Omega^j_{\mathrm{Sh}^{\mathrm{tor}}}(\log D_{\mathrm{red}}) \otimes \nabla^{\mathrm{sub}}(k,k)) = 0$$

for all $i + j > 3$ and $k < 0$. In the case $j = 3$, we get

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(k-3, k-3)) = 0$$

for all $i > 0$. In the case $j = 2$, we get

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(k-1, k-3)) = 0$$

for all $i > 1$. In the case $j = 1$, we get

$$H^i(\mathrm{Sh}^{\mathrm{tor}}, \nabla^{\mathrm{sub}}(k, k-2)) = 0$$

for all $i > 2$.

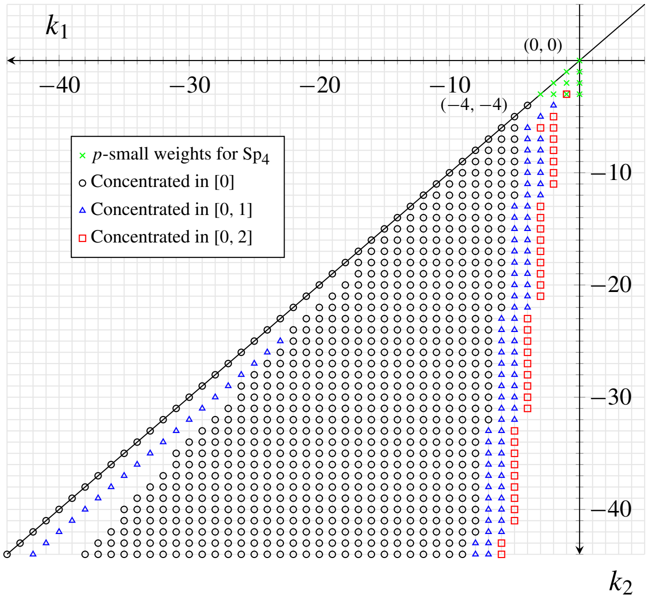

Figure 6. $g = 2$, $p = 11$. Notice that since the orbitally p-close condition is less restrictive with $p = 11$ than with $p = 5$, we are able to access more weights. However, if we look far enough, we notice the same phenomenon with the two regions corresponding to $I = \{(1, -1)\}$ and $I = \varnothing$.

7. Degeneration algorithm

7.1. ***Presentation.*** From the description of the degeneration of the different spectral sequences in the case of the Siegel threefold, it is clear that an algorithm implemented on a computer could be useful to make the vanishing results more explicit. We present an algorithm written in SageMath that uses our main result (Theorem 6.16) to compute new vanishing results from known ones. See github.com/ThibaultAlexandre/vanishing-results-over-the-siegel-variety

The algorithm depends on the following parameters:

(1) The genus $g \geq 2$ (the case $g = 1$ is obvious).

(2) A prime p such that $p > g^2$.

(3) A set of known vanishing result $\mathcal{C}_{\mathrm{van}}$ for each cohomological degree.

(4) The integer e that appears on the spectral sequence (2).

(5) A subset $I_0 \subset \Delta$ for the choice of the flag bundle $Y^{\mathrm{tor}}_{I_0}$ over the Siegel modular variety.

In the special case where $e = 0$, our algorithm does not need any vanishing result for the degeneration process as the spectral sequence (2) is concentrated on one row. In the special case where $e = d - 1$, the degeneration is automatic as it is given by the vanishing of the coherent cohomology in degree $i > d$. Then, these results can be used to run the algorithm with $e = 1$ and with differents $I_0 \subset \Delta$ and so on.

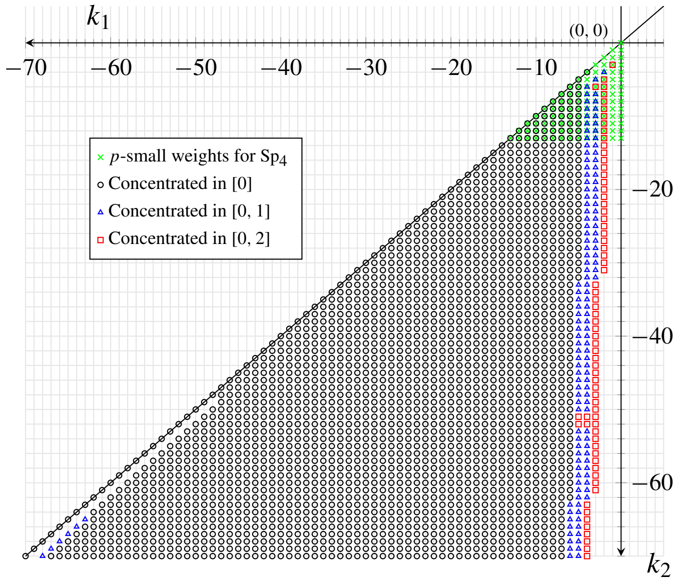

Figure 7. $g = 2$, $p = 31$.

Our SageMath code defines a class SiegelVariety with some methods that can be used to compute vanishing results. We create the Siegel threefold X over $\mathbb{F}_7$.

```
In [1]: X = SiegelVariety(g = 2, p = 7)
```

If the next line returns True, it means that the automorphic line bundle $\mathcal{L}_{(-2,-8)}$ is D-ample on the complete flag variety Y over X.

```
In [2]: X.ample([],[-2,-8])
```

```
Out[2]: True
```

The next line compute vanishing results for characters $\lambda = (k_1, k_2)$ with $-50 \le k_2 \le k_1 \le 0$ using the function $g_{I_0,e}$ in the case where $I_0 = \varnothing$ and $e = 0$. The results are registered in the list C_{van}. It returns True if the algorithm has found new vanishing results.

```
In [3]: X.compute([], e = 0, kmin = -50, kmax = 0)
```

```
Out[3]: True
```

The next line runs the compute method for each $I_0 \subset I$ and $0 \le e \le d-1$. We only need to specify the range of characters $\lambda = (k_1, k_2)$ we want to consider. It returns True if the algorithm has found new vanishing results. You may want to run this command several times until it returns False.

```
In [4]: X.computeAll(-50, 0)
```

```
Out[4]: True
```

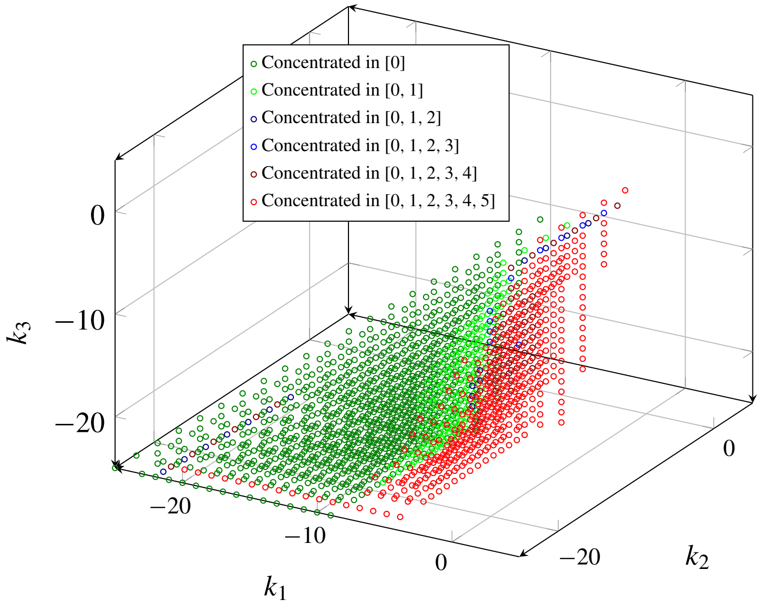

Figure 8. $g = 3$, $p = 11$.

The next line returns True if we know that

$$H^i(X, \nabla^{\mathrm{sub}}(-4, -6)) = 0$$

for all $i > 1$.

```
In [5]: X.vanishes(1,(-4,-6))
```

```
Out[5]: True
```

If the next line returns False, it means we don't know if

$$H^i(X, \nabla^{\mathrm{sub}}(-4, -6)) = 0$$

for all $i > 0$.

```
In [6]: X.vanishes(0,(-4,-6))
```

```
Out[6]: False
```

7.2. ***Explicit vanishing for $G = \mathrm{Sp}_4$.*** We plot some vanishing results we have obtained for the Siegel threefold with our algorithm. We have also added the p-small characters for Sp_4 with a twist by $-w_0$ to have them in the antidominant Weyl chamber.

7.3. ***Explicit vanishing for $G = \mathrm{Sp}_6$.*** We plot some vanishing results we have obtained in the case $g = 3$ with our algorithm. The weights live in a three-dimensional space and we need six different labels.

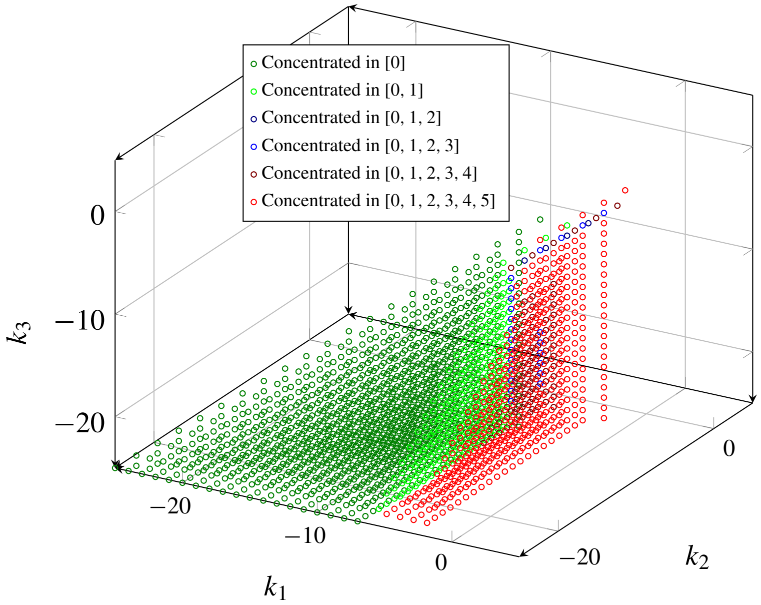

Figure 9. $g = 3$, $p = 691$.

Acknowledgements

I heartily thank Benoit Stroh for encouraging me to write this paper and for explaining me the relevance of generalized Hasse invariants to obtain positivity results on the flag bundle over the Siegel modular variety. I am very grateful to Jean-Stefan Koskivirta for answering my questions on the theory of G-Zips. I also thank Diego Berger, Yohan Brunebarbe and Arnaud Eteve for very helpful discussions.

References

[Andreatta 2023] F. Andreatta, “On two mod p period maps: Ekedahl–Oort and fine Deligne–Lusztig stratifications”, *Math. Ann.* **385**:1-2 (2023), 511–550. MR Zbl

[Ash et al. 1975] A. Ash, D. Mumford, M. Rapoport, and Y. Tai, *Smooth compactification of locally symmetric varieties*, Lie Groups Hist. Frontiers Appl. **4**, Math. Sci. Press, Brookline, MA, 1975. MR Zbl

[Brunebarbe et al.] Y. Brunebarbe, W. Goldring, J.-S. Koskivirta, and B. Stroh, “Ample automorphic bundles on zip-schemes”, in preparation.

[Cartier 1957] P. Cartier, “Une nouvelle opération sur les formes différentielles”, *C. R. Acad. Sci. Paris* **244** (1957), 426–428. MR Zbl

[Conrad] B. Conrad, “Applications of base change for coherent cohomology”, course notes, available at https://tinyurl.com/conradbasechange.

[Deligne 1970] P. Deligne, *Équations différentielles à points singuliers réguliers*, Lecture Notes in Math. **163**, Springer, 1970. MR Zbl

[Deligne and Illusie 1987] P. Deligne and L. Illusie, “Relèvements modulo p^2 et décomposition du complexe de de Rham”, *Invent. Math.* **89**:2 (1987), 247–270. MR Zbl

[Donkin 1985] S. Donkin, *Rational representations of algebraic groups: tensor products and filtration*, Lecture Notes in Math. **1140**, Springer, 1985. MR Zbl

[Esnault and Viehweg 1992] H. Esnault and E. Viehweg, *Lectures on vanishing theorems*, DMV Seminar **20**, Birkhäuser, Basel, 1992. MR Zbl

[Faltings and Chai 1990] G. Faltings and C.-L. Chai, *Degeneration of abelian varieties*, Ergebnisse der Math. (3) **22**, Springer, 1990. MR Zbl

[Goldring and Koskivirta 2019a] W. Goldring and J.-S. Koskivirta, "Strata Hasse invariants, Hecke algebras and Galois representations", *Invent. Math.* **217**:3 (2019), 887–984. MR Zbl

[Goldring and Koskivirta 2019b] W. Goldring and J.-S. Koskivirta, "Stratifications of flag spaces and functoriality", *Int. Math. Res. Not.* **2019**:12 (2019), 3646–3682. MR Zbl

[Harris 1985] M. Harris, "Arithmetic vector bundles and automorphic forms on Shimura varieties, I", *Invent. Math.* **82**:1 (1985), 151–189. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Grad. Texts in Math. **52**, Springer, 1977. MR Zbl

[Jantzen 2003] J. C. Jantzen, *Representations of algebraic groups*, 2nd ed., Math. Surv. Monogr. **107**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl

[Kempf et al. 1973] G. Kempf, F. F. Knudsen, D. Mumford, and B. Saint-Donat, *Toroidal embeddings, I*, Lecture Notes in Math. **339**, Springer, 1973. MR Zbl

[Lan 2012] K.-W. Lan, "Toroidal compactifications of PEL-type Kuga families", *Algebra Number Theory* **6**:5 (2012), 885–966. MR Zbl

[Lan and Suh 2012] K.-W. Lan and J. Suh, "Vanishing theorems for torsion automorphic sheaves on compact PEL-type Shimura varieties", *Duke Math. J.* **161**:6 (2012), 1113–1170. MR Zbl

[Lan and Suh 2013] K.-W. Lan and J. Suh, "Vanishing theorems for torsion automorphic sheaves on general PEL-type Shimura varieties", *Adv. Math.* **242** (2013), 228–286. MR Zbl

[Lazarsfeld 2004] R. Lazarsfeld, *Positivity in algebraic geometry, I: Classical setting: line bundles and linear series*, Ergebnisse der Math. (3) **48**, Springer, 2004. MR Zbl

[Mathieu 1990] O. Mathieu, "Filtrations of G-modules", *Ann. Sci. École Norm. Sup.* (4) **23**:4 (1990), 625–644. MR Zbl

[Moonen and Wedhorn 2004] B. Moonen and T. Wedhorn, "Discrete invariants of varieties in positive characteristic", *Int. Math. Res. Not.* **2004**:72 (2004), 3855–3903. MR Zbl

[Oort 2001] F. Oort, "A stratification of a moduli space of abelian varieties", pp. 345–416 in *Moduli of abelian varieties* (Texel Island, Netherlands, 1999), edited by C. Faber et al., Progr. Math. **195**, Birkhäuser, Basel, 2001. MR Zbl

[Pink et al. 2011] R. Pink, T. Wedhorn, and P. Ziegler, "Algebraic zip data", *Doc. Math.* **16** (2011), 253–300. MR Zbl

[Pink et al. 2015] R. Pink, T. Wedhorn, and P. Ziegler, "F-zips with additional structure", *Pacific J. Math.* **274**:1 (2015), 183–236. MR Zbl

[Polo and Tilouine 2002] P. Polo and J. Tilouine, "Bernstein–Gelfand–Gelfand complexes and cohomology of nilpotent groups over $\mathbb{Z}_{(p)}$ for representations with p-small weights", pp. 97–135 in *Cohomology of Siegel varieties*, Astérisque **280**, Soc. Math. France, Providence, RI, 2002. MR Zbl

[Riche 2016] S. Riche, *Geometric representation theory in positive characteristic*, Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand II, 2016, available at https://theses.hal.science/tel-01431526.

[Stacks 2005–] "The Stacks project", electronic reference, 2005–, available at http://stacks.math.columbia.edu.

[Wedhorn 1999] T. Wedhorn, "Ordinariness in good reductions of Shimura varieties of PEL-type", *Ann. Sci. École Norm. Sup.* (4) **32**:5 (1999), 575–618. MR Zbl

[Zhang 2018] C. Zhang, "Ekedahl–Oort strata for good reductions of Shimura varieties of Hodge type", *Canad. J. Math.* **70**:2 (2018), 451–480. MR Zbl

Communicated by Frank Calegari
Received 2023-02-12 Revised 2023-12-20 Accepted 2024-02-13

alexandre.thx@gmail.com
Institut de Mathématiques de Jussieu-Paris Rive Gauche, Sorbonne Université, Paris, France

ALGEBRA AND NUMBER THEORY 19:1 (2025)
https://doi.org/10.2140/ant.2025.19.195

Super-Hölder vectors and the field of norms

Laurent Berger and Sandra Rozensztajn

Let E be a field of characteristic p. In a previous paper of ours, we defined and studied super-Hölder vectors in certain E-linear representations of $\mathbb{Z}_p$. In the present paper, we define and study super-Hölder vectors in certain E-linear representations of a general p-adic Lie group. We then consider certain p-adic Lie extensions K_∞/K of a p-adic field K, and compute the super-Hölder vectors in the tilt of K_∞. We show that these super-Hölder vectors are the perfection of the field of norms of K_∞/K. By specializing to the case of a Lubin–Tate extension, we are able to recover $E((Y))$ inside the Y-adic completion of its perfection, seen as a valued E-vector space endowed with the action of $\mathcal{O}_K^\times$ given by the endomorphisms of the corresponding Lubin–Tate group.

Introduction

Let E be a field of characteristic p, for example a finite field. In our paper [Berger and Rozensztajn 2022], we defined and studied super-Hölder vectors in certain E-linear representations of the p-adic Lie group $\mathbb{Z}_p$. These vectors are a characteristic p analogue of locally analytic vectors. They allowed us to recover $E((X))$ inside the X-adic completion of its perfection, seen as a valued E-vector space endowed with the action of $\mathbb{Z}_p^\times$ given by $a \cdot f(X) = f((1+X)^a - 1)$.

In the present paper, we define and study super-Hölder vectors in certain E-linear representations of a general p-adic Lie group. We then consider certain p-adic Lie extensions K_∞/K of a p-adic field K, and compute the super-Hölder vectors in the tilt of K_∞. We show that these super-Hölder vectors are the perfection of the field of norms of K_∞/K. By specializing to the case of a Lubin–Tate extension, we are able to recover $E((Y))$ inside the Y-adic completion of its perfection, seen as a valued E-vector space endowed with the action of $\mathcal{O}_K^\times$ given by the endomorphisms of the corresponding Lubin–Tate group.

MSC2020: primary 11S15, 11S20, 11S80; secondary 11S31, 12J25, 13J05, 22E35.
Keywords: Super-Hölder function, locally analytic function, field of norms, pro-p-group, perfectoid ring, decompletion, Lubin–Tate group, Mahler basis, p-adic wavelet.

We now give more details about the contents of our paper. Let Γ be a p-adic Lie group. It is known that Γ always has a uniform open pro-p subgroup G. Let G be such a subgroup, and let $G_i = G^{p^i}$ for $i \geqslant 0$. Let M be an E-vector space, endowed with a valuation val_M such that $\mathrm{val}_M(xm) = \mathrm{val}_M(m)$ if $x \in E^\times$. We assume that M is separated and complete for the val_M-adic topology. We say that a function $f : G \to M$ is super-Hölder if there exist constants $e > 0$ and $\lambda, \mu \in \mathbb{R}$ such that $\mathrm{val}_M(f(g) - f(h)) \geqslant p^\lambda \cdot p^{ei} + \mu$ whenever $gh^{-1} \in G_i$, for all $g, h \in G$ and $i \geqslant 0$. If M is now endowed with an action of G by isometries, and $m \in M$, we say that m is a super-Hölder vector if the orbit map $g \mapsto g \cdot m$ is a super-Hölder function $G \to M$. We let $M^{G\text{-}e\text{-sh},\lambda}$ denote the space of super-Hölder vectors for given constants e and λ as in the definition above. The space of vectors of M that are super-Hölder for a given e is independent of the choice of the uniform subgroup G, and denoted by $M^{e\text{-sh}}$. When $G = \mathbb{Z}_p$ and $e = 1$, we recover the definitions of [Berger and Rozensztajn 2022]. If Γ is a p-adic Lie group and $e = 1$, we get an analogue of locally $\mathbb{Q}_p$-analytic vectors. If K is a finite extension of $\mathbb{Q}_p$, Γ is the Galois group of a Lubin–Tate extension of K, and $e = [K : \mathbb{Q}_p]$, we seem to get an analogue of locally K-analytic vectors.

From now on, assume that $p \neq 2$. Let K be a p-adic field and let K_∞/K be an almost totally ramified p-adic Lie extension, with Galois group Γ of dimension $d \geqslant 1$. The tilt of K_∞ is the fraction field $\widetilde{\mathbb{E}}_{K_\infty}$ of $\varprojlim_{x \mapsto x^p} \mathcal{O}_{K_\infty}/p$. It is a perfect complete valued field of characteristic p, endowed with an action of Γ by isometries. The field $\widetilde{\mathbb{E}}_{K_\infty}$ naturally contains the field of norms $X_K(K_\infty)$ of the extension K_∞/K, and it is known that $\widetilde{\mathbb{E}}_{K_\infty}$ is the completion of the perfection of $X_K(K_\infty)$. We have the following result (Theorem 2.2.3).

Theorem A. *We have* $\widetilde{\mathbb{E}}_{K_\infty}^{d\text{-sh}} = \bigcup_{n \geqslant 0} \varphi^{-n}(X_K(K_\infty))$.

Assume now that K is a finite extension of $\mathbb{Q}_p$, with residue field k, and let LT be a Lubin–Tate formal group attached to K. Let K_∞ be the extension of K generated by the torsion points of LT, so that $\mathrm{Gal}(K_\infty/K)$ is isomorphic to $\mathcal{O}_K^\times$. The field of norms $X_K(K_\infty)$ is isomorphic to $k((Y))$, and $\mathcal{O}_K^\times$ acts on this field by the endomorphisms of the Lubin–Tate group: $a \cdot f(Y) = f([a](Y))$. Let $d = [K : \mathbb{Q}_p]$. The following (Theorem 3.2.1) is a more precise version of Theorem A in this situation.

Theorem B. *If* $j \geqslant 1$, *then* $\widetilde{\mathbb{E}}_{K_\infty}^{1+p^j\mathcal{O}_K\text{-}d\text{-sh},dj} = k((Y))$.

If $K = \mathbb{Q}_p$ and K_∞/K is the cyclotomic extension, Theorem B was proved in [Berger and Rozensztajn 2022]. A crucial ingredient of the proof of this theorem was Colmez' analogue of Tate traces for $\widetilde{\mathbb{E}}_{K_\infty}$. If the Lubin–Tate group is of height $\geqslant 2$, there are no such traces (we state and prove a precise version of this assertion in Section 3.2). Instead of Tate traces, we use a theorem of Ax and a precise characterization of the field of norms $X_K(K_\infty)$ inside $\widetilde{\mathbb{E}}_{K_\infty}$ in order to prove Theorem A.

As an application of Theorem B, we compute the perfectoid commutant of Aut(LT). If $b \in \mathcal{O}_K^\times$ and $n \in \mathbb{Z}$, then $u(Y) = [b](Y^{q^n})$ is an element of $\widetilde{\mathbb{E}}_{K_\infty}^+$ that satisfies the functional equation $u \circ [g](Y) = [g] \circ u(Y)$ for all $g \in \mathcal{O}_K^\times$. Conversely, we prove the following (Theorem 3.3.1).

Theorem C. *If* $u \in \widetilde{\mathbb{E}}_{K_\infty}^+$ *is such that* $\mathrm{val}_Y(u) > 0$ *and* $u \circ [g] = [g] \circ u$ *for all* $g \in \mathcal{O}_K^\times$, *there exists* $b \in \mathcal{O}_K^\times$ *and* $n \in \mathbb{Z}$ *such that* $u(Y) = [b](Y^{q^n})$.

In the last section, we give a characterization of super-Hölder functions on a uniform pro-p group in terms of their Mahler expansions (Theorem 4.3.4). In order to do so, we prove some results of independent interest on the space of continuous functions on $\mathcal{O}_K^d$ with values in a valued E-vector space M as above.

At the end of [Berger and Rozensztajn 2022], we suggested an application of super-Hölder vectors for the action of $\mathbb{Z}_p$ to the p-adic local Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$. We hope that this general theory of super-Hölder vectors, especially in the Lubin–Tate case, will have applications to the p-adic local Langlands correspondence for other fields than $\mathbb{Q}_p$.

1. Super-Hölder functions and vectors

In this section, we define super-Hölder vectors inside a valued E-vector space M endowed with an action of a p-adic Lie group Γ. The definition is very similar to the one that we gave for $\Gamma = \mathbb{Z}_p$ in [Berger and Rozensztajn 2022]. The main new technical tool is the existence of uniform open subgroups of Γ. These uniform subgroups look very much like $\mathbb{Z}_p^d$ in a sense that we make precise.

1.1. ***Uniform pro-p groups.*** Uniform pro-p groups are defined at the beginning of Section 4 of [Dixon et al. 1991]. We do not recall the definition, nor the notion of rank of a uniform pro-p group, but rather point out the following properties of uniform pro-p groups. A coordinate (below) is simply a homeomorphism.

Proposition 1.1.1. *If G is a uniform pro-p group of rank d, then*:

(1) $G_i = \{g^{p^i}, g \in G\}$ *is an open normal* (*and uniform*) *subgroup of G for $i \geqslant 0$.*

(2) *We have $[G_i : G_{i+1}] = p^d$ for $i \geqslant 0$.*

(3) *There is a coordinate $c : G \to \mathbb{Z}_p^d$ such that $c(G_i) = (p^i\mathbb{Z}_p)^d$ for $i \geqslant 0$.*

(4) *If $g, h \in G$, then $gh^{-1} \in G_i$ if and only if $c(g) - c(h) \in (p^i\mathbb{Z}_p)^d$.*

Proof. Properties (1)–(4) are proved in Section 4 of [Dixon et al. 1991]. Alternatively, a uniform pro-p group G has a natural integer valued p-valuation ω such that (G, ω) is saturated [Klopsch 2005, Remark 2.1]. Properties (1)–(4) are then proved in Section 26 of [Schneider 2011]. □

For example, the pro-p group $\mathbb{Z}_p^d$ is uniform for all $d \geqslant 1$.

Lemma 1.1.2. *If G is a uniform pro-p group, and H is a uniform open subgroup of G, there exists $j \geqslant 0$ such that $G_{i+j} \subset H_i$ for all $i \geqslant 0$.*

Proof. This follows from the fact that $\{G_i\}_{i\geqslant 0}$ forms a basis of neighborhoods of the identity in G. □

A p-adic Lie group is a p-adic manifold that has a compatible group structure. For example, $\mathrm{GL}_n(\mathbb{Z}_p)$ and its closed subgroups are p-adic Lie groups. We refer to [Schneider 2011] for a comprehensive treatment of the theory. Every uniform pro-p group is a p-adic Lie group. Conversely, we have the following.

Proposition 1.1.3. *Every p-adic Lie group Γ has a uniform open subgroup G, and the rank of G is the dimension of Γ.*

Proof. See Interlude A of [Dixon et al. 1991, pages 97–98]. □

Proposition 1.1.4. *Let G be a pro-p group of finite rank, and N a closed normal subgroup of G. There exists an open subgroup G' of G such that G', $G' \cap N$ and $G'/G' \cap N$ are all uniform.*

Proof. This is stated and proved on page 64 of [Dixon et al. 1991] (their H is our G'). □

1.2. ***Super-Hölder functions and vectors.*** Let M be an E-vector space, endowed with a valuation val_M such that $\mathrm{val}_M(xm) = \mathrm{val}_M(m)$ if $x \in E^\times$. We assume that M is separated and complete for the val_M-adic topology. Throughout this section, G denotes a uniform pro-p group.

Definition 1.2.1. We say that $f : G \to M$ is super-Hölder if there exist constants $\lambda, \mu \in \mathbb{R}$ and $e > 0$ such that $\mathrm{val}_M(f(g) - f(h)) \geqslant p^\lambda \cdot p^{ei} + \mu$ whenever $gh^{-1} \in G_i$, for all $g, h \in G$ and $i \geqslant 0$.

Remark 1.2.2. If $G = \mathbb{Z}_p$ and $e = 1$, we recover the functions defined in [Berger and Rozensztajn 2022, Section 1.1]; see also Remark 1.12 of [loc. cit.].

In the above definition, e will usually be equal to either 1 or $\dim(G)$.

We let $\mathcal{H}_e^{\lambda,\mu}(G, M)$ denote the space of functions such that $\mathrm{val}_M(f(g) - f(h)) \geqslant p^\lambda \cdot p^{ei} + \mu$ whenever $gh^{-1} \in G_i$, for all $g, h \in G$ and $i \geqslant 0$, and $\mathcal{H}_e^\lambda(G, M) = \bigcup_{\mu\in\mathbb{R}} \mathcal{H}_e^{\lambda,\mu}(G, M)$ and $\mathcal{H}_e(G, M) = \bigcup_{\lambda\in\mathbb{R}} \mathcal{H}_e^\lambda(G, M)$.

If M, N are two valued E-vector spaces, and $f : M \to N$ is an E-linear map, we say that f is Hölder-continuous if there exists $c > 0$, $d \in \mathbb{R}$ such that $\mathrm{val}_N(f(x)) \geqslant c \cdot \mathrm{val}_M(x) + d$ for all $x \in M$.

Proposition 1.2.3. *If $\pi : M \to N$ is a Hölder-continuous linear map, we get a map $\mathcal{H}_e(G, M) \to \mathcal{H}_e(G, N)$.*

Proof. Take $c, d \in \mathbb{R}$ of Hölder continuity for π, $f \in \mathcal{H}_e^{\lambda,\mu}(G, M)$, and $g, h \in G$ with $gh^{-1} \in G_i$. We have $\mathrm{val}_N(\pi(f(g)) - \pi(f(h))) \geqslant c \cdot \mathrm{val}_M(f(g) - f(h)) + d \geqslant cp^\lambda \cdot p^{ei} + (\mu + d)$, so that $\pi \circ f \in \mathcal{H}_e^{\lambda',\mu'}(G, N)$ with $p^{\lambda'} = cp^\lambda$, and $\mu' = \mu + d$. □

Proposition 1.2.4. *If $\alpha : G \to H$ is a group homomorphism, we get a map $\alpha^* : \mathcal{H}_e(H, M) \to \mathcal{H}_e(G, M)$.*

Proof. By definition of the subgroups G_i and H_i, we have $\alpha(G_i) \subset H_i$ for all i. Take $f \in \mathcal{H}_e^{\lambda,\mu}(H, M)$, and $g, h \in G$ with $gh^{-1} \in G_i$. We have $\mathrm{val}_M(f(\alpha(g)) - f(\alpha(h))) \geqslant p^\lambda \cdot p^{ei} + \mu$ as $\alpha(g)\alpha(h)^{-1} \in H_i$, so that $\alpha^*(f) = f \circ \alpha \in \mathcal{H}_e^{\lambda,\mu}(G, M)$. □

Proposition 1.2.5. *Suppose that M is a ring, and that $\mathrm{val}_M(mm') \geqslant \mathrm{val}_M(m) + \mathrm{val}_M(m')$ for all $m, m' \in M$. If $c \in \mathbb{R}$, let $M_c = M^{\mathrm{val}_M \geqslant c}$:*

(1) *If $f \in \mathcal{H}_e^{\lambda,\mu}(G, M_c)$ and $g \in \mathcal{H}_e^{\lambda,\nu}(G, M_d)$, and $\xi = \min(\mu + d, \nu + c)$, then $fg \in \mathcal{H}_e^{\lambda,\xi}(G, M_{c+d})$.*

(2) *If $\lambda, \mu \in \mathbb{R}$, then $\mathcal{H}_e^{\lambda,\mu}(G, M_0)$ is a subring of $C^0(G, M)$.*

(3) *If $\lambda \in \mathbb{R}$, then $\mathcal{H}_e^\lambda(G, M)$ is a subring of $C^0(G, M)$.*

Proof. Items (2) and (3) follow from item (1), which we now prove. If $x, y \in G$, then

$$(fg)(x) - (fg)(y) = (f(x) - f(y))g(x) + (g(x) - g(y))f(y),$$

which implies the claim. □

We now assume that M is endowed with an E-linear action by isometries of G. If $m \in M$, let $\mathrm{orb}_m : G \to M$ denote the function defined by $\mathrm{orb}_m(g) = g \cdot m$.

Definition 1.2.6. Let $M^{G\text{-}e\text{-sh},\lambda,\mu}$ be those $m \in M$ such that $\mathrm{orb}_m \in \mathcal{H}_e^{\lambda,\mu}(G, M)$, and let $M^{G\text{-}e\text{-sh},\lambda}$ and $M^{G\text{-}e\text{-sh}}$ be the corresponding sub-E-vector spaces of M.

Remark 1.2.7. We assume that G acts by isometries on M, but not that G acts continuously on M, namely that $G \times M \to M$ is continuous. However, let M^{cont} denote the set of $m \in M$ such that $\mathrm{orb}_m : G \to M$ is continuous. It is easy to see that M^{cont} is a closed sub-E-vector space of M, and that $G \times M^{\mathrm{cont}} \to M^{\mathrm{cont}}$ is continuous; compare with Section 3 of [Emerton 2017]. We then have $M^{\mathrm{sh}} \subset M^{\mathrm{cont}}$.

Lemma 1.2.8. *If $m \in M$, then $m \in M^{G\text{-}e\text{-sh},\lambda,\mu}$ if and only if for all $i \geqslant 0$, we have $\mathrm{val}_M(g \cdot m - m) \geqslant p^\lambda \cdot p^{ei} + \mu$ for all $g \in G_i$.*

Proof. If $m \in M$, then $m \in M^{G\text{-}e\text{-sh},\lambda,\mu}$ if and only if the function orb_m is in $\mathcal{H}_e^{\lambda,\mu}(G, M)$, that is, for all g, h with $gh^{-1} \in G_i$, we have $\mathrm{val}_M(g \cdot m - h \cdot m) \geqslant p^\lambda \cdot p^{ei} + \mu$. As G acts by isometries, we have $\mathrm{val}_M(g \cdot m - h \cdot m) = \mathrm{val}_M(h^{-1}g \cdot m - m)$. The result follows, as $h^{-1}g = h^{-1} \cdot gh^{-1} \cdot h \in G_i$. □

Lemma 1.2.9. *The space $M^{G\text{-}e\text{-sh},\lambda,\mu}$ is a closed sub-E-vector space of M.*

Lemma 1.2.10. *If $i_0 \geqslant 0$, and $m \in M$ is such that $\mathrm{val}_M(g \cdot m - m) \geqslant p^\lambda \cdot p^{ei} + \mu$ for all $g \in G_i$ with $i \geqslant i_0$, then $m \in M^{G\text{-}e\text{-sh},\lambda}$.*

Proof. Take $i < i_0$, and let R_i be a set of representatives of $G_{i_0} \backslash G_i$. This is a finite set, so there exists $\mu_i \in \mathbb{R}$ such that $\mathrm{val}_M(r \cdot m - m) \geqslant p^\lambda \cdot p^{ei} + \mu_i$ for all $r \in R_i$. If $g \in G_i$, it can be written as $g = hr$ for some $h \in G_{i_0}$ and $r \in R_i$. We then have $g \cdot m - m = hr \cdot m - h \cdot m + h \cdot m - m$, so that $\mathrm{val}_M(g \cdot m - m) \geqslant \min(\mathrm{val}_M(r \cdot m - m), \mathrm{val}_M(h \cdot m - m))$ (recall that G acts by isometries), so $\mathrm{val}_M(g \cdot m - m) \geqslant \min(p^\lambda \cdot p^{ei} + \mu_i, p^\lambda \cdot p^{ei_0} + \mu) \geqslant p^\lambda \cdot p^{ei} + \min(\mu, \mu_i)$ as $i_0 > i$. If μ' is the min of μ and the μ_i for $0 \leqslant i < i_0$, then $m \in M^{G\text{-}e\text{-sh},\lambda,\mu'}$. □

Recall that if $k \geqslant 0$, then G_k is also a uniform pro-p group.

Lemma 1.2.11. *If $k \geqslant 0$ then $M^{G\text{-}e\text{-sh},\lambda} = M^{G_k\text{-}e\text{-sh},\lambda+k}$.*

Proof. Note that $(G_k)_i = G_{i+k}$. The inclusion $M^{G\text{-}e\text{-sh},\lambda} \subset M^{G_k\text{-}e\text{-sh},\lambda+k}$ is obvious, and the reverse inclusion follows from Lemma 1.2.10. □

Proposition 1.2.12. *The space $M^{H\text{-}e\text{-sh}}$ does not depend on the choice of a uniform open subgroup $H \subset G$.*

Proof. Let H and H' be uniform open subgroups of G. The group $H \cap H'$ contains an open uniform subgroup by Proposition 1.1.3, so to prove the proposition, we can further assume that $H' \subset H$. We then have $H'_i \subset H_i$ for all i, so that if $m \in M^{H\text{-}e\text{-sh},\lambda,\mu}$, then $m \in M^{H'\text{-}e\text{-sh},\lambda,\mu}$. This implies that

$M^{H\text{-}e\text{-sh},\lambda} \subset M^{H'\text{-}e\text{-sh},\lambda}$. Conversely, by Lemma 1.1.2, there exists j such that $H_j \subset H'$. The previous reasoning implies that $M^{H'\text{-}e\text{-sh},\lambda} \subset M^{H_j\text{-}e\text{-sh},\lambda}$. Lemma 1.2.11 now implies that $M^{H_j\text{-}e\text{-sh},\lambda} = M^{H\text{-}e\text{-sh},\lambda-j}$.

These inclusions imply the proposition. □

Definition 1.2.13. If Γ is a p-adic Lie group that acts by isometries on M, we let $M^{e\text{-sh}} = M^{G\text{-}e\text{-sh}}$ where G is any uniform open subgroup of Γ.

Remark 1.2.14. If $e \leqslant f$, then $M^{f\text{-sh}} \subset M^{e\text{-sh}}$.

Recall that G is a uniform pro-p group. If a closed normal subgroup N of G acts trivially on M, then G/N acts on M.

Proposition 1.2.15. *If a closed normal subgroup N of G acts trivially on M, then $M^{G\text{-}e\text{-sh}} = M^{G/N\text{-}e\text{-sh}}$.*

Proof. By Proposition 1.1.4, G has an open subgroup G' such that G' and G'/N' are uniform (where $N' = G' \cap N$). By Proposition 1.2.12, we have $M^{G\text{-}e\text{-sh}} = M^{G'\text{-}e\text{-sh}}$ and $M^{G/N\text{-}e\text{-sh}} = M^{G'/N'\text{-}e\text{-sh}}$. Let $\pi : G' \to G'/N'$ denote the projection. We have $\pi(G'_i) = (G'/N')_i$ for all i. Hence if $m \in M$, then $\mathrm{val}_M(g \cdot m - m) \geqslant p^\lambda \cdot p^{ei} + \mu$ for all $g \in G'_i$ if and only if $\mathrm{val}_M(\pi(g) \cdot m - m) \geqslant p^\lambda \cdot p^{ei} + \mu$ for all $\pi(g) \in (G'/N')_i$. □

Proposition 1.2.16. *Suppose that M is a ring, and that $g(mm') = g(m)g(m')$ and $\mathrm{val}_M(mm') \geqslant \mathrm{val}_M(m) + \mathrm{val}_M(m')$ for all $m, m' \in M$ and $g \in G$:*

(1) *If $v \in \mathbb{R}$ and $m, m' \in M^{G\text{-}e\text{-sh},\lambda,\mu} \cap M^{\mathrm{val}_M \geqslant v}$, then $m \cdot m' \in M^{G\text{-}e\text{-sh},\lambda,\mu+v}$.*

(2) *If $m \in M^{G\text{-}e\text{-sh},\lambda,\mu} \cap M^\times$, then $1/m \in M^{G\text{-}e\text{-sh},\lambda,\mu-2\,\mathrm{val}_M(m)}$.*

Proof. Item (1) follows from Proposition 1.2.5 and Lemma 1.2.8. Item (2) follows from

$$g\left(\frac{1}{m}\right) - \frac{1}{m} = \frac{m - g(m)}{g(m)m}.$$ □

2. The field of norms

Let K be a p-adic field, and let K_∞ be an algebraic Galois extension of K, whose Galois group G is a p-adic Lie group of dimension $\geqslant 1$. We assume that K_∞/K is almost totally ramified, namely that the inertia subgroup of G is open in G. Let $d = \dim(G)$ and let $\ell = p^d$. Let $\widetilde{\mathbb{E}}^+_{K_\infty}$ denote the ring $\varprojlim_{x \mapsto x^\ell} \mathcal{O}_{K_\infty}/p$. This is a perfect domain of characteristic p, which has a natural action of G. The map $(y_j)_{j \geqslant 0} \mapsto (y_{di})_{i \geqslant 0}$ gives an isomorphism between $\varprojlim_{x \mapsto x^p} \mathcal{O}_{K_\infty}/p$ and $\widetilde{\mathbb{E}}^+_{K_\infty}$, so that $\widetilde{\mathbb{E}}^+_{K_\infty}$ is the ring of integers of the tilt of $\hat{K}_\infty$; see Section 3 of [Scholze 2012].

If $x = (x_i)_{i \geqslant 0}$, and $\hat{x}_i$ is a lift of x_i to $\mathcal{O}_{K_\infty}$, then $\ell^i \mathrm{val}_p(\hat{x}_i)$ is independent of $i \geqslant 0$ such that $x_i \neq 0$. We define a valuation on $\widetilde{\mathbb{E}}^+_{K_\infty}$ by $\mathrm{val}_{\mathrm{E}}(x) = \lim_{i \to +\infty} \ell^i \mathrm{val}_p(\hat{x}_i)$.

The aim of this section is to compute $(\widetilde{\mathbb{E}}^+_{K_\infty})^{d\text{-sh}}$. Given Definition 1.2.13, we assume from now on (replacing K by a finite subextension if necessary) that G is uniform and that K_∞/K is totally ramified. Let k denote the common residue field of K and K_∞.

2.1. ***The field of norms.*** Let $\mathcal{E}(K_\infty)$ denote the set of finite extensions E of K such that $E \subset K_\infty$. Let $X_K(K_\infty)$ denote the set of sequences $(x_E)_{E \in \mathcal{E}(K_\infty)}$ such that $x_E \in E$ for all $E \in \mathcal{E}(K_\infty)$, and $\mathrm{N}_{F/E}(x_F) = x_E$ whenever $E \subset F$ with $E, F \in \mathcal{E}(K_\infty)$.

If $n \geqslant 0$, let $K_n = K_\infty^{G_n}$ so that $[K_{n+1} : K_n] = \ell$, $\{K_n\}_{n \geqslant 0}$ is a cofinal subset of $\mathcal{E}(K_\infty)$, and $X_K(K_\infty) = \varprojlim_{\mathrm{N}_{K_n/K_{n-1}}} K_n$. If $x = (x_n)_{n \geqslant 0} \in X_K(K_\infty)$, let $\mathrm{val_E}(x) = \mathrm{val}_p(x_0)$.

Theorem 2.1.1. *Let K and K_∞ be as above*:

(1) *If $x, y \in X_K(K_\infty)$, then $\{\mathrm{N}_{K_{n+j}/K_n}(x_{n+j} + y_{n+j})\}_{j \geqslant 0}$ converges for all $n \geqslant 0$.*

(2) *If we set $(x+y)_n = \lim_{j \to +\infty} \mathrm{N}_{K_{n+j}/K_n}(x_{n+j} + y_{n+j})$, then $x + y \in X_K(K_\infty)$, and the set $X_K(K_\infty)$ with this addition law, and componentwise multiplication, is a field of characteristic p.*

(3) *The function $\mathrm{val_E}$ is a valuation on $X_K(K_\infty)$, for which it is complete.*

(4) *If $\varpi = (\varpi_n)_{n \geqslant 0}$ is a norm compatible sequence of uniformizers of $\mathcal{O}_{K_n}$, the valued field $X_K(K_\infty)$ is isomorphic to $k((\varpi))$ (with $\mathrm{val}(\varpi) = \mathrm{val}_p(\varpi_0)$).*

Proof. By a result of Sen [1972], K_∞/K is strictly APF in the terminology of Section 1.2 of [Wintenberger 1983]; see 1.2.2 of [loc. cit.]. The theorem is then proved in Section 2 of [loc. cit.]. □

Let $X_K^+(K_\infty) = \varprojlim_{\mathrm{N}_{K_n/K_{n-1}}} \mathcal{O}_{K_n}$ be the ring of integers of the valued field $X_K(K_\infty)$.

If $c > 0$, let $I_n^c = \{x \in \mathcal{O}_{K_n}$ such that $\mathrm{val}_p(x) \geqslant c\}$. If $m, n \geqslant 0$, the map $\mathcal{O}_{K_n}/I_n^c \to \mathcal{O}_{K_{m+n}}/I_{m+n}^c$ is well-defined and injective.

Proposition 2.1.2. *There exists $c(K_\infty/K) \leqslant 1$ such that if $0 < c \leqslant c(K_\infty/K)$, then*

$$\mathrm{val}_p(\mathrm{N}_{K_{n+k}/K_n}(x)/x^{[K_{n+k}:K_n]} - 1) \geqslant c$$

for all $n, k \geqslant 0$ and $x \in \mathcal{O}_{K_{n+k}}$.

Proof. See [Wintenberger 1983] as well as Section 4 of [Cais and Davis 2015]. The result follows from the fact (see 1.2.2 of [Wintenberger 1983]) that the extension K_∞/K is strictly APF. One can then apply 1.2.1, 4.2.2 and 1.2.3 of [Wintenberger 1983]. □

Using Proposition 2.1.2, we get a map $\iota : X_K^+(K_\infty) \to \varprojlim_{x \mapsto x^\ell} \mathcal{O}_{K_\infty}/I_\infty^c$ given by

$$(x_n)_{n \geqslant 0} \in \varprojlim_{\mathrm{N}_{K_n/K_{n-1}}} \mathcal{O}_{K_n} \mapsto (\bar{x}_n)_{n \geqslant 0}.$$

Let $\varprojlim_{x \mapsto x^\ell} \mathcal{O}_{K_n}/I_n^c$ denote the set of $(x_n)_{n \geqslant 0} \in \varprojlim_{x \mapsto x^\ell} \mathcal{O}_{K_\infty}/I_\infty^c$ such that $x_n \in \mathcal{O}_{K_n}/I_n^c$ for all $n \geqslant 0$.

Proposition 2.1.3. *Let $0 < c \leqslant c(K_\infty/K)$ be as in Proposition 2.1.2*:

(1) *The natural map $\widetilde{\mathbb{E}}_{K_\infty}^+ \to \varprojlim_{x \mapsto x^\ell} \mathcal{O}_{K_\infty}/I_\infty^c$ is a bijection.*

(2) *The map $\iota : X_K^+(K_\infty) \to \varprojlim_{x \mapsto x^\ell} \mathcal{O}_{K_\infty}/I_\infty^c = \widetilde{\mathbb{E}}_{K_\infty}^+$ is injective and isometric.*

(3) *The image of ι is $\varprojlim_{x \mapsto x^\ell} \mathcal{O}_{K_n}/I_n^c$.*

Proof. See [Wintenberger 1983] and Section 4 of [Cais and Davis 2015]. We give a few more details for the convenience of the reader. Item (1) is classical; see for instance Proposition 4.2 of [Cais and Davis 2015]. The map ι is obviously injective and isometric. For (3), choose $x = (x_n)_{n\geqslant 0} \in \varprojlim_{x\mapsto x^\ell} \mathcal{O}_{K_n}/I_n^c$, and choose a lift $\hat{x}_n \in \mathcal{O}_{K_n}$ of x_n. One proves that $\{\mathrm{N}_{K_{n+j}/K_n}(\hat{x}_{n+j})\}_{j\geqslant 0}$ converges to some $y_n \in \mathcal{O}_{K_n}$, and that $(y_n)_{n\geqslant 0} \in X_K^+(K_\infty)$ is a lift of $(x_n)_{n\geqslant 0}$. See Section 4 of [loc. cit.] for details, for instance the proof of Lemma 4.1. □

Proposition 2.1.3 allows us to see $X_K^+(K_\infty)$, and hence $\varphi^{-n}(X_K^+(K_\infty))$ for all $n \geqslant 0$, as a subring of $\widetilde{\mathbb{E}}^+_{K_\infty}$.

Proposition 2.1.4. *The ring* $\bigcup_{n\geqslant 0} \varphi^{-n}(X_K^+(K_\infty))$ *is dense in* $\widetilde{\mathbb{E}}^+_{K_\infty}$.

Proof. See Section 4.3 of [Wintenberger 1983]. □

2.2. ***Decompleting the tilt.*** We now compute $(\widetilde{\mathbb{E}}^+_{K_\infty})^{d\text{-sh}}$. Since Proposition 2.2.1 below is vacuous if $p = 2$, we assume in this section that $p \neq 2$.

Proposition 2.2.1. *If* $0 < c \leqslant 1 - 1/(p-1)$, *and* $x \in \mathcal{O}_{K_\infty}$ *is such that* $\mathrm{val}_p(g(x) - x) \geqslant 1$ *for all* $g \in G_n$, *then the image of* x *in* $\mathcal{O}_{K_\infty}/I_\infty^c$ *belongs to* $\mathcal{O}_{K_n}/I_n^c$.

Proof. If $\mathrm{val}_p(g(x) - x) \geqslant 1$ for all $g \in \mathrm{Gal}(K^{\mathrm{alg}}/K_n)$, then by Theorem 1.7 of [Le Borgne 2010] (an optimal version of a theorem of Ax), there exists $y \in K_n$ such that $\mathrm{val}_p(x - y) \geqslant 1 - 1/(p-1)$. This implies the proposition. □

Proposition 2.2.2. *If* $c = p^\gamma$ *is as above, then* $X_K^+(K_\infty) \subset (\widetilde{\mathbb{E}}^+_{K_\infty})^{G\text{-}d\text{-sh},\gamma,0}$.

Proof. Take $x = (x_n)_{n\geqslant 0} \in \varprojlim_{x\mapsto x^\ell} \mathcal{O}_{K_n}/I_n^c$. If $g \in G_i$, then $g(x_n) = x_n$ for $n \leqslant i$, so that $\mathrm{val}_{\mathrm{E}}(gx - x) \geqslant p^{di} p^\gamma$. □

Theorem 2.2.3. *We have*:

(1) $(\widetilde{\mathbb{E}}^+_{K_\infty})^{G\text{-}d\text{-sh},0,0} \subset X_K^+(K_\infty)$.

(2) $(\widetilde{\mathbb{E}}^+_{K_\infty})^{d\text{-sh}} = \bigcup_{n\geqslant 0} \varphi^{-n}(X_K^+(K_\infty))$ *and* $\widetilde{\mathbb{E}}^{d\text{-sh}}_{K_\infty} = \bigcup_{n\geqslant 0} \varphi^{-n}(X_K(K_\infty))$.

Proof. Take $c \leqslant \min(c(K_\infty/K), 1 - 1/(p-1))$. Take $x = (x_n)_{n\geqslant 0} \in \varprojlim_{x\mapsto x^\ell} \mathcal{O}_{K_\infty}/p$. If $n \geqslant 0$ and $x \in (\widetilde{\mathbb{E}}^+_{K_\infty})^{G\text{-}d\text{-sh},0,0}$, then $\mathrm{val}_{\mathrm{E}}(g(x) - x) \geqslant p^{dn}$ if $g \in G_n$. This implies that $\mathrm{val}_p(g(x_n) - x_n) \geqslant 1$ if $g \in G_n$. By Proposition 2.2.1, the image of x_n in $\mathcal{O}_{K_\infty}/I_\infty^c$ belongs to $\mathcal{O}_{K_n}/I_n^c$. Hence the image of x in $\varprojlim_{x\mapsto x^\ell} \mathcal{O}_{K_\infty}/I_\infty^c$ belongs to $\varprojlim_{x\mapsto x^\ell} \mathcal{O}_{K_n}/I_n^c$. By Proposition 2.1.3, x belongs to $X_K^+(K_\infty)$. This proves (1).

Since $\mathrm{val}_{\mathrm{E}}(\varphi(x)) = p \cdot \mathrm{val}_{\mathrm{E}}(x)$, item (2) follows from (1) and Propositions 2.2.2 and 1.2.16. □

Remark 2.2.4. We have $\widetilde{\mathbb{E}}^{d\text{-sh}}_{K_\infty} \subset \widetilde{\mathbb{E}}^{1\text{-sh}}_{K_\infty}$. The field $\widetilde{\mathbb{E}}^{1\text{-sh}}_{K_\infty}$ contains the field of norms $X_K(L_\infty)$ of any p-adic Lie extension L_∞/K contained in K_∞. Indeed, $\widetilde{\mathbb{E}}_{L_\infty} \subset \widetilde{\mathbb{E}}_{K_\infty}$ and if $e = \dim \mathrm{Gal}(L_\infty/K)$, then $X_K(L_\infty) \subset \widetilde{\mathbb{E}}^{e\text{-sh}}_{L_\infty} \subset \widetilde{\mathbb{E}}^{1\text{-sh}}_{K_\infty}$ (see Proposition 1.2.15).

Can one give a description of $\widetilde{\mathbb{E}}^{1\text{-sh}}_{K_\infty}$, for example along the lines of Section 5 of [Berger 2016]?

3. The Lubin–Tate case

We now specialize the constructions of the previous section to the case when K_∞ is generated over K by the torsion points of a Lubin–Tate formal group.

3.1. ***Lubin–Tate formal groups.*** Let K be a finite extension of $\mathbb{Q}_p$ of degree d, with ring of integers $\mathcal{O}_K$, inertia index f, ramification index e, and residue field k. Let $q = p^f = \mathrm{Card}(k)$ and let π be a uniformizer of $\mathcal{O}_K$. Let LT be the Lubin–Tate formal $\mathcal{O}_K$-module attached to π; see [Lubin and Tate 1965]. We choose a coordinate Y on LT. For each $a \in \mathcal{O}_K$ we get a power series $[a](Y) \in \mathcal{O}_K[\![Y]\!]$, that we now see as an element of $k[\![Y]\!]$. In particular, $[\pi](Y) = Y^q$. Let $S(T,U) \in k[\![T,U]\!]$ denote the reduction mod π of the power series giving the addition law in LT in that coordinate. Recall that $S(T,0) = T$ and $S(0,U) = U$.

Lemma 3.1.1. *If $a, b \in \mathcal{O}_K$ and $i \geqslant 0$, then* $\mathrm{val}_Y([a + p^i b](Y) - [a](Y)) \geqslant p^{di}$.
Furthermore, $[1+\pi^i](Y) = Y + Y^{q^i} + \mathrm{O}(Y^{q^i+1})$.

Proof. We have $[\pi](Y) = Y^q$, so $\mathrm{val}_Y([\pi](Y)) \geqslant p^f$. Writing $p = u\pi^e$ for a unit u, we see that $\mathrm{val}_Y([p^i b](Y)) \geqslant p^{di}$ if $b \in \mathcal{O}_K$. If $a, b \in \mathcal{O}_K$ and $i \geqslant 0$, then $[a + bp^i](Y) = S([a](Y), [bp^i](Y))$. We have $S(T,U) = T + U + TU \cdot R(T,U)$, so that $[a+bp^i](Y) - [a](Y) = S([a](Y), [bp^i](Y)) - [a](Y) \in [bp^i](Y) \cdot k[\![Y]\!]$. This implies the first result.

The second claim follows likewise from the fact that $[1+\pi^i](Y) = S(Y, [\pi^i](Y)) = Y + [\pi^i](Y) + Y \cdot [\pi^i](Y) \cdot R(Y, [\pi^i](Y))$. □

Let $\mathbb{E} = k((Y))$. Let $\mathbb{E}_n = k((Y^{1/q^n}))$ and let $\mathbb{E}_\infty = \bigcup_{n \geqslant 0} \mathbb{E}_n$. These fields are endowed with the Y-adic valuation val_Y, and we let $\mathbb{E}_\star^+$ denote the ring of integers of $\mathbb{E}_\star$. The group $\mathcal{O}_K^\times$ acts on $\mathbb{E}_n$ by $a \cdot f(Y^{1/q^n}) = f([a](Y^{1/q^n}))$.

Lemma 3.1.2. *If $j \geqslant 1$ ($j \geqslant 2$ if $p = 2$), then $1 + p^j\mathcal{O}_K$ is uniform, and $(1 + p^j\mathcal{O}_K)_i = 1 + p^{i+j}\mathcal{O}_K$.*

Proof. The map $1 + p^j\mathcal{O}_K \to \mathcal{O}_K$, given by $x \mapsto p^{-j} \cdot \log_p(x-1)$, is an isomorphism of pro-p groups taking $1 + p^{i+j}\mathcal{O}_K$ to $p^i\mathcal{O}_K$. □

Recall that $d = [K : \mathbb{Q}_p]$, that $f = [k : \mathbb{F}_p]$, and that $q = p^f$.

Proposition 3.1.3. *We have* $\mathbb{E}_n^+ = (\mathbb{E}_n^+)^{1+p^j\mathcal{O}_K\text{-}d\text{-}\mathrm{sh},dj-fn,0}$.

Proof. If $b \in \mathcal{O}_K$ and $i, j \geqslant 0$, then by Lemma 3.1.1, we have

$$\mathrm{val}_Y([1 + p^{i+j}b](Y^{1/q^n}) - Y^{1/q^n}) \geqslant 1/q^n \cdot p^{d(i+j)} = p^{dj-fn} \cdot p^{di}.$$

Lemma 3.1.2 then implies that $Y^{1/q^n} \in (\mathbb{E}_n^+)^{1+p^j\mathcal{O}_K\text{-}d\text{-}\mathrm{sh},dj-fn,0}$. The lemma follows from Proposition 1.2.16 and Lemma 1.2.9. □

Corollary 3.1.4. *We have* $\mathbb{E} = \mathbb{E}^{1+p^j\mathcal{O}_K\text{-}d\text{-}\mathrm{sh},dj}$.

Proof. This follows from Proposition 3.1.3 with $n = 0$, and Proposition 1.2.16. □

Proposition 3.1.5. *If $\varepsilon > 0$, then* $k[\![Y]\!]^{1+p^j\mathcal{O}_K\text{-}d\text{-}\mathrm{sh},dj+\varepsilon} \subset k[\![Y^p]\!]$.

Proof. Take $f(Y) \in k[\![Y]\!]$. There is a power series $h(T, U) \in k[\![T, U]\!]$ such that

$$f(T+U) = f(T) + U \cdot f'(T) + U^2 \cdot h(T, U).$$

If $m \geqslant 0$, Lemma 3.1.1 implies that $[1+\pi^m](Y) = Y + Y^{q^m} + \mathrm{O}(Y^{q^m+1})$. Therefore,

$$f([1+\pi^m](Y)) = f(Y) + (Y^{q^m} + \mathrm{O}(Y^{q^m+1})) \cdot f'(Y) + \mathrm{O}(Y^{2q^m}).$$

If $f(Y) \notin k[\![Y^p]\!]$, then $f'(Y) \neq 0$. Let $\mu = \mathrm{val}_Y(f'(Y))$. The above computations imply that $\mathrm{val}_Y(f([1+\pi^{ei+ej}](Y)) - f(Y)) = p^{dj} \cdot p^{di} + \mu$ for $i \gg 0$.

This implies the claim, since $\pi^e \mathcal{O}_K = p\mathcal{O}_K$. □

Corollary 3.1.6. *We have* $\mathbb{E}_\infty^{1+p^j\mathcal{O}_K\text{-}d\text{-sh},dj-fn} = \mathbb{E}_n$.

Proof. We prove that, more generally,

$$\mathbb{E}_\infty^{1+p^j\mathcal{O}_K\text{-}d\text{-sh},dj-\ell} = k((Y^{1/p^\ell})).$$

Take $f(Y^{1/p^m}) \in (\mathbb{E}_\infty^+)^{1+p^j\mathcal{O}_K\text{-}d\text{-sh},dj-\ell}$ where $f(Y) \in k[\![Y]\!]$. Since $\mathrm{val}_Y(h^p) = p \cdot \mathrm{val}_Y(h)$ for all $h \in \widetilde{\mathbb{E}}^+$, we have $f^{p^m}(Y) \in (\mathbb{E}_\infty^+)^{1+p^j\mathcal{O}_K\text{-}d\text{-sh},dj-\ell+m}$, where $f^{p^m}(Y) \in E[\![Y]\!]$ is $f^{p^m}(Y) = f(Y^{1/p^m})^{p^m}$. If $m \geqslant \ell+1$, then Proposition 3.1.5 implies that $f^{p^m}(Y) \in E[\![Y^p]\!]$, so that $f(Y) = g(Y^p)$, and $f(Y^{1/p^m}) = g(Y^{1/p^{m-1}})$. This implies the claim. □

3.2. ***Decompletion of*** $\widetilde{\mathbb{E}}$***.*** Since we use the results of Section 2.2, we once more assume that $p \neq 2$. Let $\widetilde{\mathbb{E}}$ denote the Y-adic completion of $\mathbb{E}_\infty$.

Theorem 3.2.1. *We have* $\widetilde{\mathbb{E}}^{1+p^j\mathcal{O}_K\text{-}d\text{-sh},dj} = \mathbb{E}$, *and* $\widetilde{\mathbb{E}}^{d\text{-sh}} = \mathbb{E}_\infty$.

Proof. Let $K_\infty = K(\mathrm{LT}[\pi^\infty])$ denote the extension of K generated by the torsion points of LT, and let $\Gamma = \mathrm{Gal}(K_\infty/K)$. The Lubin–Tate character χ_π gives rise to an isomorphism $\chi_\pi : \Gamma \to \mathcal{O}_K^\times$. For $n \geqslant 1$, let $K_n = K(\mathrm{LT}[\pi^n])$. If $(\pi_n)_{n \geqslant 1}$ is a compatible sequence of primitive π^n-torsion points of LT, then π_n is a uniformizer of $\mathcal{O}_{K_n}$, $\varpi = (\pi_n)_{n \geqslant 0}$ belongs to $\varprojlim_{\mathrm{N}_{K_n/K_{n-1}}} \mathcal{O}_{K_n}$, and $X_K(K_\infty) = k((\varpi))$ by Theorem 2.1.1. If $g \in \Gamma$, then $g(\varpi) = [\chi_\pi(g)](\varpi)$, so that if we identify Γ and $\mathcal{O}_K^\times$, then $X_K(K_\infty) = \mathbb{E}$ with its action of $\mathcal{O}_K^\times$. Proposition 2.1.4 implies that $\widetilde{\mathbb{E}} = \widetilde{\mathbb{E}}_{K_\infty}$ as valued fields with an action of (an open subgroup of) $\mathcal{O}_K^\times$. We can therefore apply Theorem 2.2.3, and get $(\widetilde{\mathbb{E}}^+)^{d\text{-sh}} = \mathbb{E}_\infty^+$. This implies the second statement. The first one then follows from Corollary 3.1.6. □

Remark 3.2.2. In the above proof, note that $K_\infty^{1+p^n\mathcal{O}_K} = K_{ne}$, so that the numbering is not the same as in Section 2.1.

Remark 3.2.3. We can define Lubin–Tate Γ-modules over $\mathbb{E}$ as in Section 3.2 of [Berger and Rozensztajn 2022]. The results proved in that section carry over to the Lubin–Tate setting without difficulty.

In Theorem 2.9 of [Berger and Rozensztajn 2022], we proved Theorem 3.2.1 above in the cyclotomic case, using Tate traces. There are no such Tate traces in the Lubin–Tate case if $K \neq \mathbb{Q}_p$. We now explain why this is so. More precisely, we prove that there is no Γ-equivariant k-linear projector $\widetilde{\mathbb{E}} \to \mathbb{E}$ if $K \neq \mathbb{Q}_p$.

Choose a coordinate T on LT such that $\log_{\mathrm{LT}}(T) = \sum_{n \geqslant 0} T^{q^n}/\pi^n$, so that $\log'_{\mathrm{LT}}(T) \equiv 1 \bmod \pi$. Let $\partial = 1/\log'_{\mathrm{LT}}(T) \cdot d/dT$ be the invariant derivative on LT. Let $\varphi_q = \varphi^f$ where $q = p^f$.

Lemma 3.2.4. *We have $d\gamma(Y)/dY \equiv \chi_\pi(\gamma)$ in $\mathbb{E}$ for all $\gamma \in \Gamma$.*

Proof. Since $\log'_{\mathrm{LT}} \equiv 1 \bmod \pi$, we have $\partial = d/dY$ in $\mathbb{E}$. Applying $\partial \circ \gamma = \chi_\pi(\gamma)\gamma \circ \partial$ to Y, we get the claim. □

Lemma 3.2.5. *If $\gamma \in \Gamma$ is nontorsion, then $\mathbb{E}^{\gamma=1} = k$.*

Proposition 3.2.6. *If $K \neq \mathbb{Q}_p$, there is no Γ-equivariant map $R : \mathbb{E} \to \mathbb{E}$ such that $R(\varphi_q(f)) = f$ for all $f \in \mathbb{E}$.*

Proof. Suppose that such a map exists, and take $\gamma \in \Gamma$ nontorsion and such that $\chi_\pi(\gamma) \equiv 1 \bmod \pi$. We first show that if $f \in \mathbb{E}$ is such that $(1-\gamma)f \in \varphi_q(\mathbb{E})$, then $f \in \varphi_q(\mathbb{E})$. Write $f = f_0 + \varphi_q(R(f))$ where $f_0 = f - \varphi_q(R(f))$, so that $R(f_0) = 0$ and $(1-\gamma)f_0 = \varphi_q(g) \in \varphi_q(\mathbb{E})$. Applying R, we get $0 = (1-\gamma)R(f_0) = g$. Hence $g = 0$ so that $(1-\gamma)f_0 = 0$. Since $\mathbb{E}^{\gamma=1} = k$ by Lemma 3.2.5, this implies $f_0 \in k$, so that $f \in \varphi_q(\mathbb{E})$.

However, Lemma 3.2.4 and the fact that $\chi_\pi(\gamma) \equiv 1 \bmod \pi$ imply that $\gamma(Y) = Y + f_\gamma(Y^p)$ for some $f_\gamma \in \mathbb{E}$, so that $\gamma(Y^{q/p}) = Y^{q/p} + \varphi_q(g_\gamma)$. Hence $(1-\gamma)(Y^{q/p}) \in \varphi_q(\mathbb{E})$ even though $Y^{q/p}$ does not belong to $\varphi_q(\mathbb{E})$. Therefore, no such map R can exist. □

Corollary 3.2.7. *If $K \neq \mathbb{Q}_p$, there is no Γ-equivariant k-linear projector $\varphi_q^{-1}(\mathbb{E}) \to \mathbb{E}$. A fortiori, there is no Γ-equivariant k-linear projector $\widetilde{\mathbb{E}} \to \mathbb{E}$.*

Proof. Given such a projector Π, we could define R as in Proposition 3.2.6 by $R = \Pi \circ \varphi_q^{-1}$. □

3.3. *The perfectoid commutant of* Aut(LT).

In Section 3.1 of [Berger and Rozensztajn 2022], we computed the perfectoid commutant of $\mathrm{Aut}(\mathbb{G}_{\mathrm{m}})$. We now use Theorem 3.2.1 to do the same for Aut(LT). We still assume that $p \neq 2$.

Theorem 3.3.1. *If $u \in \widetilde{\mathbb{E}}^+$ is such that $\mathrm{val}_Y(u) > 0$ and $u \circ [g] = [g] \circ u$ for all $g \in \mathcal{O}_K^\times$, there exists $b \in \mathcal{O}_K^\times$ and $n \in \mathbb{Z}$ such that $u(Y) = [b](Y^{q^n})$.*

Recall that a power series $f(Y) \in k[\![Y]\!]$ is separable if $f'(Y) \neq 0$. If $f(Y) \in Y \cdot k[\![Y]\!]$, we say that f is invertible if $f'(0) \in k^\times$, which is equivalent to f being invertible for composition (denoted by $\circ$). We say that $w(Y) \in Y \cdot k[\![Y]\!]$ is nontorsion if $w^{\circ n}(Y) \neq Y$ for all $n \geqslant 1$. If $w(Y) = \sum_{i \geqslant 0} w_i Y^i \in k[\![Y]\!]$ and $m \in \mathbb{Z}$, let $w^{(m)}(Y) = \sum_{i \geqslant 0} w_i^{p^m} Y^i$. Note that $(w \circ v)^{(m)} = w^{(m)} \circ v^{(m)}$.

Proposition 3.3.2. *Let $w(Y) \in Y + Y^2 \cdot k[\![Y]\!]$ be a nontorsion series, and let $f(Y) \in Y \cdot k[\![Y]\!]$ be a separable power series. If $w^{(m)} \circ f = f \circ w$ for some $m \in \mathbb{Z}$, then f is invertible.*

Proof. This is a slight generalization of [Lubin 1994, Lemma 6.2]. Write

$$\begin{aligned} f(Y) &= f_n Y^n + \mathrm{O}(Y^{n+1}), \\ f'(Y) &= g_j Y^j + \mathrm{O}(Y^{j+1}), \\ w(Y) &= Y + w_r Y^r + \mathrm{O}(Y^{r+1}), \end{aligned}$$

with $f_n, g_j, w_r \neq 0$. Since w is nontorsion, we can replace w by $w^{\circ p^\ell}$ for $\ell \gg 0$ and assume that $r \geqslant j+1$. We have

$$w^{(m)} \circ f = f(Y) + w_r^{(m)} f(Y)^r + \mathrm{O}(Y^{n(r+1)}) = f(Y) + w_r^{(m)} f_n^r Y^{nr} + \mathrm{O}(Y^{nr+1}).$$

If $j = 0$, then $n = 1$ and we are done, so assume that $j \geqslant 1$. We have

$$\begin{aligned} f \circ w &= f(Y + w_r Y^r + \mathrm{O}(Y^{r+1})) \\ &= f(Y) + w_r Y^r f'(Y) + \mathrm{O}(Y^{2r}) \\ &= f(Y) + w_r g_j Y^{r+j} + \mathrm{O}(Y^{r+j+1}). \end{aligned}$$

This implies that $nr = r + j$, hence $(n-1)r = j$, which is impossible if $r > j$ unless $n = 1$. Hence $n = 1$ and f is invertible. □

Lemma 3.3.3. *If $u \in \widetilde{\mathbb{E}}^+$ is such that* $\mathrm{val}_X(u) > 0$ *and $u \circ [g] = [g] \circ u$ for all $g \in \mathcal{O}_K^\times$, then $u \in (\widetilde{\mathbb{E}}^+)^{d\text{-sh}}$.*

Proof. The group $\mathcal{O}_K^\times$ acts on $\widetilde{\mathbb{E}}^+$ by $g \cdot u = u \circ [g]$. By lemmas 3.1.1 and 3.1.2, the function $g \mapsto [g] \circ u$ is in $\mathcal{H}_d^\lambda(1 + p\mathcal{O}_K, \widetilde{\mathbb{E}}^+)$, where $p^\lambda = \mathrm{val}_Y(u)$. □

Proof of Theorem 3.3.1. Take $u \in \widetilde{\mathbb{E}}$ such that $\mathrm{val}_Y(u) > 0$ and $u \circ [g] = [g] \circ u$ for all $g \in \mathcal{O}_K^\times$. By Lemma 3.3.3 and Theorem 3.2.1, there is an $m \in \mathbb{Z}$ such that $f(Y) = u(Y)^{p^m}$ belongs to $Y \cdot k[\![Y]\!]$ and is separable. Take $g \in 1 + \pi\mathcal{O}_K$ such that g is nontorsion, and let $w(Y) = [g](Y)$ so that $u \circ w = w \circ u$. We have $f \circ w = w^{(m)} \circ f$. By Proposition 3.3.2, f is invertible. In addition, $f \circ w = w^{(m)} \circ f$ if $w(Y) = [g](Y)$ for all $g \in \mathcal{O}_K^\times$. Hence $f_0 \cdot \bar{g} = \bar{g}^{p^m} \cdot f_0$, so that $a^{p^m} = a$ for all $a = \bar{g} \in k$. This implies that $\mathbb{F}_q \subset \mathbb{F}_{p^{|m|}}$, so that $m = fn$ for some $n \in \mathbb{Z}$. Hence $w^{(m)} = w$, and $f \circ [g] = [g] \circ f$ for all $g \in \mathcal{O}_K^\times$. Theorem 6 of [Lubin and Sarkis 2007] implies that $f \in \mathrm{Aut}(\mathrm{LT})$. Hence there exists $b \in \mathcal{O}_K^\times$ such that $u(Y) = [b](Y^{q^n})$. □

4. Mahler expansions and super-Hölder functions

In Section 1.3 of [Berger and Rozensztajn 2022], we proved an analogue of Mahler's theorem for continuous functions $\mathbb{Z}_p \to M$, and then gave a characterization of super-Hölder functions in terms of their Mahler expansions. We now indicate how these results generalize to functions $G \to M$ for a uniform pro-p group G. Given the definition of super-Hölder functions and the existence of a coordinate $c : G \to \mathbb{Z}_p^d$ as in Proposition 1.1.1, it is enough to study functions $\mathbb{Z}_p^d \to M$. We generalize the setting a little bit, and study functions $\mathcal{O}_K^d \to M$ where K is a finite extension of $\mathbb{Q}_p$. Let K be such a field, fix a uniformizer π of $\mathcal{O}_K$ and let k be the residue field of K. Let $q = \mathrm{Card}(k)$.

4.1. ***Good bases and wavelets.*** Let $X = \mathcal{O}_K^d$, which we endow with the valuation $\mathrm{val}_X(x_1, \dots, x_d) = \min_i \mathrm{val}_\pi(x_i)$. For $n \geqslant 0$, let $X_n = \pi^n X = \{x \in X, \mathrm{val}_X(x) \geqslant n\}$.

We endow X with the val_X-adic topology. For any set Y, we denote by $\mathrm{LC}(X, Y)$ the set of locally constant functions $X \to Y$. For $n \geqslant 0$ we denote by $\mathrm{LC}_n(X, Y)$ the subset of elements of $\mathrm{LC}(X, Y)$

that factor through X/X_n. Let $I = \bigcup_{n \geqslant 0} I_n$ be a set of indices, where $I_n \subset I_{n+1}$ for all $n \geqslant 0$, and $\mathrm{Card}(I_n) = \mathrm{Card}(X/X_n) = q^{nd}$. Let E be a field of characteristic p.

Definition 4.1.1. A family $\{h_i\}_{i \in I}$ is a good basis of $\mathrm{LC}(X, E)$ if it is a basis of the E-vector space $\mathrm{LC}(X, E)$ such that for all $n \geqslant 0$, $\{h_i\}_{i \in I_n}$ is a basis of $\mathrm{LC}_n(X, E)$.

Let M be (as usual) an E-vector space with a valuation val_M, such that $\mathrm{val}_M(ax) = \mathrm{val}_M(x)$ for all $a \in E^\times$ and $x \in M$. We assume that M is separated and complete for the val_M-adic topology.

Proposition 4.1.2. *Every $f \in \mathrm{LC}_n(X, M)$ can be written uniquely as $\sum_{i \in I_n} h_i \cdot m_i$ for some elements $m_i \in M$. Moreover*, $\inf_{x \in X} \mathrm{val}_M(f(x)) = \inf_{i \in I_n} \mathrm{val}_M(m_i)$.

Proof. Let $\{\delta_x\}_{x \in X/X_n}$ be the basis of $\mathrm{LC}_n(X, E)$ defined as follows: δ_x is the characteristic function of $x + X_n$. Then $f \in \mathrm{LC}_n(X, M)$ is equal to $\sum_{x \in X/X_n} \delta_x \cdot f(x)$.

As $\{h_i\}_{i \in I_n}$ is also a basis of $\mathrm{LC}_n(X, E)$, we can write $\delta_x = \sum_{i \in I_n} a_{i,x} h_i$ for some elements $a_{i,x} \in E$. We now have $f = \sum_{i \in I_n} h_i \cdot m_i$ where $m_i = \sum_{x \in X/X_n} a_{i,x} f(x)$. This formula implies that $\inf_{i \in I_n} \mathrm{val}_M(m_i) \geqslant \inf_{x \in X} \mathrm{val}_M(f(x))$.

On the other hand we can also write $h_i = \sum_{x \in X/X_n} b_{x,i} \delta_x$ for some elements $b_{x,i} \in E$, so that $f(x) = \sum_{i \in I_n} b_{x,i} m_i$. This implies that $\inf_{i \in I_n} \mathrm{val}_M(m_i) \leqslant \inf_{x \in X} \mathrm{val}_M(f(x))$. □

We now give an example of a particularly nice good basis of $\mathrm{LC}(X, E)$, the basis of wavelets; see Section I.3 of [Colmez 2010] and Section 2.1 of [de Shalit 2016]. Let $\mathcal{T}$ be a set of representatives of X/X_1 in X, chosen so that the representative of 0 is 0. For each $n \geqslant 0$, let $\mathcal{R}_n$ be the set of representatives of X/X_n defined as follows: $\mathcal{R}_0 = \{0\}$, and for $n \geqslant 1$, $\mathcal{R}_n = \left\{\sum_{i=0}^{n-1} \pi^i x_i,\ x_i \in \mathcal{T} \text{ for all } i\right\}$. We have $\mathcal{R}_1 = \mathcal{T}$, and $\mathcal{R}_n \subset \mathcal{R}_{n+1}$ for all n. Let $\mathcal{R} = \bigcup_{n \geqslant 0} \mathcal{R}_n$. If $r \in \mathcal{R}$ let $\ell(r)$ be the smallest n such that $r \in \mathcal{R}_n$. For $r \in \mathcal{R}$, let χ_r be the characteristic function of the closed disc $r + X_{\ell(r)} = \{x \in X, \mathrm{val}_X(x - r) \geqslant \ell(r)\}$.

Proposition 4.1.3. *The set $\{\chi_r\}_{r \in \mathcal{R}}$ is a good basis of* $\mathrm{LC}(X, E)$.

Proof. We prove that for all $n \geqslant 0$, the set $\{\chi_r\}_{r \in \mathcal{R}_n}$ is a basis of $\mathrm{LC}_n(X, E)$. Consider the basis $\{\delta_r\}_{r \in \mathcal{R}_n}$ of $\mathrm{LC}_n(X, E)$, where δ_r is the characteristic function of $r + X_n$. We have

$$\chi_r = \sum_{r' \in \mathcal{R}_{n-\ell(r)}} \delta_{r + \pi^{\ell(r)} r'}.$$

This implies that if we write $\mathcal{R}_n = (\mathcal{R}_n \setminus \mathcal{R}_{n-1}) \sqcup \cdots \sqcup (\mathcal{R}_1 \setminus \mathcal{R}_0) \sqcup \mathcal{R}_0$ and we express the family $\{\chi_r\}_{r \in \mathcal{R}_n}$ in terms of the basis $\{\delta_r\}_{r \in \mathcal{R}_n}$, we get a unipotent matrix. This shows that $\{\chi_r\}_{r \in \mathcal{R}_n}$ is also a basis of $\mathrm{LC}_n(X, E)$. □

4.2. ***Expansions of continuous functions.*** We show that every continuous function $X \to M$ has a convergent expansion along a good basis of X, and prove some continuity estimates in terms of the coefficients of the expansion. If $\{m_i\}_{i \in I}$ is a family of M, we say that $m_i \to 0$ if $\inf_{i \notin I_n} \mathrm{val}_M(m_i) \to +\infty$ as $n \to +\infty$.

Theorem 4.2.1. *Let $\{h_i\}_{i\in I}$ be a good basis of* $\mathrm{LC}(X, E)$.

If $\{m_i\}_{i\in I}$ is a family of M such that $m_i \to 0$, the function $f : X \to M$ given by $f = \sum_{i\in I} h_i \cdot m_i$ belongs to $C^0(X, M)$, and $\inf_{x\in X} \mathrm{val}_M(f(x)) = \inf_{i\in I} \mathrm{val}_M(m_i)$.

Conversely, if $f \in C^0(X, M)$, there exists a unique family $\{m_i(f)\}_{i\in I}$ of elements of M such that $m_i(f) \to 0$ and such that $f = \sum_{i\in I} h_i \cdot m_i(f)$.

Proof. Let $\{m_i\}_{i\in I}$ be a family of M such that $m_i \to 0$. If $f_n = \sum_{i\in I_n} h_i \cdot m_i$, then $f_n \in C^0(X, M)$, and f is the uniform limit of the f_n. We have $\inf_X \mathrm{val}_M(f_n(x)) = \inf_{i\in I_n} \mathrm{val}_M(m_i)$ by Proposition 4.1.2. Since $m_i \to 0$, we have $\inf_{i\in I} \mathrm{val}_M(m_i) = \inf_{i\in I_n} \mathrm{val}_M(m_i)$ for $n \gg 0$. Hence $\inf_X \mathrm{val}_M(f_n(x)) = \inf_{i\in I} \mathrm{val}_M(m_i)$ for $n \gg 0$. Since $\inf_{x\in X} \mathrm{val}_M(f(x)) = \lim_n \inf_x \mathrm{val}_M(f_n(x))$, we have $\inf_{x\in X} \mathrm{val}_M(f(x)) = \inf_{i\in I} \mathrm{val}_M(m_i)$.

We now prove the converse. Let $M_n = \{m \in M, \mathrm{val}_M(m) \geqslant n\}$, let $\pi_n : M \to M/M_n$ be the projection, and for each n, fix a lift $\psi_n : M/M_n \to M$. Take $f \in C^0(X, M)$, and let $f_n = \psi_n \circ \pi_n \circ f$. As f and f_n coincide modulo M_n, f is the uniform limit of the f_n. On the other hand, $\pi_n \circ f$ is locally constant, and therefore so is f_n. As X is compact, there exists some $k(n) \geqslant 0$ such that $f_n \in \mathrm{LC}_{k(n)}(X, M)$. By Proposition 4.1.2, we can write $f_n = \sum_{i\in I} h_i \cdot m_{i,n}$, where $m_{i,n} = 0$ if $i \notin I_{k(n)}$. We have $\mathrm{val}_M(m_{i,n} - m_{i,n'}) \geqslant \min(n, n')$ by construction, so that for each i, the sequence $\{m_{i,n}\}_n$ converges to some $m_i \in M$. Moreover, if $i \notin I_{k(n)}$, then $\mathrm{val}_M(m_i) \geqslant n$, so that $m_i \to 0$. The continuous function $\sum_{i\in I} h_i \cdot m_i$ is the uniform limit of the f_n, so that finally $f = \sum_{i\in I} h_i \cdot m_i$. □

Proposition 4.2.2. *Take $f \in C^0(X, M)$ and $t \in \mathbb{Z}_{\geqslant 0}$. If $\{h_i\}_{i\in I}$ is a good basis of* $\mathrm{LC}(X, E)$, *and we write $f = \sum_i h_i \cdot m_i$ with $m_i \to 0$, then $\inf_{i\notin I_t} \mathrm{val}_M(m_i)$ depends only on f and not on the choice of the good basis.*

Proof. Fix two good bases $\{h_i\}_{i\in I}$ and $\{h'_i\}_{i\in I}$ of $\mathrm{LC}(X, E)$. There exists a family $\{\lambda_{i,j}\}_{(i,j)\in I\times I}$ of elements of E such that $h_i = \sum_j \lambda_{i,j} h'_j$ for all i. Moreover, if $i \in I_c$ then $\lambda_{i,j} = 0$ for all $j \notin I_c$. Now write $f = \sum_{i\in I} h_i \cdot m_i(f) = \sum_{i\in I} h'_i \cdot m'_i(f)$. We also have

$$f = \sum_i \Big(\sum_j \lambda_{i,j} h'_j\Big) \cdot m_i(f) = \sum_j h'_j \cdot \Big(\sum_i \lambda_{i,j} m_i(f)\Big),$$

so that $m'_j(f) = \sum_i \lambda_{i,j} m_i(f)$. If $j \notin I_t$, then $m'_j(f) = \sum_{i\notin I_t} \lambda_{i,j} m_i(f)$, as $\lambda_{i,j} = 0$ if $i \in I_t$ and $j \notin I_t$. This implies that $\inf_{j\notin I_t} \mathrm{val}_M(m'_j(f)) \geqslant \inf_{i\notin I_t} \mathrm{val}_M(m_i(f))$.

By symmetry, we get that $\inf_{j\notin I_t} \mathrm{val}_M(m'_j(f)) = \inf_{i\notin I_t} \mathrm{val}_M(m_i(f))$. □

Theorem 4.2.3. *Take $f \in C^0(X, M)$ and $t \in \mathbb{Z}_{\geqslant 0}$.*

If $\{h_i\}_{i\in I}$ is a good basis of $\mathrm{LC}(X, E)$, *and we write $f = \sum_i h_i \cdot m_i$ with $m_i \to 0$, then*

$$\inf_{i\notin I_t} \mathrm{val}_M(m_i) = \inf_{\substack{x,y\in X \\ \mathrm{val}_X(x-y)\geqslant t}} \mathrm{val}_M(f(x) - f(y)).$$

Proof. Let $C_t(f) = \inf_{x,y\in X, \mathrm{val}_X(x-y)\geqslant t} \mathrm{val}_M(f(x) - f(y))$ and $B_t(f) = \inf_{i\notin I_t} \mathrm{val}_M(m_i)$.

If $x \in X$ and $z \in X_t$, then $f(x+z) - f(x) = \sum_{i \in I} (h_i(x+z) - h_i(z)) \cdot m_i(f)$. As $h_i \in \mathrm{LC}_t(X, E)$ for $i \in I_t$, the above equality gives us

$$f(x+z) - f(x) = \sum_{i \notin I_t} (h_i(x+z) - h_i(z)) \cdot m_i(f).$$

This implies that $C_t(f) \geqslant B_t(f)$.

We now prove the converse inequality. By Proposition 4.2.2, $B_t(f)$ is independent of the choice of a good basis, and we choose the wavelet basis of Proposition 4.1.3. Write $f = \sum_{r \in \mathcal{R}} \chi_r \cdot m_r(f)$, so that we want to show that $\mathrm{val}_M(m_r(f)) \geqslant C_t(f)$ for all $r \notin \mathcal{R}_t$. If $x \in X$, define $g_x : X \to M$ by $g_x(z) = f(x + \pi^t z) - f(x)$, and write $g_x = \sum_{r \in \mathcal{R}} \chi_r \cdot m_r(g_x)$. For each $r \in \mathcal{R}$, we can write uniquely $r = r_t + \pi^t s$ with $r_t \in \mathcal{R}_t$, where $s = 0$ if $r \in \mathcal{R}_t$, and $s \neq 0 \in \mathcal{R}_{\ell(r)-t}$ if $r \notin \mathcal{R}_t$. For $x \in \mathcal{R}_t$ and $r \notin \mathcal{R}_t$, the map $z \mapsto \chi_r(x + \pi^t z) - \chi_r(x)$ is the zero function if $r_t \neq x$, and is χ_s if $r_t = x$. This implies that if $x \in \mathcal{R}_t$, then

$$\begin{aligned} g_x(z) &= \sum_{r \in \mathcal{R}} (\chi_r(x + \pi^t z) - \chi_r(x)) \cdot m_r(f) \\ &= \sum_{r \notin \mathcal{R}_t} (\chi_r(x + \pi^t z) - \chi_r(x)) \cdot m_r(f) \\ &= \sum_{s \notin \mathcal{R}_0} \chi_s(z) \cdot m_{x + \pi^t s}(f). \end{aligned}$$

Therefore if $x \in \mathcal{R}_t$, then $m_0(g_x) = 0$ and $m_s(g_x) = m_{x+\pi^t s}(f)$ if $s \neq 0$. We have $\inf_{s \in \mathcal{R}} \mathrm{val}_M(m_s(g_x)) = \inf_{z \in X} \mathrm{val}_M(g_x(z)) \geqslant C_t(f)$, so that $\mathrm{val}_M(m_s(g_x)) \geqslant C_t(f)$ for all $x \in X$ and $s \in \mathcal{R}$. This implies that for all $x \in \mathcal{R}_t$ and $s \neq 0$, $\mathrm{val}_M(m_{x+\pi^t s}(f)) \geqslant C_t(f)$. Hence for all $r \notin \mathcal{R}_t$, we have $\mathrm{val}_M(m_r(f)) \geqslant C_t(f)$. □

4.3. ***Mahler bases.*** We now construct some other examples of good bases. For $n \geqslant 0$, let $\mathrm{Int}_n(\mathcal{O}_K)$ denote the set of polynomials $f(T) \in K[T]$ such that $\deg(P) \leqslant n$ and $f(\mathcal{O}_K) \subset \mathcal{O}_K$. Recall (see for instance Section 1.2 of [de Shalit 2016]) that a Mahler basis for $\mathcal{O}_K$ is a sequence $\{h_n\}_{n \geqslant 0}$ with $h_n(T) \in K[T]$ of degree n, and such that $\{h_0, \dots, h_n\}$ is a basis of the free $\mathcal{O}_K$-module $\mathrm{Int}_n(\mathcal{O}_K)$ for all $n \geqslant 0$. For example, if $K = \mathbb{Q}_p$, we can take $h_n(T) = \binom{T}{n}$. Let $\{h_n\}_{n \geqslant 0}$ be a Mahler basis for $\mathcal{O}_K$. Each h_n defines a function $\mathcal{O}_K \to \mathcal{O}_K$ and hence $\mathcal{O}_K \to k$. Let $I = \mathbb{Z}_{\geqslant 0}$ and let $I_n = \{0, \dots, q^n - 1\}$ for $n \geqslant 0$.

Proposition 4.3.1. *If $\{h_n\}_{n \geqslant 0}$ is a Mahler basis for $\mathcal{O}_K$, then $\{h_i\}_{i \in I}$ is a good basis of* $\mathrm{LC}(\mathcal{O}_K, k)$.

Proof. By Theorem 1.2 of [de Shalit 2016], $\{h_0, \dots, h_{q^m - 1}\}$ is a basis of the k-vector space $\mathrm{LC}_m(\mathcal{O}_K, k)$ for all $m \geqslant 0$. This implies the claim. □

We now specialize to $K = \mathbb{Q}_p$. Write $\mathbb{N}$ for $\mathbb{Z}_{\geqslant 0}$ and $\boldsymbol{n}$ for an element $(n_1, \dots, n_d) \in \mathbb{N}^d$. For each $\boldsymbol{n} \in \mathbb{N}^d$, we denote by $h_{\boldsymbol{n}}$ the function $\mathbb{Z}_p^d \to E$ given by $(x_1, \dots, x_d) \mapsto \binom{x_1}{n_1} \cdots \binom{x_d}{n_d}$. For $m \in \mathbb{Z}_{\geqslant 0}$, let $I_m = \{\boldsymbol{n} \in \mathbb{N}^d \text{ such that } \max(n_1, \dots, n_d) \leqslant p^m - 1\}$.

Proposition 4.3.2. *The functions $\{h_{\boldsymbol{n}}\}_{\boldsymbol{n} \in \mathbb{N}^d}$ form a good basis of* $\mathrm{LC}(\mathbb{Z}_p^d, \mathbb{F}_p)$.

Proof. The claim follows from Proposition 4.3.1 for $K = \mathbb{Q}_p$, and Lemma 4.3.3 below. □

Lemma 4.3.3. *If X and X' are as in Section 4.1, and $\{h_i\}_{i\in I}$ and $\{h'_j\}_{j\in J}$ are good bases of* $\mathrm{LC}(X, E)$ *and* $\mathrm{LC}(X', E)$, *then $\{h_i \otimes h'_j\}_{(i,j)\in I\times J}$ is a good basis of* $\mathrm{LC}(X \times X', E)$, *with $(I \times J)_n = I_n \times J_n$.*

Let G be a uniform pro-p group, and let $c : G \to \mathbb{Z}_p^d$ be a coordinate as in Proposition 1.1.1. The theorem below follows from Proposition 4.3.2 and Theorems 4.2.1 and 4.2.3.

Theorem 4.3.4. *If $\{m_{\boldsymbol{n}}\}_{\boldsymbol{n}\in\mathbb{N}^d}$ is a sequence of M such that $m_{\boldsymbol{n}} \to 0$, the function $f : G \to M$ given by $f(g) = \sum_{\boldsymbol{n}\in\mathbb{N}^d} \binom{c_1(g)}{n_1} \cdots \binom{c_d(g)}{n_d} m_{\boldsymbol{n}}$ belongs to $C^0(G, M)$. We have* $\inf_{g\in G} \mathrm{val}_M(f(g)) = \inf_{\boldsymbol{n}\in\mathbb{N}^d} \mathrm{val}_M(m_{\boldsymbol{n}})$.

Conversely, if $f \in C^0(G, M)$, there exists a unique sequence $\{m_{\boldsymbol{n}}(f)\}_{\boldsymbol{n}\in\mathbb{N}^d}$ such that $m_{\boldsymbol{n}}(f) \to 0$ and such that $f(g) = \sum_{\boldsymbol{n}\in\mathbb{N}^d} \binom{c_1(g)}{n_1} \cdots \binom{c_d(g)}{n_d} m_{\boldsymbol{n}}(f)$.

We have $f \in \mathcal{H}_e^{\lambda,\mu}(G, M)$ if and only if for all $i \geqslant 0$, we have $\mathrm{val}_M(m_{\boldsymbol{n}}(f)) \geqslant p^\lambda \cdot p^{ei} + \mu$ *whenever* $\max(n_1, \ldots, n_d) \geqslant p^i$.

Remark 4.3.5. The first two assertions in the above theorem also follow from Theorem 1.2.4 in Section III of [Lazard 1965] (we thank Konstantin Ardakov for pointing this out).

We finish by considering the case $G = \mathcal{O}_K$ for K a finite extension of $\mathbb{Q}_p$, and working with a Mahler basis for $\mathcal{O}_K$. Let K be a finite extension of $\mathbb{Q}_p$ as before. Assume that E is an extension of k. Let $\{h_n\}_{n\geqslant 0}$ be a Mahler basis for $\mathcal{O}_K$. If $f \in C^0(\mathcal{O}_K, M)$, write $f = \sum_{n\geqslant 0} h_n m_n(f)$ with $m_n(f) \to 0$. Let e denote the ramification index of K.

Proposition 4.3.6. *If $f = \sum_{n\geqslant 0} h_n m_n(f)$ as above, then $f \in \mathcal{H}_t^{\lambda,\mu}(\mathcal{O}_K, M)$ if and only if* $\mathrm{val}_M(m_n(f)) \geqslant p^\lambda \cdot p^{ti} + \mu$ *whenever $n \geqslant p^{di}$.*

Proof. This follows from Theorem 4.2.3, since $\mathrm{val}_p(x - y) \geqslant i$ if and only if $\mathrm{val}_\pi(x - y) \geqslant ei$, and since $q^e = p^d$. □

In this situation we can also define a slightly different version of super-Hölder functions. We say that a function $f : \mathcal{O}_K \to M$ is in $\mathcal{H}_{K,t}^{\lambda,\mu}(\mathcal{O}_K, M)$ if $\mathrm{val}_M(f(x) - f(y)) \geqslant p^\lambda \cdot p^{ti} + \mu$ whenever $\mathrm{val}_\pi(x - y) \geqslant i$. We then have

$$\mathcal{H}_{te}^{\lambda+t(e-1),\mu}(\mathcal{O}_K, M) \subset \mathcal{H}_{K,t}^{\lambda,\mu}(\mathcal{O}_K, M) \subset \mathcal{H}_{te}^{\lambda,\mu}(\mathcal{O}_K, M).$$

In particular, $\mathcal{H}_{K,t}(\mathcal{O}_K, M) = \mathcal{H}_{te}(\mathcal{O}_K, M)$. If $K/\mathbb{Q}_p$ is unramified then $\mathcal{H}_{K,t}^{\lambda,\mu}(\mathcal{O}_K, M) = \mathcal{H}_t^{\lambda,\mu}(\mathcal{O}_K, M)$. Moreover we have the following criterion:

Proposition 4.3.7. *If $f = \sum_{n\geqslant 0} h_n m_n(f)$ as above, then $f \in \mathcal{H}_{K,t}^{\lambda,\mu}(\mathcal{O}_K, M)$ if and only if* $\mathrm{val}_M(m_n(f)) \geqslant p^\lambda \cdot p^{ti} + \mu$ *whenever $n \geqslant q^i$.*

Example 4.3.8. For all $n \geqslant 0$, there exists $c_n(T) \in \mathrm{Int}_n(\mathcal{O}_K)$ such that $[a](Y) = \sum_{n\geqslant 0} c_n(a) Y^n$. This implies that $\mathrm{val}_Y(m_n(a \mapsto [a](Y))) \geqslant n$, so that the function $a \mapsto [a](Y)$ is in $\mathcal{H}_d^{0,0}(\mathcal{O}_K, E[\![Y]\!])$, and in $\mathcal{H}_{K,f}^{0,0}(\mathcal{O}_K, E[\![Y]\!])$ where $q = p^f$.

References

[Berger 2016] L. Berger, “Multivariable (φ, Γ)-modules and locally analytic vectors”, *Duke Math. J.* **165**:18 (2016), 3567–3595. MR Zbl

[Berger and Rozensztajn 2022] L. Berger and S. Rozensztajn, “Decompletion of cyclotomic perfectoid fields in positive characteristic”, *Ann. H. Lebesgue* **5** (2022), 1261–1276. MR Zbl

[Cais and Davis 2015] B. Cais and C. Davis, “Canonical Cohen rings for norm fields”, *Int. Math. Res. Not.* **2015**:14 (2015), 5473–5517. MR Zbl

[Colmez 2010] P. Colmez, “Fonctions d’une variable p-adique”, pp. 13–59 in *Représentations p-adiques de groupes p-adiques, II: Représentations de* $\mathrm{GL}_2(\mathbb{Q}_p)$ *et* (φ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Soc. Math. France, Paris, 2010. MR Zbl

[Dixon et al. 1991] J. D. Dixon, M. P. F. du Sautoy, A. Mann, and D. Segal, *Analytic pro-p groups*, Lond. Math. Soc. Lect. Note Ser. **157**, Cambridge Univ. Press, 1991. MR Zbl

[Emerton 2017] M. Emerton, *Locally analytic vectors in representations of locally p-adic analytic groups*, Mem. Amer. Math. Soc. **1175**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Klopsch 2005] B. Klopsch, “On the Lie theory of p-adic analytic groups”, *Math. Z.* **249**:4 (2005), 713–730. MR Zbl

[Lazard 1965] M. Lazard, “Groupes analytiques p-adiques”, *Inst. Hautes Études Sci. Publ. Math.* **26** (1965), 389–603. MR Zbl

[Le Borgne 2010] J. Le Borgne, “Optimisation du théorème d’Ax–Sen–Tate et application à un calcul de cohomologie galoisienne p-adique”, *Ann. Inst. Fourier* (*Grenoble*) **60**:3 (2010), 1105–1123. MR Zbl

[Lubin 1994] J. Lubin, “Non-Archimedean dynamical systems”, *Compos. Math.* **94**:3 (1994), 321–346. MR Zbl

[Lubin and Sarkis 2007] J. D. Lubin and G. Y. Sarkis, “Extrinsic properties of automorphism groups of formal groups”, *J. Algebra* **315**:2 (2007), 874–884. MR Zbl

[Lubin and Tate 1965] J. Lubin and J. Tate, “Formal complex multiplication in local fields”, *Ann. of Math.* (2) **81** (1965), 380–387. MR Zbl

[Schneider 2011] P. Schneider, *p-adic Lie groups*, Grundl. Math. Wissen. **344**, Springer, 2011. MR Zbl

[Scholze 2012] P. Scholze, “Perfectoid spaces”, *Publ. Math. Inst. Hautes Études Sci.* **116** (2012), 245–313. MR Zbl

[Sen 1972] S. Sen, “Ramification in p-adic Lie extensions”, *Invent. Math.* **17** (1972), 44–50. MR Zbl

[de Shalit 2016] E. de Shalit, “Mahler bases and elementary p-adic analysis”, *J. Théor. Nombres Bordeaux* **28**:3 (2016), 597–620. MR Zbl

[Wintenberger 1983] J.-P. Wintenberger, “Le corps des normes de certaines extensions infinies de corps locaux: applications”, *Ann. Sci. École Norm. Sup.* (4) **16**:1 (1983), 59–89. MR Zbl

Communicated by Wee Teck Gan
Received 2023-03-16 Accepted 2024-02-13

laurent.berger@ens-lyon.fr *UMPA, ENS de Lyon, UMR 5669 du CNRS, Lyon, France*

sandra.rozensztajn@ens-lyon.fr *UMPA, ENS de Lyon, UMR 5669 du CNRS, Lyon, France*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 19 No. 1 2025